
“Departamento de Estadística”

CONTRIBUCIONES AL META-ANÁLISIS DE PRUEBAS
DIAGNÓSTICAS EN ENFERMEDADES DE BAJA
PREVALENCIA

T E S I S

MEMORIA QUE, PARA OPTAR AL GRADO DE:
Doctor en Estadística Multivariante Aplicada

PRESENTA:

Johny Javier Pambabay Calero

DIRECTORAS:

Dra. M^a Purificación Galindo Villardón

Dra. Ana Belén Nieto Librero

Salamanca, España, 2019

A **DIOS**,
a mi familia
y a mis amig@s.

Reconocimientos

Quiero agradecer en primer lugar a las autoridades de la Escuela Superior Politécnica del Litoral, Ph.D. Cecilia Paredes, Ph.D. Paúl Herrera, Mgter. Marcos Mendoza, MSc Oswaldo Valle y Ph.D. Ramón Espinel que han hecho posible la realización del trabajo presentado en esta memoria de tesis por su ayuda brindada.

Muy especialmente a mis directoras de tesis, Doctoras, Puri Galindo y Ana Nieto, por la acertada orientación, el soporte y discusión crítica que me permitió un buen aprovechamiento en el trabajo realizado, y que esta tesis llegara a buen término. Asimismo, agradezco a Ana Belén Sánchez por su apoyo personal y humano, ya que ha sido un motor fundamental para el presente trabajo.

Un trabajo de investigación es siempre fruto de ideas, proyectos y esfuerzos previos que corresponden a otras personas. En este caso mi más sincero agradecimiento a los profesores y colaboradores del Departamento de Estadística de la USAL por sus enseñanzas, tiempo y colaboración prestada para la realización de esta investigación.

Pero un trabajo de investigación es también fruto del reconocimiento y del apoyo vital que ofrecen las personas que nos estiman, sin el cual no tendría la fuerza y energía que me animan a crecer como persona y profesional.

Gracias a mis padres, por sus sabios consejos y valores, los cuales me ha permitido ser una persona capaz y útil a la sociedad y especialmente por haber estado a mi lado en todo momento. A mis hermanas, porque con ellas compartí una infancia feliz, que guardo en el recuerdo y es un aliento para seguir adelante.

Pero, sobre todo, gracias a mi esposa, Karin y a mi hijo, Matías por su paciencia, comprensión y solidaridad con este proyecto, por el tiempo que me han concedido, un tiempo “robado” a la historia familiar. Sin su apoyo este trabajo nunca se habría escrito y, por eso, este trabajo es también el suyo.



VNiVERSiDAD
D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

Dra. M^a Purificación Galindo Villardón

Dra. Ana Belén Nieto Librero

Profesoras del Departamento de Estadística de la Universidad de Salamanca

CERTIFICAN

Que D. Johny Javier Pambabay Calero, Ingeniero en Estadística Informática, ha realizado en el Departamento de Estadística de la Universidad de Salamanca, bajo su dirección, el trabajo que para optar al Grado de Doctor, presenta con el título: **Contribuciones al Meta-análisis de pruebas diagnósticas en enfermedades de baja prevalencia**; y para que conste, firma el presente certificado en Salamanca, a 26 de Noviembre de 2019.

Fdo: M^a Purificación Galindo Villardón

Fdo: Ana Belén Nieto Librero

Resumen

La capacidad discriminante de una prueba se expresa comúnmente en términos de sensibilidad y especificidad, y por lo general, existe una relación de compromiso entre estas dos medidas, ya que un umbral creciente para definir la positividad de la prueba provoca una disminución de la sensibilidad y un aumento de la especificidad. Los métodos recomendados para el meta-análisis de pruebas diagnósticas, como el modelo bivalente, se centra en la estimación de una sensibilidad y especificidad resumen en un umbral común, mientras que el modelo HSROC se centran en la estimación de una curva resumen a partir de estudios que han utilizado umbrales diferentes. Sin embargo, estos modelos no informan la media general, sino más bien la media de un estudio central y tienen diferencias en la estimación de la correlación entre la sensibilidad y la especificidad cuando el número de estudios en el meta-análisis es pequeño y/o cuando la varianza entre los estudios es relativamente grande. Para solucionar los problemas antes mencionados se considera el uso de estructuras que pueden ser modeladas usando funciones de cópula como una alternativa al modelado de dependencia.

En este trabajo, se presentan los pasos a seguir para llevar a cabo un meta-análisis de pruebas diagnósticas en situaciones de baja prevalencia, mediante un novedoso esquema de decisión que se fundamenta en el análisis y estudio de la modelización jerárquica y cópulas. Con el modelo HSROC se simuló meta-análisis de enfermedades de baja prevalencia con diferentes puntos de corte y, a partir de las medidas simuladas, se construyeron predictores y una variable respuesta, con el fin de modelizar un aprendizaje automático que sugiera el mejor modelo estadístico a la hora de resumir los resultados de un meta-análisis de pruebas diagnósticas.

El problema de la multicolinealidad fue abordado, para ello se identificó los predictores inflactados, mediante la propuesta de un algoritmo que calcula la matriz de correlación de los predictores e identifica todas las correlaciones por pares que estén por debajo de un umbral.

Finalmente, dentro del análisis expuesto, se propone un novedoso diagrama de flujo en conjunto con un algoritmo, que le permita al investigador construir un modelo estadístico con algoritmos de aprendizaje automático, con la finalidad de obtener resultados confiables tanto en la predicción como en clasificación.

Índice general

Índice de figuras	XV
Índice de tablas	XIX
1. Introducción	1
1.1. Objetivos	4
1.1.1. Objetivo General	4
1.1.2. Objetivos Específicos	5
1.2. Motivación	5
1.3. Planteamiento del problema	6
1.4. Metodología	7
1.4.1. Métodos Estadísticos	7
1.4.2. Evaluación de la heterogeneidad	7
1.4.3. Efecto umbral	7
1.4.4. Modelos bivariantes, jerárquicos	8
1.5. Contribuciones	9
1.6. Estructura de la tesis	9
2. Meta-Análisis	11
2.1. Definición	11
2.2. Orígenes y Antecedentes	11
2.3. Meta-Análisis en la Medicina	12
2.4. Etapas o Fases de un Meta-Análisis	13
2.4.1. Formulación del Problema	14
2.4.2. Búsqueda de la Literatura	15
2.4.3. Codificación de Estudios	19
2.4.4. Medidas de los Resultados	20
2.4.5. Análisis e Interpretación de los Resultados	26
3. Meta-Análisis en Pruebas Diagnósticas	31
3.1. Pruebas Diagnósticas	31
3.1.1. Clasificación de las enfermedades por diagnosis	31

ÍNDICE GENERAL

3.1.2.	Criterios de calidad de las pruebas	32
3.1.3.	Diseño de una investigación para evaluar una prueba diagnóstica	32
3.1.4.	Parámetros de valoración de pruebas diagnósticas	33
3.2.	Análisis de Sensibilidad y Especificidad en las medidas resumen del análisis diagnóstico	49
3.3.	Modelos Estadísticos para integración de pruebas	52
3.3.1.	Enfoque SROC	53
3.3.2.	Modelo Bivariante	67
3.3.3.	Modelo de Regresión Binomial	71
3.3.4.	Regresión Jerárquica para Meta-análisis de las evaluaciones en la precisión de las pruebas diagnósticas-HSROC	73
3.3.5.	Meta-análisis bayesiano para la precisión de un test en ausencia de la prueba oro estándar	81
3.3.6.	Unificación de modelos para Meta-Análisis de la precisión diagnóstica	86
3.3.7.	Relación entre el modelo bivariante y HSROC	87
3.3.8.	Relación de los parámetros de los modelos sin covarianzas	88
3.3.9.	Relación de los parámetros de los modelos con covarianza	90
3.3.10.	Enfoque de inferencia y salidas de los modelos	90
3.3.11.	Modelo Estadístico para I^2	92
3.3.12.	Meta-análisis para estudios de precisión diagnóstica: un modelo estadístico utilizando distribuciones beta-binomiales y cópulas bivariantes	97
3.3.13.	Algoritmo y Esquema de Decisión para meta-análisis de pruebas diagnósticas en enfermedades de baja prevalencia	111
4.	Análisis de Resultados	125
4.1.	Precisión Diagnóstica con MADA para Meta-Análisis	125
4.1.1.	Introducción	125
4.1.2.	Obteniendo el paquete	126
4.1.3.	Ingresando los datos	126
4.1.4.	Celdas vacías	129
4.1.5.	Estadística descriptiva	129
4.1.6.	Gráficos descriptivos	133
4.1.7.	Enfoque univariante	135
4.1.8.	Odds Ratio de Diagnóstico (DOR)	136
4.1.9.	Modelo Proporcional Hazards	138
4.1.10.	Enfoque Bivariante	141
4.1.11.	Comparando curvas SROC	144
4.1.12.	Meta-Regresión bivariante	146
4.2.	Precisión diagnóstica con Stata para Meta-Análisis	149
4.2.1.	Descripción	149
4.2.2.	Sintaxis de los comandos de Stata	151

4.2.3.	Gestiones de base de datos	152
4.2.4.	Tipos de datos	152
4.2.5.	Cargar un archivo de datos en formato Stata	153
4.2.6.	Opciones en Stata	154
4.2.7.	Guardando los resultados	158
4.2.8.	Ejemplo de un conjunto de datos (dataset)	159
4.2.9.	Modelos de predicción y diagnósticos	163
4.2.10.	Evaluación de la calidad	165
4.2.11.	Asociación bivalente	168
4.2.12.	Resúmenes de las estimaciones de rendimiento	169
4.2.13.	Heterogeneidad estadística	169
4.2.14.	Forest-Plot para analizar la sensibilidad y especificidad	171
4.2.15.	Curva SROC	172
4.2.16.	Meta-Regresión	173
4.2.17.	Test de regresión lineal de asimetría-Gráfico Funnel Plot	176
4.2.18.	Fagan Plot (Nomograma de Bayes)	178
4.2.19.	Diagrama de dispersión de los Odds ratio de Máxima Verosimilitud	179
4.2.20.	Valores predictivos y gráfico de probabilidad modificatoria	180
4.3.	Precisión diagnóstica con HSROC para Meta-Análisis	182
4.3.1.	Introducción	182
4.3.2.	Meta-Análisis de una prueba oro de referencia estándar asumiendo independencia condicional	182
4.3.3.	Preparación de los datos	183
4.3.4.	Ejecutando la simulación Cadenas de Markov Monte Carlo (MCMC)	186
4.3.5.	Interpretando los archivos de salida	187
4.3.6.	Estadística descriptiva	187
4.3.7.	Resumen gráfico	191
4.4.	Simulación de datos derivados de un modelo HSROC suponiendo independencia condicional entre el índice y las pruebas de referencia.	195
4.5.	Precisión diagnóstica mediante el paquete CopulaDTA para Meta-Análisis	200
4.5.1.	Marco de referencia y software	200
4.5.2.	El paquete CopulaDTA	200
4.5.3.	Diagnóstico del modelo	201
4.5.4.	Comparación y selección de modelos	202
4.5.5.	Dataset	202
4.5.6.	Modelización	204
4.5.7.	Comparación de modelos	222
4.6.	Meta-análisis de precisión diagnóstica utilizando el programa informático SAS	224
4.6.1.	Modelos Jerárquicos	224
4.6.2.	Modelo Bivalente	225
4.6.3.	Modelo HSROC	231

4.6.4. Heterogeneidad y análisis de regresión utilizando el modelo biva- riante	235
4.6.5. Heterogeneidad y análisis de regresión utilizando el modelo HSROC	239
4.6.6. Criterios para selección de modelos	242
4.6.7. Comparando pruebas índices	245
5. Aprendizaje Automático	253
5.1. Introducción	253
5.2. Orígenes del Aprendizaje Automático	254
5.3. Cómo “aprenden” las máquinas	256
5.4. Tipos de datos de entrada	258
5.5. Tipos de algoritmos de aprendizaje automático	259
5.6. Relacionando datos de entrada con algoritmos	262
5.6.1. Comprendiendo la clasificación del vecino más cercano	263
5.6.2. Aprendizaje probabilístico-Clasificación usando Naive Bayes	267
5.6.3. Comprender los árboles de decisión	270
5.6.4. Comprender las reglas de clasificación	274
5.6.5. Métodos Black Box - Redes Neuronales y Máquinas de Soporte Vectorial	277
5.6.6. Evaluación del rendimiento del modelo	297
5.6.7. ¿Qué tiene de malo la Exactitud?	299
5.6.8. ¿Qué tiene de malo Precision/Recall?	300
5.6.9. El Estadístico Kappa	301
5.6.10. La medida F	301
5.6.11. Análisis ROC	302
5.6.12. Estimación del rendimiento futuro	303
5.6.13. Afinación de los modelos para mejorar desempeño (rendimiento)	307
5.7. Elección del Modelo Estadístico en Meta-análisis de Pruebas Diagnósti- cas mediante Aprendizaje Automático	308
5.7.1. Algoritmo de Máquinas de Vectores Soporte para Big Data	312
5.7.2. Vectores de Soporte para elección del modelo estadístico	317
Conclusiones	341
A. Código/Manuales/Publicaciones	343
A.1. Estadística descriptiva de los parámetros a estimar	343
A.2. Gráficos de dispersión, distribución muestral a posteriori, datos MRI	347
A.3. Gráficos de densidad de los parámetros inter-estudio, datos MRI	354
A.4. Código en SAS para el modelo bivariante, datos Nishimura et al. (136)	361
A.5. Código en SAS para el modelo HSROC, datos Nishimura et al. (136)	365
A.6. Código en SAS comparación de pruebas, datos Schuetz et al. (167)	370
A.7. Código en R, generación de tablas tetracóricas, con patrón oro	378
A.8. Código en R, generación de tablas tetracóricas, 1 punto de corte	380
A.9. Código en R, generación de tablas tetracóricas, 2 puntos de corte	382

A.10.Código en R, generación de tablas tetracóricas, 3 puntos de corte	384
A.11.Código en R, ajustando modelos estadísticos a las tablas tetracóricas simuladas	386
A.12.Código en R, para etiquetar modelos a los meta-análisis simulados	394
B. Artículos científicos publicados, Estancia de Investigación y Asistencia a Eventos Académicos Internacionales	397
B.1. An Alternative To The Cochran-(Q) Statistic For Analysis Of Heterogeneity In Meta-Analysis Of Diagnostic Tets Based On HJ Biplot	397
B.2. Meta-Analysis in DTA with Hierarchical Models Bivariate and HSROC: Simulation Study	407
B.3. Meta-Análisis del Estudio de la Precisión Diagnóstica empleando Modelización de Efectos Aleatorios	420
B.4. Certificado de Estancia de Investigación, Primer Evaluador	422
B.5. Certificado de Estancia de Investigación, Segundo Evaluador	424
B.6. Asistencia a la Primera Escuela AME: Ciencia de Datos del 23 al 25 de Septiembre de 2019 en la UNAM, CDMX, México	426
B.7. Certificate of Presentation: For presentation the work “Hierarchical modeling and copulas for the study of diagnostic accuracy” in the IV International Workshop on Proximity Data, Multivariate Analysis and Classification on 25 and 26 of April 2019 at Unversity Salamanca	428
B.8. Presentación Ponencia Oral en el ciclo de Conferencia en el marco del Programa de Estadística Multivariante organizado en conjunto entre ESPOL y USAL, en Guayaquil del 21 al 29 de septiembre 2018	430
B.9. Constancia como Ponente en el 33 Foro Nacional de Estadística y 13 Congreso Latinoamericano de Sociedades Estadística, celebrado del 1 al 5 de Octubre de 2018 en Guadalajara, Jalisco, México	432
B.10.Participación como Expositor en el ”VI Encuentro Iberoamericano de Biometría”, realizado en Sangolquí-Ecuador, el 15, 16 y 17 de Noviembre del 2017	434
B.11.Comunicación de Simposio en el III Congreso Nacional de Psicología, realizado en Oviedo-Asturias-España, desde 3 al 7 de julio 2017	436
B.12.Ponente de Simposio en el XXVI Simposio de Estadística 2016, realizado del 8 al 12 de agosto de 2016 en Sicelejo, Sucre, Colombia	438
Bibliografía	441

Índice de figuras

1.1. Tipos de revisiones según Sackett (159)	3
1.2. Tipos de revisiones. Adaptado de Madhukar Pai	4
2.1. Número de publicaciones sobre Meta-análisis en PubMed (1979-2018)	13
2.2. Etapas de meta-análisis	14
2.3. Metodología PICO	14
2.4. Población objetivo y población accesible en un meta-análisis	16
2.5. Selección de Estudio - Diagrama de flujo	18
2.6. Interpretación coeficiente Kappa.	18
2.7. Diagrama que resume la metodología meta-partición.	29
3.1. Criterio de calidad de las pruebas	32
3.2. Diseño para evaluar una prueba diagnóstica	33
3.3. Variación de la razón de verosimilitud en función de la sensibilidad para diferentes pruebas cuya especificidad es igual a 0,90	40
3.4. Variación de las razones de verosimilitud en función de la especificidad para diferentes pruebas cuya sensibilidad es igual a 0.80	41
3.5. Especificidad Vs Sensibilidad	42
3.6. Especificidad Vs Sensibilidad	42
3.7. Cocientes de Probabilidad, según Evidence-bases Medicine Working Group	43
3.8. Gráfica de Fagan, datos midas_example	44
3.9. Curva ROC, datos simulados	47
3.10. Curva SROC con varios valores de α ($\beta = 0$), datos simulados	58
3.11. Curva SROC con varios valores de β ($\alpha = 2$), datos simulados	59
3.12. Diagrama de enfoques HSROC, Bivariante, SROC, elaboración propia	96
3.13. Densidad y gráfico de contorno Cópula de Gauss	108
3.14. Densidad y gráfico de contorno Cópula de Clayton	109
3.15. Densidad y gráfico de contorno Cópula de Clayton	110
3.16. Densidad y gráfico de contorno Cópula de Clayton	111
3.17. Esquema metodológico para el meta-análisis en pruebas diagnósticas	121
4.1. Gráficos forest-plot, para sensibilidad y especificidad, datos AuditC	134

ÍNDICE DE FIGURAS

4.2. Gráfico ponderado crosshair con colores arbitrarios y regiones de confianza en forma de elipse, datos AuditC	135
4.3. Forest-plot, logaritmos DOR´s, ajuste DSL, datos AuditC	138
4.4. Resumen gráfico con modelo PHM, datos AuditC	141
4.5. Curva SROC, modelo de Reitsma, datos AuditC	144
4.6. Comparación de curvas SROC para SAQ (entrevistados) y IAQ (entrevistadores), datos IAQ y SAQ	146
4.7. Ambiente de trabajo en Stata	150
4.8. Menú y paneles en Ambiente de trabajo Stata	151
4.9. Cargar archivo desde barra de Menú en Stata	153
4.10. Comando para leer en disco duro la ruta del archivo <i>midas_example</i>	159
4.11. Acceso a <i>midas_example.dta</i> desde las pantallas	160
4.12. Comando para usar los datos de <i>midas_example.dta</i>	160
4.13. Instalación de Paquete midas desde el repositorio oficial	163
4.14. Bayes empírico predicho frente a los resultados de las pruebas observadas, datos <i>midas_example</i>	164
4.15. Representación gráfica de la normalidad bivalente, bondad de ajuste basada en residuales y los análisis de detección de influencia y atípicos, datos <i>midas_example</i>	165
4.16. Instalación del paquete QUADAS desde el repositorio oficial	166
4.17. Gráfico de barras de la evaluación de la calidad para cada artículo, incluidos los criterios modificados QUADAS, datos <i>midas_example</i>	167
4.18. Diagrama de caja bivalente de los estudios agrupados con siete valores atípicos que sugiere indirectamente un menor grado de heterogeneidad, datos <i>midas_example</i>	168
4.19. Salida de la estadística de heterogeneidad, datos <i>midas_example</i> , elaboración propia	170
4.20. Diagrama forest plot para cada estudio específico (eje derecho) y la media de sensibilidad y especificidad con las estadísticas de heterogeneidad correspondientes, datos <i>midas_example</i>	171
4.21. Resumen de la curva ROC con las regiones de confianza y predicción alrededor de la media de sensibilidad y el punto de especificidad, datos <i>midas_example</i>	173
4.22. Forest plot de una meta-regresión univariante múltiple, datos <i>midas_example</i>	176
4.23. Gráfico Funnel Plot con la línea de regresión superpuesta, datos <i>midas_example</i>	177
4.24. Gráfica de Fagan, datos <i>midas_example</i>	179
4.25. Razón de verosimilitud y diagrama de dispersión, datos <i>midas_example</i>	180
4.26. Gráfica de Probabilidad modificada, datos <i>midas_example</i>	181
4.27. Lista de configuraciones generales de la simulación	188
4.28. Resumen de las distribuciones utilizadas	188
4.29. Resumen de las distribuciones utilizadas	190

4.30. Resumen de las estimaciones de los parámetros intra-estudio	190
4.31. Gráfico de dispersión, de la distribución muestral a posteriori de un sólo parámetro, datos MRI	192
4.32. Gráficos de densidad de algunos de los parámetros inter-estudios, datos MRI.	193
4.33. Curva SROC, datos MRI	194
4.34. Convergencia de la Mezcla según cadenas, Cópula Gauss	206
4.35. Forest Plot de sensibilidad y especificidad a priori por estudios con intervalos al 95 % de confianza	208
4.36. Forest Plot de sensibilidad y especificidad a posteriori por estudios con intervalos al 95 % de confianza	209
4.37. Forest Plot de sensibilidad y especificidad a priori y a posteriori por estudios con intervalos al 95 % de confianza	210
4.38. Trace Plot de la sensibilidad y especificidad media a posteriori de los datos de telomerasa estimados por cópulas de Gaussian, Clayton 90° (C90), 270° (C270), Farlie-Gumbel-Morgenstern (FGM) y lade Frank basada en distribuciones beta bivariantes y normal bivalente (BRMA).	218
4.39. Gráfico de la sensibilidad y especificidad meta-analítica a posteriori y los intervalos correspondientes de 95 % de confianza según lo estimado por las cópulas de, Gaussian, Clayton 90° (C90) y 270° (C270), Farlie-Gumbel-Morgenstern (FGM) y la de Frank basada en distribuciones bivariadas bivariantes y el modelo normal bivalente (BRMA) para los datos de la telomerasa.	222
4.40. Forest Plot, datos Nishimura et al. (136)	228
4.41. Scatter Plot, datos Nishimura et al. (136)	229
4.42. Medidas resumen, ajuste bivalente, datos Nishimura et al. (136)	230
4.43. Curva SROC, ajuste bivalente, datos Nishimura et al. (136)	231
4.44. Medidas resumen, ajuste HSROC, datos Nishimura et al. (136)	234
4.45. Curva SROC, ajuste HSROC, datos Nishimura et al. (136)	234
4.46. Medidas resumen con covariable “CCP”, ajuste Bivalente, datos Nishimura et al. (136)	238
4.47. Curva SROC, ajuste Bivalente con covariable CCP, datos Nishimura et al. (136)	239
4.48. La precisión relativa depende de los valores de especificidad, las curvas se entrelazan	240
4.49. El grupo 1 domina en todos los valores de especificidad, las curvas no se cruzan.	241
4.50. Medidas resumen según método, ajuste HSROC, datos Nishimura et al. (136)	244
4.51. Curva SROC, ajuste HSROC según método, datos Nishimura et al. (136)	245
4.52. Medidas resumen, comparación de índices, ajuste bivalente, datos Schuetz et al. (167)	249
4.53. Curva SROC, ajuste Bivalente según Test, datos Schuetz et al. (167)	250

ÍNDICE DE FIGURAS

5.1. Ciclo del aprendizaje automático	255
5.2. Pasos en el proceso de aprendizaje	257
5.3. Un ejemplo de aprendizaje supervisado y no supervisado	261
5.4. Límite de decisión algoritmo K-NN, k “grande” y “pequeño”	265
5.5. Árbol de decisión	270
5.6. Funcionamiento de una neurona	279
5.7. Funcionamiento de una neurona artificial	279
5.8. Función de activación por pasos	281
5.9. Función de activación por pasos	281
5.10. Tipos de funciones de activación	282
5.11. Topología de una red neuronal simple	284
5.12. Topología de una red neuronal compleja	285
5.13. Tipología de una red con múltiples capas	285
5.14. Tipología de una red recurrente	286
5.15. Gradiente descendente	289
5.16. SVM, hiperplano dos dimensiones	290
5.17. SVM, Varios hiperplanos en dos dimensiones	291
5.18. SVM, Vectores Soporte y Margen Máximo	292
5.19. SVM, Convex Hull	293
5.20. SVM, variable “slack”	294
5.21. SVM, Kernel no lineal	295
5.22. Partición de los datos en muestra de entrenamiento y prueba	304
5.23. Partición de los datos en muestra de entrenamiento, prueba y validación	305
5.24. Proceso de simulación de Meta-análisis y selección del mejor modelo estadístico	311
5.25. Meta-análisis etiquetas según modelo estadístico	312
5.26. Exactitud e Índice Kappa según modelos de aprendizaje	317
5.27. Importancia de los predictores en el modelo SVM	323
5.28. Error de Validación , hiperparámetro C	327
5.29. Exactitud para diferentes valores de C, kernel lineal	330
5.30. Error de validación, según costo y grado del polinomio	332
5.31. Error de validación, según costo y factor gamma	336
5.32. Accuracy, según costo kernel radial	337
5.33. Accuracy y Kappa, según kernel	338
5.34. Diagrama de Decisión para realizar un aprendizaje automático, según naturaleza de los datos	339

Índice de tablas

1.1. Revisiones narrativas vs. Revisiones sistemáticas. Adaptado de Cook et al. (29)	2
2.1. Estrategias de Búsquedas en Pub-Med sobre Meta-Análisis	12
2.2. Definiciones de riesgos y Odds Ratio	24
2.3. Presencia o ausencia del factor de riesgo	25
3.1. Tabla de doble entrada, ejemplo hipotético	34
3.2. Explicación de la construcción de los principales parámetros en la descripción de una prueba diagnóstica	48
3.3. Tabla resumen de datos	52
3.4. AUC_{hom} , Q^* , y sus diferencias para varios valores de odds ratio diagnóstico: caso homogéneo	64
3.5. Resumen de modelos estadísticos en pruebas diagnósticas	112
3.6. Resumen de las funciones empleadas en R para enfoque univariante . . .	113
3.7. Resumen de las funciones empleadas en R para enfoque bivariante . . .	115
3.8. Resumen de las funciones empleadas en Stata para enfoque bivariante .	116
3.9. Resumen de la macro empleada en SAS para pruebas diagnósticas . . .	117
3.10. Resumen de las funciones empleadas en R para enfoque jerárquico-HSROC	118
3.11. Resumen de las principales medidas y gráficos para el enfoque univariante, bivariante y HSROC	119
4.1. Datos del estudio i-ésimo en una tabla de 2×2	126
4.2. Orden de las variables a,b,c y d	152
4.3. Tipos de datos en Stata	152
4.4. Interpretación valores AUROC	172
4.5. Media a posteriori, intervalo de confianza al 95 % , tamaño de muestra efectiva y factor de reducción de escala potencial \hat{R} para las medias marginales y parámetros de correlación según estimación por cópulas de, Gaussian, Clayton 90° (C90) y 270° (C270), Farlie-Gumbel-Morgenstern (FGM) y Frank explicadas en distribuciones beta bivalentes y el modelo normal bivalente (BRMA) para los datos de telomerasa.	220

ÍNDICE DE TABLAS

5.1. Principales tipos de algoritmos de aprendizaje	262
5.2. Fortalezas y Debilidades del algoritmo K-NN	264
5.3. Fortalezas y Debilidades del algoritmo Naive Bayes	268
5.4. Fortalezas y Debilidades del algoritmo C5.0	272
5.5. Fortalezas y Debilidades del algoritmo 1R	275
5.6. Fortalezas y Debilidades del algoritmo RIPPER	276
5.7. Fortalezas y Debilidades del algoritmo “backpropagation”	288
5.8. Fortalezas y Debilidades de los kernels no lineales	296
5.9. Matriz de Confusión	299
5.10. Medidas estimadas a los Meta-análisis simulados	310
5.11. Criterios para etiquetar el mejor modelo estadística para cada meta-análisis simulado	311
5.12. Tipos de medidas estadísticas empleadas para la construcción de variables predictoras	318
5.13. Descripción de variables según su tipo	318
5.14. Porcentaje de los errores de entrenamiento y prueba según kernel del modelo SVM	321

Capítulo 1

Introducción

El crecimiento exponencial de la literatura médica y el uso cada vez más extendido de las tecnologías de la información y la comunicación (TIC), junto con la dispersión de la literatura científica, dificulta al investigador, y sobre todo al profesional sanitario, el acceso a la información relevante. Meta-análisis es una revisión sistemática que incorpora una estrategia estadística para integrar resultados de varios estudios en una sola estimación.

La necesidad de manejar grandes volúmenes de datos, con características importantes como la falta de actualización de la información, o la pobre calidad de las fuentes de información no son situaciones desconocidas en el campo médico-sanitario, por lo que obligan al experto a investigar artículos de revisión en lugar de artículos originales. Es así, que las revisiones proveen las herramientas necesarias para informarse y actualizarse en el menor tiempo posible. En la práctica se ha desarrollado una revisión, la revisión narrativa, la cual hace uso de métodos informales y hace un análisis muy subjetivo, Sáenz and Ausejo (163). En las revisiones narrativas existen defectos, tales como que los investigadores no especifican el procedimiento que se ha utilizado para focalizar y evaluar los resultados que han ayudado a obtener sus conclusiones y la más importante es que no se es capaz de reproducir, verificar y contrastar las conclusiones de las revisiones Sáenz and Ausejo (163). Al nombrar el término de revisiones hay que tener claro que existen dos tipos, revisión y revisión sistemática, Sackett (159). La primera hace referencia a todo intento que sintetice los resultados y las conclusiones de dos o más publicaciones relacionadas con un tema dado, la segunda revisión, se centra en la identificación exhaustiva a través de toda la literatura de un tema dado, valorando su calidad y sintetizando sus resultados.

1. INTRODUCCIÓN

Característica	Revisión Narrativa	Revisión Sistemática
Pregunta de investigación	Amplia. No estructurada	Estructurada, clara, concreta y centrada en un problema clínico bien definido.
Búsqueda bibliográfica.	No detallada. No sistemática.	Búsqueda detallada, sistemática y explícita.
Selección de fuentes de información.	No orientada a localizar todos los estudios disponibles. Probabilidad alta de sesgo	
Selección de artículos.	No hay criterios de selección. No reproducible. Probabilidad alta de sesgo.	Selección basada en criterios explícitos. Aplicación uniforme de los criterios de selección, exclusión a todos los artículos.
Valoración de la calidad de los estudios.	No hay valoración.	Valoración, evaluación crítica de la calidad metodológica de los estudios.
Síntesis.	A menudo resumen subjetivo, cualitativo, sin un estimador estadístico.	Basada en la calidad metodológica de los estudios. A menudo resumen cuantificado por un estimador estadístico
Interpretación.	A veces basada en la evidencia. Frecuentemente basada en opiniones personales.	Generalmente basada en la evidencia

Tabla 1.1: Revisiones narrativas vs. Revisiones sistemáticas. Adaptado de Cook et al. (29)

Las diferencias entre las revisiones narrativas y revisiones sistemáticas aparecen recogidas en la Tabla 1.1. Dado el creciente volumen de información disponible, ha motivado la creación de técnicas que ayuden a los médicos a filtrar y extraer información de calidad y clínicamente válida como la lectura crítica, Oxman et al. (140) y las revisiones sistemáticas. Las características esenciales de toda revisión sistemática, Martín Rodero (121), son las siguientes:

- Objetivos claramente definidos con criterios, previamente establecidos, de selección (inclusión/exclusión) de los estudios.
- Metodología explícita y reproducible.
- Búsqueda sistemática y exhaustiva que trate de identificar todos los estudios que cumplan los criterios de elegibilidad.
- Evaluación de la validez de los resultados de los estudios incluidos, por ejemplo a través de la evaluación del riesgo de sesgo.

-
- Síntesis y presentación sistemática de las características y resultados de los estudios incluidos.

Una de las fortalezas del método científico es que una investigación bajo este contexto debe ser replicable. Es por tal motivo que entre lo que se ha estudiado en el pasado y las futuras investigaciones se tome en consideración la revisión de la literatura y la resolución de las contradicciones empíricas para así validar la consistencia o robustez de un determinado hecho empírico, Galindo-Villardón (59). Sackett et al. (160) presentan definiciones claras de revisiones narrativas, sistemáticas y meta-análisis, ver Figura 1.1. A partir de la Figura 1.1, se podría expandir hacia un cuarto nivel los tipos de revisiones, ver Figura 1.2, en la cual se encontrarían los meta-análisis de datos individuales de pacientes, Martín Rodero (121).

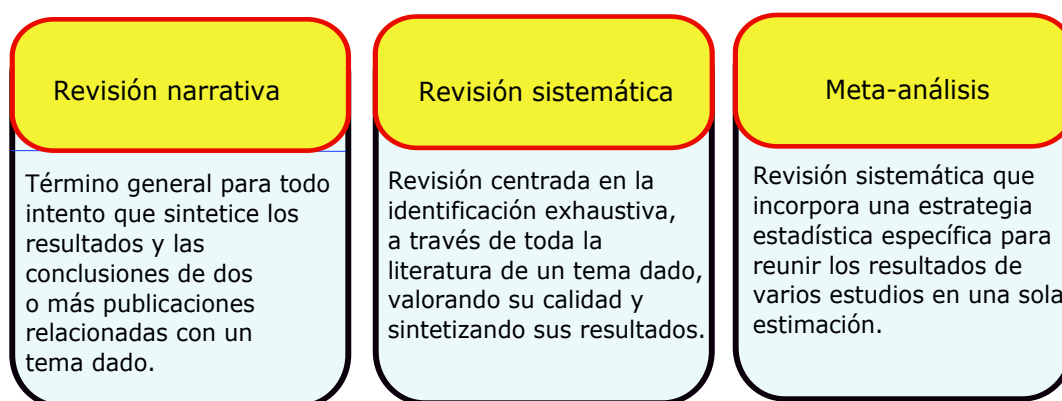


Figura 1.1: Tipos de revisiones según Sackett (159)



Figura 1.2: Tipos de revisiones. Adaptado de Madhukar Pai

En un meta-análisis se diferencian 6 etapas, Rosenthal (152), todas ellas esenciales. En este trabajo centraremos la atención en la fase de integración de resultados y abordaremos la problemática particular asociada a las pruebas diagnósticas utilizadas en screening masivos de enfermedades de baja prevalencia, pero cuyo diagnóstico precoz es esencial, es decir, cuando los verdaderos positivos son muy bajos en comparación con los verdaderos negativos (autismo, por ejemplo).

1.1. Objetivos

1.1.1. Objetivo General

Este trabajo tiene por objetivo general proponer procedimientos alternativos a las soluciones clásicas de meta-análisis, en ausencia de prueba oro, específicamente diseñadas para los casos en los que la eficacia es muy alta pero los falsos positivos también.

1.1.2. Objetivos Específicos

- Realizar una exhaustiva revisión bibliográfica que nos permita conocer el estado del arte en relación a la evaluación de las pruebas diagnósticas cuando los estudios son heterogéneos, presentan diferentes puntos de corte y no se dispone de prueba oro.
- Analizar la problemática asociada a los meta-análisis de pruebas diagnósticas utilizadas en enfermedades con baja prevalencia, donde no es posible utilizar la exactitud para valorar la prueba.
- Evaluar el desempeño de la prueba Q, el índice I^2 y el HJ Biplot Galindo-Villardón (60) como herramientas para detectar las posibles causas de la heterogeneidad en un espacio bidimensional.
- Realizar un estudio comparativo entre el Modelo de Moses Moses et al. (132), el modelo Bivariante Reitsma et al. (148) y el modelo jerárquico HSROC Rutter and Gatsonis (158).
- Realizar un estudio crítico comparativo de los programas MIDAS (STATA) Dwamena (45), MADA (R) Doebler (40), y el HSROC (R) Schiller et al. (166), por ser los de más uso en este contexto.
- Presentar los pasos a seguir para llevar a cabo un meta-análisis de tests diagnósticos en situaciones de baja prevalencia, que pueda servir de referencia a los investigadores de este campo y nos permita construir un modelo de **aprendizaje autónomo** que sea una herramienta útil para el investigador en el momento de integrar los resultados de un meta-análisis de test diagnóstico.
- Desarrollar los algoritmos necesarios para la implantación de los modelos propuestos.
- Aplicar los diferentes modelos a datos reales o simulados y realizar un análisis comparativo.

1.2. Motivación

El diagnóstico correcto de una alteración (enfermedad) es de interés eminente y principal en la psicología y la medicina. Las pruebas diagnósticas son herramientas comunes para discernir la presencia o ausencia de una condición, o para examinar a los pacientes que están en riesgo de desarrollar una enfermedad. Muchas pruebas diagnósticas se basan en puntuaciones obtenidas a partir de cuestionarios breves o se fundamentan en un biomarcador único. Por lo tanto, no siempre darán un diagnóstico correcto. Cuando se dispone de estudios primarios que evalúan la calidad de una prueba diagnóstica,

la realización de un meta-análisis de diagnóstico se ha convertido en una herramienta clave para investigar la información disponible sobre una prueba diagnóstica Egger et al. (48), Hasselblad and Hedges (78), Schiller et al. (166), Sutton et al. (179). En un estudio diagnóstico primario, la calidad de una prueba diagnóstica se mide a menudo en términos de la sensibilidad (tasa de verdaderos positivos) y la especificidad (tasas de verdaderos negativos = $1 - \text{tasa de falsos positivos}$) de la prueba.

1.3. Planteamiento del problema

Para poder operar con las razones de verosimilitudes en el cálculo de probabilidades, estas deben transformarse en ventajas (odds), estos cálculos se simplifican recurriendo a nomogramas (de Fagan). La sensibilidad y la especificidad varían en función del punto de corte que se utilice para separar los valores normales del resto. En estos casos, los resultados pueden representarse gráficamente como una curva ROC, que permite conocer las características de la prueba según diferentes puntos de corte y que puede utilizarse para elegir el más adecuado. Un estimador de la relevancia global de una prueba puede ser el área bajo la curva (AUC). Es preciso tener en cuenta que la información de que se dispone sobre la validez de las pruebas diagnósticas procede de diferentes poblaciones. Por tanto, las estimaciones obtenidas en dichos estudios están sujetas a variabilidad aleatoria y si los estudios han sido diseñados incorrectamente, a sesgos. La variabilidad de las mediciones va a estar influida por múltiples factores que interesa conocer y controlar. Entre ellos, tiene especial importancia distinguir las variaciones intra e inter estudio.

Debido a esta particularidad, se debe conservar la naturaleza bivariante de los datos, modelizando la sensibilidad y especificidad de manera conjunta. Durante la última década se han establecido dos modelos: un modelo jerárquico Rutter and Gatsonis (158) y un modelo bivariante Reitsma et al. (148).

Concomitantemente con lo anterior, el campo de la estadística está sufriendo constantemente desafíos por los problemas que la ciencia y la industria presentan. Con el advenimiento de las computadoras y la era de la información, los problemas estadísticos han crecido tanto en tamaño como en complejidad. Los desafíos en las áreas de almacenamiento de datos, organización y búsqueda han generado un nuevo campo: “**minería de datos**”; y el trabajo del estadístico es darle sentido a todo esto: extraer patrones y tendencias importantes y entender cuáles son estos datos, a esto se llama “aprendizaje estadístico”. Los problemas de aprendizaje pueden ser categorizados como supervisados o no supervisados. En el aprendizaje supervisado, el objetivo es predecir el valor de una medida de resultado basada en una serie de medidas de insumos; en el aprendizaje no

supervisado, no hay una medida de resultado, y el objetivo es describir las asociaciones y los patrones entre un conjunto de medidas de entrada.

1.4. Metodología

La metodología a utilizar para cumplir con el propósito de la investigación, será de naturaleza analítica y comparativa.

1.4.1. Métodos Estadísticos

En general, el meta-análisis es un proceso en dos etapas. En un primer paso se estiman los resultados de cada estudio, aunque en el caso de la evaluación de pruebas diagnósticas cada estudio es resumido por una pareja de índices que describen la validez de la prueba. Habitualmente, estos dos índices son sensibilidad y especificidad, o bien los cocientes de probabilidad positivo y negativo. En un segundo paso se deben calcular índices globales de validez para lo que se han propuesto diversos métodos, que se expondrán más adelante. En pruebas diagnósticas el supuesto de homogeneidad metodológica en los estudios no se cumple y es por ello importante la evaluación de la heterogeneidad.

1.4.2. Evaluación de la heterogeneidad

Valorar la eventual presencia de heterogeneidad estadística en los resultados puede realizarse (de manera clásica) presentando la sensibilidad y especificidad de cada estudio en un forest plot. En estos gráficos se representan los estimadores de los índices junto con sus intervalos de confianza, y suelen presentarse pareados y ordenados de acuerdo a uno de los índices. Una fuente característica de heterogeneidad es la que surge porque los estudios incluidos pueden haber usado diferentes umbrales para definir qué es un resultado positivo. Este efecto es conocido como **efecto umbral**.

1.4.3. Efecto umbral

Para explorar esta fuente de variación es útil representar en una gráfica los pares de sensibilidad y especificidad de cada estudio en un plano ROC. En este plano, la zona más cercana a la esquina superior izquierda asume buen rendimiento diagnóstico mientras que la zona central, la diagonal en la que sensibilidad y especificidad son iguales, representa una nula capacidad diagnóstica. Si existiera efecto umbral, los puntos en el

espacio ROC reflejaría una curva curvilínea. Cambiando el umbral de positividad de una prueba se obtendría una mayor (o menor) sensibilidad con el consiguiente efecto contrario sobre la especificidad.

Los métodos estadísticos más robustos propuestos para el meta-análisis tienen en cuenta esta correlación entre sensibilidad y, lo hacen mediante la estimación de una curva ROC resumen (SROC) de los estudios incluidos. Sin embargo, en limitadas ocasiones los resultados de los estudios primarios son homogéneos y puede descartarse la presencia tanto de efecto umbral como de otras fuentes de heterogeneidad. En esta situación, el resultado global de la revisión podría obtenerse a partir de la combinación ponderada de los índices de los estudios individuales. Como siempre, esta combinación puede hacerse mediante un modelo de efectos fijos o bien mediante un modelo de efectos aleatorios, dependiendo de la magnitud de la heterogeneidad. Se han propuesto varios métodos estadísticos para estimar la curva SROC. El primero de ellos, propuesto por Moses et al. (132), se basa en estimar una regresión lineal entre dos variables creadas a partir de los índices de validez de cada estudio.

1.4.4. Modelos bivariantes, jerárquicos

El modelo de Moses presenta algunas limitaciones. Por una parte, no tiene en cuenta la distinta precisión con la que se estimaron la sensibilidad y la especificidad en cada estudio y tampoco incorpora la heterogeneidad entre estudios. Para superar estas limitaciones, se han propuesto modelos de regresión más complejos. El primero de ellos es un modelo de efectos aleatorios, bivalente Reitsma et al. (148), que parte de la asunción de que los logit de sensibilidad y especificidad siguen una distribución normal bivalente. El modelo contempla la eventual correlación entre ambos índices, modeliza la distinta precisión con la que han sido estimadas la sensibilidad y la especificidad e incorpora una fuente de heterogeneidad adicional debida a la varianza entre estudios. El segundo modelo, se refiere al enfoque HSROC o modelo jerárquico Rutter and Gatsonis (158). Es similar al modelo anterior, salvo que hace explícita la relación existente entre sensibilidad y especificidad a través del umbral. Un tercer modelo es un modelo mixto de cópula que es una extensión del modelo lineal general mixto de Chu et al. (24), utilizando más bien una representación copular de la distribución de efectos aleatorios con distribuciones marginales normales y beta. Nikoloulopoulos (135) demuestra que en meta-análisis de los estudios de precisión de la prueba diagnóstica son un campo de aplicación primordial para los modelos de cópula, ya que el supuesto tradicional de normalidad multivariante es inválido en este contexto. Existe una gran variedad de paquetes estadísticos que pueden utilizarse para realizar los análisis descriptos., como SAS y Stata, que a través de una serie de macros programadas por usuarios facilitan la obtención de los modelos descritos. Las macros de, Stata se llama midas, y la macro desarrollada para SAS se denominada MetaDas Takwoingi and Deeks (182). Otros programas específicos para el meta-análisis de estudios de pruebas diagnósticas son las librerías mada, CopulaREMADA Nikoloulopoulos (135) y HSROC del paquete

estadístico R.

1.5. Contribuciones

La principal contribución de este trabajo es, presentar los pasos a seguir para llevar a cabo un meta-análisis de tests diagnósticos en situaciones de baja prevalencia, como también, construir un modelo de “**aprendizaje autónomo**” que se fundamentará en datos simulados del modelo HSROC que permita al investigador decidir qué modelo jerárquico seleccionar a la hora de integrar los resultados de un meta-análisis de test diagnóstico.

1.6. Estructura de la tesis

La presente investigación estudia el diagnóstico correcto de una condición específica, la cual es de interés eminente y principal en la psicología y la medicina. La utilidad de una prueba diagnóstica depende fundamentalmente de su validez y de su fiabilidad, pero también de su rendimiento clínico y de su coste. Si se evaluara la práctica clínica, observaremos que a menudo se emplean procedimientos diagnósticos de los que desconocemos estos parámetros. Para conocer esta información debemos recurrir a los estudios de evaluación de pruebas diagnósticas publicados. Sin embargo, en muchos de estos estudios no siempre encontramos información válida, relevante o aplicable a un entorno clínico, por lo que, la realización de un meta-análisis de diagnóstico se ha convertido en una herramienta clave para investigar la información disponible sobre una prueba diagnóstica.

El presente trabajo, tiene la siguiente estructura. En el Capítulo 1 se enuncia una breve introducción al tema de Meta-análisis, se plantan los objetivos del estudio, como también la motivación de la investigación con su respectivo planteamiento de la problemática a estudiar para posteriormente detallar la metodología y las contribuciones.

En el Capítulo 2 se analizan y detallan los conceptos importantes en meta-análisis y pruebas diagnósticas, luego en el Capítulo 3 se estudian los métodos estadísticos que se emplean en las pruebas diagnósticas, y finalmente en los Capítulos 4, 5, y 5.7.2 se presentan los resultados, aprendizaje autónomo, discusiones y las conclusiones más relevantes.

Capítulo 2

Meta-Análisis

2.1. Definición

La revisión sistemática nos lleva al meta-análisis, Galindo-Villardón (59), el cual puede definirse como el análisis estadístico de los resultados de estudios individuales con el fin de integrar los resultados, es decir, combina información estadística de estudios similares y realiza análisis estadístico de los resultados generales. En resumen, el meta-análisis puede ser definido como una revisión cuantitativa y síntesis de los resultados de estudios relacionados, los cuales son independientes, Martín Rodero (121).

2.2. Orígenes y Antecedentes

Agrupar investigaciones independientes no es algo nuevo, en 1904 Karl Pearson formuló un procedimiento para sintetizar los resultados de estudios que no tenían potencia con el fin de llegar a conclusiones significativas en términos estadísticos, Casares (17). El objetivo de un meta-análisis es llevar a cabo un análisis sistemático e integral de la información, sobre un conjunto de resultados de investigaciones empíricas que estudian un mismo problema, con el fin de llegar a conclusiones que puedan generalizarse, para optimizar el proceso cognoscitivo e investigativo en la toma de decisiones, cuando la incertidumbre está presente, sea esta científica, empresarial, educacional, social, etcétera, Avilés Merens et al. (5). En el año 1976 el Psicólogo americano Gene V. Glass, utiliza por primera vez el término meta-análisis para dar a conocer estudios estadísticos de una serie de resultados mediante ensayos clínicos individuales con el fin de integrarlos, Martín Rodero (121). Mediante la combinación de información a través de diferentes estudios, un análisis integrado tendrá más poder estadístico para detectar el tamaño del efecto que un análisis basado en un solo estudio. Cuando varios estudios tienen

resultados contradictorios, un meta-análisis puede ser utilizado para estimar un efecto promedio de estudios asociados, Normand (137).

2.3. Meta-Análisis en la Medicina

En Medicina, una de las primeras publicaciones que pueden ser calificadas de meta-análisis es un trabajo de Beechers sobre la evaluación del efecto placebo publicado en la revista JAMA en el año 1955, Egger and Smith (46). Actualmente existe un crecimiento en las publicaciones de meta-análisis, lo que refleja una importancia muy interesante en estas revisiones logrando aceptación en el campo de la medicina, Martín Rodero (121).

Es importante notar como han evolucionado en el tiempo las revisiones en el tema del meta-análisis, así encontramos la Figura 2.1, generada con la estrategia que se observa en la Tabla 2.1.

1. "meta-analysis"[PT] OR "meta-analysis as Topic"[Mesh]
2. meta-analysis[Title/Abstract] OR metaanalysis[Title/Abstract]
3. #1 OR #2

Tabla 2.1: Estrategias de Búsquedas en Pub-Med sobre Meta-Análisis

Nótese en la Figura 2.1, un crecimiento exponencial en las publicaciones de meta-análisis, a partir del año 1979 hasta 2018, algo importante que recalcar, es que este comportamiento creciente coincide con el nacimiento y desarrollo del estudio de la medicina basada en la evidencia.

El meta-análisis en gran medida se ha empleado para analizar y combinar salidas de ensayos clínicos aleatorizados, pero en la última década es frecuente notar aplicaciones importantes en evaluación de pruebas diagnósticas.

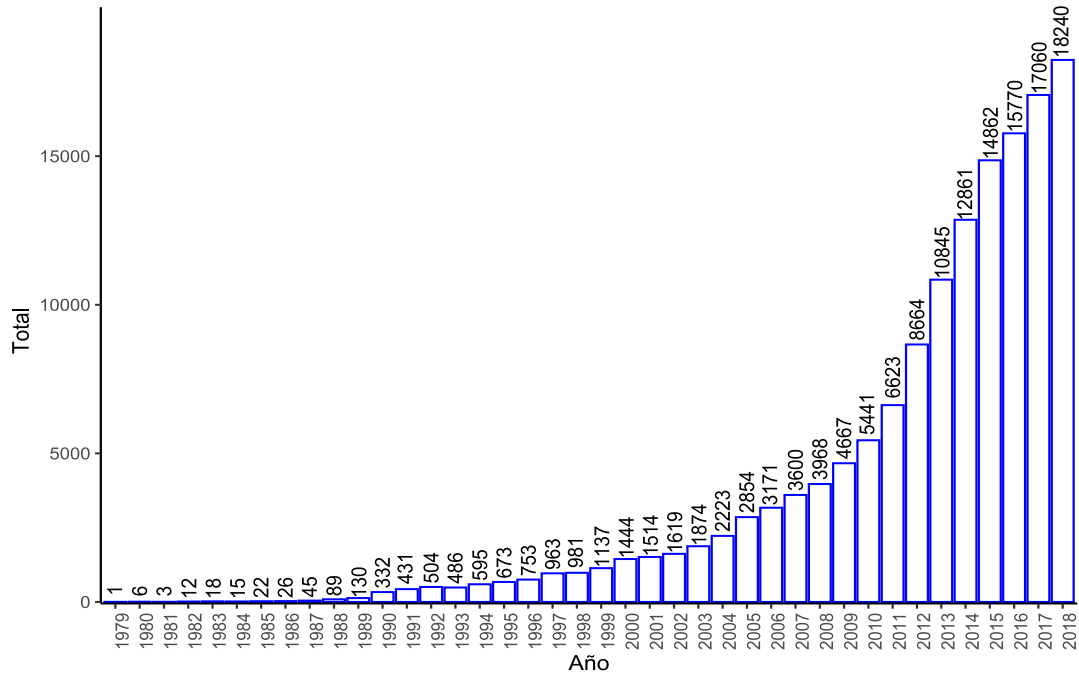


Figura 2.1: Número de publicaciones sobre Meta-análisis en PubMed (1979-2018)

2.4. Etapas o Fases de un Meta-Análisis

El termino meta-análisis es conocido como revisión cuantitativa, integración de investigación o análisis cuantitativo de dominios de investigación. La National Library of Medicine, utiliza el término meta-análisis desde 1989. Para elaborar un meta-análisis robusto y confiable, el mismo debe tener una planificación correcta y ha de contar con un protocolo que determine con claridad los objetivos y las etapas que deben de realizarse, Martín Rodero (121). El protocolo puede contar de seis fases, Martín-Vallejo (122). Existen cinco etapas por las que debe atravesar una síntesis de investigación, observe Figura 2.2.

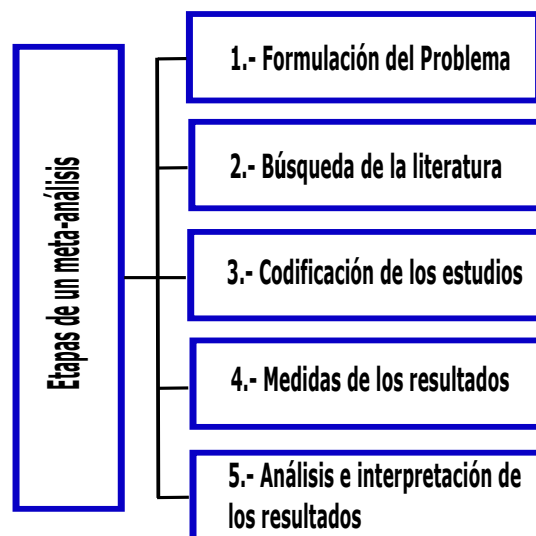


Figura 2.2: Etapas de meta-análisis

2.4.1. Formulación del Problema

En este paso se debe establecer con máxima rigurosidad posible el tema sobre el que se va a recoger la información, se deben tener claros los objetivos de la investigación y se deben tener definidos los conceptos y las variables de estudio. Para tener un buen tema de investigación se deben contemplar cuatro características de acuerdo a la metodología PICO, Oxman et al. (140), véase Figura 2.3.

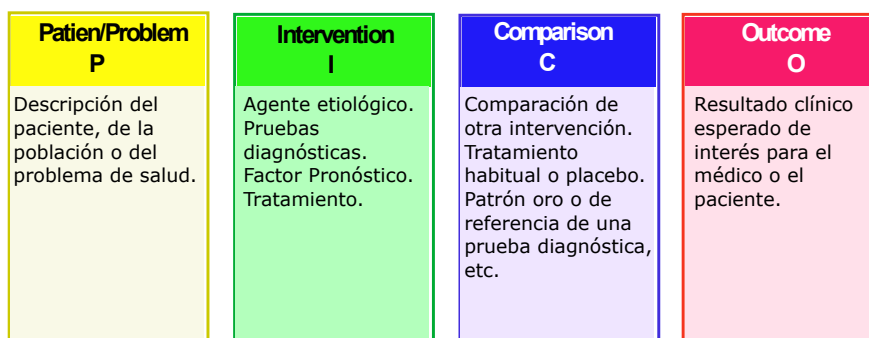


Figura 2.3: Metodología PICO

En otras palabras, hay que circunscribir el objeto de la revisión, definiendo con precisión las variables y conceptos implicados. Los objetivos con los que debe contar un meta-análisis se establecen de manera clara, en Gisbert and Bonfill (64),

1. Resumir los resultados de los estudios seleccionados y realizar una valoración cualitativa y cuantitativa.
2. Resolver la incertidumbre.
3. Mejorar la precisión del estimador del efecto testado i.e., aumentar el efecto estadístico.
4. Comparar y contrastar los resultados divergentes de los estudios individuales.
5. Responder preguntas nuevas que los estudios originales no habían planteado o no podían responder por no tener suficiente poder estadístico (análisis de subgrupos).
6. Orientar futuros trabajos de investigación, es decir, generar hipótesis. Resumir los resultados de los estudios seleccionados y realizar una valoración cualitativa y cuantitativa.

2.4.2. Búsqueda de la Literatura

Consiste en identificar todos los estudios que se han realizado sobre el tema de estudio. Es importante en un meta-análisis definir la población objetivo y la población accesible; la primera está constituida por todos los entes o individuos que el revisor espera poder representar en el estudio y la segunda incluye a aquellos individuos que el revisor puede hallar o localizar para el estudio. Se debe tener claro que la población accesible debe ser un subconjunto o una adecuación de la población objetivo, por ello para asegurar esta adecuación se debe contar con dos fuentes de información, fuentes informales y fuentes primarias, sino se cumple lo anterior es decir la población objetivo y accesible no coinciden, observe Figura 2.4. Esto puede suponer una amenaza importante a la validez de los resultados, Galindo-Villardón (59), Martín Rodero (121).

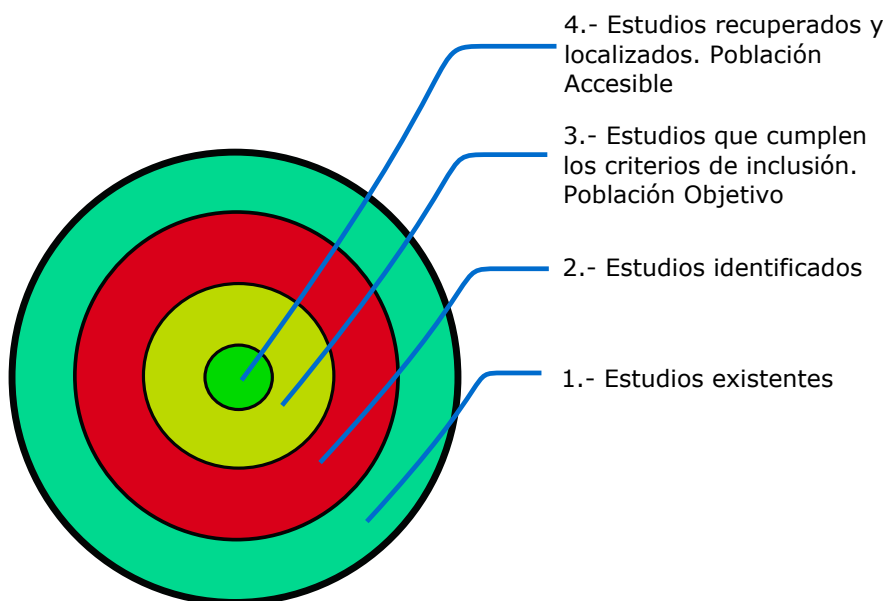


Figura 2.4: Población objetivo y población accesible en un meta-análisis

En las fuentes informales, se pueden encontrar trabajos presentados en congresos profesionales y las conexiones que se hagan con expertos en el tema de interés, con el fin de obtener material nuevo o no publicado por parte de un investigador que trabaje en el campo de estudio; las fuentes primarias son búsquedas o consultas de bibliografías personales y de revistas de investigación, las dos fuentes no son excluyentes entre sí.

En lo relacionado a las localizaciones de estudios existen dos técnicas establecidas, Cooper (30), el procedimiento ascendente y el descendente. En el primero se localizan investigaciones previas en listas de referencias de estudios existentes y disponibles, en cambio en el segundo procedimiento se ubican trabajos consultando índices de citas. La búsqueda por computadoras puede incluir los dos procedimientos antes mencionados, el mismo es más rápido y eficaz.

Existe una crítica a la integración de investigación, y es llamada sesgo de selección editorial a favor de estudios con resultados significativos. Este problema ha sido abordado por Rosenthal (153), pero la solución no es definitiva, la misma se utiliza para estimar la tolerancia (fail-safe number) de un meta-análisis a resultados nulos, es decir, se estima el número de resultados nulos que deben existir en los archivos editoriales para alterar los resultados significativos de un meta-análisis.

En el informe de cualquier meta-análisis se debe especificar claramente la fuente o fuentes de información utilizadas (base de datos electrónicas), periodos de búsquedas

y fecha de la última búsqueda, también si fuera el caso se deben incluir las búsquedas manuales realizadas.

Estrategia de Búsqueda

El procedimiento más utilizado es el efectuado en bases de datos electrónicas. El mismo no es tan simple, ya que existe una gran cantidad de revistas en las diferentes bases de datos. Por tal motivo debe realizarse una búsqueda en las bases de datos bibliográficas como MEDLINE, EMBASE, SCISEARCH. También existe otras bases de datos muy personalizadas como Current Contents, Cochrane Library, National Research Register, etc, y otras transversales como Web of Science, Martín Rodero (121).

Identificación y Selección de Estudios

Los parámetros de inclusión y exclusión de los estudios deben estar establecidos en un protocolo, los mismos deberán estar enmarcados dentro de los objetivos propuestos en el estudio. Se debe considerar, tamaño de la muestra, diseño del estudio, etc. Asimismo se debe decidir el número de evaluadores, la especialización de los mismos y la forma de resolver las disputas u opiniones contrarias que se produzcan. Es importante que en todas la fases de localización y selección de los estudios se debe documentar de manera correcta mediante un diagrama de flujo los artículos seleccionados en cada fase, Martín Rodero (121), note la Figura 2.5. Si existiera diferencia entre los evaluadores, se aplica el índice estadístico kappa, el mismo mide la concordancia de ambos evaluadores, es decir, hasta qué punto los evaluadores coinciden en su evaluación, Martín Rodero (121), el cual se puede interpretar de la forma que se explica en la Figura 2.6.

2. META-ANÁLISIS

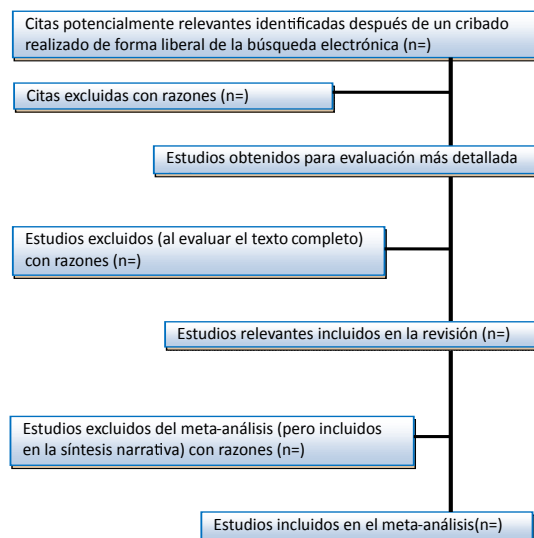


Figura 2.5: Selección de Estudio - Diagrama de flujo

Para evaluar la calidad de un meta-análisis, son elementos importantes: a) la evaluación de la validez de los estudios; b) la detección de la presencia de sesgos y c) la utilización de escalas de calidad de los estudios, Landis and Koch (104).

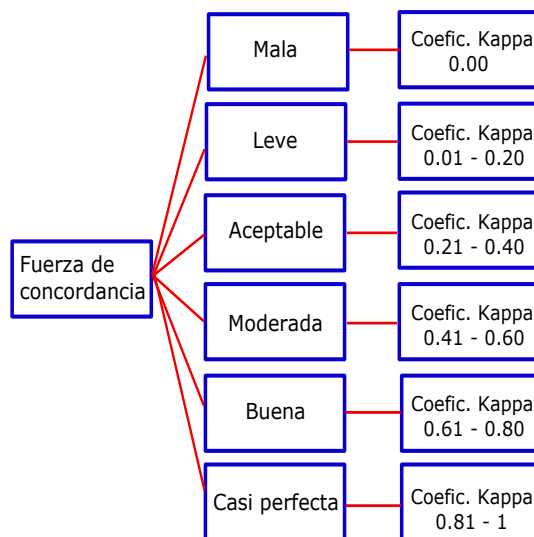


Figura 2.6: Interpretación coeficiente Kappa.

La validez de un meta-análisis debe garantizar la detección y eliminación de errores sistemáticos y sesgo. Existen dos tipos de validez, la interna y la externa. La validez interna se refiere aquello para lo cual ha sido diseñado, al referirse a los conceptos de

exactitud, fiabilidad y reproducibilidad. La validez externa es la facilidad de generalizar los resultados de un estudio a personas o poblaciones distintas.

En lo referente al sesgo, el mismo se relaciona con las debilidades de la validez interna de una investigación, se cuestiona si los resultados obtenidos son verdaderos en la población a la cual se pretende generalizar los resultados. En la práctica existen cuatro tipos de sesgo, Martín Rodero (121), sesgo de selección, de realización, de desgaste y de detección. El primer sesgo ocurre cuando existen diferencias en la asignación de los participantes al grupo de control y al de tratamiento; el sesgo de realización es causado por las diferencias de los cuidados que reciben los pacientes; el tercero ocurre cuando se dan bajas entre los participantes de estudios afectando los grupos de comparación y el último sesgo es ocasionado por los errores en la medición de los resultados entre los grupos que se comparan.

2.4.3. Codificación de Estudios

El revisor debe codificar las características de los estudios que están relacionados con los resultados del tópico de estudio. Glass distingue dos características, lo metodológico y lo sustantivo. La primera característica se refiere al número de pacientes en cada estudio, años de publicación del estudio, criterios de inclusión/exclusión, tipo de cegado, número de hombres/mujeres, edades de los pacientes (media, desviación típica, etc.) y la segunda característica considera aspectos como valoración de los análisis estadísticos, valoración del diseño, valoración de la presentación de los resultados, adecuación del tipo de cegado, etc.

Para formalizar el proceso de codificación se realiza un cuestionario que incluya las características relevantes, pero este proceso de codificación presenta un problema de medida, por lo que está sujeto a fiabilidad y validez del estudio. Para que el proceso de codificación sea coherente, se debe elaborar un manual de codificación que contenga lo más detallado posible las definiciones de las características, las normas y reglas de codificación. Este proceso debe ser conducido por expertos codificadores capacitados con anterioridad. El mismo evaluará una muestra aleatoria de todos los estudios incluidos en el meta-análisis. No existe un procedimiento único para codificar los estudios a combinar. Existe un formulario de manera general que puede contener variables, de Dios and Santamaría (31), como:

1. Identificación del estudio: autores, país, fuente de financiación.
2. Fecha de publicación del estudio.
3. Fecha de inicio y final del trabajo.
4. Problema estudiado o hipótesis formulada.

5. Características de los individuos: sexo, edad.
6. Diseño del estudio: experimental (con aleatorización o no) u observacional (prospectivo o retrospectivo).
7. Exposición o tratamiento estudiado: dosis, pauta.
8. Co-exposición o co-tratamiento.
9. Resultados que permiten medir los efectos principales y secundarios: resultados brutos y medidas del efecto ajustadas con sus intervalos de confianza o valores “p” asociados e interacciones representativas.
10. Valoración de la calidad o validez de los estudios.
11. Susceptibilidad del estudio a posibles sesgos, aspectos destacables y limitaciones del estudio, y explicaciones que los autores presentan de los resultados.

La parte fundamental de la codificación es materializar el conjunto de peculiaridades que tiene cada uno de los estudios recuperados, Martín Rodero (121). En el proceso de la codificación se pueden presentar problemas de fiabilidad y puede estar latente el sesgo. Para evitar la presencia de sesgo se recomienda que existan al menos dos codificadores. Mientras mayor sea la cantidad de codificadores, la fiabilidad efectiva se incrementa y por ende los errores aleatorios de cada codificador se anulan con los errores de los otros. Galindo-Villardón (59), Martín-Vallejo (122).

2.4.4. Medidas de los Resultados

Para la integración de los resultados de un conjunto de estudios es preciso que los mismos sean cuantificados en una escala de medida común, ya que si las variables dependientes empleadas en los diferentes estudios difieren entre sí, no son comparables entre sí. Es por eso que se aplican tres estrategias, la primera hace referencia a “vote counting”, la segunda el uso de niveles de significancia (p-valor) y la tercera emplea el uso de medidas de la magnitud del efecto (o tamaño del efecto - TE).

VOTE COUNTING

Los resultados encontrados se categorizan de tres maneras, Martín-Vallejo (122).

- Resultados significativos en los cuales la diferencia de medias es positiva.
- Resultados significativos en los cuales la diferencia de medias es negativa.
- Resultados no significativos.

En esta estrategia se examinan todos los estudios que aporten relevancia sobre una variable independiente de interés. Se tabula el número de estudios que incluyen dentro de cada una de las tres categorías citadas anteriormente. Si un gran número de estudios se ubica dentro de una de las categorías anteriores y muy pocas en las dos restantes, se infiere que esa categoría modal es asumida como la mejor estimación de la verdadera relación entre la variable dependiente y la independiente, Light and Smith (110). El inconveniente de este método es que puede llevar a conclusiones erróneas ya que no se toma en cuenta el tamaño de la muestra, la magnitud del efecto no es analizada, Martín Rodero (121).

Valor p

El valor “**p**” hace referencia a la probabilidad de que la diferencia observada en una muestra haya ocurrido al azar siendo los grupos comparados realmente semejantes, es decir, bajo la hipótesis nula, esta estrategia presente una debilidad, no proporciona información en relación a la magnitud del efecto experimental. La significancia estadística no provee inferencia sobre la intensidad de la relación (tamaño del efecto), por lo cual es interesante seleccionar un indicador que conlleve a la significancia clínica a través de la estimación de la magnitud del efecto.

Tamaño del Efecto

Las medidas más informativas del resultado de una investigación es el tamaño o magnitud del efecto, Green and Hall (68), pero fue Glass el responsable de introducir las medidas del TE en la integración cuantitativa de resultados de investigación.

El “**TE**” es un valor concreto, distinto de cero, cuanto mayor es este valor, tanto mayor es el grado en que se manifiesta el fenómeno bajo estudio, esto es cuando la hipótesis nula es falsa. El “**p**” valor permite determinar si un efecto experimental es distinto de cero, el TE permite conocer en que medida es distinto de cero. Para estimar el TE existen varios procedimientos, entre los cuales tenemos la diferencia de medias, diferencias de medias estandarizadas, Glass (66), etc. Dependiendo de la naturaleza de la investigación que se pretenda integrar, se puede hacer uso de varios índices o indicadores de la magnitud el efecto, Castillo (18), entre los más importantes se encuentran los siguientes:

1. Coeficiente de Correlación.- Es una medida estandarizada de la relación entre dos variables, es decir mide el grado de intensidad o de relación. Tenemos el coeficiente de correlación de Pearson y de Spearman. Estos índices pueden tomar valores entre **-1** hasta **+1**. Valores cercanos a **+1** indican una fuerte relación positiva o directa, valores cercanos a **-1** denota una relación negativa o inversa

2. META-ANÁLISIS

entre las variables, si el índice es igual a cero se puede inferir ausencia de relación lineal entre las variables.

2. Diferencia de Medias.- Estima la magnitud de la diferencia de medias, y se encuentran los índices de Cohen, Glass y Hedge, Martín-Vallejo (122).

Se puede calcular la diferencia de media y la medida asociada de precisión para cada estudio, Normand (137). Sea i el índice de estudio, T el tratamiento, C el grupo de control, n_{Ti} y n_{Ci} , los respectivos tamaños muestrales de las poblaciones.

Una posible medida es la diferencia de media $Y = \bar{X}_{Ti} - \bar{X}_{Ci}$ (diferencia de media del tratamiento i con el grupo de control i) con error estándar denotado por S_i y estimado como $S_i = S_{pi}^2 \left(\frac{1}{n_{Ti}} + \frac{1}{n_{Ci}} \right)$, donde $S_{pi}^2 = \frac{(n_{Ti} - 1)S_{Ti}^2 + (n_{Ci} - 1)S_{Ci}^2}{n_{Ti} + n_{Ci} - 2}$, nótese que S_{Ti}^2 y S_{Ci}^2 son las varianzas muestrales de los tratamientos y grupo de control respectivamente para i -ésimo estudio.

Otro estimador común para la medida del efecto es la diferencia de media estandarizada, la misma se estima como la diferencia de media dividida para la variabilidad de las medidas, es decir, si asumimos normalidad en los tratamientos y grupos de control con sus medias μ^T y μ^C respectivamente y varianzas iguales σ^2 , en otras palabras $Y_{ij}^T \sim N(\mu^T, \sigma^2)$ e $Y_{ij}^C \sim N(\mu^C, \sigma^2)$, se tiene el estimador insesgado $\delta = \frac{\mu^T - \mu^C}{\sigma}$, donde δ se puede interpretar como la pérdida o ganancia de la fracción de variabilidad de la medida. Un estimador de δ denotado por Hedges'g es definido como $h_i = \frac{\bar{Y}_i^T - \bar{Y}_i^C}{S_p}$, la razón de dividir por la desviación estándar es tener una medida resumen libre de unidades (adimensional) debido que el éxito en cada estudio se mide de diferente manera; donde el estimador de la varianza de h_i es $\frac{1}{n_{Ti}} + \frac{1}{n_{Ci}} + \frac{\hat{\delta}^2}{2(n_{Ti} + n_{Ci})}$, donde $\hat{\delta}^2$ es el estimador muestral de δ^2 .

3. Estimación de Riesgo.- La estimación se obtiene a partir de una tabla de doble entrada (2X2), entre los cuales tenemos diferencia de riesgo, el riesgo relativo, Odds-Ratio (OR) y el número necesario de pacientes a tratar (NNT). La diferencia de riesgo es fácil de interpretar, son proporciones (individuos en una población que tienen la enfermedad o el evento de interés en un momento dado), que pueden tomar valores de 0 hasta 1 (es decir, es la probabilidad de estar enfermo en el instante en que se mide la prevalencia) y siguen aproximadamente una distribución normal para tamaños muestrales pequeños o modestos, Normand (137).

Los riesgos relativos y odds ratios suelen ser analizados en la escala logarítmi-

ca, pero, a diferencia de la diferencia de riesgo, no se definen valores límites o fronteras. El riesgo relativo (RR), indica la probabilidad de que se desarrolle la enfermedad en los expuestos a un factor de riesgo, en relación al grupo de los no expuestos, si $RR = 1$ se infiere que no hay asociación entre la enfermedad y el factor de riesgo, si es mayor que 1 el factor de riesgo favorece la enfermedad y si RR es menor que 1 indica un factor protector, Martínez-González et al. (123).

Es importante notar que el RR no puede ser utilizado en estudios retrospectivos ya que no se conocen las probabilidades condicionadas de presentar la enfermedad ya que se fija de entrada los casos totales con enfermedad y los casos totales sin enfermedad y no es posible conocer la proporción real en los subgrupos estudiados. Odd-Ratio es una razón, en la cual el numerador es la probabilidad “p” de que ocurra un suceso y el denominador es la probabilidad de que el suceso no ocurra, toma valores entre cero e infinito, tiene como fin saber si un suceso es más probable que su complemento, Martínez-González et al. (123).

Por otro lado, NNT es el número de individuos que hay que atender con el tratamiento experimental para producir una mejoría respecto al control, es decir cuántos pacientes hay que tratar para obtener un resultado positivo; lo ideal sería que NNT sea igual a uno, es decir que todos mejoran con el tratamiento y ninguno con el control, cuanto más grande sea NNT o se aleje de la unidad menos eficaz es el tratamiento, Martínez-González et al. (123). Según, Normand (137) la diferencia de riesgo, el riesgo relativo y Odds ratio se definen como se observa en la Tabla 2.2.

	Diferencia de Riesgo	Riesgo Relativo	Odds Ratio
Parámetro	$D = P_T - P_C$	$R = \frac{P_T}{P_C}$	$\Omega = \frac{P_T(1 - P_T)}{P_C(1 - P_C)}$
Estimador	$d_i = \hat{p}_{Ti} - \hat{p}_{Ci}$	$r_i = \frac{\hat{p}_{Ti}}{\hat{p}_{Ci}}$	$\omega_i = \frac{\hat{p}_{Ti}\hat{q}_{Ci}}{\hat{q}_{Ti}\hat{p}_{Ci}}$
Error Estándar	$S_{d_i} = \sqrt{\frac{\hat{p}_{Ti}\hat{q}_{Ti}}{n_{Ti}}}$	$S \log(r_i) = \frac{\hat{q}_{Ti}}{n_{Ti}\hat{p}_{Ti}} + \frac{\hat{q}_{Ci}}{n_{Ci}\hat{p}_{Ci}}$	$S \log(\omega_i) = \sqrt{\frac{1}{n_a} + \frac{1}{n_b} + \frac{1}{n_c} + \frac{1}{n_d}}$

Tabla 2.2: Definiciones de riesgos y Odds Ratio

Donde T denota el tratamiento, C el grupo de control, $q_i = 1 - p_i$, n_{Ti} y n_{Ci} representan el número de pacientes en el grupo de tratamiento como de con-

trol, y a , b , c y d representan el número de observaciones en la tabla de doble entrada que presentan la presencia y ausencia de la enfermedad y factor de riesgo.

Otra manera de observar los riesgos es mediante una tabla de doble entrada (2x2), donde las columnas representan la presencia o ausencia de la enfermedad y las filas la presencia o ausencia del factor de riesgo según Martínez-González et al. (123), mayor detalle note Tabla 2.3.

	Enfermedad			
	Enfermos	No Enfermos		
Factor de Riesgo	Presente	VP (a)	FP (b)	VP + FP
	Ausente	FN (c)	VN (d)	FN + VN
		VP + FN	FP + VN	VP + FN + FP + VN

Tabla 2.3: Presencia o ausencia del factor de riesgo

Donde VP, FP, FN y VN, denotan los verdaderos positivos, falsos positivos, falsos negativos y verdaderos negativos respectivamente, luego el riesgo relativo se define como la ecuación 2.1.

$$RR = \frac{P(E|+)}{P(E|-)} = \frac{\frac{a}{a+b}}{\frac{c}{a+d}} = \frac{\text{Incidencia en expuesto}}{\text{Incidencia en no expuesto}} \quad (2.1)$$

Donde, la probabilidad condicional $P(E|+)$ se interpreta como la probabilidad de que un paciente esté enfermo cuando la prueba da positiva y $P(E|-)$ representa la probabilidad de que el paciente esté enfermo cuando el test es negativo. Con relación a los Odds Ratio, se define el odd ratio prospectivo y retrospectivo con las ecuaciones 2.2 y 2.3, Martínez-González et al. (123). Donde $P(No E|+)$ y $P(No E|-)$ denotan la probabilidad de no estar enfermo, dado que el test es positivo y la probabilidad de no estar enfermo dado que la prueba fue negativa respectivamente. Asimismo $P(+|No E)$ y $P(-|No E)$ representan la probabilidad que dio positivo en el test dado que no está enfermo y la probabilidad de que el test fue negativo dado que no está enfermo respectivamente.

$$OR = \frac{\frac{P(E|+)}{P(No E|+)}}{\frac{P(E|-)}{P(No E|-)}} = \frac{\frac{\frac{a}{a+b}}{\frac{b}{a+b}}}{\frac{\frac{c}{c+d}}{\frac{d}{c+d}}} = \frac{ad}{bc} = \frac{\text{Razón enfermos en los expuestos}}{\text{Razón enfermos en los no expuestos}} \quad (2.2)$$

$$OR = \frac{\frac{P(+|E)}{P(-|E)}}{\frac{P(+|No E)}{P(-|No E)}} = \frac{\frac{\frac{a}{a+c}}{\frac{c}{a+c}}}{\frac{\frac{b}{b+d}}{\frac{d}{b+d}}} = \frac{ad}{bc} = \frac{\text{Razón de expuestos al riesgo en enfermos}}{\text{Razón de expuestos al riesgo en sanos}} \quad (2.3)$$

2.4.5. Análisis e Interpretación de los Resultados

Una vez cuantificados los estudios mediante niveles de significancia o estimadores de TE, los resultados son transformados en un índice global, el cual debe ser representativo del conjunto de estudio. Por otro lado para combinar estadísticamente los resultados en un meta-análisis existen tres modelos, el modelo de efectos fijos (fixed effects), el modelo de efectos aleatorios (random effects) y el modelo de efectos mixtos (mixed effects).

Modelo de Efectos Fijos (fixed effects)

Asume que en cada estudio analizado se mide el mismo parámetro subyacente y que no hay variación en la medición entre los estudios Normand (137), es decir, las diferencias que se observan entre los estudios se deben únicamente a variabilidad aleatoria. Una definición matemática la encontramos en Normand (137) dice, que cada resumen estadístico Y_i es una realización de una población de estudio, la cual se estima con media θ , donde θ es el parámetro principal de interés y se asume que existen k estudios independientes ($i = 1, 2, \dots, k$). Se asume que $E[Y_i] = \theta$ y $S_i^2 = Var(Y_i)$ la varianza resumen del estudio i . Cuando se tienen tamaños de estudio moderadamente grandes Normand (137) cada Y_i se distribuye asintóticamente normal (por el teorema del límite

central) y aproximadamente insesgada, es decir $Y_i \sim N(\theta, S_i^2)$ para $(i = 1, 2, \dots, k)$ y asume conocida la varianza S_i^2 . El parámetro central de interés es θ que cuantifica el promedio de los efectos del tratamiento.

Modelo de Efectos Aleatorios(random effects)

No asume la homogeneidad entre los resultados de los estudios. Se suelen aplicar cuando los estudios tienen un grado de heterogeneidad superior al esperado por simple variabilidad aleatoria, Martín Rodero (121). Según Normand (137), en el modelo se asume que cada resumen estadístico del estudio Y_i analizado son independientes y pueden ser representados como una distribución normal con media θ_i y varianza S_i^2 , es decir $Y_i | \theta_i, S_i^2 \stackrel{ind}{\sim} N(\theta_i, S_i^2)$.

Del mismo modo, cada media de los estudios que se analizan θ_i , se asumen que proviene de una gran población normal de efectos de media θ y varianza τ^2 es decir $\theta_i | \theta, \tau^2 \stackrel{ind}{\sim} N(\theta, \tau^2)$, donde θ y τ^2 son llamados hiper-parámetros y representan la media del efecto del tratamiento y la variación inter-estudio respectivamente. Nótese, que dado los hiper-parámetros de la distribución de Y_i , después de promediar los efectos de cada estudio específico es normal con media θ y varianza $S_i^2 + \tau^2$. Como en el caso del modelo de efectos fijos, el parámetro central a estimar era θ , asimismo la variación inter-estudio τ^2 , juega un rol importante y debe ser estimada.

Adicionalmente la media del tratamiento del efecto es posible estimarla mediante los efectos de los estudios específicos, θ_i , los cuales son útiles para identificar particularmente la eficacia de los estudios. La distribución θ_i , condicionada a los datos observados y a los hiper-parámetros es denotada como en la ecuación 2.4, donde $Y = (Y_1, Y_2, \dots, Y_k)$ y B_i es definido como $\frac{S_i^2}{S_i^2 + \tau^2}$, denominado comúnmente “**shrinkage factor**”.

Cuanto más grande sea la variación inter-estudio τ^2 , menor será B_i de los efectos de los estudios observados, porque $0 \leq B_i \leq 1$ y la media de θ_i de la ecuación 2.4, es un compromiso de la media de los efectos del tratamiento θ y el resumen estadístico del estudio observado Y_i . Cuando $\tau^2 = 0$, “**shrinkage**” toma su valor máximo con $B_i = 1$ tal que $\theta_1 = \theta_2 = \dots = \theta_k = \theta$, entonces el modelo de efectos aleatorios corresponde al modelo de efectos fijos.

$$\theta_i | Y, \theta, \tau \sim N(B_i \theta + (1 - B_i) Y_i, S_i^2 (1 - B_i)) \quad (2.4)$$

Modelo de Efectos Mixtos (mixed effects)

El modelo de efectos aleatorios se basa en una hipótesis poco realista, asumiendo que los estudios provienen de una gran población de estudios y que la variabilidad entre estudios puede ser representada por una sola varianza. El modelo de efectos mixtos puede considerarse como una regresión de efectos aleatorios que explica la heterogeneidad adicional debido a variables externas por lo que, podría ser la solución al problema de la heterogeneidad, ya que este enfoque incluye una variación aleatoria adicional tomando en cuenta diferentes factores o variables que podrían explicar la variabilidad entre los estudios. De esta manera, la meta-regresión y la meta-partición proporcionan una alternativa para explorar las asociaciones entre las características y los resultados de los estudios.

Meta-Regresión

El término meta-regresión hace referencia a los análisis en los cuales las características de los estudios, o las características de los sujetos que intervienen en ellos se utilizan como variables independientes (explicativas) en una ecuación de regresión simple o múltiple, en la cual la variable dependiente (a explicar) es el resultado de cada estudio (efecto medio), Martín Rodero (121), Molinero (130).

Meta-Partición

Meta-partición es un procedimiento recursivo que identifica sub-grupos mutuamente excluyentes y exhaustivos de una población de tamaños de efectos Ortega et al. (139). Es un método que se utiliza en meta-análisis para explicar la heterogeneidad de los tamaños de efectos con variables moderadoras. Por lo tanto, se consideran dos tipos de variables: (1) las variables de respuesta que proporcionan las estimaciones de los tamaños del efecto, y (2) los factores que influyen en los tamaños del efecto, que pueden ser variables cualitativas o cuantitativas. El criterio consiste en ajustar la heterogeneidad de los tamaños muestrales de los efectos y particionar esta variabilidad mediante un algoritmo que encuentre el mejor moderador de cada partición. El mejor moderador es el que explica la mayor cantidad de variabilidad entre los subconjuntos posibles de cada partición Ortega et al. (139), note Figura 2.7.

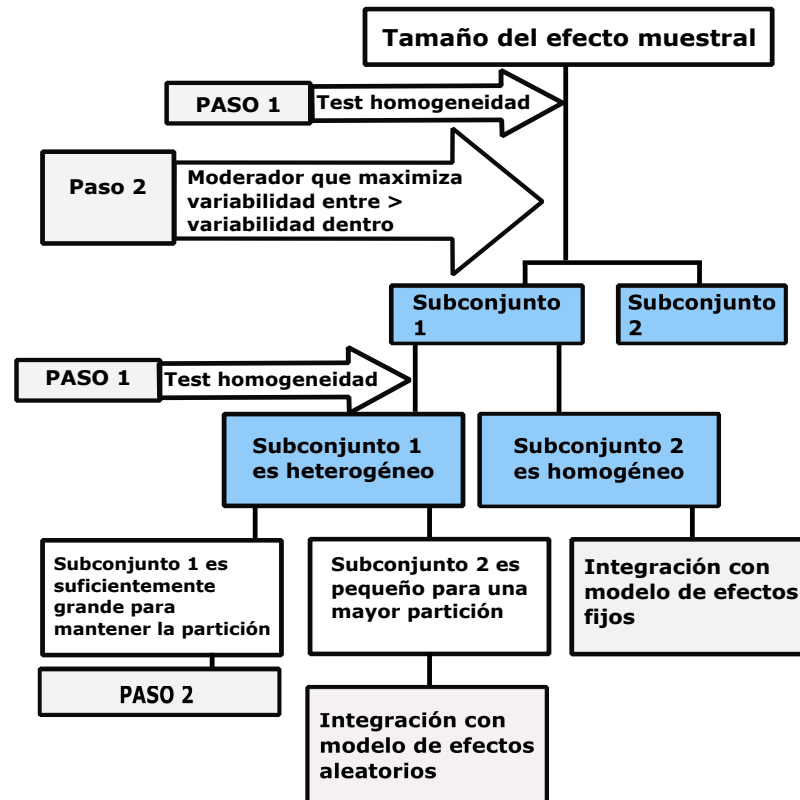


Figura 2.7: Diagrama que resume la metodología meta-partición.

El paso 1 consiste en la evaluación de la heterogeneidad con la prueba de homogeneidad y el valor de I^2 . El paso 2 consiste en seleccionar el mejor moderador de acuerdo con el valor de p , y la etapa 3 implica integrar la magnitud del efecto en los subconjuntos finales.

Como se sabe la medicina es una ciencia que involucra probabilidades y el manejo de la incertidumbre, la misma se extiende no sólo a las actividades preventivas, terapéuticas y pronósticos sino también a las diagnósticas, es por eso que en el Capítulo 3 se analizan los principales conceptos de pruebas diagnósticas desde un punto de vista estadístico.

Meta-Análisis en Pruebas Diagnósticas

3.1. Pruebas Diagnósticas

Las pruebas diagnósticas son pruebas para determinar si un individuo tiene o no cierta enfermedad (u otra característica). Las estimaciones numéricas que se obtengan de la misma pueden depender de la prevalencia (proporción de individuos que tienen la enfermedad o el evento de interés en el momento dado) y de las características clínicas de los pacientes que se hayan estudiado.

3.1.1. Clasificación de las enfermedades por diagnosis

Dependiendo del grado de avance de la enfermedad en el paciente, las enfermedades pueden clasificarse en tres tipos básicos.

- **Tipo I.-** La peor equivocación que se pueden cometer en el diagnóstico es un Falso Negativo. En esta clasificación están las enfermedades que son curables si se detectan a tiempo.
- **Tipo II.-** La equivocación que se puede cometer en el diagnóstico es un Falso Positivo, es decir la enfermedad es grave, pero difícilmente curable o sin remisión.
- **Tipo III.-** Las enfermedades que no pueden ser clasificadas como de Tipo I o Tipo II. Ejemplo el SIDA.

3.1.2. Criterios de calidad de las pruebas

La calidad de la prueba, es la capacidad de la prueba para medir efectivamente lo que se quiere medir. Una prueba diagnóstica es considerada de buena calidad si es capaz de clasificar correctamente a los miembros de la población a la que se aplica, dando resultados positivos en las personas enfermas y negativas en los carentes de la enfermedad de estudio. La calidad de una prueba diagnóstica se mide en términos de validez y seguridad.

- **Validez.-** Compara la prueba con otra prueba de probada eficacia en la enfermedad de estudio, a este proceso de comparación se le llama Patrón Áureo (Gold Estándar) o criterio de referencia, Pita Fernández and Pértegas Díaz (145).
- **Seguridad.-** Capacidad para predecir la presencia o ausencia de la enfermedad y se mide en términos de los valores predictivos positivos y negativos, Pita Fernández and Pértegas Díaz (145), observe Figura 3.1.

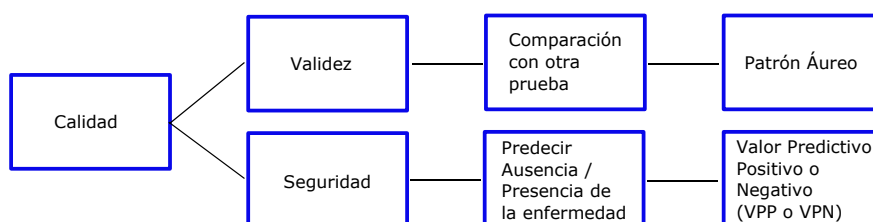


Figura 3.1: Criterio de calidad de las pruebas

Es importante notar que existen tres tipos de pruebas:

- Las que tienen respuesta dicotómica (Positiva +, Negativa -), ejemplo, pruebas de embarazo.
- Las pruebas que poseen respuesta policotómica, ejemplo, estudios del cáncer.
- Las que poseen respuesta cuantitativa, ejemplo, prueba Glucosa/diabetes.

3.1.3. Diseño de una investigación para evaluar una prueba diagnóstica

En este punto, se trata de realizar una investigación descriptiva, en la cual un grupo de participantes con características determinadas acorde a la prueba a probar es sometido a la prueba y concomitantemente al patrón áureo, note Figura 3.2. En toda investigación, existe sesgo, para minimizar el mismo se deben considerar, tres aspectos

importantes: la selección de la muestra, comparación e interpretación de los resultados.

- **Selección de la muestra.-** La muestra debe contemplar participantes que caractericen todas las formas clínicas del daño buscado, guardando la proporción en que la misma se presenta en la práctica cotidiana.
- **Comparación.-** Debe establecerse con un patrón de referencia confiable y apropiado. El patrón áureo debe reunir esas características y a todos los participantes se le deben realizar las dos pruebas (la que se está investigando y el patrón áureo). Al utilizar una prueba costosa como patrón áureo se imposibilita llenar las casillas c y d de la tabla de doble entrada, por lo que no es posible estimar la sensibilidad y la especificidad, a este error de procedimiento (error sistemático o sesgo) se lo denomina error de verificación (work-up bias).
- **Interpretación de los resultados.-** La lectura de los resultados de la prueba analizada debe ser efectuada, sin tener conocimiento de los resultados del patrón áureo, para minimizar o evitar el sesgo hacia una coincidencia espuria.

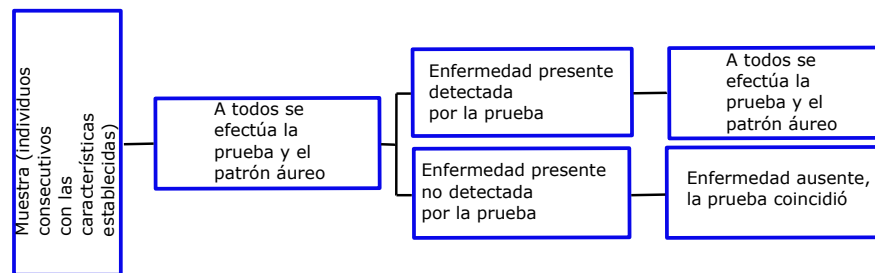


Figura 3.2: Diseño para evaluar una prueba diagnóstica

3.1.4. Parámetros de valoración de pruebas diagnósticas

Supóngase que se quiere evaluar una prueba analítica para el diagnóstico de una enfermedad, obteniéndose los resultados, que se observan en la Tabla 3.1. Los 27 pacientes que padecen la enfermedad dieron positivo en la prueba, se consideran verdaderos positivos (VP). Los 6 pacientes que están enfermos y que dieron negativo en la prueba de laboratorio son considerados falsos negativos (FN). Los tres pacientes que no tienen la enfermedad y que sus pruebas clínicas denotan un resultado positivo se denominan falsos positivos (FP). Y los 164 pacientes que no tienen la enfermedad y cuyos resultados de laboratorio fueron negativos se catalogan como verdaderos negativos (VN). A continuación se procederá a definir los parámetros de valoración de una prueba de diagnóstico.

1. **Sensibilidad.-** Mide el porcentaje de individuos enfermos correctamente diagnosticados, es decir, es la probabilidad de que el test sea positivo en presencia de

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

la enfermedad, Burgueño et al. (15).

Definamos los siguientes eventos:

- **E:** Individuo con la condición de interés
- **NE:** Individuo sin la condición de interés
- **T:** Prueba diagnóstica
- **T+:** Resultado positivo del test
- **T-:** Resultado negativo del test
- **P(E):** Probabilidad a priori de que un paciente elegido al azar tenga la enfermedad, también denominada prevalencia.

	Enfermedad			
		Enfermos	No Enfermos	
Prueba Diagnóstica	Presente	27 (VP) a	3 (FP) b	30 (TP)
	Ausente	6 (FN) c	164 (VN) d	170 (TN)
		33 (TE)	167 (TNE)	200 (T)

Tabla 3.1: Tabla de doble entrada, ejemplo hipotético

La sensibilidad se puede medir como; $Se = P(T + |E)$, es decir la probabilidad de que el test sea positivo dado que el paciente está enfermo. En otras palabras, mide el porcentaje de individuos enfermos correctamente diagnosticados, e indica la capacidad de la prueba para detectar a un sujeto enfermo. Siguiendo con la Tabla 3.1 la sensibilidad será:

$$Se = \frac{VP}{TE} \quad (3.1)$$

En el ejemplo, 81,8% de los pacientes enfermos dieron positivo en la prueba, por lo que habrá un total de 18,2% de falsos negativos.

2. **Especificidad.-** Es el porcentaje de resultados negativos en pacientes que no padecen la enfermedad, es decir, mide la probabilidad de que el test sea negativo en individuos sanos, $Es = P(T - |NE)$.

En otros palabras, valora la capacidad de una prueba para detectar correctamente individuos sanos, una alta especificidad indica una baja frecuencia de falsos positivos, Burgueño et al. (15); siguiendo el ejemplo de la Tabla 3.1 la especificidad se pueden medir como:

$$Es = \frac{VN}{TNE} \quad (3.2)$$

El 98,2% de individuos sanos dieron negativo en la prueba, y habrá un 1,8% de falsos positivos. Queda claro que la calidad de la prueba puede ser representada por la sensibilidad y especificidad del test, pero en la práctica clínica lo que interesa es responder a las preguntas, ¿cuál es la probabilidad de que esté verdaderamente enfermo si un sujeto ha resultado positivo, $P(E|T+)$ y por el contrario, ¿cuál es la probabilidad de que realmente esté sano si el sujeto resultó negativo en la prueba $P(NE|T-)$, estas dos probabilidades se pueden calcular utilizando el Teorema de Bayes, siempre y cuando se conozca la prevalencia de que un sujeto este enfermo antes de ser sometido al test.

Para encontrar estas probabilidades denominemos P la probabilidad pre-prueba de padecer la enfermedad, Se la sensibilidad de la prueba y Es la especificidad de la misma, lo cual permitirá estimar la probabilidad post-prueba de que el sujeto esté enfermo cuando resultó positivo o de que esté sano cuando resultó negativo.

3. **Valor Predictivo Positivo (VPP).**- Es el porcentaje de pacientes enfermos entre todos los pacientes con resultados positivos, Burgueño et al. (15). Valora la probabilidad de que una prueba positiva diagnostique correctamente a un individuo enfermo, luego el VPP queda definido como se nota en la ecuación 3.3.

$$P(E|T+) = \frac{P(T+|E) \times P(E)}{P(T+|E) \times P(E) + P(T+|NE) \times P(NE)} = \frac{Se \times P(E)}{Se \times P(E) + (1 - Se) \times (1 - P(E))} = \frac{VP}{VP + FP} \quad (3.3)$$

$$P(E|T+) = \frac{VP}{TP}$$

4. **Valor Predictivo Negativo (VPN).**- Indica la frecuencia de pacientes no enfermos entre todos los pacientes con resultado negativo Burgueño et al. (15). Pondera la probabilidad de que una prueba negativa diagnostique correctamente a un individuo sano, note ecuación 3.4.

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

$$P(NE|T-) = \frac{P(T-|NE) \times P(NE)}{P(T-|NE) \times P(NE) + P(T-|E) \times P(E)} = \frac{Es \times P(NE)}{Es \times P(NE) + (1 - Es) \times (1 - P(E))} = \frac{VN}{VN + FN} \quad (3.4)$$

$$P(NE|T-) = \frac{VN}{TN}$$

Luego si seguimos con el ejemplo de la Tabla 3.1, tenemos que, $VPP = \frac{27 \times 100}{30} = 90\%$ y $VPN = \frac{164 \times 100}{170} = 96,5\%$. Un VPP de 90% indica que de cada 100 pacientes que dan la prueba positiva 90 padecen la enfermedad. Un VPN de 96,5% indica que de cada 100 pruebas negativas 96,5 pertenecen a individuos sanos.

En resumen un test altamente sensible proporciona pocos falsos negativos y es bueno para descartar, de manera análoga un test altamente específico proporciona pocos falsos positivos y es bueno para confirmar. Es importante conocer las estimaciones para los intervalos de confianza de los valores predictivos positivo y negativo, los cuales se estiman con las expresiones 3.5 y 3.6, donde $\varepsilon = 1,96$.

$$VPP \pm \varepsilon \times \sqrt{\frac{VPP \times (1 - VPP)}{VP + FP}} \quad (3.5)$$

$$VPN \pm \varepsilon \times \sqrt{\frac{VPN \times (1 - VPN)}{FN + VN}} \quad (3.6)$$

Los valores predictivos dependen mucho de la prevalencia de la enfermedad en la población (cociente entre enfermos y población), no así la sensibilidad y especificidad, que son independientes de la misma. VPP y VPN interesan en la práctica clínica, pero presentan un inconveniente, es que dependen de la proporción de enfermos en la muestra estudiada, por lo que es necesario utilizar otros tipos de índices como las razones de verosimilitud que no dependan de esa proporción de enfermos.

5. **Punto de Corte.-** Para entender la elección de los puntos de corte en las pruebas diagnósticas es necesario introducir el concepto de coste (o utilidad) en la realización de las pruebas. El costo puede ser interpretado en un sentido estricto

de dinero o días perdidos por pacientes saludables, pero también se puede pensar en una combinación de varios componentes Metz (127).

El costo medio (promedio) de la realización de una prueba diagnóstica debe incluir, en primer lugar, el precio promedio que se debe pagar para la realización de la prueba; a este costo lo denotaremos por C_o . Al coste C_o hay que añadirle los costes médicos de cada tipo de decisión diagnóstica.

Como se trata de estimar un coste medio, hay que ponderar la decisión de cada uno de estos costos mediante las probabilidades involucradas en cada tipo de decisión en cuestión. Por lo que el coste medio \bar{C} resultante de la prueba diagnóstica queda expresado en la ecuación 3.7, Metz (127).

$$\bar{C} = C_o + C_{VP}P(VP) + C_{VN}P(VN) + C_{FP}P(FP) + C_{FN}P(FN) \quad (3.7)$$

Donde P , representa la probabilidad de clasificar un caso como verdadero positivo - $P(VP)$, verdadero negativo - $P(VN)$, falso positivo - $P(FP)$ y falso negativo - $P(FN)$. C representa el coste promedio de una decisión medica de clasificar un paciente como, verdadero positivo - C_{VP} , verdadero negativo - C_{VN} , falso positivo - C_{FP} y falso negativo - C_{FN} .

La probabilidad $P(VP)$ puede ser reemplazada por $P(E)P(T+|E) = Se \cdot P(E)$. Similarmente, $P(VN) = P(NE)P(T-|NE) = Es \cdot P(NE)$; $P(FN) = P(NE)P(T-|E) = (1 - Se) \cdot P(E)$; y $P(FP) = P(NE)P(T+|NE) = (1 - Es) \cdot P(NE)$.

Además, $P(T-|E) = 1 - P(T+|E)$ y $P(T-|NE) = 1 - P(T+|NE)$. Por lo tanto la expresión 3.7 puede ser expresada como la ecuación 3.8.

$$\begin{aligned} \bar{C} = & - \{[C_{FN} - C_{VP}]P(E)\}P(T+|E) + \{[C_{FP} - C_{VN}]P(NE)\}P(T+|NE) + \\ & + \{C_o + C_{VN}P(NE) + C_{FN}P(E)\} \end{aligned} \quad (3.8)$$

De la ecuación 3.8 se puede inferir que \bar{C} depende proporcionalmente de C_o y de las probabilidades $P(T+|E)$ y $P(T+|NE)$ que representan a la Se y $(1 - Es)$ las cuales son las coordenadas de una curva ROC.

Dado que el tomador de decisiones puede cambiar estas cantidades en conjunto, es decir, puede mover la Se y $(1 - Es)$ a lo largo de la curva ROC mediante el uso

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

de umbrales diferentes, el costo promedio dependerá del umbral utilizado. Por lo tanto, en términos de análisis coste-beneficio el mejor punto de funcionamiento en una curva ROC se obtiene minimizando el coste medio (o maximizando los beneficios) en una situación particular de diagnóstico.

Si se deriva implícitamente \bar{C} en términos de $P(T+|NE)$, es decir $P(T+|E)$ es una función de $P(T+|NE)$ e igualando a cero se obtiene la ecuación 3.9.

$$\begin{aligned} \frac{d\bar{C}}{dP(T+|NE)} &= -\{[C_{FN} - C_{VP}]P(E)\} \frac{dP(T+|E)}{dP(T+|NE)} + \{[C_{FP} - C_{VN}]P(NE)\} = 0 \\ &\text{despejando } \frac{dP(T+|E)}{dP(T+|NE)} \\ &\therefore \\ \frac{dP(T+|E)}{dP(T+|NE)} &= \frac{C_{FP} - C_{VN}}{C_{FN} - C_{VP}} \cdot \frac{P(NE)}{P(E)} \end{aligned} \tag{3.9}$$

Para considerar el efecto de la prevalencia en el espacio ROC, supóngase que se estudia una enfermedad rara en una población cualquiera, es decir, $P(E)$ es cercano a 0, por lo que $P(NE)$ será cercana a uno. Luego el tomador de decisión deberá trabajar en la parte inferior izquierda de la curva ROC donde TVP es pequeña pero TFP es aún más pequeña. Pero si se considera una enfermedad común $P(NE)$ será cercana a 0, y el mejor punto de operación de la curva ROC se ubicará en la parte superior derecha del espacio ROC - donde TVP es alta pero TFP es aún más.

Si se analizan los costos de la ecuación 3.9 y se nota que $C_{FP} - C_{VN}$ es mayor que $C_{FN} - C_{VP}$ (cuando se aplica un tratamiento o varias pruebas diagnósticas a pacientes saludables, siendo estas medidas perjudiciales a los mismos pero de poco beneficio a pacientes enfermos), entonces el valor de la pendiente es grande. A la inversa, si $C_{FP} - C_{VN}$ es menor que $C_{FN} - C_{VP}$, es decir, se aplicó un tratamiento o algunas pruebas a pacientes sanos, pero las mismas fueron inofensivas para estos pacientes y beneficiosas a pacientes enfermos, entonces el tomador de decisión debería operar en la parte superior derecha del espacio ROC, donde la TVP debería mantenerse con valores altos e incluso a expensas de la TFP .

6. **Razones de Verosimilitud.-** Al cociente entre la tasa de verdaderos positivos y la tasa de falsos positivos se le denomina cociente de probabilidad positiva (**LR+**); también recibe el nombre de cociente de probabilidad positivo o cociente de verosimilitud de una prueba positiva o poder predictivo de una prueba positiva; indica mejor capacidad para diagnosticar la presencia de la enfermedad en

pacientes que presentan la alteración en comparación con los que no tienen la enfermedad (deseable que sea lo mayor posible).

Al cociente entre la tasa de verdaderos negativos y la tasa de falsos negativos se le denomina cociente de probabilidad negativo (**LR-**) también conocido como cociente de verosimilitud negativo o índice de eficiencia pronóstico de una prueba negativa. Indica cuanto menos probable es un resultado negativo en los enfermos en comparación con los pacientes que no presentan la alteración; el inverso de este cociente se llama poder (o potencia) predictiva de una prueba negativa, note ecuaciones 3.10 y 3.11 respectivamente.

$$LR+ = \frac{P(T+|E)}{P(T+|NE)} = \frac{Se}{1 - Es} \quad (3.10)$$

$$LR- = \frac{P(T-|E)}{P(T-|NE)} = \frac{1 - Se}{Es} \quad (3.11)$$

La importancia o utilidad de los cocientes $LR+$ y $LR-$ frente a los VPP y VPN de la prueba se fundamenta en que estos LR no dependen de la proporción de enfermos en la muestra sino de la sensibilidad y especificidad.

Por otro lado si se estima la probabilidad pre-prueba de que un sujeto padezca la enfermedad, utilizando los cocientes de probabilidad, al realizar la prueba se podrá corregir ese valor, de tal manera que la probabilidad aumente o disminuya según que el resultado sea positivo o negativo, aplicando la fórmula 3.12, donde P es la probabilidad a priori o pre-prueba de la enfermedad, LR es el cociente de probabilidad (positivo en el caso de que deseáramos calcular la probabilidad de que padezca la enfermedad o negativo en caso contrario) y P_{post} es la probabilidad post prueba.

$$P_{post} = \frac{P \times LR}{1 + P \times (LR - 1)} \quad (3.12)$$

7. **Variabilidad de las razones de verosimilitud.**-Si la especificidad es constante, las razones de verosimilitud positivas $LR+ = \frac{Se}{1 - Es}$ son función lineal de la sensibilidad. Cuando la sensibilidad aumenta, $LR+$ aumenta y viceversa, Pita Fernández and Pértegas Díaz (145). La razón de verosimilitud negativa $LR- = \frac{1 - Se}{Es}$ es igualmente una función lineal de sensibilidad pero negativa, cuando aumenta la sensibilidad disminuye $LR-$, Pita Fernández and Pértegas Díaz (145).

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

Nótese que, cuando $Se = 1 - Es = 0,1$, Figura 3.3, las dos razones de verosimilitud son iguales a 1. Luego la prueba no tiene ningún valor de diagnóstico. El valor diagnóstico del resultado positivo o negativo de la prueba crece linealmente cuando la sensibilidad aumenta de 0,1 a 1.

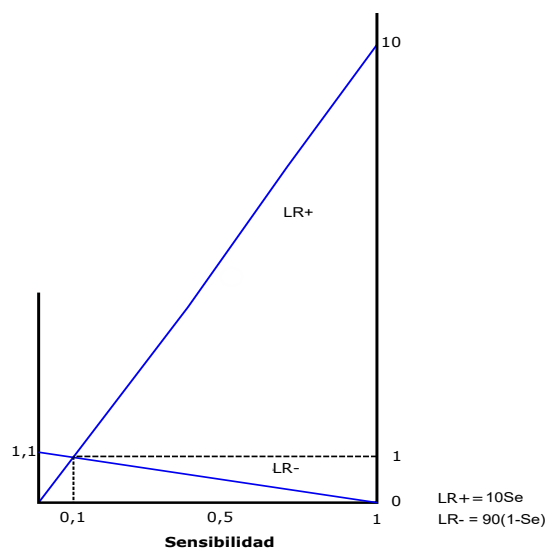


Figura 3.3: Variación de la razón de verosimilitud en función de la sensibilidad para diferentes pruebas cuya especificidad es igual a 0,90

En la Figura 3.4, se pone de manifiesto que cuando $Es = 1 - Se = 0,8$, las dos razones de verosimilitud son iguales a 1 y por lo tanto la prueba carece de utilidad diagnóstica. Se nota que la progresión de $LR+$ es muy rápida cuando la especificidad es superior a 0,7. Se dice que la eficacia diagnóstica de una prueba en la investigación de una enfermedad E está ligada al crecimiento de su especificidad.

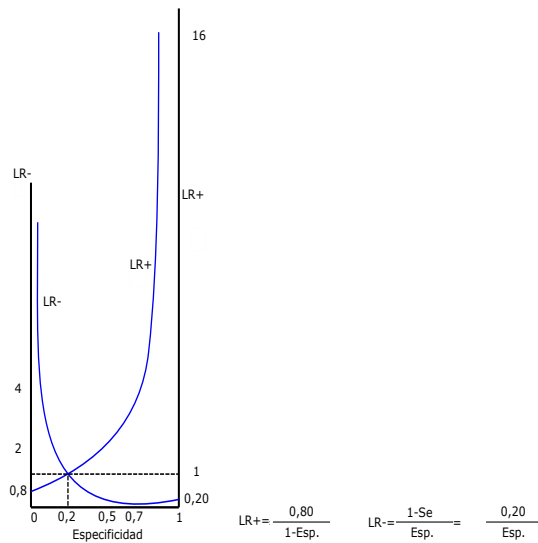


Figura 3.4: Variación de las razones de verosimilitud en función de la especificidad para diferentes pruebas cuya sensibilidad es igual a 0.80

Comparando los cocientes de probabilidades de dos pruebas diagnósticas existen 4 posibles resultados que se pueden visualizar en un gráfico. En el eje Y ubicamos la sensibilidad y el eje X se representa $1 - \text{especificidad}$. Supóngase ahora que se quiere comparar la prueba A con la prueba B , cuya sensibilidad es de 0,9 y especificidad 0,5. Las zonas de la Figura 3.5, corresponden a las siguientes situaciones:

1. En esta región la prueba B es siempre superior a la A tanto para confirmar presencia como ausencia de enfermedad.
2. Prueba B solo es superior a la A para confirmar ausencia de enfermedad.
3. Prueba B siempre es inferior a A
4. B será superior a A para confirmar la presencia de la enfermedad.

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

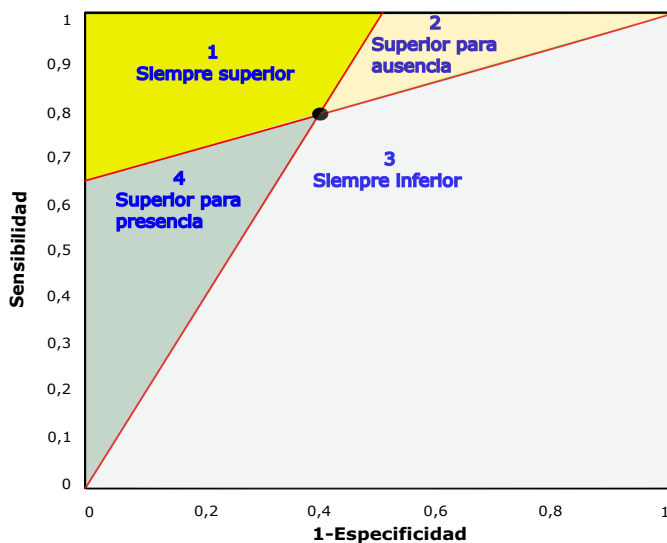


Figura 3.5: Especificidad Vs Sensibilidad

Si se desea comparar la prueba *A* con otra prueba *B* cuya sensibilidad es 0.9 y especificidad 0.5, dichas pruebas pueden ser representadas mediante la Figura 3.6, obteniendo que *B* cae en la zona 2, lo cual indica que es una prueba mejor que *A* para validar o confirmar ausencia de la enfermedad.

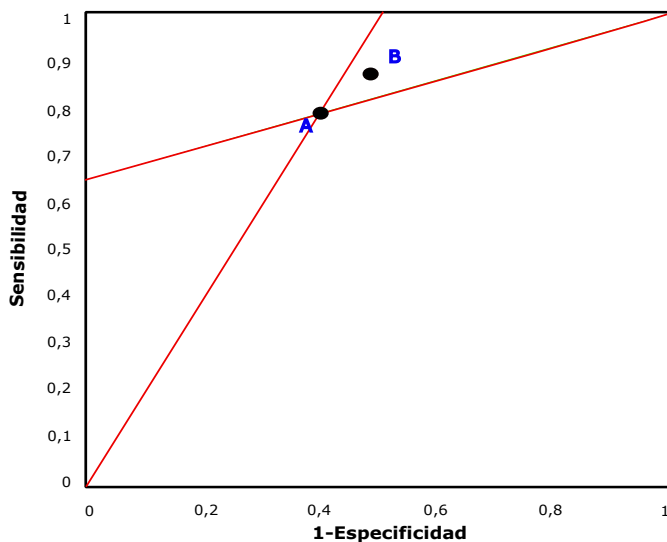


Figura 3.6: Especificidad Vs Sensibilidad

Para comparar los parámetros de dos pruebas diagnósticas hay que considerar al-

gunos criterios que las estimaciones realizadas podrán estar sometidos a posibles errores de muestreo, con lo cual habrá que realizar un contraste de hipótesis estadística para inferir si las diferencias encontradas son suficientemente importantes y no se deben al azar. Si la razón de verosimilitud se acerca a 1, la probabilidad diagnóstica es la misma antes y después de la prueba. Bajo este escenario el test es inútil y no tiene capacidad discriminatoria. Cuanto más se aleje de la unidad la razón de verosimilitud, más interés presenta la prueba. Según Evidence-bases Medicine Working Group, los cocientes de probabilidad tienen la interpretación que se observa en la Figura 3.7.

<p>LR+ > 10 o LR- < 0.1</p> <p>Generan cambios amplios y a menudo concluyentes desde una probabilidad pre-prueba hasta una probabilidad post-prueba</p>	<p>5 < LR+ < 10 y 0.1 < LR- < 0.2</p> <p>Generan cambios moderados desde la probabilidad pre-prueba hasta la probabilidad post-prueba</p>	<p>2 < LR+ < 5 y 0.05 < LR- < 0.2</p> <p>Generan cambios pequeños (pero en ocasiones importantes) de la probabilidad</p>	<p>1 < LR+ < 2 y 0.5 < LR- < 0.1</p> <p>Alteran la probabilidad en un grado insignificante (y rara vez importante)</p>
--	--	---	---

Figura 3.7: Cocientes de Probabilidad, según Evidence-bases Medicine Working Group

Las ventajas de utilizar o emplear la razón de verosimilitud, es que es independiente de la prevalencia, reflejando de manera correcta el valor de la prueba independientemente del grupo al que se le aplique, resumiendo la información en un solo parámetro.

Los cocientes de probabilidad permiten transformar la probabilidad pre-prueba a post-prueba mediante la relación $odds_{post} = LR_T \times odds_{pre}$, donde odds es el cociente entre la probabilidad de que ocurra un evento y la probabilidad de que no ocurra, por lo tanto se tiene la ecuación 3.13.

$$odds = \frac{P}{1 - P} \tag{3.13}$$

$$P = \frac{odds}{1 + odds}$$

Los sub-índices post y pre señalan la post-prueba y pre-prueba y el sub-índice T representa el resultado obtenido positivo o negativo. Fagan (52), describió un

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

nomograma para el teorema de Bayes, basado en la capacidad de convertir al teorema en una función sumatoria lineal simple. El mismo es una herramienta que permite calcular la probabilidad post-prueba una vez que se conoce la probabilidad pre-prueba y la razón de verosimilitud. El nomograma tiene tres columnas, en la primera columna se tiene la probabilidad pre-prueba, la segunda o la del centro es la razón de verosimilitud y en la columna de la derecha se localiza la probabilidad post-prueba.

Al observar la Figura 3.8 podemos notar que el médico ha establecido un diagnóstico de sospecha de 25 %, es decir la probabilidad a priori, luego si un paciente obtuvo un test positivo con el nomograma podemos inferir que tendría la enfermedad con una probabilidad del 84 %. Por otro lado, si el test fue negativo, se puede inferir que el paciente no tendrá la enfermedad con una probabilidad del 9 %. En otras palabras con el nomograma se puede notar la probabilidad post-prueba.

Un ejemplo del uso del nomograma con tres pruebas para el diagnóstico de Sepsis se lo observa en Seqc (172). Donde se detalla como a partir de una sospecha de la enfermedad (probabilidad a priori) se pasa a una probabilidad post-prueba e indicando además que prueba es más efectiva en la detección de la alteración.

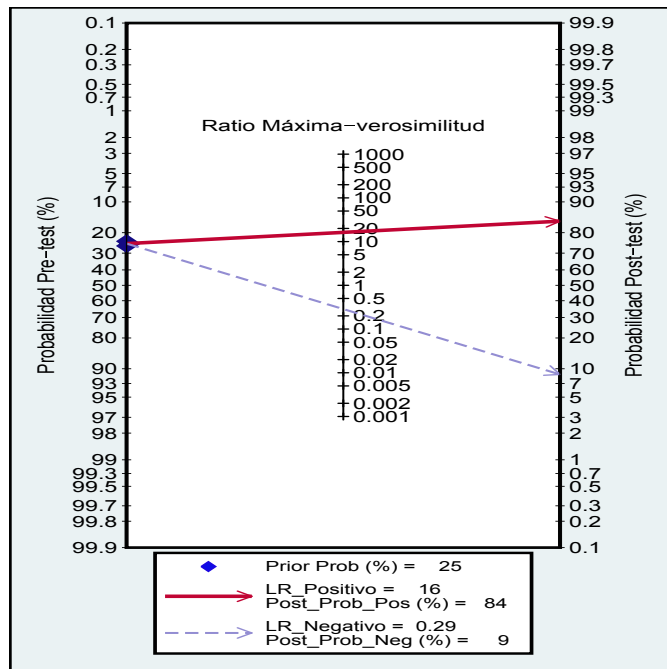


Figura 3.8: Gráfica de Fagan, datos midas_example

Para entender el Nomograma se debe entender la versión de la probabilidad del

Teorema de Bayes, donde E , representa enfermedad, \bar{E} , significa ausencia de enfermedad y T es un resultado específico de la prueba; es decir, $Odds\ postprueba = Odds\ preprueba \times LR$, una manera de interpretarla matemáticamente es la que se observa en la ecuación 3.14.

$$\begin{aligned} O(E|T) &= O(E) \times \frac{P(T|E)}{P(T|\bar{E})} = \text{si tomamos logaritmo tenemos } \log(\text{Odds postprueba}) \\ &= \log(O(E|T)) = \log(Oddspreprueba) \times \log(LR) \\ &= \log(O(E)) + \log\left(\frac{P(T|E)}{P(T|\bar{E})}\right) = P(preprueba) \times LR \end{aligned}$$

\therefore

$$\log(O(E|T)) = \log(P(preprueba)) + \log(LR) \quad (3.14)$$

8. **Odds Ratio Diagnóstica.-** Es un estadístico utilizado en los estudios epidemiológicos que mide la fuerza de asociación entre el factor de riesgo y la enfermedad. En nuestro caso es útil para denotar el grado de intensidad entre el resultado de una prueba y la enfermedad, el índice no se ve influenciado por la prevalencia y se estima como la razón entre odds de estar enfermo si la prueba da positivo y la odds de no estar enfermo si da negativo. Su estimación se obtiene mediante la ecuación 3.15.

$$DOR = \frac{\frac{VP}{FN}}{\frac{FP}{VN}} = \frac{LR+}{LR-} = \frac{\frac{VPP}{1-VPP}}{\frac{1-VPN}{VPN}} \quad (3.15)$$

Los valores DOR pueden variar desde cero hasta el infinito (cuanto más alto o grande sea DOR , mejor será la prueba). Si $DOR = 1$, es decir, no es discriminante la prueba, la misma es inútil. Si $DOR \geq 1$, es más probable que la prueba de positivo en el caso de enfermos que en sanos, Seqc (172).

