# Propuesta de modelado multigrupo y longitudinal en la Teoría de Respuesta al Item para estimar la habilidad en Matemática de los alumnos de la UNLPam

XVII Congreso Dr. Antonio Monteiro

7 al 9 de junio de 2023

Departamento de Matemática - Universidad Nacional del Sur

María Virginia Piergentili & María Cristina Martín

# Introducción

La **habilidad en Matemática** puede definirse como la capacidad de obtener, procesar y retener información matemática (Krutetskii 1976, Vilkomir y ODonoghue 2009 ) o como la capacidad de aprender y dominar nuevas ideas matemáticas (Koshy et al., 2009 ) (citado en Karsenty 2014)

El trabajo realizado por Roldán (2022) se planteaba como objetivo estudiar la habilidad en Matemática de los alumnos inscriptos en alguna de las carreras de la Facultad de Ciencias Exactas y Naturales (FCEyN) de la Universidad Nacional de La Pampa (UNLPam) que requieren del cursado de al menos una asignatura referida a una Matemática introductoria y que no debieron aprobar ningún exámen de ingreso a la carrera para comenzar a cursar la asignatura.

# Introducción

La habilidad en matemática es un constructo inobservable que requiere ser estudiado a partir de otras variables.

En la actualidad, la Teoría de Respuesta al Ítem (TRI) es una de las herramientas utilizada para estudiar este tipo de variables no observadas, también llamadas, **variables latentes** utilizando cuestionarios o tests, que determinan las llamadas **variables manifiestas** (ítems).

Para este estudio se aplicó un test denominado Prueba Diagnótico en Matemática. Dicha prueba estaba compuesta por 10 ítems con una sola respuesta correcta entre cuatro posibles.

Apéndice A

# Prueba Diagnóstico en Matemática

**Ítem 1**

Indique el resultado de la siguiente operación:

$$\sqrt{36+64}\cdot(-2)+3^2$$

○ $-11$ ○ 19 ○ 37 ○ 70

**Ítem 2**

¿Cuál debe ser el valor de $x$ para que se cumpla la igualdad?

$$8-6x=32$$

○ 16 ○ 4 ○ $-4$ ○ 30

# Introducción

**Variables manifiestas:** Resultado obtenido en el $i$-ésimo ítem($i = 1, 2, \ldots, 10$) de la PDM 2020 por un estudiante inscripto para cursar, en el año 2020, alguna Matemática introductoria que se dicta en la FCEyN de la UNLPam.

- Cualitativa nominal - dicotómica.
- Nivel de medición: Nominal.
- $U_{ji} = \begin{cases} 1 & \text{si la respuesta del individuo } j \text{ al ítem } i \text{ es correcta} \\ 0 & \text{si la respuesta del individuo } j \text{ al ítem } i \text{ es incorrecta} \end{cases}$

**Variable latente:** Habilidad en Matemática de un estudiante inscripto para cursar, en el año 2020, alguna Matemática introductoria que se dicta en la FCEyN de la UNLPam.

- Cuantitativa continua.
- Nivel de medición: De intervalos.
- $\theta \sim N(\mu_\theta, \sigma^2_\theta)$

# Modelo Logístico de 2 parámetros

Suponiendo que $\theta \sim N(0,1)$ y empleando la Estimación por Máxima Verosimilitud Marginal (EMVM), resultó el Modelo Logístico de 2 parámetros (ML2) el más apropiado para estimar la habilidad en Matemática.

Sean

- $j$ : subíndice asociado al individuo, $j = 1, \ldots, n$
- $i$ : subíndice asociado al ítem, $i = 1, \ldots, k$
- $\theta_j$ : habilidad (rasgo latente) del $j$-ésimo individuo
- $\varsigma_i = (a_i, b_i)$ : vector de parámetros del ítem $i$ donde
  $a_i$ es el parámetro de discriminación
  $b_i$ es el parámetro de dificultad
- $P_{ji} = P(U_{ji} = 1 | \theta_j, \varsigma_i)$ es la probabilidad de que el individuo $j$ con habilidad $\theta_j$ responda correctamente al ítem $i$.

