Décimocuartas Jornadas en Estadísticas e Informática
ECUADOR 2007

Modelos de respuestas al Item usando WinBUGS

Dr. Jorge Luis Bazán
jlbazan@pucp.edu.pe

Guayaquil, Octubre 17-18, 2007

- ASPECTOS GENERALES DE LA EVALUACION
- EL MODELO DE RASCH

Parte 1:
ASPECTOS GENERALES DE EVALUACIÓN:

1. Evaluación -Ejemplos
2. Naturaleza Jerárquica de la Evaluación
3. Naturaleza Latente de la Evaluación
4. HLM o Modelo Multinivel
5. Teoría de Respuesta al ítem
6. Extensiones y Referencias

1. Evaluación

Es importante para:

- Modernización del Estado (SNIP, OIOE)
- Proceso de Descentralización
- Mejoras de eficiencia (Servicios, Sectores Productivos, Mercados-competitividad)
- Mejora en la calidad del sistema educativo
- Mejores criterios de selección y evaluación de estudiantes y profesores

La evaluación tiene que ver con:

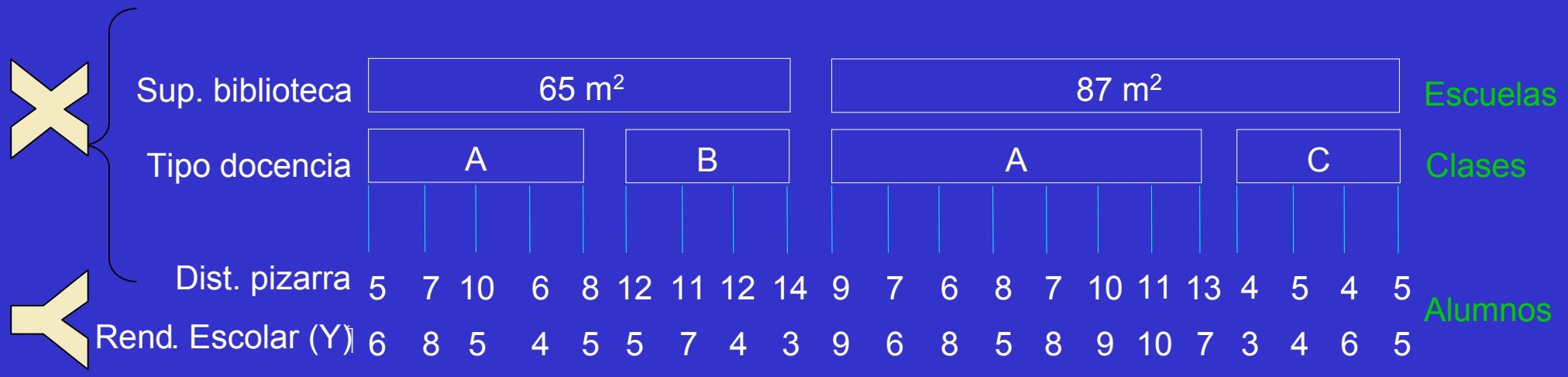
- Desarrollo de instrumentos (escalas, cuestionarios y sus propiedades)
- Mejor definición de los propósitos de investigación (objetivos y resultados)
- Mejores modelos matemáticos y estadísticos
- Mejores sistemas informáticos para base de datos, análisis y aplicaciones de pruebas
- Desarrollo de criterios para el diseño de políticas usando sus resultados

Esto induce al desarrollo metodológico de la evaluación y a un debate sobre su aplicaciones.

En el Perú no hay suficientes personas calificadas en el manejo técnico de la evaluación.

2. Naturaleza jerárquica de la evaluación

	Educacion	Clinicas	Encuestas	Longitudinal
Nivel 3	Escuelas	Centros	Conglomerados	Organizaciones
Nivel 2	Clases	Terapeut	Entrevistadores	Individuos
Nivel 1	Alumnos	Pacientes	Encuestados	Ocasiones

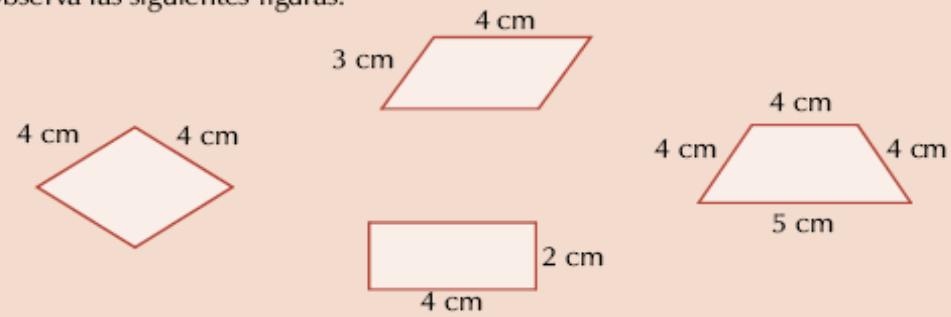


¿ Modelo de regresión lineal múltiple ?

3. Naturaleza latente de la evaluación

Ítem Nº 10

Observa las siguientes figuras.



¿Cuál es la suma de todos los lados del **rombo**?

