

*Distribuciones asimétricas para el trazo latente en  
modelos de teoría de respuesta al ítem con múltiples  
poblaciones*

SAIRA YANOHA DÍAZ PEDRAZA  
LIC. EN MATEMÁTICAS Y ESTADÍSTICA  
CÓDIGO: 2949675627



UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE ESTADÍSTICA  
BOGOTÁ, D.C.  
31 DE JULIO DE 2018

*Distribuciones asimétricas para el trazo latente en  
modelos de teoría de respuesta al ítem con múltiples  
poblaciones*

SAIRA YANOHA DÍAZ PEDRAZA  
LIC. EN MATEMÁTICAS Y ESTADÍSTICA  
CÓDIGO: 2949675627

DISERTACIÓN PRESENTADA PARA OPTAR AL TÍTULO DE MAGÍSTER EN  
CIENCIAS ESTADÍSTICA

DIRECTOR  
ÁLVARO MAURICIO MONTENEGRO DÍAZ, PH.D.

LÍNEA DE INVESTIGACIÓN  
TEORÍA DE RESPUESTA AL ÍTEM



UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE ESTADÍSTICA  
BOGOTÁ, D.C.  
31 DE JULIO DE 2018

## **Título en español**

Distribuciones asimétricas para el trazo latente en modelos de teoría de respuesta al ítem con múltiples poblaciones

## **Title in English**

Asymmetric distributions latent variables in item response theory models with multiple populations

**Resumen:** En este documento, se propone un modelo para analizar datos provenientes de una prueba cuya población esta dividida en múltiples subpoblaciones. Se presenta un modelo de teoría de respuesta al ítem TRI con distribución normal asimétrica no centrada para los trazos latentes para múltiples poblaciones. Se muestran los resultados de la aplicación del modelo propuesto a una muestra de datos reales provenientes del Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado (PERCE), que fue aplicada por el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación; cuyos datos fueron suministrados por el ICFES. Para evaluar la construcción del modelo se realizó un estudio de simulación; el cual mostró que el algoritmo de estimación se encuentra bien especificado dado que recupera los parámetros de los conjuntos de datos simulados. Se concluye mediante una prueba de bondad de ajuste que cuando se tiene en cuenta la asimetría de los trazos latentes el modelo se ajusta tanto a los parámetros de los ítems como a los trazos de los individuos. Por otro lado, la implementación se realizó haciendo de Stan (Carpenter et al., 2017) que es un lenguaje de programación probabilística que se ha convertido en la herramienta que se esta utilizando en la actualidad para modelos y cálculos estadísticos de alto rendimiento

**Abstract:** In this this thesis, it is proposed a model to analyze data coming from a test whose population is divided into multiple subpopulations. We introduce an item theory model (IRT) with asymmetric normal distribution for the latent traits. The model is applied to a real case data from the First International Comparative Study on Language, Mathematics and Factors Associates in the third and fourth grades (PERCE), which was applied by Latin American Laboratory for Evaluation of the Quality of Education. Data data were supplied by Instituto Colombiano de Evaluacion de la Educación (ICFES). To evaluate the algorithm is well specified since it retrieves the parameters of the simulated data sets. We compare the proposed model with multi-group model (MGM), Bock and Zimoski, (1997). The results show that when the asymmetry of the latent traits is taken into account the proposed model has better fit for both the parameters of the items and the latent traits of the individuals. The implementation was done by using Stan (Carpenter et al., 2017), which is a probabilistic programming language that has become the tool that is currently being used for high performance statistical models and calculations.

**Palabras clave:** TRI, trazo latente, modelos asimétricos, múltiples grupos, Stan

**Keywords:** IRT, latent variable, asymmetric models, multiple groups, Stan

# Nota de aceptación

Trabajo de tesis

---

Jurado

---

Jurado

---

Jurado

---

Director  
Álvaro Mauricio Montenegro Díaz

Bogotá, D.C., 31 07 de 2018

---

---

## Dedicado a

---

---

A mi familia: mis padres Miriam y Norberto, mis hermanos Dilan y David y a mi esposo Carlos y mi hijo Christopher.

---

---

## Agradecimientos

---

---

Quisiera iniciar agradeciéndole a Dios por permitirme cumplir esta meta a pesar de las dificultades. A mis padres Miriam y Norberto por apoyarme y darme palabras de aliento. A mi esposo Carlos y a mi hijo Christopher porque sin ayuda no hubiese sido posible y por comprender tantos momentos de ausencia. Finalmente quisiera agradecerle a mi director de tesis, el profesor Álvaro Mauricio Montenegro Díaz, por su apoyo, su comprensión y sus conocimientos.

---

---

# Índice general

---

---

|  |            |
|--|------------|
| <b>Índice general</b>  | <b>I</b>   |
| <b>Índice de tablas</b>  | <b>III</b> |
| <b>Índice de figuras</b>   | <b>IV</b>  |
| <b>Introducción</b>  | <b>V</b>   |
| <b>1. Marco teórico</b>  | <b>1</b>   |
| 1.1. Distribución Normal asimétrica . . . . .  | 1          |
| 1.2. Modelos clásicos de la TRI . . . . .  | 2          |
| 1.3. Modelo Normal asimétrico . . . . .  | 4          |
| 1.4. Modelo MGM . . . . .  | 5          |
| <b>2. Modelo TRIAM</b>   | <b>7</b>   |
| 2.1. Función de probabilidad del modelo TRIAM . . . . .                                      | 7          |
| 2.2. Especificación de las distribuciones a priori . . . . .                                 | 8          |
| 2.3. Distribución posterior . . . . .  | 9          |
| <b>3. Estimación de los parámetros del modelo TRIAM</b>                                      | <b>10</b>  |
| <b>4. Aplicación a datos reales</b>  | <b>17</b>  |
| 4.1. Aplicación del modelo TRIAM con distribución asimétrica para el trazo latente . . . . . | 17         |
| 4.2. Estimación de los parámetros del modelo TRIAM . . . . .                                 | 20         |
| 4.3. Ajuste del modelo . . . . .   | 25         |
| <b>5. Discusión</b>  | <b>27</b>  |

---

|   |    |
|---|----|
| Conclusiones  | 30 |
| Trabajos futuros  | 32 |
| Apéndice A. Introducción a Stan   | 33 |
| Apéndice B. Resultados simulaciones   | 39 |
| Apéndice C. Resultados Estimaciones individuales                                    | 42 |
| Apéndice D. Código modelo TRIAM en Rstan  | 45 |
| Apéndice E. Resumen estadísticas a posteriori de estimaciones modelo TRIAM en Rstan | 49 |
| Bibliografía  | 53 |

---

---

## Índice de tablas

---

---

|  |    |
|--|----|
| 3.1. Matriz de datos subpoblación 1 . . . . .  | 13 |
| 3.2. Matriz de datos subpoblación 2 . . . . .  | 13 |
| 4.1. Correlaciones, error absoluto y ECM valores estimados vs simulados . . . . .  | 18 |
| 4.2. Estimaciones para los parámetros de localización $\mu$ , asimetría $\gamma$ y $d$ de Colombia modelo TRIAM . . . . .                                | 21 |
| 4.3. Estimaciones para los parámetros de localización $\mu$ , asimetría $\gamma$ y $d$ de Argentina modelo TRIAM . . . . .                               | 22 |
| 4.4. Estimaciones para los parámetros de localización $\mu$ , asimetría $\gamma$ y $d$ de Venezuela modelo TRIAM . . . . .                               | 22 |
| 4.5. Estimaciones para los parámetros de localización $\mu$ , asimetría $\gamma$ y $d$ de Cuba modelo TRIAM . . . . .                                    | 22 |
| 4.6. Estimaciones para la medias, las varianzas y las modas para cada subpoblación, Colombia=1, Argentina=2, Venezuela=3 y Cuba=4, modelo TRIAM. . . . . | 23 |
| 4.7. Correlaciones entre los trazos latentes estimados y el puntaje clásico . . . . .  | 24 |
| 4.8. Resultados de la prueba de bondad de ajuste para los items modelo TRIAM p-valores Bayesianos . . . . .  | 26 |
| 4.9. Resultados de la prueba de bondad de ajuste para los items modelo MGM p-valores Bayesianos . . . . .  | 26 |
| 1. Estimaciones individuales parámetro de discriminación . . . . .   | 43 |
| 2. Estimaciones individuales parámetro de dificultad . . . . .   | 44 |

---

---

## Índice de figuras

---

---

|  |    |
|--|----|
| 1.1. Curva normal asimétrica para diferentes valores del parámetro de localización $\mu$ y del parámetro de asimetría $\gamma$ . . . . .                                 | 2  |
| 3.1. Muestra de las cadenas estimadas trazos latentes . . . . .  | 16 |
| 4.1. Gráficos de dispersión de los valores simulados versus los valores estimados con el modelo TRIAM . . . . .  | 19 |
| 4.2. Densidades trazos latentes estimados individualmente modelo TRI asimétrico . . . . .  | 20 |
| 4.3. Densidades trazos latentes estimados de Colombia, Venezuela, Argentina y Cuba con el modelo TRIAM . . . . .   | 23 |
| 4.4. Puntaje clásico vs. trazos latentes estimados. Arriba a la izquierda Colombia, a la derecha Venezuela, abajo a la izquierda Argentina y a la derecha Cuba . . . . . | 24 |
| 4.5. Densidades trazos latentes estimados de Colombia, Argentina, Venezuela y Cuba con el modelo MGM . . . . .   | 25 |
| 1. Diagnóstico gráfico de convergencia . . . . .   | 38 |

---

---

## Introducción

---

---

En este trabajo se estudia un modelo Bayesiano de teoría de respuesta al ítem para múltiples poblaciones con distribución asimétrica para los trazos latentes de cada subpoblación; el modelo se denominará TRIAM. En el modelo TRIAM se asume que los individuos están agrupados en subpoblaciones las cuales son generadas por distribuciones asimétricas y se hace uso de la distribución “*normal asimétrica no centrada*” para los trazos latentes.

Es usual que en los modelos TRI se asuma que los trazos o rasgos latentes son variables aleatorias que siguen una distribución normal estándar (ver por ejemplo, Albert, 1992; Patz y Junker 1999); pero aceptar que los trazos latentes se distribuyen simétricamente cuando hay ausencia de simetría implica obtener estimaciones poco confiables o sesgadas (Azevedo, Bolfarine, Andrade, 2011). Por lo anterior, el supuesto de normalidad en los modelos TRI ha sido ampliamente cuestionado, principalmente cuando se estudian variables como habilidades académicas y otras variables latentes referentes al comportamiento humano (Micceri 1989; Samejima 1997,2000).

Uno de los primeros autores en evidenciar que el supuesto de simetría es cuestionable fue Micceri (1989), quien argumenta que en muchos conjuntos de datos psicométricos el supuesto de normalidad para los trazos latentes solo es asumido por conveniencia. Posteriormente, Seong (1990) demuestra que en los modelos dicotómicos una especificación incorrecta de la distribución de las variables latentes genera errores de estimación en los parámetros de los ítems y en los trazos de los individuos. De igual forma Azevedo, Andrade y Fox (2009a,2009b) muestran grupos de datos en los cuales la normalidad y simetría de los trazos latentes no se deberían asumir. En este sentido, investigadores como Bazan, Branco y Bolfarine (2006, 2014), Azevedo, Bolfarine y Andrade (2011), entre otros, han propuesto estrategias para introducir distribuciones asimétricas tanto para el trazo latente como para la curva característica del ítem. Estos modelos hacen uso de la distribución normal asimétrica y normal asimétrica centrada respectivamente. Por ejemplo, Bazán, Branco y Bolfarine (2006) proponen un modelo de teoría de respuesta al ítem asimétrico, denominado “*skew-normal ogive model*”. En este modelo se considera la función de distribución acumulada de la distribución normal asimétrica propuesta por Azzalini (1985) para la curva característica del ítem (CCI). Los mismos autores en (2014) proponen una familia de modelos extendiendo el modelo planteado por ellos mismos en 2006, considerando la distribución normal asimétrica (Azzalini, 1985) para la CCI y para el trazo latente. Otro tipo de modelos TRI en los que se ha incluido la asimetría son los modelos para múltiples poblaciones, por ejemplo, Santos Azevedo y Bolfarine (2013) plantean un modelo de

---

TRI para múltiples grupos usando la distribución normal asimétrica con parametrización centrada para los trazos latentes bajo un marco Bayesiano.

En los modelos TRI multigrupos se tiene el modelo propuesto por Bock y Zimoski (1997) denominado modelo TRI multigrupo MGM (por sus siglas en inglés multi-group model). Este modelo se consolida como un modelo de referencia para el análisis de datos provenientes de individuos organizados en subpoblaciones; el objetivo principal es estimar conjuntamente los parámetros del ítem y la distribución del trazo latente de los individuos en cada una de las poblaciones. En el MGM, se asume que los trazos latentes son variables aleatorias que siguen distribuciones normales simétricas diferentes en cada población. En este sentido, autores como Kirisci L., Tse-chi, Yu L. (2001), resaltan la importancia de usar este tipo de modelos, argumentando que cuando no se tiene en cuenta la estructura de subpoblaciones se incluye sesgo en las estimaciones. Otros autores que han desarrollado modelos TRI para múltiples grupos son: Azevedo, Andrade y Fox (2012) que proponen un modelo de teoría de respuesta al ítem Bayesiano generalizado de TRI para múltiples grupos, en este modelo el objetivo principal es explorar las ventajas de un procedimiento de estimación MCMC y herramientas Bayesianas de ajuste del modelo MGM de Bock y Zimoski (1997); Santos, Azevedo y Bolfarine (2013) plantean modelos de teoría de respuesta al ítem con distribución normal asimétrica centrada para el trazo latente y muestran que el modelo con la distribución asimétrica centrada se ajusta mejor a los datos que el modelo TRI con la distribución normal estándar.

En el modelo desarrollado en este trabajo se usa la distribución normal asimétrica no centrada para los trazos latentes y se estima un modelo para múltiples poblaciones. Se sigue una estrategia similar a la usada por Santos, Azevedo y Bolfarine (2013); ellos estiman un modelo multigrupos usando como distribución de los trazos latentes la distribución normal asimétrica centrada, para identificabilidad del modelo y realizar comparaciones entre los trazos latentes. En el modelo TRIAM se asumen que los trazos latentes siguen una distribución normal asimétrica y se sigue una estrategia de estimación Bayesina. Para implementar el modelo TRI se hace uso de Stan (Carpenter et al., 2017), que es un lenguaje de programación probabilística de estadística Bayesiana que implementa un eficiente algoritmo Hamiltoniano Monte Carlo (HMC; Neal, 2011) y el muestreador no-U-turn (NUTS; Hoffman y Gelman, 2014). Un lenguaje de programación probabilística es un lenguaje de programación diseñado para describir modelos probabilísticos y luego realizar inferencias; está fuertemente relacionado a modelos gráficos y a redes Bayesianas, pero es más flexible.

Stan recibe su nombre en honor a Stanislaw Ulam, pionero del método Monte Carlo; surge de la necesidad de aplicar inferencia Bayesiana a modelos como los lineales generalizados multinivel discutidos en Gelman y Hill (2007). Al usar algoritmos como Jags, Bugs, entre otros o al intentar realizar extensiones del muestreador de Gibbs, encontraron dificultades como no convergencia de las cadenas luego de más de cien mil iteraciones. Algunas de las ventajas de Stan sobre otros algoritmos de estadística Bayesiana son el manejo de grandes volúmenes de datos, la facilidad para la implementación de los modelos, la convergencia de las cadenas, los tiempos de estimación, entre otros.

Entre otras ventajas Stan permite el uso de prior impropias, además, debido a que el HMC explora el espacio de los parámetros más eficientemente que el muestreador de Gibbs y que el algoritmo Metropolis (Gelman, Carlin, Stern, & Rubin, 2014, p. 300), es considerado más eficiente que otros software para estadística Bayesiana como Bugs, Winbugs o Jags

(Luo Y., Jiao H., 2017). Por otro lado, los modelos escritos en Stan son compilados a  $C_{++}$  lo que hace que sea más rápido y permite la traducción a otros lenguajes de programación. Otra de las ventajas de Stan es que cuenta con interfaces para los lenguajes del análisis de datos más comunes como los son: R, Python, Shell, MATLAB, Julia, Mathematica, Scala y Stata; además, se ejecuta en los sistemas operativos más usados: Linux, Windows y Mac.

El modelo desarrollado surge del análisis de los datos del Primer Estudio Internacional Comparativo sobre Lenguaje, Matemáticas y Factores Asociados en Tercero y Cuarto grado (PERCE) aplicada por el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (Casassus & Cusato, 1998). Esta prueba se implementó con el fin de generar recomendaciones para la ejecución de políticas públicas con miras a la calidad educativa en cada uno de los 13 países participantes (Casassus & Cusato, 1998, pp.7). La prueba se diseñó con intención de medir habilidades referentes a matemáticas y lenguaje en niños de tercer y cuarto grado. Se diseñaron dos cuestionarios tipificados como A y B (de los cuales cada estudiante respondió uno al azar), cada uno con el mismo número de ítems pero con ítems diferentes. La prueba de matemáticas consta de 32 ítems y la de lenguaje consta de 19. En el diseño de la prueba se tuvo en cuenta los currículos de los países participantes y fue aprobada en consenso por los coordinadores de educación de cada país.

Casassus y Cusato mencionan en el informe técnico de la prueba PERCE publicado en 2001 que al menos uno de los países participantes (Cuba) muestra un comportamiento atípico comparado con el desempeño de los demás países; este comportamiento se evidencia también en investigaciones posteriores, por ejemplo, Antonio y Montenegro (2014) mencionan que los trazos latentes de Cuba evidencian asimetría. Es por esto, que se surgió la necesidad de diseñar un modelo que estimará los trazos latentes teniendo en cuenta la asimetría y las diferencias entre las poblaciones participantes en la prueba. Es importante resaltar que utilizar el modelo adecuado en campos como la educación, se evidencia, por ejemplo, cuando los resultados de las estimaciones tienen repercusiones en las políticas públicas de no sólo de un país sino de una región, como es el caso de la prueba PERCE. En ese estudio se usó un modelo de Rasch para la estimación de los trazos latentes (Casassus & Cusato, 1998), posteriormente autores como Antonio y Montenegro (2014) cuestionan el uso de este modelo, argumentan que es incorrecto usarlo desde el punto de vista teórico.

Por otro lado, el modelo TRIAM es aplicable en general a situaciones en las que existan múltiples poblaciones y se quiera medir variables que se distribuyan asimétricamente; estos datos son usuales en estudios realizados en áreas como psicología, educación, medicina, entre otros; en las cuales los individuos se encuentran agrupados por características comunes como género, nivel educativo, edad, patología, entre otros. Dentro de las variables latentes que pueden ser asimétricamente distribuidas, se pueden mencionar la ansiedad (Zaider et al., 2001), la depresión (Riddle, Blais y Hess, 2002), las habilidades escolares (Hashimoto, 2002), entre otras.

Este trabajo está organizado como sigue: en el capítulo 1 se presenta una revisión de literatura. En el capítulo 2 se presenta el modelo y la especificación Bayesiana. En el capítulo 3 se hace una introducción a la estimación en Stan mediante un ejemplo; en

---

el capítulo 4 se muestra el código en Stan para la estimación de los trazos latentes y los parámetros de los items. En el capítulo 5 se realiza un estudio de simulación y en el capítulo 6 se muestra una aplicación a datos reales. Finalmente se presentan algunas conclusiones y se realizan algunas sugerencias para investigaciones futuras.

# CAPÍTULO 1

---

---

## Marco teórico

---

---

### 1.1. Distribución Normal asimétrica

Propuesta por Azzalini (1985), enuncia que si una V.A  $X$  sigue una distribución normal asimétrica. Su función de densidad está dada por:

$$f_{(\gamma)}(x) = \frac{2}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left(\gamma \left(\frac{x - \mu}{\sigma}\right)\right), \quad (1.1)$$

donde:

- $\phi(\cdot)$  y  $\Phi(\cdot)$ : denotan, respectivamente la densidad y la función de distribución de la distribución normal estándar.
- $\mu$ : parámetro de localización.
- $\sigma$ : parámetro de escala.
- $\gamma$ : parámetro de forma o de asimetría.

