



Universidad Autónoma de Querétaro  
Facultad de Ingeniería  
Maestría en Matemáticas Aplicadas

TEORÍA DE RESPUESTA AL ÍTEM: UNA APLICACIÓN EDUCATIVA  
**TESIS**

Que como parte de los requisitos para obtener el grado de  
Maestra en Matemáticas Aplicadas

**Presenta:**

Rocío Minerva Hidalgo Flores

**Dirigida por:**

Dr. Eduardo Castaño Tostado

SINODALES

Dr. Eduardo Castaño Tostado  
Presidente

Dr. Humberto Suzán Azpiri  
Secretario

Dr. Víctor Larios Osorio  
Vocal

Dra. Rebeca del Rocío Peniche Vera  
Suplente

M. en D. M/ Carmen Sosa Garza  
Suplente

Dr. Gilberto Herrera Ruiz  
Director de la Facultad

Firma

Firma

Firma

Firma

Firma

Dr. Luis Gerardo Hernández Sandoval  
Director de Investigación y Posgrado

Centro Universitario  
Querétaro, Qro.  
Junio de 2008  
México

## RESUMEN

En este trabajo se exploran los modelos de Teoría de Respuesta al Ítem que se aplican a datos binarios: expresiones matemáticas de los modelos, interpretación de los parámetros y métodos de estimación de parámetros. Se intenta ajustar uno de estos modelos a un conjunto de datos generado por una aplicación del Examen de Habilidades y Conocimientos Básicos, EXHCOBA. Al realizar la evaluación del cumplimiento de los supuestos de la Teoría de Respuesta al Ítem en los datos, se encuentra que varios de los ítems no cumplen el supuesto de monotonicidad. Se observa también que el conjunto de ítems utilizado en este estudio, 20 ítems de Matemáticas para el Cálculo, no es unidimensional. Se concluye que el supuesto de unidimensionalidad resulta bastante restrictivo. Sin embargo la Teoría de Respuesta al Ítem sigue siendo útil en el diseño de ítems adecuados al propósito de la prueba.

**(Palabras clave:** Teoría de Respuesta al Ítem, datos binarios, parámetro, evaluación de supuestos)

## SUMMARY

In this work we explore Item Response Theory models which are applied to binary data; mathematical expressions of the models, interpretation of the parameters and parameter estimation methods. We attempt to adjust one of these models to a group of data generated by an application of the Skills and Basic Knowledge Exam (EXHCOBA, from its initials in Spanish). Upon assessing the assumptions of the Item Response Theory related to the data, we found that several of the items do not comply with the monotonicity assumption. It can also be seen that the group of items used in this study, 20 items from Mathematics for Calculus, is not unidimensional. We conclude that the assumption of unidimensionality turns out to be quite restrictive. Nevertheless, the Item Response Theory continues to be useful in the design of items that are appropriate for the purpose of the test.

**(Key words:** Item Response Theory, binary data, parameters, assessment of assumptions)

## Dedicatoria

A mi esposo e hijos por su comprensión y cariño. A mis padres, por su ejemplo y amor. Sé que Dios tuvo que ver en todo lo que me fue dado, de modo que agradecer a las personas es agradecerle a él.

## Agradecimientos

Deseo agradecer a todas las personas a mi alrededor que de diferentes maneras me ayudaron a realizar este trabajo. En primer lugar al Dr. Eduardo Castaño Tostado por su dirección, motivación y paciencia. A la doctores Marthagloria Morales Garza, Ma. de los Ángeles Guzmán Molina y Augusto Peón Solís por apoyarme, en mi decisión de cursar la maestría, durante sus respectivos periodos al frente de la dirección de la Facultad de Ciencias Políticas y Sociales. Al Dr. Gilberto Herrera y a la Dra. Rocío Peniche Vera por su ayuda y confianza. A los profesores y compañeros de la maestría por su ayuda y amistad.

## ÍNDICE GENERAL

|  | Página |
|--|--------|
| Resumen.....   | ii     |
| Summary.....   | iii    |
| Dedicatorias.....  | iv     |
| Agradecimientos.....   | v      |
| Índice.....  | vi     |
| Índice de Tablas.....  | viii   |
| Índice de Figuras.....   | ix     |
| 1. INTRODUCCIÓN.....   | 1      |
| 2. FUNDAMENTO TEÓRICO.....   | 3      |
| Antecedentes de la teoría de respuesta al ítem.....                                    | 3      |
| 2.1.1. Relación entre teoría de respuesta al ítem y teoría clásica del test.....       | 5      |
| 2.1.2. Los supuestos de la TRI.....  | 6      |
| 2.2 Modelos matemáticos usados en la TRI.....  | 7      |
| 2.2.1. El modelo ojiva normal de 3 parámetros.....                                     | 7      |
| 2.2.2. Modelo logístico de 3 parámetros.....   | 9      |
| 2.2.3. Modelo logístico de 1 parámetro.....  | 11     |
| 2.2.4. Modelos no paramétricos.....  | 11     |
| 2.2.5. Funciones de información.....   | 13     |
| 2.2.5.1. Función de información del ítem.....  | 13     |
| 2.2.5.2. Función de información de la prueba.....                                      | 15     |
| 2.3. Evaluación de los supuestos de la TRI.....  | 16     |
| 2.3.1. Evaluación del supuesto de monotonicidad.....                                   | 16     |
| 2.3.2. Evaluación del supuesto de unidimensionalidad.....                              | 17     |
| 2.3.3. Evaluación del supuesto de independencia local.....                             | 19     |
| 2.4. Procedimientos de estimación de parámetros.....                                   | 20     |
| 2.4.1. Estimación MV de los parámetros del ítem, modelo ojiva normal 2 parámetros..... | 21     |
| 2.4.2. Estimación MV de los parámetros del ítem, modelo logístico 2 parámetros.....    | 22     |
| 2.4.3. Estimación MV de los parámetros de los examinados, modelo ojiva normal.....     | 23     |
| 2.4.4. Estimación MV de los parámetros de los examinados, modelo logístico.....        | 23     |
| 2.4.5. Estimación Bayesiana de los parámetros del ítem.....                            | 24     |

|  |    |
|--|----|
| 2.4.6. Estimación marginalizada Bayesiana de los parámetros del ítem.....                                    | 24 |
| 2.4.7. Estimación Bayesiana de los parámetros de los examinados.....   | 25 |
| 2.4.7.1. Estimación modal de Bayes.....  | 25 |
| 2.4.7.2. Estimación de Bayes EAP.....  | 26 |
| 2.4.8. Estimación de parámetros de ítemes y de examinados vía muestreador de Gibbs .....                     | 26 |
| 2.4.8.1. El muestreador Gibbs de Albert.....   | 27 |
| 3. APLICACIÓN Y RESULTADOS.....  | 34 |
| 3.1. Descripción de la prueba.....   | 34 |
| 3.2. Descripción de los datos.....   | 35 |
| 3.3. Estimación de los parámetros.....   | 36 |
| 3.4. Evaluación de supuestos de la TRI.....  | 37 |
| 4. CONCLUSIONES.....   | 41 |
| 5. BIBLIOGRAFÍA.....   | 45 |
| APÉNDICES  |    |
| A. Aspectos computacionales.....   | 46 |
| B. Datos utilizados.....   | 48 |
| C. Curvas características de los ítemes, usando TESTGRAF.....  | 51 |
| D. Estimación de parámetros de ítemes y examinados usando el modelo ojiva normal 2 parámetros y WINBUGS..... | 58 |
| E. Curvas características de ítemes, usando el modelo ojiva normal 2 parámetros y WINBUGS.....               | 62 |
| Índice   |    |

## Índice de Tablas

| <b>Tabla</b> | <b>Página</b> |
|--------------|---------------|
| 1            | 56            |

## Índice de Figuras

| <b>Figura</b> | <b>Página</b> |
|---------------|---------------|
| 1             | 12            |
| 2             | 51            |
| 3             | 60            |
| 4             | 62            |
| 5             | 63            |
| 6             | 67            |
| 7             | 68            |

# CAPÍTULO 1

## INTRODUCCIÓN

Actualmente la medición de habilidades académicas de individuos es muy importante para las instituciones educativas y los gobiernos de los países del mundo. Inmersos como estamos en la globalización, es una práctica común comparar el aprovechamiento escolar de los estudiantes de un determinado país con los de otros países, para lo cual se requieren mediciones que permitan hacer esto de manera adecuada, es decir medir a través de diferentes grupos de individuos y de diferentes tests. En muchas universidades, el ingreso de un aspirante depende en buena medida de la medición que obtuvo como resultado de un examen de admisión, por ejemplo. Las instituciones educativas de nivel superior desean contar con instrumentos que les permitan captar a los aspirantes con mayores probabilidades de obtener resultados académicos óptimos, de aquí que se requiera contar con herramientas para el diseño y la evaluación de los reactivos del examen y del examen en su conjunto, el instrumento de medición. Se requiere que el instrumento utilizado sea de tal forma que la estimación de la habilidad de un examinado no dependa del resto de los examinados ni de las preguntas del examen. Es aquí donde interviene la Teoría de Respuesta al Ítem, la cual permite estimar el nivel de habilidad académica de un examinado al mismo tiempo que la precisión de la misma, es decir obtener un índice que describa la confiabilidad de cada estimación. Pero no sólo eso sino que también permite caracterizar la calidad de cada ítem del examen y del examen mismo como un todo. Básicamente, la teoría de respuesta al ítem se basa en modelos matemáticos que relacionan la habilidad del examinado con la propensión de éste a responder acertadamente a un reactivo del examen.

La teoría cuenta con herramientas para verificar los supuestos del modelo, inferir sobre los parámetros de los examinados y de los ítems y finalmente producir indicadores de la efectividad de la medición del examen.

En realidad, la teoría de respuesta al ítem es aplicable en todas aquellas situaciones donde se utilice un test para medir un rasgo psicológico de los individuos, en particular el rasgo puede ser una habilidad académica como es el caso de este trabajo.

El tipo de reactivo en el examen analizado aquí es el conocido como de opción múltiple, el cual es calificado de manera dicotómica: correcto o incorrecto.

El presente trabajo tiene como **objetivo** la aplicación de la Teoría de Respuesta al Ítem a un conjunto de datos real proveniente de la aplicación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA) en el año 2006, en la Univer-

sidad Autónoma de Querétaro. A fin de aplicar la Teoría se estudian los diferentes métodos para estimar los parámetros del modelo, los parámetros de habilidad de los examinados y verificar los supuestos de la teoría y el ajuste del modelo; lo cual implica el uso de varios programas de cómputo. Los utilizados aquí son: TESTGRAF, el cual fue realizado por Jim Ramsay de la Universidad de McGill, en Canada, WINBUGS y TESTFACT.

La aplicación de la teoría a los datos atraviesa por el análisis de los ítemes que componen el examen, y la pertinencia de reactivos de estas características en el mismo. También se analiza la distribución de las habilidades estimadas de los sujetos examinados y el comportamiento de los errores estándar de las estimaciones a través del intervalo de la habilidad para diagnosticar en qué valores de la habilidad, la prueba está midiendo con mayor efectividad, para luego analizar la concordancia de estas características con la finalidad del examen.

# CAPÍTULO 2

## FUNDAMENTO TEÓRICO

La Teoría de Respuesta al Ítem (TRI) es muy extensa, de modo que la revisión que se hace está más dirigida hacia las aplicaciones en educación y en particular a ítemes calificados dicotómicamente; esto es ítemes de opción múltiple a los cuales se les asigna calificación de 1 si la opción elegida por el examinado fue la correcta y 0 si la opción elegida fue incorrecta. Enseguida se hará un breve recorrido por su historia, sus fundamentos, los modelos utilizados y los métodos de estimación de los parámetros.

### 2.1. Antecedentes de la teoría de respuesta al ítem

La TRI tiene su antecedente histórico en la Teoría Clásica del Test.

Baker y Kim (2004) mencionan que los orígenes de los tests basados en los ítemes se pueden ver en el trabajo de Binet y Simon, quienes usaron una presentación tabular de la relación funcional entre la proporción de respuestas correctas a un ítem y la edad cronológica de niños. En este texto se describe brevemente el experimento de Binet y Simon para ejemplificar el concepto de la curva característica como resultado de una regresión. El procedimiento involucra ciertos pasos: se especificó una variable criterio (que en este caso es la edad cronológica), variando de 5 a 12 años con incrementos de 1 año. Se administró un ítem de interés a determinado número de niños en cada nivel de edad. La respuesta de cada sujeto se calificó como correcta o incorrecta, es decir de manera dicotómica. Se obtuvo la proporción de respuestas correctas en cada nivel de edad y se presentó en forma tabular.

Como se menciona en Baker y Kim (2004), en 1916, Terman revisó la escala Binet-Simon y graficó la proporción de respuestas correctas como función de la edad y ajustó una línea suave a los puntos de la gráfica. Esta línea ajustada es lo que actualmente se conoce como Curva Característica del Ítem (CCI).

Ambos, Terman y Binet, usaron el principio de que con el incremento de la edad, debería haber un crecimiento correspondiente en la proporción de respuestas correctas.

En Bryce (2000) se menciona que en 1943, un trabajo de Lawley extendió el análisis estadístico de las propiedades de curva ojiva normal y describió procedimientos para obtener estimaciones máxima verosimilitud de los parámetros de la CCI y que en 1952, Fred Lord introdujo la idea de rasgo o variable latente y diferenció la variable latente de los puntajes observados de la prueba. Menciona también

que en 1950 Lazarsfeld describió a las variables no observables como un resumen de las interrelaciones observadas entre las respuestas a los ítems. Bock y varios estudiantes colaboradores de la Universidad de Chicago, desarrollaron métodos de estimación efectiva y programas de cómputo como Bilog, Multilog, Parscale y Testfact. En 1960 Georg Rasch discutió la necesidad de crear modelos estadísticos que mantuvieran la propiedad de objetividad específica, esto es que los parámetros de las personas y de los ítems pudieran estimarse por separado pero que sean comparables bajo una métrica similar. Rasch inspiró la aplicación de modelos. Los trabajos de Lawley y Lord establecieron los conceptos básicos de la teoría psicométrica basada en los ítems, lo que ahora se conoce como la Teoría de Respuesta al Ítem.

*Expectativas y desarrollos recientes de la TRI*

Bock (1997) comenta sobre lo que considera son las expectativas sobre el desarrollo futuro de la TRI. Pone énfasis en tres aspectos principalmente:

a) La extensión de la estimación de la distribución latente a grupos múltiples de examinados.

b) La posible interacción entre las características del ítem y las características del grupo, esto es el funcionamiento diferencial del ítem (DIF) con respecto a grupos demográficos.

c) Quizás el desarrollo más prometedor sea el de las pruebas adaptativas por computadora. Las CAT permiten acortar el tiempo de la prueba manteniendo un nivel preespecificado de precisión de la medición. Con el advenimiento de las computadoras personales con alta capacidad multimedia las CAT son una práctica real. En las CAT se estima provisionalmente el nivel de destreza del examinado después de cada presentación de un ítem, el procedimiento dirige la selección del siguiente ítem hacia aquellos ítems de un banco, que dan más información en ese nivel de destreza (un banco de ítems es una colección de ítems para los cuales se conocen sus curvas características). Simultáneamente se calculan intervalos de confianza para el nivel de destreza verdadero hasta que el intervalo de confianza sea aceptablemente pequeño o bien hasta que se alcance el límite fijo de presentaciones. Ya que la computadora tiene la capacidad de desplegar una gran variedad de ítems visual y auditivamente y la respuesta se puede elegir por medio de un ratón o un teclado, estas pruebas son más enriquecedoras que cualquiera que se piense con sólo papel y lápiz. El uso de las CAT es muy ventajoso cuando se tiene prisa por medir, por ejemplo en mediciones relacionadas con la salud de los pacientes.

Reise et al. (2005) aborda las ventajas que tiene la TRI sobre la TCT, ilustra cómo la modelación con TRI puede mejorar las mediciones psicológicas y señala que estas ventajas harán que se use cada vez más la TRI en las aplicaciones. También, se percibe que uno de los grandes retos que enfrenta la TRI es la falta de software amigable y ampliamente disponible que dificulta su uso en un amplio sector de profesionales que trabajan con tests psicológicos. Otro de los aspectos

que actualmente se estudian es el efecto del tiempo insuficiente para responder los ítems de la prueba. Se ha documentado, Oshima (1994) que en exámenes largos como el TOEFL y el SAT ocurre este problema. En un examen puede suceder que algunos examinados no contesten algunos ítems debido a que ya no les alcanzó el tiempo para hacerlo, sin embargo en la TRI se asume que la respuesta del examinado al ítem sólo depende de la habilidad del mismo.

En Oshima (1994) se hacen simulaciones para determinar cómo afecta el problema de los ítems no contestados a las estimaciones y se concluye que se ven afectadas las estimaciones de los parámetros de los ítems y de las habilidades. Existen diversos modelos propuestos para esta situación. Por ejemplo, Goegebeur, De Boeck, Molenberghs y del Pino (2006) proponen modelos para tratar con respuestas omitidas usando un enfoque basado en influencia local.

Sinharay (2005) comenta que existe escasez de trabajos sobre verificación del modelo desde una perspectiva Bayesiana y analiza las consecuencias prácticas de la falta de ajuste del modelo. Argumenta que los paquetes de software estándar carecen de índices para medir la bondad del ajuste de los modelos de la TRI que sean confiables y que esto empeora cuando se habla de modelos TRI bayesianos, por lo cual ésta es un área que requiere más estudio.

### 2.1.1. Relación entre teoría de respuesta al ítem y teoría clásica del test

Como ya se mencionó arriba, la teoría clásica es el antecedente de la teoría de respuesta al ítem, aunque actualmente aun no ha dejado de usarse la teoría clásica del test.

Según Lord (1980), aunque en la práctica no se puedan administrar muchas formas de una prueba a un solo individuo para así obtener una mejor muestra de su comportamiento, conceptualmente es útil pensar que se hace ésto y que el individuo permanece sin cambio a través del proceso. La calificación promedio del individuo sobre un conjunto de formas de prueba es un concepto útil. La calificación obtenida por un examinado, en una forma de prueba se denota por  $X$ , entonces  $X$  se puede considerar como una variable aleatoria que variará según la forma de la prueba que se administre. Esta variable aleatoria, llamada la calificación observada, tiene alguna distribución de frecuencias usualmente desconocida. La media de esta distribución se llama la calificación verdadera del individuo y es denotada por  $T$ .

En la teoría clásica se intenta estimar la calificación verdadera del examinado, aquella que se observaría si se aplicara un número grande de veces el test al examinado y suponiendo que el examinado no aprende nada nuevo entre una y otra aplicación del test. Al promediar las calificaciones (número de ítems acertados) obtenidas en los tests se observaría la calificación verdadera del examinado. Este

modelo se expresa como sigue:

$$X = T + \xi$$

donde  $\xi$  es un término llamado el error de medición. En este modelo, que es lineal en  $T$ , el valor esperado de la calificación observada es la calificación verdadera,

$$E(X) = T$$

por lo cual se supone que el valor esperado del error es cero.

En la teoría de respuesta al ítem se intenta medir el nivel de un rasgo latente, denotado por  $\theta$ , que el examinado posee y que es el que hace que el examinado responda a los ítems de la manera observada. Este rasgo se supone subyace a los ítems del test y no es observable. Este rasgo es una variable latente continua por lo que sus valores teóricos van de  $-\infty$  a  $\infty$ . La relación entre la probabilidad de respuesta correcta al ítem  $i$ ,  $P_i(\theta)$  y  $\theta$  se modela como una función  $f$  que depende tanto de  $\theta$  como de los parámetros que caracterizan al ítem:

$$P_i(\theta) = f(\theta, \text{parámetros del ítem } i).$$

La variable latente es un continuo unidimensional que explica la covarianza entre las respuestas de los ítems. Las personas con los niveles más altos de  $\theta$  tienen una probabilidad más alta de responder acertadamente un ítem (Bryce, 2000). Los modelos TRI usan las respuestas a los ítems para obtener estimaciones de  $\theta$ , así como para estimar los parámetros del ítem, ya que cada ítem está caracterizado por uno o más parámetros.

La diferencia entre las dos teorías consiste en que mientras la teoría clásica del test se basa en el puntaje total del test, la teoría de respuesta al ítem se basa en cada ítem. Esta diferencia de enfoques hace que existan ventajas de la teoría de respuesta al ítem sobre la teoría clásica, ya que al estar la TRI basada en el ítem, es posible cuantificar la contribución de información al test en cada nivel de habilidad, de manera individual para cada ítem. En la teoría clásica la contribución de la información de un ítem es la misma en cualquier valor de la habilidad. Otra ventaja de la TRI es que la estimación de la habilidad de un examinado no depende del conjunto de ítems usados en el test, mientras que en la teoría clásica sí. Para una revisión detallada de las ventajas de la TRI, se puede consultar Bryce (2000).

### 2.1.2. Los supuestos de la TRI

Los modelos de la TRI están basados en la suposición de que los ítems están midiendo primordialmente una sola variable latente  $\theta$ , que toma valores desde  $-\infty$  hasta  $+\infty$ . A esto se le conoce como supuesto de *unidimensionalidad*. Este supuesto se puede evaluar realizando un Análisis de Factores. Para pruebas que tienen muchos ítems, la suposición de unidimensionalidad puede ser no

realista, sin embargo Cooke y Michie (1997) reportan que los modelos de la TRI son moderadamente robustos al alejamiento de la unidimensionalidad. Existen los modelos TRI multidimensionales, pero sus modelos así como documentación informativa y software amigable aun están en desarrollo (Bryce, 2000).

En los modelos de la TRI se supone que, dado  $\theta$ , las respuestas a los ítemes son independientes unas de otras, este es el supuesto de *independencia local*. Esto significa que la única relación entre los ítemes es explicada por la relación condicional con la variable latente  $\theta$ . La violación de esta suposición puede resultar en que se estimen parámetros que son diferentes de lo que serían si los datos fueran localmente independientes, entonces si la selección de los ítemes para construir el cuestionario basada en estas estimaciones no cumpliera con la independencia local, las inferencias serían incorrectas. Las suposiciones de unidimensionalidad e independencia local están relacionadas de modo que los reactivos identificados como localmente dependientes aparecerán como una dimensión separada en un análisis de factor (Bryce, 2000).

Para algunos modelos de la TRI, se supone que la variable latente es normalmente distribuida dentro de la población. Sin esta suposición, las estimaciones de  $\theta$  para algunos patrones de respuesta extremos (por ejemplo individuos examinados que no contestan correctamente ninguno de los ítemes de la escala o cuestionario) no tienen valores finitos resultando en estimaciones inestables de los parámetros.

La TRI supone que las CCI's de todos y cada uno de los ítemes es monótona creciente pues a mayor nivel de habilidad del examinado, mayor probabilidad de responder correctamente al ítem.

## 2.2. Modelos matemáticos usados en la TRI

En la TRI se describe la relación entre la habilidad del individuo y la probabilidad de respuesta correcta a un ítem a través de una expresión o modelo matemático, la cual recibe el nombre de Función de Respuesta al Ítem y la curva correspondiente es la CCI. La teoría cuenta con varios modelos matemáticos para las CCI's. En las expresiones de los modelos aparecen parámetros, cuyos valores determinan la forma y localización de la curva característica y sus propiedades técnicas, lo cual permite juzgar la calidad y pertinencia del ítem en el test. Aunque la TRI cuenta con modelos para ítemes politómicos, los modelos matemáticos presentados en este trabajo son utilizados para ítemes dicotómicos pues son el tipo de ítemes que se usaron en el examen EXHCOBA.