A) 17 cm
B) 16 cm
C) 14 cm
D) 12 cm

Resultados: A) 33% **B) 54%** C) 5,5% D) 5,5% Omitidos) 2,1%

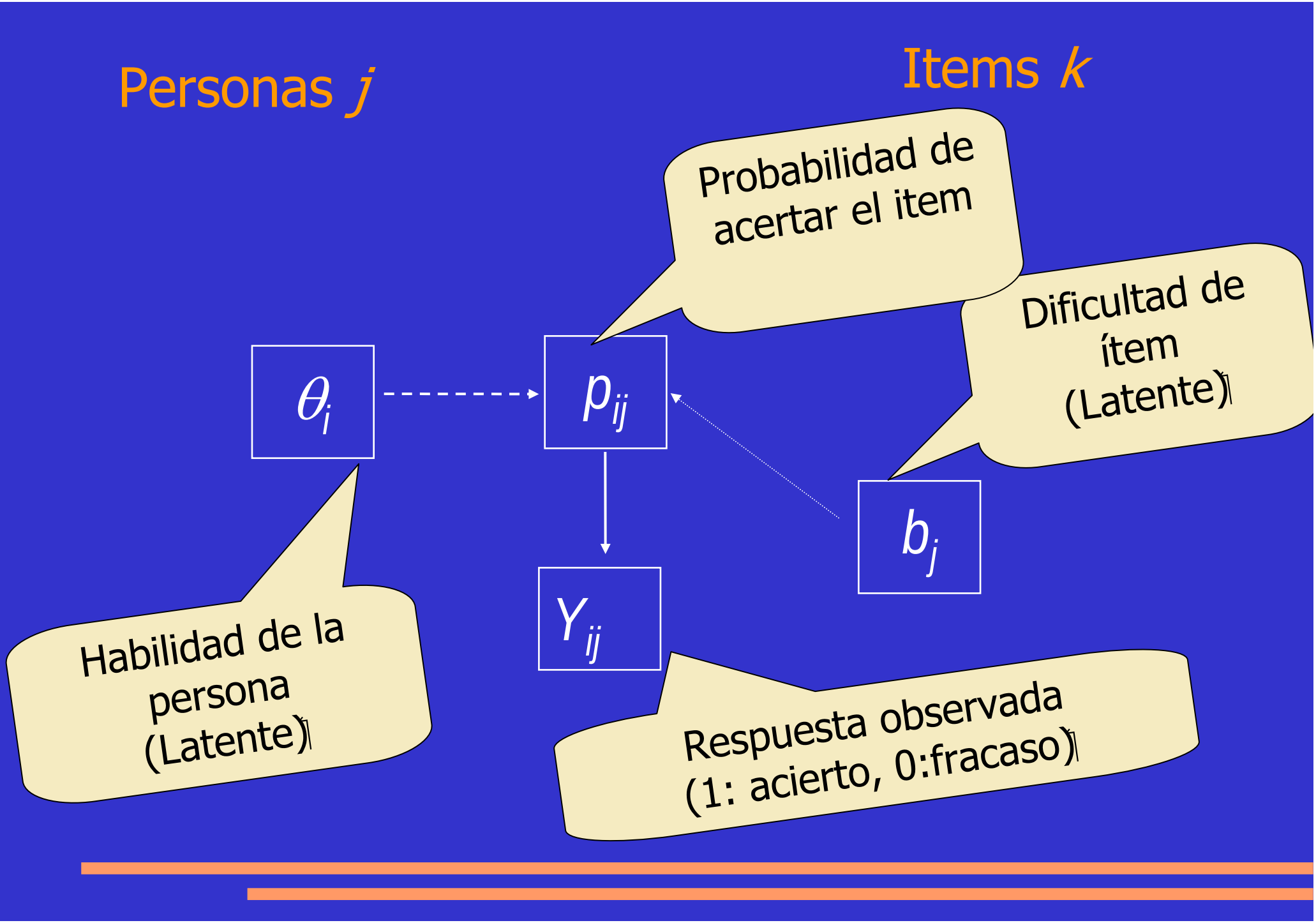
Ítem Nº 30

Un tanque recibe 4,5 litros de agua por minuto. ¿Cuántos litros de agua tendrá el tanque en una hora y media?

A) 4 050 litros
B) 405 litros
C) 7,2 litros
D) 6,75 litros

Resultados: A) 38,9% **B) 22,4%** C) 12,5% D)14,7% Omitidos) 11,4%

http://www.minedu.gob.pe/umc/publicaciones/boletines/Boletin-13.pdf


Personas j
Items k
Probabilidad de acertar el item
Dificultad de ítem (Latente)
θ_i
p_{ij}
b_j
Y_{ij}
Habilidad de la persona (Latente)
Respuesta observada (1: acierto, 0:fracaso)

- Análisis de las variables explicativas: El Análisis de Factores Asociados

MODELOS JERARQUICO O MULTINIVEL

- Análisis de la variable a explicar: Los Análisis psicométricos de la variable rendimiento

MODELOS DE RESPUESTA AL ITEM

4. HLM o modelo Multinivel

Sistema jerárquico de 2 ecuaciones de regresión

Nivel 1: ***X*** = Variable independiente

Y = Variable dependiente

Nivel 2: ***Z*** = Variable independiente

Nivel 1

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij}$$

Nivel 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j}$$

$$Y_{ij} = \underbrace{\gamma_{00} + \gamma_{01} Z_j + \gamma_{10} X_{ij} + \gamma_{11} X_{ij} Z_j}_{\text{Parte determinista}} + \underbrace{u_{1j} X_{ij} + u_{0j} + e_{ij}}_{\text{Parte aleatoria}}$$

Esta familia de modelos han sido difundido en la literatura como modelos HLM (ver Goldstein, 1995).

Esta metodología representa una mejora en la estrategia del empleo de regresión múltiple en los estudios de rendimiento.

Esto define los modelos de multinivel, es decir, un nivel donde se explica el comportamiento de los estudiantes y otro nivel, el nivel de grupo, donde se modelan los factores del grupo o escuela.

5. Teoría de Respuesta al Item (modelo 2L)

$$Y_{ij} \sim Bern(p_{ij})$$

$$i = 1, \ldots, n, \quad j = 1, \ldots k$$

$$p_{ij} = F(m_{ij})$$

$$m_{ij} = a_j(\theta_i - b_j)$$

$$p(Y|\theta, \alpha, \beta) = \prod_{i=1}^{n} \prod_{j=1}^{k} \left[F(m_{ij})\right]^{Y_{ij}} \left[1 - F(m_{ij})\right]^{1-Y_{ij}}$$

$$\theta_i \sim N(\mu, \sigma^2)$$

Una familia de modelos psicométricos para determinar la probabilidad de acertar todo tipo de item en función de características de las personas y de los ítemes que son consideradas latentes, tomando en cuenta las respuestas obtenidas en cada pregunta.

(Ver Van der Linden y Hambleton, 1996)

Complementa y mejora el llamado análisis clásico de items que incluye la validez (correlación item-test), dificultad, discriminación e índice de no respuesta.

