

DEPARTAMENTO DE ESTADÍSTICA

**MÁSTER EN ANÁLISIS AVANZADO
DE DATOS MULTIVARIANTES**



**VNiVERSiDAD
D SALAMANCA**

TRABAJO FIN DE MÁSTER

Detección del funcionamiento diferencial
de los ítems en el test MMSE en la
población española

ARMANDO GONZÁLEZ SÁNCHEZ

**TUTOR:
JOSÉ LUIS VICENTE VILLARDÓN**

Salamanca, julio de 2015

UNIVERSIDAD DE SALAMANCA
Departamento de Estadística
Máster en Análisis Avanzado de Datos Multivariantes
Trabajo Fin de Máster

Detección del funcionamiento diferencial de los ítems en el test MMSE en la población española

En función de las variables edad, sexo y nivel educativo

Autor: Armando González Sánchez

Tutor: José Luis Vicente Villardón

2015



Dpto. de Estadística
Universidad de Salamanca

JOSÉ LUIS VICENTE VILLARDÓN

Profesor del Departamento de Estadística de la Universidad de Salamanca

CERTIFICA que don Armando González Sánchez ha realizado en la Universidad de Salamanca, bajo su dirección, el trabajo que para optar al título de Máster en Análisis de Datos Multivariantes presenta con título **Detección del funcionamiento diferencial de los ítems en el test MMSE en la población española**, autorizando expresamente su lectura y defensa.

Y para que conste, firma el presente certificado en Salamanca a 13 de julio de 2015.

José Luis Vicente Villardón

DEDICATORIA

Dedico el fruto de este esfuerzo a mi madre,
principal patrocinadora de mis actividades académicas.

Con cariño, tu hijo.

Tabla de contenido

1.- INTRODUCCIÓN.....	8
2.- MINI MENTAL STATE EXAMINATION.....	9
2.1- EXTENSIÓN DEL MINI MENTAL.....	10
2.2.- EN QUÉ CONSISTE EL MINI MENTAL.....	11
2.3.- VERSIONES.....	11
2.4.- VENTAJAS.....	13
2.5.- EL TEST NO SIEMPRE RESPONDE COMO SE ESPERA.....	14
2.6.- NO NORMALIDAD.....	15
2.7.- EDUCACIÓN.....	16
3.- MÉTODO.....	17
3.1.- TEORÍA CLÁSICA DE LOS TEST.....	17
3.2.- TEORÍA DE RESPUESTA AL ÍTEM.....	20
3.2.1.- SUPUESTOS DE LA TRI.....	24
3.3.- ANÁLISIS DEL FUNCIONAMIENTO DIFERENCIAL DE LOS ÍTEMS.....	25
3.4.- REGRESIÓN LINEAL.....	27
3.5.- REGRESIÓN LOGÍSTICA.....	29
3.6.- REGRESIÓN LOGÍSTICA BINARIA.....	29
3.7.- RECORRIDO HISTÓRICO DE LA TÉCNICA.....	34
3.8.- EL PROBLEMA DE LA SEPARACIÓN.....	37
3.8.1.- DIF NO UNIFORME CON SPSS.....	39
3.8.2.- DIF UNIFORME CON SPSS.....	40
3.8.3.- PROPUESTAS PARA EL PROBLEMA DE LA SEPARACIÓN.....	44
4.- MUESTRA.....	48
5.- METODOLOGÍA.....	49
6.- RESULTADOS.....	50
6.1- RESULTADOS CON LA VARIABLE EDUCACIÓN.....	51
6.1.1.- DIF NO UNIFORME Y EDUCACIÓN.....	55
6.1.2.- DIF UNIFORME Y EDUCACIÓN.....	56
6.2.- RESULTADOS CON LA VARIABLE EDAD.....	56
6.2.1.- DIF NO UNIFORME Y EDAD.....	58
6.2.2.- DIF UNIFORME Y EDAD.....	58
6.3.- RESULTADOS CON LA VARIABLE SEXO.....	58
6.3.1.- DIF NO UNIFORME Y SEXO.....	59
6.3.2.- DIF UNIFORME Y SEXO.....	60
7.- CONCLUSIONES.....	60
8.- AGRADECIMIENTOS.....	61

9.- REFERENCIAS BIBLIOGRÁFICAS	62
10.- APÉNDICE	71

Resumen: El Mini Mental State Examination es un test avalado y muy utilizado que se usa para detectar deterioro cognitivo. Aunque ha sido ampliamente estudiado, no lo ha sido tanto en referente a su funcionamiento diferencial en la población española. Este trabajo tiene por objetivo detectar el funcionamiento diferencial de los ítems de este test en función a tres variables propuestas: nivel educativo, edad y sexo. Para ello se utiliza una regresión logística binaria en la que se han corregido los parámetros con la reducción del sesgo de Firth para solucionar el problema de la separación. En todas las variables estudiadas se ha encontrado funcionamiento diferencial.

Palabras clave: MMSE, DIF, regresión logística binaria.

Abstract: The Mini Mental State Examination is a very used, extended and verified test, which is used to detect cognitive impairment. Although it has been widely studied, it has not been studied so much referring to their differential functioning in the Spanish population. This work aims to detect the differential item functioning in terms of these three proposed variables: education level, age and sex. For this, a binary logistic regression was used, in which the parameters were corrected using Firth's bias reduction to solve the problem of separation. It found differential functioning in all the studied variables.

Keywords: MMSE, DIF, binary logistic regression.

1.- INTRODUCCIÓN

No sería posible comparar la estatura de los varones y de las mujeres si el metro con el que se mide no tuviese las mismas propiedades en ambas poblaciones. De igual forma no debería compararse el deterioro cognitivo de grupos de sujetos diferentes entre sí, si el test no tuviera las mismas propiedades métricas en los grupos analizados (Prieto, Delgado, Perea y Ladera, 2011)

Prieto y cols. (2011) exponen que el Mini Mental State Examination (MMSE) (véase apéndice) se aplica a poblaciones diversas. Para que un test pueda ser aplicado a diferentes grupos, tiene que carecer de comportamiento diferencial de los ítems (DIF por sus siglas en inglés), esto quiere decir que en un test sin DIF no le tienen que influir las variables sociodemográficas como edad, sexo o nivel de estudios. Un ítem presenta DIF asociado a un grupo cuando sujetos con el mismo valor en la variable medida, pero procedentes de grupos distintos, tienen distinta probabilidad de resolver dicho ítem correctamente.

El DIF en las pruebas de un test cognitivo breve puede tener serias implicaciones para la validez de las conclusiones de la prueba (Crane y cols., 2006).

El presente estudio pretende analizar el DIF del test Mini Mental en la población española, ya que se ha señalado esta necesidad y así establecer las relaciones de las puntuaciones del test con las variables sociodemográficas como el género, la raza/etnia, ocupación, ingresos y residencia geográfica (Crum, Anthony, Bassett y Folstein, 1993), además de que ya se ha estudiado el comportamiento diferencial de los ítems en población italiana con el test MMSE y las variables edad, nivel de estudios, sexo y ocupación encontrándose diferencias en el test italiano (Crane y cols., 2006), pero no ha sido estudiado en la población española con las variables edad, sexo y nivel de estudios, que es el objetivo del presente trabajo.

También ha habido precedentes del estudio del DIF en el MMSE relativo a la educación con técnicas psicométricas (Jones & Gallo, 2001, 2002; Küçükdeveci, Kutlay, Elhan y Tennant, 2005; Teresi y cols., 1995).

Para conseguir detectar el DIF se propone hacerlo con regresión logística en el programa estadístico SPSS (IBM Corp., 2012). Pero se detecta un problema: la separación en la regresión y el SPSS no puede solucionarlo. Por tanto se buscan alternativas encontrando la solución en el paquete *logistf* (Heinze, Ploner, Dunkler y Southworth, 2013) del programa R (R Core Team, 2014).

2.- MINI MENTAL STATE EXAMINATION

En 1975 se publicó el primer artículo sobre el MMSE (Folstein, Folstein y McHugh, 1975). Nació en el New York Hospital-Wetchester-Division de la universidad Cornell, donde de inmediato se convirtió, incluso antes de la publicación oficial, en el instrumento estándar de evaluación cognoscitiva; lo mismo ocurrió en el Johns Hopkins Hospital de Baltimore, donde después trabajaron los autores. El MMSE extendió su campo de aplicación además de a la práctica clínica, a la docencia y a la investigación, particularmente con pacientes en edad geriátrica (Lobo, Escobar, Ezquerra y Seva Díaz, 1980).

En el artículo original del Mini Mental State (MMS) se propone un test rápido, que solo requiere entre 5 y 10 minutos para su aplicación, solucionando así el problema que supone que pacientes ancianos, particularmente aquellos que cursan con síndrome de delirios o demencia, cooperen por un período reducido de tiempo (Folstein y cols., 1975). Este artículo es probablemente la referencia más repetida en ciencias de la salud (Nilsson, 2007). El test

tiene por población diana a los pacientes geriátricos, es decir, personas mayores de 65 años o pacientes adultos con alteraciones neurológicas donde se consideran los años de escolaridad que éste posea (desde 0 a 12 años o más de escolaridad) (González-Hernández y cols., 2009). Fue concebido y creado originalmente para evaluar y describir de forma sistemática y cuantitativa el estado mental de pacientes psiquiátricos hospitalizados y monitorizar el cambio del mismo (Nieuwenhuis-Mark, 2010). Y en su versión española (Lobo y cols., 1980) tiene como finalidad la detección rápida del deterioro de funciones cognitivas.

2.1- EXTENSIÓN DEL MINI MENTAL

A partir de la presentación del test se ha extendido y ha sido adaptado a distintos países y traducido a distintos idiomas, entre los que se encuentra el español (Lobo y cols., 1980), portugués (Guerreiro y cols., 1994), japonés (Ishizaki y cols., 1998), chino (Sahadevan, Lim, Tan y Chan, 2000), turco (Güngen, Ertan, Eker, Yaşar y Engin, 2001), italiano (Noale, Limongi y Minicuci, 2005), hebreo (Werner, Heinik, Mendel, Reicher y Bleich, 1999), coreano (Kang, Duk y Hahn, 1997), cantonés (Chiu, Lee, Chung y Kwong, 1994), hindi (Ganguli y cols., 1995), griego (Fountoulakis, Tsolaki, Chantzi y Kazis, 2000), persa (Ansari, Naghdi, Hasson, Valizadeh, y Jalaie, 2010), francés (Law & Wolfson, 1995) y en total más de 50 idiomas (Carnero-Pardo, 2013).

Es bien conocido que la adaptación de los test neuropsicológicos a poblaciones diferentes continúa siendo un reto hoy en día, ya que las características sociodemográficas y culturales son aspectos que influyen notablemente en el rendimiento de los participantes (Contador, Fernández-Calvo, Ramos, Tapias-Merino y Bermejo-Pareja, 2010; Hannay, Howieson, Loring, Fischer y Lezak, 2004; Pedraza y cols., 2012).

El uso del Mini-mental es el instrumento recomendado por las principales guías de práctica clínica [p. ej. The The Quality Standards Subcommittee of the American Academy of Neurology (QSS of the ANN)] recomienda que instrumentos como el MMSE sean considerados para la detección de la demencia en sujetos con sospecha de deterioro cognitivo (Petersen, Stevens, Ganguli, Tangalos, Cummings y DeKosky, 2001), NICE (Guideline, 2007), Canadá (Chertkow y cols., 2008), España (Grupo de trabajo de la Guía, 2010).

Actualmente, es el TCB (Test Cognitivo Breve/screening) usado comunmente por la mayor parte de los psicogeriatras de EE.UU y Canadá (Shulman y cols., 2006). Además, es posiblemente el TCB más conocido en el mundo. Originalmente diseñado para evaluar el

deterioro cognitivo en la población de edad avanzada, se ha convertido en una de las primeras pruebas de cribado para hacer un diagnóstico de demencia (Nieuwenhuis-Mark, 2010).

El primer sorprendido del éxito y la difusión del mini-mental es su propio creador (Folstein, 1990), que reconoce que continúa sorprendiéndoles que una combinación particular de ítems que ha sido concebido en una sola noche sea tan útil para los clínicos y los epidemiólogos de tan diferentes países.

2.2.- EN QUÉ CONSISTE EL MINI MENTAL

El MMSE consta de 30 ítems que se clasifican en 6 dominios cognitivos: orientación temporal, orientación espacial, registro de información, atención y cálculo, memoria y lenguaje (Prieto y cols., 2011). También pueden clasificarse en grupos de ítems: orientación temporal (5 ítems), orientación espacial (5 ítems), fijación (3 ítems), atención y cálculo (5 ítems), memoria (3 ítems), nominación (2 ítems), repetición (1 ítem), comprensión (3 ítems), lectura (1 ítem), escritura (1 ítem), dibujo (1 ítem). Además en la versión de Lobo y cols. (1999) también se evalúa el nivel de consciencia en: alerta/responde, obnubilación, estupor, en coma/no responde.

El test no cuenta con ningún ítem original y, según su autor, se creó seleccionando aquellos ítems con significado clínico para el autor y que podían ser puntuados con poca interpretación (Folstein, 1990) entre los que se incluían algunos incluidos en otros instrumentos [p. ej. Escala de Blessed (Blessed, Tomlinson y Roth, 1968)].

El MMSE completo o abreviado es útil para predecir la enfermedad de Alzheimer en fase ligera. Sin embargo, no se recomienda para su uso como un instrumento de diagnóstico ya que un resultado negativo no la descarta. No obstante, se recomienda como una herramienta para identificar aquellos pacientes que necesitan un seguimiento más extenso (Tierney, Szalai, Dunn, Geskani y McDowell, 2000).

2.3.- VERSIONES

Existen varias versiones del test Mini Mental, originariamente introducido por Folstein y cols. (1975).

- Las traducciones en las que se estabiliza culturalmente algunas preguntas y en concreto la versión española cambia la serie de los 7 del MMSE original (restar de 100 siete y hacerlo sucesivamente cinco veces) por la resta de 30 menos tres, cinco veces; dispone de una versión de 35 puntos y otra de 30 (Lobo y cols., 1999).
- Otra versión es el MMSE de 37 puntos (MMSE-37), que está adaptado a personas de bajo nivel cultural y que proviene de un proyecto multicéntrico gestionado por la World Health Organization (WHO- AAD Project, Proyecto OMS-demencias asociadas a la edad) y es consecuencia de querer una herramienta válida que pueda ser aplicada de forma transnacional, realizable con personas de bajo nivel cultural y con la que pueda detectarse demencia (Amaducci y cols., 1991). De esta versión de 37 puntos se ha estudiado la fiabilidad interobservador y test-retest (Tapias-Merino y cols., 2010). El MMSE-37 añade la figura de un hombre que levanta las manos (se pide al paciente que lo imite) y el dibujo de dos círculos cruzados (más fácil que dos pentágonos del original), se sustituye el deletreo de la palabra 'mundo' al revés, poco empleada en español, por la repetición al revés de cinco dígitos: 1-3-5-7-9 (prueba de atención y comprensión del lenguaje), y se elimina la frase 'no ifs, and or buts' (ni síes, íes o peros/ni sí, ni no, ni pero), sin dificultad de comprensión y repetición en español, y se sustituye por '...en un trigal había tres tigres' (frase con poco sentido y de pronunciación difícil). El MMSE-37 se validó para el cribado de demencia en un estudio multipaís (España, Malta y Chile) (Baldereschi y cols., 1994) y posteriormente se ha validado en otras muestras poblacionales (Quiroga y cols., 2004; Villanueva-Iza y cols., 2003) y se ha utilizado en el estudio poblacional NEDICES en España (Bermejo Pareja y cols., 2008a, 2008b, 2009).
- También está la versión telefónica (Roccaforte, Burke, Bayer y Wengel, 1992). En la que se han sustituido las órdenes «coja el papel con la mano derecha», «dóblelo por la mitad con las 2 manos» y «deposítelo en la mesa» por «dígame hola», «golpee el auricular 3 veces» y «dígame "ya vuelvo a estar con usted"». Las otras modificaciones respecto a la versión presencial se realizaron en 2 ítems de orientación espacial y en el ítem de denominación. Concretamente, respecto a la orientación espacial se sustituyó la pregunta «¿dónde estamos en este momento?» por «¿cuál es su dirección de correo?» para los pacientes con domicilio propio y por la pregunta «¿cómo se llama el centro donde usted vive?» para los pacientes institucionalizados. También se sustituyó la pregunta «¿en qué piso/planta estamos?» por la pregunta «¿cuál es su número de teléfono?». Respecto a la denominación, se sustituyó la denominación por contraste visual del reloj y el bolígrafo por la pregunta «dígame cómo se llama la cosa que está utilizando para hablar conmigo» (Garre-Olmo y cols., 2008).

- The Modified Mini Mental State Examination (3MS) es una versión modificada del examen Mini Mental State (MMS). El 3MS está diseñado para probar una variedad más amplia de las funciones cognitivas, cubrir una amplia gama de niveles de dificultad, y mejorar la fiabilidad y validez. Además conserva la brevedad, facilidad de administración, y la puntuación objetiva del original, pero amplía el rango de calificaciones posibles. También añade cuatro preguntas adicionales (fecha y lugar de nacimiento, fluidez verbal, similitudes, y el retraso de recuerdo de las palabras). Por último proporciona directrices adicionales de puntuación, y aumenta el rango de puntuaciones de los ítems para producir una puntuación máxima total de 100 puntos (Teng & Chui, 1987).
- El MMSE-blind es una versión del MMSE para discapacitados visuales. En ella se han omitido todos los ítems que requieren procesamiento de imágenes y se ha hecho una transformación lineal para interpretar las puntuaciones (Busse, Sonntag, Bischof, Matschinger y Angermeyer, 2002).

2.4.- VENTAJAS

Contextualmente este test ha sido un antes y un después en la evaluación clínica de las personas mayores ya que la alta prevalencia de demencias en la población general geriátrica (Hofman y cols., 1991; Lobo y cols., 1999) demandaba la necesidad de un instrumento fiable, válido y sencillo para la detección de las funciones cognitivas (Lobo y cols., 1999).

Una de sus principales ventajas es su extendido uso. De aquí se deriva que se dispone de una ingente cantidad de datos, es ampliamente conocido y aplicado, existen muchos estudios y por ello puede compararse su puntuación entre estudios y entre sujetos así como el seguimiento y valoración de la respuesta al tratamiento (Carnero-Pardo, 2013). Es un instrumento de fácil uso que puede ser aplicado y evaluado por personal no cualificado y que se enseña en la mayoría de centros académicos (Powsner & Powsner, 2005). Por último es de señalar que la estructura del MMSE asegura la evaluación de diferentes dominios cognitivos (Tombaugh & McIntyre, 1992).

