



FACULTAD DE  
CIENCIAS ECONÓMICAS  
Y DE ADMINISTRACIÓN



UNIVERSIDAD  
DE LA REPÚBLICA  
URUGUAY

UNIVERSIDAD DE LA REPÚBLICA  
FACULTAD DE CIENCIAS ECONÓMICAS Y DE  
ADMINISTRACIÓN

TRABAJO DE FINAL DE GRADO PARA LA  
LICENCIATURA EN ESTADÍSTICA

El Modelo de Rasch desde la perspectiva de  
los Modelos Lineales Generalizados Mixtos,  
un abordaje Clásico y Bayesiano, con aplicación  
a un Test de Comportamiento utilizado en  
identificar el Trastorno del Espectro Autista

Estudiante: Miguel Barreto

Tutores: Dr. R. Álvarez-Vaz y Mag. F. Massa

Julio 2023

Montevideo, Uruguay

Esta página ha sido intencionalmente dejada en blanco.

# Tribunal

Dr. Ramón Álvarez-Vaz ..... FCEA, UdelaR.

Dr. Juan José Goyeneche ..... FCEA, UdelaR.

Mag. Fernando Massa ..... FCEA, UdelaR.

Dr. Marco Scavino ..... FCEA, UdelaR.

Esta página ha sido intencionalmente dejada en blanco.

“A la voluntad, más que a la inteligencia, se enderezan nuestros consejos, porque tenemos la convicción de que aquélla, como afirma cuerdamente Payot, es tan educable como ésta, y creemos además que toda obra grande, en arte como en ciencia, es el resultado de una gran pasión puesta al servicio de una gran idea.”

REGLAS Y CONSEJOS SOBRE LA  
INVESTIGACIÓN CIENTÍFICA  
LOS TÓNICOS DE LA VOLUNTAD  
SANTIAGO RAMÓN Y CAJAL

Esta página ha sido intencionalmente dejada en blanco.

# Agradecimientos

Quiero comenzar por agradecer a mis tutores, Dr. Ramón Álvarez-Vaz y Mag. Fernando Massa, por aceptar acompañarme en este desafío, por la confianza y guía, han sido una piedra angular en este proceso.

Al Dr. Marco Scavino y el Dr. Juan Kalemkerian, integrantes de la Comisión de Carrera en la Licenciatura en Estadística, gracias por su apoyo.

Seguidamente, quiero agradecer a mi familia, que son el sostén permanente e incondicional, y a mis amigos, por estar presentes a pesar de la distancia.

A mi juicio, si he llegado hasta aquí, buena parte se lo debo a tres docentes e investigadores, quienes dejaron una huella indeleble durante mi pasaje por sus clases. El Dr. Gonzalo Perera, mi primer profesor de estadística, quien sembró la semilla de la curiosidad por esta disciplina. Durante aquel año compartí, en algunas ocasiones, inolvidables charlas, siempre atento a escuchar, trabajador incesante en la búsqueda de alternativas que permitieran mejorar el acercamiento de sus estudiantes a la teoría y la práctica. A la Dra. Erna Frins y el Dr. José Ferrari (fallecido en 2021), quienes me enseñaron el maravilloso mundo de la óptica aplicada, así como también el trabajo en el laboratorio, actividad ésta que resultó fundamental para forjar la paciencia, resiliencia y tenacidad, cualidades que serían gravitantes en mis primeros trabajos en el ámbito académico. A Uds. muchas gracias por sus enseñanzas y consejos.

Gracias a mi amigo PhD (c) Gabriel Merlo, juntos comenzamos a estudiar en la Licenciatura en Estadística, luego nuestros caminos tomaron distintos rumbos, y ojalá pronto se puedan volver a cruzar.

Cometería una enorme injusticia al no hacer mención de los contribuyentes, muchos de ellos personas sencillas y de humilde condición, ya que gracias a sus aportes fue posible formarme en la educación pública. A Uds. quiero expresar un especial agradecimiento.

El camino no ha sido fácil, al transitarlo fue necesario sortear grandes dificultades. En éste conocí a personas que, quizás con fundados argumentos, no confiaron en mí, y de su parte sentí que me cerraron puertas. A ellos también quiero agradecer, ya que, sin imaginarlo me ayudaron a desafiar la adversidad, afirmar la personalidad, y templar el carácter.

No quiero dejar pasar la oportunidad para agradecer a MSc. Inés Urrestarazu por su apoyo y esas palabras de aliento cuando parecía que abandonaría la Licenciatura en Estadística, probablemente no imaginó que llegaron en el momento justo. También, gracias al equipo docente del IESTA, he disfrutado mucho de sus clases y el tiempo compartido. En particular, voy a extrañar mucho el dote pedagógico y sentido del humor del Dr. Juan José Goyeneche, quien me acercó a la estadística bayesiana.

Sería injusto de mi parte no agradecer a todos los que de alguna manera, a lo largo de estos años, fueron parte de mi formación en la Licenciatura en Estadística, en especial a mis compañeras de estudio Ec. Ana Vignolo y Prof. Lucía Tafernaberry.

## Resumen

La psicometría es el área que se dedica al estudio y desarrollo de herramientas de medición en psicología. Uno de los problemas que aborda es la medición de variables latentes. En este sentido, la Teoría de los Test (TT) es una de las herramientas que brinda un marco de trabajo a la hora de diseñar instrumentos de medida. En el ámbito de la TT, es la Teoría de Respuesta al Ítem (TRI) una técnica moderna cuya flexibilidad permite analizar las respuestas y las características de los ítems en un test. Un subconjunto de modelos en la TRI puede encontrarse bajo el paraguas del denominado Modelo de Rasch y sus extensiones, cuyas propiedades, así como también su sencillez, los convierten en atractivos cuando se trabaja con problemas reales. A pesar de haber sido desarrollados en forma independiente, los modelos propuestos en la TRI pueden ser representados a través de un modelo lineal generalizado mixto (GLMM). Este aspecto es sumamente relevante, ya que es posible hacer uso de los resultados y algoritmos desarrollados para los GLMM en la aplicación de un problema específico. En el presente Trabajo de Fin de Grado estaremos abordando el Modelo de Rasch, desde un enfoque tradicional, así como también desde la perspectiva de los GLMM. Además, ambas estrategias son aplicadas a un test de comportamiento utilizado en la identificación del Trastorno del Espectro Autista (TEA). Los resultados obtenidos muestran el potencial del Modelo de Rasch, y la posibilidad de ser extendidos en experimentos que involucren a la población en Uruguay.

## Palabras Claves

Teoría de Respuesta al Ítem (TRI), Modelo de Rasch, Modelos Lineales Generalizados Mixtos, Inferencia Bayesiana.

## Key Words

Item Response Theory (IRT), Rasch Model, Generalized Linear Mixed Model (GLMM), Bayesian Inference.

Esta página ha sido intencionalmente dejada en blanco.

# Índice general

<b>Agradecimientos</b>	<b>7</b>
<b>Resumen</b>	<b>9</b>
<b>Palabras Claves</b>	<b>9</b>
<b>Índice de Figuras</b>	<b>15</b>
<b>Índice de Tablas</b>	<b>17</b>
<b>Lista de Abreviaturas</b>	<b>19</b>
<b>1. Introducción</b>	<b>23</b>
1.1. Antecedentes . . . . .	24
1.2. Objetivo . . . . .	26
1.3. Aportes del TFG . . . . .	27
1.4. Estructura del TFG . . . . .	28
<b>2. Teoría de Respuesta al Ítem en Psicometría</b>	<b>29</b>
2.1. Introducción . . . . .	30
2.2. Teoría de Respuesta al Ítem . . . . .	31
2.2.1. TRI en Modelos con Respuesta Dicotómica . . . . .	32
2.2.2. Modelo Unidimensional Ojiva Normal . . . . .	34
2.2.3. Modelo Unidimensional Logístico . . . . .	37
2.2.4. Modelo Unidimensional 1PL . . . . .	38
2.2.5. Modelo Unidimensional 2PL . . . . .	40
2.3. Modelo de Rasch . . . . .	43
2.4. Observaciones Finales . . . . .	47
<b>3. Estimación de los Parámetros y Ajuste en el Modelo Logístico</b>	<b>49</b>
3.1. Introducción . . . . .	50
3.2. Método de Máxima Verosimilitud en la Estimación de los Parámetros de Discriminación y Dificultad . . . . .	51
3.2.1. Modelo 2PL . . . . .	52
3.3. Método de Máxima Verosimilitud en la Estimación del Rasgo Latente . . . . .	54

3.3.1. Modelo 2PL . . . . .	55
3.4. Métodos de Estimación Conjunta, Condicional y Marginal . . . . .	58
3.4.1. Máxima Verosimilitud Conjunta . . . . .	58
3.4.2. Máxima Verosimilitud Condicional . . . . .	59
3.4.3. Máxima Verosimilitud Marginal . . . . .	62
3.5. Ajuste del Modelo de Rasch . . . . .	63
3.5.1. Test $\chi^2$ de Pearson . . . . .	64
3.5.2. Test de Razón de Verosimilitudes . . . . .	65
3.5.3. Índices . . . . .	65
3.6. Observaciones Finales . . . . .	66
<b>4. Modelos Lineales Generalizados Mixtos y Modelo de Rasch</b>	<b>67</b>
4.1. Introducción . . . . .	68
4.2. Modelo Lineales Generalizados Mixtos . . . . .	68
4.3. Modelo Logit Mixto y Modelos de Respuesta al Ítem . . . . .	71
4.4. Modelos Jerárquicos y Modelos de Respuesta al Ítem . . . . .	73
4.5. Estimación de los parámetros del Modelo . . . . .	76
4.6. Estimación bayesiana de los parámetros en el Modelo Jerárquico . . . . .	77
4.7. Selección del Modelo . . . . .	80
4.7.1. Factor de Bayes - BF . . . . .	81
4.7.2. Criterio de Información Bayesiana - BIC . . . . .	82
4.7.3. Criterio de Información Devianza - DIC . . . . .	83
4.7.4. WAIC . . . . .	84
4.7.5. LOOCV . . . . .	85
4.8. Observaciones Finales . . . . .	86
<b>5. Aplicación del Modelo de Rasch a un Test utilizado en la identificación del TEA</b>	<b>89</b>
5.1. Introducción . . . . .	90
5.2. Descripción de los Datos . . . . .	91
5.3. Descripción de los Datos . . . . .	93
5.4. Elección de Variables . . . . .	95
5.5. Modelos Propuestos . . . . .	95
5.6. Selección de Modelos y Diagnóstico . . . . .	97
5.7. Resultados . . . . .	100
5.8. Evaluación de los Modelos . . . . .	102
5.9. Observaciones finales . . . . .	102
<b>6. Conclusiones</b>	<b>103</b>
6.1. Conclusiones . . . . .	104
6.2. Trabajos a Futuro . . . . .	105
<b>Apéndices</b>	<b>107</b>

<b>A. Resultados Generales</b>	<b>109</b>
A.1. Familia de Distribuciones . . . . .	109
A.2. Familia Exponencial . . . . .	109
A.3. Función de Verosimilitud . . . . .	110
A.4. Condiciones de Regularidad FI . . . . .	111
A.5. Función Score de Fisher . . . . .	111
A.6. Función Información de Fisher . . . . .	112
A.7. Métodos Numéricos . . . . .	113
A.8. Estadísticos Suficientes . . . . .	114
<b>B. Modelos Logísticos Dicotómicos - Estimación de los Parámetros</b>	<b>117</b>
B.1. Método de Máxima Verosimilitud - Estimación de los Parámetros de Discriminación y Dificultad en el Modelo 2PL . . . . .	117
<b>C. Extensiones al Modelo de Rasch</b>	<b>123</b>
C.1. Introducción . . . . .	123
C.2. Modelo de Test Lineal Logístico - LLTM . . . . .	124
C.3. Modelo de Test Lineal Logístico para Cambios - LLTMC . . . . .	125
C.4. Modelo de Rasch Regresión Latente . . . . .	126
<b>D. Métodos de Cadenas de Markov Monte Carlo</b>	<b>127</b>
D.1. Introducción . . . . .	127
D.2. Algoritmo Metropolis Hasting . . . . .	128
D.3. Muestreo de Gibbs . . . . .	129
D.4. Análisis de la Convergencia . . . . .	130
<b>E. Modelo Lineales Mixtos</b>	<b>131</b>
E.1. Introducción . . . . .	131
E.1.1. REML . . . . .	134
E.1.2. Estimación de los parámetros del modelo . . . . .	135
E.1.3. Ecuaciones de Henderson . . . . .	140
<b>F. Test de Comportamiento - Identificación de TEA</b>	<b>143</b>
F.1. El Test . . . . .	143
<b>G. Resultados Complementarios</b>	<b>145</b>
G.1. Análisis Descriptivo . . . . .	145
G.2. Modelo de Rasch . . . . .	147
G.3. Modelo de Rasch de Regresión Latente - Inferencia Clásica . . . . .	149
G.4. Modelo de Rasch de Regresión Latente - Inferencia Bayesiana . . . . .	151
G.5. Curva ROC . . . . .	153
<b>H. Código en R</b>	<b>155</b>
<b>Bibliografía</b>	<b>163</b>

Esta página ha sido intencionalmente dejada en blanco.

# Índice de Figuras

2.1. Los modelos Teoría Clásica de los Test (TCT) y Teoría de Respuesta al Ítem (TRI), propiedades psicométricas y métodos. Fuente: Rodríguez Jiménez, Rosero Burbano, Botía Sanabria, and Duarte Mateus (2011). . . . .	31
2.2. Ejemplo de curva característica de del ítem (Curva Característica del Ítem (CCI)), probabilidad $P(\theta)$ vs $\theta$ habilidad o rasgo latente. . . . .	33
2.3. Ejemplo del modelo ojiva normal dado por Ec. 2.2, donde se han representado gráficamente tres niveles de rasgo latente ( $\theta_j$ ) distintos. . . . .	36
2.4. Ejemplo de curva característica de del ítem (CCI), probabilidad ( $P(\theta)$ ) vs habilidad o rasgo latente ( $\theta$ ), para dos parámetros de dificultad. La curva color rojo corresponde a $b = 0$ y la curva color azul a $b_i = 1,5$ . Las curvas se han generado en R utilizando el código de ShinyItemAnalysis Martinkova and Drabinova (2018) y modificado por el autor de este Trabajo de Fin de Grado (TFG). . . . .	39
2.5. Ejemplo de curva característica de del ítem (CCI), probabilidad ( $P(\theta)$ ) vs habilidad o rasgo latente ( $\theta$ ), para dos valores del parámetro de discriminación. La curva CCI1 corresponde a $a = 1$ y $b = 0$ y la CCI2 a $a = 5$ y $b = 0$ . Las curvas se han generado en R utilizando el código de ShinyItemAnalysis Martinkova and Drabinova (2018) y modificado por el autor de este TFG. . . . .	42
5.1. Boxplot: variación del Score total observado en los participantes. . . . .	94
5.2. Adaptación del método método posterior predictive simulation aplicado a los modelos $m\{1,2,3,4\}$ _rasch_glm. . . . .	99
5.3. Cantidad de respuestas correctas, según el diagnóstico recibido, agrupadas por Sexo y categoría de edad. . . . .	101
5.4. Estimación de los parámetros asociado a los examinados. . . . .	101
F.1. Aplicación ASDTests disponible en <a href="https://www.asdtests.com/#">https://www.asdtests.com/#</a> . . . . .	143
G.1. Porcentaje de examinados diagnosticados, a través del test, con TEA agrupados por sexo. . . . .	145
G.2. Cantidad de respuestas correctas, según el diagnóstico recibido, agrupadas por Sexo y categoría de edad. . . . .	145
G.3. Cantidad de respuestas correctas, según el diagnóstico recibido, agrupadas por Sexo e ictericia (Jaundice). . . . .	146
G.4. Curvas características de los ítems. . . . .	147

G.5. Curvas de información vs Rasgo Latente. . . . .	147
G.6. Residuos del modelo m2_rasch_glm. . . . .	150
G.7. Residuos del modelo m2_rasch_glm agrupados por Sexo. . . . .	150
G.8. Distribución a posteriori de los parámetros y convergencia de las cadenas. . . . .	152
G.9. Predictivas a posteriori. . . . .	152
G.10. Curva ROC. . . . .	153

# Índice de Tablas

2.1. Ejemplo de datos recolectados en un test con n ítems en donde ha participado J examinados. . . . .	33
4.1. Clasificación de modelos de respuesta la ítem. La función $\eta(\cdot)$ corresponde a la inversa de la función link. Para cada modelo se presenta el valor de dicha función discriminando entre la componente asociada a el examinado y el ítem. Fuente: De Boeck and Wilson (2004) . . . . .	73
5.1. Ítems del test “Autistic Spectrum Disorder Screening Data for Toddlers” Thabtah (2020). . . . .	92
5.2. Descripción de las variables que forman parte del conjunto de datos publicado por Thabtah (2020). . . . .	92
5.3. Medidas de resumen para las variables Age_Mons y Score. . . . .	93
5.4. Cantidad de participantes en cada categoría para las variables Sex, Jaundice, Class.ASD.Traits., Family_mem_with_ASD, Age_Cat y Who.completed.the.test. . . . .	93
5.5. Promedio y desvío estándar (SD) del Score total obtenido en el test, agrupados según el diagnóstico (Class.ASD.Traits.), desglosado por: Todos los participantes, Sexo, ictericia (Jaundice), familiar diagnosticado con TEA (Family_mem_with_ASD), Categoría de edad (Age_cat). . . . .	94
5.6. Modelos propuestos, variables involucradas y tipo de inferencia utilizada en la estimación de los parámetros . . . . .	96
5.7. Modelos propuestos, variables involucradas y tipo de inferencia utilizada en la estimación de los parámetros . . . . .	97
5.8. Valores obtenidos de AIC, BIC y Test de Razón de Verosimilitudes correspondientes a los modelos MRRL-GLMM cuyos parámetros han sido estimados utilizando ML a través de la aproximación de Lapace. . . . .	97
5.9. Valores obtenidos de DIC y LOOCV correspondientes a los modelos MRRL-GLMM cuyos parámetros han sido estimados utilizando inferencia bayesiana. . . . .	98
5.10. Resultados del Test de Wald aplicado al modelo m2_rasch_glmm. . . . .	98
5.11. Resultados del Test de Kolmogorov Smirnov y Jarque Bera aplicado a los residuos del modelo m2_rasch_glmm. . . . .	98
5.12. Estimación de los parámetros en los modelos seleccionados. . . . .	100
5.13. Evaluación de la capacidad predictiva de los modelos. . . . .	102

E.1. Estimación de efectos fijos y predicción de efectos aleatorios en Linear Mixed Models (LMM) cuando $\Sigma$ es conocida. . . . .	136
E.2. Estimación de efectos fijos y predicción de efectos aleatorios en LMM cuando $\Sigma$ es desconocida. . . . .	140
F.1. Preguntas en el Test. . . . .	144
G.1. Resultados del Test de Wald aplicado al modelo m_rasch. . . . .	147

