UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA

"TEORÍA DE RESPUESTA AL ÍTEM Y MODELOS DE RASCH"

PRESENTADO POR:

MANUEL ALEXANDER ABREGO PERLA

YESSICA BEATRIZ SÁNCHEZ ROSALES

PARA OPTAR AL GRADO DE:

LICENCIADO(A) EN ESTADÍSTICA

ASESOR:

Dr. JOSÉ NERYS FUNES TORRES

CIUDAD UNIVERSITARIA, FEBRERO DE 2007.

UNIVERSIDAD DE EL SALVADOR

RECTORA:

Dra. María Isabel Rodríguez

SECRETARIA GENERAL:

Licda. Alicia Margarita Rivas de Recinos

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

DECANO:

MSc. José Héctor Elías Días

SECRETARIA:

Licda. Marta Noemy Martínez de Rosales

ESCUELA DE MATEMÁTICA

DIRECTOR:

Lic. Mauricio Hernán Lovo Córdova

TRABAJO DE GRADUACIÓN APROBADO POR:

COORDINADOR:

Lic. Mauricio Hernán Lovo Córdova

ASESOR:

Dr. José Nerys Funes Torres

DEDICATORIAS

En primer lugar doy gracias a Dios por haberme permitido culminar mi carrera profesional, a mi querida madre Vidalia Dionicia Perla por su sacrificio y ser el pilar fundamental para alcanzar mi meta; asimismo, le doy gracias a mis tíos principalmente a: Francisco Perla y Jorge Adalberto Perla por haberme brindado todo su apoyo cuando más lo necesitaba; a mis queridísimos abuelos: María Luisa Perla e Inés Cecíles (en paz descansen) por sus consejos y apoyo moral en toda mi vida de estudiante y de igual manera quiero expresar mi agradecimiento a mi asesor de tesis, Dr. José Nerys Funes Torres por su paciencia y enseñanza al desarrollar este trabajo de graduación.

Manuel Alexander Abrego Perla

Doy gracias a Dios por haberme permitido culminar mi carrera profesional, a mis queridos y abnegados padres: Licda. Juana de Jesús Sánchez Rosales y Prof. José Reyes Sánchez Echeverría por sus inmensos sacrificios diarios que hicieron para ayudarme a lograrlo, ya que con su amor y comprensión logre alcanzar mi meta; a mis hermanos por sus palabras de aliento; a mis tíos y padrinos por sus consejos; a mi abuela Margarita Rosales por sus cuidados y apoyo moral, y finalmente a mis amigos por su gran apoyo en mis estudios.

Yessica Beatriz Sánchez Rosales

ÍNDICE

CONTENIDO **PÁG.**

CAPÍTULO III: APLICACIÓN DEL MODELO DE RASCH

ANEXOS

INTRODUCCIÓN

Actualmente es frecuente encontrar diversas investigaciones en el área de la educación donde se hace uso de diferentes metodologías con el fin de obtener resultados que muestren la realidad del sistema educativo. Cuando el objeto de estudio en una disciplina científica involucra a personas o grupos sociales, los retos para la medición se presentan desde el momento mismo en que se intentan definir las características o fenómenos a ser investigados. Los test (cuestionarios) son conceptualizaciones, construcciones mentales que usan los evaluadores, investigadores o estadísticos para lograr describir o explicar aspectos que se desean estudiar en los individuos.

Medir el nivel de conocimiento a través de un test que uno o varios sujetos tienen acerca de una disciplina ha sido una necesidad por décadas anteriores, sobre todo en el ámbito de la educación. Muchos investigadores tanto educativos como en otras disciplinas (Psicología, Sociología, etc.) han utilizado diversas teorías y modelos, una de estas teorías ha sido la Teoría Clásica de los Test (TCT), que se utiliza para conocer las puntuaciones concretas de uno o varios individuos y a la vez medir la capacidad de medición del test. Sin embargo, esta teoría presenta algunas deficiencias.

Para obtener medidas que sean independientes del instrumento y los sujetos evaluados surge la Teoría de Respuesta al Ítem (TRI). La TRI tiene como objetivos obtener mediciones que no varíen en función del instrumento utilizado, disponer de instrumentos de medida que no dependan de los objetos medidos, es decir, que sean invariantes respecto a los sujetos evaluados. Dentro de esta teoría se encuentra el Modelo de Rasch, el cual nos permite medir un atributo en una única dimensión donde se sitúan conjuntamente los ítems y las personas, además se puede obtener la probabilidad de que una persona responda correctamente a uno o varios ítems dependiendo de su capacidad o rasgo.

La Teoría de Respuesta al Ítem y el Modelo de Rasch están siendo ampliamente aplicados para el análisis de los resultados de la prueba de ingreso que se realiza en las universidades de diferentes países del mundo, como México, EEUU, Chile, Argentina, Dinamarca, Suecia, entre otros. Obteniendo indicadores confiables para orientar las acciones y programas de mejoramiento de la calidad de la enseñanza. Así como, la toma de decisión en la ubicación de las carreras de los estudiantes que solicitan ingresar por primera vez a la universidad.

Estos aspectos serán abordados en el presente trabajo de graduación que está estructurado en tres capítulos: Análisis de Ítems, Modelos de Rasch y Aplicación del Modelo de Rasch.

En el primer capítulo se estudia las teorías que se utilizan para el análisis de ítems: la Teoría Clásica del Test (TCT) y Teoría de Respuesta al Ítem (TRI). En la TCT se presenta los supuestos básicos, algunos indicadores más importantes y la forma de cómo obtenerlos, y en la TRI se estudia las características de los principales modelos más difundidos, así como también las ventajas que ofrece la TRI con respecto a la TCT.

En el segundo capítulo se estudia los modelos de Rasch: el modelo de Rasch para datos dicotómicos y el modelo de crédito parcial (MCP) considerado como el modelo general de Rasch, el cual trabaja con datos politómicos (categóricos). Se presentan las características principales y la estimación de los parámetros de dichos modelos, así como también las ventajas de utilizar el modelo de Rasch con respecto a la TCT.

En el tercer capítulo se desarrollan dos aplicaciones del modelo de Rasch. En la primera aplicación se dispone de una base de datos con las notas por cada problema o ítem del primer parcial de matemática I, año 2006, que fue impartida por docentes de la Escuela de Matemática, de la Facultad de Ciencias Naturales y Matemática de la Universidad de El Salvador, en el cual se hará uso del modelo de crédito parcial (MCP), ya que los resultados de los estudiantes fueron agrupados en categorías. En la segunda aplicación se dispone de los datos obtenidos en el Test de Nuevo Ingreso, UES 2007, donde se hará uso del modelo dicotómico de Rasch. Asimismo, se presenta las conclusiones y recomendaciones a partir de los resultados obtenidos.

Finalmente, dentro de este trabajo se presentan las referencias bibliográficas que se utilizó para el desarrollo de este trabajo de graduación y los anexos de algunos resultados obtenidos mediante el programa BIGSTEPS.

CAPÍTULO I: ANÁLISIS DE ÍTEMS

1.1 INTRODUCCIÓN

Desde el origen de la medición educativa se reconoce la influencia decisiva de la psicometría[1]. Esta surge, a partir del desarrollo progresivo de un conjunto de métodos y técnicas implicadas en la medición de las variables psicológicas. De la cual podemos reconocer cuatro grandes áreas temáticas en la psicometría:

1) Teoría de la medida, que tiene como objetivo establecer las condiciones y propiedades de las asignaciones numéricas que pueden realizarse.
2) Teoría de los tests, que aborda la lógica y los modelos matemáticos subyacentes en la construcción y uso de los tests.
3) Técnicas multivariadas, que junto con la tecnología estadística resultan imprescindibles para la construcción y análisis de los instrumentos de medida.

De estos aspectos resaltamos a continuación los relativos a las teorías de la medida y su impacto en el desarrollo de modalidades de tests y técnicas de prueba progresivamente mejor adaptadas a sus funciones diagnósticas como instrumentos de medida educativa. El punto de partida no puede ser otro que reconocer la dificultad de este tipo de medida, como paso previo a su abordaje desde los diversos modos de aproximación teórico-conceptual a los tests. Concretamente, el diseño de instrumentos para la medida de tests educativos y psicológicos presenta algunos problemas:

- *No existe una única aproximación a la medida de un test* que sea universalmente aceptada. Existe siempre la posibilidad de que dos teóricos, seleccionen diferentes tipos de conducta para la definición operativa de un mismo test.
- *Las medidas educativas y psicológicas están basadas en muestras limitadas de conductas*. Delimitar el número de elementos y la variedad de expresiones en que se manifiesta un test, es uno de los principales problemas en el desarrollo de un instrumento de medida.
- *La medida obtenida siempre tiene error.* Al estar basada en muestras limitadas en contenidos y recogidas en un momento en el tiempo, está afectada por todos los errores del muestreo.
- *Falta de escalas con origen y unidades de medida bien definidas*.

[1] La psicometría se ocupa de los problemas de medición en psicología utilizando la estadística como pilar básico para la elaboración de teorías y para el desarrollo de métodos y técnicas específicas de medición.

- *Dificultad de identificar un test con total independencia así como sus definiciones operativas.*

La teoría de los tests aborda estos problemas y plantea métodos y alternativas para su solución. Las distintas teorías y sus desarrollos fundamentan avances en la construcción y validación de instrumentos de medida concretos. Además esta teoría puede verse como un esquema conceptual formalizado para hacer inferencias a partir de las puntuaciones de las personas en los tests y permitir la toma de decisiones que impliquen el constructo objeto de la medición. La teoría de los tests proporciona los fundamentos para la elaboración y uso de los tests, aportando además las herramientas para examinar las propiedades métricas de las mediciones obtenidas con ellos.

A principio del siglo XX, el principal reto en el campo de la investigación pedagógica era crear tests propiamente escolares y elaborar escalas de comprobación de conocimientos. "La línea iniciada por Rice en 1910 origina la creación de escalas de redacción, de ortografía y de cálculo aritmético". Los primeros trabajos empíricos en el campo educativo corresponden a test de instrucción, inteligencia y escalas métricas de escritura; también se puede decir que estas preocupaciones surgen estrechamente ligadas al nacimiento de la Pedagogía experimental como disciplina científica. Este período de la medición educativa se caracteriza por su énfasis en la medición de las personas. Los instrumentos típicos de esta orientación son *tests psicométricos de aplicación individual*.

Entre los años 30 y 40 se construyen y afinan las hoy clásicas *baterías de aptitudes*, que conforman lo que entonces se consideraban como "componentes fundamentales del funcionamiento inteligente", referido a aspectos como: comprensión verbal, fluidez verbal, aptitud numérica, aptitud espacial, rapidez perceptiva y razonamiento general. La técnica del análisis factorial permite estructurar multidimensionalmente la genérica puntuación global de inteligencia.

Paralelamente al desarrollo de Tests Cognoscitivos, los *Test de Personalidad* también se benefician de los avances técnicos del análisis factorial y otras técnicas multivariadas afines. A estos avances se añaden nuevas orientaciones como los *Tests Proyectivos*. Estos test proyectivos se caracterizan porque el sujeto examinado proyecta inconcientemente en las respuestas, sus propios rasgos de personalidad (afectos, pulsiones, conflictos pasados y presentes,…).

El desarrollo de la *teoría de los tests* va de la mano, con la evolución de tests concretos que van surgiendo. De hecho, ambos desarrollos se influyen mutuamente e incluso a veces los tests como instrumentos se han anticipado a su fundamentación teórica. Como hemos señalado con anterioridad, la teoría de los tests constituye un área que provee de enfoques y modelos para el análisis y construcción de instrumentos de medida, identificándose tres orientaciones básicas: la teoría clásica de los tests (TCT), teoría de la generalizabilidad (TGT) y teoría de respuesta al ítem (TRI). Estas tres orientaciones aportan modelos diferentes para la construcción y validación de tests, aunque todas siguen vigentes, tienen un recorrido progresivo y secuencial.

La Teoría Clásica de los Tests (TCT) ha sido el modelo dominante durante gran parte del siglo XX y aún hoy tiene una vigencia notable en las aplicaciones prácticas. Esta teoría surge con el objetivo central de desarrollar un modelo estadístico que fundamente las puntuaciones de los tests y permita la estimación de los errores de medida asociados al proceso de medición. El nacimiento de la TCT se sitúa en los primeros trabajos de Spearman. No obstante, las obras de Guilford (1936) y Gulliksen (1950), suponen la integración y sistematización del cuerpo de conocimientos de esta teoría, incluyendo temas sobre validez, fiabilidad, análisis de ítems, etc. El modelo que asume esta teoría plantea que la puntuación empírica del sujeto en un test consta de dos componentes: puntuación verdadera y error de medida. A partir de estos conceptos y de los supuestos asociados se estima el grado de fiabilidad y validez de los instrumentos de medida. A esta teoría se debe el desarrollo de múltiples técnicas estadísticas multivariadas y métodos para el cálculo de la fiabilidad de un instrumento: análisis factorial, componentes principales, formas paralelas, test-retest, las dos mitades, etc. En este marco, también se producen avances fundamentales respecto a la noción de validez, tipos y procedimientos para garantizarla.

Las principales dificultades que presenta la TCT se refieren a que la medición de un sujeto en una variable depende del test utilizado y del grupo normativo. En primer lugar, la dependencia de la medida en función del instrumento utilizado plantea problemas para tratar de establecer la equivalencia entre las puntuaciones de dos tests distintos que midan una misma variable. En segundo lugar, las propiedades del instrumento de medida (ítems y test) están en función de los sujetos a los que se aplica. Los índices de dificultad, discriminación y consistencia interna dependen de los grupos de sujetos y su tamaño en el cálculo.

La teoría de la generalizabilidad de los tests (TGT) se considera una extensión de la teoría clásica de los tests, en tanto que reconoce el mismo modelo relativo a la existencia de errores junto a la puntuación verdadera, pero intenta resolver algunos de estos problemas mediante la aplicación del análisis de varianza para analizar dichas fuentes de error de medida. Esta teoría se plantea como una alternativa en la forma de conceptualizar la fiabilidad de los instrumentos de medida y de su varianza de error, reconociendo múltiples fuentes de error y posibilitando su integración en una estructura global. Fue desarrollada por L.J. Cronbach y colaboradores. Esta teoría introduce una serie de modificaciones importantes a la TCT (muestreo de fuentes de error, medidas aleatoriamente paralelas, puntuación del universo, generalizabilidad, etc.)

Igualmente, aunque los orígenes de la *teoría de respuesta al ítem* (TRI) se sitúan en los trabajos pioneros de Richardson, Lawley, Tucker, Lord y Birnbaum de los años cincuenta, su expansión se produce a partir de los años sesenta con la aparición del libro de Rasch (1960) y las contribuciones de Birnbaum en el libro de Lord y Novick (1968). Sin embargo, el desarrollo y expansión de los ordenadores durante la década de los 80 constituye un factor decisivo para la materialización de esta teoría en modelos y aplicaciones concretas.

Al igual que la teoría de la generalizabilidad, la TRI trata de resolver los problemas de la teoría clásica: proporcionar mediciones invariantes respecto de los instrumentos utilizados y disponer de instrumentos de medida cuyas propiedades no dependan de los objetos medidos. Esta teoría aporta todo un conjunto de avances técnicos para la construcción y análisis de los tests: las funciones de información de los ítems y del test, errores típicos de medida distintos para cada nivel de la variable medida o el establecimiento de bancos de ítems con parámetros estrictamente definidos, lo que posibilita el uso de tests adaptados al nivel de los sujetos con exploraciones más exhaustivas y rigurosas en función de las características de éstos. La TRI revitalizará áreas tales como los bancos de ítems o los tests referidos al criterio.

En la exposición de este breve recorrido sobre la medición, los años sesenta destacan como un período de gran actividad teórica en el marco de la medida y las teorías de los tests. El impacto de estos nuevos enfoques sobre la medida y, en concreto, sobre las teorías de los tests fue significativo, sobre todo por la aparición de todo un conjunto de nuevas temáticas (clima del centro, ambiente, organización social, resolución colectiva de problemas, ...) para evaluar mediante pruebas y otras técnicas de escalamiento, también retos para los tests (individuales y colectivos); todavía hoy es así en nuevos campos educativos.

La noción de medida también va cambiando; de entender la medida de forma estable, cuyo objetivo es la medición de la posición de una persona o evento en un momento preciso, *se ha pasado a una concepción evolutiva*, en la que el objetivo es registrar los cambios o progresos a través del tiempo. Los objetivos educativos a los que sirven también han cambiado: junto a usos estandarizados de la medida a nivel institucional o individual (admisión de alumnos, valorar el nivel de logros, evaluar programas, centros o instituciones, etc.) surgen otras aplicaciones tales como diagnosticar problemas de forma inmediata o establecer perfiles y trayectorias de aprendizaje.

Otro de los factores, que es necesario destacar en el recorrido diacrónico por la teoría y práctica de las medidas basadas en técnicas de prueba o tests, es *el auge de las tecnologías y su aplicación en la medición*. La incorporación del ordenador en el desarrollo científico en general constituye uno de los avances significativos del siglo XX y su impacto en la medición también es considerable; especialmente la tecnología informática facilita el desarrollo de modelos matemáticos y técnicas estadísticas multivariadas antes prácticamente inviables. En particular, el desarrollo más espectacular (además de todos los campos de la medida hasta ahora tratados) trata sobre la aplicación de ordenadores en el desarrollo exponencial de la TRI, en sus posibilidades reales de aprovechamiento social; posibilitando, los *tests adaptativos informatizados (TAIs)*, el perfeccionamiento de instrumentos en general e incluso interpretación y elaboración de informes. Sus posibilidades son enormes no sólo en la *administración del test*, sino también en la construcción y *selección de ítems* y en la calidad y precisión de la información. Su origen no es tan nuevo: pero dada la complejidad de los cálculos para su aplicación solo empezó a difundirse y utilizarse gracias a programas de computación específicos como BIGSTEP, LOGIST, BILOG, entre otros.

1.2 LA TEORÍA CLÁSICA DE LOS TESTS (TCT)

1.2.1 ASPECTOS BÁSICOS Y FUNDAMENTOS

La teoría clásica de los tests, también denominada como teoría basada en el modelo de la medida verdadera, centra sus aportaciones más significativas en la estimación de indicadores de los aspectos de fiabilidad y validez, así como de sus relaciones. En este modelo, cuando administramos un tests a un sujeto, la medida observada representa su habilidad en una muestra particular de ítems a resolver en una situación particular, bajo condiciones particulares. Algunos factores pueden afectar al desarrollo del sujeto, diferentes condiciones, diferentes tiempos, etc. Si fuese posible administrar múltiples veces el test al mismo sujeto obteniendo las diferentes medidas, bajo todas las condiciones, tiempos y modalidades de ítems posibles, la media de todas las puntuaciones así observadas supone (teóricamente) una estimación insesgada de la habilidad o nivel de rasgo que se está midiendo. Dicha media se define como medida verdadera. Sin embargo, esta teoría asume que toda medida basada en una muestra particular de ítems está afectada por un error (error al azar de medida) que la hace ligeramente diferente de la medida verdadera; lo que matemáticamente se puede expresar mediante un modelo lineal para la formalización de las puntuaciones:

$$X_i\ (puntuación\ observada) = V_i\ (puntuación\ verdadera) + e_i\ (error\ de\ medida) \qquad (1.1)$$

Esta expresión indica que la puntuación empírica directa de una persona en un test (X) está compuesta de dos componentes hipotéticos: el nivel del rasgo o puntuación verdadera de la persona (V) y el error de medida (e_i) que se comete al medir el rasgo con el test. El error de medida se considera una variable aleatoria compuesta por diferentes factores (propios del sujeto, del test y externos a ambos) que hacen que su puntuación empírica no sea exactamente su nivel de rasgo. Por tanto, el error de medida se establece como la diferencia entre la puntuación empírica y la verdadera:

$$e_i = X_i - V_i \qquad (1.2)$$

En definitiva, mientras que para una persona la puntuación verdadera (su nivel de habilidad o nivel de rasgo medido, no observable) se asume invariante o constante, la puntuación observada y los errores son variables aleatorias con varianza distinta de cero.

Este modelo permite considerar las esperanzas matemáticas *E* de ("promedio de" o "valor esperado de") cada término de la ecuación para el caso en que se ha desarrollado una muestra amplia de mediciones repetidas, donde:

$$E(X) = E(V) + E(e) \tag{1.3}$$

En los supuestos del modelo se asume que el error de medida es una desviación no sistemática o aleatoria de la puntuación verdadera, por tanto, en una muestra amplia de medidas la esperanza matemática del error tiende a ser cero, si la muestra es infinita $E(e) = 0$. Lo que implica $E(X) = E(V)$; siendo, por definición, la esperanza matemática del valor observado $E(X)$ la verdadera medida.

En resumen, estos supuestos implican que:

a) los sujetos tienen un nivel de rasgo o puntuación verdadera que, si no hubiera error de medida, coincidiría con la puntuación obtenida;

b) al existir un error aleatorio en las puntuaciones obtenidas por un sujeto en los tests, estas puntuaciones difieren de forma aleatoria de las puntuaciones verdaderas; y,

c) si se aplica varias veces, o formas equivalentes, el test a un sujeto la puntuación media resultante se aproximaría mucho a su nivel o puntuación verdadera en la habilidad o rasgo medido. Toda esta teoría supone un marco idóneo para el desarrollo y progresivo afinamiento de los conceptos e indicadores de fiabilidad y validez que se abordan posteriormente. En el marco de la teoría clásica, estos indicadores de fiabilidad y validez (operativizados mediante coeficientes de correlación) suponen una guía de conocimiento sistemático para la construcción, selección y mejora de pruebas y el desarrollo de inferencias desde las puntuaciones obtenidas con las mismas.

La siguiente tabla muestra un resumen sobre la formulación del modelo clásico y los supuestos matemáticos básicos que lo sustentan:

Tabla 1.1: Formulación del Modelo Matemático de la TCT

Modelo:	**$X = V + e$;** la puntuación empírica observada se compone de la puntuación verdadera más el error de medida.
Supuestos:	**1. $V = E(x)$;** la puntuación verdadera es la esperanza matemática de la puntuación empírica observada; consecuentemente, $E(e) = 0$. **2. $\rho(v, e) = 0$;** no existe correlación entre las puntuaciones verdaderas (niveles de rasgo) y los errores de medida. Son aspectos independientes o no sistemáticamente asociados. **3. $\rho(e_j, e_k) = 0$;** si en una población conociéramos los errores de medida de cada individuo en dos test diferentes (*j* y *k*), dada su condición de aleatoriedad, la correlación entre ambas variables también sería nula.
Definición:	Dos tests, j y k, se denominan paralelos si la varianza de los errores es la misma en ambos $[\sigma^2(e_j) = \sigma^2(e_k)]$ y también lo son las puntuaciones verdaderas de los sujetos ($V_j = V_k$).

1.2.2 FIABILIDAD Y VALIDEZ DE LA MEDIDA EN EDUCACIÓN

Los conceptos de fiabilidad y validez son centrales en el problema de la medida. En este sentido, parece razonable abordar primeramente la fiabilidad y los procedimientos básicos para su estudio, principalmente desde la perspectiva clásica de la medición que genera los métodos fundamentales; posteriormente, abordamos la validez y los tipos de la misma que se generan desde diversos marcos conceptuales y de estudio metodológico.

1.2.2.1 *Concepto y Tipos de Fiabilidad: Procedimientos de Cálculo* e *Interpretación*

El origen del término *fiabilidad* (o confiabilidad) se atribuye a los artículos publicados a inicios del siglo XX por Spearman en torno a la idea de determinar de forma objetiva la medida de la inteligencia y los desarrollos correlacionales para establecer las relaciones de asociación entre variables. No obstante, parece que su contribución sólo supuso un punto de partida desde el cual otros investigadores van a establecer factores relacionados con la fiabilidad estableciendo métodos diversos para su estudio y estimación.

La idea básica y original de fiabilidad deriva de las consideraciones sobre la medida que establece la Teoría Clásica de los Tests, respecto de que una determinada puntuación empírica está compuesta de la verdadera puntuación y de un error de medida aleatorio cuya esperanza matemática es cero dadas múltiples repeticiones del proceso de medida. Además todos los temas que comprende la teoría clásica, la fiabilidad es sin duda el tema central en torno al cual se han estructurado todos los demás.

Como señalan diversos autores, en términos *absolutos* la fiabilidad puede definirse como la *proporción de la varianza total* de las puntuaciones obtenidas con un procedimiento *que es varianza verdadera*; esto es, igual que se descompone la puntuación observada en puntuación verdadera y error aleatorio de medida (X = V + e), cuya correlación es nula, también pueden descomponerse sus varianzas ($S_t^2 = S_v^2 + S_e^2$). Diversos procedimientos basados en el análisis de la varianza (ANOVA) permiten descomponer y estimar las varianzas residuales o de error y totales, permitiendo por tanto, expresar la fiabilidad en términos de varianza [$1 - (S_e^2 / S_t^2)$]. No obstante, debido a que ello implica hacer mención a un concepto teórico (la puntuación verdadera) no disponible, al menos en los campos de las ciencias sociales, de forma empírica separada del error de medida; muy diversos autores han trabajado con preferencia en los enfoques *relativos* de la fiabilidad, basados siempre en situaciones de paralelismo de series de datos (normalmente dos) y expresados mediante coeficientes de correlación. Operativamente la fiabilidad puede definirse como el coeficiente de correlación entre dos conjuntos de puntuaciones obtenidas independientemente, en formas paralelas de la prueba, para un mismo grupo.

En el campo de la educación nada es tan constante, ni los fenómenos se repiten más que aparentemente, por lo tanto, los investigadores han desarrollado otro conjunto de estrategias que nos hablan sobre la *fiabilidad relativa* de la medida desde diversos planos. Estos planos, dadas dos series de puntuaciones medidas en un mismo grupo de sujetos o una sola serie a descomponer o analizar internamente, son:

1) ***La estabilidad o constancia***. Si disponemos de las puntuaciones de n personas en un test y, después de transcurrido un tiempo, volvemos a medir a las mismas personas en el mismo test, cabe suponer que siendo el test altamente fiable, deberíamos obtener una correlación de Pearson elevada entre ambas mediciones. Dicha correlación entre la evaluación test y la evaluación retest (r_{XX}) se denomina *coeficiente de estabilidad,* e indicará tanta mayor

estabilidad de la prueba cuanto más cercano a uno sea. De esta forma si los niveles de rasgo de las personas no han variado a lo largo del tiempo transcurrido entre las dos aplicaciones, podemos decir que el test proporciona bastantes garantías respecto a la precisión con la que mide, dado que una persona concreta obtiene puntuaciones muy parecidas en las dos aplicaciones. En este caso el coeficiente de correlación entre los resultados del test y el retest sería próximo a uno.

2) ***Equivalencia*** de las puntuaciones o resultados obtenidos por los mismos individuos sobre la base de dos pruebas paralelas o instrumentos considerados equivalentes o intercambiables para la medida del mismo rasgo. En este caso, la correlación estima la fiabilidad denominada *coeficiente de equivalencia* dando lugar al *procedimiento de formas paralelas*, el que consiste en diseñar un test y una segunda versión del mismo, denominada forma paralela, que intenta evaluar o medir lo mismo que el test original pero con diferentes ítems, obteniendo medias y varianzas probabilísticamente similares. La correlación de Pearson entre las puntuaciones obtenidas en una misma muestra en dos formas paralelas se considera el coeficiente de equivalencia de cualquiera de ellas, e indicará el grado en que pueden considerarse equivalentes.

3) ***Consistencia interna*** o coherencia de las puntuaciones obtenidas en el marco de un mismo procedimiento de medida parte del supuesto de que todos los elementos de un procedimiento o los ítems de una prueba conducen a la medida de un mismo rasgo o porciones coherentes del mismo. Lo más usual es obtener la consistencia entre dos mitades del test (método de dos mitades) o entre tantas partes como elementos tenga la prueba (intercorrelación de elementos). Desde este punto de vista, una prueba unitaria puede ser aleatoriamente descompuesta en partes, normalmente dos, dando lugar al *procedimiento de las mitades*. Este procedimiento consiste en dividir el test en dos mitades equivalentes (normalmente una con los elementos pares y otra con los impares). Para cada sujeto se obtiene la puntuación directa en ambas mitades. Disponemos entonces de dos variables *(P e I)*, cuya correlación de Pearson r_{PI} indica su grado de relación. Su resolución implica considerar la fórmula del *procedimiento de Spearman-Brown* sobre el coeficiente de correlación de Pearson entre ambas mitades, dada la homogeneidad de varianzas. La razón de dividir el test en la mitad par y la impar es garantizar su equivalencia.

También puede llevarse el planteamiento al extremo y considerarse cada elemento de la prueba como un subconjunto de la misma, dando lugar el *método de intercorrelación de elementos*. Una vertiente de este método es considerar el coeficiente medio que se obtendría al dividir las pruebas en infinitos pares de mitades diferentes, dando lugar a los *procedimientos de Kuder y Richardson (KR-20 y KR-21)*. Otra vertiente de este método es el *procedimiento alfa de Cronbach* que debe interpretarse como un indicador del grado de covariación entre los ítems, el cual tendrá un valor alto (cercano a 1) cuando los ítems covaríen fuertemente entre si; asumirá valores cercanos a cero, si los ítems son linealmente independientes (si covarían de forma escasa). Matemáticamente, α puede asumir valores negativos.

Tabla 1.2: Tipos y procedimientos de fiabilidad.

<table>
<tr><th>CONCEPTO DE FIABILIDAD</th><th colspan="2">PROCEDIMIENTOS DE CÁLCULO</th></tr>
<tr><td>Estabilidad</td><td colspan="2">Test-retest</td></tr>
<tr><td>Equivalencia</td><td colspan="2">Formas paralelas</td></tr>
<tr><td rowspan="2">Consistencia interna</td><td>De las mitades</td><td>Spearman-Brown</td></tr>
<tr><td>Intercorrelación de elementos</td><td>Kuder y Richardson</td></tr>
</table>

1.2.2.2 Concepto y Tipos de Validez: Procedimientos de Estudio e Indicadores

El concepto de validez es muy importante en el marco de las técnicas e instrumentos de diagnóstico; tanto, como en cualquier otra actividad científica que va a implicar tomar decisiones para acción educativa. Hay que tener en cuenta que el concepto de fiabilidad, antes tratado, es una condición necesaria de la medida, pero no suficiente; en tanto, hemos de aportar pruebas sobre el valor y utilidad de las puntuaciones (o informaciones), cuya adecuación hemos de garantizar más allá del hecho de que las mismas sean precisas.

En un sentido básico y genérico, *la validez en la medida hace referencia sencillamente al grado en que un procedimiento de medida recoge, precisamente, aquello que pretende medir y no otras cosas*. Este concepto intuitivo es muy útil didácticamente y está muy extendido, por su simplicidad.

Diversos autores (Kerlinger, 1975; Messick, 1980; 1981; 1988; Angoff, 1988; Hernández Pina, 1993; Morales, 2000; entre otros) recogen la opinión de que el propio concepto de validez tiene muy diversas facetas y es en alguna medida controvertido y difuso en las fuentes bibliográficas.

En este sentido, los trabajos sobre el concepto de validez se ven impulsados en los años 40 por la publicación del trabajo de Jenkis (1946) titulado "Validity for what?" que la postulaba como criterio de calidad que debía apoyarse en una justificación empírica; al tiempo, que señalaba la dificultad de identificar criterios externos en el ámbito psicológico y pedagógico que la fundamenta.

En este clima aparecen un conjunto de nuevos trabajos y conceptos de validez: "validez factorial" (Guilford, 1946), "validez aparente" y "validez cruzada" (Mosier, 1947); Cronbach (1949) establece la distinción entre "validez lógica" y "validez empíricamente comprobada"; "validez intrínseca" Gullikson (1950); etc. Pese a la heterogeneidad de facetas de la validez que muestra este ambiente científico, existen conceptos que ya están claramente consolidados, lo que permite pronunciarse en el sentido de proponer *tres enfoques* que resumen las aportaciones básicas sobre el problema de la validez de los procedimientos de medida: de contenido, de criterio (concurrente y predictiva) y de constructo.

1) ***La validez de contenido***, referida a la adecuación de la medida a un cuerpo definido de contenido que consta en su dominio de un conjunto de temas y procesos. Sobre todo en pruebas de rendimiento (por ejemplo, pruebas de inteligencia, de aptitudes, etc….) y en pruebas de conocimientos (cuestionarios para evaluar el rendimiento en una materia escolar o en una especialidad temática concreta), tiene sentido justificar que el conjunto de ítems que forman el test conforman una muestra representativa del universo de contenidos que interesa evaluar. Un test de conocimientos de Química en octavo grado, por ejemplo, debería incluir cuestiones representativas de los diferentes núcleos de contenidos que oficialmente deben impartirse en ese nivel de estudios. Sería una prueba poco válida si incluye demasiadas cuestiones de unos temas y muy pocas de otros. Para justificar, aunque solo sea racionalmente, que un test posee validez de contenido, debe quedar bien definido el universo o dominio conductual de referencia: especificar claramente cuales son los contenidos de Química que debe conocer un alumno de octavo grado, cuales son las componentes que interesa considerar en un cuestionario de cultura general, que tipo de conocimientos y destrezas son las pertinentes para medir el nivel básico de inglés, etc. En definitiva nos referimos a explicitar claramente los objetivos de la evaluación y la importancia que se quiere dar a cada uno, lo que determinará la cantidad de cuestiones a incluir referidas a cada uno de esos objetivos. En definitiva, la validez de contenido es un tema particular del muestreo: si deseamos realizar inferencias sobre el rendimiento de las personas en una población de contenido determinada, el test debe incluir una muestra representativa de

dichos contenidos. El proceso de validación de contenidos es eminentemente lógico, si bien pueden utilizarse jueces expertos en el tema para valorar la congruencia entre los diversos ítems y los diversos objetivos. Existen procedimientos cuantitativos diversos para que cada experto valore el grado en que un ítem sirve para evaluar el objetivo al que corresponde. El procedimiento cuantitativo más sencillo sería el siguiente:

- Especificar los diversos objetivos que pretenden evaluar.
- Elaborar varios ítems para cada objetivo.
- Seleccionar una muestra de expertos en el contenido del test.
- Pedirles que, según su opinión, asignen cada ítem al objetivo que pretende medir.
- Seleccionar los ítems en los que los expertos manifiestan mayor acuerdo en sus clasificaciones.

En este proceso se alude de forma necesaria a la participación de expertos tanto en la definición rigurosa de los dominios como para la valoración de la relevancia y representatividad de los elementos considerados en el procedimiento de medida. Por lo tanto, la discusión de este tipo de validez tiene componentes básicamente conceptuales y teóricos más que empíricos.

2) ***La validez referida al criterio***, el criterio de validación (*Y*) es una medida diferente del test para reflejar el mismo rasgo u otro muy relacionado, de tal manera que si el test mide lo que se pretende, debería correlacionar de forma elevada con el criterio. Por ejemplo, un criterio para validar un test de unos trabajadores podrían valorar el grado de motivación de cada uno y utilizar estas valoraciones como el criterio de validación de un test de motivación laboral; el total de ventas en dólares que realizan los vendedores puede ser un buen criterio para validar un test de aptitud para la venta. Sobre todo cuando se pretende utilizar el test para pronosticar determinados criterios de rendimiento (por ejemplo, el rendimiento escolar en un nivel dado, el total de ventas que se van a conseguir, el aprovechamiento de un cursillo o la mejora en un proceso terapéutico) conviene que el test se relacione muy estrechamente con un criterio externo. Este criterio externo debe ser una medida fiable del rendimiento que se quiere pronosticar con el test: calificaciones escolares, total de ventas producidas en un determinado periodo, estimaciones de un terapeuta de las mejoras conseguidas por cada persona, etc. A la correlación entre las puntuaciones en el test *(X)* y en el criterio *(Y)* se le denomina *coeficiente de validez,* lo designamos como r_{XY} e indicará el grado en el que el test sirve para pronosticar con precisión el rendimiento en el criterio.

El coeficiente de validez es una correlación de Pearson y, por tanto, su interpretación más inmediata se fundamenta en el denominado *coeficiente de determinación***,** que es simplemente el cuadrado de la correlación y que indica la proporción de varianza del criterio que podemos pronosticar con el test. Así, un test con un coeficiente de validez de 0.5 indicará que un 25% ($(0.5)^2 = 0.25$ ó R cuadrado) de la variabilidad o diferencias individuales en el criterio, mientras que el 75% restante se debe a variables diferentes al test.

Algunos conceptos fundamentales de la regresión lineal simple son:

- Coeficiente de determinación, el cual puede expresarse de la siguiente manera:

$$r_{xy}^2 = \frac{S_{y'}^2}{S_y^2} = 1 - \frac{S_{y-y'}^2}{S_y^2} \qquad (1.4)$$

Donde S_y^2 es la varianza del criterio

$S_{y'}^2$ es la varianza de los pronósticos

$S_{y-y'}^2$ es la varianza de los errores de pronóstico.

- Ecuación de regresión de *Y* sobre *X* en la escala directa:

$$Y_i^{'} = (\bar{Y} - r_{xy}\frac{S_y}{S_x}\bar{X}) + r_{xy}\frac{S_y}{S_x}X_i \qquad (1.5)$$

- Mediante esta expresión podemos estimar la puntuación directa en el criterio de una determinada persona pero, como es conocido, esa estimación será tanto más precisa cuanto mayor sea la correlación entre test y criterio. Estadísticamente, resulta más apropiada una estimación por intervalos realizada con cierta probabilidad, para lo cual aplicaremos la siguiente expresión:

$$Y_i^{'} \pm Z_{1-\alpha/2} S_{y-y'} \qquad (1.6)$$

Donde $Z_{1-\alpha/2}$ es el valor Z, de la normal (0,1), asociado a la probabilidad establecida y $S_{y-y'}$ es el error típico de estimación.

Ejemplo: A una muestra de 15 estudiantes de la Escuela de Biología que cursan la materia de Matemática I, se les aplica un test con el objetivo de evaluar el conocimiento de las funciones trigonométricas. Al docente encargado se le solicita que haga una valoración (de 0 a 10 puntos) de la nota que él espera de cada estudiante de acuerdo al desempeño observado. Estas valoraciones hacen la función de criterio *(Y)*. Los resultados del conocimiento de las funciones trigonométricas y en el criterio fueron los siguientes:

Tabla 1.3: Resultados del Test.

Estudiante	X	Y
	nota obtenida	nota esperada
1	2	5
2	3.5	5
3	9.4	7.5
4	6	8
5	1.7	4
6	2	4
7	6.4	8
8	8.5	7.5
9	3.5	5
10	6	6
11	6.4	5
12	4.3	5
13	2	4
14	2.5	4
15	3	5
Media	4.48	5.53
Varianza	6.186	2.231

El coeficiente de validez del test es r_{XY} = 0.84, lo que significa que el test de las funciones trigonométricas explica un 71% (R cuadrado) de las diferencias en las valoraciones del docente sobre el desempeño de sus estudiantes.

Si queremos pronosticar puntualmente la puntuación en el criterio del estudiante número 15, aplicando la ecuación (1.5) de regresión obtenemos:

$$Y'_{15} = 4.78$$

Para realizar la estimación por intervalo para este mismo estudiante, con probabilidad 0.95, obtenemos de la tabla de la normal (0,1) el valor $Z_{1-\alpha/2} = 1.96$ y calculamos el error típico de estimación:

$$S_{y-y'} = S_y \sqrt{1-r_{xy}^2} = 1.49365324\sqrt{1-0.713} = 0.80$$

Y el intervalo será:

$$4.78 \pm (1.96)(0.80) \begin{cases} \nearrow 3.21 \\ \searrow 6.34 \end{cases}$$

Diremos entonces, con probabilidad de 0.95, que la puntuación del alumno 15 en el criterio se encontrará entre 3.21 y 6.34.

3) ***La validez de constructo***. Un constructo es un concepto elaborado por los teóricos de la psicología para explicar el comportamiento humano. Inteligencia fluida, extraversión, autoconcepto, asertividad, motivación intrínseca… son constructos que forman parte de teorías psicológicas y que precisan de indicadores observables para su estudio. En muchas ocasiones, estos indicadores son los ítems de un test, y debe comprobarse empíricamente que resultan adecuados para reflejar el constructo de referencia.

☑ *Estrategias para la validez de constructo.*

La validez de constructo incluye la planificación y ejecución de determinados estudios de investigación orientados a comprobar empíricamente que un test mide realmente el constructo o rasgo que pretendemos medir. Aunque los métodos a emplear son sin duda variados, así como las técnicas estadísticas para analizar los datos, podemos encontrar un común denominador a todos ellos, que se sintetiza en las siguientes fases:

- *Formular hipótesis relevantes* (extraídas de deducciones teóricas o del sentido común) en las que aparezca el constructo que pretendemos evaluar con el test. En definitiva, una hipótesis de trabajo consiste en poner en relación dos o más variables. Pues bien, una de esas variables ha de ser el constructo que pretendemos medir con el test. Imaginemos, por ejemplo, que un investigador esta interesado en validar una prueba de motivación intrínseca-extrínseca que ha construido. Por lo que formula la siguiente hipótesis:

  H_0: Las personas motivadas intrínsecamente (por el mero placer que les supone la ejecución de determinadas tareas) deben rendir mejor en actividades escolares que las personas motivadas por razones extrínsecas (deseos de alcanzar determinada nota o determinado refuerzo externo).

- *Efectuar en la práctica mediciones* oportunas de las variables o constructos involucrados en las hipótesis. La medición del constructo de interés se realizará con la prueba diseñada a tal efecto, que es la que pretendemos validar.

Continuando con el ejemplo, el investigador tienen que demostrar empíricamente que mide auténticamente el constructo motivacional que se pretende y podría proceder de la siguiente manera:

a) Aplicar el test a un grupo amplio de alumnos del nivel escolar apropiado.
b) Recoger información de cada alumno sobre su nivel intelectual, su calificación académica media en el último curso y las horas que dedica al estudio.
c) Formar dos grupos diferentes (A y B), de tal manera que ambos tengan un mismo nivel intelectual medio y que ocupen un número similar de horas en el estudio, pero que el grupo A tenga niveles altos de motivación intrínseca y el B niveles altos de motivación extrínseca.
d) Comparar el rendimiento académico de los dos grupos. Si la hipótesis de partida fuera cierta, el grupo A debería rendir significativamente más que el grupo B, con lo cual se aportaría información sobre la validez del test. Desde luego, si el test no midiera motivación, sería improbable que se verificase la hipótesis de trabajo.

- *Determinar si se verifican o no las hipótesis planteadas.* En el caso de que así sea, queda confirmado mediante una investigación que el test mide el constructo de interés ya que, de lo contrario, no habría razones lógicas para que se cumplieran las hipótesis formuladas. Si las hipótesis planteadas no se confirman, no podemos decir que el test no es válido, ya que puede ser debido a que las hipótesis no estaban planteadas de manera adecuada, lo cual exigiría una revisión de la teoría subyacente.

Esta idea de validez implica que la medida ha de ponerse a prueba y explorarse empíricamente como cualquier hipótesis de investigación. Se hace necesario partir de un *modelo teórico que especifique las relaciones teóricas*, entre el rasgo latente o constructo teóricamente considerado y los elementos e indicadores específicos, *que se deben satisfacer empíricamente*. En este sentido, un *constructo* hace referencia a un rasgo, atributo o cualidad no observable directamente, sino que es inferible a través de una teoría. Ello implica que la validez del test no puede expresarse empíricamente mediante indicadores básicos como un coeficiente de correlación simple sino que pertenece a la *estructura dimensional teórica del procedimiento o* *instrumento de medida*, lo que se explora más convenientemente mediante técnicas multivariantes complejas tales como el análisis de componentes principales, el análisis factorial, los modelos de escalamiento multidimensional, el análisis cluster de variables, etc.

☑ *Validez de constructo factorial*

El análisis factorial es una técnica estadística multivariante que sirve para estudiar los factores (dimensiones) que subyacen a las relaciones entre varias variables. Normalmente toma como datos de partida la matriz de correlaciones entre las *n* variables que interesa analizar. Como información final, proporciona una matriz de tamaño *n x p*, denominada matriz factorial rotada. Esta matriz contiene las *saturaciones* de cada variable en cada una de los *"p"* factores extraídas, y que son las correlaciones de Pearson entre cada variable y cada factor. Dicho análisis se realiza bajo dos objetivos:

1) Determinar cual es el número de factores que mide un test y descubrir cual es el significado de cada una.
2) Obtener la puntuación de cada sujeto en cada factor.

Normalmente, el número de factores que mide un test es mucho menor que el de ítems. Para descubrir su significado y darles sentido es necesario fijarse en las variables que saturan de forma elevada en cada dimensión. Cuando el investigador se enfrenta con la tarea de dar significado a una dimensión debe realizar un proceso inferencial para encontrar el nexo de unión entre variables que manifiestan correlaciones elevadas en dimensión. Además, los diferentes factores (dimensiones) extraídos no tienen la misma importancia. Cada uno explica una determinada cantidad de la varianza total de los ítems, que se expresa porcentualmente, y que indica la importancia de esa dimensión para dar cuenta de la covariación entre las variables. Si un factor explica un porcentaje elevado de la varianza total, eso es síntoma de que es una dimensión importante a la hora de descubrir las relaciones entre las variables originales.

Por ejemplo:

Un psicólogo ha elaborado una prueba de cinco ítems para evaluar la actitud hacia las nuevas tecnologías por parte de las personas mayores. Los ítems, que se responden en una escala de siete categorías ordenas (desde 1: "muy de acuerdo" hasta 7: "muy en desacuerdo") son los siguientes:

Ítem 1: El uso de teléfonos móviles puede hacerme la vida más fácil.

Ítem 2: Los aparatos modernos son demasiado caros.

Ítem 3: Me gustaría tener una agenda electrónica.

Ítem 4: Utilizaría más el teléfono móvil si fuera mas barato.

Ítem 5 Gracias a Internet podemos resolver muchos problemas.

Los 5 ítems se aplicaron a una muestra de 200 personas. La matriz de correlaciones entre ellos se sometió a un análisis factorial, obteniéndose los siguientes resultados. Esta matriz contiene las saturaciones, es decir, la correlación de cada ítem con cada uno de los factores que mide el test:

Tabla 1.4: Matriz de Saturaciones

Ítem	Factor 1	Factor 2
1	0.845	-0.126
2	-0.201	0.803
3	0.672	0.012
4	0.052	-0.615
5	0.713	-0.143
% de varianza total explicada	34%	21%

Hay dos factores fundamentales que explican las relaciones entre los 5 ítems. Supongamos que se triplican las puntuaciones en los ítems; la varianza total sería cinco, que es la suma de la varianza de cada ítem. El factor 1 explica un 34% de la varianza total, el factor 2 explica un 21% de la varianza total. En el factor 1 se obtienen saturaciones altas de los ítems 1, 3 y 5, que indican si la persona considera que las nuevas tecnologías pueden ser útiles para mejorar su calidad de vida. El ítem 2 tiene una saturación negativa porque manifiesta una actitud contraria hacia las nuevas tecnología. Por lo tanto, el factor 1 puede denominarse "*Actitud hacia las nuevas tecnologías como medio para mejorar la calidad de vida*". En el factor 2 se obtienen saturaciones elevadas (el valor absoluto) los ítems 2 y 4, mientras que el resto de saturaciones son cercanas a cero. El hecho de que el ítem 2 tenga una saturación positiva y el 4 negativa significa que las personas con puntuación alta en el factor 2 tienden a estar de acuerdo con el ítem 2 y en desacuerdo con el 4. Este segundo factor podría etiquetarse "*Sensibilidad hacia el gasto que supone utilizar las nuevas tecnologías*". Por lo tanto podemos decir que en la prueba que elaboró el psicólogo, se están midiendo principalmente dos factores los cuales explican un 55% de la varianza de los ítems.

La aplicación del análisis factorial aporta información sobre los factores que estamos midiendo con un determinado test, es decir, proporciona información sobre la validez de la prueba.

De esta manera Messick, señala: "*la validez de constructo es el concepto unificador de validez que integra consideraciones de contenido y de criterio, en un marco general para probar hipótesis racionales acerca de las relaciones teóricamente relevantes*" En este sentido esta perspectiva unificadora nos lleva a una importante conclusión: la validez de contenido, de criterio y de constructo no son diferentes tipos de validez, sino diferentes facetas o aspectos de la misma. Estas tres facetas no son sino caminos distintos para obtener las evidencias de la validez; se trata de aproximaciones a la misma, siendo único el concepto de validez. En este marco, la validez ya no es una característica de los instrumentos, ni se valida un test o un procedimiento, sino *las consecuencias que se derivan de la interpretación de las medidas y el uso que se hace de las mismas en un proceso de acumulación de evidencias* que puede llevarse por muy diversos caminos orientados por los objetivos del diagnóstico.

1.2.3 LA TEORÍA DE LA GENERALIZABILIDAD DE LOS TESTS (TGT): FUNDAMENTOS Y APORTACIONES

Cronbach, Gleser, Nanda & Rajaratnam (1972), han desarrollado la Teoría de la Generalizabilidad, asumiendo que hay otras fuentes de variación además de las diferencias individuales e integrando cada una de estas fuentes de variación en una estructura global, que permite aplicaciones particulares de la teoría estadística del muestreo.

La teoría de la Generalizabilidad de los tests se plantea como una alternativa para superar la concepción unívoca y aleatoria del error de medida propuesta en los supuestos de la teoría clásica, por lo que aporta un nuevo marco para el estudio de la fiabilidad y la investigación de las pruebas. En la TGT se conciben múltiples fuentes de error que pueden estimarse separadamente (así como las diferentes iteraciones); ello, en respuesta a las críticas que pueden realizarse al concepto de fiabilidad inherente a los supuestos de la teoría clásica, en la cual, las fuentes de error están indeterminadas. En definitiva, se trata de ampliar la formulación matemática del modelo clásico ($X_i = V_i + e_i$) para el caso en que es posible aislar múltiples fuentes de variación del error separadas de la variación verdadera:

$$X_i = V_i + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ . \\ . \\ . \\ e_n \end{pmatrix} \tag{1.7}$$

La teoría de la Generalizabilidad es como una extensión sobre la base del mismo modelo clásico de medida, pero sosteniendo que el error de medida puede analizarse, al menos en parte, diferenciando explícitamente distintos componentes del error, cada uno de ellos asociado con algún aspecto manipulable de una prueba: los contenidos, condiciones de aplicación, calificadores, etc. En dicha teoría la aplicación del análisis de varianza permite determinar los componentes de la varianza atribuibles a las diversas condiciones. Ello en la práctica permite integrar el estudio de las fuentes del error en una estructura explicativa global. Como consecuencia, esta teoría aporta una forma de pensar distinta sobre la fiabilidad de las medidas y de su varianza de error, así como sobre sus relaciones con el concepto de validez.

La consideración clásica de fiabilidad plantea problemas en la toma de decisiones sobre la calidad de un instrumento; es un problema de inferencia, en tanto que un alto coeficiente de fiabilidad global no determina el grado de generalización de las condiciones particulares de observación a un universo de condiciones. La teoría de la generalizabilidad considera las observaciones sobre un objeto de medida (habitualmente un sujeto) como una muestra de un universo de observaciones, todas las cuales proporcionan el mismo tipo de información acerca de dicho objeto de medida. El universo de todas las observaciones que son igualmente útiles y relevantes para quien haya de tomar decisiones, apoyadas en una determinada medida, constituye su universo de generalización. Las condiciones en que se realizan las observaciones en un universo de generalización varían, por ejemplo, pueden ser hechas por diferentes observadores y en diferentes ocasiones.

El objetivo de esta teoría es desglosar, en cualquier tipo de medición la variabilidad real de la variabilidad del error. La filosofía básica que subyace a la teoría de la generalizabilidad es que "un investigador se pregunta acerca de la precisión o fiabilidad de una medición dado que se desea generalizar de observaciones reales a cualquier tipo de observaciones a las que pertenezcan". Y el eje central de la teoría de la Generalizabilidad se encuentra en los componentes de varianza, dado que su magnitud nos aporta información sobre fuentes de error que están afectando una medición conductual. Dado el carácter abstracto, la teoría estadística puede aplicarse a cualquiera que fuese el objeto de estudio. Y es por ello que la Teoría de la Generalizabilidad constituye una teoría general de la medida.

Algunas ventajas que se pueden mencionar acerca de este enfoque con respecto a la TCT son:

1) Los supuestos en los que se basa son menos restrictivos que los de la teoría clásica, únicamente requiere el muestreo aleatorio de individuos y condiciones de medida.

2) Reconoce de forma explícita las fuentes de error de medida así como las interacciones entre ellas.

3) Sirve para desarrollar y optimizar diseños de medida en estudios posteriores. Es decir, mediante el estudio sistemático de las diferentes fuentes de error podemos desarrollar un diseño de medida que reduzca el error total en los estudios posteriores.

4) Permite estudios en muestras estratificadas.

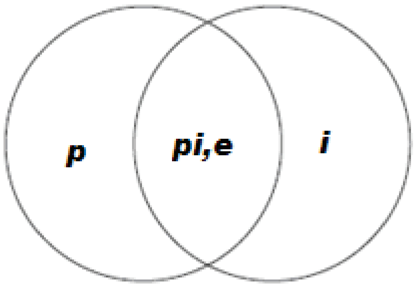
Entre las aportaciones más significativas de esta teoría de los tests consta la creación de un nuevo marco general para la evaluación y desarrollo de pruebas que implica una nueva terminología conceptual que supera los clásicos conceptos de puntuación verdadera y error.

En la TGT se utilizan otros términos, por ejemplo: el término *población* designa los objetos de medida y el de *universo de generalización* el conjunto de las condiciones implicadas en la generalización que quiere realizarse de la observación particular. Las condiciones de la misma clase se agrupan en *facetas* (variables), este concepto designa los factores que determinan fuentes de variación cada una de cuyas manifestaciones posibles se denomina *nivel de la faceta*. El diseño de una faceta tendría cuatro fuentes de variación o error. Por ejemplo, en un test de matemáticas, la primer fuente sería la diferencia en el rendimiento de los alumnos, objeto de la medición y que se llama la varianza verdadera; la segunda fuente de error sería la misma dificultad de los ítems; la tercera, la interacción de las diferencias de las personas con los ítems y la cuarta, las fuentes de error no identificados o aleatorios. Estos distintitos tipos de error se discriminan con el uso del ANOVA. El término *puntuación universo* representa la característica que se desea medir o estimar al aplicar un test a un sujeto. También se aporta un coeficiente de fiabilidad denominado *coeficiente de generalizabilidad*, que estima en que medida se puede generalizar, desde la media observada en unas condiciones, a la media de todas las observaciones posibles.

En la TGT se distinguen dos tipos de facetas: *facetas al azar* y *facetas fijas*, se piensa que el número de condiciones es infinito en una faceta al azar. Es decir, se asume que la prueba es una muestra al azar de un número muy grande de posibles condiciones. La combinación de una faceta al azar y una faceta fija nos da como resultado un estudio de generalizabilidad de *facetas mixtas*. Los cálculos de los coeficientes de generalizabilidad varían en estos tres tipos de estudio (para más detalles consultar: Dato N. M. de Gruijter & Leo J. Th. van der Kamp, october 2005, STATISTICAL TEST THEORY FOR EDUCATION AND PSYCHOLOGY). Pero en la mayoría de casos se asume por definición que se tiene un diseño con facetas aleatorias, ya que se considera a las personas y los ítems como una muestra al azar.

Supongamos que tenemos un diseño *p X i* (personas e ítems) de *una* sola faceta al azar, ítems. Podemos representar este diseño a través del diagrama de Venn que se muestra a continuación:

Figura 1.1: Diagrama de Venn para el Diseño *p X i*



Podemos observar que la interacción se encuentra en el segmento donde los círculos de ítems y personas se solapan. La ventaja de este tipo de diagrama es que se puede observar los componentes de varianza involucrados en descomposición de las puntuaciones observadas.

La variación de la puntuación universo o efectos de las personas (X_{pi}) se llama componente de varianza para las personas, denotado σ_p^2. También tenemos un componente de varianza para los ítems, denotado σ_i^2 y el componente varianza residual denotado por $\sigma_{pi,e}^2$. El efecto del residuo refleja la combinación del error aleatorio y la interacción. La variación de X_{pi} sobre *p* e *i* es:

$$\sigma^2(X_{pi}) = \sigma_p^2 + \sigma_i^2 + \sigma_{pi,e}^2 \qquad (1.8)$$

Como se ha mencionado anteriormente, los componentes de varianza pueden estimarse a través de una tabla ANOVA. Las observaciones de las puntuaciones del diseño pueden escribirse como se muestra en la figura 1.2

Figura 1.2: Representación de los Datos

PERSONAS	ÍTEMS					
	I_1	…	I_i	…	I_{ni}	
P_1	X_{11}	…	X_{1i}	…	X_{1ni}	$X_{1.}$
.	.		.		.	.
.	.		.		.	.
P_p	X_{p1}	…	X_{pi}	…	X_{pni}	$X_{p.}$
.	.		.		.	.
.	.		.		.	.
P_{np}	X_{np1}	…	X_{npi}	…	X_{npni}	X_{np}
	$X_{.1}$	…	$X_{.i}$	…	$X_{.ni}$	$X_{..}$

Cada dato de la tabla mostrado en la figura anterior representa la puntuación obtenida por cada persona en un determinado ítem. En la primera columna tenemos a las personas o sujetos a los cuales se les ha aplicado un determinado test y en la columna de la derecha tenemos los promedios para las personas sobre los ítems. En la primera fila tenemos los ítems que comprende el test y en la última fila tenemos las puntuaciones medias para los ítems promediados sobre las personas.

Para procesar los componentes de varianza para este diseño p X i usamos la tabla ANOVA. Se empieza por calcular las sumas de cuadrados para las personas e ítems respectivamente. La suma de cuadrados para las personas se obtiene restando las puntuaciones medias de cada persona de la media general ($x..$), estas elevadas al cuadrado. Luego se suman estas desviaciones cuadradas multiplicadas por el número de ítems. La expresión nos queda de la siguiente manera:

$$SS_p = \sum_{p=1}^{n_p} n_i (x_{p.} - x..)^2 \qquad (1.9)$$

Donde

$x_{p.}$: Puntuación media de cada persona

n_i : Número de ítems en el diseño

De forma similar la expresión para la suma de cuadrados para los ítems es:

$$SS_i = \sum_{i=1}^{n_i} n_p (x_{.i} - x..)^2 \qquad (1.10)$$

Donde:

$x_{.i}$: Es la puntuación media para cada ítem

n_p : Número personas en el diseño

En la suma de cuadrados totales, en vez de las puntuaciones medias, se reemplazará cada puntuación obtenida por cada persona en todos los ítems. La expresión para la suma de cuadrados totales es:

$$SS_T = \sum_{p=1}^{n_p}\sum_{i=1}^{n_i}(x_{pi} - x..)^2 \qquad (1.11)$$

Donde

x_{pi} : es la puntuación obtenida por un sujeto en un determinado ítem.

Para obtener la suma de cuadrados para los errores, resulta más sencillo restar de la suma de cuadrados totales, la suma de cuadrados de las personas y de los ítems. De la siguiente forma:

$$SS_E = SS_T - SS_p - SS_i \qquad (1.12)$$

Las medias de cuadrados para las personas son obtenidas con la suma de cuadrados para las personas entre los grados de libertad que le corresponde a esta suma, $n_p - 1$. La suma de cuadrados para los ítems se obtiene de forma similar. El ANOVA completo se resume en la siguiente figura donde observamos las sumas de cuadrados y las medias de cuadrados correspondientes a las personas, ítems y el error:

Figura 1.3: Tabla ANOVA para un Diseño de Una Faceta

Fuente de Variación	Suma de Cuadrados *SS*	Grados de Libertad *df*	Medias de Cuadrados *MS*	Esperanza *MS*: *EMS*
Personas (*p*)	SS_p	n_p-1	$MS_p=SS_p/df_p$	$\sigma^2_{pi,e} + n_i\sigma^2_p$
Ítems (*i*)	SS_i	n_i-1	$MS_i=SS_i/df_i$	$\sigma^2_{pi,e} + n_p\sigma^2_i$
Error	$SS_{pi,e}$	(n_p-1)(n_i-1)	$MS_{pi,e}=SS_{pi,e}/df_{pi,e}$	$\sigma^2_{pi,e}$
Total	$\sum_{p=1}^{n_p}\sum_{i=1}^{n_i}(X_{pi} - x..)^2$			

$MS_{pi,e}$ es una estimación de $\sigma^2_{pi,e}$. De la tabla anterior podemos obtener una estimación de σ^2_p:

$$\hat{\sigma}_p^2 = (MS_p - MS_{pi,e}) / n_{i.} \tag{1.13}$$

El coeficiente de generalizabilidad se expresa de la siguiente manera:

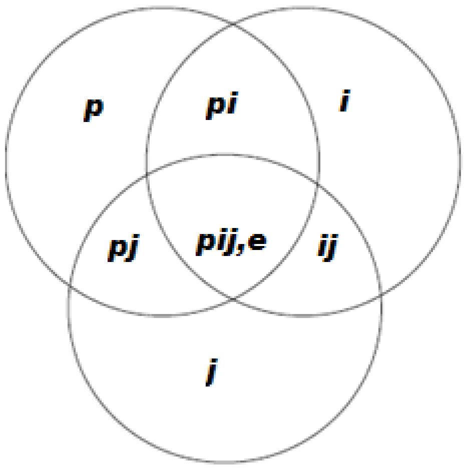
$$\rho_{\mathrm{Re}l}^2 = \frac{\sigma_p^2}{\sigma_p^2 + (\sigma_{pi,e}^2 / n_i)} \tag{1.14}$$

También conocido como correlación intra-clase. El tamaño de este coeficiente nos da información sobre la exactitud con la que pueden hacerse comparaciones entre las personas. Este coeficiente involucra medidas relativas por lo que es denotado por *(Rel)*. La estimación de la expresión anterior en términos de medias de cuadrados es:

$$\rho_{\mathrm{Re}l}^2 = \frac{MS_p - MS_{pi,e}}{MS_p} \tag{1.15}$$

Si consideramos un diseño de dos facetas (ítems y materias) siempre al azar, el diagrama de Venn donde se puede observar los componentes de varianza es:

Figura 1.4: Diagrama de Venn para el Diseño de Dos Facetas



Los círculos nos representan las personas *(p)*, los ítems *(i)* y las materias *(j)*, considerando como facetas a los ítems y las materias. Podemos observar que existen tres componentes principales, tres componentes para las interacciones y un componente para el error. Se observa que a medida que se aumenta el número de facetas van surgiendo nuevas interacciones y componentes de varianza que calcular. Las sumas y medias de cuadrados se calculan de forma similar al caso donde se tenía una sola faceta.

La tabla ANOVA para un diseño de dos facetas aleatorias será:

Figura 1.5: ANOVA para un Diseño de Dos Facetas

Fuente de Variación	*df*	*EMS*
Personas (*p*)	n_p-1	$\sigma^2_{pij,e} + n_i\sigma^2_{pj} + n_j\sigma^2_{pi} + n_in_j\sigma^2_p$
Ítems (*i*)	n_i-1	$\sigma^2_{pij,e} + n_p\sigma^2_{ij} + n_j\sigma^2_{pi} + n_pn_j\sigma^2_i$
Materia (*j*)	n_j-1	$\sigma^2_{pij,e} + n_p\sigma^2_{ij} + n_i\sigma^2_{pj} + n_pn_i\sigma^2_j$
Interacción *pi*	(n_p-1)(n_i-1)	$\sigma^2_{pij,e} + n_j\sigma^2_{pi}$
Interacción *pj*	(n_p-1) (n_j-1)	$\sigma^2_{pij,e} + n_i\sigma^2_{pj}$
Interacción *ij*	(n_i-1) (n_j-1)	$\sigma^2_{pij,e} + n_p\sigma^2_{ij}$
Error	(n_p-1) (n_i-1) (n_j-1)	$\sigma^2_{pij,e}$
Total		

El coeficiente de generalizabilidad para el diseño de dos facetas es una extensión del coeficiente de generalizabilidad para una faceta, en términos de las medias de cuadrados es:

$$\rho^2_{\mathrm{Re}l} = \frac{MS_p - MS_{pi} - MS_j + MS_{pij,e}}{MS_p} \tag{1.16}$$

A medida que se aumenta el número de facetas en un diseño de la TGT, tanto la tabla ANOVA como los coeficientes de generalizabilidad se hacen más complicadas sus estimaciones. Por tanto, se observa que el objetivo de dicha teoría es establecer múltiples fuentes de error los cuales quedan en la tabla ANOVA.