Los autores Tombaugh y Mc Intyre (1992) encontraron fiabilidad que oscilaba entre 0,95 y 0,80 además de una sensibilidad (proporción de diagnósticos positivos correctos) de al menos el 87% y un valor de predictivo positivo de 79%.

2.5.- EL TEST NO SIEMPRE RESPONDE COMO SE ESPERA

Se ha reconocido que las puntuaciones del MMSE deberían ser interpretadas teniendo en cuenta el nivel de educación y edad del paciente (Crum y cols., 1993).

A pesar de su popularidad, el MMSE tiene varias deficiencias psicométricas, incluyendo fallos por omisión de las personas con déficits cognitivos leves, sensibilidad limitada a lesiones focales, falsas alarmas frecuentes en personas con bajo nivel de educación, más fallos por omisión en los individuos con altos niveles de educación, falta de instrucciones estandarizadas para la prueba y fallos referidos a los criterios de puntuación (Teng & Chui, 1987). No obstante en el test se señala que puntuaciones bajas no deberían ser utilizadas para hacer un diagnóstico de demencia o para discriminar entre diversas formas de demencia u otros trastornos (Lobo y cols., 1999). También señala que en relación a los “falsos negativos” algún paciente con nivel educacional e intelectual muy elevados puede puntuar por encima del punto de corte más común si, por ejemplo, tiene un deterioro incipiente de su rendimiento cognoscitivo. Y las lesiones focales pueden no ser detectadas si se tiene en consideración, exclusivamente, la puntuación total del MMSE; por el contrario, el fallo sobre algunos ítems específicos del instrumento pueden alterar sobre un déficit real.

Y sobre los falsos positivos, en el artículo del test (Lobo y cols., 1999) se explica que además de los casos de edad avanzada o escolaridad deficiente, puede deberse a psicastenia o negativismo. A veces puede deberse a depresiones, aunque éstas también producen en casos graves déficits cognoscitivos reales.

Se ha señalado que es un TCB que sin embargo no identifica desórdenes específicos (Crum y cols., 1993). Afirmación confirmada ya que también se ha encontrado que el funcionamiento diferencial de 11 de los 30 ítems no es igual para pacientes con enfermedad de Párkinson que para pacientes con demencia tipo Alzheimer que para sujetos normales por lo que la comparación de las puntuaciones no estaría justificada (Prieto y cols., 2011).

George, Landerman, Blazer y Anthony (1991) encontraron que aunque no hay diferencia en cuanto al sexo, se encontraron diferencias entre razas/etnias. Se encontró mayores tasas de deterioro cognitivo grave en personas de color con los blancos o hispanos.

Algunos informes han mostrado la relación entre el nivel educativo y la edad con la puntuación total del MMSE, esta información ha sido difícil de interpretar en el campo clínico porque normalmente se necesita evaluar la puntuación individualmente en los pacientes de diferentes edades y niveles educativos. De todas formas no se han publicado las puntuaciones en función de la edad y el nivel educativo en a partir de un muestra

representativa, lo que permitiría a los clínicos una forma de contextualizar las puntuaciones de sus pacientes (Bleecker, Bolla-Wilson, Kawas y Agnew, 1988).

El Mini mental contiene múltiples ítems que no aportan apenas capacidad discriminativa al conjunto, en especial en deterioro cognitivo y demencias leves. El mini mental completo no es mejor que un grupo reducido de sus ítems o versiones abreviadas del mismo (Tierney y cols., 2000; Haubois y cols., 2011; Schultz-Larsen, Lomholt y Kreiner, 2007).

Es de destacar que en las instrucciones originales no se especifican las palabras a repetir (Folstein y cols., 1975). Esto supone un gran inconveniente ya que queda a la elección del explorador, porque el recuerdo es sensible a las palabras empleadas y está influido por factores como frecuencia de uso, categoría semántica, concreción, dificultad fonética, número de sílabas, imaginabilidad, familiaridad, existiendo, en consecuencia, elementos que se recuerdan más fácilmente que otros (McCarten, Anderson, Kuskowski, McPherson y Borson, 2011). No obstante, en la versión española (Lobo y cols., 1999) sí se detallan estas instrucciones solucionándose así este problema. También se proponen series alternativas de palabras para su recuerdo, alternativas para la orientación espacial “¿Sabe en qué provincia estamos? (comunidad autónoma)”.

El test presenta ítems condicionales, de tal forma que si no se supera el primer ítem de cálculo numérico, será mucho más difícil de superar el siguiente ítem, ya que se parte del anterior.

Por último este test está protegido y limitado en cierta medida por el hecho de estar protegido con copyright de la edición española por TEA ediciones S.A.

2.6.- NO NORMALIDAD

Existe una amplia evidencia experimental que demuestra que los resultados obtenidos en muchas pruebas psicométricas no siguen una distribución normal (Micceri, 1989). Y las puntuaciones del MMSE no son diferentes, ya que Crum y cols. (1993) encontraron que no se corresponden a una distribución normal y por lo tanto no se aplican el uso de los supuestos de la teoría de la normalidad.

2.7.- EDUCACIÓN

Se han encontrado más variabilidad en puntuaciones de personas con bajo nivel de educación (entre 0 y 4 años de escolaridad) que en personas con alto nivel de educación (más de 12 años de escolaridad) (Crum y cols., 1993).

Pese a que en las instrucciones originales el punto de corte se establece en 23 en el que ese punto de corte o menos supondría la probabilidad de presencia de un déficit cognoscitivo y un examen más exhaustivo (Lobo y cols., 1999), algunos se han dado cuenta de la necesidad de ajustar las puntuaciones del MMSE porque las personas con un bajo nivel de educación pueden no ser capaces de hacer bien el test pese a tener una buena capacidad cognitiva. De ser cierto esto se estaría cometiendo un sesgo al hacer inferencias sobre población utilizando este test (Magaziner, Bassett y Hebel, 1987; O'connor, Pollitt, Treasure, Brook y Reiss, 1989; Kittner y cols., 1986).

Ya se señala en el test que debe tenerse especial cautela en los grupos de edad más avanzada y escolaridad muy deficiente (Lobo y cols., 1999).

Se ha demostrado que las personas con bajo nivel de educación y con baja puntuación en el MMSE también fueron menos capaces de realizar las tareas de la vida cotidiana, lo que sugiere que estos individuos estén de hecho más dañados cognitivamente. El menor grado de variabilidad en el grupo de personas con alto nivel de educación puede reflejar un tipo de efecto techo o una menor sensibilidad del MMSE de detectar daño cognitivo en estos individuos (Bassett & Folstein, 1991).

Una importante limitación del mini mental en cuanto a la educación es que no puede aplicarse a analfabetos al exigir 2 de sus ítems saber leer y escribir. Es también una limitación el hecho de incluir una tarea de lápiz y papel (copia el dibujo), no por la complejidad instrumental que pueda suponer, sino porque las tareas de lápiz y papel, aunque teóricamente son accesibles a los sujetos analfabetos y con bajo nivel educativo, generan aversión y rechazo en estas personas y la ejecución en ellas de éstas es muy pobre (Carnero-Pardo, 2013).

El perfil de la persona analfabeta en España responde a dos fenómenos. Uno residual, que es el de mayores de 65 años que tuvieron que ponerse a trabajar muy jóvenes o a cuidar a familiares. Y otro fenómeno marginal vinculado a las minorías étnicas o inmigrantes (Viñao, Moreno Martínez y Navarro García, 2009). Ese es uno de los motivos por los que la tasa de alfabetización de la población adulta (porcentaje de la población de 15 años y más que puede tanto leer como escribir frases simples sobre su vida cotidiana) se estima en 98,10% para el año 2015. (UNESCO, 2015)

3.- MÉTODO

3.1.- TEORÍA CLÁSICA DE LOS TEST

El nacimiento formal de la Teoría de los Test puede ubicarse en los primeros trabajos de Spearman realizados en la primera década en los que da soporte teórico a la integración de las diferentes pruebas de que constan los test, así como a las puntuaciones que se obtienen con ellas estableciendo los fundamentos de la Teoría Clásica de los Test (TCT) (Muñiz, 1992).

Thurstone y Chave (1929) propusieron un método de escalamiento denominado intervalos aparentemente iguales. Estos autores lo emplearon para construir escalas de aptitudes, es decir, de estímulos psicológicos. Una aptitud es una valoración de un objeto aprendida socialmente que se manifiesta en creencias, sentimientos y conductas. En este método se propone la redacción de un conjunto de 100 ítems para que, después de un cribado se quede en el número de ítems que creían mejor para un test: 30.

La TCT tiene como objetivo cuantificar la fiabilidad de los test. Que se refiere al grado de error que afecta a las medidas, es decir, a la precisión del test. Dicha precisión se analiza en función de la estabilidad de las sucesivas puntuaciones que pueden alcanzar una misma persona en varias aplicaciones de un mismo test.

La fiabilidad del test es cuestión del grado en que reducimos los errores ya que todos los test cometen error en mayor o menor medida.

Spearman (1904) con sus primeros trabajos origina la TCT también conocido como modelo de la puntuación verdadera o de los test paralelos. La ecuación básica del modelo es:

$$X = V + E$$

Donde

- X es la puntuación empírica u observada (puntuación directa del sujeto en el test).
- V es la puntuación verdadera o puntuación libre de error (nivel real de las personas en el área medida en el test).
- E es el error aleatorio de medida.

Tiene como supuestos los siguientes:

- La puntuación verdadera de un sujeto en un test es igual a la esperanza matemática de sus puntuaciones empíricas. $V = \bar{X} = \varepsilon(X)$

- Las puntuaciones verdaderas y los errores aleatorios de medida son independientes, es decir, su correlación es nula. $r_{VE} = 0$

- Los errores de medida cometidos en un mismo test o en dos test diferentes son independientes. $r_{e_i e_j} = 0$

Estos supuestos no pueden ser comprobados empíricamente, lo que quiere decir que no permiten deducir la cantidad de error que afecta a una determinada puntuación en un test.

El modelo así formulado no tiene relevancia práctica ya que no pueden hacerse operaciones con ellos, porque de una u otra forma todos ellos contienen elementos no observables empíricamente. Así pues Spearman propone la técnica de los test paralelos, la cual dice que dos test serán paralelos cuando además de medir el mismo rasgo teórico, presenten los mismos índices psicométricos, o sea, que si un test es paralelo a otro, ambos medirán la misma conducta. Y si se cumplen las condiciones suficientes

$$S_{e_1}^2 = S_{e_2}^2$$

$$V_1 = V_2$$

y las condiciones necesarias

$$S_{x_1}^2 = S_{x_2}^2$$

$$X_1 = X_2$$

$$r_{x_1 x_1} = r_{x_2 x_2}$$

Los dos test tendrán la misma varianza de los errores y las puntuaciones verdaderas de los sujetos serán iguales a ambos.

La TCT tiene limitaciones que viene librando su consecuente, la TRI:

- La TCT está orientada a la puntuación global en el test $X = V + E$. Así dos personas han podido tener la misma puntuación en el test, pero no es necesario que acierten los mismos ítems, así pueden tener idéntica puntuación pero su conducta no refleja dicha puntuación. Es decir, la TCT no representa bien las respuestas de los sujetos. En la TRI la puntuación de cada sujeto corresponde a un ítem concreto lo que implica mayor eficacia pronosticada de la conducta observada.
- En la TCT las propiedades psicométricas del ítem dependen de la muestra de sujetos a la que lo apliquemos, mientras que el objetivo de la TRI es que no exista tal dependencia.
- En la TCT la puntuación de un sujeto depende de la muestra de ítems. Las puntuaciones de diferentes test no serían equiparables. Así aparecen problemas con los estudios trasversales: no se puede medir al sujeto con el mismo test a lo largo del tiempo. En la TRI se desea que dependa de su nivel de habilidad invariante con respecto a los ítems.
- Las medidas en la TCT no consideran las diferencias entre los ítems (miden sujetos, los ítems permanecen invariantes). Resolver un ítem fácil puntúa lo mismo que resolver un ítem difícil.
- La diferente métrica para ítems y para sujetos impide interpretar las interacciones entre ambos. No se conoce qué ítems ha superado el sujeto, así las puntuaciones se interpretan en base al grupo de pertenencia del sujeto, no a su nivel en la variable.
- Los supuestos centrales son poco verosímiles e indemostrables: definición de test paralelos, invarianza del error típico de medida... (Lord, Novick y Birnbaum, 1968).
- En la TCT no se mide con la misma fiabilidad a lo largo de la variable, ni se evalúa la fiabilidad a nivel de la persona, sino a nivel del grupo.
- El nivel de medida en la TCT es arbitrario, no llegando a alcanzar el grado de intervalo por no poseer una unidad de medida constante.

Así pues, se ha dado paso a la teoría de respuesta al ítem.

3.2.- TEORÍA DE RESPUESTA AL ÍTEM

El término Teoría de la Respuesta a los Ítems se debe a Lord (1980) y constituye un nuevo enfoque en la teoría de los test que permite resolver ciertos problemas de medición psicológica inatacables desde la TCT. El nombre de TRI proviene de que este enfoque se basa en las propiedades de los ítems más que en las del test global.

Tiene por objetivos:

- Obtener mediciones que no varíen en función del instrumento utilizado, que sean invariantes respecto a los test empleados. Esto quiere decir que si se intenta medir la depresión, la medida obtenida no deberá cambiar si se utiliza un test u otro, la puntuación de ese rasgo es la suya y no otra. En caso de que salgan diferentes puntuaciones se estaría midiendo rasgos diferentes.
- Disponer de instrumentos de medida cuyas propiedades no dependan de los objetos medidos, sean invariantes respecto de los sujetos evaluados. Así se mediría el rasgo en sí, solo el rasgo y no otras cosas.
- Los modelos han de ser expresados a nivel de los ítems. Este enfoque se centra en los ítems y no tanto en el conjunto de ellos conformando un test, por eso se tratan individualmente.
- Eliminación del requerimiento de la definición de test paralelos.
- Medición de la precisión (fiabilidad) para cada nivel de habilidad. Una medición cuantificable.

Puede estudiarse el ítem a través de la Curva Característica del Ítem (CCI) la cual es una distribución de probabilidad condicional que indica cuál es la probabilidad de resolver correctamente un ítem para los distintos niveles de habilidad de los sujetos. Puede hacerse una representación gráfica de una probabilidad condicional. Dicha probabilidad es la de poder dominar un ítem en función del nivel de la persona en la variable. Dicho de otra forma, un test se compone de ítems, si estos ítems corresponden a una sola dimensión y difieren en

dificultad podrían escalarse los sujetos. Así podrían clasificarse a las personas en función de su nivel de habilidad en determinado rasgo. De esta forma una persona con su capacidad numérica mermada tendría más dificultades en superar los ítems de cálculo del test. Esta probabilidad se representa en forma de gráfico, en la que en el eje de ordenadas representa la probabilidad de acertar dicho ítem $P(\theta)$ y en el eje de abscisas se representan los valores de la variable que mide el ítem, denominado por θ . Quedando constituida una escala de intervalo con media 0 y desviación típica 1 cuyos valores oscilan entre $\pm\infty$, aunque entre ± 5 se suelen recoger todos los casos. Así el nivel de habilidad necesario para acertar el ítem queda representando una función en la que se muestra un crecimiento gradual de las proporciones de respuesta afirmativa.

Por su puesto la CCI no es “directamente observable” ya que tampoco lo es θ ; pero puede ser inferida a través de las respuestas de los sujetos a los ítems del test (García-Cueto, 1993).

Como ejemplo del uso y generación del gráfico podemos suponer que después de pasar un test a muchos sujetos obtenemos las puntuaciones del test, por ejemplo 80 como mínimo y 120 como máximo. Se podría representar el rendimiento fijándonos en todas las personas que obtuvieron puntuación de 80 (en este caso 132 personas) Vemos cuántas personas de las anteriores han acertado el ítem (pongamos que han sido 2 personas) y realizamos una proporción ($2/132 = 0.02$). Repetimos el proceso con las personas que han obtenido la siguiente puntuación (81) obteniendo otra proporción, (por ejemplo 0,02). Este proceso se repite hasta completarlo con las que obtuvieron 120 (la proporción fue 0.99). Quedando así una gráfica en la que muestra la proporción de acertar el ítem en función del grupo de personas al que se pertenezca.

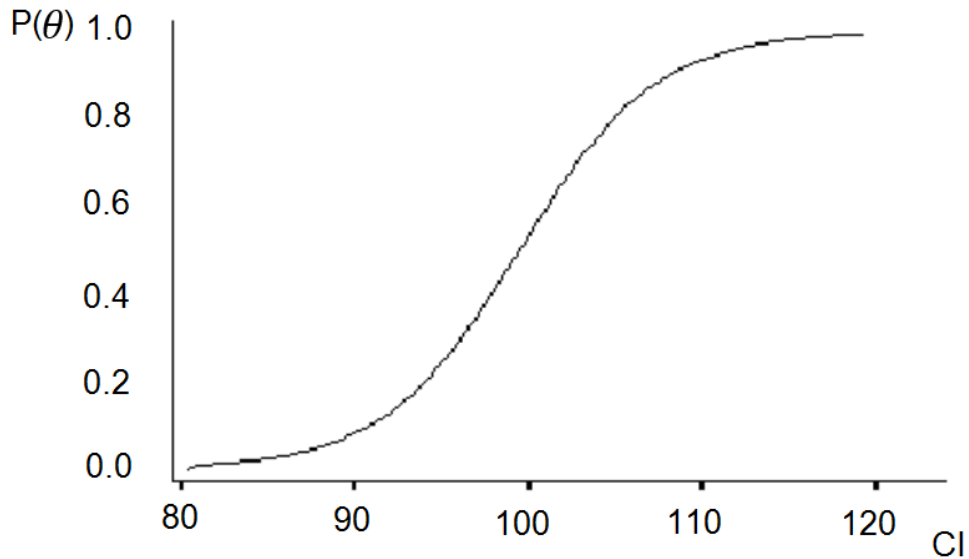


Figura 1. Curva Característica del Ítem. Proporción de acertar un ítem dentro de una muestra de personas.

Si en este ejemplo se midiese el cociente intelectual de las personas, quedaría una gráfica que representase que cuanto mayor es el cociente intelectual, mayor probabilidad de acertar el ítem. Así para una puntuación de 100 tendría una probabilidad de acertar de 0,5 y a una puntuación de 120 le correspondería una probabilidad de 0,99.