### 2.2.1. El modelo ojiva normal de 3 parámetros

Birnbaum (1968) introdujo el modelo ojiva normal de 3 parámetros, cuya expresión está dada como

$$P_i(\theta) = c_i + (1 - c_i) \int_{-\infty}^{a_i(\theta - b_i)} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt. \quad i = 1, \dots, n. \quad (2.1)$$

En la ecuación (2.1)  $P_i(\theta)$  es la probabilidad de acertar el ítem  $i$ . En esta ecuación dicha probabilidad es modelada por una función que involucra cuatro parámetros, tres de ellos,  $a_i$ ,  $b_i$ ,  $c_i$ , describen el funcionamiento del ítem y el cuarto es el parámetro de habilidad del examinado  $\theta$ . El parámetro de pseudo azar ( $c_i$ ) que describe la probabilidad de que un sujeto con nivel de habilidad baja ( $\theta = -\infty$ ) responda correctamente un ítem relativamente difícil, sólo por azar, es decir adivinando la opción correcta. Esto es posible ya que los ítemes son de opción múltiple y es perfectamente posible que si el examinado no conoce la respuesta correcta al ítem intente seleccionar una opción al azar. De aquí se deduce que teóricamente pueda asumir valores en el intervalo  $[0, 1]$ . El parámetro  $b_i$ , se conoce como parámetro de localización y describe la dificultad del ítem. Entre más difícil sea un ítem,  $P(\theta)$  se incrementará para valores altos de  $\theta$ . El parámetro de dificultad indica en qué lugar de la escala de habilidad el ítem discrimina o separa mejor pues los ítemes fáciles separan mejor entre los examinados de habilidad baja mientras que ítemes difíciles funcionan entre examinados de habilidad alta por lo cual se dice que la dificultad es un índice de localización del ítem (Baker y Kim, 2004), por lo tanto  $b_i$  puede asumir los mismos valores teóricos que  $\theta$ . La fórmula para determinar el valor del parámetro  $b_i$  depende del número de parámetros del modelo, por lo cual más adelante se aclara cómo determinar la dificultad del ítem cuando se vaya viendo cada modelo en específico. El parámetro  $a_i$  se conoce como parámetro de discriminación, su valor está relacionado con la pendiente de la curva en su punto de inflexión. Describe qué tan bien el ítem puede diferenciar entre examinados de habilidad por debajo de la localización y examinados de habilidad por arriba de la localización. Representa el grado en que la respuesta del ítem varía con el nivel de habilidad, es decir el escarpamiento de la CCI en su parte central. Entre más escarpada sea la curva en la región central, más discriminativo es el ítem. Las curvas características casi horizontales tienen discriminación prácticamente nula pues la probabilidad de respuesta correcta es constante para todos los niveles de habilidad.

Si se habla de la CCI correspondiente a la opción correcta (también existe una CCI para la opción incorrecta) y se asume que la CCI es monótona creciente en todo su dominio, los valores teóricos para  $a_i$  son los valores en el intervalo  $[0, +\infty]$ . Las curvas características de los modelos de la TRI tienen forma de ojiva o S elongada.

La curva descrita por el modelo ojiva normal de 3 parámetros, se muestra en la Figura 1. En ella puede observarse la identificación de los 3 parámetros del ítem.

En la curva característica mostrada en la Figura 1, se observa que  $c$  es la

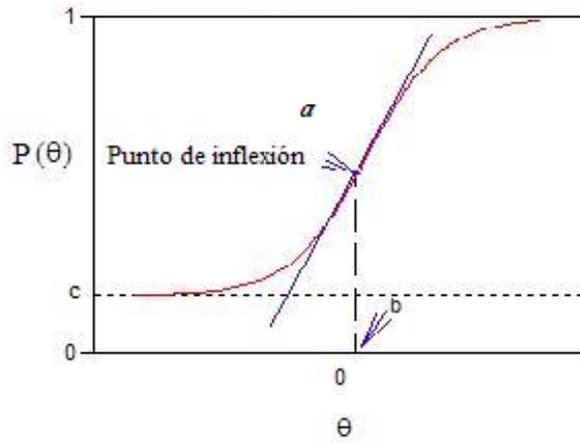


Figura 2.1: Figura 1. Curva Característica del Ítem con tres parámetros.

altura de la asíntota inferior,  $b$  es el nivel de habilidad donde se encuentra el punto de inflexión, donde la probabilidad de respuesta correcta es  $(1 + c)/2$  y  $a$  es proporcional a la pendiente de la tangente a la curva en el punto de inflexión. El valor de la pendiente es:  $a(1 - c)/\sqrt{2\pi}$ , (Lord, 1980). Observe que la integral en (2.1) es la función de distribución Normal estándar acumulada evaluada en el punto  $a_i(\theta - b_i)$ .

El modelo ojiva normal de 2 parámetros se tiene en la ecuación (2.1) cuando  $c_i = 0$ , esto es

$$P_i(\theta) = \int_{-\infty}^{a_i(\theta - b_i)} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \quad (2.2)$$

con lo que el concepto de pseudo azar no se considera en este modelo.

### 2.2.2. Modelo logístico de tres parámetros

El modelo logístico de 3 parámetros está dado por

$$P_i(\theta) = c_i + (1 - c_i) \frac{\exp[Da_i(\theta - b_i)]}{1 + \exp[Da_i(\theta - b_i)]} \quad i = 1, \dots, n. \quad (2.3)$$

donde el parámetro  $c$  tiene la misma interpretación que en el modelo ojiva normal, lo mismo sucede con el parámetro de localización,  $b$ . La curva logística tiene su punto de inflexión en  $\theta = b$ . Si  $c = 0$ , entonces  $b$  es el nivel de habilidad para el cual la probabilidad de respuesta correcta es  $1/2$ . Cuando  $c \neq 0$ ,  $b$  es el nivel de habilidad que corresponde al punto medio entre  $c$  y  $1$ . Esto es, si  $c \neq 0$ ,  $b$  es el nivel

de habilidad para el cual la probabilidad de respuesta correcta es  $\frac{1+c}{2}$ . El parámetro  $a$ , como en el caso del modelo ojiva normal, es proporcional a la pendiente de la recta tangente a la curva en el punto de inflexión; esta pendiente para la curva logística es  $0,425a(1-c)$ . La constante  $D$  es igual a 1.702 y su función es hacer que la curva logística (2.3) y la curva ojiva normal (2.1) concuerden de tal manera que la diferencia en valor absoluto entre ellas sea a lo más 0.01.

Una justificación teórica para el uso del modelo ojiva normal se encuentra en un trabajo realizado por Lord y Novick en 1968 (Baker y Kim, 2004). La justificación descansa en la regresión de una variable ítem denotada por  $\Gamma_i$ , sobre una variable de habilidad  $\theta$ . La variable ítem  $\Gamma_i$  es una variable aleatoria hipotética continua no observable sobre una población de sujetos, que representa una propensión del sujeto a responder correctamente al ítem. Ésta varía en  $-\infty < \Gamma_i < +\infty$ . Los valores grandes y positivos de  $\Gamma_i$  indican que un sujeto tiene propensión alta a responder correctamente al ítem y valores grandes negativos, lo contrario. Se puede seleccionar un valor crítico  $\gamma_i$ , llamado umbral, tal que cuando  $\Gamma_i \geq \gamma_i$ , el ítem se califica como correcto,  $u_i = 1$ ; cuando  $\Gamma_i < \gamma_i$ , el ítem se califica como incorrecto,  $u_i = 0$ .

Existen varias suposiciones, como en la teoría tradicional de regresión.

1. La relación de  $\Gamma_i$  sobre  $\theta$  es lineal.
2. En cada valor de  $\theta$ , existe una distribución condicional de  $\Gamma_i$  que es normal con media  $\mu_j$  y varianza  $\sigma_j^2$ .
3. La varianza  $\sigma_j^2$  es independiente de  $\theta$  y constante para todos los valores de  $\theta$ .
4. Los  $\theta$  tratados, son cantidades fijas que no tienen errores de medición.

Las medias de la distribución condicional de  $\Gamma_i$  caen a lo largo de la línea de regresión  $\mu'_i | \theta$ . En un nivel dado de habilidad  $\theta$ , la distribución condicional es intersectada por la línea que representa al umbral  $\gamma_i$ . El área de la distribución condicional por arriba de  $\gamma_i$  representa la probabilidad de respuesta correcta  $P_i(\theta)$  para sujetos con nivel de habilidad  $\theta$ . Si se obtienen los valores de  $P_i(\theta)$  para los correspondientes valores de  $\theta$  en todo su rango y si se grafican estos valores  $P_i(\theta)$  como una función de  $\theta$ , la curva que resulta de esto es la ojiva normal.

En la justificación pragmática, simplemente se supone que la relación funcional entre la probabilidad de respuesta correcta y la variable de habilidad  $\theta$ , sigue la ojiva normal.

El modelo Logístico de 2 parámetros que es atribuido también a Birnbaum, es la ecuación (2.3) con  $c_i = 0$ , esto es

$$P_i(\theta) = \frac{\exp[Da_i(\theta - b_i)]}{1 + \exp[Da_i(\theta - b_i)]} \quad i = 1, \dots, n. \quad (2.4)$$

### 2.2.3. Modelo logístico de 1 parámetro

Rasch en 1960 introdujo el modelo conocido como modelo Rasch, el cual es equivalente matemáticamente al modelo Logístico de dos parámetros, ecuación (2.4), con parámetro de discriminación  $a_i$  fijo e igual a 1, parámetro de dificultad  $b_i$  y sin la constante  $D$ . La expresión del modelo de Rasch es la siguiente:

$$P_i(\theta) = \frac{\exp(\theta - b_i)}{1 + \exp(\theta - b_i)} \quad i = 1, \dots, n. \quad (2.5)$$

En este modelo es más fácil entender la relación entre la habilidad y la probabilidad de acierto al ítem pues cuando el nivel de habilidad del examinado,  $\theta$ , es igual a la dificultad del ítem  $b_i$ , entonces la probabilidad de acertar el ítem es  $P_i(\theta) = 0,5$ , mientras que cuando  $\theta < b_i$ , entonces  $P_i(\theta) < 0,5$  y cuando  $\theta > b_i$ ,  $P_i(\theta) > 0,5$ .

### 2.2.4. Modelos no paramétricos

Existen los modelos no paramétricos, los cuales utilizan un supuesto relajado de monotonicidad de la CCI. Este es el caso del modelo utilizado por el programa TESTGRAF (Ramsay, 2000), el cual se utilizó en el análisis de datos del examen EXHCOBA. A fin de entender el procedimiento que utiliza TESTGRAF, se denota el vector de respuestas del examinado  $a$ -ésimo como  $(y_{a1}, y_{a2}, \dots, y_{an})$  donde  $a = 1, \dots, N$  y donde  $n$  es el número de ítems del examen. Recuerde que las respuestas son ceros y unos.

En TESTGRAF el proceso para estimar la CCI no requiere conocer los valores verdaderos de habilidad de los examinados  $\theta_a$  o sus estimaciones  $\hat{\theta}_a$ . El proceso consiste de los siguientes pasos:

1. Se asigna el rango  $r_a$  al  $a$ -ésimo examinado, a partir de  $x_a = \sum_{i=1}^n y_{ai}$ . Si existen empates en las calificaciones  $x_a$ , se rompe el empate asignando el rango de manera aleatoria.

2. Se reemplazan los rangos  $r_a$  por los cuantiles de la distribución normal estándar,  $q_a$ , entendiéndose a los cuantiles como aquellos valores que dividen el área de la función de densidad normal estándar en  $N + 1$  áreas del mismo tamaño  $1/(N + 1)$ . Se usan estos valores como si fueran los valores verdaderos de habilidad  $\theta_a$ .

3. Se ordenan los vectores de respuesta de los examinados  $(y_{a1}, y_{a2}, \dots, y_{an})$  de acuerdo a su rango de habilidad estimado  $r_a$ . Es decir el  $a$ -ésimo vector de respuestas en la prueba ordenada,  $(y_{(a)1}, \dots, y_{(a)n})$  es el del  $a$ -ésimo examinado por el tamaño de  $x_a$ .

4. Para el  $i$ -ésimo ítem se estima  $P_i(\theta)$  suavizando la relación entre los valores 0-1, de  $y_{ia}$  y los valores de  $\theta_a$ .

El suavizamiento se realiza usando un kernel Gaussiano con un parámetro de ancho de banda  $h$  (el ancho del intervalo dentro del cual se promediarán los valores de  $\theta$ ) dado por default en TESTGRAF pero que el usuario puede modificar. Para la opción correcta del ítem  $i$ , se estima  $P_i(\theta)$  como un promedio ponderado de los valores  $y_{ia}$  de todos los  $N$  examinados. Es decir, se usa la estimación

$$P_i(\theta) = \sum_{a=1}^N w_a(\theta) y_{ai} \quad (2.6)$$

con

$$w_a(\theta) = \frac{K[(q_a - \theta)/h]}{\sum_k K[(q_k - \theta)/h]} \quad (2.7)$$

de modo que la estimación suavizada viene a ser

$$P_i(\theta) = \frac{\sum_a K\left[\frac{q_a - \theta}{h}\right] y_{ai}}{\sum_a K\left[\frac{q_a - \theta}{h}\right]}.$$

El valor del parámetro  $h$  que usa TESTGRAF en la ecuación (2.7) es por default  $h = 1.1N^{-1/5}$ . En Ramsay (1991) se justifica el valor de  $h$  argumentando que se ha visto en varios experimentos que usan una distribución Normal estándar para la habilidad y un kernel Gaussiano, que  $h = N^{-1/5}$  funciona bien. El kernel Gaussiano es  $K(u) = \exp(-u^2/2)$ . Con el fin de economizar tiempo de cómputo, pues en algunas aplicaciones  $N$  puede ser muy grande, en lugar de promediar los valores  $y_{ia}$  de todos los examinados, se eligen  $Q$  valores de despliegue igualmente espaciados, en la escala de la habilidad. TESTGRAF usa por default  $Q = 51$ , lo que origina que la escala se divida en 50 intervalos de igual ancho, como en un proceso de cribado.  $Q = 51$  implica que se toman valores  $\theta$  de  $-2.5$  a  $2.5$  con incrementos de  $0.1$ . Entonces para  $q = 1$ ,  $\theta_q = -2.5$ , mientras que para  $q = Q = 51$ ,  $\theta_q = 2.5$ . Suponga que se desea estimar la curva característica del ítem en el valor de despliegue  $\theta_q$ . TESTGRAF calcula el promedio de los valores  $y_{ia}$  para valores de  $\theta_a$  que caen dentro de los límites dados por los centros de los intervalos adyacentes  $[\theta_{q-1}, \theta_q]$  y  $[\theta_q, \theta_{q+1}]$ ; es decir dentro de  $(\theta_{q-1} + \theta_q)/2$  y  $(\theta_q + \theta_{q+1})/2$ . Para el valor de despliegue más pequeño, se toma el promedio de los valores  $y_{ia}$ , correspondientes a aquellos  $\theta_a$  que caen por debajo del centro del primer intervalo, y para el valor de despliegue más grande se toma el promedio de los valores que caen por arriba del centro del último intervalo. Denote estos  $Q$  promedios por  $P_{iq}$ . Al mismo tiempo, se calcula el área bajo la curva Normal estándar entre estos centros de intervalos. Denote estas áreas por  $\phi_q$ . Los valores promedio son entonces suavizados por la ecuación

$$P_i(\theta_q) = \sum_{r=1}^Q w_{rq} P_{ir}. \quad (2.8)$$

Los pesos

$$w_{rq} = \frac{\phi_r K [(\theta_r - \theta_q) / h]}{\sum_{s=1}^Q K [(\theta_s - \theta_q) / h]} \quad (2.9)$$

no dependen del ítem ni de la opción, por lo cual se pueden calcular una sola vez y almacenar en una matriz para ser usados en los cálculos de todas las CCI.

El programa permite usar dos tipos de variables de habilidad como variables independientes en el eje horizontal de las gráficas de las CCI's. Una de ellas es la calificación esperada de la prueba y es el número esperado de ítemes correctos que un examinado en un nivel de habilidad particular tendrá. Su expresión es la siguiente:

$$\eta(\theta) = \sum_{i=1}^n P_i(\theta) \quad (2.10)$$

donde  $P_i(\theta)$  es la probabilidad de respuesta correcta al ítem  $i$  al nivel de habilidad  $\theta$ . La segunda variable consta de los cuantiles de la distribución Normal estándar. En las gráficas del presente trabajo se usó la segunda variable, es decir los cuantiles de la Distribución Normal estándar.

### 2.2.5. Funciones de información

Es de interés examinar las varianzas de las habilidades estimadas, sobre la escala de habilidad completa porque de esta manera se aprecia en qué niveles de habilidad mide con más precisión el test. Esto se realiza a través de la función de información de la prueba y de las funciones de información de los ítemes. La función de información de la prueba está compuesta por la contribución de información de los ítemes, dicha contribución se da a través de sus funciones de información. La función de información del ítem, al igual que la de la prueba, señala en qué lugar de la escala de habilidad mide con más precisión.

El concepto de la cantidad de información  $I(\theta)$  se debe a R. A. Fisher y es el recíproco de la varianza de una estimación  $I(\theta) = 1/\sigma_\theta^2$ . Entonces, entre más grande es la varianza, menos precisa es la estimación de  $\theta$  y menos información se tiene del nivel de habilidad desconocido del examinado (Baker y Kim, 2004).

#### 2.2.5.1. Función de información del ítem

La expresión matemática explícita para la cantidad de información del ítem depende del modelo empleado para la CCI y de los parámetros de las CCI's. Birnbaum en 1968 definió la función de información de la prueba como

$$I(\theta) = \sum_{i=1}^n \frac{[P'_i(\theta)]^2}{P_i(\theta) Q_i(\theta)}, \quad (2.11)$$

esta expresión puede interpretarse como la suma de las contribuciones de información de los ítemes. A la contribución del ítem se le conoce como la función de

información del ítem

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta) Q_i(\theta)}.$$

Ya que la cantidad de información se puede determinar en cada nivel de habilidad,  $\theta$ , se puede graficar a  $I_i(\theta)$  y a  $I(\theta)$  como una función de la habilidad. En el caso de la función de información de la prueba, el análisis de la gráfica de  $I(\theta)$ , permite encontrar el valor de  $\theta$  donde la información es máxima. Esto indica el valor de habilidad donde el test mide con más precisión.

La función de información del ítem para distintos modelos de la CCI (Baker y Kim, 2004) es:

Ojiva Normal

$$I_i(\theta) = a_i^2 \left\{ \frac{[h_i(\theta)]^2}{P_i(\theta) Q_i(\theta)} \right\},$$

Logístico de 2 parámetros

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta) Q_i(\theta)} = a_i^2 P_i(\theta) Q_i(\theta),$$

Rasch

$$I_i(\theta) = P_i(\theta) Q_i(\theta),$$

Logístico de 3 parámetros

$$I_i(\theta) = a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[ \frac{P_i(\theta) - c_i}{1 - c_i} \right]^2 = a_i^2 P_i(\theta) Q_i(\theta) \left[ \frac{P_i^*(\theta)}{P_i(\theta)} \right]^2$$

donde  $P_i^*(\theta) = \frac{P_i(\theta) - c_i}{1 - c_i}$ . La expresión para la función de información bajo un modelo de tres parámetros involucra un término extra resultante del parámetro de pseudo azar,  $c_i$ .

Cuando se usan los modelos Normal y Logístico de dos parámetros para la CCI, el máximo ocurre cuando  $\theta = b_i$ . El máximo de la función de información del ítem bajo estos modelos es (Baker y Kim, 2004):

(a) Ojiva Normal

$$\max [I_i(\theta)] = a_i^2 \frac{[.3989]^2}{(.5)(.5)} = .64a_i^2$$

(b) Logístico

$$\max [I_i(\theta)] = a_i^2 (.5)(.5) = .25a_i^2$$

(c) Rasch

$$\max [I_i(\theta)] = (.5)(.5) = 0.25$$

donde  $a_i$  es el parámetro de discriminación.

En todos los casos, la cantidad de información puede ser bastante pequeña para niveles de habilidad que se desvían considerablemente de  $b_i$ . Esto significa que la estimación de la habilidad es más precisa cuando la dificultad del ítem,  $b_i$ , coincide con la habilidad del examinado.

El máximo de la función de información bajo el modelo Logístico de tres parámetros de Birnbaum no se da en  $\theta = b_i$ . Birnbaum derivó la siguiente expresión para el nivel de habilidad en el cual se da el máximo:

$$\theta_{\text{máx}} = b_i + \frac{1}{a_i} \log \left( \frac{1 + \sqrt{1 + 8c_i}}{2} \right).$$

Cuando el parámetro de pseudo azar,  $c_i$  es cero, el término  $\frac{1 + \sqrt{1 + 8c_i}}{2}$  produce un valor de 1. En este caso, la función de información del modelo coincide con la función de información del modelo Logístico de dos parámetros. Como resultado la función de información del modelo Logístico de dos parámetros es la cota superior para la función de información del modelo Logístico de tres parámetros, dados los mismos valores de  $a_i$  y  $b_i$ .

### 2.2.5.2. Función de información de la prueba

Como ya se mencionó, la función de información de la prueba proporciona una medida de la precisión con la cual los  $n$  ítems están estimando la habilidad a lo largo de toda la escala de habilidad y es la suma de las funciones de información de los  $n$  ítems. Entonces, se pueden obtener las expresiones para la función de información de la prueba bajo diversos modelos, usando los resultados de arriba para la cantidad de información del ítem. Las fórmulas son las siguientes (Baker y Kim, 2004):

(a) Ojiva Normal

$$I(\theta) = \sum_{i=1}^n \frac{\alpha_i^2 [h_i(\theta)]^2}{P_i(\theta) Q_i(\theta)}$$

(b) Logístico de 2 parámetros

$$I(\theta) = \sum_{i=1}^n \alpha_i^2 P_i(\theta) Q_i(\theta)$$

(c) Rasch

$$I(\theta) = \sum_{i=1}^n P_i(\theta) Q_i(\theta)$$

(d) Logístico de 3 parámetros

$$I(\theta) = \sum_{i=1}^n \alpha_i^2 P_i(\theta) Q_i(\theta) \left[ \frac{P_i^*(\theta)}{P_i(\theta)} \right]^2.$$

La forma de la función de información de la prueba depende de la mezcla de valores de los parámetros de los ítemes en la prueba.

En el caso del modelo no paramétrico usado por Ramsay (2000) en TESTGRAF para las CCI, se indica que la estimación de la función de información presenta problemas estadísticos muy serios y que la mejor alternativa para su estimación aún está siendo investigada. Puede ser que los problemas para estimar la función de información a que se refiere Ramsay tengan que ver con  $P'_i(\theta)$ , véase la ecuación (2.11), pues en el caso no paramétrico solo se tienen puntos  $(\theta_j, P_i(\theta_j))$  y no una curva como en el caso de los modelos paramétricos. La forma en que Ramsay estima las habilidades de los examinados es primero estimando los parámetros del modelo logístico cuadrático y una vez construidas las CCI con estos parámetros, las utiliza para estimar por el método de Máxima Verosimilitud las habilidades  $\theta_a$ .