Una parametrización alternativa del parámetro de asimetría está dada por:

$$d = \frac{\gamma}{(1 + \gamma^2)^{1/2}}, \quad (1.2)$$

donde  $-1 \leq d \leq 1$ . La media y la varianza están dadas por:  $E(X) = \mu + \sigma d \sqrt{\frac{2}{\pi}}$ ,  $V(X) = \sigma^2(1 - \frac{2}{\pi}d^2)$ .

La V.A  $Z = \frac{(X-\mu)}{\sigma}$  sigue una distribución normal estándar asimétrica, con función de densidad dada por:

$$f_{\gamma}(z) = 2\phi(z)\Phi(\gamma z), \quad (1.3)$$

donde:

- $\phi(\cdot)$  y  $\Phi(\cdot)$ : denotan, respectivamente la densidad y la función de distribución de la distribución normal estándar.

y función de distribución acumulada dada por,

$$F_{\gamma}(z) = \int_{-\infty}^z 2\phi(t)\Phi(\gamma t)dt \quad (1.4)$$

En la gráfica 1.1 se muestran algunas densidades de la distribución normal asimétrica para diferentes valores del parámetro de localización  $\mu$  y del parámetro de asimetría  $\gamma$

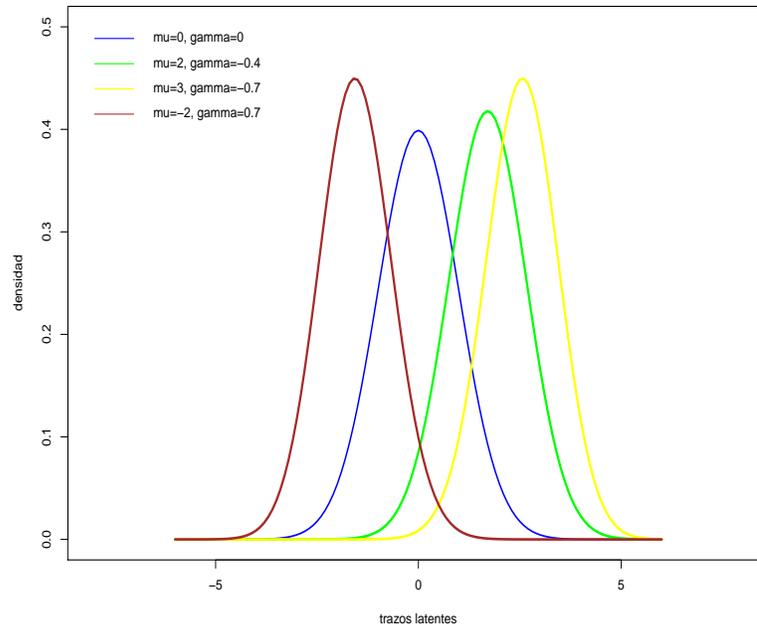


FIGURA 1.1. Curva normal asimétrica para diferentes valores del parámetro de localización  $\mu$  y del parámetro de asimetría  $\gamma$

## 1.2. Modelos clásicos de la TRI

Las teorías de los test se centran en establecer la relación entre la habilidad de un individuo (variable latente) y su puntuación observada en un test. Tradicionalmente la medición de los trazos se venía realizando haciendo uso de la teoría clásica de los test (TCT), cuya limitación principal es que sus estimaciones dependen de los individuos y de los instrumentos. En la actualidad, se hace uso de la teoría de respuesta al ítem (TRI), la cual brinda soluciones a las principales limitaciones de la TCT al conseguir medidas invariantes respecto de los sujetos medidos y de los instrumentos utilizados (Muñiz, 2010).

Los modelos propuestos en la TRI se pueden clasificar según cuatro características principales, como lo son:

1. Naturaleza de los ítems: dicotómicos o politómico.
2. Número de poblaciones: una o más poblaciones.

3. Dimensión del trazo latente: unidimensional o multidimensional.
4. Distribución del trazo latente: simétrico o asimétrico.

### Modelo TRI Unidimensional

Un modelo TRI es unidimensional (TRIU) cuando se asume que la respuesta de un individuo a un ítem resulta de la interacción de los parámetros del ítem y un único trazo  $\theta$  (trazo latente del individuo) para cada respondiente. El modelo TRIU se escribe de manera general (Reckase 2009, pp. 12):

$$P(Y = y | \theta) = h(\theta, \xi_j, y), \quad (1.5)$$

donde  $Y$  corresponde a las respuestas de un individuo a los ítems de la prueba,  $y$  es la respuesta dicotómica (1 si es correcta y en caso contrario 0),  $h$  representa la función que describe la relación entre los parámetros del modelo y la probabilidad de respuesta del individuo,  $\xi_j$  corresponde al vector de parámetros de los ítems de la prueba.

La forma general del modelo TRIU de tres parámetros con enlace logístico, está dada por:

$$p_j(\theta_i) = P(Y_{ij} = 1 | \theta_i, a_j, b_j) = c_j + \frac{(1 - c_j)}{1 + e^{-a_j(\theta_i - b_j)}}, \quad (1.6)$$

con  $i = 1, \dots, I$  y  $j = 1, \dots, J$ , donde:

- $Y_{ij}$  es la respuesta de la persona  $i$  al ítem  $j$ .
- $Y_{ij} = 1$  si el individuo  $i$  responde correctamente al ítem  $j$ .
- $\theta_i$  es el parámetro del trazo latente unidimensional para la persona  $i$ .
- $a_j$  parámetro de discriminación del ítem  $j$ .
- $b_j$  es el parámetro de dificultad o de posición del ítem  $j$ .
- $c_j$  es la probabilidad de respuesta correcta por azar.

En los modelos TRI la curva característica del ítem - CCI, es una representación gráfica de la probabilidad de respuesta correcta a un ítem con parámetros de discriminación, dificultad y acierto casual, de acuerdo al valor del trazo latente. En el modelo definido por la ecuación 1.6 cuando  $c_j$  y  $a_j$  son constantes para todo  $j$ , con  $j = 1, \dots, J$ , en particular  $c_j = 0$  y  $a_j = 1$  se tiene un modelo TRI de un parámetro del ítem. Cuando  $c_j = 0$  para todo  $j$ , con  $j = 1, \dots, J$ , se tiene un modelo TRI 2P (de 2 parámetros), aquí el parámetro de dificultad corresponde al valor del trazo latente que tiene exactamente la probabilidad del 50% de contestar correctamente y el parámetro de discriminación se relaciona con la inclinación de la CCI. Se dice que un ítem discrimina eficientemente si al haber un pequeño cambio en el valor de un trazo latente con referencia a otro (aumenta o disminuye) la probabilidad de respuesta correspondiente cambia significativamente.

### 1.3. Modelo Normal asimétrico

Bazán, Branco y Bolfarine (2006) proponen un modelo TRI con distribución asimétrica para la curva característica denominado normal asimétrico TRI SNO, definido como sigue:

$$Y_{ij} \mid \theta_i, \eta_j \sim \text{Bern}(p_{ij}), \quad (1.7)$$

$$p_{ij} = F_{d_j}(m_{ij}), \quad (1.8)$$

$$m_{ij} = \alpha_j(\theta_i - \beta_j), \quad (1.9)$$

con,  $\alpha_j > 0$ ,  $-\infty < \beta_j < \infty$ ,  $-\infty < \theta_i < \infty$  y  $d_j = \frac{\gamma_j}{\sqrt{1 + \gamma_j^2}}$ , por lo tanto,  $-1 < d_j < 1$ ,

donde:

- $F_{d_j}(\cdot)$ : denota la función de distribución acumulada de una distribución normal estándar asimétrica.
- $\gamma_j$ : parámetro de asimetría del ítem  $j$ .
- $\eta_j = (\alpha_j, \beta_j)$ : conjunto de parámetros del ítem  $j$ .
- $\theta_i$ : trazo latente del individuo  $i$ .

Posteriormente, Bazán, Branco y Bolfarine (2014) extienden su modelo TRI SNO considerando distribuciones asimétricas para el trazo ó variable latente, nuevamente hacen uso de la distribución normal asimétrica. Indican que desde una perspectiva Bayesiana la especificación de una distribución para el trazo latente  $\theta_i$ , es equivalente a la especificación de una distribución a priori. A continuación se especifica la a priori para los trazos latentes:

$$\theta_i \sim SN(\mu, \sigma^2, \gamma), i = 1, \dots, n$$

con  $-\infty < \mu < \infty$ ,  $\sigma^2 > 0$  y  $-\infty < \gamma < \infty$ .

Para la construcción de la verosimilitud los autores tomaron como referencia la verosimilitud de su modelo inicial TRI SNO propuesto en 2006, de esta forma enuncian la verosimilitud completa para el modelo TRI SNO, con  $D = (Z^*, V, y)$ , como:

$$L(\mu, \eta, \lambda \mid D) = \prod_{i=1}^n \prod_{j=1}^k \phi^*(Z_{ij}^*) I(Z^*, Y_{ij}) \phi(V_{ij} I(V_{ij} > 0)), \quad (1.10)$$

donde  $\phi^*(\cdot)$  denota la función densidad de probabilidad de la distribución normal con media  $m_{ij} - d_j v_{ij}$ , varianza  $1 - d_j^2$  y  $I(Z_{ij}^*, Y_{ij}) = I(Z_{ij}^* > 0) I(Y_{ij} = 1) + I(Z_{ij}^* \leq 0) I(Y_{ij} = 0)$ .

Además,  $Z_{ij}^* = m_{ij} + e_{ij}$ ,  $m_{ij} = \alpha_j(\theta_i - \beta_j)$  y  $e_{ij} = -d_j V_{ij} - (1 - d_j^2)^{1/2} W_{ij}$ , con  $W_{ij} \sim N(0, 1)$  y  $V_{ij} \sim HN(0, 1)$  (la distribución half-normal o normal recortada).

Finalmente, Bazán, Branco, Bolfarine (2014) proponen una familia de modelos TRI asimétricos teniendo en cuenta el modelo TRI SNO propuesto por ellos mismos en 2006, el cual implementan considerando 4 escenarios:

- Distribución “skew-normal” para la curva característica y para el trazo latente, llamado modelo “Skew-Probit-Skew-Normal” (SPSN).
- Distribución “skew-normal” para la curva característica y normal para el trazo latente, llamado modelo “Skew-Probit-Normal” (SPN).
- Distribución normal para la curva característica y skew-normal para el trazo latente, llamado modelo “Probit-Skew-Normal” (PSN).
- Distribución normal para la curva característica y normal para el trazo latente, llamado modelo “Probit-Norma” (PN).

## 1.4. Modelo MGM

Bock y Zimowski (1997) plantean un modelo TRI para múltiples grupos que lo denominan MGM por sus siglas en inglés (multi-group model), con el cual brindan una estrategia para analizar datos provenientes de conjuntos numéricos organizados en subpoblaciones, en el modelo se da importancia a la selección de los grupos de individuos, de manera que se puede definir la distribución del trazo latente de un grupo específico independiente de los demás. En este modelo se asume que los grupos son independientes entre si y que cada individuo pertenece a un único grupo. Además, se esperan comportamientos similares para los individuos dentro de cada grupo y diferencias significativas entre los grupos (Azevedo, Andrade y Fox, 2012). En el modelo MGM se pueden tener diferentes grupos de individuos respondiendo la misma prueba o pruebas diferentes con ítems comunes.

Los autores hicieron uso de la distribución normal simétrica para los trazos latentes, además, fijaron la distribución del trazo latente de un grupo referencia (usualmente con media 0 y varianza 1) con el fin de que el modelo sea identificable.

Sea  $Y_{ijk}$  la respuesta del individuo  $j$ , del grupo  $k$  al ítem  $i$  (1 si la respuesta es correcta 0 en caso contrario),  $j = 1, 2, \dots, n_k$ ,  $i = 1, \dots, I$  y  $k = 1, \dots, K$ .

$$\begin{aligned}
 Y_{ijk} \mid (\theta_{jk}, \zeta_i) &\sim \text{Bernoulli}(P_{ijk}) \\
 P_{ijk} = P(Y_{ijk}=1 \mid \theta_{jk}, \zeta_i) &= \Phi(a_i \theta_{jk} - b_i) \\
 \theta_{jk} \mid \eta_{\theta_k} &\sim N(\mu_{\theta_k}, \psi_{\theta_k}),
 \end{aligned}$$

donde:

- $\theta_{jk}$  : trazo latente del individuo  $j$  del grupo  $k$
- Usualmente  $\theta_{jk} \mid (\mu_{\theta_k}, \psi_{\theta_k}) \sim^{i.i.d} N(\mu_{\theta_k}, \psi_{\theta_k})$ , donde  $N(\cdot, \cdot)$  representa una distribución normal con  $\varepsilon(\theta) = \mu_{\theta_k}$ ,  $\nu(\theta_k) = \psi_{\theta_k}$  (0 y 1 para el grupo de referencia, respectivamente, para la identificación del modelo).

- 
- $\Phi(\cdot)$ : representa la función de distribución de una normal.
  - $\zeta_i$ : vector de parámetros del grupo  $i$ .
  - $a_i$ : parámetro de discriminación del ítem  $i$ .
  - $b_i$ : parámetro de dificultad del ítem  $i$ .
  - $\eta_{\theta_k}$ : vector de parámetros de la población  $k$ .

---

---

## Modelo TRIAM

---

---

En este trabajo se propone un modelo de teoría de respuesta al ítem, con distribución asimétrica para el trazo latente para múltiples poblaciones, que se denominará TRIAM. Este modelo fue pensado para ser aplicado a datos provenientes de test contestados por poblaciones subdivididas en grupos de individuos.

En el modelo TRIAM se asume que:

1. La población consta de  $N$  individuos divididos en  $G$  grupos, cada grupo de tamaño  $n_g$ , con  $g = 1, \dots, G$ . Además  $N = \sum_{g=1}^G n_g$ .
2. Los trazos latentes se denotan  $\theta_{ig}$ , este representa el trazo latente del individuo  $i$  de la subpoblación  $g$ , con  $i = 1, \dots, n_g$ ,  $g = 1, \dots, G$  y  $n_g = n_1, \dots, n_G$ .
3. Se aplica una única prueba compuesta por  $J$  ítems con  $j = 1, \dots, J$ . Los parámetros de los ítems se denotan  $\eta_j = (\alpha_j, \beta_j)'$ , donde  $\alpha_j$  y  $\beta_j$  representan los parámetros de discriminación y de dificultad del ítem  $j$  respectivamente;  $\eta = (\eta_1, \dots, \eta_J)$ , los parámetros de los ítems son comunes a todas las subpoblaciones dado que todos los individuos responden la misma prueba.
4. Cada subpoblación  $g$  de los trazos proviene de una distribución normal asimétrica  $\theta_{ig} \sim NA(\mu_{\theta_g}, \sigma, \gamma_{\theta_g})$ , con  $i = 1, \dots, n_g$  y  $g = 1, \dots, G$  y  $\xi_{\theta_g} = (\mu_{\theta_g}, \sigma, \gamma_{\theta_g})$  corresponde al conjunto de los parámetros de la población  $g$ .
5. Se asume independencia entre grupos y entre individuos del mismo grupo.
6. Dado un trazo latente fijo las respuestas de un mismo individuo son independientes.

### 2.1. Función de probabilidad del modelo TRIAM

Sea  $Y_{igj}$  la variable aleatoria dicotómica que denota la respuesta del  $i$ -ésimo individuo con  $i = 1, \dots, n_g$  de la población  $g$ ,  $g = 1, \dots, G$  al ítem  $j$  con  $j = 1, \dots, J$ .  $Y_{igj}$  toma el valor de 1 si el individuo responde correctamente o 0 en otro caso. El modelo TRIAM

cuando se asigna una distribución normal asimétrica específica a los trazos latentes de cada subpoblación, está dado por:

$$\begin{aligned} Y_{igj} | \theta_{ig}, \eta_j &\sim Ber(p_{igj}) \\ p_{igj} &= P(Y_{igj} = y_{igj} | \theta_{ig}, \eta_j) \end{aligned} \quad (2.1)$$

$$P(Y_{igj} = 0 | \theta_{ig}, \eta_j) = 1 - P(Y_{igj} = 1 | \theta_{ig}, \eta_j) = \frac{1}{1 + e^{\alpha_j(\theta_{ig} + \beta_j)}}, \quad (2.2)$$

donde:

- $Y_{igj}$ : es la respuesta del  $i$ -ésimo individuo de la subpoblación  $g$  al ítem  $j$ , con  $i = 1, \dots, n_g$ ,  $g = 1, \dots, G$  y  $j = 1, \dots, J$ .
- $\theta_{ig}$ : es el trazo latente del individuo  $i$  de la subpoblación  $g$ .
- $\alpha_j$ : es el parámetro de discriminación del  $j$ -ésimo ítem.
- $\beta_j$ : es el parámetro de dificultad del  $j$ -ésimo ítem.
- $\eta_j$  es el conjunto de parámetros del  $j$ -ésimo ítem.

La probabilidad  $p_{igj} = P(Y_{igj} = y_{igj} | \theta_{ig}, \eta_j)$  es la probabilidad condicional de respuesta correcta dado el  $i$ -ésimo valor del trazo  $\theta_{ig}$  y el parámetro del  $j$ -ésimo  $\eta_j = (\alpha_j, \beta_j)$ . Se tiene un grupo de  $N$  individuos divididos en  $G$  grupos respondiendo a  $J$  ítems, cada grupo con  $n_g$  individuos y  $N = \sum_{g=1}^G n_g$ . Por otro lado, de manera general se puede definir un modelo

TRIAM de tres parámetros, pero en este trabajo se introduce un modelo TRIAM de 2 parámetros del ítem, que es el definido en la ecuación 2.2.

### Transformación de parámetros

Se realizó una transformación del parámetro de discriminación  $\alpha$ , dado que en HMC es importante que los parámetros no tengan restricciones y básicamente tomen valores en  $\mathbb{R}$  (equipo desarrollador de Stan, 2016). El parámetro de discriminación  $\alpha$  transformado es:

$$\alpha_j = \log(1 + \exp(\alpha_j^*)), \text{ con } j = 1, \dots, J$$

## 2.2. Especificación de las distribuciones a priori

En el modelo TRIAM se propone que los trazos latentes tengan como a priori una distribución normal asimétrica:

$$\theta_{ig} \sim SN(\mu_{\theta_g}, \sigma, \gamma_{\theta_g}), \quad i = 1, \dots, n_g \text{ y } g = 1, \dots, G.$$

con  $-\infty < \mu < \infty$ ,  $\sigma^2 > 0$  y  $-\infty < \gamma < \infty$ , donde  $SN$  representa la distribución normal asimétrica de parámetros: de localización ( $\mu_{\theta_g}$ ), de escala ( $\sigma$ ) y parámetro de asimetría

( $\gamma_{\theta_g}$ ). Para que el modelo TRIAM sea identificable se fija la distribución de los trazos latentes de una de las subpoblaciones con  $\mu = 0$  y se fija el parámetro de escala para todas las subpoblaciones  $\sigma = 1$ :

$$\begin{aligned}\theta_{i1} &\sim SN(0, 1, \gamma_{\theta_g}), \\ \theta_{ig} &\sim SN(\mu_{\theta_g}, 1, \gamma_{\theta_g}), \quad i = 1, \dots, n_g \text{ y } g = 2, \dots, G.\end{aligned}$$

A continuación se especifican las a priori para los parámetros de los items  $\alpha_j$ ,  $\beta_j$  y para los hiperparámetros  $\mu_{\theta_g}$ ,  $\gamma_{\theta_g}$ ,  $\mu_\beta$  y  $\sigma_\beta$ :

$$\begin{aligned}\theta_{ig} &\sim SN(\mu_{\theta_g}, 1, \gamma_{\theta_g}) \\ \mu_{\theta_g} &\sim N(0, \frac{1}{4}) \\ \gamma_{\theta_g} &\sim N(0, 1) \\ \alpha_j^* &\sim N(0, \frac{1}{4}) \\ \beta_j &\sim N(\mu_\beta, \sigma_\beta) \\ \mu_\beta &\sim N(0, \frac{1}{4}) \\ \sigma_\beta &\sim Cauchy(\frac{1}{10}, \frac{1}{10}).\end{aligned}$$

### 2.3. Distribución posterior

Para el caso del modelo TRIAM la distribución posterior total de los parámetros  $\theta$ ,  $\eta$ , los hiperparámetros de las subpoblaciones y de los items  $\xi_\eta$  y  $\xi_\xi$  respectivamente, dados los datos  $Y$ , es:

$$\begin{aligned}p(\Theta, \eta, \xi_\theta | y, \xi_\eta, \xi_\xi) &\propto \left\{ \prod_{g=1}^G \prod_{i=1}^{n_g} \prod_{j=1}^J P(y_{igj} | \theta_{ig}, \eta_j) \right\} \left\{ \prod_{g=1}^G \prod_{i=1}^{n_g} p(\theta_{ig} | \xi_{\theta_g}) \right\} \\ &\times \left\{ \prod_{j=1}^J p(\eta_j | \xi_\eta) \right\} \left\{ \prod_{g=1}^G p(\xi_{\theta_g} | \xi_\xi) \right\},\end{aligned}\tag{2.3}$$

donde,  $\xi_{\theta_g}$  representa el conjunto de parámetros de la población  $g$ ,  $\xi_\eta$  y  $\xi_\xi$  representan los hiperparámetros asociados a  $\eta$  y  $\xi_\theta$ , respectivamente.

