



PONTIFICIA  
**UNIVERSIDAD**  
**CATÓLICA**  
DEL PERÚ

## ESCUELA DE POSGRADO

# MODELOS DE TEORÍA DE RESPUESTA AL ÍTEM MULTIDIMENSIONAL CON UNA APLICACIÓN PSICOLÓGICA

Tesis para optar el grado de Magíster en Estadística

### AUTOR

Martín Ludgardo Malaspina Quevedo

### ASESOR

Dr. Jorge Luis Bazán Guzmán

LIMA - PERÚ

2016

## Dedicatoria

A mis queridos padres.



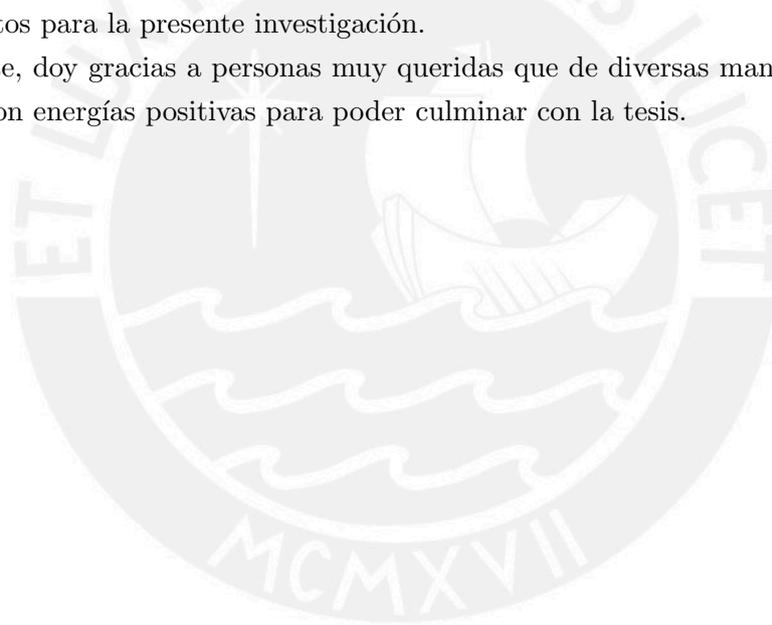
## Agradecimientos

Agradezco infinitamente a mi familia, en especial a mis padres, por aprender tanto de ellos y sentir que siempre me apoyaron con mucho amor en todas las aventuras estadísticas, locuras psicológicas y desafíos educativos de la maestría.

Doy gracias especiales a mi asesor Jorge Bazán, que pese a la distancia física, supo guiarme adecuadamente en el desarrollo de esta tesis con sus observaciones significativas y buenos consejos. Asimismo, agradezco a los profesores Luis Valdivieso y Cristian Bayes por todo lo enseñado en la maestría y el apoyo en la presente investigación.

De igual manera, agradezco a la investigadora Ana Aparicio, quien me cedió amablemente la base de datos para la presente investigación.

Finalmente, doy gracias a personas muy queridas que de diversas maneras me apoyaron y transmitieron energías positivas para poder culminar con la tesis.



## Resumen

La presente investigación, dentro del contexto de la Teoría de Respuesta al Ítem (TRI), estudia un modelo multidimensional logístico compensatorio de dos parámetros (M2PL) para ítems dicotómicos. Para ello, se explican teóricamente los métodos de estimación más conocidos para los parámetros de los ítems y de los rasgos latentes de las personas, priorizando el método bayesiano mediante Cadenas de Markov de Monte Carlo (MCMC). Estos métodos de estimación se exploran mediante implementaciones computacionales con el software R y R2WinBUGS. La calidad de las respectivas estimaciones de los parámetros se analiza mediante un estudio de simulación, en el cual se comprueba que el método de estimación más robusto para el modelo propuesto es el bayesiano mediante MCMC. Finalmente, el modelo y el método de estimación elegidos se ilustran mediante una aplicación que usa un conjunto de datos sobre actitudes hacia la estadística en estudiantes de una universidad privada de Colombia.

**Palabras-clave:** TRI Multidimensional, Ítems dicotómicos, Estimación Bayesiana mediante MCMC, Actitudes hacia la Estadística.

## Abstract

The current research, in the context of Item Response Theory (IRT), presents a study of a multidimensional two parameter logistic compensatory model (M2PL) for dichotomous items. This model is theoretically explained with well-known methods for estimation item parameters and latent traits of people, prioritizing the Bayesian method using Markov Chain Monte Carlo (MCMC). These estimation methods are also explored using computer software implementations with R and R2WinBUGS. The quality of the respective estimates of the parameters is analyzed using a simulation study in which is verified that Bayesian estimate method by MCMC is the most robust for the proposed model. Finally, the model and the method of estimation chosen are illustrated by an application using a database about attitudes in students towards statistics in a private university in Colombia.

**Keywords:** Multidimensional IRT, Dichotomous items, Bayesian estimation by MCMC, Attitudes toward Statistics.

# Índice general

<b>Lista de abreviaturas</b>	<b>VIII</b>
<b>Lista de símbolos</b>	<b>IX</b>
<b>Índice de figuras</b>	<b>X</b>
<b>Índice de cuadros</b>	<b>XII</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos	2
1.2. Organización del trabajo	2
<b>2. Consideraciones preliminares</b>	<b>4</b>
2.1. Modelos de Teoría de Respuesta al Ítem Unidimensional	4
2.1.1. Modelos para pruebas con ítems dicotómicos	5
2.1.2. Limitaciones de los modelos de TRIU	7
2.2. Análisis de la dimensionalidad	7
2.2.1. Análisis Factorial Exploratorio	8
2.2.1.1. Matriz de correlaciones	8
<b>3. Modelos Multidimensionales de Teoría de Respuesta al Ítem</b>	<b>11</b>
3.1. Modelos de TRIM para la interacción entre una persona y los ítems del test	12
3.1.1. Modelo de estudio para pruebas con ítems dicotómicos	13
3.2. Estimación de los parámetros del ítem y los rasgos latentes de la persona	15
3.2.1. Estimación Clásica	16
3.2.1.1. Estimación de parámetros de los ítems	16
3.2.1.2. Estimación de los rasgos latentes de las personas	17
3.2.2. Estimación Bayesiana	18
3.2.2.1. Estimación conjunta mediante métodos de MCMC	18
3.2.2.2. Diagnóstico de Convergencia del Método de MCMC	20
3.3. Implementación computacional	22
<b>4. Estudio de Simulación</b>	<b>24</b>
4.1. Condiciones de la simulación	24
4.1.1. Criterios para evaluar las estimaciones en la simulación	24
4.2. Métodos de estimación a estudiar	25
4.3. Análisis de la calidad de las estimaciones	25

4.3.1. Conclusión del estudio de simulación . . . . .	27
<b>5. Aplicación</b>	<b>28</b>
5.1. Instrumento de medición . . . . .	28
5.2. Análisis clásico de la prueba . . . . .	29
5.3. Análisis de la dimensionalidad del constructo actitudes hacia la Estadística .	30
5.4. Estimación de parámetros mediante el modelo estudiado . . . . .	33
5.4.1. Estimación de los parámetros usuales del M2PL de TRIM . . . . .	33
5.4.2. Análisis del poder discriminativo (MDISC) y la dificultad (MDIFF) del ítem del constructo actitudes hacia la Estadística . . . . .	37
5.4.3. Importancia de los ítems en las dimensiones del constructo actitudes hacia la Estadística . . . . .	42
5.4.4. Estimación de las dimensiones del constructo actitudes hacia la Estadística . . . . .	50
5.4.5. Análisis de los rasgos latentes hallados según algunas características de los evaluados . . . . .	52
5.4.6. Comparación con puntajes de la Teoría Clásica de los Test . . . . .	54
<b>6. Conclusiones</b>	<b>59</b>
6.1. Conclusión . . . . .	59
6.2. Sugerencias para investigaciones futuras . . . . .	61
<b>Referencias</b>	<b>62</b>
<b>A. Programas en R y R2WinBUGS</b>	<b>65</b>
<b>B. Resultados de convergencia</b>	<b>73</b>

## Lista de abreviaturas

TCT	Teoría Clásica de los Test .
TRI	Teoría de Respuesta al Ítem .
TRIU	Teoría de Respuesta al Ítem Unidimensional .
1PL	Modelo logístico de un parámetro .
2PL	Modelo logístico de dos parámetros .
3PL	Modelo logístico de tres parámetros .
AF	Análisis Factorial .
TRIM	Teoría de Respuesta al Ítem Multidimensional .
M2PL	Modelo multidimensional logístico compensatorio de dos parámetros .
MDIFF	Índice de dificultad en el caso multidimensional .
MDISC	Índice de discriminación en el caso multidimensional .
MVM	Máxima Verosimilitud Marginal .
EAP	Esperanza a posteriori .
EM	Esperanza - Maximización .
MCMC	Cadenas de Markov de Monte Carlo .
AEC	Escala de Actitudes hacia la Estadística .
T	Estadístico T student .
F	Estadístico F de Fisher .
Sig.	Nivel de significancia .

## Lista de símbolos

- $\theta$  Parámetro de un rasgo latente de la persona.
- $\eta$  Vector de parámetros de los ítems.
- $Y$  Puntuación en el ítem de la prueba.
- $y$  Valor posible para la puntuación en la prueba.
- $a$  Parámetro asociado a la discriminación del ítem.
- $b$  Parámetro asociado a la dificultad del ítem.
- $c$  Parámetro asociado a la adivinación del ítem.
- $\theta$  Vector de rasgos latentes de la persona.
- $\Theta$  Vector que contiene a los vectores  $\theta$  y  $\eta$
- $a$  Vector de discriminación del ítem en TRIM.
- $d$  Medida escalar asociada a la dificultad del ítem en TRIM.
- $B$  Índice de dificultad del ítem en TRIM.
- $A$  Índice de discriminación del ítem en TRIM.
- $n$  Cantidad de personas que responden los ítems.

## Índice de figuras

2.1. Esquema del coeficiente de correlación tetracórica . . . . .	10
3.1. Representación de la multidimensionalidad entre y dentro de cada ítem para un caso de dos dimensiones. . . . .	12
5.1. Gráfico de sedimentación . . . . .	31
5.2. Distribución de ítems según los factores encontrados . . . . .	33
5.3. Diagrama de cajas de las distribuciones a posteriori de $a_1$ en el modelo M2PL utilizando el método MCMC para los ítems de la escala de actitudes hacia la Estadística . . . . .	34
5.4. Diagrama de cajas de las distribuciones a posteriori de $a_2$ en el modelo M2PL utilizando el método MCMC para los ítems de la escala de actitudes hacia la Estadística . . . . .	35
5.5. Diagrama de cajas de las distribuciones a posteriori de $d$ en el modelo M2PL utilizando el método MCMC para los ítems de la escala de actitudes hacia la Estadística . . . . .	37
5.6. Diagrama de cajas de las distribuciones de MDISC en el modelo M2PL utilizando el método MCMC para los ítems de la escala de actitudes hacia la Estadística . . . . .	39
5.7. Diagrama de cajas de las distribuciones de MDIFF en el modelo M2PL utilizando el método MCMC para los ítems de la escala de actitudes hacia la Estadística . . . . .	41
5.8. Diagrama de dispersión de los ítems en función a MDISC y MDIFF . . . . .	42
5.9. Diagrama de cajas de las distribuciones de $a_1^*$ en el modelo M2PL utilizando el método MCMC para los ítems de la escala de actitudes hacia la Estadística . . . . .	44
5.10. Diagrama de cajas de las distribuciones de $a_2^*$ en el modelo M2PL utilizando el método MCMC para los ítems de la escala de actitudes hacia la Estadística . . . . .	46
5.11. Diagrama de distribución de ítems en función a las dimensiones halladas de actitudes hacia la Estadística . . . . .	48
5.12. Histograma de puntuaciones de $\theta_1$ . . . . .	51
5.13. Histograma de puntuaciones de $\theta_2$ . . . . .	51
5.14. Diagrama de caja de $\theta_1$ . . . . .	51
5.15. Diagrama de caja de $\theta_2$ . . . . .	51
5.16. Diagrama de dispersión de las puntuaciones de $\theta_1$ y $\theta_2$ . . . . .	52

5.17. Diagrama de dispersión de los índices de dificultad de TRIM (MDIFF) y TCT (Dificultad) . . . . .	54
5.18. Diagrama de dispersión de los índices de discriminación de TRIM (MDISC) y TCT (Discriminación) . . . . .	55
5.19. Diagrama de dispersión de las puntuaciones de $\theta_1$ y $MR1$ . . . . .	56
5.20. Diagrama de dispersión de las puntuaciones de $\theta_2$ y $MR2$ . . . . .	57



## Índice de cuadros

4.1. Comparación de las distancias entre valores estimados y valores simulados considerando diferentes software para estimar un modelo logístico bidimensional	26
4.2. Medidas de ajustes de la calidad de las estimaciones de los parámetros de los ítems con diferentes software	27
5.1. Ítems de actitudes hacia la Estadística	29
5.2. Matriz de correlaciones tetracóricas	31
5.3. Matriz de componentes rotados	32
5.4. Varianza explicada por los factores hallados	32
5.5. Medidas de resumen de la distribución a posteriori de $a_1$ en el modelo M2PL utilizando el método MCMC	34
5.6. Medidas de resumen de la distribución a posteriori de $a_2$ en el modelo M2PL utilizando el método MCMC	35
5.7. Medidas de resumen de la distribución a posteriori de $d$ en el modelo M2PL utilizando el método MCMC	36
5.8. Medidas de resumen de MDISC en el modelo M2PL utilizando el método MCMC	38
5.9. Medidas de resumen de MDIFF en el modelo M2PL utilizando el método MCMC	40
5.10. Medidas de resumen de $a_1^*$ en el modelo M2PL utilizando el método MCMC	43
5.11. Medidas de resumen de $a_2^*$ en el modelo M2PL utilizando el método MCMC	45
5.12. Estimación de medias para $a_1^*$ y $a_2^*$ de los ítems de actitudes hacia la Estadística	47
5.13. Ítems más relacionados con la primera dimensión hallada de actitudes hacia la Estadística	49
5.14. Ítems más relacionados con la segunda dimensión hallada de actitudes hacia la Estadística	49
5.15. Medidas resumen de las estimaciones de los rasgos latentes utilizando el método MCMC	50
5.16. Comparación según el género usando los puntajes de las dimensiones de las actitudes hacia la Estadística	53
5.17. Comparación según la facultad usando los puntajes de las dimensiones de las actitudes hacia la Estadística	53
5.18. Tabla de contingencia de las puntuaciones de $\theta_1$ (deciles) y $MR1$	57
5.19. Tabla de contingencia de las puntuaciones de $\theta_2$ (deciles) y $MR2$	58

## Capítulo 1

### Introducción

En el campo de la psicología y la educación, las investigaciones, generalmente se realizan con personas, lo cual implica desafíos para la medición. Ante estos, los estudios psicométricos, apoyado en métodos estadísticos, ayudan a ofrecer bases teóricas y técnicas cuantitativas para construir mejores instrumentos de medición en estos campos (Montero, 2000).

Tradicionalmente, se ha utilizado en mayor medida la Teoría Clásica de los Tests (TCT) como principal modelo psicométrico para construir y analizar ítems de diversas pruebas. Sin embargo, sus limitaciones justificaron complementar estos análisis estadísticos con modelos de Teoría de Respuesta al Ítem (TRI), los cuales se centran más en las propiedades individuales de los ítems que en las propiedades globales del test, permitiendo de esta manera construir pruebas más adecuadas y eficientes (Montero y Jiménez, 2013).

Los modelos más usados de TRI suponen solo un rasgo latente para explicar la probabilidad de respuesta al ítem (modelos unidimensionales o TRIU). Estos modelos de TRIU o Rasch están teniendo mayor presencia en los últimos años en el campo de la medición educativa nacional. Por ejemplo, son usados por el Ministerio de Educación para la Evaluación Censal de Estudiantes (ECE) que realizan cada año. Otra institución que hace uso de este tipo de modelos es la Pontificia Universidad Católica del Perú (PUCP) al momento de aplicar y analizar la información de sus evaluaciones de admisión.

A raíz de esto, se han desarrollado diversos trabajos de investigación en la PUCP, por ejemplo, el documento elaborado por los profesores Bazán, Valdivieso y Calderón (2010) sobre el Enfoque Bayesiano en Modelos de Teoría de Respuesta al Ítem. Asimismo, se han desarrollado tesis como la de Chincaro (2010), donde se analiza estadísticamente modelos Rasch unidimensionales y una aplicación con este enfoque en una prueba de comprensión lectora. En años posteriores se han seguido desarrollando investigaciones estadísticas sobre TRI en la PUCP, como el de Flores (2012) que hace un estudio de Modelos Testlet Logísticos y Logístico de Exponente Positivo. Por otra parte, Tarazona (2013) estudió Modelos Alternativos de Respuesta Graduada con aplicaciones en la calidad de servicios.

En la presente investigación se pretende continuar aportando al desarrollo de esta línea de estudio. Teniendo en cuenta que los modelos mencionados anteriormente asumen unidimensionalidad en la variable latente y que existen varias pruebas psicológicas y educativas que poseen variables que no pueden ser consideradas con una sola dimensión, en el sentido de no evaluar un único rasgo latente para la persona (Embretson y Reise, 2000), surge con mucho interés de estudio en el campo psicométrico los modelos de Teoría de Respuesta al

Ítem Multidimensional (TRIM), los cuales asumen la presencia de más de un rasgo latente en la persona para contestar un ítem de una prueba (Quaresma, 2014).

De esta manera, la presente investigación aporta al estudio de la TRIM explicando aspectos teóricos de los modelos y sus diversas formas de estimación (clásico y bayesiano). Asimismo, se plantea un análisis de la calidad de las estimaciones de los parámetros mediante una simulación. Finalmente, se desarrolla una aplicación en el campo de la psicología, para lo cual se utiliza un enfoque bayesiano y herramientas computacionales.

### 1.1. Objetivos

El objetivo general de la tesis es estudiar un modelo estadístico de la Teoría de Respuesta al Ítem Multidimensional con variables dicotómicas en la medición de una prueba psicológica. Se revisa teóricamente sus propiedades y métodos de estimación, enfatizando la metodología bayesiana. Asimismo, se exploran estos métodos de estimación mediante el software R y R2WinBUGS haciendo un estudio de simulación, en el cual se analiza la calidad de las estimaciones de los parámetros. Finalmente, se desarrolla una aplicación mediante implementaciones computacionales a un conjunto de datos que miden las actitudes hacia la Estadística en estudiantes universitarios colombianos (Pérez, Aparicio, Bazán y Abdounur, 2015). De manera específica, los objetivos son:

- Revisar la literatura acerca de las propuestas de modelos de TRIU y TRIM más comúnmente usados.
- Estudiar teóricamente las propiedades y métodos de estimación del modelo de TRIM elegido.
- Implementar computacionalmente un método de estimación bayesiana del modelo de TRIM propuesto mediante el software R y R2WinBUGS.
- Realizar un estudio de simulación acerca del modelo propuesto para analizar la calidad de las estimaciones de los parámetros.
- Aplicar el modelo propuesto a un conjunto de datos reales del ámbito psicológico, específicamente a una prueba que mide actitudes hacia la Estadística en estudiantes de una universidad privada de Colombia.

### 1.2. Organización del trabajo

En el Capítulo 2, se presentan las consideraciones preliminares relevantes como la Teoría de Respuesta al Ítem Unidimensional (TRIU) y el análisis de la dimensionalidad mediante el Análisis Factorial; estos temas servirán para entender y explicar mejor el desarrollo del tema a tratar en la presente investigación.

En el Capítulo 3, se explican los modelos multidimensionales propuestos de TRI dicotómicos; asimismo, se propone un modelo específico para el cual se explican algunos métodos de estimación y algunas de sus posibles implementaciones computacionales. En el Capítulo 4, se muestra un estudio de simulación en el cual se analiza la calidad de las estimaciones de las variables latentes desde el punto de vista clásico y bayesiano. En el Capítulo 5, se presenta

## CAPÍTULO 1. INTRODUCCIÓN

una aplicación del modelo propuesto de TRIM dicotómico en el ámbito de la Psicología, específicamente en una prueba que mide actitudes hacia la Estadística en estudiantes universitarios colombianos (Pérez et al., 2015). Finalmente, en el Capítulo 6 se presentan y discuten algunas conclusiones (metodológicas y de la aplicación) obtenidas en este trabajo. Se analizan las ventajas y desventajas de los métodos propuestos, y se hacen algunas sugerencias o recomendaciones para investigaciones futuras.

En el Apéndice A, se presentan los programas y algoritmos en R y R2WinBUGS que fueron utilizados en la simulación y en la aplicación al conjunto de datos real. En el Apéndice B se muestra los resultados y gráficos de convergencia de la aplicación.



## Capítulo 2

### Consideraciones preliminares

La Teoría de Respuesta al Ítem (TRI) ha sido reconocida como una de las principales contribuciones al desarrollo de la medición educativa y psicológica durante el siglo XX (Lee, 2012). La TRI es un marco general para la especificación de las funciones matemáticas que describen las interacciones de las personas y los ítems de una prueba. Estos modelos muestran la relación entre la habilidad o rasgo latente (simbolizado por  $\theta$ ) medido por un instrumento y una respuesta al ítem. Esta teoría es muy usada cuando se han cuantificado constructos de interés a través de cuestionarios que poseen ítems con respuestas dicotómicas o de más de dos categorías de respuesta (Demars, 2010).

Vale resaltar que la Teoría de Respuesta al Ítem Multidimensional (TRIM) es una consecuencia del análisis de las dimensiones y la Teoría de Respuesta al Ítem Unidimensional (TRIU). Como la forma en que los resultados del análisis de la TRIM se interpretan son mucho más parecidos a la TRIU (Reckase, 2009), este capítulo se enfocará más en la introducción de la TRIU con especial énfasis en los componentes que se pueden generalizar cuando se desarrollen los modelos TRIM. Asimismo, pero de manera más breve, se desarrollarán algunos aspectos teóricos relevantes del Análisis Factorial (AF) que pueden contribuir para el desarrollo del modelo estadístico elegido.

#### 2.1. Modelos de Teoría de Respuesta al Ítem Unidimensional

La TRIU comprende varios modelos que tienen la premisa básica de que las interacciones de una persona con los ítems de la prueba pueden ser adecuadamente representados por una expresión matemática que contiene un único parámetro que describe las características de la persona (Embretson y Reise, 2000). La representación básica de un modelo TRIU es dada en (2.1). En esta ecuación,  $\theta$  representa el único parámetro que describe las características de la persona, denominado también rasgo latente,  $\boldsymbol{\eta}$  representa un vector de parámetros que describen la características de los ítems de la prueba,  $Y$  representa la puntuación en el ítem de la prueba,  $y$  es un valor posible para la puntuación, y  $f$  es una función que describe la relación entre los parámetros y la probabilidad de la respuesta,  $P(Y = y)$ .

$$P(Y = y | \theta) = f(\theta, \boldsymbol{\eta}) \quad (2.1)$$

El supuesto de unidimensionalidad o de tener una única aptitud o rasgo  $\theta$  para explicar los resultados de las personas y las relaciones entre los ítems es una fuerte suposición. Un considerable número de investigaciones se han dedicado a la determinación de si esta suposi-

ción es razonable cuando se modela un conjunto de datos en particular (Reckase, 2009). En la realidad es complicado que este supuesto se cumpla exactamente debido a los múltiples factores que están presentes en el preciso momento de dar respuestas a una prueba; sin embargo, con este tipo de modelos se puede hablar de una aptitud fundamental o rasgo dominante que explique las respuestas al test (Martínez, Hernández y Hernández, 2006).

Junto con el anterior supuesto, estos modelos también asumen independencia, es decir, que la respuesta de una persona a un ítem no influye en la respuesta a un ítem producido por otra persona y también la respuesta de una persona a un ítem no afecta las tendencias de esa misma persona para responder de una manera particular a otro ítem (Reckase, 2009). Esto significa que estadísticamente existe independencia local o condicional al rasgo latente medido. Este supuesto está muy relacionado con la unidimensionalidad, ya que cuando se cumple, el espacio latente se define únicamente por un rasgo o aptitud.

La implicación del supuesto anterior es que para una persona dada o varias con el mismo parámetro  $\theta$ , la probabilidad conjunta de dar un determinado patrón de respuesta a un conjunto de ítems es igual al producto de las probabilidades de respuesta de la persona a los ítems individuales (Demars, 2010). Así la probabilidad de la matriz completa de las respuestas de  $n$  personas para  $I$  ítems en una prueba, es dada por lo siguiente:

$$P(\mathbf{Y} = \mathbf{y} \mid \boldsymbol{\theta}) = \prod_{j=1}^n \prod_{i=1}^I P(y_{ij} \mid \theta_j) \quad (2.2)$$

Estos dos supuestos son los principales y más importantes de la TRI; sin embargo, también existen otros de carácter más general, como el de monotonicidad, el cual asume que dar una puntuación máxima a un ítem aumenta a medida que se incrementa el nivel de aptitud o rasgo latente de la persona (Embretson y Reise, 2000).

De esta manera, con los supuestos señalados, se define de manera general los modelos de TRIU. Este tipo de modelos es común usarlos con pruebas que contengan ítems dicotómicos.

### 2.1.1. Modelos para pruebas con ítems dicotómicos

En este contexto, hay solo dos categorías y la probabilidad de una puntuación de 0 viene a ser igual a 1 menos la probabilidad de una puntuación de 1. En pruebas académicas la respuesta correcta se codifica usualmente como 1, en pruebas psicológicas la categoría que indica los niveles más altos del constructo suele codificarse con 1. La probabilidad de una respuesta con máximo puntaje o igual a 1 se expresa como una función de  $\theta$ . Cuando la probabilidad se determina para un determinado valor de  $\theta$ , esta puede ser interpretada como la probabilidad de una respuesta con máximo puntaje o igual a 1 para una persona elegida de manera aleatoria de un grupo de participantes con ese valor de  $\theta$  (Demars, 2010).

Los modelos de TRI más usados para analizar ítems dicotómicos son los modelos logísticos, los cuales cambian dependiendo del número de parámetros usados para resumir las características de los ítems de la prueba:

- El modelo logístico de un parámetro (1PL)

Uno de los modelos logísticos más interesantes se debe al matemático G. Rasch que aproximó el análisis de los datos de pruebas desde la teoría de la probabilidad, aunque

lo hizo desde un marco teórico diferente al de otros autores de estos modelos. Wright y Stone por el año 1979 introdujeron el modelo como la razón de probabilidades de acierto a un ítem, dado un nivel de aptitud (Embretson y Reise, 2000). De esta manera, este modelo sugerido tiene un parámetro para describir las características de la persona y un parámetro para describir las características del ítem (Martínez, Hernández y Hernández, 2006), y se puede representar de la siguiente manera:

$$P(Y_{ij} = y_{ij} | \theta_j) = f(\theta_j, b_i, y_{ij}) \quad (2.3)$$

donde  $y_{ij}$  es el puntaje para una persona  $j$  en el ítem  $i$  (0 ó 1),  $\theta_j$  es el parámetro que describe la característica relevante de la persona  $j$  y  $b_i$  es un parámetro que describe las características relativas al ítem  $i$  (que suele considerarse como la dificultad del ítem). De esta manera, si se basa el modelo en una transformación logarítmica de las escalas de los parámetros usados, la ecuación matemática del modelo resulta:

$$P(Y_{ij} = 1 | \theta_j, b_i) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}} \quad (2.4)$$

Dado que el modelo usa la función de distribución acumulada logística y solo tiene un único parámetro para los ítems, este se denomina un modelo logístico de un parámetro de TRI o Modelo de Rasch.