En la CCI pueden diferenciarse los siguientes parámetros: a , β , c y d (véase figura 3):

- El parámetro “ a ” se denomina índice de discriminación (pendiente o inclinación máxima) y su valor es proporcional a la pendiente de la recta tangente a la CCI en el punto de máxima pendiente de ésta (donde $b = \theta$). Así, cuanto mayor sea la pendiente, mayor será el índice de discriminación y como se intenta que la discriminación sea máxima, cuanto mayor sea el parámetro “ a ”, mejor será el ítem. Teóricamente puede alcanzar valores desde menos infinito a más infinito, pero todos los ítems con valores de a negativos se eliminan de los test.

Si θ sigue una distribución $n(0,1)$ y no hay aciertos ni fallos al azar, (Lord, 1980), el valor del parámetro a viene dado aproximadamente por la siguiente expresión:

$$a \cong \frac{r_b}{\sqrt{1 - r_b^2}}$$

Donde r_b es la correlación biserial entre el ítem y el test, es decir, r_b es una medida de la discriminación del ítem en la TCT.

- El parámetro “b” (β) también llamado índice de dificultad o “localización” cuando se habla de test de ejecución típica (en el que se recogen actitudes o sentimientos de las personas) es el valor de θ correspondiente al punto de máxima pendiente de la CCI que coincide con el valor en habilidad correspondiente a una probabilidad de acertar el ítem de 0,5 [$P(\theta)=0.5$] si no se dan aciertos o fallos por azar y de

$$P(\theta) = \frac{(1 + c)}{2}$$

teniendo en cuenta el parámetro c. La figura 1 representa en el eje de las ordenadas la probabilidad de acertar el ítem y en el eje de las abscisas la habilidad necesaria para acertar el ítem. En esta figura, el valor de máxima pendiente coincide con la habilidad o desempeño cero.

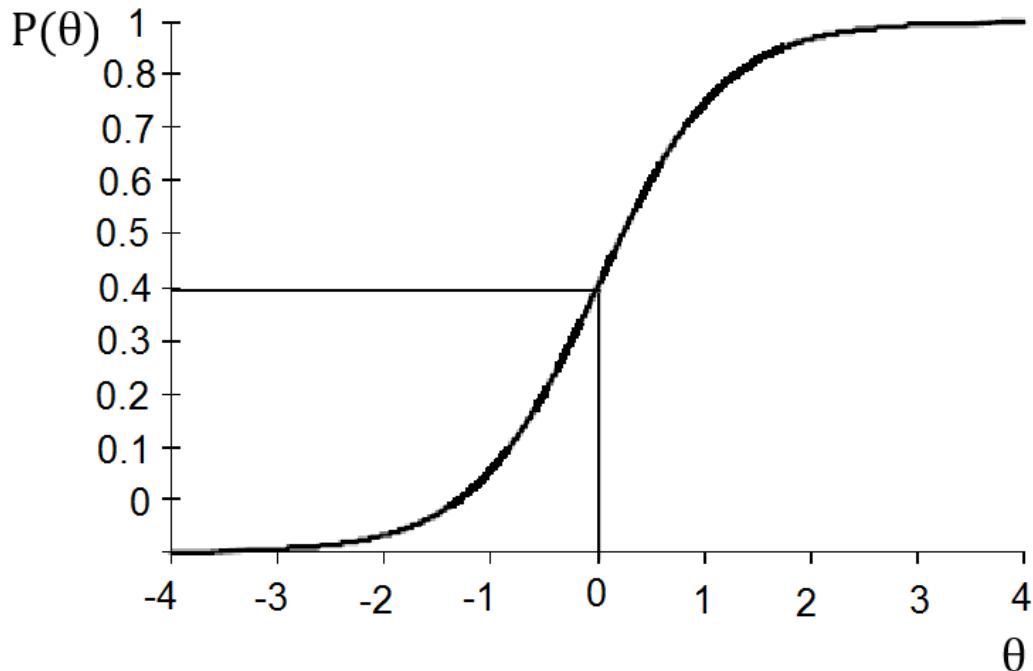


Figura 2. Curva característica de un ítem.

- El parámetro “c” (también se llama pseudoazar) representa la probabilidad de acertar el ítem al azar cuando “no se sabe nada”, acierto por adivinación, es decir el valor de

$P(\theta)$ cuando $\theta = -\infty$. En otros términos, es el valor asintótico de la CCI cuando θ tiende a $-\infty$. Atendiendo a este parámetro, los mejores ítems serán aquellos cuyo parámetro “c” sea cero.

- Parámetro “d” trata de explicar a los sujetos con un alto nivel de aptitud en la variable pero no superan un ítem que se supone más fácil. Esto se da si un sujeto ha superado un ítem difícil, pero no un ítem fácil. Tratándose del mismo constructo, sería una incongruencia a explicar. Sería, en cierto modo lo opuesto a acertar el ítem por azar. Más que tener en cuenta este parámetro, habrá que cuidar que su valor sea cero. No parece que dicho parámetro añada mucha más información sobre la habilidad de los sujetos que la proporcionada por los otros tres.

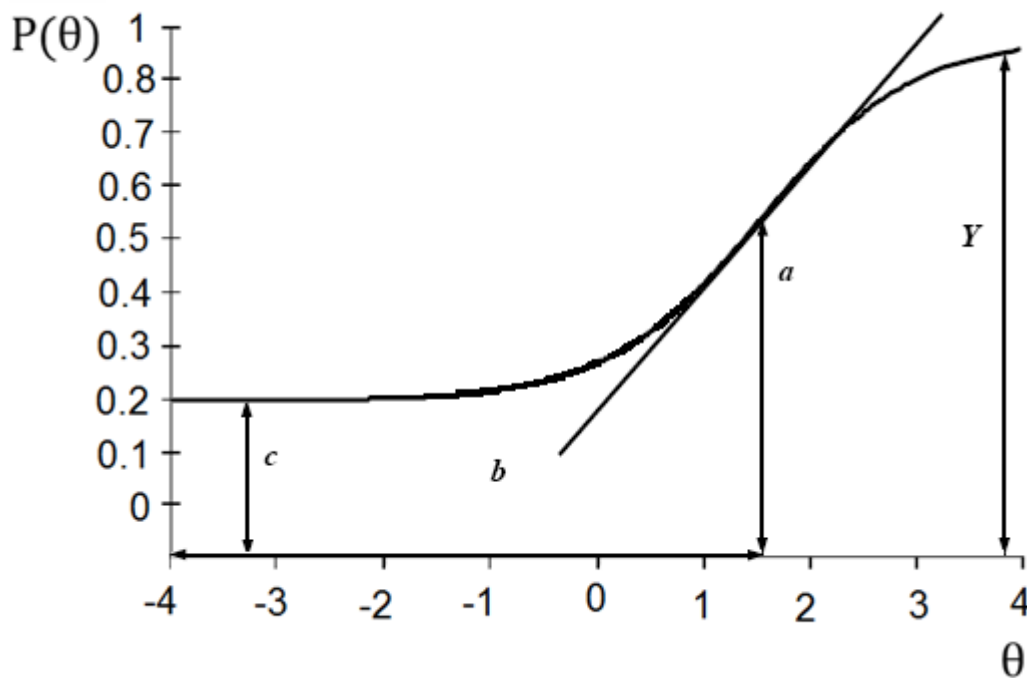


Figura 3. Parámetros de la Curva Característica de un Ítem. ($\theta \sim N(0,1)$ y, $a=1$, $b=1.5$, $c=0.2$ e $Y=0.9$).

3.2.1.- SUPUESTOS DE LA TRI

- Unidimensionalidad. Los sujetos y los ítems pueden graduar objetivamente en la misma dimensión, es decir, dependen de un único atributo. A esto se le llamaría escalamiento conjunto o de respuestas. Las correlaciones entre los ítems pueden ser fundamentalmente explicadas por un solo factor, por una única dimensión. Esto es porque es más sencillo

explicar la probabilidad de dar una respuesta en función de una sola variable y no de varias. Para comprobar la unidimensionalidad comúnmente se ha llevado a cabo mediante análisis factorial (Ferrando, 1996; León, 2012).

- Independencia local. Cuando se fija o se controla en nivel en el rasgo medido, las respuestas a los ítems son independientes; las correlaciones entre los ítems no difieren de cero. Acertar un ítem no quiere decir acertar el siguiente en dificultad. El cumplimiento de este supuesto significa que las respuestas de los sujetos se deben esencialmente a su distinto nivel en un único atributo. Por lo tanto, si se cumple la unidimensionalidad de eso se deriva matemáticamente que existe independencia local entre los ítems.

La probabilidad de que un sujeto con aptitud θ supere unas respuestas determinadas en un test ($i = 1, 2, 3, \dots, n$) queda igual al producto de probabilidades de respuesta de cada uno de los ítems por separado en función al nivel de aptitud, esto queda expresado por:

$$P(X_1, X_2, X_3, \dots, X_n | \theta) = \prod_{i=1}^n P(X_i | \theta)$$

en la que X_i es la respuesta de un sujeto al ítem i -enésimo.

3.3.- ANÁLISIS DEL FUNCIONAMIENTO DIFERENCIAL DE LOS ÍTEMS

En la variante anglosajona puede encontrarse en la literatura DIF de funcionamiento diferencial de los ítems, o también como “*item bias*”, o sesgo aunque sería una acepción errónea ya que el término sesgo implica el que un grupo posee una “injusta” ventaja sobre otro. Sin embargo, los grupos pueden presentar diferencias entre ellos en sus respuestas ante un ítem debido a razones diferentes al sesgo (García & Marañón, 1996).

Podemos hablar de la presencia de DIF cuando miembros de ambos grupos con igual nivel de competencia o habilidad poseen diferentes probabilidades de acertar o superar el ítem en cuestión (Lord, 1980).

La detección del DIF tiene como objetivos:

- Determinar si la diferencia de respuesta a un ítem entre grupos de personas (varones/mujeres, nativos/inmigrantes, etc.) se debe a diferencias en el constructo medido o a variables extrañas.
- Identificar diferencias cualitativas en las manifestaciones de un atributo (p. ej. saber si el atributo cálculo numérico funciona igual para hombres que para mujeres).
- Construcción de test con una validez generalizada para diversos grupos (a base de eliminar los ítems que presente comportamiento diferencial).

Para una mejor aclaración de los conceptos que engloban esta técnica, a continuación se definen los centrales:

- Impacto: diferencias globales de las respuestas a un ítem que se deben a que los grupos que difieren en el constructo medido.
- Funcionamiento diferencial o sesgo: la diferencia entre los grupos de sujetos se debe a diferencias entre atributos distintos (rasgos diferentes) al que se pretende medir con el ítem.
- Indicador de DIF: las personas de distintos grupos, pese a tener el mismo nivel en el constructo, tienen distinta probabilidad de una respuesta.
- Grupo focal (GF): el grupo que es el foco de interés del análisis (grupo minoritario).
- Grupo de referencia (GR): el grupo que sirve como base de comparación (grupo mayoritario).

Tipos de DIF (Mellenbergh, 1982):

- DIF uniforme: la probabilidad de una cierta respuesta es siempre mayor para uno de los grupos a través de todos los niveles del atributo, esto quiere decir que siempre es mayor en todos los niveles del rasgo. Si se representase en una CCI del GF y la CCI del GR no se cruzarían en ningún momento.
- DIF no uniforme: la diferencia entre las probabilidades de cada grupo de cierta respuesta cambia de signo (positivo o negativo) a lo largo de los niveles del atributo. Esto quiere decir que en algún nivel del atributo cambia la probabilidad de una cierta respuesta para los grupos. Un posible ejemplo sería que algunos hombres tuviesen menor probabilidad

que las mujeres de acertar un ítem, pero otros hombres tuvieran una mayor probabilidad de acertar ese ítem que las mujeres. Al representar las CCI, ambas curvas se cruzarían.

- o Dentro del DIF no uniforme, Swaminathan & Rogers (1990) hacen una distinción entre simétrico y mixto. El FDI no uniforme simétrico se da cuando las diferencias de probabilidad de contestar correctamente un ítem entre los dos grupos se cancelan, o sea, que el parámetro b se mantiene constante y el parámetro a varía entre ambos grupos. Se representaría la CCI de cada grupo cuando se cruzan estas curvas en el centro.

- o El DIF no uniforme mixto se da cuando las diferencias de probabilidad de contestar el ítem no se anulan, o sea, los parámetros a y b son desiguales para ambos grupos. Se representaría con un cruce como en el DIF simétrico, pero éste no sería simétrico.

3.4.- REGRESIÓN LINEAL

Toma el nombre de dos conceptos: de la regresión a la media, la cual es el fenómeno en el que si una variable toma un valor extremo, en las siguientes mediciones tenderá a suavizarse ese extremo, por lo que ese valor extremo no lo será tanto. Así si los hijos de una persona son demasiado altos o demasiado bajos, los descendientes de éstos o consecutivos no lo serán tanto, porque no se puede crecer o decrecer hasta el infinito; el término lineal se utiliza por estar basado en una explicación simplificada de la realidad en la que destaca por su agilidad y extenso soporte teórico.

Se engloba dentro del modelo lineal general. Es una técnica estadística usada para estudiar la relación entre variables. Con dos variables (regresión simple) o más de dos variables (regresión múltiple) una regresión lineal sirve para explorar y cuantificar la relación entre la variable dependiente y una o más variables independientes, también sirve para desarrollar una ecuación lineal para poder diseñar un modelo explicativo.

El modelo de regresión lineal relaciona la variable dependiente γ con \mathbf{N} variables explicativas

$X_n (n = 1, \dots, N)$ o transformaciones de éstas de tal forma que

$$\gamma = \sum \beta_n X_n + \varepsilon$$

donde ε es el error aleatorio.

Para que la ecuación quede completamente especificada es necesario un conjunto de observaciones, así en la observación j -enésima queda el modelo como sigue:

$$y = \sum \beta_n X_{nj} + \varepsilon_j$$

El modelo de regresión lineal clásico tiene unas hipótesis en las que se apoya para poder desarrollarse. Éstos son:

- Nulidad de la esperanza matemática. Los valores de las variables explícitas no tomarán valores positivos o negativos de forma sistemática, sino que será de forma aleatoria teniendo como su valor esperado el cero.

$$E(\varepsilon_j) = 0$$

- Homocedasticidad. En la que todos los términos tienen la misma varianza (la cual es desconocida). Para todo t queda así la ecuación:

$$\text{Var}(\varepsilon_t) = E(\varepsilon_t - E\varepsilon_t)^2 = E\varepsilon_t^2 = \sigma^2$$

en la que la dispersión de cada ε_t en torno a su valor esperado es siempre la misma.

- Incorrelación. Las covarianzas entre las diferentes perturbaciones no están correlacionadas. Por consiguiente el valor de la perturbación para cualquier observación muestral no está influenciado por los valores de las perturbaciones correspondientes a otras observaciones muestrales. Para todo t y S , con t distinto de S la fórmula quedaría como sigue:

$$\text{COV}(\varepsilon_t, \varepsilon_S) = (\varepsilon_t - E\varepsilon_t)(\varepsilon_S - E\varepsilon_S) = E\varepsilon_t\varepsilon_S = 0$$

- Regresores no estocásticos, o sea, que las variables exógenas sean de carácter determinista. Así que al repetir el análisis con una muestra nueva, los valores de las variables explicativas serían los mismos que a los iniciales.

- Inexistencia de relaciones lineales exactas entre los regresores.

- No existen errores de especificación en el modelo ni errores de medida en las variables explicativas. $T > k + 1$

- Normalidad de las perturbaciones

$$\varepsilon \sim N(0, \sigma^2)$$

Y unos supuestos necesarios para lograr un modelo de regresión, que son:

- Relación lineal entre variables.
- Independencia de los errores en la medición de las variables explicativas.
- Homocedasticidad o varianza constante de los errores.
- Que la esperanza matemática de los errores sea cero.
- El error total debe ser la suma de todos los errores por separado.

3.5.- REGRESIÓN LOGÍSTICA

Cornfield, Gordon y Smith (1961) perfilaron la técnica de la regresión logística. Posteriormente Walter y Duncan (1967) la utilizaron de la misma forma en la que se utiliza en la actualidad, para estimar la probabilidad de ocurrencia de un proceso en la que viene influenciado por las variables que componen el modelo propuesto. Debido al mejoramiento de los ordenadores, esta técnica se ha visto en auge dando paso a que en la actualidad se conoce como un método estándar en el que puede encontrarse en muchos manuales como el de Agresti (1996).

Con esta técnica se permite evaluar la influencia de cada una de las variables independientes sobre la variable dependiente o también llamada variable de respuesta controlando el efecto del resto de variables. De esta forma el modelo se compondrá de una variable dependiente o variable de respuesta Y , la cual puede ser dicotómica o politómica, además de las variables independientes X_1, X_2, X_3, \dots las cuales pueden ser cualitativas o cuantitativas.

Con esta técnica puede estudiarse cómo influye en la probabilidad de ocurrencia un suceso (dicotómico o politómico) la presencia o ausencia de factores y el nivel de estos factores.

3.6.- REGRESIÓN LOGÍSTICA BINARIA

El modelo más simple de regresión es estudiar el efecto de una sola variable X sobre la variable de respuesta Y , siendo ésta dicotómica. De esta manera, puede considerarse como una probabilidad en la que los individuos que presentan el evento se les asigna una probabilidad observada de 1, y a los que no presentan dicho evento, se les asigna una probabilidad de 0. No es fácil dar con un modelo que pueda describir del todo bien la realidad, aunque puede describirse la situación de un modo lo más aceptable posible.

La regresión logística ofrece técnicas para analizar aquellos datos que no se ajustan a los modelos estadísticos tradicionales, ya que la regresión lineal no es aplicable porque la variable de respuesta solo presenta dos valores.

Para ilustrar esta técnica, se expone un famoso ejemplo que hace Preston (2006) en el que explica a sus alumnos la utilidad de la regresión logística con un ajuste que explica cómo los senadores de EE. UU. votan sobre un proyecto de ley que afecta directamente al sector automovilístico. Y se intenta explicar en función de la afiliación política (partido) de cada senador, del estado al que representa y de la cantidad de dinero total que ha recibido el senador por parte de las compañías automovilísticas en su período de campaña para ser elegido como senador. Así pues, se intenta saber el peso de estas variables sobre la decisión de votar sí o no.

Los datos con los que trabaja son:

- Variable de respuesta: votar sí o no.
- Variables explicativas: afiliación al partido (republicano, demócrata o independiente), el estado al que representa cada senador y la cantidad total de dinero de las contribuciones recibidas por los fabricantes de automóviles.

A simple vista puede verse que existe una mayor concentración de los votos positivos en quienes recibieron mayor cantidad de dinero en su campaña electoral y de éstos parece que abundan más los republicanos. La variable dinero recibido ha sido transformada en " $\log_{10} * \text{dinero} + 1$ " para confeccionar el gráfico.