# Lista de Abreviaturas

- 1PL** One Parameter Logistic. 38–41, 43, 44, 50, 58, 64, 76, 78, 80
- 2PL** Two Parameter Logistic. 40–44, 47, 50–52, 54, 55, 58, 59, 62–64, 74, 76, 78, 80, 105, 117, 118
- 3PL** Three Parameter Logistic. 43, 47, 58, 59, 62, 63, 76
- 4PL** Four Parameter Logistic. 47
- AIC** Akaike Information Criterion. 82, 83, 85, 97
- BF** Bayes Factor. 81, 82
- BIC** Bayesian Information Criterion. 82, 83, 85, 97
- BLP** Best Linear Predictor. 137, 138
- BLUE** Best Linear Unbiased Estimator. 136
- BLUP** Best Linear Unbiased Predictor. 137, 138
- BP** Best Predictor. 137, 138
- CCI** Curva Característica del Ítem. 15, 32–35, 37, 39, 41–44, 50, 55, 58, 123
- cdp** cumulative distribution function. 36, 41, 58
- CML** Conditional Maximum Likelihood. 58, 59, 61, 62, 66, 95, 100
- DIC** Deviance Information Criterion. 83–85, 97
- DIF** Differential Item Functioning. 72
- elppd** expected log pointwise predictive density. 84
- EM** Expectation–Maximization. 63
- FCEA** Facultad de Ciencias Económicas y Administración. 25

**GIMESE** Grupo Interdisciplinario en Métodos Estadísticos en Salud y Economía. 25

**GLM** Generalized Linear Models. 69–71, 73

**GLMM** Generalized Linear Mixed Models. 25–28, 68–73, 77, 86, 87, 95, 97, 104

**IESTA** Instituto de Estadística. 25

**iid** independientes e idénticamente distribuidos. 35, 85

**INE** Instituto Nacional de Estadística. 28, 90

**IS** Importance Sampling. 85, 86

**JML** Joint Maximum Likelihood. 58, 59, 62, 66

**LLTM** Linear Logistic Test Model. 61, 72, 124–126

**LLTMC** Linear Logistic Test Model for Change. 125, 126

**LMM** Linear Mixed Models. 18, 68, 131, 135, 136, 139, 140

**LOO** Leave-One-Out. 86

**LOOCV** Leave-One-Out Cross-Validation. 85, 97

**lppd** log pointwise predictive density. 84

**M-H** Metropolis Hastings. 77, 79, 80, 127–130

**MCMC** Markov chain Monte Carlo. 68, 77, 79, 83, 87, 127

**ML** Maximum Likelihood. 54, 95, 100, 117, 134, 135, 139, 140

**MML** Marginal Maximum Likelihood. 58, 62, 66

**MR** Modelo de Rasch. 43–46, 95, 123, 125

**MRRL-GLMM** Modelo de Rasch de Regresión Latente - GLMM. 95, 96, 102

**MSE** Mean Squared Error. 138

**OR** Odds Ratio. 38, 42, 46

**PSIS** Pareto Smoothed Importance Sampling Leave-One-Out. 86

**RAE** Real Academia Española. 30

**REML** Restricted Maximum Likelihood. 134, 135, 139

**TCT** Teoría Clásica de los Test. 15, 24, 25, 30–32, 39, 47

**TEA** Trastorno del Espectro Autista. 26–28, 90, 91, 93–95, 99, 105, 143

**TFG** Trabajo de Fin de Grado. 15, 26–28, 32, 34, 39, 42, 47, 50, 69, 70, 87, 91, 95, 102, 123

**TRI** Teoría de Respuesta al Ítem. 15, 24–28, 30–32, 34, 39, 41, 46, 47, 50, 64–66, 68, 71, 74, 76, 77, 80–82, 86, 104, 105

**VA** Variable Aleatoria. 32, 35, 38–41, 50, 112, 127

**VAs** Variables Aleatorias. 127

**WAIC** Widely Applicable Information Criterion. 84, 85

Esta página ha sido intencionalmente dejada en blanco.

# Capítulo 1

## Introducción

## 1.1. Antecedentes

En muchas situaciones de investigación aplicada las variables de interés no son directamente observables, es decir, el problema involucra a las denominadas variables latentes. A pesar de su peculiar característica, su existencia puede ser inferida a través de una serie de manifestaciones, conocidas como variables manifiestas, que cuentan con la virtud de ser cuantificables (De Ayala, 2009, Chap. 1). Un ejemplo de ello es la variable “inteligencia de los individuos”, donde a través de un conjunto de test (o cuestionarios) se obtienen medidas indirectas que, entre otras cosas, permiten caracterizar y conformar agrupamientos de personas (Baker, 2001, Chap. 1).

La psicometría, según la Psychometric Society, es el área dedicada a la investigación y el desarrollo de herramientas de medición cualitativas en psicología. Dentro de dicha área, muchos de los problemas abordados involucran variables latentes. En este sentido, algunos de los instrumentos que se han diseñado son los tests, los cuales, dependiendo del objetivo que se persiga, tendrán una estructura, un conjunto de preguntas (abiertas o cerradas), un conjunto de respuestas admisibles, etc.

En general, las técnicas utilizadas en el proceso de medir variables latentes involucran una correspondencia entre un conjunto de niveles en la variable de interés y valores en una escala como consignan (Wright and Masters, 1982, Chap. 1 y 3), (Dunn-Rankin, Knezek, Wallace, and Zhang, 2004, Chap. 2). Actualmente, dentro de la psicometría, los frameworks que permiten establecer dicha correspondencia son: la teoría clásica de los test y la teoría de respuesta al ítem.

La TCT nace, según algunos autores como Muñiz (2000), en los trabajos de Spearman a comienzo del siglo XX. Ésta se basa en el hecho que el instrumento (un test) utilizado para evaluar a un conjunto de personas, no es preciso e introduce un error. Esto quiere decir que, el score total (o resultado) del test es una medida empírica imprecisa, y en consecuencia, los apartamientos respecto al valor real deben de ser tenidos en cuenta a la hora de elaborar conclusiones.

Bajo este escenario, donde el instrumento que permite inferir una variable latente es impreciso, se propone utilizar un modelo estadístico lineal para modelar los desvíos respecto al verdadero valor. Es decir, el score total observado (score empírico), correspondiente a cada participante del test, está formado por dos componentes: la primera, corresponde al verdadero valor del score total, mientras que la segunda, es el error introducido por el instrumento. Además, se asume que los errores son independientes e idénticamente distribuidos, no correlacionados entre sí y con el score observado.

Una de las ventajas de la TCT radica en el hecho de su sencillez, tanto en su aplicación, así como también en la interpretación de los resultados. Sin embargo, carece de una propiedad relevante que invalida su aplicación en forma generalizada. Esto es, el modelo no es invariante respecto del instrumento, es decir, al ser aplicado en distintos tests, los resultados obtenidos no son comparables. A pesar de ello, la TCT ha sido utilizada en distintas disciplinas y aún sigue formando parte en los textos clásicos de psicometría.

En la TRI, concebida en forma independiente y casi en simultáneo por Lord (1952) y Rasch (1960), se realiza un abordaje diferente respecto a la TCT. Mediante la propuesta de un nuevo modelo estadístico se abre un abanico de opciones que posibilita el análisis de las respuestas observadas, así como también ciertas características de las preguntas que forman parte del test. En este sentido, el centro de interés no se encuentra únicamente en el score total de cada participante, sino en sus

respuestas y los ítems del test. Esto último es lo que hace factible, a través de un mismo modelo, comparar distintos instrumentos. Algunos autores, por ejemplo, De Ayala (2009), Martínez Arias, Hernández Lloreda, and Hernández Lloreda (2014) o Muñiz (2018), la denominan teoría moderna de los tests.

El Modelo de Rasch y sus extensiones puede considerarse como un subconjunto de los modelos desarrollados en la TRI. Sus propiedades y sencillez lo han convertido en una herramienta atractiva en el análisis de test. Como veremos en el siguiente capítulo, a pesar de poseer una expresión algebraica idéntica al propuesto por Lord, la forma en que Rasch elabora la solución al mismo problema es partiendo de un enfoque completamente diferente.

Un aspecto relevante a tener presente es el siguiente, modelar las características de los participantes y la estructura del test al mismo tiempo no es gratuito, y trae consigo complejizar el modelo. Es decir, los modelos desarrollados dentro de la TRI presentan una estructura más compleja respecto al modelo lineal de la TCT. En consecuencia, la inferencia estadística requerirá de técnicas más sofisticadas. Esto se ve reflejado en los algoritmos que se han propuesto durante el desarrollo de la TRI, tanto desde la estadística frecuentista, así como también bajo la estadística bayesiana.

Resulta oportuno señalar que los modelos propuestos bajo el paraguas de la TRI pueden ser abordados a través de la teoría desarrollada en los Generalized Linear Mixed Models (GLMM). Esta característica es interesante ya que, a pesar de ser teorías desarrolladas en forma independiente y persiguiendo distintos objetivos, es casi inmediato representar cualquiera de los modelos desarrollados en la TRI como un GLMM. Por lo tanto, es admisible la aplicación de todas las herramientas y los resultados desarrollados dentro de la teoría de modelos generalizados mixtos, a un subconjunto de problemas pertenecientes a la teoría de los test.

A pesar que el desarrollo de la TRI comienza a mediados del siglo XX, es en las últimas décadas, debido al aumento en capacidad de cómputo y el avance tecnológico, en que toma mayor relevancia y aplicación a problemas reales. No debemos perder de vista que, en general, se suelen analizar test donde el número de participantes es elevado, y en consecuencia, la cantidad de parámetros a estimar es elevada, lo cual resultó ser una limitante durante muchos años.

En la actualidad, podemos encontrar aplicaciones de la TRI en diversas áreas. Por ejemplo, en medicina destacamos el estudio de: el Trastorno del Espectro Autista Cordier, Munro, Wilkes-Gillan, R, Parsons, and Joosten (2019); Kuhfeld and A. (2018); Lundin, Kosidou, and Dalman (2019); Mazefsky, Yu, White, Siegel, and Pilkonis (2018); Murray, Booth, McKenzie, and Kuenssberg (2016); Seung, Ji, Kim, Sung, Youn, Hong, H., Lee, H., and Youm (2015); la enfermedad de Alzheimer Ard, Galasko, and Edland (2013); Balsis, Unger, Bengé, Geraci, and Doody (2012); Balsis, Choudhury, Geraci, and Patrick (2018); Le, Lancashire, Hirst, and Jokel (2011); Umeda-Kameyama, Mori, Wada-Isoe, Kikuchi, Kojima, Kagimura, Ueki, Watabe, Kudoh, Akishita, and Nakamura (2018); Verma and Markey (2014); o la epilepsia Birbeck, Sehyun, Hays, and Vickrey (2000); Goodwin, Ferro, and Speechley (2018).

Dentro del Instituto de Estadística (IESTA) de la Facultad de Ciencias Económicas y Administración (FCEA) el grupo Grupo Interdisciplinario en Métodos Estadísticos en Salud y Economía (GIMESE) se encuentra investigando, desde distintas perspectivas, problemas que involucran en análisis de test. En esta dirección, algunos trabajos realizados pueden encontrarse en Alvarez-Vaz (2020); Alvarez-Vaz, Massa, Lorenzo, and Fabruccini (2018); Debera and Nalbarte (2006); Massa (2020); Massa, Camaño, and Álvarez Vaz (2018); Massa, Berberian, and Álvarez-Vaz (2019).

Motivados en la relevancia intrínseca de la TRI y la diversidad de aplicaciones en las que ésta puede contribuir, nos proponemos abordar dicha teoría con un enfoque distinto al habitual. Es decir, analizaremos la TRI de forma tradicional e incorporaremos la perspectiva desde los GLMM. Sumado a esto, serán presentadas las técnicas de inferencia que provienen de la estadística frecuentista y bayesiana. Además, aplicaremos los métodos aquí desarrollados a un test de comportamiento utilizado en la identificación del Trastorno del Espectro Autista (TEA).

## 1.2. Objetivo

El objetivo del presente TFG consiste en el estudio de los modelos de TRI, en particular profundizaremos en el Modelo de Rasch definido como un GLMM. Es decir, será objeto de investigación aquellos casos donde, mediante un Modelo de Rasch, pueda describirse el comportamiento de la variable dependiente (respuesta al ítem dicotómica), utilizando efectos fijos y aleatorios.

Los objetivos específicos propuestos son:

- Revisión bibliográfica de la TRI: los denominados modelos de un parámetro, dos parámetros y el Modelo de Rasch, así como también el vínculo con los GLMM.
- Análisis de las bibliotecas disponibles en R que posibiliten la aplicación a un problema real.
- Implementación en R del Modelo de Rasch, desde un enfoque clásico así como también desde la perspectiva de los GLMM, con aplicación a un problema real.

Los supuestos son:

- Los modelos y el test a analizar son de respuesta dicotómica.
- La variable latente es unidimensional.
- La aplicación de los modelos propuestos a un problema real se realizará haciendo uso de los datos proporcionados por Thabtah (2020), que corresponden a un test de comportamiento utilizado en la identificación del TEA.

### 1.3. Aportes del TFG

El presente TFG realiza un incipiente aporte en:

- Exponer los principales resultados de la TRI, para el caso de los modelos que son objeto de estudio, en forma unificada a partir de las publicaciones realizados principalmente por Fischer and Molenaar (1995), Baker and Kim (2004), De Boeck and Wilson (2004), Fox (2010) y Levy and Mislevy (2016), así como también otros autores.
- Revisión de la evolución de la TRI a lo largo de los años, aspecto éste que nos llevó a encontrar erratas en algunos desarrollos de resultados, citas a autores, y citas a resultados.
- Incorporar, en forma breve e introductoria, la visión de la Teoría de los Test desde una óptica que proviene del ámbito de la psicología, algo que no es habitual en los textos de estadística.
- Aplicación del Modelo de Rasch, bajo el enfoque frecuentista y bayesiano, al conjunto de datos generado por Thabtah (2020). De acuerdo a nuestra investigación, este acercamiento al problema aún no ha sido publicado a nivel de revistas arbitradas o conferencias.

En lo que refiere a la evolución de la TRI, en particular los modelos de un parámetro, dos parámetros y el modelo de Rasch, a lo largo de nuestro trabajo hemos sido diligentes en revisar las citas y los resultados que provenían de diversas publicaciones. Sin embargo, sabemos que aún queda trabajo por hacer para llegar a una completa revisión del estado del arte.

El Modelo de Rasch es una herramienta que ha sido utilizada en test que permiten identificar personas con el TEA. En este sentido, por ejemplo, Lundqvist and Lindner (2017) han diseñado un test y lo han aplicado a pacientes con y sin diagnóstico de TEA en Suecia. Por lo tanto, la aplicación de dicho modelo al análisis del TEA no es novedoso. Sin embargo, creemos que sí lo es la forma en que se aborda el problema, es decir, bajo la perspectiva de los GLMM haciendo uso de la inferencia clásica y bayesiana.

Un aspecto relevante a señalar está asociado a la dificultad en el acceso a datos. En las publicaciones que hemos consultado los datos analizados no se encuentran disponibles, lo cual imposibilita la comparación entre los modelos que se han propuesto y los que proponemos. Esto nos condujo a elegir un conjunto de datos, cuyo acceso es público, y que son el resultado de un proyecto de investigación del Manukau Institute of Technology en Nueva Zelanda. Dicho proyecto persiguió el fin de desarrollar un test, y una plataforma de recolección de datos, que posibilitara la investigación del TEA Thabtah (2017a,1,1,2,1,1).

Existen variadas publicaciones que han hecho uso de los datos del Manukau Institute of Technology, aunque la mayoría de éstas utilizan algoritmos de machine learning en el análisis. Por ejemplo, la aplicación de árboles de decisión, k-nearest neighbors, random forest, redes neuronales, regresión logística, support vector machine, etc, puede encontrarse en Al-diabat (2018); Bala, Ali, Satu, Hasan, and Moni (2022); Erkan and Thanh (2019); Jannat and Canavan (2021); Jayakumaran and Sweetlin (2020); Raj and Masood (2020); Saihi and Alshraideh (2021); Singh, Farooqui, Sattler,

Usua, and Helde (2021); Thabtah (2017c); Thabtah, Kamalov, and Rajab (2018); Thabtah, Abdelhamid, and Peebles (2019). Una visión crítica sobre la aplicación de estos algoritmos y sus implicancias en el diagnóstico del TEA es desarrollada por Thabtah (2019a). Hasta este momento, de acuerdo a nuestra investigación, no hemos encontrado publicaciones donde los datos elegidos (Thabtah, 2020), hayan sido analizados utilizando algún modelo proveniente de la TRI. Esto hace que nuestra aproximación al problema sea una nueva contribución.

Sumado a lo anterior, teniendo en cuenta que durante el presente año se llevará a cabo el censo en Uruguay, y de acuerdo a la información oficial<sup>123</sup> del Instituto Nacional de Estadística (INE), por primera vez se realizará un relevamiento de la población diagnosticada con TEA. Entonces, se abren nuevas oportunidades para investigar el TEA, utilizando (o replicando) la herramienta desarrollada por Thabtah (2017b), en una muestra probabilística tomada en nuestro país, aplicando los modelos aquí desarrollados, así como también extensiones a éstos o nuevas propuestas.

## 1.4. Estructura del TFG

El presente TFG se estructura de la siguiente manera: en el Capítulo 2 estudiaremos los principales aspectos de la teoría de los test, introduciendo el modelo ojiva normal y los modelos de uno y dos parámetros, así como también el Modelo de Rasch y sus extensiones; a lo largo del Capítulo 3 se realiza un estudio formal de las técnicas utilizadas en la inferencia de los parámetros; en el Capítulo 4 se presenta el modelo de uno y dos parámetros de la TRI y el Modelo de Rasch bajo la perspectiva de los GLMM; por último, el Capítulo 5 está dedicado a la aplicación de las técnicas abordadas a un test de diagnóstico del TEA, mientras que en el Capítulo 6 se presentan las conclusiones y los trabajos a futuro.

---

<sup>1</sup><https://censo2023.uy/nombre-tiene-diagnostico-de-autismo-o-trastorno-del-espectro-autista-tea/>.

<sup>2</sup><https://webtest.censo2023.uy/wp-content/uploads/2023/01/Comparativo-de-censos-1963-2011-1.xlsx>

<sup>3</sup><https://preprod.censo2023.uy/wp-content/uploads/2023/02/Cuestionario-Censo-personas-viviendas-hogares-2011.pdf>

## Capítulo 2

# Teoría de Respuesta al Ítem en Psicometría

## 2.1. Introducción

La psicometría es un área multidisciplinaria que, de acuerdo a Psychometric Society Psychometric Society, se dedica al estudio de la teoría y las técnicas de medición cuantitativa en psicología, educación y ciencias sociales. Rao and Sinharay (2007) han realizado una reconstrucción histórica atribuyendo las primeras investigaciones en el área a C. F. Gauss y F. W. Bassel en la primera mitad del siglo XIX; sus trabajos estuvieron focalizados en las diferencias individuales en la percepción aplicados a la astronomía Meneses, Barrios, Bonillo, Cosculluela, Lozano, Turbany, and Valero (2013). Price 2016 da cuenta de las primeras investigaciones a partir de las publicaciones de F. Galton en la segunda mitad del siglo XIX, quien inspirado en los trabajos de C. Darwin se propuso clasificar a las personas según habilidades. Sin embargo, su consolidación como disciplina científica data de la década de 1930 (Meneses et al., 2013).