Algunos autores consideran a esta teoría como una ampliación de los diseños de experimento, debido a las similitudes que existen entre ellos; ya que, tanto en la Teoría de la Generalizabilidad como en los modelos de los Diseños de Experimentos parten de un modelo estadístico lineal donde se toma en cuenta la interacción, además de tener en común denominador al análisis de varianza, aunque ambos difieren en la interpretación de los resultados que corresponden a la estimación. Lo que en un principio pudiera pensarse que son semejanzas y disimilitudes no es otra cosa que la realidad ofrecida en la mayoría de los trabajos de investigación que utilizan una u otra técnica (una, los Diseños de Experimentos, utilizada mayoritariamente en estudios experimentales y la otra, Teoría de la Generalizabilidad, en estudios observacionales).

1.3 LA TEORÍA DE RESPUESTA A LOS ÍTEMS (TRI)

1.3.1 PRINCIPIOS

La Teoría del Rasgo Latente o Teoría de Respuesta a los Ítems (TRI) aparece como alternativa y nuevo modo de pensar los desarrollos y la calidad de los tests basados en las teorías clásicas. La cual constituye un nuevo enfoque que permite superar algunas limitaciones de la TCT. Esta teoría pretende subsanar las graves limitaciones que hacen que los resultados obtenidos de la aplicación de un ítem dependan de las circunstancias de aplicación, de las características del test y de los sujetos examinados. En este sentido se construye todo un marco teórico cuyos orígenes se encuentran en los trabajos de Richardson (1936); sin embargo, la creación de la TRI se atribuye a Lawley (1943), Tucker (1946), Lord (1952), Birnbaum (1958) y Rasch (1960). De modo independiente, Lord y Rasch desarrollan el modelo teórico de la teoría del rasgo latente; el primero plantea la teoría de la curva característica de los ítems, el segundo desarrolla el modelo logístico de respuesta a los ítems dicotómicos de un parámetro (“b” –índice de dificultad). Ambas aproximaciones parten del supuesto de que las puntuaciones obtenidas por un sujeto en un ítem de un tests dependen directamente del “nivel de rasgo” o grado en que posee la habilidad definida por dicho rasgo latente o constructo medido. Esto supone establecer una relación funcional entre el nivel de habilidad o nivel de rasgo de un individuo y su probabilidad de acertar el ítem.

La expansión de la TRI se produce fundamentalmente a partir de los años 60 en que se publica la obra de Rasch (1960) y, sobretodo, la obra de Lord y Novik (1968) en que se recogen las aportaciones de Birnbaum sobre fundamentación estadística de estos modelos. La progresiva extensión del uso de ordenadores y el diseño de programas que resuelven los complejos procesos estadísticos a que se someten los datos (mediante estudios de patrones multivariantes de respuestas a los ítems) hacen viable este enfoque, que llega a nuestros días claramente potenciado por las nuevas tecnologías. Desde un punto de vista teórico sobre la medición, la TRI supone el avance actual más significativo en la posibilidad de obtener medidas invariantes respecto de los instrumentos utilizados y de los objetos (sujetos) evaluados. Una diferencia claramente observable de la TRI respecto de anteriores aproximaciones es que centra su estudio sobre los ítems y sus propiedades paramétricas (discriminación, dificultad, pseudoazar,...) más que sobre las propiedades globales de los tests. Para varios autores como Muñiz, los modelos tienen que partir de supuestos más restrictivos, más fuertes, que los de la teoría clásica, por lo que a veces se denominan con el nombre genérico de *teoría fuerte de los tests*, frente a la *teoría clásica o débil*.

Otra diferencia principal de la TCT y la TRI es que la relación entre el valor esperado de la TCT es de tipo lineal, mientras que en la TRI las relaciones pueden ser funciones de tipo exponencial, tales como de Poisson, de la ojiva normal, el Modelo de Rasch o los modelos logísticos de *1, 2 ó 3 parámetros.*

1.3.2 ¿QUÉ ES UN MODELO DE LA TRI?

Se puede decir que un modelo TRI es una conceptualización, que partiendo de ciertos conceptos básicos de medición y usando las herramientas de la estadística y la matemática, busca encontrar una descripción teórica para explicar el comportamiento de datos empíricos derivados de la aplicación de un instrumento psicométrico. Los parámetros estimados por el modelo permiten entonces evaluar la calidad técnica de cada uno de los ítems por separado y del instrumento como un todo y a la vez estimar el nivel que cada examinado presenta en el constructo de interés. En un modelo de TRI se asume que hay una variable latente o constructo, no observable directamente y que se desea estimar para cada examinado a partir de las respuestas suministradas por éste en el instrumento de medición. Además se asume que para cada ítem o pregunta el comportamiento de las respuestas dadas por los examinados puede ser modelado mediante una función matemática que se denomina Curva Característica del Ítem o CCI. Otros conceptos fundamentales en TRI son la Función de Información y el Error Estándar de Medición, que serán explicados posteriormente y que son "Conceptos Fundamentales en TRI".

1.3.3 CONDICIONES DE APLICACIÓN

En cuanto a los requisitos que se deben cumplir para ajustar un modelo de TRI es necesario mencionar que debe contarse con los datos derivados de la aplicación del instrumento en muestras relativamente grandes de sujetos (como mínimo 300 personas). En general las respuestas a los ítems deben ser dicotomizadas, es decir clasificadas usando los códigos 0 ó 1. Esta codificación dicotómica resulta obvia cuando se están analizando los resultados de pruebas para medir conocimientos o habilidades intelectuales. Pero es también posible dicotomizar las respuestas a ítems que se contestan en una escala ordinal, tipo Likert, por ejemplo, asignando el 1 a las categorías que representen mayor intensidad o nivel en el constructo de interés, y el 0 a las que representen niveles más bajos.

1.3.4 POSTULADOS Y SUPUESTOS DE LA TRI

La fundamentación teórica que comprenden los modelos de la TRI están basados en los siguientes postulados y supuestos que se presentan a continuación:

Postulados básicos de la TRI:

a) El resultado de un examinado en un ítem puede ser explicado por un conjunto de factores llamados rasgos latentes o aptitudes que se simbolizan por θ.
b) La relación entre la respuesta de un sujeto a un ítem y el rasgo latente que subyace puede escribirse como una función monótona creciente que se llama función característica del ítem o curva característica del ítem (CCI). Esta función especifica que a medida que la aptitud aumenta la probabilidad de una respuesta correcta al ítem también aumenta.
c) Las estimaciones de la aptitud (θ) obtenidos con distintos ítem serían iguales y las estimaciones de los parámetros de los ítems obtenidos en distintas muestras de examinados serán iguales. Es decir que en la TRI los parámetros de aptitud y de los ítems son invariantes (los parámetros que caracterizan a un ítem no dependen de la distribución del rasgo latente de los encuestados. El parámetro que caracteriza a un encuestado no depende del conjunto de ítems del test).

Además de los postulados que se han mencionado anteriormente, podemos mencionar los supuestos de esta teoría:

1) *La unidimensionalidad del rasgo latente:* Los ítems que constituyen un test deben medir una sola aptitud o rasgo. El cumplimiento de este supuesto puede no ser exacto debido a que factores tanto cognoscitivos, como de personalidad, así como relacionados a la propia administración del test. Si existe un componente dominante que influya en las respuestas estaremos frente a un rasgo latente, en ese caso se cumple el supuesto de unidimensionalidad del modelo TRI. Caso contrario estaremos en presencia de más de un rasgo latente (multidimencional).

Este supuesto no se cumple totalmente porque el rendimiento en un test está influido por variables cognitivas, como la personalidad, motivación, ansiedad, etc. Por lo que, en la práctica, es cuestión de grado, y no puede afirmarse categóricamente si un conjunto de ítems son o no unidimensionales. Un método propuesto por Rectase (1979) se basa en aplicación del análisis factorial y consiste en estudiar la varianza explicada por el primer factor extraído de la matriz de correlaciones entre ítems. En la práctica, cuando el primer factor explica más de un 25% de la varianza total, tras haber eliminado los ítems con saturaciones inferiores a 0.10, se considera que se cumple el supuesto de unidimensionalidad.

2) *Independencia local:* Las respuestas de un examinado a cualquier par de ítems son independientes y no existe relación entre las respuestas de un examinado a diferentes ítems. Así, las aptitudes específicas en el modelo son los mismos factores que influyen sobre las respuestas a los ítems del test. De esta manera la probabilidad del tipo de respuesta a un conjunto de ítems es igual al producto de las probabilidades asociadas con las respuestas del examinado a los ítems individuales. Así, por ejemplo, si alguien tiene probabilidad de 0.5 de responder correctamente cada uno de dos ítems, la probabilidad de que el individuo responda correctamente a ambos ítems es de (0.5)(0.5) = 0.25.

Este supuesto se cumple usualmente si los ítems no están "encadenados" entre sí. Además no debe haber influencia de lo contestado por una persona sobre lo contestado por otra, lo cual se logra evitando la copia o plagio entre los sujetos al responder la prueba.

Cuando se cumplen estos supuestos con los datos que se están trabajando y encontramos un modelo apropiado, se obtienen una serie de características deseadas como:

- Las estimaciones del rasgo latente del encuestado no dependen del test.
- Las estimaciones de los parámetros de los ítems no depende de la muestra de encuestados utilizada.
- Los rasgos latentes estimados obtenidos de diferentes conjuntos de ítems serán las mismas (si no se consideran errores de medición).
- Los parámetros de los ítems estimados obtenidos de diferentes muestras serán los mismos (si no se consideran errores de medición).

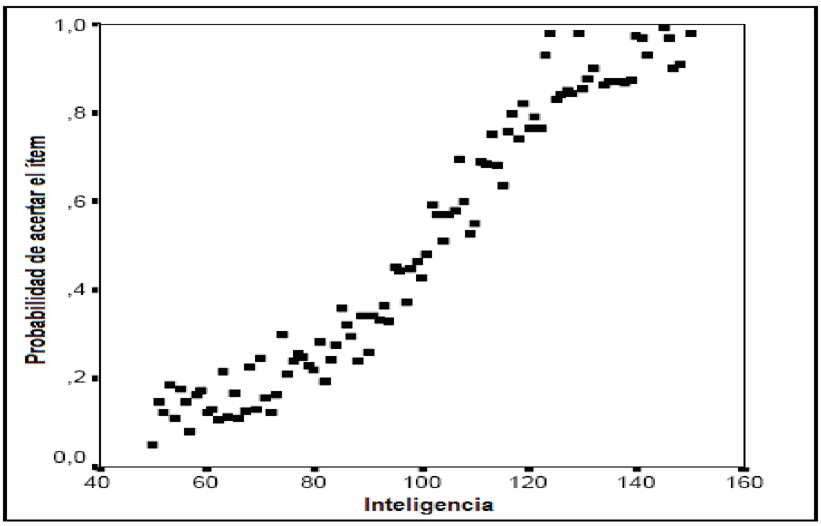
1.3.5 LA CURVA CARACTERÍSTICA DEL ÍTEM (CCI)

La Curva Característica de un ítem expresa la relación funcional entre el nivel de rasgo o habilidad que se mide *(θ)* y la probabilidad de responderlo correctamente $P(\theta)$, lo que supone una función relacional de las habilidades latentes (no observables) y las habilidades manifiestas (observadas en el ítem). Por tanto, estas curvas son el medio que disponemos para representar gráfica y matemáticamente dicha relación. Aunque Lord no la desarrolló en sus primero trabajos, se adopta más generalmente como CCI la función logística, por la simplicidad de los procedimientos matemáticos. La CCI viene definida, más comúnmente, por tres parámetros:

dificultad del ítem (b); discriminación (a); y, la probabilidad de acierto al azar, también denominada pseudoazar (c). No obstante, otras muchas funciones son posibles para definir la relación funcional que establece la CCI entre el rasgo latente y la habilidad manifiesta, dándose lugar a muy diversos modelos de la TRI. Los tres modelos más desarrollados son los modelos de $1(b)$, $2(b,a)$ y $3(b,a,c)$ con sus parámetros respectivamente.

La Curva Característica del Ítem nos indica que la probabilidad de acertar un ítem solo depende de los valores de la variable medida por el ítem, de modo que los sujetos con distinta puntuación en la variable tendrán distintas probabilidades de superar un determinado ítem. Consideremos el siguiente ejemplo: supongamos que tenemos un test que mide la inteligencia y que ha sido aplicado a muchísimas personas (100,000, por ejemplo). Supongamos que la menor y mayor puntuación obtenidas en el test son de 50 y 150. Vamos a representar el rendimiento en un ítem concreto en la siguiente forma: nos fijamos en todas las personas que han obtenido la puntuación de 50 (supongamos que son 132). Vemos cuantas personas de las anteriores han acertado el ítem (supongamos que han sido 5) y calculamos la proporción (5/132=0.04). Hacemos lo mismo con los que obtuvieron en el test 51 puntos (y obtenemos una proporción, supongamos que 0.15),…hasta llegar con las que obtuvieron 150 (la proporción fue de 0.99). La siguiente gráfica nos muestra la proporción de aciertos en el grupo de personas que obtuvo en el test 50 puntos, 51 puntos,…,150 puntos.

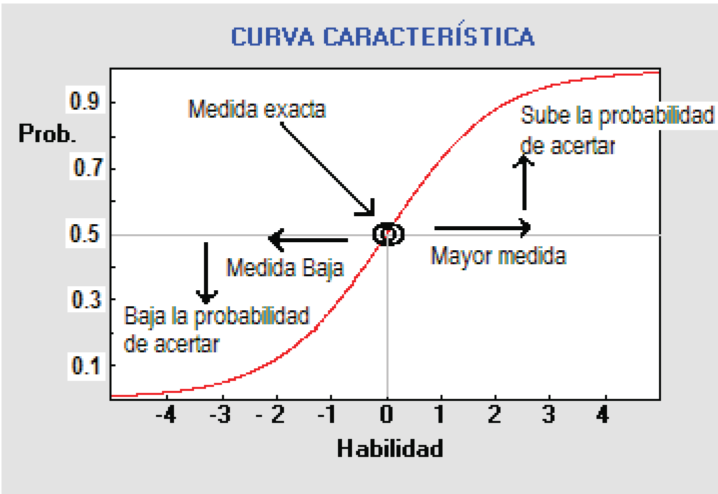
Figura 1.6: Curva Característica del Ítem Empírica



La escala que se presenta en el eje de las abscisas en la figura 1.6 corresponde con la puntuación obtenida por las personas. En este ejemplo podemos ver que cuanto mayor es el cociente intelectual de las personas, mayor es la proporción de aciertos en el ítem. A una puntuación de 100 le corresponde una proporción de 0.45; mientras que a una de 150 le corresponde una proporción de 0.99.

En la figura 1.6 tenemos una CCI empírica, pero en la TRI se necesita resumir la información que contiene cada CCI empírica en una fórmula o modelo en el que uno, dos o tres valores resuman la información contenida en la CCI empírica. Por lo tanto, un paso inexcusable es optar por un buen modelo que sea una buena descripción del rendimiento de los ítems. En la siguiente figura se presenta una Curva Característica del Ítem con algunas características:

Figura 1.7: Curva Característica del Ítem



La escala que se presenta en el eje de las abscisas corresponde a la habilidad de las personas en unidades de lógitos (unidad de medida que se desarrollará en el capítulo II). El punto donde el ítem tiene una medida exacta corresponde con el punto de inflexión de la Curva Característica del Ítem. El punto de inflexión identifica el lugar donde la curva cambia de concavidad: a la derecha del punto de inflexión la curva tiene concavidad hacia abajo, a la izquierda del punto de inflexión la concavidad es hacia arriba. En la curva que aparece en la figura 1.7 este punto de inflexión coincide con $P = 0.5$.

1.3.6 MODELOS PARA LA CURVA CARACTERÍSTICA DEL ÍTEM

Existen muchos modelos que se utilizan según el comportamiento de la Curva Característica del Ítem, sin embargo los más desarrollados y difundidos son:

- ✓ El Modelo Logístico de un Parámetro.
- ✓ Modelo Logístico de dos Parámetros.
- ✓ Modelo Logístico de tres Parámetros.

A continuación se presentan las principales características de cada uno.

1) Modelo Logístico de un Parámetro: a este modelo se le conoce como modelo de 1 parámetro de Thurstone, aunque tradicionalmente se le llama modelo de Rasch, pero existen ciertas diferencias con el verdadero modelo de Rasch, ya que se pueden destacar algunas diferencias entre ambos respecto a su motivación, sujetos, ítems, y evaluación del ajuste, entre otros aspectos (como se verá en el capítulo 2); debido a las disimilitudes que existe en ambos, el modelo de un parámetro se considera como un *modelo descriptivo*, y el modelo de Rasch que también es un modelo de un parámetro se considera un modelo preescriptivo (ya que hace un estudio más profundo acerca de la estructura de los datos).

Este modelo fue ampliamente desarrollado por Wright (1977) y Wright y Stone (1979) y es el más simple de todos. Este nos indica que la probabilidad de acertar un ítem depende solamente del nivel de dificultad de dicho ítem y el nivel del sujeto en la variable medida (nivel de rasgo o habilidad).

La expresión matemática es:

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}} = \frac{1}{1 + e^{-D(\theta - b_i)}} \quad ; i = 1, 2, 3, ..., n \qquad (1.17)$$

Donde:

$P_i(\theta)$: es la probabilidad de que un examinado elegido al azar con aptitud θ conteste correctamente el ítem i-ésimo.

$\theta =$ Nivel de habilidad del sujeto (Parámetro desconocido).

$b_i =$ Índice de dificultad del ítem i-ésimo.

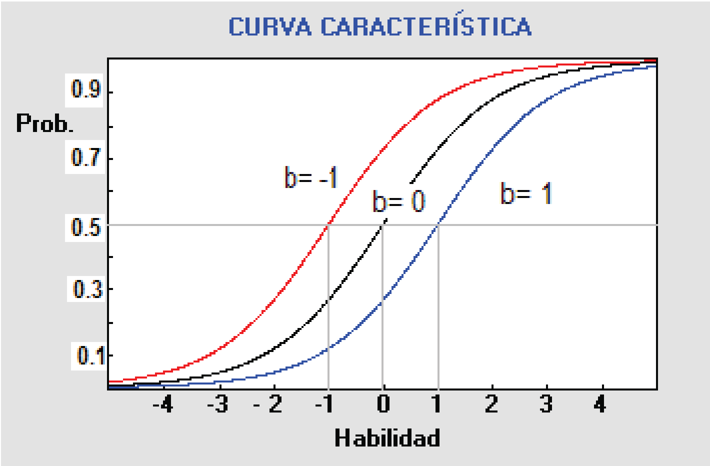
e = base de los logaritmos neperianos

D = constante (D=1.7 ó 1)

El parámetro (b) de dificultad del ítem es el punto en la escala de aptitud (θ) cuya probabilidad de respuesta correcta es 0.5. Por tanto, cuando más grande es el valor de (b), mayor es la aptitud requerida para que el examinado tenga una $P(\theta)=0.5$ de resolver correctamente el ítem.

Los ítems con $b=-3$ son muy fáciles, los ítems con $b=3$ son muy difíciles. En la figura 1.8 se representa el gráfico para 3 ítems con distinto nivel de dificultad y una constante $D=1.7$

Figura 1.8: Curvas Características de Ítems con Distinto Nivel de Dificultad (b)



La probabilidad que observamos en la figura 1.8 en el eje de la ordenada se calcula mediante la expresión (1.17), donde se toma en cuenta el nivel de dificultad del ítem y la habilidad de la persona (θ). El parámetro de dificultad del ítem (b) y la habilidad de la persona (θ) se estiman mediante el método de máxima verosimilitud (como se verá posteriormente en el capítulo II para el modelo de Rasch).

Como se observa en el gráfico anterior, el ítem más a la izquierda con $b=-1$ es el más fácil de los tres, mientras que el más difícil es el que tiene un índice de dificultad de $b=1$. En este modelo se presenta la característica de que las CCI son paralelas.

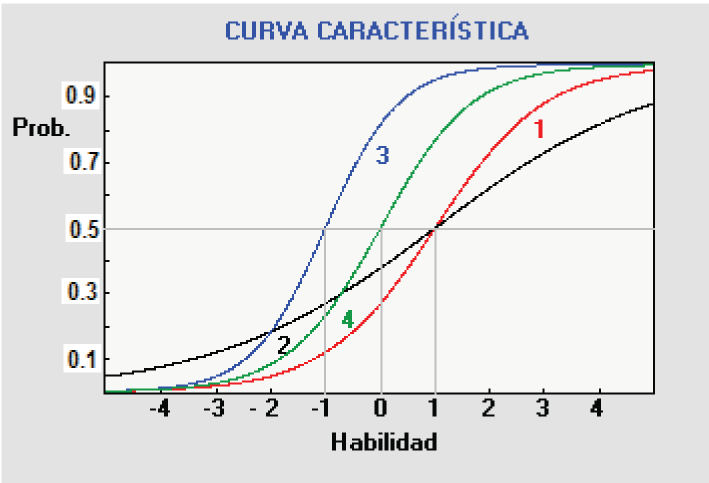
2) Modelo Logístico de dos Parámetros: Lord (1968,1980) fue el primero en elaborarlo, pero lo hizo basándose en una distribución normal. Actualmente este modelo es poco usado por la complicación matemática. En vez de este se sustituyó el modelo de dos parámetros de la ojiva normal por una función logística que tiene la ventaja de ser más conveniente para manejar. El modelo de la ojiva normal supone integración mientras que el modelo logístico no. Este modelo modificado esta dado por la siguiente ecuación:

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} = \frac{1}{1 + e^{-Da_i(\theta - b_i)}} \quad ; i = 1, 2, 3, \ldots, n \tag{1.18}$$

Aquí (b) es, igual al de la ecuación (1.17), el parámetro de posición o dificultad. El factor D=1.7 es un valor arbitrario introducido para que la función logística sea ajustada a la ojiva normal con exactitud de 0.01. Además hay un segundo parámetro (a) que es el de discriminación el cual es la *pendiente* de la curva característica del ítem en el punto (b). El parámetro (a) se encuentra generalmente entre 0.5 y 2. Los ítems con pendiente mayor son más útiles para separar los examinados en distintos niveles de aptitud, que los ítems de menor pendiente. Este modelo es una generalización del modelo de un parámetro. En la figura 1.9 podemos ver las CCI de cuatro ítems con las siguientes características:

Ítem 1; $b = 1 \quad a = 1$, Ítem 2; $b = 1 \quad a = 0.5$, Ítem 3; $b = -1 \quad a = 1.5$, Ítem 4; $b = 0 \quad a = 1.2$

Figura 1.9: CCI para el Modelo de Dos Parámetros



Como puede apreciarse en el gráfico anterior las CCI no son paralelas como en el modelo de Rasch sino que ciertos casos pueden cruzarse. Naturalmente los parámetros (a) y (b) pueden estimarse y también puede estimarse $P(\theta)$ pero es un proceso complicado hacerlo de forma manual, por lo que solo puede realizarse con programas de computación. Supongamos que tenemos un ítem (por ejemplo, ítem 55) para los que hemos obtenido (a) y (b) y queremos saber la probabilidad en distintos puntos para trazar la CCI. El proceso en este caso sería el siguiente:

Ítem 55 $a = 1.8$ $b = 1$

¿Cuál es la probabilidad del ítem en los valores de θ = -3, -2, -1, 0, 1, 2, 3?

Aplicando nuestros valores a la ecuación anterior, con θ=-3 tenemos:

$$P_i(\theta) = \frac{e^{1.7*1.8(-3-1)}}{1+e^{1.7*1.8(-3-1)}} = 0.9978$$

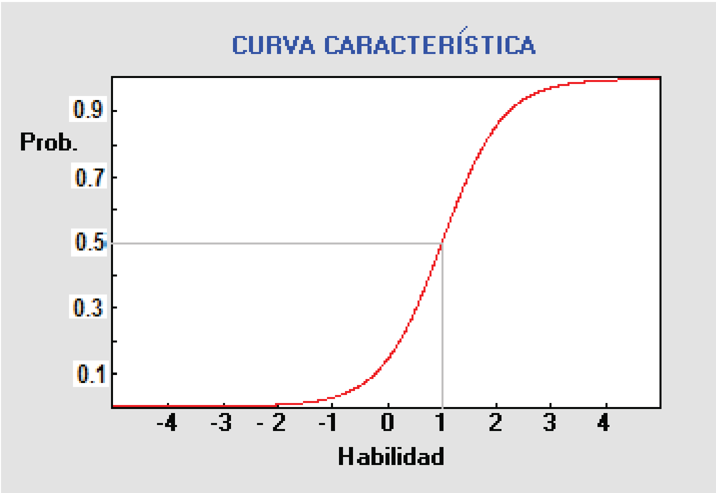
Repetimos esta operación para los distintos puntos de θ y podemos dibujar la curva característica del ítem como se muestra a continuación:

P (3)=0.9978 , *P (2)*=0.9950 , *P (1)*=0.500

P (0)=0.0450 , *P (-1)*=0.0020 , *P (-2)*=0.0000

P (-3)=0.0000

Figura 1.10: CCI para el Modelo de 2 Parámetros con $a = 1.8$ y, $b = 1$

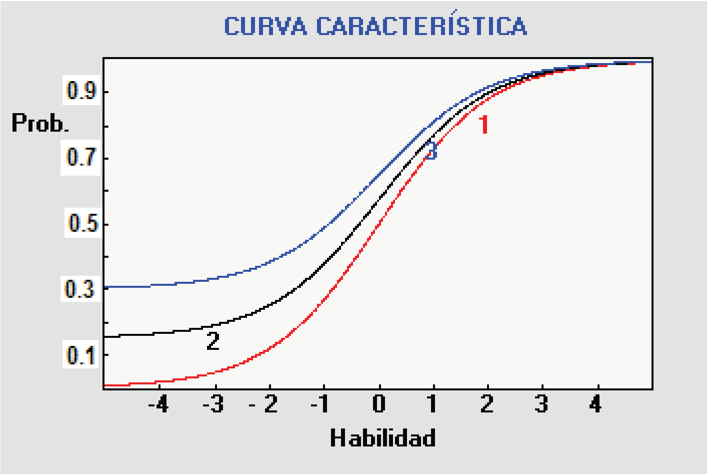


3) *Modelo Logístico de tres Parámetros:* este modelo es una generalización de los modelos anteriores. Este modelo añade el índice relativo a la probabilidad de acertar el ítem cuando el sujeto desconoce la respuesta correcta (pseudoazar). El parámetro (c) proporciona una posible asíntota no nula para la CCI y representa la probabilidad de respuesta correcta al ítem de los encuestados de bajo nivel de rasgo latente. Se incorpora para representar el comportamiento al más bajo nivel del continuo de la variable latente, en donde la "adivinación" o seudo azar puede ser factor significativo ante los test de respuestas múltiples. La expresión matemática es:

$$P_i(\theta) = c_i + \frac{(1-c_i)e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} = c_i + \frac{(1-c_i)}{1+e^{-Da_i(\theta-b_i)}} \tag{1.19}$$

En la gráfica 1.11 podemos observar la CCI de varios ítems con los mismos valores de $a=1$, $b=0$, $D=1.7$ y (c) con valores de 0, 0.15, 0.30.

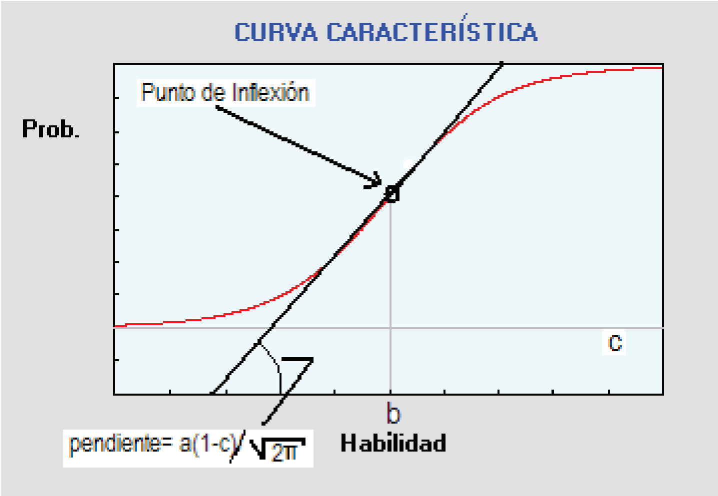
Figura 1.11: CCI para un Modelo de 3 Parámetros



Es claro que el parámetro (c) puede ser interesante para describir a un ítem; pero de ahí aceptar que el ítem es respondido por todos los alumnos con el mismo patrón de adivinación hay un gran trecho. Muchos autores consideran incorrecto suponer que (c) representa adivinación sistemática, ya que no es posible demostrar que todos los alumnos contestan con el mismo patrón de azar. No hay duda que en un grupo de personas que contestan un test puede haber quienes intenten contestar los ítems al azar, pero seguramente también habrá personas que contestarán muy pocas preguntas por adivinación y muchos más que responderán sin adivinar.

Hasta aquí hemos visto que la curva característica puede describirse por medio de tres parámetros: b (dificultad), a (discriminación) y c (Adivinación sistemática o pseudoazar). A continuación se hace la representación geométrica de estos parámetros y la relación que existe entre ellos.

Figura 1.12: Representación Geométrica de los Tres Parámetros de la CCI



El punto en el que se encuentra el valor de (b), corresponde con el *"punto de inflexión"*. Cuando (c) es igual a 0 el punto de inflexión coincide con *P=0.5*, en caso contrario el punto se encuentra al centro del intervalo entre *c* y *1*. Puede establecerse esta fórmula general:

$$\text{punto de inflexión: } (c+1)/2 \tag{1.19.1}$$

En la misma figura 1.12 se señala la relación entre la pendiente de la curva logística sobre el punto de inflexión y los parámetros *a* y *c*. A su vez, la pendiente se calcula con la tangente del ángulo que forma la *"parte recta"* de la curva logística y el eje horizontal.

$$a = \text{pendiente}\sqrt{2\pi}\,/(1-c) \tag{1.19.2}$$

Para el cálculo del parámetro (c) no hay un valor preciso para el cual debe identificarse. Un método manual para saber el valor de dicho parámetro es cortar la CCR por debajo de -3 lógitos a -4 lógitos (conceptos que serán estudiados en el capítulo II).

En la mayoría de situaciones el modelo logístico de 3 parámetros se ajusta a cualquier test, pero algunos evaluadores cuando sienten que dicho modelo no les llena sus expectativas, optan por buscar otros tipos de modelos. Un modelo alternativo es el *ajustar un polinomio de tercer grado* a los puntos, esto se debe a que una cúbica se parece mucho al comportamiento descrito por una curva característica del ítem.

Una razón mucho más fuerte y de origen matemático se debe a que toda función continua puede desarrollarse en serie de Taylor, lo cual quiere decir que dicha función se transforma a una función trascendente en otra de tipo algebraica.

Todos estos modelos sirven para aquellos test en los que se pueden considerar una respuesta correcta como 1 y la incorrecta como 0. Además de estos modelos, Barton y Lord (1981) describen el modelo logístico unidimensional dicotómico de cuatro parámetros, el cual se muestra a continuación:

$$P_i(\theta) = c_i + (g_i - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \tag{1.20}$$

Donde g_i representa el parámetro de descuido del ítem.

Este nuevo modelo tiene poca utilización en la práctica debido a su complicación matemática y a la dificultad para estimar los parámetros. Si bien los modelos dicotómicos básicos suponen una aproximación didáctica útil a los modelos de la TRI, esta teoría va hoy mucho más allá. Ya que existen otros modelos promisorios como el de las respuestas graduadas de Samejima (1973) y el de Bock (1972) para escalas nominales. El modelo de respuestas graduadas de Samejima prefiere hablar en sentido plural de las curvas características de las categorías de respuesta del ítem; pues, incluso en los ítems dicotómicos puede representarse, junto a la CCI que pone en relación θ con la $P(\theta)$, su CCI complementaria para $Q(\theta)$, que supone la curva para la probabilidad de respuesta incorrecta con idéntico índice paramétrico de dificultad y el mismo índice de discriminación con signo opuesto. No obstante, dadas las restricciones que la cuestión de la unidimensionalidad establece en el campo de la educación y la psicología humana, se están desarrollando muy diversos modelos TRI para el tratamiento de datos multidimensionales.

El modelo de escalas nominales de Bock (1972) plantea un modelo para ítems con respuesta politómica, el modelo de respuesta nominal, introduciendo una CCI para cada categoría de respuesta posible cuya condición básica es que la suma de las probabilidades sea 1; una variación de éste es el modelo de respuesta graduada de Samejima (1969). También se desarrollan modelos multidimensionales de TRI (Embretson, 1991).

Como se ha mencionado anteriormente la Teoría de Respuesta al Ítem se complica mucho para la estimación (o calibración) de los parámetros de los modelos, además de que se necesita contar con muestras grandes de sujetos (*n>300*) que hacen posible el ajuste de cualquier modelo de uno, dos o tres parámetros (a excepción del modelo de 1 parámetro que se puede ajustar también para muestras pequeñas).

1.4 TEORÍA DE RESPUESTA AL ÍTEM VRS. TEORÍA CLÁSICA DEL TEST

Un test no es un instrumento de medición como un metro, un termómetro o un velocímetro que proporcionan mediciones directas en una escala numérica. Un test debe considerarse más bien como una serie de pequeños experimentos, en que el examinador registra una serie de respuestas del examinado y estas respuestas no son mediciones directas sino que proporcionan los datos de los cuales se pueden inferir mediciones. Por lo tanto como en cualquier experimento, surge el problema de controlar el error experimental. Por esta razón cuando evaluamos en una determinada área a un grupo de sujetos es necesario obtener mediciones tanto de los sujetos como de los ítems que se les administra.

En un test siempre se propone establecer inferencias sobre los rasgos psicológicos de los sujetos (no observables) basándose en la información que manifiestan en las respuestas. La Teoría de Respuesta al Ítem como la Teoría Clásica de los Test consideran que cada individuo lleva asociado un parámetro individual, que en la TRI se le denomina aptitud, incluyendo cualquier rasgo psicológico, y en la TCT se denomina puntaje verdadero

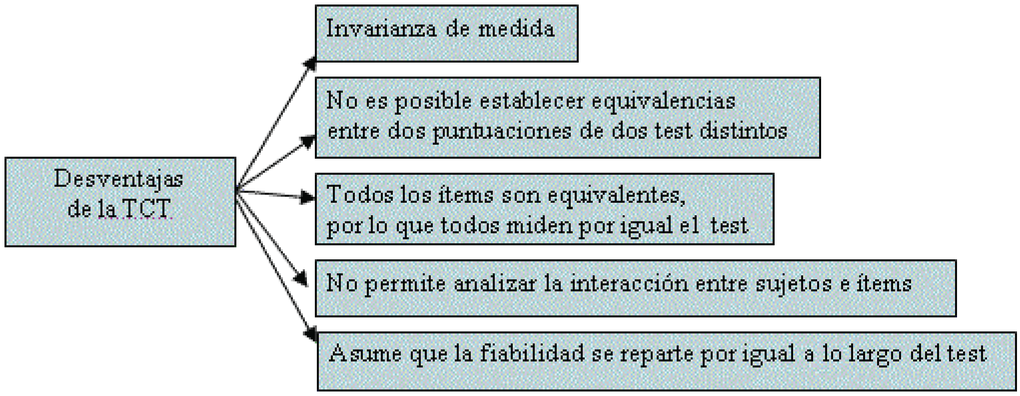
La Teoría de Respuesta al Ítem es uno de los campos con mayor proyección dentro del ámbito de la medida psicológica y educación. La TRI no contradice las asunciones fundamentales de la TCT sino que hace asunciones adicionales que permitirán responder a las cuestiones que la TCT no podía.

1.4.1 DESVENTAJAS DE LA TCT

En la TCT se produce la paradoja que un ítem es fácil o difícil según la aptitud de los examinados y la aptitud de los examinados depende de que los ítems del test sean fáciles o difíciles. Es decir que en la teoría clásica las características de los ítems son dependientes del grupo. Esto suele traer muchos problemas. Por ejemplo, consideremos dos examinados que contestan correctamente el 50% de las preguntas de dos test que difieren en dificultad. ¿Pueden ser considerados igualmente capaces?, claramente no. Además la teoría clásica esta orientada a todo el test y por lo tanto no nos permite pronosticar como responderá un individuo a un ítem en particular.

A pesar que la TCT surgió a comienzos del siglo pasado, aun en nuestros días es muy utilizada en los diversos ámbitos de la educación. Algunos inconvenientes principales de esta teoría se muestran en la siguiente figura:

Figura 1.13: Desventajas de la Teoría Clásica de los Test



La invarianza de la medida (las mediciones de un instrumento de medida deben ser independientes de los objetos medidos), este inconveniente se refleja en dos problemas concretos:

1) la medición de las variables psicológicas no es independiente del instrumento que se utiliza para medirla;

2) las propiedades de los instrumentos no son independendientes de los sujetos a los que se aplican.

Además, no es posible establecer equivalencias entre las puntuaciones de dos test diferentes que midan una misma variable, este es otro de los problemas de la TCT. La TCT parte del supuesto de considerar la mayor parte de los casos de un test como una muestra seleccionada de un universo de ítems equivalentes unos a otros, que permiten ser considerados indicadores similares del test que medimos, de ahí que se puede utilizar como procedimiento de acumulación de puntos, lo que nos lleva a otra limitación de esta teoría, ya que en una misma puntuación de un test puede deberse a distintos patrones de respuesta, y haciendo uso de la TCT no podemos analizar las interacciones entre los sujetos y los ítems, además el presuponer que todos los ítems son equivalentes implica que todos los sujetos utilizan las mismas operaciones mentales y para todos los ítems, el problema esta en que no se tienen en cuenta las diferencias individuales ni la diferencia de la dificultad de los ítems. Otra limitación que se encuentra en la teoría clásica del test se debe a la fiabilidad del instrumento de medida, según esta teoría, la fiabilidad se reparte por igual a lo largo del test pero desde otros modelos se ha podido comprobar que no se mide con la misma fiabilidad en los distintos niveles de la variable.

1.4.2 VENTAJAS DE LA TRI

Las limitaciones presentadas en la sección anterior impulsan el surgimiento de nuevos modelos, algunos de ellos no eran más que extensiones del modelo lineal de Spearman asumido en la TCT (como la Teoría de la Generalizabilidad) y otros surgen enmarcados dentro de un nuevo marco teórico, entre los que destaca la Teoría de Respuesta al Ítem (TRI), que permitirá solventar las limitaciones de la Teoría Clásica de los Test, aunque esta teoría no es reciente, su expansión se produce a partir de los años 80 con la difusión de los ordenadores, una herramienta que será imprescindible debido a la complejidad de los cálculos matemáticos.

La TRI tiene como objetivo obtener mediciones que no varíen en función del instrumento utilizado, disponer de instrumentos de medida que no dependen de los objetos medidos, es decir, que sean invariantes respecto a los sujetos evaluados y avances técnicos como funciones de información de los ítems y del test, errores típicos de medida diferentes para cada nivel de la variable medida.

Una ventaja principal de la TRI es su invarianza, en un doble sentido: invarianza de los ítems respecto a posibles diferentes distribuciones de la habilidad o del rasgo, e invarianza de la habilidad medida a partir de diferentes conjuntos de ítems. Haremos un breve comentario sobre cada tipo de invarianza. Si las condiciones de aplicación de la TRI se cumplen, ha de ocurrir lo siguiente:

- Sea cual sea la distribución de los niveles de rasgo obtendremos las mismas estimaciones de los parámetros de los ítems. Esta propiedad se cumple también en otros ámbitos. Por ejemplo, en Estadística, si se cumplen los supuestos de la regresión lineal, se llega a los mismos parámetros cuando se ajusta la recta de regresión a toda la población o sólo a parte de ella. Análogamente, los parámetros de los ítems deberán ser los mismos si éstos se han aplicado a un grupo de personas con alto nivel de rasgo, o a un grupo con niveles bajos. Es decir, los parámetros de los ítems serán los mismos sea cual sea la distribución de los niveles de habilidad de la muestra en los que se han aplicado.

- El nivel de habilidad de una persona puede ser obtenido a partir de conjuntos de ítems distintos. Algunas de las aplicaciones de la TRI descansan precisamente en esta propiedad.

Otras ventajas y propiedades que podemos mencionar son:

- Intentan establecer para cada ítem la probabilidad de ser contestado correctamente (CCI).
- La probabilidad de respuesta correcta depende de la habilidad del examinado, y de las características propias de las preguntas, tales como: dificultad, discriminación y probabilidad de ser acertada al azar por un sujeto de muy baja habilidad.
- La posibilidad de construir curvas de información para cada ítem lo que permite la optimización de la selección de las preguntas para una prueba con objetivos específicos.
- Mientras mayor es la información que proporciona una pregunta en un determinado nivel de habilidad, mejor es el grado de precisión con que se estima ese nivel de habilidad, esto permite construir pruebas muy ajustadas al propósito que se persigue.

Estos modelos de la TRI relacionan a sujetos e ítems de modo interactivo lo que permite localizar al mismo tiempo en un continuo psicológico que representa a la variable a sujetos e ítems, el proceso de medición se puede representar como la localización de personas e ítems en una línea recta. Así, la posición de las personas en la línea dependerá de sus respuestas a los ítems del test, del mismo modo los ítems tendrán distintas localizaciones dependiendo de su nivel de dificultad.

Aunque la Teoría de Respuesta al Ítem se muestra como un método capaz de enfrentarse a las deficiencias planteados de la TCT, estos modelos teóricamente se solapan para entender el funcionamiento del test más que competir entre ellos. Y debido a que en la TRI se hacen suposiciones más fuertes que en la TCT, los supuestos de dicha teoría no se cumplen muchas veces por la naturaleza de la disciplina a medir, por lo que hace aconsejable no utilizarla.

CAPÍTULO II: MODELOS DE RASCH

2.1 INTRODUCCIÓN

El uso de los test psicométricos se ha basado principalmente en la Teoría Clásica de los Test (TCT), un modelo simple, flexible y muy conocido, pero con muchas limitaciones. Esto motivó al matemático danés George Rasch (1901-1980) a desarrollar un método estadístico que permitiera la medida de los resultados educativos de los estudiantes, mediante el diseño racional de exámenes. Identificó una familia de modelos de medida probabilística que completó las perspectivas de la TCT. La característica estadística que distingue al modelo de Rasch de otros modelos de la TRI es que los parámetros de las personas e ítems están algebraicamente separados y dan lugar a estadísticos suficientes.

Rasch inició sus investigaciones con el estudio de la distribución de los errores (de lectura de los encuestados) en observaciones en Dinamarca tras la segunda guerra mundial. Si el test había sido bien elegido, surgirán pocos errores, por lo que era razonable representar la distribución de los errores de lectura por una función de probabilidad de Poisson (multiplicativa). La elección de este modelo de Poisson multiplicativo se debía a las propiedades que cualifican como un modelo de medida. El análisis de las propiedades que debían cumplir las medidas, le llevan a Rasch al uso de modelos aditivos exponenciales (modelos de medida): pasando de los modelos de Poisson multiplicativo al modelo logístico.

Rasch planteo su modelo dentro de la denominada teoría de la medición conjunta, que parte de la consideración de que las medidas (observables) fundamentales tienen estructura aditiva. Algunos matemáticos como Perline (1979) ha demostrado que los modelos pertenecientes a la familia identificada por Rasch son los únicos modelos compatibles con los principios de la medida conjunta en el caso probabilístico.

En el presente capítulo se abordan los siguientes temas:

1. *Los fundamentos básicos del modelo de Rasch*: donde se estudiarán las características principales, la escala y unidad de medida utilizada, el método del Escalograma de Guttman, la medida y ajuste de dicho modelo y algunos conceptos que son muy importantes como la estimación de los parámetros y función de información.

2. *El Modelo de Crédito Parcial*: se conocerá este modelo que también es conocido como el modelo politómico de Rasch por su generalizabilidad y además se estudiará la estimación de los parámetros de los sujetos e ítems.
3. Y por último, la importancia de usar el modelo de Rasch en comparación con la Teoría Clásica de los Test (TCT).

2.2 FUNDAMENTOS BÁSICOS DEL MODELO DE RASCH

2.2.1 EL MODELO DE RASCH

En 1960 el matemático danés George Rasch propuso un modelo de medida que permite solventar muchas de las deficiencias de la TCT y construir pruebas mucho más adecuadas y eficientes. El modelo de Rasch es un modelo uniparamétrico, es decir, un solo parámetro de medición ligado al número de respuestas. Este parámetro corresponde, de acuerdo con Rasch, con una sola dimensión relativa a una sola escala a medir tanto en el desempeño de la persona (habilidad, conocimientos, etc.) como en la calidad de un ítem (relacionada con la dificultad). Existe una ventaja aparente en usar una sola escala para los ítems y para las personas, porque ambos se pueden comparar directamente en dicha escala única.

De acuerdo con lo que puede esperarse de una buena medición, el instrumento de medida no debería verse influido por los objetos que mide. Como mencionó Thurstone: "Si la regla de medida brinda diferentes valores cuando se usa en una superficie rugosa, en una pintura o en una hoja de papel, entonces se puede dudar de la calidad de la regla. Dentro del rango de objetos para el cual se usa un instrumento de medida, su función debe ser independiente del objeto medido".

El modelo de Rasch trata de usar este razonamiento: supóngase las preguntas *i,j* que se van analizar y que se espera que sean independientes de los sujetos medidos. Si se comparan las probabilidades de aciertos y errores que tiene la persona *n* en estos ítems se tiene:

$$\frac{p(i=si, j=no)}{p(i=no, j=si)} = \left(\frac{p_{ni}}{q_{ni}}\right)\left(\frac{q_{nj}}{p_{nj}}\right) \tag{2.1}$$

Donde p_{ni} es la probabilidad de aciertos de la persona *n* en la pregunta *i*, q_{ni} es la probabilidad de falla en la misma pregunta. Ya sabemos que $q_{ni} = 1 - p_{ni}$

La expresión anterior debe ser válida independientemente de que personas se emplean para su cálculo. Hágase la misma fórmula para la persona *m*, entonces:

$$\frac{p(i = si, j = no)}{p(i = no, j = si)} = \left(\frac{p_{mi}}{q_{mi}}\right)\left(\frac{q_{mj}}{p_{mj}}\right) \tag{2.2}$$

Y, forzosamente, la expresión (2.1) debería ser igual a la expresión (2.2) porque se ha dicho que las medidas de los ítems son independientes de los sujetos. Queda por tanto:

$$\left(\frac{p_{ni}}{q_{ni}}\right)\left(\frac{q_{nj}}{p_{nj}}\right) = \left(\frac{p_{mi}}{q_{mi}}\right)\left(\frac{q_{mj}}{p_{mj}}\right) \tag{2.3}$$

Ahora establezcamos como “orígenes” de referencia de las preguntas *j*=*k* (pregunta cualquiera) y *m*=*k* (persona cualquiera). Con estos “orígenes”, la pregunta *i* se calibrará respecto a *k* y la persona *n* se medirá igualmente respecto a *k*. Puede usarse el mismo valor como origen por la hipótesis básica del modelo que consiste en usar una sola escala tanto para las personas como para los ítems. El origen además se va a ubicar, por comodidad, al centro de la escala en una medida de 0, que corresponde con un valor de *P*=0.5. Con ello se ubica el centro de los ítems en el 50% de dificultad y el de las personas en el 50% de dominio o habilidad.

Esto simplifica a la expresión (2.3), ya que $p_{mj}=q_{mj}=p_{kk}=q_{kk}=0.5$
La expresión (2.3) se escribe:

$$\left(\frac{p_{ni}}{q_{ni}}\right)\left(\frac{q_{nk}}{p_{nk}}\right) = \left(\frac{p_{ki}}{q_{ki}}\right) \tag{2.4}$$

A su vez:

$$\left(\frac{p_{ni}}{q_{ni}}\right) = \left(\frac{p_{nk}}{q_{nk}}\right)\left(\frac{p_{ki}}{q_{ki}}\right) \tag{2.5}$$

Se aprecia que el segundo miembro define dos cocientes: uno sobre la persona *n* referida al origen *k*, y otro sobre la pregunta *i* referida al origen *k*, con lo cual se estima un valor del primer miembro que es la interacción entre la persona *n* y la pregunta *i*.

La expresión (2.5) puede hacerse lineal tomando logaritmos en ambos miembros, quedando:

$$\ln\left(\frac{p_{ni}}{q_{ni}}\right) = \ln\left(\frac{p_{nk}}{q_{nk}}\right) + \ln\left(\frac{p_{ki}}{q_{ki}}\right) \tag{2.6}$$

El primer miembro $\ln\left(\frac{p_{ni}}{q_{ni}}\right)$ corresponde con la medida en lógitos de la persona *n* al responder el ítem *i*.

El primer término del segundo miembro $\ln\left(\frac{p_{nk}}{q_{nk}}\right)$ es la medida en lógitos de la persona *n* referida al origen *k*, esto lo denominamos θ_v (como se verá en las secciones siguientes).

El segundo término del segundo miembro $\ln\left(\frac{p_{ki}}{q_{ki}}\right)$ se puede escribir también: $\ln\left(\frac{p_{ki}}{q_{ki}}\right) = -\ln\left(\frac{q_{ki}}{p_{ki}}\right)$, y $\ln\left(\frac{q_{ki}}{p_{ki}}\right)$ es la calibración en lógitos del ítem *i* referida al origen *k*, que se denomina b_i (como se verá en las secciones siguientes).

Entonces podemos escribir:

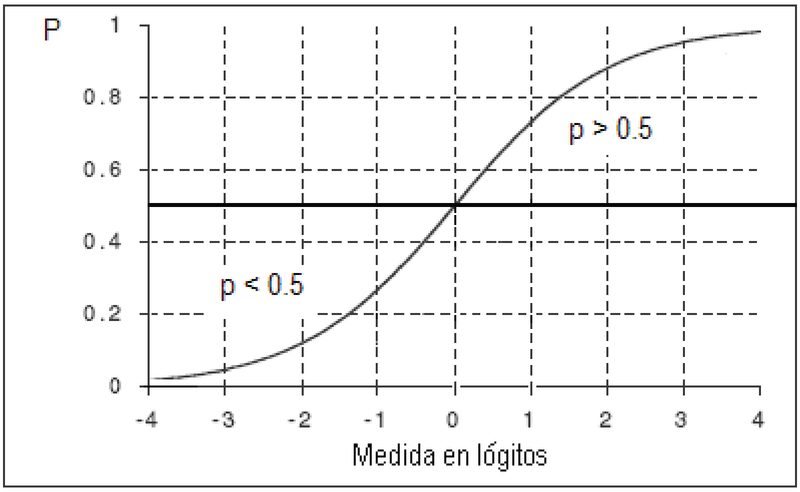
$$\ln\left(\frac{p_{ni}}{q_{ni}}\right) = \theta_v - b_i \tag{2.7}$$

Si resolvemos usando antilogaritmos, llegamos a obtener $P_{vi} = \frac{e^{(\theta_v - b_i)}}{1 + e^{(\theta_v - b_i)}}$, la cual denotaremos como:

$$P_i(\theta_v) = \frac{e^{(\theta_v - b_i)}}{1 + e^{(\theta_v - b_i)}} \tag{2.8}$$

Esta función se representa como se muestra en la siguiente figura:

Figura 2.1: Curva del Modelo de Rasch



La expresión (2.7) es el modelo de Rasch, mientras que la expresión (2.8) es el modelo de Thurstone de 1 parámetro. Se demuestra que el modelo de Rasch es el que corresponde con la exigencia de independencia que propone Thurstone (Se trata por lo tanto de una condición suficiente).

De acuerdo con esta deducción, el modelo generalizado de 3 parámetros no es suficiente si no redundante. Más concretamente podríamos decir que no es posible encontrar más de un parámetro a partir de la información disponible. Siguiendo con la deducción la expresión (2.7) no solamente es suficiente si no necesaria para la independencia de la medida.

Este modelo propuesto por Rasch se fundamenta en los siguientes supuestos:

1. El atributo que se desea medir puede representarse en una única dimensión en la que se situarían conjuntamente las personas y los ítems.
2. El nivel de la persona en el atributo y la dificultad del ítem determinan la probabilidad de que la respuesta sea correcta. Si el control de la situación es adecuado, esta expectativa es razonable y así debe representarla el modelo matemático elegido.

La ecuación (2.7) indica que el cociente entre la probabilidad de una respuesta correcta y la probabilidad de una respuesta incorrecta a un ítem, $\left(\frac{P_i(\theta_v)}{1-P_i(\theta_v)}\right)$, es una función de la diferencia en el atributo entre el nivel de la persona (θ_v) y el nivel del ítem (b_i). Así cuando una persona responde a un ítem equivalente a su umbral de competencia, tendrá la misma probabilidad de una respuesta correcta y una respuesta incorrecta, $\left(\frac{P_i(\theta_v)}{1-P_i(\theta_v)} = 0.5/0.5\right)$. En este caso, el logaritmo natural de $\frac{P_i(\theta_v)}{1-P_i(\theta_v)}$, refleja que la dificultad del ítem es equivalente al nivel de competencia de la persona $(\theta_v - b_i = 0)$. Si la competencia del sujeto es mayor que la requerida por el ítem $(\theta_v - b_i > 0)$, la probabilidad de una respuesta correcta será mayor que la de una respuesta incorrecta. Por el contrario, si la competencia del sujeto es menor que la requerida por el ítem $(\theta_v - b_i < 0)$, la probabilidad de una respuesta correcta será menor que la de una respuesta incorrecta.

Una formulación más conocida del modelo de Rasch, se deriva de la predicción de la probabilidad de responder correctamente al ítem a partir de la diferencia en el atributo entre el nivel de la persona (θ_v) y el nivel del ítem (b_i) (ecuación 2.8).

Los valores escalares de las personas y los ítems pueden expresarse en distintas métricas. La más utilizada es la escala de *logit*, que es el logaritmo natural de $\frac{P_i(\theta_v)}{1-P_i(\theta_v)}$, es decir, $\theta_v - b_i$. La localización del punto 0 de la escala es arbitraria. En la tradición de Rasch, se suele situar dicho punto en la dificultad media de los ítems. En este caso, es muy sencilla la interpretación de los parámetros de las personas (los valores de θ_v mayores que 0 indican que las personas tienen una probabilidad superior a 0.50 de éxito en los ítems de dificultad media). Aunque, la escala *logit* puede adoptar valores entre más y menos infinito, la gran mayoría de los casos se sitúa en el rango ± 3 ó ± 5 y muy exageradamente entre ± 8.

2.2.2 CARACTERÍSTICAS DEL MODELO DE RASCH

El modelo de Rasch es una de las propuestas dentro de la Teoría de Respuesta al Ítem (TRI) el cual tiene las siguientes características:

1. La Unidimensionalidad asume que los ítems del instrumento de medida reflejan solo un rasgo latente subyacente. Planteada la unidimensionalidad como hipótesis operativa en el proceso de ajuste de los datos al modelo, la obtención del correspondiente índice de

consistencia interna nos indicará si se valida la hipótesis. La validez del test esta vinculada a la identificación de la unidimensionalidad en un sentido sustantivo. El modelo de Rasch, como un modelo de rasgo latente, considera que un test unidimensional permitirá la localización de los ítems en función del parámetro de los ítems y los sujetos encuestados en función del parámetro de su rasgo latente.

2. Cuando un sujeto responde a un ítem equivalente a su umbral de competencia, la respuesta tiene la misma probabilidad para una respuesta errónea que para un acierto. Si la probabilidad de una respuesta correcta es mayor que la de una respuesta incorrecta, la competencia del sujeto será mayor que la requerida por el ítem. Por el contrario si la probabilidad de una respuesta correcta es menor que la incorrecta, la competencia del sujeto será menor que la requerida por el ítem.
3. La localización del punto cero de la escala es arbitrario. En este modelo se suele situar en el punto de dificultad media de los ítems.
4. A pesar de que la escala esta expresada en un logaritmo natural, es frecuente que sea multiplicado por 1.7 con el fin de conseguir una distribución normal con media 0, desviación típica igual a 1 y un rango de distribución entre ± 3.

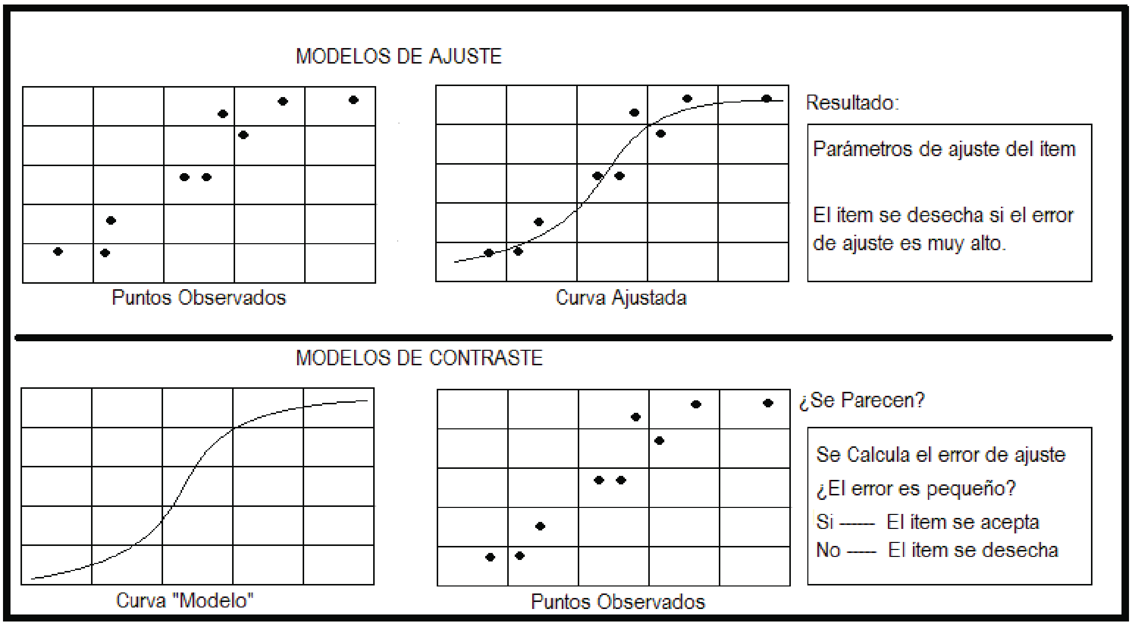
2.2.3 EL MODELO DE RASCH. UN MODELO DE CONTRASTE

Como se vio en el capítulo I, existen varios modelos para aproximarse a la curva característica de los ítems. De estos modelos se deben distinguir dos esquemas básicos:

a) *Modelo de Ajuste:* modelos que buscan una función que pase lo más cercanamente posible por el conjunto de puntos observados.
b) *Modelos de Contraste:* modelos que definen una curva de propiedades y características deseables para los ítems, contra la cual se contrastan los puntos observados.

Es importante ver la diferencia esencial de ambos modelos porque eso dará una primera luz sobre la calidad de cada uno de ellos. En la siguiente figura se observa la diferencia de los dos enfoques:

Figura 2.2: Diferencias entre los Modelos de Ajuste y los Modelos de Contraste



En el caso de los Modelos de Ajuste se tiene un conjunto de puntos y lo que se pretende es construir una curva que ajuste lo mejor posible a dichos puntos. Bajo este procedimiento siempre será posible obtener una curva que “ajuste” con los puntos dados. El ajuste será bueno o malo dependiendo de qué tan alineados se encuentren los puntos, pero siempre podrá obtenerse una curva de ajuste.

El modelo de ajuste más conocido es el modelo logístico generalizado de 3 parámetros y recientemente cuando éste falla se trata de ajustar un polinomio de grado 3 (ajuste con una cúbica).

Los Modelos de Contraste especifican una forma “deseable” para la curva característica de los ítems. Esta forma deseable es teórica y se basa en hipótesis que establecen los examinadores tratando de representar el comportamiento “Ideal” de un ítem. Una vez aceptada la curva se trata de ver si los puntos se parecen a ella. Para poder ver si hay ajuste o no, se calcula un error de aproximación o error de ajuste. En función del error de ajuste se toma la decisión de aceptar el ítem cuando se “ajusta al modelo” o rechazarlo cuando “no se ajusta al modelo”.

Este segundo enfoque, como vemos es más riguroso que el primero, en el sentido de que no busca ajustar una curva a cualquier precio; en cambio trata de averiguar si los puntos obtenidos se ajustan al modelo.

El modelo de Rasch esta dentro de los Modelos de Contraste, ya que no busca ajustar una curva indiscriminadamente. Si los puntos se ajustan como establece dicho modelo ¡Que bueno!, porque el ítem se comportará como establece dicho modelo y conoceremos todas sus propiedades. En cambio, si los puntos no se ajustan a la curva teórica, entonces se debe rechazar el ítem porque no se comporta como establece dicho modelo. No necesariamente el comportamiento que dice dicho modelo es el que nos puede gustar, pero es un modelo que tiene algunas limitaciones y alcances. Es claro que bajo este enfoque habrá muchos ítems que serán rechazados.

El modelo de Rasch establece que solamente puede obtenerse (b_i) a partir de la información disponible. En este modelo se cancela el uso de los siguientes parámetros:

a) *Adivinación sistemática*: para el modelo de Rasch puede haber adivinación individual, pero no hay adivinación sistemática para todos los sujetos.
b) *Discriminación*: relacionada con la inclinación de la logística. Todas las curvas tienen la misma inclinación en la zona central, por lo que se dice que en el modelo de Rasch todos los ítems, independientemente de su dificultad, discriminan igualmente bien.

2.2.4 EL LÓGITO. UNA NUEVA UNIDAD DE MEDIDA

Cuando se habla de evaluación de conocimientos es común emplear cuestionarios que se aplican a estudiantes para conocer el número de aciertos al conjunto de preguntas o ítems utilizados. Se acostumbra a reportar el "*grado de dominio"* de cada estudiante en términos de *"número de aciertos"* o *"porcentaje de aciertos"*. Independientemente de la conveniencia o no de usar estos indicadores, se afirma que la escala que se genera tiene varios defectos:

a) Cuando se dice que un estudiante obtuvo 0 aciertos, ¿Quiere decir que tiene cero conocimientos?
b) Cuando un estudiante obtiene todos los aciertos ¿Se puede afirmar que conoce todo?
c) Un estudiante obtiene 80 aciertos y otro 40 aciertos ¿Se puede decir que un estudiante sabe el doble del otro?

Es evidente que la escala hace referencia exclusivamente al rasgo evaluado, por ejemplo: conocimientos de Ciencias Sociales de tercer ciclo, Capacidad Analítica de Solución de Problemas en Matemática, etc. Un alumno que obtiene 100 aciertos en Ciencias Sociales no necesariamente obtendrá 100 aciertos en Matemática. La escala hace referencia exclusivamente al rasgo medido, que en la "jerga" de la evaluación se le denomina *"rasgo latente".* Se dice que no es posible determinar con precisión el grado de dominio en una determinada materia, pero si se puede explorar (por medio de algún muestreo) el conjunto de conocimientos de dicha materia, que el investigador supone que solamente tratan de esa disciplina.

Una pregunta de Matemática mide (o cuando menos trata de medir) un cierto conocimiento de la materia, pero la pregunta se ve influida por la forma de redactar que tenga el evaluador, por presentación del test, por el entorno del alumno, por la forma de aplicar el test, etc. Por ello se dice que se esta tratando de medir el "rasgo latente en el estudiante" que también es un "rasgo latente del ítem". La medida de este rasgo tiene error y es de interés emitir un dictamen de la medida y el error de la medición para que la evaluación que se haga al estudiante sea lo más justa y precisa posible. Por lo anterior, cuando se habla de 0 aciertos o del 100% de aciertos se tiene un desconocimiento preciso del rasgo latente, aunque se puede decir que, se trata de una aproximación o estimación de la medida del dominio del sujeto en el rasgo. La medida debe en consecuencia, referirse, señalarse o ubicarse en una escala para poder comparar el dominio de un sujeto con otros, o el dominio de un sujeto respecto a un criterio de aceptación o de competencia previamente establecido. En este momento entramos, por lo tanto, en el problema de definir la escala.

Es importante que la escala sea tan extensa que permita medir todos los sujetos, pero no tanto que resulte extremadamente amplia y, por lo tanto, se vuelva inútil. Por ejemplo, los límites naturales de la escala Celsius de temperaturas se definen en 0 y 100 ºC. Un termómetro en esta escala es conveniente para medir los puntos de fusión del hielo y de la ebullición del agua, pero se vuelve muy poco útil si se desea emplear para medir temperatura corporal de una persona, ya que en este caso solo se requiere un intervalo menor, entre 35 y 42 ºC.