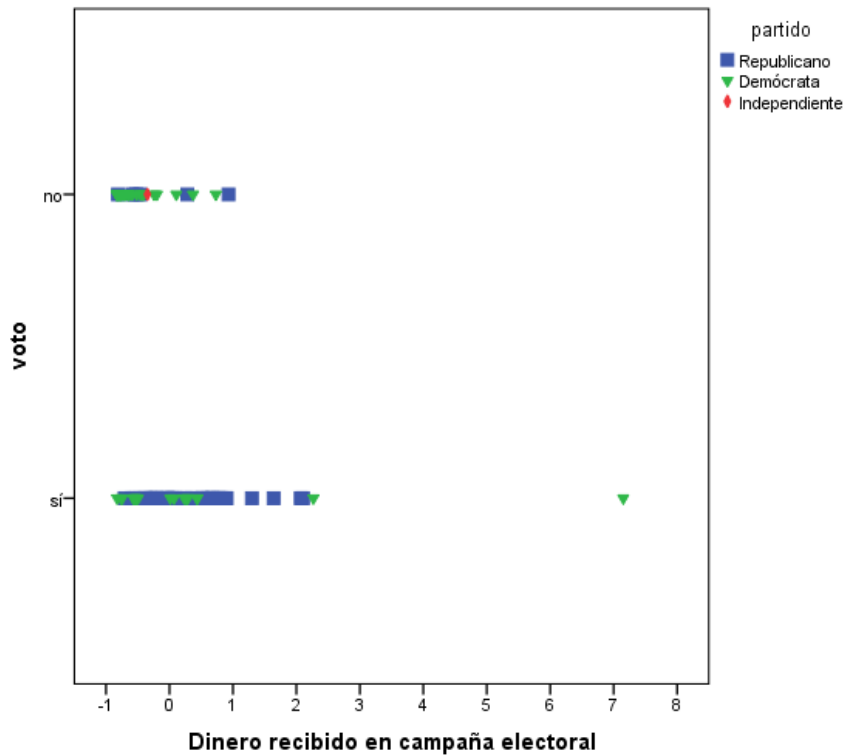


Figura 4. Gráfico de las respuestas de los congresistas en función de su partido de referencia y de la cantidad de dinero recibido en campaña electoral.

También puede evidenciarse en un diagrama de tallo y hojas en el que en el eje central es el llamado tallo que representa en este caso \$10.000. Así un ejemplo de interpretación sería que solo dos demócratas han recibido \$700 en la campaña o que un republicano ha recibido \$35.000. Y como caso extremo existe un demócrata que ha obtenido \$133.000. Se tiene una base más extensa en el grupo de demócratas, mientras que el de republicanos no es tan ancha, pero es más elevada. En cuanto al independiente, no se muestra en el gráfico, pero solo existe un sujeto y éste cobró \$8.000.

Republicanos		Demócratas
	13	3
	12	
	11	
	10	
	9	
	8	
	7	
	6	
	5	1
881	4	
5	3	0116
98876655543222	2	04557888
988855443332000	1	9
99987766544420	0	000000000011111122233344445556677

Figura 5. Diagrama de tallo y hojas que representa las frecuencias y de este modo se registra y ordena el monto de cada senador.

También en el gráfico que se muestra, puede verse la proporción de voto positivo, en la que los votos positivos aumentan, cuando aumenta las subvenciones recibidas.

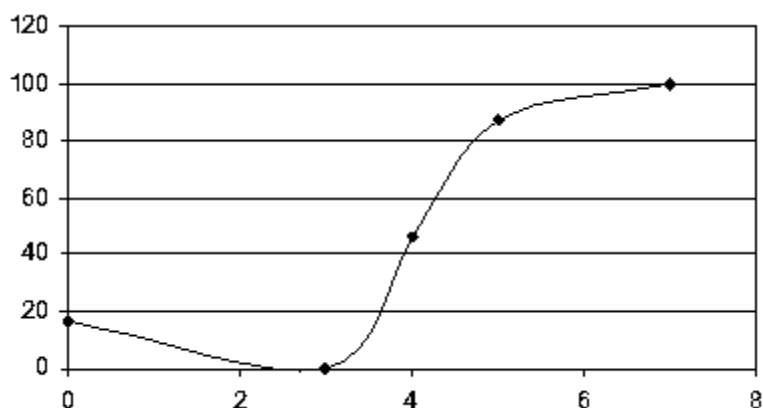


Figura 6. Este gráfico representa en el eje de ordenadas el voto positivo (100) o negativo (0) y su distribución de probabilidad en función del eje de abscisas que representa el dinero recaudado.

En el modelo que propone Scott se centra solo en los republicanos y demócratas, ya que solo tiene un caso de un senador independiente.

- AP. Toma como variable dicotómica la afiliación política (AP) (0 si es demócrata, 1 si es republicano).
- D. El dinero transformado lo toma como variable continua (D).
- p. Es la probabilidad de votar positivo.

Por tanto, el modelo quedaría como sigue:

$$\text{logit}(p) \equiv \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 AP + \beta_2 D + \beta_3 AP \times D$$

Y una vez aplicado el modelo pueden evaluarse las tablas.

Predictor	β	SE	Z	p	Odds	95% IC	
					Ratio	Mín	Máx
Constante	-3.20274	1.64536	-1.95	0.052			
Partido político	-8.04599	6.46223	-1.25	0.213	0.00	0.00	101.49
Dinero	0.61841	0.35038	1.76	0.078	1.86	0.93	3.69
Interacción	2.01934	1.30583	1.55	0.122	7.53	0.58	97.39

Tabla 1. Resultado de la regresión logística incluida la interacción.

Una vez descartado que no existe interacción, $p_{\text{Interacción}} > 0,05$ se decide no incluirla en el modelo, por lo que quedaría como sigue:

$$\text{logit}(p) \equiv \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 AP + \beta_2 D$$

De esta forma se tiene la siguiente tabla

Predictor	β	SE	Z	p	Odds	95% IC	
					Ratio	Mín	Máx
Constante	-4.51004	2.00965	-2.24	0.025			
Partido político	1.91463	0.55837	3.43	0.001	6.78	2.27	20.27
Dinero	0.89897	0.42306	2.12	0.034	2.46	1.07	5.63

Tabla 2. Resultado de la regresión logística excluida la interacción.

En la que se muestra que efectivamente aporta información al modelo el hecho de que se sea de un partido político u otro y de la cantidad de dinero obtenida para su campaña electoral. Y quedaría demostrado por tanto que estas variables influyen en el voto.

Quedaría así la ecuación del modelo:

$$\text{logit}(p) \equiv \ln\left(\frac{p}{1-p}\right) = -4,510 + 1,914AP + 0,899D$$

Así, puede interpretarse que un republicano que haya recibido \$10.000, tras hacer la transformación queda como $\log(100001) \cong 5$

$$\text{logit}(p) = -4,510 + 1,914 + 0,899 * 5 = 1,899$$

Las probabilidades quedan entonces en $e^{1,899} = 6,6825$ Así que puede decirse que

existe una probabilidad de $\frac{6,6825}{(1+6,6825)} = 0,8698$ veces de que se vote sí siendo

demócrata habiendo recibido esa cantidad de dinero.

Cuantificando el “efecto demócrata” puede hacerse una separación entre cuánto difieren las probabilidades de voto en función al partido de pertenencia. De esta forma con el coeficiente

1,914 se tiene que $e^{1,914} = 6,7844$ Así que la probabilidad de votar sí es 6,7844

veces mayor siendo republicano que demócrata con la misma cantidad de contribución en la campaña.

Gracias a estos análisis, puede calcularse, por ejemplo cuánto necesitan invertir las compañías automovilísticas en las campañas de los senadores para asegurar las probabilidades de estar a su favor en futuras votaciones del senado.

3.7.- RECORRIDO HISTÓRICO DE LA TÉCNICA

Ya en el año 1985 (Shepard, Camilli y Williams, 1985) señalan que para la detección del DIF se han propuesto numerosos métodos que abarcan desde aproximaciones centradas en el análisis de varianza hasta el empleo de las técnicas derivadas de la TRI. Y ya García-Cueto (1993) exponía que los procedimientos para el estudio del sesgo de los test basados en la TRI aparecen como más adecuados y más potentes que cualquier otro tipo de procedimiento.

García y Marañón (1996) opinan que los mejores y más prometedores métodos son los basados en la TRI junto con el método de Mantel-Haenzel (Holland & Thayer, 1988). Aunque posteriormente el análisis de regresión logística se ha consolidado como una de las técnicas más utilizadas en la detección de ítems con funcionamiento diferencial respecto a grupos que difieren en variables lingüísticas, culturales o sociales (Gómez-Benito, Hidalgo, Padilla y González, 2005).

Este método es fácil de implementar y tiene asociado un test de significancia, pero no resulta útil ni adecuado para la detección de DIF no uniforme (Hambleton & Rogers, 1989; Rogers & Swaminathan, 1993). Y para detectar el DIF no uniforme se han sugerido modelos log-lineales (Shepard, Camilli y Williams, 1985).

Swaminathan y Rogers (1990) proponen un procedimiento de regresión logística para el análisis de DIF, el cual es efectivo para detectar si existe DIF uniforme o no uniforme. Está expresado por:

$$P(u = 1|X, G) = \frac{e^z}{(1 + e^z)}$$

donde

$$z = \beta_0 + \beta_1 X + \beta_2 G + \beta_3 (XG)$$

Aquí X es el nivel de aptitud observado de un individuo y G es la pertenencia a uno de los grupos (1 para el grupo de referencia y 0 para el grupo focal). El parámetro β_2 es la diferencia entre los grupos en el desempeño del ítem, mientras que el parámetro β_3 corresponde a la interacción entre aptitud y grupo. Un ítem presenta DIF uniforme cuando el efecto del grupo (G) resulta estadísticamente significativo, siempre que la interacción de la habilidad y el grupo (XG) no ejerza efecto alguno sobre el ítem, o sea cuando $\beta_2 \neq 0$ y $\beta_3 = 0$. Además presentará DIF no uniforme si la interacción resulta estadísticamente significativa, o sea

$$\beta_3 \neq 0.$$

Esta forma de detección del DIF salva la necesidad de tener un grupo focal, ya que no se limita a comparar dos grupos, sino que pueden compararse con más de un grupo focal, lo que la hace ideal para variables con más de una categoría. Además esta técnica no tiene ninguna limitación conocida para trabajar con variables continuas, lo que la hace atractiva en ese sentido.

El equipo de Swanson utiliza la regresión logística como técnica para detectar el DIF (Swanson, Clauser, Case, Nungester y Featherman, 2002), aunque no se logra la distribución esperada con muestras de ítems muy discriminantes, la regresión logística se demuestra como más poderosa para detectar DIF no uniforme y tan potente como el procedimiento de Mantel-Haenszel para la detección de DIF uniforme (Rogers & Swaminathan, 1993). Además es una técnica promovida por diferentes estudios para extender su uso (Balluerka, Plewis, Gorostiaga y Padilla, 2015). Incluso se habla de regresión logística de clase latente como método de tercera generación para detectar el DIF (Zumbo y cols., 2015).

Con la regresión logística se han propuesto varias estrategias:

1ª Comparación de modelos anidados. Se ajustan tres modelos en etapas; la primera etapa se ajusta el *modelo base* de ausencia de DIF y se introduce en la ecuación la puntuación total del sujeto en el test (X) (la respuesta al ítem sólo depende de la habilidad del sujeto); la segunda etapa es la correspondiente al *modelo de DIF uniforme* en la que se añade a la ecuación una variable correspondiente al grupo de pertenencia (G); la tercera y última etapa consiste en añadir a la ecuación la interacción entre el grupo y la puntuación total del test resultando en la valoración del ajuste del modelo de *DIF no uniforme* o también llamado

modelo completo. En la tabla 3 pueden verse los modelos a ajustar para la detección del DIF y su correspondiente parametrización. En este proceso de ir añadiendo componentes a la ecuación, se evalúa la mejora del modelo.

Modelo básico	Etapa 1 (modelo 1)	$z = \beta_0 + \beta_1 X$
Modelo de DIF uniforme	Etapa 2 (modelo 2)	$z = \beta_0 + \beta_1 X + \beta_2 G$
Modelo de DIF no uniforme	Etapa 3 (modelo 3)	$z = \beta_0 + \beta_1 X + \beta_2 G + \beta_3 (XG)$

Tabla 3. Modelos de regresión logística para analizar en la detección del DIF.

2ª Prueba simultánea de DIF uniforme y no uniforme. Se sopesan ambas hipótesis para la su detección y no diferencia entre ellas, así que no se distingue si se ha detectado DIF, qué tipo de DIF se ha detectado. Se obtiene comparando el valor de verosimilitud del modelo básico con el modelo con interacción o completo.

3ª Ajuste del modelo completo. En el que se incluyen todos los términos, se evalúa su significación en los coeficientes de Wald. Pueden salir varios resultados: que solo β_2 sea significativo, lo que indicaría DIF uniforme; que solo sea significativo el coeficiente de la interacción β_3 indicaría DIF no uniforme simétrico y si tanto β_2 como β_3 sin significativos el ítem estaría afectado por DIF no uniforme asimétrico.

Estrategia.	Método.	Objetivo de la prueba.
Comparación de modelos anidados.	1°. Comparar el valor de la verosimilitud del modelo 1 con el del modelo 2.	Detectar DIF uniforme.
	2°. Comparar el valor de la verosimilitud del modelo 2 con el del modelo 3.	Detectar DIF no uniforme.
Prueba simultánea de DIF uniforme y no uniforme.	Comparar el valor de verosimilitud del modelo 1 con el del modelo 3.	Detectar DIF sin especificar su tipo.
Ajustar el modelo completo.	1°. Estimar si solo β_2 es significativo en el modelo 3.	Detectar DIF uniforme.
	2°. Estimar si solo β_3 es significativo en el modelo 3.	Detectar DIF no uniforme simétrico.
	3°. Estimar si tanto β_2 como β_3 son significativos en el modelo 3.	Detectar DIF no uniforme asimétrico.

Tabla 4. Estrategias para la detección de DIF usando regresión logística.

3.8.- EL PROBLEMA DE LA SEPARACIÓN.

En realidad no es un problema como tal, porque los grupos se separan perfectamente y eso implementado en un modelo, hace que el modelo sea aún mejor. Pero los modelos con los que se trabaja se enfocan en que si una variable es demasiado buena, seguramente se encuentre contaminada por la variable objetivo.

Lo que ocurre con la regresión logística cuando hay una separación es que los estimadores máximo verosímiles no existen y tienden a infinito. Este fenómeno se da en el proceso de ajuste de un modelo logístico, es la división completa de la respuesta (aquí codificada como presente o ausente). Se observa cuando los estimadores del modelo convergen a un valor infinito. O dicho de otro modo, existe separación completa cuando dos grupos son divididos completamente. De esta manera pueden aparecer en el análisis valores anómalos. De tal forma que si la variable de respuesta se distribuye en 0 y 1, en uno de los grupos se obtiene en todos los puntos de la muestra el evento de interés y en el otro grupo ocurre este evento sin que exista una mezcla de éxitos o fracasos, o sea, en el caso binario existe región de respuestas negativas o codificadas como 0 (R1), región de separación en las que no existen

respuestas mezcladas (RS) y región de respuestas positivas o codificadas como 1 (R2). Así se separa totalmente los éxitos de los fracasos o superación o no superación del ítem (Correa, Juan y Valencia, 2011).

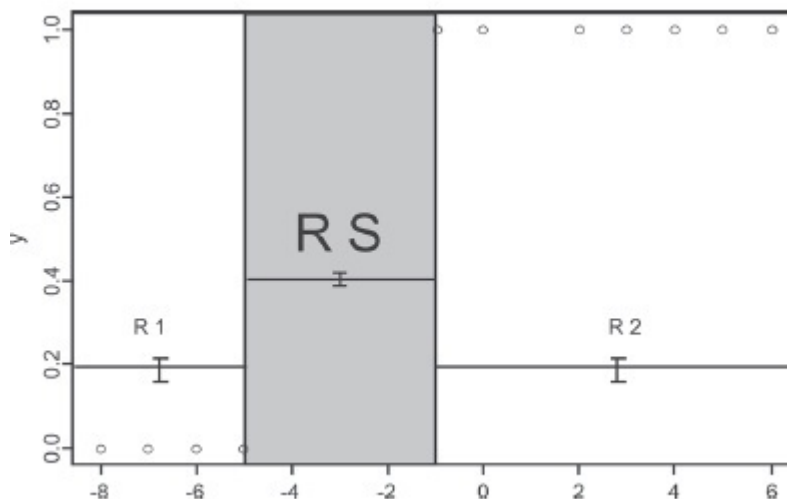


Figura 7. Explicación visual de separación de los grupos en las respuestas sobre un ítem donde la zona R1 representa las puntuaciones 0, la zona RS puntuaciones mezcladas y zona R2 puntuaciones positivas.

La separación normalmente se da en muestras con covariables muy predictivas. Entonces si un determinado grupo tiene siempre unas respuestas y no hay variabilidad en las mismas, al estimar el modelo, dará una separación perfecta, lo que implica que la estimación no sea buena y no podrá calcularse la máxima verosimilitud en el modelo.

El caso más simple es la separación en una tabla 2x2 en el que alguna de las casillas de las frecuencias sea cero. No es muy común que se trate este problema aunque existiese y ni siquiera se alerta de ello. El programa SPSS (IBM Corp., 2012) tanto en su versión 21 como en las sucesivas hasta la fecha, es inservible al utilizar la regresión logística por el mero hecho de que cuando un sujeto contesta todas sus respuestas como positivas o negativas, o sea, cuando los datos son demasiado homogéneos, el modelo que propone este programa da una separación perfecta y el modelo tiende a infinito.

Firth (1993) propone un método para reducir el sesgo producido en la regresión logística en el que siempre conduce a estimaciones de los parámetros finitos. En él trata de ofrecer una solución ideal a través de estimaciones finitas de los parámetros. Los test de razón de verosimilitudes también pueden ser adaptados, como los test de Wald y los intervalos de confianza al método penalizado.

Heinze y Schemper (2002) muestran que la separación depende del tamaño de la muestra, además del número de factores de riesgo dicotómicos, la magnitud de los Odds Ratio que se asocian con ellos y también depende del grado de equilibrio en su distribución.