# Modelo Logístico de 2 parámetros

## Modelo Logístico de 2 parámetros

- Componente aleatoria:
$$U_{ji} \sim Ber(P_{ji}) \qquad j = 1, \ldots, n; \;\; i = 1, \ldots, k$$
- Componente sistemática:
$$\upsilon_{ji} = a_i(\theta_j - b_i) \qquad j = 1, \ldots, n; \;\; i = 1, \ldots, k$$
- Función de enlace:
$$logit(P_{ji}) = \upsilon_{ji} \qquad j = 1, \ldots, n; \;\; i = 1, \ldots, k$$

Al tratarse de un Modelo Logístico de dos parámetros, la Curva Característica del Ítem está representada por:

$$P_{ji}(\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}$$

# Motivación

La hipótesis corroborada por el trabajo de Roldán fue que: *"La mayoría de los estudiantes que ingresan a la FCEyN de la UNLPam no poseen conocimientos en Matemática suficientes para un cursado satisfactorio y posterior aprobación de las asignaturas relacionadas con una Matemática introductoria"*.

Muchos de los casos de abandono universitario en la FCEyN de la UNLPam se corresponden con la dificultad en el aprendizaje de los saberes básicos que hacen a la Matemática(Roldán 2022).

Del análisis descriptivo se observó que la mayoría de los estudiantes tenían en general una estimación muy baja de la habilidad, lo cual apoya la hipótesis de insuficiencia en el nivel medio de Matemática requerido para cursar de manera satisfactoria las materias relacionadas con esta Ciencia Exacta.

# Motivación

Se recolectó información de ciertas características: ingresantes o recursantes, sexo, carrera, lugar de origen(localidad y pcia). Lo que se hizo con estas variables fue obtener las medidas resumen para la variable habilidad en matemática, por ejemplo, discriminada por carrera.

Fueron identificados aquellos estudiantes que necesitan un mayor acompañamiento para el aprendizaje de la disciplina, sin embargo, no se realizó un seguimiento. De esta identificación, por ejemplo, se observa que prácticamente la mitad de los aspirantes a Ingeniería en Recursos Naturales y Medio Ambiente presenta niveles muy bajos de Habilidad en Matemática.

Si bien existen los casos particulares interesa estudiar si hay algún grupo de estudiantes que muestra menor habilidad en Matemática teniendo en cuenta algunas de las características observadas de los estudiantes.

Todo esto hace pensar que la habilidad de los alumnos podría ser estudiada a lo largo del tiempo, y además, considerando distintos grupos.

# Modelo Longitudinal

La habilidad en Matemática, por lo que significa, es una variable que tiene sentido estudiar a lo largo del tiempo. El objetivo de este modelo es estimar la habilidad en Matemática para un grupo de individuos a través de varias condiciones de evaluación(tests, tiempos).

Se asume que el vector de habilidades de la población en las $T$ condiciones de evaluación $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_T)'$ tienen distribución continua multivariada con vector de parámetros $\boldsymbol{\eta}$ y fdp $g(\boldsymbol{\theta}|\boldsymbol{\eta})$.

Considerando además:

- $t$ : subíndice asociado al test, $t = 1, \ldots, T$
- $\theta_{jt}$ : habilidad (rasgo latente) del $j$-ésimo individuo en el test $t$
- $U_{jit} = \begin{cases} 1 & \text{si la respuesta del individuo } j \text{ al ítem } i \text{ del test } t \text{ es correcta} \\ 0 & \text{si la respuesta del individuo } j \text{ al ítem } i \text{ del test } t \text{ es incorrecta} \end{cases}$
- $P_{jit} = P(U_{jit} = 1|\theta_{jt}, \varsigma_i)$ es la probabilidad de que el individuo $j$ con habilidad $\theta_j$ responda correctamente al ítem $i$ del test $t$.

# Modelo Logístico longitudinal de 2 parámetros

El test $t$ está conformado por $k_t$ ítems y se tienen $k$ ítems en el total $T$ de tests, con $k \leq \sum_{t=1}^{T} k_t$.

Los test que se utilicen deben estar planificados de manera tal que tengan ítems en común tal que sean comparables las escalas de los parámetros de los ítems y poblacionales.

La prueba utilizada por Roldán estaba formado por 10 ítems. Con lo cual, las restantes pruebas que se les administren a los estudiantes deberán incluir algunos de los ítems que conformaban la PDM.