## CAPÍTULO 3

---

---

### Estimación de los parámetros del modelo TRIAM

---

---

A continuación se muestra el código usado para la estimación de los parámetros de los ítems y los trazos latentes con el modelo TRIAM desarrollado en RStan.

#### Bloque de Datos

En el modelo TRIAM los parámetros que deben especificarse son: número de poblaciones, vector con tamaño de cada subpoblación, número total de estudiantes o individuos, número total de ítems; para efecto de la identificabilidad del modelo se fija la media de la primera población y el parámetro de escala, es por ello que en este bloque se incluyen estas dos constantes (líneas 4 y 5).

```
1 data {
2 int<lower=0> G;// número de grupos o subpoblaciones
3 int<lower=0> N;// número total de estudiantes
4 int<lower=0> mu;// media fija
5 int<lower=0> sigma;// param escala
6 int<lower=0> n_g[G];//tamaño de subpoblaciones
7 int<lower=0> J;// número total de ítems
8 int<lower=0,upper=1> Y1[(n_g[1]),J]; //matriz de datos P1
9 int<lower=0,upper=1> Y2[(n_g[2]),J]; //matriz de datos P2
10 int<lower=0,upper=1> Y3[(n_g[3]),J]; //matriz de datos P3
11 int<lower=0,upper=1> Y4[(n_g[4]),J]; //matriz de datos P4
13 }
```

#### Bloque de parámetros

Los parámetros  $\theta$ ,  $\alpha_a$  ( $\alpha^*$ ) y  $\beta$  corresponden a los trazos latentes de los individuos, una transformación del parámetro de discriminación y el parámetro dificultad, respectivamente, se definen como vectores. El parámetro  $\alpha_a$  ( $\alpha^*$ ) es una transformación del parámetro de discriminación y no se hace necesario colocar la restricción de que debe ser mayor que 0. En este caso el modelo TRIAM se aplica a un conjunto de datos con cuatro subpoblaciones, en consecuencia aparecen las medias de tres de ellas dado que la primera se ha fijado (líneas 23, 24 y 25), se especifica que son reales. En caso de que

algún parámetro deba ser mayor que cero, se debe indicar, como se muestra a continuación, línea programación debe iniciar con “*real < lower = 0 >*” (línea 30), seguido del nombre del parámetro.

```

14 parameters {
15 vector[(n_g[1])] theta1;
16 vector[(n_g[2])] theta2;
17 vector[(n_g[3])] theta3;
18 vector[(n_g[4])] theta4;
20 vector[J] alpha_a;
21 vector[J] beta;
22 real mu_beta;
23 real mu_theta2;
24 real mu_theta3;
25 real mu_theta4;
26 real gamma1;
27 real gamma2;
28 real gamma3;
29 real gamma4;
30 real<lower=0> sigma_beta;
31 }

```

### Bloque de parámetros transformados (Opcional)

Este bloque es opcional, en este caso se realizó una transformación del parámetro de discriminación.

```

32 transformed parameters{
33 vector[J] alpha;
34 for(j in 1:J){
35 alpha[j]=log(1+exp(alpha_a[j]));
36 }
37 }

```

### Bloque del modelo

En esta sección se asignan las a priori para los parámetros e hiperparámetros y se especifica el modelo a estimar. Para el caso del modelo TRIAM se asigna una priori normal asimétrica para los trazos latentes de cada subpoblación y las correspondientes a priori para los parámetros de los items e hiperparámetros enunciadas en el capítulo 2 (ver sección 2.2; líneas de la 51 a la 54). Finalmente se plantea el modelo a estimar que en este caso corresponde al modelo TRIAM (líneas de 55 a la 72).

```

39 model {
40 theta1 ~ skew_normal(mu,sigma, gamma1);
41 gamma1 ~ normal(0,1);

```

```

42 theta2 ~ skew_normal(mu_theta2,sigma, gamma2);
43 mu_theta2 ~ normal(0,0.25);
44 gamma2 ~ normal(0,1);

...

51 alpha_a ~ normal(0,0.25);
52 beta ~ normal(mu_beta,sigma_beta);
53 mu_beta ~ normal(0,0.25);
54 sigma_beta ~ cauchy(0.01,0.01);

55 {for(i in 1:(n_g[1])){
56 for(j in 1:J){
57 Y1[i,j] ~ bernoulli_logit((alpha[j]*(theta1[i] - 58 beta[j])));
59 }}}
60 {for(i in 1:(n_g[2])){
...
68 {for(i in 1:(n_g[4])){
69 for(j in 1:J){
70 Y4[i,j] ~ bernoulli_logit((alpha[j]*(theta4[i] - beta[j])));
71 }}}
72}

```

### Bloque de generación de cantidades

Este bloque permite el cálculo de nuevas variables y obtener sus correspondientes distribuciones posterior. En la TRI se usa esencialmente para calcular el modelo basado en la log-Verosimilitud, su calculo requiere el uso de la combinación de la función usada en la etapa del muestreo y el sufijo log (Luo Y.,Jiao H.,2017). En el modelo TRIAM, se usa la función *bernoulli\_logit\_lpmf* (línea 69).

Se usa para calcular el LOO y el WAIC, criterios de comparación de modelos Bayesianos descritos en la sección anterior. Por otro lado, se debe tener en cuenta que un código stan se finaliza con una línea en blanco; además, si se desea chequear la sintaxis del modelo se puede hacer antes de ejecutarlo, pero esto no consolida un diagnóstico de convergencia.

```

63 generated quantities {
64 vector[J] log_lik[N];
65 for (i in 1: N){
66 for(j in 1:J){
67 real eta;
68 eta = (alpha[j]*theta[i] -d[j]);
69 log_lik[i, j] = bernoulli_logit_lpmf(Y[i,j] | eta);
70 }}
71 }
72

```

**Rstan**

Inicialmente se fija como directorio la misma carpeta en la que se encuentra guardado el archivo stan, en este caso se denominó triam.stan. para usar Rstan se deben cargar varias librerías (líneas 2 a 5) y además se incluyen librerías adicionales para importar y exportar datos, entre otros (líneas 6 y 7).

```
#CARGANDO STAN
1 setwd("C:/Rtesis/TRIAM_DOC")
2 library(rstan)
3 rstan_options(auto_write= TRUE)
4 options(mc.cores = parallel::detectCores())
5 library(loo)
6 library(writexl)
7 library(readxl)
```

Es importante indicar que para la aplicación del modelo TRIAM se organizan los datos de cada subpoblación en matrices separadas (ver tablas 3.1 y 3.2), en las filas se tienen los individuos por subpoblación y en las columnas los items, los trazos latentes estimados se obtienen en vectores igualmente separados.

TABLA 3.1. Matriz de datos subpoblación 1

| Subpoblación | i1 | i2 | i3 | i4 | i5 | i6 | ... | iJ |
|--------------|----|----|----|----|----|----|-----|----|
| 1            | 1  | 0  | 0  | 1  | 0  | 0  |     | 0  |
| 1            | 0  | 0  | 0  | 0  | 0  | 0  |     | 0  |
| 1            | 0  | 0  | 0  | 1  | 0  | 0  |     | 0  |
| 1            | 0  | 0  | 0  | 0  | 1  | 0  |     | 0  |
| 1            | 0  | 0  | 0  | 1  | 0  | 0  |     | 0  |
| 1            | 0  | 0  | 0  | 0  | 0  | 0  | ... | 1  |
| 1            | 0  | 1  | 0  | 0  | 0  | 0  |     | 0  |

TABLA 3.2. Matriz de datos subppoblación 2

| Subpoblación | i1 | i2 | i3 | i4 | i5 | i6 | ... | iJ |
|--------------|----|----|----|----|----|----|-----|----|
| 2            | 1  | 1  | 0  | 1  | 1  | 0  | ... | 0  |
| 2            | 1  | 1  | 1  | 1  | 1  | 0  |     | 1  |
| 2            | 1  | 1  | 1  | 1  | 1  | 0  |     | 1  |
| 2            | 1  | 1  | 1  | 1  | 1  | 1  |     | 1  |
| 2            | 1  | 1  | 1  | 1  | 1  | 0  |     | 1  |

A continuación se cargan los datos a R para luego especificarlos en el modelo Stan, estos se incluirán en el bloque de datos. En el caso específico del modelo TRIAM las matrices de datos se importaron desde Excel, en esta, la primera columna corresponde a la codificación de la subpoblación (aunque esta se puede obviar dado que no es de importancia en la especificación de los datos).

```
#CARGANDO DATOS
```

```
8 datos1 <- read_excel("C:/Rtesis/triam/MUESTRA_spn.xlsx",sheet =
"COLM", col_names = FALSE)
9 datos2 <- read_excel("C:/Rtesis/triam/MUESTRA_spn.xlsx",sheet =
"ARGM", col_names = FALSE)
10 datos3 <- read_excel("C:/Rtesis/triam/MUESTRA_spn.xlsx",sheet =
"VENM", col_names = FALSE)
11 datos4 <- read_excel("C:/Rtesis/triam/MUESTRA_spn.xlsx",sheet =
"CUBM", col_names = FALSE)
```

En el modelo TRIAM es necesario crear un vector con los tamaños de cada subpoblación (líneas 12-16) y el número total de subpoblaciones (línea 17). Además, se establece las dimensiones de la matriz de datos, número de columnas corresponde al total de ítems y el número de filas o la sumatoria de los tamaños de las subpoblaciones al total de individuos. Igualmente se especifican las constantes a incluir (líneas 20 y 21)

```
#TAMAÑOS DE LAS SUBPOBLACIONES
```

```
#####
```

```
12 n1=dim(datos1)[1]
13 n2=dim(datos2)[1]
14 n3=dim(datos3)[1]
15 n4=dim(datos4)[1]
16 n_g=c(n1,n2,n3,n4)#vector con los tamaños de las subpoblaciones
17 N = n1+n2+n3+n4 #Total de individuos
18 G= length(n_g)
19 J=dim(datos1)[2]
20 mu_col=0 #media subpoblación fija, constante en el modelo
21 sigma= 1 # Parámetro de escala, constante en el modelo
```

Antes de iniciar la estimación se debe crear una lista con los datos a incluir el en bloque de datos.

```
#DATOS PARA STAN
```

```
22 data_irt = list(n_g = n_g,G=G,N = I, mu_col=mu_col, sigma=sigma, J=J,
Y1=datos1, Y2=datos2, Y3=datos3, Y4=datos4)
```

Por último, se tiene la especificación del modelo, como ya se había mencionado para implementar un modelo Stan se hace uso de la función `stan`, algunos de sus argumentos son:

- `file`: Archivo “.stan” que contiene el algoritmo.
- `data`: lista de datos
- `iter`: número de iteraciones, por defecto stan fija el warmup o calentamiento en  $\frac{iter}{2}$

- chains: el número de cadenas que se desea estimar

```
#MODELO STAN
```

```
23 irt_2pltriam = stan(file = 'triam.stan', data = data_irt, iter = 3000,
chains = 3)
```

### Salidas en Rstan

Una vez realizada la estimación se puede hacer uso de herramientas como: *pairs* (línea 24), *print* (línea 24), *traceplot* (línea 31).

```
#DIAGNÓSTICO GRAFICO DE CONVERGENCIA
23 pairs(irt_2pltriam,pars="alpha")
#EXTRACCIÓN PARÁMETROS
24 print(irt_2pltriam,par="theta")
...
#GRAFICA CADENAS
31 traceplot(irt_2pltriam, pars= c("theta[3]","theta[17]","theta[25]"),
,inc_warmup = FALSE)
```

A continuación se muestra una salida de Rstan, se puede observar un resumen de las estadísticas a posteriori del parámetro de discriminación, los intervalos de credibilidad, entre otros. Se puede evidenciar que bajo el criterio de Gelman y Rubin (1992), las cadenas mostraron buena convergencia. En el Apéndice D. se muestran algunos resultados de las estadísticas a posteriori del parámetro  $\beta$  de dificultad de los items, de los trazos latentes y de los parámetros de localización  $\mu$ , asimetría  $\gamma$  para cada subpoblación.

```
Inference for Stan model: triam.
```

```
3 chains, each with iter=5000; warmup=2500; thin=1;
post-warmup draws per chain=2500, total post-warmup draws=7500.
```

|           | mean | sd   | 2.5% | 25%  | 50%  | 75%  | 97.5% | n_eff | Rhat |
|-----------|------|------|------|------|------|------|-------|-------|------|
| alpha[1]  | 1.09 | 0.05 | 0.98 | 1.05 | 1.08 | 1.12 | 1.19  | 7500  | 1    |
| alpha[2]  | 1.11 | 0.06 | 1.01 | 1.08 | 1.11 | 1.15 | 1.23  | 7500  | 1    |
| alpha[3]  | 0.96 | 0.05 | 0.87 | 0.93 | 0.96 | 0.99 | 1.05  | 7500  | 1    |
| alpha[4]  | 1.30 | 0.08 | 1.15 | 1.25 | 1.30 | 1.36 | 1.47  | 7500  | 1    |
| alpha[5]  | 0.76 | 0.04 | 0.69 | 0.74 | 0.76 | 0.79 | 0.84  | 7500  | 1    |
| alpha[6]  | 0.49 | 0.03 | 0.43 | 0.47 | 0.49 | 0.51 | 0.54  | 7500  | 1    |
| alpha[7]  | 1.06 | 0.05 | 0.97 | 1.03 | 1.06 | 1.09 | 1.15  | 7500  | 1    |
| alpha[8]  | 0.91 | 0.05 | 0.81 | 0.87 | 0.91 | 0.95 | 1.02  | 7500  | 1    |
| alpha[9]  | 0.83 | 0.04 | 0.75 | 0.80 | 0.83 | 0.86 | 0.91  | 7500  | 1    |
| alpha[10] | 1.26 | 0.07 | 1.13 | 1.21 | 1.26 | 1.30 | 1.39  | 7500  | 1    |
| alpha[11] | 1.00 | 0.05 | 0.90 | 0.97 | 1.00 | 1.04 | 1.11  | 7500  | 1    |
| alpha[12] | 0.79 | 0.04 | 0.72 | 0.77 | 0.79 | 0.82 | 0.87  | 7500  | 1    |
| alpha[13] | 0.94 | 0.04 | 0.86 | 0.91 | 0.94 | 0.97 | 1.03  | 7500  | 1    |
| alpha[14] | 0.57 | 0.03 | 0.51 | 0.55 | 0.57 | 0.59 | 0.63  | 7500  | 1    |
| alpha[15] | 1.12 | 0.05 | 1.01 | 1.08 | 1.12 | 1.15 | 1.22  | 7500  | 1    |

|           |      |      |      |      |      |      |      |      |   |
|-----------|------|------|------|------|------|------|------|------|---|
| alpha[16] | 1.10 | 0.05 | 1.01 | 1.07 | 1.10 | 1.13 | 1.20 | 7500 | 1 |
| alpha[17] | 0.77 | 0.04 | 0.70 | 0.74 | 0.77 | 0.79 | 0.84 | 7500 | 1 |
| alpha[18] | 0.93 | 0.04 | 0.85 | 0.90 | 0.93 | 0.96 | 1.02 | 7500 | 1 |
| alpha[19] | 0.89 | 0.04 | 0.82 | 0.87 | 0.89 | 0.92 | 0.97 | 7500 | 1 |
| alpha[20] | 1.08 | 0.05 | 1.00 | 1.05 | 1.08 | 1.12 | 1.18 | 7500 | 1 |
| alpha[21] | 0.90 | 0.04 | 0.82 | 0.87 | 0.90 | 0.93 | 0.98 | 7500 | 1 |
| alpha[22] | 0.58 | 0.03 | 0.52 | 0.56 | 0.58 | 0.60 | 0.65 | 7500 | 1 |
| alpha[23] | 0.85 | 0.04 | 0.77 | 0.82 | 0.85 | 0.88 | 0.94 | 7500 | 1 |
| alpha[24] | 1.12 | 0.06 | 1.01 | 1.09 | 1.12 | 1.16 | 1.24 | 7500 | 1 |
| alpha[25] | 1.06 | 0.05 | 0.96 | 1.02 | 1.06 | 1.09 | 1.16 | 7500 | 1 |
| alpha[26] | 0.67 | 0.04 | 0.60 | 0.64 | 0.67 | 0.69 | 0.74 | 7500 | 1 |
| alpha[27] | 0.90 | 0.05 | 0.81 | 0.87 | 0.90 | 0.93 | 0.99 | 7500 | 1 |
| alpha[28] | 0.70 | 0.03 | 0.63 | 0.67 | 0.69 | 0.72 | 0.76 | 7500 | 1 |
| alpha[29] | 1.08 | 0.05 | 0.99 | 1.05 | 1.08 | 1.12 | 1.18 | 7500 | 1 |
| alpha[30] | 1.15 | 0.05 | 1.04 | 1.11 | 1.15 | 1.19 | 1.26 | 7500 | 1 |
| alpha[31] | 0.66 | 0.03 | 0.59 | 0.64 | 0.66 | 0.68 | 0.73 | 7500 | 1 |
| alpha[32] | 0.76 | 0.04 | 0.69 | 0.73 | 0.76 | 0.78 | 0.83 | 7500 | 1 |

Samples were drawn using NUTS(diag\_e) at Thu May 31 02:39:42 2018.  
 For each parameter, `n_eff` is a crude measure of effective sample size,  
 and `Rhat` is the potential scale reduction factor on split chains (at  
 convergence, `Rhat=1`).

En el gráfico 3.1 se puede observar varias cadenas de los trazos latentes estimados con el modelo TRIAM, se puede inferir buena convergencia.

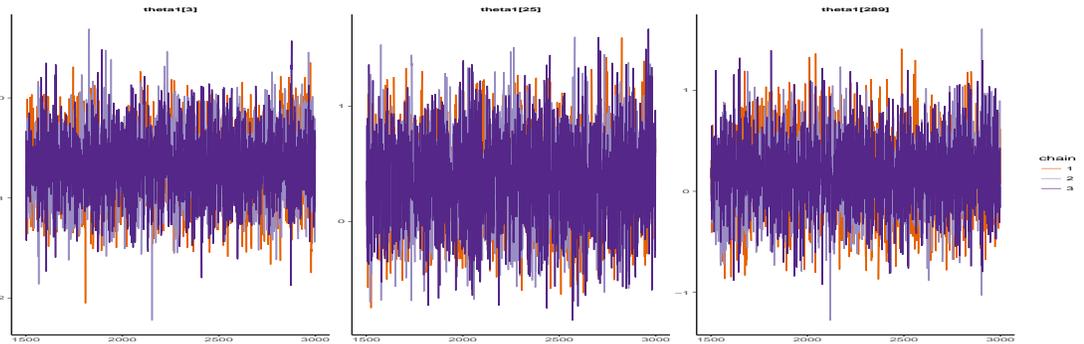


FIGURA 3.1. Muestra de las cadenas estimadas trazos latentes

---

---

## Aplicación a datos reales

---

---

Como se menciona en la introducción el modelo fue diseñado para estimar los datos provenientes de la prueba PERCE (Casassus & Cusato, 1998). Estos datos fueron estudiados inicialmente por Antonio y Montenegro (2014) quienes propusieron un modelo TRI jerárquico multidimensional, ellos usaron una muestra equivalente al 6 % de los datos provenientes de Colombia, Cuba y Bolivia. En este trabajo se amplía el tamaño de muestra, se analiza el comportamiento de otros países, se tiene en cuenta la estructura de subpoblaciones y se incluye la asimetría para los trazos latentes. De forma general se puede señalar que una de las aplicaciones del modelo TRIAM son las pruebas que tienen por objetivo medir la habilidad de los estudiantes en determinadas áreas del conocimiento, ya que, como lo afirma Hashimoto (2002) los trazos referentes a las habilidades escolares pueden ser asimétricamente distribuidas y que además los individuos estén divididos en subpoblaciones.