El problema de medir, en diversas áreas entre ellas la psicometría, es en la actualidad objeto de discusión y da lugar a opiniones contrapuestas. Medir es “comparar una cantidad con su respectiva unidad”, según la definición de la Real Academia Española (RAE). Nummaly y Bernstein 1994 asocian el concepto de medir al proceso de asignar símbolos a objetos, con el fin de representar cantidades de atributos mediante números, por ejemplo, mediante el uso de una escala; o decidir si un objeto pertenece a cierta categoría que cuenta con un conjunto de atributos. Dichos objetos en el campo de la psicología y biología generalmente son personas o animales, aunque, como es de esperar, no sólo a éstos se restringe en otras disciplinas.

A la hora de medir se requiere de, al menos, contar con instrumentos de medición junto con un proceso sistemático, estandarizado y reproducible, que permita comparar resultados. Pensemos, por ejemplo, en aplicaciones donde se busca estudiar el comportamiento de una especie, la escasez de dichos instrumentos o sus limitaciones imponen restricciones en el desarrollo de la investigación que no siempre son sencillas de sortear. Este problema, además, puede verse acentuado cuando las variables de interés son latentes, es decir, no son directamente observables.

Los test forman parte del conjunto de herramientas utilizadas en psicometría con el fin de medir, en forma indirecta, un fenómeno. Yela (1996) los define como el reactivo que, aplicado a un sujeto, revela el grado de instrucción, aptitud o manera de ser; al mismo tiempo afirma que, son los test psicométricos el instrumento que permite medir el grado de algún rasgo mental. Su utilización, junto con otras herramientas complementarias, contribuyen al proceso de recabar evidencia sobre las hipótesis que han sido formuladas como parte de la aplicación del método científico. Debido a la relevancia adquirida, se ha desarrollado lo que se conoce como Teoría de los Test.

La Teoría de los Test brinda un marco teórico sobre el cual se ha desarrollado, entre otros, el diseño de test, los modelos estadísticos y sus aplicaciones. Dicho marco permite establecer el vínculo funcional entre las variables observables a partir del test y aquellas que son latentes (Barbero García, García Cueto, Holgado Tello, and Vila Abad, 2015). Dos de los enfoques más destacados son la TCT y la TRI, siendo esta última la que ha contado con un fuerte impulso a nivel de investigación en los últimos años, con aplicaciones que trascienden a áreas como psicología, educación o medicina.

Los modelos TCT y TRI, más allá de perseguir el mismo objetivo y ser un instrumento en el análisis de propiedades psicométricas, cada uno presenta ventajas y desventajas respecto al otro. El primero nace a partir de los trabajos de Spearman a principios del siglo XX, quien a partir de

un modelo lineal sencillo analiza el score (o puntaje) que obtienen las personas al realizar un test; mientras que el segundo surge a partir de los trabajo de F. Lord y M. Novick, donde se proponen modelos estadísticos más complejos que permitan afrontar algunas de las carencias del enfoque clásico (Muñiz, 2010).

Rodriguez Jiménez et al. (2011) presenta un excelente resumen sobre como cada uno prioriza e identifica las propiedades psicométricas<sup>1</sup> (ver Figura 2.1), así como también los procedimientos de valoración o medición.

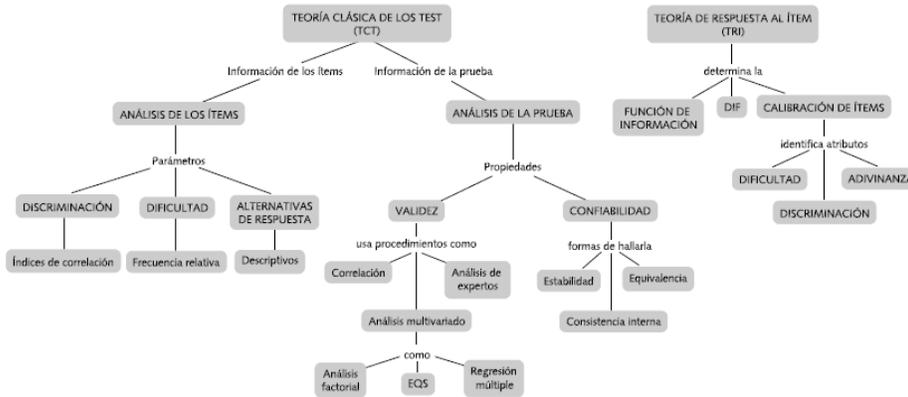


Figura 2.1: Los modelos TCT y TRI, propiedades psicométricas y métodos. Fuente: Rodriguez Jiménez et al. (2011).

Por último, vale la pena hacer mención que la psicometría cuenta con cinco pilares: teoría de la medición, teoría de los test, escalamiento psicológico, escalamiento psicofísico, y técnicas de estadística multivariada. Un análisis detallado puede encontrarse en Muñiz (2018) o Rao and Sinharay (2007).

## 2.2. Teoría de Respuesta al Ítem

La TRI es considerada la teoría moderna de los test, un paradigma de medición en psicometría (De Ayala, 2019), o un marco matemático que permite analizar conjuntamente las capacidades de las personas y los ítem Wu, Davis, Domingue, Piech, and Goodman (2020). Es una alternativa a TCT que permite entrelazar el rasgo latente ( $\theta$ ), que ha sido medido por el instrumento psicométrico, y la respuesta al ítem. En la presente sección estaremos abordando algunos aspectos de dicha teoría, así como también su base teórica.

En su génesis, TRI nace como una herramienta para dar solución al problema de medir variables latentes dentro del área de psicometría. Bock y Moustaki (2007, Chap. 11) atribuyen su origen a la búsqueda de soluciones al problema de medición en educación, donde el fin de era lograr medir el

<sup>1</sup>EQS corresponde a Ecuaciones Estructurales, mientras que DIF a Funcionamiento Diferencial del Ítem.

rendimiento individual de los estudiantes. Su gesta es posterior a la TCT, algunos autores, como De Ayala (2009) o van der Linden (2016); van der Linden and Hambleton (1997), atribuyen los primeros trabajos en el área a Thurstone (1925,2); posteriormente surgen las contribuciones de Lord (1952) y Rasch (1960).

Lord (1980,Chap. 2) establece que, TRI parte de una declaración matemática sobre cómo las respuestas a los ítems en un test dependen del nivel de habilidades de los examinados. Donde el vínculo entre ambos esta dado por la función de respuesta al ítem. En la actualidad, TRI puede ser visto como una familia de modelos probabilísticos que posibilita relacionar habilidades de los individuos con la probabilidad de responder correctamente a un ítem.

Vale la pena observar que en TRI no se persigue el objetivo de predecir cómo responderá un individuo a los ítem en un test, sino buscar construir herramientas estadísticas que posibiliten modelar características psicométricas independientemente del test en consideración. En este sentido mediante la TRI es posible inferir el nivel de habilidad en los examinados.

Una posible clasificación (Mair, 2018, Chap. 1) para los modelos TRI es de acuerdo a la naturaleza de las respuestas (dicotómicas, politómicas); o la dimensión del rasgo latente (unidimensional, multidimensional). En el presente TFG nuestro enfoque estará en respuestas dicotómicas y rasgo latente unidimensional.

En la literatura de teoría de test, cuando se trabaja con  $\theta \in \mathbb{R}$  se dice que nos encontramos bajo la asunción de unidimensionalidad. Por lo tanto, sólo una habilidad es medida a través del conjunto de ítems que forman parte del test. Esto es admisible siempre que exista un componente (o factor) dominante que influye en la performance de los examinados (Hambleton, Swaminathan, and Rogers, 1991, Chap. 2). Los modelos desarrollados bajo estas condiciones se los denomina modelos unidimensionales. Esta es una de las hipótesis sobre la cual se construye la TRI. La segunda hipótesis está ligada al concepto de independencia local, es decir, se asume que las respuestas de un examinado a cada ítem son independientes entre sí.

### 2.2.1. TRI en Modelos con Respuesta Dicotómica

Supongamos que los ítem admiten respuestas dicotómicas,  $\theta$  es unidimensional y existe independencia local. Además, supongamos que los examinados pueden ser ubicados en concordancia con una escala, de acuerdo a sus habilidades (o rasgo latente), y que la probabilidad de responder correctamente a un ítem depende únicamente de su habilidad. Bajo este escenario, la TRI propone que el vínculo entre  $\theta$  y la probabilidad que tienen los examinados de responder correctamente a un ítem  $P(\theta)$  esta dado por la CCI. Siendo ésta un piedra fundamental sobre la cual se construye la teoría.

Al igual que hicimos antes, supongamos que un test esta formado por  $i = 1, 2, \dots, n$ . ítems,  $j = 1, 2, \dots, J$ . son los examinados y las respuestas a cada ítem se modelan utilizando una Variable Aleatoria (VA) binaria  $Y_{ij} \in \{0, 1\}$ . En otras palabras, asumiremos que las respuestas tienen un score igual a 1 cuando son correctas y 0 en caso contrario. El vector aleatorio de respuestas del  $j$ -ésimo participante lo denotaremos como  $\mathbf{Y}_j \in \{0, 1\}^n$  (vector columna), mientras que el vector de las respuestas proporcionadas por éste lo representaremos como  $\mathbf{y}_j = (y_{1j}, y_{2j}, \dots, y_{nj})$ , donde  $y_{ij} \in \{0, 1\}$ .

Bajo esta notación, una vez que se ha realizado el test y se han registrado las respuestas de los participantes, la información obtenida puede representarse como en la Tabla 2.1, Donde  $y_{\cdot j} = \sum_{i=1}^n y_{ij}$ ,  $0 \leq y_{\cdot j} \leq n$ ,  $\forall j = 1, 2, \dots, J$  representa el score total del j-ésimo participante, mientras que  $y_{i\cdot} = \sum_{j=1}^J y_{ij}$ ,  $0 \leq y_{i\cdot} \leq J$ ,  $\forall i = 1, 2, \dots, n$  corresponde a la cantidad de respuestas correctas que han proporcionado los examinados en el i-ésimo ítem.

Tabla 2.1: Ejemplo de datos recolectados en un test con n ítems en donde ha participado J examinados.

Ítems	Participantes						Total por Ítem
	$P_1$	$P_2$	...	...	$P_{J-1}$	$P_J$	
Ítem 1	$y_{11}$	$y_{12}$			$y_{1,J-1}$	$y_{1J}$	$y_{1\cdot}$
Ítem 2	$y_{21}$	$y_{22}$			$y_{2,J-1}$	$y_{2J}$	$y_{2\cdot}$
⋮							
⋮							
⋮							
Ítem n	$y_{n1}$	$y_{n2}$			$y_{n,J-1}$	$y_{nJ}$	$y_{n\cdot}$
Score Total	$y_{\cdot 1}$	$y_{\cdot 2}$			$y_{\cdot, J-1}$	$y_{\cdot J}$	

De Ayala (2009, Chap 1.) y Luzardo (2017, Cap. 1), establecen que la primer referencia al concepto de CCI proviene de Tucker (1946), posteriormente llega el trabajo de Lord and Novick (1968). En la Figura 2.2 se muestra su forma genérica, donde podemos observar que existe una relación funcional entre los niveles de habilidad ( $\theta$ ) y la probabilidad de responder correctamente ( $P(\theta)$ ). En un test tendremos una CCI asociada a cada ítem, que a grandes rasgos conservará la forma de “S” y se diferenciarán entre sí en la velocidad crecimiento/decrecimiento, así como también en el nivel de habilidad para el cual la probabilidad de responder correctamente es igual 0.5.

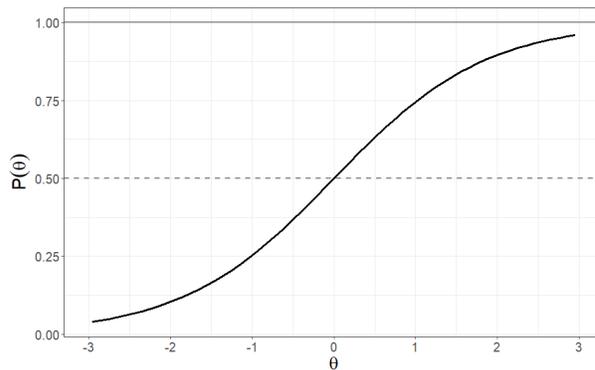


Figura 2.2: Ejemplo de curva característica de del ítem (CCI), probabilidad  $P(\theta)$  vs  $\theta$  habilidad o rasgo latente.

Potencialmente la escala del rasgo latente puede variar de  $-\infty$  a  $+\infty$ , sin embargo, bajo la hipótesis que dicha variable proviene de una población normal estandarizada, el rango de variación que suele utilizarse es  $[-3, 3]$  o  $[-4, 4]$  (Debera and Nalbarte, 2006; Hambleton et al., 1991). Si observamos la Figura 2.2, tenemos que, aquellos individuos con menor habilidad ( $\theta \leq 0$ ) tendrán menor probabilidad de responder correctamente, tendiendo ésta a cero a medida que  $\theta$  decrece. Mientras que aquellos con mayor habilidad ( $\theta > 0$ ) tendrán mayor probabilidad de responder correctamente y a medida que  $\theta$  crece dicha probabilidad tiende a uno.

La CCI representa el score medio que pueden obtener los examinados en función del rasgo latente. Existen dos propiedades sumamente relevantes asociadas a dicha función, una de ellas tiene que ver con su monotonía respecto de  $\theta$ . La segunda propiedad radica en que CCI puede ser definida como un miembro de la familia de posición y escala (Baker and Kim, 2004, Chap. 1), donde los parámetros involucrados son:  $b$ , que corresponde al parámetro de posición, expresado en las mismas unidades que  $\theta$ , en consecuencia con un rango de variación de  $(-\infty, +\infty)$  y corresponde al punto en que la probabilidad de responder correctamente el ítem es igual a 0,5;  $a$ , correspondiente al parámetro de escala de la curva, siendo su rango de variación teórico  $(0, +\infty)$ . En la siguiente sección se presentará formalmente y llegaremos a que éste corresponde al llamado modelo de dos parámetros.

Antes de continuar con la presentación formal de los modelos, es de orden observar que hasta ahora hemos supuesto que cada ítem en un test cuenta con una respuesta que se asume es la correcta. En test vinculados al área de educación es sencillo encontrar múltiples ejemplos de esta situación, pensemos en la prueba de ingreso que se realiza todos los años en nuestra facultad a cada nueva generación. Sin embargo, en otras áreas, como en medicina, un test puede ser un instrumento utilizado para elaborar un diagnóstico. Por lo tanto, cada ítem no posee una respuesta correcta, sino que las respuestas brindadas por los pacientes permiten confirmar (descartar) la presencia (ausencia) de ciertos síntomas. Este hecho no invalida el uso de la TRI, en estos casos es necesario interpretar las respuestas obtenidas como una forma de corroborar aquello que se sospecha como verdadero (Desjardins and Bulut, 2018, Chap. 5). Otro caso particular, común en la práctica real, son aquellas situaciones donde los participantes no responden, es decir, estamos frente a la ausencia de respuesta en algunos ítems. Esta casuística no será abordada en el presente TFG, un desarrollo en profundidad puede encontrarse en (Lord, 1980).

### 2.2.2. Modelo Unidimensional Ojiva Normal

El modelo unidimensional ojiva normal fue propuesto por Lord (1952), quien consideró que el comportamiento de la CCI presenta la misma forma que la curva de la distribución normal acumulada. Fue el primer modelo propuesto, presentado en 1959, donde en la parametrización fueron considerados los parámetros de dificultad y discriminación del ítem.

Backer establece que el modelo ojiva normal puede deducirse desde dos enfoques, uno más pragmática y otro que se basa en el formalismo teórico (Baker and Kim, 2004, Chap. 1). Comencemos por el primero de ellos, el cual nos permite acercarnos al modelo desde la intuición.

Si asumimos que el comportamiento de la CCI presenta similares características que la curva de una distribución normal acumulada, ésta última podría ser calibrada utilizando la proporción de respuestas correctas y los niveles de habilidad o rasgo latente. Dichos valores presentan un

estrecho vínculo, como veremos más en detalle en la siguiente sección, con dos atributos que posee la CCI y que corresponden al parámetro de dificultad y al parámetro de discriminación del ítem respectivamente.

Por otro lado, sabemos que la distribución  $N(\mu, \sigma^2)$  pertenece a la familia de posición y escala, con parámetro de posición  $\mu$  y  $1/\sigma$  el parámetro de escala, siendo la Ec. 2.1 la expresión analítica de la función de distribución en el caso de la normal estandarizada.

$$\Phi(z) = \int_{-\infty}^{z=(\theta-\mu)/\sigma^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \quad (2.1)$$

Si se quiere utilizar el modelo ojiva normal para modelar la CCI, una opción sencilla es igualar los parámetros de posición y escala de ambas curvas. Observar que la Ec. 2.1 no es utilizada en el sentido tradicional, es decir, para obtener la probabilidad acumulada, sino como una expresión analítica que modela una curva.

Este es un camino muy simplista del problema a la hora de modelar la CCI, sin embargo, de forma sencilla e intuitiva nos permite encontrar el vínculo entre dos modelos que, en principio, parecerían estar completamente disociados. Un desarrollo riguroso del modelo ojiva normal ha sido desarrollado por Lord (1980) y Lord and Novick (1968), donde la idea central, a grandes rasgos, consiste en construir un modelo de regresión lineal donde la variable independiente es el rasgo latente ( $\theta$ ) y la dependiente es una VA continua ( $\Upsilon \in \mathbb{R}$ ) que modela la propensión de un individuo a responder en forma correcta un ítem del test.

Siguiendo la línea de razonamiento propuesta por D. Bock y I. Moustaki (2007, Chap. 15), supongamos que  $Y_{ij} \in \{0, 1\}$  es una VA dicotómica que modela la respuesta al  $i$ -ésimo ítem de un test,  $i = 1, 2, \dots, n$ , por parte del  $j$ -ésimo participante,  $j = 1, 2, \dots, J$ . Es decir,  $Y_{ij} = 1$  si la respuesta fue la correcta y 0 en caso contrario. Sea  $\theta_j \in \mathbb{R}$  el rasgo latente que se desea medir en el  $j$ -ésimo individuo y  $\Upsilon_{ij}$  la propensión de éste a responder en forma correcta al  $i$ -ésimo ítem. Además, supongamos que existe un umbral  $\zeta_i$  (constante) tal que si  $\Upsilon_{ij} > \zeta_i$ , entonces  $Y_{ij} = 1$ , es decir, el participante responde correctamente al ítem. Mientras que si  $\Upsilon_{ij} < \zeta_i$ , entonces  $Y_{ij} = 0$ . Se asume, además, que con probabilidad nula ocurre el suceso  $\Upsilon_{ij} = \zeta_i$  (Lord, 1980, Chap. 3). Bajo estas condiciones, la propuesta de Lord se basa en el hecho que al medir el valor de  $\theta_j$ , dicha medida está afectada de un error aleatorio, proponiendo el siguiente modelo lineal:

$$\Upsilon_{ij} = a_i(\theta_j - b_i) + \epsilon_{ij} \quad (2.2)$$

Donde  $\epsilon_{ij}$  son errores aleatorios, que pueden considerarse independientes e idénticamente distribuidos (iid) con distribución  $N(0, 1)$ , mientras que  $a_i$  y  $b_i$  son dos parámetros característicos de cada ítem denominados parámetro de discriminación del ítem y parámetro de dificultad del ítem. El primero de ellos aporta información sobre la capacidad de ítem para discriminar entre los distintos niveles de  $\theta$ , como veremos a continuación es proporcional a la pendiente de la CCI. El segundo nos brinda información sobre los niveles del rasgo latente, donde la probabilidad de responder correctamente es mayor o igual a 0,5 en cada ítem.