# Modelo Logístico longitudinal de 2 parámetros

## Modelo Logístico longitudinal de 2 parámetros

- Componente aleatoria:

$$U_{jit} \sim Ber(P_{jit}) \qquad j = 1, \ldots, n; \;\; i = 1, \ldots, k; \;\; t = 1, \ldots, T$$

- Componente sistemática:

$$\upsilon_{jit} = a_i(\theta_{jt} - b_i) \qquad j = 1, \ldots, n; \;\; i = 1, \ldots, k; \;\; t = 1, \ldots, T$$

- Función de enlace:

$$logit(P_{jit}) = \upsilon_{jit} \qquad j = 1, \ldots, n; \;\; i = 1, \ldots, k; \;\; t = 1, \ldots, T$$

Al tratarse de un Modelo Logístico longitudinal de dos parámetros, la Curva Característica del Ítem está representada por:

$$P_{jit}(\theta_{jt}) = \frac{1}{1 + e^{-a_i(\theta_{jt} - b_i)}}$$

# Estimación

Como tanto los parámetros poblacionales como los de los ítems son desconocidos, se propone una estimación conjunta de los mismos. Generalmente esa estimación es realizada por el **Método de Máxima Verosimilitud** a través de la aplicación de algún proceso iterativo como Newton-Rapson o Scoring de Fisher con los abordajes usuales, estimación conjunta (parámetros de los ítems y habilidades) o en dos etapas (primero la estimación de los parámetros de los ítems y posteriormente, de las habilidades) (Tavares, 2001).

El cuidado que hay que tener en la estimación conjunta es que, para evitar un problema de inidentificabilidad, se debe adoptará al test $t = 1$ como referencia considerando $\mu_1 = 0$ y $\sigma_1^2 = 1$

El primer paso para la estimación es plantear la función de verosimilitud.

# Estimación

Sean:

- $\boldsymbol{U}_{j.t} = (U_{j1t}, U_{j2t}, \ldots, U_{jn_t t})'$ : vector de respuestas del individuo $j$ en el test $t$
- $\boldsymbol{U}_{j..} = (\boldsymbol{U}'_{j,1}, \boldsymbol{U}'_{j,2}, \ldots, \boldsymbol{U}'_{j.T})'$ : vector de respuestas del individuo $j$ en todos los tests
- $\boldsymbol{U}_{...} = (\boldsymbol{U}'_{1..}, \boldsymbol{U}'_{2..}, \ldots, \boldsymbol{U}'_{n..})'$ : vector de todas las respuestas

Lo natural sería considerar:

$$L(\varsigma, \eta) = P\left(\boldsymbol{U}_{...}|\varsigma, \eta\right) = \prod_{j=1}^{n} P(\boldsymbol{U}_{j..}|\varsigma, \eta)$$

pero cuando el número de individuos es grande con relación al número de ítems, puede haber ventajas computacionales en trabajar con el número de ocurrencias de los diferentes **patrones de respuesta**.

# Estimación

Dado el test $t$ con $k_t$ ítems (donde cada ítem tiene dos posibles respuestas: 0 ó 1) y $k_c$ el número de respuestas para cada individuo, hay un total de

$$S = 2^{k_c} \quad \text{respuestas posibles (patrones de respuesta)}$$

$j$ será ahora el subíndice asociado a un patrón de respuesta en vez de a un individuo.
Sea $r_j$ el número de ocurrencias del patrón de respuesta $j$ y sea $s \leq \min(n, S)$ el número de patrones de respuesta con $r_j > 0$.

Al trabajar con $s$ se excluyen los casos donde todas las respuestas son incorrectas. Resulta:

$$n = \sum_{j=1}^{s} r_j$$

# Estimación

Sea $R = (r_1, r_2, \ldots, r_s)'$ el vector de frecuencia de los patrones de respuesta. Considerando la independencia entre las respuestas de los diferentes individuos, $R$ sigue una distribución multinomial, es decir:

$$P(\boldsymbol{R}|\varsigma, \boldsymbol{\eta}) = \frac{n!}{\prod_{j=1}^{s} r_j!} \prod_{j=1}^{s} (P(\boldsymbol{U}_{j..}|\varsigma, \boldsymbol{\eta}))^{r_j}$$

Luego, resulta:

$$L(\varsigma, \boldsymbol{\eta}) = \frac{n!}{\prod_{j=1}^{s} r_j!} \prod_{j=1}^{s} (P(\boldsymbol{U}_{j..}|\varsigma, \boldsymbol{\eta}))^{r_j}$$

A partir de esto, la log-verosimilitud es:

$$\ln L(\varsigma, \boldsymbol{\eta}) = C + \sum_{j=1}^{s} r_j \ln (P(\boldsymbol{U}_{j..}|\varsigma, \boldsymbol{\eta})) \quad \text{donde } C \text{ es una constante}$$