- El modelo logístico de dos parámetros (2PL)

Birnbaum (1968) propone un modelo ligeramente más complejo que el anterior añadiendo un parámetro de discriminación en el ítem ( $a_i$ ). Así, la ecuación que expresa este modelo es la siguiente:

$$P(Y_{ij} = 1 | \theta_j, a_i, b_i) = \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}} \quad (2.5)$$

Mientras que para el modelo logístico de un parámetro no importa qué ítems contesta correctamente la persona y la estimación de  $\theta$  depende solo del número de respuestas correctas, para el modelo logístico de dos parámetros, los ítems particulares de una prueba contestados correctamente afectan la estimación de  $\theta$  (Reckase, 2009). El modelo logístico de dos parámetros se aproxima a una distribución normal acumulativa, llamada una ojiva normal. El uso de la función ojiva normal tiene la desventaja de que requiere integración matemática y suele ser bastante complejo, por lo que es más práctico trabajar con funciones logísticas (Demars, 2010).

- El modelo logístico de tres parámetros (3PL)

Birnbaum y Lord (1980) modificaron el modelo de dos parámetros para dar cabida a un tercer parámetro que representa la contribución de la adivinación a la respuesta correcta. El modelo resultante es una transformación del modelo logístico de dos parámetros. La expresión matemática del modelo, donde  $c_i$  es el parámetro de asíntota inferior para el ítem  $i$ , es la siguiente:

$$P(Y_{ij} = 1 \mid \theta_j, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}} \quad (2.6)$$

El modelo logístico de tres parámetros con asíntotas inferiores únicas para cada ítem puede conducir a problemas de estimación. Para evitar esto se suele estimar una asíntota inferior común para todos los ítems o para un grupo de ítems similares. Cabe señalar que la dificultad de los ítems tiene un significado diferente en este modelo. A pesar de que la dificultad del ítem aún ocurre en el punto de inflexión de la gráfica, éste ya no coincidirá con el nivel de rasgo en el que la probabilidad de éxito es 0.5 pues el punto de inflexión se desplazará por la asíntota inferior debido al cambio que produce el parámetro  $c_i$  (Embretson y Reise, 2000).

### 2.1.2. Limitaciones de los modelos de TRIU

Como menciona Reckase (2009), los modelos expuestos anteriormente tienen la ventaja de tener una interpretación matemática menos compleja, de tener diversos ejemplos de aplicación y de tener evidencia de robustez ante las violaciones de los supuestos. Sin embargo, se sabe que la interacción actual entre personas y los ítems de una prueba no son tan simples como se asume en estos modelos. Es muy probable que las personas tengan más de un rasgo latente para explicar las respuestas de los ítems de una prueba y que los enunciados planteados en los ítems de la prueba requieran numerosos rasgos latentes para determinar una respuesta que involucre un puntaje máximo para un ítem. Por esta razón, aunque los modelos unidimensionales de TRI han demostrado ser útiles en determinadas condiciones, existe una necesidad de modelos de TRI más complejos que puedan reflejar con mayor precisión esta complejidad de las interacciones entre las personas y los ítems de una prueba.

Una forma de aumentar la capacidad de los modelos de TRI para describir mejor las interacciones de personas con los ítems de un test es asumir que las personas varían en un rango amplio de rasgos latentes y que los ítems requieren de éstos para ser contestados. De esta manera, esta extensión de modelos de TRIU describirá la interacción de los vectores de rasgos latentes de las personas con las características de los ítems de la prueba, en lugar de solo asumir un único parámetro de rasgo latente. Por esta razón, estos modelos con múltiples parámetros para la persona, se denominan modelos de Teoría de Respuesta al Ítem Multidimensional (Reckase, 2009).

## 2.2. Análisis de la dimensionalidad

Como lo manifiesta Montenegro (2010), los modelos multidimensionales requieren de la previa especificación de la dimensión del espacio de los rasgos latentes para su uso. Cuando las pruebas están divididas en subpruebas, cada una de ellas se diseña para medir un rasgo latente, lo cual reflejaría una dimensión de la prueba. Sin embargo, es posible que las dimensiones teóricas de la prueba no coincidan exactamente con las dimensiones encontradas mediante los datos observados.

Para estimar las dimensiones se suele usar la técnica del Análisis Factorial (AF), la cual, generalmente, es usada para encontrar un conjunto de variables latentes (factores) capaces de explicar la variabilidad de un número mayor de variables observadas; básicamente es un

procedimiento estadístico para reducir dimensiones y de esta manera encontrar las variables fundamentales que intervienen en la explicación de ciertos fenómenos (Levy y Varela, 2003). A continuación, se explicará brevemente la parte conceptual del análisis factorial exploratorio y las matrices de correlaciones usadas para ello dependiendo del tipo de variables a analizar.

### 2.2.1. Análisis Factorial Exploratorio

El objetivo de este análisis es explorar los datos para descubrir las dimensiones fundamentales subyacentes en las variables medidas sobre estos datos. Spearman por el año 1904 comenzó a desarrollar esta teoría planteando el modelo matemático de AF y los métodos de obtención de factores (Levy y Varela, 2003). Lo que busca este modelo es representar un grupo de variables observadas ( $X_i$ ) a través de otro conjunto de variables llamadas factores (variables latentes) por medio de un modelo lineal. Para esto, se asume el supuesto de que las variables observadas están estandarizadas y que los factores que representan a las variables pueden ser factores comunes presentes en todas las variables ( $f_j$ ) o factores específicos que forman parte de una sola variable ( $u_k$ ). De esta manera, el modelo busca explicar la varianza común presente en las variables (partiendo de una matriz de correlaciones) en términos de los factores  $f$ , los factores  $u$  solo dan cuenta de la varianza específica de cada variable (Giacconi, 2012).

El modelo factorial teórico es el siguiente:

$$X_i = \lambda_{i1} \cdot f_1 + \lambda_{i2} \cdot f_2 + \dots + \lambda_{im} \cdot f_m + L_{ii} \cdot u_i \quad (2.7)$$

donde:

$\lambda_{ij}$  = Coeficientes de la matriz  $\Lambda$  (matriz de cargas); ésta relaciona a los factores comunes con las variables observadas.

$L_{ii}$  = Coeficientes de la matriz  $L$ , que relacionan a los factores específicos con las variables observadas.

De esta manera las variables observadas se escriben como combinación lineal de los factores comunes y de los factores específicos. Finalmente, si se denota a  $X = (X_1, \dots, X_p)$ ,  $f = (f_1, \dots, f_m)$  y  $u = (u_1, \dots, u_p)$ , el modelo se puede representar de forma matricial de la siguiente manera:

$$X = \Lambda \cdot f + L \cdot u \quad (2.8)$$

#### 2.2.1.1. Matriz de correlaciones

El modelo factorial se aplica con mayor frecuencia usando como insumo la matriz de correlaciones de Pearson, la cual asume que las variables son cuantitativas y por lo general con distribución normal. Sin embargo, en contextos psicológicos, las variables manifiestas suelen tener un carácter cualitativo y en la mayoría de casos suelen ser el producto de la aplicación de una prueba en la cual se usaron escalas tipo Likert, por lo cual, se recomienda el uso de correlaciones policóricas si las variables son de carácter ordinal o tetracóricas en el caso de variables dicotómicas. (Valdivieso, Bayes y Tarazona, 2014).

### Correlación tetracórica

La correlación tetracórica brinda un valor de asociación entre dos variables latentes distribuidas normalmente cuando se dicotomizan en función a algún punto de corte o umbral. De esta forma, para obtener estas correlaciones se asume una variable aleatoria  $X^*$  continua asociada a la respuesta  $X$  dicotómica, siendo 1 (definido arbitrariamente) como la presencia subyacente de un atributo y 0 como la ausencia de este atributo, de tal manera que:

$$X_p = \begin{cases} 0 & \text{si } X_p^* \leq t_p, \\ 1 & \text{si } X_p^* > t_p \end{cases} \quad (2.9)$$

donde  $t_p$  es un parámetro desconocido denominado punto de corte o umbral asociado a un ítem  $p$ .

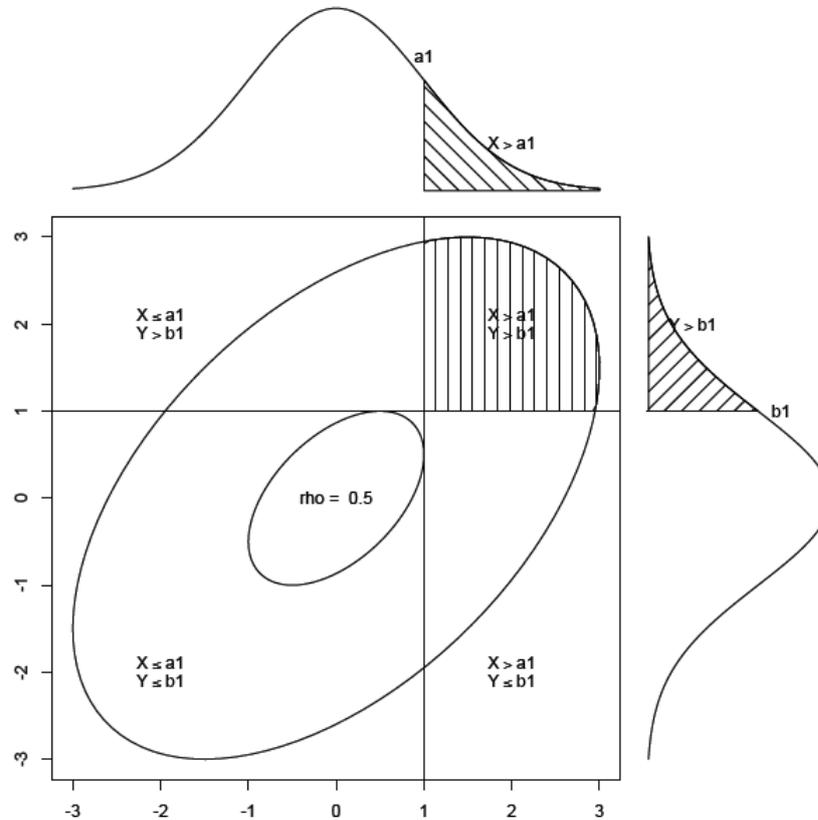
Por lo antes señalado, es factible realizar un análisis sobre las variables continuas  $X^*$  ya que al aplicar técnicas de análisis factorial con estas variables no se necesitará conocer sus valores sino solo la matriz de correlaciones. De esta manera, se estimarán las correlaciones de las variables  $X^*$  a través de los datos observados de las variables  $X$ ; estas correlaciones estimadas son las correlaciones tetracóricas de las variables  $X$  (Giacconi, 2012).

El coeficiente de correlación tetracórica entre dos variables dicotómicas  $X$  e  $Y$  es la estimación de la correlación entre las variables latentes  $X^*$  e  $Y^*$ , para las cuales se asume que tienen una distribución conjunta normal bivariada. Para explicarlo de manera más formal, se asume que  $X$  presenta la categoría 0 si  $X^* \leq a_1$  e  $Y$  presenta la categoría 0 si  $Y^* \leq b_1$ , donde  $a_1$  y  $b_1$  son los umbrales. De esta manera, el vector aleatorio  $(X^*, Y^*)$  presenta una distribución normal bivariada de vector de medias cero, varianzas unitarias y coeficiente de correlación  $\rho$ . Para estimar este último coeficiente se puede aplicar el método de máxima verosimilitud, donde la función de verosimilitud a maximizar viene dada por la correspondiente a la de la distribución multinomial de la tabla de contingencia de  $X$  e  $Y$  (Valdivieso, Bayes y Tarazona, 2014). El logaritmo de esta función estaría dado por:

$$L = \sum_{i=0}^1 \sum_{j=0}^1 N_{ij} \log(p_{ij}), \quad (2.10)$$

donde, por ejemplo,  $p_{11}$  es la probabilidad de que la distribución normal multivariada anterior tome valores en el rectángulo  $]a_1, +\infty[ \times ]b_1, +\infty[$  y  $N_{11}$  denota al número de valores de  $(X, Y)$  que toman la categoría 1 en ambas variables. Un esquema de correlación tetracórica estaría representado por el siguiente gráfico (Revelle, 2015):

Figura 2.1: Esquema del coeficiente de correlación tetracórica



La tabla de contingencia  $2 \times 2$  está pensada como una doble dicotomía de una distribución normal bivariada como se puede observar en la Figura 2.1, donde la función de densidad de la normal bivariada en forma de campana está encima de la tabla de contingencia. En función a esto, el coeficiente de correlación tetracórica sería el valor del parámetro para el cual los volúmenes de esta doble dicotomización de la distribución normal bivariada igualan las probabilidades conjuntas de la tabla de contingencia (Ekstrom, 2011).

## Capítulo 3

# Modelos Multidimensionales de Teoría de Respuesta al Ítem

Los modelos de Teoría de Respuesta al Ítem Multidimensional (TRIM) fueron desarrollados después de largos años de estudio e investigación, en respuesta a la dificultad del cumplimiento del supuesto de unidimensionalidad del rasgo latente requerido por los modelos de la TRI más usados. Esto es notorio en pruebas psicológicas de ejecución típica que al evaluarse no tienden a ser unidimensionales (Abal, Lozzia, Aguerri, Galibert y Atorresi, 2010). Estos modelos multidimensionales están basados en el supuesto de que las personas requieren más de una habilidad o rasgo latente básico  $\theta$  para contestar un ítem de una prueba (Quaresma, 2014).

Esta clase de modelos asumen que las respuestas observadas de las personas a una prueba, son consecuencia de la interacción entre el conjunto de parámetros de los ítems de la prueba y un conjunto de rasgos latentes de las personas (Antonio, 2013). Esto quiere decir, como mencionan Reckase (2009) y Fragoso y Curi (2013), que estos modelos están basados en la premisa de que la función matemática incluye como parámetros a dos vectores:

- Uno  $\theta$  con las múltiples características de la persona que describen las habilidades, actitudes y/o conocimientos (rasgos latentes) que el individuo aporta a la prueba.
- Otro  $\eta$  con las características del ítem que describen generalmente la dificultad del ítem de la prueba y la sensibilidad del ítem para poder diferenciar las características de las personas.

Dimensiones múltiples proporcionan mayor ajuste de los datos de respuesta al ítem cuando las personas difieren sistemáticamente de los elementos. En muchos modelos multidimensionales, múltiples parámetros de discriminación de los ítems representan el impacto de las dimensiones sobre temas específicos (Embretson y Reise, 2000).

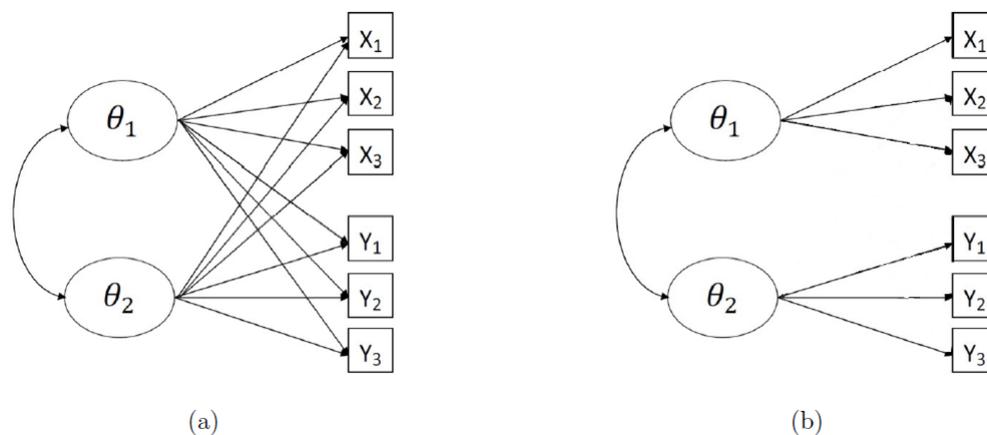
En adelante, se explorarán teóricamente los modelos de TRIM de interacción entre la persona y el ítem de la prueba que cumplan con tener dos categorías de puntuación (variable dicotómica) en la medición de la prueba; asimismo, se abordarán algunos métodos de estimación clásica y, sobre todo, bayesiana y sus implementaciones computacionales.

### 3.1. Modelos de TRIM para la interacción entre una persona y los ítems del test

La elección de un modelo multidimensional depende esencialmente de la forma como los ítems se asocian dentro del test, lo que implica un patrón de interacción entre las dimensiones latentes y los rasgos latentes de las personas al contestar los ítems del test y la relación entre las dimensiones latentes y los ítems (multidimensionalidad entre ítems o multidimensionalidad dentro de cada ítem) (Hasting y Hohler, 2009).

La multidimensionalidad dentro y entre cada ítem para el caso de dos dimensiones se puede observar en la Figura 3.1 (Quaresma, 2014):

Figura 3.1: Representación de la multidimensionalidad entre y dentro de cada ítem para un caso de dos dimensiones.



En la subfigura (a) los ítems se relacionan con ambas dimensiones y en la subfigura (b) los ítems se relacionan solo con su respectiva dimensión.

Estos tipos de multidimensionalidad son frecuentemente analizados con modelos compensatorios o no compensatorios, dependiendo de la naturaleza de las interacciones entre las dimensiones (Quaresma, 2014). Como lo manifiesta Fragoso y Curi (2013) y Reckase (2009), estos son los dos tipos principales de modelos de la TRIM, los cuales se definen por la forma en que el vector  $\theta$  de rasgos latentes se combina con las características del ítem para especificar la probabilidad de respuesta al ítem:

- *Modelos compensatorios*: estos que se ilustran en la subfigura (a) de la Figura 3.1 se basan en una combinación lineal de las componentes del vector  $\theta$ . Esta combinación lineal se utiliza con una ojiva normal o con una forma logística para especificar la probabilidad de una respuesta. Esta combinación puede producir la misma suma con varias combinaciones de las componentes de  $\theta$ . Si una componente es baja, la suma puede ser la misma si otra componente es lo suficientemente alta. Es importante mencionar que estos modelos tienen estrecha relación con el análisis factorial y son los más usados en la literatura de la TRIM.
- *Modelos no compensatorios o parcialmente compensatorios*: estos que se ilustran en la subfigura (b) de la Figura 3.1 separan las tareas cognitivas en partes y utilizan un

modelo unidimensional para analizar cada parte. La probabilidad de respuesta correcta para el ítem es el producto de las probabilidades de cada parte. El uso del producto de probabilidades tiene como resultado características no lineales para esta clase de modelos.

La representación general de un modelo de TRIM es dada según la ecuación que se muestra a continuación:

$$P_i(\mathbf{Y} = \mathbf{y} \mid \boldsymbol{\theta}) = f(\boldsymbol{\theta}, \boldsymbol{\eta}_i, \mathbf{y}) \quad (3.1)$$

donde  $\boldsymbol{\eta}$  representa un vector de parámetros que describen las características de los ítems de la prueba,  $\mathbf{Y}$  representa la puntuación en el ítem de la prueba,  $\mathbf{y}$  es un valor posible para la puntuación y  $f$  es una función que describe la relación entre las características de las personas ( $\boldsymbol{\theta}$ ) y la probabilidad de responder obteniendo el máximo puntaje (Reckase, 2009). Cabe resaltar, que los supuestos asumidos para los modelos de TRIU de independencia local y monotonicidad también se asumen para los de TRIM.

### 3.1.1. Modelo de estudio para pruebas con ítems dicotómicos

Algunas de las motivaciones para desarrollar los modelos de TRIM, según Reckase (2009), proceden de los intentos por resolver el problema de realizar un análisis factorial a un conjunto de datos dicotómicos. Por esta razón, los modelos de TRIM para ítems dicotómicos han aparecido en diversas investigaciones desde la década de 1980.

El modelo elegido para la siguiente investigación, debido al interés de analizar datos de pruebas psicológicas de ejecución típica (poco estudiadas por modelos de TRI), es el modelo multidimensional logístico compensatorio de dos parámetros (M2PL); ya que en estos contextos las personas evaluadas no deberían tener motivos a priori para responder los ítems al azar.

Este modelo propuesto, como manifiestan Fragoso y Curi (2013), Reckase (2009) y Quarresma (2014), es una extensión multidimensional del modelo unidimensional 2PL, en el que el exponente  $a(\theta - b)$ , de la forma lineal  $a\theta + d$ , tomará ahora la forma  $\mathbf{a}\boldsymbol{\theta}' + d$ , siendo  $\mathbf{a}$  un vector de  $1 \times m$  del parámetro vectorial de discriminación del ítem y  $\boldsymbol{\theta}$  es un vector de  $1 \times m$  coordenadas de la persona, siendo  $m$  el número de dimensiones en el espacio de coordenadas. Asimismo, el intercepto  $d$  es una medida escalar asociada a la dificultad del ítem. La forma del modelo, con opciones de respuesta dicotómicas, estaría dada por:

$$P_i(\boldsymbol{\theta}_j) = P(Y_{ij} = 1 \mid \boldsymbol{\theta}_j, \mathbf{a}_i, d_i) = \frac{e^{\mathbf{a}_i \boldsymbol{\theta}_j' + d_i}}{1 + e^{\mathbf{a}_i \boldsymbol{\theta}_j' + d_i}} \quad (3.2)$$

El exponente en este modelo puede ser descompuesto para mostrar la manera en que los elementos de los vectores  $\mathbf{a}$  y  $\boldsymbol{\theta}$  interactúan:

$$\mathbf{a}_i \boldsymbol{\theta}_j' + d_i = a_{i1}\theta_{j1} + a_{i2}\theta_{j2} + \cdots + a_{im}\theta_{jm} + d_i = \sum_{l=1}^m a_{il}\theta_{jl} + d_i \quad (3.3)$$

El exponente es una función lineal con los elementos del vector  $\boldsymbol{\theta}$  con el parámetro  $d$  como intercepto y los elementos del vector  $\mathbf{a}$  como parámetros de pendiente, de esta manera

la expresión en el exponente es lineal en una dimensión  $m$ . Una propiedad interesante e importante de este modelo es que si el exponente se ajusta a un valor constante  $k$ , todos los vectores  $\theta$  que satisfacen la expresión  $k = \mathbf{a}_i \theta'_j + d_i$  caen a lo largo de una línea recta y todos ellos tendrán la misma probabilidad de respuesta correcta para el modelo (Reckase, 2009).

Para explicar mejor lo expuesto en el párrafo anterior, se asumirán solo dos dimensiones y se analizará el caso en que  $k = 0$ . De esta forma se encontrará que la probabilidad de acierto será siempre 0.5. Si se tiene un vector  $\mathbf{a} = [a_1, a_2]$  y el parámetro  $d$ , entonces del exponente  $k = 0 = a_1\theta_1 + a_2\theta_2 + d$ , se tendría que:

$$\theta_2 = -a_1\theta_1/a_2 - d/a_2 \quad (3.4)$$

Reckase (2009), manifiesta que si se logra estimar los valores del vector  $\mathbf{a}$  y el parámetro  $d$ , entonces se podría graficar una recta en el plano  $(\theta_1, \theta_2)$ . De esta forma, se encuentra una propiedad relevante de este modelo de TRIM, que demuestra que todas las personas con un vector  $\theta$  que se encuentran en la recta, tendrán una probabilidad de 0.5 de contestar correctamente el ítem. Cuando las coordenadas  $(\theta_1, \theta_2)$  son interpretadas como habilidades, entonces la característica antes encontrada indica que una alta capacidad en una dimensión puede compensarse con una baja capacidad en la otra dimensión. Es por esto, que este tipo de modelos se denominan compensatorios.

Las respuestas de los ítems se pueden asumir como señales parciales de los rasgos latentes de las personas. Si una prueba está formada por  $I$  ítems, el patrón de respuestas de la persona será un vector en el espacio  $\{0, 1\}^I$ . De esta manera, el vector de rasgos latentes de una persona es una representación de su patrón de respuesta en un espacio euclidiano de dimensiones reducidas (Montenegro, 2010).

En este tipo de modelos, como se vio anteriormente, se calcula un parámetro de intercepto  $d$  para cada ítem, el cual es un escalar. En este caso, éste no sería más un parámetro de dificultad como suele interpretarse en los modelos de TRIU, ya que éste no da un indicador único de la dificultad del ítem. En lugar de esto, el índice de dificultad para el caso multidimensional se define como:

$$MDIFF_i = B_i = \frac{-d_i}{\sqrt{\sum_{l=1}^m a_{il}^2}} \quad (3.5)$$

donde  $d_i$  es el parámetro de intercepto del ítem  $i$ ,  $a_{il}$  denota el parámetro de discriminación del ítem  $i$  en la dimensión  $l$  y  $m$  denota el número de dimensiones del espacio de rasgos latentes. El valor de  $B_i$  o MDIFF tiene la misma interpretación que en los modelos unidimensionales (Reckase, 2009).

El índice de discriminación multidimensional para el ítem  $i$  se define como:

$$MDISC_i = A_i = \sqrt{\sum_{l=1}^m a_{il}^2} \quad (3.6)$$

Esto representaría la norma del vector de discriminación. Además, como  $A_i$  tiene la misma forma matemática que el denominador de  $B_i$ , también se podría expresar la dificultad del

ítem multidimensional como  $B_i = -d_i/A_i$ , el cual es un simil de la TRIU.

La mejor forma para interpretar los parámetros de discriminación según [Fragoso y Curi \(2013\)](#) es transformar o estandarizar el coeficiente para obtener un rango posible entre 0 y 1, de tal forma que se pueda interpretar de forma muy parecida a un análisis factorial. El cálculo de estos nuevos coeficientes ( $a_i^*$ ) estaría dado por:  $a_i^* = a_i/MDISC_i$

### 3.2. Estimación de los parámetros del ítem y los rasgos latentes de la persona

La estimación de los parámetros en los modelos de TRI han sido una ciencia progresiva en los últimos 60 años, pasando de técnicas de estimación heurística hasta métodos más avanzados como el bayesiano mediante Cadenas de Markov de Monte Carlo ([Baker y Kim, 2004](#)).

Los modelos de TRI poseen un gran número de parámetros a ser estimados, por un lado están los que se refieren a los ítems (parámetros estructurales) y los que se refieren a los rasgos latentes de las personas (parámetros incidentales), los cuales aumentarán si el tamaño de muestra es mayor ([Fragoso, 2010](#)).

Cuando se pretende estimar los parámetros en los modelos de TRIM la dificultad aumenta, tanto por la complejidad del modelo como por la mayor cantidad de parámetros a estimar. Asimismo, [Bock y Aitkin \(1981\)](#) señalan que las estimaciones mediante máxima verosimilitud en estos modelos no necesariamente cumplen con las propiedades de ausencia de sesgo, consistencia y eficiencia, como si ocurre en los modelos más simples como los TRIU ([Baker y Kim, 2004](#)).