## 2.3. Evaluación de los supuestos de la TRI

### 2.3.1. Evaluación del supuesto de monotonidad

Como ya se mencionó la Teoría de Respuesta al Ítem supone que todas las curvas características de los ítemes son monótonas crecientes en todo el dominio de  $\theta$ , es decir que a medida que el nivel del examinado aumenta también aumenta la probabilidad de acertar el ítem. Este supuesto se puede explorar para cada ítem, graficando para un ítem particular, el nivel de habilidad contra la proporción de respuestas correctas obtenidas por los examinados que tuvieron ese nivel de habilidad. Se espera observar en la gráfica que los puntos tengan un comportamiento monótono creciente, es decir que cuando el nivel de habilidad aumenta, la proporción de aciertos también aumenta. Otra manera de verificarlo es observar las CCI obtenidas por un método no paramétrico, ya que estos métodos no suponen la monotonidad de las CCI en su construcción. En este caso también debe observarse en la gráfica que los puntos o la curva ajustada a dichos puntos, tengan un comportamiento monótono creciente.

La naturaleza de los modelos de la TRI implican que todas y cada una de las CCI correspondientes a los ítemes del examen son funciones monótonas crecientes. En las CCI paramétricas, este comportamiento se da porque la función logaritmo y la distribución normal acumulada son funciones monótonas crecientes. Sin embargo, es perfectamente posible en algunas aplicaciones, como se muestra aquí, que algunos ítemes puedan tener problemas para satisfacer este supuesto. Sijtsma y Junker (2006) mencionan que no hay una razón lógicamente convincente por la cual se dé en datos reales y que es frecuente encontrar que las CCI decrezcan significativamente en uno o más intervalos de  $\theta$ . Argumentan que esto puede deberse a que los examinados en diversos niveles de habilidad implementen diferentes estrategias que varían en grado de complejidad y asertividad y que por ejemplo los

examinados con niveles bajos de  $\theta$  pueden usar una estrategia simple e incorrecta, mientras los de nivel medio pueden ser una estrategia innecesariamente compleja y parcialmente correcta y los examinados de nivel alto una estrategia simple y correcta. También dicen que puede suceder que para algunos ítemes, no todos, la compleja estrategia intermedia de  $\theta$  lleve a los examinados con mayor frecuencia a despistarlos con una respuesta incorrecta que una simple e incorrecta estrategia que por accidente produzca varias respuestas correctas, justo como sucedería con el lanzamiento de una moneda. Esto tendría como resultado que se observe un ligero descenso de la CCI en la parte central de la distribución de  $\theta$ .

### 2.3.2. Evaluación del supuesto de unidimensionalidad de los ítemes

En este estudio el rasgo latente que se mide es la habilidad en Matemáticas para el Cálculo. Una de las herramientas estadísticas utilizada frecuentemente para verificar este supuesto es el Análisis de Factores, el cual es adaptado para funcionar con variables binarias. En el Análisis de Factores el foco de interés es la explicación de la estructura de covarianza o de correlación, o de ambas, entre las variables medidas (Johnson, 2000).

El Análisis de Factores estándar se aplica a variables con nivel de medición intervalar al menos, éste no es el caso de los ítemes dicotómicos. Lo que se hace en este caso es aplicar el Análisis de Factores sobre la matriz de correlaciones tetracóricas estimadas, las cuales son las estimaciones máximo verosímil de las correlaciones de Pearson que habría entre las variables continuas originales (Ferrando, 1996).

El coeficiente de correlación tetracórica se usa cuando ambas variables son dicotómicas pero se hipotetiza que en realidad ambas variables son continuas y normalmente distribuidas (Manual de TESTFACT, Wood R., Wilson D., Gibbons R., Schilling S., Muraki E. y Bock D., 2003), de modo que la dicotomía se hace artificialmente. La fórmula involucra a la función trigonométrica coseno. La fórmula es  $R_{tet} = \cos(180^\circ / (1 + \sqrt{(BC/AD)})$  (Wikipedia, La Enciclopedia Libre) si las respuestas de una pareja de ítemes  $i, j$  calificados dicotómicamente se acomodan en una tabla de contingencia de 2x2 como la siguiente, donde 0 indica respuesta incorrecta, 1 indica respuesta correcta y A,B,C y D son las frecuencias de las respuestas a ambos ítemes:

$$\begin{array}{cc} & 0 & 1 \\ 0 & A & B \\ 1 & C & D \end{array} .$$

El análisis de Factores se puede aplicar de dos maneras a los datos. El Análisis de Factores Exploratorio (AFE) y el Análisis de Factor Confirmatorio (AFC). En el AFE no se impone ninguna estructura hipotetizada de la prueba en el análisis, mientras que en el AFC sí se hipotetiza una estructura y se formulan hipótesis que se pueden evaluar directamente. El AFC es particularmente apropiado cuando se

desea evaluar hipótesis de estructura simple o casi simple, en la cual los ítemes son determinados por una o primordialmente una dimensión.

Los métodos de Análisis de Factor Lineal analizan las correlaciones entre ítemes para identificar factores que puedan usarse para explicar las relaciones observadas entre los ítemes. El número de factores determina la dimensionalidad de la prueba y las cargas de factor determinan qué ítemes deben usarse como indicadores para interpretar cada factor. Sin embargo el análisis de factores para ítemes dicotómicos presenta una serie de problemas, entre otras cosas por el uso de correlaciones tetracóricas.

Adicionalmente se hipotetiza que para cada par de ítemes las dos correspondientes variables latentes continuas tienen una distribución normal bivariada (Wood R., Wilson D., Gibbons R., Schilling S., Muraki, E. y Bock, D., 2003).

Desde un punto de vista aplicado, el uso de correlaciones tetracóricas en el Análisis de Factores presenta problemas. Un primer problema es que la matriz de correlaciones tetracóricas no siempre es definida positiva por lo cual la matriz se somete a un proceso de “suavizamiento” que la convierte en definida positiva. Un segundo problema se da cuando en algunas celdas de las tablas de contingencia entre cada par de ítemes aparecen frecuencias muy bajas e incluso pueden ser cero. TESTFACT llama par de ítemes *no válido* cuando existe alguna frecuencia marginal 0 en la tabla de contingencia. Ejemplos de pares de ítemes no válidos son los siguientes, donde  $C$  denota la respuesta correcta e  $I$ , la incorrecta:

$$\begin{array}{cc} & C \quad I \\ C & \quad 0 \end{array}$$

$$\begin{array}{cc} I & \quad 0 \\ & C \quad I \end{array}$$

$$\begin{array}{ccc} C & 0 & 0 \\ I & & \end{array}$$

$$\begin{array}{cc} & C \quad I \\ C & \quad 0 \\ I & \quad 0 \end{array}$$

En estos casos, TESTFACT sustituye una correlación tetracórica de 1 en la matriz de correlaciones tetracóricas (Manual del TESTFACT).

En este programa el problema de las raíces latentes negativas se corrige sustituyendo el valor de la raíz negativa por cero o por un valor positivo pequeño. Luego multiplica cada raíz latente por un factor que hace que la suma de las raíces latentes vuelva a ser igual al número de ítemes. La matriz de correlaciones tetracóricas con las raíces latentes corregidas se llama *matriz de correlaciones suavizada*.

Según Lord (1980) un procedimiento burdo para examinar la unidimensionalidad de un cuestionario es calcular las raíces latentes de la matriz de correlaciones tetracóricas con las comunales estimadas colocadas en la diagonal y (1) si la primer raíz es grande comparada con la segunda y (2) la segunda raíz no es mucho

más grande que cualquiera de las otras, entonces los ítemes son aproximadamente unidimensionales.

### 2.3.3. Evaluación del supuesto de independencia local

La propiedad de independencia local de los ítemes afirma que la probabilidad de éxito sobre un ítem depende de los parámetros del ítem y de la habilidad,  $\theta$ , del examinado y nada más. Si el modelo es cierto, la habilidad de la persona es todo lo que se necesita saber para determinar su probabilidad de éxito sobre un ítem específico. Si conocemos la habilidad del examinado, cualquier conocimiento sobre su éxito o falla sobre otros ítemes no agregará nada a esta determinación (Lord, 1980). Si  $u_i = 0$  o  $1$  denota la calificación del ítem  $i$ , entonces el supuesto de independencia local se puede escribir de manera compacta como

$$P(u_i = 1 | \theta) = P(u_i = 1 | \theta, u_j, u_k, \dots) \quad (i \neq j, k, \dots).$$

Una afirmación matemáticamente equivalente de independencia local es que la probabilidad condicional de éxito sobre todos los ítemes es igual al producto de las probabilidades condicionales de éxito separadas, esto es

$$P[u_1 = 1, \dots, u_n = 1 | \theta] = \prod_{i=1}^n P(u_i = 1 | \theta).$$

Lord (1980) hace notar que la independencia local se sigue automáticamente de la unidimensionalidad, no es un supuesto adicional. La independencia local requiere que cualquier par de ítemes sean no correlacionados cuando  $\theta$  está fijo. Una forma de verificar este supuesto es mediante la utilización del estadístico Mantel-Haenzel. En Sinharay (2005) se expone el método, el cual es descrito enseguida. Para cada par de ítemes es posible definir un cociente de momios condicional sobre la calificación bruta de los examinados y luego combinarlos. Se define el cociente de momios para un par de ítemes condicional sobre la calificación remanente (esto es, la calificación bruta sobre la prueba, obtenida al excluir los dos ítemes)  $r$  como

$$OR_r = \frac{n_{11r}n_{00r}}{n_{10r}n_{01r}},$$

donde  $n_{kk'r}$  es el número de sujetos con calificación remanente  $r$  que obtuvieron una calificación  $k$  sobre un ítem y  $k'$  sobre el otro,  $k, k' = 0, 1$ . Entonces, es posible combinarlos en un cociente de momios condicional acumulado conocido como el estadístico Mantel-Haenzel, como:

$$MH = \frac{\sum_r n_{11r}n_{00r}/n_r}{\sum_r n_{10r}n_{01r}/n_r},$$

donde  $n_r$  es el número de examinados que obtuvieron la puntuación remanente  $r$ . Este estadístico debe ser útil en la detección de falta de independencia local en un modelo unidimensional de TRI. Si la independencia local se cumple, la covarianza condicional entre las calificaciones de los dos ítemes es cercana a cero y el estadístico MH debe acercarse a 1; si la independencia local se viola, la covarianza condicional es positiva (lo que significa que MH será mayor a 1) para ítemes dentro de un conglomerado y negativa (estadístico MH menor que 1) para pares de ítemes entre conglomerados. Entonces, si la independencia local se viola, se espera que el modelo unidimensional de la TRI subvalore al estadístico MH para ítemes que pertenecen a un mismo conglomerado y sobrevalore para ítemes de diferentes conglomerados. Aquí se debe entender como "ítemes dentro de un conglomerado", aquellos ítemes que están influenciados por un solo rasgo que es diferente del rasgo que la prueba intenta medir, mientras los "ítemes de diferentes conglomerados" son aquellos que están influenciados por diferentes rasgos.

## 2.4. Procedimientos de estimación de parámetros

Primeramente debe ser claro que se necesita estimar dos conjuntos de parámetros: los  $N$  parámetros de habilidad de los examinados ( $\theta_1, \dots, \theta_N$ ) y los parámetros de los  $n$  ítemes. El número de estos últimos depende de cuántos parámetros tiene el modelo que se use. Por ejemplo si se usa un modelo de tres parámetros entonces se tienen que estimar  $3n$  parámetros de los ítemes.

Un texto donde se pueden consultar los principales procedimientos de estimación de parámetros en la TRI, tanto de habilidad como del ítem, es Baker y Kim (2004). Para esta sección se resumió la información acerca de algunos de los métodos de estimación, presentados en esta fuente. Si se desea ver los detalles de los métodos, se recomienda acudir a dicho texto.

Se presenta el método de Estimación Máxima Verosimilitud (MV) de los parámetros del ítem, para el modelo Ojiva Normal y Logístico, ambos de dos parámetros. Luego se presentan los mismos métodos anteriores pero para la estimación de los parámetros de los examinados. Luego se presenta el método Marginalizado de Bayes para los parámetros del ítem y dos métodos Bayesianos para la estimación de los parámetros de los examinados. Finalmente se presenta la alternativa del Muestreador de Gibbs (para la estimación de parámetros de ítemes y de examinados) que es una alternativa Bayesiana implementada en WINBUGS. Todos los procedimientos de estimación se pueden ver con detalle en Baker y Kim (2004). Es precisamente de este texto, de donde se extrae toda la información que se presenta en los métodos de estimación de esta sección.

### 2.4.1. Estimación MV de los parámetros del ítem, modelo ojiva normal 2 parámetros

En las ecuaciones (2.2) hasta (2.5) se observa que la probabilidad de acierto de un ítem en realidad depende del nivel de habilidad y de los parámetros del ítem. En el método de estimación MV, para estimar los parámetros del ítem se requiere conocer los parámetros de habilidad y para estimar los parámetros de habilidad se requiere conocer los del ítem. Por esta razón, en esta subsección y en la siguiente se suponen conocidos los parámetros de los examinados  $\theta_j$ , con  $j = 1, \dots, N$ .

El método MV utiliza la parametrización de una recta en su forma intercepción-pendiente ( $Z_j = \zeta + \lambda\theta_j$ ). Bajo el modelo Ojiva Normal  $P_j = P(\zeta, \lambda, \theta_j) = \int_{-Z_j}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$  con  $Z_j = \zeta + \lambda\theta_j$ . Como primer paso del proceso se forman grupos de examinados de acuerdo a su nivel de habilidad. Se forman  $k$  grupos de  $f_j$  sujetos que poseen calificación de habilidad conocida  $\theta_j$ ,  $j = 1, 2, \dots, k$  los cuales son tomados al azar de una población de personas.

De los  $f_j$  sujetos que tienen habilidad  $\theta_j$ ,  $r_j$  sujetos responden correctamente al ítem (se usa un sólo ítem) y  $f_j - r_j$  responden incorrectamente. Sea  $R = (r_1, r_2, \dots, r_k)$  el vector del número de respuestas correctas observadas.

El logaritmo de la verosimilitud es

$$L = \log [P(R)] = \text{constante} + \sum_{j=1}^k r_j \log P_j + \sum_{j=1}^k (f_j - r_j) \log Q_j. \quad (2.12)$$

Las estimaciones máxima verosimilitud de los parámetros desconocidos  $\zeta$  y  $\lambda$  están basadas en la información proporcionada por la muestra y son los valores que maximizan la función logaritmo de verosimilitud  $L$  de la ecuación (2.12); es decir, satisfacen las ecuaciones:

$$\frac{\partial L}{\partial \zeta} = 0 \text{ y } \frac{\partial L}{\partial \lambda} = 0. \quad (2.13)$$

Para resolver las ecuaciones de verosimilitud (2.13), se emplea un procedimiento iterativo basado en una serie de Taylor. Se dan aproximaciones  $\hat{\zeta}_t, \hat{\lambda}_t$  para  $t = 0$  y se calculan  $\hat{\zeta}_{t+1}, \hat{\lambda}_{t+1}$ , utilizando incrementos, de modo que  $\hat{\zeta}_{t+1} = \hat{\zeta}_t + \Delta\hat{\zeta}$  y  $\hat{\lambda}_{t+1} = \hat{\lambda}_t + \Delta\hat{\lambda}$ .

La ecuación iterativa para obtener estimaciones de  $\zeta$  y  $\lambda$ , conocida como ecuación puntuación scoring de Fisher es:

$$\begin{bmatrix} \hat{\zeta} \\ \hat{\lambda} \end{bmatrix}_{t+1} = \begin{bmatrix} \hat{\zeta} \\ \hat{\lambda} \end{bmatrix}_t - \begin{bmatrix} -\sum_{j=1}^k f_j W_j & -\sum_{j=1}^k f_j W_j \theta_j \\ -\sum_{j=1}^k f_j W_j \theta_j & -\sum_{j=1}^k f_j W_j \theta_j^2 \end{bmatrix}_t^{-1} \begin{bmatrix} \sum_{j=1}^k f_j W_j v_j \\ \sum_{j=1}^k f_j W_j v_j \theta_j \end{bmatrix}_t \quad (2.14)$$

donde  $W_j = h_j^2 / P_j Q_j$ ,  $v_j = (p_j - P_j) / h_j$ ,  $p_j = r_j / f_j$  y  $h_j = \frac{1}{\sqrt{2\pi}} e^{-Z_j^2/2}$ .

La solución iterativa de la ecuación (2.14) se conoce como "método de puntuación Scoring" de Fisher. Las ecuaciones para calcular los incrementos son:

$$\Delta \hat{\lambda}_t = \frac{\sum_{j=1}^k f_j W_j v_j \theta_j - \left[ \sum_{j=1}^k f_j W_j \theta_j \sum_{j=1}^k f_j W_j v_j / \sum_{j=1}^k f_j W_j \right]}{\sum_{j=1}^k f_j W_j \theta_j^2 - \left[ \left( \sum_{j=1}^k f_j W_j \theta_j \right)^2 / \sum_{j=1}^k f_j W_j \right]} \quad (2.15)$$

y

$$\Delta \hat{\zeta}_t = \frac{\bar{v} \sum_{j=1}^k f_j W_j \theta_j^2 - \bar{\theta} \sum_{j=1}^k f_j W_j v_j \theta_j}{\sum_{j=1}^k f_j W_j \theta_j^2 - \left[ \left( \sum_{j=1}^k f_j W_j \theta_j \right)^2 / \sum_{j=1}^k f_j W_j \right]}. \quad (2.16)$$

Una vez que se estiman los parámetros de la línea de regresión  $\zeta$  y  $\lambda$ , se obtienen las estimaciones de los parámetros del ítem por medio de las ecuaciones:

$$\begin{aligned} \hat{a}_i &= \hat{\lambda} \\ \hat{b}_i &= -\frac{\hat{\zeta}}{\hat{\lambda}} \end{aligned} \quad (2.17)$$

donde  $\hat{a}_i$  es la estimación del parámetro de discriminación del ítem  $i$  y  $\hat{b}_i$  es la estimación del parámetro de dificultad del ítem  $i$ .

#### 2.4.2. Estimación MV de los parámetros del ítem, modelo logístico 2 parámetros

Para el modelo Logístico, la probabilidad de responder acertadamente al ítem, que tiene un examinado de habilidad  $\theta_j$ , es

$$P_j = P(\zeta, \lambda, \theta) = \frac{1}{1 + e^{-(\zeta + \lambda \theta_j)}}. \quad (2.18)$$

Sea  $Z_j = \zeta + \lambda \theta_j$  la desviación logística, esto es, el logit. La función logaritmo de la verosimilitud, es la misma que en el modelo Ojiva Normal, ecuación (2.12) donde  $W_j = P_j Q_j$  y  $v_j = (p_j - P_j) P_j Q_j$  en el caso del modelo Logístico.

Para este modelo la ecuación que determina el proceso iterativo para estimar los parámetros es

$$\begin{bmatrix} \hat{\zeta} \\ \hat{\lambda} \end{bmatrix}_{t+1} = \begin{bmatrix} \hat{\zeta} \\ \hat{\lambda} \end{bmatrix}_t - \begin{bmatrix} -\sum_{j=1}^k f_j W_j & -\sum_{j=1}^k f_j W_j \theta_j \\ -\sum_{j=1}^k f_j W_j \theta_j & -\sum_{j=1}^k f_j W_j \theta_j^2 \end{bmatrix}_t^{-1} \begin{bmatrix} \sum_{j=1}^k f_j W_j v_j \\ \sum_{j=1}^k f_j W_j v_j \theta_j \end{bmatrix}_t \quad (2.19)$$

donde  $v_j = (p_j - P_j) / P_j Q_j$ . Esta ecuación es idéntica a la ecuación (2.14), por lo cual las expresiones para los incrementos  $\Delta \hat{\zeta}$  y  $\Delta \hat{\lambda}$  son las mismas que las expresiones (2.15) y (2.16).

### 2.4.3. Estimación MV de los parámetros de los examinados, modelo ojiva normal

El método de estimación MV de los parámetros de habilidad supone que

- 1) se conocen los parámetros de los  $n$  ítemes,
- 2) que los examinados son objetos independientes,
- 3) que la habilidad se puede estimar para un examinado a la vez y
- 4) que todos los ítemes de la prueba se modelan utilizando CCI's de la misma familia.

Suponiendo ahora conocidos los parámetros del ítem, se estiman los parámetros de habilidad de los examinados. Un examinado dado responde a los  $n$  ítemes de una prueba y las respuestas se califican de manera dicotómica,  $u_{ij} = 0, 1$ , donde  $i = 1, \dots, n$  y  $j$  designa al examinado  $j = 1, \dots, N$ , produciendo un vector (de longitud  $n$ ) de respuestas a los ítemes denotado por  $U_j = (u_{1j}, u_{2j}, u_{3j}, \dots, u_{nj} \mid \theta_j)$ . Para simplificar la notación, sea  $P_i(\theta_j) = P_{ij}$ ,  $Q_i(\theta_j) = Q_{ij}$ ,  $Z_{ij} = \zeta_i + \lambda_i \theta_j$ ,  $h_{ij} = (1/\sqrt{2\pi})e^{-Z_{ij}^2/2}$  y  $W_{ij} = h_{ij}^2/P_{ij}Q_{ij}$  donde  $P_{ij}$  es el modelo Ojiva Normal de 2 parámetros. El logaritmo de la función de verosimilitud, es

$$L = \log P(U_j \mid \theta_j) = \sum_{i=1}^n [u_{ij} \log P_{ij} + (1 - u_{ij}) \log Q_{ij}]. \quad (2.20)$$

Para un examinado particular, se establece la ecuación de Newton-Raphson que es resuelta iterativamente por la estimación máxima verosimilitud de la habilidad del examinado:

$$\left[ \hat{\theta}_j \right]_{t+1} = \left[ \hat{\theta}_j \right]_t - \left[ \frac{\sum_{i=1}^n \lambda_i W_{ij} \left( \frac{u_{ij} - P_{ij}}{h_{ij}} \right)}{-\sum_{i=1}^n \lambda_i^2 W_{ij}} \right]_t. \quad (2.21)$$

La varianza muestral de  $\theta_j$  es dada por

$$S_{\hat{\theta}_j}^2 = \frac{1}{\sum_{j=1}^n \lambda_i^2 W_{ij}}$$

### 2.4.4. Estimación MV de los parámetros de los examinados, modelo logístico

Si se está trabajando con un modelo Logístico de dos parámetros para la CCI, ecuación (2.4), la ecuación de puntuación Scoring de Fisher es

$$\left[ \hat{\theta}_j \right]_{t+1} = \left[ \hat{\theta}_j \right]_t - \left[ \frac{\sum_{i=1}^n \lambda_i W_{ij} \left( \frac{u_{ij} - P_{ij}}{P_{ij}Q_{ij}} \right)}{-\sum_{i=1}^n \lambda_i^2 W_{ij}} \right]_t. \quad (2.22)$$

La varianza muestral de  $\widehat{\theta}_j$  es dada por

$$S_{\widehat{\theta}_j}^2 = \frac{1}{\sum_{i=1}^n \lambda_i^2 W_{ij}}. \quad (2.23)$$

#### 2.4.5. Estimación Bayesiana de los parámetros del ítem

El método de estimación MV tiene algunos problemas. Entre estos, que las estimaciones de los parámetros y de los ítemes pueden asumir valores no razonables. La alternativa Bayesiana evita estas dificultades.

Las alternativas Bayesianas en la TRI, se pueden distinguir por la característica de si las estimaciones de los parámetros del ítem tienen lugar antes o después de la marginalización (esto es, la integración) sobre los parámetros de habilidad o sin marginalización (Baker y Kim, 2004).