9. **Eficiencia diagnóstica de una prueba.-** Representa la proporción de resultados correctos entre el total de resultados. Es la probabilidad de que un paciente sea clasificado correctamente por la prueba; es decir, en el numerador aparecerán los enfermos con resultado positivo y los sanos con test negativo, mientras que en el denominador se encuentran todos los sujetos o pacientes. La eficiencia se puede categorizar en tres tipos: eficiencia, eficiencia si se considera la prevalencia y eficiencia estándar, nótese ecuaciones 3.16, 3.17 y 3.18 respectivamente.

$$\frac{VP + VN}{VP + VN + FP + FN} = \frac{a + d}{a + b + c + d} \quad (3.16)$$

$$\frac{P(E) \times Se}{1 - P(E) \times Es} \quad (3.17)$$

$$\frac{Se + Es}{2} \quad (3.18)$$

10. **Curvas “receiver-operating characteristics”(ROC).**- Si el resultado de una prueba es una variable continua (cantidad en la sangre de glucemia en mg/dl), hay que decidir cuál resultado es positivo para obtener el diagnóstico apropiado; entonces se debe establecer un punto de corte. El mismo determinará la sensibilidad y la especificidad de la prueba, Burgueño et al. (15). Para establecer dicho punto de corte se recurre a las curvas de rendimiento diagnóstico (ROC). Es un gráfico que relaciona la sensibilidad y la especificidad. En el eje Y se ubican la sensibilidad o los verdaderos positivos, en el eje X se sitúan los falsos positivos ($1 - Es$).

Una prueba con discriminación perfecta, posee una curva ROC que pasa por la esquina superior izquierda, donde Se y Es toman sus máximos valores ($Se = Es = 1$). Una curva ROC que tiene una línea diagonal de 45° , desde la esquina inferior izquierda hasta la esquina superior derecha es característico de una prueba con bajo poder discriminatorio, véase Figura 3.9.

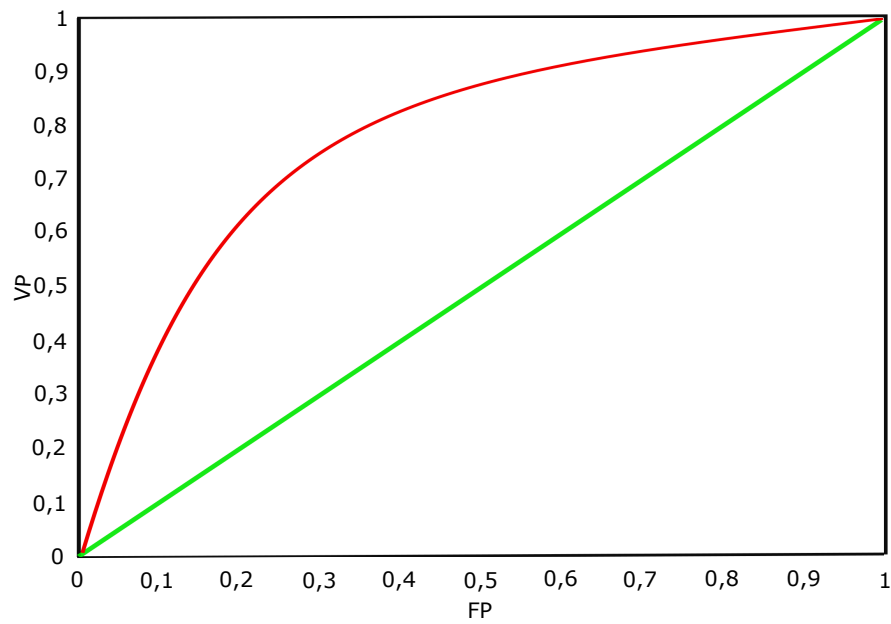


Figura 3.9: Curva ROC, datos simulados

Por último, la Tabla 3.2 resume lo antes definido con relación a la construcción de los principales parámetros de una prueba diagnóstica.

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

Estatus de la condición				
Resultado de la prueba	Presente/positivo	Ausente/negativo	Totales	Mar-ginales
Test Positivo	a-(VP)	b-(FP)	a+b	
Test Negativo	c-(FN)	d-(VN)	c+d	
Totales Marginales	a+c-(Paciente con la enfermedad)	b+d-(Pacientes sin la enfermedad)		
Parámetro	Fórmula	Definición		
Sensibilidad	$a/(a + c)$	La probabilidad de que un individuo enfermo tenga un test +.		
Especificidad	$d/(b + d)$	Es la proporción de sanos que tienen una prueba negativa.		
Valor Predictivo positivo	$a/(a + b)$	Es la probabilidad de que un individuo con resultado positivo en la prueba tenga la enfermedad.		
Valor Predictivo negativo	$d/(c + d)$	Es la probabilidad de que un individuo con resultado negativo en la prueba no tenga la enfermedad.		
LR+	$\frac{\text{sensibilidad}}{1 - \text{especificidad}}$	Describe cuánto más probable es un resultado positivo de la prueba en el grupo de los enfermos en comparación con el grupo que no tiene la enfermedad.		
LR-	$\frac{1 - \text{sensibilidad}}{\text{especificidad}}$	Describe cuántas veces es menos probable un resultado negativo de la prueba en el grupo de los enfermos en comparación con el grupo no afectado.		
Exactitud	$\frac{a + d}{a + b + c + d}$	La probabilidad de que el resultado del test prediga correctamente la presencia o ausencia de la enfermedad.		
DOR	$\frac{ad}{bc} = \frac{LR+}{LR-}$	Describe cuántas veces es más probable de obtener un resultado positivo en la prueba en un enfermo en lugar de una persona no afectada.		

Tabla 3.2: Explicación de la construcción de los principales parámetros en la descripción de una prueba diagnóstica

3.2. Análisis de Sensibilidad y Especificidad en las medidas resumen del análisis diagnóstico

Los estudios sobre la precisión o exactitud diagnóstica son de vital importancia en la evaluación de las tecnologías diagnósticas médicas, es decir ayudan a la validez y reproducibilidad de los procedimientos que son utilizados por el personal de la salud para realizar la mejor aproximación posible a la condición real de los pacientes, Guyatt et al. (71), Knottnerus and Muris (99), Sackett and Haynes (161), y también en la precisión de los estudios, para medir el nivel de concordancia entre los resultados de una prueba, con relación a un patrón de referencia, en un proceso de evaluación. Existen diferentes medidas para la estimación de la precisión diagnóstica Griner et al. (69), Habbema et al. (73), pero la mayoría de los estudios actuales de la precisión diagnóstica se realizan mediante la estimación de la sensibilidad y/o especificidad o en combinación con otras medidas Honest and Khan (89).

Debido a que la gran mayoría de los trabajos de diagnóstica muestran estimaciones de sensibilidad y especificidad, el enfoque que ha seguido el meta-análisis se ha centrado en estas medidas Deeks (32), Honest and Khan (89), Irwig et al. (93), Kardaun et al. (97), Lijmer et al. (112), Moses et al. (132), Rutter and Gatsonis (158), Walter and Jadad (194). La integración de la sensibilidad y especificidad no es sencilla, debido a que están correlacionadas negativamente en muchos estudios; la necesidad de una revisión sistemática y la sintonización de las evidencias publicadas sobre la precisión de las pruebas diagnósticas, se ha incrementado. La información de estas revisiones es un elemento clave en la toma de decisiones en el campo clínico y en las políticas de salud sobre el uso de pruebas diagnósticas; también es esencial para guiar el proceso de desarrollo de la tecnología y la evaluación del diagnóstico en la medicina.

La curva SROC (Summary Receiver Operator Characteristic - SROC), se ha convertido en el método principal en los estudios de meta-análisis, ya que informa tanto la sensibilidad como la especificidad, Hasselblad and Hedges (78), Irwig et al. (94), Kardaun et al. (97), Littenberg and Moses (114), Midgette et al. (129), Moses et al. (132), Walter (193). La curva SROC convierte cada par de valores de sensibilidad y especificidad en una medida de la precisión Glas et al. (65). La desventaja de contar con una sola medida de la precisión diagnóstica, es que no tiene la capacidad para detectar a los enfermos (sensibilidad) y la identificación de los sanos (especificidad). La discriminación entre estas medidas (sensibilidad y especificidad) es muy importante para determinar el empleo adecuado (óptimo) de una prueba en la práctica clínica. El modelo bivalente que propone Reitsma et al. (148), tiene la ventaja de preservar la naturaleza bidimensional de los datos subyacentes y logra una estimación de la sensibilidad y especificidad, capturando una posible correlación (negativa) entre estas dos medidas.

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

Las revisiones diagnósticas empiezan con un conjunto de sensibilidades y especificidades de estudios individuales. Un enfoque intuitivo podría ser separar en dos conjuntos los valores de la sensibilidad y especificidad, usando el método estándar para proporciones, pero sin embargo la relación entre la sensibilidad y la especificidad es inversa (**negativa**) dentro de los estudios, por lo cual si se ignora este detalle podría ser inapropiado, Moses et al. (132), Rutter and Gatsonis (158), Walter and Jadad (194). Una posible causa de esta correlación negativa entre la sensibilidad y la especificidad es que los estudios pueden haber usado diferentes umbrales para definir resultados positivos y negativos de la prueba. En algunos casos, esto puede deberse a una diferencia explícita; por ejemplo, los estudios que utilizaron diferentes puntos de corte para clasificar a una medición (continua) bioquímica como positiva o negativa.

En otras situaciones, se presentan variaciones implícitas en los umbrales entre los estudios, debido a diferencias en los observadores, laboratorios o equipos. A diferencia de otras fuentes de variación, una diferencia de umbral conduce a un patrón particular entre sensibilidad y especificidad. Este patrón es bien conocido a partir de estudios que notan el efecto de distintos puntos de cortes en pruebas bioquímicas con resultados continuos, Begg and McNeil (12), Hilden (85), Sorribas et al. (177). Al reducir el valor del punto de corte, esto ocasionará más pacientes con un resultado positivo, no solo aumentará el número de verdaderos positivos, sino también el número de falsos positivos. Esto significa que la sensibilidad será mayor, a expensas de la especificidad. Esta relación negativa “(**trade-off**)” conduce a una curva cóncava, y esto ocurre cuando se grafica la sensibilidad versus uno menos la especificidad, lo cual genera la curva SROC. En muchos estudios que involucran la curva ROC, la sensibilidad se conoce como verdaderos positivo (tasa de verdaderos positivos) (TVP) y uno menos la especificidad como falsos positivos (tasa de falsos positivos) (TFP).

En los siguientes secciones se mostrará como la SROC, enfoque bivariante, modelo de regresión binomial, HSROC, y meta-análisis bayesiano para la precisión de un test, en ausencia de una prueba de referencia para la estimación de la sensibilidad y especificidad, pueden mostrar una posible correlación negativa. Es importante señalar que el enfoque SROC comienza con la representación de los pares ordenados de la sensibilidad y especificidad de cada estudio; la parte central de este enfoque es encontrar una curva suave mediante la unión de los pares ordenadas antes mencionados, asimismo es importante notar que el paso principal consiste en aplicar una transformación a las tasas de verdaderos positivos (*TVP*) (*sensibilidad*) y a las tasas de falsos positivos (*TFP*)($1 - \textit{especificidad}$). Por otro lado el modelo bivariante, utiliza como punto de partida los conjuntos de puntos de la sensibilidad y especificidad de los meta-análisis, asume que los valores de la sensibilidad y especificidad de cada estudio (después de aplicarle la transformación logit) de un meta-análisis sigue una distribución normal.

En el modelo binomial, el número de pruebas positivas de cada estudio, siguen distribuciones binomiales independientes, en donde la probabilidad de una prueba positiva queda expresada como una transformación logit. El modelo de regresión binomial implica una relación lineal entre $\text{logit}(\text{TVP})$ y $\text{logit}(\text{TFP})$.

El análisis de regresión jerárquica es una extensión del modelo de regresión binomial para explicar más detalladamente la variabilidad dentro y entre estudios de las tasas de VP y FP. El método bayesiano, es una extensión del HSROC, asume que la variación de la sensibilidad y especificidad en los estudios aparece debido al empleo de diferentes puntos de corte para definir una prueba positiva y/o diferencias en la precisión del diagnóstico, cuando no existe una prueba de referencia.

En general, se puede decir que los métodos bayesianos no han sido aprovechados en gran medida en el campo de la salud. El método más común empleado en las pruebas diagnósticas es el **SUMMARY RECEIVER OPERATOR CHARACTERISTIC (SROC)**. Las pruebas diagnósticas son ampliamente usadas en la medicina para determinar el estatus de una enfermedad, es decir, determinar si la enfermedad está presente o ausente. Las pruebas se basan en un resultado continuo subyacente, para determinar si un valor está por encima de un umbral y así considerar indicios de la enfermedad, por tal motivo en la Sección 3.3 se analizan diferentes enfoques estadísticos para la valoración de una prueba diagnóstica.

3.3. Modelos Estadísticos para integración de pruebas

En meta-análisis, el objetivo está en integrar los resultados de los diferentes artículos en los que se evalúa una prueba diagnóstica. La información de partida es una tabla tetracórica en la cual aparecen los VP , los FP , los FN y los VN . La tabla muestra la concordancia entre el resultado de la prueba en forma binaria y la información de la enfermedad. Cuando se trabaja en ensayos clínicos, el tamaño del efecto se mide generalmente en términos de *OddRatio*, incluso cuando se trabaja con pruebas diagnósticas, pero cuando estamos ante enfermedades de muy baja prevalencia (**100 de cada 10 000 por ejemplo**), donde además no existe una prueba oro y la decisión de que el sujeto está enfermo se hace a partir de un punto de corte subjetivo, las técnicas estándar no dan resultados óptimos.

Existen distintas alternativas en la literatura. A continuación se mostrará el enfoque univariante denominado SROC, el enfoque bivariante, el modelo de regresión binomial, el modelo jerárquico HSROC, y el meta-análisis bayesiano, todos ellos para evaluar la precisión de un test en ausencia de una prueba de referencia y para llevar a cabo la estimación de la sensibilidad y especificidad cuando ambas presentan una correlación significativa.

Supóngase que se tienen m estudios que contribuyen al vector z_i que denota las covarianzas en los niveles de estudio ($i = 1, 2, \dots, m$) y una tabla resumen de datos de 2×2 , que muestra la concordancia entre el resultado de la prueba en forma binaria y la información de la enfermedad de manera definitiva (o patrón de referencia) y sea la notación de la Tabla 3.3; donde, los resultados de los verdaderos y falsos positivos de las pruebas se definen en la ecuaciones 3.19 y 3.20 respectivamente.

	Enfermedad	
	Sí	No
Prueba diagnóstica	Sí=1	y_{i11} y_{i01}
	No=0	y_{i10} y_{i00}
Total	n_{i1}	n_{i0}

Tabla 3.3: Tabla resumen de datos

$$\widehat{VP} = \frac{y_{i11}}{n_{i1}} \quad (3.19)$$

$$\widehat{FP} = \frac{y_{i01}}{n_{i0}} \quad (3.20)$$

3.3.1. Enfoque SROC

Los test diagnósticos son ampliamente usados en medicina para determinar el status de una enfermedad; es decir, para determinar si la enfermedad esta presente o ausente. En general se supone que el test diagnóstico esta basado en un contenido latente para el cual, valores por encima de un punto de corte (**specified threshold**) nos llevan a declarar al individuo enfermo. Para investigar el efecto de un punto de corte en la sensibilidad y la especificidad, los investigadores presentan los resultados en forma de curva ROC (**Receiver Operator Characteristic**) donde la probabilidad de los verdaderos positivos (**sensibilidad**) es presentada en ordenadas frente a la probabilidad de falsos positivos (**1- Especificidad**) que se representa en abscisas.

Una forma de resumir el comportamiento de un test diagnóstico desde múltiples estudios es calculando la media de la sensibilidad y la media de la especificidad pero estas medidas resumen no son válidas si hay heterogeneidad entre los estudios y desafortunadamente los test para detectar heterogeneidad no son muy potentes Midgette et al. (129). Además la sensibilidad y la especificidad dentro de cada estudio están inversamente relacionadas y dependen del punto de corte. Todo esto implica que la sensibilidad media y la especificidad media no son medidas adecuadas Irwig et al. (93). Para corregir este posible sesgo, Moses et al. (132) desarrollaron la técnica para calcular la **SROC** (**Summary ROC**) la cual relaciona el punto de corte con la precisión de la prueba, vía regresión lineal.

El enfoque **Summary Receiver Operator Characteristic (SROC) curve approach** comienza con el trazado de los pares ordenados de la sensibilidad y especificidad de cada estudio. El objetivo principal es encontrar una curva suave a partir de estos puntos. El paso principal es transformar la tasa de verdaderos positivos (*TVP*) (**sensibilidad**) y la tasa de falsos positivos (*TFP*) (**1-especificidad**) para que la relación se vuelva lineal y así poder realizar un ajuste para los puntos dados, Moses et al. (132).

Se utilizan las transformaciones 3.21 y 3.22 para *TVP* y *TFP*. Se define *D* como la diferencia de los valores de las transformaciones *logit* de *TVP* y *TFP*, mientras que *S* es la suma de estas transformaciones.

$$D_i = \ln \left(\frac{TV P_i}{1 - TV P_i} \right) - \ln \left(\frac{TF P_i}{1 - TF P_i} \right) = \ln (DOR) \quad (3.21)$$

$$S_i = \ln \left(\frac{TV P_i}{1 - TV P_i} \right) + \ln \left(\frac{TF P_i}{1 - TF P_i} \right) \quad (3.22)$$

$$D_i = \alpha + (\beta S_i) + \varepsilon_i \quad (3.23)$$

D es el logaritmo de la razón del odd ratio diagnóstico “**(diagnostic odds ratio)**” (**DOR**) para cada estudio. Es un indicador o índice de la precisión diagnóstica, y muestra con que frecuencia se produce un resultado positivo en los pacientes con la condición de interés, en comparación con los pacientes sin la enfermedad. S se refiere al umbral de cada prueba. Tiene un valor de cero en un estudio cuando la sensibilidad es igual a la especificidad. S es positiva en los estudios donde la sensibilidad es mayor que la especificidad, y es negativa cuando la especificidad toma valores más grandes que la sensibilidad. Una recta de regresión lineal se ajusta a los puntos de las transformaciones 3.21 y 3.22, que nota como el logaritmo DOR varía en los umbrales implícitos.

El modelo 3.23 es ajustado utilizando mínimos cuadrados ponderados o no ponderados, Moses et al. (132) y ε_i es el error aleatorio para el estudio i . La interpretación del intercepto y de la pendiente del modelo 3.23 se realiza con mayor detalle en la Sección 3.3.1. Cuando el DOR varía con S , el coeficiente β , no tiene una interpretación directa, pero influye en la forma de la curva **ROC** Walter (193).

Algunos aspectos de la variabilidad entre los estudios pueden ser manejados dentro del enfoque $SROC$. Los niveles de covariación de los estudios se pueden incorporar de manera sencilla en la ecuación 3.23 para proporcionar un análisis exploratorio de los efectos de las características del estudio y de la especificidad de la prueba. El modelo puede ser estimado por mínimos cuadrados ordinarios o ponderados o algún método de regresión robusta. Los pesos pueden ser utilizados para cuantificar la diferencia que existen entre estudios, pero no captura de manera simultánea la diferencia en el tamaño de la muestra dentro de los pacientes con la enfermedad (n_{i1}) y los pacientes que no tienen la enfermedad (n_{i0}). Estos dos tamaños muestrales afectan a la precisión de las tasas de verdaderos positivos y falsos positivos. En la práctica, los modelos ponderados y no ponderados pueden producir resultados muy diferentes, y no hay una manera clara de elección entre ellos, Rutter and Gatsonis (158).

El punto Q^* es el punto en la curva SROC donde la sensibilidad y especificidad son iguales. Desafortunadamente, el punto Q^* puede dar valores de sensibilidad y especificidad fuera de los rangos establecidos en los estudios analizados. Estos puntos Q^* , suelen utilizarse para probar la diferencia de precisión (exactitud) global entre las pruebas diagnósticas. La justificación, es que el punto Q^* remueve el efecto de posibles diferencias en los umbrales comparando los **odds ratio diagnóstico** con valores específicos de S .

El coeficiente β en la ecuación 3.23 representa la dependencia de la precisión de la prueba en el umbral. Si $\beta \approx 0$ entonces los estudios pueden referirse como homogéneos, y pueden ser resumidos por un DOR global, notando que $\alpha = \ln(DOR)$. Si $\beta \neq 0$, los estudios son referidos como heterogéneos con respecto al DOR , en este caso el valor de α puede ser estimado como $\ln(DOR)$ cuando $S = 0$. El modelo está generalmente validado para el supuesto de distribución logística. Si los resultados de las pruebas subyacentes en realidad tienen distribución logística, S para un estudio dado puede ser expresado como una función del punto de corte que da lugar al análisis de la sensibilidad y especificidad para ese estudio, Moses et al. (132).

Si las varianzas de las distribuciones de los resultados de las pruebas de los casos y no casos son iguales, entonces la **SROC** es simétrica respecto a la línea diagonal, donde $VP = 1 - FP$. Si las varianzas son desiguales, la SROC resultante será asimétrica. No se hacen suposiciones acerca de la distribución de S . Las covariables se pueden añadir al modelo para evaluar si la exactitud de la prueba varía sistemáticamente con otros factores relacionados con el estudio, Deeks (32), Fahey et al. (53).

Si la estimación de la exactitud de la prueba para cada estudio se pondera por la inversa de su varianza, esto supone que el error de muestreo es la única fuente de variabilidad entre los estudios. Sin embargo, este supuesto puede ser poco realista, dado que los estudios tienden a variar en un sin número de aspectos, incluyendo el espectro de la enfermedad, las condiciones bajo las que se administró la prueba, y otros estudios y características de los pacientes que podrían afectar a la precisión de ensayo, Lijmer et al. (111). A pesar de que los parámetros α y β son tanto fijos porque se supone que son constantes en todos los estudios, aplicando pesos iguales a estudios (es decir, ajustando una regresión no ponderada) se ha demostrado empíricamente que son consistentes con el supuesto de un efecto aleatorio, Moses et al. (132), Rutter and Gatsonis (157). Esto es porque tanto las variaciones intra y la inter de los estudios se tienen en cuenta, lo que da relativamente más peso a los estudios más pequeños, como ocurriría en un modelo de efectos aleatorios.

Irwig et al. (93) recomienda el análisis no ponderado, porque los pesos basados en varianzas de las muestras pueden dar demasiado énfasis a los estudios inexactos, y

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

por lo tanto dar una curva SROC sesgada. Moses et al. (132) también recomienda el enfoque no ponderado. Una vez que la regresión se ha ajustado, se pueden invertir las transformaciones, representadas por las ecuaciones 3.21 y 3.22 y por lo tanto deducir la relación entre VP y FP quedando expresado como la ecuación 3.24, Walter (193).

$$\begin{aligned} \ln\left(\frac{VP}{1-VP}\right) - \ln\left(\frac{FP}{1-FP}\right) &= \alpha + \beta \left(\ln\left(\frac{FP}{1-FP}\right) + \ln\left(\frac{FP}{1-FP}\right) \right) \\ (1-\beta) \ln\left(\frac{VP}{1-VP}\right) &= \alpha + (1-\beta) \ln\left(\frac{FP}{1-FP}\right) \end{aligned}$$

si elevamos a la ecuación por e

$$\frac{VP}{1-VP} = e^{\frac{\alpha}{1-\beta}} \cdot \left(\frac{FP}{1-FP}\right)^{\frac{1+\beta}{1-\beta}}$$

si agrupamos por el término VP

$$\left(1 + e^{\frac{\alpha}{1-\beta}} \cdot \left(\frac{FP}{1-FP}\right)^{\frac{1+\beta}{1-\beta}}\right) \cdot VP = e^{\frac{\alpha}{1-\beta}} \cdot \left(\frac{FP}{1-FP}\right)^{\frac{1+\beta}{1-\beta}}$$

despejando el término VP

∴

$$VP = \frac{e^{\frac{\alpha}{1-\beta}} \cdot \left(\frac{FP}{1-FP}\right)^{\frac{1+\beta}{1-\beta}}}{1 + e^{\frac{\alpha}{1-\beta}} \cdot \left(\frac{FP}{1-FP}\right)^{\frac{1+\beta}{1-\beta}}}$$

(3.24)

La expresión 3.24 expresa TVP en términos de TFP y por lo tanto define toda la curva SROC. Si bien puede haber interés en identificar puntos particulares en la curva, también es útil tener una medida global resumen del comportamiento de la curva. Una medida adecuada es el área bajo la curva (**AUC**), que se puede calcular con la ecuación 3.25, Walter (193).

$$\int_0^1 \frac{e^{\frac{\alpha}{1-\beta} \cdot \left(\frac{x}{1-x}\right)^{\frac{1+\beta}{1-\beta}}}}{1 - e^{\frac{\alpha}{1-\beta} \cdot \left(\frac{x}{1-x}\right)^{\frac{1+\beta}{1-\beta}}}} dx \quad (3.25)$$

En general, la **AUC** debe ser estimada empleando algún método numérico para aproximar. La adopción de un índice resumen es útil para la información sucinta de un determinado conjunto de datos, especialmente cuando los datos limitados impiden la identificación fiable de determinados puntos de la curva. La medida **AUC** es ampliamente utilizado en el análisis **ROC**, en el que se puede interpretar como la probabilidad de que los valores de prueba para un par aleatorio de individuos enfermos y no enfermos se clasifican correctamente. También representa el promedio (no ponderado) de TVP sobre todos los valores posibles de *TFP*. Las AUC son también un resumen candidato natural para un análisis **SROC**. El **AUC parcial** también se ha propuesto como una medida resumen, siendo este el área bajo una parte restringida de la curva correspondiente a los valores *TFP* de interés clínico, o en los que se encuentran los datos del estudio.

Comportamiento empírico de SROC

La Figura 3.10 muestra un conjunto de tres curvas simétricas ROC con $\beta = 0$, que ocurre cuando los estudios son homogéneos y por lo tanto no presentan ninguna relación entre *DOR* y el umbral *S*. Las curvas de la Figura 3.10 se obtuvieron utilizando evaluación numérica de la ecuación 3.24 para $\alpha = 1, 5, \alpha = 2$, y $\alpha = 3$ que corresponden a *DOR* = 4.5, 7.4, y 20.1, respectivamente. Al incrementarse el valor de α la curva *SROC* se acerca a su ideal, cerca de la esquina superior izquierda. Si $\alpha \rightarrow \infty$, $AUC \rightarrow 1$, lo que indicaría una prueba perfecta con 100% de sensibilidad y especificidad y no hay errores en la distinción de casos y no casos. En contraste, si $\alpha \simeq 0$ o *DOR* = 1 la curva se encuentra cerca de la diagonal $VP = FP$ en el espacio *SROC*, luego $AUC = \frac{1}{2}$, por lo que se infiere que la prueba no realiza ninguna mejora que el azar, Walter (193).

Para complementar, si $\alpha < 0$ es decir *DOR* < 1, la prueba discrimina casos y no casos en direcciones equivocadas, dando resultados peores que el azar. Si *DOR* $\rightarrow 0$, es decir $AUC \rightarrow 0$ la curva se encuentra cerca de la esquina inferior derecha del espacio *SROC*. Tales situaciones son poco probables que ocurran en la práctica, Walter (193).

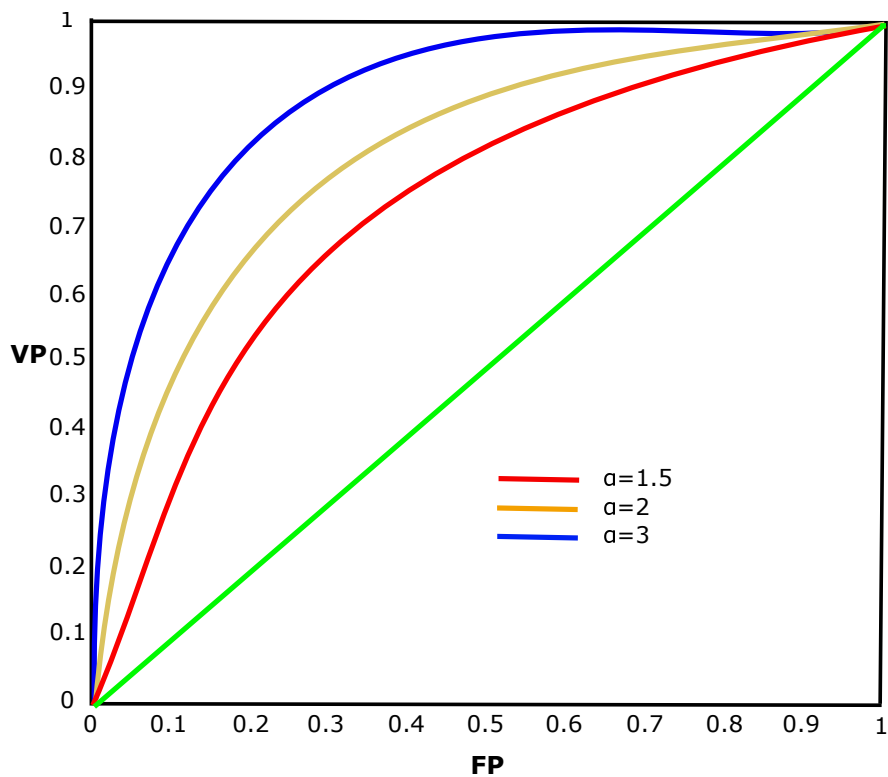


Figura 3.10: Curva SROC con varios valores de α ($\beta = 0$), datos simulados

Se considera el caso de estudios heterogéneos, modelo 3.23 , donde la precisión diagnóstica depende del umbral $\beta \neq 0$. La Figura 3.11 muestra tres curvas **SROC** derivadas de la ecuación 3.24 , todas con $\alpha = 2$, pero diferentes valores de β .

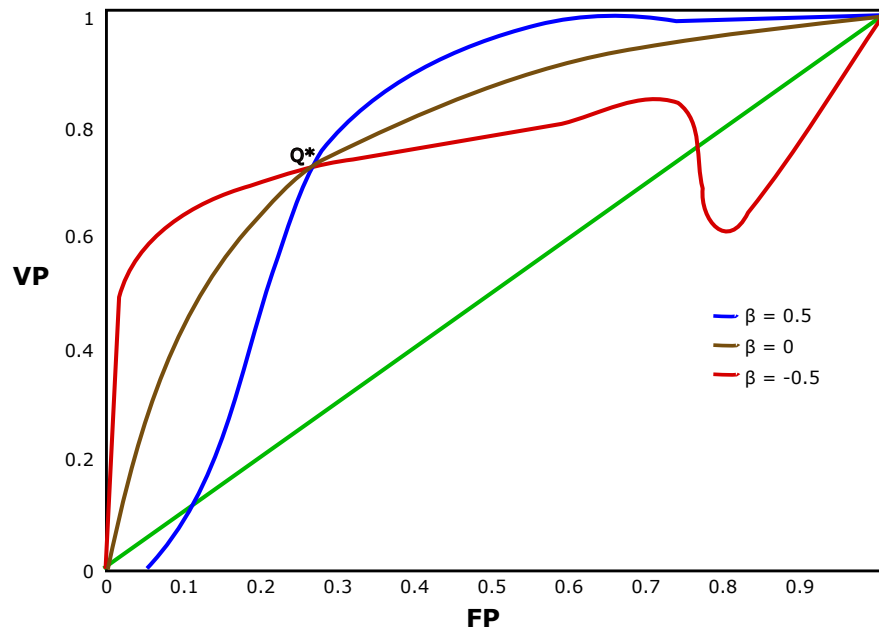


Figura 3.11: Curva SROC con varios valores de β ($\alpha = 2$), datos simulados

Los valores distintos de cero para β producen curvas asimétricas, si $\beta > 0$ tiene una elevación menos pronunciada que la curva simétrica con $\beta = 0$ hasta que la intercepta y a partir del punto de intersección se eleva más abruptamente para alcanzar valores relativamente altos de VP para valores altos de FP. Las curvas ROC con $\beta < 0$ exhiben el comportamiento opuesto, por ejemplo, la curva con $\beta = -0.5$ en la Figura 3.11. Curiosamente, como se sugiere en la Figura 3.11 y como lo muestra Moses et al. (132), la familia de curvas definidas por un valor fijo cruzan por un punto común, que se encuentra en la antidiagonal, donde $TVP = 1 - TFP$, o la sensibilidad es igual a la especificidad, teniendo como coordenadas las ecuaciones, 3.26 y 3.27, Walter (193) que son obtenidas mediante las ecuaciones 3.21, 3.22 y 3.23. Nótese, que para obtener las ecuaciones siguientes se asume que la ecuación 3.22 es igual a cero.

$$\ln\left(\frac{VP}{1-VP}\right) + \ln\left(\frac{FP}{1-FP}\right) = 0$$

$$\ln\left(\frac{VP}{1-VP}\right) - \ln\left(\frac{FP}{1-FP}\right) = \alpha \text{ si sumamos las dos ecuaciones, se tiene que:}$$

$$2 \ln\left(\frac{VP}{1-VP}\right) = \alpha \text{ aplicando logaritmo a la ecuación, se tiene que:}$$

$$VP + VP\left(e^{\alpha/2}\right) = e^{\alpha/2} \text{ despejando VP, se tiene que:}$$

∴

$$VP = \frac{e^{\alpha/2}}{1 + e^{\alpha/2}} \tag{3.26}$$

$$FP = \frac{1}{1 + e^{\alpha/2}} \tag{3.27}$$

Moses et al. (132) denotó el valor de VP en este punto por Q^* . En el análisis ROC, Q^* se ha sugerido como una medida resumen más. Corresponde con el punto más cercano a la esquina superior izquierda ideal del espacio SROC en curvas simétricas. El hecho de que todas las curvas SROC con un valor dado pasen por el mismo punto Q^* significa que Q^* no transmite ninguna información estadística adicional más allá de odds ratio. Sin embargo, el uso de Q^* puede estar motivado por su sencilla interpretabilidad, teniendo en cuenta que Q^* es el punto de la curva donde $VP = 1 - FP$; por lo tanto, representa el umbral de diagnóstico en el que la probabilidad de un diagnóstico correcto es constante para todos los sujetos.

Como consecuencia de asumir el modelo 3.23, $\beta \neq 0$, la curva SROC tiene una región en la que $VP < FP$, que se encuentra por debajo de la diagonal principal. Esta región puede verse, por ejemplo, cerca de la esquina inferior izquierda de la curva de $\beta = 0.5$ en la Figura 3.11. En esta región, la prueba teóricamente estaría clasificando los casos y no casos peor que al azar. Sin embargo, en la práctica, esta región es muy pequeña. Incluso si existe presencia relativamente fuerte de heterogeneidad con $\beta > 0$, la parte “impropia” de la curva SROC (donde $TVP < TFP$) incluye valores muy bajos de TVP , y estos valores corresponden a umbrales que es poco probable que sean aceptables en la práctica clínica.

Si la heterogeneidad ocurre con $\beta < 0$, la parte impropia de la curva corresponde a valores altos de FP , lo que de nuevo no tiene relevancia clínica, vea Walter (193) para más detalle. Cuando $\beta \rightarrow 1$ la curva SROC ajustada se vuelve progresivamente más

empinada, y en el caso límite en $\beta = 1$ degenera a una línea vertical. Esto podría ocurrir, por ejemplo, si no hay variación en FP entre los estudios. Sin embargo, pasaría por el punto común Q^* definida por las ecuaciones 3.26 y 3.27.

Si $\beta > 1$, se invierte la curva y muestra una relación negativa de VP y FP . Esto es inverosímil, en la práctica, excepto tal vez por casualidad en muestras pequeñas. Un comportamiento similar se observa si β se aproxima a -1 . Si $\beta = -1$, la curva es horizontal, indicando que no hay relación de VP y FP , esto podría ocurrir si no hubiera variación en VP entre los estudios.

Si $\beta < -1$, la curva se invierte (a una forma diferente) y sugiere una relación negativa entre VP y FP . A pesar de ello, todas las curvas para $|\beta| > 1$ poseen un valor común Q^* , dado por las ecuaciones 3.26 y 3.27. En la práctica, valores $|\beta| > 1$ son poco probables, Walter (193). La experiencia empírica sugiere que los conjuntos de datos de meta-análisis práctico a menudo tienen β cercano a 0, y son raras veces mayor que 0.5 en valor absoluto, Walter (193).

Área Bajo la Curva y Q^*

El AUC es un índice popular del rendimiento global de la prueba, Beck and Shultz (10), DeLong et al. (34). Los rangos de las AUC van, desde 1 para una prueba perfecta que diagnostica correctamente siempre; a 0, para una prueba que no diagnostica correctamente Hanley (74), Hanley and McNeil (75), Hilden (85), Metz et al. (128).

En los estudios individuales, AUC se puede interpretar como la probabilidad de que la prueba categoriza correctamente un caso de los no casos, n elegido al azar con respecto a sus valores de prueba, Bamber (9). Se propone el área bajo la curva “AUC” para que cumpla con la misma función que en los meta-análisis, suponiendo que la curva **SROC** transmite con precisión el rendimiento de pruebas a nivel de los sujetos que se analizan en la prueba. Aunque se requiere la integración numérica en general para obtener AUC, algunos casos especiales son de interés porque con ellos se obtienen expresiones analíticas exactas y resultados comparativos.

Se sabe que AUC se incrementa con el valor de α , manteniendo fijo β . Mediante el examen de la ecuación AUC, se puede demostrar que para un valor dado de α , AUC se maximiza cuando $\beta = 0$, lo que implica que el AUC es óptima en los estudios “grandes” homogéneos, Walter (193). Por otra parte, el AUC es simétrica en β , por lo que los valores negativos de β producen el mismo valor de AUC en valor absoluto. Si $\alpha = \beta = 0$ de la ecuación 3.25, se tiene que $\int_0^1 x dx = \frac{1}{2}$.

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

Es evidente que si $AUC = \frac{1}{2}$ cuando $\alpha = 0$, inclusive si $\beta \neq 0$, así que cuando $AUC = \frac{1}{2}$ indica el rendimiento general al azar para cualquier conjunto de estudios. En el caso homogéneo $\beta = 0$, la ecuación 3.25 queda expresada como la ecuación 3.28, Walter (193), obteniendo una solución exacta, mediante la ecuación 3.29, Walter (193).

$$\int_0^1 \frac{e^{\alpha \cdot \left(\frac{x}{1-x}\right)}}{1 + e^{\alpha \cdot \left(\frac{x}{1-x}\right)}} dx \quad (3.28)$$

$$\int_0^1 \frac{e^\alpha \cdot \left(\frac{x}{1-x}\right)}{1 + e^\alpha \cdot \left(\frac{x}{1-x}\right)} dx = e^\alpha \int_0^1 \frac{\frac{x}{\cancel{1-x}}}{\frac{1-x + e^\alpha x}{\cancel{1-x}}} dx = e^\alpha \int_0^1 \frac{x}{1 + x(e^\alpha - 1)} dx$$

$u = 1 + x(e^\alpha - 1)$; realizando cambio de variable

$$dx = \frac{du}{e^\alpha - 1}$$

si $x = 0$ se tiene que $u = 1$

si $x = 1$ se tiene que

$$u = 1 + e^\alpha - 1 = e^\alpha; \text{ luego se tiene que:} \tag{3.29}$$

$$\begin{aligned} &= \frac{e^\alpha}{(e^\alpha - 1)^2} \int_1^{e^\alpha} \frac{u-1}{u} du \\ &= \frac{e^\alpha}{(e^\alpha - 1)^2} \left[\int_1^{e^\alpha} du - \int_1^{e^\alpha} \frac{du}{u} \right] \\ &= \frac{e^\alpha}{(e^\alpha - 1)^2} \left[u \Big|_1^{e^\alpha} - \ln(u) \Big|_1^{e^\alpha} \right] \\ &= \frac{e^\alpha}{(e^\alpha - 1)^2} [e^\alpha - 1 - \ln(e^\alpha)] \end{aligned}$$

$$\therefore AUC_{hom} = \frac{DOR}{(DOR - 1)^2} \cdot [(DOR - 1) - \ln(DOR)]$$

Donde AUC_{hom} indica el AUC para estudios homogéneos, y $DOR = e^\alpha$. Si $\alpha = 0$ u $DOR = 1$,

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

entonces $AUC_{hom} = \frac{1}{2}$. La expresión 3.29 puede ser usada para evaluar AUC para estudios homogéneos, sin necesidad de utilizar integración numérica. La Tabla 3.4 muestra el valor de AUC, Q^* , y su diferencia para un rango de valores de DOR en el caso homogéneo. Note que AUC disminuye con el aumento de β , y que la curva cuando $\beta \rightarrow 1$ pasa a través del punto común Q^* , a partir de las ecuaciones 3.26 y 3.27, se deduce un límite inferior para el AUC en curvas con un determinado valor de α , que se expresa mediante la ecuación 3.30, Walter (193), que es equivalente a la ecuación 3.26.

$$Q^* = \frac{e^{\alpha/2}}{1 + e^{\alpha/2}} = \frac{\sqrt{OR}}{1 + \sqrt{OR}} \quad (3.30)$$

Odds ratio	AUC_{hom}	Q^*	$AUC_{hom} - Q^*$
0.5	0.386	0.414	-0.028
1	0.500	0.500	0.000
1.5	0.567	0.511	0.017
2	0.614	0.586	0.028
3	0.676	0.634	0.042
4	0.717	0.667	0.051
5	0.747	0.691	0.056
10	0.827	0.760	0.067
20	0.887	0.817	0.069
30	0.913	0.846	0.068
40	0.929	0.863	0.065
50	0.939	0.876	0.063

Tabla 3.4: AUC_{hom} , Q^* , y sus diferencias para varios valores de odds ratio diagnóstico: caso homogéneo

El valor de Q^* de la ecuación 3.30 y el máximo valor de AUC_{hom} de la ecuación 3.29 proveen límites inferiores y superiores fácilmente computables para AUC para cualquier valor dado de $\alpha > 0$. Las evaluaciones numéricas de las expresiones 3.29 y 3.30 demuestran que, en el caso

homogéneo, la diferencia entre AUC y Q^* aumentan para valores moderados de DOR , pero nunca es más de alrededor del 7%, Walter (193).

La diferencia máxima se produce con un valor de $\alpha = 2.85$ ($DOR = 17.3$), que es el valor solución de una ecuación trascendental. Para valores mayores de α , la diferencia disminuye muy lentamente, con un valor límite $AUC_{hom} - Q^* = 0$ cuando $\alpha \rightarrow \infty$. La integración numérica de la ecuación 3.25 para el caso heterogéneo permite evaluar el impacto de la heterogeneidad en el valor de AUC. Para un valor fijo de DOR , AUC disminuye lentamente a medida que se incrementa β desde cero (estudios homogéneos) a valores más grandes (el aumento de heterogeneidad). Sin embargo, la dependencia de las AUC en β es débil, y el efecto dominante en el AUC es el valor de DOR (o α). Para $|\beta| < 0.4$, el cambio porcentual en comparación con el AUC caso homogéneo es de menos de 2%. En consecuencia, AUC_{hom} proporciona una buena aproximación a AUC incluso en estudios heterogéneos Walter (193).

Error Estándar de AUC y Q^*

Considérese la variación muestral de \widehat{AUC} , de la ecuación 3.25. Nótese que es una función de regresión de los parámetros α y β y que por lo tanto la variabilidad en \widehat{AUC} es una función de regresión de la variación de $\hat{\alpha}$ y $\hat{\beta}$, usando el método delta -se emplea para obtener la distribución asintótica de una función de variables aleatorias usando la primera derivada para obtener una aproximación por una serie de Taylor de primer orden- para obtener una aproximación a la varianza de \widehat{AUC} , se tiene la ecuación 3.31, Walter (193) y de la ecuación 3.25 se obtienen las expresiones 3.32 y 3.33, Walter (193).

$$var(\widehat{AUC}) = \left(\frac{\partial AUC}{\partial \alpha}\right)^2 var(\hat{\alpha}) + \left(\frac{\partial AUC}{\partial \beta}\right)^2 var(\hat{\beta}) + 2\left(\frac{\partial AUC}{\partial \alpha}\right)\left(\frac{\partial AUC}{\partial \beta}\right)cov(\hat{\alpha}, \hat{\beta}) \quad (3.31)$$

$$\frac{\partial AUC}{\partial \alpha} = \left(\frac{1}{1-\beta}\right) e^{\frac{\alpha}{1-\beta}} \int_0^1 \frac{\left(\frac{x}{1-x}\right)^p}{\left[1 + \left(\frac{x}{1-x}\right)^p e^{\frac{\alpha}{1-\beta}}\right]^2} dx \quad (3.32)$$

$$\frac{\partial AUC}{\partial \beta} = \left(\frac{1}{1-\beta}\right)^2 e^{\frac{\alpha}{1-\beta}} \int_0^1 \frac{\left(\frac{x}{1-x}\right)^p \left[\alpha + 2 \ln\left(\frac{x}{1-x}\right)\right]}{\left[1 + \left(\frac{x}{1-x}\right)^p e^{\frac{\alpha}{1-\beta}}\right]^2} dx \quad (3.33)$$

Donde $p = \frac{1+\beta}{1-\beta}$. En general la evaluación de la $var(\widehat{AUC})$, requiere integración numérica sin embargo, para el caso especial de los estudios homogéneos donde $\beta = 0$, una expresión es posible utilizando el método delta, que se podrá obtener mediante la ecuación 3.34, Walter (193).

De la ecuación 3.34 se tiene que $SE(\widehat{AUC}_{hom})$ es simétrico con $\ln(DOR)$ para valores $DOR > 1$. Cuando $DOR = 1$, la expresión 3.34 se degenera, pero se puede utilizar la regla de L'Hôpital y se puede demostrar que en la vecindad de $DOR = 1$, se tiene la expresión $SE(\widehat{AUC}) \approx \frac{SE(\hat{\alpha})}{6}$, Walter (193). El método delta también provee una estimación para el error estándar de Q^* , donde se tiene la expresión $SE(\hat{Q}^*) = \frac{\sqrt{DOR}}{2(\sqrt{DOR} + 1)} SE(\hat{\alpha})$, Walter (193).

$$SE(\widehat{AUC}_{hom}) = \frac{DOR}{(DOR - 1)^3} [(DOR + 1) \ln(DOR) - 2(DOR - 1)] SE(\hat{\alpha}) \quad (3.34)$$

3.3.2. Modelo Bivariante

Hay que señalar que el modelo SROC no cuantifica el error en S , que puede ser una fuente de sesgo en las estimaciones de parámetros y los resúmenes de la curva, Baker (8).

Un enfoque alternativo para la construcción de una curva de SROC fue propuesto por Kardaun et al. (97), que supone que $\left(\ln\left(\frac{TV P_i}{1 - TV P_i}\right); \ln\left(\frac{TF P_i}{1 - TF P_i}\right)\right)$ sigue una distribución normal bivariante y postuló una relación lineal entre los dos componentes de la media bivariante.

El principio de máxima verosimilitud es empleado en este enfoque para la estimación de la pendiente y el intercepto de este modelo. La desventaja del uso del **DOR** como parámetro de resultado para las estimaciones es que al combinar sensibilidad y especificidad en un único indicador se pierde el valor relativo de cada uno de éstos.

El modelo bivariante utiliza un punto de partida diferente para los pares ordenados de sensibilidad y especificidad de los meta-análisis. En vez de transformar estas dos medidas en un indicador diagnóstico de la precisión, como en el enfoque SROC, el modelo bivariante preserva las dos dimensiones de los datos a través del análisis y se fundamenta en el siguiente razonamiento Kardaun et al. (97), Van Houwelingen et al. (187) y Van Houwelingen et al. (188).

Se asume que los valores de la sensibilidad de los estudios individuales (después de la transformación logit) de un meta-análisis siguen una distribución normal alrededor del valor de la media y con cierta cantidad de variabilidad alrededor de la misma media. Este es un enfoque de efectos aleatorios, similar al que se usa en los ensayos terapéuticos, para incorporar la variabilidad no explicada en el análisis.

Esta variación en la sensibilidad subyacente entre los estudios puede estar relacionada con las diferencias que subsisten en la población de estudio, las diferencias implícitas en el umbral, o variaciones imperceptibles en el protocolo del ensayo clínico. Las mismas consideraciones se aplican a las especificidades de estos estudios. La presencia potencial de una correlación (negativa) entre la sensibilidad y la especificidad dentro de los estudios se aborda mediante la

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

incorporación explícita de esta correlación en el análisis.

La combinación de la distribución normal y las transformaciones logit de la sensibilidad y especificidad, reconocen la posible correlación entre ellas, lo que lleva a una distribución normal bivalente Kotz et al. (100), Van Houwelingen et al. (187). Además de la variabilidad entre los estudios en las sensibilidades y especificidades subyacentes, también existe variación debida al muestreo. Si los estudios difieren en tamaño, la variación es debida probablemente al tamaño de la muestra (muestras pequeñas).

El modelo bivalente incorpora la precisión, donde la sensibilidad y especificidad se miden en cada estudio mediante el enfoque de Van Houwelingen et al. (187). Es decir, en estudios con una estimación más precisa de la sensibilidad, tendrán peso más alto en el análisis de sensibilidad. Una descripción más técnica del modelo bivalente es la siguiente.

Sean los estudios individuales ($I = 1, 2, 3, \dots, k$) que muestran sensibilidad $p(A_i)$, determinada en N_A individuos con la condición de interés y especificidad $p(B_i)$ medida en N_B individuos sin esta condición de interés. Sea $\theta_{A,i}$ la transformación logit de la sensibilidad en el estudio i , y $\theta_{B,i}$ la transformación logit de la especificidad. Supóngase, que los verdaderos valores logit de sensibilidad de estos estudios se distribuyen normalmente alrededor de algún valor medio θ_A , con una variabilidad entre estudios de σ_A^2 . Se asume el supuesto de efectos aleatorios para la especificidad de los estudios; donde θ_B , denota el valor medio *logit* de la especificidad y σ_B^2 la varianza entre estudios de los logit de las especificidades.

Se incorpora de manera explícita la posible correlación entre la sensibilidad (*logit*) y especificidad dentro de los estudios. La combinación de las dos distribuciones normales que se pueden correlacionar lleva al modelo normal bivalente, 3.35. Nótese, la representación de Σ en la ecuación 3.36. Es importante notar que Σ y σ_{AB} representan las covarianzas de la transformación *logit* de la sensibilidad y especificidad. El modelo bivalente se extiende mediante la incorporación de la precisión de la sensibilidad y especificidad que se ha estimado en cada estudio. Cuando N_A y N_B son **grandes** y $0 < p_{A,i}; p_{B,i} < 1$, se tiene que la estimación correspondiente para la varianza de los logit de la sensibilidad y especificidad en cada estudio viene dada por

las ecuaciones 3.37 y 3.38 respectivamente. Si se trata a las varianzas observadas de los *logit* de la sensibilidad y especificidad como cantidades fijas, un enfoque estándar en meta-análisis se puede escribir como la ecuación 3.39, donde C_i es una matriz diagonal de S_i^2 , ecuación 3.40. Este modelo puede ser ajustado por el método de máxima verosimilitud, utilizando **SAS**, **proc mixed**, **R mada**, **stata** entre otros. La salida estándar del modelo bivalente incluye, la media *logit* de la sensibilidad θ_A y especificidad θ_B con sus errores estándar y sus intervalos de confianza al 95%; y las estimaciones entre estudios de la variabilidad de los *logit* de la sensibilidad σ_A^2 y especificidad σ_B^2 y la covarianza entre ellas σ_{AB} .

$$\begin{pmatrix} \theta_{A,i} \\ \theta_{B,i} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_A \\ \theta_B \end{pmatrix}, \Sigma \right) \quad (3.35)$$

$$\Sigma = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix} \quad (3.36)$$

$$S_{A,i}^2 = \frac{1}{n_{A,i} p_{A,i} (1 - p_{A,i})} \quad (3.37)$$

$$S_{B,i}^2 = \frac{1}{n_{B,i} p_{B,i} (1 - p_{B,i})} \quad (3.38)$$

$$\begin{pmatrix} \hat{\theta}_{A,i} \\ \hat{\theta}_{B,i} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_A \\ \theta_B \end{pmatrix}, C_i \right) \quad (3.39)$$

$$C_i = \begin{pmatrix} S_{A,i}^2 & 0 \\ 0 & S_{B,i}^2 \end{pmatrix} \quad (3.40)$$

$$LR+ = \frac{\frac{\exp(\theta_A)}{1 + \exp(\theta_A)}}{1 - \frac{\exp(\theta_B)}{1 + \exp(\theta_B)}} \quad (3.41)$$

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

$$LR- = \frac{1 - \frac{\exp(\theta_A)}{1 + \exp(\theta_A)}}{\frac{\exp(\theta_B)}{1 + \exp(\theta_B)}} \quad (3.42)$$

Basado en los parámetros antes descritos se pueden calcular otras medidas de interés, como, los ratios de máxima verosimilitud para test positivos y negativos, ecuaciones 3.41 y 3.42, la correlación entre el *logit* de la sensibilidad y especificidad, estimada como $\frac{\sigma_{AB}}{\sqrt{\sigma_A^2}\sqrt{\sigma_B^2}}$. Los parámetros del modelo bivalente se estiman en un modelo único que incorpora una posible correlación entre la sensibilidad y especificidad. Hasta cierto punto, los modelos de diagnóstico bivalente pueden ser expresados como un análisis longitudinal con dos mediciones (sensibilidad y especificidad) dentro de cada persona (correspondiente a estudios individuales), Reitsma et al. (148). El modelo bivalente puede ser visto como una mejora y ampliación del enfoque SROC, ya que puede producir los siguientes resultados, Reitsma et al. (148):

- Estima la cantidad de variación entre los estudios mediante la sensibilidad y especificidad; además provee el grado de correlación entre la sensibilidad y la especificidad. Esto proporciona información importante acerca de los resultados de heterogeneidad entre los estudios donde se pueda notar una diferencia de un umbral implícito.
- Produce estimaciones de sensibilidad y especificidad con sus respectivos intervalos de confianza al 95%. Estos intervalos toman en cuenta la heterogeneidad entre los estudios más allá del azar (modelo de efectos aleatorios).
- Los parámetros de la distribución bivalente permiten calcular ya sea una elipse de confianza alrededor de las medias de los *logit* de los valores de sensibilidad y especificidad o una elipse de predicción para los valores sensibilidad y especificidad, teniendo en cuenta la posible correlación (negativa) entre ambas.
- Se pueden calcular otras medidas derivadas de la sensibilidad y especificidad, como los odds ratio de diagnóstico y las razones de máxima verosimilitud.
- Las covariaciones pueden ser agregadas al modelo bivalente, lo cual conduce a separar los efectos sobre la sensibilidad y especificidad, pero los efectos sobre los odds ratio de diagnóstico estarán presentes. Esto quiere decir que se puede probar de manera explícita si la sensibilidad o especificidad o ambas son diferentes entre dos tecnologías de diagnóstico.

3.3.3. Modelo de Regresión Binomial

Un modelo de regresión para el estudio en el campo del meta-análisis para relacionar el par ordenado de $(VP_i; FP_i)$ se discutió por primera vez en la referencia, Rutter and Gatsonis (158) y fue motivado por la formulación de un análisis de regresión ordinal para la curva ROC, Baker (8), McCullagh (124), Tosteson and Begg (186).

Si W denota el grado de sospecha de la presencia de una anomalía, denotada en una escala categórica ordinal con J categorías, el modelo paramétrico ROC es equivalente al modelo de regresión ordinal $g = (\theta_j + \alpha x) \exp(-\beta x)$. Donde x es una covariable que denota el verdadero estado de la enfermedad (binario). La base conceptual del modelo es la suposición de las respuestas observadas W que representan una categorización de una variable latente, con una función enlace $g(\cdot)$ que relaciona la parte aleatoria y sistemática del modelo. El **probit link** (que transforma probabilidades en valores estándar de la función de distribución normal) es una variable **latente gaussiana** que se utiliza comúnmente para estudios individuales de la curva ROC, Harbord et al. (76). Se usa un logit link bajo el supuesto de un modelo logit, para estimar los parámetros de la regresión mediante los logaritmos de los odds ratios. En el ajuste binomial con $J = 2$, el uso de la relación logit no requiere la asunción de una variable latente subyacente-log normal. Un modelo de regresión ordinal con una logistic link y $J = 2$, se puede utilizar en los estudios de meta-análisis que reportan el par ordenado $(VP_i; FP_i)$, Rutter and Gatsonis (158).

Bajo el modelo binomial, el número de pruebas positivas de cada estudio, $y_{ij1}, i = 1, 2, \dots, m, j = 0, 1$, siguen distribuciones binomiales independientes, es decir $y_{ij1} \sim Binomial(n_{ij}, \pi_{ij})$, en el que la probabilidad de una prueba positiva queda expresada o modelizada mediante la ecuación 3.43. Donde θ_i representan los puntos de corte, y los valores de VP y FP se incrementan al incrementar los puntos de corte. Por lo tanto, los puntos de corte explican la relación negativa (trade-off) entre los VP y FP. Llamamos a α , parámetro de precisión, ya que mide la diferencia entre VP y FP .

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

$$\text{logit}(\pi_{ij}) = (\theta_i + \alpha_i X_{ij}) \exp(-\beta X_{ij}) \quad (3.43)$$

Cuando $\beta = 1$, α modeliza al logaritmo de los odds ratio como constante para una prueba positiva de un paciente que tiene la enfermedad versus un paciente que no tiene la enfermedad. Llamamos β el parámetro de escala, ya que permite notar la diferencia que existe en las varianzas de los pacientes con y sin la enfermedad de una población de estudio. Si $\beta \neq 1$, los valores de VP y FP aumentan a un ritmo diferente al variar el valor de θ , lo que produce curvas ROC asimétricas. El modelo de regresión binomial se estima por máxima verosimilitud y puede cuantificar una tasa de error para los valores de VP y FP .

El modelo de regresión binomial, ecuación 3.43, implica una relación lineal entre $\text{logit}(VP)$ y $\text{logit}(FP)$. El modelo anterior, asume que en todos los estudios, las pruebas (test) tienen la misma precisión y parámetros de escala, con diferencias entre los estudios debido a los diferentes umbrales de positividad θ_i . Estos modelos de regresión binomial se pueden formular para permitir efectos de covarianza de los niveles de estudio, Rutter and Gatsonis (158). Sin embargo, debido a que cada estudio contribuye sólo con dos grados de libertad, existen limitaciones prácticas sobre el número de covariables que pueden ser incluidas en estos modelos.

3.3.4. Regresión Jerárquica para Meta-análisis de las evaluaciones en la precisión de las pruebas diagnósticas-HSROC

El objeto del modelo jerárquico es obtener estimaciones resúmenes significativas de la sensibilidad y especificidad y propone un mejor manejo de la variabilidad dentro y entre estudio, aplicando un enfoque bayesiano para la estimación de los parámetros.

Este enfoque bayesiano tiene un precio, se basa en las cadenas de Markov¹, simulaciones de Monte Carlo², lo cual requiere programación, simulaciones y una evaluación de la convergencia del modelo y una adecuación y síntesis de los resultados de la simulación. El modelo se divide en tres niveles, Rutter and Gatsonis (158).

Nivel I (Variación dentro del estudio)

El número de falsos y verdaderos positivos de la prueba del estudio i -ésimo, y_{i01} e y_{i11} , se suponen que son independientes y siguen una distribución binomial, donde la probabilidad de que la prueba sea positiva viene dada por la ecuación 3.43. Donde X_{ij} , denota el verdadero estado de la enfermedad para los casos que se encuentran en la ij -ésima celda. Bajo este enfoque jerárquico (HSROC), se permiten criterios de positividad, θ_i y un parámetro de exactitud α_i , que pueden variar en cada estudio, ecuación 3.43. Debido a que la estimación del parámetro de escala β requiere información de más de un estudio, β se supone que es constante en todos los estudios. Si se permite que β varíe en todos los estudios, los parámetros en cada estudio serían identificables o estimados sólo a través de sus distribuciones a priori.

¹Una cadena de Markov es una sucesión de ensayos similares u observaciones en la cual cada ensayo tiene el mismo número finito de resultados posibles y en donde la probabilidad de cada resultado para un ensayo dado depende sólo del resultado del ensayo inmediatamente precedente y no de cualquier resultado previo.

²El método Montecarlo es un método numérico que permite resolver problemas físicos y matemáticos mediante la simulación de variables aleatorias. El método MonteCarlo fue bautizado así por su clara analogía con los juegos de ruleta de los casinos, el más célebre de los cuales es el de Montecarlo, casino cuya construcción fue propuesta en 1856 por el príncipe Carlos III de Mónaco, siendo inaugurado en 1861.

Nivel II (Variación entre estudios)

Los parámetros de los niveles de estudio en 3.43, α_i y θ_i , se asumen que están distribuidos normalmente, con una media determinada por una función lineal de las covariaciones de los niveles de estudio. En el caso de una sola covarianza Z que afecta al punto de corte y al parámetro de precisión, el modelo puede ser escrito como se nota en la ecuación 3.44, Rutter and Gatsonis (158). Los coeficientes de γ y λ del modelo sistemático se diferencian en el criterio de positividad y precisión en todos los estudios, debido a la covarianza Z . Sin embargo, se puede hacer una formulación más general del modelo, es decir, se incluye más de una covariable y diferentes entre sí, que se utilizan para los puntos de corte y precisión en las ecuaciones de la regresión.

$$\left. \begin{aligned} \theta_i | \Theta, \gamma, Z_i, \sigma_\theta^2 &\sim N(\Theta + \gamma Z_i, \sigma_\theta^2) \\ \alpha_i | \Lambda, \lambda, Z_i, \sigma_\alpha^2 &\sim N(\Lambda + \lambda Z_i, \sigma_\alpha^2) \end{aligned} \right\} \text{Condicionales independientes} \quad (3.44)$$

Nivel III

La especificación del modelo jerárquico se completa con la elección de las distribuciones a priori de los parámetros restantes, ecuación 3.45, Rutter and Gatsonis (158). Los parámetros Θ , Λ , β , γ , λ , σ_θ^2 y σ_α^2 se asume que son mutuamente independientes y los restantes parámetros μ_θ , μ_γ , ξ_θ , μ_α , μ_λ , ξ_α y μ_β se fijan de tal manera que reflejen rangos plausibles. La curva SROC, puede ser obtenida a partir de los valores esperados de $\Lambda + \lambda Z$ y β . Si se le asigna el valor de $\frac{1}{2}$ para los casos que tienen la enfermedad y $-\frac{1}{2}$ para los casos que no tienen la enfermedad, para un valor dado de covarianza Z_i , la tasa de verdaderos positivos queda expresada en 3.46, Rutter and Gatsonis (158). Luego la curva SROC es dibujada, colocando en un plano los pares $(FP, VP(FP))$, para $FP \in [0, 1]$.

$$\left. \begin{array}{l}
 \Theta \sim \text{Uniforme} [\mu_{\theta_1}, \mu_{\theta_2}] \\
 \gamma \sim \text{Uniforme} [\mu_{\gamma_1}, \mu_{\gamma_2}] \\
 \sigma_{\theta}^2 \sim \Gamma^{-1} [\xi_{\theta_1}, \xi_{\theta_2}] \\
 \Lambda \sim \text{Uniforme} [\mu_{\alpha_1}, \mu_{\alpha_2}] \\
 \lambda \sim \text{Uniforme} [\mu_{\lambda_1}, \mu_{\lambda_2}] \\
 \sigma_{\alpha}^2 \sim \Gamma^{-1} [\xi_{\alpha_1}, \xi_{\alpha_2}] \\
 \beta \sim \text{Uniforme} [\mu_{\beta_1}, \mu_{\beta_2}]
 \end{array} \right\} \text{Mutuamente independientes} \quad (3.45)$$

$$VP(FP) = \text{logit}^{-1} \left[\left(\text{logit}(FP) \exp \left\{ \frac{E(\beta)}{2} \right\} + E(\Lambda + \lambda Z_i) \right) \exp \left\{ \frac{E(\beta)}{2} \right\} \right] \quad (3.46)$$

Ajuste del Modelo HSROC

La inferencia del modelo HSROC se fundamenta en una distribución a posteriori de los parámetros del modelo. Debido a que los modelos que se consideran no son conjugados¹, es decir, son expresiones de forma cerradas, las distribuciones a posteriori no existen. Luego estos valores a posteriori son estimados mediante una simulación de la distribución a posteriori, utilizando cadenas de Markov y Monte Carlo (MCMC), Gelfand and Smith (62). Estos valores simulados de las distribuciones a posteriori se utilizan para estimar la distribución marginal de interés, tal como distribuciones a posteriori de los parámetros particulares.

Las distribuciones condicionales de los parámetros del nivel II Θ , Λ , σ , θ^2 , σ , σ^2 , γ , y λ tienen una distribución normal estándar. Por ejemplo, la distribución condicional de Λ dado λ , z , σ , σ^2 , α , μ_{α} es proporcional a una distribución normal con media $\sum_{i=1}^m \frac{\alpha_i - \lambda z_i}{m}$ y varianza $\frac{\sigma_{\alpha}^2}{m}$ en el rango $[\mu_{\alpha_1}, \mu_{\alpha_2}]$. Las varianzas de los parámetros son previas conjugadas, es decir, una previa que al combinarla (conjugarla) con la verosimilitud, la posteriori resultante pertenece a la misma familia de distribuciones que la previa, ecuación 3.47.

¹Una distribución previa (a priori) tal que al combinarla (conjugarla) con la verosimilitud, la posterior resultante pertenece a la misma familia de distribuciones que la previa, entonces se dice que es una previa conjugada.

$$\sigma_{\alpha}^2 | \lambda, z, \alpha, \Lambda, \xi_{\alpha} \sim \Gamma_{-1} \left(\left(\xi_{\alpha_1} + \frac{m}{2} \right); \left(\frac{1}{2} \sum_{i=1}^m (\alpha_1 - \Lambda - \lambda z_i)^2 + \frac{1}{\xi_{\alpha_2}} \right)^{-1} \right) \quad (3.47)$$

De manera similar se obtienen las distribuciones condicionales de Θ y σ_{θ}^2 . La distribución condicional de $\lambda | \Lambda, \sigma_{\alpha}^2, \alpha, \mu_{\lambda}$ es proporcional a una distribución normal con media $\frac{\sum_{i=1}^m z_i (\alpha_i - \Lambda)}{\sum_{i=1}^m z_i^2}$ y varianza $\frac{\sigma_{\alpha}^2}{\sum_{i=1}^m z_i^2}$ en el rango $[\mu_{\lambda_1}, \mu_{\lambda_2}]$. La distribución condicional del nivel I con parámetros $\theta_i, \alpha_i, \beta$ no está distribuida de manera estándar. La distribución condicional de la precisión (exactitud) de los estudios específicos, $\alpha_i | y_i, n_i, z_i, \Lambda, \sigma_{\alpha}^2, \theta_i, \beta$, se obtiene multiplicando una binomial con una normal, Rutter and Gatsonis (158), note expresión 3.48. La distribución condicional de θ_i , tiene la misma forma que la anterior expresión. La distribución condicional de los parámetros de escala $\beta | y, n, Z, \theta, \alpha, \mu_{\beta}$, es proporcional al producto de $2 \times m$ binomiales, Rutter and Gatsonis (158), note expresión 3.49, con probabilidad en el intervalo $[\mu_{\beta_1}, \mu_{\beta_2}]$.

$$\exp \left[-\frac{(\alpha_i - (\Lambda + \lambda z_i))^2}{2\sigma_{\alpha}^2} \right] \prod_{j=1}^2 \binom{n_{ij}}{y_{ij}} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{(n_{ij} - y_{ij})} \quad (3.48)$$

$$\prod_{i=1}^2 \prod_{j=1}^2 \binom{n_{ij}}{y_{ij}} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{(n_{ij} - y_{ij})} \quad (3.49)$$

Elección a priori

Los rangos para Θ , Λ , β deben ser seleccionados de tal manera que reflejen conocimiento de las modalidades de diagnósticos de las revisiones que se están analizando. En general, el intervalo $[-10, 10]$ cubre todos los valores posibles de Θ , Rutter and Gatsonis (158). De manera similar, el intervalo $[-5, 5]$ cubre todos los valores posibles para β . Pero, debido a que se espera que los resultados de las pruebas sean positivas en pacientes que padezcan la enfermedad, el intervalo $[-2, 20]$ cubrirá todos los valores posibles de Λ , Rutter and Gatsonis (158). La selección a priori de la distribución gamma inversa Γ^{-1} para los parámetros de la varianza entre estudios, σ_{θ}^2 y σ_{α}^2 es más complicada. La distribución muestral es potencialmente sensible a la distribución a priori, la cual es usada para modelizar los parámetros de la varianza. La gamma inversa a

priori que explica los parámetros de la varianza debería ser seleccionada de tal forma que refleje un rango apropiado de los parámetros de precisión y los puntos de corte específicos del estudio.

El objetivo de una previa (priori) adecuada para los parámetros de la varianza, es seleccionar una distribución que no asigne valores de probabilidades no muy grandes (pocos realistas). Afortunadamente la elección a priori de γ , es realista ya que se fundamenta en los rangos de los puntos de corte y de los parámetros de precisión, que se han identificado anteriormente.

Parámetros de estimación

El objetivo de la estimación es la descripción de la distribución a posteriori de los parámetros del modelo y los resúmenes estadísticos que son funciones de los parámetros del modelo. Los intervalos de confianza al 95 % son estimados empíricamente por los percentiles a posteriori del 2,5 % y 97,5 % de la simulación. La simetría de la distribución a posteriori es aproximada por las medias de los valores simulados de (Θ, Λ, β) . La asimetría de la distribución a posteriori se aproxima por la mediana de los valores simulados de $(\sigma_\theta^2, \sigma_\alpha^2)$, Rutter and Gatsonis (158).

Diagnóstico

En este punto se analizan las estadísticas de diagnóstico para evaluar una mala especificación del modelo, un sobre ajuste del modelo y en general para identificar datos atípicos. Se pueden seguir para este punto las sugerencias de Weiss (199). Debido a que el número de verdaderos positivos y falsos positivos dentro de los estudios siguen una distribución binomial, se puede comprobar si la mala especificación del modelo se debe a la evaluación de la distribución a priori de cada estudio específico de los parámetros θ y α .

Nótese, que ambos parámetros se distribuyen de una manera normal. Si se asume que, $\gamma = \lambda = 0$, las sumas de cuadrados $S_\theta = \frac{\sum_i (\theta_i - \Theta)^2}{\sigma_\theta^2}$ y $S_\alpha = \frac{\sum_i (\alpha_i - \Lambda)^2}{\sigma_\alpha^2}$, siguen una distribución χ^2 con m grados de libertad, donde m es el número de estudios en el meta-análisis. Si se asume que $\gamma \neq 0$ y $\lambda \neq 0$ en el modelo, la suma de cuadrados tendrá la forma $S_\alpha = \frac{\sum_i (\alpha_i - \Lambda - \lambda Z_i)^2}{\sigma_\alpha^2}$; si las probabilidades en las colas son muy grandes o muy pequeñas, Rutter and Gatsonis (158) sugieren que hay una mala especificación de la distribución a priori. Se tiene que evaluar el

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

supuesto de independencia condicional entre θ y α , estimando estos parámetros directamente de los datos y realizando una inspección mediante un gráfico de dispersión y correlación.

Estas estimaciones se basan en la fórmula 3.43 con $\beta = 0$ y son representadas en las ecuaciones, 3.50, 3.51, donde “cc” denota corrección de continuidad o de Yates¹. Para la evaluación global de la bondad de ajuste del modelo se suelen usar dos estadísticos de discrepancia chi-cuadrado.

El primero, se basa en las estimaciones de los recuentos (counts), note ecuación 3.52, donde y_{ij1} , denota el número de sujeto i que dió positivo en la prueba pero que no tiene la enfermedad ($j = 0$) o que padece la enfermedad ($j = 1$) en el grupo. D_{counts}^2 , se compara con una distribución χ_{2m}^2 . El segundo estadístico de la bondad de ajuste global es el que se fundamenta en la corrección de continuidad de los logaritmos de los odds ratios, y este se estima como en la ecuación 3.53. Donde $\log(OR_{cc})_i$, es la corrección de continuidad de los logaritmos de los odds ratio del i -ésimo estudio.

Resumen estadístico

La estimación **MCMC**, permite estimar los parámetros de las funciones antes descritas, por lo que los resultados pueden ser presentados utilizando una variedad de medidas de resumen. En los resúmenes aparecen medidas como las tasas de verdaderos positivos, falsos positivos y los ratios de máxima verosimilitud, Boyko (14). En el contexto del meta-análisis, estas medidas son estimaciones a posteriori de las combinaciones de las informaciones de los estudios.

Las tasas de VP y FP se estiman como aparecen en las ecuaciones 3.54 y 3.55; donde los ratios de máxima verosimilitud positivos y negativos respectivamente, se estiman a partir de las ecuaciones 3.56 y 3.57.

¹Si la tabla de contingencia (tabla de 2x2), cuenta con un grado de libertad, el valor estadístico esta sobre-estimado. En tal caso, se aplica una corrección en la fórmula del cálculo de χ^2 , que se denomina corrección de Yates. Esta corrección consiste en restar 0.5 cuando la frecuencia observada es mayor que la esperada, y sumar 0.5 en caso contrario.

El enfoque jerárquico del resumen ROC (HSROC), es un modelo que combina el par de estimación de la sensibilidad y especificidad de múltiples estudios que incorpora la variabilidad dentro y entre los estudios en comparación con el modelo de efectos fijos del resumen ROC (SROC). Esto permite flexibilidad en las estimaciones de los resúmenes estadísticos.

El HSROC, describe la variabilidad dentro de los estudios usando una distribución binomial del número de pruebas positivas con pacientes que tienen la enfermedad y no. La variación en los criterios de positividad y la precisión es modelizada usando una distribución normal, con una regresión lineal en la media que permite la dependencia de la covarianza de los niveles en los estudios.

Distribuciones con colas más pesadas (como la t o de cauchy) pueden ser utilizadas en el segundo nivel del modelo jerárquico en vez de una normal. El modelo puede ser extendido para permitir la dependencia entre el criterio de positividad y la precisión. El modelo HSROC, representado por las ecuaciones 3.50 y 3.51, provee una cuantificación más realista de la variabilidad dentro de los estudios que el enfoque de efectos fijos.

$$\theta_i^{(0)} = \frac{\text{logit}(VP_i^{cc}) - \text{logit}(FP_i^{cc})}{2} \quad (3.50)$$

$$\alpha_i^{(0)} = \text{logit}(VP_i^{cc}) - \text{logit}(FP_i^{cc}) \quad (3.51)$$

$$D_{counts}^2 = \sum_i \sum_j \frac{y_{ij1} - E(y_{ij1}|model, data)}{E(y_{ij1}|model, data)}^2 \quad (3.52)$$

$$D_{lg(OR)}^2 = \sum_i \frac{\lg(OR_{cc})_i - E(\lg(OR_{cc})_i|model, data)}{\sqrt{Var(\lg((OR_{cc})_i)|model, data)}}^2 \quad (3.53)$$

$$VP = \text{logit}^{-1} \left(\left(\Theta + \frac{\Lambda}{2} \right) \exp^{-\frac{\beta}{2}} \right) \quad (3.54)$$

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

$$FP = \text{logit}^{-1} \left(\left(\Theta - \frac{\Lambda}{2} \right) \exp^{-\frac{\beta}{2}} \right) \quad (3.55)$$

$$LR+ = \frac{VP}{FP} \quad (3.56)$$

$$LR- = \frac{1 - VP}{1 - FP} \quad (3.57)$$

3.3.5. Meta-análisis bayesiano para la precisión de un test en ausencia de la prueba oro estándar

Se puede decir que el modelo de Moses et al. (132) describe un meta-análisis como odds ratio de diagnóstico (DOR). El enfoque HSROC, Rutter and Gatsonis (158) expresa la sensibilidad y la probabilidad de falsos positivos de cada estudio como una función de un modelo normal bivalente subyacente, mientras que el modelo bivalente de Reitsma et al. (148) supone que el vector (logit(sensibilidad), logit(especificidad)) sigue una distribución normal bivalente. En ausencia de covarianza, las funciones de verosimilitud del HSROC y del modelo bivalente son algebraicamente equivalentes, proporcionando “pooled” idénticos de sensibilidad y especificidad y estimaciones de varianza entre estudios Harbord et al. (76). El modelo HSROC conduce de forma más natural a una curva SROC, mientras que el modelo bivalente genera un “pooled” de sensibilidad y especificidad.

Walter and Jadad (194), proponen una clase de modelo latente, donde asume que el test bajo evaluación y el test de referencia son considerados variables latentes (no observables), del estado verdadero de la enfermedad. Estos autores demostraron que el modelo es identificable al asumir que la sensibilidad y la especificidad siguen siendo la misma, pero la prevalencia varía entre los estudios. En la práctica, la asunción de que la sensibilidad y la especificidad sean idénticas en todos los estudios es difícil de justificar, dada la variabilidad de la población y los aspectos individuales de los diseños de estudios. Chu et al. (22), describen un modelo más general donde la sensibilidad y especificidad de la prueba, tanto en la fase de evaluación y la prueba de referencia, así como la prevalencia, se tratan como efectos aleatorios. Este modelo puede ser conceptualizado como una extensión del modelo de Reitsma et al. (148) para el caso donde no exista el patrón o prueba oro. Sadatsafavi et al. (162), describen un modelo donde un parámetro (i.e., sensibilidad o especificidad de uno de los test) varía en los estudios. Ninguno de los modelos analizados anteriormente se ajusta cuando los estudios utilizan patrones de referencias diferentes en los estudios seleccionados.

Este método es una extensión del HSROC. Asume que la variación de la sensibilidad y especificidad en los estudios aparece debido al empleo de diferentes puntos de corte para definir una prueba positiva y/o diferencias en la precisión del diagnóstico, cuando no existe una prueba de

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

referencia.

Sean J estudios de diagnóstico que son incluidos en un meta-análisis, y que cada estudio provee una tabla de doble entrada, que relaciona el test bajo evaluación T_1 y el test de referencia T_2 . Las dos medidas se asume que son variables dicotómicas, tomando el valor de 1 cuando el test es positivo y 0 cuando es negativo. Ambas variables se supone que son medidas imperfectas de una variable latente subyacente dicotómica D , para el verdadero estado de la enfermedad. Sea t_{1j} y t_{2j} , vectores de resultados del estudio j para T_1 y T_2 respectivamente. Donde la sensibilidad de la prueba de referencia es definida como $S_2 = P(T_2 = 1|D = 1)$ y su especificidad se define por $C_2 = P(T_2 = 0|D = 0)$.

Rutter and Gatsonis (158), notan que los resultados de T_1 se basan en una variable latente subyacente. Se asume que la variable latente Z_1 sigue una distribución normal y que un resultado positivo en T_1 corresponde a valores altos de Z_1 en comparación con un resultado negativo. El modelo asume que entre pacientes con $D = 0$, $Z_1 \sim N\left(-\frac{\alpha_j}{2}, \exp\left(-\frac{\beta}{2}\right)\right)$.

Este modelo puede ser conceptualizado como un modelo de regresión binomial con un enlace probit (probit link). Dentro del j -ésimo estudio, la diferencia en medias de estas dos distribuciones es α_j , y el ratio de su desviación estándar es $\exp(\beta)$. Cada estudio asume un punto de corte diferente, θ_j , que es un valor positivo. Se define una distribución a priori jerárquica sobre la diferencia de medias (o precisión diagnóstica), $\alpha_j \sim N(\Lambda, \sigma_\alpha^2)$, permitiendo una variación en la distribución Z_1 en cada estudio. Del mismo modo, una jerárquica a priori $\theta_j \sim N(\Theta, \sigma_\theta^2)$, que permite una variación de los puntos de corte en los estudios. Esta estructura es equivalente a un modelo jerárquico de dos niveles, un nivel dentro del estudio para los parámetros θ_j y α_j y un nivel para estudiar la variabilidad entre los estudios con parámetros Λ , Θ , y β , los cuales son comunes en todos los estudios.

En base a las asunciones anteriores, la sensibilidad T_1 en el j -ésimo estudio es dada por, $S_{1j} = \Phi\left(-\frac{\theta_j^{-\alpha_j/2}}{\exp(\beta/2)}\right)$, mientras que la especificidad está dada por, $C_{1j} = \Phi\left(\frac{\theta_j^{\alpha_j/2}}{\exp(-\beta/2)}\right)$. Por lo tanto, valores crecientes de θ_j , inducen a una correlación negativa entre la sensibilidad y la especificidad de T_1 a través de los estudios, mientras que aumento de los valores α_j , provoca

correlación positiva entre ellos.

Un índice global para la prueba de la sensibilidad y especificidad puede resumirse como, $\Phi\left(-\frac{\Theta^{-\Lambda/2}}{\exp(\beta/2)}\right)$ y $\Phi\left(\frac{\Theta^{-\Lambda/2}}{\exp(\beta/2)}\right)$ respectivamente. La utilidad de un solo “pooled” de estimación dependerá del grado de heterogeneidad entre los estudios, Macaskill et al. (119).

Estimación

Dendukuri et al. (37), proponen la siguiente estimación: La función de verosimilitud de los datos observados de los J estudios puede ser expresada en términos de la sensibilidad y especificidad de cada test, y la prevalencia en el j -ésimo estudio, $P((D = 1|Estudio = j) = \pi_j)$, puede ser expresada como en la ecuación 3.58. Para llevar a cabo estimación bayesiana, se tienen que especificar distribuciones a priori (previas) sobre el conjunto de parámetros desconocidos. La estrategia es emplear distribuciones previas para la mayoría de los parámetros, utilizando valores a priori de los parámetros. Las previas (priori) para Λ , Θ , y β son seleccionadas de tal manera que la distribución marginal resultante del “pooled” de la sensibilidad y especificidad sea aproximadamente normal en $(0, 1)$.

El “pooled” de diferencias de medias para los parámetros es asumido con una densidad previa (priori) $\Lambda \sim U(-3, 3)$. El logaritmo del ratio entre las dos desviaciones estándar, β , se asume que sigue una distribución uniforme, es decir $U(-0.75, 0.75)$. El “pooled” del punto de corte de los parámetros, Θ , se asume que sigue una distribución uniforme, $U(-1.5, 1.5)$. Los parámetros σ_β y σ_θ se distribuyen como uniforme, $U(0, 2)$. Para los parámetros π_j , S_2 y C_2 se emplea una distribución a priori (previa) beta, si se trata de una distribución objetivo a priori se emplea una Beta $(1, 1)$, Dendukuri et al. (37).

El número total de grados de libertad disponibles es $3J$, con cada estudio contribuyendo 3 grados de libertad. El número total de parámetros a estimar es al menos $J + 7$. Por lo tanto, un mínimo de 4 estudios se requeriría para estimar razonablemente este modelo sin ningún tipo de distribuciones previas informativas. Dado que dos parámetros, S_{1j} y C_{1j} , son definidos como una función que depende de tres parámetros, α_j, θ_j y β , se asume que β es la misma en todos los estudios, para evitar problemas de identificabilidad.

Se puede emplear un paquete de R, HSROC, “**An R package for Bayesian meta-analysis of diagnostic test accuracy**” para la estimación de los parámetros antes mencionados, algoritmo Schiller et al. (166). Para evaluar la convergencia de los modelos se utilizan cinco cadenas diferentes de 50.000 iteraciones a partir de los valores iniciales y se calcula el estadístico de Gelman-Rubin para comparar la variabilidad dentro y entre las cadenas Gelman and Rubin (63).

La dependencia condicional entre el índice y análisis de referencia

En ausencia de una referencia estándar de oro es importante ajustar la dependencia condicional entre múltiples pruebas realizadas sobre los mismos temas, Dendukuri et al. (37), con el fin de ajustar la correlación no explicada entre los ensayos dentro de cada clase latente, Dendukuri and Joseph (36). El modelo en 3.58 se puede extender al ajustar la dependencia condicional de distintas maneras, Dendukuri et al. (35). Una de las maneras de ajustar la dependencia condicional es sumando la covarianza entre la sensibilidad del índice y el test de referencia $cov(S_j)$ en el j -ésimo estudio y entre su especificidad $cov(C_j)$ en el j -ésimo estudio como lo proponen Chu et al. (22), Dendukuri et al. (35), Dendukuri and Joseph (36), teniendo que la probabilidad conjunta entre estos dos test en el j -ésimo estudio se puede representar como se expresa en la ecuación 3.59. Luego se establece una distribución a priori uniforme para la covarianza de los parámetros; $cov(s_j) \sim U(0, \min(S_{1j}, S_2) - S_{1j}S_2)$, $cov(c_j) \sim U(0, \min(C_{1j}, C_2) - C_{1j}C_2)$.

$$\begin{aligned}
 L(\Theta, \Lambda, S_2, C_2, \sigma_\alpha^2, \sigma_\theta^2, \beta, \pi_j, \alpha_j, j = 1, 2, \dots, J | t_{1j}, t_{2j}, j = 1, 2, \dots, J) = \\
 \prod_{j=1}^J \left[\pi_j \Phi \left\{ -\frac{\theta_j^{-\alpha_j/2}}{\exp(\beta/2)} \right\} S_2 + (1 - \pi_j) \Phi \left\{ -\frac{\theta_j^{\alpha_j/2}}{\exp(-\beta/2)} \right\} (1 - C_2) \right]^{t_{ij} \cdot t_{2j}} \times \\
 \times \left[\pi_j \Phi \left\{ -\frac{\theta_j^{-\alpha_j/2}}{\exp(\beta/2)} \right\} (1 - S_2) + (1 - \pi_j) \Phi \left\{ -\frac{\theta_j^{\alpha_j/2}}{\exp(-\beta/2)} \right\} C_2 \right]^{t_{ij} \cdot (1 - t_{2j})} \times \\
 \times \left[\pi_j \Phi \left\{ \frac{\theta_j^{-\alpha_j/2}}{\exp(\beta/2)} \right\} S_2 + (1 - \pi_j) \Phi \left\{ \frac{\theta_j^{\alpha_j/2}}{\exp(-\beta/2)} (1 - C_2) \right\} \right]^{(1 - t_{ij}) \cdot t_{2j}} \times \\
 \times \left[\pi_j \Phi \left\{ \frac{\theta_j^{-\alpha_j/2}}{\exp(\beta/2)} \right\} (1 - S_2) + (1 - \pi_j) \Phi \left\{ \frac{\theta_j^{\alpha_j/2}}{\exp(-\beta/2)} C_2 \right\} \right]^{(1 - t_{ij}) \cdot (1 - t_{2j})} \quad (3.58)
 \end{aligned}$$

$$\begin{aligned} P(T_1 = u, T_2 = v | \text{Estudio} = j) &= \pi_j \left[S_{ij}^u (1 - S_{ij})^{(1-u)} S_2^v (1 - S_2)^{(1-v)} (-1)^{|u-v|} \text{cov}(S_j) \right] + \\ &= +(1 - \pi_j) \left[C_{ij}^{(1-u)} (1 - C_{ij})^u C_2^{(1-v)} (1 - C_2)^v + (-1)^{|u-v|} \text{cov}(C_j) \right] \end{aligned} \tag{3.59}$$

3.3.6. Unificación de modelos para Meta-Análisis de la precisión diagnóstica

Existe un creciente interés en las revisiones sistemáticas y meta-análisis de los datos de los estudios de la precisión de diagnóstico, Bossuyt et al. (13), Deeks (32), Devillé et al. (39), Gluud and Gluud (67), Khan et al. (98), Tatsioni et al. (185), Whiting et al. (203). Los datos de cada estudio primario se resumen en una tabla de 2x2, sobre la base de resultados de una prueba dicotómica contra el verdadero estado de la enfermedad, de la cual se extraen medidas conocidas como, la sensibilidad y la especificidad.

Los métodos estadísticos antes estudiados y propuestos por Dukic and Gatsonis (44), Moses et al. (132), Reitsma et al. (148), Rutter and Gatsonis (158), Siadaty and Shu (174), reflejan dos principales características, primero una correlación negativa entre la sensibilidad y la especificidad es explicada por los diferentes tipos de umbrales en las pruebas, Deeks (32), Moses et al. (132). Segundo se espera una sustancial heterogeneidad entre los estudios, la cual debe ser incorporada en los modelos, Lijmer et al. (111).

El enfoque inferencial de estos métodos es sujeto de ciertos debates. Algunos autores, proponen estimaciones de las medidas de sensibilidad y especificidad o regiones de predicción, dentro de las cuales pueden contener estudios con poca credibilidad, Reitsma et al. (148), mientras que otros sugieren que en presencia de heterogeneidad, los resultados de los meta-análisis deben presentarse como una curva (SROC), Rutter and Gatsonis (158).

Algunos autores han propuesto un método de generación de una curva SROC, utilizando una regresión lineal simple, que es el enfoque más utilizado con frecuencia, Littenberg and Moses (114). Asimismo, los supuestos de la regresión lineal no se cumplen, por lo que el método es una aproximación. También existe incertidumbre al momento de elegir los pesos de la regresión, Rutter and Gatsonis (158), Walter (193). Dos métodos estadísticos se han propuesto, para superar estos problemas descritos anteriormente, Reitsma et al. (148), Rutter and Gatsonis (158), el enfoque bivalente y el HSROC respectivamente.

3.3.7. Relación entre el modelo bivalente y HSROC

El modelo bivalente permite la inclusión de covarianzas que afectan a la sensibilidad y/o especificidad, mientras que el enfoque HSROC permite introducir covariaciones a los parámetros de precisión y/o los umbrales de las pruebas. Un modelo HSROC que introduce covariaciones a la precisión y a los umbrales de las pruebas es equivalente a un enfoque bivalente si en el modelo se introducen las mismas covariaciones del modelo HSROC pero a nivel de sensibilidad y especificidad Harbord et al. (76). Una interpretación más formal se detalla a continuación:

Sea $b = \exp\left(\frac{\beta}{2}\right)$, entonces el modelo HSROC, ecuación 3.43, nivel I queda expresado en las ecuaciones 3.60 y 3.61.

$$\text{logit}(\pi_{i1}) = b^{-1} \left(\theta_i + \frac{1}{2} \alpha_i \right) \quad (3.60)$$

$$\text{logit}(\pi_{i0}) = b \left(\theta_i - \frac{1}{2} \alpha_i \right) \quad (3.61)$$

Si en el modelo bivalente, ecuación 3.35, reemplazamos θ_{Ai} y θ_{Bi} por μ_{Ai} y μ_{Bi} , las cuales representan las transformaciones logit de la sensibilidad y especificidad, se obtienen las ecuaciones 3.62 y 3.63 respectivamente. Luego, se pueden relacionar las variables aleatorias que forman la base de los dos modelos, de la manera que se aprecia en las expresiones 3.64 y 3.65. Este par de ecuaciones, describen como μ_{Ai} y μ_{Bi} son combinaciones lineales de dos variables aleatorias, θ_i y α_i , que en el modelo HSROC son variables normales independientes. Luego se puede inferir que cualquier par de combinaciones lineales de variables aleatorias con distribuciones normales, tiene una distribución normal bivalente, Dudewicz and Mishra (43). Se sabe que la distribución conjunta de μ_{Ai} y μ_{Bi} es normal bivalente, en el modelo HSROC.

$$\mu_{Ai} = \text{logit}(\pi_{i1}) \quad (3.62)$$

$$\mu_{Bi} = \text{logit}(1 - \pi_{i0}) = \ln\left(\frac{1 - \pi_{i0}}{\pi_{i0}}\right) = -\ln\left(\frac{\pi_{i0}}{1 - \pi_{i0}}\right) = -\text{logit}(\pi_{i0}) \quad (3.63)$$

$$\mu_{Ai} = b^{-1}\left(\theta_i + \frac{1}{2}\alpha_i\right) \quad (3.64)$$

$$\mu_{Bi} = -b\left(\theta_i - \frac{1}{2}\alpha_i\right) \quad (3.65)$$

Luego, el modelo HSROC es equivalente a un modelo bivariante. Las ecuaciones 3.64 y 3.65 se pueden expresar en forma matricial, 3.66. Invertiendo el resultado anterior se tiene la ecuación 3.67. Donde S , es la matriz de transformación asociada, con los cambios desde las coordenadas del modelo bivariante (transformaciones logit de la sensibilidad y especificidad) hasta las coordenadas del modelo HSROC (puntos de corte y parámetros de precisión). Nótese que la matriz S , no es ortogonal $S^{-1} \neq S^T$.

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} = S^{-1} \begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix} \text{ donde } S^{-1} = \begin{pmatrix} b^{-1} & \frac{1}{2}b^{-1} \\ -b & \frac{1}{2}b \end{pmatrix} \quad (3.66)$$

$$\begin{pmatrix} \theta_{Ai} \\ \alpha_{Bi} \end{pmatrix} = S \begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \text{ donde } S = \begin{pmatrix} \frac{1}{2}b & -\frac{1}{2}b \\ b & b^{-1} \end{pmatrix} \quad (3.67)$$

3.3.8. Relación de los parámetros de los modelos sin covarianzas

Para expresar la relación entre los parámetros de los dos modelos, sin covarianzas, en términos de la matriz de transformación S , tomando la esperanza y la varianza de la ecuación 3.67, se tienen las expresiones, 3.68 y 3.69. La asunción del modelo HSROC de que θ_i y α_i no están correlacionadas, es decir los elementos fuera de la diagonal principal son ceros, fija el valor de b y por lo tanto la matriz de transformación S , donde S es una matriz no ortogonal que diagonaliza la matriz de varianza y covarianza del modelo bivariante.

$$E\left(\begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix}\right) = E\left(S\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix}\right) = \begin{pmatrix} \Theta \\ \Lambda \end{pmatrix} = S\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix} \quad (3.68)$$

$$\text{var}\left(\begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix}\right) = \text{var}\left(\alpha\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix}\right) = \begin{pmatrix} \sigma_\theta^2 & 0 \\ 0 & \sigma_\alpha^2 \end{pmatrix} = S \cdot \text{cov}\left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}\right) S^T = S\Sigma_{AB}S^T \quad (3.69)$$

Si se desarrollan las operaciones del lado derecho de la ecuación 3.69, se tiene la ecuación 3.70. Donde, los elementos fuera de la diagonal principal de 3.70, son iguales a cero si $b = \sqrt{\frac{\sigma_B}{\sigma_A}}$ o su equivalente $\beta = \log\left(\frac{\sigma_B}{\sigma_A}\right)$. Luego el parámetro de forma del modelo HSROC β , está determinado únicamente por la variación de los logit de la sensibilidad y especificidad del modelo bivalente. Se puede inferir de las ecuaciones 3.68 y 3.69, los otros parámetros del HSROC con el modelo bivalente, 3.71. Reemplazando $b = \sqrt{\frac{\sigma_B}{\sigma_A}}$ en la ecuación 3.71, se obtienen las relaciones 3.72 y 3.73. Reemplazando $b = \sqrt{\frac{\sigma_B}{\sigma_A}}$ en 3.70, se obtiene las ecuaciones 3.74 y 3.75.

$$\begin{pmatrix} \sigma_\theta^2 & 0 \\ 0 & \sigma_\alpha^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{4}b^2\sigma_A^2 - \frac{1}{2}\sigma_{AB} + \frac{1}{4}b^{-2}\sigma_B^2 & \frac{1}{2}b^2\sigma_A^2 - \frac{1}{2}b^{-2}\sigma_B^2 \\ \frac{1}{2}b^2\sigma_A^2 - \frac{1}{2}b^{-2}\sigma_B^2 & b^2\sigma_A^2 + 2\sigma_{AB} + b^{-2}\sigma_B^2 \end{pmatrix} \quad (3.70)$$

$$\begin{pmatrix} \Theta \\ \Lambda \end{pmatrix} = \begin{pmatrix} \frac{1}{2}b\mu_A - \frac{1}{2}b^{-1}\mu_B \\ b\mu_A + b^{-1}\mu_B \end{pmatrix} \quad (3.71)$$

$$\Theta = \frac{1}{2} \left[\sqrt{\frac{\sigma_B}{\sigma_A}} \mu_A - \sqrt{\frac{\sigma_A}{\sigma_B}} \mu_B \right] \quad (3.72)$$

$$\Lambda = \sqrt{\frac{\sigma_B}{\sigma_A}} \mu_A + \left[\sqrt{\frac{\sigma_B}{\sigma_A}} \right]^{-1} \mu_B \quad (3.73)$$

$$\sigma_\theta^2 = \frac{1}{2} (\sigma_A\sigma_B - \sigma_{AB}) \quad (3.74)$$

$$\sigma_\alpha^2 = 2 (\sigma_A\sigma_B + \sigma_{AB}) \quad (3.75)$$

3.3.9. Relación de los parámetros de los modelos con covarianza

Si el modelo bivalente se extiende, incluyendo una covarianza que afecta a la sensibilidad y especificidad, las ecuaciones 3.67 y 3.68 pueden ser escritas como 3.76 y 3.77. Por lo tanto, un modelo bivalente, que se le agregue una o más covarianzas que afecten la sensibilidad y la especificidad es equivalente a un modelo HSROC, donde las mismas covarianzas afectan tanto a los parámetros de precisión como a los puntos de corte. Sin embargo, si en el modelo bivalente se incluyen covarianzas que afecten a la sensibilidad o especificidad, no será equivalente a un modelo HSROC, a menos que las restricciones se impongan sobre la relación entre los coeficientes del modelo HSROC.

$$E \begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix} = S \begin{pmatrix} \mu_A + v_A Z_i \\ \mu_B + v_B Z_i \end{pmatrix} \quad (3.76)$$

$$E \begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix} = S \begin{pmatrix} \Theta + \gamma Z_i \\ \Lambda + \lambda Z_i \end{pmatrix} \quad (3.77)$$

3.3.10. Enfoque de inferencia y salidas de los modelos

Modelo HSROC

El modelo HSROC, genera una curva SROC, al permitir que el parámetro θ varíe, mientras que el parámetro de precisión α se mantiene fijo en su media. Para el caso donde no exista covarianza, la sensibilidad esperada para una especificidad dada viene dada por 3.78, Macaskill (118), Rutter and Gatsonis (158). Rutter and Gatsonis (158), sugieren que la curva se restrinja dentro del rango de las especificidades estimadas de los estudios. Si $\beta = 0$, la curva es simétrica (*sensibilidad = especificidad*). Esta curva ROC, no representa la incertidumbre de los parámetros estimados, pero representa la variabilidad del umbral.

$$\text{logit}(\text{sensibilidad}) = \Lambda \exp^{\beta/2} - \exp^{\beta} \text{logit}(\text{especificidad}) \quad (3.78)$$

Modelo Bivariante

Reitsma et al. (148), sugieren regiones de confianza y de predicción para la curva ROC, que pueden ser construidas utilizando las estimaciones de un modelo bivariante. A medida que la correlación entre la sensibilidad y la especificidad aumentan, los intervalos de confianza para la media de los logit de la sensibilidad y especificidad (μ_A, μ_B) se separan, lo cual puede ser engañoso; por lo cual es preferible utilizar una región de confianza elíptica para ambos parámetros. Tal región se genera utilizando una representación paramétrica, ecuaciones 3.79 y 3.80, Douglas (41).

$$\mu_A = \hat{\mu}_A + S_A C \cos(t) \quad (3.79)$$

$$\mu_B = \hat{\mu}_B + S_B C \cos(t + \arccos(r)) \quad (3.80)$$

Donde S_A, S_B , son las estimaciones de los errores estándar de μ_A y μ_B y r es la estimación de su correlación y variando t , $0 \leq t \leq 2\pi$ se generan los bordes de las elipses. La constante “ C ” se denomina, constante de borde de la elipse, Alexandersson et al. (2). Asintóticamente, provee una región de confianza de $100(1 - \alpha\%)$, con $C = \sqrt{\chi_{2,\alpha}^2}$. Cuando el número de estudios es pequeño, es recomendable utilizar una medida más conservadora para la región de confianza dada por $C = \sqrt{2F_{2,(n-2);\alpha}}$, donde n es el número de estudios y $F_{2,(n-2);\alpha}$ sigue una distribución de Fisher, con 2 y $(n-2)$ grados de libertad, Chew (21), Dudewicz and Mishra (43).

También es posible construir una región de predicción que incluya los verdaderos valores de la sensibilidad y especificidad de estudios futuros. La matriz de covarianzas de los verdaderos valores de los logit de la sensibilidad y especificidad de estudios futuros quedan expresados como $\Sigma_{AB} = \text{var}\left(\begin{matrix} \hat{\mu}_A \\ \hat{\mu}_B \end{matrix}\right)$.

3.3.11. Modelo Estadístico para I^2

Cuando se estima el tamaño del efecto de cada estudio individual, el meta-análisis los integra calculando un tamaño del efecto promedio, ajustando la heterogeneidad alrededor de la media estimada. En general, el modelo estadístico más realista es el que integra la estimación del efecto del meta-análisis en un modelo de efectos aleatorios, porque incorpora las fuentes de heterogeneidad entre los estudios.

La cuantificación del grado de heterogeneidad entre una colección de estudios es uno de los aspectos más problemáticos de un meta-análisis. Es importante porque puede afectar a la decisión sobre la elección del modelo estadístico fijo o de efectos aleatorios. Por otro lado, si se encuentra una variabilidad significativa, se pueden detectar las variables potenciales responsables de esta variabilidad. La varianza entre los estudios, τ^2 , es el parámetro en el modelo estadístico que representa principalmente a la verdadera heterogeneidad entre los verdaderos efectos de los estudios. Por lo tanto, un buen procedimiento para determinar si existe una verdadera heterogeneidad entre una colección de estudios debe ser relacionado positivamente con τ^2 . Al mismo tiempo, no debe ser afectado por el número de estudios, y debe ser libre de escala con el fin de ser comparable entre los meta-análisis que han aplicado diferentes índices.

Sea d el mejor estimador del parámetro del tamaño del efecto δ , ecuación 3.81, siendo \bar{Y}_E y \bar{Y}_C las medias muestrales del grupo experimental y control respectivamente y S un estimador de la desviación estándar intra grupo, Hedges (80), estimada por la ecuación 3.82, Hedges (80), donde S_E^2 , S_C^2 , n_E y n_C representan las varianzas muestrales y los tamaños de la muestra de los grupos experimental y control respectivamente. El término $c(m)$ es un factor de corrección para el sesgo positivo sugerido para estandarizar la diferencia de medias cuando el tamaño de la muestra es pequeño y es estimado por la ecuación 3.83, Hedges (80), donde $m = n_E + n_C - 2$. La varianza muestral del índice d es estimada por la ecuación 3.84, Hedges (80).

$$d = c(m) \frac{\bar{Y}_E - \bar{Y}_C}{S} \quad (3.81)$$

$$S = \sqrt{\frac{(n_E - 1) S_E^2 + (n_C - 1) S_C^2}{n_E + n_C - 2}} \quad (3.82)$$

$$c(m) = 1 - \frac{3}{4 \cdot m - 1} \quad (3.83)$$

$$S_d^2 = \frac{n_E + n_C}{n_E \cdot n_C} + \frac{d^2}{2(n_E + n_C)} \quad (3.84)$$

Sea T_i , la estimación del i -ésimo efecto de una colección de k estudios ($i = 1, 2, 3, \dots, k$). Donde T_i corresponde a la ecuación 3.81. En un modelo de efectos aleatorios cada efecto T_i , es estimador del tamaño del efecto paramétrico θ_i con varianza condicional σ_i^2 y estimada por $\hat{\sigma}_i^2$. El estimador condicional de la varianza $\hat{\sigma}_i^2$ viene dado por la ecuación 3.84. Luego el modelo puede ser propuesto $T_i = \theta_i + \varepsilon_i$, donde ε_i , son independientes y distribuidos normalmente con media cero y varianza σ_i^2 , $\varepsilon_i \sim N(0, \sigma_i^2)$. La varianza condicional representa la variabilidad intra estudio, que es la variabilidad producida por el error del muestreo.

A su vez, el parámetro del tamaño del efecto θ_i , se refiere a un parámetro de efecto distribuido con media μ_θ y varianza incondicional τ^2 . Donde cada parámetro puede ser definido como $\theta_i = \mu_\theta + u_i$, se asume que los errores u_i son normales e independientes con media cero y varianza τ^2 [$u_i \sim N(0, \tau^2)$]. El valor τ^2 , representa el verdadero valor de heterogeneidad entre los efectos del estudio producidos por innumerables factores sustantivos (tipo de tratamiento, características del paciente), metodológicos (tipo de diseño, tamaño de la muestra, desgaste, etc.) y características de los estudios, Lipsey (113). Por otro lado T_i , se distribuye normalmente con media μ_θ y varianza $\tau^2 + \sigma_i^2$ [$T_i \sim N(\mu_\theta, \tau^2 + \sigma_i^2)$].

Cuando no existe variabilidad entre los efectos estimados, la varianza de la variabilidad inter estudio es cero ($\tau^2 = 0$), existiendo variabilidad por el error de muestreo, que es representada en el modelo por la varianza incondicional intra estudio, σ_i^2 . En este caso todos los estudios estiman un único parámetro para el tamaño del efecto $\theta_i = \theta$ y el modelo estadístico simplificado queda expresado como $T_i = \theta + \varepsilon_i$, llegándose a convertir en un modelo de efectos fijos.

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

El modelo de efectos fijos puede ser considerado como un caso particular de un modelo de efectos aleatorios, cuando no existe variabilidad inter estudio, y como consecuencia la estimación del efecto T_i , se ve afectada por el error de muestreo σ_i^2 siguiendo una distribución normal con media $\theta = \mu_\theta$ y varianza σ_i^2 [$T_i \sim N(\mu_\theta, \sigma_i^2)$], cuando el tamaño de la muestra es **grande**.

La estimación usual del tamaño promedio del efecto consiste en estimar cada efecto estimado, T_i , por el inverso de su varianza w_i , note ecuación 3.85. En el modelo de efectos fijos, el peso del factor es estimado como $w_i = \frac{1}{\hat{\sigma}_i^2}$, y cuando el modelo es de efectos aleatorios w_i , es estimado como $w_i = \frac{1}{\tau^2 + \hat{\sigma}_i^2}$. Para el índice \mathbf{d} la estimación de la varianza intra estudio es denotada por σ_i^2 y definida por la ecuación 3.84. Un estimador para la varianza inter estudio τ^2 , se basa en el método de los momentos propuesto por DerSimonian and Laird (38), ecuación 3.86. El valor de c se estima mediante la ecuación 3.87, donde w_i son los pesos del factor para el i -ésimo estudio asumiendo un modelo de efectos fijos ($w_i = \frac{1}{\hat{\sigma}_i^2}$), k es el número de estudios y Q es el test estadístico para la heterogeneidad propuesto por Cochran (26) y definido por Hedges (80), mediante la ecuación 3.88, donde \bar{T} se la estima mediante la ecuación 3.85.

$$\bar{T} = \frac{\sum_i w_i T_i}{\sum_i w_i} \quad (3.85)$$

$$\tau^2 = \begin{cases} \frac{Q - (k - 1)}{c} & \text{para } Q > k - 1 \\ 0 & \text{para } Q \leq k - 1 \end{cases} \quad (3.86)$$

$$c = \sum w_i - \frac{\sum w_i^2}{\sum w_i} \quad (3.87)$$

$$Q = \sum w_i (T_i - \bar{T})^2 \quad (3.88)$$

Un procedimiento para detectar si hay una significativa heterogeneidad entre un conjunto de estudios debe estar relacionado directamente con el valor de τ^2 y a la vez libre de escalas, con el objetivo de ser comparable. Si se asume que las varianzas condicional o intra estudio, σ_i^2 son conocidas, entonces bajo la hipótesis de homogeneidad ($H_0 : \delta_1 = \delta_2 = \dots = \delta_k; o H_0 : \tau^2 = 0$), el estadístico Q tiene una distribución χ^2 con $k - 1$ grados de libertad. Por lo tanto, para valores

superiores de Q con relación al punto crítico para un nivel de significación dada (α), permite rechazar la hipótesis nula y concluir que no es estadísticamente significativa la variación inter estudio.

Un problema con el estadístico Q es que su potencia estadística depende del número de estudios, siendo muy bajo o muy alto para un pequeño o un gran número de estudios. Para resolver los problemas del estadístico Q y la no compatibilidad de la varianza inter estudio, τ^2 entre los meta-análisis con diferentes tamaño de efectos (métricas distintas), Higgins and Thompson (83), han propuesto el índice I^2 . El índice I^2 , expresión 3.89 cuantifica el grado de heterogeneidad de un conjunto de tamaños de efectos, comparando el valor de Q con su valor esperado asumiendo homogeneidad, con $k - 1$ grados de libertad.

$$I^2 = \begin{cases} \frac{Q - (k - 1)}{Q} \times 100\% & \text{para } Q > k - 1 \\ 0 & \text{para } Q \leq k - 1 \end{cases} \quad (3.89)$$

Cuando el estadístico Q es menor que sus grados de libertad, entonces I^2 es truncado a cero. El índice I^2 puede ser interpretado como un porcentaje de heterogeneidad, esto es, la parte total de variación que es debida a la varianza inter estudios $\hat{\tau}^2$. Existe una relación directa entre τ^2 e I^2 y puede ser formalizada a partir de las ecuaciones 3.86 y 3.89, mediante la expresión 3.90.

$$I^2 = \frac{c \cdot \hat{\tau}^2}{Q} \quad (3.90)$$

Es importante notar que el modelo bivalente representa una mejora al enfoque univariante (SROC). Por otro lado el enfoque bivalente si se le añade una o más covariaciones que afecten la sensibilidad y la especificidad será equivalente a un modelo HSROC, vea Figura 3.12.

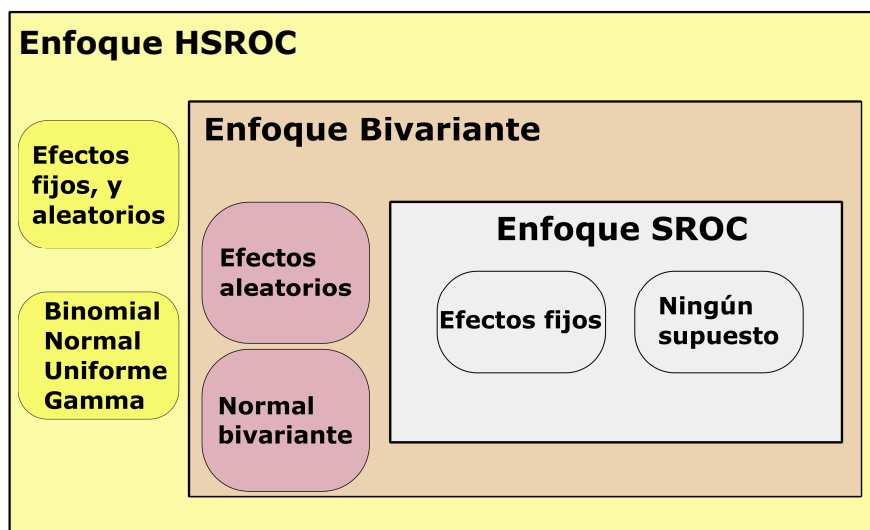


Figura 3.12: Diagrama de enfoques HSROC, Bivariante, SROC, elaboración propia

Los modelos antes analizados (bivariante y HSROC), no informan la media global, sino más bien la media de un estudio central y tienen diferencias estimando la correlación entre sensibilidad y especificidad cuando el número de estudios en el meta-análisis es pequeño y / o cuando la varianza entre estudios es relativamente grande, Riley et al. (149). Recientemente se ha propuesto un modelo mixto de cópula que modelizan la relación entre la sensibilidad y especificidad mediante una estructura funcional utilizando el concepto de cópulas, capturando la dependencia bivalente de los datos de manera natural sin recurrir a transformaciones. Por lo expresado anteriormente, en la Sección 3.3.12 se amplía con mayor detalle dicho concepto.

3.3.12. Meta-análisis para estudios de precisión diagnóstica: un modelo estadístico utilizando distribuciones beta-binomiales y cópulas bivariantes

En esta sección, se analizan las estrategias para combinar los resultados de las pruebas con el fin de aumentar la precisión diagnóstica, usando el concepto de cópula. La cópula es un concepto estadístico bien conocido para modelar la dependencia de variables aleatorias. Una cópula es una función de distribución conjunta cuyos marginales están todos distribuidos uniformemente y se puede usar para modelizar la dependencia por separado de las distribuciones marginales, es decir, es una función que aproxima el comportamiento conjunto de variables aleatorias a partir de sus comportamientos individuales.

A partir de la cópula es posible construir medidas de dependencia, diferentes al coeficiente de correlación de Pearson, cuyas definiciones se presentarán más adelante, e identificar características propias de cada una de ellas.

Si bien los métodos estadísticos para el meta-análisis de los ensayos clínicos están bien desarrollados y comprendidos hoy en día, todavía hay desafíos cuando se realizan meta-análisis de estudios sobre la exactitud diagnóstica. Esto se debe principalmente a la naturaleza bivalente de la respuesta, donde la información sobre sensibilidad y especificidad debe resumirse teniendo en cuenta su correlación dentro de un solo ensayo. Esta correlación puede ser inducida por poblaciones o entornos heterogéneos en los diferentes ensayos, pero se deriva principalmente del equilibrio entre sensibilidad y especificidad cuando el umbral para definir la positividad de la prueba varía, Leeflang et al. (106).

Los métodos estadísticos estudiados en las secciones anteriores, tienen algunas desventajas. No funcionan en la escala original de sensibilidad y especificidad, por lo que se recurre a una transformación en la escala logit, la cual generalmente depende de la distribución normal bivalente, permitiendo una sola estructura de correlación. Lo que es más importante, la estimación de máxima verosimilitud es complicada porque la función de verosimilitud correspondiente no tiene forma cerrada, sino que es un producto de integrales, que no pueden ser resueltos analíticamente.

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

te. Esto requiere la integración numérica, Verde (190), o algún método de aproximación, Menke (126), Paul et al. (142). Además, estos modelos no informan la media global, sino más bien la media de un estudio central y tienen diferencias estimando la correlación entre sensibilidad y especificidad cuando el número de estudios en el meta-análisis es pequeño y/o cuando la varianza entre estudios es relativamente grande, Riley et al. (149).

La modelización conjunta de la sensibilidad y la especificidad de los estudios utilizando distribuciones **beta bivariantes** mediante **cópulas** superan las dificultades antes mencionadas. Dado que tanto la sensibilidad como la especificidad toman valores en el espacio de intervalo $(0, 1)$, es una opción más natural utilizar una distribución **beta** para describir su distribución a través de los estudios, sin necesidad de ninguna transformación. La distribución **beta** se **conjuga** con la **distribución binomial** y por lo tanto es fácil integrar los efectos aleatorios analíticamente dando lugar a las distribuciones marginales **beta-binomiales**.

Este enfoque de cópulas presenta la siguiente estructura. En el primer nivel de la jerarquía, dada la sensibilidad y especificidad de cada estudio, se utilizan dos distribuciones binomiales para describir la variación en el número de verdaderos positivos y verdaderos negativos entre los individuos enfermos y sanos, respectivamente. En el segundo nivel, se modelizan las sensibilidades y especificidades no observadas usando una distribución bivalente, es decir se fundamenta en una representación de cópulas para la distribución de efectos aleatorios.

Medidas de dependencia

En la práctica, el coeficiente de correlación lineal de Pearson es la medida de dependencia más usada. Sin embargo, como se muestra posteriormente, esta medida presenta ciertos inconvenientes. A continuación se describen varias definiciones de medidas y conceptos de dependencia y sus implicaciones.

Definición 3.3.1 (Coeficiente de Correlación de Pearson) *Una de las aproximaciones más utilizadas para analizar relaciones de dependencia es el coeficiente de correlación de Pearson, ρ , el cual mide la relación lineal existente entre un vector de variables aleatorias. Para la pareja*

de variables aleatorias (X_i, Y_j) , el coeficiente de correlación de Pearson, está definido por:

$$\rho_{i,j} = \frac{Cov(X_i, X_j)}{Var(X_i)Var(X_j)} \quad (3.91)$$

El coeficiente de correlación de Pearson tiene diferentes ventajas, las cuales lo han hecho una medida de dependencia de amplia aceptación en muchos contextos, principalmente porque es fácil de calcular, ya que sólo se necesita estimar los dos primeros momentos de los datos observados. Además, es invariante ante transformaciones afines positivas ¹. Finalmente, una de las grandes ventajas que tiene el coeficiente de correlación de Pearson es su relación con la función de distribución normal multivariante, en donde resume toda la relación de dependencia existente entre las variables aleatorias Embrechts et al. (51).

Sin embargo, es necesario ser cauteloso cuando se utiliza el coeficiente de correlación como medida de dependencia, ya que puede presentar ciertos inconvenientes. Algunos de estos inconvenientes son tratados por Embrechts et al. (51) y entre los cuales se resaltan los siguientes:

- **El coeficiente de correlación es una medida de dependencia escalar.** Al ser una medida escalar, este resume toda la relación de dependencia entre los factores de riesgo en un solo número. Aunque esto facilita la interpretación, puede tener inconvenientes en la toma de decisiones.
- Variables perfectamente dependientes, no significa que el coeficiente de correlación sea igual a 1. McNeil et al. (125).
- Que el coeficiente de correlación sea cero no implica independencia entre los factores de riesgo.
- El coeficiente de correlación es invariante únicamente ante transformaciones afines positivas.
- El coeficiente de correlación únicamente se encuentra definido para parejas de variables aleatorias que tengan varianzas finitas, Embrechts et al. (51).

¹Una transformación afín positiva T sobre X_i está definida de la siguiente forma: $T(X_i) = aX_i + b$, $a > 0$

Cóputas

Para evitar los problemas anteriormente mencionados, es posible modelar la dependencia entre d factores a través de una función que satisface algunas propiedades particulares que la hacen idónea para cumplir esta tarea. Esta aproximación tiene la ventaja de presentar el concepto de dependencia como una estructura que describa completamente la relación entre los factores, en lugar de tratar de resumirla en un sólo número. No obstante, pensar en una estructura de este tipo implica una mayor complejidad en su interpretación.