TAMAÑO MUESTRAL	NÚMERO DE FACTORES DE RIESGO	$B_x^* = 1 : 1$				$B_x^* = 1 : 4$			
		Odds ratio				Odds ratio			
		1	2	4	16	1	2	4	16
30	3	0	3	10	53	17	25	43	74
	5	2	7	24	75	30	41	58	85
	10	12	38	78	98	56	71	86	98
50	3	0	0	1	18	2	5	15	46
	5	0	0	2	32	6	9	22	53
	10	0	1	20	78	10	19	36	74

Tabla 4. Muestra la probabilidad de separación en tanto por ciento, para la regresión logística con factores de riesgo dicotomizados. Ejemplo basado en 1000 muestras. El saldo marginal esperado de las respuestas y no respuestas está Fijado en 1:1.

Utilizando el programa SPSS no detecta DIF cuando en realidad lo hay.

Se ha seleccionado el ítem 18 del test MMSE “Si tiene 30 pesetas y me va dando de 3 en 3, ¿cuántas pesetas le van quedando?” Este ítem en esta variable tiene funcionamiento diferencial uniforme, pero no se evidencia en las tablas que ha sacado SPSS. De todas formas se seguirá el proceso para evidenciar las fallas.

3.8.1.- DIF NO UNIFORME CON SPSS

En un primer momento se evalúa si existiera o no un funcionamiento diferencial no uniforme en el ítem, por lo que se considera, además de las covariables “puntuación total del test” y “educación”, la interacción entre ambas “puntuación total del test * educación”. Resultando en la siguiente salida de SPSS:

Variables en la ecuación

	β	E.T.	Wald	Grados de libertad	Sig.	Exp(β)	I.C. 95% para EXP(β)	
							Inferior	Superior
							Puntuación MMSE	0,000
Educación			1,151	3	0,765			
Estudios mínimos	-51,132	168667,271	0,000	1	1,000	0,000	0,000	.
Estudios primarios	-41,555	168667,271	0,000	1	1,000	0,000	0,000	.
Bachiller y estudios superiores	-45,947	168667,271	0,000	1	1,000	0,000	0,000	.
Interacción educación con puntuación del MMSE			1,138	3	0,768			
Interacción de estudios mínimos con puntuación MMSE	1,121	5870,523	0,000	1	1,000	3,069	0,000	.
Interacción de estudios primarios con puntuación MMSE	0,788	5870,523	0,000	1	1,000	2,199	0,000	.
Interacción de bachiller y estudios superiores con la puntuación del MMSE	0,969	5870,523	0,000	1	1,000	2,636	0,000	.
Constante	21,204	168667,271	0,000	1	1,000	1616640		
						210,820		

a. Variables introducidas en el paso 1: Puntuación del MMSE, Educación, Interacción de la educación y de la puntuación del MMSE.

Tabla 5. Resultados del programa SPSS en el que se muestra la regresión logística y se incluye como variable dependiente el acertar o no el ítem y como variables independientes la variable educación subdividida en los diferentes grados de estudios, la puntuación total del test y la interacción entre la educación y la puntuación total del test.

En un primer momento ninguna interacción ha resultado significativa, así que sin hacer ningún otro análisis, se pasa a analizar el DIF uniforme o el modelo sin interacción.

3.8.2.- DIF UNIFORME CON SPSS

En este modelo, el programa se atreve a afirmar que selecciona correctamente el 87% de los casos:

Tabla de clasificación*

Observado	Pronosticado			
	18 Cálculo: 65		Porcentaje correcto	
	0	1		
Paso 1	0	99	30	76,7
	1	15	201	93,1
Porcentaje global				87,0

a. El valor de corte es 0,500

Tabla 6. Clasificación del modelo logístico.

De las 99+30 personas que han no han superado el ítem, 99 de ellas han sido bien clasificadas, es decir el porcentaje de aciertos es del:

$$\frac{99}{99 + 30} * 100 = 76,7\%$$

Y de las 15+201 personas que han superado el ítem, 201 de ellas se ha sido correctamente clasificadas por el modelo, así que su porcentaje de aciertos se calcula de igual manera:

$$\frac{201}{201 + 15} * 100 = 93,1\%$$

Y haciendo el porcentaje global se obtiene que:

$$\frac{99 + 201}{99 + 201 + 30 + 15} * 100 = 87\%$$

El modelo afirma que clasifica correctamente el 87% de los casos globales en el ítem. Esto es que usando la ecuación de regresión y de los datos observados, se realiza una predicción y ofrece un valor pronosticado. La predicción se compara con los valores observados de la forma en que aparece en la tabla. Cuantos más casos correctamente clasificados, mejor predictor se estimará el modelo y el porcentaje de casos correctamente clasificados será mayor. En este caso, el programa afirma que clasifica correctamente el 87% de todos los casos, incluyendo los clasificados en el ítem como acertados y los clasificados como fallados. Este es un porcentaje elevado y se daría por bueno el análisis, pero en realidad no es así, ya que al ver la tabla de variables en la ecuación pueden verse anomalías:

VARIABLES EN LA ECUACIÓN

	β	E.T.	Wald	Grados de libertad	Sig.	Exp(β)	I.C. 95% para EXP(β)	
							Inferior	Superior
Puntuación MMSE	0,896	0,103	76,219	1	0,000	2,449	2,003	2,994
Educación			4,504	3	0,212			
Estudios mínimos	-20,608	7386,674	0,000	1	0,998	0,000	0,000	.
Estudios primarios	-19,846	7386,674	0,000	1	0,998	0,000	0,000	.
Bachiller y estudios superiores	-19,481	7386,674	0,000	1	0,998	0,000	0,000	.
Constante	-3,319	7386,674	0,000	1	1,000	0,036		

a. Variables introducidas en el paso 1: Puntuación del MMSE y educación en sus diferentes categorías.

Tabla 7. Resultados del programa SPSS en el que se muestra la regresión logística y se incluye como variable dependiente el acertar o no el ítem y como variables independientes la variable educación subdividida en los diferentes grados de estudios y la puntuación total del test.

Ningún subgrupo resulta como significativo, aunque utilizando otros métodos se sabe que comparando el grupo de personas que tienen bachiller y personas que tienen estudios superiores, con las personas analfabetas y que solo saben leer y escribir se tiene una diferencia achacable al funcionamiento diferencial entre los subgrupos.

El estadístico de Wald se encarga de contrastar la hipótesis de que uno de los coeficientes, cuando funciona solo, es distinto de 0 (media 0 y desviación típica 1). Viene descrito por la fórmula:

$$W(b_j) = \left(\frac{\hat{b}_j}{\sigma(b_j)} \right)^2$$

Se distribuye como una χ^2 con un grado de libertad; de esta manera, todos los coeficientes que tengan un $W(b_j) > \chi^2$ serán significativos y se estimará que debe ser incluido en el modelo.

Si fuese significativo significaría que el coeficiente es diferente de 0 y se estimaría integrarlo en el modelo. Pero puede proporcionar falsas ausencias de significación o incremento de error tipo II siempre y cuando los errores típicos asociados sean grandes y justamente es lo que se da.

En la columna del intervalo de confianza para el exponencial de β , puede verse que se encuentran mínimos, pero no se encuentran máximos. Por eso el programa devuelve un punto (.), porque el valor obtenido no es alcanzable para el programa y lo considera infinito.

Para visionar qué ocurre exactamente, puede evidenciarse con el siguiente gráfico de caja y bigotes.

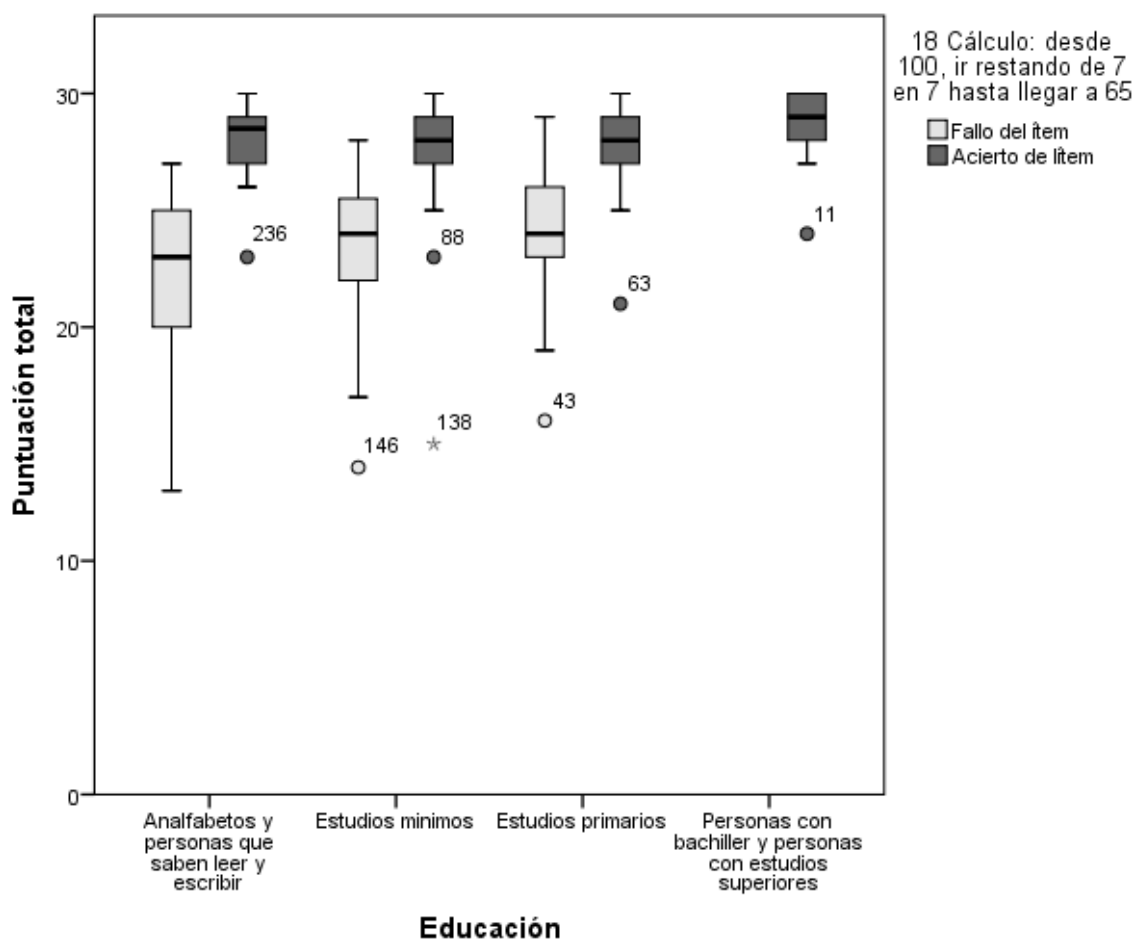


Figura 8. Diagrama de caja y bigotes que representa la educación y la puntuación total. Se evidencia ausencia de puntuaciones de fallo en las personas con bachiller y personas con estudios superiores.

La variable educación se divide en 5 subgrupos y se compara el primero con los demás. Pero en los datos recogidos ninguna persona del grupo superior presenta ninguna respuesta de

fallo al ítem. Por eso no existe el gráfico de caja y bigotes que debería aparecer en gris claro en la columna de personas que tienen bachiller y personas con estudios superiores.

3.8.3.- PROPUESTAS PARA EL PROBLEMA DE LA SEPARACIÓN

La escasez de datos puede hacer que se presente este problema, así que después de comprobar que no se pueden obtener más datos para el análisis, en los que con suerte puedan obtenerse frecuencias absolutas diferentes de cero y después de descartar la opción de agregar un par de observaciones ficticias (Correa y cols., 2011), se opta por buscar alternativas al problema de la separación. Éstas vienen de la mano de Firth y del método bayesiano usando la función *MCMClogit* que viene dentro de la función *MCMCpack* (Martin, Quinn y Park, 2011).

EL MÉTODO BAYESIANO

Requiere datos con distribución binomial así que se presentará de este modo:

$$L(\beta|Y, X) = \prod_{i=1}^n f(y_i|X, \beta) = \prod_{i=1}^n (\pi(x_i))^{y_i} (1 - \pi(x_i))^{1-y_i}$$

donde $\pi(x_i) = \frac{1}{1+e^{-x_i'\beta}}$

Quedando la función de verosimilitud como sigue:

$$L(\beta|Y, X) = \prod_{i=1}^n \left[\left(\frac{1}{1+e^{-x_i'\beta}} \right)^{y_i} \left(1 - \frac{1}{1+e^{-x_i'\beta}} \right)^{1-y_i} \right]$$

Lo cual lleva a

$$L(\beta|Y, X) = \prod_{i=1}^n \left[\left(\frac{e^{x_i'\beta}}{e^{x_i'\beta} + 1} \right)^{y_i} \left(\frac{1}{e^{x_i'\beta} + 1} \right)^{1-y_i} \right] = \prod_{i=1}^n \left[\frac{(e^{x_i'\beta})^{y_i}}{e^{x_i'\beta} + 1} \right]$$

La función a priori para los parámetros del modelo a estimar es una distribución bivalente:

$$\beta \sim MN(\beta_0, \Sigma) \text{ donde se usará la matriz de precisión } \mathbf{T} = \Sigma^{-1}$$

Así la función a posteriori es:

$$\xi(\beta | \text{datos}) \propto MN(\beta_0, \Sigma) * L(\beta | Y, X)$$

Y de esta forma:

$$\xi(\beta | \text{datos}) \propto L(\beta | Y, X) = \prod_{i=1}^n \left[\frac{(e^{X_i' \beta})^{y_i}}{e^{X_i' \beta} + 1} \right] * \exp^{-\frac{1}{2}(\beta - \beta_0)' T (\beta - \beta_0)}$$

Puede desarrollarse usando el algoritmo *Metropolis*, integrado dentro del paquete de R *MCMCpack*.

EL MÉTODO DE FIRTH

El modelo óptimo que puede llegar a estimarse viene dado por los parámetros $\hat{\beta}$ y se obtienen por la solución a la ecuación:

$$\frac{\partial \log V(\beta)}{\partial \beta} \equiv U(\beta) = 0$$

en la que V es la función de verosimilitud. Y actualizando el vector de parámetros en cada iteración, se obtiene que:

$$\beta_{l+1} = \beta_l + I(\beta_l)^{-1} U(\beta_l) = \beta_l + (X' V_l X)^{-1} X' (y - \pi)$$

donde $I(\beta_l)$ es la matriz de información.

Firth plantea basar las estimaciones en ecuaciones modificadas para el j -ésimo parámetro con objeto de reducir el sesgo de la muestra.

$$U^*(\beta_j) = U(\beta_j) + \frac{1}{2} \text{traza} \left[I(\beta_l)^{-1} \left\{ \frac{\partial I(\beta)}{\partial \beta_j} \right\} \right]$$

De esta forma, la función $U^*(\beta_j)$ es la función modificada y está relacionada con la función de verosimilitud corregida. Y así es como se logra eliminar el sesgo de los estimadores máximo verosímiles.

Aplicando esta idea al modelo logístico se tiene que el vector de primeras derivadas

$$U(\beta_l) = X'(y - \pi)$$

Puede reemplazarse en la ecuación por

$$U^*(\beta_l) = X' \left(y - \pi + \frac{1}{2} h - h * \pi \right)$$

Y en esta ecuación * pasaría a ser el producto de dos vectores elemento a elemento y donde h es un vector formado por los elementos de la diagonal de la matriz de influencia de las observaciones.

$$H = V^{1/2} X (X' V_l X)^{-1} X' V^{1/2}$$

y

$$V = \text{diag}\{\pi_i(1 - \pi_i)\}$$

Pasando a ser cada iteración del algoritmo como sigue:

$$\beta_{l+1} = \beta_l + I(\beta_l)^{-1} U^*(\beta_l) = \beta_l + (X' V_l X)^{-1} X' \left(y - \pi + \frac{1}{2} h - h * \pi \right)$$

De esta manera se elimina por completo el problema de la separación, pero no se han solucionado todos los problemas, ya que al igual que con el método lineal general, pueden

ocurrir problemas de multicolinealidad o distribución de variables independientes casi degeneradas. La base de la estimación de los errores estándar son las raíces de los elementos de la diagonal de $I(\beta_l)^{-1}$. De forma análoga, el método de máxima verosimilitud para datos multimodales en muestras pequeñas puede verse en los trabajos de Shelley B. Bull (Bull, Mak y Greenwood, 2002).

Propone la ecuación que se tiene como de indicadora para medir la gravedad de la separación, en el que el modelo logístico se basa con una sola variable predictora así:

$$\pi_i = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x))}$$

También es importante explicar el índice de longitud J_ℓ ya que relaciona el rango de la región de separación (\mathcal{R}_S) y el rango completo de la variable predictora x (\mathcal{R}).

$$J_\ell = \frac{\mathcal{R}_S}{\mathcal{R}}$$

Este índice es un indicador que compara la longitud del área en la que no existen observaciones con el área entera (rango completo de la variable predictora X). Es un indicador normalizado en el que cuanto más próximo a cero está, la separación no sería muy grave, pero la separación sería severa si $J_\ell = 1$, ya que $\mathcal{R}_S \rightarrow \mathcal{R}$, es el motivo por el que no es fácil encontrar el verdadero signo de β_1 , la probabilidad del modelo verdadero puede ser tanto creciente como decreciente, lo que merece considerar tanto el total de éxitos como el de fracasos además de la naturaleza del problema.

Comparando la gravedad de la separación, Correa (2011) explica en una tabla las estimaciones de las diferencias entre los parámetros $E(\hat{\beta}_0)$ y $E(\hat{\beta}_1)$ con los dos métodos propuestos más un tercero que trata en agregar puntuaciones ficticias a los datos para que no exista la separación. Al método bayesiano se le agregó datos ficticios a un 28% de la región de separación. Al método clásico también se le agregó el mismo par de puntuaciones ficticias

mostrando el sesgo encontrado en cada caso. Para todas las soluciones probadas, se han implementado valores de $\hat{\beta}_0 = 0,1$ y $\hat{\beta}_1 = 0,2$.

# Pares de observaciones ficticias	N=20		N=40		N=60	
	Método de Firth					
	$E(\hat{\beta}_0)$ (%sesgo)	$E(\hat{\beta}_1)$ (%sesgo)	$E(\hat{\beta}_0)$ (%sesgo)	$E(\hat{\beta}_1)$ (%sesgo)	$E(\hat{\beta}_0)$ (%sesgo)	$E(\hat{\beta}_1)$ (%sesgo)
1	0,1860 (82,6)	0,12367 (38,16)	0,2015 (101,46)	0,1546 (22,71)	0,13981 (0,498)	0,08258 (0,55)
Método Bayesiano MCMClogit, con ficticias						
1	0,06681 (33,19)	0,2242 (12,085)	0,1294 (29,4)	0,2553 (27,65)	0,1631 (0,63)	0,1 (0,49997)
Método Clásico, con ficticias						
1	0,15853 (58,53)	0,1064 (46,817)	0,181 (80,777)	0,135 (32,478)	0,14498 (0,45)	0,090 (0,55)

Tabla 8. Comparación de las estimaciones de los parámetros utilizando diferentes técnicas para solucionar la separación.