Supóngase que se va a medir la estatura de un conjunto de personas y para ello se traza una marca en un muro a una altura aproximadamente de 1.5 m. ¿Qué puede decirse de esta escala? Resulta evidente que no tiene extensión. Cuando se haga la medición solo se podrá identificarse a una persona que este por arriba o por debajo de la marca, pero su medida es indefinida.

Para mejorar la medida se trazan cinco marcas por arriba y por debajo de la inicial, separadas por 1 cm. cada una. Ahora se tiene una escala de 10 cm. en total, centrada alrededor de la marca de 1.5 m. en cierta manera podemos decir que la escala ha mejorado, ya que se dispone de mayor número de marcas, pero sigue careciendo de extensión suficiente y no será factible medir a las personas por arriba de 1.55 cm. ni abajo de 1.45 cm. podemos ver que una escala mal diseñada solo permitirá identificar que un sujeto esta por arriba de la marca establecida, o por abajo, pero no faculta al evaluador para estimar que tan arriba o que tan debajo de la marca se encuentra el sujeto.

Si se traduce el caso anterior a lo que ocurre en un examen, se debería seguir exactamente la misma lógica. Es conveniente disponer de ítems en toda la gama de dificultades y no solamente ítems centrados al 50% de dificultad, con objeto de poder medir el dominio de cada persona con precisión. Con una escala completa si se puede ubicar a las personas y estimar que tanto sabe o que tanto no sabe.

Es evidente que podemos intuir que un alumno que obtiene 80 aciertos es probablemente mejor que otro que obtuvo 40 aciertos, pero ¿80 es el doble de conocimientos que 40? La respuesta es no. En esta escala no puede garantizarse que 4 unidades representen el doble del rasgo que 2 unidades. El problema es que las divisiones de la escala no están ubicadas a distancias constantes, ni las cantidades son directamente proporcionales. Se dice por ello que la escala no es lineal.

Con el objeto de mejorar el concepto de la escala, se propone el uso de una nueva unidad de medida lineal (en este caso 4 unidades es el doble de 2 unidades) y que facilite cubrir todo el rasgo medido.

2.2.4.1 El Lógito y la Medida de una Persona

El lógito es una traducción libre de la unidad de medida definida en ingles *"logit"*. La palabra *"logit"* es una forma abreviada de *"log odd ratio unit"* que se traduce por *"unidades en logaritmo del momio"*. Se le denomina momio o apuesta a la combinación del éxito *P* y el fracaso *Q* el cual es una estimación de la *"Medida"* de una persona. Por lo tanto el logaritmo del momio nos dará como resultado la medida de una persona en lógitos.

Esta nueva unidad de medida natural fue propuesta inicialmente por George Rasch para su modelo, el cual tiene como características principales que se encuentra en una escala sin extremos y centrada en el valor cero, además, es una escala lineal. Esta medida tiene la ventaja de que limita los valores de medida a intervalos razonables y que permite tomar en cuenta tanto el éxito como el fracaso en una sola medida. En general, aunque el rango de la escala va de menos infinito a infinito, el rango operativo es de tres lógitos a ambos lados del valor central.

Para comprender mejor el lógito y la medida de una persona, se presentará el siguiente ejemplo: Supóngase que hay una materia *M* formada por 100 objetivos específicos y que el alumno tiene un dominio de un 60% en los conocimientos de dicha materia. Si se le administra un examen de 100 ítems (una pregunta por objetivo), esperaríamos que el alumno *"A"* contestara correctamente a 60 de las preguntas y no pudiera contestar 40 de ellas. Esto puede interpretarse como que el alumno tiene una probabilidad de 0.6 (60/100) de respuestas correctas y una probabilidad de 0.4 (40/100) de fallas.

Ahora, su momio está dado por (aciertos dividido entre errores):

$$momio(P) = P/Q$$
$$momio(0.6) = (0.6/0.4) = 1.5$$

Podemos decir que si corremos las apuestas para ver como va a salir este alumno en el examen, las apuestas estarían de 1 a 1.5

El logaritmo natural del momio es la medida del alumno (Denominado $\theta(P)$):

$$\theta(P) = \ln(P/Q) = \ln(1.5) = 0.4055 \ \textit{Lógitos,}$$

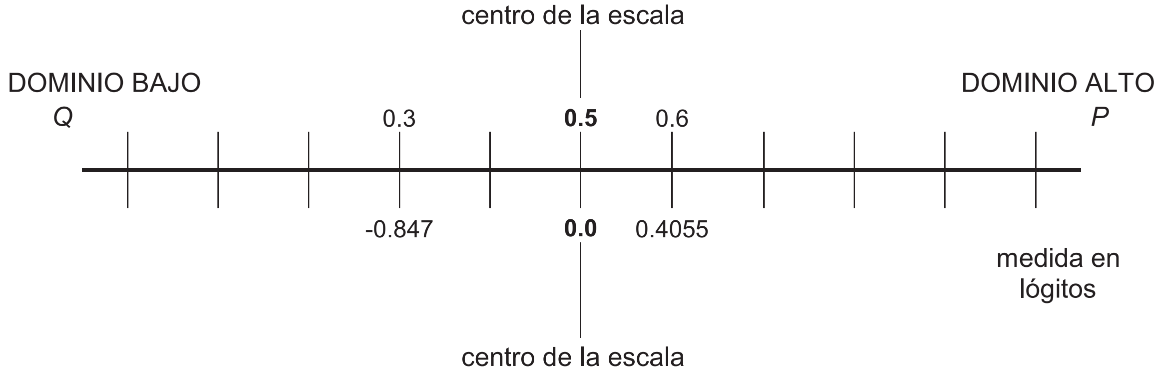
donde $\theta(P)$ representa la medida del alumno ante una probabilidad de 0.6 de respuesta correcta.

Ahora supongamos a un Alumno "B" con 30% de dominio. Al aplicar el proceso anterior obtenemos su medida de la siguiente manera:

$$\theta(0.3) = \ln(0.3/0.7) = -0.8473 \ \textit{Lógitos}$$

Si alguno de los dos alumnos tuviera un dominio del 50%, les correspondería una medida con el valor de 0 lógitos en la nueva escala. De esta manera el alumno "A" el cual tiene un dominio superior del 50% su medida resulta ser mayor que cero (0.4055 lógitos). De tal manera que el alumno "B" al tener un dominio inferior al 50% tiene una medida menor que cero (-0.8473 lógitos), lo cual se ilustra en la siguiente figura.

Figura 2.3: Escala de Valores Estimados de Probabilidad y la Medida del Alumno en Lógitos



En la figura anterior se puede observar la ubicación de ambos estudiantes (A y B).

La medida en lógitos hace referencia a la probabilidad de respuesta correcta en un examen dado. Cuando un estudiante obtiene medida 0, se dice que tiene probabilidad 0.5 de responder correctamente el test (cuestionario o examen). Cuando un estudiante tiene medida de 0.4055 lógitos, tiene una probabilidad de responder correctamente 0.6.

Cuando se tiene una medida en lógitos puede obtenerse la probabilidad de respuesta resolviendo la fórmula dada en la definición de medida:

$$\theta(P) = \ln(P/Q) \tag{2.9}$$

Haciendo uso de las propiedades de los logaritmos se obtiene:

$$P(\theta) = \frac{e^{\theta}}{1+e^{\theta}} \tag{2.10}$$

Ahora el argumento entre paréntesis nos indica que la probabilidad ha sido calculada para una medida dada θ. Por ejemplo, para una persona que tiene una medida θ de 1.1 lógitos, la probabilidad de respuesta es de:

$$P(1.1) = \frac{e^{1.1}}{1+e^{1.1}} = 0.75$$

Además, a partir de la fórmula de la medida se tiene que los límites van a ser fijados por los valores extremos de p, que varía entre 0 y 1:

Si *P=0, Q=1*, se tiene $\theta = \ln(0/1) = -\infty$

Si *P=1, Q=0*, se tiene $\theta = \ln(1/0) = +\infty$

2.2.4.2 El Lógito y la Medida de un Ítem

En el caso anterior se trabajó con la escala en lógitos para las personas. La misma medición se puede plantear para los ítems, por lo que los ítems también se pueden medir en lógitos.

En diversos aspectos de la vida utilizamos balanzas de precisión, termómetros u otros aparatos especiales para medir un objeto o un fenómeno dado. El instrumento de medida que se emplea en la evaluación de aprendizaje es el ítem o test. A la medida de los ítems se le denomina *"calibración"*. Así se dice que se está calibrando un ítem y se está midiendo a las personas. Esta nomenclatura es convencional y se justifica en el sentido que el instrumento que sirve para medir, por ejemplo una balanza de precisión, se tiene que "calibrar" en un laboratorio especializado; una vez calibrado sirve para medir el peso de las personas.

Supongamos que en un examen se incluyen tres ítems. Uno de 40% de dificultad, otro de 50% de dificultad y otro más de 70% de dificultad. Aquí debe observarse que conforme se hace más grande el valor de la dificultad se dice que el ítem es más difícil: un ítem de 0 de dificultad quiere decir que no es difícil (su dificultad es nula), mientras que otro del 100% de dificultad es totalmente imposible de ser respondido. La dificultad para esta metodología es el inverso del "Grado de dificultad"[1] clásico. En lo sucesivo se hablará de dificultad, que indica que tan difícil es un ítem y se dejará "Grado de dificultad" para el concepto clásico.

La calibración del ítem (b) se define por la expresión:

$$b(Q) = \ln(Q/P) \tag{2.11}$$

A partir de esta definición se puede ver que el momio empleado para la calibración del ítem es el cociente de fallas entre aciertos.

$$momio(b) = \ln(Q/P) \tag{2.12}$$

En la calibración el momio se refiere a las fallas, en la medición de las personas se refería al éxito o acierto. Por ejemplo: Determinar la calibración de los ítems que tienen 40%, 50% y 70% de dificultad, respectivamente.

Para el primer ítem, Q=0.4, P=0.6 y su calibración es: $b(0.4) = \ln(0.4/0.6) = -0.4055$

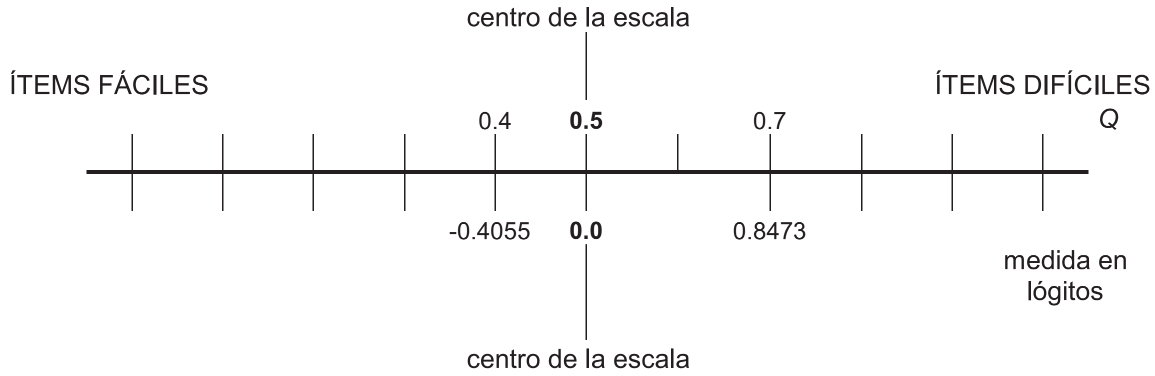
Para el segundo ítem, Q=0.5, P=0.5 y su calibración es: $b(0.5) = \ln(0.5/0.5) = 0$

---

[1] Recordemos que el Grado de dificultad se define como el cociente de las respuestas correctas entre el número de personas que contestan el ítem. Mientras más personas contestan correctamente, el ítem se hace más fácil y el Grado de dificultad aumenta; si pocas personas contestan correctamente, se dice que el ítem es difícil, sin embargo el grado de dificultad disminuye. Por ello el Grado de dificultad clásico es un índice inverso.

Y para el tercer ítem, Q=0.7, P=0.3 y su calibración es: $b(0.7) = \ln(0.7/0.3) = 0.8473$

Figura 2.4: Dificultad y Escala en Lógitos



Observemos que un ítem que se encuentra en el 50% de dificultad tiene calibración de 0 lógitos. Un ítem difícil tendrá una calibración mayor que cero, mientras que un ítem fácil tendrá una calibración menor que cero. Mientras más difícil sea el ítem más *"positivo"* y grande será; en cambio, los ítems más fáciles tendrán valores más pequeños y *"negativos"*.

La dificultad como su nombre lo indica, tiene que ver con la posibilidad de que los alumnos fallen o no respondan correctamente; por ello la medida (b) es función de *Q* fallas y no de *P*. Si se hubiera planteado (b) en función de *P*, entonces hablaríamos de "*facilidad*" del ítem y tendríamos un valor equivalente al Grado de Dificultad Clásico. Del mismo modo que se resolvió para la medida de la persona, se puede tener la expresión explícita para la probabilidad de respuesta dada una calibración. Así se tiene:

$$P(b) = \frac{e^{-b}}{1+e^{-b}} \tag{2.13}$$

Que se parece a la anterior empleada para (θ), pero se observa que ahora los exponentes del número *"e"* son negativos. Esta última expresión puede escribirse en forma equivalente:

$$P(b) = \frac{1}{1+e^{b}} \tag{2.14}$$

Por ejemplo: si se sabe que un ítems tiene una calibración de -0.3 lógitos, ¿Cuál es su probabilidad de respuesta?

Al sustituir en la fórmula se tiene:

$$P(b) = \frac{1}{1+e^{-0.3}} = 0.5744$$

El ítem es ligeramente más fácil que el "ítem medio" al 50% de dificultad. Por ello tiene una calibración menor que cero (-0.3) y por lo tanto una probabilidad de respuesta mayor que 0.5 (0.5744).

2.2.5 EL ESCALOGRAMA DE GUTTMAN

El escalograma de Guttman es conocido por los nombres de escalamiento acumulativo y análisis de escalograma, permite ordenar ítems y sujetos sobre una dimensión acumulativa subyacente. En el caso ideal, la puntuación total en una escala Guttman permite reproducir las respuestas dadas ante cada uno de los ítems que configuran la escala. Cuando estamos midiendo una sola variable o dimensión el escalograma es una buena herramienta que nos permite ver algunas cosas interesantes con relación al ítem y con relación a las respuestas de las personas.

Podemos localizar a una persona que contestó todo el cuestionario y podemos localizar a la persona que contestó menos. La construcción ayuda por lo tanto a ubicar a los alumnos de manera más rápida. Este escalograma permite hacer conjeturas respecto a la adivinación personal. Es probable que un sujeto haya adivinado a cierto ítem, pero seguramente otro sujeto no adivinó en sus respuestas.

Una forma de representar las respuestas de las personas a un test, consiste en hacer una tabla en donde se anotarán en cada renglón las personas y en las columnas las respuestas en cada ítem. Se trata de ordenar a las personas en orden descendente de aciertos (de la que tiene más aciertos a la que contestó menos) y los ítems en forma descendente de aciertos o ascendente de fallas (del ítems que tiene más respuestas correctas o más fácil, al que tiene menos respuestas correctas o más difícil). Tomemos el caso de 5 personas que contestan un test de 5 ítems:

Tabla 2.1: Resultados de las 5 personas ante un test de 5 ítems

Persona	***Ítem***				
	1	***2***	***3***	***4***	***5***
Alberto	1	1	1	0	0
Gonzalo	1	0	1	1	1
Luisa	0	0	1	1	1
Laura	0	0	0	0	1
Verónica	1	0	0	0	1

Las respuestas que aparecen en la tabla son obtenidas de acuerdo con el nivel de dominio o de habilidad que tiene cada persona al enfrentarse a cada uno de los ítems. Esperamos que las respuestas sigan algún "*patrón*" del tipo: a mayor dominio debe haber mayor número de respuestas. Sin embargo un alumno de cualquier nivel de dominio puede fallar en un ítem "*fácil*" por diversos motivos (error de lectura, distracción, malas instrucciones en la pregunta, etc.). Se dice entonces que los valores de esta tabla son estocásticos, porque están asociados con una cierta probabilidad.
A continuación se presenta el Escalograma de Guttman elaborado mediante el proceso descrito anteriormente:

Tabla 2.2: Escalograma de Guttman

Persona	**Ítem**				
	5	**1**	**3**	**4**	**2**
Gonzalo	1	1	1	1	0
Luisa	1	0	1	1	0
Alberto	0	1	1	0	1
Verónica	1	1	0	0	0
Laura	1	0	0	0	0

A partir del escalograma podemos apreciar que la persona que contestó más ítems del test es Gonzalo y podemos localizar a la persona que contestó menos ítems del test que es Laura. La construcción ayuda por lo tanto a ubicar a las personas de manera más rápida.

Vemos que algunas personas contestan los ítems fáciles y a partir de un cierto ítem dejan de contestar. Esto lo señalamos en color celeste dentro de la tabla 2.2. Se puede afirmar que estas personas contestan de acuerdo con "patrón lógico" (Gonzalo, Verónica y Laura), a saber: solamente contestan las preguntas debajo de su nivel de dominio, una vez alcanzado dicho nivel los ítems son más difíciles de lo que pueden contestar, y a partir de ahí no pueden responder correctamente. Podemos ver una diferencia sustancial entre Luisa y Alberto: ambos tienen 3 aciertos, pero Luisa esta en el "patrón lógico" y Alberto en cambio responde un ítem que es muy difícil. Podríamos conjeturar que Alberto contestó el ítem 2 por adivinación (o tal vez se le pregunto acerca de una cosa que sabía)

Este escalograma permite hacer conjeturas respecto a la adivinación personal. Es probable que un sujeto haya adivinado a cierto ítem, pero seguramente otro sujeto no adivinó en sus respuestas.

Conviene aclarar las hipótesis involucradas en el Escalograma de Guttman:

a) ***La variable es unidimensional***. Mide un solo rasgo. No se permite juntar o analizar rasgos de diversas dimensiones en una misma tabla. Dicho de otro modo: no se deben analizar en el mismo Escalograma las respuestas de por ejemplo Habilidad Verbal y las de Matemáticas.
b) ***La variable es ordenada***. El rasgo tiene una relación directa con la dificultad del ítem. Por ello el conjunto de ítems puede ordenarse en dificultad de tal forma que un ítem donde contestan 10 personas es más difícil que el ítem que es contestado por 20 personas y, a su vez, más difícil que el ítem contestado por 80 personas.
c) ***La variable es inclusiva***. Si una persona contesta un ítem con un nivel de dificultad de 30% entonces sería de esperarse que posee el dominio de los ítems con niveles de dificultad iguales o menores del 30% de dificultad.

De lo anterior se observa que dada una habilidad θ_v y un ítem de dificultad b_i,

- Si un θ_v es mayor que b_i entonces el ítem se le hará "fácil" a la persona y por lo tanto tiene probabilidad de acertar.
- Si θ_v es menor que b_i entonces el ítem se le hará difícil a la persona y por lo tanto tiene probabilidad de fallar.

Podemos considerar que el escalograma identifica a los ítems y a las personas de tal modo que se tiene una configuración como la que muestra en la siguiente figura:

Figura 2.5: Escalograma de Guttman “ideal”

1	1	1	1	1												1	1	1	1
1	1	1	1	1												1	1	1	0
1	1	1	1	1												1	1	0	0
1	1	1	1	1												1	0	0	0
1	1	1	1	1										1	1	0	0	0	0
1	1	1	1	1										1		0	0	0	0
1	1	1	1	1													0		
1	1	1	1	1													0		
1	1	1	1	1							1	0	0				0		
1	1	1	1	1							0						0		
1	1	1	1	1													0		
1	1	1	1	1													0		
1	1	1	1	1													0		
1	1	1	1	1													0		
1	1	1	1	1													0		
1	1	1	1	1	0												0		
1	1	1	1	0	0												0		
1	1	1	0	0													0		
1	1	0	0														0		
1	0	0	0														0		

Llamaremos “*diagonal*” del Escalograma a la línea que se traza de la esquina inferior izquierda hacia la esquina superior derecha y que divide al Escalograma aproximadamente en dos zonas triangulares. Observamos que arriba de la diagonal se esperan aciertos y que abajo se esperan errores. Es claro que las respuestas son estocásticas, por lo que no todos los valores arriba de la diagonal serán 1, habrá necesariamente algunos errores normales, atribuibles a fallas de comprensión de la pregunta, buenas opciones distractoras en un ítem de opción múltiple, etc. Igualmente se puede esperar que no todas las respuestas debajo de la diagonal sean 0. Habrá presencia de aciertos y fallas en forma estocástica.

Definimos el Escalograma Perfecto al que es perfectamente determinista, que cumple la no estocasticidad, es decir que garantiza que arriba de la diagonal se tienen siempre aciertos y debajo de la diagonal se tienen siempre errores, como el que se muestra en la figura 2.5.

Es prácticamente imposible encontrar un *Escalograma Perfecto*. Por lo tanto, no debemos forzar a los cuestionarios o exámenes a que se vuelvan escalogramas perfectos ni esperar que se nos presenten. Es más: si un estadístico llega a encontrar un *Escalograma Perfecto* al analizar los datos de su examen, entonces se verá en serios problemas para usar esta herramienta dentro del pre-proceso del análisis de Rasch, como veremos posteriormente.

2.2.5.1 El "Pre-proceso" del Escalograma de Guttman

Recordemos que un ítem que es contestado por un 50% de alumnos corresponde con una dificultad de 0 (sería el centro de la escala). Al trabajar en lógitos se sabe que un ítem de calibración 1 está una unidad por arriba del origen y tiene una distancia de 2 unidades por arriba de otro ítem de dificultad -1. Si encontramos un ítem que contestan muy pocas personas podríamos tener una medida de -2.5, por ejemplo.

Pero esta construcción de la escala tiene un problema. ¿Cuánto mide el ítem que no es contestado por ningún alumno? Podemos afirmar que está por abajo del ítem cuya calibración es -2.5 pero ¿Qué tan abajo?, ¿podría ser -3?, ¿tal vez llegue a ser -4, -4.5?. Nótese que cualquier ítem no contestado implica carecer de pistas para ubicarlo en la escala. Desde el punto de vista del análisis clásico de ítems estos ítems de respuesta nula son desechables, pero no permiten hacer una medida precisa y, además, no discriminan entre los sujetos. Desde el punto de vista de lo que estamos construyendo en esta tesis, un ítem de estas características también es desechable, porque un ítem de respuesta nula resulta "imposible" de medirse y ubicarse en la escala.

El mismo problema lo tenemos con los ítems que son contestados por la totalidad de las personas. ¿Qué tan arriba están ubicados? estos ítems también son desechables. Los ítems descritos hasta ahora se denominarán "*ítems extremos*".

Algo similar ocurre con las personas y sus medidas. No es posible ubicar a la persona que contesta todo el cuestionario o examen ni a la que no contesta nada. ¿Qué tan alto está la primera? ¿Qué tan bajo queda la segunda?. Las personas a que hacemos referencia se denominarán "personas extremas".

¿Qué debemos hacer en estos casos? Hay varias opciones, pero la que mencionaremos aquí es la de hacer un *"pre-proceso"* de los datos, que consiste en eliminar del análisis a los ítems extremos, así como a las personas extremas. Con este "*pre-proceso*" se eliminan los casos que presentan problemas de ubicación en la escala. El conjunto de datos que resta ya puede ser analizado y ubicado en su correcto valor de medida.

Tomemos el caso de un grupo de 5 personas que contestan un test de 6 ítems:

Tabla 2.3: Escalograma de Guttman Perfecto

Persona	***Ítem***					
	1	***2***	***3***	***4***	***5***	***6***
Juan	1	1	1	1	1	1
Luís	1	1	1	1	1	0
Memo	1	1	1	1	0	0
Lola	1	1	1	0	0	0
Marta	1	1	0	0	0	0

Hagamos el pre-proceso para eliminar ítems extremos y personas extremas. Los pasos son los siguientes:

1. Eliminar los ítems extremos 1 y 2
2. Eliminar la persona extrema "Juan"

Con ello se obtiene esta nueva tabla:

Tabla 2.4: Resultados del Paso 1

Persona	***Ítem***			
	3	***4***	***5***	***6***
Luís	1	1	1	0
Memo	1	1	0	0
Lola	1	0	0	0
Marta	0	0	0	0

Repetimos los pasos con esta nueva tabla:

3. Eliminar ítem extremo 6
4. Eliminar persona extrema "Marta"

Se tiene esta nueva tabla:

Tabla 2.5: Resultados del Paso 2

Persona	***Ítem***		
	3	***4***	***5***
Luís	1	1	1
Memo	1	1	0
Lola	1	0	0

Se repiten los pasos con esta nueva tabla:

5. Eliminar ítem extremo 3
6. Eliminar persona extrema "Luís"

Se tiene esta nueva tabla:

Tabla 2.6: Resultados del Paso 3

Persona	**Ítem**	
	4	**5**
Memo	1	0
Lola	0	0

En el paso 4 obligará a eliminar a Lola y al ítem 5, con lo cual se deberá quitar finalmente a Memo. Resulta curioso que en un Escalograma Perfecto el "*pre-proceso*" obliga a deshacerse a todos los datos. Realmente se trata de un problema muy serio. Por ejemplo, los programas BIGSTEPS y WINSTEPS para análisis de Rasch fallan durante el pre-proceso cuando se enfrentan a un Escalograma Perfecto. En este caso deberá recurrirse a los procedimientos clásicos de análisis de ítems: grado de dificultad, poder de discriminación, etc.

2.2.5.2 El Concepto de "Error"

A partir del Escalograma podemos ver que hay alumnos que responden con lo que hemos denominado un "*patrón lógico*" y otros que tienen respuestas inesperadas. Entendemos como respuestas inesperadas los aciertos que ocurren en ítems más difíciles que la medida de la persona o bien respuestas incorrectas en preguntas más fáciles que la medida de la persona.

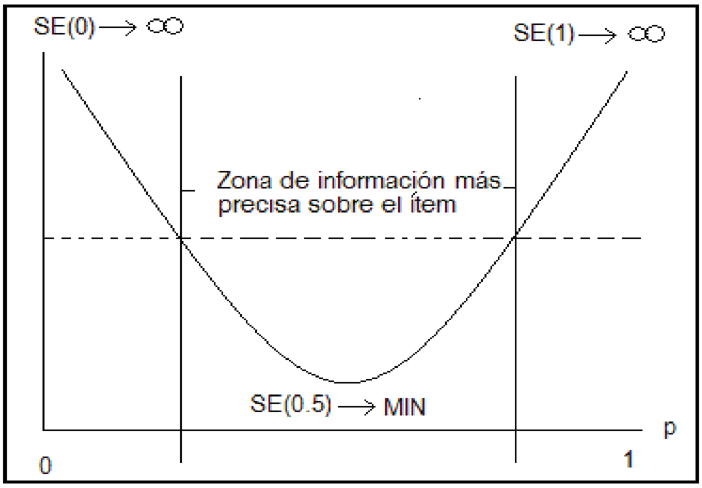
Hay autores que sugieren este procedimiento: Construir el escalograma de Guttman, identificar la diagonal que separa a las dos zonas triangulares de acierto y error, finalmente hacer un conteo simple de respuestas correctas inesperadas (1 en el triángulo inferior) y de respuestas inesperadas (0 en el triángulo superior); a este conteo le denominan "error". Desde luego este enfoque es deficiente.

En el análisis de Rasch se define como error a la distancia entre la respuesta y la medida de la persona. Por lo tanto decimos que una respuesta inesperada alejada de la medida de la persona conduce a un error grande. Una respuesta inesperada cerca de la medida de la persona indica un error pequeño. El error debe medir, por lo tanto *la distancia entre la respuesta y la medida de la persona.*

2.2.5.3 El Error Estándar

Una vez que se calcula una medida debe determinarse su error estándar de medida, que en el modelo de Rasch es mínimo cuando el ítem está cerca de *P=0.5* y tiende al infinito para *P=0* y *P=1*. Esto indica que las medidas son más precisas cuando el ítem se acerca a los valores "*centrales*" y conduce a medidas imprecisas en los extremos. De hecho se puede decir que se conocen mejor los ítems centrales y no se conoce nada (o muy poco) sobre los ítems donde todos contestan o nadie contesta.

Figura 2.6: Error Estándar



¿Qué quiere decir este error estándar?. Que la medida nunca es precisa, si no que hay error debido a los diversos patrones que tienen las personas al contestar un cuestionario. Sabemos que el error tiene que ver con la diferencia de la respuesta observada y la probabilidad de respuesta esperada. Debido a lo estocástico de las respuestas la diferencia se hace más grande en los extremos y más pequeña al centro.

2.2.5.4 El Error y la Distancia

¿Cómo mediremos la distancia entre la respuesta y la medida? El modo de hacerlo es relativamente simple. Hemos dicho que tanto la calibración de los ítems como la medida de las personas se representan en el mismo eje, expresado en lógitos. La distancia entre la respuesta y la medida se obtiene simplemente por la diferencia de la calibración y de la medida en lógitos:

$$Dis\tan cia = \theta_v - b_i \tag{2.15}$$

Donde θ_v es la medida en lógitos del dominio de una persona y b_i indica la calibración en lógitos de la dificultad de un ítem.

Por lo anterior concluimos que lo importante no es la calibración por sí misma o la medida por sí misma, sino la diferencia entre ellas. En función de esta diferencia se tendrá una probabilidad de respuesta. Dicho de manera simbólica:

$$\begin{aligned} &Si\ \ \theta_v = b_i\ ,\ \ \theta_v - b_i = 0 \rightarrow P = P(b_i) \\ &Si\ \ \theta_v > b_i\ ,\ \ \theta_v - b_i > 0 \rightarrow P > P(b_i) \\ &Si\ \ \theta_v < b_i\ ,\ \ \theta_v - b_i < 0 \rightarrow P < P(b_i) \end{aligned} \tag{2.16}$$

La expresión que relaciona la probabilidad de respuesta con la distancia $\theta_n - b_i$ tiene, esta forma:

$$\ln(P_i(\theta_v)/Q_i(\theta_v)) = \theta_v - b_i \tag{2.17}$$

Esta forma es una condición suficiente para conocer la probabilidad de respuesta de un alumno de medida θ_v ante un ítem de calibración b_i; es suficiente porque no se requiere de más parámetros para estimar la probabilidad de respuesta. Más aún, resulta claro que a partir de la información disponible en el Escalograma de Guttman no es factible estimar otros parámetros en forma clara y fidedigna. La fórmula 2.17 no es otra cosa que el mismo modelo de Rasch, presentado en la expresión 2.7. El modelo de Rasch resulta de manera natural al analizar el error a partir del Escalograma de Guttman. Con ayuda del modelo de Rasch, las expresiones 2.16 se pueden escribir así:

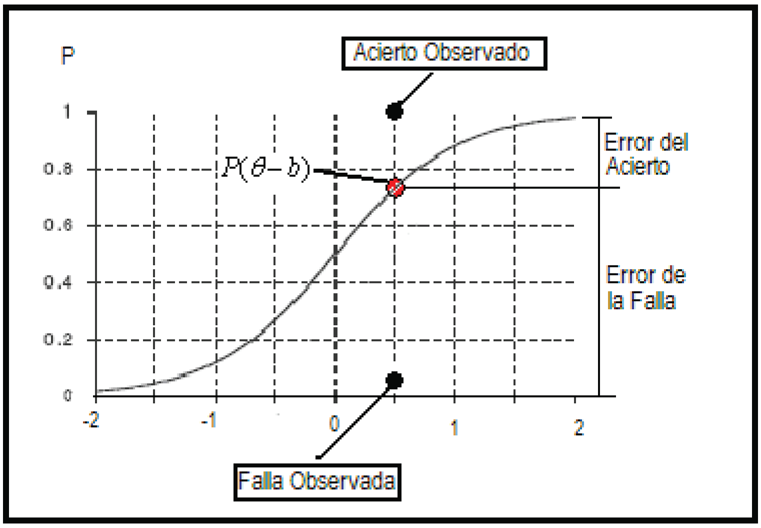
$$\begin{aligned} &Si\ \ \theta_v = b_i\ ,\ \ \theta_v - b_i = 0 \rightarrow P_i(\theta_v) = 0.5 \\ &Si\ \ \theta_v > b_i\ ,\ \ \theta_v - b_i > 0 \rightarrow P_i(\theta_v) > 0.5 \\ &Si\ \ \theta_v < b_i\ ,\ \ \theta_v - b_i < 0 \rightarrow P_i(\theta_v) < 0.5 \end{aligned} \tag{2.18}$$

El error se calculará por medio de la diferencia de respuesta observada respecto a la probabilidad de respuesta ante un ítem dado.

$$error = respuesta\ observada - P(\theta_v - b_i) \qquad (2.19)$$

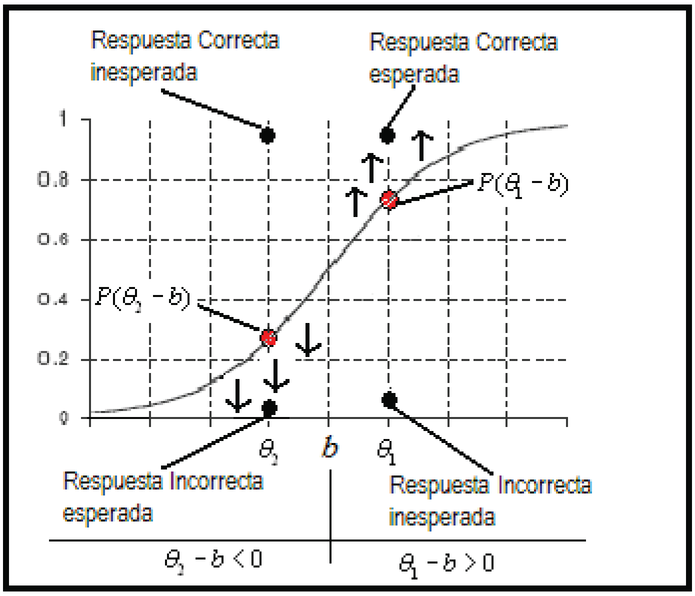
Este error se puede representar en forma gráfica. El término $P(\theta_v - b_i)$ corresponde con la curva del Modelo de Rasch (en forma amplia puede emplearse la Curva Característica del Ítem). Representemos un ítem de dificultad = 0 (figura 2.7). La respuesta observada es, en principio, correcta o incorrecta, por lo que se representa por medio de puntos en 1 ó 0, respectivamente.

Figura 2.7: Error y Distancia



La persona de medida $\theta = 0.5$ puede contestar acertadamente o en forma incorrecta, lo que se representa en la figura por dos circulitos (en 1 y en 0, respectivamente). La curva reporta la probabilidad de respuesta $P(\theta - b) = 0.75$. Si la persona contesta correctamente se tiene un error positivo (Error del acierto= 1-0.75 = 0.25), si contesta incorrectamente el error es negativo (error de la falla = 0-0.75), cada error señala la distancia que separa a $P(\theta - b)$ de la respuesta observada. Se puede extender esta discusión al caso de dos personas de medida θ_1 y θ_2, por arriba y por debajo de b respectivamente (figura 2.8).

Figura 2.8: Zonas de Respuesta Esperadas e Inesperadas



Si $\theta_v - b_i > 0$, la persona tienen un dominio mayor que la dificultad del ítem y son esperadas las respuestas correctas, mientras que son inesperadas las incorrectas. Contrariamente, si $\theta_v - b_i < 0$, la persona tiene un dominio menor que la dificultad del ítem, por lo que pueden esperarse respuestas incorrectas y son inesperadas las respuestas correctas. En la curva de la figura 2.8 se ilustra esta "tendencia" de respuestas esperadas en la dirección donde el error es más pequeño, de tal modo que cuando la curva tiene valores de *P* inferiores a 0.5 las respuestas "*esperadas*" tienden a ser incorrectas, mientras que cuando la curva tienen valores de *P* superiores a 0.5 las respuestas "esperadas" tienden a ser correctas. La tabla siguiente hace un resumen de lo que se acaba de comentar:

Tabla 2.7: Tabla Resumen

Caso	θ, b	**probabilidad** $P(\theta - b)$	**Respuesta observada**	**Clasificación**		**Observación**	**Error**
				Esperada	**Inesperada**		
1	*θ>b*	>0.5	1	X		como P>0.5 se espera que resp=1	*E=1-P(θ-b)* *E<0.5*
2	*θ>b*	>0.5	0		X	como P>0.5 es inesperado que resp=0	*E=0-P(θ-b)* *E>0.5*
3	*θ<b*	<0.5	1		X	como P<0.5 es inesperado que resp=1	*E=1-P(θ-b)* *E>0.5*
4	*θ<b*	<0.5	0	X		como P<0.5 se espera que resp=0	*E=0-P(θ-b)* *E<0.5*
5	*θ=b*	0.5	cualquiera			Como P=0.5 se puede esperar cualquier respuesta.	*E=0.5*

Otra forma de ver el error consiste en determinar en todos los casos la diferencia entre el valor observado (0,1) y la probabilidad estimada. Se considera que la respuesta es inesperada si esta diferencia es mayor que 0.5, en caso contrario se puede decir que la respuesta corresponde con el valor esperado.

Los errores tienen signo, por lo que no se acostumbra hacer su suma en forma directa, lo más recomendable es hacer la suma de los valores absolutos o elevando los errores al cuadrado. Asimismo es costumbre obtener valores medios de los errores y la raíz cuadrada de estos errores medios. En todos estos casos se trata de hacer un estimado sin signo del error total que tienen el conjunto de respuestas de un sujeto ante un grupo de ítems dado. Se pueden establecer estas fórmulas para el error:

Figura 2.9: Fórmulas para el Error

EA	= error absoluto	$= \sum_{i=1}^{N} abs(error_i)$
EAM	= error absoluto medio	$= \left[\sum_{i=1}^{N} abs(error_i) \right] / N$
EC	= error cuadrático	$= \sum_{i=1}^{N} (error_i)^2$
ECM	= error cuadrático medio	$= \left[\sum_{i=1}^{N} (error_i)^2 \right] / N$
RECM	= raíz de ECM	$= \sqrt{\sum_{i=1}^{N} (error_i)^2 / N}$

Donde *N* es el número de ítem.

El error que el evaluador debe emplear dependerá de sus necesidades, del propósito de la evaluación o de la costumbre. Finalmente este conjunto de fórmulas brinda varias formas de estimar un error y cada expresión brinda un valor diferente para expresar una misma cosa.

Por ejemplo: Se aplica un examen de 4 ítems, cuya calibración esta indicada por la tabla:

ítems	b_i
1	-1.10
2	-0.30
3	0.90
4	1.30

Dos alumnos de medida θ=1.0 contestan a estos ítems como se muestra en la tabla siguiente:

ítems	Respuesta del alumno X	Respuesta del alumno Y
1	1	1
2	0	1
3	1	1
4	0	0

Se estimaran los errores absolutos, cuadráticos y medios que tienen las respuestas de ambos alumnos:

Solución:

Para estimar el error se usará la expresión (2.19) y se construirá una tabla para cada uno de los sujetos, recordando que en ambos casos $\theta = 1.0$

a) Cálculo del error del alumno X

(1) ítems	(2) b_i	(3) $\theta_v - b_i$	(4) P($\theta_v - b_i$)	(5) Respuesta del alumno	(6) Error	(7) Error Absoluto *I Ei I*	(8) Error Cuadrático
1	-1.1	2.1	0.8909	1	0.1091	0.1091	0.0119
2	-0.3	1.3	0.7850	0	-0.7858	0.7858	0.6175
3	0.9	0.1	0.5250	1	0.4750	0.4750	0.2256
4	1.3	-0.3	0.4256	0	-0.4256	0.4256	0.1811
					TOTAL	1.7955	1.0362

EA	=Total Col (7)	=1.7955	Error Absoluto
EAM	=Total Col (7)/4	=0.4489	Error Absoluto medio
EC	=Total Col (8)	=1.0362	Error Cuadrático
ECM	= Total Col (8)/4	=0.2590	Error Cuadrático medio
RECM	= Raíz cuadrada de ECM	=0.5090	Raíz de error cuadrático medio

b) Cálculo del error para el alumno Y

(1) ítems	(2) b_i	(3) $\theta_v - b_i$	(4) $P(\theta_v - b_i)$	(5) Respuesta del alumno	(6) Error	(7) Error Absoluto I Ei I	(8) Error Cuadrático
1	-1.1	2.1	0.8909	1	0.1091	0.1091	0.0119
2	-0.3	1.3	0.7858	1	0.2142	0.2142	0.0459
3	0.9	0.1	0.5250	1	0.4750	0.4750	0.2256
4	1.3	-0.3	0.4256	0	-0.4256	0.4256	0.1811
					TOTAL	1.2239	0.4645

EA	=Total Col (7)	=1.2239	Error Absoluto
EAM	=Total Col (7)/4	=0.3060	Error Absoluto medio
EC	=Total Col (8)	=0.4645	Error Cuadrático
ECM	= Total Col (8)/4	=0.1161	Error Cuadrático medio
RECM	= Raíz cuadrada de ECM	=0.3408	Raíz de error cuadrático medio

El alumno Y tiene menor error que el *X*, debido a que éste tuvo una respuesta incorrecta inesperada en el ítem 2.

2.2.5.5 El Residuo Estandarizado

El error se ha calculado hasta ahora con respecto a $P(\theta_v, b_i)$ que es el valor esperado de respuesta del sujeto, donde se tiene una varianza esperada:

$$Varianza = s^2 = P_i(\theta_v)Q_i(\theta_v) = P_i(\theta_v)(1 - P_i(\theta_v)) = P(\theta_v - b_i)\left[1 - P(\theta_v - b_i)\right] \quad (2.20)$$

Se divide el error de la fórmula (2.19) por la raíz cuadrada de esta varianza (desviación Standard), se obtiene la expresión:

$$Z = (X - P_i(\theta_v))/s \sim N(0,1) \quad (2.21)$$

Que se denomina "*Residuo Estandarizado*" por su semejanza con la definición de la variable estandarizada Z. Esta variable se distribuye en forma normal con media 0 y con desviación estándar 1; la variable Z^2 (residuo cuadrático) se distribuye en forma de χ_1^2 (ji cuadrada con un grado de libertad).

Esta expresión se puede escribir explícitamente de esta manera:

$$Z = \frac{(X - P_i(\theta_v))}{\sqrt{P_i(\theta_v)(1 - P_i(\theta_v))}} \text{ donde } P = P(\theta_v - b_i) \quad (2.22)$$

Con ayuda del residuo estandarizado se tiene las siguientes expresiones:

Figura 2.10: Expresiones del Residuo Estandarizado

RE	=Residuo estandarizado	$= \sum_{i=1}^{N} Z_i$
REM	=Residuo estandarizado medio	$= \left[\sum_{i=1}^{N} Z_i \right] / N$
RC	=Residuo cuadrático	$= \sum_{i=1}^{N} (Z_i)^2$
RCM	=Residuo cuadrático medio	$= \left[\sum_{i=1}^{N} (Z_i)^2 \right] / N$
RRCM	=Raíz del RCM	$= \sqrt{\left[\sum_{i=1}^{N} (Z_i)^2 \right] / N}$

El uso del residuo estandarizado tiene varias ventajas. En particular se trata de un estimador sistemático que puede compararse fácilmente contra algún valor de cotejo que se elija. Comúnmente se emplea el valor de 1 como referencia.

Siguiendo con el ejemplo anterior, se determinará los residuos estandarizados y los demás valores para los datos del alumno X e Y.

a) Alumno X

(1) ítems	(2) b_i	(3) $\theta_v - b_i$	(4) $P(\theta_v - b_i)$	(5) Q=1- $P(\theta_v - b_i)$	(6) $\sqrt{P_i(\theta_v)Q_i(\theta_v)}$	(7) Respuestas	(8) Z	(9) Z^2
1	-1.1	2.1	0.8909	0.1091	0.3118	1	0.3499	0.1225
2	-0.3	1.3	0.7850	0.2142	0.4102	0	-1.9155	3.6693
3	0.9	0.1	0.5250	0.4750	0.4994	1	0.9512	0.9048
4	1.3	-0.3	0.4256	0.5744	0.4944	0	-0.8607	0.7408
						TOTAL	-1.4751	5.4374
						MEDIA	-0.3688	1.3594

RE	=Total Col(8)	=-1.4751	Error absoluto
REM	=Total Col(8)/4	=-0.3680	Error absoluto medio
RC	=Total Col(9)	=5.4374	Error Cuadrático
RCM	=Total Col(9)/4	=1.3594	Error Cuadrático medio
RRCM	=Raíz cuadrada de RCM	=1.1659	Raíz del error cuadrático medio

b) Alumno Y

(1) ítems	(2) b_i	(3) $\theta_v - b_i$	(4) P($\theta_v - b_i$)	(5) Q=1- P($\theta_v - b_i$)	(6) $\sqrt{P_i(\theta_v)Q_i(\theta_v)}$	(7) Respuestas	(8) Z	(9) Z^2
1	-1.1	2.1	0.8909	0.1091	0.3118	1	0.3499	0.1225
2	-0.3	1.3	0.7858	0.2142	0.4102	1	0.5220	0.2725
3	0.9	0.1	0.5250	0.475	0.4994	1	0.9512	0.9048
4	1.3	-0.3	0.4256	0.5744	0.4944	0	-0.8607	0.7408
						TOTAL	0.9625	2.0406
						MEDIA	0.2406	0.5102

RE	=Total Col(8)	=0.9625	Error absoluto
REM	=Total Col(8)/4	=0.2406	Error absoluto medio
RC	=Total Col(9)	=2.0406	Error Cuadrático
RCM	=Total Col(9)/4	=0.5102	Error Cuadrático medio
RRCM	=Raíz cuadrada de RCM	=0.7143	Raíz del error cuadrático medio

De nuevo se observa que el alumno Y tiene menores errores, en este caso errores estandarizados.

El cálculo de errores y residuos que hemos hecho aquí se sujeta a la posibilidad de que la tabla de respuestas no conduzca a un escalograma perfecto. Antes de hacer estos cálculos debe, por lo tanto, hacerse el pre-proceso, de tal manera de que la tabla resultante no tenga “error” por sí misma. Tanto el error, como el residuo estandarizado y sus valores medios, se emplearán cuando se hable de calidad y ajuste con el Modelo de Rasch.

2.2.6 EL MODELO DE RASCH. MEDIDA Y AJUSTE

Podemos apreciar algunas características del modelo de la expresión (2.7), las cuales son:

a) La medida en lógitos de una persona ante un ítem está dado en función de la diferencia $\theta_v - b_i$.

b) El valor θ_v es independiente del test aplicado; se entiende por θ_v la medida de una persona referida al origen común *k*. Se trata de la media real de una persona que debe poderse estimar independientemente de cualquier instrumento de medida que se desee.

c) El valor b_i es la calibración del ítem, referida a un origen común *k*. Esta calibración es independiente del conjunto de personas a quienes se aplicó el instrumento y representa la dificultad real del ítem.

El modelo propuesto por Rasch busca eliminar el problema de la métrica que se tiene en las medidas tradicionales o clásicas (porcentajes de aciertos, número bruto de ítems correctos, etc.). Se trata de una métrica lineal, donde una distancia de 1 lógito indica una unidad, independientemente de la posición en la cual se encuentre una persona o un ítem.

En la métrica de Rasch se trata de detectar y cuantificar una "estructura" entre los datos: personas, ítems, por ejemplo, siempre que esta "estructura" exista. La estructura es un concepto de tipo predictivo, donde se puede afirmar que una persona de bajo rendimiento tiene una probabilidad baja de acertar a las preguntas difíciles, o que una persona de alto rendimiento tiene una alta probabilidad de acertar a las preguntas fáciles.

Debido a que el modelo de Rasch es probabilista se establece de antemano que van a escasear los datos, o que los datos son insuficientes e, inclusive, que tengan datos procedentes de un problema mal planteado. Bajo su enfoque se estudia la probabilidad de que los datos –insuficientes o escasos- establezcan una estructura independiente, además de las posibles fluctuaciones que pueden tenerse en la toma de datos (errores en la aplicación, diferencias regionales, diferencias raciales, problemas personales, etc.).

Atendiendo que el modelo de Rasch produce curvas características del ítem que nunca se cortan, se puede decir que el modelo permite obtener medidas que definen constructos claros. El cruce de las

CCI impide que se defina un constructo asociado a la medida. La medida se representa en una escala e indica niveles de dominio, más allá de las propiedades ordinales de los modelos de ajuste.

La teoría de Rasch, por lo tanto, puede catalogarse entre las "teorías fuertes", ya que permite identificar estructuras entre los datos a pesar de todos los errores inherentes al proceso de medición. Con el método propuesto por Rasch se pueden hacer análisis para dos o más variables, dicotómicas o politómicas, etc.

Si hacemos una recapitulación, podemos anotar estas hipótesis generales del modelo de Rasch:

a) La escala que se maneja debe ser lineal.
b) La escala debe permitir realizar medidas de los ítems y personas referidas a un origen común.
c) La variable que se mide representa la diferencia entre la habilidad de la persona y la dificultad del instrumento.
d) La variable es unidimensional, ordenada e inclusiva, de acuerdo con el modelo del Escalograma de Guttman.
e) Las medidas son estocásticas, con una probabilidad de ocurrencia tanto en aciertos como en errores.

Hay que recordar que el patrón hipotético de Rasch, de acuerdo con el modelo de Guttman es que la probabilidad *p* es una función de θ_v y b_i, tal que:

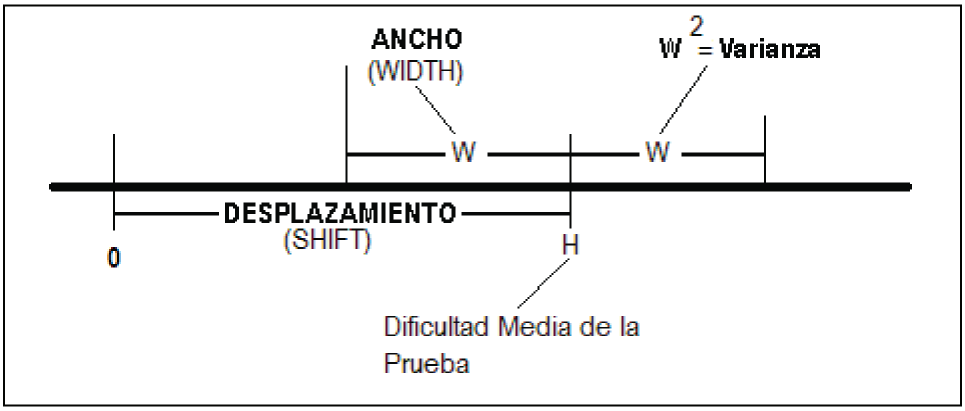
Si $\theta_i > \theta_j$, dada una dificultad b cualquiera, se cumple: $p(\theta_i, b) > p(\theta_j, b)$

Si $b_i < b_j$, dada una medida θ cualquiera, se cumple: $p(\theta, b_i) > p(\theta, b_j)$

2.2.6.1 La Independencia de la Medida

La independencia de la medida es una de las ventajas que ofrece el modelo de Rasch y que se conoce como *"objetividad específica"* que se verá en la última parte de este capítulo. Un método que se utiliza para garantizar que la medida es independiente de la muestra e independiente del instrumento, es que se modifican los valores de acuerdo con la media obtenida en el cuestionario por las personas. Esta modificación consiste en un "desplazamiento" hacia la media, de tal modo que se tiene un lógito de 0 en la media de los ítems y las personas. Se puede centrar por desplazamiento en los ítems o en las personas, pero generalmente se tiene en los ítems; es un desplazamiento de origen directamente, si la media de los ítems está en 0.5 por ejemplo, se centra restando 0.5 a todos los ítems. Si la media del test es de -0.5, el desplazamiento se obtiene sumando 0.5, al hacer este desplazamiento la media del test queda en 0. Una vez hecho esto, las medidas de las personas se desplazan en la misma cantidad.

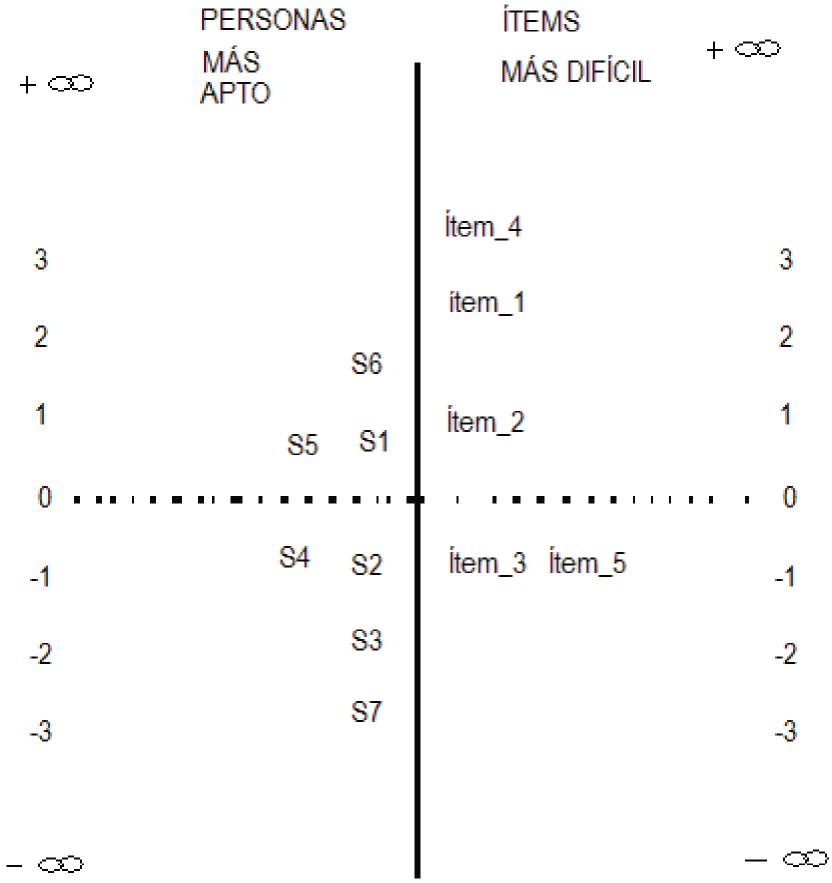
Figura 2.11: Desplazamiento en la Medida



La figura anterior nos muestra de una forma gráfica, el desplazamiento que se hace de los ítems y los sujetos hacia la media del test. Es gracias a este procedimiento que se pueden representar en una misma escala las medidas de las personas y las calibraciones de los ítems.

El mapa de ítems y personas (figura 2.12) nos sirve para observar la distribución de los parámetros de los ítems y las personas; además de poder observar que tan bien o mal se encuentra la escala y de esta manera poder agregar o eliminar ítems del test. Para entender mejor esta herramienta, consideraremos el siguiente mapa de distribución conjunta, donde tenemos a siete sujetos y cinco ítems.

Figura 2.12: Métrica Común para Ítems y Personas (Mapa de ítems y personas)



El programa BIGSTEPS utiliza "*X*" para representar las estimaciones de los parámetros de los ítems y sujetos, en el caso que sean demasiados utiliza "#". Lo ideal es que en este tipo de gráficos se pueda observar la forma de la normal en la parte de las personas, mientras que en la parte de los ítems, es recomendable que estos se comporten de manera uniforme. De la figura (2.12) podemos observar que los sujetos aparentemente tienen una distribución normal; en cambio en los ítems podemos observar que se encuentran bastante espaciados (entre el ítem 3 y el ítem 2; en el ítem 2 y el ítem 1), por lo tanto sería recomendable añadir más ítems para poder cubrir mejor la escala, además se observa que el ítem 3 y el ítem 5 se solapan entre si, esto nos esta indicando que hay que modificar o eliminar uno de estos dos ítems ya que está midiendo la misma parte del test. Podemos observar también que el ítem 4 y el ítem 1 son los más difíciles y sobrepasan el nivel de dominio de los sujetos. Los ítems que podrían ser contestados correctamente son el ítem 2, ítem 3 y ítem 5. Mientras que el sujeto 7 podemos observar que el dominio que posee es muy bajo en comparación de los ítems, por lo tanto es de esperarse que no pueda responder correctamente los ítems que comprende el test.

El desplazamiento ("SHIFT" en inglés) es un valor en términos de varianza que hace que la medida de un alumno se refiera a la dificultad del test. Por ejemplo, pensemos en un alumno de nivel medio que resuelve un test difícil. Debido a la dificultad del examen sus resultados pueden ser bajos y estaríamos tentados a decir que su media es baja. De modo contrario le ocurriría al mismo alumno ante un examen fácil, creeríamos que su media es mayor simplemente porque el instrumento le permite tener mejores resultados.

Si hacemos referencia a la dificultad del instrumento, es claro que en el primer caso debemos "aumentar" la calificación del alumno para que refleje que le tocó resolver un examen difícil. En el segundo caso debemos "reducir" la calificación del alumno porque su examen resultó fácil. El resultado de aumentar una calificación baja o el reducir una calificación alta debe conducir a la misma medida de la persona.
Como estamos trabajando en una escala lineal, está permitido hacer el desplazamiento (hacia arriba o hacia abajo), con el objeto de obtener la medida "real" del alumno.

2.2.6.2 El Ajuste o Control de Calidad del Modelo

Para poder estudiar la presencia o ausencia de "estructura" entre los datos se emplea el concepto de "ajuste" ("FIT" en inglés), para verificar que tan bien se ajustan los resultados obtenidos en el experimento o medición con el modelo de Rasch.

El ajuste es un proceso de cálculo que permite estimar la calidad de los resultados comparados contra el modelo. Por medio de unos parámetros de ajuste se hace el "control de calidad", que indica el grado en el cual las variables dadas (o las categorías de dichas variables) permiten identificar y definir una "estructura" entre los datos. Puede decirse que el "ajuste" permite medir la manera en que se "coordinan" o "correlacionan" los datos en la estructura.

Hay varios parámetros de ajuste. Se sugiere el uso de χ^2 para controlar la calidad de los datos en el análisis de Rasch. En particular se trata de valores referidos a los residuos medios estandarizados y, más específicamente, el cuadrado de dichos residuos, que se han visto anteriormente.

Las fórmulas dadas anteriormente, solo requieren una pequeña modificación, con el objeto que se emplee un valor de referencia de fácil interpretación. En este caso se sugiere utilizar el valor 1 como referencia. Al hacer esta modificación se tiene el valor cuadrático medio de ajuste con esperanza de 1. Los valores superiores a 1 indican que los patrones de respuesta no están de acuerdo con las hipótesis de que el cuestionario permite identificar una estructura general entre los datos.

La primera medida de ajuste se denomina "OUTFIT" (del inglés "Outlier-sensitive fit statistics") y que podríamos traducir libremente como "ajuste externo". Se trata de un valor sensible al comportamiento inesperado que afecta a los ítems cuya dificultad esta lejos del nivel de habilidad de la persona.

El Outfit se calcula con la fórmula:

$$OUTFIT = \sum_{i=1}^{N} Z^2 / N \tag{2.23}$$

(Esta fórmula es la misma que el correspondiente residuo cuadrático medio)

La otra forma de calcular el ajuste es con el "INFIT" (del inglés "Information-weighted fit statistics") y que podríamos traducir muy libremente como "ajuste interno". Se trata de un valor sensible al comportamiento inesperado que afecta a los ítems cuya dificultad esta cerca del nivel de habilidad de la persona. Se calcula con la expresión que hace un promedio ponderado de los residuos estandarizados:

$$INFIT = \frac{\sum_{i=1}^{N} Z^2 \sigma^2}{\sum_{i=1}^{N} \sigma^2} = \frac{\sum_{i=1}^{N} (X - E(X))^2}{\sum_{i=1}^{N} \sigma^2} \tag{2.24}$$

donde

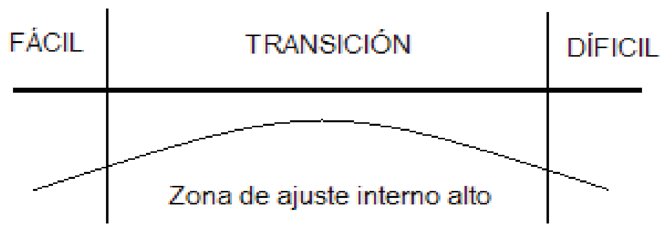
X: es el valor observado

E(X): es el valor esperado basado en parámetros de medida del modelo

σ^2 : es la varianza modelada alrededor del valor esperado

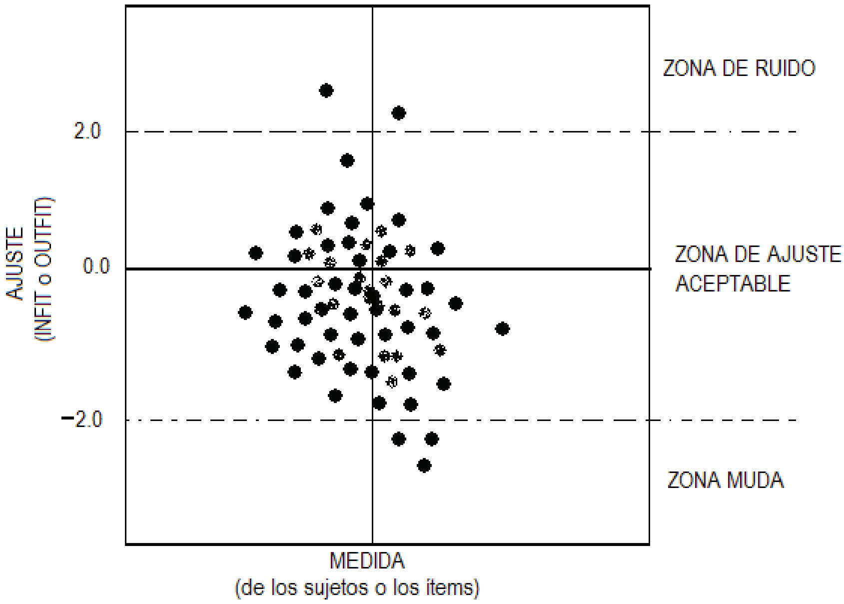
El ajuste interno (INFIT) es aplicable a las variables que se encuentran en la zona intermedia de la escala. Permite identificar los patrones inesperados entre observaciones que corresponden a la zona de estudio y, por ello es sensible a los patrones que se encuentran dentro de lo esperado. En forma esquemática, si se tiene la escala de calibración que va de fácil a difícil, el ajuste interno se enfoca a la zona media o de transición entre los extremos.

Figura 2.13: Forma Esquemática del INFIT



Bajo este parámetro es de esperar que en la zona fácil y en la zona difícil los patrones correspondan con el modelo (muchas respuestas en la zona fácil y pocas respuestas en la zona difícil). En la zona de transición se esperan mezclas estocásticas de "éxitos" y "fracasos" o "aciertos" y "errores". Si los datos corresponden con el modelo, el ajuste interno corresponderá con lo esperado, es decir con un valor cercano a 1. Los puntos correspondientes a las calibraciones de los ítems o las medidas de las personas se representan en un plano de medida (de los sujetos o ítems) versus el residuo cuadrático medio o residuo estandarizado. En este caso se obtiene una nube de puntos que se distribuyen de manera normal respecto al origen.

Figura 2.14: Nube de Ajuste de los Datos



La interpretación del "ajuste" (FIT) es como sigue:

1. Valores cercanos a 0, dentro del intervalo (-1, 1) indican un ajuste razonable al modelo estocástico.
2. Valores inferiores a -2 indican demasiado "determinismo" en el patrón de respuestas o "poca estocasticidad", digamos que "no hay errores" en el patrón de respuestas (Se tiene el patrón del escalograma de Guttman ideal). A esto se le llama "patrón mudo" ("MUTE" en inglés).
3. Valores superiores a 2 indican demasiada estocasticidad, mucha posibilidad de azar, o demasiadas respuestas inesperadas. A este caso se le denomina "patrón ruidoso" o " patrón de ruido" ("NOISY" en inglés)

El modelo de Rasch no se compromete a rangos específicos de valores, pero los anteriores son muy utilizados en la práctica. Disponemos de un conjunto de aceptación de los ítems: aceptaremos los ítems cuyo ajuste (INFIT o OUTFIT) este entre -2 y 2 lógitos.