La función mencionada en el párrafo anterior es conocida como cópula, la cual se define de la siguiente manera:

Definición 3.3.2 (Nelsen (133)) Una función $C : [0, 1]^d \rightarrow [0, 1]$ es una cópula d -dimensional, si:

- $C(u_1, \dots, u_d)$ es creciente en cada componente u_i
- $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ para todo $i \in \{1, 2, \dots, d\}$, $u_i \in [0, 1]$
- Para todo $(a_1, \dots, a_d), (b_1, \dots, b_d) \in [0, 1]^d$ con $a_i \leq b_i$ se tiene que:

$$\sum_{i_1=1}^2 \dots \sum_{i_d=1}^2 (-1)^{i_1+\dots+i_d} C(u_{1i_1}, \dots, u_{di_d}) \geq 0 \quad (3.92)$$

Donde $u_{j1} = a_j$ y $u_{j2} = b_j$ para todo $j \in \{1, 2, \dots, d\}$

- Para cada cópula $C(u_1, \dots, u_d)$ se tienen las siguientes cotas:

$$\max \left\{ \sum_{i=1}^d u_i + 1 - d, 0 \right\} \leq C(u) \leq \min \{u_1, \dots, u_d\} \quad (3.93)$$

Las cópulas son una herramienta valiosa para construir la función de distribución conjunta de múltiples variables aleatorias. Para obtener las funciones de distribución acumulada correspondientes, es necesario el Teorema de **Sklar**.

Teorema 3.3.1 (Teorema de Sklar, Nelsen (133) o Sklar (175)) Sea H una función de distribución conjunta con marginales $F_i(x)$. Entonces, existe una cópula C tal que para todo $x_i \in \mathbb{R} \cup \{-\infty, \infty\}$, $i = 1, 2, \dots, d$:

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (3.94)$$

Si las marginales son continuas, entonces C es única. Por otro lado, C está definida únicamente en el $\text{Ran}(F_1) \times \dots \times \text{Ran}(F_d)$ donde $\text{Ran}(F_i)$ denota el rango de la función F_i . Si C es una cópula y $F_i, i = 1, 2, \dots, d$ son funciones de distribuciones, entonces, 3.94 es una función de distribución conjunta con marginales $F_i(x)$.

La demostración del teorema es dada por varios autores, ver por ejemplo Schweizer and Sklar (170) o Rüschendorf (156)

Modelización de la dependencia

El concepto de cópulas permite modelizar la dependencia entre las distribuciones marginales a través de los parámetros de la cópula. Estos parámetros de asociación se pueden transformar en medidas de correlación, como los coeficiente de correlación de Spearman “ $\rho_s(X_1, X_2)$ ” o el de Kendall “ $\tau(X_1, X_2)$ ”. Schweizer and Wolff (171), mostraron que ambas medidas podrían usarse para la descripción de la dependencia. A continuación, los dos coeficientes de correlación se explican con más detalles con respecto a su relación con las cópulas.

Definición 3.3.3 (Coeficiente de correlación de Kendall) Sean (X_1, Y_1) y (X_2, Y_2) variables aleatorias independientes e idénticamente distribuidas “iid”. Entonces, τ es dado por:

$$\tau = \tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \quad (3.95)$$

En el contexto de cópulas, el τ de Kendall es definido por Nelsen (133) de la siguiente manera:

Teorema 3.3.2 Sean X e Y variables aleatorias con cópula C . El τ de Kendall es dado por:

$$\tau_{X,Y} = \tau_C = Q(C, C) = 4 \iint_{I^2} C(u, v) dC(u, v) - 1 \quad (3.96)$$

Donde $Q = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]$.

Demostración: Sea F y G las distribuciones marginales de X_1, X_2 e Y_1, Y_2 respectivamente. C_1 y C_2 son las cópulas de (X_1, Y_1) y (X_2, Y_2) . Por lo tanto, se deduce que $H_1(x, y) =$

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

$C_1(F(x), G(y))$ y $H_2(x, y) = C_2(F(x), G(y))$. Q denota la diferencia entre las probabilidades de concordancia y discordancia, en la cual son concordantes si $(X_i - X_j)(Y_i - Y_j) > 0$ y discordantes si $(X_i - X_j)(Y_i - Y_j) < 0$ Nelsen (133). Para variables aleatorias continuas se aplica que $P[(X_1 - X_2)(Y_1 - Y_2) < 0] = 1 - P[(X_1 - X_2)(Y_1 - Y_2) > 0]$ y por lo tanto:

$$Q = 2P[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1 \quad (3.97)$$

Además, $P[(X_1 - X_2)(Y_1 - Y_2) > 0] = P[(X_1 > X_2, Y_1 > Y_2)] + P[(X_1 < X_2, Y_1 < Y_2)]$ se puede lograr integrando la distribución de uno de los vectores (X_1, Y_1) o X_2, Y_2 . Empleando (X_1, Y_1) , se tiene:

$$\begin{aligned} P[X_1 > X_2, Y_1 > Y_2] &= P[X_2 < X_1, Y_2 < Y_1] \\ &= \iint_{\mathbb{R}^2} P[X_2 \leq x, Y_2 \leq y] dC_1(F(x), G(y)) \\ &= \iint_{\mathbb{R}^2} C_2(F(x), G(y)) dC_1(F(x), G(y)) \end{aligned}$$

Sea $u = F(x)$ y $v = G(y)$. De lo que obtenemos,

$$P[X_1 > X_2, Y_1 > Y_2] = \iint_{I^2} C_2(u, v) dC_1(u, v)$$

Por otro lado, es claro que:

$$\begin{aligned} P[X_1 < X_2, Y_1 < Y_2] &= \iint_{\mathbb{R}^2} P[X_2 > x, Y_2 > y] dC_1(F(x), G(y)) \\ &= \iint_{\mathbb{R}^2} [1 - F(x) - G(y) + C_2(F(x), G(y))] dC_1(F(x), G(y)) \\ &= \iint_{\mathbb{R}^2} [1 - u - v + C_2(u, v)] dC_1(u, v) \end{aligned}$$

Es evidente que $E(U) = E(V) = 1/2$, porque C_1 es la distribución conjunta del par (U, V) que se distribuyen de acuerdo a una variable aleatoria uniforme $(0, 1)$. Por lo tanto,

$$P[X_1 < X_2, Y_1 < Y_2] = 1 - \frac{1}{2} - \frac{1}{2} + \iint_{I^2} C_2(u, v) dC_1(u, v) = \iint_{I^2} C_2(u, v) dC_1(u, v)$$

Por lo que se obtiene la siguiente expresión,

$$P[(X_1 - X_2)(Y_1 - Y_2) > 0] = 2 \iint_{I^2} C_2(u, v) dC_1(u, v)$$

En consecuencia, empleando la expresión 3.97,

$$Q = 2P[(X_1 - X_2)(Y_1 - Y_2)] - 1 = 4 \iint_{I^2} C_2(u, v) dC_1(u, v) - 1$$

Por último, se obtiene el τ de Kendall:

$$\tau_{X,Y} = \tau_C = Q = Q(C, C) = 2P[(X_1 - X_2)(Y_1 - Y_2)] - 1 = 4 \iint_{I^2} C_2(u, v) dC_1(u, v) - 1 \quad (3.98)$$

□

Definición 3.3.4 (Coeficiente de correlación de Spearman) Sea (X_1, Y_1) , (X_2, Y_2) y (X_3, Y_3) variables aleatorias “*iid*”. Entonces, el coeficiente de correlación de Spearman es definido como:

$$\rho_{X,Y} = \rho_S = 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0])$$

Teorema 3.3.3 Sean X e Y dos variables aleatorias continuas con su correspondiente cópula C . El coeficiente de correlación de Spearman es dado por:

$$\rho_{X,Y} = \rho_C = 3Q(C, \Pi) = 12 \iint_{I^2} uv dC(u, v) - 3 = 12 \iint_{I^2} C(u, v) dudv - 3$$

Donde $\Pi = \Pi(u, v) = uv$.

Demostración: La demostración es simple, empleando la ecuación 3.98, se tiene que:

$$\begin{aligned} \rho_{X,Y} &= 3Q(C, \Pi) \\ &= 12 \iint_{I^2} uv dC(u, v) - 3 \\ &= 12 \iint_{I^2} C(u, v) dudv - 3 \end{aligned}$$

□

Modelización estadística

Se supone que cada estudio individual ($i = 1, 2, \dots, I$) en el meta-análisis reporta una tabla 2×2 con valores de verdaderos positivos VP_i , verdaderos negativos VN_i , falsos positivos FP_i

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

y falsos negativos FN_i y además con medidas resumen establecidas por la sensibilidad Se_i y especificidad Es_i . Se asume además que VP_i y VN_i se distribuyen de acuerdo a una distribución binomial, Kuss et al. (103), vea ecuación 3.99.

$$\begin{aligned} VP_i | Se_i &\sim \text{Binomial}(VP_i + FN_i, Se_i) \\ VN_i | Es_i &\sim \text{Binomial}(VN_i + FP_i, Es_i) \end{aligned} \quad (3.99)$$

Las distribuciones “**marginales**” para la sensibilidad y especificidad se asumen como distribuciones beta con parámetros α , β , es decir, haciendo una suposición sobre Se y Es en su escala original $[0, 1]$, Kuss et al. (103), vea ecuación 3.100.

$$\begin{aligned} Se &\sim \text{Beta}(\alpha_{Se}, \beta_{Se}) \\ Es &\sim \text{Beta}(\alpha_{Es}, \beta_{Es}) \end{aligned} \quad (3.100)$$

La correspondiente función de densidad para la sensibilidad es definida por la ecuación 3.101.

$$f(u, \alpha_{Se}, \beta_{Se}) = \begin{cases} \frac{\Gamma(\alpha_{Se} + \beta_{Se})}{\Gamma(\alpha_{Se})\Gamma(\beta_{Se})} u^{\alpha_{Se}-1} (1-u)^{\beta_{Se}-1} & 0 < u < 1 \\ 0 & \text{en otro caso} \end{cases} \quad (3.101)$$

Análogamente definida para la especificidad, donde $\Gamma(x)$ es la función Gamma con $\Gamma(n+1) = n!$ para $n \in \mathbb{Z}^+$.

Los valores esperados $E(u)$ para las distribuciones betas son definidos como $\alpha_*/(\alpha_* + \beta_*)$ y describen los parámetros meta-analíticos de interés (Se , Es). Adicionalmente, las estimaciones de las varianzas $(\alpha_*\beta_*) / [(\alpha_* + \beta_* + 1)(\alpha_* + \beta_*)^2]$ que explican la variabilidad debido a la heterogeneidad en los estudios.

La distribución *beta* se utiliza debido a su flexibilidad y al hecho de que está conjugada a la distribución binomial. La ventaja es que se obtiene distribuciones *beta-binomiales* a posteriori que son comúnmente conocidas.

Teorema 3.3.4 (La distribución beta se conjuga con la distribución binomial.)

Demostración: [empleando la sensibilidad, vea 3.100] Sea $(VP = d|Se = u) \sim \text{Binomial}(D, u)$, $u \in [0, 1]$, $d \in \mathbb{N}$ y Se se distribuye de acuerdo a una *beta*.

El Teorema de Bayes lleva a:

$$f(Se = u|VP = d) = \frac{P(VP = d|Se = u) f(u)}{\int_0^1 P(VP = d|Se = u) f(u) du}$$

Donde f es la densidad de la distribución beta y $D = VP + FN$. Luego:

$$\begin{aligned} P(VP = d|Se = u) f(u) &= \binom{D}{d} u^d (1-u)^{D-d} \frac{\Gamma(\alpha_{Se} + \beta_{Se})}{\Gamma(\alpha_{Se}) \Gamma(\beta_{Se})} u^{\alpha_{Se}-1} (1-u)^{\beta_{Se}-1} \\ &= \binom{D}{d} u^{d+\alpha_{Se}-1} (1-u)^{D-d+\beta_{Se}-1} \frac{\Gamma(\alpha_{Se} + \beta_{Se})}{\Gamma(\alpha_{Se}) \Gamma(\beta_{Se})} \end{aligned}$$

$$\begin{aligned} \int_0^1 P(VP = d|Se = u) f(u) du &= \binom{D}{d} \frac{\Gamma(\alpha_{Se} + \beta_{Se})}{\Gamma(\alpha_{Se}) \Gamma(\beta_{Se})} \underbrace{\int_0^1 u^{d+\alpha_{Se}-1} (1-u)^{D-d+\beta_{Se}-1} du}_{\frac{\Gamma(d + \alpha_{Se}) \Gamma(D - d + \beta_{Se})}{\Gamma(D + \alpha_{Se} + \beta_{Se})}} \\ &= \binom{D}{d} \frac{\Gamma(\alpha_{Se} + \beta_{Se})}{\Gamma(\alpha_{Se}) \Gamma(\beta_{Se})} \frac{\Gamma(d + \alpha_{Se}) \Gamma(D - d + \beta_{Se})}{\Gamma(D + \alpha_{Se} + \beta_{Se})} \end{aligned}$$

$$\begin{aligned} f(Se = u|VP = d) &= \frac{\binom{D}{d} \frac{\Gamma(\alpha_{Se} + \beta_{Se})}{\Gamma(\alpha_{Se}) \Gamma(\beta_{Se})} u^{d+\alpha_{Se}-1} (1-u)^{D-d+\beta_{Se}-1}}{\binom{D}{d} \frac{\Gamma(\alpha_{Se} + \beta_{Se})}{\Gamma(\alpha_{Se}) \Gamma(\beta_{Se})} \frac{\Gamma(d + \alpha_{Se}) \Gamma(D - d + \beta_{Se})}{\Gamma(D + \alpha_{Se} + \beta_{Se})}} \\ &= \begin{cases} u^{d+\alpha_{Se}-1} (1-u)^{D-d+\beta_{Se}-1} \frac{\Gamma(d + \alpha_{Se}) \Gamma(D - d + \beta_{Se})}{\Gamma(D + \alpha_{Se} + \beta_{Se})} & , 0 < u < 1 \\ 0 & \text{en otro caso} \end{cases} \end{aligned}$$

Finalmente, obtenemos una distribución a posteriori beta para $(Se = u|VP = d)$ con:

$$(Se = u|VP = d) \sim \text{Beta}(\alpha_{Se} + u; \beta_{Se} + D - u)$$

Por lo tanto, la distribución a posteriori pertenece a la misma familia que la distribución a priori. □

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

Finalmente, se logran obtener distribuciones **beta-binomiales** Held (81) para VP y VN :

$$\begin{aligned}
 P(VP = d) &= \int_0^1 P(VP = d|Se = u) f(u; \alpha_{Se}, \beta_{Se}) du \\
 &= \binom{D}{d} \frac{\Gamma(\alpha_{Se} + \beta_{Se})}{\Gamma(\alpha_{Se}) \Gamma(\beta_{Se})} \underbrace{\int_0^1 u^{d+\alpha_{Se}-1} (1-u)^{D-d+\beta_{Se}-1} du}_{\frac{\Gamma(d + \alpha_{Se}) \Gamma(D - d + \beta_{Se})}{\Gamma(D + \alpha_{Se} + \beta_{Se})}} \\
 &= \binom{D}{d} \frac{\Gamma(\alpha_{Se} + \beta_{Se})}{\Gamma(\alpha_{Se}) \Gamma(\beta_{Se})} \frac{\Gamma(d + \alpha_{Se}) \Gamma(D - d + \beta_{Se})}{\Gamma(D + \alpha_{Se} + \beta_{Se})}
 \end{aligned}$$

Análogamente, definimos para VN . La modelización presentada es un modelo de efectos aleatorios. Esto significa que cada sujeto “o estudio” tiene dos parámetros “Se y Es”. Estos parámetros siguen una distribución común. En consecuencia, se tiene una distribución marginal de forma cerrada, en contraste con la mayoría de otros modelos de efectos aleatorios tratados en secciones anteriores.

Para modelizar la correlación potencial entre VP y VN , aplicamos el concepto de cópulas. En nuestro caso, se construye una función de distribución acumulada bidimensional “(cdf)”. Por lo tanto, por el Teorema de Sklar, Sklar (175):

$$H(x_1, x_2) = C(F_{Se}(x_1), F_{Es}(x_2))$$

Ahora se determina la función de densidad de probabilidad (pdf) simplemente diferenciando la función de distribución conjunta. Obtenemos (considerando la derivación de Radon-Nikodym):

$$h(x_1, x_2) = f_{Se} f_{Es} c_{12}(F_{Se}(x_1), F_{Es}(x_2))$$

Donde c_{12} , denota la densidad de una cópula bivalente. La densidad se trata como una función de verosimilitud. Nótese, que en el caso de la cópula, la verosimilitud se determina analíticamente y tiene una forma cerrada. Por lo tanto, los métodos estándar de máxima verosimilitud se utilizan para la estimación de parámetros.

Una ventaja del modelo de la cópula es que, en principio, hay una gran cantidad de cópulas que permiten diferentes estructuras de asociación. Esto es un contraste con el modelo estándar donde se usa una distribución normal bivalente como una sola estructura de correlación. Los

tipos de cópulas potenciales solo están limitados por el rango de asociación. Este rango varía desde “**perfectamente negativo**” (-1) a “**perfectamente positivo**” ($+1$).

Existen un rico número de cópulas que podrían ser utilizadas, cada una resultando en un nuevo modelo para el meta-análisis de estudios de precisión diagnóstica. Para ser específico, utilizamos las siguientes cópulas, la cópula bivalente, Gauss, Clayton, Farlie-Gumbel-Morgenstern y Frank.

Definición 3.3.5 (Cópula de Gauss) *La Cópula de Gauss C_G es definida por:*

$$C_G(u_1, u_2) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2) | \Gamma)$$

Donde $\Phi_2(\cdot | \Gamma)$ es la función de distribución de una Normal Estándar Bivalente $N(0, \Gamma)$ con matriz de correlación Γ . Φ^{-1} hace referencia a la inversa de una función normal estándar univariante. Se puede mostrar la densidad de una cópula de Gauss es, Xue-Kun Song (207):

$$c_G(u_1, u_2) = |\Gamma|^{-1/2} e^{\frac{1}{2} q^T (I_2 - \Gamma^{-1}) q}$$

Donde $q = (q_1, q_2)^T$ representa un vector de valores normales estandarizados $q_j = \Phi^{-1}(u_j)$, $j = 1, 2$ e I_2 es una matriz identidad de dimensión 2.

La cópula de Gauss es un miembro de la clase de cópulas elípticas que se puede generalizar a más de dos dimensiones. El parámetro de asociación debe interpretarse como una correlación de Pearson de los valores normales estandarizados; esto es numéricamente cercano a una correlación de Spearman de las marginales originales Xue-Kun Song (207).

Esta cópula, es simétrica y tiene dependencia nula en las colas o muy baja, a no ser que las dos variables tengan correlación igual a 1, note Figura 3.13.

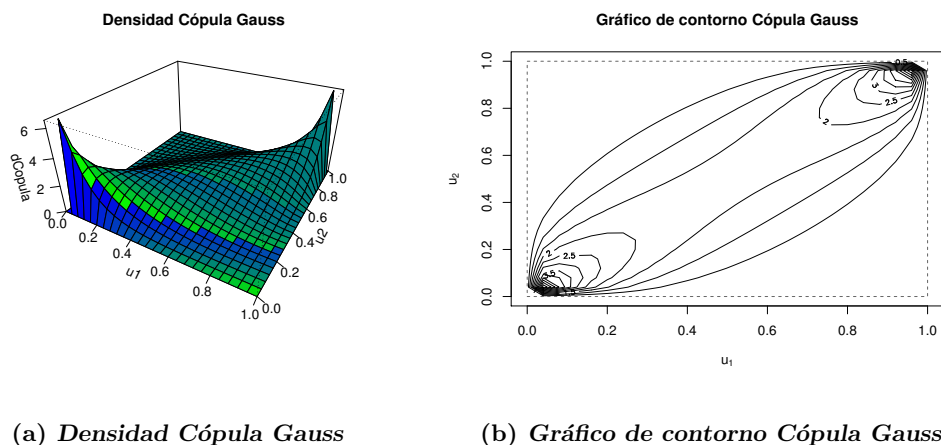


Figura 3.13: Densidad y gráfico de contorno Cópula de Gauss

Definición 3.3.6 (Cópula de Clayton) La cópula de Clayton es definida por Clayton (25) de la siguiente manera:

$$C_C(u_1, u_2, \theta_C) = \left(\max(u_1^{-\theta_C} + u_2^{\theta_C} - 1, 0) \right)^{-\frac{1}{\theta_C}}, \theta_C \in [-1, \infty) \setminus \{0\}$$

Su densidad es dada por la expresión siguiente:

$$c_C(u_1, u_2, \theta_C) = (1 + \theta_C) (u_1 u_2)^{-\theta_C - 1} \left(u_1^{-\theta_C} + u_2^{-\theta_C} - 1 \right)^{-2 - \frac{1}{\theta_C}}$$

para el caso continuo. La cópula de Clayton es miembro de la clase de cópulas de Arquímedes.

Para medir la correlación entre las marginales, se usa Tau de Kendall, $\tau = \frac{\theta_C}{\theta_C + 2}$

Una cópula de Clayton tiene dependencia simétrica en las colas. De hecho, tiene dependencia igual a cero en la cola superior pero dependencia positiva en la cola inferior cuando $\theta_C > 0$, lo que hace la cópula de Clayton pueda recoger distintos grados de dependencia, con dependencia perfecta positiva cuando $\theta_C \rightarrow \infty$, vea Figura 3.14

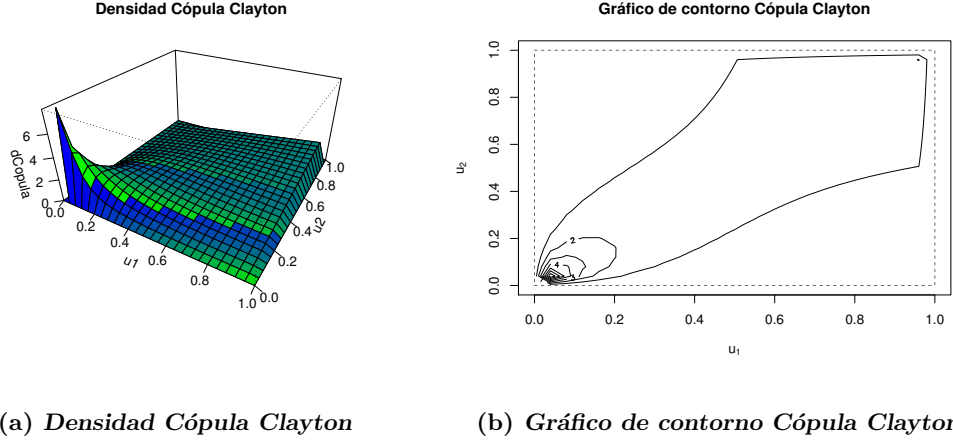


Figura 3.14: Densidad y gráfico de contorno Cópula de Clayton

Definición 3.3.7 (Cópula de Frank) *Esta cópula pertenece a la familia de cópulas de Arquímedes, fue introducida por Frank (57). La forma funcional y la densidad de la cópula están dadas por las siguientes ecuaciones:*

$$C_F(u_1, u_2, \theta_F) = -\frac{1}{\theta_F} \log \left[1 + \frac{(e^{-\theta_F u_1} - 1)(e^{-\theta_F u_2} - 1)}{e^{\theta_F} - 1} \right]$$

$$c_F(u_1, u_2, \theta_F) = \frac{\theta_F (1 - e^{-\theta_F}) e^{-\theta_F(u_1+u_2)}}{[1 - e^{-\theta_F} - (1 - e^{-\theta_F u_1})(1 - e^{-\theta_F u_2})]^2}$$

Dado que $\theta_F \in \mathbb{R}$, tanto la correlación positiva y negativa puede ser modelizada, por lo que esta es una de las cópulas más completa. Cuando $\theta_F = 0$, la sensibilidad y la especificidad son independientes. Para $\theta_F > 0$, la sensibilidad y la especificidad muestran dependencia en el cuadrante positivo y dependencia en el cuadrante negativo cuando $\theta_F < 0$, nótese Figura 3.15. La correlación de Spearman y el tau de Kendall pueden expresarse en términos de θ_F como:

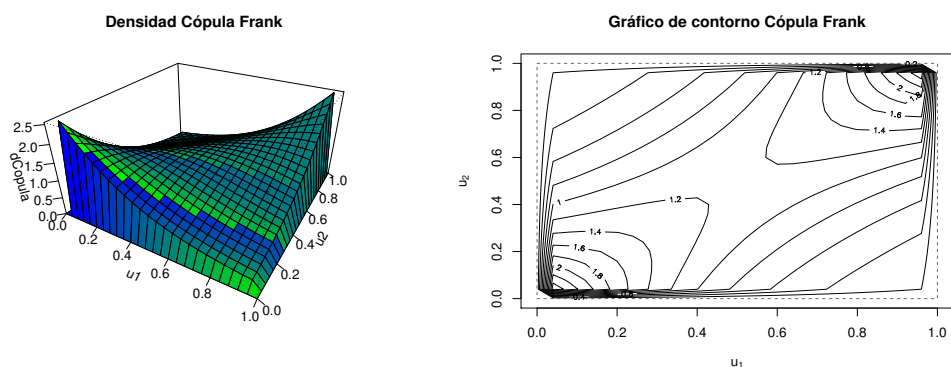
$$\rho_S = 1 - 12 \frac{D_1(\theta_F) - D_2(\theta_F)}{\theta_F}$$

$$\tau = 1 + 4 \frac{D_1(\theta_F) - 1}{\theta_F}$$

Donde $D_j(\delta)$ es la función de Debye definida como:

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

$$D_j(\delta) = \frac{j}{\delta^j} \int_0^\delta \frac{t^j}{e^t - 1} dt, j = 1, 2.$$



(a) Densidad Cópula Frank

(b) Gráfico de contorno Cópula Frank

Figura 3.15: Densidad y gráfico de contorno Cópula de Clayton

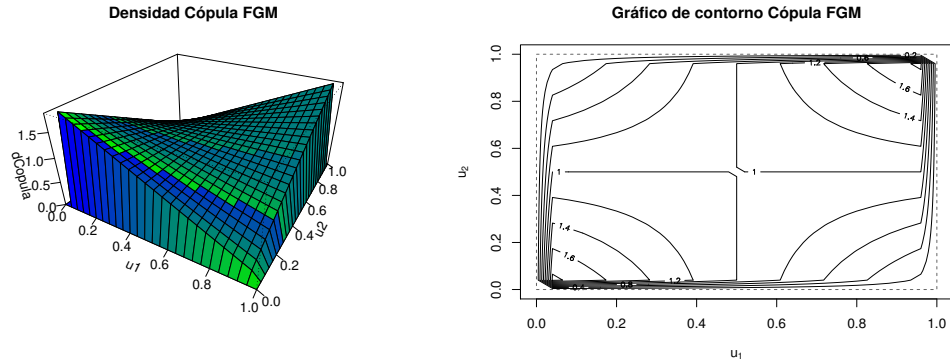
Definición 3.3.8 (Cópula de Farlie-Gumbel-Morgenstern - FGM) Esta cópula estudiada por Farlie (54), Gumbel (70) y Morgenstern (131) se define como:

$$C_{FGM}(u_1, u_2, \theta_{FGM}) = (u_1 u_2) (1 + \theta_{FGM} (1 - u_1) (1 - u_2))$$

$$c_{FGM}(u_1, u_2, \theta_{FGM}) = [1 + \theta_{FGM} (2u_1 - 1) (2u_2 - 1)]$$

Debido a que $\theta_{FGM} \in (-1, 1)$, vea Figura 3.16, la correlación de Spearman y el tau de Kendall se expresan en términos de θ_{FGM} como $\theta_{FGM}/3$ y $2\theta_{FGM}/9$ respectivamente, haciendo que esta cópula sólo sea apropiada para los datos con dependencia débil puesto que:

$$|\rho_S| \leq 1/3$$



(a) *Densidad Cópula FGM* (b) *Gráfico de contorno Cópula FGM*

Figura 3.16: Densidad y gráfico de contorno Cópula de Clayton

Como se observa en los gráficos anteriores, una cópula no tiene el aspecto habitual de las funciones de densidad, es decir, generalmente es elevada en las colas, indicando la importancia de la dependencia en las colas, y reducida en el centro del rango de valores de las variables aleatorias.

3.3.13. Algoritmo y Esquema de Decisión para meta-análisis de pruebas diagnósticas en enfermedades de baja prevalencia

Los métodos antes descritos, se pueden resumir en la Tabla 3.5, donde se detallan las bondades y supuestos que encierra cada modelo. Para entender el funcionamiento de los modelos analizados en este capítulo, es necesario emplear algún programa estadístico, en nuestro caso se utilizará R, Stata y Sas, para obtener las salidas correspondientes que cada enfoque genera, vea Tabla 3.6. La misma muestra las principales medidas que se estiman para el enfoque univariante y la función que se debe emplear en R, mayor detalle vea la Sección 4.1.7 del Capítulo 4.

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

Enfoque	Descripción	Supuestos	Estimación de Parámetros
SROC	Describe un meta-análisis como odds ratio diagnóstico (DOR). Transforma la sensibilidad y especificidad en un indicador de la precisión. No cuantifica el error en S, que puede ser una fuente de sesgo en las estimaciones de parámetros y los resúmenes de la curva. Reconoce únicamente variación debido al tamaño de la muestra.	Ninguno	Mínimos cuadrados
Bivariante	Supone que el vector $(\text{logit}(\text{sensibilidad}); \text{logit}(\text{especificidad}))$ sigue una distribución normal bivalente. Preserva las dos dimensiones de los datos a través del análisis. Utiliza un enfoque de efectos aleatorios.	Binomial condicionada y Normal bivarian-	Principio de máxima verosimilitud
HSROC	Expresa la sensibilidad y la probabilidad de falsos positivos de cada estudio como una distribución normal bivalente. Pone un mejor manejo de la variabilidad dentro y entre estudios, aplicando un enfoque bayesiano para la estimación de los parámetros	Nivel I (intra) - Binomial. Nivel II (inter) - Normal Nivel III (parámetros) - Uniforme y Gamma	Simulación. Método Montecarlo de Cadena de Markov
Cópulas	Modelización conjunta de la sensibilidad y la especificidad del estudio utilizando distribuciones beta bivalente	Utiliza la idea de tener distribuciones marginales beta para sensibilidad y especificidad, resultando en distribuciones marginales binomiales correspondientes para verdaderos positivos y verdaderos negativos, y vinculando a los marginales por distribuciones de cópula	Cópula gaussiana bivalente de Frank FGM Cópula de Clayton

Tabla 3.5: Resumen de modelos estadísticos en pruebas diagnósticas

Función-R	Método	Efectos	Estimación	Gráficos
madauni	Mantel-Haenszel - MH	Fijos	Media Ponderada DOR	
madauni	DerSimonian and Laird - DSL	Aleatorio	Media Ponderada logarit- mos DOR. \mathbf{Q}, I^2, τ^2	Forest Plot con diamante
phm	Modelo proporcional de Hazards PHM	Fijos y aleatorios	θ, τ, AUC	ROCellipse con bandas de confianza y crosshair

Tabla 3.6: Resumen de las funciones empleadas en R para enfoque univariante

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

Para el enfoque bivalente se tiene la Tabla 3.7, las salidas para este enfoque se analizan con mayor detalle en la Sección 4.1.10 del Capítulo 4. Con relación al paquete Stata y Sas, se tienen la Tabla 3.8 y 3.9, las mismas muestran las funciones a utilizar para realizar un análisis de meta-análisis en pruebas diagnósticas, mayor detalle se aprecia en la Sección 4.2 del Capítulo 4. La Tabla 3.10, denota las principales salidas del modelo HSROC en R, mayor detalle se analiza en la Sección 4.3 del Capítulo siguiente, por último, de lo expresado anteriormente se puede construir la Tabla 3.11, donde se muestran las principales medidas y gráficos para los enfoques antes descritos.

Función-R	Método	Efectos	Estimación	Gráficos
reitsma	The restricted maximum likelihood (REML)	Fijos y Aleatorios	Efectos Fijos: tsen, tfp. Efectos aleatorios: tsen, tfp, sen, fp, σ_A^2 , σ_B^2 , σ_{AB} , σ_{BA} . HSROC: Θ , Λ , σ_β^2 , σ_α^2 , α . AUC	Curva SROC con región de confianza
reitsma	REML bivariente	Meta-regresión Aleatorios	Efectos fijos: tsen, tfp, sen, fp, σ_A^2 , σ_B^2 , σ_{AB} , σ_{BA} .	

Tabla 3.7: Resumen de las funciones empleadas en R para enfoque bivariente

Función	Método	Efectos	Estimación	Gráficos
Stata				
medias	Principio de máxima verosimilitud	Fijos y Aleatorios	<ul style="list-style-type: none"> Sensibilidad Especificidad LR+ LR- DOR Q I^2 ICC Sensibilidad ICC Especificidad Efectos de heterogeneidad entre estudios AUC. 	<ul style="list-style-type: none"> Forest Plot de Sen. y Esp. de Ajuste Bondad de ajuste Normal bivalente Anomalías de influencia Detección de atípicos Gráfico de barras (calidad) Gráfica de caja bivalente Curva SROC con regiones de confianza y predicción Funnel Plot Fagan Plot Diagrama de dispersión odd ratio Gráfico de probabilidad modificatorio

Tabla 3.8: Resumen de las funciones empleadas en Stata para enfoque bivalente

Macro en Sas	Método	Efectos	Estimación	Gráficos
MetaDAS	Emplea una sintaxis Proc NLMIXED, la cual utiliza la estimación de máxima verosimilitud mediante cuadratura gaussiana y un algoritmo de optimización cuasi Newton dual como el optimizador predeterminado.	Fijos y Aleatorios	<ul style="list-style-type: none"> • Sensibilidad • Especificidad • LR+ • LR- • DOR • Información de las covariables • Estado de la convergencia • Medidas de ajuste del modelo • Estimación de los parámetros del enfoque HSROC y bivariente 	<ul style="list-style-type: none"> • Histograma y gráfico de probabilidad normal de las estimaciones bayesianas empíricas de los efectos aleatorios.

Tabla 3.9: Resumen de la macro empleada en SAS para pruebas diagnósticas

Función en R	Método	Efectos	Estimación	Gráficos
HSROC,	Método de Monte Carlo	Fijos y Aleatorios	<ul style="list-style-type: none"> Tamaño de la muestra 	<ul style="list-style-type: none"> Gráfica de sensibilidad
HSROC-Summary	de Cadena de Markov - MCMC		<ul style="list-style-type: none"> Información a priori de $\Theta, \Lambda, \sigma_\alpha, \sigma_\theta$ Información de parámetros inter-estudio, estimación, deviance estándar, τ_i de cada parámetro) MC_error con sus respectivos IC, $\Theta, \Lambda, \beta, \sigma_\alpha, \sigma_\beta, S$ y C 	<ul style="list-style-type: none"> Gráfica de densidad (seguimiento de cada parámetro) Gráfica a posteriori (función de cada parámetro) Curva SROC
			<ul style="list-style-type: none"> Información de parámetros intra-estudio, θ, α, prevalencia, estimador, deviance, MC_error, sensibilidad, especificidad con sus respectivos IC Valores predictivos a posteriori de sensibilidad y especificidad con sus respectivas deviance, MC_error e IC. 	

Tabla 3.10: Resumen de las funciones empleadas en R para enfoque jerárquico-HSROC

Medida - Gráfico	Stata	Mada-R	HSROC-R	CÓPULAS	SAS
Sensibilidad IC	X	X	X	X	X
Especificidad IC	X	X	X	X	X
LR+ IC	X				X
LR- IC	X				X
DOR IC	X	X			X
Q, I^2	X	X			
Eger's Index	X				
Curva SROC, regiones de confianza	X	X	X		
Forest Plot con diamante	X	X		X	
Forest Plot sin diamante					
Hair cross - ROCcellipses		X			
Parámetros del modelo			X		X
HSROC					

Tabla 3.11: Resumen de las principales medidas y gráficos para el enfoque univariante, bivalente y HSROC

Finalmente, presentamos un esquema de decisión claro, Figura 3.17, para usar los programas

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

R, Stata y SAS. En otras palabras, situaciones donde exista correlación entre sensibilidad y especificidad, usamos el programa R paquete MADA, que considera los modelos de efectos fijos, aleatorios y mixtos, muestra resúmenes clásicos, medidas de heterogeneidad y gráficas (foresplot y RocEllipse). Si existieran diferentes puntos de corte en el meta-análisis, usamos el programa STATA paquete MIDAS, que produce salidas bivariantes para heterogeneidad y gráficas (fagan, boxPlot bivariante y SROC). Si se pretende analizar el efecto de covariables que influyan sobre el umbral, la precisión o la curva SROC, empleamos los programas R y SAS mediante los paquetes HSROC, CopulaDTA y MetaDas, los mismos modelizan el enfoque jerárquico de efectos fijos y aleatorios, mostrando la estimación de parámetros desde distribuciones a priori y posteriori.

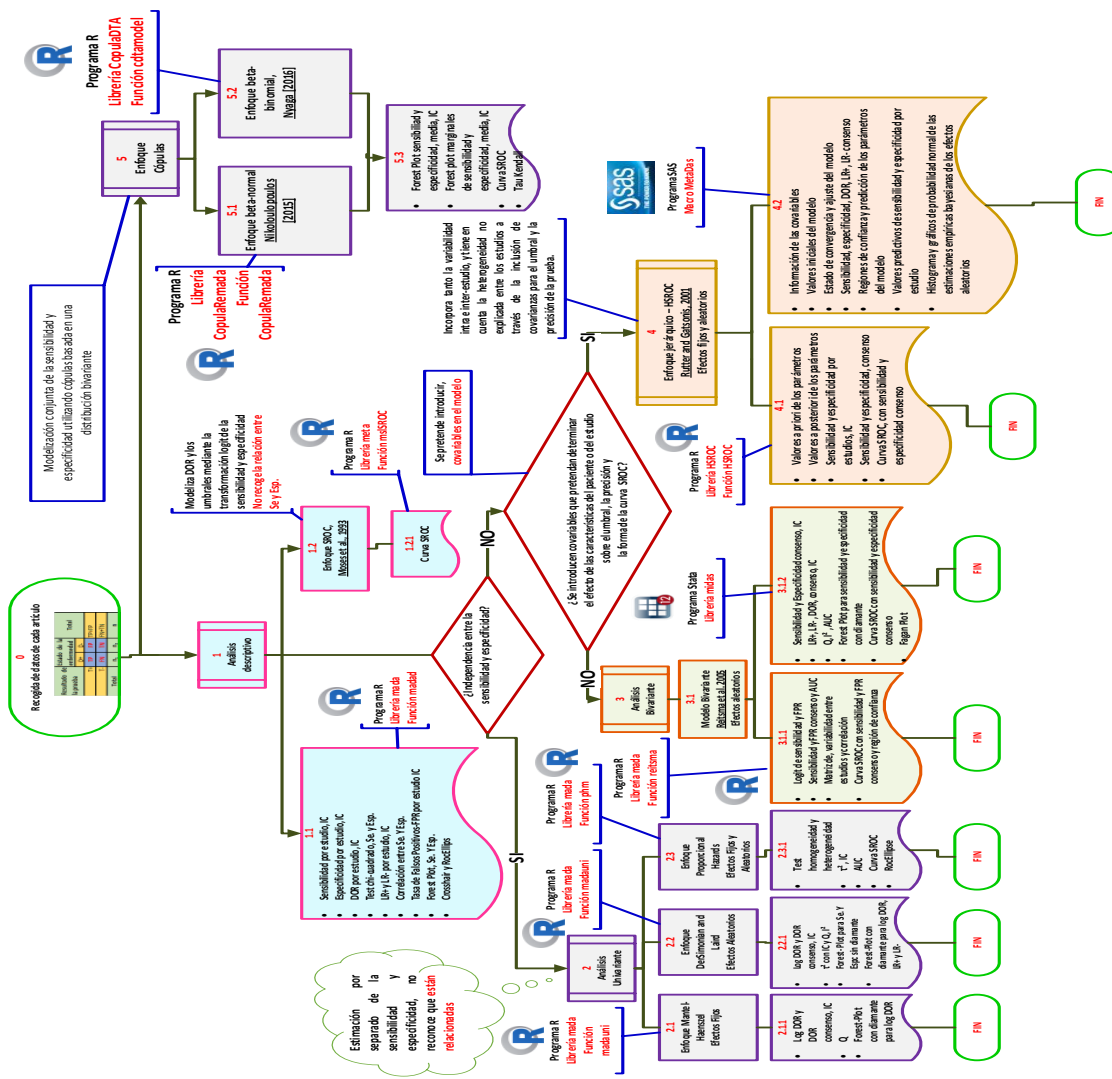


Figura 3.17: Esquema metodológico para el meta-análisis en pruebas diagnósticas

La interpretación de la Figura 3.17 la podemos resumir en los siguientes pasos y en el Algoritmo 1.

1. Realizar un análisis descriptivo de los estudios con el lenguaje R, mediante las librerías **madad** y **msISORC** empleando las funciones **madad** y **msISORC** respectivamente.

3. META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

2. Si existe independencia entre la sensibilidad y especificidad, se procede a realizar un análisis univariante, empleando las funciones **madauni** y **phm** de la librería **mada** del lenguaje R. Este análisis emplea los modelos de Mantel-Haenszel (efectos fijos), DerSimonian-Laird (efectos aleatorios) y el enfoque proporcional de Hazards (efectos fijos y aleatorios).
3. Si la sensibilidad y especificidad están relacionadas, es decir, existen diferentes puntos de corte en el meta-análisis y los datos se ajustan a una distribución normal bivariante, se procede a un análisis bivariante empleando los lenguajes R y Stata mediante las librerías **mada** y **midas**. Nótese, que para poder utilizar el enfoque bivariante en R, se utiliza la función **reitsma**.
4. Si se pretende determinar el efecto de las características o del estudio sobre el umbral, la precisión y la forma de la curva SROC, se procede a utilizar un enfoque jerárquico HSROC, los datos se ajustan a este enfoque jerárquico empleando los paquetes **HSROC** y **MetaDas** de los lenguajes R y SAS respectivamente.
5. Si el número de estudios en el meta-análisis es relativamente “pequeño” (no se cumple el supuesto de normalidad bivariante de las transformaciones logit de la sensibilidad y especificidad de los estudios) y/o la relación entre las varianzas inter e intra estudio es cercana a cero, se procede a utilizar la modelización de cópulas, empleando el paquete **CopulaRemada** o **CopulaDTA** del lenguaje R.

Resultado: Selección del mejor modelo para un meta-análisis de pruebas diagnósticas

Data: Recolectar los meta-análisis de pruebas diagnósticas

Análisis descriptivo, Modelo de Moses ▷ DOR por cada estudio, gráfico de crosshair y forestplot de sensibilidad y especificidad

si *independencia entre sensibilidad y especificidad* **entonces**

| realizar análisis univariante ▷ modelos MH, DSL y PHM con librería mada de
| Rstudio

en otro caso

| *si las transformaciones logit de se. y esp. se distribuyen como normal bivalente*

| **entonces**

| | **si** *se introduce covariable en el modelo* **entonces**

| | | ajustar modelo HSROC ▷ invocar librería HSROC de Rstudio o
| | | MacroMetaDas de SAS

| | **en otro caso**

| | | ajustar modelo bivalente ▷ invocar librería mada de Rstudio o librería
| | | midas de Stata

| | **fin**

| **en otro caso**

| | ajustar modelos de cópulas ▷ invocar librería CopulaRemada o CopulaDTA de
| | Rstudio

| **fin**

fin

Algoritmo 1: Algoritmo para realizar un meta-análisis de pruebas diagnósticas

Una interpretación de los análisis realizados en los Capítulos 2 y 3 se muestra en el Capítulo 4, dando relevancia a la parte conceptual, metodológica y empírica de los modelos y resultados.

Análisis de Resultados

4.1. Precisión Diagnóstica con MADA para Meta-Análisis

El paquete R-mada es una herramienta para la precisión diagnóstica de un meta-análisis. En contraposición con un meta-análisis univariante, el meta-análisis de diagnóstico requiere un enfoque bivariante. El paquete implementa el enfoque de Reitsma et al. (148), en ausencia de covarianza que es equivalente al enfoque de Rutter and Gatsonis (158).

4.1.1. Introducción

Si bien existen muchos métodos para meta-análisis de diagnóstico, aún no es un procedimiento rutinario. Una de las razones puede ser la complejidad del enfoque bivariante, pero otra de las razones es que algunos paquetes para el análisis de meta-análisis, por ejemplo Comprehensive Meta-Analysis and RevMan (Biostat, Inc. 2006; The Nordic Cochrane Centre 2011), no incluyen un ajuste apropiado del modelo para el meta-análisis de diagnóstico. Rutter and Gatsonis (158), le da un enfoque bivariante, pero también existen modelos para meta-análisis de diagnóstico con enfoque bayesiano como por ejemplo WinBUGS, Lunn et al. (117) u OpenBUGS, Lunn et al. (116).

El paquete estadístico “**mada**” escrito para ejecutarse en R (R Core Team 2012), ofrece algunos enfoques actuales para el meta-análisis de diagnóstico, como también funciones que permiten obtener estadísticas descriptivas y gráficos.

4.1.2. Obteniendo el paquete

Si R está instalado y se cuenta con una conexión a internet, el paquete se puede instalar desde el repositorio (CRAN - The Comprehensive R Archive Network). En la dirección <http://r-forge.r-project.org/projects/mada/> se encuentra la versión más actualizada, pero las versiones estables están disponibles desde el CRAN. El paquete se puede cargar escribiendo los siguientes comandos.

```
>install.packages("mada")  
> library("mada")
```

4.1.3. Ingresando los datos

En los estudios primarios de diagnóstico, los resultados de un procedimiento estándar de oro, definen la presencia o ausencia de una condición y el resultado de una prueba diagnóstica (normalmente algún tipo de procedimiento de bajo coste). Los datos de un estudio primario pueden ser reportados en una tabla de 2×2 , Tabla 4.1.

		Enfermedad-Condición	
		Sí	No
Prueba diagnóstica	Sí	y_{i11}	y_{i01}
	No	y_{i10}	y_{i00}
Total		n_{i1}	n_{i0}

Tabla 4.1: Datos del estudio i -ésimo en una tabla de 2×2

Los valores de y_{i11} , y_{i01} , denotan los valores de verdaderos positivos (VP) y falsos positivos (FP) respectivamente, y los valores y_{i10} y y_{i00} muestran los valores de falsos negativos (FN) y verdaderos negativos (VN). Las medidas de precisión diagnóstica, se estiman a partir de una tabla de 2×2 . Usando la nomenclatura de la Tabla 4.1, se tienen las ecuaciones 4.1, 4.2 y 4.3.

$$\theta_{Ai} = \frac{y_{i11}}{n_{i1}} = \text{sensibilidad del } i\text{-ésimo estudio} \quad (4.1)$$

$$1 - \theta_{Bi} = \frac{y_{i01}}{n_{i0}} = \text{tasa de falsos positivos del } i\text{-ésimo estudio} \quad (4.2)$$

$$\theta_{Bi} = \frac{y_{i00}}{n_{i0}} = \text{especificidad del } i\text{-ésimo estudio} \quad (4.3)$$

Básicamente, todas las funciones de *mada* necesitan una tabla de 2×2 . De la fórmula 4.1 se tiene que $y_{i11} = \theta_{Ai}n_{i1}$. Si en un estudio primario se reporta una sensibilidad de 0.944 y que 142 personas tuvieron la condición, se puede calcular y , pero como el resultado anterior no es un entero, se necesita redondearlo. Note las siguientes líneas de comando, para lograr lo antes mencionado:

```
> y <- 142 * .944
> y

[1] 134.048

> round(y)

[1] 134

>
```

Se debe asumir que los valores VP, FP, FN y VN son conocidos para cada estudio. Una buena manera de organizar la información en R, es mediante data frame, el cual puede contener varias variables. Donde cada fila del data frame corresponde a un estudio primario. Como ejemplo se ingresarán en un data frame información de seis estudios de un meta-análisis del AUDIT-C (una prueba corta para detectar problemas de alcohol), Kriston et al. (101), es decir, se tiene la siguiente salida:

```
> AuditC6 <- data.frame(VP = c(47, 126, 19, 36, 130, 84),
+ FN = c(9, 51, 10, 3, 19, 2),
+ FP = c(101, 272, 12, 78, 211, 68),
+ VN = c(738, 1543, 192, 276, 959, 89))
> AuditC6
```

4. ANÁLISIS DE RESULTADOS

```
VP FN  FP  VN
1  47  9 101 738
2 126 51 272 1543
3  19 10  12  192
4  36  3  78  276
5 130 19 211  959
6  84  2  68   89
```

Nótese que algunas de las funciones centrales de *mada* aceptan cuatro vectores de frecuencias (VP, FN, FP, VN) como entrada. Sin embargo, es conveniente para almacenar no sólo las frecuencias observadas, sino también los nombres de estudio en la misma trama de datos. El comando que se muestra a continuación, permite realizar lo antes mencionado.

```
> AuditC6$names <- c("Estudio 1", "Estudio 2", "Estudio 3",
+ "Estudio 4", "Estudio 5", "Estudio 6")
> AuditC6$names

[1] "Estudio 1" "Estudio 2" "Estudio 3" "Estudio 4" "Estudio 5" "Estudio 6"
```

El conjunto completo de datos del ejemplo se encuentra en *mada* con 14 estudios, como se muestra a continuación.

```
> data("AuditC")
> AuditC

VP FN  FP  VN
1  47  9 101 738
2 126 51 272 1543
3  19 10  12  192
4  36  3  78  276
5 130 19 211  959
6  84  2  68   89
7  68  0 112  423
8 752  0 3226 2977
9  59  5  55  136
```

10	142	50	571	2788
11	137	24	107	358
12	57	3	103	437
13	34	1	21	56
14	152	51	88	264

4.1.4. Celdas vacías

En el análisis de las tablas de 2×2 , se puede dar el caso que ciertas celdas tengan ceros, lo cual puede conducir a problemas o artificios estadísticos, ya que no se pueden estimar ciertos ratios. En lugar de que las celdas contengan ceros, se hace uso de la corrección por continuidad. Una sugerencia como corrección por continuidad es utilizar el valor de 0.5. Esta estrategia la utiliza *mada* por defecto.

4.1.5. Estadística descriptiva

Las estadísticas descriptivas para un conjunto de datos incluyen la sensibilidad, especificidad y tasa de falsos positivos de los estudios primarios y sus ratios de verosimilitud positivo y negativo ($LR+$, $LR-$) y sus odds ratios de diagnóstico, Glas et al. (65). Están definidas como, $LR+ = \frac{Q_A}{1 - Q_B} = \frac{\text{sensibilidad}}{\text{tasa de falsos positivos}}$, $LR- = \frac{1 - Q_A}{Q_B}$ y $DOR = \frac{LR+}{LR-} = \frac{VP \cdot VN}{FN \cdot FP}$. Todas estas medidas se calculan fácilmente usando la función “**madad**”, junto con sus intervalos de confianza. La función “**madad**” usa las fórmulas proporcionadas por Deeks (32). *madad* utiliza un test de χ^2 para evaluar la heterogeneidad de las sensibilidades y especificidades. La hipótesis nula en ambos casos es que todos son iguales. Por último, la correlación de las sensibilidades y las tasas de falsos positivos se calcula para dar una pista de si existe un problema con los puntos de corte. Se pueden observar estas medidas con la siguiente sentencia:

```
> madad(AuditC)
```

```
Descriptive summary of AuditC with 14 primary studies.  
Confidence level for all calculations set to 95 %  
Using a continuity correction of 0.5 if applicable
```

4. ANÁLISIS DE RESULTADOS

Diagnostic accuracies

```
sens  2.5% 97.5%  spec  2.5% 97.5%
[1,] 0.833 0.716 0.908 0.879 0.855 0.899
[2,] 0.711 0.640 0.772 0.850 0.833 0.866
[3,] 0.650 0.471 0.795 0.939 0.898 0.964
[4,] 0.912 0.785 0.967 0.779 0.733 0.819
[5,] 0.870 0.807 0.915 0.819 0.796 0.840
[6,] 0.971 0.912 0.991 0.566 0.489 0.641
[7,] 0.993 0.934 0.999 0.790 0.754 0.822
[8,] 0.999 0.994 1.000 0.480 0.468 0.492
[9,] 0.915 0.822 0.962 0.711 0.643 0.770
[10,] 0.738 0.672 0.795 0.830 0.817 0.842
[11,] 0.849 0.786 0.896 0.769 0.729 0.805
[12,] 0.943 0.854 0.979 0.809 0.773 0.840
[13,] 0.958 0.838 0.990 0.724 0.616 0.811
[14,] 0.748 0.684 0.802 0.749 0.702 0.792
```

Test for equality of sensitivities:

X-squared = 272.3603, df = 13, p-value = <2e-16

Test for equality of specificities:

X-squared = 2204.8, df = 13, p-value = <2e-16

Diagnostic OR and likelihood ratios

```
DOR   2.5%    97.5%  posLR  2.5%  97.5%  negLR  2.5%  97.5%
[1,]   36.379 17.587    75.251  6.897  5.556   8.561  0.190  0.106  0.339
[2,]   13.913  9.818    19.717  4.736  4.100   5.470  0.340  0.270  0.429
[3,]   28.600 11.133    73.469 10.660  5.862  19.384  0.373  0.229  0.608
[4,]   36.732 11.925   113.144  4.127  3.320   5.129  0.112  0.041  0.306
[5,]   30.361 18.440    49.986  4.817  4.201   5.523  0.159  0.105  0.240
[6,]   44.162 12.077   161.486  2.240  1.868   2.687  0.051  0.015  0.173
[7,]  515.729 31.687  8393.755  4.730  4.009   5.581  0.009  0.001  0.145
[8,] 1388.854 86.752 22234.711  1.922  1.876   1.968  0.001  0.000  0.022
```

4.1 Precisión Diagnóstica con MADA para Meta-Análisis

```
[9,] 26.607 10.523 67.271 3.167 2.506 4.001 0.119 0.053 0.266
[10,] 13.768 9.865 19.216 4.341 3.879 4.857 0.315 0.249 0.400
[11,] 18.716 11.572 30.271 3.679 3.079 4.397 0.197 0.136 0.284
[12,] 69.444 23.113 208.648 4.927 4.099 5.922 0.071 0.026 0.196
[13,] 60.442 10.948 333.681 3.477 2.411 5.014 0.058 0.012 0.277
[14,] 8.850 5.949 13.165 2.982 2.448 3.632 0.337 0.264 0.430
```

Correlation of sensitivities and false positive rates:

```
rho 2.5 % 97.5 %
0.677 0.228 0.888
```

La función **madad** tiene una gama de opciones con respecto a los detalles de los cálculos. Por ejemplo, se pueden calcular intervalos de confianza al 80%, mediante la sentencia `madad(AuditC, level = 0.80)`.

```
> madad(AuditC, level = 0.80)
```

Descriptive summary of AuditC with 14 primary studies.

Confidence level for all calculations set to 80 %

Using a continuity correction of 0.5 if applicable

Diagnostic accuracies

```
sens 10% 90% spec 10% 90%
[1,] 0.833 0.761 0.887 0.879 0.864 0.893
[2,] 0.711 0.665 0.752 0.850 0.839 0.860
[3,] 0.650 0.533 0.751 0.939 0.914 0.957
[4,] 0.912 0.838 0.955 0.779 0.749 0.806
[5,] 0.870 0.831 0.901 0.819 0.805 0.833
[6,] 0.971 0.938 0.987 0.566 0.516 0.616
[7,] 0.993 0.964 0.999 0.790 0.767 0.812
[8,] 0.999 0.997 1.000 0.480 0.472 0.488
[9,] 0.915 0.860 0.950 0.711 0.667 0.751
[10,] 0.738 0.696 0.777 0.830 0.821 0.838
[11,] 0.849 0.809 0.881 0.769 0.743 0.793
```

4. ANÁLISIS DE RESULTADOS

[12,] 0.943 0.892 0.970 0.809 0.786 0.829

[13,] 0.958 0.892 0.985 0.724 0.655 0.784

[14,] 0.748 0.707 0.784 0.749 0.719 0.778

Test for equality of sensitivities:

X-squared = 272.3603, df = 13, p-value = <2e-16

Test for equality of specificities:

X-squared = 2204.8, df = 13, p-value = <2e-16

Diagnostic OR and likelihood ratios

DOR	10%	90%	posLR	10%	90%	negLR	10%	90%	
[1,]	36.379	22.618	58.513	6.897	5.987	7.944	0.190	0.130	0.277
[2,]	13.913	11.077	17.475	4.736	4.310	5.204	0.340	0.293	0.396
[3,]	28.600	15.433	53.001	10.660	7.210	15.760	0.373	0.271	0.513
[4,]	36.732	17.603	76.651	4.127	3.579	4.757	0.112	0.058	0.216
[5,]	30.361	21.914	42.063	4.817	4.405	5.268	0.159	0.121	0.208
[6,]	44.162	18.918	103.093	2.240	1.989	2.523	0.051	0.023	0.113
[7,]	515.729	83.223	3195.961	4.730	4.245	5.270	0.009	0.002	0.056
[8,]	1388.854	226.547	8514.419	1.922	1.892	1.952	0.001	0.000	0.008
[9,]	26.607	14.508	48.797	3.167	2.718	3.690	0.119	0.070	0.201
[10,]	13.768	11.072	17.121	4.341	4.033	4.672	0.315	0.270	0.368
[11,]	18.716	13.667	25.630	3.679	3.275	4.134	0.197	0.155	0.250
[12,]	69.444	33.825	142.573	4.927	4.369	5.557	0.071	0.036	0.138
[13,]	60.442	19.778	184.715	3.477	2.737	4.417	0.058	0.021	0.161
[14,]	8.850	6.826	11.474	2.982	2.621	3.392	0.337	0.287	0.395

Correlation of sensitivities and false positive rates:

rho 2.5 % 97.5 %

0.677 0.228 0.888

Las salidas de “madad” pueden ser asignadas a un objeto. Por ejemplo, las tasas de falsos positivos, con sus respectivos intervalos de confianza, pueden ser extraídos utilizando el constructo

\$, mediante la sentencia `AuditC.d <- madad(AuditC)`, vea la siguiente salida.

```
> AuditC.d <- madad(AuditC)
> AuditC.d\ $fpr

$ \fpr
[1] 0.12083333 0.15005507 0.06097561 0.22112676 0.18061486 0.43354430
[7] 0.20988806 0.52006770 0.28906250 0.17008929 0.23068670 0.19131238
[13] 0.27564103 0.25070822

$ \fpr.ci
2.5%      97.5%
[1,] 0.10050071 0.1446182
[2,] 0.13436927 0.1672182
[3,] 0.03560834 0.1024938
[4,] 0.18106960 0.2671547
[5,] 0.15963660 0.2036817
[6,] 0.35875648 0.5114869
[7,] 0.17753840 0.2463665
[8,] 0.50762738 0.5324832
[9,] 0.22957510 0.3568250
[10,] 0.15776394 0.1831681
[11,] 0.19473290 0.2710443
[12,] 0.16038974 0.2265879
[13,] 0.18879306 0.3835508
[14,] 0.20834216 0.2984416
```

4.1.6. Gráficos descriptivos

Para los datos de AUDIT-C, la prueba χ^2 , sugiere la heterogeneidad entre la sensibilidad y la especificidad. El correspondiente forest plot, note Figura 4.1, confirma esta afirmación, mediante las sentencias:

4. ANÁLISIS DE RESULTADOS

```
> forest(madad(AuditC), type = "sens")
> forest(madad(AuditC), type = "spec")
```

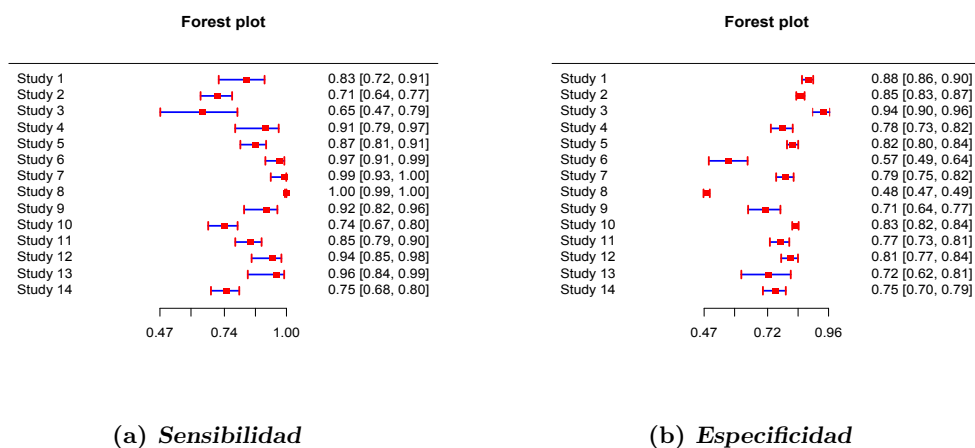


Figura 4.1: Gráficos forest-plot, para sensibilidad y especificidad, datos AuditC

Además de los gráficos anteriores, mada provee una variedad de gráficos para estudiar los datos en el espacio ROC. Hay que tener en cuenta que para fines de exploración es a menudo útil emplear el color y otros detalles con los que R cuenta para dar a los gráficos. Existen dos grandes niveles de gráficos en mada: “crosshair” que producen gráficos “crosshair”, Phillips et al. (144), y ROCellipse. El gráfico crosshair puede producir colores a los crosshair y hace que los crosshairs sean más amplios mientras se incrementa el tamaño de la muestra, el gráfico se genera con las sentencias `rs <- rowSums(AuditC)`, `weights <- 4 * rs / max(rs)`, `crosshair(AuditC, xlim = c(0,0.6), ylim = c(0.4,1), col = 1:14, lwd = weights)`, vea Figura 4.2.

```
> rs <- rowSums(AuditC)
> pesos <- 4 * rs / max(rs)
> crosshair(AuditC, xlim = c(0,0.6), ylim = c(0.4,1),
+ col = 1:14, lwd = pesos)
```

Asimismo, muestra un gráfico ROCellipse, el cual detalla una región de confianza que describe la incertidumbre de la pareja sensibilidad y especificidad. Estas regiones son elipses sobre el espacio logit ROC, y son generadas con las sentencias, `ROCellipse(AuditC, pch =)`, `points(fpr(AuditC), sens(AuditC))`, vea Figura 4.2.


```
> ROCellipse(AuditC, pch = "")  
> points(fpr(AuditC), sens(AuditC))
```

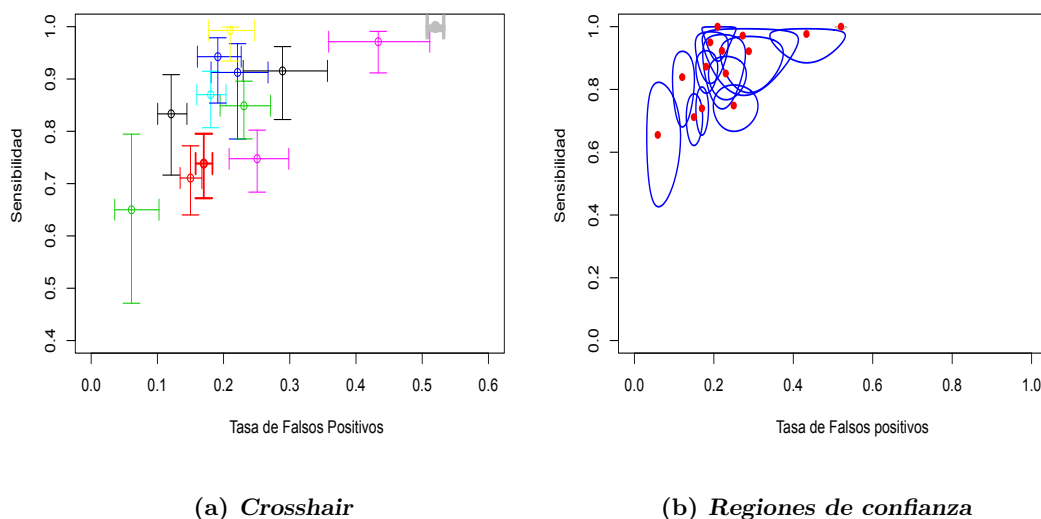


Figura 4.2: Gráfico ponderado crosshair con colores arbitrarios y regiones de confianza en forma de ellipse, datos AuditC

El gráfico crosshair es una combinación de un gráfico ROC y un forest-plot que demuestra la relación bivalente y el grado de heterogeneidad entre la sensibilidad y la tasa de falsos positivos.

4.1.7. Enfoque univariante

Antes de la llegada del enfoque bivalente, por Rutter and Gatsonis (158) y Reitsma et al. (148), algunos enfoques univariantes eran muy populares. El enfoque bivalente no es recomendable cuando el tamaño de la muestra es muy pequeño. El modelo bivalente de Reitsma et al. (148), tiene 5 parámetros, que serían claramente demasiados para un número pequeño de estudios.

4.1.8. Odds Ratio de Diagnóstico (DOR)

El enfoque DOR, desarrollado por Glas et al. (65) es estimado por el método de Mantel-Haenszel (MH) véase por ejemplo Deeks (32). Mada proporciona los resultados correspondientes. El modelo de efectos fijos subyacente tiene la forma, $DOR_i = \mu + \varepsilon_i$, donde μ es la media de los DOR's y ε_i , son errores independientes con varianza específica. El estimador MH es una media ponderada de los DOR's observados de los estudios primarios y es robusto para celdas con valores nulos. Se estima como $\hat{\mu} = \sum_i \frac{w_i^{MH} DOR_i}{\sum_i w_i^{MH}}$, donde $w_i^{MH} = \frac{y_{i01}(y_{i10})}{n_{i1} + n_{i0}}$ son los pesos de Mantel-Haenszel. Un estimador para un modelo de efectos aleatorios sigue el enfoque de DerSimonian and Laird (38). El modelo se formula en términos de los logaritmos de los DOR's, luego se tiene que $\log(DOR_i) = \mu + \varepsilon_i + \delta_i$, donde μ es la media de los $\log(DOR_i)$ y $\varepsilon_i + \delta_i$, son independientes con media cero, y la varianza ε_i , denotada por σ_i^2 , se estima como $\hat{\sigma}_i^2 = \frac{1}{y_{i11}} + \frac{1}{y_{i10}} + \frac{1}{y_{i01}} + \frac{1}{y_{i00}}$, y la varianza de δ_i , denotada por τ^2 , cuyo valor tiene que ser estimado.

El "DSL" es un estimador ponderado, por lo que se tiene que $\hat{\mu} = \sum_i \frac{w_i^{DSL} DOR_i}{\sum_i w_i^{DSL}}$, donde $w_i^{DSL} = \frac{1}{\hat{\sigma}_i^2 + \tau^2}$ y τ^2 es estimado mediante el estadístico Q de Cochran. La función "madauni" puede ayudar a estimar τ^2 , mediante la sentencia `fit.DOR.DSL <- madauni(AuditC)`.

```
> (ajuste.DOR.DSL <- madauni(AuditC))
```

```
Call:
```

```
madauni(x = AuditC)
```

```
DOR tau^2
```

```
26.337 0.311
```

Asimismo, se puede obtener una estimación del DOR mediante el método MH con la sentencia,

```
ajuste.DOR.MH <- madauni(AuditC, method = "MH")
```

```
> (ajuste.DOR.MH <- madauni(AuditC, method = "MH"))
```

```
Call:
```

```
madauni(x = AuditC, method = "MH")
```

```
DOR
17.93335
```

La función `madauni` no es muy informativa, así que empleando la sentencia `summary(ajuste.DOR.DSL)`, se puede obtener más detalle para el enfoque DSL, como se muestra a continuación.

```
> (summary(ajuste.DOR.DSL))

Call:
madauni(x = AuditC)

Estimates:
DSL estimate  2.5 % 97.5 %
DOR           26.337 17.971 38.596
lnDOR         3.271  2.889  3.653
tau^2         0.311  0.000  3.787
tau           0.557  0.000  1.946

Cochran's Q: 19.683 (13 df, p = 0.103)
Higgins' I^2: 33.955%
```

Además de notar los intervalos de confianza, se observa también el estadístico Q Cochran (26) y el índice I^2 Higgins et al. (84). Elaborar un gráfico de forest-plot de los logaritmos de los DOR's es muy sencillo, se utiliza el método forest de la clase `madauni` mediante la sentencia `forest(ajuste.DOR.DSL)`, como se muestra en la Figura 4.3.

```
> forest(ajuste.DOR.DSL)
```

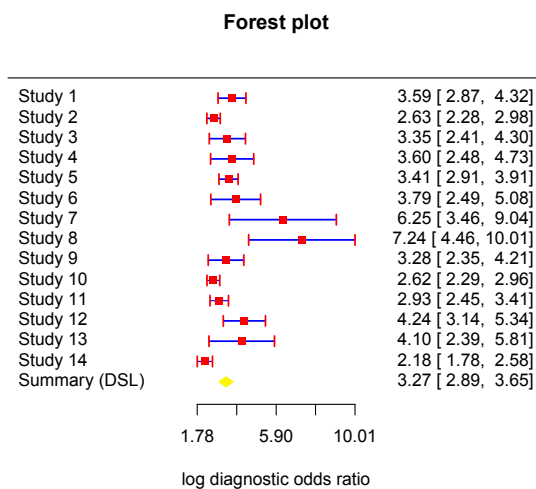


Figura 4.3: Forest-plot, logaritmos DOR's, ajuste DSL, datos AuditC

4.1.9. Modelo Proporcional Hazards

El modelo proporcional hazards (PHM; vea Holling et al. (87)), se construye sobre la asunción de una curva ROC simple y se asumen las condiciones del modelo de Lehmann, Le (105). Sea θ_{Ai} y $1 - \theta_{Bi}$, los valores de la sensibilidad y la tasa de falsos positivos del i -ésimo estudio respectivamente. Entonces la relación entre θ_{Ai} y $1 - \theta_{Bi}$ es: $\theta_{Ai} = (1 - \theta_{Bi})^{\theta_H}$, donde θ_H es el parámetro de la precisión del diagnóstico. Cuanto más pequeño sea el valor de θ , mayor será el área bajo la curva ROC, lo cual indica que el test es más preciso en el diagnóstico. Para estimar θ , el estimador *APMLE* es implementado en mada, para el caso de homogeneidad (efectos fijos), se realiza con el comando `fit.phm.homo <- phm(AuditC, hetero = FALSE)`.

```
> (ajuste.phm.homo <- phm(AuditC, hetero = FALSE))
```

Call:

```
phm.default(data = AuditC, hetero = FALSE)
```

Coefficients:

```
theta
```

```
0.004586893
```

Para el caso de heterogeneidad, se realiza con el comando `ajuste.phm.het <- phm(AuditC)`.

```
> (ajuste.phm.het <- phm(AuditC))
```

Call:

```
phm.default(data = AuditC)
```

Coefficients:

```
theta      taus_sq  
0.084631351 0.003706143
```

La salida de la función **phm**, donde el método “summary” es más informativo, se realiza mediante la sentencia `summary(ajuste.phm.homo)`.

```
> summary(ajuste.phm.homo)
```

Call:

```
phm.default(data = AuditC, hetero = FALSE)
```

```
Estimate      2.5 %      97.5 %  
theta 0.004586893 0.003508507 0.00566528
```

Log-likelihood: -61.499 on 1 degrees of freedom

AIC: 125

BIC: 125.6

Chi-square goodness of fit test (Adjusted Profile Maximum Likelihood under homogeneity)

data: x

Chi-square = 222.47, df = 1, p-value < 2.2e-16

```
AUC 2.5 % 97.5 % pAUC 2.5 % 97.5 %  
0.995 0.997 0.994 0.994 0.995 0.992
```

4. ANÁLISIS DE RESULTADOS

La prueba de bondad de ajuste χ^2 rechaza la homogeneidad. Es decir, la hipótesis de heterogeneidad se acepta, y se comprueba con la sentencia `summary(ajuste.phm.het)`.

```
> summary(ajuste.phm.het)

Call:
phm.default(data = AuditC, hetero = TRUE)

Estimate      2.5 %      97.5 %
theta  0.084631351  0.047449859  0.121812844
taus_sq 0.003706143 -0.001277798  0.008690085

Log-likelihood: 31.121 on 2 degrees of freedom
AIC: -58.2
BIC: -57

Chi-square goodness of fit test (Adjusted Profile Maximum Likelihood
under heterogeneity)

data: x
Chi-square = 13.726, df = 2, p-value = 0.3185

AUC  2.5 % 97.5 %  pAUC  2.5 % 97.5 %
0.922 0.955 0.891 0.891 0.937 0.848
```

La estimación de θ_H , se obtiene a partir de la curva SROC. Graficando la curva en conjunto con las bandas de confianza obtenidas de los intervalos de confianza de θ_H a partir de `summary` mediante las sentencias `plot(fit.phm.het, xlim = c(0,0.6), ylim = c(0.4,1))`, `ROCellipse(AuditC, add = TRUE)`.

```
> plot(fit.phm.het, xlim = c(0,0.6), ylim = c(0.4,1))
> ROCellipse(AuditC, add = TRUE)
```

Nótese, que la curva SROC, no se extrapola más allá del rango de los datos originales. El

área bajo la curva (AUC), forma parte de la salida del summary. Para PHM, es calculada por $AUC = \frac{1}{\theta_H + 1}$. note Figura 4.4.

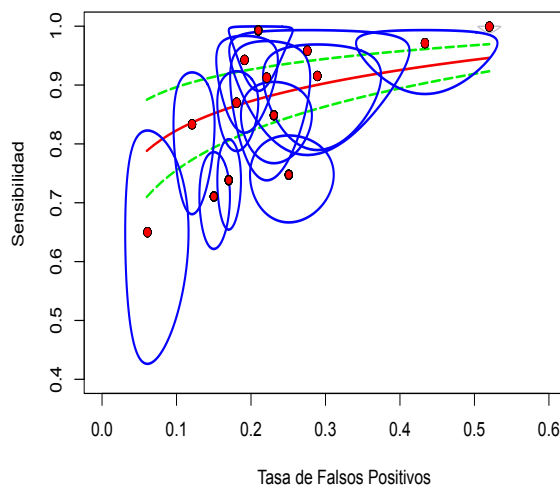


Figura 4.4: Resumen gráfico con modelo PHM, datos AuditC

4.1.10. Enfoque Bivariante

Típicamente, la sensibilidad y la especificidad de una prueba diagnóstica dependen de los puntos de corte que se elijen en cada estudio, es decir, al incrementar la sensibilidad, la especificidad disminuye. En un entorno meta-analítico, la relación entre estas medidas es negativa. Es decir, existe una correlación positiva entre la sensibilidad y la tasa de falsos positivos. Dado que estas cantidades están relacionadas entre sí, el enfoque bivariante para meta-análisis en la precisión del diagnóstico ha tenido acogida, Arends et al. (4), Harbord et al. (76), Reitsma et al. (148), Rutter and Gatsonis (158), Van Houwelingen et al. (187).

Se asume que la sensibilidad y la tasa de falsos positivos siguen un modelo binomial condicionado y las transformaciones logit de la sensibilidad y la tasa de los falsos positivos un modelo normal bivalente. Existen dos maneras de llegar al modelo final: un modelo no lineal mixto o un modelo lineal mixto, Arends et al. (4). El último enfoque se encuentra en *mada* como

4. ANÁLISIS DE RESULTADOS

la función **reitsma**. Se puede entender lo anteriormente expresado como una meta-regresión multivariante y el paquete **mvmeta** sirve para este enfoque, Gasparrini et al. (61).

Sea θ_{Ai} y $1 - \theta_{Bi}$, la sensibilidad y la tasa de falsos positivos del i -ésimo estudio respectivamente, donde θ_{Ai} y $1 - \theta_{Bi}$ denotan sus estimaciones a partir de las frecuencias observadas. Entonces, dado un modelo binomial que se asume condicional sobre θ_{Ai} , la varianza de $\text{logit}(\theta_{Ai})$ puede ser aproximada como $\frac{\hat{\theta}_{Ai}(1 - \hat{\theta}_{Ai})}{n_{i1}}$ y la varianza de $\text{logit}(1 - \theta_{Bi})$ como $\frac{(1 - \hat{\theta}_{Bi})(\hat{\theta}_{Bi})}{n_{i0}}$. Para los niveles dentro de los estudios, se asume condicional sobre θ_{Ai} y $1 - \theta_{Bi}$. La variación observada puede ser descrita por estas varianzas y un modelo normal.

Sea D_i una matriz de 2×2 con dos varianzas en su diagonal principal. Sobre los niveles de estudio se supone una media global $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}^T$ y la matriz de covarianzas Σ , vea ecuación 4.4, la cual describe la heterogeneidad del par ordenado $(\text{logit}(\hat{p}_i), \text{logit}(\hat{u}_i))$. Luego el modelo del i -ésimo estudio es $(\text{logit}(\hat{p}_i), \text{logit}(\hat{u}_i))^T \sim N(\mu, \Sigma + D_i)$, y el modelo en meta es similar a las otras funciones. Las salidas se las obtiene mediante la sentencia `fit.reitsma <- reitsma(AuditC)`.

```
> (ajuste.reitsma <- reitsma(AuditC))

Call: reitsma.default(data = AuditC)

Fixed-effects coefficients:
tsens      tfpr
(Intercept) 2.0997 -1.2637

14 studies, 2 fixed and 3 random-effects parameters
logLik      AIC      BIC
31.5640 -53.1279 -46.4669
```

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma \\ \sigma & \sigma_2^2 \end{pmatrix} \quad (4.4)$$

Las salidas del objeto *reitsma* son “escasas”, pero se puede obtener más información con *summary*, `summary(fit.reitsma)`.


```

> summary(ajuste.reitsma)

Call: reitsma.default(data = AuditC)

Bivariate diagnostic random-effects meta-analysis
Estimation method: REML

Fixed-effects coefficients
Estimate Std. Error      z Pr(>|z|) 95%ci.lb 95%ci.ub
tsens.(Intercept)  2.100      0.338 6.215  0.000  1.438  2.762 ***
tfpr.(Intercept)  -1.264     0.174 -7.249  0.000 -1.605 -0.922 ***
sensitivity        0.891        -      -      -      0.808  0.941
false pos. rate    0.220        -      -      -      0.167  0.285
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Variance components: between-studies Std. Dev and correlation matrix
Std. Dev tsens  tfpr
tsens    1.175 1.000  .
tfpr     0.638 0.854 1.000

logLik    AIC    BIC
31.564 -53.128 -46.467

AUC: 0.887
Partial AUC (restricted to observed FPRs and normalized): 0.861

HSROC parameters
Theta      Lambda      beta sigma2theta sigma2alpha
-0.083     3.262     -0.610     0.695     0.218

```

Nótese que la sensibilidad y la tasa de falsos positivos son las transformaciones de μ_1 y μ_2 . Luego, se puede graficar la curva SROC de este modelo. Por defecto, la representación del par ordenado de la sensibilidad y la tasa de falsos positivos se representan con una región de

4. ANÁLISIS DE RESULTADOS

confianza. Es posible graficar la curva SROC un poco más “gruesa” con el argumento `sroclwd`, así como agregar un título, mediante las sentencias:

```
> plot(ajuste.reitsma, sroclwd = 2,  
+ main = "SROC curve (bivariate model) for AUDIT-C data")  
> points(fpr(AuditC), sens(AuditC), pch = 2)  
> legend("bottomright", c("data", "summary estimate"), pch = c(2,1))  
> legend("bottomleft", c("SROC", "conf. region"), lwd = c(2,1))
```

Por defecto, la curva SROC no extrapola más allá del rango de los datos originales, observe Figura 4.5.

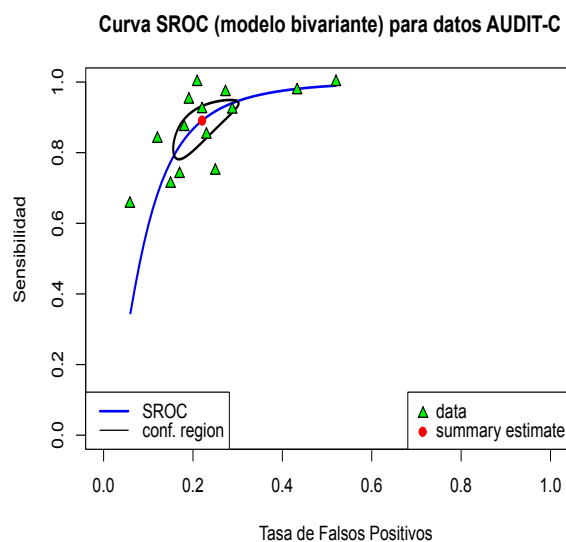


Figura 4.5: Curva SROC, modelo de Reitsma, datos AuditC

4.1.11. Comparando curvas SROC

En este apartado se mostrará como comparar curvas SROC. Patrick et al. (141), condujeron un meta-análisis para averiguar la eficacia de los cuestionarios, los cuales fueron cumplimentados por el entrevistado (SAQ) y por el entrevistador (IAQ) para detectar el consumo de nicotina. El conjunto de datos está constituido por los subconjuntos de datos IAQ y SAQ. Primero los

subconjuntos de datos se ajustan a un modelo bivalente mediante las sentencias `data(IAQ)`, `data(SAQ)`, `fit.IAQ <- reitsma(IAQ)`, `fit.SAQ <- reitsma(SAQ)`.

```
> data("IAQ")
> data("SAQ")
> ajuste.IAQ <- reitsma(IAQ)
> ajuste.SAQ <- reitsma(SAQ)
```

Luego se grafica la curva SROC de los ajustes, comenzando con el ajuste de IAQ y agregando después al gráfico el ajuste de SAQ, mediante las sentencias:

```
> plot(ajuste.IAQ, xlim = c(0,.5), ylim = c(.5,1),
+ main = "Comparison of IAQ and SAQ")
> lines(sroc(ajuste.SAQ), lty = 2)
> ROCellipse(ajuste.SAQ, lty = 2, pch = 2, add = TRUE)
> points(fpr(IAQ), sens(IAQ), cex = .5)
> points(fpr(SAQ), sens(SAQ), pch = 2, cex = 0.5)
> legend("bottomright", c("IAQ", "SAQ"), pch = 1:2, lty = 1:2)
```

Nótese que el argumento `lty` es utilizado para distinguir las curvas de ambos ajustes, vea Figura 4.6.

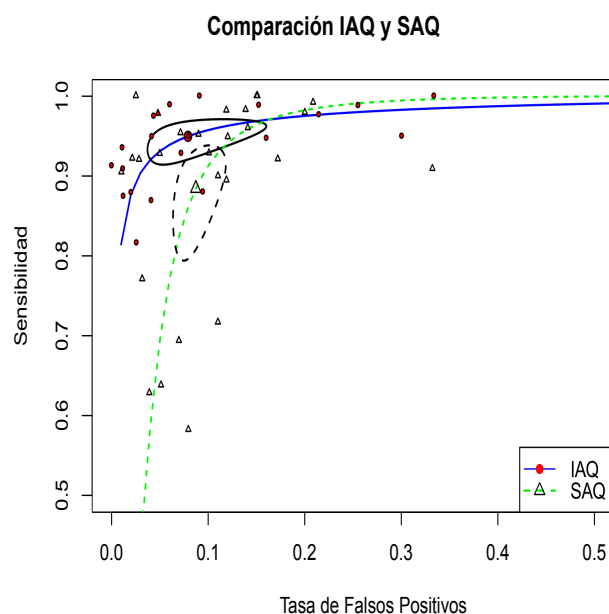


Figura 4.6: Comparación de curvas SROC para SAQ (entrevistados) y IAQ (entrevistadores), datos IAQ y SAQ

4.1.12. Meta-Regresión bivalente

Usaremos el conjunto de datos de Patrick et al. (141), para demostrar el enfoque bivalente de meta-regresión. Los datos se cargan mediante la sentencia `data("smoking")`. El conjunto de datos contiene las mismas variables que el subconjunto de datos IAQ y SAQ, pero con tipos de variables diferentes, y se puede verificar mediante la sentencia `summary(smoking$type)`.

```
> data("smoking")
> summary(smoking$type)
```

```
IAQ SAQ
20  31
```

Se utiliza “*type*” como factor de covariable en la meta-regresión de diagnóstico. Nótese, que el lado izquierdo del objeto tiene que tener la forma `cbind(tsens, tfpr)`, donde `tsens` y `tfpr` representan las transformaciones de la sensibilidad y la tasa de falsos positivos respectivamente, generando la salida mediante las sentencias:

```

> ajuste.smoking.type <- reitsma(smoking,
+ formula = cbind(tsens, tfpr) ~ type)
> summary(ajuste.smoking.type)

Call: reitsma.default(data = smoking, formula = cbind(tsens, tfpr) ~
type)

Bivariate diagnostic random-effects meta-analysis
Estimation method: REML

Fixed-effects coefficients
Estimate Std. Error      z Pr(>|z|) 95%ci.lb 95%ci.ub
tsens.(Intercept)   3.007      0.307  9.809   0.000   2.406   3.608 ***
tsens.typeSAQ      -0.999      0.395 -2.531   0.011  -1.772  -0.225  *
tfpr.(Intercept)  -2.407      0.243 -9.912   0.000  -2.883  -1.931 ***
tfpr.typeSAQ       0.019      0.307  0.061   0.952  -0.583   0.621
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Variance components: between-studies Std. Dev and correlation matrix
Std. Dev tsens  tfpr
tsens    1.286 1.000   .
tfpr     1.003 0.510 1.000

logLik      AIC      BIC
126.622 -239.244 -220.869

```

La salida anterior puede ser interpretada de la siguiente manera: El valor de z para los coeficientes de regresión asociado a la sensibilidad es significativa, indicando que la realización de una entrevista ofrece una mejor sensibilidad. Es interesante notar que las tasas de falsos positivos no indican ningún efecto. Una vez que se utiliza meta-regresión, no se pueden trazar curvas SROC razonables, ya que se necesitan valores para la covarianzas. Luego, los valores de AUC no tienen sentido.

4. ANÁLISIS DE RESULTADOS

Se puede comparar el ajuste de dos meta-regresiones bivariantes con la prueba del ratio de máxima verosimilitud. Para esto, se necesita que los modelos se ajusten por el método de máxima verosimilitud, y como la prueba del ratio de máxima verosimilitud emplea la teoría asintótica, la cual solo es válida si se emplea este método de estimación. Lo anteriormente mencionado se realiza mediante las sentencias:

```
> ajuste.smoking.ml.type <- reitsma(smoking,  
+ formula = cbind(tsens, tfpr) ~ type, method = "ml")  
> ajuste.smoking.ml.intercept <- reitsma(smoking,  
+ formula = cbind(tsens, tfpr) ~ 1,  
+ method = "ml")  
> anova(ajuste.smoking.ml.type, ajuste.smoking.ml.intercept)
```

Likelihood-ratio test

Model 1: cbind(tsens, tfpr) ~ type

Model 2: cbind(tsens, tfpr) ~ 1

ChiSquared Df Pr(>ChiSquared)

7.858 2 0.0197 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

La meta-regresión confirma que el “type” explica algo de heterogeneidad entre los estudios primarios.

4.2. Precisión diagnóstica con Stata para Meta-Análisis

4.2.1. Descripción

Midas es un programa exhaustivo de rutinas estadísticas y gráficas para entender el meta-análisis de las pruebas de diagnóstico en Stata, el cual es un paquete de software estadístico creado en 1985 por StataCorp. El índice y la prueba de referencia (prueba oro) son dicotómicas. La síntesis de los datos primarios es desarrollada dentro de una regresión bivalente de efectos mixtos, centrada en hacer inferencia sobre la media de la sensibilidad y especificidad. El enfoque bivalente fue desarrollado para ensayos de tratamiento en meta-análisis, Van Houwelingen et al. (187, 188) y modificado para resumir los datos de una prueba de diagnóstico usando una aproximación normal dentro de los estudios, Arends et al. (4), Reitsma et al. (148), Riley et al. (149), Riley et al. (150), Riley et al. (151). Una interpretación binomial del modelo bivalente, se obtiene a partir de, Arends et al. (4), Chu and Cole (23), Riley et al. (150), donde se asume que distribuciones binomiales son independientes para los verdaderos positivos y verdaderos negativos condicionadas a la sensibilidad y especificidad en cada estudio.

La estimación basada en máxima verosimilitud para el enfoque binomial es desarrollada adoptando una cuadratura gaussiana, usando el comando `xtnlogit` o `gllamm`, Rabe-Hesketh et al. (146, 147). Midas permite analizar la heterogeneidad (relacionado con los umbrales y covarianza) y los sesgos de publicación o precisión.

En general, bajo el paquete estadístico Stata se trabaja mediante lenguaje de programación, y conviene recordar que es **case-sensitive**, es decir, diferencia entre letras mayúsculas y minúsculas. En relación a la capacidad del manejo de grandes volúmenes de datos, el paquete Stata requiere hacer una copia a la base de datos con la que se va a trabajar en la memoria RAM del computador donde se trabajará, por lo tanto, la memoria disponible del computador de trabajo debe estar acorde con el tamaño de las bases de datos a utilizar.

El ambiente de trabajo se subdivide en cuatro ventanas, review, variables, command y result. En la ventana básica **review** se detalla el listado de los comandos ejecutados desde que se inició

4. ANÁLISIS DE RESULTADOS

la aplicación; en la ventana **variables** aparecen las variables contenidas en la base de datos cargada; en la ventana **command** se escriben las instrucciones que se le da a la aplicación y por último en la ventana **result**, aparecen los resultados de las órdenes ejecutadas, note Figura 4.7. En la esquina inferior izquierda de la pantalla, Figura 4.8, se tiene el directorio de trabajo. Para cambiar de directorio se puede realizar de dos maneras: escribiendo en la sub-ventana command el comando **cd** seguido de la nueva ruta o directorio, por ejemplo **cd c;** o vía ventana ingresando a la opción **Change Working Directory** ubicada en la opción File de la barra menú.

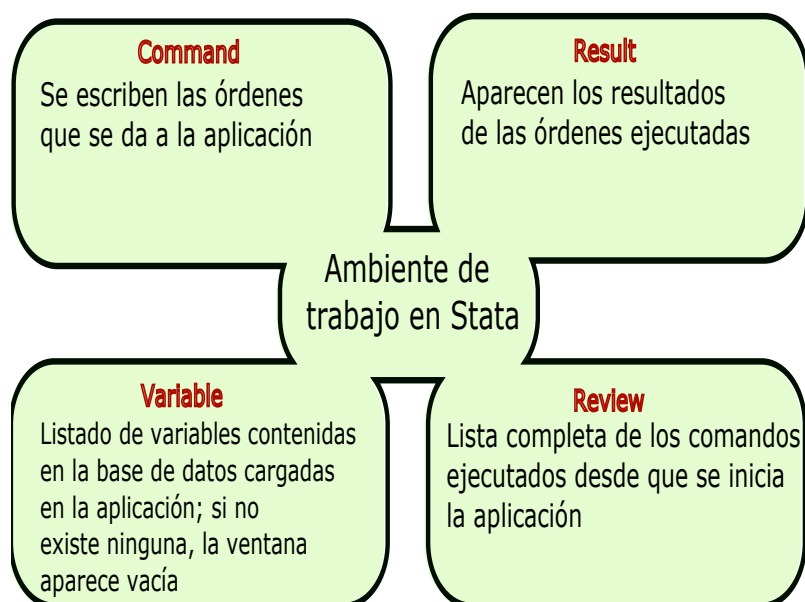


Figura 4.7: Ambiente de trabajo en Stata

En la Sección 4.2.8 se trabajará con el archivo de datos *midas_example.dta*, el cual se obtuvo como parte de una revisión sistemática de un meta-análisis, el cual analizaba:

- El rendimiento de tomografías por emisión de positrones axilar (FDG-PET) en el cáncer de mama.
- La identificación de la calidad de los estudios primarios.
- La cuantificación de los resultados globales de clasificación (sensibilidad y especificidad), el poder discriminatorio (odds ratio de diagnóstico) y el valor informativo (ratios de máxima verosimilitud de diagnóstico).

- La evaluación de las características técnicas de la prueba y sesgo de publicación en las estimaciones de precisión diagnóstica.

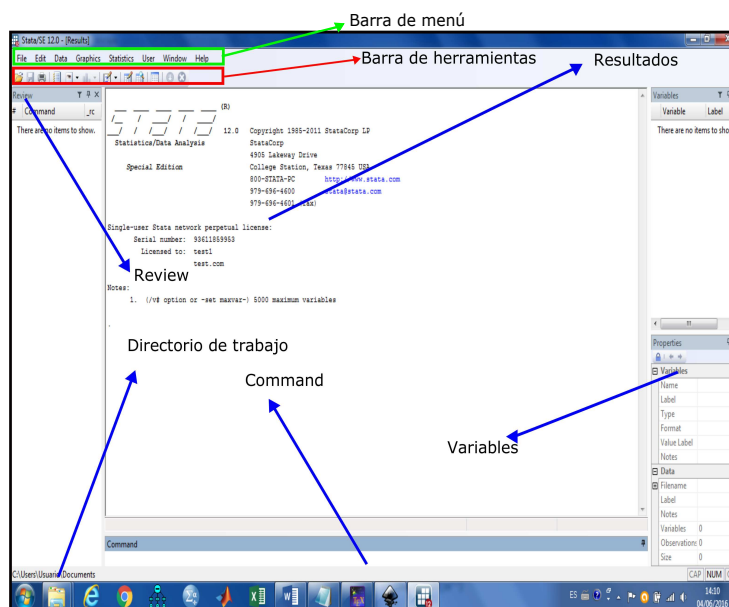


Figura 4.8: Menú y paneles en Ambiente de trabajo Stata

4.2.2. Sintaxis de los comandos de Stata

Stata cuenta con un lenguaje de programación avanzado que respeta normas de sintaxis. Cualquier sentencia debe contener la sintaxis “[**prefix :**] **command** [**varlist**] [**=exp**] [**if**] [**in**] [**weight**] [**using filename**] [, **options**]”. Pero la mayoría de las veces se utiliza de la siguiente forma, mucho más simplificada, “[**command** [**varlist**] [**if**] [**in**] [**weight**] [, **options**]”, donde todo lo que aparece entre corchetes “[...]” es opcional, “**if**” seguida de una expresión lógica que indica que sólo los datos que verifiquen dicha condición serán incluidos en el análisis, **in** sirve para indicar el rango de observaciones que deseamos analizar, **weight** sirve para indicar una variable de ponderación y **options** son las opciones específicas del comando que estemos utilizando. Para el caso de las pruebas de diagnóstico se tendría lo siguiente “[**midas varlist** [**min=4 max=4**] [**if**] [**in**] [**weight**] [, **options**]”, donde en **varlist** se almacenarán los datos de la tabla de contingencia de 2×2 de los índices y resultados de la prueba de referencia.

4. ANÁLISIS DE RESULTADOS

El usuario provee los datos de un arreglo rectangular para los elementos a , b , c y d . Cada fila de datos contiene una tabla de datos 2×2 , es decir, los datos de cada estudio. **varlist** debe contener los valores de las variables a , b , c y d en el orden de la Tabla 4.2.

Prueba de referencia			
	Positivo		Negativo
Test	Positivo	a=verdaderos positivos	c=falsos negativos
	Negativo	c=falsos positivos	d=verdaderos negativos

Tabla 4.2: Orden de las variables a,b,c y d

4.2.3. Gestiones de base de datos

Los archivos de datos en *Stata* se denominan **dataset**. Un dataset es una tabla, donde las columnas representan variables y las filas observaciones o casos. Para conocer la naturaleza de los datos contenidos en el dataset se utiliza la instrucción `description`.

4.2.4. Tipos de datos

Al utilizar el comando **describe**, aparece una variable denominada **storage type** que indica el formato de almacenamiento, es decir, el número de bytes, es decir, la precisión de la variable, Tabla 4.3.

Tipos	Menor valor	Mayor valor	Valor cercano a 0	Bytes	Precisión
Byte	-127	100	± 11	1	
Int	-32,767	32,767	± 1	2	
Long	-2,147,483,647	2,147,483,647	± 1	4	
Float	$-1.70141173319 \times 10^{38}$	$1.70141173319 \times 10^{38}$	$\pm - 10^{-36}$	4	3.795×10^{-8}
Double	$-8.9884656743 \times 10^{307}$	$8.9884656743 \times 10^{307}$	$\pm - 10^{-323}$	8	1.414×10^{-16}

Tabla 4.3: Tipos de datos en Stata

4.2.5. Cargar un archivo de datos en formato Stata

Para cargar un archivo de datos en formato Stata (extensión *.dta), se tienen dos opciones: realizarlo desde la barra de menú o mediante una línea de código en la ventana command. Desde la barra de menú vea Figura 4.9, y con instrucción en la ventana command mediante la sentencia `use nombre del archivo`.

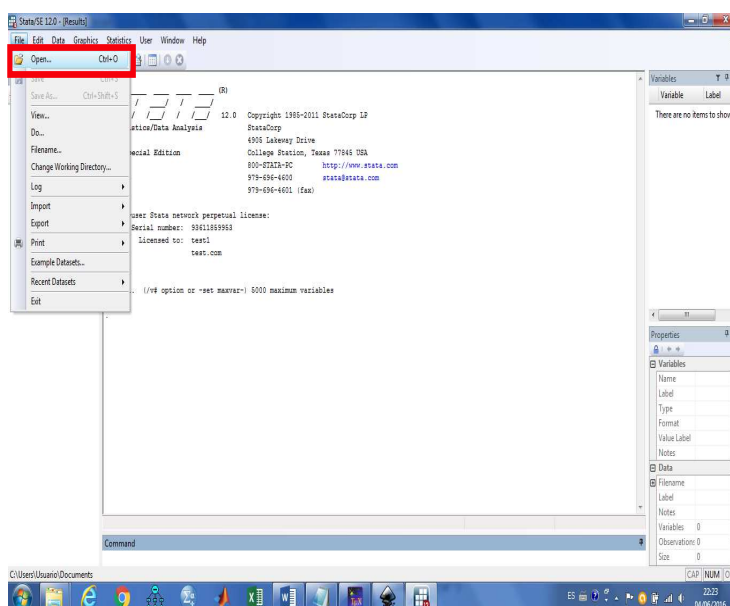


Figura 4.9: Cargar archivo desde barra de Menú en Stata

Si existe un archivo de datos en memoria, la aplicación no permitirá cargarlo en memoria pues borraría el anterior. Por ello, debería de utilizarse la opción clear para liberar de la memoria el anterior archivo de datos `use "nombre del archivo", clear`. La sintaxis de este comando es bastante flexible, permitiendo controlar qué parte del archivo se desea cargar `use [varlist] [if] [in] using filename [, clear nolabel]`.

4.2.6. Opciones en Stata

Modelamiento y opciones de Post-Estimación

- **nip** (número entero) especifica el número de puntos de integración utilizados para la estimación de máxima verosimilitud basada en una cuadratura gaussiana. Valores altos mejoran la precisión a expensas del tiempo de ejecución. Usando `xtmelogit` con `nip(1)`, el modelo se estima mediante una aproximación Laplaciana. Esto disminuye el tiempo computacional, dando resultados confiables en las estimaciones para modelos de efectos fijos. Pero puede producir estimaciones sesgadas de las componentes de la varianza.
- **ebpred(for|roc)** genera un forest-plot y una curva ROC empírica Bayesiana frente a las estimaciones observadas de sensibilidad y especificidad.
- **modchk(gof|bvn|inf|out|all)** provee una forma gráfica para verificar capacidades; **modchk(gof)** despliega gráficos de cuartiles de la prueba de bondad de ajuste de los residuos; **modchk(bvn)** muestra un gráfico de probabilidades de la chi-cuadrado de las distancias al cuadrado de Mahalanobis para evaluar los supuestos de la normal bivalente; **modchk(inf)** realiza un spikeplot para evaluar observaciones particularmente influyentes mediante la distancia de Cook; **modchk(out)** muestra un gráfico de dispersión para la comprobación de los valores atípicos utilizando efectos aleatorios de los valores pronosticados tipificados y **modchk(all)** provee un gráfico que integra los cuatro gráficos anteriores.

Opciones de evaluación de la calidad

- **qtab(varlis)** crea una tabla que muestra las frecuencias de la calidad metodológica de los ítems.
- **qbar(varlis)** calcula las puntuaciones de la calidad, específica para cada estudio y traza un gráfico de barras de la calidad metodológica. `qlab` se puede combinar con `qtab` o `qbar`, para usar etiquetas en las variables para tablas o gráficos de barras de los ítems metodológicos.

Gráficos exploratorios

- **bivbox** implementa una generalización bivalente de un gráfico de caja para datos univariantes similar a un gráfico de caja bivalente, Rousseeuw et al. (154). Es empleado para detectar asimetría y valores atípicos.
- **chiplot** crea un gráfico chiplot, Fisher and Switzer (56) para juzgar si los índices de rendimiento emparejados son independientes.

Opciones del reporte principal

- **results(all)** provee un resumen estadístico del desempeño de los índices, de la variación entre estudios, la prueba estadística del ratio de máxima verosimilitud y otros test globales de homogeneidad.
- **results(het)** provee un resumen estadístico, de la variación entre estudios, la prueba estadística del ratio de máxima verosimilitud y otros test globales de homogeneidad.
- **results(sum)** provee un resumen estadístico para el desempeño de los índices, es decir, sensibilidad/especificidad, ratios de máxima verosimilitud positivo/negativo, y resultados de diagnósticos/odds ratios.
- **table(dss|dlor|dlr)** crea una tabla de las estimaciones del desempeño de los estudios específicos con resúmenes estadísticos de las estimaciones de las medidas y resultados de homogeneidad (chi-cuadrado), y el test de inconsistencia I^2 . **dss|dlor|dlr** representan las medidas de desempeño de las parejas de sensibilidad/especificidad, ratios de máxima verosimilitud positivo/negativo, y resultados de diagnósticos/odds ratio.

Opciones de Forest Plot

- **id(varlis)** provee etiquetas para estudios que permiten hasta 4 variables.
- **bforest(dss|dlr|dlor)** crea un resumen gráfico con estudios específicos (box) y overall (diamante) con sus respectivos intervalos de confianza para cada par de índices usando `graph combine`.

4. ANÁLISIS DE RESULTADOS

- **uforest(dss|dlr|dlor)** crean resúmenes de gráficos univariantes donde los puntos estimados y los intervalos de confianza pueden extenderse. Fordata añade estimaciones de los desempeños de los estudios específicos y sus intervalos de confianza al 95 %.
- **forstats** añade estadísticos de heterogeneidad siguiendo estimaciones puntuales.

Opciones de la curva ROC

- **rocplane** grafica los datos observados en un espacio ROC para la evaluación visual del efecto umbral, una fuente de heterogeneidad única para meta-análisis de diagnóstico. Cuanto mayor sea el valor de corte, mayor será la especificidad y menor la sensibilidad. Esta interdependencia entre estas medidas basada en la variabilidad del umbral puede ser probada a priori utilizando una prueba de correlación de rango como el rho de Spearman. En **midas**, la proporción de variación debido a los efectos de umbral se calcula como el coeficiente de correlación al cuadrado a partir del parámetro de covarianza entre los estudios.
- **sroc(none|pred|conf|both)** grafica los datos observados en forma de puntos, resumen de la sensibilidad y especificidad y curva SROC con o sin las regiones de confianza o predicción.
- **sroc(nnoc|pnoc|cnoc|bnoc)** grafica los datos observados en forma de puntos, resumen de la sensibilidad y especificidad con o sin los contornos de confianza y predicción. No se grafica la curva SROC.

Opciones de heterogeneidad

- **galb(tpr|tnr|dlor|lrp|lrn)** produce el gráfico de Galbraith de las transformaciones estandarizadas de los logit de (TVP, TFP) y el inverso de sus precisiones. Una recta de regresión que pasa por el origen es estimada, junto con una región de 95 % de confianza. Los estudios fuera de estas regiones se consideran datos atípicos.
- **regvars(varlist)** permite realizar un análisis univariante de meta-regresión de una o múltiples covariables dicotómicas o continuas, reportando los resultados en una tabla y forest-plot.

Opciones de sesgo de publicación

- **pubbias** cuando esta opción se invoca, Stata realiza una regresión de los log de los odds ratios con el inverso de la raíz de los tamaños muestrales como una prueba para la asimetría del gráfico, el cual tiene forma de embudo en un meta-análisis de diagnóstico. Un coeficiente diferente de cero indica que el sesgo del estudio es no significativo (p-value < 0.10).

Opciones de utilidad clínica

- **fagan(0-0.99)** crea un gráfico mostrando la relación entre la probabilidad a priori especificada por el usuario, el ratio de máxima verosimilitud (combinación de sensibilidad y especificidad) y la probabilidad de prueba a posteriori.
- **pddam(lbp ubp)** produce una línea gráfica de la probabilidad de prueba a posteriori versus la probabilidad a priori, los valores se encuentran entre 0 y 1 usando los resúmenes de los ratios de máxima verosimilitud.
- **lrmatrix** crea un gráfico de dispersión entre los ratios de máxima verosimilitud positivo y negativo. El gráfico se divide en cuadrantes basados en la intensidad o relación de los umbrales para determinar medidas informativas del test.

Opciones de misceláneos

- **level(#)** especifica los niveles de significancia para las pruebas estadísticas, regiones de confianza, regiones de predicción, e intervalos de confianza.
- **mscale(#)** afecta al tamaño de los puntos estimados en los forest-plots.
- **scheme(string)** permite la elección del esquema gráfico.
- **textscale(#)** permite la elección del tamaño del texto para gráficos, en especial con respecto a las etiquetas de los forest-plots.
- **zcf(#)** define una corrección de continuidad fija para cuando un estudio contiene una celda con cero.

4.2.7. Guardando los resultados

Midas guarda los resultados como $r()$:

- $r(\mathbf{mtp\!r})$ Media de la sensibilidad.
- $r(\mathbf{mtp\!rse})$ Error estándar de la media de la sensibilidad.
- $r(\mathbf{mt\!nr})$ Media de la especificidad.
- $r(\mathbf{mt\!nrse})$ Error estándar de la media de la especificidad.
- $r(\mathbf{ml\!rp})$ Media del ratio de máxima verosimilitud del resultado de una prueba positiva.
- $r(\mathbf{ml\!rpse})$ Error estándar de la media del ratio de máxima verosimilitud del resultado de una prueba positiva
- $r(\mathbf{ml\!rn})$: Media del ratio de máxima verosimilitud del resultado de una prueba negativa.
- $r(\mathbf{ml\!rnse})$ Error estándar de la media del ratio de máxima verosimilitud del resultado de una prueba negativa.
- $r(\mathbf{md\!or})$ Media de los odds ratio de diagnóstico.
- $r(\mathbf{md\!orse})$ Error estándar de la media de los odds ratio de diagnóstico.
- $r(\mathbf{AUC})$ Área bajo la curva SROC.
- $r(\mathbf{AUClo})$ Límite inferior del área bajo la curva SROC.
- $r(\mathbf{AUC\!hi})$ Límite superior del área bajo la curva SROC.
- $r(\mathbf{rho})$ Correlación entre los logits de la sensibilidad y especificidad.
- $r(\mathbf{reffs1})$ Varianza de los logit de la sensibilidad.
- $r(\mathbf{reffs1se})$ Error estándar de la varianza del logit de la sensibilidad.
- $r(\mathbf{reffs2})$ Varianza del logit de la especificidad.
- $r(\mathbf{reffs2se})$ Error estándar de la varianza del logit de la especificidad.
- $r(\mathbf{Islrt})$ Índice de inconsistencia global del ratio de la prueba de máxima verosimilitud.
- $r(\mathbf{Islrtlo})$ Límite inferior del índice global de inconsistencia.
- $r(\mathbf{Islr\!thi})$ Límite superior del índice global de inconsistencia.

4.2.8. Ejemplo de un conjunto de datos (dataset)

Para acceder al conjunto de datos *midas_example.dta*, se puede realizar de dos maneras, vía sentencias o por ventanas. Vía sentencia: seleccionar la ruta donde se encuentra el archivo *midas_example.dta* con la sentencia, en nuestro caso `cd "C:\Program Files (x86)\Stata12\ado\base\m"`, Figura 4.10 o vía ventana, Figura 4.11, si se utiliza sentencia se debe cargar el archivo con la sentencia `. use midas_example`, Figura 4.12.

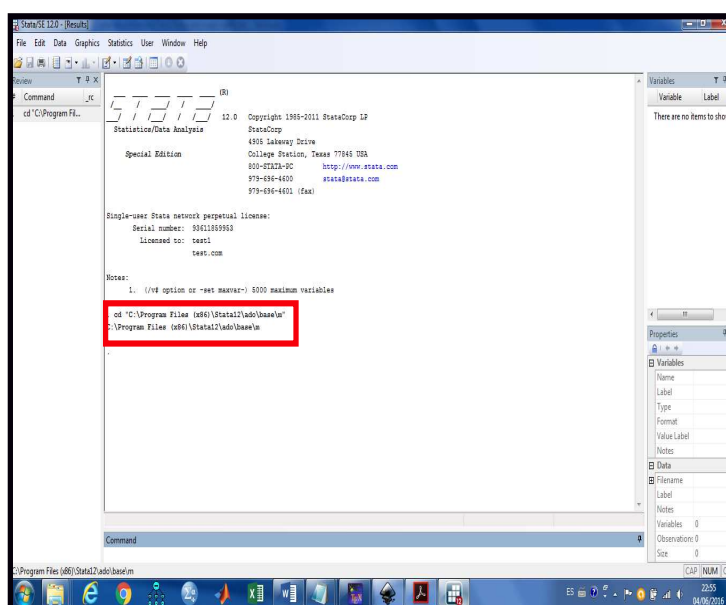


Figura 4.10: Comando para leer en disco duro la ruta del archivo *midas_example*

4. ANÁLISIS DE RESULTADOS

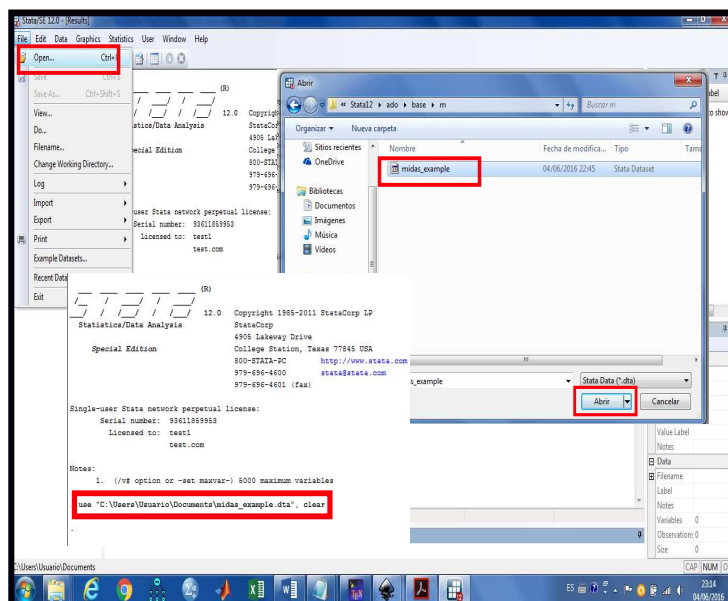


Figura 4.11: Acceso a *midas_example.dta* desde las pantallas

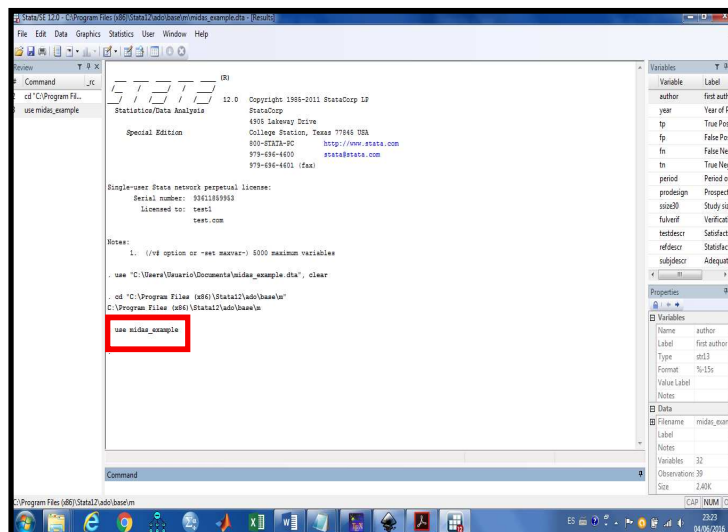


Figura 4.12: Comando para usar los datos de *midas_example.dta*

Para conocer la naturaleza de los datos contenidos en el dataset se utilizan las sentencias, `describe` o la sentencia `list author year pmet` `sampsize tp fp fn tn in 1\20, sep(1) ab(32) abs noo`.

4.2 Precisión diagnóstica con Stata para Meta-Análisis

```
. describe
```

```
Contains data from C:\Program Files (x86)\Stata12\ado\base\m\midas_example.dta
```

```
obs:          39
vars:         32          2 Feb 2009 14:41
size:        2,457
```

storage	display	value	label	variable label
variable name	type	format		
<hr/>				
author	str13	%-15s		first author
year	int	%8.0g		Year of Publication
tp	int	%8.0g		True Positive
fp	byte	%8.0g		False Positive
fn	float	%9.0g		False Negative
tn	float	%9.0g		True Negative
period	byte	%8.0g		Period of publication(before or
after 1998)				
prodesign	byte	%10.0g		Prospective design
ssize30	byte	%7.0g		Study size greater than 30
fulverif	byte	%8.0g		Verification of all by same
method				
testdescr	byte	%9.0g		Satisfactory description of index
test				
refdescr	byte	%9.0g		Satisfactory description of ref
test				
subjdescr	byte	%9.0g		Adequate description of study
subjects				
report	byte	%7.0g		Satisfactory reporting of results
consel	byte	%7.0g		Consecutive selection of subjects
brdspect	byte	%10.0g		Broad spectrum of disease
blinded	byte	%7.0g		Blinded interpretation of ref and
index tests				
sampsize	int	%8.0g		Number Data Units Per Study
qscore	byte	%9.0g		Quality score
age	byte	%8.0g		Mean Age
fast	byte	%8.0g		Prior Fasting of at least 4-6
hours				
res	byte	%8.0g		Resolution of PET Camera
dose	int	%8.0g		Appropriate dose used
upttime	byte	%8.0g		Uptake Period specified
acquitime	byte	%8.0g		Duration of image acquisition
testcrit	byte	%8.0g		Test criteria described

4. ANÁLISIS DE RESULTADOS

multiobs	byte	%8.0g	Multiple Observers
attcor	byte	%9.0g	Attenuation correction of PET
images			
axonly	byte	%8.0g	Dedicated axillary imaging
pmet	float	%9.2g	Study-specific Prevalence of
axillary Metastases			
Analysis	str7	%9s	Unit of Data Analysis(Patient
versus Nodes)			
patient	byte	%10.0g	Mode of analysis

Sorted by:

```
. list author year pmet sampsize tp fp fn tn in 1/20, sep(1) ab(32) abs noo
```

author	year	pmet	sampsize	tp	fp	fn	tn
Tse	1992	.7	10	4	0	3	3
Adler1	1993	.47	18	8	0	1	10
Hoh	1993	.64	14	6	0	3	5
Crowe	1994	.5	20	9	0	1	10
Avril	1996	.47	51	19	1	5	26
Bassa	1996	.81	16	10	0	3	3
Scheidhauer	1996	.5	18	9	1	0	8
Utech	1996	.35	124	44	20	0	60
Adler2	1997	.38	50	19	11	0	20
Palmedo	1997	.3	20	5	0	1	14
Noh	1998	.54	24	12	0	1	11
Smith	1998	.42	50	19	1	2	28
Rostom	1999	.65	74	42	0	6	26

Yutani1	1999	.38	26	8	0	2	16
Hubner	2000	.27	22	6	0	0	16
Ohta	2000	.59	32	14	0	5	13
Yutani2	2000	.42	38	8	0	8	22
Greco	2001	.43	167	68	13	4	82
Schirrmester	2001	.3	113	27	6	7	73
Yang	2001	.33	18	3	0	3	12

4.2.9. Modelos de predicción y diagnósticos

Para poder utilizar las sentencias de midas, se debe instalar en STATA con la sentencia `findit midas`, después se procede a dar click a cada uno de los recuadros rojos, Figura 4.13, con lo cual midas queda instalado.