### 4.1. Aplicación del modelo TRIAM con distribución asimétrica para el trazo latente

Para la aplicación a del modelo TRIAM se seleccionaron cuatro países: Argentina, Colombia, Venezuela y Cuba. Se seleccionó a Colombia con el interés de observar su desempeño y realizar comparaciones con los de Argentina y Venezuela; Cuba fue seleccionado debido a que muestra un desempeño atípico (Casassus & Cusato, 1998), además Antonio y Montenegro (2014) señalan que existe evidencia de asimetría. La muestra seleccionada corresponde al 22 % de los estudiantes que presentaron la prueba distribuidos de la siguiente manera: 847 estudiantes de Colombia, 845 estudiantes de Argentina, 748 estudiantes de Venezuela y 798 estudiantes de Cuba para un total de 3238 estudiantes. Es importante tener en cuenta que en este caso las respuestas ausentes o no respuestas fueron calificadas como cero. Posteriormente el tamaño de muestra se redujo, ya que, se realizaron algunos exclusiones para la estimación; se excluyeron los puntajes perfectos y los puntajes nulos (dado que esta prueba no los estaba midiendo correctamente).

Como se menciona anteriormente se realizaron algunas exclusiones de la muestra. En Argentina, Colombia y en Venezuela se eliminaron 1, 2 y 6 individuos respectivamente cuyos

puntajes fueron cero; en el caso de Cuba se realizaron 98 exclusiones, 95 que corresponden a puntajes perfectos y 3 a puntajes nulos.

Inicialmente se muestran los resultados de un estudio de simulación, posteriormente se presentan los resultados de las estimaciones por país haciendo uso de un modelo TRI asimétrico de dos parámetros con enlace logístico (con distribución asimétrica para el trazo latente), a continuación se muestran los resultados de la implementación del modelo TRIAM y del modelo MGM (Bock y Zimoski,1997), finalmente se muestra la prueba de bondad de ajuste (Fox, 2010) para el modelo TRIAM y MGM.

### Estudio de simulación

En esta sección se presentan los resultados de un estudio de simulación realizado con el objetivo de evaluar la calidad de recuperación de los parámetros de los ítems y de trazos latentes, utilizando el código en Rstan del modelo TRIAM (que se encuentra en el Anexo C, detallado en capítulo 3).

Para simular los datos se tomaron las estimaciones de los datos provenientes de la prueba PERCE realizados con el modelo TRIAM. A los parámetros de los ítems  $\alpha_j$ ,  $\beta_j$  para  $j = 1, \dots, J$  y los trazos latentes de los individuos  $\theta_{ig}$  con  $i = 1, \dots, n_g$ ,  $g = 1, \dots, G$ , se les sumo una uniforme  $(-0.25, 0.25)$ . Con estos parámetros y trazos latentes se simularon 100 test (el código fue desarrollado en R, este puede ser descargado en <https://www.r-project.org/>). Para recuperar los parámetros se tomaron los promedios de las estimaciones.

Para lograr la convergencia de las cadenas (se estimaron tres cadenas por cada parámetro e hiperparámetro) fueron suficientes 3000 muestras de las distribuciones posterior más las de calentamiento (1500). Para la estimación se usó un computador con procesador Intel Core i7, con una velocidad de 3.1 GHz con una memoria Ram de 8 Gb.

Se calculo la correlación, el error absoluto y el error cuadrático medio(ECM) entre los parámetros simulados y recuperados, además, de que se tuvieron en cuenta los intervalos de credibilidad que muestra Rstan, los resultados de algunas simulaciones se muestran en la tabla 4.1, se puede ver que las correlaciones para los parámetros de los ítems simulados y estimados oscilan entre  $[0.98, 0.99]$  y para los trazos latentes oscilan entre  $[0.88, 0.90]$ .

TABLA 4.1. Correlaciones, error absoluto y ECM valores estimados vs simulados

| Parámetro       | Cor    | Error absoluto | sd     | ECM    |
|-----------------|--------|----------------|--------|--------|
| $\alpha$        | 0.984  | 0.0743         | 0.0555 | 0.0084 |
| $\beta$         | 0.997  | 0.015          | 0.1204 | 0.0142 |
| $\mu_\theta$    | 0.9998 | 0.0499         | 0.0336 | 0.0032 |
| $\gamma_\theta$ | 0.9836 | 0.0928         | 0.0125 | 0.0725 |
| $\theta_1$      | 0.8929 | 0.2973         | 0.2351 | 0.2111 |
| $\theta_2$      | 0.9046 | 0.3251         | 0.2423 | 0.1643 |
| $\theta_3$      | 0.8759 | 0.3064         | 0.2305 | 0.1470 |
| $\theta_4$      | 0.8825 | 0.3914         | 0.2866 | 0.2352 |

Para estimar los parámetros del modelo fueron necesarias 3000 iteraciones par lograr la convergencia<sup>1</sup>, esta se evidenció con el criterio de Gelman y Rubin (1992) que muestra Stan. Además, se analizaron los intervalos de credibilidad y los diagnósticos gráficos (ver resultados completos Anexo A.)

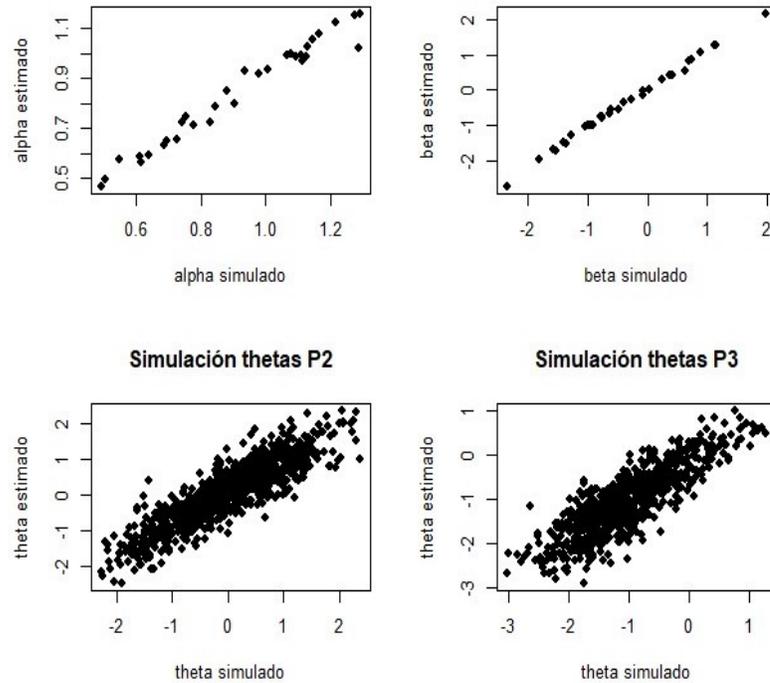


FIGURA 4.1. Gráficos de dispersión de los valores simulados versus los valores estimados con el modelo TRIAM

En la figura 4.1 se muestran los gráficos de dispersión de los valores simulados versus los valores estimados con el modelo propuesto. Se observa que el rango de los valores simulados es similar al de los valores estimados, además existe una alta correlación entre estos. Por lo anterior, se puede afirmar que los valores estimados son cercanos a los valores simulados; se evidencia además, que la correlación es más alta para los parámetros de dificultad que para los de discriminación.

### Estimaciones por país

Inicialmente se aplicó un modelo TRI con distribución asimétrica para el trazo latente por separado para cada subpoblación; con el fin observar las estimaciones de los parámetros de los items en cuanto a invarianza en las diferentes poblaciones; además, de observar el comportamiento de los trazos latentes, en esta sección se muestran los resultados de la aplicación de un modelo 2PL con distribución normal asimétrica para los trazos latentes, siguiendo la estrategia dada en Bazán, et. al (2011) cambiando la función de enlace. Los modelos se aplicaron a la muestra de la prueba PERCE descrita anteriormente.

<sup>1</sup>En un procesador intel core i7 con 8Gb, 3 cadenas de 3000 iteraciones cada una, tardaron 1,59 horas en promedio

En cuanto a las estimaciones de los parámetros de discriminación se realizaron comparaciones para los cuatro países, encontrándose que para Colombia tuvieron valores diferentes en los ítems 1 y 22; para el caso de Argentina los valores varían en los ítems 8 y 11; para Venezuela los ítems 4, 7, 16, 19, 20, 29, 30 tuvieron valores diferentes y finalmente para Cuba los siguientes ítems tuvieron valores diferentes 9, 12, 14, 25, 29 y 32 (los resultados se pueden ver en el Apéndice 5). Para el parámetro de dificultad se puede decir que en general los valores obtenidos fueron muy diferentes para los cuatro países (los resultados se pueden ver en el Apéndice 5). Se observó que para Cuba se obtuvieron valores del parámetro de discriminación muy bajos en comparación de los Venezuela que fueron los más altos. Teóricamente no se deberían presentar estas diferencias, una de las razones de que se presenten puede ser debido a que las poblaciones provienen de diferentes poblaciones.

En la figura 4.2 se muestran las densidades de los trazos latentes estimados, arriba a la izquierda Colombia, a la derecha Argentina, abajo a la izquierda Cuba y a la derecha Venezuela. Se observa diferencias en las densidades de Venezuela y Cuba con respecto a las de Colombia y Argentina.

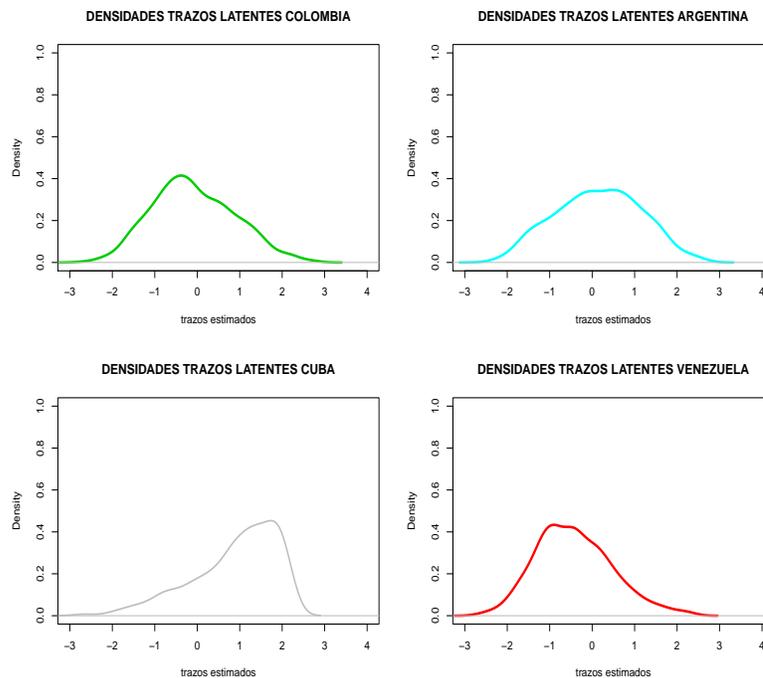


FIGURA 4.2. Densidades trazos latentes estimados individualmente modelo TRI asimétrico

## 4.2. Estimación de los parámetros del modelo TRIAM

### Parámetros de los ítems

Para realizar la estimación de los parámetros con el modelo TRIAM se hizo uso del código de Rstan que se encuentra en el Apéndice 5. Las estimaciones del parámetro de discriminación oscilan entre 0.4863 y 1,3085, las del parámetro de dificultad se encuentran entre -2.3104 y 1,96. De la estimación de los parámetros de los ítems se puede resaltar que el ítem

6 tiene la discriminación más baja y la dificultad más alta, el ítem 4 tiene la discriminación más alta y la dificultad más baja.

### Estimación (predicción) de los trazos latentes

En la figura 4.3 se muestran las densidades de los trazos latentes estimados conjuntamente (se estiman simultáneamente los trazos latentes de los cuatro países), se puede observar que las densidades difieren entre los países (Colombia, Argentina, Venezuela y Cuba), la diferencia más notoria se observa en la densidad de Cuba (a la derecha), que de acuerdo a lo observado por Casassus y Cusato (2001) muestra un comportamiento diferente con respecto a los demás países participantes y a lo encontrado por Antonio y Montenegro (2014) quienes sugieren que su distribución es asimétrica. En concordancia con estos hallazgos se encontró que efectivamente su distribución es asimétrica. De la misma forma se observa que Venezuela (a la izquierda) muestra un comportamiento asimétrico pero inverso al de Cuba, para el caso de Venezuela la mayoría de sus trazos latentes son bajos. Por último se observa comportamiento similar en las densidades de Colombia y Argentina, aunque Argentina esta un poco más a la derecha.

En las figuras 4.2, 4.3 y 4.5 se muestran las densidades de los trazos latentes estimados con el modelo TRI asimétrico (por país), con el modelo TRIAM y con el modelo MGM respectivamente. Se puede concluir que las tres representaciones tienen semejanza, por ejemplo, la distribución de los trazos latentes de Argentina en las tres estimaciones parece ser simétricas, mientras que la de Cuba, Colombia y Venezuela muestran asimetría. Se observa, por ejemplo, que Cuba y Venezuela tienen asimetrías opuestas; siendo Cuba el país con el mejor desempeño, Venezuela y Colombia muestran desempeños más bajos.

En las tablas 4.2, 4.3, 4.4 y 4.5 se muestran las estimaciones para los parámetros ( $\mu_g$ : parámetro de localización,  $\gamma_g$ : parámetro de asimetría,  $d_j$ : transformación del parámetro de asimetría (ver ecuación 1.2),  $\bar{\theta}_g$ : estimación de la media poblacional,  $\sigma_g^2$ : estimación de la varianza poblacional; con  $g = 1, \dots, 4$ ) de las 4 poblaciones: Colombia, Argentina, Venezuela y Cuba respectivamente. En estas se puede observar que al comparar las estimaciones, de las medias poblacionales de los trazos latentes de los estudiantes de los cuatro países, los estudiantes de Venezuela son los que tienen el desempeño más bajo y los de Cuba el desempeño más alto. En correspondencia con lo anterior se observa que el parámetro  $d$  para Venezuela es de -0,69 y la moda de los trazos latentes es de -0,97, que es inferior a la moda de los estudiantes de Cuba (1,805).

TABLA 4.2. Estimaciones para los parámetros de localización  $\mu$ , asimetría  $\gamma$  y  $d$  de Colombia modelo TRIAM

| PARÁMETRO  | MEDIA     | SD      | 2.5 %    | 25 %     | 50 %      | 75 %    | 97.5 %  |
|------------|-----------|---------|----------|----------|-----------|---------|---------|
| $\mu_1$    | 0         | NA      | NA       | NA       | NA        | NA      | NA      |
| $\gamma_1$ | -0,603306 | 0,111   | -0,81255 | -0,67978 | -0,606529 | -0,5296 | -0,3778 |
| $d_1$      | -0,511498 | 0,07078 | -0,63062 | -0,56219 | -0,518595 | -0,4680 | -0,3534 |

De igual modo, se observa que Colombia se encuentra un poco más a la derecha de Venezuela (ver tabla 4.2) y se puede afirmar que los trazos latentes son efectivamente asimétricos,

dado que el valor del parámetro  $d$  es  $-0,511$  (también es negativo al igual que en Venezuela y Argentina) que es significativo según la literatura.

En cuanto a las estimaciones de los parámetros de Argentina se puede mencionar que parece ser simétrica, dado que el valor del parámetro  $d$  es  $-0,006$  que de acuerdo con la literatura no es significativo. Se observa además que las estimaciones de la media y la moda son muy cercanas ( $0,03599$  y  $0,03596$ , respectivamente).

TABLA 4.3. Estimaciones para los parámetros de localización  $\mu$ , asimetría  $\gamma$  y  $d$  de Argentina modelo TRIAM

| PARÁMETRO  | MEDIA     | SD      | 2.5 %    | 25 %     | 50 %      | 75 %    | 97.5 % |
|------------|-----------|---------|----------|----------|-----------|---------|--------|
| $\mu_2$    | -0,0311   | 0,13312 | -0,2926  | -0,1219  | -0,02864  | 0,06271 | 0,2186 |
| $\gamma_2$ | -0,006287 | 0,1557  | -0,30717 | -0,11508 | -0,009103 | 0,1035  | 0,2959 |
| $d_2$      | -0,006129 | 0,15086 | -0,29363 | -0,11433 | -0,009102 | 0,1030  | 0,2837 |

TABLA 4.4. Estimaciones para los parámetros de localización  $\mu$ , asimetría  $\gamma$  y  $d$  de Venezuela modelo TRIAM

| PARÁMETRO  | MEDIA     | SD      | 2.5 %    | 25 %     | 50 %      | 75 %     | 97.5 %  |
|------------|-----------|---------|----------|----------|-----------|----------|---------|
| $\mu_3$    | -0,5141   | 0,08024 | -0,6723  | -0,567   | -0,51444  | -0,46054 | -0,3547 |
| $\gamma_3$ | -0,821777 | 0,1223  | -1,06212 | -0,90382 | -0,822405 | -0,7399  | -0,5832 |
| $d_3$      | -0,62984  | 0,05755 | -0,72808 | -0,67053 | -0,63519  | -0,5948  | -0,5038 |

TABLA 4.5. Estimaciones para los parámetros de localización  $\mu$ , asimetría  $\gamma$  y  $d$  de Cuba modelo TRIAM

| PARÁMETRO  | MEDIA    | SD      | 2.5 %    | 25 %    | 50 %     | 75 %    | 97.5 % |
|------------|----------|---------|----------|---------|----------|---------|--------|
| $\mu_4$    | 1,6963   | 0,10516 | 1,4953   | 1,6247  | 1,69581  | 1,76684 | 1,9035 |
| $\gamma_4$ | 0,13989  | 0,0751  | -0,00898 | 0,08908 | 0,140416 | 0,1902  | 0,2879 |
| $d_4$      | 0,137436 | 0,07243 | -0,00898 | 0,08873 | 0,139052 | 0,1869  | 0,2767 |

Los valores de los parámetros de asimetría y el parámetro transformado  $d$  de los trazos latentes de los estudiantes es diferente para Cuba (ver tabla 4.5) comparado con el de los otros tres países, dado que este es el único cuyo parámetro de asimetría es de signo positivo, además, como ya se mencionó también tiene la media y la moda más altas. Esto va en correspondencia con lo encontrado en estudios realizados anteriormente. Puesto que con base en el valor del parámetro  $d$  y según la literatura la distribución de los trazos latentes es asimétrica.