Bajo la hipótesis de independencia local y suponiendo que no existe correlación entre las respuestas de los participantes, tenemos que:

$$\Upsilon_{ij}|\theta_j \stackrel{iid}{\sim} N(a_i(\theta_j - b_i), 1) \quad (2.3)$$

Lo que equivale a decir:

$$Z_{ij}|\theta_j = \Upsilon_{ij} - a_i(\theta_j - b_i) \stackrel{iid}{\sim} N(0, 1) \quad (2.4)$$

La expresión anterior nos conduce a que la probabilidad de responder correctamente el  $i$ -ésimo ítem por parte del  $j$ -ésimo participante ( $P_i(\theta_j)$ ), dado  $a_i$ ,  $b_i$  y un nivel de habilidad  $\theta_j \in \mathbb{R}$ , esta dada por:

$$\begin{aligned} P_i(\theta_j) &= P(Y_{ij} = 1|\theta_j) = P(\Upsilon_{ij} > \zeta_i|\theta_j) \\ &= \int_{z=\zeta_i - a_i(\theta_j - b_i)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \int_{-\infty}^{z=-\zeta_i + a_i(\theta_j - b_i)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \Phi(z(\theta_j)) \end{aligned} \quad (2.5)$$

Siendo  $\Phi(z(\theta_j))$  la cumulative distribution function (cdf) de la distribución normal, que en este modelo debe interpretarse<sup>2</sup> como una función que depende de  $\theta_j$ , es decir, de la habilidad o rasgo latente del individuo. Observar que en el desarrollo de la Ec. 2.5 hemos utilizados la propiedad  $1 - \Phi(z) = \Phi(-z)$ , es decir, simetría de la distribución normal.

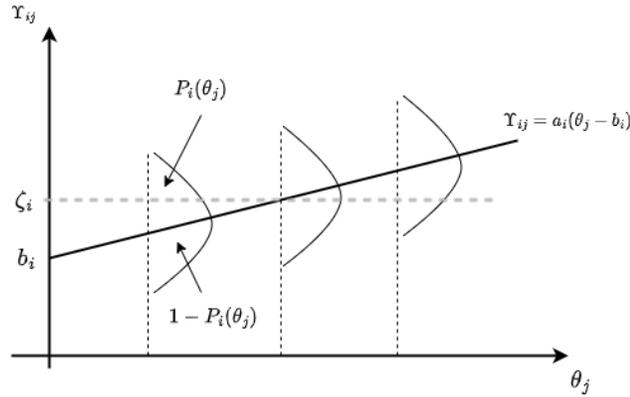


Figura 2.3: Ejemplo del modelo ojiva normal dado por Ec. 2.2, donde se han representado gráficamente tres niveles de rasgo latente ( $\theta_j$ ) distintos.

En la Figura 2.3 se muestra un ejemplo del modelo dado por la Ec. 2.2, donde se ha representado en forma genérica tres niveles de  $\theta$ ; además se incluye el potencial rango de variación de  $Y_{ij}$  utilizando

<sup>2</sup>En el modelo ojiva normal  $\Phi(z(\theta_j))$  no debe interpretarse como se hace habitualmente, es decir, como la probabilidad acumulada.

la campana sobre cada línea vertical. La probabilidad  $P_i(\theta_j)$  corresponde al área bajo la campana para aquellos valores tal que  $\Upsilon_{ij} > \zeta_i$ , mientras que  $Q_i(\theta_j) = 1 - P_i(\theta_j)$  es el área bajo la campana por debajo del umbral. Se puede observar que a cada nivel de habilidad  $\theta_j$  le corresponde una probabilidad distinta de responder correctamente al ítem, lo mismo sucede con la probabilidad de responder en forma incorrecta. A medida que nos movemos hacia la derecha en el eje de las abscisas, el área correspondiente a  $P_i(\theta_j)$  aumenta, lo cual está en la misma dirección de la intuición, es decir, mayores niveles de habilidad por parte del examinado se corresponden con un incremento en la probabilidad de responder correctamente al ítem. En contrapartida, a medida que disminuye el nivel de habilidad del examinado, es menos probable que responda correctamente.

La CCI en el modelo ojiva normal se modela utilizando la expresión para  $P_i(\theta_j)$  (Ec. 2.5), a éste también se lo conoce como modelo de dos parámetros, ya que  $a_i, b_i$  forman parte de la expresión analítica. Dicha función es creciente en  $\theta_j$ , con un punto de inflexión en  $\theta_j = b_i$  correspondiente a  $P_i(\theta_j) = 0,5$ , siendo  $a_i/\sqrt{2\pi}$  la pendiente en dicho punto. Un caso particular del modelo de dos parámetros es cuando  $a_i = 1 \forall i = 1, 2, \dots, n.$ , es decir, las curvas características de todos los ítems presentan la misma pendiente, en consecuencia se obtiene el llamado modelo de un parámetro.

Existe una extensión al modelo de dos parámetros donde se incorpora un tercer parámetro ( $c_i$ ), que permite reflejar el efecto de adivinación (guessing) en el acierto de la respuesta por parte del examinado. La expresión analítica esta dada por:

$$\begin{aligned} P_i(\theta_j) &= P(Y_{ij} = 1|\theta_j) = P(\Upsilon_{ij} > \zeta_i|\theta_j) \\ &= c_i + (1 - c_i) \int_{z=\zeta_i - a_i(\theta_j - b_i)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \end{aligned} \quad (2.6)$$

A pesar de su potencial el modelo ojiva normal presenta la desventaja de ser más difícil trabajar desde el punto de vista matemático y computacional, ya que involucran el proceso de integración. Es por ello que en la práctica resultan más adecuados los denominados modelos logísticos (Hambleton et al., 1991, Chap. 1). En las siguientes secciones analizaremos dichos modelos, así como también su vínculo con el ojiva normal.

### 2.2.3. Modelo Unidimensional Logístico

El modelo logístico es una alternativa al modelo ojiva normal, donde se utiliza la función logística con el fin de modelar a la CCI. La propuesta surge, como veremos en la siguiente sección, a través de dos concepciones distintas del problema, una de ellas desarrollada por A. Birnbaum y otra por G. Rasch.

Recordemos que la función logística  $g(x)$  esta dada por la Ec. 2.7, ésta es creciente en el recorrido de  $x$ , presenta una cota inferior en  $g(x) = 0$  y una superior en  $g(x) = 1$ . Además,  $x = 0$  es un punto de inflexión tal que  $g(0) = 0,5$ . Podemos observar que el recorrido de  $g(x)$  es el intervalo  $(0, 1)$ , por lo tanto, dicha función transforma  $\mathbb{R}$  en  $(0, 1)$ , siendo ésta una aplicación biyectiva.

$$g(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}} \quad , \quad x \in \mathbb{R} \quad (2.7)$$

La expresión general del modelo logístico esta dado por la Ec. 2.8, donde  $P_i(\theta_j)$  es la probabilidad de responder en forma correcta al ítem  $i$  dado que  $\theta_j$  es el rasgo latente del  $j$ -ésimo participante.

$$P_i(\theta_j) = \frac{e^{L(\theta_j)}}{1+e^{L(\theta_j)}} = \frac{1}{1+e^{-L(\theta_j)}} \quad , \quad \theta_j \in \mathbb{R} \quad (2.8)$$

Algunas ventajas que presenta respecto a su predecesor, el modelo ojiva normal, se encuentra en el hecho que nos libera de realizar el cálculo integral. Además, teniendo en cuenta que la probabilidad de responder en forma incorrecta el ítem es  $Q_i(\theta_j) = 1 - P_i(\theta_j)$ , entonces  $\ln(P_i(\theta_j)/Q_i(\theta_j))$  es una función lineal en  $L(\theta_j)$ . A este resultado se llega de forma sencilla a partir de la definición de Odds Ratio (OR) y la función<sup>3</sup> logit (Ec. 2.10), como se puede observar en la expresión 2.11

$$OR = \frac{P(Y_{ij} = 1|\theta_j)}{P(Y_{ij} = 0|\theta_j)} \quad (2.9)$$

$$\ln(OR) = \ln \left( \frac{P(Y_{ij} = 1|\theta_j)}{P(Y_{ij} = 0|\theta_j)} \right) \quad (2.10)$$

$$L(\theta_j) = \ln \left( \frac{e^{L(\theta_j)}}{1+e^{L(\theta_j)}} \cdot \frac{1+e^{L(\theta_j)}}{1} \right) \quad (2.11)$$

A lo largo del tiempo se han desarrollado distintos modelos logísticos siendo los más relevantes aquellos que, al igual que el ojiva normal, utilizan los parámetros de discriminación, dificultad y adivinación. A continuación abordaremos cada uno de ellos.

#### 2.2.4. Modelo Unidimensional 1PL

El modelo de un parámetro logístico o One Parameter Logistic (1PL) corresponde al caso en que  $L(\theta_j)$  es función del rasgo latente  $\theta_j$  y  $b_i$  correspondiente al parámetro de dificultad ítem. Este es un caso particular del modelo inicialmente propuesto Birnbaum (1968) denominado modelo de dos parámetros.

La propuesta de Birnbaum consiste en modelar  $P_i(\theta_j)$ , probabilidad de responder correctamente al  $i$ -ésimo ítem por parte del  $j$ -ésimo examinado, a través de una función lineal de su habilidad y la dificultad del ítem. En otras palabras, dado un nivel del rasgo latente, la dificultad del ítem es la única característica que influye en la performance de sujeto (Hambleton et al., 1991, Chap. 2).

Supongamos que el test cuenta con  $i = 1, 2, \dots, n$ . ítem y  $j = 1, 2, \dots, J$ . son los examinados, tal como representamos en la Tabla 2.1, entonces la expresión analítica para el modelo 1PL es:

$$P_i(\theta_j) = P(Y_{ij} = 1|\theta_j, b_i) = \frac{e^{(\theta_j - b_i)}}{1+e^{(\theta_j - b_i)}} \quad (2.12)$$

Siendo  $P(Y_{ij} = 1|\theta_j)$  la probabilidad que el examinado  $j$ , con un nivel de habilidad  $\theta_j$ , responda en forma correcta al  $i$ -ésimo ítem,  $L(\theta_j) = \theta_j - b_i$ . El parámetro  $b_i \in \mathbb{R}^+$ , al igual que antes, corresponde a la dificultad del  $i$ -ésimo ítem. Notar que  $Y_{ij}$  es una VA dicotómica ( $Y_{ij} \in \{0, 1\}$ )

<sup>3</sup>Utilizaremos  $\ln$  para denotar el logaritmo neperiano.

que modela la respuesta del examinado, tomando el valor 1 cuando la respuesta es correcta y 0 en caso contrario.

A partir de la expresión 2.12 es sencillo obtener la probabilidad que un participante  $j$  responda en forma incorrecta al ítem  $i$  dado un nivel de habilidad  $\theta$ :

$$P(Y_{ij} = 0|\theta_j, b_i) = 1 - P(Y_{ij} = 1|\theta_j, b_i) = \frac{1}{1+e^{(\theta_j-b_i)}} \quad (2.13)$$

Sea  $y_{ij}$  la realización de la VA  $Y_{ij}$ , entonces en forma compacta podemos expresar las Ecs. 2.12 y 2.13 de la siguiente manera:

$$P_i(\theta_j) = P(Y_{ij} = y_{ij}|\theta_j, b_i) = \frac{e^{(\theta_j-b_i)y_{ij}}}{1+e^{(\theta_j-b_i)}} \quad (2.14)$$

Observar que  $b_i$  es para  $P(Y_{ij} = 1|\theta_j)$  el parámetro de posición y corresponde al punto de inflexión de la curva CCI o de máxima pendiente, es decir,  $P(Y_{ij} = 1|\theta_j = b_i, b_i) = 0,5$ . Además, vale la pena notar que el concepto de dificultad en TRI difiere al que se utiliza en TCT, donde en esta última se encuentra asociado a la proporción de respuestas correctas al ítem observada entre todos los examinados.

En la Figura 2.4 se han representado dos ejemplos de CCI para el modelo 1PL, donde el rasgo latente varía en el intervalo  $[-4, +4]$ . y el parámetro de discriminación toma los valores  $b = 0$  y  $b = 1,5$ , curvas denominadas CCI1 y CCI2 respectivamente. Podemos observar que la curva CCI2 es igual a la curva CCI1 trasladada. Además, en el nivel de habilidad  $\theta = 0$ , la probabilidad de responder correctamente al ítem cuando  $b = 0$  es igual a 0.5, mientras que si  $b = 1,5$  ésta toma un valor de 0.15 aproximadamente.

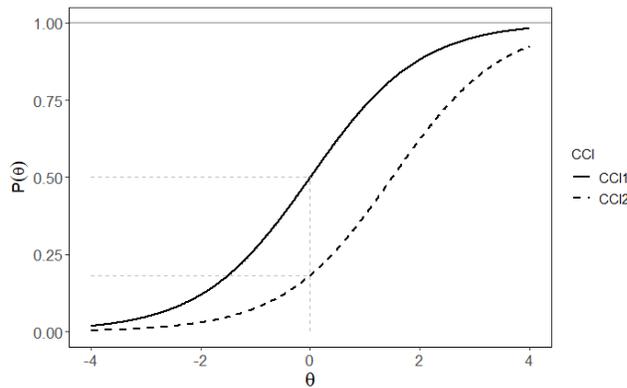


Figura 2.4: Ejemplo de curva característica de del ítem (CCI), probabilidad ( $P(\theta)$ ) vs habilidad o rasgo latente ( $\theta$ ), para dos parámetros de dificultad. La curva color rojo corresponde a  $b = 0$  y la curva color azul a  $b_i = 1,5$ . Las curvas se han generado en R utilizando el código de ShinyItemAnalysis Martinkova and Drabinova (2018) y modificado por el autor de este TFG.

Una propiedad interesante del modelo 1PL es la siguiente, la diferencia de la función logit en dos participantes ( $j, k$ ), en un ítem  $i$  cualquiera, es función de la diferencia de sus habilidades. A este resultado se llega utilizando las 2.9, y las propiedades de los logaritmos:

$$\ln(OR_j) - \ln(OR_k) = \ln \left( \frac{P(Y_{ij} = 1|\theta_j)}{P(Y_{ij} = 0|\theta_j)} \right) - \ln \left( \frac{P(Y_{ik} = 1|\theta_k)}{P(Y_{ik} = 0|\theta_k)} \right) = \theta_j - \theta_k \quad (2.15)$$

Haciendo uso de la notación aquí presentada (Tabla 2.1, Ec. 2.14) y bajo la hipótesis de independencia local, tenemos que la probabilidad de observar, en el  $j$ -ésimo examinado con un nivel de habilidad  $\theta$ , un patrón de respuestas  $\mathbf{y}'_j = (y_{1j}, y_{2j}, \dots, y_{nj})$ ,  $y_{ij} \in \{0, 1\}$  esta dada por la siguiente expresión:

$$P(Y_j = \mathbf{y}_j | \theta_j, b_i) = \prod_{i=1}^n P(Y_{ij} = y_{ij} | \theta_j, b_i) = \prod_{i=1}^n \frac{e^{(\theta_j - b_i)y_{ij}}}{1 + e^{(\theta_j - b_i)}} \quad (2.16)$$

Donde  $Y_j \in \{0, 1\}^n$  es un vector aleatorio cuyas componentes modelan las  $n$  respuestas del  $j$ -ésimo participante. Observar que la Ec. 2.16 corresponde a una función de verosimilitud  $L(\theta | \mathbf{y}_j)$ .

El resultado anterior puede extenderse para todos los participantes del test, bajo la hipótesis que sus respuestas son independientes, de la siguiente manera: tomando como referencia la Tabla 2.1, definimos  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_J)$  el vector de habilidades,  $\mathbf{b} = (b_1, b_2, \dots, b_n)$  el vector de parámetros de dificultad de cada ítem,  $Y \in \{0, 1\}^{n \times J}$  es la matriz de respuestas formada por los vectores columna  $Y_j$ ,  $j = 1, 2, \dots, J$ , son sus columnas; entonces, la probabilidad de observar una matriz de respuestas  $Y = (\mathbf{y}_1, \mathbf{y}_1, \dots, \mathbf{y}_J)$  es:

$$P(Y = \mathbf{y} | \boldsymbol{\theta}, \mathbf{b}) = \prod_{j=1}^J \prod_{i=1}^n P(Y_{ij} = y_{ij} | \theta_j, b_i) = \prod_{j=1}^J \prod_{i=1}^n \frac{e^{(\theta_j - b_i)y_{ij}}}{1 + e^{(\theta_j - b_i)}} \quad (2.17)$$

Al comienzo de la presente sección hacíamos referencia al hecho que 1PL es un caso particular del modelo de dos parámetros, el cual presentaremos a continuación. Además, las Ecs. 2.16 y 2.17 presentan las mismas características que las correspondientes al Modelo de Rasch, el cual abordaremos más adelante y probaremos un importante resultado en lo que refiere al estadístico suficiente para estimar el parámetro de dificultad.

### 2.2.5. Modelo Unidimensional 2PL

El modelo logístico de dos parámetros o Two Parameter Logistic (2PL) surge como una alternativa al modelo ojiva normal, brindando una solución que permite desligarse del cálculo integral. Algunos autores como Santisteban 2009; 2001, Bock y Moustaki en (2007, Chap 15) atribuyen las primeras publicaciones sobre el modelo de dos parámetros a Birnbaum (1957,5,5), posteriormente, de acuerdo a Bartolucci, Bacci y Gnaldi (2015, Chap. 3), P. Mair (2018, Chap. 4), podemos encontrar los modelos logísticos desarrollado por Birnbaum (1968).

Al igual que antes, consideramos un test con  $i = 1, 2, \dots, n$  ítems, en donde  $j = 1, 2, \dots, J$  son los examinados y  $\theta_j$  el nivel de habilidad que éstos poseen. Sea  $Y_{ij} \in \{0, 1\}$  la VA que utilizaremos para modelar la respuesta del  $j$ -ésimo examinado al  $i$ -ésimo ítem. Bajo esta notación, la expresión analítica del modelo 2PL esta dada por:

$$P_i(\theta_j) = P(Y_{ij} = 1|\theta_j, a_i, b_i) = \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}} \quad (2.18)$$

Al parámetro  $a_i$  se lo denomina parámetro de discriminación del ítem, corresponde a la pendiente de la CCI en el punto  $b_i$ . Los ítems con valores grandes de  $a_i$  son mejores para clasificar participantes de acuerdo a sus habilidades. En el caso ideal la pendiente es infinita y se logra una clasificación perfecta. En otras palabras, es posible construir una regla de decisión tal que el nivel de habilidad  $\theta_j = b_i$  permite discriminar dos grupos, aquellos examinados con probabilidad igual a 1 de responder en forma correcta al ítem, respecto de aquellos con probabilidad nula de responder en forma correcta. Tener presente que el rango teórico de variación de  $a_i$  es  $(-\infty, +\infty)$ , sin embargo, los rangos negativos no se tienen en cuenta en TRI y en la práctica  $(0, 2)$  es el rango habitual (Hambleton et al., 1991).