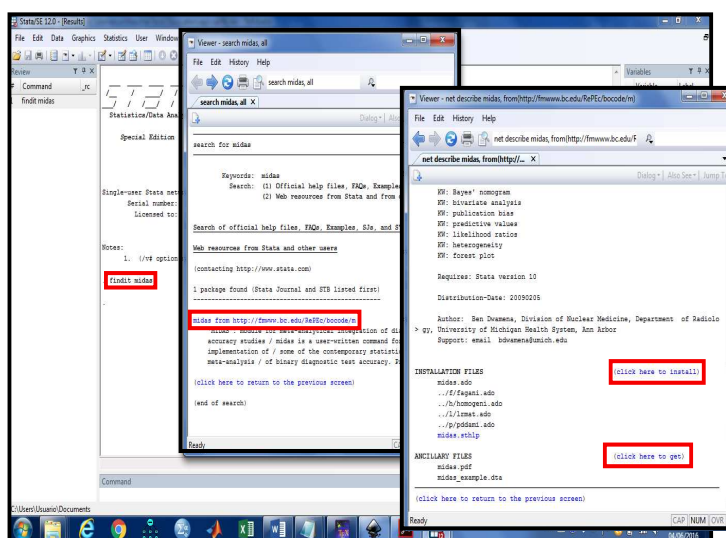


Figura 4.13: Instalación de Paquete midas desde el repositorio oficial

4. ANÁLISIS DE RESULTADOS

Las predicciones de las estimaciones a posteriori se pueden obtener a partir de estimaciones de parámetros y estimaciones empíricas de Bayes de efectos aleatorios, donde los errores típicos se estiman con el método delta. Los logit de las predicciones pueden ser transformados para obtener predicciones de sensibilidad y especificidad. Midas genera un forest-plot o una curva ROC de una empírica bayesiana versus estimaciones observadas de sensibilidad y especificidad usando `ebpre(for|roc)`. Utilizando la sentencia `midas tp fp fn tn, eb(for)`, se obtiene lo anteriormente mencionado, Figura 4.14 .

```
. midas tp fp fn tn, eb(for)
```

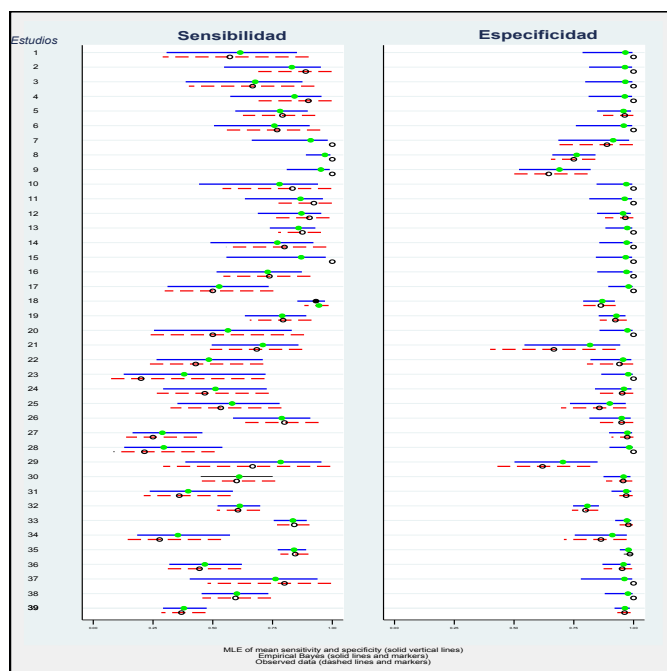


Figura 4.14: Bayes empírico predicho frente a los resultados de las pruebas observadas, datos `midas_example`

Los modelos de diagnóstico no suelen ser empleados de manera masiva, debido a que muchos investigadores no tienen la base estadística y otros consideran al meta-análisis como un procedimiento para tratar datos en lugar de un ajuste del modelo, Sutton and Higgins (180). Sin embargo, la aplicación de complejos modelos de máxima verosimilitud para el meta-análisis es importante para evaluar la posible mala especificación del modelo, bondad de ajuste, etc. **Midas** proporciona un modelo gráfico de evaluación de capacidades. Asimismo, cuenta con

gráfico de cuartiles de residuales basado en la bondad de ajuste, gráficos de probabilidades de la chi-cuadrado de las distancias de Mahalanobis al cuadrado, que permite, la evaluación de las hipótesis de la normal bivalente; gráficos spikeplot que detectan observaciones particularmente influyentes utilizando la distancia de Cook y diagramas de dispersión para la comprobación de los valores atípicos utilizando los pronósticos tipificados de efectos aleatorios con la sintaxis `midas tp fp fn tn, modchk(all)`, vea Figura 4.15.

```
. midas tp fp fn tn, modchk(all)
```

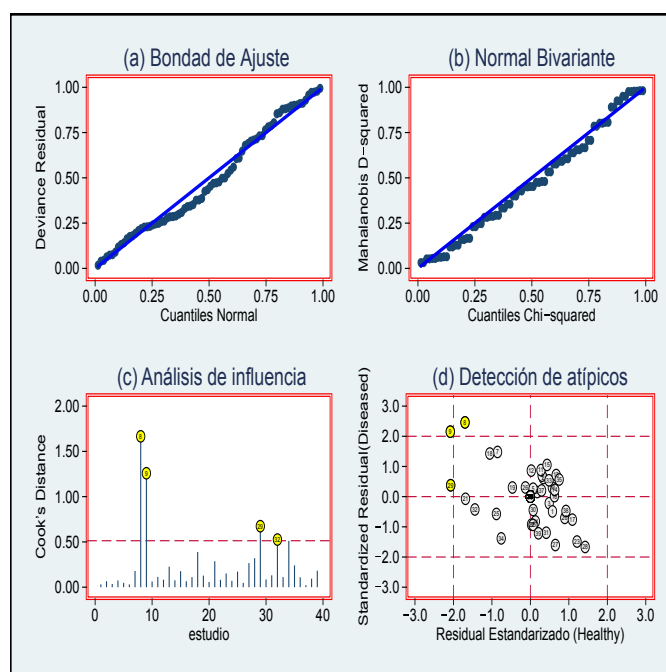


Figura 4.15: Representación gráfica de la normalidad bivalente, bondad de ajuste basada en residuales y los análisis de detección de influencia y atípicos, datos `midas_example`

4.2.10. Evaluación de la calidad

La calidad metodológica de un estudio es el grado en que se pueden mostrar todos los aspectos del diseño y realización de un estudio, para protegerlo contra sesgo sistemático y sesgo no sistemático, que puedan surgir en los estudios que no están bien diseñados y en los errores de inferencia. La herramienta para evaluar la calidad de los estudios de diagnóstico es (QUADAS) Whiting et al. (200, 201, 202, 204). Es una herramienta rigurosamente construida y validada que

4. ANÁLISIS DE RESULTADOS

puede ser utilizada por los investigadores para nuevas revisiones sistemáticas. La herramienta QUADAS contiene 14 ítems que cubren el espectro del paciente: patrón de referencia, sesgo de progresión de la enfermedad, verificación y revisión del sesgo, sesgo de revisión clínica, sesgo de incorporación, ejecución de la prueba, retiros de los estudios y los resultados intermedios. Los métodos posibles para hacer frente a las diferencias de calidad son el análisis de sensibilidad, análisis de subgrupos o meta-análisis de regresión. Antes de utilizar QUADAS hay que instalarlo en STATA, con el mismo procedimiento que se empleó para la instalación de midas, es decir, se escribe la sentencia `findit quadas`, Figura 4.16.

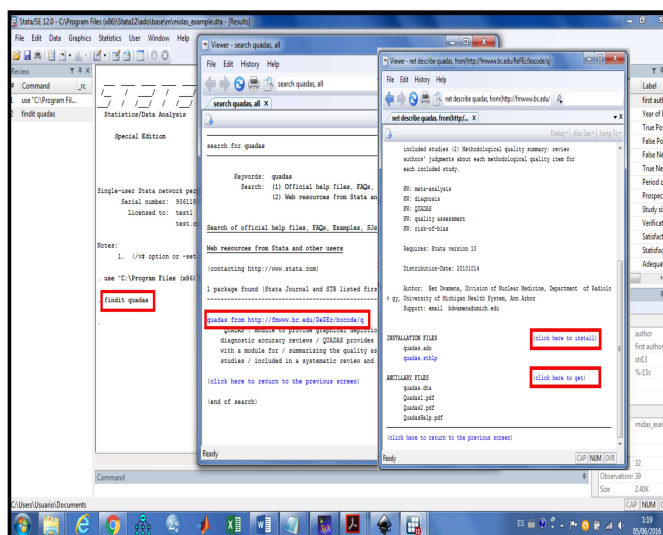


Figura 4.16: Instalación del paquete QUADAS desde el repositorio oficial

La evaluación de la calidad puede resumirse para cada elemento QUADAS como un gráfico de barras, note Figura 4.17, el cual se genera mediante la sentencia:

```
. midas tp fp fn tn, qbar(prodesign ssize30 fulverif testdescr refdescr subjdes
> cr report brdspect blinded) qlab
```

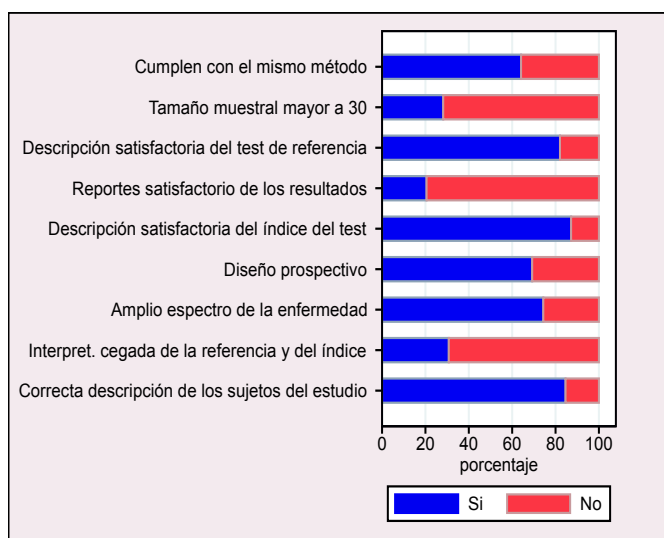



Figura 4.17: Gráfico de barras de la evaluación de la calidad para cada artículo, incluidos los criterios modificados QUADAS, datos midas_example

Como alternativa al uso de **qbar**, existe **qtab (lista de variables)** la cual produce tabla de frecuencias de las puntuaciones de calidad. La tabla se genera mediante la sentencia:

```
. midas tp fp fn tn, qtab(prodesignn ssize30 fulverif testdescr refdescr subjes
> cr report brdspect blinded) qlab
```

METHODOLOGICAL QUALITY ASSESSMENT

Criterion	Yes	Yes (%)	No	No (%)
Prospective design	27	69	12	31
Study size greater than 30	11	28	28	72
Verification of all by same method	25	64	14	36
Satisfactory description of index test	34	87	5	13
Satisfactory description of ref test	32	82	7	18
Adequate description of study subjects	33	85	6	15
Satisfactory reporting of results	8	21	31	79
Broad spectrum of disease	29	74	10	26
Blinded interpretation of ref and index	12	31	27	69

4.2.11. Asociación bivalente

Midas utiliza un modelo de efectos aleatorios bivalente que modeliza la sensibilidad y especificidad, por lo que se espera que el desempeño de este par de medidas sean independientes. El diagrama de cajas bivalente describe el grado de interdependencia, incluyendo la ubicación central y cualquier valor atípico. El óvalo interior representa la distribución media de los datos. El óvalo exterior representa la región de confianza al 95 %. La figura se obtiene con la sintaxis `midas tp fp fn tn, bivbox`, vea Figura 4.18.

```
. midas tp fp fn tn, bivbox
```

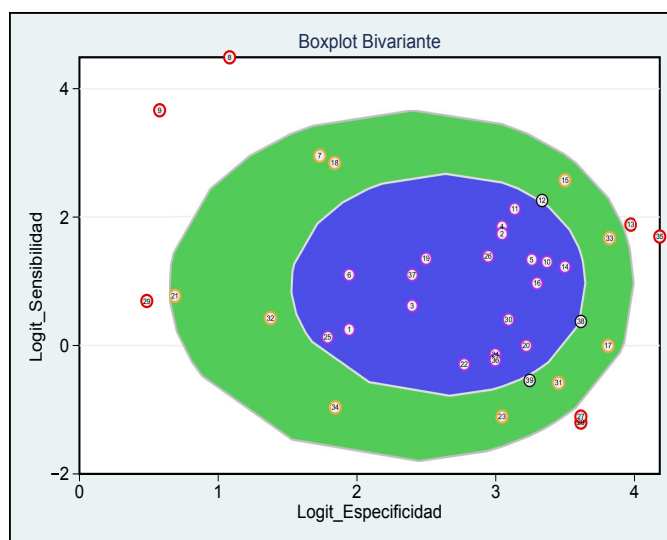


Figura 4.18: Diagrama de caja bivalente de los estudios agrupados con siete valores atípicos que sugiere indirectamente un menor grado de heterogeneidad, datos `midas_example`

El gráfico muestra una asimetría de las medias de rendimiento de la prueba hacia una mayor especificidad con menor sensibilidad, proporcionando evidencia indirecta de cierta variabilidad del umbral.

4.2.12. Resúmenes de las estimaciones de rendimiento

Los resúmenes de las estimaciones de la sensibilidad y especificidad y sus intervalos de confianza al 95 % pueden ser calculados después de la transformación anti-log de la media de los logaritmos de la sensibilidad y especificidad y sus errores estándar. Estos intervalos tienen en cuenta la heterogeneidad entre los estudios más allá del azar (modelo de efectos aleatorios). Esto es posible con la sentencia `midas tp fp fn tn, res(sum) nip(1)`.

```
. midas tp fp fn tn, res(sum) nip(1)
```

SUMMARY PERFORMANCE ESTIMATES

Parameter	Estimate	95% CI
Sensitivity	0.73 [0.63, 0.80]
Specificity	0.96 [0.92, 0.97]
Positive Likelihood Ratio	16.2 [9.7, 27.3]
Negative Likelihood Ratio	0.29 [0.21, 0.39]
Diagnostic Odds Ratio	57 [30, 107]

4.2.13. Heterogeneidad estadística

El impacto de la heterogeneidad no observada tradicionalmente se evalúa estadísticamente mediante el I^2 , el cual describe el porcentaje de variación total de los estudios que se debe a la heterogeneidad en lugar del azar, Higgins and Thompson (83). I^2 se puede calcular a partir de la heterogeneidad estadística y los grados de libertad, Higgins and Thompson (83). Por otra parte, para los modelos mixtos, es el coeficiente de correlación intra-clase expresado como un porcentaje. En midas se implementa por separado para la sensibilidad y especificidad. I^2 puede tomar valores entre 0 % y 100 %. Un valor de 0 % indica ausencia de heterogeneidad y valores por encima de 50 % se puede considerar heterogeneidad significativa. La ventaja principal del I^2 radica en que no depende intrínsecamente del número de estudios en el meta-análisis, Higgins and Thompson (83). La salida de la heterogeneidad en midas, se realiza con la sentencia `midas tp fp fn tn, res(het) nip(1)`, Figura 4.19.

4. ANÁLISIS DE RESULTADOS

```
. midas tp fp fn tn, res(het) nip(1)
```

HETEROGENEITY STATISTICS

Heterogeneity (Chi-square): LRT_Q = 146.822, df =2.00, LRT_p =0.000

Inconsistency (I-square): LRT_I2 = 99, 95% CI = [98- 99]

Proportion of heterogeneity likely due to threshold effect = 0.11

Interstudy variation in Sensitivity: ICC_SEN = 0.31, 95% CI = [0.18- 0.45]

Interstudy variation in Sensitivity: MED_SEN = 0.76, 95% CI = [0.70- 0.83]

Interstudy variation in Specificity: ICC_SPE = 0.29, 95% CI = [0.12- 0.45]

Interstudy variation in Specificity: MED_SPE = 0.75, 95% CI = [0.68- 0.84]

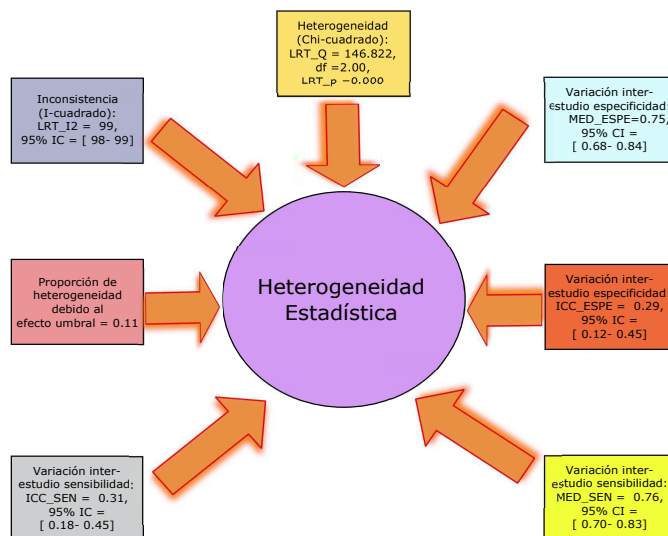


Figura 4.19: Salida de la estadística de heterogeneidad, datos midas_example, elaboración propia

4.2.14. Forest-Plot para analizar la sensibilidad y especificidad

En midas, los forest-plot se pueden crear para cada parámetro de rendimiento de la prueba de manera individual o pueden aparecer como par ordenado de la sensibilidad y especificidad, Figura 4.20. Lo anterior se puede visualizar con la sentencia

```
. midas tp fp fn tn, texts(0.60) bfor(dss) id(author year) ford fors
```

Es importante notar que los Forest Plot de sensibilidad y especificidad ayudan en la evaluación de la heterogeneidad de los aspectos individuales de la exactitud de la prueba, pero no permiten la evaluación inmediata de si la variación observada es la esperada por la relación de dos variables, es decir, la sensibilidad disminuye a medida que la especificidad se eleva, Phillips et al. (144).

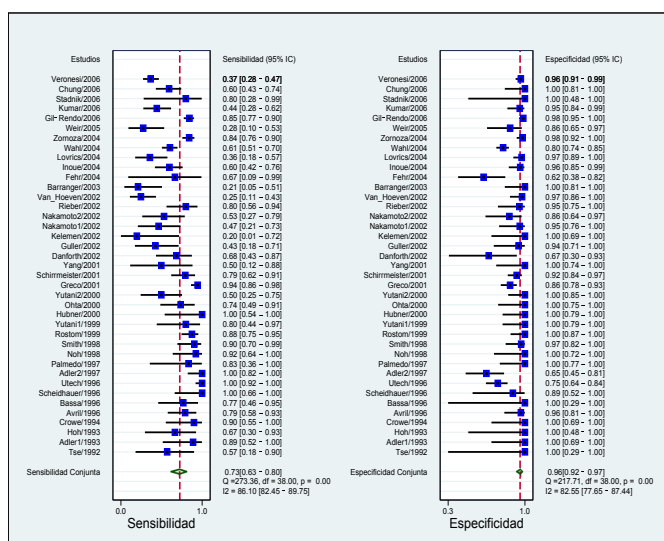


Figura 4.20: Diagrama forest plot para cada estudio específico (eje derecho) y la media de sensibilidad y especificidad con las estadísticas de heterogeneidad correspondientes, datos midas_example

4.2.15. Curva SROC

Sobre la base de las estimaciones de los parámetros del modelo bivalente, las curvas ROC resumen pueden ser expresadas como líneas de regresión que relacionan los logit de la sensibilidad y la especificidad. En *midas* las estimaciones logit derivadas de la sensibilidad, especificidad y sus varianzas respectivas se utilizan para construir una curva jerárquica SROC. Así se puede construir la curva ROC con la sentencia `midas tp fp fn tn, sroc(both)`, note Figura 4.21.

```
. midas tp fp fn tn, sroc(both)
```

La línea entrecortada alrededor del punto resumen representa la región de confianza al 95 %, el área bajo la curva (AUROC), es utilizada como una medida global del desempeño de la prueba. El *AUROC* es el promedio de *TVP* sobre la *TFP*. La Tabla 4.4 se ha sugerido para dar una interpretación a los valores de AUROC, Swets (181).

Intervalos de AUROC	Interpretación
$0.5 \leq \text{AUROC} < 0.7$	Moderado
$0.7 \leq \text{AUROC} < 0.9$	Intermedio
$0.9 \leq \text{AUROC} < 1$	Precisión

Tabla 4.4: Interpretación valores AUROC

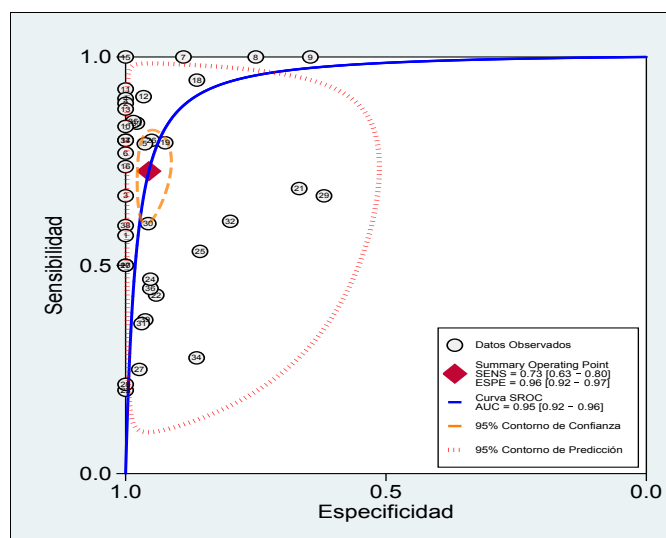


Figura 4.21: Resumen de la curva ROC con las regiones de confianza y predicción alrededor de la media de sensibilidad y el punto de especificidad, datos midas_example

4.2.16. Meta-Regresión

Meta-regresión usa los métodos de regresión que incorpora los efectos de la covarianza de los factores sobre las medidas resúmenes de desempeño y ha sido empleada para explorar la heterogeneidad entre estudios terapéuticos. En los estudios de diagnóstico, la heterogeneidad en la sensibilidad y especificidad puede ser resultado de muchas causas relacionadas con las definiciones de las pruebas estándar y referencias, las características operativas del test, métodos de recolección de datos y características del paciente.

Las covarianzas pueden ser introducidas en el modelo como cualquier medida de rendimiento de la prueba. El tamaño de la muestra corresponderá al número de estudio en el análisis, para tamaño de muestra pequeño se limita el poder de la regresión para detectar efectos significativos.

Las siguientes sintaxis generan lo antes mencionado, note Figura 4.22:

```
. midas tp fp fn tn, reg(prodesign age qscore sampsize fulverif testdescr refde
> scr clindescr report spectrum blindref blindtest)
variable clindescr not found
in option regvars()
r(111);
```

4. ANÁLISIS DE RESULTADOS

```
. midas tp fp fn tn, reg(prodesign ssize30 fulverif testdescr refdescr subjdesc
> r report brdspect blinded)
```

Estimating Covariate Effect Of: prodesign

Estimating Covariate Effect Of: ssize30

Estimating Covariate Effect Of: fulverif

Estimating Covariate Effect Of: testdescr

Estimating Covariate Effect Of: refdescr

Estimating Covariate Effect Of: subjdesc

Estimating Covariate Effect Of: report

Estimating Covariate Effect Of: brdspect

Estimating Covariate Effect Of: blinded

Sensitivity and Specificity

Parameter	category	nstudies	Sensitivity	p1	Specificity	p2
prodesign	Yes	27	0.73 [0.63 - 0.84]	0.40	0.94 [0.91 -0.97]	0.00
	No	12	0.71 [0.55 - 0.87]	.	0.98 [0.95 -1.00]	.
ssize30	Yes	11	0.74 [0.60 - 0.89]	0.35	0.94 [0.90 -0.99]	0.00
	No	28	0.72 [0.61 - 0.83]	.	0.96 [0.94 -0.99]	.
fulverif	Yes	25	0.73 [0.62 - 0.84]	0.29	0.95 [0.93 -0.98]	0.01
	No	14	0.72 [0.57 - 0.87]	.	0.96 [0.92 -0.99]	.
testdescr	Yes	34	0.72 [0.63 - 0.82]	0.51	0.95 [0.92 -0.98]	0.22
	No	5	0.74 [0.51 - 0.97]	.	0.98 [0.96 -1.00]	.
refdescr	Yes	32	0.72 [0.63 - 0.82]	0.38	0.95 [0.93 -0.98]	0.09
	No	7	0.75 [0.55 - 0.95]	.	0.97 [0.93 -1.00]	.
subjdesc	Yes	33	0.68 [0.59 - 0.78]	0.01	0.95 [0.93 -0.98]	0.35
	No	6	0.90 [0.80 - 1.00]	.	0.98 [0.93 -1.00]	.

4.2 Precisión diagnóstica con Stata para Meta-Análisis

report	Yes	8	0.76	[0.58 - 0.93]	0.57	0.92	[0.85 -0.99]	0.00
	No	31	0.72	[0.62 - 0.82]	.	0.96	[0.94 -0.99]	.
brdspect	Yes	29	0.74	[0.65 - 0.84]	0.82	0.95	[0.92 -0.98]	0.02
	No	10	0.66	[0.46 - 0.86]	.	0.97	[0.94 -1.00]	.
blinded	Yes	12	0.54	[0.37 - 0.71]	0.00	0.95	[0.91 -0.99]	0.01
	No	27	0.79	[0.71 - 0.87]	.	0.96	[0.93 -0.99]	.

Joint Model

Parameter	category	LRTChi2	Pvalue	I2	I2lo	I2hi
prodesign	Yes	2.73	0.26	27	0	100
	No
ssize30	Yes	0.67	0.71	0	0	100
	No
fulverif	Yes	0.03	0.99	0	0	100
	No
testdescr	Yes	1.79	0.41	0	0	100
	No
refdescr	Yes	0.61	0.74	0	0	100
	No
subjdescr	Yes	6.75	0.03	70	34	100
	No
report	Yes	2.75	0.25	27	0	100
	No
brdspect	Yes	1.15	0.56	0	0	100
	No
blinded	Yes	8.35	0.02	76	48	100
	No

4. ANÁLISIS DE RESULTADOS

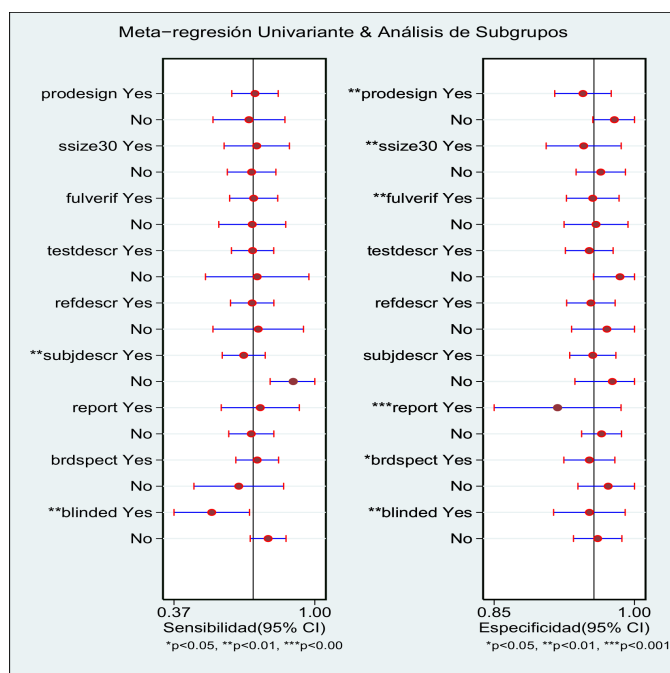


Figura 4.22: Forest plot de una meta-regresión univariante múltiple, datos midas_example

4.2.17. Test de regresión lineal de asimetría-Gráfico Funnel Plot

El gráfico funnel plot es generalmente considerado una buena herramienta de exploración para investigar el sesgo de publicación. Se genera mediante el trazado de la medida del tamaño del efecto versus una medida de la precisión. Si la gráfica es simétrica, se infiere que no existe el sesgo de publicación. Sin embargo, la evaluación por este método puede ser subjetiva. Se han desarrollado métodos de regresión lineal paramétricos y no paramétricos para probar formalmente la simetría del gráfico funnel plot, Begg and Mazumdar (11), Egger and Smith (47), Harbord et al. (77), Macaskill et al. (120), Peters et al. (143), Rücker et al. (155), Schwarzer et al. (168, 169).

El uso de estos métodos para evaluar el sesgo de publicación en los estudios de pruebas diagnósticas puede producir resultados engañosos, Deeks et al. (33), Song et al. (176). Las pruebas formales para el sesgo de publicación pueden ser llevadas a cabo por una regresión lineal, donde, la variable independiente son los logaritmos de los odds ratio de diagnóstico y la variable

dependiente es $\frac{1}{\sqrt{\text{tamaño real de la muestra}}}$, donde un P-valor < 0.10 para la pendiente de la recta de regresión indica asimetría significativa, Deeks et al. (33). Antes de ejecutar la sentencia `midas tp fp fn tn, pubbias`, se debe instalar **mylabels** como en los casos anteriores. La salida también genera un gráfico, note Figura 4.23.

```
. midas tp fp fn tn, pubbias
```

STATISTICAL TESTS FOR SMALL STUDY EFFECTS/PUBLICATION BIAS

yb	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Bias	-3.206976	4.332084	-0.74	0.464	-11.98461 5.57066
Intercept	4.255449	.5420326	7.85	0.000	3.157187 5.353711

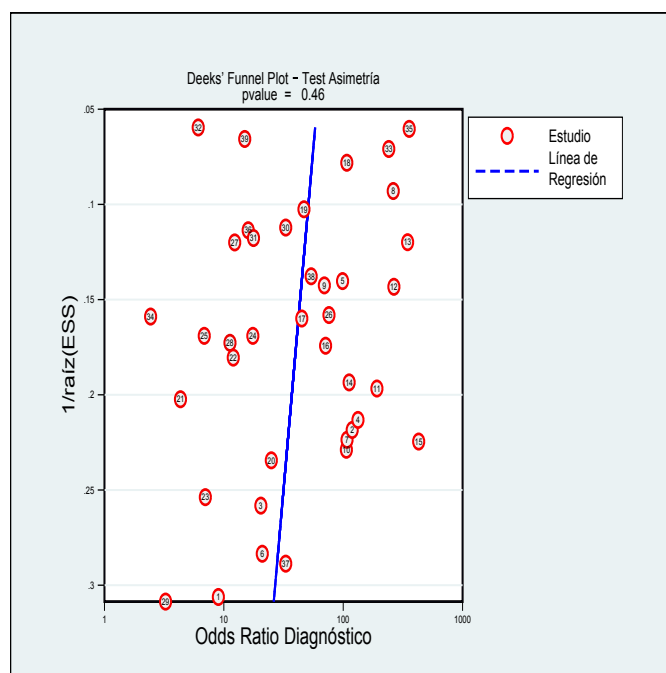


Figura 4.23: Gráfico Funnel Plot con la línea de regresión superpuesta, datos `midas_example`

El valor p de 0.46 para el coeficiente de la pendiente sugiere la simetría en los datos y una baja

probabilidad de sesgo de publicación. Sin embargo se conoce que el test tiene baja potencia, Deeks et al. (33).

4.2.18. Fagan Plot (Nomograma de Bayes)

La utilidad clínica o relevancia del paciente de una prueba de diagnóstico es evaluada usando los ratios de máxima verosimilitud para calcular la probabilidad post-test (PTP). Basada en el teorema de Bayes, este concepto se muestra en el nomograma de Fagan, Fagan (52). Cuando el teorema de Bayes se expresa en términos de los logaritmos de los odds ratio, el logaritmo a posteriori es una función lineal de los logaritmos a priori de los odds ratio y de los logaritmos de los ratios de máxima verosimilitud. Un gráfico de Fagan consiste en un eje vertical del lado izquierdo de los log odds ratio a priori, un eje en el centro del gráfico que representa a los log de los odds ratio de máxima verosimilitud y un eje vertical en el lado derecho de la gráfica que denota los log de los odds ratio a posteriori. Las líneas son dibujadas desde el lado izquierdo, pasando por el centro hasta el lado derecho del gráfico, con la sentencia `midas tp fp fn tn, fagan(0.25)`, note Figura 4.24. Al observar la Figura 4.24 podemos notar, que si un paciente obtuvo un test positivo, con el nomograma podemos inferir que tendría la enfermedad con una probabilidad del 84%. Por otro lado, si el test fue negativo, se puede inferir que el paciente no tendrá la enfermedad con una probabilidad del 9%.

```
. midas tp fp fn tn, fagan(0.25)
```

```
FAGAN'S (BAYESIAN) NOMOGRAM
```

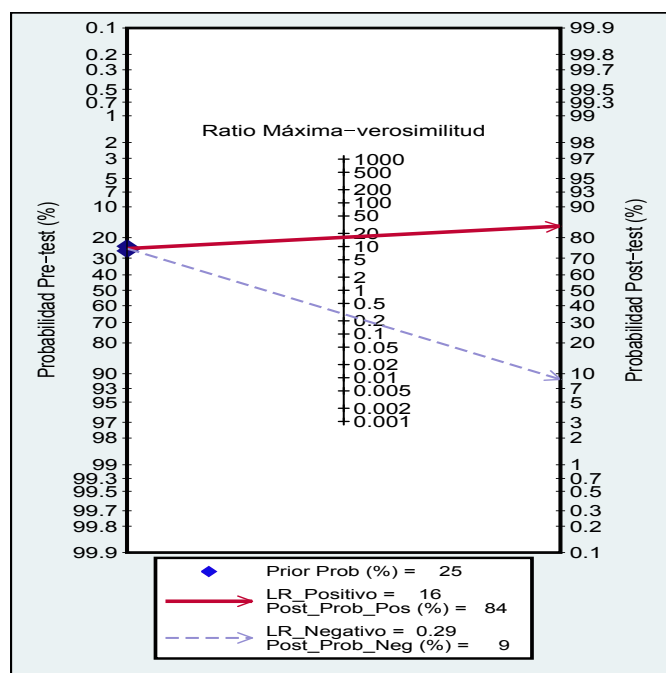


Figura 4.24: Gráfica de Fagan, datos midas_example

4.2.19. Diagrama de dispersión de los Odds ratio de Máxima Verosimilitud

Los resúmenes pueden ser representados de manera gráfica en forma de matriz o por un diagrama de dispersión de los odds ratios de máxima verosimilitud, Stengel et al. (178). En él se definen cuadrantes sobre la base de los umbrales establecidos:

- Cuadrante superior izquierdo-CSI, ratio de máxima verosimilitud positivo > 10 , ratio de máxima verosimilitud negativo < 0.1 ; Exclusión & Confirmación.
- Cuadrante superior derecho-CSD, ratio de máxima verosimilitud positivo > 10 , ratio de máxima verosimilitud negativo > 0.1 ; Solo confirmación.
- Cuadrante inferior izquierdo-CII, ratio de máxima verosimilitud positivo < 10 , ratio de máxima verosimilitud negativo < 0.1 ; Solo exclusión.
- Cuadrante inferior derecho-CID, ratio de máxima verosimilitud < 10 , ratio de máxima verosimilitud negativo > 0.1 ; No exclusión o confirmación.

4. ANÁLISIS DE RESULTADOS

Para generar el gráfico se utiliza la sentencia `midas tp fp fn tn, lrmats`, note Figura 4.25.

```
. midas tp fp fn tn, lrmats
```

LIKELIHOOD RATIO SCATTERGRAM

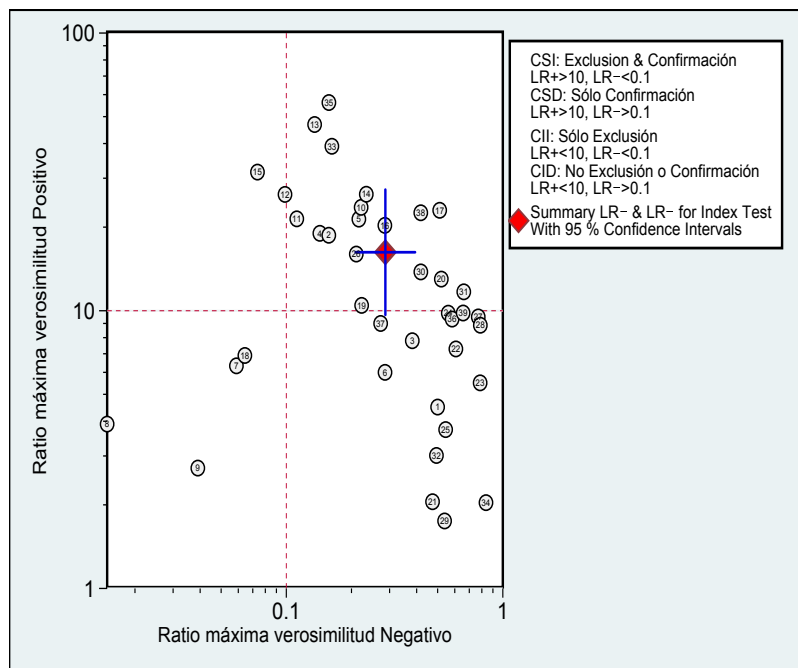


Figura 4.25: Razón de verosimilitud y diagrama de dispersión, datos midas_example

El gráfico de dispersión de los ratio de máxima verosimilitud muestra los puntos de la razón de verosimilitud obtenidos como funciones de la media de la sensibilidad y especificidad, Leeflang et al. (106). El cuadrante superior derecho sugiere que la prueba es útil para la confirmación de la presencia de la enfermedad (cuando es positiva) y no para su exclusión (cuando es negativa).

4.2.20. Valores predictivos y gráfico de probabilidad modificatoria

La probabilidad condicional de la enfermedad cuando la prueba es positiva o negativa, los llamados valores predictivos positivos (negativos), son de importancia crítica para la aplicación clínica de un procedimiento de diagnóstico. Estas medidas no dependen solamente de la sensibilidad y especificidad sino también de la prevalencia de la enfermedad. La gráfica de probabilidad

modificatoria es un análisis de sensibilidad gráfica del valor predictivo de un proceso continuo que define prevalencia baja para poblaciones de alto riesgo. Se representan curvas separadas para las pruebas positivas y negativas, con la sentencia `midas tp fp fn tn, pddam(0.25 0.75)`, note Figura 4.26.

```
. midas tp fp fn tn, pddam(0.25 0.75)
```

CONDITIONAL PROBABILITY PLOT AND UNCONDITIONAL PREDICTIVE VALUES

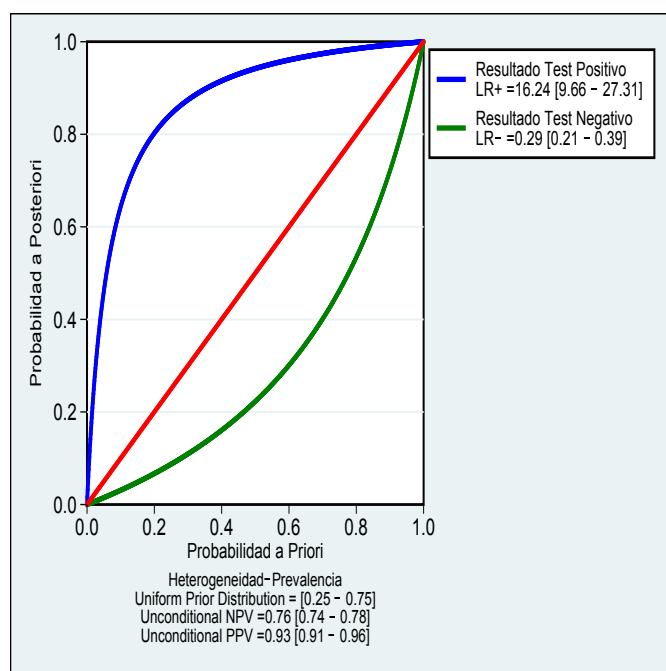


Figura 4.26: Gráfica de Probabilidad modificada, datos `midas_example`

4.3. Precisión diagnóstica con HSROC para Meta-Análisis

4.3.1. Introducción

Esta sección está destinada a servir como una guía para el uso del paquete HSROC que es empleado en un entorno de software estadístico libre. Por lo tanto, el objetivo principal será describir las funciones en el paquete a través de diferentes ejemplos, sin énfasis en la teoría estadística detrás de ellos. Para profundizar más en la parte matemática, consultar en Dendukuri et al. (37).

El paquete *HSROC* consta de 4 funciones y 2 conjuntos de datos. Las dos funciones principales son **HSROC** y **HSROCSummary** y las 2 restantes funciones son funciones secundarias.

- **HSROC**, el cual debe ejecutarse primero, se utiliza para implementar el algoritmo de Cadenas de Markov Monte Carlo.
- **HSROCSummary** produce el resumen de parámetros para el modelo HSROC.
- **simdata** simula un conjunto de datos basado en el modelo de meta-análisis de diagnóstico sobre HSROC.
- **beta.parameter** devuelve los parámetros de la función de probabilidad Beta (α, β) correspondiente a un rango dado.

En las siguientes secciones se presenta un ejemplo para explicar cómo hacer uso de las funciones dentro del paquete de HSROC. Se basa en una prueba bajo supervisión que es comparada con un ensayo de referencia perfecta cuando ambas pruebas son independientes dado el verdadero estado de la enfermedad.

4.3.2. Meta-Análisis de una prueba oro de referencia estándar asumiendo independencia condicional

Primero se empieza instalando el paquete y la librería *HSROC* en R, mediante las sentencias:


```
> install.packages("HSROC")  
> library("HSROC")
```

Empezamos con el modo más simple, donde se asume que la prueba de referencia es la prueba oro (i.e, la sensibilidad y especificidad de la prueba de referencia ambas igual a 100%). Para este ejemplo, se utilizan los datos de la resonancia magnética (MR) a partir de 10 estudios realizado por Scheidler et al. (165). Este es un subconjunto del conjunto de datos utilizado en el documento que describe el modelo *HSROC* de Rutter and Gatsonis (158).

4.3.3. Preparación de los datos

Los datos de las imágenes de resonancia magnética están incluidos dentro del paquete y pueden ser cargados y leídos con las líneas de comando `data("MRI")` y `MRI`. La salida de las sentencias previas se nota a continuación.

```
> data(MRI)  
> MRI  
  
++ +- -+ --  
1  9  2  2  44  
2  3  6  5  32  
3  3  2  1  16  
4  3  1 12  44  
5  1  1  6  16  
6  7  2 22 167  
7 12  4  4  29  
8 23  5 14 230  
9  8  5  5  53  
10 16  2  2  22
```

Las columnas ++, +-, -+, --, representan los resultados de las tablas cruzadas entre MRI (prueba de evaluación) y muestras Histológicas/citológicas obtenidas por la cirugía de los ganglios linfáticos, biopsia (prueba de referencia). Con el fin de estimar los parámetros del modelo

4. ANÁLISIS DE RESULTADOS

condicional independiente se utiliza la función HSROC. Para ver los parámetros de la función HSROC se ejecuta la línea de comando `args(HSROC)`, cuya salida es:

```
> args(HSROC)

function (data, iter.num, init = NULL, sub_rs = NULL, first.run = TRUE,
path = getwd(), refresh = 100, prior.SEref = NULL, prior.SPref = NULL,
prior.PI = c(0, 1), prior.LAMBDA = c(-3, 3), prior.THETA = c(-1.5,
1.5), prior.sd.alpha = list(0, 2, "sd"), prior.sd.theta = list(0,
2, "sd"), prior.beta = c(-0.75, 0.75))
NULL
```

La función anterior, resulta de extraer una muestra de la distribución a posteriori del modelo de una vía de la distribución de la simulación de Cadena de Markov y Monte Carlo (MCMC). La función contiene una serie de argumentos, los cuales determinan las distribuciones a priori, y tienen por defecto ciertos valores. En particular, los valores por defecto **prior.SEref = NULL** y **prior.SPref = NULL** definen una referencia a la prueba de oro estándar. La simulación de **Cadenas Markov y Monte Carlo (MCMC)** requiere de valores iniciales que se proporcionarán en este ejemplo. Por lo tanto, lo único que se altera de los valores predeterminados en este ejemplo, es el argumento **init**. Cuando este argumento se deja igual a su valor predeterminado, los valores iniciales requeridos para la simulación de la Cadena de Markov y Monte Carlo (MCMC) son seleccionados al azar por la función *HSROC*, basado en las distribuciones a priori.

El argumento **init** es un objeto tipo lista compuesto por los valores iniciales de los parámetros intra estudio como primer elemento de la lista y los valores iniciales de los parámetros inter estudios como el segundo elemento de la lista. El elemento intra estudio debe ser un objeto tipo matriz donde cada columna corresponde a un parámetro diferente y cada fila corresponde a un diferente estudio, mientras que el elemento inter estudio debe ser un vector con cada elemento correspondiente a un parámetro diferente. Supóngase que se desean los siguientes valores iniciales para cada uno de los parámetros intra estudio α_i (precisión diagnóstica del estudio i), θ_i (el valor de corte para un resultado positivo en el estudio i), S_{1i} (sensibilidad del estudio i para la prueba en evaluación), C_{1i} (especificidad del estudio i para la prueba en evaluación) y π_i (la prevalencia del estudio i). Para lograr la inicialización de los valores antes mencionados se deben realizar las siguientes sentencias:

```
> init.alpha = c(2.51, 2.54, 3.81, 2.41, 2.64, 2.70, 3.31, 3.39, 3.11, 2.99)
> init.theta = c(-0.51, -0.39, 0.33, -2.06, -0.14, -0.08, 1.11, 0.38, -0.86,
+ -0.38)
> init.s1 = rep(0.9,10)
> init.c1 = rep(0.9,10)
> init.pi = c(0.38, 0.17, 0.78, 0.07, 0.74, 0.84, 0.52, 0.95, 0.07, 0.56)
```

Luego se debe crear una matriz de valores iniciales intra-estudio, mediante el comando:

```
> init_within = cbind(init.alpha, init.theta, init.s1, init.c1, init.pi)
```

El orden de los valores iniciales de la función **cbind** es importante y no debe ser alterado. Es decir, la colocación de **init.theta** antes que **init.alpha**, con el comando:

```
> init_within = cbind(init.theta, init.alpha, init.s1, init.c1, init.pi)
```

Esto dará lugar a la inicialización de los parámetros α_i y θ_i . Esto podría dar lugar a una convergencia más lenta del muestreo en el caso que los valores de partida no sean adecuados. Ahora, para los parámetros inter estudios Θ (el valor de la media global para el punto de corte para pruebas positivas), σ_θ (el punto de corte para la desviación estándar inter estudios), Λ (la exactitud de los diagnósticos generales), σ_α (la desviación estándar de la diferencia de medias inter-estudios), y β (el logaritmo del ratio de la desviación estándar de los resultados de las pruebas entre los pacientes con la enfermedad y entre los pacientes sin la enfermedad), se supondrán los siguientes valores de partida:

```
> init.THETA = -0.16
> init.sig.theta = 0.75
> init.LAMBDA = 2.58
> init.sig.alpha = 0.5
> init.beta = 0.25
```

Luego se debe crear el vector de valores iniciales inter estudios, con la sentencia siguiente:

4. ANÁLISIS DE RESULTADOS

```
> init_between = c(init.THETA, init.sig.theta, init.LAMBDA,  
                  + init.sig.alpha, init.beta)
```

El orden en el vector anterior debe mantenerse. Por último, simplemente hay que fusionar el intra-estudio y los valores iniciales inter-estudios anteriores en una lista, con la sentencia:

```
> init = list(init_within, init_between)
```

4.3.4. Ejecutando la simulación Cadenas de Markov Monte Carlo (MCMC)

Para completar “**la llamada a la función**” necesitamos proveer dos argumentos adicionales (**data** e **iter.num**) los datos y el número de iteraciones, para el conjunto de datos y el número de iteraciones en la simulación de Cadenas de Markov y Monte Carlo (MCMC). En este ejemplo, vamos a ejecutar la Simulación para 50.000 iteraciones, llamando a la función:

```
> HSROC(data=MRI, iter.num=50000, init=init )
```

```
Starting the Gibbs sampler for 50000 iterations ...
```

```
100 iterations completed out of 50000 ...
```

```
200 iterations completed out of 50000 ...
```

```
300 iterations completed out of 50000 ...
```

```
⋮
```

```
49900 iterations completed out of 50000 ...
```

```
The files created during the Gibbs sampler process are in "C:/Trabajo/..."
```

En lugar de manejar grandes matrices de muestras a posteriori para cada parámetro dentro de R, que crecen en tamaño y en número de iteraciones que se incrementan, la función crea archivos de texto en el directorio de trabajo predeterminado, o en cualquier directorio de trabajo especificado por el argumento de ruta. Esto ayudará a aumentar la velocidad de cálculo con el número de iteraciones. Una vez que la función ha alcanzado el número de iteraciones indicadas, aparecerá un

mensaje, `The files created during the Gibbs sampler process are in "C:/Trabajo/..."`, donde **C:** \... es el directorio de trabajo donde se crean y guardan los archivos tipo texto.

4.3.5. Interpretando los archivos de salida

La función **HSROCSummary** se puede utilizar para obtener las estadísticas descriptivas y gráficos del ejemplo anterior. Los argumentos de la función se obtienen mediante la sentencia `args(HSROCSummary)`.

```
> args(HSROCSummary)

function (data, burn_in = 0, iter.keep = NULL, Thin = 1, sub_rs = NULL,
point_estimate = c("median", "mean"), summary.path = getwd(),
chain = getwd(), tv = NULL, digit = 6, print_plot = FALSE,
plot.ind.studies = TRUE, cred_region = TRUE, predict_region = TRUE,
col.pooled.estimate = "red", col.predict.region = "blue",
lty.cred.region = "dotdash", lty.predict.region = "dotted",
region_level = 0.95, trunc_low = 0.025, trunc_up = 0.025)
NULL
```

El argumento **data** de la función sirve para referenciar al conjunto de datos. En nuestro caso el argumento **burn_in** indica cuántas iteraciones se deben realizar antes del cálculo de las estimaciones. El argumento **Thin** define la reducción del intervalo. Finalmente `print_plot = TRUE` permite la creación de herramientas gráficas para ayudar al usuario a evaluar si se ha logrado la convergencia de **MCMC**.

4.3.6. Estadística descriptiva

La función **HSROCSummary** devuelve un resumen de los resultados de la interfaz de usuario de R y más detalles en un archivo de texto en el directorio de trabajo. La interfaz de usuario de R enumera las estimaciones puntuales y la densidad más alta a posteriori con su respectivo intervalo de confianza del 95% para los parámetros inter-estudios e intra-estudios. El archivo de texto se guarda en el directorio de trabajo y se divide en tres secciones.

4. ANÁLISIS DE RESULTADOS

La primera sección tiene configuraciones y ajustes generales. El número de Cadenas de Markov es de 1 en este ejemplo. Un total de 50.000 iteraciones se utilizaron con un punto de quiebre de 10.000 dentro de cada cadena. Cada valor por segundo es guardado para las estimaciones (la disminución del intervalo). Esto nos deja con una muestra de 20.000 valores para extraer la distribución a posteriori de cada parámetro que interesa. Esta sección también muestra la ubicación del archivo en el disco duro, la fecha de creación del resumen a presentar y la naturaleza de la prueba de referencia (perfecta o imperfecta). El conjunto de datos también está en la lista, vea Figura 4.27.

```

Summary for n = 20000
Number of chains = 1
Number of iteration within a chain = 50000 burn in within each chain = 10000
Thinning interval = 2
Total Number of Iteration kept = 20000
File Location : C:/users/usuario/documents
Date : 2018-06-06 01:23:10
Perfect reference standard
-----
SAMPLE SIZE
-----
Study  Total  +-+--+--+
Study 1  57  10  2  44
Study 2  46  1  0  32
Study 3  22  0  1  16
Study 4  60  1  10  44
Study 5  108  2  10  107
Study 6  289  12  4  270
Study 7  272  21  1  250
Study 8  71  8  0  53
Study 9  42  16  2  22
    
```

Figura 4.27: Lista de configuraciones generales de la simulación

La siguiente sección de salida proporciona un resumen de las distribuciones anteriores utilizadas en este ejemplo. La distribución a priori utilizada para la prevalencia es una distribución Beta (1,1), esta es equivalente a una uniforme a priori de la prevalencia, permitiendo a todos los valores posibles de ser igualmente probables. El logaritmo de la proporción entre las dos desviaciones estándar, β se supone que sigue una distribución U (-0,75, 0,75). La media general de corte Θ sigue una distribución uniforme desde -1,5 hasta 1,5. La precisión del diagnóstico combinado Λ sigue una distribución uniforme de -3 a 3. Finalmente, los parámetros σ_α y σ_θ se supone que ambas siguen una distribución U(0, 2), vea Figura 4.28.

Finalmente, en esta sección se enumeran los estadísticos descriptivos de los parámetros del modelo. Adicional a la estimación puntual e intervalos más altos de la densidad a posteriori (HPD), también incluye el error y la desviación estándar de la muestra a posteriori en la simulación de Monte Carlo (MC). El error MC se calcula por el método de medias por lote descrito por Ntzoufras (138). Por esta razón, el usuario puede observar que no se entregará ninguna estimación si es inferior a 100 iteraciones, teniendo en cuenta la condición de parada y afinamiento del intervalo, generando un mensaje de error, si se emplea menos de 100 iteraciones, `HSROCSummary(data = MRI, burn_in = 1, Thin = 1, i = 50, print_plot = T)`.

```

PRIOR INFORMATION
-----
Prior of prevalence (pi) is beta ( 1 , 1 ), <<> pi in [ 0 , 1 ]
Prior of beta is uniform( -0.75 , 0.75 )
Prior of THETA is uniform( -1.5 , 1.5 )
Prior of LAMBDA is uniform( -3 , 3 )
Prior of sigma_alpha is uniform( 0 , 2 )
Prior of sigma_theta is uniform( 0 , 2 )
    
```

Figura 4.28: Resumen de las distribuciones utilizadas

```
> Error in HSROCSummary(data = MRI, burn_in = 1, Thin = 1, i = 50, print_plot = T):  
You don't have enough iterations to estimate the MC error.  
After taking into account the "burn in" and "thinning interval",  
you need at least 100 iterations to proceed.
```

4. ANÁLISIS DE RESULTADOS

La suma del error MC (Monte Carlo) y el error estándar proporcionan medios para examinar la precisión de la estimación. Uno podría querer ejecutar la simulación MCMC hasta que el error MC sea suficientemente pequeño. Por ejemplo, un criterio deseable podría ser tener un error MC para cada parámetro menor al 10% de su desviación estándar posterior. Las estimaciones de los parámetros intra-estudios e inter-estudio se observan en las Figuras 4.29, 4.30, mayor detalle note **Apéndice A.1**.

BETWEEN-STUDY parameters (Point estimate = median)					
	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
THETA	0.928372	0.228954	0.011726	0.469583	1.365921
LAMBDA	2.106404	0.340238	0.039787	1.420487	2.752459
beta	-0.415436	0.264169	0.021773	-0.749537	0.120237
sigma.alpha	0.155597	0.296133	0.006616	5.ee-05	1.090918
sigma.theta	0.512763	0.188981	0.00337	0.230685	0.921319
S overall	0.55776	0.102648	0.00121	0.349044	0.752855
C overall	0.947616	0.023182	0.000312	0.890218	0.981592

Figura 4.29: Resumen de las distribuciones utilizadas

Por ejemplo, el punto de corte para el parámetro Θ fue estimado en 0.928372 y un intervalo de (0.469583, 1.365921) al 95% de confianza para el valor más alto de densidad-posterior. La desviación estándar y el error MC fue

WITHIN-STUDY PARAMETERS					
	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
theta					
Study 1	0.717434	0.269566	0.013283	0.200333	1.252638
Study 2	0.761313	0.263249	0.013039	0.24062	1.267722
Study 3	0.628859	0.322931	0.011389	-0.032148	1.273068
Study 4	1.479933	0.276003	0.012005	0.915156	2.002552
Study 5	1.318358	0.34507	0.01275	0.678148	2.033561
Study 6	1.580732	0.213855	0.012518	1.158249	1.998937
Study 7	0.523696	0.252035	0.013539	0.02064	1.000077
Study 8	1.084833	0.191897	0.013538	0.704547	1.453734
Study 9	0.749527	0.23466	0.013352	0.274661	1.19451
Study 10	0.436283	0.279488	0.0128	-0.120885	0.979411
alpha					
Study 1	2.46736	0.461626	0.01854	1.605338	3.413924
Study 2	1.546263	0.507919	0.018999	0.527815	2.492834
Study 3	2.091117	0.502131	0.018307	1.06714	3.070978
Study 4	1.924161	0.498969	0.025052	0.869954	2.825598
Study 5	1.767675	0.578755	0.023734	0.506963	2.73959
Study 6	2.124205	0.432142	0.027618	1.245024	2.956996
Study 7	2.07992	0.395889	0.016659	1.265104	2.815339
Study 8	2.55992	0.364965	0.021331	1.806921	3.225344
Study 9	2.028452	0.392633	0.017885	1.257335	2.793918
Study 10	2.430977	0.445813	0.016628	1.60232	3.348222
Prevalence					
Study 1	0.200152	0.051609	0.000367	0.106957	0.306094
Study 2	0.184025	0.056113	0.00042	0.085939	0.300807
Study 3	0.200346	0.051228	0.000601	0.059669	0.363462
Study 4	0.255629	0.055353	0.000375	0.150532	0.364656
Study 5	0.303807	0.089419	0.000619	0.137763	0.481396
Study 6	0.148772	0.025197	0.00016	0.100658	0.198386
Study 7	0.332415	0.065783	0.000413	0.207833	0.462481
Study 8	0.137648	0.02098	0.00014	0.099626	0.18145
Study 9	0.188975	0.045677	0.000336	0.103775	0.279726
Study 10	0.431199	0.07431	0.000587	0.286243	0.574721

Figura 4.30: Resumen de las estimaciones de los parámetros intra-estudio

estimado en 0.228954 y 0.011726 respectivamente. Vemos que el error MC es menor que el 10% de la desviación estándar posterior para Θ . Otros cálculos se incluyen como los de Λ , β , σ_θ y σ_α . Las estimaciones de la sensibilidad y la especificidad agrupada, denotan S en general y C general, y se obtienen como funciones de Θ , Λ y β . Un análisis similar de los datos de las imágenes de MR fue realizado por Rutter and Gatsonis (158) utilizando un modelo HSROC con una función de enlace logit. Reportaron una estimación de la sensibilidad agrupada de 0.541 con un intervalo de (0.286, 0.771) al 95% de confianza con igualdad de colas y una estimación de la especificidad agrupada de 0.953 con un intervalo de

(0.907 , 0.981) al 95 % de confianza con igualdad de colas. A pesar de que ambos modelos utilizan una función de enlace diferente, es alentador notar que las estimaciones de S General y C general son satisfactoriamente cercanas a las estimaciones de Rutter y Gatsonis. La diferencia se nota en los intervalos de confianza, dada por el hecho de que la función HSROCSummary retorna un intervalo al 95 % de confianza para el valor más alto de densidad posterior. La Figura 4.30 muestra una parte de las estimaciones de los parámetros inter-estudios. Se nota que las estimaciones para estudios individuales se agrupan por parámetros. Por ejemplo, las estimaciones de $\theta_1, \theta_2, \theta_3, \dots, \theta_{10}$ (punto de corte en el estudio 1, 2, 3, $\dots, 10$), se muestran en la primera sub-sección de los parámetros intra-estudios.

4.3.7. Resumen gráfico

La función HSROCSummary crea varios gráficos para que el usuario pueda juzgar si la estadística descriptiva es fiable. Entre ellos, una gráfica de seguimiento para cada parámetro, puede ser utilizada para evaluar la convergencia del muestreo. Cada tramo de la gráfica es un gráfico de dispersión de la muestra posteriori de un solo parámetro versus el número de la iteración de la simulación. Las gráficas de sensibilidad de MR en los estudios 1, 2, 3, 6, 7 y 8 se muestra en la Figura 4.31, para mayor detalle observe **Apéndice A.2**. Para el ejemplo, notamos que la convergencia se logra con bastante rapidez. Otro resumen gráfico producido por la función HSROCSummary es el diagrama de densidad. En él se traza una gráfica suavizada de la función de densidad a posteriori para cada parámetro, Figura 4.32, mayor detalle vea **Apéndice A.3**. La función también produce una curva (SROC). La curva SROC resume la relación entre sensibilidad y (1 - especificidad) en los estudios, teniendo en cuenta la heterogeneidad inter-estudios. La curva SROC para los datos se muestra en la Figura 4.33, los estudios individuales se representan por un círculo claro. El radio del círculo es proporcional al tamaño de la muestra del estudio. El círculo rojo marca la combinación de la sensibilidad y especificidad a través de los 10 estudios en este meta-análisis. La predicción al 95 % de la región se define por el punteado de la curva azul. El trazo y puntos de la curva de color rojo marcan el límite de la región creíble 95 % para las estimaciones agrupadas de sensibilidad y especificidad a través de los 10 estudios.

4. ANÁLISIS DE RESULTADOS

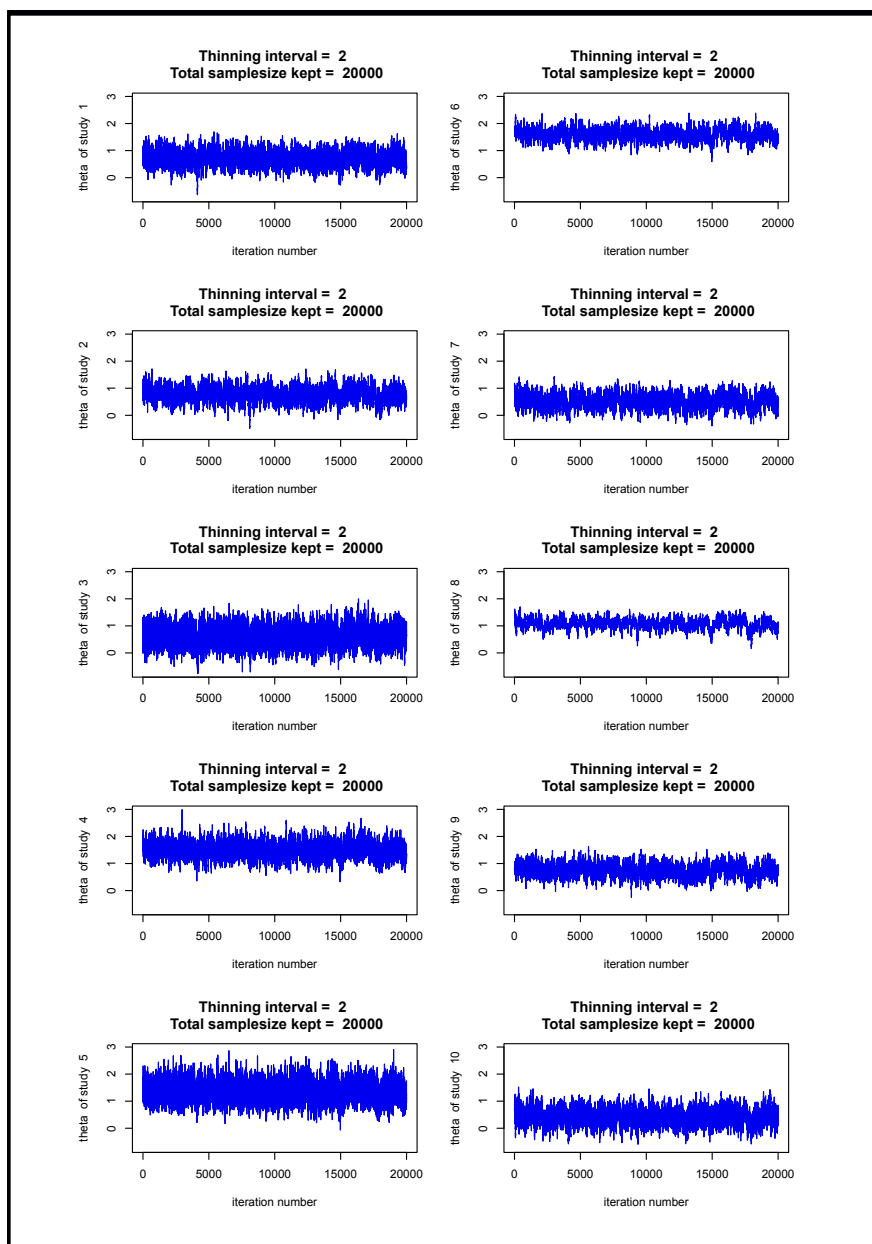


Figura 4.31: Gráfico de dispersión, de la distribución muestral a posteriori de un sólo parámetro, datos MRI

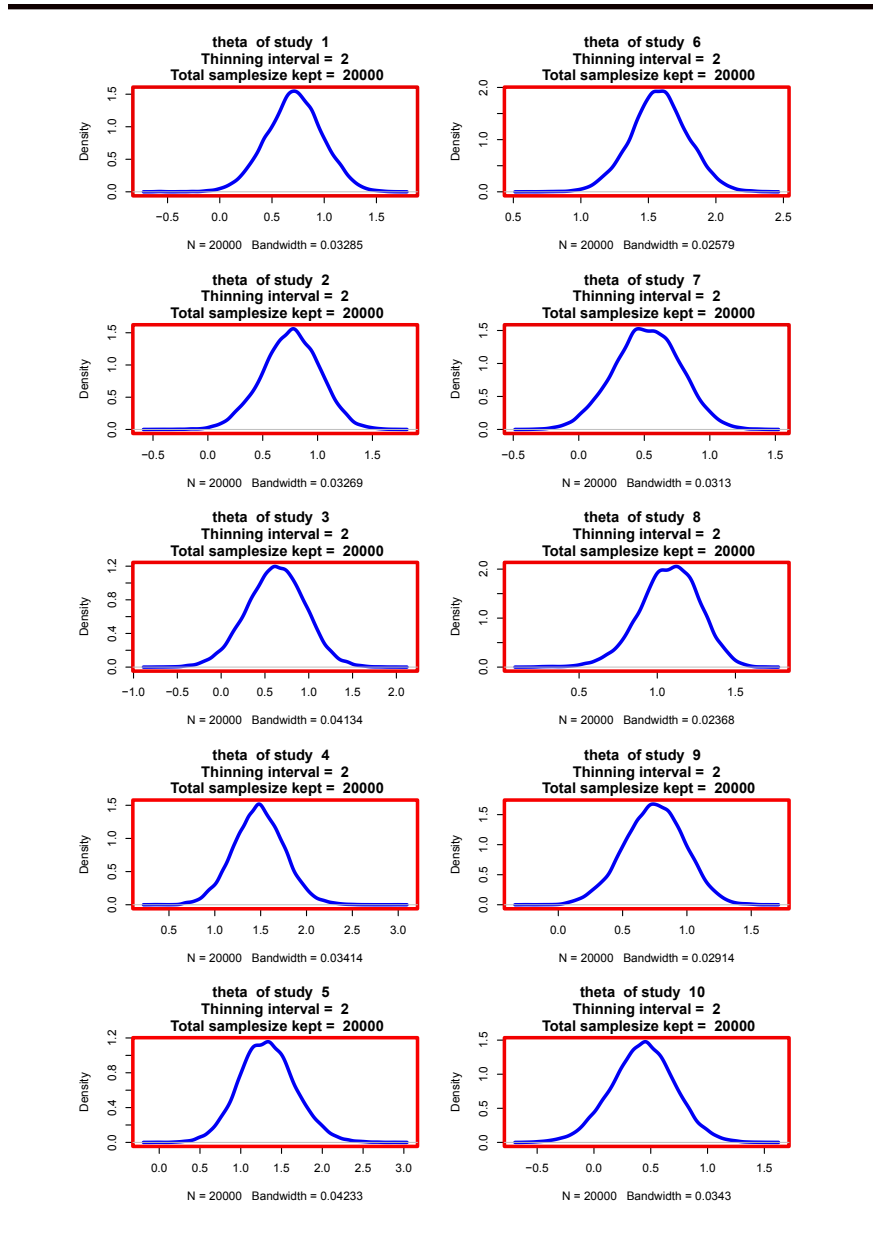


Figura 4.32: Gráficos de densidad de algunos de los parámetros inter-estudios, datos MRI.

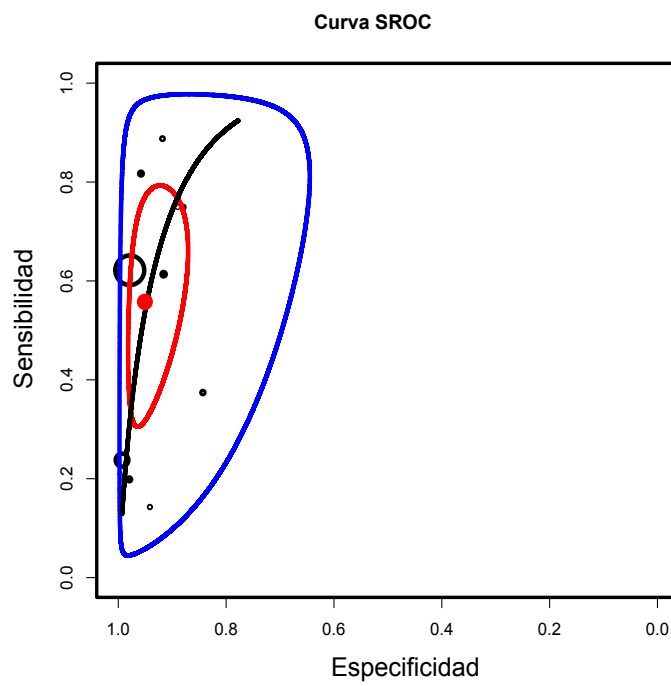


Figura 4.33: Curva SROC, datos MRI

4.4. Simulación de datos derivados de un modelo HSROC suponiendo independencia condicional entre el índice y las pruebas de referencia.

El paquete *HSROC* contiene una función que permite la simulación de un conjunto de datos provenientes de meta-análisis de diagnóstico. Los argumentos de la función son:

```
> args(simdata)

function (N, n, n.random = FALSE, sub_rs = NULL, prev, se_ref = NULL,
sp_ref = NULL, T, range.T = c(-Inf, Inf), L, range.L = c(-Inf,
Inf), sd_t, sd_a, b, path = getwd())
NULL
```

El usuario debe especificar los valores de los parámetros entre estudios $(\Theta, \Lambda, \beta, \sigma)$, el número de estudios deseados y el número de individuos dentro de cada estudio. Además, la sensibilidad y la especificidad de cada prueba de referencia deben especificarse en el caso de pruebas estándar que no sean de oro. Supongamos que queremos generar un conjunto de datos de 6 estudios con 20 individuos dentro de cada estudio y con los siguientes parámetros:

```
> beta = 0.25
> LAMBDA = 2.5
> sd_alpha = 0.75 ;
> THETA = -0.2
> sd_theta = 0.50 ;
> pi = c(0.15, 0.25, 0.10, 0.12, 0.22, 0.18)
```

La siguiente llamada generaría los datos deseados bajo la suposición de que se utiliza una prueba de referencia perfecta en cada estudio.

```
> sim.data$Data
Los resultados se ven como:
```

4. ANÁLISIS DE RESULTADOS

```
++ +- -+ --
[1,] 1 0 1 18
[2,] 8 6 0 6
[3,] 1 8 0 11
[4,] 1 1 0 18
[5,] 5 0 0 15
[6,] 5 11 0 4
```

```
> sim.data$`Whithin study parameters`
```

```
alpha      theta      ++      -- prev
[1,] 3.951347 -0.28563782 0.9770113 0.9722577 0.15
[2,] 1.720171 0.56546250 0.6025696 0.9468833 0.25
[3,] 1.881614 0.19450001 0.7449280 0.9008607 0.10
[4,] 2.930092 0.51383079 0.7993901 0.9875310 0.12
[5,] 3.655323 -0.44238565 0.9774280 0.9417603 0.22
[6,] 3.740089 0.06045237 0.9448626 0.9856484 0.18
```

```
> sim.data$`Between study parameters`
```

```
THETA sigma theta      LAMBDA sigma alpha      beta Overall ++
-0.2000000 0.5000000 2.5000000 0.7500000 0.2500000 0.8996607
Overall --
0.8829387
```

```
> sim.data$`Reference standard`
```

```
[1] "PERFECT"
```

El argumento n se puede modificar, es decir, $n = c(20, 50, 25, 35, 105, 15)$ por lo que cada estudio tiene un tamaño de muestra único proporcionado por el usuario, lo que da como resultado los datos a continuación:

```
> sim.data = simdata(N=6, n=c(20,50,25,35,105,15) , prev=pi, T=THETA,
+                    L=LAMBDA,sd_t=sd_theta, sd_a=sd_alpha, b=beta)
> sim.data$Data
```

4.4 Simulación de datos derivados de un modelo HSROC suponiendo independencia condicional entre el índice y las pruebas de referencia.

```
++ +- -+ --
[1,]  1  4  1 14
[2,] 15 19  3 13
[3,]  3  2  0 20
[4,]  8 23  0  4
[5,] 20  1  2 82
[6,]  2  7  0  6
```

Alternativamente, se puede seleccionar un tamaño de muestra único al azar dentro de un rango como sigue:

```
> n=sample(seq(50,250,1),6,replace=F)
> sim.data$Data
```

```
++ +- -+ --
[1,] 26  5  1 137
[2,] 67 38  7 136
[3,] 23 18  4 178
[4,] 21 33  6 140
[5,] 30  3  8 116
[6,] 17  6  0  62
```

En cada caso descrito anteriormente, la función también creará 3 archivos en el directorio especificado por el argumento de la ruta que resume el conjunto de datos. Esos archivos se pueden usar dentro de la función *HSROCSummary* para comparar las estimaciones de parámetros con sus valores reales. Se puede obtener ayuda adicional sobre esta función escribiendo:

```
> help(simdata)
```

Para simular un conjunto de datos para un modelo *HSROC* con referencias estándar imperfectos, debemos proporcionar los valores de la sensibilidad y especificidad de cada prueba de referencia. Supongamos que queremos generar datos para 6 estudios con 2 referencias estándar diferentes donde la primera prueba de referencia se aplicará en los primeros 4 estudios, mientras que la otra prueba se aplicará en los 2 estudios restantes.

4. ANÁLISIS DE RESULTADOS

```
> REFSTD = list(2, 1:4, 5:6)
> s2 = c(0.6, 0.75)
> c2 = c(0.95, 0.7)
```

En el ejemplo anterior, la sensibilidad y la especificidad de la primera prueba de referencia son 60 % y 95 % respectivamente. La sensibilidad y la especificidad de la segunda prueba de referencia son del 75 % y 70 %, respectivamente. Para obtener un conjunto de datos de 6 estudios con 200 individuos dentro de cada estudio suponemos los mismos parámetros entre estudios que en la sección 4.4, obteniendo lo siguiente:

```
> simdata(N=6, n=200, sub_rs=REFSTD, se_ref=s2, sp_ref=c2, prev=pi,
+ T=THETA, L=LAMBDA, sd_t=sd_theta, sd_a=sd_alpha, b=beta)
```

```
$Data
```

```
++ +- -- --
[1,] 20 99 3 78
[2,] 32 25 11 132
[3,] 13 56 7 124
[4,] 13 5 11 171
[5,] 40 14 45 101
[6,] 32 45 38 85
```

```
$`Whithin study parameters`
```

```
alpha      theta      ++      -- prev
[1,] 1.950865 -0.9478363 0.9551780 0.5124732 0.15
[2,] 3.400859 -0.4647693 0.9719836 0.9192712 0.25
[3,] 2.430639 -0.6796673 0.9527694 0.7280657 0.10
[4,] 3.671882 0.1273654 0.9341987 0.9869499 0.12
[5,] 3.261340 -0.2786336 0.9540013 0.9372460 0.22
[6,] 2.338588 -0.6737186 0.9480733 0.7127922 0.18
```

```
$`Between study parameters`
```

```
THETA sigma theta      LAMBDA sigma alpha      beta Overall ++
-0.2000000 0.5000000 2.5000000 0.7500000 0.2500000 0.8996607
```


4.4 Simulación de datos derivados de un modelo HSROC suponiendo independencia condicional entre el índice y las pruebas de referencia.

Overall --

0.8829387

\$`Reference standard`

1 2

s2 0.60 0.75

c2 0.95 0.70

4.5. Precisión diagnóstica mediante el paquete CopulaDTA para Meta-Análisis

4.5.1. Marco de referencia y software

En el marco bayesiano, el analista actualiza una opinión (“información previa”) de un parámetro basado en los datos observados, mientras que en el marco frecuentista, el analista investiga el comportamiento de los parámetros estimados en muestras hipotéticas repetidas de una determinada población. Debido a su flexibilidad y el uso de simulaciones **MCMC**, la modelización más compleja a menudo se puede implementar más fácilmente dentro del marco bayesiano. Mediante la manipulación de las distribuciones a priori, la inferencia bayesiana puede eludir los problemas de identificación, mientras que los algoritmos de aproximación numérica en la inferencia frecuentista sin distribuciones previas pueden quedar atascados por problemas de identificación.

Sin embargo, los métodos bayesianos suelen requerir experiencia estadística y paciencia porque las simulaciones MCMC son computacionalmente intensivas. Por el contrario, la mayoría de los métodos frecuentistas han sido envueltos en “procedimientos” estándar que requieren menos conocimientos estadísticos y habilidades de programación. Además, los métodos frecuentista son optimizados con estimación de máxima verosimilitud (MLE) que tienen tiempos de ejecución mucho más cortos en comparación con las simulaciones MCMC.

4.5.2. El paquete CopulaDTA

El paquete **CopulaDTA** es una extensión de *rstan* que es la interfaz de *R* a Stan, Carpenter et al. (16) para los datos de precisión de la prueba de diagnóstico. *Stan* es un lenguaje de programación probabilístico que ha implementado Hamilton Monte Carlo (MHC), Hoffman and Gelman (86). El paquete facilita la aplicación de modelos complejos y su visualización dentro del marco bayesiano con tiempos de ejecución mucho más cortos.

JAGS, es un motor de muestreo de propósito general extensible alternativo a Stan, Hornik et al. (90). Entender “JAGS” requiere conocimientos de *C++* para ensamblar un módulo de

biblioteca de vínculos dinámicos (DLL). Por experiencia, configurar y construir el módulo es una tarea desalentadora y tediosa, especialmente en el sistema operativo de Windows. Los anteriores inconvenientes, junto con el hecho de que Stan tiende a converger con menos iteraciones inclusive con valores iniciales pocos confiables en relación con JAGS, lo cual, hace preferible el motor de muestreo **Stan MCMC**.

El paquete CopulaDTA está disponible a través de la Red *CRAN* en `http://CRAN.R-project.org/package=CopulaDTA`. Con una conexión a Internet activa, el paquete CopulaDTA se instala y carga en *R* con los siguientes comandos:

```
> install.packages("CopulaDTA", dependencies = TRUE)
> library(CopulaDTA)
```

El paquete **CopulaDTA** proporciona funciones de distribuciones bivariantes *beta-binomiales* construidas como un producto de dos distribuciones marginales *beta* y densidades de cópula discutidas en el Capítulo 3. El paquete también proporciona **forest plot** para un modelo con covariables categóricas o sólo con intercepción. Dada la función de cópula elegida, una distribución beta-binomial es empleada por la función **cdtamodel** que devuelve un objeto **cdtamodel**. El ajuste de función principal toma el objeto **cdtamodel** y fija el modelo al conjunto de datos dado y devuelve un objeto **cdtafit** para el cual se proporcionan gráficos y medidas resumen.

4.5.3. Diagnóstico del modelo

Para evaluar la convergencia del modelo, la mezcla y la estacionalidad de las cadenas, es necesario comprobar el factor de reducción de escala potencial \hat{R} , el tamaño de muestra efectivo (ESS), el error de MCMC y graficar los parámetros. Cuando todas las cadenas alcanzan la distribución posteriori objetivo, se espera que la varianza posteriori estimada esté próxima a la varianza dentro de la cadena de tal manera que la relación sea cercana a 1 indicando que las cadenas son estables y han alcanzado la distribución objetivo. Un gran \hat{R} indica mala mezcla y que se necesitan más iteraciones. El tamaño efectivo de la muestra indica cuánta información se tiene realmente sobre un determinado parámetro. Cuando las muestras se correlacionan automáticamente, se espera menos información de la distribución posteriori de nuestros parámetros

4. ANÁLISIS DE RESULTADOS

que si las muestras fueran independientes. El ESS cercano al total de iteraciones posteriores al calentamiento es una indicación de una menor autocorrelación y una buena mezcla de las cadenas. Las simulaciones con ESS más alto tienen errores estándar más bajos y estimaciones más estables. Dado que la distribución posterior es simulada, existe la posibilidad de que la aproximación esté desactivada en alguna cantidad; el error de Monte Carlo (MCMC) cercano a 0 indica que es probable que se haya alcanzado la distribución objetivo.

4.5.4. Comparación y selección de modelos

Para comparar los modelos se utilizará el Criterio de Información de **Watanabe-Alkaike (WAIC)** Watanabe (198), una herramienta de comparación de modelos reciente para medir la precisión predictiva de los modelos en el marco bayesiano. “WAIC” puede ser visto como una mejora del Criterio de Información de Desviación (DIC) que, aunque popular, se sabe que tiene algunos problemas. WAIC es una herramienta totalmente bayesiana, que se aproxima mucho a la validación cruzada bayesiana, es invariante a la reparametrización y puede usarse tanto para modelos simples como jerárquicos y de mezcla.

4.5.5. Dataset

Telomerase data

Glas et al. (65), revisaron sistemáticamente la sensibilidad y especificidad de la citología y otros marcadores incluyendo la **telomerasa** para el diagnóstico primario del cáncer de vejiga. Presentaron una distribución normal bivalente a la transformación logit de la sensibilidad y especificidad de cada estudio permitiendo la heterogeneidad entre los estudios. De los 10 estudios incluidos, informaron que la telomerasa tenía una sensibilidad y una especificidad de 0.75 [0.66 0.74] y 0.86 [0.71 0.94], respectivamente. Concluyeron que la telomerasa no era lo suficientemente sensible como para ser recomendada para el uso diario. Este conjunto de datos está disponible dentro del paquete mediante los siguientes comandos.

```
> data(telomerase)
> telomerase
```

ID	VP	VN	FN	FP
1	25	25	8	1
2	17	11	4	3
3	88	31	16	16
4	16	80	10	3
5	40	137	17	1
6	38	24	9	6
7	23	12	19	0
8	27	18	6	2
9	14	29	3	3
10	37	7	7	22

Donde, ID es el identificador del estudio, TP es el número de verdaderos positivos, TN es el número de verdaderos negativos, FN representa a los falsos negativos y FP es el número de falsos positivos.

ASCUS triage data

Arbyn et al. (3), realizaron una revisión Cochrane sobre la exactitud de la prueba del virus del papiloma humano y la repetición de la citología para el triaje de las mujeres con un frotis de Papanicolaou equívoco para diagnosticar precáncer de cuello uterino. Ellos usaron el modelo BRMA (bivariante) en SAS usando METADAS en 10 estudios donde emplearon ambas pruebas. Se informó de una sensibilidad absoluta de 0.909[0.857 0.944] y 0.715[0.629 0.788] para HC2 y la citología, respectivamente. La especificidad fue de 0,607[0.539 0.68] y 0.684[0.599 0.758] para HC2 y la citología de repetición, respectivamente. Estos datos se utilizan para demostrar cómo el modelo de intercepción sólo se extiende en una configuración de meta-regresión. Este conjunto de datos también está disponible dentro del paquete mediante los siguientes comandos.

```
> data(ascus)
> ascus
```

Carga los datos en el entorno R y genera la siguiente salida, donde $Test$ es una variable explicativa que muestra el tipo de prueba de triaje, $StudyID$ es el identificador del estudio, TP es

4. ANÁLISIS DE RESULTADOS

el número de verdaderos positivos, FP es el número de falsos positivos, TN es el número de verdaderos negativos, FN es el número de falsos negativos.

	Test	StudyID	VP	FP	VN	FN
1	RepC	Andersson 2005	6	14	28	4
2	RepC	Bergeron 2000	8	28	71	4
3	RepC	Del Mistro 2010	20	191	483	7
4	RepC	Kulasingham 2002	20	74	170	6
5	RepC	Lytwyn 2000	4	20	26	2
6	RepC	Manos 1999	48	324	570	15
7	RepC	Monsonogo 2008	10	18	168	15
8	RepC	Morin 2001	14	126	214	5
9	RepC	Silverloo 2009	24	43	105	10
10	RepC	Solomon 2001	227	1132	914	40
11	HC2	Andersson 2005	6	17	25	4
12	HC2	Bergeron 2000	10	38	61	2
13	HC2	Del Mistro 2010	27	154	566	2
14	HC2	Kulasingham 2002	23	115	129	3
15	HC2	Lytwyn 2000	4	19	33	1
16	HC2	Manos 1999	58	326	582	7
17	HC2	Monsonogo 2008	22	110	72	2
18	HC2	Morin 2001	17	88	253	2
19	HC2	Silverloo 2009	34	65	81	2
20	HC2	Solomon 2001	256	1050	984	11

4.5.6. Modelización

El paquete *CopulaDTA* modeliza diferentes estructuras de correlación que resultan en cinco diferentes distribuciones beta-binomiales bivariantes . La estructura de correlación se especifica indicando $copula \sim \sim$ “*gauss*” o, “*fgm*”, “*c90*”, “*c270*”, “*frank*” en la función *fitcopula*.

```
> gauss.1 <- cdtamodel("gauss")
```

El código para ajustar el modelo y las opciones del modelo se muestran de la siguiente manera:

```
> print(gauss.1)
> str(gauss.1)
```

La distribución beta-binomial bivalente de la cópula gaussiana se ajusta a los datos de la telomerasa con el siguiente código:

```
> fitgauss.1 <- fit(
+   gauss.1,
+   data = telomerase,
+   SID = "ID",
+   iter = 10000,
+   warmup = 1000,
+   thin = 30,
+   seed = 3)
```

De forma predeterminada, se realizan 3 cadenas con 3 núcleos **y no es necesario especificarlas a menos que se indique lo contrario**. A partir del código anterior, se extraen 28000 muestras de cada una de las 3 cadenas, se descartan las primeras 1000 muestras y después cada 30 se mantienen tal que cada cadena tiene 900 post-warm-up calados posteriores al calentamiento, lo que hace un total de 2700 post-warm-up. El valor de la semilla, *seed* = 3, especifica un generador de números aleatorios para permitir la reproducibilidad de los resultados.

Los gráficos siguientes muestran una mezcla satisfactoria de las cadenas y la convergencia.

```
> traceplot(fitgauss.1)
```

4. ANÁLISIS DE RESULTADOS

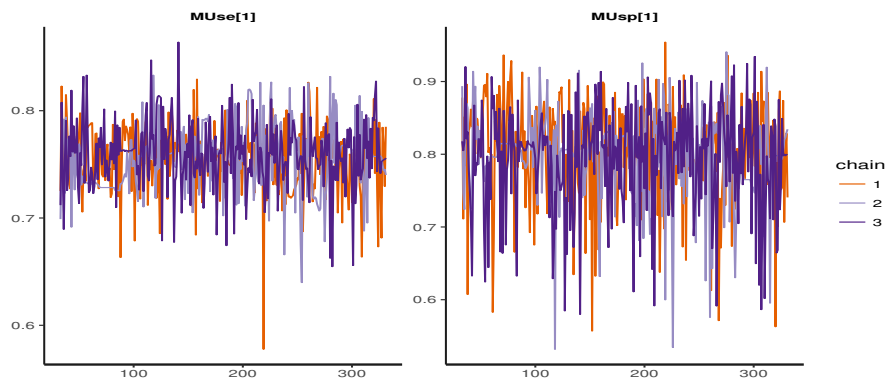


Figura 4.34: Convergencia de la Mezcla según cadenas, Cópula Gauss

A continuación, se obtienen las estimaciones resumen del modelo de la siguiente manera.

```
> print(fitgauss.1, digits = 4)
```

Posterior marginal mean and median sensitivity and specificity
with 95% credible intervals

Parameter	Mean	Lower	Median	Upper	n_eff	Rhat
MUse[1] Sensitivity	0.756404	0.6904965	0.757350	0.81593	572.6	1.006
MUsp[1] Specificity	0.798385	0.6310427	0.806124	0.90634	697.8	1.005
ktau[1] Correlation	-0.858247	-0.9857824	-0.916157	-0.42862	294.9	1.006
Varse[1] Var(Sens)	0.006524	0.0007091	0.005446	0.01939	477.2	1.010
Varsp[1] Var(Spec)	0.048764	0.0150084	0.044097	0.11080	395.1	1.007

Model characteristics

Copula function: gauss, sampling algorithm: NUTS(diag_e)

Formula(1): MUse ~ 1

Formula(2): MUsp ~ 1

Formula(3): Omega ~ 1


```
3 chain(s)each with iter=10000; warm-up=1000; thin=30.  
post-warmup draws per chain=300;total post-warmup draws=900.
```

Predictive accuracy of the model

```
Log point-wise predictive density (LPPD): -38.0313  
Effective number of parameters: 7.1842  
Watanabe-Akaike information Criterion (WAIC): 90.4310
```

A partir de la salida anterior, n_{eff} y $Rhat$ confirman una correcta mezcla de las cadenas con poca autocorrelación. La sensibilidad meta-analítica $MU_{se}[1]$ y la especificidad $MU_{sp}[1]$ son 0.7568[0.6904 0.8166] y 0.7983[0.6171 0.9064] respectivamente. La variabilidad entre los estudios con relación a sensibilidad y especificidad se estimaron con los valores 0.0062[0 0.0195] y 0.0048[0.0136 0.1220], respectivamente. La correlación tau de Kendall entre sensibilidad y especificidad se estima en $-0.8202[-0.9861 - 0.3333]$.

El siguiente comando produce una serie de gráficos forest plot.

```
> plot(fitgauss.1)
```

\$G1

\$G2

\$G3

4. ANÁLISIS DE RESULTADOS

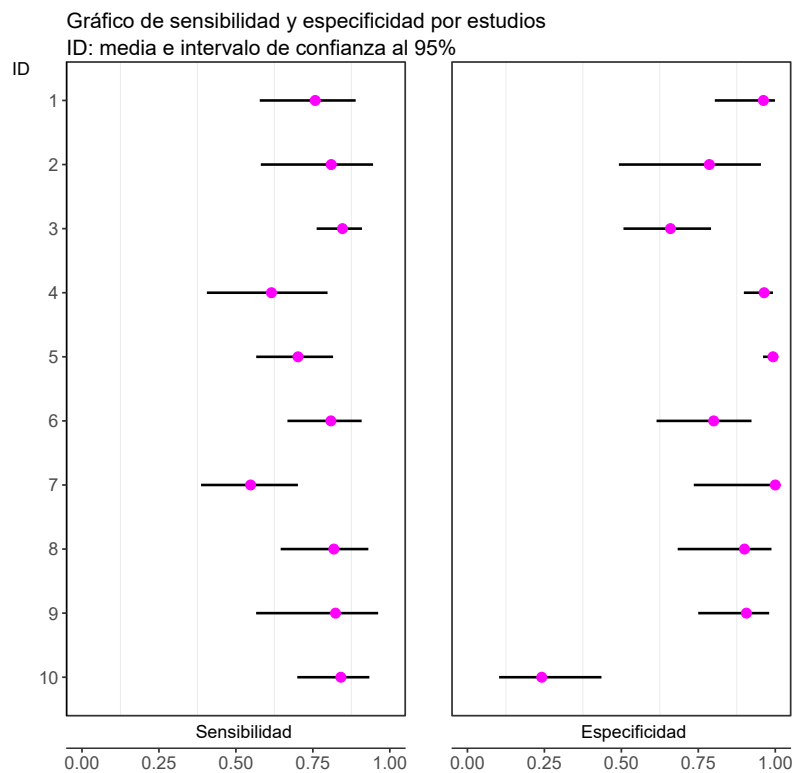


Figura 4.35: Forest Plot de sensibilidad y especificidad a priori por estudios con intervalos al 95 % de confianza

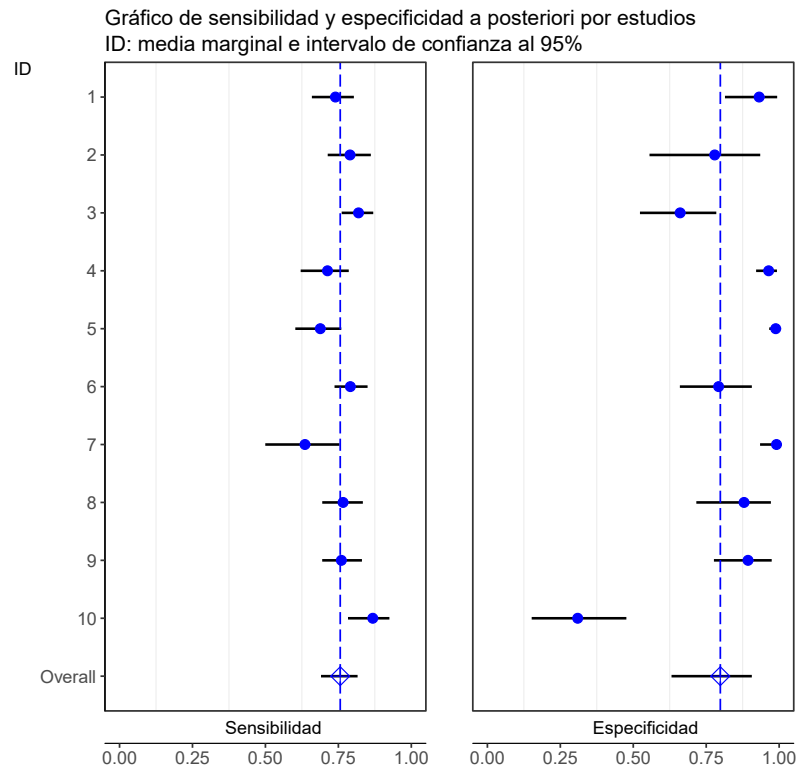


Figura 4.36: Forest Plot de sensibilidad y especificidad a posteriori por estudios con intervalos al 95 % de confianza

4. ANÁLISIS DE RESULTADOS

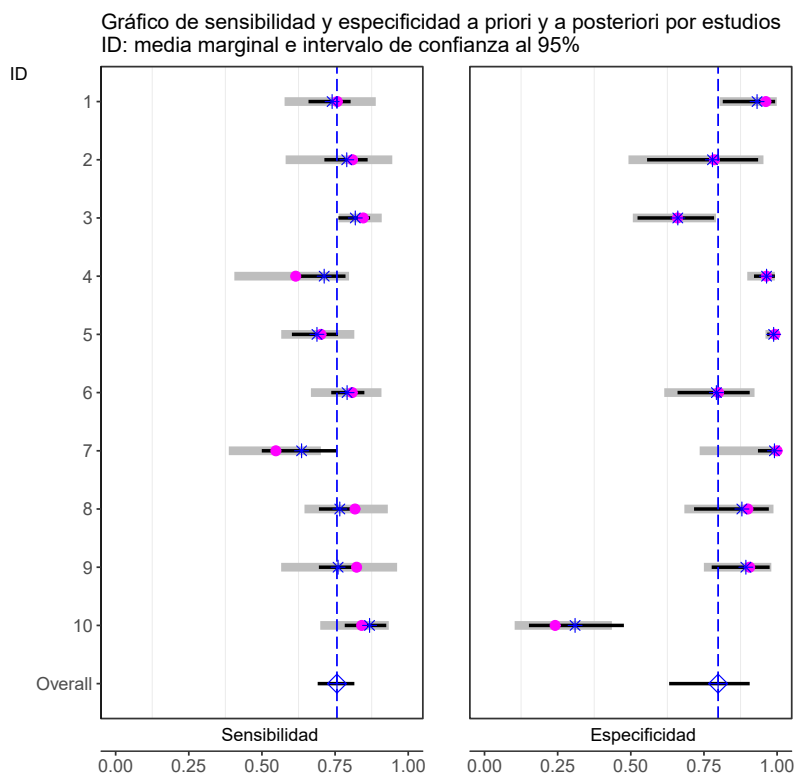


Figura 4.37: Forest Plot de sensibilidad y especificidad a priori y a posteriori por estudios con intervalos al 95 % de confianza

§G1, Figura 4.35 es un gráfico de sensibilidad y especificidad por cada estudio (puntos color magenta) con sus correspondientes intervalos de confianza al 95 % (líneas negras). §G2, Figura 4.36 es un gráfico para la sensibilidad y especificidad a posteriori de los estudios (puntos color azul) y con un valor único para las distribuciones marginales “(diamante azul)” de la sensibilidad y especificidad con sus correspondientes intervalos de confianza del 95 % (líneas color negras). §G3, Figura 4.37 es un gráfico de sensibilidad y especificidad a posteriori de cada estudio (estrellas azules) y marginal (diamante azul) con sus correspondientes intervalos de confianza al 95 % (líneas negras). También se presentan la sensibilidad y especificidad específicas del estudio (puntos magenta) y sus correspondientes intervalos de confianza exactos al 95 % (líneas grises gruesas).

Como se observa en las gráficas 4.35, 4.36 y 4.37, la sensibilidad y especificidad a posteriori

son menos extremas y variables que las a priori. En otras palabras, existe una “contracción” hacia la sensibilidad y especificidad (media global), ya que las estimaciones posteriores específicas del estudio dependen de la estimación global y, por lo tanto, también de los demás estudios.

Las otras cuatro distribuciones beta bivariantes basadas en cópula se ajustan de la siguiente manera:

```
> c90.1 <- cdtamodel(copula = "c90")
> fitc90.1 <- fit(c90.1,
+               data=telomerase,
+               SID="ID",
+               iter=10000,
+               warmup=1000,
+               thin=30,
+               seed=718117096)

> c270.1 <- cdtamodel(copula = "c270")
> fitc270.1 <- fit(c270.1,
+               data=telomerase,
+               SID="ID",
+               iter=10000,
+               warmup=1000,
+               thin=20,
+               seed=3)

> fgm.1 <- cdtamodel(copula = "fgm")
> fitfgm.1 <- fit(fgm.1,
+               data=telomerase,
+               SID="ID",
+               iter=10000,
+               warmup=1000,
+               thin=20,
+               seed=3)
```

4. ANÁLISIS DE RESULTADOS

```
> frank.1 <- cdtamodel(copula = "frank")
> fitfrank.1 <- fit(frank.1,
+                 data=telomerase,
+                 SID="ID",
+                 iter=10000,
+                 warmup=1000,
+                 thin=20,
+                 seed=1959300748)
```

Para fines de comparación, el modelo recomendado actual; el BRMA (bivariante), que usa marginales normales también se ajusta a los datos, aunque no es parte del paquete **CopulaDTA**. El modelo se expresa primero en lenguaje de Stan en el siguiente código y se almacena dentro del entorno R como una cadena de caracteres llamada **BRMA1**.

```
> BRMA1 <- "
+ data{
+   int<lower = 0> N;
+   int<lower = 0> tp[N];
+   int<lower = 0> dis[N];
+   int<lower = 0> tn[N];
+   int<lower = 0> nondis[N];
+ }
+ parameters{
+   real etarho;
+   vector[2] mul;
+   vector<lower = 0>[2] sigma;
+   vector[2] logitp[N];
+   vector[2] logitphat[N];
+ }
+ transformed parameters{
+   vector[N] p[2];
+   vector[N] phat[2];
+
+   real MU[2];
```

```

+   vector[2] mu;
+   real rho;
+   real ktau;
+   matrix[2,2] Sigma;
+
+   rho = tanh(etarho);
+   ktau = (2/pi())*asin(rho);
+
+   for (a in 1:2){
+       for (b in 1:N){
+           p[a][b] = inv_logit(logitp[b][a]);
+           phat[a][b] = inv_logit(logitphat[b][a]);
+       }
+       mu[a] = inv_logit(mul[a]);
+   }
+
+   MU[1] = mean(phat[1]);
+   MU[2] = mean(phat[2]);
+
+   Sigma[1, 1] = sigma[1]^2;
+   Sigma[1, 2] = sigma[1]*sigma[2]*rho;
+   Sigma[2, 1] = sigma[1]*sigma[2]*rho;
+   Sigma[2, 2] = sigma[2]^2;
+ }
+ model{
+   etarho ~ normal(0, 10);
+   mul ~ normal(0, 10);
+   sigma ~ cauchy(0, 2.5);
+
+   for (i in 1:N){
+       logitp[i] ~ multi_normal(mul, Sigma);
+       logitphat[i] ~ multi_normal(mul, Sigma);
+   }

```

4. ANÁLISIS DE RESULTADOS

```
+
+   tp ~ binomial(dis,p[1]);
+   tn ~ binomial(nondis, p[2]);
+
+ }
+ generated quantities{
+   vector[N*2] loglik;
+
+   for (i in 1:N){
+     loglik[i] = binomial_lpmf(tp[i] | dis[i], p[1][i]);
+   }
+   for (i in (N+1):(2*N)){
+     loglik[i] = binomial_lpmf(tn[i-N] | nondis[i-N], p[2][i-N]);
+   }
+ }
+ "
```

A continuación, se preparan los datos creando una lista de la siguiente manera:

```
> datalist = list(
+   tp = telomerase$TP,
+   dis = telomerase$TP + telomerase$FN,
+   tn = telomerase$TN,
+   nondis = telomerase$TN + telomerase$FP,
+   N = 10)
```

En el bloque de datos se especifican las dimensiones y nombres de las variables en el conjunto de datos, aquí N indica el número de estudios en el conjunto de datos. El bloque de parámetros introduce los parámetros desconocidos a estimar. Estos son η , un escalar que representa la transformada de Fisher del parámetro de asociación ρ , μ un vector 2×1 que representa la media de sensibilidad y especificidad en la escala logit para un estudio central donde el efecto aleatorio es cero, σ un vector 2×1 que representa la desviación estándar entre el estudio de la sensibilidad y la especificidad en la escala logit, \logitp una matriz $N \times 2$ de sensibilidad para el estudio en la primera columna y especificidad en la segunda columna en la escala logit,

y *logitphat* una matriz $N \times 2$ de sensibilidad pronosticada en la primera columna y la especificidad en la segunda columna en escala logit.

Los parámetros se transforman adicionalmente en el bloque de “parámetros transformados”. Aquí, p es una matriz $2 \times N$ de sensibilidad en la primera columna y especificidad en la segunda columna después de la transformación inversa logit de *logitp*, y *phat* es una matriz $2 \times N_s$ de sensibilidad predicha en la primera columna y especificidad en la segunda columna después de la transformación logit inversa de *logitphat* para ser utilizada en el cálculo de la sensibilidad y especificidad meta-analítica. Mu es un vector 2×1 que representa la media de sensibilidad y especificidad para un cierto estudio con un efecto aleatorio igual a 0, MU es un vector 2×1 que contiene la sensibilidad y especificidad meta-analítica, $Sigma$ una matriz 2×2 que representa la matriz de varianza-covarianza de sensibilidad y especificidad en la escala logit, ρ y κ son escalares que representan la correlación de tau de Pearson y Kendall respectivamente. Las distribuciones anteriores para todos los parámetros y la probabilidad de datos se definen en el bloque del modelo. Finalmente, en el bloque de cantidades generadas, *loglik* es un vector $(2N_s) \times 1$ de la probabilidad de logaritmo necesaria para calcular el *WAIC*.

Luego, se llama a la función *stan* del paquete *rstan* para traducir el código a C++, se compila el código y toma muestras de la distribución a posteriori de la siguiente manera.

```
> brma.1 <- stan(model_code = BRMA1,  
+             data = datalist,  
+             chains = 3,  
+             iter = 5000,  
+             warmup = 1000,  
+             thin = 10,  
+             seed = 3,  
+             cores = 3)
```

Las estimaciones de los parámetros se extraen y la convergencia de la cadena y la autocorrelación se examinan con el siguiente código.

```
> print(brma.1, pars=c('MU', 'mu', 'rho', "Sigma"), digits=4, prob=c(0.025, 0.5, 0.975))
```

4. ANÁLISIS DE RESULTADOS

```
Inference for Stan model: 2895953836ae0fbfc163aba8e0d162c2.
3 chains, each with iter=5000; warmup=1000; thin=10;
post-warmup draws per chain=400, total post-warmup draws=1200.
```

	mean	se_mean	sd	2.5%	50%	97.5%	n_eff	Rhat
MU[1]	0.7532	0.0021	0.0504	0.6343	0.7573	0.8392	550	1.0037
MU[2]	0.7920	0.0042	0.1127	0.5149	0.8099	0.9532	731	1.0016
mu[1]	0.7656	0.0015	0.0371	0.6891	0.7644	0.8384	624	1.0038
mu[2]	0.8950	0.0025	0.0756	0.6816	0.9112	0.9823	930	1.0023
rho	-0.9387	0.0081	0.1323	-0.9997	-0.9886	-0.5690	266	1.0080
Sigma[1,1]	0.3297	0.0104	0.2764	0.0533	0.2533	1.0615	709	1.0115
Sigma[1,2]	-1.2375	0.0351	0.9200	-3.6019	-0.9787	-0.2586	686	1.0127
Sigma[2,1]	-1.2375	0.0351	0.9200	-3.6019	-0.9787	-0.2586	686	1.0127
Sigma[2,2]	5.7514	0.1873	4.5836	1.6390	4.4223	17.0807	599	1.0119

```
Samples were drawn using NUTS(diag_e) at Sat Apr 28 11:46:43 2018.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

La sensibilidad meta-analítica ($MU[1]$) y la especificidad ($MU[2]$) y los intervalos de confianza al 95 % son 0.7532[0.6343 0.8392] y 0.7920[0.5149 0.9532] respectivamente. Esto difiere de lo que publicaron los autores, cuyos valores fueron (0.75[0.66 0.74] y 0.86[0.71 0.94]). Los autores ajustaron una distribución normal bivalente estándar a los valores de la transformación logit de sensibilidad y especificidad de los estudios, lo que permitió la heterogeneidad entre los estudios y descartó el nivel más alto del modelo jerárquico. Debido a esto, los autores tuvieron que usar una corrección de continuidad de 0.5 ya que el séptimo estudio había “observado” especificidad igual a 1, un problema no encontrado en el modelo jerárquico. En segundo lugar, los autores no informan los valores meta-analíticos, sino que informan la sensibilidad ($mu[1]$) y la especificidad ($mu[2]$) media para un estudio hipotético particular con efecto aleatorio igual a cero, que en nuestro caso es 0.7656[0.6891 0.8384] y 0.8950[0.6816 0.9823] respectivamente y es comparable a lo que informaron los autores. Esta discrepancia entre MU y mu de hecho aumentará con el aumento de la variabilidad entre los estudios. Los valores, ($MU[1]$) y

($\mu[1]$) son similares porque la variabilidad entre estudios de la sensibilidad en la escala logit ($\Sigma[1,1]$) es pequeña 0.3297[0.0533 1.0615]. ($\mu[2]$) es considerablemente menor que ($\mu[2]$) como resultado de la heterogeneidad sustancial entre estudios en especificidad en la escala logit ($\Sigma[2,2] = 5.7514[1.6390 17.0807]$).

La figura siguiente muestra una mezcla de cadenas satisfactoria con poca autocorrelación para la mayoría de los modelos ajustados, excepto los modelos de cópula de Clayton.

```
> f1.1 <- traceplot(fitgauss.1)
> f1.2 <- traceplot(fitc90.1)
> f1.3 <- traceplot(fitc270.1)
> f1.4 <- traceplot(fitfgm.1)
> f1.5 <- traceplot(fitfrank.1)
> #draws <- as.array(brma.1, pars=c('MU'))
> f1.6 <- rstan::stan_trace(brma.1, pars=c('MU'))

> library(Rmisc)

> multiplot(f1.1, f1.2, f1.3, f1.4, f1.5, f1.6, cols=2)
```

4. ANÁLISIS DE RESULTADOS

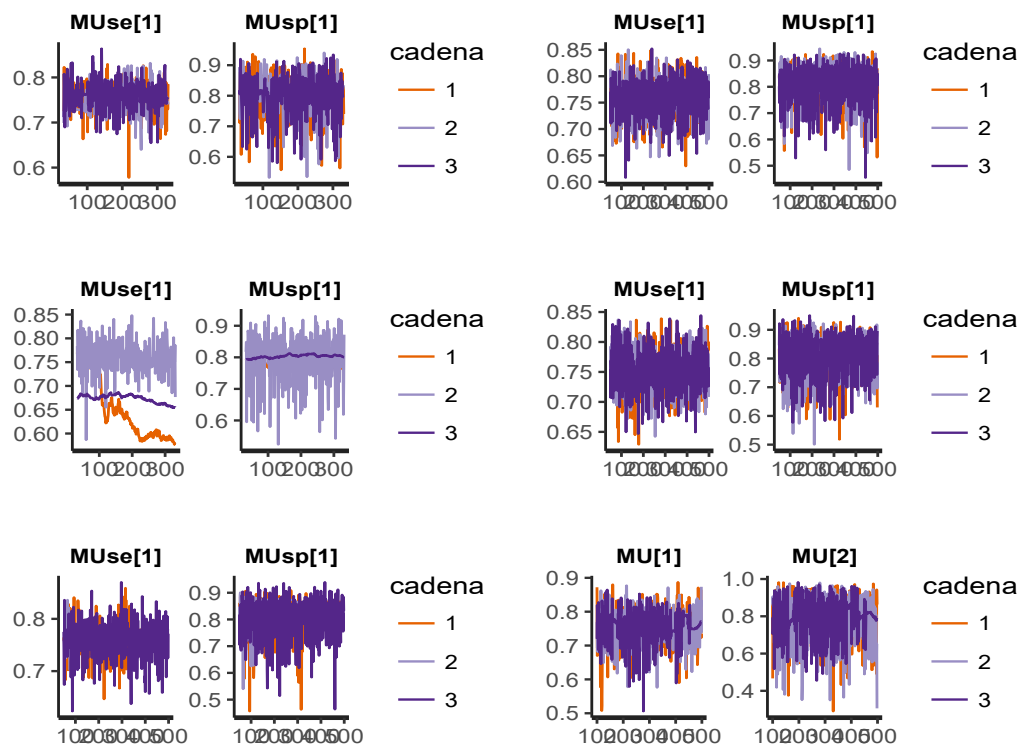


Figura 4.38: Trace Plot de la sensibilidad y especificidad media a posteriori de los datos de telomerasa estimados por cópulas de Gaussian, Clayton 90° (C90), 270° (C270), Farlie-Gumbel-Morgenstern (FGM) y lade Frank basada en distribuciones beta bivariantes y normal bivalente (BRMA).

La sensibilidad y especificidad media estimadas para todas las distribuciones ajustadas se muestran en la tabla a continuación.

```
> brma.summary1 <- data.frame(Parameter=c("Sensitivity",
"Specificity", "Correlation", "Var(1Sens)",
"Sigma[1,2]", "Sigma[2,1]", "Var(1Spec)"),
summary(brma.1, pars=c('MU', 'ktau', 'Sigma'))$summary[,c(1, 4, 6, 8:10)])
> brma.summary1 <- brma.summary1[-c(5,6),]
> names(brma.summary1)[2:5] <- c("Mean", "Lower", "Median", "Upper")
> library(loo)
```

4.5 Precisión diagnóstica mediante el paquete CopulaDTA para Meta-Análisis

```
> Table1 <- cbind(Model=rep(c("Gaussian", "C90", "C270", "FGM", "Frank", "BRMA"), each=5),
rbind(summary(fitgauss.1)$Summary,
summary(fitc90.1)$Summary,
summary(fitc270.1)$Summary,
summary(fitfgm.1)$Summary,
summary(fitfrank.1)$Summary,
brma.summary1),
WAIC = t(data.frame(rep(c(summary(fitgauss.1)$WAIC[1],
summary(fitc90.1)$WAIC[1],
summary(fitc270.1)$WAIC[1],
summary(fitfgm.1)$WAIC[1],
summary(fitfrank.1)$WAIC[1],
loo::waic(extract_log_lik(brma.1, parameter_name="loglik"))[3]), each=5))))

> rownames(Table1) <- NULL
> print(Table1, digits=4)
```

4. ANÁLISIS DE RESULTADOS

Modelo	Parámetro	Media	Inferior	Superior	n_eff	Rhat	WAIC
Gaussian	Sensibilidad	0.756404	6.905e-01	8.159e-01	572.630	1.0057	90.43
	Especificidad	0.798385	6.310e-01	9.063e-01	697.829	1.0057	
	Correlación	-0.858247	-9.858e-01	-4.286e-01	294.938	1.0065	
C90	Sensibilidad	0.692787	5.858e-01	8.058e-01	2.678	2.3483	89.80
	Especificidad	0.789717	6.680e-01	8.858e-01	27.537	1.0520	
	Correlación	-0.229573	-9.772e-01	-2.808e-18	1.884	2.0396	
C270	Sensibilidad	0.761263	6.988e-01	8.138e-01	227.667	1.0269	110.04
	Especificidad	0.826975	6.614e-01	8.964e-01	9.318	1.1135	
	Correlación	-0.429526	-9.850e-01	-9.942e-18	4.390	1.4996	
FGM	Sensibilidad	0.758029	6.868e-01	8.199e-01	1264.625	0.9983	95.51
	Especificidad	0.807785	6.548e-01	9.081e-01	1338.751	0.9997	
	Correlación	-0.190256	-2.222e-01	2.222e-01	1304.596	1.0038	
Frank	Sensibilidad	0.756100	6.880e-01	8.154e-01	1350.000	0.9995	90.17
	Especificidad	0.811109	6.474e-01	9.094e-01	1350.000	1.0002	
	Correlación	-0.702213	-8.523e-01	1.000e+00	1350.000	NaN	
BRMA	Sensibilidad	0.753172	6.343e-01	8.392e-01	549.734	1.0037	-43.83
	Especificidad	0.792032	5.149e-01	9.532e-01	731.158	1.0016	
	Correlación	-0.840375	-9.835e-01	-3.853e-01	168.022	1.0181	

Tabla 4.5: Media a posteriori, intervalo de confianza al 95% , tamaño de muestra efectiva y factor de reducción de escala potencial \hat{R} para las medias marginales y parámetros de correlación según estimación por cópulas de, Gaussian, Clayton 90° (C90) y 270° (C270), Farlie-Gumbel-Morgenstern (FGM) y Frank explicadas en distribuciones beta bivariantes y el modelo normal bivalente (BRMA) para los datos de telomerasa.

Los resultados se presentan gráficamente de la siguiente manera.