## Ecuaciones de estimación de los parámetros de los ítems

Las ecuaciones de estimación para los parámetros de los ítems son:

$$\frac{\partial(\ln L(\varsigma, \boldsymbol{\eta}))}{\partial \varsigma_i} = 0; \quad i = 1, 2, \ldots, k.$$

Luego de desarrollar la derivada resulta:

$$\frac{\partial(\ln L(\varsigma, \boldsymbol{\eta}))}{\partial \varsigma_i} = \sum_{j=1}^{s} r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \boldsymbol{\tau}_i} (U_{jit} - P_{it}) \left( \frac{\partial P_{it}}{\partial \varsigma_i} \right) \frac{1}{P_{it} Q_{it}} \right\} \frac{P(\boldsymbol{U}_{j..} | \boldsymbol{\theta}, \eta) g(\boldsymbol{\theta} | \boldsymbol{\eta})}{P(\boldsymbol{U}_{j..} | \varsigma, \boldsymbol{\eta})} d\boldsymbol{\theta}$$

Y se tiene que:

$$\frac{\partial P_{it}}{\partial b_i} = -a_i(1 - c_i) P_{it} Q_{it} \quad \text{y} \quad \frac{\partial P_{it}}{\partial a_i} = (\theta_t - b_i) P_{it} Q_{it}$$

# Estimación

En resumen, las ecuaciones de estimación para los parámetros $a_i$ y $b_i$ son respectivamente:

$$\begin{cases} \displaystyle\sum_{j=1}^{s} r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \boldsymbol{\tau}_i} (U_{jit} - P_{it})(\theta_t - b_i) \right\} \frac{P(\boldsymbol{U}_{j..}|\boldsymbol{\theta}, \eta) g(\boldsymbol{\theta}|\boldsymbol{\eta})}{P(\boldsymbol{U}_{j..}|\varsigma, \boldsymbol{\eta})} d\boldsymbol{\theta} = 0 \\ \displaystyle -a_i \sum_{j=1}^{s} r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \boldsymbol{\tau}_i} (U_{jit} - P_{it}) \right\} \frac{P(\boldsymbol{U}_{j..}|\boldsymbol{\theta}, \eta) g(\boldsymbol{\theta}|\boldsymbol{\eta})}{P(\boldsymbol{U}_{j..}|\varsigma, \boldsymbol{\eta})} d\boldsymbol{\theta} = 0 \end{cases}$$

## Ecuaciones de estimación de los parámetros poblacionales

Las ecuaciones de estimación para los parámetros poblacionales son:

$$\frac{\partial(\ln L(\varsigma, \boldsymbol{\eta}))}{\partial \boldsymbol{\eta}} = 0$$

Luego de desarrollar la derivada resulta:

$$\frac{\partial(\ln L(\varsigma, \boldsymbol{\eta}))}{\partial \boldsymbol{\eta}} = \sum_{j=1}^{s} r_j \int_{\mathbb{R}^T} \left( \frac{\partial}{\partial \boldsymbol{\eta}} \ln g(\boldsymbol{\theta}|\boldsymbol{\eta}) \right) \frac{P(\boldsymbol{U}_{j..}|\boldsymbol{\theta}, \eta) g(\boldsymbol{\theta}|\boldsymbol{\eta})}{P(\boldsymbol{U}_{j..}|\varsigma, \boldsymbol{\eta})} d\boldsymbol{\theta}$$

Para la estimación será necesario proponer una distribución para la habilidad. Varias distribuciones pueden ser propuestas para las habilidades. Una de las distribuciones de mayor interés es la Normal.

# Estimación

Se considera entonces que $\boldsymbol{\theta}$ tiene distribución normal $T$-variada con vector de medias $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_T)'$, matriz de covarianzas $\boldsymbol{\Sigma}$ y
$g(\boldsymbol{\theta}|\boldsymbol{\eta}) = \dfrac{1}{(2\pi)^{T/2}|\Sigma|^{1/2}} \cdot e^{-1/2(\boldsymbol{\theta}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})}$

Propuesto esto, se sabe entonces que los parámetros a estimar son $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$, con lo cual se deben obtener las derivadas del logaritmo de la verosimilitud respecto tanto ambos parámetros y esas derivadas igualarlas a 0 para encontrar la estimación.