Al igual que hicimos antes, podemos expresar en forma compacta  $P(Y_{ij} = 1|\theta_j, a_i, b_i)$  y  $Q(Y_{ij} = 1|\theta_j, a_i, b_i) = 1 - P(Y_{ij} = 1|\theta_j, a_i, b_i)$ ; para ello utilizaremos la VA  $Y_{ij} \in \{0, 1\}$  que modela la respuestas del j-ésimo individuo al i-ésimo ítem, por lo tanto la expresión analítica es:

$$P(Y_{ij} = y_{ij}|\theta_j, a_i, b_i) = \frac{e^{a_i(\theta_j - b_i)y_{ij}}}{1 + e^{a_i(\theta_j - b_i)}} \quad (2.19)$$

El vínculo entre los modelos 1PL, 2PL y ojiva normal (Ec. 2.5) puede ser visto como un re-escalado, ya que, de acuerdo a Birnbaum (1968), si se utiliza el resultado de Haley (1952) tenemos que:

$$|\Phi(x) - \Psi(1,7x)| < 0,01 \quad \forall x \in \mathbb{R}. \quad (2.20)$$

En forma equivalente:

$$|\Phi(x/1,7) - \Psi(x)| < 0,01 \quad \forall x \in \mathbb{R}. \quad (2.21)$$

Siendo  $\Phi(x)$  la cdp normal y  $\Psi(x)$  la cdp logística que se define<sup>4</sup>:

$$\Psi(x) = \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-x}} \quad (2.22)$$

La Ec. 2.20 o 2.21 nos dice que ambas cdp, normal y logística, difieren menos de 0,01 en forma uniforme para todo el rango de variación de la variable independiente. Es por ello que muchas veces en la literatura, por ejemplo DeMars (2010) o Lord and Novick (1968), podemos encontrar que el modelo 2PL se define utilizando la expresión algebraica dada por la Ec. 2.18 y un factor multiplicativo igual a 1.7, además de los ya mencionados parámetros discriminación y de dificultad.

En la Figura 2.5 se han representado dos ejemplos de la CCI para el modelo 2PL, donde  $[-4, +4]$  es el intervalo en donde varía el rasgo latente. Los valores de los parámetros utilizados son:  $a = 1$  y  $b = 0$  en la curva CCI1; y  $a = 5$  y  $b = 0$  en la curva CCI2. Podemos observar que, para el mismo valor del parámetro de dificultad, la curva CCI2 posee una pendiente mayor respecto de la curva CCI1, este resultado está en consonancia con el hecho que dicho modelo posee un parámetro de discriminación mayor.

---

<sup>4</sup>No confundir con la Ec. 2.26 en donde hemos utilizado la misma notación

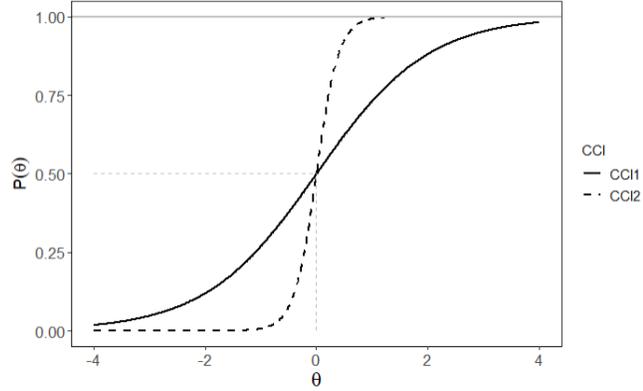


Figura 2.5: Ejemplo de curva característica de del ítem (CCI), probabilidad ( $P(\theta)$ ) vs habilidad o rasgo latente ( $\theta$ ), para dos valores del parámetro de discriminación. La curva CCI1 corresponde a  $a = 1$  y  $b = 0$  y la CCI2 a  $a = 5$  y  $b = 0$ . Las curvas se han generado en R utilizando el código de ShinyItemAnalysis Martinkova and Drabinova (2018) y modificado por el autor de este TFG.

Continuando con el análisis del modelo 2PL, bajo las mismas hipótesis y utilizando la misma estrategia que en la sección anterior, es posible obtener la probabilidad de observar en el vector de respuestas  $\mathbf{y}'_j = (y_{1j}, y_{2j}, \dots, y_{nj})$  en el  $j$ -ésimo participante:

$$\begin{aligned}
 P(Y_j = \mathbf{y}_j | \theta_j, \mathbf{a}, \mathbf{b}) &= L(\mathbf{y}_j | \theta_j, \mathbf{a}, \mathbf{b}) = \prod_{i=1}^n (P(Y_{ij} = y_{ij}))^{y_{ij}} \times (1 - P(Y_{ij} = y_{ij}))^{1-y_{ij}} \\
 &= \frac{e^{\theta_j \sum_{i=1}^n a_i y_{ij} - \sum_{i=1}^n a_i b_i y_{ij}}}{\prod_{i=1}^n (1 + e^{a_i(\theta_j - b_i)})} \quad (2.23)
 \end{aligned}$$

Donde  $\mathbf{b} = (b_1, b_2, \dots, b_n)$ ,  $\mathbf{a} = (a_1, a_2, \dots, a_n)$ , mientras que el término  $y_j^* = \sum_{i=1}^n a_i y_{ij}$  se lo denomina score ponderado, ya que el score de cada ítem se encuentra ponderado por el parámetro de discriminación del ítem.

Por último, dado que estamos bajo la hipótesis de independencia local y suponiendo que las respuestas de los examinados no están correlacionadas, de forma sencilla es posible extender el resultado anterior a la probabilidad de observar todos los vectores de respuestas de la Tabla 2.1. Como resultado se obtiene la Ec. 2.24, siendo  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_J)$  el rasgo latente de los  $J$  examinados.

$$\begin{aligned}
 P(\mathbf{Y} = \mathbf{y} | \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}) &= L(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_J) | \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}) \\
 &= \frac{e^{\sum_{j=1}^J \theta_j y_{\cdot j} - \sum_{i=1}^n a_i b_i y_{ij}}}{\prod_{j=1}^J \prod_{i=1}^n (1 + e^{a_i(\theta_j - b_i)})} \quad (2.24)
 \end{aligned}$$

Al incluir en el modelo el parámetro de discriminación se pierde la propiedad objetividad específica, en consecuencia, la diferencia entre los OR ya no depende únicamente de la diferencia de niveles de

habilidad entre individuos para cualquier ítem, o de la diferencia entre los parámetros de dificultad entre dos ítems para un mismo individuo. Además, también se pierde la propiedad sobre los estadísticos suficientes, es decir, la suma del score por participante y por ítem ya no son estadísticos conjuntamente suficientes para el nivel de habilidad y el parámetro de discriminación. En consecuencia, observamos que existe un compromiso contrapuesto entre las propiedades intrínsecas de 1PL e incorporar un nuevo parámetro en el modelo.

### 2.3. Modelo de Rasch

El Modelo de Rasch (MR) lleva el nombre, de acuerdo Andersen (1977) Andersen and Wøhlk (2001), del matemático danés G. Rasch, quien propuso a comienzos de la década de 1960 un nuevo enfoque en el análisis de la CCI Rasch (1960). Su trabajo se realiza en forma independiente, y casi en simultáneo, a la propuesta de Lod y Birnbaum de los modelos 2PL y Three Parameter Logistic (3PL). Más allá que el MR puede ser visto como un caso particular del modelo de Birnbaum, como veremos a continuación, su concepción es a partir de una perspectiva diferente.

Supongamos que  $P(Y_{ij} = 1|\xi_j)$  es la probabilidad que tiene el  $j$ -ésimo individuo de responder correctamente al  $i$ -ésimo ítem, dado un nivel del rasgo latente  $\xi_j$ , mientras  $P(Y_i = 0|\xi) = 1 - P(Y_i = 1|\xi)$  es la probabilidad de responder en forma incorrecta. La propuesta de Rasch, según Andersen (1977), White (1976) y Maydeu-Olivares and Montaña (2013), consiste en definir  $P(Y_{ij} = 1|\xi_j)$  de la siguiente manera:

$$P(Y_i = 1|\xi_j) = h(\Omega) \quad (2.25)$$

Siendo  $\Omega = (\xi_j, \delta_i) \in \mathbb{R}^{2+}$ , donde  $\xi_j$  el rasgo latente del examinado,  $\delta_i$  es un parámetro definido como dificultad del  $i$ -ésimo ítem y  $h(\cdot)$  (definida en Ec. 2.26) es una transformación biyectiva de  $\mathbb{R}$  a el intervalo  $[0, 1]$ .

$$h(x) = \frac{x}{1+x} \quad (2.26)$$

Específicamente el modelo propuesto por Rasch es:

$$P(Y_i = 1|\xi_j) = \frac{\xi_j}{\xi_j + \delta_i} \quad (2.27)$$

El modelo de la Ec. 2.27 puede interpretarse de la siguiente manera: la probabilidad de responder correctamente a un ítem es una función que depende únicamente del rasgo latente del sujeto y un parámetro intrínseco a cada ítem. Además, el vínculo con el modelo 2PL (Ec. 2.12) es simplemente la transformación  $\xi = e^{a\theta}$  y  $\delta_i = e^{ab_i}$ , siendo  $a$  el parámetro de discriminación (idéntico para todos los ítems) y  $b_i$  el parámetro de dificultad. En otras palabras, el MR se construye sobre la hipótesis que todos los ítems presentan igual poder de discriminación  $a_i = a \forall i = 1, 2, \dots, n.$ , lo que equivale a decir, las CCI presentan la misma pendiente.

Frecuentemente en la práctica se trabaja con  $a = 1$  (Santisteban Requena, 2009; Santisteban Requena and Alvarado, 2001), por lo tanto, los resultados presentados en la sección anterior para el modelo 1PL son también aplicables al MR. Además, en la literatura el modelo suele presentarse a través de las siguientes expresiones:

$$P_i(\theta_j) = P(Y_{ij} = 1|\theta_j, b_i) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}} \quad , \quad \theta_j, b_i \in \mathbb{R} \quad (2.28)$$

$$Q_i(\theta_j) = P(Y_{ij} = 0|\theta_j, b_i) = 1 - P(Y_{ij} = 1|\theta_j, b_i) = \frac{1}{1 + e^{(\theta_j - b_i)}} \quad (2.29)$$

Desde una perspectiva puramente matemática el MR y el modelo 2PL (cuando  $a = 1$ ) hacen uso de idénticas expresiones analíticas para modelar la CCI, sin embargo, parten de concepciones distintas en el ajuste de la curva. El primero de éstos, se basa en la hipótesis que el parámetro de discriminación es igual para todos los ítems, mientras que en el segundo se busca el mejor ajuste del modelo en base a los datos observados sin restringir a un único valor el parámetro de discriminación (Andrich, 2004; Columbia Public Health). Ellis en su tesis (Ellis, 2011, Chap. 2) nos dice que el MR es confirmatorio ya que los datos deben de ajustarse al modelo, mientras que 2PL es un modelo exploratorio debido a que el modelo se ajusta a los datos.

Vale la pena señalar que, al igual que antes, si las respuestas de los participantes siguen la notación utilizada en la Tabla 2.1, entonces un patrón de respuestas  $\mathbf{y}_j$  puede modelarse como  $n$  experimentos Bernoulli. Por lo tanto, bajo la hipótesis de independencia local, la probabilidad de observar dicho patrón es:

$$P(Y = \mathbf{y}|\xi) = \prod_{i=1}^n P(Y_i = 1|\xi)^{y_i} \cdot P(Y_i = 0|\xi)^{1-y_i} \quad (2.30)$$

Luzardo (2017) y McDonald (2011) resaltan algunas bondades interesantes del modelo 1PL, entre ellas se destacan: a) la probabilidad de observar en el  $j$ -ésimo participante un patrón de respuestas  $\mathbf{y}_j = (y_{1j}, y_{2j}, \dots, y_{nj})$ , o en forma equivalente la función de verosimilitud  $L(\theta_j|\mathbf{y}_j)$ , presenta una expresión sencilla; b) el score total del examinado es un estadístico suficiente para  $\theta_j$ , a éste lo notaremos como  $T = t(Y_j)$ . Es decir, la probabilidad de observar un vector de respuestas dado ( $P(Y_j = \mathbf{y}_j)$ ) depende únicamente del score total. Además, éste es un estadístico mínimo suficiente. En nuestro caso, dado que  $y_i \in \{0, 1\}$ ,  $\forall i = 1, 2, \dots, n$ , entonces el score total de  $j$ -ésimo examinado es  $s_j = t(\mathbf{y}_j) = y_{.j} = \sum_{i=1}^n y_{ij}$ .

Una demostración para el estadístico suficiente fue propuesta por Andersen (1977), quien utilizó algunos de sus resultados ya publicados en 1973a. La propuesta se basa en el Teorema 1 y el Corolario 1 que presentamos a continuación.

**Teorema 1.** *Si existe un estadístico mínimo suficiente para  $\theta_j$ , que es independiente de  $b_i$ ,  $i = 1, 2, \dots, n$ , entonces  $P(Y_{ij} = 1|\theta_j, b_i)$  tiene la forma:*

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

para ciertas funciones  $\varphi(\theta)$ ,  $\psi(b_i)$ .

*Dem. ver Teorema 1. pág. 72 (Andersen, 1977).*

**Corolario 1.** *Si existe un estadístico mínimo suficiente para el parámetro  $\theta_j$  que es independiente de  $b_i$ ,  $i = 1, 2, \dots, n$ , entonces el score total es el estadístico mínimo suficiente y el modelo es el modelo de Rasch.*

*Dem. ver Corolario 1. pág. 72 (Andersen, 1977).*

Un camino alternativo, que nos conduzca a obtener el estadístico mínimo suficiente para  $\theta_j$  es, utilizar la definición de familia exponencial y el Teorema de Factorización de Neyman (ver Apéndice A). De esta forma se obtiene que  $y_{\cdot j} = \sum_{i=1}^n y_{ij}$  es un estadístico suficiente para  $\theta_j$  (Olive, 2014, Chap. 4, Example 4.8).

Continuando en esta línea de razonamiento, otro resultado relevante enlaza el vínculo entre el estadístico suficiente y un subconjunto de ítems en el test. Fischer y Molenaar prueban que, bajo las hipótesis del MR, si  $y_{\cdot j}$  es un estadístico suficiente para estimar el rasgo latente  $\theta_j$  en un test con  $n$  ítems, entonces  $y_{\cdot j}^{(S)} = \sum_{i=1}^n y_{ij} I_S$  es también un estadístico suficiente para  $\theta_j$  para algún  $S$ . Donde  $S$  es un subconjunto (no vacío) de los  $n$  ítems, e  $I_S = \{0, 1\}$  es la función indicatriz que, toma el valor 1 si el ítem pertenece al subconjunto  $S$  y cero en caso contrario (Fischer and Molenaar, 1995, Lemma 2.1).

Supongamos que  $J$  examinados participan de un test formado por  $n$  ítems, donde cada uno de éstos tiene un score igual a 1 punto cuando la respuesta es correcta y 0 en caso contrario. Sea  $y_{ij} \in \{0, 1\}$  la respuesta al ítem  $i$  del participante  $j$ , por lo tanto, el score total de dicho participante esta dado por  $y_{\cdot j} = \sum_{i=1}^n y_{ij}$ ,  $0 \leq y_{\cdot j} \leq n$ ,  $\forall j = 1, 2, \dots, n$ . Además utilizaremos la siguiente notación  $y_{i\cdot} = \sum_{j=1}^J y_{ij}$ ,  $0 \leq y_{i\cdot} \leq J$ ,  $\forall i = 1, 2, \dots, n$ . Observar que a la cantidad  $y_{i\cdot}$  corresponde a la cantidad de respuestas correctas que han proporcionado los examinados en el  $i$ -ésimo ítem. Esta información puede representarse en una matriz de  $n \times J$ , en la Tabla 2.1 puede observarse un ejemplo.

Al igual que antes, la Ec. 2.28 nos permite modelar la probabilidad de responder correctamente al  $i$ -ésimo ítem por parte del  $j$ -ésimo participante dado que  $\theta_j$  es su nivel de habilidad. Sea  $Y = (Y_1, Y_2, \dots, Y_J)$  el vector aleatorio que representa el score total (o la cantidad de respuestas correctas, debido a que estamos considerando ítems binarios) de cada examinado. Se puede observar que  $Y_{ij} \sim Ber(P_{ij})$ , con  $P_{ij} = P(Y_{ij} = y_{ij} | \theta_j, b_i)$ . Bajo estas condiciones, siguiendo la línea de razonamiento de G. Fisher (2007, Chap. 16), Barndorff-Nielsen (2014, Ch. 9, Example 9.14) y R. Sundberg (2019, Chap. 12), tenemos que la probabilidad de observar en el participante  $j$  un patrón de respuestas  $\mathbf{y}'_j = (y_{1j}, y_{2j}, \dots, y_{nj})$  esta dado por:

$$\begin{aligned} P(\mathbf{Y}_j = \mathbf{y}_j | \theta_j, \mathbf{b}) &= L(\mathbf{y}_j | \theta_j, \mathbf{b}) = \prod_{i=1}^n (P(Y_{ij} = y_{ij}))^{y_{ij}} \times (1 - P(Y_{ij} = y_{ij}))^{1-y_{ij}} \\ &= \frac{e^{\theta_j y_{\cdot j} - \sum_{i=1}^n b_i y_{ij}}}{\prod_{i=1}^n (1 + e^{(\theta_j - b_i)})} \end{aligned} \quad (2.31)$$

Donde  $\mathbf{b} = (b_1, b_2, \dots, b_n)$  corresponde al vector de parámetros de dificultad de cada ítem,  $y_{\cdot j}$  es el score que obtuvo el  $j$ -ésimo participante en el test. Además, en la Ec. 2.31 hemos utilizado la Ec. 2.14, es decir, la hipótesis de independencia local, así como también propiedades de las potencias.

El resultado anterior lo podemos extender, bajo las hipótesis del MR, para todos los participantes del test, según lo que se ha registrado en la Tabla 2.1, de la siguiente manera:

$$P(\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_1, \dots, \mathbf{y}_J) | \boldsymbol{\theta}, \mathbf{b}) = L((\mathbf{y}_1, \mathbf{y}_1, \dots, \mathbf{y}_J) | \boldsymbol{\theta}, \mathbf{b}) = \frac{e^{\sum_{j=1}^J \theta_j y_{\cdot j} - \sum_{i=1}^n b_i y_{i\cdot}}}{\prod_{j=1}^J \prod_{i=1}^n (1 + e^{(\theta_j - b_i)})} \quad (2.32)$$

En el desarrollo de la expresión anterior, nuevamente, se ha utilizado la hipótesis de independencia local e independencia de respuestas entre participantes. Observar también que el término  $y_i$  corresponde a la cantidad de examinados que respondieron en forma correcta al  $i$ -ésimo ítem y  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_J)$  es el vector de habilidades.

