

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE GRADUADOS



Modelo de Rasch dicotómico con aplicación a la Educación

TESIS PARA OPTAR POR EL GRADO DE MAGISTER EN
ESTADÍSTICA

Presentado por:

Omar Antonio Chincaro Del Coral

Asesor: Dr. Jorge Luis Bazán Guzmán

Miembros del jurado:

Dr. Luis Hilmar Valdivieso Serrano

Dra. Mery Elizabeth Doig Camino

Dr. Jorge Luis Bazán Guzmán

Lima, Mayo 2010

Dedicatoria

Dedico la presente tesis a la Pontificia Universidad Católica del Perú (PUCP) y a mis padres, principalmente a mi madre quien constantemente me estuvo apoyando en la elaboración de mi tesis.



Agradecimientos

- Agradezco a mi asesor el Dr. Jorge Luis Bazán quien me guió en el aspecto metodológico de los Modelos Rasch.
- Agradezco a los profesores de la Maestría de Estadística de la Pontificia Universidad Católica del Perú (PUCP), quienes me brindaron el soporte teórico / práctico para el desarrollo del modelo.



Resumen

En investigaciones de origen cuantitativo generalmente se emplean instrumentos de medición que generan base de datos dicotómicas, en la cual cada persona responde las preguntas o ítems del instrumento. Subyacente a estas respuestas existen variables no observables o variables latentes que caracterizan a las personas evaluadas y a los ítems del instrumento de medición utilizado. En este trabajo se modeló la probabilidad de responder correctamente al ítem en función a sus parámetros mediante el uso de los modelos logísticos o modelos de Rasch. Considerando las respuestas a estas variables latentes de las personas, de los ítems, y sus supuestos se estimó los parámetros a partir de la función de verosimilitud del modelo. En esta tesis se mostró diferentes métodos de estimación como el de Máxima Verosimilitud Marginal (MVM) que depende de las puntuaciones que se obtenga en cada ítem, el de Máxima Verosimilitud Condicional (MVCOND) que considerara patrones de respuesta, el método de Máxima Verosimilitud Conjunta (MVC) y el método Bayesiano utilizando Cadenas de Markov y métodos de Monte Carlo (MCMC) como el algoritmo Gibbs Sampling. El Método Bayesiano fue analizado bajo dos esquemas: adaptative rejection sampling (ARS) y el data argumentation gibbs sampling (DAGS). Con estos métodos se estimaron los parámetros de los ítems y las personas evaluadas, los cuales se compararon con estudios de simulación determinándose que el mejor método de estimación es el Bayesiano. El método bayesiano presenta las estimativas más precisas considerando diferentes escenarios de tamaño de muestra y número de ítems frente a los otros métodos de estimación. Adicionalmente no tiene restricción en la estimación frente a valores extremos y finalmente es un método conjunto que estima al mismo tiempo habilidades y dificultades a diferencia de otros métodos que sólo estiman dificultades u otros que estiman ambos pero baja precisión. Finalmente se realizó una aplicación del modelo en el ámbito educacional.

Palabras-clave: Variable latentes, Modelos logísticos, Modelos Rasch, Función de Máxima Verosimilitud, Máxima Verosimilitud Marginal, Máxima Verosimilitud Condicional, Máxima Verosimilitud Conjunta, Patrones de respuesta, Método Bayesiano, Cadenas de Markov y Monte Carlo, Algoritmo Gibbs Sampling, Adaptative Rejection Sampling (ARS), Data Argumentation Gibbs Sampling (DAGS).

Índice general

Lista de Abreviaturas	VIII
Lista de Símbolos	IX
Índice de figuras	X
Índice de cuadros	XI
1. Introducción	1
1.1. Consideraciones Preliminares	1
1.2. Objetivos	2
1.3. Organización del Trabajo	2
2. El Modelo de Rasch	3
2.1. Historia	3
2.2. El Modelo de Rasch dicotómico	4
2.2.1. Definición del Modelo	4
2.2.2. Función de verosimilitud de los modelos de Rasch dicotómicos	7
2.2.3. Caracterización y propiedades de los modelos de Rasch dicotómicos	9
2.2.4. Información de Fisher del modelo de Rasch	11
3. Estimación del modelo de Rasch	15
3.1. Métodos de estimación clásica del modelo de Rasch dicotómico	15
3.1.1. Estimación por métodos de Máxima Verosimilitud Conjunta	15
3.1.2. Estimación por métodos de Máxima Verosimilitud Condicional	16
3.1.3. Estimación por métodos de Máxima Verosimilitud Marginal	18
3.2. Métodos de estimación bayesiana para el modelo de Rasch dicotómico	20
3.2.1. MCMC usando adaptive rejection sampling (ARS)	21
3.2.2. MCMC usando data argumentation Gibbs sampling (DAGS)	22
4. Estudio de Simulación	24
4.1. Condiciones de la simulación	24
4.1.1. Algoritmo para generar una data dicotómica	24
4.1.2. Criterios para evaluar la simulación	25
4.2. Métodos de estimación a estudiar	25

4.3. Escenarios de Estudio	25
4.3.1. Número pequeño de ítems y tamaño de muestra variable	26
4.3.2. Número grande de ítems y tamaño de muestra variable	28
4.3.3. Conclusión del estudio de simulación	29
5. Aplicación	30
5.1. Descripción del problema de investigación	30
5.2. Tipo de instrumento de medición	30
5.3. Diseño muestral	31
5.4. Recolección de información y base de datos	33
5.5. Análisis clásico de la prueba	33
5.5.1. Proporción de aciertos	33
5.5.2. Índice de discriminación	33
5.5.3. Confiabilidad de la prueba	34
5.6. Estimación de parámetros bajo el modelo de Rasch	37
5.6.1. Estimación de las dificultades de ítems desde un punto de vista clásico	37
5.6.2. Estimación de las dificultades de ítems desde un punto de vista bayesiano bajo el esquema DAGS	38
5.6.3. Estimación de las habilidades de los estudiantes desde un punto de vista bayesiano bajo el esquema DAGS	44
5.6.4. Función de información y de la prueba para la aplicación	48
6. Conclusiones	49
6.1. Conclusiones	49
6.2. Sugerencias para investigaciones futuras	50
A. Anexos A	52
A.1. Suficiencia estadística	52
A.1.1. Teorema de Factorización de Neyman	52
A.2. Formulas de cuadratura de Gauss Hermite	54
A.2.1. Polinomios de Hermite	54
A.2.2. Relaciones entre los polinomios de Hermite	55
A.3. Métodos computaciones de aproximación: Newton Raphson	55
A.4. Método de Jackniffe	56
A.5. Algoritmo Gibbs Sampling	56
B. Anexos B	59
B.1. Programa para generar una base de datos dicotómica	59
B.2. Programa para graficar una Curva Característica del Ítem, Función de Información y Mapa de ítems	60
B.3. Programa Métodos de estimación	61
B.3.1. Métodos de estimación Máxima Verosimilitud Marginal y Condicional	61
B.3.2. Métodos de estimación Máxima Verosimilitud Conjunta	62
B.3.3. Métodos de estimación Bayesiana bajo el esquema ARS	63

B.3.4. Métodos de estimación Bayesiana bajo el esquema DAGS	64
Bibliografía	65



Lista de Abreviaturas

MVC	Máxima Verosimilitud Conjunta.
MVM	Máxima Verosimilitud Marginal.
MVCOND	Máxima Verosimilitud Condicional.
MB	Método Bayesiano.
MSE	Cuadrado Medio del error.
RSME	Raíz cuadrada del Cuadrado Medio del error.
CORR	Correlación.
GL	Grados de Libertad.
F	Estadístico F de Fisher.
T	Estadístico T student.
SIG	Nivel de significancia.
ID	Indice de dificultad.

Lista de Símbolos

β	Dificultad de los ítems.
θ	Habilidad de los estudiantes.
r_i	Puntuaciones obtenidas por la i-ésima persona evaluada.
Z_j	Puntuaciones para el j-ésimo ítem obtenidas por todas las personas evaluadas.
ρ_{pj}	Índice de discriminación.
μ_a	Promedio de las puntuaciones cuando el estudian ha acertado al ítem.
μ_x	Promedio de las puntuaciones obtenidas por los estudiantes.
σ_x	Promedio de las puntuaciones obtenidas por los estudiantes.
p_j	Proporción de aciertos para ítem j.
q_j	Proporción de fallas para ítem j.
α	Coficiente de confiabilidad: Alfa de Cronbach.
σ_i^2	Varianza de los ítems.
σ_x^2	Varianza de los puntajes totales.

Índice de figuras

2.1. Curva Característica del Ítem	6
2.2. Curva Logística o Curva Característica del Ítem	10
2.3. Función de información del ítem	13
2.4. Función de información de la prueba	14
5.1. Histograma de puntajes	36
5.2. Puntajes promedios por Región Natural	37
5.3. Dificultad de ítems por el MB	39
5.4. CCI con habilidad Rasch Bayesiana	40
5.5. Curvas características de ítems 10 y 12	42
5.6. Mapa de ítems	43
5.7. Asociación entre la dificultad Rasch y la proporción de aciertos	44
5.8. Habilidad promedio de los estudiantes por el MB según Regiones Naturales	45
5.9. Habilidad de los estudiantes por el MB según regiones naturales	46
5.10. Mapa de ítems por Regiones Naturales	47
5.11. Funciones de información de los ítems	48
5.12. Funciones de información de la prueba	48
B.1. Programa para generar una base de datos dicotómica	59
B.2. Programa Máxima Verosimilitud Marginal y Condicional	60
B.3. Programa Máxima Verosimilitud Marginal y Condicional	61
B.4. Programa Máxima Verosimilitud Conjunta	62
B.5. Programa estimación Bayesiana bajo el esquema ARS	63
B.6. Programa estimación Bayesiana bajo el esquema DAGS	64

Índice de cuadros

2.1. Resumen de la función de información de la prueba	14
4.1. Estimaciones de los parámetros de los ítems para el escenario 1 y 2	27
4.2. Medidas de ajustes de los ítems para el escenario 1 y 2	27
4.3. Medidas de ajustes de las personas para el escenario 1 y 2	28
4.4. Medidas de ajustes de los ítems para el escenario 3 y 4	28
4.5. Medidas de ajustes de las personas para el escenario 3 y 4	29
5.1. Distribución de los instrumentos de medición	31
5.2. Distribución ítems en los textos del instrumentos de medición	31
5.3. Tamaño de muestra y errores muestrales por región natural	32
5.4. Proporción de acierto y discriminación de los ítems	34
5.5. Distribución de puntuaciones de los evaluados	35
5.6. Estadísticos de la puntuación total	36
5.7. Puntuaciones promedio de los evaluados por Región Natural	36
5.8. Estimación clásica de las dificultades de los ítems bajo el modelo de Rasch . .	38
5.9. Estimación bayesiana de las dificultades de los ítems bajo el modelo de Rasch	39
5.10. Prueba de homogeneidad de varianza de las habilidades de los estudiantes según regiones naturales	45
5.11. ANOVA de la habilidades de los estudiantes según regiones naturales	45
5.12. Comparaciones múltiples de las habilidades de los estudiantes entre las re- giones naturales	46

Capítulo 1

Introducción

1.1. Consideraciones Preliminares

En investigaciones de origen cuantitativo generalmente se emplean instrumentos de medición los cuales generan bases de datos dicotómicas, donde los sujetos evaluados responden a preguntas o ítems del instrumento de medición, asociados a estas respuestas se presentan variables latentes que son características de los individuos las cuales denotaremos como θ_i donde i denota al i -ésimo del sujeto; y de las preguntas del instrumento de medición que denotaremos como β_j donde j es la j -ésima pregunta.

Tal como proponen Bartholomew y Knoot (1999), existen diferentes tipos de modelos de variables latentes, dependiendo si las variables latentes del modelo son categóricas o continuas. Dos de los más importantes modelos de rasgos latentes en orden histórico son la teoría clásica de los test (TCT) (Lord y Baker, 1968) y la teoría de respuesta al ítem (TRI) (Baker, 1992).

Los modelos de Teoría de Respuesta al Ítem (TRI), se ocupan del análisis estadístico de los datos obtenidos de las respuestas a las preguntas del instrumento de medición por parte de los sujetos evaluados, en categorías que son mutuamente excluyentes (Back y Moustataki). Históricamente estos modelos han sido desarrollados en el ámbito de la Psicometría, pero actualmente tienen un creciente desarrollo en las áreas de Estadística, Computación y Probabilidad debido a la importante cantidad de problemas que se plantean con el desarrollo y aplicaciones de esta clase de modelos. La importancia del estudio de modelos TRI subyace también en sus múltiples aplicaciones crecientes en campos tan diversos como la Educación, la Psicología, la Política, la Economía, la Sociología, el Marketing, la Ingeniería Genética entre otras.

La Teoría de Respuesta al Ítem que estudiamos en este trabajo, considera que cada individuo lleva asociado un parámetro individual y un parámetro a cada pregunta (Kohan, 1999) θ_i y β_j respectivamente. La ventaja de estos modelos es que dan la posibilidad de tener mediciones invariantes respecto de los instrumentos de medición utilizados y las personas evaluadas (Muñiz, 1977).

Existe una gama de modelos de Teoría de respuesta al Ítem (TRI), nosotros nos enfocaremos en los Modelos TRI de un parámetro que según la literatura son llamados Modelos de Rasch (Bond y Fox, 2005). En el Perú estos modelos son usados por el Ministerio de Educación,

para reportar el rendimiento escolar en pruebas nacionales e internacionales.

La aplicación, en este caso es en el ámbito educacional, se utilizan como instrumentos de medición a pruebas, las cuales contienen ítems. A la variable latente que caracteriza a los estudiantes se le llama habilidad y a la que caracteriza a los ítems se le llama dificultad del ítem.

1.2. Objetivos

El objetivo general de la tesis es presentar los modelos de Rasch dicotómicos incluyendo su racionalidad, principales propiedades y métodos de estimación así como el uso de estudios de simulación y la aplicación a datos reales. De manera específica:

- Presentar y derivar las propiedades del modelo de Rasch dicotómico.
- Presentar la estimación de los modelos de Rasch dicotómico desde la perspectiva clásica y bayesiana.
- Realizar estudios de simulación acerca del modelo propuesto en diferentes escenarios y comparando los métodos de estimación propuestos en la tesis.
- Aplicar el modelo a un conjunto de datos reales en el ámbito de la Educación, en la cual se busca determinar las habilidades de los estudiantes con el mejor método de estimación.

1.3. Organización del Trabajo

En el Capítulo 2, presentamos conceptos previos a los modelos de Rasch dicotómico, definiciones del modelo, caracterizaciones y propiedades.

En el Capítulo 3, presentamos los diferentes métodos de estimación de las variables latentes desde la perspectiva clásica y bayesiana del modelo de Rasch.

En el Capítulo 4, presentamos simulaciones bajo escenarios diferentes en la cual analizaremos las estimaciones de las variables latentes tanto desde el punto de vista clásico como bayesiano.

En el Capítulo 5, presentamos las aplicaciones del modelo de Rasch dicotómico en el ámbito de la Educación.

En el Capítulo 6 discutimos algunas conclusiones obtenidas en este trabajo. Analizamos las ventajas y desventajas de los métodos propuestos, y sugerencias para investigaciones futuras.

En el Anexo A, presentamos demostraciones y pruebas con más detalles.

Finalmente en el Anexo B presentamos los programas utilizados en las aplicaciones a conjuntos de datos reales.

Capítulo 2

El Modelo de Rasch

Se diseña toda encuesta o prueba con la finalidad de establecer inferencias sobre los rasgos de las personas (no observables) basándose en la información que manifiestan en sus respuestas; en la literatura estos rasgos son llamados variables latentes (Fisher, 1974). Para identificar estas variables latentes se han propuestos diversos modelos estadísticos. En este estudio nos centraremos en los Modelos de Rasch dicotómicos en los que se modela con la función logística la probabilidad de responder correctamente a un ítem considerando el parámetro de las personas y el parámetro de los ítems.

Estos modelos centran su análisis de las preguntas del instrumento de medición, específicamente de la interacción entre una persona y un ítem. Establecen la probabilidad de respuesta de una persona ante un ítem, en función de las variables latentes que se estiman; es decir, los parámetros de las personas θ y los ítems β .

2.1. Historia

George Rasch (1901 - 1980) fue un matemático danés, estadístico, y psicometrista, famoso por el desarrollo de una clase de medición, los modelos conocidos como modelos de Rasch. Estudió con R.A. Fisher y también brevemente con Ragnar Frisch, fue elegido miembro de la Instituto Internacional de Estadística en 1948.

En 1919, Rasch, comenzó a estudiar matemáticas en la Universidad de Copenhague. Completó una maestría en 1925 y recibió un doctorado en ciencias matemáticas en 1930. Rasch se casó en 1928. Incapaz de encontrar trabajo como matemático en la década de 1930, se dedicó a trabajar como consultor de estadística. En este cargo, trabajó en una serie de problemas, incluyendo problemas de crecimiento biológico.

Después de Segunda Guerra Mundial, Rasch se involucró en psicometría, como consultor de proyectos del Grupo de Psicología Militar del Ministerio de Defensa Danés y del Ministerio de Bienestar Social Danés. El punto de partida para el desarrollo de sus modelos nace en 1952 cuando tuvo que analizar datos de seguimiento basado en la lectura de registros.

Rasch introduce el Modelo de Poisson Multiplicativo, donde el número de palabras de un texto T_j leído por el evaluado S_i en un periodo fijo, se denota α_{ij} como el número de errores leídos en el texto j por el i -ésimo sujeto. T_j fue considerado como una variable de Poisson. En este caso el parámetro λ_{ij} representa la velocidad de lectura del texto por parte del evaluado.

Para explicar el rendimiento de lectura en términos del individuo y la propiedad del texto, Rasch clasificó λ_{ij} en factores de la persona ξ_i , la habilidad de lectura del evaluado como factor ϵ_j , obteniendo el siguiente modelo:

$$P(\alpha_{ij}) = \frac{\lambda_{ij}^{\alpha_{ij}}}{\alpha_{ij}!} e^{-\lambda_{ij}} = \frac{(\xi_i \epsilon_j)^{\alpha_{ij}}}{\alpha_{ij}!} e^{-\xi_i \epsilon_j} \quad (2.1)$$

Análogamente, si la misma persona tiene que leer un nuevo texto T_k , bajo otras condiciones similares y las respuestas de ambas pruebas son estocásticamente independientes, la probabilidad de que el evaluado responda α_{ik} palabras del del texto es:

$$P(\alpha_{ik}) = \frac{(\xi_i \epsilon_k)^{\alpha_{ik}}}{\alpha_{ik}!} e^{-\xi_i \epsilon_k} \quad (2.2)$$

Cuando Rasch en 1952/1953 llegó a analizar los datos de prueba de inteligencia para el Grupo de Psicología Militar, trató de llevar a los mismos principios en el área de análisis de ítems. En primer lugar, buscó una adecuada función de respuesta al ítem (IRF), en la cual modelaba la probabilidad de responder correctamente en función a las características del evaluado y del ítem. Considerando el Modelo Multiplicativo de Poisson y haciendo reparametrizaciones en este modelo logró modelar la probabilidad de responder positivamente al ítem, denominando a la características de los evaluados: *habilidad* y a la característica del ítem: *dificultad*.

2.2. El Modelo de Rasch dicotómico

Los Modelos de Rasch dicotómicos modelan la probabilidad de responder correctamente al ítem, en la cual se estiman las variables latentes como la habilidad y la dificultad de los ítems del instrumento de medición (Fischer, G. H.(2007)).

2.2.1. Definición del Modelo

Para definir el Modelo de Rasch dicotómico presentamos la data:

<i>Preguntas</i>						
<i>Personas</i>	I_1	I_2	I_3	.	.	I_k
P_1	1	1	0	.	.	1
P_2	1	0	1	.	.	0
P_3	0	1	1	.	.	1
.
.
.
P_n	1	1	1	.	.	1

Donde las filas son las personas evaluadas y las columnas son las preguntas del instrumento de medición. Los datos toman valores de 0 (en caso de responder negativamente a la pregunta) y 1 (en caso de responder positivamente a la pregunta) y sigue una distribución de Bernoulli:

Sea θ_i la característica del evaluado (variable latente), β_j la característica del ítem (variable latente), Y_{ij} respuesta observada del i - ésimo evaluado para j - ésimo ítem que toma valores de 1 (acertar a la pregunta) o 0 (fallar a la pregunta) y p_j probabilidad de acertar al j - ésimo ítem. El Modelo Rasch queda definido de la siguiente forma:

$$Y_{ij} \sim \text{Bernoulli}(p_{ij}) \tag{2.3}$$

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \theta_i - \beta_j \tag{2.4}$$

o equivalentemente considerando

$$\frac{p_{ij}}{1 - p_{ij}} = e^{(\theta_i - \beta_j)}$$

La ecuación (2.4) puede ser reemplazado por:

$$p_{ij} = P(Y_{ij}|\theta_i, \beta_j) = \frac{e^{[y_{ij}(\theta_i - \beta_j)]}}{1 + e^{(\theta_i - \beta_j)}} \tag{2.5}$$

Donde los parámetros para las personas $\theta_i \in \mathbb{R}$ y los ítems $\beta_j \in \mathbb{R}$ para todo $i = 1 \dots n$ y $j = 1 \dots k$.