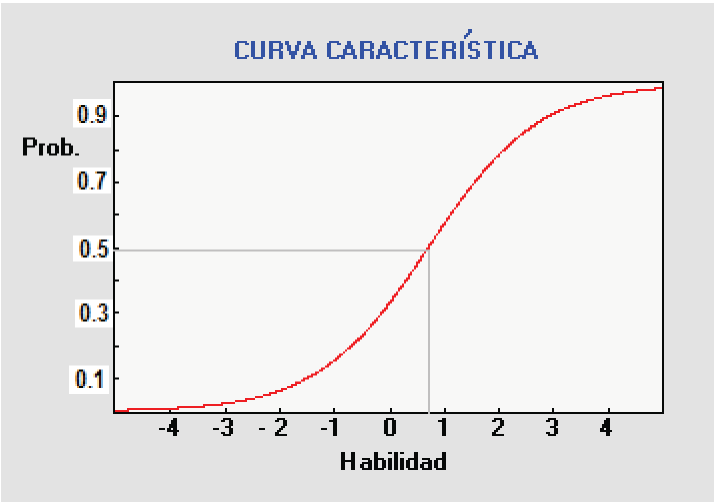
2.2.7 ESTIMACIÓN DE PARÁMETROS

Seleccionado un modelo de TRI (Específicamente el Modelo de Rasch), hay que aplicar el test a una muestra amplia y estimar los parámetros de cada ítem b_i (nivel de dificultad del ítem) y

θ_v (nivel de habilidad de cada sujeto), a partir de la matriz de respuestas obtenidas. Si tenemos, por ejemplo, diez ítems que miden un mismo rasgo, los podemos aplicar a una muestra de 300 personas. La matriz de datos tendrá 300 filas y 10 columnas, siendo cada fila la secuencia de unos (aciertos) y ceros (errores) de cada persona de la muestra. Al aplicar el modelo de Rasch, tendremos que estimar los 10 parámetros de los ítems (es decir "b_i" de cada ítem) y 300 parámetros de las personas (los 300 valores de " θ_v ", uno por persona). La estimación de parámetros es el paso que nos permite llegar de las respuestas conocidas de las personas a los ítems, a los valores desconocidos de los parámetros de los ítems y de los niveles de rasgo.

En el modelo de Rasch, se obtienen las estimaciones de los parámetros b_i y de los niveles de θ_v con los que la matriz de datos encontrada tiene la máxima probabilidad. Supongamos, por ejemplo, que tenemos un test compuesto por tan sólo dos ítems, y se lo aplicamos a un sujeto y éste acierta el primero y falla el segundo. A partir de estas respuestas, la estimación máximo-verosímil de θ_v se puede explicar de forma gráfica, como lo hacemos a continuación (en este ejemplo, para simplificar la explicación, suponemos que los parámetros de los ítems son conocidos). Como el sujeto ha acertado el primer ítem, podemos calcular, mediante su CCI, la probabilidad de que esto ocurra para cada nivel de θ_v. Gráficamente, para un ítem cuyo único parámetro es $b_1 = 0.7$ tenemos:

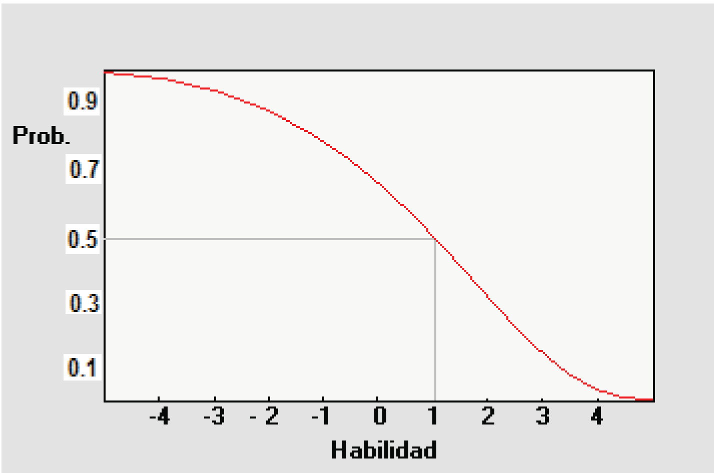
Figura 2.15: Curva Característica del Ítem 1



Como se observa en la figura anterior, si sólo hubiera respondido a ese ítem, podemos ver que no existe un único valor de θ_v para el que la probabilidad del suceso encontrado (acierto en el primer ítem) sea máxima. Por el contrario, son infinitos los valores de θ_v para los que la CCI alcanza el valor máximo 1.

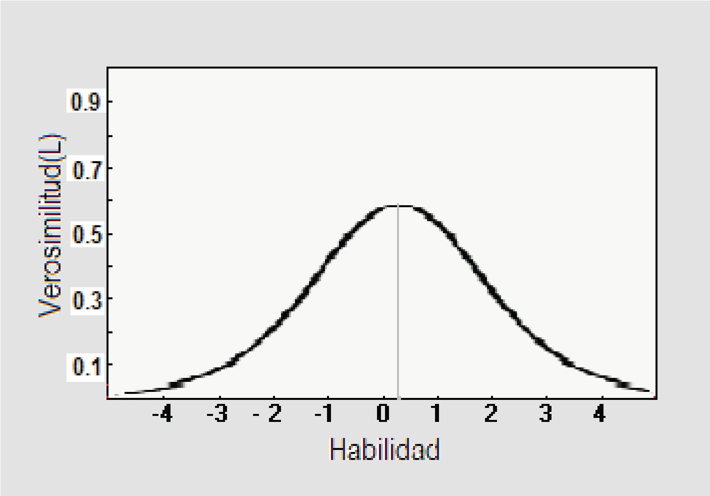
Como el sujeto ha fallado el segundo ítem, a partir de su CCI podemos calcular la probabilidad de que esto ocurra para cada uno de los valores de θ_v. En concreto, como la probabilidad de fallar (*Q*) se puede obtener a partir de la probabilidad de acertar (*Q* = *1-P*), podremos representar la probabilidad de error en el segundo ítem como se muestra en la siguiente gráfica. Observemos que la siguiente gráfica no corresponde con una CCI, ya que una CCI debe ser monótonamente creciente y además para cada valor de θ_v (habilidad del sujeto) se ha presentado la probabilidad de error, y no de acierto, como se exige en una CCI. Ahora, supongamos que tenemos la siguiente gráfica para el ítem 2 donde el único parámetro de dificultad es $b_2 = 1$.

Figura 2.16: Curva de Probabilidad de Error del Ítem 2



El eje de las ordenadas en la gráfica anterior nos indica la probabilidad de fallar al ítem 2. Y como podemos observar, es más probable que fallen el ítem los sujetos con niveles bajos de habilidad que los sujetos con niveles altos (cosa bastante lógica). Por lo tanto, si el sujeto sólo hubiese respondido a este ítem, de nuevo son infinitos los valores de θ_v que maximizan la probabilidad del suceso encontrado (error en el segundo ítem). Como ha respondido a dos ítems, el valor estimado de θ_v para este sujeto sería aquel que haga más probable el resultado obtenido (acertar el primer ítem y fallar el segundo). Según el supuesto de independencia local que existe en la TRI, ambos sucesos son independientes y, por lo tanto, la probabilidad de que ocurran ambos conjuntamente es igual al producto de las probabilidades de acertar el primero (P_1) por la de fallar el segundo (Q_2). Si representamos gráficamente la función $L=(P_1)(Q_2)$ para cada valor de θ_v, correspondiente al ejemplo que venimos comentando, obtendríamos una curva parecida a la siguiente:

Figura 2.17: Gráfico de la Función de Verosimilitud $L=(P_1)(Q_2)$



En este caso vemos que el θ_v que hace más probable el resultado obtenido (acierto en el primer ítem y fallo en el segundo) es algo mayor que cero. De hecho, 0.15 será el θ_v estimado para la habilidad de este sujeto.

En general, una persona responderá a un número de ítems (mayor de dos) y producirá una particular secuencia de unos y ceros. La probabilidad de obtener tal secuencia de aciertos y errores se puede escribir como:

$$L=\prod_{i=1}^{k}\left(\prod_{v=1}^{n}P_i(\theta_v)^{X_{vi}}\left[Q_i(\theta_v)\right]^{1-X_{vi}}\right) \qquad (2.25)$$

donde:

X_{vi}: Resultado observado para el sujeto *v* en el ítem *i* (1, acierto; 0, fallo).

$P_i(\theta_v)$: Probabilidad de acierto en el ítem *i* para el sujeto *n* con habilidad θ_v.

$Q_i(\theta_v)$: Probabilidad de error en el ítem *i* para el sujeto *n* con habilidad θ_v.

k: Número máximo de ítems.

n: Número máximo de sujetos.

El nivel de habilidad (θ_v) estimado por el método de máxima verosimilitud será el valor de θ_v para el que la expresión anterior alcanza su máximo valor.

Supongamos el siguiente ejemplo: Un test consta de 4 ítems, cuyos parámetros, según el modelo de Rasch, son -1, 0, 1 y 2. Una persona completa el test y acierta los tres primeros ítems y falla el cuarto. Se obtendrá el valor de la función de verosimilitud (*L*), para los siguientes valores de θ_n : -3, -2, -1, 0, 1 y 2, 3 . ¿Cuál de los valores anteriores maximiza *L*?.

Aplicando la expresión para el modelo de Rasch (2.8), se obtiene la probabilidad de acierto para cada ítem y cada uno de los valores de θ_v :

Tabla 2.8: Tabla de Probabilidades

ítems	b_i	θ_v						
		-3	-2	-1	0	1	2	3
1	-1	0.03	0.15	0.50	0.85	0.97	0.99	0.99
2	0	0.01	0.03	0.15	0.50	0.85	0.97	0.99
3	1	0.01	0.01	0.03	0.15	0.50	0.85	0.97
4	2	0.01	0.01	0.01	0.03	0.15	0.50	0.85

La función de verosimilitud, *L*, al haber acierto en los 3 primeros ítems y fallo en el último, será la siguiente:

$$L = (P_1^1 Q_1^0)(P_2^1 Q_2^0)(P_3^1 Q_3^0)(P_4^0 Q_4^1) = (P_1)(P_2)(P_3)(Q_4)$$

Aplicando la expresión (2.25) a cada uno de los valores de θ_v se obtienen los siguientes resultados:

L(3) = (0.99)(0.99)(0.97)(1-0.85) = 0.14

L(2) = (0.99)(0.97)(0.85)(1-0.50) = 0.41

Los restantes valores de *L* son *L*(1)= 0.35, *L*(0)= 0.06, *L*(-1) = *L*(-2) = *L*(-3) = 0.00, por lo tanto, de los siete valores de θ_v considerados, el valor que maximiza *L* es θ = 2.

Cuando se trata de estimar en una situación real el nivel de rasgo, no se hace una búsqueda restringida a unos cuantos valores, se necesita hallar el valor de θ_v que maximiza *L* de entre todos los posibles valores.

En el caso del modelo de Rasch no existen estrictamente fórmulas que permitan obtener las estimaciones de manera directa. En el ejemplo de las monedas se sabe que el estimador máximo-verosímil de la proporción poblacional es la proporción muestral. En el modelo de Rasch, al no existir tales fórmulas, las estimaciones se obtienen por métodos numéricos, mediante programas de ordenador como (Bigsteps, Winsteps, XCalibre, Quest, etc). En el caso más general se establece una función *L* que depende de los parámetros de los ítems y de los niveles de rasgo. Los programas de ordenador contienen algoritmos que encuentran el conjunto de estimaciones para el que la función *L* alcanza el valor máximo. Los parámetros de los ítems y los niveles de rasgo de las personas serán los valores dados por el programa de ordenador para una matriz de respuestas particular.

El método de máxima verosimilitud que mayormente ocupan los ordenadores para el cálculo de los parámetros se muestra a continuación:

Supongamos que X_{vi} es una variable aleatoria dicotómica i.i.d, tal que si θ_v y b_i son conocidos entonces

$$P(X_{vi} / \theta_v, b_i) = P_i(\theta_v)^{X_{vi}} \left[1 - P_i(\theta_v)\right]^{1-X_{vi}} \tag{2.26}$$

Ahora, para el caso de *k* ítems y un sujeto la función *L* (máxima verosimilitud) estará dada por la siguiente expresión:

$$L = \prod_{i=1}^{k} \left(P(X_{1i} / \theta_1, b_i)\right) = \prod_{i=1}^{k} \left(P_i(\theta_1)^{X_{1i}} \left[1 - P_i(\theta_1)\right]^{1-X_{1i}}\right) \tag{2.27}$$

Para el caso de un ítem y *n* sujetos la función de máxima verosimilitud estará dada por:

$$L = \prod_{v=1}^{n} \left(P(X_{v1} / \theta_v, b_1)\right) = \prod_{v=1}^{n} \left(P_1(\theta_v)^{X_{v1}} \left[1 - P_1(\theta_v)\right]^{1-X_{v1}}\right) \tag{2.28}$$

Ahora, cuando se tiene *k* ítems y *n* sujetos tenemos que la función de máxima verosimilitud es:

$$L = \prod_{i=1}^{k} \left(\prod_{v=1}^{n} P_i(\theta_v)^{X_{vi}} \left[1 - P_i(\theta_v)\right]^{1-X_{vi}} \right) \tag{2.29}$$

Al aplicar logaritmo natural y sus propiedades tenemos:

$$\ln L = \ln\left\{\prod_{i=1}^{k}\left(\prod_{v=1}^{n} P_i(\theta_v)^{X_{vi}}\left[1-P_i(\theta_v)\right]^{1-X_{vi}}\right)\right\}$$
$$\ln L = \sum_{i=1}^{k}\sum_{v=1}^{n}\left\{(X_{vi})\ln P_i(\theta_v) + (1-X_{vi})\ln\left[1-P_i(\theta_v)\right]\right\} \tag{2.30}$$

Derivando la expresión anterior con respecto a θ_v (habilidad del sujeto), tenemos:

$$\frac{\partial \ln L}{\partial \theta_v} = \sum_{i=1}^{k}\left\{\left(\frac{X_{vi}}{P_i(\theta_v)}\right)\left(\frac{\partial P_i(\theta_v)}{\partial \theta_v}\right) - \left(\frac{1-X_{vi}}{1-P_i(\theta_v)}\right)\left(\frac{\partial P_i(\theta_v)}{\partial \theta_v}\right)\right\}$$
$$\frac{\partial \ln L}{\partial \theta_v} = \sum_{i=1}^{k}\left\{\frac{X_{vi}-P_i(\theta_v)}{P_i(\theta_v)\left[1-P_i(\theta_v)\right]}\right\}\left(\frac{\partial P_i(\theta_v)}{\partial \theta_v}\right) \tag{2.31}$$

Según el modelo de Rasch la probabilidad de acierto de un ítem está dada por $P_i(\theta_v) = \dfrac{e^{(\theta_v - b_i)}}{1+e^{(\theta_v - b_i)}}$,

entonces la expresión para la derivada de $\dfrac{\partial P_i(\theta_v)}{\partial \theta_v}$ está dada por:

$$\frac{\partial P_i(\theta_v)}{\partial \theta_v} = \frac{\partial}{\partial \theta_P}\left(\frac{e^{(\theta_v - b_i)}}{1+e^{(\theta_v - b_i)}}\right) = \frac{e^{(\theta_v - b_i)}\left(1+e^{(\theta_v - b_i)}\right) - e^{2(\theta_v - b_i)}}{\left(1+e^{(\theta_v - b_i)}\right)^2}$$
$$\frac{\partial P_i(\theta_v)}{\partial \theta_v} = \left(\frac{e^{(\theta_v - b_i)}}{\left(1+e^{(\theta_v - b_i)}\right)}\right)\left(\frac{1}{\left(1+e^{(\theta_v - b_i)}\right)}\right) = P_i(\theta_v)\left\{1-P_i(\theta_v)\right\} \tag{2.32}$$

Ahora:

$$\frac{\partial \ln L}{\partial \theta_v} = \sum_{i=1}^{k}\left\{\frac{X_{vi}-P_i(\theta_v)}{P_i(\theta_v)\left[1-P_i(\theta_v)\right]}\right\}P_i(\theta_v)\left\{1-P_i(\theta_v)\right\}$$
$$\frac{\partial \ln L}{\partial \theta_v} = \sum_{i=1}^{k}\left\{X_{vi}-P_i(\theta_v)\right\} \tag{2.33}$$

A continuación se deriva con respecto a (b_i) (dificultad del ítem) la ecuación (2.30), obteniendo el siguiente resultado:

$$\frac{\partial \ln L}{\partial b_i} = \sum_{v=1}^{n}\left\{\frac{X_{vi}-P_i(\theta_v)}{P_i(\theta_v)\left[1-P_i(\theta_v)\right]}\right\}\frac{\partial P_i(\theta_v)}{\partial b_i} \tag{2.34}$$

donde

$$\frac{\partial P_i(\theta_v)}{\partial b_i} = \frac{\partial}{\partial b_i}\left(\frac{e^{(\theta_v - b_i)}}{1+e^{(\theta_v - b_i)}}\right) = \frac{-e^{(\theta_v - b_i)}(1+e^{(\theta_v - b_i)}) + e^{2(\theta_v - b_i)}}{\left(1+e^{(\theta_v - b_i)}\right)^2}$$

$$\frac{\partial P_i(\theta_v)}{\partial b_i} = \left(\frac{e^{(\theta_v - b_i)}}{1+e^{(\theta_v - b_i)}}\right)\left(\frac{-1}{1+e^{(\theta_v - b_i)}}\right) = -P_i(\theta_v)\left\{1-P_i(\theta_v)\right\} \qquad (2.35)$$

Ahora,

$$\frac{\partial \ln L}{\partial b_i} = -\sum_{v=1}^{n}\left\{\frac{X_{vi} - P_i(\theta_v)}{P_i(\theta_v)\left[1-P_i(\theta_v)\right]}\right\}P_i(\theta_v)\left\{1-P_i(\theta_v)\right\}$$

$$\frac{\partial \ln L}{\partial b_i} = -\sum_{v=1}^{n}\left\{X_{vi} - P_i(\theta_v)\right\} \qquad (2.36)$$

Igualando a cero las ecuaciones (2.33) y (2.36) y sustituyendo $P_i(\theta_v)$, se tiene el siguiente sistema de ecuaciones, llamadas ecuaciones normales:

$$\frac{\partial \ln L}{\partial \theta_v} = \sum_{i=1}^{k}\left\{X_{vi} - P_i(\theta_v)\right\} = \sum_{i=1}^{k}\left\{\frac{X_{vi} + X_{vi}e^{(\theta_v - b_i)} - e^{(\theta_v - b_i)}}{1+e^{(\theta_v - b_i)}}\right\} = 0 \quad ; v = 1,2,...,n$$

$$\frac{\partial \ln L}{\partial b_i} = -\sum_{v=1}^{n}\left\{X_{vi} - P_i(\theta_v)\right\} = -\sum_{v=1}^{n}\left\{\frac{X_{vi} + X_{vi}e^{(\theta_v - b_i)} - e^{(\theta_v - b_i)}}{1+e^{(\theta_v - b_i)}}\right\} = 0 \quad ; i = 1,2,...,k \qquad (2.37)$$

A partir del sistema de ecuaciones normales resulta muy difícil obtener expresiones explicitas para los estimadores de los parámetros de los ítems y los sujetos. Por tanto, se utilizan métodos numéricos para obtener los valores de los estimadores de los parámetros con algún grado de precisión (generalmente el de Newton-Raphson).

Para introducir el método de Newton-Raphson que permite aproximar los ceros reales de una función, es necesario que ésta sea continua (el denominador nunca se hace 0) y derivabable en un intervalo $\left[a,b\right]$, lo cual se cumple con el sistema de ecuaciones (más detalles consultar cualquier libro de Cálculo o Análisis Numérico). Partiremos de los resultados obtenidos por el método de máxima verosimilitud para encontrar la ecuación punto pendiente de la siguiente manera:

$$\frac{\partial \ln L}{\partial \theta_v} \cong \frac{\partial \ln L}{\partial \theta_v}\bigg|_{\theta_v^0, b_i^0} + \frac{\partial^2 \ln L}{\partial \theta_v^2}\bigg|_{\theta_v^0, b_i^0}(\theta_v - \theta_v^0) = \frac{\partial \ln L}{\partial \theta_v}\bigg|_{\theta_v^0, b_i^0} + \frac{\partial^2 \ln L}{\partial \theta_v^2}\bigg|_{\theta_v^0, b_i^0}(\theta_v) - \frac{\partial^2 \ln L}{\partial \theta_v^2}\bigg|_{\theta_v^0, b_i^0}(\theta_v^0) \qquad (2.38)$$

Igualmente se obtiene la expresión para (b_i):

$$\frac{\partial \ln L}{\partial b_i} \cong \frac{\partial \ln L}{\partial b_i}\bigg|_{\theta_v^0, b_i^0} + \frac{\partial^2 \ln L}{\partial b_i^2}\bigg|_{\theta_v^0, b_i^0}(b_i - b_i^0) = \frac{\partial \ln L}{\partial b_i}\bigg|_{\theta_v^0, b_i^0} + \frac{\partial^2 \ln L}{\partial b_i^2}\bigg|_{\theta_v^0, b_i^0}(b_i) - \frac{\partial^2 \ln L}{\partial b_i^2}\bigg|_{\theta_v^0, b_i^0}(b_i^0) \qquad (2.39)$$

Como $\frac{\partial \ln L}{\partial \theta_v} = 0$, $\frac{\partial \ln L}{\partial b_i} = 0$, se puede despejar (θ_v) y (b_i) de las expresiones (2.38) y (2.39) respectivamente, llegando a las siguientes expresiones:

$$\begin{aligned} \theta_v &= \theta_v^0 - \frac{\partial \ln L}{\partial \theta_v}\bigg|_{\theta_v^0, b_i^0} \left[\frac{\partial^2 \ln L}{\partial \theta_v^2}\bigg|_{\theta_v^0, b_i^0} \right]^{-1} \\ b_i &= b_i^0 - \frac{\partial \ln L}{\partial b_i}\bigg|_{\theta_v^0, b_i^0} \left[\frac{\partial^2 \ln L}{\partial b_i^2}\bigg|_{\theta_v^0, b_i^0} \right]^{-1} \end{aligned} \qquad (2.40)$$

De las expresiones (2.37) se tiene que:

$\frac{\partial^2 \ln L}{\partial \theta_v^2} = \frac{\partial}{\partial \theta_v}\left[\sum_{i=1}^{k} \{ X_{vi} - P_i(\theta_v) \} \right] = -\sum_{i=1}^{k} \left\{ \frac{\partial P_i(\theta_v)}{\partial \theta_v} \right\}$ y sustituyendo la expresión (2.32) se llega a:

$$\frac{\partial^2 \ln L}{\partial \theta_v^2} = -\sum_{i=1}^{k} \{ P_i(\theta_v) \{ 1 - P_i(\theta_v) \} \} \qquad (2.41)$$

Además,

$\frac{\partial^2 \ln L}{\partial b_i^2} = \frac{\partial}{\partial b_i}\left[-\sum_{v=1}^{n}\left\{X_{vi} - P_i(\theta_v)\right\}\right] = -\sum_{v=1}^{n}\left\{\frac{\partial P_i(\theta_v)}{\partial b_i}\right\}$ y sustituyendo la expresión (2.35) se llega a:

$$\frac{\partial^2 \ln L}{\partial b_i^2} = \sum_{v=1}^{n}\left\{P_i(\theta_v)\left\{1 - P_i(\theta_v)\right\}\right\} \tag{2.42}$$

Finalmente la estimación de los parámetros se obtiene del siguiente sistema de ecuaciones:

$$\begin{aligned} \theta_v &= \theta_v^0 + \frac{\left.\sum_{i=1}^{k}\left\{X_{vi} - P_i(\theta_v)\right\}\right|_{\theta_v^0, b_i^0}}{\left.\sum_{i=1}^{k}\left\{P_i(\theta_v)\left\{1 - P_i(\theta_v)\right\}\right\}\right|_{\theta_v^0, b_i^0}} \\ b_i &= b_i^0 + \frac{\left.\sum_{v=1}^{n}\left\{X_{vi} - P_i(\theta_v)\right\}\right|_{\theta_v^0, b_i^0}}{\left.\sum_{v=1}^{n}\left\{P_i(\theta_v)\left\{1 - P_i(\theta_v)\right\}\right\}\right|_{\theta_v^0, b_i^0}} \end{aligned} \tag{2.43}$$

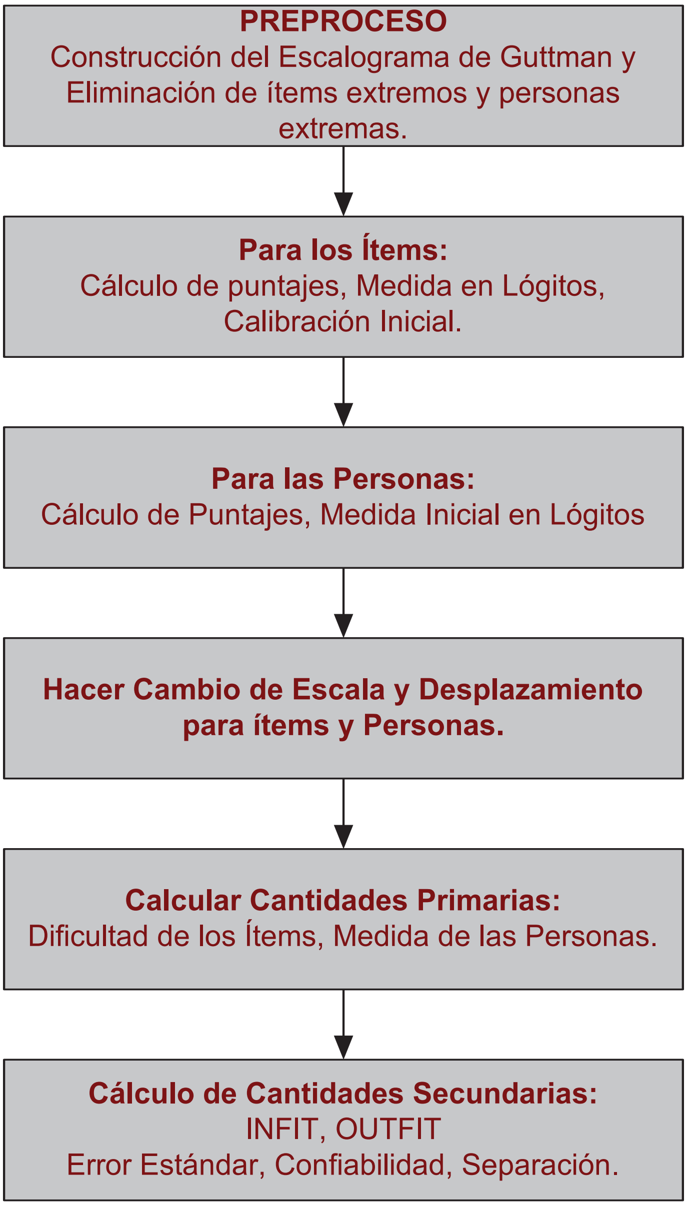
De las ecuaciones (2.43) se observa que es necesaria la asignación de un valor inicial de (θ_v^0) y (b_i^0), para luego obtener un estimador de los parámetros. Para la obtención de valores iniciales de la medida de los sujetos y calibración de los ítems, se utiliza el escalograma de Guttman y luego se aplica el método de Newton-Raphson para obtener los estimadores de los parámetros del modelo con algún grado de precisión, el cual está dado por:

$$\left|\theta_v - \theta_v^0\right| < \xi \tag{2.44}$$

El valor de precisión que se utiliza en el programa de BIGSTEPS (software que se utilizará en capítulo III) es de 0.001, aunque este valor puede ser modificado al gusto del usuario así como también el máximo número de iteraciones que se hacen.

El esquema de aplicación del modelo de Rasch en cualquier software sigue estos pasos:

Figura 2.18: Esquema de Aplicación del Modelo de Rasch por Cualquier Programa



Estos pasos se efectúan de forma iterativa y el tiempo de cálculo depende de la velocidad de la computadora, de la eficiencia del algoritmo así como del tamaño del test y del número de personas que lo contestan.

Ejemplo de simulación:

Supongamos que se ha examinado a un grupo de 5 estudiantes, con un test de 10 ítems, las respuestas obtenidas del test son dicótomicas, quedando así la matriz de datos *5X10:*

Tabla 2.9: Resultados del Test

Ítems / Sujetos	Ítems									
	1	2	3	4	5	6	7	8	9	10
Sujeto 1	1	1	0	1	0	0	1	0	1	0
Sujeto 2	0	0	1	0	1	0	0	0	1	0
Sujeto 3	1	1	0	0	1	0	0	0	0	1
Sujeto 4	0	1	0	0	0	1	1	1	0	1
Sujeto 5	0	1	1	0	0	1	1	1	0	1

De la matriz de datos se debe hacer el pre-proceso del escalograma de Guttman con el fin de eliminar ítems y personas extremas, en este caso no se eliminan ítems ni personas por tanto el escalograma de Guttman es el siguiente:

Tabla 2.10: Escalograma de Guttman

Sujetos	ítems									
	2	7	10	1	3	5	6	8	9	4
Sujeto 5	1	1	1	0	1	0	1	1	0	0
Sujeto 1	1	1	0	1	0	0	0	0	1	1
Sujeto 4	1	1	1	0	0	0	1	1	0	0
Sujeto 3	1	0	1	1	0	1	0	0	0	0
Sujeto 2	0	0	0	0	1	1	0	0	1	0

A partir del escalograma de Guttman se pueden analizar muchas situaciones en cuanto a los sujetos e ítems, por ejemplo podemos ver que el ítem más fácil es el 2 y el más difícil es el ítem 4; además se observa que el sujeto 5 es el que ha contestado correctamente la mayoría de los ítems del test y por el contrario el sujeto 2 ha contestado correctamente el menor número de ítems.

Con este escalograma se obtienen los valores iniciales de los sujetos e ítems. Para los sujetos el valor inicial de la medida se obtiene a través de la expresión (2.9) y para los ítems a través de la expresión (2.11), como por ejemplo: el sujeto 5 tiene un nivel de habilidad del test del 60%, por tanto, su medida inicial será:

$$\theta_5(P) = \ln(0.6/0.4) = \ln(1.5) = 0.4055 \text{ lógitos}$$

Ahora para obtener la medida o calibración inicial del ítem 2 será:

$$b_2(Q) = \ln(0.2/0.8) = \ln(0.25) = -1.3863 \text{ lógitos}$$

De esta manera al obtener los valores iniciales por cada sujeto e ítems se presentan en la siguiente tabla:

Tabla 2.11: Valores Iniciales de los Estimadores de los Parámetros

ÍTEMS		***SUJETOS***	
Número de Entrada	Medida inicial del ítem	Número de Entrada	Medida inicial de los Sujetos
ítem_2	-1.3863	Sujeto_5	0.4055
ítem_7	-0.9163	Sujeto_1	0.0000
ítem_10	-0.9163	Sujeto_4	0.0000
ítem_1	-0.9163	Sujeto_3	-0.4055
ítem_3	-0.5108	Sujeto_2	-0.8473
ítem_5	-0.5108		
ítem_6	-0.5108		
ítem_8	-0.5108		
ítem_9	-0.5108		
ítem_4	-0.2231		

Estos son los valores iniciales con que comienza a trabajar el algoritmo de Newton-Raphson. Para la estimación de los parámetros de los sujetos (θ_v) y de los ítems (b_i) se ha utilizado el programa BIGSTEPS, el cual utiliza dicho método para encontrar con mayor precisión los estimadores de los parámetros, la siguiente tabla nos muestra los estimadores:

Tabla 2.12: Estimación de los Parámetros de los Ítems y los Sujetos

ÍTEMS		***SUJETOS***	
Número de Entrada	Medida y error estándar del ítem (b_i)	Número de Entrada	Medida y error estándar de los Sujetos (θ_v)
ítem_4	1.28 (1.14)	Sujeto_5	0.46 (0.68)
ítem_5	0.26 (0.94)	Sujeto_1	0.02 (0.67)
ítem_3	0.26 (0.94)	Sujeto_4	0.02 (0.67)
ítem_9	0.26 (0.94)	Sujeto_3	-0.44 (0.68)
ítem_1	0.26 (0.94)	Sujeto_2	-0.93 (0.73)
ítem_6	0.26 (0.94)		
ítem_8	0.26 (0.94)		
ítem_7	-0.60 (0.94)		
ítem_10	-0.60 (0.94)		
ítem_2	-1.62 (1.14)		

Los valores que aparecen en paréntesis nos indican los errores de estimación de los estimadores de los parámetros. En la tabla anterior podemos observar de los resultados que son un reflejo del comportamiento de los sujetos y los ítems en el escalograma de Guttman (Tabla 2.10). De las estimaciones de (θ_v) y (b_i) podemos observar que el ítem más difícil para los sujetos es el ítem 4 y el más fácil es el ítem 2; mientras que el sujeto con mayor habilidad es el sujeto 5, y el sujeto con menor habilidad ha sido el sujeto 2. Como nos encontramos en una escala lineal podemos hacer comparaciones entre sujetos e ítems (una ventaja del modelo de Rasch), por tanto, se espera que el sujeto con mayor habilidad (sujeto 5) conteste correctamente todos los ítems a excepción del ítem 4, en otras palabras este ítem habrá que modificarlo o eliminarlo del test ya que es casi imposible que sea contestado correctamente. Los sujetos con medida de 0.02 (sujeto 1 y sujeto 4) se espera que contesten correctamente los ítems 7, 10 y 2.

Los estimadores de la tabla (2.12) se encuentran ordenados en forma descendente. Para encontrar estos valores, el programa BIGSTEPS ha hecho 10 iteraciones por cada ítem y por cada sujeto utilizando el algoritmo de Newton-Raphson y un criterio de convergencia de 0.001. Es evidente que al efectuar estas iteraciones en forma manual resulta muy trabajoso, por lo tanto el ordenador nos facilita estos procedimientos.

2.2.8 FUNCIÓN DE INFORMACIÓN (FI)

Una vez aplicado un conjunto de ítems y estimado el nivel de habilidad de un sujeto, la TRI (específicamente el modelo de Rasch) nos permite calcular el error típico de estimación (S_e) de esa persona en el test aplicado. Esto es una diferencia fundamental con la TC, que asume que el error es el mismo para todos los sujetos. El error típico de estimación nos dice la precisión con que hemos estimado un determinado parámetro. A mayor error, menos precisión. Su tamaño depende de varios factores:

1) Número de ítems aplicados: en general, al aumentar la longitud (número de ítems) del test disminuye S_e.

2) La diferencia entre θ_v y b_i: cuanto más próximo este el índice de dificultad de los ítems (b_i), menor será S_e.

La varianza de las puntuaciones θ_v estimadas $(Var(\hat{\theta}_v))$, tiene su origen mediante el teorema de Cramer-Rao donde la desigualdad se convierte en una igualdad estricta. Para el modelo de Rasch, la función de información y varianza del estimador se obtiene de la siguiente manera:

Por definición se tiene que la información de un parámetro o función de información del test aplicado sobre el parámetro se define como la inversa de $Var(\hat{\theta}_v)$ como se muestra en la siguiente expresión:

$$I(\theta_v) = E\left\{\left(\frac{\partial \ln P(X_{vi}/\theta_v, b_i)}{\partial \theta_v}\right)^2\right\} \text{ y } Var(\hat{\theta}_v) = S_e^2 = I(\hat{\theta}_v)^{-1} = \frac{1}{E\left\{\left(\frac{\partial \ln P(X_{vi}/\theta_v, b_i)}{\partial \theta_v}\right)^2\right\}} \quad (2.45)$$

De la expresión (2.26) se tiene

$$\frac{\partial \ln P(X_{vi}/\theta_v, b_i)}{\partial \theta_v} = X_{vi} - P_i(\theta_v) \quad (2.46)$$

y

$$I_i(\hat{\theta}_v) = E\left\{(X_{vi} - P_i(\theta_v))^2\right\} = P_i(\theta_v)\left(1 - P_i(\theta_v)\right) \quad (2.47)$$

$$I(\hat{\theta}_v) = \sum_{i=1}^{k} I_i(\theta_v) = \sum_{i=1}^{k} P_i(\theta_v)\left(1 - P_i(\theta_v)\right) \quad (2.48)$$

Por tanto:

$$Var(\hat{\theta}_v) \cong \frac{1}{\left\{\sum_{i=1}^{k} P_i(\theta_v)\left(1 - P_i(\theta_v)\right)\right\}} \quad (2.49)$$

La varianza anterior nos dice como es de importante la variación entre los valores de θ_v. Cuanto menor sea esta varianza, indicará que más podemos fiar del test; pues sabemos que son pocas las diferencias entre los valores estimados y el verdadero. El error típico de estimación de θ_v es la desviación típica de las puntuaciones de θ_v estimadas, es decir,

$$S_e = \sqrt{S_e^2} \tag{2.50}$$

El error típico de estimación permite obtener el intervalo de confianza en el que, con probabilidad predeterminada, se ha de encontrar el nivel de habilidad de la persona. Utilizando la ecuación (2.21) se tiene un intervalo de confianza ($100(1-\alpha)\%$) para (θ_v) de la siguiente manera:

$$\hat{\theta}_v - Z_{\alpha/2}(S_e) \leq \theta \leq \hat{\theta}_v + Z_{\alpha/2}(S_e) \tag{2.51}$$

donde

$$S_e = \sqrt{\frac{1}{\left\{ \sum_{i=1}^{k} \left(\frac{e^{(\theta_v - b_i)}}{1+e^{(\theta_v - b_i)}} \right) \left(1 - \frac{e^{(\theta_v - b_i)}}{1+e^{(\theta_v - b_i)}} \right) \right\}}}$$

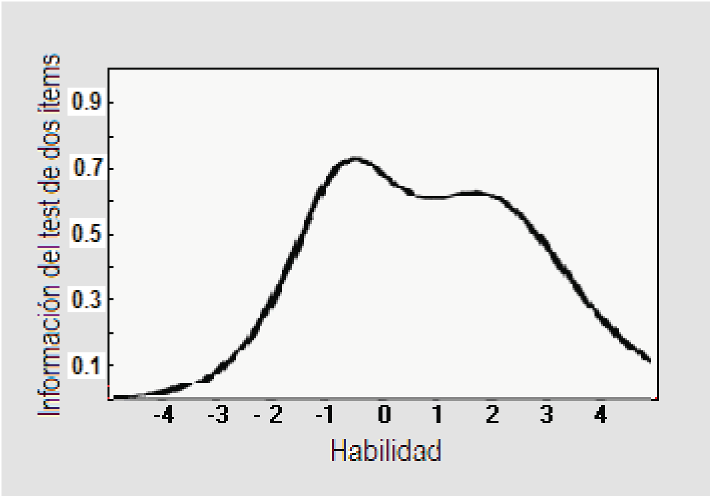
Para el ejemplo anterior, si el estimador de (θ_v) para el sujeto 5 es de 0.46 y su error típico de estimación es de 0.68, el cual se obtiene mediante la expresión:

$$S_e = \sqrt{\frac{1}{\left\{ \sum_{i=1}^{10} \left(\frac{e^{(0.46 - b_i)}}{1+e^{(0.46 - b_i)}} \right) \left(1 - \frac{e^{(0.46 - b_i)}}{1+e^{(0.46 - b_i)}} \right) \right\}}} = 0.68$$

entonces el nivel de rasgo de dicha persona se encuentra entre -0.8728 (0.46-(1.96)*(0.68)=-0.8728) y 1.7928 (0.46+(1.96)*(0.68)=1.7928), con un nivel de confianza del 95%.

Si se calcula $I(\theta_v)$ para todos los niveles de θ_v y se presenta gráficamente se obtiene una curva como la que se muestra en la siguiente figura:

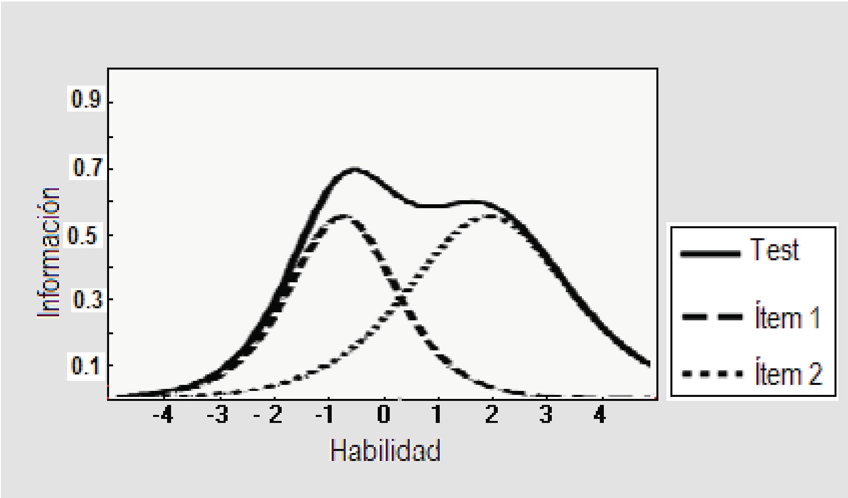
Figura 2.19: Gráfico de la Función de Información para Dos Ítems



Vemos que este test (compuesto por dos ítems, cuyos parámetros son $b_1 = -0.7$ y $b_2 = 2$) aporta más información para valores de θ_v alrededor de -0.5. La función de información tiene una gran importancia en la utilización de los test, ya que nos permite elegir aquel que aporte más información en el intervalo de θ_v que estemos interesados en medir. También es muy útil en la construcción del test. A partir de un banco de ítems calibrados (es decir, de los que hemos estimado sus parámetros) podemos seleccionar aquellos que permitan que la función de información se ajuste a unos objetivos determinados.

Todos los conceptos anteriores referidos a la función de información del test son aplicables también a cada uno de los ítems por separado. De hecho la función de información del test es la suma de las funciones de información de cada uno de los ítems que lo componen como se muestra en la expresión (2.48). Cada ítem puede representarse gráficamente y ver a que nivel de θ_v proporcionan más información. La siguiente gráfica muestra la función de información de los dos ítems que forman el test y la función de información del test.

Figura 2.20: Funciones de Información de los Ítems y del Test



Esto nos permite elegir los ítems más adecuados en cada momento en función de nuestras necesidades. Por ejemplo, si queremos llevar a cabo una selección de personal en la que sólo vamos a elegir unos pocos sujetos muy competentes, a partir de un banco de ítems previamente calibrado, podríamos elegir aquellos ítems que proporcionan más información para niveles altos de (θ_v). Esto nos permite reducir enormemente el número de ítems de un test sin perder precisión al estimar (θ_v).

2.3 GENERALIZACIÓN DEL MODELO DE RASCH

Los modelos politómicos se utilizan en el contexto de las ciencias sociales para medir capacidades, actitudes o rasgos psicológicos. Un ítem es cada una de las preguntas o tareas que se le proponen a un sujeto, y un conjunto de ítems es un test. Los modelos politómicos asumen que el número posible de respuestas a cada ítem es mayor de dos. Por ejemplo, en un ítem de conocimientos de álgebra las respuestas podrían catalogarse como “mal”, ”regular”, ”bien”. En un ítem de actitudes las categorías podrían ser “muy en desacuerdo”, “en desacuerdo”, “de acuerdo”, o “muy de acuerdo”. Los modelos politómicos de respuesta al ítem permiten estimar las propiedades de los ítems y los niveles del rasgo de las personas a partir de la matriz de respuesta obtenida por una muestra de sujetos en un test.

El modelo dicotómico de Rasch visto en los apartados anteriores, se aplica a datos de muy distinta procedencia. En cambio al trabajar con datos politómicos se explora en una complejidad mayor que el dicotómico debido a que las categorías de respuesta pueden estar o no ordenadas. Por ejemplo, si las categorías de respuesta a un ítem de actitudes son: en desacuerdo, neutral y de acuerdo, existe un orden implícito en estas respuestas. Esto no sucede así en un ítem de opción múltiple en los que existe una sola respuesta correcta y las demás se consideran incorrectas[2]. La existencia de distintos tipos de datos politómicos ha hecho que se propongan distintos modelos. Actualmente con el desarrollo de la tecnología ha sido posible el desarrollo de muchos modelos politómicos, entre estos se encuentran modelos que son considerados de la familia de Rasch por poseer las mismas características del modelo dicotómico (independencia de la medida, estadísticos suficientes, etc.), pero la generalización del modelo de Rasch para datos politómicos con parámetros unidimensionales se le atribuye al modelo de Crédito Parcial (MCP), por su generalización. Aunque de éste pueden derivarse otros modelos como el modelo Escala de Calificación que es un caso particular del modelo de Crédito Parcial.

Los datos que podemos analizar son como los que se muestran en la figura 2.21.

Figura 2.21: Tipos de Datos para el Análisis de Rasch

	Ítems					
Individuo	**1**	**2**	**…**	***i***	**…**	***k***
1	0	1		0		0
2	1	1		0		1
:						
i						
:						
n	1	1		1		0

Modelo de Rasch Dicotómico

	Ítems					
Individuo	**1**	**2**	**…**	***i***	**…**	***k***
1	0	3		2		1
2	3	0		0		4
:						
i						
:						
n	2	1		4		2

Modelo de Rasch politómico

$v \in \{1, 2, 3, ..., n\}$: Índice del individuo

$i \in \{1, 2, 3, ..., k\}$: Índice del ítem

$h \in \{0, 1, 2, ..., m\}$: Modalidad de respuesta por el ítem

[2] Este tipo de ítems se analiza con el modelo de respuesta nominal (Bock, 1972) no visto en este trabajo.

2.3.1 MODELO DE CRÉDITO PARCIAL (MCP)

El modelo MCP fue desarrollado por Masters (1982), a partir del modelo dicotómico de Rasch. Éste es una generalización del modelo dicotómico, que se puede aplicar en los contextos en los cuales las puntuaciones sucesivas son números enteros que representan las categorías del nivel o de la magnitud de aumento de un rasgo latente, tal como capacidad de aumento, evaluaciones psicológicas y así sucesivamente. Además, teóricamente los ítems pueden tener distintas categorías.

Una crítica de este modelo de Rasch es que tras su finalización, éste es restrictivo porque no permite que cada ítem tenga distinto valor de discriminación, al igual que el caso del modelo logístico de dos parámetros (Birnbaum, 1968). La especificación de la discriminación es uniforme, sin embargo, es una característica necesaria de este modelo para sostener sus propiedades que lo hace un modelo accesible. El modelo de Rasch requiere que la discriminación sea uniforme a través de interacciones entre las personas y los ítems dentro de un marco especificado de la referencia (es decir contexto experimental).

El Modelo de Crédito Parcial es un modelo de medida que tiene uso potencial en cualquier contexto en el cual el objetivo sea medir un rasgo o una capacidad con un proceso en el cual las respuestas a los ítems *se anoten* con números enteros sucesivos. Por ejemplo, el modelo es aplicable al uso de las escalas de Likert, escalas de grado y a los ítems educativos para los cuales la puntuación categórica más alta se utiliza para indicar niveles de aumento de la capacidad o del logro.

Para el análisis de datos categóricos ordinales, puede utilizarse el modelo de Rasch, el cual posee una importante característica: a diferencia de otros modelos de la TRI, el modelo de Rasch politómico asume que si eliminan una de las categorías del ítem cambian las probabilidades de todas las categorías. Esto implica que para emitir una respuesta el sujeto no evalúa cada categoría por separado sino todas en conjunto.

Por ser un modelo de Rasch, los parámetros de los ítems son separables de los parámetros de las personas, si bien están en una misma escala, y existen estadísticos suficientes para su estimación. Este modelo puede aplicarse en test de capacidad. Los ítems pueden tener un formato de respuesta abierta y puntuarse con valores: 0 (respuesta incorrecta), 1 (parcialmente correcta) y 2 (correcta). En un test de actitudes puede utilizarse una escala cerrada con categorías ordenadas (en desacuerdo, neutral, de acuerdo). Este modelo también puede aplicarse a ítems dicotómicos que se refieren a todos ellos a un mismo texto. Por ejemplo, en un test de comprensión lectora, la persona debe leer un fragmento de texto y responder a cinco preguntas que se puntúan cada una como 0 y 1. En este conjunto de cinco ítems, es bastante probable que no se cumpla el supuesto de independencia local, que requiere la TRI, por basarse todos sobre el mismo texto. Una solución al problema es un tratamiento politómico del ítem, en el que se puntúa 0 si hacen mal las cinco preguntas; 1, si hace una bien;… 5, si hace las cinco bien.

A modo de ejemplo, supongamos el test de matemáticas en el que aparezca el siguiente ítem con un formato de respuesta abierta:

Calcule la integral

$$\boxed{I = \int x e^x dx}$$

La solución se comprende de los siguientes cuatro pasos:

Paso 1º: Identificar los factores

$$u = x$$
$$dv = e^x dx$$

Paso 2º: Calcular

$$du = dx$$
$$v = e^x$$

Paso 3º: Aplicar la integración por partes

$$\begin{aligned} \int u\,dv &= uv - \int v\,du \\ &= xe^x - \int e^x dx \\ &= (x-1)e^x \end{aligned}$$

Paso 4º: Añadir constante de integración

$$I == (x-1)e^x + K$$

Este ítem se puntúa de 0 a 4, según cuantos pasos haya completado el sujeto correctamente. Además los pasos han de ser aplicados en forma sucesiva. Por ejemplo, no es posible completar correctamente el paso 3 sin haber completado antes correctamente el paso 1 y 2. Supongamos que un sujeto aplica correctamente el paso 3 pero lo hace sobre una función incorrecta como resultado debe haberse equivocado en el paso 2. En este caso su puntuación sería 1, por haber acertado el paso 1 y fallado en el 2; es decir, una vez que el sujeto se ha equivocado en el paso 2, se considera incorrecto todo el resto del problema.

Por tanto, el MCP puede aplicarse cuando la resolución de un ítem implica realizar varios pasos sucesivos, y la puntuación recoge cuantos pasos se han completado correctamente.

2.3.2 DESCRIPCIÓN DEL MODELO

El modelo asume que la probabilidad de que el sujeto complete cada paso puede formalizarse mediante el modelo de Rasch. Supongamos un ítem *i* de 3 pasos en el que las posibles puntuaciones son de 0, 1, 2 y 3.

Solo las personas que hayan acertado los pasos 1 y 2 se enfrentan al paso 3. La probabilidad de acertarlo ϕ_3, o la probabilidad de que esas personas obtengan un 3 en vez de un 2, será el cociente entre la probabilidad de que den el paso y obtengan un 3 $(P(3))$ y la suma de $P(2)$ y $P(3)$. Se asume, además que dicha probabilidad viene dada por el modelo de Rasch con parámetro b_{i3}. Es decir,

$$\phi_3 = \frac{P(3)}{P(2)+P(3)} = \frac{e^{(\theta_v - b_{i3})}}{1+e^{(\theta_v - b_{i3})}} \tag{2.52}$$

Siendo θ_v el nivel del rasgo para el sujeto *v*, b_{i3} la dificultad del paso 3, y $P(X=h)$, en lo sucesivo $P(h)$, la probabilidad de que la persona obtenga, la puntuación *h* al enfrentarse al paso 3 (*h*=2,3).

Supongamos que una persona ha completado correctamente el paso 1. Al enfrentarse al paso 2 puede acertar o fallar. La probabilidad de acertar, ϕ_2, que será la probabilidad de que obtenga un 2 en vez de 1, se asume que viene dada por el modelo de Rasch con parámetro b_{i2}.

Es decir,

$$\phi_2 = \frac{P(2)}{P(1)+P(2)} = \frac{e^{(\theta_v - b_{i2})}}{1+e^{(\theta_v - b_{i2})}} \tag{2.53}$$

Análogamente, cuando una persona se enfrente al primer paso, puede acertarlo o fallarlo. La probabilidad de acertarlo, ϕ_1, que será la probabilidad de que obtenga un 1 en vez de un 0, se asume que viene dada por el modelo de Rasch con parámetro b_{i1}.

Es decir,

$$\phi_1 = \frac{P(1)}{P(0)+P(1)} = \frac{e^{(\theta_v - b_{i1})}}{1+e^{(\theta_v - b_{i1})}} \tag{2.54}$$

De la expresión anterior se sigue que:

$$P(1) = \left[e^{(\theta_v - b_{i1})}\right]P(0) \tag{2.55}$$

En efecto tomando los recíprocos en (2.54), queda

$$\frac{1}{\phi_1} = \frac{P(0)+P(1)}{P(1)} = \frac{1+e^{(\theta_v - b_{i1})}}{e^{(\theta_v - b_{i1})}}$$
$$\frac{1}{\phi_1} = \frac{P(0)}{P(1)} + 1 = \frac{1}{e^{(\theta_v - b_{i1})}} + 1$$

De donde se obtiene (2.55).

En las fórmulas (2.52), (2.53) y (2.54), aplicando el proceso de (2.55), se obtiene que

$$\begin{aligned} P(1) &= \left[e^{(\theta_v - b_{i1})}\right]P(0) \\ P(2) &= \left[e^{(\theta_v - b_{i2})}\right]P(1) \\ P(3) &= \left[e^{(\theta_v - b_{i3})}\right]P(2) \end{aligned} \tag{2.56}$$

Además, dado que $\sum_{h=0}^{3} P(h) = 1$, como vamos a ver a continuación, con algo de álgebra, nos queda la expresión general del modelo. Las tres igualdades anteriores pueden también ponerse así:

$$\begin{aligned} P(1) &= \left[e^{(\theta_v - b_{i1})}\right]P(0) \\ P(2) &= \left[e^{(\theta_v - b_{i2})}\right]P(1) = \left[e^{(\theta_v - b_{i2})}\right]\left[e^{(\theta_v - b_{i1})}\right]P(0) \\ P(3) &= \left[e^{(\theta_v - b_{i3})}\right]P(2) = \left[e^{(\theta_v - b_{i3})}\right]\left[e^{(\theta_v - b_{i2})}\right]\left[e^{(\theta_v - b_{i1})}\right]P(0) \end{aligned} \tag{2.57}$$

En general,

$$P(h) = \left[e^{\sum_{r=1}^{h}(\theta_v - b_{ir})} \right] P(0) \tag{2.58}$$

Además,

Como $P(0) + P(1) + P(2) + P(3) = 1$, tendremos que

$$P(0) + \left[e^{\sum_{r=1}^{1}(\theta_v - b_{ir})} \right] P(0) + \left[e^{\sum_{r=1}^{2}(\theta_v - b_{ir})} \right] P(0) + \left[e^{\sum_{r=1}^{3}(\theta_v - b_{ir})} \right] P(0) = 1 \tag{2.59}$$

O, que es lo mismo,

$$P(0) \left[1 + e^{\sum_{r=1}^{1}(\theta_v - b_{ir})} + e^{\sum_{r=1}^{2}(\theta_v - b_{ir})} + e^{\sum_{r=1}^{3}(\theta_v - b_{ir})} \right] = 1 \tag{2.60}$$

De donde se obtiene que

$$P(0) = \frac{1}{1 + e^{\sum_{r=1}^{1}(\theta_v - b_{ir})} + e^{\sum_{r=1}^{2}(\theta_v - b_{ir})} + e^{\sum_{r=1}^{3}(\theta_v - b_{ir})}} \tag{2.61}$$

Para hacer más compacta la notación se define $\theta_v - b_{ir} = 0$, por lo que la expresión (2.61) se puede escribir como:

$$P(0) = \frac{1}{\sum_{l=0}^{3} e^{\sum_{r=0}^{l}(\theta_v - b_{ir})}} \tag{2.62}$$

Teniendo en cuenta (2.58), nos queda:

$$P(h) = \frac{e^{\sum_{r=0}^{h}(\theta_v - b_{ir})}}{\sum_{l=0}^{3} e^{\sum_{r=0}^{l}(\theta_v - b_{ir})}} \tag{2.63}$$

En general para un ítem puntuado con *m+1* posibles valores, la probabilidad del resultado *h* (*h=0, 1,…, m*) en el modelo de Crédito Parcial es:

$$P(h)=\frac{e^{\sum_{r=0}^{h}(\theta_v-b_{ir})}}{\sum_{l=0}^{m}e^{\sum_{r=0}^{l}(\theta_v-b_{ir})}} \tag{2.64}$$

Con $\theta_v - b_{ih} = 0$.

Esta probabilidad tiene dos interpretaciones: puede entenderse como la probabilidad de que una persona, con nivel de rasgo θ_v, produzca cualquiera de las *m+1* respuestas, o como la proporción de personas que respondan determinada opción del total de personas con nivel de rasgo θ_v.

Ejemplo:

Los parámetros de un ítem *i* de 4 categorías son $b_{i1} = -0.5, b_{i2} = 1.2,$ y $b_{i3} = 1.5$. Las probabilidades de puntuar 0, 1, 2 y 3 para una persona con nivel de rasgo $\theta = 1$, serán:

$$P(0)=\frac{1}{1+e^{(1+0.5)}+e^{(1+0.5)+(1-1.2)}+e^{(1+0.5)+(1-1.2)+(1-1.5)}}=0.09$$

$$P(1)=\frac{e^{(1+0.5)}}{1+e^{(1+0.5)}+e^{(1+0.5)+(1-1.2)}+e^{(1+0.5)+(1-1.2)+(1-1.5)}}=0.39$$

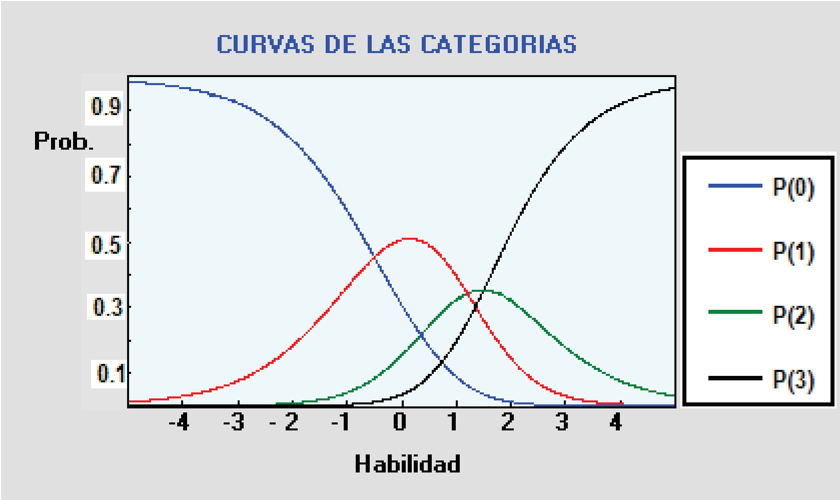
$$P(2)=\frac{e^{(1+0.5)+(1-1.2)}}{1+e^{(1+0.5)}+e^{(1+0.5)+(1-1.2)}+e^{(1+0.5)+(1-1.2)+(1-1.5)}}=0.32$$

$$P(3)=\frac{e^{(1+0.5)+(1-1.2)+(1-1.5)}}{1+e^{(1+0.5)}+e^{(1+0.5)+(1-1.2)}+e^{(1+0.5)+(1-1.2)+(1-1.5)}}=0.20$$

Cada persona ha de dar una de las 4 respuestas. Por lo tanto $\sum_{h=0}^{3} P(h) = 1$. La dificultad del ítem se obtiene promediando la dificultad de las categorías, para este ítem su dificultad es 0.7333.

Procediendo así para los demás valores de θ_v, obtendríamos las funciones de respuesta a las categorías (FRC), que muestra la siguiente figura:

Figura 2.21: Funciones de Respuesta de las Categóricas para
$b_{i1} = -0.5, b_{i2} = 1.2,$ y $b_{i3} = 1.5$



2.3.3 SIGNIFICADO DE LOS PARÁMETROS

El MCP se aplica al modelo de Rasch para la resolución de cada paso de un problema determinado. Por tanto, b_h puede entenderse como la *dificultad del paso* "*h*". En el ejemplo numérico anterior, $b_{i1} = -0.5$. Las personas con ese nivel de rasgo tienen una probabilidad de 0.5 de no dar el paso 1 (y puntuar 0 en el ítem) o darlo (y enfrentarse al segundo paso). Diríamos que es un paso fácil. El nivel de rasgo que han de tener los que den el paso 1 para tener una probabilidad de 0.5 de no dar el paso 2 (y puntuar 1) o darlo correctamente (y enfrentarse al paso 3), es de 1.2 (pues $b_{i2} = 1.2$); por lo que este paso es difícil; aun lo es más el paso 3, pues $b_{i3} = 1.5$. En este ejemplo, los tres valores de b_{ir} están ordenados de, menor a mayor; pero a diferencia de otros modelos de la TRI, en este modelo los valores de b_{ir} no han de estar necesariamente ordenados. Los pasos deben estar en secuencia, pues solo se enfrenta uno al paso *h* si se ha hecho bien el paso *h-1*, pero sus dificultades no han de estar ordenadas.

El valor b_h es también el punto de corte de las curvas que corresponden a las categorías $h-1$ y h. En efecto, igualando $P(h-1)$ a $P(h)$, tenemos:

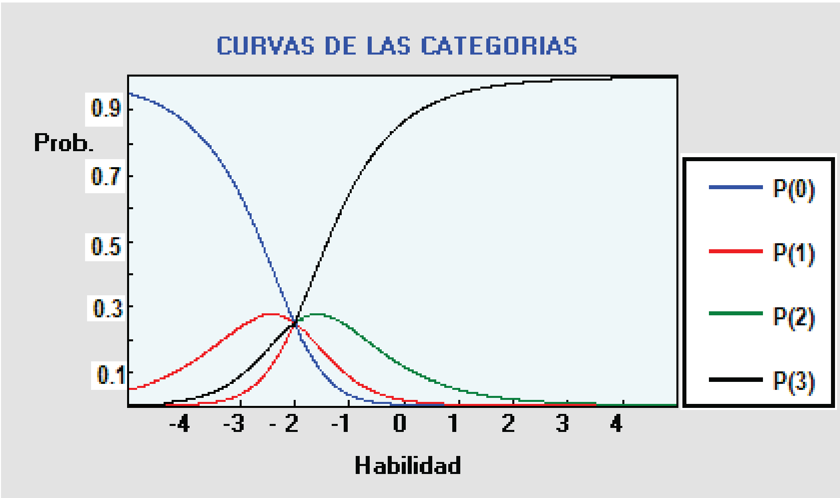
$$e^{\sum_{j=0}^{h-1}(\theta - b_j)} = e^{\sum_{j=0}^{h}(\theta - b_j)} \qquad (2.65)$$

De la expresión anterior se sigue que $(\theta - b_h) = 0$.

Por tanto, en $\theta = b_h$ coinciden ambas curvas. En el ejemplo numérico las personas con $\theta = -0.5$ tienen la misma probabilidad de puntuar 0 ó 1. Las personas con $\theta = 1.2$ tienen la misma probabilidad de puntuar 1 ó 2, y las personas con $\theta = 1.5$ tienen la misma probabilidad de puntuar 2 ó 3.

Supongamos un ítem cuyos tres pasos tengan la misma dificultad. En la figura 2.22 los tres valores de b_{ir} han sido -2.

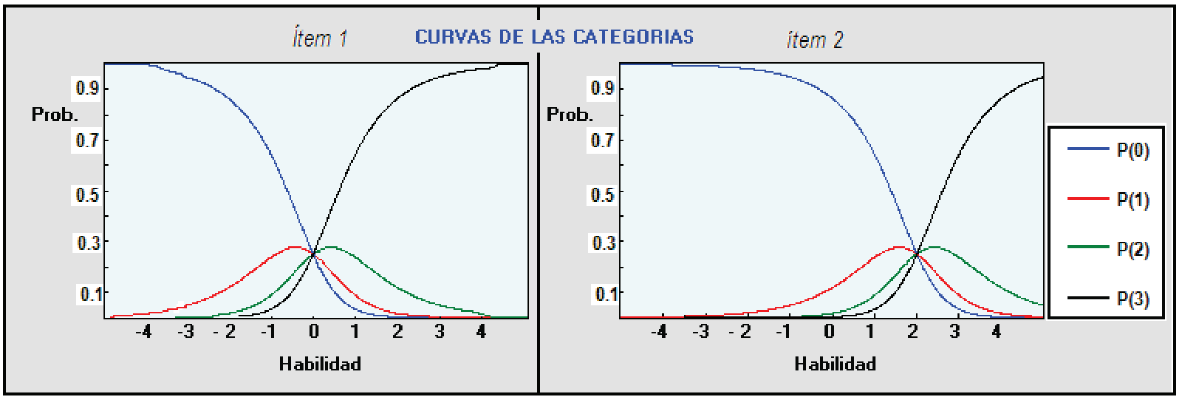
Figura 2.22: Funciones de Respuesta de un Ítem con $b_{i1} = b_{i2} = b_{i3} = -2$



La figura muestra que las personas con un nivel de rasgo inferior a -2 lo más probable es que obtenga cero puntos. La probabilidad que obtenga cero puntos es tanto mayor cuanto menor sea nivel de rasgo. Si tiene un nivel de rasgo inferior a -2 es poco probable de acertar el paso 1, que resulta difícil para estos sujetos (pues $\theta < b_{i1} = -2$). Será menos probable acertar el paso 1 y fallar el dos (es decir, puntuar 1); y menos probable acertar el 1, acertar el 2 y fallar el tres (es decir, puntuar 2); y aun menos acertar los tres pasos (es decir, puntuar 3). La figura 2.22 muestra que, para las personas con niveles de rasgo menores a -2, las funciones de respuesta se ordenan como se acaba de indicar. Las personas con nivel superior a -2 lo más probable es que obtengan tres puntos. La probabilidad es un tanto mayor cuanto mayor sea su θ. Las personas que completen el paso 1 se enfrentan al paso 2, que es muy fácil (pues, $\theta > b_{i2} = -2$), por lo que es muy probable que lo acierten y que se enfrenten en al paso 3, que vuelve a ser muy fácil (pues $\theta > b_{i3} = -2$) y será muy probable que también lo acierten, con lo que su puntuación en el ítem será muy probablemente 3. En resumen se trataría de un ítem muy fácil, pues las mayorías de las personas obtendrían un 3 en este ítem, dado que las mayorías de las personas tienen niveles de rasgo mayores a -2 y de estas, la mayoría tendría un 3.