Para la estimación de los parámetros del modelo estadístico elegido (M2PL) se puede utilizar el método clásico, estimando primero los parámetros de los ítems mediante Máxima Verosimilitud Marginal (MVM) y el algoritmo de Esperanza - Maximización (EM), para luego estimar los rasgos latentes de la persona, lo cual se puede resolver con métodos de Máxima Verosimilitud y Esperanza a Posteriori (EAP). Sin embargo, se utilizará el método bayesiano, ya que análisis previos e investigaciones anteriores coinciden en concluir que éste método es más adecuado cuando se estiman parámetros en los modelos de TRI ([González, 2010](#)). Asimismo, el enfoque bayesiano logra realizar extensiones naturales en los modelos de TRI y permite una descripción inicial de los parámetros mediante la especificación de la priori ([Fox, 2010](#)).

Un método efectivo para la estimación bayesiana es considerar métodos de generación de distribuciones a partir de Cadenas de Markov de Monte Carlo (MCMC), la cual logra una estimación conjunta de los parámetros de los ítems y de los rasgos latentes de las personas en el contexto de TRI ([Fragoso y Curi, 2013](#)).

Con fines ilustrativos, se comenzará explicando brevemente una de las formas de estimación clásica y luego se centrará el análisis en el método de estimación bayesiano conjunto mediante MCMC.

### 3.2.1. Estimación Clásica

#### 3.2.1.1. Estimación de parámetros de los ítems

Bock y Aitkin (1981) fueron los primeros en proponer un método factible para estimar los parámetros de los ítems de pruebas mediante un método similar al del algoritmo EM. Sin embargo, este método suele ser adecuado solo cuando existen pocas o moderadas soluciones factoriales y el número de cuadraturas por dimensión disminuye a medida que el número de factores aumenta (Chalmers, 2012).

Según Fragoso (2010), si denotamos con  $j = 1, \dots, n$  a las personas que responden a los ítems  $i = 1, \dots, I$ ,  $\theta_j$  al vector de rasgos latentes del  $j$ -ésimo individuo,  $y_j = (y_{1j}, \dots, y_{Ij})$  a un patrón de respuestas dicotómicas del individuo  $j$  y  $P_i(\theta_j)$  al modelo propuesto anteriormente (en 3.2), se podrá establecer la siguiente función de verosimilitud para la persona  $j$

$$L_j(\theta) = P(Y_j = y_j \mid \theta_j, a_i, d_i) = \prod_{i=1}^I [P_i(\theta_j)]^{y_{ij}} [1 - P_i(\theta_j)]^{1-y_{ij}} \quad (3.7)$$

Si se considera que  $\theta_j$  tiene una distribución normal  $p$ -variada, la probabilidad incondicional de que el individuo  $j$  presente un patrón de respuestas  $y_j$  viene dada por:

$$\tilde{P}_j = P(Y_j = y_j \mid a_i, d_i) = \int_{\theta} L_j(\theta) g(\theta) d\theta \quad (3.8)$$

Dada la presencia de la densidad normal ( $g(\theta)$ ), la última integral puede aproximarse usando el método de cuadratura gaussiana a un  $\tilde{P}_l$ , la cual consiste en un agrupamiento de muestras en torno a ciertos niveles de rasgos latentes, lo que induce a escribir una función de verosimilitud por una distribución multinomial (Fragoso, 2010).

Si  $r_l$  denota la frecuencia de un patrón de respuesta  $y_l$ , para cada uno de los  $s$  patrones de respuestas distintos, la función de verosimilitud del modelo de TRIM se define como:

$$L = \prod_{l=1}^s \tilde{P}_l^{r_l} \quad (3.9)$$

La derivada de la función log-verosimilitud con respecto a un parámetro del ítem,  $v_i$  ( $a_i$  o  $d_i$ ), es dada por

$$\begin{aligned} \frac{\partial \log(L)}{\partial v_i} &= \sum_{l=1}^s r_l \frac{\partial \tilde{P}_l}{\tilde{P}_l} \left( \frac{\partial \tilde{P}_l}{\partial v_i} \right) \\ &= \sum_{l=1}^s \frac{r_l}{\tilde{P}_l} \int_{\theta} \left( \frac{L_l(\theta)}{[P_i(\theta)]^{y_{li}} [1 - P_i(\theta)]^{1-y_{li}}} \cdot \frac{\partial [P_i(\theta)]^{y_{li}} [1 - P_i(\theta)]^{1-y_{li}}}{\partial v_i} g(\theta) \right) d\theta \\ &= \sum_{l=1}^s \frac{r_l}{\tilde{P}_l} \int_{\theta} \left( \frac{y_{li} - P_i(\theta)}{P_i(\theta)[1 - P_i(\theta)]} \right) L_l(\theta) \frac{\partial P_i(\theta)}{\partial v_i} g(\theta) d\theta. \end{aligned} \quad (3.10)$$

Definiendo

$$\bar{R}_l = \sum_{i=1}^s \frac{r_l y_{li} L_l(\theta)}{\tilde{P}_l} \quad (3.11)$$

y

$$\bar{n} = \sum_{l=1}^s \frac{r_l L_l(\boldsymbol{\theta})}{\tilde{P}_l}, \quad (3.12)$$

tenemos que la derivada de la función log-verosimilitud puede ser re-escrita como

$$\frac{\partial \log(L)}{\partial v_i} = \int_{\boldsymbol{\theta}} \frac{\bar{R}_l - \bar{n} P_i(\boldsymbol{\theta})}{P_i(\boldsymbol{\theta}) [1 - P_i(\boldsymbol{\theta})]} \cdot \frac{\partial P_i(\boldsymbol{\theta})}{\partial v_i} g(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (3.13)$$

expresión que se podría aproximar con una técnica llamada de cuadratura de Gauss-Hermite, la cual reemplaza el problema de buscar la suma del área que se encuentra bajo la curva continua por el simple problema de sumar las áreas de un número finito de rectángulos que se aproxima al área bajo la curva (Baker y Kim, 2004).

Las ecuaciones definidas en (3.11) y (3.12) representan la frecuencia esperada de aciertos del ítem  $i$  y el número esperado de personas con el número de rasgos latentes determinados. Estas ecuaciones se logran encontrar con el paso E del algoritmo EM usado en la obtención de las estimaciones. Asimismo, el paso M es presentado en la ecuación (3.13) o en su aproximación de cuadratura de Gauss-Hermite, en la cual se maximiza utilizando el método de aceleración del algoritmo EM (Fragoso, 2010). El proceso de EM se repite hasta que el cambio entre iteraciones cae por debajo de cierta tolerancia pre-especificada (Chalmers, 2012).

### 3.2.1.2. Estimación de los rasgos latentes de las personas

Como menciona Fragoso (2010), suponiendo que se conocen los parámetros de los ítems, la estimación del vector de rasgos latentes para las personas puede ser encontrada resolviendo las ecuaciones de verosimilitud para  $k = 1, \dots, p$

$$\frac{\partial l(\theta_s | \mathbf{a}, \mathbf{d}, \mathbf{Y})}{\partial \theta_k} = 0 \quad (3.14)$$

donde  $l(\theta_s | \mathbf{a}, \mathbf{b}, \mathbf{Y})$  es la función log-verosimilitud del  $s$ -ésimo término de respuesta observada,  $s = 1, \dots, S$  es condicionada a una matriz de respuestas observadas  $\mathbf{Y}$  y los parámetros de los ítems  $\mathbf{a} = (a_1, \dots, a_I)$  y  $\mathbf{d} = (d_1, \dots, d_I)$  se obtienen como en la subsección anterior.

Sin embargo, la estimación de rasgos latentes por máxima verosimilitud implica la resolución de  $pS$  ecuaciones no lineales o maximizar  $S$  funciones de verosimilitud, lo cual puede ser demasiado complicado. Una alternativa es la estimación bayesiana por la EAP. Se aprovecharían las estimaciones de cuadratura por máxima verosimilitud marginal y con la premisa de normalidad multivariada del vector de rasgos latentes se puede estimar el  $k$ -ésimo componente del vector de rasgos latentes,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k, \dots, \theta_p)$  por el valor esperado de la distribución a posteriori del vector de rasgos latentes, dado por:

$$\hat{\theta}_{ks} = \frac{\int_{RP} \theta_k L_s(\boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\tilde{P}_s}, \quad (3.15)$$

siendo  $\tilde{P}_s$  la probabilidad marginal del  $s$ -ésimo término de respuesta definida en (3.8). La aproximación por cuadratura de Gauss-Hermite de (3.15) viene dada por:

$$\theta_{ks} \approx \frac{\sum_{m \in \mathbf{K}} K_{mk} L_s(\mathbf{K}_m) A(\mathbf{K}_m)}{\tilde{P}_s}, \quad (3.16)$$

para la cual se usan aproximaciones definidas para un conjunto de puntos de cuadratura  $p$ -dimensional  $\mathbf{K}$ .

### 3.2.2. Estimación Bayesiana

#### 3.2.2.1. Estimación conjunta mediante métodos de MCMC

Para estimar los parámetros de los ítems y los rasgos latentes de manera conjunta se usa una perspectiva bayesiana, lo cual resulta conveniente, ya que refleja mejor el problema original de la TRIM, el cual es estimar los rasgos latentes de las personas y de los ítems de manera simultánea y completa. Asimismo, con esta perspectiva se tratan a todos los parámetros (para ítems y personas) como variables aleatorias, de esta forma, la no identificabilidad acerca de sus valores es recogida por una especificación de una distribución a priori para ellas (Bazán, Valdivieso y Calderón, 2010).

En la inferencia bayesiana aplicada a la TRIM se considerará que  $\Theta$  es un vector aleatorio no determinístico (que contiene a  $\theta$  y  $\eta$ ) y por lo tanto tiene una distribución de probabilidades asociado. Por lo tanto, se puede asignar información preliminar a  $\Theta$ , lo cual se sintetiza proponiendo una distribución a priori para  $\Theta$ , denominada  $f(\Theta)$ .

Asimismo, los datos se organizan en la función de verosimilitud  $L(\mathbf{y}|\Theta)$ , donde  $\mathbf{y}$  es un vector de datos observados del vector aleatorio  $\mathbf{Y}$ , cuya distribución depende de  $\Theta$ . Usando el teorema de Bayes se podría obtener la distribución a posteriori de  $\Theta$  dados los siguientes datos:

$$f(\Theta|\mathbf{y}) = \frac{f(\Theta, \mathbf{y})}{f(\mathbf{y})} = \frac{L(\mathbf{y}|\Theta)f(\Theta)}{f(\mathbf{y})} \quad (3.17)$$

Como la distribución marginal  $f(\mathbf{y})$  no depende de  $\Theta$  y la distribución a posteriori es proporcional a la verosimilitud y a la priori, la ecuación se puede simplificar a:

$$f(\Theta|\mathbf{y}) \propto L(\mathbf{y}|\Theta)f(\Theta) \quad (3.18)$$

La distribución a posteriori  $f(\Theta|\mathbf{y})$  contiene toda la información relevante del vector desconocido  $\Theta$  dada la data observada  $\mathbf{y}$ . De esta forma, toda inferencia estadística puede deducirse de  $f(\Theta|\mathbf{y})$  al considerarse un resumen adecuado. Estos resúmenes suelen escribirse de la siguiente forma integral:

$$I = \int g(\Theta)f(\Theta|\mathbf{y})d\Theta. \quad (3.19)$$

El problema con esta ecuación es que generalmente es muy difícil o imposible encontrar  $I$  y más aún, como es el caso, si  $\Theta$  es multidimensional (Bazán et al., 2010).

Por lo antes descrito, el método bayesiano que se suele usar para estimar los parámetros es el de simulación de Cadenas de Markov de Montecarlo (MCMC), este método ha incrementado su popularidad para la estimación de los parámetros de modelos estadísticos

complejos debido a la efectividad y disminución en los costes informáticos. En el paradigma de esta estimación se asume una distribución de probabilidad a priori para el parámetro que uno va a estimar y obtener de esta manera posibles valores de dichos parámetros según la distribución de probabilidad de estos parámetros encontrada en la base de datos observada. Esta distribución es la a posteriori y resume estadísticas de los valores muestreados que se pueden utilizar para obtener estimaciones puntuales e intervalos de credibilidad para los parámetros del modelo estadístico (Fragoso, 2010).

La lógica de estos métodos se basa en diseñar iterativamente una cadena de Markov para  $\Theta$  de tal manera que  $f(\Theta|\mathbf{y})$  sea su distribución ergódica estacionaria. Empezando en algún estado inicial  $\Theta_0$ , la idea es simular un número suficientemente grande  $M$  de transiciones bajo la cadena de Markov y registrar los correspondientes estados simulados  $\Theta_t$ . Es posible demostrar, bajo ciertas condiciones de regularidad, que la media muestral ergódica

$$\hat{I} = \frac{1}{M} \sum_{t=1}^M g(\Theta_t) \quad (3.20)$$

converge a la integral deseada anteriormente señalada, dando de esta forma una buena aproximación de  $I$ . Por esto, el desafío de los métodos MCMC es precisar una cadena de Markov adecuada con la distribución a posteriori  $f(\Theta|\mathbf{y})$  como su distribución estacionaria y decidir cuando detener la simulación (Bazán et al., 2010).

Si se tiene una distribución conjunta  $p(\boldsymbol{\theta}, \boldsymbol{\eta})$  donde  $\boldsymbol{\theta}$  es el vector del rasgo latente y  $\boldsymbol{\eta}$  el vector de los parámetros de los ítems. El objetivo es encontrar la distribución a posteriori conjunta, tal como:  $p(\boldsymbol{\theta}, \boldsymbol{\eta} | \mathbf{Y}) \propto p(\mathbf{Y} | \boldsymbol{\theta}, \boldsymbol{\eta}) p(\boldsymbol{\theta}, \boldsymbol{\eta})$ . Con el fin de encontrar una distribución conjunta de este tipo, se corre la Cadena de Markov, con un núcleo de transición, la probabilidad de pasar a un nuevo estado  $(\boldsymbol{\theta}^{t+1}, \boldsymbol{\eta}^{t+1})$ , dado el estado actual de la cadena  $(\boldsymbol{\theta}^t, \boldsymbol{\eta}^t)$ . Existen dos núcleos de transición conocidos, el muestreo de Gibbs y el esquema Metropolis-Hasting, los cuales son por construcción invariantes con respecto a la distribución a posteriori buscada (Lee, 2012).

Fragoso (2010), menciona que para obtener estimaciones de los parámetros de interés, se construyen núcleos de transición que convenientemente produzcan cadenas teniendo una distribución a posteriori de interés como sus distribuciones estacionarias. Así, a partir de un cierto momento  $t^*$ , los elementos de  $M_t$  pueden ser considerados muestras aleatorias de una distribución a posteriori en cuestión.

El valor de  $t^*$  es elegido a partir de la convergencia de la cadena. Las observaciones generadas antes de la interacción  $t^*$ , es decir, antes de la convergencia, son descartadas de ese período y se denominan “burn in”. Las demás observaciones son usadas para realizar inferencias y obtener estimaciones de los parámetros de interés. Existen diferentes métodos para determinar el número de iteraciones a ser descartadas en una convergencia de cadena.

Para obtener las muestras de los parámetros se utiliza un método de Muestreador de Gibbs que obtiene muestras de distribución estacionaria de cadena utilizando un procedimiento iterativo:

1. Simulación de un valor inicial para la cadena  $(\mathbf{a}^{(0)}, \mathbf{d}^{(0)}, \boldsymbol{\theta}^{(0)})$

2. Simulación de la observación de la muestra  $\theta_{jk}^{(1)}$  de la distribución condicional completa  $\pi(\theta|\mathbf{a}^{(0)}, \mathbf{d}^{(0)}, \theta^{(0)})$  obtenida a partir de la distribución estacionaria  $\pi(\cdot)$  para todo  $j, k$  en  $j = 1, \dots, N$  y  $k = 1, \dots, p$ .
3. Simulación de  $d_i^{(1)}$  de la distribución  $\pi(d|\mathbf{a}^{(0)}, \mathbf{d}^{(0)}, \theta^{(1)})$  para todo  $i = 1, \dots, I$
4. Simulación de  $a_{ik}^{(1)}$  de la distribución  $\pi(a|\mathbf{a}^{(0)}, \mathbf{d}^{(0)}, \theta^{(1)})$  para todo  $i = 1, \dots, I, k = 1, \dots, p$
5. Utilizando el punto  $(\mathbf{a}^{(1)}, \mathbf{d}^{(1)}, \theta^{(1)})$  en el paso 1, repita el procedimiento.

Sin embargo, obtener las distribuciones condicionales de cada parámetro no es una tarea sencilla. Frecuentemente, las distribuciones no tienen una forma analítica establecida. Para eso se utilizará el algoritmo de Metropolis - Hastings para obtener las distribuciones deseadas.

El algoritmo de Metrópolis-Hastings consiste en tomar un núcleo de transición mas conveniente  $q(\mathbf{a}^{(0)}, \mathbf{d}^{(0)}, \theta^{(0)}, \mathbf{a}^{(1)}, \mathbf{d}^{(1)}, \theta^{(1)})$ , generar una observación de esa distribución  $(\mathbf{a}^*, \mathbf{d}^*, \theta^*)$  y calcular la probabilidad de aceptación de un valor generado como un próximo movimiento de cadena, dada por

$$\alpha(\mathbf{a}^{(0)}, \mathbf{d}^{(0)}, \theta^{(0)}, \mathbf{a}^{(*)}, \mathbf{d}^{(*)}, \theta^{(*)}) = \min \left\{ \frac{\pi(\mathbf{a}^{(*)}, \mathbf{d}^{(*)}, \theta^{(*)})q(\mathbf{a}^*, \mathbf{d}^*, \theta^*, \mathbf{a}^{(0)}, \mathbf{d}^{(0)}, \theta^{(0)})}{\pi(\mathbf{a}^{(0)}, \mathbf{d}^{(0)}, \theta^{(0)})q(\mathbf{a}^{(0)}, \mathbf{d}^{(0)}, \theta^{(0)}, \mathbf{a}^{(*)}, \mathbf{d}^{(*)}, \theta^{(*)})}, 1 \right\}. \quad (3.21)$$

Un valor aleatorio  $U$  de una distribución uniforme de intervalo  $(0,1)$ , donde  $U \leq \alpha$ , un valor de observación actualizado como  $(\mathbf{a}^*, \mathbf{d}^*, \theta^*)$

Un problema de estimación por MCMC es la alta correlación entre las observaciones, por esta razón se genera una adecuación modificando los algoritmos de bloques de ciertos conjuntos de parámetros como el vector de discriminación de rasgos latentes.

Se suelen especificar las distribuciones a priori, según Patz y Junker (1999), de la siguiente manera: para los rasgos latentes, es usual plantear una distribución normal estandar,  $\theta_j \sim N(0, I_p)$ . Para cada componente de los vectores de la discriminación se asumen distribuciones log normal  $(1, 2)$ ,  $a_{ik} \sim \text{lognormal}(1, 2)$ , y para los parámetros de dificultad se suele tomar una distribución normal estandar,  $d_i \sim N(0, 1)$ .

### 3.2.2.2. Diagnóstico de Convergencia del Método de MCMC

Una parte crucial al usar los métodos de MCMC en la estimación de parámetros es evaluar cuan bien se está desarrollando el algoritmo de MCMC, es decir, evaluar en qué etapa la distribución de los valores de los parámetros producidos por la cadena de Markov pueden ser considerados como los de la distribución estacionaria de la cadena, la cual es la distribución a posteriori de los parámetros dados los datos (Bartholomew, Knott y Moustaki, 2011). Sin evidencia de haber llegado correctamente a la distribución de destino (distribución estacionaria), las inferencias realizadas a partir de éste método podrían ser cuestionadas (Lee, 2012). Varios estudios sugieren diversas maneras de verificar la convergencia, entre ellas las más conocidas son la de los criterios de Geweke (1992) y Gelman-Rubin (1992), las cuales permiten inferir la convergencia basándose solo en las muestras. Se determina un período de

## CAPÍTULO 3. MODELOS MULTIDIMENSIONALES DE TRI

*burn-in* y de intervalo entre las iteraciones para minimizar las autocorrelaciones según el criterio de Raftery-Lewis (1995).

Se simplificará la notación usando  $\eta_t$  para denotar una muestra de un parámetro de un modelo obtenido por alguno de los métodos de MCMC comentados anteriormente.

El criterio de Raftery-Lewis se basa en un proceso de estimación de un cuantil fijado  $q = P(\pi(\eta_t) \leq u)$  de la distribución a posteriori con un error  $r$  de probabilidad de cobertura  $s$ . Se construye la secuencia de variables aleatorias

$$Z_t = \begin{cases} 1, & \text{si } \eta_t \leq u \\ 0, & \text{c.c.} \end{cases} \quad (3.22)$$

sobre las cuales se obtienen subsecuencias, para  $k > 1$

$$Z_t^{(k)} = Z_{1+(t-1)k}, \quad (3.23)$$

los cuales no son más que los valores de secuencia espaciados de  $k$  valores.

Raftery y Lewis entonces asumen que la autocorrelación entre valores  $Z_t^{(k)}$  decrece a medida que  $k$  aumenta, y que para valores suficientemente grandes de  $k$ , la secuencia se comporta como una cadena de Markov. De esta manera, las variables aleatorias son ajustadas como el resultado de una cadena de Markov de primer orden (el próximo valor de secuencia depende apenas de un valor anterior), de una cadena de Markov de segundo orden (el valor que sigue está determinado por los dos últimos valores), y el modelo escogido utilizando algún método de selección de modelos.

El período de *burn-in* está determinado por el número de iteraciones necesarias para que se obtengan valores suficientemente próximos a la distribución estacionaria de la cadena de Markov construida. Tal número acostumbra ser pequeño, siendo utilizada frecuentemente en la literatura la regla práctica de descartar el primer uno por ciento de iteraciones como *burn-in*.

Otro criterio para verificar la convergencia de cadenas basado en muestras es el que desarrolla Geweke. Para esto se dividen las iteraciones en dos partes, la primera es el 10 por ciento o más ( $n_a$ ) y la última parte que es el 50 por ciento o más ( $n_b$ ), y se toman las diferencias de las medias para cada parte y se divide por el error estándar, encontrando de esta forma los puntajes  $Z$  de un test estadístico (Lee, 2012).

Se podría representar de la siguiente forma:

$$\bar{\eta}_a = \sum_{i=m}^{m+n_a} \eta_i \quad (3.24)$$

$$\bar{\eta}_b = \sum_{i=N^*-n_b}^{N^*+n_b} \eta_i$$

para los cuales se calcula el valor

$$z_G = \frac{\bar{\eta}_a - \bar{\eta}_b}{\sqrt{\hat{\sigma}_a - \hat{\sigma}_b}},$$

donde  $\hat{\sigma}_a, \hat{\sigma}_b$  son las varianzas muestrales calculadas para los  $n_a$  y  $n_b$  elementos de la muestra.

El valor de  $z_G$  tiende a una distribución normal, cuyos valores se utilizan para evaluar la significancia de  $z_G$ . Generalmente, se supone una convergencia de cadena para valores de  $z_G$  entre -1.96 y 1.96.

Andrew Gelman y Rubin (1992) usan múltiples secuencias de cadenas para estimar la varianza, llamada la escala potencial de reducción de factor (PSRF). Si este valor es alto, la convergencia de la cadena de Markov no es la adecuada y se necesitan más iteraciones. Si el valor de PSRF es cercano a 1, entonces la cadena de Markov es parecida a la distribución estacionaria (Fragoso, 2010). La varianza entre cadenas es calculada por

$$B = \frac{C}{N^* - 1} \sum_{c=1}^C (\bar{\eta}^{(c)} - \bar{\eta})^2, \quad (3.25)$$

donde  $\bar{\eta}^c$  es la medida muestral de los elementos de la cadena  $c$ , y  $\bar{\eta}$  es la media muestral de todos los valores muestreados.

También se calcula la varianza entre las muestras de cada cadena por

$$W = \frac{1}{C(N^* - 1)} \sum_{c=1}^C \sum_{N^*}^{t=1} (\eta_t^{(c)} - \bar{\eta}^{(c)})^2. \quad (3.26)$$

Entonces, se estima la varianza del parámetro  $\eta$  por

$$\hat{\sigma}_\eta = \left(1 - \frac{1}{N^*}\right) W + \frac{1}{N^*} B. \quad (3.27)$$

Si las cadenas convergen, las ecuaciones mencionadas anteriormente serán buenos estimadores para la varianza de  $\eta$ .

Gelman y Rubin plantean un valor de

$$\hat{R} = \sqrt{\frac{\hat{\sigma}_\eta}{W}}, \quad (3.28)$$

que son mayores que 1, pero que tiende a 1 en la medida que  $N^* \rightarrow \infty$ . Los autores sugieren valores menores a 1.2 para detectar convergencia.

### 3.3. Implementación computacional

Varios paquetes de R pueden usarse para estimar los parámetros en diversos modelos de TRI, sin embargo, un número considerable de ellos solo analizan modelos de TRI unidimensionales.

Uno de los paquetes más usados y creados en R para estimar parámetros de TRIM con un método estadístico clásico, tanto con modelos exploratorio y confirmatorios, usando el método de máxima verosimilitud, es el *mirt*, el cual analiza conjuntos de datos dicotómicos y/o politómicos usando variables latentes bajo el paradigma de la TRI y ajusta modelos logísticos unidimensionales y multidimensionales de uno o más parámetros usando el algoritmo tradicional de EM. Bock y Aitkin (1981) fueron los primeros en proponer un método factible de estimación para los parámetros del ítem usando un método similar al de EM

(Chalmers, 2012). Otro paquete que puede estimar de forma clásica los parámetros de TRIM pero hasta un máximo de dos dimensiones es el paquete *ltm*, el método de estimación que utiliza es el de MVM usando la regla de cuadratura de Gauss-Hermite (Rizopoulos, 2006).

Se exploraron y analizaron métodos de estimación clásica disponibles en el programa R con los paquetes mencionados anteriormente, sin embargo, no se mostraron consistentes con los resultados obtenidos en función a un software comercial usado y validado (IRTPRO). Por esta razón, se dió mayor énfasis a la inferencia bayesiana por ser más estable.

Se implementó, mediante inferencia bayesiana, el método de estimación conjunta MCMC mediante códigos BUGS para TRI (Curtis, 2010), ya que suele ser más preciso que estimaciones clásicas e incluso bayesianas en este tipo de modelos. Es probable que el software más utilizado para la aplicación de la inferencia Bayesiana sea el software WinBUGS (Bayesian inference Using Gibbs Sampling), realizado por Spiegelhalter, Thomas, Best, y Lunn (2003) y distribuido por la MRC Biostatistics Unit at Cambridge. Este paquete es una contribución importante y se basa en un conjunto de algoritmos computacionales generales que pueden ser utilizados para estimar los modelos específicos usando una sintaxis de un modelo definido. Este software usa el muestreo de Gibbs y el algoritmo de Metropolis Hasting para generar cadenas de Markov mediante un muestreo de distribuciones condicionales completas (Sturtz, Ligges y Gelman, 2005).