TABLA 4.6. Estimaciones para la medias, las varianzas y las modas para cada subpoblación, Colombia=1, Argentina=2, Venezuela=3 y Cuba=4, modelo TRIAM.

| PARÁMETRO          | MEDIA    | SD      | 2.5 %   | 25 %     | 50 %     | 75 %     | 97.5 %  |
|--------------------|----------|---------|---------|----------|----------|----------|---------|
| $\bar{\theta}_1$   | -0,40812 | 0,05648 | -0,5032 | -0,44856 | -0,41378 | -0,37343 | -0,282  |
| $\hat{\sigma}_1^2$ | 0,8303   | 0,04475 | 0,7468  | 0,7988   | 0,8288   | 0,8606   | 0,9205  |
| $Mod_1$            | -0.38908 | 0.04852 | -0.4663 | -0.42427 | -0.39542 | -0.36044 | -0.2768 |
| $\bar{\theta}_2$   | -0,03599 | 0,0737  | -0,1714 | -0,08645 | -0,04023 | 0,01128  | 0,1202  |
| $\hat{\sigma}_2^2$ | 0,9855   | 0,01834 | 0,9347  | 0,9796   | 0,9925   | 0,9983   | 1       |
| $Mod_2$            | -0.03596 | 0.07358 | -0.1709 | -0.08638 | -0.04032 | 0.01128  | 0.1200  |
| $\bar{\theta}_3$   | -1,01661 | 0,06972 | -1,1438 | -1,06583 | -1,01917 | -0,97075 | -0,8717 |
| $\hat{\sigma}_3^2$ | 0,7453   | 0,04515 | 0,6625  | 0,7138   | 0,7431   | 0,7748   | 0,8384  |
| $Mod_3$            | -0.97809 | 0.06980 | -1.1074 | -1.02646 | -0.98069 | -0.93240 | -0.8330 |
| $\bar{\theta}_4$   | 1,80599  | 0,08812 | 1,6408  | 1,74424  | 1,80345  | 1,86326  | 1,9847  |
| $\hat{\sigma}_4^2$ | 0,9846   | 0,01323 | 0,9513  | 0,9778   | 0,9877   | 0,995    | 0,9999  |
| $Mod_4$            | 1.80545  | 0.08814 | 1.6404  | 1.74346  | 1.80286  | 1.86294  | 1.9845  |

En la tabla 4.6 se observan las estimaciones de las medias, varianzas y modas de cada subpoblación, que son comparables entre si, se puede observar que Cuba tiene el desempeño más alto, en contra parte los estudiantes de Venezuela obtienen los desempeños más bajos. Además, también resaltar que los valores de la varianza de todas las poblaciones esta alrededor de 1, esto se debe a que el parámetro de escala se fijo para todas las subpoblaciones en 1.

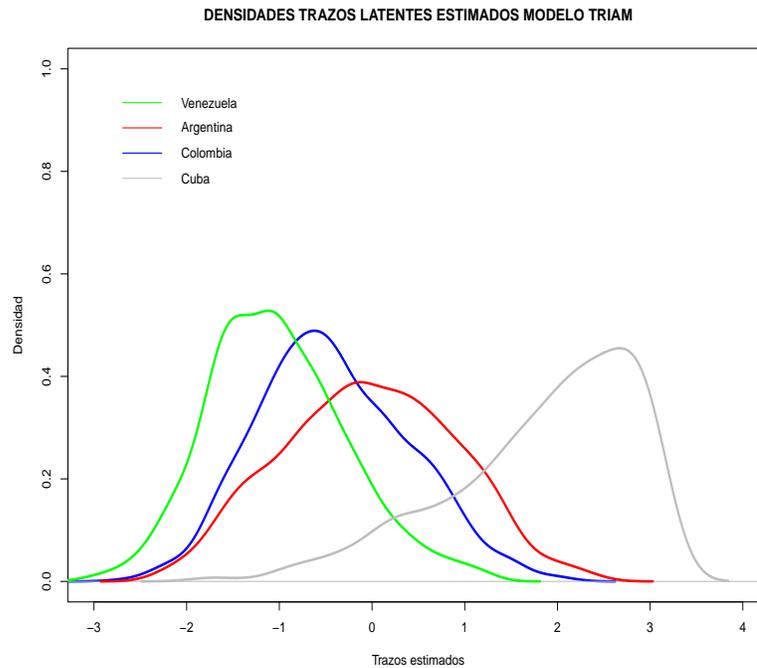


FIGURA 4.3. Densidades trazos latentes estimados de Colombia, Venezuela, Argentina y Cuba con el modelo TRIAM

Como lo afirma Muñiz (2010) la teoría clásica de los test y la TRI no son teorías opuestas sino complementarias, es por esto que se calculan las correlaciones entre los trazos estimados y el puntaje clásico, los resultados se muestran en la tabla 4.7. Se encontró que las correlaciones oscilan entre 0.9831 y 0.9942, se observa que la correlación más baja fue la de Cuba, con un valor de 0.9831.

TABLA 4.7. Correlaciones entre los trazos latentes estimados y el puntaje clásico

| Puntaje clásico/Trazos estimados | $\theta$  |
|----------------------------------|-----------|
| COLOMBIA                         | 0.9942505 |
| VENEZUELA                        | 0.9923715 |
| ARGENTINA                        | 0,9919966 |
| CUBA                             | 0.9831179 |

En la figura 4.4 se muestran las correlaciones entre los trazos latentes estimados y los puntajes clásicos por país. Se observa que Colombia, Argentina y Venezuela tienen pocos puntajes clásicos altos. Además se evidencia que las correlaciones son altas, lo que implica que estudiantes con puntajes clásicos iguales tuvieron trazos latentes similares, aunque como se evidenció anteriormente no son iguales.

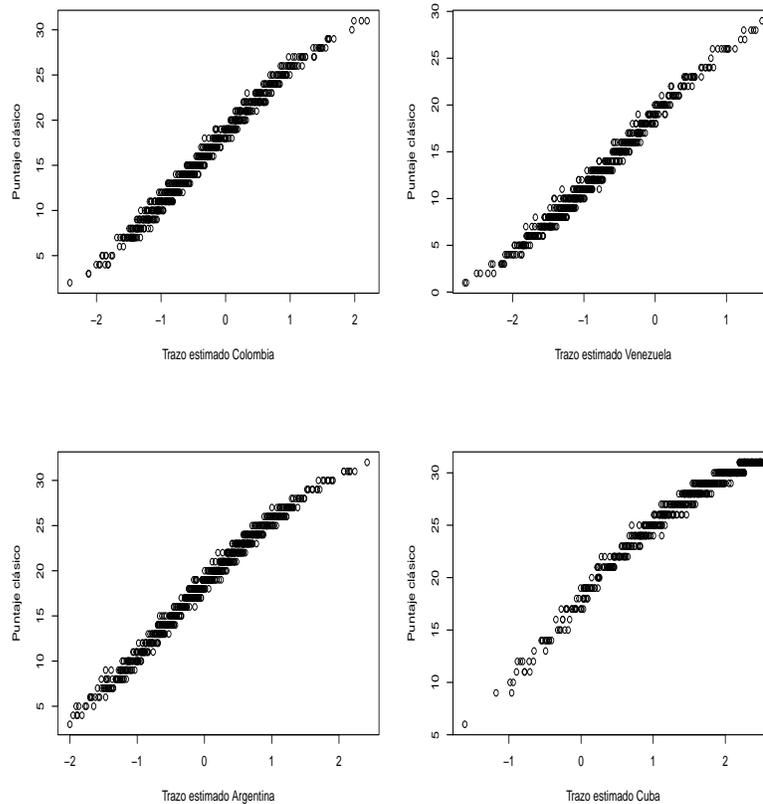


FIGURA 4.4. Puntaje clásico vs. trazos latentes estimados. Arriba a la izquierda Colombia, a la derecha Venezuela, abajo a la izquierda Argentina y a la derecha Cuba

Con el propósito de comparar las estimaciones obtenidas con el modelo TRIAM al conjunto de datos seleccionados y establecer mediante la prueba de bondad de ajuste si el modelo TRIAM se adapta mejor a los datos que el modelo MGM, se usó el modelo MGM para realizar estimaciones de los trazos latentes al mismo conjunto de datos seleccionado; las densidades de los trazos latentes estimados con el modelo MGM se muestran en la figura 4.5.

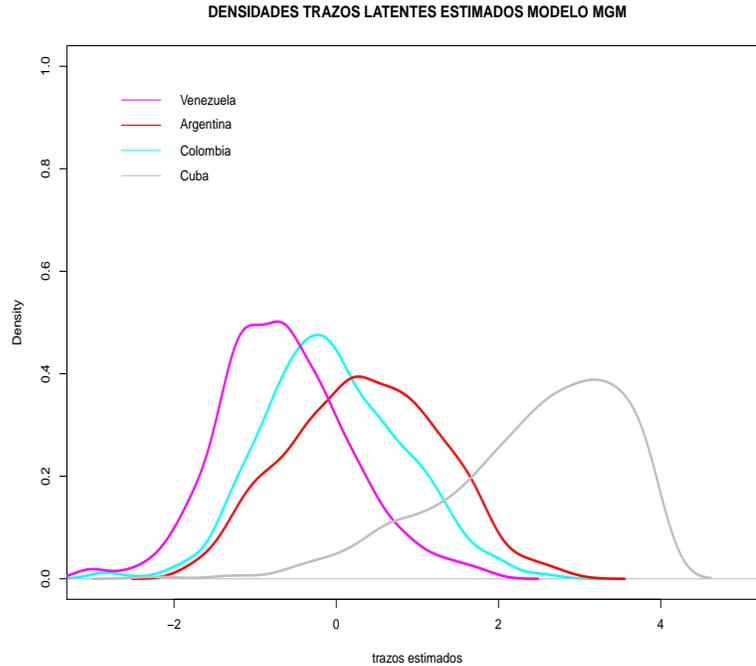


FIGURA 4.5. Densidades trazos latentes estimados de Colombia, Argentina, Venezuela y Cuba con el modelo MGM

### 4.3. Ajuste del modelo

En esta sección se describe el criterio de bondad de ajuste usado para el modelo TRIAM. Para evaluar el ajuste del modelo se usó la evaluación de las predicciones (Fox, 2010). El método consiste en generar predicciones también llamadas medidas de discrepancia (Gelman et al., 1996) y comparar los datos replicados con los datos observados (Fox, 2010, pp.118). Se asume como hipótesis nula que los datos de respuesta se distribuyen condicionalmente  $p(y|\xi)$  y los parámetros desconocidos tienen distribución a priori  $p(\theta, \xi)$ , no se tiene una hipótesis alterna específica. Un p-valor (Bayesiano) dado un modelo  $M$  se puede estimar:

$$p_0(y_{obs}) = P(T(y_{rep}, \theta) \geq T(y_{obs}, \theta) | M)$$

Para evaluar el ajuste del modelo TRIAM con los datos observados se calculó un p-valor Bayesiano, para los ítems y para los individuos con las estimaciones obtenidas para el conjunto de datos reales usando el modelo TRIAM. En la tabla 4.8 se muestran los resultados de los p-valores Bayesianos para los ítems; se puede observar que todos los

valores están alrededor de 0.5; esto significa que todos los items pasan la prueba de bondad de ajuste (basados en p-valores Bayesianos). Además, se calculó un p-valor Bayesiano para estimar la compatibilidad del modelo con los datos observados de cada uno de los individuos, se encontró que el 0.639% por ciento de los p-valores son menores a 0.05 o mayores que 0,95 y con el criterio de 0.1 o del 0,9, se encontraron 4.4% de p-valores menores. Por lo anterior, se considera que el modelo TRIAM se ajusta a los datos de acuerdo con Fox (2010).

TABLA 4.8. Resultados de la prueba de bondad de ajuste para los items modelo TRIAM p-valores Bayesianos

|       |       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|-------|
| 0.314 | 0.290 | 0.355 | 0.250 | 0.467 | 0.584 | 0.348 | 0.400 |
| 0.422 | 0.247 | 0.359 | 0.442 | 0.372 | 0.573 | 0.298 | 0.327 |
| 0.454 | 0.364 | 0.411 | 0.322 | 0.397 | 0.586 | 0.408 | 0.317 |
| 0.318 | 0.509 | 0.401 | 0.491 | 0.320 | 0.278 | 0.531 | 0.464 |

### Comparación modelo TRIAM y modelo MGM

Por otro lado, se realizó el mismo procedimiento descrito anteriormente con las estimaciones obtenidas con el modelo MGM al mismo conjunto de datos reales, los p-valores obtenidos para los items se pueden ver en la tabla 4.9, en esta se observa que varios items no pasan el test, es decir, el modelo no ajusta bien a los items. Al igual que con el modelo TRIAM se calculó un p-valor Bayesiano para estimar la compatibilidad de los modelos con los datos observados, encontrando que el 3.8% los p-valores menores a 0.05 y con el criterio de 0.1, se encontraron 7.64% de p-valores menores. Con base en estos resultados se puede decir que el modelo MGM no es valido.

TABLA 4.9. Resultados de la prueba de bondad de ajuste para los items modelo MGM p-valores Bayesianos

|       |       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|-------|
| 0.148 | 0.059 | 0.352 | 0.000 | 0.531 | 1.000 | 0.058 | 0.848 |
| 0.396 | 0.001 | 0.342 | 0.417 | 0.392 | 1.000 | 0.192 | 0.148 |
| 0.518 | 0.382 | 0.719 | 0.251 | 0.448 | 0.531 | 0.601 | 0.028 |
| 0.256 | 0.951 | 0.664 | 1.000 | 0.290 | 0.189 | 0.999 | 0.773 |

Aunque Stan brinda la posibilidad de calcular los criterios LOO y WAIC en este caso no se estimaron para ninguno de los dos modelos debido a que por la cantidad de datos es poco práctico.

---

---

## Discusión

---

---

Se introdujo un modelo Bayesiano de teoría de respuesta al ítem diseñado para ser aplicado a datos provenientes de test cuya población esta subdividida en múltiples grupos, como lo son los datos provenientes de la prueba PERCE. Se asumió que los trazos latentes tienen distribuciones asimétricas diferentes en cada subpoblación; en trabajos previos se evidenció comportamiento atípico para algunas subpoblaciones (ver por ejemplo Casassus y Cusato, 2001; Antonio y Montenegro, 2014) por lo que se intuía que las distribuciones de los trazos latentes debían ser asimétricas. Se mostró que al tener en cuenta la asimetría en la distribución de los trazos el modelo ajusta mejor y en consecuencia se mejoran las estimaciones.

Inicialmente, se realizaron estimaciones por separado para cada subpoblación aplicando un modelo TRI con asimetría para el trazo latente haciendo uso de la distribución normal asimétrica. Se evidenció que al aplicar este tipo de modelos los parámetros de los items no son invariantes. Esta situación no sólo se evidencia en la prueba PERCE sino que como lo mostró Fox (2010) al estudiar datos provenientes de la prueba PISA, también en esta existen diferencias en los parámetros de los items entre poblaciones; es decir, los parámetros de los items no son invariantes. El uso del modelo TRIAM permite superar esta dificultad dado que los parámetros de los items se estiman conjuntamente para todas las subpoblaciones, es decir, son invariantes.

Para evaluar el algoritmo de estimación propuesto se analizó la calidad de recuperación de los parámetros de los items y el ajuste del modelo a los datos. Se realizó el estudio de simulación, para verificar la calidad de la recuperación de los parámetros; se encontró que recupera adecuadamente los parámetros de los items. De igual manera, se realizaron pruebas de bondad de ajuste, que permitió comprobar el adecuado ajuste del modelo a los datos. Además, el modelo propuesto TRI para múltiples grupos con distribución asimétrica para los trazos latentes, es más flexible que los modelos TRI usados en la estimación de los trazos latentes en estudios como la prueba PERCE, ya que tiene en cuenta la estructura de subpoblaciones, que es usual en este tipo de análisis y permite modelar la asimetría presente cuando se miden habilidades humanas (Hashimoto, 2002).

En general, una de las áreas de aplicación del modelo TRIAM es en el ámbito educativo, a nivel internacional se le ha venido dando especial importancia a la medición de

habilidades académicas, algunas de las finalidades son el indagar acerca de la capacidad de los estudiantes para utilizar los conocimientos y habilidades frente a los desafíos que plantea la sociedad actual. Estos resultados más allá de obtener indicadores de desempeño individuales se realizan con el fin de diseñar estrategias que van desde la modificación de contenidos curriculares y establecimiento de estándares de desempeño hasta la creación de políticas públicas con el fin de racionalizar y focalizar recursos hacia procesos que favorecerán la calidad y equidad de la educación. Por ejemplo, la OCDE (2006, pp. 29) argumenta que las evaluaciones internacionales aportan información para diseñar y poner en práctica políticas públicas adecuadas para la atención de las problemáticas descubiertas y mejorar la calidad de los sistemas educativos.

En torno a lo anterior, se han aplicado pruebas internacionales como las TIMSS (Trends in International Mathematics and Science Study), PERCE (Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado), PISA (Program for International student assessment), entre otras. Con el fin de cumplir los objetivos de la aplicación de estas pruebas se espera que los resultados sean comparables entre los países participantes; por esto se procura que los instrumentos sean válidos en todos los contextos de los países participantes, por ejemplo, en las pruebas TIMSS los currículos de los países participantes se tienen en cuenta en el marco teórico y a la hora de seleccionar los dominios de los contenidos temáticos de los instrumentos; en las pruebas PISA el instrumento de evaluación fue gestado con la colaboración de los mismos países que habrían de aplicarlo, y ello contribuyó a que se atendiera la universalidad en el enfoque del examen y en la pertinencia cultural; en la prueba PERCE el contenido de la prueba PERCE fue aprobado por los coordinadores de educación de cada uno de los países participantes, para garantizar que los contenidos del test estuvieran incluidos en los currículos de todos los países.

Es evidente que se realizan esfuerzos para que los instrumentos de medición sean los adecuados; pero es de igual importancia usar el modelo adecuado para la estimación de los resultados. En este sentido los modelos TRI son una herramienta apropiada que favorece la calidad de las estimaciones, estos han sido usados para estimar los trazos latentes en algunas de las pruebas mencionadas, como es caso de la PERCE y las TIMSS, pero, es conveniente indagar dentro de los modelos TRI existentes el más adecuado. En este sentido, es importante contar con instrumentos de fácil implementación que se adapten al uso de nuevas herramientas, que aseguren la calidad de las estimaciones como consecuencia del uso del modelo pertinente.

Por ejemplo, en la prueba PERCE se usó un modelo Rasch (Casassus y Cusato, 2011) y Antonio y Montenegro (2014) mostraron que no es adecuado usar este modelo, debido a que los datos no cumplen los supuestos para que el modelo Rasch sea aplicado desde el punto de vista teórico; además, se mostró que al menos uno de los países participantes en la prueba PERCE tenía distribución asimétrica para los trazos latentes. La estrategia usada por los evaluadores de la prueba PERCE para generar una sola escala para todos los países, habiendo encontrado diferencias entre ellos, fue el uso de un modelo de Rasch. Extrajeran una muestra regional especial de datos, para las formas A y B, de los grados tercero y cuarto en cada país. El tamaño de la muestra fue de aproximadamente 300 estudiantes para cada combinación de formas y grados para cada país, en lenguaje y matemáticas. La meta era aplicar la Escala de Rasch a un grupo representativo de estudiantes y dar a cada

---

país igual participación en la definición de la escala. Las muestras (aproximadamente 7.200 para cada prueba) se consideraron lo suficientemente grandes para generar una escala de Rasch precisa (Casassus & Cusato 2001, pp. 183-184).

En este trabajo se propone el modelo TRIAM como alternativa para el análisis de pruebas internacionales como las mencionadas. Este fue diseñado para modelar datos de pruebas aplicadas a poblaciones divididas en grupos. Con el uso del modelo TRIAM se tienen en cuenta las distribuciones de los trazos latentes, y permite modelar la asimetría. Además, se estima un conjunto de parámetros de los items que al ser calibrado con toda la población permite hacer comparaciones entre los parámetros de las distribuciones a posteriori de las subpoblaciones. En el capítulo 6, se mostró que la aplicación de este modelo a una muestra de datos de la prueba PERCE, ajusta bien a los datos, mientras que el modelo MGM no. Igualmente, se verificó la presencia de asimetría en los trazos latentes.

Por otro lado, la implementación del modelo TRIAM se realizó haciendo uso de Stan, que es un lenguaje de programación probabilístico muy poderoso que entre otros beneficios facilita la estimación, muestra cadenas mejor comportadas, requiere menos iteraciones para la convergencia de las cadenas, entre otras. Además cuenta con la interface Rstan, como se evidencia en Apéndice 5 (código completo implementación modelo TRIAM), su codificación es muy sencilla en aspectos como incluir los datos, especificar a priori, realizar transformaciones, entre otras.