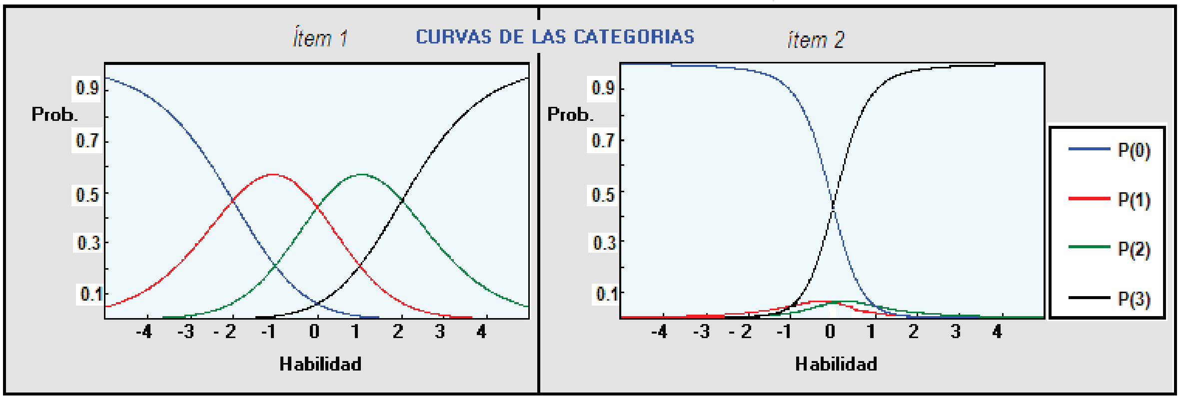
La figura 2.23 muestra las FRCs de dos nuevos ítems. En cada uno coinciden los tres valores de b_{ir}. Las nuevas FRCs coinciden con la mostrada en la figura 2.22, si bien ahora están centradas en 0 y en 2, que son valores de b_{ir} comunes. Podemos también comprobar que las cuatro curvas se cruzan en dichos valores. El ítem de la izquierda tendría dificultad media y el de la derecha dificultad alta.

Figura 2.23: FRCs de dos Ítems
Ítem 1: $b_{11} = b_{12} = b_{13} = 0$
Ítem 2: $b_{21} = b_{22} = b_{23} = 2$



En la siguiente figura se muestra la importancia del orden de los valores de b_{ir} en dos ítems.

Figura 2.24: FRCs de dos Ítems
Ítem 1: $b_{11} = -2,\ b_{12} = 0,\ b_{13} = 2$
Ítem 2: $b_{21} = 2,\ b_{22} = 0,\ b_{23} = -2$



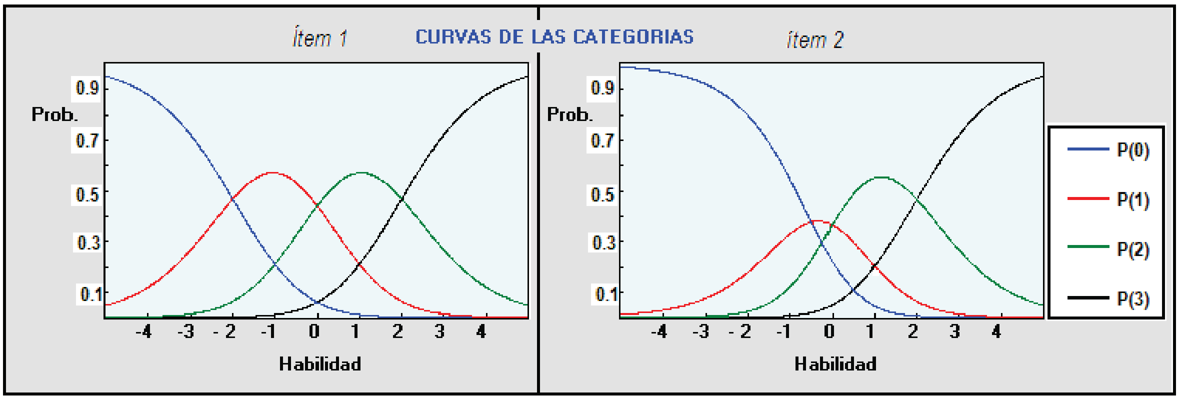
En el gráfico de la izquierda de la figura 2.24 tenemos los valores de $b_{i1} = -2$, $b_{i2} = 0$, $b_{i3} = 2$ de las FRCs los cuales están ordenados de menor a mayor dificultad. En este caso, las cuatro categorías son máximamente probables para ciertos rangos de θ. En concreto, para $\theta < b_{i1} = -2$, la categoría más probable es la 0. Para $b_{i1} = -2 < \theta < 0 = b_{i2}$, la más probable es la 1. Para $b_{i2} = 0 < \theta < 2 = b_{i3}$, la más probable es la 2. Para $\theta > 2 = b_{i3}$, la más probable es la 3. Cuando los valores de b_{ir} no están ordenados de menor a mayor dificultad, hay categorías que no son máximamente probables para ningún valor de θ. En la misma figura 2.24, el gráfico de la derecha, con valores $b_{i1} = 2$, $b_{i2} = 0$, $b_{i3} = -2$, las FRCs de las categorías 1 y 2 no son máximamente probables para ningún valor de θ.

Si cambiamos el parámetro -2 por -0.5, la figura 2.25 muestra que la curva de la primera categoría se desplaza a la derecha, indicando que aumenta la probabilidad de obtener la puntuación 0 en el ítem y que disminuye la probabilidad de obtener 1, 2 y 3, muy especialmente la primera. Nótese que pasar de -2 a -0.5 supone incrementar la dificultad de este primer paso. Cuando el parámetro b_{ir} del primer paso es -2, cabe esperar que casi todos los sujetos den el primer paso y que solo la mitad de los que lo han dado den también el segundo ($b = 0$) quedando entonces muchos sujetos con calificación de 1 en el ítem. Sin embargo, cuando aumentamos el parámetro -0.5, son mucho menos los que dan el primer paso, por lo que serán también mucho menos los que queden con la calificación de 1.

Figura 2.25: FRCs de Dos Ítems

Ítem 1: $b_{11} = -2,\ b_{12} = 0\ \ y\ \ b_{13} = 2$

Ítem 2: $b_{21} = -0.5,\ b_{22} = 0\ \ y\ \ b_{23} = 2$



2.3.4 FUNCIÓN DE INFORMACIÓN (FI)

La función de información del ítem nos dice cuanto ayuda el ítem a estimar cada nivel de rasgo. Por lo tanto, se interpreta un ítem si consideramos simultáneamente sus FRCs y su FI.

Siguiendo la definición expuesta en la sección 2.2.8 del tema del modelo dicotómico de Rasch, se llega a que la función de información de un ítem *i* es:

$$I_j(\theta) = \sum_{h=0}^{m} \frac{[P'(h)]^2}{P(h)} \tag{2.66}$$

En esta expresión, *P'* es la primera derivada de *P*, donde *P* viene dada por (2.64). La expresión que resulta al calcular la función de información de dicho modelo es:

$$I_j(\theta) = \sum_{h=0}^{m} \left\{ \frac{e^{\sum_{r=0}^{h}(\theta_v - b_{ir})} \left[h \sum_{h=0}^{m} e^{\sum_{r=0}^{h}(\theta_v - b_{ir})} - \sum_{h=0}^{m} l e^{\sum_{r=0}^{h}(\theta_v - b_{ir})} \right]^2}{\left[\sum_{l=0}^{m} e^{\sum_{r=0}^{l}(\theta_v - b_{ir})} \right]^3} \right\} \tag{2.67}$$

Las figuras (2.26) y (2.27) muestran las FRCs y la FIs de dos ítems:

Figura 2.26: FRCs y FI de un Ítem con $b_{11} = -2 \ y \ b_{12} = 4$

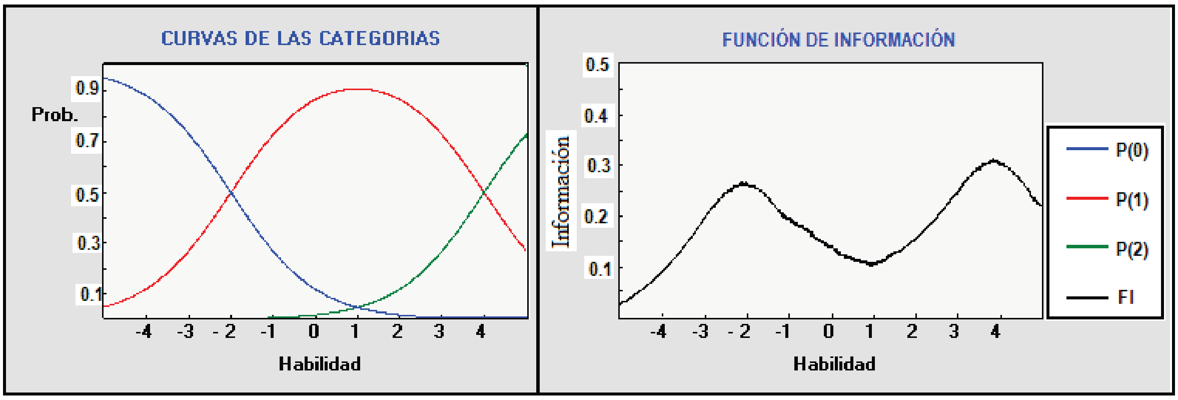
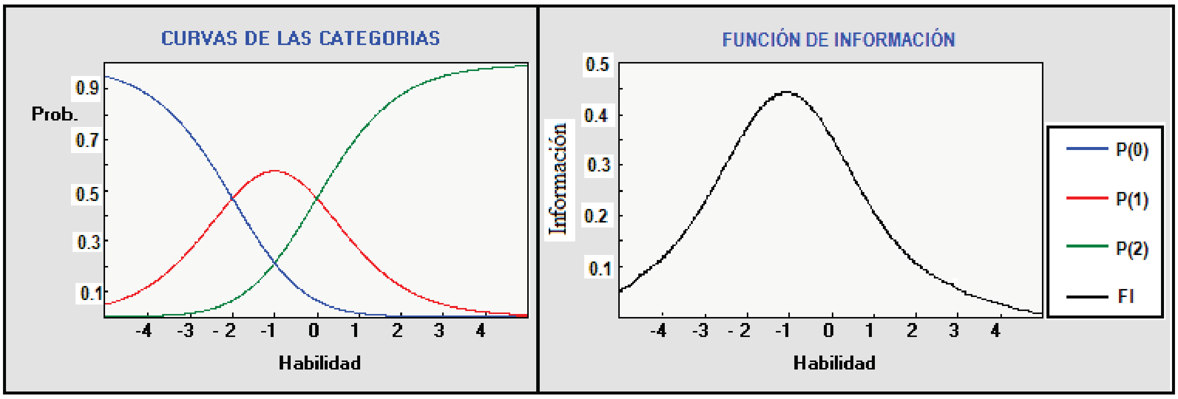


Figura 2.27: FRCs y FI de un Ítem con $b_{21} = -2 \ y \ b_{22} = 0$



Las figuras muestran que la información se acumula en los niveles de rasgo próximos a los valores de b_{ir}. Sin embargo, la relación entre dichos valores y donde ocurre la información máxima depende de varias características de dichos valores de b_{ir} como:

1. El rango o diferencia entre el valor mayor y menor de b_{ir}
2. Del número de valores de b_{ir} que no están ordenados secuencialmente, y
3. De la diferencia entre estos valores b_{ir}.

2.3.5 ESTIMACIÓN DE PARÁMETROS PARA EL MODELO POLITÓMICO DE RASCH

Como se ha mencionado anteriormente, el MCP es un modelo politómico de Rasch; sin embargo en los trabajos realizados por George Rasch (1961), describe una generalización del modelo politómico con parámetros unidimensionales tanto para los sujetos e ítems, el cual la principal justificación es la separabilidad de los parámetros[3]. A partir de la expresión (2.64), podemos observar que el modelo MCP puede ser llevado al modelo descrito por Rasch (2.68). Por lo tanto, para poder estimar los parámetros de los sujetos e ítems, se trabajará con la expresión general descrita por Rasch y de esta manera se obtendrá una forma general de estimar los parámetros de los sujetos e ítems. El modelo politómico de Rasch se observa en la siguiente expresión:

[3] Si el lector esta interesado en obtener mayor información en cuanto a la estimación de los parámetros y otros modelos puede consultar el libro “RASCH MODELS, foundations, recent developments and applications”, año 1995.

$$P(X_{vi}=h)=\frac{e^{(w_h\theta_v+\beta_{ih})}}{\sum_{l=0}^{m}e^{(w_l\theta_v+\beta_{il})}} \tag{2.68}$$

donde

h : es la respuesta observada en la categoría *h* para el sujeto *v* en el ítem *i*.

β_{ih} : es el parámetro de dificultad del ítem *i* en la categoría *h*

θ_v : el parámetro de habilidad del sujeto *v*

w : nos indica los pesos de las categorías (este generalmente es fijado por el diseñador del test)

Generalmente los pesos de las categorías (w) son fijados por el diseñador del test, pero esto no impide que también puedan ser estimados mediante el método de máxima verosimilitud. Es evidente que la expresión (2.64) puede llevarse a la forma del modelo descrito por Rasch en la expresión anterior. Observamos que el numerador de la expresión (2.68) se encuentra en función de la categoría que ha sido escogida por el sujeto en un ítem determinado, y el denominador es la suma de todas las alternativas del ítem (al igual que la expresión 2.64).

Además Rasch sugiere que para la estimación de los parámetros de los ítems se utilice el método de máxima verosimilitud condicional (CML) condicionado a los parámetros individuales de los sujetos θ , y para la estimación de los parámetros de los sujetos se puede hacer mediante el método de máxima verosimilitud; estos dos métodos se complementan con la aplicación del algoritmo de Newton-Raphson.

A continuación se presenta el procedimiento para la estimación de los parámetros de los sujetos y seguidamente se presenta el método propuesto por Rasch para la estimación de los parámetros de los ítems.

2.3.5.1 Estimación de los Parámetros de los Sujetos del Modelo Politómico de Rasch

Asumiendo independencia local para la expresión (2.68), tenemos que la función de máxima verosimilitud es:

$$L=\prod_{v=1}^{n}\prod_{i=1}^{k}\prod_{h=0}^{m}(P_{vih})^{x_{vih}} \qquad (2.69)$$

donde

n : Número de sujetos

k : Número de ítems

m : Máxima opción en el ítem

Al aplicar logaritmo natural nos da la siguiente expresión:

$$\ln L=\sum_{v=1}^{n}\sum_{i=1}^{k}\sum_{h=0}^{m}(x_{vih})\left[\ln(P_{vih})\right] \qquad (2.70)$$

Derivando la expresión anterior con respecto a θ_v tenemos:

$$\frac{\partial \ln L}{\partial \theta_v}=\sum_{i=1}^{k}\sum_{h=0}^{m}\left(\frac{x_{vih}}{P_{vih}}\right)\left(\frac{\partial P_{vih}}{\partial \theta_v}\right) \qquad (2.71)$$

La expresión para la derivada de $\frac{\partial P_{vih}}{\partial \theta_v}$ está dada por:

$$\frac{\partial P_{vih}}{\partial \theta_v}=\frac{\partial}{\partial \theta_v}\left[\frac{e^{(w_h\theta_v+\beta_{ih})}}{\sum_{l=0}^{m}e^{(w_l\theta_v+\beta_{il})}}\right]$$

$$\frac{\partial P_{vih}}{\partial \theta_v}=\frac{e^{(w_h\theta_n+\beta_{ih})}\left[(w_h)\sum_{l=0}^{m}e^{(w_l\theta_v+\beta_{il})}-\sum_{l=0}^{m}(w_l)e^{(w_l\theta_v+\beta_{il})}\right]}{\left(\sum_{l=0}^{m}e^{(w_l\theta_v+\beta_{il})}\right)^2} \qquad (2.72)$$

Ahora:

$$\frac{\partial \ln L}{\partial \theta_v} = \sum_{i=1}^{k}\sum_{h=0}^{m}\left\{\frac{x_{vih}}{\frac{e^{(w_h\theta_v+\beta_{ih})}}{\sum_{l=0}^{m} e^{(w_l\theta_v+\beta_{il})}}}\right\}\left\{\frac{e^{(w_h\theta_v+\beta_{ih})}\left[(w_h)\sum_{l=0}^{m} e^{(w_l\theta_v+\beta_{il})} - \sum_{l=0}^{m}(w_l)e^{(w_l\theta_v+\beta_{il})}\right]}{\left(\sum_{l=0}^{m} e^{(w_l\theta_v+\beta_{il})}\right)^2}\right\} \quad (2.73)$$

$$\frac{\partial \ln L}{\partial \theta_v} = \sum_{i=1}^{k}\sum_{h=0}^{m}\left\{\frac{x_{vih}\left[(w_h)\sum_{l=0}^{m} e^{(w_l\theta_v+\beta_{il})} - \sum_{l=0}^{m}(w_l)e^{(w_l\theta_v+\beta_{il})}\right]}{\left(\sum_{l=0}^{m} e^{(w_l\theta_v+\beta_{il})}\right)}\right\}$$

Igualando a cero (2.73) se tiene la siguiente ecuación normal:

$$\frac{\partial \ln L}{\partial \theta_v} = \sum_{i=1}^{k}\sum_{h=0}^{m}\left\{\frac{x_{vih}\left[(w_h)\sum_{l=0}^{m} e^{(w_l\theta_v+\beta_{il})} - \sum_{l=0}^{m}(w_l)e^{(w_l\theta_v+\beta_{il})}\right]}{\left(\sum_{l=0}^{m} e^{(w_l\theta_v+\beta_{il})}\right)}\right\} = 0 \quad ; v = 1,2,...,n \qquad (2.74)$$

Observamos que para la estimación de los parámetros de los sujetos resulta muy difícil obtener expresiones explicitas para los estimadores de los sujetos, por lo tanto el algoritmo de Newton-Rapshon es una opción numérica para la obtención de los estimadores.

Basándonos en la expresión (2.38) para el caso de datos dicotómicos, tenemos la siguiente expresión:

$$\frac{\partial \ln L}{\partial \theta_v} = \frac{\partial \ln L}{\partial \theta_v\big|_{\theta_v^0,\beta_{ih}^0}} + \frac{\partial^2 \ln L}{\partial \theta_v^2\big|_{\theta_v^0,\beta_{ih}^0}}(\theta_v - \theta_v^0)$$

$$\theta_v = \theta_v^0 - \left[\frac{\partial \ln L}{\partial \theta_v\big|_{\theta_v^0,\beta_{ih}^0}}\right]\left[\frac{\partial^2 \ln L}{\partial \theta_v^2\big|_{\theta_v^0,\beta_{ih}^0}}\right]^{-1} \qquad (2.75)$$

De la expresión (2.74) se encontrará $\left(\frac{\partial^2 \ln L}{\partial \theta_v^2\big|_{\theta_v^0,\beta_{ih}^0}}\right)$, de la siguiente manera:

$$\frac{\partial}{\partial\theta_v}\left[\frac{\partial \ln L}{\partial\theta_v}\right]=\frac{\partial}{\partial\theta_v}\left\{\sum_{i=1}^{k}\sum_{h=0}^{m}\left[\frac{x_{vih}\left((w_h)\sum_{l=0}^{m}e^{(w_l\theta_v+\beta_{il})}-\sum_{l=0}^{m}(w_l)e^{(w_l\theta_v+\beta_{il})}\right)}{\left(\sum_{l=0}^{m}e^{(w_l\theta_v+\beta_{il})}\right)}\right]\right\} \tag{2.76}$$

Obteniendo:

$$\frac{\partial^2 \ln L}{\partial\theta_v^2\Big|_{\theta_v^0,\beta_{ih}^0}}=\sum_{i=1}^{k}\sum_{h=0}^{m}\left\{\frac{x_{vih}\left[(w_h)(A)(B)-(A)(C)-(w_h)(A)(B)+(B)^2\right]}{(A)^2}\right\} \tag{2.77}$$

Donde:

$$A=\left(\sum_{l=0}^{m}e^{(w_l\theta_v+\beta_{il})}\right)$$

$$B=\left(\sum_{l=0}^{m}(w_l)e^{(w_l\theta_v+\beta_{il})}\right)$$

$$C=\left(\sum_{l=0}^{m}(w_l)^2\,e^{(w_l\theta_v+\beta_{il})}\right)$$

Finalmente la estimación de los parámetros para los sujetos se obtiene de la siguiente ecuación:

$$\theta_v=\theta_v^0-\frac{\sum_{i=1}^{k}\sum_{k=1}^{m}\left[\frac{x_{vih}\left((w_h)\sum_{l=0}^{m}e^{(w_l\theta_v+\beta_{il})}-\sum_{l=0}^{m}(w_l)e^{(w_l\theta_v+\beta_{il})}\right)}{\left(\sum_{l=0}^{m}e^{(w_l\theta_v+\beta_{il})}\right)}\right]}{\sum_{i=1}^{k}\sum_{h=0}^{m}\left\{\frac{x_{vih}\left[(w_h)(A)(B)-(A)(C)-(w_h)(A)(B)+(B)^2\right]}{(A)^2}\right\}} \tag{2.78}$$

Donde A, B, y C son los mismos descritos en la expresión (2.77).

2.3.5.2 Estimación de Parámetros de los Ítems del Modelo Politómico de Rasch

Para estimar los parámetros de los ítems partiremos de la expresión (2.68). Observamos que el denominador no depende de las respuestas observadas *h*, lo que significa que es solo una función de del vector $\beta_i=(\beta_{i0},...,\beta_{im})$ y θ_v. Por lo tanto podemos poner:

$$C(\theta_v, \beta_i) = \sum_{l=0}^{m} e^{(w_l\theta_v + \beta_{il})} \tag{2.79}$$

Conviene escribir las respuestas reales para el individuo S_v como la selección del vector $(x_{vi0}, ..., x_{vim})$, donde $x_{vih} = 1$, si la respuesta es *h*, y 0 en el caso contrario. Por la independencia local tenemos que la función de verosimilitud es:

$$L = \prod_{v=1}^{n}\prod_{i=1}^{k}\prod_{h=0}^{m} \frac{e^{x_{vih}(w_h\theta_v + \beta_{ih})}}{C(\theta_v, \beta_i)} \tag{2.80}$$

Al aplicar el logaritmo natural tenemos el siguiente resultado:

$$\begin{aligned} \ln L &= \sum_{v=1}^{n}\sum_{i=1}^{k}\sum_{k=1}^{m}(w_h\theta_v + \beta_{ih})x_{vih} - \sum_{v=1}^{n}\sum_{i=1}^{k}\ln C(\theta_v, \beta_i) \\ \ln L &= \sum_{v=1}^{n}\theta_v\sum_{h=0}^{m} w_h x_{v.h} + \sum_{i=1}^{k}\sum_{h=0}^{m}\beta_{ih}x_{.ih} - \sum_{v=1}^{n}\sum_{i=1}^{k}\ln C(\theta_v, \beta_i) \end{aligned} \tag{2.81}$$

Donde

$$x_{v.h} = \sum_{i=1}^{k} x_{vih} \qquad y \qquad x_{.ih} = \sum_{v=1}^{n} x_{vih}$$

Así, acorde a los resultados estándar para la familia exponencial, la puntuación

$$r_v = \sum_{h=0}^{m} w_h x_{v.h}$$

Es un estadístico minimal suficiente para θ_v. Similarmente, los totales del ítem $x_{.ih}$ son minimal suficientes para los parámetros de los ítems β_{ih}. Los estimadores de máxima verosimilitud son obtenidos como la solución simultanea de las ecuaciones (2.82) y (2.83),

$$r_v = E[R_v], \text{ para } v = 1, 2, ..., n \tag{2.82}$$

Con

$$R_v = \sum_{h=0}^{m} w_h X_{v.h},$$

y

$$x_{.ih} = E[X_{.ih}], \; para\, i = 1, 2, ..., k; \quad h = 0, 1, 2, ..., m \tag{2.83}$$

Andersen (1970), demuestra que los estimadores de máxima verosimilitud para los parámetros de los ítems son inconsistentes cuando *n* es grande y *k* y *m* son fijos.

Como una alternativa Rasch sugiere la estimación de los parámetros de los ítems por medio del método de máxima verosimilitud condicional (CML), donde la condición es con respecto a los estadísticos suficientes para los parámetros individuales $\theta_1, \theta_2, ..., \theta_n$; bajo esta condición los estimadores CML son consistentes.

La distribución condicional de la matriz $\{X_{.ih}\}$, $i = 1, 2, ..., k$; $h = 0, 1, 2, ..., m$, dado los estadísticos minimal suficiente $r_1, ...r_n$ para $\theta_1, \theta_2, ..., \theta_n$ es nuevamente de la familia exponencial. Los totales de los ítems son suficientes para β también en este modelo condicional. Por lo tanto, los estimadores de CLM son (por la teoría de la familia exponencial estándar) obtenidas como soluciones de las ecuaciones:

$$x_{.ih} = E\left[X_{.ih} / r_1, ...r_n\right], \ \ para\, i = 1, 2, ..., k \ ; h = 0, 1, 2, ..., m \tag{2.84}$$

donde los valores medios $E\left[... / r_1, ...r_n\right]$ referidos a la distribución condicional dada las puntuaciones $r_1, ...r_n$.

La derivación de estas ecuaciones es simple. Puesto que,

$$X_{.ih} = \sum_{v=1}^{m} X_{vih}$$

Solo necesitamos derivar los valores medios de X_{vih}, dada la puntuación individual r_v. La probabilidad conjunta de $\{X_{vih}\}$, para $i = 1, ..., k$ y $h = 0, ..., m$, es:

$$\begin{aligned}
f(x_{v10}, ..., x_{vkm}) &= P(X_{n10} = x_{n10}, ...X_{vkm} = x_{vkm}) \\
f(x_{v10}, ..., x_{vkm}) &= \prod_{i=0}^{k} \prod_{h=0}^{m} \left[e^{\left(w_h \theta_v + \beta_{ih}\right)} \right]^{x_{vih}} C^{-1}(\theta_v, \beta_i) \\
f(x_{v10}, ..., x_{vkm}) &= e^{\left(r_v \theta_v + \sum_{i=1}^{k} \sum_{h=0}^{m} \beta_{ih} x_{vih} \right)} \prod_{i=1}^{k} C^{-1}(\theta_v, \beta_i)
\end{aligned} \tag{2.85}$$

Por lo tanto, la probabilidad marginal de la puntuación $r_v = \sum_{i=1}^{k} \sum_{h=0}^{m} w_h x_{vih}$ es:

$$f(r_v) = e^{(r_v \theta_v)} \sum_{(rv)} e^{\left(\sum_{i=1}^{k} \sum_{h=0}^{m} \beta_{ih} x_{vih} \right)} \prod_{i=1}^{k} C^{-1}(\theta_v, \beta_i) \tag{2.86}$$

donde (r_v) es sobre todos los patrones factibles de respuesta $\{x_{vih}\}$ para los cuales el puntaje es r_v. La suma en (2.86) es denotada por:

$$\gamma_{r_v}(\beta) = \sum_{(rv)} e^{(\sum_{i=1}^{k}\sum_{h=0}^{m}\beta_{ih}x_{vih})} \tag{2.87}$$

definiéndolo así, la función γ, la probabilidad condicional de las respuestas $(x_{v10},...,x_{vkm})$, dado r_v para el individuo S_v, es obtenida de (2.85) y (2.86) como:

$$f(x_{v10},...,x_{vkm}/r_v) = \frac{e^{\sum_{i=1}^{k}\sum_{h=0}^{m}\beta_{ih}x_{vih}}}{\gamma_{r_v}(\beta)} \tag{2.88}$$

De esta expresión se sigue inmediatamente que:

$$P(X_{vih}=1/r_v) = e^{(\beta_{ih})}\frac{\sum_{(r_v-w_h)} e^{\left(\sum_{j\neq i}\sum_{h=0}^{m}\beta_{jh}x_{vjh}\right)}}{\gamma_{r_v}(\beta)} \tag{2.89}$$

La expresión (2.88) se define sobre todos los patrones de respuesta $\{x_{vih}\}$ para la cual la puntuación es r_v y $x_{vih}=1$, obviamente, si $x_{vih}=1$, la puntuación es $r_v - w_h$ en los faltantes $k-1$ ítems. Así tenemos un resultado importante:

$$P(X_{vih}=1/r_v) = e^{(\beta_{ih})}\frac{\gamma_{r_v-w_h}^{(i)}(\beta)}{\gamma_{r_v}(\beta)} \tag{2.90}$$

donde $\gamma^{(i)}$ es la medida de la función γ omitiendo el ítem I_i. Es claro que:

$$\begin{aligned} E\left[X_{.ih}/r_1,...,r_n\right] &= \sum_{v=1}^{n} E\left[X_{vih}=1/r_v\right] \\ E\left[X_{.ih}/r_1,...,r_n\right] &= e^{(\beta_{ih})}\sum_{r} n_r \frac{\gamma_{r_v-w_h}^{(i)}(\beta)}{\gamma_{r_v}(\beta)} \end{aligned} \tag{2.91}$$

donde n_r es el número de individuos con $r_v = r$, y consecuentemente las ecuaciones CML son:

$$x_{.ih} = e^{(\beta_{ih})} \sum_r n_r \frac{\gamma^{(i)}_{r_v - w_h}(\beta)}{\gamma_{r_v}(\beta)} \text{, para } i = 1,...,k \; y \; h = 0,...,m \tag{2.92}$$

Es importante notar que (2.92) ha sido derivado sin asumir ninguna estructura en la matriz $\{\beta_{ih}\}$ de los parámetros del ítem y solo asumiendo que la puntuación categóricas son conocidas. Esto explica, que para muchos modelos politómicos la estimación de los parámetros de los ítems se haga a través del método de máxima verosimilitud condicional.

Ahora, aplicando logaritmo natural a la expresión (2.92), se tiene:

$$\begin{gathered} \ln x_{.ih} = \beta_{ih} + \ln \sum_r n_r \frac{\gamma^{(i)}_{r_v - w_h}(\beta)}{\gamma_{r_v}(\beta)} \\ \ln x_{.ih} - \beta_{ih} - \ln \sum_r n_r \frac{\gamma^{(i)}_{r_v - w_h}(\beta)}{\gamma_{r_v}(\beta)} = 0 \end{gathered} \tag{2.93}$$

Así

$$F_{ih}(\beta) = \ln(x_{.ih}) - \beta_{ih} - G_{ih}(\beta) = 0 \tag{2.94}$$

donde $G_{ih}(\beta) = \ln \sum_r n_r \frac{\gamma^{(i)}_{r - w_h}(\beta)}{\gamma_r(\beta)}$, por otra parte, observemos que:

$$\begin{gathered} F_{ih}(\beta^n) \cong F_{ih}(\beta^0) + \left[\beta^n - \beta^0\right] \left\{ \frac{\partial F_{ik}(\beta^0)}{\partial \beta_{jk}} \right\} \\ \left[F_{ih}(\beta^n) - F_{ih}(\beta^0)\right] \left\{ \frac{\partial F_{ik}(\beta^0)}{\partial \beta_{jk}} \right\}^{-1} \cong \left[\beta^n - \beta^0\right] \end{gathered} \tag{2.95}$$

finalmente el procedimiento de Newton-Rapshon es basado en el algoritmo:

$$\beta^n = \beta^0 + \left[F_{ih}(\beta^n) - F_{ih}(\beta^0)\right] \left\{ \frac{\partial F_{ik}(\beta^0)}{\partial \beta_{jk}} \right\}^{-1} \tag{2.96}$$

donde β^0 y β^n son vectores de estimadores obtenidos en iteraciones 0 y n, respectivamente.

La estimación de los parámetros de los ítems para el modelo politómico de Rasch es diferente en comparación al modelo de de Rasch dicotómico, debido a que el parámetro de dificultad puede descomponerse en los parámetros de dificultad de las categorías. La mayoría de programas para el análisis de Rasch utilizan el método de máxima verosimilitud condicional para los parámetros de los ítems y el método de máxima verosimilitud para la estimación de los parámetros de los sujetos.

Ejemplo de Simulación:

Supongamos que se ha examinado a un grupo con un test de 4 ítems, las respuestas obtenidas del test son politómicas (0, 1, y 2), donde la categoría 0 nos indica una respuesta incorrecta, 1 nos indica parcialmente correcta, y 2 nos indica una respuesta correcta. Quedando una matriz de datos *10X4.*

Tabla 2.11: Resultados del Test

Ítems / Sujetos	Ítems			
	1	2	3	4
Sujeto 1	0	1	2	1
Sujeto 2	1	2	0	1
Sujeto 3	2	0	1	2
Sujeto 4	1	0	1	2
Sujeto 5	1	0	1	2
Sujeto 6	1	1	2	0
Sujeto 7	0	0	2	1
Sujeto 8	1	2	0	2
Sujeto 9	2	2	0	1
Sujeto 10	1	0	1	2

Para estimar los parámetros de los sujetos e ítems se utilizará el programa BIGSTEPS, el cual nos proporciona el escalograma de Guttman donde podemos apreciar el ordenamiento de los sujetos e ítems. Los resultados los podemos observar en la tabla 2.12.

Tabla 2.12: Escalograma de Guttman

Sujetos	ítems			
	4	1	2	3
Sujeto 3	2	2	0	1
Sujeto 8	2	1	2	0
Sujeto 9	1	2	2	0
Sujeto 1	1	0	1	2
Sujeto 2	1	1	2	0
Sujeto 4	2	1	0	1
Sujeto 5	1	1	2	0
Sujeto 6	0	1	1	2
Sujeto 10	2	1	0	1
Sujeto 7	1	0	0	2

El escalograma nos da una primera apreciación acerca de qué sujetos y qué ítems son los más difíciles y los más fáciles. También, como se mencionó para el modelo de Rasch para ítems dicotómicos, es el primer paso para obtener los valores de las estimaciones iniciales.

Las estimaciones que se obtienen para los ítems y sujetos mediante el programa BIGSTEPS se presentan en las siguientes tablas:

Tabla 2.13: Estimaciones de los Parámetros de las Categorías e Ítems

Categorías \ Ítems	Dificultad y error estándar del paso (b_{ir})			Dificultad y error estándar del ítem $(b_i)^{**}$
	0*	1	2	
3	X	0.40 (0.66)	0.19 (0.70)	0.29 (0.39)
1	X	-0.99 (0.80)	1.28 (0.80)	0.14 (0.51)
2	X	0.80 (0.66)	-0.52 (0.66)	0.14 (0.36)
4	X	-1.53 (1.06)	0.38 (0.65)	-0.57 (0.50)

*: en este paso no se encuentra ningún valor debido a que no existe dificultad en fallar el paso 0.

**: los estimadores de los parámetros de los ítems son obtenidos promediando los estimadores de la dificultad de los pasos.

Tabla 2.14: Medida de los Estudiantes

Sujetos	Medida y error estándar de la habilidad (θ_v)
Sujeto 3	0.50 (0.68)
Sujeto 8	0.50 (0.68)
Sujeto 9	0.50 (0.68)
Sujeto 1	0.05 (0.67)
Sujeto 2	0.05 (0.67)
Sujeto 4	0.05 (0.67)
Sujeto 5	0.05 (0.67)
Sujeto 6	0.05 (0.67)
Sujeto 10	0.05 (0.67)
Sujeto 7	-0.41 (0.71)

En la tabla 2.13 podemos observar la dificultad de cada paso en las categorías junto con su desviación estándar en el paréntesis. Cada ítem tiene 2 pasos ya que se cuenta con tres categorías, además en la última columna se presenta la dificultad del ítem en general el cual es el promedio de todas las dificultades de los pasos que comprende el ítem. En la tabla 2.14 se presenta la medida de cada sujeto. Estas estimaciones se han obtenido mediante el programa BIGSTEPS el cual ha aplicado el método de máxima verosimilitud condicional para los ítems y el método de máxima verosimilitud para los estudiantes.

Podemos observar que el ítem 3 es el de mayor dificultad entre los cuatro ítems con medida 0.29, mientras que el ítem más fácil es el ítem 4 con una medida de -0.57. Sin embargo, los sujetos que están más aptos para contestar todos los ítems son Sujeto 3, 8 y 9. Mientras que los sujetos 1, 2, 4, 5 y 6, solo están aptos a contestar el ítem 4 el cual tiene menor medida. Mientras que el sujeto 7 que tiene la menor medida, solo podría esperarse que llegue hasta el paso 1 de los ítems 1 y 4.

En cuanto a los pasos que presenta cada ítem podemos ver que en el ítem 3 el paso más difícil es el paso 1, con lo cual estamos diciendo que la mayoría de estudiantes puntuaría 0, aquellos estudiantes que logren pasar el paso 1, se enfrentarían al paso 2 que tiene menor dificultad. En el ítem 1 podemos observar que la mayoría de estudiantes podrían llegar hasta el paso 1, sin embargo es muy difícil que la mayor parte de los estudiantes logre una puntuación de 2 ya que la dificultad del paso 2 es bastante alta (1.28). En el ítem 2, observamos que el paso más difícil es el 1, una vez superado este paso los sujetos se enfrentarían al paso 2 que tiene un menor nivel de dificultad, por lo tanto, se espera que los sujetos que superen el paso 1, responderán correctamente el ítem. Por último, en el ítem 4, los estudiantes obtendrían una puntuación de 1, ya que el paso 1 no presenta mayor dificultad y donde tendrían problemas es en el paso 2.

2.4 IMPORTANCIA DE USAR EL MODELO DE RASCH

En primer lugar conviene aclarar que el modelo clásico mide a los ítems con un esquema frecuentista de los resultados observados y no a través de una curva característica, por lo que la discriminación no es función de la pendiente de la curva si no en el conteo de respuestas correctas e incorrectas. Por ello no se puede garantizar que los criterios empleados en la teoría clásica brindan resultados similares a los que emite el modelo de Rasch.

Cuando un ítem no se comporta de manera adecuada (monótonamente creciente) se sabe que no se cumplen las hipótesis del modelo de Rasch.

1. Unidimensionalidad (posiblemente esté midiendo más de un solo rasgo).
2. Orden e inclusividad (posiblemente el ítem no tiene validez de contenido).

A esto se agrega:

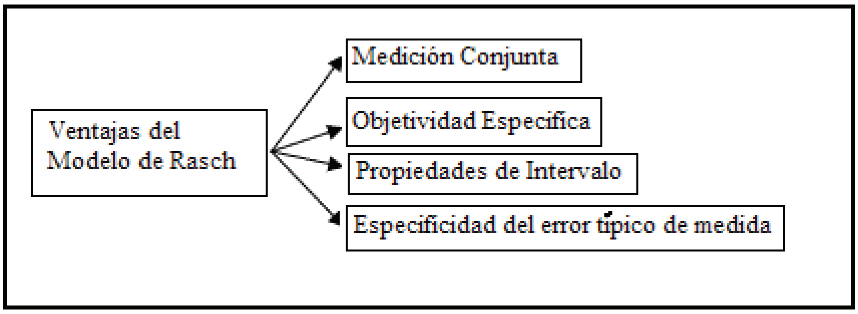
1. Posibilidad de respuestas al azar.
2. Influencia externas de respuestas.

Por otra parte, estas hipótesis no pueden ser verificadas por los métodos de la TCT.

2.4.1 VENTAJAS PRINCIPALES DEL MODELO DE RASCH RESPECTO A LA TCT

El modelo de Rasch, de entre los posibles modelos de la TRI, se destaca por tener las siguientes ventajas como se muestra en la siguiente figura:

Figura 2.22: Ventajas del Modelo de Rasch



Medición Conjunta: significa que los parámetros de las personas y los ítems se expresan en las mismas unidades y se localizan en el mismo continuo. En primer lugar esta propiedad confiere al modelo de Rasch un carácter más realista que el de la TCT, puesto que no es razonable mantener el supuesto de invarianza de los ítems: es obvio que no todos los ítems miden la misma cantidad del constructo. En segundo lugar, esta característica permite medir la interacción entre las personas y los ítems. En consecuencia, la interpretación de las puntuaciones no se fundamenta en normas de grupo, sino en la identificación de los ítems que las personas tienen una alta o baja probabilidad de resolver correctamente. Esta característica dota al modelo de Rasch con una gran riqueza diagnóstica.

Objetividad específica: una medida solo puede ser considerada válida y generalizable si no depende de las condiciones específicas con que ha sido obtenida. Es decir, la diferencia entre dos personas en un atributo no debe depender de los ítems específicos con los que sea estimada. Igualmente, la diferencia entre dos ítems no debe depender de las personas específicas que se utilicen para cuantificarla. Esta propiedad fue denominada objetividad específica por Rasch (1977).

Supóngase que dos personas de distinto nivel contestan al mismo ítem. De acuerdo con la ecuación (2.7):

$$\ln\left[P_i(\theta_1)/1-P_i(\theta_1)\right]=\theta_1-b_i,\ y\ \ln\left[P_i(\theta_2)/1-P_i(\theta_2)\right]=\theta_2-b_i$$

La diferencia entre ambas personas será igual a:

$$\ln(P_i(\theta_1)/1-P_i(\theta_1))-\ln(P_i(\theta_2)/1-P_i(\theta_2))=(\theta_1-b_i)-(\theta_2-b_i)=\theta_1-\theta_2$$

De forma similar, si la misma persona contesta a dos ítems de diferente dificultad:

$$\ln(P_1(\theta_v)/1-P_1(\theta_v))=\theta_v-b_1,\ y\ \ln(P_2(\theta_v)/1-P_2(\theta_n))=\theta_v-b_2$$

La diferencia en dificultad de ambos ítems será igual a:

$$\ln(P_1(\theta_v)/1-P_1(\theta_v))-\ln(P_2(\theta_v)/1-P_2(\theta_v))=(\theta_v-b_1)-(\theta_v-b_2)=b_1-b_2$$

En consecuencia si los datos se ajustan al modelo, las comparaciones entre personas son independientes de los ítems administrados y las estimaciones de los parámetros de los ítems no estarán influenciadas por la distribución de la muestra que se usa para la calibración. Hay que recordar que en la TCT las puntuaciones de las personas dependen de los ítems administrados y la dificultad de los ítems puede variar entre grupos de personas. En la propiedad de objetividad específica se fundamentan aplicaciones psicométricas muy importantes como la equiparación de puntuaciones obtenidas con distintos test, la construcción de bancos de ítems y los test adaptados al sujeto.

Propiedades de Intervalo: es importante notar que la interpretación de las diferencias en la escala es la misma a lo largo del atributo medido. Es decir, a diferencias iguales entre un sujeto y un ítem le corresponden probabilidades idénticas de una respuesta correcta. Por ello, la escala *logit* tiene propiedades de intervalo. Por el contrario, en la TCT las puntuaciones son casi siempre ordinales. La métrica de intervalar es de gran importancia, puesto que es una condición necesaria para usar con rigor los análisis paramétricos más frecuentemente empleados en las ciencias sociales (análisis de varianza, regresión, etc.) y, además, garantiza la invarianza de las puntuaciones diferenciales a lo largo del continuo.

Especificidad del error típico de medida: la objetividad específica no implica que la precisión de las estimaciones de los parámetros sea similar en distintos conjuntos de ítems y de personas. Si los ítems son fáciles, se estimarán con más precisión los parámetros de los sujetos de bajo nivel. De forma similar, si los sujetos son de alto nivel, se estimarán con mayor precisión los parámetros de los ítems difíciles. En la TCT, se supone que los test miden con la misma fiabilidad en todas las regiones de la variable. El modelo de Rasch no asume este supuesto tan poco verosímil. Permite por el contrario: i) Cuantificar la cantidad de información con la que se mide en cada punto de la dimensión y ii) Seleccionar los ítems que permiten incrementar la información en regiones del atributo previamente especificadas.

CAPÍTULO III: APLICACIÓN DEL MODELO DE RASCH

3.1 INTRODUCCIÓN

Los modelos de Rasch son modelos que nos sirven para obtener mediciones que no varíen en función del instrumento utilizado, disponer de instrumentos de medida que no dependan de los objetos medidos, es decir, que sean invariantes respecto a los sujetos evaluados. En la actualidad los modelos de Rasch están siendo ampliamente utilizados en países desarrollados como: EE.UU., Chile, Argentina, Dinamarca, etc. Con el fin de poder obtener estimaciones del nivel de habilidad de los sujetos en una determinada área y al mismo tiempo conocer las propiedades básicas de los ítems que son aplicados. Estos modelos generalmente son aplicados en el campo de la educación, sin embargo, en la actualidad países como: España y Brasil están aplicando el modelo de Rasch en el área de la economía y la salud respectivamente.

En este capítulo se hará una aplicación del modelo politómico de Rasch, más conocido como el modelo de Crédito Parcial, así como también, una aplicación del modelo dicotómico. La aplicación del modelo politómico se debe a que la base de datos con que disponemos ha sido categorizada ordenadamente, donde un estudiante con una puntuación de 2 en un ítem determinado tuvo que haber superado los pasos 0 y 1 del mismo ítem.

Para realizar dicha aplicación, se dispone de una base de datos con las notas por cada problema o ítem del primer parcial de matemática I, año 2006, que fue impartida por docentes de la Escuela de Matemática, de la Facultad de Ciencias Naturales y Matemática de la Universidad de El Salvador. Los ítems del primer parcial de matemática I tratan de conceptos básicos como: propiedades numéricas, propiedades algebraicas, factoreo, etc. Para realizar dicha aplicación se han categorizado las notas, de tal manera que la categoría más baja indique un nivel mínimo de conocimientos de los sujetos y la más alta indica un buen nivel de conocimientos de los sujetos, para resolver el problema. Asimismo se dispone de los resultados del Test de Nuevo Ingreso, UES, 2007, donde se hará la aplicación del modelo dicotómico de Rasch.

En los capítulos anteriores hemos estudiado algunas limitaciones que obtenemos con la aplicación de la TCT, entre ellas la dependencia que existe entre el test y los sujetos, y el estudio global que se hace con respecto al test. Con la aplicación del modelo de Rasch politómico se trata de resolver esta situación; obtener medidas que sean independientes del instrumento de medida y de los sujetos. Además ver si en realidad con el test aplicado se esta midiendo el objetivo que se persigue con esta evaluación (evaluar el nivel de habilidad de los estudiantes en conceptos básicos de matemática).

Para esta aplicación se hará uso el programa BIGSTEPS v. 2.82 el cual es una versión libre que puede ser descargado directamente de Internet en *http://www.winsteps.com* y que esta diseñado especialmente para el análisis de Rasch. Este programa trabaja en un entorno de MS-DOS y los resultados nos los brinda en un procesador Word.

De esta manera los objetivos que se pretenden lograr con dicha aplicación son:

Objetivo general.

- ☑ Realizar una aplicación de la “Teoría de respuesta al Ítem y Modelos de Rasch”, con las notas obtenidas en cada problema por cada estudiante que cursó la asignatura de Matemática I, en el ciclo I, 2006. Asimismo con los resultados del Test de Nuevo Ingreso, UES 2007.

Este objetivo general se desagrega en los siguientes objetivos específicos:

- ✓ Obtener una estimación de la medida (habilidad sobre los fundamentos básicos de matemática) de cada estudiante sometido a la prueba. De esta manera podemos observar qué estudiantes tienen o han adquirido buenas bases acerca de los fundamentos básicos de matemática.

- ✓ Obtener una estimación de la medida o calibración de los ítems. Se mostrará que ítems son los más fáciles o los más difíciles con el objeto de ser eliminados o modificado para futuras evaluaciones.

- ✓ Determinar probabilidades de acierto de acuerdo con el nivel de habilidad presentado por cada estudiante, así como también se valorará si el test esta acorde al nivel de conocimiento de dichos estudiantes.

- ✓ Obtener estadísticos de ajuste con el fin de conocer que tanto se ajustan los datos al modelo de Rasch.

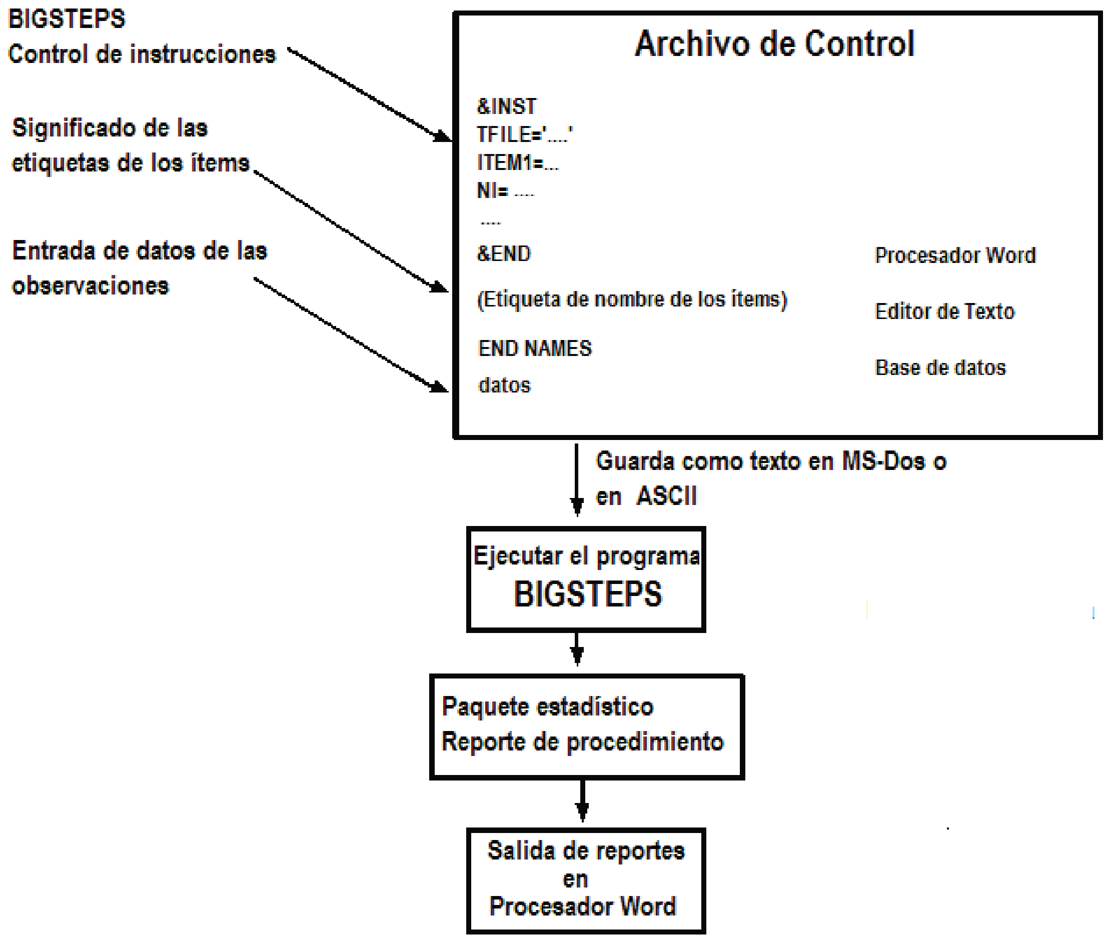
- ✓ Realizar un análisis acerca de las tablas y gráficas que el programa BIGSTEPS nos proporciona.

3.2 DESCRIPCIÓN DEL SOFTWARE A UTILIZAR (Programa BIGSTEPS versión 2.82)

El cálculo de la medida, de la calibración y de los valores de ajuste se realiza por medio de la computadora siguiendo los esquemas presentados hasta aquí. Se requiere de una computadora ya que la cantidad de operaciones se resuelven excesivamente rápidas.

BIGSTEPS es un programa creado por John M. Linacre y Benjamin D. Wright (1998) para el análisis de Rasch. La capacidad de trabajo del programa es de 32,000 personas y 3,000 ítems por lo que el tiempo de ejecución del procesamiento de los datos dependerá del espacio y la capacidad del microprocesador de la computadora. El esquema de trabajo de este programa se muestra en la siguiente figura:

Figura 3.1: Flujo de Trabajo Usado por BIGSTEPS



El programa BIGSTEPS esta basado en MS-DOS por lo que en principio puede ser un poco tedioso su uso, ya que se necesitan definir los comandos para procesar los datos. Este programa implementa la formulación de Rasch por medio de versiones modificadas de algoritmos, para la estimación de los parámetros, como *PROX* y *UCON*. Existen varios algoritmos que hacen el cálculo de los parámetros descriptivos de los ítems y las personas. Tratándose de diferentes métodos, ocurre que se puede llegar a valores diferentes para la calibración y para la medida. El programa BIGSTEPS comienza por una estimación central para cada medida de la persona, calibración del ítem, a menos que ésta sea predeterminada por el analista. Aunque no hay una normalización de estos métodos, el programa BIGSTEPS, por una parte, utiliza una versión interactiva del algoritmo PROX (aproximación normal) el cual permite obtener estimadores iniciales bajo la hipótesis de que los ítems y las personas se distribuyen normalmente a la vez es usado para alcanzar una convergencia del patrón de los datos.

Por otra parte, el programa también utiliza el algoritmo UCON el cual se utiliza, generalmente, para refinar o mejorar los resultados obtenidos con PROX, además para obtener estimaciones más exactas, de los errores estándar y de los estadísticos de ajuste (OUTFIT: ajuste externo e INFIT: ajuste interno). En este algoritmo no se hace hipótesis alguna sobre la distribución de los ítems o de las personas, por lo que pueden distribuirse en forma no normal. La implementación del UCON (máxima verosimilitud condicional, máxima verosimilitud conjunta) es un método usado correspondientemente con el método de Newton-Raphson para mejorar las estimaciones.

En definitiva el programa BIGSTEPS esta diseñado para construir medidas de Rasch para las respuestas de un conjunto de personas a un conjunto de ítems. Las respuestas pueden estar gravadas como letras o números y cada respuesta gravada puede ser de uno o dos caracteres. Los caracteres alfanuméricos, no diseñados como respuestas legítimas, son tratados como datos ausentes. Las respuestas a un ítem puede ser dicotómicas (“bueno”/”malo”, “si”/”no”), o pueden ser en escala de grado (“bueno”/”mejor”/”gustar”, “no estar de acuerdo”/”neutro”/”estar de acuerdo”) o tener créditos parciales u otra estructura jerárquica. Los ítems pueden ser agrupados en conjunto o en subconjuntos de uno o más ítems que comparten una estructura de respuesta.

Las salidas del programa consisten en una variedad de gráficos y de tablas adecuados para la importación a reportes escritos. Los estadísticos también pueden ser escritos en archivos de datos para la importación a otro software. Los estadísticos de ajuste (INFIT y OUTFIT) son reportados como la medias de cuadrados residuales, que tienen una aproximación a la distribución Chi-cuadrado, estos son también reportados en forma normal estándar (0,1).

3.2.1 ARCHIVO DE CONTROL: ÓRDENES SINTÁCTICAS BÁSICAS

Para el procesamiento de los datos hay que crear un *archivo de control*, éste es elaborado en un editor de texto como Worpad o Bloc de Notas, donde se especifica el número de ítems, número de sujetos, la posición de columna donde comienza la información de los ítems y etiquetas, el título del trabajo, estructura de los datos, etc. Así mismo, se especifica la ubicación de los datos (en caso que los datos no se encuentren en el mismo archivo de control), y el nombre de los archivos donde se van a enviar los resultados (datos de salida). La figura 3.2 nos muestra claramente las partes que consta un archivo de control.

Figura 3.2: Estructura del Archivo de Control para la Presente Aplicación

```
; Este archivo se llama TESIS.TXT
&INST                  ; inicio de órdenes
; Formato de entrada de datos
TITLE='TESIS'          ; Nombre que contendrá el encabezado de tablas
                       y gráficos
NI=7                   ; Número de ítems a utilizar
ÍTEM1=1                ; Posición de la columna donde comienza la
                       introducción de las respuestas a los ítems de
                       cada sujeto
NAME1=9                ; Posición de la columna donde comienza el
                       nombre del sujeto
XWIDE=1                ; Ancho de respuesta
NAMLEM=30              ; Máxima longitud que utiliza el nombre de los
                       sujetos
NAMLMP=20              ; Máxima longitud que utiliza el nombre de los
                       ítems en el mapa de ítems-personas
MUCON=10               ; Máximo número de iteraciones en el algoritmo
                       UCON
CODES=012345           ; códigos de respuesta válidas
CFILE=*                ; Etiqueta de los códigos de respuesta válidas
0 Deficiente
1 Necesita Mejorar
2 Regular
3 Suficiente              ; Área de codificación de las respuestas
4 Muy Bueno                 a los ítems
5 Excelente
*
```

```
TFILE=*                 ; Tablas o gráficos solicitados
 1                      ; Mapas de medición conjunta
 3                      ; Resumen de estadísticas del análisis
 7                      ; Ajuste de los estudiantes
 11                     ; Ajuste de los ítems
 17                     ; Medida de los ítems
 22                     ; Escalograma de Guttman
 *
 PERSON=ESTUDIANTES     ; etiqueta de los sujetos
 ÍTEM=PARCIAL 1         ; etiqueta de los ítems
 PFILE=tesis1.PF        ; archivo de salida para los resultados de los
                        sujetos
 IFILE=tesis1.IF        ; archivo de salida para los resultados de los
                        ítems
 &END
 Prop. de los números racionales e irrac. (ítem_1)
 Prop. de los números reales (ítem_2)
 Transf. de un decimal a un racional  (ítem_3)
 Simp. de un num. racional  (ítem_4)                     ;etiqueta de
 Prop. de los exponentes en expreciones racio. (ítem_5)  los ítems (7
 Simp. de expresiones con núm. Irrac. 1   (ítem_6)       en total)
 Simp. y prod. de expresiones con núm. Irrac. (ítem_7)
 END NAMES
 2200050 E1
 3253300 E2
 2234200 E3
 1200000 E4
 ………………                      ;Datos
 ………………
 4255230 E174
 2355100 E175
 2205130 E176
```

A partir del archivo de control para la aplicación que se realizará, pueden observarse varios comandos que son utilizados para el análisis de Rasch. A continuación se presentan algunos comandos e instrucciones generales:

Figura 3.3: Comando e Instrucciones Generales

Comando de inicio:	*&INST*
Explicación de ordenes:	*;*
Comando de finalización de ordenes:	*&END*
Comando de finalización de etiquetas de los ítems:	*END NAMES*
Inicio y finalización de información	*

Figura 3.4: Entrada de Datos

DATA=*."Extensión del archivo"	Nombre del fichero de datos (en caso que los datos se encuentren en otro archivo)
NAME1=	Primera columna donde se inicia la información identificativa de las personas
NAMLEN=	Número total de columnas que incluye información identificativa de las personas, por ejemplo 30.
ÍTEM1=	Primera columna donde inicia la información de los ítems
NI=	Número de ítems
XWIDE=	Columnas por ítem

Figura 3.5: Selección de Datos

CODES=	Códigos de datos válidos, 01 (dicotómico) ó 012345 (politómico escala 0 a 5)

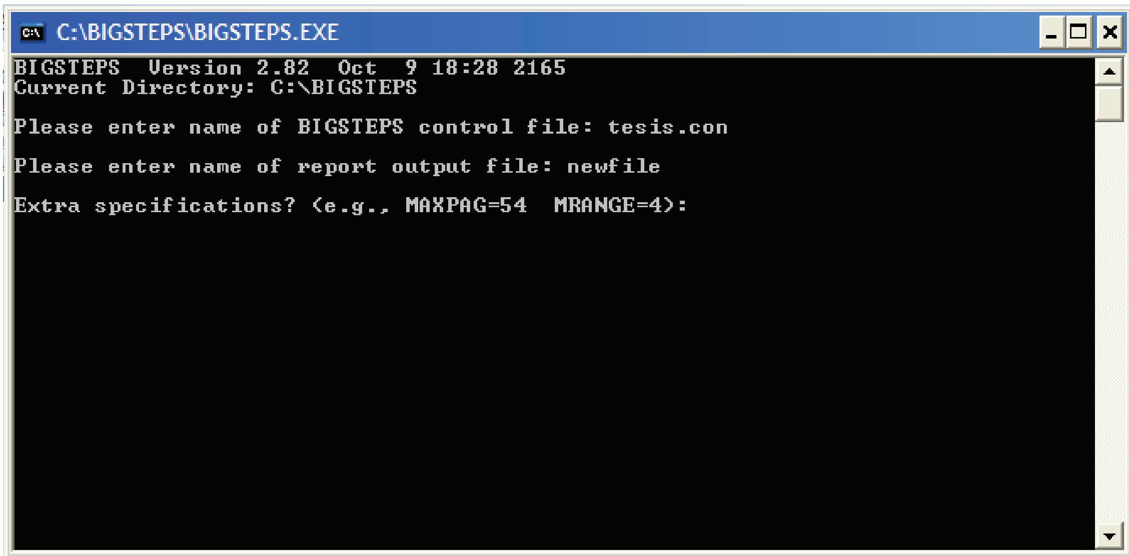
Figura 3.6: Tablas de Resultados

ÍTEM=	Título o etiqueta de los ítems
CFILE=	Titulo o etiquetas de las categorías
PERSON=	Titulo o etiqueta de los sujetos
TITLE=	Titulo o etiqueta de trabajo
TFILE=	Tablas o gráficos solicitadas

Los comandos anteriores son los más básicos para usar BIGSTEPS, pero si el lector está interesado por conocer y profundizar en el manejo de dicho programa puede utilizar la guía de usuario que los creadores del programa brindan y que se encuentra disponible como archivo PDF en www.winsteps.com/bigsteps.htm.

Por último falta conocer como ejecutar el programa, lo cual daremos una breve explicación: Una vez instalado BIGSTEPS y creado el *archivo de control*, es necesario buscar el directorio que fue creado por dicho programa, generalmente se encuentra en C:\BIGSTEPS; dentro de este directorio, se busca el archivo llamado del mismo nombre (BIGSTEPS.EXE) y damos doble clic, aparecerá una pantalla en MS- DOS (Figura 3.7) que nos pide el nombre del *archivo de control* que se debe de encontrar en el mismo directorio donde esta instalado el programa, junto con la extensión con la que fue hecho (por ejemplo: *.txt, *.con, etc.). Posteriormente pide el nombre del archivo donde se guardarán nuestros resultados (debe completarse aunque no es necesario haber creado un archivo, podemos darle un nombre como NEWFILE y este archivo se creará automáticamente dentro de la misma carpeta). La siguiente figura nos muestra la pantalla para el programa BIGSTEPS donde se ha introducido el archivo de control para nuestra aplicación:

Figura 3.7: Ubicación del Archivo de Control Dentro del Programa BIGSTEPS



Si no hay especificaciones extras, se le da *enter* en la última línea de pantalla anterior y se creará automáticamente el archivo NEWFILE conteniendo los resultados que le hemos especificado al programa. Finalmente, el archivo NEWFILE ha sido almacenado en el directorio C:\BIGSTEPS, es decir en la carpeta que el mismo programa crea dentro de su computadora. Si nuestro archivo de control esta elaborado adecuadamente, el programa no nos dará ningún problema y podremos obtener resultados confiables.

3.3 DESCRIPCIÓN Y CODIFICACIÓN DE LOS DATOS

PARTICIPANTES

Se han analizado las respuestas de 176 estudiantes de la materia de Matemática I, que fue impartida por docentes de la Escuela de Matemática, en el ciclo I, 2006.