Existen otras implementaciones en el lenguaje BUGS como JAGS y OpenBUGS, sin embargo, hay otras interfaces que el software R incluye, entre ellas están el R2WinBUGS, rjags y BRugs (Curtis, 2010). Para el presente estudio se usará el paquete R2WinBUGS, el cual resume las inferencias y los diagnósticos de convergencia en tablas y gráficos, asimismo guarda las simulaciones en conjunto para facilitar el acceso en R.

Asimismo, el paquete *coda* realizado por Plummer, Best, Cowles y Vines (2004) es muy útil para el análisis de los resultados de WinBUGS, este paquete suele proporcionar funciones para el diagnóstico de convergencia, calcular estimaciones de Monte Carlo y realizar diversos gráficos relevantes para un mejor análisis (Sturtz, Ligges y Gelman, 2005).

Finalmente, cabe resaltar que para poder implementar el uso computacional a través del R y R2WinBUGS para las estimaciones de los parámetros del modelo de manera más efectiva y evitar problemas numéricos en el conjunto de datos usados en la aplicación (probabilidades encontradas pueden ser muy cercanas a 1 o muy cercanas a 0 y esto hace que los valores en la distribución tiendan a  $+\infty$  o  $-\infty$  respectivamente), se procedió a utilizar un procedimiento práctico (propuesto inicialmente por Ntzoufras (2009) para un caso de regresión en el caso probit) en el algoritmo del modelo planteado en la función logit para evitar probabilidades extremas. Se propuso truncar las colas entre  $(-\xi, \xi)$  del enlace logit usando la siguiente sintaxis en el WinBUGS:

---

```
logit(p[i,j])<- m[i,j]
m[i,j]<- ma[i,j]*(1-step(abs(ma[i,j])-xi))-xi*step(-xi-ma[i,j])+xi*step(ma[i,j]-xi)
y[i, j] ~ dbern( p[i, j] )
```

---

donde  $ma[i, j]$  es la ecuación definida por  $a1[j] * theta1[i] + a2[j] * theta2[i] - d[j]$  y  $\xi = 5$  es el valor de truncación.

## Capítulo 4

### Estudio de Simulación

En el presente capítulo, se presentará una simulación bajo un escenario planteado, en el cual se analizará la calidad de las estimaciones de las variables latentes desde un enfoque clásico y bayesiano.

#### 4.1. Condiciones de la simulación

Para realizar la siguiente simulación se tomó como base las estimaciones de los parámetros encontrados por [Fragoso \(2010\)](#), en su estudio sobre estudiantes universitarios brasileños y sus puntuaciones dicotómicas en el Inventario de Depresión de Beck. De esta forma, se consideran valores prefijados para los parámetros asociados a la discriminación y dificultad de los  $I = 21$  ítems en función a lo encontrado anteriormente por el autor y se indica que se trata de un constructo medido de forma bidimensional. Por otro lado, se simuló una muestra de  $n = 1000$  valores para el constructo latente bidimensional a partir de una distribución normal bivariada, con un vector de medias nulo y una matriz de varianzas igual a la identidad. El detalle del algoritmo utilizado se puede encontrar en el Apéndice A.

Por otro lado, se utilizó un modelo logístico multidimensional compensatorio de dos parámetros (M2PL) para ajustar las respuestas de los 1000 estudiantes a los 21 ítems. De esta forma, los parámetros del modelo planteado serán estimados con estas respuestas utilizando diferentes métodos para poder encontrar cual de estas técnicas se aproxima mejor en su estimación a los valores de los parámetros dados.

##### 4.1.1. Criterios para evaluar las estimaciones en la simulación

Existen varios criterios para medir el ajuste de los parámetros estimados donde  $\beta_l$  representará cualquier parámetro posible prefijado o simulado,  $\hat{\beta}_l$  su parámetro estimado y  $k$  el número de ítems, entre ellos se encuentran:

- El error cuadrático medio (MSE)

$$MSE = \sum_{l=1}^k \frac{(\beta_l - \hat{\beta}_l)^2}{k} \quad (4.1)$$

- La raíz cuadrada del error cuadrático medio (RMSE)

$$RMSE = \sqrt{\frac{\sum_{l=1}^k (\beta_l - \hat{\beta}_l)^2}{k}} \quad (4.2)$$

- La correlación entre valores simulados y estimados (CORR)

$$CORR = \frac{\sum_{l=1}^k (\beta_l - \bar{\beta}_l) (\hat{\beta}_l - \bar{\hat{\beta}}_l)}{\sqrt{\sum_{l=1}^k (\beta_l - \bar{\beta}_l)^2} \sqrt{\sum_{l=1}^k (\hat{\beta}_l - \bar{\hat{\beta}}_l)^2}} \quad (4.3)$$

## 4.2. Métodos de estimación a estudiar

Se procedió a estimar los parámetros de los ítems y de los rasgos latentes de las personas del modelo propuesto, para esto se utilizaron métodos de estimación clásicos y bayesianos.

Para la parte clásica se implementó computacionalmente la estimación con R mediante los paquetes *ltm* y *mirt*. Estos paquetes estiman los parámetros usando principalmente MVM y el algoritmo de EM. Asimismo, se usó un software comercial llamado IRTPRO, el cual suele utilizar para estimar los parámetros las técnicas de MVM y EAP. Para la parte bayesiana, como lo plantea Bazán (2014), se usó el software R y WinBUGS (R2WinBUGS) a fin de estimar los parámetros mediante MCMC. Los algoritmos utilizados en R para cada método de estimación se encuentran en el Apéndice A.

Con el estudio de simulación se busca analizar y precisar la calidad del funcionamiento de los métodos de estimación sujetos a un contexto similar al que se abordará en la aplicación más adelante.

## 4.3. Análisis de la calidad de las estimaciones

Se considera un escenario en el cual la muestra es de  $n = 1000$  estudiantes y la cantidad de ítems dicotómicos es de 21, asimismo, se sabe que la prueba tiene dos dimensiones.

Para este escenario planteado se analizaron 4 maneras diferentes de estimar los parámetros computacionalmente. Para los métodos de estimación clásica se utilizaron los paquetes *ltm* y *mirt* (que se encuentran en R) y el software comercial IRTPRO; para la estimación bayesiana se utilizó el programa WinBUGS y R.

Para efectos de la comparación entre las diferentes formas de estimar los parámetros de discriminación y dificultad de los ítems con sus valores (parámetros) prefijados se tomó respectivamente como indicadores los siguientes cálculos que representarían distancias:  $(a_1 - \hat{a}_1)^2 = da_1^2$ ,  $(a_2 - \hat{a}_2)^2 = da_2^2$  y  $(d - \hat{d})^2 = dd^2$

ITEM	ltm			mirt			IRTPRO			WinBUGS		
	$da_1^2$	$da_2^2$	$dd^2$									
1	0.03	1.38	0.00	0.04	0.01	0.01	0.00	0.03	0.01	0.02	0.01	0.00
2	0.14	1.56	0.04	0.12	0.00	0.06	0.00	0.01	0.05	0.02	0.00	0.04
3	0.06	4.23	0.01	0.13	0.00	0.00	0.06	0.32	0.00	0.00	0.18	0.01
4	0.99	1.89	0.01	0.41	0.08	0.00	0.00	0.11	0.00	0.06	0.09	0.01
5	0.00	8.40	0.16	1.44	0.66	0.20	0.86	0.07	0.20	0.97	0.01	0.11
6	0.02	0.96	0.01	0.04	0.00	0.00	0.02	0.10	0.00	0.00	0.07	0.01
7	0.89	4.01	0.01	0.38	0.04	0.02	0.40	0.27	0.02	0.10	0.14	0.01
8	0.07	1.89	0.00	0.02	0.12	0.00	0.01	0.01	0.00	0.04	0.01	0.00
9	0.69	1.45	0.07	0.13	0.22	0.10	0.01	0.41	0.09	0.14	0.32	0.06
10	0.26	0.56	0.00	0.36	0.10	0.00	0.04	0.09	0.00	0.00	0.08	0.00
11	0.35	0.68	0.00	0.27	0.01	0.00	0.02	0.00	0.00	0.01	0.01	0.00
12	0.49	1.01	0.02	0.29	0.02	0.01	0.01	0.01	0.01	0.02	0.01	0.02
13	0.18	1.11	0.00	0.57	0.01	0.00	0.11	0.01	0.00	0.01	0.01	0.00
14	0.03	0.28	0.00	0.27	0.01	0.00	0.08	0.01	0.00	0.02	0.01	0.00
15	0.30	1.86	0.00	0.36	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
16	0.59	0.48	0.01	0.42	0.02	0.01	0.03	0.00	0.01	0.01	0.00	0.00
17	0.81	0.83	0.01	0.58	0.05	0.00	0.03	0.00	0.00	0.01	0.01	0.01
18	0.10	0.36	0.00	0.17	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
19	0.16	0.02	0.00	0.32	0.10	0.01	0.05	0.01	0.00	0.02	0.02	0.00
20	0.04	0.52	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.01	0.00	0.00
21	0.25	0.11	0.00	0.39	0.11	0.00	0.06	0.04	0.00	0.00	0.05	0.00

Cuadro 4.1: Comparación de las distancias entre valores estimados y valores simulados considerando diferentes software para estimar un modelo logístico bidimensional

Como se puede observar en el cuadro anterior las estimaciones son similares, principalmente, entre el software IRTPRO y el WinBUGS, ya que la mayoría de sus distancias de los parámetros estimados son cercanas a cero. El paquete *mirt* también logra tener distancias cercanas a cero en los parámetros  $a_2$  y  $d$ , sin embargo, el número crece en la distancia de  $a_1$ . Por otro lado, el paquete *ltm* es el que peor estima ya que presenta valores más alejados del cero al comparar las estimaciones con los valores simulados.

Para poder resumir mejor esta información observada se presenta la siguiente tabla con medidas estadísticas que facilitan determinar la calidad de las estimaciones definidas en la sección 4.1.

Medida Estadística	ltm			mirt			IRTPRO			WinBUGS		
	$a_1$	$a_2$	d	$a_1$	$a_2$	d	$a_1$	$a_2$	d	$a_1$	$a_2$	d
MSE	0.31	1.60	0.02	0.32	0.08	0.02	0.09	0.07	0.02	0.07	0.05	0.01
RMSE	0.55	1.26	0.13	0.57	0.27	0.14	0.30	0.27	0.14	0.26	0.22	0.12
CORR	0.81	0.59	0.99	0.93	0.90	0.99	0.95	0.85	0.99	0.96	0.91	0.99

Cuadro 4.2: Medidas de ajustes de la calidad de las estimaciones de los parámetros de los ítems con diferentes software

En el cuadro anterior, se muestra de forma más evidente lo señalado anteriormente sobre la calidad de las estimaciones. Se sabe que mientras las medidas MSE y RMSE sean más cercanas a cero y el coeficiente CORR sea más cercano a uno, la estimación realizada será más optima. Por esta razón, se puede evidenciar que el WinBUGS es el que mejor estima, seguido del software IRTPRO. Los paquetes *mirt* y *ltm* de R no logran estimar adecuadamente según lo observado en las medidas estadísticas halladas.

Dado que el WinBUGS es el que mejor estima los parámetros de los ítems, se procedió a estimar los rasgos latentes de cada persona y calcular las respectivas medidas de ajuste, obteniéndose un MSE de 0.00 y un RMSE de 0.01. De esta forma, se comprueba que con el uso del R2WinBUGS las estimaciones de los parámetros de los ítems y de las personas son de buena calidad.

#### 4.3.1. Conclusión del estudio de simulación

Luego del presente estudio de simulación realizado, se observa que el método de estimación con mejores indicadores de ajustes y, por lo tanto, más robusto para el modelo planteado (M2PL), es el método bayesiano mediante MCMC (implementado computacionalmente con el WinBUGS). Éste método muestra mayor precisión que los otros métodos clásicos en las estimaciones de los parámetros de los ítems y una muy buena estimación de los rasgos latentes de las personas. Cabe resaltar que las estimaciones realizadas con el software comercial IRTPRO también son adecuadas y muy similares a las realizadas mediante el WinBUGS, pero tiene la desventaja de que no es un software de libre uso como el del caso bayesiano.

## Capítulo 5

# Aplicación

En éste capítulo, se presenta la aplicación del modelo propuesto (M2PL) a un conjunto de datos reales orientados en el ámbito de la psicología, específicamente se trata de una prueba que mide actitudes hacia la Estadística en una muestra de estudiantes de una universidad privada de Colombia (Pérez et al., 2015). Se realizará un análisis previo clásico de la prueba de forma breve y, posteriormente, un análisis mediante el enfoque de TRIM. La estimación de los parámetros de los ítems (“dificultad” y discriminación) y de los rasgos latentes de las personas se realizarán mediante inferencia bayesiana con el método de MCMC.

### 5.1. Instrumento de medición

La Escala de Actitudes hacia la Estadística de Cazorla et al - AEC (Cazorla, Silva, Vendramini y Brito, 1999), fue adaptada a partir de una escala de actitudes en relación a las matemáticas creada por Aiken (1974). Es una escala de tipo Likert, compuesta por 20 ítems, 10 afirmativos y 10 negativos. Inicialmente fue considerada unidimensional, pero luego de diversos estudios psicométricos, como se muestra en Campos, Bonafe, Dovigo y Maroco (2013), Estrada (2011) y Campos, Bonafe, Dovigo y Maroco (2010), se encontró que dos dimensiones representan mejor el constructo.

Este instrumento se aplicó en el año 2013 a una muestra de 545 universitarios colombianos entre 17 y 25 años de edad, que cursaban el primer o segundo semestre académico de una universidad privada de la ciudad de Bogotá, en las siguientes facultades: Internacional de Administración y Marketing (EIAM), Ciencias Exactas e Ingeniería, y Economía (todas ellas cuentan en su plan de estudios con la disciplina de Estadística en el tercer o cuarto semestre). Cabe resaltar que los estudiantes de la muestra no habían llevado ni estaban llevando el curso de Estadística a nivel universitario (Pérez et al., 2015).

La prueba aplicada presentó una escala Likert de respuesta de 1 (totalmente en desacuerdo) hasta 5 (totalmente de acuerdo); sin embargo, para poder aplicar los métodos descritos en el presente trabajo y analizar las respuestas de los participantes con el modelo propuesto, estas respuestas se cambiaron a un formato binario o dicotómico. El valor 0 reemplazó las respuestas 1, 2 y 3, lo cual indica que el participante (estudiante) está en desacuerdo sobre ese elemento; asimismo, el valor 1 reemplazó las respuestas 4 y 5, por lo tanto esta puntuación representa un máximo acuerdo del estudiante por el ítem propuesto. Cabe destacar que 10 ítems son planteados de manera negativa por lo que se invirtió la puntuación para ellos. De esta forma, mientras más respuestas iguales a 1 en la respuesta de un universitario, se

interpretaría como que este tiene mejores actitudes hacia la Estadística.

Los 20 ítems de la prueba, según [Cazorla et al. \(1999\)](#), son:

Ítem	Enunciado
1	Yo quedo terriblemente tenso(a) en la clase de Estadística
2	Yo no gusto de Estadística y me asusta tener que hacer el curso de Estadística
3	Yo creo que la Estadística es muy interesante y gusto de las clases de Estadística
4	La Estadística es fascinante y divertida
5	La Estadística me hace sentir seguro(a) y es al mismo tiempo estimulante
6	Cuando estudio Estadística mi cabeza “queda en blanco” y no consigo pensar claramente
7	Yo tengo una sensación de inseguridad cuando me esfuerzo en Estadística
8	La Estadística me deja inquieto(a), descontento, irritado(a) e impaciente
9	El sentimiento que yo tengo con relación a la Estadística es bueno
10	La Estadística me hace sentir como si estuviese perdido(a) en una selva de números y sin encontrar la salida
11	La Estadística es algo que yo aprecio grandemente
12	Cuando yo escucho la palabra Estadística, yo tengo un sentimiento de aversión (rechazo)
13	Yo encaro la Estadística con un sentimiento de indecisión, que es resultado del miedo de no ser capaz en Estadística
14	Yo gusto realmente de la Estadística
15	La Estadística es una de las materias que yo realmente gusto de estudiar en la universidad
16	Pensar sobre la obligación de resolver un problema de Estadística me deja nervioso(a)
17	Yo nunca guste de la Estadística y es la materia que más me da miedo
18	Yo quedo más feliz en la clase de Estadística que en la clase de cualquier otra materia
19	Yo me siento tranquilo(a) en Estadística y gusto mucho de esa materia
20	Yo tengo una reacción definitivamente positiva con relación a la Estadística: yo gusto y aprecio esa materia

Cuadro 5.1: Ítems de actitudes hacia la Estadística

## 5.2. Análisis clásico de la prueba

En la teoría de las variables latentes, la estimación del puntaje total o score se realiza a partir de información de las variables observadas o manifiestas. La prueba o test es un instrumento de medición cuantitativo que se elabora siguiendo ciertos procesos de definición, validación y consistencia. De esta manera se busca conseguir una prueba que recoja información precisa para poder estimar un puntaje total adecuado para cada participante ([Flores, 2012](#)).

La validación es un proceso en el cual se verifica que la medida obtenida se acerca a la definición planteada. Asimismo, la confiabilidad se puede definir como la estabilidad de los resultados y la calidad de la representación de los puntajes observados al puntaje verdadero.

Este indicador estadístico, de consistencia interna, puede ser medido a través del cálculo del alfa de Cronbach.

Desde la TCT, el análisis de los ítems tiene como primer alcance realizar una descripción de las características estadísticas de los parámetros de un ítem de la prueba, sin embargo, el ítem individual solo es de interés a través del efecto que tiene sobre el puntaje total de la prueba. Por esta razón, como en este análisis se asume multidimensionalidad del constructo dado, este análisis clásico pierde precisión.

- Confiabilidad interna de los ítems

Esta medida estadística permite dar información sobre el grado de estabilidad, precisión o consistencia que presentan las puntuaciones de la prueba de determinado constructo. Se puede calcular la confiabilidad interna de los datos mediante el alfa de Cronbach con la siguiente fórmula:

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

donde  $k$  representa el número de ítems,  $\sigma_i^2$  la varianza de los puntajes para el ítem  $i$  y  $\sigma_x^2$  la varianza de los puntajes totales.

Al calcular el alfa de Cronbach para la prueba de ítems dicotómicos se obtiene el coeficiente de Kuder-Richardson (KR), con un valor = 0.89, lo cual verifica que la prueba aplicada a los estudiantes de Colombia tiene una consistencia interna adecuada y por lo tanto sus resultados son confiables. Sin embargo, no se reporta la correlación ítem-total pues no se conocen las dimensiones evaluadas, lo cual debería ser abordado luego de estudiar la dimensionalidad. Por esta razón, hacer un análisis de correlación ítem-total con una dimensión única que incluya a todos los ítems podría llevar a otras interpretaciones.

### 5.3. Análisis de la dimensionalidad del constructo actitudes hacia la Estadística

La prueba de actitudes hacia la Estadística de Cazorla et al., originalmente se planteó como una prueba unidimensional (Cazorla et al., 1999); sin embargo, a lo largo de los años se ha ido estudiando mejor la psicometría de las puntuaciones de la prueba, como se puede observar en Turik (2010), Campos et al. (2010), Estrada (2011) y Campos et al. (2013), donde se concluye que el constructo medido por el instrumento usado se ajusta mejor desde una perspectiva bidimensional .

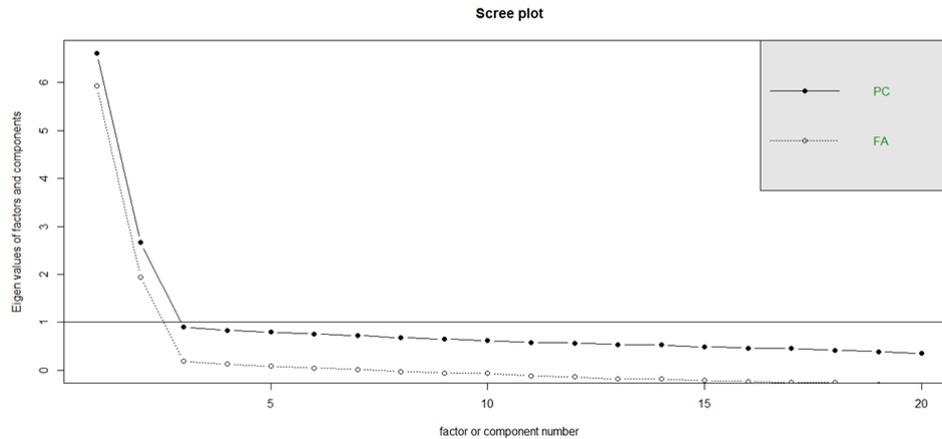
Análisis psicométricos previos mediante TCT de la prueba en la muestra usada (sin dicotomizar las variables de respuesta) determinaron mediante análisis factorial que esta es bidimensional (Aparicio, 2015). Por esta razón, se podría asumir la multidimensionalidad (dos dimensiones) del constructo a medir.

Para verificar que se cumple la multidimensionalidad con la base dicotómica para el presente estudio, se calculó un coeficiente de Kaiser - Meyer - Olkin (KMO) = 0.93 y en

la prueba de Esfericidad de Barlett se encontró un valor de  $\chi^2 = 3796$  ( $p < 0.01$ ); estos resultados indican que sería pertinente hacer un análisis factorial exploratorio.

Para analizar la cantidad de dimensiones presentes en esta base de ítems dicotómicos se procedió a usar un gráfico de sedimentación, ver Figura 5.1, el cual sugiere el uso de 2 factores para medir el constructo:

Figura 5.1: Gráfico de sedimentación



Luego de esto, se procedió a realizar un análisis factorial exploratorio con la siguiente matriz de correlaciones tetracóricas, ver Cuadro 5.2:

Ítem	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1																			
2	.63	1																		
3	.41	.63	1																	
4	.37	.40	.68	1																
5	.22	.23	.58	.72	1															
6	.62	.62	.45	.26	.18	1														
7	.50	.53	.38	.29	.10	.63	1													
8	.53	.57	.47	.22	.23	.65	.69	1												
9	.37	.53	.62	.54	.45	.43	.25	.39	1											
10	.53	.51	.46	.31	.21	.66	.61	.62	.46	1										
11	.34	.33	.59	.69	.64	.32	.22	.22	.47	.21	1									
12	.42	.53	.51	.39	.30	.58	.49	.56	.46	.49	.38	1								
13	.56	.52	.46	.33	.28	.62	.63	.56	.34	.59	.31	.61	1							
14	.33	.36	.68	.72	.62	.29	.33	.33	.67	.45	.67	.50	.40	1						
15	.39	.36	.65	.64	.57	.34	.27	.25	.49	.34	.66	.42	.38	.79	1					
16	.54	.55	.37	.30	.17	.52	.57	.53	.35	.44	.27	.49	.50	.30	.27	1				
17	.56	.58	.46	.43	.25	.57	.52	.61	.43	.57	.32	.62	.56	.45	.34	.63	1			
18	.15	.13	.42	.57	.63	.00	.05	.11	.32	.01	.59	.16	.00	.51	.53	.10	.00	1		
19	.43	.42	.65	.67	.62	.33	.27	.32	.55	.45	.61	.44	.41	.75	.69	.42	.43	.58	1	
20	.36	.38	.66	.69	.67	.27	.30	.31	.57	.39	.69	.43	.38	.73	.65	.34	.46	.53	.80	1

Cuadro 5.2: Matriz de correlaciones tetracóricas

Esta matriz de correlaciones muestra las asociaciones entre todos los ítems y es el principal insumo para proceder con el análisis factorial, el cual se realizó asumiendo la bidimensionalidad del constructo y con una rotación oblimin dada la asociación teórica que existe entre ambas dimensiones del constructo. Se encontró lo siguiente:

	MR1	MR2
Ítem 1	0.10	<b>0.67</b>
Ítem 2	0.11	<b>0.70</b>
Ítem 3	<b>0.63</b>	0.31
Ítem 4	<b>0.82</b>	0.03
Ítem 5	<b>0.84</b>	-0.12
Ítem 6	-0.07	<b>0.85</b>
Ítem 7	-0.09	<b>0.81</b>
Ítem 8	-0.06	<b>0.82</b>
Ítem 9	<b>0.52</b>	0.29
Ítem 10	0.04	<b>0.75</b>
Ítem 11	<b>0.79</b>	-0.01
Ítem 12	0.20	<b>0.67</b>
Ítem 13	0.06	<b>0.73</b>
Ítem 14	<b>0.81</b>	0.11
Ítem 15	<b>0.77</b>	0.07
Ítem 16	0.04	<b>0.67</b>
Ítem 17	0.10	<b>0.72</b>
Ítem 18	<b>0.82</b>	-0.31
Ítem 19	<b>0.79</b>	0.13
Ítem 20	<b>0.81</b>	0.09

Cuadro 5.3: Matriz de componentes rotados

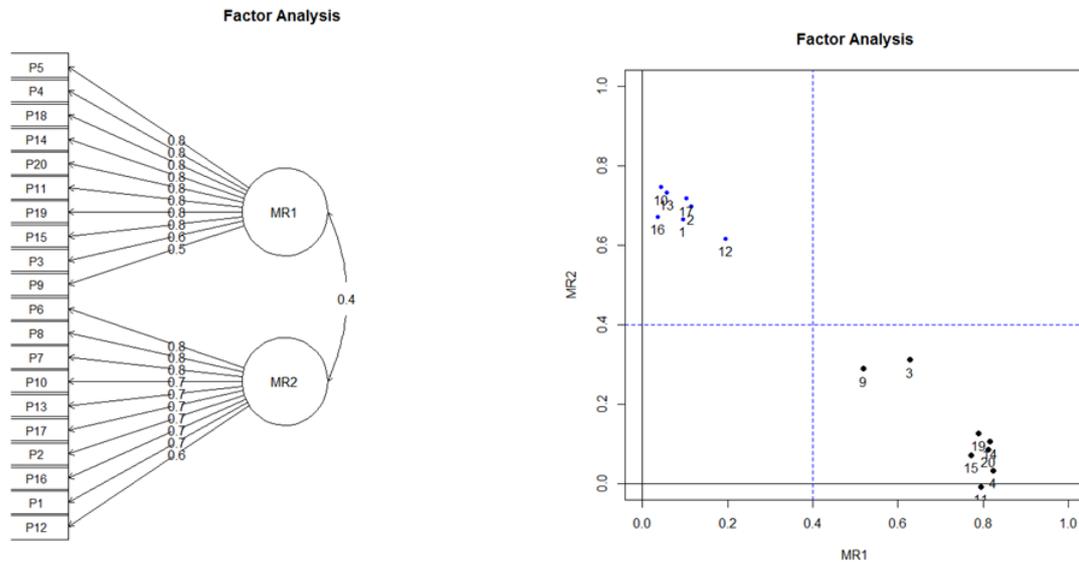
En este cuadro se logra distinguir la presencia de cada ítem en cada dimensión en función a las cargas factoriales calculadas.