Luego:

$$\frac{\partial(\ln L(\varsigma, \boldsymbol{\eta}))}{\partial \boldsymbol{\eta}} = -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\eta}}\left(\ln(|\boldsymbol{\Sigma}|)\right) - \frac{1}{2}\frac{\partial}{\partial \boldsymbol{\eta}}\left[(\boldsymbol{\theta} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu})\right]$$

# Estimación

Para llegar a las ecuaciones de estimación para $\boldsymbol{\eta}$ basta con obtener las expresiones de las derivadas primeras de $\ln(|\boldsymbol{\Sigma}|)$ y de $\boldsymbol{V} = \boldsymbol{\Sigma}^{-1}$ y para hacer esto se consideran distintas estructuras de covarianzas.

- ***Matriz de Covarianza Diagonal***
  Varianzas en la diagonal (no necesariamente iguales) y correlaciones nulas.

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 \end{pmatrix} \equiv \boldsymbol{diag}\left\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_T^2\right\}$$

# Estimación

- ***Matriz de Covarianza Uniforme***
  Todas las varianzas son iguales a $\sigma^2$ en las ecuaciones de las $T$ condiciones de evaluación, y entre cualquier par de habilidades $(\theta_t, \theta_s)$ la correlación es la misma e igual a $\rho\sigma^2$.

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho & \cdots & \sigma^2\rho \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho & \cdots & \sigma^2\rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma^2\rho & \sigma^2\rho & \sigma^2\rho & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix}$$

- ***Matriz de Simetría compuesta***
  Sólo las habilidades en los tiempos inmediatamente anteriores y posteriores a un tiempo particular son dependientes entre sí (iguales las covarianzas). La correlación entre dos tiempos inmediatos es $\rho\sigma^2$.

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma^2 & \sigma^2\rho & 0 & 0 & \cdots & 0 \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho & 0 & \cdots & 0 \\ 0 & \sigma^2\rho & \sigma^2 & \sigma^2\rho & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & \rho & 0 & 0 & \cdots & 0 \\ \rho & 1 & \rho & 0 & \cdots & 0 \\ 0 & \rho & 1 & \rho & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

- ***Matriz de Covarianza de Toeplitz***
  Supone que cualquier par de respuestas que estén igualmente separadas en el tiempo tienen la misma correlación. Esta estructura solo es apropiada cuando las mediciones se realizan en intervalos de tiempo iguales (o aproximadamente iguales).

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \begin{pmatrix}
\sigma^2 & \sigma^2\rho_1 & \sigma^2\rho_2 & \cdots & \sigma^2\rho_{T-1} \\
\sigma^2\rho_1 & \sigma^2 & \sigma^2\rho_1 & \cdots & \sigma^2\rho_{T-2} \\
\sigma^2\rho_2 & \sigma^2\rho_1 & \sigma^2 & \cdots & \sigma^2\rho_{T-3} \\
\vdots & \vdots & & \vdots & \ddots & \vdots \\
\sigma^2\rho_{T-1} & \sigma^2\rho_{T-2} & \sigma^2\rho_{T-3} & \cdots & \sigma^2
\end{pmatrix} \\
&= \sigma^2 \begin{pmatrix}
1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\
\rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\
\rho_2 & \rho_1 & 1 & \cdots & \rho_{T-3} \\
\vdots & \vdots & & \vdots & \ddots & \vdots \\
\rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \cdots & 1
\end{pmatrix}
\end{aligned}
$$

- ***Matriz de Covarianza AR(1)***
  Las varianzas no cambian con el tiempo, pero las correlaciones disminuyen monótonamente.

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \begin{pmatrix}
\sigma^2 & \sigma^2\rho & \sigma^2\rho^2 & \cdots & \sigma^2\rho^{T-1} & \\
\sigma^2\rho & \sigma^2 & \sigma^2\rho & \cdots & \sigma^2\rho^{T-2} & \\
\sigma^2\rho^2 & \sigma^2\rho & \sigma^2 & \cdots & \sigma^2\rho^{T-3} & \\
\vdots & \vdots & & \vdots & \ddots & \vdots \\
\sigma^2\rho^{T-1} & \sigma^2\rho^{T-2} & \sigma^2\rho^{T-3} & \cdots & \sigma^2 &
\end{pmatrix} \\
&= \sigma^2 \begin{pmatrix}
1 & \rho & \rho^2 & \cdots & \rho^{T-1} & \\
\rho & 1 & \rho & \cdots & \rho^{T-2} & \\
\rho^2 & \rho & 1 & \cdots & \rho^{T-3} & \\
\vdots & \vdots & & \vdots & \ddots & \vdots \\
\rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 &
\end{pmatrix}
\end{aligned}
$$

