

UNIVERSIDAD DE SALAMANCA
Departamento de Estadística y Matemática Aplicadas



MÉTODOS ESTADÍSTICOS
EN
META-ANÁLISIS

Javier Martín Vallejo

Directora: M^a Purificación Galindo Villardón

1995

**METODOS ESTADISTICOS
EN
META-ANALISIS**

Memoria que para optar al Grado de Doctor por
el Departamento de Estadística y Matemática
Aplicadas de la Universidad de Salamanca,
presenta:

Javier Martín Vallejo.

Salamanca
1995



Dpto. de Estadística y Matemática Aplicadas
Universidad de Salamanca

M^a PURIFICACION GALINDO VILLARDON

*Profesora Titular del Area de Estadística e I.O.
de la Universidad de Salamanca*

CERTIFICA: Que **D. Javier Martín Vallejo**, Licenciado en Biología, ha realizado en el Departamento de Estadística y Matemática Aplicadas de la Universidad de Salamanca, bajo su dirección, el trabajo que para optar al Grado de Doctor, presenta con el título: "*Métodos Estadísticos en Meta-Análisis*", y para que conste, firma el presente certificado en Salamanca el 21 de Septiembre de 1995.

A mis Padres,
y a M^a José.

AGRADECIMIENTOS

Quiero expresar mi más sincero agradecimiento a la Profesora **Dra. M^a Purificación Galindo Villardón**, directora de este trabajo, por su dedicación continua, por sus consejos, no sólo científicos, y muy especialmente por la confianza que ha depositado en mi en todo momento. Su labor, en mi formación, ha sido fundamental. Ha conseguido convertirme en un Estadístico seguro de mi mismo.

A mis compañeros del **Departamento de Estadística y Matemática Aplicadas** por su ayuda, sus consejos, las discusiones estadísticas y "no estadísticas" a la hora del café, por los momentos buenos que hemos compartido y también, porqué no, por los malos, que han sido muchos y que han servido para unirnos de una forma indescriptible. Especialmente a José Luis, Inmaculada, M^a José, Carmelo, Ricardo, Antonio y Santiago porque de todos ellos hay algo en este trabajo.

Al profesor **Dr. José Antonio Castro Posada** por ser nuestro primer nexo con el meta-análisis.

A **Dra. Nora Donaldson** y **Dr. Mike Pidcock**, profesores de la *School of Computing in Mathematical Sciences* (Oxford Brookes University) y al **Dr. Max Parmar** del *Medical Research Council* (Cambridge), por sus sabios consejos, que han sido para mi de un incalculable valor.

A **Dr. Dirk Bauwens** del *Institute of Nature Conservation* de Hasselt por preocuparse en todo momento de enviarme, sin demora, la pocas referencias de meta-análisis que encontré en el mundo de la Ecología, así como por su continuo apoyo.

A **mis padres**, que día a día me han apoyado en mi carrera profesional y mi vida personal.

A **M^a José García-Oliva** que continuamente me ha apoyado en mis decisiones y ha soportado los momentos difíciles de este trabajo. Gracias por todo ello y por tantas cosas más....

Por último, un agradecimiento especial a tres personas gracias a las cuales conocí la estadística: Puri, Inmaculada y José Luis. Su proyecto es ahora también mi proyecto y el de un grupo de jóvenes investigadores que formamos una gran familia. Con mucho cariño, Gracias.

*It is clear that meta-analysis,
like rock and roll, is here to stay.*
IYENGAR y GREENHOUSE, 1988

Indice

INTRODUCCION 1**CAPITULO 1. META-ANÁLISIS**

1.1. ¿QUE ES UN META-ANALISIS?	7
1.2. RESEÑA HISTORICA	9
1.3. CLASIFICACION DE LOS METODOS META-ANALITICOS .	14
1.4. OBJETIVO DEL META-ANALISIS	18
1.5. ESTADO ACTUAL DEL META-ANALISIS	
1.5.1.- INTRODUCCION.....	20
1.5.2.- ANALISIS BIBLIOGRAFICO	20
1.6. ETAPAS DE UN META-ANALISIS	26

CAPITULO 2. FORMULACION DEL PROBLEMA

2.1. DEFINICION DE LOS OBJETIVOS	28
2.2. PROTOCOLO	29
2.3. CALIDAD DE LOS ESTUDIOS	31

CAPITULO 3. BUSQUEDA DE LA LITERATURA

3.1. INTRODUCCION	33
3.2. POBLACION OBJETIVO Y POBLACION ACCESIBLE	35
3.3. FUENTES DE INFORMACION	36
3.4. ETAPAS EN LA BUSQUEDA DE LA LITERATURA	39

CAPITULO 4. CODIFICACION DE LOS ESTUDIOS

4.1. DESCRIPCION DE LAS CARACTERISTICAS DIFERENCIALES DE LOS ESTUDIOS	42
4.2. ESTUDIO DE LA FIABILIDAD EN LA CODIFICACION	45
4.3. CONTRIBUCIONES A LA FASE DE CODIFICACION DESDE UNA PERSPECTIVA MULTIVARIANTE	
4.3.1.-ANALISIS MULTIVARIANTE DE LOS RESULTADOS DE LA CODIFICACION	48
4.3.2.- EL METODO BIPLLOT COMO TECNICA DE CARACTERIZACION DE ESTUDIOS E IDENTIFICACION DE JUECES DISCORDANTES	50
4.4. APLICACION PRACTICA DE LAS APORTACIONES TEORICAS EN LA FASE DE CODIFICACION	56

CAPITULO 5. MEDIDA DE LOS RESULTADOS

5.1. INTRODUCCION	65
5.2. "VOTE COUNTING"	66
5.3. EL NIVEL DE SIGNIFICACION	67
5.4. EL TAMAÑO DEL EFECTO	68
5.4.1.-LA DIFERENCIA MEDIA TIPIFICADA	69
5.4.2.- COMPARACIONES DE ESTIMADORES	73
5.4.3.- EL COEFICIENTE DE CORRELACION	85
5.4.4.- TAMAÑO DEL EFECTO EN TABLAS 2x2	88
5.4.5.- ESTIMADORES NO PARAMETRICOS	93

CAPITULO 6. ANALISIS DE LOS RESULTADOS

6.1. INTRODUCCION	99
6.2. METODOS GRAFICOS EN ESTUDIOS META-ANALITICOS	100
6.2.1.-GRAFICOS DE DISTRIBUCION DE FRECUENCIAS	101
6.2.2.- DIAGRAMA DE DISPERSION	101
6.2.3.- GRAFICOS DE ODDS-RATIO (O RIESGO RELATIVO)	106
6.2.4.- GRAFICO RADIAL	108
6.2.5.-GRAFICOS DE DISTRIBUCION DE FRECUENCIAS	101
6.3. METODOS MULTIVARIANTES PARA LA EXPLORACION DE LA VARIABILIDAD DE LOS ESTUDIOS	112
6.3.1.-METODOS MULTIVARIANTES GRAFICOS	113
6.3.1.i. METODOS BASADOS EN LA ROTACION DEL ESPACIO INICIAL	114
6.3.1.ii. METODOS BASADOS EN LA OBTENCION DE UNA IMAGEN EUCLIDEA	116
6.3.1.iii. CALIDAD DE LA REPRESENTACION EN DIMENSION REDUCIDA .	117
6.3.1.iv. APLICACION DE UNO DE LOS METODOS MULTIVARIANTES A UN CASO REAL	125
6.3.2.- METODOS GRAFICOS MULTIVARIANTES	131
6.3.2.i. ESTRELLAS O POLIGONOS	131
6.3.2.ii. CARAS DE CHERNOFF	134
6.3.2.iii. CURVAS DE ANDREWS	139

6.4. COMBINACION DE NIVELES DE PROBABILIDAD	142
6.4.1.-TECNICAS DE COMBINACION	
6.4.1.i. METODO DE TIPPETT	143
6.4.1.ii. METODO DE FISHER	144
6.4.1.iii. METODO DE LA SUMA DE NIVELES DE PROBABILIDAD	145
6.4.1.iv. METODO DEL CONTRASTE DEL NIVEL p MEDIO	146
6.4.1.v. METODO DE LA SUMA DE PUNTUACIONES Z	146
6.4.1.vi. METODO DE LA SUMA DE PUNTUACIONES Z PONDERADAS	150
6.4.1.vii. METODO DEL CONTRASTE DE LA PUNTUACION Z MEDIA	151
6.4.1.viii. METODO DE WINER	152
6.4.1.ix. METODO LOGIT	153
6.4.2.- CRITICA A LA COMBINACION DE NIVELES DE PROBABILIDAD	153
6.5. INTEGRACION DEL TAMAÑO DEL EFECTO	156
6.5.1.-METODOS GENERALES DE INTEGRACION DEL TAMAÑO DEL EFECTO	
6.5.1.i. MODELO DE EFECTOS FIJOS	157
6.5.1.ii. MODELO DE EFECTOS ALEATORIOS	163
6.5.2.-INTEGRACION DE LA DIFERENCIA MEDIA TIPIFICADA	167
6.5.3.-INTEGRACION DEL COEFICIENTE DE CORRELACION	171
6.5.4.-INTEGRACION DEL TAMAÑO DEL EFECTO EN TABLAS 2x2	172
6.5.4.i. INTEGRACION DE ODDS-RATIO	157
6.5.4.ii. INTEGRACION DE DIFERENTES PROPORCIONES	181
6.6. HETEROGENEIDAD DE LOS TAMAÑOS DEL EFECTO	187
6.7. UNA CONTRIBUCION A LA BUSQUEDA DE LA HETEROGENEIDAD EN LA FASE DE ANALISIS DE RESULTADOS DESDE UNA PERSPECTIVA MULTIVARIANTE	
6.7.1.-EL METODO JERARQUICO PARA LA BUSQUEDA DE LA HETROGENEIDAD	
6.7.1.i. PLANTEAMIENTO TEORICO	194

6.7.1.ii. ELECCION DE LOS NIVELES DE PROBABILIDAD	203
6.7.1.iii. MODELO VECTORIAL PARA RELACIONAR LA ORDENACION DE LOS ESTUDIOS	207

Pág

6.8. COMPARACION DE LAS TECNICAS DE HEDGES Y OLKIN PARA LA BUSQUEDA DE LA HETEROGENEIDAD, Y DE LA ALTERNATIVA MULTIVARIANTE PROPUESTA

6.8.1.-INTRODUCCION	211
6.8.1.-APLICACION DEL METODO DE BUSQUEDA DE LA HETEROGENEIDAD	213

CAPITULO 7. PROBLEMAS DEL META-ANALISIS

7.1. EL SESGO DE PUBLICACION	225
7.1.1.-SOLUCIONES PROPUESTAS	227
7.1.1.i. METODO DEL "NUMERO DE FALLOS NULOS"	227
7.1.1.ii. METODOS BASADOS EN LOS MUESTREOS TRUNCADOS	230
7.1.1.iii. METODO DE MUESTREO INVARIANTE EN BANCOS DE DATOS	232
7.1.2.-COMPARACION DE LOS PROCEDIMIENTOS DE ESTIMACION DEL SESGO DE PUBLICACION	232
7.2. PROBLEMA DE LA HETEROGENEIDAD	240
7.3. PROBLEMA DE INDEPENDENCIA DE LOS TAMAÑOS DEL EFECTO	242

CAPITULO 8. UNA CONTRIBUCION AL META-ANALISIS DE DATOS MULTIVARIANTES DESDE UNA PERSPECTIVA MULTIVARIANTE

8.1. INTRODUCCION	244
8.2. METODOS PROCRUSTES	
8.2.1.-PROCRUSTES PARA 2 CONFIGURACIONES	246
8.2.2.- PROCRUSTES PARA MAS DE 2 CONFIGURACIONES.....	249
8.2.2.i. MODELOS PINDIS	250
8.2.2.ii. MODELOS DE PONDERACION DE LAS DIMENSIONES	251
8.2.2.iii. MODELOS DE PONDERACION DE LOS ESTIMULOS	252

8.3. METODOS DE COMPARACION DE SUBESPACIOS	
8.3.1.-INTEGRACION DE LA INFORMACION DE 2 GRUPOS	254
8.3.2.- GENERALIZACION AL CASO EN QUE SEA NECESARIO INTEGRAR LA INFORMACION DE MAS DE 2 GRUPOS	259
8.3.3.- COMPARACION E INTEGRACION DE SUBESPACIOS OBTENIDOS TRAS UN ANALISIS DE COMPONENTES PRINCIPALES A 2 MATRICES DE DATOS REALES	263
8.3.4.- COMPARACION E INTEGRACION DE VARIOS SUBESPACIOS OBTENIDOS TRAS UN ANALISIS DE COMPONENTES PRINCIPALES: DATOS REALES	272
 CONCLUSIONES	 287
 BIBLIOGRAFIA	 290

INTRODUCCION

Uno de los pilares fundamentales sobre los que descansa la investigación es la acumulación del conocimiento. Los avances realizados en la actualidad, en el mundo científico, están avalados por los miles de estudios precedentes que han intentado describir los distintos procesos.

La recopilación bibliográfica sobre la información preliminar a un estudio es fundamental para comprender, en toda su extensión, la problemática que está presente en el Area de Investigación.

La vía formal para la evaluación y síntesis de la información ha estado cubierta por las "revisiones narrativas" en las cuales el revisor recopila la información que considera relevante en su campo de investigación, y a partir de la cual extrae las conclusiones generales en el tema de estudio.

Sin embargo, muchos autores han descrito las limitaciones de este tipo de revisiones: son subjetivas, no pueden considerar resultados cuantitativos, el revisor se ve obligado a procesar una gran cantidad de información...

Estas limitaciones impulsaron el desarrollo de las revisiones cuantitativas que culminó con la propuesta de GLASS en 1976, denominada **Meta-análisis**.

El auge de los métodos meta-analíticos es incuestionable, lo que no parece tan claro es su origen.

OLKIN escribió en 1990: *"Se ha alcanzado un punto donde psicólogos, sociólogos, educadores y médicos opinan que el meta-análisis comenzó a desarrollarse en su particular disciplina"*.

Las publicaciones relacionadas con el tema crecen a tal velocidad que es prácticamente imposible mantener actualizada la información, crecen y crecen las aplicaciones sobre todo en Psicología, Medicina y CC. Sociales, sin embargo, los avances metodológicos no son verdaderamente notables.

Es destacable el hecho de que ese crecimiento no se ha producido en especialidades tan importantes de la Biología como la Ecología, la Zoología o la Botánica. No obstante, en 1995, ARNQVIST y WOOSTER publican una reflexión sobre la importancia de las técnicas meta-analíticas como base para la acumulación de información cuantitativa en Ecología y Evolución.

Esta asincronía, ya detectada en el momento de comenzar este trabajo, hace ahora 5 años, fue uno de los motores que me impulsó a aceptar el reto de abordar una investigación de esta envergadura.

Es bien conocido que en estas Areas de la Biología el investigador necesita el manejo de técnicas estadísticas y sin embargo es poco frecuente encontrar equipos multidisciplinarios, cosa que es bastante frecuente, por ejemplo, en los ensayos clínicos en donde estadísticos y expertos interactúan en los diferentes proyectos. Quizás ésto ha motivado el desarrollo explosivo en Medicina y el bajo impacto en esta otra parte de las Ciencias de la Vida.

Otra laguna importante había sido detectada en el momento de comenzar este trabajo: los estudios meta-analíticos han estado dirigidos a la integración de

resultados univariantes y bajo un enfoque univariante

Sin embargo, es precisamente en las ramas de la Biología antes citada, donde los métodos multivariantes han alcanzado mayor desarrollo.

El auge de los métodos meta-analíticos, la falta de trabajos meta-analíticos en Biología, y el contar con el apoyo de un equipo de Bioestadísticos especializados en Análisis Multivariante fueron el caldo de cultivo en el que se gestó mi decisión de aceptar como tema de Tesis, a pesar de mi condición de Biólogo, "Métodos Estadísticos en Meta-análisis".

El trabajo ha sido duro pero verdaderamente fascinante y ha contribuido de manera decisiva a esta metamorfosis que me ha transformado de biólogo aficionado a la Estadística, en estadístico amante de la Biología.

Es paradójico también, que el único estudio encontrado en la bibliografía donde se intente la integración de representaciones multidimensionales, pertenezca al campo de la Ecología (WARWICK y CLARKE, 1993). Sin embargo, no es realmente un meta-análisis de técnicas multivariantes.

Esta puede ser otra razón por la que los biólogos no han sentido atracción por el Meta-análisis: la metodología está desarrollada para tratar de integrar resultados de estudios basados en análisis univariantes; sin embargo, el análisis de sus datos requiere, de manera generalizada, estudios multivariantes.

Las ideas sobre las que se han desarrollado las aportaciones teóricas desde la perspectiva multivariante se han extraído de la literatura especializada ya que no se ha encontrado ni una sola referencia en las búsquedas con la palabra clave meta-análisis (o equivalentes).

Las ideas de GABRIEL y GALINDO, se han incorporado a la fase codificación y análisis de resultados del meta-análisis , las ideas de MORGAN y SONQUIST en la fase de análisis de resultados, las ideas de KRANOWSKY y las ideas de GOWER nos han servido para abordar el problema del meta-análisis desde una perspectiva multivariante y suponen la primera contribución en este sentido.

CAPITULO 1

META-ANALISIS

1.1. ¿QUE ES UN META-ANALISIS?:

GLASS (1976) definió el meta-análisis como "*el conjunto de métodos cuantitativos para la comparación y combinación de resultados a partir de estudios analíticos individuales y similares*". Dicha definición la resumió con una reseña mas concisa como "*el análisis de los análisis*" (GLASS y cols. 1981). Son múltiples las definiciones de Meta-análisis encontradas en la literatura especializada de las diferentes áreas. Sin embargo, todas ellas presentan un mismo punto de partida referido a la **agregación de resultados cuantitativos extraídos de trabajos científicos individuales**. The National Library of Medicine en la inclusión de dicho termino en la lista de palabras indexadas especifica la definición de Meta-análisis como (ver DICKERSIN y cols., 1990) :

"Un método cuantitativo de combinación de resultados procedentes de estudios independientes (usualmente obtenidos a partir de la literatura publicada) que sintetiza los sumarios y conclusiones con aplicación principalmente en las áreas de investigación y medicina. Pueden ser utilizados para evaluar la efectividad, nuevos proyectos de estudios, etc., El método consiste en cuatro pasos: una profunda revisión de la literatura; calculo de un tamaño del efecto para cada estudio; determinación de un tamaño del efecto conjunto a partir de la combinación ponderada de los tamaños del efecto individuales; cálculo del sesgo de publicación (número de estudios no publicados con conclusiones opuestas necesarios para cambiar los resultados publicados) para valorar la confianza del tamaño del efecto conjunto"

Otras acepciones conocidas para denominar este tipo de revisiones son: **revisión**

cuantitativa, (GREEN y HALL, 1984), **integración de investigaciones**, (WALBERG y HAERTEL, 1980), **análisis cuantitativo de dominios de investigación** (ROSENTHAL, 1980), **pooling** (PETO y cols., 1977) u **overview**, estas dos últimas expresiones son más conocidas en el campo de la Medicina y más específicamente para la combinación de resultados derivados de ensayos clínicos aleatorizados.

1.2. RESEÑA HISTORICA

La primera aproximación de combinación de la evidencia de diferentes pruebas científicas, fue los experimentos realizados en el campo de la física por Heyl, en 1930, dirigidos a la determinación de la constante gravitacional. En estos estudios, Heyl combina la información obtenida a partir de los resultados de diferentes experimentos mediante medias ponderadas (OLKIN, 1990).

Sin embargo, los primeros estudios metodológicos formales en agrupación de resultados se deben a investigadores de la talla de FISHER (1932), PEARSON (1933) y TIPPETT (1931). Estos científicos plantean la combinación de los varios niveles de probabilidad procedentes de diferentes experimentos en un intento por obtener resultados globales a partir de test individuales.

En años sucesivos, el planteamiento de agregación alcanza un mayor desarrollo en el campo de la experimentación agrícola. La imposibilidad de controlar diferentes factores (climáticos, físicos, etc.) en una misma zona de experimentación condujo a la combinación de múltiples resultados recogidos en diferentes lugares bajo un mismo diseño. COCHRAN (1937), YATES y COCHRAN (1938) estudian la combinación de los resultados de un particular tratamiento en diversos centros agrícolas con el objetivo de controlar dichos factores. En posteriores estudios (COCHRAN, 1943, 1954) profundizan en las técnicas estadísticas para la agrupación de experimentos, así como, en su filosofía.

MANTEL y HAENSZEL (1959) proponen un método para la combinación de los resultados de tablas 2x2 procedentes de la estratificación según una variable explicativa relacionada con las variables medidas. Esta aportación, junto a las de PETO y PETO (1972), en relación a la modificación del "*logrank test*", y PETO y cols. (1977) proporcionaron la bases metodológicas para su posterior uso en revisiones cuantitativas de Medicina. STJERNWARD (1974) y CHALMERS y cols. (1977) son los primeros investigadores que aplican el procedimiento propuesto por Mantel-Haenszel en la combinación de los resultados de varios ensayos clínicos.

Paralelamente, en el área de Psicología, GHISELLI (1949, 1955) y BLOOM (1964) promedian varios coeficientes de correlación extraídos de diversos estudios sobre un tópico concreto. WILKINSON (1951) interpreta, a partir de la distribución binomial, múltiples resultados cuantitativos resumidos en dos categorías: resultados significativos y no significativos. Esta aportación establece las bases para el procedimiento de "*vote-counting*" utilizado como complemento cuantitativo de las revisiones clásicas. Una contribución fundamental al desarrollo del Meta-análisis procede de COHEN (1962) que asienta las bases del desarrollo de las técnicas de acumulación para resultados cuantitativos: define el tamaño del efecto y postula la agrupación de resultados con el fin de incrementar la potencia estadística de los análisis.

LIGHT y SMITH (1971) proponen una alternativa a los métodos narrativos de integración. Se considera como el primer intento metodológico formal en la combinación y acumulación de los resultados extraídos de un conjunto de experimentos y uno de los antecedentes más relevantes en el planteamiento de los métodos Meta-analíticos. Este procedimiento se puede resumir en cuatro etapas:

- Incorporar al análisis sólo aquellos estudios que son metodológicamente replicables.
- Recuperar los datos originales de los cuales se han obtenido los resultados en los estudios.
- Determinar agrupaciones naturales en las cuales los datos puedan ser reagrupados.
- Por último, investigar las diferencias entre las agregaciones reconocidas para comparar medias, varianzas, relaciones entre variables independientes y dependientes e interacciones sujetos-tratamiento y poder explicar las variaciones entre los diferentes estudios.

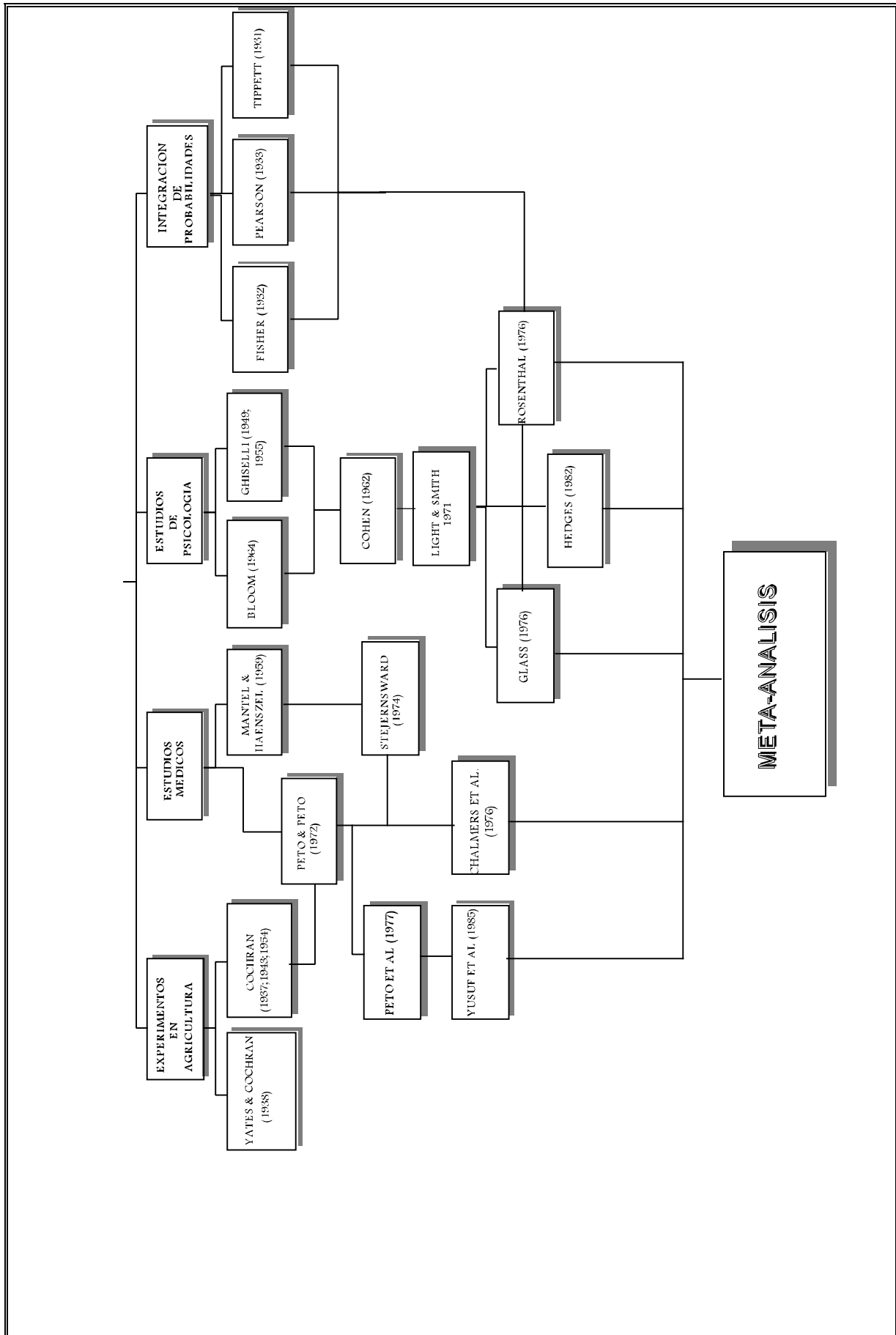
Este tipo de revisiones podrían ser englobadas dentro de los **análisis secundarios** donde es necesario conocer los datos originales de cada estudio de investigación.

Estos hechos son el caldo de cultivo para que GLASS (1976) acuñara el nombre de **meta-análisis** en el campo de las Ciencias Sociales. Este término se ha generalizado en las diferentes áreas del mundo científico. SACKS y cols. (1985) introducen la acepción de Meta-análisis en el campo de la Medicina.

A pesar de la utilización más habitual de otras expresiones ya admitidas en la literatura científica médica para describir la integración de resultados, como **overview** o **pooling**, la acepción de meta-análisis se ha extendido en esta área. PETO (1987a) enjuicia dicha expresión como poco afortunada para describir los métodos que aglutinan los datos procedentes de diferentes investigaciones y defiende la expresión "**overview**".

Como demanda al incremento de publicaciones de investigación médica que incluyen en el título el término de Meta-análisis "The National Library of Medicine" introduce la palabra meta-análisis en 1989 como (Medical Subject Heading).

En el esquema siguiente se ha conformado un árbol que pretende resumir los estudios más importantes que consideramos claves en el desarrollo del meta-análisis.



1.3. CLASIFICACIÓN DE LOS METODOS META-ANALITICOS

Todos los antecedentes históricos en la integración de resultados cuantitativos desembocaron en un conjunto de métodos estadísticos dentro de un nuevo marco formal, en el mundo de las revisiones. Sin duda, los trabajos metodológicos de GLASS en el mundo de las ciencias sociales y PETO en el campo de la Medicina, son los que han impulsado la práctica y con ella, el desarrollo de las técnicas meta-analíticas, cada vez más sofisticadas.

El meta-análisis introdujo, a partir de la idea inicial propuesta por Glass, un conjunto de modificaciones en función de las diferentes críticas lanzadas a los planteamientos teóricos y prácticos. BANGERT-DROWNS (1986) describe cinco aproximaciones de los métodos meta-analíticos en el campo de las ciencias sociales que pueden ser ampliadas si se consideran las aportaciones y modificaciones en Medicina. La clasificación presentada se ha estructurado en función de los objetivos, criterios de selección de estudios, unidad de análisis y métodos de análisis propuestos y se ha ampliado considerando las aportaciones y modificaciones en otras áreas:

- (GLASS, 1976): Se propone conocer la información que proporciona la literatura sobre un tema seleccionado. Busca toda la información disponible independientemente de su calidad, calcula los tamaños del efecto para cada variable dependiente del estudio y compara dichos tamaños en relación a un conjunto de variables

preestablecidas.

- (MANSFIELD y BUSSE, 1977). Presenta modificaciones en relación a tres puntos:
 - Restringe el objetivo propuesto a la efectividad del tratamiento estudiado.
 - Aplica criterios estrictos a la hora de seleccionar los estudios.
 - Obtiene de cada estudio un único valor.

- (ROSENTHAL, 1976; COOPER, 1979). Se distingue del anterior en la incorporación de nueva información mediante la acumulación de niveles de probabilidad y presenta un estudio de fiabilidad de los estudios en relación al sesgo de publicación.

- (HEDGES, 1982; ROSENTHAL y RUBIN, 1982a; HUNTER y SCHMIDT, 1978, 1990). Introducen los test de homogeneidad para poder reconocer posibles fuentes de variabilidad de los datos. Añaden control sobre fuentes de posibles errores que pueden afectar a los datos sobre los que se obtienen los diferentes estadísticos para la revisión.

- (CHALMERS y cols., 1977; YUSUF y cols., 1985). Sólo intervienen aquellos ensayos donde la asignación de los individuos a los grupos es aleatoria. Evitan, en lo posible, la estratificación debida a variables externas que caractericen los estudios.

- (EDDY, 1989; EDDY y cols., 1992). Intenta modelizar el proceso de integración a través del cálculo de múltiples parámetros

referentes al tamaño del efecto, evidencia adicional, sesgo de validez interna y externa.

Elegida una distribución de partida, este método pretende definir una distribución de probabilidad para el parámetro de interés. El cálculo de dicha distribución se basa en la combinación de los parámetros referentes al sesgo y la información de los estudios mediante análisis bayesianos.

En la actualidad los estudios meta-analíticos aplicados se planifican en función de características procedentes de los diversos tipos de meta-análisis descritos. Esto es debido a las diferentes conexiones entre áreas de investigación que se refleja en reuniones interdisciplinarias, cuyos objetivos se dirigen a unificar las características determinantes del meta-análisis como se puede reflejar en las diferentes reuniones acontecidas cuyo tema principal es el meta-análisis.¹

Una nueva clasificación, más general, puede derivar de los estudios llevados a cabo por el *ADVANCED OVARIAN CANCER TRIALISTS GROUP* (1991) y el *EARLY BREAST CANCER TRIALIST'S COLLABORATIVE GROUP* (1992) ambos realizan un Meta-análisis de los datos individuales de los distintos ensayos revisados. A partir de estos nuevos estudios, STEWART y PARMAR (1993) distinguen dos clases de Meta-análisis en función de los datos que se pretenden analizar:

1

- **Meeting of Systematic Reviews.** Organizado por el *British Medical Journal* y el *United Kingdom Cochrane Centre*. Organizado en Londres (1993)

- **Workshop on the future of meta-analysis.** *Committee on National Statistics of the National Research Council*. USA. (1986). Esta reunión derivó a la creación de un proyecto de la misma organización para la publicación de una colección de libros referentes a estudios teóricos y prácticos del Meta-análisis. *The future of Meta-analysis* (1990) y *Meta-analysis for explanation. A casebook* (1992) son dos de los libros editados hasta la fecha, resultado de dicho proyecto.

- Se dedicó un número completo de la revista *Statistics in Medicine* (YUSUF y Col 1987) a partir de una reunión de trabajo sobre las revisiones cuantitativas, celebrada en 1986.

- Utiliza como datos los resultados cuantitativos extraídos de las publicaciones o informes de investigación.
- : Analiza los datos individuales procedentes de cada uno de los estudios localizados en la revisión.

Las ventajas de esta nueva aproximación se resumen en la posibilidad del control del sesgo debido a exclusiones de pacientes; mayor potencia en los estudios de estratificación, y en el caso de análisis de supervivencia se pueden analizar los puntos de tiempo individualmente prolongando los tiempos de supervivencia (GELBER y GOLDBIRSCH, 1991; STEWART y PARMAR, 1993; CLARKE y STEWART, 1994).

El MAP comparte la misma estructura y objetivos de la revisión cuantitativa propuesta por LIGHT y SMITH (1971) y al igual que éste, el problema principal de este tipo de revisiones es la disponibilidad de los datos reales.

La antigüedad de los estudios y el recelo de los investigadores a la hora de prestar los datos impide que se pueda disponer de los mismos, lo que produce un sesgo que afecta a los resultados (GELBER y GOLDBIRSCH, 1991).

Realmente las revisiones con datos reales se podrían encuadrar dentro de los estudios multicéntricos. En dichos estudios varios investigadores plantean un estudio conjunto sobre un tópico común, en diferentes lugares. El diseño de protocolo, las variables a medir, etc. son iguales para todos los ensayos. De una forma paralela, en el meta-análisis con datos reales, los investigadores aportan sus trabajos y plantean un estudio conjunto con un mismo protocolo.

1.4. OBJETIVOS DEL META-ANALISIS

Cuando se plantea la realización de un meta-análisis, el investigador debe preguntarse la intención de la revisión. A partir de la intención se definen los diferentes objetivos que se intenta cubrir con el meta-análisis. Se puede resumir en dos alternativas generales posibles:

1.- El revisor quiere describir la literatura existente sobre un problema concreto, intentando responder la pregunta de *¿Qué dice la investigación disponible sobre el efecto de los tratamientos X's?* (GLASS, y cols., 1981; KULIK y KULIK, 1982).

En este tipo de meta-análisis los objetivos propuestos podrían sintetizarse así:

- Organizar y extraer la información sobre el tema seleccionado. No prejuzga los resultados en términos de calidad de estudio.
- Examinar la relación entre tamaños del efecto y características predeterminadas.
- Busca conclusiones generales sobre el tema investigado, si bien examina las características diferenciales de cada trabajo.

2.- El revisor intenta incrementar el tamaño de muestra en relación a una hipótesis específica para determinar una estimación general del efecto de un tratamiento. En este caso la pregunta a responder será ¿Cuál es el efecto real del tratamiento? (PETO y cols., 1977; HEDGES y OLKIN, 1985; ROSENTHAL y RUBIN, 1982b). Los objetivos planteados son más específicos:

- Proveer estimaciones del tamaño del efecto.
- Resolver el desacuerdo existente en ciertos campos.
- Responder a cuestiones que no se plantean en los estudios individuales.
- Confirmar la información referente a las hipótesis planteadas.
- Búsqueda de posibles errores que afecten a los resultados.
- Desarrollo de nuevas ideas (hipótesis) para abrir nuevas líneas de investigación.

1.5. ESTADO ACTUAL DEL META-ANÁLISIS

1.5.1 INTRODUCCIÓN

Los estudios meta-analíticos han proliferado con gran intensidad a partir de su planteamiento por Glass. Este incremento se ha producido sobre todo a nivel teórico práctico. En el campo aplicado, la publicación de estudios meta-analíticos ha sido notable en la mayoría de áreas de investigación científica. Sin embargo, el desarrollo es mayor en aquellos campos donde el meta-análisis surgió como un método alternativo de revisión científica: Psicología y Medicina.

Las prestaciones ofrecidas por el mundo informático facilita al usuario el acceso a las múltiples publicaciones que se producen en el área de investigación correspondiente. Esa ventaja nos ha permitido la realización de un estudio bibliográfico para estimar el número de publicaciones de meta-análisis en las diferentes áreas de investigación. El número de dichos artículos puede proporcionar una aproximación del impacto de dichas revisiones en el mundo científico.

1.5.2 ANÁLISIS BIBLIOGRÁFICO

Para este estudio se eligieron las bases de datos que generalmente los

investigadores suelen consultar para la búsqueda de las publicaciones de interés en las diferentes áreas.

Las bases de datos consultadas fueron

- *Medline.*
- *Psychology literature.*
- *Sociofile.*
- *Biological Abstract.*
- *American Economic Association's Journal of Economic Literature.*

En la siguiente tabla aparecen el número de publicaciones encontradas relacionadas con el meta-análisis y el año de aparición del primer artículo.

Base de datos	Años de búsqueda	1er Año de aparición	Nº de artículos
<i>Medline</i>	1983-94	1983	2304
<i>Psychlit</i>	1974-94	1976	1367
<i>Sociofile</i>	1974-94	1980	21
<i>A.E.A.</i>	1969-94	1985	25
<i>Biological Abstract</i>	1987-1994	1987	450

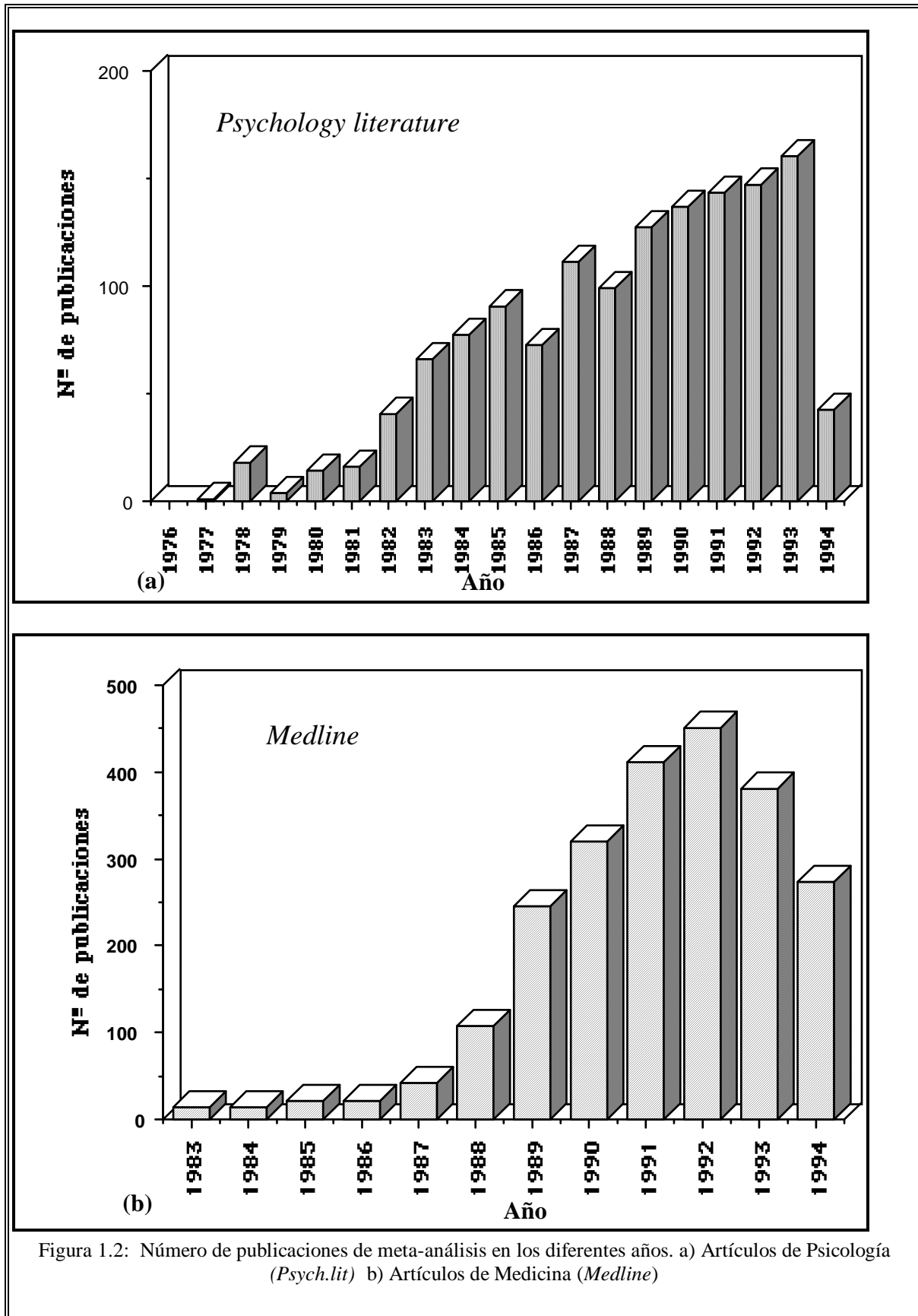
Dada la importancia cualitativa y cuantitativa de este método en el campo de la Medicina y Psicología, se ha procedido a un seguimiento anual de la incidencia de publicación del Meta-análisis en dichas áreas. En la figura 1.2. aparece el número de publicaciones en distintos años.

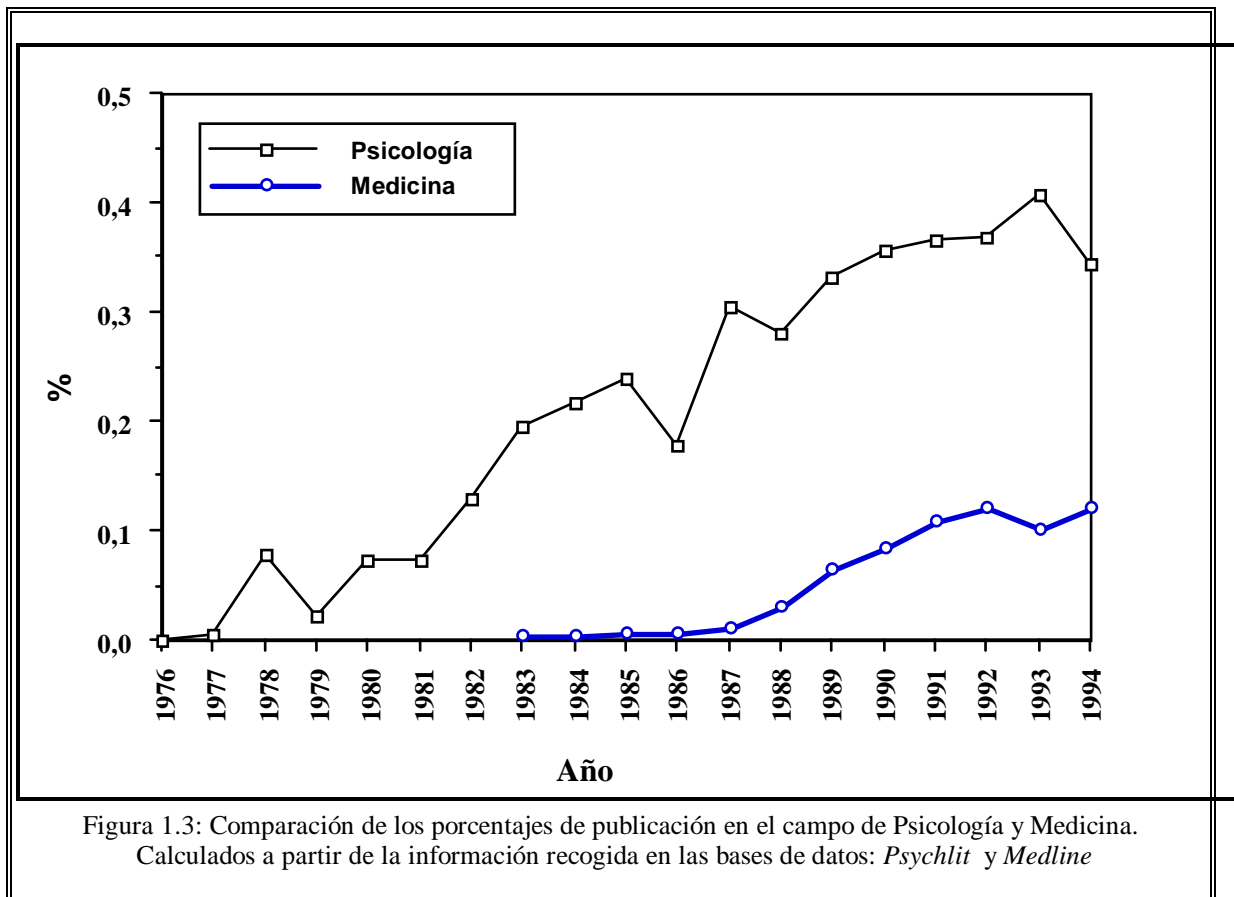
Se observa el incremento producido en los años noventa donde se recogen el 60,2% de los meta-análisis médicos. En el área de Psicología las aplicaciones son mas constantes a lo largo del tiempo y sobre todo a partir de los años ochenta. En ambos gráficos se observa el rápido crecimiento de este método en los años noventa donde hay una incorporación anual a las bases de datos de 100 artículos en Psicología y 300 en Medicina.

En la figura 1.3 se representa el estudio comparativo en porcentaje referente al número de publicaciones total recogidas por la base de datos en los distintos años. Así, en la base de datos *Psychology* se calcula un 0,23% de artículos dedicados al meta-análisis frente a un 0,057% obtenido en la *Medline* .

La diferencias varían cuando se realiza una nueva búsqueda informática mediante la palabra clave *Overview*, sinónimo de meta-análisis en Medicina. El número de articulos encontrados con esta acepción fue de 9222 que, sumadas a las ya obtenidas, incrementa el porcentaje a un 0,28%, muy similar al obtenido para el caso de las publicaciones en psicología, si bien, en la acepción de *overview* se han englobado revisiones narrativas denominadas con esta palabra.

En el caso concreto de las publicaciones en Biología el número de Meta-análisis es relativamente grande debido a las áreas de investigación convergentes con los estudios médicos sobre todo en nutrición, microbiología y genética. Sin embargo, la incidencia de este tipo de revisiones en otras áreas de investigación dentro de la Biología como Ecología, Zoología o Botánica es muy escasa. Así, en la búsqueda en la base de datos *Zoological Record* proporcionó un total de cinco publicaciones de meta-análisis desde 1974. En caso de búsquedas similares para el resto de áreas, el resultado fue muy similar.





Las primeras publicaciones en el campo de Ecobiología se deben a JARVINEN (1991) el cual realiza un meta-análisis sobre el efecto que produce la edad de la hembra en la fecha y tamaño de la puesta de *Parus major* y *Ficedula hypoleuca* y MARCHANT y MCGREW (1991) sobre la funciones de lateralidad en el cerebro de primates.

En años sucesivos se produce una diversificación en distintos campos de conocimiento como: reproducción en aves (VANDERWERF, 1992), competencia (GUREVITCH y cols., 1992), comportamiento de primates (ISELL y YOUNG, 1993); entomología y parasitología (POULIN, 1994; TONHASCA y BYRNE, 1994) y contaminación (DAVIS, 1993; AGARD y cols., 1993; WARWICK y CLARKE, 1993; FRENI, 1994).

FERNANDEZ-DUQUE y VALEGGIA (1994) y, ARNQVIST y WOOSTER (1995) comentan la importancia del meta-análisis como un valiosa herramienta en las áreas de ecología y evolución.

En España la estudios revisados sobre meta-análisis tanto metodológicos como prácticos son menos abundantes. GOMEZ (1987) y ANGUERA y SANCHEZ (1982) hacen una descripción metodológica del meta-análisis referido al área de psicología. GOMEZ y cols. (1989); GALINDO (1991); DELGADO-RODRIGUEZ y cols.(1991; 1992) son trabajos pioneros en la descripción de los métodos meta-analíticos referidos a la investigación médica de ámbito nacional. En los años 1990-1994 las aplicaciones de las técnicas meta-analíticas en el campo de Medicina son más habituales. Así, en una búsqueda en *Medline* de publicaciones sobre aplicaciones de meta-análisis cuyo origen es de instituciones nacionales fueron encontrados 21 artículos en las diferentes áreas.

Por el contrario, en las distintas búsquedas en las bases de datos biológicas mencionadas anteriormente **no aparece ningún artículo referido a meta-análisis** procedente de centros de investigación nacionales.

1.6. ETAPAS DE UN META-ANALISIS

Al igual que en un proyecto de investigación, en los estudios meta-analíticos se establecen una serie de etapas las cuales deben ser minuciosamente controladas para desarrollar correctamente la revisión cuantitativa. Varios autores (COOPER, 1982; ROSENTHAL, 1984; SACKS y cols., 1987; KUTTY, 1991) han propuesto diferentes estadios en función del campo en el que se realiza su investigación. Las diferencias entre las distintas proposiciones son de tipo formal y no sustancial.

Una de las clasificaciones más generales es la propuesta por ROSENTHAL en su trabajo de 1984, donde define las etapas en función del paralelismo que debe existir entre una investigación primaria y un meta-análisis. Se diferencian seis etapas básicas por las que debe discurrir cualquier síntesis de investigación:

- 1.- Formulación del problema.**
- 2.- Búsqueda de la literatura.**
- 3.- Codificación de los estudios.**
- 4.- Medida de los resultados.**
- 5.- Análisis e interpretación de datos.**
- 6.- Publicación del estudio.**

CAPITULO 2

FORMULACION DEL PROBLEMA

2.1. DEFINICION DE LOS OBJETIVOS

Análogamente a como se hace en las investigaciones primarias, cuando se lleva a cabo una revisión, sea o no cuantitativa, el primer paso consiste en especificar, con la mayor rigurosidad posible, el tema concreto sobre el que se va a recoger la información. Se define un objetivo primario en el cual se hace referencia al impacto que produce cierta intervención en un contexto dado como el efecto de cierto tratamiento en una enfermedad, el impacto de un plan educativo en el desarrollo infantil o el efecto de una campaña publicitaria en las ventas de un producto, etc. Adicionalmente, se puede presentar un conjunto de objetivos secundarios relacionados que ayudan a la comprensión del problema planteando interacciones, estudio de subpoblaciones, etc.

A la hora de definirse un objetivo debe tenerse en cuenta que un concepto definido en términos muy restringidos puede afectar a su generalidad y a las posibles inferencias; por el contrario si se define de forma amplia puede no tener una representación concreta de la población de estudios, proporcionando una generalización exagerada de los resultados.

2.2. PROTOCOLO

En todo estudio meta-analítico debe considerarse un plan de investigación o *protocolo* el cual tiene que tener presente la definición precisa y concisa de los objetivos sobre el tema a revisar (CURLETTE y CANELLA, 1985; SACKS y Col., 1987; JENICEK, 1987).

El protocolo debe ser preparado antes de comenzar cada revisión y debe especificar los procedimientos que van a ser utilizados en cada una de las etapas de dicho meta-análisis. Este plan de estudios debería presentar la siguiente estructura:

- Una introducción del tema elegido para revisión donde se centra el estudio y se plantea el impacto del mismo.
- Definición de los objetivos a cubrir distinguiendo entre primarios y secundarios. Se deben especificar las variables que se van a analizar.
- Definición de los criterios de inclusión de los trabajos o experimentos que formarán parte del estudio meta-analítico. Estos criterios deben tener presentes los objetivos planteados en el estudio.
- Identificación de las fuentes disponible y útiles para la búsqueda de la literatura.

- Descripción de la información extraíble de cada uno de los estudios. Se incluirá el cuestionario de codificación donde se definen las características cualitativas diferenciales de cada estudio.
- Descripción de los análisis que se aplicarán para el cálculo del tamaño del efecto, así como los análisis paralelos del estudio de homogeneidad, estudios de subpoblaciones, etc.

Las fases segunda y tercera delimitan el universo de estudios que deben ser recuperados para la realización de la revisión. Estas etapas son fundamentales para la ulterior definición del sesgo de publicación por lo cual deben ser detalladas con rigurosidad.

2.3. CALIDAD DE LOS ESTUDIOS

Hay una fuerte controversia sobre si los estudios con deficiencias metodológicas deben ser excluidos a priori del meta-análisis; ciertos autores defienden la inclusión de todos, aunque dándoles un peso en función de su calidad metodológica (GLASS y cols., 1981; STRUBE y HARTMAN, 1984; CURLETTE y CANELLA, 1985; FLEISS, 1987). En otra dirección hay investigadores que propugnan la incorporación de los estudios que presenten cierta calidad valorada por un conjunto de jueces. (JENICEK, 1989; CHALMERS y LAU, 1993). En una posición intermedia PETO (1987b) propugna un mínimo requerimiento, que los experimentos procedan de asignaciones aleatorias a los grupos, para que los estudios intervengan en las revisiones meta-analíticas. Para este autor esta condición evita una importante fuente de sesgo procedente del control de los factores que pueden influir en los resultados y que no están controlados en los estudios primarios lo cual evita que influyan en los resultados finales obtenidos en el meta-análisis.

CAPITULO 3

BUSQUEDA DE LA LITERATURA

3.1. INTRODUCCION

La recuperación del conjunto de publicaciones o ensayos que intervienen en la revisión es una de las etapas más largas e importantes del meta-análisis. En esta fase el revisor está obligado a conseguir el mayor número de estudios sobre el tópico de investigación que pretende revisar. RACHMAN y WILSON (1980) juzgan como inútiles, incluso contraproducentes, los resultados de un meta-análisis, si la búsqueda de la literatura es incompleta o procedente de una exploración sesgada.

De esta etapa deriva uno de los problemas más importantes que puede afectar al meta-análisis y que más adelante se estudiará con detalles: el sesgo procedente de la recuperación de los estudios. Este tipo de sesgo generalmente se conoce con el nombre de *sesgo de publicación*. Los principales factores que inducen a una recuperación incompleta de los estudios son:

- La selección editorial a favor de estudios con resultados significativos (CARVER, 1978; SMITH, 1980; COURSOLO y WAGNER, 1986). ROSENTHAL, 1979, afirma en la pág. 638 de su libro que en la literatura científica publicada existe, aproximadamente, un 5% de trabajos que cometen errores Tipo I mientras que permanecen sin publicar el 95% de los estudios con resultados no significativos.

Si un meta-análisis sólo incluye literatura publicada es probable que cometa un importante error, favoreciendo la significación de los resultados.

- La tendencia, por parte de los editores, a publicar trabajos cuyos resultados coinciden con una hipótesis establecida. Aquellos que la refutan es poco

probable conseguir su publicación o la admisión en revistas de impacto que facilite su recuperación (BEGG y BERLIN, 1988).

- El punto de vista que el revisor tiene sobre un tema determinado puede sesgar la selección hacia áreas o revistas que están más cerca de su especialidad (CHALMERS y cols., 1990; HATTIE y cols., 1984).

Un análisis de esta problemática en el contexto de los ensayos clínicos puede encontrarse en SIMES, 1986b; y DIKERSIN y cols., 1987.

El paso siguiente a la definición del problema objeto de investigación consiste en la identificación de todos los estudios que han tratado dicho tópico y por lo tanto describen una población de estudios a identificar.

3.2. POBLACION OBJETIVO Y POBLACION ACCESIBLE

La población de estudios está directamente delimitada por las definiciones conceptuales y operativas de la etapa anterior. Debido a las limitaciones que aparecen en la recuperación de los estudios se debe distinguir entre población objetivo y población accesible:

- **Población objetivo**. Está compuesta por todos los estudios que el revisor espera poder representar en el estudio. La definen los objetivos y criterios de elección propuestos en el protocolo.
En la práctica resulta materialmente imposible encontrar todos los elementos de una población objetivo.
- **La población accesible**. Incluirá aquellos estudios que el revisor es capaz de localizar. Derivado de esta limitación nos encontraremos con el problema del sesgo de publicación.

Es requisito fundamental en un meta-análisis definir la población objetivo y la población accesible así como determinar en qué medida pueden diferir, pues si ambas no coinciden, cosa que ocurre en la práctica totalidad de los meta-análisis, ésto supone una importante amenaza para la validez de los resultados.

3.3. FUENTES DE INFORMACION

Tres fuentes de información deben ser utilizadas para asegurar la adecuación de la población accesible a la población objetivo (COOPER, 1984; GLASS y cols., 1981; HUNTER y cols., 1982).

- **Fuentes informales:** El llamado procedimiento de “*colegio invisible*”; es una red de comunicación informal entre científicos que investigan en un mismo tópico. El revisor debe intentar contactar con estos expertos investigadores para obtener el material bibliográfico que puedan proporcionarle. Es un método fructífero para acceder a nuevos estudios o trabajos no publicados (MULLEN, 1989).
- **Fuentes Primarias:** Esta estrategia utiliza las bibliografías y referencias de los estudios relevantes que han sido ya recuperados para la localización de nuevos estudios. COOPER (1982) distingue los siguientes tipos fundamentales de técnicas para la localización de estudios:
 - *El Procedimiento ascendente (ancestry approach):* consiste en localizar investigaciones previas que figuran en listas de referencias de estudios ya disponibles (una revisión, una tesis, etc.). El revisor general parte de una publicación del año que se plantea el estudio meta-analítico y a partir de ella puede obtener las citas de los estudios más relevantes en el tópico elegido.
 - *El procedimiento descendente:* consiste en localizar trabajos en los

cuales se citan los documentos ya recuperados, por ejemplo, consultando índices de citas del área de interés. Es un buen complemento al método anterior.

Estos métodos de búsqueda presenta el inconveniente de las citas irrelevantes, poco útiles para la revisión, y en las cuales el investigador invierte mucho tiempo para su recuperación (MULLEN, 1989).

- **Servicio de abstracts.** Este procedimiento permite localizar estudios asociados a determinadas palabras clave en las distintas bases de datos que existen para las áreas de conocimiento (*Medline, Biological Abstracts, Educacional Resources Information Center, Psychological Abstracts, Sociofile, American Economic Association's Journal of Economic Literature, etc.*).

Es importante que el autor conozca bien la estructura de la bases de datos para identificar el dominio que comprende la búsqueda: el año en que comienza a almacenar información, tipo de trabajos que comprende (artículos, editoriales, cartas, tesis, etc.) cuales son los campos de búsqueda (autor, año, palabra clave, etc.), problemas que se presentan en cada campo.

Una parte importante de esta búsqueda es definir con detalle las palabras claves (*Key words*) para identificar el mayor número de publicaciones en el tema que se está revisando. Las claves de identificación tienen que compartir la característica de ser habituales y de uso común en la literatura propia del tema elegido.

Un inconveniente de este procedimiento resulta del método exacto de identificación definido por las diversas bases de datos. El ordenador reconoce la palabra exacta pero no los plurales o las palabras derivadas, así por ejemplo, si se busca la palabra meta-análisis en una base de datos, ésta

sólo identificará las publicaciones donde aparece este término pero no sus derivaciones metaanálisis, meta-analíticos, etc.

Varios autores (SACKS y cols., 1987; MULLEN, 1989; SAWKA y PRITCHARD, 1992; PETITTI, 1994) advierten de la problemática del uso de la búsqueda computerizada si bien hoy en día es la más rápida y eficaz. Sin embargo, debe intentarse la búsqueda no sólo con consulta en bases de datos ya que según comprobaron DIKERSIN y cols. (1985) la búsqueda computerizada no llega a cubrir ni siquiera los 2/3 de los trabajos efectuados.

3.4. ETAPAS EN LA BUSQUEDA DE LA LITERATURA

Una buena búsqueda de la literatura debe utilizar el conjunto de todas las posibles fuentes de información ya que no son excluyentes entre si. PETITTI (1994) presenta una lista de los pasos generales que se utilizan para la recuperación de los estudios de investigación en un meta-análisis:

- 1.- Búsqueda en los archivos personales.
- 2.- Búsqueda de la literatura en las bases de datos relacionadas con el tópico:
 - a) Inspección de títulos y resúmenes. En esta fase se produce la primera eliminación de aquellos estudios irrelevantes.
 - b) Recuperación de los estudios válidos.
 - c) Revisión sistemática de los estudios recuperados y eliminación de aquéllos en los que no es posible conseguir la información necesaria. Se intenta contactar con el investigador principal para recuperar la información no disponible en los artículos.
- 3.- Revisión de las referencias citadas en los artículos recuperados en las fases anteriores.
- 4.- Consulta con los expertos para la obtención de aquellos estudios que no han sido publicados.

En la figura 3.1. aparece un esquema en el que se sintetiza las diferentes fases en la búsqueda de la bibliografía.

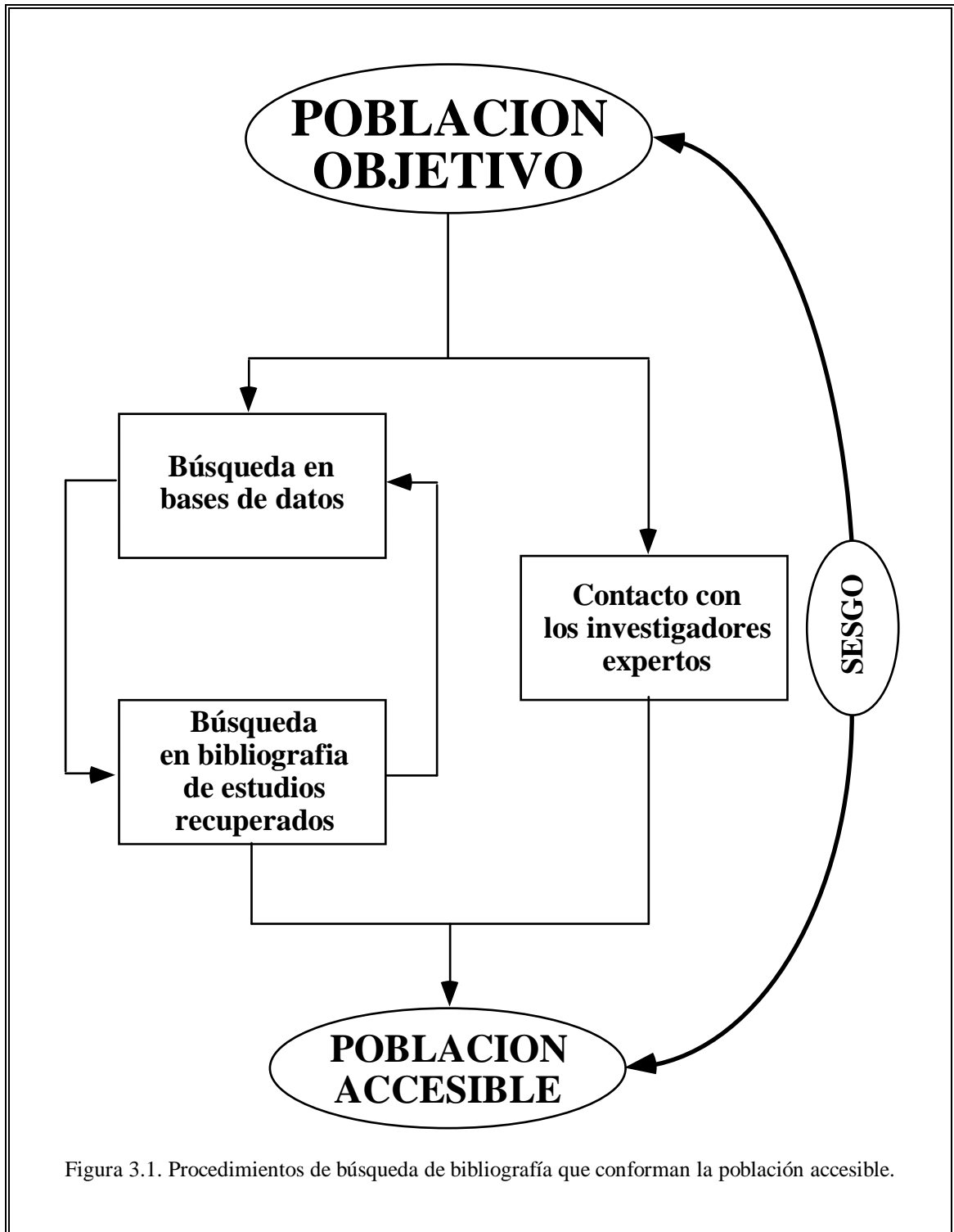


Figura 3.1. Procedimientos de búsqueda de bibliografía que conforman la población accesible.

CAPITULO 4

CODIFICACION DE LOS ESTUDIOS

4.1. DESCRIPCIÓN DE LAS CARACTERÍSTICAS DIFERENCIALES DE LOS ESTUDIOS

El propósito de cualquier meta-análisis es extraer el estado de conocimiento que se posee sobre algún ámbito de investigación; dicha información puede ser cuantitativa o cualitativa. JENICEK (1987) hace distinción entre el *meta-análisis cualitativo* definido por el estudio de las características diferenciales de los trabajos y el *meta-análisis cuantitativo* procedente del análisis de los resultados de cada trabajo.

En las revisiones meta-analíticas los *datos* provienen de ensayos de diferente procedencia y calidad que puede arrojar resultados contradictorios. El revisor debe codificar las características de los estudios que supuestamente afecten a dichos resultados.

Este proceso de codificación suele formalizarse elaborando un cuestionario que incluya todas las características relevantes: **objetivas**, como el número de pacientes, año de publicación, etc. y **subjetivas**, relacionadas con la calidad de los estudios y las cuales son valoradas por un conjunto de jueces. El objetivo de la codificación es concretar el conjunto de peculiaridades que definen cada uno de los estudios recuperados. Las características diferenciales pueden ser útiles a la hora de explicar la heterogeneidad de los resultados, así como para posibles análisis de la calidad de los mismos.

GLASS y cols. (1981) distinguen entre **características metodológicas** y **características sustantivas**; las primeras se refieren a aspectos más generales de la investigación, tales como la fecha y fuente de publicación, tipo de diseño de investigación, número de individuos perdidos, etc. Las características sustantivas son específicas del tópico de investigación que se estudie; por ejemplo el tipo de droga administrada, la cantidad de la misma, la edad de los sujetos, etc.

Paralelamente CHALMERS y cols., (1981) ofrece unas líneas de trabajo en la fabricación de los cuestionarios para el estudio de la calidad en los estudios de Medicina. Este cuestionario se divide en tres bloques principales:

- Items referentes al diseño del estudio: Selección de los pacientes, descripción de las características basales de los pacientes, valoración del método de aleatorización, etc.
- Items referentes a los análisis: Tipo de test estadísticos utilizados, la potencia estadística, tipo de estimación, etc.
- Items referentes a la presentación: Tabulación de los resultados, gráficos de los mismos, etc.

A cada ítem se le asigna una valoración de forma que el primer bloque suma un total de 60 puntos, el segundo un total de 40 y el tercero 20.

Autores como COOPER (1984), MULLEN y ROSENTHAL (1985) y, STRUBE y HARTMANN (1984) advierten el peligro de extraer conclusiones sobre la relación entre características y resultados de los estudios; teniendo en cuenta que las unidades del meta-análisis no han sido muestreadas de forma aleatoria e independiente a partir de una población conocida; los resultados de

aplicar técnicas estadísticas inferenciales utilizando las características de los estudios deberían considerarse únicamente como orientativos; en consecuencia, si el análisis de la relación entre características y resultados sugiere alguna relación novedosa y teóricamente relevante, el revisor debe procurar que sea investigada a nivel primario ya que el meta-análisis establece hipótesis plausibles, pero no las prueba (STRUBE y HARTMAN, 1984).

Métodos empíricos para detectar “*outliers*” o “*influential observations*” juegan un papel importante en la evaluación de los datos; estos métodos a veces revelan problemas en los estudios que fueron inicialmente pasados por alto.