	MR1	MR2
Varianza	0.31	0.30
Varianza acumulada	0.31	0.61
Proporción explicada	0.51	0.49

Cuadro 5.4: Varianza explicada por los factores hallados

Se muestra que el porcentaje de varianza acumulada explicada por ambos factores es de más de 60%, lo cual es bastante bueno en el área de la psicología; además cada dimensión por separado explica un porcentaje muy similar de la varianza lo cual reforzaría la premisa de la bidimensionalidad de la prueba. Los gráficos a continuación, representan la relación y segmentación de ítems en función a los factores propuestos en los análisis anteriores.

Figura 5.2: Distribución de ítems según los factores encontrados



#### 5.4. Estimación de parámetros mediante el modelo estudiado

Para seleccionar como se ajusta el modelo planteado a los datos observados se utilizó el modelo logístico multidimensional compensatorio de dos parámetros (M2PL), suponiendo dos dimensiones para el vector de rasgos latentes. Para este modelo compensatorio se ajustarán los datos utilizando el Método de estimación Bayesiano MCMC descrito en el capítulo 3. Asimismo, son utilizados para este modelo las distribuciones a priori especificadas anteriormente y recomendadas por Patz y Junker (1999): para los parámetros asociados a la discriminación se usará  $a_{ik} \sim \text{Lognormal}(1, 2)$ , para el escalar asociado a la dificultad, se representará mediante  $d_i \sim N(0, 1)$  y para los rasgos latentes de las personas se usará  $\theta_j \sim N(0, I_p)$ .

Para la implementación del modelo elegido se utilizó el software R y R2WinBUGS. El algoritmo de este modelo es presentado en el Apéndice A. Siguiendo el análisis de [Fragoso y Curi \(2013\)](#), se utilizaron 105 000 iteraciones, tomando un “burn in” de 5000 e intervalos de 50 iteraciones entre los valores utilizados para minimizar la autocorrelación. Adicionalmente, en el presente estudio, se hizo uso de una sola cadena.

Asimismo, se consideró un análisis de convergencia usando el criterio de Geweke, el cual fue expuesto en el capítulo 3. Los resultados, descritos en el Apéndice B, indican que las estimaciones convergieron adecuadamente.

##### 5.4.1. Estimación de los parámetros usuales del M2PL de TRIM

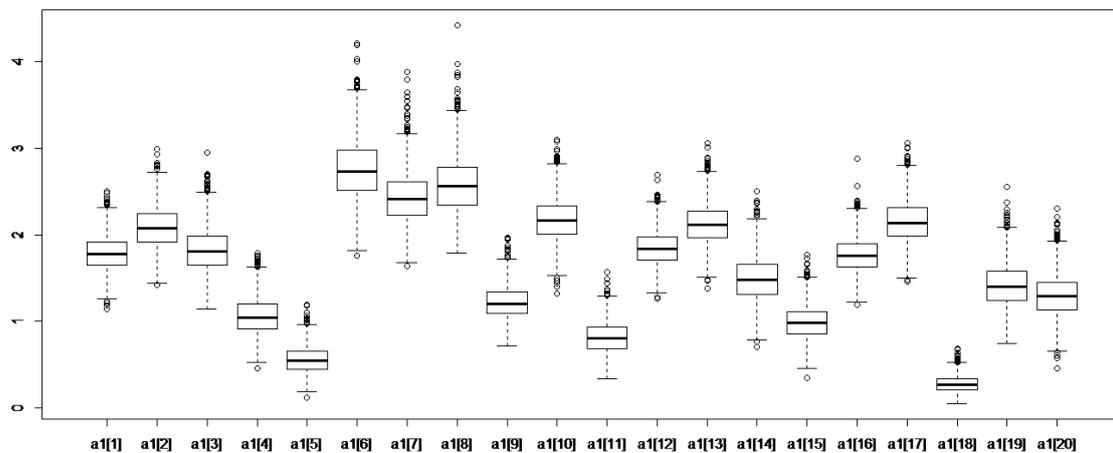
Las estimaciones de los componentes del vector de discriminación  $\mathbf{a}$  del ítem ( $a_1$  y  $a_2$ ) pueden ser interpretadas en términos de la capacidad del ítem de evaluar específicamente el rasgo latente. Estos elementos del vector están relacionados con la pendiente de la superficie de respuesta al ítem en la dirección de la correspondiente dimensión del rasgo latente ([Reckase, 1996](#)).

A continuación, se presentan las medidas resumen (Cuadro 5.5) y gráficos de cajas (Figura 5.3) del primer elemento ( $a_1$ ) del vector de discriminación, el cual está relacionado con la primera dimensión del rasgo latente estudiado:

	Media	D.E.	$P_{2.5}$	$P_{50}$	$P_{97.5}$
Ítem 1	1.79	0.20	1.41	1.78	2.23
Ítem 2	2.09	0.23	1.68	2.07	2.57
Ítem 3	1.82	0.26	1.35	1.81	2.38
Ítem 4	1.06	0.22	0.68	1.05	1.52
Ítem 5	0.55	0.15	0.27	0.55	0.88
Ítem 6	2.76	0.34	2.17	2.73	3.49
Ítem 7	2.43	0.29	1.90	2.41	3.05
Ítem 8	2.58	0.32	2.02	2.56	3.25
Ítem 9	1.22	0.19	0.88	1.20	1.62
Ítem 10	2.17	0.25	1.69	2.16	2.70
Ítem 11	0.81	0.18	0.48	0.80	1.19
Ítem 12	1.84	0.21	1.46	1.83	2.27
Ítem 13	2.13	0.24	1.69	2.12	2.65
Ítem 14	1.49	0.26	1.01	1.48	2.05
Ítem 15	0.99	0.20	0.61	0.98	1.41
Ítem 16	1.77	0.20	1.39	1.75	2.18
Ítem 17	2.15	0.25	1.71	2.14	2.65
Ítem 18	0.28	0.09	0.13	0.27	0.49
Ítem 19	1.42	0.25	0.97	1.40	1.97
Ítem 20	1.30	0.25	0.86	1.29	1.82

Cuadro 5.5: Medidas de resumen de la distribución a posteriori de  $a_1$  en el modelo M2PL utilizando el método MCMC

Figura 5.3: Diagrama de cajas de las distribuciones a posteriori de  $a_1$  en el modelo M2PL utilizando el método MCMC para los ítems de la escala de actitudes hacia la Estadística



Se observa que el ítem 6 (cuando estudio Estadística mi cabeza “queda en blanco” y no consigo pensar claramente) es el que tiene la mayor pendiente en la primera dimensión,

esto quiere decir que este ítem muestra más sensibilidad para encontrar diferencias en esta dimensión del rasgo latente ( $\theta_1$ )

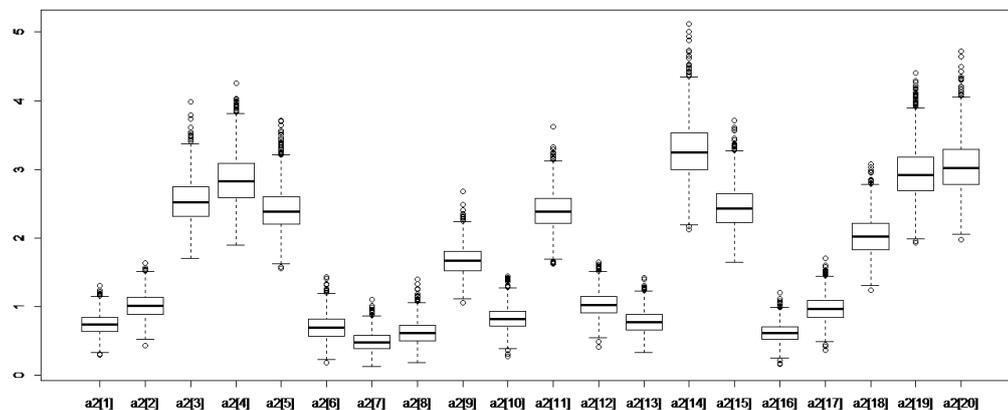
Por el contrario, el ítem 18 (yo quedo más feliz en la clase de Estadística que en la clase de cualquier otra materia) es el que tiene menor pendiente en esta dimensión, por lo tanto, será el ítem que muestre menor sensibilidad para encontrar diferencias en  $\theta_1$ .

A continuación, se presentan las medidas resumen (Cuadro 5.6) y gráficos de cajas (Figura 5.4) del segundo elemento ( $a_2$ ) del vector de discriminación, el cual está relacionado con la segunda dimensión del rasgo latente estudiado:

	Media	D.E.	$P_{2.5}$	$P_{50}$	$P_{97.5}$
Ítem 1	0.74	0.15	0.46	0.73	1.04
Ítem 2	1.02	0.18	0.68	1.01	1.39
Ítem 3	2.53	0.31	1.97	2.52	3.18
Ítem 4	2.84	0.37	2.20	2.82	3.64
Ítem 5	2.41	0.31	1.86	2.39	3.08
Ítem 6	0.69	0.18	0.37	0.69	1.07
Ítem 7	0.49	0.14	0.24	0.47	0.79
Ítem 8	0.62	0.17	0.30	0.61	0.98
Ítem 9	1.67	0.21	1.28	1.66	2.12
Ítem 10	0.82	0.17	0.50	0.82	1.19
Ítem 11	2.40	0.28	1.88	2.39	3.00
Ítem 12	1.03	0.18	0.70	1.02	1.39
Ítem 13	0.78	0.17	0.48	0.77	1.12
Ítem 14	3.28	0.41	2.56	3.25	4.15
Ítem 15	2.45	0.31	1.90	2.43	3.10
Ítem 16	0.61	0.14	0.36	0.61	0.90
Ítem 17	0.97	0.18	0.64	0.97	1.34
Ítem 18	2.03	0.29	1.51	2.02	2.67
Ítem 19	2.95	0.37	2.30	2.92	3.75
Ítem 20	3.05	0.38	2.38	3.02	3.87

Cuadro 5.6: Medidas de resumen de la distribución a posteriori de  $a_2$  en el modelo M2PL utilizando el método MCMC

Figura 5.4: Diagrama de cajas de las distribuciones a posteriori de  $a_2$  en el modelo M2PL utilizando el método MCMC para los ítems de la escala de actitudes hacia la Estadística



Se observa que el ítem 14 (yo gusto realmente de la Estadística) es el que tiene la mayor pendiente en la segunda dimensión, esto quiere decir que este ítem muestra más sensibilidad para encontrar diferencias en esta dimensión del rasgo latente ( $\theta_2$ )

Por el contrario, el ítem 7 (yo tengo una sensación de inseguridad cuando me esfuerzo en Estadística) es el que tiene menor pendiente en esta dimensión, por lo tanto, será el ítem que muestre menor sensibilidad para encontrar diferencias en  $\theta_2$ .

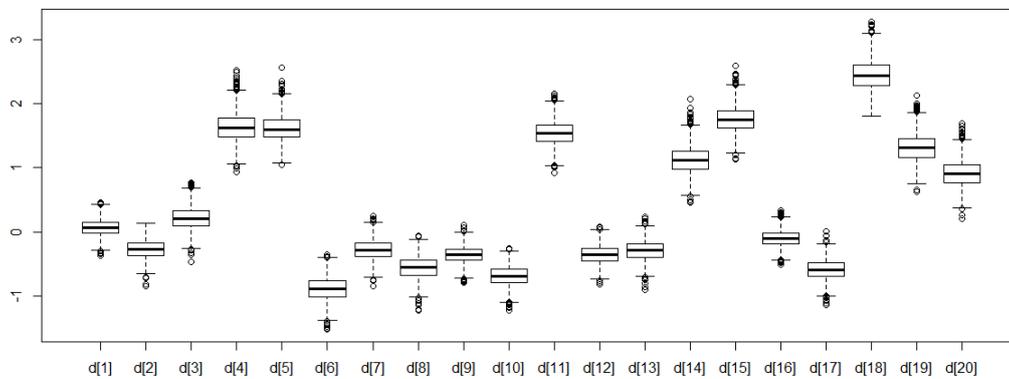
Por otro lado, se tiene al parámetro  $d$  del modelo planteado, el cual está relacionado con la “dificultad” de los ítems de la prueba. Es importante recordar que el valor de este parámetro no se puede interpretar directamente de la misma forma como se realiza en los modelos de TRIU ya que en la forma planteada del modelo multidimensional usado,  $d$  sería el intercepto en la ecuación del exponente que explica parte del modelo.

A continuación, se presentan las medidas resumen (Cuadro 5.7) y gráficos de cajas (Figura 5.5) del intercepto ( $d$ ):

	Media	D.E.	$P_{2.5}$	$P_{50}$	$P_{97.5}$
Ítem 1	0.07	0.13	-0.19	0.06	0.31
Ítem 2	-0.27	0.15	-0.55	-0.27	0.01
Ítem 3	0.21	0.18	-0.13	0.21	0.55
Ítem 4	1.64	0.22	1.23	1.62	2.12
Ítem 5	1.62	0.20	1.26	1.60	2.02
Ítem 6	-0.89	0.18	-1.27	-0.88	-0.54
Ítem 7	-0.28	0.16	-0.59	-0.28	0.03
Ítem 8	-0.56	0.17	-0.90	-0.56	-0.24
Ítem 9	-0.36	0.14	-0.63	-0.35	-0.09
Ítem 10	-0.69	0.15	-0.99	-0.69	-0.40
Ítem 11	1.55	0.19	1.19	1.55	1.93
Ítem 12	-0.35	0.14	-0.63	-0.35	-0.09
Ítem 13	-0.30	0.15	-0.59	-0.29	-0.03
Ítem 14	1.13	0.21	0.74	1.12	1.57
Ítem 15	1.76	0.21	1.38	1.76	2.20
Ítem 16	-0.10	0.13	-0.35	-0.10	0.14
Ítem 17	-0.59	0.15	-0.91	-0.59	-0.30
Ítem 18	2.46	0.24	2.04	2.44	2.97
Ítem 19	1.32	0.21	0.93	1.31	1.75
Ítem 20	0.91	0.20	0.54	0.90	1.30

Cuadro 5.7: Medidas de resumen de la distribución a posteriori de  $d$  en el modelo M2PL utilizando el método MCMC

Figura 5.5: Diagrama de cajas de las distribuciones a posteriori de  $d$  en el modelo M2PL utilizando el método MCMC para los ítems de la escala de actitudes hacia la Estadística



Se observa que el ítem 18 (yo quedo más feliz en la clase de Estadística que en la clase de cualquier otra materia) presenta valores más altos. Por el contrario, el ítem 6 (cuando estudio Estadística mi cabeza “queda en blanco” y no consigo pensar claramente) presenta los valores más bajos. Como se mencionó, este parámetro (escalar) está relacionado con la “dificultad” de los ítems pero no se puede interpretar directamente.

#### 5.4.2. Análisis del poder discriminativo (MDISC) y la dificultad (MDIFF) del ítem del constructo actitudes hacia la Estadística

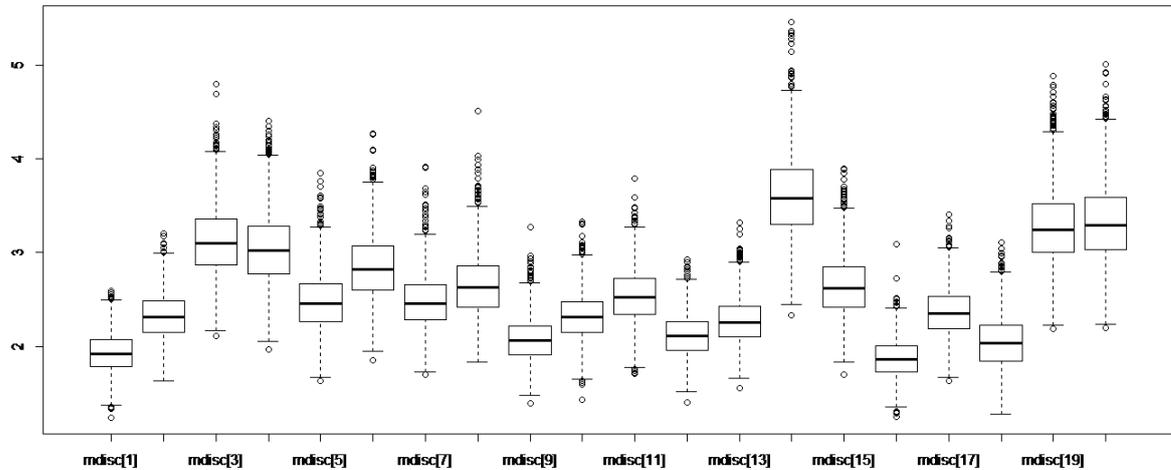
En las estimaciones de los parámetros de discriminación mencionados anteriormente ( $a_1$  y  $a_2$ ) solo se tienen medidas para una dimensión en particular, sin embargo, si se combinan las dimensiones se tendrá un índice que discrimine de una manera más global. Ésta es la llamada discriminación multidimensional o poder discriminativo (MDISC).

A continuación, se presentan las medidas resumen (Cuadro 5.8) y diagramas de cajas (Figura 5.6) de MDISC:

	Media	D.E.	$P_{2.5}$	$P_{50}$	$P_{97.5}$
Ítem 1	1.94	0.21	1.56	1.93	2.39
Ítem 2	2.33	0.25	1.87	2.31	2.83
Ítem 3	3.13	0.36	2.51	3.10	3.89
Ítem 4	3.04	0.38	2.37	3.02	3.88
Ítem 5	2.48	0.31	1.94	2.46	3.14
Ítem 6	2.85	0.34	2.26	2.83	3.60
Ítem 7	2.48	0.29	1.96	2.46	3.09
Ítem 8	2.66	0.32	2.08	2.64	3.35
Ítem 9	2.07	0.24	1.64	2.06	2.59
Ítem 10	2.33	0.26	1.84	2.32	2.89
Ítem 11	2.54	0.29	2.00	2.53	3.15
Ítem 12	2.11	0.22	1.70	2.11	2.58
Ítem 13	2.27	0.24	1.81	2.26	2.80
Ítem 14	3.61	0.43	2.87	3.58	4.51
Ítem 15	2.65	0.32	2.08	2.62	3.34
Ítem 16	1.88	0.20	1.49	1.87	2.30
Ítem 17	2.37	0.26	1.90	2.36	2.90
Ítem 18	2.05	0.29	1.53	2.04	2.69
Ítem 19	3.28	0.40	2.59	3.25	4.15
Ítem 20	3.32	0.41	2.61	3.29	4.21

Cuadro 5.8: Medidas de resumen de MDISC en el modelo M2PL utilizando el método MCMC

Figura 5.6: Diagrama de cajas de las distribuciones de MDISC en el modelo M2PL utilizando el método MCMC para los ítems de la escala de actitudes hacia la Estadística



Se observa que el ítem 14 (yo gusto realmente de la Estadística) es el que tiene la mayor medida de tendencia central, esto quiere decir que este ítem muestra más sensibilidad para encontrar diferencias en todo el rasgo latente estudiado.

Por el contrario, el ítem 16 (pensar sobre la obligación de resolver un problema de Estadística me deja nervioso(a)) es el que tiene la menor medida, por lo tanto, será el ítem que muestre menor sensibilidad para encontrar diferencias en el constructo estudiado.

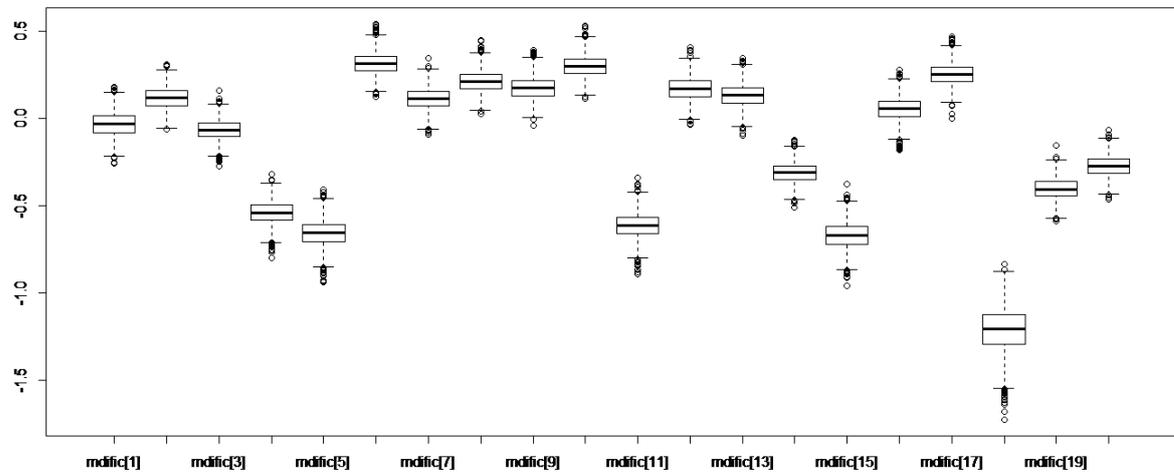
Anteriormente, se presentaba el parámetro estimado  $d$  del modelo planteado, el cual no se podía interpretar directamente. Por esta razón, se recurre al índice de dificultad (MDIFF) que se puede interpretar de una manera equivalente al parámetro de dificultad estimado de un modelo de TRIU. Mientras menor sea su valor para un ítem, éste indicará que es más importante para tener mejores actitudes hacia la estadística.

A continuación, se presentan las medidas resumen (Cuadro 5.9) y gráficos de cajas (Figura 5.7) de MDIFF:

	Media	D.E.	$P_{2.5}$	$P_{50}$	$P_{97.5}$
Ítem 1	-0.03	0.07	-0.16	-0.03	0.09
Ítem 2	0.12	0.06	0.00	0.11	0.23
Ítem 3	-0.07	0.06	-0.18	-0.07	0.05
Ítem 4	-0.54	0.06	-0.67	-0.54	-0.42
Ítem 5	-0.66	0.08	-0.82	-0.65	-0.52
Ítem 6	0.31	0.06	0.19	0.31	0.44
Ítem 7	0.11	0.06	-0.01	0.11	0.24
Ítem 8	0.21	0.06	0.09	0.21	0.34
Ítem 9	0.17	0.07	0.05	0.17	0.31
Ítem 10	0.30	0.06	0.17	0.30	0.42
Ítem 11	-0.61	0.07	-0.76	-0.61	-0.47
Ítem 12	0.17	0.07	0.04	0.17	0.30
Ítem 13	0.13	0.06	0.01	0.13	0.26
Ítem 14	-0.31	0.06	-0.43	-0.31	-0.20
Ítem 15	-0.67	0.07	-0.82	-0.67	-0.53
Ítem 16	0.05	0.07	-0.08	0.06	0.18
Ítem 17	0.25	0.06	0.13	0.25	0.37
Ítem 18	-1.21	0.12	-1.47	-1.21	-0.99
Ítem 19	-0.40	0.06	-0.52	-0.41	-0.28
Ítem 20	-0.28	0.06	-0.39	-0.27	-0.16

Cuadro 5.9: Medidas de resumen de MDIFF en el modelo M2PL utilizando el método MCMC

Figura 5.7: Diagrama de cajas de las distribuciones de MDIFF en el modelo M2PL utilizando el método MCMC para los ítems de la escala de actitudes hacia la Estadística

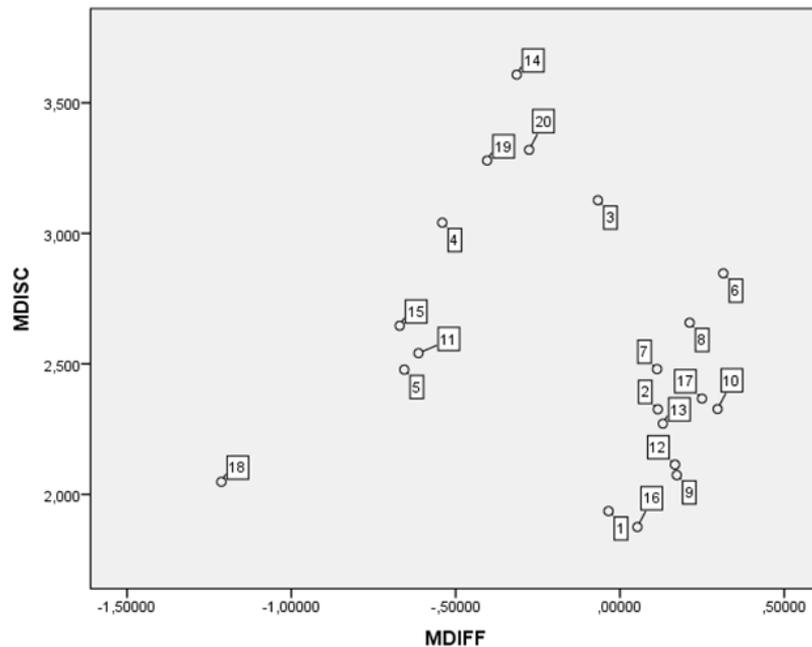


Los ítems 6 (cuando estudio Estadística mi cabeza “queda en blanco” y no consigo pensar claramente) y 10 (la Estadística me hace sentir como si estuviese perdido(a) en una selva de números y sin encontrar la salida) presentan las medias más positivas, por lo tanto se puede afirmar que estos ítems son los menos valorados para tener una mejor actitud hacia la estadística.

Por el contrario, el ítem 18 (yo quedo más feliz en la clase de Estadística que en la clase de cualquier otra materia) presenta la media más baja. Esto representaría que es el ítem más valorado para tener una mejor actitud hacia la estadística.

Si se relacionan ambas medidas en un gráfico de dispersión (Figura 5.8) se puede evidenciar mejor lo analizado anteriormente:

Figura 5.8: Diagrama de dispersión de los ítems en función a MDISC y MDIFF



Lo más resaltante sería que el ítem 18 se encuentra más alejado del resto de ítems y con puntajes muy bajos en MDIFF y MDISC, también se observa que el ítem 14 es el que discrimina mejor de los ítems y presenta un MDIFF ligeramente por debajo del promedio. Asimismo, llama la atención que el grupo de ítems con puntajes positivos en MDIFF tienen valores muy similares y un MDISC entre 2 y 2.5 en su mayoría.

#### 5.4.3. Importancia de los ítems en las dimensiones del constructo actitudes hacia la Estadística

Para los parámetros de discriminación del ítem por dimensión, la mejor forma de interpretarlos es incluyendo al poder discriminativo MDISC, de esta manera se logra calcular un número entre los valores 0 y 1, los cuales son análogos a las cargas de un análisis factorial (Fragoso y Curi, 2013). De manera similar, se usará un puntaje de corte (0.6) para clasificar los ítems a determinada dimensión.