Stan es la herramienta que se está utilizando en la actualidad para modelos y cálculos estadísticos de alto rendimiento; dado que usa muestreadores más eficientes en comparación con otros software Bayesianos como Bugs, Winbugs o Jags, como lo son el NUTS ó el HMC. Algunos de sus usos son el modelado estadístico, el análisis de datos y la predicción en diferentes ciencias (Equipo desarrollador de Stan, 2017). Una de las dificultades que se encontró fue que la curva de aprendizaje al usar una herramienta nueva es bastante pesada pero se ve compensada como ya se mencionó por la facilidad de la implementación; otra de sus ventajas en este sentido, es la posibilidad de tener diagnósticos precisos sobre los errores cometidos en la codificación, ya que es posible conocer en que línea del código se cometieron, adicionalmente se puede encontrar asesoría en los foros en línea, en los que participan miembros del equipo desarrollador de Stan como Bob Carpenter y también usuarios que han adquirido cierta destreza y ayudan a que otros usuarios superen su dificultades.

---

---

## Conclusiones

---

---

- El propósito en pruebas internacionales como las TIMSS, PERCE, PISA, entre otras, es realizar comparaciones entre los países participantes; en el análisis de estas es usual aplicar el modelo de Rash; pero los mismos expertos en algunas ocasiones han mencionado que su uso no es adecuado, dado que los conjuntos de datos no cumplen los supuestos.
- Autores como Antonio y Montenegro han usado modelos TRI para la estimación de los trazos en este tipo de datos, evidenciaron la necesidad de incluir asimetría; es por esto que se propone el modelo TRIAM como alternativa para el análisis de este tipo de pruebas internacionales. En la modelación (aplicación a datos de la prueba PERCE) se encontró que efectivamente las distribuciones de los trazos de al menos tres países de los países seleccionados son asimétricas dado que los valores de los parámetros de asimetría son  $(-0,51, -0,01, -0,63, 0,14)$ ; lo que indica que hay diferencias en los parámetros de las poblaciones.
- Para evaluar el algoritmo de estimación propuesto se quizá comparar con el modelo MGM (Bock y Zimoski, 1997) que era el más apropiado, al realizar las pruebas de bondad de ajuste se encontró que este modelo no ajusta al conjunto de datos.
- Aunque con el modelo TRIAM se obtienen estimaciones del trazo latente de cada individuo, el interés se centra en realizar comparaciones entre las subpoblaciones (para el caso de la aplicación, países). Las estimaciones de los trazos de los individuos obtenidas con el modelo TRIAM son comparables dentro de cada subpoblación pero no entre subpoblaciones.

Si se desea o requieren realizar comparaciones entre individuos de diferentes poblaciones, se plantea como trabajo posterior desarrollar un algoritmo para homogeneizar los puntajes de toda la población.

- Stan es la herramienta que se está utilizando en la actualidad para modelos y cálculos estadísticos de alto rendimiento, dado que entre otros beneficios ya mencionados aporta todos los resultados necesarios para la evaluación de un modelo, además, se ha desarrollado una comunidad académica abierta de apoyo y retroalimentación entorno a su implementación. Es por esto que se recomienda se motive su uso en la universidad en los diferentes programas.
- Como trabajos futuros se plantea ampliar la utilización de modelos para múltiples poblaciones multidimensionales, en la actualidad ya se han planteado modelos similares que modelan la asimetría pero para una única población.

- 
- Utilización del modelo en otro tipo de problemas en áreas como psicología, medicina y demás en las que es de interés analizar variables referentes al comportamiento humano.

---

---

## Trabajos futuros

---

---

- Utilizar otras distribuciones asimétricas para los trazos latentes como: la distribución gamma, lognormal, t-asimétrica, entre otras.
- Estimar un modelo TRI con distribución asimétrica para trazos latentes multidimensionales, que implicaría el uso de la distribuciones asimétricas multidimensionales.
- Desarrollar un procedimiento para poder equipar los puntajes de todos los países.

---

---

## Apéndice A. Introducción a Stan

---

---

Stan surgió por la necesidad de aplicar la inferencia Bayesiana para estimar modelos que haciendo uso de Jags, Winbugs o Bugs mostraban problemas de convergencia después de muchas iteraciones (más de 100 000; Equipo desarrollador de Stan, 2017, p. X). Es un lenguaje de programación probabilística de estadística Bayesiana que facilita la implementación de complejos algoritmos haciendo uso del muestreador No-U-Turn por sus siglas en inglés NUTS (Hoffman & Gelman, 2012) ó una extensión del Hamiltoniano Monte Carlo HMC (Neal, 2011). Entre otras ventajas Stan permite el uso de prior impropias, además, debido a que el HMC explora el espacio de los parámetros más eficientemente que el muestreador de Gibbs y que el algoritmo Metropolis (Gelman, Carlin, Stern, & Rubin, 2014, p. 300), es considerado más eficiente que otros software para estadística Bayesiana como Bugs, Winbugs o Jags (Luo Y.,Jiao H.,2017).

Los modelos escritos en Stan son compilados a  $C_{++}$  lo que hace que sea más rápido y permite la traducción a otros lenguajes de programación. Otra de las ventajas de Stan es que cuenta con interfaces para los lenguajes del análisis de datos más comunes como los son: R, Python, Shell, MATLAB, Julia, Mathematica, Scala y Stata; además, se ejecuta en los sistemas operativos más usados: Linux, Windows y Mac.

El equipo desarrollador de Stan (2017) señala que cuando se especifica una función de log-densidad en el lenguaje de programación probabilístico de Stan, se obtiene:

- Completa Inferencia estadística Bayesiana con muestreo MCMC (NUTS, HMC)
- Inferencia Bayesiana aproximada con inferencia variable (ADVI)
- Estimación por máxima verosimilitud penalizada con optimización (L-BFGS)

Además ofrece métodos de comparación de modelos basados en la log-verosimilitud como lo son el LOO (leave-one-out cross validation; Vehtari, Gelman, & Gabry, 2016 ) y el WAIC (widely available information criterion; Watanabe, 2010), estos se pueden calcular incluyendo en la programación en Stan un bloque denominado “generated quantities”, del cual se hará mención más adelante.

El cálculo del LOO y el WAIC puede ser complejo y no trivial por lo que usualmente se deben usar métodos de aproximación. Por otro lado, el LOO y el WAIC son métodos de estimación puntual “out-of-sample” (dejando uno fuera) de precisión de predicción del ajuste de un modelo Bayesiano y usan la log-verosimilitud evaluada en las simulaciones posterior de los valores de los parámetros. Estos métodos de comparación tienen varias ventajas con respecto a las estimaciones simples de error predictivo tales como AIC (Akaike information criterion) y DIC (deviance information criterion; Vehtari, Gelman y Gabry, 2016).

Para implementar un modelo haciendo uso del paquete Rstan, que es la interface de Stan para R, que se usa en el presente trabajo se especifica un archivo Stan que se debe guardar con la extensión .stan y otro con los comandos en R, los dos archivos son complementarios. Aunque se pueden unificar, no es recomendable, ya que al estar en archivos separados se pueden identificar fácilmente errores de programación o realizar las modificaciones necesarias sin alterar los dos simultáneamente.

El paquete Rstan está disponible en <https://github.com/stan-dev/rstan/wiki/RStan-Getting-Started>, junto con las instrucciones de instalación y ejemplos. Si se desea otra interface se puede acceder consultando <http://mc-stan.org/users/interfaces/index.html>.

A continuación se describe cada componente del modelo.stan (código en Stan), bloques de: datos, parámetros, transformación de parámetros (opcional), modelo y generación de cantidades (opcional). Para ilustrarlo se muestra un ejemplo tomado de Gelman et al. (2003), disponible en <https://github.com/stan-dev/rstan/wiki/RStan-Getting-Started> que denominó ocho escuelas. En este el autor quería observar el efecto de los instructores o profesores en 8 escuelas.

### **Bloque de Datos**

En este primer bloque se colocan los datos relevantes para el modelo, en el ejemplo, se especifican: número de escuelas, efectos estimados del tratamiento y desviación estándar de los efectos estimados.

```
data {  
  int<lower=0> J; // number of schools  
  real y[J]; // estimated treatment effects  
  real<lower=0> sigma[J]; // s.e. of effect estimates  
}
```

### **Bloque de parámetros**

En este segundo bloque se realiza una lista de los parámetros e hiperparámetros del modelo especificando el tipo de parámetro (real, entero o vector) y sus restricciones. En caso de que algún parámetro por ejemplo deba ser mayor que cero, se debe indicar en línea programación, esta debe iniciar con “*real < lower = 0 >*”, seguido del nombre del parámetro.

```
parameters {
  real mu;
  real<lower=0> tau;
  real eta[J];
}
```

### Bloque de parámetros transformados (Opcional)

Este bloque es opcional, en este caso se realizó una transformación. Es importante mencionar que los parámetros transformados no se deben enunciar en el bloque anterior.

```
transformed parameters {
  real theta[J];
  for (j in 1:J)
    theta[j] = mu + tau * eta[j];
}
```

### Bloque del modelo

En esta sección se asignan las a priori para los parámetros e hiperparámetros de ser necesario y se especifica el modelo a estimar.

```
model {
  target += normal_lpdf(eta | 0, 1);
  target += normal_lpdf(y | theta, sigma);
}
```

### Bloque de generación de cantidades

Este bloque permite el cálculo de nuevas variables y obtener sus correspondientes distribuciones posterior. En la TRI se usa esencialmente para calcular el modelo basado en la log-Verosimilitud, su calculo requiere el uso de la combinación de la función usada en la etapa del muestreo y el sufijo log (Luo Y.,Jiao H.,2017). En el ejemplo no se incluyó.

### Rstan

Inicialmente se fija como directorio la misma carpeta en la que se encuentra guardado el archivo stan, que en este caso se denominó 8schols.stan. La función principal del paquete Rstan es “*stan*”, para su uso se deben cargar varias librerías como se muestra a continuación.

```
#CARGANDO STAN
setwd("C:/ejm")
library(rstan)
rstan_options(auto_write= TRUE)
options(mc.cores = parallel::detectCores())
library(loo)
```

A continuación se cargan los datos o se preparan en R para luego especificarlos en el modelo Stan, estos se incluirán en el bloque de datos del archivo “.stan” descrito en la sección anterior. Para este ejemplo los datos se prepararon directamente en R.

```
#PREPARANDO DATOS
schools_dat <- list(J = 8,
y = c(28, 8, -3, 7, -1, 1, 18, 12),
sigma = c(15, 10, 16, 11, 9, 11, 10, 18))
```

Por último se tiene la especificación del modelo, como ya se había mencionado para implementar un modelo Rstan se hace uso de la función `stan`, algunos de sus argumentos son:

- `file`: Archivo “.stan” que contiene el algoritmo.
- `data`: lista de datos.
- `iter`: número de iteraciones, por defecto stan fija el warmup o calentamiento en  $\frac{iter}{2}$ .
- `chains`: el número de cadenas que se desea estimar.

```
#MODELO STAN
fit <- stan(file = '8schools.stan', data = schools_dat,
iter = 1000, chains = 4)
```

### Salidas en Rstan

Una vez realizada la estimación se puede hacer uso de herramientas como: *pairs* (diagnóstico gráfico de convergencia de las cadenas), *print* (extracción de parámetros), *extract* devuelve una lista de arreglos con la síntesis de la distribución posterior de los parámetros, *traceplot* (gráfico de las cadenas).

```
print(fit)
plot(fit)
pairs(fit, pars = c("mu", "tau", "lp_"))

la <- extract(fit, permuted = TRUE) # return a list of arrays
mu <- la$mu

### return an array of three dimensions: iterations, chains, parameters
a <- extract(fit, permuted = FALSE)

### use S3 functions on stanfit objects
a2 <- as.array(fit)
m <- as.matrix(fit)
d <- as.data.frame(fit)
```

A continuación se muestra una salida de Rstan, en este se puede observar que el paquete ofrece una síntesis de las estadísticas (media, desviación estándar, cuartiles, intervalos de credibilidad) de la distribución posterior de cada parámetro. Además, incluye otros resultados como *n-eff*, que se puede interpretar como el número efectivo de muestras para la distribución posterior sobre la cual se realizó la inferencia. La columna Rhat muestra el diagnóstico de convergencia de Gelman y Rubin (1992), este calcula el “potencial scale reduction factor” *PSRF* por sus siglas en inglés, una buena convergencia corresponde a valores alrededor de 1. Se observa que el Rhat es de 1, esto indica la convergencia del modelo; si se desean calcular otros diagnósticos de convergencia, Rstan brinda las herramientas para calcularlos a través del paquete de R ShinyStan (Equipo desarrollador de Stan, 2016).

```
> print(fit, digits = 1)
Inference for Stan model: 8schools.
4 chains, each with iter=1000; warmup=500; thin=1;
post-warmup draws per chain=500, total post-warmup draws=2000.
```

|          | mean   | se_mean | sd   | 2.5%   | 25%    | 50%    | 75%    | 97.5%  | n_eff | Rhat |
|----------|--------|---------|------|--------|--------|--------|--------|--------|-------|------|
| mu       | 8.00   | 0.15    | 4.86 | -1.42  | 4.90   | 7.93   | 11.19  | 17.60  | 1094  | 1    |
| tau      | 6.50   | 0.17    | 5.46 | 0.22   | 2.58   | 5.28   | 8.92   | 19.93  | 978   | 1    |
| eta[1]   | 0.40   | 0.02    | 0.94 | -1.52  | -0.26  | 0.42   | 1.04   | 2.26   | 2000  | 1    |
| eta[2]   | 0.00   | 0.02    | 0.85 | -1.72  | -0.55  | -0.01  | 0.57   | 1.71   | 2000  | 1    |
| eta[3]   | -0.15  | 0.02    | 0.93 | -1.98  | -0.81  | -0.14  | 0.47   | 1.71   | 2000  | 1    |
| eta[4]   | -0.05  | 0.02    | 0.87 | -1.75  | -0.65  | -0.08  | 0.54   | 1.71   | 1838  | 1    |
| eta[5]   | -0.35  | 0.02    | 0.86 | -1.98  | -0.92  | -0.38  | 0.19   | 1.53   | 2000  | 1    |
| eta[6]   | -0.21  | 0.02    | 0.89 | -1.90  | -0.78  | -0.24  | 0.36   | 1.63   | 2000  | 1    |
| eta[7]   | 0.36   | 0.02    | 0.90 | -1.39  | -0.22  | 0.35   | 0.96   | 2.20   | 2000  | 1    |
| eta[8]   | 0.07   | 0.02    | 0.94 | -1.83  | -0.54  | 0.09   | 0.66   | 1.97   | 2000  | 1    |
| theta[1] | 11.42  | 0.23    | 8.37 | -2.04  | 6.16   | 10.24  | 15.45  | 32.00  | 1382  | 1    |
| theta[2] | 7.99   | 0.14    | 6.17 | -4.52  | 4.08   | 8.05   | 11.82  | 20.34  | 2000  | 1    |
| theta[3] | 6.42   | 0.17    | 7.74 | -11.03 | 2.24   | 7.13   | 11.16  | 20.49  | 2000  | 1    |
| theta[4] | 7.55   | 0.14    | 6.34 | -5.50  | 3.74   | 7.68   | 11.33  | 20.27  | 2000  | 1    |
| theta[5] | 5.19   | 0.14    | 6.28 | -8.29  | 1.35   | 5.74   | 9.31   | 16.50  | 2000  | 1    |
| theta[6] | 6.03   | 0.15    | 6.68 | -9.15  | 2.37   | 6.57   | 10.19  | 18.58  | 2000  | 1    |
| theta[7] | 10.70  | 0.15    | 6.72 | -1.88  | 6.20   | 10.35  | 14.73  | 25.69  | 2000  | 1    |
| theta[8] | 8.39   | 0.18    | 7.56 | -7.18  | 4.25   | 8.30   | 12.67  | 24.30  | 1838  | 1    |
| lp__     | -39.48 | 0.09    | 2.56 | -45.49 | -40.93 | -39.26 | -37.73 | -35.02 | 732   | 1    |

Samples were drawn using NUTS(diag\_e) at Sun May 27 19:12:35 2018.  
 For each parameter, *n\_eff* is a crude measure of effective sample size,  
 and Rhat is the potential scale reduction factor on split chains (at  
 convergence, Rhat=1).

En la figura 1 se muestra un ejemplo de los gráficos de convergencia para los parámetros *mu*, *tau* y *lp\_\_* extraídos mediante la función *pairs*.

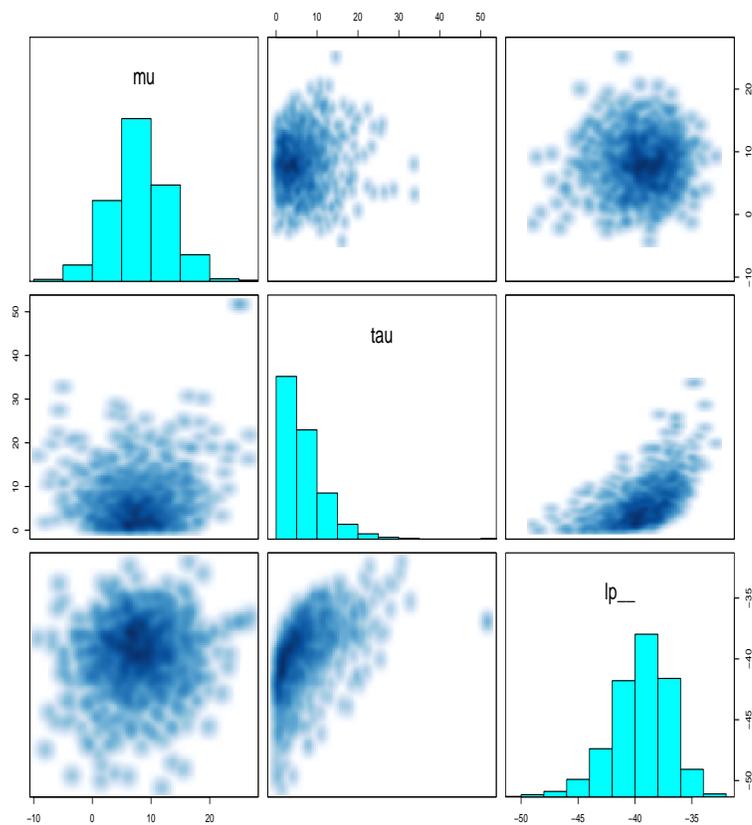


FIGURA 1. Diagnóstico gráfico de convergencia

# APÉNDICE

---

---

## Apéndice B. Resultados simulaciones

---

---

A continuación se muestran algunos resultados del estudio de simulación para el parámetro de discriminación, dificultad, medias y parámetro de asimetría de los trazos latentes

Estimaciones parámetros de discriminación

Inference for Stan model: triam.