INSTRUMENTO

Con el test se pretende medir el manejo a los sistemas numéricos entre ellos las operaciones algebraicas y las aritméticas. Este instrumento de prueba se constituye en 3 partes. La primera parte consta de propiedades de los números racionales e irracionales, la segunda parte consta de propiedades de los números reales y la tercera y última parte esta compuesta por cinco ítems donde el estudiante tiene que resolver los problemas que se le presentan. El test esta conformado por 7 ítems los que se identifican por medio de una descripción breve, que hace referencia al problema sobre el cual trata cada uno, esto se observa a continuación:

Propiedades de los números racionales e irracionales (ítem 1)

Propiedades de los números reales (ítem 2)

Transformación de un decimal a un racional (ítem 3)

Simplificación de un número racional (ítem 4)

Propiedades de los exponentes en expresiones racionales (ítem 5)

Simplificación de expresiones con números irracionales (ítem 6)

Simplificación y producto de expresiones con números irracionales (ítem 7)

PREPARACIÓN DE LOS DATOS.

El conjunto de datos está formado por una matriz de 176 estudiantes y 7 ítems o problemas (M_{176X7}), de cada estudiante se tienen la nota obtenida por ítem, los cuales son datos continuos como se muestra en la siguiente tabla:

Tabla 3.1: Representación de la Base de Datos

Codigo de Estudiante	NOTAS POR PROBLEMA						
	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6	Ítem 7
E1	0.60	0.50	0.00	0.00	0.00	1.50	0.00
E2	0.80	0.50	1.50	1.00	0.80	0.00	0.00
E3	0.40	0.60	0.80	1.20	0.50	0.00	0.00
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
E174	1.00	0.40	1.50	1.50	0.50	1.00	0.00
E175	0.60	0.80	1.50	1.50	0.10	0.00	0.00
E176	0.60	0.50	0.00	1.50	0.30	0.80	0.00

Para poder realizar la aplicación del modelo de Rasch es necesario codificar la base de datos de tal manera que tengamos datos enteros, con categorías ordenadas y con las propiedades de los datos originales.

Los datos se clasificarán en escala likert, en los cuales los valores enteros crecientes nos indicarán un nivel de aumento en la puntuación obtenida, por lo que las categorías son las siguientes:

0. Deficiente $(p = 0)$
1. Necesita Mejorar $\left(0 < p \leq 2.5\right)$
2. Regular $\left(2.5 < p \leq 5\right)$
3. Bueno $\left(5 < p \leq 7\right)$
4. Muy Bueno $\left(7 < p \leq 9\right)$
5. Excelente $\left(9 < p \leq 10\right)$

De esta manera la nueva base nos queda de la siguiente forma:

Tabla 3.2: Codificación de la Base de Datos

Codigo de Estudiante	CATEGORIAS POR PROBLEMA						
	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6	Ítem 7
E1	2	2	0	0	0	5	0
E2	3	2	5	3	3	0	0
E3	2	2	3	4	2	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
E174	4	2	5	5	2	3	0
E175	2	3	5	5	1	0	0
E176	2	2	0	5	1	3	0

Como se observa, los valores crecientes de categorías indican un nivel de aumento en la habilidad de los estudiantes. De esta manera los datos pueden ser procesados para su posterior análisis. Para poder analizar los datos con modelos politómicos se le debe proporcionar al programa BIGSTEPS las opciones concretas escogidas por cada sujeto en cada Ítem (figura 3.8):

Figura 3.8: Fichero de Datos con Respuestas Directas de Individuos

```
2200050 E1
3253300 E2
2234200 E3
1200000 E4
..................
4255230 E174
2355100 E175
2205130 E176
```

Cada línea representa los datos de un individuo. En las primeras columnas se colocan las respuestas a cada uno de los ítems (en este caso, 7 ítems). Cada una de las siguientes columnas se utilizan para identificar a los individuos (en este caso, el identificador ocupa las últimas 4 columnas). En nuestro caso, los ítems tienen 6 opciones. Se puede observar que el sujeto 1 ha obtenido una puntuación de 2 en el ítem 1 e ítem 2 (2: regular), y no ha resuelto el ítem 3, 4, 5 y 7 (debido a que tiene una puntuación de 0), y ha resuelto correctamente el problema 6 (5: Excelente). Además, podemos observar que el último estudiante ha obtenido una puntuación de 2 en el ítem 1 y 2, y no ha podido resolver el ítem 3 y el ítem 7; sin embargo, el ítem 4 lo ha resuelto correctamente, y en el ítem 5 ha obtenido una puntuación de 1 (1: necesita mejorar), y seguidamente en el ítem 6 ha obtenido una puntuación de 3 (3: bueno).

3.4 APLICACIÓN DEL MODELO DE RASCH

En el capítulo I, se estudiaron los supuestos que se deben cumplir en la TRI para la aplicación de cualquier modelo que pertenezca a ella. En el modelo de Rasch *la unidimensionalidad* de los datos *y la independencia local* son supuestos que deben de cumplirse para poder obtener conclusiones válidas.

3.4.1 CUMPLIMIENTO DE LOS SUPUESTOS

Una de las formas de verificar la unidimensionalidad es el hecho de aplicar el análisis factorial a la base de datos con que se cuenta, dado que a través de la varianza explicada podemos verificar si el primer factor explica más de un 25% de la varianza total, si es así, se considera que se cumple con el supuesto de unidimensionalidad.

Tabla 3.3: Varianza Explicada

Factores	% de Varianza	% de Varianza Acumulada
1	51.43	51.43
2	11.47	62.9
3	9.51	72.41
4	9.3	81.71
5	6.92	88.63
6	6.79	95.43
7	4.57	100

Como se observa de la tabla anterior, el factor uno explica el 51.43% del total de variabilidad de los resultados de los ítems, por lo tanto podemos decir que los datos están midiendo un solo factor (datos unidimensionales) y este factor[1] es: *operaciones elementales con los números.* El supuesto dos trata de la *independencia local*, por las características del parcial I, se observa que los ítems no dependen entre sí, además se asumirá que no hubo copia entre los estudiantes.

Bajo el cumplimiento de los supuestos de:

a) Unidimensionalidad, es decir, que el test mide exactamente un constructo
b) Independencia local: esto es, las respuestas de los ítems son independientes entre si.

Podemos aplicar el modelo de Rasch politómico, para el análisis de los resultados del primer parcial de Matemática I.

3.4.2 RESULTADOS OBTENIDOS CON EL PROGRAMA BIGSTEPS (PRIMER PARCIAL DE MATEMÁTICA I)

A continuación se presenta una visión parcial de los resultados, ya que el programa BIGSTEPS proporciona un gran número de reportes. Aquí solo se van a presentar los resultados más importantes de los cuales se dará una interpretación general. Ahora bien, los resultados obtenidos al hacer la aplicación del modelo de Rasch politómico a los datos del primer parcial de Matemática I son los siguientes:

3.4.2.1 Tabla de Convergencia

La tabla 3.4 nos da un reporte ilustrativo sobre la convergencia de los resultados de los estudiantes de Matemática I. La tabla 3.4 se divide en dos partes, la parte superior nos indica los resultados del algoritmo PROX (algoritmo de aproximación normal) donde la primera columna

[1] En la práctica, cuando el primer factor explica más de un 25% de la varianza total se considera que se cumple el supuesto de unidimensionalidad.

muestra el número de iteraciones que realizó para calcular las estimaciones de los parámetros. La segunda columna indica el conteo activo, que presenta el número de elementos participantes en el proceso de estimación luego de eliminar las puntuaciones perfectas (por ejemplo un estudiante que tiene la puntuación de 5 en todos los ítems o al contrario un estudiante que tenga la puntuación de 0 en todos los ítems). Esto lo hace para los estudiantes, los ítems y para las categorías. En la tercera y última columna se encuentra el máximo cambio en lógitos para las medidas de los estudiantes y representa como su nombre lo dice el máximo cambio en lógitos en la estimación para los estudiantes y para los pasos significa el máximo cambio en logit en cualquier estimación de la dificultad de los pasos de cada ítem.

Tabla 3.4 : Tabla de Convergencia

Íteracción PROX	Conteo Activo			Máximo cambio en lógitos	
	Estudiante	Ítems	Categorías	Medidas	Pasos
1	176	7	6	-3.5264	2.6114
2	172	7	6	0.2914	1.7582
3	172	7	6	-0.0639	-0.0624

Iteracción UCON	Máxima puntación residual	Máximo cambio en lógitos	Categoria residual
1	62.01	-0.6394	-17.04
2	-47.34	-0.2307	12.44
3	10.62	-0.1399	-5.80
4	8.88	-0.1013	-5.00
5	7.52	-0.0931	3.82
6	6.73	-0.0922	3.53
7	8.15	-0.0698	2.90
8	10.47	-0.0642	-1.80
9	5.62	-0.0383	0.95
10	3.16	-0.0253	-0.75

Residuales Estandarizados N(0,1) Media: -0.01 Desviación Estándart: 0.98

Obsérvese que como resultado del conteo activo de los 176 estudiantes que en un principio iniciaron, al final de las 3 iteraciones el programa BIGSTEPS conservó solamente 172 estudiantes los cuales podrán ser analizados. Por otra parte, todos los ítems han podido calibrarse correctamente ya que al final de las iteraciones siguen conservándose los mismos 7 ítems, al igual que las categorías. El máximo cambio en lógitos que se obtuvo en la última iteración es del orden de tres centésimas, lo cual indica que los valores tienen una aceptable precisión.

Para afinar los resultados obtenidos con la implementación del algoritmo PROX, el programa BIGSTEPS aplica el método de Máxima Verosimilitud o el método de Máxima Verosimilitud Condicional junto con el algoritmo de Newton-Raphson (que se le denomina algoritmo UCON) como se muestra en la segunda parte de la tabla 3.4. En la primera columna observamos el número de iteraciones que realizó este algoritmo para calcular las estimaciones de los parámetros. La máxima puntuación residual se encuentra en la segunda columna, esta medida surge de la diferencia entre la puntuación observada y la puntuación esperada del modelo. De la misma forma en la tercera columna encontramos el máximo cambio en lógitos, que como su nombre lo dice representa el máximo cambio en lógitos en cualquier estimación de los parámetros de estudiantes o ítems. En la cuarta y última columna encontramos la categoría residual que representa el máximo residuo el cual surge de la diferencia entre el valor observado y el valor esperado; para cualquier escala de las categorías que comprende el ítem. Es bueno mencionar que en esta columna los valores menores a 0.5 no tienen un valor significativo.

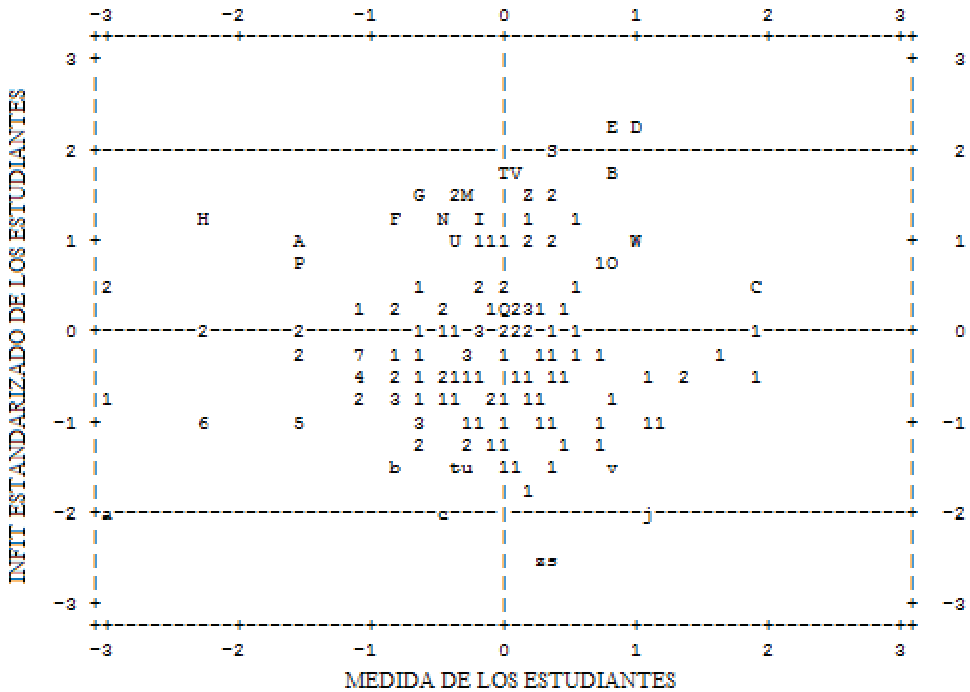
El reporte indica el residuo máximo que se obtiene en los puntajes; nótese el decrecimiento gradual de este residuo, lo cual hace que el proceso interactivo sea convergente. El máximo cambio en lógitos que se obtuvo crece gradualmente, por lo que la convergencia es aceptable. Podemos observar que los residuos estandarizados con una media de -0.01 y desviación estándar de 0.98 por lo cual se puede decir que los datos presentan un buen ajuste porque se aproximan a la normal con media 0 y desviación estándar 1. En la categoría residual podemos observar que existen valores menores a 0.5 lógitos, por lo que podemos decir existen defectos o irregularidades en los datos, estas situaciones se particularizaran en los resultados que se presentan posteriormente.

En general, la tabla 3.4 nos indica con que precisión se han obtenido las estimaciones de los parámetros de los sujetos, ítems y los pasos que comprende cada ítem, de tal manera que cada vez que se modifique el test, esta tabla indicará la precisión de los resultados.

3.4.2.2 Gráfico de Ajuste de los Estudiantes

En la siguiente figura podemos observar como se comporta el estadístico de ajuste INFIT (ajuste interno), con respecto a la medida de los estudiantes. En el eje de las abscisas podemos ver la medida de los estudiantes y en el eje de las ordenadas tenemos el ajuste interno estandarizado de los estudiantes. Cada letra o número indica el estadístico de ajuste correspondiente a un estudiante con cierto nivel de habilidad. En la figura 3.9 se observa tres bandas, se espera que la mayor parte de los datos se concentren en la banda central (entre -2 y 2 lógitos) para decir que presentan un buen ajuste del modelo. Sin embargo, si se observan datos en la banda superior (INFIT>2 lógitos), podríamos decir que existen estudiantes que presentan demasiada estocasticidad en las respuestas; caso contrario, si observamos datos en la banda inferior (INFIT<-2) estaríamos en presencia de estudiantes que presentan demasiado determinismo en sus respuestas. Esta figura nos ayuda a conocer de una forma general el comportamiento de los datos que se analizan.

Figura 3.9: Ajuste Interno de los Estudiantes



En efecto, de la figura anterior podemos observar que la mayoría de elementos o datos correspondientes a los estudiantes se encuentran en la zona media que corresponde a -2 y 2 lógitos. Sin embargo, podemos observar que existen estudiantes que se encuentran en la banda superior e inferior respectivamente, por lo cual podemos decir que existe desajuste interno en las medidas de algunos estudiantes.

Asimismo el programa BIGSTEPS reporta otros gráficos que corresponde al ajuste de los datos (FIT), donde al igual que la figura anterior, se presentan tres bandas, una superior, una zona media y una zona inferior tanto para estudiantes como para los ítems. Estos gráficos se muestran en el ANEXO I y también se observa que en el ajuste externo de los estudiantes tenemos datos en la zona superior e inferior, pero la mayoría de datos se encuentran en la zona media y por lo tanto podemos decir que la mayoría de estudiantes presentan un buen ajuste externo al modelo de Rasch. De igual manera podemos observar los gráficos de ajuste de los ítems en el ANEXO I, donde podemos observar que para el ajuste interno (INFIT) tenemos un ítem (simbolizado por “a”) que se encuentra entre la zona media e inferior (INFIT=2) y en el ajuste externo de los ítems (OUTFIT) observamos que el mismo ítem se encuentra en la zona inferior, lo cual indica que existe un ítem que presenta desajuste al modelo. Estos gráficos nos dan un primer acercamiento del ajuste de los datos, pero para determinar que estudiantes o ítems son los que presentan estos desajustes al modelo se utilizarán otras técnicas que se verán posteriormente.

3.4.2.3 Análisis de los Pasos por Categoría

En la tabla 3.5 la primera columna muestra el número de ítem del análisis, de la segunda a la séptima columna tenemos el número de estudiantes que obtuvieron las puntuaciones de *0: Deficiente, 1: Necesita Mejorar, 2: Regular, 3: Suficiente, 4: Muy Bueno, 5: Excelente*, en cada ítem (tabla de frecuencias). La suma, por filas, de estas frecuencias habrá de ser, en cada ítem, 172, el número de los estudiantes que han podido ser analizados por el modelo. De la columna octava a la décima segunda tenemos los estimadores del nivel de dificultad de cada paso, por ítem, y entre paréntesis, los errores de estimación obtenidos mediante el programa BIGSTEPS.

Tabla 3.5: Dificultad de los Pasos de Cada Categoría

ítem	Deficiente (0)	Necesita Mejorar (1)	Regular (2)	Suficiente (3)	Muy Bueno (4)	Excelente (5)	Paso 1	Paso 2	Paso 3	Paso 4	Paso 5
1	2	14	101	33	12	10	-4.20(0.83)	-2.88(0.30)	0.92(0.19)	1.27(0.27)	0.89(0.37)
2	4	38	75	31	14	10	-3.99(0.58)	-1.35(0.21)	0.76(0.19)	1.11(0.26)	1.05(0.37)
3	57	6	41	18	11	39	1.39(0.22)	-2.33(0.21)	0.70(0.20)	0.62(0.22)	-0.85(0.23)
4	75	15	13	13	10	46	0.91(0.20)	-0.17(0.21)	-0.06(0.22)	0.41(0.22)	-1.13(0.22)
5	96	20	19	10	13	14	1.08(019)	-0.03(0.21)	0.83(0.24)	0.23(0.27)	0.70(0.34)
6	101	6	13	17	4	31	2.31(0.21)	-0.92(0.21)	-0.18(0.23)	1.76(0.25)	-1.47(0.26)
7	114	8	11	17	0	22	2.25(0.20)	-0.33(0.22)	-0.14(0.24)	X	0.34(0.28)

Como se mencionó en el capítulo anterior, los parámetros que se estiman son *m-1* categorías, en este caso tenemos 6 categorías, por lo tanto se estiman los primeros 5 parámetros como se observa en la tabla anterior.

En el ítem 1 (*Propiedades de los números racionales e irracionales)*, se observa que el primer paso es muy fácil, ya que tiene un valor de -4.20 lógitos, por tanto se espera que la mayoría de estudiantes superen o hagan bien el paso 1 y se enfrenten al paso 2. La dificultad no aumenta significativamente al enfrentarse al paso 2, y se espera que superen este paso aquellos estudiantes con habilidad o nivel de conocimiento sobre la competencia que evalúa el ítem 1, mayor o igual a -2.88 lógitos. Luego, los estudiantes que superen el paso 2, se enfrentaran a un paso más difícil (paso 3) y se espera que únicamente superen el paso 3 los estudiantes con nivel de conocimiento mayor o igual 0.92, sobre la competencia que evalúa el ítem 1. Los estudiantes que superen el paso 3, se enfrentaran a un paso más difícil (paso 4), y se espera que este paso lo superen aquellos estudiantes con nivel de habilidad mayor o igual a 1.27, sobre la competencia que evalúa el ítem 1. Finalmente, los estudiantes que hayan superado el paso 4, se enfrentaran a un paso menos difícil, que es el paso 5, y se espera que esté sea superado por los estudiantes que superaron el paso 4.

En el ítem 2 *(Propiedades de los números reales)*, se observa que el primer paso es muy fácil, ya que tiene un valor de -3.99 lógitos, por lo tanto se espera que la mayoría de estudiantes superen este paso y se enfrenten al paso 2. La dificultad no aumenta significativamente al enfrentarse al paso 2, y se espera que superen este paso aquellos estudiantes con habilidad sobre la competencia que evalúa el ítem, mayor o igual a -1.35. Luego, los estudiantes que superen el paso 2, se enfrentaran a un paso más difícil (paso 3) y se espera que únicamente superen el paso 3, los estudiantes con nivel de conocimiento de mayor o igual a 0.76 lógitos, sobre la competencia que evalúa el ítem. Los estudiantes que superen el paso 3, se enfrentan a un paso más difícil (paso 4), y se espera que este paso lo superen aquellos estudiantes con niveles de habilidad mayores o iguales 1.11 lógitos. Finalmente, los estudiantes que hayan superado el paso 4, se enfrentaran a un paso menos difícil, que es el paso 5, y se espera que este paso sea superado por todos los estudiantes que superaron el paso 4, ya que se requiere un nivel de conocimiento de 1.05 lógitos, sobre la competencia que evalúa el ítem 2.

En el ítem 3 *(Transformación de un decimal a un racional)*, se observa que el primer paso es muy difícil, ya que tiene valor de 1.39 lógitos, por lo tanto se espera que este paso sea superado por estudiantes con un nivel de habilidad de 1.39 lógitos para superar el paso 1 y enfrentarse al paso 2. La dificultad disminuye considerablemente al enfrentarse al paso 2, y se espera que este paso lo superen todos los estudiantes que superaron el paso 1, ya que se requiere un nivel de habilidad mayor o igual a -2.33 lógitos. Luego, los estudiantes que hayan superado el paso 1 y paso 2, se enfrentaran al paso 3, y se espera que superen este paso todos los estudiantes que tengan un nivel de habilidad mayor o igual a 0.70 lógitos, por tanto se espera que este paso sea superado por todos los estudiantes. Los estudiantes que superen el paso 3, se enfrentan a un paso menos difícil (paso 4), y se espera que este paso sea superado por aquellos estudiantes con habilidad de 0.62 lógitos, sobre la competencia que evalúa el ítem. Finalmente, los estudiantes se enfrentan a un paso fácil (paso 5), ya que se requiere un nivel de conocimiento o habilidad de -0.85 lógitos. Por último podemos decir que la mayor dificultad se encuentra en el paso 1, una vez superado esté paso, se espera que los estudiantes puedan resolver cada paso que comprende el ítem.

En el ítem 4 *(Simplificación de un número racional)*, se observa que los estudiantes se enfrentan a un paso difícil, con dificultad de 0.91 lógitos, por tanto se espera que este paso sea superado por aquellos estudiantes con habilidad o un nivel de conocimiento mayor o igual a 0.91 lógitos y se enfrenten al paso 2. La dificultad disminuye en el paso 2, y se espera que éste sea superado por todos los estudiantes que superaron el paso 1, debido que solo se requiere un nivel de conocimiento de -0.17 lógitos sobre la competencia que evalúa el ítem. Luego, los estudiantes que superen el paso 1 y paso 2 respectivamente, se enfrentan a un paso fácil (paso 3), y se espera que este paso sea superado por todos los estudiantes que superaron los pasos anteriores, ya que se requiere una habilidad o nivel de conocimiento de -0.06 lógitos sobre la competencia que evalúa el ítem. Seguidamente, los estudiantes se enfrentan a un paso difícil (paso 4), y se espera que este paso sea superado por aquellos estudiantes con un nivel de habilidad de 0.41, por lo tanto, se espera que sea superado por todos los estudiantes que superaron el paso 1. Finalmente el paso 5 es el más fácil, ya que se requiere un nivel de conocimiento de -1.13 lógitos; por lo tanto se espera que no presente ningún problema para todos los estudiantes que superaron el paso 1 (ya que el paso 1 es el más difícil).

En el ítem 5 *(Propiedades de los exponentes en expresiones racionales)*, se observa que el primer paso es difícil, ya que tiene un valor de 1.08 lógitos, por tanto se espera que solo los estudiantes con un nivel de habilidad de 1.08 lógitos superen este paso y se enfrenten al paso 2. La dificultad disminuye al enfrentarse al paso 2, y se espera que todos los estudiantes que superaron el paso 1 no tengan ningún problema en superar el paso 2, el cual tiene un nivel de dificultad de -0.03 lógitos. Luego, los estudiantes que superaron el paso 1 y paso 2 respectivamente, se enfrentan a un paso más difícil que el paso 2 (paso 3), se espera que este paso sea superado por aquellos estudiantes con un nivel de conocimiento mayor o igual a 0.83 lógitos, por tanto no presenta ningún problema para los estudiantes que superaron el paso 1, ya que el paso 1 tiene mayor dificultad. Al enfrentarse al paso 4, la dificultad disminuye con respecto al paso 3, ya que se espera que sea superado por los estudiantes con un nivel de habilidad mayor o igual a 0.23 lógitos, por tanto, al igual que el paso 3, no presenta ningún problema para los estudiantes que superaron el paso 1. Finalmente, en el paso 5 se espera que sea superado por los estudiantes con un nivel de habilidad mayor o igual a 0.70 lógitos sobre la competencia que evalúa el ítem; por tanto, al igual que los pasos anteriores, no presenta ningún problema para los estudiantes que superaron el paso 1.

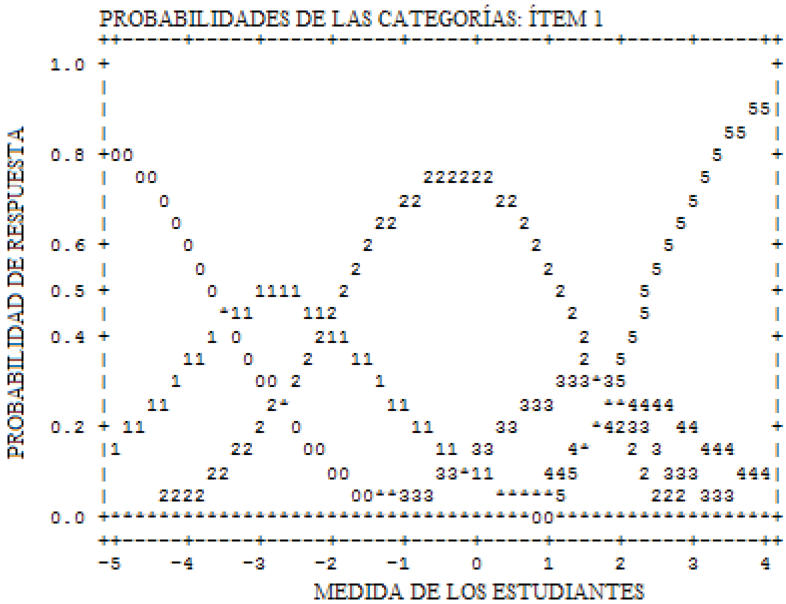
En el ítem 6 *(Simplificación de expresiones con números irracionales)*, se observa que el primer paso es muy difícil, ya que tiene un valor de 2.31 lógitos, por tanto, se espera que muy pocos estudiantes logren superar o hagan bien el paso 1 y se enfrenten al paso 2. La dificultad disminuye al enfrentarse al paso 2, y se espera que este paso sea superado por los estudiantes que superaron el paso 1, ya que se requiere un nivel de conocimiento de -0.92 sobre la competencia que evalúa el ítem. Luego, los estudiantes que superaron el paso 1 y paso 2 respectivamente, se enfrentan a un paso fácil (paso 3), y se espera que este paso sea superado por los estudiantes, ya que se requiere de un nivel de conocimiento menor que el paso 1 (-0.18 lógitos<2.31 lógitos) sobre la competencia que evalúa el ítem. Al enfrentarse al paso 4, la dificultad aumenta y se espera que sea superado por estudiantes con un nivel de conocimiento mayor o igual 1.76 lógitos; por tanto, al igual que el paso 3, no presenta problema para los estudiantes que superaron el paso 1. Finalmente, el paso 5 se espera que sea superado por estudiantes con un nivel de conocimiento mayor o igual a -1.47 lógitos; por tanto, al igual que los pasos anteriores, no presenta ningún problema para los estudiantes que superaron el paso 1.

Finalmente, el ítem 7 *(Simplificación y producto de expresiones con números irracionales)*, se observa que el primer paso es muy difícil, ya que tiene un valor de 2.25 lógitos, por tanto, se espera que pocos estudiantes superen este paso. Los estudiantes que superaron el paso 1, se enfrentan al paso 2, en el cual se espera que éste sea superado por estudiantes con un nivel de habilidad mayor o igual a -0.33 lógitos sobre la competencia que evalúa el ítem; por tanto no presenta ningún problema para los estudiantes. Al igual que el paso 2, la dificultad del paso 3 no presenta ningún problema para los estudiantes que superaron el paso 1, ya que se requiere un nivel de conocimiento mayor o igual a -0.14 lógitos sobre la competencia que evalúa el ítem. En el paso 4 podemos observar en la tabla que no se presenta ningún valor, esto debido a que no hubieron estudiantes que obtuvieran una puntuación de 4 (Muy Bueno) en este ítem; en otras palabras, no existe dificultad al pasar de la puntuación de 3 a la puntuación 4. Finalmente, al pasar de la puntuación 4 a la puntuación 5 (paso 5), se espera que este paso sea superado por estudiantes con un nivel de habilidad mayor o igual a 0.34 lógitos sobre la competencia que evalúa el ítem; por tanto, al igual que los pasos anteriores, no presenta ningún problema para los estudiantes que superaron el paso 1.

El valor que está entre paréntesis a la par de cada estimador representa el error estándar de los estimadores y puede ser utilizado como medida de precisión de la estimación de los parámetros. A menor error de estimación, el estimador se encuentra más próximo del verdadero valor del parámetro. Es decir, más seguros podremos estar de que obtendríamos valores parecidos al parámetro, de repetir la evaluación del estudiante en las mismas condiciones. Si quisiéramos saber cual es la dificultad del ítem, éste se calcula como el promedio de los estimadores de los pasos, por ejemplo, para el ítem 5 su nivel de dificultad será 0.56 ($(1.08-0.03+0.83+0.23+0.70)/(5)=0.562$), la dificultad de cada ítem se analizará en las siguientes secciones.

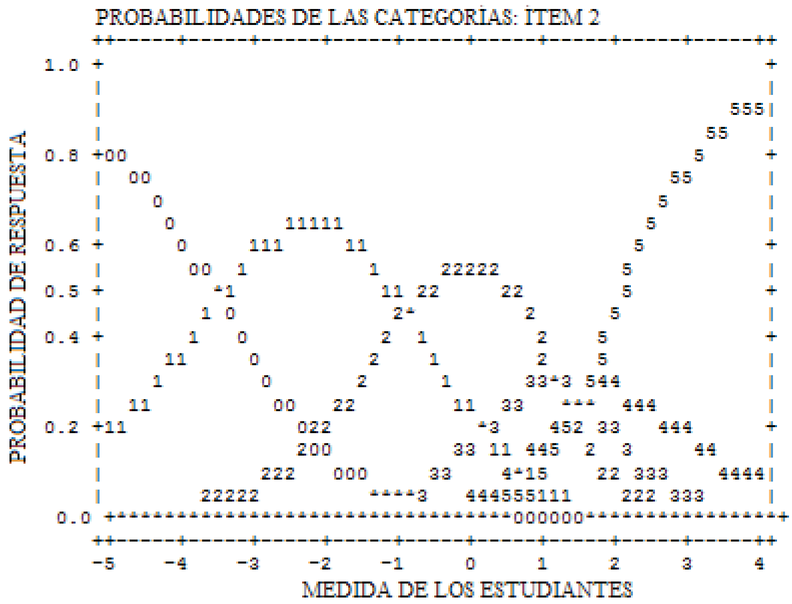
Para tener una mejor compresión de los resultados obtenidos por las categorías de cada ítem se presentan las funciones de respuestas categóricas. En los siguientes gráficos, el eje de las abscisas indica la medida o nivel de habilidad de los sujetos, y el eje de las ordenadas la probabilidad de respuesta.

Figura 3.10: Gráfico de las Funciones de Respuesta Categóricas para el Ítem 1



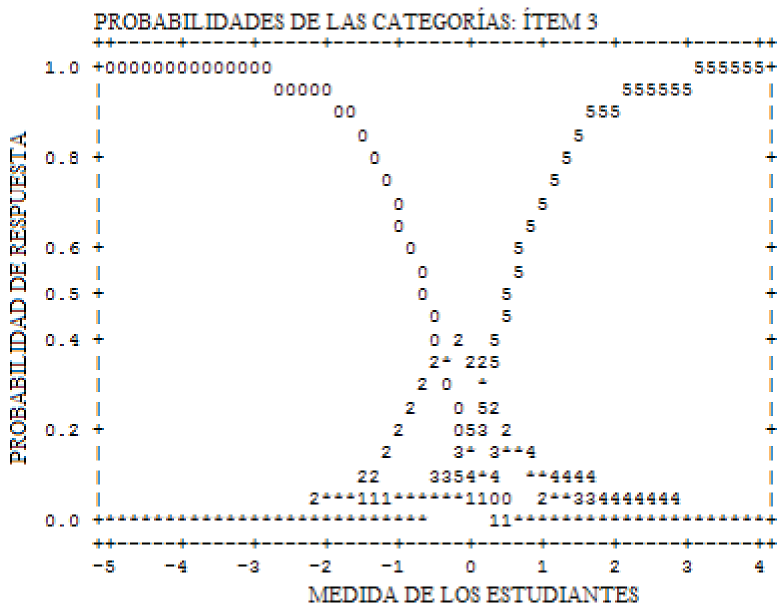
Del gráfico anterior podemos observar que existen varios intervalos donde las respuestas son máximas. Por ejemplo, para estudiantes con niveles de conocimiento menores que -3.5 lógitos ($\theta < b = -3.5$), es más probable que lleguen a obtener una puntuación de 0 (no resolver ninguna parte del problema). Mientras que para estudiantes que tengan un nivel de conocimiento entre -3.5 lógitos y -2 lógitos, lo más probable es que obtengan una puntuación de 1 (resolver correctamente hasta el paso 1) como se observa en el gráfico. Podemos observar también, que para estudiantes con niveles de conocimiento mayores a 1.5 lógitos ($\theta > b = 1.5$) lo más probable es que obtengan una puntuación de 5 (resolver correctamente todo el problema).

Figura 3.11: Gráfica de las Funciones de Respuestas Categóricas para el Ítem 2



Se observa que para niveles de habilidad menores -3.5 lógitos, es más probable que los estudiantes obtengan una puntuación de 0 en el ítem 2 (no resuelvan ninguna parte del problema). Mientras que para niveles de habilidad que se encuentren entre -3.5 lógitos y -1 lógitos, los estudiantes obtendrían una puntuación de 1 (resuelva correctamente el paso 1) y se observa que para estudiantes con niveles de habilidad entre -1 y 1 lógitos, lo más probable es que obtengan una puntuación de 2 (resuelvan correctamente el paso 1 y paso 2). Y los estudiantes con niveles de habilidad mayores a 1.5 lógitos es más probable que resuelvan correctamente el ítem 2.

Figura 3.12: Gráfica de las Funciones de Respuesta para el Ítem 3



En el gráfico de la figura anterior podemos observar algunas categorías donde son máximamente probables, por ejemplo: se observa que estudiantes con niveles de habilidad menores que -0.5 lógitos, es más probable que obtengan un puntuación de 0 (no resuelvan ninguna parte del problema); y para estudiantes con niveles de habilidad entre -0.5 lógitos y 0.4 lógitos es más probable que obtengan una puntuación de 2 (resuelvan correctamente los pasos 1 y 2); y estudiantes con niveles de habilidad mayores a 0.4, es más probable que obtengan una puntuación de 5 (resuelvan correctamente el ítem 3). Si queremos saber las probabilidades para un sujeto de un determinado nivel de habilidad para este ítem, se tendría que aplicar la teoría vista en el capítulo II. Supongamos que tenemos a un estudiante con un nivel de habilidad 1 lógito ($\theta = 1$), y los parámetros estimados para cada paso de este ítem son: $b_{31} = -1.39$, $b_{32} = -2.33$, $b_{33} = 0.70$, $b_{34} = 0.62$, $b_{35} = -0.85$ como se observa en la tabla 3.6 entonces las probabilidades las obtenemos de la siguiente manera:

$$P(0) = \frac{1}{1 + e^{(1-1.39)} + e^{(1-1.39)+(1+2.33)} + e^{(1-1.39)+(1+2.33)+(1-0.70)} + e^{(1-1.39)+(1+2.33)+(1-0.70)+(1-0.62)} + e^{(1-1.39)+(1+2.33)+(1-0.70)+(1-0.62)+(1+0.85)}}$$
$$P(0) = 0.0031$$

La probabilidad de que un estudiante obtenga una puntuación de 1 con un nivel de habilidad igual a 0 es de 0.0031

Para calcular la probabilidad de que un estudiante obtenga una puntuación de 1, se calcula de la siguiente manera:

$$P(1) = \frac{e^{(1-1.39)}}{1 + e^{(1-1.39)} + e^{(1-1.39)+(1+2.33)} + e^{(1-1.39)+(1+2.33)+(1-0.70)} + e^{(1-1.39)+(1+2.33)+(1-0.70)+(1-0.62)} + e^{(1-1.39)+(1+2.33)+(1-0.70)+(1-0.62)+(1+0.85)}}$$
$$P(1) = 0.0021$$

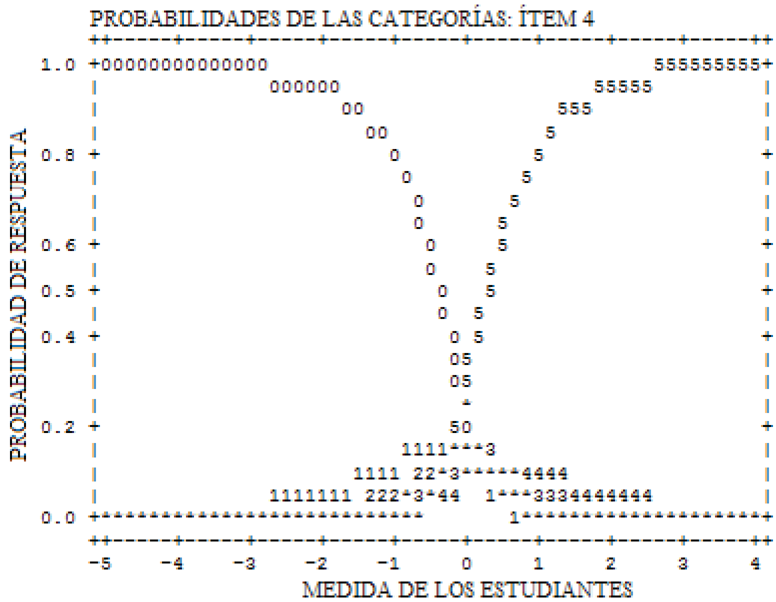
Asimismo en las demás probabilidades obtenemos: $P(2) = 0.05, P(3) = 0.07, P(4) = 0.11, P(5) = 0.74$

En efecto la suma de estas probabilidades nos tendrá que dar el valor de 1.

$$\sum_{i=1}^{5} P(i) = 0.0031 + 0.0021 + 0.05 + 0.07 + 0.11 + 0.74 = 1$$

Las funciones de respuesta categóricas para el ítem 4 se muestran a continuación:

Figura 3.13: Gráfico de las Funciones de Respuestas Categóricas para el Ítem 4



El gráfico de la figura anterior podemos observar que las categorías 1, 2, 3 y 4 no son máximamente probables para ningún valor de θ. Podemos observar que para niveles de habilidad menores a 0 lógitos, es más probable que los estudiantes obtengan una puntuación de 0 (Deficiente); y mientras que para estudiantes con niveles de habilidad mayores a 0 lógitos, es más probable que los estudiantes obtengan una puntuación de 5 (Excelente). Sin embargo, para llegar al paso 5 hay que superar los pasos 1, 2, 3 y 4 donde los pasos más difíciles son el paso 1 y paso 4, por tanto, pocos estudiantes llegaran al paso 5. Un estudiante con un nivel de habilidad 0 lógitos, podrían resolver este ítem correctamente como incorrectamente como se puede apreciar en la figura anterior.

Supongamos que queremos conocer la probabilidad de que un sujeto con nivel de habilidad -2 obtenga una puntuación de 0; además tenemos otro estudiante con nivel de habilidad 2 y queremos saber cual es la probabilidad de que obtenga una puntuación de 5. Los resultados se presentan a continuación:

Con $\theta = -2$, la probabilidad de que un estudiante puntué 0 es:

$$P(0)=\frac{1}{1+e^{(-2-0.91)}+e^{(-2-0.91)+(-2+0.17)}+e^{(-2-0.91)+(-2+0.17)+(-2+0.06)}+e^{(-2-0.91)+(-2+0.17)+(-2+0.06)+(-2-0.41)}+e^{(-2-0.91)+(-2+0.17)+(-2+0.06)+(-2-0.41)+(-2+1.13)}}$$

$$P(0)=0.9392$$

Ahora con $\theta=2$, la probabilidad de que un estudiante puntué 5 es:

$$P(5)=\frac{e^{(2-0.91)+(-2+0.17)+(2+0.06)+(2-0.41)+(2+1.13)}}{1+e^{(2-0.91)}+e^{(2-0.91)+(2+0.17)}+e^{(2-0.91)+(2+0.17)+(2+0.06)}+e^{(2-0.91)+(2+0.17)+(2+0.06)+(2-0.41)}+e^{(2-0.91)+(2+0.17)+(2+0.06)+(2-0.41)+(2+1.13)}}$$

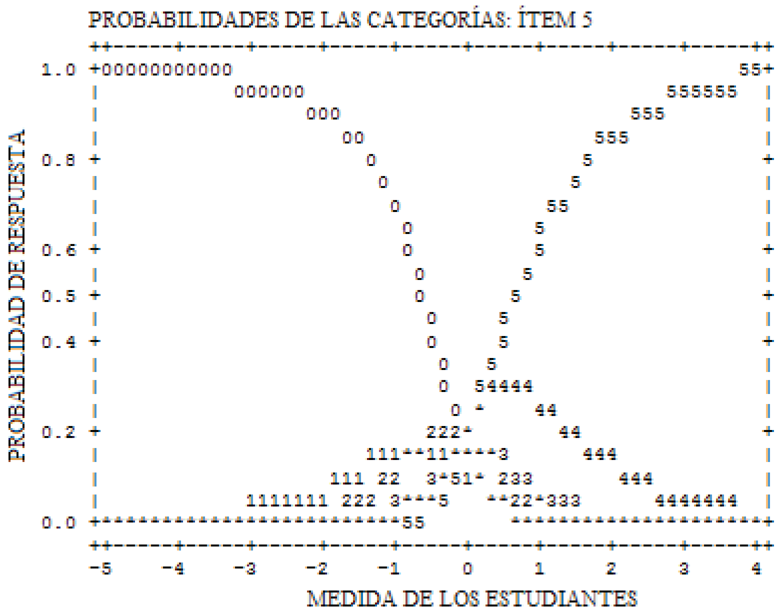
$$P(5)=0.9488$$

Por tanto se observa que las probabilidades van acorde a las funciones de respuesta categóricas.

Como podemos observar de los resultados anteriores, al calcular la probabilidad de que un estudiante falle este ítem con un nivel de habilidad -2 lógitos es cercana a 1 (0.94) e igualmente un estudiante con habilidad 2 lógitos (0.95), por tanto, al observar la dificultad de cada paso (Tabla 3.5), el paso más difícil es el 1 con 0.94 lógitos, entonces esto nos indica que si el estudiante supera el paso 1, resolverá correctamente todo el ítem 4; si no supera el primer paso, este estudiante no resolverá este ítem y obtendrá la puntuación de 0 (Deficiente) lo cual se observa en la figura 3.13.

El gráfico para las funciones de respuesta categóricas para el ítem 5 lo podemos observar a continuación:

Figura 3.14: Gráfico de las Funciones de Respuestas Categóricas para el Ítem 5



De igual manera en la figura anterior podemos observar que las categorías 1, 2, 3 y 4 no son máximamente probables para ningún valor de θ. Podemos observar que para niveles de habilidad menores a 0, es más probable que los estudiantes obtengan una puntuación de 0 (Deficiente); y mientras que para estudiantes con niveles de habilidad mayores a 0, es más probable que los estudiantes obtengan una puntuación de 5 (Excelente). Sin embargo, para llegar al paso 5 hay que superar los pasos 1, 2, 3 y 4 donde el paso más difícil es el paso 1 con dificultad de 1.08 lógitos, por tanto, pocos estudiantes llegaran al paso 5. Si un estudiante supera el paso 1 en este ítem, es seguro que resolverá todo el problema (puntuación de 5); mientras que si un estudiante no logra superar el paso 1, no podrá resolver dicho ítem. Un estudiante con nivel de habilidad 0 lógito, podría resolver este ítem correctamente como incorrectamente como se puede apreciar en la figura anterior.

Las funciones de respuestas categóricas para el Ítem 6 y 7 se presentan en el ANEXO II, debido a que el comportamiento de las curvas es similar a las del ítem 4 e ítem 5, además se obtienen resultados similares ya que las categorías 1,2, 3 y 4 no son máximamente probables para ningún valor de θ.

3.4.2.4 Mapa de Distribución Conjunta (Mapa de ítems y estudiantes)

De acuerdo con la medida y la calibración, sabemos que se pueden referir los resultados a una misma métrica o escala, por esta razón como tercer punto tenemos el mapa de distribución conjunta o mapa de ítems y personas. El gráfico es diseñado verticalmente, al lado izquierdo tenemos a los estudiantes donde las personas más capaces se sitúan en la parte superior; al lado derecho tenemos la ubicación de los ítems y los ítems difíciles se sitúan en la parte superior.

Figura 3.15: Mapa de ítems y Personas

```
ESTUDIANTES                 ÍTEMS
              <frequ>|<less>
    3             ##  +
                      |
                      |
                      |
                      |
                      |
                      |
                      |
    2            .#   +
                      |
                      |
                  .  Q|
                      |
                  #   |
                  .   |
                 .#   |
    1             #   +Q
                .##   |
                  #   |
                  #  S|
               .###   |S  item_5   item_7
             .#####   |
             ######   |   item_6
             ######   |
    0        ######   +M  item_4
            #######   |   item_3
              .####   |
               .###  M|
               ####   |S  item_2
             .#####   |
                      |   item_1
              #####   |
   -1                 +Q
            #######   |
                      |
                     S|
             .#####   |
                      |
                      |
                      |
   -2                 +
                      |
              .####   |
                     Q|
                      |
                      |
                      |
                      |
   -3             ##  +
               <rare>|<more>
```

En el gráfico de la figura anterior, los estudiantes están representados por el símbolo numeral (#). La columna de la parte izquierda ubica la medida de habilidad de los estudiantes a lo largo de la variable. La columna de la parte derecha, localiza la calibración de los ítems a lo largo de la variable. Al buscar un margen equivalente de los ítems a lo largo de la variable (en el eje Y) se observa que existen zonas en donde faltan ítems (ítems con niveles de dificultad menores -1 lógito e ítems con niveles de dificultad mayores a 1 lógito), también se observa que hay acumulación de ítems entre -1 y 1 lógitos. En este caso la escala o los ítems tienen defectos, ya que se esperaría que los ítems se distribuyeran uniformemente a lo largo de toda la escala (recta). Además, el ítem 5 *(Propiedades de los exponentes en expresiones racionales)* y el ítem

7 *(Simplificación y producto de expresiones con números irracionales)* se traslapan entre si, esto nos indica que estos dos ítems están midiendo la misma parte del examen o tienen el mismo nivel de dificultad; lo mismo podríamos decir para el ítem 3 *(Transformación de un decimal a un racional)* y el ítem 4 *(Simplificación de un número racional)* ya que el nivel de dificultad de ambos ítems son muy próximos (ítem 3: -0.09 lógitos, ítem 4: -0.01 lógitos). Por tanto, lo recomendable es eliminar o modificar uno de los ítems en ambos casos. También podemos ver la necesidad de agregar más ítems en el examen con distintos niveles de dificultad (ítems más fáciles e ítems más difíciles), ya que se observa que no existen ítems con niveles de dificultad menores a -1 lógito y mayores a 1 lógito y debido a esto, el examen de Matemática I no esta midiendo de manera adecuada a todos los estudiantes.

Otro aspecto importante que nos muestra el mapa es el comportamiento de los estudiantes que se esperaría que estos se aproximen a una normal. De la figura 3.15 podemos observar que los estudiantes no se aproximan a una normal, lo que nos indica que la mayoría de estudiantes no han comprendido los temas evaluados. Además podemos observar que la mayoría de estudiantes de Matemática I, únicamente habían entendido el 50% o menos del tema evaluado y los ítems únicamente podían ser resueltos correctamente por aquellos estudiantes que hubiesen entendido el 40% o más del tema evaluado. Por lo tanto, es necesario modificar los problemas que comprende el examen y modificar la metodología de la enseñanza con el objetivo de mejorar los resultados en futuros exámenes para los nuevos estudiantes.

3.4.2.5 Medida (calibración) de los Ítems

En este apartado se estudia la medida de los ítems (Calibración), para entender como se encuentran estructuradas las tablas de estas medidas, seguiremos el mismo método utilizado anteriormente donde se explica la información contenida en cada columna. Para describir la tabla 3.6, iniciaremos comentando que la primera columna muestra el número de ítem en el análisis, en la segunda columna se muestra la puntuación bruta correspondiente al parámetro (total de personas que han respondido al menos una opción en cada ítem), en la tercera columna tenemos la medida estimada de cada Ítem (nivel de dificultad), vale la pena mencionar que si el puntaje del Ítem es extremo (sea resuelto correctamente o incorrectamente por todos los estudiantes), su medida es estimada como un MÁXIMO (puntuación perfecta) o un MÍNIMO (puntuación de cero). En la cuarta columna tenemos el error estándar de la estimación, inmediatamente la quinta y sexta columna nos muestran los estadísticos de ajuste al modelo (INFIT y OUTFIT). En la última columna se presenta un estimado de la discriminación del Ítem, llamada punto-biserial. Esta es una correlación, la cual mide la relación entre las puntuaciones de los ítems de cada estudiante y el total de la puntuación de los estudiantes en el test, el valor que debe aparecer ha de ser positivo, en caso contrario existe un claro desajuste de los ítems y menor importancia tendrá el valor del nivel obtenido (esta correlación es opcional).

Tabla 3.6: Medidas de los Ítems

Ítems	Puntuación Bruta	Medida	Error Estándar	INFIT		OUTFIT		Corr. Punt. Bis.
				MNSQ	ZSTD	MNSQ	ZSTD	
ítem_1	413	-0.80	0.1	1.25	1.8	1.23	1.5	A 0.33
ítem_3	381	-0.09	0.06	1.13	1.2	0.97	-0.1	B 0.53
ítem_4	350	-0.01	0.05	1.12	1.1	1.06	0.2	C 0.51
ítem_6	254	0.30	0.06	1.09	0.7	1.05	0.1	D 0.54
ítem_7	169	0.53	0.07	1.01	0.1	0.9	-0.2	c 0.54
ítem_5	210	0.56	0.06	0.79	-1.8	0.71	-1.3	b 0.66
ítem_2	387	-0.48	0.09	0.79	-1.9	0.76	-2.2	a 0.65
Media	309	0.00	0.07	1.03	0.20	0.95	-0.30	
S.D.	90	0.47	0.02	0.16	1.40	0.17	1.10	

Del cuadro anterior podemos observar qué ítems son los más fáciles y cuales los más difíciles. Los ítems con medida negativa se puede decir que son los más fáciles del test, pero podemos considerarlos aceptables debido a que se encuentran cerca del valor de 0 (donde el nivel de dificultad del ítem es del 50%). Mientras que al observar qué ítem es el más difícil es el ítem 5 el cual tiene la mayor medida en lógitos (0.56). También es necesario observar que todos los ítems se encuentran muy próximos a 0, con lo cual reafirmamos que es necesario agregar más ítems con distintos niveles de dificultad al primer examen de Matemática I.

Ahora ¿qué tan bien ajustados se encuentran estos ítems al modelo de Rasch? , En el caso de la discriminación punto-biserial podemos ver que todos los valores son positivos lo cual nos indicarían que los ítems se comportan como el modelo esperaba, sin embargo, tenemos que ver, también, los estadísticos de ajuste; para ello tenemos que recordar que disponemos de un conjunto de aceptación de los ítems: aceptaremos aquellos que se encuentren entre -2 y 2 lógitos. Al observar el ajuste de los ítems (FIT), el ítem 2 que se refiere a "*Propiedades de los números reales*" presenta un INFIT de -1.9 y un OUTFIT de -2.2 lo que indica que en el ajuste externo de este ítem, se presenta bastante determinismo o poca estocasticidad en las respuestas de los estudiantes. En los demás ítems podemos observar que tanto el ajuste interno (INFIT) como su ajuste externo (OUTFIT) no presentan desajuste en los datos, ya que los valores se encuentran dentro del intervalo de aceptación (-2 lógitos y 2 lógitos).

3.4.2.6 Medida (habilidad) de los Estudiantes

En la tabla 3.7 se presenta la estimación de la habilidad de los estudiantes, esta tabla posee las mismas columnas que la correspondiente a la medida de los ítems, por esta razón nos limitaremos a presentarla. En primer lugar hay que resaltar que los estudiantes E81, E107, E123 y E136 han obtenido las mayores medidas en el parcial I, de Matemática I. Podemos decir que estos estudiantes forman la categoría más alta (Excelente) en cuanto al nivel de conocimientos en el área evaluada (según la TCT). El programa BIGSTEPS no puede obtener una estimación de los estadísticos de ajuste, debido a que los estudiantes presentan un patrón igual en todos los ítems que han contestado, por esta razón el programa les ha asignado una medida de 3.11 lógitos:

Tabla 3.7: Estudiantes con Máximas Medidas

Estudiante	Puntuación Bruta	Medida	Error Estándar	INFIT		OUTFIT		Corr. Punt. Bis.
				MNSQ	ZSTD	MNSQ	ZSTD	
E81	35	3.11	1.36	MÁXIMA MEDIDA ESTIMADA				
E107	35	3.11	1.36	MÁXIMA MEDIDA ESTIMADA				
E123	35	3.11	1.36	MÁXIMA MEDIDA ESTIMADA				
E136	35	3.11	1.36	MÁXIMA MEDIDA ESTIMADA				

La puntuación bruta de estos estudiantes nos indica que en todos los ítems han obtenido una puntuación de 5 (Excelente), lo cual nos lleva a obtener una puntuación bruta de 35, que es precisamente la suma de la puntuación 5 en los 7 ítems que forman el test. Por tal razón el programa indica que son los que han obtenido la máxima medida (esto se observa en el escalograma de Guttman, ANEXO III). Para analizar estos estudiantes, se tendría que aplicar los índices conocidos dentro de la TCT que se estudió en el capítulo II, ya que se presenta un escalograma de Guttman Perfecto. La medida de estos estudiantes nos indican que tienen una probabilidad mayor al 90% de contestar correctamente el test (prácticamente el test no presenta ningún nivel de dificultad para ellos), esta probabilidad la podemos obtener mediante las expresiones vistas en la sección 2.2.4.1 del capítulo II. En conclusión para estos estudiantes lo recomendable sería darles un tratamiento adecuado al nivel de conocimientos que ellos presentan.

En segundo lugar, se analiza los estudiantes que han tenido una puntuación “*muy buena*”, ya que sus medidas son las que les siguen a los estudiantes que han tenido las máximas puntuaciones, estas medidas van desde 1.08 lógitos hasta 1.94 lógitos. En la siguiente tabla podemos ver las medidas por estudiante:

Tabla 3.8: Estudiantes con Medidas que se Consideran *"Muy Buenas"*

Estudiante	Puntuación Bruta	Medida	Error Estándar	INFIT		OUTFIT		Corr. Punt. Bis.
				MNSQ	ZSTD	MNSQ	ZSTD	
E96	32	1.94	0.64	1.81	0.60	3.42	1.00	0.54
E108	32	1.94	0.64	1.05	0.00	0.57	-0.30	-0.16
E106	31	1.62	0.52	0.84	-0.20	0.80	-0.20	-0.20
E89	30	1.40	0.44	0.58	-0.60	0.54	-0.50	-0.10
E127	30	1.40	0.44	0.60	-0.60	0.65	-0.40	0.36
E18	29	1.22	0.39	0.38	-1.10	0.32	-1.00	0.09
E59	28	1.08	0.36	0.16	-1.90	0.18	-1.50	0.04
E64	28	1.08	0.36	0.66	-0.50	0.51	-0.70	0.25
E65	28	1.08	0.36	0.48	-0.90	0.45	-0.80	-0.35

Como podemos observar en la tabla anterior, el error nos podría servir para saber en que intervalo se encuentra el verdadero parámetro de dicho sujeto, por ejemplo para el estudiante E89 tenemos un intervalo de confianza del 95% de:

$$1.40-(1.96)(0.44)\leq\theta\leq1.40+(1.96)(0.44)$$
$$0.5376\leq\theta\leq2.2624$$

Lo que nos indica que de 100 muestras que tomemos de forma repetitiva, 95 de ellas tendrán el verdadero valor del parámetro (nivel de habilidad) entre 0.5376 y 2.2624. Por último, podemos mencionar que la probabilidad de contestar correctamente el test por estos estudiantes es aproximadamente mayor al 75%.

Existe una gran mayoría de estudiantes que se encuentran con niveles de medida alrededor de 0 lógitos (-0.82 lógitos y 0.96 lógitos), los cuales pueden verse en la sección 3.4.2.9. El estudiante que más sobresale en esta categoría es el E6 con una medida de 0.96 lógitos y el estudiante con menor medida o nivel de conocimientos es el E168 con -0.82 lógitos. De todos los estudiantes podemos decir que el nivel de conocimientos de ellos es *"Bueno"* ya que se encuentran alrededor de la media de 0 lógitos; recordemos que si un estudiante tiene un lógito de 0 podemos decir que tiene una probabilidad del 50% de que conteste correctamente el test.

Además existe un buen grupo de estudiantes con niveles de habilidad que podemos categorizarlos como *"Regular"* ya que sus niveles de conocimiento se encuentran por debajo de la medida (-1.52 lógitos y lógitos -1.09), la cual se puede observar en la siguiente tabla:

Tabla 3.9: Estudiantes con Medidas Consideradas *"Regular"*

Estudiante	Puntuación Bruta	Medida	Error Estándar	INFIT		OUTFIT		Corr. Punt. Bis.
				MNSQ	ZSTD	MNSQ	ZSTD	
E39	5	-1.09	0.58	0.55	-0.50	0.63	-0.30	0.63
E41	5	-1.09	0.58	0.35	-0.80	0.26	-0.70	0.73
E62	5	-1.09	0.58	1.28	0.20	1.05	0.00	0.66
E70	5	-1.09	0.58	0.81	-0.20	0.85	-0.10	0.66
E75	5	-1.09	0.58	0.59	-0.40	0.70	-0.20	0.71
E94	5	-1.09	0.58	0.83	-0.10	0.46	-0.50	0.77
E103	5	-1.09	0.58	0.83	-0.10	0.46	-0.50	0.77
E111	5	-1.09	0.58	0.81	-0.20	0.85	-0.10	0.66
E114	5	-1.09	0.58	0.48	-0.60	0.84	-0.10	0.40
E132	5	-1.09	0.58	0.55	-0.50	0.63	-0.30	0.63
E133	5	-1.09	0.58	0.42	-0.70	0.47	-0.50	0.47
E139	5	-1.09	0.58	0.83	-0.10	0.46	-0.50	0.77
E143	5	-1.09	0.58	0.74	-0.20	1.07	0.00	0.42
E11	4	-1.52	0.75	0.24	-0.90	0.16	-0.70	0.64
E26	4	-1.52	0.75	0.73	-0.20	2.06	0.40	0.31
E29	4	-1.52	0.75	2.21	0.70	1.26	0.10	0.71
E36	4	-1.52	0.75	0.71	-0.30	1.46	0.20	0.50
E57	4	-1.52	0.75	0.24	-0.90	0.16	-0.70	0.64
E84	4	-1.52	0.75	2.64	0.90	4.20	1.00	-0.04
E118	4	-1.52	0.75	0.93	-0.10	0.74	-0.10	0.58
E134	4	-1.52	0.75	0.24	-0.90	0.16	-0.70	0.64
E155	4	-1.52	0.75	0.88	-0.10	0.61	-0.20	0.72
E169	4	-1.52	0.75	0.24	-0.90	0.16	-0.70	0.64
E173	4	-1.52	0.75	0.24	-0.90	0.16	-0.70	0.64

El estudiante con mayor medida es el E39 y el estudiante con menor medida es E173. De esta categoría podemos decir que la probabilidad que conteste el test correctamente es menor al 30%.

Por último, hay estudiantes que han obtenido medidas por debajo de -2.00 lógitos, esto nos indica que estos estudiantes han obtenido los peores resultados y en la mayoría de los ítems del test obtuvieron una calificación nula, por tal razón podemos categorizar estos estudiantes como *"Deficientes"*. La tabla se presenta a continuación:

Tabla 3.10: Estudiantes con Medidas Consideradas *"Deficientes"*

Estudiante	Puntuación Bruta	Medida	Error Estándar	INFIT		OUTFIT		Corr. Punt. Bis.
				MNSQ	ZSTD	MNSQ	ZSTD	
E31	3	-2.26	0.97	0.18	-1.10	0.09	-0.60	0.62
E38	3	-2.26	0.97	0.18	-1.10	0.09	-0.60	0.62
E54	3	-2.26	0.97	3.10	1.20	1.24	0.10	0.46
E55	3	-2.26	0.97	0.18	-1.10	0.09	-0.60	0.62
E69	3	-2.26	0.97	0.18	-1.10	0.09	-0.60	0.62
E73	3	-2.26	0.97	0.18	-1.10	0.09	-0.60	0.62
E142	3	-2.26	0.97	1.01	0.00	0.42	-0.30	0.56
E165	3	-2.26	0.97	0.18	-1.10	0.09	-0.60	0.62
E91	2	-3.29	1.04	1.47	0.50	0.47	-0.10	0.47
E135	2	-3.29	1.04	0.07	-2.00	0.03	-0.40	0.64
E22	1	-4.47	1.19	0.51	-0.70	0.16	-0.20	0.47

La probabilidad de que estos estudiantes resuelvan correctamente el test es menor al 10%, por tanto, es recomendable para estos estudiantes darles un curso especial para retroalimentar los conocimientos evaluados en el examen, ya que sus niveles de conocimiento son muy bajos.

La siguiente tabla muestra un resumen de los estadísticos de ajuste interno (INFIT) y ajuste externo (OUTFIT), esta tabla solo muestra los estudiantes que tienen los mayores o menores estadísticos de ajuste. La siguiente tabla esta ordenada por los estadísticos de ajuste y su estructura es similar a las tablas anteriores.