6.1 Referencias HLM

- Bryk, A., & Raudenbush, S. W. (1992). *Hierarchical Linear Models for Social and Behavioral Research: Applications and Data Analysis Methods.* Newbury Park, CA: Sage.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition.* Newbury Park, CA: Sage.
- Goldstein, H. (1995). *Multilevel Statistical Models.* New York: Halstead.

6.2 Referencias TRI

- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory.* Newbury Park: Sage.
- Baker, F. and Kim, S. (2004). Item Response Theory: Parameter Estimation Techniques . Marcel Dekker Inc. New York
- Van der Linden, W. y Hambleton, R. K (Editores) (1996). Handbook of Modern Item Response Theory.. New York: Springer

IRT o TRI

- http://edres.org/irt/
- http://work.psych.uiuc.edu/irt/tutorial.asp
- http://psychcentral.com/psypsych/Item_response_theory

HLM o Multilevel

- http://multilevel.ioe.ac.uk/
- http://www.bvs.sld.cu/revistas/ems/vol18_2_04/ems01204.htm

Software e Libros

- http://www.ssicentral.com http://www.assess.com

Parte 2:

EL MODELO DE RASCH Y SUS APLICACIONES:

1. Supuestos del Modelo
2. Derivación Funcional
3. Curvas características de los ítems
4. Propiedades
5. Software
6. Estimación Bayesiana
7. Ejemplos
8. Otros modelos TRI

1. Supuestos del Modelo

1. Relación funcional

$$p_{ij} = P(Y_{ij} = 1 / \theta_i, \beta_j) = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}}$$

$i = 1, \cdots, n$

$j = 1, \cdots, k$

b_j : Dificultad del item *j*

θ_i : Habilidad de la persona *i*

p_{ij}: probabilidad de que la persona *i* acierte el item *j*

2) Independencia condicional

$Y_{ij} / \theta_i, \beta_j$ son independientes para todo *i* y todo *j*

2. Derivación de la relación funcional

- Si la persona *i* es *k* veces más hábil que la persona *i´*
- Si el item *j* es *k* veces más díficil que el item j´

$$\eta_i = k\eta_{i'}$$

$$\delta_i = k\delta_{j'}$$

$$\varepsilon_{ij} = \frac{\eta_i}{\delta_{j'}}$$

- Si una persona tiene más habilidad su probabilidad de acierto es mayor
- Si un ítem es más difícil, la probabilidad de acierto es menor

$$\uparrow \eta_i \Rightarrow \uparrow p_{ij}$$

$$\downarrow \delta_j \Rightarrow \uparrow p_{ij}$$

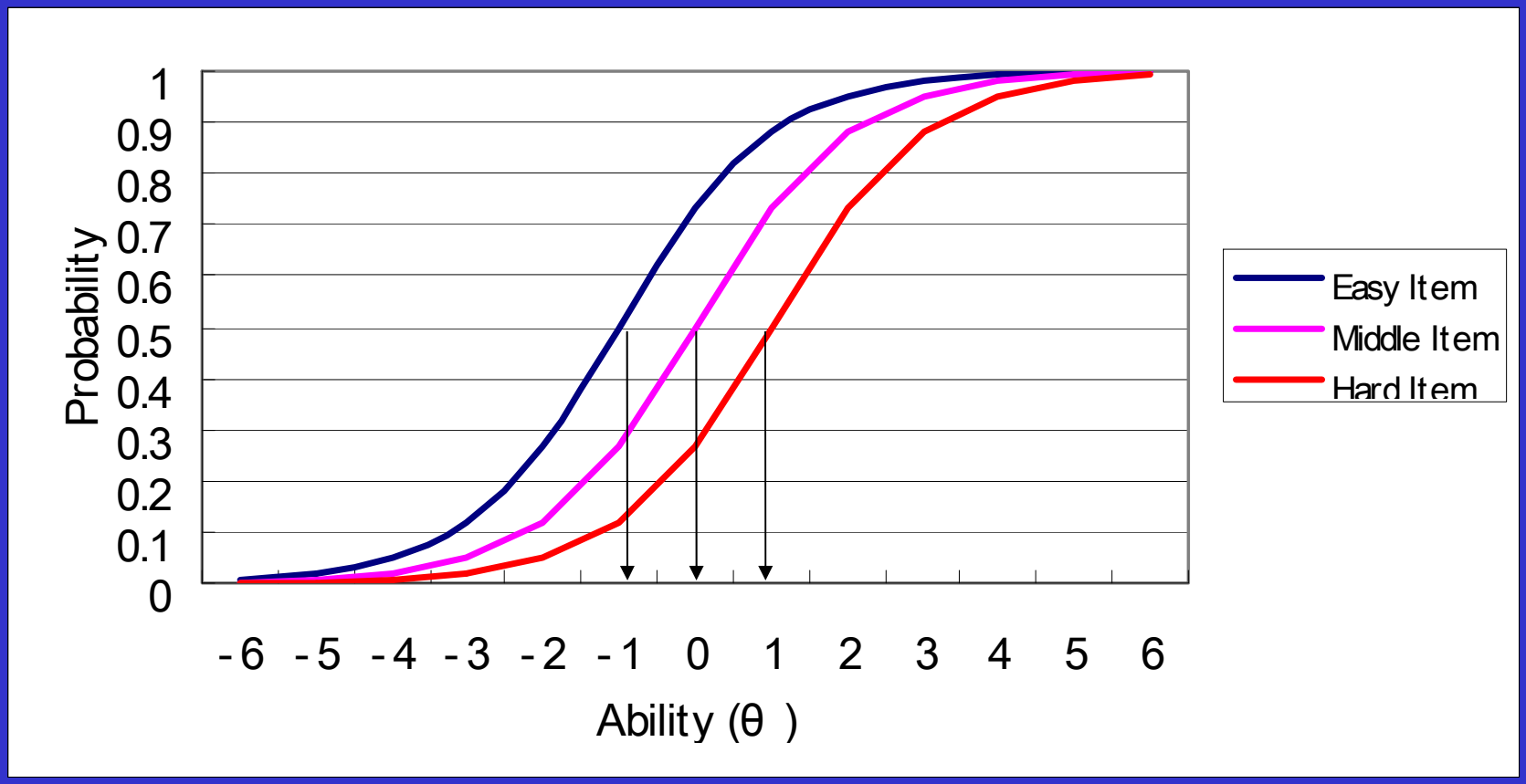
$$p_{ij} = \frac{\eta_i}{\eta_i + \delta_{j'}}$$