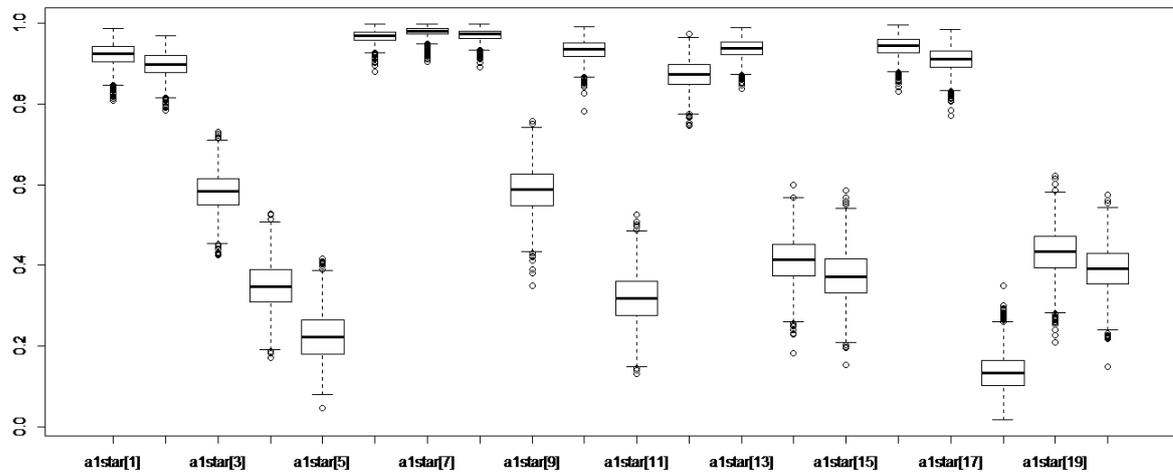
Estos valores calculados se denominarán  $a_1^*$  para la primera dimensión y  $a_2^*$  para la segunda dimensión del constructo estudiado.

A continuación, se presentan las medidas resumen (Cuadro 5.10) y gráficos de cajas (Figura 5.9) para  $a_1^*$ :

	Media	D.E.	$P_{2.5}$	$P_{50}$	$P_{97.5}$
Ítem 1	0.92	0.03	0.86	0.92	0.97
Ítem 2	0.90	0.03	0.83	0.90	0.95
Ítem 3	0.58	0.05	0.48	0.58	0.68
Ítem 4	0.35	0.06	0.23	0.35	0.47
Ítem 5	0.22	0.06	0.11	0.22	0.35
Ítem 6	0.97	0.02	0.93	0.97	0.99
Ítem 7	0.98	0.01	0.95	0.98	1.00
Ítem 8	0.97	0.01	0.93	0.97	0.99
Ítem 9	0.59	0.06	0.47	0.59	0.69
Ítem 10	0.93	0.02	0.88	0.93	0.97
Ítem 11	0.32	0.06	0.20	0.32	0.44
Ítem 12	0.87	0.04	0.80	0.87	0.94
Ítem 13	0.94	0.02	0.88	0.94	0.98
Ítem 14	0.41	0.06	0.30	0.41	0.52
Ítem 15	0.37	0.06	0.25	0.37	0.50
Ítem 16	0.94	0.02	0.89	0.94	0.98
Ítem 17	0.91	0.03	0.85	0.91	0.96
Ítem 18	0.14	0.05	0.06	0.13	0.24
Ítem 19	0.43	0.06	0.32	0.43	0.54
Ítem 20	0.39	0.06	0.28	0.39	0.50

Cuadro 5.10: Medidas de resumen de  $a_1^*$  en el modelo M2PL utilizando el método MCMC

Figura 5.9: Diagrama de cajas de las distribuciones de  $a_1^*$  en el modelo M2PL utilizando el método MCMC para los ítems de la escala de actitudes hacia la Estadística



Se observa que los ítems 6 (cuando estudio Estadística mi cabeza “queda en blanco” y no consigo pensar claramente), 7 (yo tengo una sensación de inseguridad cuando me esfuerzo en Estadística) y 8 (la Estadística me deja inquieto(a), descontento, irritado(a) e impaciente) presentan valores de tendencia central más altos y muy cercanos a uno. Esto quiere decir que estos ítems representan mejor la primera dimensión  $\theta_1$ .

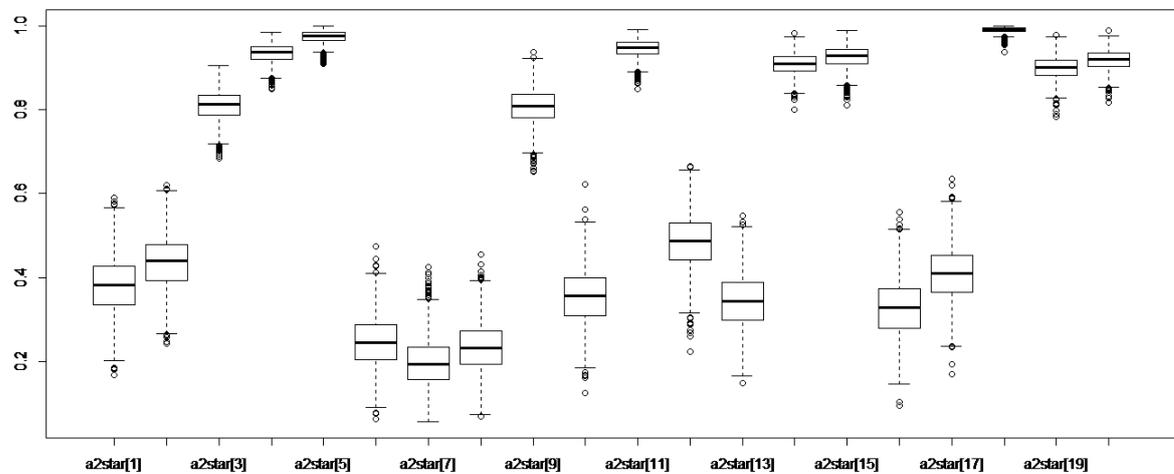
Por el contrario, los ítems 18 (yo quedo más feliz en la clase de Estadística que en la clase de cualquier otra materia) y 5 (la Estadística me hace sentir seguro(a) y es al mismo tiempo estimulante) tienen menores medidas, por lo tanto, representan en menor medida a la dimensión  $\theta_1$ .

De forma similar, se presentarán los estadísticos (Cuadro 5.11) y gráficos de cajas (Figura 5.10) para  $a_2^*$ :

	Media	D.E.	$P_{2.5}$	$P_{50}$	$P_{97.5}$
Ítem 1	0.38	0.07	0.25	0.38	0.51
Ítem 2	0.44	0.06	0.31	0.44	0.55
Ítem 3	0.81	0.03	0.74	0.81	0.88
Ítem 4	0.93	0.02	0.88	0.94	0.97
Ítem 5	0.97	0.01	0.94	0.98	0.99
Ítem 6	0.24	0.06	0.13	0.24	0.36
Ítem 7	0.20	0.06	0.10	0.19	0.32
Ítem 8	0.23	0.06	0.12	0.23	0.36
Ítem 9	0.81	0.04	0.72	0.81	0.88
Ítem 10	0.35	0.06	0.23	0.36	0.48
Ítem 11	0.95	0.02	0.90	0.95	0.98
Ítem 12	0.49	0.06	0.35	0.49	0.61
Ítem 13	0.34	0.06	0.22	0.34	0.47
Ítem 14	0.91	0.03	0.85	0.91	0.95
Ítem 15	0.92	0.03	0.87	0.93	0.97
Ítem 16	0.33	0.07	0.20	0.33	0.46
Ítem 17	0.41	0.06	0.28	0.41	0.53
Ítem 18	0.99	0.01	0.97	0.99	1.00
Ítem 19	0.90	0.03	0.84	0.90	0.95
Ítem 20	0.92	0.02	0.87	0.92	0.96

Cuadro 5.11: Medidas de resumen de  $a_2^*$  en el modelo M2PL utilizando el método MCMC

Figura 5.10: Diagrama de cajas de las distribuciones de  $a_2^*$  en el modelo M2PL utilizando el método MCMC para los ítems de la escala de actitudes hacia la Estadística



Al analizar los ítems más y menos influyentes para esta dimensión se detectaron los mismos ítems de la primera dimensión pero influyendo de manera contraria. Por ejemplo, se observa que los ítems 18 y 5 presentan promedios más altos y cercanos a 1; esto quiere decir que estos ítems representan mejor la segunda dimensión  $\theta_2$ . Por el contrario, los ítems 6, 7 y 8 tienen las menores medias, por lo tanto, representan de forma menos considerable a  $\theta_2$ .

Para comparar mejor estos valores y poder interpretar mejor las dimensiones encontradas se presenta el siguiente cuadro:

Ítem	Enunciado	$a_1^*$	$a_2^*$
1	Yo quedo terriblemente tenso(a) en la clase de Estadística	<b>0.92</b>	0.38
2	Yo no gusto de Estadística y me asusta tener que hacer el curso de Estadística	<b>0.90</b>	0.44
3	Yo creo que la Estadística es muy interesante y gusto de las clases de Estadística	0.58	<b>0.81</b>
4	La Estadística es fascinante y divertida	0.35	<b>0.93</b>
5	La Estadística me hace sentir seguro(a) y es al mismo tiempo estimulante	0.22	<b>0.97</b>
6	Cuando estudio Estadística mi cabeza “queda en blanco” y no consigo pensar claramente	<b>0.97</b>	0.24
7	Yo tengo una sensación de inseguridad cuando me esfuerzo en Estadística	<b>0.98</b>	0.20
8	La Estadística me deja inquieto(a), descontento, irritado(a) e impaciente	<b>0.97</b>	0.23
9	El sentimiento que yo tengo con relación a la Estadística es bueno	0.59	<b>0.81</b>
10	La Estadística me hace sentir como si estuviese perdido(a) en una selva de números y sin encontrar la salida	<b>0.93</b>	0.35
11	La Estadística es algo que yo aprecio grandemente	0.32	<b>0.95</b>
12	Cuando yo escucho la palabra Estadística, yo tengo un sentimiento de aversión (rechazo)	<b>0.87</b>	0.49
13	Yo encaro la Estadística con un sentimiento de indecisión, que es resultado del miedo de no ser capaz en Estadística	<b>0.94</b>	0.34
14	Yo gusto realmente de la Estadística	0.41	<b>0.91</b>
15	La Estadística es una de las materias que yo realmente gusto de estudiar en la universidad	0.37	<b>0.92</b>
16	Pensar sobre la obligación de resolver un problema de Estadística me deja nervioso(a)	<b>0.94</b>	0.33
17	Yo nunca guste de la Estadística y es la materia que más me da miedo	<b>0.91</b>	0.41
18	Yo quedo más feliz en la clase de Estadística que en la clase de cualquier otra materia	0.14	<b>0.99</b>
19	Yo me siento tranquilo(a) en Estadística y gusto mucho de esa materia	0.43	<b>0.90</b>
20	Yo tengo una reacción definitivamente positiva con relación a la Estadística: yo gusto y aprecio esa materia	0.39	<b>0.92</b>

Cuadro 5.12: Estimación de medias para  $a_1^*$  y  $a_2^*$  de los ítems de actitudes hacia la Estadística

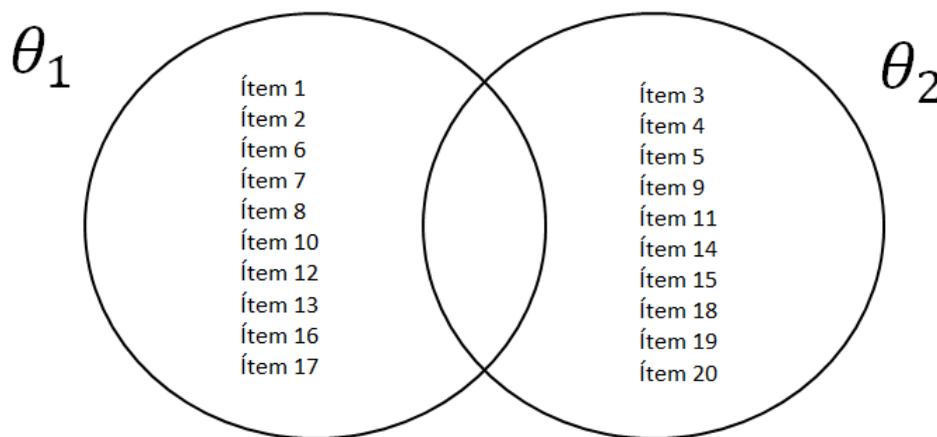
De esta manera, se confirma la bidimensionalidad del rasgo latente de actitudes hacia la estadística porque se evidencia la existencia de ítems con cargas significativas en por lo menos una de las dimensiones. Una gran ventaja es que con los modelos de TRIM uno puede relacionar los ítems con ambos rasgos latentes asociados, de esta forma es posible encontrar características del ítem que indiquen aspectos no tan específicos o estudiados del constructo (Fragoso y Curi, 2013).

Al evaluar los ítems que explican en mayor medida la primera dimensión del rasgo latente, considerando un punto de corte de 0.6, se puede incluir a los ítems 1, 2, 6, 7, 8, 10, 12, 13, 16 y 17. Esto indicaría que esta dimensión se asocia a una valoración más negativa de la Estadística, con manifestaciones de inseguridad, ansiedad, temor e insatisfacción, lo cual puede ir generando cierto rechazo a ésta materia.

Por otro lado, los ítems 3, 4, 5, 9, 11, 14, 15, 18, 19 y 20 están más asociados con la segunda dimensión, la cual manifiesta aspectos más positivos hacia la Estadística, como lo son la felicidad, seguridad, gusto, motivación y diversión, las cuales generan actitudes más favorables hacia la Estadística.

Una representación gráfica que ayuda a resumir y visualizar mejor la clasificación de los ítems relativa a las dos dimensiones adoptadas, usando los valores de la tabla anterior, sería la siguiente:

Figura 5.11: Diagrama de distribución de ítems en función a las dimensiones halladas de actitudes hacia la Estadística



Asimismo, se puede complementar la información anterior ordenando los ítems con mayor carga en las dimensiones 1 y 2, los cuadros serían los siguientes:

Ítem	Enunciado	$a_1^*$
7	Yo tengo una sensación de inseguridad cuando me esfuerzo en Estadística	<b>0.98</b>
8	La Estadística me deja inquieto(a), descontento, irritado(a) e impaciente	<b>0.97</b>
6	Cuando estudio Estadística mi cabeza “queda en blanco” y no consigo pensar claramente	<b>0.97</b>
16	Pensar sobre la obligación de resolver un problema de Estadística me deja nervioso(a)	<b>0.94</b>
13	Yo encaro la Estadística con un sentimiento de indecisión, que es resultado del miedo de no ser capaz en Estadística	<b>0.94</b>
10	La Estadística me hace sentir como si estuviese perdido(a) en una selva de números y sin encontrar la salida	<b>0.93</b>
1	Yo quedo terriblemente tenso(a) en la clase de Estadística	<b>0.92</b>
17	Yo nunca gusto de la Estadística y es la materia que más me da miedo	<b>0.91</b>
2	Yo no gusto de Estadística y me asusta tener que hacer el curso de Estadística	<b>0.90</b>
12	Cuando yo escucho la palabra Estadística, yo tengo un sentimiento de aversión (rechazo)	<b>0.87</b>

Cuadro 5.13: Ítems más relacionados con la primera dimensión hallada de actitudes hacia la Estadística

Ítem	Enunciado	$a_2^*$
18	Yo quedo más feliz en la clase de Estadística que en la clase de cualquier otra materia	<b>0.99</b>
5	La Estadística me hace sentir seguro(a) y es al mismo tiempo estimulante	<b>0.97</b>
11	La Estadística es algo que yo aprecio grandemente	<b>0.95</b>
4	La Estadística es fascinante y divertida	<b>0.93</b>
15	La Estadística es una de las materias que yo realmente gusto de estudiar en la universidad	<b>0.92</b>
20	Yo tengo una reacción definitivamente positiva con relación a la Estadística: yo gusto y aprecio esa materia	<b>0.92</b>
14	Yo gusto realmente de la Estadística	<b>0.91</b>
19	Yo me siento tranquilo(a) en Estadística y gusto mucho de esa materia	<b>0.90</b>
9	El sentimiento que yo tengo con relación a la Estadística es bueno	<b>0.81</b>
3	Yo creo que la Estadística es muy interesante y gusto de las clases de Estadística	<b>0.81</b>

Cuadro 5.14: Ítems más relacionados con la segunda dimensión hallada de actitudes hacia la Estadística

Dados estos resultados donde se aprecia una segmentación de los ítems clara por cada dimensión, según los valores de  $a_1^*$  y  $a_2^*$ , se puede concluir que el constructo estudiado de actitudes hacia la Estadística es bidimensional. Asimismo, en función a lo hallado y lo encontrado por otros autores como [Turik \(2010\)](#) y [Campos et al. \(2013\)](#), se podría definir la primera dimensión ( $\theta_1$ ) como “Autoconcepto y afectividad negativa hacia la Estadística” ya que los principales ítems expresan reacciones emocionales negativas y creencias sobre su incapacidad para la Estadística. Por otro lado, la segunda dimensión ( $\theta_2$ ) se podría denominar “Autoconcepto y afectividad positiva hacia la Estadística”, ya que los ítems están más asociados a estados emocionales positivos hacia la estadística y su aprendizaje, lo cual contribuye con una mejor imagen y capacidad de sí mismo frente a la estadística.

#### 5.4.4. Estimación de las dimensiones del constructo actitudes hacia la Estadística

Respecto a los parámetros estimados de los rasgos latentes de las personas se encontraron medias cercanas a cero debido a la distribución a priori dada. Para poder interpretar mejor estas puntuaciones se realizó una conversión lineal para tener resultados con una media alrededor de 500 y todas las puntuaciones positivas, para esto se multiplicó el rasgo latente por 100 y se le sumó 500. Luego de esto, se obtuvieron las siguientes medidas resumen para las dos dimensiones señaladas anteriormente:

Medida	$\theta_1$	$\theta_2$
Media	499.48	494.38
Mediana	501.03	489.02
D.E.	85.50	83.32
Asimetría	-0.11	0.38
Curtosis	-0.84	-0.78

Cuadro 5.15: Medidas resumen de las estimaciones de los rasgos latentes utilizando el método MCMC

Las estimaciones de las medias de los rasgos latentes pueden ser interpretadas en términos de distribuciones de respuestas de las personas dentro de los grupos de ítems asociados en mayor medida a cada rasgo latente. Por ejemplo, el vector estimado de rasgos latentes de la persona número 28 es (449.24, 545.64), lo cual indica un valor de la primera componente de este vector por debajo de la media y un valor por encima de la media en la segunda componente.

Las distribuciones y dispersión de los puntajes de cada dimensión se pueden observar con mayor detalle en los siguientes gráficos:

Figura 5.12: Histograma de puntuaciones de  $\theta_1$

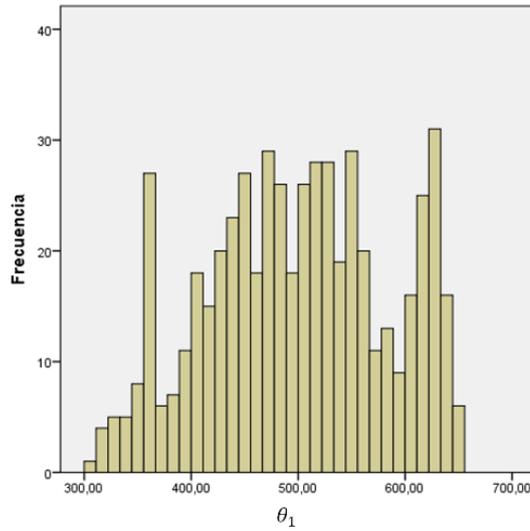


Figura 5.13: Histograma de puntuaciones de  $\theta_2$

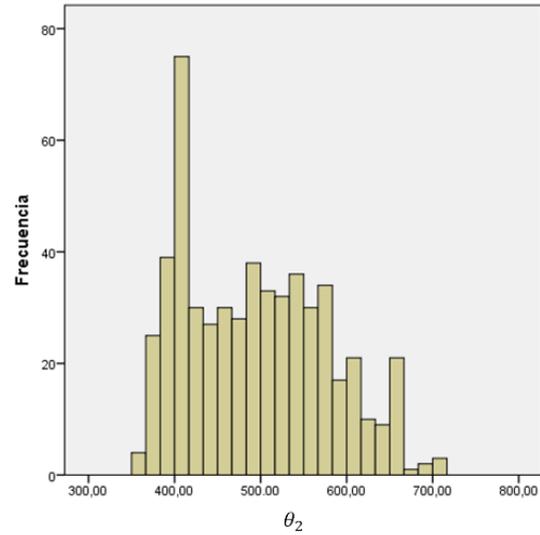


Figura 5.14: Diagrama de caja de  $\theta_1$

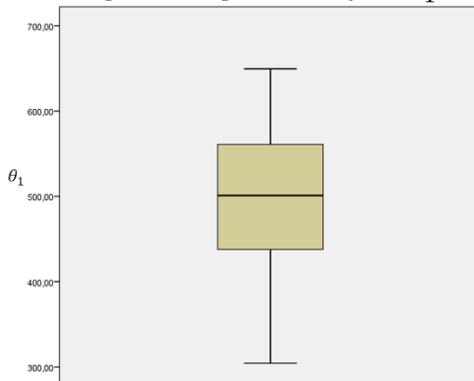
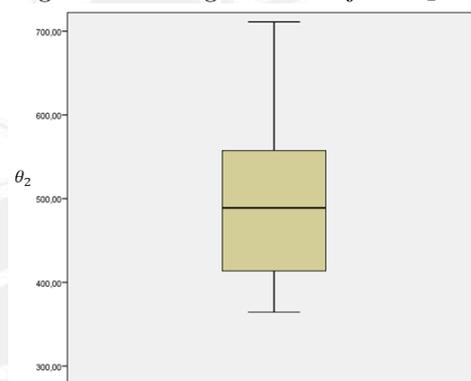
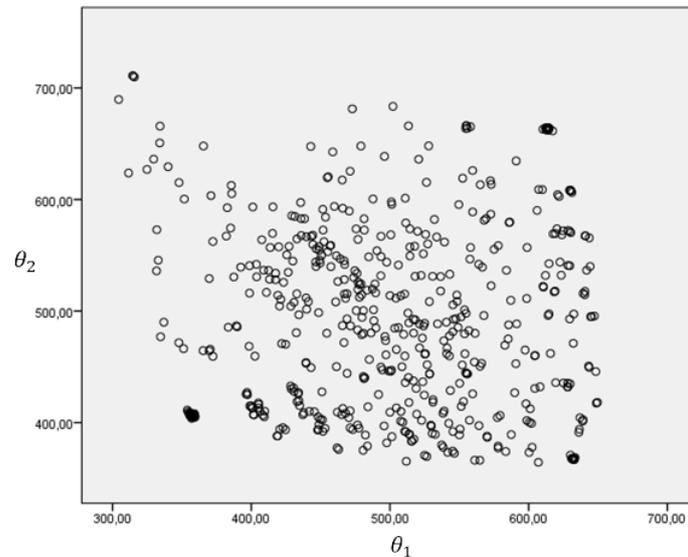


Figura 5.15: Diagrama de caja de  $\theta_2$



Como se observa, el comportamiento de las medidas de las dimensiones encontradas varía en la muestra estudiada.

Para relacionar ambas dimensiones se recurrió a un coeficiente de correlación lineal de Pearson, sin embargo el coeficiente encontrado fue 0.028 ( $p > 0.05$ ), lo cual demuestra que no existe una correlación lineal significativa entre estas dos dimensiones. Este resultado se puede complementar con en el siguiente gráfico:

Figura 5.16: Diagrama de dispersión de las puntuaciones de  $\theta_1$  y  $\theta_2$ 


Estos resultados muestran mayor detalle de las puntuaciones en cada dimensión por parte del grupo participante, también, permiten verificar que el constructo estudiado de actitudes hacia la Estadística es bidimensional y no existe una correlación significativa entre las dos dimensiones. Asimismo, como se mencionó anteriormente, se puede denominar a la primera dimensión ( $\theta_1$ ) como “Autoconcepto y afectividad negativa hacia la Estadística” y a la segunda dimensión ( $\theta_2$ ) como “Autoconcepto y afectividad positiva hacia la Estadística” para los siguientes análisis estadísticos comparativos entre diferentes grupos.

#### 5.4.5. Análisis de los rasgos latentes hallados según algunas características de los evaluados

Los análisis comparativos se realizarán con las puntuaciones en las dimensiones encontradas en la prueba de actitudes hacia la Estadística considerando características de interés como género y facultad. Esto servirá a la vez para comparar los resultados con lo encontrado por la investigación de Pérez et al. (2015).

Al realizar un contraste de hipótesis de normalidad de los puntajes de ambas dimensiones del constructo mediante la prueba de Kolmogorov - Smirnov, se encuentra que las distribuciones no tienden a una curva normal, por tal razón, se usarán pruebas no paramétricas para las comparaciones. Cabe resaltar que en el estudio de Pérez et al. (2015) se encontró que los puntajes del constructo unidimensional sí tienden a una distribución normal.

- **Comparación actitudes hacia la Estadística de acuerdo al género de los evaluados**

Para poder comparar los puntajes de las dimensiones halladas entre hombres y mujeres se procedió a realizar un contraste U de Mann Whitney y se encontraron los siguientes resultados:

Prueba	Dimensiones	Género	n	Mediana	R.I.	U	Sig.
AEC	Autoc. y afec. negativa	Masculino	350	496.31	120.11	30791	.058
		Femenino	195	513.33	139.58		
	Autoc. y afec. positiva	Masculino	350	494.86	145.38	33262	.624
		Femenino	195	480.25	141.04		

Cuadro 5.16: Comparación según el género usando los puntajes de las dimensiones de las actitudes hacia la Estadística

De acuerdo a los resultados observados, los puntajes de las actitudes hacia la Estadística no presentan diferencias significativas entre ambos grupos. Sin embargo, hay más probabilidad de encontrar diferencias significativas entre hombres y mujeres en la primera dimensión, siendo las mujeres las que tienen más autoconcepto y afectividad negativa hacia la Estadística.

Estos resultados son similares de cierta forma con lo encontrado por Pérez et al. (2015) cuando compararon estos grupos contemplando solo el puntaje total de la prueba (unidimensional). Sin embargo, con el análisis descrito en el presente estudio se encuentra que con los modelos de TRIM se logra analizar y explicar de forma más detallada y precisa las interacciones entre las personas y los ítems de una prueba multidimensional.

- **Comparación de actitudes hacia la Estadística de acuerdo a la facultad de los evaluados**

En este caso se realizó una prueba de Kruskal Wallis dado que se deben comparar más de dos grupos independientes. Se encontró lo siguiente:

Prueba	Dimensiones	Facultad	n	Mediana	R.I.	$\chi^2$	Sig.
AEC	Autoconcepto y afectividad negativa	EIAM	267	496.03	123.48	3.86	.145
		C. Exactas e Ingeniería	219	511.43	143.29		
		Economía	59	501.03	108.48		
	Autoconcepto y afectividad positiva	EIAM	267	480.08	129.41	9.03	.011
		C. Exactas e Ingeniería	219	501.63	155.07		
		Economía	59	511.29	122.58		

Cuadro 5.17: Comparación según la facultad usando los puntajes de las dimensiones de las actitudes hacia la Estadística

Del cuadro anterior, se observa que no existen diferencias significativas entre las facultades en la dimensión de autoconcepto y afectividad negativa hacia la Estadística; sin embargo, sí se encuentran diferencias significativas entre algunas de las facultades en la dimensión de autoconcepto y afectividad positiva hacia la Estadística. Por ello, se realizaron pruebas U Mann Whitney en esta última dimensión para identificar, específicamente, entre qué facultades se encontraban las diferencias significativas.