```
> g1 <- ggplot(Table1[Table1$Parameter %in% c("Sensitivity", "Specificity"),],
+           aes(x=Model,
+             y= Mean)) +
```

```
+ geom_point(size=3) +
+ theme_bw() +
+ coord_flip() +
+ facet_grid(~ Parameter, switch="x") +
+ geom_errorbar(aes(ymin=Lower,
+                   ymax=Upper),
+              size=.75,
+              width=0.15) +
+ theme(axis.text.x = element_text(size=13, colour='black'),
+       axis.text.y = element_text(size=13, colour='black'),
+       axis.title.x = element_text(size=13, colour='black'),
+       strip.text = element_text(size = 13, colour='black'),
+       axis.title.y= element_text(size=13, angle=0, colour='black'),
+       strip.text.y = element_text(size = 13, colour='black'),
+       strip.text.x = element_text(size = 13, colour='black'),
+       plot.background = element_rect(fill = "white", colour='white'),
+       panel.grid.major = element_blank(),
+       panel.background = element_blank(),
+       strip.background = element_blank(),
+       axis.line.x = element_line(color = 'black'),
+       axis.line.y = element_line(color = 'black')) +
+ scale_y_continuous(name="Mean [95% equal-tailed credible intervals]",
+                   limits=c(0.45,1.1),
+                   breaks=c(0.5, 0.75, 1),
+                   labels = c("0.5", "0.75", "1"))
> g1
```

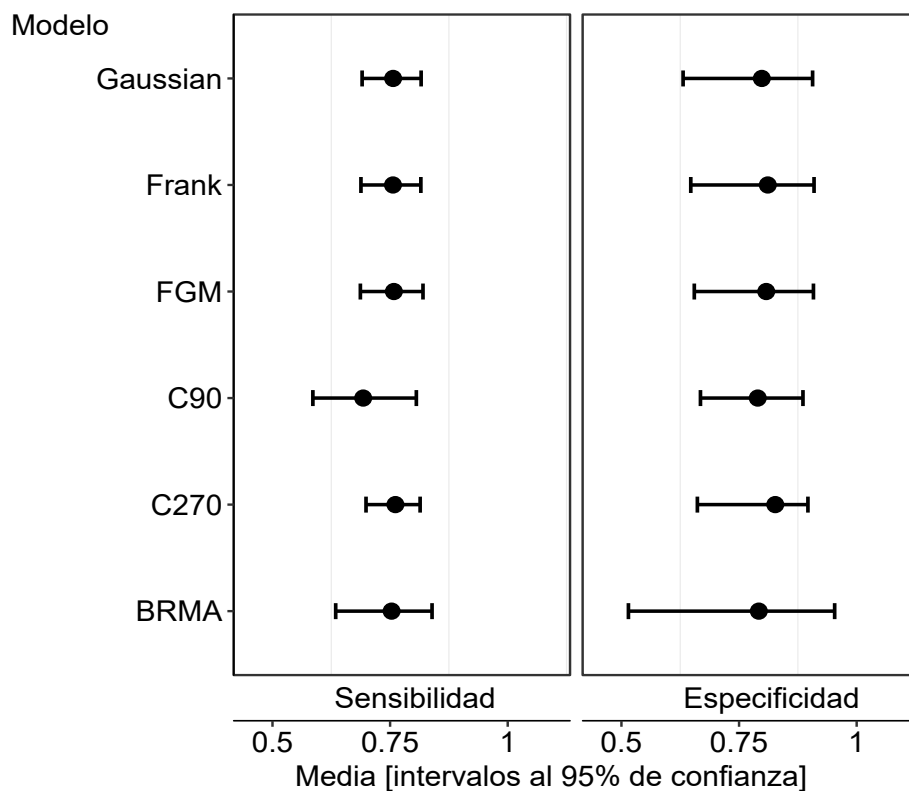


Figura 4.39: Gráfico de la sensibilidad y especificidad meta-analítica a posteriori y los intervalos correspondientes de 95 % de confianza según lo estimado por las cópulas de, Gaussian, Clayton 90° (C90) y 270° (C270), Farlie-Gumbel-Morgenstern (FGM) y la de Frank basada en distribuciones bivariadas bivariantes y el modelo normal bivalente (BRMA) para los datos de la telomerasa.

4.5.7. Comparación de modelos

La Tabla 4.5 muestra que la correlación estimada por el modelo BRMA y la cópula gaussiana beta bivalente son más extremas pero comparables con las estimaciones de la cópula de Frank, y Clayton 90° y 270°. Por otro lado están las estimaciones a partir de la cópula FGM y esto se debe a las restricciones sobre el parámetro de asociación en la cópula, donde los valores están dentro de $[2/9]$. En la Figura 4.39 la sensibilidad y especificidad media marginal de las distribuciones beta bivalentes son comparables con diferencias sutiles en sus intervalos de confianza a pesar de diferencias en la estructura de correlación.

Glas et al. (65) y Riley et al. (149) estimaron el parámetro de correlación de Pearson en el modelo BRMA ρ como -1 dentro del marco frecuentista. Utilizando la estimación de máxima verosimilitud, Riley et al. (150) demostraron que la correlación entre estudios del BRMA es a menudo estimada como $+/- 1$. La Tabla 4.5 muestra una correlación estimada de Pearson de $-0.840375[-0.9835 \quad -0.3853]$. Esto se debe a que los métodos **bayesianos** no están influenciados por el tamaño de la muestra y por lo tanto son capaces de manejar casos de tamaños de muestras pequeñas con menos problemas.

Esencialmente, los seis modelos son equivalentes en el primer nivel de jerarquía y difieren en la especificación de las distribuciones previas para la sensibilidad y especificidad de los estudios. Por lo tanto, los modelos deben tener el mismo número de parámetros, en cuyo caso tiene sentido comparar los logaritmos de las densidades predictivas. Según **WAIC**, este indica que el modelo BRMA ajusta de mejor manera los datos analizados.

4.6. Meta-análisis de precisión diagnóstica utilizando el programa informático SAS

4.6.1. Modelos Jerárquicos

El metanálisis de los estudios de precisión diagnóstica requiere el uso de métodos más avanzados que el meta-análisis de los estudios de intervención. Esto se debe al hecho de que el rendimiento de la prueba se informa típicamente en términos de la sensibilidad y especificidad de las dos medidas y, a menudo, existe una correlación negativa entre ellas a medida que cambia el umbral de la positividad de la prueba. Además, generalmente hay una heterogeneidad considerable entre los estudios en el metanálisis. El enfoque de modelado jerárquico o multinivel toma en cuenta tanto dentro como entre las variaciones de estudio. Se han propuesto dos modelos, a saber, el modelo de característica operativa de resumen jerárquico (HSROC) y el modelo de efectos aleatorios bivariante. Se ha demostrado que ambos modelos están estrechamente relacionados y son idénticos en situaciones comunes.

Ambos modelos jerárquicos implican distribuciones estadísticas en dos niveles. En el primer nivel, modelizan los valores de las celdas en las tablas de 2×2 extraídas de cada estudio usando distribuciones binomiales y transformaciones logísticas (log-odds). En el nivel superior, se supone que los efectos del estudio aleatorio explican la heterogeneidad en la precisión de la prueba diagnóstica entre los estudios más allá de los que se explican por variabilidad de muestreo en el primer nivel. El modelo bivariante y el modelo HSROC son matemáticamente equivalentes cuando no se incluyen covariables, pero difieren en sus parametrizaciones. La parametrización Bivariante modeliza la sensibilidad, la especificidad y la correlación entre ellos directamente, mientras que en el HSROC la parametrización modeliza las funciones de sensibilidad y especificidad para definir una curva SROC.

Tanto del modelo bivariante como del modelo HSROC producen salidas como:

- Curva SROC.
- El punto de operación de resumen, (es decir, valores de resumen de sensibilidad y espe-

cificidad).

- Una región de confianza del 95 % alrededor del punto de operación resumido.
- Una región de predicción del 95 %

Esta región de predicción es una forma de ilustrar el alcance de la heterogeneidad estadística al representar una región dentro de la cual, asumiendo que el modelo es correcto, tenemos 95 % de confianza de que la verdadera sensibilidad y especificidad de un estudio futuro caiga dentro de esta región.

Para poder emplear estos modelos jerárquicos se utilizará el procedimiento “**PROC NLMIXED**” de SAS. Este procedimiento se ajusta a modelos mixtos lineales generalizados que utilizan métodos basados en la verosimilitud. Requiere una ecuación de regresión y una declaración de parámetros con sus estimaciones iniciales (**valores iniciales**). Estos valores iniciales son necesarios para el proceso iterativo y es esencial seleccionar unos “buenos” valores para evitar un tiempo de cálculo excesivamente largo y también facilitar la convergencia. **Proc NLMIXED** tiene la capacidad de buscar valores iniciales sobre una cuadrícula de posibles valores para el proceso de inicialización. Se pueden obtener predicciones empíricas de Bayes para los efectos aleatorios y las estimaciones de las funciones de los parámetros, el método delta se utiliza para estimar el error estándar. **NLMIXED** no puede ajustar una gran cantidad de efectos aleatorios “solo menores que 5” y está limitado a 2 niveles.

4.6.2. Modelo Bivalente

Resumiendo lo analizado en el Capítulo 3, el método Bivalente modeliza la sensibilidad y la especificidad directamente. Se puede considerar que el modelo tiene dos niveles correspondientes a la variación dentro y entre los estudios. En el primer nivel, se supone que la variabilidad dentro del estudio para sensibilidad y especificidad sigue una distribución binomial. Para sensibilidad (denotada por A), el número de pruebas positivas de cada estudio $y_{Ai} \sim B(n_{Ai}, \pi_{Ai})$ donde n_{Ai} se refiere al número de individuos con la condición de interés evaluados y π_{Ai} la probabilidad de un resultado positivo del test en el grupo de estudio i . Similarmente, para la especificidad (denotada por B), el número de pruebas negativas de cada estudio se distribuye de acuerdo a una binomial $y_{Bi} \sim B(n_{Bi}, \pi_{Bi})$, donde n_{Bi} denota el número total de individuos

4. ANÁLISIS DE RESULTADOS

sanos evaluados y π_{Bi} la probabilidad de un resultado negativo del test en el grupo de estudio i .

El par de sensibilidad y especificidad para cada estudio debe modelarse conjuntamente en el nivel uno (dentro del estudio) del análisis porque están vinculados por características de estudio compartidas, incluido el umbral de positividad. En el nivel superior, se supone que las transformaciones *logit* de las sensibilidades tienen una distribución normal con media μ_A y varianza σ_A^2 , mientras que las transformaciones *logit* de las especificidades se distribuyen de acuerdo a una normal con media μ_B y varianza σ_B^2 . Su correlación se incluye modelando ambas distribuciones normales mediante una distribución normal bivalente:

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \Sigma \right)$$

$$\Sigma = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}$$

Donde σ_A^2 y σ_B^2 describen la variabilidad entre estudios del verdadero valor de las transformaciones *logit* de la sensibilidad y especificidad y σ_{AB} es la covarianza entre el *logit* de la sensibilidad y especificidad. El modelo también puede ser parametrizado empleando la correlación $\rho_{AB} = \sigma_{AB} / (\sigma_A \sigma_B)$ que puede ser más interpretable que la covarianza. Por lo tanto, el modelo bivalente tiene cinco parámetros cuando no se incluyen covariables: μ_A , μ_B , σ_A^2 , σ_B^2 y ρ_{AB} .

La inclusión de un parámetro de correlación en el modelo permite la compensación “**trade-off**” que existe entre la sensibilidad y especificidad, ya que el umbral de positividad de la prueba varía según los estudios. Cuando la variación entre los estudios surge a través de dicha compensación, se espera que esta correlación sea negativa, pero la correlación puede ser positiva si hay otras fuentes de heterogeneidad.

Ejemplo 1: AntiCCp para el diagnóstico de la artritis reumática, gráficos descriptivos

Estos datos provienen de una revisión realizada por Nishimura et al. (136) de anti-cyclic citrullinated peptide antibody (anti-CCP). La referencia estándar se basó en los criterios del Colegio Americano de Reumatología de 1987 (ACR). Treinta y siete estudios se incluyeron en el meta-análisis.

Los *forest-plot* se muestran en la Figura 4.40, la cual denota los estudios en orden alfabético y los valores de las medidas de la tabla 2×2 (TP, FP, FN, y TN) para cada estudio que formará la base para análisis estadísticos. Las estimaciones específicas del estudio de sensibilidad y especificidad son mostrados, con sus intervalos de confianza del 95 %. Estas estimaciones (e intervalos de confianza) también se muestran gráficamente. La característica más llamativa de esta figura es la gran dispersión (indicada por el ancho del intervalo de confianza) y la variabilidad (indicada por la dispersión de estimaciones puntuales) en sensibilidad que en especificidad. Los estudios se pueden ordenar de diferentes maneras (por ejemplo, en orden creciente de sensibilidad) para proporcionar un representación de cualquier asociación entre sensibilidad y especificidad.

4. ANÁLISIS DE RESULTADOS

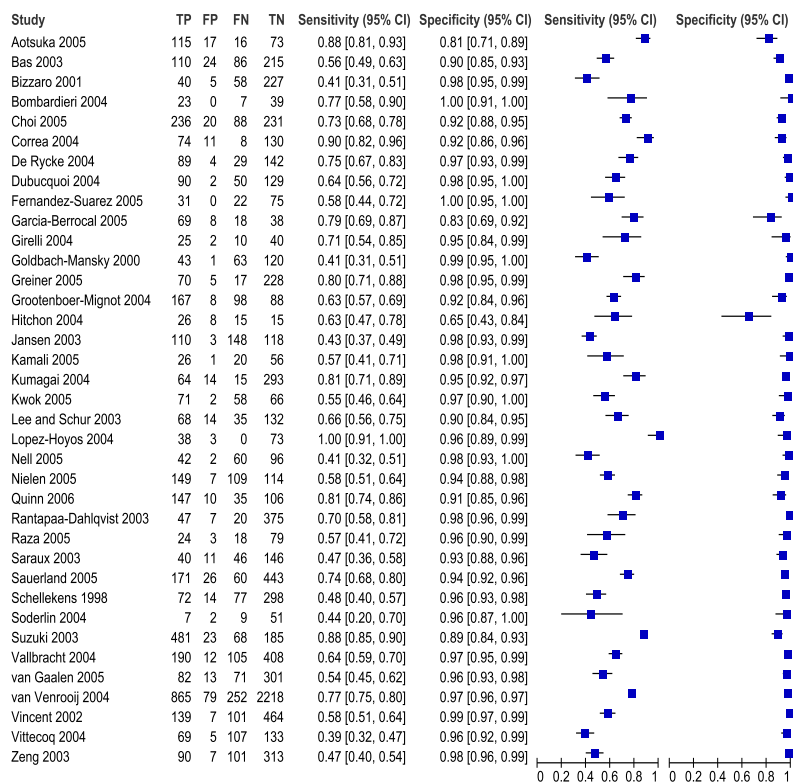


Figura 4.40: Forest Plot, datos Nishimura et al. (136)

El diagrama de dispersión *ROC* que se muestra en la Figura 4.41 muestra que la mayor variabilidad se concentra en la sensibilidad que en la especificidad. La información de Covariables (por ejemplo, generación de CCP) se puede usar para distinguir entre los estudios diferentes subgrupos (por ejemplo, CCP1 vs CCP2) permitiendo agregar curvas *ROC* que se pueden superponer para un análisis descriptivo. Sin embargo, un análisis estadístico más formal es requerido para proporcionar estimaciones de resumen de la precisión y la heterogeneidad.

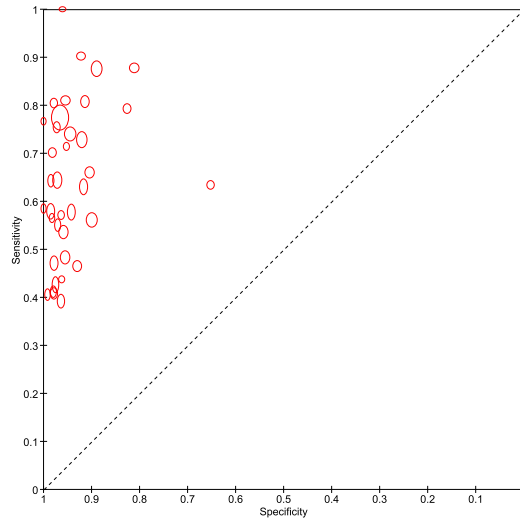


Figura 4.41: Scatter Plot, datos Nishimura et al. (136)

Antes de continuar con estos análisis estadísticos, el autor de la revisión debe decidir si es apropiado enfocarse en un punto de resumen o una curva resumen en los análisis estadísticos. Esto estará determinado por los umbrales utilizados por los estudios para definir un resultado positivo de la prueba.

Ahora emprendemos la primera etapa de un análisis estadístico formal de los datos del anticuerpo anti-cyclic citrullinated peptide antibody (anti-CCP). Supondremos que la prueba anti-CCP se considera positiva si se detecta cualquier anticuerpo anti-CCP y esa detección puede considerarse un umbral común, tiene sentido centrarse en las estimaciones resúmenes de sensibilidad y especificidad. Las estimaciones de los parámetros del modelo bivalente se muestran en la Figura 4.42, para ver código en SAS obsérvese Apéndice A.4.

4. ANÁLISIS DE RESULTADOS

Estadísticos de ajuste							
Verosimilitud	-2 Log						545.6
AIC (mejor más pequeño)							555.6
AICC (mejor más pequeño)							556.4
BIC (mejor más pequeño)							563.6

Estimadores de parámetros								
Parámetro	Error		DF	Valor t	Pr > t	Límites de confianza al 95% Gradiente		
	Estimador	estándar						
msens	0.6534	0.1275	35	5.13	<.0001	0.3946	0.9122	3.959E-6
mspec	3.1090	0.1459	35	21.31	<.0001	2.8128	3.4052	3.472E-8
s2usens	0.5426	0.1463	35	3.71	0.0007	0.2455	0.8397	-6.62E-6
s2uspec	0.5717	0.1873	35	3.05	0.0043	0.1914	0.9520	1.36E-6
covsesp	-0.2704	0.1199	35	-2.26	0.0304	-0.5137	-0.02710	-1.59E-6

Matriz de covarianza de estimadores de parámetro					
	msens	mspec	s2usens	s2uspec	covsesp
msens	0.01625	-0.00741	0.000890	-0.00004	-0.00004
mspec	-0.00741	0.02128	-0.00006	0.004287	-0.00116
s2usens	0.000890	-0.00006	0.02142	0.003997	-0.00874
s2uspec	-0.00004	0.004287	0.003997	0.03509	-0.01184
covsesp	-0.00004	-0.00116	-0.00874	-0.01184	0.01436

Figura 4.42: Medidas resumen, ajuste bivalente, datos Nishimura et al. (136)

Las estimaciones de los parámetros en los cuadros anteriores pueden ser ingresadas a *RevMan* para producir una medida resumen, región de confianza y de predicción al 95% como se muestra en la Figura 4.43. La gráfica de salida bivalente en *RevMan* requiere: resumen de logit (sensibilidad) que es 0.6534, la estimación resumida para logit (especificidad) que es 3.1090; y las varianzas de los efectos aleatorios para logit (sensibilidad), logit (especificidad) y su covarianza que son 0.5426, 0.5717 y -0.2704 respectivamente. Para el cálculo de las regiones de confianza y de predicción se requieren el error estándar de las estimaciones de resumen para logit (sensibilidad), logit (especificidad) y su covarianza, que son 0.1275, 0.1459 y -0.00741 respectivamente.

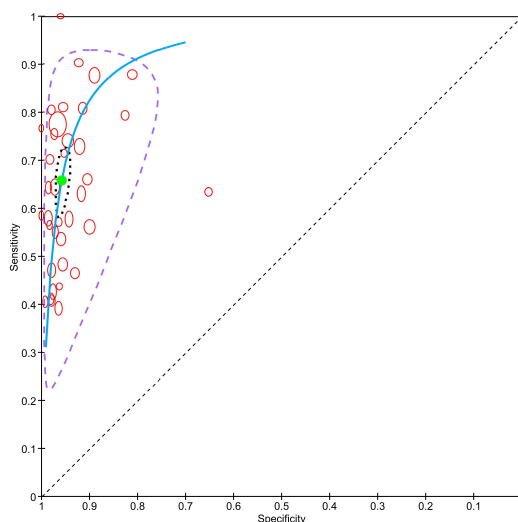


Figura 4.43: Curva SROC, ajuste bivalente, datos Nishimura et al. (136)

Los coeficientes de varianza indican una heterogeneidad similar en las sensibilidades y especificidades. La magnitud de la heterogeneidad también es evidente en el tamaño de la región de predicción en el gráfico SROC. La estimación resumida de sensibilidad y especificidad se muestra con el punto verde sólido. La sensibilidad y especificidad en este punto pueden ser estimadas a partir de la transformación inversa de las estimaciones logit, para obtener una sensibilidad y especificidad de 0.66 y 0.96 respectivamente. El gráfico 4.43 muestra un valor atípico potencial, **Hitchon 2004** con una sensibilidad de 0,63 y una especificidad de 0,65. Se puede realizar un análisis de sensibilidad para evaluar la influencia de este estudio en las estimaciones resumidas.

4.6.3. Modelo HSROC

El modelo *HSROC*, se basa en un modelo de regresión logística de escala latente. El modelo *HSROC* supone que hay una curva *ROC* subyacente en cada estudio con parámetros α y β que caracterizan la precisión y la asimetría de la curva, de forma similar (aunque técnicamente distinto) a los parámetros α y β en el método de regresión lineal de Moses. A diferencia del modelo de Moses, el modelo *HSROC* está obligado a proporcionar una curva *ROC* donde la sensibilidad no puede disminuir drásticamente a medida que aumenta la especificidad.

4. ANÁLISIS DE RESULTADOS

La precisión, definida en términos de $\ln(DOR)$, determina la posición de la curva resumen. Al igual que con el método de regresión SROC, cada estudio contribuye datos en un solo umbral para el análisis. La tabla tetracórica para cada estudio surge de la dicotomización en un umbral de positividad denotado por θ . Se supone que los parámetros α y θ varían entre los estudios: se supone que ambos parámetros tienen distribuciones normales como en el meta-análisis convencional de efectos aleatorios.

El modelo HSROC también considera dos niveles correspondientes a la variación dentro y entre los estudios. En el primer nivel, el número de individuos enfermos con resultados positivos se denomina y_{i1} para el i -ésimo estudio, y el número correspondiente de no enfermos con resultados positivos se indica con y_{i2} . Para cada estudio (i), se supone que el número de pruebas positivas en cada grupo de enfermedad (j) sigue una distribución binomial tal que $y_{ij} \sim B(n_{ij}, \pi_{ij})$, $j = 1, 2$ donde n_{ij} y π_{ij} respectivamente representan el número total de evaluados y la probabilidad de un resultado positivo de la prueba. El número de pruebas positivas en cada pareja enferma y no enferma se analiza conjuntamente dentro de cada estudio en el nivel uno del análisis.

El modelo toma la forma $\text{logit}(\pi_{ij}) = (\theta_i + \alpha_i \text{dis}_{ij}) \exp(-\beta \text{dis}_{ij})$ donde dis_{ij} representa el verdadero estado de enfermedad (codificado como -0.5 para el no enfermo y 0.5 para el enfermo) teniendo en cuenta la variabilidad dentro del estudio en el nivel uno. Usando la terminología usual para este modelo, generalmente nos referimos a θ_i como el valor medio para el umbral de positividad calculado como el promedio del log odds de un resultado de prueba positiva para el enfermo y el log odds de un resultado de prueba positiva para los grupos no enfermos en el estudio i . α_i (el $\ln(DOR)$ para el estudio i) representa una medida de precisión diagnóstica en el estudio que incorpora tanto la sensibilidad como la especificidad para ese estudio. El parámetro de escala (β) proporciona una asimetría a la curva SROC al permitir que la precisión varíe con el umbral dado que cada estudio contribuye solo a una estimación de sensibilidad y especificidad en un umbral único, por lo que es necesario suponer que la forma de la verdadera curva ROC subyacente en cada estudio es la misma y, por lo tanto, el valor β se ajusta como un efecto fijo.

El umbral y la precisión diagnóstica para cada estudio se especifican como efectos aleatorios y se

supone que son independientes (no correlacionadas) y normalmente distribuidas. El parámetro de precisión tiene media Λ y varianza σ_α^2 , mientras que el parámetro de positividad (umbral) cuenta con media Θ y varianza σ_β^2 . El parámetro de forma (β) se estima utilizando los datos de los estudios en conjunto y, suponiendo efectos aleatorios distribuidos normalmente para la precisión de la prueba. Cuando no se incluyen covariables en el modelo, el ajuste HSROC tiene cinco parámetros, Λ , Θ , σ_α^2 y σ_β^2 .

Se puede construir una curva ROC resumida a partir del modelo HSROC eligiendo un rango de valores de $1 - \textit{especificidad}$ y usando el parámetro de ubicación promedio estimado (Λ) y el parámetro de escala (β) para calcular los valores correspondientes para la sensibilidad. La sensibilidad esperada dado un valor de falso positivo (FPR) ($1 - \textit{especificidad}$) está dada por:

$$\textit{sensibilidad} = 1 / \left[1 + \exp \left(- \left(\Lambda e^{-0.5\beta + \textit{logit}(1 - \textit{especificidad}) e^{-\beta}} \right) \right) \right]$$

Cuando $\beta = 0$, la precisión de la prueba se puede resumir por Λ , que representa la precisión esperada ($\log(DOR)$), y la curva de resumen resultante será simétrica.

Ejemplo 2: Factor reumatoide como marcador de artritis reumática

En este ejemplo, se investiga el rendimiento diagnóstico del factor reumatoide (FR) como marcador de la artritis reumática (AR). Los 50 estudios incluidos en el análisis se tomaron de la misma revisión que en el Ejemplo 1 (Nishimura et al. (136)). El estándar de referencia se basó de nuevo en los criterios revisados de 1987 por el “American College of Rheumatology (ACR) or clinical diagnosis”.

El límite de positividad de la prueba para *RF* varió entre los estudios en un rango de 3 a 100 U/ml. La dispersión en el umbral utilizado para definir la positividad de la prueba entre los estudios se refleja en la variabilidad en el estudio de estimaciones específicas de sensibilidad y especificidad que se muestran en el gráfico de SROC, vea Figura 4.45. Debido a la variación en el umbral entre los estudios, una curva ROC resumida es apropiada para resumir estos datos. El modelo HSROC se usó para estimar una curva de resumen usando Proc NLMIXED (SAS) generando la siguiente salida, observe código en Apendice A.5.

4. ANÁLISIS DE RESULTADOS

Estadísticos de ajuste								
	Verosimilitud	-2 Log						806.9
	AIC	(mejor más pequeño)						816.9
	AICC	(mejor más pequeño)						817.6
	BIC	(mejor más pequeño)						826.5
Test para normalidad								
Estimadores de parámetros								
Parámetro	Estimador	Error estándar	DF	Valor t	Pr > t	Límites de confianza al 95%		Gradiente
alpha	2.6016	0.1862	48	13.97	<.0001	2.2273	2.9759	2.227E-6
theta	-0.4370	0.1469	48	-2.98	0.0046	-0.7323	-0.1417	4.573E-6
beta	0.2267	0.1624	48	1.40	0.1691	-0.09978	0.5532	-1.16E-6
s2ua	1.3014	0.3046	48	4.27	<.0001	0.6890	1.9137	-6.42E-7
s2ut	0.5423	0.1237	48	4.39	<.0001	0.2936	0.7909	-6.99E-6

Figura 4.44: Medidas resumen, ajuste HSROC, datos Nishimura et al. (136)

Las estimaciones de los parámetros mostrados en la Figura 4.44 se pueden ingresar en RevMan para trazar la curva de resumen como se muestra en la Figura 4.45; 2.6016 estima la media de los efectos aleatorios para la precisión (es decir, Λ), 0.4370 estima la media de los efectos para el umbral (θ), 0.2267 estima el parámetro de forma (β), 1.3014 estima la varianza de los efectos aleatorios para precisión y 0.5423 estima la varianza de los efectos aleatorios para el umbral. La curva resultante muestra el compromiso esperado entre la sensibilidad y la especificidad a través de los umbrales.

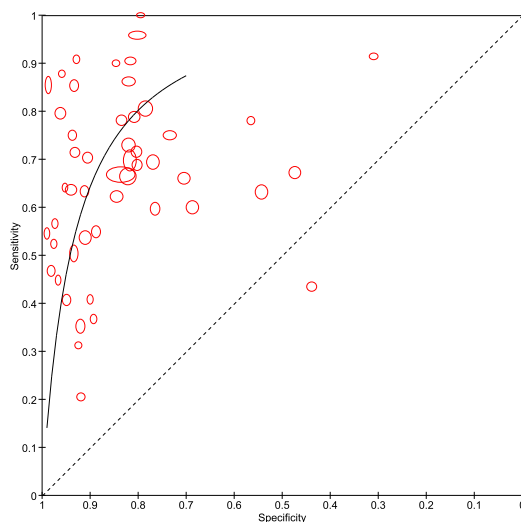


Figura 4.45: Curva SROC, ajuste HSROC, datos Nishimura et al. (136)

Al interpretar los resultados del análisis, es importante tener en cuenta que RF constituye parte de los criterios de ACR . Por lo tanto, existe un riesgo de sesgo en la curva estimada

ya que la prueba de índice se incorpora en la referencia estándar. Esto podría resultar en una sobre-estimación de la precisión diagnóstica de la *RF*, y podría dar como resultado una imagen distorsionada de la conveniencia de usar *RF* como una primera prueba para resolver la incertidumbre en un caso sospechoso de artritis reumatoide.

4.6.4. Heterogeneidad y análisis de regresión utilizando el modelo bivariente

El modelo bivariente permite que las covariables afecten la sensibilidad y/o especificidad. Suponiendo que tenemos un único estudio de nivel de covariable Z que pueda afectar tanto la sensibilidad como la especificidad, entonces el modelo se puede extender de la siguiente manera:

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_A + \nu_A Z_i \\ \mu_B + \nu_B Z_i \end{pmatrix}, \Sigma \right)$$

Como antes, Σ representa la matriz de covarianza para los efectos aleatorios para los logit de sensibilidad y especificidad. Si la covariable explica algo de la heterogeneidad en la sensibilidad y / o especificidad entonces esperaríamos que se reduzca la varianza estimada para uno o ambos efectos aleatorios. El parámetro de covarianza (correlación) estimado también puede cambiar.

Suponiendo que tenemos una covariable de nivel de estudio binario (Z) codificada como 0 o 1 para representar los dos grupos de estudios, μ_A estima la sensibilidad *logit* esperada en el punto resumen para el grupo de referencia ($Z = 0$), y $\mu_A + \nu_A$ estima la sensibilidad *logit* esperada en el punto resumen para el otro grupo ($Z = 1$). Por lo tanto, $\exp(\nu_A)$ estima el odds ratio para la sensibilidad en el grupo 1 en relación con el grupo de referencia. La sensibilidad esperada se estima como $\exp(\nu_A) / (1 + \exp(\nu_A))$ para el grupo de estudios referente, y como $\exp(\mu_A + \nu_A) / (1 + \exp(\mu_A + \nu_A))$ para el otro grupo. Las comparaciones de especificidad entre los dos grupos de estudios siguen el mismo enfoque como se describe anteriormente sobre la base de $\mu_B + \nu_B$. El ajuste del modelo, con y sin los parámetros adicionales ν_A y ν_B , se puede utilizar para evaluar si la covariable está asociada con la sensibilidad y/o especificidad. Esta prueba conjunta tendrá 2 grados de libertad si Z es binaria. También se pueden realizar pruebas separadas de la significación estadística de la covariable con sensibilidad y especificidad, primero para evaluar si ν_A difiere de 0 (un resultado significativo indica que hay evidencia de

4. ANÁLISIS DE RESULTADOS

que la sensibilidad difiere entre los dos grupos de estudios) y en segundo lugar si ν_B difiere de 0 (un resultado significativo indica que hay evidencia de que la especificidad difiere entre los dos grupos de estudios).

El modelo se extiende fácilmente para permitir más de una covariable. Sin embargo, esto puede no ser factible en la práctica si el número de estudios no es grande. Además, es importante tener en cuenta que una covariable solo puede estar asociada a la sensibilidad y no a la especificidad, o viceversa. No es necesario que las mismas covariables estén adaptadas tanto a la sensibilidad como a la especificidad, aunque este suele ser el caso. Cuando se permite que una covariable (o covariables) afecte tanto a la sensibilidad como a la especificidad, el modelo bivariante es equivalente a un modelo HSROC en el que la covariable o las covariables afectan tanto la precisión como el umbral de positividad pero no el parámetro de forma. Sin embargo, usando las estimaciones del modelo Bivariante para probar el efecto de las covariables en en la forma y posición de la curva SROC no es sencilla. El uso de la parametrización del modelo HSROC permite que esto se haga de una manera más directa.

Suele suponerse que la varianza de los efectos aleatorios (y su correlación en el caso del modelo bivariante) no están asociadas con la covariable. Esta es probablemente una suposición razonable en la mayoría de los análisis que investigan la heterogeneidad en la precisión de la prueba para una prueba de índice único. Sin embargo, para los análisis que comparan diferentes índices de pruebas, es menos probable que se mantenga esta suposición.

Continuación del Ejemplo 1: Investigando la heterogeneidad del desempeño diagnóstico de anti-CCP

Los estudios incluidos en la revisión para evaluar el rendimiento diagnóstico de anti-CCP utilizaron dos generaciones diferentes del ensayo: primera generación (*CCP1*, 8 estudios) y segunda generación (*CCP2*, 29 estudios). Una covariable binaria para la generación de *CCP* con el coeficiente *se2* para la sensibilidad, y el coeficiente *sp2* para la especificidad, las cuales se agregaron al modelo. La covariable se codificó como 0 para *CCP1* y 1 para *CCP2*. Permitir que el rendimiento de la prueba varíe según la generación de *CCP* en el modelo dio como resultado un valor -2Log Likelihood de -533.4 , una reducción de 12.2 en comparación con el modelo que no

contenía ninguna covariable. Por lo tanto, hay evidencia estadística (chi-cuadrado = 12.2, 2df, $P = 0.002$) de que el rendimiento de la prueba está asociado con la generación de CCP, pero se requiere más investigación para determinar si esta asociación es para sensibilidad, especificidad o ambas.

Las estimaciones de parámetros requeridos para graficar los puntos y regiones de resumen en RevMan pueden volver a ser extraídos de la salida Proc NLMIXED. Las varianzas de los efectos aleatorios para logit (sensibilidad) y logit (especificidad), y su covarianza son comunes para ambos generaciones de CCP. Para el grupo de referencia (CCP1 en este caso) el resumen de las estimaciones para logit (sensibilidad), logit (especificidad), los errores estándar correspondientes y la covarianza se muestran en la Figura 4.46. El logit (sensibilidad) para CCP2 es estimado por $msens + se2$, y el logit (especificidad) es estimado por $mspec + sp2$. Los errores estándar y la covarianza de estas medidas se pueden obtener usando el comando ESTIMATE en Proc NLMIXED como se muestra en el Apéndice A.4. Alternativamente, una forma simple de obtener estos resultados es reajustar el modelo utilizando CCP2 como el grupo de referencia (codificado como 0) y CCP1 como el otro grupo (codificado como 1). El ajuste del modelo y los resultados para los efectos aleatorios serán los mismos, pero las estimaciones para “msens” “mspec” serán ahora para CCP2 y, por lo tanto, las estimaciones requeridas se podrán extraer de la salida estándar. El gráfico resultante se muestra en la Figura 4.47.

4. ANÁLISIS DE RESULTADOS

Estadísticos de ajuste								
Verosimilitud	-2 Log							533.4
AIC (mejor más pequeño)								547.4
AICC (mejor más pequeño)								549.1
BIC (mejor más pequeño)								558.6

Estimadores de parámetros								
Parámetro	Estimador	Error estándar	DF	Valor t	Pr > t	Límites de confianza al 95%		Gradiente
msens	-0.09653	0.2203	35	-0.44	0.6640	-0.5438	0.3507	0.000317
mspec	3.4467	0.2982	35	11.56	<.0001	2.8412	4.0522	-0.00005
s2usens	0.3598	0.1022	35	3.52	0.0012	0.1524	0.5673	8.325E-6
s2uspec	0.5399	0.1802	35	3.00	0.0050	0.1742	0.9057	0.000159
covsesp	-0.1969	0.09836	35	-2.00	0.0532	-0.3965	0.002825	-0.00004
se2	0.9626	0.2513	35	3.83	0.0005	0.4523	1.4728	0.000319
sp2	-0.4302	0.3377	35	-1.27	0.2111	-1.1158	0.2554	-0.00004

Matriz de covarianza de estimadores de parámetro							
	msens	mspec	s2usens	s2uspec	covsesp	se2	sp2
msens	0.04854	-0.02464	-0.00012	-0.00001	-0.00003	-0.04855	0.02465
mspec	-0.02464	0.08895	-0.00002	0.004772	-0.00065	0.02463	-0.08834
s2usens	-0.00012	-0.00002	0.01044	0.002118	-0.00440	0.000693	-0.00005
s2uspec	-0.00001	0.004772	0.002118	0.03246	-0.00860	-0.00006	-0.00039
covsesp	-0.00003	-0.00065	-0.00440	-0.00860	0.009674	0.000100	-0.00091
se2	-0.04855	0.02463	0.000693	-0.00006	0.000100	0.06317	-0.03160
sp2	0.02465	-0.08834	-0.00005	-0.00039	-0.00091	-0.03160	0.1141

Figura 4.46: Medidas resumen con covariable “CCP”, ajuste Bivariante, datos Nishimura et al. (136)

Según las regiones de confianza en la Figura 4.47, está claro que la sensibilidad varía según la generación, pero no la especificidad. Las estimaciones resumen es de las especificidades fueron: 0.97 (*IC del 95 %* : 0.95;0.98) para la CCP1 y 0.95 (*IC del 95 %* : 0.94;0.97) para CCP2. Las estimaciones resumenes de la sensibilidad fueron 0.48 (*IC del 95 %* : 0.37;0.58) para CCP1 y 0.70 (*IC del 95 %* : 0.65;0.75) para CCP2. Estos resultados indican una mejora en la sensibilidad, sin pérdida de especificidad para la generación 2 en comparación con la generación 1 CCP. Comparando el resultado de este modelo con el del modelo sin covariables (ver SubSección 4.6.2.1), es claro que las varianzas de los efectos aleatorios se han reducido, particularmente para la sensibilidad.

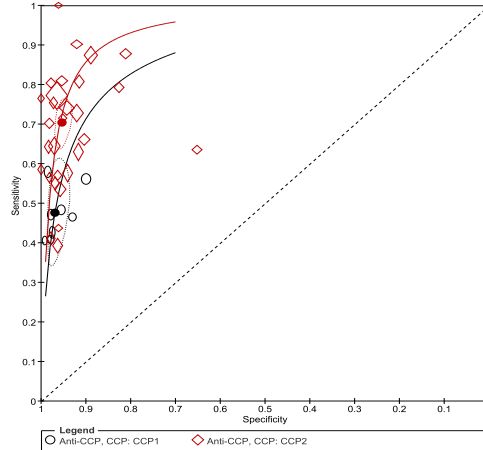


Figura 4.47: Curva SROC, ajuste Bivariante con covariable CCP, datos Nishimura et al. (136)

4.6.5. Heterogeneidad y análisis de regresión utilizando el modelo HS-ROC

El modelo HSROC permite agregar covariables para explorar la heterogeneidad en la positividad de la prueba (umbral), la posición de la curva (precisión) y la forma de la curva. Una covariable puede estar asociada con estas tres medidas, pero no con los tres parámetros del modelo.

Suponiendo que se tiene una covariable binaria (Z) codificada como 0 o 1 para representar los dos grupos de estudios, entonces el modelo HSROC puede extenderse para estimar las probabilidades *log* de una prueba positiva para el estudio i y el grupo de enfermedad j de la siguiente manera:

$$\text{logit}(\pi_{ij}) = ((\theta_i + \gamma Z_i) + (\alpha_i + \lambda Z_i) \text{dis}_{ij}) \exp(-(\beta + \delta Z_i) \text{dis}_{ij})$$

Donde, γ , λ y δ son asumidos como efecto fijos. Por lo tanto, la distribución de los efectos aleatorios para el umbral y la precisión ahora están dados por $\theta_i \sim N(\Theta + \gamma Z_i, \sigma_\theta^2)$ y

4. ANÁLISIS DE RESULTADOS

$\alpha_i \sim N(\Lambda + \lambda Z_i, \sigma_\alpha^2)$ respectivamente. El parámetro de forma para las curvas de resumen para los dos grupos se estima como β para el grupo de estudios de referencia ($Z = 0$) y $\beta + \alpha$ para el otro grupo ($Z = 1$). Si la covariable explica algo de la heterogeneidad en el umbral y / o la precisión, entonces esperaríamos que se reduzca la varianza estimada para uno o ambos efectos aleatorios.

El primer paso sería investigar la forma de la curva resumen. Si $\delta \neq 0$, entonces la forma de la curva resumen difiere para los dos grupos de estudios, lo que significa que la precisión relativa de la prueba variará con el umbral. La Figura 4.49 representa el escenario más complejo, y el modelo generalmente no se simplificará más. En la práctica, es difícil detectar una diferencia estadísticamente significativa en la forma de la curva entre los grupos porque el número de estudios en cada grupo suele ser limitado. Además, es importante cuando se investiga la forma considerar el efecto de estudios atípicos y potencialmente influyentes. Donde una curva se encuentra constantemente sobre otra, hay evidencia de una precisión superior, aunque la diferencia entre las curvas variará entre los umbrales. Si las curvas se cruzan, entonces la interpretación de qué curva muestra una precisión superior dependerá del umbral.

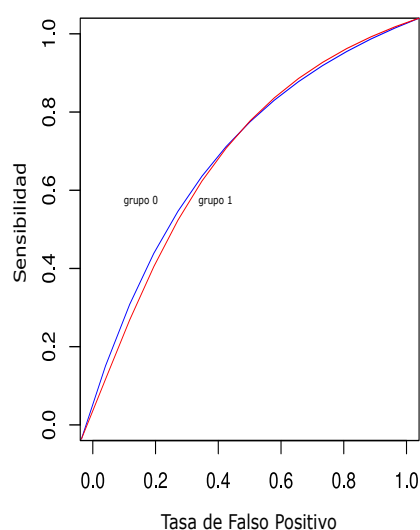


Figura 4.48: La precisión relativa depende de los valores de especificidad, las curvas se entrelazan

Si, de acuerdo con la evidencia estadística, se tiene similitud en las formas de las curvas, se puede suponer que $\delta = 0$, entonces la covariable puede eliminarse. Las curvas SROC estimadas para los dos grupos tendrán la misma forma, aunque no sean simétricas, Figura 4.49. y la precisión relativa de diagnóstico de las dos curvas se puede resumir utilizando el odds ratio de diagnóstico ($DOR = e\lambda$). El DOR será constante en todos los posibles valores de θ . Luego, el modelo se puede simplificar aún más y se puede suponer que ambas curvas son simétricas, es decir, $\beta = 0$, el DOR nuevamente proporciona una medida de precisión relativa.

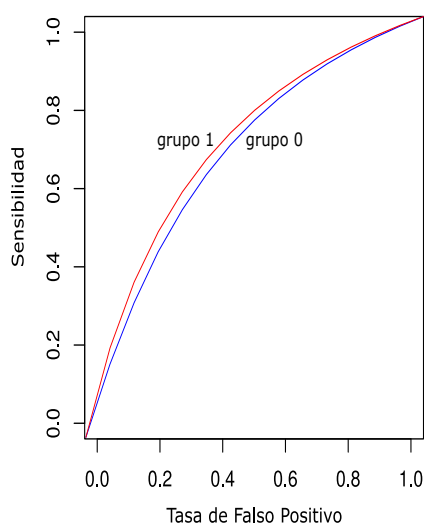


Figura 4.49: El grupo 1 domina en todos los valores de especificidad, las curvas no se cruzan.

Si se puede asumir que las curvas tienen la misma forma (tanto asimétrica como simétrica), entonces la pregunta es si la covariable está asociada con la precisión. Si hay evidencia de que $\lambda \neq 0$, entonces el DOR proporciona una estimación de la precisión diagnóstica global. Esto correspondería a una clara separación entre las curvas SROC para los dos grupos. Alternativamente, $\lambda = 0$ implica que no hay separación entre las curvas y no hay asociación entre la covariable y la precisión.

Si se puede suponer que $\lambda = 0$, entonces el modelo se puede simplificar aún más eliminando la covarianza para el parámetro de la exactitud que dará como resultado una única curva de

resumen (suponiendo que la forma de la curva es la misma para los dos grupos de estudios). Una asociación entre la covariable y el parámetro umbral (es decir, $\gamma \neq 0$) indicaría que la tasa de positividad de la prueba subyacente para los dos grupos de estudios es diferente. Tal asociación es a menudo difícil de interpretar a menos que se pueda asumir que las curvas tienen la misma forma y precisión.

Cuando el punto de corte real para definir una prueba positiva está disponible para cada estudio, se puede ajustar como covariable al parámetro umbral para permitir la estimación de la sensibilidad y especificidad en la curva de resumen en un punto de corte seleccionado. Sin embargo, esto presupone una relación funcional particular entre el umbral, la sensibilidad y especificidad.

4.6.6. Criterios para selección de modelos

Independientemente del modelo que se use, los revisores deben especificar qué estrategia de modelización se utilizará para agregar o eliminar covariables. La decisión sobre si una covariable debe incluirse en el modelo puede basarse en parte en pruebas estadísticas. El “software” comúnmente utilizado para ajustar estos modelos, por ejemplo es *SAS*, el mismo, proporciona el estadístico de Wald y los p – valores correspondientes para cada variable en el modelo. Un valor p basado en el estadístico de la razón de probabilidad chi-cuadrado es generalmente más confiable. El estadístico chi-cuadrado se calcula como el cambio en -2Log likelihood cuando se agrega (o elimina) una covariable en el modelo. El efecto de agregar (o eliminar) covariables para el ajuste del modelo se debe fundamentar en ciertas medidas como, el criterio de información de Akaike (AIC) o el criterio de información Bayesian (BIC). También se puede utilizar el criterio de información de desviación (DIC) que se usa comúnmente para la simulación de Cadenas de Markov y Monte Carlo (MCMC).

Las pruebas de razón de verosimilitud también se pueden usar para evaluar la significancia en términos de la varianza para los efectos aleatorios en cualquier modelo, o permitir que la varianza se relacione con la precisión de la prueba proporcionando un mejor ajuste en el modelo.

Continuación del Ejemplo 2: Investigando la heterogeneidad en la precisión diagnóstica del factor reumatoide

Se investigará si la técnica de laboratorio utilizada para medir RF está asociada con el rendimiento diagnóstico. De los 50 estudios, 15 usaron nefelometría (N), 16 de aglutinación de látex (LA), 16 de ELISA, un estudio utilizó hemoaglutinación de RA y 2 no informaron el método utilizado. El análisis se limita a estudios que utilizaron N, LA o ELISA. El modelo HSROC se utilizó debido a la variación en el umbral utilizado para la positividad de la prueba en todos los estudios. Las covariables (variables de indicador para la técnica, usando AL como la categoría de referencia) se incluyeron en el modelo para evaluar si la precisión, el umbral o la forma de la curva SROC variaban con la técnica.

El valor -2Log likelihood para el modelo más complejo que incluyó covariables para la forma de la curva, precisión y parámetros de umbral fue 752.9. El aumento en la probabilidad de -2Log fue insignificante (un aumento a 753.1) cuando la covariable para la forma se eliminó del modelo ($\text{chi-cuadrado} = 753.1 - 752.9 = 0.2$, 2 df, $P = 0.90$). Las estimaciones de parámetros para el modelo que asume una forma común se dan a continuación, y las curvas de HSROC correspondientes se muestran en la Figura 4.51.

Las estimaciones de α , θ y β se pueden ingresar en RevMan para obtener la curva resumen para el grupo de referencia (LA). Las varianzas de los efectos aleatorios para el umbral y la precisión son comunes a las tres técnicas, como lo es el parámetro de forma β . Sin embargo, las estimaciones de parámetros de umbral y exactitud para ELISA están dadas por $\theta + t1$ y $\alpha + a1$ respectivamente, y para N están dadas por $\theta + t2$ y $\alpha + a3$, vea Figura 4.50.

4. ANÁLISIS DE RESULTADOS

Estadísticos de ajuste								
Verosimilitud	-2 Log							753.1
AIC	(mejor más pequeño)							771.1
AICC	(mejor más pequeño)							773.2
BIC	(mejor más pequeño)							787.7

Estimadores de parámetros									
Parámetro	Error			DF	Valor t	Pr > t	Límites de		Gradiente
	Estimador	estándar					confianza al 95%		
alpha	2.4552	0.3245	45	7.57	<.0001	1.8017	3.1087	-0.00040	
theta	-0.5490	0.2137	45	-2.57	0.0136	-0.9794	-0.1186	0.000139	
beta	0.1995	0.1702	45	1.17	0.2472	-0.1432	0.5423	-0.00018	
s2ua	1.2865	0.3109	45	4.14	0.0002	0.6602	1.9128	-0.00038	
s2ut	0.4786	0.1139	45	4.20	0.0001	0.2492	0.7080	0.000620	
a1	0.2483	0.4408	45	0.56	0.5760	-0.6395	1.1361	-0.00038	
a2	0.3328	0.4439	45	0.75	0.4573	-0.5612	1.2269	0.000093	
t1	-0.1962	0.2614	45	-0.75	0.4568	-0.7227	0.3303	-0.00017	
t2	0.4960	0.2627	45	1.89	0.0654	-0.03301	1.0250	0.000366	

Figura 4.50: Medidas resumen según método, ajuste HSROC, datos Nishimura et al.

(136)

En la Figura 4.51, parece que *LA* puede ser menos precisa que los otros 2 métodos, sin embargo, la eliminación de la covariable para la precisión (coeficientes *a1* y *a2*) del modelo tiene un efecto insignificante en el ajuste del modelo ($\chi^2 = 753.7 - 753.1 = 0.6$) con 2 grados de libertad y un *p*-valor de 0.74, indicando que no existe evidencia estadística de una diferencia en la precisión diagnóstica de la RF de acuerdo con la técnica. Esto indica que es razonable ajustar un solo resumen ROC para RF.

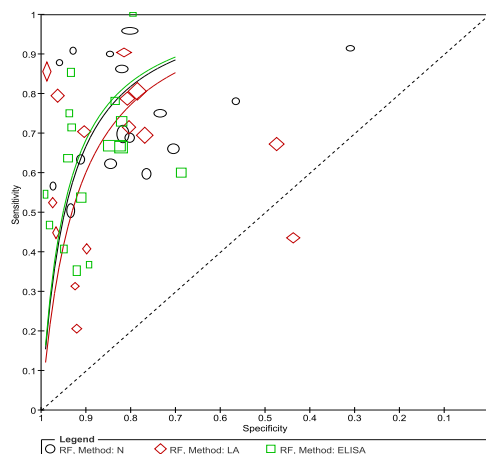


Figura 4.51: Curva SROC, ajuste HSROC según método, datos Nishimura et al. (136)

4.6.7. Comparando pruebas índices

Para muchas revisiones de diagnóstico, un objetivo clave es comparar la precisión diagnóstica de dos pruebas índice que pueden usarse para diagnosticar la misma afección. En esta sección, la atención se centrará en la comparación de dos pruebas índice, pero el enfoque puede ampliarse para permitir más de dos pruebas.

En general, se adoptan dos enfoques para las comparaciones de prueba. El primer enfoque utiliza datos de precisión de todos los estudios elegibles que han sido evaluados por una o dos pruebas. El segundo enfoque restringe el análisis a los estudios que han sido evaluados por ambas pruebas en los mismos individuos o que han asignado al azar individuos para someterse a una o a las dos pruebas. El segundo enfoque tiene ventajas porque es menos probable que la comparación sea sesgada debido a factores de confusión y, por lo tanto, se debe confiar en estos resultados siempre que sea posible. Sin embargo, el número de estudios que informan tales comparaciones es a menudo muy limitado, lo que significa que tal análisis puede no ser factible o solo puede considerarse como un análisis de sensibilidad.

Comparaciones de prueba basadas en estudios disponibles

A menudo, muchos de los estudios disponibles evalúan solo una de las pruebas de interés. Al utilizar todos los estudios que han evaluado al menos una de las pruebas, maximizamos la cantidad de estudios en el análisis. Sin embargo, es probable que los estudios sean heterogéneos en términos de diseño y características del paciente que están asociados con la precisión de la prueba y, por lo tanto, la confusión puede ser un problema. En los análisis preliminares, esto se puede resolver comparando las pruebas dentro de subgrupos de estudios que son homogéneos con respecto a factores de confusión potenciales importantes, como el diseño del estudio o el espectro de la enfermedad. El valor y la viabilidad de dichos análisis exploratorios se verán afectados por la cantidad de estudios disponibles y por la ausencia o inconsistencia de informes entre los estudios de información sobre posibles factores de confusión.

Para la comparación de dos pruebas índice, el tipo de prueba se representa mediante una covariable binaria que se utiliza para identificar la prueba que dio lugar a cada tabla tetracórica incluida en el análisis. Los factores de confusión pueden ser ajustados potencialmente, sin embargo esto a menudo es difícil de hacer en la práctica porque la cantidad de estudios es pequeña y/o los datos sobre los factores de confusión importantes pueden estar mal registrados o incompletos.

Tanto el modelo bivariante como el modelo HSROC se pueden utilizar para investigar la precisión relativa de dos pruebas índice. Sin embargo, como se señaló anteriormente, la elección del enfoque dependerá de la naturaleza de los datos disponibles, y la interpretación de los resultados también dependerá de qué enfoque se utilice.

Comparación de pruebas usando el modelo Bivariante

Si, para cada prueba índice, los estudios disponibles han utilizado un punto de corte consistente en una escala continua u ordinal para definir la positividad de la prueba, entonces el modelo bivariante proporciona un marco apropiado para las comparaciones de prueba. También puede ser razonable asumir un punto de corte consistente cuando una prueba comprende un “test kit” que produce resultados positivos y negativos (como una línea coloreada que aparece en un dispositivo). Al adoptar la misma estrategia descrita anteriormente (4.6.4), se puede incluir una

covariable binaria para el tipo de prueba para investigar si la sensibilidad esperada y / o la especificidad difieren entre las pruebas.

Se debe tener cuidado con la interpretación de los resultados de dicho modelo, particularmente si el punto de corte común para la positividad de la prueba para cualquier prueba se aplica a una escala continua u ordinal. Cualquier inferencia hecha sobre la precisión diagnóstica relativa a las dos pruebas solo es válida en el punto de corte elegido para cada una de las pruebas y no se puede extrapolar a otros puntos de corte posibles. Cuando se informan otros puntos de corte, el análisis se puede repetir utilizando los datos disponibles para investigar la precisión diagnóstica relativa de las pruebas en esos puntos de corte alternativos.

Debido a que se están analizando datos de precisión de prueba para dos pruebas índice, puede que no sea razonable asumir que las varianzas de los efectos aleatorios para logit (sensibilidad) y logit (especificidad) sean las mismas para las dos pruebas. El modelo bivariente puede ampliarse para permitir que la varianza de los efectos aleatorios (ambos) dependan de la covariable que está asociada al tipo de prueba. Esto también afectará la correlación estimada entre ellos. Estadísticamente, la estimación de las varianzas de los efectos aleatorios para logit (sensibilidad) y logit (especificidad) y la correlación entre ellos está sujeta a un mayor nivel de incertidumbre que para los principales parámetros de interés. Sin embargo, si se basan en gráficas preliminares donde se captura el nivel de sensibilidad y especificidad en el espacio ROC para los estudios, y se observan marcadas diferencias en la heterogeneidad entre los estudios para las dos pruebas, es aconsejable evaluar si se asumen varianzas iguales de efectos aleatorios para las dos pruebas. Esto generalmente se hace comparando el ajuste de los modelos alternativos (las varianzas dependen o no de la covariable para el tipo de prueba) usando una prueba de razón de verosimilitud. Una comparación de las principales estimaciones de interés entre los modelos alternativos también es útil para evaluar si las conclusiones sobre la sensibilidad y/o la especificidad de las pruebas son sólidas a las suposiciones sobre las variaciones de los efectos aleatorios. La investigación no será factible si la cantidad de estudios es pequeña.

Es habitual que la mayoría de los estudios en este enfoque para el análisis de las comparaciones de prueba hayan sido evaluados por una prueba, pero se puede dar el caso que algunos estudios han evaluados por ambas pruebas. Si la proporción de estudios que han sido evaluados por ambas

4. ANÁLISIS DE RESULTADOS

pruebas es muy pequeña, entonces es poco probable que los resultados de las dos pruebas en un estudio afecten a los resultados finales. Aunque esto se hace a menudo en la práctica, tal enfoque no se recomienda si la proporción de estudios que evalúan ambas pruebas no es pequeña porque es probable que resulte en errores estándar inapropiados para los parámetros de sensibilidad y especificidad. En ese caso, los pares ordenados sensibilidad / especificidad para ambas pruebas índice de cada estudio deben estar en el nivel uno del análisis, y una covariable binaria para el tipo de prueba incluida para identificar a qué tabla de 2×2 corresponde cada prueba.

Ejemplo 3: CT versus MRI para el diagnóstico de enfermedad arterial coronaria

Schuetz et al. (167), evaluó el rendimiento del diagnóstico de la tomografía computarizada (CT) multicorte y la resonancia magnética (MRI) para el diagnóstico de la enfermedad arterial coronaria (CAG). Los estudios prospectivos que evaluaron CT o MRI (o ambos), utilizaron angiografía coronaria convencional (CAG) como la referencia estándar y utilizaron el mismo umbral para la estenosis de arteria coronaria clínicamente significativa (una reducción del diámetro del 50 % o superior) en la revisión. Un total de 103 estudios proporcionaron una tabla de 2×2 para una o ambas pruebas y se incluyeron en el meta-análisis: 84 estudios evaluaron solo la CT, 14 evaluaron solo la MRI y 5 estudios evaluaron la CT y la MRI. (Ver el Apéndice A.6 para los datos y código en SAS).

Debido a que los estudios se seleccionaron en base a un umbral común para la estenosis de la arteria coronaria clínicamente significativa, el modelo bivariante se utilizó para la síntesis de datos y la comparación de la prueba. El enfoque sigue de cerca el método ilustrado en el Ejemplo 1 para explorar la heterogeneidad usando el modelo Bivariante.

Se agrega una covariable binaria (tipo de prueba) al modelo que se codifica como 0 si la tabla 2×2 es para MRI (la categoría de referencia), y se codifica como 1 si la tabla 2×2 es para CT. Los cinco estudios que evaluaron ambas pruebas aportan una tabla de 2×2 para cada prueba, por lo tanto, hay 19 estudios incluidos para la RM y 89 estudios incluidos para la TC. Permitir que el rendimiento de la prueba varíe según el tipo de prueba resultó en un valor -2Log Likelihood de 953.0, una reducción de 42.5 en comparación con el modelo que no contenía ninguna covariable.

Por lo tanto, existe evidencia estadística (chi-cuadrado = 42.5, 2df, $P < 0.001$) de que la sensibilidad y/o especificidad están asociadas con el tipo de prueba. La eliminación de la covariable para la sensibilidad del modelo (chi-cuadrado = $976.7 - 953.0 = 23.7$, 1df, $P < 0.001$) muestra una fuerte evidencia estadística de una diferencia en la sensibilidad entre las dos pruebas. Del mismo modo, la eliminación de la covarianza para la especificidad del modelo (chi-cuadrado = $976.2 - 953.0 = 23.2$, 1df, $P < 0.001$) muestra una fuerte evidencia estadística de una diferencia en la especificidad entre las dos pruebas.

La salida en SAS para el modelo que permite que tanto la sensibilidad como la especificidad varíen según la prueba es (Figura 4.52):

Estadísticos de ajuste								
Verosimilitud	-2 Log							953.0
AIC (mejor más pequeño)								967.0
AICC (mejor más pequeño)								967.5
BIC (mejor más pequeño)								985.5

Estimadores de parámetros								
Parámetro	Estimador	Error estándar	DF	Valor t	Pr > t	Límites de confianza al 95% Gradiente		
msens	2.1771	0.2457	101	8.86	<.0001	1.6896	2.6645	0.000046
mspec	0.8754	0.2111	101	4.15	<.0001	0.4566	1.2942	-0.00008
s2usens	0.8749	0.2293	101	3.82	0.0002	0.4201	1.3297	0.000033
s2uspec	0.8447	0.1696	101	4.98	<.0001	0.5084	1.1811	-4.31E-6
covsesp	0.1803	0.1384	101	1.30	0.1956	-0.09424	0.4548	-0.00002
se_CT	1.3033	0.2625	101	4.97	<.0001	0.7827	1.8240	0.000053
sp_CT	1.0415	0.2154	101	4.84	<.0001	0.6143	1.4687	-0.00005

Matriz de covarianza de estimadores de parámetro							
	msens	mspec	s2usens	s2uspec	covsesp	se_CT	sp_CT
msens	0.06038	0.005241	0.01100	0.000342	0.002242	-0.05262	-0.00404
mspec	0.005241	0.04457	0.000518	0.003360	0.001651	-0.00376	-0.03861
s2usens	0.01100	0.000518	0.05257	0.000694	0.007609	0.003839	-0.00060
s2uspec	0.000342	0.003360	0.000694	0.02875	0.005537	-0.00023	0.000438
covsesp	0.002242	0.001651	0.007609	0.005537	0.01915	0.001155	-0.00110
se_CT	-0.05262	-0.00376	0.003839	-0.00023	0.001155	0.06889	0.004479
sp_CT	-0.00404	-0.03861	-0.00060	0.000438	-0.00110	0.004479	0.04638

Matriz de covarianza de estimadores adicionales			
Etiqueta	Cov1	Cov2	
logitsens CT	0.02403	0.001916	
logitspec CT	0.001916	0.01373	

Figura 4.52: Medidas resumen, comparación de índices, ajuste bivalente, datos Schuetz et al. (167)

Las estimaciones de logit (sensibilidad) y logit (especificidad) para la categoría de referencia (MRI) están dados por *msens* y *mspec*, respectivamente. Estas estimaciones, sus errores estándar y su covarianza se muestran la Figura 4.52. El comando *ESTIMATE* en SAS se ha utilizado para obtener las estimaciones correspondientes para CT y las varianzas de los efectos

4. ANÁLISIS DE RESULTADOS

aleatorios y su covarianza, vea Figura 4.52.

Las estimaciones de la Figura 4.52 se pueden ingresar a RevMan para producir un diagrama de dispersión ROC con puntos de operación resumidos para MRI y CT y sus regiones de confianza superpuestas como se muestra en la Figura 4.53 donde los símbolos negros representan CT y el símbolo rojo representan MRI. Debido a la gran cantidad de estudios para CT, la estimación resumen y la región de confianza son difíciles de ver.

A partir del resultado anterior, los límites de confianza del 95 % para las estimaciones resumen para logit (sensibilidad) y logit (especificidad) se pueden encontrar en las columnas tituladas “Límites de confianza al 95 %”. Al usar la transformación inversa, las estimaciones resumen para la sensibilidad son 0.90 (*IC del 95 %* : 0.84; 0.93) para MRI y 0.97 (*IC del 95 %* : 0.96; 0.98) para la CT. Las estimaciones resumen para la especificidad 0.71 (*IC del 95 %* : 0.61a0.78) para MRI y 0.87 (*IC del 95 %* : 0.84a0.90) para CT. Por lo tanto, en base a este análisis, existe una fuerte evidencia de que CT tiene mayor sensibilidad y especificidad que MRI, para detectar estenosis de la arteria coronaria clínicamente significativa.

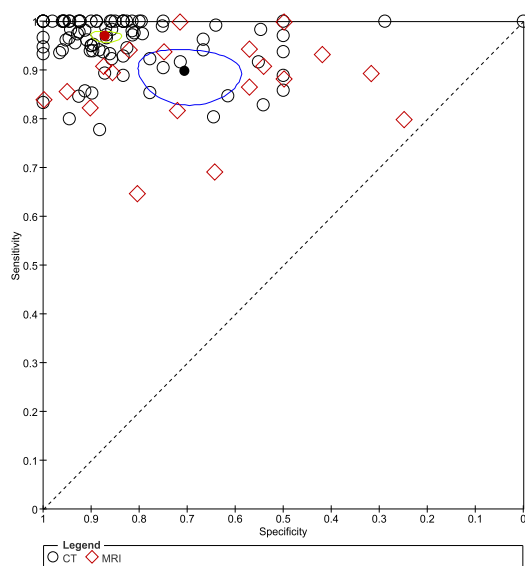


Figura 4.53: Curva SROC, ajuste Bivariante según Test, datos Schuetz et al. (167)

En el siguiente capítulo, presentamos un algoritmo de aprendizaje automático para el meta-análisis de pruebas diagnósticas, el mismo se ajustará a la Figura 3.17. Dado que los conjuntos

de datos médicos, como muchos otros conjuntos de datos del mundo real exhiben una distribución no simétrica; sin embargo, este no es el único problema a resolver en este tipo de conjuntos de datos, también debemos considerar otros problemas además de la escasa precisión de clasificación causada por distintas distribuciones estadísticas. Por lo tanto, proponemos una estrategia diferente basada en algoritmos de clasificación, que permita clarificar al investigador que enfoque implementar a la hora de estimar medidas resúmenes en el meta-análisis de pruebas diagnósticas.

Aprendizaje Automático

5.1. Introducción

Muchos conjuntos de datos del mundo real muestran una distribución no simétrica. La detección de fraude, el ataque a redes computacionales y el diagnóstico médico son ejemplos de este tipo de conjuntos de datos; sin embargo, frente a otras aplicaciones de aprendizaje autónomo, el problema de diagnóstico médico no termina una vez que tenemos un modelo para clasificar las nuevas instancias. Es decir, si la instancia se clasifica como enferma (la clase más importante) el conocimiento generado debe ser capaz de proporcionar al personal médico un punto de vista novedoso sobre el problema dado. Esto podría ayudar a aplicar un tratamiento médico a tiempo para evitar, retrasar o disminuir la incidencia de la enfermedad. Entonces, además de la precisión de clasificación, también deberíamos considerar la comprensión del conocimiento de diagnóstico. Además, debemos considerar un problema adicional, la selección de atributos relevantes (o factores de riesgo). Deberíamos centrarnos en los atributos modificables (pueden modificarse con tratamiento médico), como la presión arterial o los niveles de colesterol, y no deberíamos considerar atributos no modificables, como la edad y el sexo (por lo general, buenos atributos para la clasificación). Esto hace aún más difícil la tarea de clasificación.

Otro tema importante es que los conjuntos de datos médicos utilizados para el aprendizaje autónomo deben ser representativos de la incidencia general de la enfermedad estudiada. Esto es importante para hacer posible el uso del conocimiento generado con otras poblaciones. Por lo tanto, las técnicas de sobre-muestreo y sub-muestreo Chawla et al. (20), Kubat et al. (102)

utilizadas frecuentemente para equilibrar las clases y mejorar la predicción de clase minoritaria de algunos clasificadores, podrían generar conocimiento sesgado que podría no ser aplicable a la población general debido a la manipulación artificial de los conjuntos de datos. Por esta razón, proponemos una estrategia diferente que intenta maximizar la precisión de clasificación de la clase minoritaria (personas enfermas) sin modificar el conjunto de datos original. Por lo tanto, cada uno de los pasos de nuestro algoritmo se guía para alcanzar este objetivo. Como estamos lidiando con problemas de clasificación binarios, la precisión de la clase mayoritaria está garantizada por defecto.

En las siguientes secciones describimos, el origen del aprendizaje autónomo, la teoría de los algoritmos de clasificación, luego presentamos una breve descripción de la metodología de nuestro algoritmo acompañado de la precisión, comprensibilidad y validez de los resultados obtenidos.

5.2. Orígenes del Aprendizaje Automático

Los gobiernos, las empresas y las personas están registrando y reportando información, desde lo monumental hasta lo mundano. Los sensores meteorológicos registran los datos de temperatura y presión, las cámaras de vigilancia vigilan las aceras y los túneles del metro, y se monitorean todo tipo de comportamientos electrónicos: transacciones, comunicaciones, amistades y muchos otros.

Esta gran cantidad de datos ha llevado a algunos a afirmar que hemos entrado en una era de **“Big Data”**, pero esto puede ser un poco inapropiado. Los seres humanos siempre han estado rodeados de grandes cantidades de datos. Lo que hace que la era actual sea única son las grandes cantidades de datos almacenados, gran parte de los cuales pueden ser accedidos directamente por las computadoras.

El campo de estudio interesado en el desarrollo de algoritmos informáticos para transformar datos en acción inteligente se conoce como **“aprendizaje automático”**. Este campo se originó en un entorno donde los datos disponibles, los métodos estadísticos y el poder de cómputo evolucionaron de manera rápida y simultánea. El crecimiento de los datos requirió poder de cómputo adicional, lo que a su vez estimuló el desarrollo de métodos estadísticos para analizar

grandes conjuntos de datos. Esto creó un ciclo de avance que permite recopilar datos aún más grandes e interesantes, Figura 5.1.

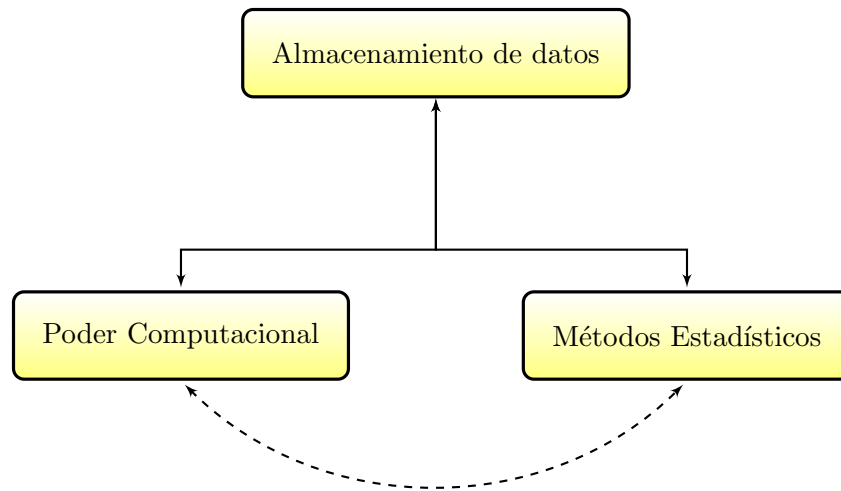


Figura 5.1: Ciclo del aprendizaje automático

Un tema muy relacionado del aprendizaje autónomo, **la minería de datos**, se preocupa por la generación de nuevos conocimientos en grandes bases de datos. La minería de datos implica una búsqueda sistemática de “nuggets” procesables. Aunque existe cierto desacuerdo sobre cómo se superponen el aprendizaje autónomo y la minería de datos, un punto potencial de distinción es que el aprendizaje autónomo se centra en enseñar a las computadoras como usar datos para resolver un problema, mientras que la minería de datos se centra en enseñar a las computadoras la detección de patrones que los humanos luego usan para resolver un problema.

Prácticamente toda la minería de datos implica el uso de aprendizaje autónomo, pero no todo el aprendizaje autónomo implica la extracción de datos. Por ejemplo, se puede aplicar el aprendizaje autónomo a datos de tráfico relacionados con automóviles para encontrar patrones relacionados con tasas de accidentes; por otro lado, si la computadora está aprendiendo cómo conducir el automóvil, esto es puramente aprendizaje autónomo sin extracción de datos.

5.3. Cómo “aprenden” las máquinas

Una definición formal de aprendizaje autónomo fue propuesta por el científico informático **Tom M. Mitchell**, quien afirma:

Que una máquina aprende siempre que sea capaz de utilizar su experiencia para mejorar el rendimiento de experiencias similares en el futuro.

Aunque esta definición es intuitiva, ignora por completo el proceso de cómo la experiencia se puede traducir exactamente en acciones futuras, y por supuesto, **aprender siempre es más fácil decirlo que hacerlo**.

Mientras que los cerebros humanos son naturalmente capaces de aprender desde el nacimiento, las condiciones necesarias para que las computadoras aprendan deben ser explícitas. Por esta razón, aunque no es estrictamente necesario comprender la base teórica del aprendizaje, esta base ayuda a comprender, distinguir e implementar algoritmos de aprendizaje autónomo. Independientemente de si el alumno es un humano o una máquina, el proceso básico de aprendizaje es similar. Se puede dividir en cuatro componentes inter-relacionados:

- **Almacenamiento de datos.-** Utiliza la observación, la memoria y el recuerdo para proporcionar una base fáctica para un razonamiento posterior.
- **Abstracción.-** Implica la traducción de datos almacenados en representaciones y conceptos más amplios.
- **Generalización.-** Usa datos abstractos para crear conocimiento e inferencias que dirigen la acción en nuevos contextos.
- **Evaluación.-** Proporciona un mecanismo de retroalimentación para medir la utilidad del conocimiento aprendido e informar mejoras potenciales.

La Figura 5.2 ilustra los pasos en el proceso de aprendizaje:

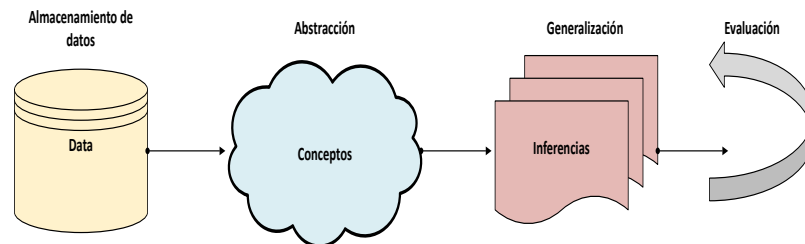


Figura 5.2: Pasos en el proceso de aprendizaje

El proceso de aprendizaje se ha conceptualizado como cuatro componentes distintos, simplemente se organizan de esta manera con fines ilustrativos. En realidad, todo el proceso de aprendizaje está vinculado. En los seres humanos, el proceso ocurre inconscientemente. Recordamos, deducimos, e intuimos de manera mental, y como este proceso está oculto, cualquier diferencia de persona a persona se atribuye a una vaga noción de subjetividad. En contraste, con las computadoras, estos procesos son explícitos, y dado que todo el proceso es transparente, el conocimiento aprendido puede examinarse, transferirse y utilizarse para futuras acciones.

Un proceso de aprendizaje comprensivo y más amplio a tareas del mundo real, implica un proceso de cinco pasos. Independientemente de la tarea en cuestión, cualquier algoritmo de aprendizaje autónomo se puede implementar siguiendo los siguientes pasos que se detallan .

1. **Recopilación de datos.-** El paso de recopilación de datos implica reunir el material de aprendizaje que utilizará un algoritmo para generar conocimiento procesable. En la mayoría de los casos, los datos deberán combinarse en una sola fuente, como un archivo de texto, una hoja de cálculo o una base de datos.
2. **Exploración y preparación de datos.-** La calidad de cualquier proyecto de aprendizaje autónomo se basa en gran medida en la calidad de sus datos de entrada. Por lo tanto, es importante aprender más sobre los datos y sus matices durante una práctica llamada exploración de datos. Se requiere trabajo adicional para preparar los datos para el proceso de aprendizaje. Esto implica arreglar o limpiar los datos llamados “messy”, eliminar datos innecesarios y recodificar los datos para que se ajusten a las entradas esperadas del aprendizaje.
3. **Entrenamiento del modelo.-** Para cuando los datos se hayan preparado para el análisis, es probable que se tenga una idea clara de realizar un proceso de aprendizaje con

los datos preparados. La tarea específica de aprendizaje autónomo elegida informará la selección de un algoritmo apropiado, y el algoritmo representará los datos en forma de modelo.

4. **Evaluación del modelo.-** Debido a que cada modelo de aprendizaje automático resulta en una solución sesgada al problema de aprendizaje, es importante evaluar qué tan bien aprende el algoritmo de su experiencia. Según el tipo de modelo utilizado, es posible que pueda evaluar la precisión del modelo utilizando un conjunto de datos de prueba o desarrollar medidas de rendimiento específicas para la aplicación prevista.
5. **Mejora del modelo.-** Si se necesita un mejor rendimiento, se hace necesario utilizar estrategias más avanzadas para aumentar el rendimiento del modelo. A veces, puede ser necesario cambiar por completo el modelo. Es posible que se necesite complementar los datos con información adicional o realizar trabajos como en el paso dos de este proceso.

5.4. Tipos de datos de entrada

La práctica del aprendizaje autónomo implica hacer coincidir las características de los datos de entrada con los “sesgos” de los enfoques disponibles. Por lo tanto, antes de aplicar el aprendizaje autónomo a los problemas del mundo real, es importante estudiar y analizar la terminología que se distingue entre los conjuntos de datos de entrada.

La **unidad de observación** se emplea para describir la entidad más pequeña con propiedades medibles de interés para un estudio. Comúnmente, la unidad de observación representa personas, objetos o cosas, transacciones, intervalos de tiempo o regiones geográficas. A veces, las unidades de observación se combinan para formar unidades tales como años-persona, que denotan casos en los que se hace un seguimiento de la misma persona durante varios años.

Los conjuntos de datos que almacenan las unidades de observación y sus propiedades se pueden imaginar como colecciones de datos que consisten en:

- **Ejemplos.-** Instancias de la unidad de observación para las que se han registrado propiedades.

- **Características-atributos.-** Propiedades registradas o atributos de ejemplos que pueden ser útiles para el aprendizaje.

Los conceptos antes mencionados, pueden agruparse bajo las siguientes particularizaciones. Para construir un algoritmo de aprendizaje que permita identificar correo electrónico no deseado, la unidad de observación podría ser mensajes de correo electrónico, los ejemplos serían mensajes específicos y las características podrían consistir en las palabras usadas en los mensajes. Para un algoritmo de detección de cáncer, la unidad de observación podría ser pacientes, los ejemplos podrían incluir una muestra aleatoria de pacientes con cáncer, y las características pueden ser los marcadores genómicos de las células de biopsia así como las características del paciente como el peso, la altura o presión sanguínea. Comúnmente los “**ejemplos y atributos**” se recopilan en formato de matriz, lo que significa que cada ejemplo tiene exactamente las mismas características.

5.5. Tipos de algoritmos de aprendizaje automático

Los algoritmos de aprendizaje autónomo se dividen en categorías de acuerdo con su propósito, por tal motivo, el análisis y estudio de las categorías de los algoritmos de aprendizaje es un primer paso esencial hacia el uso de los datos para impulsar la acción deseada.

Un **modelo predictivo** se usa para tareas que implican, como su nombre lo indica, la predicción de un valor usando otros valores en el conjunto de datos, es decir, es un modelo basado en estadísticas inferenciales. El algoritmo de aprendizaje intenta descubrir y modelizar la relación entre la característica objetivo (la característica que se predice) y otras características. A pesar del uso común de la palabra “predicción” para implicar un pronóstico, los modelos predictivos no necesariamente prevén eventos en el futuro. Por ejemplo, un modelo predictivo podría usarse para predecir eventos pasados, como la fecha de concepción de un bebé usando los niveles hormonales actuales de la madre. Los modelos predictivos también se pueden usar en tiempo real para controlar los semáforos durante las horas pico.

Debido a que los modelos predictivos reciben instrucciones claras sobre lo que necesitan aprender y cómo se pretende que lo aprendan, el proceso de entrenamiento de un modelo predictivo

5. APRENDIZAJE AUTOMÁTICO

se conoce como aprendizaje supervisado. La supervisión no se refiere a la participación humana, sino más bien al hecho de que los valores objetivo proporcionan una manera para que la persona que esta implementando el aprendizaje autónomo “**learner**” sepa qué tan bien ha aprendido la tarea deseada. Expuesto más formalmente, dado un conjunto de datos, un algoritmo de aprendizaje supervisado intenta optimizar una función (el modelo) para encontrar la combinación de valores de características que dan como resultado el objetivo buscado o deseado, Figura 5.3.

La tarea de aprendizaje automático supervisado que se usa a menudo para predecir a qué categoría pertenece un ejemplo se conoce como clasificación. Es fácil pensar en usos potenciales para un clasificador. Por ejemplo, se puede predecir si:

- Un mensaje de correo electrónico es spam.
- Una persona tiene cáncer.
- Un equipo de fútbol ganará o perderá.
- Un solicitante incumplirá un préstamo

En la clasificación, la característica objetivo a predecir es un atributo categórico conocido como clase, y se divide en categorías llamadas niveles. Una clase puede tener dos o más niveles, y los niveles pueden ser ordinales o no. Debido a que la clasificación es tan ampliamente utilizada en el aprendizaje automático, existen muchos tipos de algoritmos de clasificación, con fortalezas y debilidades adecuadas para diferentes tipos de datos de entrada.

Un modelo descriptivo se usa para tareas que se beneficiarían del conocimiento adquirido a partir de datos que han sido resumidos mediante formas nuevas e interesantes. A diferencia de los modelos predictivos que predicen un objetivo de interés, en un modelo descriptivo, ninguna característica individual es más importante que cualquier otra. De hecho, debido a que no hay un objetivo para aprender, el proceso de entrenamiento de un modelo descriptivo se denomina aprendizaje no supervisado, Figura 5.3.

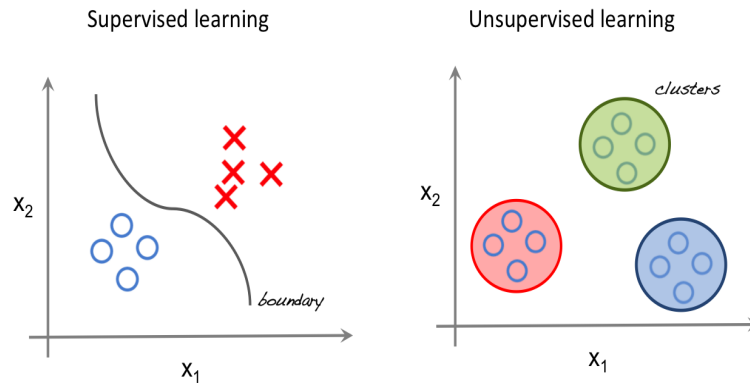


Figura 5.3: Un ejemplo de aprendizaje supervisado y no supervisado

Estas definiciones podría aplicarse también a, la tarea de modelización descriptiva llamada reconocimiento de patrones se usa para identificar asociaciones útiles dentro de los datos. El reconocimiento de patrones a menudo se usa en los datos de compras transaccionales de los minoristas. Aquí, el objetivo es identificar los artículos que con frecuencia se compran juntos, de modo que la información aprendida se pueda utilizar para refinar las tácticas de marketing. Por ejemplo, si un minorista se entera de que los trajes de baños se compran comúnmente al mismo tiempo que las gafas de sol, el minorista podría repositionar los artículos más de cerca en la percha o ejecutar una promoción para “vender” a los clientes los artículos asociados.

La tarea de la modelización descriptiva de dividir un conjunto de datos en grupos homogéneos se denomina agrupamiento “**clustering**”. En el sentido descrito, esto se usa para el **análisis de segmentación** que identifica grupos de personas con un comportamiento similar o información demográfica, de modo que las campañas publicitarias se pueden adaptar para audiencias particulares. Aunque la máquina es capaz de identificar los conglomerados, se requiere la intervención humana para interpretarlos.

Por último, una clase de algoritmos de aprendizaje automático conocidos como **meta-learners** no está ligada a una tarea específica de aprendizaje, sino que está enfocada en **aprender a aprender** de manera más efectiva, que tiene raíces con la psicología cognitiva. Un algoritmo de meta-aprendizaje utiliza el resultado de algunos aprendizajes para informar el aprendizaje adicional. Esto puede ser beneficioso para problemas muy desafiantes o cuando el rendimiento de un algoritmo predictivo debe ser lo más preciso posible.

5.6. Relacionando datos de entrada con algoritmos

Los aspectos clave del aprendizaje automático podrían enumerarse de acuerdo a los algoritmos, de la siguiente forma:

Modelo	Tarea de aprendizaje
Algoritmos de aprendizaje supervisado	
Vecino más cercano	Clasificación
Naive Bayes	Clasificación
Árboles de decisión	Clasificación
Reglas de Clasificación	Clasificación
Regresión Lineal	Predicción numérica
Árboles de regresión	Predicción numérica
Árboles de decisión	Predicción numérica
Redes neuronales	Dual use
Máquinas de soporte vectorial	Dual use
Algoritmos de aprendizaje no supervisado	
Reglas de asociación	Reconocimientos de patrones
k-means clustering	Agrupamiento
Algoritmos de meta-learning	
Agregación de bootstrap (Bagging),	Supervisado y No Supervisado
Impulsar (Boosting)	Supervisado y No Supervisado
Random forest	Supervisado y No Supervisado

Tabla 5.1: Principales tipos de algoritmos de aprendizaje

Los algoritmos de aprendizaje automático están agrupados de acuerdo a la similitud de enfoque y salida del algoritmo. Esto ayudará a crear casos de aplicación para un conjunto más diverso de problemas. Otra ventaja de organizar algoritmos de esta manera es la facilidad de trabajar con las bibliotecas de algunos programas como R por ejemplo, que están diseñadas para contener todas las funciones relevantes / similares en una única biblioteca.

5.6.1. Comprendiendo la clasificación del vecino más cercano

Muchos enfoques intentan estimar la distribución condicional de Y dado X , y luego clasificar una observación dada a la clase con la probabilidad estimada más alta. Uno de estos métodos es el de los vecinos más cercanos (K-nearest neighbors, KNN). Dando un número entero positivo k y una observación de prueba x_0 , el clasificador KNN primero identifica los puntos k en los datos de entrenamiento que son más cercanos a x_0 , representados por η_0 . Luego estima la probabilidad condicional para la clase j como la fracción de puntos en η_0 cuyos valores de respuesta son iguales a j .

$$Pr(Y = j|X = x_0) = \frac{1}{k} \sum_{i \in \eta_0} I(y_i = j), \quad (5.1)$$

donde $I(y_i = j)$ es una variable índice que es igual a 1 si $y_i = j$. En términos generales, los clasificadores de vecinos más cercanos son adecuados para tareas de clasificación, donde las relaciones entre las características y las clases objetivo son numerosas, complicadas o extremadamente difíciles de entender, sin embargo, los elementos del tipo de clase similar tienden a ser bastante homogéneos. Finalmente, $K - NN$ aplica la regla de Bayes y clasifica la observación de la prueba x a la clase con mayor probabilidad.

El algoritmo k-NN(Nearest Neighbors)

El concepto de los vecinos más cercanos para clasificación se puede clarificar mediante el algoritmo k-vecinos más cercanos (k-NN). Aunque este es quizás uno de los algoritmos más sencillos de aprendizaje automático, todavía se usa ampliamente. Las fortalezas y debilidades del algoritmo son las siguientes, Tabla 5.2.

Fortalezas	Debilidades
Simple y efectivo	No produce un modelo, lo que limita la capacidad de comprender cómo se relacionan las características con la clase.
No hace suposiciones sobre la distribución subyacente de los datos.	Requiere la selección de una k apropiada
Fase de entrenamiento rápido	Fase de clasificación lenta
	Las características nominales y los datos faltantes requieren procesamiento adicional

Tabla 5.2: Fortalezas y Debilidades del algoritmo K-NN

El algoritmo $k - NN$ recibe su nombre del hecho de que utiliza información de los k vecinos más cercanos, lo cual le permite etiquetar ejemplos que no han sido clasificados. La letra k es una variable que indica el número de vecinos más cercanos. Después de elegir k , el algoritmo requiere un conjunto de datos de entrenamiento compuesto por ejemplos que se han clasificado en varias categorías, mediante una variable nominal. Luego, para cada registro no etiquetado en el conjunto de datos de prueba, $k - NN$ identifica k registros en los datos de entrenamiento que son los “más cercanos” en similitud.

Medición de similitud con la distancia

El algoritmo $k - NN$ utiliza la distancia euclidiana, la cual se especifica mediante la Ecuación 5.2, donde p y q son los “ejemplos” que se compararán, cada uno con n características. El término p_1 se refiere al valor de la primera característica del ejemplo p , mientras que q_1 se refiere a la primera característica del ejemplo q . La Ecuación 5.2 implica comparar los valores de cada característica.

$$dist(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (5.2)$$

Elegir un k apropiado

La decisión de cuántos vecinos usar para $k - NN$ determina qué tan bien se generalizará el modelo para datos futuros. El equilibrio entre el sobre-ajuste y el sub-ajuste de los datos de entrenamiento es un problema conocido como compensación sesgo-varianza (bias-variance tradeoff).

Elegir una k grande reduce el impacto o la varianza causada por los datos ruidosos, pero puede sesgar el aprendizaje corriendo el riesgo de ignorar patrones pequeños pero importantes. Al considerar un valor muy grande de k , tan grande como el número total de observaciones en los datos de entrenamiento, dará como resultado que la clase más común siempre será la más representativa. En consecuencia, el modelo siempre predeciría la clase con mayor frecuencia que se repita, independientemente de los vecinos más cercanos. En el extremo opuesto, el uso de un único vecino más cercano permite que los datos ruidosos o valores atípicos influyan indebidamente en la clasificación de los ejemplos. Obviamente, el mejor valor k está en algún lugar entre estos dos extremos.

La Figura 5.4, de manera general, muestra cómo el límite de decisión (representado por una línea recta o curva) se ve afectado por valores k más grandes o más pequeños. Los valores más pequeños permiten límites de decisión más complejos que se ajustan mejor a los datos de entrenamiento. La figura ilustra, que la representación del verdadero concepto subyacente no está claramente establecido por el límite recto o curvo.

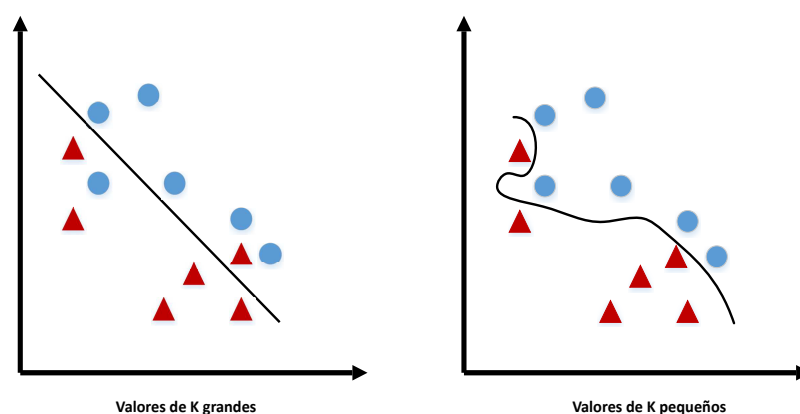


Figura 5.4: Límite de decisión algoritmo K-NN, k “grande” y “pequeño”

En la práctica, la elección de k depende de la dificultad del concepto a aprender y del número

de registros en los datos de entrenamiento. Una práctica común es comenzar con k igual a la raíz cuadrada del número de ejemplos de entrenamiento. Sin embargo, tales reglas pueden no siempre dar como resultado el mejor valor de k . Un enfoque alternativo es probar varios valores k en una variedad de conjuntos de datos de prueba y elegir el que ofrezca el mejor rendimiento de clasificación.

Preparación de datos para usar el algoritmo k-NN

Las características se transforman típicamente en valores estándares (adimensional) antes de aplicar el algoritmo k-NN. La razón de este paso es que la fórmula de distancia depende en gran medida de cómo se miden las características. En particular, si ciertas características tienen valor mucho más grande que las otras, las mediciones de distancia estarán fuertemente dominadas por las características con rangos mayores. La solución es “re-escalar” las características reduciendo o expandiendo su rango de modo que cada una contribuya de manera relativamente igual a la fórmula de distancia. El método tradicional de re-escalar características para k-NN es la normalización **min-max**. Este proceso transforma una característica tal que todos sus valores caen en un rango entre 0 y 1. La fórmula para normalizar una característica se expresa en la Ecuación 5.3.

$$X_{nueva} = \frac{X - \text{mín}(X)}{\text{máx}(X) - \text{mín}(X)} \quad (5.3)$$

Otra transformación importante es “z-score”, que se define como, la resta del valor medio de la característica X y dividida por la desviación estándar de X , Ecuación 5.4.

$$X_{nueva} = \frac{X - \mu}{\sigma} = \frac{X - \text{mean}(X)}{\text{StdDev}(X)} \quad (5.4)$$

¿Por qué el algoritmo k-NN es “lazy”?

Los algoritmos de clasificación basados en K-NN se consideran algoritmos de aprendizaje “lazy” porque, técnicamente, no se produce abstracción. Bajo la definición estricta de aprendizaje, un aprendizaje “lazy” no está realmente aprendiendo nada. Simplemente almacena los datos de entrenamiento textualmente. Esto permite que la fase de entrenamiento, que en realidad no

está entrenando nada, ocurra muy rápidamente. Por supuesto, la desventaja es que el proceso de hacer predicciones tiende a ser relativamente lento en comparación con el entrenamiento.

Como el aprendizaje basado en instancias no construye un modelo, se dice que el método está en una clase de métodos de aprendizaje no paramétricos, no se conocen parámetros sobre los datos. Sin generar teorías sobre los datos subyacentes, los métodos no paramétricos limitan nuestra capacidad para comprender cómo el clasificador está utilizando los datos. Por otro lado, esto permite al “learner” encontrar patrones naturales en lugar de intentar ajustar los datos en una forma funcional preconcebida y potencialmente sesgada.

5.6.2. Aprendizaje probabilístico-Clasificación usando Naive Bayes

Supóngase que se desea clasificar una observación en una de las k clases, donde $k \geq 2$. En otras palabras, la variable respuesta Y puede asumir k posibles valores. Sea π_k la probabilidad a priori de que una observación esté asociada con la categoría k de la variable respuesta. Sea además, $f_k(x) = Pr(X = x|Y = k)$ ¹ que denota la función de densidad de X para una observación que proviene de la categoría k . Luego el Teorema de Bayes establece que:

$$Pr(Y = k|X = x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^K \pi_l f_l(x)} \quad (5.5)$$

Los clasificadores basados en métodos Bayesianos utilizan datos de entrenamiento para calcular una probabilidad observada de cada resultado en función de la evidencia provista por los valores de las características. Cuando el clasificador se aplica posteriormente a datos sin etiqueta, usa las probabilidades observadas para predecir la clase más probable para las nuevas características. Es una idea simple, pero resulta en un método que a menudo tiene resultados a la par con algoritmos más sofisticados. De hecho, los clasificadores Bayesianos se han utilizado para:

- Clasificación de texto, como el filtrado de correo electrónico no deseado (spam).
- Detección de intrusiones o anomalías en redes informáticas.
- Diagnóstico de condiciones médicas dado un conjunto de síntomas observados.

¹Técnicamente esta definición sólo es correcta si X es una variable aleatoria discreta. Si X es continua entonces $f_k(x) dx$ correspondería a la probabilidad de que X caiga en una pequeña región dx alrededor de x .

Normalmente, los clasificadores Bayesianos se aplican mejor a problemas en los que la información de numerosos atributos se debe considerar simultáneamente para estimar la probabilidad general de un resultado. Si bien muchos algoritmos de aprendizaje autónomo ignoran las características que tienen efectos débiles, los métodos bayesianos utilizan toda la evidencia disponible para cambiar sutilmente las predicciones. Si gran cantidad de características tienen efectos relativamente menores, en conjunto, su impacto combinado podría ser bastante grande.

El algoritmo de Naive Bayes

El algoritmo de Naive Bayes describe un método para aplicar el teorema de Bayes a los problemas de clasificación. Aunque no es el único método de aprendizaje automático que utiliza métodos Bayesianos, pero es el más común. Las fortalezas y debilidades de este algoritmo se detallan en la Tabla 5.3.

Fortalezas	Debilidades
Simple, rápido y muy efectivo.	Se basa en una suposición a menudo defectuosa de características igualmente importantes e independientes.
Lo hace bien con datos ruidosos y faltantes.	No es ideal para conjuntos de datos con muchas características numéricas.
Requiere relativamente pocos ejemplos para entrenamiento, pero también funciona bien con un gran número de ejemplos	Las probabilidades estimadas son menos confiables que las clases predichas
Fácil de obtener la probabilidad estimada para una predicción	

Tabla 5.3: Fortalezas y Debilidades del algoritmo Naive Bayes

El algoritmo de Naive Bayes se denomina así porque hace suposiciones “ingenuas” sobre los datos. En particular, el algoritmo supone que todas las características del conjunto de datos son igualmente importantes e independientes. Estas suposiciones rara vez son ciertas en la mayoría de las aplicaciones del mundo real. Sin embargo, en la mayoría de los casos cuando se violan

estas suposiciones, Naive Bayes todavía se desempeña bastante bien. Esto es cierto incluso en circunstancias extremas en las que se encuentran fuertes dependencias entre las características. Debido a la versatilidad y precisión del algoritmo en muchos tipos de condiciones, Naive Bayes suele ser un fuerte candidato para las tareas de aprendizaje de clasificación.

Clasificación con Naive Bayes

El algoritmo Naive Bayes asume la independencia entre los eventos. Específicamente, supone independencia condicional de clase, lo que significa que los eventos son independientes siempre que estén condicionados al mismo valor de clase. Asumir independencia condicional permite simplificar la fórmula usando la regla de probabilidad para eventos independientes, que establece que $P(A \cap B) = P(A) \cdot P(B)$.

Luego, la clasificación de Naive Bayes se puede resumirse mediante la fórmula 5.6. Es decir, la probabilidad del nivel L para la clase C , dada la evidencia provista por las características F_1 a F_n , es igual al producto de las probabilidades de cada evidencia condicionada al nivel de clase, la probabilidad previa del nivel de clase y un factor de escala $\frac{1}{Z}$, que convierte los valores de verosimilitud en probabilidades:

$$P(C_L|F_1, F_2, \dots, F_n) = \frac{1}{Z} p(C_L) \prod_{i=1}^n p(F_i|C_L) \quad (5.6)$$

El estimador de Laplace

Si la probabilidad de la Ecuación 5.6 es cero, existe un problema, y esto puede surgir si un evento nunca ocurre para uno o más niveles de la clase. Una solución a este problema implica usar algo llamado el estimador de **Laplace**, que lleva el nombre del matemático francés Pierre-Simon Laplace. El estimador de Laplace básicamente agrega un pequeño número a cada uno de los recuentos en la tabla de frecuencias, lo que garantiza que cada característica tenga una probabilidad distinta de cero de ocurrir con cada clase. Por lo general, el estimador de Laplace se establece en 1, lo que garantiza que cada combinación de características de clase se encuentre en los datos al menos una vez.

5.6.3. Comprender los árboles de decisión

Los árboles de decisión son clasificadores potentes, que utilizan una estructura jerárquica para modelizar las relaciones entre las características y los posibles resultados. Es decir, es un método analítico que a través de una representación esquemática de las alternativas disponible, que facilita la toma de mejores decisiones, especialmente cuando existen riesgos, costos, beneficios y múltiples opciones. El nombre se deriva de la apariencia del modelo parecido a un árbol y su uso es amplio en el ámbito de la toma de decisiones bajo incertidumbre.

Como se ilustra en la Figura 5.5, esta estructura ganó su nombre debido al hecho de que refleja cómo un árbol literal comienza en un tronco ancho, que, si se sigue hacia arriba, se divide en ramas cada vez más estrechas. De la misma manera, un clasificador de árbol de decisiones utiliza una estructura de decisiones de ramificación, que canaliza ejemplos en un valor de clase predicho final.

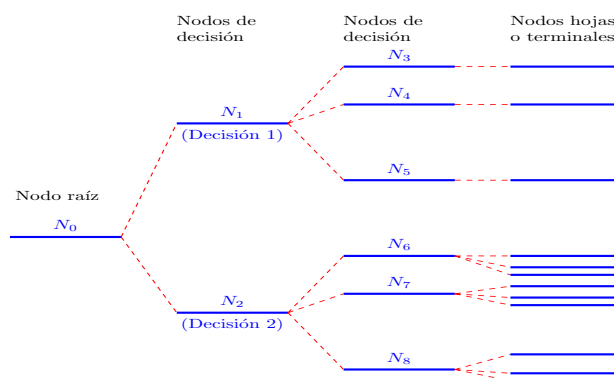


Figura 5.5: Árbol de decisión

El algoritmo de decisión C5.0

El algoritmo C5.0 utiliza algoritmos de división comunes que incluyen entropía basada en la obtención de información. La relación de ganancia es robusta y consistentemente da una mejor elección de pruebas para un gran conjunto de datos, en relación con otros algoritmos Elsalamy and Elsayad (50). El modelo funciona dividiendo la muestra en función del atributo que proporciona la máxima ganancia de información. Se construyen submuestras, normalmente basada en un atributo diferente, luego este proceso se repite hasta que las submuestras no puedan

dividirse más. Finalmente, se reexaminan las divisiones de nivel más bajo y se eliminan o podan las que no contribuyen significativamente al valor del modelo.

Existen numerosas implementaciones de árboles de decisión, pero una de las implementaciones más conocidas es el algoritmo *C5.0*. Este algoritmo fue desarrollado por el informático J. Ross Quinlan como una versión mejorada de su algoritmo anterior, *C4.5*, Salzberg (164) que a su vez es una mejora sobre su algoritmo Iterative Dichotomiser 3 (ID3).

El algoritmo *C5.0* se ha convertido en el estándar de la industria para producir árboles de decisión, porque funciona bien en la mayoría de los casos. En comparación con otros modelos avanzados de aprendizaje automático, como Neural Networks y Support Vector Machines, los árboles de decisión construidos por *C5.0* generalmente tienen un rendimiento similar, pero son mucho más fáciles de comprender e implementar. Además, como se muestra en la Tabla 5.4, las debilidades del algoritmo son relativamente menores y se pueden evitar en gran medida.

5. APRENDIZAJE AUTOMÁTICO

Fortalezas	Debilidades
Un clasificador multi-propósito que funciona bien en la mayoría de los problemas.	Los modelos de árbol de decisión a menudo están sesgados hacia divisiones en características que tienen una gran cantidad de niveles.
Proceso de aprendizaje altamente automático, que puede manejar características numéricas o nominales, así como datos faltantes.	Es fácil sobre-ajustar o sub-ajustar el modelo.
Excluye características no relevantes	Puede tener problemas para modelar algunas relaciones debido a la dependencia de divisiones paralelas al eje
Se puede usar en grandes y pequeños volúmenes de datos	Pequeños cambios en los datos de entrenamiento pueden generar grandes cambios en la lógica de decisión
Resultados en un modelo que puede interpretarse sin un fondo matemático (para árboles relativamente pequeños)	Los árboles grandes pueden ser difíciles de interpretar y las decisiones que toman pueden parecer contradictorias
Más eficiente que otros modelos complejos	

Tabla 5.4: Fortalezas y Debilidades del algoritmo C5.0

Elegir la mejor división

Hay varias medidas de pureza que se pueden usar para identificar al mejor candidato para dividir el árbol de decisión. C5.0 usa entropía, un concepto tomado de la teoría de la información que cuantifica la aleatoriedad o desorden dentro de un conjunto de valores de clase. Los conjuntos con alta entropía son muy diversos y brindan poca información sobre otros elementos que también pueden pertenecer al conjunto, ya que no existe una similitud aparente. El árbol de decisiones espera encontrar divisiones que reduzcan la entropía y, en última instancia, aumenten

la homogeneidad dentro de los grupos.

Típicamente, la entropía se mide en **bits**. Si solo hay dos clases posibles, los valores de entropía pueden variar de 0 a 1. Para n clases, la entropía varía de 0 a $\log(n)$. En cada caso, el valor mínimo indica que la muestra es completamente homogénea, mientras que el valor máximo indica que los datos son tan diversos como sea posible, y ningún grupo tiene incluso una pequeña pluralidad. En términos matemáticos, la entropía se especifica mediante la ecuación 5.7.

$$Entropía(S) = \sum_{i=1}^c -p_i \log_2(p_i) \quad (5.7)$$

En la ecuación 5.7, para un segmento dado de datos (S), el término c se refiere al número de niveles de clase y p_i se refiere a la proporción de valores que caen en el nivel de clase i .

Se emplea entropía para determinar la característica óptima a dividir, el algoritmo calcula el cambio en la homogeneidad que resultaría de una división en cada característica posible, que es una medida conocida como ganancia de información. La ganancia de información para una característica F se calcula como la diferencia entre la entropía en el segmento antes de la división (S_1) y las particiones resultantes de la división (S_2), ecuación 5.8.

$$InfoGain(F) = Entropía(S_1) - Entropía(S_2) \quad (5.8)$$

Una complicación es que después de una división, los datos se dividen en más de una partición. Por lo tanto, la función para calcular “Entropía (S)” necesita considerar la entropía total en todas las particiones. Esto se puede establecer mediante la ecuación 5.9

$$Entropía(S) = \sum_{i=1}^n Entropía(p_i) \quad (5.9)$$

En términos simples, la entropía total de la suma de la entropía de cada una de las n particiones ponderadas por la proporción de ejemplos que caen en la partición (w_i).

Cuanto mayor sea la ganancia de información, se crearán grupos más homogéneos después de dividir la característica. Si la ganancia de información es cero, no hay reducción en la entropía para dividir esta característica. Por otro lado, la ganancia de información máxima es igual a la

entropía anterior a la división. Esto implicaría que la entropía después de la división es cero, lo que significa que la división da como resultado grupos completamente homogéneos.

Las fórmulas anteriores asumen características nominales, pero los árboles de decisión también usan la ganancia de información para dividir las características numéricas. Para hacerlo, una práctica común es probar varias divisiones que dividen los valores en grupos mayores o menores que un umbral numérico. Esto reduce la característica numérica a una característica categórica de dos niveles que permite que la ganancia de información se calcule como de costumbre. El punto de corte numérico que produce la mayor ganancia de información se elige para la división.

5.6.4. Comprender las reglas de clasificación

Las reglas de clasificación representan el conocimiento en forma de declaraciones lógicas *if-else* que asignan una clase a ejemplos sin etiqueta. Se especifican en términos de un antecedente y un consecuente; estos forman una hipótesis que establece que “si esto sucede, entonces eso sucede”. Una regla simple podría indicar, “si el disco duro está haciendo un clic, entonces está a punto de fallar”. El antecedente comprende ciertas combinaciones de valores de características, mientras que el consecuente especifica el valor de clase a asignar cuando se cumplen las condiciones de la regla.

El aprendizaje de las reglas de clasificación a menudo se usa de manera similar al aprendizaje de los árboles de decisión. Al igual que los árboles de decisión, se pueden usar para aplicaciones que generan conocimiento para acciones futuras, tales como:

- Identificar las condiciones que conducen a una falla de hardware en dispositivos mecánicos.
- Describir las características clave de los grupos de personas para la segmentación del cliente.
- Encontrar condiciones que preceden a grandes caídas o aumentos en los precios de las acciones en el mercado de valores.

Por otro lado, las reglas de aprendizaje ofrecen algunas ventajas claras sobre los árboles para algunas tareas. A diferencia de un árbol, que debe aplicarse de arriba a abajo a través de una serie de decisiones, las reglas son proposiciones que pueden leerse como una declaración

de hecho. Además, los resultados de una regla de decisión pueden ser más simples, directos y fáciles de entender que un árbol de decisión basado en los mismos datos.

Los algoritmos de reglas de clasificación utilizan una **heurística** conocida como separada y conquista. El proceso implica identificar una regla que cubre un subconjunto de ejemplos en los datos de entrenamiento y luego separar esta partición de los datos restantes. A medida que se agregan las reglas, los subconjuntos adicionales de los datos se separan hasta que se cubre todo el conjunto de datos y no quedan más ejemplos.

El algoritmo 1R

El algoritmo 1R es simple en la mayoría de los aspectos. Clasifica los atributos según la tasa de error (en el conjunto de entrenamiento), a diferencia de las medidas basadas en la entropía utilizadas en el algoritmo *C4*. Trata todos los atributos valorados numéricamente como continuos y utiliza un método directo para dividir el rango de valores en varios intervalos y maneja los valores perdidos tratándolos como un valor legítimo, Holte (88).

Dicho algoritmo, es una mejora al algoritmo ZeroR (cero reglas) ya que selecciona una sola regla. Aunque esto puede parecer demasiado simplista, tiende a funcionar mejor de lo que cabría esperar. La precisión de este algoritmo puede acercarse a la de los algoritmos mucho más sofisticados para muchas tareas del mundo real. Las fortalezas y debilidades se muestra en la Tabla 5.5.

Fortalezas	Debilidades
Genera una regla general única, fácil de entender y legible para los humanos	Utiliza solo una característica.
A menudo funciona muy bien	Probablemente demasiado simplista
Puede servir como punto de referencia para algoritmos más complejos	

Tabla 5.5: Fortalezas y Debilidades del algoritmo 1R

La forma en que funciona este algoritmo es simple. Para cada característica, 1R divide los datos en grupos en función de valores similares de la característica. Luego, para cada segmento,

5. APRENDIZAJE AUTOMÁTICO

el algoritmo predice la clase de la mayoría. La tasa de error para la regla basada en cada característica se calcula y la regla con el menor número de errores se elige como la única regla.

Obviamente, este algoritmo de aprendizaje de reglas puede ser demasiado básico para algunas tareas, por lo que una regla más sofisticada podría ser útil.

El algoritmo RIPPER

Ripper (Repeated Incremental Pruning to Produce Error Reduction) Cohen (28) es una regla de inducción basada en la reducción de errores de poda Fürnkranz and Widmer (58). Ripper incluye tres etapas de procesamiento de reglas: construcción, optimización y limpieza Witten et al. (205). En la primera etapa, un conjunto de datos de entrenamiento se divide en conjuntos de crecimiento y poda. Las reglas se construyen sobre la base de un conjunto creciente. Posteriormente, estas reglas se reducen con la ayuda de un juego de poda. En la etapa de optimización, todas las reglas son reconsideradas para reducir el error en todo el conjunto de datos. En la fase de limpieza, se verifica si cada regla de longitud de descripción (DL) aumenta el conjunto completo de reglas y el conjunto de datos. El cálculo de DL permite cuantificar la complejidad de varios conjuntos de reglas, que describen un conjunto de datos ?). En caso de que la regla aumente la DL, se borra.

Como se describe en la Tabla 5.6, las fortalezas y debilidades de RIPPER son generalmente comparables a los árboles de decisión. El principal beneficio es que pueden dar como resultado un modelo un poco más parsimonioso.

Fortalezas	Debilidades
Genera reglas fáciles de entender y legibles para el ser humano	Puede dar lugar a reglas que parecen desafiar el sentido común o el conocimiento experto
Eficiente en conjuntos de datos grandes y ruidosos	No es ideal para trabajar con datos numéricos
Generalmente produce un modelo más simple que un árbol de decisión comparable	Podría no funcionar tan bien con modelos más complejos

Tabla 5.6: Fortalezas y Debilidades del algoritmo RIPPER

El algoritmo RIPPER puede crear reglas mucho más complejas que 1R, ya que considera más de una característica. Esto significa que puede crear reglas con múltiples antecedentes. Esto mejora la capacidad del algoritmo para modelar datos complejos, pero al igual que los árboles de decisión, significa que las reglas pueden volverse cada vez más difíciles de comprender.

5.6.5. Métodos Black Box - Redes Neuronales y Máquinas de Soporte Vectorial

Redes Neuronales

Una red neuronal es un modelo de regresión de dos etapas o un modelo de clasificación, típicamente representado por un diagrama de red. Para la clasificación de la clase K , existen K unidades, con la k -ésima unidad modelizando la probabilidad de la clase K . Asimismo, existen K variables respuesta denotadas por Y_k , $k = 1, 2, \dots, K$, cada una de ellas siendo codificadas 0 o 1 para la k -ésima clase.

Se construyen Z_m variables como combinación lineal de las variables predictoras, luego las variables de respuestas Y_k son modelizadas como funciones de combinaciones lineales de Z_m ,

$$Z_m = \sigma(\alpha_{0m} + \alpha_m^T X), \quad m = 1, 2, \dots, M, \quad (5.10)$$

$$T_k = \beta_{0k} + \beta_k^T Z, \quad k = 1, 2, \dots, K, \quad (5.11)$$

$$f_k(X) = g_k(T), \quad k = 1, 2, \dots, K \quad (5.12)$$

donde $Z = (Z_1, Z_2, \dots, Z_M)$, y $T = (T_1, T_2, \dots, T_K)$. La función de activación $\sigma(\nu)$ es usualmente la función sigmoide. La función $g_k(T)$ permite una transformación final del vector T ; para el caso de regresión normalmente se elige la función de identidad $g_k(T) = T$.

Z_m , se denominan unidades ocultas porque los valores Z_m no se observan directamente. Podemos pensar en Z_m como una expansión básica de las entradas originales X ; la red neuronal es entonces un modelo lineal estándar, o modelo lineal multilogit, usando estas transformaciones como entradas. Nótese que si σ es la función de identidad, entonces el modelo completo colapsa a un modelo lineal en las entradas. Por lo tanto, se puede pensar en una red neuronal como una generalización no lineal del modelo lineal, tanto para la regresión como para clasificación.

5. APRENDIZAJE AUTOMÁTICO

En otras palabras, una Red Neuronal Artificial (RNA) modeliza la relación entre un conjunto de señales de entrada y una señal de salida usando un modelo derivado de la comprensión de cómo un cerebro biológico responde a los estímulos sensoriales. Así como un cerebro usa una red de células inter-conectadas llamadas neuronas para crear un procesador paralelo masivo, RNA usa una red de neuronas o nodos artificiales para resolver problemas de aprendizaje. La complejidad de las RNA también ha aumentado tanto que ahora se aplican con frecuencia a problemas más prácticos, como:

- Programas de reconocimiento de voz y escritura a mano, como los utilizados por los servicios de transcripción de correo de voz y las máquinas clasificadoras de correo postal.
- La automatización de dispositivos inteligentes, como los controles ambientales de un edificio de oficinas o los autos sin conductor y los “drones” auto-guiados.
- Modelos sofisticados de patrones climáticos, resistencia a la tensión, dinámica de fluidos y muchos otros fenómenos científicos, sociales o económicos.

En términos generales, las RNA son algoritmos versátiles que se pueden aplicar a casi cualquier tarea de aprendizaje: clasificación, predicción numérica e incluso reconocimiento de patrones no supervisados.

Las RNA se aplican mejor a problemas donde los datos de entrada y los datos de salida están bien definidos o al menos son bastante simples, sin embargo, el proceso que relaciona la entrada con la salida es extremadamente complejo.

De neuronas biológicas a artificiales

Debido a que las RNA se diseñaron intencionalmente como modelos conceptuales de la actividad del cerebro humano, es útil primero comprender cómo funcionan las neuronas biológicas. Como se ilustra en la Figura 5.6, las señales entrantes son recibidas por las dendritas de la célula a través de un proceso bioquímico. El proceso permite ponderar el impulso según su importancia relativa o frecuencia. A medida que el cuerpo de la célula comienza a acumular las señales entrantes, se alcanza un umbral en el que la célula se dispara y la señal de salida se transmite a través de un proceso electro-químico por el axón. En las terminales del axón, la señal eléctrica se procesa nuevamente como una señal química que se pasa a las neuronas vecinas a través de un pequeño espacio conocido como sinapsis.

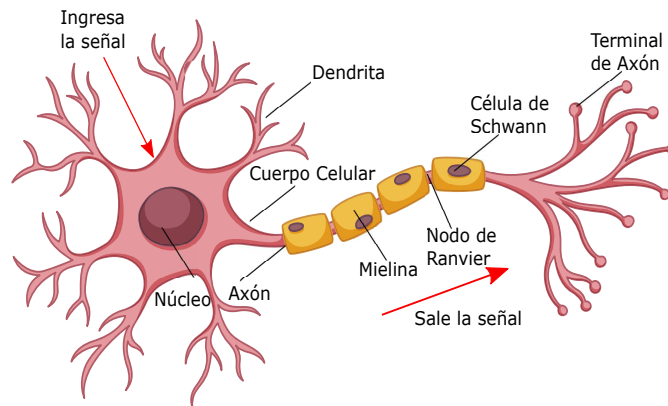


Figura 5.6: Funcionamiento de una neurona

El modelo de una sola neurona artificial se puede entender en términos muy similares al modelo biológico. Como se muestra en la Figura 5.7, un diagrama de red dirigido define una relación entre las señales de entrada recibidas por las dendritas (variables X) y la señal de salida (variable y). Al igual que con la neurona biológica, la señal de cada dendrita se pondera (valores W) de acuerdo con su importancia. Las señales de entrada son sumadas por el cuerpo de la célula y la señal se transmite de acuerdo con una función de activación indicada por f :

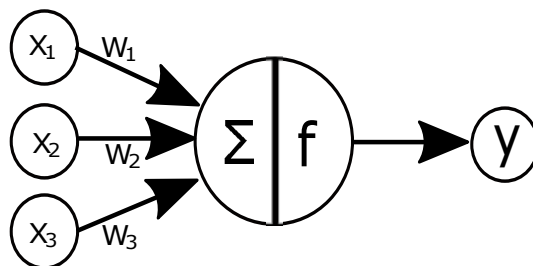


Figura 5.7: Funcionamiento de una neurona artificial

Una neurona artificial típica con n dendritas de entrada puede representarse mediante la Ecuación 5.13. Los w_i pesos permiten que cada una de las n entradas (denotadas por x_i) contribuyan una cantidad mayor o menor a la suma de las señales de entrada. El total neto es utilizado por la función de activación $f(x)$, y la señal resultante, $y(x)$, es el axón de salida.

$$y(x) = f\left(\sum_{i=1}^n w_i x_i\right) \quad (5.13)$$

Las redes neuronales usan neuronas definidas de esta manera como bloques de construcción para

5. APRENDIZAJE AUTOMÁTICO

construir modelos complejos de datos. Aunque existen numerosas variantes de redes neuronales, cada una se puede definir en términos de las siguientes características:

- Una función de activación, que transforma las señales de entrada combinadas de una neurona en una sola señal de salida para ser transmitidas más allá en la red.
- Una topología de red (o arquitectura), que describe el número de neuronas en el modelo, así como el número de capas y la forma en que están conectadas.
- El algoritmo de entrenamiento que especifica cómo se establecen los pesos de conexión para inhibir o “estimular” las neuronas en proporción a la señal de entrada.

Funciones de activación

La función de activación es el mecanismo por el cual la neurona artificial procesa la información entrante y la pasa a través de la red. Así como la neurona artificial se modeliza a partir de la versión biológica, también lo es la función de activación modelada según el diseño de la naturaleza.

En el caso biológico, la función de activación podría imaginarse como un proceso que implica sumar la señal de entrada total y determinar si cumple el cierto umbral. Si es así, la neurona pasa la señal; de lo contrario, no hace nada. En términos generales, esto se conoce como una función de activación de umbral, ya que da como resultado una señal de salida solo una vez que se ha alcanzado un umbral de entrada especificado. La Figura 5.8 representa una función umbral típica; en este caso, la neurona se dispara cuando la suma de las señales de entrada es al menos cero. Debido a que su forma se asemeja a una escalera, a veces se denomina función de activación por pasos.

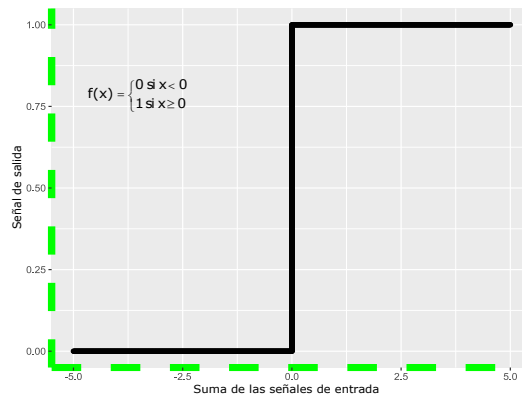


Figura 5.8: Función de activación por pasos

Aunque la función de activación del umbral es interesante debido a su paralelismo con la biología, rara vez se usa en redes neuronales artificiales. Liberadas de las limitaciones de la bioquímica, las funciones de activación se pueden elegir en función de su capacidad para demostrar características matemáticas deseables y modelar con precisión las relaciones entre los datos.

Una de la alternativa más comúnmente utilizada es la función de activación **sigmoidea** (más específicamente, el “sigmoide logístico”) que se muestra en la Figura 5.9. Tenga en cuenta que en la fórmula que se muestra, e es la base del logaritmo natural (aproximadamente 2.72). Su forma es como una “S” y guarda cierta similitud con la función de activación de umbral, la señal de salida ya no es binaria; los valores de salida pueden caer en cualquier lugar en el rango de 0 a 1. Además, el “sigmoide” es diferenciable, lo que significa que es posible calcular la derivada en todo el rango de entradas, característica crucial para crear algoritmos de optimización (RNA) eficientes.

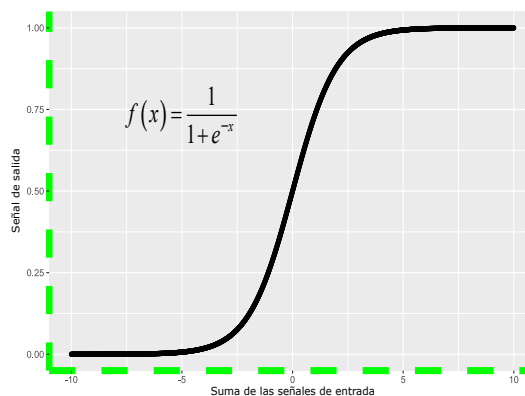


Figura 5.9: Función de activación por pasos

5. APRENDIZAJE AUTOMÁTICO

Aunque “**sigmoide**” es quizás la función de activación más comúnmente utilizada y, a menudo, se utiliza de forma predeterminada, algunos algoritmos de redes neuronales permiten una selección de alternativas. Una selección de tales funciones de activación se muestra en la Figura 5.10.

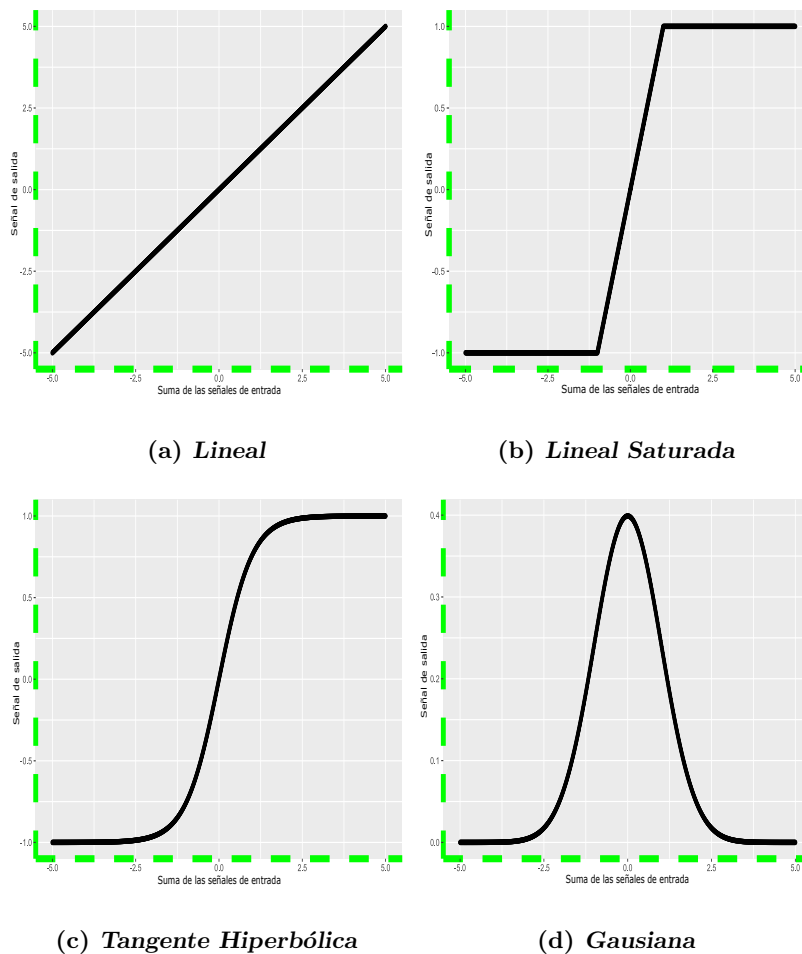


Figura 5.10: Tipos de funciones de activación

El detalle principal que diferencia estas funciones de activación es el rango de la señal de salida. Típicamente, este es uno de $(0, 1)$, $(-1, +1)$ o $(-\infty, +\infty)$. La elección de la función de activación polariza la red neuronal de modo que pueda adaptarse a ciertos tipos de datos de manera más adecuada, lo que permite la construcción de redes neuronales especializadas. Por ejemplo, una función de activación lineal da como resultado una red neuronal muy similar a un

modelo de regresión lineal, mientras que una función de activación gaussiana da como resultado un modelo llamado red de función de base radial (RBF). Cada uno de estas funciones tienen fortalezas más adecuadas para ciertas tareas de aprendizaje y otras no.

Es importante reconocer que para muchas de las funciones de activación, el rango de valores de entrada que afectan la señal de salida es relativamente **“estrecho”**. Por ejemplo, en el caso de función **sigmoide**, la señal de salida es siempre casi 0 o 1 para una señal de entrada por debajo de -5 o por encima de $+5$, respectivamente. La compresión de la señal de esta manera da como resultado una señal saturada en los extremos alto y bajo, del mismo modo que al encender un amplificador de guitarra se obtiene un sonido distorsionado debido al recorte de los picos de las ondas de sonido. Debido a que esto esencialmente comprime los valores de entrada en un rango más pequeño de salidas, las funciones de activación como la sigmoide a veces se denominan funciones de **“aplastamiento”**.

La solución al problema de aplastamiento es transformar todas las entradas de red neuronal de modo que los valores de las características caigan dentro de un rango pequeño alrededor de 0. Típicamente, esto implica estandarizar o normalizar las características. Al restringir el rango de valores de entrada, la función de activación tendrá acción en todo el rango, evitando que las características de gran valor como el ingreso del hogar dominen las características de pequeño valor, como la cantidad de niños en el hogar. Un beneficio lateral es que el modelo también puede ser más rápido de entrenar, ya que el algoritmo puede iterar más rápidamente a través del rango de valores de entrada.

Topología de la red

La capacidad de una red neuronal para aprender está enraizada en su topología, o en los patrones y estructuras de las neuronas interconectadas. Aunque existen innumerables formas de arquitectura de red, se pueden diferenciar por tres características clave:

- El número de capas
- Si la información en la red puede viajar hacia atrás
- La cantidad de nodos dentro de cada capa

La topología determina la complejidad de las tareas que la red puede aprender. Generalmente,

las redes más grandes y más complejas son capaces de identificar patrones más sutiles y límites de decisión complejos. Sin embargo, la potencia de una red no solo depende del tamaño de la red, sino también de la forma en que se organizan las unidades.

El número de capas

Para definir la topología, necesitamos una terminología que distinga las neuronas artificiales en función de su posición en la red. La Figura 5.11 ilustra la topología de una red muy simple. Un conjunto de neuronas llamadas nodos de entrada recibe señales no procesadas directamente de los datos de entrada. Cada nodo de entrada es responsable de procesar una sola característica en el conjunto de datos; el valor de la característica será transformado por la función de activación del nodo correspondiente. Las señales enviadas por los nodos de entrada son recibidas por el nodo de salida, que usa su propia función de activación para generar una predicción final (indicada aquí como p).

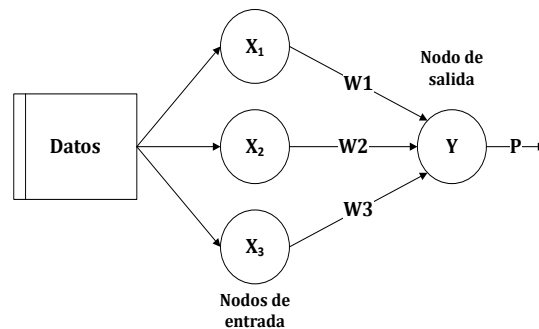


Figura 5.11: Topología de una red neuronal simple

Una forma obvia de crear redes más complejas es agregando capas adicionales. Como se muestra en la Figura 5.12, una red multicapa agrega una o más capas ocultas que procesan las señales de los nodos de entrada antes de llegar al nodo de salida. La mayoría de las redes de múltiples capas están completamente conectadas, lo que significa que cada nodo en una capa está conectado a cada nodo en la siguiente capa, pero esto no es necesario.

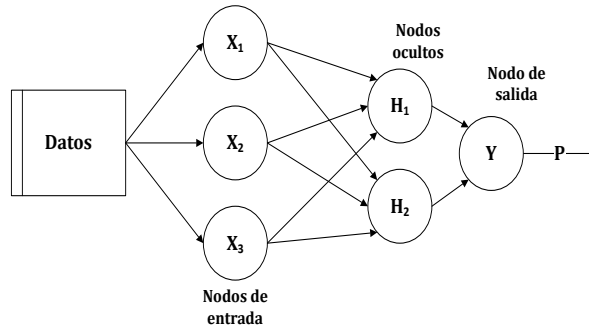


Figura 5.12: Topología de una red neuronal compleja

La dirección del viaje de información

Como se nota en las Figuras 5.11 y 5.12, las puntas de flecha indican señales que viajaban en una sola dirección. Las redes en las que la señal de entrada se alimenta continuamente en una dirección desde la conexión hasta que alcanza la capa de salida se denominan redes “**feedforward**”.

A pesar de la restricción en el flujo de información, las redes “**feedforward**” ofrecen una sorprendente cantidad de flexibilidad. Por ejemplo, se puede variar el número de niveles y nodos en cada nivel, se pueden modelar múltiples resultados simultáneamente o se pueden aplicar múltiples capas ocultas. Una red neuronal con múltiples capas ocultas se denomina Red neuronal profunda (DNN) y la práctica de la formación de dicha red a veces se denomina aprendizaje profundo, ver Figura 5.13.

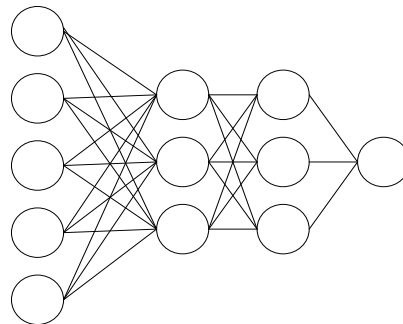


Figura 5.13: Topología de una red con múltiples capas

Por el contrario, una red recurrente permite que las señales viajen en ambas direcciones utilizando bucles. Esta propiedad, que refleja más de cerca cómo funciona una red neuronal biológica,

permite aprender patrones extremadamente complejos. La adición de una memoria a corto plazo, o “**delay**”, aumenta inmensamente el poder de las redes recurrentes. En particular, esto incluye la capacidad de comprender las secuencias de eventos a lo largo de un período de tiempo. Esto podría usarse para la predicción del mercado de valores, la comprensión del habla o el pronóstico del clima. Una red recurrente simple se describe mediante la Figura 5.14:

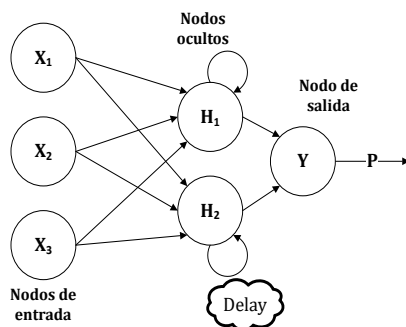


Figura 5.14: Tipología de una red recurrente

A pesar de su potencial, las redes recurrentes siguen siendo en gran parte teóricas y rara vez se utilizan en la práctica.

La cantidad de nodos en cada capa

Además de las variaciones en el número de capas y la dirección del recorrido de la información, las redes neuronales también pueden variar en complejidad por la cantidad de nodos en cada capa. El número de nodos de entrada está predeterminado por la cantidad de características en los datos de entrada. De manera similar, el número de nodos de salida está predeterminado por el número de resultados a modelar o el número de niveles de clase en el resultado. Sin embargo, el usuario debe decidir la cantidad de nodos ocultos antes de entrenar el modelo.

Lamentablemente, no existe una regla confiable para determinar el número de neuronas en la capa oculta. El número apropiado depende de la cantidad de nodos de entrada, la cantidad de datos de entrenamiento, la cantidad de datos ruidosos y la complejidad de la tarea de aprendizaje, entre muchos otros factores. En general, las topologías de red más complejas con un mayor número de conexiones de red permiten el aprendizaje de problemas más complejos. Un mayor número de neuronas dará como resultado un modelo que refleja más de cerca los datos de entrenamiento, pero esto conlleva un riesgo de **sobreajuste**; puede generalizarse pobremente

a datos futuros. Las redes neuronales grandes también pueden ser costosas desde el punto de vista de la computación y de entrenamiento lento.

Entrenamiento de redes neuronales con propagación hacia atrás

La topología de red es una pizarra en blanco que por sí sola no ha aprendido nada. Como un niño recién nacido, debe ser entrenado con experiencia. A medida que la red neuronal procesa los datos de entrada, las conexiones entre las neuronas se fortalecen o debilitan, de forma similar a cómo se desarrolla el cerebro de un bebé a medida que él o ella experimenta el entorno. Los pesos de conexión de la red se ajustan para reflejar los patrones observados a lo largo del tiempo.

El entrenamiento de una red neuronal mediante el ajuste de pesos de conexión es muy intensivo computacionalmente. En consecuencia, aunque se habían estudiado durante décadas anteriores, las redes neuronales, rara vez se aplicaban a las tareas de aprendizaje del mundo real hasta mediados y finales de la década de 1980, cuando se descubrió un método eficiente de entrenamiento de una red. El algoritmo, que usó una estrategia de errores de propagación inversa, ahora se conoce simplemente como **“backpropagation”** (propagación hacia atrás).

Aunque todavía es notoriamente lento en relación con muchos otros algoritmos de aprendizaje automático, el método de “backpropagation” provocó un resurgimiento del interés en redes neuronales. Como resultado, las redes de “feedforward” de múltiples capas que usan el algoritmo de “backpropagation” son ahora comunes en el campo de la minería de datos. Dichos modelos ofrecen las siguientes fortalezas y debilidades, Tabla 5.7

5. APRENDIZAJE AUTOMÁTICO

Fortalezas	Debilidades
Se puede adaptar a problemas de clasificación o predicción numérica	Extremadamente computacionalmente intenso y lento de entrenar, particularmente si la topología de red es compleja
Capaz de modelar patrones más complejos que casi cualquier algoritmo	Muy propenso a sobre-ajustar los datos de entrenamiento
Hace pocas suposiciones sobre las relaciones subyacentes de los datos	Resulta en un modelo complejo de caja negra que es difícil, si no imposible de interpretar

Tabla 5.7: Fortalezas y Debilidades del algoritmo “backpropagation”

En su forma más general, el algoritmo de “backpropagation” itera a través de muchos ciclos de dos procesos. Cada ciclo se conoce como una época. Debido a que la red no contiene conocimiento a priori (existente), las ponderaciones de inicio se establecen típicamente al azar. Luego, el algoritmo itera a través de los procesos, hasta que se alcanza un criterio de parada. Cada época en el algoritmo de “backpropagation” incluye:

- Una fase hacia adelante en la que las neuronas se activan en secuencia desde la capa de entrada a la capa de salida, aplicando los pesos y la función de activación de cada neurona a lo largo del camino. Al alcanzar la capa final, se produce una señal de salida.
- Una fase hacia atrás en la que la señal de salida de la red que resulta de la fase directa se compara con el valor objetivo verdadero en los datos de entrenamiento. La diferencia entre la señal de salida de la red y el valor verdadero da como resultado un error que se propaga hacia atrás en la red para modificar los pesos de conexión entre las neuronas y reducir los errores futuros.

Con el tiempo, la red utiliza la información enviada hacia atrás para reducir el error total de la red. Sin embargo, queda una pregunta: porque la relación entre las entradas y salidas de cada neurona es compleja, ¿cómo determina el algoritmo cuánto debería cambiarse un peso? La respuesta a esta pregunta implica una técnica llamada gradiente descendente. Conceptualmente, funciona de manera similar a cómo un explorador atrapado en la jungla podría encontrar un camino hacia el agua. Al examinar el terreno y caminar continuamente en la dirección con la

mayor pendiente descendente, el explorador eventualmente alcanzará el valle más bajo, que probablemente sea un lecho de río.

En un proceso similar, el algoritmo de “backpropagation” utiliza la derivada de la función de activación de cada neurona para identificar el gradiente en la dirección de cada uno de los pesos entrantes, de ahí la importancia de tener una función de activación diferenciable. El gradiente sugiere qué tan abrupto se reducirá o aumentará el error para un cambio en el peso. El algoritmo intentará cambiar los pesos que resultan en la mayor reducción de error en una cantidad conocida como tasa de aprendizaje. Cuanto mayor sea la velocidad de aprendizaje, más rápido el algoritmo intentará descender por los gradientes, lo que podría reducir el tiempo de entrenamiento a riesgo de sobrepasar el valle, Figura 5.15.

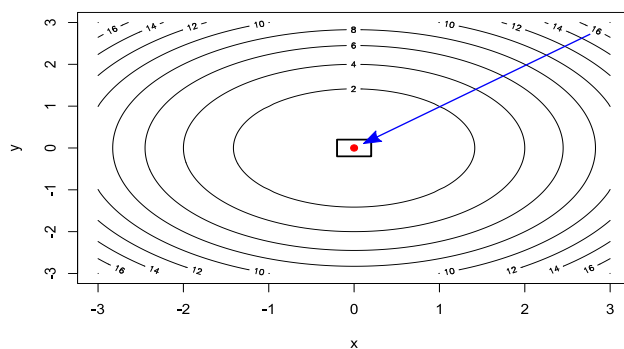


Figura 5.15: Gradiente descendente

Máquinas vectorial de soporte - support vector machine SVM

El objetivo de SVM es crear un límite plano llamado hiperplano, que divide el espacio para crear particiones bastante homogéneas en ambos lados. De esta forma, el aprendizaje SVM combina aspectos del aprendizaje vecino más cercano basado en instancia presentado en la Sección 5.6.1 y el modelo de regresión lineal. La combinación es extremadamente poderosa, permitiendo a las “SVM” modelizar relaciones altamente complejas.

Aunque las matemáticas básicas que manejan SVM han existido por décadas, recientemente han estallado en popularidad. Esto está, por supuesto, enraizado en su desempeño de vanguardia, pero quizás también debido al hecho de que los galardonados algoritmos SVM se han implemen-

tado en varias bibliotecas populares y bien respaldadas en muchos lenguajes de programación.

SVM se pueden adaptar para usar con casi cualquier tipo de tarea de aprendizaje, incluidas la clasificación y la predicción numérica. Muchos de los éxitos clave del algoritmo han venido en reconocimiento de patrones. Las aplicaciones notables incluyen:

- Clasificación de datos de expresión genética de “microarrays” en el campo de la bioinformática para identificar cáncer u otras enfermedades genéticas.
- Categorización de texto, como la identificación del idioma utilizado en un documento o la clasificación de documentos por tema.
- La detección de eventos raros pero importantes, como falla del motor de combustión, brechas de seguridad o terremotos.

Clasificación con hiperplanos

Como se señaló anteriormente, SVM emplea un límite llamado hiperplano para dividir los datos en grupos de valores de clase similares. Es así, que la Figura 5.16 representa un hiperplano (de dos dimensiones) que separa grupos de círculos en dos. Debido a que los círculos se pueden separar perfectamente por la línea recta o la superficie plana, se dice que son linealmente separables. SVM también pueden extenderse a problemas donde los puntos no son separables linealmente.

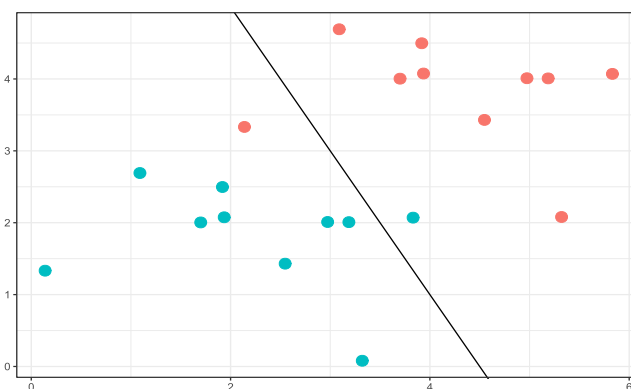


Figura 5.16: SVM, hiperplano dos dimensiones

En dos dimensiones, la tarea del algoritmo SVM es identificar una línea que separe las dos

clases. Como se muestra en la Figura 5.17, hay más de una opción de línea divisoria entre los distintos círculos. Tres de esas posibilidades se etiquetan a, b y c. ¿Cómo elige el algoritmo?

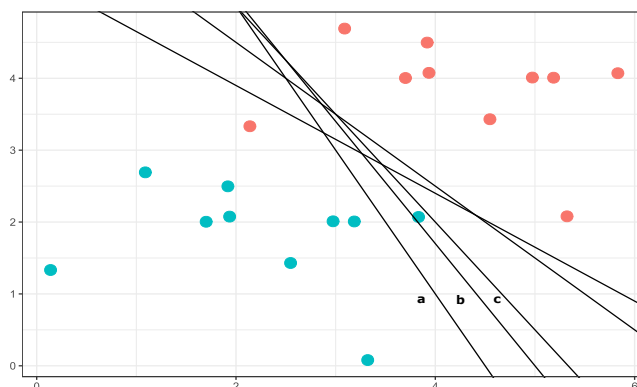


Figura 5.17: SVM, Varios hiperplanos en dos dimensiones

La respuesta a esa pregunta implica una búsqueda del **hiperplano de margen máximo** (MMH) que crea la mayor separación entre las dos clases. Aunque cualquiera de las tres líneas (a, b o c) que separan los círculos clasificaría correctamente todos los puntos de datos, es probable que la línea que conduce a la mayor separación generalice la mejor a los datos futuros. El margen máximo mejorará las posibilidades de que, a pesar del ruido aleatorio, los puntos permanezcan en el lado correcto del límite.

Los vectores de soporte (indicados por las flechas en la Figura 5.18) son los puntos de cada clase que son los más cercanos al MMH; cada clase debe tener al menos un vector de soporte, pero es posible tener más de uno. Usando solo los vectores de soporte, es posible definir el MMH. Esta es una característica clave de los SVM; los vectores de soporte proporcionan una forma muy compacta de almacenar un modelo de clasificación, incluso si el número de características es extremadamente grande.

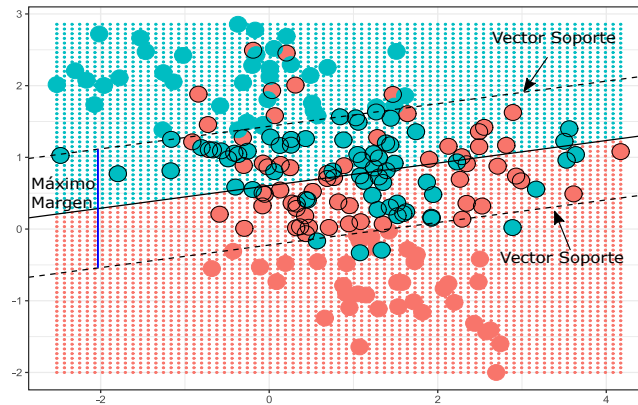


Figura 5.18: SVM, Vectores Soporte y Margen Máximo

El algoritmo para identificar los vectores de soporte se basa en la geometría vectorial e involucra algunas matemáticas bastante complicadas que están fuera del alcance de este trabajo. Sin embargo, los principios básicos del proceso son bastante sencillos, más detalle James et al. (95).

Datos linealmente separables

Es más fácil entender cómo encontrar el margen máximo bajo la suposición de que las clases son linealmente separables. En este caso, el MMH está lo más lejos posible de los límites externos de los dos grupos de puntos de datos. Estos límites externos se conocen como el casco convexo (convex hull). El MMH es entonces la bisectriz perpendicular de la línea más corta entre los dos cascos convexos, Figura 5.19. Sofisticados algoritmos de computadora que utilizan una técnica conocida como optimización cuadrática son capaces de encontrar el margen máximo de esta manera.

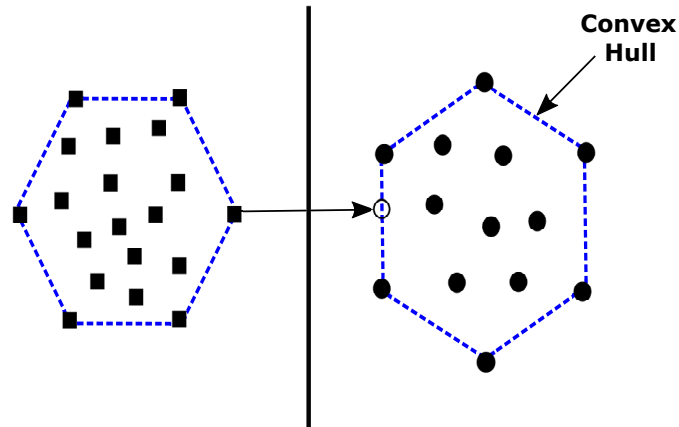


Figura 5.19: SVM, Convex Hull

Un enfoque alternativo (pero equivalente) implica una búsqueda a través del espacio de cada hiperplano posible para encontrar un conjunto de dos planos paralelos que dividen los puntos en grupos homogéneos, pero que están separados lo más posible. Para comprender este proceso de búsqueda, necesitaremos definir exactamente qué queremos decir con un hiperplano. En el espacio n -dimensional, se usa la ecuación 5.14.

$$\vec{w} \cdot \vec{x} + b = 0 \quad (5.14)$$

Donde, w es un vector de n pesos, es decir, $\{w_1, w_2, \dots, w_n\}$ y b es un número único conocido como el sesgo. El sesgo es conceptualmente equivalente al término de intersección en la forma pendiente-intersección de un modelo de regresión lineal. Usando la ecuación 5.14, el objetivo del proceso es encontrar un conjunto de pesos que especifique dos hiperplanos, ecuación 5.15.

$$\begin{cases} \vec{w} \cdot \vec{x} + b \geq +1 \\ \vec{w} \cdot \vec{x} + b \leq -1 \end{cases} \quad (5.15)$$

También se requiere que estos hiperplanos se especifiquen de forma que todos los puntos de una clase caigan por encima del primer hiperplano y todos los puntos de la otra clase queden debajo del segundo hiperplano. Esto es posible siempre que los datos sean linealmente separables. La geometría del vector define la distancia entre estos dos planos mediante la ecuación 5.16.

$$\frac{2}{\|\vec{w}\|} \quad (5.16)$$

Aquí, $\|w\|$ indica la norma euclidiana (la distancia desde el origen al vector \vec{w}). Dado que $\|w\|$ está en el denominador, se requiere maximizar la distancia minimizando $\|w\|$. La tarea normalmente se expresa como un conjunto de restricciones, ecuación 5.17

$$\begin{cases} \text{mín } \frac{1}{2} \|\vec{w}\|^2 \\ \text{s.t. } y_i (\vec{w} \cdot \vec{x}_i - b) \geq 1, \forall \vec{x}_i \end{cases} \quad (5.17)$$

La ecuación 5.17, no es muy complicada de entender conceptualmente. Básicamente, la primera línea implica que necesitamos minimizar la norma euclidiana (al cuadrado y dividida por dos para facilitar el cálculo). La segunda línea apunta que esto está sujeto a (s.t.), es decir, que cada uno de los puntos esté correctamente clasificado. Tenga en cuenta que y indica el valor de la clase (transformado en $+1$ o -1).

Datos no linealmente separables

En esta modelización se emplea una variable “slack”, que crea un margen suave que permite que algunos puntos caigan en el lado incorrecto del margen. La Figura 5.20 ilustra dos puntos que caen en el lado equivocado de la línea con los correspondientes términos “slack”(denotados con la letra griega ξ).

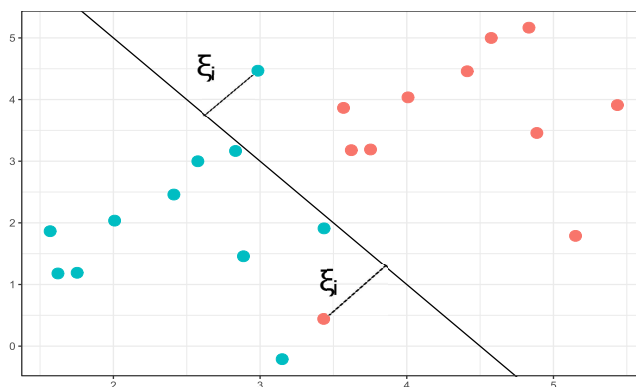


Figura 5.20: SVM, variable “slack”

Se penaliza (costo denotado por C) a todos los puntos que infringen las restricciones, y en lugar de encontrar el margen máximo, el algoritmo intenta minimizar el costo total. Por lo tanto, el algoritmo optimizara la ecuación 5.18.

$$\begin{cases} \min \frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t. } y_i (\vec{w} \cdot \vec{x} - b) \geq 1 - \xi_i, \forall \vec{x}_i, \xi_i \geq 0 \end{cases} \quad (5.18)$$

Cuanto mayor sea el parámetro de costo, más difícil será la optimización para lograr una separación del 100 por ciento. Por otro lado, un parámetro de menor costo pondrá el énfasis en un margen general más amplio. Es importante establecer un equilibrio entre estos dos para crear un modelo que generalice bien los datos futuros, mayor detalle James et al. (95).

Kernels para espacios no lineales

En muchas aplicaciones, las relaciones entre variables no son lineales, y como se ha analizado, un SVM aún puede ser entrenado en dichos datos mediante la adición de una variable de holgura, lo que permite que algunos ejemplos se clasifiquen erróneamente. Sin embargo, esta no es la única forma de abordar el problema de la no linealidad. Una característica clave de SVM es su capacidad para asignar el problema a un espacio de mayor dimensión utilizando un proceso conocido como el truco del kernel “kernel trick”. Al hacerlo, una relación no lineal de repente puede parecer bastante lineal, Figura 5.21.

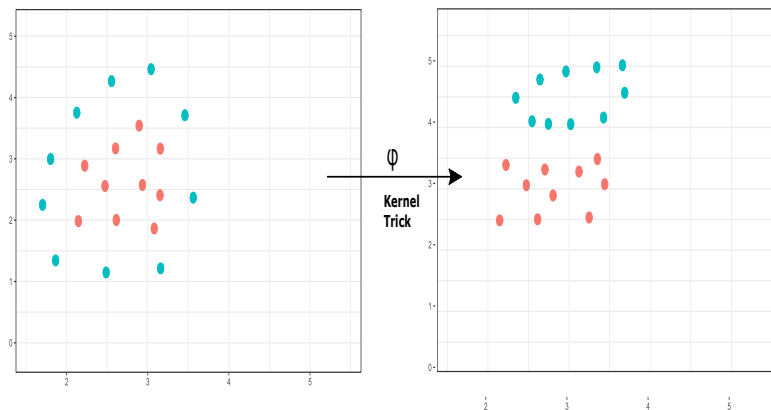


Figura 5.21: SVM, Kernel no lineal

SVM con kernels no lineales añaden dimensiones adicionales a los datos para crear una separación de esta manera. Básicamente, el truco del kernel implica un proceso de construcción de nuevas características que expresan relaciones matemáticas entre las características medidas.

5. APRENDIZAJE AUTOMÁTICO

Esto permite a SVM aprender conceptos que no se midieron explícitamente en los datos originales. SVM con kernels no lineales son clasificadores extremadamente potentes, aunque tienen algunas desventajas, como se muestra Tabla 5.8.

Fortalezas	Debilidades
Se puede usar para problemas de clasificación o predicción numérica	Encontrar el mejor modelo requiere probar varias combinaciones de kernels y parámetros del modelo
No demasiado influenciado por datos ruidosos y no muy propenso a sobreajustar	Puede ser lento para entrenar, especialmente si el conjunto de datos de entrada tiene una gran cantidad de características o ejemplos
Puede ser más fácil de usar que las redes neuronales, particularmente debido a la existencia de varios algoritmos SVM bien soportados	Resulta en un modelo complejo de caja negra que es difícil, si no imposible, interpretar
Gana popularidad debido a su alta precisión y a sus “éxitos” en la minería de datos	

Tabla 5.8: Fortalezas y Debilidades de los kernels no lineales

Las funciones kernel, por lo general, se expresan, por la letra griega ϕ , es decir, $\phi(x)$, es un mapeo de los datos en otro espacio. Por lo tanto, la función de kernel general aplica alguna transformación a los vectores característicos x_i y x_j y los combina utilizando el producto interior entre un par de vectores, por lo que su resultado será un escalar, ecuación 5.24.

$$K(\vec{x}_i, \vec{x}_j) = \phi(\vec{x}_i) \cdot \phi(\vec{x}_j) \quad (5.19)$$

Algunas de las funciones del kernel más comúnmente utilizadas son las siguientes:

- El kernel lineal no transforma los datos en absoluto. Por lo tanto, se puede expresar simplemente como el producto interior de las características, ecuación 5.20.

$$K(\vec{x}_i, \vec{x}_j) = x_i \cdot x_j \quad (5.20)$$

- El kernel polinomial de grado d agrega una transformación no lineal simple a los datos, ecuación 5.21.

$$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1)^d \quad (5.21)$$

- El kernel sigmoide algo análogo a una red neuronal. Las letras griegas *kappa* y *delta* se usan como parámetros del kernel, ecuación 5.22.

$$K(\vec{x}_i, \vec{x}_j) = \tanh(\kappa \vec{x}_i \cdot \vec{x}_j - \delta) \quad (5.22)$$

- El kernel RBF gaussiano es similar a una red neuronal. El kernel RBF funciona bien en muchos tipos de datos y se cree que es un inicio razonable para muchas tareas de aprendizaje, ecuación 5.23.

$$K(\vec{x}_i, \vec{x}_j) = e^{-\frac{\|\vec{x}_i - \vec{x}_j\|^2}{\sigma^2}} \quad (5.23)$$

No existe una regla confiable para unir un kernel a una tarea de aprendizaje en particular. El ajuste depende en gran medida del concepto o tarea que debe aprender, así como de la cantidad de datos de capacitación y las relaciones entre las características. Dicho esto, en muchos casos, la elección del kernel es arbitraria, ya que el rendimiento puede variar ligeramente.

5.6.6. Evaluación del rendimiento del modelo

El proceso de evaluación de los algoritmos de aprendizaje automático es muy similar al proceso de evaluación de los estudiantes. Dado que los algoritmos tienen diferentes puntos fuertes y débiles, los datos de pruebas deben distinguir entre los diferentes aprendizajes. De manera óptima, el proceso de evaluación incluiría las siguientes medidas cuidadosas, como se sugiere en Elazmeh et al. (49).

1. Decidir qué propiedades “interesantes” de los clasificadores deben ser medidas, y elegir una métrica de evaluación.
2. Decidir qué método de estimación de la confianza debe utilizarse para validar nuestros resultados.
3. Verificar que los supuestos hechos por la métrica de evaluación y la estimación de confianza se cumplan.

5. APRENDIZAJE AUTOMÁTICO

4. Ejecutar el método de evaluación, utilizando la métrica elegida y el método de estimación de la confianza, y analizar sus resultados.
5. Interpretar estos resultados con respecto al dominio de referencia.

Por lo antes mencionado, la Evaluación del Aprendizaje de Máquina se dividirán en dos partes:

- El desempeño o rendimiento de las métricas empleadas
- El método de estimación para la confianza elegida

La primera parte considerará lo que sucede cuando se toman malas decisiones en cuanto a las métricas de desempeño (Paso 1 prácticamente se ignora). En la segunda parte se analizará lo que ocurre cuando no se respetan las suposiciones hechas sobre el método de estimación de la confianza elegido (no se tienen en cuenta los pasos 2 y 3).

Cabe señalar que el enfoque de la evaluación en el aprendizaje automático es muy diferente de la forma en que se abordan los métodos de evaluación en la mayoría de los campos aplicados, como la economía, la psicología, la biología, etc. Esta disparidad puede explicarse por el hecho de que, en el aprendizaje automático, se da más importancia a la creación de algoritmos sofisticados para resolver problemas que a las observaciones que conducen a este algoritmo o a sus pruebas, mientras que la investigación en los otros campos mencionados se basa más en pruebas exhaustivas de las teorías existentes, antes de sugerir otras nuevas.

Si se han descuidado los pasos 1, 2 y 3 anteriormente mencionados, pueden surgir problemas en la fase de evaluación. Una vez que se ejecuta un algoritmo de aprendizaje, vale la pena señalar que se pueden calcular muchas cantidades a partir de los resultados obtenidos. Dado que el propósito de la evaluación es ofrecer formas sencillas y convenientes de juzgar el rendimiento de un sistema de aprendizaje o de una hipótesis y/o de compararlo con otros, los métodos de evaluación pueden considerarse como resúmenes del rendimiento de los sistemas. Además, los mismos datos pueden resumirse de diversas maneras y, a veces, con diferentes supuestos, lo que da lugar a diferentes métricas de evaluación del rendimiento o pruebas de evaluación de la confianza. Existen varias métricas de evaluación del rendimiento, las más comúnmente utilizadas son: Exactitud, Precisión/Recall y Análisis ROC. Se centran en varios aspectos de un clasificador, pero, al mismo tiempo, ignoran otros, asignándoles los mismos valores de rendimiento, clasificadores que exhiben comportamientos extremadamente diferentes.

Todo el análisis siguiente se basará en la definición de una matriz de confusión. La matriz de confusión que utilizamos se muestra en la Tabla 5.9. Las definiciones de las medidas consideradas figuran a continuación, basadas en la tabla antes citada. El análisis ROC utiliza las medidas de *Recall* y *TFP* como se explicará más adelante.

- $Exactitud = \frac{VP + VN}{P + N}$
- $Precisión = \frac{VP}{VP + FP}$
- $Recall = \frac{VP}{P}$
- Tasa de Falso Positivo TFP = $\frac{FP}{N}$
- Tasa de Error = 1-Exactitud, la cual indica la proporción de clasificados incorrectamente.

Una matriz de confusión es una tabla que clasifica las predicciones de acuerdo a si coinciden con el valor real. Una de las dimensiones de la tabla indica las posibles categorías de valores previstos, mientras que la otra dimensión indica lo mismo para los valores reales.

		Clase Verdadera	
		Positivo	Negativo
Clase Hipotética	Sí	Verdaderos Positivos VP	Falso Positivo FP
	No	Falso Negativo FN	Verdadero Negativo VN
Total		$P = VP + FN$	$N = FP + VN$

Tabla 5.9: Matriz de Confusión

La precisión (también conocida como valor predictivo positivo) se define como la proporción de ejemplos positivos que son verdaderamente positivos; en otras palabras, cuando un modelo predice la clase positiva. Por otro lado, “recall” es una medida de cuán completos son los resultados, es decir, representa la sensibilidad.

5.6.7. ¿Qué tiene de malo la Exactitud?

“Exactitud” es la medida de evaluación más sencilla e intuitiva para los clasificadores: puesto que el interés central es crear clasificadores que no cometan ningún error, ¿qué es más sencillo que

contar el número de errores que cometen y compararlos en base a esta medida? Si bien se trata de una solución sencilla, ampliamente utilizada, cabe señalar que Exactitud no distingue entre los tipos de errores que comete (Falsos Positivos y Falsos Negativos). Si bien esto es aceptable si el conjunto de datos contiene tantos ejemplos de ambas clases (es decir, si es un conjunto de datos equilibrado) y si la precisión es muy alta, es difícil imaginar una aplicación que, en cualquier otro caso, no necesite distinguir entre los dos tipos de errores. De hecho, en ciertas circunstancias, ignorar el tema puede conducir a resultados catastróficos como en el caso de un problema de clasificación médica cuyo objetivo es discriminar entre pacientes cancerosos (clase positiva) y no cancerosos (clase negativa).

5.6.8. ¿Qué tiene de malo Precision/Recall?

Aunque estas medidas no son tan simples e intuitivas como lo es la Exactitud, todavía tienen una interpretación relativamente sencilla. Precisión evalúa hasta qué punto el clasificador clasificó correctamente los ejemplos como positivos, mientras que Recall evalúa hasta qué punto lo fueron todos los ejemplos que necesitaban ser clasificados como positivos. Es decir, son medida de “éxito” de la predicción cuando las clases están muy desequilibradas. En la recuperación de información, la precisión es una medida de la relevancia de los resultados, mientras que “recall” es una medida de cuántos resultados verdaderamente relevantes se obtienen.

La precisión ayuda cuando los costos de los falsos positivos son altos. Así que asumamos que el problema involucra la detección de cáncer de piel. Si tenemos un modelo que tiene una precisión muy baja, entonces a muchos pacientes se les dirá que tienen melanoma, y eso incluirá algunos diagnósticos erróneos. Muchas pruebas adicionales y estrés están en juego. Cuando los falsos positivos son demasiado altos, aquellos que monitorean los resultados aprenderán a ignorarlos después de ser bombardeados con falsas alarmas.

Recall ayuda cuando el costo de los falsos negativos es alto. ¿Y si necesitamos detectar misiles nucleares entrantes? Un falso negativo tendría consecuencias devastadoras. Por tal motivo, se necesita optimizar una métrica que combine **Precisión** y **Recall**.

La siguiente sección se centra en el Análisis ROC, que corrige los problemas de Precisión, Exactitud y Recall, pero introduce, sin embargo, un nuevo problema.

A partir de las medidas de desempeño antes citadas, se pueden construir otras medidas, como

las siguientes:

5.6.9. El Estadístico Kappa

El estadístico Kappa ajusta la Exactitud teniendo en cuenta la posibilidad de una predicción correcta sólo por azar. Esto es especialmente importante para los conjuntos de datos con un desequilibrio de clase severo, porque un clasificador puede obtener una alta precisión simplemente adivinando siempre la clase más frecuente.

Los valores de Kappa van de 0 a un máximo de 1, lo que indica una perfecta concordancia entre las predicciones del modelo y los valores reales. Los valores inferiores a uno indican una concordancia imperfecta. Dependiendo de cómo se utilice un modelo, la interpretación del estadístico Kappa puede variar. Una interpretación común se muestra a continuación:

- Pobre concordancia = menos de 0.20
- Regular concordancia = 0.20 a 0.40
- Moderada concordancia = 0.40 a 0.60
- Buena concordancia = 0.60 a 0.80
- Muy Buena concordancia = 0.80 a 1.00

La Ecuación 5.24 permite estimar el estadístico Kappa. En esta fórmula, $Pr(a)$ se refiere a la proporción del acuerdo real y $Pr(e)$ al acuerdo esperado entre el clasificador y los valores reales, bajo el supuesto de que fueron elegidos al azar.

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \quad (5.24)$$

5.6.10. La medida F

Una medida del rendimiento del modelo que combina la precisión y “recall” en un solo número se conoce como la medida F (también llamada a veces $F1\ score$ o puntuación $F\ score$). La medida F combina precisión y recall utilizando la media armónica, un tipo de media que se utiliza para las tasas de cambio. Se utiliza la media armónica en lugar de la media aritmética

común, ya que tanto la precisión como “recall” se expresan como proporciones entre cero y uno, lo que puede interpretarse como tasas. La Ecuación 5.25 permite estimar su valor.

$$\text{F-score} = \frac{2 \times \text{Precisión} \times \text{Recall}}{\text{Recall} + \text{Precisión}} \quad (5.25)$$

5.6.11. Análisis ROC

El análisis ROC es, una vez más, un método que no es tan simple y directo como la precisión, pero que, sin embargo, sigue teniendo un atractivo intuitivo. Hay que pensar en ello en dos sentidos:

- La medida que utiliza
- La manipulación que hace de estas medidas

Las medidas que utiliza son Recall VP/P y la Tasa de Falsos Positivos FP/N . El análisis ROC traza TFP en el eje x de un gráfico y Recall en el eje y . Los puntos del gráfico pueden interpretarse como un reflejo de un mayor rendimiento si están situados simultáneamente cerca de 0 en el eje x y cerca de 1 (o 100%) en el eje y . Esto se debe a que cuanto más se acerca un punto a 0 en el eje x , menor es su tasa de falsos positivos; y cuanto más se acerca a 1 un punto en el eje y , mayor es su recuperación.

Con estas medidas establecidas, ROC Analysis considera que un clasificador está compuesto de dos partes. El primero (la parte difícil, que también define al clasificador) realiza una especie de transformación de datos que acerca los ejemplos similares de la misma clase. La segunda parte (mucho más fácil de manejar) decide dónde exactamente se deben colocar los límites entre los diferentes grupos establecidos en la primera parte. El algoritmo es entonces probado en varios valores de la TFP permitiendo que los límites se muevan desde ubicaciones que representan una TFP muy pequeña a ubicaciones que representan una TFP muy grande. Esto permite que un gráfico ROC vea a qué costos (Tasa FP , es decir, falsas alarmas) se pueden lograr beneficios (*Recall*, es decir, también Tasa de aciertos). Para una descripción más detallada del Análisis ROC, por favor, véase Fawcett (55). Véase también la discusión en Drummond and Holte (42) para una discusión más detallada de la relación entre falsas alarmas y tasa de aciertos.

En otras palabras, la curva (ROC) se utiliza comúnmente para examinar la compensación entre la detección de los verdaderos positivos, evitando al mismo tiempo los falsos positivos. Como se puede sospechar por el nombre, las curvas ROC fueron desarrolladas por ingenieros en el campo de las comunicaciones. Alrededor de la época de la Segunda Guerra Mundial, el radar y los operadores de radio utilizaron curvas ROC para medir la capacidad de un receptor para discriminar entre señales verdaderas y falsas alarmas. La misma técnica es útil hoy en día para visualizar la eficacia de los modelos de aprendizaje automático.

Cuanto más cerca se encuentre la curva del clasificador perfecto, mejor será la identificación de los valores positivos. Esto puede ser medido usando un estadístico conocido como el área bajo la curva ROC (AUC abreviada). Las AUC tratan el diagrama ROC como un cuadrado bidimensional y miden el área total bajo la curva ROC. Las AUC van de 0.5 (para un clasificador sin valor predictivo) a 1.0 (para un clasificador perfecto). Una convención para interpretar los puntajes de las AUC utiliza un sistema similar al de las calificaciones de las letras académicas:

- **A:** Excepcional = 0.9 to 1.0
- **B:** Excelente = 0.8 to 0.9
- **C:** Aceptable = 0.7 to 0.8
- **D:** Pobre = 0.6 to 0.7
- **E:** No discrimina = 0.5 to 0.6

5.6.12. Estimación del rendimiento futuro

Algunos programas presentan matrices de confusión y medidas de rendimiento durante el proceso de construcción del modelo. El propósito de estas estadísticas es proporcionar información sobre el error de sustitución del modelo, que ocurre cuando los datos de entrenamiento se predicen incorrectamente a pesar de que el modelo se construye directamente a partir de estos datos. Esta información puede ser utilizada como un diagnóstico aproximado para identificar desempeño deficiente. En lugar de basarse en un error de sustitución, una mejor práctica es evaluar el comportamiento de un modelo, en datos que aún no ha visto. Hemos utilizado este enfoque en los capítulos anteriores cuando dividimos los datos disponibles en un conjunto para entrenamiento y otro para pruebas. Sin embargo, no siempre es ideal crear conjuntos de datos de

formación y de pruebas. Por ejemplo, en una situación en la que sólo se dispone de un pequeño grupo de datos, es posible que no se desee reducir aún más la muestra.

El método de retención

El procedimiento de partición de datos en conjuntos de datos de entrenamiento y pruebas se conoce como el método de retención. Como se muestra en la Figura 5.22, el conjunto de datos de entrenamiento se utiliza para generar el modelo, que luego se aplica al conjunto de datos de prueba para generar predicciones para su evaluación. Por lo general, alrededor de un tercio de los datos se utiliza para las pruebas y dos tercios para el entrenamiento, pero esta proporción puede variar dependiendo de la cantidad de datos disponibles. Para asegurar que los datos del entrenamiento y de las pruebas no tengan diferencias sistemáticas, sus ejemplos se dividen aleatoriamente en dos grupos.

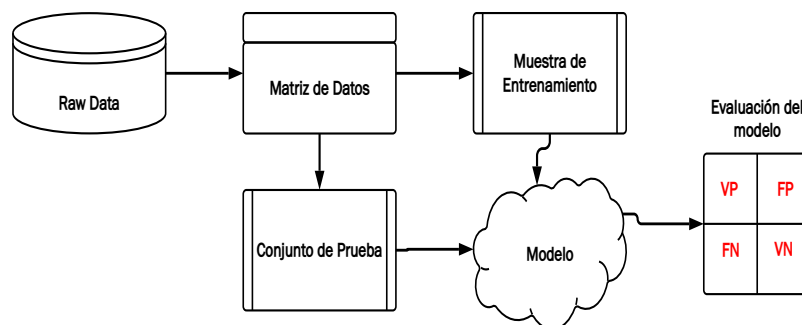


Figura 5.22: Partición de los datos en muestra de entrenamiento y prueba

Para que el método de retención resulte en una estimación verdaderamente Exacta sobre un desempeño “futuro” en ningún momento debe permitirse que el rendimiento del conjunto de datos de prueba influyan en el modelo. Para evitar este problema, es mejor dividir los datos originales de manera que, además de los conjuntos de datos de entrenamiento y los conjuntos de datos de prueba, se disponga de un conjunto de datos de validación. El conjunto de datos de validación se utilizaría para iterar y refinar el modelo o modelos elegidos, dejando que el conjunto de datos de prueba se utilice sólo una vez como paso final para informar de una tasa de error estimada para futuras predicciones. Una división típica entre entrenamiento, prueba y validación sería 50 por ciento, 25 por ciento y 25 por ciento, respectivamente.

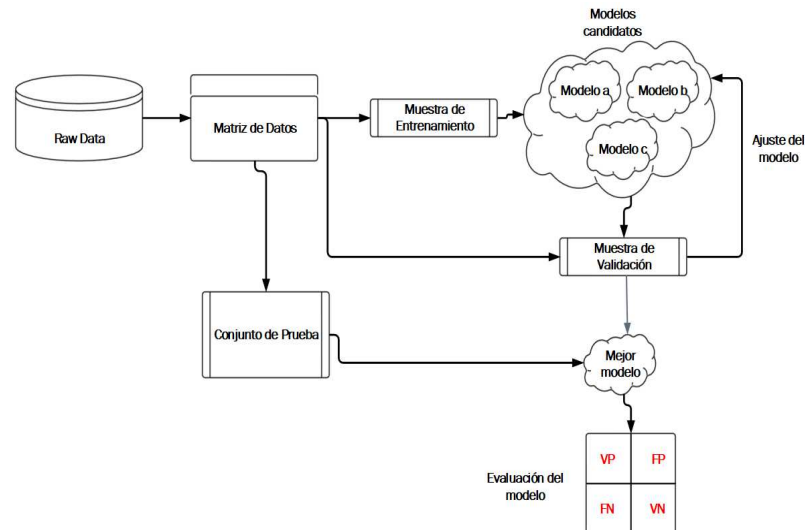


Figura 5.23: Partición de los datos en muestra de entrenamiento, prueba y validación

A medida que se utilizan múltiples muestras retenidas, es menos probable que el modelo se entrene o se pruebe con datos no representativos. Ampliaremos esta idea en la siguiente sección.

Validación Cruzada

La retención repetida es la base de una técnica conocida como k -fold cross-validation (o k -fold CV), que se ha convertido en el estándar de la industria para estimar el rendimiento del modelo. En lugar de tomar muestras aleatorias repetidas que podrían potencialmente utilizar el mismo registro más de una vez, k -fold CV divide aleatoriamente los datos en k para separar completamente las particiones aleatorias llamadas fold.

Aunque k se puede establecer en cualquier número, la convención más común es usar Validación cruzada de 10 veces (10 fold CV). ¿Por qué 10 pliegues? La razón es que la evidencia empírica sugiere que hay poco beneficio adicional en el uso de un número mayor. Para cada uno de los 10 “folds” (cada uno de los cuales comprende el 10 por ciento del total de los datos), una máquina de aprendizaje se basa en el 90 por ciento restante de los datos. El pliegue (fold) coincide con el 10 por ciento que se utiliza para la evaluación del modelo. Después del proceso de entrenamiento y evaluación el modelo se ha ajustado 10 veces (con 10 combinaciones diferentes de entrenamiento/prueba).

Una alternativa un poco menos frecuente a k-fold CV se conoce como bootstrap sampling, bootstrap o bootstrapping para abreviar.

Bootstrap sampling

En términos generales, se refieren a los métodos estadísticos que utilizan muestras aleatorias de datos para estimar las propiedades de un conjunto mayor. Cuando este principio se aplica al rendimiento del modelo de aprendizaje automático, implica la creación de varios conjuntos de datos de entrenamiento y pruebas seleccionadas al azar, que luego se utilizan para estimar los estadísticos de rendimiento. Los resultados de los diversos conjuntos de datos aleatorios se promedian para obtener una estimación final del rendimiento futuro.

Entonces, ¿qué hace que este procedimiento sea diferente de k-fold CV? Considerando que la validación cruzada divide los datos en particiones separadas en las que cada ejemplo sólo puede aparecer una vez, el bootstrap permite que los ejemplos se seleccionen varias veces a través de un proceso de muestreo con reemplazo. Esto significa que del conjunto de datos original de n ejemplos, el procedimiento de bootstrap creará uno o más nuevos conjuntos de datos de entrenamiento que también contendrá n ejemplos, algunos de los cuales se repiten. Los datos de pruebas correspondientes se construyen a partir del conjunto de ejemplos que no se seleccionaron del respectivo conjunto de datos de entrenamiento.

El bootstrap es una herramienta estadística ampliamente aplicable y extremadamente poderosa que puede ser utilizada para cuantificar la incertidumbre asociada con un estimador o método de aprendizaje automático dado. Sin embargo, el poder del bootstrap reside en el hecho de que se puede aplicar fácilmente a una amplia gama de métodos de aprendizaje automático, incluyendo algunos para los que es difícil obtener una medida de variabilidad y que no son producidos automáticamente por un software estadístico. Una ventaja de bootstrap sobre la validación cruzada es que tiende a funcionar mejor con conjuntos de datos muy pequeños. Además, el muestreo bootstrap tiene aplicaciones que van más allá de la medición del rendimiento. En particular, en la siguiente sección se hablará cómo se pueden utilizar los principios del muestreo bootstrap para mejorar el rendimiento del modelo.

5.6.13. Afinación de los modelos para mejorar desempeño (rendimiento)

Algunos problemas de aprendizaje se adaptan bien a los modelos presentados en los capítulos anteriores. En tales casos, puede que no sea necesario dedicar mucho tiempo a iterar y refinar el modelo; puede que funcione lo suficientemente bien como está. Por otro lado, algunos problemas son intrínsecamente más difíciles. Los conceptos subyacentes a “aprender” pueden ser extremadamente complejos, requiriendo un entendimiento de muchas relaciones sutiles, o pueden ser afectados por la variación aleatoria, haciendo difícil definir la señal dentro del ruido.

Desarrollar modelos que funcionen extremadamente bien en problemas difíciles es todo un arte como lo es una ciencia. A veces un poco de intuición es útil cuando se trata de identificar áreas en las que se puede mejorar el rendimiento. En otros casos, la búsqueda de mejoras requerirá un enfoque de fuerza bruta y de ensayo y error. Por supuesto, el proceso de búsqueda de numerosas mejoras posibles puede ser ayudado por el uso de programas automatizados.

El proceso de ajuste de las opciones del modelo para identificar el mejor ajuste se denomina “tuning” de parámetros. El ajuste automatizado de parámetros requiere que se consideren tres preguntas:

1. ¿Qué tipo de modelo de aprendizaje de máquina (e implementación específica) debe ser entrenado con los datos?
2. ¿Qué parámetros del modelo se pueden ajustar y con qué intensidad para encontrar los ajustes óptimos?
3. ¿Qué criterios deben utilizarse para evaluar los modelos y encontrar al mejor candidato?

Responder a la primera pregunta implica encontrar una correspondencia adecuada entre la tarea de aprendizaje automático. Obviamente, esto requiere una comprensión de la amplitud y profundidad de los modelos de aprendizaje de máquina. También puede ayudar a trabajar en un proceso de eliminación. Casi la mitad de los modelos pueden ser eliminados dependiendo de si la tarea es clasificación o predicción numérica; otros pueden ser excluidos en función del formato de los datos o de la necesidad de evitar los modelos de caja negra, etc.

Abordar la segunda pregunta es un asunto dictado en gran medida por la elección del modelo, ya que cada algoritmo utiliza un conjunto único de parámetros. El objetivo de **automatic**

tuning es buscar un conjunto de modelos candidatos que comprenden una matriz, o cuadrícula, de combinaciones de parámetros. Debido a que no es práctico buscar cualquier combinación concebible, sólo se utiliza un subconjunto de posibilidades para construir la cuadrícula.

El tercer y último paso en el ajuste automático del modelo consiste en identificar el mejor modelo entre los candidatos, mediante estadísticos como exactitud y kappa κ (para clasificadores) y R^2 o ECM – error cuadrático medio (para modelos numéricos). También es posible utilizar medidas como sensibilidad, especificidad o el área bajo la curva ROC (AUC).

Por lo anteriormente señalado, procederemos a construir un máquina de aprendizaje que permita sugerir el mejor modelo estadístico para la estimación de la precisión diagnóstica en un meta-análisis de pruebas diagnósticas.

5.7. Elección del Modelo Estadístico en Meta-análisis de Pruebas Diagnósticas mediante Aprendizaje Automático

Dentro del análisis expuesto, y compartiendo el procedimiento propuesto por Takwoingi et al. (183) se crearán meta-análisis de pruebas diagnósticas de baja prevalencia mediante el modelo HSROC porque tiene una mayor flexibilidad a la hora de elegir los parámetros Kriston et al. (101), con la finalidad de crear un modelo estadístico mediante algoritmos de aprendizaje automático, que permita sugerir el mejor modelo estadístico a la hora de resumir los resultados de un meta-análisis. Las especificaciones de los escenarios fueron diseñadas para reproducir situaciones realistas encontradas en el meta-análisis de estudios de precisión diagnóstica.

Se investigaron meta-análisis con diferentes números de estudios $k = 5, 10, 20, 35$. El tamaño de un estudio en cada meta-análisis fue muestreado aleatoriamente de una distribución uniforme $U(200; 2000)$ dada una prevalencia subyacente p , los individuos dentro de cada estudio fueron clasificados aleatoriamente como enfermos o no diagnosticados, y se asignó un valor de resultado de prueba continua que fue muestreado aleatoriamente, Bachmann et al. (6). Para determinar las tablas tetracóricas correspondientes, se utilizó el paquete HSROC del programa estadístico R, mediante el empleo de la función `simdata`, analizada en la Sección 4.4. Para crear la tabla de 2×2 para cada estudio, los individuos fueron clasificados como verdaderos positivos, falsos negativos,

falsos positivos o verdaderos negativos basados en el resultado de la prueba y el estado de la enfermedad. Las prevalencias de los estudios se ajustaron mediante una distribución uniforme con parámetros 0.01 y 0.35, es decir, la prevalencia de dichos estudios tomó valores menores o iguales a 35%. Los algoritmos de la simulación de las tablas tetracóricas se pueden observar en los Apéndices: [A.7](#), [A.8](#), [A.9](#), [A.10](#). En resumen, se generaron 23 043 estudios en 1 186 datasets (meta-análisis).

Una vez generados los meta-análisis de pruebas diagnósticas, el siguiente paso es aplicar los algoritmos analizados en Capítulo 4, el código en *R* se observa en el Apéndice [A.11](#). Este paso es importante ya que las métricas estimadas (observar Tabla [5.10](#)) permitirá seleccionar el mejor modelo estadístico para cada meta-análisis simulado, lo antes dicho se puede resumir en la Figura [5.24](#) y Tabla [5.11](#). El algoritmo en *R* se puede observar en el Apéndice [A.12](#)

Métrica	Modelo
DOR	DSL y MH
Correlación	DSL, MH, Reitsma
Área bajo la curva	PHM hetero y homo, Reitsma
Variabilidad entre estudios	DSL, PHM
Índice de Higgins	DSL
Sensibilidad, Especificidad	DSL, Copula, Reitsma
Ratios de Verosimilitud	Reitsma
Transformaciones logit de sensibilidad y especificidad	DSL, HSROC, Reitsma
Variabilidad de transformaciones logit de sensibilidad y especificidad	Reitsma
Variabilidad y correlación de sensibilidad y especificidad	Cópula
Forma de la curva SROC, puntos de corte, DOR y variabilidad entre y dentro estudios	HSROC

Tabla 5.10: Medidas estimadas a los Meta-análisis simulados

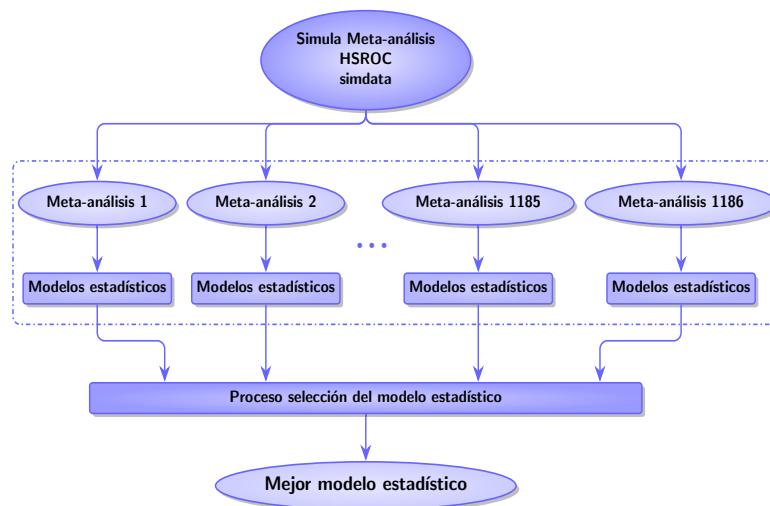


Figura 5.24: Proceso de simulación de Meta-análisis y selección del mejor modelo estadístico

Modelo	Descripción
Bivariante	Menor DOR de todos los meta-análisis simulados y AUC mayor a 0.5
Cópula	Si Correlación de efectos aleatorios es menor que cero o menor que tau de kendall
Cópula	Si Sensibilidad y especificidad (modelos jerárquicos) son menores a sensibilidad y especificidad de los modelos de copulas
HSROC	Si la variabilidad de los efectos aleatorios del meta-análisis seleccionado es menor a la variabilidad (mediana) de todos los meta-análisis simulados

Tabla 5.11: Criterios para etiquetar el mejor modelo estadística para cada meta-análisis simulado

Como resultado de los criterios establecidos en la Tabla 5.11, se etiquetaron 338 (28 %) meta-

análisis para el modelo Bivariante, 450 (38 %) para el modelo HSROC y 398 (34 %) para Cópula, ver Figura 5.25.

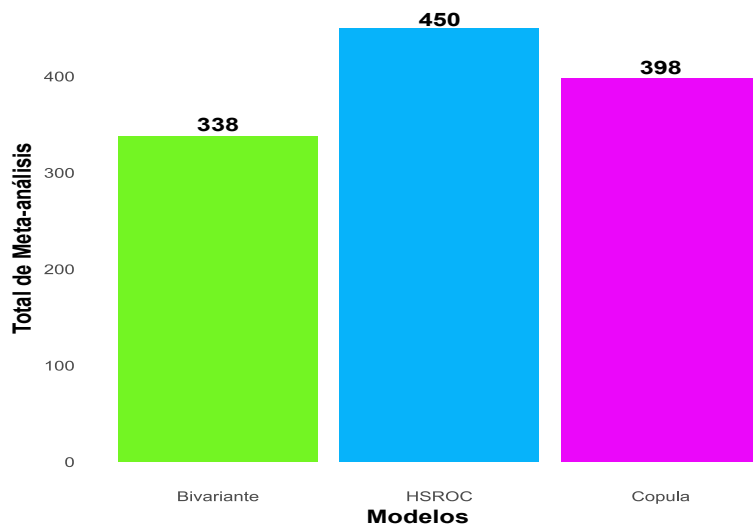


Figura 5.25: Meta-análisis etiquetas según modelo estadístico

Dado que a cada meta-análisis se le ha asignado un modelo estadístico que permita estimar medidas resúmenes (sensibilidad, especificidad, etc), es procedente construir una máquina de aprendizaje que sugiera el modelo adecuado que debe ajustarse a cada meta-análisis simulado, por tal motivo es necesario analizar el algoritmo que se emplea en las Máquinas de Vectores de Soporte y que viene incorporado en el paquete `caret` del programa estadístico R, el cual utilizaremos para construir la modelización.

5.7.1. Algoritmo de Máquinas de Vectores Soporte para Big Data

La elección de este tema se fundamenta por el grado de multidimensionalidad de los datos simulados. En esta sección trataremos de discutir uno de los principales algoritmos utilizados para la modelización de SVM que vienen incorporados en el paquete `caret`: Dual Coordinate Descent metho (DCD) Hsieh et al. (91).

La diferencia entre DCD y otros algoritmos es que en el DCD, el gradiente se calcula sobre el error muestral global; es decir, sobre toda la muestra. Se trata de un algoritmo de descenso por coordenadas, que en cada iteración minimiza un subproblema de una única variable; además

resuelve el problema dual (es decir mediante la aplicación de multiplicadores de Lagrange) Hsieh et al. (91). En cambio en otros algoritmos se calcula un gradiente local sobre el error muestral local. La ventaja DCD frente a otros es que las condiciones de convergencia se entienden mejor.

Algoritmo DCD

DCD puede ser utilizado tanto en clasificación como en regresión, pero para nuestro caso explicaremos el algoritmo para el problema de clasificación.

El problema que queremos resolver es el problema clásico de las Máquinas de Vectores Soporte para clasificación y cuya formulación primal es:

$$\min_w \frac{1}{2} w^T w + C \sum_{i=1}^n \varepsilon(w; x_i, y_i) \quad (5.26)$$

Sujeta a las restricciones de la ecuación 5.18 y donde ε es la función de pérdida. Tras una serie de cálculos mediante la teoría lagrangiana obtenemos el dual de la ecuación 5.26 como:

$$\begin{cases} \min_{\alpha} f(\alpha) = \frac{1}{2} \alpha^T \hat{Q} \alpha - e^T \alpha \\ \text{s.t } 0 \leq \alpha_i < U, \forall_i \end{cases} \quad (5.27)$$

Donde $\hat{Q} = Q + D$, con D matriz diagonal y $U = C$, y $Q_{ij} = y_i y_j x_i^T x_j$. El algoritmo Dual Coordinate Descend (DCD) empieza con un punto inicial $\alpha^0 \in R^n$, tras lo que se genera una secuencia de vectores $\{\alpha^k\}_{k=0}^{\infty}$.

Por lo tanto, en cada paso del algoritmo se necesita inicializar $\alpha^{k,i}$ a $\alpha^{k,i+1}$, siendo $\alpha^{k,i} = [\alpha_i^{k+1}, \dots, \alpha_{i-1}^{k+1}, \alpha_i^k, \dots, \alpha_n^k]$, luego es evidente que resulta un planteamiento del problema univariante, es decir,

$$\begin{cases} \min_d f(\alpha^{k,i} + d e_i) \\ \text{s.t } 0 \leq \alpha^{k,i} + d < U, \end{cases} \quad (5.28)$$

donde $e_i = [0, \dots, 0, 1, 0, \dots, 0]^T$.

Es evidente, que la forma cuadrática de la función objetivo de ecuación 5.28 se puede escribir en terminos de d , es decir,

$$f(\alpha^{k,i} + de_i) = \frac{1}{2} \hat{Q}_{ii} d^2 + \nabla_i f(\alpha^{k,i}) d + \text{constante}, \quad (5.29)$$

donde $\nabla_i f$ es la componente i del gradiente total de f

Por tanto, si derivamos la ecuación 5.29 con respecto a d e igualamos a cero, se observa que tiene un mínimo en $d = -\frac{\nabla_i f(\alpha^{k,i})}{\hat{Q}_{ii}}$, donde la actualización para cada iteración queda expresada como:

$$\alpha_i^{k,i+1} = \min \left(\max \left(\alpha_i^{k,i} - \frac{\nabla_i f(\alpha^{k,i})}{\hat{Q}_{ii}}, 0 \right), U \right), \quad (5.30)$$

si $\hat{Q}_{ii} > 0$. Nótese que se emplea la expresión máximo y mínimo para estar seguro que $0 \leq \alpha_i^{k,i+1} \leq C$. Luego el algoritmo se puede describir como se muestra a continuación:

Resultado: Estima hiperparámetros de la función de pérdida

Entrada: Dado α y su correspondiente $w = \sum y_i \alpha_i x_i$

mientras α no sea óptimo: **hacer**

$n \leftarrow 1$;

mientras $n \leq l$ **hacer**

$G = y_i w_i x_i - 1 + D_{ii} \alpha_i$;

$$PG = \begin{cases} \min(G, 0) & \text{si} \\ \max(G, 0) & \text{si } \alpha_i = U \\ G & \text{si } 0 < \alpha_i < U \end{cases}$$

si $|PG \neq 0|$ **entonces**

$\alpha_i \leftarrow \alpha_i$;

$\alpha_i \leftarrow \min \left(\max \left(\alpha_i - \frac{G}{\hat{Q}_{ii}}, 0 \right), U \right)$;

$w \leftarrow w + (\alpha_i - \hat{\alpha}_i) y_i x_i$

en otro caso

fin

$n \leftarrow n + 1$

fin

fin

Algoritmo 2: Algoritmo DCD para SVM

Resumiendo, se debe calcular Q_{ii} y $\nabla_i f(\alpha^{k,i})$, expresiones representadas por:

$$\hat{Q}_{ii} = x_i^T x_i \quad (5.31)$$

$$\nabla_i f(\alpha_i^{k,i}) = \hat{Q}_i \alpha_i - 1 = \sum_{j=1}^n \hat{Q}_{ij} \alpha_j - 1 = \sum_{j=1}^n y_i y_j x_i x_j \alpha_j - 1 \quad (5.32)$$

Por lo tanto, el coste computacional para hacer cada evaluación del núcleo puede ser reducido en el caso de usar el núcleo lineal como:

$$w = \sum_{j=1}^n y_j \alpha_j x_j \quad (5.33)$$

Luego el gradiente de la ecuación 5.32 viene dado por:

$$\nabla f(\alpha) = y_i w^T x_i - 1 + D_{ii} \alpha_i \quad (5.34)$$

SVM sobre conjuntos no balanceados

Uno de los principales problemas que enfrentan los algoritmos del aprendizaje automático, es el relacionado con las clases o las categorías de la variable respuesta, lo cual dificulta la generación de un buen modelo de clasificación con técnicas tradicionales, He and Garcia (79).

Por lo anteriormente mencionado, resulta muy importante el error de generalización que afecta a la clase de menor peso en el conjunto de datos a analizar. En el caso de SVM Vapnik (189), propone una opción novedosa para tratar estos problemas, al considerar una muestra del conjunto de entrenamiento para construir el modelo de clasificación.

SVM es una técnica de aprendizaje automático que se fundamenta en el aprendizaje estadístico Herbrich (82), Vapnik (189), siendo aplicada con éxitos en algunos campos de la ciencia. Como se dijo anteriormente, el espacio de hipótesis de esta técnica de aprendizaje automático son los hiperplanos que durante el entrenamiento se busca aquel que tenga un margen máximo de separación entre clases.

Para conjuntos con categorías desbalanceadas en la variable respuesta, los resultados empíricos de SVM a diferencias de otros técnicas de aprendizaje automático producen inferencias confiables Akbani et al. (1), Imam et al. (92). En la misma línea, se ha mostrado también que el

hiperplano construido por SVM en presencia de conjuntos no balanceados tiene aproximadamente la misma orientación que el hiperplano ideal Wu and Chang (206).

Por lo antes expuesto, podemos decir que, el problema de la generalización para datos no balanceados está relacionado con el sesgo, ya que las instancias positivas se ubicarán lejos de este límite ideal; es decir, SVM aprende en un límite que está muy cercano a la clase minoritaria. Las estrategias que se han propuesto en la literatura, en el caso de SVM, para analizar conjuntos de datos no balanceados pueden ser clasificadas de acuerdo a:

1. **Pre-procesamiento:** Se aplican técnicas de sobre-muestreo Bae et al. (7), Nguyen et al. (134), sub-muestreo Li et al. (109), Vivaracho (192), Yu et al. (209) o una combinación de éstas Castro et al. (19), Vilariño et al. (191). Algunos trabajos aplican métodos de muestreo con combinación de clasificadores Kang and Cho (96), Liu et al. (115), Waske et al. (197), Yu et al. (209).
2. **Entrenamiento:** Se asignan diferentes costos a través del parámetro C Cohen et al. (27), Yang et al. (208). También se incluyen métodos para modificar la matriz kernel de acuerdo al desbalance observada en la distribución de los datos Wu and Chang (206). Otros esquemas combinan estrategias basadas en costo con técnicas de muestreo Akbani et al. (1), Tang et al. (184) y con combinación de clasificadores Wang and Japkowicz (195).
3. **Post-procesamiento:** Se ajusta el límite de decisión aprendido por SVM de tal forma que suministre un buen margen de separación para la clase positiva Imam et al. (92), Li et al. (108), Shanahan and Roma (173). También se incluyen las técnicas que incorporan un módulo de post-procesamiento para darle una interpretación diferente a las salidas de la SVM, de tipo probabilístico Wang and Zheng (196) o fuzzy Li et al. (107).

En general, las propuestas recogidas (que incluyen el paquete caret del programa estadístico R) en la literatura pretenden la modificación del vector de pesos w en la función de decisión, o bien, la determinación de un nuevo sesgo o umbral, por tal motivo en la siguiente sección se analizan los vectores de soporte para la construcción del modelo de clasificación.

5.7.2. Vectores de Soporte para elección del modelo estadístico

Se ejecutaron tres tipos de algoritmos de aprendizaje para los 1186 meta-análisis de pruebas diagnósticas generados mediante la función HSROC del programa estadístico R: Regresión Logística (LOGIT), Redes Neuronales (NNET) y Vectores de Soporte (SVM). Con el algoritmo de SVM se obtuvieron errores de entrenamiento y de prueba menores en comparación con los otros modelos, asimismo el índice de Kappa resulto mayor con SVM en comparación con las otras técnicas. Por lo antes mencionado, se eligió SVM para el proceso de sugerir el modelo mas adecuado, ver Figura 5.26.

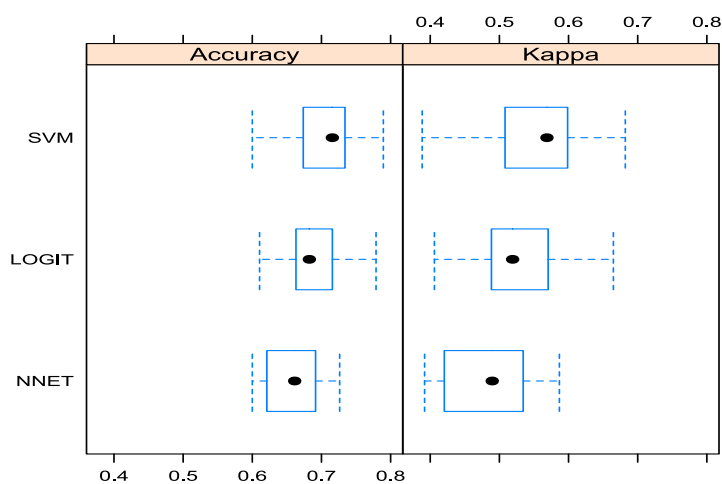


Figura 5.26: Exactitud e Índice Kappa según modelos de aprendizaje

Se “construyeron” variables predictoras para ajustar el modelo de aprendizaje automático, estas variables fueron construidas a partir de las transformaciones **logit** de los valores de la tabla tetracórica (Sensibilidad y Especificidad), dichas variables fueron construidas a partir de medidas estadísticas como: centrales, no centrales, dispersión, distribución y asociación, con la finalidad de evitar el problema de la **Multicolinealidad** en la modelización, , ver Tabla 5.12

5. APRENDIZAJE AUTOMÁTICO

Tipo de medida	Descripción
Centrales	Media y mediana
No centrales	Percentiles (primer, y tercer cuartil), número de estudios e individuos en el meta-análisis
Dispersión	Varianza y desviación estándar
Distribución	Coefficiente de asimetría y kurtosis
Asociación	Coefficiente de correlación

Tabla 5.12: Tipos de medidas estadísticas empleadas para la construcción de variables predictoras

De lo enunciado en el párrafo anterior, resultaron 26 variables predictoras; estas variables se pueden resumir en:

Tabla 5.13: Descripción de variables según su tipo

Medida	Variable
Centrales	1. Media de las transformaciones logit de sensibilidad
	2. Media de las transformaciones logit de sensibilidad
	3. Media de las transformaciones logit de sensibilidad
	4. Mediana de las transformaciones logit de sensibilidad
	5. Mediana de las transformaciones logit de especificidad.
	6. Mediana de las transformaciones logit de tasas de falsos positivos

Continúa siguiente página

5.7 Elección del Modelo Estadístico en Meta-análisis de Pruebas Diagnósticas
mediante Aprendizaje Automático

Tabla 5.13 – *Descripción de variables según su tipo*

Medida	Variable	
No centrales	7. Total de verdaderos positivos y falsos negativos.	
	8. Total de estudios en el meta-análisis.	
	9. Total de individuos en el meta-análisis.	
	10. Cuantil 1 de las transformaciones logit de sensibilidad.	
	11. Cuantil 1 de las transformaciones logit de especificidad.	
	12. Cuantil 1 de las transformaciones logit de tasas de falsos positivos	
	13. Cuantil 3 de las transformaciones logit de sensibilidad.	
	14. Cuantil 3 de las transformaciones logit de especificidad.	
	15. Cuantil 3 de las transformaciones logit de tasas de falsos positivos	
	Dispersión	16. Desviación estándar de las transformaciones logit de sensibilidad.
		17. Desviación estándar de las transformaciones logit de especificidad.
		18. Desviación estándar de las transformaciones logit de tasa de falsos positivos.

Continúa siguiente página

5. APRENDIZAJE AUTOMÁTICO

Tabla 5.13 – *Descripción de variables según su tipo*

Medida	Variable
Distribución	19. Sesgo de las transformaciones logit de sensibilidad.
	20. Sesgo de las transformaciones logit de especificidad.
	21. Sesgo de las transformaciones logit de tasa de falsos positivos.
	22. Curtosis de las transformaciones logit de sensibilidad.
	23. Curtosis de las transformaciones logit de especificidad.
Asociación	24. Curtosis de las transformaciones logit de tasa de falsos positivos
	25. Correlación entre las transformaciones logit de sensibilidad y especificidad.
	26. Correlación entre las transformaciones logit de sensibilidad y tasa de falsos positivos.

Emplearemos límites “lineales” y “no lineales” entre las clases (modelo a ajustar en el meta-análisis) para así aumentar el espacio de las variables predictoras mediante funciones polinómicas o con términos de interacción denominados “kernels”, un enfoque computacionalmente más eficiente. Los kernels son funciones que transforman un espacio de pocas dimensiones en un espacio de dimensiones mayores mediante transformaciones complejas de los datos. También puede definirse como una función que cuantifica la similitud entre dos observaciones en un nuevo espacio dimensional. Generaremos modelos basados en SVM con tres tipos de kernel: lineal, polinómico y radial, que predigan qué tipo de modelo (Bivariante, HSROC y Copula) hay que utilizar en cada meta-análisis simulado, en función del conjunto de predictores.

El uso de modelos complejos de clasificación y regresión se está volviendo cada vez más común en distintas ramas de la ciencias. Por tal motivo, para emplear los límites antes mencionados es necesario recurrir a un paquete estadístico que pueda ajustar los modelos antes citados. El

lenguaje R tiene un rico conjunto de funciones de modelización tanto para la clasificación como para la regresión. El paquete **caret**, abreviatura de clasificación y entrenamiento de regresión, se empleará para construir la máquina de aprendizaje mediante SVM. La Tabla 5.14 muestra los errores de entrenamiento y prueba de los kernel empleados en la modelización de SVM.

Kernel	Training	Test
Lineal	28.27 %	27.31 %
Polinomial	12.03 %	33.25 %
Radial	25.53 %	29.41 %

Tabla 5.14: Porcentaje de los errores de entrenamiento y prueba según kernel del modelo SVM

Preparación de los datos

Dado que existe una cantidad mínima de datos para usar en el entrenamiento, afinación y evaluación de los modelos, una de las primeras tareas es determinar cómo se deben utilizar las muestras. Estadísticamente, el uso más eficiente de los datos es entrenar el modelo usando todas las muestras y usar el remuestreo (por ejemplo, la validación cruzada, el bootstrap, etc.) para evaluar la eficacia del modelo. Dependiendo de cómo se vaya a utilizar el modelo, es posible que se necesite un conjunto de muestras de prueba (test)/validación para poder caracterizar el rendimiento del modelo a partir de datos que no se utilizaron en el entrenamiento con el modelo. Antes de proceder a generar los modelos, dividimos el set de datos en un grupo de entrenamiento (para el ajuste de los modelos) y otro de test (para la evaluación de los mismos). Esta división dependerá de la cantidad de observaciones con las que contemos y la seguridad con la que queramos obtener la estimación del test error.

Se realizará una división inicial de los datos en conjuntos de entrenamiento y pruebas (test). El conjunto de prueba se utilizará únicamente para evaluar el desempeño (por ejemplo, para comparar modelos) y el equipo de entrenamiento se utilizará para ajustar el modelo. La función **sample.split** se puede utilizar para crear muestras aleatorias estratificadas del conjunto de datos simulado. En este caso, el 80% de los datos se utilizarán para el entrenamiento de los modelos y el resto será usado para evaluar el desempeño del modelo, Guyon (72). La función crea las particiones aleatorias dentro de cada clase para que la distribución global de la clase se

5. APRENDIZAJE AUTOMÁTICO

conserve lo mejor posible.