La alternativa descansa en el Teorema de Bayes, el cual proporciona una forma de expresar la probabilidad condicional. En esta alternativa combina información que proviene de datos muestrales, con probabilidades obtenidas usando información a priori acerca de la distribución del conjunto de parámetros desconocidos. Una aplicación del teorema de Bayes produce una distribución de probabilidad posterior que es proporcional al producto de la función de verosimilitud y la distribución de probabilidad a priori. Esta distribución de probabilidad posterior se usa para hacer inferencias acerca de los parámetros desconocidos (Lindley, 1971).

#### 2.4.6. Estimación marginalizada Bayesiana de los parámetros del ítem

Se desea hacer inferencias acerca de todos los parámetros desconocidos, en este caso,  $\theta, \tau, \xi$  y  $\eta$ , donde  $\theta$  es el vector de parámetros de habilidad,  $\xi$  el vector de parámetros del ítem,  $\tau$  y  $\eta$  son los parámetros de las distribuciones de  $\theta$  y  $\xi$  respectivamente. A  $\tau$  y  $\eta$  se les conoce como hiperparámetros.

La solución del siguiente sistema de ecuaciones proporciona las estimaciones modales de Bayes, BME's, de los parámetros del ítem:

$$\frac{\partial}{\partial v_i} [\log L(U | \xi, \tau)] + \frac{\partial}{\partial v_i} [\log g(\xi)] = 0, \quad (2.24)$$

donde  $v_i$ ; es un parámetro del ítem  $i$ , como puede ser  $a_i, b_i$  o  $c_i$  y  $U = (U_1, U_2, \dots, U_N)$  con  $U_j = (u_{1j}, u_{2j}, \dots, u_{nj})$  el vector de respuestas del examinado  $j$  a los  $n$  ítemes. Se tiene una ecuación (2.24) para cada parámetro del ítem en el modelo que se utilice. Bajo el modelo Logístico de tres parámetros, uno debe especificar la forma de las distribuciones a priori de  $a_i, b_i$  y  $c_i$ , respectivamente.

Para la distribución a priori de la discriminación,  $a_i$  el programa PC-BILOG asume que cada  $a_i$  tiene una distribución a priori lognormal sobre el rango  $0 <$

$a_i < \infty$ . Esto porque en la mayoría de las aplicaciones de pruebas, las  $a_i$  son generalmente mayores de cero, suponiendo que las CCI son crecientes y refiriéndose a la opción correcta, sugiriendo que la distribución de las  $a_i$  puede ser modelada por una distribución unimodal y positivamente sesgada parecida a la lognormal (Baker y Kim, 2004).

Para la distribución a priori de la dificultad  $b_i$ , utiliza una distribución normal y para el parámetro de pseudo azar  $c_i$ , se usa una distribución Beta, ya que los valores de  $c_i$  están en el intervalo  $[0, 1]$ .

### 2.4.7. Estimación Bayesiana de los parámetros de los examinados

Los parámetros obtenidos para el ítem se consideran como los valores verdaderos en los procedimientos de estimación de los parámetros de habilidad. Se presentan dos procedimientos de estimación de los parámetros de habilidad. Estos procedimientos se deben a Bock y Aitkin (1981) y son: la *Estimación Modal de Bayes* y la *Estimación de Bayes EAP* (Esperado A Posteriori).

#### 2.4.7.1. Estimación modal de Bayes

Este procedimiento también llamado máximo a posteriori se basa en la forma del Teorema de Bayes:

$$g(\theta_j | U_j, \xi) \propto L(U_j | \theta_j, \xi) g(\theta). \quad (2.25)$$

En este procedimiento se asume que la distribución a priori  $g(\theta)$  es normal con hiperparámetros dados por  $\mu_\theta$  y  $\sigma_\theta^2$ . Tomando el logaritmo en cada lado de (2.25) se tiene

$$\log g(\theta_j | U_j, \xi) \propto \log L(U_j | \theta_j, \varepsilon) + \log g(\theta), \quad (2.26)$$

donde el componente de verosimilitud  $L(U_j | \theta_j, \varepsilon)$  está dado por

$$L(U_j | \theta_j, \varepsilon) = \prod_{i=1}^n P_i(\theta)^{u_{ij}} Q_i(\theta)^{1-u_{ij}}. \quad (2.27)$$

Aquí se asume que  $P_i(\theta)$  es el modelo Logístico de tres parámetros.

Se usa la ecuación puntuación Scoring de Fisher para obtener iterativamente la estimación modal de Bayes de la habilidad.

$$\left[ \hat{\theta}_j \right]_{(t+1)} = \left[ \hat{\theta}_j \right]_{(t)} - [\Lambda_{\theta\theta}]_{(t)}^{-1} \times [L_\theta]_{(t)}, \quad (2.28)$$

donde  $L_\theta$  y  $\Lambda_{\theta\theta}$  son dadas por

$$L_\theta = \sum_{i=1}^n a_i \left[ \frac{P_i(\theta_j) - c_i}{P_i(\theta_j)(1 - c_i)} \right] [u_{ij} - P_i(\theta_j)] - \left( \frac{\theta_j - \mu_\theta}{\sigma_j^2} \right)$$

$$\Lambda_{\theta\theta} = - \sum_{i=1}^n a_i^2 \left[ \frac{P_i(\theta_j) - c_i}{1 - c_i} \right]^2 \frac{Q_i(\theta_j)}{P_i(\theta_j)} - \frac{1}{\sigma_\theta^2}. \quad (2.29)$$

La ecuación (2.28) requiere comenzar con un valor inicial de  $\theta_j = \theta_j^{(0)}$ . Hay una ecuación (2.28) para cada examinado. Este procedimiento de estimación tiene la ventaja de que siempre converge sin importar los patrones de respuesta, de modo que aunque el examinado responda todos los ítemes correcta o incorrectamente, la estimación modal de Bayes siempre se puede obtener (Baker y Kim, 2004).

#### 2.4.7.2. Estimación de Bayes EAP

La estimación de Bayes es la media de la distribución posterior de  $\theta$ , dado el patrón de respuestas observadas  $U_j = [u_{1j}, u_{2j}, \dots, u_{nj}]'$ . Esto se puede aproximar por el método de la cuadratura Gaussiana

Para un individuo muestreado aleatoriamente de una población con distribución de habilidad  $g(\theta)$ ,  $P(U_j)$ , está dado por:

$$P(U_j) = \int_{\theta} P(U_j | \theta) g(\theta) d\theta$$

que es la probabilidad no condicionada. La esperanza de  $\theta_j$  dado  $U_j$  y  $\xi$  es

$$E(\theta_j | U_j, \xi) = \frac{\int \theta_j g(\theta) \prod_{i=1}^n P_i(\theta_j)^{u_{ij}} Q_i(\theta_j)^{1-u_{ij}} d\theta}{\int g(\theta) \prod_{i=1}^n P_i(\theta_j)^{u_{ij}} Q_i(\theta_j)^{1-u_{ij}} d\theta}. \quad (2.30)$$

Se usa el procedimiento de cuadratura Hermite-Gauss para aproximar la distribución normal  $g(\theta)$ , en el numerador y denominador de la ecuación (2.30).

#### 2.4.8. Estimación de parámetros de ítemes y de examinados vía muestreador de Gibbs

Este procedimiento es implementado por WINBUGS, el cual se utilizó en la aplicación. El Muestreador de Gibbs es un miembro de la clase de procedimientos conocidos como Cadenas Monte Carlo por Cadenas de Markov (MCMC).

Antes de examinar el Muestreador de Gibbs, se menciona la estructura básica de las Cadenas Monte Carlo por Cadenas de Markov.

Sea  $W$  un vector de variables aleatorias. El rango de dichas variables aleatorias se llama Espacio Estado. Supóngase que es factible generar una sucesión de  $W$ 's,

$$\{W^{(0)}, W^{(1)}, \dots, W^{(t)}, W^{(t+1)}, \dots\}, \quad (2.31)$$

tal que para cada tiempo  $t$ , la siguiente etapa  $W^{(t+1)}$  es muestreada de una distribución  $p(W^{(t+1)} | W^{(t)})$  que depende de la etapa actual de la cadena. Esta

sucesión se llama una cadena de Markov y  $p(W^{(t+1)} | W^{(t)})$  se llama el kernel transición de la Cadena Monte Carlo por Cadena de Markov y puede denotarse como  $k(W^{(t)}, W^{(t+1)})$ . Las cadenas de Markov se construyen a partir de un kernel transición  $k$ , una densidad de probabilidad condicional que

$$W^{(t+1)} \sim k(W^{(t)}, W^{(t+1)}). \quad (2.32)$$

Las cadenas encontradas en aplicaciones de Cadenas Monte Carlo por Cadenas de Markov tienen una fuerte propiedad de estabilidad. Por construcción existe una distribución de probabilidad estacionaria; esto es, una distribución  $\pi$  tal que si  $W^{(t)} \sim \pi$ , la  $W^{(t+1)} \sim \pi$ , siempre que el kernel  $k$  permita todo movimiento libre sobre el espacio estado.

Para la inferencia Bayesiana, la cadena de Markov se puede definir en una forma en que la distribución estacionaria  $\pi$  sea la distribución posterior. Por ejemplo,  $\pi(W)$  puede ser vista como  $p(\theta, \xi | Y)$ , donde  $\theta$  es el vector de parámetros de habilidad,  $\xi$  es el vector de parámetros del ítem y  $Y$  es la matriz de respuestas a los ítems.

El kernel transición del Muestreador de Gibbs en el contexto de la TRI, se puede definir como

$$k\left[\left(\theta^{(t)}, \xi^{(t)}\right), \left(\theta^{(t+1)}, \xi^{(t+1)}\right)\right] = p\left(\theta^{(t+1)} | \xi^{(t)}, Y\right) p\left(\xi^{(t+1)} | \theta^{(t+1)}, Y\right) \quad (2.33)$$

(Baker y Kim, 2004). Ya que

$$p(\theta | \xi, Y) = \frac{p(Y | \theta, \xi) p(\theta, \xi)}{\int p(Y | \theta, \xi) p(\theta, \xi) d\theta} \quad (2.34)$$

y

$$p(\xi | \theta, Y) = \frac{p(Y | \theta, \xi) p(\theta, \xi)}{\int p(Y | \theta, \xi) p(\theta, \xi) d\xi}. \quad (2.35)$$

Enseguida se presenta el método Muestreador Gibbs de Albert (1992) aplicado a la estimación de parámetros en el modelo Ojiva Normal de 2 parámetros.

#### 2.4.8.1. El muestreador Gibbs de Albert

Supóngase que los datos se modelan usando el modelo Ojiva Normal de dos parámetros. Dentro del método de Cadenas Monte Carlo por Cadenas de Markov, la alternativa de Gibbs para la estimación conjunta de los  $2n$  parámetros del ítem de una prueba y los  $N$  parámetros de habilidad involucra una cadena de Markov. En cada etapa o ciclo  $t$  de la cadena de Markov, el muestreo de parámetros de la distribución posterior conjunta de los  $2n + N$  parámetros es simulado por muestreo de la distribución posterior condicional de cada parámetro. De este modo, en lugar de muestrear una distribución posterior conjunta multidimensional, se muestrea cada una de las  $2n + N$  distribuciones unidimensionales.

Entonces, el par de parámetros de cada ítem es muestreado de su distribución posterior, condicional sobre los valores de todos los demás parámetros de ítemes y de examinados. Los resultantes  $2n + N$  valores se guardan como si fueran una muestra de su distribución conjunta posterior. Los valores de parámetros contenidos en la muestra o extraídos en la etapa  $t$  sirven como valores condicionales en la siguiente etapa  $t + 1$  de la cadena de Markov. Después de obtener una serie secuencial de estas etapas, los parámetros muestreados deben aproximarse a aquellos obtenidos de la distribución posterior conjunta subyacente y la cadena de Markov habrá alcanzado entonces una etapa estable o habrá llegado a la convergencia. Una vez que se alcanza la convergencia, se continúa el proceso de muestreo sobre un número grande de tales etapas para determinar empíricamente la distribución posterior conjunta de los  $2n + N$  parámetros. Se usa una medida de tendencia central como es la media de la distribución posterior marginal de cada parámetro, como una estimación de ese parámetro.

Enseguida se describe cómo se implementa el método de Albert. Aparte de la restricción de que las discriminaciones del ítem deben ser mayores que cero, no se emplean aprioris para los parámetros del ítem.

El probit está dado por

$$Z_{ij} = \alpha_i \theta_j - \gamma_i, \quad (2.36)$$

donde  $\alpha_i$  es el parámetro de discriminación del ítem,  $i = 1, 2, \dots, n$ ,  $\gamma_i$  es el parámetro de dificultad del ítem y  $\theta_j$  es el nivel del rasgo para el examinado  $j$ , con  $j = 1, \dots, N$ . Esta es una parametrización atípica de la dificultad del ítem. La parametrización usual en la TRI es

$$Z_{ij} = \alpha_i (\theta_j - \beta_i) = \alpha_i \theta_j + \zeta_i, \quad (2.37)$$

donde  $\zeta_i$  es la intercepción y  $\beta_i = -\zeta_i/\alpha_i$  es la dificultad del ítem. Entonces, el parámetro de dificultad del ítem de Albert es

$$\gamma_i = \alpha_i \beta_i = \alpha_i (-\zeta_i/\alpha_i) = -\zeta_i.$$

La dificultad del ítem es el negativo de la intercepción usual y

$$\beta_i = \frac{\gamma_i}{\alpha_i}$$

es la dificultad del ítem usual de la TRI.

Las calificaciones dicotómicas de los ítemes se almacenan en una matriz  $Y$  de dimensión  $N \times n$ . Los parámetros de un ítem son dados por  $\xi_i = (\alpha_i, \gamma_i)$  y la matriz de todos los parámetros del ítem es dada por  $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ . El vector de parámetros de habilidad es  $\theta = (\theta_1, \theta_2, \dots, \theta_N)$ . Sea  $Z$  el vector de variables aleatorias independientes  $Z = (Z_{11}, Z_{12}, \dots, Z_{ij}, \dots, Z_{nN})$ .

La implementación del componente Monte Carlo en este Muestreador de Gibbs involucra tres procesos de muestreo:

- (1) un muestreo de probits que dependen de las respuestas del examinado
- (2) un muestreo de parámetros de habilidad, y
- (3) un muestreo de  $2n$  parámetros del ítem.

Los valores iniciales son  $\alpha_i = 2$ ,  $\gamma_i = \sqrt{5}$  veces la desviación normal unitaria co-rrespondiente a la dificultad del ítem de la teoría clásica de la prueba y  $\theta_j = 0$  (Albert, 1992). El muestreo de probits se describe a continuación. Usando los valores actuales de todos los parámetros, se establecen las distribuciones condicionales de las  $Z_{ij}$  con medias  $\eta_{ij} = \alpha_i\theta_j - \gamma_i$  y varianza uno y cada una se divide en dos partes. El área por arriba de  $Z_{ij} = 0$  corresponde a la probabilidad de una respuesta correcta  $P_{ij}$  y el área por debajo a la probabilidad de una respuesta incorrecta  $Q_{ij}$ . Cuando un examinado responde correctamente al ítem,  $Y_{ij} = 1$  y  $Z_{ij}$  es muestreada aleatoriamente de la parte de la distribución normal por arriba de 0; una respuesta incorrecta resulta en un valor  $Z_{ij}$  que es muestreado aleatoriamente por debajo de 0. Este proceso se realiza para todo elemento del vector  $Z$ . El proceso resulta en una matriz de probits que dependen del valor actual de  $\xi, \theta$ , y la matriz de respuestas  $Y$ . Se describe ahora cómo se realiza el muestreo de la distribución posterior condicional de la habilidad.

Se asume que las habilidades  $\theta_j$  tienen distribución independiente. La función de verosimilitud para  $\theta_j$  es la forma normal con media

$$\tilde{\theta}_j = \frac{\sum_{i=1}^n \alpha_i (Z_{ij} + \gamma_i)}{\sum_{i=1}^n \alpha_i^2} \quad (2.38)$$

con varianza

$$v = \frac{1}{\sum_{i=1}^n \alpha_i^2}. \quad (2.39)$$

Esto se combina con una distribución normal a priori para producir una distribución posterior normalmente distribuida de  $\theta_j$  con media

$$\bar{\theta}_j = \frac{\tilde{\theta}_j/v + \mu/\sigma^2}{1/v + 1/\sigma^2} \quad (2.40)$$

y varianza

$$v_p = \frac{1}{1/v + 1/\sigma^2}, \quad (2.41)$$

donde los parámetros de la normal a priori están fijos en  $\mu = 0$  y  $\sigma^2 = 1$ . Los cálculos prosiguen computando  $\tilde{\theta}$  para un examinado y luego obteniendo  $\bar{\theta}$  y  $v_p$ . Luego se genera una desviación normal aleatoria y un valor muestreado  $\theta_j$  está dado por

$$\theta_j^{(t)} = RN\sqrt{v_p} + \bar{\theta}_j, \quad (2.42)$$

donde  $t$  designa el ciclo de muestreo y  $RN$  es una desviación normal aleatoria. Los pasos de arriba se realizan para todos los  $N$  examinados, produciendo un nuevo vector de parámetros de habilidad.

El último proceso es el muestreo de las distribuciones posteriores condicionales de parámetros del ítem. Se muestrean los parámetros del ítem en el vector  $\xi$ , condicional sobre los valores actuales en  $Z$  y en  $\theta$ . Este proceso consta de 4 pasos:

1. La matriz  $X$  se define como

$$X = \begin{bmatrix} \hat{\theta}_1 & -1 \\ \hat{\theta}_2 & -1 \\ \hat{\theta}_3 & -1 \\ \vdots & \vdots \\ \hat{\theta}_N & -1 \end{bmatrix} \quad (2.43)$$

y la matriz Hessiana está dada por

$$X'X = \begin{bmatrix} \sum_{i=1}^N \hat{\theta}_i^2 & -\sum_{i=1}^N \hat{\theta}_i \\ -\sum_{i=1}^N \hat{\theta}_i & N \end{bmatrix}. \quad (2.44)$$

2. Se realiza una solución mínimos cuadrados ordinarios para la estimación simultánea de los parámetros en  $\xi$ , usando

$$\tilde{\xi} = (X'X)^{-1} (X'Z),$$

produciendo una matriz  $2 \times n$  de valores de parámetros del ítem,

$$\tilde{\xi} = \begin{bmatrix} \tilde{\alpha}_1 & \tilde{\alpha}_2 & \cdots & \tilde{\alpha}_n \\ \tilde{\gamma}_1 & \tilde{\gamma}_2 & \cdots & \tilde{\gamma}_n \end{bmatrix}.$$

3. Se realiza una factorización de Cholesky,

$$A = chol(X'X)^{-1},$$

donde  $A$  es una matriz triangular superior,  $A'$  es su transpuesta y  $(X'X)^{-1}$  es la varianza posterior de  $\xi_i$  para cada uno de los  $n$  ítems. La matriz  $A$  es la misma para todos los ítems de la prueba.

4. Se generan dos desviaciones normales aleatorias  $RN_1$  y  $RN_2$ . Los valores de los parámetros del ítem para un ítem del ciclo de muestreo  $t$ , está dada por

$$\begin{bmatrix} \alpha_i^{(t)} \\ \gamma_i^{(t)} \end{bmatrix} = A' \begin{bmatrix} RN_1 \\ RN_2 \end{bmatrix} + \begin{bmatrix} \tilde{\alpha}_i \\ \tilde{\gamma}_i \end{bmatrix}. \quad (2.45)$$

Dado que se impone la restricción  $\alpha_i > 0$  sobre los  $\alpha_i^{(t)}$ , el paso 4 se repite hasta que dicha restricción se cumpla. El paso 4 se realiza para todos los  $n$  ítems.

Al término de estos cuatro pasos, se ha producido un nuevo conjunto de valores de parámetros en  $\theta$  y en  $\xi$ . Estos valores se almacenan como una extracción de la distribución posterior conjunta de los parámetros. El proceso descrito se repite en cada ciclo de la cadena de Markov hasta obtener el número de muestras necesarias para formar la distribución posterior marginal empírica de cada parámetro. En la alternativa MCMC se necesita para la cadena de Markov alcanzar la estabilidad antes de retener las muestras. Cuando se alcanza la estabilidad, se dice que la cadena de Markov alcanzó la convergencia.

Un problema con el proceso de la cadena de Markov es que las extracciones sucesivas no son independientes pues los valores del parámetro en las etapas presentan autocorrelación. Una solución a esto es retener un subconjunto de muestras que estén espaciadas, por ejemplo cada décima muestra. Al finalizar este proceso, se tiene una distribución posterior marginal separada para los parámetros de cada ítem y para los parámetros de habilidad de todos los examinados. Se calcula la media de cada distribución y se reporta como la estimación Muestreador de Gibbs del parámetro dado.

Recapitulando, el método Muestreador de Gibbs usa los siguientes cuatro pasos básicos:

1. Se especifican distribuciones condicionales totales y métodos de muestreo para los parámetros inobservables
2. Se proporcionan valores iniciales
3. Se moniteran las salidas
4. Se calculan resúmenes estadísticos para las cantidades de interés.

Enseguida se describe con detalle cada uno de los 4 pasos anteriores.

En el paso 1 se especifica el modelo que sirve como entrada para BUGS. En un conjunto de datos de respuesta al ítem, las respuestas  $Y_{ij}$  son independientes condicional sobre sus parámetros en  $P_{ij}$ . Para el ítem  $i$  y el examinado  $j$ , cada  $P_{ij}$  es una función del parámetro de habilidad  $\theta_j$ , del parámetro de dificultad  $b_i$  y del parámetro de discriminación  $a_i$ , bajo un modelo de dos parámetros. Se asume con el propósito de determinar la métrica de la escala, que los  $\theta_j$  se extraen independientemente y provienen de una distribución normal estándar. La relación entre los parámetros del ítem y del examinado con la probabilidad de acierto, determinada por el modelo que se use para la CCI, y las respuestas del examinado aparecen en la Figura 2.

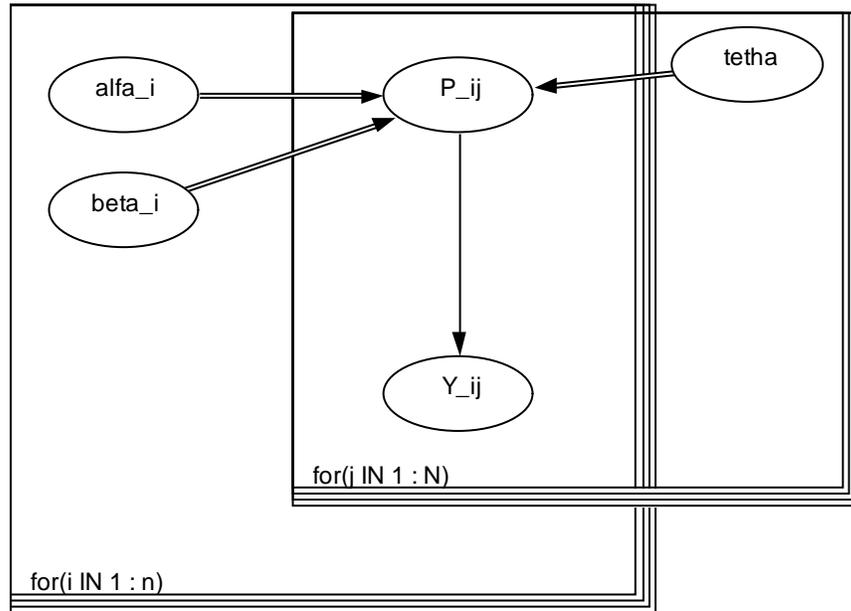


Figura 2. Esquema del modelo TRI de dos parámetros en WINBUGS.