4.2. ESTUDIO DE LA FIABILIDAD EN LA CODIFICACION

Los valores de calidad pueden ser utilizados para descartar aquellas publicaciones con valores muy bajos o nulos en el conjunto de los ítems (JENICEK, 1989; SAUKA y PRITCHARD, 1992), o bien esos valores se utilicen como ponderaciones utilizadas en el cálculo del tamaño del efecto global en el caso que la calidad esté relacionada con el tamaño del efecto (MOSTELLER y CHALMERS, 1992; LAIRD y MOSTELLER, 1990).

El proceso de codificación de las características subjetivas de los estudios constituye un problema de medida que depende de los diferentes jueces que realizan la codificación y por tanto está sujeto a las deficiencias de fiabilidad y validez del estudio.

La falta de fiabilidad en la codificación puede deberse a falta de calidad de los informes o a errores del codificador. Si las reglas de codificación no son suficientemente explícitas, distintos codificadores de un mismo estudio pueden hacer juicios diferentes (GLASS y cols., 1981).

Para proteger la fiabilidad del proceso de codificación debe elaborarse un libro de codificación que incluya lo más exhaustivamente posible las definiciones de las características, las normas y las reglas de codificación.

También es recomendable que el proceso de codificación se lleve a cabo por un equipo de codificadores previamente entrenados para la perfecta interpretación y utilización del libro de codificación. Este equipo evaluará una muestra aleatoria de todos los estudios incluidos en el meta-análisis.

Debe hacerse un estudio sistemático de la fiabilidad inter-codificadores. Cuanto mas grande sea el número de codificadores mayor será la fiabilidad efectiva, ya que los errores aleatorios de cada codificador se contrarrestan con los errores de los otros.

ROSENTHAL (1984) estima la fiabilidad efectiva **R** mediante la expresión:

$$\mathbf{R} = \frac{\mathbf{nr}}{\mathbf{1 + (n - 1)r}}$$

n es el número de codificadores y **r** la fiabilidad media entre los **n** codificadores; es decir, la media de los $n(n-1)/2$ coeficientes de correlación.

Esta expresión puede ser utilizada no sólo para el coeficiente de correlación de Pearson, sino también con el de Spearman, la correlación biserial-puntual o el coeficiente phi.

Obviamente si el número de codificadores es elevado el cálculo de la correlación media es tedioso.

Un procedimiento alternativo recogido también en la bibliografía clásica consiste en aplicar un análisis de la varianza de medidas repetidas, donde los estudios corresponden a las observaciones y los codificadores constituyen el factor de variación.

Por este procedimiento es posible estimar la fiabilidad efectiva de la siguiente forma:

$$R = \frac{MC_{\text{entre estudios}} - MC_{\text{error}}}{MC_{\text{entre estudios}}}$$

Ambos estimadores de la fiabilidad efectiva proporcionan resultados idénticos, salvo errores de redondeo.

CONTRIBUCIONES A LA FASE DE CODIFICACION DESDE UNA PERSPECTIVA MULTIVARIANTE

4.3.1. ANALISIS MULTIVARIANTE DE LOS RESULTADOS DE LA CODIFICACION

El objetivo principal de la fase de codificación suele ser la obtención de un índice de calidad de los estudios que, posteriormente, puede ser utilizado como variable para contrastar si el tamaño del efecto está relacionado con dicho índice y la calidad de los diferentes estudios puede comprometer los resultados.

Tres son los principales problemas de análisis previos a la construcción de un índice de calidad:

1. Investigar cuales son las cuestiones mal definidas o confusas para los jueces.
2. Investigar si los jueces realizan juicios congruentes.
3. Sintetizar las distintas puntuaciones en una única medida.

Si tenemos en cuenta el carácter multivariante de los cuestionarios de calidad, es evidente que los análisis univariantes propuestos en la bibliografía son insuficientes.

En lo que sigue trataremos simultáneamente los distintos problemas desde una perspectiva multivariante.

Sea \mathbf{x}_{ijk} la valoración realizada por el j -ésimo juez de la cuestión k -ésima en el i -ésimo estudio. Si disponemos de I estudios, J jueces y K cuestiones podemos organizar los datos en una matriz \mathbf{X} de IJ filas por K columnas conteniendo todas

las valoraciones realizadas.

La tabla completa de datos sería de la forma:

		Cuestiones				
Estudios	Jueces	1	...	k	K
1	1	x_{111}		x_{11k}		x_{11K}

	J	x_{1J1}		x_{1Jk}		x_{1JK}
...
i	1	x_{i11}		x_{i1k}		x_{i1K}

	j	x_{ij1}		x_{ijk}		x_{ijK}

	J	x_{iJ1}		x_{iJk}		x_{iJK}
...
I	1	x_{I11}		x_{I1k}		x_{I1K}

	J	x_{IJ1}		x_{IJk}		x_{IJK}

Si tomamos como hipótesis de partida que los jueces realizan juicios coherentes entre si, la puntuación común asignada sería la media de las puntuaciones asignadas por los J jueces. Así, llamaremos \bar{x}_{ik} a la media de las puntuaciones dadas por los J jueces en la k-ésima cuestión para el i-ésimo estudio.

$$\bar{x}_{ik} = \sum_{j=1}^J x_{ijk}$$

Las medias pueden organizarse en una matriz $\bar{\mathbf{X}}$ de I filas por K columnas.

Si las preguntas están bien formuladas y la matriz \bar{X} es correcta, un índice de calidad puede obtenerse combinando las distintas puntuaciones en una sola mediante el primer eje de una técnica factorial, como por ejemplo el Análisis de Componentes Principales (PEARSON, 1901), o los métodos Biplot (GABRIEL, 1971), o el Análisis Factorial de Correspondencias (BENZECRI, 1973), entre otros.

4.3.2. EL METODO BILOT COMO TECNICA DE CARACTERIZACION DE ESTUDIOS E IDENTIFICACION DE JUECES DISCORDANTES.

Las filas de X y las de \bar{X} se pueden representar como puntos en un espacio K dimensional de forma que ambas pueden representarse de forma exacta (mediante ejes cartesianos) en un espacio r -dimensional donde r es el rango de la matriz \bar{X} .

Sin pérdida de generalidad podemos suponer que la matriz original de datos X está centrada por columnas con lo cual la matriz \bar{X} también está centrada por columnas.

La representación directa de esta bigeometría no es útil en la práctica ya que r suele ser bastante mayor de 2 o 3.

Para resolver el problema buscaremos la mejor aproximación a \bar{X} de rango 2.

Partiendo de la descomposición en valores singulares de la matriz a representar. (GOOD, 1969)

$$\bar{\mathbf{X}} = \mathbf{U} \Sigma \mathbf{V}'$$

donde \mathbf{U} son los vectores singulares por la izquierda que coinciden con los vectores propios de $\bar{\mathbf{X}} \bar{\mathbf{X}}'$, \mathbf{V} son los vectores singulares por la derecha y coinciden con los vectores propios de $\bar{\mathbf{X}}' \bar{\mathbf{X}}$ y $\Sigma = \text{diag}(\lambda_1, \dots, \lambda_r)$ es una matriz diagonal que contiene los valores singulares ordenados de mayor a menor. Los valores singulares coinciden con los valores propios de $\bar{\mathbf{X}} \bar{\mathbf{X}}'$ y $\bar{\mathbf{X}}' \bar{\mathbf{X}}$ que son iguales.

La mejor aproximación a rango 2 vendría dada por los dos primeros valores y vectores singulares.

$$\bar{\mathbf{X}} \cong \mathbf{U}_{(2)} \Sigma_{(2)} \mathbf{V}'_{(2)}$$

donde el subíndice (2) significa las dos primeras columnas de la matriz correspondiente.

La bondad del ajuste de la aproximación en rango 2 puede calcularse mediante la expresión

$$\frac{\lambda_1 + \lambda_2}{\sum_{i=1}^r \lambda_i}$$

El tomar las dos primeras columnas hace que la aproximación sea de rango 2 y la matriz tenga una representación exacta en 2 dimensiones.

Si consideramos la factorización de la matriz $\bar{\mathbf{X}}$ en la forma:

$$\bar{\mathbf{X}} \cong \mathbf{GH}'$$

con $\mathbf{G} = \mathbf{U} \Sigma$ y $\mathbf{H} = \mathbf{V}$, y representamos las filas de \mathbf{G} y las filas de \mathbf{H} como puntos (o vectores) en un subespacio de dimensión 2, el producto escalar de las filas de \mathbf{G} y las filas de \mathbf{H} aproxima los elementos de la matriz original

$$\begin{aligned} \bar{x}_{ij} &\cong \mathbf{g}'_i \mathbf{h}_j = \\ &\text{Proy}(\mathbf{g}_i / \mathbf{h}_j) (\mathbf{S}) \|\mathbf{h}_j\| = \\ &\text{Proy}(\mathbf{h}_j / \mathbf{g}_i) (\mathbf{S}) \|\mathbf{g}_i\| \end{aligned}$$

Existen otras alternativas de representación. Para detalle ver GABRIEL (1971,1981) y GALINDO(1986).

La utilidad de la representación se deduce a partir de sus propiedades, que se pueden sintetizar de la siguiente forma:

a) La similaridad entre dos estudios medida a través de la distancia euclídea entre sus puntuaciones se aproxima en el gráfico mediante la distancia euclídea entre dos marcadores fila \mathbf{g}

Si denotamos con $\bar{\mathbf{x}}_i$ a una fila de la matriz $\bar{\mathbf{X}}$ la similaridad entre dos filas puede medirse mediante la distancia euclídea entre las mismas como

$$\mathbf{d}^2(\mathbf{i}, \mathbf{h}) = (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_h)' (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_h)$$

que puede aproximarse en el gráfico mediante

$$\begin{aligned} \mathbf{d}^2(\mathbf{i}, \mathbf{h}) &= (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_h)' (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_h) \\ &= (\mathbf{g}_i - \mathbf{g}_h)' (\mathbf{g}_i - \mathbf{g}_h) \end{aligned}$$

Es decir, la distancia euclídea entre dos estudios en el espacio original se aproxima mediante la distancia euclídea entre marcadores fila.

Esta propiedad permite identificar estudios con características similares, y clasificarlos de acuerdo con sus propiedades.

b) Los marcadores fila pueden interpretarse como la proyección de las observaciones originales sobre el espacio definido por las dos primeras columnas de \mathbf{V} ; por lo tanto pueden calcularse también en la forma:

$$\mathbf{G} = \bar{\mathbf{X}} \mathbf{V}_{(2)} = \bar{\mathbf{X}} \mathbf{H}$$

Esta propiedad permite proyectar las observaciones originales, sobre las que se calcularon las medias, en el espacio bidimensional de representación de $\bar{\mathbf{X}}$ simplemente mediante

$$\mathbf{J} = \mathbf{X} \mathbf{V}_{(2)} = \mathbf{X} \mathbf{H}$$

Estas coordenadas permiten detectar los jueces que realizan juicios no congruentes en algunos de los ítems ya que es posible aproximar las puntuaciones originales mediante el producto escalar

$$x_{ij} \cong \mathbf{j}'_i \mathbf{h}_j$$

c) La bondad del ajuste de tales aproximaciones puede medirse mediante

$$\frac{\sum_{i=1}^I \sum_{k=1}^K (x_{ik} - \mathbf{j}'_i \mathbf{h}_k)^2}{\sum_{i=1}^I \sum_{k=1}^K x_{ij}^2}$$

que será, en general, menor que la obtenida para la matriz de medias; sin embargo, presenta información mas importante desde el punto de vista práctico, ya que es posible valorar gráficamente la variabilidad de las réplicas.

d) Las coordenadas para las columnas pueden interpretarse como la proyección del sistema de referencia canónico en el hiperespacio original sobre el subespacio definido por las dos primeras columnas de \mathbf{V} , por lo que definen un sistema de referencia oblicuo en un espacio de dos dimensiones (GOWER, 1995)

$$\mathbf{H} = \mathbf{I}_p \mathbf{V}_{(2)} = \mathbf{H}$$

donde \mathbf{I}_p es la matriz identidad de orden p.

Otra posibilidad alternativa a la anterior sería la de considerar el Biplot de la matriz de datos completa

$$\mathbf{X} \cong \mathbf{J} \mathbf{H}'$$

calculada a partir de la descomposición en valores singulares de la matriz \mathbf{X} .

Las propiedades de este biplot son similares a las propiedades del biplot de medias aunque la atención se centra ahora en la aproximación de los datos originales. La bondad del ajuste global será más pequeña que en el biplot de las medias.

La matriz de medias puede proyectarse sobre el biplot de \mathbf{X} tomando

$$\mathbf{G} = \bar{\mathbf{X}} \mathbf{H}$$

Lo cual es equivalente a calcular las medias correspondientes de los marcadores para las filas.

4.4. APLICACIÓN PRÁCTICA DE LAS APORTACIONES TEORICAS EN LA FASE DE CODIFICACION

Sea la siguiente matriz en la que se consideran 4 estudios, 3 jueces y 3 cuestiones. El análisis de esta aplicación tiene un carácter aclaratorio, metodológico, por eso lo hemos propuesto con dimensiones pequeñas (4x3x3). Evidentemente, una matriz de datos reales requeriría mas tiempo de análisis, pero el desarrollo de los mismos conceptos. Para ilustrar el procedimiento propuesto se han incluido dos valoraciones muy diferentes de las demás que podrían considerarse como outliers y que deberían detectarse claramente. Se trata de la valoración realizada en el estudio 1 por el juez 3 en la variable 2 (13) con un valor anormalmente pequeño y la valoración realizada en el estudio 4 por el juez 2 en la variable 1 (42) con un valor anormalmente grande.

Estudios	Jueces	variables			etiqueta
		v1	v2	v3	
<i>e1</i>	<i>1</i>	3	4	5	11
	<i>2</i>	2	5	5	12
	<i>3</i>	3	1	3	13
<i>e2</i>	<i>1</i>	3	3	4	21
	<i>2</i>	3	3	3	22
	<i>3</i>	1	2	4	23
<i>e3</i>	<i>1</i>	2	4	2	31
	<i>2</i>	2	3	1	32
	<i>3</i>	2	3	2	33
<i>e4</i>	<i>1</i>	1	1	1	41
	<i>2</i>	5	1	2	42
	<i>3</i>	2	1	1	43

Tabla.4.1 La matriz de datos ha sido construida ordenando los estudios con puntuaciones de mayor a menor e incluyendo algunas valoraciones diferentes señaladas en la matriz con negritas.

a) La matriz de puntuaciones medias \bar{X} es:

	V1	V2	V3
E1	2.6667	3.3333	4.3333
E2	2.3333	2.6667	3.6667
E3	2.0000	3.3333	1.6667
E4	2.6667	1.0000	1.3333

El gráfico de la página siguiente muestra la representación biplot de la matriz de medias sobre la que se han proyectado las observaciones individuales, la tabla 4.2 muestra los resultados del biplot: coordenadas de filas y columnas de la matriz de medias, coordenadas para las observaciones originales y bondades de ajuste para la aproximación de las medias y de los valores originales.

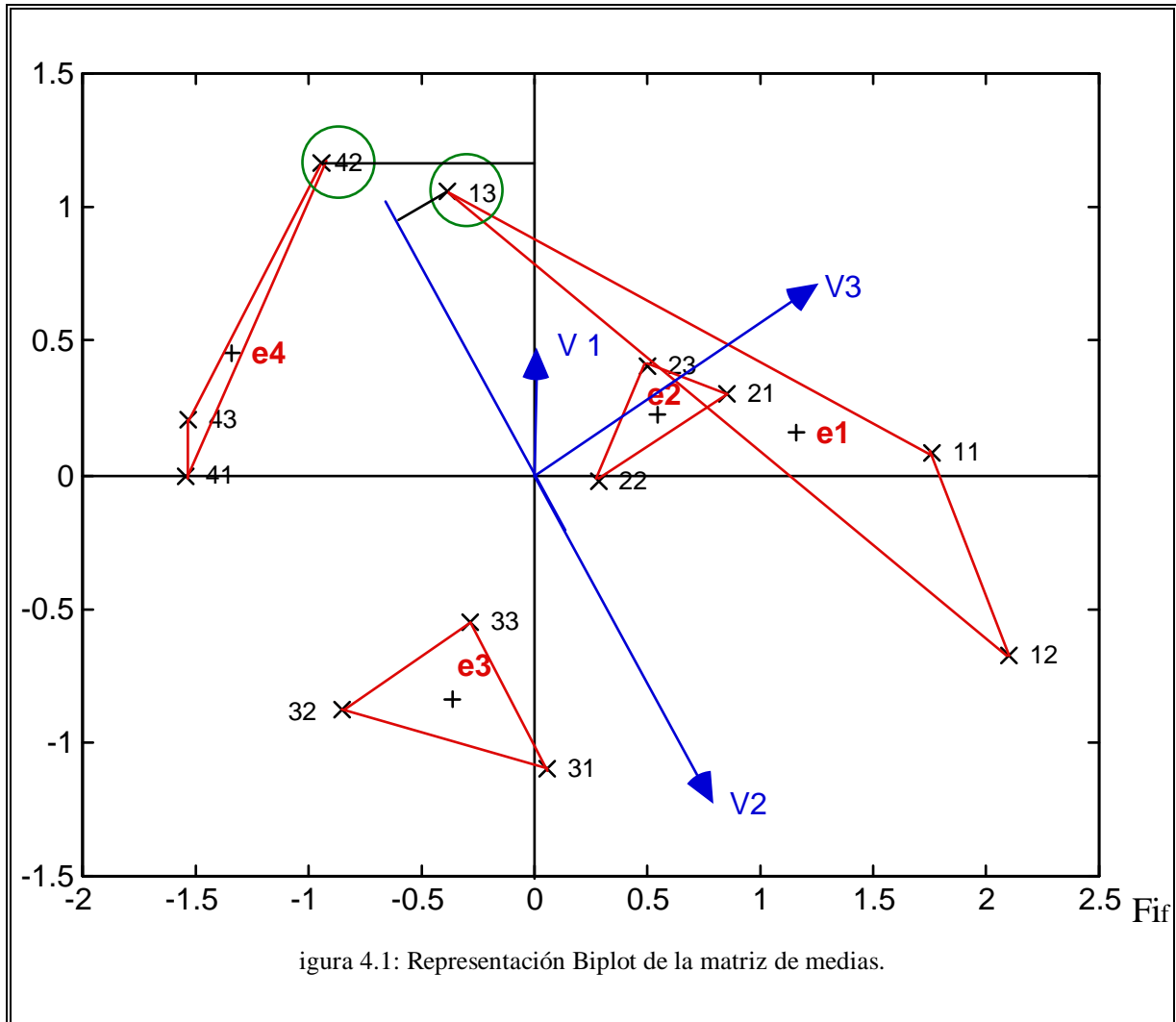
G =			H =			J =		
e1	1.7466	0.2331	v1	0.0111	0.3236	11	1.7646	0.0790
e2	0.8288	0.3469	v2	0.5125	-0.8143	12	2.0967	-0.6746
e3	-0.5504	-1.2676	v3	0.8586	0.4819	13	-0.3910	1.0587
e4	-2.0250	0.6876				21	0.8565	0.2992
						22	0.2879	-0.0199
						23	0.5024	0.4099
						31	0.0513	-1.0927
						32	-0.8568	-0.8725
						33	-0.2881	-0.5534
						41	-1.5430	-0.0082
						42	-0.9449	1.1682
						43	-1.5356	0.2061
inercia =	61.6086		94.0265			100.0000		

Bondad del ajuste para la matriz de medias = 99.2691

Bondad del ajuste para la matriz completa = 81.2181

Tabla 4.2: Resultados del Biplot de la matriz de medias

La bondad de ajuste para la matriz de medias ha sido alta (99.2691) por lo que las aproximaciones sobre el gráfico son buenas. Para los elementos de la matriz original la bondad del ajuste es un poco más baja (81.2181) como cabe esperar debido a que, en este caso, el interés se centra en el ajuste de las medias.



En la figura 4.1 se han unido con líneas los valores de los distintos jueces para el mismo estudio. Se pone claramente de manifiesto como los estudios 2 y 3 presentan menor variabilidad alrededor de la media y los jueces realizan valoraciones similares. Para los estudios 1 y 4 las respuestas son más variables poniéndose de manifiesto como son los puntos 13 y 42 los que ocupan posiciones discordantes.

Se comprueba también como es el eje 1 el que ordena los estudios de acuerdo con su puntuación por lo que las *coordenadas* sobre este eje pueden servir como un *índice de calidad*. Si se desea un índice de calidad con una variable discreta en lugar de continua, bastaría con dividir el eje en varios segmentos iguales y asignar cada estudio al segmento correspondiente.

El índice de calidad obtenido de esta forma puede utilizarse después para comprobar si el tamaño del efecto depende de la calidad de los estudios.

La obtención de cuáles son las variables responsables de las diferencias es un poco más compleja ya que, como puede observarse en el gráfico, hay una cierta indeterminación en las posiciones. Los dos puntos aparecen muy próximos en la representación cuando realmente sus valoraciones son erróneas debido a dos variables diferentes.

En un segundo paso se ha realizado el análisis biplot de la matriz de datos completa calculando después las medias. Los resultados aparecen en la tabla 4.3.

La bondad del ajuste es sensiblemente menor para las medias y de un orden comparable para la matriz completa.

Los resultados son muy similares a los obtenidos con el procedimiento anterior, aunque quizás las diferencias entre las observaciones están ahora más claras.
Figura 4.2

La observación 42 se proyecta en la parte superior de la variable 1, mientras que las otras dos observaciones del estudio 4 lo hacen en la parte inferior. La observación 13 se proyecta sobre la parte inferior de la variable 2.

J =			H =			G=		
11	2.6372	0.6694	v1	0.0192	0.8595	e1	0.9209	0.2347
12	3.2759	-0.5856	v2	0.6579	-0.3955	e2	0.4050	0.1047
13	-0.8423	1.2086	v3	0.7528	0.3237	e3	-0.1799	-0.5478
21	1.2264	0.7412				e4	-1.1460	0.2084
22	0.4736	0.4175						
23	0.5300	-0.5823						
31	0.3594	-1.1613						
32	-1.0513	-1.0895						
33	-0.2985	-0.7657						
41	-2.3864	-1.1579						
42	-1.5567	2.6039						
43	-2.3672	-0.2984						

inercia = 47.2270 77.5829 100.0000

Bondad del ajuste de la matriz completa = 86.2487

Bondad del ajuste para la matriz de medias = 49.4573

Tabla 4.3: Resultados del biplot de la matriz de datos brutos

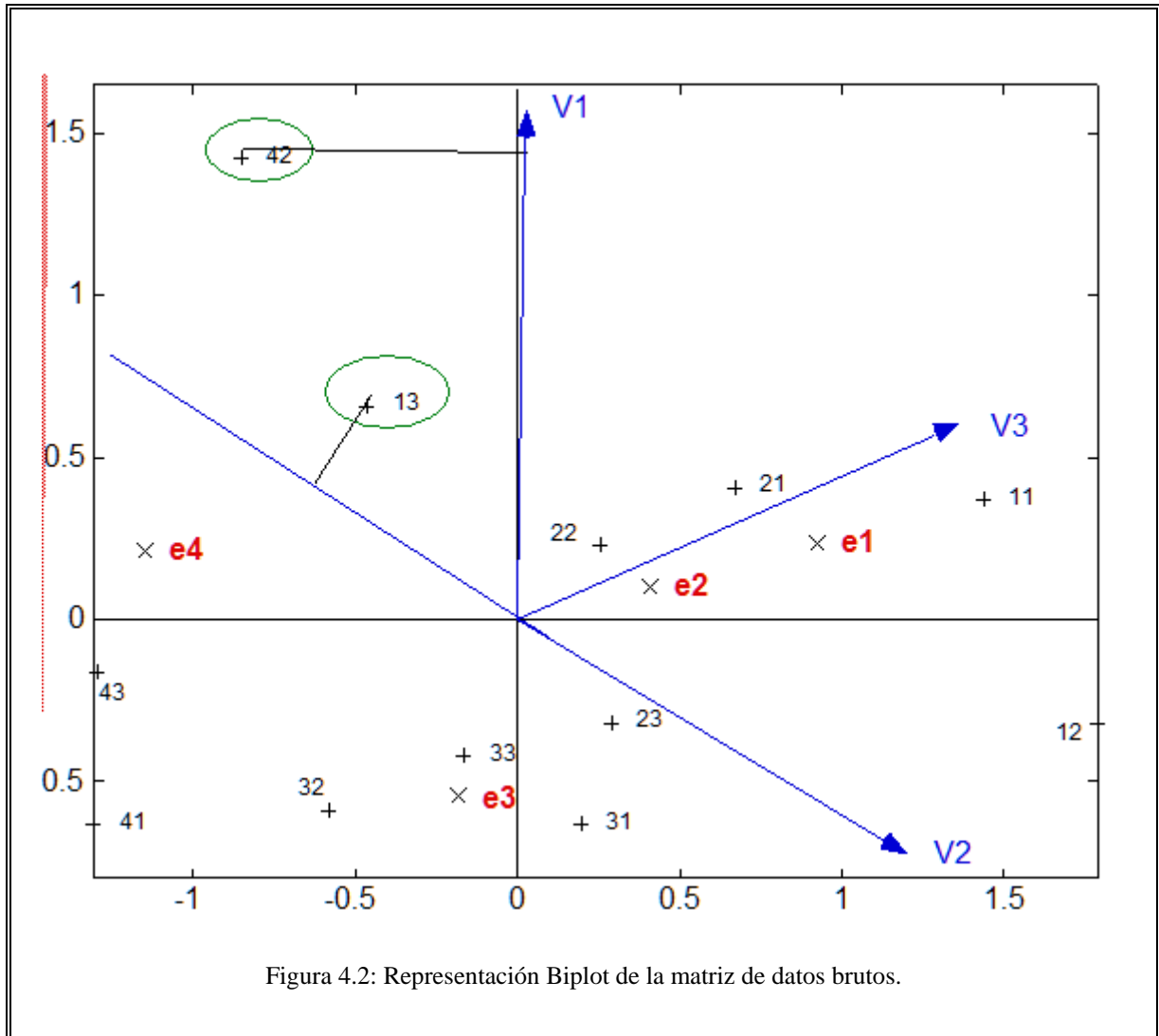


Figura 4.2: Representación Biplot de la matriz de datos brutos.

b) Supongamos ahora que se han corregido las incongruencias de los jueces y se ha obtenido las valoraciones normalizadas (Tabla 4.4).

Estudios	Jueces	variables			etiqueta
		v1	v2	v3	
<i>e1</i>	<i>1</i>	3	4	5	11
	<i>2</i>	2	5	5	12
	<i>3</i>	3	4	3	13
<i>e2</i>	<i>1</i>	3	3	4	21
	<i>2</i>	3	3	3	22
	<i>3</i>	1	2	4	23
<i>e3</i>	<i>1</i>	2	4	2	31
	<i>2</i>	2	3	1	32
	<i>3</i>	2	3	2	33
<i>e4</i>	<i>1</i>	1	1	1	41
	<i>2</i>	1	1	2	42
	<i>3</i>	2	1	1	43

Tabla 4.4: Tabla de datos de las valoraciones de los jueces después de haber corregido las incongruencias.

El nuevo ejemplo se ha construido para mostrar el aspecto que tendría la representación en el supuesto de que todos los jueces hagan juicios similares, que es lo que se espera en un meta-análisis real.

La bondad de ajuste de la matriz de medias es ahora 99.8731% y la bondad de ajuste de la matriz completa es 91.2542%.

En la figura 4.3 se observa el posicionamiento de los diferentes estudios en función de la calidad de los mismos, los jueces próximos a los estudios y las variables o ítems responsables de esta configuración.

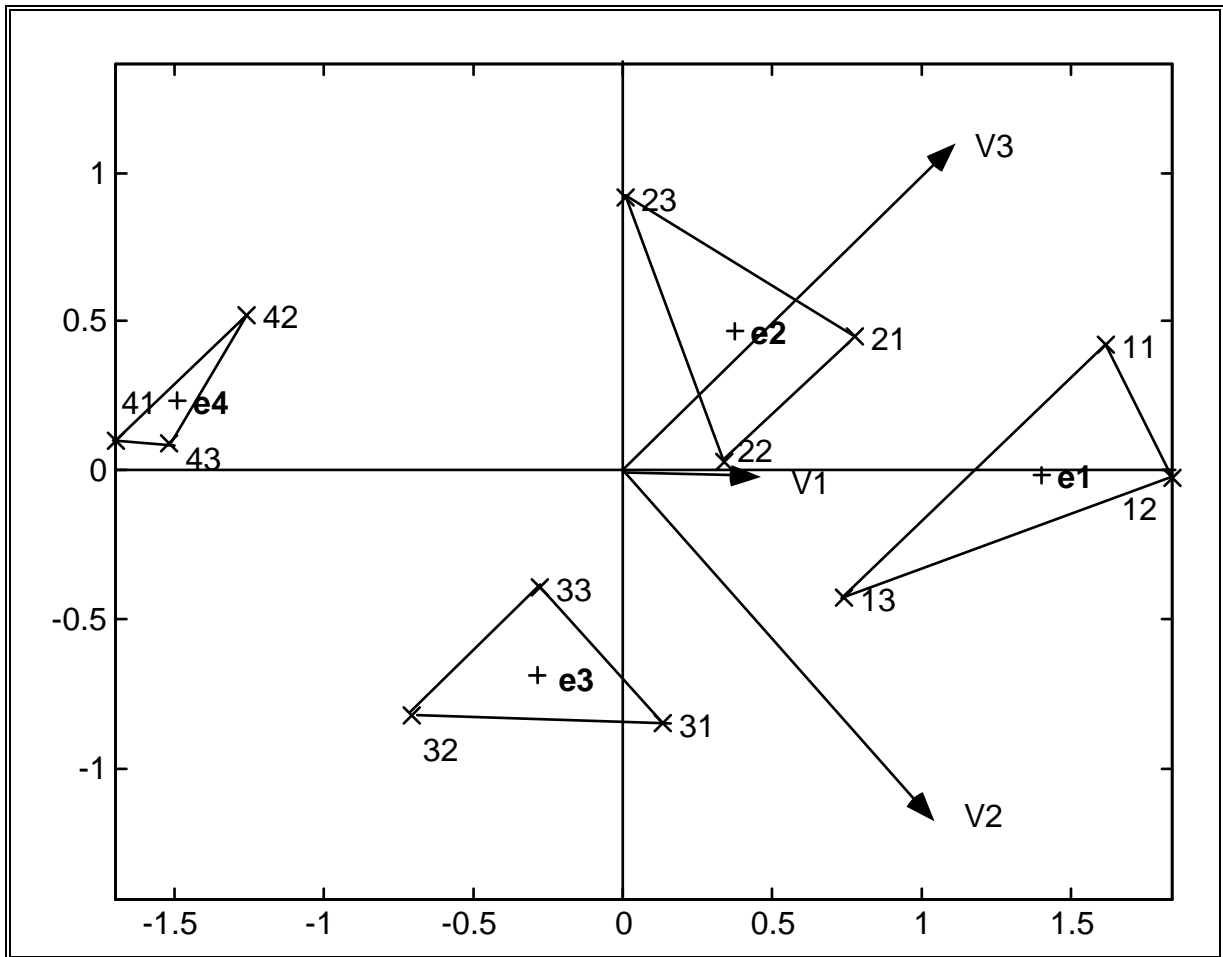


Figura 4.3 : Representación biplot de las medias de la matriz corregida.

CAPITULO 5

MEDIDA DE LOS RESULTADOS

5.1.INTRODUCCION

En una investigación primaria la unidad de análisis de datos la constituyen las observaciones empíricas, pero en una revisión cuantitativa la unidad de análisis procede de los estadísticos ya analizados en los estudios primarios.

Algunos autores opinan que los datos del meta-análisis deben ser los mismos estudios; es decir, cada estudio debe estar representado por un solo dato en el meta-análisis para evitar el problema de dependencia entre los datos. Esto podría producir una cierta ambigüedad puesto que un solo informe de investigación suele contener múltiples contrastes relacionados con una misma hipótesis.

Otros adoptan una postura más flexible que consiste en considerar que un mismo estudio puede adoptar varios resultados interesantes y por tanto, ser introducidos por separado en el meta-análisis.

Para poder integrar los resultados de un conjunto de estudios es preciso que se expresen en una escala de medida común. En general, las variables dependientes empleadas en los estudios difieren entre si, razón por la cual no son directamente comparables.

Este problema se ha intentado solucionar mediante la aplicación de tres estrategias diferentes:

- a) El uso del "vote counting"
- b) El uso de niveles de significación.
- c) El uso de medidas de la magnitud del efecto

5.2. "VOTE COUNTING".

Es una primera aproximación descrita por LIGHT y SMITH (1971) y utilizada en las revisiones tradicionales para valorar los resultados cuantitativos. Consiste en clasificar los resultados disponibles dentro de tres categorías.

- 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.

Si la mayoría de los resultados pertenecen a una de las tres categorías y pocos estudios a las otros dos, entonces la categoría modal es declarada como ganadora. *Esta categoría modal es asumida como la mejor estimación de la verdadera relación entre la variable dependiente y la independiente.*

En particular, el efecto se considera mayor que cero cuando una mayoría de trabajos, por ejemplo más de un tercio, presentan resultados estadísticamente significativos en una dirección.

HEDGES y OLKIN (1985) comentan las deficiencias de este método y plantean los problemas que surgen cuando los investigadores desean integrar K estudios, los cuales proporcionan un estadístico T (ej.: una t de Student o un coeficiente de correlación). Partiendo del supuesto de la existencia del efecto, calculan la probabilidad de que el 'vote-counting' detecte un efecto para varios tamaños de muestra (n) y tamaños del efecto (δ) utilizando el criterio típico de $1/3$. Estas probabilidades son tabuladas y se observa que cuando el tamaño del efecto es pequeño, la probabilidad de equivocarse con este método es próxima a uno.

5.3.EL NIVEL DE SIGNIFICACION.

Los niveles de probabilidad han sido utilizados en la integración de resultados debido a su amplia difusión en los trabajos e informes de investigación. Si disponemos del nivel de significación exacto p que corresponde a cada contraste de significación realizado acerca de una hipótesis común, el conjunto de niveles p unilaterales así obtenido es susceptible de ser analizado estadísticamente ya que convierte los resultados de los estudios a una misma medida.

Esta estrategia presenta, sin embargo, varios inconvenientes; quizás el más claro sea que no porta información en relación a la magnitud del efecto experimental y la relación existente entre las pruebas estadísticas y los tamaños muestrales, por lo tanto, no es aconsejable utilizarlo en un meta-análisis si no va acompañado de alguna medida del tamaño del efecto. (HUNTER y cols., 1982; ROSENTHAL, 1978, 1984, WOLF, 1986)

5.4.EL TAMAÑO DEL EFECTO

GREEN y HALL (1984) afirman que la medida más informativa del resultado de una investigación es el tamaño o magnitud del efecto; fué GLASS (1976) el principal responsable de la introducción de las medidas del tamaño del efecto (TE) a la integración cuantitativa de resultados de investigación.

Cuando la hipótesis nula es falsa, lo es en algún grado específico; es decir, el tamaño del efecto es un valor concreto, distinto de cero, de la población.

Cuanto mayor es este valor, tanto mayor es el grado en que se manifiesta el fenómeno bajo estudio.

El nivel de significación estadística sólo permite determinar si un efecto experimental es distinto de cero; el tamaño del efecto permite conocer en qué medida es distinto de cero (COOPER, 1984).

Una amplia gama de indicadores del tamaño del efecto pueden encontrarse en la literatura, tanto paramétricos, como no paramétricos.(FRIEDMAN, 1978; FLEISS, 1969, SERLIN y cols., 1982; PEECE, 1983)

Los dos tipos de índices más utilizados en meta-análisis son los basados en la diferencia media entre dos condiciones experimentales y los basados en el coeficiente de correlación. En el campo de la medicina y más concretamente en los ensayos clínicos dos de las medidas más utilizadas, a parte de la diferencia de medias, como indicadores del tamaño del efecto son el riesgo relativo (odds-ratio) y la diferencia de proporciones.

5.4.1. LA DIFERENCIA MEDIA TIPIFICADA.

Partimos de la comparación de dos grupos experimentales (ej.: un grupo tratamiento frente a un control) donde suponemos que las observaciones de ambos grupos siguen distribuciones Normales con media μ_T y μ_C respectivamente, y varianza común σ .

COHEN (1967) definió la diferencia media tipificada δ como:

$$\delta = \frac{\mu_T - \mu_C}{\sigma}$$

La diferencia de medias es realmente el valor que define el tamaño del efecto, sin embargo, sólo podría ser utilizado como tal, en el caso que cada experimento que interviene en la revisión estuviera medido en una misma escala y unidades. Esta situación es bastante difícil de encontrar en el contexto meta-analítico.

Se han propuesto varios estimadores de este parámetro en función de las propiedades que cumplen y del estimador elegido de la varianza. GLASS (1976) propone como estimador, **D**, del parámetro poblacional δ la diferencia de medias muestrales de ambos grupos dividido por la desviación típica del grupo control.

$$\mathbf{D} = \frac{\bar{y}_T - \bar{y}_C}{S_C}$$

La argumentación para el empleo de la desviación típica del grupo control está fundada en el hecho de evitar las posibles incongruencias que se presenten al comparar varios grupos de tratamientos con la misma media, frente a un grupo control.

Si se utiliza el promedio de las desviaciones típicas de los dos grupos o del grupo experimental como denominador, puede dar lugar a tamaños del efecto

diferentes, aun cuando sus diferencias de medias sean iguales. Con la solución de GLASS las estimaciones de D de los distintos tratamientos serían idénticas. Además, evita el obstáculo de la heterogeneidad de varianzas, problema que podría afectar a la interpretación de este tamaño del efecto.

HEDGES (1981), propone un estimador de la diferencia media estandarizada de la población partiendo de los supuestos iniciales y considerado más exacto que el propuesto por GLASS. Este estimador, g , estandariza en relación a la media ponderada de las dos varianzas intragrupo:

$$g = \frac{\bar{y}_T - \bar{y}_C}{S_*}$$

donde

$$S_* = \sqrt{\frac{(n_T - 1)S_T^2 + (n_C - 1)S_C^2}{n_T + n_C - 2}}$$

Tanto D como g se distribuyen, salvo un factor de corrección, aproximadamente como una distribución t no central, así cuando multiplicamos estos estimadores por $\sqrt{n^*}$, cada uno sigue una distribución t no central con parámetro $\sqrt{n^*} \delta$ y sus respectivos grados de libertad n_{C-1} y $n_T + n_C - 2$.

Siendo n^* :

$$n^* = \frac{n_T n_C}{n_T + n_C} \quad 1$$

Partiendo de esta distribución, HEDGES (1981, 1982) demuestra que el estimador g está sesgado y que dicho sesgo se hace mayor cuando consideramos el estimador D y propone un estimador (d) insesgado, suficiente y con varianza mínima.

$$d = c(m)g$$

cuya varianza es:

$$\sigma_d^2 = \frac{c(m)^2 m (1 + n^* \delta^2)}{(m - 2)n^*} - \delta^2$$

donde $m = n_T + n_C - 2$ y la constante $c(m)$ está definido como:

$$c(m) = \frac{\Gamma\left(\frac{m}{2}\right)}{\sqrt{\frac{m}{2}} \Gamma\left(\frac{m-1}{2}\right)} \quad 2$$

1

$$\left. \begin{array}{l} Y_T \equiv N(\mu_T, \sigma) \\ Y_C \equiv N(\mu_C, \sigma) \end{array} \right\} \longrightarrow \left. \begin{array}{l} \bar{Y}_T \equiv N\left(\mu_T, \frac{\sigma}{\sqrt{n_T}}\right) \\ \bar{Y}_C \equiv N\left(\mu_C, \frac{\sigma}{\sqrt{n_C}}\right) \end{array} \right\}$$

$$\frac{\bar{Y}_T - \bar{Y}_C}{\sigma} \equiv N\left(\delta, \sqrt{\frac{1}{n_T} + \frac{1}{n_C}}\right) \longrightarrow \frac{\bar{Y}_T - \bar{Y}_C}{\sigma} \sqrt{\frac{n_T n_C}{n_T + n_C}} \equiv N(\delta \sqrt{n^*}, 1)$$

Si consideramos la definición de la distribución t no central con parámetro δ :

$$t = \frac{Y}{\sqrt{X/v}} \left\{ \begin{array}{l} Y \equiv N(\delta, 1) \\ X \equiv \chi^2 \text{ con } v \text{ g.l.} \end{array} \right.$$

entonces el estimador propuesto por Hedges se distribuye aproximadamente, salvo el factor de corrección, como una t no central de parámetro d y $(n_T + n_C - 2)$ grados de libertad.

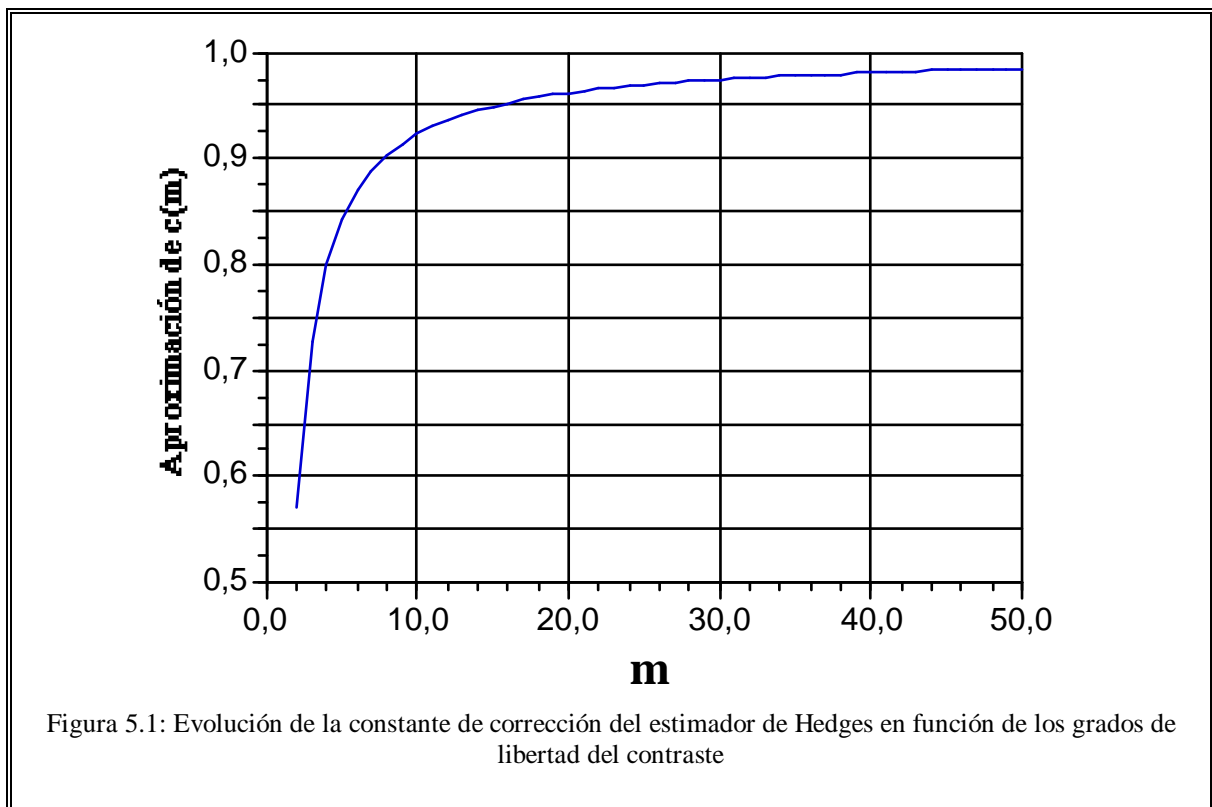
2

Γ = Función gamma

Una aproximación del valor de $c(m)$ es:

$$c(m) \approx 1 - \frac{3}{4m - 1}$$

Obviamente, cuando los tamaños muestrales se van incrementando, entonces, ambos estimadores propuestos por Hedges toman valores muy similares ya que el valor de $c(m)$ tiende a la unidad (figura 5.1).



El estimador d presenta varios problemas a la hora de su utilización en las revisiones cuantitativas. En primer lugar, la dificultad en el cálculo del propio estimador, así como el de su varianza; un segundo problema relacionado con el anterior, es la dificultad que supone realizar inferencia sobre el parámetro estudiado y por último, el inconveniente de la dependencia de la varianza del propio parámetro que se está estimando.

Para evitar algunos de estos impedimentos, HEDGES (1981) afirma que el estimador \mathbf{d} se distribuye asintóticamente con una distribución normal $\mathbf{N}(\delta, \sigma_d)$ cuando los tamaños muestrales son grandes,

donde:

$$\sigma_d^2 = \frac{\mathbf{n}_T + \mathbf{n}_C}{\mathbf{n}_T \mathbf{n}_C} + \frac{\delta^2}{2(\mathbf{n}_T + \mathbf{n}_C)}$$

A partir de esta distribución podemos calcular el intervalo de confianza clásico para el parámetro de la diferencia media estandarizada, si bien, no se ha resuelto un problema, la varianza depende de parámetro que queremos estimar.

La solución es tomar un estimador de la varianza considerando en su cálculo el estimador \mathbf{d} :

$$\hat{\sigma}_d^2 = \frac{\mathbf{n}_T + \mathbf{n}_C}{\mathbf{n}_T \mathbf{n}_C} + \frac{\mathbf{d}^2}{2(\mathbf{n}_T + \mathbf{n}_C)}$$

Por lo tanto, el intervalo de confianza es de la forma:

$$\mathbf{I}_\delta^{1-\alpha} = \left[\mathbf{d} \pm z_{\alpha/2} \hat{\sigma}_d \right]$$

5.4.2. COMPARACION DE ESTIMADORES.

Los estimadores descritos hasta el momento dependen de aproximaciones a distribuciones normales cuando los tamaños muestrales son grandes. HEDGES (1982) realiza un estudio de simulación para valorar la exactitud de la aproximación del estimador (\mathbf{d}) para muestras pequeñas.

En este sentido, se ha realizado un estudio de características similares para

conocer las diferencias entre el estimador insesgado propuesto por Hedges y el estimador sesgado mediante una aproximación para muestras grandes a la distribución normal original cuya varianza es:

$$\sigma_g^2 = \frac{n_T + n_C}{n_T n_C}$$

El estudio de simulación se fundamenta en la obtención de un conjunto de muestras aleatorias a partir de poblaciones infinitas con distribución normal (control y tratamiento) que configuran diferentes tamaños del efecto poblacionales δ (0.25, 0.5, 0.75, 1, 1.5) y con distintos tamaños de muestra (20, 30, 50, 80, 120). Se establece el supuesto que tanto el grupo control como el tratamiento presentan el mismo número de observaciones. En cada una de las muestras se calculan los estimadores \mathbf{g} y \mathbf{d} así como sus respectivas varianzas. El objetivo radica en la comparación de los porcentaje de intervalos, calculados para tres diferentes niveles de confianza (0.90, 0.95, 0.99), que no contienen al parámetro poblacional.

En una primera fase se realizó un estudio de estabilidad para determinar el número de muestras idóneo para optimizar la fiabilidad del estudio y el tiempo invertido en la generación de las muestras debido a las limitaciones de capacidad de los paquetes informáticos. Se plantearon varias pruebas con diferentes números de muestras y para los dos tamaños del efecto más alejados.

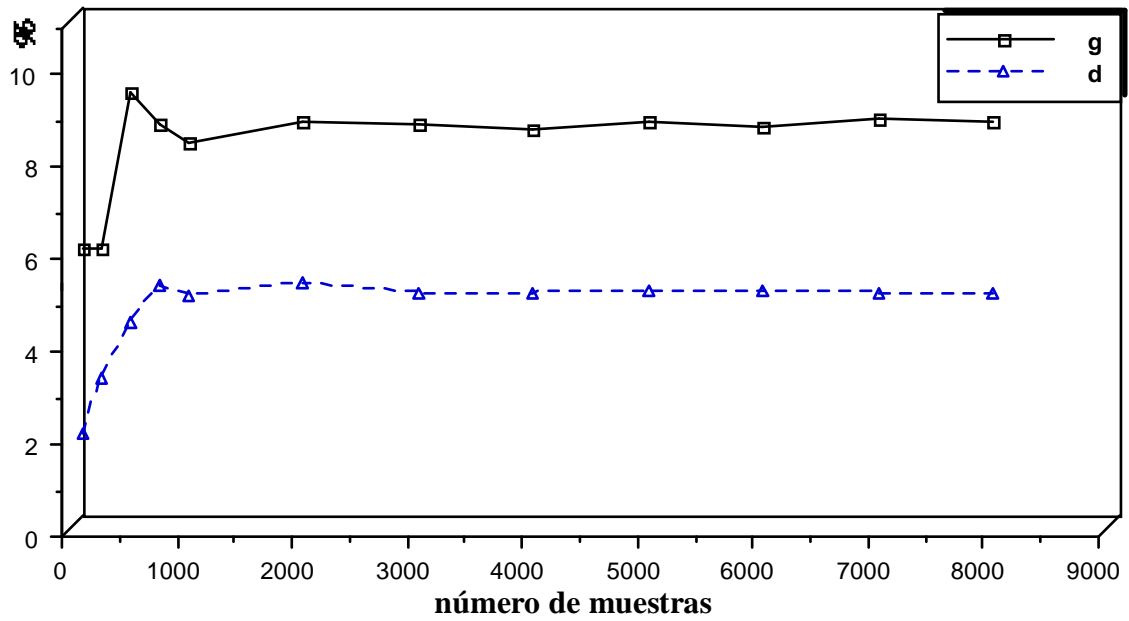
En las figuras 5.2.a y 5.2.b se puede observar como a partir de las 1000 muestras se estabiliza la estimación del parámetro poblacional, si bien, no se descarta la posibilidad de obtener un porcentaje muy pequeño de resultados aberrantes. Una vez establecido el número de muestras que intervienen en el estudio se calculan los intervalos de confianza para obtener el porcentaje de

intervalos que no contienen a la diferencia de medias estandarizada poblacional.

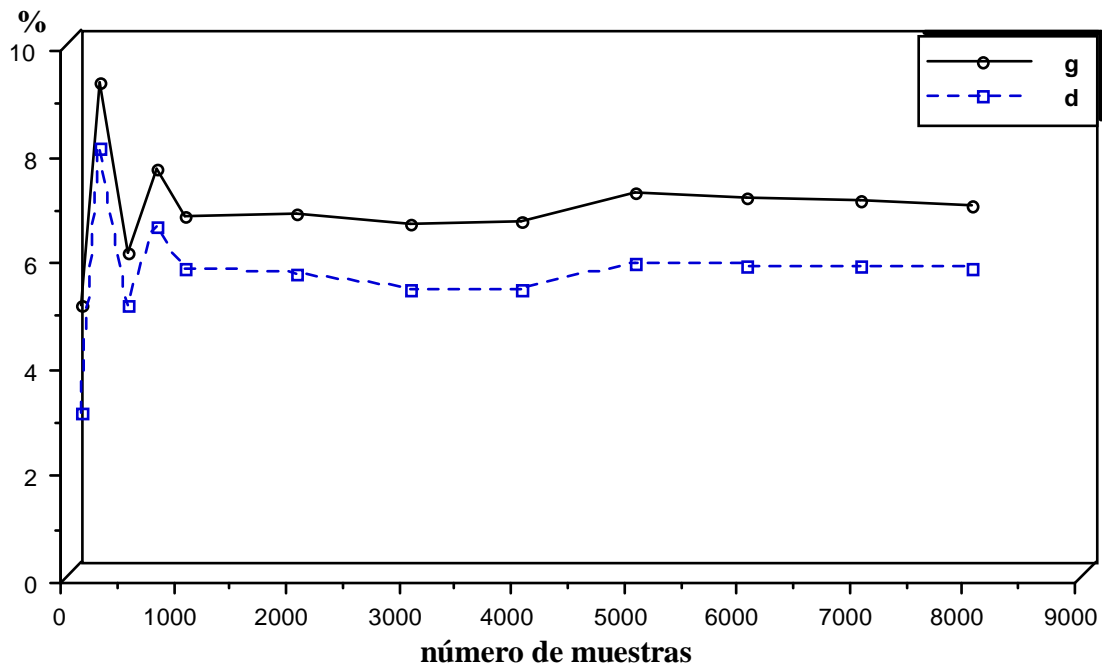
Los resultados a partir de el estimador g (tabla 5.1) arrojan un alejamiento de los porcentajes con respecto a los niveles de significación considerados. Además, según se incrementa el tamaño del efecto la diferencia aumenta de forma que cuando el tamaño del efecto toma el valor de 1.5 las diferencias son importantes.

Así el 60% de los resultados obtenidos para el niveles de significación 0.05 presentan porcentajes situados entre dos puntos de distancia del 5%. Este porcentaje se incrementa al 72% cuando se utiliza el nivel de significación del 0.01 y disminuye al 44% al 0.10.

Estas discrepancias a los diferentes niveles de confianza se deben a la acentuación o amortiguación del sesgo por el decremento o incremento de la amplitud del intervalo.



(a)



(b)

Figura 5.2: Estudio de estabilidad para la elección del número de muestras en la simulación. Se representa en eje de ordenadas el porcentaje de intervalos que no contienen a los parámetros poblacionales 1.5 (a), 0,25 (b) y en eje de abscisas el número de muestras.

Cuando se analizan los resultados basados en el estimador de Hedges (tabla 5.2) se advierte que los porcentajes calculados se encuentran próximos a los niveles de significación. Así, casi la totalidad de los porcentajes para los distintos niveles de significación se encuentra situados a en la franja de dos puntos de sus respectivos valores de significación.

Tamaño muestral	Media	Varianza	Proporción de intervalos que no contienen a δ a nivel de significación fijo		
			0,10	0,05	0,01
$\delta=0,25$					
20	0,274	0,229	11,5	6,7	1,8
30	0,252	0,145	10,7	6,5	1,7
50	0,250	0,078	9,9	4,5	1,1
80	0,258	0,053	10	4,6	1,4
120	0,258	0,035	10,6	6	1,2
$\delta=0,50$					
20	0,530	0,262	13,4	8,9	3,5
30	0,532	0,153	11,8	6,8	1,9
50	0,512	0,084	11,9	5,3	1,0
80	0,555	0,045	10	4,6	1,1
120	0,503	0,033	8,7	4,5	1,0
$\delta=0,75$					
20	0,776	0,261	14,7	8,6	2,9
30	0,772	0,158	12,6	6,8	2,5
50	0,763	0,095	12	6,6	2,6
80	0,744	0,053	12,3	6,3	1,5
120	0,752	0,038	12,5	6,1	2,4
$\delta=1,00$					
20	1,053	0,271	13,8	8,8	3,3
30	1,021	0,178	13,8	8,7	3,4
50	1,023	0,097	12,4	7,9	2,7
80	1,011	0,057	12,3	6,3	1,5
120	1,010	0,035	9,8	4,7	1,2
$\delta=1,50$					
20	1,559	0,293	15,8	9,7	3,5
30	1,535	0,188	16,8	11,1	3,1
50	1,525	0,115	16,4	10,2	3,5
80	1,500	0,068	14,8	9,4	3,0
120	1,507	0,043	12,7	8,3	2,3

Tabla 5.1: Porcentaje de intervalos que no contienen al tamaño del efecto poblacional señalado a diferentes tamaños muestrales y niveles de significación. El estimador utilizado es g , y la varianza se considera la aproximación normal para grandes muestras.

Tamaño muestral	Media	Varianza	Proporción de intervalos que no contienen a δ a nivel de significación fijo		
			0,10	0,05	0,01
$\delta=0,25$					
20	0,262	0,21	10,6	5,7	1,8
30	0,245	0,138	9,4	5,8	1,3
50	0,246	0,075	9,8	5,3	1,1
80	0,255	0,052	9,8	4,4	1,4
120	0,257	0,034	10,2	5,6	1,1
$\delta=0,50$					
20	0,508	0,241	12,1	6,8	3,1
30	0,052	0,145	10,4	5,9	1,5
50	0,504	0,082	6,1	4,4	0,08
80	0,549	0,044	8,9	4,1	1,0
120	0,499	0,033	7,7	4,1	1,0
$\delta=0,75$					
20	0,743	0,239	12,4	6,4	2,0
30	0,751	0,150	10,2	5,3	1,6
50	0,750	0,092	9,8	5,6	1,8
80	0,737	0,052	9,6	4,7	1,0
120	0,748	0,038	11,3	5,6	1,6
$\delta=1,00$					
20	1,009	0,249	9,9	6,6	1,7
30	0,993	0,168	11,3	6,6	2,2
50	1,006	0,094	10,8	6,0	1,8
80	1,001	0,056	9,5	4,7	1,0
120	1,004	0,034	8,6	3,6	0,8
$\delta=1,50$					
20	0,149	0,269	10	5,3	1,5
30	1,494	0,178	11,8	5,4	0,9
50	1,500	0,111	11,3	6,0	1,3
80	1,485	0,067	9,6	5,7	1,3
120	1,498	0,042	10,3	5,0	1,3

Tabla 5.2: Porcentaje de intervalos que no contienen al tamaño del efecto poblacional señalado a diferentes tamaños muestrales y niveles de significación. Se utiliza el estimador y varianza propuesto por Hedges.

En la figura 5.3 se pone de manifiesto las diferencias existentes entre los porcentajes calculados a través de ambos estimadores respecto al nivel de significación del 5%. Resaltar el incremento de estas diferencias para el estimador no corregido con tamaños del efecto de gran magnitud (1-1.5) los cuales alcanzan un aumento de hasta seis puntos en relación al nivel considerado. Para tamaños muestrales pequeños en cada estudio (20-30 individuos) las discrepancias también son substanciales. En el gráfico correspondiente al estimador insesgado de Hedges (figura 5.3.b) estas diferencias se aminoran presentando valores cercanos al 5%, si bien, las discrepancias son más notables para tamaños muestrales de 20 individuos y efectos de pequeña magnitud. En la figura 5.4. se comparan los porcentajes procedentes del uso de ambos estimadores donde se advierte con mayor claridad el incremento de las discrepancias entre ambos estimadores con la acentuación del efecto del tratamiento y cuando los tamaños muestrales son pequeños.

Los resultados aportados por la simulación nos muestra que *el estimador de Hedges proporciona mejores aproximaciones cuando el tamaño del efecto es grande o bien cuando el número de individuos de cada estudio es pequeño*, si bien, para este último suceso este estimador tampoco proporciona valores muy ajustados. Sin embargo, el estimador de Hedges presenta una serie de inconvenientes que pueden afectar al estimador global. El más importante, como se ha apuntado con anterioridad, se refiere a la dependencia de la varianza del parámetro poblacional, difícilmente disponible para el revisor. Los resultados arrojados por la simulación proceden de una estimación exacta de la varianza y por lo tanto sólo proporcionan información a nivel teórico. En la práctica el estimador de la varianza se calcula a partir del valor obtenido del tamaño del efecto lo que se manifestará en un sesgo sobre la estimación global.

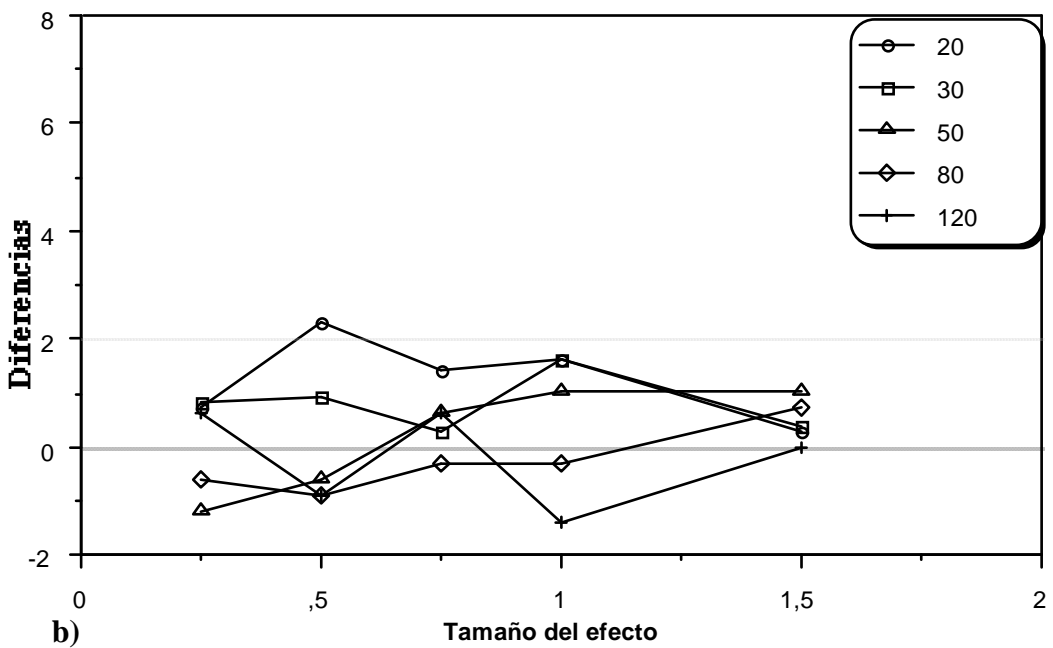
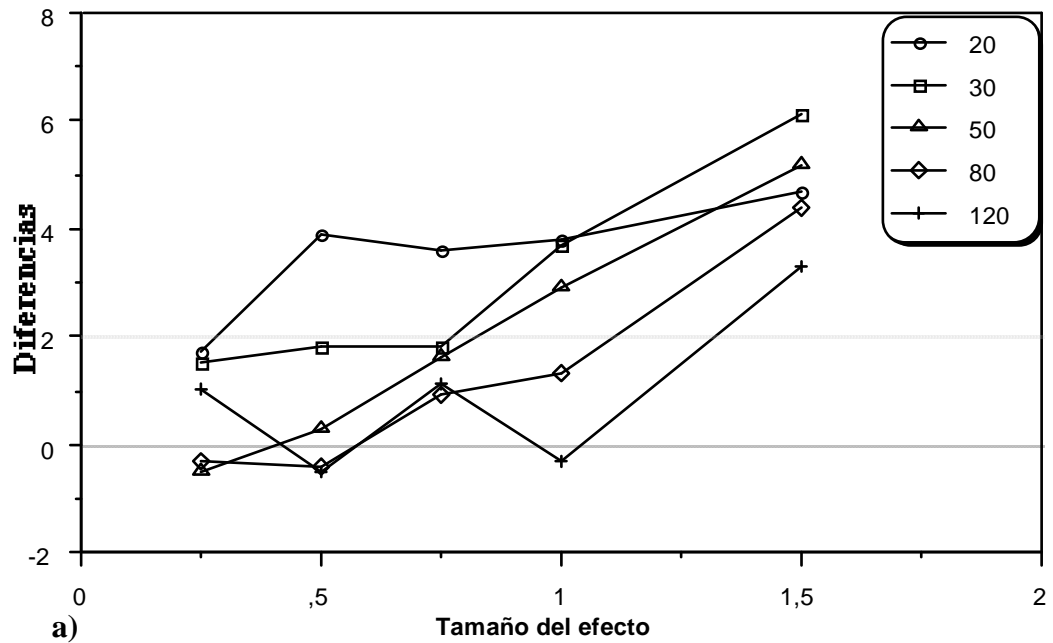
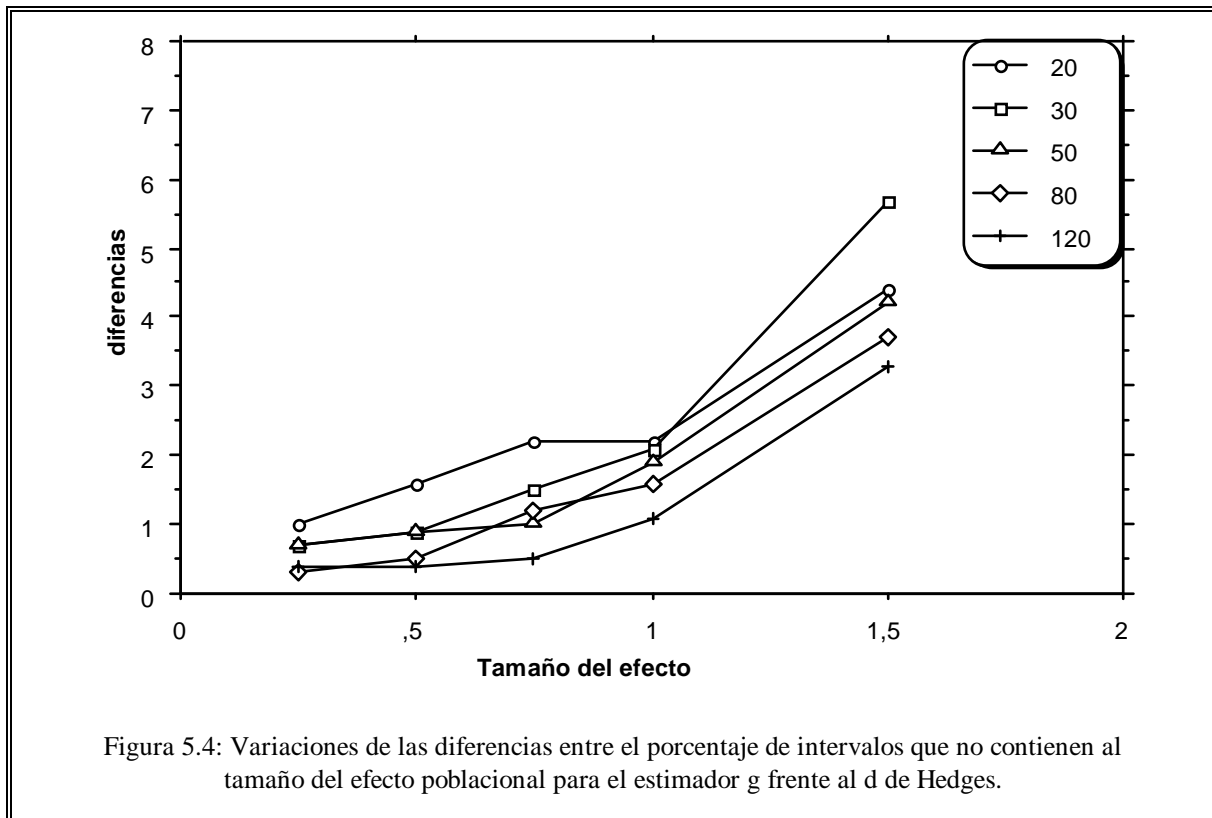


Figura 5.3: Variaciones de las diferencias entre el porcentaje de intervalos que no contienen al tamaño del efecto poblacional al nivel de confianza del 95% y el porcentaje esperado (5%) tanto para el estimador no corregido (a) como para el corregido (b)



El caso mas claro se presenta cuando los estudios basados en pocas observaciones arrojan grandes efectos para los tratamientos; por consiguiente, al calcular el estimador global a través de un sistema de ponderaciones basadas en el inverso de la varianza, esta clase de estudios presentan ponderaciones infravaloradas que aportan una estimación sesgada del tamaño del efecto. En este sentido, BEGG y BERLIN (1988), a partir de revisiones como la de MOERTEL (1984) y, POCOCK y cols. (1986), apuntan la influencia del tamaño de las muestras en relación con el tamaño del efecto, de manera que estudios basados en un escaso número de observaciones producen efectos de gran magnitud.