```
> # divide los datos en training and test data
> set.seed(1234)
> split <- sample.split(datos$model, SplitRatio = 0.80)
> training_set <- subset(datos, split == TRUE)
> test_set <- subset(datos, split == FALSE)
> prop.table(table(training_set$model))
```

Bivariante	HSROC	Copula
28.48%	37.97%	33.55%

```
> prop.table(table(test_set$model))
```

Bivariante	HSROC	Copula
28.57%	37.81%	33.62%

En los modelos lineales, el método tradicional para reducir la multicolinealidad es identificar los predictores inflactados utilizando el factor de inflactación de la varianza (VIF). Para cada variable, este estadístico mide el aumento en la variación de la estimación de parámetros del modelo en comparación con la situación óptima (es decir, un diseño ortogonal). Esta es una técnica aceptable cuando se utilizan modelos lineales y hay más observaciones que predictores. En otros casos, puede no ser tan apropiado.

Como alternativa, podemos calcular la matriz de correlación de los predictores y utilizar un algoritmo para eliminar un subconjunto de los predictores problemáticos de tal manera que todas las correlaciones por pares estén por debajo de un umbral, en nuestro caso 0.90. El

Algoritmo 3 resume lo antes dicho.

Resultado: Eliminar predictores con problemas de multicolinealidad

repita

- Encuentra el par de predictores con el mayor valor de correlación en terminos absoluto;
- Para ambos predictores, calcule la correlación media entre cada predictor y todos las otras variables;
- Marque la variable con la mayor correlación media ;
- Elimine esta fila y columna de la matriz de correlación;

hasta que ninguna correlación esté por encima de un umbral;

Algoritmo 3: Algoritmo para tratar la multicolinealidad

El Algoritmo 3 se puede utilizar para determinar el conjunto mínimo de predictores que se pueden eliminar, de modo que las correlaciones por parejas estén por debajo de un umbral específico. Nótese que, si dos variables tienen una alta correlación, el algoritmo determina cuál de ellas está involucrada con la mayor parte de las correlaciones por pares y se elimina. Para nuestro caso de los meta-análisis simulados, podemos evaluar qué predictores tuvieron el mayor impacto sobre el modelo (support vector classifier), ver Figura 5.27.

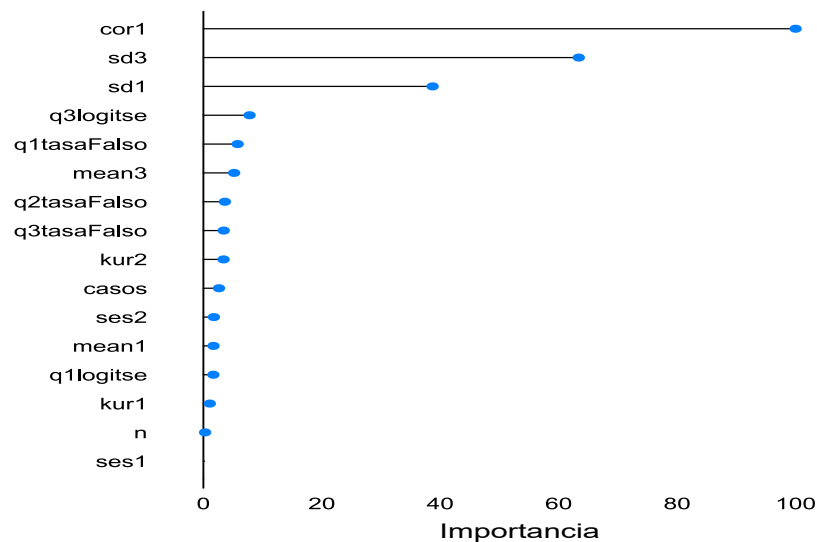


Figura 5.27: Importancia de los predictores en el modelo SVM

Los predictores que dan lugar a correlaciones absolutas por pares superiores a 0.90 pueden eliminarse utilizando la función **findCorrelation**. Esta función devuelve un índice de números

5. APRENDIZAJE AUTOMÁTICO

de columna para su eliminación.

```
> #Eliminando la colinealidad
> trainDescr<-training_set[,1:26]
> testDescr<-test_set[,1:26]
> trainClass <- training_set[,27]
> testClass <- test_set[,27]
> ncol(training_set[,1:26])
```

[1] 26

```
> descrCorr <- cor(trainDescr[,1:26])
> highCorr <- findCorrelation(descrCorr, 0.90)
> trainDescr <- trainDescr[, -highCorr]
> testDescr <- testDescr[, -highCorr]
> ncol(trainDescr)
```

[1] 16

Es evidente en la Figura 5.27 que los predictores con mayor relevancia o aporte al ajustar un “support vector classifier”, son:

1. cor1, correlación entre los logit de sensibilidad y especificidad
2. sd3, desviación estándar de la transformación logit de falsos positivos
3. sd1, desviación estándar de la transformación logit de sensibilidad
4. q3logitse, cuantil 3 de la transformación logit de la sensibilidad
5. q1tasaFalso, cuantil 1 de la tasa de falsos positivos
6. mean3, media de la transformación logit de los falsos positivos
7. q2tasaFalso, cuantil 1 de la tasa de falso positivos
8. q3tasaFalso, cuantil 3 de la tasa de falso positivos
9. kur2, curtosis de la transformación logit de especificidad
10. casos, total de verdaderos positivos y falsos negativos

11. *ses2*, sesgo de la transformación logit de especificidad
12. *mean1*, media de la transformación logit de sensibilidad
13. *q1logitse*, cuantil 1 de la transformación logit de sensibilidad
14. *kur1*, curtosis de la transformación logit de sensibilidad
15. *n*, número de estudios en el meta-análisis
16. *ses1*, sesgo de la transformación logit de sensibilidad

En otras palabras, de 26 variables predictoras el Algoritmo ?? redujo a 16 predictoras que aportan información relevante a la modelización de un “support vector classifier”, siendo las variables *cor1* y *sd3* con mayor aporte al modelo. Es claro que existe un problema de variables predictoras relacionadas, lo que también es evidente, que al incluir solo en la modelización las variables de la Figura 5.27 los resultados no mejoraron (llegando a obtener incluso errores mayores) los porcentajes errores de entrenamiento y prueba de la Tabla 5.14, por tal motivo en la secciones siguientes se utilizarán los 26 predictores para los distintos *kernels* que se empelarán.

A la hora de ajustar un “support vector classifier”, es importante tener en cuenta que el hiperparámetro *C* (cost) controla el equilibrio “bias-varianza” y la capacidad predictiva del modelo, ya que determina la severidad que se permite respecto a las violaciones sobre el margen. En otras palabras, necesitamos fijar un margen de separación entre observaciones a priori. Por ello es recomendable evaluar distintos valores del mismo mediante validación cruzada y escoger el valor óptimo.

Ajuste del Modelo, kernel lineal

Para ajustar un “support vector classifier”, el kernel indicado en la función “*tune()*” ha de ser lineal. La función ***tune()*** realiza una validación cruzada para un conjunto de datos de interés (datos de entrenamiento), para utilizar esta función, pasamos información relevante sobre el modelo que se está considerando. El siguiente comando indica que queremos modelizar un SVM con un kernel lineal, usando un rango de valores para el parámetro costo e indicando que hay que **escalar** los datos a través del argumento **scale** .

```
> #---- Entrenando modelo ----  
> # comenzando con datos de entrenamiento y kernel lineal
```

5. APRENDIZAJE AUTOMÁTICO

```
> set.seed(32)
> tuning <- tune(svm, model~., data = training_set, kernel = "linear",
+               ranges = list(cost = c(0.001, 0.01, 0.1, 1, 5, 10, 50)),
+               scale = TRUE)
```

Podemos acceder fácilmente a los errores de validación cruzada para el kernel lineal, mediante el comando `summary()`. Con dicho comando, se pueden observar los errores de validación para cada de “cost”. El valor de coste que resulta en el menor error de validación (0.3028) es $cost = 0.1$, vea Figura 5.28.

```
> summary(tuning)
Parameter tuning of 'svm':

- sampling method: 10-fold cross validation

- best parameters:
cost
0.1

- best performance: 0.3028443

- Detailed performance results:
cost      error dispersion
1 1e-03 0.3934378 0.06222650
2 1e-02 0.3080403 0.04726678
3 1e-01 0.3028443 0.04789641
4 1e+00 0.3102128 0.04263591
5 5e+00 0.3112542 0.04074573
6 1e+01 0.3144233 0.04096269
7 5e+01 0.3144233 0.04096269
```

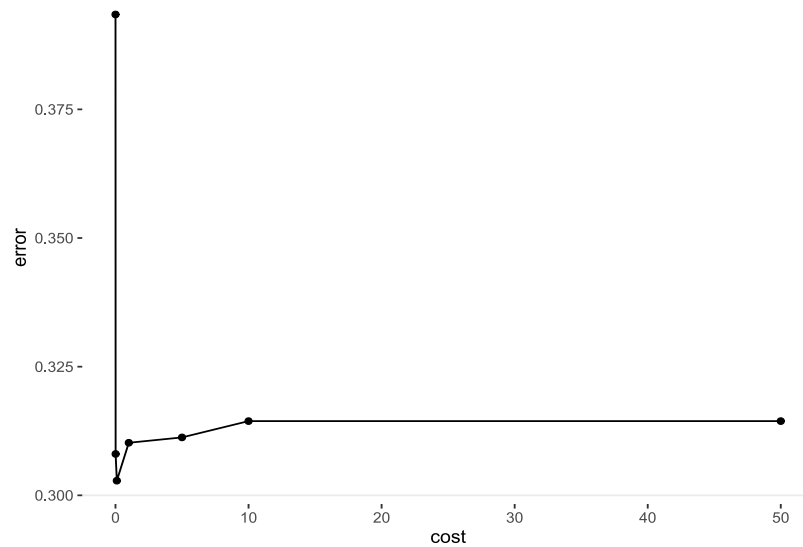



Figura 5.28: Error de Validación , hiperparámetro C

La función *tune()* almacena el mejor modelo obtenido, y el número de vectores soporte para cada clase.

```
> # seleccionando mejor modelo del kernel lineal  
> mejor.modelo.svc <- tuning$best.model  
> summary(mejor.modelo.svc)
```

Call:

```
best.tune(method = svm, train.x = model ~ ., data = training_set,  
ranges = list(cost = c(0.001, 0.01, 0.1, 1, 5, 10, 50)), kernel = "linear",  
scale = TRUE)
```

Parameters:

SVM-Type: C-classification

SVM-Kernel: linear

cost: 0.1

Number of Support Vectors: 612

5. APRENDIZAJE AUTOMÁTICO

```
( 212 207 193 )
```

```
Number of Classes: 3
```

```
Levels:
```

```
Bivariante HSROC Copula
```

Luego, el mejor modelo obtenido sería equivalente a ajustar:

```
> modelo.svc <- svm(model ~ ., data = training_set,  
+                   kernel = "linear",  
+                   cost = 0.1,  
+                   scale = TRUE)
```

Con un valor de $cost = 0.1$, el 27,31 % de las observaciones son incorrectamente clasificadas. El error de entrenamiento sobrestima el error de test, ya que el modelo es mejor prediciendo las observaciones de prueba, Tabla 5.14.

Como se ha comentado anteriormente para nuestros datos, el “preprocesamiento” es una tarea obligatoria, es decir, centrar y escalar los datos. Después del “preprocesamiento”, estos convierten nuestros datos de entrenamiento, a valores alrededor de 0 y con desviación estándar de 1. Luego es un modelo lineal, ajustándose con un valor $C = 0.01$, ver Figura 5.29, estimado las siguientes medidas: $exactitud = 0.70$ y $kappa = 0.54$.

```
Support Vector Machines with Linear Kernel
```

```
948 samples
```

```
26 predictor
```

```
3 classes: 'Bivariante', 'HSROC', 'Copula'
```

```
Pre-processing: centered (26), scaled (26)
```

```
Resampling: Cross-Validated (10 fold)
```

```
Summary of sample sizes: 854, 853, 853, 853, 853, 853, ...
```

```
Resampling results across tuning parameters:
```

5.7 Elección del Modelo Estadístico en Meta-análisis de Pruebas Diagnósticas mediante Aprendizaje Automático

C	Accuracy	Kappa
1e-03	0.6666517	0.5028944
1e-02	0.7014670	0.5473085
1e-01	0.6961702	0.5388617
1e+00	0.6993393	0.5439995
5e+00	0.6961702	0.5391207
1e+01	0.6993505	0.5441429
5e+01	0.6951288	0.5377677

Accuracy was used to select the optimal model using the largest value.
The final value used for the model was $C = 0.01$.

Usando la matriz de confusión, podemos observar estadísticas de nuestros resultados. La misma muestra que la precisión de nuestro modelo, usando los datos de pruebas es de 72.69 %.

Confusion Matrix and Statistics

Reference

Prediction	Bivariante	HSROC	Copula
Bivariante	48	10	14
HSROC	15	74	15
Copula	5	6	51

Overall Statistics

Accuracy : 0.7269

95% CI : (0.6656, 0.7825)

No Information Rate : 0.3782

P-Value [Acc > NIR] : < 2e-16

Kappa : 0.5867

Mcnemar's Test P-Value : 0.02773

5. APRENDIZAJE AUTOMÁTICO

Statistics by Class:

Class:	Bivariante	HSROC	Copula
Sensitivity	0.7059	0.8222	0.6375
Specificity	0.8588	0.7973	0.9304
Pos Pred Value	0.6667	0.7115	0.8226
Neg Pred Value	0.8795	0.8806	0.8352
Prevalence	0.2857	0.3782	0.3361
Detection Rate	0.2017	0.3109	0.2143
Detection Prevalence	0.3025	0.4370	0.2605
Balanced Accuracy	0.7824	0.8098	0.7839

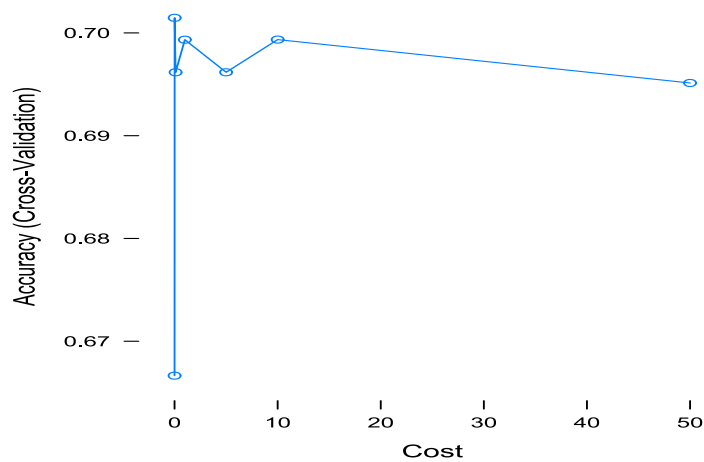


Figura 5.29: Exactitud para diferentes valores de C , kernel lineal

Además del hiperparámetro de penalización C , en los modelos SVM es necesario especificar el hiperparámetro γ (para kernel radial) o el grado de polinomio (para kernel polinómico). Es también importante optimizar estos dos últimos, previo ajuste del modelo.

Ajuste del Modelo, kernel polinomial

Con un kernel polinómico, los hiperparámetros óptimos que reducen el error de validación son $cost = 10$, $degree = 3$, Figura 5.30.

```
> #---- Entrenando modelo ----  
> # comenzando con datos de entrenamiento y kernel polinomial  
> set.seed(32)  
> tuning <- tune(svm, model ~ ., data = training_set,  
+               kernel = "polynomial",  
+               ranges = list(cost = c(0.001, 0.01, 0.1, 1, 5, 10, 50),  
+               degree = c(2, 3, 4)),  
+               scale = TRUE)  
> summary(tuning)
```

Parameter tuning of 'svm':

- sampling method: 10-fold cross validation

- best parameters:

cost degree

10 3

- best performance: 0.3354311

- Detailed performance results:

cost	degree	error	dispersion
1	1e-03	2 0.6202016	0.05929080
2	1e-02	2 0.6202016	0.05929080
3	1e-01	2 0.5337514	0.05318391
4	1e+00	2 0.4672452	0.04726797
5	5e+00	2 0.4419933	0.05285778
6	1e+01	2 0.4324524	0.04287904
7	5e+01	2 0.4504367	0.04333352

5. APRENDIZAJE AUTOMÁTICO

8	1e-03	3	0.6202016	0.05929080
9	1e-02	3	0.5875140	0.05543643
10	1e-01	3	0.4272004	0.07667353
11	1e+00	3	0.3575588	0.05626988
12	5e+00	3	0.3470437	0.05401052
13	1e+01	3	0.3354311	0.04095787
14	5e+01	3	0.3850056	0.03992682
15	1e-03	4	0.6202016	0.05929080
16	1e-02	4	0.6086114	0.05564172
17	1e-01	4	0.5527436	0.05204909
18	1e+00	4	0.4893393	0.06900990
19	5e+00	4	0.4693729	0.04833123
20	1e+01	4	0.4736282	0.03954577
21	5e+01	4	0.4620493	0.03800096

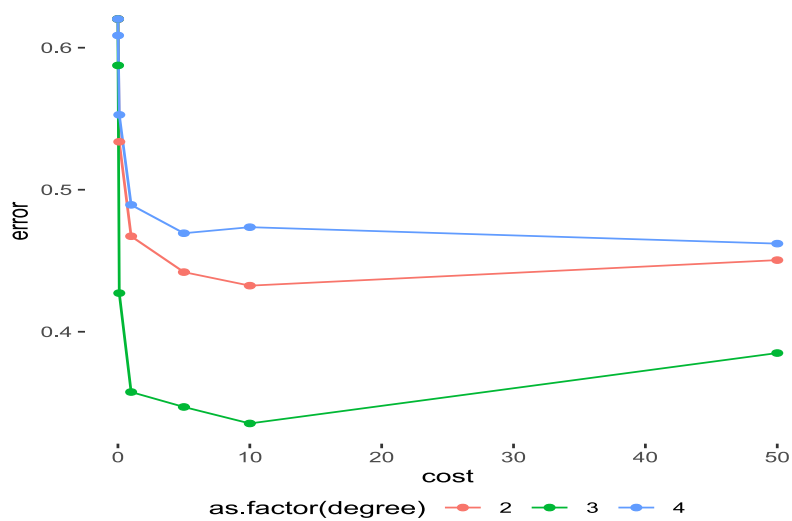


Figura 5.30: Error de validación, según costo y grado del polinomio

El modelo ajustado con un kernel polinómico de grado 3 ha aumentado los vectores soporte de 612 (kernel lineal) a 630.

```
> # Modelo SVM kernel polinomico
> modelo.svm.poli <- svm(model ~ ., data = training_set,
```

5.7 Elección del Modelo Estadístico en Meta-análisis de Pruebas Diagnósticas mediante Aprendizaje Automático

```
+             kernel = "polynomial",
+             cost = 10,
+             degree = 3,
+             scale = TRUE)
> summary(modelo.svm.poli)
```

Call:

```
svm(formula = model ~ ., data = training_set, kernel = "polynomial",
cost = 10, degree = 3, scale = TRUE)
```

Parameters:

SVM-Type: C-classification

SVM-Kernel: polynomial

cost: 10

degree: 3

coef.0: 0

Number of Support Vectors: 630

(216 197 217)

Number of Classes: 3

Levels:

Bivariante HSROC Copula

Con un valor de $cost = 10$, el 33,25% de las observaciones son incorrectamente clasificadas. El error de entrenamiento subestima el error de test 12,03, ya que el modelo es mejor prediciendo las observaciones de entrenamiento, Tabla 5.14.

Usando la matriz de confusión, podemos observar estadísticas de nuestros resultados. La misma muestra que la precisión de nuestro modelo, usando los datos de entrenamiento es de 67.65%.

5. APRENDIZAJE AUTOMÁTICO

Confusion Matrix and Statistics

Reference

Prediction	Bivariante	HSROC	Copula
Bivariante	50	14	19
HSROC	10	67	17
Copula	8	9	44

Overall Statistics

Accuracy : 0.6765

95% CI : (0.613, 0.7355)

No Information Rate : 0.3782

P-Value [Acc > NIR] : < 2e-16

Kappa : 0.5134

McNemar's Test P-Value : 0.05481

Statistics by Class:

Class: Bivariante Class: HSROC Class: Copula

Sensitivity	0.7353	0.7444	0.5500
Specificity	0.8059	0.8176	0.8924
Pos Pred Value	0.6024	0.7128	0.7213
Neg Pred Value	0.8839	0.8403	0.7966
Prevalence	0.2857	0.3782	0.3361
Detection Rate	0.2101	0.2815	0.1849
Detection Prevalence	0.3487	0.3950	0.2563
Balanced Accuracy	0.7706	0.7810	0.7212

Ajuste del Modelo, kernel radial

Con un kernel radial, los hiperparámetros que reducen el error de validación son $cost = 5$, $gamma = 0,01$, ver Figura 5.31.