En la Figura 2, cada variable o cantidad en el modelo aparece como un nodo en la gráfica. Las flechas dobles denotan una relación funcional, de modo que  $P_{ij}$  es una función de  $\alpha_i$  (el parámetro de discriminación),  $\beta_i$  (el parámetro de dificultad) y  $\theta_j$  (el parámetro de habilidad). Las flechas simples significan que los nodos están unidos por una dependencia probabilística ó función liga, en este caso las respuestas  $Y_{ij}$  se distribuyen Bernoulli con parámetro  $P_{ij}$ .

El método Muestreador de Gibbs implementado en BUGS trabaja extrayendo iterativamente muestras de la distribución condicional completa de los parámetros del modelo usando el algoritmo de muestreo de rechazo adaptativo, dado que las distribuciones condicionales totales para los modelos Ojiva Normal y Logístico son log-cóncavas (Baker y Kim, 2004). Para muestrear  $\theta$  a partir de  $p(\theta | \xi, Y) \propto p(Y | \theta, \xi) p(\theta, \xi)$  el muestreo de rechazo requiere una función envolvente  $P(\theta)$ . La función envolvente es una distribución propuesta conveniente que puede definirse fácilmente. Las muestras, digamos  $\theta^*$ , se extraen a partir de la densidad proporcional a  $P(\theta)$  y cada  $\theta^*$  se acepta con probabilidad  $p(\theta^* | \xi, Y) / P(\theta^*)$ . Es decir que se genera una variable aleatoria uniforme y  $\theta^*$  se aceptará sólo si  $u \leq p(\theta^* | \xi, Y) / P(\theta^*)$  o rechazada en caso contrario. De manera similar se hace con  $\xi$ . En BUGS las distribuciones propuestas se seleccionan automáticamente cuando las densidades objetivo son log-cóncavas.

Para analizar los datos, se debe especificar la forma en que se relacionan los nodos. Bajo el modelo Logístico de dos parámetros, la probabilidad de que el

examinado  $j$  responda correctamente al ítem  $i$  es

$$P_{ij} = \frac{1}{1 + \exp[-a_i D(\theta_j - b_i)]}$$

esto también se puede expresar como

$$\log it(P_{ij}) = a_i(\theta_j - b_i)$$

ya que las  $Y_{ij}$  tienen distribución Bernoulli con parámetro  $(P_{ij})$ .

En el caso del modelo Ojiva Normal de dos parámetros,  $P_{ij}$  se define como

$$P_{ij} = \Phi[a_i(\theta_j - b_i)]$$

donde  $\Phi$  es la distribución acumulada normal estándar. Para completar la especificación del modelo deben definirse distribuciones a priori para los parámetros. En los dos modelos que se han mencionado en esta subsección, generalmente se usan a priori normales y en el caso de  $\theta$ , la distribución normal estándar.

En el paso 2 se especifican los valores iniciales  $\omega^{(0)} = (\theta^{(0)}, \xi^{(0)})$  para arrancar el algoritmo. Raftery indicó que los valores iniciales extremos podrían llevar a un proceso muy largo de estabilización. Por otro lado, puede ser que los resultados finales sean sensibles a la elección de valores iniciales por lo que es aconsejable hacer varias corridas usando diferentes valores iniciales para detectar si hay sensibilidad (Baker y Kim, 2004).

En el paso 3 se realiza el monitoreo de las salidas del programa para verificar si el proceso de la cadena de Markov ha llegado a la convergencia. Esto es necesario saberlo para detener el muestreo y usar los resultados para estimar los parámetros de interés. Es por eso que en el monitoreo de las salidas se pueden observar gráficas y resúmenes estadísticos de las cantidades de interés. Gelman y Rubin propusieron un diagnóstico para la convergencia de los algoritmos MCMC. Para cada parámetro, el estadístico de Gelman y Rubin estima la reducción en la estimación acumulada de la varianza si las corridas continuaran indefinidamente. El valor de este estadístico debe ser cercano a 1 para estar razonablemente seguros de que la convergencia se ha dado.

Finalmente, en el paso 4 se obtienen resúmenes estadísticos para los parámetros de interés. Se puede obtener la media posterior, así como también el intervalo de probabilidad posterior y la desviación estándar posterior del Muestreador de Gibbs para cada parámetro.

## CAPÍTULO 3

### APLICACIÓN Y RESULTADOS

#### *Objetivos de la Tesis*

El objetivo del trabajo de tesis, es el siguiente:

Extraer información acerca de los examinados y de los ítems del examen EXHCOBA mediante el ajuste de un modelo probabilístico de la TRI.

#### 3.1. Descripción de la prueba

Los datos utilizados en esta prueba son producto de la aplicación del examen de admisión a las licenciaturas de la Universidad Autónoma de Querétaro en el mes de junio de 2006, denominado Examen de Conocimientos y Habilidades Básicas, EXHCOBA por sus siglas, el cual es creado por la Universidad de Baja California. El examen consta de 190 preguntas, la distribución de las cuales se muestra en la Tabla 1.

Tabla 1. Distribución de las preguntas en las áreas de conocimiento, del examen EXHCOBA.

| Nivel        | Sección                                 | Asignatura                      | No. de preg |
|--------------|---|---------------------------------|-------------|
| Primaria     | Habilidades Básicas                     | Hab. Verbales                   | 30          |
|              |   | Hab. Cuantitativas              | 30          |
| Secundaria   | Conocimientos Básicos                   | Español                         | 15          |
|              |   | Matemáticas                     | 15          |
|              |   | Cs. Naturales                   | 20          |
|              |   | Cs. Sociales                    | 20          |
|              |   | Matemáticas para el Cálculo     | 20          |
| Bachillerato | Conocimientos Básicos para especialidad | Matemáticas para la Estadística | 20          |
|              |   | Física                          | 20          |
|              |   | Química                         | 20          |
|              |   | Biología                        | 20          |
|              |   | Cs. Sociales                    | 20          |
|              |   | Humanidades                     | 20          |
|              |   | Lenguaje                        | 20          |
|              |   | Cs. Económico-admivas.          | 20          |

Los primeros 130 reactivos (primaria y secundaria) son contestados por todos los examinados. En el caso del nivel bachillerato, cada carrera tiene asignadas tres asignaturas, que completan los 190 reactivos que contesta el examinado.

### 3.2. Descripción de los datos

Del conjunto de datos completo, se utilizó solamente el subconjunto correspondiente a las licenciaturas del área de ingeniería. En el caso de las carreras de ingeniería, las asignaturas de nivel Bachillerato en las cuales son evaluados son: Matemáticas para el Cálculo, Física y Química. El orden de las preguntas es el que presenta la Tabla 1 por lo cual las preguntas correspondientes a Matemáticas para el Cálculo son de la 131 a la 150. Estas fueron las columnas utilizadas como datos en la aplicación. Es decir se utilizaron las columnas de la 131 a la 150 de los aspirantes a ingresar a las licenciaturas de Ingeniería. En realidad, no se utilizaron los registros de todos los aspirantes (que son alrededor de 1000), sino una muestra de 109 de éstos. En el análisis de conjuntos de datos de la TRI generalmente no se toman muestras de datos de examinados, sino que se utilizan todos los registros. En el caso de la aplicación presente se utilizó una muestra de los registros por una razón: la TRI requiere que a los examinados se les hayan presentado los mismos ítems y en el examen EXHCOBA no a todos los aspirantes se les presentaron los mismos ítems. Sin embargo, los aspirantes cuyo número de ficha terminó en 2, sí contestaron los mismos ítems, de modo que se seleccionaron los registros de aspirantes con esta característica, lo cual equivale en la práctica, a tomar una muestra. Los datos utilizados forman una matriz de 109 renglones por 20 colum-

nas y son binarios. Un dato 0 significa que la respuesta al ítem fue incorrecta y un 1 significa que fue correcta.

Los 109 registros utilizados aparecen en el Apéndice B. Cada renglón de la matriz corresponde a las respuestas de un examinado.

### 3.3. Estimación de los parámetros

Se analizaron los datos del examen EXHCOBA con el programa TESTGRAF de Jim Ramsay (no paramétrico) y con el programa WINBUGS (paramétrico Bayesiano). Posteriormente se compararon y complementaron los resultados de ambos programas. Se utilizó el programa TESTGRAF para estimar de modo no paramétrico las CCI's (ecuaciones 2.6 y 2.7). Este proceso no atraviesa por la estimación de parámetros del ítem pues no lo requiere. En contraparte, se estimaron los parámetros de las CCI's usando el paquete estadístico WINBUGS, el cual utiliza un enfoque paramétrico de Estadística Bayesiana. Para esto, se usaron los modelos: ojiva normal de 2 parámetros (ecuación 2.2) y ojiva normal de 3 parámetros (ecuación 2.1). Se consideró conveniente el modelo ojiva normal porque las correlaciones tetracóricas son inapropiadas para distribuciones no normales de la habilidad; también son inapropiadas cuando la función de respuesta al ítem no es una ojiva normal (Lord, 1980). El programa TESTGRAF construye las CCI's usando un procedimiento de suavizamiento por kernel, descrito en la sección 2.3.2.

El modelo ojiva normal de 2 parámetros es la ecuación (2.2), donde  $\theta$  es la habilidad en Matemáticas para el Cálculo en esta aplicación,  $a_i$  es el parámetro de discriminación del ítem  $i$ ,  $b_i$  es el parámetro de dificultad del ítem y  $P_i(\theta)$  es la probabilidad de que un examinado con nivel de habilidad  $\theta$  responda correctamente el ítem  $i$ . El modelo ojiva normal de tres parámetros aparece en la ecuación (2.1), en la cual  $c_i$  es el parámetro de pseudo azar del ítem  $i$ .

WINBUGS estima los parámetros del ítem y de habilidad usando el algoritmo Muestreador de Gibbs. Las distribuciones a priori utilizadas en esta aplicación fueron: para el parámetro de habilidad,  $\theta$ , la distribución Normal estándar, para el parámetro de discriminación,  $a$ , la distribución normal truncada por la izquierda para admitir sólo valores positivos, con media  $\mu = 0,5$  y varianza  $\sigma = 1$ , para el parámetro de dificultad del ítem,  $b$ , se utilizó también la distribución Normal estándar. Se usó como valor inicial de  $\hat{a}$  y  $\hat{b}$ , el vector constante  $(0,5, 0,5, \dots, 0,5)$  de dimensión  $20 \times 1$ . Estas especificaciones pueden verse en el código utilizado en WINBUGS, el cual se encuentra en el Apéndice A. Las distribuciones a priori utilizadas fueron elegidas de acuerdo a lo utilizado por el programa PC-BILOG, comentado en Baker y Kim, (2004). Por ejemplo, suponiendo que se habla del parámetro de discriminación,  $a$ , de la CCI correspondiente a la opción correcta, sus valores posibles son no negativos por lo que se eligió una distribución a priori

normal truncada. En el caso del parámetro de dificultad,  $b$ , dicho parámetro puede asumir valores de  $-\infty$  a  $+\infty$  al igual que la habilidad  $\theta$ , por lo cual se eligió la densidad Normal estándar como a priori para  $b$ . Para el parámetro de pseudo azar, el cual puede tomar valores acotados por 0 y 1 se utilizó la distribución a priori Beta con parámetros Beta(5,17). Una distribución Beta con estos parámetros tiene una media de 0.2 que corresponde a la probabilidad de adivinar la respuesta correcta por azar cuando los ítems tienen 5 opciones, ya que  $1/5 = 0.2$ . Los ítems del examen EXHCOBA tienen 5 opciones.

Las estimaciones de los parámetros del ítem y de examinados, usando el modelo ojiva normal de 2 parámetros con WINBUGS aparecen en el apéndice D y las 20 curvas características de los ítems construidas con las estimaciones de los parámetros de los ítems aparecen en el Apéndice E. Las CCI's correspondientes al modelo ojiva normal de 3 parámetros no aparecen debido a que como se verá enseguida, este modelo no ofreció un mejor ajuste que el modelo de dos parámetros, sin embargo el código que se utilizó para generar las estimaciones aparece en el Apéndice A.

### 3.4. Evaluación de los supuestos de la TRI

En este estudio utilizamos el Análisis de Factores para determinar si existe un sólo factor subyacente dominante en los datos. Se usó el software TESTFACT versión 4.0 para Windows (Wood R., Wilson D., Gibbons R., Schilling S., Muraki E. y Bock D.).

Con el propósito de evaluar el supuesto de unidimensionalidad, con el programa TESTFACT se obtuvo la matriz de correlaciones tetracóricas estimadas y los valores de las raíces latentes de dicha matriz. Se realizó una gráfica de las 10 raíces latentes más grandes. Véase la Figura 3.

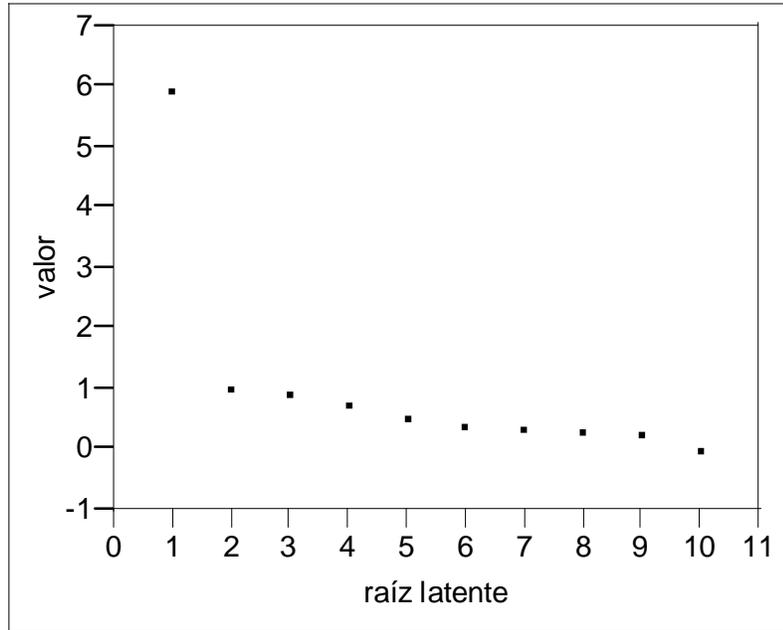


Figura 3. Gráfica de las 10 raíces latentes más grandes de la matriz de correlaciones tetracóricas.

En la Figura 3 se aprecia que el primer factor domina al resto y la segunda raíz es apenas superior a 1, lo cual apoyaría que el supuesto de unidimensionalidad se cumple, según Lord (1980). Sin embargo, dado que este procedimiento es burdo, no es suficiente evidencia por lo que se aplicaron otros 2 criterios para verificar la unidimensionalidad. Con TESTFACT se probó la hipótesis nula  $H_0$  : de que 1 factor proporciona una descripción adecuada de los datos versus la hipótesis alternativa  $H_1$  : de que 2 factores describen adecuadamente los datos.

El estadístico ji-cuadrado

$$\chi^2 = \sum_{j=1}^{N_R} W_j \log \left( \frac{W_j}{W_T \times p_j} \right)$$

con grados de libertad  $ndf = (N_R - 1) - [(nfac + 1)n - nfac(nfac - 1) / 2]$

se calcula bajo  $H_0$  y bajo  $H_1$ , lo mismo se hace con los grados de libertad, luego se toma la diferencia. En este caso la diferencia fue 34.87 con 19 grados de libertad y resultó significativo al nivel  $.02 < valor - P < .01$ . Esto significa que bajo esta prueba, un modelo de 2 factores es más apropiado que el de un solo factor, lo que contradice que los ítemes sean unidimensionales. Debido a que la conclusión del análisis de unidimensionalidad no es clara, se realizó una prueba adicional. Se simuló respuestas dicotómicas a ítemes con las características de los ítemes analizados (se proporcionan parámetros de dificultad y cargas de factor) suponiendo que sólo existe un factor subyacente (un solo rasgo latente). Los

parámetros de dificultad y cargas de factor se obtuvieron de correr el TESTFACT con un modelo ojiva normal de 3 y 2 parámetros. De estas respuestas simuladas se obtuvo la matriz de correlaciones tetracóricas y se obtuvo la matriz de diferencias entre las correlaciones tetracóricas observadas y las predichas suponiendo el modelo de un factor. Se analizaron los residuos mediante un histograma, ajustando un modelo ojiva normal de tres parámetros (Figura 4) y un modelo ojiva normal de dos parámetros (Figura 5). De estos histogramas se espera que la distribución sea unimodal, simétrica, con media 0 y desviación estándar pequeña.

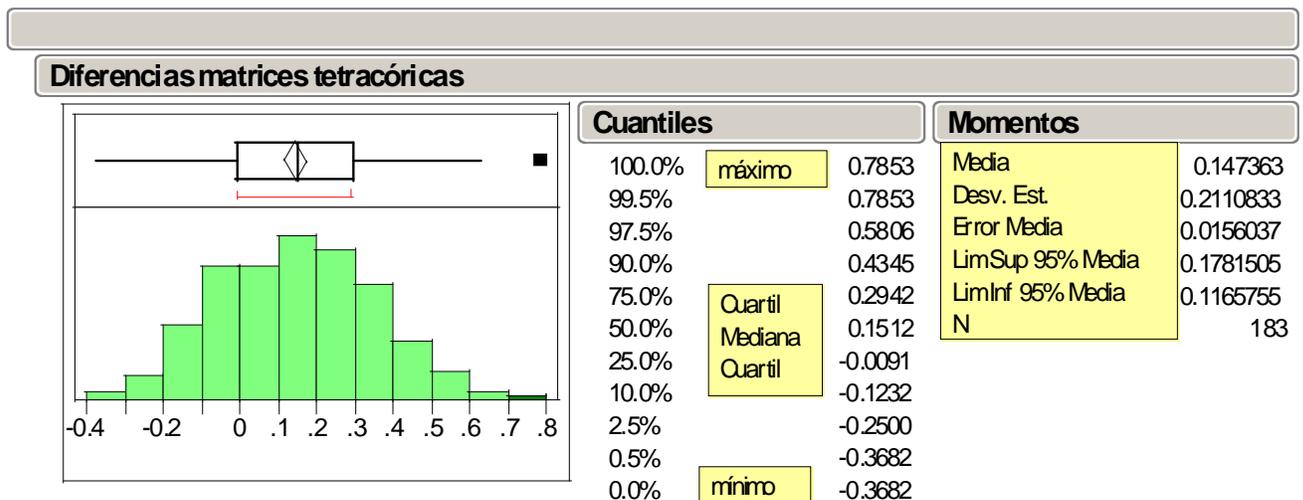


Figura 4. Diferencias de la matriz de correlaciones tetracóricas observada menos la predicha bajo el modelo de 1 factor.

La matriz predicha, en el caso de la Figura 4, se obtiene previo ajuste de un modelo ojiva normal de 3 parámetros.

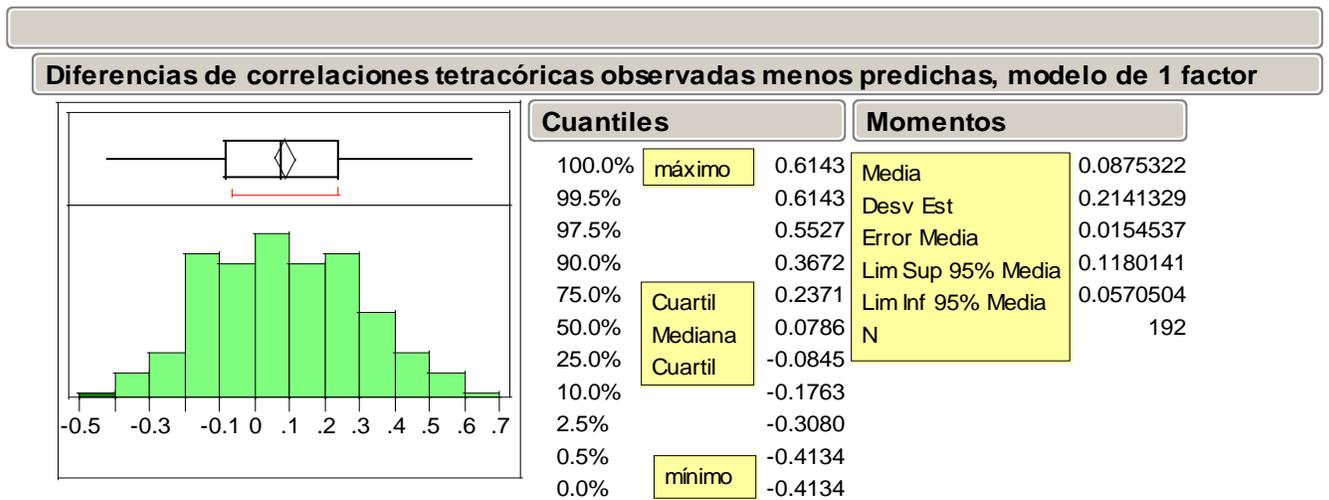


Figura 5. Diferencias de la matriz de correlaciones tetracóricas observada menos la predicha bajo el modelo de 1 factor.

La matriz predicha, en el caso de la Figura 5, se obtiene previo ajuste de un modelo ojiva normal de 2 parámetros. En las Figuras 4 y 5 se observa que los valores de las diferencias son ligeramente sesgadas a la derecha en ambos casos pero en la Figura 5 tienen menor dispersión, el rango es menor pero también las barras del centro son más altas que en la Figura 4, lo cual indica que el modelo ojiva normal de 2 parámetros contradice con menor fuerza el supuesto de unidimensionalidad. Fue esto lo que hizo que se realizaran las otras dos pruebas, la gráfica de raíces latentes y la prueba de hipótesis, con el modelo ojiva normal de 2 parámetros.

A fin de comprobar el supuesto de monotonidad de las curvas características de los ítems, se obtuvieron las CCI's con el TESTGRAF, el cual usa un método no paramétrico ya descrito anteriormente. Las 20 gráficas de las CCI's aparecen en el Apéndice C. En ellas notamos que seis no son monótonas crecientes, las que corresponden a los ítems 3, 6, 12, 14, 16 y 17. Sin embargo es en las CCI's correspondientes a los ítems 3, 12, 14 y 17 donde el comportamiento decreciente es más pronunciado.

## CAPÍTULO 4

### CONCLUSIONES

En las CCI's obtenidas con TESTGRAF encontramos que seis de ellas no son monótonas crecientes: la 3, 6, 12, 14, 16 y 17, aunque en las CCI's correspondientes al ítem 6, 16 y 17 el comportamiento decreciente no es muy pronunciado, mientras en las CCI's 3, 12, 14 y 17 sí lo es. El uso de un modelo paramétrico sin verificar el supuesto de monotonicidad de las CCI's hace que el investigador ignore completamente la existencia de comportamientos decrecientes de las CCI's en algunos intervalos de habilidad. La alternativa no paramétrica permite explorar este aspecto.

En este estudio utilizamos el Análisis de Factores para determinar si existe un sólo factor subyacente dominante en los datos. Se usó el software TESTFACT versión 4.0 para Windows (Wood R., Wilson D., Gibbons R., Schilling S., Muraki E y Bock D.)

Como ya se mencionó antes, la TRI supone que bajo los ítems utilizados subyace uno y sólo un rasgo latente. Esto es lo que se trata de verificar.