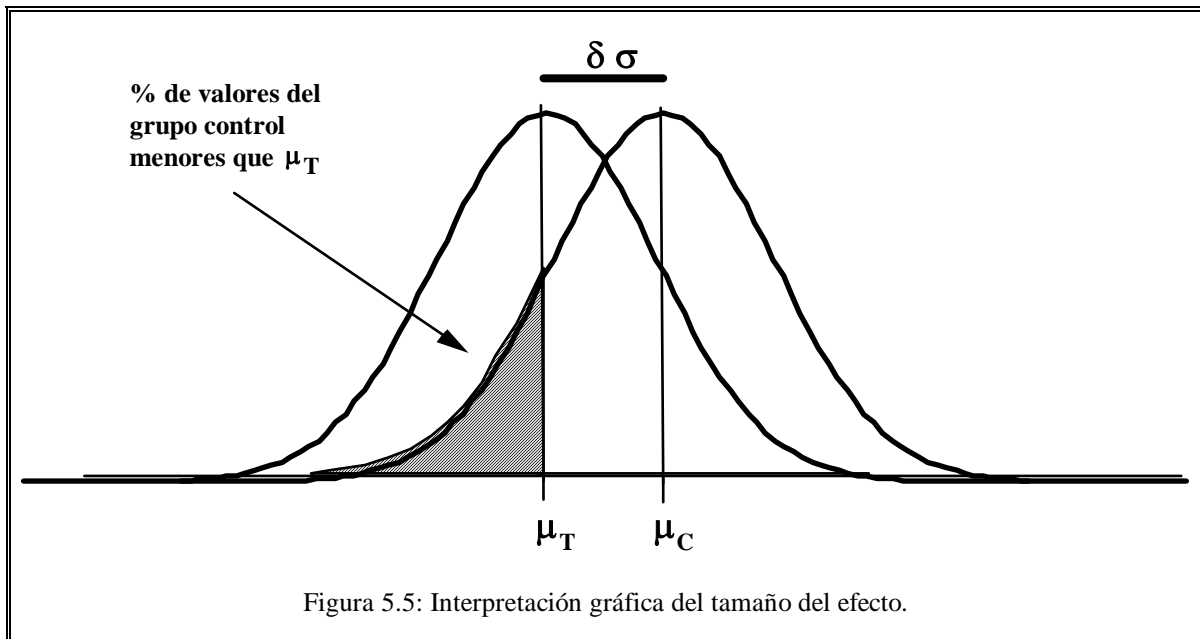
Por todo lo expuesto consideramos de mayor utilidad el empleo del estimador **g** siempre que el investigador considere que el efecto con el que se van a enfrentar es moderado o pequeño y los tamaños muestrales globales superan las 30 observaciones. En caso contrario, este estimador produce estimaciones con

sesgos grandes y el uso del estimador de Hedges proporciona óptimos resultados, si bien se tendrá que realizar un seguimiento individualizado de cada resultado extraído para evitar los posibles efectos debidos a la dependencia de la varianza.

Algunos autores (HEDGES y OLKIN, 1983 a y b; KRAEMER, 1983) para evitar estos problemas proponen una serie de transformaciones, de la forma del seno hiperbólico o transformación en funciones del coeficiente de correlación, que permitan la estabilización de la varianza, pero que no parecen haber tenido gran aceptación en la bibliografía.

Otro estimador de la diferencia media tipificada es el obtenido mediante el método de máxima verosimilitud, el cual proporciona estimadores asintóticamente eficientes, consistentes e invariantes (HEDGES y OLKIN, 1985).

La elección y difusión de la diferencia media tipificada es debida a su sencillez, y principalmente a sus posibles interpretaciones útiles en los diferentes campos de investigación. GLASS (1976) considera la interpretación de el tamaño del efecto como el grado de solapamiento existente entre ambas distribuciones. Así, **D** es el valor estandarizado de la media del grupo experimental en la distribución del grupo control, indicando qué porcentaje de valores del grupo control son más o menos efectivos que el valor medio del tratamiento (figura 5.5). Otra interpretación es la comparación directa en número de veces que es mayor o menor este estadístico cuando se estudian una serie de tratamientos frente a un control o un grupo referencia.



Uno de los problemas con el que se enfrenta el revisor a la hora de utilizar la diferencia media tipificada en la integración de resultados, es la dificultad de encontrar este estadístico en los artículos o informes de investigación. A partir de la información básica como el tamaño muestral, media y varianza de los diferentes grupos se puede calcular sin ninguna dificultad. Sin embargo, en varios trabajos esta información no se encuentra explícitamente y tenemos que recurrir a simples transformaciones para obtener la diferencia media tipificada.

Cuando el diseño de la investigación incluye solo dos grupos (experimental y control) el estadístico de contraste más común es la t .

Es posible estimar g si al menos el estudio proporciona el valor de t y los tamaños muestrales de ambos grupos.

- Para **muestras independientes**, la diferencia media tipificada viene dada directamente por la expresión:

$$g = t \sqrt{\frac{1}{n_c} + \frac{1}{n_T}}$$

Si $n_T = n_C$, entonces
$$g = t \sqrt{\frac{2}{n}}$$

- Para **muestras relacionadas**, suponiendo conocido el coeficiente de correlación r_{XY} ,

$$g = t \sqrt{\frac{2}{n} (1 - r_{XY}^2)}$$

En varias ocasiones los investigadores no aportan el valor de t sino sólo el nivel de significación alcanzado y los grados de libertad. Si el nivel de significación es exacto, puede obtenerse el valor del estadístico de contraste consultando las tablas de la distribución correspondiente y calculando g mediante las relaciones descritas más arriba.

Cuando el diseño experimental incluye **más de dos grupos** suele utilizarse como estadístico de contraste la prueba F de Snedecor. Un diseño de este tipo es susceptible de asimilar tantas estimaciones de g como comparaciones dos a dos admita el diseño; o bien, puede obtenerse un valor g para la comparación de cada grupo experimental con un grupo de control común.

En cualquier caso, la información mínima necesaria para su cálculo incluye las medias de cada grupo, los tamaños muestrales y el resultado de la razón F global para poder obtener la desviación típica intragrupos S .

- Si los grupos son independientes, S , se obtiene por

$$S = \sqrt{MC_{\text{error}}}$$

La MC_{error} puede obtenerse mediante

$$MC_{\text{error}} = \frac{MC_{\text{trat.}}}{F}$$

siendo

$$MC_{\text{trat.}} = \frac{\sum_j n_j (y_{.j} - y_{..})^2}{r - 1}$$

donde n_j es el tamaño muestral del grupo j , r es el número de grupos, $y_{.j}$ es la media del grupo j e $y_{..}$ es la media global.

- La obtención de la desviación típica intra-grupo cuando disponemos de un diseño de un factor A con medidas repetidas requiere que conozcamos las fuentes de variación (ANOVA) correspondientes. Una estimación de S consiste en combinar la variación debida a los sujetos con la variación debida a la interacción del factor con los sujetos mediante:

$$S = \sqrt{\frac{(SC_{\text{intersujetos}} + SC_{A \times \text{sujetos}})}{g.l. \text{ intersujetos} + g.l. A \times \text{sujetos}}}$$

5.4.3 EL COEFICIENTE DE CORRELACION

Un índice alternativo a la diferencia media tipificada para estimar el tamaño del efecto (TE) es el coeficiente de correlación, ampliamente utilizado para conocer el grado de asociación entre dos variables distribuidas normalmente.

HUNTER y cols. (1982) y, ROSENTHAL (1984) recomiendan el uso de este coeficiente en meta-análisis debido a que es invariante por transformaciones lineales, y a una mayor accesibilidad que la diferencia media tipificada la cual no suele estar incluida en los informes de investigación. Por lo tanto es un buen candidato del tamaño del efecto a través de los diversos estudios.

Cuando el coeficiente de correlación es diferente de cero su distribución se alejará tanto mas de la normal cuanto mas pequeño sea el tamaño de muestra n y su varianza dependerá del coeficiente poblacional. Así, cuando los estudios integrados no tienen todos el mismo tamaño muestral los errores de las varianzas individuales pueden diferir en alto grado.

Debido a estos inconvenientes se aconseja utilizar la transformación Z (FISHER, 1921) que resuelve este problema:

$$Z_i = \frac{1}{2} \log \left(\frac{1 + r_i}{1 - r_i} \right)$$

La Z -transformación estabiliza la varianza y cuando el tamaño muestral es grande Z sigue aproximadamente una $N(\zeta, 1/n)$.

Una aproximación más exacta, cuando tenemos valores moderados de n es la utilización de la varianza asintótica $1/n-3$ en lugar de $1/n$.

Por lo tanto el intervalo de confianza será:

$$I_{Z_p}^{1-\alpha} = \left[\hat{z}_p \pm \frac{Z_{\alpha/2}}{\sqrt{n-3}} \right]$$

Al investigador realmente le interesa la estimación para el coeficiente de correlación para sus posibles interpretaciones, por lo tanto, se realiza la transformación correspondiente para obtener los extremos del intervalo en forma de coeficientes de correlación:

$$\rho_I = \frac{e^{2z_p^I} - 1}{e^{2z_p^I} + 1} ; \rho_S = \frac{e^{2z_p^S} - 1}{e^{2z_p^S} + 1}$$

En muchas ocasiones es posible encontrar en los informes de investigación estadísticos tan dispares como la correlación biserial puntual, el coeficiente phi, contrastes ji-cuadrado en tablas de contingencia, pruebas t , análisis de la varianza y otras alternativas.

5.4.4. TAMAÑO DEL EFECTO EN TABLAS 2x2.

En el campo de la Medicina es habitual contrastar si hay relación entre la presencia o ausencia de un factor (un tratamiento o un factor de riesgo) y la presencia o ausencia de una respuesta (una enfermedad, fallecimiento, mejoría, etc.), situaciones cuya información podría sintetizarse en una tabla de este tipo:

		RESPUESTA		
		B (+)	\bar{B} (-)	
FACTOR	A	a (p ₁₁)	b (p ₁₂)	n _{1.} (p _{1.})
	\bar{A}	c (p ₂₁)	d (p ₂₂)	n _{2.} (p _{2.})
		n _{.1} (p _{.1})	n _{.2} (p _{.2})	N

El problema surge a la hora de elegir una medida que proporcione información sobre el tamaño del efecto.

En un principio, cualquiera de los índices de asociación propuestos por los autores serán válidos para tal propósito. Sin embargo, hay que examinar sus características para obtener el estimador más apropiado para su integración dentro de la revisión cuantitativa.

La medida de asociación debe presentar una aceptación y uso generalizado en la literatura científica, una interpretación clara y un conjunto de propiedades que permitan hacer inferencias.

Uno de los candidatos, debido a su interpretación, es el **riesgo relativo** definido como la razón de la ocurrencia de la respuesta B cuando el factor A se encuentra presente y/o ausente:

$$RR = \frac{P(B/A)}{P(B/\bar{A})} = \frac{p_{11}/p_{1.}}{p_{21}/p_{2.}} = \frac{p_{11}p_{2.}}{p_{21}p_{1.}}$$

El estimador se obtiene:

$$rr = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{a \cdot (c+d)}{b \cdot (a+b)}$$

El problema que presenta utilizar el riesgo relativo como tamaño del efecto es su restrictivo campo de aplicación, ya que sólo puede estimarse en estudios prospectivos o de cohortes¹, sin poder ser calculado en los estudios retrospectivos o estudios casos/control²

Un índice alternativo es el **odds-ratio** (CORNFIELD, 1951) propuesto en un principio como una aproximación del riesgo relativo y que se define como la razón de los odds:

$$\left. \begin{array}{l} \Omega_A = \frac{P(B/A)}{P(\bar{B}/A)} \\ \Omega_{\bar{A}} = \frac{P(B/\bar{A})}{P(\bar{B}/\bar{A})} \end{array} \right\} \longrightarrow \psi = \frac{\Omega_A}{\Omega_{\bar{A}}} = \frac{p_{11}p_{22}}{p_{12}p_{21}}$$

Por lo tanto el estimador obtenido es:

$$\hat{\psi} = \frac{a \cdot d}{b \cdot c}$$

Si la proporción de individuos que presentan la respuesta es pequeño, entonces, el odds-ratio es muy próximo al riesgo relativo, por lo que también se le conoce con el nombre de **riesgo relativo aproximado**.

¹ Se parte de dos grupos con y sin el factor presente y se observa el número de individuos que muestran o no la respuesta.

² Se parte de dos grupos con y sin presencia del resultado y se explora que individuos poseen o no el factor de riesgo.

Además, el odds-ratio se puede calcular de la misma forma en los estudios retrospectivos presentando el mismo estimador:

$$\left. \begin{array}{l} \Omega_B = \frac{P(A/B)}{P(\bar{A}/B)} \\ \Omega_{\bar{B}} = \frac{P(A/\bar{B})}{P(\bar{A}/\bar{B})} \end{array} \right\} \longrightarrow \psi = \frac{\Omega_B}{\Omega_{\bar{B}}} = \frac{p_{11}p_{22}}{p_{12}p_{21}} ; \hat{\psi} = \frac{a \cdot d}{b \cdot c}$$

La varianza del odds-ratio, como ocurre en el resto de los estimadores del tamaño del efecto, depende del propio parámetro poblacional. Para solucionar este inconveniente se considera una transformación del tipo de logaritmo neperiano ($\ln \psi = \gamma$) obteniendo la varianza de este nuevo estimador como:

$$\sigma_{\hat{\gamma}}^2 = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

Por lo tanto, el intervalo de confianza en muestras de tamaño grande será:

$$I_{\gamma}^{1-\alpha} = \left[\hat{\gamma} \pm z_{\alpha/2} \sigma_{\hat{\gamma}}^2 \right] \xrightarrow{\text{transformando}} \begin{cases} \psi_I = e^{\left[\hat{\gamma} - z_{\alpha/2} \sigma_{\hat{\gamma}}^2 \right]} \\ \psi_S = e^{\left[\hat{\gamma} + z_{\alpha/2} \sigma_{\hat{\gamma}}^2 \right]} \end{cases}$$

El inconveniente del odds-ratio como tamaño del efecto es su interpretación no tan clara como la del riesgo relativo, si bien se le puede dar un sentido similar.

Otros autores (COHEN, 1967; Der SIMONIAN y LAIRD, 1986) utilizan como estimador del tamaño del efecto la **diferencia de proporciones de ocurrencia** de un evento en dos grupos (ej.: tratamiento y control):

$$\hat{\theta} = \hat{p}_T - \hat{p}_C$$

y varianza:

$$\hat{\sigma}_{\hat{\theta}}^2 = \frac{\hat{p}_T(1-\hat{p}_T)}{n_T} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_C}$$

La ventaja de este tamaño del efecto frente a los propuestos anteriormente como son el riesgo relativo y el odds-ratio es una interpretación más simple y más útil a la hora de intentar conocer el impacto del tratamiento o el factor de riesgo. Además, su presencia es más amplia y generalizada en los diferentes informes de investigación.

Los investigadores, para evitar las posibles interferencias de variables "confounding", emplean ensayos controlados de muestras apareadas (samples matched). En este tipo de ensayos se seleccionan pares de individuos con características similares a partir de los cuales se realiza la asignación aleatoria de los tratamientos.

		CONTROLES		
CASOS		Factor presente	Factor ausente	
Factor presente	a (p ₁₁)	b (p ₁₂)	n_{1.} (p _{1.})	
Factor ausente	c (p ₂₁)	d (p ₂₂)	n_{2.} (p _{2.})	
	n_{1.} (p _{1.})	n_{2.} (p _{2.})	N	

A partir de este tipo de datos el cálculo de los estimadores de los distintos índices considerados como tamaños del efecto es diferente. MANTEL y HAENSZEL (1959) y CORNFIELD y HAENSZEL (1960) estudiaron la forma de obtener una estimación del odds-ratio cuando las muestras son apareadas proponiendo el siguiente estimador:

$$\hat{\psi} = \frac{b}{c}$$

y su desviación estándar:

$$\hat{\sigma}_{\hat{\psi}} = \hat{\psi} \sqrt{\frac{1}{b} + \frac{1}{c}}$$

En el caso de considerar la diferencia de proporciones el estimador de interés será:

$$\hat{\theta} = \hat{p}_T - \hat{p}_C = \frac{b - c}{n}$$

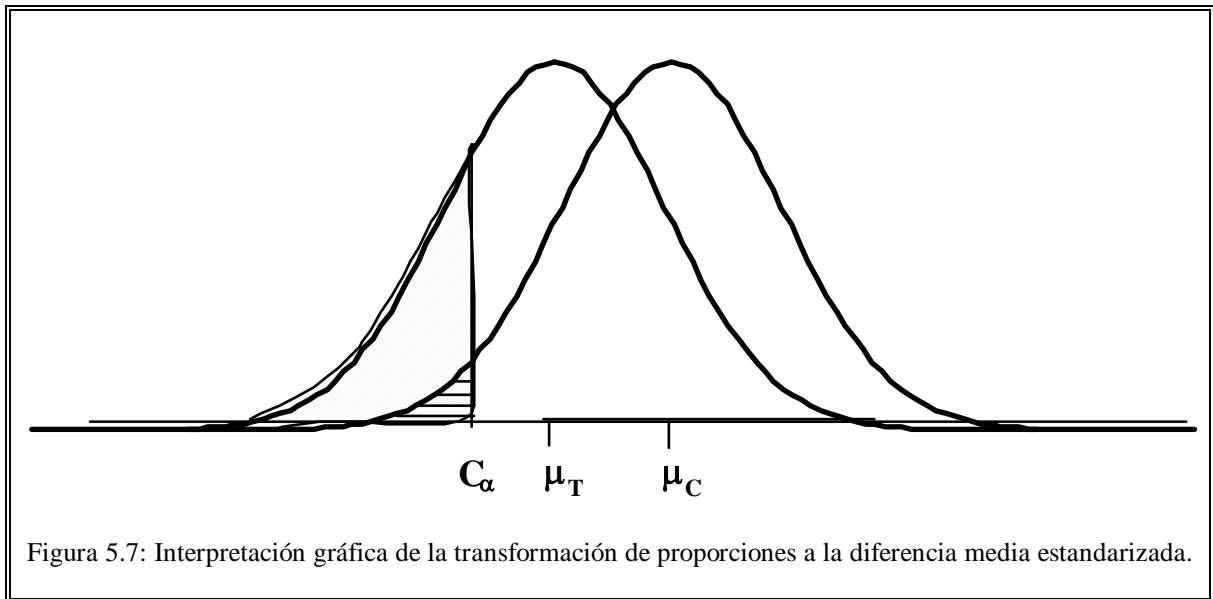
y de varianza:

$$\hat{\sigma}_{\hat{\theta}}^2 = \frac{\sqrt{b+c}}{n}$$

GLASS y cols., 1981 presentan un método para obtener una estimación de la diferencia media tipificada a partir del diseño de tablas 2x2. Utiliza la transformación probit de las proporciones en los valores Z, siempre que pueda considerarse una variable dependiente subyacente con distribución Normal.

$$\left. \begin{array}{l} \hat{p}_T \xrightarrow{\phi^{-1}(p_T)} z_T \\ \hat{p}_C \xrightarrow{\phi^{-1}(p_C)} z_C \end{array} \right\} \longrightarrow z_C - z_T = \hat{\Delta}$$

Consideramos un tratamiento frente a un control donde la variable dependiente se mide de forma dicotómica en mejorar/igual y la clasificación se hace en función de una o varias variables continuas. Se asume la existencia de un punto de corte (C_α) a partir del cual si el valor de un individuo está por debajo nos lleva a declarar la mejoría del individuo (figura5.7):



Por lo tanto, el investigador observa unas proporciones p_t y p_c que transformamos en valores Z , considerando que las varianzas poblacionales son iguales, obtenemos un estimador de la diferencia media tipificada.

$$\left. \begin{aligned} \hat{p}_T &= \int_{-\infty}^{z_T} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \xrightarrow{\phi^{-1}(p)} z_T = \frac{C_\alpha - \bar{X}_T}{S_E} \\ \hat{p}_C &= \int_{-\infty}^{z_C} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \xrightarrow{\phi^{-1}(p)} z_C = \frac{C_\alpha - \bar{X}_C}{S_C} \end{aligned} \right\} \xrightarrow{\text{Varianzas homogéneas}} z_C - z_T = \frac{\bar{X}_T - \bar{X}_C}{S_C} = \hat{\Delta}$$

5.4.5. ESTIMADORES NO PARAMÉTRICOS

Se han estudiado los principales estimadores del tamaño del efecto cuando los supuestos de partida se cumplen en los datos originales. Sin embargo, muchas veces el investigador se encuentra en situaciones donde no es posible aplicar los test clásicos a los datos brutos, con lo cual no sería apropiado la combinación de los tamaños del efecto descritos.

La asimetría de los datos es uno de los principales factores que pueden afectar a los estimadores produciendo distorsiones en la realidad del problema que se está estudiando; por otra parte, si la normalidad no se cumple la interpretación de la diferencia media tipificada de solapamiento de distribuciones normales no tiene sentido.

KRAEMER y ANDREWS (1982) proponen una solución no paramétrica, posteriormente extendida por HEDGES y OLKIN (1984), para resolver los posibles defectos producidos por el desviaciones del supuesto de Normalidad. Los autores desarrollan la teoría en base a los diferentes diseños que puede encontrar el investigador (pre-post intervención, tratamiento-control, pre-post /tratamiento-control) y suponiendo que los estudios disponen de los valores para los individuos antes (x_i) y después (y_i) de la intervención (ej.: un test, administración de un fármaco, un análisis, etc.) tanto en el grupo control como en el grupo tratamiento.

Se considera el diseño más general :

CONTROL		TRATAMIENTO	
PREINT.	POSTINT.	PREINT.	POSTINT.
X_{C1}	Y_{C1}	X_{T1}	Y_{T1}
.	.	.	.
.	.	.	.
.	.	.	.
X_{Ci}	Y_{Ci}	X_{Tj}	Y_{Tj}
.	.	.	.
.	.	.	.
.	.	.	.
X_{Cn}	Y_{Cn}	X_{Tn}	Y_{Tn}

Los valores de X se suponen ordenados tanto en el grupo control como en el grupo tratamiento:

$$x_{C1} \leq \dots \leq x_{Ci} \leq \dots x_{Cn}$$

$$y_{T1} \leq \dots \leq y_{Tj} \leq \dots y_{Tn}$$

Si el diseño que el investigador está estudiando es el típico de datos apareados, antes y después de un tratamiento (pre/post), la estimación del tamaño del efecto será la transformación probit del porcentaje de individuos antes de la intervención que presentan valores mayores o menores, dependiendo del objetivo del investigador, que la mediana de los valores de los individuos después de la intervención:

$$\hat{p} = \frac{n_{x_{ij} \geq \text{Med}(Y_T)}}{n_c} \xrightarrow{\Phi^{-1}(\hat{p})} z$$

Siendo $\Phi^{-1}(p)$ la inversa de la función de distribución de la Normal.

Generalmente, en diseños con tamaños muestrales pequeños, puede ocurrir que obtengamos valores de p que sean 0 ó 1 con el consiguiente problema a la hora de obtener los valores z.

Para evitar estos tamaños del efecto extremos KRAEMER y ANDREWS (1982) proponen considerar $\frac{n}{(n+1)}$ en lugar de 1 ó $\frac{1}{(n+1)}$ en lugar de 0.

Otras transformaciones que amplían o disminuyen el rango de estos valores extremos pueden ser utilizadas siempre que se tenga en cuenta la variable que estamos estudiando.

En el diseño Pre-post/tratamiento-control, en el cual se quiere eliminar el efecto placebo, el procedimiento es similar al caso anterior. Se calculan los valores **z** tanto para el grupo tratamiento como para el grupo control y su diferencia nos

proporciona un estimador del efecto real del tratamiento de la forma:

$$\mathbf{z}_T - \mathbf{z}_C = \mathbf{z}_R$$

Por último, en el diseño tratamiento-control el tamaño del efecto se obtiene de la transformación de la proporción de valores del grupo control que son mayores o menores que la mediana del grupo tratamiento.

Otro factor que puede afectar a la exactitud de los resultados es la presencia de “outliers” en los valores originales de las investigaciones; para esta situación HEDGES y OLKIN, (1984) proponen un índice que implica estimar las medias de las poblaciones experimental y control a través de las medianas de los grupos respectivos, obteniendo el estimador del efecto como la razón de la diferencia de las medianas y el rango o una combinación lineal de estadísticos ordenados.

$$\delta' = \frac{\text{Med}_T - \text{Med}_C}{\sigma'}$$

siendo $\sigma' = a_2 y_{C_2} + \dots + a_{n-1} y_{C_{n-1}}$

donde $y_{C_1} \leq y_{C_2} \leq \dots \leq y_{C_{n-1}}$ son los valores ordenados del grupo Control.

Los coeficientes óptimos a_2, \dots, a_{n-1} se eligen de modo que se minimice la varianza de σ' .

El estadístico δ' es un estimador de δ que no está afectado por las observaciones más extremas.

El mayor inconveniente de estos métodos no paramétricos a la hora de estimar el tamaño del efecto para su posterior análisis e integración en un meta-análisis es la disponibilidad de los datos. El problema es similar al presentado por los análisis de cluster y por lo tanto con las mismas restricciones que este tipo de revisión.

CAPITULO 6

ANALISIS DE LOS RESULTADOS

6.1. INTRODUCCION

Establecida la unidad de medida de la revisión, uno de los objetivos fundamentales que persigue el revisor es sintetizar los resultados para obtener un índice global y representativo del conjunto de estudios, suponiendo que los datos integrados son lo suficientemente homogéneos para que esta media global sea representativa del total de estudios; pretende además conjeturar nuevas relaciones con variables externas que puedan influir en la distribución y agrupación de los resultados obtenidos a través de los estudios.

De manera similar a como ocurre en las investigaciones primarias, el meta-análisis exige al revisor que adopte las mismas normas de rigor científico en el análisis e interpretación de un conjunto de resultados de estudios diferentes (COOPER, 1982, 1984; HEDGES, 1984a).

Las técnicas estadísticas que se utilizan en el análisis de los datos son de diferente naturaleza dependiendo del objetivo que se pretenda alcanzar, al igual que en los estudios primarios; si bien, en el Meta-análisis el enfoque exploratorio defendido por TUKEY (1977) es básico para la descripción de resultados utilizando estadísticos robustos y métodos gráficos.

GLASS y cols. (1981) desarrolla todos los trabajos desde este punto de vista exploratorio utilizando técnicas descriptivas que resumieran la información obtenida y descubrieran relaciones que pudieran afectar a los resultados, advirtiendo del peligro de realizar inferencia, dado el tipo de datos con el que se enfrenta el revisor meta-analítico.

Es común encontrar estudios meta-analíticos que aplican las pruebas estadísticas convencionales con naturaleza inferencial (ANOVAs, Análisis de

Regresión, etc.) como si se tratara de investigaciones primarias. Sin embargo, son probables serias violaciones de las hipótesis de base: los estudios meta-analíticos no suelen ser independientes pues un mismo estudio puede proporcionar varias estimaciones del Tamaño del Efecto (TE); el supuesto de homoscedasticidad raramente se cumple, ya que las varianzas de los TEs son inversamente proporcionales a sus tamaños muestrales (COOPER, 1984; HEDGES y OLKIN, 1985); por estas razones, deben existir serias dudas en relación a la validez de un meta-análisis basado en procedimientos estadísticos inferenciales convencionales.

6.2. METODOS GRAFICOS EN ESTUDIOS META-ANALITICOS.

Existe un fuerte paralelismo entre los análisis de datos de los estudios meta-analíticos y de los ensayos primarios de investigación, por lo que el enfoque exploratorio propuesto por TUKEY (1977), a partir de los métodos gráficos como herramienta fundamental en el análisis de los datos, es valido igualmente para las revisiones cuantitativas.

La aplicación de estos métodos puede facilitar la presentación de los resultados, la exploración de la posibles causas externas que pueden afectar a la heterogeneidad de los mismos, el descubrimiento de posibles tendencias en los tamaños del efecto, y la determinación de posibles sesgos que pueden afectar a las conclusiones finales (LIGHT y PILLEMER, 1984; JENICEK, 1987; GLASS y cols., 1981; ROSENTHAL, 1984).

Los gráficos habituales que aparecen en el meta-análisis varían desde métodos sencillos y clásicos hasta aquellos más específicos debido a las particularidades de los estudios meta-analíticos.

6.2.1 GRAFICOS DE DISTRIBUCION DE FRECUENCIAS.

Este tipo de representaciones nos proporcionan información de como se comportan los tamaños del efecto; cuál es la forma de la distribución (simétrica o asimétrica) y si existen uno o mas tamaños del efecto dependiendo de las posibles acumulaciones de frecuencias en un punto o en varios. Entre los gráficos mas utilizados se encuentran los **histogramas** (GUREVITCH y cols., 1992), **stem-leaf** (ver WILLIAMS y cols., 1982; STOCK y cols., 1983) y **box-plot**, o **notchet-plot**, (ver GLASS y cols., 1981). Este último tipo de gráfico permite realizar un estudio comparativo del comportamiento de los tamaños del efecto en los diferentes grupos generados por una variable externa que influye en el valor que toman los efectos.

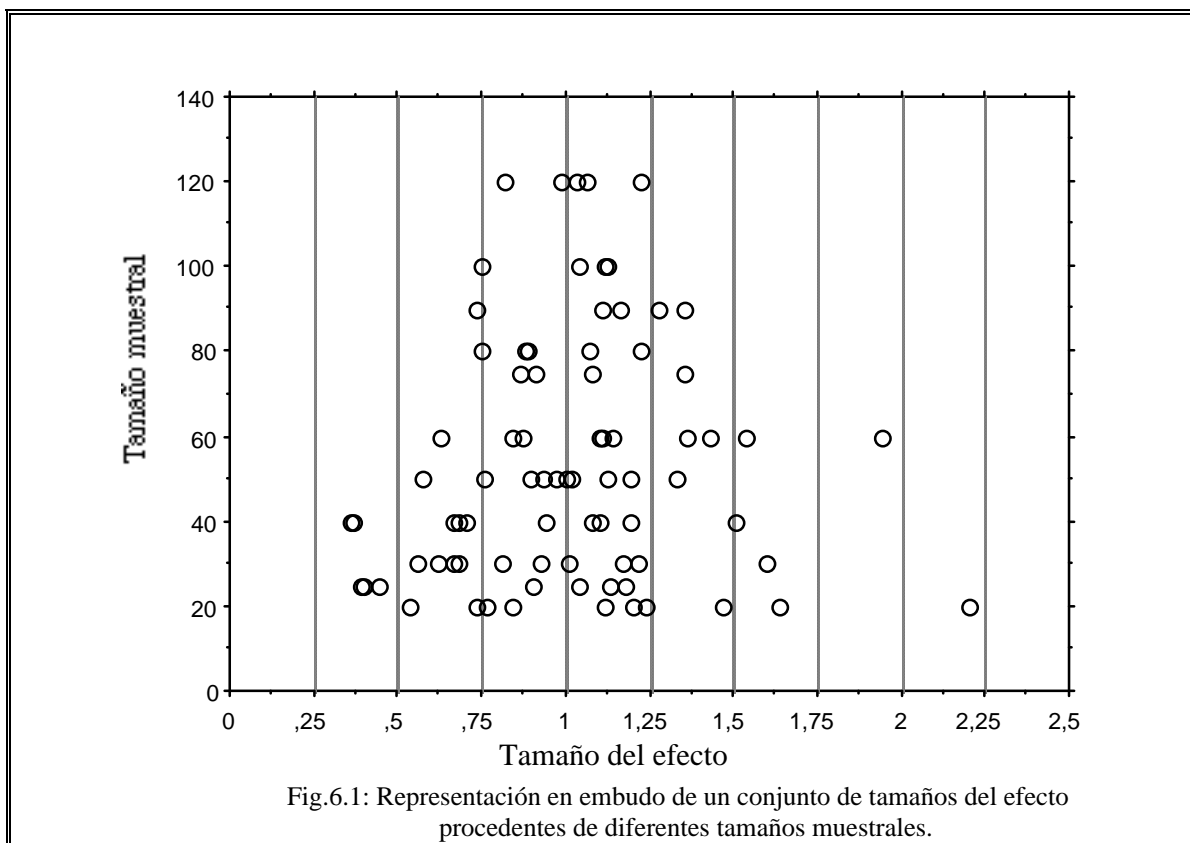
6.2.2.DIAGRAMA DE DISPERSION.

Este gráfico generalmente es utilizado para detectar posibles tendencias de los tamaños del efecto debidas a factores externos como puede ser el año de publicación, la dosis del tratamiento, edad de los pacientes, etc. (ver SMITH y cols., 1980; MESSICK y JUNGEBLUT, 1981).

Un método peculiar derivado del diagrama de dispersión es el descrito por LIGHT y PILLEMER (1984), el cual se denomina "**funnel display**" y cuyo termino en castellano puede acuñarse como "diagrama en embudo".

Esta representación se basa en la propiedad de consistencia de varios estimadores donde el error muestral disminuye a medida que aumentamos el tamaño muestral y por consiguiente se obtienen estimadores más próximos al parámetro poblacional.

En el "funnel display" se representa el tamaño del efecto (eje de abcisas) frente al tamaño muestral (eje de ordenadas) del que procede. La forma de la nube de puntos que se espera bajo la suposición de consistencia, aparece extendida sobre el eje de abcisas para valores pequeños de tamaño muestral y convergerá hacia el valor real del efecto conforme se incrementa el número de observaciones (ver figura 6.1).



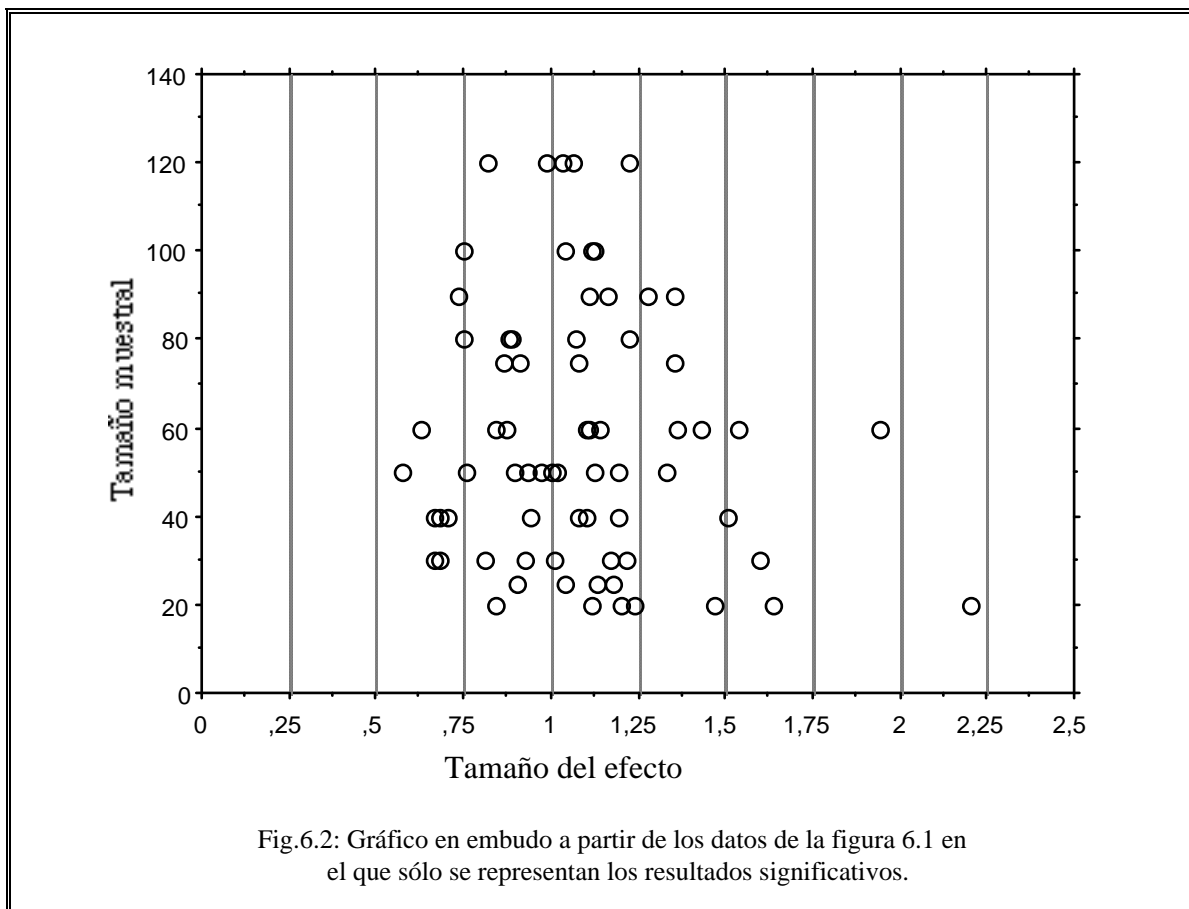
La principal utilidad de este tipo de gráfico es el estudio y detección del sesgo; ante la presencia de este problema el revisor encontrará formas degenerativas de la nube de puntos.

El fundamental sesgo que afecta a las revisiones cuantitativas es el causado por la dificultad de la exploración de estudios cuyos resultados son no significativos. Así, en presencia del sesgo de publicación el revisor únicamente dispone de trabajos significativos que proporcionan representaciones del diagrama que se alejan de la forma habitual.

Si el tamaño del efecto es próximo a cero, entonces aparecerá la zona central del cono despejada de puntos; si, por el contrario, el tamaño del efecto está más alejado del efecto nulo, el extremo de la nube adyacente al valor cero, estará truncado produciendo una forma asimétrica.

Hemos realizado un estudio de simulación en el que se obtuvieron varias muestras de diferente número de observaciones de dos poblaciones Normales (control y tratamiento), cuyo tamaño del efecto poblacional era igual a uno. En la figura 6.1 aparece el funnel display completo para todos los puntos.

Sin embargo, en la figura 6.2 sólo se representaron aquellos resultados que fueron significativos. Se puede observar como la forma del cono es anómala debido a la ausencia de los puntos en el lado izquierdo lo que produce una forma más triangular que cónica, lo cual haría pensar al revisor la presencia del sesgo de publicación.

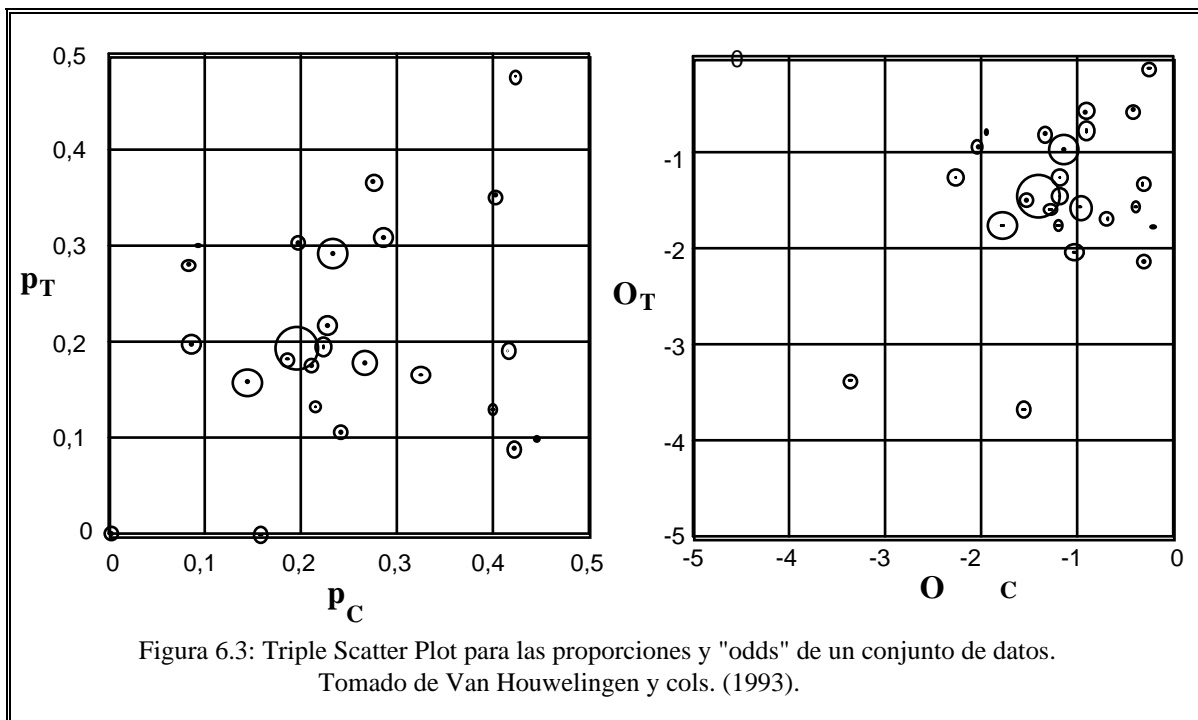


Otro gráfico basado en el diagrama de dispersión al cual se incorpora una nueva variable es el "**triple scatter plot**" (ASCOMBE, 1973) donde los puntos son representados mediante un círculo cuyo diámetro depende de esa nueva variable. Este gráfico se utiliza generalmente en meta-análisis para la presentación de proporciones u odds en estudios epidemiológicos, si bien, puede ser extensible a la diferencia media tipificada.

En el eje de abcisas se representa la proporción o el odds correspondiente al grupo control y en el eje de ordenadas los valores correspondientes al grupo experimental, definiendo el tamaño del círculo las ponderaciones correspondientes a cada par de valores de interés procedentes de las diferentes publicaciones.

A través este tipo de gráfico se puede observar la presencia de posibles "outliers" o agrupaciones de estudios, así como, el comportamiento diferente de los estudios dependiendo de las ponderaciones.

En la figura 6.3 podemos observar un ejemplo de este tipo de representación obtenido del estudio de VAN HOUWELINGEN y cols. (1993) donde dibuja los valores de proporciones y logaritmos neperianos de los "odds" para un mismo conjunto de valores. Se puede advertir la presencia de dos tamaños del efecto cuyo comportamiento difiere del resto, si bien es menos claro en el caso de la utilización de proporciones debido a la diferencia de escala utilizada en un caso y en el otro.



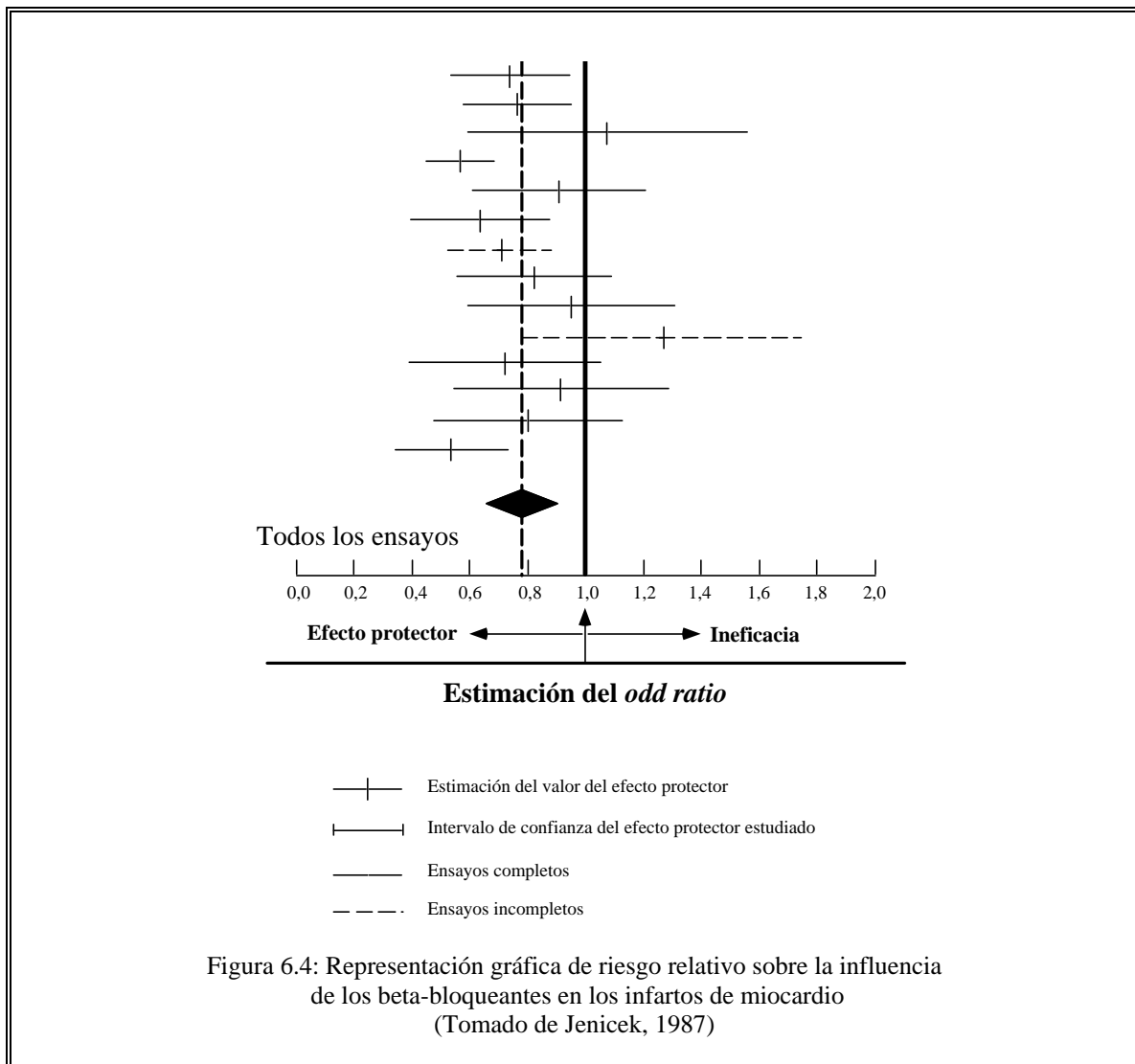
6.2.3.GRAFICO DE ODDS RATIO (O RIESGO RELATIVO).

Este gráfico (figura 6.4), si bien su aplicación es casi exclusiva en la presentación de los diferentes resultados de ensayos clínicos en términos de odds-ratio o riesgo relativo, puede aplicarse para la representación del resto de estimadores del tamaño del efecto como son el coeficiente de correlación o la diferencia de medias estandarizada. Es un método muy extendido y un ejemplo puede encontrarse en COLLINS y cols., 1990; SAWKA y PRITCHARD, 1992 o en GELBER y GOLDHIRSH, 1991 o en PIEDBOIS y cols. 1992 por citar algunos.

El gráfico consiste en la representación de los intervalos de confianza, tanto para cada uno de los resultados individuales como para el estimador global.

Cada intervalo individual se representa mediante una línea cuyo configuración puede variar para expresar información adicional sobre algunas de las características importantes del ensayo. Dentro del intervalo se suele marcar el estimador.

El intervalo del resultado global suele representarse mediante un rombo cuyos vértices verticales representan al estimador y del cual parte una línea discontinua a lo largo de todo el gráfico. Paralela a esta línea discontinua se representa otra línea continua que hace referencia a la ausencia de efecto del tratamiento



Esta representación permite al lector obtener una impresión visual de los resultados el cual puede ser más útil que la simple inspección de los resultados estadísticos. Sin embargo GALBRAITH (1988a) apunta el inconveniente de la escala lineal en la representación de odds-ratio. En este caso la longitud del intervalo es proporcional a la estimación del odds ratio y por lo tanto puede parecer engañoso al considerar los intervalos con mayor amplitud menos informativos.

$$I = \frac{ad}{bc} e^{\pm 2\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}}$$

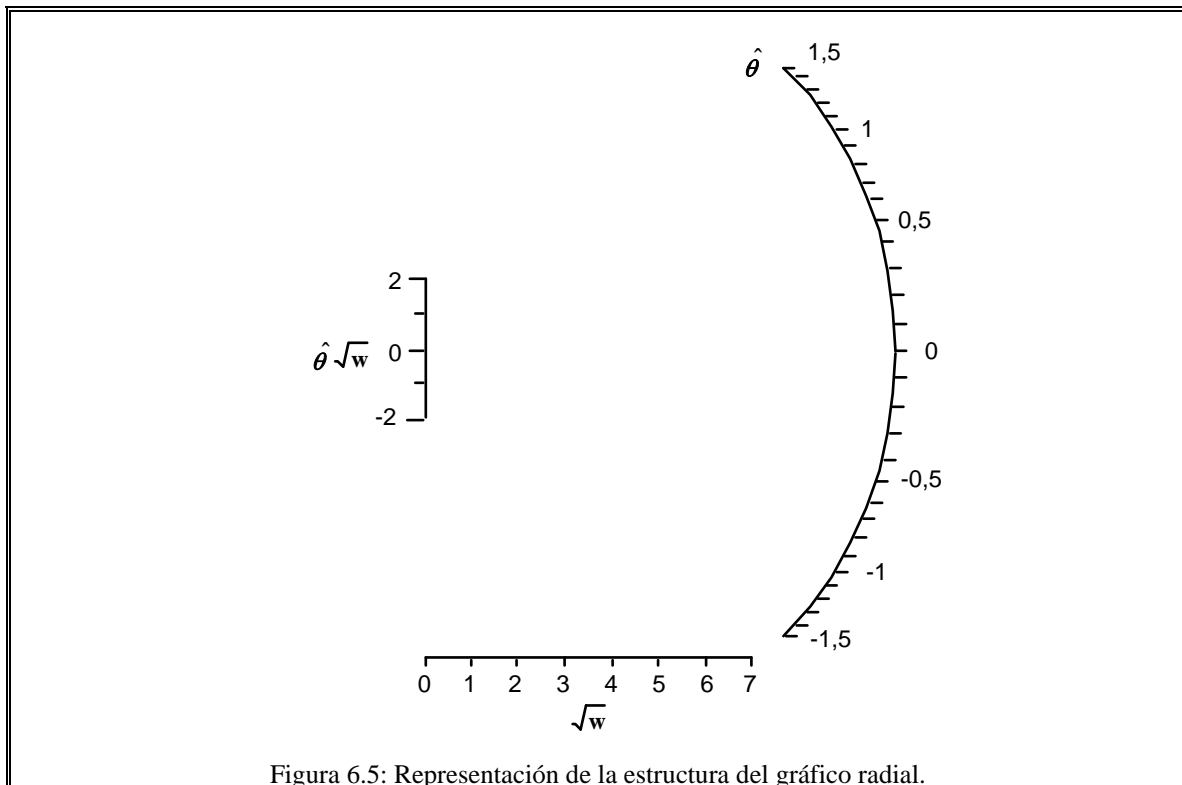
BAILEY (1987) intenta evitar el problema representando directamente los logaritmos de los odds-ratio, pese a que esta solución presenta el inconveniente de la interpretación real de dichas estimaciones. GALBRAITH (1988 a), propone una solución intermedia en la cual se representa la longitud de los intervalos obtenida mediante el uso de los logaritmos pero se refieren a la escala lineal en el gráfico.

6.2.4. GRAFICO RADIAL.

El gráfico radial (GALBRAITH, 1988a) nos permite representar los diferentes tamaños del efecto de cada estudio, así como, el estimador global representante del conjunto de todos ellos. Es útil a la hora de explorar la heterogeneidad y ver que estudios son los responsables de la misma. Además, evita el problema de escala ocasionado en la representación habitual de los odds-ratio (GALBRAITH, 1988b).

Esta representación es un diagrama de dispersión donde se introduce un nuevo eje radial (figura 6.5.). El eje de abcisas representa los valores de la precisión de cada estudio, y por lo tanto, esta relacionado con las ponderaciones utilizadas para calcular el estimador global; el eje de ordenadas representa las estimaciones individuales estandarizadas. En el eje radial se representan los valores de las estimaciones del tamaño del efecto $\hat{\theta}_i$, mediante el punto de corte

con el eje radial de la recta que pasa por el origen y el punto de coordenadas x_i, y_i .



La estimación mínimo cuadrática de la pendiente, correspondiente a la recta de regresión que pasa por el origen, es igual al estimador global del tamaño del efecto :

$$\hat{\theta} = \frac{\sum_i \hat{\theta}_i w_i}{\sum_i w_i} \quad 1$$

$$b = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{\sum_i \theta_i \sqrt{w_i} \sqrt{w_i}}{\sum_i (\sqrt{w_i})^2} = \frac{\sum_i \theta_i w_i}{\sum_i w_i}$$

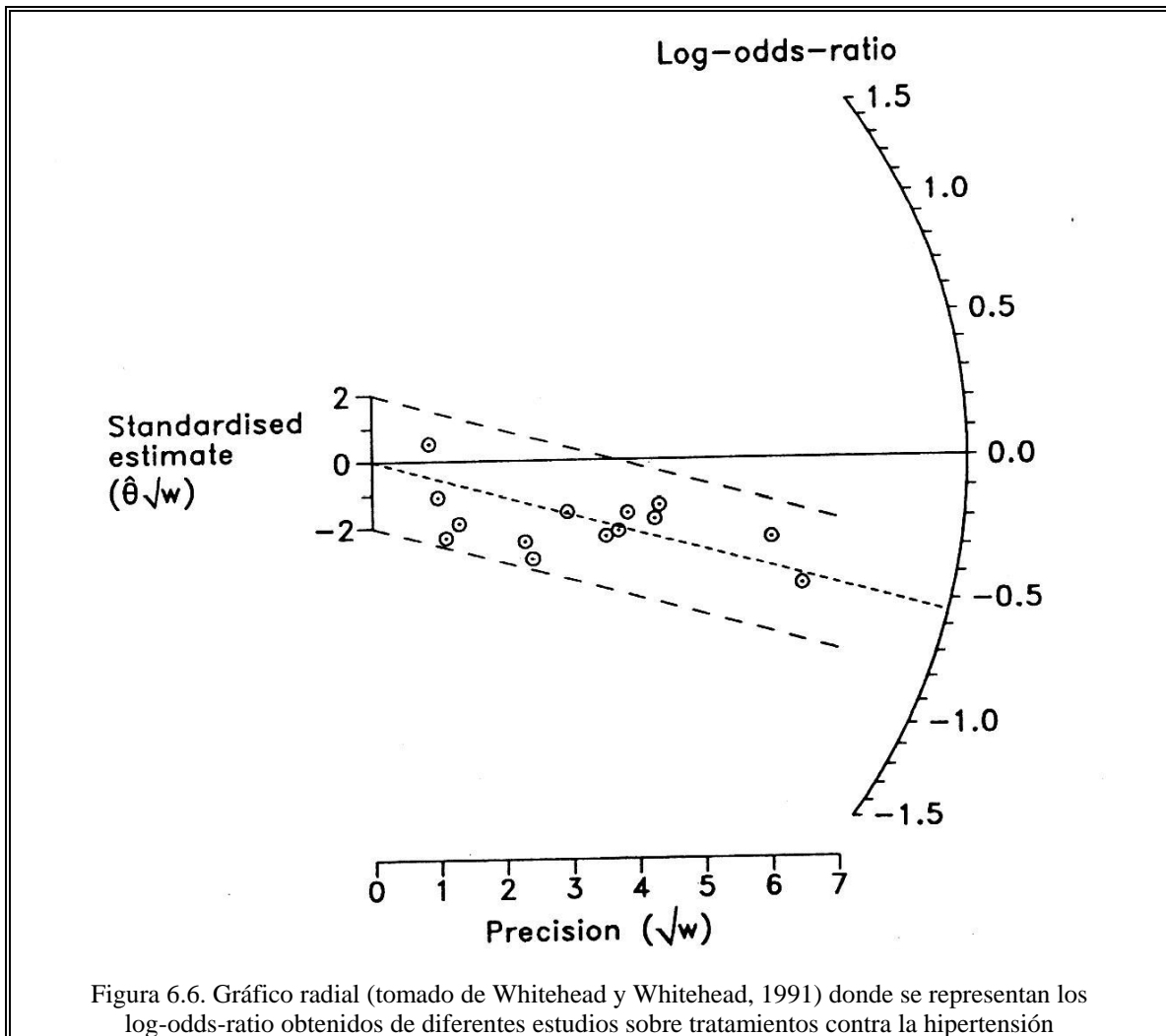
Los valores de la variable correspondientes al eje de ordenadas poseen varianza² unidad lo que permite definir un intervalo de confianza al 95% que define una banda donde podemos situar las diferentes observaciones individuales.

$$y = \hat{\theta}x \pm 2$$

Este intervalo procede de la simple idea de considerar el intervalo correspondiente a cada observación particular $\hat{\theta}_i$ y realizar la oportuna transformación (GALBRIATH, 1988b). En la figura 6.6 observamos la representación de un conjunto de log-odds-ratio donde el valor del estimador global viene dado por la intersección de la línea de puntos con el eje radial. Se observa como la totalidad de las observaciones son consistentes al encontrarse dentro de la banda de confianza.

En resumen, el gráfico nos permite observar los valores que toman los diferentes estimadores individuales y el global, además de conocer la precisión de los mismos con sólo fijarse en la situación del punto en la nube; por último, podemos conocer aquellos tamaños del efecto posibles causantes de la heterogeneidad, si se encuentran fuera de la banda de confianza señalada en el gráfico.

$$\text{var}(y_i) = \text{var}(\hat{\theta}_i w_i) = \frac{\text{var}(\hat{\theta}_i)}{\sigma_{\hat{\theta}_i}^2} = 1$$



6.3. METODOS MULTIVARIANTES PARA LA EXPLORACION DE LA VARIABILIDAD DE LOS ESTUDIOS

Los métodos multivariantes sirven para analizar, describir e interpretar un conjunto de observaciones multidimensionales, en el sentido de que numerosas observaciones son obtenidas para cada individuo o unidad de muestreo. Así, en este tipo de técnicas nos enfrentamos ante matrices de datos procedentes de la ordenación de las medidas para n individuos u objetos sobre p variables.

En meta-análisis, el entorno de las características (metodológicas y sustantivas) que rodean a los resultados, definen un espacio p -dimensional que determina y caracteriza a cada una de los estudios revisados. Este contorno es importante para el ulterior análisis de los tamaños de efecto. Por ello, como se ha comentado, la etapa de codificación cobra un especial interés a la hora de describir y cuantificar ese entorno. El conjunto de características pueden reconocerse como posibles variables explicativas de la heterogeneidad de los resultados. En relación con este proceso será de gran interés la introducción de métodos multivariantes que nos ayuden a visualizar las características de los diferentes estudios para una conexión con el proceso de análisis de resultados.

En una intensa búsqueda de bibliografía no se ha encontrado ninguna referencia al uso de métodos multivariantes en el análisis de los datos meta-analíticos, todas las referencias se reducen a la utilización de modelos de regresión múltiple ponderados, donde las variables independientes son las características

de los estudios y la dependiente los tamaños del efecto (HEDGES y OLKIN, 1985; GREENLAND, 1987).

En las técnicas multivariantes se pueden distinguir claramente dos grandes grupos de técnicas: los Métodos Multivariantes Gráficos y los Métodos Gráficos Multivariantes. Los primeros son métodos de diagnóstico que sintetizan la información de una matriz de datos mediante procesos algebraicos complejos y representan la información del hiperespacio de partida en un subespacio de dimensiones reducidas, en forma gráfica. Los métodos gráficos multivariantes efectúan simplemente una transcripción geométrica de los datos correspondientes a un conjunto de variables.

6.3.1.MÉTODOS MULTIVARIANTES GRAFICOS

Cuando se dispone de un conjunto de observaciones sobre las que se mide un número elevado de variables, es difícil la interpretación simultánea de todas ellas.

Como se ha comentado, en la etapa de codificación se obtienen un conjunto de variables que pueden identificar los distintos trabajos y ayudan a la explicación de las posibles diferencias existentes en los estudios. Por lo tanto, es particularmente interesante en este contexto, el estudio de las características de las distintas fuentes como causa de la posible heterogeneidad de los tamaños del efecto.

Para el estudio simultáneo de las variables resultantes de la codificación objetiva son útiles las técnicas multivariantes denominadas de ordenación, en las que se obtiene un conjunto de coordenadas en un espacio euclídeo que representan a los estudios y a las variables.

En todas ellas, el primer eje de la representación "contiene" mayor cantidad de información mientras que el resto de los ejes van siendo progresivamente de menor importancia. La "cantidad" de información suele medirse en términos de variabilidad absorbida y suele calcularse como el cociente entre la suma de cuadrados de las coordenadas sobre cada uno de los ejes y la suma de cuadrados de las coordenadas sobre todos los ejes.

Dos son los tipos de métodos de ordenación de acuerdo con la forma de la matriz de datos originales:

- Métodos basados en la rotación del espacio inicial.
- Métodos basados en la obtención de una imagen euclídea.

6.3.1.i.MÉTODOS BASADOS EN LA ROTACIÓN DEL ESPACIO INICIAL.

Son aquellos en los que se parte de una matriz \mathbf{X} de n individuos por p variables y cada uno de los individuos (variables) se considera como un punto en un espacio p -dimensional (n -dimensional) cuyas coordenadas son los valores que toma cada individuo en la variable correspondiente.

La técnica multivariante de ordenación consiste en rotar el conjunto original de ejes hasta que el primero tenga la propiedad de absorber la

mayor parte de la variabilidad, el segundo eje se toma perpendicular al primero y de forma que absorba la mayor parte de la variabilidad restante y así sucesivamente. Será posible obtener r ejes, donde r es el rango de la matriz de datos de partida. Obtenemos así una nueva matriz \mathbf{G} de n filas por r columnas que contiene las coordenadas de los n individuos en los r nuevos ejes. Se trata básicamente de una rotación del sistema de referencia inicial y su cálculo suele estar basado en la descomposición en valores singulares de la matriz de datos o en la descomposición espectral de matrices de covarianzas, correlaciones o productos cruzados.

De los r ejes posibles se seleccionan los dos o tres primeros sabiendo que la pérdida de información es mínima y que los dos puntos similares, y por tanto próximos en el espacio inicial estarán también próximos en el espacio de dimensión reducida.

Proyectando la base canónica del espacio original sobre el espacio de dimensión reducida se obtienen un conjunto de coordenadas para las variables \mathbf{H} que pueden interpretarse como un sistema de referencia oblicuo y que pueden interpretarse, junto con las coordenadas de los individuos, en términos de producto escalar tal y como se describía en la teoría de los métodos Biplot del capítulo de codificación.

Dependiendo del tipo de datos de los que se disponga originalmente (continuos, tablas de contingencia, varas variables discretas, estructura de grupos, etc.), de las transformaciones realizadas sobre los datos originales (centrados, estandarizados, conversión a perfiles, cálculo de los residuales estudentizados tras el ajuste de modelos aditivos, etc.), y

de las masas asignadas a cada uno de los puntos, se obtienen diferentes métodos de análisis.

Dentro de este apartado, se incluyen métodos como el Análisis de Componentes Principales (sobre datos estandarizados, centrados o brutos), el Análisis de Correspondencias Simples, el Análisis de Correspondencias Múltiples, Biplots, Análisis Canónico de Poblaciones, Análisis Canónico de Correspondencias, etc.

Si la proyección de las variables no es ortogonal sino oblícua, teniendo en cuenta la variabilidad de los ejes y la relación de las variables con los mismos, se obtienen técnicas como el Análisis de Correspondencias simétrico o el HJ-Biplot (GALINDO, 1986)

Revisiones de la mayor parte de estos métodos pueden encontrarse en libros generales sobre Análisis Multivariante como: CUADRAS (1991), SEBER (1984), MARDIA y cols. (1979), etc.

6.3.1.ii.MÉTODOS BASADOS EN LA OBTENCIÓN DE UNA IMAGEN EUCLÍDEA

Son aquellos en los que se parte de una matriz **D** de disimilaridades (o similaridades) observadas entre n individuos o variables.

El propósito es ahora obtener un conjunto de n puntos en un espacio de r dimensiones (en una matriz de coordenadas **G** o **H**) de forma que las interdistancias entre los n puntos aproximen las distancias observadas o

una función de las mismas. En este caso, en general, ya no es posible el cálculo simultáneo de coordenadas para las filas y las columnas.

En este apartado se encuentran todas las técnicas del Análisis de Proximidades (Multidimensional Scaling, MDS).

6.3.1.iii.CALIDAD DE LA REPRESENTACION EN DIMENSION REDUCIDA.

Los métodos factoriales gráficos presentan los resultados en forma de diagramas de dispersión, generalmente en un subespacio de dimensión 2, aunque la configuración original sea de dimensión mayor. Al proyectar se produce una pérdida de información que puede distorsionar las configuraciones iniciales.

El primer problema a tener en cuenta es el número de dimensiones necesarias para obtener una representación adecuada en dimensión reducida.

En la literatura tenemos varios procedimientos que nos permiten la búsqueda del número de dimensiones necesarias para describir de forma óptima la nube de puntos. Los métodos están descritos inicialmente para los Análisis de Componentes principales o de Correspondencias, pero pueden extenderse a los Métodos Biplot o cualquier otra técnica de las pertenecientes al grupo basado en la rotación del espacio original:

» BARTLETT (1950) propone un test de igualdad de los últimos (p - k) valores propios (α^2). El test ha tenido diversas modificaciones; especificamos aquí la descrita por ANDERSON (1963)

$$\chi^2 = -(\nu) \sum_{j=k+1}^p \{\ln(\alpha_j^2)\} + (\nu)(p - k) \ln \left\{ \sum_{j=k+1}^p \frac{\alpha_j^2}{(p - k)} \right\}$$

donde χ^2 sigue una distribución Ji-cuadrado con $(1/2)(p-k-1)(p-k+2)$ grados de libertad y ν representa a los grados de libertad asociados con la matriz de covarianzas.

El test puede construirse de forma secuencial, introduciendo nuevas componentes hasta que se acepte la hipótesis de igualdad. Para ver más detalles sobre el test y las condiciones de aplicación ver, por ejemplo, CUADRAS (1991).

» Muchos autores, por ejemplo MARDIA y cols. (1979), utilizan como regla de selección la inclusión de factores hasta conseguir explicar un determinado porcentaje de la varianza total (normalmente el 90 ó el 95%). No existe ninguna regla formal que permita una selección adecuada del porcentaje óptimo, por lo que depende del problema concreto a estudiar.

» CATELL (1966) propone hacer un diagrama de los valores propios en orden decreciente. Si marcamos el punto de inflexión de la curva de los valores propios, tomaremos aquellos ejes cuyos valores propios preceden dicho punto de inflexión.

La técnica es usada con mucha frecuencia si bien tampoco tiene una regla para decidir cual es el punto de corte óptimo. No es fácil obtener el punto, si la curva muestra un descenso suave.

Si la diferencia entre el primer y el segundo valor propio es muy elevada, los efectos de la escala del gráfico pueden enmascarar la diferencia real entre el resto de las raíces. Para evitar este problema, usual en la práctica, puede tomarse el logaritmo del valor propio como cantidad a representar.

» Otra técnica común usada en la práctica e implementada en la mayor parte de los programas de ordenador estándar, es seleccionar aquellas raíces que superan el valor medio $\sum_{i=1}^p \frac{\alpha_i^2}{p}$. Si se calculan los ejes con

datos centrados y reducidos la media será 1 y el criterio será tomar aquellas componentes cuyas raíces sean mayores que 1. JOLLIFFE (1986), asegura que este punto de corte es demasiado alto y no tiene en cuenta la variabilidad muestral; basándose en estudios de simulación propone utilizar el 70% de ese valor medio.

Una vez seleccionado el número de ejes de la representación interpretaremos las posiciones de los puntos con respecto a dichos ejes.