```
> #---- Entrenando modelo ----  
> # comenzando con datos de entrenamiento y kernel polinomial  
> #Kernel radial  
> #Ajuste del modelo  
> set.seed(32)  
> tuning <- tune(svm, model ~ ., data = training_set,  
+               kernel = "radial",  
+               ranges = list(cost = c(0.001, 0.01, 0.1, 1, 5, 10, 50),  
+               gamma = c(0.01, 0.1, 1, 5, 10)),  
+               scale = TRUE)  
> summary(tuning)
```

Parameter tuning of 'svm':

- sampling method: 10-fold cross validation

- best parameters:

cost gamma

5 0.01

- best performance: 0.3037962

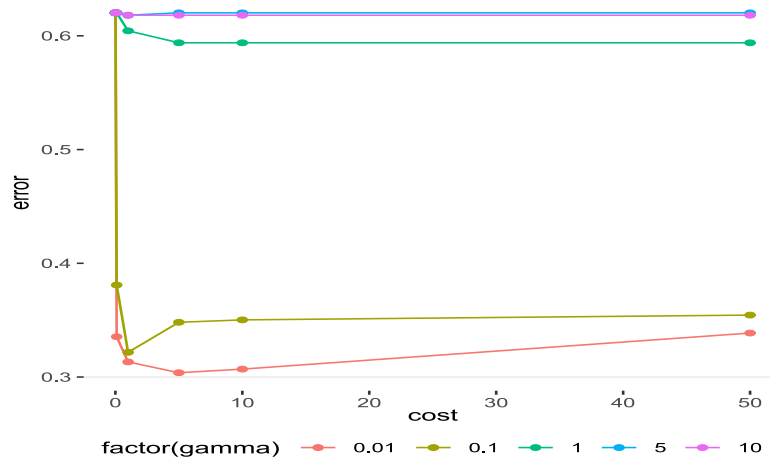


Figura 5.31: Error de validación, según costo y factor gamma

Con el uso de un kernel radial ($cost = 5$ y $gamma = 0,01$), el error de test es de 29,41 %.

Usando la matriz de confusión, podemos observar estadísticas de nuestros resultados. La misma muestra que la precisión de nuestro modelo, usando los datos de entrenamiento es de 70,59%, ver Figura 5.32.

Confusion Matrix and Statistics

Reference

Prediction	Bivariante	HSROC	Copula
Bivariante	50	14	17
HSROC	13	69	14
Copula	5	7	49

Overall Statistics

Accuracy : 0.7059

95% CI : (0.6436, 0.763)

No Information Rate : 0.3782

P-Value [Acc > NIR] : < 2e-16

5.7 Elección del Modelo Estadístico en Meta-análisis de Pruebas Diagnósticas mediante Aprendizaje Automático

Kappa : 0.5571

Mcnemar's Test P-Value : 0.03043

Statistics by Class:

Class: Bivariante	Class: HSR0C	Class: Copula	
Sensitivity	0.7353	0.7667	0.6125
Specificity	0.8176	0.8176	0.9241
Pos Pred Value	0.6173	0.7188	0.8033
Neg Pred Value	0.8854	0.8521	0.8249
Prevalence	0.2857	0.3782	0.3361
Detection Rate	0.2101	0.2899	0.2059
Detection Prevalence	0.3403	0.4034	0.2563
Balanced Accuracy	0.7765	0.7921	0.7683

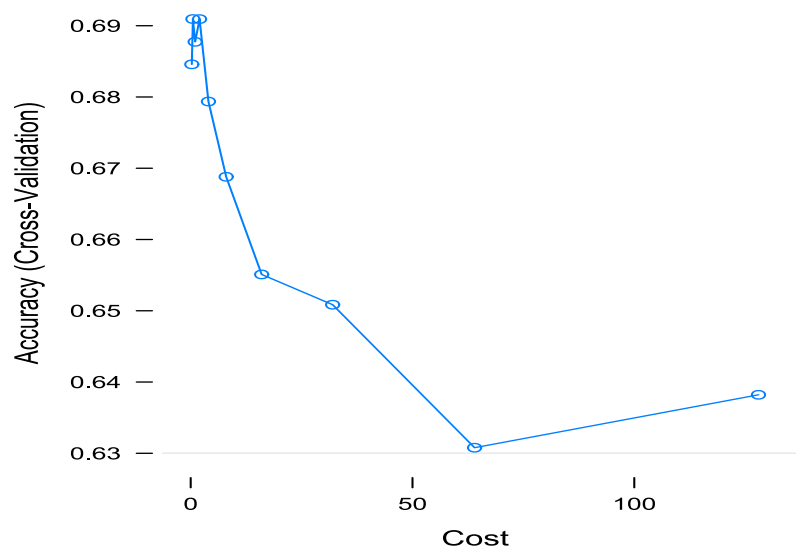


Figura 5.32: Accuracy, según costo kernel radial

Por lo anteriormente mencionado, es evidente que el kernel que mejor ajusta los datos es el polinomial, ya que tiene valores de “accuracy” y “kappa” mayores en relación a los dos kernels empleados en la modelización, ver Figura 5.33

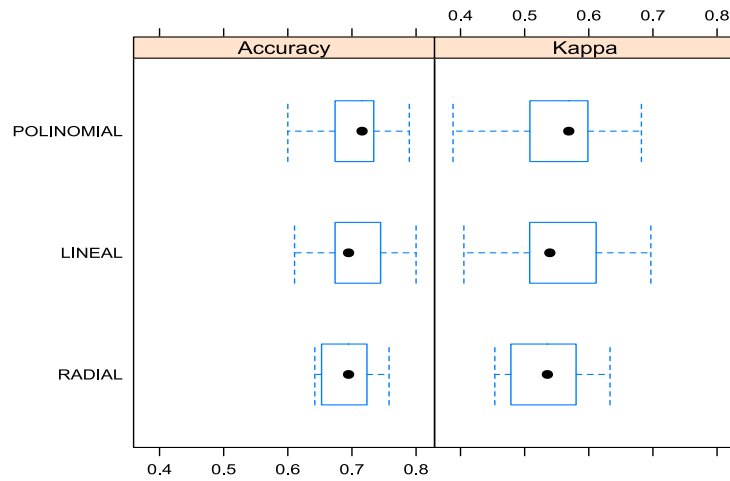


Figura 5.33: Accuracy y Kappa, según kernel

Finalmente, de los algoritmos analizados y estudiados en el presente capítulo, proponemos un diagrama de decisión con relación a la elección del tipo de aprendizaje automático y la naturaleza de los datos, ver Figura 5.34. Asimismo, la descripción de la Figura 5.34 se puede analizar mediante el Algoritmo 4.

5.7 Elección del Modelo Estadístico en Meta-análisis de Pruebas Diagnósticas mediante Aprendizaje Automático

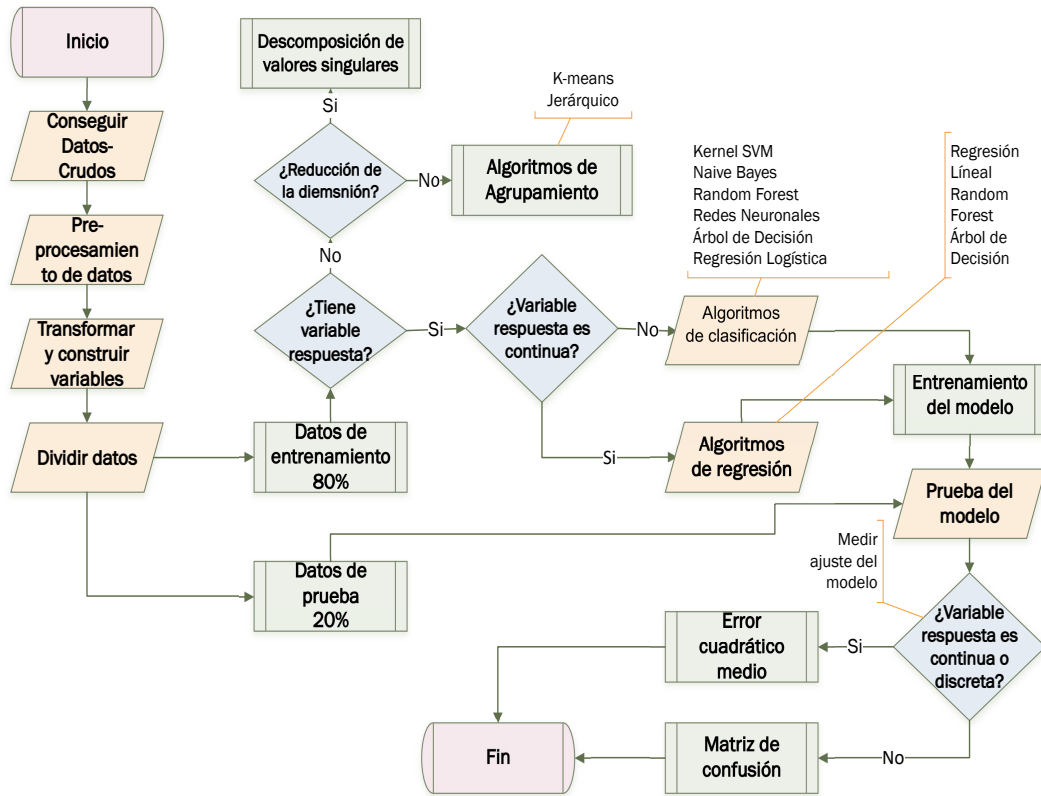


Figura 5.34: Diagrama de Decisión para realizar un aprendizaje automático, según naturaleza de los datos

5. APRENDIZAJE AUTOMÁTICO

Resultado: Realizar un aprendizaje automático, según naturaleza de los datos

Data: Conseguir Datos crudos

Pre-procesar los datos ▷ Limpieza, Integración, Transformación de datos, Reducción y Discretización de datos ;

Transformar y construir variables ▷ crear nuevas características (variables) predictoras a partir de las existentes ;

Dividir los datos en datos de entrenamiento 80 % y prueba 20 % ;

si datos tienen variable respuesta entonces

si *variable respuesta es continua entonces*

 | Emplee algoritmos de regresión y entrene el modelo

en otro caso

 | Emplee algoritmos de clasificación y entrene el modelo

fin

Ajuste modelo con datos prueba;

si *variable respuesta es continua entonces*

 | medir ajuste del modelo mediante error cuadrático medio

en otro caso

 | medir ajuste del modelo mediante matriz de confusión

fin

en otro caso

si *reducción de la dimensión entonces*

 | Descomposición de valores singulares y PCA

en otro caso

 | Emplee algoritmos de agrupamiento

fin

fin

Algoritmo 4: Algoritmo para realizar un aprendizaje automático, según naturaleza de los datos

Conclusiones

1. La modelización estadística de meta-análisis para pruebas diagnósticas se fundamenta en el modelo de Moses, el cual utiliza transformaciones logit para las tasas de verdaderos positivos y falsos positivos, mediante un modelo de regresión lineal que relaciona la precisión de la prueba y la proporción de resultados positivos.
2. Este enfoque asume homogeneidad entre los resultados de los estudios. Cuando los mismos tienen un grado de heterogeneidad superior al esperado por simple variabilidad aleatoria, es necesario recurrir a la modelización de efectos aleatorios.
3. Los modelos jerárquicos los cuales modelizan, la correlación entre sensibilidad y especificidad, y los efectos aleatorios, superan las deficiencias del modelo de Moses. Los dos enfoques principales -el modelo bivalente y el modelo HSROC- difieren en la parametrización, pero ambos son matemáticamente equivalentes cuando no se incluyen covariables.
4. Estos modelos jerárquicos suponen un comportamiento normal bivalente entre las transformaciones logit de sensibilidad y especificidad. Cuando el supuesto de normalidad bivalente no se cumple, se recurre a la modelización de cópulas, las cuales utilizan distribuciones beta-binomiales para caracterizar las distribuciones marginales de los verdaderos positivos y negativos, es decir presenta una mayor flexibilidad para capturar la dependencia funcional entre sensibilidad y especificidad.
5. En esta investigación desarrollamos un diagrama meta-analítico que guiará a los investigadores en la búsqueda del modelo óptimo y así garantizar resultados de alta calidad.
6. La elección adecuada del modelo estadístico óptimo, para resumir los resultados de un meta-análisis, requiere conocimientos estadístico-matemáticos avanzados que generalmente no tienen los investigadores meta-analíticos. Por esta razón una de las contribuciones

5. APRENDIZAJE AUTOMÁTICO

más relevantes de esta investigación es la creación de un mecanismo de elección automática usando “machine learning” estadístico.

7. El aprendizaje automático sobre el que se asientan las decisiones, está alimentado con la información proporcionada por 1186 meta-análisis de pruebas diagnósticas, de baja prevalencia, con distintos puntos de corte. De esta forma, la máquina de aprendizaje es capaz de seleccionar el mejor modelo estadístico sin que el investigador necesite una alta formación estadística-matemática.

Apéndice A

Código/Manuales/Publicaciones

A.1. Estadística descriptiva de los parámetros a estimar

Summary for N = 20000

Number of chains = 1
 Number of iteration within a chain = 50000 Burn in within each chain = 10000
 Thinning interval = 2
 Total number of iteration kept = 20000

File location : C:/Users/Usuario/Documents

Date : 2016-06-06 01:23:10

Perfect reference standard

SAMPLE SIZE

	Total	++	+-	-+	--
Study 1	57	9	2	2	44
Study 2	46	3	6	5	32
Study 3	22	3	2	1	16
Study 4	60	3	1	12	44
Study 5	24	1	1	6	16
Study 6	198	7	2	22	167
Study 7	49	12	4	4	29
Study 8	272	23	5	14	230
Study 9	71	8	5	5	53
Study 10	42	16	2	2	22

PRIOR INFORMATION

Prior of prevalence (pi) is beta (1 , 1), \Leftrightarrow pi in [0 , 1]

Prior of beta is Uniform(-0.75 , 0.75)

Prior of THETA is Uniform(-1.5 , 1.5)

Prior of LAMBDA is Uniform(-3 , 3)

Prior of sigma_alpha is uniform(0 , 2)

Prior of sigma_theta is uniform(0 , 2)

BETWEEN_STUDY parameters (Point estimate = median)

	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
THETA	0.928372	0.228954	0.011726	0.469583	1.365921
LAMBDA	2.106404	0.340238	0.019787	1.420487	2.752459
beta	-0.415436	0.264169	0.021773	-0.749537	0.120237
sigma.alpha	0.515597	0.290335	0.006616	5.6e-05	1.090918
sigma.theta	0.512763	0.188981	0.00337	0.230685	0.921319
S overall		0.55776	0.102648	0.00121	0.349044
C overall		0.947616	0.023182	0.000312	0.899218

WITHIN-STUDY PARAMETERS

theta

Summary for N = 20000
Estimate Standard_Dev MC_error C.I._lower C.I._upper

Study 1	0.717434	0.269566	0.013283	0.200333	1.252638
Study 2	0.761313	0.263249	0.013039	0.24062	1.267722
Study 3	0.628859	0.332931	0.011389	-0.032148	1.273068
Study 4	1.479935	0.276003	0.012005	0.916126	2.002552
Study 5	1.318358	0.34507	0.01275	0.678148	2.033561
Study 6	1.580732	0.213855	0.012518	1.158249	1.998937
Study 7	0.523696	0.252035	0.013539	0.02064	1.000077
Study 8	1.084833	0.191897	0.013538	0.704547	1.453734
Study 9	0.749527	0.23466	0.013352	0.274661	1.19431
Study 10	0.436283	0.279488	0.0128	-0.120885	0.979411

alpha

Estimate Standard_Dev MC_error C.I._lower C.I._upper

Study 1	2.46736	0.461626	0.01854	1.605338	3.413924
Study 2	1.546263	0.507919	0.018999	0.527815	2.492814
Study 3	2.091117	0.502131	0.018207	1.06714	3.070678
Study 4	1.924161	0.498969	0.025052	0.869954	2.825598
Study 5	1.767675	0.578755	0.023734	0.506963	2.73959
Study 6	2.124205	0.432142	0.027618	1.245024	2.956996
Study 7	2.07992	0.395889	0.016659	1.265104	2.815339
Study 8	2.55992	0.364965	0.021331	1.806921	3.225544
Study 9	2.028452	0.392633	0.017885	1.257335	2.793918
Study 10	2.430977	0.445813	0.016628	1.60232	3.348222

Prevalence

Estimate Standard_Dev MC_error C.I._lower C.I._upper

Study 1	0.200152	0.051609	0.000367	0.106957	0.306094
Study 2	0.184025	0.056113	0.00042	0.085939	0.300807
Study 3	0.200346	0.081228	0.000601	0.059669	0.363462
Study 4	0.255629	0.055353	0.000375	0.150532	0.364656
Study 5	0.303807	0.089419	0.000619	0.137763	0.48196
Study 6	0.148772	0.025197	0.00016	0.100658	0.198386
Study 7	0.332415	0.065783	0.000413	0.207833	0.462481
Study 8	0.137648	0.02098	0.00014	0.099626	0.18145
Study 9	0.188975	0.045677	0.000336	0.103775	0.279726
Study 10	0.431199	0.07431	0.000587	0.286243	0.574721

Sensitivity of test 1 (S1)

Estimate Standard_Dev MC_error C.I._lower C.I._upper

Study 1	0.73265	0.114159	0.001984	0.508552	0.939193
Study 2	0.503716	0.15282	0.004078	0.207627	0.783176
Study 3	0.689266	0.154176	0.002652	0.374697	0.952972
Study 4	0.263379	0.102553	0.002416	0.08324	0.472445
Study 5	0.290994	0.141337	0.002812	0.041681	0.566245
Study 6	0.264194	0.075501	0.002041	0.129942	0.421196
Study 7	0.73125	0.097905	0.002059	0.534789	0.905117
Study 8	0.591832	0.076565	0.001938	0.442881	0.741026
Study 9	0.621113	0.113705	0.002667	0.397975	0.835105
Study 10	0.826643	0.083212	0.002049	0.656092	0.96268

Specificity of test 1 (C1)

Estimate Standard_Dev MC_error C.I._lower C.I._upper

Study 1	0.944918	0.029476	0.000613	0.883059	0.989993
Study 2	0.892726	0.04737	0.001147	0.794837	0.968542

Summary for N = 20000						
Study	3	0.91476	0.052183	0.000844	0.800297	0.986408
Study	4	0.977134	0.019736	0.000422	0.934283	0.998795
Study	5	0.963833	0.036067	0.00067	0.882996	0.999173
Study	6	0.984777	0.00881	0.00023	0.965322	0.997326
Study	7	0.899674	0.044179	0.001189	0.805666	0.969912
Study	8	0.973305	0.010456	0.000292	0.951125	0.99041
Study	9	0.925114	0.030508	0.000709	0.858719	0.972403
Study	10	0.911237	0.047686	0.00113	0.811584	0.984596

Posterior predictive value of sensitivity of test under evaluation (S1)

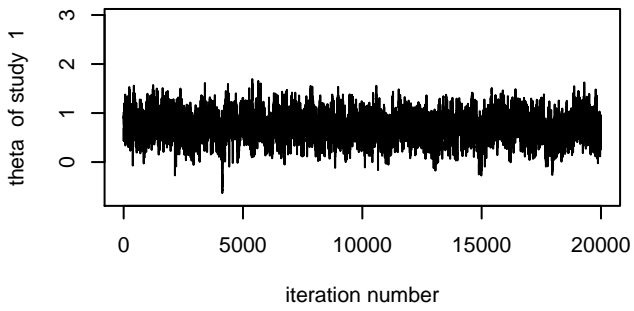
	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
Sensitivity	0.549645	0.223674	0.001684	0.126226	0.971096

Posterior predictive value of specificity of test under evaluation (C1)

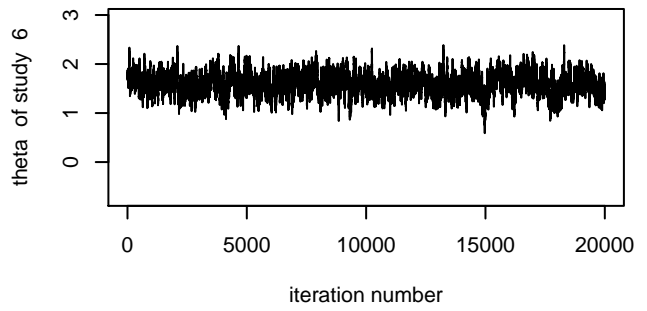
	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
Specificity	0.973323	0.095454	0.002266	0.759643	1

A.2. Gráficos de dispersión, distribución muestral a posteriori, datos MRI

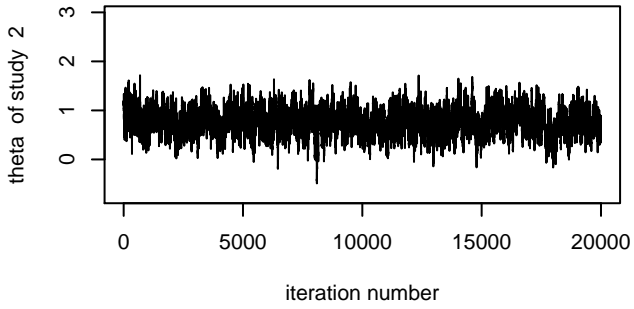
Thinning interval = 2
Total sample size kept = 20000



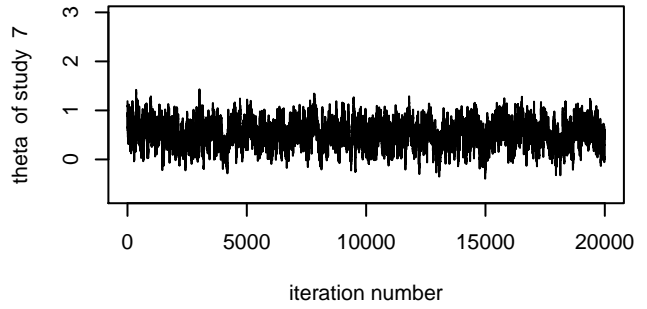
Thinning interval = 2
Total sample size kept = 20000



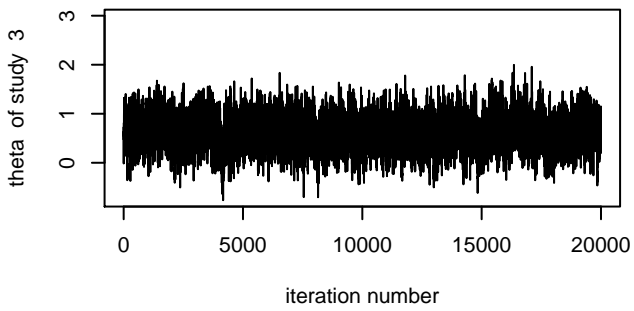
Thinning interval = 2
Total sample size kept = 20000



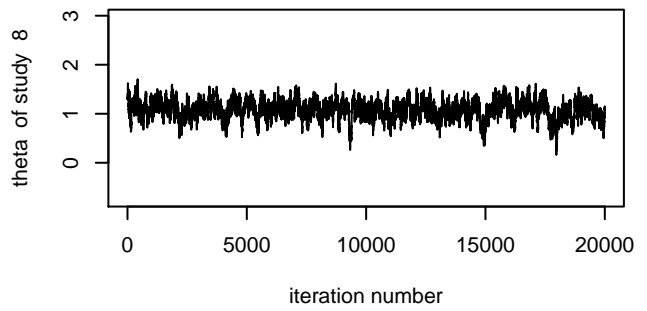
Thinning interval = 2
Total sample size kept = 20000



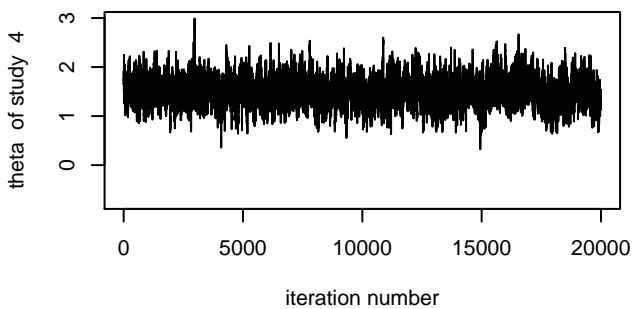
Thinning interval = 2
Total sample size kept = 20000



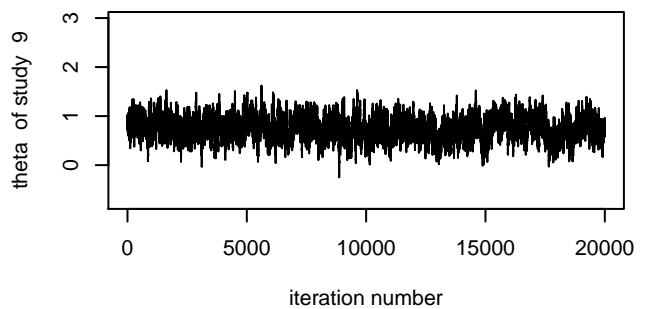
Thinning interval = 2
Total sample size kept = 20000



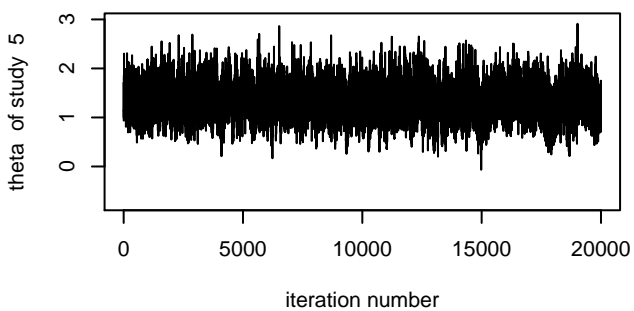
Thinning interval = 2
Total sample size kept = 20000



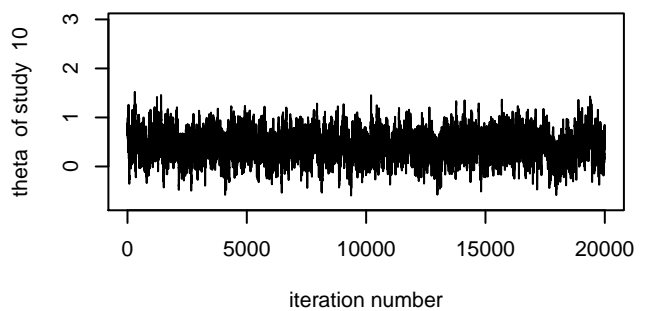
Thinning interval = 2
Total sample size kept = 20000



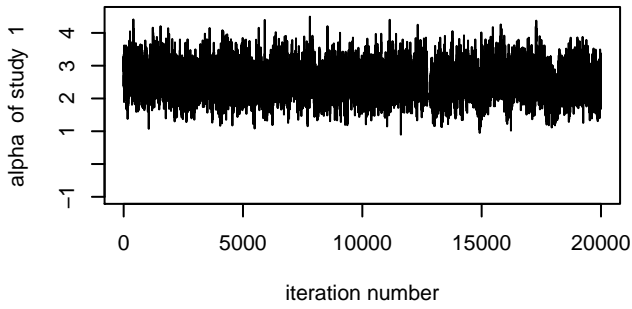
Thinning interval = 2
Total sample size kept = 20000



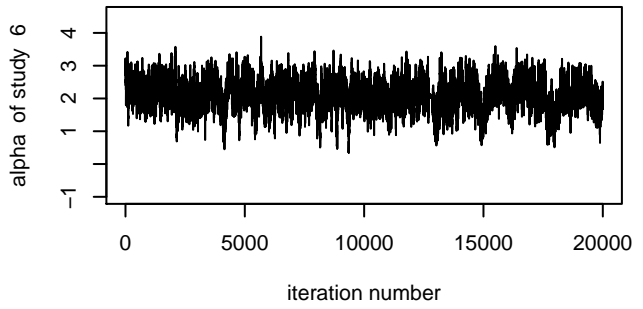
Thinning interval = 2
Total sample size kept = 20000



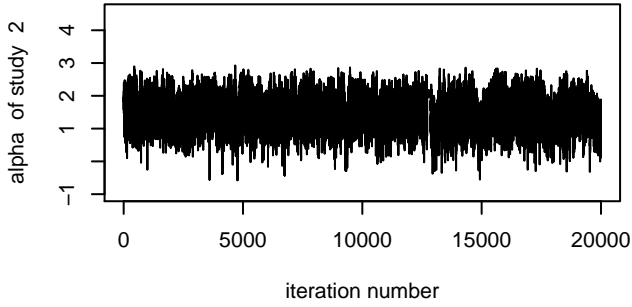
Thinning interval = 2
Total sample size kept = 20000



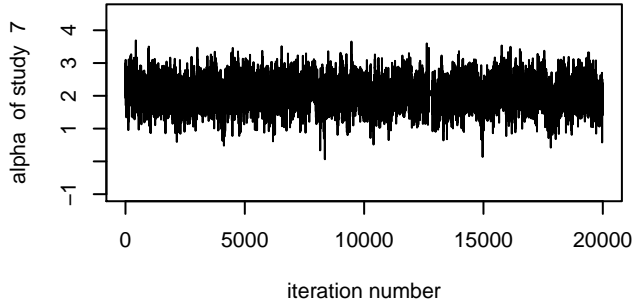
Thinning interval = 2
Total sample size kept = 20000



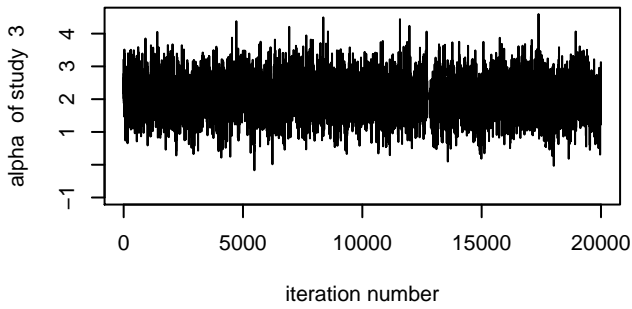
Thinning interval = 2
Total sample size kept = 20000



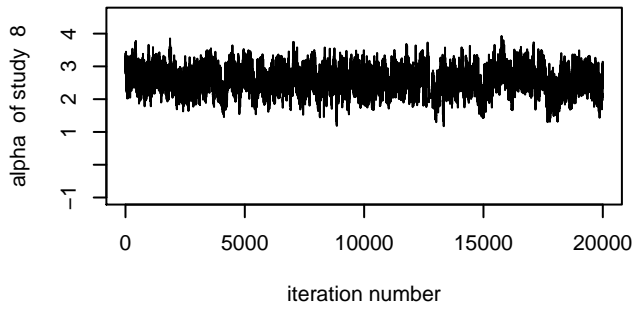
Thinning interval = 2
Total sample size kept = 20000



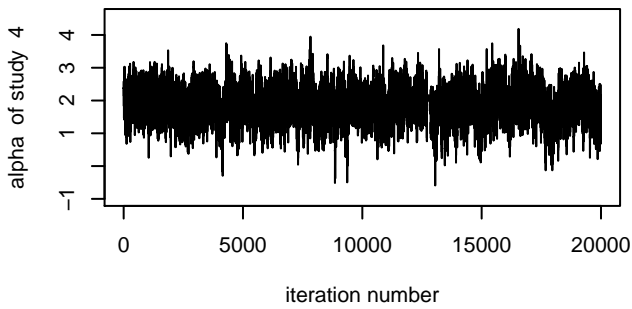
Thinning interval = 2
Total sample size kept = 20000



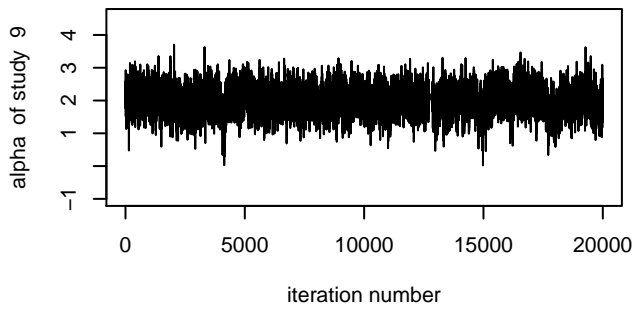
Thinning interval = 2
Total sample size kept = 20000



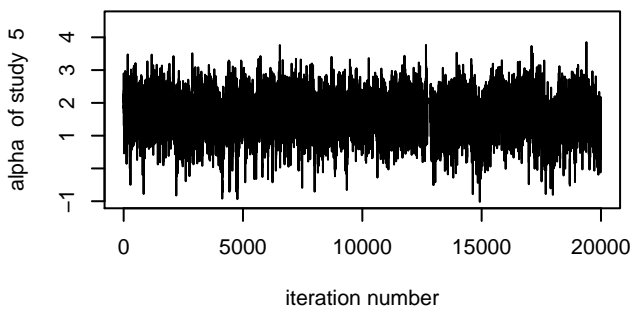
Thinning interval = 2
Total sample size kept = 20000



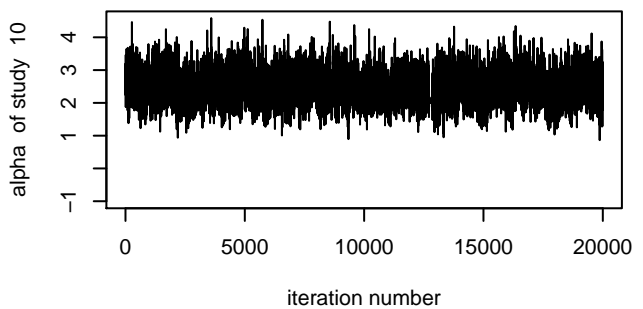
Thinning interval = 2
Total sample size kept = 20000



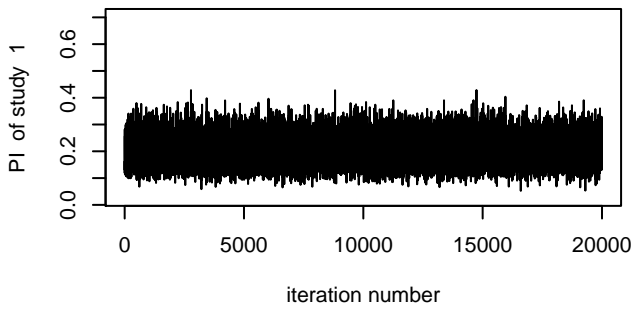
Thinning interval = 2
Total sample size kept = 20000



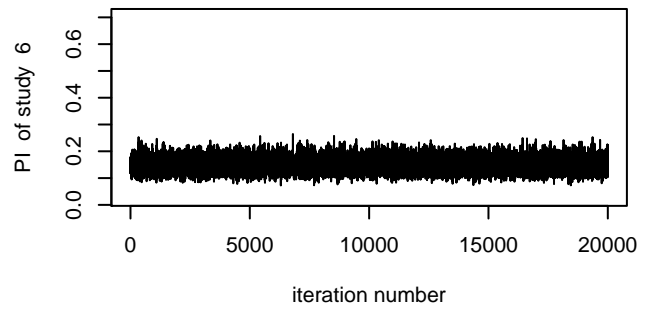
Thinning interval = 2
Total sample size kept = 20000



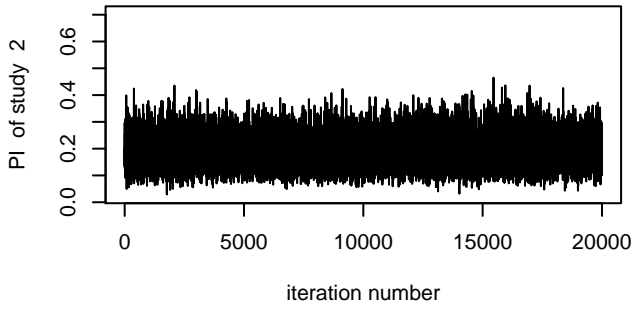
Thinning interval = 2
Total sample size kept = 20000



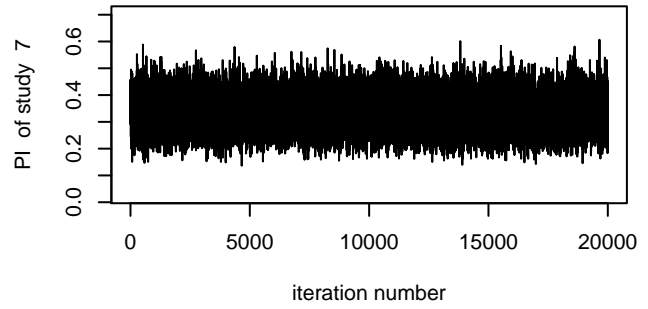
Thinning interval = 2
Total sample size kept = 20000



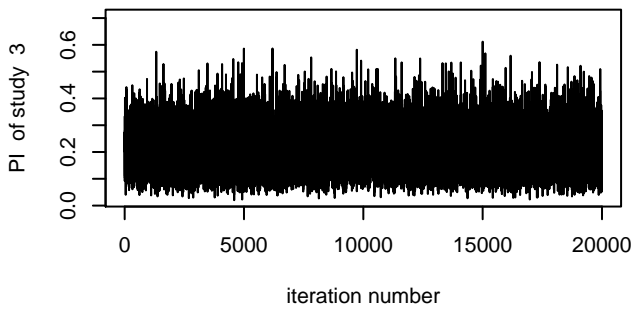
Thinning interval = 2
Total sample size kept = 20000



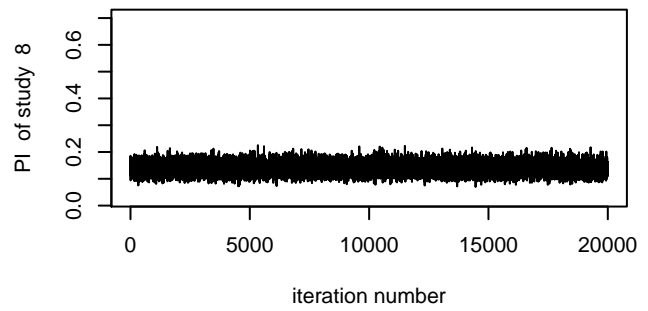
Thinning interval = 2
Total sample size kept = 20000



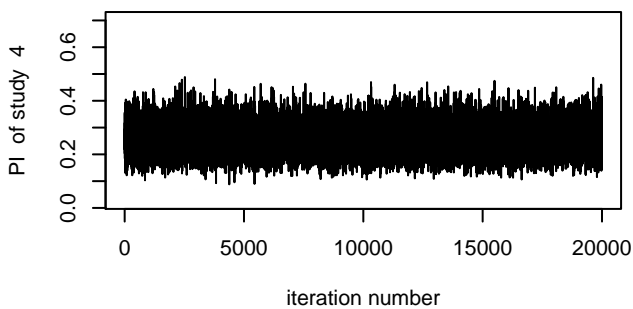
Thinning interval = 2
Total sample size kept = 20000



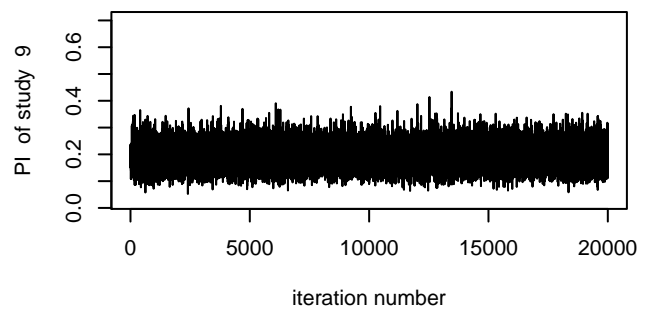
Thinning interval = 2
Total sample size kept = 20000



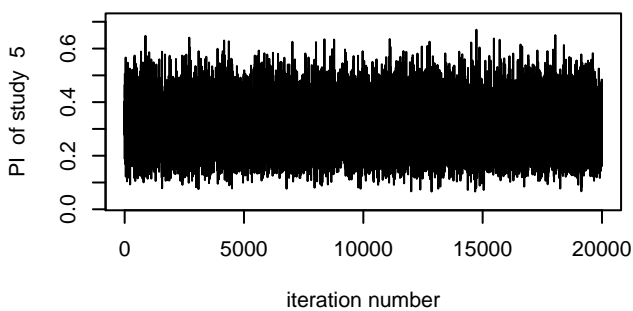
Thinning interval = 2
Total sample size kept = 20000



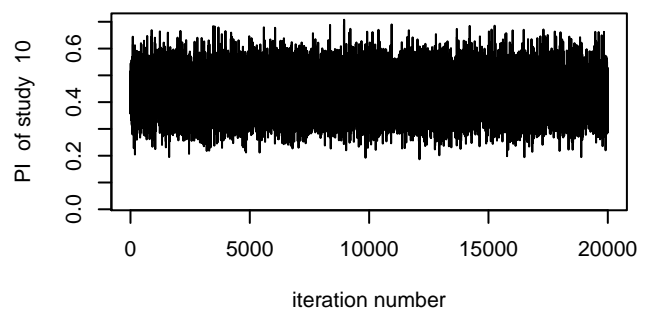
Thinning interval = 2
Total sample size kept = 20000



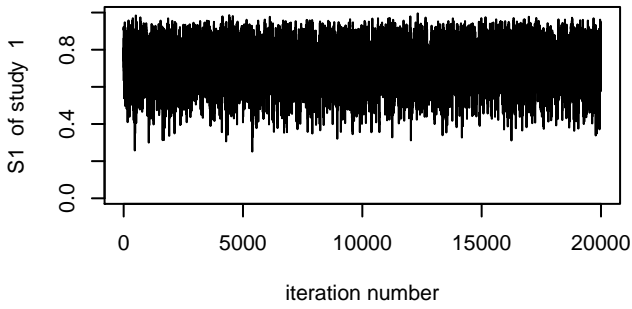
Thinning interval = 2
Total sample size kept = 20000



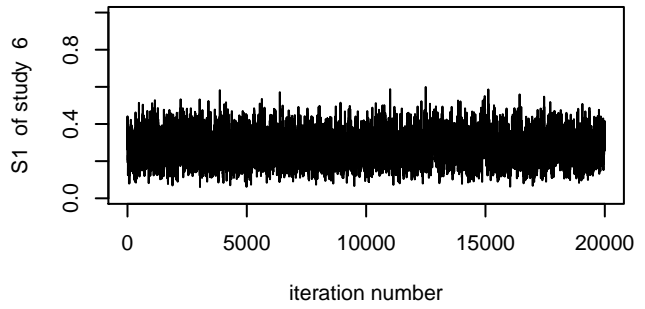
Thinning interval = 2
Total sample size kept = 20000



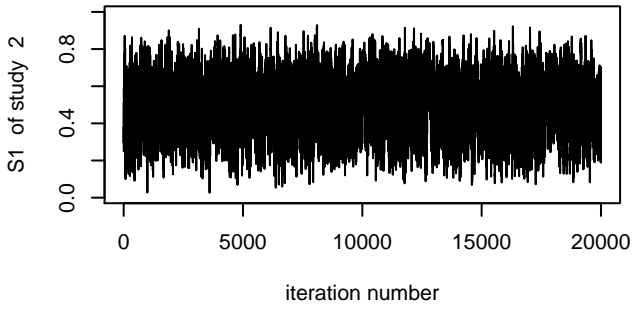
Thinning interval = 2
Total sample size kept = 20000



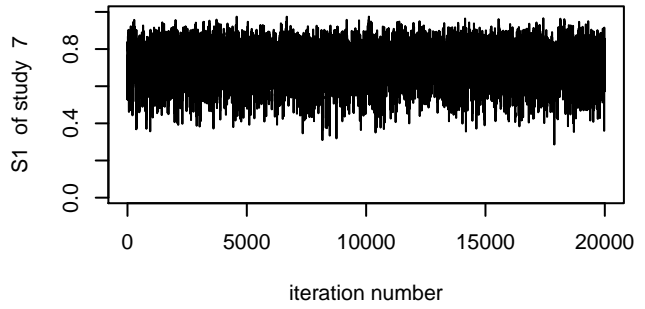
Thinning interval = 2
Total sample size kept = 20000



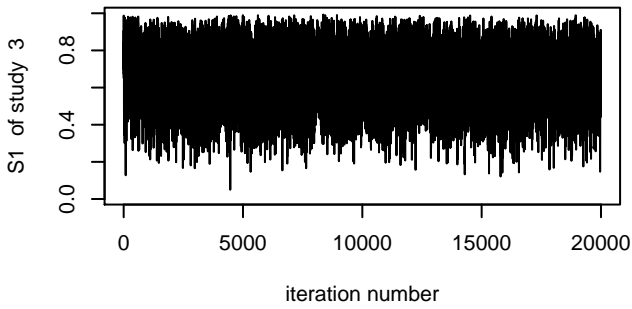
Thinning interval = 2
Total sample size kept = 20000



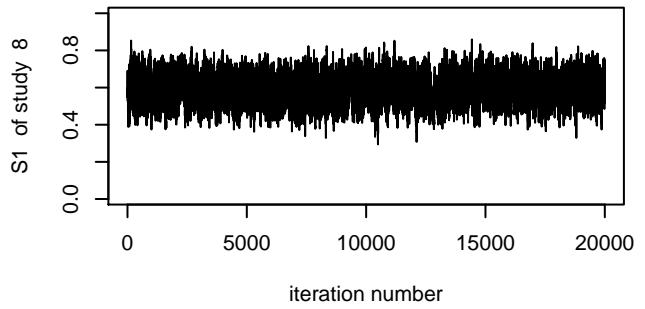
Thinning interval = 2
Total sample size kept = 20000



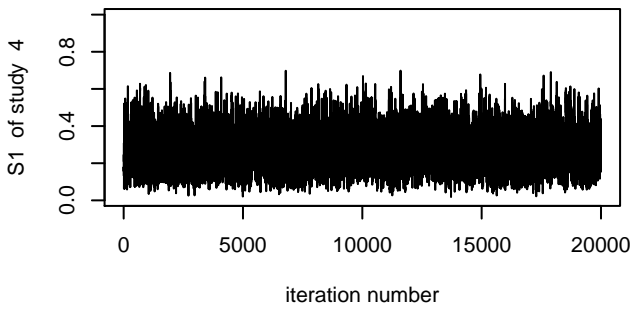
Thinning interval = 2
Total sample size kept = 20000



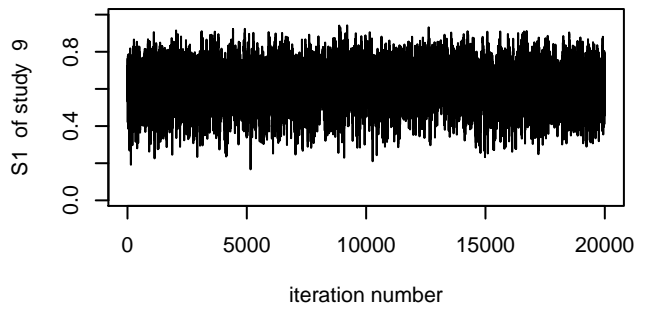
Thinning interval = 2
Total sample size kept = 20000



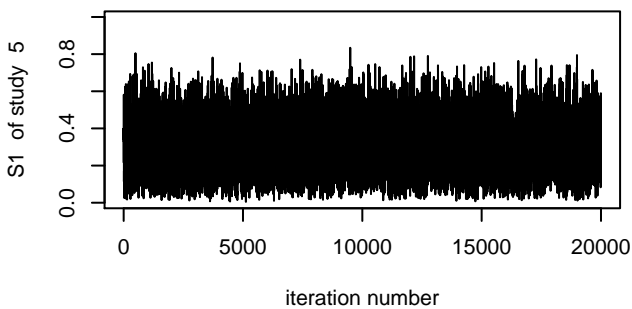
Thinning interval = 2
Total sample size kept = 20000



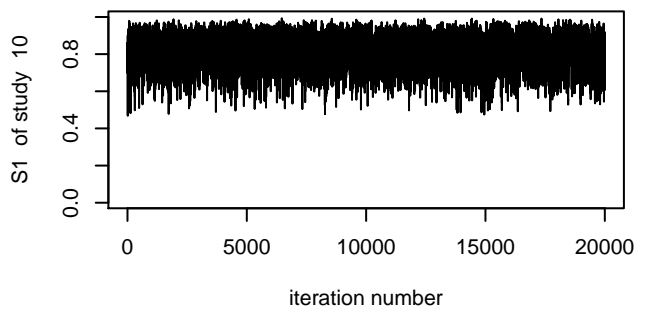
Thinning interval = 2
Total sample size kept = 20000



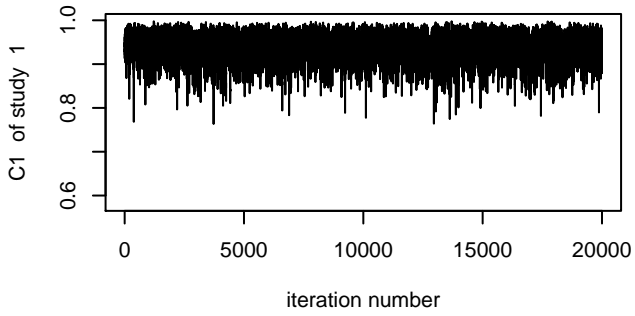
Thinning interval = 2
Total sample size kept = 20000



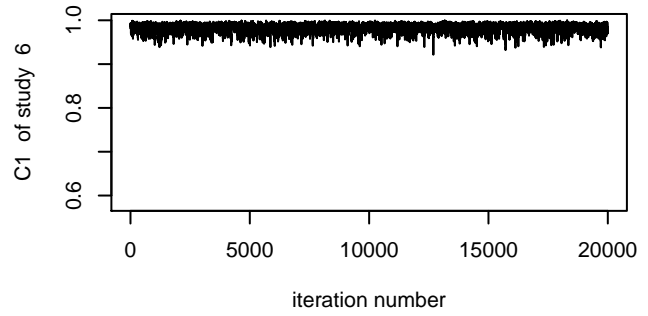
Thinning interval = 2
Total sample size kept = 20000



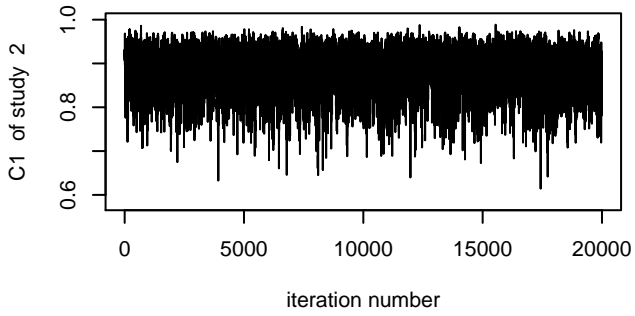
Thinning interval = 2
Total sample size kept = 20000



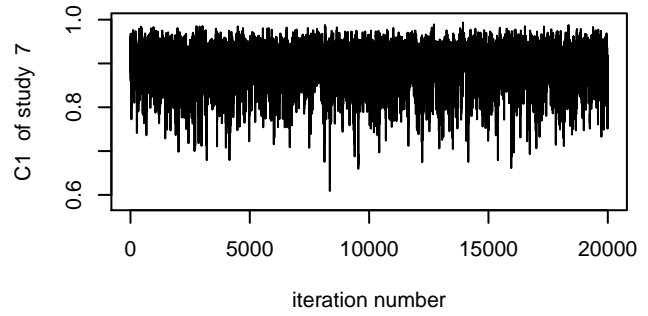
Thinning interval = 2
Total sample size kept = 20000



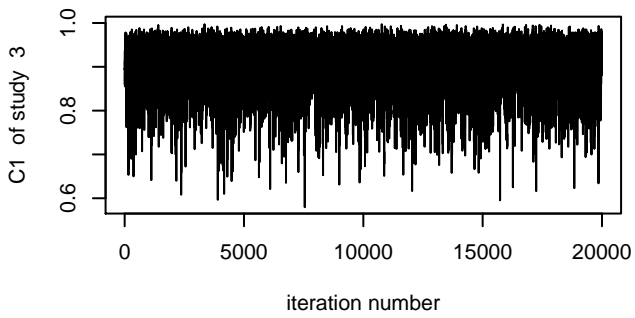
Thinning interval = 2
Total sample size kept = 20000



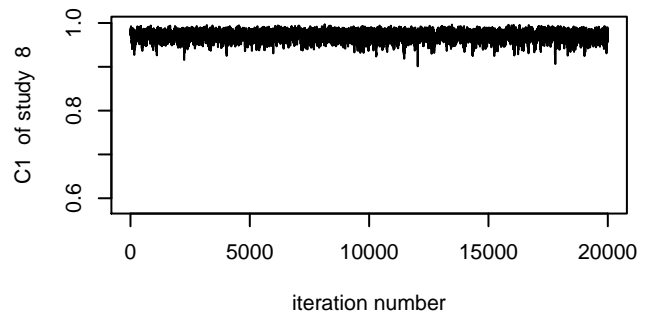
Thinning interval = 2
Total sample size kept = 20000



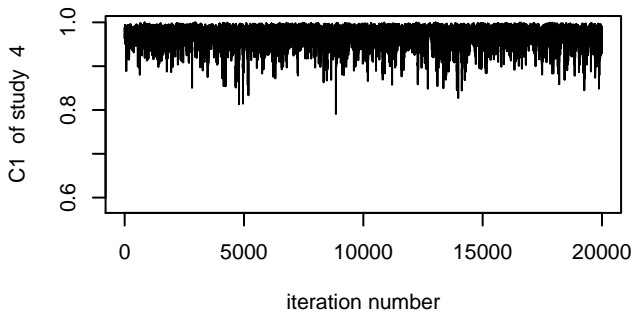
Thinning interval = 2
Total sample size kept = 20000



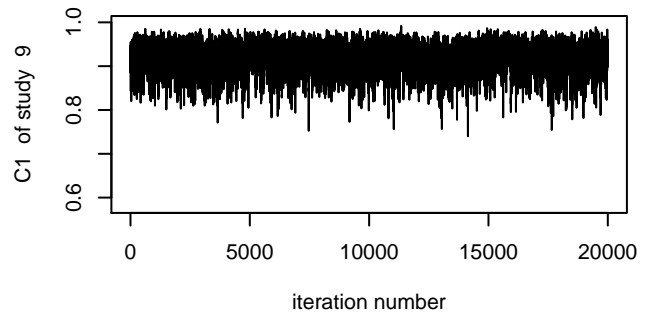
Thinning interval = 2
Total sample size kept = 20000



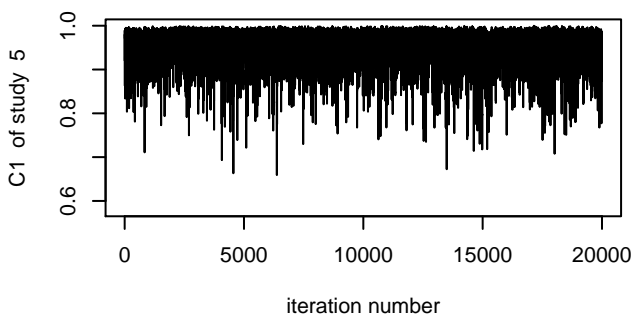
Thinning interval = 2
Total sample size kept = 20000



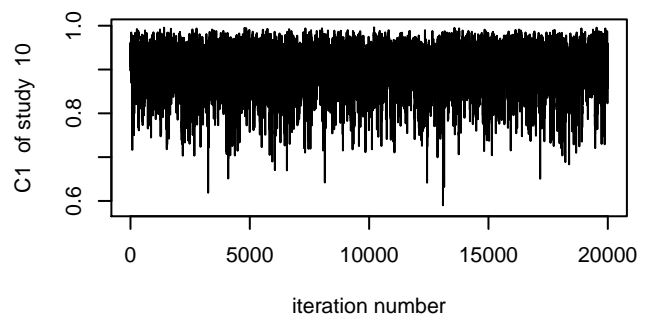
Thinning interval = 2
Total sample size kept = 20000

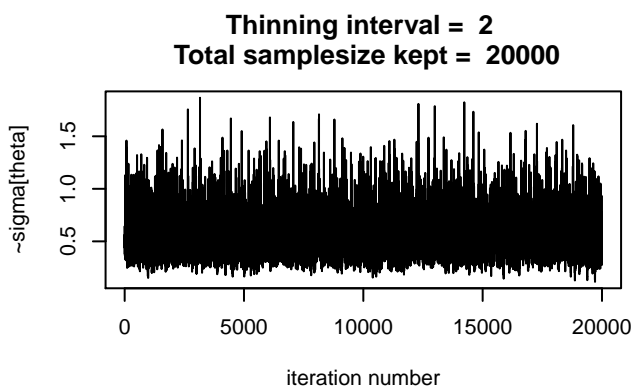
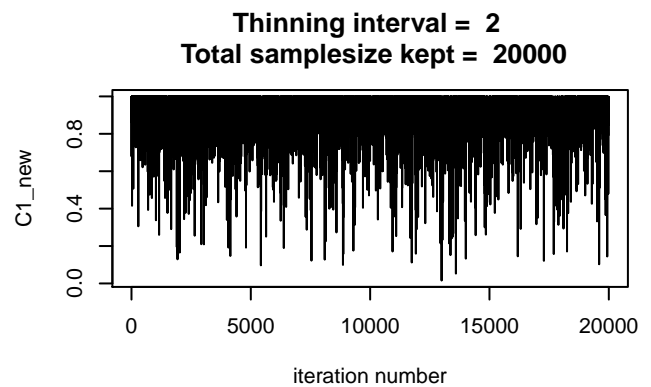
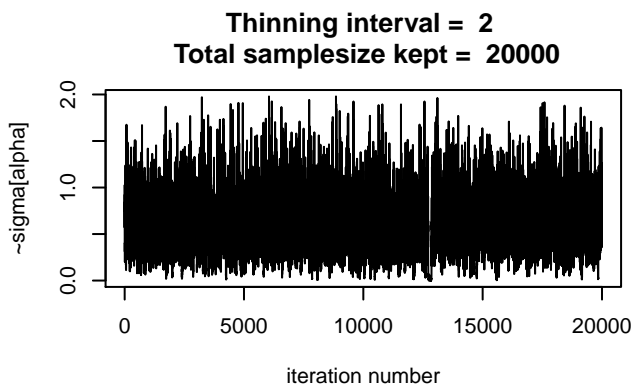
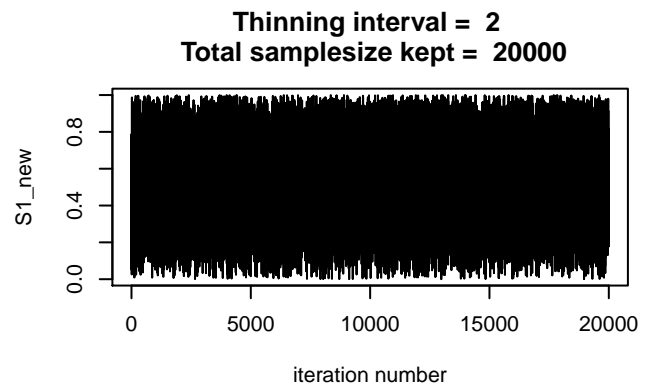
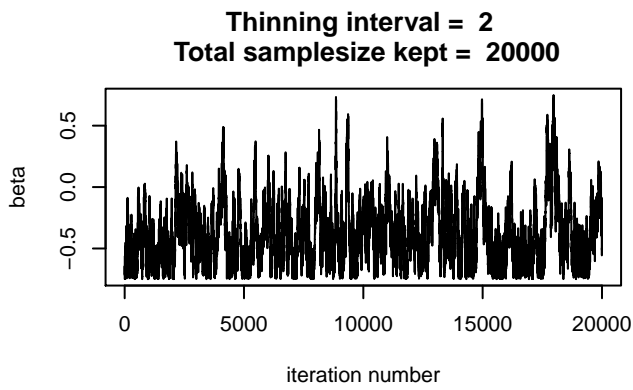
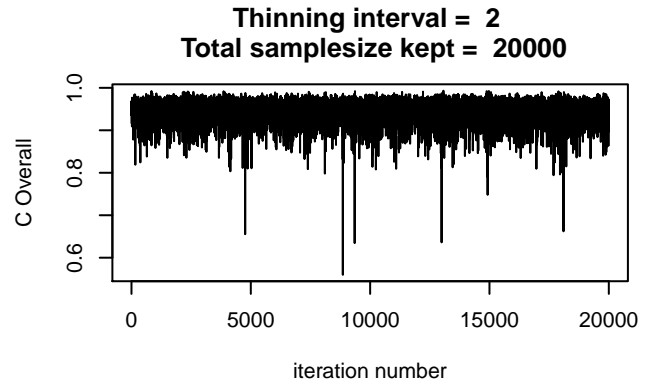
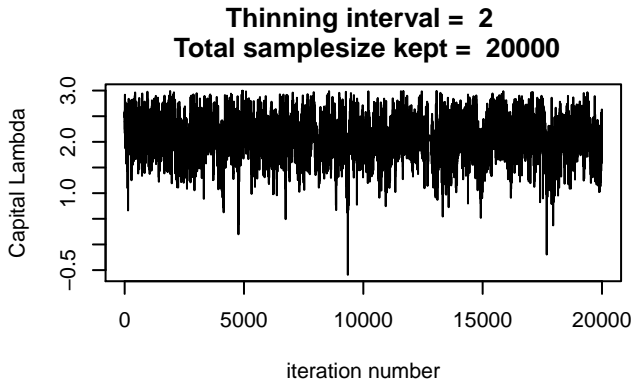
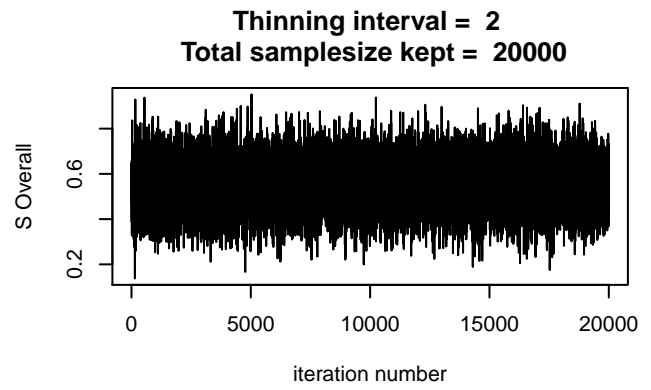
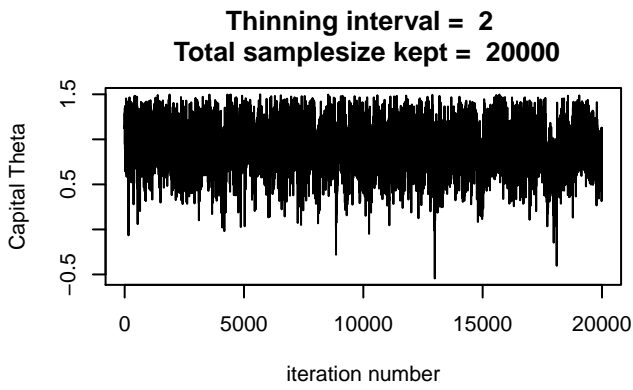


Thinning interval = 2
Total sample size kept = 20000

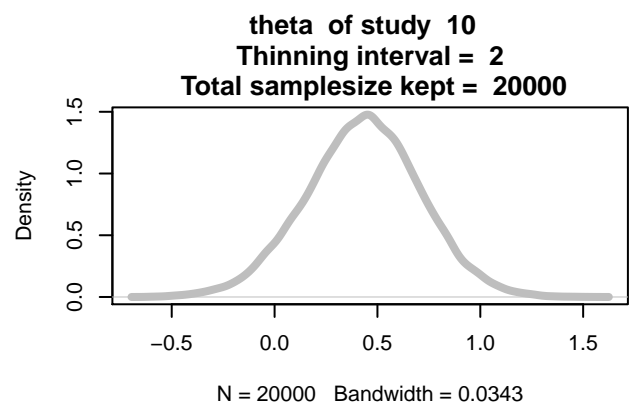
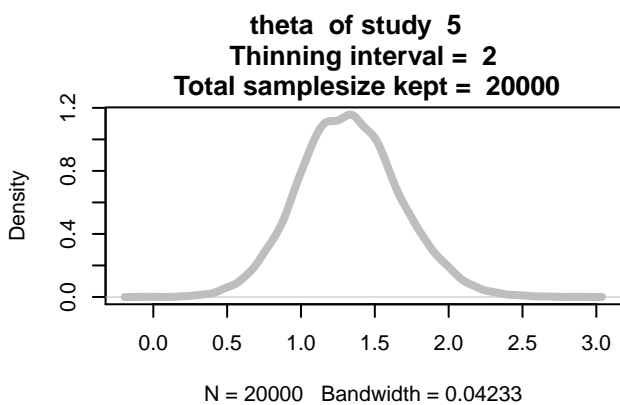
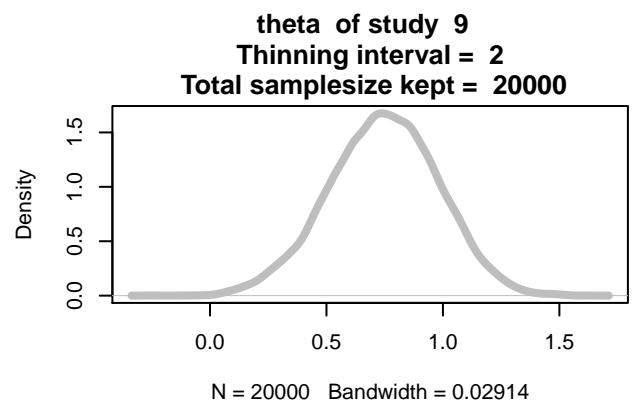
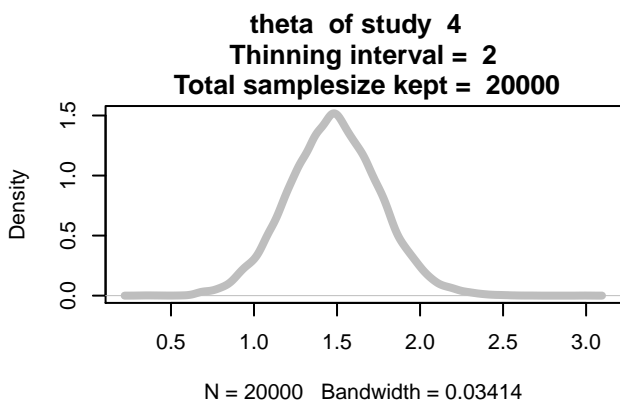
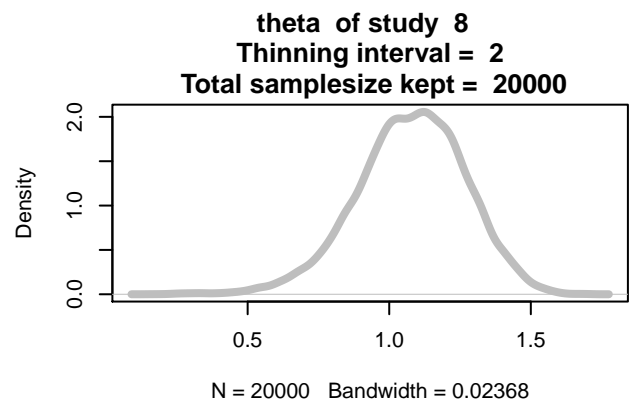
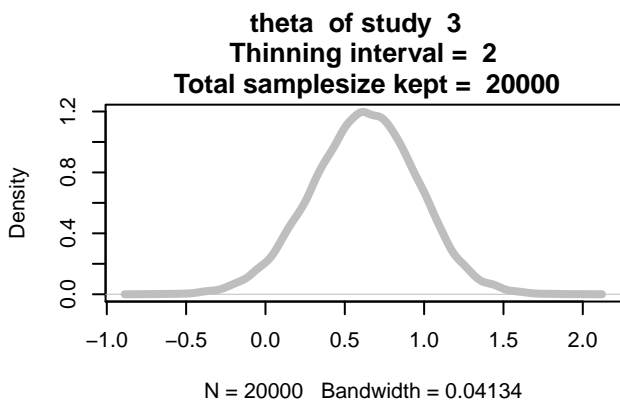
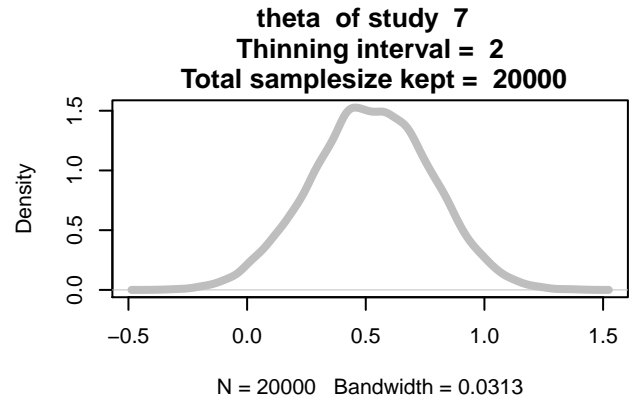
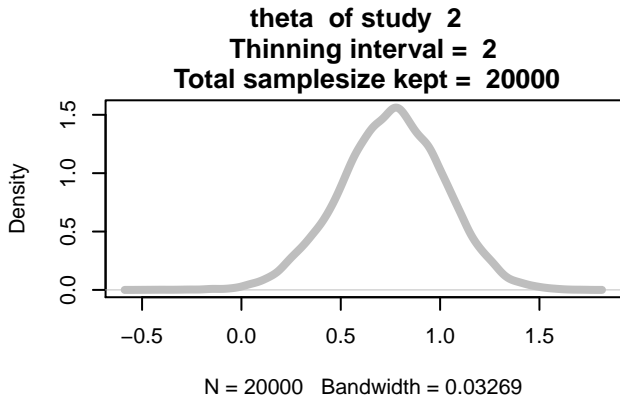
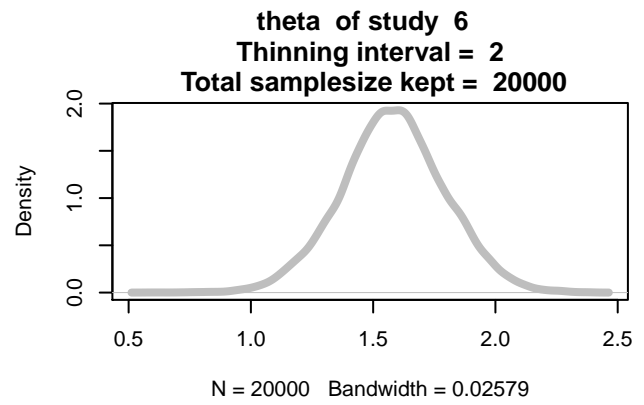
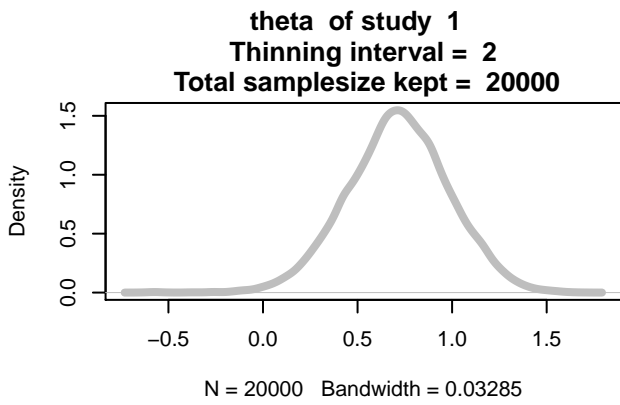


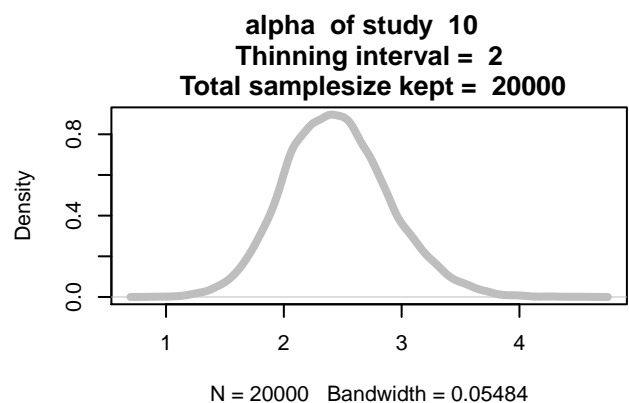
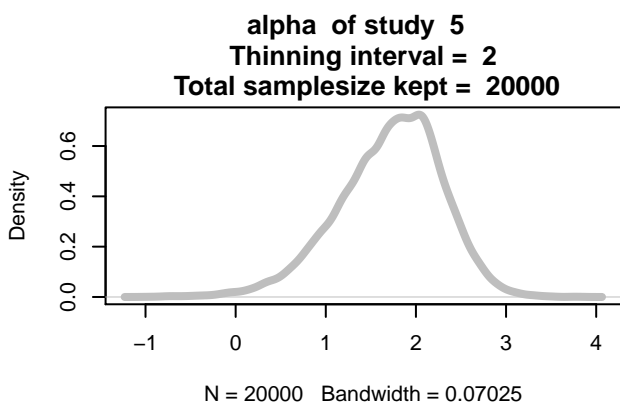
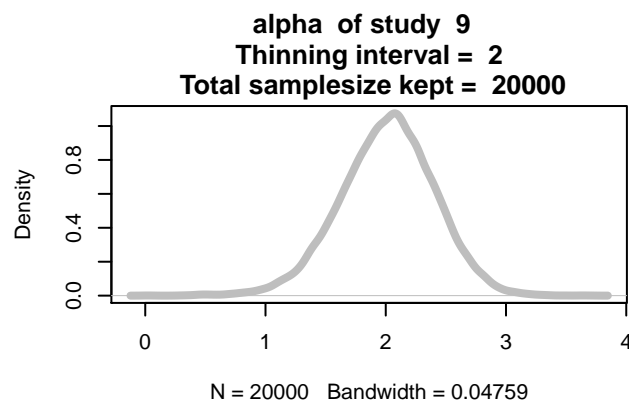
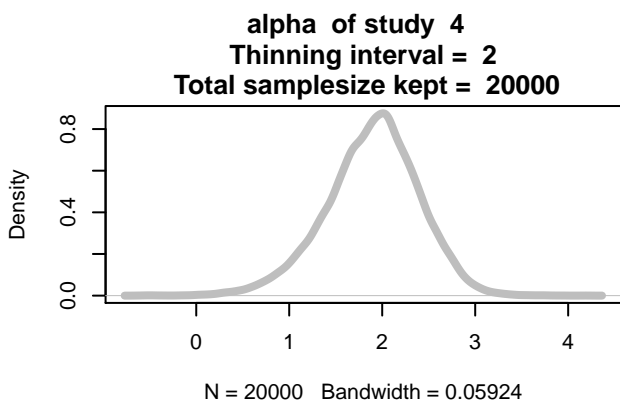
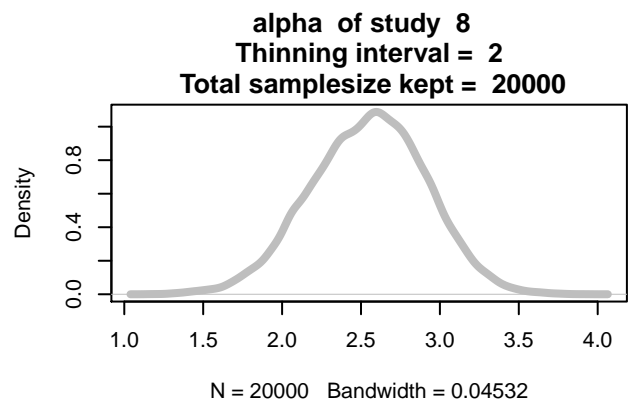
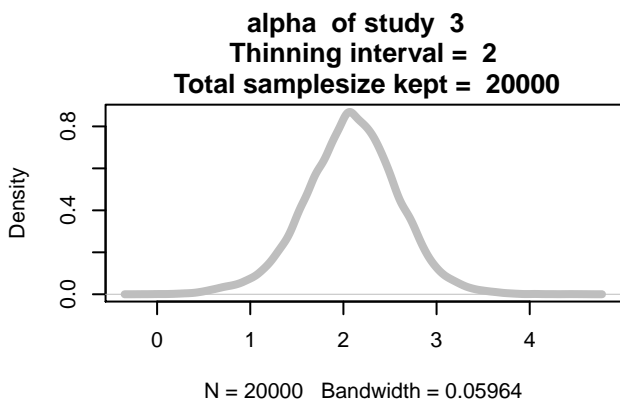
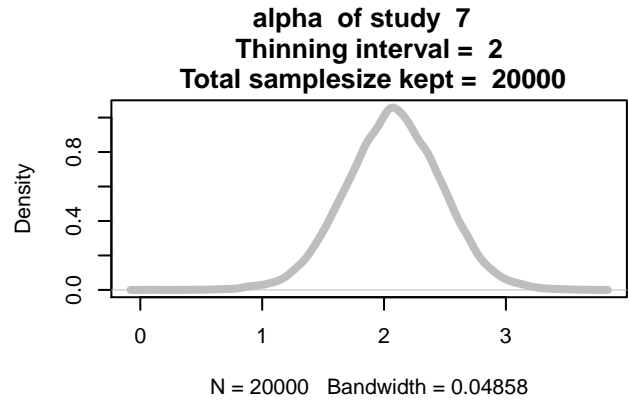
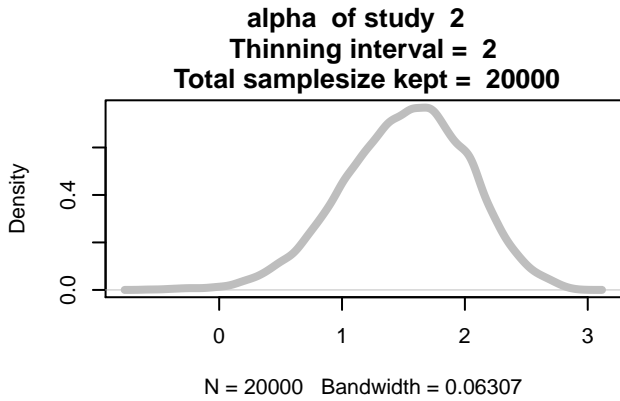
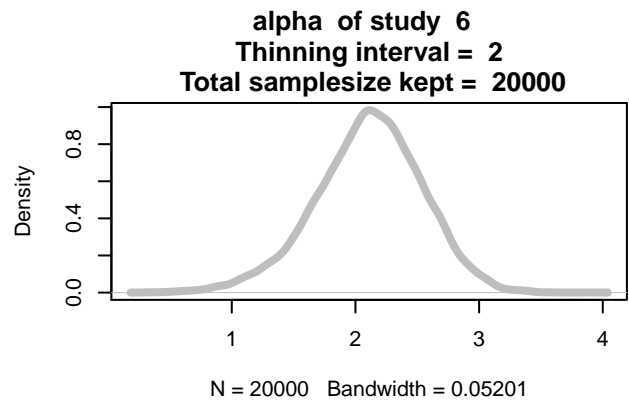
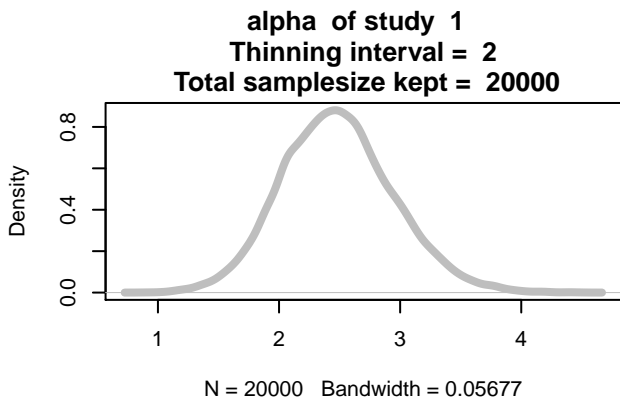
Thinning interval = 2
Total sample size kept = 20000



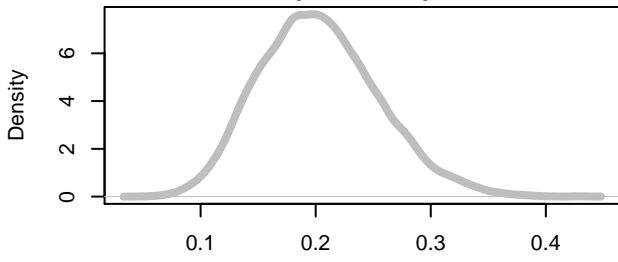


**A.3. Gráficos de densidad de los parámetros inter-estudio,
datos MRI**



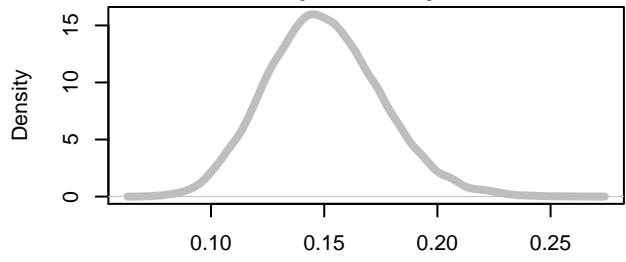


PI of study 1
Thinning interval = 2
Total sample size kept = 20000



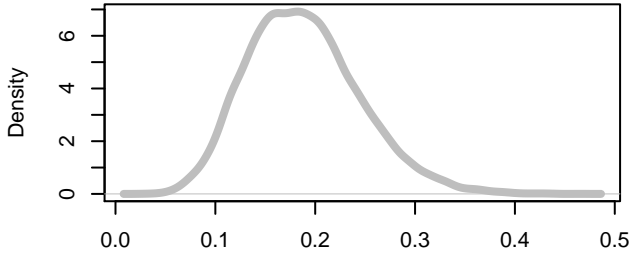
N = 20000 Bandwidth = 0.006409

PI of study 6
Thinning interval = 2
Total sample size kept = 20000



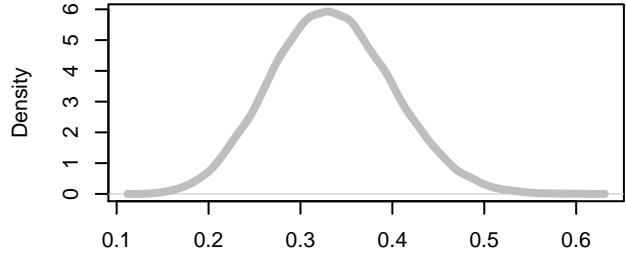
N = 20000 Bandwidth = 0.003123

PI of study 2
Thinning interval = 2
Total sample size kept = 20000



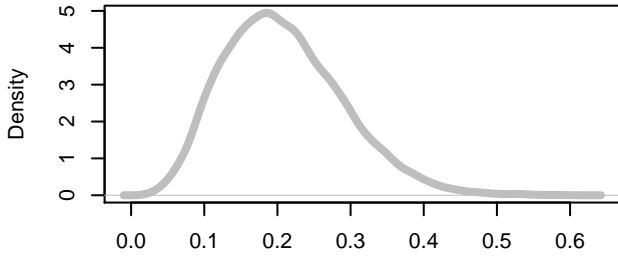
N = 20000 Bandwidth = 0.006968

PI of study 7
Thinning interval = 2
Total sample size kept = 20000



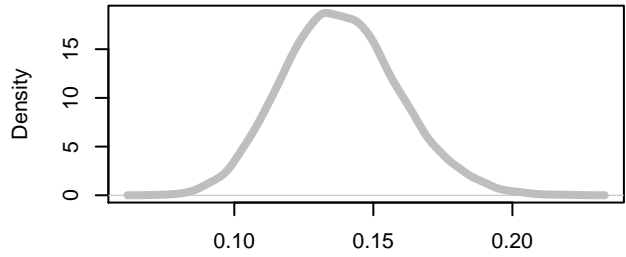
N = 20000 Bandwidth = 0.008169

PI of study 3
Thinning interval = 2
Total sample size kept = 20000



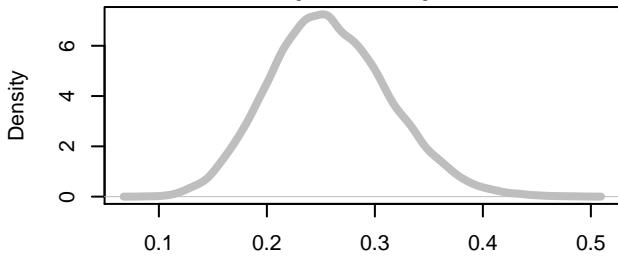
N = 20000 Bandwidth = 0.01009

PI of study 8
Thinning interval = 2
Total sample size kept = 20000



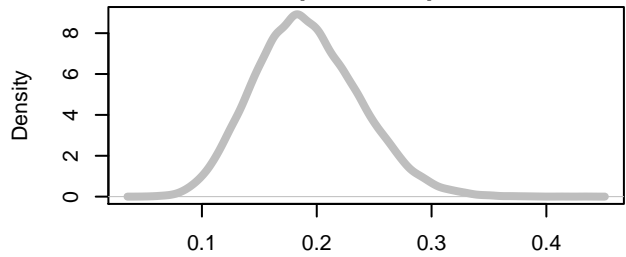
N = 20000 Bandwidth = 0.002605

PI of study 4
Thinning interval = 2
Total sample size kept = 20000



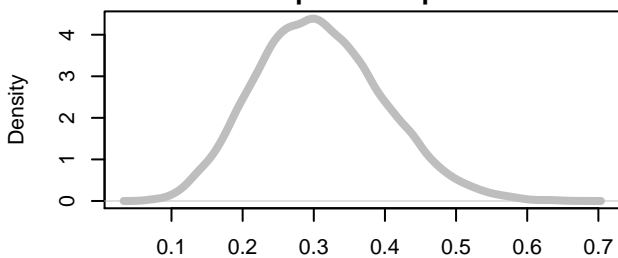
N = 20000 Bandwidth = 0.006873

PI of study 9
Thinning interval = 2
Total sample size kept = 20000



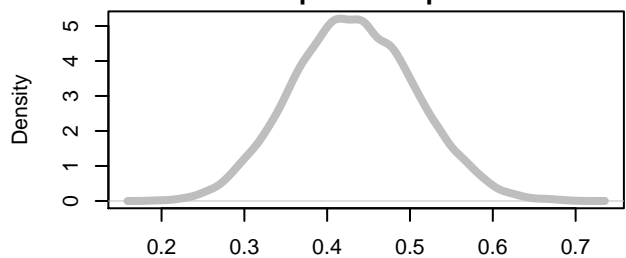
N = 20000 Bandwidth = 0.005672

PI of study 5
Thinning interval = 2
Total sample size kept = 20000

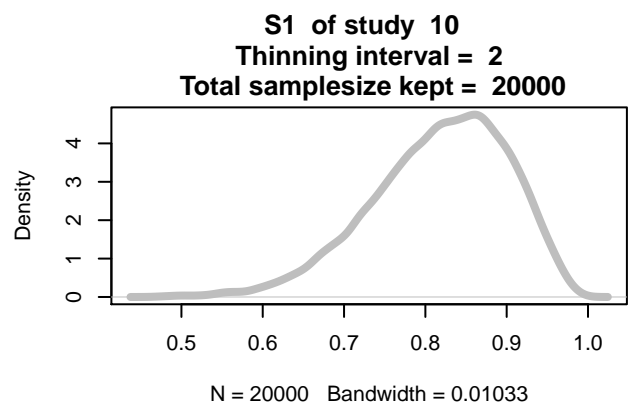
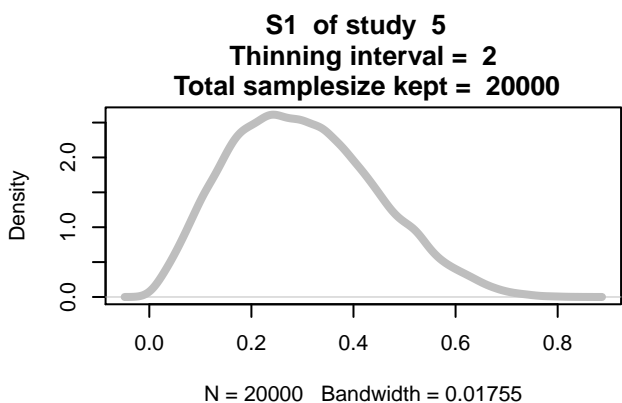
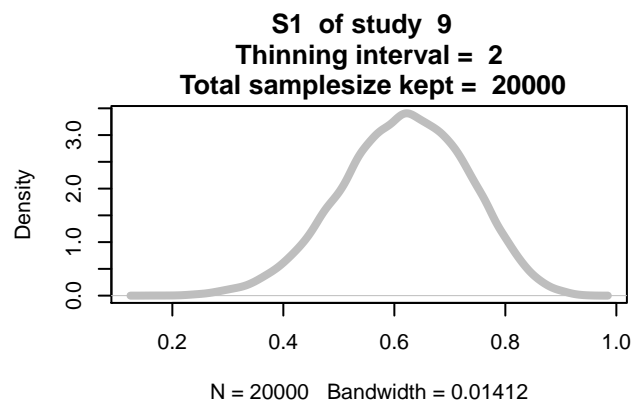
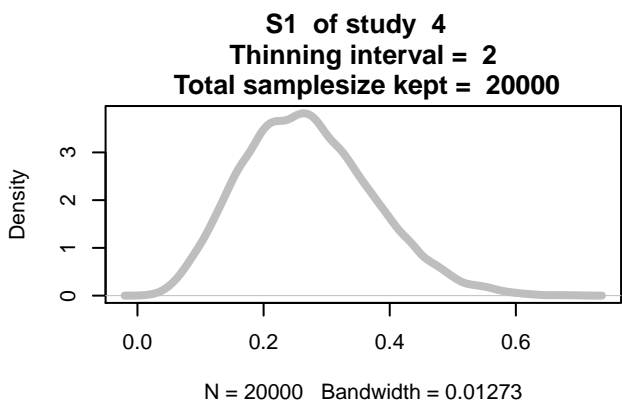
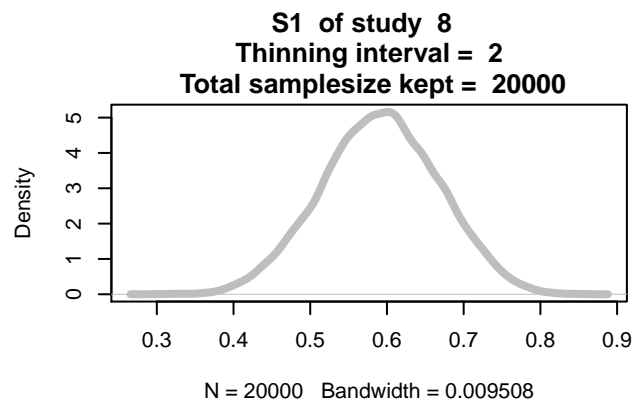
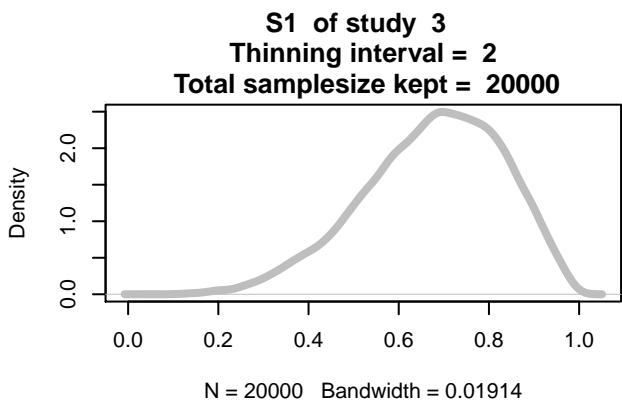
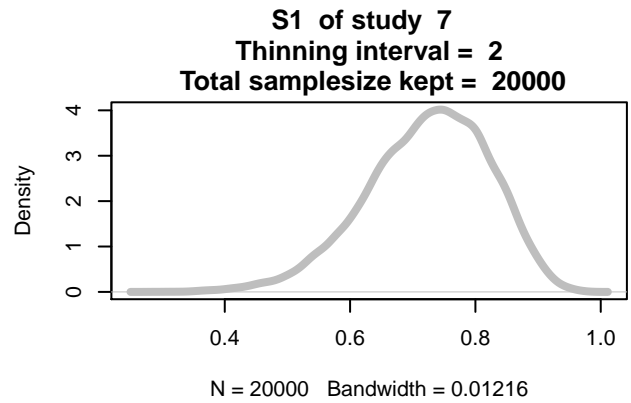
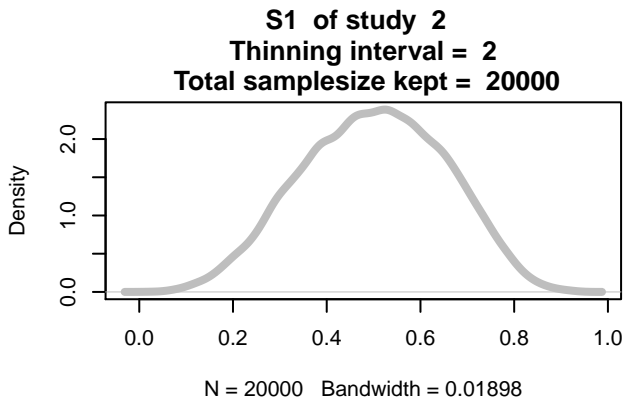
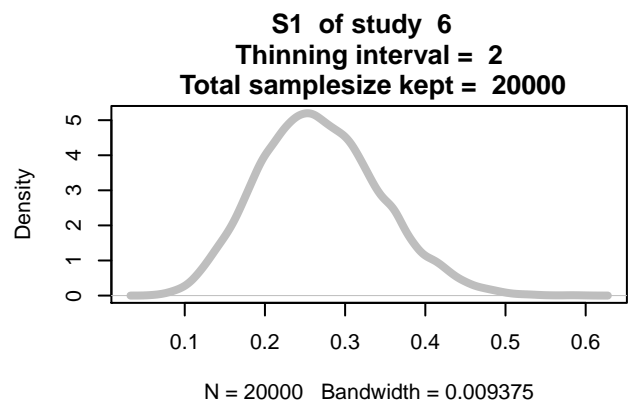
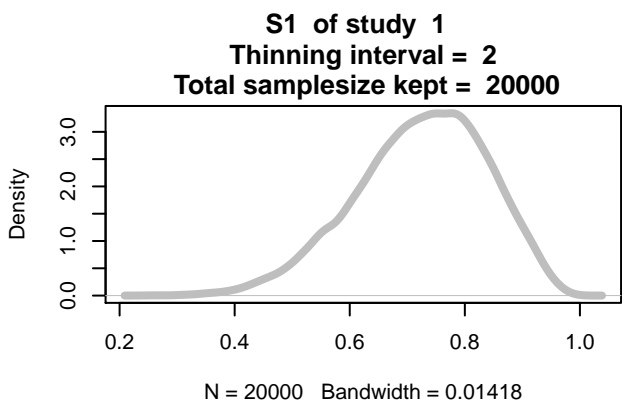


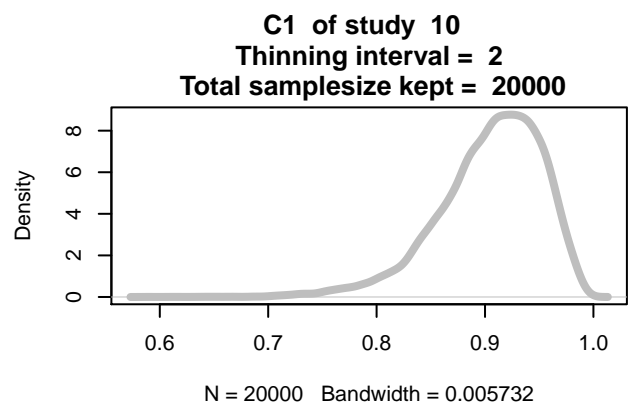
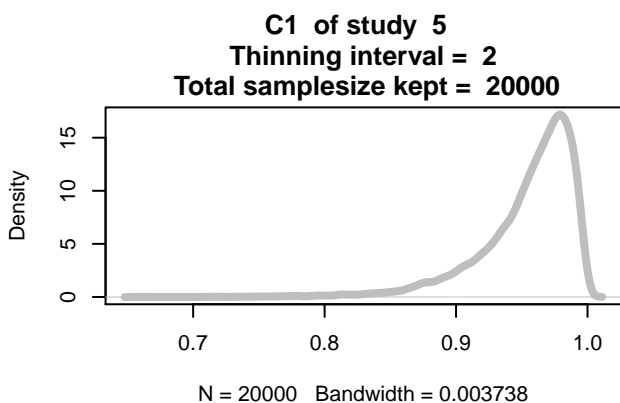
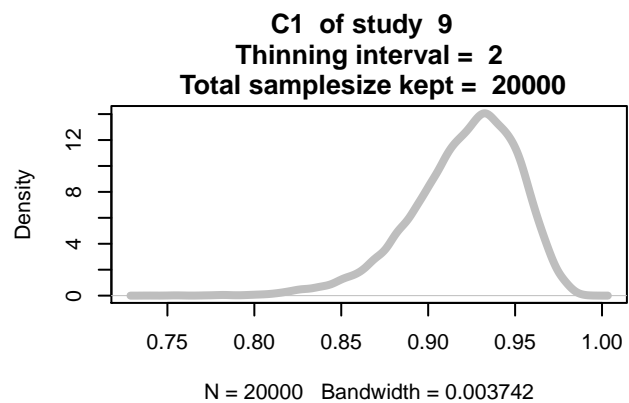
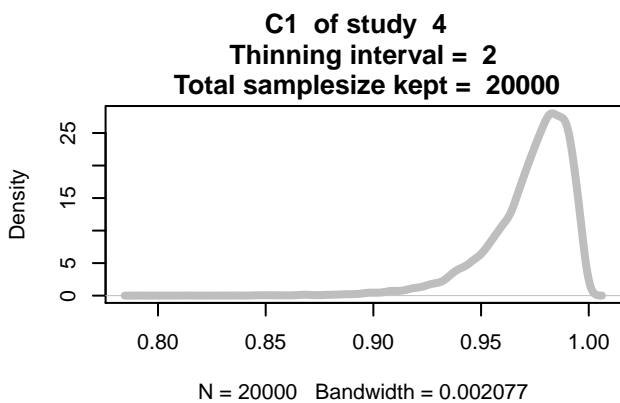
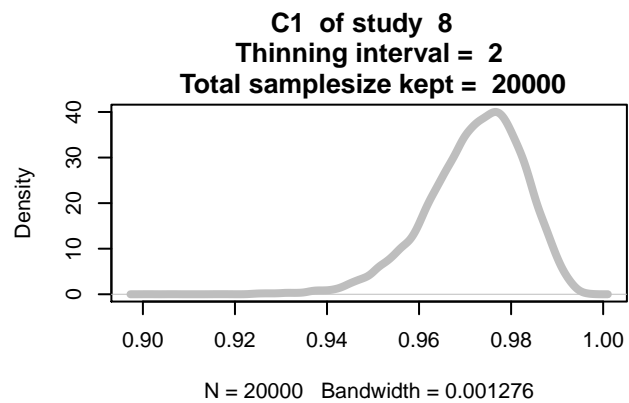
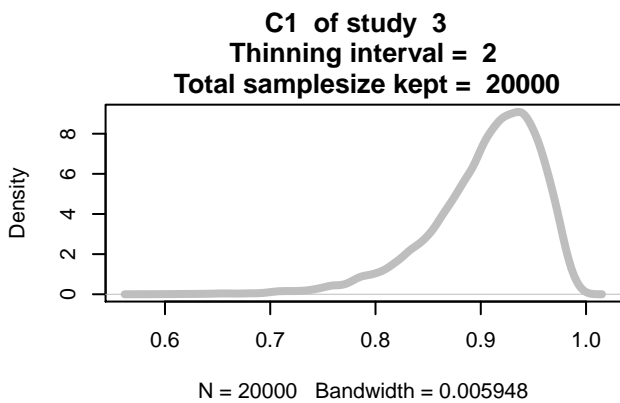
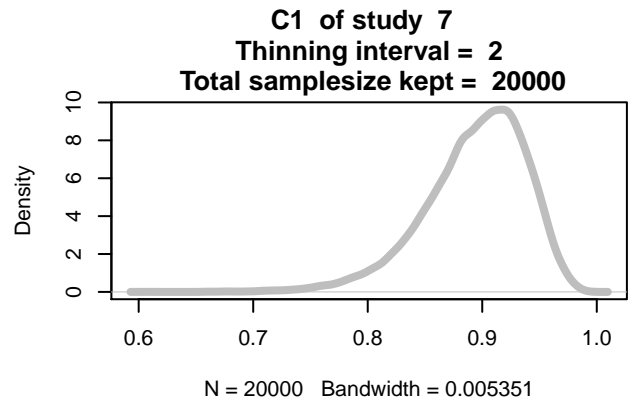
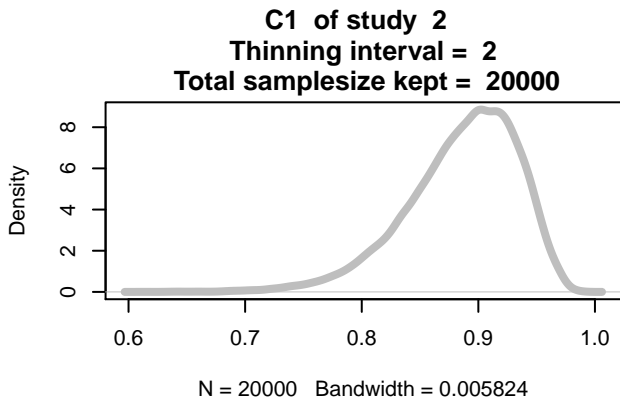
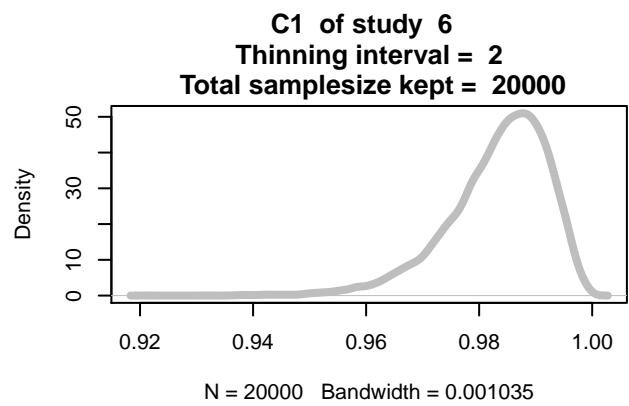
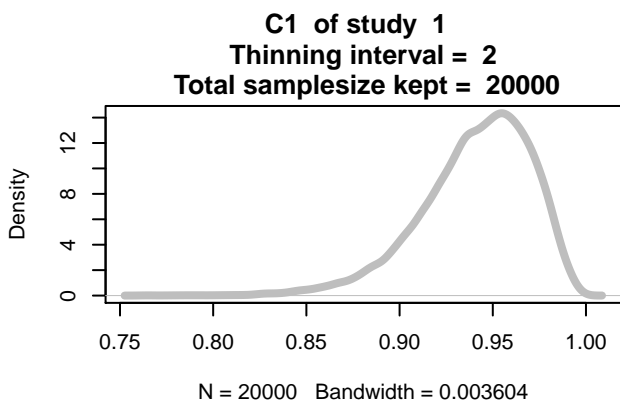
N = 20000 Bandwidth = 0.0111

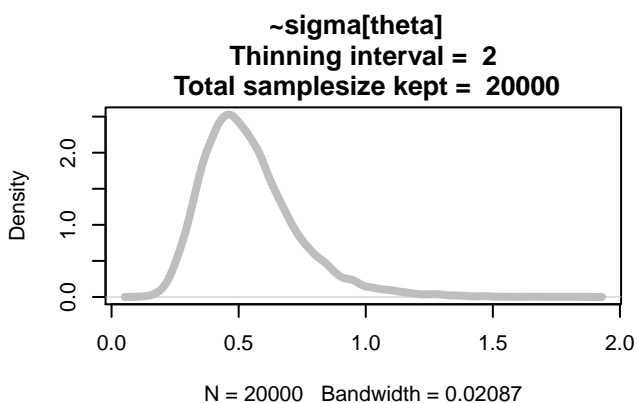
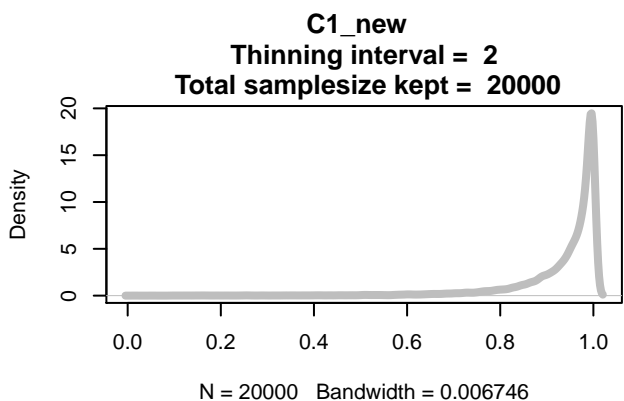
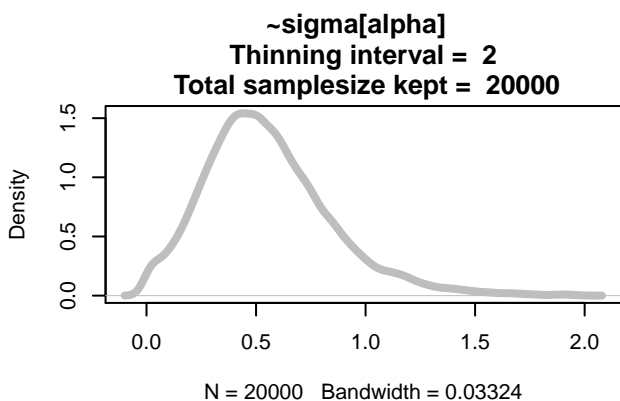
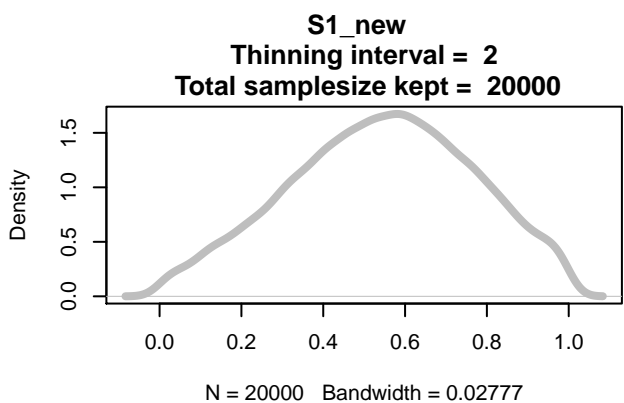
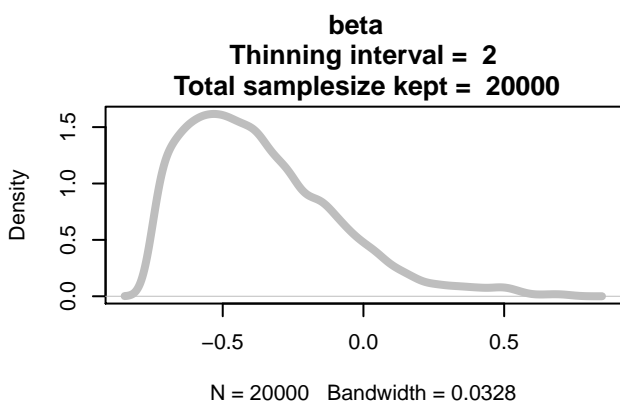
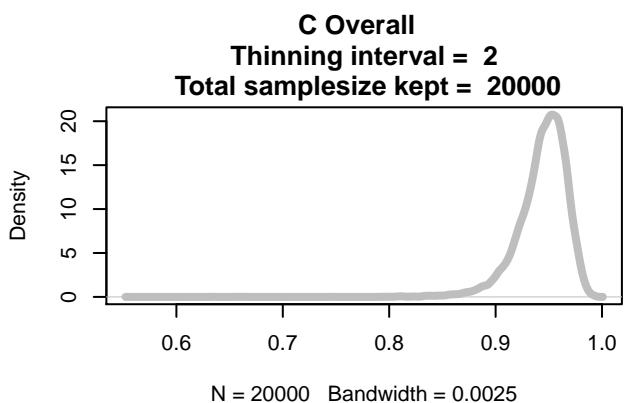
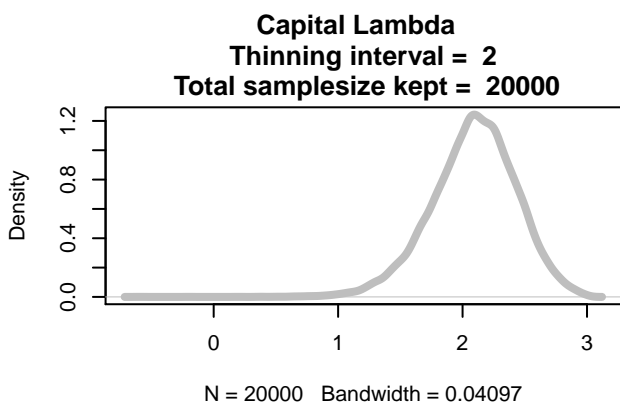
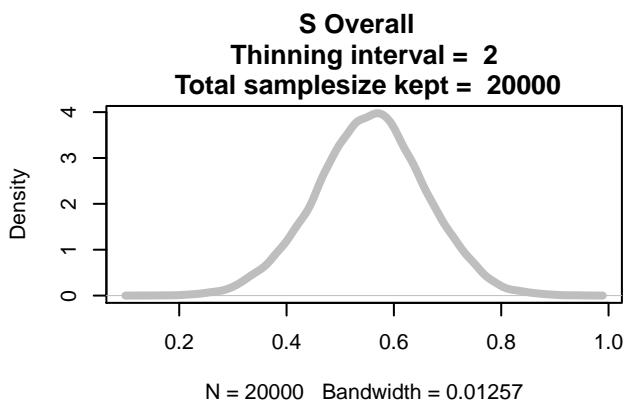
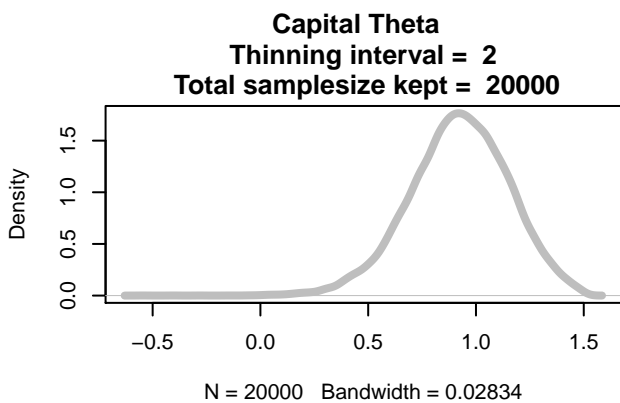
PI of study 10
Thinning interval = 2
Total sample size kept = 20000



N = 20000 Bandwidth = 0.009227







A.4. Código en SAS para el modelo bivariante, datos Nishimura et al. (136)

```
ODS TRACE ON;
/* Importar Datos */
proc import out=nishimura
datafile='E:/Trabajo/Universidad de Salamanca/Doctorado/SAS/petra/
Chapter 10 datasets and SAS files/StatRep-master/supplement/nishimura CCP.csv'
dbms=csv
replace;
getnames=yes;
run;

data nishimura_accp;
set nishimura;
where test='Anti-CCP';
run;

ods graphics on;

/* Crear dos registros por separados para los resultados verdaderos en cada estudio,
data nishimura_accp;
set nishimura_accp;
sens=1; spec=0; true=tp; n=tp+fn; output;
sens=0; spec=1; true=tn; n=tn+fp; output;
run;

/* Asegurarse que ambos registros de un estudio estén agrupados. */
proc sort data=nishimura_accp;
by study_id ;
run;
```

```

/* Ejecutar el modelo Bivariado sin covariables
La opción "cov" imprime una matriz de covarianza para todas las estimaciones de parámetros de

proc nlmixed data=nishimura_accp cov ecov ;

/* especificar valores iniciales para todos los parámetros que deben estimarse y garantizar que

parms msens=1 to 2 by 0.5 mspec=2 to 4 by 0.5 s2usens=0.2 s2uspec=0.6 covsesp=0;
bounds s2usens>=0;
bounds s2uspec>=0;
logitp = (msens + usens)*sens + (mspec + uspec)*spec;
p = exp(logitp)/(1+exp(logitp));
model true ~ binomial(n,p);

/* usens y uspec representan los efectos aleatorios. Se supone que ambos se distribuyen normal
random usens uspec ~ normal([0 , 0],[s2usens,covsesp,s2uspec])
subject=study_id out=randeffs;

/* Estimaciones adicionales que son funciones de los parámetros del modelo pueden estimarse: p
estimate 'logLR+' log((exp(msens)/(1+exp(msens)))/(1-(exp(mspec)/(1+exp(mspec)))));
estimate 'logLR-' log((1-(exp(msens)/(1+exp(msens))))/(exp(mspec)/(1+exp(mspec))));

run;

/* Compruebe la suposición de normalidad de los efectos aleatorios */
proc univariate data=randeffs plot normal;
class effect;
var estimate;
run;

/* Cree una variable ficticia para la generación de CCP, codificada como 0 para'CCP1' (la gener

```

```

data nishimura_accp;
set nishimura_accp;
ccpg=0;
if ccp_generation ="CCP2" then ccpg=1;
run;

/* añadir la covariable CCPG al modelo para permitir que tanto la sensibilidad como la

proc nlmixed data=nishimura_accp cov ecov;
parms msens=1 mspec=2 s2usens=0.2 s2uspec=0.6 covsesp=0 se2=0 sp2=0;
bounds s2usens>=0;
bounds s2uspec>=0;
logitp=(msens+usens+se2*ccpg)*sens+(mspec+uspec+sp2*ccpg)*spec;
p = exp(logitp)/(1+exp(logitp));
model true ~ binomial(n,p);
random usens uspec ~ normal([0 , 0], [s2usens,covsesp,s2uspec])
subject=study_id out=randeffs;

/* Estimar logit(sensibilidad) y logit(especificidad) para CCP2 (su correlación se pro

estimate 'logitsens CCP2' msens + se2;
estimate 'logitspec CCP2' mspec + sp2;
estimate 'logLR+ CCP1' log((exp(msens)/(1+exp(msens)))/(1-(exp(mspec)/(1+exp(mspec)))))
estimate 'logLR- CCP1' log((1-(exp(msens)/(1+exp(msens))))/(exp(mspec)/(1+exp(mspec))))
estimate 'logLR+ CCP2' log((exp(msens+se2)/(1+exp(msens+se2)))/(1-(exp(mspec+sp2)/(1+ex
estimate 'logLR- CCP2' log((1-(exp(msens+se2)/(1+exp(msens+se2))))/(exp(mspec+sp2)/(1+e
run;

/* comprobar la suposición de normalidad de los efectos aleatorios*/
proc univariate data=randeffs plot normal;
class effect;

```

A. CÓDIGO/MANUALES/PUBLICACIONES

```
var estimate;
```

```
run;
```

```
ods graphics off;
```

A.5. Código en SAS para el modelo HSROC, datos Nishimura et al. (136)

```
/* Importar datos */

proc import out=nishimura
datafile='E:/Trabajo/Universidad de Salamanca/Doctorado/SAS/petra/Chapter 10 datasets and SAS
dbms=csv
replace;
getnames=yes;
run;

/* seleccione sólo los estudios que han evaluado bajo la prueba RF */
data nishimura_RF;
set nishimura;
where test='RF';
run;

proc print;
run;

data nishimura_RF;
set nishimura_RF;

/* Crear registros separados para los grupos de personas enfermas y no enfermas en cada estudio */
dis=0.5; pos=tp; n=tp+fn; output;
dis=-0.5; pos=fp; n=tn+fp; output;
run;

/* Asegurarse de que ambos registros de un estudio estén agrupados. */
proc sort data=nishimura_RF;
by study_id dis;
```

A. CÓDIGO/MANUALES/PUBLICACIONES

```
run;
```

```
/* Ejecutar modelo HSR0C de Rutter y Gatsonis sin covariables. Solicitar matrices de covarianza para
```

```
proc nlmixed data=nishimura_RF ecov cov ;
```

```
/* establecer valores iniciales para todos los parámetros del modelo a estimar */
```

```
parms alpha=2 theta=0 beta=0 s2ua=0 s2ut=0 ;
```

```
logitp = (theta + ut + (alpha + ua)*dis) * exp(-(beta)*dis);
```

```
p = exp(logitp)/(1+exp(logitp));
```

```
model pos ~ binomial(n,p);
```

```
/* se supone que los efectos aleatorios para la exactitud (ua) y el umbral (ut) están aproximadamente
```

```
random ut ua ~ normal([0,0],[s2ut,0,s2ua]) subject=study_id out=randeffs;
```

```
run;
```

```
/* Cree dos variables ficticias para permitir los tres métodos de medición de RF. LA es el método de
```

```
data nishimura_RF;
```

```
set nishimura_RF;
```

```
if method_of_measurement ne "ELISA" and method_of_measurement ne "Nephelometry" and  
method_of_measurement ne "LA" then delete;
```

```
rfm1=0; rfm2=0; ;
```

```
if method_of_measurement = "ELISA" then rfm1=1;
```

```
if method_of_measurement = "Nephelometry" then rfm2=1;
```

```
run;
```



```
/* Asegurarse de que ambos registros de un estudio estén agrupados. */
proc sort data=nishimura_RF;
by study_id dis;
run;

/* incluir covariables para permitir que la precisión, el umbral y la forma varíen según el estudio */

proc nlmixed data=nishimura_RF ecov cov ;

parms alpha=2 theta=0 beta=0 s2ua=1 s2ut=1 a1=0 a2=0 t1=0 t2=0 b1=0 b2=0 ;

logitp = (theta + ut + t1*rfm1 + t2*rfm2 + (alpha + ua + a1*rfm1 + a2*rfm2)*dis)*
exp(-(beta + b1*rfm1 + b2*rfm2)*dis);

p = exp(logitp)/(1+exp(logitp));

model pos ~ binomial(n,p);

random ut ua ~ normal([0,0],[s2ut,0,s2ua]) subject=study_id out=randeffs;

/* estimaciones de parámetros para los métodos de medición de RF; */
estimate 'alpha ELISA' alpha + a1;
estimate 'theta ELISA' theta + t1;
estimate 'beta ELISA' beta + b1;
estimate 'alpha Nephelometry' alpha + a2;
estimate 'theta Nephelometry' theta + t2;
estimate 'beta Nephelometry' beta + b2;
run;

/* simplificar el modelo para asumir que las tres curvas tienen la misma forma */
proc nlmixed data=nishimura_RF ecov cov ;
```

A. CÓDIGO/MANUALES/PUBLICACIONES

```
parms alpha=2 theta=0 beta=0 s2ua=1 s2ut=1 a1=0 a2=0 t1=0 t2=0 ;
```

```
logitp = (theta + ut + t1*rfm1 + t2*rfm2 + (alpha + ua + a1*rfm1 + a2*rfm2)*dis)*  
exp(-(beta)*dis);
```

```
p = exp(logitp)/(1+exp(logitp));
```

```
model pos ~ binomial(n,p);
```

```
random ut ua ~ normal([0,0],[s2ut,0,s2ua]) subject=study_id out=randeffs;
```

```
/* estimaciones de parámetros para los métodos de medición de RF; */
```

```
estimate 'alpha ELISA' alpha + a1;
```

```
estimate 'theta ELISA' theta + t1;
```

```
estimate 'alpha Nephelometry' alpha + a2;
```

```
estimate 'theta Nephelometry' theta + t2;
```

```
run;
```

```
/* comprobar la hipótesis de normalidad de los efectos aleatorios */
```

```
proc univariate data=randeffs plot normal;
```

```
class effect;
```

```
var estimate;
```

```
run;
```

```
/* este modelo asume que las tres curvas tienen la misma forma y posición. La posición es la misma p
```

```
proc nlmixed data=nishimura_RF ecov cov ;
```

```
parms alpha=2 theta=0 beta=0 s2ua=1 s2ut=1 t1=0 t2=0 ;
```

```
logitp = (theta + ut + t1*rfm1 + t2*rfm2 + (alpha + ua)*dis)*
```

```
exp(-(beta)*dis);

p = exp(logitp)/(1+exp(logitp));

model pos ~ binomial(n,p);

random ut ua ~ normal([0,0],[s2ut,0,s2ua]) subject=study_id out=randeffs;
run;

/* comprobar la hipótesis de normalidad de los efectos aleatorios */
proc univariate data=randeffs plot normal;
class effect;
var estimate;

run;
```

A.6. Código en SAS comparación de pruebas, datos Schuetz et al. (167)

```
proc import out=schuetz
datafile='E:\Trabajo\Universidad de Salamanca\Doctorado\SAS\petra\Chapter 10 datasets and SAS files\S
dbms=csv
replace;
getnames=yes;
run;
```

/* Cree dos registros separados para los resultados verdaderos en cada estudio, el primero para el g
La variable sens es un indicador que toma el valor 1 si true=true positivo y 0 si no, la variable s

```
data schuetz;
set schuetz;
testtype=0;
if test ="CT" then testtype=1;
sens=1; spec=0; true=tp; n=tp+fn; output;
sens=0; spec=1; true=tn; n=tn+fp; output;
run;
```

```
/* Asegurarse de que ambos registros de un estudio estén agrupados. */
proc sort data=schuetz;
by study_id test;
run;
```

```
/* Ejecutar el modelo Bivariante sin covariables
La opción "cov" solicita que se imprima una matriz de covarianza para todas las estimaciones de pará
```

```
proc nlmixed data=schuetz cov ecov;

parms msens=2 mspec=1 s2usens=0 s2uspec=0 covsesp=0 ;

logitp=(msens+usens)*sens+(mspec+uspec)*spec;

p = exp(logitp)/(1+exp(logitp));

model true ~ binomial(n,p);

random usens uspec ~ normal([0,0],[s2usens,covsesp,s2uspec]) subject=study_id out=randeffs;

run;

/* Modelo bivalente con prueba como covariable utilizando el tipo de prueba como variable 1

proc nlmixed data=schuetz cov ecov;

parms msens=2 mspec=1 s2usens=0 s2uspec=0 covsesp=0 se_CT=1 sp_CT=0;

logitp=(msens+usens+se_CT*testtype)*sens+(mspec+uspec+sp_CT*testtype)*spec;

p = exp(logitp)/(1+exp(logitp));

model true ~ binomial(n,p);

random usens uspec ~ normal([0,0],[s2usens,covsesp,s2uspec]) subject=study_id out=randeffs;

/* Estimación de logit(sensibilidad), y logit(especificidad) */
estimate 'logitsens CT' msens + se_CT;
estimate 'logitspec CT' mspec + sp_CT;
```

```
run;

/* Compruebe la suposición de normalidad de los efectos aleatorios */
proc univariate data=randeffs plot normal;
class effect;
var estimate;
run;

/* Modelo bivariante con efecto del tipo de prueba sólo en la sensibilidad */

proc nlmixed data=schuetz cov ecov;

parms msens=2 mspec=1 s2usens=0 s2uspec=0 covsesp=0 se_CT=1 ;

logitp=(msens+usens+se_CT*testtype)*sens+(mspec+uspec)*spec;

p = exp(logitp)/(1+exp(logitp));

model true ~ binomial(n,p);

random usens uspec ~ normal([0,0],[s2usens,covsesp,s2uspec]) subject=study_id out=randeffs;

run;

/*Modelo bivariante con efecto del tipo de ensayo en la especificidad */

proc nlmixed data=schuetz cov ecov;

parms msens=2 mspec=1 s2usens=0 s2uspec=0 covsesp=0 sp_CT=0;

logitp=(msens+usens)*sens+(mspec+uspec+sp_CT*testtype)*spec;
```

```
p = exp(logitp)/(1+exp(logitp));

model true ~ binomial(n,p);

random usens uspec ~ normal([0,0],[s2usens,covsesp,s2uspec]) subject=study_id out=randeffs;

run;

/* COMPARACIONES DIRECTAS */

/*Crear un nuevo conjunto de datos de estudios con una comparación dentro del estudio de TC
"indirecta" es una variable binaria en el conjunto de datos codificada 1 si el estudio evalu

data schuetz_direct;
set schuetz;
where indirect=0;
run;

/* Ajuste del modelo bivariante sin covariable */
proc nlmixed data=schuetz_direct cov ecov qpoints=10;

parms msens=2 mspec=1 s2usens=0 s2uspec=0 covsesp=0 ;

bounds s2usens>=0;
bounds s2uspec>=0;

logitp=(msens+usens)*sens+(mspec+uspec)*spec;

p = exp(logitp)/(1+exp(logitp));

model true ~ binomial(n,p);
```

```
random usens uspec ~ normal([0,0],[s2usens,covsesp,s2uspec]) subject=study_id out=randeffs;

run;

/* Ajuste del modelo bivariante con efectos fijos */
proc nlmixed data=schuetz_direct cov ecov qpoints=10;

parms msens=2 mspec=1;

logitp=(msens)*sens+(mspec)*spec;

p = exp(logitp)/(1+exp(logitp));

model true ~ binomial(n,p);

run;

/* Ajuste del modelo bivariante con efecto aleatorio sólo para especificidad */

proc nlmixed data=schuetz_direct cov ecov qpoints=10;

parms msens=2 mspec=1 s2uspec=0;

logitp=(msens)*sens+(mspec+uspec)*spec;

p = exp(logitp)/(1+exp(logitp));

model true ~ binomial(n,p);
random uspec ~ normal([0],[s2uspec]) subject=study_id out=randeffs;

run;

/* Ajuste del modelo bivariante con efecto aleatorio sólo para la sensibilidad */
```



```
proc nlmixed data=schuetz_direct cov ecov qpoints=10;

parms msens=2 mspec=1 s2usens=0;

logitp=(msens+usens)*sens+(mspec)*spec;

p = exp(logitp)/(1+exp(logitp));

model true ~ binomial(n,p);
random usens ~ normal([0],[s2usens]) subject=study_id out=randeffs;

run;

/* Ajuste el modelo bivariante con covariante para el tipo de prueba tanto en sens como en
proc nlmixed data=schuetz_direct cov ecov qpoints=10;

parms msens=2 mspec=1 s2uspec=0 se_CT=0 sp_CT=0;

bounds s2uspec>=0;

logitp=(msens+se_CT*testtype)*sens+(mspec+uspec+sp_CT*testtype)*spec;

p = exp(logitp)/(1+exp(logitp));

model true ~ binomial(n,p);

random uspec ~ normal([0],[s2uspec]) subject=study_id out=randeffs;

/* Estimación de logit(sensibilidad) y logit(especificidad) */
estimate 'logitsens CT' msens + se_CT;
estimate 'logitspec CT' mspec + sp_CT;
```

```
run;
```

```
/* Ajustar el modelo bivariante con covariable para el tipo de ensayo sobre la especificidad. Efectos aleatorios sólo para la especificidad */  
proc nlmixed data=schuetz_direct cov ecov qpoints=10;
```

```
parms msens=2 mspec=1 s2uspec=0 sp_CT=0;
```

```
bounds s2uspec>=0;
```

```
logitp=(msens)*sens+(mspec+uspec+sp_CT*testtype)*spec;
```

```
p = exp(logitp)/(1+exp(logitp));
```

```
model true ~ binomial(n,p);
```

```
random uspec ~ normal([0],[s2uspec]) subject=study_id out=randeffs;
```

```
run;
```

```
/* Ajustar el modelo bivariante con covariable para el tipo de prueba de sensibilidad. Efectos aleatorios sólo para la especificidad */  
proc nlmixed data=schuetz_direct cov ecov qpoints=10;
```

```
parms msens=2 mspec=1 s2uspec=0 se_CT=0;
```

```
bounds s2uspec>=0;
```

```
logitp=(msens+se_CT*testtype)*sens+(mspec+uspec)*spec;
```

```
p = exp(logitp)/(1+exp(logitp));
```

```
model true ~ binomial(n,p);
```

```
random uspec ~ normal([0],[s2uspec]) subject=study_id out=randeffs;
```

```
run;
```

A.7. Código en R, generación de tablas tetracóricas, con patrón oro

```
> #Simula número de estudios (Meta-análisis),
> #una distribución uniforme entre (5-35)
> N<-trunc(runif(1,min=5,max=35))
> #N
> #Simula número de individuos en cada estudio (Meta-análisis),
> #una distribución uniforme entre (200-2000)
> #n<-trunc(runif(1,min=1000,max=4000))
> n <- seq(200,2000,1)
> #n
> #Simula N baja prevalencia
> #Según OMS frecuencia menor o igual
> #Un caso por cada 10,000 personas
> pi<-runif(N,min=0.01,max=0.35)
> #pi
>
> beta<-runif(1,min=-0.75,max=0.75)
> #beta = 0.25
> #LAMBDA = 2.5
> LAMBDA<-runif(1,min=-0.75,max=0.75)
> #sd_alpha = 0.75
> sd_alpha<-runif(1,min=0,max=2)
> #THETA = -0.2
> THETA<-runif(1,min=-1.5,max=1.5)
> #sd_theta = 0.50
> sd_theta<-runif(1,min=0,max=2)
> #REFSTD = list(2, 1:trunc(N/2),trunc(N/2)+1:N)
> #se1=runif(1,min=0.213,max=1)
> #se2=runif(1,min=0.213,max=1)
> #esp1=runif(1,min=0.818,max=1)
```

```
> #esp2=runif(1,min=0.818,max=1)
> #s2 = c(se1, se1)
> #c2 = c(esp1, esp1)
> sim.data <-simdata(N, n,n.random = TRUE,prev=pi,
T=THETA, L=LAMBDA,sd_t=sd_theta, sd_a=sd_alpha, b=beta)
> #sim.data$Data
```

A.8. Código en R, generación de tablas tetracóricas, 1 punto de corte

```
> #Simula número de estudios (Meta-análisis),
> #una distribución uniforme entre (5-35)
> N<-trunc(runif(1,min=5,max=35))
> #N
> #Simula número de individuos en cada estudio (Meta-análisis),
> #una distribución uniforme entre (200-2000)
> #n<-trunc(runif(1,min=1000,max=4000))
> n <- seq(200,2000,1)
> #n
> #Simula N baja prevalencia
> #Según OMS frecuencia menor o igual
> #Un caso por cada 10,000 personas
> pi<-runif(N,min=0.01,max=0.35)
> #pi
>
> beta<-runif(1,min=-0.75,max=0.75)
> #beta = 0.25
> #LAMBDA = 2.5
> LAMBDA<-runif(1,min=-0.75,max=0.75)
> #sd_alpha = 0.75
> sd_alpha<-runif(1,min=0,max=2)
> #THETA = -0.2
> THETA<-runif(1,min=-1.5,max=1.5)
> #sd_theta = 0.50
> sd_theta<-runif(1,min=0,max=2)
> REFSTD = list(2, 1:trunc(N/2),trunc(N/2)+1:N)
> se1=runif(1,min=0.213,max=1)
> se2=runif(1,min=0.213,max=1)
> esp1=runif(1,min=0.818,max=1)
```

```
> esp2=runif(1,min=0.818,max=1)
> s2 = c(se1, se1)
> c2 = c(esp1, esp1)
> sim.data <-simdata(N, n,n.random = TRUE,
sub_rs=REFSTD, se_ref=s2, sp_ref=c2,prev=pi,
T=THETA, L=LAMBDA,sd_t=sd_theta, sd_a=sd_alpha, b=beta)
> #sim.data$Data
```

A.9. Código en R, generación de tablas tetracóricas, 2 puntos de corte

```
> N<-trunc(runif(1,min=5,max=35))
> #N
> #Simula número de individuos en cada estudio (Meta-análisis),
> #una distribución uniforme entre (200-2 000)
> #n<-trunc(runif(1,min=1000,max=4000))
> n <- seq(200,2000,1)
> #n
> #Simula N baja prevalencia
> #Según OMS frecuencia menor o igual
> #Un caso por cada 10,000 personas
> pi<-runif(N,min=0.01,max=0.35)
> #pi
>
> beta<-runif(1,min=-0.75,max=0.75)
> #beta = 0.25
> #LAMBDA = 2.5
> LAMBDA<-runif(1,min=-0.75,max=0.75)
> #sd_alpha = 0.75
> sd_alpha<-runif(1,min=0,max=2)
> #THETA = -0.2
> THETA<-runif(1,min=-1.5,max=1.5)
> #sd_theta = 0.50
> sd_theta<-runif(1,min=0,max=2)
> REFSTD = list(2, 1:trunc(N/2),trunc(N/2)+1:N)
> se1=runif(1,min=0.213,max=1)
> se2=runif(1,min=0.213,max=1)
> while(se1 == se2){
+   se2=runif(1,min=0.213,max=1)
+ }
```



```
> esp1=runif(1,min=0.818,max=1)
> esp2=runif(1,min=0.818,max=1)
> while(esp1 == esp2){
+   esp2=runif(1,min=0.818,max=1)
+ }
> s2 = c(se1, se2)
> c2 = c(esp1, esp2)
> sim.data <-simdata(N, n,n.random = TRUE,
sub_rs=REFSTD, se_ref=s2, sp_ref=c2,prev=pi,
T=THETA, L=LAMBDA,sd_t=sd_theta, sd_a=sd_alpha, b=beta)
> #sim.data$Data
```

A.10. Código en R, generación de tablas tetracóricas, 3 puntos de corte

```
> #Simula número de estudios (Meta-análisis),
> #una distribución uniforme entre (5-35)
>
> N<-trunc(runif(1,min=5,max=35))
> #N
> #Simula número de individuos en cada estudio (Meta-análisis),
>
> n <- seq(200,2000,1)
> #n
> #Simula N baja prevalencia
> #Según OMS frecuencia menor o igual
> #Un caso por cada 10,000 personas
> pi<-runif(N,min=0.01,max=0.35)
> #pi
>
> beta<-runif(1,min=-0.75,max=0.75)
> #beta = 0.25
> #LAMBDA = 2.5
> LAMBDA<-runif(1,min=-0.75,max=0.75)
> #sd_alpha = 0.75
> sd_alpha<-runif(1,min=0,max=2)
> #THETA = -0.2
> THETA<-runif(1,min=-1.5,max=1.5)
> #sd_theta = 0.50
> sd_theta<-runif(1,min=0,max=2)
> REFSTD = list(3, 1:trunc(N/3),trunc(N/3)+1:trunc(N*2/3), trunc(N*2/3)+1:N)
> se1=runif(1,min=0.213,max=1)
> se2=runif(1,min=0.213,max=1)
> se3=runif(1,min=0.213,max=1)
```

```
> while((se1 == se2)& (se1 == se3)&(se2 == se3)){
  + se2=runif(1,min=0.213,max=1)
  + se3=runif(1,min=0.213,max=1)
  + }
> esp1=runif(1,min=0.818,max=1)
> esp2=runif(1,min=0.818,max=1)
> esp3=runif(1,min=0.818,max=1)
> while((esp1 == esp2)& (esp1 == esp3)&(esp2 == esp3)){
  + esp2=runif(1,min=0.818,max=1)
  + esp3=runif(1,min=0.818,max=1)
  + }
> s2 = c(se1, se2,se3)
> c2 = c(esp1, esp2,esp3)
> sim.data <-simdata(N, n,n.random = TRUE,
sub_rs=REFSTD, se_ref=s2, sp_ref=c2,prev=pi,
T=THETA, L=LAMBDA,sd_t=sd_theta, sd_a=sd_alpha, b=beta)}
> #sim.data$Data
```

A.11. Código en R, ajustando modelos estadísticos a las tablas tetracóricas simuladas

```
> nq=15
> gl=gauss.quad.prob(nq,"uniform")
> mgrid<- meshgrid(gl$n,gl$n)
> datos<-read.csv("matriz_final_todo.csv")
> #View(datos)
> fila=1
> wb = createWorkbook()
> #wb1 = createWorkbook()
> sheet = createSheet(wb, "Sheet 1")
> datos1<-datos[,2:6]
> #i<-1
> #attach(datos1)
> #for (i in 1:1600){
  > flag<-filter(datos1,Id_meta==1)
  > #View(flag)
  > flag[,1]<-NULL
  > #str(flag)
  > fit.DOR.MH <- madauni(flag, method = "MH")
  > #fit.DOR.MH$coefficients
  >
  >
  > fit.DOR.DSL <- madauni(flag,method = "DSL")
  > tau_2<-as.data.frame(fit.DOR.DSL$tau_sq)
  > Q_temp<-as.data.frame(fit.DOR.DSL$CQ)
  > Q<-(Q_temp[1,]-nrow(flag)+1)/Q_temp[1,]
  > fit.phm.homo <- phm(flag, hetero = FALSE)
  > AUC_homo<-1/(fit.phm.homo$coefficients+1)
  > fit.phm.het <- phm(flag)
  > temp<-as.data.frame(fit.phm.het$coefficients)
```

A.11 Código en R, ajustando modelos estadísticos a las tablas tetracóricas simuladas

```
> AUC_hete<-1/(temp[1,]+1)
> tau_phm<-temp[2,]
> fit.reitsma <- reitsma(flag)
> #summary(list(fit.reitsma))
> #(fit.reitsma)
> #summary(fit.reitsma)
> test<-summary(fit.reitsma)
> std.dev<-as.data.frame(sqrt(diag(fit.reitsma$Psi)))
> #test$vcov
> corFixed<-as.data.frame(test$corFixed)
> #corFixed[2,1]
> corRandom<-as.data.frame(test$corRandom)
> #corRandom[2,1]
> AUC_reitsma<-as.data.frame(test$AUC)
> #AUC_reitsma[1,1]
> LR<-(SummaryPts(fit.reitsma))
> hsroc<-as.data.frame(test$coef_hsroc)
> #hsroc[1,2]
> Q<-as.data.frame(Q)
> addDataFrame(exp(fit.DOR.MH$coefficients),
sheet=sheet,startRow = fila,startColumn=1, col.names = FALSE)
> addDataFrame(fit.DOR.MH$descr$cor_sens_fpr,
sheet=sheet,startRow = fila, startColumn=3, col.names = FALSE)
> addDataFrame(exp(fit.DOR.DSL$coefficients),
sheet=sheet,startRow = fila, startColumn=5, col.names = FALSE)
> addDataFrame(tau_2[1,1],
sheet=sheet,startRow = fila,startColumn=8, col.names = FALSE)
> addDataFrame(fit.DOR.DSL$descr$cor_sens_fpr,
sheet=sheet,startRow = fila,startColumn=10, col.names = FALSE)
> addDataFrame(Q[1,1],sheet=sheet,startRow = fila,
startColumn=12, col.names = FALSE)
> addDataFrame(AUC_homo,sheet=sheet,
startRow = fila,startColumn=14, col.names = FALSE)
```

```
> addDataFrame(AUC_hete, sheet=sheet,
startRow = fila, startColumn=16, col.names = FALSE)
> addDataFrame(tau_phm, sheet=sheet,
startRow = fila, startColumn=18, col.names = FALSE)
> addDataFrame(test$coefficients[1,1], sheet=sheet,
startRow = fila, startColumn=20, col.names = FALSE)
> addDataFrame(test$coefficients[2,1], sheet=sheet,
startRow = fila, startColumn=22, col.names = FALSE)
> addDataFrame(mean(fit.DOR.DSL$descr$sens$sens), sheet=sheet,
startRow = fila, startColumn=24, col.names = FALSE)
> addDataFrame(mean(fit.DOR.DSL$descr$spec$spec), sheet=sheet,
startRow = fila, startColumn=26, col.names = FALSE)
> addDataFrame(mean(fit.DOR.DSL$descr$fpr$fpr), sheet=sheet,
startRow = fila, startColumn=28, col.names = FALSE)
> addDataFrame(fit.reitsma$converged, sheet=sheet,
startRow = fila, startColumn=30, col.names = FALSE)
> addDataFrame(AUC_reitsma[1,1], sheet=sheet,
startRow = fila, startColumn=32, col.names = FALSE)
> addDataFrame(AUC_reitsma[1,2], sheet=sheet,
startRow = fila, startColumn=34, col.names = FALSE)
> addDataFrame(test$coefficients[3,1], sheet=sheet,
startRow = fila, startColumn=36, col.names = FALSE)
> addDataFrame(test$coefficients[4,1], sheet=sheet,
startRow = fila, startColumn=38, col.names = FALSE)
> addDataFrame(test$coefficients[1,2], sheet=sheet,
startRow = fila, startColumn=40, col.names = FALSE)
> addDataFrame(test$coefficients[2,2], sheet=sheet,
startRow = fila, startColumn=42, col.names = FALSE)
> addDataFrame(corFixed[2,1], sheet=sheet,
startRow = fila, startColumn=44, col.names = FALSE)
> addDataFrame(corRandom[2,1], sheet=sheet,
startRow = fila, startColumn=46, col.names = FALSE)
> addDataFrame(mean(LR$posLR), sheet=sheet,
```

A.11 Código en R, ajustando modelos estadísticos a las tablas tetracóricas simuladas

```
startRow = fila, startColumn=48, col.names = FALSE)
> addDataFrame(mean(LR$negLR), sheet=sheet,
startRow = fila, startColumn=50, col.names = FALSE)
> addDataFrame(mean(LR$invnegLR), sheet=sheet,
startRow = fila, startColumn=52, col.names = FALSE)
> addDataFrame(mean(LR$DOR), sheet=sheet,
startRow = fila, startColumn=54, col.names = FALSE)
> addDataFrame(std.dev[1,1], sheet=sheet,
startRow = fila, startColumn=56, col.names = FALSE)
> addDataFrame(std.dev[2,1], sheet=sheet,
startRow = fila, startColumn=58, col.names = FALSE)
> addDataFrame(hsroc[1,1], sheet=sheet,
startRow = fila, startColumn=60, col.names = FALSE)
> addDataFrame(hsroc[1,2], sheet=sheet,
startRow = fila, startColumn=62, col.names = FALSE)
> addDataFrame(hsroc[1,3], sheet=sheet,
startRow = fila, startColumn=64, col.names = FALSE)
> addDataFrame(hsroc[1,4], sheet=sheet,
startRow = fila, startColumn=66, col.names = FALSE)
> addDataFrame(hsroc[1,5], sheet=sheet,
startRow = fila, startColumn=68, col.names = FALSE)
> #addDataFrame(est.b5$gradient[5,1], sheet=sheet,
startRow = fila, startColumn=162, col.names = FALSE)
>
>     if (AUC_reitsma>0.5){
+     addDataFrame(1, sheet=sheet,
startRow = fila, startColumn=70,
col.names = FALSE)
+     }else {
+     addDataFrame(0, sheet=sheet,
startRow = fila, startColumn=70, col.names = FALSE)
+     }
>     if (AUC_reitsma>AUC_hete){
```

```
+           addDataFrame(1, sheet=sheet,
startRow = fila, startColumn=72, col.names = FALSE)
+           }else {
+           addDataFrame(0, sheet=sheet,
startRow = fila, startColumn=72, col.names = FALSE)
+           }
>           if ((test$coefficients[3,1]> mean(fit.DOR.DSL$descr$sens$sens)) &
((1-test$coefficients[4,1])>
mean(fit.DOR.DSL$descr$spec$spec))){
+           addDataFrame(1,
sheet=sheet,startRow = fila, startColumn=74, col.names = FALSE)
+           }else {
+           addDataFrame(0, sheet=sheet,
startRow = fila, startColumn=74, col.names = FALSE)
+           }
>           if (corRandom[2,1]>fit.DOR.DSL$descr$cor_sens_fpr) {
+           addDataFrame(1, sheet=sheet,
startRow = fila, startColumn=76, col.names = FALSE)
+           }else {
+           addDataFrame(0, sheet=sheet,
startRow = fila, startColumn=76, col.names = FALSE)
+           }
>           try(CopulaREMADA.beta(flag$TP,flag$FN,flag$FP,flag$TN,
gl,mgrid,qcondbvn,tau2par.bvn),silent = TRUE)

>           success<-try(CopulaREMADA.beta(flag$TP,flag$FN,flag$FP,flag$TN,
gl,mgrid,qcondbvn,tau2par.bvn),silent = TRUE)
>           if (class(success)=="try-error") {
+           estimate<-c(NaN,NaN,NaN,NaN,NaN);
+           gradient<-c(NaN,NaN,NaN,NaN,NaN);
+           est.b1<-data.frame(estimate,gradient);
+           a<-est.b1$estimate[1];
+           b<-est.b1$estimate[2];
```

A.11 Código en R, ajustando modelos estadísticos a las tablas tetracóricas simuladas