Se encontró que el autoconcepto y afectividad positiva hacia la Estadística es significativamente menor en la facultad de EIAM ( $Me = 480.08$ ) respecto a las otras dos facultades. Con la facultad de Ciencias Exactas e Ingeniería ( $Me = 501.63$ ) se encuentran las siguientes estadísticas:  $U = 25305$ ,  $Z = -2.55$ ,  $p = .011$ ; y respecto a la facultad de Economía ( $Me = 511.29$ ), se encuentra lo siguiente:  $U = 6388$ ,  $Z = -2.27$ ,  $p = .023$ .

Al aplicar la prueba a posteriori de Bonferroni se encontró que la única diferencia significativa ( $p = .04$ ) se encuentra entre las facultades de EIAM e Ingeniería, siendo esta última la que obtiene mayores puntajes en la dimensión de autoconcepto y afectividad positiva hacia la Estadística.

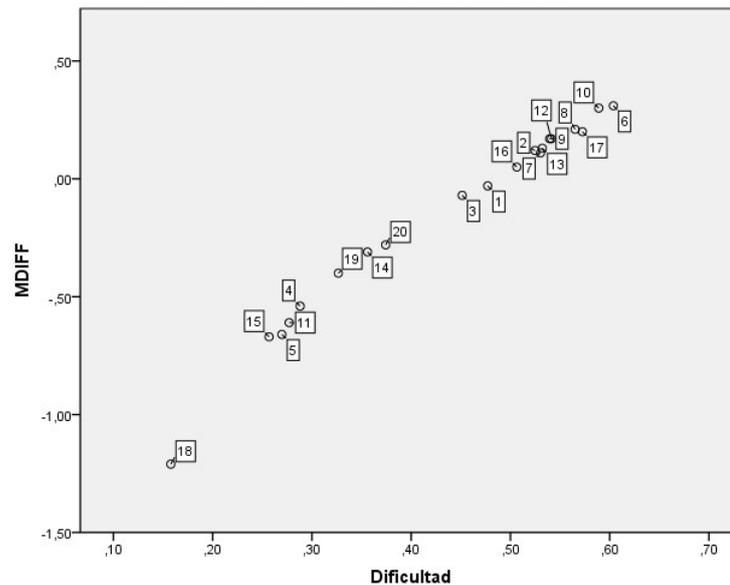
Cabe destacar, que Pérez et al. (2015) también encontraron diferencias significativas (de medias) en el constructo unidimensional entre las facultades; sin embargo, con el análisis bidimensional mostrado en esta investigación se puede precisar mejor donde se presenta en mayor medida esta diferencia, siendo ésta en la dimensión más positiva de la actitud hacia la Estadística.

#### 5.4.6. Comparación con puntajes de la Teoría Clásica de los Test

- Comparación de índices de “dificultad” de los ítems

El diagrama de dispersión de puntos entre el índice multidimensional de dificultad (MDIFF) y el índice de dificultad de TCT presenta una relación lineal alta y significativa ( $r = .98$ ,  $p < .05$ ) como se puede observar a continuación:

Figura 5.17: Diagrama de dispersión de los índices de dificultad de TRIM (MDIFF) y TCT (Dificultad)

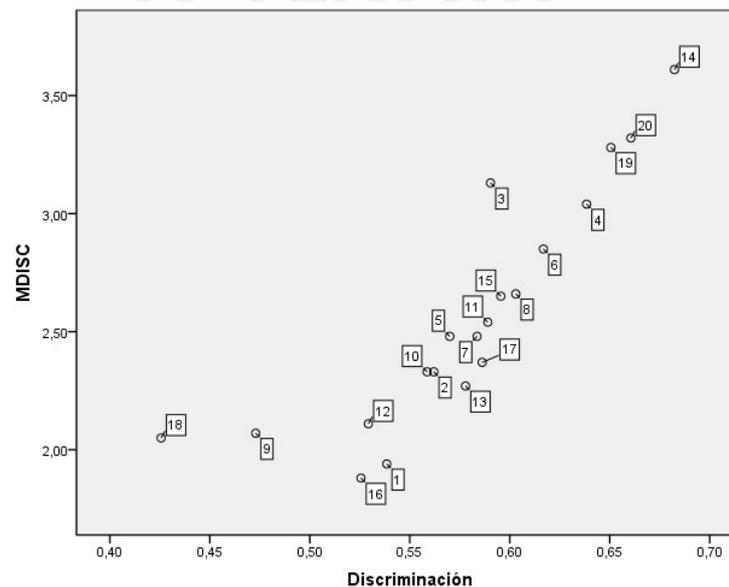


Como en análisis anteriores, se evidencia que el ítem 18 es el que más se aleja del resto del grupo de ítems de la prueba, tanto con el enfoque de TRIM como con TCT. Sin embargo, este ítem igual se ajusta adecuadamente a la asociación lineal mostrada entre ambos enfoques en índices de dificultad.

- **Comparación de índices de discriminación de los ítems**

El diagrama de dispersión de puntos entre el índice multidimensional de discriminación (MDISC) y el índice de discriminación de TCT también presenta una relación lineal alta y significativa ( $r = .86, p < .05$ ). Sin embargo, esta relación lineal es menor al caso anterior pues hay algunos ítems que no se ajustan tan bien a la linealidad percibida, esto se puede observar a continuación:

Figura 5.18: Diagrama de dispersión de los índices de discriminación de TRIM (MDISC) y TCT (Discriminación)



En el gráfico mostrado se puede observar que tres ítems se alejan un poco más del grupo que se ajusta mejor a una recta. El ítem más alejado es nuevamente el ítem 18 (Yo quedo más feliz en la clase de Estadística que en la clase de cualquier otra materia), seguido del ítem 19 (Yo me siento tranquilo(a) en Estadística y gusto mucho de esa materia), asimismo ambos ítems presentan puntajes bajos de discriminación tanto con TRIM como con TCT. Por otro lado, el ítem 3 (Yo creo que la Estadística es muy interesante y gusto de las clases de Estadística) posee un puntaje de discriminación por encima del promedio, pero su puntaje calculado con TRIM tiende a ser mayor que el calculado con TCT.

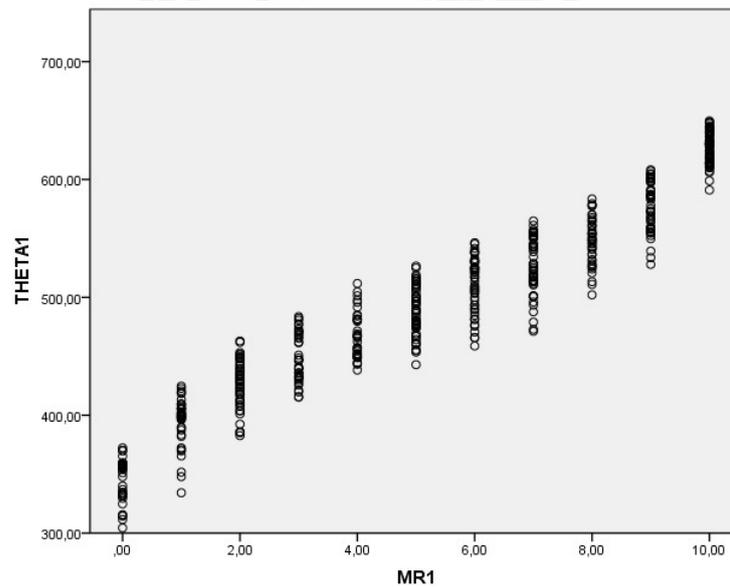
Es difícil saber exactamente a qué se deben los cambios identificados en los tres ítems mencionados anteriormente, sin embargo, estos tres ítems mantienen ciertas características en común: son ítems positivos, se encuentran en la misma dimensión (actitudes más favorables hacia la estadística) y en los tres casos la TCT subestima sus poderes de discriminación respecto a la TRIM.

Esto último podría estar asociado, en parte, a que el ítem 18 según la TRIM es el ítem que está más relacionado con la dimensión  $\theta_2$  (Cuadro 5.14) y los ítems 19 y 3 son los que puntuaron más en términos de relación con la dimensión contraria a la que pertenecen (Cuadro 5.12). Asimismo, según la TCT el poder de discriminación es un cálculo más simple y depende solo de la correlación entre el ítem y el puntaje total de la dimensión, por lo que el valor de discriminación de un ítem será más bajo mientras más diferente sea del grupo de ítems de su dimensión.

- **Comparación de puntajes de las dimensiones halladas mediante TRIM y TCT**

Los puntajes de la primera dimensión mantienen una relación lineal positiva como se observa en el siguiente gráfico:

Figura 5.19: Diagrama de dispersión de las puntuaciones de  $\theta_1$  y MR1



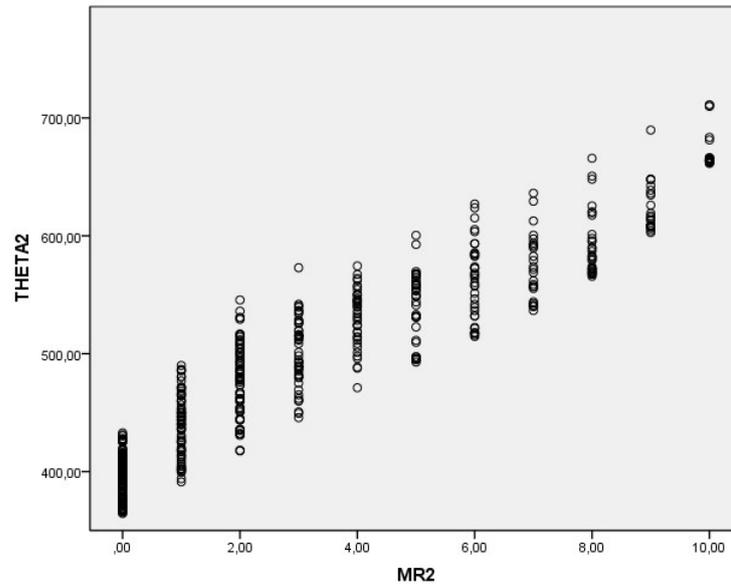
Por otro lado, un cuadro de contingencia dividiendo los puntajes hallados con TRIM en deciles brinda mayor información de la coincidencia como se observa a continuación:

Cuadro 5.18: Tabla de contingencia de las puntuaciones de  $\theta_1$  (deciles) y  $MR1$

	MR1											Total
	0	1	2	3	4	5	6	7	8	9	10	
$\theta_1$ (deciles) 1	49	7	0	0	0	0	0	0	0	0	0	56
2	0	30	19	4	0	0	0	0	0	0	0	53
3	0	1	26	19	8	1	0	0	0	0	0	55
4	0	0	5	11	17	14	5	2	0	0	0	54
5	0	0	0	5	10	23	11	6	0	0	0	55
6	0	0	0	0	3	16	15	15	5	0	0	54
7	0	0	0	0	0	2	17	13	19	4	0	55
8	0	0	0	0	0	0	0	11	20	23	0	54
9	0	0	0	0	0	0	0	0	0	23	32	55
10	0	0	0	0	0	0	0	0	0	0	54	54
Total	49	38	50	39	38	56	48	47	44	50	86	545

Los puntajes de la segunda dimensión también mantienen una relación lineal positiva como se observa en el gráfico siguiente:

Figura 5.20: Diagrama de dispersión de las puntuaciones de  $\theta_2$  y  $MR2$



En el cuadro de contingencia dividiendo los puntajes hallados con TRIM en deciles se evidencia que son pocas las personas que no coinciden en puntajes similares entre el enfoque de TRIM y TCT.

Cuadro 5.19: Tabla de contingencia de las puntuaciones de  $\theta_2$  (deciles) y  $MR2$ 

	MR2											Total
	0	1	2	3	4	5	6	7	8	9	10	
$\theta_2$ (deciles) 1	53	2	0	0	0	0	0	0	0	0	0	55
2	45	9	0	0	0	0	0	0	0	0	0	54
3	39	14	2	0	0	0	0	0	0	0	0	55
4	3	27	19	5	0	0	0	0	0	0	0	54
5	0	16	23	13	3	0	0	0	0	0	0	55
6	0	1	19	15	8	9	2	0	0	0	0	54
7	0	0	6	15	13	7	10	4	0	0	0	55
8	0	0	1	0	15	18	9	7	4	0	0	54
9	0	0	0	1	1	3	12	11	19	8	0	55
10	0	0	0	0	0	0	3	3	7	17	24	54
Total	140	69	70	49	40	37	36	25	30	25	24	545

Luego del análisis de comparación entre la TCT y TRIM se puede evidenciar que los resultados tienen similitudes, sin embargo, existen diferencias más evidentes en los índices de discriminación de los ítems y en las puntuaciones obtenidas en las dos dimensiones.

## Capítulo 6

# Conclusiones

### 6.1. Conclusión

Los modelos unidimensionales de TRI han demostrado ser útiles en determinadas condiciones, sin embargo, existe una necesidad de modelos de TRI que puedan reflejar con mayor precisión la complejidad de las interacciones entre las personas y los ítems de una prueba. Los modelos multidimensionales amplían las limitaciones de aplicación de TRI con múltiples rasgos latentes, con lo cual se puede lograr un poder explicativo superior a los modelos unidimensionales sin necesidad de requerir de un implemento computacional más complejo.

#### Conclusiones metodológicas

- En el presente trabajo de tesis se desarrolló el modelo de Teoría de Respuesta al Ítem Multidimensional logístico compensatorio de dos parámetros de respuesta binaria (M2PL).
- Se exploraron y analizaron métodos de estimación clásica disponibles en el programa R mediante los paquetes *mirt* y *ltm*; sin embargo, no se mostraron consistentes con los resultados obtenidos en función a un software comercial (*IRTPRO*). Por esta razón, se dió mayor énfasis a la inferencia bayesiana por ser más estable y de libre uso.
- La estimación del modelo se realizó mediante inferencia bayesiana con el método de MCMC, usando el software R y R2WinBUGS. Cabe resaltar que este tipo de procedimientos computacionales de estimación para modelos de TRIM suele demorar por la gran cantidad de parámetros a estimar, tanto para los ítems como para los rasgos latentes de las personas.
- Para la implementación computacional mediante MCMC es necesario considerar un número elevado de iteraciones (aproximadamente cien mil) aunque los resultados demoren más; esto ayudará a determinar mejor la estimación y evaluar de forma más precisa la convergencia de los resultados, lo cual es necesario para poder confiar en lo encontrado.
- Pueden ocurrir ciertos problemas numéricos en la implementación computacional ya que las probabilidades encontradas pueden ser muy cercanas a 1 o muy cercanas a 0, lo que hace que los valores en la distribución tiendan a  $+\infty$  o  $-\infty$  respectivamente. Por esta razón, se recurre a un artificio del algoritmo del modelo, el cual asigna un valor de truncación fijo (Ntzoufras, 2009).

- En el estudio de simulación, para analizar la calidad de las estimaciones de los parámetros, se comprobó que el método de estimación con mejores indicadores de ajustes y, por lo tanto, más robusto para el modelo planteado, es el bayesiano mediante MCMC. Asimismo, las estimaciones realizadas con el software comercial *IRTPRO* también fueron adecuadas y muy similares a las realizadas mediante el *R2WinBUGS*, sin embargo tiene la desventaja de que no es un software de libre uso.

### Conclusiones de la aplicación

- La aplicación fue realizada con datos proporcionados por [Pérez et al. \(2015\)](#) sobre actitudes hacia la Estadística en una muestra de 545 estudiantes entre 17 y 25 años de edad, que cursaban el primer o segundo semestre académico de una universidad privada de Colombia en las siguientes facultades: Internacional de Administración y Marketing (EIAM), Ciencias Exactas e Ingeniería, y Economía. El instrumento consta de 20 ítems con formato de respuesta tipo escala Likert, sin embargo, para fines del presente estudio se convirtieron a respuestas binarias o dicotómicas.
- Se planteó el modelo M2PL debido a la bidimensionalidad teórica y porque el tercer parámetro (adivinación) no sería adecuado incluirlo debido a que la prueba psicológica usada implica que las personas evaluadas no tengan motivos a priori para responder los ítems al azar.
- La estimación de los parámetros mediante el método MCMC permite concluir que los ítems que más discriminan (considerando MDISC) son el 14 (Yo gusto realmente de la Estadística) en mayor medida y luego el 20 (Yo tengo una reacción definitivamente positiva con relación a la Estadística: yo gusto y aprecio esa materia) y 19 (Yo me siento tranquilo(a) en Estadística y gusto mucho de esa materia), por el contrario, los que menos discriminan son el ítem 16 (Pensar sobre la obligación de resolver un problema de Estadística me deja nervioso(a)) y el 1 (Yo quedo terriblemente tenso(a) en la clase de Estadística). Asimismo, el ítem que es considerado como índice de mejores actitudes hacia la Estadística (considerando MDIFF) sería principalmente el ítem 18 (Yo quedo más feliz en la clase de Estadística que en la clase de cualquier otra materia). Por otro lado, los ítems menos considerados para tener mejores actitudes hacia la estadística serían el 6 (Cuando estudio Estadística mi cabeza “queda en blanco” y no consigo pensar claramente) y el 10 (La Estadística me hace sentir como si estuviese perdido(a) en una selva de números y sin encontrar la salida).
- Se confirmó la bidimensionalidad de los resultados del constructo medido, cada dimensión está conformada por 10 ítems con cargas significativas por dimensión en función a  $a_1^*$  o  $a_2^*$ . Asimismo, en función a lo hallado y lo encontrado por otros autores como [Campos et al. \(2013\)](#), [Estrada \(2011\)](#), [Turik \(2010\)](#) y [Campos et al. \(2010\)](#), se podría definir la primera dimensión ( $\theta_1$ ) como “Autoconcepto y afectividad negativa hacia la Estadística” y la segunda dimensión ( $\theta_2$ ) como “Autoconcepto y afectividad positiva hacia la Estadística”.

- En la muestra estudiada no se encontraron diferencias significativas entre hombres y mujeres en ninguna dimensión del constructo estudiado. Por otro lado, se encontró que la única diferencia significativa entre facultades se presentaba en la dimensión de “Autoconcepto y afectividad positiva hacia la Estadística”, siendo la de Ingeniería mayor a la de EIAM.
- Sobre la comparación entre la TCT y TRIM, se puede evidenciar que los resultados con ambos enfoques tienen similitudes, sin embargo, existen diferencias más evidentes en los índices de discriminación de los ítems y en las puntuaciones obtenidas en las dos dimensiones.

## 6.2. Sugerencias para investigaciones futuras

- Si bien el conjunto de datos es dicotómico, se podría analizar también en su estado original de escala Likert, para lo cual se usarían categorías para los puntajes de los ítems. De esta manera, el modelo de TRI sugerido por Kelderman y Rijkens (1994) es el politómico o modelo para escalas, denominado generalmente como modelos de crédito parcial, para los cuales también existen extensiones que consideran múltiples rasgos latentes. Estudios relacionados a este enfoque se pueden encontrar en [Tarazona \(2013\)](#), [Ostini y Nering \(2006\)](#) y [Hoskens y De Boeck \(2001\)](#).
- Los modelos multidimensionales pueden también ser utilizados para estudios longitudinales ([te Marvelde, J., Glas, Landeghem y Damme, 2006](#)), por lo cual, el conjunto de datos presentados puede servir para el inicio de una investigación como esta.
- Respecto al prolongado tiempo de estimación bayesiana, sería adecuado buscar alternativas computacionales para superar esta limitación.
- En el área de medición psicológica en Perú aún se usa mayoritariamente la TCT, la cual es una teoría bastante limitada en estos tiempos. Por esta razón, esta tesis pretende mostrar, evaluar e incentivar análisis psicométricos más completos a través de la TRIM que permitan dar mayor información sobre las características de los ítems. Asimismo, se pretende resaltar la ventaja de usar inferencias mediante métodos bayesianos, los cuales están asumiendo un rol importante en el contexto de investigaciones psicológicas al lograr solucionar problemas que la inferencia clásica no logra resolver ([González, 2010](#)).
- Es importante seguir analizando tests psicológicos de ejecución típica mediante la TRI, ya que existe escasa aplicación y más aún con modelos de TRIM, los cuales suelen ser más representativos para el análisis psicométrico de este tipo de pruebas psicológicas ([Abal et al., 2010](#)).

## Referencias

- Abal, F., Lozzia, G., Aguerri, M., Galibert, H. y Atorresi, H. (2010). La escasa aplicación de la teoría de respuesta al ítem en tests de ejecución típica, *Revista colombiana de psicología* **19(1)**: 111–122.
- Antonio, D. (2013). *Modelo multidimensional de teoría de respuesta al ítem jerárquico*, Master's thesis, Universidad Nacional de Colombia.
- Aparicio, A. (2015). *Avaliacao das Atitudes no curso Estatística: contextos universitarios Latino-Americanos.*, PhD thesis, Universidade de Sao Paulo.
- Baker, F. B. y Kim, S. (2004). *Item Response Theory - Parameter Estimation Techniques*, Marcel Dekkerl.
- Bartholomew, D., Knott, M. y Moustaki, I. (2011). *Latent Variable Models and Factor Analysis*, Wiley.
- Bazán, J. (2014). Enfoque Bayesiano nos Modelos de Teoria da Resposta ao Item, *Technical report*, Associacao Brasileira de Estatística - Simposio Nacional de Probabilidade e Estatística (SINAPE).
- Bazán, J., Valdivieso, L. y Calderón, A. (2010). Enfoque Bayesiano en Modelos de Teoría de Respuesta al ítem, *Technical report*, Departamento de Ciencias, Sección Matemática. Pontificia Universidad Católica del Perú.
- Bock, R. y Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an em algorithm, *Psychometrika* **46(4)**: 443–459.
- Campos, J., Bonafe, F., Dovigo, L. y Maroco, J. (2010). Escala de Atitudes em relacao a Estatística: Estudo de validacao, *Technical report*, Associacao Brasileira de Estatística - Simposio Nacional de Probabilidade e Estatística (SINAPE).
- Campos, J., Bonafe, F., Dovigo, L. y Maroco, J. (2013). Avaliacao psicométrica da escala de atitudes em relacao a estadística, *Revista Brasileira de Biometria* **31(2)**: 327–337.
- Cazorla, I., Silva, C., Vendramini, C. y Brito, M. (1999). Adaptacao e validacao de uma escala de atitudes em relacao a estadística, *Anais da Conferencia Internacional: Experiencias e perspectivas do ensino de Estadística, desafios para o seculo XXI* pp. 45–58.
- Chalmers, R. (2012). mirt: A multidimensional item response theory package for the R environment, *Journal of Statistical Software* **48(6)**: 1–29.

- Chincaro, O. (2010). *Modelo de Rasch dicotómico con aplicación a la educación*, Master's thesis, Pontificia Universidad Católica del Perú.
- Curtis, S. (2010). BUGS code for item response theory, *Journal of Statistical Software* **36(1)**: 1–34.
- Demars, C. (2010). *Item Response Theory*, Oxford.
- Ekstrom, J. (2011). *The Phi-coefficient, the Tetrachoric Correlation Coefficient, and the Pearson-Yule Debate*, Department of Statistics, UCLA.
- Embretson, S. y Reise, S. (2000). *Item Response Theory for Psychologists*, Lawrence Erlbaum Associates.
- Estrada, A. (2011). Instrumentos de medición de actitudes hacia la Estadística, *Investigación en Educación Matemática. Comunicación de los Grupos de Investigación de la SEIEM* pp. 233–251.
- Flores, S. (2012). *Modelos testlet logísticos y logísticos de exponente positivo para pruebas de comprensión de textos*, Master's thesis, Pontificia Universidad Católica del Perú.
- Fox, J. P. (2010). *Bayesian Item Response Modeling*, Springer.
- Fragoso, T. (2010). *Modelos multidimensionais da teoria de resposta ao item*, Master's thesis, Universidade de Sao Paulo.
- Fragoso, T. y Curi, M. (2013). Improving psychometric assessment of the Beck Depression Inventory usin Multidimensional Item Response Theory, *Biometrical Journall* **55(4)**: 527–540.
- Giaconi, V. (2012). *Aportes del análisis psicométrico a la comprensión de la estructura del conocimiento matemático para enseñar*, Master's thesis, Universidad de Chile.
- González (2010). Bayesian Methods in Psychological Research: The case of IRT, *International Journal of Psychological Research* **3(1)**: 164–176.
- Hasting, J. y Hohler, J. (2009). Multidimensional IRT models for the assesment of competencies, *Studies in Educationa Evaluation* **35**: 57–63.
- Hoskens, M. y De Boeck, P. (2001). Multidimensional componential item response theory models for polytomous items, *Applied Psychological Measurement* **25**: 19–37.
- Lee, J. (2012). *Multidimensional Item Response Theory: an investigation of interaction effects between factors on item parameter recovery using Markov Chain Monte Carlo*, PhD thesis, Michigan State University.
- Levy, J. y Varela, J. (2003). *Análisis Multivariable para las Ciencias Sociales*, Pearson.
- Martínez, M., Hernández, M. y Hernández, M. (2006). *Psicometría*, Alianza Editorial.

- Montenegro, A. (2010). *Multidimensional Item Response Theory Models where the Ability has a Latent Linear Structure*, PhD thesis, Universidad Nacional de Colombia.
- Montero, E. (2000). Teoría de respuesta a los ítemes: una moderna alternativa para el análisis psicométrico de instrumentos de medición, *Revista de Matemática: Teoría y Aplicaciones* **7(1-2)**: 217–228.
- Montero, E. y Jiménez, K. (2013). Aplicación del modelo de Rasch en el análisis psicométrico de una prueba de diagnóstico en matemática, *Revista digital Matemática, Educación e Internet* **13(1)**: 1–24.
- Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*, Wiley.
- Ostini, R. y Nering, M. (2006). *Polytomous Item Response Theory Models*, Sage.
- Pérez, L., Aparicio, A., Bazán, J. y Abdounur, O. (2015). Actitudes hacia la estadística en estudiantes de una universidad privada en Colombia, *Educación Matemática* **27(3)**: 111–149.
- Quaresma, E. (2014). *Modelagem para construcao de escalas avaliativas e classificatorias em exames seletivos utilizando teoria da resposta ao item uni e multidimensional*, PhD thesis, Universidade de Sao Paulo.
- Reckase, M. (1996). A Linear Logistic Multidimensional Model for Dichotomous Item Response Data. In: Van der Linden, W. J. and Hambleton, R. K. (Eds.), *Handbook of Modern Item Response Theory* pp. 271–286.
- Reckase, M. (2009). *Multidimensional Item Response Theory*, Springer.
- Revelle, W. (2015). *psych: Procedures for Personality and Psychological Research*, Northwestern University, Evanston. R package version 1.5.8.
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses, *Journal of Statistical Software* **17(5)**: 1–25.
- Sturtz, S., Ligges, U. y Gelman, A. (2005). R2WinBUGS: A Package for Running WinBUGS from R, *Journal of Statistical Software* **12(3)**: 1–16.
- Tarazona, E. (2013). *Modelos alternativos de respuesta graduada con aplicaciones en la calidad de servicios*, Master's thesis, Pontificia Universidad Católica del Perú.
- te Marvelde, J., Glas, C., Landeghem, G. y Damme, J. (2006). Application of multidimensional item response theory models to longitudinal data, *Educational and Psychological Measurement* **66(1)**: 5–34.
- Turik, C. (2010). *Análise de atitudes de alunos universitarios em relacao a estatística por meio da teoria da resposta ao item (TRI)*, Master's thesis, Pontificia Universidade Católica do Rio Grande do Sul.
- Valdivieso, L., Bayes, C. y Tarazona, E. (2014). *Modelos de variables latentes. Notas de clase*, Escuela de Posgrado, Maestría en Estadística. Pontificia Universidad Católica del Perú.