Usualmente, los autores que trabajan con los métodos de este tipo reducen sus interpretaciones a gráficos bidimensionales aunque, en la práctica, dos dimensiones pueden no ser suficientes para describir correctamente el conjunto de datos.

GABRIEL (1981) considera la posibilidad de realizar representaciones tridimensionales que denomina BIMODELS; tales representaciones en

tres dimensiones no pueden dibujarse directamente sobre un papel plano, y es necesaria la utilización de medios técnicos más sofisticados.

En general, a pesar del esfuerzo realizado, las representaciones obtenidas son difíciles de interpretar y, en muchos casos, inútiles. Si el número de ejes seleccionados es mayor de tres, la situación es aún más complicada.

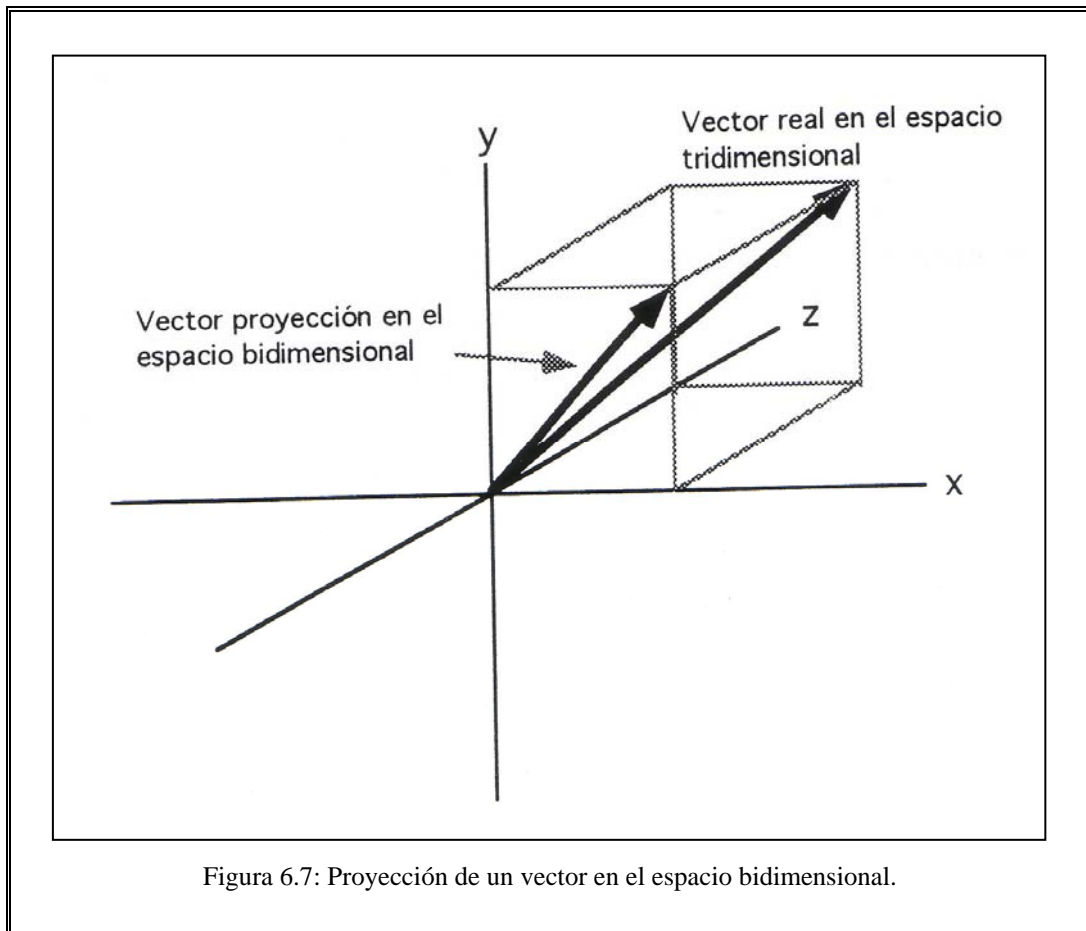
La forma más sencilla de interpretar la posición de los puntos en más de dos dimensiones, consiste en proyectar en planos sucesivos, es decir, representar primero el eje 1 con el 2, luego el 1 con el 3, etc.

El problema ahora es que todos los planos distintos del 1-2 ya no tienen las propiedades óptimas de las proyecciones en los planos de mejor ajuste. Es necesario disponer de alguna ayuda adicional que permita saber cual es la información que contiene cada uno de los planos.

La calidad de la representación de la proyección de la nube de puntos en los planos utilizados para la interpretación puede medirse mediante el cociente entre la suma de los dos valores propios correspondientes y la suma de todos ellos, tal y como se hacía en el mejor subespacio de dimensión 2.

Sin embargo, esta medida es una aproximación global para todos los puntos, lo cual no implica que sea una medida de la calidad de la representación de cada punto en particular.

La figura 6.7 muestra la proyección de uno de los puntos de la nube en un espacio bidimensional.



Supongamos que se trata de la proyección de uno de los vectores que representa a una variable. El ángulo que forman el vector “variable” y el eje horizontal, está relacionado con la correlación entre ellos.

La figura 6.8 muestra la situación esquematizada. El coseno al cuadrado del ángulo se puede tomar como medida de la relación entre la variable y el eje.

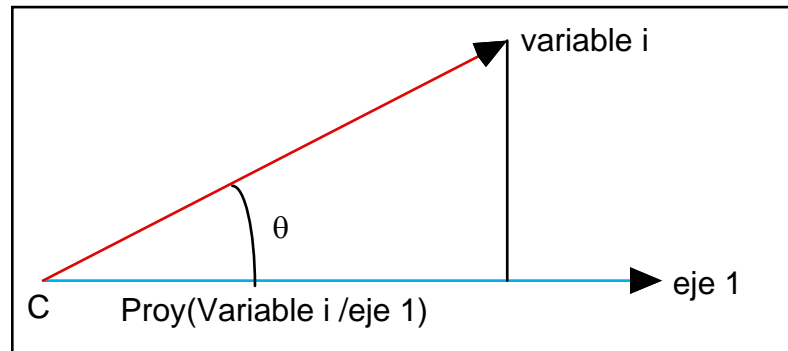


Figura 6.8: Relación geométrica entre el vector variable y el factor.

$$\cos^2(\theta) = \frac{\| \text{proy}(\text{variable } i, \text{ eje } 1) \|^2}{\| \text{variable } i \|^2}$$

A esta medida la denominaremos CALIDAD DE LA REPRESENTACION del punto i sobre el eje factorial, siguiendo una terminología similar a la que utiliza BENZECRI (1973) para conceptos similares. (CLR_{i1}).

Esta cantidad puede calcularse también a partir del producto escalar entre el vector i y un vector cualquiera en la dirección del eje.

Denotamos con $\mathbf{h}_{(i)} = (h_{i1}, \dots, h_{ir})'$ al vector correspondiente al punto i en el espacio multidimensional. Por conveniencia, tomamos como vector en la dirección del eje, al vector que tiene una única coordenada distinta de 0, $\mathbf{e}_1 = (0, \dots, \alpha_1, \dots, 0)$, se tiene que

$$\text{CLR}_{i1} = \cos^2(\theta_1) = \frac{(\mathbf{h}_{(i)}' \mathbf{e}_1)^2}{(\mathbf{h}_{(i)}' \mathbf{h}_{(i)}) (\mathbf{e}_1' \mathbf{e}_1)} = \frac{(h_{i1}^2 \alpha_1^2)}{\left(\sum_k h_{ik}^2 \right) (\alpha_1^2)} = \frac{h_{i1}^2}{\sum_k h_{ik}^2}$$

La calidad de la representación es una medida relativa, ya que la suma de las calidades de la representación de cada elemento sobre todos los ejes factoriales es 1.

La calidad de la representación con respecto a un plano se mide de la misma manera, es decir, como el coseno al cuadrado del ángulo que forman el vector y el plano. Este coseno al cuadrado es la suma de los cosenos al cuadrado de los ángulos con los ejes que forman el plano.

$$\cos^2 (\theta_{lK}) = \cos^2 (\theta_l) + \cos^2 (\theta_K)$$

Por tanto, la calidad de la representación del elemento es una medida aditiva que puede calcularse para la proyección en cualquier plano factorial, sin más que sumar las calidades de representación con respecto a los ejes factoriales que lo forman.

La medida de la bondad de ajuste de cada punto particular tiene aun otra interpretación: La calidad de la representación es la parte de la dispersión del elemento explicada por cada uno de los ejes, por lo que se denomina también “contribución relativa del factor l al elemento i” (CRFE_{il}).

Basta tener en cuenta que CLR_{il} es la coordenada correspondiente al cuadrado dividida por la longitud al cuadrado del vector en el espacio multidimensional y que, la dispersión generalizada de cada uno de los elementos está directamente relacionada con la longitud del vector.

Las definiciones de calidad de representación y de contribución relativa del factor al elemento son idénticas para las filas y las columnas de la matriz de datos.

Teniendo en cuenta que la posición de los puntos puede ser sólo aparente en las distintas proyecciones, solamente se interpretarán correctamente,

en cada una de ellas, aquellos puntos cuya calidad de representación sea buena.

Así, la forma usual en que deben presentarse los análisis de tipo factorial ha de incluir una tabla en la que se muestren las contribuciones o calidades de representación, de cada uno de los ejes factoriales a los distintos elementos (fila y columna).

No existe ningún criterio objetivo que nos permita decidir cual es el valor de la calidad que debe tomarse como aceptable, por lo que este tipo de decisión queda a juicio del investigador. Obviamente la interpretación de los valores aceptables no puede ser la misma en los ejes de mejor ajuste que en ejes residuales.

Otra interpretación importante de la contribución del factor al elemento, como variabilidad de la variable absorbida por el factor, es que los elementos con contribuciones altas de un factor son elementos exclusivos de dicho factor y no de los demás, lo cual permite interpretar claramente las posiciones de los puntos en términos de las variables originales correspondientes.

6.3.1.iv. APLICACION DE UNO DE LOS MÉTODOS MULTIVARIANTES A UN CASO REAL.

Como una ilustración de la aplicación de los métodos multivariantes gráficos consideramos los resultados obtenidos de un estudio meta-analítico realizado por CANNER (1987) y publicado en la revista "*Statistics in Medicine*".

Información recogida en la Publicación.

Este estudio revisa seis ensayos clínicos (que denominaremos simplemente con los números del uno al seis) sobre la influencia de la aspirina como tratamiento en las enfermedades coronarias. La respuesta se mide a partir de la mortalidad debida al total de las posibles causas y el tamaño del efecto elegido es el logaritmo del odds-ratio.

A partir de los estudios distingue un conjunto de características referentes al diseño de los ensayos como posibles variables explicativas de la heterogeneidad identificada. Las características elegidas son:

- * **Números de pacientes.**
- * **Media de edad de los pacientes participantes en el estudio.**
- * **Sexo de los participantes** (varones, mujeres o ambos).
- * **Dosis diaria (mg).**
- * **Nº de tomas diaria.**
- * **Tiempo transcurrido desde la ocurrencia del infarto y la incorporación al ensayo.**
- * **Duración del seguimiento de los pacientes.**

El autor encuentra heterogeneidad entre los distintos ensayos y comprueba que, eliminando el sexto estudio, dicha heterogeneidad desaparece. Examina descriptivamente estas características relatando simplemente los valores que toma cada trabajo en las distintas variables y concluye que la divergencia no está explicada por las características recogidas.

En un análisis adicional, con los datos originales de cada ensayo, ajusta los resultados individuales a características basales y observa que la heterogeneidad puede ser explicada principalmente por los años de incorporación al estudio y el tiempo de permanencia. Dichas variables están recogidas como características de los ensayos y sí conseguir los valores brutos es la mejor solución, tal vez se podría extraer mayor información mediante técnicas multivariantes y no con una simple observación de los datos.

Otra conclusión que se deriva del estudio es el alto grado de homogeneidad presente en los ensayos 2, 5 y 6, en relación a otro tipo de respuestas como son el embolismo pulmonar, los problemas gastrointestinales, la presión arterial elevada y la apoplejía.

Análisis Multivariante de los datos de CANNER

Se parte de una matriz de seis filas (ensayos) por siete columnas(variables) y somete a un HJ-biplot (GALINDO, 1986) cuyo descripción formal se ha presentado detalladamente en el capítulo 4.

En la figura 6.9 se representan las configuraciones definidas por los tres primeros ejes que absorben una inercia del 90,16, con una bondad de ajuste para filas y columnas de 85.1783. Los vectores representan a las columnas o variables y los círculos a las filas o ensayos. Debido a la diferencias en la escala de medidas la variables fueron estandarizadas.

En la figura 6.9.a, se representa el primer plano definido por los dos primeros ejes. Si nos fijamos en el primer factor se observa que las variables más importantes en su configuración son *Dosis diaria* y *Número de tomas*, ya que presentan un alta contribución relativa de este factor al elemento y bajas en el resto (ver tabla 6.1). De la misma manera, el segundo de los ejes quedaría configurado por el *número de pacientes* y *duración de seguimiento*.

	EJE I	EJE II	EJE III
<i>Nº de pacientes</i>	40	820	5
<i>Edad</i>	670	270	0
<i>Sexo</i>	480	220	270
<i>Nº de Tomas</i>	740	11	92
<i>Duración seguim.</i>	140	600	33
<i>Tiempo de incorp.</i>	1	3	990
<i>Dosis diarias</i>	910	7	9

Tabla 6.1.- Contribuciones del factor al elemento de las variables.

Si ahora centramos la atención en las filas se observa que el **ensayo uno** queda en posiciones extremas con respecto a eje uno y también, aunque en menor medida, con respecto a eje dos.

Realmente, este ensayo se caracteriza por tomar valores relativos pequeños en todas las variables. Si nos fijamos en el **ensayo 6** vemos que se diferencia del resto con respecto al factor. En este plano, el análisis nos pone en alerta sobre las diferencias del sexto ensayo y además, su separación está relacionada con la variable *duración del seguimiento* lo que concuerda con la conclusión extraída por el autor después de estudiar los datos brutos.

Analizando la información del segundo plano formado por los ejes 1-3 (figura 6.9.b), observamos que aparece de nuevo el **ensayo uno** diferenciándose del resto. El **ensayo dos** toma posiciones extremas respecto el eje tres, definido por el tiempo de incorporación al ensayo.

Es el tercer plano factorial definido por los factores 2-3, el que nos plantea una situación que concuerda con las conclusiones de CANNER en su artículo. El eje dos marca un gradiente diferencial donde el **ensayo seis** toma posiciones extremas en relación al resto, excepto el **ensayo cinco** que aparece en posiciones intermedias. En contraposición, el eje tres diferencia el **ensayo dos** del resto.

En conclusión: el ensayo causante de la heterogeneidad es detectado y además se comprueba el comportamiento diferente de los **ensayos dos y cinco**, señalados por el autor. Automáticamente conocemos que las variables causantes de esta configuración están relacionadas con la duración del seguimiento y el tiempo de incorporación, variables importantes a la hora de explicar las causas del comportamiento.

Obviamente, aparecen otras diferencias con respecto al ensayo uno y a otras variables. Nosotros hemos partido de una solución conocida y a partir de esas conclusiones hemos seguido la pista a ensayos y variables importantes.

Esta aplicación constata la capacidad de los métodos factoriales para detectar los estudios causantes de la heterogeneidad. Posteriormente habrá que analizar las variables *sospechosas* mediante un estudio minucioso y ver si realmente son las responsables de dicha heterogeneidad, como ha sido el caso de este meta-análisis.

Para valorar en su justa medida la contribución de esta familia de métodos multivariantes, hemos de pensar en estudios reales a gran escala, con un elevado número de ensayos y de características, en los cuales el estudio en los términos planteados por CANNER sería absolutamente inviable.

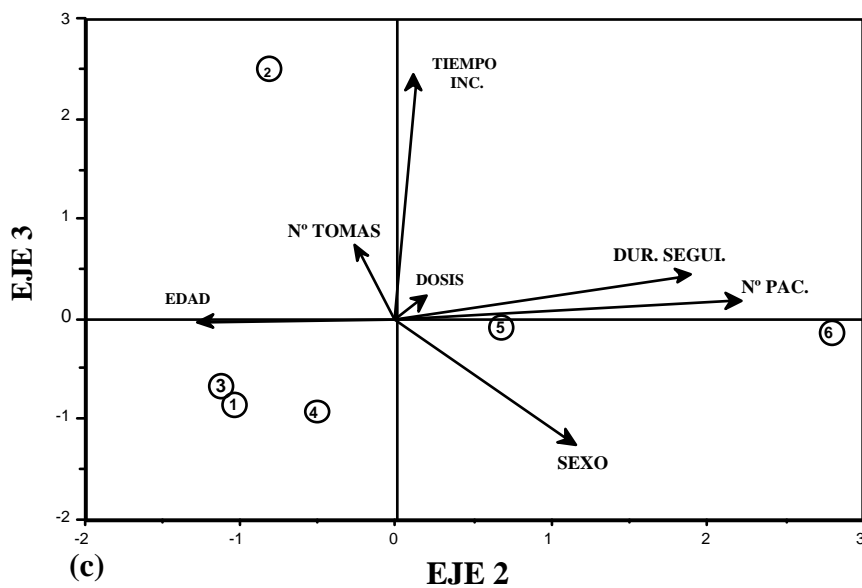
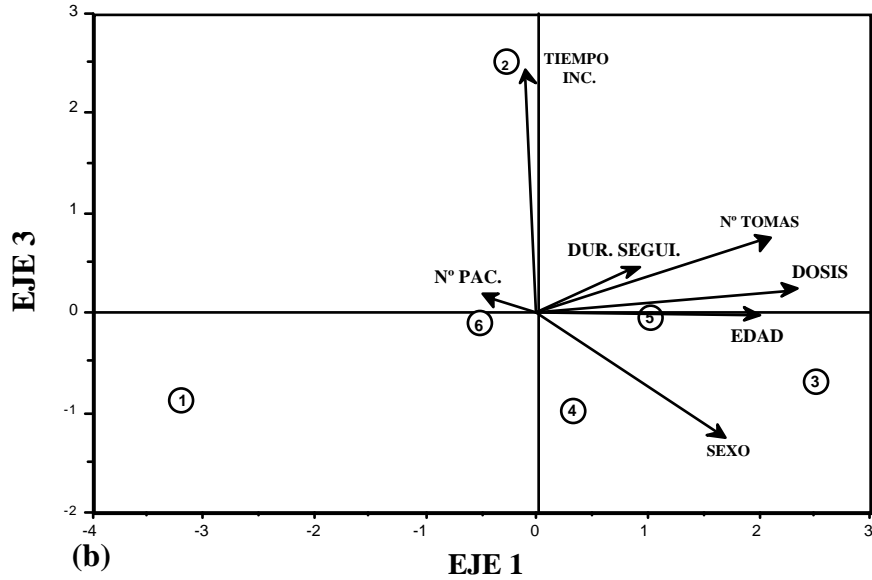
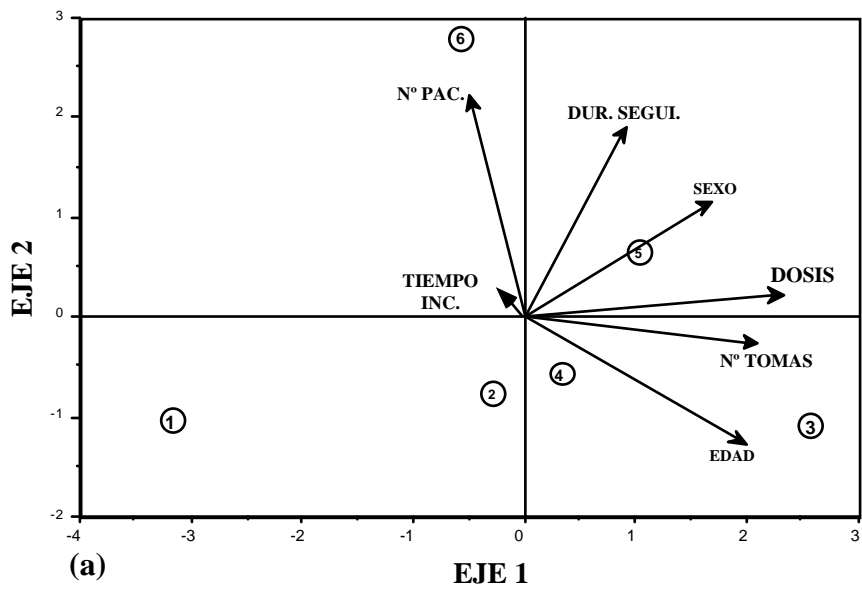


Figura 6.9: Representación HJ-biplot de los planos 1-2 (a), 1-3 (b) y 2-3 (c)

6.3.2.METODOS GRAFICOS MULTIVARIANTES

Se han estudiado aquellos gráficos más relevantes para presentar los resultados y reconocer los posibles tipos de sesgos que se producen en un meta-análisis. Sin embargo, a diferencia de los estudios primarios, no se ha encontrado ninguna revisión cuantitativa que utilice métodos gráficos multivariantes, los cuales poseen una gran aceptación para el análisis de datos (EVERITT, 1994).

En estudios primarios, los gráficos multivariantes son de gran utilidad a la hora de presentar los resultados como confirmación de análisis más complejos, como cualquier tipo de análisis multivariante (AVILA-ZARZA, 1993). La representación gráfica de las características de los estudios nos puede permitir confirmar gráficamente la posible heterogeneidad que se produce entre los tamaños del efecto en relación con las características que los rodean.

Dentro del abanico de posibles gráficos multivariantes (CHAMBERS y cols., 1983; EVERITT, 1978, 1994) nos referiremos a aquellos que presentan mayor difusión dentro de la literatura científica y del mismo modo pueden servir de herramienta en las revisiones cuantitativas para el análisis y presentación de resultados.

6.3.2.i.ESTRELLAS O POLIGONOS

Pertenece a la familia de iconos que intentan definir patrones o estructuras de interés del conjunto de datos (CHAMBERS y KLEINER, 1982). En esta representación cada variable es representada mediante la longitud a

partir de un centro común. Los extremos de los radios son usualmente conectados para formar diferentes formas de estrellas o de polígonos. La asignación de las variables a los radios es importante porque de ella depende la configuración final¹.

Un procedimiento que se propone para el reconocimiento de las figuras se basa en relativizar los datos en base al mayor valor y así obtener un radio máximo de uno (figura 6.10). A partir de un polígono referencia donde todas las variables toman el valor máximo, se superponen los polígonos correspondientes a cada individuo y se obtiene en cada icono una escala de referencia para comparar los radios de las diferentes estrellas

Representamos los seis ensayos del estudio meta-analítico realizado por CANNER (1987), mediante el método de los polígonos o estrellas (figura 6.10).

Se puede observa similitud entre en los ensayos uno, tres, cuatro y cinco, si bien el ensayo uno toma valores más pequeños para la mayoría de las variables y el resto se distinguen en la longitud de un radio.

Por el contrario, los estudios más diferentes se corresponden al ensayo seis y dos. Si seguimos la asignación de las variables a los distintos radios se puede observar como el ensayo seis se diferencia en el tiempo de seguimiento y número de pacientes. El ensayo cinco se distingue de las configuraciones uno, tres y cuatro en el tiempo de seguimiento, mientras que el ensayo dos principalmente en el tiempo de entrada.

1

Este tipo de gráficos, que en teoría son prometedores, en la práctica no suelen resultar muy útiles por las limitaciones de los programas informáticos, ya que éstos sólo presentan de forma automática el contorno de las estrellas, y la identificación de la asignación dada a cada variable se convierte en un proceso francamente laborioso.

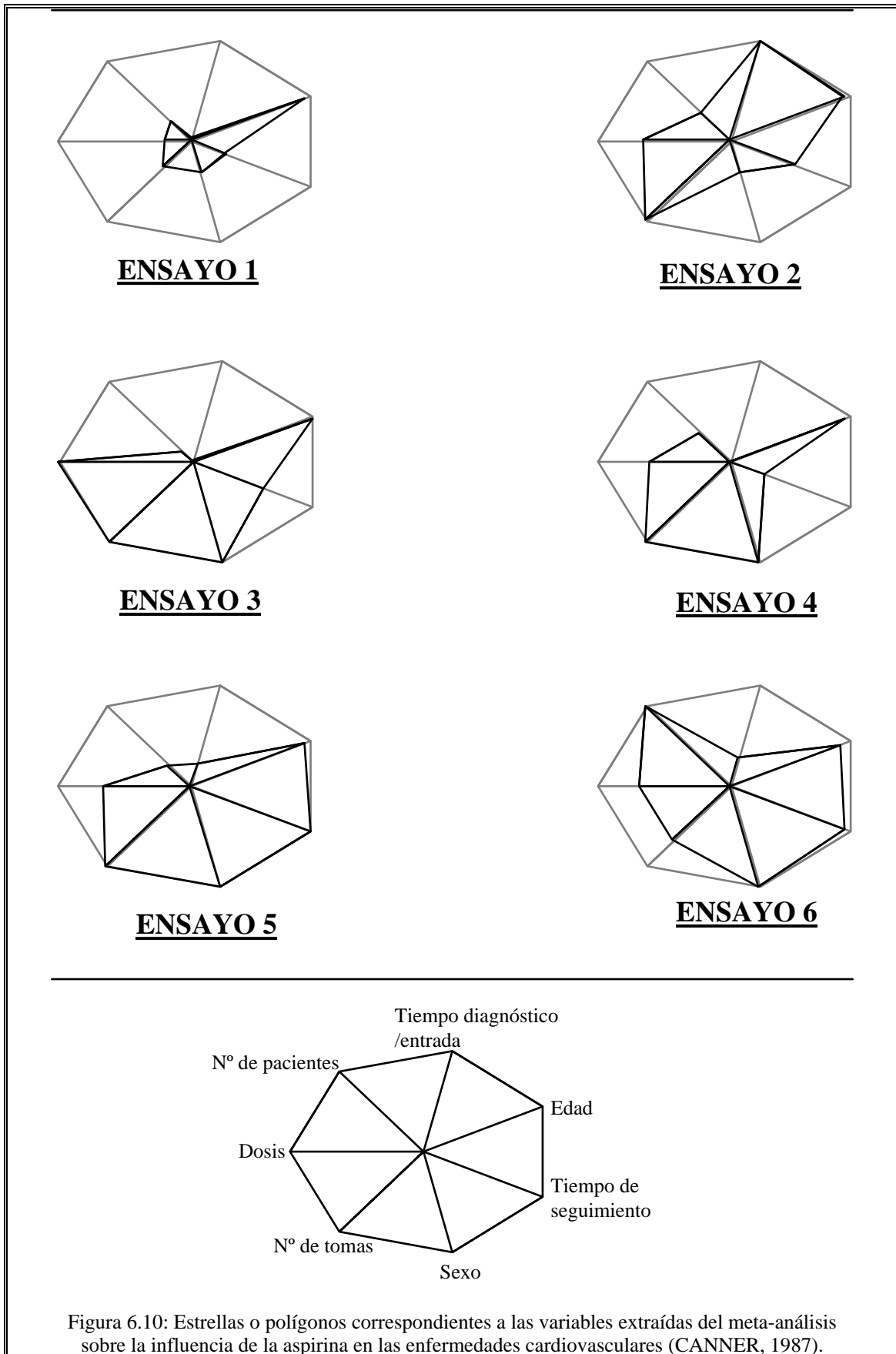
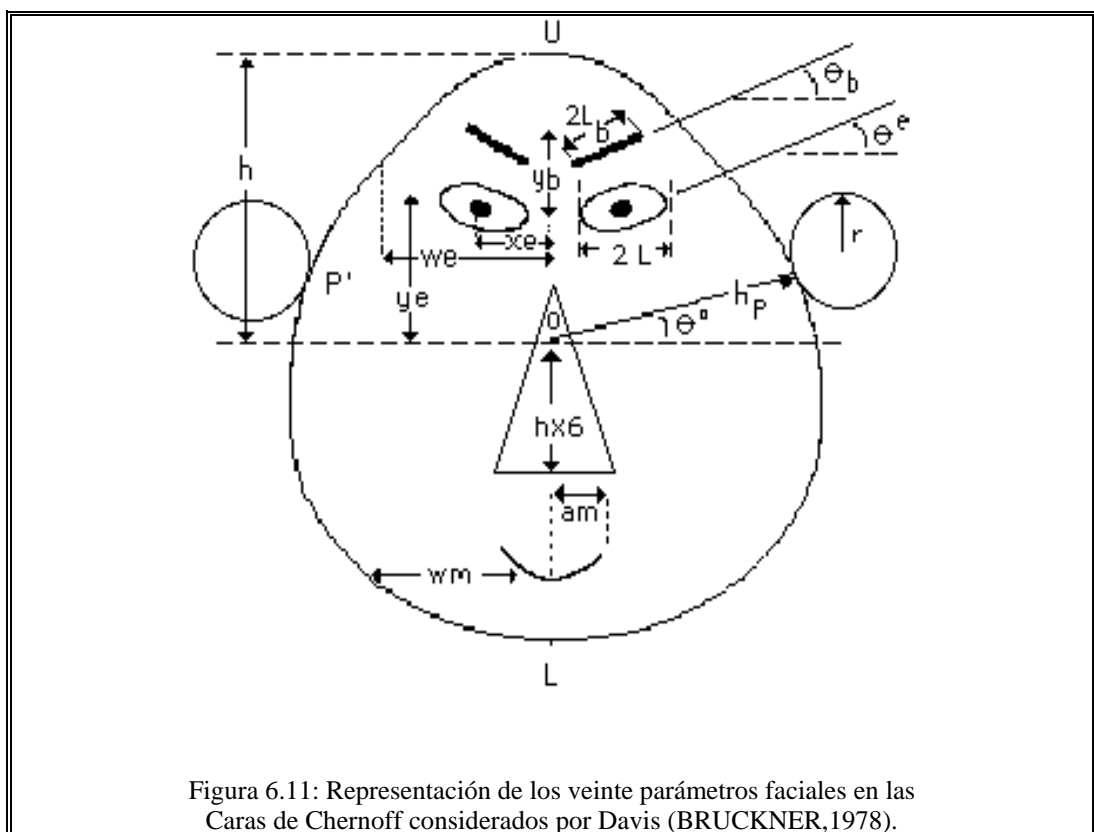


Figura 6.10: Estrellas o polígonos correspondientes a las variables extraídas del meta-análisis sobre la influencia de la aspirina en las enfermedades cardiovasculares (CANNER, 1987).

Como se puede comprobar, este gráfico confirma perfectamente los resultados obtenidos con el método biplot.

6.3.2.ii CARAS DE CHERNOFF

Es un método sencillo y útil para representar un conjunto de variables de manera impactante. Consiste en la asignación de cada una de las variables a un rasgo de una cara, de manera que esta cualidad facial sea proporcional al valor que toma la variable en el individuo que esta representando. Las caras, originariamente, se definieron por 18 parámetros representados por distancias y ángulos (CHERNOFF, 1973), si bien a lo largo de su aplicación e implementación a paquetes informáticas han sido ampliados (figura 6.11.).



El resultado final es un conjunto de caras con diferentes facciones correspondientes a cada uno de los individuos. Esto permite la formación subjetiva de agrupaciones a partir del parecido entre las diferentes caras.

La asignación de cada rasgo facial a las variables depende del objetivo que pretende cubrir el investigador. Varios autores defienden la asignación aleatoria de las variables (ALONSO, 1982), y así, evitar los posibles efectos de la subjetividad de los investigadores. Esta asignación estará generalmente vinculada a los estudios exploratorios donde el investigador desconoce la estructura de sus datos. En esta situación la eficacia de este tipo de gráficos es cuestionada debido a la dificultad de obtención de resultados admisibles (AVILA-ZARZA, 1993). Por el contrario, en estudios confirmatorios o bien en exploratorios donde el investigador posee información adicional, otros autores sostienen la asignación intencionada de las variables que refleje claramente la información extractada del estudio (HUFF y BLACK, 1978).

En meta-análisis consideramos que la utilidad de este método se centra en el marco fijado por los análisis confirmatorios. El objetivo principal se reduce a la presentación de las variables diferenciales que caracterizan los estudios una vez realizados los análisis correspondientes.

Para realizar la representación de los ensayos en función de las variables elegidas se utilizó el programa SYSTAT. En la tabla 6.2 se muestra el orden de aparición y asignación de las variables a los diferentes rasgos que aparece en este programa el cual incluye las variantes de Davis (BRUCKNER, 1978) y Everitt (EVERITT, 1978):

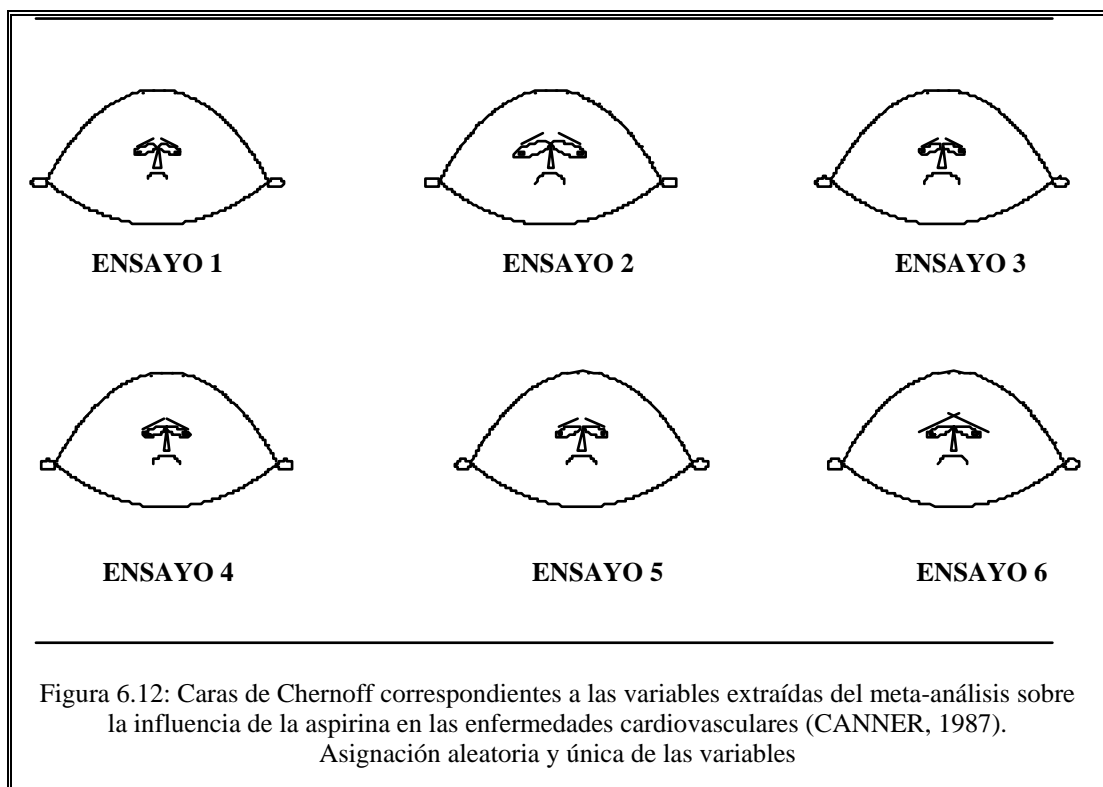
Nº VARIABLE	ORDEN DE LAS FACCIÓNES
1.	Curvatura de la Boca
2.	Angulo de la Ceja
3.	Anchura de la Nariz (DAVIS)
4.	Longitud de la Nariz
5.	Longitud de la Boca
6.	Altura del centro de la Boca
7.	Separación de los Ojos
8.	Altura del centro de los Ojos
9.	Inclinación de los Ojos
10.	Excentricidad de los Ojos
11.	Longitud media de los Ojos
12.	Posición media de las Pupilas
13.	Altura de la Ceja
14.	Longitud de las Cejas
15.	Altura de la Cara
16.	Excentricidad elipse superior Cara
17.	Excentricidad elipse inferior Cara
18.	Nivel de la Oreja
19.	Radio de la Oreja (DAVIS)
20.	Longitud del Pelo (EVERITT)

Tabla 6.2. Asignación de las variables a las facciones de las caras de Chernoff.

Se observa que el número de rasgos de asignación es mayor que el conjunto de variables disponibles en la revisión. Hay fundamentalmente dos opciones para resolver este inconveniente; la primera se basa en la asignación de las siete variables al mismo número de rasgos y el resto se mantendrán constantes (ALONSO, 1982); la otra alternativa estriba en la asignación múltiple de variables a todo el conjunto de rasgos. Esta última opción destaca mejor las diferencias entre estudios al trazar caras que presentan formas mas divergentes.

A partir de los diversos procedimientos de asignación se han representado los ensayos clínicos revisados en función de lo que consideramos los métodos mas extremos:

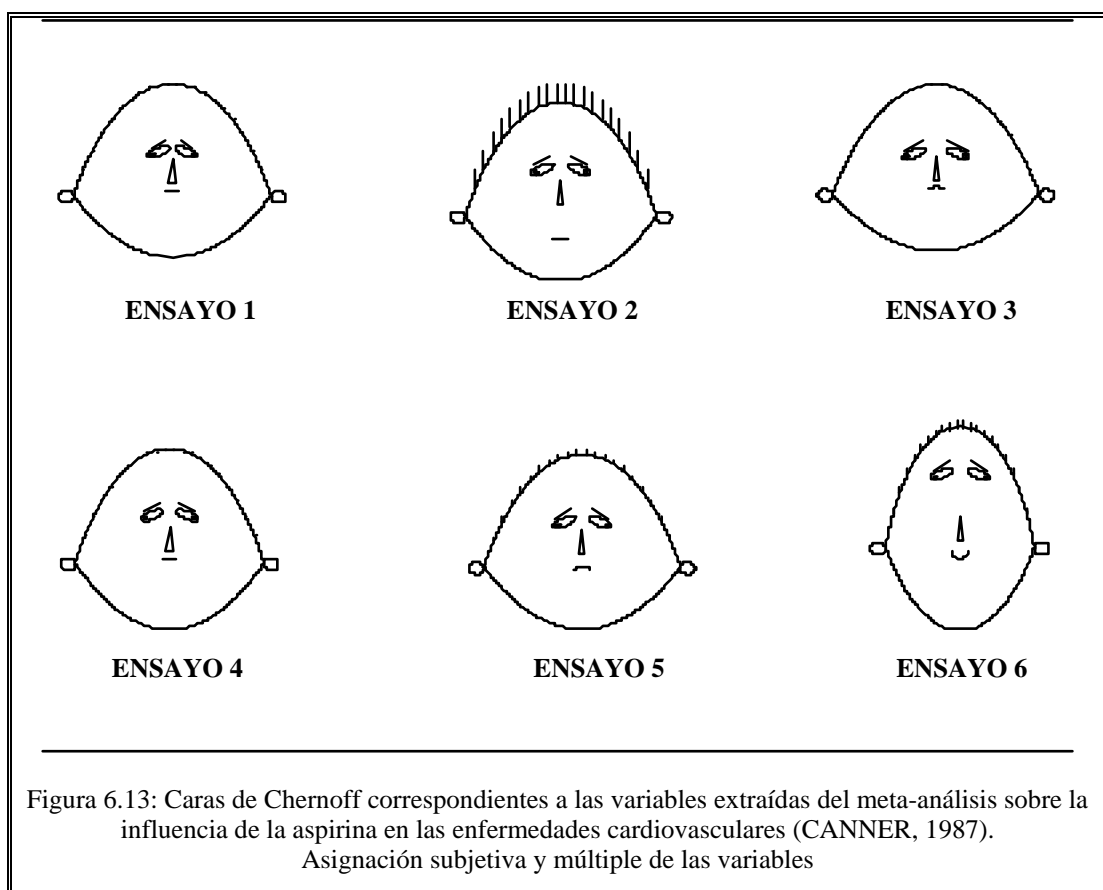
- Una asignación aleatoria donde cada variable viene determinada por un único rasgo y el resto se mantiene constantes (figura 6.12.)
- Asignación subjetiva donde cada variable es asignada a varios rasgos.(figura 6.13)



En los resultados arrojados por el meta-análisis se destaca la heterogeneidad de los estudios donde el **ensayo seis** se identifica como el culpable de dicha heterogeneidad. El análisis detallado de las representaciones de las caras mostraría que variable son culpables de las

diferencias entre el estudio seis y el resto, no obstante, en la asignación subjetiva de las variables (figura 6.13) estas discrepancias son más perceptibles y confirman con mayor claridad las divergencias entre los estudios. Además, en esta representación se diferencia el ensayo dos y en menos medida el cinco, lo que confirmaría los resultados anteriores.

Esto apoyaría la afirmación de AVILA-ZARZA (1993) sobre el poder confirmatorio de los gráficos multivariantes y el escaso poder exploratorio.



6.3.3.iii. CURVAS DE ANDREWS

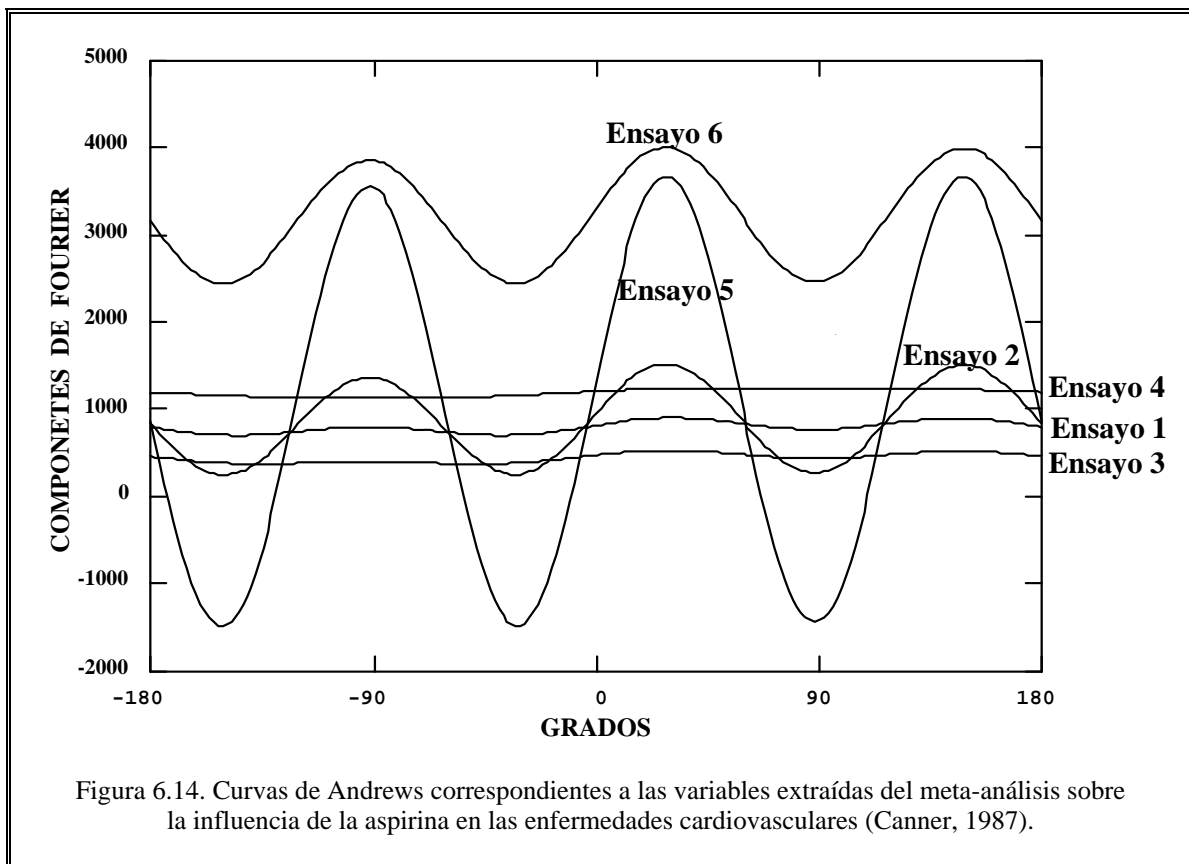
Este método (ANDREWS, 1972), consiste en una representación de datos multivariantes donde cada individuo es caracterizado mediante un grafo de función periódica:

$$f_x(t) = \frac{1}{\sqrt{2}} x_1 + x_2 \text{sen}(t) + x_3 \text{cos}(t) + x_4 \text{sen}(2t) + x_5 \text{cos}(2t) \dots$$

para valores de t comprendidos entre $-\pi < t < \pi$.

El resultado final es un conjunto de curvas que recogen la información multidimensional de cada individuo (estudio) en una única dimensión. La representación de las curvas preserva la distancia original entre los individuos, esta propiedad es importante a la hora de la interpretación del gráfico. Si las curvas son diferentes para el rango de valores t , entonces hay variables que separan a dichos individuos. Esto permite buscar fácilmente valores del parámetro t para la discriminación de grupos.

La representación de las curvas correspondientes al meta-análisis efectuado por CANNER (figura. 6.14), permite reconocer qué estudio (ensayo seis) es el culpable de la heterogeneidad y se pueden reconocer las dos agrupaciones (ensayos dos, cinco y seis frente a uno, tres y cuatro) correspondientes a los análisis ejecutados en la revisión cuantitativa. Una vez más esta configuración apoya las afirmaciones realizadas en las representaciones anteriores.



Como se observa en el gráfico, las curvas reflejan con mayor claridad los resultados que las representaciones procedentes de las caras de Chernoff. Sin embargo, las curvas de Andrews presenta varios inconvenientes entre los que podemos resaltar aquellos que influyen en los aspectos relacionados con la interpretación:

- La dificultad de reconocer las posibles agrupaciones posibles y sus diferencias cuando el número de individuos es grande. CHERNOFF (1973) propone para evitar este problema trazar la curvas en coordenadas polares con lo cual se consigue figuras similares a polígonos irregulares o estrellas.

- Las posibles diferencias en las representaciones debidas al orden de introducción de las variables. En algunas ocasiones las diversas asignaciones producen configuraciones completamente diferentes. Una solución consiste en asignar siempre las primeras x_i a las variables con mayor dispersión obtenidas a partir, por ejemplo, de un análisis de componentes principales (AVILA-ZARZA, 1993). Otra posible solución se basa en un estudio de estabilidad donde se realiza un conjunto de pruebas con diferentes asignaciones y se estudia las variaciones halladas.
- La pérdida de información sobre la relación de nuestros datos y las configuraciones de las curvas.

6.4. COMBINACION DE NIVELES DE PROBABILIDAD

La combinación de probabilidades de diferentes test es uno de los primeros procedimientos estadísticos desarrollados para sintetizar cuantitativamente los resultados de un conjunto de estudios (TIPPET, 1931; FISHER, 1932; PEARSON, 1933), además de representar una de las dos formas con mayor relevancia en la integración de resultados a partir de múltiples test de hipótesis (ROSENTHAL, 1983; STRUBE y HARTMANN, 1983; STRUBE y MILLER, 1986).

Su planteamiento teórico es bastante simple; cada test de hipótesis unilateral contribuye con una probabilidad asociada al nivel de significación. Los diferentes procedimientos utilizan ese valor de probabilidad o alguna transformación, para obtener una estimación de la probabilidad global, considerando que la hipótesis nula compuesta es cierta (todas la hipótesis nulas individuales son ciertas).

Los métodos estadísticos disponibles para la integración de niveles de probabilidad de diferentes estudios independientes, se extienden desde una variedad de procedimientos para englobar los niveles de probabilidad hasta test estadísticos del tipo de la t o de la Z , e incluso introduciendo diferentes tipos de ponderaciones.

6.4.1. TECNICAS DE COMBINACION

6.4.1.i. METODO DE TIPPETT.

Esta técnica (TIPPETT, 1931) basada en un planteamiento muy simple, fue la primera en la combinación de niveles de probabilidad. Parte de $p_1, \dots, p_i, \dots, p_k$ niveles de probabilidad independientes procedentes de k test estadísticos continuos y permite contrastar la *hipótesis nula compuesta* de que ningún nivel p es significativo, frente a la hipótesis alternativa de que al menos uno sí lo es.

El estadístico de contraste es $p_{[i]}$, definido como el nivel de probabilidad más pequeño de todo el conjunto de k niveles,

$$P_{[i]} = \text{mín } [p_1, p_2, \dots, p_i, \dots, p_k]$$

Una vez fijado un nivel de significación, α , la hipótesis nula queda rechazada si se cumple:

$$p_{[i]} < 1 - \sqrt[k]{1 - \alpha}$$

6.4.1.ii. METODO DE FISHER.

Este procedimiento (FISHER, 1932), combina las probabilidades de una serie de K contrastes independientes sobre una hipótesis común con niveles de probabilidad unilaterales $p_1, p_2, \dots, p_i, \dots, p_k$.

Si se considera que cada p_i sigue una distribución uniforme cuando la hipótesis nula es cierta y se tiene en cuenta la relación entre dicha distribución y la ji-cuadrado, entonces, el logaritmo neperiano del producto de las probabilidades multiplicado por -2 se distribuye según una ji-cuadrado con $2k$ grados de libertad cuando la hipótesis nula de cada contraste es cierta.

$$-2 \sum_{i=1}^k \ln p_i$$

ROSENTHAL, 1978; 1980; MOSTELLER y BUSH, 1954 señalan algunos inconvenientes del uso de este método cuando el número de estudios es muy grande, ya que conlleva el rechazo de la hipótesis nula de forma rutinaria y es además muy sensible a la no independencia de los niveles de p . Así mismo, destaca la inconsistencia que se puede producir cuando se combinan estudios igualmente significativos pero en direcciones opuestas.

6.4.1.iii. METODO DE LA SUMA DE NIVELES DE PROBABILIDAD.

Este método (EDGINGTON, 1972a), a partir de la suma de los K diferentes niveles de probabilidad y suponiendo cierta la hipótesis nula, calcula la probabilidad de conseguir una suma de probabilidades igual o menor a la obtenida a partir de los K experimentos ($S = \sum_1^k p_i$).

Si se consideran dos estudios, la probabilidad que interesa viene dada por el área de un triángulo donde la base y altura son iguales a la suma de los dos niveles de probabilidad.

Si son tres estudios la probabilidad será el volumen de la pirámide formada por las aristas iguales a la suma de probabilidades y si son K estudios la probabilidad global del volumen K -dimensional es:

$$p = \frac{S^k}{K!} \quad \text{Donde} \quad S = \sum_{i=1}^k p_i$$

La técnica es aconsejable sólo cuando el número de estudios es bajo, pues de lo contrario la suma S excede la unidad y en este caso la probabilidad global tiende a ser demasiado conservadora, alcanzando en algunas situaciones valores mayores que uno.

EDGINGTON, propone una corrección para evitar el problema de valores mayores que uno basada en la combinación ordenada de valores en una distribución uniforme, la fórmula de la probabilidad global será en ese caso:

$$\frac{S^k}{k!} - \binom{k}{1} \frac{(S-1)^k}{k!} + \binom{k}{2} \frac{(S-2)^k}{k!} - \binom{k}{3} \frac{(S-3)^k}{k!} + \dots$$

La serie alterna el signo hasta que el valor del numerador es menor que la suma S .

6.4.1.iv. METODO DEL CONTRASTE DEL NIVEL p MEDIO

EDGINGTON (1972b), propone un segundo método que permite integrar varios estudios evitando las restricciones de la propuesta anterior y más simple de calcular.

ROSENTHAL (1978) comenta las deficiencias de la técnica cuando el número de niveles de probabilidad (p_i) es inferior a cuatro, si bien, STRUBE y MILLER (1986) consideran que a partir de dos niveles de probabilidad, el método presenta resultados aceptables.

El procedimiento consiste en la transformación de la media de los niveles de probabilidad en una puntuación Z , basándose en la afirmación de que cada uno de los niveles de probabilidad sigue una distribución uniforme con media 0,5 y varianza 1/12 supuesta la hipótesis nula.

$$Z = (0,5 - \bar{p}) \cdot \sqrt{12K}$$

6.4.1.v. METODO DE LA SUMA DE PUNTUACIONES Z .

A este método debido a STOUFFER y cols. (1949), también se le conoce con otros nombres como método Stouffer (WOLF, 1986) o método de la normal inversa (HEDGES y OLKIN, 1985).

Es una de las técnicas para englobar niveles de probabilidad más intuitivas y extendidas en las revisiones cuantitativas.

Consiste en la transformación probit de los niveles p_i en puntuaciones Z e integrar estos valores mediante la siguiente expresión:

$$Z = \frac{\phi^{-1}(p_1) + \dots + \phi^{-1}(p_k)}{\sqrt{k}} = \frac{\sum_{i=1}^k z_i}{\sqrt{k}}$$

Bajo la evidencia de la hipótesis nula, el valor Z sigue una distribución Normal estándar a partir de la cual podemos restituir la transformación para conseguir el valor global p .

ROSENTHAL, 1978 advierte del peligro de aplicar este método en ciertas condiciones debido a que la técnica presenta varianzas unidad lo que puede producir distorsiones en el error tipo I ó II. Además, este método es sensible a la influencia de los posibles niveles de probabilidad aberrantes. Unos pocos niveles extremos pueden variar el resultado final en contra de la dirección marcada por la mayoría de los niveles de probabilidad.

SANER (1994) presenta una modificación de esta técnica para evitar esta influencia de valores extremos. La corrección consiste simplemente en la eliminación de los $(r-1)$ valores más pequeños y los $(k-s)$ más grandes. Donde:

$$1 \leq r \leq s \leq k$$

Entonces se define el estadístico como:

$$Z(r,s) = \sum_{i=r}^s a_i Z_i$$

Donde:

$$a_i = \begin{cases} \frac{1}{\sqrt{s-r-1}}, & \text{si } i = r, \dots, s \\ 0 & \text{en otro caso} \end{cases}$$

Este método alcanza resultados más conservadores que el propuesto por Stouffer (SANER, 1994). Sin embargo, el mayor inconveniente es la eliminación de información la cual produce un cierto sesgo, así como, el carácter subjetivo en la elección de r y s .

A partir del método de Stouffer se presentan varias técnicas para combinar y comparar los diferentes niveles de probabilidad a partir de estudios independientes (ROSHENTAL 1978, 1983; ROSENTHAL y RUBIN , 1979) e incluso en el caso que no se cumpla este supuesto de independencia (STRUBE, 1985).

Una primera suposición para integrar los p_i es considerar que todos son homogéneos y son representantes de un mismo contraste. ROSENTHAL y RUBIN (1979) presentan el siguiente estadígrafo basado en el propuesto por COCHRAN (1954) para combinar varios resultados:

$$X_k^2 = \sum_{i=1}^k (Z_i - \bar{Z})^2$$

Dicho estadígrafo sigue una χ^2 con $k-1$ grados de libertad cuando las puntuaciones Z son independientes. Mediante este contraste el revisor puede estudiar la diversidad de sus resultados, así como las posibles causas de la misma.

Si los estudios presentan valores π_i heterogéneos, ROSENTHAL y RUBIN (1979) aplican un test basado en los contrastes polinomiales ortogonales (ver COCHRAN y COX, 1974) donde una comparación de los valores Z es una combinación lineal de esos mismos valores pudiendo plantear $k-1$ combinaciones lineales correspondientes a $k-1$ contrastes:

$$\tilde{Z} = \lambda_1 Z_1 + \dots + \lambda_i Z_i + \dots + \lambda_k Z_k$$

y donde λ es un número fijo que debe cumplir que su suma es igual a cero, así como, la suma de los productos cruzados para cada par de conjunto λ s es igual a cero:

$$\sum_i \lambda_i = 0 \quad \text{y} \quad \sum_i \lambda_{ij} \lambda_{ik} = 0$$

Entonces el estadígrafo de contraste se obtiene:

$$Z_s = \frac{\sum_i \lambda_i Z_i}{\sqrt{\sum_i \lambda_i^2}}$$

Z_s para tamaños muestrales grandes sigue aproximadamente una distribución Normal.

ROSENTHAL y RUBIN (1979) aplican este tipo de contraste para detectar posibles dependencias de variables externas con los niveles de probabilidad que pueden afectar a la homogeneidad de los datos. Es decir, si se supone la existencia de una variable que presenta una

ordenación ascendente o descendente se pretende conocer si los niveles de probabilidad varían en el mismo sentido que esa variable exógena.

STRUBE (1985) presenta una generalización de estos contrastes a la situación donde los resultados son dependientes, introduciendo una estimación de la correlación que presentan los diferentes resultados:

$$Z_s = \frac{\sum_i \lambda_i Z_i}{\sqrt{\sum_i \lambda_i^2 + 2 \sum_i \lambda_i \lambda_j r_{ij}}} \quad (i \neq j)$$

6.4.1.vi. METODO DE LA SUMA DE PUNTUACIONES Z PONDERADAS.

Esta modificación del método de Stouffer, es una técnica (MOSTELLER y BUSH , 1954) que consiste en ponderar cada puntuación tipificada por un criterio a juicio del investigador, por ejemplo, el tamaño de la muestra o algún índice de validez interna obtenido a través del estudio de codificación de los diferentes trabajos que se quieren englobar.

$$Z = \frac{\sum_i w_i Z_i}{\sqrt{\sum_i w_i^2}}$$

w_i es la ponderación aplicada.

Al igual que en el método de Stouffer, se puede aplicar la modificación de Saner y evitar la influencia de valores aberrantes con altas ponderaciones. Si se pondera por el tamaño de muestra estamos introduciendo doblemente este factor ya que los niveles de p dependen

de los tamaños muestrales y por lo tanto el estimador vendrá determinado exclusivamente por los ensayos de gran tamaño.

6.4.1.vii. METODO DEL CONTRASTE DE LA PUNTUACION Z MEDIA

Se trata de otra modificación propuesta por MOSTELLER y BUSH (1954), que consiste en presentar un test t sobre la puntuación Z media obtenida a partir de los valores Zs correspondientes a los diferentes niveles de probabilidad.²

$$t = \frac{\sum_i z_i}{\sqrt{\frac{S_z^2}{k}}} \quad \text{siendo} \quad S_z^2 = \frac{\sum_i (z_i - \bar{z})^2}{k-1}$$

ROSENTHAL (1978); STRUBE (1985) y STRUBE y MILLER (1986) señalan que esta técnica presenta baja potencia estadística y recomiendan no utilizarla sin un mínimo de 10 estudios para combinar.

² _____

Lo que realmente sigue una t de Student es:

$$t = \frac{\frac{\sum_i z_i}{\sqrt{k}}}{\sqrt{\frac{\sum_i (z_i - \bar{z})^2}{k-1}}} \xrightarrow{\text{multiplicamos y dividimos por } \sqrt{k}} t = \frac{\frac{\sum_i z_i}{\sqrt{k}} \cdot \frac{\sqrt{k}}{\sqrt{k}}}{\sqrt{\frac{\sum_i (z_i - \bar{z})^2}{k-1}}} =$$

$$= \frac{\frac{\sum_i z_i}{\sqrt{k} \cdot \sqrt{k}}}{\sqrt{\frac{\sum_i (z_i - \bar{z})^2}{k-1}}} = \frac{\sum_i z_i}{\sqrt{\frac{S_z^2}{k}}}$$

6.4.1.viii. METODO DE WINER.

El método de Winer (WINER, 1971), es similar al planteamiento propuesto en el método de Stouffer (para tamaños muestrales grandes se obtienen los mismos resultados) y el del nivel p medio.

Combina contrastes independientes que provienen directamente de la distribución muestral de estadísticos t en una puntuación Z basándose en la varianza de la distribución t, obteniendo el estadístico:

$$Z = \frac{\left(\sum_i t_i \right)}{\sqrt{\sum_i \frac{g \cdot l_i}{g \cdot l_i - 2}}}$$

El valor Z sigue aproximadamente una distribución Normal estándar cuando los niveles de probabilidad provienen de tests t asociados a grados de libertad mayores de diez. Una vez obtenido el valor Z se transforma a partir de la distribución estándar en el nivel de probabilidad global.

Evidentemente este procedimiento no puede aplicarse en el caso de que los tamaños muestrales sean menores de tres puesto que en ese caso el denominador sería nulo o incluso negativo.

6.4.1.ix. METODO LOGIT

Otro método para combinar k niveles de probabilidad independientes ha sido desarrollado por GEORGE (1977) y consiste en transformar cada nivel p_i en un unidad logit y combinarlos mediante el estadístico L :

$$L = \sum_i \ln\left(\frac{p}{1-p_i}\right)$$

La distribución de L cuando la hipótesis compuesta es cierta es aproximadamente una t de Student con $5k+4$ grados de libertad excepto una constante:

$$C = \sqrt{\frac{3(5k+4)}{\pi^2(5k+2)k}} \quad L^* = C \cdot |L|$$

6.4.2. CRITICA A LA COMBINACION DE NIVELES DE PROBABILIDAD

Todos estos métodos están contrastados cuando los supuestos de partida (como Normalidad, independencia entre los niveles p_i , equivalencia de varianzas y tamaños muestrales) se cumplen. Sin embargo, cuando las condiciones de los estudios se alejan de estas premisas se producen distorsiones en la potencia de las diferentes técnicas con la consecuente precaución a la hora de inferir conclusiones con respecto al tópico que se está estudiando.

Los inconvenientes más habituales son:

- ***La independencia de los niveles de probabilidad***, ya que es característica la multiplicidad de resultados en un mismo estudio.

En este caso, STRUBE (1985) sugiere un método alternativo basado en el método de Stouffer en el cual incluye la información de la correlación que existe entre las dos o más variables dependientes que intervienen en el estudio.

En Psicología esta información puede estar disponible debido a que normalmente trabajan con las medidas de diferentes tests los cuales han sido estudiados con anterioridad conjuntamente y puede conocerse una estimación de la correlación que existe entre ellos. Sin embargo, en otros campos de investigación es casi imposible disponer de una medida de la correlación que existe entre los resultados. Una solución es calcular a partir de los diferentes p_i de un mismo estudio un único nivel de probabilidad que los integre, siendo éste el representante de ese estudio.

- ***La influencia de los tamaños muestrales en los niveles de probabilidad***. Este inconveniente puede sesgar la estimación global de modo que, si existen varios p_i procedentes de diferentes tamaños muestrales puede ocurrir que la estimación global este influenciada en exceso por los tamaños muestrales.

MOSTELLER y BUSH (1954), con su método, intentan considerar la información del tamaño muestral introduciendo una ponderación que refleje los diferentes tamaños. Sin embargo, esta solución produce una sobrestimación del p global ya que generalmente en cada estudio original a mayores tamaños de muestra obtenemos

mayores valores p_i y posteriormente en la técnica de integración, damos mayor peso a estos niveles de probabilidad. Por lo tanto, consideramos que las ponderaciones en la integración de niveles de probabilidades deben tener en cuenta la calidad interna del estudio, así como, de la existencia de diferentes tamaños que pueden alterar los resultados.

- *El nivel de probabilidad global sólo nos proporciona la información sobre si aceptamos o rechazamos la hipótesis nula compuesta, sin indicarnos cual es el tamaño del efecto o si esas diferencias en los niveles de probabilidad se deben a todos los estudios, a un conjunto de ellos o a un único nivel de probabilidad.*

Para evitar el problema es interesante acompañar al nivel de probabilidad global con una medida del tamaño del efecto que nos informe de la magnitud y la dirección del efecto (ROSENTHAL, 1984).

6.5. INTEGRACION DEL TAMAÑO DEL EFECTO

Los métodos de combinación de los diferentes tamaños del efecto son los más extendidos dentro de la literatura científica (ver entre otros, GLASS y cols., 1981; ROSENTHAL, 1984; HEDGES y OLKIN, 1985; EVERITT, 1992; PAUL y DONNER 1992, FLEISS, 1993). El tamaño del efecto nos proporciona mayor información sobre la magnitud del mismo, no depende del tamaño muestral y nos permite la posibilidad de formular afirmaciones inferenciales (GREEN y HALL, 1984; HEDGES y OLKIN, 1985).

Hay varios enfoques en el estudio del tamaño del efecto global dependiendo del área de investigación donde el revisor realice la integración de los resultados. Sin embargo, la base y el trasfondo de los diferentes métodos es similar puesto que el objetivo es conseguir un estimador de la magnitud del efecto y estudiar su representatividad con respecto al conjunto de todos los estudios.

En relación a esta diversidad de técnicas en los distintos campos de aplicación, se pretende desarrollar un planteamiento general del proceso analítico de los tamaños del efecto con un ulterior estudio de sus particularidades en las diferentes áreas de investigación.

6.5.1. METODOS GENERALES DE INTEGRACION DEL TAMAÑO DEL EFECTO

6.5.1.i. MODELO DE EFECTOS FIJOS.

Supongamos un parámetro θ que hace referencia al tamaño del efecto (diferencia media tipificada, odds-ratio, correlación, etc.) y puede ser estimado a partir de k estudios independientes.

Consideramos $\hat{\theta}_1, \dots, \hat{\theta}_k$, las estimaciones obtenidas a partir de la bibliografía recogida, las cuales siguen una distribución Normal $N(\theta, \sigma_{\hat{\theta}_i})$, y suponemos que los distintos trabajos estiman a un único parámetro poblacional:

$$\theta \quad (\theta_1 = \dots = \theta_i = \dots = \theta_k = \theta).$$

Bajo estos supuestos, un primer estimador de la magnitud del efecto global es la media aritmética de los diferentes estimadores obtenidos:

$$\hat{\theta} = \frac{\sum_{i=1}^k \hat{\theta}_i}{k}$$

Este estimador es adecuado cuando los diferentes tamaños del efecto están calculados con el mismo número de observaciones; sin embargo, es difícil de conseguir esta condición en los estudios meta-analíticos, por lo cual, se buscará un estimador que incorpore esta información.

Una solución es considerar como factor de ponderación el inverso de la varianza que proporciona un estimador en el que se pondera con mayor

peso a los tamaños del efecto más precisos.

Obviamente los estimadores con mayor varianza presentan menores ponderaciones debido a su menor precisión, además de ser calculados con menores tamaños de muestra, debido a la relación existente entre tamaños muestrales y las varianzas de los estimadores de interés.