$$p_{ij} = \frac{\varepsilon_{ij}}{1+\varepsilon_{ij}} \Rightarrow \varepsilon_{ij} = \frac{p_{ij}}{1-p_{ij}} \Rightarrow \ln \varepsilon_{ij} = \ln \frac{p_{ij}}{1-p_{ij}}$$

$$\Rightarrow \ln \eta_i - \ln \delta_j = \ln \frac{p_{ij}}{1-p_{ij}} \Rightarrow \theta_i - \beta_j = \ln \frac{p_{ij}}{1-p_{ij}}$$

$$\Rightarrow p_{ij} = P(Y_{ij} = 1 / \theta_i, \beta_j) = \frac{e^{\theta_i - \beta_j}}{1+e^{\theta_i - \beta_j}}$$

3. Curvas características de ítems



4. Propiedades

- Unidimensionalidad
- Medición conjunta
- Objetividad específica
- Medición de intervalo

5. Métodos de Estimación

Basados en Máxima verosimilitud:
Máxima verosimilitud conjunta, máxima verosimilitud condicional, máxima verosimilitud marginal

Basados en Inferencia Bayesiana:
sin o moda a posteriori, MCMC y con MCMC: Gibss Sampling, Metropolis Hasting

Máxima verosimilitud conjunta

$$p(\mathbf{Y}|\theta,\xi) = \prod_{i=1}^{N}\prod_{j=1}^{n} P_j(\theta_i)^{Y_{ij}}\, Q_j(\theta_i)^{1-Y_{ij}} = l(\boldsymbol{\theta},\boldsymbol{\xi}|\mathbf{Y}) \ ,$$

Estimación Bayesiana sin MCMC o de moda a posteriori

$$p(\boldsymbol{\theta},\boldsymbol{\xi}|\mathbf{Y}) = \frac{p(\mathbf{Y}|\boldsymbol{\theta},\boldsymbol{\xi})\,p(\boldsymbol{\theta},\boldsymbol{\xi})}{p(\mathbf{Y})} \propto l(\boldsymbol{\theta},\boldsymbol{\xi}|\mathbf{Y})\,p(\boldsymbol{\theta},\boldsymbol{\xi}) \ ,$$

Máxima verosimilitud Condicional

$$\frac{p(\mathbf{Y}_i|\theta_i,\boldsymbol{\xi})}{p(R_i|\theta_i,\boldsymbol{\xi})} = p(\mathbf{Y}_i|R_i,\boldsymbol{\xi}) \ ,$$

$$l(\boldsymbol{\xi}|\mathbf{Y},\mathbf{R}) \ .$$

$$R_i = \sum_j Y_{ij}$$

Máxima verosimilitud marginal

$$p(\mathbf{Y}_i|\boldsymbol{\xi}) = \int p(\mathbf{Y}_i|\theta_i,\boldsymbol{\xi})\,p(\theta_i)\,d\theta_i \ ,$$

$$p(\mathbf{Y}|\xi) = \prod_{i=1}^{N} p(\mathbf{Y}_i|\xi) = l(\boldsymbol{\xi},\mathbf{Y}) \ .$$