La probabilidad de responder correctamente al ítem es:

$$P(Y_{ij} = 1|\theta_i, \beta_j) = \frac{e^{(\theta_i - \beta_j)}}{1 + e^{(\theta_i - \beta_j)}} \tag{2.6}$$

y la probabilidad de fallar o responder incorrectamente al ítem es:

$$P(Y_{ij} = 0|\theta_i, \beta_j) = \frac{1}{1 + e^{(\theta_i - \beta_j)}} \tag{2.7}$$

Si en el modelo el parámetro de la persona evaluada θ_i es menor al parámetro del ítem β_j decimos que la probabilidad de responder correctamente al ítem $P(Y_{ij} = 1|\theta_i, \beta_j)$ es menor a 0.5 con lo cual se induce que existe una chance menor al 50% de acertar al j - ésimo ítem y es considerado algo difícil para la i - ésima persona evaluada. Y si el parámetro de la persona evaluada θ_i es mayor al el parámetro del ítem β_j decimos que la probabilidad de responder correctamente al ítem $P(Y_{ij} = 1|\theta_i, \beta_j)$ es mayor a 0.5 con lo cual se induce que existe una chance mayor al 50% de acertar al j - ésimo ítem y ha sido algo fácil para la i - ésima persona evaluada.

Curva característica del ítem

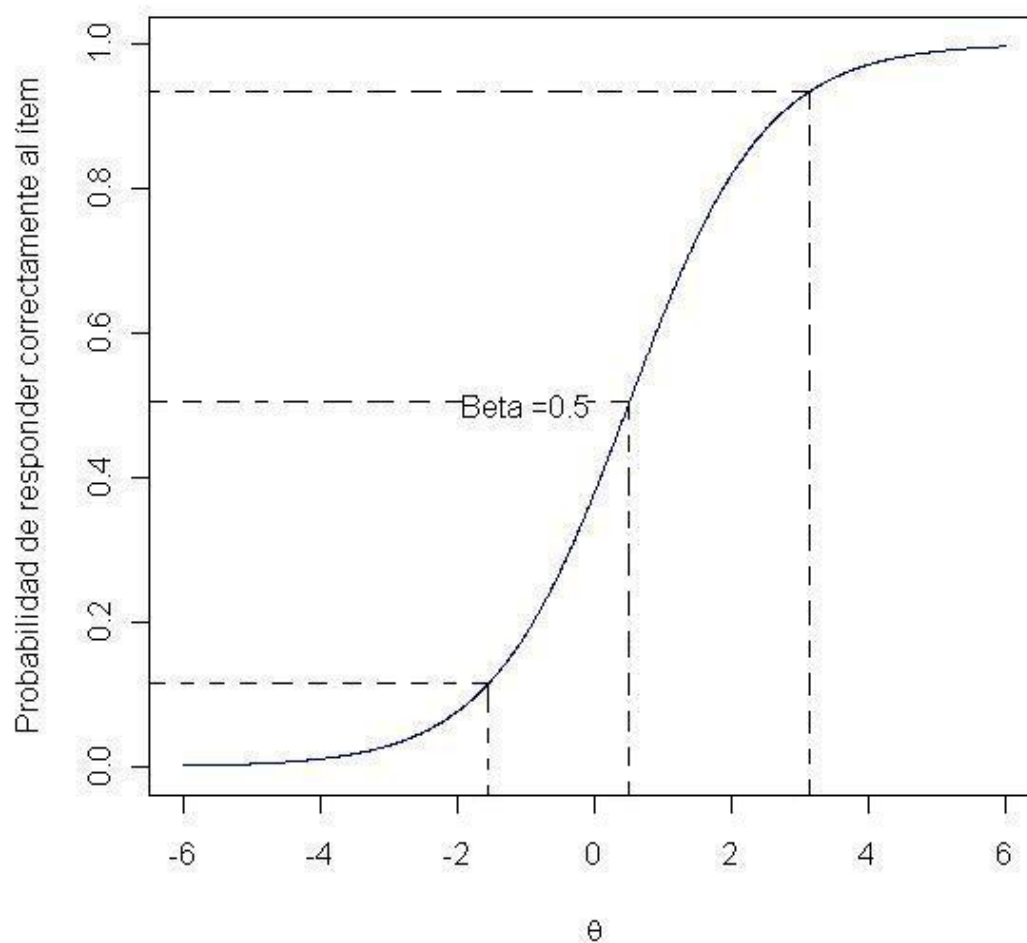


Figura 2.1: Curva Característica del Ítem

De la figura 2.1 Considerando $\beta = 0,5$, observamos que conforme el parámetro de la persona evaluada aumente entonces aumenta la probabilidad de responder correctamente el ítem.

Supuestos del modelo de Rasch

1. El modelo de Rasch planteado en la ecuación (2.5) explica la ocurrencia de la matriz de datos que contiene respuestas dicotómicas de una muestra de n personas y k ítems que mide la misma variable latente θ . Asume que las personas tienen un valor real denotado por θ_i el cual representa su posición sobre la variable latente. Cada ítem tiene un valor real que son los parámetros de los ítems β_j (Fisher, 2007). De esta manera cada uno de estos parámetros son diferentes y únicos, aunque pueden ser confrontados en la misma escala asumiendo que ambos toman valores en toda la recta. Podemos decir entonces que el modelo es unidimensional.
2. A partir de la curva característica del ítem observamos que es una función no decreciente de $\theta \in \mathbb{R}$. Entonces se asume que el modelo es monótono.
3. Para valores fijo del parámetro θ todas las respuestas son estocasticamente independientes, así la probabilidad de responder a los k ítems del instrumento de medición es el producto de la k probabilidades. Dicho de otra manera:

Si θ_i fijo e $Y_i = (Y_{i1}, Y_{i2}, Y_{i3}, \dots, Y_{ik})$ entonces

$$P(Y_i|\theta_i, \beta_j) = P(Y_{i1}|\theta_i, \beta_j) \times \dots \times P(Y_{ik}|\theta_i, \beta_j)$$

$$P(Y_i|\theta_i, \beta_j) = \prod_{j=1}^k F(\theta_i - \beta_j)^{y_{ij}} [1 - F(\theta_i - \beta_j)]^{1-y_{ij}}$$

Donde $F(\cdot)$ es la función de distribución acumulada, el cual lo reemplazamos por la función logística tenemos entonces:

$$P(Y_i|\theta_i, \beta) = \prod_{j=1}^k P(y_{ij}|\theta_i, \beta_j) = \prod_{j=1}^k \frac{e^{[y_{ij}(\theta_i - \beta_j)]}}{1 + e^{(\theta_i - \beta_j)}} \quad (2.8)$$

En psicometría esto se conoce como independencia local.

2.2.2. Función de verosimilitud de los modelos de Rasch dicotómicos

Considerando el supuesto de independencia de los ítems o en la literatura psicométrica llamado independencia local, planteamos la función de verosimilitud del modelo de Rasch dicotómico para la i -ésima persona evaluada (Baker, 1992):

Sea $\beta = (\beta_1, \beta_2, \beta_3, \dots, \beta_k)$

$$L(Y_i|\theta_i, \beta) = \prod_{j=1}^k [F_Y(\theta_i, \beta_j)]^{y_{ij}} [1 - F_Y(\theta_i, \beta_j)]^{1-y_{ij}}$$

Reemplazando por la función logística:

$$L(Y_i|\theta_i, \beta) = \prod_{j=1}^k \left[\frac{e^{(\theta_i - \beta_j)}}{1 + e^{(\theta_i - \beta_j)}} \right]^{y_{ij}} \left[\frac{1}{1 + e^{(\theta_i - \beta_j)}} \right]^{1 - y_{ij}}$$

Resolviendo:

$$L(Y_i|\theta_i, \beta) = \prod_{j=1}^k \frac{[e^{(\theta_i - \beta_j)}]^{y_{ij}}}{1 + e^{(\theta_i - \beta_j)}}$$

$$L(Y_i|\theta_i, \beta) = \prod_{j=1}^k P(Y_{ij} = y_{ij}|\theta_i, \beta_j)$$

$$L(Y_i|\theta_i, \beta) = \frac{e^{\sum_{j=1}^k y_{ij}(\theta_i - \beta_j)}}{\prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]}$$

$$L(Y_i|\theta_i, \beta) = \frac{e^{\theta_i \sum_{j=1}^k y_{ij} - \sum_{j=1}^k y_{ij} \beta_j}}{\prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]}$$

De los datos mostrados en la tabla (2.2.1) obtenemos las puntuaciones obtenidas por la i -ésima persona evaluada:

$$\Gamma_i = \sum_{j=1}^k y_{ij} \tag{2.9}$$

y las puntuaciones para el j -ésimo ítem obtenidas por todas las personas evaluadas:

$$Z_j = \sum_{i=1}^n y_{ij} \tag{2.10}$$

Entonces la función de verosimilitud para la i -ésima persona evaluada queda expresada de la siguiente forma:

$$L(Y_i|\theta_i, \beta) = \frac{e^{\theta_i \Gamma_i - \sum_{j=1}^k y_{ij} \beta_j}}{\prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]} \tag{2.11}$$

Finalmente la función de verosimilitud para todas las personas evaluadas con todos los ítems del instrumento de medición queda expresado de la siguiente forma:

Sea \mathbf{Y} la matriz de orden $n \times k$ que contiene las respuestas de las n personas evaluadas con los k ítems del instrumento de medición, $\theta = (\theta_1, \theta_2, \theta_3, \dots, \theta_n)$ los parámetros de las n personas evaluadas y $\beta = (\beta_1, \beta_2, \beta_3, \dots, \beta_k)$ los parámetros de los k ítems del instrumento de medición

$$L(\mathbf{Y}|\theta, \beta) = \prod_{i=1}^n L(Y_i|\theta_i, \beta)$$

$$L(\mathbf{Y}|\theta, \beta) = \frac{e^{\sum_{i=1}^n \theta_i \Gamma_i - \sum_{i=1}^n \sum_{j=1}^k y_{ij} \beta_j}}{\prod_{i=1}^n \prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]}$$

$$L(\mathbf{Y}|\theta, \beta) = \frac{e^{\sum_{i=1}^n \theta_i \Gamma_i - \sum_{j=1}^k Z_j \beta_j}}{\prod_{i=1}^n \prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]} \quad (2.12)$$

Otra forma de expresar la función de verosimilitud es a través de su logaritmo neperiano:

$$l(\mathbf{Y}|\theta, \beta) = \ln[L(\mathbf{Y}|\theta, \beta)]$$

$$l(\mathbf{Y}|\theta, \beta) = \sum_{i=1}^n \theta_i \Gamma_i - \sum_{j=1}^k Z_j \beta_j - \ln \prod_{i=1}^n \prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]$$

$$l(\mathbf{Y}|\theta, \beta) = \sum_{i=1}^n \theta_i \Gamma_i - \sum_{j=1}^k Z_j \beta_j - \sum_{i=1}^n \sum_{j=1}^k \ln [1 + e^{(\theta_i - \beta_j)}] \quad (2.13)$$

2.2.3. Caracterización y propiedades de los modelos de Rasch dicotómicos

A partir de la ecuación (2.6), graficamos la curva, dicha curva en la literatura psicométrica se denominada *Curva Característica del Ítem* (Fischer, 2007):

Curva característica del ítem

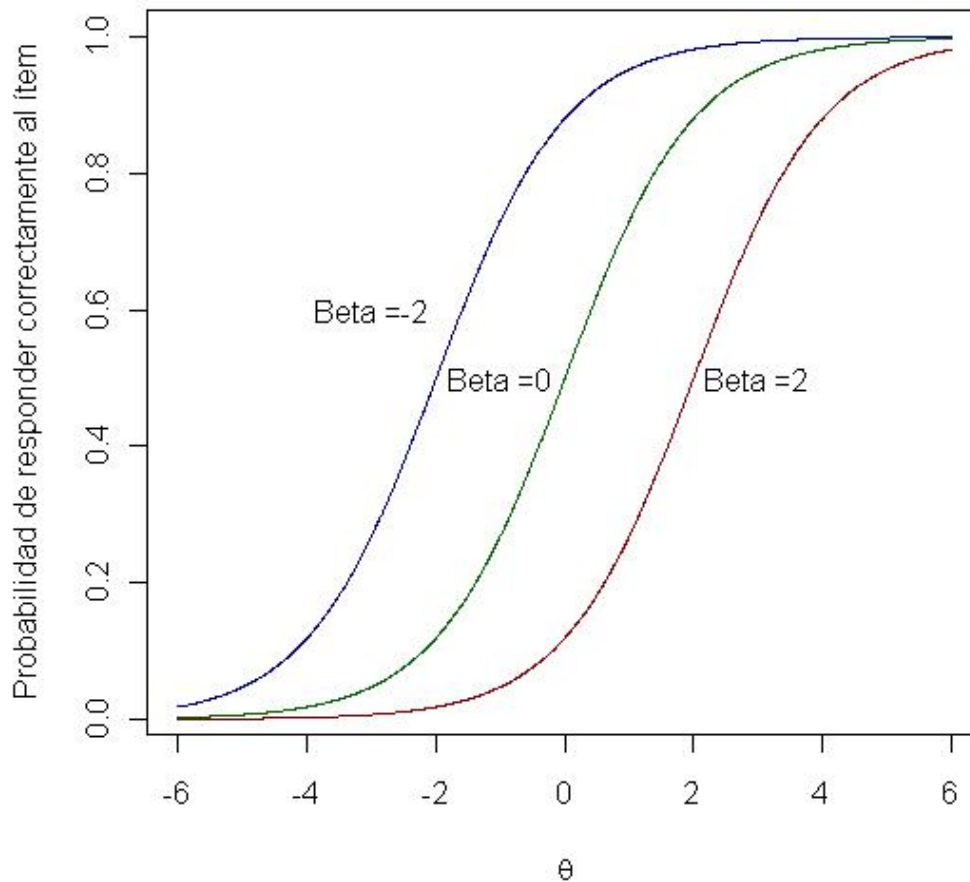


Figura 2.2: Curva Logística o Curva Característica del Ítem

De la figura 2.2 se pueden obtener la siguientes propiedades:

1. La función $f_i(\theta) = P(Y_{ij} = 1|\theta_i, \beta_j)$ es una función creciente. Formalmente, del gráfico si: $\theta_1 < \theta_2$ entonces $f_i(\theta_1) < f_i(\theta_2)$.
2. El $\lim_{\theta \rightarrow -\infty} f_i(\theta) = 0$ y el $\lim_{\theta \rightarrow \infty} f_i(\theta) = 1$
3. Como $f_i(\theta)$ es una función de probabilidad entonces: $0 \leq P(Y_{ij} = 1|\theta_i, \beta_j) \leq 1$ y $0 \leq P(Y_{ij} = 0|\theta_i, \beta_j) \leq 1$.
4. Como $f_i(\theta)$ es una función de probabilidad entonces: $P(Y_{ij} = 0|\theta_i, \beta_j) + P(Y_{ij} = 1|\theta_i, \beta_j) = 1$.
5. A partir de la función de verosimilitud mostrada en la ecuación (2.12), por el teorema de Factorización de Neyman (ver Anexo A) mostramos que $(\Gamma_i = \sum_{j=1}^k y_{ij})$, la puntuación obtenida por cada estudiante evaluado, es una estadística suficiente de su parámetro θ_i .

6. Bajo una determinada habilidad, a mayor dificultad menor probabilidad de responder correctamente al ítem y viceversa.

2.2.4. Información de Fisher del modelo de Rasch

La información de Fisher es aquella en la cual el estimador alberga la mayor cantidad de información posible. La información de Fisher de $\theta \in \Theta$, se denota por $I(\theta)$ y es tal que:

$$I(\theta) = E\left[\frac{\partial}{\partial\theta} \ln(f(X, \theta))\right]^2$$

O equivalentemente:

$$I(\theta) = E\left[-\frac{\partial^2}{\partial\theta^2} \ln(f(X, \theta))\right]$$

Partiendo de la definición de la información de Fisher de un modelo estadístico, definimos la función de Información de un ítem:

Sea k el número de preguntas del instrumento de medición y sea $\theta \in \mathbb{R}$ el parámetro de las personas y sea $j = 1, 2, 3, \dots, k$ la identificación de los ítems. Se define la función de información de cada ítem ($I_j(\theta)$) como:

$$I_j(\theta) = E\left[-\frac{\partial^2 \ln(P_j(y, \theta))}{\partial\theta^2}\right] \quad (2.14)$$

donde:

$$P_j(y, \theta) = \frac{e^{y(\theta - \beta_j)}}{1 + e^{(\theta - \beta_j)}} \quad (2.15)$$

$P_j(y, \theta)$ es la probabilidad de responder bajo los parámetros de las personas y los ítems. $P_j(\theta)$ y $Q_j(\theta)$ es la probabilidad de responder correctamente y fallar al ítem, respectivamente.

$$P_j(\theta) = \frac{e^{(\theta - \beta_j)}}{1 + e^{(\theta - \beta_j)}}$$

$$Q_j(\theta) = \frac{1}{1 + e^{(\theta - \beta_j)}}$$

Tomando logaritmos neperiano a la ecuación (2.15):

$$\ln P_j(y, \theta) = y(\theta - \beta_j) - \ln[1 + e^{(\theta - \beta_j)}]$$

Tomando primera, segunda derivada y resolviendo:

$$\frac{\partial \ln P_j(y, \theta)}{\partial\theta} = y - \frac{e^{(\theta - \beta_j)}}{1 + e^{(\theta - \beta_j)}}$$

$$\frac{\partial^2 \ln P_j(y, \theta)}{\partial \theta} = -\frac{e^{(\theta-\beta_j)}}{[1 + e^{(\theta-\beta_j)}]^2}$$

$$-\frac{\partial^2 \ln P_j(y, \theta)}{\partial \theta} = \frac{e^{(\theta-\beta_j)}}{[1 + e^{(\theta-\beta_j)}]^2}$$

Se obtiene la información de Fisher del estimador de cada persona evaluada mediante:

$$I_j(\theta) = \frac{e^{(\theta-\beta_j)}}{[1 + e^{(\theta-\beta_j)}]^2} \quad (2.16)$$

$$I_j(\theta) = P_j(\theta)Q_j(\theta) \quad (2.17)$$

Teniendo en cuenta la información de Fisher, se define la función de información del test como:

$$I(\theta) = \sum_{j=1}^k I_j(\theta) \quad (2.18)$$

Remplazando con la información de Fisher del estimador de la persona evaluada:

$$I(\theta) = \sum_{j=1}^k P_j(\theta)Q_j(\theta) \quad (2.19)$$

Reemplazando con la probabilidad de responder correctamente e incorrectamente al ítem, obteniendo finalmente la información del test o de la prueba:

$$I(\theta) = \sum_{j=1}^k \frac{e^{(\theta-\beta_j)}}{[1 + e^{(\theta-\beta_j)}]^2} \quad (2.20)$$

Los estimadores de las personas evaluadas θ son asintóticamente insesgados cuando los instrumentos de medición muchos ítems (Fischer, G. H.(2007)).

$$Var(\theta) = \frac{1}{I(\theta)} + h(k^{-1}) \quad (2.21)$$

Donde $h(\cdot)$ es una función que depende de los k - ítems y representa al sesgo del estimador. Considerando como grande la cantidad de preguntas del instrumento de medición entonces:

$$Var(\theta) \approx \frac{1}{I(\theta)} \quad (2.22)$$

De lo cual observamos que la función de información está relacionado con la varianza del estimador, es decir mayor función de información implica menor varianza del estimador y viceversa.

Teniendo en cuenta esta relación otra utilidad importante de la función de información es el

área bajo su curva para un especificado intervalo del estimador de los evaluados.