El estimador óptimo del parámetro poblacional sería de la forma:

$$\hat{\theta} = \frac{\sum_{i=1}^k \hat{\theta}_i \cdot w_i}{\sum_{i=1}^k w_i}$$

donde

$$w_i = \frac{1}{\sigma_{\hat{\theta}_i}^2}$$

La distribución del estimador será una Normal de media θ y de varianza el inverso de la raíz cuadrada de la suma de las ponderaciones³. Es un estimador máximo verosímil ya que $\hat{\theta}_i \equiv N(\theta, \sigma_{\hat{\theta}_i})$

La función de verosimilitud será:

3

$$\hat{\theta}_i \equiv N(\theta, \sigma_{\hat{\theta}_i}) \xrightarrow{\text{si multiplicamos por una cte y luego sumamos}} \sum_{i=1}^k \frac{\hat{\theta}_i}{\sigma_{\hat{\theta}_i}^2} \equiv N\left(\theta \sum_{i=1}^k \frac{1}{\sigma_{\hat{\theta}_i}^2}, \sqrt{\sum_{i=1}^k \frac{1}{\sigma_{\hat{\theta}_i}^2}}\right)$$

entonces,

$$\frac{\sum_{i=1}^k \frac{\hat{\theta}_i}{\sigma_{\hat{\theta}_i}^2}}{\sum_{i=1}^k \frac{1}{\sigma_{\hat{\theta}_i}^2}} \equiv N\left(\theta, \frac{1}{\sum_{i=1}^k \frac{1}{\sigma_{\hat{\theta}_i}^2}}\right)$$

$$L(\hat{\theta}_i) = \prod_{i=1}^k \frac{1}{\sqrt{2\pi} \cdot \sigma_{\hat{\theta}_i}^2} \cdot e^{-\frac{1}{2} \cdot \frac{(\hat{\theta}_i - \theta)^2}{\sigma_{\hat{\theta}_i}^2}} = \prod_{i=1}^k \frac{1}{\sqrt{2\pi} \cdot \sigma_{\hat{\theta}_i}^2} \cdot e^{-\frac{1}{2} \sum_{i=1}^k \frac{(\hat{\theta}_i - \theta)^2}{\sigma_{\hat{\theta}_i}^2}}$$

tomamos logaritmos neperianos:

$$\ln L(\hat{\theta}_i) = \sum_{i=1}^k \frac{1}{\sqrt{2\pi} \cdot \sigma_{\hat{\theta}_i}^2} - \frac{1}{2} \sum_{i=1}^k \frac{(\hat{\theta}_i - \theta)^2}{\sigma_{\hat{\theta}_i}^2}$$

Derivamos con respecto a θ e igualamos a cero:

$$\begin{aligned} \frac{\partial \ln L(\hat{\theta}_i)}{\partial \theta} &= -\frac{1}{2} \sum_{i=1}^k \frac{2(\hat{\theta}_i - \theta)(-1)}{\sigma_{\hat{\theta}_i}^2} = 0 \Rightarrow \sum_{i=1}^k \frac{\hat{\theta}_i}{\sigma_{\hat{\theta}_i}^2} - \theta \sum_{i=1}^k \frac{1}{\sigma_{\hat{\theta}_i}^2} = 0 \Rightarrow \\ &\Rightarrow \hat{\theta} = \frac{\sum_{i=1}^k \frac{\hat{\theta}_i}{\sigma_{\hat{\theta}_i}^2}}{\sum_{i=1}^k \frac{1}{\sigma_{\hat{\theta}_i}^2}} \end{aligned}$$

Obviamente este estimador ponderado será máximo verosímil siempre que la varianza no dependa del parámetro poblacional.

Esta media ponderada de los tamaño de efecto será un buen estimador del parámetro poblacional siempre que se cumpla el supuesto de partida de que los tamaños del efecto son homogéneos, es decir, que $\hat{\theta}$ sea representativo del conjunto de los estudios revisados.

Se pueden confeccionar un conjunto de contrastes encadenados que den información de si los tamaños del efecto son homogéneos y si el parámetro poblacional es diferente de cero:

- 1) $H_0: \theta_1 = \dots = \theta_i = \dots = \theta_k = \theta$
- 2) $H_0: \theta = 0$

Partimos de nuevo del supuesto de que el tamaño del efecto $\hat{\theta}_i$ sigue una distribución normal $N(\theta, \sigma_{\hat{\theta}_i})$. Cuando la hipótesis nula del contraste en el estudio i-ésimo es cierta, entonces:

$$\mathbf{X}_i = \frac{\hat{\theta}_i}{\sigma_{\hat{\theta}_i}} = \hat{\theta}_i \cdot \sqrt{w_i} \equiv N(0,1)$$

y por lo tanto:

$$\mathbf{X}_i^2 = w_i \hat{\theta}_i^2$$

sigue una distribución χ^2 con un grado de libertad.

El revisor está interesado en el análisis de integración todos los grupos, por lo tanto calculamos:

$$\mathbf{X}_T^2 = \sum_{i=1}^k w_i \hat{\theta}_i^2$$

Si no existe efecto en ninguno de los de los k grupos, entonces \mathbf{X}_T^2 sigue una distribución ji-cuadrado con k grados de libertad.

Si el valor obtenido es significativo, entonces concluiremos que hay efecto del tratamiento o tratamientos en alguno de los k grupos. Para conocer si este efecto es consistente a lo largo de los k grupos o si varía de un grupo a otro descompondremos \mathbf{X}_T^2 en dos componentes a partir de la propiedad aditiva de la χ^2 :

$$\mathbf{X}_T^2 = \mathbf{X}_H^2 + \mathbf{X}_M^2$$

La cantidad X_H^2 evalúa el grado de homogeneidad o igualdad entre la k medidas, y la cantidad X_M^2 valora la significación del tamaño medio del efecto. El término X_M^2 se calcula de forma similar a X_T^2 . La magnitud del efecto global considerando todos los grupos es la media ponderada de los k tamaños del efecto. Es decir:

$$\hat{\theta} = \frac{\sum_{i=1}^k \hat{\theta}_i \cdot w_i}{\sum_{i=1}^k w_i}$$

Bajo la hipótesis nula de que el efecto global es cero, $\hat{\theta}$ sigue una distribución normal de media cero y error estándar :

$$\sigma_{\hat{\theta}} = \frac{1}{\sqrt{\sum_i w_i}}$$

Por lo tanto:

$$X_M = \frac{\hat{\theta}}{\sigma_{\hat{\theta}}} = \frac{\sum_i w_i \hat{\theta}_i}{\sqrt{\sum_i w_i}}$$

sigue aproximadamente una distribución Normal estándar. Por consiguiente:

$$X_M^2 = \hat{\theta}^2 \sum_i w_i = \frac{(\sum_i w_i \hat{\theta}_i)^2}{\sum_i w_i}$$

sigue aproximadamente una ji-cuadrado con un grado de libertad.

Una vez obtenidos X_T^2 y X_M^2 , el término X_H^2 es fácilmente calculado por sustracción de la igualdad de la que hemos partido:

$$X_H^2 = X_T^2 - X_M^2 = \sum_i w_i \hat{\theta}_i^2 - \hat{\theta}^2 \sum_i w_i$$

Una expresión equivalente es:

$$X_H^2 = \sum_i w_i (\hat{\theta}_i - \hat{\theta})^2$$

X_H^2 se distribuye aproximadamente como una ji-cuadrado con K-1 grado de libertad.

Los valores de los estadígrafos se han obtenido bajo el supuesto de que la hipótesis nula de los diferentes contrastes es cierta. Sin embargo, el estadígrafo X_H^2 nos proporciona información en relación a si los tamaños del efecto son homogéneos e iguales a un parámetro poblacional fijo⁴.

⁴

 Bajo la hipótesis $H_0: \theta_1 = \dots = \theta_i = \dots = \theta_k = \theta$

$$\begin{array}{ll} \hat{\theta}_i \equiv N(\theta, w_i^{-1/2}) & i = 1, \dots, k \\ (\hat{\theta}_i - \theta)\sqrt{w_i} \equiv N(0, 1) & \\ \sum_{i=1}^k (\hat{\theta}_i - \theta)^2 w_i \equiv \chi_k^2 & \end{array} \qquad \begin{array}{ll} \hat{\theta} \equiv N\left(\theta, \left(\sum_i w_i\right)^{-1/2}\right) \\ (\hat{\theta} - \theta)\sqrt{\sum_i w_i} \equiv N(0, 1) \\ \text{y además} \\ (\hat{\theta} - \theta)^2 \sum_{i=1}^k w_i \equiv \chi_1^2 \end{array}$$

el test para contrastar H_0 es:

$$X_H^2 = \sum_i (\hat{\theta}_i - \hat{\theta})^2 w_i$$

Se demuestra que:

$$\begin{aligned} \sum_i (\hat{\theta}_i - \theta)^2 w_i &= \sum_i \left[(\hat{\theta}_i - \hat{\theta}) - (\hat{\theta} - \theta) \right]^2 w_i = \\ &= \sum_i (\hat{\theta}_i - \hat{\theta})^2 w_i + \sum_i (\hat{\theta} - \theta)^2 w_i - \underbrace{2 \sum_i (\hat{\theta}_i - \hat{\theta})(\hat{\theta} - \theta) w_i}_= 0, \text{ bajo el supuesto de independencia} \\ &= X_H^2 + (\hat{\theta} - \theta)^2 \sum_{i=1}^k w_i \Rightarrow \end{aligned}$$

$$X_H^2 = \underbrace{\sum_i (\hat{\theta}_i - \hat{\theta})^2 w_i}_{\sim \chi_k^2} - \underbrace{(\hat{\theta} - \theta)^2 \sum_{i=1}^k w_i}_{\sim \chi_1^2}$$

Por lo tanto, X_H^2 se distribuye como una chi-cuadrado con K-1 grados de libertad.

La suma de cuadrados de los tamaños del efecto es la que realmente interesa debido a la suposición usual de la existencia de un tamaño del efecto. Aunque en varias situaciones esto no ocurre y se piensa en la existencia de un efecto cuando éste no está presente.

6.5.1.ii. MODELO DE EFECTOS ALEATORIOS.

Hasta el momento se ha estudiado el análisis de los resultados extraídos de los estudios revisados bajo el supuesto de que todos estos estimadores, una vez convertidos a una misma escala de medida, eran estimadores de un mismo parámetro poblacional. La heterogeneidad se consideraba fruto de la influencia de factores externos que producían un efecto sobre los datos. El marco de referencia de los análisis se puede englobar en la suposición de que los efectos son fijos. Otra concepción de situar el contexto de las diferencias que aparecen en los informes de investigación es considerar que cada estudio representa a un tamaño efecto que forma parte de la población de los mismos, es decir, se examinan los k estudios revisados como una muestra a partir de una población de estudios (DerSIMONIAN y LAIRD, 1982; HEDGES y OLKIN, 1985; WHITEHEAD y WHITEHEAD, 1991; FLEISS, 1993; PETITTI, 1994).

De acuerdo con esta concepción no hay un único parámetro del efecto sino que existen un conjunto de efectos $(\theta_1, \dots, \theta_i, \dots, \theta_k)$ que representan a un grupo de observaciones de una población con distribución $N(\theta, \tau^2)$; el ámbito descrito se puede presentar como una aproximación de un modelo de efectos aleatorios donde se pretende estimar el representante, θ , de la población de los diferentes efectos.

Al igual que en el modelo de efectos fijos se supone que los estimadores obtenidos de cada experimento o prueba sigue una distribución $N(\theta_i, \sigma_{\hat{\theta}_i}^2)$ donde cada θ_i sigue una distribución normal de parámetros θ y τ^2 por lo tanto la distribución marginal del estimador es una $N(\theta, \sigma_{\hat{\theta}_i}^2 + \tau^2)$:

$$\hat{\theta}_i \xrightarrow{\sim N(\theta_i, \sigma_{\hat{\theta}_i}^2)} \sim N(\theta, \sigma_{\hat{\theta}_i}^2 + \tau^2)$$

Tras la aceptación de estos supuestos se puede estimar la varianza debida a la diferencia de los estudios mediante el calculo de la esperanza del test Q (DerSIMONIAN y LAIRD, 1986).

$$Q = \sum_i (\hat{\theta}_i - \hat{\theta})^2 w_i = \sum_i (\hat{\theta}_i - \theta)^2 w_i - (\hat{\theta} - \theta)^2 \sum_i w_i$$

$$E(Q) = \sum_i w_i \text{var}(\hat{\theta}_i) - \left(\sum_i w_i \right) \text{var}(\hat{\theta})$$

La varianza del estimador $\hat{\theta}$ puede derivarse del estimador ponderado para el modelo de efectos fijos:

$$\text{var}(\hat{\theta}) = \frac{\sum_i w_i^2 \text{var}(\hat{\theta}_i)}{\left(\sum_i w_i \right)^2} = \frac{\sum_i w_i^2 (\sigma_{\hat{\theta}_i}^2 + \tau^2)}{\left(\sum_i w_i \right)^2} = \frac{1}{\sum_i w_i} + \frac{\tau^2 \sum_i w_i^2}{\left(\sum_i w_i \right)^2}$$

Sustituyendo en la expresión anterior:

$$E(Q) = \sum_i w_i (\sigma_{\hat{\theta}_i}^2 + \tau^2) - \left(\sum_i w_i \right) \left[\frac{1}{\sum_i w_i} + \frac{\tau^2 \sum_i w_i^2}{\left(\sum_i w_i \right)^2} \right] =$$

$$= k + \tau^2 \sum_i w_i - 1 - \frac{\tau^2 \sum_i w_i^2}{\sum_i w_i}$$

Por lo tanto:

$$E(Q) = (k - 1) + \tau^2 \left(\sum_i w_i - \frac{\sum_i w_i^2}{\sum_i w_i} \right)$$

A partir de la expresión obtenida se puede calcular el estimador de τ^2 :

$$\hat{\tau}^2 = \frac{Q_\delta - (k - 1)}{\sum_i w_i - \frac{\sum_i w_i^2}{\sum_i w_i}}$$

El test Q puede contrastar la hipótesis de $\tau^2 = 0$ (HEDGES y OLKIN, 1985; WHITEHEAD y WHITEHEAD, 1991). Si el valor estimado de la varianza entre los estudios es negativa o igual a cero se puede admitir el modelo de efectos fijos. Por el contrario, si $\hat{\tau}^2 > 0$ entonces calcularemos un estimador global $\hat{\theta}^*$ donde se tendrá en cuenta la variabilidad debida a los estudios en las ponderaciones:

$$w_i^* = \frac{1}{\sigma_{\theta_i}^2 + \hat{\tau}^2}$$

Una vez obtenidos los nuevos valores de las ponderaciones se calculará el estimador del representante del universo de tamaños del efecto, así como, el intervalo de confianza correspondiente.

El modelo de efectos aleatorios puede ser admitido como el patrón que mejor describe el panorama donde se desarrollan las revisiones meta-analíticas (DerSIMONIAN y LAIRD, 1982; FLEISS y GROSS, 1991; CARLIN, 1992).

Así, en un meta-análisis donde el conjunto de estudios se diferencian en un sinnúmero de características, desde el tratamiento hasta el tipo de pacientes, es lógico admitir la existencia de una población de efectos de la cual obtenemos una muestra representada por cada uno de los estudios revisados. Asimismo, las posibilidades inferenciales, incluso de predicción de posibles valores futuros, son más generales si comparamos con el modelo de efectos fijos (DEMETS, 1987; BAILE, 1987; BERLIN y cols., 1989,).

Sin embargo, PETO (1987b) y THOMPSON y POCOCK (1991) se refieren al estudio exclusivo de los grupos o investigaciones revisadas con hipótesis que hagan referencia a los mismo ya que son los realmente interesantes y rechaza hipótesis inferenciales más ambiciosas que son difíciles de contrastar con fiabilidad. Además, se debe admitir que el conjunto de estudios obtenidos y sus tamaños del efecto no representan una muestra aleatoria del universo de efectos posible, de modo que el revisor procura recoger el mayor número posible de estudios referentes a los objetivos planteados.

El formato de recolección de "datos" ocasiona que el poder inferencial del análisis se refiera al conjunto de estudios que intervienen en la revisión y por lo tanto el modelo de efectos fijos nos permite inferir a partir de los estudios obtenidos (DEMETS, 1987; BERLIN y cols., 1989; JONES, 1995). Además, si nos referimos al carácter exploratorio del meta-análisis el estudio de heterogeneidad en el modelo de efectos fijos puede proporcionar una panorámica global de posibles factores exógenos que afectan o introducen distorsiones en el tamaño del efecto que pueden aportar nuevas ideas en los protocolos de investigaciones futuras.

La polémica filosófica referente a la utilización de un patrón de efectos fijos o aleatorios se debe reducir a un control minucioso de la heterogeneidad de los resultados, con el control de los posibles factores que pueden afectar a la variabilidad, y si ésta no puede explicarse por el conjunto de características que definen a los estudios, entonces se debe tener presente el modelo de efectos aleatorios, aunque para algunos autores ninguna aproximación es buena cuando la heterogeneidad presente es grande (POCOCK y HUGHES, 1990).

6.5.2. INTEGRACION DE LA DIFERENCIA MEDIA TIPIFICADA.

Partimos de k estudios revisados a partir de los cuales se obtienen como estimadores del tamaño del efecto la diferencia media tipificada $\hat{\delta}_1, \dots, \hat{\delta}_k$. Se asume que cada experimento son replicas de un mismo ensayo y por lo tanto consideramos la hipótesis nula de igualdad de diferencias medias tipificadas:

$$\delta_1 = \dots = \delta_i = \dots = \delta_k = \delta$$

El estimador global de la diferencia media tipificada es la media ponderada donde los pesos vienen definidos por la varianza del estimador:

$$\hat{\delta} = \frac{\sum_{i=1}^k \hat{\delta}_i \cdot w_i}{\sum_{i=1}^k w_i}$$

El problema se plantea en la elección ponderación que depende de la varianza y por lo tanto varía en función de los posibles estimadores de la misma.

HEDGES (1982) y HEDGES y OLKIN (1985) discuten sobre la utilización de los diferentes estimadores dependientes de los pesos que presentan características óptimas para ser utilizado.

Un primer estimador esta basado en que las ponderaciones sean independientes del tamaño del efecto poblacional. Se elige la varianza de los estimadores suponiendo muestras grandes, o bien, eliminando de la expresión de la varianza de d_i , el sumando que depende de la diferencia media tipificada poblacional :

$$w_i = \frac{n_{T_i} n_{C_i}}{n_{T_i} + n_{C_i}} \quad \text{ó} \quad w_i = \frac{n_i^* (m_i - 4)}{c(m_i)^2 m_i}$$

Otra aproximación es utilizar estimaciones de la varianza donde se sustituye el parámetro poblacional por su estimación \mathbf{d}_i , o bien por algún estimador obtenido en estudios anteriores o en pruebas piloto realizadas con este fin.

$$\hat{\sigma}_{d_i}^2 = \frac{n_{T_i} + n_{C_i}}{n_{T_i} n_{C_i}} + \frac{\mathbf{d}_i^2}{2(n_{T_i} + n_{C_i})}$$

HEDGES (1982) discute sobre las ventajas e inconvenientes de ambos estimadores presentando al primer estimador como insesgado pero menos preciso que el segundo el cual presenta sesgo; por ello describe un método iterativo para conseguir reducir este sesgo.

Consiste en recalcular de nuevo la estimación de la varianza pero una vez obtenido el estimador $\bar{\mathbf{d}}_p$ a partir de los \mathbf{d}_i . Este proceso se repite hasta que los estimadores obtenido difieren en pequeñas cantidades, aunque con valores grande de N generalmente el estimador converge en la primera iteración.

Todos los estimadores revisados hasta el momento están basados en tamaños del efecto que provienen de estudios donde se supone que el número de observaciones global es mayor de diez. Generalmente, los estudios están basados en tamaños de muestras óptimos para aplicar esta clase de estimadores, a pesar, de la posible presencia de algún o algunos trabajos que no cumplen esta condición y cuyas deficiencias quedan enmascaradas por el resto. Sin embargo, podemos encontrar algún tópico haya dificultad de conseguir el suficiente número de observaciones en el conjunto de las investigaciones realizadas. Por ello, HEDGES y OLKIN (1985) proponen un estimador de la varianza a partir de la media aritmética, \bar{d} , de los diferentes d_j s obtenidos de los estudios:

$$\hat{\sigma}_{\bar{d}}^2 = a_i + b_i \bar{d}$$

Donde:

$$a_i = \frac{m_i c(m_i)^2}{n_i^* (m_i - 1)}$$

$$b_i = \frac{m_i c(m_i)^2 - (m_i - 1)}{(m_i - 1)}$$

La elección de un estimador u otro depende de las diferentes condiciones de las revisiones. La mayoría de los trabajos meta-analíticos utilizan para el cálculo de la diferencia media tipificada el estimador **D** de Glass (SMITH y GLASS, 1980; WILLIAMS y cols., 1982) o el estimador **d** de Hedges (GIACONIA y HEDGES, 1983; GUREVITCH y cols., 1992).

En situaciones donde los mayoría de los estudios tienen tamaños muestrales grandes ($N > 30$) el mejor estimador global es el basado en ponderaciones cuya varianza no dependa del parámetro poblacional evitando posibles sesgos e influencias de el tamaño del efecto en el cálculo de la varianza.

Si los tamaños muestrales en una gran mayoría de los estudios no son demasiado grandes ($N < 30$) la ponderación óptima será la propuesta por Hedges, la cual depende del parámetro poblacional. En este caso, HEDGES y OLKIN (1985) advierten de los posibles sesgos que se pueden producir debido a la subponderación de los estudios con tamaños del efecto grandes y la sobreponderación de los estimadores con valores pequeños.

Una vez calculado el estimador global del tamaño del efecto, contrastamos la homogeneidad de nuestros datos entorno al efecto global:

$$Q = \sum_i \frac{(d_i - d.)^2}{v(d_i)}$$

Si a un determinado nivel α , Q supera el valor crítico, la hipótesis nula queda rechazada, y en consecuencia el modelo de media global no se ajusta a los datos; es decir, no es adecuado representar el conjunto de estimaciones de la diferencia media tipificada mediante el efecto global, $d.$, ya que no estiman el mismo parámetro δ y por lo tanto podremos buscar las diferentes causas de esta heterogeneidad mediante los análisis generales propuestos en el parte final de este capítulo.

Si la hipótesis nula se acepta, se puede estimar δ mediante un intervalo de confianza de la forma :

$$P \left[d. - Z_{\alpha/2} \sqrt{\sigma_d^2} < \delta < d. + Z_{\alpha/2} \sqrt{\sigma_d^2} \right] = 1 - \alpha$$

$Z_{\alpha/2}$ es el correspondiente percentil de la distribución normal y $V(d.)$ la varianza del TE medio que se obtiene por:

$$\sigma_d^2 = \frac{1}{\sum_i \frac{1}{\sigma_{d_i}^2}}$$

6.5.3. INTEGRACION DEL COEFICIENTE DE CORRELACION.

Lo anteriormente expuesto, puede generalizarse al caso en que tengamos $r_1, r_2, \dots, r_i, \dots, r_k$ coeficientes de correlación obtenidos de k estudios independientes sobre una hipótesis común .

Al igual que en el apartado anterior el modelo del coeficiente de correlación global supone que los k estudios estiman un coeficiente de correlación paramétrico ρ

$$H_0 : \rho_1 = \dots = \rho_k = \rho$$

Para la integración de los coeficientes de correlación generalmente los autores proponen el uso de la transformación Z de Fisher debido a las ventajas que proporciona esta transformación a nivel inferencial (ROSENTHAL, 1984 ; HEDGES y OLKIN, 1985; FLEISS, 1993).

Puede obtenerse un coeficiente de correlación global ponderando las transformaciones de cada coeficiente por la inversa de sus respectivas varianzas.

La Z de Fisher, ponderada media, se puede obtener por

$$Z = \frac{\sum_i Z_{r_i} w_i}{\sum_i w_i} = \frac{\sum_i Z_{r_i} (n_i - 3)}{\sum_i (n_i - 3)}$$

recordamos que la varianza de la Z de Fisher es $1/(n-3)$

Contrastamos la homogeneidad de los estimadores mediante el estadístico:

$$Q = \sum_i (Z_i - Z.)^2 (n_i - 3)$$

La estimación del intervalo de confianza para el coeficiente de correlación poblacional ρ se obtiene mediante:

$$I_{Z.}^{\alpha} = [Z. \pm Z_{\alpha/2} \sqrt{\sigma_{z.}^2}]$$

siendo la varianza:

$$V(Z.) = \frac{1}{\sum_i (n_i - 3)}$$

6.5.4. INTEGRACION DEL TAMAÑO DEL EFECTO EN TABLAS 2x2

La revisión de estudios, generalmente en el campo de la epidemiología, donde las variables de estudio son dicotómicas es un extenso campo debido al número de métodos propuestos para la integración de resultados procedentes de tablas 2x2 (ver por ejemplo; FLEISS, 1981; EVERITT, 1992; ARMITAGE y BERRY, 1992).

En estudios procedentes de análisis de tablas 2x2 es fácil reconstruir los datos brutos a partir de la información cuantitativa que presentan, por ello, varios

métodos están basados en valores originales en lugar de índices de asociación o contrastes.

Un conjunto de estos métodos agrupan los datos originales, y a partir del conjunto total, contrastan la asociación entre las variables o estiman el grado de asociación. Sin embargo, estos estimadores pueden producir resultados poco satisfactorios para los investigadores, como la obtención de estimadores globales que indican no asociación, cuando los estimadores puntuales muestran todo lo contrario.

Este tipo de integración produce tamaños del efecto globales, odds-ratio en particular, con un sesgo hacia la independencia (GART, 1992). Por consiguiente, las técnicas más desarrolladas y contrastadas para englobar resultados son las basadas en el cálculo de la media ponderada de los resultados particulares y los respectivos contrastes de homogeneidad y asociación.

6.5.4.i. INTEGRACION DE ODDS-RATIO

Partimos de un conjunto de odds-ratios de diferentes estudios, $\hat{\psi}_1, \dots, \hat{\psi}_i, \dots, \hat{\psi}_k$; un vez más, se considera que estiman un único tamaño del efecto poblacional ψ . Después de transforman los odds-ratios a sus logaritmos neperianos, debido a las ventajas inferenciales ya comentadas y se calcula el estimador global como:

$$\ln(\hat{\psi}) = \hat{\gamma} = \frac{\sum_{i=1}^k \hat{\gamma}_i \cdot w_i}{\sum_{i=1}^k w_i}$$

Donde:

$$w_i = \frac{1}{\sigma_{\hat{\gamma}_i}^2}$$

y la varianza del estimador es:

$$\sigma_{\hat{\gamma}}^2 = \frac{1}{\sum_i \frac{1}{\sigma_{\hat{\gamma}_i}^2}}$$

Una vez calculado el estimador del efecto, podemos realizar los diferentes contrastes de asociación y homogeneidad de los odds-ratio para el estudio de su representatividad. Sin embargo, este estimador arroja buenos resultados, siempre y cuando el número de observaciones de cada estudio sea grande y el número de estratos (publicaciones) pequeños (FLEISS, 1981).

MANTEL y HAENSZEL (1959) propusieron un nuevo estimador de integración de odds-ratio que es más robusto a la deficiencia de tamaño muestral. Esta basado en las observaciones originales de cada tabla las cuales se integran para obtener un estimador global:

$$R_{MH} = \frac{\sum_i \frac{a_i d_i}{n_i}}{\sum_i \frac{b_i c_i}{n_i}}$$

Cuando el número de individuos es grande ambos estimadores tienden a coincidir.

El problema que presenta R_{MH} es la dificultad en la estimación y elección de las diferentes varianzas propuestas (MANTEL y HANKEY, 1975; MIETTINEN, 1976; MANTEL, 1977; ROBINS y cols., 1986)

debido tanto a los problemas de cálculo como a su validez, lo cual afecta a la elaboración del intervalo de confianza y a las diferentes pruebas inferenciales.

La estimación del tamaño del efecto global deriva de la prueba de significación (MANTEL y HAENSZEL, 1959) para contrastar la hipótesis nula de independencia en el conjunto de experimentos. El contraste, al igual que el test exacto de Fisher, supone que las frecuencias siguen una distribución Hipergeométrica, considerando los marginales fila y columna fijos, a partir de la cual se obtiene el valor esperado y la varianza de la frecuencia de interés (COX y SNELL, 1989). Conociendo:

$$E(a_i) = \frac{n_{i.} \cdot n_{.1}}{N_i}$$

$$VAR(a_i) = \frac{n_{i.} \cdot n_{i2} \cdot n_{.1} \cdot n_{.2}}{N_i^2 (N_i - 1)}$$

el estadígrafo de contraste es:

$$X_{MH}^2 = \frac{\left(\sum_i a_i - \sum_i E(a_i) \right)^2}{\sum_i VAR(a_i)}$$

Bajo la hipótesis de independencia X_{MH}^2 sigue una distribución Ji-cuadrado con un grado de libertad.

A partir de la prueba de significación de Mantel y Haenszel deriva el método de Peto (YUSUF y cols., 1985) el cual obtiene un estimador "no ponderado" como la diferencia entre la frecuencia observada y esperada, $a_i - E(a_i)$, y la varianza:

$$\frac{\sum_i [a_i - E(a_i)]}{\sum_i \text{VAR}(a_i)} \equiv N(0, 1)$$

Dicha expresión es el primer paso del método Newton-Rapshon para la obtención del estimador máximo verosímil condicional del logaritmo del odds-ratio común, tomando como valor inicial en el proceso el cero. Así, se consigue un test eficiente de la hipótesis nula de no asociación como un simple y útil estimador del logaritmo de los odds-ratio, si esa hipótesis es rechazada (YUSUF y cols., 1985). De forma análoga se demuestra que se puede obtener un estimador del logaritmo del odds-ratio en cada experimento:

$$\frac{[a_i - E(a_i)]}{\text{VAR}(a_i)}$$

y la varianza de este estimador es:

$$\sigma_{\hat{\gamma}_i^P}^2 = \frac{1}{\text{VAR}(a_i)}$$

Considerando esta afirmación se puede presentar el estimador global como una media ponderada por los inversos de la varianza:

$$\begin{aligned} \hat{\gamma}^P = \ln \hat{\psi} &= \frac{\sum_i \frac{[a_i - E(a_i)]}{\text{VAR}(a_i)} \cdot w_i}{\sum_i w_i} = \frac{\sum_i \frac{[a_i - E(a_i)]}{\text{VAR}(a_i)} \cdot \frac{1}{\sigma_{\hat{\gamma}_i^P}^2}}{\sum_i \frac{1}{\sigma_{\hat{\gamma}_i^P}^2}} = \\ &= \frac{\sum_i [a_i - E(a_i)]}{\sum_i \text{VAR}(a_i)} \end{aligned}$$

La estimación de la varianza es:

$$\hat{\sigma}_{\hat{\gamma}^p}^2 = \frac{1}{\sum_i \text{VAR}(\mathbf{a}_i)}$$

El estimador es muy similar al propuesto por Mantel y Haenszel y presenta las mismas ventajas de robustez ante la presencia de bajo tamaños de muestra, si bien la varianza de este estimador es fácil de calcular proporcionando mayor flexibilidad para posteriores análisis.

GREENLAND y SALVAN (1990) discuten la aplicación de ambos métodos en el campo de la epidemiología concluyendo que el método de Peto proporción sobrestimaciones de los odds-ratio cuando los totales de filas y columnas no son balanceados (sobre todo si una de las frecuencias de la tabla representa la mayor parte de las observaciones) y cuando el efecto producido por la intervención se aleja de valores próximos a uno. Estos autores aconsejan la aplicación de una adaptación del método de Mantel-Haenszel utilizando la varianza propuesta por ROBINS y cols.,(1986) para obtener un estimador ponderado por la inversa de las varianzas, si bien estas varianzas están calculadas bajo el supuesto de la hipótesis nula de no asociación.

La varianza propuesta por ROBINS y cols. (1986), para este estimador, es:

$$\sigma_{\gamma_{MH}}^2 = \frac{\sum_i F_i}{2 \sum_i R_i^2} + \frac{\sum_i G_i}{2 \sum_i R_i S_i} + \frac{\sum_i H_i}{2 \sum_i S_i^2}$$

Donde:

$$R_i = \frac{\mathbf{a}_i \mathbf{d}_i}{N_i} ; \quad S_i = \frac{\mathbf{b}_i \mathbf{c}_i}{N_i} ; \quad F_i = \frac{R_i (\mathbf{a}_i + \mathbf{d}_i)}{N_i}$$

$$G_i = \frac{R_i (\mathbf{b}_i + \mathbf{c}_i) + S_i (\mathbf{a}_i + \mathbf{d}_i)}{N_i} ; \quad H_i = \frac{S_i (\mathbf{b}_i + \mathbf{c}_i)}{N_i}$$

Una vez obtenida la estimación del tamaño del efecto mediante cualquiera de los procedimientos expuestos, podemos contrastar la hipótesis de unicidad de parámetro mediante el test de homogeneidad:

$$Q_\gamma = \sum_i (\hat{\gamma}_i - \hat{\gamma})^2 w_i$$

En el caso de utilizar el método de Peto es más conocida la expresión :

$$Q_\gamma = \sum_i \frac{[a_i - E(a_i)]^2}{\text{VAR}(a_i)} - \frac{\left\{ \sum_i [a_i - E(a_i)] \right\}^2}{\sum_i \text{VAR}(a_i)}$$

Si se concluye que el conjunto de estimadores representan a un único parámetro poblacional entonces el siguiente paso es la construcción de el intervalo de confianza:

$$I_\gamma^\alpha = \left[\hat{\gamma} \pm Z_{\alpha/2} \sqrt{\sigma_{\hat{\gamma}}^2} \right]$$

Otro conjunto de estimadores de el odds-ratio global bajo las condiciones de independencia y homogeneidad derivan de las estimaciones máximo verosímiles tanto condicionales como no condicionales en base a las distribuciones binomial e hipergeométrica respectivamente⁵ (COX y

⁵ _____

La función de verosimilitud del estimador máximo verosímil:

$$L = \prod_{i=1}^2 \prod_{k=1}^k P_{ik}^{X_{ik}} Q_{ik}^{N_{ik}-X_{ik}}, \quad P_{2k} = \frac{P_{1k}}{\psi Q_{1k} + P_{1k}}$$

La función verosimilitud condicional de los marginales fijos T:

$$L' = \prod_{k=1}^k \frac{\binom{N_{1k}}{X_{1k}} \binom{N_{2k}}{T_k - X_{1k}} \psi^{X_{1k}}}{\sum_{u=a_k}^{b_k} \binom{N_{1k}}{u} \binom{N_{2k}}{T_k - u} \psi^u},$$

Donde $a = \max(0, T_k - N_{2k})$ y $b_k = \min(T_k - N_{1k})$

cols., 1989; BRESLOW, 1981; HAUCK, 1984).

Los estudios de simulación que comparan este tipo de estimadores con los expuestos reflejan pequeñas diferencias, tanto en eficiencia del estimador como en precisión a la hora de estimar los intervalos de confianza (BRESLOW y LIANG, 1982, HAUCK, 1984; EMERSON, 1994).

En la actualidad, debido a los soportes informáticos fácilmente disponibles para los investigadores, el problema de cálculo iterativo de los estimadores máximo verosímiles deja de ser un inconveniente. La mayoría de los paquetes estadísticos avanzados permiten el cálculo de dichos estimadores.

Sin embargo, hasta ahora las aplicaciones prácticas de este tipo de estimadores son inexistentes y la robustez de los mismos no están contrastados como ocurre en los estimadores propuestos por Mantel-Haenszel o Peto. Quizás las ventajas puedan apreciarse en los casos donde las condiciones de aplicación de estos últimos estimadores proporcionan malos resultados.

Cuando los datos obtenidos se alejan de las restricción de independencia los resultados pueden verse afectados en mayor grado que si se utilizan los estimadores habituales (LIANG, 1985).

Otra familia de estimadores son aquellos que provienen de métodos no iterativos y que son asintóticamente óptimos bajo los supuestos de

independencia y homogeneidad (EJIGOU, 1990; PIGEOT, 1990; 1991). Al igual que los anteriores es necesario contrastar la robustez de estos estimadores mediante su aplicación a estudios reales.

BRESLOW y LIANG (1982) proponen un estimador Jackknife, basado en el método original propuesto por QUENOUILLE (1949). A partir del logaritmo del estimador de Mantel y Haenszel R_{MH} se calcula el estimador Jackknife mediante los pseudovalores que se obtienen de omitir el resultado de un estudio completo en diferentes etapas.

$$J_i = k \ln R_{MH} - (k-1) \ln R_{MH,i}$$

El estimador será:

$$J_{R_{MH}} = \frac{\sum_{i=1}^k J_i}{k}$$

y varianza:

$$\text{Var}(J_{R_{MH}}) = \frac{1}{k(k-1)} \sum_{i=1}^k (J_k - J_{R_{MH}})^2$$

PIGEOT (1991) propone una variación del método donde la diferencia estriba en la obtención de los pseudovalores. En esta aproximación se omite cada una de las observaciones que aparecen en las distintas tablas. Así, se consigue tantos pseudovalores como valores intervienen en la revisión. El cálculo es más aparatoso aunque la estructura es la misma.

$$J_{a,i} = N R_{MH} - (N-1) R_{MH_{a,i}}, \quad N = \sum_i^k N_i$$

donde:

$$R_{MH_{a,i}} = \frac{\frac{(a_i - 1)d_i}{N_i} + \sum_{j=1, j \neq i}^k \frac{a_j d_j}{N_j}}{\frac{b_i c_i}{N_i} + \sum_{j=1, j \neq i}^k \frac{b_j c_j}{N_j}}$$

De la misma manera pueden ser obtenidos el resto de pseudovalores J_{bi} , J_{ci} , J_{di} . Así el nuevo estimador Jackknife se expresa como:

$$J_P = \frac{1}{N} \sum_{i=1}^k a_i J_{a,i} + b_i J_{b,i} + c_i J_{c,i} + d_i J_{d,i}$$

y la varianza:

$$V(J_P) = \frac{1}{N} \frac{1}{N-1} \sum_{i=1}^k a_i (J_{a,i} - J_P)^2 + b_i (J_{b,i} - J_P)^2 + c_i (J_{c,i} - J_P)^2 + d_i (J_{d,i} - J_P)^2$$

Estos estimadores presentan una tendencia a subestimar el verdadero odds ratio cuando sus valores son grandes y el problema se complica cuando las tablas contienen pocos valores por celda (PIGEOT, 1991).

Este método se puede generalizar a todo el conjunto de estimadores del tamaño del efecto como la diferencia media tipificada o el coeficiente de correlación, etc.

6.5.4.ii. INTEGRACION DE DIFERENCIA DE PROPORCIONES.

Un primer test para contrastar la existencia de diferencia de proporciones extraídas de un conjunto de muestras de datos binarios fue propuesto por COCHRAN (1954). Este estadígrafo toma como ponderaciones el inverso de la suma de los inversos de los tamaños muestrales y presenta la ventaja de ser sensible a detectar pequeñas diferencias sistemáticas entre las proporciones de tal magnitud que la diferencia de sus logit sea constante (SNEDECOR y COCHRAN, 1971).

Se considera:

$$\mathbf{d}_i = \hat{\mathbf{p}}_{i1} - \hat{\mathbf{p}}_{i2} \quad \text{y} \quad \mathbf{w}_i = \frac{\mathbf{n}_{i1}\mathbf{n}_{i2}}{\mathbf{n}_{i1} + \mathbf{n}_{i2}}$$

El estadígrafo se define como:

$$\frac{\sum_i \mathbf{d}_i \mathbf{w}_i}{\sqrt{\sum_i \mathbf{w}_i \hat{\mathbf{p}}_{0i} \hat{\mathbf{q}}_{0i}}}$$

donde:

$$\hat{\mathbf{p}}_{0i} = \frac{\mathbf{n}_{i1} \hat{\mathbf{p}}_{i1} + \mathbf{n}_{i2} \hat{\mathbf{p}}_{i2}}{\mathbf{n}_{i1} + \mathbf{n}_{i2}}$$

Si los tamaños muestrales de los diferentes grupos son grandes el estadígrafo sigue una distribución Normal estándar a partir de la cual podemos contrastar la hipótesis de que no existe un efecto a un nivel de significación determinado.

Para la construcción del test se puede definir un estimador del efecto global como la media ponderada donde los pesos están definidos de la misma forma que en el estadígrafo de contraste:

$$\mathbf{d}_c = \frac{\sum_i \mathbf{d}_i \mathbf{w}_i}{\sum_i \mathbf{w}_i}$$

y varianza:

$$\hat{\sigma}_{dc}^2 = \frac{\sum_i w_i \hat{p}_{0i} \hat{q}_{0i}}{\left(\sum_i w_i \right)^2}$$

El estimador medio ponderado y su varianza es un artefacto para derivar el estadígrafo de contraste de modo que si éste fuera significativo, la varianza del efecto medio estimado no sería representativa al ser obtenida bajo el supuesto de igualdad de proporciones.

La varianza que corresponde al estimador ponderado es:

$$\hat{\sigma}_{dc}^2 = \frac{\sum_i \left(\frac{\hat{p}_{i1} \hat{q}_{i1}}{n_{i1}} + \frac{\hat{p}_{i2} \hat{q}_{i2}}{n_{i2}} \right) w_i^2}{\left(\sum_i w_i \right)^2}$$

Otro problema añadido al método es la imposibilidad de aplicación del test de homogeneidad tal como se ha definido debido a la estructura de las ponderaciones definidas. Este procedimiento a causa de sus limitaciones es poco utilizado y su importancia trasciende del contraste de hipótesis a partir del cual han derivado la gran mayoría de los métodos de integración de datos binario, así como, sus diferentes test para contrastar la asociación global.

El estimador, derivado del anterior y más extendido en las aplicaciones prácticas, es el definido como la media ponderada, donde los pesos son el inverso de la varianza de la diferencia de proporciones (ROSENTHAL y

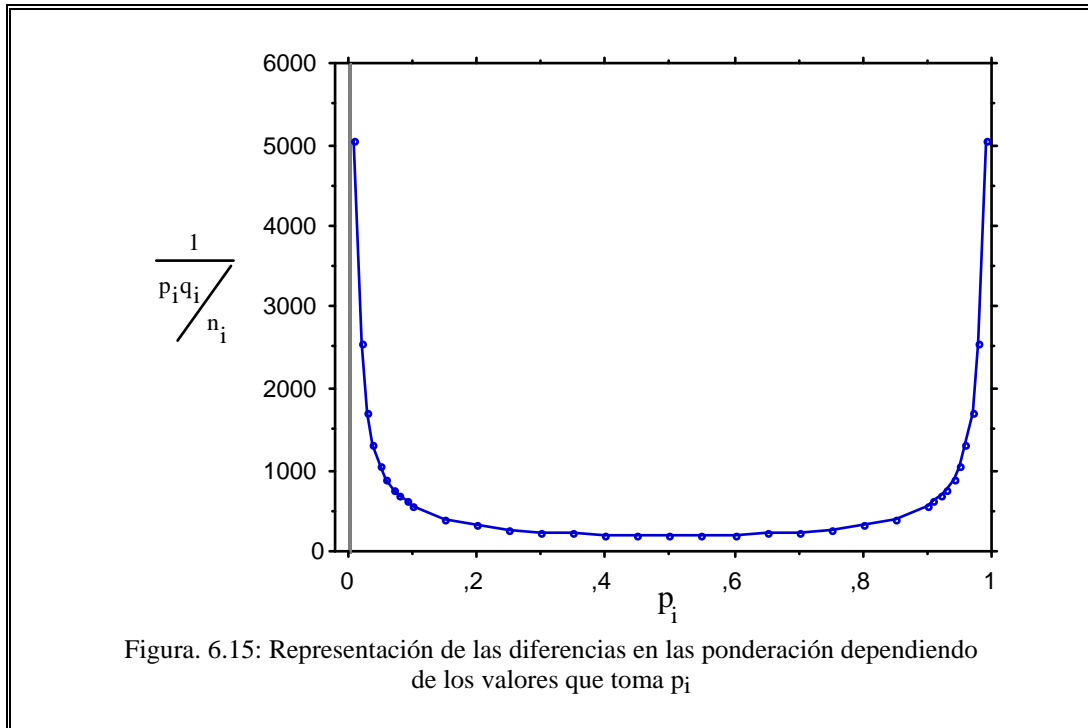
RUBIN, 1982a; DerSIMONIAN y LAIRD, 1986):

$$d_p = \frac{\sum_i d_i w_i}{\sum_i w_i} ; \quad w_i = \frac{1}{\frac{\hat{p}_{i1} \hat{q}_{i1}}{n_{i1}} + \frac{\hat{p}_{i2} \hat{q}_{i2}}{n_{i2}}}$$

Como en procedimientos anteriores se parte de la hipótesis de igualdad de magnitud del efecto en los diferentes estudios, lo cual implica la aplicación del test de homogeneidad para contrastar tal afirmación. Una vez contrastada esta hipótesis, si los resultados son no significativos, se obtienen los extremos de los intervalos de confianza:

$$I_{\delta}^2 = [d_p \pm Z_{\alpha/2} \sigma_{d_p}^2]$$

Un inconveniente de este estimador es la dependencia existente entre la varianza y las proporciones muestrales, lo que ocasiona distorsiones en situaciones particulares (ver figura 6.15).



Así, cuando los valores de los p_{ij} son próximos a cero ó a uno y los tamaños de muestra difieren en gran medida de un grupo a otro se puede producir un sesgo en la estimación del efecto global al presentar mayores pesos este tipo de experimentos. (ROSENTHAL y RUBIN, 1982a ; EVERITT, 1992; MOSTELLER y CHALMERS, 1992).

Este problema, incluso puede afectar los resultados obtenidos al aplicar el test de homogeneidad, lo que origina posibles contradicciones tales como concluir la existencia de heterogeneidad cuando los diferentes resultados particulares presentan odds-ratios iguales. Este problema fue abordado por FLEISS (1981) en relación a un procedimiento de integración similar que propuso a partir del test de Cochran y el cual critica en este sentido.

DerSIMONIAN y LAIRD (1986) proponen un estimador considerando el método de efectos aleatorios donde parten del supuesto de que cada estudio presenta su propio tamaño del efecto δ_i con distribución Normal $N(\delta, \tau)$ e intentan estimar el representante de esta población de efectos.

Recordamos la estimación de la varianza τ^2 la cual se incorpora a las ponderaciones.

$$\hat{\tau}^2 = \frac{Q_{\hat{\delta}} - (k-1)}{\sum_i w_i - \frac{\left(\sum_i w_i\right)^2}{\sum_i w_i}}$$

El nuevo estimador de la magnitud del efecto será la media ponderada donde se incorpora la varianza debida a los diferentes estudios:

$$w_i = \frac{1}{\hat{\sigma}_{d_i}^2 + \hat{\tau}_i^2}$$

BERLIN y cols. (1989) estudian las diferencias entre el estimador de Peto y el basado en efectos aleatorios propuesto por DerSimonian y Laird, concluyendo que ambos estimadores alcanzan resultados similares especialmente cuando la heterogeneidad no es muy grande.

Los factores que influyen en la decisión de elegir un estimador u otro debe depender de los objetivos que pretende cubrir en la revisión. Así, la interpretabilidad de dicho estimador juega un papel importante a la hora de trabajar con la diferencia de proporciones puesto que ofrece mayor utilidad en la valoración del impacto del tratamiento o intervención que el odds-ratio (BERLIN y cols., 1989; EDDY y cols., 1992). No obstante, ante el uso de este estimador se debe prestar atención a las ponderaciones cuando las proporciones estimadas sean pequeñas o grandes.

Por otro lado, el uso de los odds-ratio es más generalizado, especialmente en el campo de la epidemiología, lo que obliga al revisor

al uso de procedimientos expresados en términos de ratio de proporciones.

Un componente importante en la elección del procedimiento de integración se sustenta sobre la etapa de codificación de los estudios. Mediante la inspección minuciosa de las características de cada uno de los informes de investigación, el revisor puede decidir cual es el mejor estimador para integrar todos los resultados considerando las ventajas y limitaciones de cada uno de ellos. Otro factor limitante es la información cuantitativa que proporcionan los estudios, así una gran mayoría presentan exclusivamente los índices de asociación o test de contraste lo cual impide la utilización de los métodos basados en las tablas 2x2.

6.6. HETEROGENEIDAD DE LOS TAMAÑOS DEL EFECTO

El test de homogeneidad, basado en el test propuesto por COCHRAN (1954), es usualmente utilizado en meta-análisis por diferentes autores (GLASS, 1981; HEDGES, 1982; ROSENTHAL y RUBIN, 1982b; WHITEHEAD y WHITEHEAD, 1991) y está diseñado para detectar situaciones en las cuales el tamaño de efecto no representa al conjunto de estudios revisados, lo cual advierte al revisor de la posible influencia de variables exógenas que caracterizan a los diferentes trabajos de investigación en los resultados obtenidos o, en el peor de los casos, del erróneo planteamiento de los objetivos al integrar resultados que nos están situados en un mismo marco de referencia. En relación a la influencia de variables externas, LIGHT y SMITH (1971) argumentan que varias de las contradicciones que aparecen en las revisiones cuantitativas podrían ser resueltas por el agrupamiento de estudios con características similares.

Cuanto más en común tengan los estudios más fácil será obtener tamaños del efecto homogéneo, además, de considerar las discrepancias entre los resultados no como un problema sino como la oportunidad para aprender sobre la efectividad de los diferentes tratamientos o intervenciones (LIGHT, 1979; PILLEMER y LIGHT, 1982).

Por el contrario, algunos autores advierten de los posibles riesgos que se pueden producir como consecuencia de la disgregación de los datos en subconjuntos definidos por significaciones pequeñas o moderadas (PETO, 1987b; COLLINS y cols., 1987; YUSUF, 1987). Al igual que en los ensayos clínicos individuales prefieren desconfiar de estos resultados por las posibles significaciones sin sentido que podrían aparecer por la disminución de los tamaños muestrales.

HEDGES y OLKIN (1985) proponen un método basado en el análisis de la varianza debido a que este método proporciona resultados deficientes porque las observaciones individuales que se utilizan en este tipo de investigaciones presentan un error dependiente del tamaño de muestra, y por lo tanto, incumple los supuestos del Análisis de la Varianza. El método asume la influencia de una o varias dimensiones⁶ categorizadas, relacionadas con las características de los estudios, en la heterogeneidad de los resultados.

Se parte de la hipótesis nula de que todos los tamaños del efecto son homogéneos en el conjunto de resultados revisados y consideramos el análisis del contraste X_H^2 anteriormente estudiado:

$$X_H^2 = \sum_i w_i (\hat{\theta}_i - \hat{\theta})^2$$

Comparamos el valor experimental con un Ji-cuadrado con k-1 grados de

⁶_____

Estas dimensiones pueden proceder de los análisis multivariantes realizado con las características de los estudios

libertad y un α considerado. Si los resultados no son significativos entonces concluiremos de la manera expuesta de que todos los tamaños del efecto son homogéneos y podremos estimar una magnitud del efecto que represente al conjunto de los estudios. Si por el contrario los resultados son significativos, entonces, a partir de la primera dimensión que se considera más relacionada con el tamaño del efecto, efectuaremos una primera división de los tamaños del efecto en p grupos considerando homogeneidad en las diferentes clases.

Análogamente al Análisis de la Varianza, fragmentamos la cantidad que hace referencia a la homogeneidad de los tamaños del efecto en una cantidad que alude a la uniformidad entre las clases y otra a la uniformidad dentro de las clases. Así:

$$\mathbf{X}_H^2 = \mathbf{X}_E^2 + \mathbf{X}_D^2$$

donde:

$$\mathbf{X}_E^2 = \sum_{i=1}^p \sum_{j=1}^{k_i} (\hat{\theta}_{i.} - \hat{\theta})^2 w_{ij} = \sum_{i=1}^p (\hat{\theta}_{i.} - \hat{\theta})^2 w_{i.} \xrightarrow{\text{Si la hipótesis de homogeneidad de tamaños del efecto es cierta}} \chi_{p-1}^2$$

$$\mathbf{X}_D^2 = \sum_{i=1}^p \sum_{j=1}^{k_i} (\hat{\theta}_{ij} - \hat{\theta}_{i.})^2 w_{ij} \xrightarrow{\text{Si la hipótesis de homogeneidad de tamaños del efecto es cierta}} \chi_{k-p}^2$$

Consideramos:

$$w_{ij} = \frac{1}{\sigma_{\hat{\theta}_{ij}}^2} \quad \hat{\theta} = \frac{\sum_{i=1}^p \sum_{j=1}^{k_i} \hat{\theta}_{ij} w_{ij}}{\sum_{i=1}^p \sum_{j=1}^{k_i} w_{ij}} \quad \hat{\theta}_{i.} = \frac{\sum_{j=1}^{k_i} \hat{\theta}_{ij} w_{ij}}{\sum_{j=1}^{k_i} w_{ij}}$$

$$w_{i.} = \sum_{j=1}^{k_i} \frac{1}{\sigma_{\hat{\theta}_{ij}}^2}$$

Si X_D^2 toma valores bajos, o es no significativo, entonces podemos considerar la existencia de p grupos homogéneos y a partir de ahí calcular los tamaños del efecto globales para cada grupo, estudiando las condiciones que han ocasionado esta división para posibles investigaciones futuras. Si esto no ocurre, es decir, X_D^2 toma valores altos o es significativo, se repite el proceso considerando una nueva dimensión y extrayendo de la suma de cuadrados total unos nuevos valores de X_E^2 y X_D^2 para el nuevo conjunto de grupos formado por las dos dimensiones (figura 6.16.). El proceso finalizará cuando X_D^2 sea lo suficientemente pequeño para considerar que los grupos que hemos obtenido a partir de las variables consideradas, son homogéneos.

Otro objetivo propuesto, que puede ayudar a conocer las posibles causas de la heterogeneidad, es averiguar qué estudio o estudios son los posibles causantes de la diversidad de los datos. HEDGES y OLKIN (1985) presentan un método sencillo derivado del anterior que puede ayudar al revisor en la búsqueda de las posibles causas de la significación y que formaría parte de la técnica jerárquica de homogeneidad propuesta por los mismos autores.

Consiste en la fragmentación de X_D^2 en los diferentes sumandos que lo componen (con sus respectivos grados de libertad) y que hacen referencia a la uniformidad de los tamaños del efecto en los diferentes grupos que se están comparando. Si uno o varios grupos presentan altas contribuciones a la suma de cuadrados dentro de los grupos esa clase puede ser el origen de los problemas de heterogeneidad y debe estudiarse con detalle.

La heterogeneidad también se ha analizado bajo el prisma de la regresión múltiple ponderada (HEDGES y OLKIN, 1985; GREENLAND, 1987), donde se consideran las características extraídas de los estudios como variable explicativas del tamaño del efecto.

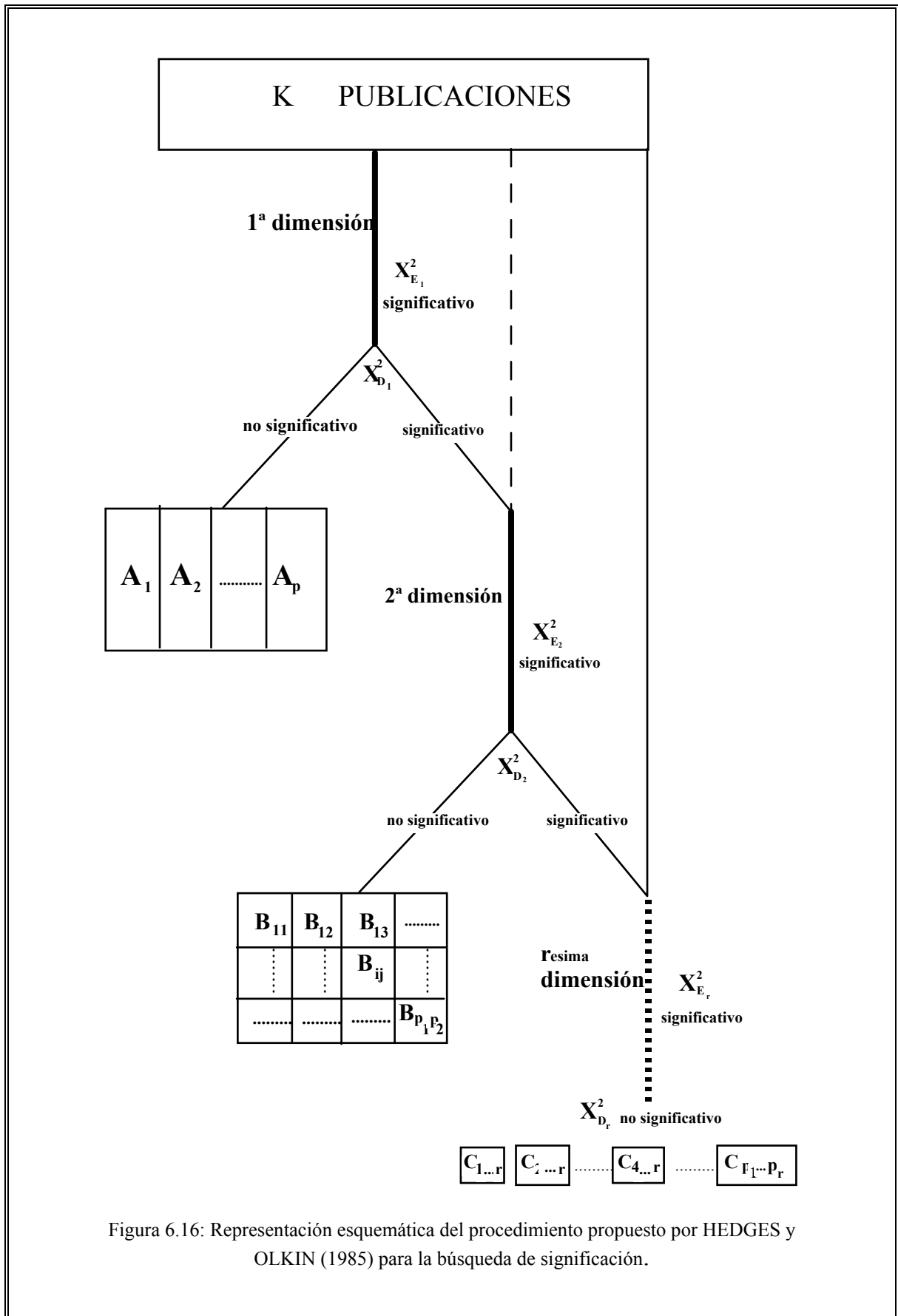


Figura 6.16: Representación esquemática del procedimiento propuesto por HEDGES y OLKIN (1985) para la búsqueda de significación.

ROSENTHAL y RUBIN (1982b) y HEDGES y OLKIN (1985) proponen la realización de contrastes ortogonales para la comparación de los diferentes tamaños del efecto, o grupos de ellos, en el mismo sentido que se utilizan para la comparación de los niveles de probabilidad. El contraste obtenido, cuando queremos comparar diferentes grupos de tamaños del efecto, es de la forma:

$$L = \frac{\sum_{i=1}^p \lambda_i \hat{\theta}_i}{\sqrt{\sum_{i=1}^p \lambda_i^2 \sigma_{\hat{\theta}_i}^2}} = \frac{\sum_{i=1}^p \lambda_i \hat{\theta}_i}{\sqrt{\sum_{i=1}^p \frac{\lambda_i^2}{w_i}}}$$

El estadígrafo L sigue una distribución Normal cuando los tamaños de muestra son grandes. Con este método se puede averiguar los grupos que difieren en la media del tamaño del efecto y estudiar las causas de esas posibles variaciones.

Los principales problemas derivados del método propuesto por HEDGES y OLKIN en el método de la búsqueda de la heterogeneidad son:

- El investigador debe conocer la importancia jerárquica de las variables exógenas influyentes en los tamaños del efecto. Es decir, el revisor debe plantearse a priori un orden de importancia para las variables que cree que están relacionadas con los tamaños del efecto.

Esta premisa raramente se cumple ya que el investigador generalmente conoce que variables pueden estar relacionadas, pero desconoce cual es el grado de importancia de las mismas.

- El investigador sólo puede reconocer los grupos que son diferentes a partir de la estructura categórica de la variable pero no identifica otras posibles agrupaciones entre las categorías definidas. Si la variable que interviene en la primera partición presenta cuatro categorías, por lo tanto cuatro grupos, puede ocurrir que los grupos uno y dos sean homogéneos pero que difieran de los grupos 3 y 4 que a la vez son homogéneos entre ellos mismos.
- El método no es un procedimiento jerárquico real ya que las particiones se realizan siempre de la población de tamaños del efecto accesible. Además este método no considera las posibles interacciones.

6.7.UNA CONTRIBUCIÓN A LA BUSQUEDA DE LA HETEROGENEIDAD EN LA FASE DE ANÁLISIS DE RESULTADOS DESDE UNA PERSPECTIVA MULTIVARIANTE.

6.7.1. METODO JERARQUICO PARA LA BUSQUEDA DE LA HETEROGENEIDAD.

6.7.1.i. PLANTEAMIENTO TEORICO

Se realiza una modificación del método con la intención de solucionar los problemas planteados, si bien el objetivo a conseguir es el propuesto por HEDGES y OLKIN (1985). La técnica se basa en los métodos **AID** (Automatic Interaction Detection) propuesto por MORGAN y SONQUIST (1963).

Se parte de una población accesible de tamaños del efecto, **M**, y un conjunto de variables externas $X_1, \dots, X_i, \dots, X_r$ que describen a los distintos estudios.

Las variables cualitativas presentan un número de categorías definidas como:

$$C_i = \{C_1^i, C_2^i, \dots, C_{c_i}^i\}$$

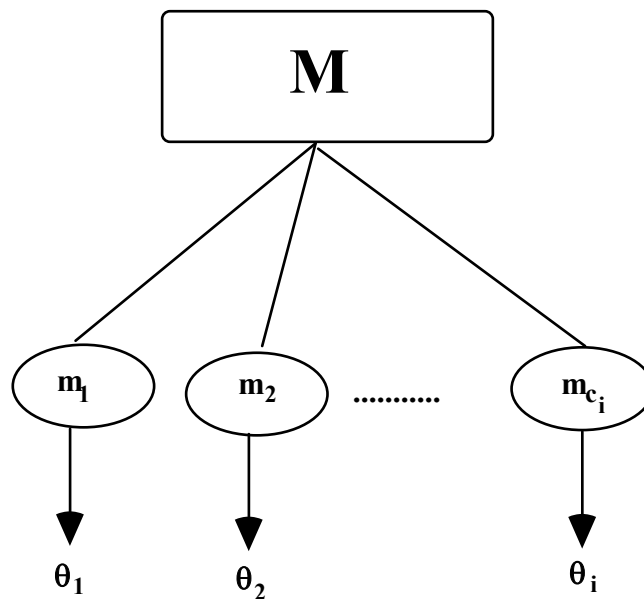
siendo C_1^i la primera categoría de la variable X_i y c_i el número de categorías

de X_i .

La finalidad del proceso es encontrar una partición m_1, m_2, \dots, m_q de M , definida a través de las variables $X_1, \dots, X_i, \dots, X_r$, que cumpla las siguientes dos condiciones:

- Dos grupos m_i y m_j sean lo más diferentes en relación a los tamaños del efecto.
- Cada parte m_i sea lo más homogénea posible con respecto a los tamaños del efecto.

Esquemáticamente:



Siguiendo las ideas de HEDGES y OLKIN (1985) definiremos una distancia total dentro de la población accesible como medida de homogeneidad global; una distancia intragrupo dentro de una subpoblación, la cual utilizaremos como una medida de homogeneidad en cada grupo y una distancia intergrupo la cual nos aporta la medida diferencial entre las distintas subpoblaciones. Por lo tanto,

la distancia intragrupo de la población \mathbf{M} , que es un valor fijo, se puede descomponer como la suma de la distancia intragrupo y la distancia intergrupo. Maximizar la distancia entre grupos es equivalente a minimizar la distancia intragrupo de modo que se cumplan las condiciones definidas anteriormente.

$$\underbrace{\sum_{i=1}^p \sum_{j=1}^{k_i} (\hat{\theta}_{ij} - \hat{\theta})^2 w_{ij}}_{\text{Total}} = \underbrace{\sum_{i=1}^p \sum_{j=1}^{k_i} (\hat{\theta}_{ij} - \hat{\theta}_i)^2 w_{ij}}_{\text{Dentro}} + \underbrace{\sum_{i=1}^p (\hat{\theta}_i - \hat{\theta})^2 w_i}_{\text{Entre}}$$

Al igual que en el método anterior las diferentes distancias están relacionadas con la distribución Ji-cuadrado.

El procedimiento a seguir para obtener los distintos grupos excluyentes y exhaustivos es de tipo secuencial.

Una vez que se ha rechazado la hipótesis de homogeneidad entre los tamaños del efecto se realizan las diferentes descomposiciones en suma de cuadrados para cada una de las variables $\mathbf{X}_1, \dots, \mathbf{X}_i, \dots, \mathbf{X}_r$.

En el caso que las variables sean continuas se puede contrastar mediante una regresión ponderada. De todas las posibles particiones óptimas elegiremos la que presenta mayor distancia entre grupos, \mathbf{X}_i , lo cual produce una partición de la población \mathbf{M} en p_i subpoblaciones $\mathbf{m}_1^i, \mathbf{m}_2^i, \dots, \mathbf{m}_{p_i}^i$ definidas por las categorías $\mathbf{C}_1^i, \mathbf{C}_2^i, \dots, \mathbf{C}_{p_i}^i$ de dicha variable.

En un segundo paso se pretende detectar dentro de esta primera partición que grupos difieren y cuales son homogéneos para constituir una nueva agrupación definitiva en q_i grupos ($q_i \leq p_i$). Esta nueva división representa la mejor agrupación para la variable elegida.

En dicho análisis se debe tener en cuenta la naturaleza de las categorías de cada una de las variables externas.

Así, en el contexto meta-analítico, se pueden definir dos patrones generales:

- **Variables monótonas**: aquellas variables cuyas categorías representan una escala ordinal. Esto implica que las agrupaciones se deben realizar de forma contigua. El número de pares posibles que se pueden obtener bajo el supuesto de monotonía con las p_i categorías de una variable es de $p_i - 1$
- **Variables libres**: aquellas variables cuyas categorías pertenecen a una escala nominal lo cual permite la agrupación libre de las categorías. El conjunto de pares posibles bajo esta presunción es de $\binom{p_i}{2}$.