Los resultados obtenidos con este software fueron en primer término las CCI obtenidas por métodos no paramétricos y estas curvas sirvieron para juzgar la calidad de los ítems del examen y también para compararlas con las CCI obtenidas por métodos paramétricos para determinar la existencia de problemas en el cumplimiento de la monotonicidad. Ciertamente es que los modelos logísticos (que en este caso son los paramétricos) describen curvas características monótonas, no se debe olvidar que estas curvas deben ajustar adecuadamente las proporciones de respuesta correcta calculadas en cada nivel de habilidad, pues este es el sentido intuitivo de las CCI. La cuestión es entonces verificar si este ajuste es adecuado en cada ítem, pues además la TRI supone que este comportamiento monótono de las CCI lo tienen todas las CCI del examen. En el caso de este estudio se encontró que cuatro de los veinte ítems del examen no cumplen el supuesto de monotonicidad, lo que indica que estos ítems deben ser modificados en su redacción o tal vez hasta excluidos. Hay un elemento de incertidumbre en esta conclusión pues el tamaño de la muestra es pequeño, 109 registros, y queda la duda de si al analizar los datos de una muestra grande, los resultados se mantienen o cambian. Con los resultados obtenidos de la muestra utilizada, aun cuando sea pequeña, el investigador debe desconfiar de las conclusiones que surjan de aplicar los modelos debido a esta falta de ajuste.

Otro de los resultados obtenidos fue el cálculo del nivel de habilidad de cada examinado junto con su error estándar, lo que permite ordenar a los examinados en términos de este nivel de habilidad y el error estándar permite saber cuán confiable es este valor. También derivado de este punto se obtuvo la distribución de los niveles de habilidad (ver Figura 6) permitiendo esto juzgar cómo fue el desempeño del grupo, en su conjunto, en el tema Matemáticas para el Cálculo. Dicha distribución es ligeramente sesgada a la derecha, es decir que los niveles de habilidad inferiores a la media son un poco más frecuentes que los superiores a la media. Esto indica un buen desempeño del grupo, lo cual es de esperarse pues se trata de aspirantes a cursar alguna licenciatura en ingeniería.

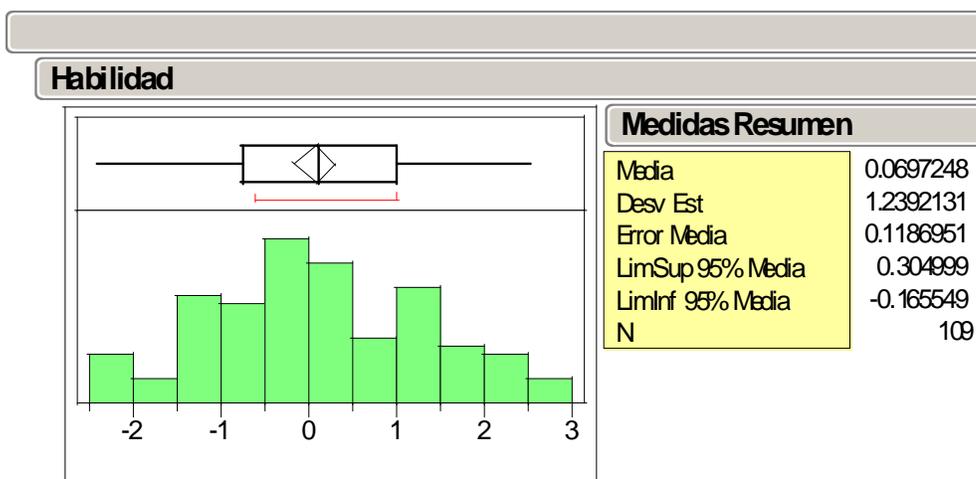


Figura 6. Histograma de los valores estimados de habilidad

Con el nivel de habilidad del examinado y su error estándar se construye una gráfica, dada por TESTGRAF, la cual permite interpretar más fácilmente la información. Enseguida se muestra una gráfica de éstas para ejemplificar su interpretación.

Otro de los elementos en la salida del programa es la gráfica de la Función de Información de la Prueba, esta gráfica sirvió para ver en qué niveles de habilidad mide mejor la prueba y observamos (ver Figura 7) que esto ocurre en los niveles de habilidad ligeramente por debajo de la media. También se determinaron gráficamente los niveles de dificultad de los 20 ítems y se observó que las dificultades se distribuyen entre -2 y +2 uniformemente sobre la escala de habilidad. Esto es una cualidad deseable de medición de los ítems.

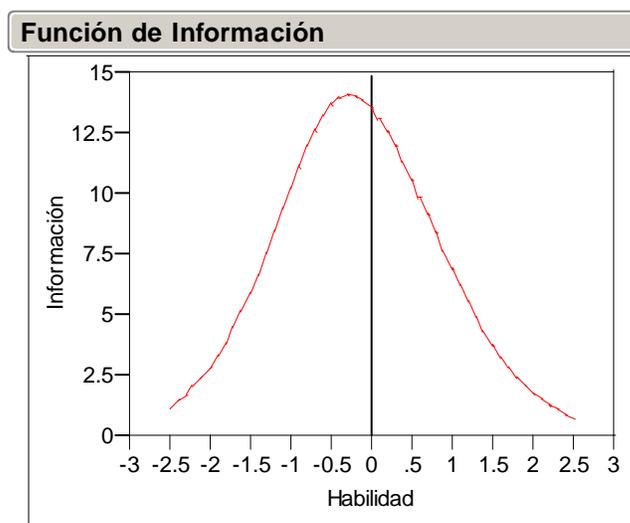


Figura 7. Gráfica de la Función de Información de la Prueba.

Respecto al supuesto de unidimensionalidad de los ítemes utilizados en el estudio se concluye que no se cumple satisfactoriamente este supuesto, lo cual se verificó usando Análisis de Factor para Ítemes binarios en el programa TEST-FACT.

Como conclusión se observa que la Teoría de Respuesta al Ítem ha sido muy útil en el análisis sobre las propiedades de medición de la prueba, lo mismo que en las propiedades de medición de los ítemes. En cuanto a la estimación de la habilidad, dado que la unidimensionalidad no es clara, debemos desconfiar de dichas estimaciones pues no se sabe a qué rasgo latente corresponden.

En el contexto de un examen de admisión personalmente pienso que debe ser de interés medir con mayor precisión las habilidades altas pues finalmente interesa admitir a los examinados con mayor habilidad. Sin embargo en este instrumento, se miden con mayor precisión las habilidades medias que las altas.

El supuesto de unidimensionalidad resulta muy restrictivo pues si los ítemes no lo satisfacen, no se pueden inferir las habilidades de los examinados. En el caso de este estudio es probable que la razón por la cual no se cumpla el supuesto sea la existencia de otro rasgo estrechamente vinculado a la habilidad en Matemáticas para el Cálculo que esté interviniendo, como podría ser una habilidad para plantear el problema.

Sin embargo es posible mejorar el examen de admisión ya que al estimar la habilidad del examinado se puede estimar con la misma precisión para cada examinado; es decir, se tendría un vector de niveles de habilidad estimados donde cualesquiera dos elementos del vector son igualmente confiables y esto facilitaría la comparación entre el desempeño de cualquier par de examinados. Esto se puede hacer implementando las Pruebas Adaptativas Informatizadas (Bock, 1997), las

cuales poseen precisamente la característica de estimar una habilidad con una precisión fija de antemano. En estas pruebas se le administra el primer ítem al examinado, se estima su habilidad y se elige dentro de un banco de ítemes previamente calibrados, el siguiente ítem que se le presentará pero este ítem debe ser adecuado a la habilidad estimada previamente del examinado. Un ítem calibrado es aquél para el cual se ha determinado el valor de sus parámetros. Por otro lado, debe recordarse que se tomó sólo un subconjunto de ítemes, aquellos que de antemano, se pensaba medían una sola habilidad. Esto significa que si se aplican Pruebas Adaptativas se obtendrá una matriz donde las columnas serían las habilidades examinadas y los renglones los niveles de habilidad del examinado en cada una de esas habilidades examinadas. Estos datos se pueden usar como input para otros análisis que se antojan multivariados por la naturaleza de los datos, brindando esto la oportunidad de explorarlos bajo otra perspectiva.

## BIBLIOGRAFÍA

- Albert, J. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics*, 17, 251-269.
- Baker, F. y Kim, S. (2004). *Item Response Theory Parameter Estimation Techniques*. Marcel Dekker, Inc.. Nueva York, USA.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring and examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp.397-479) Reading, MA: Adisson-Wesley.
- Bock, D. (1997). A Brief History of Item Response Theory. *Educational Measurement: Issues and Practice* 16 (4) 21-33.
- Bock, R. y Aitkin M. (1981). Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. *Psychometrika*, 46, 443-459.
- Bryce, B. (2000). An Introduction to Modern Measurement Theory. <http://appliedresearch.cancer.gov/areas/cognitive/inmt.pdf>.
- Cooke, D. y Michie, C. (1997). An item response theory analysis of the Hare Psychopathy Checklist. *Psychological Assessment*, 9, 3-14.
- Datos del examen de ingreso a licenciaturas de la UAQ (EXHCOBA), periodo junio de 2006.
- Ferrando, P. (1996). Evaluación de la unidimensionalidad de los ítems mediante análisis factorial. *Psicothema* 8, 397-410.
- Goegebeur, Y., De Boeck, P., Molenberghs G. y Del Pino, G. (2006). A local-influence-based diagnostic approach to a speeded item response theory model. *Applied Statistics*. 55 (5) 647-676.
- Johnson, D. (2000). *Métodos Multivariados Aplicados al Análisis de Datos*. International Thomson Editores.
- Lindley, D. (1971). *Bayesian statistics: A review*. Philadelphia: Society for Industrial and Applied Mathematics.
- Lord, F. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Erlbaum Associates. Hillsdale, Nueva Jersey.
- Ramsay, J. O. (2000). *TestGraf: A Program for the Graphical Analysis of Multiple Choice Test and Questionnaire Data*. Universidad de McGill.
- Reise, S. P., Ainsworth, A. T. y Haviland, M. G. (2005). Fundamentals, Applications, and Promise in Psychological Research. *Current Directions in psychological science*. 14 (2) 95-101.
- Sijtsma, K. and Junker, B. (2006). Item Response Theory: Past Performance, Present Developments, and Future Expectations. *Behaviormetrika*. 33, 75-102.
- Sinharay, S. (2005). Assessing Fit of Unidimensional Item Response Theory Models Using a Bayesian Approach. *Journal of Educational Measurement*, 42, 4, 375-394.
- Oshima T. C. (1994). The effect of Speededness on Parameter Estimation in Item Response Theory. *Journal of Educational Measurement*. 31 (3) 200-219.
- Wikipedia, The Free Encyclopedia. <http://en.wikipedia.org>
- Wood, R., Wilson D., Gibbons R., Schilling S., Muraki, E. y Bock, D., (2003). *Manual de TESTFACT*

## APÉNDICES

### A. Aspectos computacionales

Código utilizado en WINBUGS para estimar los parámetros del ítem y del examinado con el modelo ojiva normal de 2 parámetros:

```
# Modelo ojiva normal de dos parámetros
# Datos del EXHCOBA, 20 preguntas sobre Matemáticas para el Cálculo
model
{
  for (i in 1:N){                                #N es el número de examinados
    theta[i]~dnorm(0,1)                          # distribución a priori para theta
    for (j in 1:T){
      p[i, j] <- phi(a[j]*(theta[i] - b[j]))      #phi es el modelo ojiva normal
      y[i, j]~dbern(p[i, j])                      #las respuestas y[i,j] tienen distribución
Bernoulli
    }
  }
  # Distribuciones apriori para los parámetros del ítem
  for(i in 1:T){                                  # T es el número de ítems que se analizan
    a[i]~dnorm(0.5,1)I(0,)                        # I(0) indica que la distrib. se trunca por la izq.
  }
  en 0
  b[i]~dnorm(0,1)
}
}
Data click on one of the arrows to open the data
Inits list(a = c(.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5), #valores iniciales de a
y b.
b = c(.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5,.5))
```

```

////////////////////////////////////
////////

```

Código utilizado en WINBUGS para estimar los parámetros del ítem y examinados con el modelo ojiva normal de 3 parámetros:

```

# Modelo ojiva normal de tres parámetros
# Datos del EXHCOBA, 20 preguntas sobre Matemáticas para el Cálculo
model
{
  for (i in 1:N){
    theta[i]~dnorm(0,1)
    for (j in 1:T){
      p[i, j] <- c[j]+((1-c[j]) * phi(a[j]*(theta[i] - b[j])))
      y[i, j]~dbern(p[i, j])
    }
  }
  # Distribuciones apriori para los parámetros del ítem
  for(i in 1:T){
    a[i]~dlnorm(0,0.5)
    b[i]~dnorm(0,1)
    c[i]~dbeta(5,17)
  }
}

```

Data click on one of the arrows to open the data

```

Inits list(a = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1), b =
c(.2,.2,.2,.2,.2,.2,.2,.2,.2,.2,.2,.2,.2,.2,.2,.2,.2,.2,.2,.2),
c=c(.1,.1,.1,.1,.1,.1,.1,.1,.1,.1,.1,.1,.1,.1,.1,.1,.1,.1,.1,.1))

```

## B. Datos utilizados

### RESPUESTAS A LOS 20 ÍTEMES DE MATEMÁTICAS PARA EL CÁLCULO, 109 EXAMINADOS.

En la matriz de datos, abajo, cada renglón de la matriz corresponde a las respuestas dadas por un examinado. Cada columna de la matriz corresponde a las respuestas de todos los examinados a un mismo ítem. Una respuesta de 1 indica que la respuesta fue correcta, mientras que una respuesta de 0 indica que fue incorrecta.

```

00101101100011000000
10100000100010001111
01111011111010111011
11111001011011001111
00011001101011001111
01111011110011011010
00111001100100001000
00100011010010011001
01111111111111011101
10101000100010001100
01101011110111111011
10111001110111001100
01111101011011001100
00001010010111000000
01101011110100110000
01111001110101101001
01110001001000000100
01111001011011111111
00100010110101001000
10101001111011101110
00111110111010001001
01001011011101001000
00000001001100001110
00111001111000001101
00000000000000000000
11100001000011011001
01101001111011101101
01000001000000011000
11111111111110111111
11011001110000011100
01010000100001011000
01101000000001001111
01001010110101000100
00010011010010101101
01111000000001101110
11011001100011001100
01110011111111111101

```

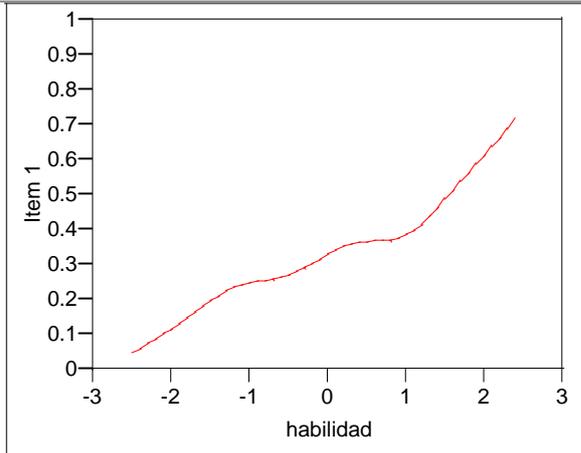
01111001111110111101  
00011011111111011111  
01011001110011101001  
00100001000111111001  
1111111111111101111  
01011001001011001110  
01001011111110011111  
01011011111001011000  
01001011111000101100  
01011001110001001110  
11111101011000001001  
00111001110011011001  
1111100111111001101  
00011101010011101110  
01110000110000101101  
0111101111111011101  
00111011111010001110  
00011011110010101111  
11100001110111101101  
00111010101111111100  
00011111111011011011  
01111011111011011100  
10100001101011001010  
00010000100001000000  
01001010111010001111  
11111001111110101101  
11111011111111111111  
11111001110011101101  
00000000000101000100  
10111011111110111111  
10100000000000000101  
11111011110110101011  
01111111111011111111  
10011000111010001001  
01110100110000001010  
0111111111111001100  
10100011010001111001  
01111101101011000100  
11111111111010111101  
01111011111111001101  
11111011111111111101  
10101001110010000100  
01111011111111111101  
00001011011010001001  
11111111110111111101

011110111111101011  
01111001001110011000  
00000010000011001000  
0111101111111111100  
11011011111010011010  
01111010111111001010  
01011000101000001101  
11111011010010101100  
01000000000000000000  
11111111111010101001  
10111011111110011100  
01010001110001011101  
01111111111110111111  
1111111010111100111  
1111111111111011110  
01111011111111111111  
01111111111011011001  
11101011110010011100  
01011011011000011110  
01001001011110100000  
11111011110010111111  
01111111111010111001  
01011001100010001101  
01001001100111001001  
11111001100010011100  
0011001111111011101  
11010001111011101101

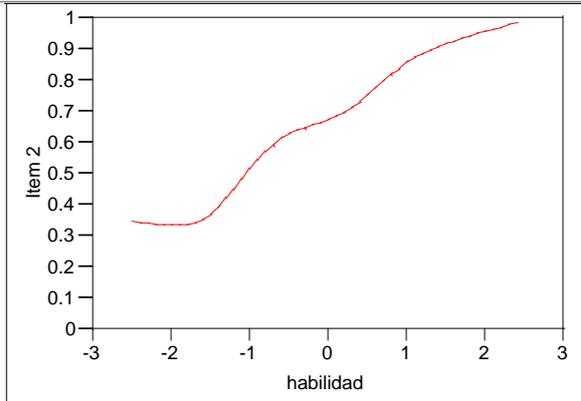
### C. Curvas Características de los Ítems, usando TESTGRAF

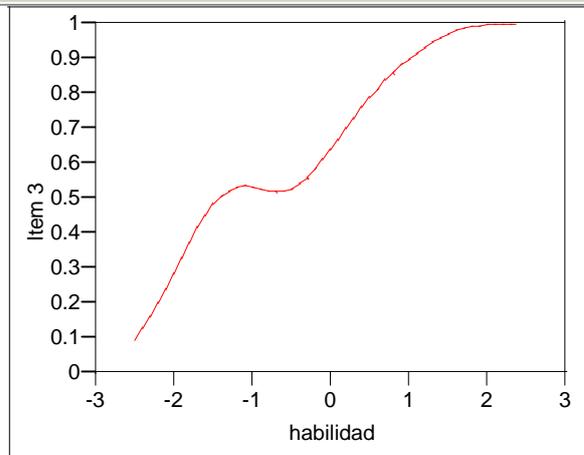
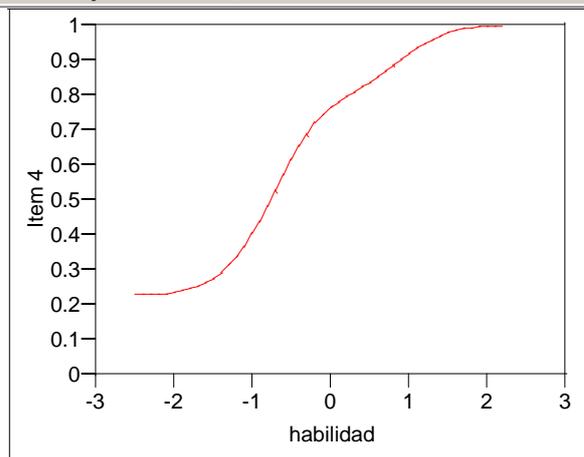
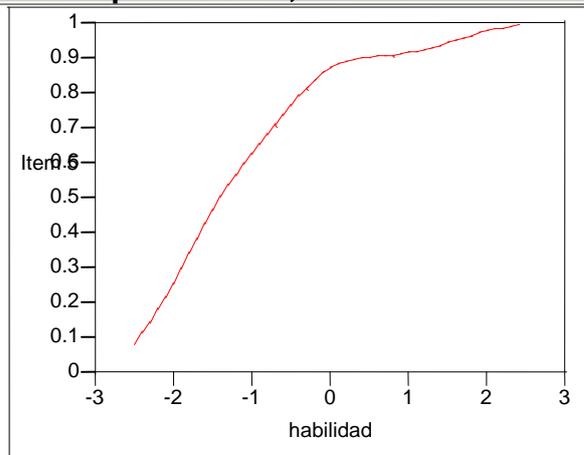
El programa TESTGRAF utiliza un modelo no paramétrico para estimar las curvas características de los ítems.

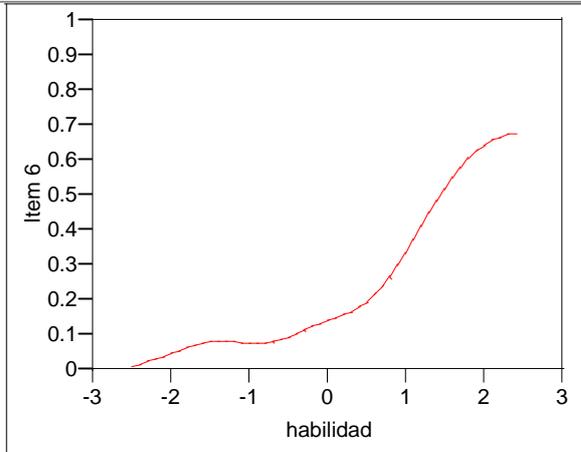
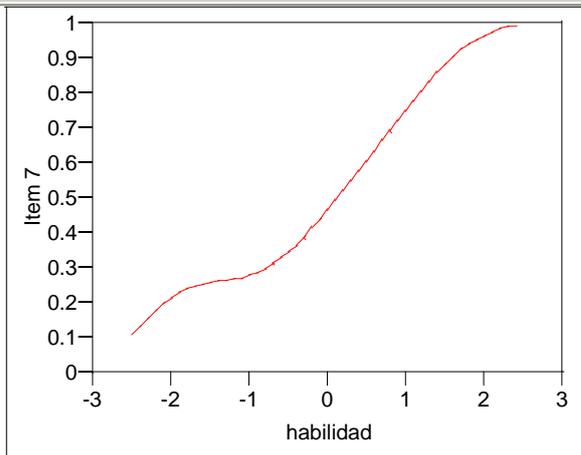
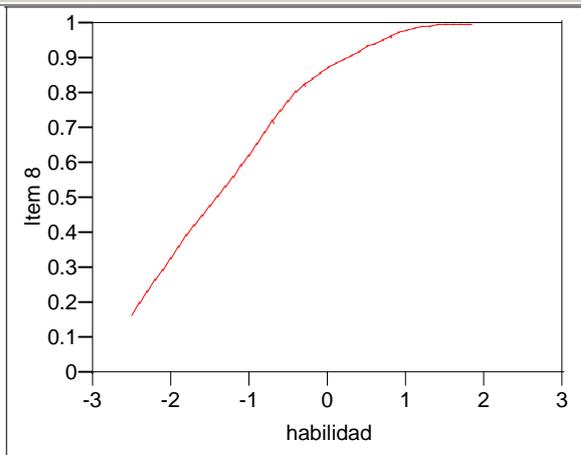
CCI no paramétrica, Item 1