Para evaluar esto, hemos graficado la función de información derivada en (2.16) para diferentes valores de beta, considerando un rango de valores de θ . En este caso tenemos tres ítems con $\beta = -2$ (ítem fácil), $\beta = 0$ (ítem no tan fácil, ni tan difícil) y $\beta = 2$ (ítem difícil)

Función de Información del ítem

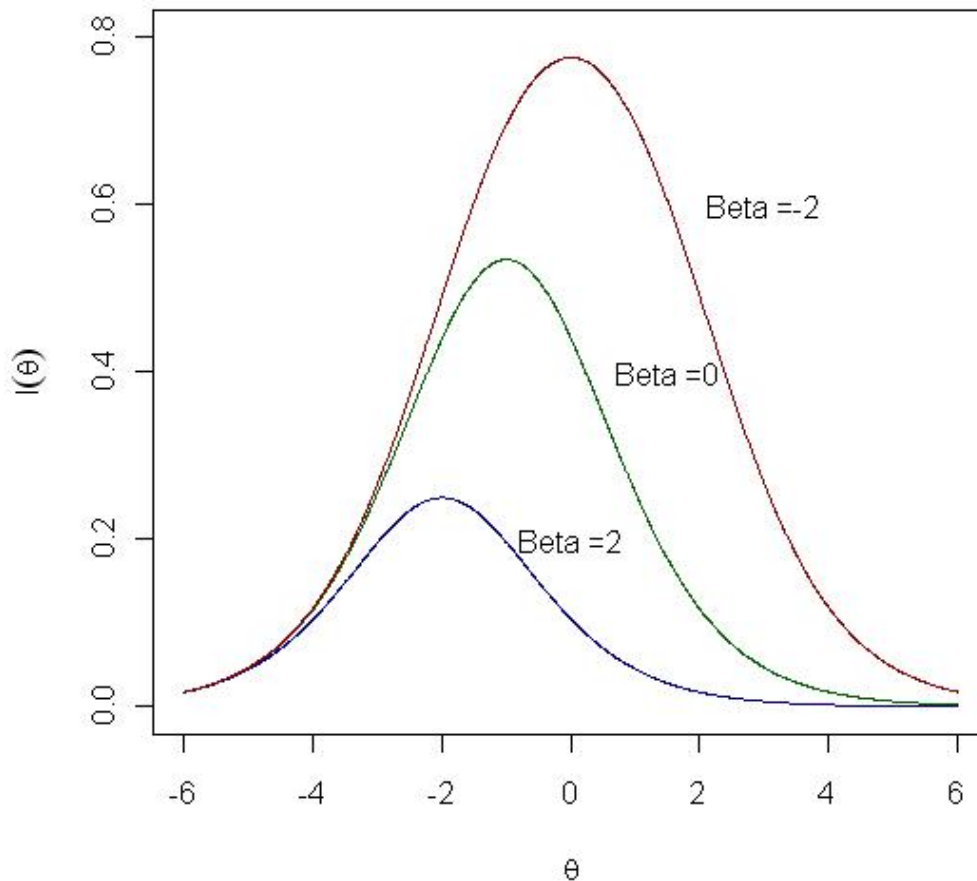


Figura 2.3: Función de información del ítem

De la figura 2.3 observamos que para un ítem difícil la función de información (curva) es menor, y está concentrada en niveles bajos de habilidad lo que indica que puede discriminar en este grupo pero no en todos los casos. Un efecto inverso se observa en los ítems fáciles, en la cual la curva de información es mayor en los niveles altos de las habilidades de los evaluados.

Graficamos la información de Fisher de un test o prueba de k ítems obtenida en la ecuación (2.20):

Función de información de la prueba

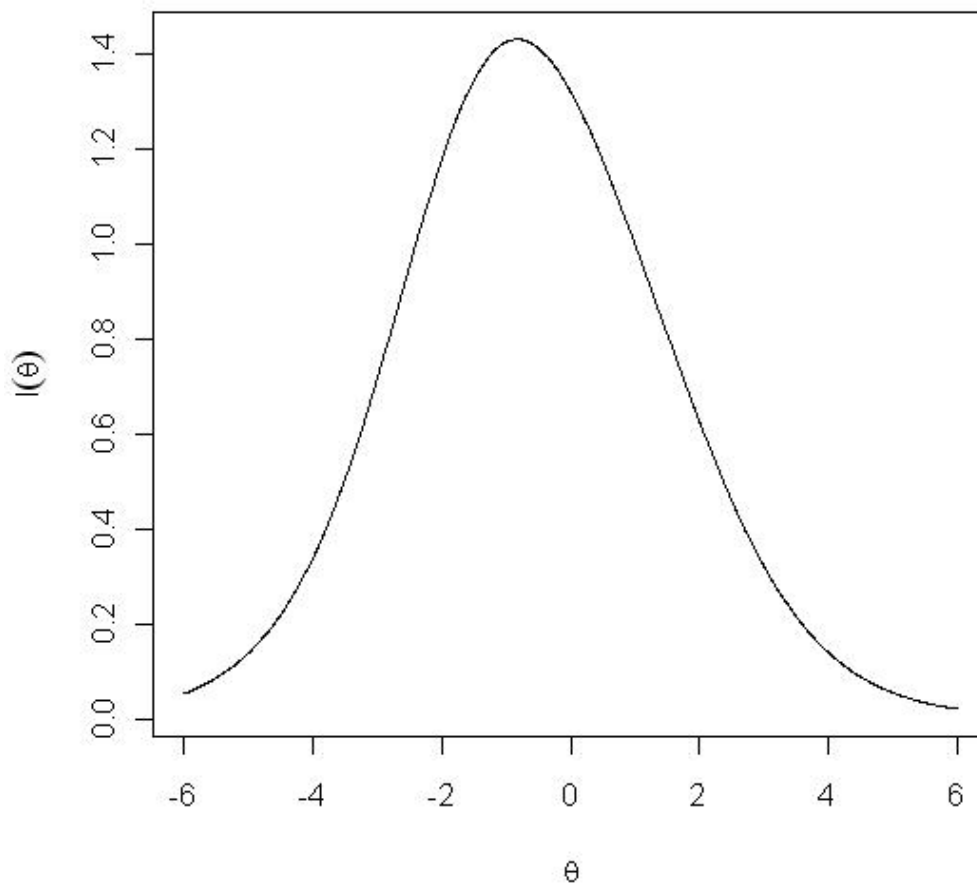


Figura 2.4: Función de información de la prueba

Integrando la ecuación (2.20) obtenemos el área bajo la curva de información de Fisher del test o prueba, el cual se muestra en el cuadro siguiente:

INTERVALO	INFORMACION	%
$[-6,0]$	9.02903	61.79 %
$]0,6]$	5.58368	38.21 %
TOTAL	14.61271	100 %

Cuadro 2.1: Resumen de la función de información de la prueba

De la figura 2.4 y el cuadro 2.1 podemos observar la función de información del test o prueba y el área bajo la curva en intervalos definidos de habilidades de los evaluados; el cual me indica que los ítems de la prueba principalmente proporcionan más información de los evaluados con baja habilidad. En particular, la cantidad de información de la prueba para niveles de habilidad en el intervalo de $(-6,0)$ es casi 62 %, mientras que el ítem que parece discriminar a los evaluados con altos niveles de habilidad es el ítem más fácil ($\beta = -2$).

Capítulo 3

Estimación del modelo de Rasch

En este capítulo presentamos métodos de estimación de los parámetros para modelos de Rasch dicotómicos. Los métodos de estimación presentados serán vistos de una perspectiva clásica y bayesiana. Dentro de la perspectiva clásica se han propuestos diversos métodos de estimación basados en la función de verosimilitud del modelo ver por ejemplo en (Kim, 2001), (Fischer y Molenaar, 1995), (Baker, 1992) y (Swaminathan y Gifford, 1982). Estos métodos son: Métodos de Máxima Verosimilitud Conjunta, Métodos de Máxima Verosimilitud Condicional y Métodos de Máxima verosimilitud Marginal. Dentro de la perspectiva bayesiana se utilizan los métodos MCMC implementando en WingBUGS.

3.1. Métodos de estimación clásica del modelo de Rasch dicotómico

Dentro de los métodos de estimación clásica que mostraremos tenemos: Métodos de Máxima Verosimilitud Conjunta, Métodos de Máxima Verosimilitud Condicional y Métodos de Máxima verosimilitud Marginal basado en la función de Máxima Verosimilitud del modelo.

3.1.1. Estimación por métodos de Máxima Verosimilitud Conjunta

Considerando la función de verosimilitud del modelo de Rasch dicotómico (Baker, 1992), (Wright, 1998), dada en la ecuación (2.12):

$$L(\mathbf{Y}|\theta, \beta) = \frac{e^{\sum_{i=1}^n \theta_i \Gamma_i - \sum_{j=1}^k Z_j \beta_j}}{\prod_{i=1}^n \prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]} \quad (3.1)$$

Otra forma de expresar la función de verosimilitud es a través de su logaritmo neperiano:

$$l(\mathbf{Y}|\theta, \beta) = \sum_{i=1}^n \theta_i \Gamma_i - \sum_{j=1}^k Z_j \beta_j - \sum_{i=1}^n \sum_{j=1}^k \ln [1 + e^{(\theta_i - \beta_j)}] \quad (3.2)$$

Para estimar por máxima verosimilitud conjunta los parámetros de las personas y los ítems, tomamos derivadas parciales respecto de cada parámetro al logaritmo neperiano de la función de verosimilitud e igualamos a 0:

$$\frac{\partial l(\mathbf{Y}|\theta, \beta)}{\partial \theta_i} = \Gamma_i - \sum_{j=1}^k \frac{e^{(\theta_i - \beta_j)}}{1 + e^{(\theta_i - \beta_j)}} = 0 \quad (3.3)$$

obtenemos:

$$\Gamma_i = \sum_{j=1}^k \frac{e^{(\theta_i - \beta_j)}}{1 + e^{(\theta_i - \beta_j)}} \quad (3.4)$$

Similarmente:

$$\frac{\partial l(\mathbf{Y}|\theta, \beta)}{\partial \beta_j} = -Z_j + \sum_{i=1}^n \frac{e^{(\theta_i - \beta_j)}}{1 + e^{(\theta_i - \beta_j)}} = 0 \quad (3.5)$$

obtenemos:

$$Z_j = \sum_{i=1}^n \frac{e^{(\theta_i - \beta_j)}}{1 + e^{(\theta_i - \beta_j)}} \quad (3.6)$$

Para obtener los valores estimados de los parámetros de las personas evaluadas (θ_i) y la de los ítems (β_j) resolvemos las ecuaciones (3.4) y (3.6) utilizando métodos numéricos y métodos computacionales como el de Newton Raphson (ver Anexo A)

Existen diferentes softwares que tienen implementado el método de estimación de Máxima Verosimilitud Conjunta. Entre los cuales tenemos: Rascal, Quest, Winsteps, Bigsteps, Facets, Facets-DOS, Minfac, Minstep, Bical, Bias, OpenStat, Mark Moulton, Smile Chien. En la actualidad el Ministerio de Educación del Perú hace uso de los modelos de Rasch bajo este método de estimación, emplearemos para el análisis software Winsteps (John M. Linacre, 2009), ver el código del programa utilizado en el Anexo B.

3.1.2. Estimación por métodos de Máxima Verosimilitud Condicional

Para este método de estimación partimos también de la función de verosimilitud para cada persona evaluada del modelo (Fischer, 2007), (Baker, 1992), (Kim, 2001), (Morgan y Gustafsson, 1980) está dada en la ecuación (2.12):

$$L(Y_i|\theta_i, \beta) = \frac{e^{\theta_i \Gamma_i - \sum_{j=1}^k y_{ij} \beta_j}}{\prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]} \quad (3.7)$$

$$L(Y_i|\theta_i, \beta) = \frac{e^{\theta_i \Gamma_i} \prod_{j=1}^k e^{y_{ij} \beta_j}}{\prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]} \quad (3.8)$$

Consideramos un conjunto de patrones de respuesta $W = (W_1, W_2, W_3, \dots)$, los patrones de respuesta son las formas de respuesta a los ítems por parte de cada persona evaluada:

$W_1 = (0, 0, 0, 0, 0)$, $W_2 = (0, 1, 1, 0, 0)$, $W_3 = (0, 1, 0, 1, 0)$, ..., $W_k = (1, 1, 1, 1, 1)$. Dicho esto la función de verosimilitud de las puntuaciones de cada persona queda expresado de la siguiente forma:

$$L(\Gamma|\beta) = \frac{e^{\theta_i \Gamma_i} \sum_W \prod_{j=1}^k e^{-y_{ij} \beta_j}}{\prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]} \quad (3.9)$$

Definimos: $\epsilon_j = e^{-\beta_j}$ y $\delta_r(\epsilon) = \sum_W \prod_{j=1}^k e^{-y_{ij} \beta_j} = \sum_W \prod_{j=1}^k \epsilon^{y_{ij}}$, donde $\delta_r(\epsilon)$ es una función simétrica elemental. Dicho de otro modo:

$$\delta_0(\epsilon) = 1$$

$$\delta_1(\epsilon) = \epsilon_1 + \epsilon_2 + \epsilon_3 + \dots + \epsilon_k$$

$$\delta_2(\epsilon) = \epsilon_1 \epsilon_2 + \epsilon_1 \epsilon_3 + \dots + \epsilon_{k-1} \epsilon_k$$

.

.

$$\delta_k(\epsilon) = \epsilon_1 \epsilon_2 \epsilon_3 \epsilon_4 \dots \epsilon_{k-1} \epsilon_k$$

Entonces la función de verosimilitud de las puntuaciones obtenida por cada persona queda expresado de la siguiente forma:

$$L(\Gamma|\beta) = \frac{e^{\theta_i \Gamma_i} \delta_r(\epsilon)}{\prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]} \quad (3.10)$$

La función de verosimilitud condicional para una persona se determina de la siguiente manera:

$$L_c(Y_i|\beta, \Gamma) = \frac{e^{-\sum_{j=1}^k y_{ij} \beta_j}}{\delta_{ri}(\epsilon)} \quad (3.11)$$

Luego la función de verosimilitud condicional para todas las personas evaluadas es:

$$L_C(Y|\beta, \Gamma) = \frac{\prod_{i=1}^n e^{-\sum_{j=1}^k y_{ij} \beta_j}}{\prod_{i=1}^n \delta_{ri}(\epsilon)}$$

$$L_C(Y|\beta, \Gamma) = \frac{e^{-\sum_{j=1}^k Z_j \beta_j}}{\prod_{i=1}^n \delta_{ri}(\epsilon)} \quad (3.12)$$

Otra forma de expresar la función de verosimilitud condicional para todas las personas evaluadas es mediante el logaritmo neperiano de su función de verosimilitud:

$$\ln L_C(Y|\beta, \Gamma) = - \sum_{j=1}^k Z_j \beta_j - \sum_{i=1}^n \ln \delta_{ri}(\epsilon) \quad (3.13)$$

Para estimar los parámetros de ítems tomamos las derivadas parciales al logaritmo nepe-

riano de la función de verosimilitud respecto del parámetro de los ítems e igualamos a 0:

$$\frac{\partial \ln L_c(Y|\beta, \Gamma)}{\partial \beta} = \frac{\partial [-\sum_{j=1}^k Z_j \beta_j - \sum_{i=1}^n \ln \delta_{ri}(\epsilon)]}{\partial \beta} = 0 \quad (3.14)$$

Desarrollando las ecuaciones via métodos numéricos obtenemos los estimados de los parámetros de los ítems.

En la tesis para este método de estimación utilizamos el software R, implementando el paquete **eRm** (Extended Rasch Modeling) desarrollado por (Mair P. y Hatzinger R, 2007), ver el código del programa utilizado en el Anexo B. Existen diferentes softwares que tienen implementado este método de estimación entre los cuales tenemos: eRm, Logimo, LPCM-WIN, RSP, WINMIRA, Digram, MULTIRA, OPLM, RASCHTEST (Modulo de STATA); vale indicar que no se pueden estimar los parámetros para casos extremos o patrones de respuesta del tipo: (0, 0, 0, ..., 0) y (1, 1, 1, ..., 1), es decir para personas que contestan todas las preguntas o no contestan ninguna, o para preguntas contestadas por todos o por nadie. Si existen estos casos, estos son eliminados del proceso de estimación.

3.1.3. Estimación por métodos de Máxima Verosimilitud Marginal

Para aplicar los métodos de estimación de máxima verosimilitud marginal asumimos que existe una distribución a priori para el parámetro de las persona evaluadas, que tiende a una distribución normal con media 0 y varianza 1 (Fischer, 2007), (Baker, 1992), (Kim, 2001), (Morgan y Gustafsson, 1980). Dicho de otra manera:

$$\theta_i \sim N(0, 1) \Rightarrow P(\theta_i) = \frac{e^{-\frac{\theta_i^2}{2}}}{\sqrt{2\pi}}$$

De la siguiente forma obtenemos la función de verosimilitud marginal para cada persona evaluada:

$$L_M(Y_i|\beta) = \int_{-\infty}^{\infty} L(Y_{ij}|\theta_i, \beta) P(\theta_i) d\theta_i \quad (3.15)$$

$$L_M(Y_i|\beta) = \int_{-\infty}^{\infty} \frac{e^{\theta_i \Gamma_i - \sum_{j=1}^k y_{ij} \beta_j}}{\prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]} \times \frac{e^{-\frac{\theta_i^2}{2}}}{\sqrt{2\pi}} d\theta_i \quad (3.16)$$

Haciendo cambio de variable y aproximando la integral mediante el algoritmo EM (Bock and Aitkin, 1981) con fórmulas de cuadraturas de Gauss - Hermite (ver Anexo A), generando m puntos, podemos expresar la función de verosimilitud marginal para cada persona evaluada de la siguiente manera:

$$L_M(Y_i|\beta) = \frac{1}{\sqrt{\pi}} \sum_{h=1}^m C_h \frac{e^{\theta_{ih} \Gamma_i - \sum_{j=1}^k y_{ij} \beta_j}}{\prod_{j=1}^k [1 + e^{(\theta_{ih} - \beta_j)}]} \quad (3.17)$$

donde

$$C_h = \frac{2^{(m-1)}m!\sqrt{\pi}}{m^2[H_{(m-1)}(\theta_i)]}$$

y $H_{(m-1)}(\theta_i)$ son polinomios Hermite.

Luego la función de verosimilitud marginal para todas las personas evaluadas se expresa como: $L_M(Y|\beta) = \prod_{i=1}^n L_M(Y_i|\beta)$.

$$L_M(Y|\beta) = e^{-\sum_{j=1}^k Z_j} \prod_{i=1}^n \sum_{h=1}^m \frac{2^{(m-1)}m!}{m^2[H_{(m-1)}(\theta_i)]} \times \frac{e^{\theta_{ih}\Gamma_i}}{\prod_{j=1}^k [1 + e^{(\theta_{ih}-\beta_j)}]} \quad (3.18)$$

Otra forma de expresar la función de verosimilitud marginal es mediante su logaritmo neperiano:

$$\ln [L_M(Y|\beta)] = -\sum_{j=1}^k Z_j \beta_j + \sum_{i=1}^n \left(\ln \left(\sum_{h=1}^m \frac{2^{(m-1)}m!}{m^2[H_{(m-1)}(\theta_i)]} \times \frac{e^{\theta_{ih}\Gamma_i}}{\prod_{j=1}^k [1 + e^{(\theta_{ih}-\beta_j)}]} \right) \right) \quad (3.19)$$

Para estimar los parámetros de los ítems tomamos la derivada parcial respecto al parámetro del ítem al logaritmo neperiano de la función de verosimilitud marginal de todas las personas evaluadas e igualamos a 0

$$\frac{\partial \ln (L_M(Y|\beta))}{\partial \beta} = 0$$

La solución a la ecuación se obtienen por métodos numéricos.

Una vez estimado los parámetros de los ítems, podemos bajo el mismo algoritmo estimar los parámetros de las personas. Para la estimación de los parámetros existen limitaciones cuando las personas evaluadas alcanzan el puntaje máximo o el mínimo los cuales son discutidos en Patz and Junker (1999) y Sahu (2002).

Otra forma de estimar los parámetros de las personas es a través de métodos Bayesianos via MCMC, las cuales serán desarrollados en el siguiente método de estimación.