## Apéndice A

### Programas en R y R2WinBUGS

#### Simulación

a) Simular parámetros

---

```
set.seed(1234)
require(mirt)

asm = matrix(c( 1.14, .70,
1.06, .97,
2.64, 0.71,
1.09, 1.46,
2.43, .46,
1.31, .34,
3.09, .43,
1.45, .30,
1.22, .96,
.80, .86,
.59, 1.03,
.70, 1.17,
.99, 1.16,
.63, .64,
1.12, 1.30,
.32, 1.21,
.53, 1.40,
.47, .73,
.15, .58,
.50, .59,
.36, .74),21,2,byrow=TRUE)

d = matrix(c(.76,-.69,-1.58,.11,-1.40,-.89,.47,.79,-2.49,-1.17,.05,
-.73,-.61,-.73,-.18,-.02,.58,-1.08,-2.26,-.82,-1.33),ncol=1)

datasetf = simdata(asm, d, 1000, itemtype = 'dich')

write.csv(datasetf, 'datasetf.csv')
```

---

## b) Función simdata

---

```

function (a, d, N, itemtype, sigma = NULL, mu = NULL, guess = 0,
upper = 1, nominal = NULL, Theta = NULL)
{ fn <- function(p, ns) sample(1L:ns, 1, prob = p)
nfact <- ncol(a)
nitems <- nrow(a)
K <- rep(0, nitems)
if (length(guess) == 1L)
guess <- rep(guess, nitems)
if (length(guess) != nitems)
stop("Guessing parameter is incorrect")
if (length(upper) == 1L)
upper <- rep(upper, nitems)
if (length(upper) != nitems)
stop("Upper bound parameter is incorrect")
if (length(itemtype) == 1L)
itemtype <- rep(itemtype, nitems)
for (i in 1L:length(K)) {
K[i] <- length(na.omit(d[i, ])) + 1L
if (itemtype[i] == "partcomp")
K[i] <- 2L
if (any(itemtype[i] == c("gpcm", "nominal", "nestlogit")))
K[i] <- K[i] - 1L
}
K <- as.integer(K)
if (any(guess > 1 || guess < 0))
stop("guess input must be between 0 and 1")
if (any(upper > 1 || upper < 0))
stop("upper input must be between 0 and 1")
guess <- logit(guess)
upper <- logit(upper)
oldguess <- guess
oldupper <- upper
guess[K > 2L] <- upper[K > 2L] <- NA

```

---

---

```

guess[itemtype == "nestlogit"] <- oldguess[itemtype == "nestlogit"]
upper[itemtype == "nestlogit"] <- oldupper[itemtype == "nestlogit"]
if (is.null(sigma))
sigma <- diag(nfact)
if (is.null(mu))
mu <- rep(0, nfact)
if (!is.null(Theta))
if (ncol(Theta) != nfact — nrow(Theta) != N)
stop("The input Theta matrix does not have the correct dimensions")
if (is.null(Theta))
Theta <- mirt rmvnorm(N, mu, sigma, check = TRUE)
if (is.null(nominal))
nominal <- matrix(NA, nitems, max(K))
data <- matrix(0, N, nitems)
a[is.na(a)] <- 0
for (i in 1L:nitems) {
if (itemtype[i] == "nestlogit") {
par <- na.omit(c(a[i, ], d[i, 1], guess[i], upper[i],
nominal[i, -1L], d[i, -1L]))
obj <- new(itemtype[i], par = par, nfact = nfact,
correctcat = 1L) }
else {
if (itemtype[i] == "gpcm") {
par <- na.omit(c(a[i, ], 0:(K[i] - 1), d[i, ],
guess[i], upper[i])) }
else {
par <- na.omit(c(a[i, ], nominal[i, ], d[i, ],
guess[i], upper[i]))
}
obj <- new(itemtype[i], par = par, nfact = nfact)
}
if (any(itemtype[i] == c("gpcm", "nominal", "nestlogit")))
obj@ncat <- K[i]
P <- ProbTrace(obj, Theta)
data[, i] <- apply(P, 1, fn, ns = ncol(P))
if (any(itemtype[i] == c("dich", "gpcm", "partcomp")))
data[, i] <- data[, i] - 1L
}
colnames(data) <- paste("Item", 1L:nitems, sep = "")
return(data) }

```

---

c) Estimar parámetros

---

```

#mirt
mod = mirt(datasetf, 2, method = 'EM')
coef(mod)
summary(mod)

#ltm
require(ltm)
fit2LM =ltm(datasetf z1+z2)
fit2LM

#WinBUGS
y=as.matrix(datasetf)
colnames(y) <- NULL

n=nrow(datasetf)
I=ncol(datasetf)

data<-list(y=y,n=n,I=I)

inits<-function(){list(a1=c(rep(1,I)),a2=c(rep(1,I)),
d=c(rep(0,I)),theta1=c(rep(0.5,n)),theta2=c(rep(0.5,n)))}

parameters<-c("a1", "a2", "d", "mdifc", "mdisc", "a1star", "a2star", "theta1", "theta2")

n.iter=105000
n.burnin=5000
n.thin=50

library(R2WinBUGS)
bd="C:/WinBUGS14/"

bayes2LMW<-bugs(data,inits=inits,parameters.to.save=parameters,
model.file="Logistic 2LM ModelF.txt",working.directory=wd,
n.chains=1,n.iter=n.iter,n.burnin=n.burnin,n.thin=n.thin,bugs.directory=bd)

```

---

## Aplicación

a) Analisis previos

---

```
#Cargar la base de datos de actitudes hacia la Estadística
require(foreign)
wd="D:\\Users \\martin.malaspina\\Desktop"
setwd(wd)
actitud=read.spss("actitud.sav")
act=data.frame(actitud)
act

#Consistencia interna (alpha de Cronbach)
require(ltm)
cronbach.alpha(act)

require(psych)
alpha(act)

#Evaluacion de la Multidimensionalidad: Análisis factorial usando correlaciones
tetracóricas

require(psych)
c=tetrachoric(act)

fap2 <- fa(r=crho,nfactors=2,rotate="oblimin")
fap2

par(mfrow=c(1,2))
fa.diagram(fap2,cut=0.4)
factor.plot(fap2, cut=0.4,ylim=c(0,1),xlim=c(0,1))
abline(h=0.4,col=4,lty=2)
abline(v=0.4,col=4,lty=2)

descri<-paf(as.matrix(act))
```

---

b) Estimacion bayesiana de los parametros

# Modelo propuesto (M2PL)

---

```

model{
for (i in 1 : n) {
for (j in 1 : I) {
logit(p[i,j])<- m[i,j]
ma[i,j]<- a1[j]*theta1[i]+a2[j]*theta2[i] - d[j]
m[i,j]<- ma[i,j]*(1-step(abs(ma[i,j])-5))-5*step(-5-ma[i,j])+5*step(ma[i,j]-5)
y[i, j ]~dbern( p[i, j] ) }
}
#prioris para los parametros de los items y medidas adicionales
for (j in 1:I) {
d[j]~dnorm(0, 1)
a1[j]~dlnorm(1, 2)
a2[j]~dlnorm(1, 2)
mdific[j] <- -d[j]/mdisc[j]
a1quad[j] <- pow(a1[j],2)
a2quad[j] <- pow(a2[j],2)
mdisc[j] <- pow(a1quad[j] + a2quad[j] ,1/2)
a1star[j]<- a1[j]/mdisc[j]
a2star[j]<- a2[j]/mdisc[j]
}
#prioris para los rasgos latentes
for (i in 1:n) {
theta1[i]~dnorm(0,1)
theta2[i]~dnorm(0,1) }
}

```

---

# Algoritmo para la estimación bayesiana utilizando MCMC

---

```

y=as.matrix(act)
colnames(y) <- NULL

n=nrow(act)
I=ncol(act)

data<-list(y=y,n=n,I=I)

inits<-function(){list(a1=c(rep(1,I)),a2=c(rep(1,I)),
d=c(rep(0,I)),theta1=c(rep(0.5,n)),theta2=c(rep(0.5,n)))}

parameters<-c("a1", "a2", "d", "mdific", "mdisc", "a1star", "a2star", "theta1", "theta2")

n.iter=105000
n.burnin=5000
n.thin=50

#Usando WinBUGS

library(R2WinBUGS)
bd="C:/WinBUGS14/"

bayes2LMW<-bugs(data,inits=inits,parameters.to.save=parameters,
model.file="Logistic 2LM Modelc.txt",working.directory=wd,
n.chains=1,n.iter=n.iter,n.burnin=n.burnin,n.thin=n.thin,bugs.directory=bd)

colnames(bayes2LMW$sims.matrix)
plot(bayes2LMW)
print(bayes2LMW, digits=2)

#Gráficos de cajas

boxplot(bayes2LMW$sims.matrix[,1:20])
boxplot(bayes2LMW$sims.matrix[,21:40])
boxplot(bayes2LMW$sims.matrix[,41:60])
boxplot(bayes2LMW$sims.matrix[,61:80])
boxplot(bayes2LMW$sims.matrix[,81:100])
boxplot(bayes2LMW$sims.matrix[,101:120])
boxplot(bayes2LMW$sims.matrix[,121:140])

```

---

#Algoritmo para el análisis de convergencia

---

```
bayes2LMW2<-bugs(data,inits=inits,parameters.to.save=parameters,
model.file="Logistic 2LM Modelc.txt",working.directory=wd,codaPkg = TRUE,
n.chains=1,n.iter=n.iter,n.burnin=n.burnin,n.thin=n.thin,bugs.directory=bd)

library("coda")
codaobject <- read.bugs(bayes2LMW2)
summary(codaobject)
effectiveSize(codaobject)

#Parámetro a1
geweke.diag(codaobject[,c(1:20)])
plot(codaobject[,c(1:20)])
#densityplot(codaobject[,c(1:20)])
#xyplot(codaobject[,c(1:20)])
#acfplot(codaobject[,c(1:20)])

#Parámetro a2
geweke.diag(codaobject[,c(41:60)])
plot(codaobject[,c(41:60)])
#densityplot(codaobject[,c(41:60)])
#xyplot(codaobject[,c(41:60)])
#acfplot(codaobject[,c(41:60)])

#Parámetro d
geweke.diag(codaobject[,c(81:100)])
plot(codaobject[,c(81:100)])
#densityplot(codaobject[,c(81:100)])
#xyplot(codaobject[,c(81:100)])
#acfplot(codaobject[,c(81:100)])

#Para grabar lo estimado y el análisis de convergencia
save.image(file="D:\\Users\\martin.malaspina\\Desktop\\actitud\\TRI.Rdata")
```

---

## Apéndice B

### Resultados de convergencia

#### Criterio de Geweke

##### Para $a_1$

a1[1]	a1[2]	a1[3]	a1[4]	a1[5]	a1[6]	a1[7]	a1[8]	a1[9]	a1[10]
0.900	0.610	-0.467	0.763	-0.015	1.902	-1.045	-1.063	0.413	-0.293
a1[11]	a1[12]	a1[13]	a1[14]	a1[15]	a1[16]	a1[17]	a1[18]	a1[19]	a1[20]
-1.156	-0.795	1.704	-1.334	1.274	0.398	0.519	0.912	-0.308	-0.959

##### Para $a_2$

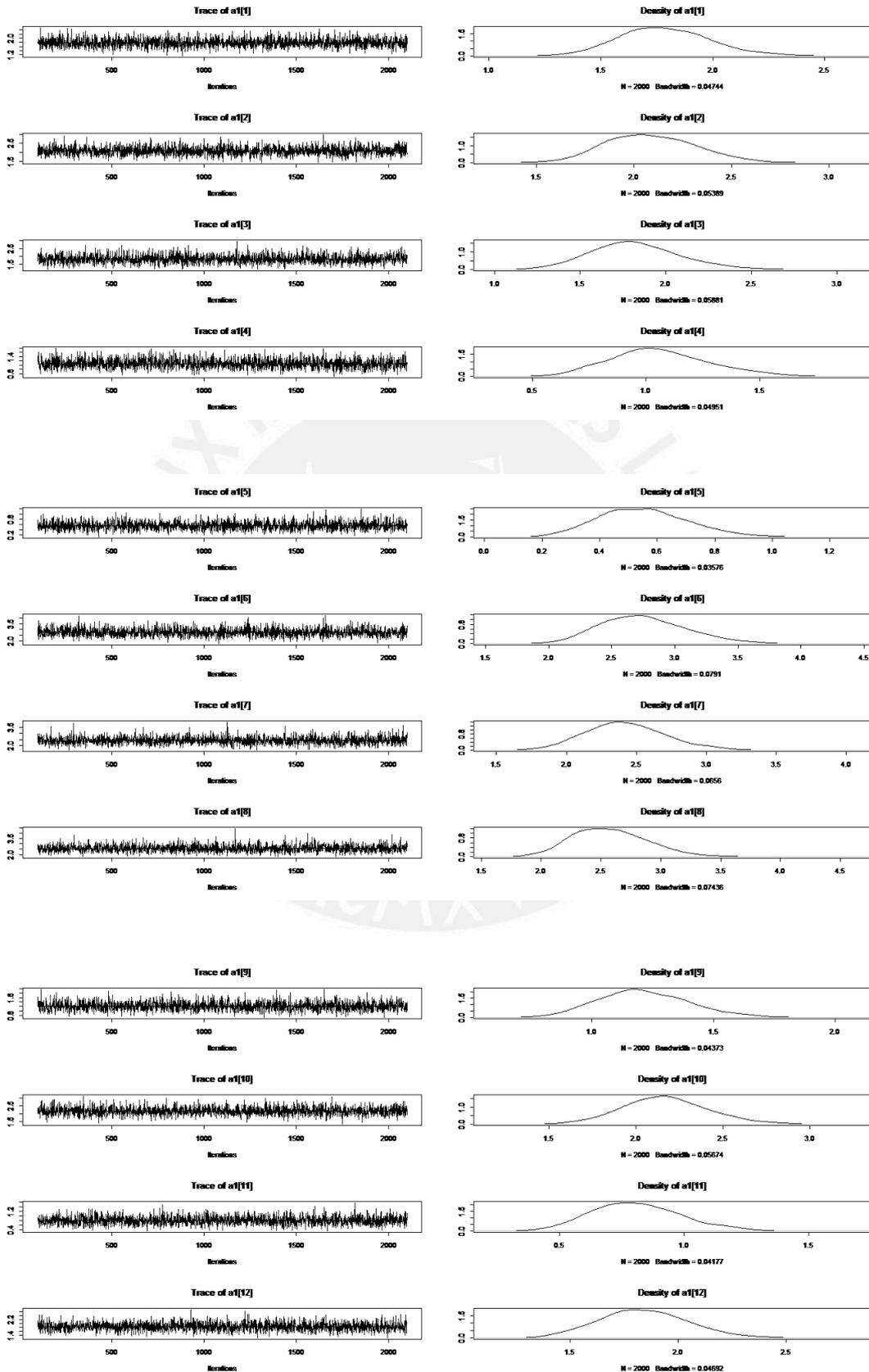
a2[1]	a2[2]	a2[3]	a2[4]	a2[5]	a2[6]	a2[7]	a2[8]	a2[9]	a2[10]
1.174	0.327	0.473	0.367	0.092	0.471	0.482	1.330	0.087	0.253
a2[11]	a2[12]	a2[13]	a2[14]	a2[15]	a2[16]	a2[17]	a2[18]	a2[19]	a2[20]
1.159	-0.338	0.347	-1.996	0.359	0.188	0.885	0.406	-0.334	-0.643

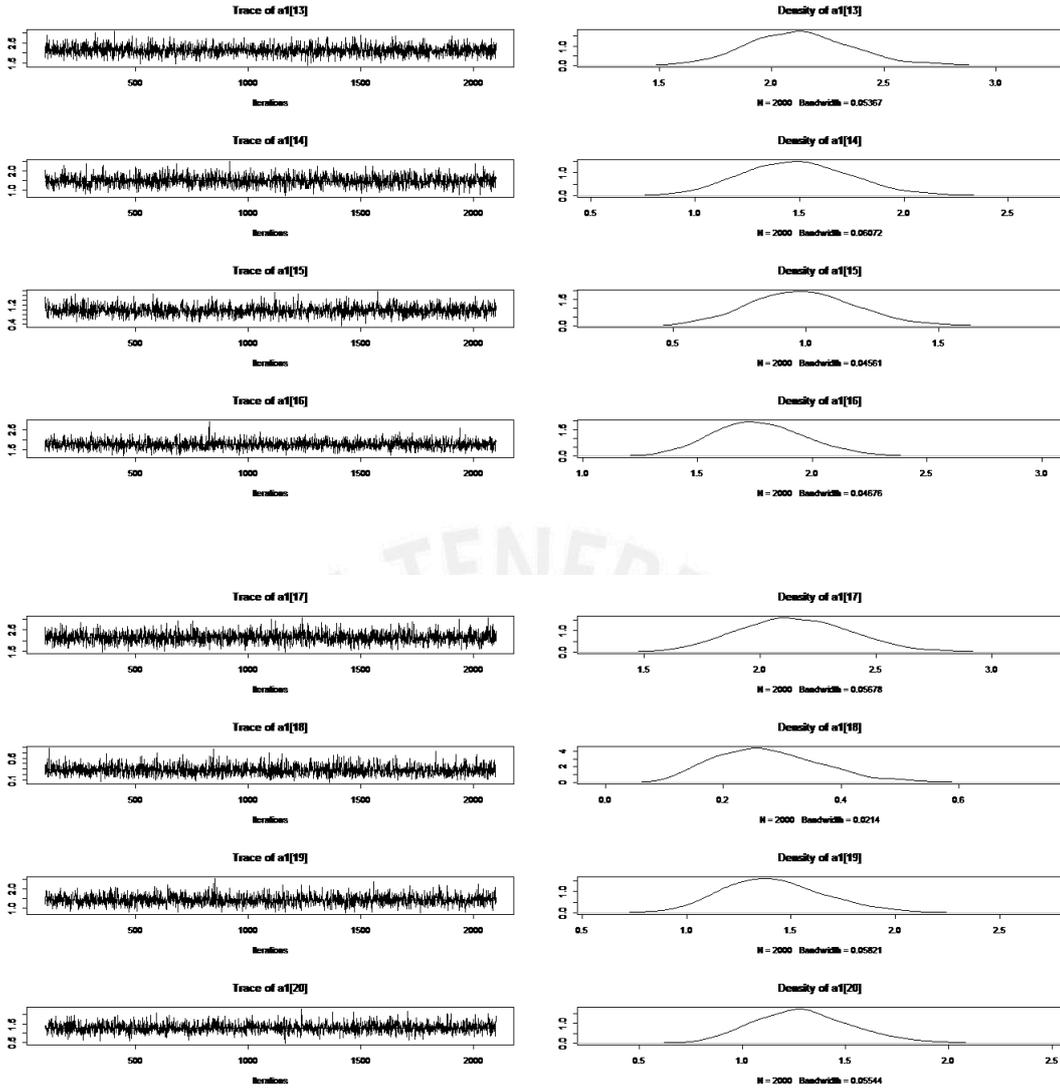
##### Para $d$

d[1]	d[2]	d[3]	d[4]	d[5]	d[6]	d[7]	d[8]	d[9]	d[10]
0.310	-0.835	-0.047	0.722	-0.498	-0.232	0.692	-1.203	0.464	-0.936
d[11]	d[12]	d[13]	d[14]	d[15]	d[16]	d[17]	d[18]	d[19]	d[20]
1.034	-0.447	-0.367	-0.008	0.761	-1.325	-1.197	0.679	0.279	0.921

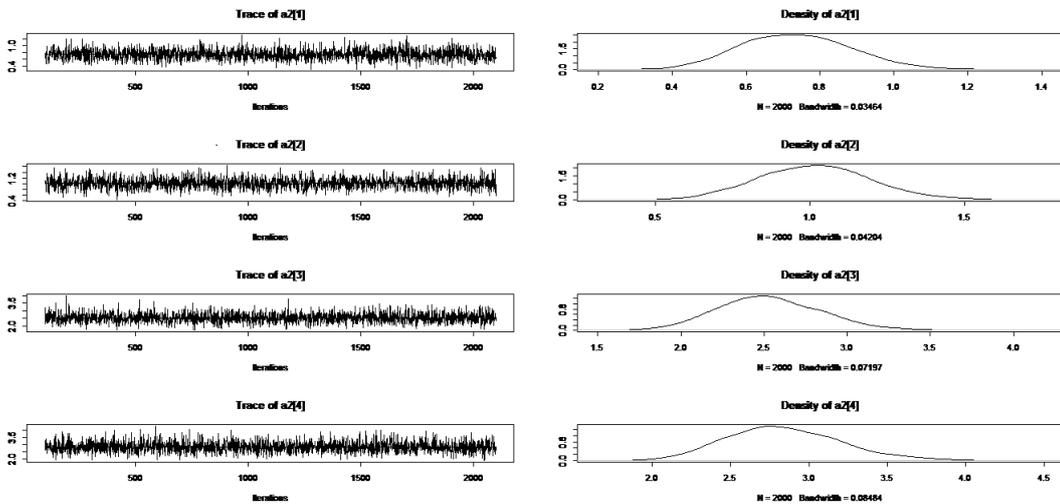
Gráficos

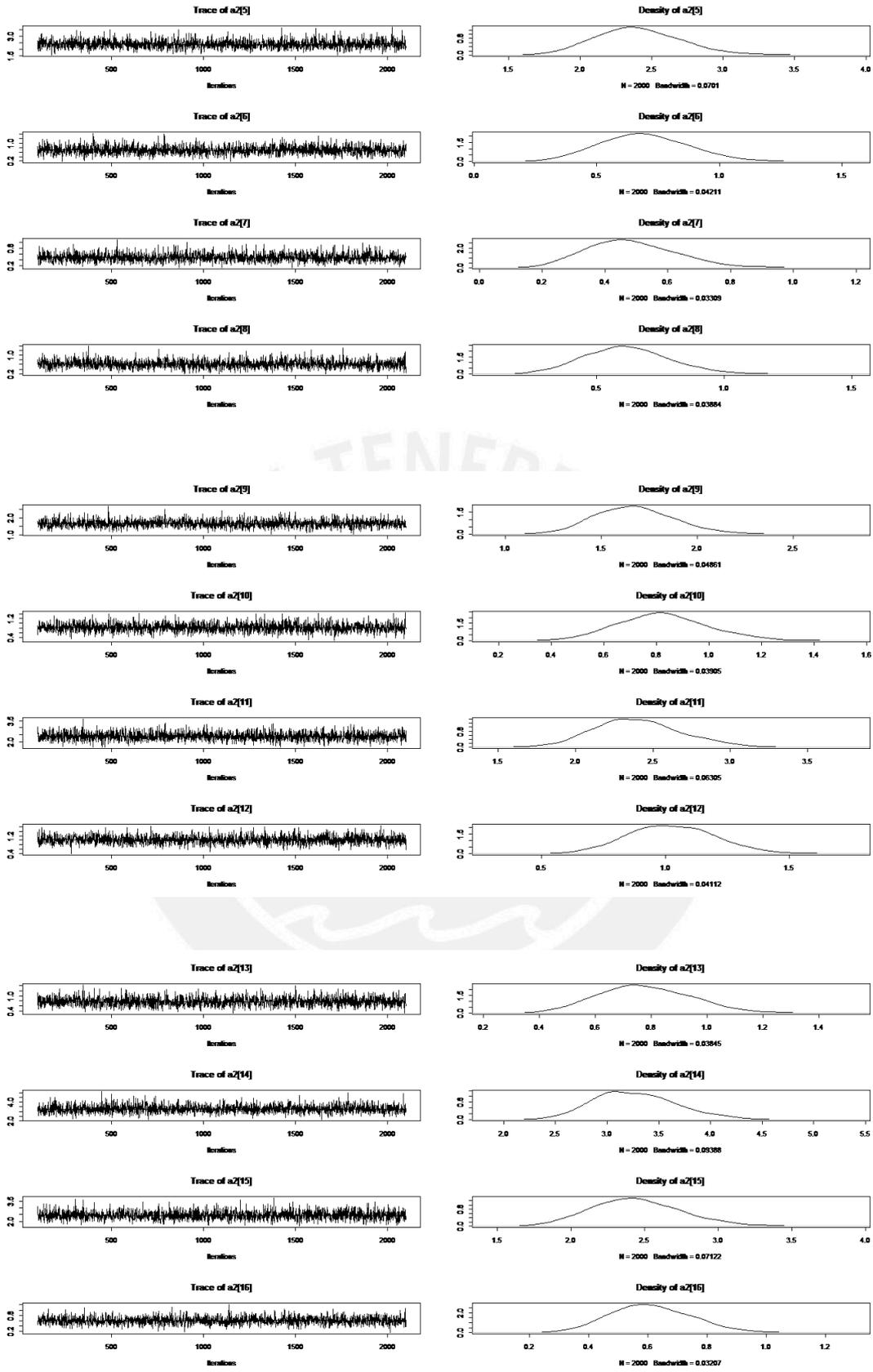
Para  $a_1$

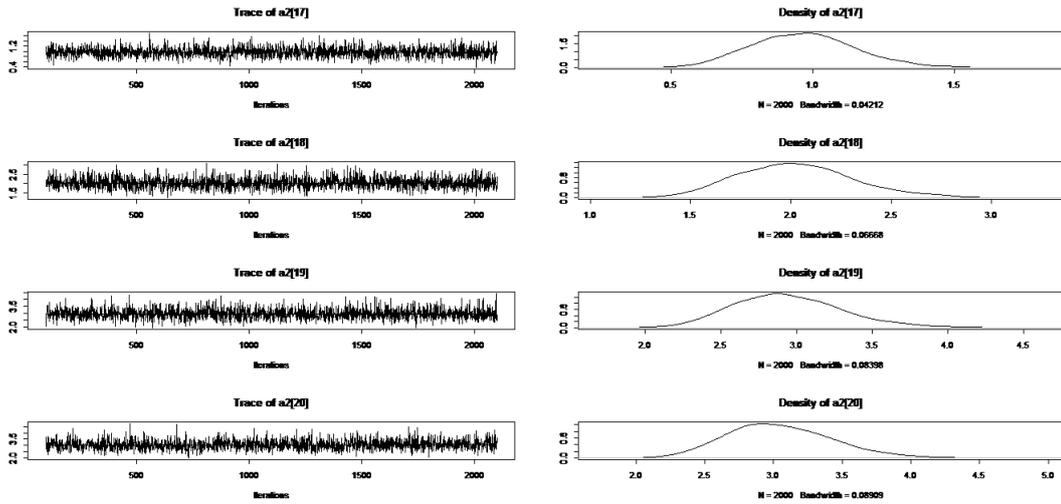




Para  $a_2$







Para d