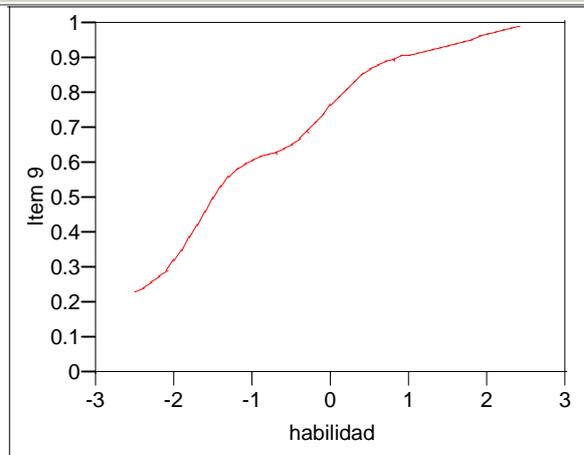
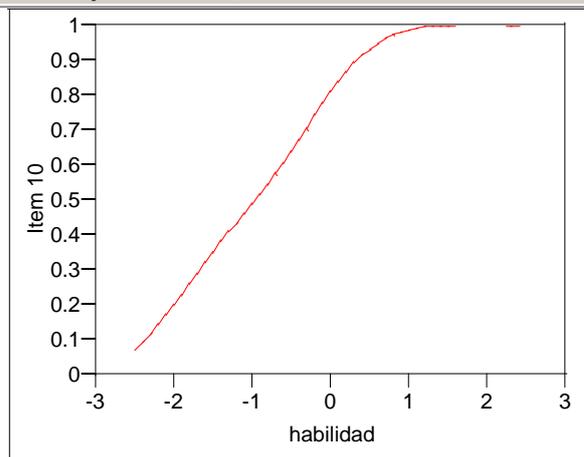
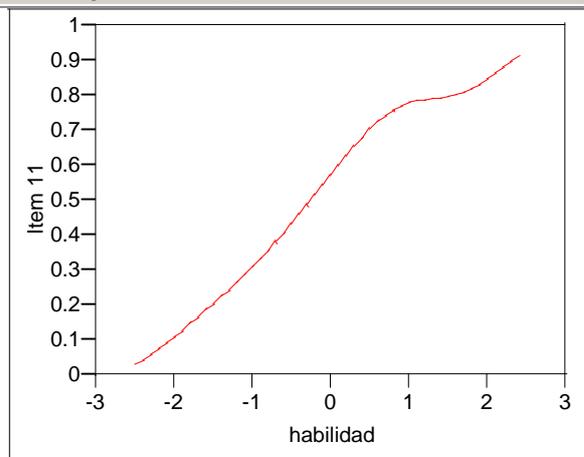


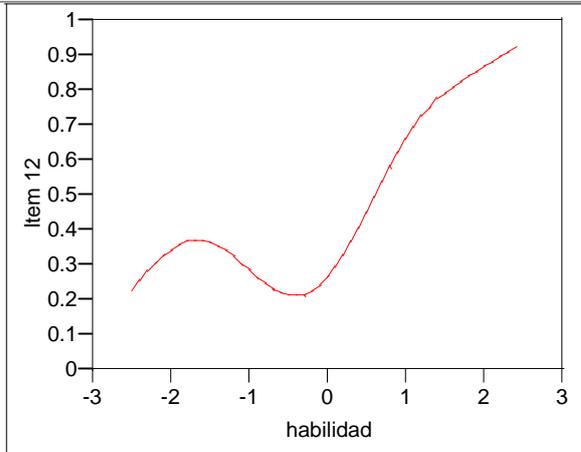
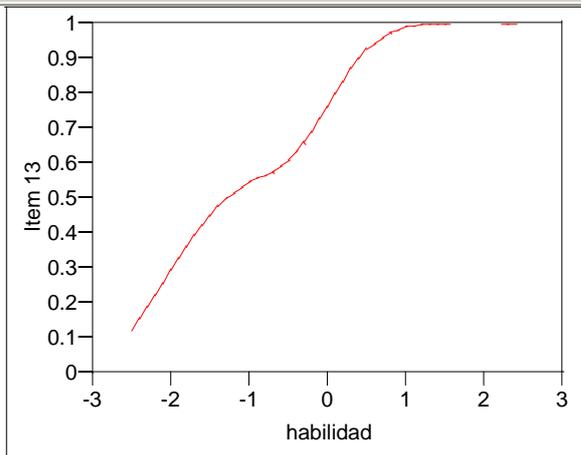
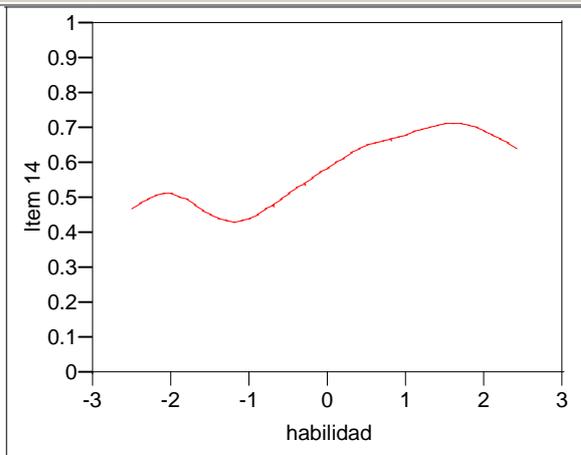
CCI no paramétrica, Item 2

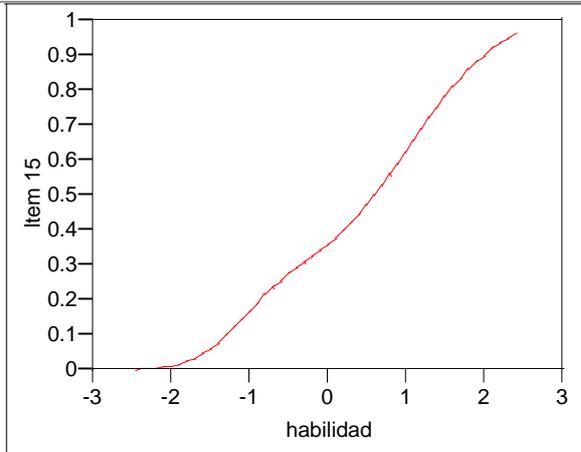
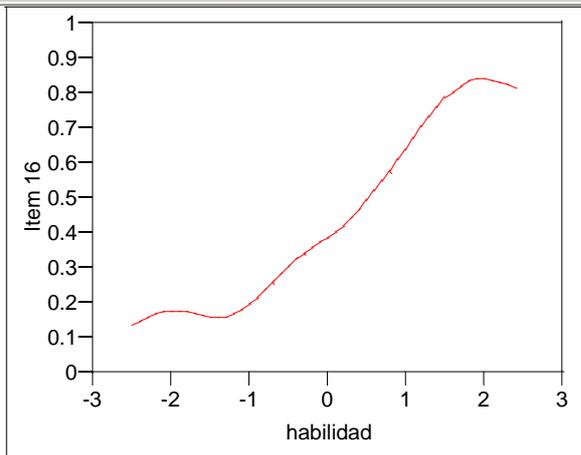
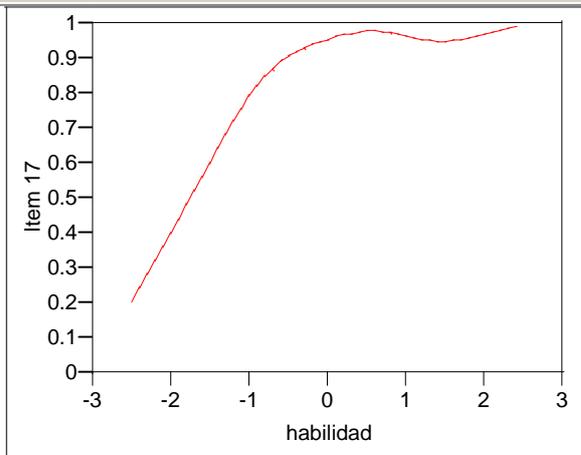


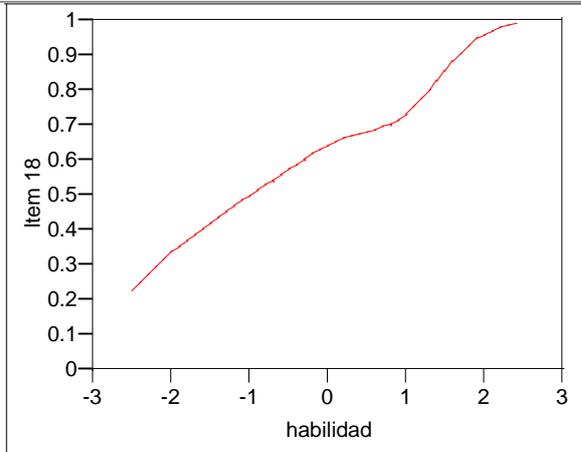
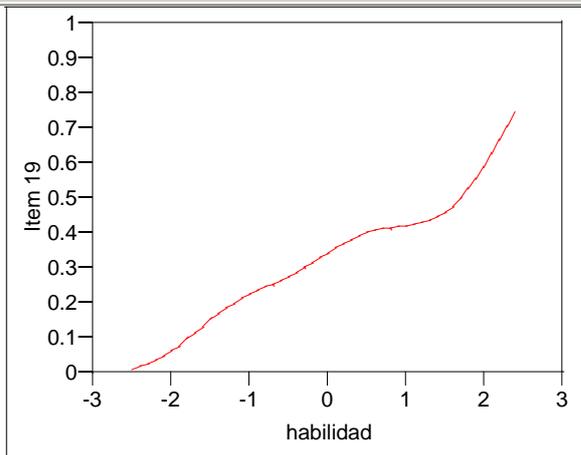
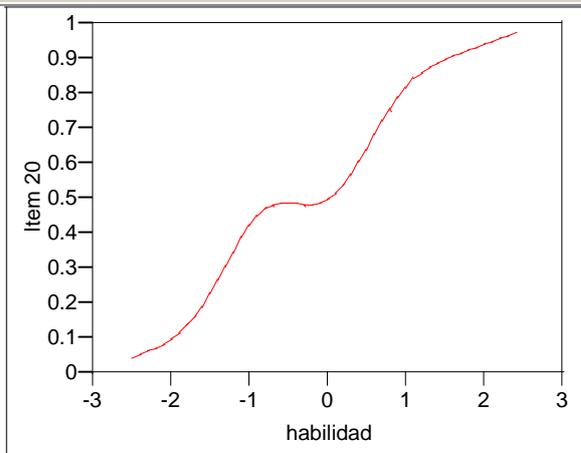
**CCI no paramétrica, Item 3****CCI no paramétrica, Item 4****CCI no paramétrica, Item 5**

**CCI no paramétrica, Item 6****CCI no paramétrica, Item 7****CCI no paramétrica, Item 8**

**CCI no paramétrica, Item 9****CCI no paramétrica, Item 10****CCI no paramétrica, Item 11**

**CCI no paramétrica, Item 12****CCI no paramétrica, Item 13****CCI no paramétrica, Item 14**

**CCI no paramétrica, Item 15****CCI no paramétrica, Item 16****CCI no paramétrica, Item 17**

**CCI no paramétrica, Item 18****CCI no paramétrica, Item 19****CCI no paramétrica, Item 20**

#### D. Estimaciones de parámetros de ítems y examinados usando el modelo ojiva normal 2 parámetros y WINBUGS

La siguiente es la salida del programa WINBUGS, al estimar los parámetros de ítems y examinados. Los nodos en la primera columna, identifican a los parámetros. Los primeros 40 nodos identifican a los parámetros de ítems y los restantes 109 a los parámetros de examinados. La segunda columna (mean) es la media de la distribución posterior y contiene las estimaciones de los correspondientes parámetros. La tercera columna (sd) contiene las desviaciones estándar de las estimaciones de los parámetros. La cuarta columna (MC error) es el error Monte Carlo. La quinta, sexta y séptima columnas proporcionan estimaciones de los cuantiles 2.5%, 50% y 97.5% respectivamente, de la distribución posterior.

| node  | mean    | sd     | MC error | 2.5%    | median  | 97.5%   |
|-------|---------|--------|----------|---------|---------|---------|
| a[1]  | 0.3568  | 0.1232 | 0.002455 | 0.1489  | 0.3452  | 0.6313  |
| a[2]  | 0.5422  | 0.1701 | 0.003641 | 0.2465  | 0.5299  | 0.9156  |
| a[3]  | 0.6966  | 0.2014 | 0.003702 | 0.3354  | 0.6821  | 1.126   |
| a[4]  | 0.9363  | 0.2586 | 0.00541  | 0.5016  | 0.9125  | 1.508   |
| a[5]  | 1.058   | 0.2895 | 0.006943 | 0.5671  | 1.031   | 1.704   |
| a[6]  | 0.685   | 0.1826 | 0.004637 | 0.3755  | 0.6683  | 1.083   |
| a[7]  | 0.8585  | 0.2336 | 0.005122 | 0.4422  | 0.8419  | 1.366   |
| a[8]  | 1.217   | 0.3322 | 0.006559 | 0.6711  | 1.18    | 1.957   |
| a[9]  | 0.8009  | 0.2138 | 0.004813 | 0.4376  | 0.7833  | 1.264   |
| a[10] | 1.638   | 0.4366 | 0.01043  | 0.9259  | 1.591   | 2.618   |
| a[11] | 0.8053  | 0.2186 | 0.003554 | 0.4137  | 0.7927  | 1.277   |
| a[12] | 0.4403  | 0.1559 | 0.002477 | 0.1649  | 0.4307  | 0.7725  |
| a[13] | 1.154   | 0.3011 | 0.006515 | 0.6534  | 1.125   | 1.824   |
| a[14] | 0.1714  | 0.1078 | 0.001542 | 0.01116 | 0.1565  | 0.413   |
| a[15] | 0.7406  | 0.2093 | 0.004294 | 0.3685  | 0.728   | 1.186   |
| a[16] | 0.68    | 0.1966 | 0.003437 | 0.3277  | 0.6693  | 1.085   |
| a[17] | 1.104   | 0.2976 | 0.007043 | 0.6251  | 1.072   | 1.782   |
| a[18] | 0.4017  | 0.1435 | 0.002207 | 0.1542  | 0.3895  | 0.7061  |
| a[19] | 0.4245  | 0.1422 | 0.002656 | 0.1837  | 0.4126  | 0.7361  |
| a[20] | 0.6662  | 0.1975 | 0.003066 | 0.3105  | 0.6562  | 1.092   |
| b[1]  | 1.246   | 0.4633 | 0.008311 | 0.4932  | 1.196   | 2.305   |
| b[2]  | -0.9347 | 0.3679 | 0.008945 | -1.816  | -0.8847 | -0.364  |
| b[3]  | -0.7389 | 0.2882 | 0.00561  | -1.416  | -0.7063 | -0.2649 |
| b[4]  | -0.723  | 0.2364 | 0.005719 | -1.255  | -0.6992 | -0.3352 |
| b[5]  | -0.9518 | 0.2475 | 0.006527 | -1.514  | -0.9272 | -0.5407 |
| b[6]  | 1.488   | 0.3657 | 0.008708 | 0.9086  | 1.442   | 2.33    |
| b[7]  | 0.01286 | 0.1935 | 0.003573 | -0.3686 | 0.01326 | 0.3991  |
| b[8]  | -1.009  | 0.2364 | 0.006077 | -1.543  | -0.9826 | -0.6129 |
| b[9]  | -0.9483 | 0.2812 | 0.007492 | -1.573  | -0.9187 | -0.4936 |
| b[10] | -0.6644 | 0.1739 | 0.004998 | -1.017  | -0.6552 | -0.3579 |
| b[11] | -0.1627 | 0.2072 | 0.004259 | -0.6007 | -0.156  | 0.2259  |
| b[12] | 0.6309  | 0.3657 | 0.005405 | 0.0244  | 0.5911  | 1.485   |
| b[13] | -0.7746 | 0.2078 | 0.005634 | -1.228  | -0.7602 | -0.4124 |
| b[14] | -0.552  | 0.6757 | 0.007804 | -1.938  | -0.5321 | 0.8428  |

| node      | mean     | sd     | MC error | 2.5%     | median    | 97.5%    |
|-----------|----------|--------|----------|----------|-----------|----------|
| b[15]     | 0.4935   | 0.2358 | 0.004786 | 0.082    | 0.4738    | 1.028    |
| b[16]     | 0.3898   | 0.2467 | 0.00486  | -0.02855 | 0.3707    | 0.9418   |
| b[17]     | -1.432   | 0.2989 | 0.007751 | -2.101   | -1.396    | -0.953   |
| b[18]     | -0.8345  | 0.4225 | 0.006534 | -1.808   | -0.7814   | -0.1652  |
| b[19]     | 1.064    | 0.4121 | 0.007423 | 0.3938   | 1.018     | 2.006    |
| b[20]     | -0.2359  | 0.2383 | 0.004557 | -0.7464  | -0.2247   | 0.1962   |
| theta[1]  | -0.7787  | 0.36   | 0.006059 | -1.505   | -0.7732   | -0.08566 |
| theta[2]  | -0.8267  | 0.3568 | 0.005754 | -1.544   | -0.8211   | -0.1478  |
| theta[3]  | 1.003    | 0.5054 | 0.006224 | 0.0783   | 0.9779    | 2.066    |
| theta[4]  | 0.2894   | 0.4146 | 0.006007 | -0.4832  | 0.2771    | 1.144    |
| theta[5]  | -0.2681  | 0.3573 | 0.005989 | -0.9625  | -0.2745   | 0.4473   |
| theta[6]  | 0.3273   | 0.4176 | 0.005679 | -0.4624  | 0.3166    | 1.187    |
| theta[7]  | -0.724   | 0.3571 | 0.005325 | -1.436   | -0.7186   | -0.02413 |
| theta[8]  | -0.3972  | 0.3565 | 0.006294 | -1.083   | -0.4018   | 0.3106   |
| theta[9]  | 1.228    | 0.5286 | 0.008577 | 0.2549   | 1.205     | 2.32     |
| theta[10] | -0.8453  | 0.3569 | 0.005782 | -1.56    | -0.8361   | -0.1602  |
| theta[11] | 0.5375   | 0.4385 | 0.007455 | -0.2786  | 0.5242    | 1.445    |
| theta[12] | 0.01643  | 0.3875 | 0.005767 | -0.7105  | 0.01008   | 0.815    |
| theta[13] | 0.117    | 0.3945 | 0.005798 | -0.6165  | 0.1042    | 0.9341   |
| theta[14] | -0.874   | 0.3515 | 0.005781 | -1.58    | -0.8665   | -0.1962  |
| theta[15] | -0.2984  | 0.358  | 0.006105 | -0.9937  | -0.3028   | 0.4283   |
| theta[16] | 0.003284 | 0.3822 | 0.005161 | -0.7239  | -0.006694 | 0.7752   |
| theta[17] | -1.024   | 0.3653 | 0.006032 | -1.772   | -1.013    | -0.3258  |
| theta[18] | 0.5733   | 0.4416 | 0.005949 | -0.2312  | 0.5565    | 1.49     |
| theta[19] | -0.8013  | 0.357  | 0.005823 | -1.537   | -0.791    | -0.1271  |
| theta[20] | 0.1366   | 0.3913 | 0.005373 | -0.6089  | 0.123     | 0.9383   |
| theta[21] | 0.06887  | 0.3853 | 0.006803 | -0.6661  | 0.06074   | 0.8604   |
| theta[22] | -0.4234  | 0.3557 | 0.005739 | -1.127   | -0.4242   | 0.2885   |
| theta[23] | -1.049   | 0.3637 | 0.005922 | -1.792   | -1.041    | -0.3657  |
| theta[24] | -0.1448  | 0.3654 | 0.005721 | -0.8528  | -0.1532   | 0.6088   |
| theta[25] | -2.399   | 0.5632 | 0.008061 | -3.608   | -2.364    | -1.423   |
| theta[26] | -0.6451  | 0.3509 | 0.00546  | -1.345   | -0.6375   | 0.02247  |
| theta[27] | 0.2405   | 0.4085 | 0.005544 | -0.5292  | 0.2329    | 1.066    |
| theta[28] | -1.222   | 0.3843 | 0.006059 | -2.015   | -1.215    | -0.4987  |
| theta[29] | 1.946    | 0.6284 | 0.007663 | 0.8092   | 1.909     | 3.263    |
| theta[30] | -0.2499  | 0.3665 | 0.005314 | -0.9592  | -0.2549   | 0.4847   |
| theta[31] | -1.085   | 0.3746 | 0.005602 | -1.851   | -1.071    | -0.3872  |
| theta[32] | -0.9266  | 0.3658 | 0.006101 | -1.684   | -0.9202   | -0.232   |
| theta[33] | -0.7757  | 0.3541 | 0.006411 | -1.481   | -0.768    | -0.08826 |
| theta[34] | -0.2653  | 0.3657 | 0.006262 | -0.9807  | -0.2673   | 0.4501   |
| theta[35] | -0.7362  | 0.3506 | 0.005219 | -1.434   | -0.7311   | -0.04904 |
| theta[36] | -0.4451  | 0.3582 | 0.005725 | -1.15    | -0.4481   | 0.2456   |
| theta[37] | 0.694    | 0.4633 | 0.008132 | -0.1648  | 0.677     | 1.657    |
| theta[38] | 0.8265   | 0.4795 | 0.00625  | -0.06084 | 0.8012    | 1.827    |