6. Software

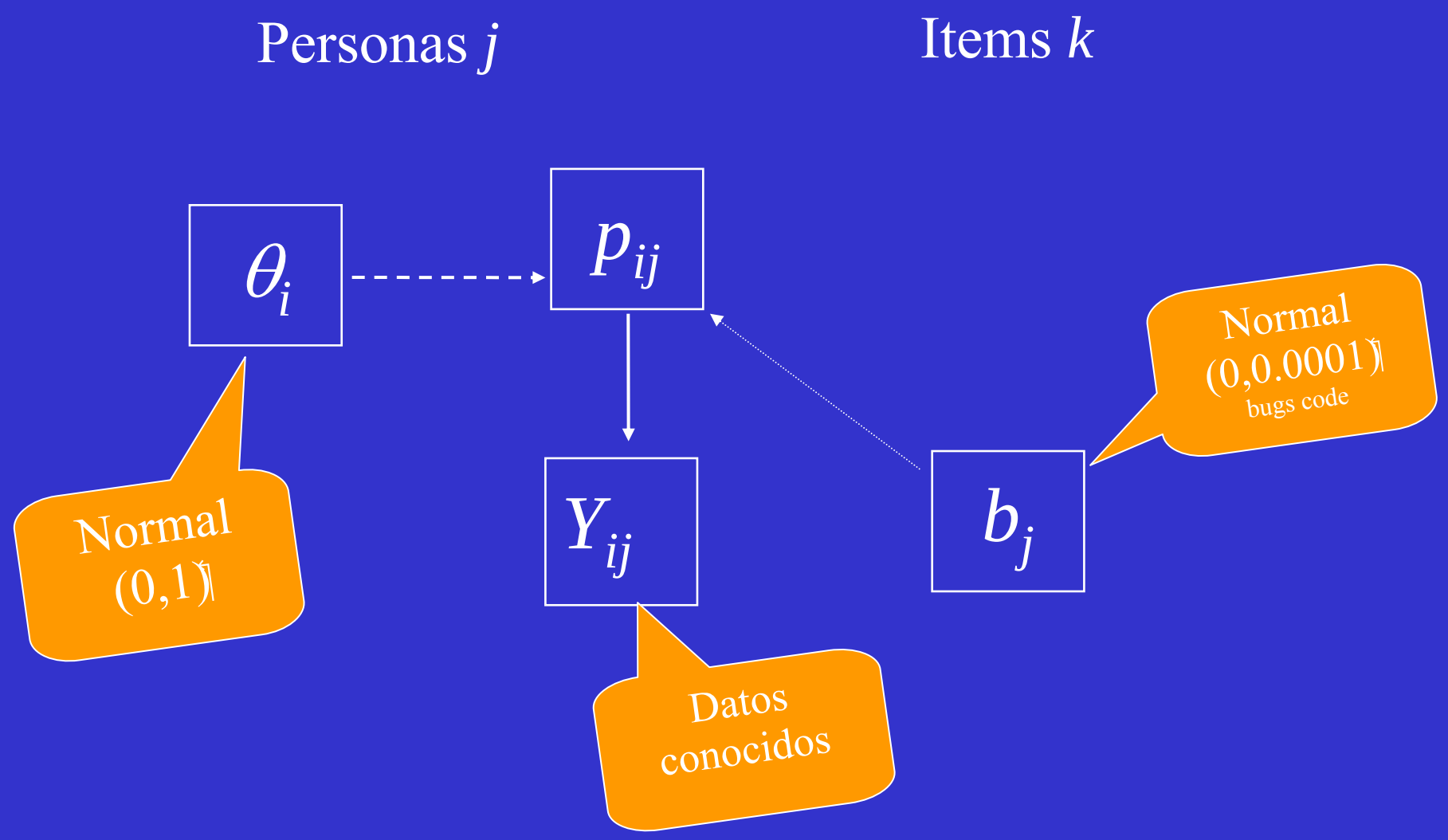
Psicométricos

- Winstep
- Rascal
- Bilog
- Conquest
- Quest
- Winmira
- RUMM2020
- Param3PL
- Logimo
- MSP
- LPCM-WIN
- RSP
- T-Rasch
- ICL-WIN
- LEM
- Multilog

Estadísticos

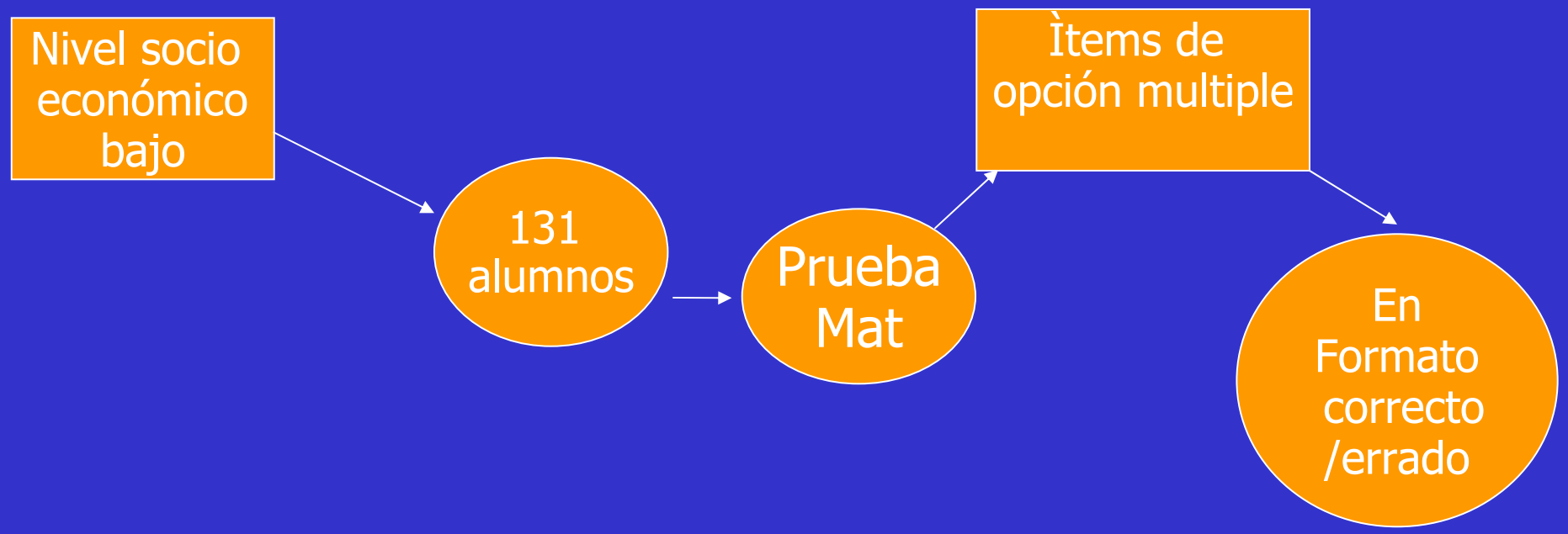
- SAS
- R
- Stata
- WinBUGS
- Systat

7. Estimación Bayesiana



8. Ejemplos
Math data

Analizar resultados de un subgrupo de alumnos de sexto grado de nivel socioeconómico alto en una prueba de Matemática de 14 ítems



Boletin-13.pdf

mathb.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

4 :

	i01	i02	i03	i04	i05	i06	i07	i08	i09	i10	i11	i12	i13	i14
1	1	1	0	1	1	0	1	1	1	1	1	0	0	1
2	1	1	1	1	1	1	1	1	1	1	1	0	1	1
3	1	1	1	0	1	0	0	1	0	1	0	0	0	1
4	1	0	0	1	1	0	1	0	1	1	0	0	1	0
5	1	1	1	1	1	0	1	1	1	1	1	1	1	1
6	1	0	1	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	0	1	1	1	1	1	1	1	1
8	1	1	0	1	0	0	1	1	1	1	1	1	1	1
9	0	1	1	1	1	1	1	1	0	1	1	1	1	1
10	1	0	1	1	1	0	1	0	1	0	1	0	1	1
11	1	0	0	1	1	0	0	1	1	1	1	0	1	1
12	1	1	0	1	0	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	0	1	1	1	1	1	0	1	1
14	1	0	0	1	1	1	0	1	0	1	0	0	1	1
15	1	1	1	1	1	1	1	1	1	0	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1
17	0	1	1	1	1	0	1	0	1	1	1	1	0	1
18	1	1	0	1	1	0	1	1	1	0	1	0	1	1
19	1	0	0	1	1	0	1	1	1	1	1	0	0	1
20	1	1	0	1	1	0	1	1	1	1	1	0	1	1
21	1	1	0	1	1	0	1	1	0	1	1	1	1	1