3 chains, each with iter=3000; warmup=1500; thin=1;

post-warmup draws per chain=1500, total post-warmup draws=4500.

|           | mean | se_mean | sd   | 2.5% | 25%  | 50%  | 75%  | 97.5% | n_eff | Rhat |
|-----------|------|---------|------|------|------|------|------|-------|-------|------|
| alpha[1]  | 0.82 | 0       | 0.05 | 0.74 | 0.79 | 0.82 | 0.85 | 0.92  | 4500  | 1    |
| alpha[2]  | 1.19 | 0       | 0.06 | 1.08 | 1.15 | 1.19 | 1.23 | 1.31  | 4500  | 1    |
| alpha[3]  | 0.77 | 0       | 0.04 | 0.70 | 0.74 | 0.77 | 0.80 | 0.85  | 4500  | 1    |
| alpha[4]  | 1.19 | 0       | 0.08 | 1.04 | 1.14 | 1.19 | 1.24 | 1.35  | 4500  | 1    |
| alpha[5]  | 0.75 | 0       | 0.04 | 0.67 | 0.72 | 0.74 | 0.77 | 0.82  | 4500  | 1    |
| alpha[6]  | 0.29 | 0       | 0.02 | 0.24 | 0.27 | 0.29 | 0.30 | 0.34  | 4500  | 1    |
| alpha[7]  | 1.00 | 0       | 0.05 | 0.91 | 0.97 | 1.00 | 1.03 | 1.09  | 4500  | 1    |
| alpha[8]  | 1.04 | 0       | 0.06 | 0.93 | 1.00 | 1.04 | 1.08 | 1.16  | 4500  | 1    |
| alpha[9]  | 0.67 | 0       | 0.04 | 0.60 | 0.64 | 0.67 | 0.69 | 0.74  | 4500  | 1    |
| alpha[10] | 0.85 | 0       | 0.05 | 0.75 | 0.81 | 0.85 | 0.88 | 0.95  | 4500  | 1    |
| alpha[11] | 0.84 | 0       | 0.05 | 0.76 | 0.81 | 0.84 | 0.87 | 0.94  | 4500  | 1    |
| alpha[12] | 0.68 | 0       | 0.04 | 0.61 | 0.66 | 0.68 | 0.71 | 0.75  | 4500  | 1    |
| alpha[13] | 0.87 | 0       | 0.04 | 0.79 | 0.84 | 0.87 | 0.90 | 0.96  | 4500  | 1    |
| alpha[14] | 0.77 | 0       | 0.04 | 0.69 | 0.74 | 0.77 | 0.79 | 0.84  | 4500  | 1    |
| alpha[15] | 1.03 | 0       | 0.05 | 0.94 | 1.00 | 1.03 | 1.07 | 1.14  | 4500  | 1    |
| alpha[16] | 1.03 | 0       | 0.04 | 0.94 | 1.00 | 1.03 | 1.06 | 1.12  | 4500  | 1    |
| alpha[17] | 0.90 | 0       | 0.04 | 0.82 | 0.87 | 0.90 | 0.93 | 0.98  | 4500  | 1    |
| alpha[18] | 0.96 | 0       | 0.04 | 0.87 | 0.93 | 0.96 | 0.99 | 1.05  | 4500  | 1    |
| alpha[19] | 0.72 | 0       | 0.04 | 0.65 | 0.69 | 0.72 | 0.74 | 0.79  | 4500  | 1    |
| alpha[20] | 0.84 | 0       | 0.04 | 0.76 | 0.81 | 0.84 | 0.86 | 0.92  | 4500  | 1    |
| alpha[21] | 0.90 | 0       | 0.04 | 0.82 | 0.87 | 0.90 | 0.93 | 0.98  | 4500  | 1    |
| alpha[22] | 0.51 | 0       | 0.03 | 0.45 | 0.49 | 0.51 | 0.53 | 0.57  | 4500  | 1    |
| alpha[23] | 0.95 | 0       | 0.05 | 0.86 | 0.92 | 0.95 | 0.98 | 1.04  | 4500  | 1    |
| alpha[24] | 1.03 | 0       | 0.05 | 0.93 | 0.99 | 1.03 | 1.06 | 1.14  | 4500  | 1    |

|           |      |   |      |      |      |      |      |      |      |   |
|-----------|------|---|------|------|------|------|------|------|------|---|
| alpha[25] | 1.13 | 0 | 0.05 | 1.03 | 1.09 | 1.13 | 1.16 | 1.23 | 4500 | 1 |
| alpha[26] | 0.69 | 0 | 0.04 | 0.62 | 0.66 | 0.69 | 0.71 | 0.76 | 4500 | 1 |
| alpha[27] | 0.77 | 0 | 0.04 | 0.69 | 0.74 | 0.76 | 0.79 | 0.85 | 4500 | 1 |
| alpha[28] | 0.49 | 0 | 0.03 | 0.43 | 0.47 | 0.49 | 0.51 | 0.55 | 4500 | 1 |
| alpha[29] | 0.88 | 0 | 0.04 | 0.81 | 0.86 | 0.88 | 0.91 | 0.97 | 4500 | 1 |
| alpha[30] | 1.08 | 0 | 0.05 | 0.98 | 1.04 | 1.08 | 1.11 | 1.18 | 4500 | 1 |
| alpha[31] | 0.64 | 0 | 0.03 | 0.57 | 0.62 | 0.64 | 0.66 | 0.71 | 4500 | 1 |
| alpha[32] | 0.53 | 0 | 0.03 | 0.47 | 0.51 | 0.53 | 0.56 | 0.60 | 4500 | 1 |

Estimaciones parámetro de dificultad

Inference for Stan model: triam.

3 chains, each with iter=1000; warmup=500; thin=1;

post-warmup draws per chain=500, total post-warmup draws=1500.

|          | mean  | se_mean | sd   | 2.5%  | 25%   | 50%   | 75%   | 97.5% | n_eff | Rhat |
|----------|-------|---------|------|-------|-------|-------|-------|-------|-------|------|
| beta[1]  | -1.32 | 0.01    | 0.09 | -1.49 | -1.39 | -1.33 | -1.26 | -1.15 | 3105  | 1.01 |
| beta[2]  | -1.27 | 0.01    | 0.08 | -1.41 | -1.32 | -1.27 | -1.22 | -1.10 | 830   | 1.01 |
| beta[3]  | -0.74 | 0.01    | 0.09 | -0.91 | -0.80 | -0.74 | -0.69 | -0.56 | 800   | 1.01 |
| beta[4]  | -2.84 | 0.01    | 0.16 | -3.16 | -2.93 | -2.83 | -2.73 | -2.55 | 3547  | 1.00 |
| beta[5]  | -0.69 | 0.01    | 0.09 | -0.86 | -0.75 | -0.69 | -0.63 | -0.51 | 1970  | 1.01 |
| beta[6]  | 1.89  | 0.01    | 0.17 | 1.58  | 1.77  | 1.88  | 2.00  | 2.24  | 1980  | 1.01 |
| beta[7]  | 1.49  | 0.01    | 0.10 | 1.31  | 1.43  | 1.49  | 1.55  | 1.71  | 3680  | 1.01 |
| beta[8]  | -2.34 | 0.00    | 0.14 | -2.62 | -2.44 | -2.34 | -2.25 | -2.09 | 4500  | 1.00 |
| beta[9]  | -0.23 | 0.01    | 0.08 | -0.37 | -0.29 | -0.23 | -0.18 | -0.05 | 670   | 1.01 |
| beta[10] | -1.57 | 0.01    | 0.08 | -1.73 | -1.62 | -1.57 | -1.51 | -1.41 | 1904  | 1.01 |
| beta[11] | -1.19 | 0.01    | 0.08 | -1.33 | -1.24 | -1.19 | -1.14 | -1.02 | 1703  | 1.01 |
| beta[12] | -0.38 | 0.01    | 0.09 | -0.56 | -0.44 | -0.38 | -0.32 | -0.18 | 2000  | 1.01 |
| beta[13] | -0.60 | 0.01    | 0.08 | -0.76 | -0.66 | -0.60 | -0.55 | -0.42 | 800   | 1.01 |
| beta[14] | 0.50  | 0.01    | 0.11 | 0.31  | 0.43  | 0.49  | 0.57  | 0.72  | 950   | 1.01 |
| beta[15] | -1.07 | 0.01    | 0.08 | -1.20 | -1.12 | -1.07 | -1.02 | -0.91 | 706   | 1.01 |
| beta[16] | 0.85  | 0.01    | 0.09 | 0.69  | 0.79  | 0.84  | 0.90  | 1.05  | 506   | 1.01 |
| beta[17] | 0.18  | 0.01    | 0.08 | 0.02  | 0.12  | 0.17  | 0.23  | 0.37  | 640   | 1.01 |
| beta[18] | -0.60 | 0.01    | 0.08 | -0.75 | -0.65 | -0.60 | -0.55 | -0.42 | 703   | 1.01 |
| beta[19] | 0.96  | 0.01    | 0.09 | 0.80  | 0.89  | 0.95  | 1.02  | 1.17  | 604   | 1.01 |
| beta[20] | 0.39  | 0.01    | 0.08 | 0.24  | 0.33  | 0.38  | 0.44  | 0.56  | 590   | 1.01 |
| beta[21] | 0.12  | 0.01    | 0.09 | -0.03 | 0.06  | 0.12  | 0.18  | 0.31  | 730   | 1.01 |
| beta[22] | -0.31 | 0.00    | 0.13 | -0.57 | -0.40 | -0.31 | -0.23 | -0.06 | 3500  | 1.00 |
| beta[23] | -1.09 | 0.01    | 0.10 | -1.28 | -1.15 | -1.09 | -1.02 | -0.88 | 1025  | 1.00 |
| beta[24] | -1.54 | 0.01    | 0.08 | -1.70 | -1.60 | -1.54 | -1.49 | -1.37 | 1900  | 1.01 |
| beta[25] | -0.95 | 0.01    | 0.08 | -1.09 | -1.00 | -0.95 | -0.90 | -0.78 | 707   | 1.01 |
| beta[26] | -0.80 | 0.01    | 0.08 | -0.95 | -0.85 | -0.80 | -0.75 | -0.62 | 705   | 1.01 |
| beta[27] | -1.23 | 0.01    | 0.09 | -1.40 | -1.28 | -1.23 | -1.17 | -1.05 | 112   | 1.01 |
| beta[28] | 1.15  | 0.01    | 0.10 | 0.97  | 1.08  | 1.14  | 1.21  | 1.37  | 701   | 1.01 |
| beta[29] | -0.01 | 0.01    | 0.08 | -0.15 | -0.07 | -0.02 | 0.03  | 0.16  | 508   | 1.01 |
| beta[30] | -0.62 | 0.01    | 0.07 | -0.75 | -0.67 | -0.62 | -0.58 | -0.46 | 610   | 1.01 |
| beta[31] | 0.95  | 0.01    | 0.10 | 0.76  | 0.88  | 0.94  | 1.01  | 1.17  | 810   | 1.01 |
| beta[32] | 0.49  | 0.01    | 0.09 | 0.34  | 0.44  | 0.49  | 0.54  | 0.68  | 606   | 1.01 |

Estimaciones parámetro de localización de los trazos latentes

Inference for Stan model: triam.

3 chains, each with iter=3000; warmup=1500; thin=1;

post-warmup draws per chain=1500, total post-warmup draws=4500.

|           | mean  | se_mean | sd   | 2.5%  | 25%   | 50%   | 75%   | 97.5% | n_eff | Rhat |
|-----------|-------|---------|------|-------|-------|-------|-------|-------|-------|------|
| mu_theta2 | -0.08 | 0.00    | 0.15 | -0.39 | -0.18 | -0.07 | 0.03  | 0.19  | 2573  | 1.00 |
| mu_theta3 | -0.67 | 0.01    | 0.08 | -0.81 | -0.72 | -0.67 | -0.62 | -0.52 | 207   | 1.01 |
| mu_theta4 | 1.60  | 0.01    | 0.11 | 1.38  | 1.53  | 1.60  | 1.68  | 1.82  | 169   | 1.01 |

Estimaciones parámetros de asimetría de los trazos latentes

Inference for Stan model: triam.

3 chains, each with iter=3000; warmup=1500; thin=1;

post-warmup draws per chain=1500, total post-warmup draws=4500.

|        | mean  | se_mean | sd   | 2.5%  | 25%   | 50%   | 75%   | 97.5% | n_eff | Rhat |
|--------|-------|---------|------|-------|-------|-------|-------|-------|-------|------|
| gamma1 | -0.79 | 0.02    | 0.12 | -1.03 | -0.87 | -0.79 | -0.71 | -0.56 | 530   | 1.01 |
| gamma2 | -0.05 | 0.00    | 0.20 | -0.42 | -0.19 | -0.06 | 0.08  | 0.33  | 2542  | 1.00 |
| gamma3 | -0.85 | 0.00    | 0.13 | -1.11 | -0.93 | -0.84 | -0.76 | -0.60 | 953   | 1.00 |
| gamma4 | 0.25  | 0.00    | 0.10 | 0.05  | 0.18  | 0.25  | 0.32  | 0.45  | 4500  | 1.00 |

## APÉNDICE

---

---

### Apéndice C. Resultados Estimaciones individuales

---

---

TABLA 1. Estimaciones individuales parámetro de discriminación

| ítem | alphaCol | alphaAr | alphaVen | alphaCub |
|------|----------|---------|----------|----------|
| 1    | 0.74     | 0.75    | 0.93     | 0.91     |
| 2    | 0.81     | 0.88    | 0.91     | 0.87     |
| 3    | 0.86     | 0.71    | 0.62     | 0.94     |
| 4    | 0.84     | 0.89    | 1.24     | 1.02     |
| 5    | 0.41     | 0.73    | 0.52     | 0.89     |
| 6    | 0.46     | 0.46    | 0.41     | 0.42     |
| 7    | 0.62     | 0.59    | 0.47     | 0.88     |
| 8    | 0.88     | 0.65    | 0.99     | 0.92     |
| 9    | 0.60     | 0.66    | 0.65     | 0.87     |
| 10   | 0.97     | 0.81    | 1.04     | 0.95     |
| 11   | 0.84     | 0.65    | 0.90     | 1.03     |
| 12   | 0.52     | 0.58    | 0.53     | 0.90     |
| 13   | 1.04     | 0.98    | 0.90     | 0.85     |
| 14   | 0.35     | 0.49    | 0.39     | 0.72     |
| 15   | 1.14     | 1.13    | 0.98     | 1.17     |
| 16   | 0.91     | 1.08    | 0.67     | 1.09     |
| 17   | 0.74     | 0.86    | 0.74     | 0.76     |
| 18   | 1.20     | 1.26    | 0.92     | 1.06     |
| 19   | 0.89     | 0.73    | 0.50     | 0.91     |
| 20   | 1.10     | 1.28    | 0.60     | 1.13     |
| 21   | 1.12     | 1.38    | 0.85     | 0.91     |
| 22   | 0.44     | 0.64    | 0.70     | 0.70     |
| 23   | 1.04     | 0.90    | 0.94     | 0.81     |
| 24   | 0.84     | 0.97    | 0.88     | 0.94     |
| 25   | 0.93     | 1.15    | 0.74     | 0.66     |
| 26   | 0.52     | 0.61    | 0.59     | 0.74     |
| 27   | 0.66     | 0.92    | 0.84     | 0.99     |
| 28   | 0.66     | 0.65    | 0.54     | 0.73     |
| 29   | 1.10     | 1.02    | 0.70     | 0.76     |
| 30   | 1.10     | 1.05    | 0.74     | 0.90     |
| 31   | 0.53     | 0.61    | 0.46     | 0.65     |
| 32   | 0.59     | 0.74    | 0.52     | 0.83     |

TABLA 2. Estimaciones individuales parámetro de dificultad

| ítem | alphaCol | alphaAr | alphaVen | alphaCub |
|------|----------|---------|----------|----------|
| 1    | 0.74     | 0.75    | 0.93     | 0.91     |
| 2    | 0.81     | 0.88    | 0.91     | 0.87     |
| 3    | 0.86     | 0.71    | 0.62     | 0.94     |
| 4    | 0.84     | 0.89    | 1.24     | 1.02     |
| 5    | 0.41     | 0.73    | 0.52     | 0.89     |
| 6    | 0.46     | 0.46    | 0.41     | 0.42     |
| 7    | 0.62     | 0.59    | 0.47     | 0.88     |
| 8    | 0.88     | 0.65    | 0.99     | 0.92     |
| 9    | 0.60     | 0.66    | 0.65     | 0.87     |
| 10   | 0.97     | 0.81    | 1.04     | 0.95     |
| 11   | 0.84     | 0.65    | 0.90     | 1.03     |
| 12   | 0.52     | 0.58    | 0.53     | 0.90     |
| 13   | 1.04     | 0.98    | 0.90     | 0.85     |
| 14   | 0.35     | 0.49    | 0.39     | 0.72     |
| 15   | 1.14     | 1.13    | 0.98     | 1.17     |
| 16   | 0.91     | 1.08    | 0.67     | 1.09     |
| 17   | 0.74     | 0.86    | 0.74     | 0.76     |
| 18   | 1.20     | 1.26    | 0.92     | 1.06     |
| 19   | 0.89     | 0.73    | 0.50     | 0.91     |
| 20   | 1.10     | 1.28    | 0.60     | 1.13     |
| 21   | 1.12     | 1.38    | 0.85     | 0.91     |
| 22   | 0.44     | 0.64    | 0.70     | 0.70     |
| 23   | 1.04     | 0.90    | 0.94     | 0.81     |
| 24   | 0.84     | 0.97    | 0.88     | 0.94     |
| 25   | 0.93     | 1.15    | 0.74     | 0.66     |
| 26   | 0.52     | 0.61    | 0.59     | 0.74     |
| 27   | 0.66     | 0.92    | 0.84     | 0.99     |
| 28   | 0.66     | 0.65    | 0.54     | 0.73     |
| 29   | 1.10     | 1.02    | 0.70     | 0.76     |
| 30   | 1.10     | 1.05    | 0.74     | 0.90     |
| 31   | 0.53     | 0.61    | 0.46     | 0.65     |
| 32   | 0.59     | 0.74    | 0.52     | 0.83     |

## APÉNDICE

---

---

### Apéndice D. Código modelo TRIAM en Rstan

---

---

A continuación se muestra el código implementado en Rstan para estimar el modelo TRIAM

```
data {
  int<lower=0> G;// número de grupos o subpoblaciones
  int<lower=0> N;// número total de estudiantes
  int<lower=0> mu_col;// media col
  int<lower=0> sigma;// param escala
  int<lower=0> n_g[G];//tamaño de subpoblaciones
  int<lower=0> J;// número total de ítems
  int<lower=0,upper=1> Y1[(n_g[1]),J];
  int<lower=0,upper=1> Y2[(n_g[2]),J];
  int<lower=0,upper=1> Y3[(n_g[3]),J];
  int<lower=0,upper=1> Y4[(n_g[4]),J];
  //int<lower=0,upper=1> Y[N,J];
}
parameters {
  vector[(n_g[1])] theta1;
  vector[(n_g[2])] theta2;
  vector[(n_g[3])] theta3;
  vector[(n_g[4])] theta4;
  vector[J] alpha_a;
  vector[J] beta;
  real mu_beta;
  real mu_theta2;
  real mu_theta3;
  real mu_theta4;
  real gamma1;
  real gamma2;
  real gamma3;
  real gamma4;
  real<lower=0> sigma_beta;
}
```

```

transformed parameters{
vector[J] alpha;
for(j in 1:J){
alpha[j]=log(1+exp(alpha_a[j]));
}
}
model {
theta1 ~ skew_normal(mu_col,sigma, gamma1);
gamma1 ~ normal(0,1);

theta2 ~ skew_normal(mu_theta2,sigma, gamma2);
mu_theta2 ~ normal(0,0.25);
gamma2 ~ normal(0,1);

theta3 ~ skew_normal(mu_theta3,sigma, gamma3);
mu_theta3 ~ normal(0,0.25);
gamma3 ~ normal(0,1);

theta4 ~ skew_normal(mu_theta4,sigma, gamma4);
mu_theta4 ~ normal(0,0.25);
gamma4 ~ normal(0,1);

alpha_a ~ normal(0,0.25);
beta ~ normal(mu_beta,sigma_beta);
mu_beta ~ normal(0,0.25);
sigma_beta ~ cauchy(0.01,0.01);

{for(i in 1:(n_g[1])){
for(j in 1:J){
Y1[i,j] ~ bernoulli_logit((alpha[j]*(theta1[i] - beta[j])));
}}}
{for(i in 1:(n_g[2])){
for(j in 1:J){
Y2[i,j] ~ bernoulli_logit((alpha[j]*(theta2[i] - beta[j])));
}}}
{for(i in 1:(n_g[3])){
for(j in 1:J){
Y3[i,j] ~ bernoulli_logit((alpha[j]*(theta3[i] - beta[j])));
}}}
{for(i in 1:(n_g[4])){
for(j in 1:J){
Y4[i,j] ~ bernoulli_logit((alpha[j]*(theta4[i] - beta[j])));
}}}
}

```

Adicional al archivo anterior se debe implementar el siguiente archivo en R

```
setwd("C:/Rtesis/triam")
```

```
#####
#CARGANDO STAN
#####

library(rstan)
rstan_options(auto_write= TRUE)
options(mc.cores = parallel::detectCores())
library(loo)
library(writexl)
library(readxl)

#####
#CARGANDO DATOS
#####

datos1 <- read_excel("MUESTRA_spn.xlsx",sheet = "COLM", col_names = FALSE)
datos2 <- read_excel("MUESTRA_spn.xlsx",sheet = "ARGM", col_names = FALSE)
datos3 <- read_excel("MUESTRA_spn.xlsx",sheet = "VENM", col_names = FALSE)
datos4 <- read_excel("MUESTRA_spn.xlsx",sheet = "CUBM", col_names = FALSE)
#####
#TAMAÑOS DE LAS SUBPOBLACIONES
#####
n1=dim(datos1)[1]
n2=dim(datos2)[1]
n3=dim(datos3)[1]
n4=dim(datos4)[1]

n_g=c(n1,n2,n3,n4)

N = n1+n2+n3+n4
G= length(n_g)
J=dim(datos1)[2]
mu_col=0 #media fija
sigma= 1 # Parámetro de escala
#####
#DATOS PARA STAN
#####

data_irt = list(n_g = n_g,G=G,N = I,mu_col=0,sigma= 1,J=J,Y1=datos1,
Y2=datos2,Y3=datos3,Y4=datos4,Y=datos)
#####
#####
#MODELO STAN
#####

irt_2pltriam = stan(file = 'triam13.stan', data = data_irt, iter = 7000,
chains = 3)
#####
#DIAGNÓSTICO GRAFICO DE CONVERGENCIA
```

```
#####  
pairs(irt_2pltriam, pars="alpha")  
#####  
#EXTRACCIÓN PARÁMETROS  
#####  
print(irt_2pltriam, par="theta2")  
print(irt_2pltriam, par="alpha")  
print(irt_2pltriam, par=c("gamma1", "gamma2", "gamma3", "gamma4"))  
print(irt_2pltriam, par="beta")  
print(irt_2pltriam, par=c("mu_theta2", "mu_theta3", "mu_theta4"))  
  
#####  
#GRAFICA CADENAS  
#####  
traceplot(irt_2pltriam, pars= c("theta1[3]", "theta1[17]", "theta1[25]",  
"theta1[39]", "theta1[389]"), inc_warmup = FALSE)  
traceplot(irt_2pltriam, pars= c("beta[7]"), inc_warmup = FALSE)  
  
traceplot(irt_2pltriam, pars= c("gamma4"), inc_warmup = FALSE)
```

## APÉNDICE

---

---

### Apéndice E. Resumen estadísticas a posteriori de estimaciones modelo TRIAM en Rstan

---

---

A continuación se muestra un resumen de las estadísticas a posteriori de los parámetros:  $\beta_j$ : dificultad del ítem,  $\theta_g$ : trazos latentes,  $\mu_g$ : parámetro de localización de los trazos latentes

Inference for Stan model: triam.