Tabla 3.11: Estadísticos de Ajustes para los Estudiantes

Estudiantes	Puntuación Bruta	Medida	Error Estándar	INFIT		OUFIT		Corr. Punt. Bis.
				MNSQ	ZSTD	MNSQ	ZSTD	
E84	4	-1.52	0.75	2.64	0.90	4.20	1.00	A-.04
E98	26	0.86	0.31	2.68	1.90	4.12	2.60	B-.09
E96	32	1.94	0.64	1.81	0.60	3.42	1.00	C .54
E6	27	0.96	0.33	3.28	2.10	3.38	2.00	D .29
E99	26	0.86	0.31	3.19	2.20	3.14	2.00	E .15
E7	6	-0.82	0.47	2.79	1.20	3.16	1.20	F .48
E61	7	-0.63	0.40	3.16	1.50	3.01	1.30	G .45
E54	3	-2.26	0.97	3.10	1.20	1.24	0.10	H .46
E93	11	-0.19	0.29	1.79	1.10	3.05	2.10	I .36
E1	9	-0.37	0.33	2.60	1.50	2.23	1.20	J .06
E151	9	-0.37	0.33	2.56	1.50	2.39	1.30	K .03
E42	19	0.34	0.25	1.72	1.50	2.42	2.40	L .61
E10	10	-0.27	0.30	2.35	1.50	1.97	1.10	M .12
E110	8	-0.49	0.36	2.27	1.10	2.35	1.10	N .16
E126	26	0.86	0.31	1.65	0.90	2.23	1.40	O-.34
E29	4	-1.52	0.75	2.21	0.70	1.26	0.10	P .71
E92	13	-0.04	0.26	1.19	0.40	2.11	1.60	Q .76
E26	4	-1.52	0.75	0.73	-0.20	2.06	0.40	R .31
E130	19	0.34	0.25	2.03	2.00	1.79	1.50	S-.78
E152	13	-0.04	0.26	1.99	1.70	2.00	1.50	T .01
E60	9	-0.37	0.33	1.94	1.00	1.09	0.10	U .50
E105	15	0.10	0.25	1.91	1.80	1.82	1.40	V-.19
E128	27	0.96	0.33	1.83	1.00	1.63	0.70	W .31
E102	10	-0.27	0.30	0.92	-0.10	1.82	1.00	X .78
E82	8	-0.49	0.36	1.29	0.30	1.76	0.70	Y-.04
E87	16	0.16	0.25	1.74	1.50	1.36	0.70	Z-.10
E52	18	0.28	0.25	0.29	-2.60	0.35	-2.10	z .40
E58	6	-0.82	0.47	0.34	-0.90	0.20	-1.00	y .81
E72	6	-0.82	0.47	0.34	-0.90	0.20	-1.00	x .81
E167	7	-0.63	0.40	0.31	-1.00	0.33	-0.90	w .63
E50	26	0.86	0.31	0.33	-1.50	0.27	-1.50	v .28
E116	10	-0.27	0.30	0.33	-1.50	0.30	-1.40	u .77
E83	9	-0.37	0.33	0.27	-1.40	0.28	-1.30	t .79
E53	20	0.41	0.25	0.27	-2.60	0.28	-2.40	s .16
E45	7	-0.63	0.40	0.28	-1.10	0.23	-1.20	r .77
E86	7	-0.63	0.40	0.26	-1.20	0.22	-1.20	q .70
E141	7	-0.63	0.40	0.26	-1.20	0.22	-1.20	p .70
E11	4	-1.52	0.75	0.24	-0.90	0.16	-0.70	o .64
E57	4	-1.52	0.75	0.24	-0.90	0.16	-0.70	n .64
E134	4	-1.52	0.75	0.24	-0.90	0.16	-0.70	m .64
E169	4	-1.52	0.75	0.24	-0.90	0.16	-0.70	l .64
E173	4	-1.52	0.75	0.24	-0.90	0.16	-0.70	k .64
E59	28	1.08	0.36	0.16	-1.90	0.18	-1.50	j .04
E31	3	-2.26	0.97	0.18	-1.10	0.09	-0.60	i .62
E38	3	-2.26	0.97	0.18	-1.10	0.09	-0.60	h .62
E55	3	-2.26	0.97	0.18	-1.10	0.09	-0.60	g .62
E69	3	-2.26	0.97	0.18	-1.10	0.09	-0.60	f .62
E73	3	-2.26	0.97	0.18	-1.10	0.09	-0.60	e .62
E165	3	-2.26	0.97	0.18	-1.10	0.09	-0.60	d .62
E67	8	-0.49	0.36	0.11	-1.90	0.13	-1.70	c .94
E144	6	-0.82	0.47	0.11	-1.50	0.10	-1.30	b .91
E135	2	-3.29	1.04	0.07	-2.00	0.03	-0.40	a .64
Media	13	-0.33	0.42	0.97	-0.20	0.95	-0.20	
S.D.	8	1.00	0.22	0.66	0.90	0.77	0.90	

En la tabla anterior los valores que aparecen de color rojo nos indican desajuste en los estudiantes. Podemos observar que los estudiantes E6 (INFIT: 2.10, OUTFIT: 2.00), E99 (INFIT: 2.20, OUTFIT: 2.00) presentan valores mayores o iguales a 2 lógitos tanto en su ajuste interno (INFIT) como en el ajuste externo (OUTFIT), por lo que podemos decir que estos estudiantes presentan demasiada estocasticidad en sus respuestas (patrón ruidoso), lo que de alguna manera podríamos decir que dichos estudiantes han resuelto el examen sin saber lo que hacen. Los estudiantes E52 (INFIT: -2.60, OUTFIT: -2.10), E53 (INFIT: -2.60, OUTFIT: -2.40) presentan valores menores a -2 lógitos en su ajuste interno (INFIT) como en su ajuste externo (OUTFIT), esto indica que dichos estudiantes presentan demasiado determinismo en sus respuestas (poca estocasticidad) es decir casi un escalograma de Guttman perfecto. Además se tiene estudiantes que solo presentan desajuste interno (INFIT), estos estudiantes son: E130 (INFIT: 2.00) y E135 (INFIT: -2.00), hay estudiantes que solo presentan desajuste externo OUTFIT, estos estudiantes son: E98 (OUTFIT: 2.60), E42 (OUTFIT: 2.40), E93 (OUTFIT: 2.10). Por lo tanto, podemos decir que estos estudiantes, presentan desajustes con respecto a sus respuestas.

3.4.2.7 Desajustes de los Estudiantes

Las medidas de los sujetos se complementan con el análisis detallado del origen del desajuste entre el valor real y el modelo. Ello permite considerar las acciones necesarias para corregir las causas del desajuste, en el caso que sea necesario. Así la siguiente tabla nos muestra los sujetos que según el modelo de Rasch presentan desajuste, y estos resultados nos ayudarán a darle mayor veracidad al análisis que anteriormente se ha hecho con los estudiantes respecto a sus estadísticos FIT (INFIT y OUTFIT).

Figura 3.16: Desajuste de los Estudiantes

```
DESAJUSTES DE LOS ESTUDIANTES    (EN ORDEN DE ENTRADA)
NÚMERO - NOMBRE -- POSICIÓN ------ MEDIDA - INFIT (ZSTD) OUTFIT

   98 E98                          .86      1.9    B     2.6
       RESPONSE:     1:  5 4 4 0 5  5 3
     Z-RESIDUAL:              -4

    6 E6                           .96      2.1    D     2.0
       RESPONSE:     1:  5 5 5 3 5  0 5
     Z-RESIDUAL:              -2   -3

   99 E99                          .86      2.2    E     2.0
       RESPONSE:     1:  5 5 2 5 5  0 5
     Z-RESIDUAL:            -2     -3

   93 E93                         -.19      1.1    I     2.1
       RESPONSE:     1:  5 3 0 0 3  0 0
     Z-RESIDUAL:         3

   42 E42                          .34      1.5    L     2.4
       RESPONSE:     1:  5 5 2 3 4  0 0
     Z-RESIDUAL:         2 2

  130 E130                         .34      2.0    S     1.5
       RESPONSE:     1:  2 2 2 0 5  4 5
     Z-RESIDUAL:              -2
```

En la tabla de la figura anterior podemos observar: el nombre del estudiante (en nuestro caso el código que le hemos designado), posición, medida del estudiante, y los estadísticos de ajuste. Así los desajustes de los estudiantes recogidos en la sección 3.4.2.6 se complementa con la expresión de qué ítem es el que genera el desajuste. Los valores residuales positivos indican que han puntuado al ítem por encima del valor que esperaba el modelo. Los valores residuales negativos indican que han sido puntuados más bajos de lo que el modelo esperaba. El estudiante E98 ha puntuado 4 unidades menos en el ítem 4 *(Simplificación de un número racional)* de lo que esperaba el modelo; podemos ver también los valores del INFIT y OUTFIT, el valor del INFIT es de 1.9 lógitos y el valor de OUTFIT es de 2.6 lógitos. Asimismo el sujeto E6 ha puntuado 2 unidades menos en el ítem 4 *(Simplificación de un número racional)* y ha puntuado 3 unidades menos en el ítem 6 *(Simplificación de expresiones con números irracionales)* de lo que el modelo esperaba. En el estudiante E99 podemos ver que los residuales son negativos en el ítem 3 *(Transformación de un decimal a un racional)* y el ítem 6 *(Simplificación de expresiones con números irracionales)*, lo que nos indica que han puntuado el ítem por debajo de lo que el modelo esperaba. El estudiante E93 ha puntuado 3 unidades más en el ítem 1 (*Propiedades de los números racionales e irracionales)* de lo que el modelo esperaba. De forma similar podemos observar los residuales del estudiante E42, ya que ha puntuado dos unidades más en el ítem 1 (*Propiedades de los números racionales e irracionales)* y el ítem 2 *(Propiedades de los números reales)* de lo que el modelo esperaba. Por último, el estudiante E130 ha puntuado dos unidades menos en el ítem 4 *(Simplificación de un número racional)*. Ante esta situación podemos decir que los estudiantes presentados en esta tabla muestran un desajuste que afectan la obtención de las medidas de los estudiantes. Con este análisis, se determina que estudiantes están afectando al diseño del instrumento de medida, por lo cual se aconseja eliminar estos sujetos de la base.

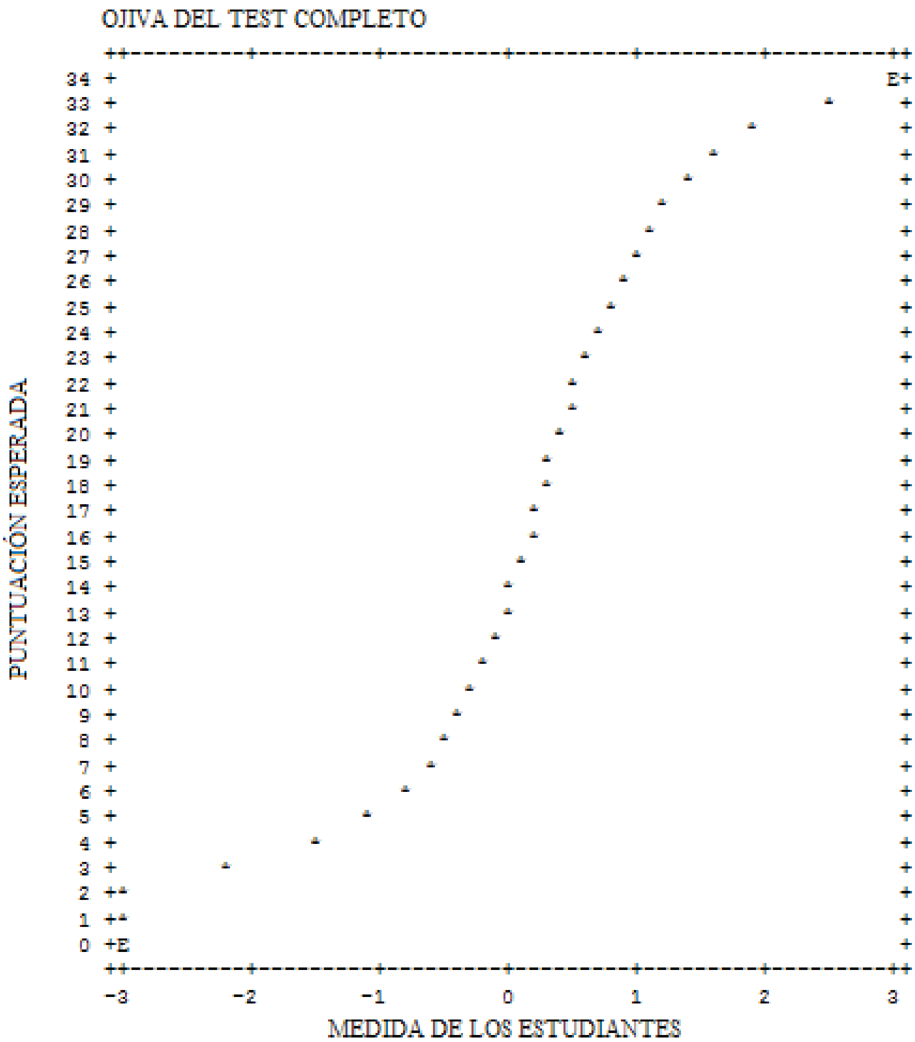
De la misma manera el programa BIGSTEPS presenta una tabla con el contenido de los ítems que presentan desajuste, en nuestro caso el programa no nos da resultados acerca de los ítems que presentan desajustes, como se observó en la sección 3.4.2.5 la mayoría de los estadísticos de ajuste interno (INFIT) y ajuste externo (OUTFIT) para los ítems se encuentran dentro del intervalo de aceptación.

En conclusión tenemos 6 estudiantes (E98, E6, E99, E93, E42 y E130) que presentan desajustes con el modelo de Rasch y no hay ítems que estén afectando la obtención de las medidas de estudiantes e ítems.

3.4.2.8 Gráficos del Test Completo

El programa BIGSTEPS proporciona una curva para el test completo donde se recogen las puntuaciones que mide el instrumento. Esta gráfica resulta de dibujar la medida del estudiante contra las puntuaciones que se pueden obtener en el test, la cual esta asociada a la dificultad del ítem. En el eje de las abscisas podemos observar la medida de los estudiantes y en el eje de las ordenadas las puntuaciones que mide el instrumento. La letra *"E"* que se observa en los extremos de la curva, nos indica la puntuación mínima y la puntuación máxima respectivamente.

Figura 3.17: Gráfica de la Curva del Test Completo



Observemos que un estudiante con medida $\theta = 0$ puede obtener 12 puntos de un total de 35 puntos, por lo que podría resolver correctamente el 34.29% del test (parcial I, Matemática I). Un estudiante con un nivel de rasgo $\theta = -1$, obtendría una puntuación de 5 unidades, lo que significa que solo podría resolver correctamente un 14.29% del test completo. Observemos que a menor medida de un estudiante, menor será la puntuación que obtendrá en el test; caso contrario sucede en un estudiante con mayor medida, que tendrá una mayor puntuación en el test. Tal es el caso de un estudiante con un nivel de rasgo 2 lógitos, que obtendrá una puntuación de 32 puntos, lo que equivale a resolver correctamente el 91.43% del primer test de Matemática I.

En la siguiente tabla se muestra los posibles puntajes que pueden ser obtenidos por el test (la mayor puntuación de 35 puntos equivale al 100% de la prueba), la medida del estudiante y su error de estimación.

Tabla 3.12: Medidas del Test Completo

Puntuación	Medida	S.E	Puntuación	Medida	S.E	Puntuación	Medida	S.E
0	-5.33E	1.53	12	-0.10	0.27	24	0.68	0.28
1	-4.45	1.19	13	-0.03	0.26	25	0.76	0.29
2	-3.27	1.04	14	0.03	0.26	26	0.85	0.31
3	-2.25	0.97	15	0.10	0.25	27	0.95	0.33
4	-1.50	0.75	16	0.16	0.25	28	1.07	0.36
5	-1.07	0.57	17	0.22	0.25	29	1.21	0.39
6	-0.81	0.47	18	0.28	0.25	30	1.38	0.44
7	-0.62	0.40	19	0.34	0.25	31	1.61	0.51
8	-0.48	0.36	20	0.40	0.25	32	1.92	0.63
9	-0.36	0.32	21	0.47	0.26	33	2.49	0.92
10	-0.27	0.30	22	0.54	0.26	34	3.09E	1.34
11	-0.18	0.28	23	0.61	0.27			

La tabla anterior muestra los puntajes que pueden ser obtenidos en el test junto con la medida del estudiante y su error de estimación. Podemos observar que el error de estimación en los extremos es más grande que en el centro, como se mencionó en el capítulo II, solo con datos que se encuentran en el centro se puede obtener un error de estimación más preciso.

Además de la curva anterior, el programa BIGSTEPS proporciona también, curvas referidas a los ítems que comprende el test. La siguiente figura responde a la pregunta ¿Cuál es el promedio evaluado que nosotros esperamos observar para una persona de medida en particular? En este gráfico se observa en el extremo izquierdo la categoría más baja, en este caso es 0, y en el extremo derecho de la figura podemos observar la categoría más alta, la cual es 5. Los dos puntos que se observan en el gráfico nos indican la zona que divide una categoría de la otra. Podemos observar también en el lado derecho de la figura, los ítems que comprende el test, se esperaría que entre cada ítem no hubieran "huecos" para decir que se está midiendo bien la escala del test.

Figura 3.18: Promedio Evaluado que se Esperaría de los Estudiantes

```
    EXPECTED SCORE: MEAN  (":" INDICATES HALF-SCORE POINT)
 -3       -2        -1         0         1         2         3
 |---------+---------+---------+---------+---------+---------|  PARCIAl 1
 0                   0   :   1  : 2: 3 :  4   :     5        5  Ítem_5
 0                      0  :  1 : 2 : 3   :    5             5  Ítem_7
 |                                                           |
 0                    0  :  1 : 2:3 : 4  :  5                5  Ítem_6
 |                                                           |
 0                 0   :  1 : 2:3 : 4  :  5                  5  Ítem_4
 0             0   :   1  : 2 : 3 :  4   :   5               5  Ítem_3
 |                                                           |
 |                                                           |
 0  1           :         2      :   3   :   4      :    5   5  Ítem_2
 |                                                           |
 0    :             2           :    3   :   4    :      5   5  Ítem_1
 |---------+---------+---------+---------+---------+---------|
 -3       -2        -1         0         1         2         3
```

En este gráfico podemos observar los espacios o "huecos" que existe entre los ítems (extremo derecho), como ya se ha comentado en el mapa de distribución conjunta no se esta midiendo bien la escala (los ítems que presentan espacios en este gráfico son los mismos que presentaban espacios en el mapa de distribución conjunta). Podemos observar que para un estudiante con medida $\theta = -1$ se espera que en promedio obtenga una puntuación de 2 (Regular) en el ítem 1, el cual se refiere a propiedades de los números racionales e irracionales. Para un estudiante con $\theta = 2.2$, se espera que en promedio obtenga una puntuación de 5 en el ítem 5 que se refiere a propiedades de los exponentes en expresiones racionales. Un estudiante con un nivel de rasgo de -0.9 se espera que en promedio falle en el ítem 5.

Otro gráfico interesante es el que presenta la respuesta más probable que se puede esperar de un sujeto con una medida particular. La estructura del gráfico es similar al mostrado anteriormente.

Figura 3.19: Respuestas más Probables de los Estudiantes

```
MOST PROBABLE RESPONSE: MODE  (BETWEEN "0" AND "1" IS "0", ETC.)
 -3       -2        -1         0         1         2         3
 |---------+---------+---------+---------+---------+---------|  PARCIAL 1
 0                                    5                      5  item_5
 0                                  5                        5  item_7
 |                                                           |
 0                                5                          5  item_6
 |                                                           |
 0                             5                             5  item_4
 0                        2      5                           5  item_3
 |                                                           |
 |                                                           |
 0                2                     3  5                 5  item_2
 |                                                           |
 02                                      3 5                 5  item_1
 |---------+---------+---------+---------+---------+---------|
 -3       -2        -1         0         1         2         3
```

Podemos observar que las categorías más probables son las categorías 2 (regular) y 5 (excelente), las categorías restantes no son mostradas por no ser consideradas respuestas más probables. Podemos ver que un estudiante con habilidad 0 podría resolver el ítem 4 *(Simplificación de un número racional)* con una puntuación de 5 (si supera las etapas del 1 al 4). Podemos observar también que un estudiante con nivel de rasgo 1.1 lógitos obtendría una puntuación de 5 tanto en el ítem 1 *(Propiedades de los números racionales e irracionales)* como en el ítem 2 *(Propiedades de los números reales)*. Sin embargo, no podemos asegurar que los ítems 4, 2 y 1 sean comprendidos con mayor facilidad por todos los estudiantes.

3.4.2.9 Ranking de las Medidas de los Estudiantes e Ítems

En la siguiente tabla se presenta el resultado final de las medidas de los estudiantes, eliminando a aquellos que han obtenido puntuaciones extremas (todos los ítems resueltos correctamente e incorrectamente). En la primera columna se presenta el orden ascendente (posición que ocupa en el Ranking el estudiante), la segunda columna muestra el código del estudiante y la tercera columna la medida obtenida por el estudiante.

Tabla 3.13: Ranking de Medidas de los Estudiantes

Corr.	Estudiante	Medida
1	E47	1.94
2	E96	1.94
3	E108	1.94
4	E106	1.62
5	E89	1.40
6	E127	1.40
7	E18	1.22
8	E59	1.08
9	E64	1.08
10	E65	1.08
11	E6	0.96
12	E128	0.96
13	E27	0.86
14	E50	0.86
15	E98	0.86
16	E99	0.86
17	E126	0.86
18	E66	0.77
19	E76	0.77
20	E43	0.69
21	E109	0.69
22	E33	0.54
23	E44	0.54
24	E100	0.54
25	E129	0.54
26	E88	0.47
27	E113	0.47
28	E174	0.47
29	E40	0.41
30	E53	0.41
31	E150	0.41
32	E161	0.41
33	E164	0.41
34	E24	0.34
35	E42	0.34
36	E71	0.34
37	E119	0.34
38	E130	0.34
39	E149	0.34
40	E52	0.28
41	E120	0.28
42	E121	0.28
43	E158	0.28
44	E166	0.28
45	E17	0.22
46	E19	0.22
47	E20	0.22
48	E63	0.22
49	E124	0.22
50	E131	0.22
51	E137	0.22
52	E2	0.16
53	E35	0.16
54	E87	0.16
55	E154	0.16
56	E175	0.16
57	E12	0.10
58	E77	0.10
59	E105	0.10
60	E115	0.10
61	E153	0.10
62	E157	0.10
63	E160	0.10
64	E9	0.03
65	E23	0.03
66	E56	0.03
67	E101	0.03
68	E159	0.03
69	E172	0.03
70	E3	-0.04
71	E46	-0.04
72	E49	-0.04
73	E92	-0.04
74	E152	-0.04
75	E176	-0.04
76	E14	-0.11
77	E16	-0.11
78	E28	-0.11
79	E85	-0.11
80	E138	-0.11
81	E8	-0.19
82	E30	-0.19
83	E79	-0.19
84	E80	-0.19
85	E93	-0.19
86	E104	-0.19
87	E140	-0.19
88	E148	-0.19
89	E163	-0.19
90	E5	-0.27
91	E10	-0.27
92	E32	-0.27
93	E51	-0.27
94	E78	-0.27
95	E90	-0.27
96	E97	-0.27
97	E102	-0.27
98	E116	-0.27
99	E1	-0.37
100	E60	-0.37
101	E74	-0.37
102	E83	-0.37
103	E146	-0.37
104	E147	-0.37
105	E151	-0.37
106	E15	-0.49
107	E37	-0.49
108	E67	-0.49
109	E82	-0.49
110	E110	-0.49
111	E117	-0.49
112	E122	-0.49
113	E171	-0.49
114	E45	-0.63
115	E61	-0.63
116	E68	-0.63
117	E86	-0.63
118	E95	-0.63
119	E112	-0.63
120	E141	-0.63
121	E145	-0.63
122	E162	-0.63
123	E167	-0.63
124	E170	-0.63
125	E7	-0.82
126	E21	-0.82
127	E25	-0.82
128	E48	-0.82
129	E58	-0.82
130	E72	-0.82
131	E125	-0.82
132	E144	-0.82
133	E156	-0.82
134	E168	-0.82
135	E13	-1.09
136	E39	-1.09
137	E41	-1.09
138	E62	-1.09
139	E70	-1.09
140	E75	-1.09
141	E94	-1.09
142	E103	-1.09
143	E111	-1.09
144	E114	-1.09
145	E132	-1.09
146	E133	-1.09
147	E139	-1.09
148	E143	-1.09
149	E11	-1.52
150	E26	-1.52
151	E29	-1.52
152	E36	-1.52
153	E57	-1.52
154	E84	-1.52
155	E118	-1.52
156	E134	-1.52
157	E155	-1.52
158	E169	-1.52
159	E173	-1.52
160	E4	-2.26
161	E31	-2.26
162	E38	-2.26
163	E54	-2.26
164	E55	-2.26
165	E69	-2.26
166	E73	-2.26
167	E142	-2.26
168	E165	-2.26
169	E34	-3.29
170	E91	-3.29
171	E135	-3.29
172	E22	-4.47

De la tabla anterior se observa que la mayor parte de los estudiantes se encuentran en el intervalo de -0.79 a 0.96 lógitos.

A continuación se presenta el Ranking de medidas de los ítems, donde la primera columna muestra la posición en que se encuentra el ítem de acuerdo a la medida obtenida, la segunda columna presenta el número de ítem y la tercera columna la medida obtenida o calibración del ítem.

Tabla 3.14: Ranking de Medidas de los Ítems

Corr.	Ítems	Medida
1	ítem_5	0.56
2	ítem_7	0.53
3	ítem_6	0.30
4	ítem_4	-0.01
5	ítem_3	-0.09
6	ítem_2	-0.48
7	ítem_1	-0.80

Donde:

Ítem_5: Propiedades de los exponentes en expresiones racionales

Ítem_7: Simplificación y producto de expresiones con números irracionales

Ítem_6: Simplificación de expresiones con números irracionales

Ítem_2: Propiedades de los números reales

Ítem_4: Simplificación de un número racional

Ítem_3: Transformación de un decimal a un racional

Ítem_1: Propiedades de los números racionales e irracionales

De esta manera se presenta el orden de medida que se obtiene al aplicar el modelo de Rasch a datos ordenados categóricamente.

3.4.3 RESULTADOS OBTENIDOS PARA EL TEST DE NUEVO INGRESO, UES 2007

En las secciones posteriores se hace un análisis acerca de los resultados obtenidos de la aplicación del Test de Nuevo Ingreso, UES 2007, clave 1, en el área de matemática. Inicialmente se contaba con una matriz de datos de 25 ítems o problemas de matemática y 6,333 aspirantes que les tocó resolver la clave 1, sin embargo, el programa BIGSTEPS que tenemos a nuestra disposición es una versión para estudiantes y es muy limitada, por tanto, se seleccionó una muestra de 500 aspirantes con los mismos 25 problemas, quedando así para su análisis. En cada sección que se desarrollará solo se hablará de "ítem" y si el lector quiere saber más información sobre cada ítem puede consultar el ANEXO IV.

3.4.3.1 Mapa de Distribución Conjunta para el Test de Nuevo Ingreso, UES 2007

De la misma manera que el mapa de ítems y estudiantes para el parcial I de Matemática I, se presenta el mapa de ítems y personas para los aspirantes a ingresar a la Universidad de El Salvador en el año 2007. El gráfico está diseñado igual que el mapa de ítems y estudiantes visto anteriormente, al lado izquierdo tenemos a las personas, donde los aspirantes más capaces se sitúan en la parte superior; al lado derecho tenemos la ubicación de los ítems, donde los ítems difíciles se sitúan en la parte superior.

Figura 3.20: Mapa de Ítems y Aspirantes

```
  ASPIRANTES          ÍTEMS
             <frequ>|<less>
 2                  +
                    |  item_30
                    |Q
                    |
                    |  item_34
                    |
                    |
 1                  +
                    |S item_26
                    |  item_39
                    |  item_32   item_35   item_47
                    |  item_31
                    |  item_48
                    |  item_37   item_49
                    |
 0                  +M item_45
                    |  item_44
                    |  item_33
                  . |
                  Q |  item_40
                 .# |  item_27   item_28   item_42   item_50
                    |  item_38   item_41   item_43
                 ## |  item_46
              .#### |S
-1                 S+
         .######### |
       .########### |
                    |
      .############ M|
                    |Q
        .########## |  item_29
                    |  item_36
-2                  +
                   S|
           .######  |
                    |
              .### Q|
-3                  +
                    |
                  . |
                    |
-4                . +
              <rare>|<more>
```

En el gráfico de la figura anterior, los estudiantes están representados por el símbolo numeral (#). La columna de la parte izquierda ubica la medida de habilidad de los estudiantes a lo largo de la variable. La columna de la parte derecha, localiza la calibración de los ítems a lo largo de la variable. Al buscar un margen equivalente de los ítems a lo largo de la variable (en el eje Y) se observa que existen zonas en donde faltan ítems (ítems con niveles de dificultad menores -2 lógitos e ítems con niveles de dificultad mayores a 2 lógitos), también se observa que hay grupos de ítems, por ejemplo, se observa que entre -1 y 1 lógitos hay tres grupos de ítems bien definidos, pero que no están midiendo a los sujetos. En este caso la escala o los ítems tienen defectos, ya que se esperaría que los ítems se distribuyeran uniformemente a lo largo de toda la escala (recta). De la misma manera, dentro de estos grupos existen ítems que se solapan, por ejemplo: los ítems 38, 41, 43; los ítems 27, 28, 42, 50; los ítems 37, 49; y los ítems 32, 35, 47, esto nos indica que estos ítems están midiendo la misma parte de la prueba de conocimiento o tienen el mismo nivel de dificultad. Por tanto, del primer grupo de ítems (ítems 38, 41 y 43) lo recomendable es eliminar o modificar dos de los tres ítems, en el segundo grupo de ítems (ítems 27, 28, 42 y 50) lo recomendable es eliminar o modificar tres de los cuatro ítems, de igual manera el tercer grupo de ítems (ítems 37 y 49) se recomienda eliminar o modificar uno de los dos ítems y del último grupo de ítems (ítems 32, 35 y 47) se recomienda eliminar o modificar dos de los tres ítems, con el fin de que midan niveles de dificultad diferentes en el test. También podemos ver la necesidad de agregar más ítems en la prueba con distintos niveles de dificultad, de tal manera de poder ir llenando los espacios vacíos que se pueden observar en la escala, además se necesitan ítems con niveles de dificultad mayores a 2 lógitos y menores a -2 lógitos, ya que se observa que no se esta midiendo toda la escala en que consiste la prueba de conocimientos.

Otro aspecto importante que nos muestra el mapa es el comportamiento de los estudiantes que se esperaría que estos se aproximen a una normal. De la figura 3.20 podemos observar que los aspirantes se aproximan a una normal, pero con niveles de conocimiento muy bajos. Además podemos observar que los aspirantes a ingresar a la Universidad de El Salvador, únicamente tienen conocimiento de un 38% de la prueba de conocimiento en la parte de matemática y para resolver la mayoría de ítems en que consiste dicha prueba es necesario conocer un 73% de los conceptos básicos de matemática. Es evidente la deficiencia que los aspirantes presentan al querer ingresar a la Universidad de El Salvador, por lo tanto, una forma de apalear esta situación es ofrecer cursos de nivelación con docentes capacitados antes de someterse a dicha prueba, donde se estudie todos los conceptos básicos de matemática y de esta manera minimizar las deserciones que ocurren cada año en las primeras materias de las carreras.

Finalmente, es recomendable que la UES defina claramente las competencias mínimas de los aspirantes a ingresar a la UES y con base a las competencias definidas elaborar un instrumento de evaluación bien calibrado.

3.4.3.2 Medida (Calibración) de los ítems del Test de Nuevo Ingreso, UES 2007

De la misma manera como se ha trabajado anteriormente para el primer parcial de la asignatura de Matemática I, se presentan los resultados para el Test de Nuevo Ingreso. Para describir la tabla 3.15, iniciaremos comentando que la primera columna muestra el número de ítem en el análisis, en la segunda columna se muestra la puntuación bruta correspondiente al parámetro (total de personas que han respondido al menos una opción en cada ítem), en la tercera columna tenemos la medida estimada de cada Ítem (nivel de dificultad), vale la pena mencionar que si el puntaje del ítem es extremo (resuelto correctamente o incorrectamente por todos los aspirantes), su medida es estimada como un MÁXIMO (puntuación perfecta) o un MÍNIMO (puntuación de cero). En la cuarta columna tenemos el error estándar de la estimación, inmediatamente la quinta y sexta columna nos muestran los estadísticos de ajuste al modelo (INFIT y OUTFIT). En la última columna se presenta un estimado de la discriminación del Ítem, llamada punto-biserial. Esta es una correlación, la cual mide la relación entre las puntuaciones de los ítems de cada aspirante y el total de la puntuación de los aspirantes en la prueba, el valor que debe aparecer ha de ser positivo, en caso contrario existe un claro desajuste de los ítems y menor importancia tendrá el valor del nivel obtenido (esta correlación es opcional).

Tabla 3.15: Medidas de los Ítems

Ítems	Puntuación Bruta	Medida	Error	INFIT		OUTFIT		Corr. Punt. Bis.
				MNSQ	ZSTD	MNSQ	ZSTD	
ítem_30	20	1.76	0.23	1.00	0.00	0.95	-0.20	0.00
ítem_34	29	1.37	0.19	1.01	0.10	1.05	0.20	-0.02
ítem_26	46	0.86	0.16	0.97	-0.20	0.81	-1.30	0.10
ítem_39	47	0.83	0.16	0.99	0.00	0.96	-0.20	0.02
ítem_32	51	0.74	0.15	1.00	0.00	1.05	0.40	-0.01
ítem_47	53	0.70	0.15	1.00	0.00	1.00	0.00	-0.01
ítem_35	55	0.65	0.14	0.99	-0.10	0.98	-0.10	0.04
ítem_31	60	0.55	0.14	1.00	0.00	1.06	0.50	0.00
ítem_48	63	0.50	0.14	1.01	0.10	0.98	-0.20	-0.01
ítem_37	66	0.44	0.13	1.03	0.30	1.07	0.60	-0.08
ítem_49	70	0.37	0.13	1.00	0.00	1.09	0.70	-0.01
ítem_45	96	-0.02	0.12	1.00	-0.10	1.00	-0.10	0.01
ítem_44	99	-0.06	0.11	1.02	0.30	1.09	1.00	-0.05
ítem_33	107	-0.16	0.11	1.00	0.00	0.96	-0.50	0.00
ítem_40	124	-0.37	0.11	1.01	0.20	1.00	0.00	-0.01
ítem_42	136	-0.50	0.10	1.01	0.10	1.01	0.10	-0.01
ítem_50	136	-0.50	0.10	1.00	0.00	0.99	-0.20	0.00
ítem_27	137	-0.51	0.10	1.00	-0.10	1.04	0.60	0.01
ítem_28	140	-0.54	0.10	0.96	-0.90	0.92	-1.40	0.08
ítem_43	141	-0.55	0.10	1.01	0.30	1.04	0.70	-0.03
ítem_41	142	-0.56	0.10	1.00	-0.10	1.03	0.50	0.01
ítem_38	150	-0.64	0.10	1.02	0.40	1.01	0.10	-0.03
ítem_46	155	-0.69	0.10	1.00	0.00	0.98	-0.30	0.00
ítem_29	277	-1.78	0.09	1.00	0.20	1.01	0.50	-0.01
ítem_36	288	-1.88	0.09	0.97	-1.20	0.96	-1.20	0.06
Media	108	0.00	0.13	1.00	0.00	1.00	0.00	
S.D.	66	0.85	0.03	0.02	0.30	0.06	0.60	

Del cuadro anterior podemos observar qué ítems son los más fáciles y cuales los más difíciles. Los ítems con medida más baja son los más fáciles del test y los ítems con medida más altas son los más difíciles del test. Podemos observar que los ítems más difíciles son el ítem 30 e ítem 34 los cuales tienen las mayores medidas en lógitos (ítem 30: 1.76 y el ítem 34: 1.37) y los ítems más fáciles son el ítem 29 e ítem 36 los cuales tienen las menores medidas en lógitos (ítem 29: -1.78 y el ítem 36:-1.88). También es de mencionar que aparte de los ítems descritos anteriormente, el resto de ítems se encuentran muy cercanos a 0 lógitos, lo cual reafirmamos que es necesario agregar más ítems con distintos niveles de dificultad al test de nuevo ingreso UES.

Ahora ¿qué tan bien ajustados se encuentran estos ítems al modelo de Rasch?, el ajuste lo observamos a través de los estadísticos de ajuste; para ello tenemos que recordar que disponemos de un conjunto de aceptación de los ítems: aceptaremos aquellos que se encuentren entre -2 y 2 lógitos. Al observar el ajuste de los ítems (INFIT: ajuste interno y OUTFIT: ajuste externo) podemos observar que tanto el ajuste interno como su ajuste externo no presentan desajuste en los datos, ya que los valores se encuentran dentro del intervalo de aceptación (-2 lógitos y 2 lógitos). Sin embargo, se puede observar la correlación punto biserial, donde se encuentran valores negativos muy pequeños, por tanto, se observa que no hay ítems que presenten desajustes significativos al modelo de Rasch.

3.4.3.3 Medida (habilidad) de los Aspirantes del Test de Nuevo Ingreso, UES 2007

En la tabla 3.16 se presenta la estimación de la habilidad de los estudiantes, esta tabla posee las mismas columnas que la correspondiente a la medida de los ítems, por esta razón nos limitaremos a presentarla. En esta tabla se presenta a los aspirantes que han obtenido las mayores medidas en el Test de Nuevo Ingreso UES.

Tabla 3.16: Aspirantes con las Medidas más Altas en el Test de Nuevo Ingreso

Aspirantes	Puntuación Bruta	Medida	Error	INFIT		OUTFIT		Corr. Punt. Bis.
				MNSQ	ZSTD	MNSQ	ZSTD	
22410	11	-0.27	0.43	1.24	1.50	1.33	1.50	0.09
17143	11	-0.27	0.43	0.89	-0.80	0.83	-0.90	0.48
3485	11	-0.27	0.43	1.04	0.30	1.05	0.20	0.27
3665	11	-0.27	0.43	1.20	1.30	1.23	1.10	0.10

La puntuación bruta de estos aspirantes nos indica que solo han podido resolver correctamente 11 ítems de los 25 ítems que consiste el Test de Nuevo Ingreso en la parte de Matemática (menos del 50%), por tanto se observa la deficiencia presentada por dichos aspirantes. La medida de estos aspirantes nos indican que tienen una probabilidad del 43% de resolver correctamente el test (esta probabilidad la podemos obtener mediante las expresiones vistas en la sección 2.2.4.1 del capítulo II). Por lo tanto, podemos esperar que un aspirante con un nivel de habilidad -0.27 lógitos solo pueda resolver los ítems 40, 42, 50, 27, 28, 43, 41, 38, 46, 29 y 36.

La tabla 3.17 nos muestra el grupo de aspirantes que han obtenido las segundas medidas más altas, ya que son las que les siguen a los aspirantes presentados en la tabla anterior.

Tabla 3.17: Aspirantes con las Segundas Medidas más Altas en el Test de Nuevo Ingreso

Aspirantes	Puntuación Bruta	Medida	Error	INFIT		OUTFIT		Corr. Punt. Bis.
				MNSQ	ZSTD	MNSQ	ZSTD	
20193	10	-0.46	0.44	1.09	0.60	1.09	0.40	0.14
25	10	-0.46	0.44	1.01	0.10	0.98	-0.10	0.37
3101	10	-0.46	0.44	0.86	-0.90	0.79	-1.00	0.52
3249	10	-0.46	0.44	1.25	1.40	1.27	1.20	0.07
3300	10	-0.46	0.44	1.28	1.60	1.30	1.30	0.03
3322	10	-0.46	0.44	0.80	-1.40	0.72	-1.40	0.58
3570	10	-0.46	0.44	0.97	-0.20	0.95	-0.20	0.41
3598	10	-0.46	0.44	0.82	-1.20	0.76	-1.20	0.56
3660	10	-0.46	0.44	0.81	-1.30	0.75	-1.30	0.57

Del cuadro anterior se observa el decrecimiento de la puntuación bruta y con ella la medida de los aspirantes. De los 25 ítems, estos aspirantes solo han podido resolver 10 ítems correctamente. El error nos podría servir para saber en que intervalo se encuentra el verdadero parámetro de dichos sujetos, por ejemplo un intervalo de confianza del 95% para estos sujetos es de:

$$-0.46-(1.96)(0.44)\leq\theta\leq-0.46+(1.96)(0.44)$$
$$-1.3224\leq\theta\leq 0.4024$$

Lo que nos indica que de 100 muestras que tomemos de forma repetitiva, 95 de ellas tendrán el verdadero valor del parámetro (nivel de habilidad) entre -1.3224 y 0.4024. Por último, podemos mencionar que la probabilidad de contestar correctamente el test por estos estudiantes es aproximadamente el 39%.

Existe una gran mayoría de estudiantes que se encuentran con niveles de medida alrededor entre -0.66 y -2.74 lógitos, los cuales pueden verse en la sección 3.4.3.6. La probabilidad de que estos aspirantes puedan resolver correctamente la parte de Matemática en el Test de Nuevo Ingreso es menor al 34%.

Por último, tenemos los aspirantes que han obtenido medidas por debajo de -3.00 lógitos, esto nos indica que estos aspirantes han obtenido los peores resultados y en la mayoría de los ítems del test solo pudieron resolver 1 ítem de los 25 que comprende el Test de Nuevo Ingreso en la parte de Matemática, la tabla se presenta a continuación:

Tabla 3.18: Aspirantes con las Medidas más Bajas en el Test de Nuevo Ingreso

Aspirantes	Puntuación Bruta	Medida	Error	INFIT		OUTFIT		Corr. Punt. Bis.
				MNSQ	ZSTD	MNSQ	ZSTD	
22479	1	-3.52	1.04	1.12	0.10	1.99	0.60	-0.12
15481	1	-3.52	1.04	1.06	0.10	0.82	-0.10	0.10
13078	1	-3.52	1.04	1.10	0.10	1.31	0.20	-0.03
10	1	-3.52	1.04	0.84	-0.20	0.26	-0.80	0.52
130	1	-3.52	1.04	0.84	-0.20	0.26	-0.80	0.52
134	1	-3.52	1.04	1.06	0.10	0.85	-0.10	0.09
3629	1	-3.52	1.04	1.06	0.10	0.81	-0.10	0.10
5171	0	-4.25	1.44	MÍNIMA MEDIDA ESTIMADA				

El programa BIGSTEPS no puede obtener una estimación de los estadísticos de ajuste para el aspirante 5171 ya que ha sido el único sujeto que no pudo resolver correctamente ningún ítem del test como se puede observar en el escalograma de Guttman (ver ANEXO V), por esta razón el programa le ha asignado una medida de -4.25 lógitos, ya que presenta un patrón sistemático en todos sus resultados en los ítems. En la tabla anterior se observan las medidas más bajas obtenidas por los aspirantes a ingresar a la UES. Por tanto, se espera que un sujeto con un nivel de habilidad menor a -3.52 lógitos tenga una probabilidad de resolver correctamente el test de nuevo ingreso menor al 1% o equivalentemente no pueda resolver ningún ítem del Test de Nuevo Ingreso en la parte de Matemática.

La siguiente tabla muestra un resumen de los estadísticos de ajuste interno (INFIT) y ajuste externo (OUTFIT), esta tabla solo muestra los estudiantes que tienen los mayores o menores estadísticos de ajuste. La siguiente tabla esta ordenada por los estadísticos de ajuste y su estructura es similar a las tablas anteriores.

Tabla 3.19: Estadísticos de Ajuste para los Aspirantes

Aspirante	Puntuación Bruta	Medida	Error	INFIT		OUTFIT		Corr. Punt. Bis.
				MNSQ	ZSTD	MNSQ	ZSTD	
3642	3	-2.26	0.64	1.21	0.40	2.79	1.80	A-.04
13072	3	-2.26	0.64	1.29	0.60	2.63	1.70	B-.19
3319	5	-1.58	0.53	1.24	0.70	2.61	2.50	C-.01
3588	5	-1.58	0.53	1.56	1.60	2.54	2.40	D-.43
22733	4	-1.89	0.57	1.42	1.00	2.45	1.90	E-.31
3395	3	-2.26	0.64	1.27	0.60	2.43	1.50	F-.15
16267	2	-2.74	0.76	1.25	0.40	2.42	1.20	G-.23
45	4	-1.89	0.57	1.28	0.70	2.41	1.90	H-.09
3631	3	-2.26	0.64	1.24	0.50	2.20	1.30	I-.09
3508	4	-1.89	0.57	1.20	0.50	2.17	1.60	J .03
150	3	-2.26	0.64	1.21	0.50	2.09	1.20	K-.04
101	4	-1.89	0.57	1.29	0.70	2.08	1.50	L-.10
22479	1	-3.52	1.04	1.12	0.10	1.99	0.60	M-.12
12183	7	-1.08	0.48	1.42	1.70	1.96	2.30	N-.18
2411	7	-1.08	0.48	1.24	1.00	1.92	2.20	O .06
3238	5	-1.58	0.53	1.10	0.30	1.90	1.60	P .19
21688	2	-2.74	0.76	1.19	0.30	1.82	0.70	Q-.09
3606	6	-1.32	0.50	1.40	1.40	1.81	1.70	R-.17
5727	5	-1.58	0.53	1.18	0.60	1.81	1.40	S .07
3534	8	-0.86	0.46	1.31	1.50	1.75	2.20	T-.01
22062	3	-2.26	0.64	1.27	0.60	1.72	0.90	U-.13
42	5	-1.58	0.53	1.01	0.10	1.69	1.30	V .30
17170	4	-1.89	0.57	1.33	0.80	1.68	1.00	W-.15
14454	8	-0.86	0.46	1.26	1.20	1.66	2.00	X .06
22428	6	-1.32	0.50	1.22	0.80	1.66	1.50	Y .05
19716	6	-1.32	0.50	1.11	0.40	1.66	1.40	Z .20
3143	6	-1.32	0.50	0.76	-1.00	0.61	-1.20	z .65
3001	4	-1.89	0.57	0.75	-0.80	0.57	-0.90	y .65
19607	4	-1.89	0.57	0.75	-0.80	0.55	-1.00	x .66
19621	4	-1.89	0.57	0.75	-0.80	0.55	-1.00	w .66
17182	7	-1.08	0.48	0.74	-1.30	0.61	-1.40	v .66
3272	5	-1.58	0.53	0.74	-0.90	0.57	-1.10	u .67
22485	8	-0.86	0.46	0.74	-1.50	0.63	-1.50	t .66
2988	4	-1.89	0.57	0.73	-0.80	0.52	-1.10	s .68
73	4	-1.89	0.57	0.73	-0.80	0.52	-1.10	r .68
3289	6	-1.32	0.50	0.73	-1.10	0.57	-1.30	q .69
18	5	-1.58	0.53	0.73	-1.00	0.55	-1.20	p .69
2966	6	-1.32	0.50	0.73	-1.20	0.57	-1.30	o .69
17168	5	-1.58	0.53	0.73	-1.00	0.54	-1.20	n .69
3595	4	-1.89	0.57	0.73	-0.80	0.51	-1.10	m .69
11377	7	-1.08	0.48	0.72	-1.40	0.59	-1.50	l .69
3402	8	-0.86	0.46	0.72	-1.60	0.61	-1.60	k .69
3647	4	-1.89	0.57	0.72	-0.90	0.50	-1.10	j .70
168	4	-1.89	0.57	0.71	-0.90	0.49	-1.20	i .71
3527	4	-1.89	0.57	0.71	-0.90	0.49	-1.20	h .71
3569	3	-2.26	0.64	0.71	-0.70	0.45	-1.00	g .70
19604	3	-2.26	0.64	0.71	-0.70	0.45	-1.00	f .70
3590	3	-2.26	0.64	0.71	-0.70	0.45	-1.00	e .70
3586	3	-2.26	0.64	0.71	-0.70	0.45	-1.00	d .70
78	4	-1.89	0.57	0.70	-0.90	0.48	-1.20	c .72
3280	3	-2.26	0.64	0.68	-0.80	0.41	-1.10	b .74
17164	2	-2.74	0.76	0.66	-0.70	0.26	-1.20	a .78
Media	5	-1.55	0.54	1.00	0.00	1.00	-0.10	
S.D.	2	0.58	0.10	0.16	0.60	0.37	0.70	

En la tabla anterior los valores que aparecen de color rojo nos indican desajuste en los aspirantes, y como se observa, solo hay aspirantes que han tenido desajuste externo. Podemos observar que los aspirantes 3319, 3588, 12183, 2411, 3534, 14454 presentan valores mayores o iguales a 2 lógitos en su ajuste externo (OUTFIT), lo que indica que estos aspirantes presentan demasiada estocasticidad en sus respuestas (patrón ruidoso), lo que de alguna manera podríamos decir que dichos estudiantes han resuelto el Test de Nuevo Ingreso en la parte de Matemática sin saber lo que hacen (lo han resuelto al azar).

3.4.3.4 Desajustes de los Aspirantes del Test de Nuevo Ingreso, UES 2007

Las medidas de los sujetos se complementan con el análisis detallado del origen del desajuste entre el valor real y el modelo. Ello permite considerar las acciones necesarias para corregir las causas del desajuste, en caso que sea necesario. Así la siguiente tabla nos muestra los aspirantes que según el modelo de Rasch presentan desajuste, y estos resultados nos ayudarán a darle mayor veracidad al análisis que anteriormente se ha hecho con los aspirantes respecto a sus estadísticos FIT (INFIT y OUTFIT).

Figura 3.21: Desajustes de los Aspirantes

```
TABLE OF POORLY FITTING ASPIRAS   (PROBLES IN ENTRY ORDER)
NÚMERO - NOMBRE -- POSICIÓN ------ MEDIDA - INFIT (ZSTD) OUTFIT

  360 3319                         -1.58      .7   C    2.5
       RESPONSE:    1:  0 0 0 0 1  1 0 0 1 0  1 0 0 0 0  0 0 0 0 0  0 0 0 0 1
     Z-RESIDUAL:                8  2     4

  450 3588                         -1.58     1.6   D    2.4
       RESPONSE:    1:  1 0 0 0 0  0 1 0 1 0  0 1 0 0 0  0 0 0 0 0  0 0 1 0 0
     Z-RESIDUAL:        3            3   4       2                      2

   32 12183                        -1.08     1.7   N    2.3
       RESPONSE:    1:  1 0 0 0 1  0 0 0 0 1  0 0 1 0 0  1 0 1 0 0  0 0 0 1 0
     Z-RESIDUAL:        2       4          2                              2

  273 2411                         -1.08     1.0   O    2.2
       RESPONSE:    1:  0 1 0 1 1  1 0 0 1 0  0 0 0 0 1  0 0 0 0 0  1 0 0 0 0
     Z-RESIDUAL:                4  2     3

  437 3534                          -.86     1.5   T    2.2
       RESPONSE:    1:  0 1 0 1 1  0 1 0 0 1  0 0 1 1 1  0 0 0 0 0  0 0 0 0 0
     Z-RESIDUAL:                3    2     2         2
```

En la tabla de la figura anterior podemos observar: el número que el programa le ha asignado, el nombre del aspirante (en nuestro caso es el número del aspirante que realizó la prueba), posición, medida del estudiante, y los estadísticos de ajuste. Así los desajustes de los estudiantes recogidos en la sección 3.4.3.3 se complementa con la expresión de qué ítem es el que genera el desajuste. Los valores residuales positivos indican que han puntuado al ítem por encima del valor que esperaba el modelo. Los valores residuales negativos indican que han sido puntuados más bajos de lo que el modelo esperaba.

El aspirante 3319 ha puntuado cinco unidades más en el ítem 30, dos unidades más en el ítem 31 y cuatro unidades más en el ítem 34 de lo que esperaba el modelo; podemos ver también los valores del INFIT y OUTFIT, el valor del INFIT es de 0.7 lógitos y el valor de OUTFIT es de 2.5 lógitos, por tanto se observa el desajuste externo de dicho aspirante.

Asimismo el aspirante 3588 ha puntuado tres unidades más en el ítem 26, tres unidades más en el ítem 32, cuatro unidades más en el ítem 34, dos unidades más en el ítem 37 y dos unidades más en el ítem 48 de lo que el modelo esperaba. El valor del INFIT es de 1.6 lógitos y el valor de OUTFIT es de 2.4 lógitos, por tanto se vuelve a observar el desajuste externo de dicho aspirante.

En el aspirante 12183 podemos ver que todos los residuales son positivos, en el ítem 26, 35 y 49 el aspirante ha puntuado dos unidades por encima de lo que el modelo esperaba, de igual manera en el ítem 30 el aspirante ha puntuado cinco unidades más de lo que el modelo esperaba. De igual manera podemos observar los estadísticos de ajuste, donde el valor del ajuste externo (OUTFIT: 2.3) nos indica desajuste del aspirante en este ítem.

Podemos observar el aspirante 2411 el cual ha puntuado cuatro unidades más en el ítem 30 de lo que el modelo esperaba, en el ítem 31 ha puntuado dos unidades más de lo que el modelo esperaba y de igual manera en el ítem 34 el aspirante ha puntuado tres unidades más de lo que el modelo esperaba. El estadístico de ajuste externo (OUTFIT: 2.2) nos indica un desajuste externo del aspirante.

De forma similar podemos observar los residuales del aspirante 3534, ya que ha puntuado dos unidades más en el ítem 32, ítem 35 y el ítem 39 de lo que el modelo esperaba y ha puntuado tres unidades más en el ítem 30 de lo que el modelo esperaba. Su estadístico de ajuste externo (OUTFIT: 2.2) nos indica un desajuste de este aspirante en dicho ítem.

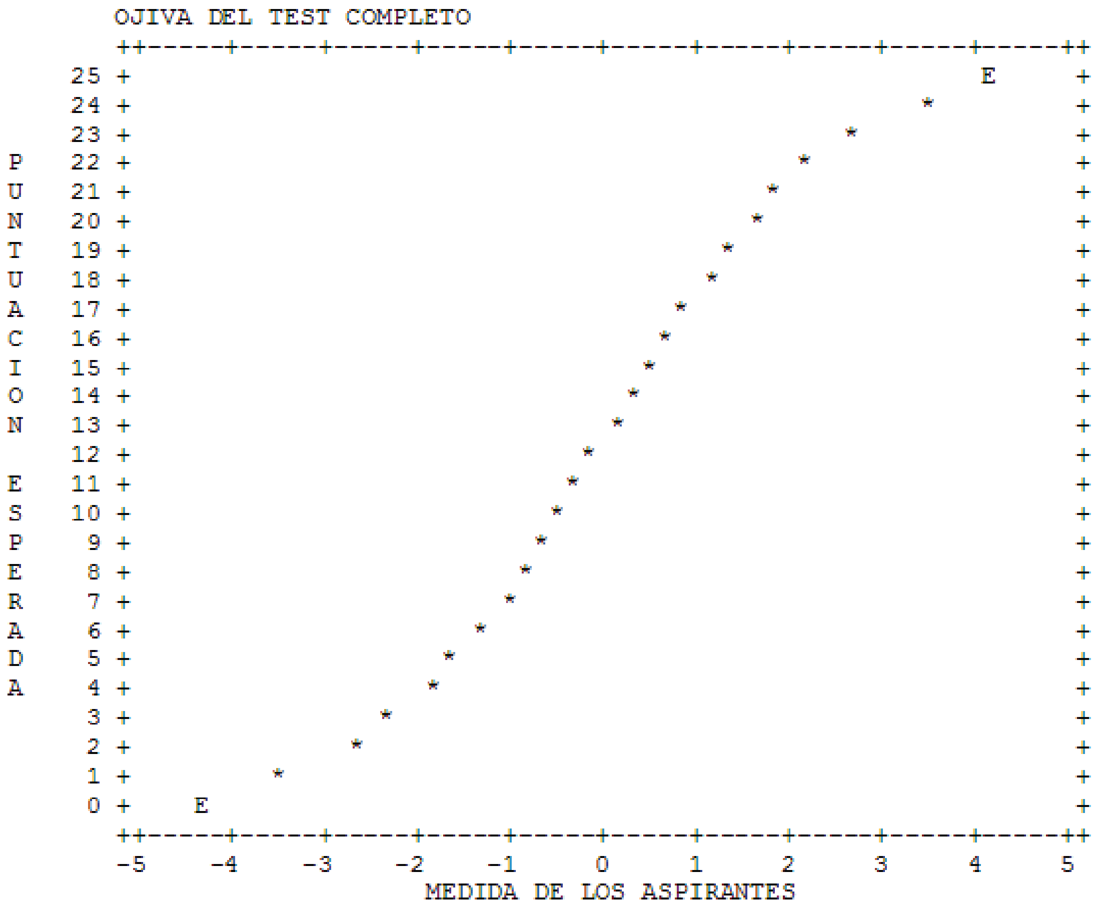
Ante esta situación podemos decir que los estudiantes presentados en esta tabla muestran un desajuste que afectan la obtención de las medidas de los estudiantes. Con este análisis, se determina que aspirantes están afectando al diseño del instrumento de medida, por lo cual se aconseja eliminar estos sujetos de la base.

En conclusión tenemos 5 aspirantes (3319, 3588, 12183, 2411 y 3534) que presentan desajustes con el modelo de Rasch y en nuestro caso no hay ítems que estén afectando la obtención de las medidas de estudiantes e ítems. Podemos decir que estos aspirantes han resuelto o respondido los ítems 26, 30, 31, 32, 34, 35, 37, 39, 48 y 49 de una manera al azar.

3.4.3.5 Gráficos del Test Completo para los Aspirantes del Test de Nuevo Ingreso, UES 2007

Esta gráfica resulta de dibujar la medida del estudiante contra las puntuaciones que se pueden obtener en el test, la cual está asociada a la dificultad del ítem. En el eje de las abscisas podemos observar la medida de los estudiantes y en el eje de las ordenadas las puntuaciones que mide el instrumento. La letra *"E"* que se observa en los extremos de la curva, nos indica la puntuación mínima y la puntuación máxima respectivamente.

Figura 3.22: Gráfica de la Curva del Test Completo



Observemos que un estudiante con medida $\theta = 0$ puede obtener 11 puntos de un total de 25 puntos, por lo que podría resolver correctamente el 44% del Test de Nuevo Ingreso UES. Un estudiante con un nivel de rasgo $\theta = -1$, obtendría una puntuación de 7 unidades, lo que significa que solo podría resolver correctamente un 28% del test completo. Observemos que a menor medida que un estudiante tenga, menor será la puntuación que obtendrá en el test; caso contrario sucede en un estudiante con mayor medida, ya que tendrá una mayor puntuación en el test. Tal es el caso de un estudiante con un nivel de rasgo 3 lógitos, ya que obtendrá una puntuación de 23 puntos, lo que equivale a resolver correctamente el 92% del Test de Nuevo Ingreso en la parte de Matemática.

En la siguiente tabla se muestra los posibles puntajes que pueden ser obtenidos en el test (la mayor puntuación de 25 puntos equivale al 100% de la prueba), la medida del estudiante y su error de estimación.

Tabla 3.20: Medidas del Test Completo

Puntuación	Medida	S.E.	Puntuación	Medida	S.E.	Puntuación	Medida	S.E.
0	-4.25E	1.44	9	-0.66	0.45	18	1.09	0.47
1	-3.52	1.04	10	-0.46	0.44	19	1.32	0.50
2	-2.74	0.76	11	-0.27	0.43	20	1.58	0.53
3	-2.26	0.64	12	-0.09	0.43	21	1.88	0.57
4	-1.89	0.57	13	0.10	0.43	22	2.25	0.64
5	-1.58	0.53	14	0.29	0.43	23	2.72	0.76
6	-1.32	0.50	15	0.48	0.44	24	3.49	1.04
7	-1.08	0.48	16	0.67	0.45	25	4.22E	1.44
8	-0.86	0.46	17	0.87	0.46			

La tabla anterior muestra los puntajes que pueden ser obtenidos en el test junto con la medida del aspirante y su error de estimación. Podemos observar que el error de estimación en los extremos es más grande que en el centro, como se mencionó en el capítulo II, solo con datos que se encuentran en el centro se puede obtener un error de estimación más preciso.

Además de la curva anterior, el programa BIGSTEPS proporciona también, curvas referidas a los ítems que comprende el test. La siguiente figura responde a la pregunta ¿Cuál es el promedio evaluado que nosotros esperamos observar para una persona de medida en particular?. En este gráfico se observa en el extremo izquierdo la categoría más baja, 0, y en el extremo derecho de la figura podemos observar la categoría más alta, la cual sería 1. Los dos puntos que se observan en el gráfico nos indican la zona que divide el error del acierto. Podemos observar también en el lado derecho de la figura, los ítems que comprende el test, se esperaría que entre cada ítem no hubieran “huecos” para decir que se está midiendo bien la escala del Test de Nuevo Ingreso en la parte de Matemática.

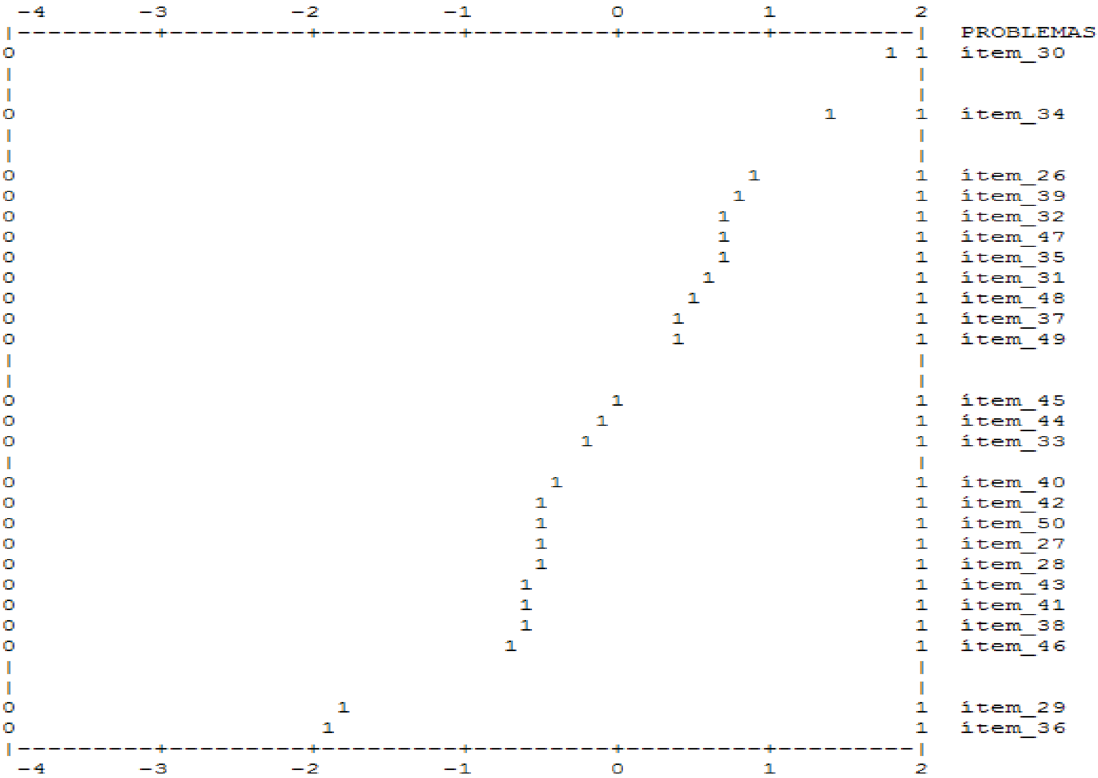
Figura 3.23: Promedio Evaluado que se Esperaría de los Aspirantes

```
EXPECTED SCORE: MEAN  (":" INDICATES HALF-SCORE POINT)
 -4      -3        -2        -1         0         1         2
|---------+---------+---------+---------+---------+---------|  PROBLEMAS
0                                              0          : 1  item_30
|                                                           |
|                                                           |
0                                          0          :     1  item_34
|                                                           |
|                                                           |
0                                     0          :          1  item_26
0                                    0          :          11  item_39
0                                   0          :          1 1  item_32
0                                   0          :          1 1  item_47
0                                   0          :          1 1  item_35
0                                  0          :          1  1  item_31
0                                 0          :          1   1  item_48
0                                0          :          1    1  item_37
0                                0          :          1    1  item_49
|                                                           |
|                                                           |
0                            0          :          1        1  item_45
0                           0          :          1         1  item_44
0                          0          :          1          1  item_33
|                                                           |
0                        0          :          1            1  item_40
0                       0          :          1             1  item_42
0                       0          :          1             1  item_50
0                       0          :          1             1  item_27
0                       0          :          1             1  item_28
0                       0         :          1              1  item_43
0                      0          :          1              1  item_41
0                      0          :          1              1  item_38
0                     0          :          1               1  item_46
|                                                           |
|                                                           |
0          0          :          1                          1  item_29
0         0          :          1                           1  item_36
|---------+---------+---------+---------+---------+---------|
 -4      -3        -2        -1         0         1         2
```

En este gráfico podemos observar los espacios o "huecos" que existe entre los ítems (extremo derecho), como ya se ha comentado en el mapa de distribución conjunta no se esta midiendo bien la escala (los ítems que presentan espacios en este gráfico son los mismos que presentaban espacios en el mapa de distribución conjunta para el Test de Nuevo Ingreso). Podemos observar que para un estudiante con medida $\theta = 1$ se espera que en promedio resuelva correctamente el ítem 44 y los que tienen un nivel de dificultad menor que dicho ítem. Para un estudiante con $\theta = 2$, se espera que en promedio resuelva todos los ítems del Test de Nuevo Ingreso en la parte de Matemática. Un estudiante con un nivel de rasgo de -1.5 se espera que en promedio falle en el ítem 33.

Otro gráfico interesante es el que presenta la respuesta más probable que se puede esperar de un sujeto con una medida particular. La estructura del gráfico es similar al mostrado anteriormente.

Figura 3.23: Respuesta más Probable de los Aspirantes



Esta gráfica permite estimar la probabilidad de respuesta de un sujeto de medida θ ante la prueba aplicada a los aspirantes. Un sujeto con un nivel de rasgo 0.8 lógitos se tiene una probabilidad de resolver el ítem 39, esto quiere decir que el sujeto debería resolver correctamente los ítems que están por debajo del ítem 39. Si no los resuelve correctamente se tratará de una falla inesperada. Por el contrario, no debería poder resolver correctamente las preguntas por arriba del ítem 39 a riesgo de decir que se trata de un acierto inesperado. Un sujeto con un nivel de rasgo de -1.8 lógitos debería resolver correctamente el ítem 29, esto quiere decir que el sujeto debería resolver correctamente el ítem 36. Si no los resuelve correctamente se tratará de una falla inesperada. Por el contrario, no debería poder resolver correctamente las preguntas por arriba del ítem 29 a riesgo de decir que se trata de un acierto inesperado.