Existen diferentes softwares que tienen implementado este método de estimación como son: LTM (Latent Variable Modeling), ConQuest, RSP, SALTUS, ICL, BILOG-MG, PARSCALE, PARDUX, RASCHTEST (modulo de STATA). Estos softwares presentan problemas en la estimación para puntuaciones totalmente correctas e incorrectas de las personas, es decir para personas que contestan todas las preguntas o no contestan ninguna, o para preguntas contestadas por todos o por nadie. Si existen estos casos, estos son eliminados del proceso de estimación. Para la tesis utilizamos el software R, implementado el paquete **LTM** (Latent Variable Modeling) propuesto por (Rizopoulos D., 2006), debido a que es un software libre. Ver el código del programa utilizado en el Anexo B.

3.2. Métodos de estimación bayesiana para el modelo de Rasch dicotómico

Para la estimación de los parámetros en un modelo de Rasch por métodos bayesianos debemos especificar las distribuciones a priori de los parámetros de las personas (θ) y de los ítems (β). Si consideramos la densidad a priori conjunta de estos parámetros, entonces la densidad a posteriori es proporcional a:

$$P(\theta, \beta|Y) \propto L(\mathbf{Y}|\theta, \beta)P(\theta, \beta) \quad (3.20)$$

Considerando la función de verosimilitud del modelo:

$$l(\mathbf{Y}|\theta, \beta) = \sum_{i=1}^n \sum_{j=1}^k y_{ij}\theta_i - \sum_{i=1}^n \sum_{j=1}^k y_{ij}\beta_j - \sum_{i=1}^n \sum_{j=1}^k \ln [1 + e^{(\theta_i - \beta_j)}] \quad (3.21)$$

Asumimos a priori que los parámetros de la personas e ítems tienen una distribución normal $\theta_i \sim N(0, 1)$ y $\beta_j \sim N(0, \sigma_\beta^2)$.

y considerando la función conjunta a priori de los parámetros a estimar como:

$$P(\theta, \beta) = \prod_{i=1}^n P(\theta_i) \prod_{j=1}^k P(\beta_j)$$

$$P(\theta, \beta) = \prod_{i=1}^n \frac{e^{-\frac{\theta_i^2}{2}}}{\sqrt{2\pi}} \times \prod_{j=1}^k \frac{e^{-\frac{\beta_j^2}{2\sigma_\beta^2}}}{\sqrt{2\pi\sigma_\beta^2}}$$

Resolviendo

$$P(\theta, \beta) = \frac{1}{\sigma_\beta^k (2\pi)^{\frac{n+k}{2}}} e^{-\frac{1}{2}(\sum_{i=1}^n \theta_i^2 + \frac{\sum_{j=1}^k \beta_j^2}{\sigma_\beta^2})} \quad (3.22)$$

$$P(\theta, \beta) \propto e^{-\frac{1}{2}(\sum_{i=1}^n \theta_i^2 + \frac{\sum_{j=1}^k \beta_j^2}{\sigma_\beta^2})} \quad (3.23)$$

Teniendo en cuenta la ecuación (3.20), reemplazando por la función de verosimilitud del modelo y la probabilidad conjunta de los parámetros, hallamos la distribución a posteriori de los parámetros a estimar:

$$P(\theta, \beta|Y) \propto \sum_{i=1}^n \sum_{j=1}^k y_{ij}\theta_i - \sum_{i=1}^n \sum_{j=1}^k y_{ij}\beta_j - \sum_{i=1}^n \sum_{j=1}^k \ln [1 + e^{(\theta_i - \beta_j)}] \times e^{-\frac{1}{2}(\sum_{i=1}^n \theta_i^2 + \frac{\sum_{j=1}^k \beta_j^2}{\sigma_\beta^2})} \quad (3.24)$$

Teniendo en cuenta la distribución conjunta encontramos las distribuciones condicionales completas

Distribución condicional completa de θ

$$P(\theta|\beta, Y) \propto L(Y|\theta, \beta)P(\theta)$$

Reemplazando con la función de máxima verosimilitud del modelo y con la distribución a priori de θ :

$$P(\theta_i|\beta, Y) \propto \frac{e^{\sum_{j=1}^k y_{ij}\theta_i - \sum_{j=1}^k y_{ij}\beta_j}}{\prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]} \times \frac{e^{-\frac{\theta_i^2}{2}}}{\sqrt{2\pi}}$$

Resolviendo

$$P(\theta_i|\beta, Y) \propto \frac{e^{\sum_{j=1}^k y_{ij}(\theta_i - \beta_j) - \frac{1}{2} \sum_{i=1}^n \theta_i^2}}{\prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]} \quad (3.25)$$

Distribución condicional completa de β

$$P(\beta|\theta, Y) \propto L(Y|\theta, \beta)P(\beta)$$

Reemplazando con la función de máxima verosimilitud del modelo y con la distribución a priori de β :

$$P(\beta_j|\theta, Y) \propto \frac{e^{\sum_{i=1}^n y_{ij}\theta_i - \sum_{i=1}^n y_{ij}\beta_j}}{\prod_{i=1}^n [1 + e^{(\theta_i - \beta_j)}]} \times \frac{e^{-\frac{\beta_j^2}{2\sigma_\beta^2}}}{\sqrt{2\pi\sigma_\beta^2}}$$

Resolviendo

$$P(\beta_j|\theta, Y) \propto \frac{e^{\sum_{i=1}^n y_{ij}(\theta_i - \beta_j) - \frac{1}{2\sigma_\beta^2} \sum_{i=1}^n \beta_j^2}}{\prod_{i=1}^n [1 + e^{(\theta_i - \beta_j)}]} \quad (3.26)$$

De las distribuciones condicionales completas para θ_i y β_j notamos que nos son estándares, es por ello que es difícil implementar un muestreo Gibbs.

Para estimar los parámetros daremos uso a los métodos MCMC y para ello consideramos dos escenarios:

3.2.1. MCMC usando adaptive rejection sampling (ARS)

El modelo de Rasch dicotómico puede ser ajustado usando MCMC. El logaritmo de las densidades de las distribuciones condicionales para cada uno de los parámetros a estimar son cóncavos (Sahu, 2001), el muestreo exacto de ellas es utilizando rejection sampling (Wilks y Wald, 1982). El ARS trabaja dinámicamente dos envolventes (una superior y una inferior) para la distribución a ser muestreada de evaluaciones sucesivas de la densidad en los puntos rechazados. El algoritmo se detiene cuando el punto propuesto ha sido aceptado. Este el

procedimiento por defecto en el software WinBUGS para el modelo de Rasch dicotómico. Ver el código del programa utilizado en el Anexo B.

3.2.2. MCMC usando data argumentation Gibbs sampling (DAGS)

Definimos e introducimos una variable aleatoria continua Z_{ij} tal que:

$$Z_{ij} = \theta_i - \beta_j + \epsilon_{ij}$$

donde $\epsilon_{ij} \sim N(0, 1)$, con la cual la Z_{ij} tiene una distribución normal con media $\theta_i - \beta_j$ y varianza 1.

$$Y_{ij} = \begin{cases} 1 & \text{Si } Z_{ij} > 0, \\ 0 & \text{Si } Z_{ij} \leq 0 \end{cases}$$

Entonces la probabilidad de dar respuesta 1 puede ser expresado como:

$$P(Y_{ij} = 1) = P(Z_{ij} > 0)$$

Considerando $D = (Z, Y)$ tal que:

$$P(D|\theta, \beta) = \prod_{i=1}^n \prod_{j=1}^k \left[\frac{e^{Z_{ij}(\theta_i - \beta_j)}}{1 + e^{(\theta_i - \beta_j)}} \right] I(Z_{ij} Y_{ij}) \quad (3.27)$$

Donde I es una función indicadora el cual toma valores de 0 y 1:

$$I(Z_{ij} Y_{ij}) = I(Z_{ij} > 0)I(Y_{ij} = 1) + I(Z_{ij} \leq 0)I(Y_{ij} = 0)$$

Sabemos por el teorema de Bayes que:

$$P(\theta, \beta|D) \propto P(D|\theta, \beta)P(\theta, \beta)$$

Reemplazando:

$$P(\theta, \beta|D) \propto \prod_{i=1}^n \prod_{j=1}^k \left[\frac{e^{Z_{ij}(\theta_i - \beta_j)}}{1 + e^{(\theta_i - \beta_j)}} \right] \times I(Z_{ij} Y_{ij}) \times e^{-\frac{1}{2}(\sum_{i=1}^n \theta_i^2 + \frac{\sum_{j=1}^k \beta_j^2}{\sigma_\beta^2})}$$

Finalmente tenemos:

$$P(\theta, \beta|D) \propto \frac{e^{\sum_{i=1}^n \sum_{j=1}^k Z_{ij}(\theta_i - \beta_j) - \frac{1}{2}(\sum_{i=1}^n \theta_i^2 + \frac{\sum_{j=1}^k \beta_j^2}{\sigma_\beta^2})}}{\prod_{i=1}^n \prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]} I(Z_{ij} Y_{ij}) \quad (3.28)$$

De las ecuaciones (3.27) y (3.28) determinamos la distribuciones condicionales completas: $f(Z_{ij}|\theta, \beta, Y)$, $f(\theta|Z_{ij}, \beta, Y)$ y $f(\beta|Z_{ij}, \theta, Y)$, las cuales estan determinados en (Maris G. y Maris E., 2002)

Con las distribuciones condicionales se puede implementar el muestreo de Gibbs (ver anexo A).

Para implementar el método Gibbs sampler. Suponemos $(t-1)$ iteraciones de los valores actuales de los parámetros del modelo que son denotados por $[Z_{ij}^{(t-1)}]$, $[\theta_i^{(t-1)}]$ y $[\beta_j^{(t-1)}]$, entonces el algoritmo Gibbs Sampling puede ser descrito como sigue:

1. Los valores de la data latente $[Z_{ij}]$ son simulados a partir de la distribución condicional completa $f(Z_{ij}|\theta, \beta, Y)$.
2. Los valores de los parámetros de las personas θ_i son simulados a partir de distribución condicional completa $f(\theta|Z_{ij}, \beta, Y)$.
3. Los valores de los parámetros de los ítems β_j son simulados a partir de la distribución condicional completa $f(\beta|Z_{ij}, \theta, Y)$.

Se puede implementar fácilmente el muestreo de Gibbs. Las rutinas en R (en MCMCpack de Martín y Quinn, 2003) y Matlab (Johnson y Albert, 1999) están disponibles en la web para el caso normal o probit; también es posible construir un script en R directamente. Nosotros preferimos en la tesis el uso del WinBugs por que nos brinda mas información y resultados de este método de estimación. Ver el código del programa utilizado en el Anexo B.

Capítulo 4

Estudio de Simulación

Se desarrollaran simulaciones sobre diferentes escenarios, para comparar los diferentes métodos de estimación y ver cual es el método mas preciso.

4.1. Condiciones de la simulación

4.1.1. Algoritmo para generar una data dicotómica

Para generar la data dicotómica bajo el Modelo Rasch: $P = P(Y = 1|\theta, \beta) = \frac{e^{(\theta-\beta)}}{1+e^{(\theta-\beta)}}$

- Fijamos un número "n" de personas evaluados y "k" ítems del instrumento de medición utilizado en la investigación.
- Simulamos los parámetros del modelo bajo una distribución normal: $\theta \sim N(0, 1)$ y $\beta \sim N(0, 1)$.
- Una vez simulado los parámetros, los valores son corregidos por la media de modo que el promedio sea 0, con la finalidad de hacer comparables los tres métodos de estimación.
- Simulamos la probabilidad de responder correctamente al ítem bajo el Modelo Rasch descrito.
- Simulamos probabilidades cualquiera (R) mediante números aleatorios que tengan una distribución entre 0 y 1.
- Si "R" es mayor que el complemento de la probabilidad Rasch simulada entonces la variable aleatoria toma valor "1", caso contrario toma valor "0".

Mediante el algoritmo mostrado simulamos la data dicotómica aleatoria. Ver el código del programa utilizado para generar una base de datos dicotómica en el Anexo B.

4.1.2. Criterios para evaluar la simulación

Una forma de medir el ajuste es a través del error cuadrático medio, definido por:

$$MSE = \sum_{j=1}^k \frac{(\beta_{simulado} - \beta_{estimado})^2}{k} \quad (4.1)$$

Otra forma es considerar la raíz cuadrada del error cuadrático medio definido por:

$$RMSE = \sqrt{\sum_{j=1}^k \frac{(\beta_{simulado} - \beta_{estimado})^2}{k}} \quad (4.2)$$

Una ultima forma es correlacionar los valores simulados y ajustados:

$$CORR = \frac{\sum_{j=1}^k (\beta_{simulado} - \bar{\beta}_{simulado})(\beta_{estimado} - \bar{\beta}_{estimado})}{\sqrt{\sum_{j=1}^k (\beta_{simulado} - \bar{\beta}_{simulado})^2} \sqrt{\sum_{j=1}^k (\beta_{estimado} - \bar{\beta}_{estimado})^2}} \quad (4.3)$$

4.2. Métodos de estimación a estudiar

En el caso de los métodos de estimación de Máxima Verosimilitud Conjunta (**MVC**) se utilizará el software Winsteps (Linacre, 2009) y (Bond y Fox, 1985), para el método de estimación de Máxima Verosimilitud Marginal (**MVM**) se utilizará el software R bajo el paquete **ltm**, para el método de estimación de Máxima Verosimilitud Condicional (**MVCOND**) se utilizará el software R bajo el paquete **eRm** (Mair y Hatzinger, 2007) y (Rizopoulos, 2006) y para los métodos de estimación Bayesiana (**MB**) se utilizará el software WinBUGS (Tantau, 2007) bajo dos esquemas: ARS (Adaptative Rejection Sampling) y DAGS (Data Argumentation Gibbs Sampling), presentados en el capítulo de estimación del modelo de Rasch. Los programas utilizados para cada método de estimación están en el Anexo B.

Con las simulaciones se busca analizar y cuantificar el funcionamiento de los métodos de estimación sujetos a diferentes variaciones del número de personas evaluadas y el número de preguntas de cualquier instrumento de medición utilizado en todo proyecto de investigación.

4.3. Escenarios de Estudio

Son 4 escenarios en los cuales se analizarán los métodos de estimación presentados, sujetos a variaciones del número de ítems del instrumento de medición y números de evaluados. Considerando los siguientes escenarios:

Numero pequeño de ítems y tamaño de muestra variable

- Escenario 1: Número pequeño de ítems y número grande de personas, en la cual se simulan 10 ítems y 1000 evaluados .

- Escenario 2: Número pequeño de ítems y número pequeño de personas, en la cual se simulan 10 ítems y 15 evaluados .

Numero grande de ítems y tamaño de muestra variable

- Escenario 3: Número grande de ítems y número grande de personas, en la cual se simulan 100 ítems y 1000 evaluados .
- Escenario 4: Número grande de ítems y número pequeño de personas, en la cual se simulan 100 ítems y 15 evaluados .

Para cada escenario de análisis consideramos los cuatro métodos de estimación propuesto: Máxima Verosimilitud Marginal (MVM), Máxima Verosimilitud Condicional (MV-COND), Máxima Verosimilitud Conjunta (MVC) y Método Bayesiano (MB). Debido a las parametrizaciones realizadas en el modulo eRm y LTM del software R y WinsSteps, es por ello que para efectos de comparabilidad entre los métodos de estimación y la información simulada corregimos los valores de los parámetros simulados por: $\beta_j^* = \beta_j - \bar{\beta}$ y $\theta_j^* = \theta_j - \bar{\theta}$. Y para la estimación bayesiana se realizó la misma corrección. Otros software realizan otras parametrizaciones, la cuales deben ser definidas adecuadamente para obtener una estimación precisa.

Se obtuvo a continuación los siguientes resultados:

4.3.1. Número pequeño de ítems y tamaño de muestra variable

Consideramos a 10 ítems frente a diversos cambios del tamaño de muestra y mostramos las estimaciones de los parámetros de los ítems en los cuatro métodos de estimación presentados:

Comparación de estimación de parámetros de los ítems

ESC	ITEM	SIM	MVM	MVCON	MVC	MB(ARS)	MB(DAGS)
1	1	-0.4842	-0.4837	-0.4826	-0.5376	-0.4840	-0.4823
	2	0.1495	0.2011	0.2012	0.2236	0.1997	0.1991
	3	-0.061	-0.0023	-0.0017	-0.0022	-0.0017	-0.0033
	4	0.9047	0.905	0.9025	1.0077	0.9033	0.9055
	5	0.1349	0.0939	0.0949	0.1052	0.0979	0.0938
	6	0.3366	0.3564	0.3569	0.3971	0.3535	0.3558
	7	-0.5638	-0.6057	-0.6051	-0.6742	-0.6036	-0.6018
	8	0.7325	0.6778	0.6767	0.7542	0.6758	0.6774
	9	-0.8073	-0.7944	-0.7951	-0.8864	-0.7956	-0.7928
	10	-0.3419	-0.3483	-0.3477	-0.3873	-0.3453	-0.3513
2	1	-0.4842	0.5734	0.5314	0.5822	0.4883	0.5020
	2	0.1495	0.2587	0.2391	0.2623	0.2503	0.227
	3	-0.061	-0.3648	-0.3387	-0.3707	-0.3429	-0.3211
	4	0.9047	1.2683	1.1786	1.2925	1.1134	1.095
	5	0.1349	-0.3648	-0.3387	-0.3707	-0.3417	-0.3354
	6	0.3366	-0.0506	-0.0477	-0.0518	-0.0637	-0.0520
	7	-0.5638	-0.0505	-0.0477	-0.0518	-0.0546	-0.0507
	8	0.7325	-0.0505	-0.0477	-0.0518	-0.0416	-0.0259
	9	-0.8073	-1.478	-1.3678	-1.5026	-1.259	-1.238
	10	-0.3419	0.2588	0.2391	0.2623	0.2517	0.1993

Cuadro 4.1: Estimaciones de los parámetros de los ítems para el escenario 1 y 2

ESC	ESTADISTICO	MVM	MVCON	MVC	MB(ARS)	MB(DAGS)
1	MSE	0.001313	0.001322	0.00479	0.001264	0.001263
	RMSE	0.036237	0.036353	0.069211	0.035550	0.035537
	CORR	0.997683	0.997671	0.997691	0.997776	0.997776
2	MSE	0.344192	0.309051	0.352834	0.288142	0.277154
	RMSE	0.586679	0.555923	0.593998	0.536788	0.526455
	CORR	0.547265	0.547385	0.547794	0.542636	0.549451

Cuadro 4.2: Medidas de ajustes de los ítems para el escenario 1 y 2

En el cuadro 4.1 mostramos las estimaciones del parámetro de los ítems con los cuatro métodos de estimación presentados, para una cantidad de ítems pequeños y muestra de personas variable. En el cuadro 4.2: Medidas de ajustes de los ítems para el escenario 1 y 2, observamos que se logra un mejor ajuste para un número pequeño de ítems siempre y cuando la cantidad de personas evaluadas es grande. También observamos que para un número pequeño de ítems sujeto a variaciones del número de personas evaluadas, que el mejor método de estimación es el método bayesiano bajo el esquema DAGS dado que muestra mejores criterios de ajuste, es uno de los métodos más robustos debido a que no presentan problemas en la estimación de los parámetros de los ítems.

Comparación de estimación de parámetros de las personas

ESC	ESTADISTICO	MVC	MB (ARS)	MB (DAGS)
1	MSE	0.6886	0.3616	0.3622
	RMSE	0.8298	0.6013	0.6018
	CORR	0.8064	0.8133	0.8128
2	MSE	0.2752	0.1876	0.1943
	RMSE	0.5246	0.4332	0.4408
	CORR	0.8441	0.8240	0.8170

Cuadro 4.3: Medidas de ajustes de las personas para el escenario 1 y 2

Del cuadro 4.3: Medidas de ajustes de las personas para el escenario 1 y 2, para un número pequeño de ítems sujeto a variaciones del número de personas observamos que el mejor método de estimación es el método bayesiano bajo los esquemas ARS y DAGS, dado que presentan mejores ajustes.

4.3.2. Número grande de ítems y tamaño de muestra variable

Consideramos a 100 ítems frente a diversos cambios del tamaño de muestra y mostramos los diferentes métodos de estimación:

Comparación de estimación de parámetros de los ítems

ESC	ESTADISTICO	MVM	MVCON	MVC	MB (ARS)	MB (DAGS)
3	MSE	0.0075	0.0074	0.0078	0.0073	0.0072
	RMSE	0.0865	0.0862	0.0882	0.0854	0.0851
	CORR	0.9965	0.9965	0.9965	0.9965	0.9965
4	MSE	5.8039	0.5725	0.4918	0.3260	0.3257
	RMSE	2.4091	0.7566	0.7013	0.5709	0.5707
	CORR	0.5545	0.7677	0.8047	0.8394	0.8404

Cuadro 4.4: Medidas de ajustes de los ítems para el escenario 3 y 4

Del cuadro 4.4: Medidas de ajustes de los ítems para el escenario 3 y 4, observamos que para un número grande ítems el ajuste disminuye si es que se reduce la muestra de personas evaluadas y para un número grande de ítems sujeto a variaciones del número de personas observamos que el mejor método de estimación es el método bayesiano bajo el esquema DAGS. Este es uno de los métodos más robustos dado que no presentan problemas en la estimación de los parámetros a comparación de los métodos de Máxima Verosimilitud Condicional y Conjunta.