En primer lugar, a partir de la Ec. 2.32, es inmediato observar que, las matrices de respuestas, como la presentada en la Tabla 2.1, con igual suma por filas y columnas, presentan idéntica expresión algebraica de la función verosimilitud.

En segundo lugar, la función de verosimilitud pertenece a la familia exponencial. Entonces, al aplicar el Teorema de Factorización<sup>5</sup> se obtiene una parametrización tal que, el vector de estadísticos ( $T = (T_\theta, T_b)$ ) depende únicamente del score que obtuvo cada participante y la cantidad de respuestas correctas en cada ítem. En otras palabras, bajo las hipótesis del MR, dichos estadísticos son conjuntamente suficientes para  $(\theta_j, b_i)$ . Por lo tanto, los totales por columna y fila ( $y_{\cdot j} = \sum_{i=1}^n y_{ij}$ ,  $y_{i \cdot} = \sum_{j=1}^J y_{ij}$ ), contienen toda la información acerca de  $\Theta = (\theta_j, b_i)$  disponible en la muestra de respuestas.

Además de la suficiencia, otra propiedad que posee el Modelo de Rasch es la denominada objetividad específica. Según Fischer (1987) e Irtel (1995), ésta fue introducida por el propio Rasch entre 1967 y 1977 quien estableció: para todo ítem  $i = 1, 2, \dots, n$ , la diferencia del cociente de probabilidades  $P_i/Q_i$  en dos examinados  $j$  y  $l$  cualesquiera, cuyas habilidades son  $\theta_j$  y  $\theta_l$  respectivamente, es función únicamente de la diferencia de habilidades. A este resultado se llega en forma inmediata aplicando la definición del modelo (Ecs. 2.28, 2.29), la definición de OR (Ec. 2.9) y la función logit (Ec. 2.10), como se muestra a continuación:

$$\ln(OR_{\theta_j, b_i}) - \ln(OR_{\theta_l, b_i}) = \theta_j - \theta_l \quad (2.33)$$

Se dice que es una propiedad objetiva ya que al comparar la razón de probabilidades, ésta sólo depende de la diferencia de habilidades. A su vez, es específica debido a que no depende del ítem en consideración (Luzardo, 2017).

Análogamente, para un individuo  $j \in \{0, 1, 2, \dots, J\}$  con un nivel de habilidad  $\theta_j$ , al comparar el cociente de probabilidades  $P_i/Q_j$  para dos ítems cualesquiera ( $b_i, b_k, \forall i, k \in \{0, 1, 2, \dots, n\}$ ), se obtiene una función que depende únicamente de la diferencia de los parámetros de dificultad de cada ítem. Procediendo igual que antes, tenemos que:

$$\ln(OR_{\theta_j, b_i}) - \ln(OR_{\theta_j, b_k}) = b_k - b_i \quad (2.34)$$

Estas propiedades junto con el estadístico suficiente son sumamente relevantes ya que pueden utilizarse en la estimación de los parámetros del modelo, así como también en la comparación de examinados e ítems.

El MR marcó un mojón importante en el desarrollo de la TRI, sin embargo, parte del supuesto que únicamente es la dificultad del ítem una característica de influencia. Este supuesto, denominado

---

<sup>5</sup>Ver Apéndice A.

homogeneidad en la dificultad del ítem, puede imponer en ciertos casos una restricción importante y en consecuencia su validez queda restringida a un subconjunto de test y participantes. A pesar de ello, su uso se ha extendido a diferentes áreas debido a la sencillez intrínseca que posee, tanto su aplicación como en la interpretación. En el Apéndice C hemos incluido un conjunto de modelos que corresponden a extensiones al modelo aquí presentado.

## 2.4. Observaciones Finales

Una diferencia importante entre el modelo de la TCT y TRI radica en el hecho que, el segundo nos permite calcular la probabilidad de responder en forma correcta un ítem por parte de un examinado, dado un nivel de habilidad; mientras que en el primero la habilidad del examinado para responder en forma correcta depende del conjunto de ítems en consideración.

Los modelos ojiva normal son poco utilizados en la práctica debido a la complejidad que implica el cálculo de la integral. Además, dado que es posible obtener buenas aproximaciones a los modelos logísticos, son éstos últimos preferidos en diversas aplicaciones.

Existen otros modelos, además de los presentados en la sección de TRI, son los denominados 3PL y Four Parameter Logistic (4PL), en donde se incorporan nuevos parámetros al modelo 2PL. De acuerdo a Santisteban Requena and Alvarado (2001), éstos han sido introducidos con el objetivo de resolver los problemas de ajuste, en particular, cuando se observan datos de examinados, que no responden correctamente a ítems que deberían resultar “fáciles” para su nivel de habilidad. Sin embargo, diversos autores cuestionan la ganancia que se obtiene, desde el punto de vista práctico, al adicionar complejidad en el modelo. En nuestro TFG el análisis de dichos modelos ha quedado fuera del alcance.

Esta página ha sido intencionalmente dejada en blanco.

## Capítulo 3

# Estimación de los Parámetros y Ajuste en el Modelo Logístico

### 3.1. Introducción

Una vez que se ha elegido el modelo a utilizar, el siguiente paso es hacer uso del conjunto de herramientas disponibles con el fin de estimar sus parámetros y validar la bondad en el ajuste. En la presente sección estaremos abordando las características de las técnicas comúnmente utilizadas, quedando fuera del alcance de nuestro TFG la estimación de parámetros en el modelo ojiva normal y los métodos no paramétricos.

Al hablar de estimación de parámetros del modelo nos estamos refiriendo a los parámetros  $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^n$  utilizados en el ajuste de la CCI, así como también de  $\boldsymbol{\theta} \in \mathbb{R}^J$ , es decir, el rasgo latente. Recordemos que, a priori, la única información disponible para este fin es el conjunto de respuestas a cada ítem del test proporcionadas por los participantes.

En la literatura, generalmente, se presentan dos estrategias para este fin, a saber, los métodos de estimación condicional y estimación conjunta. El primero de ellos consiste en estimar el rasgo latente bajo la hipótesis que los parámetros de dificultad y discriminación son conocidos o viceversa; mientras que en el segundo la estimación de todos los parámetros se realiza en forma conjunta, siendo éste el más utilizado en la práctica. Bajo el alcance de los problemas en la TRI, dicha estimación se realiza ítem a ítem.

Los métodos antes mencionados se basan en la hipótesis que el rasgo latente es un valor fijo asociado a cada individuo. En la literatura se lo suele denominar efecto fijo o fixed-effect. Sin embargo, en ciertos problemas debido a las características del experimento, o cuando se quiere lograr estimaciones consistentes, esta asunción no es válida. Por lo tanto, es necesario modelar al rasgo latente como una VA que proviene de un modelo de probabilidad con cierta función de distribución. En estos casos se dice que el modelo es de efectos mixtos, donde  $\boldsymbol{\theta}$  es un efecto aleatorio o random-effect y los parámetros  $(\mathbf{a}, \mathbf{b})$  son los efectos fijos. Este último enfoque será presentando en el siguiente capítulo.

Una vez que se han estimado los parámetros involucrados en el modelo, necesitamos herramientas que nos ayuden a evaluar la bondad en el ajuste por parte del modelo a los datos observados. En este sentido existen, en líneas generales, dos grupos de herramientas: gráficas y estadísticas. Las segundas serán objeto de estudio en el presente capítulo. Una excelente descripción sobre las primeras puede encontrarse, por ejemplo, en Bartolucci et al. (2015).

A continuación estaremos abordando las técnicas de estimación y evaluación de modelos. El desarrollo que subsigue esta basado principalmente en los trabajos de Baker and Kim (2004), Bartolucci et al. (2015), Cai and Thissen (2014), Fischer and Molenaar (1995), Melià (2003) y Santisteban Requena and Alvarado (2001). Los resultados que se exponen son para el caso del modelo 2PL, sin embargo, el procedimiento es análogo para los modelos 1PL.

## 3.2. Método de Máxima Verosimilitud en la Estimación de los Parámetros de Discriminación y Dificultad

La propuesta de utilizar el método de máxima verosimilitud surge de Maxwell (1959), quien utiliza el modelo logístico, como alternativa al modelo ojiva normal, a la hora de estimar los parámetros. Para ello se basó en los trabajos de Berkson (1951,5,5), quien recientemente había publicado una serie de resultados sobre función logística. En la presente sección analizaremos el método máxima verosimilitud, aplicado al modelo 2PL, con el objetivo de estimar el parámetro de discriminación y dificultad de cada ítem en el test cuando el rasgo latente es conocido.

Supongamos que un test esta formado por  $n$  ítems dicotómicos y nos interesa estimar los parámetros del modelo 2PL para el  $i$ -ésimo ítem,  $i = 1, \dots, n$ . Además, vamos a suponer que los participantes están agrupados en  $J$  grupos de acuerdo al nivel de rasgo latente (conocido)  $\theta_j$ ,  $j = 1, \dots, J$ . En cada grupo hay  $f_j$  individuos y denotaremos con  $r_j$  a la cantidad de respuestas correctas que han proporcionado, por lo tanto,  $f_j - r_j$  es la cantidad de respuestas incorrectas. Por último, llamaremos  $R = (r_1, r_2, \dots, r_n)$  al vector de respuestas correctas observadas.

En el  $j$ -ésimo grupo, la proporción de respuestas correctas ( $p_j$ ) esta dada por:

$$p_j = p(\theta_j) = \frac{r_j}{f_j} \quad (3.1)$$

En consecuencia, la proporción de respuestas incorrectas es:

$$q_j = q(\theta_j) = \frac{f_j - r_j}{f_j} \quad (3.2)$$

Bajo la hipótesis que las respuestas de los examinados son independientes, podemos suponer que  $r_j \sim Bin(f_j, p_j)$ . Haciendo uso de Ec. 2.18 tenemos que la probabilidad de observar en el  $i$ -ésimo ítem,  $i \in \{1, 2, \dots, n\}$ , el vector  $R$  de respuestas correctas es:

$$P(Y = R) = L = \prod_{j=1}^J \binom{f_j}{r_j} P_j^{r_j} Q_j^{f_j - r_j} \quad (3.3)$$

Donde, de acuerdo a la notación antes utilizada,  $P_j = P_i(\theta_j)$ ,  $Q_j = Q_i(\theta_j) = 1 - P_i(\theta_j)$  y  $L$  corresponde a la distribución conjunta o función de verosimilitud. Aplicando el método de máxima verosimilitud (ver Apéndice A) buscamos maximizar  $L$ , es decir, encontrar  $\hat{a}_i, \hat{b}_i$  donde  $L(a_i, b_i | \theta_j)$  se maximiza, que equivale a maximizar su logaritmo<sup>1</sup>.

$$l = \ln(L) = \sum_{j=1}^J \binom{f_j}{r_j} + \sum_{j=1}^J r_j \ln(P_j) + \sum_{j=1}^J (f_j - r_j) \ln(Q_j) \quad (3.4)$$

La expresión anterior representa a la función objetivo, mientras que las restricciones forman parte del recorrido de los parámetros involucrados.

<sup>1</sup>Recordemos que la función logaritmo es monótona creciente, por lo tanto maximizar  $L$  es equivalente a maximizar  $\ln L$ .

### 3.2.1. Modelo 2PL

En la práctica, cuando se trabaja con el modelo 2PL, se suele realizar una reparametrización de la Ec. 2.18 definiendo  $\nu_i = -a_i b_i$ , por lo tanto  $P_i(\theta_j)$  puede expresarse de la siguiente manera:

$$P_j = P_i(\theta_j) = P(Y_{ij} = 1 | \theta_j, a_i, b_i) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad (3.5)$$

$$Q_j = Q_i(\theta_j) = P(Y_{ij} = 0 | \theta_j, a_i, b_i) = \frac{1}{1 + e^{a_i(\theta_j - b_i)}} \quad (3.6)$$

Maximizar 3.4 implica buscar  $\hat{a}_i = \hat{a}_i(\theta_1, \dots, \theta_J)$ ,  $\hat{\nu}_i = \hat{\nu}_i(\theta_1, \dots, \theta_J)$  tal que  $\partial l / \partial a_i = 0$ ,  $\partial l / \partial \nu_i = 0$  y se debe de cumplir el criterio del Hessiano. Esto nos lleva a resolver el sistema de ecuaciones:

$$\frac{\partial l}{\partial a_i} \Big|_{(\hat{a}_i, \hat{\nu}_i)} = 0 \quad , \quad \frac{\partial l}{\partial \nu_i} \Big|_{(\hat{a}_i, \hat{\nu}_i)} = 0 \quad (3.7)$$

Sea  $W_j = P_j Q_j$  los pesos de Urban-Müller, si derivamos 3.5, 3.6 y sustituimos en 3.7 tenemos que:

$$\frac{\partial l}{\partial a_i} = \sum_{j=1}^J \theta_j f_j W_j \left( \frac{p_j - P_j}{P_j Q_j} \right) \quad , \quad \frac{\partial l}{\partial \nu_i} = \sum_{j=1}^J f_j W_j \left( \frac{p_j - P_j}{P_j Q_j} \right) \quad (3.8)$$

$$\frac{\partial^2 l}{\partial a_i^2} = \sum_{j=1}^J f_j W_j \quad , \quad \frac{\partial^2 l}{\partial \nu_i^2} = \sum_{j=1}^J \theta_j^2 f_j W_j \quad , \quad \frac{\partial^2 l}{\partial \nu_i \partial a_i} = \frac{\partial^2 l}{\partial a_i \partial \nu_i} = \sum_{j=1}^J \theta_j f_j W_j \quad (3.9)$$

La búsqueda de soluciones a el sistema (no lineal) de ecuaciones 3.7, que maximicen  $l$ , se realiza a través de algoritmos de aproximaciones sucesivas. Uno de los procedimientos utilizados es el método de Newton-Raphson<sup>2</sup>, que consiste en iterar haciendo uso de la serie de Taylor de primer orden en torno a  $(\hat{a}_i, \hat{\nu}_i)$ , punto que maximiza  $l = l(a_i, \nu_i)$ .

En el proceso de iteración<sup>3</sup> se asume que  $\hat{a}_i^{(k+1)} = \hat{a}_i^{(k)} + \Delta \hat{a}_i^{(k)}$  y  $\hat{\nu}_i^{(k+1)} = \hat{\nu}_i^{(k)} + \Delta \hat{\nu}_i^{(k)}$ ,  $k = 0, 1, 2, \dots$ , donde  $\hat{a}_i^{(k)}$  y  $\hat{\nu}_i^{(k)}$  son las aproximaciones a  $\hat{a}_i$  y  $\hat{\nu}_i$  (soluciones al sistema 3.7), mientras que  $\Delta \hat{a}_i^{(k)}$  y  $\Delta \hat{\nu}_i^{(k)}$  son los errores en la aproximación. Por lo tanto, la serie de Taylor, expresada en forma matricial, en torno a  $(\hat{a}_i^{(k)}, \hat{\nu}_i^{(k)})$  de las funciones  $\partial l / \partial a_i$  y  $\partial l / \partial \nu_i$  es:

$$\begin{bmatrix} \frac{\partial l}{\partial a_i} \\ \frac{\partial l}{\partial \nu_i} \end{bmatrix}_{(k+1)} = \begin{bmatrix} \frac{\partial l}{\partial a_i} \\ \frac{\partial l}{\partial \nu_i} \end{bmatrix}_{(k)} + \begin{bmatrix} \frac{\partial^2 l}{\partial a_i^2} & \frac{\partial^2 l}{\partial \nu_i \partial a_i} \\ \frac{\partial^2 l}{\partial a_i \partial \nu_i} & \frac{\partial^2 l}{\partial \nu_i^2} \end{bmatrix}_{(k)} \cdot \begin{bmatrix} \Delta \hat{a}_i^{(k)} \\ \Delta \hat{\nu}_i^{(k)} \end{bmatrix} \quad (3.10)$$

Los subíndices  $(k+1)$  y  $(k)$  indican que las derivadas parciales son evaluadas en  $(\hat{a}_i^{(k+1)}, \hat{\nu}_i^{(k+1)})$  y  $(\hat{a}_i^{(k)}, \hat{\nu}_i^{(k)})$  respectivamente. Si  $(\hat{a}_i^{(k+1)}, \hat{\nu}_i^{(k+1)})$  es solución de la Ec. 3.7, entonces el lado izquierdo

<sup>2</sup>Una descripción más completa de este método puede encontrarse en A.

<sup>3</sup>El desarrollo completo de los conceptos aquí presentados se encuentra en el Apéndice B, Sec. B.1.

de la igualdad es cero. Por lo tanto, se obtiene un sistema lineal en  $\Delta\hat{a}_i^{(k)}$  y  $\Delta\hat{\nu}_i^{(k)}$  cuya solución es:

$$\begin{aligned} \begin{bmatrix} \Delta\hat{a}_i^{(k)} \\ \Delta\hat{\nu}_i^{(k)} \end{bmatrix} &= - \begin{bmatrix} \frac{\partial^2 l}{\partial \hat{a}_i^2} & \frac{\partial^2 l}{\partial \hat{\nu}_i \partial \hat{a}_i} \\ \frac{\partial^2 l}{\partial \hat{\nu}_i \partial \hat{a}_i} & \frac{\partial^2 l}{\partial \hat{\nu}_i^2} \end{bmatrix}_{(k)}^{-1} \cdot \begin{bmatrix} \frac{\partial l}{\partial \hat{a}_i} \\ \frac{\partial l}{\partial \hat{\nu}_i} \end{bmatrix}_{(k)} \\ &= - \begin{bmatrix} -\sum_{j=1}^J \theta_j^2 f_j W_j & -\sum_{j=1}^J \theta_j f_j W_j \\ -\sum_{j=1}^J \theta_j f_j W_j & -\sum_{j=1}^J f_j W_j \end{bmatrix}_{(k)}^{-1} \cdot \begin{bmatrix} \sum_{j=1}^J \theta_j f_j W_j \left( \frac{p_j - P_j}{P_j Q_j} \right) \\ \sum_{j=1}^J f_j W_j \left( \frac{p_j - P_j}{P_j Q_j} \right) \end{bmatrix}_{(k)} \end{aligned} \quad (3.11)$$

En la expresión anterior hemos utilizado las Ecs. 3.8 y 3.9. Ahora bien, si esta aproximación es buena, entonces  $\Delta\hat{a}_i^{(k)} < \epsilon_{a_i}$  y  $\Delta\hat{\nu}_i^{(k)} < \epsilon_{\nu_i}$ , para cualquier  $\epsilon_{a_i}, \epsilon_{\nu_i}$  arbitrarios pequeños y en consecuencia hemos encontrado la solución, en caso contrario se continúa iterando.