3 chains, each with iter=5000; warmup=2500; thin=1;

post-warmup draws per chain=2500, total post-warmup draws=7500.

|          | mean  | se_mean | sd   | 2.5%  | 25%   | 50%   | 75%   | 97.5% | n_eff | Rhat |
|----------|-------|---------|------|-------|-------|-------|-------|-------|-------|------|
| beta[1]  | -1.11 | 0.01    | 0.08 | -1.26 | -1.16 | -1.11 | -1.06 | -0.96 | 2108  | 1.01 |
| beta[2]  | -1.32 | 0.01    | 0.08 | -1.46 | -1.37 | -1.32 | -1.27 | -1.16 | 2290  | 1.01 |
| beta[3]  | -0.44 | 0.01    | 0.08 | -0.58 | -0.49 | -0.44 | -0.39 | -0.28 | 1702  | 1.00 |
| beta[4]  | -2.31 | 0.00    | 0.11 | -2.52 | -2.38 | -2.31 | -2.24 | -2.11 | 3098  | 1.01 |
| beta[5]  | -0.56 | 0.01    | 0.08 | -0.71 | -0.61 | -0.56 | -0.50 | -0.39 | 2330  | 1.01 |
| beta[6]  | 1.97  | 0.01    | 0.15 | 1.69  | 1.87  | 1.96  | 2.06  | 2.28  | 3505  | 1.00 |
| beta[7]  | 1.25  | 0.01    | 0.09 | 1.09  | 1.19  | 1.25  | 1.31  | 1.44  | 1097  | 1.01 |
| beta[8]  | -2.01 | 0.00    | 0.11 | -2.22 | -2.08 | -2.01 | -1.94 | -1.80 | 6400  | 1.01 |
| beta[9]  | -0.25 | 0.01    | 0.08 | -0.40 | -0.30 | -0.26 | -0.20 | -0.09 | 1970  | 1.01 |
| beta[10] | -1.58 | 0.00    | 0.08 | -1.73 | -1.63 | -1.58 | -1.52 | -1.42 | 2050  | 1.01 |
| beta[11] | -1.35 | 0.01    | 0.08 | -1.50 | -1.40 | -1.35 | -1.29 | -1.19 | 2408  | 1.01 |
| beta[12] | -0.23 | 0.01    | 0.08 | -0.38 | -0.28 | -0.23 | -0.18 | -0.06 | 2020  | 1.01 |
| beta[13] | -0.45 | 0.01    | 0.08 | -0.59 | -0.50 | -0.45 | -0.40 | -0.29 | 1800  | 1.01 |
| beta[14] | 0.59  | 0.01    | 0.10 | 0.41  | 0.52  | 0.59  | 0.65  | 0.79  | 2702  | 1.00 |
| beta[15] | -0.92 | 0.01    | 0.07 | -1.06 | -0.97 | -0.93 | -0.88 | -0.77 | 1089  | 1.01 |
| beta[16] | 0.64  | 0.01    | 0.08 | 0.49  | 0.58  | 0.63  | 0.69  | 0.80  | 1706  | 1.01 |
| beta[17] | -0.04 | 0.01    | 0.08 | -0.20 | -0.10 | -0.05 | 0.01  | 0.12  | 1807  | 1.01 |
| beta[18] | -0.37 | 0.01    | 0.08 | -0.51 | -0.42 | -0.37 | -0.32 | -0.22 | 1093  | 1.01 |
| beta[19] | 0.96  | 0.01    | 0.09 | 0.79  | 0.90  | 0.96  | 1.02  | 1.15  | 2207  | 1.01 |
| beta[20] | 0.27  | 0.01    | 0.08 | 0.13  | 0.22  | 0.26  | 0.32  | 0.42  | 1700  | 1.01 |
| beta[21] | 0.15  | 0.01    | 0.08 | 0.00  | 0.09  | 0.14  | 0.20  | 0.31  | 1801  | 1.01 |
| beta[22] | -0.20 | 0.01    | 0.09 | -0.38 | -0.26 | -0.20 | -0.14 | -0.02 | 2075  | 1.01 |

|          |       |      |      |       |       |       |       |       |      |      |
|----------|-------|------|------|-------|-------|-------|-------|-------|------|------|
| beta[23] | -0.80 | 0.01 | 0.08 | -0.94 | -0.85 | -0.80 | -0.74 | -0.64 | 2200 | 1.01 |
| beta[24] | -1.48 | 0.00 | 0.08 | -1.63 | -1.53 | -1.48 | -1.43 | -1.32 | 2067 | 1.00 |
| beta[25] | -0.81 | 0.01 | 0.07 | -0.95 | -0.86 | -0.82 | -0.77 | -0.66 | 1840 | 1.01 |
| beta[26] | -0.97 | 0.00 | 0.09 | -1.14 | -1.03 | -0.97 | -0.91 | -0.79 | 3031 | 1.00 |
| beta[27] | -1.13 | 0.01 | 0.08 | -1.29 | -1.19 | -1.13 | -1.08 | -0.97 | 2503 | 1.00 |
| beta[28] | 0.94  | 0.01 | 0.10 | 0.76  | 0.88  | 0.94  | 1.01  | 1.15  | 2708 | 1.01 |
| beta[29] | -0.08 | 0.01 | 0.07 | -0.22 | -0.13 | -0.09 | -0.04 | 0.07  | 1730 | 1.01 |
| beta[30] | -0.81 | 0.01 | 0.07 | -0.94 | -0.86 | -0.81 | -0.76 | -0.66 | 1079 | 1.01 |
| beta[31] | 0.79  | 0.01 | 0.09 | 0.61  | 0.72  | 0.78  | 0.85  | 0.98  | 2704 | 1.00 |
| beta[32] | 0.29  | 0.01 | 0.09 | 0.13  | 0.23  | 0.28  | 0.34  | 0.47  | 2201 | 1.01 |

Samples were drawn using NUTS(diag\_e) at Thu May 31 02:39:42 2018.  
 For each parameter, n\_eff is a crude measure of effective sample size,  
 and Rhat is the potential scale reduction factor on split chains (at  
 convergence, Rhat=1).

Resumen estadísticas a posteriori del parámetro de localización para los trazos latentes de cada subpoblación.

Inference for Stan model: triam.  
 3 chains, each with iter=5000; warmup=2500; thin=1;  
 post-warmup draws per chain=2500, total post-warmup draws=7500.

|           | mean  | se_mean | sd   | 2.5%  | 25%   | 50%   | 75%   | 97.5% | n_eff | Rhat |
|-----------|-------|---------|------|-------|-------|-------|-------|-------|-------|------|
| mu_theta2 | -0.03 | 0.00    | 0.13 | -0.28 | -0.12 | -0.02 | 0.07  | 0.23  | 7500  | 1.01 |
| mu_theta3 | -0.51 | 0.00    | 0.08 | -0.66 | -0.57 | -0.52 | -0.46 | -0.36 | 288   | 1.01 |
| mu_theta4 | 1.70  | 0.01    | 0.10 | 1.51  | 1.64  | 1.70  | 1.77  | 1.91  | 284   | 1.01 |

Resumen estadísticas a posteriori de las estimaciones del parámetro de asimetría para los trazos latentes de cada subpoblación.

Inference for Stan model: triam.  
 3 chains, each with iter=5000; warmup=2500; thin=1;  
 post-warmup draws per chain=2500, total post-warmup draws=7500.

|        | mean   | se_mean | sd   | 2.5%  | 25%   | 50%   | 75%   | 97.5% | n_eff | Rhat |
|--------|--------|---------|------|-------|-------|-------|-------|-------|-------|------|
| gamma1 | -0.603 | 0.02    | 0.11 | -0.81 | -0.68 | -0.61 | -0.53 | -0.38 | 1330  | 1.01 |
| gamma2 | -0.006 | 0.01    | 0.15 | -0.31 | -0.12 | -0.01 | 0.10  | 0.29  | 2542  | 1.00 |
| gamma3 | -0.821 | 0.00    | 0.12 | -1.06 | -0.90 | -0.82 | -0.74 | -0.58 | 953   | 1.00 |
| gamma4 | 0.140  | 0.01    | 0.08 | -0.01 | 0.09  | 0.14  | 0.19  | 0.29  | 4500  | 1.00 |

Resumen estadísticas a posteriori de las estimaciones de los trazos latentes de cada subpoblación.

Estimaciones trazos latentes Colombia:



|             |       |      |      |       |       |       |       |       |      |   |
|-------------|-------|------|------|-------|-------|-------|-------|-------|------|---|
| theta3[96]  | -1.87 | 0.00 | 0.42 | -2.72 | -2.15 | -1.86 | -1.57 | -1.07 | 7500 | 1 |
| theta3[97]  | -1.82 | 0.00 | 0.41 | -2.64 | -2.08 | -1.80 | -1.53 | -1.03 | 7500 | 1 |
| theta3[98]  | -1.82 | 0.00 | 0.41 | -2.67 | -2.10 | -1.82 | -1.54 | -1.06 | 7500 | 1 |
| theta3[99]  | -1.85 | 0.00 | 0.42 | -2.70 | -2.13 | -1.85 | -1.57 | -1.07 | 7500 | 1 |
| theta3[100] | -1.89 | 0.00 | 0.41 | -2.74 | -2.16 | -1.88 | -1.62 | -1.09 | 7500 | 1 |

Estimaciones trazos latentes Cuba

Inference for Stan model: triam.

3 chains, each with iter=5000; warmup=2500; thin=1;

post-warmup draws per chain=2500, total post-warmup draws=7500.

|             | mean  | se_mean | sd   | 2.5%  | 25%   | 50%   | 75%   | 97.5% | n_eff | Rhat |
|-------------|-------|---------|------|-------|-------|-------|-------|-------|-------|------|
| theta4[1]   | -1.69 | 0.00    | 0.41 | -2.53 | -1.96 | -1.68 | -1.41 | -0.91 | 7500  | 1    |
| theta4[2]   | -1.75 | 0.00    | 0.42 | -2.61 | -2.03 | -1.74 | -1.47 | -0.93 | 7500  | 1    |
| theta4[3]   | -1.69 | 0.00    | 0.42 | -2.52 | -1.96 | -1.68 | -1.41 | -0.89 | 7500  | 1    |
| theta4[4]   | -0.97 | 0.00    | 0.39 | -1.74 | -1.22 | -0.97 | -0.71 | -0.21 | 7500  | 1    |
| theta4[5]   | -1.19 | 0.00    | 0.38 | -1.95 | -1.45 | -1.19 | -0.93 | -0.47 | 7500  | 1    |
| ...         |       |         |      |       |       |       |       |       |       |      |
| theta4[97]  | 0.68  | 0.00    | 0.43 | -0.14 | 0.39  | 0.68  | 0.97  | 1.58  | 7500  | 1    |
| theta4[98]  | 0.76  | 0.00    | 0.43 | -0.07 | 0.47  | 0.75  | 1.04  | 1.64  | 7500  | 1    |
| theta4[99]  | 0.74  | 0.00    | 0.43 | -0.08 | 0.45  | 0.74  | 1.03  | 1.62  | 7500  | 1    |
| theta4[100] | 0.51  | 0.00    | 0.42 | -0.29 | 0.22  | 0.50  | 0.78  | 1.34  | 7500  | 1    |

---

---

## Bibliografía

---

---

- [1] Albert, J. H. (1992), *Bayesian estimation of normal ogive item response curves using Gibbs sampling*. Journal of educational statistics, 17(3), 251–269.
- [2] Antonio, D. & Montenegro, A.(2014), *Modelo Multidimensional de Teoría de Respuesta al Item Jerárquico*. In CONBRATRI Congresso Brasileiro de Teoria da Resposta ao Item (No. 1, pp. 1–30).
- [3] Azzalini, A. (1985)., *A class of distributions which includes the normal ones* . Scandinavian journal of statistics, 171–178.
- [4] Azevedo, C. L. N., Andrade, D. F., & Fox, J. P. (2009)., *sampling estimation procedure and model-fit for multiple groups IRT model* . Technical Report.
- [5] Azevedo, C. L. N., Fox, J. P., & Andrade, D. F. (2009), *Gibbs sampling for Bayesian estimation and model fit assessment for a longitudinal IRT model*. Technical Report.
- [6] Azevedo, C. L., Andrade, D. F., & Fox, J. P. (2012)., *A Bayesian generalized multiple group IRT model with model-fit assessment tools*. Computational statistics & data analysis, 56(12), 4399–4412.
- [7] Azevedo, C. L., Bolfarine, H., & Andrade, D. F. (2011). , *Bayesian inference for a skew-normal IRT model under the centred parameterization* . Computational Statistics & Data Analysis, 55(1), 353–365.
- [8] Bazán, J. L., Branco, M. D., & Bolfarine, H. (2006), *A skew item response model* . Bayesian analysis, 1(4), 861–892.
- [9] Bazán, J. L., Branco, M. D., & Bolfarine, H. (2014). , *Extensions of the skew-normal ogive item response model*. Brazilian Journal of Probability and Statistics, 1–23.
- [10] Bock, R. D., & Zimowski, M. F. (1997). , *Multiple group IRT*. In Handbook of modern item response theory (pp. 433–448). Springer, New York, NY.
- [11] Casassus & Cusato, S. (1998), *Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados en tercero y cuarto grado* . Santiago de Chile, LLECE UNESCO–Santiago.
- [12] Casassus, J., Cusato, S., Froemel, J., & Palafox, J. (2001). , *Primer estudio internacional comparativo sobre lenguaje, matemáticas y factores asociados, para alumnos de tercer y cuarto*. Santiago de Chile, LLECE UNESCO–Santiago.

- 
- [13] Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ? & Riddell, A. (2017). , *Stan: A probabilistic programming language* . Journal of statistical software, 76(1).
- [14] Equipo desarrollador de Stan (2017), *Stan modeling language users guide and reference manual*. (Version 2.17.0). Retrieved from <http://mc-stan.org/documentation/>
- [15] Fox, J. P. (2010). , *Bayesian item response modeling: Theory and applications* . Springer Science & Business Media.
- [16] Gelman, A., Meng, X. L., & Stern, H. (1996)., *Posterior predictive assessment of model fitness via realized discrepancies*. Statistica sinica, 733–760.
- [17] Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014) , *Bayesian analysis*.
- [18] Gelman, A., & Hill, J. (2007), *Data analysis using regression and multilevelhierarchical models* . (Vol. 1). New York, NY, USA: Cambridge University Press.
- [19] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013)., *Bayesian data analysis*. CRC press.
- [20] Gelman, A., & Rubin, D. B. (1992). , *Inference from iterative simulation using multiple sequences*. Statistical science, 457-472.
- [21] Hashimoto, Y. (2002). , *Motivation and willingness to communicate as predictors of reported l2 use: The Japanese ESL context*. Second Language Studies 20, 29–70.
- [22] Hoffman, M. D., & Gelman, A. (2014). , *The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo*. Journal of Machine Learning Research, 15(1), 1593–1623.
- [23] Kirisci, L., Hsu, T. C., & Yu, L. (2001)., *Robustness of item parameter estimation programs to assumptions of unidimensionality and normality*. Applied psychological measurement, 25(2), 146–162.
- [24] Luo, Y., & Jiao, H. (2017). , *Using the Stan Program for Bayesian Item Response Theory*. Educational and Psychological Measurement, 0013164417693666.
- [25] Micceri, T. (1989). , *The unicorn, the normal curve, and other improbable creatures*. Psychological bulletin, 105(1), 156.
- [26] Muñoz, J. (2010). , *Las teorías de los tests: teoría clásica y teoría de respuesta a los ítems*. Papeles del psicólogo, 31(1).
- [27] Neal, R. M. (2011)., *MCMC using Hamiltonian dynamics*. In S. Brooks, A. Gelman, & X.-L.Meng (Eds.), *Handbook of Markov Chain Monte Carlo* . (Vol. 2, pp. 113–162). New York, NY: CRC Press.
- [28] Patz, R. J., & Junker, B. W. (1999), *A straightforward approach to Markov chain Monte Carlo methods for item response models*. Journal of educational and behavioral Statistics, 24(2), 146–178.
- [29] Reckase, M. (2009)., *Multidimensional item response theory*. (Vol. 150). New York, NY: Springer.

- 
- [30] Riddle, A. S., Blais, M. R., & Hess, U. (2002), *A multi-group investigation of the CES-D's measurement structure across adolescents, young adults and middle-aged adults*. *Cirano*.
- [31] Sahu, S. K. (2002). , *Bayesian estimation and model choice in item response models*. *Journal of Statistical Computation and Simulation* 72, 217–232. MR1909259
- [32] Samejima, F. (1997), *Departure from normal assumptions: A promise for future psychometrics with substantive mathematical modeling*. *Psychometrika* 62, 471–493.
- [33] Samejima, F. (2000). , *Logistic positive exponent family of models: Virtue of asymmetric item characteristics curves*. *Psychometrika* 65, 319–335.
- [34] Santos, J. R., Azevedo, C. L., & Bolfarine, H. (2013), *A multiple group item response theory model with centered skew-normal latent trait distributions under a Bayesian framework*. *Journal of Applied Statistics*, 40(10), 2129–2149.
- [35] Curran, D. P., & Seong, C. M. (1990)., *Atom-transfer addition, annulation, and macrocyclization reactions of iodomalnonitriles*. *Journal of the American Chemical Society*, 112(25), 9401–9403.
- [36] Vehtari, A., Gelman, A., & Gabry, J. (2017)., *Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC*. *Statistics and Computing*, 27(5), 1413–1432.
- [37] Watanabe, S. (2010), *Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory*. *Journal of Machine Learning Research*, 11(Dec), 3571–3594.
- [38] Zaider, T. I., Heimberg, R. G., Fresco, D. M., Schneier, F. R. and Liebowitz, M. R. (2003), *Evaluation of the Clinical Global Impression Scale among individuals with social anxiety disorder*. *Psychological Medicine* 33, 611–622.