Comparación de estimación de parámetros de las personas

ESC	ESTADISTICO	MVC	MB (ARS)	MB (DAGS)
3	MSE	0.0584	0.0543	0.0541
	RMSE	0.2416	0.2330	0.2325
	CORR	0.9739	0.9737	0.9738
4	MSE	0.0465	0.0342	0.0355
	RMSE	0.2157	0.1849	0.1885
	CORR	0.9752	0.9755	0.9746

Cuadro 4.5: Medidas de ajustes de las personas para el escenario 3 y 4

Del cuadro 4.5: Medidas de ajustes de las personas para el escenario 3 y 4, para un número grande de ítems sujeto a variaciones del número de personas observamos que el mejor método de estimación es el método bayesiano bajos los esquemas ARS y DAGS, dado que presenta mejores ajustes que el método de máxima verosimilitud conjunta. Además este uno de los métodos mas robustos dado que no presentan problemas en la estimación de los parámetros.

4.3.3. Conclusión del estudio de simulación

Del estudio de simulación realizado observamos que el método de estimación mas robusto es el Método Bayesiano, dado que a comparación de los otros métodos de estimación mencionados no presentan problemas en la estimación de parámetros de ítems ni de personas. El Método de Máxima Verosimilitud Marginal y Condicional no puede estimar los parámetros de los ítems cuando las personas han respondido correctamente a todos los ítems o haber fallado todos los ítems; para estimar los parámetros de las personas hay que aplicar métodos bayesianos. En cambio el Método de Máxima Verosimilitud Conjunta, para estos casos sobrestima los parámetros de las personas. Observamos que el método de estimación mas optimo es el Método Bayesiano bajos los esquemas ARS y DAGS, dado que presenta mejores indicadores de ajustes para los 4 escenarios analizados. Haciendo comparaciones en los escenarios 1 y 2 se logra tener un mejor ajuste cuando se tienen muestras grandes para numero de ítems pequeños. Cuando la cantidad de ítems es grande, sujeto a variaciones del tamaño de muestra el ajuste es similar. Al comparar los escenarios 2 y 4 observamos que cuando se tienen tamaños pequeños de muestra se logra un mejor ajuste siempre y cuando la cantidad de ítems es grande.

Capítulo 5

Aplicación

5.1. Descripción del problema de investigación

La aplicación de la tesis, esta orientado al ámbito educacional, en la cual mediante las respuestas de los estudiantes de primero de secundaria de las instituciones educativas estatales a una prueba de Comprensión Lectora a nivel nacional, se estimarán las variables latentes: habilidades que tienen los estudiantes para responder correctamente a un ítem y dificultad de los ítems de la prueba representado mediante un modelo logístico o modelo de Rasch, que modela la probabilidad de responder correctamente a un ítem de la prueba, bajo métodos de estimación desde una perspectiva clásica y bayesiana. Dentro de los métodos de estimación clásica presentamos los métodos de Máxima Verosimilitud Marginal, Máxima Verosimilitud Condicional y Máxima Verosimilitud Conjunta. Dentro de los Métodos Bayesianos presentamos a aquellos basados en los métodos de MCMC por el algoritmo Gibbs Sampling para la estimación de las habilidades de los estudiantes.

5.2. Tipo de instrumento de medición

Para estimar los parámetros del modelo y obtener las respuestas de los estudiantes se utilizó como instrumento de medición una prueba de Comprensión Lectora de primer grado de secundaria, utilizada en el Proyecto de Movilización por la Comprensión Lectora (DIGBER, 2009). En el Proyecto de Movilización por la Comprensión Lectora, se evaluaron a los estudiantes de los 5 grados de educación secundaria provenientes de las instituciones estatales de todas las regiones del país. Se consideraron 5 pruebas, una prueba para cada grado de educación secundaria y cada prueba esta compuesta por dos formas las cuales se detallan:

GRADO	FORMA	N ÍTEMES
1	1	17
	2	18
2	1	21
	2	18
3	1	24
	2	24
4	1	26
	2	26
5	1	25
	2	24

Cuadro 5.1: Distribución de los instrumentos de medición

Para la tesis se trabajó con la prueba correspondiente a la segunda forma de primero de secundaria, el cual consta de 18 ítems. La prueba esta compuesta por textos que agrupan varios ítems distribuidos de la siguiente manera:

TEXTOS	ÍTEMES
1	1
	2
	3
2	4
	5
	6
	7
	8
	9
3	10
	11
	12
	13
	14
4	15
	16
	17
	18

Cuadro 5.2: Distribución ítems en los textos del instrumentos de medición

Cada ítem consta de 4 alternativas de respuesta (A, B, C, D), en la cual si el estudiante tiene que escoger solo una de las alternativas, si es la respuesta correcta al ítem es codificada como 1 caso contrario como 0.

5.3. Diseño muestral

Para efectos de comparabilidad entre los 4 métodos de estimación que mostraremos, analizaremos a todos los estudiantes de primer grado de educación secundaria de instituciones educativas estatales, que hayan sido evaluados con el segundo cuadernillo y hayan

respondido todas las preguntas de la prueba. Esto nos dá 1536 estudiantes de las distintas regiones del Perú. El tipo de muestreo utilizado es un muestreo complejo de varias etapas. La prueba que utilizamos en la tesis ha sido parte del Proyecto Nacional por la Comprensión Lectora, en la cual se evaluaron a los estudiantes de los 5 grados de educación secundaria de las instituciones educativas estatales en todas las Regiones del Perú (DIGBER, 2009).

Población Objetivo

La población objetivo está conformada por los estudiantes primer grado de educación secundaria de instituciones educativas estatales.

Unidades de muestreo

En la investigación las unidades de muestreo son: instituciones educativas estatales de primero de secundaria de todas las regiones.

Unidades de análisis

La unidad de análisis es un estudiante de primer grado de educación secundaria de una institución educativa seleccionada.

Tamaño de muestra y errores muestrales

Hay que tener en cuenta los errores muestrales dado que son de vital importancia para las inferencias que realizaremos en la investigación. Los errores muestrales nos van a indicar el nivel de precisión de las inferencias a realizar. Además hay que tener en cuenta que los errores muestrales dependen mucho del presupuesto que se tenga, ya que es posible lograr minimizar el error muestral agrandando la muestra. Para obtener los errores del muestreo, teniendo en cuenta que es un muestreo complejo, se aplicaron técnicas de remuestreo, concretamente métodos Jackknife (ver Anexo A)

REGION NATURAL	N MUESTRA	E JACKKNIFFE
COSTA	721	0.0271
SIERRA	518	0.0382
SELVA	297	0.0443

Cuadro 5.3: Tamaño de muestra y errores muestrales por región natural

Consideramos el tipo de muestreo observamos que en las regiones naturales los errores muestrales son menores al 5 % con lo cual garantizamos una buena precisión en las inferencias de los estimadores en las regiones naturales.

Se consideraron en la **Costa** los siguientes departamentos: Callao, Lambayeque, Lima, Ica, La Libertad, Moquegua, Arequipa, Tacna, Ancash, Tumbes y Piura. En la **Sierra**: Junin, Puno, Apurimac, Cusco, Cajamarca, Huanuco, Ayacucho, Pasco y Huancavelica. Y en la **Selva**: Amazonas, Loreto, Madre de Dios, San Martín y Ucayali.

5.4. Recolección de información y base de datos

Para obtener la base de datos del trabajo se utilizaron fichas ópticas, en la cual los estudiantes marcan sus respuestas a cada ítem de la prueba. Para minimizar errores ya están predefinidos en la ficha óptica la región, provincia, distrito, nombre de la institución educativa, el código modular, código de la institución educativa, el grado, número de estudiantes. Con la lectura de las fichas ópticas se obtiene la base de todas las respuestas de los estudiantes evaluados.

Una vez obtenida la base de datos se realiza un control de calidad de la data, con el instrumento de medición que tiene las respuestas de los estudiantes y con listas de asistencia. Las listas de asistencia son documentos en los cuales se registran los nombres y apellidos de los estudiantes que asistieron y no asistieron a rendir la prueba, el tipo de forma que rindieron, la institución educativa, el departamento, la provincia, el distrito, la sección evaluada y el total de pruebas evaluadas en esa sección (DIGBER, 2009).

Finalmente se califica la prueba, en la cual la respuesta correcta es codificada como 1 y las incorrecta es codificada como 0; obteniéndose de este modo una base de datos dicotómica.

5.5. Análisis clásico de la prueba

Según Muñiz (1998), para analizar los ítems y la prueba se consideraron los siguientes criterios:

5.5.1. Proporción de aciertos

Con la proporción de aciertos (ACIERTOS) analizamos que tanto han acertado o no al ítem, la forma de cuantificarlo es a través de la frecuencia de respuesta a los ítems por parte de los estudiantes evaluados:

$$ACIERTOS_j = \frac{a_j}{n} \quad (5.1)$$

Donde a_j es la cantidad de estudiantes que acertaron correctamente al ítem j y "n" es el número de estudiantes evaluados.

Desde un punto de vista de la Teoría Clásica de los Test (TCT), se dice que si la proporción de aciertos es mayor o igual a 0.8 entonces el ítem ha resultado fácil al estudiante. Y cuanto la proporción de aciertos más se aproxime a 0 decimos que el ítem resultó difícil al estudiante evaluado.

5.5.2. Índice de discriminación

Se dice que un ítem tiene poder discriminativo si distingue, discriminan entre aquellos estudiantes que puntúan alto en la prueba y los que puntúan bajo. Se define como la correlación entre las puntuaciones de los estudiantes en el ítem y sus puntuaciones totales obtenido en la prueba. Como se van a correlacionar una variable dicotómica con una variable de escala de razón el tipo de correlación a utilizar es la correlación tetracórica y es definida como sigue:

$$\rho_{pj} = \frac{\mu_a - \mu_x}{\sigma_x} \sqrt{\frac{p}{q}} \quad (5.2)$$

Donde

μ_a es promedio de las puntuaciones cuando el estudian ha acertado al ítem.

μ_x es el promedio de las puntuaciones obtenidas por los estudiantes.

σ_x es el promedio de las puntuaciones obtenidas por los estudiantes.

p_j es la proporción de aciertos para ítem j.

q_j es la proporción de fallas para ítem j.

Desde una perspectiva de la teoría clásica de los test (TCT), si el índice de discriminación es menor a 0.25 entonces el ítem no discrimina.

5.5.3. Confiabilidad de la prueba

Medimos la confiabilidad de la prueba mediante el alfa de cronbach, el cual se define como:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_x^2} \right) \quad (5.3)$$

Donde: k : Número de ítems.

σ_i^2 : Varianza de los ítems.

σ_x^2 : Varianza de los puntajes totales.

Mostramos los resultados del análisis clásico en función a los criterios presentado:

ITEM	ACIERTOS	DISCRIMINACION
1	0.6302	0.3166
2	0.8164	0.4715
3	0.9408	0.4265
4	0.8392	0.5500
5	0.7826	0.5501
6	0.8750	0.4923
7	0.6426	0.4423
8	0.8503	0.4741
9	0.4089	0.2097
10	0.9466	0.3665
11	0.7617	0.4476
12	0.4564	0.3724
13	0.7806	0.4871
14	0.4915	0.3594
15	0.2969	0.3266
16	0.8320	0.4420
17	0.3626	0.3698
18	0.3177	0.3298

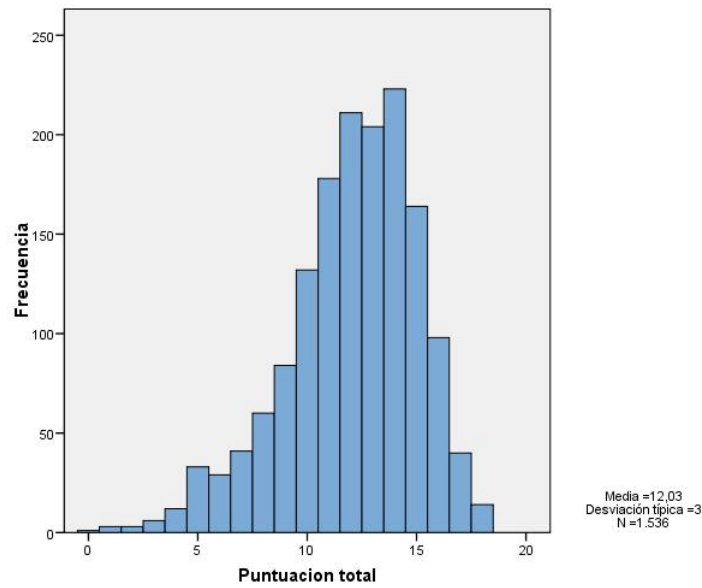
Cuadro 5.4: Proporción de acierto y discriminación de los ítems

Del cuadro 5.4, observamos que los ítems 2, 3, 4, 6, 8, 10 y 16 han sido bastante acertados por parte de los estudiantes de primero de secundaria, dado que las probabilidades de acertar al ítem son superiores a 0.8 y se acercan a 1 y los ítems 15 y 18 no han sido muy acertados por los estudiantes evaluados dado que las probabilidades de acertar al ítem se acercan a 0. Observamos que las correlaciones son mayores a 0.25, con lo cual concluimos que los ítems de la prueba discriminan. De la prueba con que se evaluaron a los estudiantes de primer grado de secundaria observamos el alfa de cronbach es 0.68795. Con lo cual podemos decir que la prueba con que fueron evaluados los estudiantes de primero de secundaria es confiable.

PUNTOS	CANTIDAD
0 puntos	1
1 puntos	3
2 puntos	3
3 puntos	6
4 puntos	12
5 puntos	33
6 puntos	29
7 puntos	41
8 puntos	60
9 puntos	84
10 puntos	132
11 puntos	178
12 puntos	211
13 puntos	204
14 puntos	223
15 puntos	164
16 puntos	98
17 puntos	40
18 puntos	14

Cuadro 5.5: Distribución de puntuaciones de los evaluados

Figura 5.1: Histograma de puntajes



ESTADISTICOS	VALOR
MEDIA	12.03
DESVIACION TIP.	3
ASIMETRIA	-0.723
ERROR TIP. DE ASIMETRIA	0.062
CURTOSIS	0.541
ERROR TIP. DE CURTOSIS	0.125

Cuadro 5.6: Estadísticos de la puntuación total

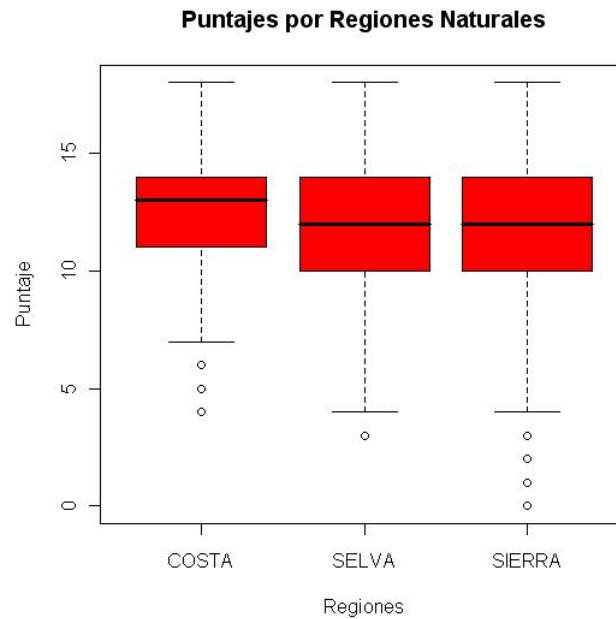
Del cuadro 5.5 y la figura 5.1, observamos que la gran mayoría de los estudiantes de primer grado de secundaria de una prueba de 18 puntos han obtenido entre 10 y 15 puntos. En el cuadro 5.6 mostramos los estadísticos de la distribución de la puntuación total obtenida por los estudiantes evaluados.

REGION NATURAL	MEDIA	DESVIACION	Q1	MEDIANA	Q3
NACIONAL	12.0319	2.9996	10	12	14
COSTA	12.4383	2.7186	11	13	14
SIERRA	11.5078	3.3348	10	12	14
SELVA	11.9600	2.8931	10	12	14

Cuadro 5.7: Puntuaciones promedio de los evaluados por Región Natural

Del cuadro 5.7 observamos que los estudiantes de las instituciones educativas estatales obtuvieron mayores puntuaciones en la Costa.

Figura 5.2: Puntajes promedios por Región Natural



De la figura 5.2, observamos que la puntuación de los estudiantes de primero de secundaria de las instituciones educativas estatales de la Costa presentan mayor concentración entre 11 y 13 puntos, mientras que los de la Sierra y Selva presentan concentración entre 10 y 14. La menor variabilidad de las puntuaciones se presentan en la Costa, y las puntuaciones mínimas en los estudiantes de la Sierra.

5.6. Estimación de parámetros bajo el modelo de Rasch

5.6.1. Estimación de las dificultades de ítems desde un punto de vista clásico

Estimaremos las habilidades de los estudiantes de primer grado de secundaria y las dificultades de los ítems de la prueba de comprensión lectora mediante los métodos de estimación de perspectiva clásica ya presentados: Métodos de Máxima Verosimilitud Marginal (MVM), Métodos de Máxima Verosimilitud Condicional (MVCOND) y Métodos de Máxima Verosimilitud Conjunta (MVC). Los resultados lo presentamos a continuación:

ITEMES	MVM	MVCOND	MVC
15	2.056	0.355	2.166
18	1.942	-2.24	2.046
17	1.707	1.455	1.799
9	1.478	1.902	1.557
12	1.248	-0.876	1.316
14	1.081	-1.209	1.14
1	0.408	0.417	0.432
7	0.345	-0.816	0.366
11	-0.33	-2.119	-0.346
13	-0.454	-0.449	-0.477
5	-0.468	1.072	-0.491
2	-0.711	-0.312	-0.748
16	-0.834	-0.972	-0.878
4	-0.893	-0.436	-0.941
8	-0.989	1.676	-1.042
6	-1.224	2.011	-1.291
3	-2.122	1.234	-2.241
10	-2.24	-0.693	-2.367

Cuadro 5.8: Estimación clásica de las dificultades de los ítems bajo el modelo de Rasch

Como se observó en el capítulo 4: estudio de simulación, cuando un número pequeño de ítems y número grande de evaluados se mostró que dentro de los métodos de estimación clásica mejor método de estimación para los ítems es el **Método de Máxima Verosimilitud Condicional (MVCOND)** y el que presentaba una asociación mas elevada era el **Método de Máxima Verosimilitud Conjunta (MVC)**, pero ambos métodos presentan problemas de estimación ante patrones de respuesta: $(0, 0, 0, \dots, 0)$ y $(1, 1, 1, \dots, 1)$ dado que sobrestiman los parámetros en el caso del método de máxima verosimilitud conjunta (MVC), y para la estimación de la habilidades de los estudiantes evaluados mediante el **Método de Máxima Verosimilitud Condicional** se necesita de métodos bayesianos. Adicionalmente se determinó que el método de estimación mas robusto es el **Método Bayesiano (MB)**, dado que estima las dificultades de los ítems y las habilidades de los alumnos evaluados sin ninguna restricciones. Es por ello nuestro interés, para esta aplicación presentar las estimaciones con el método bayesiano.