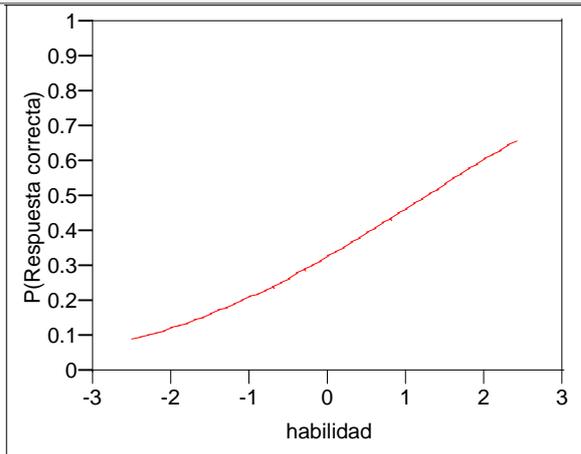
| node      | mean     | sd     | MC error | 2.5%     | median   | 97.5%   |
|-----------|----------|--------|----------|----------|----------|---------|
| theta[39] | 0.6193   | 0.4474 | 0.005687 | -0.2117  | 0.6018   | 1.557   |
| theta[40] | 0.02619  | 0.3875 | 0.0061   | -0.7024  | 0.01504  | 0.795   |
| theta[41] | -0.5879  | 0.3511 | 0.005575 | -1.281   | -0.5851  | 0.09439 |
| theta[42] | 1.61     | 0.5747 | 0.006936 | 0.5354   | 1.584    | 2.826   |
| theta[43] | -0.4392  | 0.3552 | 0.00564  | -1.138   | -0.4424  | 0.2709  |
| theta[44] | 0.432    | 0.4232 | 0.007423 | -0.3439  | 0.4108   | 1.313   |
| theta[45] | -0.04339 | 0.3736 | 0.005432 | -0.7605  | -0.04985 | 0.7038  |
| theta[46] | -0.1771  | 0.3667 | 0.005625 | -0.8802  | -0.1892  | 0.5688  |
| theta[47] | -0.3441  | 0.3589 | 0.005508 | -1.039   | -0.3469  | 0.3621  |
| theta[48] | -0.05073 | 0.3753 | 0.00611  | -0.7555  | -0.0576  | 0.7182  |
| theta[49] | 0.05251  | 0.3915 | 0.005528 | -0.6994  | 0.04069  | 0.843   |
| theta[50] | 0.5389   | 0.4461 | 0.006751 | -0.2915  | 0.5198   | 1.466   |
| theta[51] | -0.06382 | 0.3756 | 0.005951 | -0.7885  | -0.0714  | 0.6927  |
| theta[52] | -0.4942  | 0.3506 | 0.005501 | -1.182   | -0.4933  | 0.1943  |
| theta[53] | 0.9471   | 0.4971 | 0.006024 | 0.04229  | 0.9258   | 1.995   |
| theta[54] | 0.28     | 0.4139 | 0.005752 | -0.4832  | 0.2673   | 1.122   |
| theta[55] | 0.2455   | 0.4032 | 0.006259 | -0.5168  | 0.2336   | 1.083   |
| theta[56] | -0.02858 | 0.3797 | 0.006164 | -0.7506  | -0.04162 | 0.7437  |
| theta[57] | -0.09732 | 0.3615 | 0.005935 | -0.7941  | -0.1041  | 0.6209  |
| theta[58] | 0.5788   | 0.4492 | 0.006404 | -0.2512  | 0.5561   | 1.521   |
| theta[59] | 0.5719   | 0.4536 | 0.006429 | -0.2576  | 0.5531   | 1.522   |
| theta[60] | -0.6256  | 0.355  | 0.005094 | -1.33    | -0.6222  | 0.07231 |
| theta[61] | -1.63    | 0.438  | 0.006849 | -2.547   | -1.61    | -0.8575 |
| theta[62] | -0.1451  | 0.3719 | 0.005691 | -0.8536  | -0.1548  | 0.6075  |
| theta[63] | 0.7118   | 0.4653 | 0.007119 | -0.1582  | 0.6948   | 1.679   |
| theta[64] | 1.637    | 0.5929 | 0.006958 | 0.5445   | 1.603    | 2.889   |
| theta[65] | 0.3946   | 0.4308 | 0.006624 | -0.3992  | 0.3851   | 1.297   |
| theta[66] | -1.941   | 0.486  | 0.007075 | -2.997   | -1.914   | -1.068  |
| theta[67] | 1.246    | 0.5308 | 0.006634 | 0.269    | 1.224    | 2.33    |
| theta[68] | -1.562   | 0.4256 | 0.007057 | -2.474   | -1.54    | -0.7742 |
| theta[69] | 0.7167   | 0.4601 | 0.006276 | -0.1096  | 0.6911   | 1.701   |
| theta[70] | 1.578    | 0.584  | 0.00623  | 0.5028   | 1.555    | 2.792   |
| theta[71] | -0.3025  | 0.3573 | 0.00516  | -0.9973  | -0.3038  | 0.4125  |
| theta[72] | -0.5891  | 0.3513 | 0.005905 | -1.289   | -0.5909  | 0.1028  |
| theta[73] | 0.7039   | 0.4629 | 0.007048 | -0.1548  | 0.6857   | 1.663   |
| theta[74] | -0.3657  | 0.3558 | 0.006243 | -1.056   | -0.3706  | 0.3536  |
| theta[75] | -0.3341  | 0.3559 | 0.006246 | -1.034   | -0.3369  | 0.3808  |
| theta[77] | 0.7022   | 0.4647 | 0.006559 | -0.1513  | 0.6796   | 1.68    |
| theta[78] | 1.413    | 0.5588 | 0.007292 | 0.4011   | 1.389    | 2.587   |
| theta[79] | -0.5393  | 0.3551 | 0.005347 | -1.225   | -0.5408  | 0.1671  |
| theta[80] | 1.259    | 0.5371 | 0.007343 | 0.2861   | 1.234    | 2.385   |
| theta[81] | -0.3011  | 0.3666 | 0.005233 | -1.002   | -0.3046  | 0.4325  |
| theta[82] | 1.297    | 0.5384 | 0.007003 | 0.3111   | 1.273    | 2.437   |
| theta[83] | 0.8824   | 0.4844 | 0.006812 | -0.00072 | 0.8574   | 1.91    |

| node       | mean     | sd     | MC error | 2.5%      | median   | 97.5%    |
|------------|----------|--------|----------|-----------|----------|----------|
| theta[84]  | -0.2853  | 0.3646 | 0.005233 | -0.9876   | -0.2906  | 0.4423   |
| theta[85]  | -1.269   | 0.3865 | 0.007066 | -2.073    | -1.249   | -0.5661  |
| theta[86]  | 0.9528   | 0.4943 | 0.007256 | 0.06969   | 0.9317   | 1.997    |
| theta[87]  | 0.4052   | 0.427  | 0.005832 | -0.3814   | 0.3827   | 1.286    |
| theta[88]  | 0.1131   | 0.3947 | 0.005648 | -0.6289   | 0.09897  | 0.9367   |
| theta[89]  | -0.7077  | 0.3505 | 0.005598 | -1.405    | -0.705   | -0.02406 |
| theta[90]  | 0.1675   | 0.4016 | 0.005919 | -0.5864   | 0.159    | 0.975    |
| theta[91]  | -2.132   | 0.5183 | 0.007449 | -3.24     | -2.105   | -1.205   |
| theta[92]  | 0.9675   | 0.504  | 0.006266 | 0.05071   | 0.9352   | 2.026    |
| theta[93]  | 0.578    | 0.4538 | 0.005603 | -0.237    | 0.5543   | 1.514    |
| theta[94]  | -0.3759  | 0.3576 | 0.006127 | -1.063    | -0.382   | 0.3337   |
| theta[95]  | 1.73     | 0.5995 | 0.007614 | 0.6338    | 1.711    | 2.956    |
| theta[96]  | 0.3745   | 0.4098 | 0.007111 | -0.3804   | 0.3606   | 1.221    |
| theta[97]  | 1.222    | 0.5306 | 0.007722 | 0.2646    | 1.197    | 2.332    |
| theta[98]  | 1.465    | 0.5686 | 0.007974 | 0.4213    | 1.44     | 2.663    |
| theta[99]  | 0.8805   | 0.4874 | 0.006183 | -0.005495 | 0.8516   | 1.922    |
| theta[100] | 0.1263   | 0.3961 | 0.006483 | -0.6093   | 0.1132   | 0.9375   |
| theta[101] | -0.07935 | 0.3726 | 0.005668 | -0.808    | -0.08308 | 0.6646   |
| theta[102] | -0.4896  | 0.3495 | 0.00552  | -1.184    | -0.4829  | 0.1897   |
| theta[103] | 0.9512   | 0.4937 | 0.006809 | 0.03461   | 0.9355   | 1.984    |
| theta[104] | 1.092    | 0.5232 | 0.007159 | 0.1291    | 1.064    | 2.168    |
| theta[105] | -0.4294  | 0.3511 | 0.005146 | -1.114    | -0.4261  | 0.2576   |
| theta[106] | -0.5587  | 0.3565 | 0.006229 | -1.268    | -0.5542  | 0.1393   |
| theta[107] | -0.2255  | 0.3627 | 0.005732 | -0.9374   | -0.2386  | 0.5035   |
| theta[108] | 0.3471   | 0.417  | 0.007072 | -0.4164   | 0.3356   | 1.204    |
| theta[109] | 0.1064   | 0.3906 | 0.006217 | -0.628    | 0.09835  | 0.9103   |

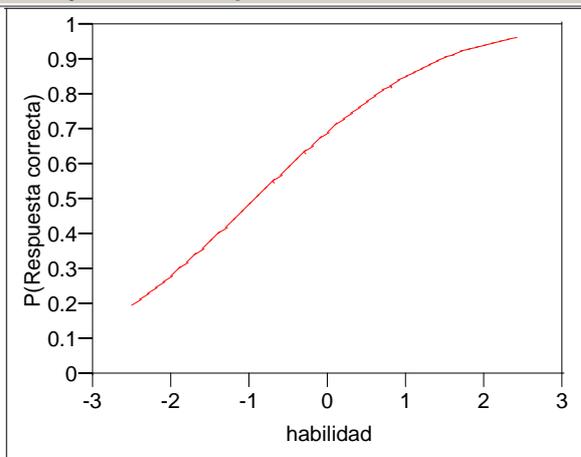
### E. Curvas Características de Ítems, usando el modelo Ojiva Normal 2 parámetros y WINBUGS.

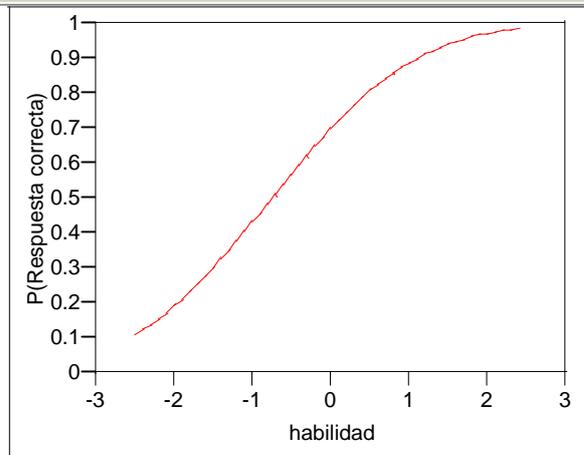
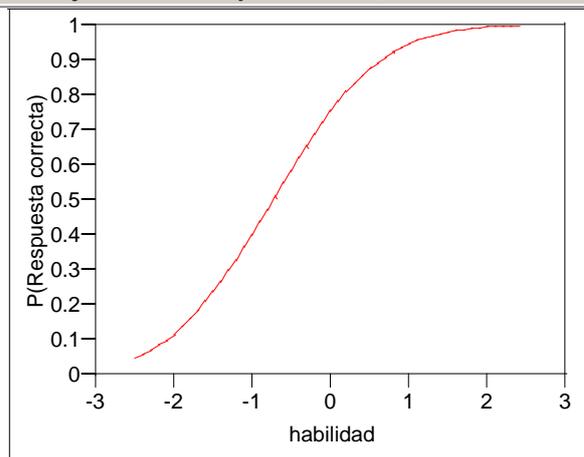
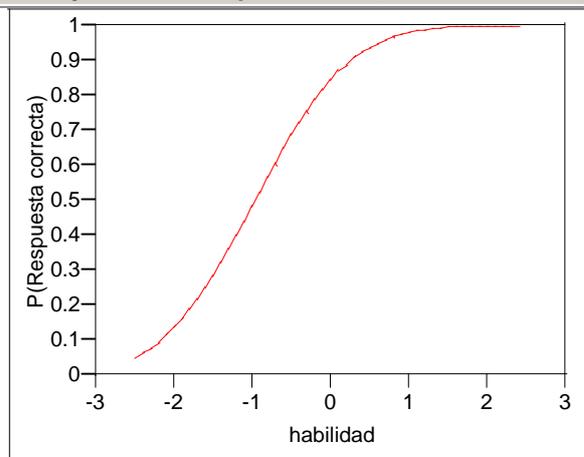
El programa WINBUGS utiliza un enfoque paramétrico Bayesiano para estimar los parámetros de las curvas características de los ítems.

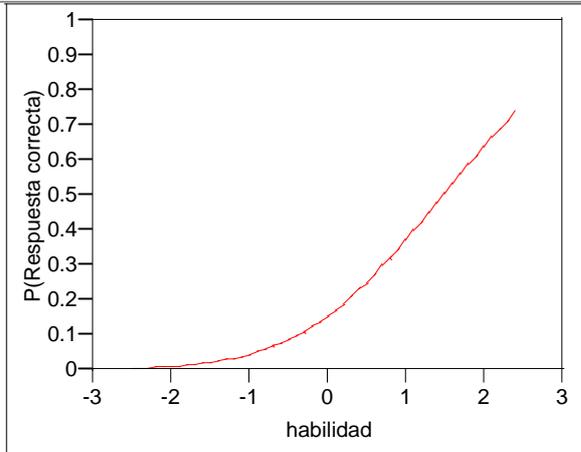
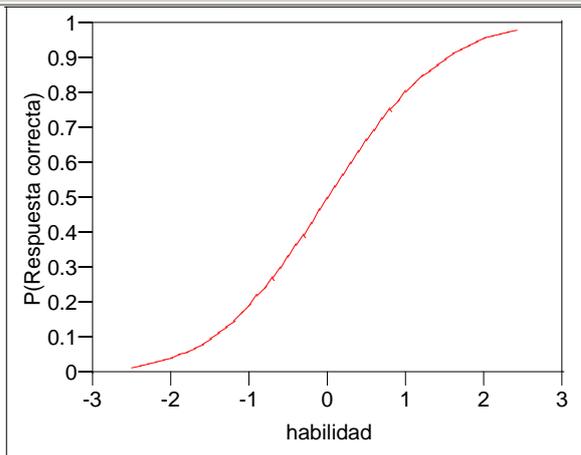
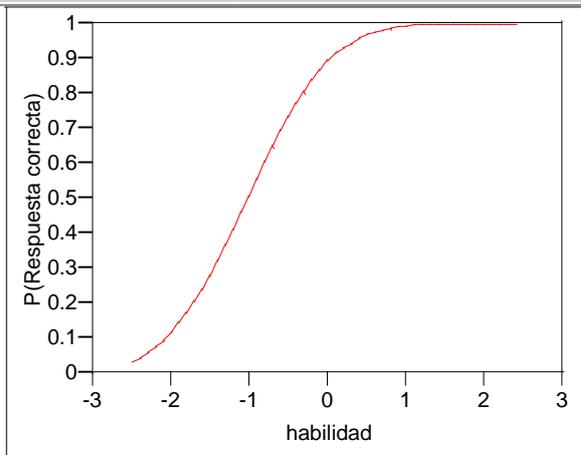
CCI Ojiva Normal 2 parámetros, Item 1

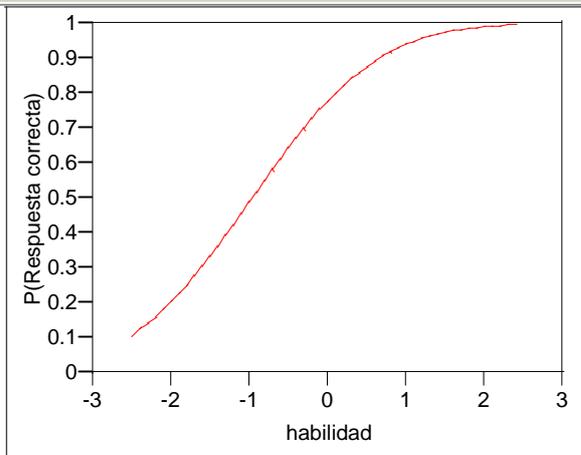
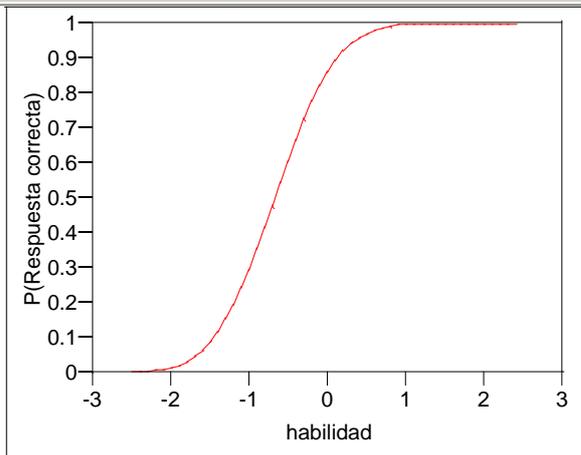
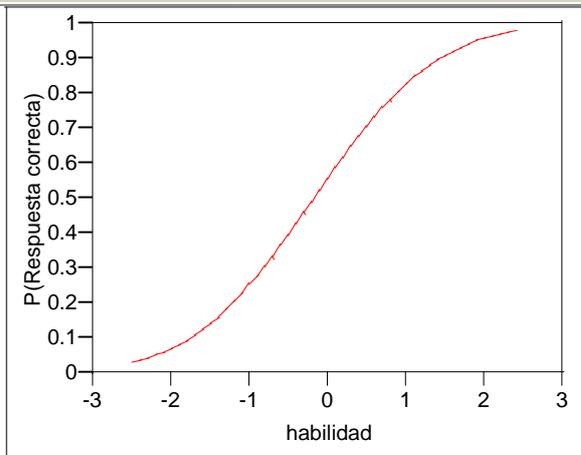


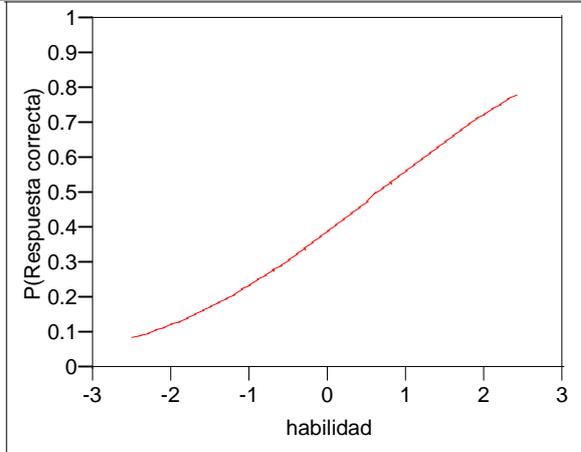
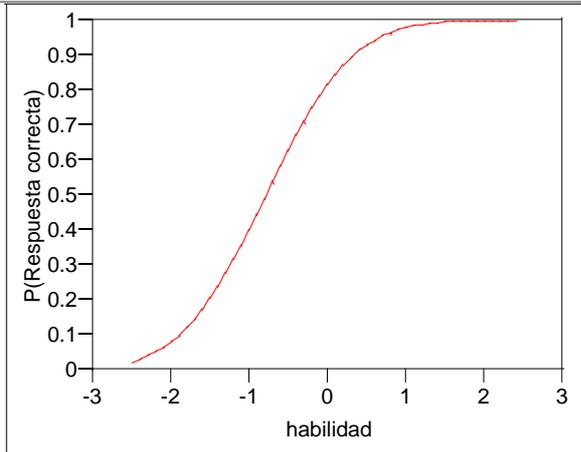
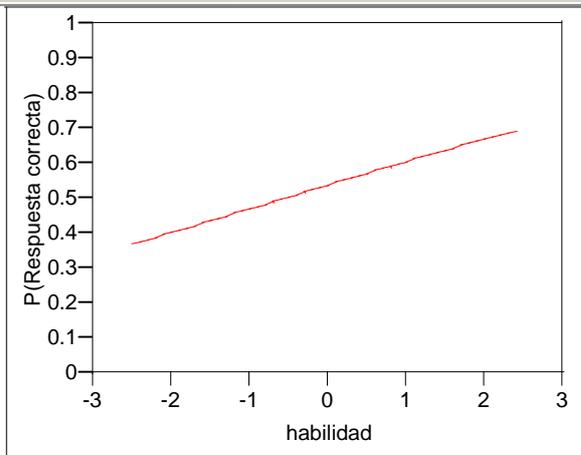
CCI Ojiva Normal 2 parámetros, Item 2

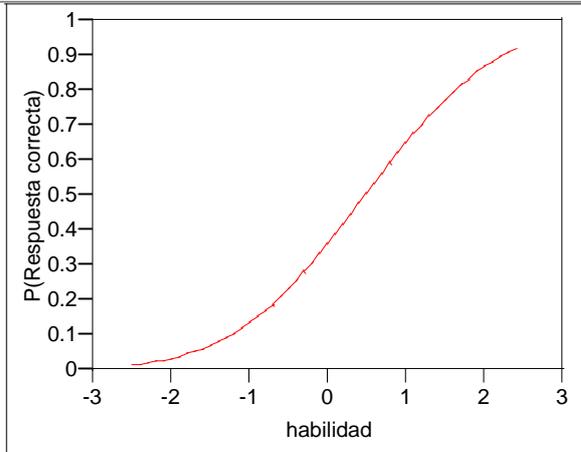
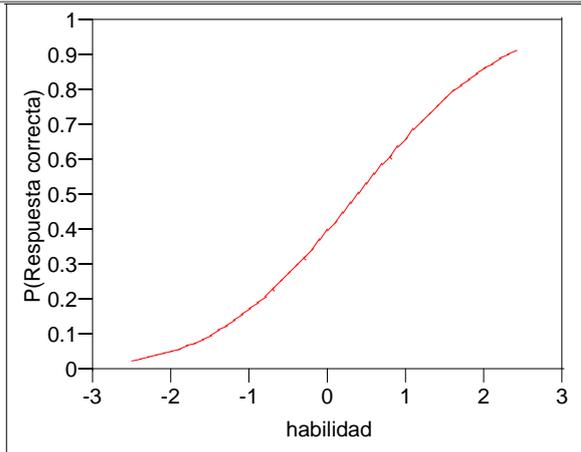
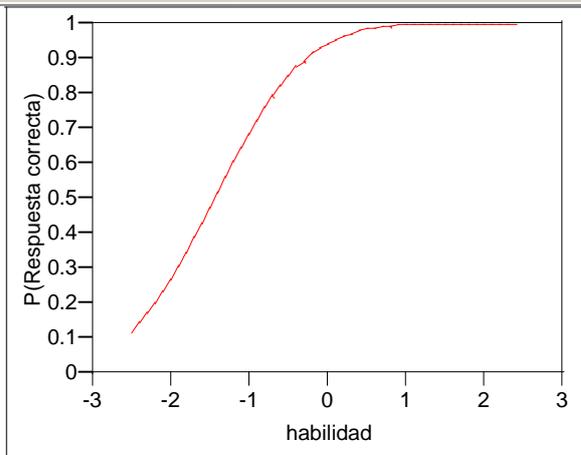


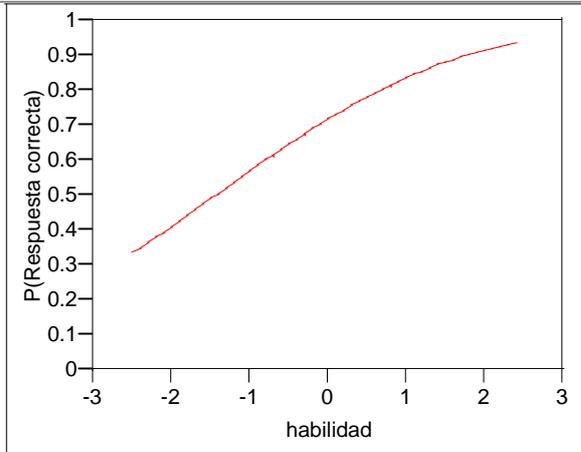
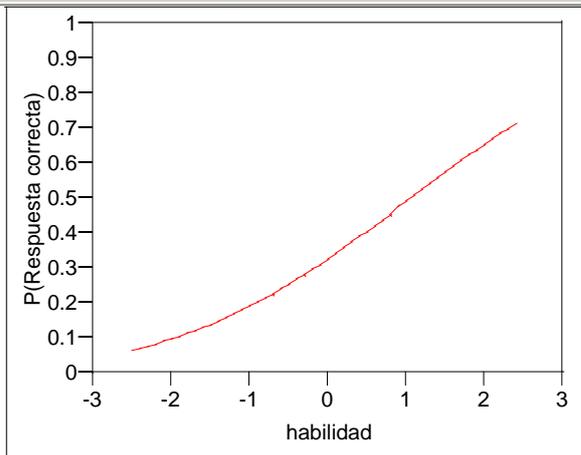
**CCI Ojiva Normal 2 parámetros, Item 3****CCI Ojiva Normal 2 parámetros, Item 4****CCI Ojiva Normal 2 parámetros, Item 5**

**CCI Ojiva Normal 2 parámetros, Item 6****CCI Ojiva Normal 2 parámetros, Item 7****CCI Ojiva Normal 2 parámetros, Item 8**

**CCI Ojiva Normal 2 parámetros, Item 9****CCI Ojiva Normal 2 parámetros, Item 10****CCI Ojiva Normal 2 parámetros, Item 11**

**CCI Ojiva Normal 2 parámetros, Item 12****CCI Ojiva Normal 2 parámetros, Item 13****CCI Ojiva Normal 2 parámetros, Item 14**

**CCI Ojiva Normal 2 parámetros, Item 15****CCI Ojiva Normal 2 parámetros, Item 16****CCI Ojiva Normal 2 parámetros, Item 17**

**CCI Ojiva Normal 2 parámetros, Item 18****CCI Ojiva Normal 2 parámetros, Item 19****CCI Ojiva Normal 2 parámetros, Item 20**