De la tabla puede concluirse que a pesar de tener pocas observaciones ficticias, existe posibilidad de obtener la solución al modelo que se plantea, aunque no en los extremos, sino en la región de separación.

La corrección de Firth está implementada en el paquete *logistf* (Heinze y cols., 2013) del programa R (R Core Team, 2014). Este paquete trata la separación restringiendo los parámetros para que no se vayan al infinito, añade una restricción usando el método de Firth (1993) para aplicar esa restricción.

Se elige el método de Firth para trabajar porque está publicado, implementado en R y es efectivo, ya que la corrección de Firth es una penalización en el tamaño de los coeficientes que impide que ninguno de ellos crezca hacia el infinito.

4.- MUESTRA

Se ha recogido información de 347 sujetos (153 varones y 194 mujeres) en la que se incluyen las puntuaciones en el test, la edad, el sexo y el nivel de estudios. Esta información ha sido

recogida entre 2007 y 2009 por un equipo de médicos en el centro de salud de pizarrales de Salamanca (España). Son sujetos que han acudido a dicho centro de salud y, según criterio médico, se ha estimado conveniente la inspección del deterioro cognitivo a través del test MMSE. Por tanto se trata de un estudio analítico observacional en el que el proceso de selección no ha sido controlado, ya que las personas acceden por su cuenta y no han sido llamadas. Por consiguiente se estudia a la población demandante de servicios.

Como no se ha utilizado ningún método en los que intervenga el azar para seleccionar a los sujetos en grupos, se trata de un muestreo no probabilístico.

No se ha utilizado ningún tipo de enmascaramiento para la consecución de los datos.

5.- METODOLOGÍA

Se utiliza una regresión logística binaria para realizar la detección del DIF.

En el caso que nos atañe tenemos los datos concernientes al acierto o al fallo de cada ítem del test además de variables demográficas como edad, sexo y nivel de estudios. Así pues se codifica como 0 que el sujeto no haya dado una respuesta adecuada para superar el ítem, por lo que se denominará fallo, mientras que se codifica como 1 cuando se ha superado el ítem y se ha obtenido una respuesta satisfactoria.

Por tanto en la regresión logística para el modelo, la variable dependiente será el acierto o el fallo del sujeto en el ítem, mientras que las variables independientes serán en este caso la edad, el sexo y el nivel de estudios. Se utilizará un modelo diferente para cada variable para no emponzañar el análisis.

El nivel de estudios es una variable independiente que puede tomar más de dos valores, así que se denominará como categórica y se distribuirá en:

- Analfabetos junto con las personas que pueden leer y escribir,
- Personas que han superado los estudios mínimos,
- Personas que hayan superado los estudios primarios,
- Y por último personas con bachiller y personas con estudios superiores.

De esta forma se obtienen cuatro grupos o categorías dentro de esta variable. Dichas categorías son mutuamente excluyentes

El sexo se codifica como una variable binaria o dicotómica, donde el hombre se codifica como 1 y la mujer como 2.

La edad en este caso se considera como una variable continua, en la que se comparan saltos consecutivos de un año.

6.- RESULTADOS

A continuación se detalla el conjunto de ítems señalando cuáles presentan funcionamiento diferencial clasificándose en uniforme o no uniforme y en relación a las variables estudiadas.

n°	Ítem	EDAD		EDUCACIÓN		SEXO	
		p No Uniforme	p Uniforme	p No Uniforme	p Uniforme	p No Uniforme	p Uniforme
1	Año	0,7438	0,5747	0,9663	0,5937	0,9367	0,9471
2	Estación	0,8632	0,2124	0,1855	0,3181	0,5715	0,3973
3	Día Mes	0,1288	<u>0,0365</u>	0,1162	0,2263	0,3817	0,7008
4	Día Semana	0,9455	0,2905	0,9088	0,8476	0,8969	0,4829
5	Mes	0,7070	0,1084	0,9235	0,6577	0,7127	0,1257
6	Región	0,9710	<u>0,0414</u>	0,9112	0,1352	0,0835	0,5411
7	Provincia	0,7076	<u>0,0344</u>	0,9680	0,2545	0,5743	0,2210
8	Ciudad	0,2318	0,6366	0,8251	0,4098	0,6309	0,7147
9	Edificio	0,3707	<u>0,0024</u>	0,1947	0,3497	0,8469	<u>0,0431</u>
10	Piso	0,6451	0,8375	0,8675	<u>0,0308</u>	0,4669	0,2874
11	Peseta	0,7306	0,5682	0,9137	0,7834	0,9036	0,3084
12	Caballo	0,7306	0,5682	0,9137	0,7834	0,9036	0,3084
13	Manzana	0,6832	0,9459	0,9393	0,9454	0,8571	0,1196
14	N93_O	<u>0,0301</u>		0,2284	0,7801	0,4244	<u>0,0322</u>
15	N86_D	<u>0,0368</u>		0,1848	0,7933	0,6004	0,0533
16	N79_N	<u>0,0007</u>		0,9054	0,2878	0,0658	<u>0,0417</u>
17	N72_U	<u>0,0050</u>		0,3057	<u>0,0493</u>	0,3604	0,0822
18	N65_M	<u>0,0064</u>		0,7357	<u>0,0244</u>	0,2264	<u>0,0093</u>
19	Peseta_R	0,0634	<u>0,0325</u>	0,3973	0,1503	0,4537	0,5670
20	Caballo_R	0,9204	0,5480	0,0596	0,9152	<u>0,0271</u>	
21	Manzana_R	<u>0,0113</u>		<u>0,0002</u>		<u>0,0052</u>	
22	Bolígrafo	0,8564	0,4298	0,9779	0,5544	0,6703	0,6393
23	Reloj	0,6956	0,2558	0,9630	0,8313	0,5461	0,5744
24	"Nada de síes"	0,2656	0,2116	0,2152	0,8888	0,3582	0,4643
25	Coja	0,2891	0,7236	0,8350	0,3636	0,4004	0,7496
26	Doble	0,9939	0,8226	0,7317	0,9591	0,6465	0,2542
27	Suelo	<u>0,0114</u>		0,5999	0,0784	0,1452	0,3334
28	Cierre Ojos	0,1291	0,7125	0,9859	0,5804	0,1351	<u>0,0017</u>
29	Frase	0,6860	0,2836	0,2785	0,9186	0,4407	0,4605
30	Pentágonos	0,8456	0,0435	0,0694	0,9352	0,4891	0,8335

Tabla 9. Valores p de la prueba de Funcionamiento Diferencial de los Ítems tanto uniforme como no uniforme en las variables edad, educación y sexo.

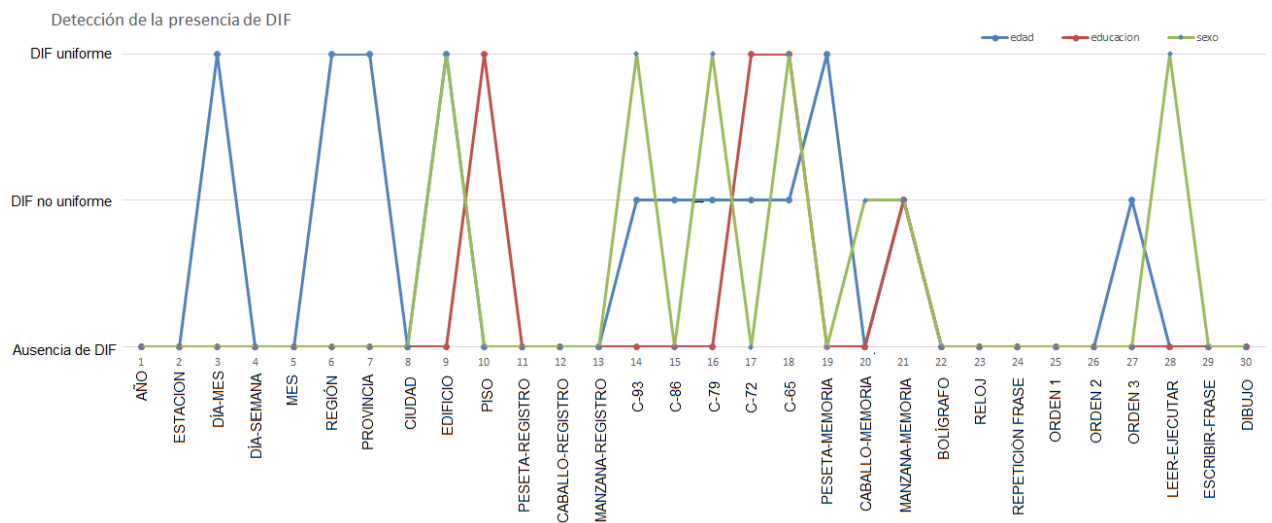


Figura 9. Gráfico de la detección de la presencia de DIF en función del sexo, educación y edad.

En la figura 9 puede verse la distribución del DIF en las variables estudiadas.

Seguidamente se pasará a analizar la detección del DIF uniforme y no uniforme en relación al nivel de estudios.

6.1- RESULTADOS CON LA VARIABLE EDUCACIÓN

Para el ítem 21 en la variable educación se ha detectado DIF no uniforme.

El análisis se hace en dos pasos:

- En el primero de ellos se busca si hubiese DIF no uniforme, si hubiese DIF no uniforme, querría decir que se ha detectado que las diferencias en la interacción son significativas. Una vez que se ha detectado el DIF no uniforme, los coeficientes de β_0 , β_1 y β_2 dejan de ser interpretables y solo pasa a interpretarse la interacción.
- En el segundo paso se evalúa la existencia de DIF uniforme. De existir DIF uniforme quisiera decir que existe una diferencia en el desempeño de los grupos, pero que esta diferencia no es solo atribuible a los propios grupos, sino a la interacción del grupo con la otra variable independiente, que tratándose del MMSE, sería la puntuación total.

Al detectarse DIF no uniforme cuando se está analizando la influencia del nivel de estudios en el la puntuación de un ítem, puede afirmarse que existe una diferencia en la probabilidad de

acertar el ítem en los diferentes niveles de estudio, pero depende además de la puntuación del test. O sea, que no es la misma dependiendo del grado de deterioro de la persona.

Una vez conocida si la diferencia es significativa o no, pueden pasar a hacerse las pertinentes conclusiones con respecto a los grupos. De esta forma, se evidencia que existen diferencias significativas entre la interacción del grupo de estudios mínimos, con el grupo con el que se compara ($p=0,0033$), que es el de analfabetos y personas que saben leer y escribir. Si el coeficiente fuese negativo, querría decir que la probabilidad de acertar el ítem es más baja en el grupo de analfabetos y personas que saben leer y escribir interaccionando con la puntuación total, pero en este caso es positivo. Así pues, quiere decir que la interacción con la puntuación total y el desempeño de las personas analfabetas y que saben leer y escribir es mayor que las personas, o sea, que las diferencias encontradas no solo dependen de la educación, sino que también dependen del deterioro.

Esta columna de coeficientes es la β del modelo completo

$$z = \beta_0 + \beta_{1.1}X + \beta_{1.2}X + \beta_{1.3}X + \beta_{1.4}X + \beta_2 + \beta_{3.1}(XG) + \beta_{3.2}(XG) + \beta_{3.3}(XG) + \beta_{3.4}(XG)$$

Así $\beta_{1.1}$ es el coeficiente en la fórmula de las personas que son analfabetas y de las que saben leer y escribir,

$\beta_{1.2}$ es el coeficiente en la fórmula de las personas con estudios mínimos,

$\beta_{1.3}$ es el coeficiente en la fórmula de las personas con estudios primarios y

$\beta_{1.4}$ es el coeficiente en la fórmula de las personas con bachiller y estudios superiores.

Mientras que $\beta_{3.1}$ es el coeficiente en la fórmula de la interacción entre personas que son analfabetas y personas que saben leer y escribir. Y de forma análoga, el segundo subíndice de β indica el nivel de estudios para la interacción.

	coeficiente	S.E.(coef)	mín 0.95	máx 0.95	Chi ²	p
(Intercepto)	-6,7949	2,9272	-13,4242	-1,7424	7,5593	0,0060
Puntuación total	0,2398	0,1135	0,0406	0,4923	5,7991	0,0160
Interacción Estudios mínimos con Puntuación total	0,1134	0,1504	-0,1951	0,4057	0,5533	0,4570
<u>Interacción Estudios primarios con Puntuación total</u>	0,5848	0,2000	0,1975	0,9947	8,6601	0,0033*
<u>Interacción Bachiller y Estudios superiores con Puntuación total</u>	2,6796	1,3007	0,8202	7,4805	11,5475	0,0007*

Prueba de razón de verosimilitud=93,70893 con 7 grados de libertad, p=0, n=345

Tabla 10.1. Prueba de Funcionamiento Diferencial de los ítems No uniforme. Ítem 21 en relación al nivel de estudios.

La significación es menor de 0,05 de Chi² indica que el modelo ayuda a explicar el evento, o lo que es lo mismo, que las variables independientes ayudan a explicar la variable dependiente y por tanto teniendo en cuenta esta variable en el modelo, ayuda a construir un modelo que se adapta más a la realidad.

	exponente(coeficiente)	1/exponente(coeficiente)
(Intercepto)	0,00110	893,29170
Puntuación total	1,27097	0,78680
Interacción Estudios mínimos con Puntuación total	1,12011	0,89277
<u>Interacción Estudios primarios con Puntuación total</u>	1,79468	0,05572
<u>Interacción Bachiller y Estudios superiores con Puntuación total</u>	14,57867	0,06859

Tabla 10.2. Prueba de Funcionamiento Diferencial de los ítems No uniforme. Ítem 21 en relación al nivel de estudios.

La columna “exponente (coeficiente)” indica el exponente del coeficiente, o lo que es lo mismo, el Odds Ratio (OR). Es una ecuación que compara sucesos. Los OR han sido calculados a partir de la siguiente fórmula:

$$OR = \frac{a/b}{c/d}$$

Donde *a* son los casos expuestos (probabilidad de acertar con estudios mínimos), *b* los controles expuestos (probabilidad de acertar siendo analfabeto y personas que saben leer y escribir), *c* los casos no expuestos (probabilidad de no acertar con estudios mínimos) y *d* los controles no expuestos (probabilidad de no acertar siendo analfabeto y personas que saben leer y escribir). Lo que podemos agrupar en una tabla 2x2 tal y como se hace en la tabla 11.

	Estudios mínimos	Analfabetos y personas que saben leer y escribir
Acertar el ítem	a	b
No acertar el ítem	c	d

Tabla 11. Disposición de los términos para calcular el OR.

De esta forma, un $OR=1$ se interpreta como que la oportunidad de ocurrencia de un evento es igual a la que ocurra otro evento. Cuanto más alejados sean los valores de 1, más fortaleza habrá en la relación descrita. Para explicar e interpretar mejor este concepto, se pone a su lado la columna "1/exponente (coeficiente)" en la cual se invierte la fracción y el grupo de referencia pasaría al numerador. Esta columna es útil cuando el OR es inferior a 1, ya que la interpretabilidad se hace más difícil. En esta columna se ubica el inverso o recíproco del exponente del coeficiente.

Una vez descartado que no existe DIF no uniforme puede pasar a hacerse la prueba para el DIF uniforme y para el Ítem 10 se ha detectado DIF uniforme en la educación.

Modelo sin Interacción

	coeficiente	S.E. (coef)	mín 0.95	máx 0.95	Chi ²	p
(Intercepto)	-6,8478070	1,8616016	-10,8167915	-3,3983808	15,2557200	0,0001
Puntuación total	0,4505096	0,08570923	0,2944601	0,6386410	36,6374820	0,0000
Estudios mínimos	-0,8978614	0,91405642	-2,9563819	0,7434137	1,0763430	0,2995
Estudios primarios	-0,9900212	0,99208512	-3,1898672	0,8355093	1,0843830	0,2977
<u>Bachiller y Estudios superiores</u>	<u>-3,7351857</u>	<u>1,2719903</u>	<u>-6,4645972</u>	<u>-1,3172490</u>	<u>8,7971340</u>	<u>0,0030*</u>

Prueba de razón de verosimilitud=39,4815 con 4 grados de libertad, $p=0,05540196$, $n=345$

Tabla 12.1. Prueba de Funcionamiento Diferencial de los ítems uniforme. Ítem 10 en relación al nivel de estudios.

De esta forma se puede afirmar que hay una diferencia en la probabilidad de acertar en los diferentes niveles de estudio, en concreto para las personas con bachiller y personas con estudios superiores con respecto a personas analfabetas y personas que saben leer y escribir, pero es la misma para cualquier puntuación del test (para cualquier nivel de deterioro cognitivo).

Al contrario de lo que se pudiera pensar, el signo del coeficiente es negativo, lo que quiere decir que personas al comparar el grupo de estudios de analfabetos y personas que solo saben leer y escribir con las personas con estudios de bachiller y superiores, superan

significativamente mejor este ítem las personas analfabetas y personas que solo saben leer y escribir.

Sobre el intervalo de confianza del OR hay que fijarse si contiene o no el valor nulo (si entre ellos está contenido el valor 0).

	exp(coef)	1/exp(coef)
(Intercepto)	0,001061782	941,8132093
Puntuación total	1,569111635	0,6373033
Estudios mínimos	0,407440075	2,4543487
Estudios primarios	0,371568798	2,6912916
<i>Bachiller y Estudios superiores</i>	<i>0,023868737</i>	<i>41,8958063</i>

Tabla 12.2 Prueba de Funcionamiento Diferencial de los ítems uniforme. Ítem 10 en relación al nivel de estudios.

De esta forma, de la tabla puede inferirse que existen 41 (casi 42) veces más de probabilidad de que una persona con analfabetismo o que su nivel de estudios solo llegue a saber leer y escribir acierte el ítem 10 frente a una persona con bachiller o una persona con estudios superiores.

Actualmente solo se está comparando el grupo de analfabetos y personas que saben leer y escribir con el resto de los grupos. Pero las comparaciones entre el resto de los grupos no se efectúan, eso tendría que hacerse en otros análisis a parte, ya que no es el objetivo de este trabajo.

6.1.1.- DIF NO UNIFORME Y EDUCACIÓN.

Solo se ha detectado un ítem que presente DIF no uniforme en la variable educación, el ítem 21, que se refiere al recuerdo de la palabra “manzana”. Como es funcionamiento diferencial no uniforme se refiere a que en la interacción del modelo se ha detectado un p-valor de 0,0002. Esta interacción es significativa cuando se compara el grupo de personas con más baja educación, el grupo de personas que tienen estudios primarios y también para el grupo de personas que tienen estudios de bachiller y estudios superiores. Pero no es significativa cuando se compara con el grupo de estudios mínimos. Las diferencias en la superación del ítem no son debidas únicamente a la edad, sino que también interviene la puntuación total del test, o sea, a la interacción.