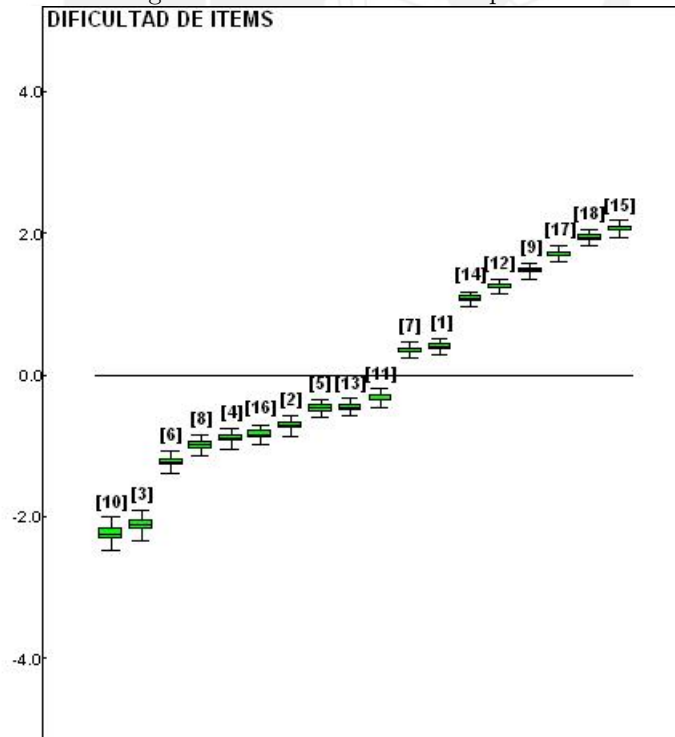
5.6.2. Estimación de las dificultades de ítems desde un punto de vista bayesiano bajo el esquema DAGS

Como se observó en el capítulo de simulación, para esta aplicación el mejor método de estimación es el método de estimación bayesiana bajo el esquema DAGS, para ello hacemos uso del software WinBugs, obteniéndose los siguientes resultados:

ITEMES	ESTIMACION	2.5 %	MEDIANA	97.5 %
15	2.05	1.933	2.049	2.169
18	1.935	1.823	1.934	2.051
17	1.703	1.593	1.703	1.815
9	1.474	1.365	1.475	1.583
12	1.245	1.137	1.246	1.352
14	1.076	0.964	1.077	1.181
1	0.402	0.288	0.403	0.513
7	0.344	0.231	0.343	0.458
11	-0.329	-0.456	-0.328	-0.207
13	-0.452	-0.578	-0.451	-0.326
5	-0.466	-0.594	-0.465	-0.341
2	-0.715	-0.852	-0.713	-0.581
16	-0.831	-0.975	-0.83	-0.69
4	-0.892	-1.031	-0.891	-0.752
8	-0.986	-1.136	-0.984	-0.838
6	-1.224	-1.375	-1.224	-1.076
3	-2.105	-2.326	-2.103	-1.901
10	-2.229	-2.465	-2.229	-2.004

Cuadro 5.9: Estimación bayesiana de las dificultades de los ítems bajo el modelo de Rasch

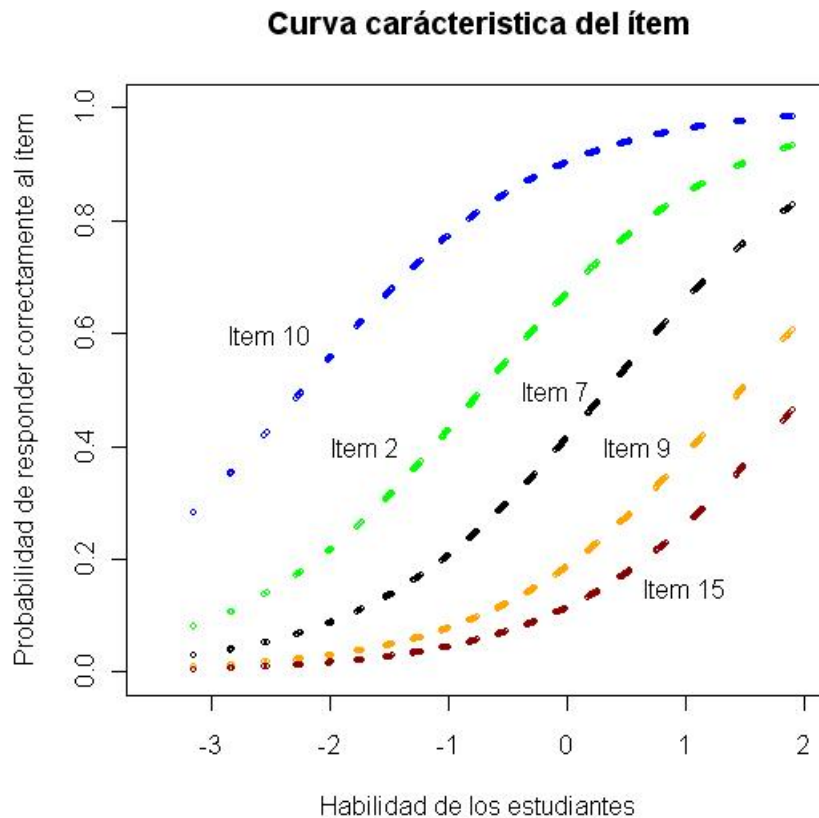
Figura 5.3: Dificultad de ítems por el MB



A través de la estimación bayesiana de las dificultades de los ítems, observamos que los ítems más difíciles son el ítems 15 y 18 y los ítems mas fáciles son los ítems 2, 16, 4, 8, 6, 3

y 10. En el gráfico de boxplot se muestra las dificultades de los ítems ordenados de los mas faciales a los mas difíciles, considerándose un nivel de dificultad de 0, observamos que 6 ítems de son los mas difíciles de la prueba de Comprensión Lectora.

Figura 5.4: CCI con habilidad Rasch Bayesiana



De la figura 5.4: Curva Característica del ítem (CCI), la curva de color azul es la del ítem 10, la verde es la del ítem 2, la negra es la del ítem 7, la naranja es la del ítem 9 y la roja es la del ítem 15, mostramos el comportamiento de los ítems desde los mas faciales hasta el mas difícil. Bajo ciertas habilidades de los estudiantes, a mas dificultad del ítem menor probabilidad de resolverlo correctamente. Observamos que para los estudiantes de primero de secundaria, tienen mayor probabilidad de resolver correctamente el ítem 10 y menor probabilidad el ítem 15.

Presentamos algunos ítems, en la cual analizamos sus curvas características:

La llegada de un nuevo bebé

El Parque de Las Leyendas, conocido zoológico de la ciudad de Lima, se enorgullece en anunciar el nacimiento de un hipopótamo el domingo 16 del presente. Esto es una clara muestra del trabajo que el parque viene realizando a favor de la preservación de especies a través de su programa de reproducción en cautiverio. Los hipopótamos "Pipo" y "Carlota", traídos de Cuba en octubre de 2003, son los padres del nuevo engreído del parque y estos, así como su cría, son evaluados constantemente en cuanto a su salud y bienestar, respetando la naturaleza animal de mantener juntos a la madre y la cría. Sumada a esta buena noticia para el parque, cabe destacar además el nacimiento de tres otorongos y un tapir, especies oriundas de nuestra selva, que también son resultado de la constante labor que realizan nuestros profesionales y expertos en zoología y veterinaria en el cuidado de cada ejemplar. El Parque de Las Leyendas invita al público en general a visitar de sus instalaciones para poder apreciar al bebé hipopótamo, a los cachorros de otorongo y al pequeño tapir, además de visitar el resto de la colección zoológica, entre la que destacan las nuevas jirafas y un león, traídos directamente desde África el día 28 del mes de abril, y un elefante blanco de la India, además de la flora y el patrimonio arqueológico del Perú.

Ítem 10: ¿De dónde fueron traídos los padres del bebé hipopótamo?

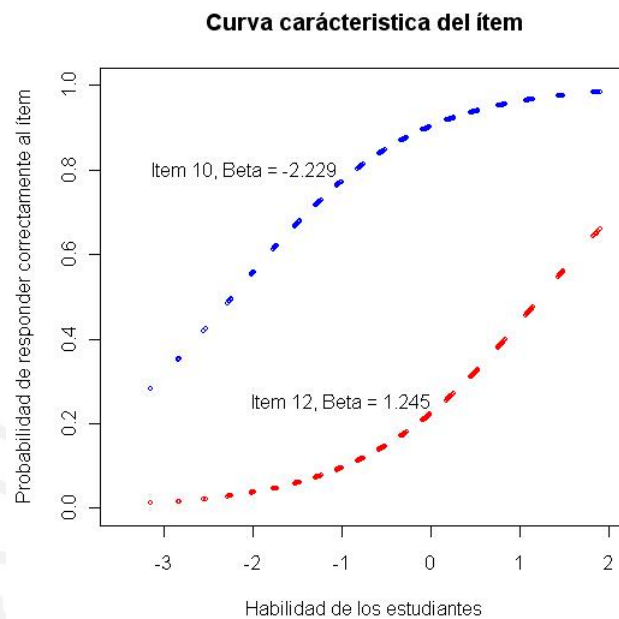
- A. De la selva peruana
- B. De Cuba
- C. De África
- D. De la India

Ítem 12. ¿Para qué fue escrito principalmente el texto anterior?

- A. Para anunciar el nacimiento de nuevos ejemplares en el parque e invitar al público a visitarlos.
- B. Para explicar que es muy importante mantener a las crías recién nacidas junto a su madre.
- C. Para informar sobre las características de los hipopótamos y su forma de reproducción en cautiverio.
- D. Para informar a los lectores sobre la importancia de cuidar la naturaleza y preservar las especies en peligro.

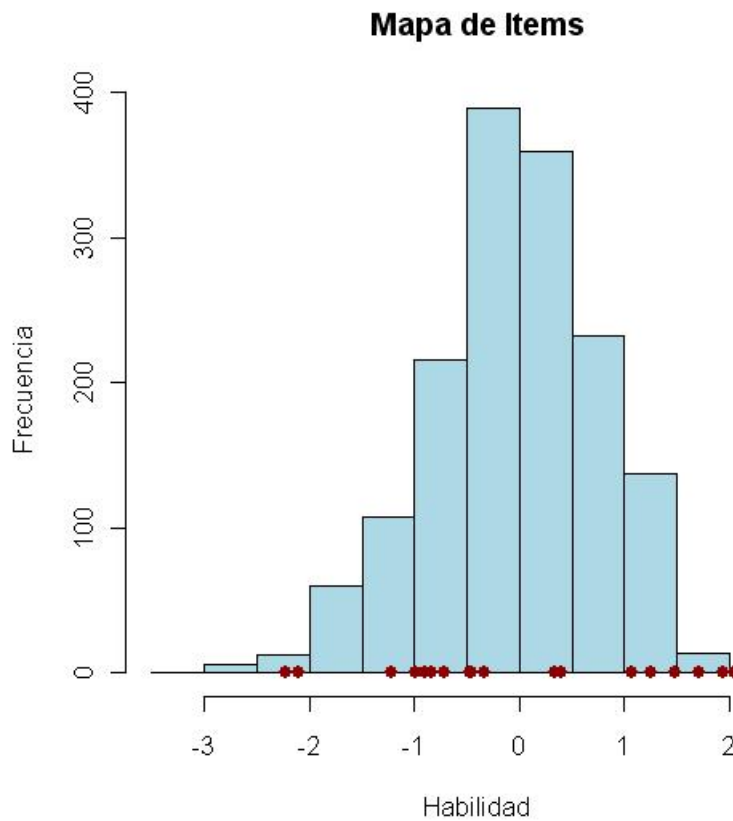
El ítem 10 refleja la capacidad que tienen los estudiantes de primero de secundaria de los colegios estatales para obtener información explícita del texto, es decir obtiene información ubicada en un párrafo del texto. Y el ítem 12 refleja la capacidad para reflexionar acerca del texto, es decir reflexiona acerca del propósito que tuvo el autor al escribir el texto o al incluir alguna de sus partes.

Figura 5.5: Curvas características de ítems 10 y 12



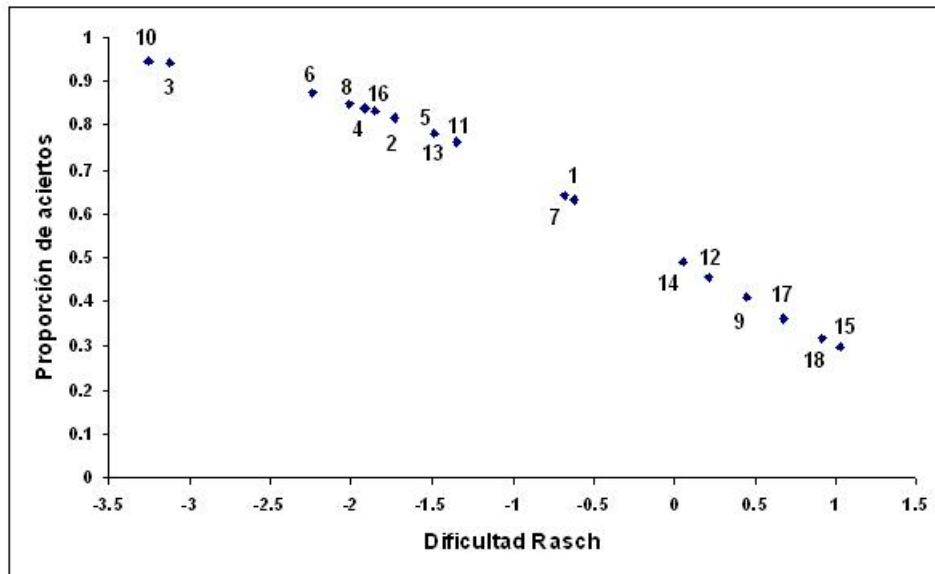
De la figura 5.5, la curva de color azul representa al ítem 10, curva la de color rojo representa al ítem 12, observamos que el ítem 12 es más difícil, dado que tiene menor probabilidad de acierto que la del ítem 10.

Figura 5.6: Mapa de ítems



De la figura 5.6: mapa de ítems, observamos que la mayoría de los estudiantes evaluados de primero de secundaria de las instituciones estatales educativas, tienen las habilidades suficiente para resolver la prueba favorablemente, aproximadamente el 48 % de los estudiantes tienen las habilidades y capacidades para resolver satisfactoriamente los ítems más difíciles de la prueba de comprensión lectora. Podemos inferir que para los estudiantes de primero de secundaria la prueba de comprensión lectora no fue muy difícil, tal como se mostró en el análisis clásico de la prueba.

Figura 5.7: Asociación entre la dificultad Rasch y la proporción de aciertos

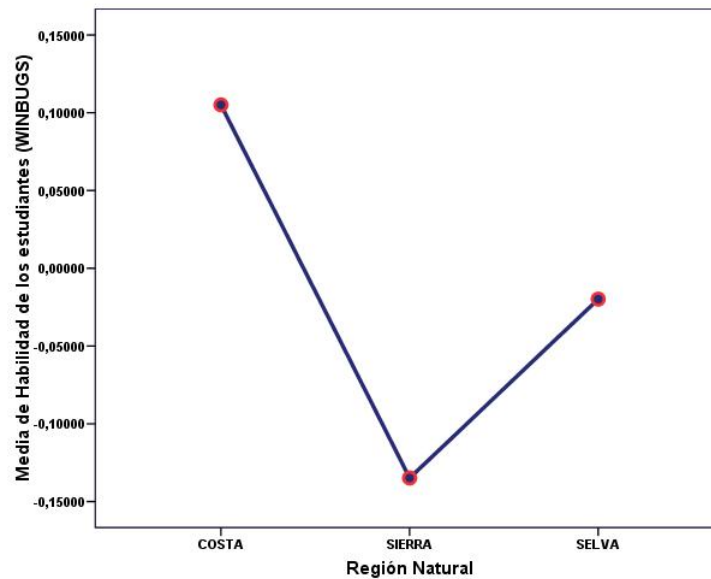


De la figura 5.7, observamos que existe una relación lineal inversa entre la proporción de acierto y la dificultad Rasch, a mayor dificultad Rasch corresponde menor proporción de aciertos.

5.6.3. Estimación de las habilidades de los estudiantes desde un punto de vista bayesiano bajo el esquema DAGS

Para la estimación bayesiana de las habilidades de los estudiantes, hacemos uso del software WinBugs. Debido a las diferencias puntuaciones obtenidas por los estudiantes de primero de secundaria de los colegios estatales es de nuestro interés analizar el comportamiento de sus habilidades en las regiones naturales (costa, sierra y selva):

Figura 5.8: Habilidad promedio de los estudiantes por el MB según Regiones Naturales



De la figura 5.8, observamos que tienen mayor habilidad promedio los estudiantes de la Costa que los de la Sierra y la Selva. Además la habilidad promedio de los estudiantes de la Selva es ligeramente mayor a los de la Sierra. En función a lo mostrado en el gráfico analizamos si las variaciones de las habilidades promedios entre las regiones naturales es estadísticamente significativa:

LEVENE	GL1	GL2	SIG
5.933	2	1533	0.0027

Cuadro 5.10: Prueba de homogeneidad de varianza de las habilidades de los estudiantes según regiones naturales

Del cuadro 5.10, mostramos el p value de la prueba de homogeneidad de varianzas, la cual es menor a 0.05, concluimos que a un nivel de significancia del 5 % existe evidencia estadística suficiente para afirmar existen diferencias significativas en las varianzas de las habilidades de los estudiante en cada region natural evaluada.

	SC	GL	CME	F	SIG
Inter-grupos	17.497	2	8.749	14.189	0.0000
Intra-grupos	945.180	1533	0.617		
Total	962.677	1535			

Cuadro 5.11: ANOVA de la habilidades de los estudiantes según regiones naturales

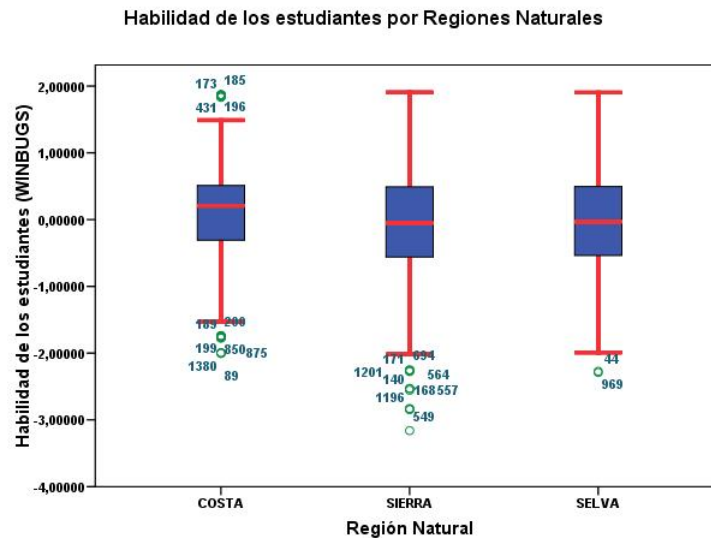
Del cuadro 5.11 mostramos que el p value de la prueba es menor a 0.05, concluimos que a un nivel de significancia del 5 % existe evidencia estadística suficiente para afirmar existen diferencias significativas en las habilidades promedios de los estudiantes evaluados en cada region natural. Como hay diferencias significativas en las varianzas y en los promedios de

las habilidades entre cada región natural, es de nuestro interés mostrar entre que regiones naturales existen diferencias significativas en la habilidad promedio mediante comparaciones múltiples pos hoc de Games Howell:

(I) REGION	(J) REGION	DIF MEDIAS (I-J)	SIG
COSTA	SIERRA	0.2393	0.000
(0.0271)	SELVA	0.1249	0.044
SIERRA	COSTA	-0.2399	0.000
(0.0382)	SELVA	-0.1150	0.122
SELVA	COSTA	-0.1249	0.044
(0.0444)	SIERRA	0.1150	0.122

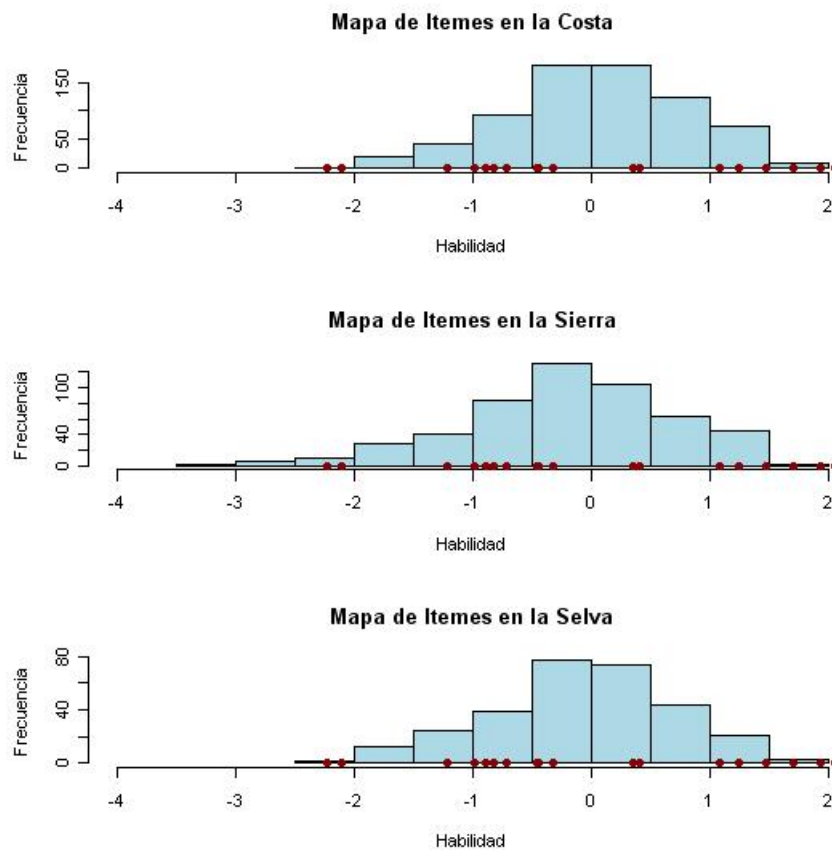
Cuadro 5.12: Comparaciones múltiples de las habilidades de los estudiantes entre las regiones naturales

Figura 5.9: Habilidad de los estudiantes por el MB según regiones naturales



En el cuadro 5.12 mostramos comparaciones pareadas de las habilidades promedios de los estudiantes entre las regiones naturales con su respectivo error estándar da cada región. Del cuadro 5.12 y la figura 5.9 concluimos que a un nivel de significancia del 5% existe evidencia estadística suficiente para afirmar que existen diferencias significativas entre las habilidades promedios de los estudiantes de la Costa con las otras regiones naturales, mostrando la costa las mayores habilidades promedios.