3.4.3.6 Ranking de las Medidas de los Aspirantes e Ítems del Test de Nuevo Ingreso, UES 2007

En la siguiente tabla se presenta el resultado final de las medidas de los aspirantes, eliminando a aquellos que han obtenido puntuaciones extremas (todos los ítems resueltos correctamente e incorrectamente). En la primera columna se presenta el orden ascendente (posición que ocupa en el Ranking el aspirante), la segunda columna muestra el número del aspirante y la tercera columna la medida obtenida por el aspirante.

Tabla 3.21: Ranking de Medidas de los Aspirantes

Corr.	Aspirantes	Medida	Corr.	Aspirantes	Medida	Corr.	Aspirantes	Medida	Corr.	Aspirantes	Medida	Corr.	Aspirantes	Medida
1	22410	-0.27	51	7	-0.86	101	112	-1.08	151	19721	-1.32	201	3294	-1.32
2	17143	-0.27	52	41	-0.86	102	117	-1.08	152	19716	-1.32	202	3287	-1.32
3	3485	-0.27	53	142	-0.86	103	126	-1.08	153	16109	-1.32	203	3289	-1.32
4	3665	-0.27	54	170	-0.86	104	2411	-1.08	154	15960	-1.32	204	3308	-1.32
5	20193	-0.46	55	3274	-0.86	105	2942	-1.08	155	18718	-1.32	205	3298	-1.32
6	25	-0.46	56	3369	-0.86	106	2953	-1.08	156	13067	-1.32	206	3305	-1.32
7	3101	-0.46	57	3402	-0.86	107	2963	-1.08	157	13080	-1.32	207	3325	-1.32
8	3249	-0.46	58	3393	-0.86	108	2968	-1.08	158	13086	-1.32	208	3326	-1.32
9	3300	-0.46	59	3515	-0.86	109	3104	-1.08	159	13105	-1.32	209	3324	-1.32
10	3322	-0.46	60	3540	-0.86	110	3134	-1.08	160	13097	-1.32	210	3343	-1.32
11	3570	-0.46	61	3551	-0.86	111	3145	-1.08	161	19590	-1.32	211	3345	-1.32
12	3598	-0.46	62	3534	-0.86	112	3109	-1.08	162	13102	-1.32	212	3385	-1.32
13	3660	-0.46	63	3557	-0.86	113	3178	-1.08	163	19581	-1.32	213	3388	-1.32
14	19289	-0.66	64	3577	-0.86	114	3166	-1.08	164	19608	-1.32	214	3398	-1.32
15	6572	-0.66	65	3654	-0.86	115	3263	-1.08	165	5209	-1.32	215	3381	-1.32
16	10763	-0.66	66	3651	-0.86	116	3276	-1.08	166	5200	-1.32	216	3472	-1.32
17	5212	-0.66	67	12701	-1.08	117	3284	-1.08	167	5196	-1.32	217	3470	-1.32
18	5206	-0.66	68	18282	-1.08	118	3312	-1.08	168	5172	-1.32	218	3499	-1.32
19	6230	-0.66	69	18128	-1.08	119	3296	-1.08	169	5174	-1.32	219	3502	-1.32
20	6222	-0.66	70	5222	-1.08	120	3349	-1.08	170	18887	-1.32	220	3533	-1.32
21	17158	-0.66	71	12183	-1.08	121	3362	-1.08	171	6233	-1.32	221	3558	-1.32
22	44	-0.66	72	12342	-1.08	122	3370	-1.08	172	22037	-1.32	222	3554	-1.32
23	3169	-0.66	73	3556	-1.08	123	3363	-1.08	173	17141	-1.32	223	3579	-1.32
24	3261	-0.66	74	20722	-1.08	124	3468	-1.08	174	17145	-1.32	224	3581	-1.32
25	3302	-0.66	75	14074	-1.08	125	3463	-1.08	175	17180	-1.32	225	3606	-1.32
26	3307	-0.66	76	10591	-1.08	126	3489	-1.08	176	17193	-1.32	226	3632	-1.32
27	3318	-0.66	77	13070	-1.08	127	3486	-1.08	177	17187	-1.32	227	3638	-1.32
28	3366	-0.66	78	13084	-1.08	128	3511	-1.08	178	17218	-1.32	228	3640	-1.32
29	3344	-0.66	79	13091	-1.08	129	3524	-1.08	179	5	-1.32	229	3641	-1.32
30	22426	-0.86	80	19588	-1.08	130	3532	-1.08	180	24	-1.32	230	3626	-1.32
31	18908	-0.86	81	19623	-1.08	131	3546	-1.08	181	43	-1.32	231	3655	-1.32
32	18738	-0.86	82	5173	-1.08	132	3582	-1.08	182	74	-1.32	232	3666	-1.32
33	14454	-0.86	83	5176	-1.08	133	3620	-1.08	183	85	-1.32	233	3652	-1.32
34	19541	-0.86	84	21041	-1.08	134	3637	-1.08	184	100	-1.32	234	5872	-1.58
35	16619	-0.86	85	22401	-1.08	135	3623	-1.08	185	132	-1.32	235	5877	-1.58
36	10794	-0.86	86	11117	-1.08	136	3646	-1.08	186	133	-1.32	236	5841	-1.58
37	10293	-0.86	87	16329	-1.08	137	3650	-1.08	187	155	-1.32	237	5727	-1.58
38	13073	-0.86	88	11377	-1.08	138	3675	-1.08	188	167	-1.32	238	17695	-1.58
39	13074	-0.86	89	6228	-1.08	139	3678	-1.08	189	2951	-1.32	239	19416	-1.58
40	13075	-0.86	90	17146	-1.08	140	3661	-1.08	190	2966	-1.32	240	4762	-1.58
41	19620	-0.86	91	17176	-1.08	141	22428	-1.32	191	2993	-1.32	241	21838	-1.58
42	10876	-0.86	92	17182	-1.08	142	20710	-1.32	192	3148	-1.32	242	13089	-1.58
43	22032	-0.86	93	17199	-1.08	143	11223	-1.32	193	3162	-1.32	243	3535	-1.58
44	17153	-0.86	94	17204	-1.08	144	20819	-1.32	194	3132	-1.32	244	18591	-1.58
45	17156	-0.86	95	9	-1.08	145	6106	-1.32	195	3143	-1.32	245	18621	-1.58
46	17233	-0.86	96	71	-1.08	146	3615	-1.32	196	3186	-1.32	246	22451	-1.58
47	11310	-0.86	97	84	-1.08	147	14500	-1.32	197	3174	-1.32	247	16415	-1.58
48	22485	-0.86	98	89	-1.08	148	10771	-1.32	198	3265	-1.32	248	10374	-1.58
49	13447	-0.86	99	91	-1.08	149	20351	-1.32	199	3247	-1.32	249	17904	-1.58
50	2	-0.86	100	99	-1.08	150	4513	-1.32	200	3270	-1.32	250	17997	-1.58

Corr.	Aspirantes	Medida
251	18341	-1.58
252	13079	-1.58
253	19583	-1.58
254	19610	-1.58
255	5203	-1.58
256	5177	-1.58
257	5165	-1.58
258	5169	-1.58
259	19056	-1.58
260	21081	-1.58
261	11169	-1.58
262	6225	-1.58
263	6211	-1.58
264	6213	-1.58
265	6214	-1.58
266	6218	-1.58
267	22057	-1.58
268	17144	-1.58
269	17161	-1.58
270	17173	-1.58
271	17174	-1.58
272	17168	-1.58
273	17194	-1.58
274	17195	-1.58
275	17201	-1.58
276	17219	-1.58
277	17225	-1.58
278	17205	-1.58
279	16235	-1.58
280	1	-1.58
281	8	-1.58
282	18	-1.58
283	33	-1.58
284	42	-1.58
285	111	-1.58
286	115	-1.58
287	123	-1.58
288	129	-1.58
289	157	-1.58
290	158	-1.58
291	159	-1.58
292	17042	-1.58
293	2960	-1.58
294	2957	-1.58
295	2958	-1.58
296	2991	-1.58
297	2977	-1.58
298	2998	-1.58
299	3003	-1.58
300	3180	-1.58
301	3231	-1.58
302	3181	-1.58
303	3251	-1.58
304	3235	-1.58
305	3238	-1.58
306	3272	-1.58
307	3278	-1.58
308	3316	-1.58
309	3317	-1.58
310	3319	-1.58
311	3314	-1.58
312	3386	-1.58
313	3456	-1.58
314	3465	-1.58
315	3452	-1.58
316	3475	-1.58
317	3476	-1.58
318	3483	-1.58
319	3512	-1.58
320	3516	-1.58
321	3518	-1.58
322	3531	-1.58
323	3548	-1.58
324	3566	-1.58
325	3588	-1.58
326	3601	-1.58
327	3610	-1.58
328	3593	-1.58
329	3644	-1.58
330	3663	-1.58
331	3664	-1.58
332	3676	-1.58
333	5443	-1.89
334	12157	-1.89
335	15143	-1.89
336	22419	-1.89
337	19584	-1.89
338	19589	-1.89
339	19605	-1.89
340	19607	-1.89
341	19619	-1.89
342	19621	-1.89
343	5210	-1.89
344	5211	-1.89
345	5182	-1.89
346	5179	-1.89
347	5164	-1.89
348	6232	-1.89
349	22035	-1.89
350	17152	-1.89
351	17149	-1.89
352	17155	-1.89
353	17170	-1.89
354	17175	-1.89
355	17183	-1.89
356	17189	-1.89
357	17202	-1.89
358	17231	-1.89
359	17238	-1.89
360	6115	-1.89
361	22733	-1.89
362	45	-1.89
363	47	-1.89
364	50	-1.89
365	72	-1.89
366	73	-1.89
367	76	-1.89
368	78	-1.89
369	88	-1.89
370	94	-1.89
371	101	-1.89
372	116	-1.89
373	137	-1.89
374	168	-1.89
375	20204	-1.89
376	2961	-1.89
377	2980	-1.89
378	2988	-1.89
379	3001	-1.89
380	2985	-1.89
381	3106	-1.89
382	2996	-1.89
383	3141	-1.89
384	3170	-1.89
385	3268	-1.89
386	3259	-1.89
387	3309	-1.89
388	3310	-1.89
389	3328	-1.89
390	3315	-1.89
391	3321	-1.89
392	3396	-1.89
393	3464	-1.89
394	3474	-1.89
395	3477	-1.89
396	3498	-1.89
397	3488	-1.89
398	3507	-1.89
399	3508	-1.89
400	3494	-1.89
401	3521	-1.89
402	3527	-1.89
403	3513	-1.89
404	3568	-1.89
405	3587	-1.89
406	3595	-1.89
407	3617	-1.89
408	3647	-1.89
409	3648	-1.89
410	3667	-1.89
411	3653	-1.89
412	3672	-1.89
413	3674	-1.89
414	20745	-2.26
415	4295	-2.26
416	13066	-2.26
417	13072	-2.26
418	13104	-2.26
419	19604	-2.26
420	19606	-2.26
421	19624	-2.26
422	19625	-2.26
423	5208	-2.26
424	5178	-2.26
425	15728	-2.26
426	6212	-2.26
427	22060	-2.26
428	22062	-2.26
429	17177	-2.26
430	17228	-2.26
431	14	-2.26
432	75	-2.26
433	83	-2.26
434	97	-2.26
435	108	-2.26
436	140	-2.26
437	150	-2.26
438	2956	-2.26
439	2973	-2.26
440	2989	-2.26
441	2986	-2.26
442	3136	-2.26
443	3102	-2.26
444	3167	-2.26
445	3242	-2.26
446	3245	-2.26
447	3253	-2.26
448	3282	-2.26
449	3292	-2.26
450	3280	-2.26
451	3320	-2.26
452	3313	-2.26
453	3348	-2.26
454	3380	-2.26
455	3365	-2.26
456	3395	-2.26
457	3454	-2.26
458	3459	-2.26
459	3569	-2.26
460	3590	-2.26
461	3586	-2.26
462	3594	-2.26
463	3631	-2.26
464	3616	-2.26
465	3645	-2.26
466	3658	-2.26
467	3642	-2.26
468	18716	-2.74
469	16267	-2.74
470	22298	-2.74
471	16000	-2.74
472	16971	-2.74
473	19571	-2.74
474	21688	-2.74
475	16534	-2.74
476	6234	-2.74
477	17164	-2.74
478	17217	-2.74
479	17235	-2.74
480	18361	-2.74
481	138	-2.74
482	141	-2.74
483	151	-2.74
484	2978	-2.74
485	2975	-2.74
486	3306	-2.74
487	3329	-2.74
488	3371	-2.74
489	3379	-2.74
490	3657	-2.74
491	3659	-2.74
492	3643	-2.74
493	22479	-3.52
494	15481	-3.52
495	13078	-3.52
496	10	-3.52
497	130	-3.52
498	134	-3.52
499	3629	-3.52

De la tabla anterior se observa que todos los aspirantes poseen medidas negativas, por lo tanto, se reafirma la deficiencia que los aspirantes traen al querer ingresar a la Universidad de El Salvador.

A continuación se presenta el Ranking de medidas de los ítems, donde la primera columna muestra la posición en que se encuentra el ítem de acuerdo a la medida obtenida, la segunda columna presenta el número de ítem y la tercera columna la medida obtenida o calibración del ítem.

Tabla 3.22: Ranking de Medidas de los Ítems

Corr.	Ítems	Medida
1	ítem_30	1.76
2	ítem_34	1.37
3	ítem_26	0.86
4	ítem_39	0.83
5	ítem_32	0.74
6	ítem_47	0.70
7	ítem_35	0.65
8	ítem_31	0.55
9	ítem_48	0.50
10	ítem_37	0.44
11	ítem_49	0.37
12	ítem_45	-0.02
13	ítem_44	-0.06
14	ítem_33	-0.16
15	ítem_40	-0.37
16	ítem_42	-0.50
17	ítem_50	-0.50
18	ítem_27	-0.51
19	ítem_28	-0.54
20	ítem_43	-0.55
21	ítem_41	-0.56
22	ítem_38	-0.64
23	ítem_46	-0.69
24	ítem_29	-1.78
25	ítem_36	-1.88

De esta manera se presenta el ranking de medidas de los estudiantes y el ranking de medidas de los ítems.

3.5 CONCLUSIONES Y RECOMENDACIONES

PARCIAL I DE LA ASIGNATURA DE MATEMÁTICA I

De la aplicación del modelo de Rasch a los datos proporcionados por el docente responsable de la asignatura de Matemática I, en el ciclo I/2006, impartida en la Facultad de Ciencias Naturales y Matemática por profesores de la Escuela de Matemática, se concluye:

1) De la matriz de datos proporcionada, solo fue posible estudiar las características de 172 estudiantes de los 176 proporcionados inicialmente; con respecto a los ítems o problemas en que consistía el test de Matemática I, fueron estudiados en su totalidad.
2) Los estudiantes de Matemática I, únicamente habían entendido el 50% o menos del tema evaluado.
3) Los problemas del parcial de Matemática I, únicamente podían ser resueltos correctamente por aquellos estudiantes que hubiesen entendido el 40% o más del tema evaluado.
4) Agregar más ítems en el test con el fin de mejorar la escala de medida ya que existen muchas partes de la escala que no son evaluadas. Además modificar o eliminar uno de los ítems que se solapan *(ítem 7: Propiedades de los exponentes en expresiones racionales* con *ítem* 5: *Simplificación y producto de expresiones con números irracionales* y *ítem 3: Transformación de un decimal a un racional* con *ítem 4: Simplificación de un número racional),* ya que tienen el mismo nivel de dificultad y hay muchos espacios en blanco en la escala de medida. Es decir, debe mejorarse la elaboración del test a fin de incluir problemas con diferentes niveles de dificultad.

5) Los estudiantes que no fue posible obtener las características con respecto al modelo de Rasch son: E81, E107, E123 y E136, los cuales han tenido una puntuación de 5 (Excelente) en todos los ítems.

6) De los estudiantes que se obtuvieron estadísticos, las personas con mayor medida fueron E96, E108, E106, E89, E127, E18, E59, E64 y E65 por el contrario las personas con menor medida son E31, E38, E54, E55, E69, E73, E142, E165, E91, E135 y E22, por lo tanto, el docente debe brindar una atención especial a los estudiantes con mayores niveles de rasgo; y en los estudiantes con medidas más bajas darles un

tratamiento especial a fin de mejorar su nivel de conocimiento del tema evaluado (en todos los ítems que obtuvieron la categoría más baja).

7) En el caso de los ítems, todos han podido ser ajustados al modelo de Rasch ya que no ha habido ítems extremos (todos los estudiantes hayan respondido correctamente o incorrectamente).

8) Los estudiantes que presentan desajustes con respecto al modelo de Rasch son: E98, E6, E99, E93, E42 y E130.

9) Del primer parcial aplicado a los estudiantes de Matemática I, el ítem 1 (*Propiedades de los números racionales e irracionales)* tiene una dificultad de -0.80 lógitos (el ítem más fácil), lo que indica que es un ítem relativamente fácil, sin embargo, el paso con mayor dificultad es el paso 4, por tanto, se espera que los estudiantes con un nivel de rasgo de 1.27 lógitos, resuelvan correctamente este ítem. El ítem 2 *(Propiedades de los números reales)*, tiene una dificultad de -0.48 lógitos, lo que indica que también es un ítem fácil, sin embargo, el paso más difícil es el paso 4, por tanto se espera que los estudiantes con un nivel de rasgo mayor o igual a 1.11 lógitos resuelvan correctamente todas las partes del ítem 2. El ítem 3 *(Transformación de un decimal a un racional)* tiene una dificultad de -0.09 lógitos, lo que indica que es un ítem relativamente fácil, sin embargo, el paso más difícil es el paso 1, por tanto se espera que los estudiantes que tengan un nivel de rasgo mayor o igual a 1.39 lógitos resuelvan correctamente todos los pasos del ítem. El ítem 4 *(Simplificación de un número racional)* presenta una dificultad de -0.01 lógitos, por tanto podemos decir que se trata de un ítem fácil, sin embargo, el paso más difícil es el paso 1, lo que nos indica que los estudiantes con un nivel de rasgo mayor o igual a 0.91 lógitos, resolverán correctamente todas las partes de dicho ítem. El ítem 5 *(Propiedades de los exponentes en expresiones racionales)*, tiene una dificultad de 0.56 lógitos (el ítem más difícil), por tanto, se considera un ítem difícil, sin embargo, el paso más difícil es el paso 1, lo que nos indica que los estudiantes con un nivel de rasgo mayor o igual a 1.08 lógitos resuelvan correctamente todas las partes de dicho ítem. El ítem 6 *(Simplificación de expresiones con números irracionales)*, tiene una dificultad de 0.30 lógitos, por lo tanto se considera un ítem difícil, sin embargo, el paso más difícil es el paso 1, lo que indica, que los estudiantes con un nivel de rasgo mayor o igual a 2.31 lógitos resolverán correctamente todas las partes de dicho ítem. Por último, el ítem 7 *(Simplificación y producto de expresiones con números irracionales)*, tiene una dificultad

de 0.53 lógitos, por tanto es considerado un ítem difícil, sin embargo, el paso más difícil es el paso 1, lo que nos indica que los estudiantes con nivel de rasgo mayor o igual a 2.25 lógitos resolverán correctamente todas las partes de dicho ítem.

10) Por último, podemos decir que según los resultados, es necesario modificar el examen de matemática I, debido a que se necesitan más ítems que midan la mayor parte de la escala e ítems con distintos niveles de dificultad, ya que los que se tienen están concentrados alrededor de 0 lógitos. Y además, podrán ser resueltos por aquellos estudiantes que han entendido el 40% o más del tema evaluado.

TEST DE NUEVO INGRESO, UES 2007

De la aplicación del Test de Nuevo Ingreso para los aspirantes a ingresar a la Universidad de El Salvador, se concluye:

1) De la matriz de datos proporcionada, solo fue posible estudiar las características de 499 aspirantes de una muestra de 500 personas; con respecto a los ítems o problemas en que consistía el Test de Nuevo Ingreso, fueron estudiados en su totalidad.

2) Los aspirantes a ingresar a la Universidad de El Salvador, únicamente tenían conocimiento de un 38% del Test de Nuevo Ingreso, UES 2007.

3) Agregar más ítems en el test (ítems con niveles de dificultad menores a -2 lógitos e ítems con niveles de dificultad mayores a 2 lógitos) con el fin de llenar los espacios vacíos en la escala del test ya que existen muchas partes que no son evaluadas. Además, existen 3 grupos de ítems bien definidos y entre ellos se solapan. Del primer grupo de ítems (ítems 38, 41 y 43) lo recomendable es eliminar o modificar dos de los tres ítems, en el segundo grupo de ítems (ítems 27, 28, 42 y 50) lo recomendable es eliminar o modificar tres de los cuatro ítems, de igual manera el tercer grupo de ítems (ítems 37 y 49) se recomienda eliminar o modificar uno de los dos ítems y del último grupo de ítems (ítems 32, 35 y 47) se recomienda eliminar o modificar dos de los tres ítems, con el fin que midan niveles de dificultad diferentes en el test., ya que tienen el mismo nivel de dificultad y hay muchos espacios en blanco en la escala de medida.

4) Del sujeto que no fue posible obtener las características con respecto al modelo de Rasch es el aspirante 5171 por no haber resuelto correctamente ningún ítem que comprende el test.

5) De los aspirantes que se obtuvieron estadísticos, las personas con mayor medida pero que siempre con niveles de conocimiento deficiente son 22410, 17143, 3485 y 3665, por el contrario las personas con menor medida son 22479, 15481, 13078,10, 130, 134 y 3629, por lo tanto, es necesario que las autoridades correspondientes brinden un cursos de nivelación en el área de matemática con docentes capacitados antes de someterse al examen de nuevo ingreso a la Universidad y de esta manera minimizar los altos grados de reprobación y niveles de frustración de los aspirantes a ingresar a la UES.

6) En el caso de los ítems, todos han podido ser ajustados al modelo de Rasch ya que no han habido ítems extremos (todos los aspirantes hayan respondido correctamente o incorrectamente).

7) Por último tenemos 5 aspirantes (3319, 3588, 12183, 2411 y 3534) que presentan desajustes con el modelo de Rasch (OUTFIT mayores a 2 lógitos) y en nuestro caso no hay ítems que estén afectando la obtención de las medidas de estudiantes e ítems. Por tanto, se puede decir que estos aspirantes han resuelto o respondido los ítems 26, 30, 31, 32, 34, 35, 37, 39, 48 y 49 de una manera al azar.

8) Aplicar a los aspirantes a ingresar a la UES un instrumento de evaluación que previamente haya sido validado.

REFERENCIAS BIBLIOGRÁFICAS

1. Abad, Francisco J.; García Gil, Carmen Beatriz; Olea, Julio; Ponsola, Vicente; Revuelta, Javier. (2004). Introducción a la Psicometria: Teoría Clássica del Test y Teoría de Respuesta al Ítem. Universidad Autónoma de Madrid.
2. Abad, José Francisco; Ponsoda, Vicente; Revuelta, Javier. (2006). Modelos Politómicos de la TRI. Editorial de La Muralla S.A.
3. Adánez Prieto, Gerardo; Delgado R., Ana. (2001). Análisis de un Test mediante el Modelo de Rasch. http://biblioteca.universia.net/irARecurso.do?
4. Andrade, Dalton F. (2005). Teoria da resposta ao item: Conceitos, Modelos e Aplicações. Departamento de Informática e Estatística – UFSC. http://www.inep.gov.br/download/saeb/2004/resultados/BRASIL.pdf.
5. Baker, Frank B. (2001). The Basic of Item Response Theory. University of Wisconsin (EEUU).
6. Barbero, Isabel; Marañón, Pedro Prieto; Suárez, Juan y Costas, Concepción. (2001). Relaciones Empíricas entre los estadísticos de la Teoría Clásica de los Test y los de la Teoría de Respuesta al Ítem. Universidad Nacional de Educación a Distancia y Universidad de La Laguna.
7. Fischer, Gerhard H. and Moolenar, Ivo W. (1995). Rasch Models: Foundations, Recent Developments, and Applicattions. Springer-Verlag.
8. Jiménez, Mariano (1995). Noticias ICI: Breves Notas Sobre el Modelo de Rasch (1). Instituto Tecnológico de San Luís Potosí (México).
9. López, Agustín Tristán (1999). Análisis de Rasch para Todos (Una Guía Simplificada para Evaluadores), Instituto Tecnológico de San Luís Potosí (México).
10. Melia, J. L. (2003). Métodos de Estimación en Teoría de la Respuesta al Ítem. Universidad de Valencia. http://www.uv.es/psicometria.
11. Meliá, José Luís; Lloret, Susana; Hontangas, Pedro. (Curso 2004-2005). Avances Metodológicos en la Medición de Habilidades Cognitivas. Universidad de Valencia, http://www.uv.es/meliajl/Docencia/Doctorat/Progdoct04_05.pdf.
12. Mendo, Antonio Hernández; Sánchez Morales, Verónica; Rodríguez, Josefina. (2005). La Teoría de Respuesta al Ítem en la Construcción de Cuestionarios en Psicología del Deporte. Universidad de Málaga (España), http://www.efdeportes.com.
13. Ostini, Remo; Nering, Michael L. (2006). Polytomous Item Response Theory Models. University of Illions.

14. Partchev, Ivailo. (2004). A visual guide to item response theory, Universität Jena.
15. Ravela, Pedro. (2003). ¿Que Significa los números de las Evaluaciones?. Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL).
16. Rojas, Margarita; Manríquez, Germán; Gatica, Yanira. (2004) .Curso de UML Basado en la Teoría de Respuesta al Ítem Multiplataforma Adaptativo. Universidad de Concepción (Chile).
17. Vidóni, Daniele. Verso una Misura Oggettiva: Criteri per la realizzazione e la valutazione di test a risposta aperta. Daniele Bidón, Comissione Europea.
18. Wiberg, Marie. (2004). Classical Test Theory vs. Item Response Theory. Univeritet UMEA.

ANEXOS

ANEXO I: GRÁFICOS DE AJUSTE DE LOS ESTUDIANTES E ÍTEMS

```
AJUSTE EXTERNO DE LOS ESTUDIANTES          tesis Dec 16 20:52 2006
INPUT:  176  ESTUDIS,  7  PARCIAS   ANALYZED: 172  ESTUDIS,  7  PARCIAS,  6
CATS    v2.82
-----------------------------------------------------------------------------
          -3         -2         -1          0          1          2          3
          ++----------+----------+----------+----------+----------+----------++
        3 +                                 |                                 +   3
          |                                 |        B                        |
          |                                 |   L                             |
    E     |                               I |                                 |
    S   2 +---------------------------------|--------E-D----------------------+   2
    T     |                                 |                                 |
    U     |                                 2V  S                             |
    D     |                        F G N2   1 1      O                        |
    I   1 +                A             2  |   2 1 1            C            +   1
          |                            Y 1  | 2        W                      |
    O     |                R         1  1 2 1 1 1                             |
    U     |                2         1 1   1313 113              1            |
    T   0 +--------H------------5-------U12-|-122-----------------------------+   0
    F     |3       2       2    4  2   1  1113 11   1      1  1  1            |
    I     |a       6            4  3 1 111 1|11  1   1     1                  |
    T     |                5    1  1 3 2112 | 2  1      2                     |
       -1 +                        2 w    111  11   1    1                    +  -1
    Z     |                        b 3  t2 12                                 |
    S     |                              u  111     1v  j                     |
    T     |                            c    |                                 |
    D  -2 +---------------------------------|--z------------------------------+  -2
          |                                 |   s                             |
          |                                 |                                 |
          |                                 |                                 |
       -3 +                                 |                                 +  -3
          ++----------+----------+----------+----------+----------+----------++
          -3         -2         -1          0          1          2          3
                                    ESTUDI MEASURE
AJUSTE INTERNO DE LOS ÍTEMS       tesis Dec 16 20:52 2006
INPUT:  176  ESTUDIS,  7  PARCIAS   ANALYZED: 172  ESTUDIS,  7  PARCIAS,  6
CATS    v2.82

          -3         -2         -1          0          1          2          3
          ++----------+----------+----------+----------+----------+----------++
        3 +                             |                                     +   3
          |                             |                                     |
          |                             |                                     |
    P     |                             |                                     |
    A   2 +-----------------------------|-------------------------------------+   2
    R     |                        A    |                                     |
    C     |                             |                                     |
    I     |                             |  B                                  |
    A   1 +                             |   C                                 +   1
          |                             |      D                              |
    I     |                             |                                     |
    N     |                             |                                     |
    F   0 +-----------------------------|---------c---------------------------+   0
    I     |                             |                                     |
    T     |                             |                                     |
          |                             |                                     |
       -1 +                             |                                     +  -1
    Z     |                             |                                     |
    S     |                             |                                     |
    T     |                             |         b                           |
    D  -2 +----------------------------a|-------------------------------------+  -2
          |                             |                                     |
          |                             |                                     |
          |                             |                                     |
       -3 +                             |                                     +  -3
          ++----------+----------+----------+----------+----------+----------++
          -3         -2         -1          0          1          2          3
                                    PARCIA MEASURE
```

```
AJUSTE EXTERNO DE LOS ÍTEMS       tesis Dec 16 20:52 2006
INPUT: 176 ESTUDIS, 7 PARCIAS   ANALYZED: 172 ESTUDIS, 7 PARCIAS, 6
CATS   v2.82

      -3        -2        -1         0         1         2         3
      ++---------+---------+---------+---------+---------+---------++
    3 +                          |                                  +   3
      |                          |                                  |
      |                          |                                  |
P     |                          |                                  |
A   2 +--------------------------|----------------------------------+   2
R     |                          |                                  |
C     |                      A   |                                  |
I     |                          |                                  |
A   1 +                          |                                  +   1
      |                          |                                  |
O     |                          |                                  |
U     |                          |   C                              |
T   0 +--------------------------|------D---------------------------+   0
F     |                          |  B     c                         |
I     |                          |                                  |
T     |                          |                                  |
   -1 +                          |                                  +  -1
Z     |                          |        b                         |
S     |                          |                                  |
T     |                          |                                  |
D  -2 +--------------------------|----------------------------------+  -2
      |                         a|                                  |
      |                          |                                  |
      |                          |                                  |
   -3 +                          |                                  +  -3
      ++---------+---------+---------+---------+---------+---------++
      -3        -2        -1         0         1         2         3
                              PARCIA MEASURE
```

ANEXO II: FUNCIONES DE RESPUESTA CATEGÓRICAS PARA EL ÍTEM 6 Y 7

```
GRÁFICO PARA EL ÍTEM 6           tesis Dec 16 20:52 2006
INPUT: 176 ESTUDIS, 7 PARCIAS    ANALYZED: 172 ESTUDIS, 7 PARCIAS, 41
CATS   v2.82
----------------------------------------------------------------------
CATEGORY PROBABILITIES: MODES - Step measures at intersections
P      ++-----+-----+-----+-----+-----+-----+-----+----+-----++
R  1.0 +000000000000000000                           55555555555+
O      |                  0000                 55555            |
B      |                      00             55                 |
A      |                        0           5                   |
B   .8 +                         0         5                    +
I      |                          0       5                     |
L      |                          0       5                     |
I      |                           0     5                      |
T   .6 +                           0    5                       +
Y      |                            0   5                       |
    .5 +                            0  5                        +
O      |                            0  5                        |
F   .4 +                             0 5                        +
       |                             05                         |
R      |                              *                         |
E      |                             50                         |
S   .2 +                             *3*3                       +
P      |                            *220 33                     |
O      |                          2*5  2*  33                   |
N      |                     1******1***4***4***44              |
S   .0 +**************************444   111*********************+
E      ++-----+-----+-----+-----+-----+-----+-----+----+-----++
       -5    -4    -3    -2    -1     0     1     2     3     4
                            ESTUDI [MINUS] PARCIA MEASURE
```

```
GRÁFICO PARA EL ÍTEM 7           tesis Dec 16 20:52 2006
INPUT: 176 ESTUDIS, 7 PARCIAS  ANALYZED: 172 ESTUDIS, 7 PARCIAS, 41
CATS   v2.82
CATEGORY PROBABILITIES: MODES - Step measures at intersections
P      ++-----+-----+-----+-----+-----+-----+-----+----+-----++
R  1.0 +00000000000000000                                   5555+
O      |                 00000                       555555     |
B      |                      00                  555           |
A      |                        0              555              |
B   .8 +                         0            5                 +
I      |                         0           5                  |
L      |                          0         5                   |
I      |                           0       5                    |
T   .6 +                           0      5                     +
Y      |                           0     5                      |
    .5 +                            0   5                       +
O      |                            0   5                       |
F   .4 +                             0 5                        +
       |                             05                         |
R      |                              * 3                       |
E      |                             5*3 333                    |
S   .2 +                             * 0    33                  +
P      |                            *  0      333               |
O      |                          **2222*2       333            |
N      |                   111*****1111110*222      3333333     |
S   .0 +************************5        11*********************+
E      ++-----+-----+-----+-----+-----+-----+-----+-----+-----++
       -5    -4    -3    -2    -1     0     1     2     3     4
                            ESTUDI [MINUS] PARCIA MEASURE
```

ANEXO III: ESCALOGRAMA DE GUTTMAN

```
TABLE 22.1 TESIS                  tesis Dec 16 20:52 2006
INPUT: 176 ESTUDIS, 7 PARCIAS   ANALYZED: 172 ESTUDIS, 7 PARCIAS, 41
CATS   v2.82
--------------------------------------------------------------------
GUTTMAN SCALOGRAM OF RESPONSES:
ESTUDI|PARCIA
      |1234675
      |-------
    81 +5555555
   107 +5555555
   123 +5555555
   136 +5555555
    47 +5454555
    96 +5555355
   108 +3555555
   106 +5345555
    89 +3545554
   127 +4455525
    18 +3455553
    59 +3445534
    64 +4355552
    65 +2355554
     6 +5553055
   128 +2555505
    27 +1455434
    50 +3355532
    98 +5440535
    99 +5525055
   126 +1552554
    66 +2254534
    76 +2255515
    43 +3355521
   109 +2351554
    33 +2543035
    44 +2325505
   100 +4422532
   129 +3552331
    88 +4452204
   113 +4335501
   174 +4255302
    40 +5234321
    53 +3324332
   150 +2205354
   161 +2255330
   164 +2355005
    24 +2154350
    42 +5523004
    71 +4224350
   119 +2250551
   130 +2220455
   149 +2213453
    52 +2253321
   120 +2452023
   121 +3355200
   158 +2255004
   166 +1255320
    17 +3350501
    19 +3252302
    20 +1332503
    63 +3225500
```

```
GUTTMAN SCALOGRAM OF RESPONSES:
ESTUDI|PARCIA
      |1234675
      |-------
   124 +3455000
   131 +2305502
   137 +2245050
     2 +3253003
    35 +2404052
    87 +2320504
   154 +1304233
   175 +2355001
    12 +3435000
    77 +2325102
   105 +1320504
   115 +2355000
   153 +2215032
   157 +2422500
   160 +3205302
     9 +4205030
    23 +2243003
    56 +4350101
   101 +3202251
   159 +2153030
   172 +2325200
     3 +2234002
    46 +4225000
    49 +2240221
    92 +5340001
   152 +3105050
   176 +2205301
    14 +2205003
    16 +2244000
    28 +2350002
    85 +2221302
   138 +2152110
     8 +2120330
    30 +2123003
    79 +2125001
    80 +2225000
    93 +5300003
   104 +2230202
   140 +2105021
   148 +3205001
   163 +2225000
     5 +2233000
    10 +2300500
    32 +4130002
    51 +2233000
    78 +2230300
    90 +1230202
    97 +3320200
   102 +4420000
   116 +2231200
     1 +2200500
    60 +2205000
    74 +2320002
    83 +2221200
   146 +2202210
   147 +3210030
   151 +2101500
```

```
GUTTMAN SCALOGRAM OF RESPONSES:
ESTUDI|PARCIA
      |1234675
      |-------
   15 +2240000
   37 +3230000
   67 +2221100
   82 +2201030
  110 +3100400
  117 +2330000
  122 +2220020
  171 +2200220
   45 +2220100
   61 +0250000
   68 +2230000
   86 +2220001
   95 +3200020
  112 +2131000
  141 +2220001
  145 +3103000
  162 +3220000
  167 +2220010
  170 +2320000
    7 +0123000
   21 +2120001
   25 +2202000
   48 +2130000
   58 +2220000
   72 +2220000
  125 +2130000
  144 +2211000
  156 +2201010
  168 +2121000
   13 +2300000
   39 +3200000
   41 +2201000
   62 +2102000
   70 +3101000
   75 +3110000
   94 +2120000
  103 +2120000
  111 +3101000
  114 +2200010
  132 +3200000
  133 +2200001
  139 +2120000
  143 +2101010
   11 +2200000
   26 +2100010
   29 +1120000
   36 +2100100
   57 +2200000
   84 +1100002
  118 +3100000
  134 +2200000
  155 +1210000
  169 +2200000
  173 +2200000
    4 +1200000
   31 +2100000
   38 +2100000
```

```
GUTTMAN SCALOGRAM OF RESPONSES:
ESTUDI|PARCIA
      |1234675
      |-------
   54 +3000000
   55 +2100000
   69 +2100000
   73 +2100000
  142 +1200000
  165 +2100000
   34 +2000000
   91 +2000000
  135 +1100000
   22 +1000000
      |-------
      |1234675
```

ANEXO IV: ÍTEMS DEL TEST DE NUEVO INGRESO, UES 2007

26. Al resolver la ecuación Log 8 + Log **x** = 3 se obtiene para **x** el valor de

A) 75
B) 125
C) 250
D) 240

27. Al simplificar la expresión $4^{3/2} + 16^{5/4} - 32^{2/5}$ se obtiene

A) 2^6
B) 2^{10}
C) 6^2
D) 8^6

28. Al multiplicar un número por 24, su valor aumenta en 1334 unidades. ¿El número es?

A) 58
B) 53.36
C) 55.58
D) 57

29. El 3% de 81 es igual al 9% de

A) 27
B) 54
C) 72
D) 90

30. Si 6 gatos cazan 6 ratones en 6 minutos, entonces el número de ratones que 30 gatos pueden cazar en 30 minutos es:

A) 6
B) 30
C) 150
D) 180

31. Dada la sucesión infinita de números: 1, 4, 9, 16, 25, 36, 49, 64,... ¿Qué número sigue después de 64?

A) 65
B) 81
C) 74
D) Cualquier número mayor que 64.

32. "Tres veces **z** más dos veces la suma de **x** y **y**" se expresa en notación algebraica, así:

A) (**z** + 3) + (**x** + 2) + **y**
B) (**z** + 3) + (**x** + **y** + 2)
C) 3**z** + 2**y** + **y**
D) 3**z** + 2(**x** + **y**)

33. Al desarrollar $(1+\mathbf{x}^2)(1-\mathbf{x}^3)$, entonces se obtiene:

A) $1-\mathbf{x}^5$
B) $1-\mathbf{x}^6$
C) $1+\mathbf{x}^2-\mathbf{x}^3$
D) $1+\mathbf{x}^2-\mathbf{x}^3-\mathbf{x}^5$

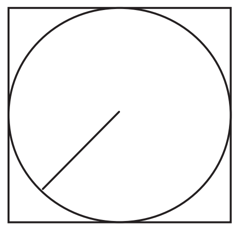
34. Factorice completamente la siguiente expresión $x^2y-25y+3x^2-75$

A) $(x+5)^2(x-5)^2(y+3)$
B) $(x+5)(x-5)(y+3)$
C) $(x+5)(x-5)(y-3)$
D) $(x+5)(x-5)(y-3)$

35. Dos atletas se encuentran a una distancia de 12 kilómetros. Si salen corriendo el uno hacia el otro, de tal forma que la rapidez del segundo es el triple que la rapidez del primero, ¿a qué distancia del punto medio del trayecto entero se cruzan?

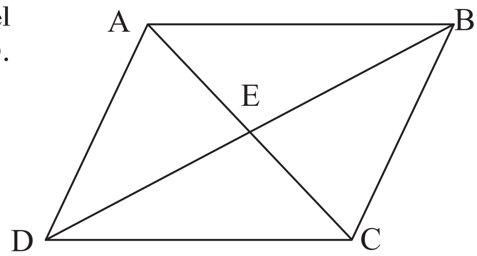
A) 6 Km.
B) 2 Km.
C) 3 Km.
D) 1 Km.

36. En la figura adjunta se tiene un círculo inscrito en un cuadrado. Si el radio del círculo mide 2 cm., entonces la medida lado del cuadrado es



A) $2\sqrt{2}$ cm
B) 2 cm
C) 4 cm
D) $\sqrt{2}$ cm

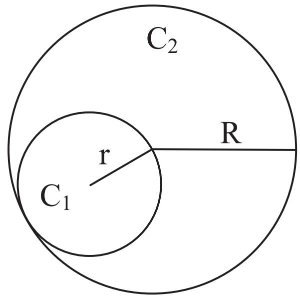
37. El cuadrilátero ABCD es un paralelogramo, E es el punto en el cual se cortan las diagonales AC y BD. Entonces los triángulos Δ ABE y Δ CDE son



A) isósceles y congruentes
B) semejantes y congruentes
C) semejantes y no congruentes
D) congruentes y no semejantes

38. En el gráfico, el círculo C_1, de radio r, está incluido en el círculo C_2, de radio R. El círculo C_1 es tangente interiormente al círculo C_2 y pasa por el centro de C_2. Si el área de C_1 mide 5 cm^2 entonces el área de C_2 es

A) 10 cm^2
B) 10π cm^2
C) 20 cm^2
D) 20π cm^2

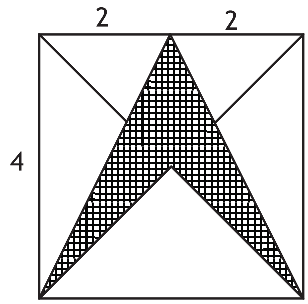


39. ¿A qué es igual el suplemento de $40^o - \alpha$?

A) $130^o + \alpha$
B) $140^o + \alpha$
C) $150^o + \alpha$
D) $120^o + \alpha$

40. ¿Cuál es el área de la zona sombreada de la figura?

A) 5
B) 7
C) 4
D) 9



41. La sombra de un monumento mide 10 m, y la de una varilla vertical de 1 m de altura, situada a su lado, mide, en el mismo momento, 40 cm; ¿qué altura tiene el monumento?

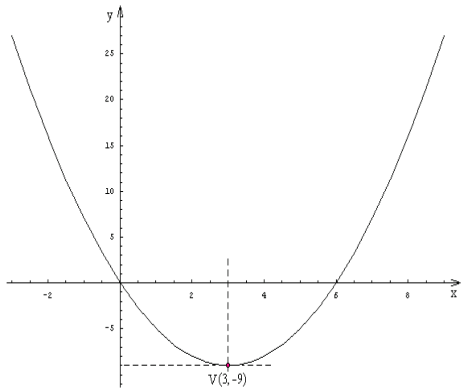
A) 4 m B) 25 m C) 40 m D) 32.4 m

42. Dados los conjuntos $A = \{x \in R \ / \ x > -2\}$ y $B = \{x \in R \ / \ x \leq 5\}$, la intersección de A con B es:

A)]-2, 5] B)]-2, 5[C) [-2, 5[D) $\varnothing$

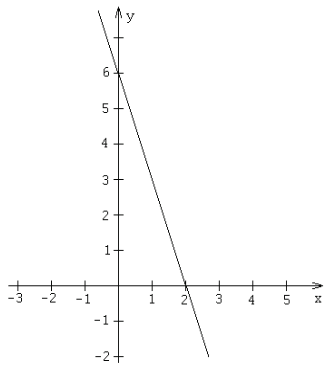
43. A continuación se muestra la gráfica de la ecuación $(y + 9) = (x - 3)^2$. ¿Cuál de los desplazamientos tiene la grafica de la ecuación $(y + 10) = (x - 3)^2$?

A) La gráfica mostrada se traslada verticalmente una unidad hacia arriba
B) La gráfica mostrada se traslada verticalmente una unidad hacia abajo
C) La gráfica mostrada se traslada horizontalmente una unidad hacia la izquierda
D) La gráfica mostrada se traslada horizontalmente una unidad hacia la derecha



44. ¿A cuál de las ecuaciones corresponde la gráfica siguiente?

A) $y = -3x + 6$
B) $y = x - 6$
C) $y = 3x - 6$
D) $y = x^2$



45. El gráfico de la función $f(x)=2(x+1)^2-1$ es una parábola vertical
A) de vértice V(1,-1), abierta hacia arriba.
B) de vértice V(-1,1), abierta hacia abajo.
C) de vértice V(2,-1), abierta hacia arriba.
D) de vértice V(-1,-1), abierta hacia arriba.

46. La función inversa de $f(x)=2x-2$ es
A) $f^{-1}(x)=-2x+2$
B) $f^{-1}(x)=\frac{1}{2}(x-1)$
C) $f^{-1}(x)=\dfrac{1}{2x-2}$
D) $f^{-1}(x)=\dfrac{x+2}{2}$

47. La nota media conseguida en una clase de 20 alumnos ha sido de 6. Diez alumnos han reprobado con nota 3 y el resto obtuvo más de 5. ¿Cuál es la nota media de los alumnos aprobados?
A) 9 B) 5 C) 4.5 D) 3

48. Se pretende ordenar a un grupo de 3 señoras y 3 señores en una línea. ¿De cuántas maneras se puede hacer si se desea que las 3 señoras permanezcan juntas?
A) 144 B) 72 C) 24 D) 36

49. Se dispone de 10 tarjetas enumeradas del 1 al 10 en una urna; se extraen dos tarjetas, ¿cuál es la probabilidad de que las dos presenten números pares cuando se extraen una tras otra si la primera extraída no se regresa a la urna?
A) 2/9 B) 5/7 C) 1/3 D) 2/5

50. Sea C(n, r) el número combinatorio. ¿Cuánto es C(6,1) + C(6,2) + C(6,3) + C(6,4) + C(6,5)?
A) 64 B) 62 C) 36 D) 63

ANEXO V: ESCALOGRAMA DE GUTTMAN PARA EL TEST DE NUEVO INGRESO

```
GUTTMAN SCALOGRAM OF RESPONSES:
ASPIRA|PROBLE
      |1 2111  121 12212 12 1
      |1413683275589042360274195
      |-------------------------
   132 +1001110010101000010001110
   159 +1101101001110101000100000
   410 +1011011101001000011010000
   488 +0100110101100100011101000
    30 +0001111110100110010000000
   217 +1101101001000011100001000
   305 +1110011100110001100000000
   333 +1011000100001110000011100
   346 +0110100000101100011010100
   372 +1110001111011010000000000
   443 +1111100001001010000011000
   453 +1100110111100100001000000
   499 +1111100110100010100000000
     7 +1100101111000100010000000
    53 +1111001010000000001010001
    65 +1011101010010000001000100
   110 +1111000001011100000010000
   113 +0110111111000010000000000
   140 +1111000110000100001100000
   144 +1100010011000010011010000
   166 +1100101001111100000000000
   222 +1100010010100100000001110
   313 +0001100111000000100111000
   337 +1111001100010000100100000
   347 +1111110000010001001000000
   350 +1100001101110000110000000
   359 +1110100010100000010100100
   377 +1011110100000100110000000
   382 +1110101100010000100100000
     3 +1000101100100100001010000
     6 +1000010101101010000001000
    10 +1000101011000110000000100
    43 +1010110000000000010110001
    44 +1100110110000001010000000
    45 +1100000111000010010010000
    57 +1010011011000000001001000
    66 +1100000100100010011001000
    77 +1110000101001000001100000
    78 +1100101101000000000010100
    79 +1010011000011000000100100
   101 +1000011011001110000000000
   139 +1010101010001010000010000
   157 +1000011101101100000000000
   164 +0100000111011000000100100
   176 +0100011111000010000000010
   196 +1000001111000001010000010
   200 +1101000010010110000000001
   202 +1101100111010000000000000
   204 +1100001010101100000010000
   208 +1110000100001101000000100
   210 +1110011000100100000000100
   219 +1000100101101000001000100
   261 +0000011001110011001000000
   270 +1111000001110000100000000
```

```
GUTTMAN SCALOGRAM OF RESPONSES:
ASPIRA|PROBLE
      |1 2111  121 12212 12 1
      |1413683275589042360274195
      |-------------------------
   334 +0101110101010100000000000
   378 +1011001011100000010000000
   394 +1101101011100000000000000
   399 +1110011100001100000000000
   430 +1100001100010000101000100
   431 +1100000000010111011000000
   434 +0110101001000100001100000
   437 +0101000100100000001011001
   438 +1000111100100001000000010
   452 +0110111010000000100001000
   478 +0101001110111000000000000
   491 +1101100100000000011000001
     1 +1000001010000010010000110
     9 +1110110100000001000000000
    22 +0001010010001010000001010
    31 +1001000010110001000000010
    32 +0001110000000010001000101
    39 +1011000010011000000000001
    48 +1100001010000100000011000
    49 +1100110000011000000100000
    50 +1110000000111000000100000
    58 +1111010010000010000000000
    75 +1100100110000000100000010
    81 +0100001011110001000000000
    85 +1100000001110100000000100
    89 +1100010000101010100000000
   103 +0001110110100000000001000
   117 +1110100001110000000000000
   119 +0101111000001000000000100
   129 +0101010010000100000000101
   131 +1000110110100001000000000
   134 +0101000011100010000000001
   136 +1110001001001000000000100
   137 +1101100111000000000000000
   146 +0101110001110000000000000
   162 +1100010101000110000000000
   174 +0011010001001001000000100
   178 +1100111001100000000000000
   185 +0111000100100000000010010
   188 +1000101100100000000110000
   212 +1101000011000000010001000
   226 +0010001101011001000000000
   234 +0010110100001100000000010
   237 +1010010000111000100000000
   238 +0110101011000010000000000
   241 +1110011000000001100000000
   246 +0111010001010000000000010
   249 +1001010011000100000100000
   251 +1101101010000100000000000
   273 +0110000100100000010000011
   274 +0100101010000000001001100
   275 +1110011000010000000100000
   279 +1110100000010100000100000
   281 +1100111100010000000000000
   299 +1100111000001001000000000
   302 +1110001000010000100000100
```

```
GUTTMAN SCALOGRAM OF RESPONSES:
ASPIRA|PROBLE
      |1 2111  121 12212 12 1
      |1413683275589042360274195
      |-------------------------
   307 +1111010000010000000000010
   310 +0111000000010001010001000
   316 +0000010001000110100100100
   318 +1010110000001000000100010
   329 +0010000100010010001010100
   335 +0110000000010101101000000
   339 +0111100000010000000001100
   354 +1000001110001000100000100
   355 +1000111100001000100000000
   375 +1110010000110000001000000
   376 +1000110000001010010010000
   379 +1110000011000100000100000
   386 +0001111010001000000010000
   402 +1101001101000000010000000
   408 +0000010011000001101010000
   412 +0100011101100000000010000
   416 +0001111000000100000001100
   420 +1100010001010001100000000
   432 +1001101000100100010000000
   435 +0100011100100000000101000
   440 +1110001000001000010001000
   447 +0001001110010000010000001
   462 +1101010000100000010001000
   467 +1101011000000000000110000
   471 +1011000110100000000010000
   474 +1101000000010100100000100
   477 +1011001001101000000000000
   496 +1101100100001000000000010
   498 +0110010110010100000000000
   500 +1100100100100010001000000
     2 +0001101100000100000000001
     4 +1001001000100010000010000
     5 +1100000110000010010000000
    14 +0111001001000000001000000
    18 +1000100000101010000000100
    20 +0100101100100010000000000
    21 +0000101010100100010000000
    23 +1101100001000000000000100
    27 +1011000000100101000000000
    28 +0110100010000011000000000
    38 +1100100011001000000000000
    46 +0101011000000000100000001
    52 +1100000111000000000100000
    55 +1110100000010000000000100
    56 +1101010000001000010000000
    74 +1011010001000000000010000
    80 +1100000110000011000000000
    82 +1100010010000001001000000
    84 +1110010000100000001000000
    86 +1110001000010000100000000
    91 +1000111000001000100000000
    92 +1110001001010000000000000
    97 +1101000100000000100000100
    98 +1101100010000001000000000
   107 +1110110000000000100000000
   111 +1000111110000000000000000
```

```
GUTTMAN SCALOGRAM OF RESPONSES:
ASPIRA|PROBLE
      |1 2111  121 12212 12 1
      |1413683275589042360274195
      |-------------------------
   114 +1011001100010000000000000
   116 +0011011010100000000000000
   118 +0101000001001000001001000
   133 +0101100000001110000000000
   142 +1000101001100100000000000
   148 +1100010010000100000100000
   158 +0010110000011100000000000
   169 +1001100000001001000001000
   177 +0010111100000100000000000
   182 +1010010000100101000000000
   189 +1000101011000100000000000
   197 +1000101010001001000000000
   209 +0011000001010001000001000
   216 +1100100001000000001001000
   221 +1001001110000010000000000
   229 +1010000110000000011000000
   235 +1010001100001000000000001
   242 +1110000001010000000100000
   254 +1111001000001000000000000
   255 +0000000111000011000100000
   264 +0100000010001101000000100
   268 +0100100001110100000000000
   280 +0010001011000000010100000
   287 +1111001000100000000000000
   291 +1110010000010000100000000
   308 +1101000000011100000000000
   309 +0111000100010100000000000
   311 +0110001100010000100000000
   315 +1110101000010000000000000
   319 +0100010001100110000000000
   321 +1100010001010100000000000
   330 +1110001100000000001000000
   332 +1010000011100000000000100
   340 +1111000000000100100000000
   343 +1010000011010000100000000
   348 +1100000100010000001000100
   349 +1101011001000000000000000
   351 +1110001000010000100000000
   356 +1010001100010000000100000
   362 +1011000001001000100000000
   364 +1101000000110000100000000
   365 +1100010010010000100000000
   373 +1100001000110000100000000
   381 +1100110000000010010000000
   383 +1100010001000001010000000
   388 +1110000010100010000000000
   390 +0100110010100000000000100
   393 +1110010001010000000000000
   395 +1100000010100000000100100
   404 +1100100000100001000010000
   411 +1110000100010000100000000
   415 +1101000000010000001001000
   426 +1100000001110100000000000
   436 +0001001001111000000000000
   439 +0001011100000110000000000
   444 +1111010000000000001000000
```

```
GUTTMAN SCALOGRAM OF RESPONSES:
ASPIRA|PROBLE
      |1 2111  121 12212 12 1
      |1413683275589042360274195
      |-------------------------
   445 +0100001000110001000100000
   446 +1101000000100000110000000
   456 +0010000100000101000000110
   465 +0011000010001001100000000
   468 +1101000010000000100000010
   469 +1000111100001000000000000
   470 +1010100010000010100000000
   472 +1100101000000000010010000
   479 +1100001000110001000000000
   489 +1110000001010100000000000
   492 +0101000101000011000000000
    15 +1100001010001000000000000
    16 +0001000010000010001010000
    17 +1111000000001000000000000
    24 +1001000000000010000010010
    26 +1010000100100100000000000
    33 +0100001001101000000000000
    37 +1000000010100001001000000
    41 +1000100000000001001001000
    42 +0100010101000000100000000
    47 +1100100100000100000000000
    61 +1001000000001010100000000
    62 +1101000000001000000100000
    63 +1101100000000000000010000
    64 +0101000011000001000000000
    67 +0011000000100101000000000
    68 +1100010100000000010000000
    69 +0101001001001000000000000
    70 +1100100000001100000000000
    72 +1110001000000000100000000
    87 +1001001000001100000000000
    99 +1100010000001000000100000
   112 +1100100001000001000000000
   120 +1100100000001000000001000
   124 +0101100010000000000001000
   125 +0110100001000000000100000
   127 +1100100001000000000010000
   130 +1100010010100000000000000
   135 +1110001000001000000000000
   145 +0000011101000000100000000
   149 +0011000010001010000000000
   151 +1000111010000000000000000
   152 +0111000100100000000000000
   153 +0100001101010000000000000
   154 +0100010010100000000001000
   160 +1101000100000000001000000
   167 +0000001101101000000000000
   171 +1000100010001000000000010
   172 +0101001010000010000000000
   180 +1101101000000000000000000
   183 +1000001000100001000001000
   184 +1000100010100000000100000
   186 +1001000000001100000100000
   191 +0100000000101100000100000
   192 +1000000100000100010100000
   193 +1010000000011000000010000
```

```
GUTTMAN SCALOGRAM OF RESPONSES:
ASPIRA|PROBLE
      |1 2111  121 12212 12 1
      |1413683275589042360274195
      |-------------------------
   206 +1100100000100100000000000
   207 +0110000110000000000001000
   211 +1101010000000000000000100
   215 +1111000000100000000000000
   218 +0110000000100000110000000
   220 +1001110000000000000000001
   245 +1000110000000110000000000
   247 +1001101010000000000000000
   250 +0010001100000100000000010
   252 +0101101100000000000000000
   265 +1100000000000010010010000
   266 +0000001001100100000010000
   267 +0100000110100000000100000
   272 +1001011000000100000000000
   277 +0110001000000100100000000
   283 +1100100000100000000001000
   284 +1100010100010000000000000
   290 +1110001000010000000000000
   293 +1110010000010000000000000
   294 +0100010110100000000000000
   296 +1001000001000110000000000
   317 +1000000011010000010000000
   320 +1100010001000000010000000
   324 +0001010001001000100000000
   325 +0100010000100100000100000
   327 +1010001010000010000000000
   328 +1010010000000000100000001
   341 +1110000100100000000000000
   344 +1100100000000100100000000
   357 +1100000001001001000000000
   358 +1000000101000100001000000
   360 +1000000000100000000100000 11
   369 +0100100000010100000100000
   389 +1000000000101000001001000
   397 +1111000000000000000010000
   401 +0000110100110000000000000
   403 +1110000000010000100000000
   406 +0000000001001100000101000
   407 +0100110000001000000100000
   409 +0000000110000100110000000
   421 +1100001000110000000000000
   423 +1100001001100000000000000
   424 +0101101010000000000000000
   428 +0000000111001100000000000
   433 +1100000100000000001100000
   448 +0110000010000010010000000
   450 +0000000000000001100010110
   454 +1100001100100000000000000
   457 +0100010011000000000010000
   458 +0001000011001010000000000
   485 +1000110001000001000000000
   486 +1000000100100000000101000
   487 +1000010001000000100000100
   497 +0110100011000000000000000
    12 +1100000000110000000000000
    35 +1001010100000000000000000
```

```
GUTTMAN SCALOGRAM OF RESPONSES:
ASPIRA|PROBLE
      |1 2111  121 12212 12 1
      |14136832755890423602741 95
      |--------------------------
   54 +00010100000000000010010000
   60 +10000000100001000000001000
   88 +10010001000100000000000000
   90 +01010000000100000010000000
   94 +01000000010000000000100100
   96 +11000100001000000000000000
  100 +00010001010010000000000000
  102 +11000100001000000000000000
  108 +01001100000000010000000000
  109 +10000000110000000010000000
  115 +00011010000000000000001000
  122 +01000001001100000000000000
  123 +01000000000001001000000100
  141 +00110001000000000010000000
  147 +00100001000000001000100000
  161 +00000100011000100000000000
  163 +00001010101000000000000000
  165 +10001000001000100000000000
  170 +00000010000001000100001000
  173 +01001001000000000001000000
  179 +00000000100010100010000000
  181 +01000001010001000000000000
  187 +00010010010000000000010000
  195 +01010000010000010000000000
  199 +00010100110000000000000000
  203 +01100000000000000110000000
  205 +00000000000001001000010010
  223 +00010100000000100000000001
  224 +00000001011000000000001000
  225 +01001000100000000010000000
  227 +01001000100010000000000000
  228 +11000101000000000000000000
  231 +11000000001100000000000000
  232 +11110000000000000000000000
  236 +11000000001100000000000000
  239 +00000100101000010000000000
  243 +00100000100000000000010010
  248 +00100100100000010000000000
  257 +11000000001100000000000000
  269 +11101000000000000000000000
  271 +10000010000000000000011000
  278 +11000000000100000001000000
  286 +01010000010100000000000000
  288 +11000100010000000000000000
  295 +11100000000100000000000000
  297 +01000001010100000000000000
  300 +11000000010100000000000000
  301 +01000001000100000001000000
  304 +11100000000000010000000000
  314 +10010000000000000100001000
  331 +11000000100000000000000100
  336 +00101001100000000000000000
  352 +10100000000000001000100000
  353 +01010000101000000000000000
  366 +10100000000000101000000000
  370 +01100100000100000000000000
```

```
GUTTMAN SCALOGRAM OF RESPONSES:
ASPIRA|PROBLE
      |1 2111  121 12212 12 1
      |1413683275589042360274195
      |-------------------------
   371 +0111000000000001000000000
   392 +0111000000000001000000000
   400 +0100000100000001000100000
   405 +0000100110000000100000000
   413 +1000001110000000000000000
   414 +0110000000100000000000100
   417 +0110001000010000000000000
   418 +1010010000000100000000000
   419 +0010010010000000000000001
   422 +1010100001000000000000000
   425 +0000001011100000000000000
   427 +1110100000000000000000000
   429 +1000000000010001000100000
   441 +0001000010000100000010000
   449 +0001000000100000000101000
   460 +1100110000000000000000000
   461 +0000101000100001000000000
   475 +1110000001000000000000000
   476 +0001100001010000000000000
   490 +1001010000000000100000000
   493 +0100000010100000010000000
   494 +1100000100000000001000000
   495 +1100000000010010000000000
     8 +0001000100000000001000000
    40 +0000001100100000000000000
    73 +1000000110000000000000000
    76 +0000001000000000000100010
    83 +1000100000010000000000000
    93 +1100000100000000000000000
    95 +1000100010000000000000000
   104 +0000100000000100000100000
   105 +0000001000010000010000000
   106 +0010000001000001000000000
   121 +0001010000000001000000000
   138 +1100000000000000100000000
   150 +0100000000001001000000000
   155 +1010000001000000000000000
   156 +0000000100000001000001000
   175 +0001010000000000000001000
   194 +0010000001100000000000000
   214 +0000101000100000000000000
   230 +0001010000000000001000000
   233 +1000000001000000010000000
   240 +1000000000001000000010000
   244 +0110001000000000000000000
   259 +0000001000001000000010000
   262 +0000010001000000000000010
   276 +0010000000100000100000000
   282 +0100100000010000000000000
   289 +0001000100000000000010000
   298 +1000100000000010000000000
   303 +0100001000001000000000000
   306 +1000110000000000000000000
   312 +1010010000000000000000000
   322 +0100000000001100000000000
   323 +0000010010000000010000000
```

```
GUTTMAN SCALOGRAM OF RESPONSES:
ASPIRA|PROBLE
      |1 2111  121 12212 12 1
      |1413683275589042360274195
      |-------------------------
  326 +0000000001000100001000000
  338 +1100000000000000000010000
  342 +0000000001010000000001000
  345 +1110000000000000000000000
  361 +1000010000000100000000000
  368 +0100000000110000000000000
  374 +0000000000101001000000000
  385 +1000000001000001000000000
  387 +1000000001001000000000000
  391 +0001000000000001000000010
  396 +0011000000000001000000000
  398 +0100000000001000001000000
  442 +1100000001000000000000000
  451 +1100000100000000000000000
  455 +1100000100000000000000000
  459 +0101010000000000000000000
  464 +0001000000001000000000010
  466 +1000000100000000010000000
  473 +0100010000010000000000000
  481 +0110000000000000010000000
  483 +0001000001000000000000001
   11 +0000000001000010000000000
   13 +0000000000000000010100000
   25 +0100100000000000000000000
   29 +0100000010000000000000000
   34 +0000100010000000000000000
   36 +0000010000000100000000000
   59 +0000000000100000000010000
  128 +0000001000001000000000000
  143 +1000000000000100000000000
  168 +1100000000000000000000000
  190 +0001000100000000000000000
  198 +0001000000100000000000000
  201 +0000100000000000010000000
  258 +0010000100000000000000000
  260 +1000000000000100000000000
  263 +0000000010000100000000000
  285 +1000010000000000000000000
  292 +1000100000000000000000000
  363 +1000000000000000001000000
  367 +0100000100000000000000000
  380 +0000010010000000000000000
  384 +0010000100000000000000000
  480 +0100000000000001000000000
  482 +0000000010000000100000000
  484 +1000100000000000000000000
   19 +0000000000000001000000000
   51 +0000010000000000000000000
   71 +0000000000001000000000000
  213 +0100000000000000000000000
  253 +0100000000000000000000000
  256 +0000000100000000000000000
  463 +0000100000000000000000000
  126 +0000000000000000000000000
      |-------------------------
```