6.1.2.- DIF UNIFORME Y EDUCACIÓN.

Los ítems 10, 17 y 18 son los ítems en los que se ha detectado el funcionamiento diferencial con respecto a la educación.

En el ítem 10 solo ha sido significativo el subgrupo de personas que tienen bachiller y estudios superiores. Lo que quiere decir que con respecto a las personas que son analfabetas y las que saben tanto leer como escribir, solo presenta funcionamiento diferencial de estos subgrupos. Además su coeficiente es negativo ($\beta = -3,7352$) lo que indica un desempeño mejor en el ítem de las personas analfabetas y de las que saben leer y escribir con respecto a las personas con bachiller y de estudios superiores. Puede decirse que es 41,8958 veces más probable acertar el ítem siendo del primer grupo con respecto al último.

Al igual que en el ítem 10, en el ítem 17 solo se ha detectado DIF uniforme para el subgrupo de personas con bachiller y estudios superiores. Pero en este caso la probabilidad de acertar el ítem siendo una persona con bachiller o estudios es mayor que con respecto a una persona que es analfabeta o que solo sabe leer y escribir. En concreto es 41,5754 veces más probable superar el ítem siendo del último grupo, que del primero.

Por último se detectó DIF uniforme en el ítem 18 refiriéndose a la educación. Se ha detectado tanto para las personas con estudios primarios, como para las personas que tienen estudios de bachiller o estudios superiores. El desempeño en ambos grupos es mejor si se tienen más estudios, pero no se ha detectado ese funcionamiento diferencial para el grupo de estudios primarios. Para estudios primarios es 3,0189 veces más probable de acertar el ítem con respecto a personas analfabetas o que solo saben leer y escribir. Para estudios de bachiller y superiores, es 35,4430 veces más probable acertar el ítem teniendo estos estudios con respecto a ser analfabeto o solo tener conocimientos de lectura y escritura.

6.2.- RESULTADOS CON LA VARIABLE EDAD

Se ha detectado DIF no uniforme en el ítem 27 con respecto a la edad.

	coeficiente	S.E. (coef)	mín 0.95	máx 0.95	Chi ²	p
(Intercepto)	-79,8319213	33,7372796	-148,48658	-13,9197396	5,608026	0,01787839
Puntuación total	3,67060171	1,36190823	1,01804466	6,44664268	7,222493	0,00719957
Edad	0,99869355	0,43577585	0,15215436	1,89061745	5,332396	0,02093259
<u>Interacción</u> <u>edad-puntuación</u> <u>total</u>	<u>-0,04423084</u>	<u>0,01745719</u>	<u>-0,07985502</u>	<u>-0,01026606</u>	<u>6,408277</u>	<u>0,01135896*</u>

Prueba de razón de verosimilitud=18.3318 con 3 grados de libertad, p=0,000375701, n=347

Tabla 13.1 Prueba de Funcionamiento Diferencial de los ítems no uniforme. Ítem 27 en relación a la edad.

Al estar la edad en años, el incremento corresponde a un año, significa un año con respecto al anterior. Esto hace que el grupo focal sea un grupo movable en el que a medida que se aumenta en edad, aumenta tanto el grupo focal como el grupo de referencia. Y siempre se comparará una edad con su anterior, aunque estemos considerando saltos de un año de edad, éstos pueden referirse a intervalos de edad más grandes, pero a efectos de interpretación será para intervalos de un año.

Por esta razón los parámetros son muy pequeños, incluso cercanos a 1. Y cuando el coeficiente es negativo, quiere decir que a edades superiores, su desempeño en la realización del ítem es peor. Por tanto, habiendo detectado DIF no uniforme, se puede afirmar que las puntuaciones además de la edad, también dependen también del deterioro. Ya que la existencia de interacción quiere decir justamente eso.

	exp(coef)	1/exp(coef)
(Intercepto)	0,0000	4,6834*10 ³⁴
Puntuación total	39,2755	0,0255
edad	2,7147	0,3684
<u>Interacción puntuación total-edad</u>	<u>0,9567</u>	<u>1,0452</u>

Tabla 13.2. Prueba de Funcionamiento Diferencial de los ítems no uniforme. Ítem 27 en relación a la edad.

Se ha detectado DIF uniforme en el ítem 3 correspondiente a la pregunta “¿Qué día del mes es hoy?”

	coeficiente	S.E.(coef)	mín 0.95	máx 0.95	Chi ²	p
(Intercepto)	-0,76856	2,91346	-6,58332	4,87396	0,06998	0,79137
Puntuación Total	0,30754	0,05220	0,20963	0,41460	41,74883	0,00000
<u>Edad</u>	<u>-0,06644</u>	<u>0,03171</u>	<u>-0,12900</u>	<u>-0,00419</u>	<u>4,37129</u>	<u>0,03655*</u>

Prueba de razón de verosimilitud=55,82523 con 2 grados de libertad, p=7,546186e-13, n=347

Tabla 14.1 Prueba de Funcionamiento Diferencial de los ítems no uniforme. Ítem 3 en relación a la edad.

	exp(coef)	1/exp(coef)
(Intercept)	0,4636804	2,1566577
Puntuación Total	1,3600753	0,7352534
<u>Edad</u>	<u>0,9357235</u>	<u>1,0686917</u>

Tabla 14.2. Prueba de Funcionamiento Diferencial de los ítems no uniforme. Ítem 3 en relación a la edad.

En DIF uniforme puede afirmarse que la diferencia en la edad solo empeora con el deterioro, ya que habiéndose descartado la interacción, podemos afirmar que existe DIF en con respecto a los saltos de un año en la edad.

En referencia a la interpretación del exponente, puede decirse que existe 1.0686917 (valor muy cercano a uno por considerarse incrementos muy pequeños de tiempo) veces más de probabilidad de acertar el ítem con edades más bajas, que con edades más altas.

6.2.1.- DIF NO UNIFORME Y EDAD.

De la tabla 10 se infiere que tanto los ítems 14, 15, 16, 17, 18, 21 y 27 presenta DIF no uniforme para la variable edad. Lo que quiere decir que el DIF no es igual en los diferentes niveles del constructo, o lo que es lo mismo, que la edad influye en el DIF, pero también influye el deterioro cognitivo. En todos los ítems se ha ido empeorando el rendimiento en función de la interacción de la edad con el deterioro, así que a mayor edad, peor desempeño diferencial.

6.2.2.- DIF UNIFORME Y EDAD.

Los ítems 3, 6, 7, 9 y 19 presentan DIF uniforme. Esto quiere decir que la diferencia en el desempeño de los ítems, puede achacarse directamente a la edad. Los incrementos son muy pequeños, en el ítem 3, la probabilidad de acertar un ítem siendo un año más joven con respecto a tener más edad es solo de 1,0687 veces más, aumenta ligeramente para el ítem 6 siendo aumentando a 1,1175 veces más probable de acertarlo siendo más joven. Para el ítem 7 es de 1,2670, para el ítem 9 es de 1,2489 y para el 19 de 1,0537 veces más. Todos ellos comparten el mismo sentido de empeoramiento: el ítem se desempeña peor conforme más edad tienen los sujetos.

6.3.- RESULTADOS CON LA VARIABLE SEXO

Se ha encontrado DIF no uniforme en el ítem 20, correspondiente al recuerdo de “caballo”.

	coeficiente	S.E. (coef)	mín 0.95	máx 0.95	Chi ²	p
(Intercepto)	-15,0958	2,8770	-21,2709	-9,8918	48,0783	0,0000
Puntuación Total	0,5649	0,1049	0,3748	0,7897	50,3768	0,0000
Mujer	7,3674	3,2598	1,2500	14,1582	5,6690	0,0173
<u>Interacción puntuación total-mujer</u>	<u>-0,2541</u>	<u>0,1202</u>	<u>-0,5035</u>	<u>-0,0277</u>	<u>4,8856</u>	<u>0,0271*</u>

Prueba de razón de verosimilitud=86,28001 con 2 grados de libertad, p=0, n=347

Tabla 15.1 Prueba de Funcionamiento Diferencial de los ítems no uniforme. Ítem 20 en relación al sexo.

El análisis se para en el primer paso, ya que se ha detectado DIF no uniforme, así que solo se interpreta la interacción. Puede decirse en este caso que la diferencia en la probabilidad de desempeñar este ítem además de la puntuación total, también interviene el sexo. Y lo hace siendo más desfavorable para el grupo de hombres que de mujeres, esto se evidencia en el signo del coeficiente.

	exp(coef)	1/exp(coef)
(Intercepto)	0,0000	3597547
Puntuación Total	1,7593	0,5684
Mujer	1583,5480	0,0006
Interacción puntuación total-mujer	0,7756	1,2893

Tabla 15.2 Prueba de Funcionamiento Diferencial de los ítems no uniforme. Ítem 20 en relación al sexo.

En el ítem 14 se ha detectado DIF uniforme en referencia al sexo.

	coeficiente	S.E.(coef)	mín 0.95	máx 0.95	Chi ²	p
(Intercepto)	-11,7478	2,3832	-17,0710	-7,4506	35,1787	0,0000
Puntuación Total	0,6667	0,1123	0,4691	0,9274	Inf	0,0000
Mujer	-1,5099	0,7701	-3,2061	-0,1198	4,5865	0,0322*

Prueba de razón de verosimilitud=86,28001 con 2 grados de libertad, p=0, n=347

Tabla 16.1. Prueba de Funcionamiento Diferencial de los ítems uniforme. Ítem 14 en relación al sexo.

Una vez descartada la interacción, puede afirmarse que existe un funcionamiento diferencial en función del sexo, en concreto los hombres desempeñan peor este ítem

	exp(coef)	1/exp(coef)
(Intercepto)	0,00	126474,90
Puntuación Total	1,95	0,51
Mujer	0,22	4,53

Tabla 16.2. Prueba de Funcionamiento Diferencial de los ítems uniforme. Ítem 14 en relación al sexo.

Interpretando esta tabla, puede decirse que existen 4,35 veces más de probabilidades de acertar el ítem siendo mujer, que siendo varón.

6.3.1.- DIF NO UNIFORME Y SEXO.

Con respecto al sexo se compara el grupo de mujeres con el de hombres, o sea, que el grupo de referencia es el de hombres. Así al interpretar el signo negativo, de su coeficiente quiere decirse que los hombres tienen un funcionamiento diferencial superior al de las mujeres en este aspecto. Así, se ha detectado funcionamiento diferencial en el ítem 20 ($\beta_3 = -0,2541$), pero como es un funcionamiento no uniforme, lo que quiere decir que las diferencias detectadas no son solamente atribuibles al sexo, sino que también interviene el grado de deterioro o la puntuación total. De esta forma se deduce que existe una probabilidad 1,2893 veces mayor de acertar el ítem siendo hombre que siendo mujer, esta

relación viene influida no solamente por el sexo, sino también por el grado de deterioro cognitivo.

En el ítem 21, el recuerdo de la palabra “manzana” también se ha detectado DIF no uniforme. Así que la interacción para este ítem es significativa a favor de los varones. Así que puede decirse que existe una probabilidad 1,6069 veces superior de acertar este ítem siendo hombre, que siendo mujer cuando la relación está siendo afectada por la puntuación total.

6.3.2.- DIF UNIFORME Y SEXO.

Para el ítem 9 el mejor desempeño de los varones puede asumirse por el DIF, así que para este ítem, el grupo de hombres se ve favorecido con respecto al de mujeres. De la misma forma ocurre con el ítem 14, con el ítem 16 y 18. Pero no así con el ítem 28, en el cual el DIF está a favor del grupo de mujeres siendo 7,0177 veces superior la probabilidad de acertar este ítem siendo mujer que siendo varón. Para el ítem 14 son 4,5262 veces más de probabilidades de acertar el ítem siendo varón que mujer. Para el ítem 16 las veces son 2,1389 y para el ítem 18 son 2,3928 veces.

7.- CONCLUSIONES

Se ha encontrado DIF en todas las variables estudiadas.

Hay un grupo de respuestas que están relacionadas entre sí, ya que si se falla una pregunta, se fallan las sucesivas. Este grupo de preguntas del test se refiere a las preguntas de cálculo numérico, las preguntas: 14, 15, 16, 17 y 18. Consisten ir restando 30 menos 3. Si se acierta la primera resta, se acertaría el ítem 14 y se pasaría al ítem 15, pero si se falla esa respuesta se pasa directamente al ítem 19. Por eso están concatenadas esas preguntas y por eso es extraño que no actúen como bloque, sino como ítems separados. Esta separación tendría más sentido cuanto mayor sea el ítem de ese bloque, pero en lo referente al sexo, que las respuestas presenten DIF de manera salteada, puede ser una línea de investigación en futuros trabajos.

De los 30 ítems 15 de ellos presentan DIF y los otros 15 no lo presentan. Se ha encontrado tanto DIF uniforme como no uniforme en las tres variables estudiadas.

En la variable edad se está estudiando como variable continua comparándose con la edad anterior en tramos de un año. Pese a ello, es la variable en la que más DIF se ha encontrado: 5 para DIF uniforme y 7 ítems presentan DIF no uniforme.

En la variable educación se está comparando el grupo de analfabetos y personas que solamente saben leer y escribir con el resto de grupos. Haciendo solo esas comparaciones se ha logrado detectar DIF en solo un ítem tratándose de DIF no uniforme y con respecto al DIF uniforme se han encontrado solamente 3 de los 30 ítems.

Por último en la variable sexo se ha encontrado DIF no uniforme en dos ítems y DIF uniforme en 5 de ellos. No debería darse este tipo de DIF en esta variable, ya que las explicaciones pueden presentar controversia. Ante esta conjetura, se sugiere que en futuras investigaciones se plantee este problema, ya que una de las posibles explicaciones es que se trate de una variable enmascarada en la que influya otra variable que explique mejor los resultados.

Para futuras investigaciones podría estudiarse la calibración del test, esto es pasar a medir a los diferentes grupos con la misma medida salvando las diferencias entre ellos.

8.- AGRADECIMIENTOS

Este trabajo se ha realizado gracias a los datos ofrecidos por Jesús Cacho, jefe del servicio del Clínico y a la cabeza del equipo de médicos del centro de salud del barrio de Pizarrales de la ciudad española de Salamanca, lugar donde se obtuvieron los datos.

9.- REFERENCIAS BIBLIOGRÁFICAS

- Agresti, A. (1996). An introduction to categorical data analysis (Vol. 135). Wiley New York.
- Amaducci, L., Baldereschi, M., Amato, M. P., Lippi, A., Nencini, P., Maggi, S., & Litvak, J. (1991). The World Health Organization cross-national research program on age-associated dementias. *Aging Clinical and Experimental Research*, 3(1), 89-96.
- Ansari, N. N., Naghdi, S., Hasson, S., Valizadeh, L., & Jalaie, S. (2010). Validation of a Mini-Mental State Examination (MMSE) for the Persian population: a pilot study. *Applied Neuropsychology*, 17(3), 190-195.
- Baldereschi, M., Meneghini, F., Quiroga, P., Alcala, C., MUSCAT, P., Gabriel, R., ... & MAMO, J. (1994, April). Cognitive versus functional screening for dementia across different countries-cross-cultural validation of the mini-mental-state-examination (MMSE) and of the pfeffer functional activities questionnaire (PFAQ) against the standardized clinical-diagnosis of dementia. In *Neurology* (Vol. 44, No. 4, pp. A365-A365). 34 Beacon Street, Boston, Ma 02108-1493: Little Brown Co.
- Balluerka, N., Plewis, I., Gorostiaga, A., & Padilla, J.-L. (2015). Examining sources of DIF in psychological and educational assessment using multilevel logistic regression. *Methodology*.
- Bassett, S. S., & Folstein, M. F. (1991). Cognitive impairment and functional disability in the absence of psychiatric diagnosis. *Psychological medicine*, 21(01), 77-84.
- Bermejo-Pareja, F., Benito-León, J., Vega, S., Medrano, M. J., Román, G. C., & Neurological Disorders in Central Spain (NEDICES) Study Group. (2008). Incidence and subtypes of dementia in three elderly populations of central Spain. *Journal of the neurological sciences*, 264(1), 63-72.
- Bermejo-Pareja, F., Benito-Leon, J., Vega, S., Olazara, J., De Toledo, M., Di, J., ... & Roma, G. C. (2009). Consistency of clinical diagnosis of dementia in NEDICES: a population-based longitudinal study in Spain. *Journal of geriatric psychiatry and neurology*.
- Bermejo-Pareja, F., Benito-León, J., Vega-Q, S., Díaz-Guzmán, J., Rivera-Navarro, J., Molina, J. A., ... & Morales-González, J. M. (2008). La cohorte de ancianos NEDICES. Metodología y principales hallazgos neurológicos. *Rev Neurol*, 46, 416-423.

- Bleecker, M. L., Bolla-Wilson, K., Kawas, C., & Agnew, J. (1988). Age-specific norms for the Mini-Mental State Exam. *Neurology*, 38(10), 1565-1565.
- Blessed, G., Tomlinson, B. E., & Roth, M. (1968). The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. *The British Journal of Psychiatry*.
- Bull, S. B., Mak, C., & Greenwood, C. M. T. (2002). A modified score function estimator for multinomial logistic regression in small samples. *Computational Statistics & Data Analysis*, 39(1), 57–74.
- Busse, A., Sonntag, A., Bischof, J., Matschinger, H., & Angermeyer, M. C. (2002). Adaptation of dementia screening for vision-impaired older persons: administration of the Mini-Mental State Examination (MMSE). *Journal of Clinical Epidemiology*, 55(9), 909–915.
- Carnero-Pardo, C. (2013). ¿Es hora de jubilar al Mini-Mental? *Neurología*, (xx). doi:10.1016/j.nrl.2013.07.003
- Chertkow, H., Massoud, F., Nasreddine, Z., Belleville, S., Joanette, Y., Bocti, C., ... Bergman, H. (2008). Diagnosis and treatment of dementia: 3. Mild cognitive impairment and cognitive impairment without dementia. *CMAJ: Canadian Medical Association Journal = Journal de l'Association Médicale Canadienne*, 178(10), 1273–1285. doi:10.1503/cmaj.070797
- Chiu, H. F., Lee, H. C., Chung, W. S., & Kwong, P. K. (1994). Reliability and validity of the Cantonese version of mini-mental state examination-a preliminary study. *Hong Kong Journal of Psychiatry*, 4(2), 25.
- Contador, I., Fernández-Calvo, B., Ramos, F., Tapias-Merino, E., & Bermejo-Pareja, F. (2010). [Dementia screening in primary care: critical review]. *Revista de neurología*, 51(11), 677-686.
- Cornfield, J., Gordon, T., & Smith, W. W. (1961). Quantal response curves for experimentally uncontrolled variables. *Bull Int Stat Inst*, 38(3), 97–115.
- Correa, M., Juan, C., & Valencia, C. (2011). La separación en regresión logística, una solución y aplicación. *Revista Facultad Nacional de Salud Pública*, 29(3), 281–288.