Figura 5.10: Mapa de ítems por Regiones Naturales

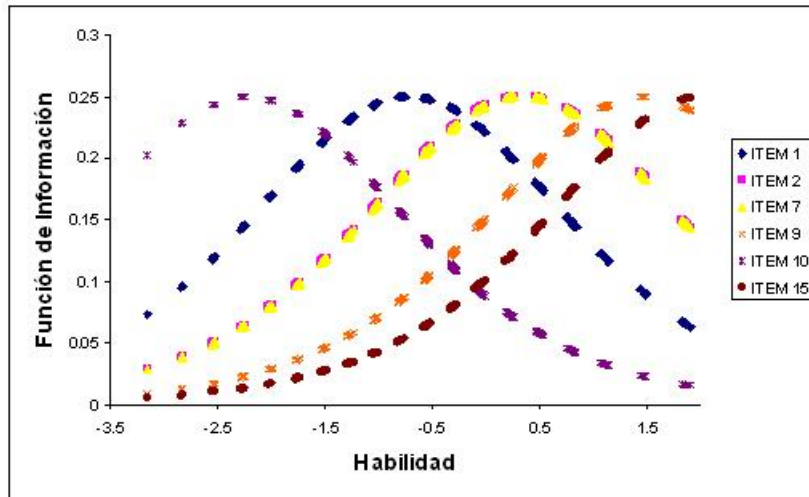


De la figura 5.10: gráfico de Mapa de ítems por Regiones Naturales, observamos que en la Sierra a pesar de que los estudiantes de primero de secundaria de las instituciones educativas estatales tienen menores habilidades promedio que los de la Costa y la Selva, están en la capacidad de resolver los ítems de la prueba inclusive los más difíciles pero; en menor proporción que los de las otras regiones naturales. La mayoría de estudiantes de la Sierra tienen capacidades para resolver los ítems más fáciles de la prueba.

5.6.4. Función de información y de la prueba para la aplicación

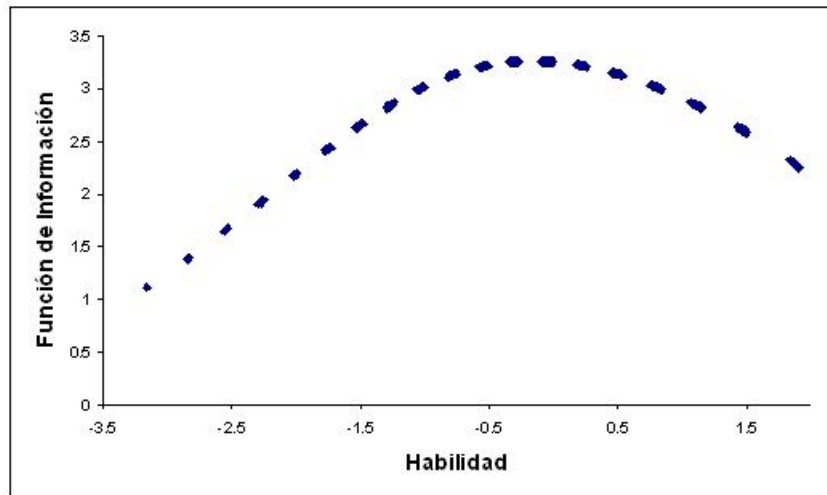
Mostramos a continuación la función de información para los ítems 2, 7, 9, 10, 15; de los cuales observamos que a menor dificultad mayor información del ítem.

Figura 5.11: Funciones de información de los ítems



A la vez mostramos la función de información de la prueba utilizada en la aplicación.

Figura 5.12: Funciones de información de la prueba



Capítulo 6

Conclusiones

6.1. Conclusiones

- La tesis presenta el Modelo de Rasch dicotómico y los métodos de estimación desde un punto de vista clásico y bayesiano empleando diferentes software disponibles en la literatura con énfasis en el software libre.
- Para estimar las habilidades de los estudiantes con el método de estimación marginal y condicional hay que aplicar procedimientos complementarios de estimación. En cambio con el método de máxima verosimilitud conjunta y el método bayesiano se estiman a la vez las habilidades de los estudiantes y la dificultad de los ítems.
- La estimación de las dificultades de los ítems incluyendo la distribución de probabilidad de habilidades y dificultades, por el método de máxima verosimilitud marginal y conjunta presentan problemas cuando los estudiantes han respondido correctamente o incorrectamente todos los ítems de la prueba, en tal sentido con estos métodos de estimación en esos casos no se pueden estimar los parámetros.
- Observamos que el Método Bayesiano es el mejor método de estimación bajo los diferentes escenarios presentados en el capítulo de estudios de simulación, es decir considerando un número pequeño de ítems y/o número grande de ítems y tamaño de muestra variable. El método bayesiano es el método más robusto, dado que estima las habilidades y dificultades de los ítems sin ningún tipo de restricciones como el caso de presencia de puntajes extremos de ítems y personas así como presenta menor error cuadrático y medio y mayor correlación con los datos simulados lo que indica que presenta menor sesgo. Sin embargo en general demanda un tiempo mayor que otros métodos de estimación.
- Del análisis clásico de los ítems en la aplicación, resultaron que los ítems 2, 3, 4, 6, 8, 10 los más acertados, y los ítems 15 y 18 los menos acertados. Todos los ítems de la prueba discriminan y este tiene un índice de confiabilidad de 0.68795, con lo cual la prueba es confiable.
- En la aplicación, mediante las estimación bayesiana de las dificultades, se concluye que

los ítems 15 y 18 son los más difíciles. Los ítems 2, 16, 4, 8, 6, 3, 10 son los ítems más fáciles.

- Existe una relación lineal entre la proporción de acierto y la dificultad Rasch. Desde la perspectiva bayesiana es posible tener intervalos de confianza para las dificultades lo que permite una mejor inferencia.
- Las habilidades promedios de los estudiantes evaluados de primero de secundaria de las instituciones educativas estatales de la Costa difiere significativamente y presentan mayores habilidades promedios que los estudiantes de la Sierra y la Selva. Tienen las capacidades de resolver la prueba de Comprensión Lectora, incluyendo a los ítems más difíciles.
- En la aplicación se observó que la estimación de habilidades de los estudiantes con el método bayesiano (MB) presenta mayor función de información que con el Método de Máxima Verosimilitud Conjunta (MVC); es decir, que alberga la mayor cantidad de información posible, con lo cual el error de estimación es menor.

6.2. Sugerencias para investigaciones futuras

- Para investigaciones futuras, para el modelo hay que tener en cuenta otros factores como las respuestas al azar de los estudiantes, la estimación en situaciones de pruebas rotadas o pruebas administradas bajo diseños incompletos así como en el caso de datos faltantes o no respuesta.
- Existen peligros de usar softwares que aplican métodos de estimación diferentes al Método Bayesiano, dado que pueden presentar sesgos e inconvenientes al estimar los parámetros cuando el número de ítems o personas son pequeños. No obstante, existe una mayor demanda computacional en la estimación bayesiana para muestras grandes.
- En la aplicación se observó que la distribución de las puntuaciones totales es asimétrica, para este caso se pueden estudiar los Modelos de Teoría de Respuesta al Ítem Sesgados.
- Se pueden realizar otros estudios de simulación complementarios para analizar el efecto de las distribuciones a prioris.
- Se pueden estudiar de manera complementaria diversos procedimientos para evaluar el ajuste de ítems y de modelos incluyendo procedimientos de análisis de residuales.
- Se puede analizar el uso de otros modelos de Teoría de Respuesta al Ítem para la aplicación presentada como los presentados en (Bazán J., Valdivieso L. y Calderón A. (2010)) .
- Alguna entidad pública o privada deberían aplicar políticas que mejoren el rendimiento en Comprensión lectora de los estudiantes de primero de secundaria de las instituciones educativas estatales de la Sierra y la Selva.

- Alguna entidad publica o privada deberían plantear ítems más difíciles con el objeto de lograr mayor exigencia en las capacidades de los estudiantes de primero de secundaria de las instituciones educativas estatales.



Apéndice A

Anexos A

A.1. Suficiencia estadística

Una estadística $T = g(X_1, X_2, X_3, \dots, X_n)$ es suficiente para el parámetro θ , si $f_{X_1, \dots, X_n | T=t}(\cdot)$ no depende de θ , para cada t valor posible de T .

A.1.1. Teorema de Factorización de Neyman

Tomando la literatura (Hogg R. V. and Craig A.T.,(1978)) y (Tanner M.A.,(1995))
Sea $(x_1, x_2, x_3, \dots, x_n)$ una muestra aleatoria de tamaño n que tiene una distribución $f(x, \theta)$, $\theta \in \Omega$. Sea $Y_1 = u_1(x_1, x_2, x_3, \dots, x_n)$ una estadística cuya función de distribución de probabilidad es $g_1(Y_1, \theta)$. Entonces Y_1 es una estadística suficiente para todo θ si y solamente si

$$\frac{f(x_1, \theta)f(x_2, \theta)\dots f(x_n, \theta)}{g_1(u_1(x_1, x_2, x_3, \dots, x_n), \theta)} = H(x_1, x_2, x_3, \dots, x_n)$$

donde $H(x_1, x_2, x_3, \dots, x_n)$ no depende de $\theta \in \Omega$.

Una estadística $T = g(X_1, X_2, \dots, X_n)$ es suficiente para el parámetro θ si y solo si, existen funciones "h" independiente de θ y "l" tales que $\forall(x_1, x_2, \dots, x_n), \forall\theta$:

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = h(x_1, x_2, \dots, x_n)l(g(x_1, x_2, \dots, x_n), \theta)$$

Suficiencia de Γ_i

Partiendo del supuesto de independencia, $\forall\beta$ fijo

$$P(X_{ij}|\theta_i\beta) = \prod_{j=1}^k P(X_{ij}|\theta_i\beta_j)$$

$$P(X_{ij}|\theta_i\beta) = \prod_{j=1}^k \frac{e^{X_{ij}(\theta_i-\beta_j)}}{1 + e^{(\theta_i-\beta_j)}}$$

Resolviendo

$$P(X_{ij}|\theta_i\beta) = \frac{e^{\sum_{j=1}^k X_{ij}\theta_i - \sum_{j=1}^k X_{ij}\beta_j}}{\prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]}$$

Reemplazamos por $\Gamma_i = \sum_{j=1}^k X_{ij}$

$$P(X_{ij}|\theta_i\beta) = \frac{e^{\theta_i\Gamma_i - \sum_{j=1}^k X_{ij}\beta_j}}{\prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]}$$

Desdoblado:

$$P(X_{ij}|\theta_i\beta) = \frac{e^{\theta_i\Gamma_i}}{\prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]} e^{-\sum_{j=1}^k X_{ij}\beta_j}$$

Hacemos :

$$l(g(\Gamma_i|\theta_i)) = \frac{e^{\theta_i\Gamma_i}}{\prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]}$$

$$h(x) = e^{-\sum_{j=1}^k X_{ij}\beta_j}$$

Con lo cual tenemos:

$$P(X_{ij}|\theta_i\beta) = h(x)l(g(\Gamma_i|\theta_i))$$

Por el teorema de Factorización de Neyman concluimos que:

$\Gamma_i = \sum_{j=1}^k X_{ij}$ es una estadística suficiente de θ_i

Suficiencia de Z_j

Partiendo de la función de verosimilitud del modelo, $\forall\theta$ fijo:

$$L(\mathbf{X}|\theta, \beta) = \frac{e^{\sum_{i=1}^n \sum_{j=1}^k X_{ij}\theta_i - \sum_{i=1}^n \sum_{j=1}^k X_{ij}\beta_j}}{\prod_{i=1}^n \prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]}$$

Reemplazamos por:

$$Z_j = \sum_{i=1}^n y_{ij}$$

Se tiene entonces:

$$L(\mathbf{X}|\theta, \beta) = \frac{e^{\sum_{i=1}^n \sum_{j=1}^k X_{ij}\theta_i - \sum_{j=1}^k Z_j\beta_j}}{\prod_{i=1}^n \prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]}$$

Desdoblado

$$L(\mathbf{X}|\theta, \beta) = e^{\sum_{i=1}^n \sum_{j=1}^k X_{ij}\theta_i} \frac{e^{-\sum_{j=1}^k Z_j\beta_j}}{\prod_{i=1}^n \prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]}$$

Hacemos:

$$h(X) = e^{\sum_{i=1}^n \sum_{j=1}^k X_{ij}\theta_i}$$

$$l(g(Z_j|\theta_i)) = \frac{e^{-\sum_{j=1}^k Z_j\beta_j}}{\prod_{i=1}^n \prod_{j=1}^k [1 + e^{(\theta_i - \beta_j)}]}$$

Por el teorema de Factorización de Neyman concluimos que:

$Z_j = \sum_{i=1}^n X_{ij}$ es una estadística suficiente de β_j

A.2. Formulas de cuadratura de Gauss Hermite

Las formulas de cuadratura de Gauss Hermite son utilizados para aproximar mediante una sumatoria a integrales indefinidas (Qing L y Pierce D (1994)). En el espacio de la integral $\langle -\infty; \infty \rangle$ se toman n particiones de igual tamaño y se aproxima la integral de la siguiente forma:

$$\int_{-\infty}^{\infty} e^{x^2} f(x) dx \approx \sum_{i=1}^n c_i f(x_i)$$

Es decir la integral se aproxima mediante la combinación lineal de la función donde:

$$C_h = \frac{2^{(n-1)} n! \sqrt{\pi}}{n^2 [H_{(n-1)}(\theta_i)]}$$

$$\forall i = 1, 2, 3, \dots, n$$

Y $H_n(x)$ es el polinomio de Hermite el cual tienen "n" raíces reales y distintas y forman una familia ortogonal en $\langle -\infty; \infty \rangle$

A.2.1. Polinomios de Hermite

Definimos los polinomios de Hermite por:

$$H_n(x) = (-1)^n e^{t^2} \frac{d^n}{dx^n} e^{-t^2}$$

$H_n(x)_{n \in \mathbb{N}}$ son polinomios de grado n. Se tiene que:

$$H_n(-x) = (-1)^n H_n(x)$$

es decir, H_n es par si n es par, e impar si n es impar.

A.2.2. Relaciones entre los polinomios de Hermite

Podemos establecer algunas relaciones entre los polinomios de Hermite:

$$H_0(x) = 1$$

$$H_1(x) = 2x$$

$$H_2(x) = 4x^2 - 2$$

$$H_3(x) = 8x^3 - 12x$$

$$H_4(x) = 16x^4 - 48x^2 + 12$$

$$H_{n+2}(x) = 2xH_{n+1}(x) - 2(n+1)H_n(x)$$

A.3. Métodos computaciones de aproximación: Newton Raph- son

Considerando la literatura de (Ross, S.M. (1990)) :

Definimos la función score:

$$g(\theta) = \frac{\partial \log L(\theta, x)}{\partial \theta}$$

Resolvemos la función score igualando a 0 obteniendo θ^* el punto máximo local tal que:

$$\frac{\partial \log(\theta_h), x}{\partial \theta_h} < 0$$

Para resolver $g(\theta)$ se utilizan aproximaciones de Taylor:

$$g(\theta) = g(\theta_0) + (\theta - \theta_0)g'(\theta_0) + (\theta - \theta_0)^2g''(\theta_0) + \dots$$

$$g(\theta) - g(\theta_0) = \sum_{i=1}^{\infty} (\theta - \theta_0)^i g^i(\theta_0)$$

$$g(\theta) \approx g(\theta_0) + (\theta - \theta_0)g'(\theta_0)$$

Si consideramos a θ_h un máximo local de $L(\theta, x)$:

$$g(\theta_h) \approx g(\theta_0) + (\theta_h - \theta_0)g'(\theta_0)$$

$$g(\theta_0) + (\theta_h - \theta_0)g'(\theta_0) = 0$$

$$\theta_h = \theta_0 - g(\theta_0)[g'(\theta_0)]^{-1}$$

$$\theta_j = \theta_{j-1} - \frac{g(\theta_{j-1})}{g'(\theta_{j-1})}$$

El algoritmo continua hasta que $|\theta_k - \theta_{k-1}| < \epsilon$ donde ϵ es una cantidad pequeña.

A.4. Método de Jackniffe

Teniendo en cuenta la literatura, ver (Efron B. and Tibshirani R.J.,(1993)): Supongamos que se tiene una muestra $X = (x_1, x_2, x_3, \dots, x_n)$ y un estimador $\hat{\theta} = s(x)$. Deseamos estimar el error estándar de $\hat{\theta}$. El método de Jackniffe se enfoca sobre muestras en la cual se deja de lado una observación en un instante dado:

$$X_{(i)} = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$$

para $i = 1, 2, \dots, n$ es llamada muestra Jackniffe. $\theta_{(i)} = s(x_{(i)})$ es la i -ésima replicación jackniffe de $\hat{\theta}$.

El error estándar mediante el método de Jackniffe se define como:

$$S(\hat{\theta})_{jack} = \sqrt{\frac{n-1}{n} \sum (\theta_{(i)} - \theta_{(.)})^2}$$

Donde:

$$\theta_{(.)} = \frac{\sum_{i=1}^n \hat{\theta}_{(i)}}{n}$$

A.5. Algoritmo Gibbs Sampling

Métodos MCMC obtienen una muestra de parámetros estimados a partir de la distribución a posteriori. Esta muestra la usamos para estimar los parámetros θ y β (Tanner, M.A (1995)).

El algoritmo consististe:

- Fijamos un número valores n cantidad de personas a evaluar y "k" cantidad de ítems del instrumento de medición.
- Fijamos un número m de iteraciones en el algoritmo.