# Estimación

- ***Matriz de Covarianza de Hankel***

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho_{12} & \rho_{12} & \cdots & \rho_{12} & \\ \rho_{12} & \sigma_2^2 & \rho_{12} & \cdots & \rho_{12} & \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ \rho_{12} & \rho_{12} & \rho_{12} & \cdots & \sigma_T^2 & \end{pmatrix}$$

<u>**Observación:**</u> En este trabajo se proponen estructuras de covarianza para las cuales su determinante y su inversa tiene una forma analítica, sin embargo podrían proponerse covarianzas no estructuradas para las cuales puedan utilizarse derivadas numéricas.

# Modelo Multigrupo

En el trabajo de Roldán se realizó un análisis exploratorio para la habilidad en Matemática discriminada por las características registradas y se identificó para aquellos alumnos con baja habilidad en Matemática, cúal era su lugar de procedencia, la carrera en la que se había inscripto, el sexo y si era ingresante o recursante. De este análisis se desprende que sería de interés estimar la habilidad en Matemática a través de un Modelo Multigrupo que considere la carrera a la que ingresa el estudiante.

El planteo del Modelo es análago al Modelo Longitudinal ya que Tavares propuso la incorporación de una componente longitudinal en la TRI a partir del trabajo desarrollado por Bock y Zimowski(1997) para estimar por grupos.

Sea $m$ el subíndice asociado a cada componente del grupo, $m = 1, \dots, M$.

# Modelo Multigrupo

## Modelo Logístico multigrupo de 2 parámetros

- Componente aleatoria:

$$U_{jim} \sim Ber(P_{jim}) \qquad j = 1, \ldots, n; \;\; i = 1, \ldots, k; \;\; m = 1, \ldots, M$$

- Componente sistemática:

$$\upsilon_{jim} = a_i(\theta_{jm} - b_i) \qquad j = 1, \ldots, n; \;\; i = 1, \ldots, k; \;\; m = 1, \ldots, M$$

- Función de enlace:

$$logit(P_{jim}) = \upsilon_{jim} \qquad j = 1, \ldots, n; \;\; i = 1, \ldots, k; \;\; m = 1, \ldots, M$$

Al tratarse de un Modelo Logístico multigrupo de dos parámetros, la Curva Característica del Ítem está representada por:

$$P_{jim}(\theta_{jm}) = \frac{1}{1 + e^{-a_i(\theta_{jm} - b_i)}}$$

# Trabajos futuros

- Construcción de nuevas pruebas que permitan evaluar a los estudiantes en principio, al comenzar la asignatura y otra prueba para evaluarlos antes de finalizado el dicactado de la materia. Estudiar si existe algun porcentaje mínimo de ítems que deben tener en común las pruebas.
- Implementación de las pruebas y aplicación de ambos modelos comparando resultados con los ya obtenidos para una única prueba.
- Proponer un único modelo que estime la habilidad para múltiples grupos en varias condiciones de evaluación, es decir, un único modelo a partir del Modelo Logístico longitudinal y el Modelo Logístico Multigrupo.

# Bibliografía

Bock, R. D. and Zimowski, M. F. (1997). Multiple Group IRT. In W.J. van der Linder e R.K. Hambleton Eds. Handbook of Modern Item Response Theory. New York: Spring-Verlag.

Fitzmaurice, G.M, Laird, N.M y Ware, J.H. (2011). Applied longitudinal analysis (2da edición). Wiley series in Probability and Statistics.

Karsenty, R. (2014). Mathematical Ability. In: Lerman, S. (eds) Encyclopedia of Mathematics Education. Springer, Dordrecht.

Roldán, J. M. (2022). La teoría de respuesta al ítem aplicada a pruebas diagnóstico de ingreso universitario. Tesis de Maestría, Universidad Nacional de Córdoba. Recuperado el 24/10/2022 de: https://rdu.unc.edu.ar/handle/11086/26763

Tavares, H. R. (2001). Teoria da Resposta ao Item para Dados Longitudinais. Tese de Doutorado. Universidade de São Paulo.
Recuperado el 02/02/2023 de:
https://www.teses.usp.br/teses/disponiveis/45/45133/tde-20210729-123610/en.php

# ¡MUCHAS GRACIAS!

virginia.piergentili@uns.edu.ar;
maritamartin11@gmail.com