Una vez que se converge a una solución, ésta puede expresarse de la siguiente manera:

$$\begin{aligned} \begin{bmatrix} \hat{a}_i^{(k+1)} \\ \hat{\nu}_i^{(k+1)} \end{bmatrix} &= \begin{bmatrix} \hat{a}_i^{(k)} \\ \hat{\nu}_i^{(k)} \end{bmatrix} + \begin{bmatrix} \Delta\hat{a}_i^{(k)} \\ \Delta\hat{\nu}_i^{(k)} \end{bmatrix} \\ &= \begin{bmatrix} \hat{a}_i^{(k)} \\ \hat{\nu}_i^{(k)} \end{bmatrix} - \begin{bmatrix} -\sum_{j=1}^J \theta_j^2 f_j W_j & -\sum_{j=1}^J \theta_j f_j W_j \\ -\sum_{j=1}^J \theta_j f_j W_j & -\sum_{j=1}^J f_j W_j \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_{j=1}^J \theta_j f_j W_j \left( \frac{p_j - P_j}{P_j Q_j} \right) \\ \sum_{j=1}^J f_j W_j \left( \frac{p_j - P_j}{P_j Q_j} \right) \end{bmatrix} \end{aligned} \quad (3.12)$$

Siendo  $\hat{a}_i^{(k+1)}$  y  $\hat{\nu}_i^{(k+1)}$  la solución buscada al sistema de Ecs. 3.7. A partir de estos valores es posible obtener la estimación para el parámetro de dificultad:  $\hat{b}_i = -\hat{\nu}_i / \hat{a}_i$ .

Berkson (1951,5) prueba que  $\sum_{j=1}^J f_j p_j$  y  $\sum_{j=1}^J f_j p_j$  son estadísticos suficientes de  $\nu_i$  y  $a_i$  respectivamente, bajo la hipótesis que  $\theta_j$  es conocido. Además, Berkson (1953) provee una expresión para la varianza de  $\hat{a}_i$ ,  $\hat{\nu}_i$  y  $\hat{b}_i$ :

$$\begin{aligned}
Var(\hat{a}_i) &= \frac{1}{\sum_{j=1}^J f_j W_j (\theta_j - \tilde{\theta})^2} \\
Var(\hat{v}_i) &= \frac{1}{\sum_{j=1}^J f_j W_j} \tilde{\theta}^2 Var(\hat{a}_i) \\
Var(\hat{b}_i) &= \frac{1}{\hat{a}_i^2} \cdot \left( \frac{1}{\sum_{j=1}^J f_j W_j} + Var(\hat{a}_i) (\hat{b}_i - \tilde{\theta})^2 \right)
\end{aligned} \tag{3.13}$$

Siendo  $\tilde{\theta}$  :

$$\tilde{\theta} = \frac{\sum_{j=1}^J f_j \theta_j W_j}{\sum_{j=1}^J f_j W_j} \tag{3.14}$$

Los resultados anteriores permiten obtener una estimación de los parámetros involucrados en el modelo 2PL, así como también su varianza.

### 3.3. Método de Máxima Verosimilitud en la Estimación del Rasgo Latente

El desarrollo del método de máxima verosimilitud presentado en la sección anterior partía de la base que el rasgo latente ( $\theta_j$ ) de los examinados era conocido. Sin embargo, en la práctica los tests son diseñados con el fin de ser un instrumento que permita estimar dicho valor. En la presente sección estaremos presentando el método de máxima verosimilitud para estimar el rasgo latente en el caso en que los parámetros del modelo (dificultad y discriminación) son conocidos.

Las hipótesis utilizadas en la estimación de  $\theta_j$  están vinculadas al hecho que los examinados pueden considerarse independientes unos de otros. Es decir, las respuestas que éstos proporcionan son independientes entre sí, sumado al hecho que es válido admitir que su habilidad es estimable a través de un test.

Al igual que hicimos antes, supondremos que  $J$  es la cantidad de examinados, el test está formado por  $n$  ítems que admiten respuestas dicotómicas, las cuales notaremos como  $y_{ij} \in \{0, 1\}$ , es decir, la respuesta al  $i$ -ésimo ítem por parte del  $j$ -ésimo examinado, donde  $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, J$ . De modo que  $\mathbf{y}_j = (y_{1j}, y_{2j}, \dots, y_{nj})^T$  es el vector columna que contiene las respuestas proporcionadas por el  $j$ -ésimo examinado en el test.

Por último, llamemos  $P_i(\theta_j)$  a la probabilidad que tiene el examinado  $j$  de responder correctamente al ítem  $i$  dado que  $\theta_j$  es su nivel de habilidad. Mientras que  $Q_i(\theta_j) = 1 - P_i(\theta_j)$  es la probabilidad de responder en forma incorrecta.

A continuación presentamos un bosquejo del método Maximum Likelihood (ML) para el caso en que los parámetros del modelo 2PL son conocidos.

### 3.3.1. Modelo 2PL

Supongamos que estamos modelando la CCI a través de un modelo 2PL, por lo tanto, utilizando los resultados de la Sec 2.2.5 tenemos que  $P_i(\theta_j)$  y  $Q_i(\theta_j)$  están dadas por las Ec.s 3.15 y 3.16 respectivamente.

$$P_i(\theta_j) = P(Y_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad (3.15)$$

$$Q_i(\theta_j) = Q(Y_{ij} = 0|\theta_j) = \frac{1}{1 + e^{a_i(\theta_j - b_i)}} \quad (3.16)$$

La probabilidad de observar el vector de respuestas  $\mathbf{y}_j$  esta dada por:

$$P(\mathbf{y}_j|\theta_j) = L(\theta_j|\mathbf{y}_j) = \prod_{i=1}^n [P_i(\theta_j)]^{y_{ij}} [Q_i(\theta_j)]^{1-y_{ij}} \quad (3.17)$$

Siendo  $L(\theta_j|\mathbf{y}_j)$  la función de verosimilitud, cuyo logaritmo es:

$$l(\theta_j|\mathbf{y}_j) = \sum_{i=1}^n \ln ([P_i(\theta_j)]^{y_{ij}} [Q_i(\theta_j)]^{1-y_{ij}}) \quad (3.18)$$

En la expresión anterior  $P_i(\theta_j)$  y  $Q_i(\theta_j)$  son funciones únicamente de  $\theta_j$ , ya que hemos supuesto que, para cada ítem  $i$ , los parámetros  $a_i$  y  $b_i$  son conocidos. Por lo tanto, al aplicar el método de máxima verosimilitud, buscamos  $\hat{\theta}_j$  que maximice  $l(\theta_j|\mathbf{y}_j)$ . Esto nos conduce a que se debe de satisfacer las siguientes igualdades<sup>4</sup>:

$$\left. \frac{\partial}{\partial \theta_j} l(\theta_j|\mathbf{y}_j) \right|_{\hat{\theta}_j} = 0 \quad , \quad \left. \frac{\partial^2}{\partial \theta_j^2} l(\theta_j|\mathbf{y}_j) \right|_{\hat{\theta}_j} < 0 \quad (3.19)$$

Las expresiones anteriores equivalen a:

$$\begin{aligned} \left. \frac{\partial}{\partial \theta_j} l(\theta_j|\mathbf{y}_j) \right|_{\hat{\theta}_j} &= \sum_{i=1}^n y_{ij} \frac{1}{P_i(\theta_j)} \frac{\partial P_i(\theta_j)}{\partial \theta_j} \Big|_{\hat{\theta}_j} + \sum_{i=1}^n (1 - y_{ij}) \frac{1}{Q_{ij}} \frac{\partial Q_{ij}}{\partial \theta_j} \Big|_{\hat{\theta}_j} = 0 \\ \left. \frac{\partial^2}{\partial \theta_j^2} l(\theta_j|\mathbf{y}_j) \right|_{\hat{\theta}_j} &= \sum_{i=1}^n y_{ij} \frac{\partial^2 P_{ij}}{\partial \theta_j^2} \Big|_{\hat{\theta}_j} + \sum_{i=1}^n y_{ij} \frac{1}{P_{ij}} \frac{\partial^2 P_{ij}}{\partial \theta_j^2} \Big|_{\hat{\theta}_j} + \\ &\quad \sum_{i=1}^n (1 - y_{ij}) \frac{\partial^2 Q_{ij}}{\partial \theta_j^2} \Big|_{\hat{\theta}_j} + \sum_{i=1}^n (1 - y_{ij}) \frac{1}{Q_{ij}} \frac{\partial^2 Q_{ij}}{\partial \theta_j^2} \Big|_{\hat{\theta}_j} < 0 \end{aligned} \quad (3.20)$$

Derivando Ec. 3.15, Ec. 3.15 respecto a  $\theta_j$  se obtiene:

<sup>4</sup>En la expresión 3.19 la notación  $\frac{\partial}{\partial \theta_j} \Big|_{\hat{\theta}_j}$  es utilizada para indicar que la derivada evaluada en un punto.

$$\frac{\partial P_{ij}}{\partial \theta_j} = a_i P_{ij} Q_{ij} \quad (3.21)$$

$$\frac{\partial Q_{ij}}{\partial \theta_j} = -a_i P_{ij} Q_{ij} \quad (3.22)$$

Sustituyendo las derivadas 3.21 3.22 en 3.19 y definiendo  $W_{ij} = P_{ij} Q_{ij}$  (pesos de Urban-Müller) obtenemos el siguiente sistema de ecuaciones no lineal:

$$\begin{aligned} \frac{\partial}{\partial \theta_j} l(\theta_j | \mathbf{y}_j) &= \sum_{i=1}^n y_{ij} \frac{1}{P_{ij}} \frac{\partial P_{ij}}{\partial \theta_j} + \sum_{i=1}^n (1 - y_{ij}) \frac{1}{Q_{ij}} \frac{\partial Q_{ij}}{\partial \theta_j} \\ &= \sum_{i=1}^n a_i y_{ij} - \sum_{i=1}^n a_i P_{ij} = \sum_{i=1}^n a_i W_{ij} \left( \frac{y_{ij} - P_{ij}}{P_{ij} Q_{ij}} \right) \end{aligned} \quad (3.23)$$

$$\begin{aligned} \frac{\partial}{\partial \theta_j} \left( \frac{\partial}{\partial \theta_j} l(\theta_j | \mathbf{y}_j) \right) &= \frac{\partial^2}{\partial \theta_j^2} l(\theta_j | \mathbf{y}_j) = - \sum_{i=1}^n a_i \frac{\partial P_{ij}}{\partial \theta_j} \\ &= - \sum_{i=1}^n a_i^2 P_{ij} Q_{ij} = - \sum_{i=1}^n a_i^2 W_{ij} \end{aligned} \quad (3.24)$$

Observar que en la Ec. 3.23, el primer sumando ( $\sum_{i=1}^n a_i y_{ij}$ ) corresponde al estadístico mínimo suficiente de  $\theta_j$  en el modelo logístico.

La búsqueda de  $\hat{\theta}_j$  (solución al problema) se realiza, como lo hicimos antes, mediante técnicas de aproximaciones sucesivas. En general suele aplicarse el algoritmo de Newton-Raphson<sup>5</sup>. Por lo tanto, al igual que antes, se asume que la solución es  $\hat{\theta}_j^{(1)} = \hat{\theta}_j^{(0)} + \Delta \hat{\theta}_j^{(0)}$ , siendo  $\Delta \hat{\theta}_j^{(0)}$  el error en la estimación.

La serie de Taylor de primer orden para la función  $\partial l / \partial \theta_j$  en torno a  $\hat{\theta}_j^{(0)}$  la es:

$$\left. \frac{\partial}{\partial \theta_j} l(\theta_j | \mathbf{y}_j) \right|_{\hat{\theta}_j^{(1)}} \cong \left. \frac{\partial}{\partial \theta_j} l(\theta_j | \mathbf{y}_j) \right|_{\hat{\theta}_j^{(0)}} + \left. \frac{\partial^2}{\partial \theta_j^2} l(\theta_j | \mathbf{y}_j) \right|_{\hat{\theta}_j^{(0)}} \Delta \hat{\theta}_j^{(0)} \quad (3.25)$$

Si  $\hat{\theta}_j^{(0)}$  es solución de Ec. 3.19, entonces el lado izquierdo de Ec. 3.25 es igual a cero. Por lo tanto, obtenemos una estimación para  $\Delta \hat{\theta}_j^{(0)}$ :

$$\Delta \hat{\theta}_j^{(0)} = - \left[ \left. \frac{\partial^2}{\partial \theta_j^2} l(\theta_j | \mathbf{y}_j) \right|_{\hat{\theta}_j^{(0)}} \right]^{-1} \cdot \left. \frac{\partial}{\partial \theta_j} l(\theta_j | \mathbf{y}_j) \right|_{\hat{\theta}_j^{(0)}} \quad (3.26)$$

<sup>5</sup>Una descripción a alto nivel de este método puede encontrarse en A.

A partir de la estimación de  $\Delta\hat{\theta}_j^{(0)}$  podemos estimar  $\hat{\theta}_j^{(1)}$  ya que  $\hat{\theta}_j^{(0)}$  es conocida y puede expresarse de la siguiente manera:

$$\begin{aligned}\hat{\theta}_j^{(1)} &= \hat{\theta}_j^{(0)} + \Delta\hat{\theta}_j^{(0)} \\ &= \hat{\theta}_j^{(0)} - \left[ \frac{\partial^2 l(\theta_j|\mathbf{y}_j)}{\partial\theta_j^2} \Big|_{\hat{\theta}_j^{(0)}} \right]^{-1} \cdot \frac{\partial l(\theta_j|\mathbf{y}_j)}{\partial\theta_j} \Big|_{\hat{\theta}_j^{(0)}}\end{aligned}\quad (3.27)$$

El algoritmo se detiene cuando el punto obtenido cumple:  $\Delta\hat{\theta}_j^{(0)} < \epsilon_{\theta_j}$ , con  $\epsilon_{\theta_j}$  arbitrario y suficientemente pequeño. En caso contrario, repetimos el procedimiento anterior. En forma genérica, en el paso  $k + 1$  la solución propuesta tomará la forma:

$$[\hat{\theta}_j]_{k+1} = [\hat{\theta}_j]_k + [\Delta\hat{\theta}_j^{(k)}] \quad (3.28)$$

Mientras que la estimación es:

$$[\hat{\theta}_j]_{k+1} = [\hat{\theta}_j]_k - \left[ \frac{\partial^2 l(\theta_j|\mathbf{y}_j)}{\partial\theta_j^2} \right]_k^{-1} \cdot \left[ \frac{\partial l(\theta_j|\mathbf{y}_j)}{\partial\theta_j} \right]_k \quad (3.29)$$

La Ec. 3.29 puede reescribirse, utilizando las Ecs. 3.23 y 3.24, de la siguiente manera:

$$[\hat{\theta}_j]_{k+1} = [\hat{\theta}_j]_k - \left[ \sum_{i=1}^n a_i W_{ij} \left( \frac{y_{ij} - P_{ij}}{P_{ij} Q_{ij}} \right) \right]_k^{-1} \cdot \left[ -\sum_{i=1}^n a_i^2 W_{ij} \right]_k \quad (3.30)$$

En muestras grandes se puede probar que la varianza de  $\hat{\theta}_j$  es:

$$Var(\hat{\theta}_j) = -\frac{1}{E\left(\frac{\partial^2 l(\theta_j|\mathbf{y}_j)}{\partial\theta_j^2}\right)} \quad (3.31)$$

Utilizando el resultado de la Ec. 3.24 la varianza puede expresarse de la siguiente manera:

$$Var(\hat{\theta}_j) = \frac{1}{\sum_{i=1}^n a_i^2 W_{ij}} \quad (3.32)$$

A pesar de la sencillez que presenta el método, en la práctica suele ser poco usado ya que en general se desconoce el valor de los parámetros de dificultad y discriminación involucrados en el test.

### 3.4. Métodos de Estimación Conjunta, Condicional y Marginal

En las secciones anteriores fueron presentadas las técnicas de estimación condicional, éstas parten de la hipótesis que el rasgo latente o los parámetros del modelo son conocidos. Sin embargo, como hemos mencionado, en la práctica es poco frecuente encontrarnos con experimentos en los que sean conocidos dichos parámetros. Es por ello que se han desarrollado métodos alternativos que permiten realizar una estimación conjunta de ambos grupos de parámetros.

Las técnicas de estimación paramétrica conjunta más utilizadas son: máxima verosimilitud conjunta o Joint Maximum Likelihood (Joint Maximum Likelihood (JML)), máxima verosimilitud condicional o Conditional Maximum Likelihood (Conditional Maximum Likelihood (CML) y máxima verosimilitud marginal o Marginal Maximum Likelihood (Marginal Maximum Likelihood (MML)). Éstas se diferencian entre sí principalmente en lo siguiente, las dos primeras modelan el rasgo latente como un efecto fijo, mientras que en la tercera es considerado como un efecto aleatorio. A continuación estaremos presentando los principales resultados asociados a cada una de ellas.

#### 3.4.1. Máxima Verosimilitud Conjunta

El método de máxima verosimilitud conjunta JML consiste en maximizar la distribución conjunta ( $P(\mathbf{y}|\boldsymbol{\theta}, \mathbf{a}, \mathbf{b})$ ) con respecto al rasgo latente ( $\theta_j$ ) y el resto de los parámetros involucrados en el modelo, es decir, dificultad del ítem ( $b_i$ ) y discriminación del ítem ( $a_j$ ). Para este fin es necesario asumir que el rasgo latente y los parámetros del modelo son efectos fijos.

En el capítulo anterior, bajo la hipótesis de independencia local, obtuvimos la distribución conjunta para cada modelo analizado (1PL, 2PL). Recordemos que fueron considerados  $n$  ítems en un test, siendo  $J$  la cantidad de examinados. Además,  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_J)$  corresponde al rasgo latente de los participantes e  $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_J)$  los vectores columna correspondientes a sus respuestas en el test. Por último,  $\mathbf{b} = (b_1, b_2, \dots, b_n)$  y  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  corresponden a los vectores de parámetros de dificultad, discriminación respectivamente para cada ítem.

En su forma general, la probabilidad de observar una matriz de respuestas como la presentada en la Tabla 2.1, es decir, para los modelos 1PL, 2PL, 3PL, esta dada por:

$$P(\mathbf{Y} = \mathbf{y}|\boldsymbol{\theta}, \mathbf{d}) = L(\boldsymbol{\theta}, \mathbf{d}|\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_J) = \prod_{i=1}^n \prod_{j=1}^J [\Psi(\theta_j, \mathbf{d}_i)]^{y_{ij}} \cdot [\Psi(\theta_j, \mathbf{d}_i)]^{1-y_{ij}} \quad (3.33)$$

Donde  $\mathbf{d}$  es el vector de parámetros que involucra a los vectores  $\mathbf{a}$  y  $\mathbf{b}$ , según el modelo que se esté considerando<sup>6</sup> y  $\Psi(\cdot)$  es la cdp logística<sup>7</sup> utilizada para modelar la CCI. A partir de la expresión anterior tenemos que el logaritmo de la función de verosimilitud es:

$$\ln \left( L \left( (\boldsymbol{\theta}, \mathbf{d}|\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_J) \right) \right) = \sum_{i=1}^n \sum_{j=1}^J (y_{ij}) \cdot P_i(\theta_j, d_i) + (1 - y_{ij}) \cdot Q_i(\theta_j, d_i) \quad (3.34)$$

<sup>6</sup>Por ejemplo, en un modelo 2PL  $d_i = (a_i, b_i)$ .