- Consideramos a

$$\theta = (\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_n^{(0)})^T$$

$$\beta = (\beta_1^{(0)}, \beta_2^{(0)}, \beta_3^{(0)}, \dots, \beta_k^{(0)})^T$$

- Simulamos en cada una de las iteraciones:

Primera iteración:

$$\theta_1^{(1)} \sim P(\theta_1 | \theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_n^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}, \beta_3^{(0)}, \dots, \beta_k^{(0)}, Y)$$

$$\theta_2^{(1)} \sim P(\theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_n^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}, \beta_3^{(0)}, \dots, \beta_k^{(0)}, Y)$$

$$\theta_n^{(1)} \sim P(\theta_n | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{n-1}^{(1)}, \beta_1^{(0)}, \beta_2^{(0)}, \beta_3^{(0)}, \dots, \beta_k^{(0)}, Y)$$

$$\beta_1^{(1)} \sim P(\beta_1 | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_n^{(1)}, \beta_2^{(0)}, \beta_3^{(0)}, \beta_4^{(0)}, \dots, \beta_k^{(0)}, Y)$$

$$\beta_2^{(1)} \sim P(\beta_2 | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_n^{(1)}, \beta_1^{(1)}, \beta_3^{(0)}, \beta_4^{(0)}, \dots, \beta_k^{(0)}, Y)$$

$$\beta_k^{(1)} \sim P(\beta_k | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{n-1}^{(1)}, \beta_1^{(1)}, \beta_2^{(1)}, \beta_3^{(1)}, \dots, \beta_{k-1}^{(1)}, Y)$$

Segunda iteración:

$$\theta_1^{(2)} \sim P(\theta_1 | \theta_2^{(1)}, \theta_3^{(1)}, \dots, \theta_n^{(1)}, \beta_1^{(1)}, \beta_2^{(1)}, \beta_3^{(1)}, \dots, \beta_k^{(1)}, Y)$$

$$\theta_2^{(2)} \sim P(\theta_2 | \theta_1^{(2)}, \theta_3^{(1)}, \dots, \theta_n^{(1)}, \beta_1^{(1)}, \beta_2^{(1)}, \beta_3^{(1)}, \dots, \beta_k^{(1)}, Y)$$

$$\theta_n^{(2)} \sim P(\theta_n | \theta_1^{(2)}, \theta_2^{(2)}, \dots, \theta_{n-1}^{(2)}, \beta_1^{(1)}, \beta_2^{(1)}, \beta_3^{(1)}, \dots, \beta_k^{(1)}, Y)$$

$$\beta_1^{(2)} \sim P(\beta_1 | \theta_1^{(2)}, \theta_2^{(2)}, \dots, \theta_n^{(2)}, \beta_2^{(1)}, \beta_3^{(1)}, \beta_4^{(1)}, \dots, \beta_k^{(1)}, Y)$$

$$\beta_2^{(2)} \sim P(\beta_2 | \theta_1^{(2)}, \theta_2^{(2)}, \dots, \theta_n^{(2)}, \beta_1^{(2)}, \beta_3^{(1)}, \beta_4^{(1)}, \dots, \beta_k^{(1)}, Y)$$

$$\beta_k^{(2)} \sim P(\beta_k | \theta_1^{(2)}, \theta_2^{(2)}, \dots, \theta_{n-1}^{(2)}, \beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}, \dots, \beta_{k-1}^{(2)}, Y)$$

Tercera iteración:

$$\theta_1^{(3)} \sim P(\theta_1 | \theta_2^{(2)}, \theta_3^{(2)}, \dots, \theta_n^{(2)}, \beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}, \dots, \beta_k^{(2)}, Y)$$

$$\theta_2^{(3)} \sim P(\theta_2 | \theta_1^{(3)}, \theta_3^{(2)}, \dots, \theta_n^{(2)}, \beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}, \dots, \beta_k^{(2)}, Y)$$

$$\theta_n^{(3)} \sim P(\theta_n | \theta_1^{(3)}, \theta_2^{(3)}, \dots, \theta_{n-1}^{(3)}, \beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}, \dots, \beta_k^{(2)}, Y)$$

$$\beta_1^{(3)} \sim P(\beta_1 | \theta_1^{(3)}, \theta_2^{(3)}, \dots, \theta_n^{(3)}, \beta_2^{(2)}, \beta_3^{(2)}, \beta_4^{(2)}, \dots, \beta_k^{(2)}, Y)$$

$$\beta_2^{(3)} \sim P(\beta_2 | \theta_1^{(3)}, \theta_2^{(3)}, \dots, \theta_n^{(3)}, \beta_1^{(3)}, \beta_3^{(2)}, \beta_4^{(2)}, \dots, \beta_k^{(2)}, Y)$$

$$\beta_k^{(3)} \sim P(\beta_k | \theta_1^{(3)}, \theta_2^{(3)}, \dots, \theta_{n-1}^{(3)}, \beta_1^{(3)}, \beta_2^{(3)}, \beta_3^{(3)}, \dots, \beta_{k-1}^{(3)}, Y)$$

m-esima iteración

$$\theta_1^{(m)} \sim P(\theta_1 | \theta_2^{(m-1)}, \theta_3^{(m-1)}, \dots, \theta_n^{(m-1)}, \beta_1^{(m-1)}, \beta_2^{(m-1)}, \beta_3^{(m-1)}, \dots, \beta_k^{(m-1)}, Y)$$

$$\theta_2^{(m)} \sim P(\theta_2 | \theta_1^{(m)}, \theta_3^{(m-1)}, \dots, \theta_n^{(m-1)}, \beta_1^{(m-1)}, \beta_2^{(m-1)}, \beta_3^{(m-1)}, \dots, \beta_k^{(m-1)}, Y)$$

$$\theta_n^{(m)} \sim P(\theta_n | \theta_1^{(m)}, \theta_2^{(m)}, \dots, \theta_{n-1}^{(m)}, \beta_1^{(m-1)}, \beta_2^{(m-1)}, \beta_3^{(m-1)}, \dots, \beta_k^{(m-1)}, Y)$$

$$\beta_1^{(m)} \sim P(\beta_1 | \theta_1^{(m)}, \theta_2^{(m)}, \dots, \theta_n^{(m)}, \beta_2^{(m-1)}, \beta_3^{(m-1)}, \beta_4^{(m-1)}, \dots, \beta_k^{(m-1)}, Y)$$

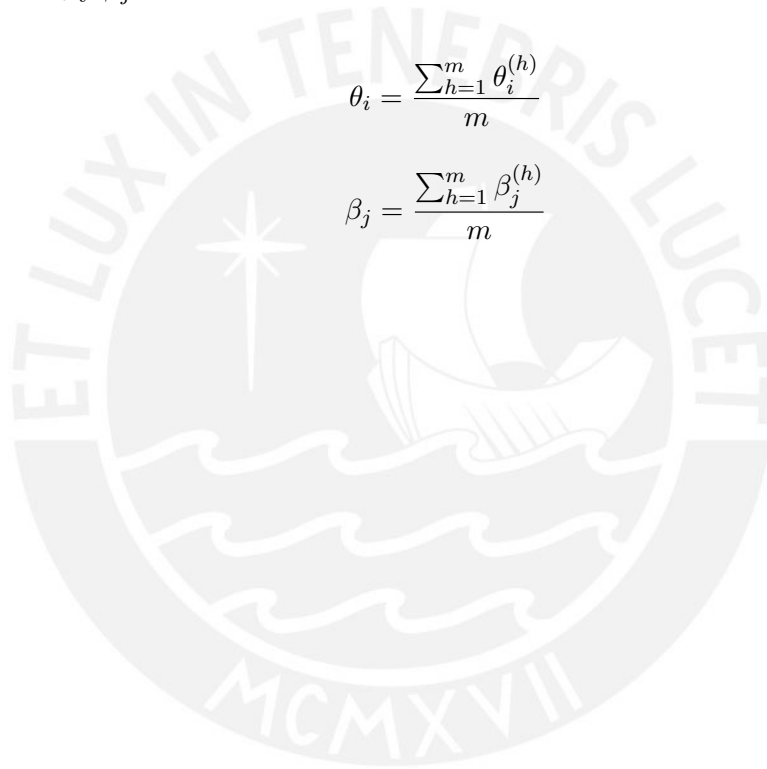
$$\beta_2^{(m)} \sim P(\beta_2 | \theta_1^{(m)}, \theta_2^{(m)}, \dots, \theta_n^{(m)}, \beta_1^{(m)}, \beta_3^{(m-1)}, \beta_4^{(m-1)}, \dots, \beta_k^{(m-1)}, Y)$$

$$\beta_k^{(m)} \sim P(\beta_k | \theta_1^{(m)}, \theta_2^{(m)}, \dots, \theta_{n-1}^{(m)}, \beta_1^{(m)}, \beta_2^{(m)}, \beta_3^{(m)}, \dots, \beta_{k-1}^{(m)}, Y)$$

- Estimamos θ_i y β_j :

$$\theta_i = \frac{\sum_{h=1}^m \theta_i^{(h)}}{m}$$

$$\beta_j = \frac{\sum_{h=1}^m \beta_j^{(h)}}{m}$$



Apéndice B

Anexos B

B.1. Programa para generar una base de datos dicotómica

```

#programa simula rasch
#####
simrasch <-function(n,k){

#seed(      )
set.seed(100) #semilla para reproducir los mismos valores
b=rnorm(k,0,1)
#optional correction for sum be 0
b=b-mean(b)

YY = matrix(0,n,k);
Y = matrix(0,n,k);

bmat = matrix(1,n,1)%*(b);

# simulation of u
u=rnorm(n);

umat =matrix(rep(u,k),n,k);
#optional correction for sum be 0
u=u-mean(u)

eta = umat - bmat;
P=exp(eta)/(1+exp(eta)) # Rasch model
# N x K probability matrix of correct answers.
#P=logis(eta)

R = matrix(runif(n*k),n,k);
Y = 1*(R > 1-P)
y = list(Y,u,b)
}

#programa principal
#####
#Especificar los valores de items y personas
k=100
n=15

y=simrasch(n,k)

datas = y[[1]]
usim=y[[2]]
bsim=y[[3]]

```

Figura B.1: Programa para generar una base de datos dicotómica

B.2. Programa para graficar una Curva Característica del Ítem, Función de Información y Mapa de ítems

```

#Programa para graficar
#####
bsitem1 = -2
bsitem2 = 0
bsitem3 = 2
u=seq(-6,6,by=0.01)
p1=plogis(u-bsitem1)
p2=plogis(u-bsitem2)
p3=plogis(u-bsitem3)
I1=p1*(1-p1)
I2=p1*(1-p2)
I3=p1*(1-p3)
I=matrix((I1+I2+I3),length(u),1)
theta=matrix(u,length(u),1)
datainfo=cbind(theta,I)

#Curva Caracteristica del Item
#####
plot(u,p1, type="l",main="Curva característica del
item",xlab=expression(theta),ylab="Probabilidad de responder correctamente al
item",col="dark blue")
text(-3, 0.6, "Beta =-2",cex = 1.0)
lines(u,p2, type="l", col="dark green")
text(-1, 0.5, "Beta =0",cex = 1.0)
lines(u,p3, type="l", col="dark red")
text(3, 0.5, "Beta =2",cex = 1.0)

#Función de Información de los ítems
#####
plot(u,I1, type="l",main="Función de Información del
item",xlab=expression(theta),ylab="Función de
información",ylim=c(0,0.8),col="dark blue")
lines(u,I2, type="l", col="dark green")
lines(u,I3, type="l", col="dark red")

#Función de Información de la prueba
#####
plot(u,I1+I2+I3, type="l",main="Función de información de la
prueba",xlab=expression(theta),ylab="Función de información",col="black")

#Mapa de Ítems
#####
habilidad_wb<-read.table("D:/HABILIDAD_WB.csv",sep=",")
dificultad_wb<-read.table("D:/DIFICULTAD_WB.csv",sep=",")
beta_wb=dificultad_wb[,2]
theta=habilidad_wb[,2]
Y=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
hist(theta,xlab='Meeasure',ylab="Frecuencia",main="Mapa de Items",col="light
blue",xlim=c(-3.5,2.7))
points(beta_wb,Y,cex=1.1,col="dark red",pch=16)

```

Figura B.2: Programa Máxima Verosimilitud Marginal y Condicional

B.3. Programa Métodos de estimación

B.3.1. Métodos de estimación Máxima Verosimilitud Marginal y Condicional

```

#Usando ltm para ajustar rasch con método marginal
#####
require(ltm)

#tiemporasch<-system.time(fitrasch <- rasch(prueba))
# l L, estima una discriminación única no necesariamente 1
tiemporasch<-system.time(
fitrasch <- rasch(prueba, constraint = cbind(ncol(prueba) + 1, 1))
# estima rasch donde la discriminación es 1.

bfitltm=matrix(-fitrasch$coefficients[,1],k,1)
#bfitltm=fitrasch$coefficients$beta.i
bfitltm
matrix(bsim,k,1)

MSEltm=sum((bsim-bfitltm)^2)/k
RMSEltm=sqrt(sum((bsim-bfitltm)^2)/k)
MSEltm
RMSEltm

plot(bsim,bfitltm,xlab="BETAS SIMULADOS",ylab="BETAS ESTIMADOS")
lines(bsim,bsim)
cor(bsim,bfitltm)

#Usando eRm para ajustar rasch con método condicional
#####
require(eRm)

tiemporascherm<-system.time(fitrascherm <- RM(prueba, sum0 = TRUE))
#restriccion para que la suma sea cero

bfiterm=matrix(-fitrascherm$betapar,k,1)
bfiterm
matrix(bsim,k,1)

MSEerm=sum((bsim-bfiterm)^2)/k
RMSEerm=sqrt(sum((bsim-bfiterm)^2)/k)
MSEerm
RMSEerm

plot(bsim,bfiterm,xlab="BETAS SIMULADOS",ylab="BETAS ESTIMADOS")
lines(bsim,bsim)
cor(bsim,bfiterm)

```

Figura B.3: Programa Máxima Verosimilitud Marginal y Condicional

B.3.2. Métodos de estimación Máxima Verosimilitud Conjunta

```
&INST
  TITLE = MAESTRIA
  DATA = D:\BASE WINSTEPS COMUNICACION.dat
  PERSON = EVALUADO
  ITEM = ITEM
  ITEM1 = 10
  NI = 18
  NAME1 = 1
  NAMELEN = 9
  CODES = 01
  XWIDE = 1
  LOCAL = Y
  DISCRIM = NO
  ASYMPTOTE = YES
  IFILE = D:\DIFF.txt
  PFILE = D:\PERS.txt
  DISFILE = D:\DISCR.txt
  UPMEAN = 0
  USCALE = 1
  UDECIM = 5
  GROUPS = 0
&END
P001
P002
P003
P004
P005
P006
P007
P008
P009
P010
P011
P012
P013
P014
P015
P016
P017
P018
END LABELS
```

Figura B.4: Programa Máxima Verosimilitud Conjunta

B.3.3. Métodos de estimación Bayesiana bajo el esquema ARS

Model Rasch bajo el esquema ARS

```

model{
  for (i in 1:n) {
    for (j in 1:k) {
      m[i,j]<-u[i]-b[j]
      p[i,j] <- exp(m[i,j])/(1+exp(m[i,j]))
      y[i,j] ~ dbern(p[i,j])
    }
  }
  #abilities priors
  for (i in 1:n) {
    u[i]~dnorm(0,1)
    uc[i] <- u[i] - mean(u[])
  }
  #items priors
  for (j in 1:k) {
    # usual priors
    #Bazan et al (2007)
    b[j] ~ dnorm(0,0.5)
    bc[j] <- b[j] - mean(b[])
  }
}

Data list(n=1536, k=18,
          y= structure(
            .Data = c(
1,1,1,1,1,1,0,1,1,1,1,1,1,1,0,0,1,0,0,
.....,
0,1,1,1,0,0,0,1,0,1,0,0,0,0,0,0,0,0),
            .Dim = c(1536,18))
          )

Inits list(b = c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0))

```

Figura B.5: Programa estimación Bayesiana bajo el esquema ARS

B.3.4. Métodos de estimación Bayesiana bajo el esquema DAGS

```

Model Rasch bajo el esquema DAGS

model
{
  for (i in 1 : n) {
    for (j in 1 : I) {
      m[i,j] <- u[i] - b[j]
      z[i,j] ~ dlogis(m[i,j],1)I(lo[y[i,j]+1],up[y[i,j]+1])
    }
  }
  #priors F for item parameters
  for (j in 1:I)
  {
    b[j] ~ dnorm(0,0.5);
    bc[j] <- b[j] - mean(b[])
  }
  #prior for latent variable
  for (i in 1:n)
  {
    u[i] ~ dnorm(0,1)
    uc[i] <- u[i] - mean(u[])
  }
  # auxiliary latent variable
  lo[1] <- -50; lo[2] <- 0;
  up[1] <- 0; up[2] <- 50;
  # mean and standard deviation for latent variable
  mu<-mean(u[])
  du<-sd(u[])
}

Data list(n=1536, k=18,
          y= structure(
            .Data = c(
1,1,1,1,1,1,0,1,1,1,1,1,1,0,0,1,0,0,
.....,
0,1,1,1,0,0,0,1,0,1,0,0,0,0,0,0,0,0),
            .Dim = c(1536,18))
          )

Inits list(b = c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0))

```

Figura B.6: Programa estimación Bayesiana bajo el esquema DAGS

Bibliografía

- [1] Albert, J.H (1992). Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling. *Journal of Educational Statistics*, **17(3)**, 251 - 269.
- [2] Albert, J. (2009). *Bayesian Computation with R* Springer.
- [3] Baker, F.B. (1992). *Item Response Theory. Parameter estimation techniques*. New York: Marcel Dekker.
- [4] Barry R.J. (1995). *Probabilidad: Un curso de nivel intermedio*. IMCA.
- [5] Bayes, C. y Bazán, J. (2010). Inferencia Bayesiana en Modelos de respuesta Binaria usando BRMUV. *I Jornada Internacional de Probabilidad y Estadística, Serie B(25)*, Lima: PUCP, Febrero.
- [6] Bazán, J., Valdivieso, L. y Calderón, A. (2010). Enfoque Bayesiano en Modelos de Teoría de Respuesta al Ítem. *I Jornada Internacional de Probabilidad y Estadística, Serie B(27)*, Lima: PUCP, Febrero.
- [7] Bond, T.G. y Fox Ch.M. (1985). *Applying The Rasch Model Fundamental in the Human Sciencies*. Springer.
- [8] Congdon, P. (2003). *Applied Bayesian Modelling*. John Wiley & Sons.
- [9] Chincaro, O. y Bazán, J. (2010). Poster: Una comparación de los métodos de estimación del Modelo Rasch. *I Jornada Internacional de Probabilidad y Estadística*. Lima: PUCP, Febrero.
- [10] De Boeck, P. y Wilson, M. (2009). *Explonatory Item Response Models* Springer.
- [11] Dirección Nacional de Educación Secundaria (2009). *Proyecto: Movilización Nacional por la Comprensión Lectora (MNCL)*. Ministerio de Educación.
- [12] Efron, B. y Tibshirani, R.J. (1993). *Rasch Models Foundations, Recent Developments, and Applications*. Chapman and Hall/CRC.
- [13] Fischer, G.H.(2007). Rasch Models. *Handbook of Statistics in Psychometrics*, **26**, 515-586. North Holland
- [14] Fischer, G.H. y Molenaar, I.W. (1995). *Rasch Models Foundations, Recent Developments, and Applications*. Springer - Verlag.

- [15] Hambleton, R.K, Swaminathan, H. y Jane, R.H. (1991). *MMSS Fundamentals of Item Response Theory*. Sage Publications, Inc.
- [16] Hogg, R.V. y Craig A.T (1978). *Introduction to Mathematical Statistics* Macmillan Publishing Co, Inc.
- [17] Johnson V. y Albert J. (2000). *Ordinal Data Modeling* Newyork, Springer.
- [18] Kim, S.H (2001). An Evaluation of a Markov Chain Monte Carlo Method for the Rasch Model. *Applied Psychological Measurement*, **25(2)**, 163 - 176.
- [19] Linacre, John M. (2009). A User's Guide to WINSTEPS. <http://www.winsteps.com/a/winsteps.pdf>
- [20] Lord, F.M. (1986). Maximum likelihood and bayesian parameter estimation in item response theory. *Journal of Educational Measurement*, **23(2)**, 157 - 162.
- [21] Mair, P y Hatzinger, R (2007). Extended Rasch modeling: The eRm Package for the Application of IRT Models in R. *Journal of Statistical Software*, **20(9)**.
- [22] Maris, G y Maris, E (2002). A MCMC - method for models with continuous latente responses. *Psychometrika*, **67(3)**, 335 - 350.
- [23] Muñiz, J. (1998). *Teoría clásica de los Test*. Pirámide.
- [24] Qing, L. y Pierce, D.A. (1994). A Note on Gauss-Hermite Quadrature. *Journal of Educational Statistics*, **81(3)**, 624 - 629
- [25] Rizopoulos, D (2006). ltm: An R Package for Latent variable Modeling. *Journal of Statistical Software*, **17(5)**.
- [26] Ross, Sheldon M. (1990). *A course in simulation* Springer.
- [27] Swaminathan, H. y Gifford, J.A. (1982). Bayesian Estimation in the Rasch Model. *Journal of Educational Statistics*, **7(3)**, 175 - 191 Springer, New York.
- [28] Tanner, M.A. (1995). *Tools for Statistical Inference*. Springer.
- [29] Tsutakawa, R.K (1988). Misunderstanding the rasch model. *Journal of Educational Statistics*, **13(2)**, 117 - 130.
- [30] Wainer, H., Morgan, A y Gustafsson J.E (1980). A review of estimation procedures for the rasch model with an eye toward longish tests. *Journal of Educational Statistics*, **5(1)**, 35 - 64.
- [31] Wright, B.D. (1977). Misunderstanding the rasch model. *Journal of Educational Statistics*, **14(3)**, 219 - 225.
- [32] Wright, B.D. (1998). *Diseño de las mejores pruebas*. Ceneval.