Una vez conocido el tipo de variable se realizan las agrupaciones por pares para contrastar la homogeneidad de dichos grupos. El contraste utilizado vuelve a ser la descomposición de la suma de cuadrados de los distintos grupos formados mediante las agrupaciones monótonas o libres. Partimos de la primera partición $m_1^i, m_2^i, \dots, m_{p_i}^i$:

- **Agrupación monótona.** (figura 6.17) Se realiza la primera agrupación entre las subpoblaciones m_1 y m_2 . A continuación se elabora el contraste correspondiente:

$$X_{T_{1 \leftrightarrow 2}}^2 = X_{E_{1 \leftrightarrow 2}}^2 + X_{D_{1 \leftrightarrow 2}}^2$$

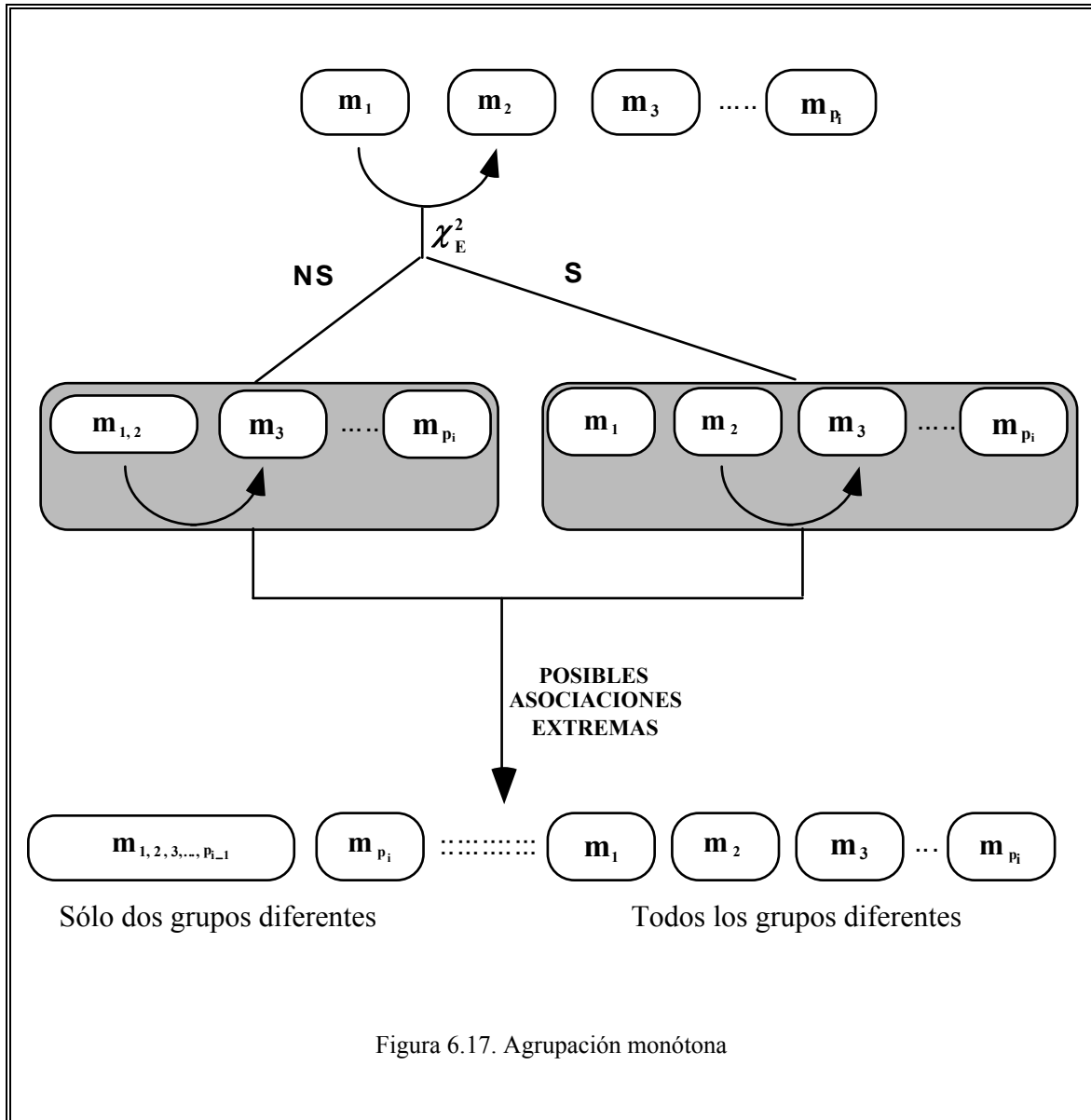


Figura 6.17. Agrupación monótona

1.- Si la suma de cuadrados entre grupos toma valores altos o es significativo, y por el contrario la distancia intragrupos es pequeña o no significativa, entonces, las subpoblaciones se mantienen separadas y continua el proceso con la siguiente comparación $\mathbf{m}_2 \times \mathbf{m}_3$:

$$\mathbf{X}_{T_{2 \leftrightarrow 3}}^2 = \mathbf{X}_{E_{2 \leftrightarrow 3}}^2 + \mathbf{X}_{D_{2 \leftrightarrow 3}}^2$$

2.- Si la suma de cuadrados entre grupos presenta valores no significativos frente a la suma de cuadrados dentro de los grupos que arroja resultados significativos, entonces se crea una nueva subpoblación, $\mathbf{m}_{1,2}$, formada por ambos grupos. El siguiente paso es contrastar la nueva asociación $\mathbf{m}_{1,2} \times \mathbf{m}_3$:

$$\mathbf{X}_{T_{1,2 \leftrightarrow 3}}^2 = \mathbf{X}_{E_{1,2 \leftrightarrow 3}}^2 + \mathbf{X}_{D_{1,2 \leftrightarrow 3}}^2$$

El proceso termina cuando se hacen todas las comparaciones factibles obteniendo las agrupaciones que presentan diferencias.

- **Agrupación libre.** El procedimiento se configura de forma similar con la única diferencia del número de agrupaciones a realizar debido al carácter nominal de la variable.

El problema que se puede presentar en este tipo de partición son las posibles incongruencias debidas a la formación de clases no excluyentes.

Para evitar este inconveniente el proceso de agrupación puede ser jerárquico; en una primera fase se realizan todos los posibles análisis de homogeneidad por parejas. Se consideran aquellos contrastes donde la distancias intergrupos no son significativas y se impone un criterio para realizar la primera agrupación.

Se pueden definir dos criterios los cuales se pueden utilizar indistintamente:

- 1.- Se elige el contraste de agrupación que presenta mayor valor de probabilidad.
- 2.- Se elige el valor más pequeño de X_E^2 no significativo después de ser divididos por los correspondientes grados de libertad para su posible comparación.

Una vez realizada la primera agrupación, en una segunda etapa, se repiten de nuevo los análisis de homogeneidad y el proceso es iterativo hasta conseguir que todos los contrastes sean significativos. (figura 6.18)

Una vez obtenida las agrupaciones de interés, tanto en el caso monótono como en el libre, comprobamos si estas nuevas subpoblaciones son homogéneas, o por el contrario alguna de ellas presentan heterogeneidad. En aquellos grupos donde se detecta heterogeneidad se repite el proceso con el resto de variables externas. De nuevo, aquella variable que presente mayor distancia intergrupar será la elegida para una segunda segmentación. El proceso se detendrá cuando los diferentes grupos obtenidos en las consiguientes particiones son todos homogéneos (Figura 6.19).

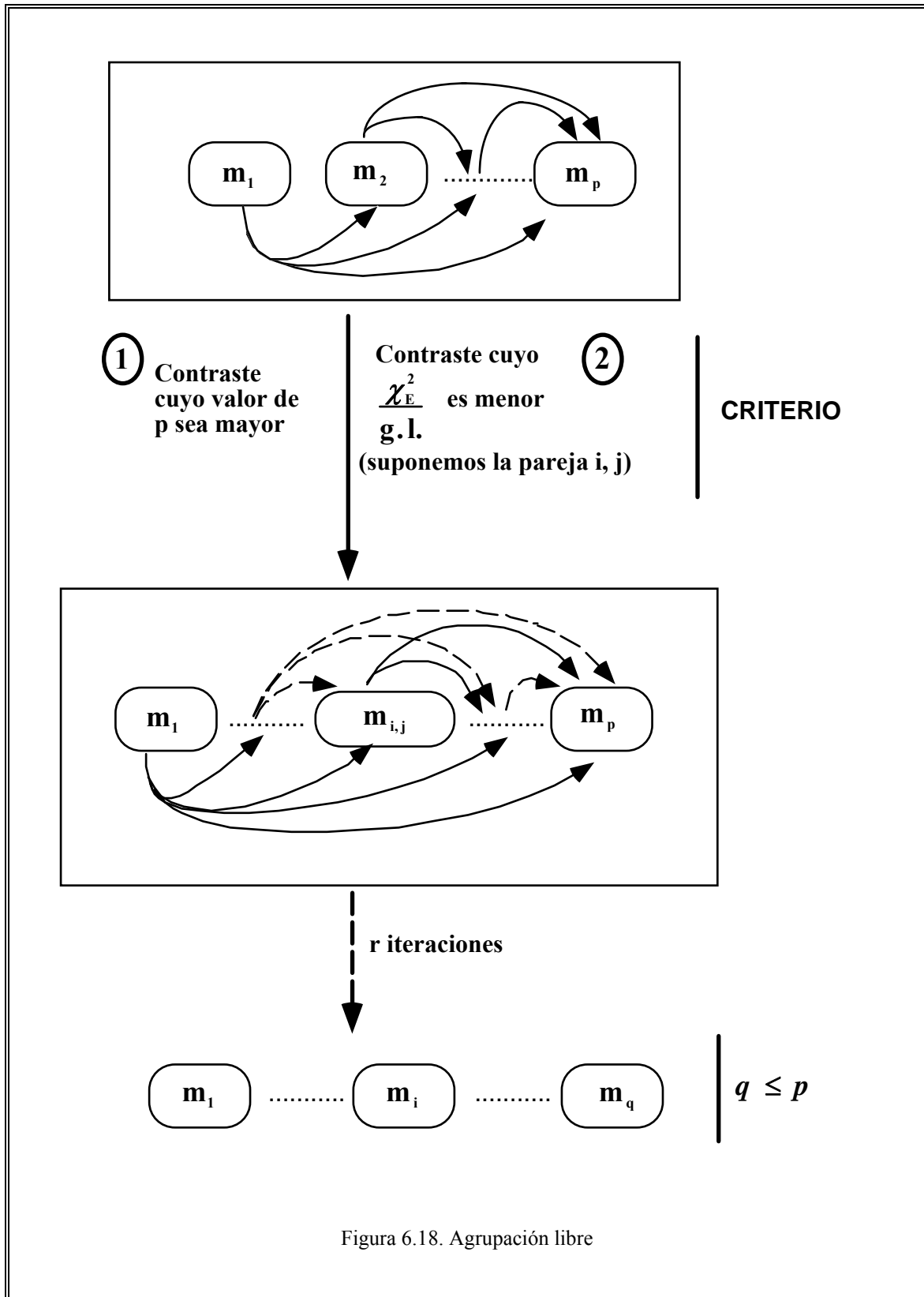
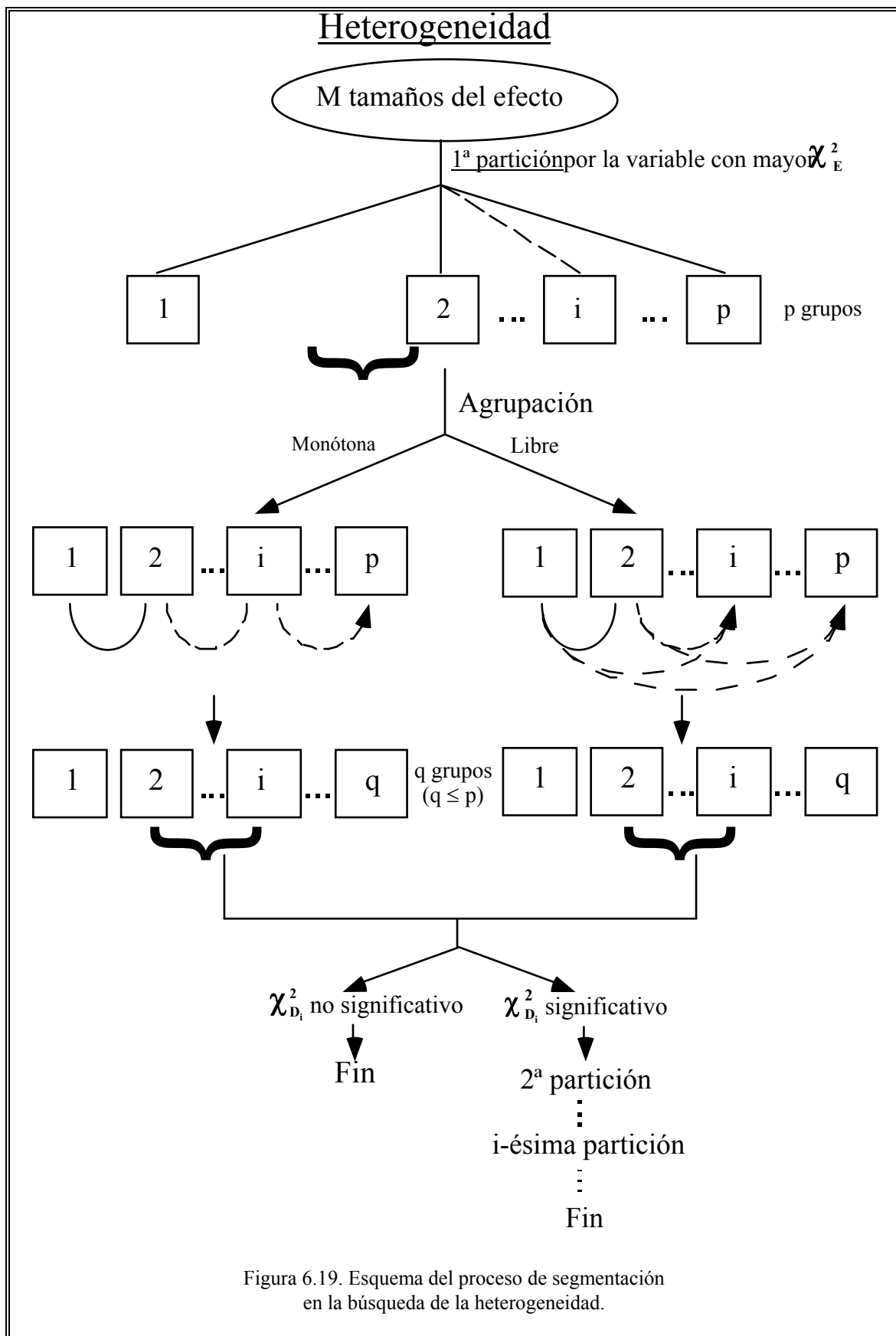


Figura 6.18. Agrupación libre



6.7.1.ii.ELECCIÓN DE LOS NIVELES DE PROBABILIDAD.

Los procedimientos basados en segmentaciones a partir de un conjunto de variables o predictores presenta el inconveniente del alto número de comparaciones que se realizan a lo largo del análisis. En esta circunstancia la probabilidad de cometer un *error de tipo I* a nivel global se incrementa. Dicha situación es bien conocida en el contexto de los contrastes tras el Anova. Una solución a este problema puede ser utilizar los métodos clásicos de corrección para el control del riesgo tipo I, es decir, la aproximación de Bonferroni con las diferentes modificaciones de dicho procedimiento para el incremento de la potencia en el ámbito de los contrastes múltiples (ver HOLM, 1979; SIMES, 1986a; HOLLAND y DIPONZIO, 1987; HOMMEL, 1989; DUNNET y TAMHANE, 1992; entre otros)

En el problema de los contrastes múltiples se parte de n hipótesis $H_{0_1}, \dots, H_{0_1}, \dots, H_{0_n}$ las cuales se contrastan mediante los estadísticos $T_1, \dots, T_i, \dots, T_n$.

Se define la hipótesis global

$$H_0 = \bigcap_{i=1}^n H_{0_i}$$

la cual es cierta si y sólo si todas las hipótesis son ciertas.

La hipótesis alternativa:

$$H_a = H_0^c = \left(\bigcap_{i=1}^n H_{0_i} \right)^c = \left(\bigcup_{i=1}^n H_{0_i}^c \right)$$

la cual es cierta si y sólo al menos una de las hipótesis individuales H_i es falsa.

Denotamos las regiones críticas para cada hipótesis por R_i , por lo tanto se rechaza cada H_i si y solo si $T_i \in R_i$.

Si centramos el problema en el contexto de la búsqueda de la heterogeneidad, entonces, la región crítica vendrá derivada de los contrastes unilaterales superiores:

$$R_i = \{T_i : T_i > C_i\}$$

Donde C_i es la constante crítica con la cual se compara el valor observado del estadístico T_i para decidir si se rechaza o no la hipótesis individual H_i . De forma equivalente en lugar de comparar los valores C_i y T_i se pueden comparar la probabilidades correspondientes, es decir, el p-valor p_i con el nivel de significación prefijado α'_i , tal que:

$$\Pr(T_i > C_i / H_i) = \alpha'_i$$

Si $T_i > C_i$, entonces:

$$\Pr(P_i \leq \alpha'_i / H_i) = \alpha'_i \quad (i=1, \dots, n)$$

Los contrastes son planteados de forma que la probabilidad de que se produzca al menos un error de tipo I se mantenga menor o igual a un nivel especificado α . Por lo tanto debe existir una relación entre el nivel global α y los niveles particulares α'_i .

La aproximación de Bonferroni se basa en la desigualdad probabilística de Boole, que afirma que para cualesquiera suceso $A_1, \dots, A_i, \dots, A_n$ se cumple que:

$$\Pr\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$$

Por lo tanto:

$$\begin{aligned} \Pr(\text{rechazar al menos una } H_i / H_i) &= \Pr\left\{\bigcup_{i=1}^n (T_i > C_i / H_i)\right\} = \\ &= \Pr\left\{\bigcup_{i=1}^n (P_i \leq \alpha'_i / H_i)\right\} \leq \sum_{i=1}^n \Pr(P_i \leq \alpha'_i / H_{0_i}) = \sum_{i=1}^n \alpha'_i \end{aligned}$$

El procedimiento de Bonferroni compara los p-valores, p_j , con los niveles críticos $\alpha'_i = \alpha/n$, donde n es el número de contrastes. Por lo tanto con la utilidad de la desigualdad de Boole permite ver que la probabilidad de rechazar al menos una hipótesis cierta es menor o igual a α :

$$\sum_{i=1}^n \alpha'_i = \sum_{i=1}^n \left(\frac{\alpha}{n}\right) = \alpha$$

La aplicación del método de Bonferroni para contrastes múltiples en el contexto de la búsqueda de la heterogeneidad presenta varios inconvenientes al igual que otros procedimientos descritos en la literatura como son los análisis de segmentación (KASS, 1980; RAMIREZ, 1995) o detección automática de la interacción (MORGAN y SONQUIST, 1963; WORSLEY, 1977; MESSENGER y MANDELL, 1972). :

- El número de contrastes no es fijo, depende del tipo de variable. Este valor depende de las categorías que presenta la variable o bien, del carácter ordinal o nominal de la misma.
- Hay contrastes que se repiten y por lo tanto la constante no es la suma de todos los contrastes sino menor.

Si el número de comparaciones es alto los niveles críticos particulares α'_i son muy pequeños y por consiguiente el criterio resulta exigente para rechazar la hipótesis, lo cual implica una potencia del contraste muy baja.

RAMIREZ (1995) presenta, en el contexto de la segmentación, una simulación para observar la influencia del número de comparaciones en el incremento del riesgo tipo I. Además, determina como se corrigen el incremento del error mediante los métodos de Bonferroni clásico y una modificación propuesta por el propio autor. En dicha simulación se observa como se produce el aumento de la probabilidad de cometer un error tipo I según se incrementan las categorías de las variable o se pasa de la asociación monótona a la libre o sin restricciones. Así, en un predictor monótono de seis categorías y una variable dependiente de cuatro, bajo el supuesto de independencia, se pasa de esperar obtener 5 asociaciones por cada 100 pruebas (nivel de significación 0,05) a conseguir 20 asociaciones de cada 100 pruebas. En el caso de un predictor sin restricciones la situación se agrava incurriendo en 47 asociaciones.

RAMIREZ destaca el hecho de que el procedimiento clásico de Bonferroni no resulta tan conservador en este contexto y sólo cuando los predictores presentan un número elevado de categorías, esta clase de penalización se convierte en conservadora sobre todo en la opción libre y es preferible utilizar la modificación de dicho método.

La situación es muy similar en la técnica de la búsqueda de la heterogeneidad en el ámbito meta-analítico. El número de contrastes puede ser grande y por lo tanto la potencia ser baja. La propuesta en la penalización es la de utilizar una penalización clásica de Bonferroni en cada una de las particiones. Así, en la cada división se penalizará por el número de comparaciones a realizar en dicha etapa.

Obviamente, la potencia global depende del número de categorías de las diferentes variables. Esto aconseja al revisor a estudiar minuciosamente el proceso de categorización. Si un estudio presenta un número elevado de variables con excesivas categorías las penalizaciones serán elevadas con lo cual la partición en grupos será excesiva, con pocos estudios en cada subpoblación y por lo tanto una situación similar a la de considerar cada estudio como independiente.

Otro problema importante es la inclusión de un gran número de variables explicativas de la posible heterogeneidad lo cual podría provocar divisiones sin sentido (GELBER y GOLDBIRSCH, 1991; JONES, 1992). Por ello al igual que en el problema anterior el investigador debe prestar especial interés para seleccionar las variables y intentar que el número no sea muy elevado.

Otra solución al problema para evitar la intervención del alto número de variables en el método, consiste en un previo estudio multivariante donde se ponga de manifiesto que variables están más relacionadas con los efectos del tamaño.

6.7.1.iii. MODELO VECTORIAL PARA RELACIONAR LA ORDENACION DE LOS ESTUDIOS

Supóngase que se ha obtenido una ordenación de los estudios contenida en una matriz $\mathbf{X}_{n \times r}$, a partir de la información obtenida de cada uno de ellos en la fase de codificación y que se traduce en un conjunto de características de cada uno de los estudios. Dicha ordenación puede haberse obtenido mediante cualquiera de las técnicas de ordenación habitualmente utilizadas (Análisis de Componentes Principales, Biplot, Análisis de Correspondencias, Multidimensional Scaling, etc...) y que se describieron en conjunto en el

apartado 6.3.1.i de este capítulo.

La ordenación "ordena" los estudios de acuerdo con las características más variables en el estudios pero éstas, en general, no tienen por qué ser las más relacionadas con el tamaño del efecto que se pretende estudiar.

Supongamos que disponemos también de una matriz $\mathbf{Z}_{n \times q}$, que contiene los tamaños del efecto de los diferentes estudios y sea $\mathbf{W}_{n \times q}$, la matriz que contiene los inversos las varianzas estimadas de los tamaños del efecto que en este caso van a ser consideradas como ponderaciones.

Generalmente $q=1$ aunque no se excluye la posibilidad de que de cada uno de los estudios se obtengan varios tamaños del efecto como apuntan HEDGES y OLKIN (1985, Cap. 10).

Tratamos de buscar aquellas direcciones del espacio de la ordenación de los estudios que tienen correlación máxima con los tamaños del efecto.

Esto es equivalente a situar sobre el gráfico q vectores (r -dimensionales) \mathbf{b}_j ($k=1, \dots, q$), uno para cada uno de los tamaños del efecto, de forma que el producto escalar de una fila de \mathbf{X} , \mathbf{x}_i , ($i=1, \dots, n$), por dicho vector $\mathbf{x}_i' \mathbf{b}_j$, aproxime, tan adecuadamente como sea posible, el elemento z_{ij} de la matriz \mathbf{Z} . Si tomamos los vectores \mathbf{b}_j como filas de una matriz $\mathbf{B}_{q \times r}$, la condición anterior puede escribirse como

$$\mathbf{X} \mathbf{B}' \approx \mathbf{Z} \quad (2)$$

donde el símbolo \approx significa aproximación mínimo cuadrática ponderada. Se trata de encontrar \mathbf{B} que minimice:

$$\mathbf{L} = \sum_{j=1}^q \sum_{i=1}^n w_{ij} (z_{ij} - \mathbf{x}_i' \mathbf{b}_j) \quad (3)$$

que puede separarse en q problemas independientes

$$L_j = \sum_{i=1}^n w_{ij} (z_{ij} - x_i' b_j) \quad (4)$$

ya que los vectores que buscamos dependen solamente de j .

El vector b_j que hace mínimo L_j puede obtenerse por regresión mínimo cuadrática ponderada como

$$b_j = (\mathbf{X}' \mathbf{D}_{\langle w_j \rangle} \mathbf{X})^{-1} \mathbf{X}' \mathbf{D}_{\langle w_j \rangle} z_j$$

donde $\mathbf{D}_{\langle w_j \rangle}$ es una matriz diagonal cuyos elementos son los de la j -ésima columna de \mathbf{W} y z_j es la j -ésima columna de \mathbf{Z} . Combinando los vectores b_j como filas de una matriz se obtiene \mathbf{B} .

Las coordenadas que buscamos pueden considerarse como coeficientes de regresión de cada una de las columnas de \mathbf{Z} sobre los ejes de la ordenación contenidos en \mathbf{X} utilizando las ponderaciones contenidas en \mathbf{W} . De acuerdo con esta definición los vectores contenidos en \mathbf{B} son las direcciones en el espacio generado por las columnas de \mathbf{X} que mejor explican el comportamiento de los tamaños del efecto.

La bondad del ajuste de la aproximación puede medirse para cada tamaño del efecto como:

$$R_j^2 = \frac{\sum_{i=1}^n w_{ij} (z_{ij} - x_i' b_j)^2}{\sum_{i=1}^n w_{ij} z_{ij}^2}$$

y la bondad del ajuste global como:

$$R^2 = \frac{\sum_{j=1}^q \sum_{i=1}^n w_{ij} (z_{ij} - x_i' b_j)^2}{\sum_{j=1}^q \sum_{i=1}^n w_{ij} z_{ij}^2}$$

Las medidas se han denominado R_j^2 porque son los coeficientes de determinación de las regresiones.

El coeficiente de determinación para cada una de las regresiones puede considerarse como una medida de la bondad del ajuste equivalente a lo que denominamos "*Contribución relativa del factor al elemento*" (ver GALINDO y CUADRAS, 1986) en los análisis internos en los que se considera un único conjunto de variables; para demostrar esta afirmación basta tener en cuenta que dicha contribución se puede interpretar como la parte de la variabilidad de la variable que viene explicada por la variabilidad del factor.

Estas contribuciones del factor a explicar la variabilidad del tamaño del efecto permiten identificar cuales son los ejes de ordenación más relacionados con el mismo y utilizar, en pasos posteriores, solamente las variables características de ese eje.

Si se dispone de un solo tamaño del efecto convendría hacer coincidir el primer eje de ordenación con la dirección del espacio más relacionada con el tamaño del efecto que es el objetivo del estudio; para ello basta calcular el ángulo entre los dos y rotar los ejes del espacio original.

Llamando \mathbf{b} al vector que contiene la dirección de mejor ajuste al tamaño del efecto y $\mathbf{e}_1=(1, 0, \dots, 0)$ al vector unitario en la dirección del primer eje, el coseno del ángulo que forman es

$$\cos \alpha = \frac{\mathbf{b}' \mathbf{e}_1}{\|\mathbf{b}\|}$$

y rotar la matriz de coordenadas \mathbf{X} con este ángulo.

Las calidades de representación y las contribuciones del factor al elemento se calculan de nuevo tal y como se explicó en el apartado correspondiente.

6.8. COMPARACION DE LAS TECNICAS DE HEDGES Y OLKIN PARA LA BUSQUEDA DE LA HETEROGENEIDAD, Y DE LA ALTERNATIVA MULTIVARIANTE PROPUESTA

6.8.1 INTRODUCCIÓN

Para ilustrar el procedimiento de partición propuesto, se han tomado los datos de VANDERWERF (1992) del que se pudieron extraer los tamaños del efecto y varias variables explicativas que utilizaremos para el proceso.

En este artículo se recoge la información de 42 estudios sobre los tamaño de puesta óptimos en pájaros. En todos ellos se contrasta la hipótesis de que parejas con un aumento en el número de huevos producen mayores porcentajes de pollos volanderos que las parejas con una nidada natural. Los resultados arrojados por los estudios de manipulación experimental en el incremento de huevos ha provocado resultados contradictorios a favor y en contra de dicha hipótesis. El autor obtiene un conjunto de variables que considera importantes a la hora de explicar la variabilidad entre los tamaños del efecto. Estas variables fueron:

- **Fecha del estudio: año.** (antes del 70, entre 71 y 80, después del 80)
- **Longitud del estudio en número de estaciones.** (continua)
- **Tamaño muestral del grupo experimental.** (continua)
- **Latitud del estudio** (Grados continua, Norte-Sur)
- **Desarrollo de los jóvenes en la especie estudiada** (precocial=1, semiprecocial=2, semialtricial=3 y altricial=4).
- **Supervivencia anual de los adultos** (continua).

- **Validez del estudio.** (0 y 1)

Se elige como tamaño del efecto a la diferencia de medias estandarizada donde el estimador de la varianza, en general, es la media ponderada de las varianzas de los grupos control y experimental, si bien en cinco estudios el autor detectó heterocedasticidad entre dichas varianzas, por lo cual seleccionó como estimador la varianza del grupo control como propone GLASS y cols. (1981).

Lamentablemente, el autor utiliza como estimador del efecto global la media sin ponderar y no proporciona la información suficiente para poder calcular dichas ponderaciones por lo que nos vimos obligados a desarrollar el proceso de la búsqueda de la heterogeneidad sin la introducción de los pertinentes ponderaciones.

Debido a que los resultados son extraídos de un meta-análisis ya realizado, las limitaciones de información nos obligan a adaptar el método a dichas limitaciones, si bien el propósito es ilustrar con datos reales las diferentes etapas y análisis que conlleva el método propuesto. Por lo tanto, los pasos generales llevados a cabo se pueden resumir:

- En cada paso de la partición se comprueba la homogeneidad del grupo que se quiere estudiar y si el grupo resulta no homogéneo se buscan aquellas variables que producen mayor heterogeneidad en el tamaño del efecto. Para simplificar el procedimiento se ha considerado que todas las variables cualitativas tienen un número fijo de categorías y no se han agrupado las mismas en cada paso del proceso.
- Para la variables cuantitativas se ha comprobado si existe relación con los

tamaños del efecto mediante regresión y si se encontraba una relación significativa, se dividía la variable en varios grupos para continuar con las particiones. El proceso puede realizarse a priori dividiendo los valores de la variable en varias categorías para después realizar los contrastes de homogeneidad.

6.8.2. APLICACION DEL METODO DE BUSQUEDA DE LA HETEROGENEIDAD

Se realiza primero el contraste de homogeneidad para conocer si los tamaños del efecto estiman un único parámetro o por el contrario existen subpoblaciones. El contraste de homogeneidad global proporciona un valor experimental de **82,5567** para una ji-cuadrado con **41** grados de libertad, $p < 0.0001$ y por tanto es un contraste altamente significativo que implica que los tamaños del efecto no son homogéneos y no se pueden integrar. Pasamos a buscar las variables causantes de esta heterogeneidad:

POBLACION TOTAL DE ESTUDIOS

1.- Relación entre el tamaño del efecto y las variables independientes.

En las tablas 6.3 y 6.4 aparecen los resultados correspondientes a los diferentes contraste para conocer que variable (continua o categórica) es la causante de la primera partición. En la variable longitud del estudio, las categorías 3 y 4 fueron agrupadas ya que en esta última estaba formada por un sólo estudio. La variable latitud se considero como categórica con dos categorías (Norte y Sur)

VARIABLES CONTINUAS:

VARIABLES	R²	p-valor
Tamaño	0.02	0.3679
Latitud	0.001	0.8415
Supervivencia	0.0629	0.1092

Tabla 6.3. Tabla de resultados para las variables continuas.

VARIABLES CUALITATIVAS:

VARIABLE	Entre	g.l.	p-valor	Dentro	g.l.	p-valor
Longitud1	10.888	2	0.0043	71.679	39	0.0011
Latitud 2	0.079	1	0.7787	82.478	40	<0.0001
Desarrollo	20.060	3	0.0002	62.496	38	0.0074
Validez	0.626	1	0.4288	81.931	40	<0.0001
Año	3.066	2	0.2159	79.490	39	0.0015

Tabla 6.4: Tabla de resultados para las variables categóricas

La variable desarrollo es la que produce mayor heterogeneidad en los datos (con un p-valor menor, o relación Entre/Dentro mayor), por lo que procedemos a separar el conjunto de estudios de acuerdo con la misma. Así, en una primera partición hemos obtenido cuatro subpoblaciones definidas por el desarrollo de los pollos: precocial, semiprecocial, semialtricial y altricial

2.-Estudio de la homogeneidad de los grupos resultantes de la división.

En la tabla 6.5 aparecen los diferentes contrastes realizados para conocer si estos grupos son homogéneos y por lo tanto grupos terminales o por el contrario alguno de ellos presentan heterogeneidad.

Se puede observar que el único grupo heterogéneo el definido por un desarrollo altricial, el resto se conforman como grupos terminales y en cada uno de ellos se estima el tamaño del efecto.

GRUPO				
Desarrollo	Test	g.l.	p-valor	Estado
Precocial				Un solo estudio (Terminal)
Semiprecocial	9.310	9	0.4092	Homogéneo (Terminal)
Semialtricial	9.532	7	0.2167	Homogéneo (Terminal)
Altricial	43.834	22	0.0037	No homogéneo (No terminal)

Tabla 6.5. Resultados de los contrastes de homogeneidad para los grupos resultantes de la primera partición.

Una vez descubierto el grupo heterogéneo el proceso se repite de igual manera para el grupo seleccionado.

SUBPOBLACION DESARROLLO ALTRICIAL**1.- Relación entre el tamaño del efecto y las variables independientes.**

Variables continuas (regresión)

VARIABLE	R²	p-valor
Tamaño	0.018	0.5466
Latitud	0	0.8044
Supervivencia	0.034	0.3973

Tabla 6.6. Tabla de resultados para las variables continuas en el grupo desarrollo altricial.

Variables cualitativas

VARIABLE	Entre	g.l.	p-valor	Dentro	g.l.	p-valor
Longitud	4.316	2	0.1156	39.518	20	0.0057
Latitud	0.128	1	0.7205	43.707	21	0.0026
Validez	0.805	1	0.3696	43.029	21	0.0031
Año	0.583	2	0.2529	43..251	21	0.0029

Tabla 6.7: Tabla de resultados para las variables categóricas en grupo desarrollo altricial

Aunque ninguna de las variables es significativa y debido a que el grupo es no homogéneo, vamos a dividir de acuerdo con la longitud que es la que produce mayor heterogeneidad.

2.-Estudio de la homogeneidad de los grupos resultantes de la división.

GRUPO					
Desarrollo	Longitud	Test	g.l.	p-valor	Estado
Altricial	1	32.124	12	0.0013	No homogéneo (No Terminal)
Altricial	2	3.726	3	0.2926	Homogéneo (Terminal)
Altricial	3	3.668	5	0.5981	Homogéneo (Terminal)

Tabla 6.8. Resultados de los contrastes de homogeneidad para los grupos resultantes de la segunda partición.

Podemos seguir dividiendo el grupo Desarrollo Altricial con una longitud de estudio de una estación.

SUBPOBLACION DESARROLLO ALTRICIAL
CON LONGITUD UNO

1.- Relación entre el tamaño del efecto y las variables independientes.

VARIABLES CONTINUAS (REGRESIÓN)

VARIABLE	R²	p-valor
Tamaño	0,52	0,005
Latitud	0,175	0,1550
Supervivencia	0	0,9003

Tabla 6.9. Tabla de resultados para las variables continuas en el grupo

desarrollo altricial con longitud uno.

Variables cualitativas:

VARIABLE	Entre	g.l.	p-valor	Dentro	g.l.	p-valor
Latitud	1.618	1	0.2034	30.506	11	0.0013
Validez	0.982	1	0.3217	31.142	11	0.0010
Año	1.452	2	0.4838	30.672	10	0.0007

Tabla 6.10: Tabla de resultados para las variables categóricas en grupo desarrollo altricial con longitud uno

La variable que produce mayor heterogeneidad en los datos es el tamaño muestral. Para hacer las particiones dividimos la variable en dos grupos (≤ 10 y > 10) y contrastamos la homogeneidad entre grupos.

2.-Estudio de la homogeneidad de los grupos resultantes de la división.

VARIABLE	Entre	g.l.	p-valor	Dentro	g.l.	p-valor
Tamaño	12.783	1	0.003	19.341	11	0.0552

Tabla 6.11: Tabla de resultados

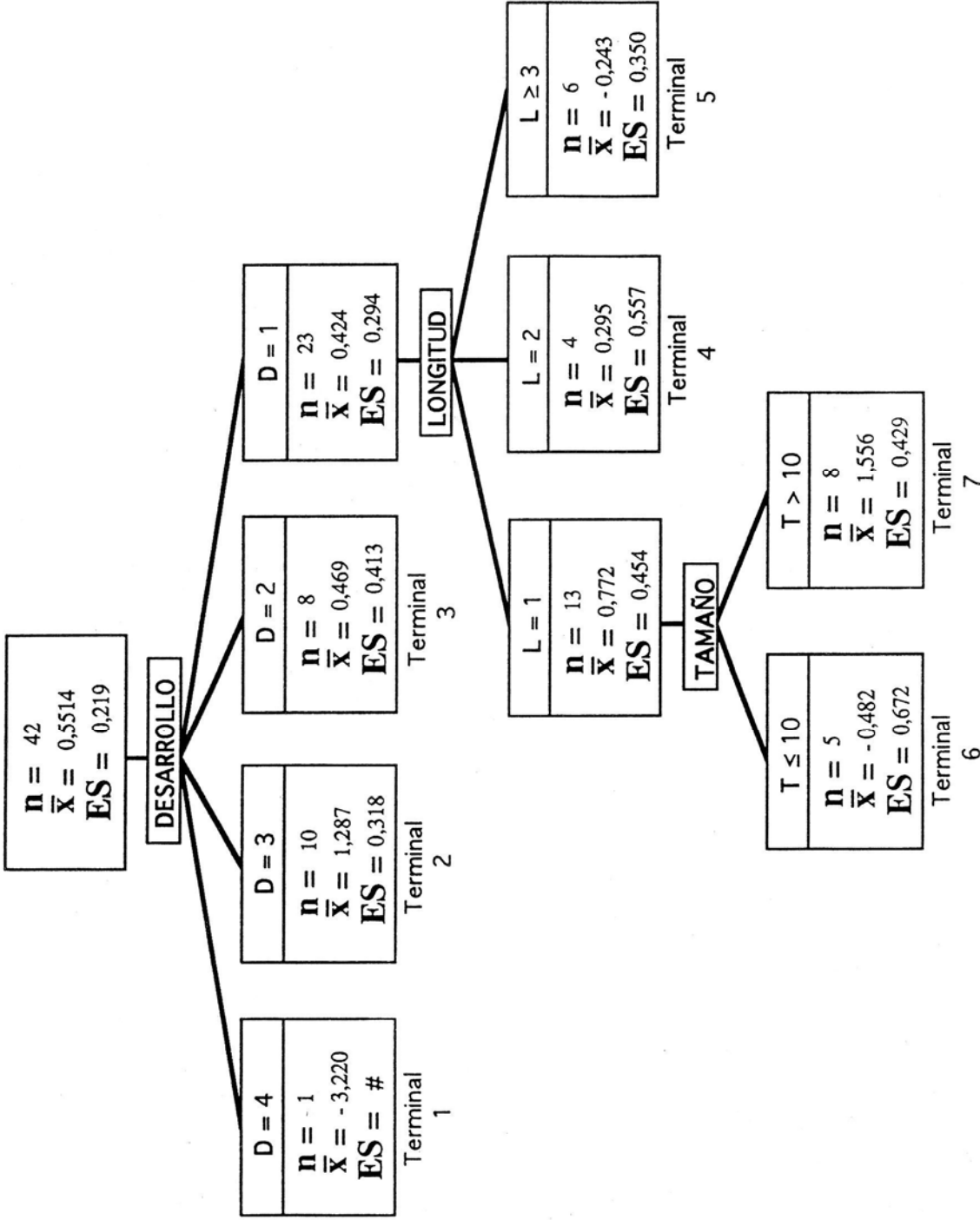
Efectivamente los grupos no son homogéneos entre si. Aunque la suma de cuadrados dentro de los grupos es no significativa realizamos los contrastes de homogeneidad para cada grupo por separado.

GRUPO						
Desarrollo	Longitud	Tamaño	Test	g.l.	p-valor	Estado
Altricial	1	≤ 10	9.02	4	0.0606	Homogéneo Terminal
Altricial	1	> 10	10.321	7	0.1711	Homogéneo Terminal

Tabla 6.12: Tabla de resultados

Por tanto se reconocen siete grupos terminales en los cuales podemos estimar el tamaño del efecto, puesto que son homogéneos. Obviamente, sólo pretendemos ilustrar el procedimiento propuesto para la búsqueda de heterogeneidad y por lo tanto no intentamos extraer conclusiones biológicas definitivas, ya que la información que disponemos sobre los estudios es limitada.

En el esquema siguiente se resumen los grupos terminales donde se refleja el valor del estimador del efecto, su varianza y el número de observaciones en cada grupo:



Los resultados obtenidos son concordantes con las conclusiones del trabajo; si bien, el investigador sólo habla en términos generales de una dependencia de los tamaños del efecto con las variables *desarrollo* y *longitud*, y no profundiza en las causas de esa dependencia.

Por último, se ha realizado un modelo de análisis factorial para comprobar si efectivamente podemos extraer de este análisis las variables mas importantes para efectuar posteriormente el análisis.

Se introduce el vector de tamaños del efecto en la ordenación proporcionada por el HJ-biplot en el cual los tres primeros ejes absorben un 83% de inercia. Se comprueba que el tamaño del efecto está relacionado con el tercer eje, el cual viene definido por las variables *desarrollo* y *longitud*. En la figura 6.20 se representa el plano 2-3 donde aparece con mayor claridad la relación entre ambas variables y el tamaño del efecto.

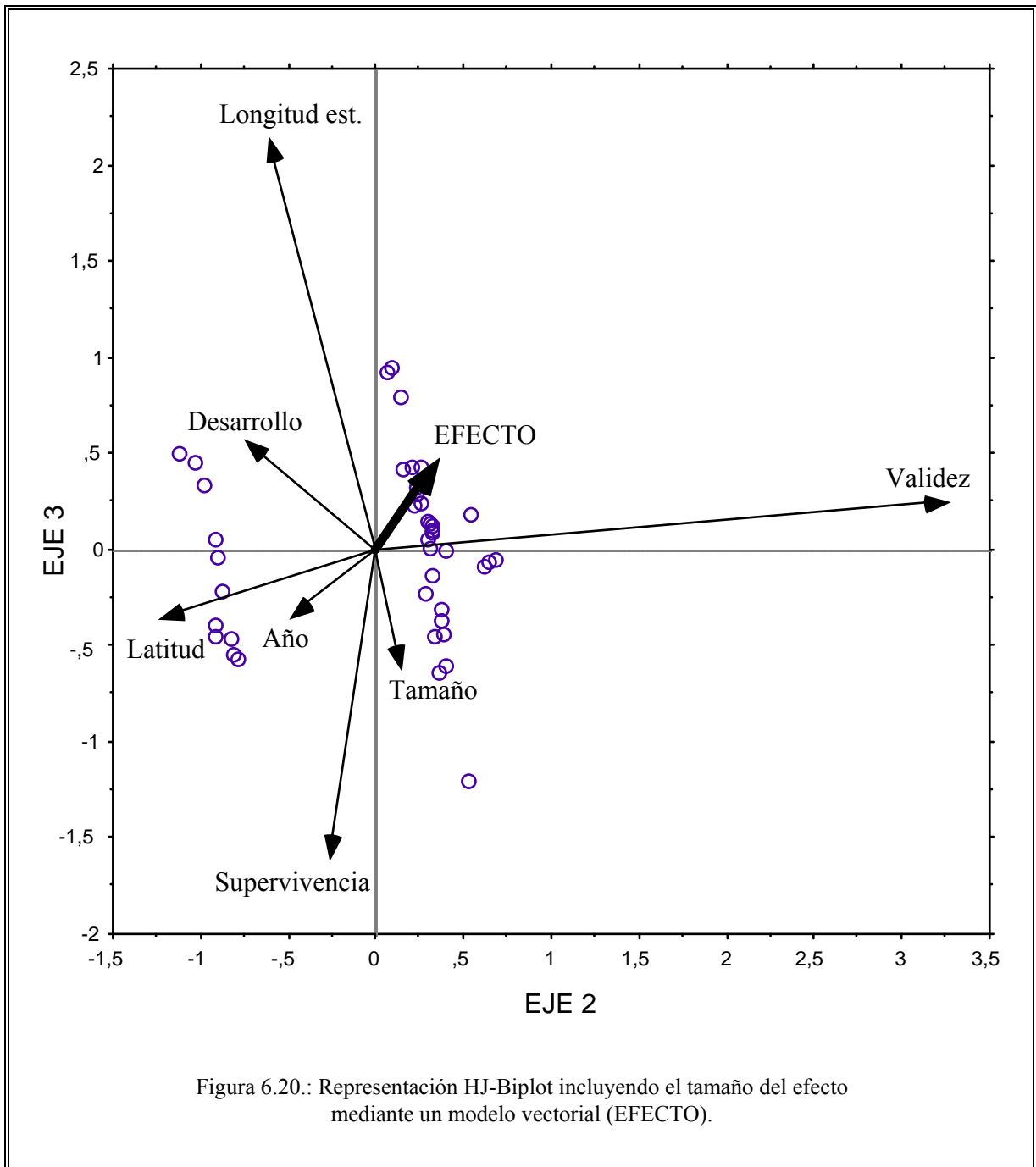


Figura 6.20.: Representación HJ-Biplot incluyendo el tamaño del efecto mediante un modelo vectorial (EFECTO).

CAPITULO 7

PROBLEMAS DEL META- ANALISIS

El meta-análisis ha planteado un debate en el foro científico entre los defensores y detractores de estas técnicas para la agrupación de información sobre un tópico de investigación (WACHTER, 1988; MANN, 1990; STRAUSS, 1992). Dada las características de este tipo de revisión cuantitativa son varios los inconvenientes que surgen debido a la fusión de los resultados de varios trabajos con rasgos diferenciales en el tiempo, en el espacio y en la forma. Entre las críticas que se han planteado en el meta-análisis, las más importantes hacen referencia al Sesgo de Publicación, la Heterogeneidad de los Resultados, e Independencia de los Tamaños del Efecto.

7.1. EL SESGO DE PUBLICACION.

BEGG & BERLIN (1988) definen el sesgo de publicación, como el error sistemático inducido en una inferencia estadística por el condicionamiento de alcanzar el estatus de publicación.

El sesgo de publicación se produce por el rechazo, tanto de los investigadores como de los editores, a publicar resultados que no presentan ninguna significación estadística, lo cual motiva que este tipo de resultados estén fuera del alcance del revisor y por lo tanto no formen parte de la denominada "población accesible".

La consecuencia de dicho sesgo se refleja en la sobrestimación del tamaño del efecto como resultado de una mayor presencia de efectos positivos en detrimento de los negativos.

En NAYLOR (1988), TAVEGGIA (1974) y CHALMERS y cols. (1987), se analiza el impacto del sesgo de publicación en los resultados del meta-análisis.

SIMES (1986b), EASTERBROOK y cols., (1991) y DICKERSIN y cols., (1992) estudian los factores que influyen en la publicación de los estudios y afirman que la naturaleza de los datos es influyente en la publicación de un estudio.

Las principales fuentes descritas que influyen en la publicación de un artículo son :

- Tamaños de muestra.
- Aleatorización de las muestras.
- Definición del protocolo.
- Orden de año de publicación.
- Estudios con contrastes múltiples.

Otra fuente importante de sesgo es la que se produce cuando los resultados de las publicaciones son defectuosos o faltantes. Es un hecho que muchas de las publicaciones sólo presentan los resultados comparando el p-valor experimental con el nivel de significación marcado, $p < 0,05$. Dichos estudios carecen de información fundamental para calcular el tamaño del efecto. En este caso, el revisor debe intentar ponerse en contacto con los autores para conseguir la información no extraíble del estudio. Evidentemente ésta no es una tarea fácil, ni hay garantía de éxito.

ROSENBERGER (1990) presenta una simple lista de recomendaciones para cumplir los mínimos requerimientos que deben aparecer en los estudios de investigación. En este sentido, SACKS y cols., (1987) evalúan la calidad de 86 meta-análisis publicados; es decir, llevó a cabo un meta-meta-análisis de 86 meta-análisis de ensayos clínicos controlados, aleatorizados, usando 23 ítems de control divididos en diferentes apartados (diseño del estudio, combinabilidad, control del sesgo, análisis estadístico, análisis de sensibilidad y aplicabilidad de los resultados) y encontraron que sólo el 28% de los estudios (es decir sólo 24 de los 86) contenían información sobre los 6 apartados. Los autores identifican grandes deficiencias en la metodología de esos meta-análisis y sugería la necesidad de mejorar la metodología meta-analítica.

Un importante aspecto de cualquier meta-análisis debería ser *el análisis de la sensibilidad* para determinar en que medida cambiarían los resultados cuando se introducen variaciones en la selección de los estudios, las características del diseño, etc. El análisis de sensibilidad también resulta útil para explicar los diferentes resultados de varias revisiones sobre un mismo tema (L'ABBE y cols., 1987).

7.1.1. SOLUCIONES PROPUESTAS

Hay varias propuestas que intentan solucionar este problema. Unas calificadas como *informales* son de tipo gráfico como la representación en embudo o piramidal (funnel display) descrita en el capítulo 6. Otras, que se describen a continuación y denominadas *formales*, abordan la cuestión desde un punto de vista cuantitativo.

7.1.1.i. METODO DEL "NUMERO DE FALLOS NULOS"

Esta técnica fue propuesta por ROSENTHAL (1979) y por ORWIN (1983); modificada por IYENGAR & GREENHOUSE (1988). ROSENTHAL (1979) pretende estimar el número de resultados no significativos (*fail-safe number*) que deberían encontrarse para alterar los resultados obtenidos en un meta-análisis. La solución no es definitiva pero puede utilizarse como una aproximación del posible sesgo que se ha producido en nuestra revisión.

Dado un grupo de k contrastes independientes de una misma hipótesis de los que se han obtenido p-valores, p_i , podemos obtener los valores z_i correspondientes a esas probabilidades y calcular la puntuación combinada z_T :

$$z_T = \sum_{i=1}^k \frac{z_i}{\sqrt{k}} \quad \text{Donde: } z_i = \phi^{-1}(p_i)$$

Se asume la existencia de k_0 estudios con resultados no significativos y que cumplen:

$$\sum_i \phi^{-1}(p_i) = 0$$

Entonces la combinación de los contrastes para los k estudios observados y k_0 no observados es:

$$z = \sum_{i=1}^k \frac{\phi^{-1}(p_i)}{\sqrt{k+k_0}} = \frac{\sqrt{k}z_T}{\sqrt{k+k_0}}$$

Fijando el valor crítico z_α al nivel de significación al α y despejando de la ecuación anterior k_0 , entonces:

$$k_0 = \frac{k(z_T^2 - z_\alpha^2)}{z_\alpha^2}$$

Es más común encontrar revisiones cuantitativas que acumulan tamaños del efecto experimental, en particular de la diferencia media tipificada d . En este sentido, ORWIN en 1983 introduce un estadístico aplicable al tamaño del efecto d , análogo al de Rosenthal, el cual estima el número de estudios no significativos necesarios para reducir el tamaño del efecto medio a un valor crítico más pequeño. La expresión es:

$$k_0 = \frac{k(d - d_c)}{d_c}$$

k es el número de estudios del meta-análisis, d la media del tamaño del efecto de los estudios integrados y d_c el valor crítico de los estudios seleccionados.

No hay consenso en relación a los valores críticos del tamaño del efecto; ORWIN sugiere que se adopten los valores propuestos por Cohen (COHEN, 1967), considerando efectos pequeños en torno a $d=0.20$, efectos medios en torno a $d=0.50$ y efectos grandes para $d \leq 0.80$.

Dependiendo del rigor del criterio que escojamos para asignar el valor de d_c serán necesarios más o menos estudios contradictorios con los tratados para modificar el sentido de los resultados.

Las críticas de este método se centran en la asunción de que todos los valores no recuperados toman valores nulos (IYENGAR & GREENHOUSE, 1988; OAKES, 1993).

IYENGAR & GREENHOUSE (1988) plantean una corrección donde parten del supuesto que los estudios no publicados siguen una distribución normal truncada. A partir de esta distribución estiman un valor medio, $M(\alpha)$, que incorporan a la expresión propuesta por Rosenthal:

$$z_\alpha = \frac{z_T + k_0 M(\alpha)}{\sqrt{k + k_0}}$$

En el mismo sentido corrige la técnica planteada por ORWIN para los tamaños del efecto.

7.1.1.ii. METODOS BASADOS EN LOS MUESTREOS TRUNCADOS.

HEDGES (1984b) calcula un estimador máximo verosímil del tamaño del efecto a partir de una función de probabilidad basada en el muestreo truncado y bajo el supuesto de publicación de todos los estudios con resultados significativos.

Se parte de la función de densidad $f(\mathbf{x} / \delta, \mathbf{n}, \alpha)$ para el estimador del tamaño del efecto, \mathbf{d} , bajo el supuesto de partida y considerando tamaños muestrales iguales:

$$f(\mathbf{x} / \delta, \mathbf{n}, \alpha) = \begin{cases} \frac{h(\mathbf{x} / \delta, \mathbf{n}, \alpha)}{A(\delta, \mathbf{n}, \alpha)} & \text{si } |\mathbf{x}| > \sqrt{\frac{2}{\mathbf{n}}} t_{\alpha; 2\mathbf{n}-2} \\ 0 & \text{en otro caso} \end{cases}$$

donde $h(\mathbf{x} / \delta, \mathbf{n}, \alpha)$ es la función de densidad de una distribución $\sqrt{\frac{2}{\mathbf{n}}}$ veces una t no central con $2\mathbf{n}-2$ grados de libertad y como parámetro central $\sqrt{\frac{\mathbf{n}}{2}}\delta$, y $A(\delta, \mathbf{n}, \alpha)$ es la probabilidad que una t no central, con $2\mathbf{n}-2$ grados de libertad, con parámetro central $\sqrt{\frac{\mathbf{n}}{2}}\delta$, sea mayor que $t_{\alpha, 2\mathbf{n}-2}$ o menor que $-t_{\alpha, 2\mathbf{n}-2}$.

A partir de la distribución truncada se construye una función de verosimilitud para estimar el tamaño del efecto poblacional, δ :

$$L(\delta) = \prod_{i=1}^k \frac{h(\mathbf{d}_i / \delta, \mathbf{n}_i, \alpha)}{A(\delta, \mathbf{n}_i, \alpha)}$$

Para el cálculo del estimador máximo verosímil se requiere

procedimientos numéricos.

HEDGES (1984b) presenta unas tablas donde provee la solución individual a la ecuación anterior para cada estudio en particular. Obteniendo el estimador global del tamaño del efecto como una media ponderada en concordancia a los estimadores habituales descritos en el capítulo de la medida de los resultados. El principal problema que plantea el método propuesto, es el supuesto de que todos los estudios censurados son no significativos. Este supuesto es demasiado estricto y difícilmente se cumple. Por ello se realizan diferentes modificaciones introduciendo funciones ponderadas que corrigen el estimador en función del criterio de no publicación de los estudios.

Así, IYENGAR & GREENHOUSE (1988) admiten la posibilidad de recolectar resultados no significativos aunque mantiene el supuesto de que todos los resultados significativos son admitidos en el estudio. RUST y cols., (1990) en lugar de considerar el punto de corte que define el proceso de censura, éste es estimado en el propio procedimiento: estima el punto de corte a partir del cual decide que son censurados o no los artículos.

En el mismo sentido DEAR & BEGG (1992) proponen un modelo semiparamétrico, en el cual una función ponderada describe la probabilidad de que los ensayos sean publicados.

Todas las modificaciones presentan el inconveniente del cálculo de los estimadores, si bien hoy en día existen paquetes estadísticos que permiten el cálculo de dichos estimadores mediante procesos numéricos fácilmente programables. Sin embargo, el inconveniente se presenta en la forma de la función ponderada cuando la información del proceso de

censura se desconoce, lo cual ocurre en la mayoría de los estudios.

7.1.1.iii. METODO DEL MUESTREO INVARIANTE EN BANCOS DE DATOS.

Este método fue propuesto por SIMES (1986c) y está basado en la búsqueda de estudios en un marco de referencia donde se pueden obtener tanto los estudios publicados como los estudios no publicados. Generalmente este marco de referencia suelen ser bancos de datos donde se registran los diferentes experimentos, por ejemplo *Internacional Cancer Resarch Data Bank* . Con esta información se puede ver las diferencias inducidas por los estudios no publicados y por lo tanto se puede calcular un factor de corrección para aquellos estudios no registrados.

Un inconveniente importante se debe, generalmente, a la inexistencia de bancos de datos en la mayoría de los temas científicos estudiados.

7.1.2. COMPARACIÓN DE LOS PROCEDIMIENTOS DE ESTIMACION DEL SESGO DE PUBLICACION

Se ha realizado una simulación para conocer cuál es el alcance del sesgo de publicación, así como de la fiabilidad de los métodos más extendidos para detectar dicho problema: el gráfico embudo, el número de fallos nulos y el estimador corregido de HEDGES.

Se desea representar una situación ideal en la cual se considera que se han realizado 100 estudios sobre un tema particular. Se establece que cada estudio

arroja un único resultado procedente de un contraste de dos medias.

Para ello se crean 100 contrastes (estudios) simulados a partir de dos muestras de igual número de observaciones y obtenidas de poblaciones Normales con parámetros, μ y σ , predeterminados y que definen un tamaño del efecto poblacional (diferencia de medias estandarizada).

Obviamente de cada contraste simulado podemos extraer un estimador del tamaño del efecto y un nivel de probabilidad.

Se establece que las cien comparaciones proceden de contrastes con diferentes grados de libertad. Por lo tanto, los cien estudios se dividen de grupos de 10 con el mismo número de observaciones totales (20, 30, 40, 50, 60, 80, 100, 150, 200, 300); así diez contrastes procederán de 20 observaciones, otros diez de 30 observaciones, etc.

Esquemáticamente:

POBLACION DE ESTUDIOS

10		10		10
estudios		estudios		estudios
de 20	de 60	de 300
observ.		observ.		observ.

Además se programan dos posibles situaciones:

- Las poblaciones Normales definen un tamaño del efecto poblacional pequeño ($\delta=0,25$).
- Las poblaciones Normales definen un tamaño del efecto poblacional grande ($\delta=0,75$).

Por lo tanto se simulan 100 contrastes bajo cada tamaño del efecto poblacional. Una vez definida la población objetivo se establecen dos criterios de eliminación de estudios mas importantes:

- ‡ sólo se consideran los estudios significativos
- ‡ sólo se recuperan los estudios con tamaños muestrales globales grandes. Se ha acordado como tamaños grandes los estudios superiores o iguales a 50 observaciones.

En la tabla 7.1 se observa los estimadores del tamaño del efecto (d) cuando se recuperen todos los estudios y cuando se produce una fuente de sesgo debida a cualquiera de los dos criterios antes citados.

Se observa como el sesgo de publicación cuando sólo se recuperan los estudios significativos es importante aunque menor cuando el tamaño del efecto es 0,75.

En el segundo de los criterios, del cual BEGG & BERLIN (1988) comentan que es una de las fuentes mas importantes que producen sesgo de publicación, se puede constatar que no afecta demasiado en la estimación del efecto e incluso presenta estimaciones mas próximas al verdadero parámetro poblacional que cuando se recuperan todos los estudios.

d			
δ	Todos los estudios	Sólo estudios Significativos	Sólo estudios con $n \geq 50$
0.25	0.2744 (k =100)	0.3838 (k =28)	0.2664 (k =70)
0.75	0.7579 (k =100)	0.7825 (k =86)	0.7571 (k =70)

Tabla 7.1: Se presentan los estimadores de los tamaños del efecto (d) para los dos efecto poblacionales (δ). Entre paréntesis aparecen el número de estudios (k) sobre los que se calcula el estimador en cada criterio (todos los estudios, estudios significativos, estudios con un $n \geq 50$).

En el estudio sobre la detección del sesgo se comprueba que **la representación en embudo** es un buen método exploratorio para determinar la existencia de un sesgo de publicación en ambos criterios de no recuperación de estudios (figuras 7.1. y 7.2.).

Se observa la aparición de formas incompletas de la estructura piramidal típica en aquellas representaciones que han sufrido la no recuperación de todos los estudios.

En el método del **número de fallos nulos** se ha obtenido el número de valores nulos o tamaños del efecto cero, necesarios para convertir los efectos estimados en cada situación (sólo se recogen los significativos o sólo se recogen los estudios con tamaños muestrales mayores o iguales de 50) en los del efecto crítico (Tabla 7.2). Para ello, hemos elegido estos valores críticos 0,25 y 0,1 para el tamaños del efecto 0,25 y 0,75, 0,5, 0,25 para el efecto de 0,75.

Si volvemos a fijarnos en la tabla 7.1 el estimador del tamaño del efecto de los estudios sólo significativos, cuando el parámetro poblacional es 0,25, es 0,3838, por lo tanto, el revisor obtendrá este valor.

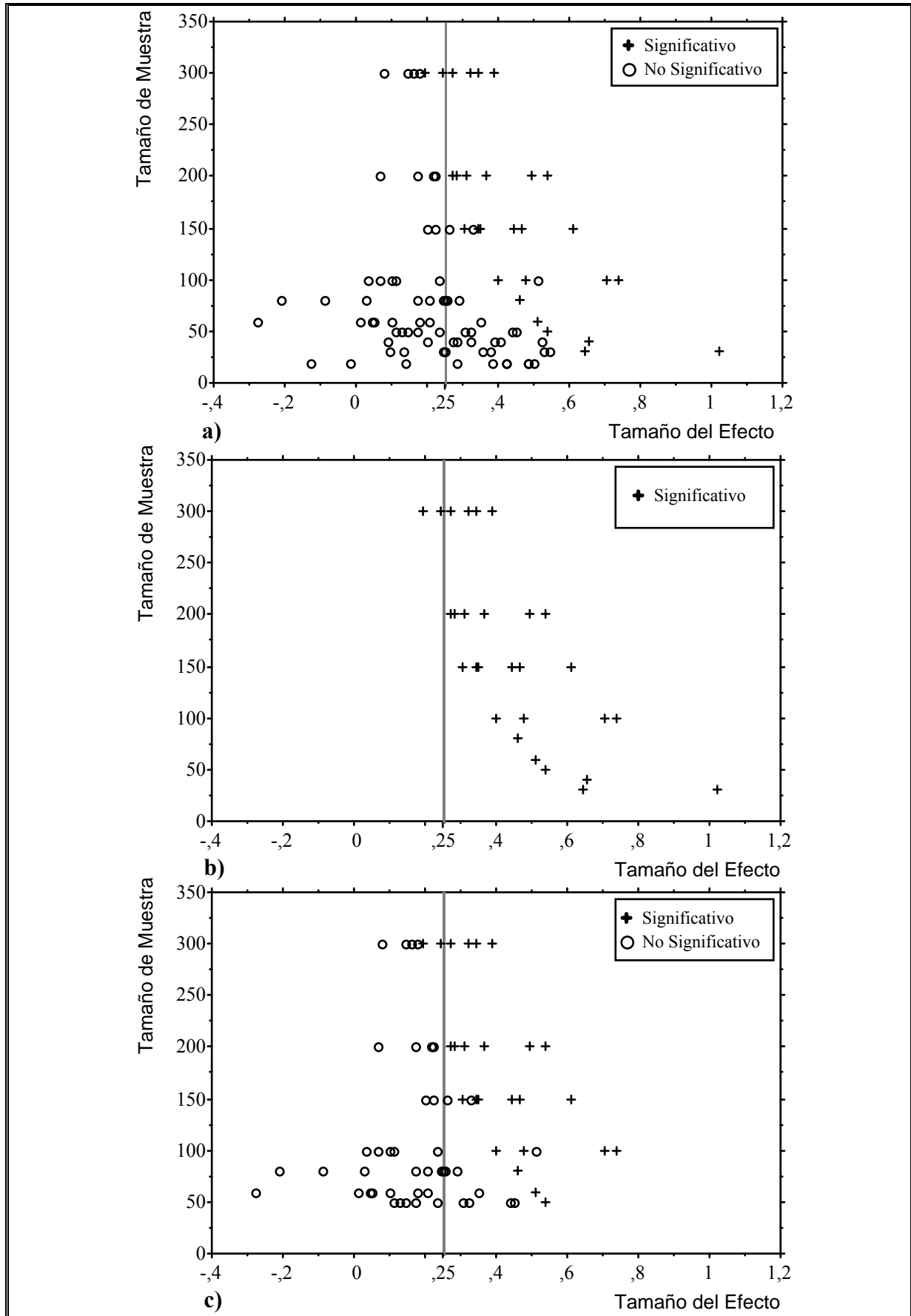


Figura 7.1: Gráfico en embudo de los tamaños del efecto para $\delta = 0,25$. a) se representan todos los estudios. b) Sólo los significativos. c) sólo los de tamaños muestrales ≥ 50 .

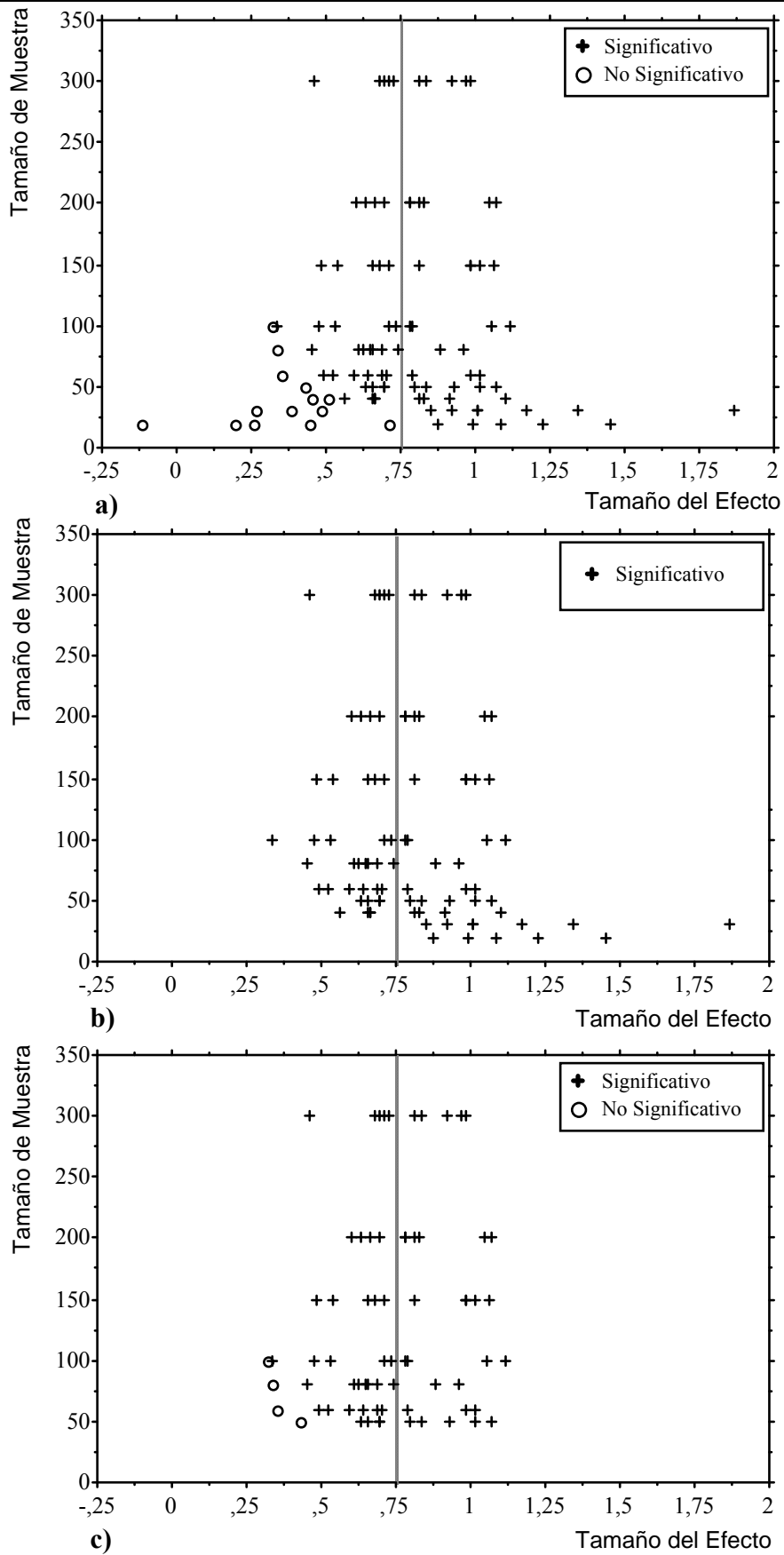


Figura 7.2: Gráfico en embudo de los tamaños del efecto para $\delta = 0,75$. a) se representan todos los estudios. b) Sólo los significativos. c) sólo los de tamaños muestrales ≥ 50 .