- Crane, P. K., Gibbons, L. E., Jolley, L., van Belle, G., Selleri, R., Dalmonte, E., & De Ronchi, D. (2006). Differential item functioning related to education and age in the Italian version of the Mini-mental State Examination. *International psychogeriatrics*, 18(03), 505-515.
- Crum RM, Anthony JC, Bassett SS, Folstein MF (1993). Population-based norms for the mini-mental state examination by age and educational level. *JAMA*, 269(18), 2386–2391. Retrieved from <http://dx.doi.org/10.1001/jama.1993.03500180078038>
- Ferrando, P. J. (1996). Evaluación de la unidimensionalidad de los ítems mediante análisis factorial. *Psicothema*, 8(2), 397–410.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1), 27–38.
- Folstein, M. (1990). The birth of the MMS. *This Week's Citation Classic*, 2, 18.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189–198.
- Fountoulakis, K. N., Tsolaki, M., Chantzi, H., & Kazis, A. (2000). Mini mental state examination (MMSE): a validation study in Greece. *American Journal of Alzheimer's Disease and Other Dementias*, 15(6), 342-345.
- Ganguli, M., Ratcliff, G., Chandra, V., Sharma, S., Gilby, J., Pandav, R., ... & Dekosky, S. (1995). A Hindi version of the MMSE: the development of a cognitive screening instrument for a largely illiterate rural elderly population in India. *International Journal of Geriatric Psychiatry*, 10(5), 367-377.
- García, M. I. B., & Marañón, P. P. (1996). Detección de funcionamiento diferencial de los ítems mediante análisis de residuales: una aplicación de la TRI. *Psicothema*, 8(1), 173-180.
- García-Cueto, E. (1993). *Introducción a la psicometría*. Siglo XXI de España.
- Garre-Olmo, J., Lax-Pericall, C., Turro-Garriga, O., Soler-Cors, O., Monserrat-Vila, S., Vilalta-Franch, J., ... López-Pousa, S. (2008). Adaptación y validez convergente de una versión telefónica del Mini-Mental State Examination. *Medicina Clínica*, 131(3), 89–95. doi:10.1157/13124011
- George, L. K., Landerman, R., Blazer, D. G., & Anthony, J. C. (1991). Cognitive impairment. *Psychiatric disorders in America: The epidemiologic catchment area study*, 291-327.

- Gómez-Benito, J., Hidalgo, M. D., Padilla, J. L., & González, A. (2005). Desarrollo informático para la utilización de la regresión logística como técnica de detección del DIF. In *Demostración informática presentada al IX Congreso de Metodología de las Ciencias Sociales y de la Salud, Granada, España*.
- González-Hernández, J., Aguilar, L., Oporto, S., Araneda, L., Vásquez, M., & Von-Bernhardi, R. (2009). Normalización del "Mini-Mental State Examination" según edad y educación, para la población de Santiago de Chile. *Revista Memoriza Com*, 3, 23–34.
- Grupo de trabajo de la Guía (2010). Guía de Práctica Clínica sobre la atención integral a las personas con enfermedad de Alzheimer y otras demencias.
- Guerreiro, M., Silva, A. P., Botelho, M. A., Leitão, O., Castro-Caldas, A., & Garcia, C. (1994). Adaptação à população portuguesa da tradução do Mini Mental State Examination (MMSE). *Revista Portuguesa de Neurologia*, 1(9), 9-10.
- Guideline, S. (2007). *The NICE - SCIE Guideline on supporting people with dementia and their carers in health and social care. The British Psychological Society and the Royal College of Psychiatrists*.
- Güngen, C., Ertan, T., Eker, E., Yaşar, R., & Engin, F. (2001). [Reliability and validity of the standardized Mini Mental State Examination in the diagnosis of mild dementia in Turkish population]. *Turk psikiyatri dergisi= Turkish journal of psychiatry*, 13(4), 273-281.
- Hambleton, R. K., & Rogers, H. J. (1989). Detecting potentially biased test items: Comparison of IRT area and Mantel-Haenszel methods. *Applied Measurement in Education*, 2(4), 313–334.
- Hannay, H. J., Howieson, D. B., Loring, D. W., Fischer, J. S., & Lezak, M. D. (2004). Neuropathology for neuropsychologists. *Neuropsychological assessment*, 4, 157-194.
- Haubois, G., Annweiler, C., Launay, C., Fantino, B., de Decker, L., Allali, G., & Beauchet, O. (2011). Development of a short form of Mini-Mental State Examination for the screening of dementia in older adults with a memory complaint: a case control study. *BMC Geriatrics*, 11(1), 59.
- Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, 21(16), 2409–2419.

- Heinze, G., Ploner, M., Dunkler, D., & Southworth, H. (2013). logistf: Firth's bias reduced logistic regression. *R Package Version, 1*.
- Hofman, A., Rocca, W. A., Brayne, C., Breteler, M. M. B., Clarke, M., Cooper, B., ... Hagnell, O. (1991). The prevalence of dementia in Europe: a collaborative study of 1980–1990 findings. *International Journal of Epidemiology, 20*(3), 736–748.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. *Test Validity, 129–145*.
- IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.
- Ishizaki, J., Meguro, K., Ambo, H., Shimada, M., Yamaguchi, S., Hayasaka, C., ... & Yamadori, A. (1998). A normative, community-based study of Mini-Mental State in elderly adults: the effect of age and educational level. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 53*(6), P359-P363.
- Jones, R. N., & Gallo, J. J. (2001). Education bias in the mini-mental state examination. *International Psychogeriatrics, 13*(03), 299–310.
- Jones, R. N., & Gallo, J. J. (2002). Education and sex differences in the mini-mental state examination effects of differential item functioning. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 57*(6), P548–P558.
- Kang, Y., Na, D. L., & Hahn, S. (1997). A validity study on the Korean Mini-Mental State Examination (K-MMSE) in dementia patients. *Journal of the Korean Neurological Association, 15*(2), 300-308.
- Kittner, S. J., White, L. R., Farmer, M. E., Wolz, M., Kaplan, E., Moes, E., ... & Feinleib, M. (1986). Methodological issues in screening for dementia: the problem of education adjustment. *Journal of chronic diseases, 39*(3), 163-170.
- Küçükdeveci, A. A., Kutlay, S., Elhan, A. H., & Tennant, A. (2005). Preliminary study to evaluate the validity of the mini-mental state examination in a normal population in Turkey. *International Journal of Rehabilitation Research, 28*(1), 77–79.
- Law, S. & Wolfson, C. (1995). Validation of a French version of an informant-based questionnaire as a screening test for Alzheimer's disease. *The British Journal of Psychiatry, 167*(4), 541-544.

- León, A. B. (2012). La unidimensionalidad de un instrumento de medición: perspectiva factorial. *Revista de Psicología*, 24(1), 53–80.
- Lobo, A., Escobar, V., Ezquerro, J., & Seva Díaz, A. (1980). "El Mini-Examen Cognoscitivo"(Un test sencillo, práctico, para detectar alteraciones intelectuales en pacientes psiquiátricos). *Revista de psiquiatría y Psicología Médica*.
- Lobo, A., Launer, L. J., Fratiglioni, L., Andersen, K., Di Carlo, A., Breteler, M. M., ... Martinez-Lage, J. (1999). Prevalence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology*, 54(11 Suppl 5), S4–9.
- Lobo, A., Marcos, G., Dia, J. L., De La Cámara, C., Ventura, T., & Morales Asin, F. (1999). Validación y estandarización del mini-examen cognoscitivo (primera edición española del Mini-Mental Status Examination) en la población geriátrica general. *Medicina clinica (Barc)*, 112(20), 767–774.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.
- Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). Statistical theories of mental test scores.
- Magaziner, J., Bassett, S. S., & Hebel, J. R. (1987). Predicting performance on the Mini-Mental State Examination: use of age-and education-specific equations. *Journal of the American Geriatrics Society*.
- Martin, A. D., Quinn, K. M., & Park, J. H. (2011). Mcmcpack: Markov chain monte carlo in r. *Journal of Statistical Software*, 42(9), 1–21.
- McCarten, J. R., Anderson, P., Kuskowski, M. A., McPherson, S. E., & Borson, S. (2011). Screening for Cognitive Impairment in an Elderly Veteran Population: Acceptability and Results Using Different Versions of the Mini-Cog. *Journal of the American Geriatrics Society*, 59(2), 309–313.
- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational and Behavioral Statistics*, 7(2), 105–118.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156.
- Muñiz, J. (1992). *Teoría clásica de los tests*. Pirámide.

- Nieuwenhuis-Mark, R. E. (2010). The death knoll for the MMSE: has it outlived its purpose? *Journal of Geriatric Psychiatry and Neurology*, 23(3), 151–157.
- Nilsson, F. M. (2007). Mini Mental State Examination (MMSE)—probably one of the most cited papers in health science. *Acta Psychiatrica Scandinavica*, 116(2), 156-157.
- Noale, M., Limongi, F., & Minicuci, N. (2005). Identification of factorial structure of MMSE based on elderly cognitive destiny: the Italian Longitudinal Study on Aging. *Dementia and geriatric cognitive disorders*, 21(4), 233-241.
- O'connor, D. W., Pollitt, P. A., Treasure, F. P., Brook, C. P. B., & Reiss, B. B. (1989). The influence of education, social class and sex on Mini-Mental State scores. *Psychological Medicine*, 19(03), 771-776.
- Pedraza, O., Clark, J. H., O'Bryant, S. E., Smith, G. E., Ivnik, R. J., Graff-Radford, N. R., ... & Lucas, J. A. (2012). Diagnostic Validity of Age and Education Corrections for the Mini-Mental State Examination in Older African Americans. *Journal of the American Geriatrics Society*, 60(2), 328-331.
- Petersen, R. C., Stevens, J. C., Ganguli, M., Tangalos, E. G., Cummings, J. L., & DeKosky, S. T. (2001). Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review). *Neurology*, 56(9), 1133 –1142. Retrieved from <http://www.neurology.org/content/56/9/1133.abstract>
- Powsner, S., & Powsner, D. (2005). Cognition, copyright, and the classroom. *American Journal of Psychiatry*, 162(3), 627–a.
- Preston, S. (2006). Assessing the US Senate vote on the corporate average fuel economy (CAFE) standard. *Journal of Statistics Education*, 14(2), n2.
- Prieto, G., Delgado, a. R., Perea, M. V., & Ladera, V. (2011). Funcionamiento diferencial de los ítems del test Mini-mental en función de la patología. *Neurología*, 26(xx), 474–480. doi:10.1016/j.nrl.2011.01.013
- Quiroga, P., Albala, C., & Klaasen, G. (2004). Validación de un test de tamizaje para el diagnóstico de demencia asociada a edad, en Chile. *Revista médica de Chile*, 132(4), 467-478.
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

- Roccaforte, W. H., Burke, W. J., Bayer, B. L., & Wengel, S. P. (1992). Validation of a telephone version of the mini-mental state examination. *Journal of the American Geriatrics Society*.
- Rogers, H. J., & Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement*, 17(2), 105–116.
- Sahadevan, S., Lim, P. P., Tan, N. J., & Chan, S. P. (2000). Diagnostic performance of two mental status tests in the older Chinese: influence of education and age on cut-off values. *International journal of geriatric psychiatry*, 15(3), 234-241.
- Schultz-Larsen, K., Lomholt, R. K., & Kreiner, S. (2007). Mini-Mental Status Examination: a short form of MMSE was as accurate as the original MMSE in predicting dementia. *Journal of Clinical Epidemiology*, 60(3), 260–267.
- Shepard, L. A., Camilli, G., & Williams, D. M. (1985). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement*, 22(2), 77–105.
- Shulman, K. I., Herrmann, N., Brodaty, H., Chiu, H., Lawlor, B., Ritchie, K., & Scanlan, J. M. (2006). IPA survey of brief cognitive screening instruments. *International Psychogeriatrics*, 18(02), 281–294. Retrieved from <http://dx.doi.org/10.1017/S1041610205002693>
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72–101.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361–370.
- Swanson, D. B., Clauser, B. E., Case, S. M., Nungester, R. J., & Featherman, C. (2002). Analysis of differential item functioning (DIF) using hierarchical logistic regression models. *Journal of Educational and Behavioral Statistics*, 27(1), 53–75.
- Tapias-Merino, E., Puertas-Martin, v., Vera-Garcia, C., Lora-Pablos, D., Revuelta-Alonso, A., & Bermejo-Pareja, f. (2010). Fiabilidad interobservador y test-retest de una versión española (MMSE-37) del test minimal de Folstein, adaptada a poblaciones de bajo nivel educativo. *Revista de neurología*, 50(11), 646-652.

- Teng, E. L., & Chui, H. C. (1987). The modified mini-mental state examination (3MS). *Can J Psychiatry, 41*(2), 114-21.
- Teresi, J. A., Golden, R. R., Cross, P., Gurland, B., Kleinman, M., & Wilder, D. (1995). Item bias in cognitive screening measures: comparisons of elderly white, Afro-American, Hispanic and high and low education subgroups. *Journal of Clinical Epidemiology, 48*(4), 473–483.
- Thurstone, L. L., & Chave, E. J. (1929). The measurement of attitude: A psychophysical method and some experiments with a scale for measuring attitude toward the Church.
- Tierney, M. C., Szalai, J. P., Dunn, E., Geslani, D., & McDowell, I. (2000). Prediction of probable Alzheimer disease in patients with symptoms suggestive of memory impairment: value of the Mini-Mental State Examination. *Archives of family medicine, 9*(6), 527.
- Tombaugh, T. N., & McIntyre, N. J. (1992). The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society.*
- UNESCO Institute for Statistics (UIS). (2015). *Education: literacy rate* [Data file]. Retrieved from <http://data.uis.unesco.org/Index.aspx?queryid=166>
- Villanueva-Iza, C., Bermejo Pareja, F., Berbel García, Á., Trincado Soriano, R., & Rivera Navarro, J. (2003). Validación de un protocolo clínico para la detección de demencia en el ámbito poblacional.
- Viñao, A., Moreno Martínez, P. L., & Navarro García, C. (2009). La alfabetización en España: un proceso cambiante de un mundo multiforme. *Perspectivas Históricas de La Educación de Personas Adultas, 3*(1).
- Walker, S. H., & Duncan, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika, 54*(1-2), 167–179.
- Werner, P., Heinik, J., Mendel, A., Reicher, B., & Bleich, A. (1999). Examining the reliability and validity of the Hebrew version of the Mini Mental State Examination. *Aging Clinical and Experimental Research, 11*(5), 329-334.
- Zumbo, B. D., Liu, Y., Wu, A. D., Shear, B. R., Olvera Astivia, O. L., & Ark, T. K. (2015). A Methodology for Zumbo's Third Generation DIF Analyses and the Ecology of Item Responding. *Language Assessment Quarterly, 12*(1), 136–151.

10.- APÉNDICE

MINI MENTAL STATE EXAMINATION (MMSE)

Basado en Folstein et al. (1975), Lobo et al. (1979)

Nombre: _____ Varón [] Mujer []
 Fecha: _____ F. nacimiento: _____ Edad: _____
 Estudios/Profesión: _____ N. Hª: _____
 Observaciones: _____

¿En qué año estamos? 0-1 ¿En qué estación? 0-1 ¿En qué día (fecha)? 0-1 ¿En qué mes? 0-1 ¿En qué día de la semana? 0-1		ORIENTACIÓN TEMPORAL (Máx.5)
¿En qué hospital (o lugar) estamos? 0-1 ¿En qué piso (o planta, sala, servicio)? 0-1 ¿En qué pueblo (ciudad)? 0-1 ¿En qué provincia estamos? 0-1 ¿En qué país (o nación, autonomía)? 0-1		ORIENTACIÓN ESPACIAL (Máx.5)
Nombre tres palabras Peseta-Caballo-Manzana (o Balón- Bandera-Arbol) a razón de 1 por segundo. Luego se pide al paciente que las repita. Esta primera repetición otorga la puntuación. Otorgue 1 punto por cada palabra correcta, pero continúe diciéndolas hasta que el sujeto repita las 3, hasta un máximo de 6 veces. Peseta 0-1 Caballo 0-1 Manzana 0-1 (Balón 0-1 Bandera 0-1 Árbol 0-1)		Nº de repeticiones necesarias FIJACIÓN-Recuerdo Inmediato (Máx.3)
Si tiene 30 pesetas y me va dando de tres en tres, ¿Cuántas le van quedando?. Detenga la prueba tras 5 sustracciones. Si el sujeto no puede realizar esta prueba, pídale que deletree la palabra MUNDO al revés. 30 0-1 27 0-1 24 0-1 21 0-1 18 0-1 (O 0-1 D 0-1 N 0-1 U 0-1 M 0-1)		ATENCIÓN- CÁLCULO (Máx.5)
Preguntar por las tres palabras mencionadas anteriormente. Peseta 0-1 Caballo 0-1 Manzana 0-1 (Balón 0-1 Bandera 0-1 Árbol 0-1)		RECUERDO diferido (Máx.3)
DENOMINACIÓN. Mostrarle un lápiz o un bolígrafo y preguntar ¿qué es esto?. Hacer lo mismo con un reloj de pulsera. Lápiz 0-1 Reloj 0-1 REPETICIÓN. Pedirle que repita la frase: "ni sí, ni no, ni pero" (o "En un trigal había 5 perros") 0-1 ÓRDENES. Pedirle que siga la orden: "coja un papel con la mano derecha, dóblelo por la mitad, y póngalo en el suelo". Coje con mano d. 0-1 dobla por mitad 0-1 pone en suelo 0-1 LECTURA. Escriba legiblemente en un papel "Cierre los ojos". Pídale que lo lea y haga lo que dice la frase 0-1 ESCRITURA. Que escriba una frase (con sujeto y predicado) 0-1 COPIA. Dibuje 2 pentágonos intersectados y pida al sujeto que los copie tal cual. Para otorgar un punto deben estar presentes los 10 ángulos y la intersección. 0-1		LENGUAJE (Máx.9)
Puntuaciones de referencia 27 ó más: normal 24 ó menos: sospecha patológica 12-24: deterioro 9-12 : demencia		Puntuación Total (Máx.: 30 puntos)

a. e. g. (1999)