```
+     c<-0;
+   } else {
+     #print("Negative number");
+     est.b1=CopulaREMADA.beta(flag$TP,flag$FN,
+ flag$FP,flag$TN,g1,mgrid,qcondbvn,tau2par.bvn);
+     # Si la condición no se cumple (FALSE);
+     est.b1$estimate<-as.data.frame(est.b1$estimate);
+     est.b1$gradient<-as.data.frame(est.b1$gradient);
+     a<-est.b1$estimate[1,1];
+     b<-est.b1$estimate[2,1];
+     c<-1
+   }
> if (class(success)=="try-error") {
+   addDataFrame( est.b1$estimate[1],
+ sheet=sheet,startRow = fila, startColumn=78,
+ col.names = FALSE)
+   addDataFrame(est.b1$estimate[2],
+ sheet=sheet,startRow = fila, startColumn=80,
+ col.names = FALSE)
+   addDataFrame(est.b1$estimate[3],
+ sheet=sheet,startRow = fila, startColumn=82,
+ col.names = FALSE)
+   addDataFrame(est.b1$estimate[4],
+ sheet=sheet,startRow = fila, startColumn=84,
+ col.names = FALSE)
+   addDataFrame(est.b1$estimate[5],
+ sheet=sheet,startRow = fila, startColumn=86,
+ col.names = FALSE)
+   addDataFrame(est.b1$gradient[1],
+ sheet=sheet,startRow = fila, startColumn=88,
+ col.names = FALSE)
+   addDataFrame(est.b1$gradient[2],
+ sheet=sheet,startRow = fila, startColumn=90,
```

```
col.names = FALSE)
+   addDataFrame(est.b1$gradient[3],
sheet=sheet,startRow = fila, startColumn=92,
col.names = FALSE)
+   addDataFrame(est.b1$gradient[4],
sheet=sheet,startRow = fila, startColumn=94,
col.names = FALSE)
+   addDataFrame(est.b1$gradient[5],
sheet=sheet,startRow = fila, startColumn=96,
col.names = FALSE)
+   } else {
+   addDataFrame( est.b1$estimate[1,1],
sheet=sheet,startRow = fila, startColumn=78,
col.names = FALSE);
+   addDataFrame(est.b1$estimate[2,1],
sheet=sheet,startRow = fila, startColumn=80,
col.names = FALSE);
+   addDataFrame(est.b1$estimate[3,1],
sheet=sheet,startRow = fila, startColumn=82,
col.names = FALSE);
+   addDataFrame(est.b1$estimate[4,1],
sheet=sheet,startRow = fila, startColumn=84,
col.names = FALSE);
+   addDataFrame(est.b1$estimate[5,1],
sheet=sheet,startRow = fila, startColumn=86,
col.names = FALSE);
+   addDataFrame(est.b1$gradient[1,1],
sheet=sheet,startRow = fila, startColumn=88,
col.names = FALSE);
+   addDataFrame(est.b1$gradient[2,1],
sheet=sheet,startRow = fila, startColumn=90,
col.names = FALSE);
+   addDataFrame(est.b1$gradient[3,1],
```

A.11 Código en R, ajustando modelos estadísticos a las tablas tetracóricas simuladas

```
sheet=sheet,startRow = fila, startColumn=92,
col.names = FALSE);
+   addDataFrame(est.b1$gradient[4,1],
sheet=sheet,startRow = fila, startColumn=94,
col.names = FALSE);
+   addDataFrame(est.b1$gradient[5,1],
sheet=sheet,startRow = fila, startColumn=96,
col.names = FALSE);
+   }
>   if ((test$coefficients[3,1]> a) &
((1-test$coefficients[4,1])> b) &(c==1)){
+       addDataFrame(1, sheet=sheet,
startRow = fila, startColumn=98, col.names = FALSE)
+   }else {
+       addDataFrame(0, sheet=sheet,
startRow = fila, startColumn=98, col.names = FALSE)
+   }
```

A.12. Código en R, para etiquetar modelos a los meta-análisis simulados

```
> datos<-read.csv("resumen_bivariate_mh_dsl.csv")
> datos1<-datos
> x <- which(datos1$DOR_mayor ==1)
> y <- which(datos1$DOR_mayor ==2)
> datos1<-mutate(datos1, modelo =0)
> datos1<-as.data.frame(datos1)
> datos1$modelo[c(x)]<-1
> datos1$modelo[c(y)]<-2
> z <- which(datos1$modelo== 2 & ((datos1$corRandom_reitsma <0) | (abs(datos1$corRand
> datos1$modelo[c(z)]<-4
> zz<-which(datos1$modelo== 1 & (datos1$Corr_DSL<0))
> datos1$modelo[c(zz)]<-4
> flag1 <- which(datos1$modelo== 2 &
+ (datos1$sens_reitsma<datos1$sens_copula) & ((1-datos1$fpr_reitsma) < datos1$spec_
> datos1$modelo[c(flag1)]<-4
> lineas_model2 <- which(datos1$modelo ==2)
> (cuantiles<-quantile
(datos1$sigmaAB_hsroc[c(lineas_model2)], prob = c(0.25, 0.5, 0.75)))

> #summary(datos1$sigmaAB_hsroc[c(lineas_model2)])
> lineas_bivariate<-which((datos1$sigmaAB_hsroc<cuantiles[2]) & datos1$modelo ==2)
> datos1$modelo[c(lineas_bivariate)]<-3
> (length(which(datos1$modelo== 1)))

> (length(which(datos1$modelo== 2)))

> (length(which(datos1$modelo== 3)))

> (length(which(datos1$modelo== 4)))

> (length((datos1$modelo)))
```

```
> lineas_model3 <- which(datos1$modelo ==2)
> (cuantiles1<-quantile (datos1$DOR_reitsma[c(lineas_model3)],
prob = c(0.25, 0.5, 0.75)))

> summary(datos1$DOR_reitsma[c(lineas_model3)])

> lineas_model4<-which((datos1$DOR_reitsma >
cuantiles1[3]) & datos1$modelo ==2)
> (length(datos1$modelo[c(lineas_model4)]))

> (length(datos1$modelo[c(lineas_model3)]))

> (length(datos1$DOR_reitsma[c(lineas_model4)]))

> View(datos1)
> length(datos1$modelo[c(lineas_model4)])

> datos1$modelo[c(lineas_model4)]<-4
> (length(which(datos1$modelo== 1)))

> (length(which(datos1$modelo== 2)))

> (length(which(datos1$modelo== 3)))

> (length(which(datos1$modelo== 4)))

> (length((datos1$modelo)))

> (length(which(datos1$modelo== 1)))

> (length(which(datos1$modelo== 2)))

> (length(which(datos1$modelo== 3)))

> (length(which(datos1$modelo== 4)))
```

```

> modelo1<-(which(datos1$modelo== 1))
> modelo2<-(which(datos1$modelo== 2))
> modelo3<-(which(datos1$modelo== 3))
> modelo4<-(which(datos1$modelo== 4))
> datos1$modelo[c(modelo1)]<-4
> datos1$modelo[c(modelo2)]<-1
> datos1$modelo[c(modelo3)]<-2
> datos1$modelo[c(modelo4)]<-3
> (length(which(datos1$modelo== 1)))

> (length(which(datos1$modelo== 2)))

> (length(which(datos1$modelo== 3)))

> (length(which(datos1$modelo== 4)))

> cuantiles5<-
+ (c<-datos1$modelo)
> flag <- which(datos1$modelo== 2 & (datos1$corRandom_reitsma <0))
> (length(datos1$modelo[c(flag)]))

> View(datos1)
> flag1 <- which(datos1$modelo== 2 &
+ (datos1$sens_reitsma<datos1$sens_copula) &
+ ((1-datos1$fpr_reitsma) < datos1$spec_copula))
> (length(datos1$modelo[c(flag1)]))

```

Apéndice B

Artículos científicos publicados, Estancia
de Investigación y Asistencia a Eventos
Académicos Internacionales

**B.1. An Alternative To The Cochran-(Q) Statistic For
Analysis Of Heterogeneity In Meta-Analysis Of Diag-
nostic Tets Based On HJ Biplot**

AN ALTERNATIVE TO THE COCHRAN-(Q) STATISTIC FOR ANALYSIS OF HETEROGENEITY IN META-ANALYSIS OF DIAGNOSTIC TESTS BASED ON HJ BIPLLOT

Johny J. Pambabay-Calero^{*.a}, Sergio A. Bauz-Olvera^{**b}, Ana B. Nieto-Librero^{***, ****c}, Ma. Purificación Galindo-Villardón^{***, ****d}, Sergio Hernández-González^{*****e}

*Escuela Superior Politécnica del Litoral, Facultad de Ciencias Naturales y Matemáticas, Guayaquil, Ecuador

**Escuela Superior Politécnica del Litoral, Facultad de Ciencias de la Vida, Guayaquil, Ecuador

***Dpto. de Estadística, Facultad de Medicina, Universidad de Salamanca, Salamanca, España

****Instituto de Investigación Biomédica (IBSAL), Salamanca, España.

*****Universidad Veracruzana Xalapa, Av. Xalapa esq. Manuel Ávila Camacho s/n, Xalapa, Veracruz, México

ABSTRACT

The possible heterogeneity among individual studies constituting a meta-analysis is traditionally evaluated using Cochran Q and Higgins I^2 statistics. However, both indices have deficiencies: The Q statistic detects heterogeneity but does not allow its quantification, whereas the I^2 index allows for quantification of heterogeneity but does not indicate which studies are responsible for it. This problem is solved by additionally using the HJ biplot of the matrix containing the information about true positives, true negatives, false positives, and false negatives for each study. This means that the information contained in such a tetrachoric table contains the joint frequency distribution of the true classification of the disease and that provided by the diagnostic test.

MSC: 62H30

KEYWORDS: Meta-analysis, effect size, heterogeneity, Q-statistic, I^2 index, HJ biplot

RESUMEN

La forma tradicional de evaluar la posible heterogeneidad de los estudios individuales que conforman un meta-análisis, es utilizando los estadísticos Q de Cochran e I^2 de Higgins. Sin embargo, ambos índices presentan deficiencias: el Q detecta heterogeneidad, pero no permite cuantificarla y el I^2 permite cuantificarla, pero no conocer cuáles son los estudios responsables de la heterogeneidad. Este problema se resuelve utilizando, además, el HJ-Biplot de la matriz que contiene la información sobre VP, VN, FP y FN, para todos y cada uno de los estudios; es decir la información contenida en la tabla tetracórica que contiene la distribución de frecuencias conjunta de la verdadera clasificación de la enfermedad y la proporcionada por el test diagnóstico.

PALABRAS CLAVES: Meta-análisis, tamaño del efecto, heterogeneidad, estadístico Q, índice I^2 , HJ biplot

1. INTRODUCTION

In a meta-analysis, the results of each study are quantified by means of an index called effect size (for example, standardized mean difference, correlation coefficient, and odds ratio, etc.) that can be applied to all studies. This allows the use of the same metric for the analysis of the results of the studies.

In general, a meta-analysis has three main objectives: (a) to test whether the results of the studies are homogeneous; (b) to obtain an overall index of the magnitude of the effect of the relationship studied, with their respective confidence intervals and their statistical significance; and (c) to analyze whether a significant heterogeneity exists among studies and to identify possible variables or moderating characteristics of the results obtained.

The purpose of this study is to compare the performance of the Q test and the I^2 index, using the HJ biplot for the analysis of heterogeneity, using data from the Audit questionnaire ([Alcohol Use Disorder Identification Test](#)). Because of the length of this questionnaire, the **Audit C** data (a short alcohol detection test, see [15]) will be used to answer only the first three questions of the AUDIT, as shown in Table 1. The table contains 14 rows and five columns: the first column identifies the respective studies with their authors and year, and the remaining

^a E-mail: jpambaba@espol.edu.ec

^b E-mail: serabauz@espol.edu.ec

^c E-mail: ananieto@usal.es

^d E-mail: pgalindo@usal.es

^e E-mail: sehernandez@uv.mx

columns show the values of true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN).

Table 1.- Datos Audit C

Estudio - Año	TP	FN	FP	TN
Aalto et al., 2006	47	9	101	738
Aertgeerts et al., 2001	126	51	272	1543
Aertgeerts et al., 2002	19	10	12	192
Bradley et al., 2003	36	3	78	276
Bradley et al., 2007	130	19	211	959
Bush et al., 1998	84	2	68	89
Gómez et al., 2006	68	0	112	423
Gordon et al., 2001	752	0	3226	2977
Gual et al., 2002	59	5	55	136
Rumpf et al., 2002	142	50	571	2788
Seale et al., 2006	137	24	107	358
Selin, 2006	57	3	103	437
Tsai et al., 2005	34	1	21	56
Tuunanen et al., 2007	152	51	88	264

This process was conducted using a program in the R environment, whose code is attached in the Appendix.

2. MATERIALS AND METHODS

2.1. The Q-statistic and I² index

The evaluation of heterogeneity in a meta-analysis is a crucial issue because the presence or absence of true heterogeneity can affect the statistical model that the individuals performing the meta-analysis (the meta-analyst) decides to apply. Therefore, when the results of the studies only vary according to one sampling error (homogeneous case), a fixed-effects model can be applied to obtain a mean effect size. Conversely, if the study results differ by more than one sampling error (heterogeneous case), the meta-analyst can assume a random-effects model, taking into account intrastudy and interstudy variability. The most common method to assess whether there is actual heterogeneity in a meta-analysis is using the Q test, a statistical test defined by [2].

Under the hypothesis of homogeneity among effect sizes, the Q-statistic follows a chi-square distribution with $k - 1$ degrees of freedom, where k is the number of studies. Not rejecting the homogeneity hypothesis usually guides the meta-analyst to adopt a fixed-effects model, because it is assumed that the sizes of the estimated effect are only differentiated by the sampling error. Conversely, the rejection of the homogeneity hypothesis may lead to the use of a random-effects model that includes both intrastudy and interstudy variability. The deficiencies of the Q-statistic are its deficient power to spot true heterogeneity between studies when a meta-analysis includes a small number of studies and its excessive power to detect insignificant variability with a large number of studies [10], [18], and [3]; i.e., its power depends on the sample size. Another strategy to quantify true heterogeneity in a meta-analysis is the estimation of the variance between studies τ^2 , which represents the true value of heterogeneity among study effects produced by innumerable substantive factors (treatment type, patient characteristics), methodological (design type, sample size, wear, etc.), and study characteristics. Because the value of τ^2 depends on the particular effect of the metric used in a meta-analysis, comparing the estimated τ^2 values from meta-analyses that have used different effect size indices (i.e., standardized mean difference, correlation coefficient and odds ratio, etc.) is not possible.

To overcome deficiencies of the Q test and τ^2 , [12]; see also [13] have proposed the I² index to evaluate heterogeneity in a meta-analysis. This index can be interpreted as a percentage of heterogeneity—i.e., the fraction of variation that is because of interstudy variance τ^2 ; this is obtained by dividing the difference of Q and $(k - 1)$ for Q. Thus, the I² index is similar to an intraclass correlation in cluster sampling [12]. For example, a meta-analysis with I² = 0 means that all variability in effect size estimates is a result of a sampling error within

the studies. On the other hand, a meta-analysis with $I^2 = 50$ means that half of the total variability among effect sizes is not caused by a sampling error but is caused by true heterogeneity between the studies. [12] proposed a tentative classification of I^2 to help interpret its magnitude; therefore, the percentages 25% ($I^2 = 25$), 50% ($I^2 = 50$), and 75% ($I^2 = 75$) would mean a low, average, and high heterogeneity, respectively.

There is a direct correlation between the I^2 index and interstudy variance τ^2 —when τ^2 increases, I^2 will also increase, and vice versa. One of the advantages of the I^2 index with respect to τ^2 is stated by [12], who state that I^2 indices obtained from meta-analysis with varying numbers of studies and varying effect metrics are directly comparable. Along with this descriptive interpretation of the I^2 index, [12] proposed a confidence interval that could be used in the same way that the Q test is used for assessing heterogeneity in meta-analyses. Therefore, if the confidence interval regarding I^2 includes the value of 0%, then the meta-analyst can support the hypothesis of homogeneity. If, on the other hand, the confidence interval does not include the value of 0%, there is no evidence of the existence of true heterogeneity. The use of the I^2 index and its confidence interval is similar to the Q test application. Because the I^2 index assesses not only the heterogeneity in the meta-analysis but also the degree of heterogeneity, it should be a more advisable procedure than the Q test to assess whether true heterogeneity exists between studies in a meta-analysis.

2.2. Biplot Method

The biplot method is a low-dimensional graphical representation of a data matrix X (with I individuals and J variables), i.e., a multivariate matrix; for more detail see [7]. The biplot is based on the singular value decomposition in vectors—i.e., it tries to reproduce the data incorporating a simultaneous representation of individuals and variables. The two most important representations proposed by [7] were the GH and JK biplots. The former offers a high quality in the representation of the columns (variables), whereas the latter offers a high representation of the rows (individuals). [8] named this type the HJ biplot and demonstrated that it is possible to represent rows and columns simultaneously on the same low-dimensional space, with a high-quality representation for individuals and variables.

The biplot method is useful for the visual inspection of data matrices because it allows the collection of atypical patterns and values, which can be graphically represented in a manner analogous to principal component analysis and discriminant analysis. Because of these advantages of the biplot method, the HJ biplot will be used to analyze the heterogeneity of the Audit C data.

Measures such as factor contributions to the element [8] will be concepts employed when analyzing the results. The proximity between individuals (rows) is interpreted as similarity; the angles formed by vector variables (columns) is interpreted as correlation, and the proximity between row and column markers is interpreted as preponderance.

2.3. Meta-analysis of diagnostic tests

When original studies that evaluate the quality of a test producing binary results are available, performing a diagnostic meta-analysis has become an important tool for investigating the information available on a test [6], [11], [19] and [20]. In a primary diagnostic study, the quality of a diagnostic test is often measured in terms of the sensitivity (TP rate) and the specificity (TN rate = $1 - \text{FP rate}$) of the test; that is, parallel to a standard gold procedure, which defines the presence of a certain condition, the diagnostic test is performed, after which the sensitivity and specificity can be calculated.

From the sensitivity and specificity, summary graphs can be made that show the variability between the studies. Thus, we have the following tools: (1) forest plot, which describes the sensitivity and specificity of each study with their respective confidence intervals; (2) crosshair, which demonstrates the bivariate relationship and degree of heterogeneity between sensitivity and FP rate; and (3) ROCellipse, which shows a region of confidence describing the uncertainty of the pair of sensitivity and specificity of each study.

2.4. Heterogeneity in meta-analysis diagnosis

The observed sensitivities and specificities can be expected to vary in the original studies that constitute a meta-analysis. This is because of two main reasons:

1. Different authors will calibrate a test differently. Given a questionnaire score or a biomarker level, an investigator will have to decide which minimum value (or maximum value) should give a positive test result.

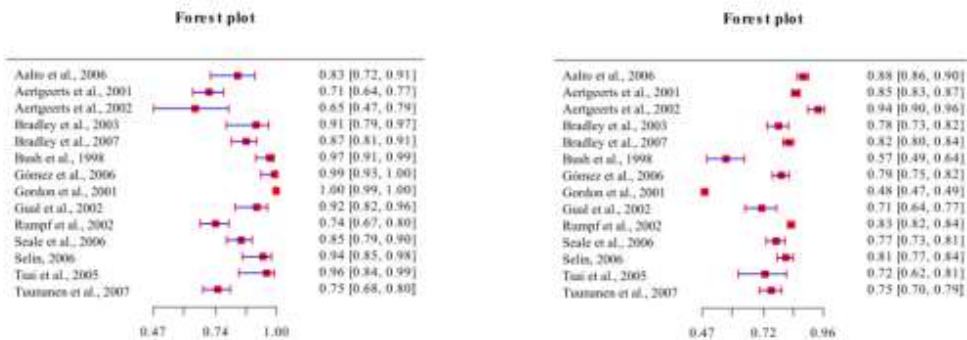
This value is known as the cutoff point. Sometimes, especially in screening for rare diseases, a cutoff value will be set to reach a certain level of sensitivity, such as 95%, which often leads to a small specificity; however, a certain type of relationship exists between sensitivity and specificity. Both approaches result in a different cutoff points. In general, a specific calibration of the population directed to a certain level of sensitivity will give rise to different cutoff points for different populations.

2. When a diagnostic test is applied to several populations, different sensitivities and specificities can be expected even if the same cutoff value is used.

In a diagnostic meta-analysis, the quality of a diagnostic test is evaluated by integrating data from individual studies, which usually include sensitivities and specificities; some of these original studies may not report the cutoff point. This challenge—i.e., nonhomogeneity and typically unknown cutoff points—is known as the cutoff point issue. Therefore, the bivariate nature of the data must be preserved, modeling sensitivity and specificity together. During the last decade, two models have been established: a hierarchical model [17] and a bivariate model [16]. However, two groups of researchers independently demonstrated [1] and [9] that hierarchical and bivariate models are equivalent in special cases (and most common) in the absence of covariance.

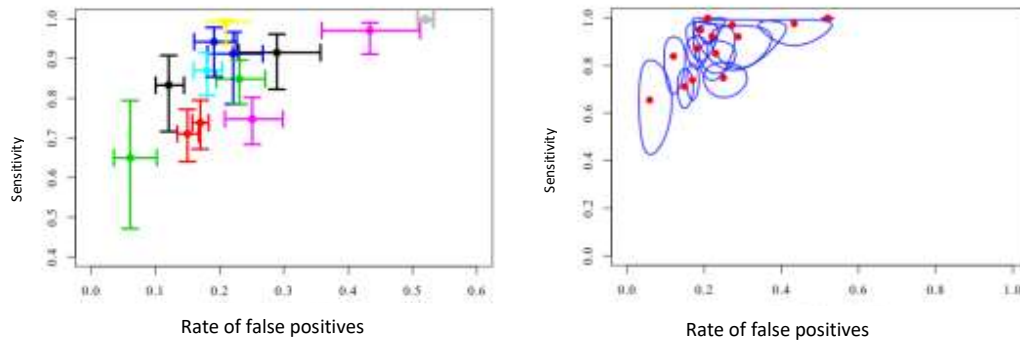
3. RESULTS

In this review, we included the Audit C data, which consists of data from 14 studies that evaluated 18,332 individuals, out of which 2071 had alcohol problems. Figure 1 shows the variation of studies in the sensitivities and specificities, despite the fact that the specificity was perfect (100%) in two studies, as shown in Figure 1 (a) and Figure 1 (b). This variability can be observed in the graphs in Figure 1 (c) and Figure (d) as crosshair and ROCellipse, respectively.



a. Forest plot for sensitivity

b. Forest plot for specificity



c. Weighted crosshair graph

d. Graph, regions of confidence in the form of ellipse

Figure 1. Forest plot, crosshair and ROC ellipse graphs, Audit C data

Table 2 shows the measures that corroborate the previous graphical analysis. Thus, there is a correlation between the sensitivity and the FP rate of 0.677, which denotes a significant relationship between both measures and a clear problem regarding the cutoff points in the studies. The true heterogeneity between studies, denoted by τ^2 , has a value of 0.311. I^2 is 33.955%, which indicates the percentage of heterogeneity among the studies because of the true variability (τ^2) between studies. In addition, the Q-statistic is available with a p-value of 0, which means that the null hypothesis of homogeneity between the studies is rejected in favor of the alternative hypothesis that there is heterogeneity among the studies analyzed.

Table 2. Correlation values, τ^2 , I^2 , and Q of Audit C data

Measure-statistic	Value	Lower limit (95%)	Higher limit (95%)	p-value
Correlation between sensitivity and specificity	0.677	0.228	0.888	-
τ^2	0.311	0.00	3.787	-
Q	4177.192	-	-	0
I^2	33.955	-	-	-

By performing the HJ biplot analysis on the data matrix described above, the following absorption of inertia (or variability explained) was obtained (see Table 3).

Table 3. Inertia absorbed by the two main axes in the HJ biplot representation, Audit C data.

Axis 1	Axis 2	Accumulated
66.46	28.69	95.15

As can be seen in Table 3, the accumulated absorption with the first two main axes is 95.15%, with a clear differentiation between axes 1 and 2. This leads us to interpret the first main plane (axes 1 and 2) together.

Figure 2 shows the simultaneous representations in the first main plane of the analysis performed. The number following the name of the author corresponds to the year that each study was conducted. Figure 2 shows the studies that form three clusters. Thus, the studies by Seale et al., 2006, Bradley et al., 2007, Aalto et al., 2006, Aertgeerts et al., 2002, Bradley et al., 2003, Gual et al., 2002, Selin, 2006, Gómez et al., 2006, Tsai et al., 2005, and Bush et al., 1998 belong to the first cluster, i.e., cluster 1 (green); the studies by Aertgeerts et al., 2001, Rumpf et al., 2002 and Tuunanen et al., 2007 belong to cluster 2 (red); finally, the study by Gordon et al., 2001 belongs to cluster 3 (blue). Note that the studies within a cluster have common characteristics among themselves, but that there is heterogeneity among the clusters.

The following tables of Relative Contribution of the Factor to the element, which show the value first for the rows (studies; Table 4) and then the columns (variables; Table 5), indicate the elements that are characteristic of axes 1 and 2. Note that these values of contributions come in a scale of 0–1000. Therefore, when an element that receives a high contribution of the axis 1 and a low contribution from the rest is an exclusive characteristic of that axis, and its interpretation will be done with respect to this axis. Values higher than a scale of 500 will

be considered for our analysis. Thus, studies by Gordon et al., 2001, Aertgeerts et al., 2002, Gual et al., 2002, Tsai et al., 2005, Bradley et al., 2003, Aalto et al., 2006, Aalto et al., 2006, Selin, 2006, and Bush et al., 1998 are representative of axis 1 as well as the variables FP, TP and TN. On the other hand, the studies by Aertgeerts et al., 2001, Tuunanen et al., 2007, Rumpf et al., 2002, Bradley et al., 2007 are representative of axis 2 as well as the variable FN.

Table 4. Relative contributions of the factor to the element-studies

Contributions rows - studies			
Studies	Axis 1	Studies	Axis 2
Gordon et al., 2001	937	Aertgeerts et al., 2001	946
Aertgeerts et al., 2002	899	Tuunanen et al., 2007	636
Gual et al., 2002	677	Rumpf et al., 2002	621
Tsai et al., 2005	635	Bradley et al., 2007	595
Bradley et al., 2003	629	Gómez et al., 2006 et al., 2007	549
Aalto et al., 2006	589	Bush et al., 1998	470
Selin, 2006	523	Selin, 2006	423
Bush et al., 1998	517	Tsai et al., 2005	362
Gómez et al., 2006	398	Bradley et al., 2003	347
Seale et al., 2006	335	Gual et al., 2002	320
Rumpf et al., 2002	274	Seale et al., 2006	170
Aertgeerts et al., 2001	53	Aalto et al., 2006	109
Tuunanen et al., 2007	31	Aertgeerts et al., 2002	85
Bradley et al., 2007	3	Gordon et al., 2001	61

Table 5. Relative contributions of the factor to the element-variables

Contributions rows - variables			
Variables	Axis 1	Variables	Axis 2
FP	940	FN	965
TP	922	TN	103
TN	787	FP	53
FN	9	TP	27

In Figure 2, it can be seen that there is atypical data corresponding to the study by **Gordon et al., 2001**, which is characterized by very high FP and TP values.

The horizontal gradient (axis 1) exclusively marks the difference between the study by **Gordon et al., 2001** and all the others, because the study by Gordon et al. shows high PF and TP values. The vertical gradient (axis 2) differentiates cluster 2 (red) studies (Aertgeerts et al., 2001, Rumpf et al., 2002 and Tuunanen et al., 2007) from cluster 1 (green) studies (Bush et al., 1998, Aertgeerts et al., 2002, Gual et al., 2002, Bradley et al., 2003, Tsai et al., 2005, Selin, 2006, Seale et al., 2006, Gómez et al., 2006 and Aalto et al., 2006, Bradley et al., 2007).

All cluster 2 (red) studies have very high FN values, but among them they differ by the values in the other three variables. Cluster 1 (green)- studies have very low values in all variables.

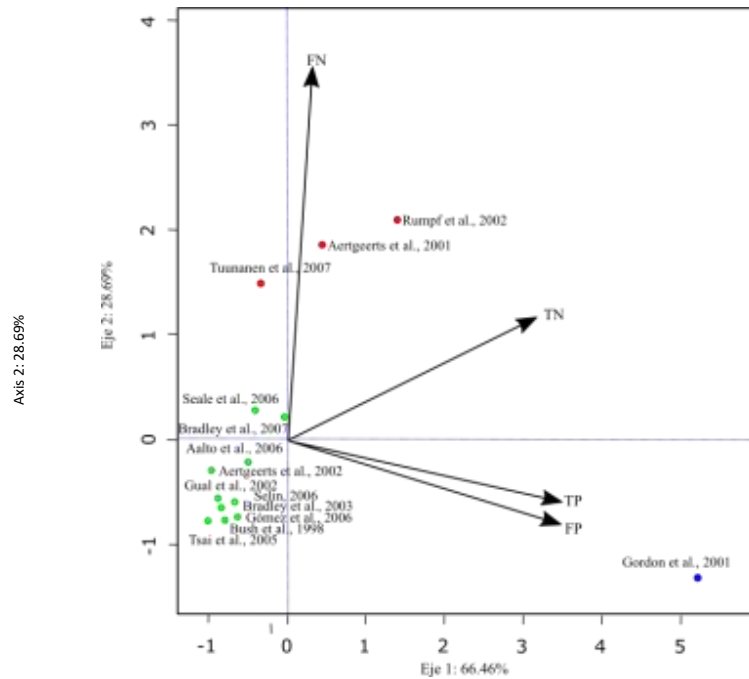


Figure 2. HJ biplot of Audit C data.

3. DISCUSSION AND CONCLUSIONS

- In this research, we analyzed the heterogeneity of the studies in meta-analysis of diagnostic tests using the Q -statistic, I^2 index, and HJ biplot. The evaluation of heterogeneity in a meta-analysis is a crucial question, because the meta-analyst's decision to select the statistical model to be applied (fixed-effects model versus random-effects model) depends on the homogeneity test employed [13]. Considering the importance of this question, the purpose of this study is to compare the behavior of the Q -statistic and the I^2 index, and to complement the information they provide with the results obtained using the HJ biplot to evaluate the heterogeneity between cluster of individual studies in a meta-analysis.

- The I^2 index behaves similarly to the Q test from an inferential point of view [14], but the I^2 index has advantages over the classical Q test—it is easily interpretable because it is a percentage and does not depend on degrees of freedom [4]. Another advantage is that it provides a way of assessing the magnitude of heterogeneity, whereas the Q test reports on the statistical significance of the homogeneity hypothesis [4]. However, these two tests do not provide information about the possible causes of heterogeneity; that is, they do not indicate whether studies are similarly responsible for the heterogeneity or which variables used in the individual studies are responsible for the heterogeneity.

- The HJ biplot method [8] provides a graphical representation of the studies and the TP, TN, FP, and FN in the same system of Cartesian axes with minimal information loss. This representation provides information on the overall structure of the studies that is complementary to that provided by Higgins' Q and I^2 statistics. The length of the vectors representing the variables informs the variability of the variables; thus, analyzing this length can detect which variables are responsible for the heterogeneity of a meta-analysis. The cosines of the angles between the vectors show the correlation between the variables they represent, thus providing useful information when choosing the appropriate model for the available data; if there is a strong correlation between FP and TP, a clear cutoff point issue becomes evident [5]. Thus, it is necessary to model these measures using a hierarchical approach [17] or a bivariate approach [16]. The points represent the different studies to be integrated into the meta-analysis of a diagnostic test. The distances between them allow the visualization of clusters of studies with homogenous characteristics. Those clusters that are further away are

responsible for the heterogeneity. Thus, by projecting such studies orthogonally on the direction of the TP, TN, FP and FN, the causes of heterogeneity can be obtained.

- The HJ biplot has proven to be a multivariate tool that is extremely useful in the analysis of meta-analysis data of diagnostic tests, both in the descriptive phase and in the search for the causes of variability. The points in Figure 2 display the different studies to be integrated into the meta-analysis of a diagnostic test.

- In conclusion, the use of the HJ biplot in the meta-analysis of diagnostic tests allowed us the following: (i) characterization of the heterogeneity of the studies regarding measures such as TP, FP, TN, and FN; (ii) identification of groups of homogeneous studies; and (iii) analysis of the relations of these measures to choose the most appropriate model that facilitates the subsequent integration of the individual characteristics of the studies analyzed.

RECEIVED: JULY, 2017
REVISED: JANUARY, 2018

REFERENCES

- [1] ARENDS, L., HAMZA, T., VAN HOUWELINGEN, J., HEIJENBROK-KAL, M., HUNINK, M., & STIJNEN, T. (2008). Bivariate random effects meta-analysis of roc curves. **Medical Decision Making**, 28, 621-638.
- [2] COCHRAN, W. G. (1954). The combination of estimates from different experiments. **Biometrics**, 10, 101-129.
- [3] CORNWELL, J. M. (1993). Monte Carlo comparisons of three tests for homogeneity of independent correlations. **Educational and Psychological Measurement**, 53, 605-618.
- [4] DEEKS, J. J., ALTMAN, D. G., & BRADBURN, M. J. (2008). Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. **Systematic Reviews in Health Care: Meta-Analysis in Context, Second Edition**, 285-312.
- [5] DOEBLER, P., HOLLING, H., & BÖHNING, D. (2012). A mixed model approach to meta-analysis of diagnostic studies with binary test outcome. **Psychological Methods**, 17, 418.
- [6] EGGER, M., DAVEY-SMITH, G., & ALTMAN, D. (Eds.). (2008). **Systematic reviews in health care: meta-analysis in context**. John Wiley & Sons, N. York.
- [7] GABRIEL, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. **Biometrika**, 58, 453-467.
- [8] GALINDO VILLARDÓN, M. (1986). Una alternativa de representación simultanea: Hj-biplot. **Qüestió. quaderns d'estadística i investigació operativa**, 10(1), 13-23.
- [9] HARBORD, R. M., DEEKS, J. J., EGGER, M., WHITING, P., & STERNE, J. A. (2006). A unification of models for meta-analysis of diagnostic accuracy studies. **Biostatistics**, 8, 239-251.
- [10] HARDY, R. J., & THOMPSON, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. **Statistics in Medicine**, 17, 841-856.
- [11] HASSELBLAD, V., & HEDGES, L. V. (1995). Meta-analysis of screening and diagnostic tests. **Psychological bulletin**, 117(1), 167.
- [12] HIGGINS, J., & THOMPSON, S. G. (2002). Quantifying heterogeneity in a meta-analysis. **Statistics in medicine**, 21, 1539-1558.

- [13] HIGGINS, J. P., THOMPSON, S. G., DEEKS, J. J., & ALTMAN, D. G. (2003). Measuring inconsistency in meta-analyses. **BMJ: British Medical Journal**, 327, 557.
- [14] HUEDO-MEDINA, T. B., SÁNCHEZ-MECA, J., MARÍN-MARTÍNEZ, F., & BOTELLA, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I² index?. **Psychological methods**, 11, 193.
- [15] KRISTON, L., H ÖLZEL, L., WEISER, A.K., BERNER, M. M., & H ÄRTER, M. (2008). Meta-analysis: are 3 questions enough to detect unhealthy alcohol use?. **Annals of Internal Medicine**, 149, 879-888.
- [16] REITSMA, J. B., GLAS, A. S., RUTJES, A. W., SCHOLTEN, R. J., BOSSUYT, P. M., & ZWINDERMAN, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. **Journal of Clinical Epidemiology**, 58, 982-990.
- [17] RUTTER, C. M. & GATSONIS, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. **Statistics In Medicine**, 20, 2865-2884.
- [18] SACKETT, P. R., HARRIS, M. M., & ORR, J. M. (1986). On seeking moderator variables in the meta-analysis of correlational data: A Monte Carlo investigation of statistical power and resistance to Type I error.
- [19] SCHULZE, R. E., HOLLING, H. E., & BÖHNING, D. E. (2003). Meta-analysis: **New Developments and Applications in Medical and Social Sciences**. Hogrefe & Huber Publishers.
- [20] SUTTON, A. J., ABRAMS, K. R., JONES, D. R., SHELDON, T. A., & SONG, F. (2003). Methods for meta-analysis in medical research. **Statistics in Medicine**, 22(19), 3111-3114.

Appendix

```
install.packages("mada")
library("mada")
data("AuditC")
AuditC

forest(madad(AuditC), type = "sens")
forest(madad(AuditC), type = "spec")

rs <- rowSums(AuditC)
weights <- 4 * rs / max(rs)
crosshair(AuditC, xlim = c(0,0.6), ylim = c(0.4,1), col = 1:14, lwd = weights)

ROCellipse(AuditC, pch = "")
points(fpr(AuditC), sens(AuditC))

library(tcltk)
library(tcltk2)
library(rgl)
library(tkrplot)
library(shapes)
library(cluster)
library(dendroextras)
library(MASS)
tclRequire(BWidget)
library(biplotbootGUI)
biplotboot(AuditC)
```


B.2. Meta-Analysis in DTA with Hierarchical Models Bivariate and HSROC: Simulation Study

Meta-Analysis in DTA with Hierarchical Models Bivariate and HSROC: Simulation Study



Sergio A. Bauz-Olvera, Johny J. Pambabay-Calero, Ana B. Nieto-Librero and Ma. Purificación Galindo-Villardón

1 **Abstract** In meta-analysis for diagnostic test accuracy (DTA), summary measures
2 such as sensitivity, specificity, and odds ratio are used. However, these measures
3 may not be adequate to integrate studies with low prevalence, which is why statistical
4 modeling based on true positives and false positives is necessary. In this context, there
5 are several statistical methods, the first of which is a bivariate random effects model,
6 part of the assumption that the logit of sensitivity and specificity follow a bivariate
7 normal distribution, the second, refers to the HSROC or hierarchical model, is similar
8 to bivariate, with the particularity that it directly models the sensitivity and specificity
9 relationship through cut points. Using simulations, we investigate the performance
10 of hierarchical models, varying their parameters and hyperparameters and proposing
11 a better management of variability within and between studies. The results of the
12 simulated data are analyzed according to the criterion of adjustment of the models
13 and estimates of their parameters.

S. A. Bauz-Olvera (✉)

Facultad de Ciencias de la Vida, Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador
e-mail: serabauz@espol.edu.ec

J. J. Pambabay-Calero

Facultad de Ciencias Naturales y Matemáticas, Escuela Superior Politécnica del Litoral,
Guayaquil, Ecuador
e-mail: jpambaba@espol.edu.ec

A. B. Nieto-Librero · M. P. Galindo-Villardón

Dpto. de Estadística, Facultad de Medicina, Universidad de Salamanca, Salamanca, Spain
e-mail: ananieto@usal.es

M. P. Galindo-Villardón

e-mail: pgalindo@usal.es

A. B. Nieto-Librero · M. P. Galindo-Villardón

Instituto de Investigación Biomédica (IBSAL), Salamanca, Spain

© Springer Nature Switzerland AG 2019

I. Antoniano-Villalobos et al. (eds.), *Selected Contributions on Statistics and Data Science in Latin America*, Springer Proceedings in Mathematics & Statistics 301,
https://doi.org/10.1007/978-3-030-31551-1_3

1

14 **Keywords** Diagnostic precision · Meta-analysis · HSROC model · Bivariate
 15 model · Low prevalence

16 **1 Introduction**

17 AQ1 One way to summarize the behavior of a diagnostic test from multiple studies is by
 18 calculating the mean sensitivity and the mean specificity, but these summary measures
 19 are not valid if there is heterogeneity between the studies and unfortunately the tests
 20 to detect heterogeneity are not very powerful [1], so that hierarchical models are
 21 necessary, which are capable of capturing this heterogeneity. In meta-analysis, the
 22 objective is to integrate the results of the different articles in which a diagnostic test
 23 is evaluated.

24 The starting information is a tetrachoric table in which measures such as true
 25 positive (TP), false positive (FP), false negative (FN), and true negative (TN) appear.
 26 Table 1 shows the concordance between the result of the test in binary form and the
 27 information on the disease (Fig. 1).

28 While the statistical methods for the meta-analysis of clinical trials are well devel-
 29 oped and understood today, there are still challenges when performing meta-analysis
 30 of data from studies on the accuracy of the diagnostic test. This is mainly due to the
 31 bivariate nature of the response, where information on sensitivity and specificity
 32 should be summarized.

33 This correlation can be induced by populations or heterogeneous environments
 34 in the different trials, but it is mainly derived from the balance between sensitivity
 35 and specificity when the threshold to define the positivity of the test varies.

36 **1.1 Hierarchical Models**

37 More rigorous approaches have been proposed statistically based on hierarchical
 38 models that overcome the limitations of the method proposed by [2]. This section
 39 briefly describes the bivariate model [3] and the hierarchical model (HSROC) [4].
 40 Both hierarchical models imply statistical distributions in two levels. At the lower
 41 level, the count of the values taken by the tetrachoric tables extracted from each study
 42 is modeled using binomial distributions and logistic transformations. At the higher
 43 level, it is assumed that the random effects of the study explain the heterogeneity

Table 1 Generic representation of the precision analysis of a diagnostic test

Patient's condition		
Result of the test	Positive	Negative
Positive	TP	FP
Negative	FN	TN

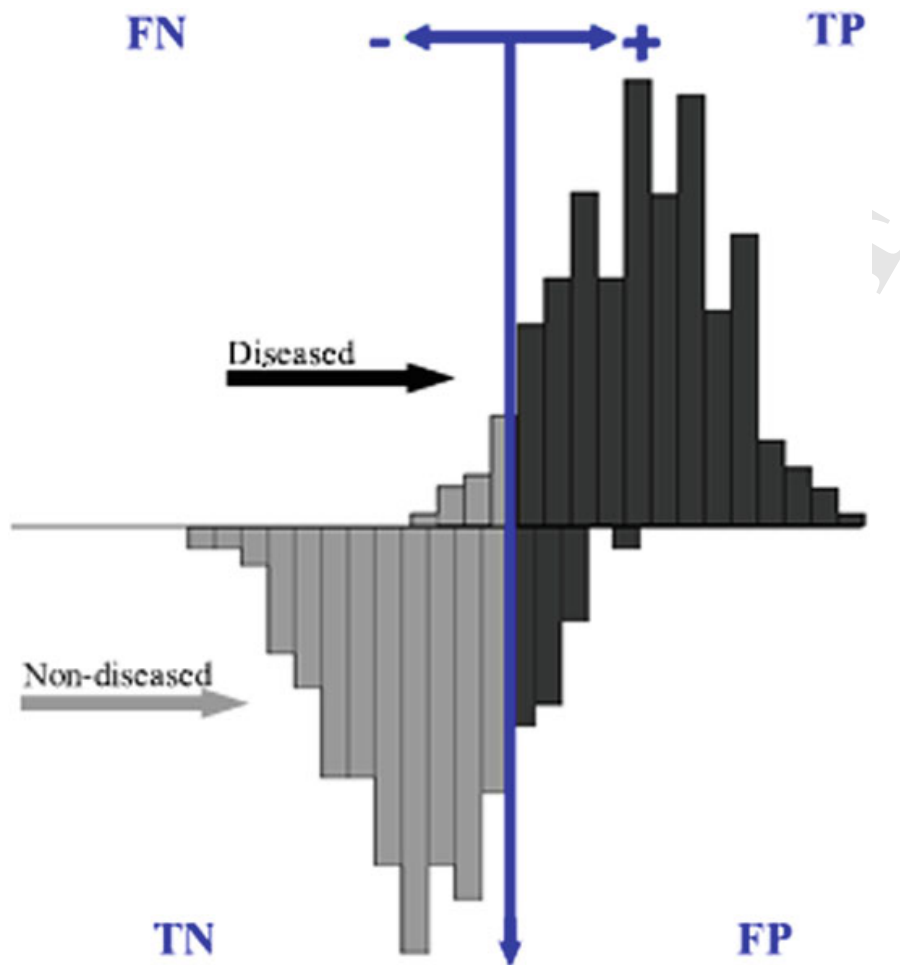


Fig. 1 Test threshold and impact on diagnostic accuracy

44 between the studies beyond what is explained by the sampling variability at the lower
 45 level. The bivariate model and the HSROC model are mathematically equivalent
 46 when covariates are not available [5, 6] but differ in their parameterizations.

47 The Bivariate parameterization models the sensitivity, the specificity and the cor-
 48 relation between them directly. While the parameterization of the model HSROC
 49 models functions of thresholds of positivity, precision, and shape of the curve to
 50 define an SROC graph. AQ2

51 **1.2 Bivariate Model**

52 The bivariate model uses a different starting point for the ordered pairs of sensitivity
 53 and specificity of the meta-analyses. Instead of transforming these two measurements
 54 into a diagnostic indicator of accuracy, as in the SROC approach, the bivariate model
 55 preserves the two dimensions of the data through analysis. It is assumed that the
 56 sensitivity values of the individual studies (after the logit transformation) of a meta-
 57 analysis follow a normal distribution around the value of the mean and with a certain
 58 amount of variability around the same mean. This is a random effects approach,

59 similar to that used in therapeutic trials, to incorporate the variability not explained
60 in the analysis. This variation in the underlying sensitivity between the studies may be
61 related to the differences that persist in the study population, the implicit differences
62 in the threshold, or imperceptible variations in the clinical trial protocol.

63 The same considerations apply to the specificities of these studies. The potential
64 presence of a (negative) correlation between sensitivity and specificity within the
65 studies is addressed by the explicit incorporation of this correlation in the analysis.
66 The combination of normal distribution and logit transformations of sensitivity and
67 specificity, recognize the possible correlation between them, leading to a normal
68 bivariate distribution [7, 8].

69 The Bivariate method models sensitivity and specificity directly. It can be consid-
70 ered that the model has two levels corresponding to the variation within and between
71 the studies. In the first level, it is assumed that the variability within the study for sen-
72 sitivity and specificity follows a binomial distribution [9] have shown that a binomial
73 probability distribution should be used to model variability within studies (especially
74 when data are scarce). A more formal definition is the following.

75 A and B are denoted for sensitivity and specificity, respectively, and the number of
76 true positives for y_{Ai} in study i , where n_{Ai} denote the total number of sick individuals
77 and π_{Ai} represents the probability that the test will give a positive result in study i .
78 Similarly, y_{Bi} the number of true negatives, where n_{Bi} denotes the total of healthy
79 individuals and π_{Bi} is the probability that the test will give a negative result in study
80 i , that is:

$$81 \quad y_{Ai} \sim \text{Binomial}(n_{Ai}, \pi_{Ai}), y_{Bi} \sim \text{Binomial}(n_{Bi}, \pi_{Bi}) \quad (1)$$

82 The specificity pair, sensitivity for each study should be modeled jointly at level
83 one (lower) since these measures are linked by study characteristics including the
84 threshold of positivity. At the upper level, it is assumed that the logit transformations
85 of the sensitivities have a normal with mean μ_A and variance σ_A^2 , while the logit
86 transformations of the specificities are distributed by a normal with mean μ_B and
87 σ_B^2 . Its correlation is included by modeling at the same time a single bivariate normal
88 distribution, which allows the joint analysis of the sensitivity and specificity, which
89 is a linear mixed model:

$$90 \quad \begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \Sigma \right), \Sigma = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix} \quad (2)$$

91 where σ_A^2 and σ_B^2 denote the variability between the logit transformations of sensi-
92 tivity and specificity, respectively, and σ_{AB} denotes the covariance between the logit
93 transformations of sensitivity and specificity. The model can also be parameterized
94 using the correlation $\rho_{AB} = \sigma_{AB}/(\sigma_A \sigma_B)$, which may be more interpretable than
95 covariance. Therefore, the bivariate model without covariations has the following
96 five parameters: μ_A , μ_B , σ_A^2 , σ_B^2 , and σ_{AB} (o ρ_{AB}). The Bivariate statistical model
97 can be simplified assuming covariances or correlations equal to zero (that is, an inde-
98 pendent variance–covariance structure), then the model is reduced to two univariate
99 random effects logistic regression models for sensitivity and specificity.

100 1.3 Model HSROC

101 The HSROC model represents a general framework for the meta-analysis of trial
 102 accuracy studies and can be seen as an extension of the SROC approach of [2] in
 103 which rates of true positives (TTP) and false positive rate (TFP) for each study are
 104 directly modeled [10]. The HSROC model is a generalized mixed nonlinear model
 105 and takes the following form:

$$106 \quad \text{logit}(\pi_{ij}) = (\theta_i + \alpha_i \text{dis}_{ij}) \exp(-\beta \text{dis}_{ij}) \quad (3)$$

107 where π_{ij} is the probability that a patient in study i with a disease condition j will
 108 obtain a positive test result. The disease state is represented by dis_{ij} , which is coded
 109 with -0.5 ($j = 0$) for individuals who do not possess the disorder (disease) and 0.5
 110 ($j = 1$) for the group of individuals with the disease for the i -th study.

111 The implicit threshold θ_i (threshold parameter or positivity criterion) and α_i the
 112 diagnostic accuracy for each study, are modeled as random effects by means of
 113 independent, variances of normal distributions: $\theta_i \sim N(\Theta, \sigma_\theta^2)$ and $\alpha_i \sim N(\Lambda, \sigma_\alpha^2)$,
 114 respectively. The modeling also includes a parameter of shape or scale β that allows
 115 the asymmetry in the SROC curve when admitting that the precision varies with the
 116 implicit threshold.

117 Therefore, the SROC curve is symmetric if $\beta = 0$ or asymmetric if $\beta \neq 0$. Each
 118 study provides a single point in the ROC space, and therefore, the estimation of β
 119 requires information from all the studies included in the meta-analysis. Thus, β is
 120 modeled as a fixed effect. The HSROC model has the following five parameters: AQ3
 121 $\Lambda, \Theta, \beta, \sigma_\alpha^2$ y σ_θ^2 ; the model is reduced to a fixed-effect model if: $\sigma_\alpha^2 = \sigma_\theta^2 = 0$.
 122 Other specifications for SROC curves based on Bivariate model functions have been
 123 proposed [5, 11], in this work we will focus our interest on the Rutter and Gatsonis
 124 model.

125 2 Methodology

126 A simulation study was carried out to compare the HSROC model with various
 127 simplifications (by eliminating the parameters of the model). We have chosen the
 128 HSROC model because it has greater flexibility when choosing parameters [12].
 129 The specifications of the scenarios were designed to reproduce realistic situations
 130 found in the meta-analysis of diagnostic accuracy studies. The effect of these factors
 131 was investigated: (1) number of studies; (2) prevalence of the disease; (3) variability
 132 between the studies in accuracy and threshold; and (4) asymmetry in the SROC curve.
 133 We only investigate methods that use a binomial distribution.

134 2.1 Generation of the Simulated Data

135 Meta-analyzes with different numbers of studies were investigated randomly ($N =$
 136 5, 10, 20, 35). The size of a study in each meta-analysis, n_j , was randomly sam-
 137 pled from a uniform distribution, $U(200; 2000)$; Given an underlying p prevalence,
 138 individuals within each study were randomly classified as sick or undiagnosed, and
 139 a continuous test result value, x , was assigned that was sampled randomly [13]. To
 140 determine the corresponding tetrachoric tables, we used the HSROC package of the
 141 statistical program R [14], by using the `simdata` function (see Fig. 2). To create the
 142 2×2 table for each study, the individuals were classified as true positives, false neg-
 143 atives, false positives or true negatives based on the result of the test and the state of
 144 the disease. The prevalences of the studies were adjusted by a uniform distribution
 145 with parameters 0 and 0.25, that is, the prevalence of these studies took values less
 146 than or equal to 25%.

147 We generated 23110 studies in 1200 independent meta-analysis datasets to allow
 148 an accurate estimation of model performance, even if a large proportion of models
 149 do not converge. In order to execute the `simdata` function, it is necessary to specify
 150 the input parameters of the function, that is, it is necessary to choose the a priori
 151 distributions of the `simdata` function, the initial parametrization was performed using
 152 the following initial conditions [15]:

$$153 \text{ Dataset } \left\{ \begin{array}{l} N \sim \text{Uniform} [5, 35] \\ n \sim \text{Uniform} [200, 2000] \\ p \sim \text{Uniform} [0.0001, 0.2499] \\ \beta \sim \text{Uniform} [-0.75, 0.75] \\ \Lambda \sim \text{Uniform} [-0.75, 0.75] \\ \sigma_{\alpha}^2 \sim \text{Uniform} [0.0001, 2] \\ \Theta \sim \text{Uniform} [-1.5; 1.5] \\ \sigma_{\theta}^2 \sim \text{Uniform} [0.0001, 2] \end{array} \right.$$

154 Additionally, in each generated meta-analysis, a covariate with categories from 1
 155 to 3 was included, through a uniform distribution $U(1, 3)$.

156 2.2 Hierarchical Models Adjusted to the Simulated Data

157 In the later part we refer to an HSROC model that contains five parameters. The
 158 following five models were adjusted to the 1200 sets of data generated. From Table 2,
 159 it follows that for an HSROC model with precision and fixed thresholds (model 5),
 160 it is only necessary to identify a priori the distributions for, Λ , Θ , and β . Similarly,
 161 the remaining four models are understood.

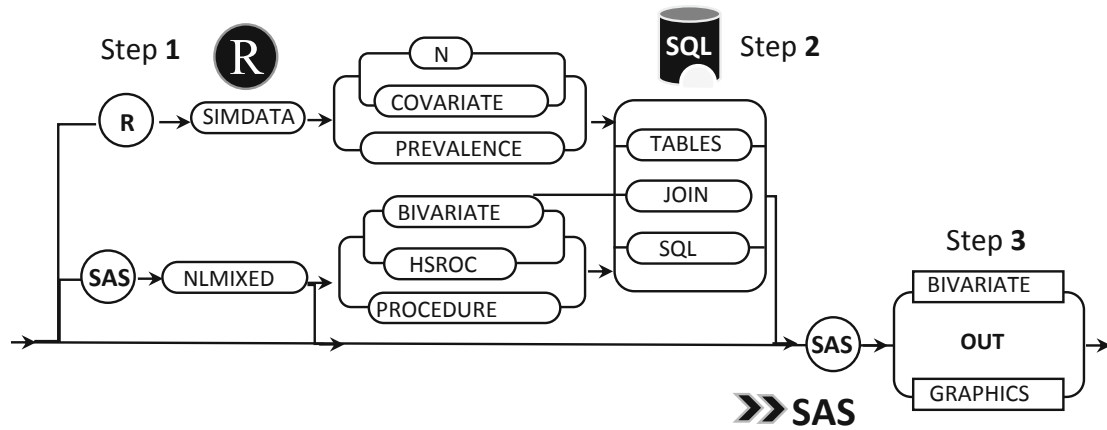


Fig. 2 General scheme of work and software required for the simulation of the data set (dataset), storage in database [16] and statistical analysis for the presentation of results [17]

Table 2 Structure of the models adjusted to the simulated data

Num.	Models	Λ	Θ	β	σ_{α}^2	σ_{θ}^2
1	Complete model HSROC	✓	✓	✓	✓	✓
2	Symmetrical HSROC model	✓	✓		✓	✓
3	HSROC model with fixed thresholds	✓	✓	✓	✓	
4	HSROC model with fixed precision	✓	✓	✓		✓
5	HSROC model with fixed threshold and accuracy	✓	✓	✓		

162 Note that due to the mathematical relationship between the bivariate model and
 163 HSROC, it is possible to find the estimates of the five parameters of the HSROC
 164 model by means of the reitsma function of the R mada package [6].

165 3 Results

166 The results of the simulated data (Table 3) show the different models that were
 167 obtained by varying the hyperparameters of the complete hierarchical model.

168 Each category of the model is compared according to the adjustment criteria
 169 $(-2\text{LogLikelihood}, \text{AIC}, \text{AICC}, \text{and BIC})$ with their respective confidence intervals.

170 For example, we see that the best model under prevalence conditions between
 171 0.01 and 25% and the selection based on the lowest information criterion of Akaike
 172 “AIC” is Model 2, in which case it corresponds to a symmetric HSROC model with
 173 precision and threshold fixed.

174 Table 4 shows the estimates of the parameters of the bivariate model through the
 175 execution of the NLMIXED procedure and the incorporation or not of a covariate.

Table 3 Adjustment of the HSROC models, with parameters in each model and adjustment measures

HSROC Model					
	Without covariate	With covariate			
		Threshold	Accuracy	Accuracy and threshold	Accuracy and threshold and shape
	Confidence intervals 95%				
Parameter estimates	Δ : (-0.37, 0.26)	(-0.34, 0.29)	(-0.29, 0.35)	(-0.33, 0.31)	(-0.33, 0.30)
	θ : (-0.31, 0.03)	(-0.32, 0.03)	(-0.28, 0.07)	(-0.30, 0.04)	(-0.30, 0.05)
	β : (0.99, 1.17)	(0.99, 1.17)	(1.00, 1.18)	(0.98, 1.17)	(0.97, 1.16)
	σ_α^2 : (2.49, 3.52)	(2.58, 3.90)	(2.40, 3.42)	(2.50, 3.53)	(2.49, 3.52)
	σ_θ^2 : (1.32, 1.68)	(1.43, 1.83)	(1.35, 1.75)	(1.38, 1.79)	(1.34, 1.73)
Goodness of fit statistics	-2logl: (1027, 1161)	(1012, 1145)	(1022, 1155)	(1014, 1147)	(1019, 1153)
	AIC: (1037, 1171)	(1024, 1157)	(1034, 1167)	(1028, 1161)	(1035, 1169)
	AICC: (1040, 1174)	(1029, 1161)	(1039, 1171)	(1035, 1168)	(1048, 1182)
	BIC: (1041, 1175)	(1029, 1162)	(1039, 1172)	(1033, 1166)	(1042, 1176)

Table 4 Adjustment of the bivariate model, bivariate model with and without covariate

Bivariate model		
	Without covariate	With covariate
	Confidence intervals 95%	
Parameter estimates	μ_A : (-4.48, -3.56)	(-4.11, -3.27)
	μ_B : (-0.12, -0.11)	(-0.30, 0.14)
	σ_A^2 : (-0.16, -0.07)	(-0.14, 0.10)
	σ_B^2 : (5.25, -12.85)	(2.58, 3.90)
	σ_{AB} : (7.05, -8.39)	(6.52, 7.53)
Fit statistics	-2logl: (372.23, 392.42)	(370.29, 390.5)
	AIC: (382.23, 402.42)	(384.29, 404.5)
	AICC: (385.07, 405.03)	(390.99, 410.57)
	BIC: (386.52, 407)	(390.3, 410.9)

176 It is important to emphasize that the incorporation of a covariate to the model
 177 provides better adjustment to the modeling (see -2Log Likelihood).

178 The convergence problems of the HSROC models were presented with greater
 179 recurrence when: (1) the estimated variance of the parameters (σ_α^2 , σ_θ^2) were very
 180 close to zero; (2) when the relation that the variance of the parameter α is twice or
 181 more the variance of the parameter θ , i.e., ($\sigma_\alpha^2/\sigma_\theta^2 > 2$) is satisfied; (3) when the
 182 variance of the random effects of the threshold and the precision are very small or

183 close to zero, this is understandable since the HSROC model was designed to model
184 the random effects.

185 4 Discussion

186 The categories of hierarchical models obtained by varying the hyperparameters
187 directly model the precision and the threshold in the generalized mixed nonlin-
188 ear model [9], with the simulation of the different scenarios and using the available
189 software, we investigate situations extremes of the hierarchical HSROC model to
190 understand its complex execution in meta-analysis for diagnostic accuracy tests.
191 Convergence problems arose when the variance of one of the random effects is close
192 to zero. This is particularly a problem for the parameterization of the bivariate model,
193 where an examination of the scatter plot can help to identify a strong heterogeneity
194 in sensitivity. We recommend a bivariate approach when the heterogeneity between
195 the included studies is moderate, and it is necessary to estimate a summary measure
196 for the sensitivity and specificity, but if the degree of heterogeneity is significant, we
197 recommend an HSROC model if the estimation of an SROC curve is necessary with
198 their respective variations.

199 Appendix

200 Availability and requirements for R Program

201 **Project name:** BivariateHSROC_Simulation

202 **Project homepage:** <https://sourceforge.net/projects/metahi/>

203 **File:** sourceCodeR.txt

204 **Operating system(s):** Microsoft Windows, Linux and Mac

205 **Programming language:** R, **License:** Open Source and free

206

207 Availability and requirements for SAS Program

208 **Project name:** BivariateHSROC_Simulation

209 **Project homepage:** <https://sourceforge.net/projects/metahi/>

210 **File:** sourceCodeSAS.txt

211 **Operating system(s):** Microsoft Windows, Linux and Mac

212 **Programming language:** SAS, **License:** Open Source and free

213

214 **References**

- 215 1. Midgette, A.S., Stukel, T.A., Littenberg, B.: A meta-analytic method for summarizing diagnos-
 216 tic test performances: receiver-operating-characteristic-summary point estimates. *Med. Decis.*
 217 *Mak.* **13**(3), 253–257 (1993)
- 218 2. Moses, L.E., Shapiro, D., Littenberg, B.: Combining independent studies of a diagnostic test
 219 into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat.*
 220 *Med.* **12**(14), 1293–1316 (1993)
- 221 3. Reitsma, J.B., Glas, A.S., Rutjes, A.W., Scholten, R.J., Bossuyt, P.M., Zwinderman, A.H.:
 222 Bivariate analysis of sensitivity and specificity produces informative summary measures in
 223 diagnostic reviews. *J. Clin. Epidemiol.* **58**(10), 982–990 (2005)
- 224 4. Rutter, C.M., Gatsonis, C.A.: A hierarchical regression approach to meta-analysis of diagnostic
 225 test accuracy evaluations. *Stat. Med.* **20**(19), 2865–2884 (2001)
- 226 5. Arends, L.R., Hamza, T.H., Van Houwelingen, J.C., Heijnenbrok-Kal, M.H., Hunink, M.G.M.,
 227 Stijnen, T.: Bivariate random effects meta-analysis of ROC curves. *Med. Decis. Mak.* **28**(5),
 228 621–638 (2008)
- 229 6. Harbord, R.M., Deeks, J.J., Egger, M., Whiting, P., Sterne, J.A.: A unification of models for
 230 meta-analysis of diagnostic accuracy studies. *Biostatistics* **8**(2), 239–251 (2006)
- 231 7. Kotz, S., Balakrishnan, N., Johnson, N.L.: Bivariate and trivariate normal distributions. *Contin.*
 232 *Multivar. Distrib.* **1**, 251–348 (2000)
- 233 8. Van Houwelingen, H.C., Arends, L.R., Stijnen, T.: Advanced methods in meta-analysis: mul-
 234 tivariate approach and meta-regression. *Stat. Med.* **21**(4), 589–624 (2002)
- 235 9. Chu, H., Guo, H., Zhou, Y.: Bivariate random effects meta-analysis of diagnostic studies using
 236 generalized linear mixed models. *Med. Decis. Mak.* **30**(4), 499–508 (2010)
- 237 10. Macaskill, P.: Empirical Bayes estimates generated in a hierarchical summary ROC analysis
 238 agreed closely with those of a full Bayesian analysis. *J. Clin. Epidemiol.* **57**(9), 925–932 (2004)
- 239 11. Chappell, F.M., Raab, G.M., Wardlaw, J.M.: When are summary ROC curves appropriate for
 240 diagnostic meta-analyses? *Stat. Med.* **28**(21), 2653–2668 (2009)
- 241 12. Kriston, L., Hölzel, L., Weiser, A.K., Berner, M.M., Härter, M.: Meta-analysis: are 3 questions
 242 enough to detect unhealthy alcohol use? *Ann. Intern. Med.* **149**(12), 879–888 (2008)
- 243 13. Bachmann, L.M., Puhan, M.A., Ter Riet, G., Bossuyt, P.M.: Sample sizes of studies on diag-
 244 nostic accuracy: literature survey. *Bmj* **332**(7550), 1127–1129 (2006)
- 245 14. Doebler, P., Holling, H.: Meta-analysis of diagnostic accuracy with mada. Retrieved at: [https://](https://cran.rproject.org/web/packages/mada/vignettes/mada.pdf)
 246 cran.rproject.org/web/packages/mada/vignettes/mada.pdf (2015)
- 247 15. Dendukuri, N., Schiller, I., Joseph, L., Pai, M.: Bayesian meta-analysis of the accuracy of a
 248 test for tuberculous pleuritis in the absence of a gold standard reference. *Biometrics* **68**(4),
 249 1285–1293 (2012)
- 250 16. Grycuk, R., Gabryel, M., Scherer, R., Voloshynovskiy, S.: Multi-layer architecture for storing
 251 visual data based on WCF and microsoft SQL server database. In: *International Conference on*
 252 *Artificial Intelligence and Soft Computing*, pp. 715–726. Springer, Cham (2015)
- 253 17. Sheu, C.F., Chen, C.T., Su, Y.H., Wang, W.C.: Using SAS PROC NLMIXED to fit item response
 254 theory models. *Behav. Res. Methods* **37**(2), 202–218 (2005)

Author Queries

Chapter 3

Query Refs.	Details Required	Author's response
AQ1	Please confirm if the corresponding author is correctly identified. Amend if necessary.	
AQ2	Please check if the sentence 'While the parameterization of the...' conveys the intended meaning.	
AQ3	Please check the term 'y' in the sentence 'The HSROC model...'.	

UNCORRECTED PROOF

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

<i>Instruction to printer</i>	<i>Textual mark</i>	<i>Marginal mark</i>
Leave unchanged	... under matter to remain	Ⓧ
Insert in text the matter indicated in the margin	∧	New matter followed by ∧ or ∧ [Ⓧ]
Delete	/ through single character, rule or underline or ┌───┐ through all characters to be deleted	∕ or ∕ [Ⓧ]
Substitute character or substitute part of one or more word(s)	/ through letter or ┌───┐ through characters	new character / or new characters /
Change to italics	— under matter to be changed	∟
Change to capitals	≡ under matter to be changed	≡
Change to small capitals	≡ under matter to be changed	≡
Change to bold type	~ under matter to be changed	~
Change to bold italic	≈ under matter to be changed	≈
Change to lower case	Encircle matter to be changed	≠
Change italic to upright type	(As above)	⊕
Change bold to non-bold type	(As above)	⊖
Insert 'superior' character	/ through character or ∧ where required	∕ or ∕ [∧] under character e.g. ∕ [∧] or ∕ [∧]
Insert 'inferior' character	(As above)	∧ over character e.g. ∧
Insert full stop	(As above)	⊙
Insert comma	(As above)	,
Insert single quotation marks	(As above)	∕ or ∕ [∧] and/or ∕ or ∕ [∧]
Insert double quotation marks	(As above)	∕ or ∕ [∧] and/or ∕ or ∕ [∧]
Insert hyphen	(As above)	∓
Start new paragraph	┌	┌
No new paragraph	~	~
Transpose	└	└
Close up	linking ○ characters	Ⓧ
Insert or substitute space between characters or words	/ through character or ∧ where required	∕ or ∕ [∧]
Reduce space between characters or words		↑

B.3. Meta-Análisis del Estudio de la Precisión Diagnóstica empleando Modelización de Efectos Aleatorios



V Congreso Internacional en Contextos Psicológicos, Educativos y de la Salud

Madrid, 21, 22 y 23 de noviembre de 2018

CERTIFICADO

A favor de:

JOHNY JAVIER PAMBABAY CALERO; SERGIO BAUZ OLVERA; ANA BELEN NIETO LIBRERO; MARIA PURIFICACION GALINDO VILLARDON

Por su contribución en la modalidad de COMUNICACIÓN en el "V CONGRESO INTERNACIONAL EN CONTEXTOS PSICOLÓGICOS, EDUCATIVOS Y DE LA SALUD", con el título:

META-ANÁLISIS DEL ESTUDIO DE LA PRECISIÓN DIAGNÓSTICA EMPLEANDO MODELIZACIÓN DE EFECTOS ALEATORIOS

Además, dicha aportación está PUBLICADA en el libro de Actas del V Congreso Internacional en Contextos Psicológicos, Educativos y de la Salud. Volumen I, con ISBN: 978-84-09-06266-9 y Depósito Legal: AL 2570-2018.

Dicho congreso se ha celebrado durante los días 21, 22 y 23 de noviembre de 2018 en Madrid, con una duración de 25 horas y organizado por el Grupo de Investigación SEJ-473 de la Universidad de Almería y la Association University of Scientific Formation Psychology and Education Research (Entidad sin fin de lucro inscrita con el número 6372 de la Sección 1, del Registro de Asociaciones de la Junta de Andalucía, y en sus estatutos sociales constan expresamente la Formación y la Investigación). Dicha actividad cuenta con la Resolución Favorable de Reconocimiento de Interés Científico-Sanitario concedida por el Ministerio de Sanidad, Servicios Sociales e Igualdad (CSV: 2FB3V-7S4YZ-PNGHV-BQKJV).

Madrid, a 23 de noviembre de 2018

Presidente del Congreso



Dr. José Jesús Gázquez Linares

Presidenta de la Association University of Scientific Formation Psychology and Education Research

Scinfoper



Dña. María del Carmen Pérez Fuentes



B.4. Certificado de Estancia de Investigación, Primer Evaluador



Universidad Veracruzana
Dirección General de Investigaciones
Centro de Estudios de Opinión y Análisis.
Región Xalapa

DRA. MA. PURIFICACIÓN GALINDO VILLARDÓN

Vicerrectora de Postgrado y Planes Especiales en Ciencias de la Salud
Universidad de Salamanca
Salamanca, España

Por este medio se hace constar que el Máster Johny Pambabay Calero con Pasaporte N° 091771755, estudiante del Doctorado en Estadística Multivariante Aplicada en la Universidad de Salamanca, España, concluyó su estancia de investigación en el Laboratorio para el Análisis de Información Generada a través de Redes Sociales en Internet (LARSI) perteneciente al Centro de Estudios de Opinión y Análisis (CEOA) de la Universidad Veracruzana (UV). La estancia se realizó desde Julio hasta septiembre 2019, al término de la estancia, se cumplieron los siguientes objetivos.

- Proponer procedimientos alternativos a las soluciones clásicas de meta-análisis, en ausencia de prueba oro, específicamente diseñadas para los casos en los que la eficacia es muy alta pero los falsos positivos también.
- Escritura de artículo científico con relación a datos del Portal Web correspondiente a la Procuraduría de Justicia de la Ciudad de México.
- Aplicar técnicas estadísticas punteras a datos no estructurados (información en internet).
- Analizar la problemática asociada a los meta-análisis de pruebas diagnósticas utilizadas en enfermedades con baja prevalencia, donde no es posible utilizar la exactitud para valorar la prueba.
- Participación de formación en la Primera Escuela AME de Ciencia de Datos de la Universidad Autónoma de México.
- Dictado de curso de Estadística Aplicada al Programa de Especialización de Estudios de Opinión de la Universidad Veracruzana.

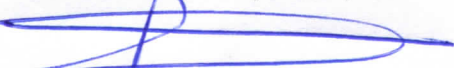
La estancia de investigación fue coordinada por el LARSI de la UV, a cargo del Dr. Carlos Adolfo Piña García y supervisado por el Dr. Rubén Flores González, Investigador de Tiempo Completo de la UV.

Estos procesos formativos permitieron consolidar la formación investigativa al Máster Johny Pambabay en las líneas de investigación que el Centro de Estudios de Opinión y Análisis maneja, es decir:

- Desarrollo y Aplicaciones de Métodos Estadísticos,
- Análisis de Representaciones Sociales en Medios de Comunicación

Se extiende la presente para los fines que se le confieran, a los treinta días del mes de septiembre del dos mil diecinueve. Sin otro particular, quedo a sus órdenes para cualquier aclaración al respecto.

"Lis de Veracruz: Arte, Ciencia, Luz"


DR. RUBÉN FLORES GONZÁLEZ
Investigador de Tiempo Completo UV



**B.5. Certificado de Estancia de Investigación, Segundo
Evaluador**



Universidad Veracruzana
Dirección General de Investigaciones
Centro de Estudios de Opinión y Análisis
Región Xalapa

DRA. MA. PURIFICACIÓN GALINDO VILLARDÓN

Vicerrectora de Postgrado y Planes Especiales en Ciencias de la Salud
Universidad de Salamanca
Salamanca, España

Por este medio se hace constar que el Máster Johny Pambabay Calero con Pasaporte N° 091771755, estudiante del Doctorado en Estadística Multivariante Aplicada en la Universidad de Salamanca, España, concluyó su estancia de investigación en el Laboratorio para el Análisis de Información Generada a través de Redes Sociales en Internet (LARSI) perteneciente al Centro de Estudios de Opinión y Análisis (CEOA) de la Universidad Veracruzana (UV). La estancia se realizó desde Julio hasta septiembre 2019, al término de la estancia, se cumplieron los siguientes objetivos.

- Proponer procedimientos alternativos a las soluciones clásicas de meta-análisis, en ausencia de prueba oro, específicamente diseñadas para los casos en los que la eficacia es muy alta pero los falsos positivos también.
- Escritura de artículo científico con relación a datos del Portal Web correspondiente a la Procuraduría de Justicia de la Ciudad de México.
- Aplicar técnicas estadísticas punteras a datos no estructurados (información en internet).
- Analizar la problemática asociada a los meta-análisis de pruebas diagnósticas utilizadas en enfermedades con baja prevalencia, donde no es posible utilizar la exactitud para valorar la prueba.
- Participación de formación en la Primera Escuela AME de Ciencia de Datos de la Universidad Autónoma de México.
- Dictado de curso de Estadística Aplicada al Programa de Especialización de Estudios de Opinión de la Universidad Veracruzana.

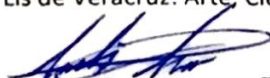
La estancia de investigación fue coordinada por el LARSI de la UV, a cargo del Dr. Carlos Adolfo Piña García y supervisado por el Dr. Rubén Flores González, Investigador de Tiempo Completo de la UV.

Estos procesos formativos permitieron consolidar la formación investigativa al Máster Johny Pambabay en las líneas de investigación que el Centro de Estudios de Opinión y Análisis maneja, es decir:

- Desarrollo y Aplicaciones de Métodos Estadísticos,
- Análisis de Representaciones Sociales en Medios de Comunicación

Se extiende la presente para los fines que se le confieran, a los treinta días del mes de septiembre del dos mil diecinueve. Sin otro particular, quedo a sus órdenes para cualquier aclaración al respecto.

“Lis de Veracruz: Arte, Ciencia, Luz”


DR. CARLOS ADOLFO PIÑA GARCIA
Coordinador LARSI



Francisco Sarabia No.
100A, Col. José Cardel,
C.P. 91030

Teléfono
8 40 35 30; 8 40 58 68;
8 40 11 71 y 8 40 29 25

Conmutador
8 42 17 00

Conm. Ext.
Ext. 14301

Correo electrónico
ceoa_ver@uv.mx

**B.6. Asistencia a la Primera Escuela AME: Ciencia de
Datos del 23 al 25 de Septiembre de 2019 en la
UNAM, CDMX, México**

La Asociación Mexicana de Estadística otorga la presente

Constancia

a

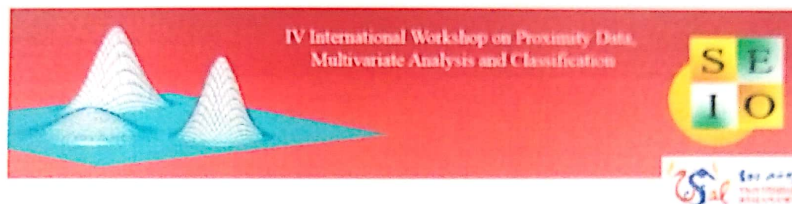
Johnny Javier Pambabay Calero

por su asistencia a la Primera Escuela AME: Ciencia de Datos
del 23 al 25 de septiembre de 2019 en la UNAM, CDMX, México.



Dr. Ramsés H. Mena
Presidente de la AME

B.7. Certificate of Presentation: For presentation the work “Hierarchical modeling and copulas for the study of diagnostic accuracy” in the IV International Workshop on Proximity Data, Multivariate Analysis and Classification on 25 and 26 of April 2019 at University Salamanca



Certificate of Presentation

To

Johny Pambabay-Calero, Sergio Bauz-Olvera, Ana Nieto-Librero and Puri Galindo-Villardón

For presenting the work

Hierarchical modeling and copulas for the study of diagnostic accuracy

In the

IV International Workshop on Proximity Data, Multivariate Analysis and Classification

Organized by Group of Multivariate Analysis and Classification of the SEIO

on 25 and 26 of April, 2019

at Department of Statistics (Universidad de Salamanca)

*José Luis Vicente Villardón
President of the Organizing Committee*

B.8. Presentación Ponencia Oral en el ciclo de Conferencia en el marco del Programa de Estadística Multivariante organizado en conjunto entre ESPOL y USAL, en Guayaquil del 21 al 29 de septiembre 2018



ESPOL

"Impulsando la sociedad del conocimiento"

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS

Certifica que el trabajo:

*Contribuciones al meta análisis de pruebas
diagnósticas en enfermedades de baja prevalencia*

Escrito por:

Johny Pambabay Calero

Fue presentado en el Ciclo de Conferencias en el marco del Programa de Estadística Multivariante organizado en conjunto con la Universidad de Salamanca, celebrado en Guayaquil del 21 al 29 de septiembre 2018.

Guayaquil, 01 de Octubre de 2018


Mgtr. Francisca Flores Nicolalde
DECANA DE LA FCNM (S)



Ph. D. Ma. Purificación Galindo Villardón
INSTRUCTORA
UNIVERSIDAD DE SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL
Vicerrectorado de Postgrado y Planes
Especiales en Ciencias de la Salud

FCNM
Facultad de Ciencias Naturales
y Matemáticas

B.9. Constancia como Ponente en el 33 Foro Nacional de Estadística y 13 Congreso Latinoamericano de Sociedades Estadística, celebrado del 1 al 5 de Octubre de 2018 en Guadalajara, Jalisco, México



33 FORO NACIONAL DE ESTADÍSTICA | CONGRESO LATINOAMERICANO DE SOCIEDADES DE ESTADÍSTICA 13

Otorgan la presente

CONSTANCIA


a
Johny Javier Pambabay Calero


Por su participación como Ponente con el tema

Modelización para meta-análisis de test diagnósticos de respuestas binarias: un enfoque desde cópulas y modelos jerárquicos, en el

33 Foro Nacional de Estadística y 13 Congreso Latinoamericano de Sociedades de Estadística

Celebrado del 1 al 5 de Octubre de 2018 en Guadalajara, Jalisco, México


Dr. Humberto Gutiérrez Pulido
Departamento de Matemáticas
CUCEI, Universidad de Guadalajara


Dr. Ramsés Humberto Mena Chávez
Presidente de la Asociación Mexicana de Estadística



B.10. Participación como Expositor en el "VI Encuentro Iberoamericano de Biometría", realizado en Sangolquí-Ecuador, el 15, 16 y 17 de Noviembre del 2017



VI Encuentro Iberoamericano
de Biometría
Quito 2017

La Universidad de las Fuerzas Armadas - ESPE,
Departamento de Ciencias de la Vida y de la Agricultura
otorgan el presente:

CERTIFICADO

A:


JOHNY PAMBABAY

Por su participación como Expositor en el "VI ENCUENTRO IBERAMERICANO DE BIOMETRIA",
realizado en Sangolquí - Ecuador, el 15, 16 y 17 de Noviembre del 2017.




E.M. Hugo Pérez Vaca
CORRECTOR DE INVESTIGACION


Dr. Omar Ruiz Barzola
PRESIDENTE IBS ECUADOR


Ing. Victor Hugo Abrit, PhD.
DIRECTOR DPTO. CIENCIAS DE LA VIDA

N°: 1027

- B.11. Comunicación de Simposio en el III Congreso Nacional de Psicología, realizado en Oviedo-Asturias-España, desde 3 al 7 de julio 2017**



III Congreso Nacional
de Psicología
3 al 7 Julio 2017 · Oviedo · Asturias

Certificado

El Consejo General de Colegios Oficiales de Psicólogos de España (COP) certifica que

Johnny Pambabay Calero

ha presentado la siguiente Comunicación de Simposio en el III Congreso Nacional de Psicología:

Modelización Bayesiana en meta análisis de pruebas diagnósticas: estado del arte

Johnny Pambabay Calero, Ana Nieto Librero, M^a Purificación Galindo-Villardón, Ana Belén Sánchez,
Johnny Pambabay Calero y Sergio Bauz Olvera

III Congreso Nacional de Psicología

3 al 7 Julio 2017 · Oviedo · Asturias · España

Francisco Santolaya Ochando

Presidente del Consejo General de Colegios Oficiales de Psicólogos

José Muñiz Fernández

Presidente del Comité Científico

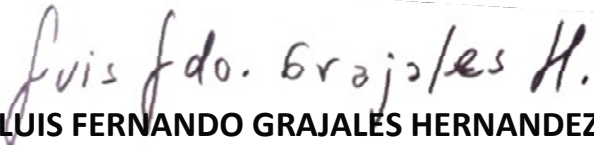
B.12. Ponente de Simposio en el XXVI Simposio de Estadística 2016, realizado del 8 al 12 de agosto de 2016 en Sicelejo, Sucre, Colombia

XXVI SIMPOSIO DE ESTADÍSTICA 2016

CERTIFICACIÓN

Johny J. Pambabay Calero con documento de identidad 917717555, participó en las actividades académicas del Simposio en calidad de **PONENTE** y presentó la(s) ponencia(s) titulada(s) “**Meta-análisis de tests diagnósticos para evaluación de enfermedades con baja prevalencia: un estudio comparativo desde HSROC, MADA y STATA**”, durante el evento que se realizó del 8 al 12 de agosto de 2016 en el Hotel Malibú.

Para constancia se firma en Sincelejo, Sucre, Colombia, el día 12 de agosto de 2016.



LUIS FERNANDO GRAJALES HERNANDEZ
Coordinador Simposio de Estadística

Bibliografía

- [1] Akbani, R., Kwek, S., and Japkowicz, N. (2004). Applying support vector machines to imbalanced datasets. In *European conference on machine learning*, pages 39–50. Springer. [315](#), [316](#)
- [2] Alexandersson, A. et al. (2004). Graphing confidence ellipses: An update of ellip for stata 8. *Stata Journal*, 4:242–256. [91](#)
- [3] Arbyn, M., Roelens, J., Simoens, C., Buntinx, F., Paraskevaidis, E., Martin-Hirsch, P. P., and Prendiville, W. J. (2013). Human papillomavirus testing versus repeat cytology for triage of minor cytological cervical lesions. *The Cochrane Library*. [203](#)
- [4] Arends, L., Hamza, T., Van Houwelingen, J., Heijnenbrok-Kal, M., Hunink, M., and Stijnen, T. (2008). Bivariate random effects meta-analysis of roc curves. *Medical Decision Making*, 28(5):621–638. [141](#), [149](#)
- [5] Avilés Merens, R., Morales Morejón, M., Sao Avilés, A., and Cañedo Andalia, R. (2004). La colaboración cochrane en cuba: Parte vii. los metanálisis.: aproximaciones útiles para su comprensión. *Acimed*, 12(4):1–1. [11](#)
- [6] Bachmann, L. M., Puhan, M. A., Ter Riet, G., and Bossuyt, P. M. (2006). Sample sizes of studies on diagnostic accuracy: Literature survey. *British Medical Journal*, 332(7550):1127–1129. [308](#)
- [7] Bae, M. H., Wu, T., and Pan, R. (2010). Mix-ratio sampling: Classifying multiclass imbalanced mouse brain images using support vector machine. *Expert Systems with Applications*, 37(7):4955–4965. [316](#)

BIBLIOGRAFÍA

- [8] Baker, F. (1992). Item response theory. statistics: Textbooks and monographs n. 129. *Item Response Theory: STATISTICS: Textbooks and Monographs n. 129*. [67](#), [71](#)
- [9] Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of mathematical psychology*, 12(4):387–415. [61](#)
- [10] Beck, J. R. and Shultz, E. K. (1986). The use of relative operating characteristic (roc) curves in test performance evaluation. *Archives of pathology & laboratory medicine*, 110(1):13–20. [61](#)
- [11] Begg, C. B. and Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, pages 1088–1101. [176](#)
- [12] Begg, C. B. and McNeil, B. (1988). Assessment of radiologic tests: control of bias and other design considerations. *Radiology*, 167(2):565–569. [50](#)
- [13] Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Moher, D., Rennie, D., De Vet, H. C., and Lijmer, J. G. (2003). The stard statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Annals of internal medicine*, 138(1):W1–12. [86](#)
- [14] Boyko, E. J. (1994). Ruling out or ruling in disease with the most sensitive or specific diagnostic test: Short cut or wrong turn? *Medical Decision Making*, 14(2):175–179. [78](#)
- [15] Burgueño, M., García Bastos, J., and González Buitrago, J. (1995). Las curvas roc en la evaluación de las pruebas diagnósticas. *Medicina clínica*, 104(17):661–670. [34](#), [35](#), [46](#)
- [16] Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1). [200](#)
- [17] Casares, L. M. (2001). Metaanálisis: claves para interpretar una herramienta de investigación controvertida. *Hipertensión y Riesgo Vascular*, 18(5):232–240. [11](#)
- [18] Castillo, I. I. (2009). Evaluación de resultados clínicos (ii): Las medidas de la significación clínica o los tamaños del efecto. *Norte de Salud Mental*, 8(34):10. [21](#)

- [19] Castro, C. L., Carvalho, M. A., and Braga, A. P. (2009). An improved algorithm for svms classification of imbalanced data sets. In *International Conference on Engineering Applications of Neural Networks*, pages 108–118. Springer. [316](#)
- [20] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357. [253](#)
- [21] Chew, V. (1966). Confidence, prediction, and tolerance regions for the multivariate normal distribution. *Journal of the American Statistical Association*, 61(315):605–617. [91](#)
- [22] Chu, H., Chen, S., and Louis, T. A. (2009). Random effects models in a meta-analysis of the accuracy of two diagnostic tests without a gold standard. *Journal of the American Statistical Association*, 104(486):512–523. [81](#), [84](#)
- [23] Chu, H. and Cole, S. R. (2006). Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of clinical epidemiology*, 59(12):1331–1332. [149](#)
- [24] Chu, H., Guo, H., and Zhou, Y. (2010). Bivariate random effects meta-analysis of diagnostic studies using generalized linear mixed models. *Medical Decision Making*, 30(4):499–508. [8](#)
- [25] Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151. [108](#)
- [26] Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, 10(1):101–129. [94](#), [137](#)
- [27] Cohen, G., Hilario, M., Sax, H., Hugonnet, S., and Geissbuhler, A. (2006). Learning from imbalanced data in surveillance of nosocomial infection. *Artificial intelligence in medicine*, 37(1):7–18. [316](#)
- [28] Cohen, W. W. (1995). Fast effective rule induction. In *Machine learning proceedings 1995*, pages 115–123. Elsevier. [276](#)
- [29] Cook, D. J., Mulrow, C. D., and Haynes, R. B. (1997). Systematic reviews: synthesis of best evidence for clinical decisions. *Annals of internal medicine*, 126(5):376–380. [xix](#), [2](#)

BIBLIOGRAFÍA

- [30] Cooper, H. M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of educational research*, 52(2):291–302. [16](#)
- [31] de Dios, J. G. and Santamaría, A. B. (2007). Revisión sistemática y metanálisis (i): Conceptos básicos. *Evidencias en pediatría*, 3(4):21. [19](#)
- [32] Deeks, J. J. (2001). Systematic reviews of evaluations of diagnostic and screening tests. *Bmj*, 323(7305):157–162. [49](#), [55](#), [86](#), [129](#), [136](#)
- [33] Deeks, J. J., Macaskill, P., and Irwig, L. (2005). The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of clinical epidemiology*, 58(9):882–893. [176](#), [177](#), [178](#)
- [34] DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, pages 837–845. [61](#)
- [35] Dendukuri, N., Hadgu, A., and Wang, L. (2009). Modeling conditional dependence between diagnostic tests: a multiple latent variable model. *Statistics in medicine*, 28(3):441–461. [84](#)
- [36] Dendukuri, N. and Joseph, L. (2001). Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics*, 57(1):158–167. [84](#)
- [37] Dendukuri, N., Schiller, I., Joseph, L., and Pai, M. (2012). Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. *Biometrics*, 68(4):1285–1293. [83](#), [84](#), [182](#)
- [38] DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3):177–188. [94](#), [136](#)
- [39] Devillé, W. L., Buntinx, F., Bouter, L. M., Montori, V. M., De Vet, H. C., Van der Windt, D. A., and Bezemer, P. D. (2002). Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC medical research methodology*, 2(1):9. [86](#)
- [40] Doebler, P. (2016). Mada: meta-analysis of diagnostic accuracy. 2015. URL <http://CRAN.R-project.org/package=mada>. R package version 0.5, 7. [5](#)
- [41] Douglas, J. (1993). Confidence regions for parameter pairs. *The American Statistician*, 47(1):43–45. [91](#)

-
- [42] Drummond, C. and Holte, R. (2005). Learning to live with false alarms. *Data Mining Methods for Anomaly Detection*, page 21. [302](#)
- [43] Dudewicz, E. J. and Mishra, S. (1988). *Modern mathematical statistics*. John Wiley & Sons, Inc. [87](#), [91](#)
- [44] Dukic, V. and Gatsonis, C. (2003). Meta-analysis of diagnostic test accuracy assessment studies with varying number of thresholds. *Biometrics*, 59(4):936–946. [86](#)
- [45] Dwamena, B. (2007). Midas: computational and graphical routines for meta-analytical integration of diagnostic accuracy studies in stata. *Ann Arbor: Division of Nuclear Medicine, Department of Radiology, University of Michigan Medical School*. [5](#)
- [46] Egger, M. and Smith, G. D. (1997). Meta-analysis. potentials and promise. *BMJ: British Medical Journal*, 315(7119):1371. [12](#)
- [47] Egger, M. and Smith, G. D. (1998). Bias in location and selection of studies. *BMJ: British Medical Journal*, 316(7124):61. [176](#)
- [48] Egger, M., Smith, G. D., and Schneider, M. (2001). *Systematic reviews of observational studies*. Wiley Online Library. [6](#)
- [49] Elazmeh, W., Japkowicz, N., and Matwin, S. (2006). Evaluating misclassifications in imbalanced data. In Furnkranz, J and Scheffer, T and Spiliopoulou, M, editor, *MACHINE LEARNING: ECML 2006, PROCEEDINGS*, volume 4212 of *Lecture Notes in Computer Science*, pages 126–137, HEIDELBERGER PLATZ 3, D-14197 BERLIN, GERMANY. Google; Humboldt Univ Berlin; Deutsch Forsch Gemeinsch; KD-Ubiq; European Off Aerosp Res & Dev; USAF Res Lab; STRATO AG; IBM; PASCAL; Mach Learning, SPRINGER-VERLAG BERLIN. 17th European Conference on Machine Learning (ECML 2006), Berlin, GERMANY, SEP 18-22, 2006. [297](#)
- [50] Elsalamony, H. A. and Elsayad, A. M. (2013). Bank direct marketing based on neural network and c5. 0 models. *International Journal of Engineering and Advanced Technology (IJEAT)*, 2(6). [270](#)
- [51] Embrechts, P., Mcneil, E., and Straumann, D. (1999). Correlation: pitfalls and alternatives. In *Risk Magazine*. Citeseer. [99](#)
-

BIBLIOGRAFÍA

- [52] Fagan, T. (1975). Nomogram for bayes's theorem. *New England Journal of Medicine*, 293:257. [43](#), [178](#)
- [53] Fahey, M. T., Irwig, L., and Macaskill, P. (1995). Meta-analysis of pap test accuracy. *American journal of epidemiology*, 141(7):680–689. [55](#)
- [54] Farlie, D. J. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47(3/4):307–323. [110](#)
- [55] Fawcett, T. (2004). ROC graphs: Notes and practical considerations for researchers. *Machine learning*, 31(1):1–38. [302](#)
- [56] Fisher, N. and Switzer, P. (2001). Graphical assessment of dependence: Is a picture worth 100 tests? *The American Statistician*, 55(3):233–239. [155](#)
- [57] Frank, M. J. (1979). On the simultaneous associativity off (x, y) and $x+y- f(x, y)$. *Aequationes mathematicae*, 19(1):194–226. [109](#)
- [58] Fürnkranz, J. and Widmer, G. (1994). Incremental reduced error pruning. In *Machine Learning Proceedings 1994*, pages 70–77. Elsevier. [276](#)
- [59] Galindo-Villardón, M. (1991). Meta-análisis: el uso de los métodos estadísticos en revisiones de estudios de investigación relacionados. *Cirugía española*, 54(3):201–3. [3](#), [11](#), [15](#), [20](#)
- [60] Galindo-Villardón, M. P. (1986). Una alternativa de representacion simultanea: Hj-biplot. *Qüestió: quaderns d'estadística i investigació operativa*, 10(1):13–23. [5](#)
- [61] Gasparrini, A., Armstrong, B., and Kenward, M. (2012). Multivariate meta-analysis for non-linear and other multi-parameter associations. *Statistics in medicine*, 31(29):3821–3839. [142](#)
- [62] Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409. [75](#)
- [63] Gelman, A. and Rubin, D. B. (1992). A single series from the gibbs sampler provides a false sense of security. *Bayesian statistics*, 4:625–631. [84](#)
- [64] Gisbert, J. P. and Bonfill, X. (2004). ¿ cómo realizar, evaluar y utilizar revisiones sistemáticas y metaanálisis? *Gastroenterología y hepatología*, 27(3):129–149. [14](#)

-
- [65] Glas, A. S., Lijmer, J. G., Prins, M. H., Bossel, G. J., and Bossuyt, P. M. (2003). The diagnostic odds ratio: a single indicator of test performance. *Journal of clinical epidemiology*, 56(11):1129–1135. [49](#), [129](#), [136](#), [202](#), [223](#)
- [66] Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, 5(10):3–8. [21](#)
- [67] Gluud, C. and Gluud, L. L. (2005). Evidence based diagnostics. *BMJ: British Medical Journal*, 330(7493):724. [86](#)
- [68] Green, B. F. and Hall, J. A. (1984). Quantitative methods for literature reviews. *Annual review of psychology*, 35(1):37–54. [21](#)
- [69] Griner, P. F., Mayewski, R. J., Mushlin, A. I., and Greenland, P. (1981). Selection and interpretation of diagnostic tests and procedures. *Annals of internal medicine*, 94(4 II). [49](#)
- [70] Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292):698–707. [110](#)
- [71] Guyatt, G. H., Tugwell, P., Feeny, D., Haynes, R., and Drummond, M. (1986). A framework for clinical evaluation of diagnostic technologies. *CMAJ: Canadian Medical Association Journal*, 134(6):587. [49](#)
- [72] Guyon, I. (1997). A scaling law for the validation-set training-set size ratio. *AT&T Bell Laboratories*, pages 1–11. [321](#)
- [73] Habbema, J. D. F., Eijkemans, R., Krijnen, P., and Knottnerus, J. A. (2002). Analysis of data on the accuracy of diagnostic tests. *The evidence base of clinical diagnosis*, pages 117–144. [49](#)
- [74] Hanley, J. A. (1998). Receiver operating characteristic (roc) curves. *Encyclopedia of biostatistics*. [61](#)
- [75] Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36. [61](#)
- [76] Harbord, R. M., Deeks, J. J., Egger, M., Whiting, P., and Sterne, J. A. (2006a). A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*, 8(2):239–251. [71](#), [81](#), [87](#), [141](#)

BIBLIOGRAFÍA

- [77] Harbord, R. M., Egger, M., and Sterne, J. A. (2006b). A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in medicine*, 25(20):3443–3457. [176](#)
- [78] Hasselblad, V. and Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological bulletin*, 117(1):167. [6](#), [49](#)
- [79] He, H. and Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, 21(9):1263–1284. [315](#)
- [80] Hedges, L. V. (1990). Olkin, I. (1985). statistical methods for meta-analysis. *Orlando, FL: Academic Press*. [92](#), [94](#)
- [81] Held, L. (2008). *Methoden der statistischen Inferenz: Likelihood und Bayes*. Spektrum, Akad. Verlag. [106](#)
- [82] Herbrich, R. (2001). *Learning kernel classifiers: theory and algorithms*. MIT press. [315](#)
- [83] Higgins, J. and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, 21(11):1539–1558. [95](#), [169](#)
- [84] Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*, 327(7414):557. [137](#)
- [85] Hilden, J. (1991). The area under the roc curve and its competitors. *Medical Decision Making*, 11(2):95–101. [50](#), [61](#)
- [86] Hoffman, M. D. and Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623. [200](#)
- [87] Holling, H., Böhning, W., and Böhning, D. (2012b). Meta-analysis of diagnostic studies based upon sroc-curves: a mixed model approach using the lehmann family. *Statistical Modelling*, 12(4):347–375. [138](#)
- [88] Holte, R. C. (1993). Very simple classification rules perform well on most commonly used datasets. *Machine learning*, 11(1):63–90. [275](#)
- [89] Honest, H. and Khan, K. S. (2002). Reporting of measures of accuracy in systematic reviews of diagnostic literature. *BMC health services research*, 2(1):4. [49](#)

-
- [90] Hornik, K., Leisch, F., and Zeileis, A. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of DSC*, volume 2, pages 1–1. 200
- [91] Hsieh, C.-J., Chang, K.-W., Lin, C.-J., Keerthi, S. S., and Sundararajan, S. (2008). A dual coordinate descent method for large-scale linear svm. In *Proceedings of the 25th international conference on Machine learning*, pages 408–415. ACM. 312, 313
- [92] Imam, T., Ting, K. M., and Kamruzzaman, J. (2006). z-svm: An svm for improved classification of imbalanced data. In *Australasian Joint Conference on Artificial Intelligence*, pages 264–273. Springer. 315, 316
- [93] Irwig, L., Macaskill, P., Glasziou, P., and Fahey, M. (1995). Meta-analytic methods for diagnostic test accuracy. *Journal of clinical epidemiology*, 48(1):119–130. 49, 53, 55
- [94] Irwig, L., Tosteson, A. N., Gatsonis, C., Lau, J., Colditz, G., Chalmers, T. C., and Mosteller, F. (1994). Guidelines for meta-analyses evaluating diagnostic tests. *Annals of internal medicine*, 120(8):667–676. 49
- [95] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer. 292, 295
- [96] Kang, P. and Cho, S. (2006). Eus svms: Ensemble of under-sampled svms for data imbalance problems. In *International Conference on Neural Information Processing*, pages 837–846. Springer. 316
- [97] Kardaun, J., Kardaun, O., et al. (1990). Comparative diagnostic performance of three radiological procedures for the detection of lumbar disk herniation. *Contributions to a rational diagnosis and treatment of lumbar disk herniation*, page 49. 49, 67
- [98] Khan, K. S., Bachmann, L. M., and ter Riet, G. (2003). Systematic reviews with individual patient data meta-analysis to evaluate diagnostic tests. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 108(2):121–125. 86
- [99] Knottnerus, J. A. and Muris, J. W. (2003). Assessment of the accuracy of diagnostic tests: the cross-sectional study. *Journal of clinical epidemiology*, 56(11):1118–1128. 49
- [100] Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). Bivariate and trivariate normal distributions. *Continuous Multivariate Distributions: Models and Applications, Volume 1, Second Edition*, pages 251–348. 68

BIBLIOGRAFÍA

- [101] Kriston, L., Hoülzel, L., Weiser, A.-K., Berner, M. M., and Hürter, M. (2008). Meta-analysis: are 3 questions enough to detect unhealthy alcohol use? *Annals of internal medicine*, 149(12):879–888. [127](#), [308](#)
- [102] Kubat, M., Matwin, S., et al. (1997). Addressing the curse of imbalanced training sets: one-sided selection. In *ICML*, volume 97, pages 179–186. Nashville, USA. [253](#)
- [103] Kuss, O., Hoyer, A., and Solms, A. (2014). Meta-analysis for diagnostic accuracy studies: a new statistical model using beta-binomial distributions and bivariate copulas. *Statistics in Medicine*, 33(1):17–30. [104](#)
- [104] Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, pages 159–174. [18](#)
- [105] Le, C. T. (2006). A solution for the most basic optimization problem associated with an roc curve. *Statistical methods in medical research*, 15(6):571–584. [138](#)
- [106] Leeflang, M. M., Deeks, J. J., Gatsonis, C., and Bossuyt, P. M. (2008). Systematic reviews of diagnostic test accuracy. *Annals of internal medicine*, 149(12):889–897. [97](#), [180](#)
- [107] Li, B., Hu, J., and Hirasawa, K. (2008a). An improved support vector machine with soft decision-making boundary. In *Proceedings of the 26th IASTED International Conference on Artificial Intelligence and Applications*, pages 40–45. ACTA Press. [316](#)
- [108] Li, B., Hu, J., and Hirasawa, K. (2008b). Support vector machine classifier with whm offset for unbalanced data. *JACIII*, 12(1):94–101. [316](#)
- [109] Li, P., Qiao, P.-L., and Liu, Y.-C. (2008c). A hybrid re-sampling method for svm learning from imbalanced data sets. In *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, volume 2, pages 65–69. IEEE. [316](#)
- [110] Light, R. and Smith, P. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard educational review*, 41(4):429–471. [21](#)
- [111] Lijmer, J. G., Bossuyt, P. M., and Heisterkamp, S. H. (2002). Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Statistics in medicine*, 21(11):1525–1537. [55](#), [86](#)

-
- [112] Lijmer, J. G., Mol, B. W., Heisterkamp, S., Bossel, G. J., Prins, M. H., van der Meulen, J. H., and Bossuyt, P. M. (1999). Empirical evidence of design-related bias in studies of diagnostic tests. *Jama*, 282(11):1061–1066. [49](#)
- [113] Lipsey, M. W. (2009). Identifying interesting variables and analysis opportunities. *The handbook of research synthesis and meta-analysis*, 2:147–158. [93](#)
- [114] Littenberg, B. and Moses, L. E. (1993). Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. *Medical Decision Making*, 13(4):313–321. [49](#), [86](#)
- [115] Liu, Y., An, A., and Huang, X. (2006). Boosting prediction accuracy on imbalanced datasets with svm ensembles. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 107–118. Springer. [316](#)
- [116] Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). The bugs project: evolution, critique and future directions. *Statistics in medicine*, 28(25):3049–3067. [125](#)
- [117] Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs—a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, 10(4):325–337. [125](#)
- [118] Macaskill, P. (2004). Empirical bayes estimates generated in a hierarchical summary roc analysis agreed closely with those of a full bayesian analysis. *Journal of clinical epidemiology*, 57(9):925–932. [90](#)
- [119] Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R., and Takwoingi, Y. (2010). Cochrane handbook for systematic reviews of diagnostic test accuracy. *Version 0.9. 0. London: The Cochrane Collaboration*. [83](#)
- [120] Macaskill, P., Walter, S. D., and Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in medicine*, 20(4):641–654. [176](#)
- [121] Martín Rodero, H. (2011). Medicina basada en la evidencia y meta-análisis: una aproximación conceptual, histórica y metodológica. [2](#), [3](#), [11](#), [12](#), [13](#), [15](#), [17](#), [19](#), [20](#), [21](#), [27](#), [28](#)
- [122] Martín-Vallejo, F. J. (1996). *Métodos estadísticos en meta-análisis*. PhD thesis, Universidad de Salamanca. [13](#), [20](#), [22](#)
-

BIBLIOGRAFÍA

- [123] Martínez-González, M., De Irala-Estévez, J., and Guillén-Grima, F. (1999). What is an odds ratio? *Medicina clínica*, 112(11):416. [23](#), [25](#)
- [124] McCullagh, P. (1980). Regression models for ordinal data. *Journal of the royal statistical society. Series B (Methodological)*, pages 109–142. [71](#)
- [125] McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton university press. [99](#)
- [126] Menke, J. (2010). Bivariate random-effects meta-analysis of sensitivity and specificity with sas proc glimmix. *Methods of information in medicine*, 49(01):54–64. [98](#)
- [127] Metz, C. E. (1978). Basic principles of roc analysis. In *Seminars in nuclear medicine*, volume 8, pages 283–298. Elsevier. [37](#)
- [128] Metz, C. E., Herman, B. A., and Shen, J.-H. (1998). Maximum likelihood estimation of receiver operating characteristic (roc) curves from continuously-distributed data. *Statistics in medicine*, 17(9):1033–1053. [61](#)
- [129] Midgette, A. S., Stukel, T. A., and Littenberg, B. (1993). A meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. *Medical Decision Making*, 13(3):253–257. [49](#), [53](#)
- [130] Molinero, L. (2008). Metaanálisis, una guía rápida para lectores y usuarios. *Hipertensión*, 25(3):108–120. [28](#)
- [131] Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitt, Math, Statist.*, 8:234–235. [110](#)
- [132] Moses, L. E., Shapiro, D., and Littenberg, B. (1993). Combining independent studies of a diagnostic test into a summary roc curve: data-analytic approaches and some additional considerations. *Statistics in medicine*, 12(14):1293–1316. [5](#), [8](#), [49](#), [50](#), [53](#), [54](#), [55](#), [56](#), [59](#), [60](#), [81](#), [86](#)
- [133] Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media. [100](#), [101](#), [102](#)
- [134] Nguyen, H. M., Cooper, E. W., and Kamei, K. (2009). Borderline over-sampling for imbalanced data classification. In *Proceedings: Fifth International Workshop on Computational Intelligence & Applications*, volume 2009, pages 24–29. IEEE SMC Hiroshima Chapter. [316](#)

-
- [135] Nikoloulopoulos, A. K. (2015). Copularemada: Copula mixed effect models for bivariate and trivariate meta-analysis of diagnostic test accuracy studies. [8](#)
- [136] Nishimura, K., Sugiyama, D., Kogata, Y., Tsuji, G., Nakazawa, T., Kawano, S., Saigo, K., Morinobu, A., Koshiha, M., Kuntz, K. M., et al. (2007). Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. *Annals of internal medicine*, 146(11):797–808. [xii](#), [xii](#), [xvii](#), [xvii](#), [xvii](#), [xvii](#), [xvii](#), [xvii](#), [xvii](#), [xvii](#), [xvii](#), [227](#), [228](#), [229](#), [230](#), [231](#), [233](#), [234](#), [238](#), [239](#), [244](#), [245](#), [361](#), [363](#), [365](#), [367](#), [369](#)
- [137] Normand, S.-L. T. (1999). Tutorial in biostatistics meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in medicine*, 18(3):321–359. [12](#), [22](#), [23](#), [26](#), [27](#)
- [138] Ntzoufras, I. (2011). *Bayesian modeling using WinBUGS*, volume 698. John Wiley & Sons. [188](#)
- [139] Ortega, Z., Martín-Vallejo, J., Mencía, A., Galindo-Villardón, M. P., and Pérez-Mellado, V. (2016). Introducing meta-partition, a useful methodology to explore factors that influence ecological effect sizes. *PloS one*, 11(7):e0158624. [28](#)
- [140] Oxman, A. D., Cook, D. J., Guyatt, G. H., Bass, E., Brill-Edwards, P., Browman, G., Detsky, A., Farkouh, M., Gerstein, H., Haines, T., et al. (1994). Users' guides to the medical literature: Vi. how to use an overview. *Jama*, 272(17):1367–1371. [2](#), [14](#)
- [141] Patrick, D. L., Cheadle, A., Thompson, D. C., Diehr, P., Koepsell, T., and Kinne, S. (1994). The validity of self-reported smoking: a review and meta-analysis. *American journal of public health*, 84(7):1086–1093. [144](#), [146](#)
- [142] Paul, M., Riebler, A., Bachmann, L., Rue, H., and Held, L. (2010). Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested laplace approximations. *Statistics in medicine*, 29(12):1325–1339. [98](#)
- [143] Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., and Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *Jama*, 295(6):676–680. [176](#)
- [144] Phillips, B., Stewart, L. A., and Sutton, A. J. (2010). 'cross hairs' plots for diagnostic meta-analysis. *Research Synthesis Methods*, 1(3-4):308–315. [134](#), [171](#)

BIBLIOGRAFÍA

- [145] Pita Fernández, S. and Pértegas Díaz, S. (2003). Pruebas diagnósticas: Sensibilidad y especificidad. *Cad Aten Primaria*, 10:120–4. [32](#), [39](#)
- [146] Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004). Gllamm manual. [149](#)
- [147] Rabe-Hesketh, S., Skrondal, A., Pickles, A., et al. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2(1):1–21. [149](#)
- [148] Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M., and Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of clinical epidemiology*, 58(10):982–990. [5](#), [6](#), [8](#), [49](#), [70](#), [81](#), [86](#), [91](#), [125](#), [135](#), [141](#), [149](#)
- [149] Riley, R. D., Abrams, K., Lambert, P., Sutton, A., and Thompson, J. (2007a). An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in medicine*, 26(1):78–97. [96](#), [98](#), [149](#), [223](#)
- [150] Riley, R. D., Abrams, K. R., Sutton, A. J., Lambert, P. C., and Thompson, J. R. (2007b). Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Medical Research Methodology*, 7(1):3. [149](#), [223](#)
- [151] Riley, R. D., Thompson, J. R., and Abrams, K. R. (2007c). An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics*, 9(1):172–186. [149](#)
- [152] Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin*, 86(3):638. [4](#)
- [153] Rosenthal, R. (1991). *Meta-analytic procedures for social research*, volume 6. Sage. [16](#)
- [154] Rousseeuw, P. J., Ruts, I., and Tukey, J. W. (1999). The bagplot: a bivariate boxplot. *The American Statistician*, 53(4):382–387. [155](#)
- [155] Rücker, G., Schwarzer, G., and Carpenter, J. (2008). Arcsine test for publication bias in meta-analyses with binary outcomes. *Statistics in medicine*, 27(5):746–763. [176](#)
- [156] Rüschenendorf, L. (2009). On the distributional transform, sklar’s theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, 139(11):3921–3927. [101](#)

- [157] Rutter, C. and Gatsonis, C. (1995). Regression methods for meta-analysis of diagnostic test data. *Academic radiology*, 2:S48–56. [55](#)
- [158] Rutter, C. M. and Gatsonis, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in medicine*, 20(19):2865–2884. [5](#), [6](#), [8](#), [49](#), [50](#), [54](#), [71](#), [72](#), [73](#), [74](#), [76](#), [77](#), [81](#), [82](#), [86](#), [90](#), [125](#), [135](#), [141](#), [183](#), [190](#)
- [159] Sackett, D. (1991). The selection of diagnostic tests. *Clinical epidemiology*, pages 47–57. [xv](#), [1](#), [3](#)
- [160] Sackett, D., Haynes, R., Guyatt, G., and Tugwell, P. (1985). Clinical epidemiology: A basic science for clinical medicine. little brown and co. *Inc., Boston*. [3](#)
- [161] Sackett, D. L. and Haynes, R. B. (2002). The architecture of diagnostic research. *Bmj*, 324(7336):539–541. [49](#)
- [162] Sadatsafavi, M., Shahidi, N., Marra, F., FitzGerald, M. J., Elwood, K. R., Guo, N., and Marra, C. A. (2010). A statistical method was used for the meta-analysis of tests for latent tb in the absence of a gold standard, combining random-effect and latent-class methods to estimate test accuracy. *Journal of clinical epidemiology*, 63(3):257–269. [81](#)
- [163] Sáenz, A. and Ausejo, M. (2000). Guía para leer e interpretar una revisión sistemática. *Rev Soc Mad Med Fam Comunit*, 2:29–36. [1](#)
- [164] Salzberg, S. L. (1994). C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993. *Machine Learning*, 16(3):235–240. [271](#)
- [165] Scheidler, J., Hricak, H., Kyle, K. Y., Subak, L., and Segal, M. R. (1997). Radiological evaluation of lymph node metastases in patients with cervical cancer: a meta-analysis. *Jama*, 278(13):1096–1101. [183](#)
- [166] Schiller, I., Dendukuri, N., and Schiller, M. I. (2015). Package ‘hsroc’. [5](#), [6](#), [84](#)
- [167] Schuetz, G. M., Zacharopoulou, N. M., Schlattmann, P., and Dewey, M. (2010). Meta-analysis: noninvasive coronary angiography using computed tomography versus magnetic resonance imaging. *Annals of internal medicine*, 152(3):167–177. [xii](#), [xvii](#), [xvii](#), [248](#), [249](#), [250](#), [370](#), [371](#), [373](#), [375](#), [377](#)

BIBLIOGRAFÍA

- [168] Schwarzer, G., Antes, G., and Schumacher, M. (2002). Inflation of type i error rate in two statistical tests for the detection of publication bias in meta-analyses with binary outcomes. *Statistics in medicine*, 21(17):2465–2477. [176](#)
- [169] Schwarzer, G., Antes, G., and Schumacher, M. (2007). A test for publication bias in meta-analysis with sparse binary data. *Statistics in medicine*, 26(4):721–733. [176](#)
- [170] Schweizer, B. and Sklar, A. (2011). *Probabilistic metric spaces*. Courier Corporation. [101](#)
- [171] Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *The annals of statistics*, pages 879–885. [101](#)
- [172] Seqc (2010). Evaluación de Pruebas Diagnósticas. <http://bit.ly/1U5kAdA/>. [Online; acceso 20-Mayo-2016]. [44](#), [45](#)
- [173] Shanahan, J. G. and Roma, N. (2003). Improving svm text classification performance through threshold adjustment. In *European Conference on Machine Learning*, pages 361–372. Springer. [316](#)
- [174] Siadaty, M. S. and Shu, J. (2004). Proportional odds ratio model for comparison of diagnostic tests in meta-analysis. *BMC medical research methodology*, 4(1):27. [86](#)
- [175] Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231. [100](#), [106](#)
- [176] Song, F., Khan, K. S., Dinnes, J., and Sutton, A. J. (2002). Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *International journal of epidemiology*, 31(1):88–95. [176](#)
- [177] Sorribas, A., March, J., and Trujillano, J. (2002). A new parametric method based on s-distributions for computing receiver operating characteristic curves for continuous diagnostic tests. *Statistics in medicine*, 21(9):1213–1235. [50](#)
- [178] Stengel, D., Bauwens, K., Sehouli, J., Ekkernkamp, A., and Porzsolt, F. (2003). Original paper: A likelihood ratio approach to meta-analysis of diagnostic studies. *Journal of Medical Screening*, 10(1):47–51. [179](#)
- [179] Sutton, A. J., Abrams, K. R., Jones, D. R., Jones, D. R., Sheldon, T. A., and Song, F. (2000). Methods for meta-analysis in medical research. [6](#)

-
- [180] Sutton, A. J. and Higgins, J. (2008). Recent developments in meta-analysis. *Statistics in medicine*, 27(5):625–650. [164](#)
- [181] Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240(4857):1285–1293. [172](#)
- [182] Takwoingi, Y. and Deeks, J. (2008). Metadas: a sas macro for meta-analysis of diagnostic accuracy studies. In *Methods for evaluating medical tests symposium. University of Birmingham*. Citeseer. [8](#)
- [183] Takwoingi, Y., Guo, B., Riley, R. D., and Deeks, J. J. (2017). Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Statistical methods in medical research*, 26(4):1896–1911. [308](#)
- [184] Tang, Y., Zhang, Y.-Q., Chawla, N. V., and Krasser, S. (2008). Svms modeling for highly imbalanced classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):281–288. [316](#)
- [185] Tatsioni, A., Zarin, D. A., Aronson, N., Samson, D. J., Flamm, C. R., Schmid, C., and Lau, J. (2005). Challenges in systematic reviews of diagnostic technologies. *Annals of internal medicine*, 142(12_Part_2):1048–1055. [86](#)
- [186] Tosteson, A. N. A. and Begg, C. B. (1988). A general regression methodology for roc curve estimation. *Medical Decision Making*, 8(3):204–215. [71](#)
- [187] Van Houwelingen, H. C., Arends, L. R., and Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in medicine*, 21(4):589–624. [67](#), [68](#), [141](#), [149](#)
- [188] Van Houwelingen, H. C., Zwinderman, K. H., and Stijnen, T. (1993). A bivariate approach to meta-analysis. *Statistics in medicine*, 12(24):2273–2284. [67](#), [149](#)
- [189] Vapnik, V. (1998). Statistical learning theory. john wiley&sons. Inc., New York. [315](#)
- [190] Verde, P. E. (2010). Meta-analysis of diagnostic test data: A bivariate bayesian modeling approach. *Statistics in medicine*, 29(30):3088–3102. [98](#)
- [191] Vilariño, F., Spyridonos, P., Vitrià, J., and Radeva, P. (2005). Experiments with svm and stratified sampling with an imbalanced problem: detection of intestinal contractions.

BIBLIOGRAFÍA

- In *International Conference on Pattern Recognition and Image Analysis*, pages 783–791. Springer. [316](#)
- [192] Vivaracho, C. E. (2006). Improving svm training by means of ntil when the data sets are imbalanced. In *International Symposium on Methodologies for Intelligent Systems*, pages 111–120. Springer. [316](#)
- [193] Walter, S. (2002). Properties of the summary receiver operating characteristic (sroc) curve for diagnostic test data. *Statistics in medicine*, 21(9):1237–1256. [49](#), [54](#), [56](#), [57](#), [59](#), [60](#), [61](#), [62](#), [64](#), [65](#), [66](#), [86](#)
- [194] Walter, S. and Jadad, A. (1999). Meta-analysis of screening data: a survey of the literature. *Statistics in medicine*, 18(24):3409–3424. [49](#), [50](#), [81](#)
- [195] Wang, B. X. and Japkowicz, N. (2010). Boosting support vector machines for imbalanced data sets. *Knowledge and information systems*, 25(1):1–20. [316](#)
- [196] Wang, H. and Zheng, H. (2008). An improved support vector machine for the classification of imbalanced biological datasets. In *International Conference on Intelligent Computing*, pages 63–70. Springer. [316](#)
- [197] Waske, B., Benediktsson, J. A., and Sveinsson, J. R. (2009). Classifying remote sensing data with support vector machines and imbalanced training data. In *International Workshop on Multiple Classifier Systems*, pages 375–384. Springer. [316](#)
- [198] Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594. [202](#)
- [199] Weiss, R. E. (1996). Bayesian model checking with applications to hierarchical models. *Unpublished Technical Report*. [77](#)
- [200] Whiting, P., Harbord, R., and Kleijnen, J. (2005). No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC medical research methodology*, 5(1):19. [165](#)
- [201] Whiting, P., Rutjes, A., Dinnes, J., Reitsma, J., Bossuyt, P., Kleijnen, J., et al. (2004a). *Development and validation of methods for assessing the quality of diagnostic accuracy studies*. Gray Pub. [165](#)

- [202] Whiting, P., Rutjes, A. W., Reitsma, J. B., Bossuyt, P. M., and Kleijnen, J. (2003). The development of quadas: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC medical research methodology*, 3(1):25. [165](#)
- [203] Whiting, P., Rutjes, A. W., Reitsma, J. B., Glas, A. S., Bossuyt, P. M., and Kleijnen, J. (2004b). Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Annals of internal medicine*, 140(3):189–202. [86](#)
- [204] Whiting, P. F., Weswood, M. E., Rutjes, A. W., Reitsma, J. B., Bossuyt, P. N., and Kleijnen, J. (2006). Evaluation of quadas, a tool for the quality assessment of diagnostic accuracy studies. *BMC medical research methodology*, 6(1):9. [165](#)
- [205] Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann. [276](#)
- [206] Wu, G. and Chang, E. Y. (2005). Kba: Kernel boundary alignment considering imbalanced data distribution. *IEEE Transactions on Knowledge & Data Engineering*, (6):786–795. [316](#)
- [207] Xue-Kun Song, P. (2000). Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics*, 27(2):305–320. [107](#)
- [208] Yang, C.-Y., Wang, J., Yang, J.-S., and Yu, G.-D. (2008). Imbalanced svm learning with margin compensation. In *International Symposium on Neural Networks*, pages 636–644. Springer. [316](#)
- [209] Yu, T., Debenham, J., Jan, T., and Simoff, S. (2006). Combine vector quantization and support vector machine for imbalanced datasets. In *IFIP International Conference on Artificial Intelligence in Theory and Practice*, pages 81–88. Springer. [316](#)