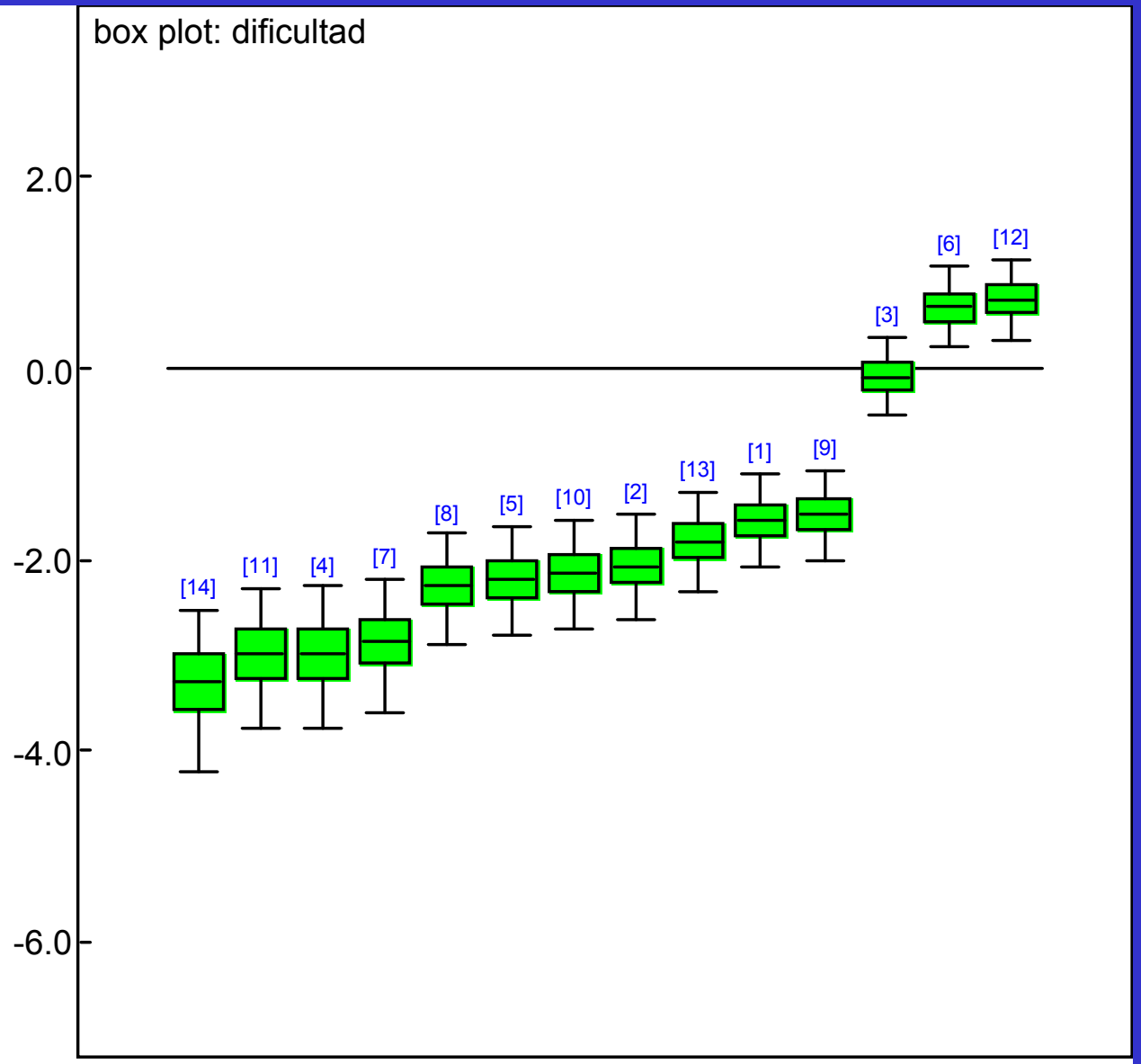
Data View / Variable View

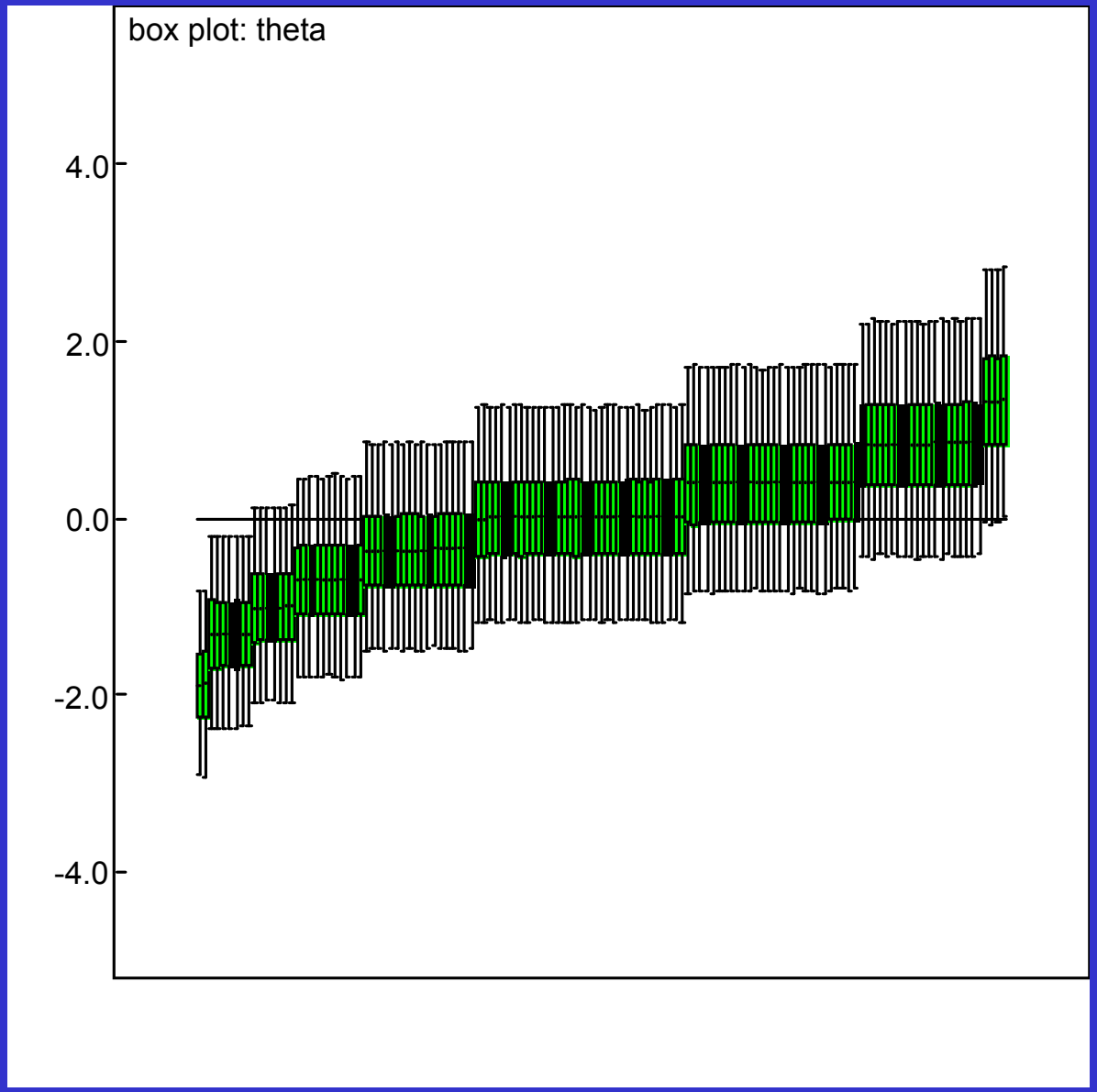
SPSS Processor is ready

BUGS code

```
 Model {
# Rasch
  for (i in 1 : n) {
    for (j in 1 : k) {
  logit(p[i, j]) <-theta[i] -b[j]
  y[i, j] ~ dbern(p[i, j])
                          }
                        }

  for (i in 1:k) {
    theta[i] ~ dnorm(0, 1)
                        }
  for (j in 1:I) {
    b[j] ~ dnorm(0, 0.0001)
                        }
   }
```


box plot: dificultad
2.0
0.0
-2.0
-4.0
-6.0
[14]
[11]
[4]
[7]
[8]
[5]
[10]
[2]
[13]
[1]
[9]
[3]
[6]
[12]


box plot: theta
4.0
2.0
0.0
-2.0
-4.0

Personas e Items en una misma escala

Modelo de Rasch mide cada item y cada persona en la misma escala

Concentrarse en los itemes que se necesitan

- Medir a todos
- Medir eficientemente

MAP ALUMNOS_PREGUNS.pdf

Parte 3:
INVESTIGACIONES CON RASCH:

1. Administración
2. Analisis Político
3. Comunicaciones
4. Educación
5. Psicología
6. Control de Calidad

InvestigacionesconRasch.pdf

Mayores detalles

Bond, T.G and Fox, C.M

Applying the Rasch Model: Fundamental Measurement in the Human Sciences

Lawrence Erlbaum Associates

ISBN: 0-8058-4252-7

Agradecimientos DAI-PUCP

Otros Modelos TRI

Item: ***Las matemáticas no son útiles para mi futuro***

a) En desacuerdo completo

b) En desacuerdo

c) Indistinto

d) De acuerdo

e) En acuerdo completo

Item: ***Escriba tres refranes y explique su significado***

a) No logra escribir ninguno

b) Lograr escrebir uno

c) Logra escribir dos

d) Logra escribir tres

Item: ***Calificativo en la redacción de una carta***

a) Insuficiente
b) Regular
c) Bueno
d) Excelente

Item: $\sqrt{9{,}0/0{,}3 - 5} = ?$ **(Masters, 1982)**

Categorias de respuesta

a) No hace nada

b) Calcula corretamente

c) Hace b) y obtiene

d) Hace c) y calcula

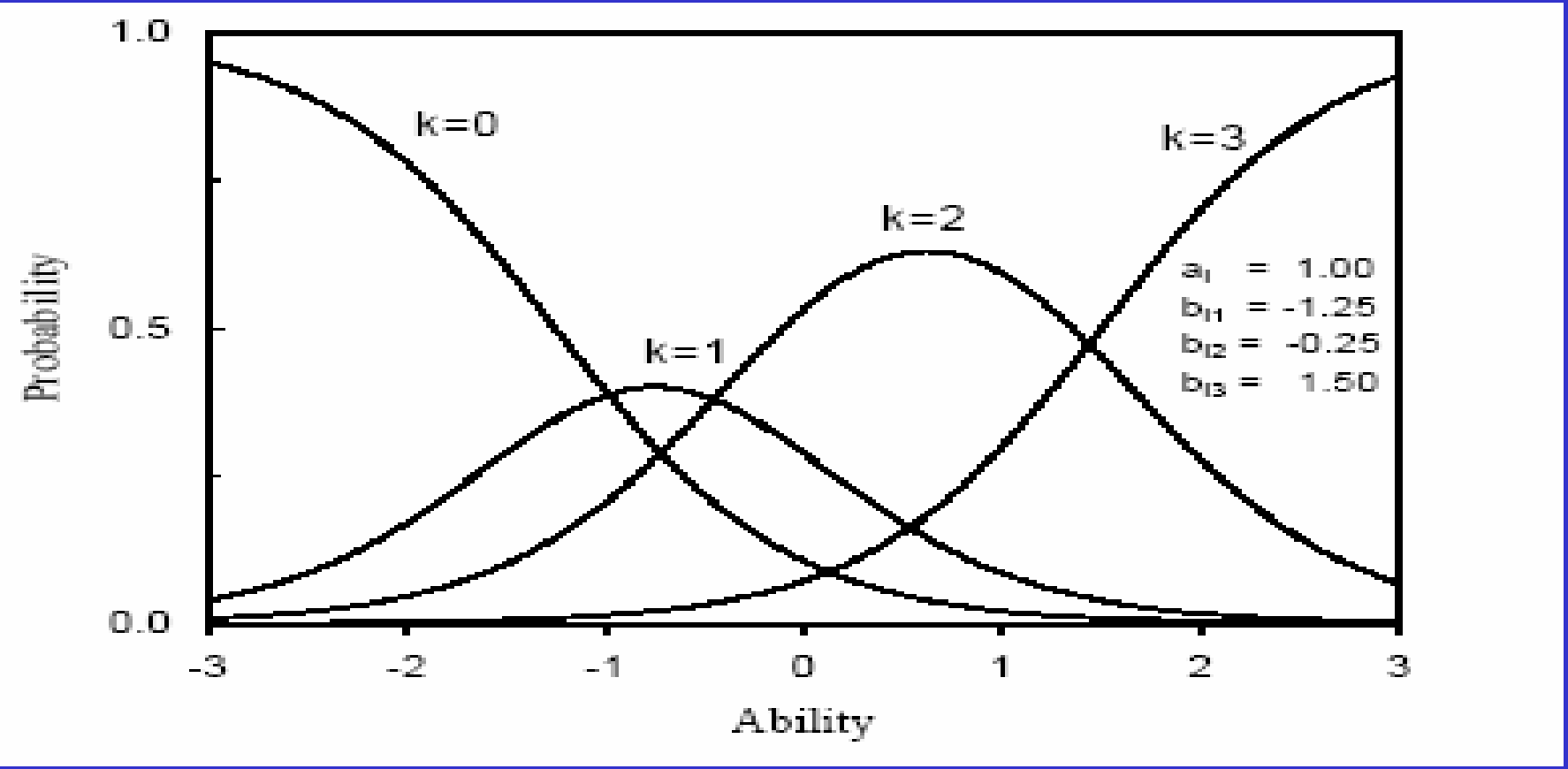
$$9{,}0/0{,}3 = 30$$

$$30 - 5 = 25$$

$$\sqrt{25} = 5$$

Modelo de credito parcial generalizado

$$P(x_i = k \mid \theta) = \frac{\exp[\sum_{s=1}^{k} a_i(\theta - b_{s_i})]}{1 + \sum_{r=1}^{k} \exp[\sum_{s=1}^{r} a_i(\theta - b_{s_i})]}$$


1.0
0.5
0.0
Probability
k=0
k=1
k=2
k=3
a_i = 1.00
b_{i1} = -1.25
b_{i2} = -0.25
b_{i3} = 1.50
-3
-2
-1
0
1
2
3
Ability

Modelos para itemes con formatos de respuestas polocotómica

- **Modelos de Categorías Nominales**
- Modelos de Respuesta para ítems de elecciones multiples
- Modelos de Escala de Rating (jerarquizaciones)
- Modelos de Respuesta Graduada
- Modelos de Crédito parcial
- Modelo Secuencial para respuestas ordenadas

Modelos para tiempo de respuesta o múltiples ensayos

- **Modelo Logístico para pruebas de tiempo límite**
- Modelos para pruebas de Rápidez y de Tiempo Límite
- Modelos de Ensayos Multiples

Modelos para habilidades múltiples o componentes cognitivos

- **Modelo de Rasch: Muled**
- Modelo de respuesta con predictores manifiestos
- Modelo Multidimensional de Ojiva Normal
- Modelo Multidimensional Logístico lineal
- Modelo de Respueta al Item multidimensional loglineal
- Modelo de Respuesta Multicomponente
- Modelo para el cambio Multidimensional Logístico Lineal

Modelos no paramétricos

- **Modelo no paramétrico para respuesta dicotómica (Mokken)**
- Modelo no paramétrico para respuesta policotómica

Modelo de Items no Monótonos

- **Modelo IRT de Coseno Hyperbólico**
- Modelo IRT para ánalisis de paralelogramo

Statistics for Social Science and Public Policy

Paul De Boeck, Mark Wilson, Editors

Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach

This edited volume gives a new and integrated introduction to item response models (predominantly used in measurement applications in psychology, education, and other social science areas) from the viewpoint of the statistical theory of generalized linear and nonlinear mixed models. Moreover, this new framework allows the domain of item response models to be co-ordinated and broadened to emphasize their *explanatory* uses beyond their standard *descriptive* uses.

The basic explanatory principle is that item responses can be modelled as a function of predictors of various kinds. The predictors can be (a) characteristics of items, of persons, and of combinations of persons and items; they can be (b) observed or latent (of either items or persons); and they can be (c) latent continuous or latent categorical. Thus, a broad range of models is generated, including a wide range of extant item response models as well as some new ones. Within this range, models with explanatory predictors are given special attention in this book, but we also discuss descriptive models. Note that the "item responses" that we are referring to are not just the traditional "test data," but are broadly conceived as categorical data from a repeated observations design. Hence, data from studies with repeated observations experimental designs, or with longitudinal designs, may also be modelled.

The book starts with a four-chapter section containing an introduction to the framework. The remaining chapters describe models for ordered-category data, multilevel models, models for differential item functioning, multidimensional models, models for local item dependency, and mixture models. It also includes a chapter on the statistical background and one on useful software. In order to make the task easier for the reader, a unified approach to notation and model description is followed throughout the chapters, and a single data set is used in most examples to make it easier to see how the many models are related. For all major examples, computer commands from the SAS package are provided which can be used to estimate the results for each model. In addition, sample commands are provided for other major computer packages.

Paul De Boeck is Professor of Psychology at K.U. Leuven (Belgium), and Mark Wilson is Professor of Education at UC Berkeley (USA). They are also co-editors (along with Pamela Moss) of a new journal entitled *Measurement: Interdisciplinary Research and Perspectives*. The chapter authors are members of a collaborative group of psychometricians and statisticians centered on K.U. Leuven and UC Berkeley.

ISBN 0-387-40275-6
9 780387 402758 >
springeronline.com

De Boeck
Wilson, Editors

Explanatory Item Response Models

Statistics for Social Science and Public Policy

Paul De Boeck
Mark Wilson
Editors

Explanatory Item Response Models

A Generalized Linear and Nonlinear Approach

Springer