En la tabla 7.2 aparecen el número de estudios con tamaños cero que necesita para que ese estimador pase a tomar un valor de 0,10, en este caso serán 79.

	$\delta=0,25$		$\delta=0,75$		
	$d_c=0,25$	$d_c=0,10$	$d_c=0,75$	$d_c=0,5$	$d_c=0,25$
Significativos	15	79	4	49	183
$n \geq 50$	4	116	1	36	142

Tabla 7.2: El número de fallos nulos d_c representa el tamaño del efecto crítico.

Obviamente, se puede comprobar como el número de estudios nulos necesarios para convertir los estimadores calculados en los valores críticos es bajo (tabla 7.2). En este procedimiento basado en los tamaños del efecto se comprueba que solamente arroja resultados claros cuando las diferencias entre el estimador calculado por el revisor y el valor crítico, definido por él mismo, presentan grandes diferencias. Así, en este estudio simulado cuando el tamaño del efecto poblacional (δ) es 0,75, el revisor rechazaría la influencia del sesgo de publicación para su estimador cuando lo comparara con el efecto crítico de 0,25, pero no estaría tan seguro cuando se compara con el $d_c=0,50$, donde serían necesarios 49 estudios para pasar del estimador calculado 0.7825 a este valor crítico.

Se ha aplicado el **procedimiento de ROSENTHAL** convirtiendo los p-valores de los contrastes a puntuaciones **z**. El número de resultados nulos arrojados por este método, elegido el criterio de la significación, son 1750 y 24950 para los estimadores provenientes de los tamaños del efecto 0,25 y 0,75 respectivamente. Se observa como el número de estudios nulos para transformar el efecto calculado en un efecto cero es grande en ambos casos.

En el método de corrección del sesgo propuesto por HEDGES (1984b) el cálculo de los estimadores máximo verosímiles se realizó mediante la tabla proporcionada en dicho estudio. La tabla muestra en filas los estimadores del tamaño del efecto estimado para cada experimento y en columnas el tamaño muestral. La intersección del efecto con su tamaño muestral proporciona los estimadores corregidos. Dado que las tablas están definidas para tamaños muestrales máximos de 100 individuos, los resultados calculados en esta simulación no proceden del archivo inicial de 100 estudios sino de un nuevo archivo con los tamaños del efecto calculados a partir de los tamaños muestrales menores o iguales a 100 individuos.

Los resultados arrojados por este procedimiento tampoco son muy claros y los estimadores siguen presentado sesgo (tabla 7.3). En el caso del tamaño del efecto 0,25 se reduce el sesgo, pero sigue existiendo una diferencia notable.

Cuando el efecto es de 0,75 el método corrige demasiado y produce una subestimación del tamaños del efecto real.

δ	d	
	sin corregir	corregidos
0.25	0.5844 (k =10)	0.4964 (k =10)
0.75	0.7755 (k =56)	0.7343 (k =56)

Tabla 7.3: Estimadores del tamaño del efecto corregidos por el procedimiento propuesto por HEDGES (1984b)

En conclusión, los métodos gráficos son buenos para explorar la posible presencia de sesgo, cuando éste depende del tamaño muestral. El método del número de fallos nulos es una técnica imprecisa, pero útil, a la hora de reconocer diferencias importantes, mientras que el estimador de HEDGES corrige el sesgo insuficientemente en tamaños del efecto pequeños y en el caso

de efectos grandes, es un método bastante conservador.

7.2. PROBLEMA DE LA HETEROGENEIDAD (Apples and Oranges Problem).

Una de las críticas más contundentes a las técnicas meta-analíticas es la intención de conseguir un estimador del efecto global procedente de la integración de trabajos de investigación con diferentes características.

En muchas ocasiones los estudios no presentan uniformidad en el tipo de diseño, o en las variables dependientes e independientes elegidas, o en los métodos de medición, o en las unidades de muestreo, etc. GLASS (1978) resuelve el problema haciendo hincapié en el objetivo que pretende cubrir con el meta-análisis y se defiende de esta crítica expresando que *"es correcto la mezcla de naranjas y manzanas cuando queremos hablar de fruta"*.

Varios autores (JENICEK, 1987; MANSFIEEL & BUSSE, 1977; SACKS y cols., 1987) propugnan, para evitar el problema de la heterogeneidad, la definición clara de criterios de selección de estudios y la eliminación de estudios debido a su calidad metodológica. Sin embargo, en contrapartida, LIGHT (1987) y FLEISS (1987) discuten el posible sesgo que se produce por desechar información que puede ayudar a comprender el problema.

Una posible solución al problema sería considerar un modelo de efectos aleatorios para analizar los datos, sin embargo, se ha demostrado que ante presencia de fuerte heterogeneidad, tanto el modelo de efectos aleatorios como

el modelo de efectos fijos, no dan soluciones satisfactorias (GREENLAND & SALVAN, 1990; THOMPSON y POCOCK, 1991).

En ensayos clínicos controlados aleatorizados, la interpretación y valoración de un meta-análisis suele ser sencilla (PETO, 1987b; COLLINS y cols., 1987) puesto que en general, si el ensayo está bien diseñado, el sesgo es pequeño y la heterogeneidad no afecta tanto a los resultados. En los estudios "caso-control" la interpretación es más complicada ya que en general están sujetos a muchas fuentes de sesgo difíciles de controlar en cada estudio, y por tanto en el estudio meta-analítico. Recientemente han aparecido publicaciones que comparan resultados de meta-análisis buscados sobre ensayos controlados y sobre estudios observacionales. (Mac MAHON y cols., 1990; COLLINS, y cols., 1990).

Como en el caso anterior, no existe una solución mágica que consiga evitar el problema de la heterogeneidad. Sin embargo, lo que el revisor tiene que tener en cuenta es la existencia de este problema e intentar valorar su magnitud y considerarlo a la hora de exponer sus conclusiones.

La codificación de las características de los trabajos puede ser clave para enfrentarse con el problema y estudiar posibles causas que producen la heterogeneidad, las cuales podrían ser evitadas en futuras investigaciones.

7.3. PROBLEMAS DE INDEPENDENCIA DE LOS TAMAÑOS DEL EFECTO.

La dependencia de los tamaños del efecto puede afectar gravemente al análisis de los resultados, incrementándose la probabilidad de cometer un error tipo I (WOLF, 1990). Hay varias fuentes de dependencia que pueden afectar a los estudios meta-analíticos.

HEDGES (1990) describe cuatro principales causas:

- Cuando los estudios usan diferentes variables o diferentes medidas de una misma variable
- Cuando en un mismo estudio existen comparaciones múltiples con un mismo grupo control o entre varios tratamientos.
- Cuando muchas muestras en un mismo estudio son usadas para estimar el tamaño del efecto.
- Cuando varios estudios son llevados a cabo por un mismo investigador.

Las soluciones a la dependencia pasan por incluir cada tamaño del efecto como si fueran independientes (la unidad de análisis es el tamaño del efecto), usar la media (la unidad de análisis es cada estudio), seleccionar el mejor de los resultados o estimar diferentes tamaños del efecto para cada medida. Otras soluciones se refieren a estimaciones de las correlaciones entre las variables o los individuos (STRUBE, 1985; ROSENTHAL y RUBIN, 1986; GLASS y cols., 1981, HEDGES y OLKIN, 1985).

CAPITULO 8

UNA CONTRIBUCION AL
META-ANALISIS DE
DATOS MULTIVARIANTES
DESDE UNA PERSPECTIVA
MULTIVARIANTE

8.1. INTRODUCCION

Los trabajos encontrados en la bibliografía sobre integración de datos multivariantes son muy escasos y están enfocados a la integración de resultados de experimentos en los que se requiere comparar uno o varios tratamientos con un control y en los que la respuesta es multivariante.

Tras chequear las bases de datos y tras revisar cuidadosamente las referencias recogidas en las más de 250 separatas en las que está basado este trabajo, estamos en condiciones de afirmar que no aparece ni un sólo trabajo metodológico, tampoco práctico, en el que se lleve a cabo la integración de resultados de Análisis Multivariante. Ni siquiera con algún método de la familia de técnicas factoriales, las cuales están extendidas y son aplicadas de forma generalizada por biólogos, psicólogos, sociólogos, economistas, etc.

El trabajo de WARWICK y CLARKE de 1993, es un primer intento, pero se limita a discutir los resultados encontrados en varios Análisis Multivariantes; no se trata realmente de un Meta-análisis.

Parece como si en los estudios meta-analíticos estuviéramos trabajando con el prisma anterior a los años 60. Sin embargo, a partir de esa década, se produjo una ruptura con la estadística utilizada hasta ese momento y se impusieron, cada vez con más fuerza, los modernos métodos de Análisis de Datos los cuales representan un conjunto de técnicas que tratan de poner de manifiesto las relaciones entre unidades taxonómicas, entre las variables que las caracterizan y entre unidades y variables.

En este campo las técnicas factoriales que representan las OTUs, en un espacio euclídeo en el que se ponen de manifiesto cluster de individuos y/o poblaciones y se pueden identificar las variables responsables de su configuración, además de conocer su importancia en la clasificación, juegan un papel crucial.

Es en este contexto donde se centran nuestras contribuciones, las cuales abordan dos perspectivas bien diferenciadas:

- (a) Integración de resultados correspondientes a poblaciones/ individuos/ objetos (unidades taxonómicas en definitiva), en la cual se requiere la comparación de configuraciones euclídeas.
- (b) Comparación de estructuras de las variables en las que se ha de llevar a cabo la comparación de las componentes.

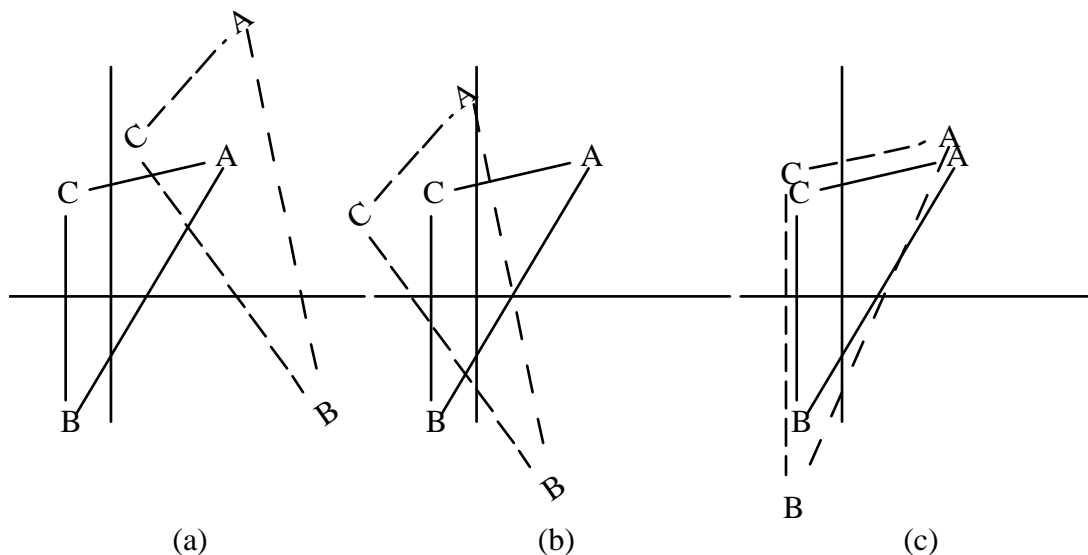
En la literatura matemática altamente especializada, este problema se aborda desde la perspectiva de los métodos "Procrustes" propuestos por GOWER (1975) y desde los modelos de Componentes Principales Comunes esbozados por KRZANOWSKI (1979, 1988). Nosotros creemos que integrar estas ideas en el contexto meta-analítico puede suponer un paso de gigante en aquellos campos en los cuales el armazón estadístico son técnicas basadas en las Componentes Principales.

8.2. METODOS PROCRUSTES

8.2.1. PROCRUSTES PARA DOS CONFIGURACIONES

Partimos de dos configuraciones de puntos (por ejemplo dos ordenaciones) que representan el mismo conjunto de unidades (por ejemplo lugares de muestreo). Una configuración se toma como fija "*target configuration*" y la otra se mueve hasta conseguir que ambas coincidan tan próximamente como sea posible.

Dos tipos de movimientos que preservan las distancias entre los puntos son las traslaciones (en el supuesto de que el centro de gravedad no sea fijo) y las rotaciones de los ejes.



Ejemplo de Análisis Procrustes.

- a) Configuraciones originales; b) Configuraciones después de centrar;
c) Configuraciones después de centrar y rotar.

(Figura adaptada de DIGBY y KEMPTON, 1987, pg. 113).

Algebráicamente, las dos configuraciones de puntos en un espacio r -dimensional vienen dadas por las matrices \mathbf{A} y \mathbf{B} (matrices $n \times r$). Si el número de dimensiones difiere, se completan con ceros.

Tomando \mathbf{A} como configuración fija ("*target matrix*") las coordenadas de \mathbf{B} deben ser reescaladas hasta conseguir un nuevo conjunto de coordenadas; es decir, una nueva matriz \mathbf{Z} que verifique que \mathbf{A} y $\mathbf{Z} = \mathbf{B}\mathbf{T}$ difieran lo menos posible.

Debemos restringir \mathbf{T} , de forma que \mathbf{T} pueda ser sólo un reflexión o una rotación, para lo cual se debe verificar que:

$$\mathbf{T}'\mathbf{T} = \mathbf{I} \Rightarrow \mathbf{T} = \mathbf{T}^{-1} \Rightarrow \mathbf{T}\mathbf{T}' = \mathbf{I}$$

O sea, \mathbf{T} debe ser una matriz ortogonal.

En este caso, obviamente, las interdistancias se mantienen, ya que para todo par \mathbf{x} , \mathbf{y} se verifica:

$$d(\mathbf{T}\mathbf{x}, \mathbf{T}\mathbf{y}) = (\mathbf{T}\mathbf{x} - \mathbf{T}\mathbf{y})' (\mathbf{T}\mathbf{x} - \mathbf{T}\mathbf{y}) = (\mathbf{x} - \mathbf{y})' \mathbf{T}' \mathbf{T} (\mathbf{x} - \mathbf{y}) = d(\mathbf{x}, \mathbf{y})$$

Hemos de considerar la restricción de que $\mathbf{T}'\mathbf{T} = \mathbf{I}$, o sea, $\mathbf{T}'\mathbf{T} - \mathbf{I} = \mathbf{0}$.

En la mayor parte de los métodos de ordenación, las escalas son arbitrarias, por lo que es muy aconsejable introducir el factor " k " directamente en la función de pérdida.

En este caso, un parámetro " k " debe ser incluido al definir la transformación, siendo de la forma $k\mathbf{B}\mathbf{T}$; así conseguiremos que las escalas de \mathbf{A} y \mathbf{B} sean comparables.

Si es necesaria una traslación de la configuración, tendremos que sumar un

vector \mathbf{t} a cada una de las filas de la matriz \mathbf{B} .

Por tanto, la función de pérdida general que incluye *rotación, escalado y traslación*, debe ser de la siguiente forma:

$$L = \text{tr} \left\{ \left[\mathbf{A} - (\mathbf{k}\mathbf{B}\mathbf{T} + \mathbf{j}\mathbf{t}') \right] \left[\mathbf{A} - (\mathbf{k}\mathbf{B}\mathbf{T} + \mathbf{j}\mathbf{t}') \right]' \right\} + \text{tr}[\Lambda(\mathbf{T}'\mathbf{T} - \mathbf{I})]$$

(\mathbf{j} es un vector de unos)

La transformación Procrustes que minimiza la función de pérdida que incluye movimientos rígidos (rotación, escalado y traslación), es de la forma:

$$\mathbf{T} = \mathbf{P}\mathbf{Q}'$$

donde \mathbf{P} y \mathbf{Q} son las matrices ortogonales que resultan de la descomposición en valores singulares de la matriz:

$$\mathbf{C} = \mathbf{B}' \left(\mathbf{I} - \frac{\mathbf{U}}{n} \right) \mathbf{A} = \mathbf{B}' \mathbf{Z}\mathbf{A}$$

con: $\mathbf{Z} = \mathbf{I} - \frac{\mathbf{U}}{n}$, $\mathbf{U} = \mathbf{j}\mathbf{j}'$, $n = \mathbf{j}'\mathbf{j}$ y siendo \mathbf{j} un vector de unos.

El factor de traslación será de la forma:

$$\mathbf{t} = \frac{(\mathbf{A} - \mathbf{k}\mathbf{B}\mathbf{T})'}{n} \mathbf{j}$$

y \mathbf{k} factor de escalado será:

$$\mathbf{k} = \frac{\text{tr}(\mathbf{T}'\mathbf{B}'\mathbf{Z}\mathbf{A})}{\text{tr}(\mathbf{B}'\mathbf{Z}\mathbf{A})}$$

Un desarrollo exhaustivo de estos métodos puede encontrarse por ejemplo en

BORG y LINGOES (1987).

8.2.2. PROCRUSTES PARA MAS DE DOS CONFIGURACIONES

Disponemos ahora de n configuraciones X_j ($j=1, \dots, n$) de orden $p \times m$, cada una de las cuales contiene las coordenadas de los mismos objetos en m dimensiones.

El problema es ahora: **¿Cuál es la similaridad o "congruencia" entre estas configuraciones?**

Esto nos conduce a otras dos preguntas:

-¿Qué transformaciones están permitidas?

-¿Qué criterio usaremos para evaluar la similaridad?

Para dar contestación a estas preguntas se pueden utilizar los modelos PINDIS de LINGOES y BORG (*Procrustean Individual Differences Scaling*) que se refieren a seis modelos en orden creciente de complejidad.

En los seis modelos se utilizan transformaciones lineales, y en todos ellos se utiliza un criterio mínimo cuadrático para evaluar la similaridad de las configuraciones transformadas.

En lugar de considerar una de las configuraciones como objetivo, supondremos ahora que existe una configuración "*centroide*" (o configuración consenso) de forma que el ajuste procrustes a este centroide es óptimo.

8.2.2.i. LOS MODELOS PINDIS de LINGOES y BORG.

El más sencillo es el análisis procrustes generalizado, originalmente desarrollado por GOWER (1975) al que nos referiremos como modelo GPA.

Sean \mathbf{R}_j una matriz ortogonal desconocida de orden $m \times m$, s_j un factor de escala (dilatación central) desconocido y \mathbf{u}_j un vector de traslación desconocido de orden $m \times 1$. \mathbf{Z} será la configuración centroide de orden $p \times m$, \mathbf{E}_j la matriz de residuales y $\mathbf{1}$ un vector $p \times 1$ de unos.

El modelo GPA puede escribirse como:

$$s_j = (\mathbf{X}_j - \mathbf{1}\mathbf{u}_j') \mathbf{R}_j = \mathbf{Z} + \mathbf{E}_j \quad \text{para } j = 1, \dots, n \quad [1]$$

Las transformaciones permitidas son:

- Transformaciones ortogonales (\mathbf{R}_j).
- Traslaciones (\mathbf{u}_j).
- Dilataciones centrales (s_j)

Las tres son transformaciones que no afectan a las distancias relativas, luego son un buen punto de comienzo cuando se desea comparar configuraciones Euclídeas que tienen la propiedad de que contienen información que no se modifica por transformaciones que preservan las distancias.

8.2.2.ii. MODELOS DE PONDERACIÓN DE LAS DIMENSIONES.

Se permite ahora que las dimensiones se extiendan o se compriman por separado.

Denotemos con \mathbf{Q}_j , \mathbf{S} y \mathbf{S}_j a las matrices ortogonales desconocidas de orden $m \times m$, con \mathbf{g}_j y \mathbf{h}_j vectores de traslación desconocidos y con \mathbf{W}_j una matriz diagonal desconocida de orden $m \times m$. Sea también \mathbf{Z} la matriz centroide.

El segundo modelo PINDIS puede escribirse:

$$(\mathbf{X}_j - \mathbf{1g}_j')\mathbf{Q}_j = (\mathbf{Z} - \mathbf{1h}_j')\mathbf{S}\mathbf{W}_j + \mathbf{E}_j \quad \text{para } j = 1, \dots, n \quad [2]$$

El tercer modelo PINDIS es:

$$(\mathbf{X}_j - \mathbf{1g}_j')\mathbf{Q}_j = (\mathbf{Z} - \mathbf{1h}_j')\mathbf{S}_j\mathbf{W}_j + \mathbf{E}_j \quad \text{para } j = 1, \dots, n \quad [3]$$

La diferencia entre ambos modelos es que en [2] \mathbf{Z} se rota sólo una vez mediante \mathbf{S} , mientras que en [3] se rota separadamente mediante cada \mathbf{S}_j .

Suponiendo que las configuraciones representan a n sujetos, la implicación psicológica del modelo [2] es que los sujetos están de acuerdo en las dimensiones subyacentes, pero difieren en la importancia que le dan a las dimensiones. En el modelo [3] los sujetos no sólo difieren en la importancia que le dan a las dimensiones, sino también en el punto de vista en el que las dimensiones de \mathbf{Z} son consideradas como importantes.

8.2.2.iii. MODELOS DE PONDERACION DE LOS ESTIMULOS

Los modelos 4 y 5 en PINDIS se caracterizan por una transformación que denominaremos "*ponderación de los estímulos*".

En estos modelos, los puntos de la configuración se unen con el origen del espacio m-dimensional y se permite que los vectores generados de esta manera sean alargados o acortados separadamente.

Sean \mathbf{g}_j , \mathbf{h} y \mathbf{h}_j vectores de traslación desconocidos de orden $m \times 1$; \mathbf{T}_j una matriz ortogonal desconocida de orden $m \times m$, y \mathbf{V}_j una matriz diagonal desconocida de orden $p \times p$.

El modelo 4 puede escribirse como:

$$(\mathbf{X}_j - \mathbf{1g}_j')\mathbf{T}_j = \mathbf{V}_j(\mathbf{Z} - \mathbf{1h}') + \mathbf{E}_j \quad \text{para } j = 1, \dots, n \quad [4]$$

y el modelo 5:

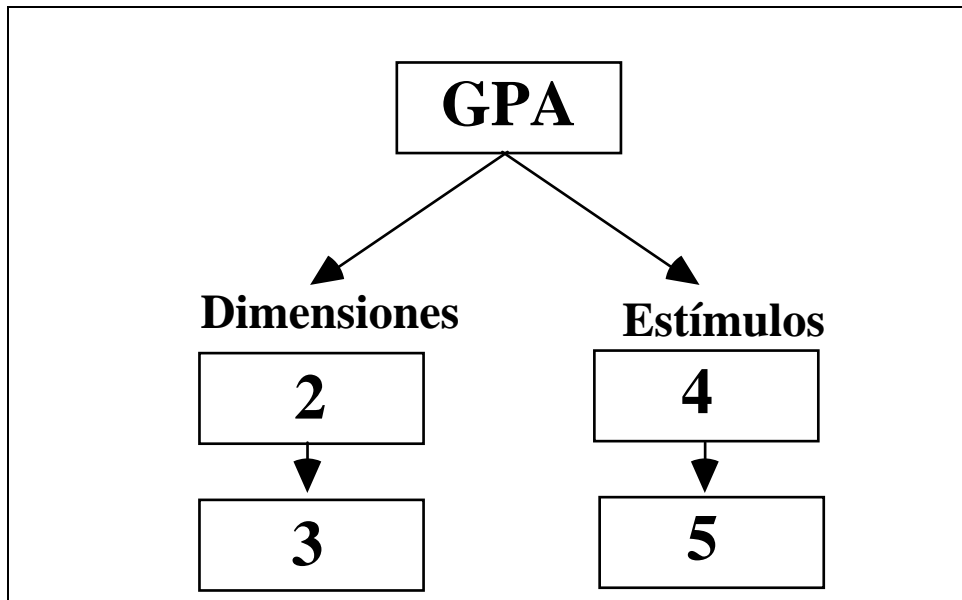
$$(\mathbf{X}_j - \mathbf{1g}_j')\mathbf{T}_j = \mathbf{V}_j(\mathbf{Z} - \mathbf{1h}_j') + \mathbf{E}_j \quad \text{para } j = 1, \dots, n \quad [5]$$

La interpretación del modelo [4] es que todos los sujetos comparten un punto de vista común en la configuración centroide \mathbf{Z} , los sujetos difieren en la centralidad que asignan a los estímulos.

En el modelo [5] además, tienen puntos de vista diferentes.

El sexto modelo combinaría la ponderación de las dimensiones y los estímulos, pero es muy difícil de interpretar en situaciones prácticas.

Jerarquía de los modelos



Los modelos pueden ajustarse secuencialmente

8.3. METODOS DE COMPARACION DE SUBESPACIOS

8.3.1. INTEGRACION DE LA INFORMACION DE DOS GRUPOS

El procedimiento descrito por KRZANOWSKI (1979, 1988) para la comparación de subespacios obtenidos por componentes principales es aplicable a muchas otras técnicas que en realidad son un análisis de Componentes Principales aplicado a la matriz de datos convenientemente transformada (Métodos Biplot, Análisis Factorial de Correspondencias, etc.).

Partimos de dos matrices de datos \mathbf{X}_1 y \mathbf{X}_2 correspondientes a las muestras de dos poblaciones sobre las que se han medido las mismas variables (x_1, \dots, x_p). Se supone que a ambas matrices se les ha aplicado el mismo tipo de transformación previa (centrado, estandarizado perfiles etc....).

Se calculan las componentes principales y suponemos que en ambos casos se obtienen k componentes. Los vectores propios (coordenadas estandarizadas para las columnas, base del nuevo sistema de referencia) correspondientes están contenidos en las matrices (de orden $p \times k$) \mathbf{H}_1 y \mathbf{H}_2 respectivamente.

Las coordenadas para las filas de las matrices de datos \mathbf{X}_1 y \mathbf{X}_2 , es decir, las coordenadas sobre las componentes principales se calculan como $\mathbf{G}_1 = \mathbf{X}_1 \mathbf{H}_1$ y $\mathbf{G}_2 = \mathbf{X}_2 \mathbf{H}_2$ respectivamente.

La comparación descriptiva de los dos conjuntos puede realizarse representando las dos primeras componentes y comparando las posiciones de los puntos, además es posible representar las variables mediante vectores utilizando las coordenadas contenidas en \mathbf{H}_1 y \mathbf{H}_2 y comparando las posiciones relativas sobre el gráfico. La representación de las variables nos dará una idea sobre la similitud entre la estructura de los ejes.

La estructura de los ejes está mejor representada utilizando los factores de carga en lugar de las coordenadas estandarizadas, por esta razón sería más conveniente representar las variables mediante las matrices en $\mathbf{H}_1 \mathbf{D}_1$ y $\mathbf{H}_2 \mathbf{D}_2$ donde \mathbf{D}_1 y \mathbf{D}_2 son las matrices diagonales de valores propios (varianzas) de las componentes principales de cada uno de los conjuntos.

Se trata de buscar una medida de lo cercanos que están los subespacios con el fin de poder compararlos. Dicha medida se basará en los ángulos ya que estamos tratando de comparar dos conjuntos de direcciones en el espacio. Sin embargo, no es suficiente con medir el ángulo entre los pares de componentes ya que éste puede ser grande incluso en el caso en el que los dos conjuntos generen el mismo subespacio (KRZANOWSKI, 1988).

Es necesario buscar primero el conjunto de ejes ortogonales que mejor haga coincidir las componentes.

De la misma manera que en el análisis procrustes las configuraciones deben transformarse previamente, seguiremos el mismo procedimiento suponiendo primero que una de las configuraciones es fija.

Definimos $\mathbf{N} = \mathbf{H}_1' \mathbf{H}_2 \mathbf{H}_2' \mathbf{H}_1$. El ángulo mínimo entre un vector del espacio de las primeras componentes principales del primer conjunto y uno casi paralelo en el espacio de las componentes principales del segundo conjunto está dado por $\cos^{-1} \sqrt{\lambda_1}$ donde λ_1 es el valor propio más grande de \mathbf{N} .

Sea λ_i el i -ésimo valor propio de \mathbf{N} , \mathbf{a}_i el vector propio asociado, $\mathbf{b}_i = \mathbf{H}_1 \mathbf{a}_i$ y $\mathbf{c}_i = \mathbf{H}_2 \mathbf{H}_2' \mathbf{b}_i$ ($i=1, \dots, k$) entonces $(\mathbf{b}_1, \dots, \mathbf{b}_k)$ forman un conjunto ortogonal de vectores en el subespacio del primer conjunto y $(\mathbf{c}_1, \dots, \mathbf{c}_k)$ forman un conjunto ortogonal de vectores en el subespacio del segundo conjunto.

Considerando los vectores \mathbf{b}_i y \mathbf{c}_i como columnas de las matrices \mathbf{B} y \mathbf{C} respectivamente, se obtienen las coordenadas para las columnas de las dos matrices originales, en los subespacios rotados, de forma que los ángulos sean mínimos.

Si, además de las componentes, son importantes los individuos, las coordenadas de los mismos en los espacios rotados son $\mathbf{X}_1 \mathbf{B}$ y $\mathbf{X}_2 \mathbf{C}$, respectivamente.

La representación conjunta de las coordenadas de los dos grupos, tras la rotación, nos permitirá conocer como se han modificado las posiciones si estas son interesantes para el estudio.

Es fácil demostrar (KRZANOWSKI, 1979) que la suma de los valores propios de \mathbf{N} es igual a la suma de los cosenos al cuadrado entre las componentes del primer grupo y las componentes del segundo.

Esta suma estará entre 0 (si los dos subespacios son completamente disjuntos), y k (si los dos subespacios son completamente coincidentes) por lo que una

medida relativa acotada entre 0 y 1 de la similaridad entre los dos subespacios es la siguiente

$$\frac{\sum_{i=1}^k \lambda_i}{k}$$

De acuerdo con los resultados anteriores, las similaridades entre los dos grupos se establecen a través de \mathbf{b}_i y \mathbf{c}_i , con λ_i midiendo la contribución de cada par a la similaridad total. Los vectores \mathbf{b}_i y \mathbf{c}_i están definidos con respecto al espacio p -dimensional original por lo que pueden interpretarse de la misma manera que las componentes principales.

El vector del espacio p -dimensional más cercano, tanto a \mathbf{b}_i como a \mathbf{c}_i , es la bisectriz del ángulo entre ellos, en el plano en el que están. Tal bisectriz viene dada por

$$\mathbf{d}_i = \{2(1 + \sqrt{\lambda_i})\} \left(\mathbf{I} + \frac{1}{\sqrt{\lambda_i}} \mathbf{H}_2 \mathbf{H}'_2 \right) \mathbf{b}_i$$

El conjunto $(\mathbf{d}_1, \dots, \mathbf{d}_k)$ que puede escribirse en una matriz \mathbf{D} puede tomarse como "media" (consenso) de los dos subespacios.

Las coordenadas de los individuos, en el subespacio medio, son $\mathbf{X}_1 \mathbf{D}$ y $\mathbf{X}_2 \mathbf{D}$ respectivamente.

En la práctica ocurre, a menudo, que los dos conjuntos de datos están caracterizados por un número distinto de componentes k_1 y k_2 . En este caso \mathbf{N} tiene solamente $k = \min(k_1, k_2)$ valores propios distintos de cero. La comparación de los dos subespacios puede realizarse a través de estos valores propios y los correspondientes vectores propios.

En resumen, hemos obtenido una medida de la similaridad de los resultados y la forma de integrar los distintos estudios cuando tratan sobre el mismo número de variables.

Información que debe aparecer en las publicaciones para que sea posible la integración de resultados tras varios análisis basados en las Componentes Principales

Para la aplicación de los resultados expuestos anteriormente es necesario encontrar en los estudios la siguiente información :

- Los datos originales, de forma que sea posible reproducir todo el proceso o bien,
- La matriz de covarianzas o de correlaciones con lo cual es posible calcular las componentes principales aunque no se disponga de una información directa sobre los individuos. o bien,
- Los valores y vectores propios de la matriz de covarianzas. o bien,
- La estructura factorial y los valores propios de forma que sea posible reconstruir los valores propios.

8.3.2. GENERALIZACION AL CASO EN QUE SEA NECESARIO INTEGRAR LA INFORMACION DE MÁS DE DOS GRUPOS

Supongamos ahora que disponemos de las componentes principales obtenidas de más de dos grupos. Esta situación será mucho más frecuente en la práctica.

Disponemos de g grupos con n_i unidades en cada uno de los grupos, en cada uno de ellos se han medido las mismas p variables y sobre cada uno se toman k componentes. (Si el número de componentes para cada uno de los grupos es diferente k_i entonces las comparaciones se realizan sobre $k = \min(k_1, \dots, k_i)$ componentes).

Denotamos con \mathbf{H}_t las matrices $p \times k$ de coeficientes de cada una de las variables sobre las componentes principales en el grupo t -ésimo. KRZANOWSKI (1979) prueba el siguiente resultado:

Sea \mathbf{b} un vector arbitrario en el espacio p -dimensional original y sea δ_t el ángulo entre \mathbf{b} y la línea más cercana en el subespacio definido por las k componentes principales del grupo t ($t = 1, \dots, g$). El valor \mathbf{b} que hace máximo

$$V = \sum_{t=1}^g \cos^2 \delta_t$$

está dado por el vector propio \mathbf{b}_1 correspondiente al mayor valor propio λ_1 de la matriz

$$\mathbf{H} = \sum_{t=1}^g \mathbf{H}_t \mathbf{H}_t'$$

Pueden tomarse muchas posibles medidas de la cercanía entre un vector \mathbf{b} y un conjunto de g subespacios k -dimensionales, pero se ha elegido esta porque la componente "media" más cercana a todos los conjuntos de componentes principales viene dada por \mathbf{b}_1 como se define en el resultado anterior.

Una medida de la discrepancia entre esta componente y el subespacio k -dimensional del t -ésimo grupo es

$$\delta_t = \cos^{-1} \{ (\mathbf{b}_1' \mathbf{H}_t \mathbf{H}_t' \mathbf{b}_1)^{1/2} \}$$

Completando la descomposición en valores y vectores propios de \mathbf{H} obtendremos el subespacio de dimensión k que más se parece a los g subespacios conjuntamente.

Los cosenos al cuadrado de los ángulos formados por cada grupo con las direcciones medias están contenidos en la matriz \mathbf{C} cuyos elementos son

$$c_{tj} = \mathbf{b}_j' \mathbf{H}_t \mathbf{H}_t' \mathbf{b}_j$$

y los ángulos correspondientes en la matriz Δ cuyos elementos son

$$\delta_{tj} = \cos^{-1} \{ (\mathbf{b}_j' \mathbf{H}_t \mathbf{H}_t' \mathbf{b}_j)^{1/2} \}$$

La suma de todos los valores propios es igual a gxp y si las componentes son iguales, solamente los k primeros valores propios son distintos de cero y todos iguales a g . Por tanto, una medida de la similaridad global de los subespacios es

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i} = \frac{\sum_{i=1}^k \lambda_i}{gk}$$

Una medida de la similaridad de todos los subespacios con cada una de las nuevas componentes principales, es decir, una medida de hasta que punto la nueva componente está presente en todos los grupos es

$$\frac{\lambda_i}{g}$$

Se verifica además que

$$\lambda_i = \sum_{t=1}^g c_{ti} = \sum_{t=1}^g \mathbf{b}'_i \mathbf{H}_t \mathbf{H}'_t \mathbf{b}_i$$

de forma que los elementos de la matriz \mathbf{C} pueden interpretarse como porcentajes de la variabilidad de cada componente explicados por cada uno de los grupos o por una dirección del subespacio de cada uno de los grupos.

Es posible calcular también alguna información suplementaria que puede ser útil en la interpretación de los resultados. El coseno del ángulo entre las componentes principales de los subespacios originales y los subespacios de los

grupos se puede interpretar como la similaridad (coeficiente de correlación) entre las mismas.

Los cosenos están contenidos en la matriz $\mathbf{H}'_j \mathbf{B}$ donde B es la matriz que tiene los vectores propios \mathbf{b}_j en columnas. Los ángulos que forman las componentes se calculan directamente a partir de los cosenos.

La proporción de la variabilidad del grupo t-ésimo explicada por la componente común j-ésima se puede calcular como

$$\frac{\mathbf{b}'_j \mathbf{X}'_t \mathbf{X}_t \mathbf{b}_j}{\text{traza}(\mathbf{X}'_t \mathbf{X}_t)}$$

la variabilidad del grupo t-ésimo explicada por todas las componentes comunes como

$$\frac{\text{traza}(\mathbf{B}' \mathbf{X}'_t \mathbf{X}_t \mathbf{B})}{\text{traza}(\mathbf{X}'_t \mathbf{X}_t)}$$

La variabilidad global explicada por la j-ésima componente común se calcula como

$$\frac{\sum_{t=1}^g \mathbf{b}'_j \mathbf{X}'_t \mathbf{X}_t \mathbf{b}_j}{\sum_{t=1}^g \text{traza}(\mathbf{X}'_t \mathbf{X}_t)}$$

8.3.3. COMPARACION E INTEGRACION DE SUBESPACIOS OBTENIDOS TRAS UN ANALISIS DE COMPONENTES PRINCIPALES A DOS MATRICES DE DATOS REALES

Se han considerado como datos los índices, con base en 1985, correspondientes a 12 países europeos en 8 ramas incluidas dentro del índice de producción industrial. Se dispone de dos matrices de datos correspondientes a los años 1988 y 1989.

Los países fueron: Bélgica (BE), Alemania (DE), Dinamarca (DK), España (ES), Francia (FR), Grecia (GR), Irlanda (IR), Italia (IT), Luxemburgo (LU), Holanda (NL), Portugal (PO), Reino Unido (UK).

Las 8 Ramas: ENERGIA (1), INDUSTRIAS EXTRACTIVAS (2), INDUSTRIAS MANUFACTURERAS (3), INDUSTRIAS MINERAS NO METALICAS (4), INDUSTRIAS DE TRANSFORMACION DE METAL (5), FABRICACION DE ARTICULOS DE METAL (6), CONSTRUCCION Y MAQUINARIA (7), INDUSTRIAS ALIMENTICIAS (8).

Los datos han sido tomados de la Tesis Doctoral de VICENTE-TAVERA (1992) donde puede leerse un descripción exhaustiva de los mismos.

Se calcularon las componentes principales de ambas matrices de datos sobre las matrices de covarianzas (centrando los datos originales), ya que se trata de datos en escalas de medida comparables.

La tabla de datos para 1988 es la siguiente

	1	2	3	4	5	6	7	8
BE	104.3	67.8	108.8	118.9	105.2	97.4	102.5	108.9
DE	99.3	88.8	107.0	104.8	107.9	108.8	105.0	105.1
DK	121.0	113.5	104.8	102.8	102.9	106.5	103.1	104.5
ES	102.4	78.6	112.9	119.7	134.0	119.7	110.3	111.0
FR	105.9	88.9	107.9	108.4	109.4	107.8	106.4	104.2
GR	113.4	101.7	102.3	110.1	101.3	100.2	81.5	95.4
IR	107.6	95.0	129.0	92.9	164.2	92.2	110.1	121.9
IT	113.5	129.6	114.0	110.2	119.8	115.9	113.4	111.6
LU	117.7	124.5	109.4	130.5	106.6	119.8	91.3	97.4
NL	89.2	84.6	108.3	120.7	103.7	112.8	99.9	108.1
PO	118.5	88.3	115.7	132.2	113.4	117.2	116.8	117.1
UK	99.3	93.5	114.3	119.0	112.8	111.6	105.6	105.7

Tabla 8.1: Datos del IPI para los doce países de la CEE y 8 ramas de actividad.

Para detalle sobre países y ramas ver el texto.

La inercia absorbida por las componentes principales es

28.64	25.11	16.94
-------	-------	-------

Las tres primeras absorben el 70%

La estructura de las componentes puede verse en los vectores propios que aparecen a continuación

	c1	c2	c3
1	0.1307	-0.3072	0.0538
2	0.3725	-0.8482	0.0316
3	-0.2629	-0.1273	0.1479
4	0.2311	0.2000	0.6678
5	-0.7388	-0.3553	0.0850
6	0.1667	-0.0339	0.5563
7	-0.2584	-0.0373	0.4142
8	-0.2851	-0.0347	0.2001

Tabla 8.2: Tres primeros vectores propios del análisis de Componentes Principales de la matriz de datos de año 1988.

La primera componente está asociada principalmente a la variable **5**, la segunda a la variable **2** y la tercera a las variables **4, 6 y 7**.

La tabla de datos para 1989 es la siguiente

	1	2	3	4	5	6	7	8
BE	101.6	56.6	113.4	128.9	111.0	99.9	113.4	116.6
DE	101.0	89.1	112.9	112.5	116.0	114.9	115.2	110.4
DK	100.5	120.0	107.2	103.8	109.1	118.6	105.0	104.4
ES	108.5	84.6	117.8	128.8	143.1	131.2	119.5	109.4
FR	108.6	86.8	112.7	110.3	114.8	114.1	114.2	106.5
GR	112.7	57.2	105.9	105.7	98.3	85.6	100.9	105.5
IR	117.3	129.2	118.5	117.4	125.6	119.8	122.5	129.2
IT	120.3	116.1	145.3	110.6	192.5	96.9	118.2	122.3
LU	122.2	135.5	118.0	165.4	116.6	136.3	95.1	104.3
NL	94.8	90.6	112.8	124.7	109.7	119.6	109.2	110.3
PO	137.2	154.8	118.3	137.6	111.7	118.8	119.0	120.8
UK	89.8	78.9	119.2	121.0	120.5	112.6	110.1	105.4

Tabla 8.3: Datos del IPI para los doce países de la CEE y 8 ramas de actividad.

Para detalle sobre países y ramas ver el texto.

La inercia absorbida por las componentes principales de la matriz de datos centrada es

32.69	24.40	14.70
-------	-------	-------

Las tres primeras absorben el 72% (Porcentaje similar al del primer grupo)

La estructura de las componentes puede verse en los vectores propios que aparecen a continuación

	c1	c2	c3
1	0.2727	-0.0104	0.1320
2	0.8432	-0.2169	0.3511
3	0.1488	0.2976	-0.1191
4	0.2263	-0.2846	-0.7582
5	0.2861	0.8286	-0.2825
6	0.2129	-0.2434	-0.3933
7	0.0496	0.1509	0.1171
8	0.1079	0.1211	0.1482

Tabla 8.4: Tres primeros vectores propios del análisis de Componentes Principales de la matriz de datos de año 1989.

La primera componente está asociada principalmente a la variable **2**, la segunda a la variable **5** y la tercera a las variables **4**. Es decir, las componentes 1 y 2 tienen estructuras similares salvo el orden de importancia. Si observamos la inercia absorbida por los dos primeros ejes en el año 1988 veremos como los valores para ambos años son similares por lo que los ejes no están claramente separados; esto implica que, efectivamente las componentes son similares luego el procedimiento de integración debería de dar como resultado cosenos altos (cercaos a 1) y ángulos pequeños.

La misma situación puede verse en la figura 8.1. en el que se representan conjuntamente las dos primeras componentes principales de los dos años en estudio.

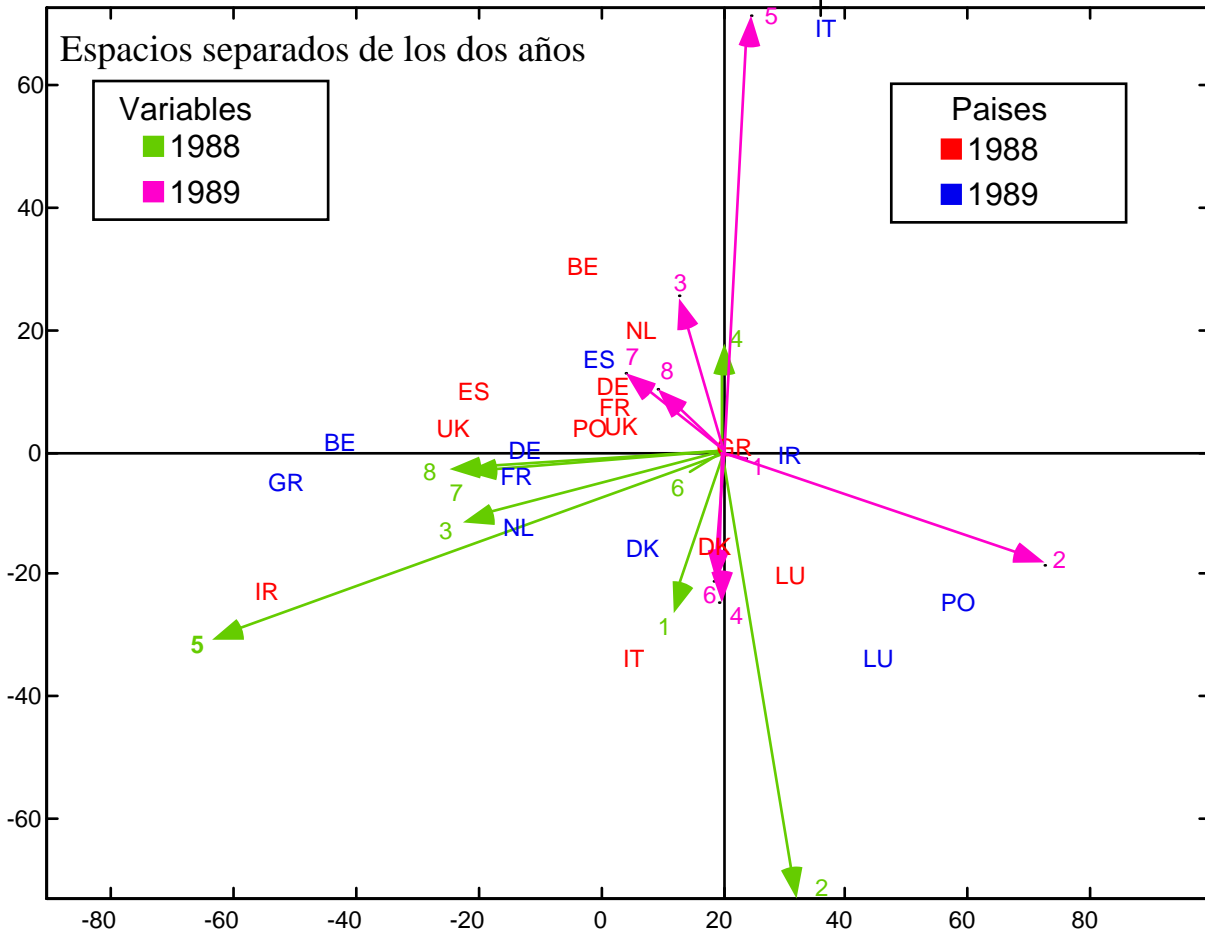


Figura 8.1: Representación gráfica conjunta de las dos primeras componentes correspondientes a los dos años de estudio.

Obsérvese como la estructura de los vectores que representan a las variables es muy similar en los dos años salvo una rotación de aproximadamente 90 grados. Tras la aplicación del procedimiento descrito en la teoría se obtuvieron los siguientes resultados

Valores propios λ_i		
0.9925	0.9746	0.5582
Cosenos $\sqrt{\lambda_i}$		
0.9963	0.9872	0.7471
Ángulos $\cos^{-1} \sqrt{\lambda_i}$		
4.9303	9.1771	43.3562

Tabla 8.5: Valores propios, coseno del ángulo entre pares de componentes de los dos subespacios y cuantificación de los ángulos

Los resultados indican que tenemos dos componente prácticamente coincidentes ya que la similaridad total entre los subespacios es: 0.8418

La estructura de las nuevas componentes para el año 1988 (matriz **B**) aparece en la tabla siguiente:

	c1	c2	c3
1	0.1323	0.3053	-0.0604
2	0.3675	0.8043	-0.2778
3	0.2580	-0.0308	0.1992
4	-0.3600	0.2847	0.5733
5	0.7517	-0.2185	0.2579
6	-0.1355	0.3501	0.4444
7	0.1660	0.0190	0.4602
8	0.2006	-0.0847	0.2741

Tabla 8.6: Estructura de las nuevas componentes para 1988.

La estructura de las nuevas componentes para el año 1989 (matriz **C**) aparece en la tabla siguiente:

	c1	c2	c3
1	0.1700	0.2339	-0.0611
2	0.3612	0.8217	-0.1784
3	0.2831	-0.0412	0.1538
4	-0.3226	0.3224	0.5259
5	0.7354	-0.2194	0.3765
6	-0.1947	0.2979	0.2694
7	0.1796	-0.0380	-0.0526
8	0.1946	0.0270	-0.0723

Tabla 8.7: Estructura de las nuevas componentes para 1989.

Obsérvese como ahora la estructura de los dos primeros vectores propios es muy similar, en el tercero la estructura es similar salvo para las variables **7** y **8**.

Los resultados se muestran, gráficamente, en la figura 8.2. donde se representan las nuevas estructuras y la proyección de las dos matrices originales

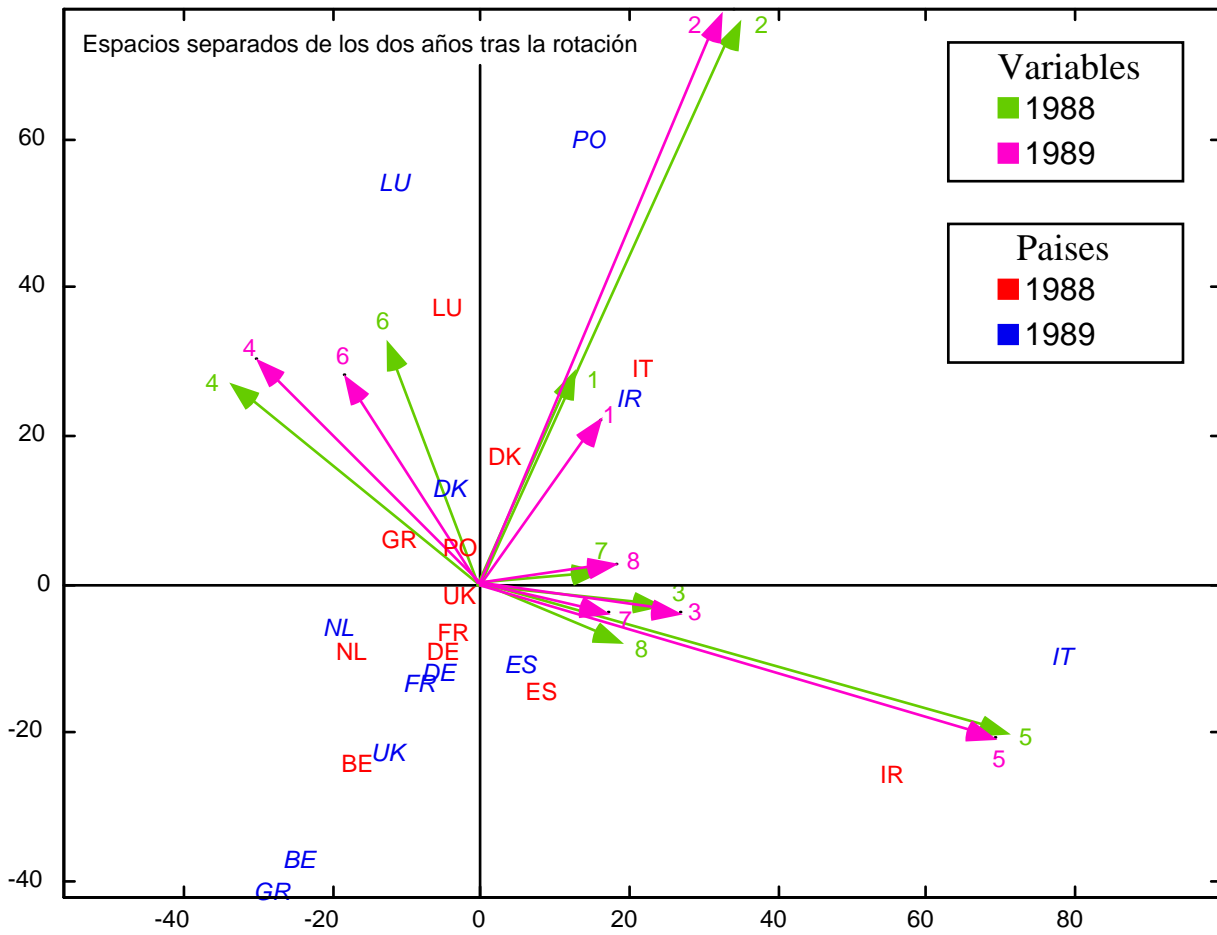


Figura 8.2: Representación de las nuevas estructuras de las matrices originales en el primer plano principal.

Es claro como la estructura de las relaciones entre las variables es muy similar para los dos años de estudio. Con respecto a los países, la mayor parte ocupa posiciones próximas en los dos años, excepto Italia e Irlanda que intercambian sus posiciones en los dos años de estudio.

La figura siguiente muestra la proyección sobre las componentes integradas.

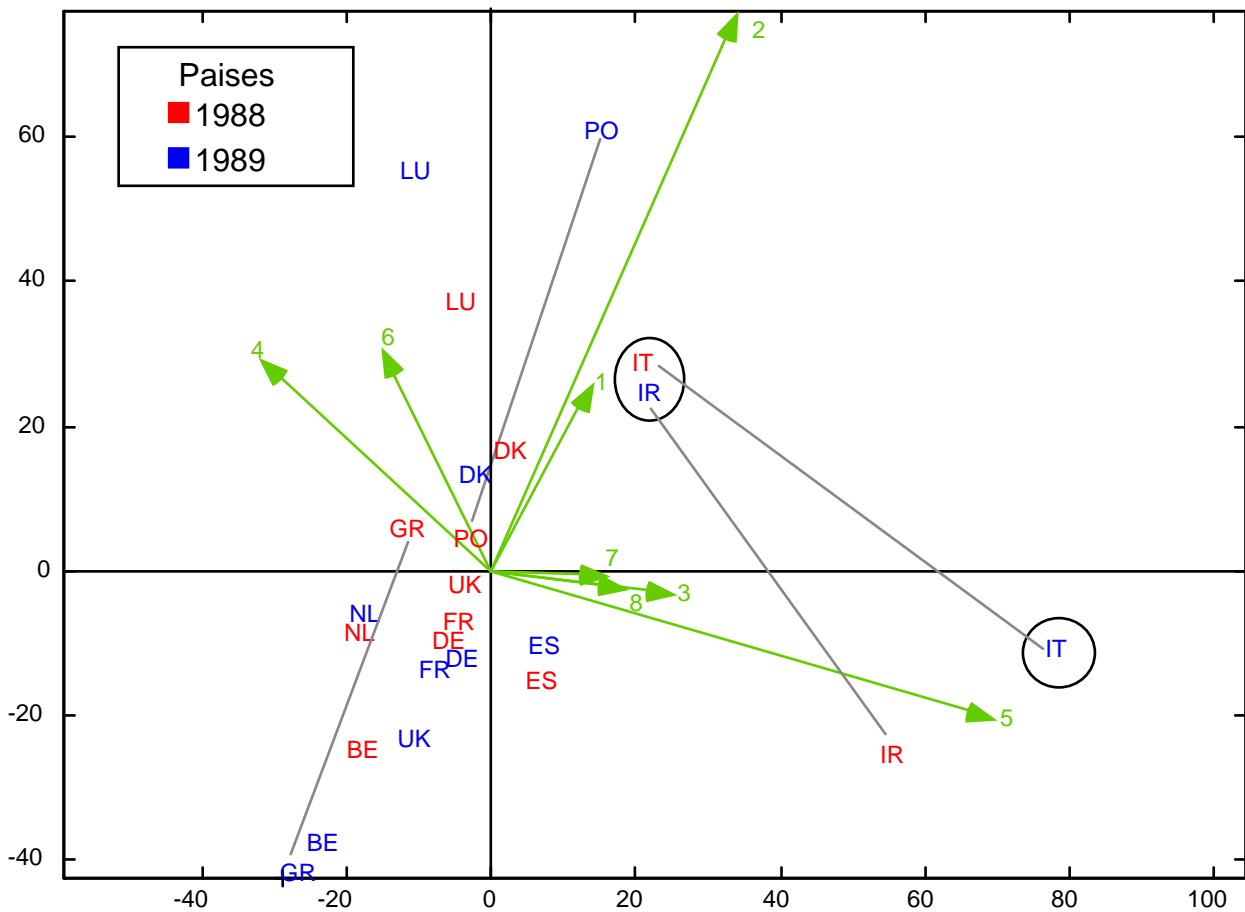


Figura 8.3: Proyección de países y ramas de actividad en los dos años de estudio sobre el espacio integrado.

8.3.4. COMPARACION E INTEGRACION DE VARIOS SUBESPACIOS OBTENIDOS TRAS UN ANALISIS DE COMPONENTES PRINCIPALES: DATOS REALES

Se han añadido a los datos del ejemplo anterior los años 1978 y 1979. Se ha suprimido Luxemburgo debido a que presentaba datos en los dos años que podían considerarse como "outliers" y se ha suprimido una de las variables que presentaba muchos datos faltantes en los dos nuevos años.

Los datos son entonces 4 matrices (correspondientes a los años 78, 79, 99 y 89) de 11 países por 7 variables, 7 ramas de actividad:

INDUSTRIAS EXTRACTIVAS (1), INDUSTRIAS MANUFACTURERAS (2), INDUSTRIAS MINERAS NO METALICAS (3), INDUSTRIAS DE TRANSFORMACION DE METAL (4), FABRICACION DE ARTICULOS DE METAL (5), CONSTRUCCION Y MAQUINARIA (6), INDUSTRIAS ALIMENTICIAS (7).

Describiremos primero las componentes principales de los 4 años por separado para realizar después la integración.

Análisis para el año 1978

La tabla de datos es la siguiente

	1	2	3	4	5	6	7
BE	115.2	93.6	134.0	92.7	95.6	89.2	78.2
DE	104.9	91.1	110.3	84.7	97.8	95.4	86.7
DK	91.6	79.4	123.7	76.8	69.3	83.0	75.4
ES	58.1	98.4	128.1	109.4	114.2	102.5	85.7
FR	113.4	99.9	106.4	99.5	104.7	99.5	89.6
GR	30.2	92.5	98.4	103.4	98.4	123.9	82.5
IR	123.6	71.8	99.7	47.8	113.1	78.7	81.9
IT	97.7	91.7	99.7	96.0	115.0	102.9	87.2
NL	106.8	87.6	103.6	85.7	92.6	90.2	90.8
PO	70.3	75.1	75.9	91.4	94.3	96.1	88.0
UK	77.1	106.1	116.0	107.5	124.9	122.2	94.1

Tabla 8.8: Datos del IPI en 1978, en los países de la CEE para las ramas de actividad consideradas.
Para detalle ver texto más arriba.

La inercia acumulada absorbida por las componentes principales es

57.4236	77.9321	92.3727
---------	---------	---------

Las tres primeras absorben, pues, el 92%.

La estructura de las componentes puede verse en los vectores propios que aparecen a continuación y en la representación gráfica correspondiente.

	c1	c2	c3
1	0.8001	-0.4011	0.2117
2	-0.1591	-0.4419	0.0089
3	0.0419	-0.5556	-0.6610
4	-0.4154	-0.3924	-0.0588
5	-0.1122	-0.3818	0.6429
6	-0.3815	-0.1681	0.2140
7	-0.0462	-0.0848	0.2360

Tabla 8.9: Vectores propios correspondientes al Análisis de Componentes Principales de la tabla de datos del año 1978.

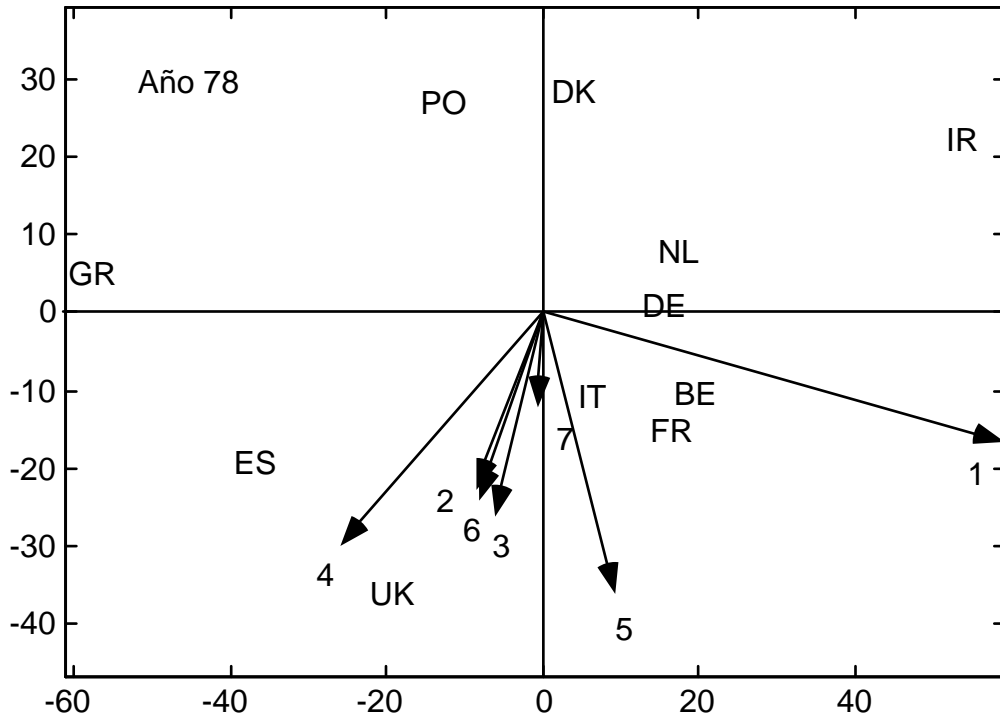


Figura 8.4: Proyección sobre el primer plano principal de países y ramas de actividad: Absorción de inercia en el plano 78%.

Análisis para el año 1979

La tabla de datos es la siguiente

	1	2	3	4	5	6	7
BE	106.9	98.2	134.4	97.2	100.5	96.4	80.7
DE	109.9	95.5	119.4	88.7	101.3	98.9	90.3
DK	96.5	82.3	120.9	80.0	77.2	84.7	78.7
ES	62.3	98.5	121.6	100.9	111.2	87.4	89.2
FR	114.2	103.9	112.8	103.1	112.7	103.3	91.5
GR	32.4	98.2	105.0	109.3	104.7	117.1	84.2
IR	138.3	77.1	107.4	54.4	131.7	87.3	84.7
IT	104.7	98.2	105.6	97.9	116.6	104.2	93.3
NL	113.0	90.3	103.0	88.2	96.7	91.1	93.8
PO	84.8	79.9	76.1	89.2	88.8	103.3	89.3
UK	98.8	105.9	116.0	105.1	122.9	118.0	95.5

Tabla 8.10: Datos del IPI en 1979, en los países de la CEE para las ramas de actividad consideradas.

Para detalle ver texto más arriba.

La inercia acumulada absorbida por las componentes principales es ahora,

55.9112	75.5320	89.3900
---------	---------	---------

Las tres primeras absorben el 89%.

La estructura de las componentes puede verse en los vectores propios y la representación gráfica que aparecen a continuación

	c1	c2	c3
1	0.8880	-0.1854	0.0143
2	-0.1282	-0.4455	-0.1033
3	0.0475	-0.3852	-0.8066
4	-0.3743	-0.3912	-0.0670
5	0.1279	-0.5985	0.4160
6	-0.1906	-0.2974	0.3507
7	0.0050	-0.1384	0.1949

Tabla 8.11: Vectores propios correspondientes al Análisis de Componentes Principales de la tabla de datos del año 1979.

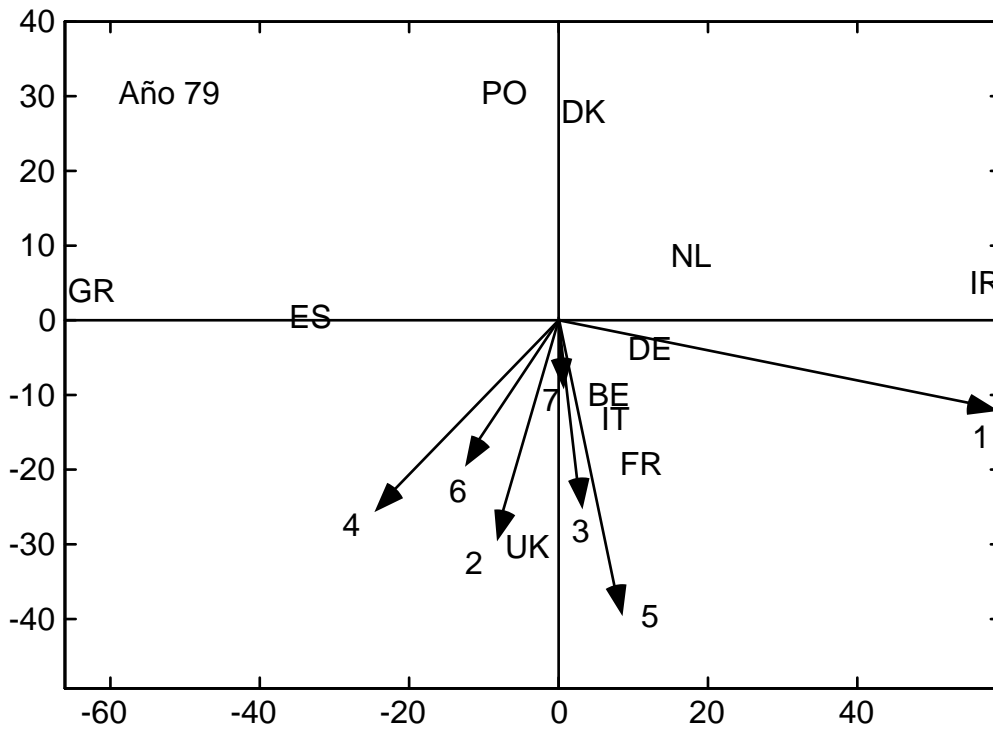


Figura 8.5: Proyección sobre el primer plano principal de países y ramas de actividad: Absorción de inercia en el plano 75.53%.

Análisis para el año 1988

La tabla de datos es la siguiente

	1	2	3	4	5	6	7
BE	67.8	108.8	118.9	105.2	97.4	102.5	108.9
DE	88.8	107.0	104.8	107.9	108.8	105.0	105.1
DK	113.5	104.8	102.8	102.9	106.5	103.1	104.5
ES	78.6	112.9	119.7	134.0	119.7	110.3	111.0
FR	88.9	107.9	108.4	109.4	107.8	106.4	104.2
GR	101.7	102.3	110.1	101.3	100.2	81.5	95.4
IR	95.0	129.0	92.9	164.2	92.2	110.1	121.9
IT	129.6	114.0	110.2	119.8	115.9	113.4	111.6
NL	84.6	108.3	120.7	103.7	112.8	99.9	108.1
PO	88.3	115.7	132.2	113.4	117.2	116.8	117.1
UK	93.5	114.3	119.0	112.8	111.6	105.6	105.7

Tabla 8.12: Datos del IPI en 1988, en los países de la CEE para las ramas de actividad consideradas. Para detalle ver texto más arriba.

La inercia acumulada absorbida por las componentes principales es

46.5822	77.0446	93.6786
---------	---------	---------

Las tres primeras absorben el 94%.

La estructura de las componentes puede verse en los vectores propios y la representación gráfica que aparecen a continuación

	c1	c2	c3
1	0.0827	-0.9326	0.2495
2	0.3084	0.0620	0.0983
3	-0.2303	0.3142	0.5317
4	0.8466	0.1004	-0.0486
5	-0.0976	0.0030	0.5838
6	0.2266	0.0791	0.5082
7	0.2595	0.1070	0.2099

Tabla 8.13: Vectores propios correspondientes al Análisis de Componentes Principales de la tabla de datos del año 1988.

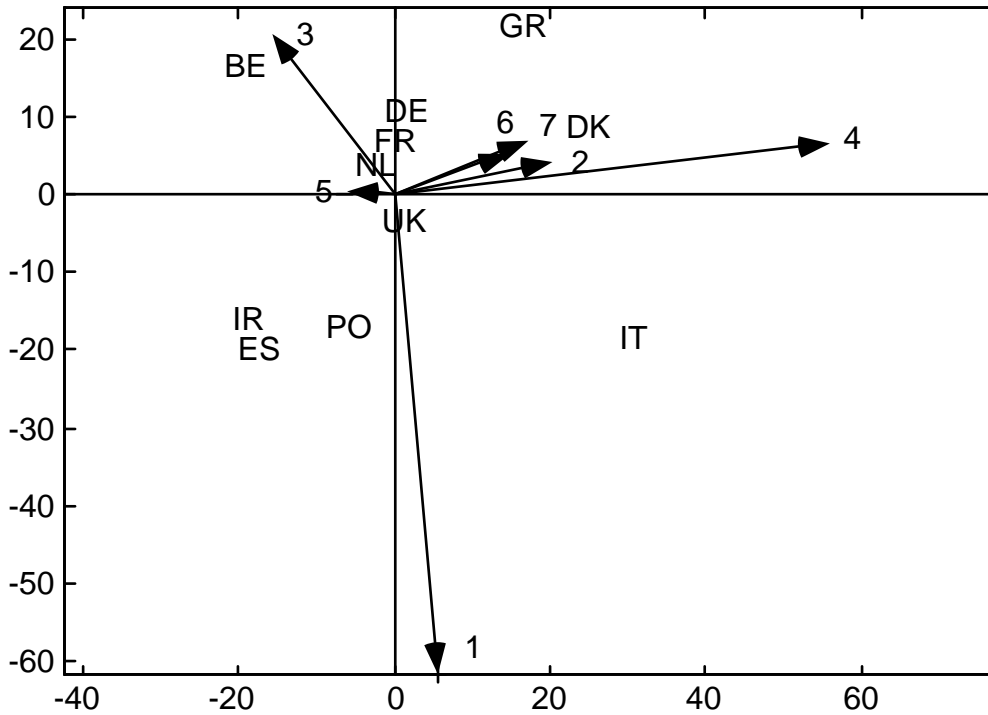


Figura 8.6: Proyección sobre el primer plano principal de países y ramas de actividad: Absorción de inercia en el plano 77%.

Análisis para el año 1989

La tabla de datos es la siguiente

	1	2	3	4	5	6	7
BE	56.6	113.4	128.9	111.0	99.9	113.4	116.6
DE	89.1	112.9	112.5	116.0	114.9	115.2	110.4
DK	120.0	107.2	103.8	109.1	118.6	105.0	104.4
ES	84.6	117.8	128.8	143.1	131.2	119.5	109.4
FR	86.8	112.7	110.3	114.8	114.1	114.2	106.5
GR	57.2	105.9	105.7	98.3	85.6	100.9	105.5
IR	116.1	145.3	110.6	192.5	96.9	118.2	122.3
IT	129.2	118.5	117.4	125.6	119.8	122.5	129.2
NL	90.6	112.8	124.7	109.7	119.6	109.2	110.3
PO	154.8	118.3	137.6	111.7	118.8	119.0	120.8
UK	78.9	119.2	121.0	120.5	112.6	110.1	105.4

Tabla 8.14: Datos del IPI en 1989, en los países de la CEE para las ramas de actividad consideradas. Para detalle ver texto más arriba.

La inercia acumulada absorbida por las componentes principales es

54.7762	85.4557	93.4354
---------	---------	---------

Las tres primeras absorben el 93.43%.

La estructura de las componentes puede verse en los vectores propios y la representación gráfica que aparecen a continuación

	c1	c2	c3
1	0.8059	-0.5069	0.2654
2	0.2150	0.2789	-0.0022
3	0.0586	-0.0949	-0.6372
4	0.4871	0.7768	-0.0886
5	0.1387	-0.2271	-0.6681
6	0.1333	0.0289	-0.2611
7	0.1627	0.0206	-0.0334

Tabla 8.15: Vectores propios correspondientes al Análisis de Componentes Principales de la tabla de datos del año 1989.

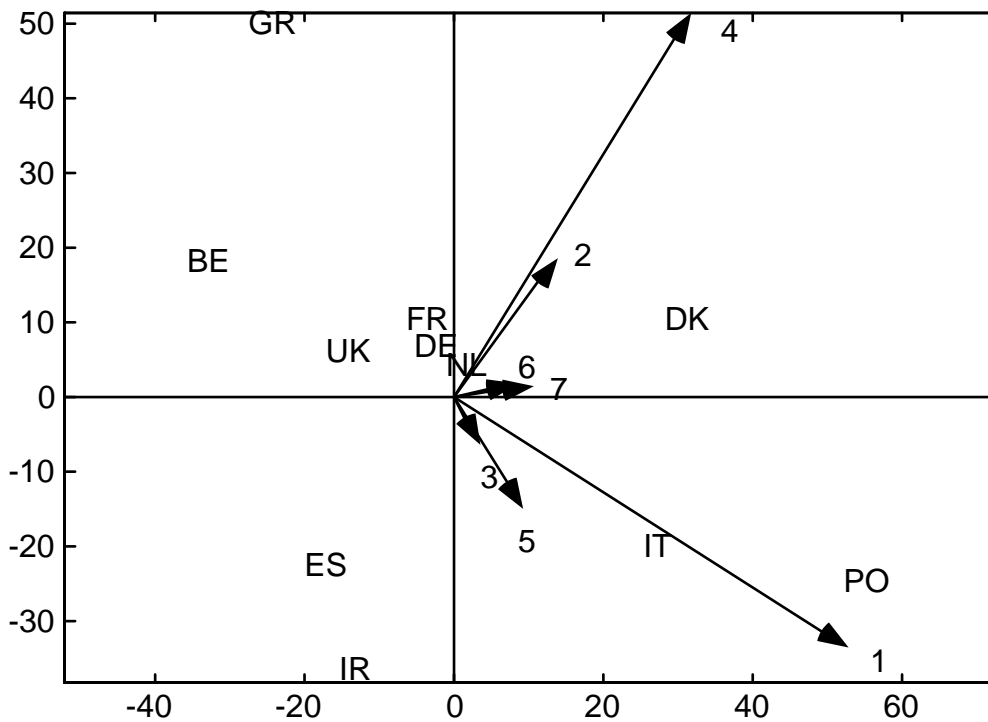


Figura 8.7: Proyección sobre el primer plano principal de países y ramas de actividad: Absorción de inercia en el plano 85%.

INTEGRACION DE LA INFORMACION

Se detallan, a continuación, los resultados encontrados para el espacio conjunto. Si las componentes de los distintos grupos fueran completamente coincidentes, tendríamos solamente $k = 3$ valores propios distintos de cero y todos ellos iguales a $g = 4$, el número de grupos.

Los valores propios fueron los siguientes

3.93	3.83	2.22	1.79	0.13	0.06	0.04
------	------	------	------	------	------	------

Los dos primeros valen prácticamente 4 lo cual quiere decir que existen dos componentes muy similares en los cuatro grupos (años de estudio). La misma información viene dada por los siguientes índices de similaridad de cada componente con los subespacios de los grupos ($\frac{\lambda_i}{g}$) pero con la ventaja de que ahora se trata de una medida acotada entre 0 y 1. Los valores encontrados fueron: 0.9818, 0.9586, 0.5551, respectivamente.

El índice de similaridad global 83.1824.

La estructura de las componentes conjuntas puede verse en los vectores propios que aparecen a continuación

	c1	c2	c3
1	<i>0.8901</i>	-0.2485	0.1089
2	-0.1255	-0.3426	0.0848
3	-0.0920	-0.3943	<i>-0.8747</i>
4	-0.3867	<i>-0.4498</i>	0.3902
5	0.1399	<i>-0.5431</i>	0.0318
6	-0.1203	-0.3640	0.1671
7	-0.0014	-0.1887	0.1860

Tabla 8.16: Vectores propios de la estructura conjunta para los 4 años de estudio.

Los cosenos al cuadrado entre los subespacios de cada uno de los años, que pueden interpretarse como índices de la similitud de cada subespacio con las componentes comunes, acotados entre 0 y 1, aparecen en la tabla siguiente:

Años	Componentes		
	c1	c2	c3
1978	0.9742	0.9798	0.5400
1979	0.9963	0.9730	0.6264
1988	0.9744	0.9255	0.6054
1989	0.9823	0.9559	0.4487

Tabla 8.17: Cosenos al cuadrado de los subespacios de cada año y el espacio de componentes comunes.

Es claro como, de nuevo se pone de manifiesto la existencia de dos componentes comunes.

Ángulos en grados entre cada subespacio y las componentes comunes, pueden ser considerados también como medidas de la similitud entre cada subespacio y las nuevas componentes aunque ya no están acotados entre 0 y 1, no obstante, se incluyen por su interés comparativo.

Años	Componentes		
	c1	c2	c3
1978	9.2444	8.1756	42.7079
1979	3.4829	9.4586	37.6772
1988	9.2025	15.8395	38.9176
1989	7.6375	12.1158	47.9468

Tabla 8.18: Medida de los ángulos entre los subespacios de cada año por separado y el espacio de componentes comunes.

En la figura siguiente se muestra el espacio conjunto con la proyección sobre el mismo de los datos correspondientes a los 4 años en estudio.

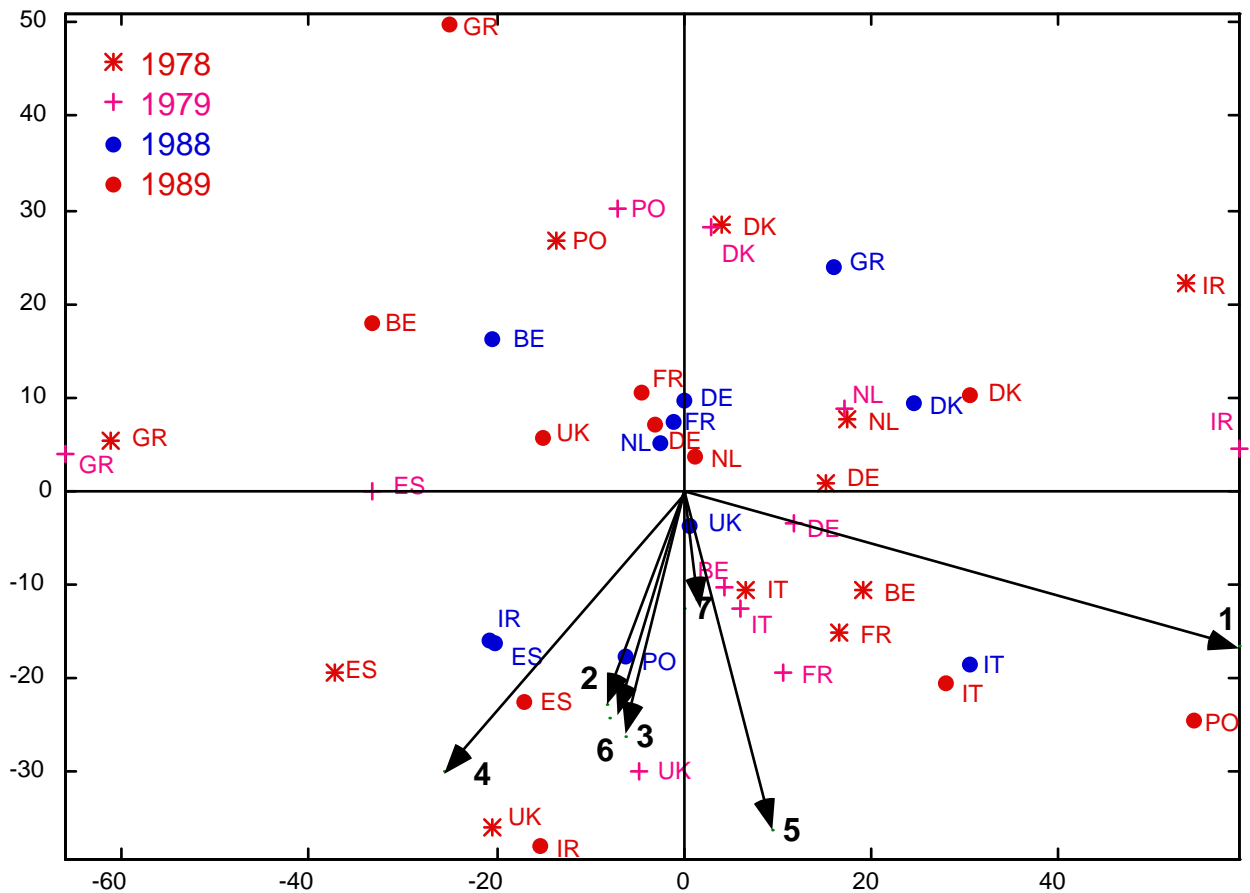


Figura 8.8: Espacio "consenso" para los cuatro años de estudio.

En el gráfico siguiente se han unido con líneas los puntos correspondientes a los cuatro años para el mismo país. Se observa claramente como, a pesar de que se han tomado las dos componentes comunes, algunos de los países presentan posiciones dispares para los puntos de los distintos años.

Muestra, una vez más, las diferencias entre la teoría y la práctica (las discrepancias entre la teoría y la práctica son algo no deseable pero real).

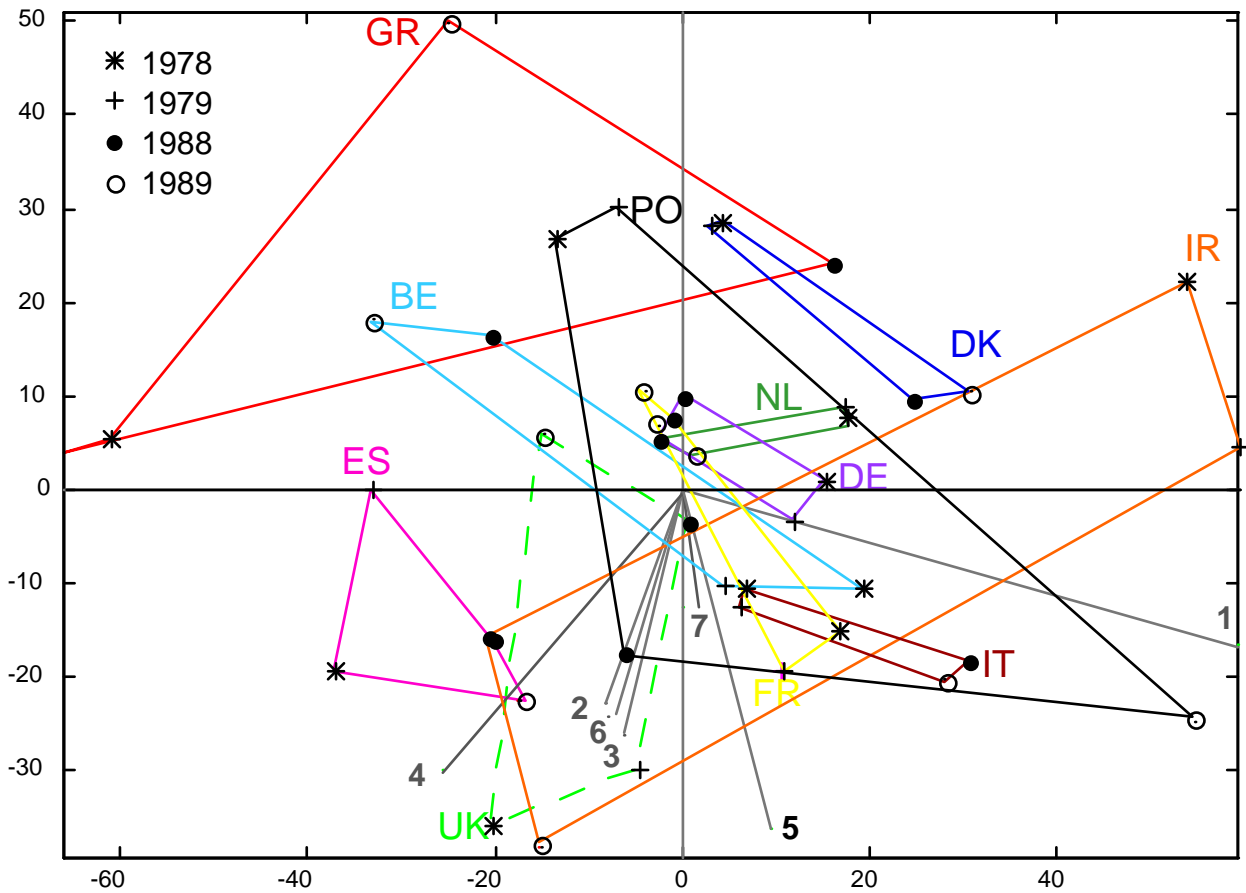


Figura 8.9: Representación de las dos primeras componentes en el espacio "consenso" con líneas que unen los puntos, para los datos de los 4 años, en cada país.

Es posible complementar los resultados anteriores con información suplementaria que puede ayudar en la interpretación de resultados. La información que se calcula a continuación no siempre estará accesible a partir de los datos de las distintas publicaciones, no obstante, dado el estado actual de datos multivariantes quizás estemos a tiempo de explicar a nuestros investigadores los datos que deberían aparecer en las publicaciones y que tendrían un incalculable valor para los investigadores futuros.

Las tablas siguientes muestran la proporción de la variabilidad de cada grupo

absorbida por cada una de las componentes comunes (y la proporción acumulada), la medida puede considerarse útil para saber hasta que punto se pierde información al sustituir los conjuntos de componentes originales por las componentes del espacio integrado.

Años	Componentes Comunes		
	c1	c2	c3
1978	50.5104	21.7788	12.2913
1979	54.8433	19.1946	11.6920
1988	30.7269	22.9110	24.9159
1989	34.5871	29.3811	12.2355

Tabla 8.19: Proporción de la variabilidad de cada grupo absorbida por cada una de las componentes comunes.

Años	Componentes Comunes		
	c1	c2	c3
1978	50.5104	72.2892	84.5806
1979	54.8433	74.0379	85.7299
1988	30.7269	53.6379	78.5538
1989	34.5871	63.9682	76.2037

Tabla 8.20: Proporción acumulada de la variabilidad de cada grupo absorbida por cada una de las componentes comunes.

Es claro que las nuevas componentes absorben una alta proporción de la variabilidad de los datos originales comparable con la que se obtenía en las componentes separadas de cada grupo, es decir, la pérdida de información con las nuevas componentes es muy pequeña.

La tabla siguiente muestra la proporción de la variabilidad global de todos los grupos absorbida por cada componente común. Se trata de una medida de la

absorción de inercia de las componentes comunes obtenidas considerando los cuatro grupos como un único conjunto de datos. Se corresponde con la inercia absorbida por las representaciones de las figuras 8.8 y 8.9.

	Comp. 1	Comp. 2	Comp. 3
Proporción	43.83	23.61	13.98
Proporción acumulada	43.83	67.44	81.42

Tabla 8.21: Proporción acumulada y proporción acumulada de la variabilidad global de todos los grupos absorbida por cada componente común.

Coseno del ángulo entre las componentes originales y las comunes (Similaridad entre las componentes originales de cada grupo y las nuevas componentes).

Año 1978	Componentes comunes		
C. Originales	c1	c2	c3
c1	0.9191	0.2345	-0.2011
c2	-0.1318	0.9312	0.1957
c3	0.3348	-0.2402	0.6792

Tabla 8.22: Coseno del ángulo entre las componentes del año 1978 y las comunes.

Año 1979	Componentes comunes		
C. Originales	c1	c2	c3
c1	0.9876	-0.0282	-0.1286
c2	0.0298	0.9859	0.0318
c3	0.1415	-0.0104	0.7803

Tabla 8.23: Coseno del ángulo entre las componentes del año 1979 y las comunes.

Año 1988	Componentes comunes		
C. Originales	c1	c2	c3
c1	-0.3125	-0.4947	0.6500
c2	-0.9148	-0.0091	-0.2987
c3	0.1998	-0.8251	-0.3060

Tabla 8.24: Coseno del ángulo entre las componentes del año 1988 y las comunes.

Año 1989

Componentes comunes		
---------------------	--	--

C. Originales	c1	c2	c3
c1	0.4998	-0.6708	0.3019
c2	-0.8131	-0.1726	0.3560
c3	0.2674	0.6901	0.4804

Tabla 8.25: Coseno del ángulo entre las componentes del año 1979 y las comunes.

De las tablas anteriores se deduce que, para los dos primeros años en estudio, las dos primeras componentes principales originales son muy similares a las componentes comunes lo que quiere decir que estos dos años presentan una estructura similar. Para el año 1988 son la segunda y la tercera componente, respectivamente, las que se corresponden con la primera y la segunda comunes, de forma que la estructura de las componentes ha cambiado, al menos en el orden de importancia ya que ahora la tercera común se corresponde con la primera original. Para el último año de estudio la segunda componente es similar a la primera común, la segunda común está entre la primera y la tercera original. Esto pone de manifiesto que con ciertas diferencias, la estructura de las componentes en el año 1989 es más similar a la estructura del año 1988 que a la estructura de los otros dos.

A pesar de las diferencias existentes en la estructura de los años, la integración sigue siendo válida ya que, a través de ella mostramos, no sólo las similitudes, sino también las diferencias existentes entre los distintos años.

Las tablas siguientes muestran el ángulo entre las componentes originales y las comunes. La información que proporcionan es básicamente la misma que la medida anterior pero ahora en términos de ángulos.

De la misma manera que antes es claro como, por ejemplo, el ángulo que forman la primera y segunda componentes de los años 1978 y 1979 con las dos

primeras componentes comunes, son pequeños lo que indica que se trata de componentes similares.

Año 1978	Componentes comunes		
C. Originales	c1	c2	c3
c1	23.2044	76.4392	101.5987
c2	97.5728	21.3805	78.7167
c3	70.4415	103.8990	47.2221

Tabla 8.26: Angulo entre las componentes originales del año 1978 y las comunes.

Año 1979	Componentes comunes		
C. Originales	c1	c2	c3
c1	9.0263	91.6181	97.3905
c2	88.2926	9.6174	88.1766
c3	81.8626	90.5973	38.7123

Tabla 8.27: Angulo entre las componentes originales del año 1979 y las comunes.

Año 1988	Componentes comunes		
C. Originales	c1	c2	c3
c1	108.2107	119.6482	49.4588
c2	156.1744	90.5188	107.3823
c3	78.4720	145.5942	107.8167

Tabla 8.28: Angulo entre las componentes originales del año 1988 y las comunes.

Año 1989	Componentes comunes		
C. Originales	c1	c2	c3
c1	60.0150	132.1258	72.4305
c2	144.3972	99.9400	69.1452
c3	74.4933	46.3624	61.2873

Tabla 8.29: Angulo entre las componentes originales del año 1989 y las comunes.

CONCLUSIONES

La exhaustiva revisión bibliográfica nos ha permitido constatar:

1)

a. Las primeras tendencias del Meta-análisis trataban de conocer la información proporcionada por la literatura sobre un tema seleccionado, independientemente de su calidad (GLASS, 1976).

b. Las ideas de ROSENTHAL (1976) y COOPER (1979) incorporan información mediante la acumulación de niveles de probabilidad, y se plantea la problemática del sesgo de publicación.

c. Las últimas tendencias lideradas por STEWART y PARMAR (1993) y CLARKE y STEWART (1994) intentan recuperar los datos individuales procedentes de cada uno de los estudios pertenecientes a la población accesible.

Esta última tendencia, muy prometedora en teoría, choca en la práctica con la falta de colaboración de los investigadores que generaron los datos.

d. Aparece una clara asincronía entre el uso de los métodos estadísticos en investigaciones primarias y los métodos estadísticos utilizados en estudios meta-analíticos. No se ha encontrado ninguna referencia en la que se lleve a cabo una integración de resultados de Análisis Multivariante.

2)

Hemos demostrado la eficacia de los Métodos Biplot en la fase de codificación permitiendo la identificación de estudios homogéneos, juicios discordantes y cuestiones mal planteadas.

3)

El estudio de simulación efectuado a partir de 1000 muestras con distribución normal, para diferentes tamaños del efecto poblacionales y con diferentes tamaños de muestra, puso de manifiesto que la corrección del estimador de HEDGES para evitar el sesgo, está justificado sólo en el caso de que el tamaño del efecto sea grande; sin embargo, cuando los tamaños del efecto son pequeños o moderados, y los tamaños de muestra grandes, el estimador clásico presenta claras ventajas en contra de lo aceptado en la bibliografía.

4)

Se ha demostrado que los Métodos Gráficos Multivariantes y los Métodos Multivariantes Gráficos son una herramienta útil en la fase de Análisis de los Resultados que permiten clasificar los estudios de acuerdo a sus características objetivas y relacionar dicha clasificación con el tamaño del efecto.

5)

Se ha demostrado la eficacia de las teorías de MORGAN y SONQUIST para determinar las causas de la heterogeneidad en el análisis de los resultados del estudio meta-analítico, las cuales son una alternativa más general que la propuesta por HEDGES y OLKIN.

6)

Hemos demostrado que los métodos procrustes propuestos por GOWER y los métodos de comparación e integración de subespacios de KRZANOWSKI pueden ser utilizados como técnicas meta-analíticas de datos multivariantes cuyos resultados vengán expresados en forma de configuraciones euclídeas o en forma de factores latentes, ampliando la metodología clásica que sólo integra vectores de tamaños del efecto.

BIBLIOGRAFIA

ADVANCED OVARIAN CANCER TRIALIST GROUP. (1991). 'Chemotherapy in advanced ovarian cancer: an overview of randomised clinical trial'. *British Medical Journal*, **303**, 884-893.

AGARD, J.B.R; GOBIN, J. y WARWICK, R.M. (1993). 'Analysis of marine macrobenthic community structure in relation to pollution, natural oil seepage and seasonal disturbance in a tropical environment (Trinidad, West Indies)'. *Marine Ecology Progress Series*, **92 (3)**, 233-243.

ALONSO, G. (1982). 'Nuevos métodos de Representación Gráfica de Datos Multivariantes'. *Publicaciones de Bioestadística y Biomatemática.*, **5**. Eunibar. Universidad de de Barcelona.

ANDREWS, D.F. (1972). 'Plots of high dimensional data'. *Biometrika*, **28**, 125-136.

ANGUERA, M.T. y SANCHEZ, P. (1982). 'Meta-análisis: conceptualización y posibilidades en la investigación experimental'. *Comunicación presentada al VII Congreso Nacional de Psicología, Santiago de Compostela, Abril*.

ARMITAGE, P. y BERRY,G. (1992). *Estadística para la Investigación Médica*. (Ed. Española). Doyma, Barcelona.

ARNQVIST, G. y WOOSTER, D. (1995). 'Meta-analysis: synthesizing research findings in ecology and evolution'. *Trends in Ecology & Evolution*, **10**, 236-240.

ASCOMBE, F.J. (1973). 'Sequential clinical trials'. *J. Am. Assoc.*, **58**, 365-384.

AVILA-ZARZA, C.A. (1993). *Métodos Gráficos Multivariantes y su Aplicación en las Ciencias de la Vida*. Tesina de Licenciatura. Universidad de Salamanca.

BAILEY, K.R. (1987). 'Inter-study defferences: How should they influence the interpretation and analysis of results?.'. *Statistics in Medicine*, **6**, 351-358.

BANGERT-DROWNS, R.L. (1986). 'Review of developments in meta-analythic method.'. *Psych. Bull.*, **99 (3)**, 388-399.

BARTLETT, M.S. (1950). 'Tests of signficance in factor Analysis'. *Br. J. Psych. Stat. Sec.* **3**, 77-85.

- BEGG, C.B. y BERLIN, J.A. (1988). 'Publication bias: A problem in interpreting medical data'. *Journal of the Royal Statistical Society, Series A* , **151**, 419-463.
- BENZECRI, J.P. et al. (1973/1982): *L'Analyse des Données. Vol 1: La Taxonomie. Vol 2: L'Analyse des Correspondances*. Dunod. Paris.
- BERLIN, J.A.; LAIRD, N. M.; SACKS, H.S. y CHALMERS, T. (1989). 'A Comparison of statistical methods for combining event rates from clinical trials'. *Statistics in Medicine*, **8**, 141-151.
- BLOOM, B.S. (1964). *Stability and Change in Human Characteristics*. Wiley. New York.
- BORG, I. y LINGOES, J. (1987): *Multidimensional Similarity Structure Analysis*. Springer Verlag. New York.
- BRESLOW, N. (1981). 'Odds ratio estimators when the data are sparse'. *Biometrika*, **68**, 73-84.
- BRESLOW, N.y LIANG, K.Y. (1982). 'The varianza of the Mantel-Haenszel estimator'. *Biometrics*, **38**, 943-952.
- BRUCKNER, L.A. (1978). 'On Chernoff faces'. In Wang, P.C.C. (ed.). *Graphical Representation of Multivariate Data*. Academic Press, pp. 93-121.
- CANNER, P. (1987). 'An overview of six clinical trials of aspirin in coronary heart disease'. *Statistics in Medicine*, **10**, 255-263.
- CARLIN, J.B. (1992). 'Meta-analysis for 2x2 tables: A bayesian approach'. *Statistics in Medicine*, **11**, 141-158.
- CARVER, R.P. (1978). 'The case against statistical significance testing'. *Harvard Educational Review*, **48 (3)**, 378-399.
- CATELL, R.B. (1966). 'The scree test for the number of factors'. *Multivariate Behav. Res.* , **1**: 245-276.

CHALMERS, T.C.; MATTA, R.J.; SMITH, H.J.R; et al. (1977). 'Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction'. *N. Engl. J. Med.*, **297**, 1091-1096.

CHALMERS, T.C.; SMITH, H.J.R; BLACKBURN, B.; SILVERMAN, B.; SCHROEDER, B.; REITMAN, D. y AMBROZ, A. (1981). 'A method for assessing the quality of a randomized control trial'. *Controlled clin. Trials* , **2**, 31-49.

CHALMERS, T.C.; LEVIN, H.; SACKS, H.S.; REITMAN, D.; BERRIER, J. y NAGALINGAM, R. (1987). 'Meta-analysis of clinical trial as a Scientific Discipline I: Control of bias and comparasion with large Co-operative trial'. *Stat. Med.*, **6**, 315-328.

CHALMERS, T.C.; FRANK, C.S. y REITMAN, D. (1990). 'Minimizing the three stages of publication bias'. *Jama*, **263**, 1392-1395.

CHALMERS, T.C. y LAU, J. (1993). 'Meta-analitic stimulus for changes in clinical trials'. *Statistical Methods in Medical Research*, **263**, 161-171.

CHAMBERS, J.M. y KLEINER, B. (1982). 'Graphical Techniques for Multivariate Data and for Clustering'. *Handbook of Statistics*. **2** (Krishnaiah, P.R. y Kanal, L.N.eds.). North-Holland P.C.

CHAMBERS, J.M.; CLEVELAND, W. S.; KLEINER, B. y TUKEY, P.A. (1983). *Graphical methods for Data Analysis*. Ed. Wadsworth. Wadsworth & Brooks/Cole Publishing Company. Statistics/Probability Series. California.

CHERNOFF, H. (1973). 'The use of faces to represent points in k-dimensional space graphically'. *Journal of the American Statistical Association*, **68**, 361-68

CLARKE, M.J. y STEWART, L.A. (1994). 'Obtaining data from randomised controlled trial: how much do we need for reliable and informative meta-analysis?'. *Lancet*, **309**, 1007-1010.

COCHRAN, W.G. (1937). 'Problems arising in the analysis of a series of similar experiments'. *Journal of the Royal Statistical Society*, **4**, 102-118.

COCHRAN, W.G. (1943). 'The comparison of different scales of measurement for experimental results'. *Annals of Mathematical Statistics*, **14**, 205-216

COCHRAN, W.G. (1954). 'The combination of estimates from different experiments'. *Biometrics*, **10**, 101-129.

COCHRAN, W.G. y COX, G.M. (1974). *Diseños Experimentales*. Trillas. México.

COHEN, J. (1962). 'The statistical power of abnormal-social psychological research: A review'. *J. Abn. Soc. Psych.*, **65**, 145-153.

COHEN, J. (1967). *The Statistical Power Analysis for Behavioral Sciences*. Academic Press. N.Y.

COLLINS, R.; GRAY, R.; GODWIN, J. y PETO, R. (1987). 'Avoidance of larger biases and larger random errors in the assessment of moderate treatments effects: The needs for systematic overviews'. *Stat. Med.*, **6**, 245-250.

COLLINS, R.; PETO, R.; MACMAHON, S.; HELBERT, P.; FIEBACH, N.H.; EBERLEIN, K.A.; GODWIN, J.; QIZILBASH, N.; TAYLOR, J.O. y HENNEKENS, C.H. (1990). Blood pressure, stroke, and coronary heart disease. Part 2, short-term reductions in blood pressure: overview of randomised drug trial in their epidemiological context. *Lancet*, **335**, 827-838.

COOPER, H.M. (1979). 'Statistical combining independent studies a meta-analysis of sex differences in conformity research'. *J. Person. Soc. Psych*, **37**, 131-146.

COOPER, H.M (1982). 'Scientific guidelines for conducting integrative research reviews'. *Rev. Educ. Res.*, **52** (2), 291-302.

COOPER, H.M (1984). *The integrative research review: a systematic approach*. Sage. Beverly Hills.

CORNFIELD, J. (1951). 'A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast and cervix'. *J. Natl. Cancer Inst.*, **11**, 1269-1275.

CORNFIELD, J. y HAENSZEL, W. (1960). 'Some aspects of retrospective studies'. *J. Chronic Dis.*, **11**, 523-534.

COURSOL, A. y WAGNER, E.E. (1986). 'Effect of positive findings on submission and acceptance rates: a note on meta-analysis bias'. *Prof. Psych. Res. & Pract.*, **17** (2), 136-137.

COX, D.R. y SNELL, E.J. (1989). *Analysis of Binary Data*.. Chapman and Hall, London. 2ª Ed.

CUADRAS, C.M. (1981/91): *Métodos de Análisis Multivariante*. EUNIBAR. Barcelona.

CURLETTE, W.L. y CANELLA, K.S. (1985). 'Going beyond the narrative summarization of research findings: the meta-analysis approach'. *Res. Nurs. & Health*, **8**, 293-301.

DAVIS, W.J. (1993). 'Contamination of coastal versus open ocean surface waters. A brief meta-analysis'. *Marine Pollution Bulletin*, **26 (3)**, 128-134.

DEAR, K.B.G. y BEGG, C. (1992). 'An approach for assessing publication bias prior to performing a meta-analysis'. *Statistical Science* , **7 (2)**, 237-245.

DELGADO-RODRIGUEZ, M.; SILLERO-ARENAS, M. y GALVEZ-VARGAS, R. (1991). 'Meta-análisis en epidemiología (primera parte): características generales'. *Gac. Sanit.*, **5 (27)**, 265-72.

DELGADO-RODRIGUEZ, M.; SILLERO-ARENAS, M. y GALVEZ-VARGAS, R.(1992). 'Meta-análisis en epidemiología (Segunda parte): metodos cuantitativos'. *Gac. Sanit.*, **6 (28)**, 30-39.

DEMETS, K.R. (1987). 'Methods for combinig randomized clinical trials: Strengths and limitations'. *Statistics in Medicine*, **6**, 341-348.

Der SIMONIAN, R. y LAIRD, N. (1982). 'Evaluating the effectc of coaching on SAT scores: a meta-analysis'. *Harvard Ed Rev.*, **53**, 1-15.

Der SIMONIAN, R. y LAIRD, N. (1986). 'Meta-analysis in clinical trials'. *Controlled Clin. Trials*, **7**, 177-188.

DICKERSIN, K.; CHAN, S.S., CHALMERS, T.C., SACKS, H.S. y SMITH, H. (1987). 'Publication bias and clinical trials'. *Controlled Clinical Trials*, **8**, 343-353.

DICKERSIN, K.; HIGGINS, K. y MEINERT, C.L. (1990). 'Identification of Meta-analysis. The Need for Standard Terminology'. *Controlled Clinical Trials*, **11**, 52-66.

DICKERSIN, K.; MIN, Y.I. y MEINERT, C.L. (1992). 'Factors influencing publication of research results. Follow-up of Applications submitted to two Institutional Review Boards'. *Jama*, **267** (3), 374-379.

DIGBY, P. G. N. & KEMPTON, R. D. (1987). *Multivariate analysis of ecological communities*. Chapman and Hall. London.

DUNNET, C. y TAMHANE, A. (1992). 'A Step-Up Multiple Test Procedure'. *Journal of the American Statistical Association*, **87**, 162-170.

EARLY BREAST CANCER TRIALIST COLLABORATIVE GROUP. (1992). 'Chemotherapy in advanced ovarian cancer: an overview of randomised clinical trials'. *Lancet*, **339**, 1-15; 71-85.

EASTERBROOK, P.J.; BERLIN, J.A.; GOPALAN, R. y MATTHEWS, D.R. (1991). 'Publication bias in clinical research'. *The Lancet*, **337**, 867-872.

EDDY, D.M. (1989). 'The confidence Profile Method: A Bayesian method for assessing health technologies'. *Oper. Res.*, **37**, 210-228.

EDDY, D.M., HASSELBLAD, V. y SHACHTER, R. (1992). *Meta-analysis by the Confidence Profile Method. The Statistical Synthesis of Evidence*. Academic Press. London.

EDGINTON, E.S. (1972a). 'An additive method for combining probability values from independent experiments'. *J. Psych.*, **80**, 351-363.

EDGINTON, E.S. (1972b). 'A normal curve method for combining probability values from independent experiments'. *J. Psych.*, **82**, 85-89.

EJIOU, A. (1990). 'Small-samples properties of Odds ratio estimators under multiple matching in Case-Control studies'. *Biometrics*, **46**, 61-69.

EMERSON, J.D. (1994). 'Combining estimates of the odds ratio: the state of the art'. *Statistical Methods in Medical Research*, **3**, 157-178

EVERITT, B.S. (1978). *Graphical techniques for multivariate data*. Heinemann Educational Books. London.

- EVERITT, B.S. (1992). *The analysis of contingency tables*. Chapman and Hall.
- EVERITT, B.S. (1994). 'Exploring multivariate data graphically: a brief review with examples'. *Journal of Applied Statistics*, Vol **21**, 3.
- FERNANDEZ-DUQUE, E. y VALEGGIA, C. (1994). 'Meta-analysis: A valuable tool in conservation research'. *Conservation Biology*, **8** (2), 555-561.
- FISHER, R.A. (1921). 'On the probable error of a coefficient of correlation deduced from small sample'. *Metron*, **1**, 1-32.
- FISHER, R.A. (1932). *Statistical Methods for Research Workers* . (4 th. ed.). Oliver & Boyd. London.
- FLEISS, J.L. (1969). 'Estimating the magnitud of experimental effects'. *Psych. Bull.*, **72** (4), 273-276.
- FLEISS, J.L. (1981). *Statistical Methods for Rates and Proportions*. Wiley. N.Y.
- FLEISS, J.L. (1987). 'Discussion contribution to light'. *RJ. Statistics in Medicine* , **6**, 221-228.
- FLEISS, J.L. (1993). 'The statistical basis of Meta-analysis'. *Statistical Methods in Medical Research*, **2**, 121-145.
- FLEISS, J.L. y GROSS, A.J. (1991). 'Meta-analysis in epidemiology, with special reference to studies of the association between exposure to enviromental tobacco smoke and lung cancer: a critique'. *Journal of Clinical Epidemiology*, **44**, 127-139
- FRENI, S.C. (1994). 'Exposure to high fluoride concentrations in drinking water is associated with decreased birth rates'. *Journal of Toxicology and Environmental Health*, **42** (1), 109-121.
- FRIEDMAN, H. (1978). 'Magnitude of experimental effect and a table for its rapid estimation'. *Psych. Bull.*, **70** (4), 245-251.
- GABRIEL, K.R. (1971). 'The biplot graphic display of matrices with application to principal component analysis'. *Biometrika*, **58**, 453-467.

GABRIEL, K.R. (1981). 'Biplot display of multivariate matrices for inspection of data and diagnosis In interpreting multivariate data'. Barnett, V.

GALBRAITH, R.F. (1988a). 'Graphical display of estimates having differing standard Errors'. *Technometrics*, **30** (3), 271-281

GALBRAITH, R.F. (1988b). 'A note on graphical presentation of estimated odds-ratios from several clinical trials'. *Statistics in Medicine*, **7**, 889-894.

GALINDO, M.P. (1986). 'Una alternativa de representación simultánea: HJ-Biplot'. *Questiio*, **10** (1), 13-23.

GALINDO, M.P. (1991). 'Meta-analysis: El uso de los métodos estadísticos en revisiones de estudios de investigación relacionados'. *Citugía Española*, **54**, 201-203.

GALINDO, M.P. y CUADRAS, C.M. (1986). 'Una extensión del método Biplot y su relación con otras técnicas'. *Publicaciones de Bioestadística y Biomatemática*. Un. de Barcelona. Nº 17.

GART, J.J. (1992). 'Pooling 2x2 Tables: Asymptotic Moments of Estimators'. *J.R.Statist. Soc. B*, **54** (2), 531-539.

GELBERT, R.D.; GOLDBIRSCHE, A. (1991). 'Meta-analysis: The fashion of summing-up evidence. Part I: Rational and Conduct'. *Annals of Oncology.*, **2**, 461-468.

GEORGE, E.O. (1977). *Combining Independent one-side and two side Statistical test: Some Theory and Applications*. Doctoral Disertation, Univ. Rochester. (Cita en HEDGES y OLKIN, 1985, pp. 40)

GHISELLI, E.E. (1949). 'The validity of commonly employed occupational tests'. *University of California Publications in Psychology*, **5**, 253-288.

GHISELLI E, E. (1955). 'The measurement of occupational aptitude'. *University of California Publications in Psychology*, **8**, 101-216

GIACONIA, R.M. y HEDGES, L.V. (1983). 'Identifying features on effective open education' original manuscript. Sage Publications, Inc.

- GLASS, G.V. (1976). 'Primary, secondary and meta-analysis of research'. *Educ. Res.*, **5**, 3-8.
- GLASS, G.V. (1978). 'In defense of generalization'. *Beh. Br. Sciences*, **3**, 394-395.
- GLASS, G.V.; Mc GAW, D. y SMITH, M.L. (1981). *Meta-analysis in Social Research..* Sage. Beverly-Hills, CA.
- GOMEZ, J. (1987). *Meta-analysis*. Promociones y Publicaciones Universitarias, S.A. Barcelona.
- GOMEZ, J.; ARTES, M. y SEGU, J.L. (1989). 'Importancia del meta-análisis como técnica cuantitativa de revisión de estudios en ciencias de la salud'. *Met. Clin.*, **93**, 295-297.
- GOOD, I.J. (1969). 'Some applications of the singular decomposition of a matrix'. *Technometrics*, **11**, 823-831.
- GOWER, J.C. (1975). 'Generalized Procrustes Analysis'. *Psychometrika*, **40**, 33-51.
- GREEN, B.F. y HALL, J.A. (1984). 'Quantitative methods for literature reviews'. *Ann. Rev. Psych.*, **35**, 37-53.
- GREENLAND, S. (1987). 'Quantitative methods in the review of epidemiologic literature'. *Epidemiologic Reviews*, **9**, 1-30.
- GREENLAND, S. y SALVAN, A. (1990). 'Bias in the one-step method for pooling study results'. *Statistics in Medicine*, **9**, 247-252.
- GUREVITCH, J.; MORROW, L.L.; WALLACE, A. y WALSH, J.S.(1992). 'A meta-analysis of competition in field experiments'. *American Naturalist*, **140** (4), 539-572.
- HATTIE, J. A. y HANSFORD, B.C. (1984). 'Meta-analysis: a reflection of problems'. *Aust. J. Psych.*, **36** (2), 239-254.
- HAUCK, W.W. (1984). 'A Comparative study of Conditional Maximun likelihood estimation of a common odds ratio'. *Biometrics*, **40**, 1117-1123.
- HEDGES, L.V. (1981). 'Distribution theory for Glass's estimator of effect size and related estimators'. *Journal of Educational Statistics*, **6**, 107-128.

- HEDGES, L.V. (1982). 'Estimation of effect size from a series of independent experiments'. *J. Psy. Bull.*, **92** (2), 490-499.
- HEDGES, L.V. (1984 a). 'Research synthesis: the state of the art'. *Int. Journal of Aging and Human Development*, **19** (2), 85-93.
- HEDGES, L.V. (1984 b). 'Estimation of effect size under nonrandom sampling: the effects of censoring studies yielding statistically insignificant mean differences'. *Journal of educational Statistics*, **9**, 61-85.
- HEDGES, L.V. (1990). 'Directions for future methodology'. in Watcher K. W. & Straf M. L. (eds) *The Future of Meta-analysis*, Russell Sage Foundation, New York págs. 3-10.
- HEDGES, L.V. y OLKIN, I. (1983 a). 'Clustering estimates of effects magnitude from independent studies'. *Psych. Bull.*, **93** (3), 563-573.
- HEDGES, L.V. y OLKIN, I. (1983 b). 'Regression models in research synthesis'. *Amer. Stat.*, **37** (2), 137-140.
- HEDGES, L.V. y OLKIN, I. (1984). 'Nonparametric estimators of effect size'. *Psy. Bull.*, **96** (3), 573-580.
- HEDGES, L.V. y OLKIN, I. (1985). *Statistical Methods for Meta-analysis*. Academic Press.
- HEYL, P.R. (1930). 'Redetermination of the constant of gravitation'. *National Bureau of Standards Journal of Research*, **5**, 1243-1290.
- HOLLAND, B. y DIPONZIO, M. (1987). 'An Improved Sequentially Rejective Bonferroni Test Procedure'. *Biometrics*, **43**, 417-423.
- HOLM, S. (1979): 'A Simple Sequentially Rejective Multiple Test Procedure'. *Scandinavian Journal of Statistics*, **6**, 65-70.
- HOMMEL, G. (1989). 'A Comparaison of Two Modified Bonferroni procedures'. *Biometrika*, **76**, 624-625.

HUFF, D.L. y BLACK, W. (1978). 'A Multivariate Graphic Display for Regional Analysis'. In Pcc. Nang (Ed.) *Graphical Representation of Multivariate Data*. pp: 199-218. Academic Press. New York.

HUNTER, J.E. y SCHMIDT, F.L. (1978). 'Diferential and single-group validity of employment test by race: A critical analysis of three recent studies'. *Journal of Applied Psychology*, **63**, 1-11.

HUNTER, J.E.; SCHMIDT, F.L. y JACKSON, G.B. (1982). *Meta-analysis: Cumulating Research Finding Across Sudies*. Sage. Beverly-Hills.

HUNTER, J.E. y SCHMIDT, F.L. (1990). *Meta-analysis: Correcting Error and Bias in Research Finding*. Sage Publications.

ISBELL, L.A. y YOUNG, T.P. (1993). 'Social and ecological influences on activity budgets of vervet monkeys and their implications for group living'. *Behavioral Ecology and Sociobiology*, **32 (6)**, 377-385

IYENGAR, S y GREENHOUSE, J.B. (1988). 'Selection models and the file drawer problem'. *Statist. Sci.*, **3**, 109-135.

JARVINEN, A. (1991). 'A Meta-analysis study of the effects of female age on laying clutch-size in the Great Tit *Parus major* and the Pied Flycatcher *Ficedula hypoleuca*'. . *Ibis*, **133**, 62-66.

JENICEK, M. (1987). *Meta-analyse en Médecine. Evaluation et Shynthèse de l'information Clinique et Epidémiologique*. St. Hyacinthe et Paris. Edisen et Maloine.

JENICEK, M. (1989). 'Meta-analysis in Medicine. Where we are and where we want to go'. *J. Clin. Epidemiol.*, **42**, 35-44.

JOLLIFFE, I.T. (1986). *Principal Component Analysis*. Springer-Verlag. New York.

JONES, D.R. (1992). 'Meta-analysis of observational epidemiological studies: a review'. *Journal of the Royal or Medicine*, **85**, 165-168.

JONES, D.R. (1995). 'Meta-analysis: weigthing the evidence'. *Statistics in Medicine*, **14**, 137-149.

KASS, G.V. (1980). 'Significance Testing in Automatic Interaction Detection'. *Applied Statistics*, **24**, 178-189.

KRAEMER, H.C. (1983). 'Theory of estimation and testing of effect sizes: Use in meta-analysis'. *Journal of Educational Statistics* , **8**, 93-101.

KRAEMER, H.C. y **ANDREWS, G.** (1982). 'A non-parametric technique for meta-analysis effect size calculation'. *Psych. Bull.*, **91** (2), 190-192.

KRZANOWSKY, W.J. (1979). 'Between groups comparison of principal components'. *J. Amer. Statist. Asoc.*, **74**, 703-707 (corrección en **76**, 1022).

KRZANOWSKY, W.J. (1988). *Principles of Multivariate Analysis..* Oxford Science Publications.

KULIK, C.L. y **KULIK, C.** (1982). 'Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings'. *American Educational Research Journal* , **19**, 415-428.

KUTTY, V.R. (1991). 'A tool for better understanding of medical research'. *Current Science*, **61** (2), 71-72.

L'ABBE, K.A.; **DETSKY, A.S.;** **O'ROURKER, K.** (1987). 'Meta-analysis in clinical research'. *Ann. Intern. Med.*, **107**, 204-223.

LAIRD, N.M., **MOSTELLER, F.** (1990). 'Some statistical methods for combining experimental results'. *Int. J. Tech. Assess Health Care.*, **6**, 5-30.

LIANG K, Y. (1985). 'Odds-ratio inference with dependent data'. *Biometrika*, **72**, 678-682.

LIGHT, R.J. (1979). 'Capitalizing on variation: How conflicting research findings can be helpful for policy'. *Educational Researcher*, **8** (3), 3-8.

LIGHT, R.J. (1987). 'Accumulating evidence from independent studies: what we can win and what we can lose'. *Statistics in Medicine*, **6**, 221-228.

LIGHT, R.J. y **SMITH, P.V.** (1971). 'Accumulating evidence: procedures for resolving contradictions among different studies'. *Harv. Educ. Rev.*, **41**, 1-26.

LIGHT, R.J. y PILLEMER, D.B. (1982). 'Numbers and narrative: combining their strengths in research reviews'. *Harv. Educ. Rev.*, **52**, 1-26.

LIGHT, R.J. y PILLEMER, D.B. (1984). *Summing Up the Science of Reviewing Research..* Harvard University Press.

MacMAHON, S.; PETO, R. y CUTLER, J. (1990). 'Blood pressure, stroke and coronary heart disease. I. Prolonged differences in blood pressure. Prospective observational studies corrected for the regression dilution bias. *Lancet*, **I**, 765-773.

MANN, C. (1990). 'Meta-analysis in the breech'. *Science*, **249**, 476-478.

MANSFIELD, R.S. y BUSSE, T.V. (1977). 'Meta-analysis in research: A rejoinder to Glass'. *Educ. Researcher*, **6**, 3.

MANTEL, N. (1977). 'Test and limits for the common odds-ratio of several 2x2 contingency tables: Methods in analogy with the Mantel-Haenszel procedure'. *J. Stat. Plann. Inf.*, **1**, 179-189.

MANTEL, N. y HAENSZEL, W. (1959). 'Statistical aspects of the analysis of data from retrospective studies of disease'. *Journal of the National Cancer Institute*, **22**, 719-748.

MANTEL, N. y HANKEY, B.F. (1975). 'The odds-ratio of a 2x2 contingency table'. *Biometrics*, **23**, 143-145.

MARDIA, K.V.; KENT, J.T. y BIBBY, J.M. (1979). *Multivariate Analysis*. Academic Press. London.

MARCHANT, L.F. y MacGREW, W.C. (1991). 'Laterality of function in apes: a meta-analysis of methods'. *Journal Of Human Evolution* , **21 (6)**, 425-438.

MESSENGER, R. y MANDELL, L. (1972). 'A model search technique for predictive nominal scale multivariate analysis. *Journal of the American Statistical Association*, **67**, 768-772.

MESSICK, S. y JUNGEBLUT, A. (1981). 'Time and method coaching for the SAT'. *Psychological bulletin*, **89 (2)**, 191-216.

- MIETTINEN, O.S. (1976). 'Estimability and estimation in case-referent studies'. *Am. J. Epidemiol*, **103**, 226-235.
- MOERTEL, C.G. (1984). 'Improving the efficiency of clinical trials: a medical perspective'. *Statistist in Medicine*, **3**, 455-466.
- MORGAN, J. y SONQUIST, J. (1963). 'Problems in the Analysis of Survey Data and a Proposal'. *Journal of the American Statistical Association*, **58**, 415-434.
- MOSTELLER, F. y BUSH, R.R. (1954). 'Selected quantitative techniques'. In G. Lindzey (ed.). *Handbook of Social Psychology: Theory and Method*. Vol. I (pp. 289-334). Addison-Wesley.
- MOSTELLER, F. y CHALMERS, T.C. (1992). 'Some progress and problems in meta-analysis of clinical trials'. *Statistical science*, **7 (2)**, 227-236.
- MULLEN, B. (1989). *Advanced BASIC Meta-analysis*. Hillsdate, NJ. Erlbaum.
- MULLEN, B. y ROSENTHAL, R. (1985). *Basic Meta-Analysis: Procedures and Programmes*. Erlbaum. U.S.
- NAYLOR, C.D. (1988). 'Two cheers for meta-analysis: Problems and opportunities in aggregating result of clinical trials'. *Can. Med. Assoc. J.*, **138**, 891-895.
- OAKES, M. (1993). 'The logic and role of meta-analysis in clinical research'. *Statistical Methods in Medical Research*, **2**, 146-158.
- OLKIN, I. (1990). 'History and goals' in Watcher K. W. y Straf M. L. (eds) *The future of Meta-analysis*, Russell Sage Foundation, New York. pp. 3-10.
- ORWIN, R.G. (1983). 'A fail-safe N for effects size in meta-analysis'. *J. Educ. Stat.*, **8 (2)**, 157-159.
- PAUL, S.R. y DONNER, A. (1992). 'Small sample performance of test of homogeneity of odds ratios in K 2x2 tables'. *Statistics in Medicine*, **11**, 159-165.
- PEARSON, K. (1901). 'On lines and planes of closest fit to systems of points in space'. *Phil. Mag.*, **2 (6)**, 559-572.

PEARSON, K. (1933). 'On a method of determining whether a samples of given size n supposed to have been draw from a parent population having a known probability integral has probably been drawn at random'. *Biometrika*, **25**, 379-410.

PEECE, P. F. W. (1983). 'A measure of experimental effect size based on success rate'. *Educ. Psych. Meas.*, **43**, 763-766.

PETITTI, D.B. (1994). *Meta-analysis, Decision Analysis and Cost-effectiveness Analysis. Methods for quantitative synthesis in Medicine*. Oxford University Press. New York.

PETO, R. (1987 a). 'Discussion during the workshop on methodological issues in overviews of randomized clinical trials. May 1976. **6**, 229.

PETO, R. (1987 b). 'Why do need "Systematic overviews of radomizer Trials?'. *Stat Med.*, **6**, 233-240.

PETO, R. y PETO, J. (1972). 'Asymptotically efficient rank invariant test procedures'. *Journal of the. Royal. Statistical Society (A)*, **135**, 185-198.

PETO, R.; PIKE, M.C.; ARMITAGE, P.; BRESLOW, N. E.; COX, D.R.; HOWARD, S.V.; MANTEL, N.; McPHERSON, K.; PETO, J. y SMITH, P. G. (1977). 'Design analysis of radommized clinical trials requiring prolonged observation of each patient. II. Analysis and examples'. *Br. j. Cancer*, **35**, 1-39.

PIEDBOIS, P.; BUYSE, M. et LE BOURGEOIS, J.P. (1992). 'La méta-analyse, instrument de synthèse et d'observation'. *Bull. cancer* , **79**, 951-956.

PIGEOT, I. (1990). 'A class of asymptotically efficient noniterative estimator of a common odds ratio'. *Biometrika*, **77**, 420-423.

PIGEOT, I. (1991). 'A Jackknife estimator of a combined Odds ratio'. *Biometrics*, **47**, 373-381.

PILLEMER, D.B. y LIGHT, R.J. (1982). 'Synthesizing outcomes: how to use research evidences from many studies'. *Harv. Educ. Rev.*, **50 (2)**, 176-195.

POCOKC, S.J; SHARPER, A.G; PHILLIPS, A.N; WALKER, A.N.y WHITEHEAD, T.P. (1986). 'High density lipoprotein cholesterol is not a major risk factor for ischaemic heart disease in British men'. *Brit. Med. J.*, **292**, 515-519.

POCOCK, S.T. y HUGHES, M.D. (1990). 'Estimation issues in clinical trials and overviews'. *Statistics in Medicine*, **9**, 657-671.

POULIN, R. (1994). 'Meta-analysis of parasite-induced behavioural changes'. *Animal Behaviour*, **48** (1), 137-146.

QUENOUILLE, (1949). 'Aproximate tests of correlation in time series'. *Journal of the Royal Statistical Society, Series B*, **11**, 68-84.

RACHMAN, S.J. y WILSON, G.T. (1980). *The Effects of Psychological Therapy*. Pergamon Press. New York.

RAMIREZ, G.J. (1995). *Contribuciones al Análisis de Segmentación*. Tesis Doctoral. Universidad de Salamanca.

ROBINS, J.M.; BRESLOW, N.E. y GREENLAND, S. (1986). 'Estimators of the Mantel-Haenszel variance consistent in the both sparse data and large-strata limiting models'. *Biometrics*, **42**, 311-323.

ROSENBERGER, J.L. (1990). 'Design of experiments for evaluating frozen sperm'. *Journal of Andrology*, **11** (2), 89-96.

ROSENTHAL, R. (1976). 'Interpersonal expectancy effects'. In R. Rosenthal, *Experimental effects in behavioral research*. New York: Irvington. pp. 440-471.

ROSENTHAL, R. (1978). 'Combining results of independent studies'. *Psych. Bull.*, **85** (1), 185-193.

ROSENTHAL, R. (1979). 'The "file drawer problem" an tolerance for null results'. *Psych. Bull.*, **86** (3), 638-641.

ROSENTHAL, R. (1980). *New Direction for Methodology of Social and Behavioral Science*. Vol. V. Jossey-Bass. San Francisco.

ROSENTHAL, R. (1983). 'Meta-analysis: toward a more cumulative social science'. In L. BICKMAN, (Ed.). *Applied Social Psychology Annual*. Vol. **IV**. pp. 65-93. Sage.

ROSENTHAL, R. (1984). *Meta-analytic procedures for social research..* Sage. Beverly Hills.

ROSENTHAL, R. y RUBIN, D.B. (1979). 'Comparing significance levels of independent studies'. *Psych..Bull.*, **86** (5), 1165-1168.

ROSENTHAL, R. y RUBIN, D.B. (1982 a). 'Further meta-analytic procedures for assessing cognitive gender differences'. *J. Educ. Psych.*, **74** (5), 707-712.

ROSENTHAL, R. y RUBIN, D.B. (1982 b). 'Comparing effect sizes of independent studies'. *Psychological Bulletin*, **92**, 638-641.

ROSENTHAL, R. y RUBIN, D.B. (1986). 'Meta-analytic procedures for combining studies with multiple effect sizes'. *Psych. Bull.*, **99** (3), 400-406.

RUST, R.T., LEHMANN, D.R. y FARLEY, J.U. (1990). 'Estimating Publication Bias in Meta-analysis', *Journal of Marketing Research*, **27**, 220-226.

SACKS, H.S; CHALMERS, T.C; BERK, A.A. y REITMAN, D. (1985). 'Should mild hypertension be treated? An attempted meta-analysis of the clinical trials'. *Mt Sinai J. Med.*, **52**, 265-270.

SACKS, H.S.; BERRIER, J.; REITMAN, D.; ANCONA-BERK, V.A. y CHALMERS, T.C. (1987). 'Meta-analysis of randomized controlled trials'. *N. Engl. J.Med.*, **316**, 450-455.

SAN MARTIN, R. y PARDO, A. (1989). *Psicoestadística. Contrastes Paramétricos y no Paramétricos*. Ediciones Pirámide. Madrid.

SANER, H. (1994). 'A conservative inverse normal test procedure for combining p-values in integrative research'. *Psychometrika*, **59**, 253-267.

SAWKA, C.A. y PRITCHARD, K.I. (1992). 'Meta-analysis: Its application to clinical medicine'. *Diagn. Oncol*, **2**, 80-89.

SEBER, G.A.F. (1984). *Multivariate Observations*. Wiley. New York.

SERLIN, R.C.; CARR, J. y MARASCUILO, L.A. (1982). 'A measure of association for selected nonparametrics procedures'. *Psych. Bull.*, **92** (3), 786-790.

SIMES, R.J. (1986 a). 'An Improved Bonferroni Procedure for Multiple Test of Significance'. *Statistics in Medicine*, **6**, 11-30.

SIMES, R.J. (1986 b). 'Publication bias: the case for an International Registry of Clinical Trials'. *J. Clin. Oncol.*, **4**, 1529-1541.

SIMES, R.J. (1986 c). 'Confronting publication bias: a cohort design for meta-analysis'. *Biometrika*, **73**, 751-754.

SMITH, M.L. (1980). 'Publication bias and meta-analysis'. *Eval. Educ.*, **4**, 22-24.

SMITH, M.L.; GLASS, G.V. y MILLER, T.I. (1980). *The Benefit of Psychotherapy*. Baltimore University Press.

SNEDECOR, G.W. y COCHRAN, W.G. (1971). *Métodos Estadísticos*. The Iowa State University Press. 3ª imp.

STEWART, L.A. y PARMAR, M.K.B. (1993). 'Meta-analysis of the literature or of individual patient data: is there a difference?'. *Lancet*, **341**, 418-422.

STJERNSWARD, J. (1974). 'Decreased survival related to irradiation of data from retrospective studies of disease'. *J. Natl. Cancer Inst.*, **22**, 1258-1286.

STOCK, W.A.; OKUN, M.A.; HARING, M.J. y WITTER, R. A. (1983). 'Age differences in subjective well-being: A meta-analysis'. Original manuscript in *Evaluation studies. Review annual..* Sage Publications, Inc. pp. 279-302.

STOUFFER, S.A. ; SUCHMAN, E.A.; De VINNEY, L.C.; STAR, S.A. y WILLIAMS, R.M. (1949). *The American Soldier*. Vol. I. Princeton, N.J. Princeton University Press.

STRAUSS, S. (1992). 'The study of studies (meta-analysis)'. *Technology Review*, **95**, 18-19.

STRUBE, M.J. (1985). 'Combining and comparing significance levels from non independent hypothesis tests'. *Psych. Bull.*, **92** (2), 334-341.

- STRUBE, M.J. y HARTMANN, D.P. (1983.). 'A critical appraisal of meta-analysis'. *Clin. Psych.*, **21**, 129-139.
- STRUBE, M.J. y HARTMANN, D.P. (1984). 'Meta-analysis: techniques, applications and functions'. In R.F. CONNER et al. (Eds.). *Evaluation Studies Review Annual*, **9**. Sage. pp. 286-299.
- STRUBE, M.J. y MILLER, R.H. (1986). 'Comparison of power rates for combined probability procedures: A simulation study'. *Psych. Bull.*, **99** (3), 407-415.
- TAVEGGIA, T.C. (1974). 'Resolving research controversy through empirical cumulation'. *Soc. Methods Res.*, **2**, 395-407.
- THOMPSON, S.G. y POCOCK, S.J. (1991). 'Can meta-analysis be trusted?'. *Lancet*, **338**, 1127-1130.
- TIPPETT, L.H. (1931). *The Method of Statistics*. Williams and Norgate. London.
- TONHASCA, A. y BYRNE, D.N. (1994). 'The effects of crop diversification of herbivorous insects: A meta-analysis approach'. *Ecological Entomology* , **19** (3). 239-244.
- TUKEY, J.W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
- VANDERWERF, E. (1992). 'Lack's clutch size hypothesis: an examination of the evidence using meta-analysis'. *Ecology*, **73** (5), 1699-1705.
- VAN HOUWELINGEN, H.C.; ZWINDERMAN, K.H. y STIJNEN, T. (1993). 'A bivariate approach to Meta-analysis'. *Statistics in Medicine*, **12**, 2273-2284.
- VICENTE-TAVERA, S. (1992). *Las técnicas de representación de datos multidimensionales en el estudio del índice de producción industrial en la C.E.E.* Tesis Docoral. Universidad de Salamanca.
- WALBERG, H.J. y HAERTEL, E.H. (1980). 'Research integration: an introduction and overview'. *Evaluation in Education*, **4**, 5-10.
- WACHTER, K.W.(1988). 'Disturbed by meta-analysis?'. *Science*, Vol. **249**, 1407-1408.

- WARWICK, R.M. y CLARKE, K.R. (1993). 'Comparing the severity of disturbance: a meta-analysis of marine macrobenthic community data'. *Marine Ecology Progress Series*, **92** (3), 221-231.
- WHITEHEAD, A. y WHITEHEAD, J. (1991). 'A General parametric approach to the Meta-analysis of randomized clinical trial'. *Statistic medicine*, **10**, 1665-1677.
- WILKINSON, B. (1951). 'A statistical consideration in psychological research'. *Psychological Bulletin*, **48**, 156-158.
- WILLIAMS, P.A. y HAERTEL, E.H. (1982). 'The impact of leisure-time television on school learning: A research synthesis'. *American Educational Research Journal.*, **19** (1), 19-50.
- WINER, B.J. (1971). *Statistical Principles in Experimental Design*. Mac. Graw Hill (2nd ed.).
- WOLF, F.M. (1986). *Meta-analysis: Quantitative Methods for Research Synthesis. Quantitative Applications in Social Sciences*. Sage Publications.
- WOLF, F.M. (1990). 'Methodological observations on bias'. In Watcher K. W. y Straf M. L. (Eds) *The Future of Meta-analysis*, Russell Sage Foundation, New York. pp. 3-10
- WORSLEY, K.J. (1977). 'A Non Parametric Extension of a Cluster Analysis Method by Scott and Knott'. *Biometrics*, **33**, 532-535.
- YATES, F. y COCHRAN, W.G. (1938). 'The analysis of groups of experiments'. *Journal of Agricultural Science*, **28**, 556-580.
- YUSUF, S.; PETO, R.; LEWIS, J.; COLLINS, R. y SLEIGHT, P. (1985). 'Beta blockade during and after myocardial infarction: an overview of the randomized trials'. *Progress in Cardiovascular Disease*, **27**, 335-371.
- YUSUF, S.; SIMON, R.; ELLEMBERG, S. (1987). 'Proceedings of Methodologic issues in overviews of randomized clinical trials'. *Stat. Med.*, **6**, 217-409.