<sup>7</sup>Se ha definido en la Ec. 2.22.

Luego, sólo resta obtener las derivadas parciales respecto a cada parámetro e igualar a cero. En esta dirección el proceso es análogo al presentado en las secciones 3.2 y 3.3, obteniéndose un sistema de ecuaciones que debe resolverse en forma simultánea. Una alternativa en la búsqueda de las estimaciones máximo verosímiles, es utilizar el método de Newton-Raphson en dos etapas. La primera de éstas consiste en estimar los parámetros del modelo, asumiendo que los ítems son independientes y la habilidad de los participantes es conocida, es decir, partiendo de un valor inicial predefinido. Luego, en la segunda etapa, se estima la habilidad de los examinados, utilizando el resultado de la primer etapa, asumiendo que éstos responden en forma independiente. El proceso continúa hasta que la diferencia entre pasos sucesivos es menor a un valor pequeño, que ha sido inicialmente fijado.

Bajo este enfoque, las ventajas y desventajas señaladas por Baker y Kim (Baker and Kim, 2004, Chap. 4) pueden resumirse en:

- El procedimiento es general y puede ser aplicado a cualquiera de los modelos, es decir, de uno, dos, o tres parámetros.
- El método no realiza una estimación simultánea de los parámetros, sino que son necesarias dos etapas, donde en cada una de ellas se debe de asumir ciertas condiciones de independencia para poder obtener la estimación conjunta.
- Existe evidencia acerca de la consistencia en las estimaciones de los parámetros, a medida que crece la cantidad de examinados, para un número fijo de ítems. Sin embargo, dicha propiedad aún no ha sido probado en modelos 2PL y 3PL.

Además, cuando se realizan la estimación en el modelo 2PL, para ciertos conjuntos de datos, el método JML puede verse afectado por problemas de estabilidad, y en consecuencia las soluciones divergen. A dicho problema en la literatura se lo conoce como el caso Heywood (Bartolucci et al., 2015, Chap. 5).

Por último, hacer notar que al problema de optimización se le debe imponer restricciones para que la estimación de los parámetros se encuentre dentro de rango esperado. Además, otras consideraciones son necesarias cuando el conjunto de datos incluye respuestas faltantes.

### 3.4.2. Máxima Verosimilitud Condicional

En la sección anterior hicimos notar que una de las desventajas presente en el método JML, es la ausencia de una demostración formal sobre la consistencia de los estimadores. Distintos métodos se han propuesto con el fin de sortear esta debilidad, uno de ellos es el denominado método de máxima verosimilitud condicional o CML.

El método CML saca provecho de las bondades que presenta la familia de distribuciones exponencial, particularmente el modelo Rasch, siendo posible obtener estimadores consistentes. Para ello, se parte del hecho que el score obtenido por parte del  $j$ -ésimo examinado ( $y_{.j} = \sum_{i=1}^n y_{ij}$ ) es un

estadístico suficiente para estimar  $\theta_j$ , es decir, la habilidad que éste presenta<sup>8</sup>. Luego, maximizando la función de verosimilitud, condicionada respecto de  $\theta_j$ , es posible estimar el parámetro  $b_i$  correspondiente a la dificultad del  $i$ -ésimo ítem.

Aplicando este enfoque al modelo de Rasch, tenemos que, si al igual que antes, se tiene un test con  $n$  ítems y son  $J$  los examinados que participan en el test, entonces, para el  $j$ -ésimo examinado, la probabilidad de observar el vector de respuestas  $\mathbf{y}'_j = (y_{1j}, y_{2j}, \dots, y_{1n})$ , condicionada a un nivel de habilidad  $\theta_j$ , esta dada por la Ec. 2.31, es decir:

$$P(\mathbf{Y}_j = \mathbf{y}_j | \theta_j, \mathbf{b}) = L(\theta_j, \mathbf{b} | \mathbf{y}_j) = \frac{e^{\theta_j y_{.j} - \sum_{i=1}^n b_i y_{ij}}}{\prod_{i=1}^n (1 + e^{(\theta_j - b_i)})} \quad (3.35)$$

Donde  $\mathbf{b} = (b_1, b_2, \dots, b_n)$  corresponde al vector de parámetros de dificultad de cada ítem. Usando este resultado podemos obtener la probabilidad de observar un vector de respuestas  $\mathbf{y}_j$ , dado que  $r_j = t(\mathbf{y}_j) = y_{.j}$  es el resultado del test y  $\theta_j$  es el nivel de habilidad. Dicha probabilidad esta dada por:

$$P(\mathbf{Y}_j = \mathbf{y}_j | \mathbf{y}_{.j}, \theta_j, \mathbf{b}) = \frac{P(\mathbf{Y}_j = \mathbf{y}_j, \mathbf{y}_{.j} | \theta_j, \mathbf{b})}{P(\mathbf{y}_{.j} | \theta_j, \mathbf{b})} = \frac{P(\mathbf{Y}_j = \mathbf{y}_j | \theta_j, \mathbf{b})}{P(\mathbf{y}_{.j} | \theta_j, \mathbf{b})} \quad (3.36)$$

El denominador de la expresión anterior es la probabilidad que el  $j$ -ésimo participante obtenga un score igual a  $r_j = t(\mathbf{y}_j) = y_{.j}$ , condicionada a  $\theta_j$ , es decir, a un nivel de habilidad dado. En un test con  $n$  ítems, existen  $\binom{n}{y_{.j}}$  posibles vectores de respuesta que el participante  $j$  puede proporcionar para obtener este score. Por lo tanto, dicha probabilidad es igual a la suma de las probabilidad para cada posible vector de respuestas<sup>9</sup>.

$$\begin{aligned} P(\mathbf{y}_{.j} | \theta_j, \mathbf{b}) &= \sum_{\forall i / t(\mathbf{y}_j) = y_{.j}} P(Y_{ij} = 1 | \theta_j, \mathbf{b}) \\ &= \sum_{\forall i / t(\mathbf{y}_j) = y_{.j}} \prod_{i=1}^n (P(Y_{ij} = y_{ij}))^{y_{ij}} \times (1 - P(Y_{ij} = y_{ij}))^{1 - y_{ij}} \\ &= \frac{e^{\theta_j y_{.j}} \left( \sum_{\forall i / t(\mathbf{y}_j) = y_{.j}} e^{-\sum_{i=1}^n b_i y_{ij}} \right)}{\prod_{i=1}^n (1 + e^{(\theta_j - b_i)})} \end{aligned} \quad (3.37)$$

A partir de las Ec. 3.36 y 3.37 se obtiene:

<sup>8</sup>Recordemos que estamos utilizando como referencia la Tabla 2.1, donde  $y_{.j} = \sum_{i=1}^n y_{ij}$  corresponde al score del  $j$ -ésimo participante,  $y_{i.} = \sum_{j=1}^J y_{ij}$  es la cantidad de respuestas correctas al  $i$ -ésimo ítem.

<sup>9</sup>Observar que en la Ec. 3.37 el índice  $i$  de la productoria en el denominador no se ve restringido a los ítems que satisfacen la condición de score total. Esto se debe a que una vez encontrado el vector de respuestas que cumple con dicha condición, el producto se realiza sobre todos los ítems.

$$P(\mathbf{Y}_j = \mathbf{y}_j | \theta_j, \mathbf{b}) = \frac{e^{-\sum_{i=1}^n b_i y_{ij}}}{\sum_{\forall i / t(\mathbf{y}_j) = \mathbf{y}_j} e^{-\sum_{i=1}^n b_i y_{ij}}} \quad (3.38)$$

La expresión anterior es independiente del  $\theta_j$ , es decir, el rasgo latente del examinado y sólo depende de los parámetros de dificultad ( $b_i$ ) de cada ítem y se restringe al score total ( $\mathbf{y}_j$ ) de interés a ser analizado. Luego, la función de verosimilitud condicional se obtiene al considerar a los  $J$  examinados, donde cada uno de ellos ha obtenido un score  $\mathbf{y}_j$ . Esto es:

$$P(\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_J) | \boldsymbol{\theta}, \mathbf{b}) = L(\boldsymbol{\theta}, \mathbf{b} | \mathbf{y}) = \frac{e^{-\sum_{j=1}^J \sum_{i=1}^n b_i y_{ij}}}{\prod_{j=1}^J \sum_{\forall i / t(\mathbf{y}_j) = \mathbf{y}_j} e^{-\sum_{i=1}^n b_i y_{ij}}} \quad (3.39)$$

Siendo  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_J)$  las habilidades de los examinados, y la matriz de respuestas corresponde a  $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_J)$ . Observar que el denominador en la expresión anterior puede compactarse<sup>10</sup>, esto se logra al tomar en cuenta aquellos examinados que han obtenido idéntico score.

La estimación máximo verosímil del vector  $\mathbf{b}$  puede obtenerse al tomar logaritmo en la Ec. 3.39, derivando respecto a cada parámetro e igualando a cero. De esta forma se obtiene un sistema de ecuaciones cuya solución puede encontrarse, por ejemplo, mediante algoritmos iterativos como Newton-Raphson. En otras palabras, el procedimiento a seguir es análogo al antes presentado en este capítulo. Un desarrollo completo es presentado por Bartolucci (Bartolucci et al., 2015, ver Sec. 5.3 y App. 5.A.2.).

En forma alternativamente, si el objetivo fuese estimar la habilidad del participante, también se podría utilizar el método CML. En este caso, se hace uso de la propiedad presente en el modelo de Rasch, es decir, la suma del score obtenido en el  $i$ -ésimo ítem por parte de todos los examinados ( $y_{i\cdot} = \sum_{j=1}^J y_{ij}$ ) es un estadístico suficiente para estimar el parámetro  $b_i$  correspondiente a la dificultad del ítem. A continuación, maximizando la función de verosimilitud, condicionada respecto de  $b_i$ , es posible estimar el parámetro  $\theta_j$ , es decir, el rasgo latente del  $j$ -ésimo examinado.

A pesar de ser una alternativa que permite obtener estimaciones consistentes cuando trabajamos con el modelo de Rasch, y de ser uno de los métodos más utilizados, respecto a Linear Logistic Test Model (LLTM), el CML, presenta algunas debilidades. Fischer nos advierte acerca dos problemas que éste presenta, el primero de ellos asociado al cómputo de las derivadas parciales y la precisión numérica, mientras que la segunda esta asociada a la velocidad de convergencia (Fischer and Molenaar, 1995, Chap. 3). Además, estos problemas se vuelven aún más críticos a medida que se incrementa el número de ítems en el test.

Antes de finalizar, hacer mención a las propiedades asintóticas del método CML. Andersen (1970) es quien propone dicho método y prueba la consistencia y normalidad de las estimaciones. Sus resultados se construyen sobre otros que fueron probados para familia de distribuciones exponencial. También Pfanzagl (1994) prueba la consistencia y además eficiencia asintótica de los estimadores.

<sup>10</sup>En la mayoría de los casos de interés  $J > n$  y cada ítem es dicotómico, entonces, en la matriz de respuestas se van a observar score idénticos entre participantes.

### 3.4.3. Máxima Verosimilitud Marginal

En la sección anterior presentamos un método alternativo, máxima verosimilitud condicional, que permite estimar los parámetros del modelo logístico, y además, resuelve el problema de consistencia presente en el método de máxima verosimilitud conjunta. Sin embargo, su aplicación se encuentra restringida al modelo de Rasch. A continuación estaremos analizando un método más general, llamado máxima verosimilitud marginal o MML, mediante el cual es posible obtener estimadores consistentes para los modelos 2PL y 3PL.

El método MML consiste en maximizar la función de verosimilitud una vez que los parámetros utilizados para modelar la habilidad se han integrado, asumiendo que éstos presentan en común una distribución de probabilidad. Dicho método nace con la propuesta de Bock y Liberbam (1970), quienes trabajando sobre el modelo ojiva normal de dos parámetros, estiman cotas para los parámetros de dificultad de los ítems haciendo uso de la técnica Gauss-Hermite Quadrature<sup>11</sup> en la integración y asumiendo que la habilidad de los examinados se distribuye  $N(0, 1)$ . Además, Bock y Liberbam sugieren utilizar el algoritmo de Newton-Raphson para obtener la solución al problema de maximización.

Bajo el enfoque de MML, los examinados son considerados como una muestra aleatoria tomada de una población cuya habilidad se distribuye según cierta  $p_h(\boldsymbol{\theta}|\boldsymbol{\tau})$ . Es decir, el vector de habilidades ( $\boldsymbol{\theta}$ ) es considerado aleatorio y  $\boldsymbol{\tau}$  es el vector de parámetros de habilidad poblacional. Por lo tanto, bajo este escenario el modelo va a incluir dos componentes, efectos fijos (dificultad de los ítems) y efectos aleatorios (habilidad de los participantes). De esta forma, al integrar la función de verosimilitud (que se obtiene a partir de la matriz de respuestas) respecto a  $\boldsymbol{\theta}$ , el problema consiste en maximizar (en los parámetros de los ítems ( $\mathbf{a}, \mathbf{b}$ )) el resultado de la integración, el cual es independiente de la habilidad de los examinados.

Observar que en JML y CML el vector de habilidades es considerado fijo, mientras que en MML la habilidad es un vector aleatorio. Por lo tanto, el rasgo latente del examinado ( $\hat{\theta}_j$ ) debe de ser reemplazado por la probabilidad de tener cierto nivel de habilidad condicionada al vector  $\mathbf{y}_j$  de respuestas, los parámetros correspondientes a los ítems ( $\mathbf{a}, \mathbf{b}$ ) y el parámetro  $\boldsymbol{\tau}$  de habilidad poblacional.

Teniendo esto presente, desde la perspectiva de la estadística bayesiana, dicha probabilidad es la denominada probabilidad a posteriori, mientras que  $p_h(\boldsymbol{\theta}|\boldsymbol{\tau})$  es la denominada distribución a priori. En consecuencia, tendremos un vector (de dimensión  $J$ ) de probabilidades a posteriori, donde cada componente corresponde a cada uno de los examinados. En forma analítica, podemos expresar esto de la siguiente manera<sup>12</sup>:

$$p_h(\theta_j|\mathbf{y}_j, \mathbf{a}, \mathbf{b}, \boldsymbol{\tau}) = \frac{p(\mathbf{y}_j|\theta_j, \mathbf{a}, \mathbf{b}) \cdot p_h(\theta_j|\boldsymbol{\tau})}{\int p(\mathbf{y}_j|\theta_i, \mathbf{a}, \mathbf{b}) \cdot p_h(\theta_i|\boldsymbol{\tau}) d\theta_i} \quad (3.40)$$

Arribados a este punto, debemos señalar que la distribución a priori del rasgo latente, en principio, es desconocida. Por lo tanto, su elección va a depender del conocimiento previo que se tenga sobre la

<sup>11</sup>Una descripción detallada sobre Gauss-Hermite Quadrature y sus aplicaciones en estadística puede encontrarse en Liu and Pierce (1994); Nia (2006).

<sup>12</sup>Hermos utilizado  $p$  para representar a las respectivas funciones de densidad (o cuantía).

población y las características del test. En ocasiones se utiliza una muestra pequeña de examinados, y mediante un test piloto se intenta inferir su forma. Además, ésta puede provenir una distribución continua o discreta<sup>13</sup>.

Bajo la hipótesis de independencia local  $P(\mathbf{y}_j|\theta_j, (\mathbf{a}, \mathbf{b}))$  es igual al producto, para los  $n$  ítems, de  $P_i(\theta_j)$  y  $Q_i(\theta_j)$  elevadas a  $y_{ij}$  e  $(1 - y_{ij})$  respectivamente, tal como hemos visto antes para los modelos 2PL y 3PL. Luego, una vez obtenida la distribución a posteriori, se sigue el procedimiento antes descrito<sup>14</sup>.

La propuesta inicial de Bock y Liberbam presentaba ciertas debilidades. Éstas estaban asociadas a los tiempos de cómputo: a) en el proceso de integración y b) el proceso de maximización. Sumado a esto, al utilizar Gauss-Hermite Quadrature, la precisión en las estimaciones depende de la cantidad de puntos que se utiliza, y resulta ser adecuada cuando la dimensión del rasgo latente es baja. Con el paso del tiempo, en la década del '80, Bock y Aitkin (Bock and Aitkin, 1981) proponen utilizar el algoritmo de Expectation–Maximization (EM) en la estimación de los parámetros de dificultad.

El nuevo enfoque de Bock y Aitkin presenta mejoras considerables respecto a los problemas identificados en la propuesta inicial. La estimación de los parámetros del ítem se realiza en dos pasos, denominados  $E$  (Expectation) y  $M$  (Maximization), aplicados a el logaritmo de la función de verosimilitud, una vez que se ha realizado la integración en el parámetro de habilidad. En el paso  $E$  se calcula el valor esperado de la verosimilitud (para cada examinado), condicionada a los datos observados y el valor de los parámetros correspondientes a los examinados. Mientras que en el paso  $M$  se maximiza dicho valor esperado con respecto a los parámetros del ítem.

Antes de finalizar, hacer mención que en los últimos años se han desarrollado distintas técnicas que permiten mejorar la convergencia y/o precisión, respecto a la propuesta Bock y Aitkin, en las estimaciones. Algunos ejemplos son: Schilling and Bock (2005) quienes muestran una mejora en la precisión y velocidad de convergencia mediante el uso de Adaptive Gauss-Hermite Quadrature; Casabianca and Lewis (2015) logran mejorar la precisión en las estimaciones, para ello utilizan loglinear smoothing en la estimación de la distribución del rasgo latente; F. and Karimi (2019) presentan un nuevo método que permite aproximar la función de verosimilitud para el caso de datos longitudinales.

### 3.5. Ajuste del Modelo de Rasch

La bondad en el ajuste del modelo es un concepto que, en el área de estadística, en ocasiones se utiliza para describir qué tan bien el modelo representa a los datos observados. Un modelo perfectamente ajustado representará muy bien a las muestras disponibles, sin embargo, probablemente resulte pobre su performance a la hora de predecir nuevas observaciones. En contrapartida, un modelo que no presente ajuste perfecto, ofrecerá mejores resultados a la hora de predecir.

En general, existen dos grupos de herramientas que permiten analizar el ajuste del modelo: gráficas y estadísticas (índices o test de hipótesis). La primera de éstas podrían considerarse menos objetivas que las segundas. A pesar de ello, debido a su sencillez en la visualización e interpretación de los

---

<sup>13</sup>Una completa descripción sobre las posibles distribuciones a priori puede encontrarse en Fox (2010).

<sup>14</sup>Un desarrollo completo puede encontrarse en Baker and Kim (2004).

resultados, cada vez más se incrementa la investigación en el campo de estadística computacional gráfica (Chambers, Cleveland, B., and Tukey, 2017).

En la presente sección, tal como adelantáramos al inicio del capítulo, estaremos analizando tres de las herramientas estadísticas frecuentemente utilizadas en la TRI. Estas son el Test  $\chi^2$  y el test de razón de verosimilitudes y el pseudo  $R^2$ .

### 3.5.1. Test $\chi^2$ de Pearson

El test  $\chi^2$  de Pearson es una de las técnicas más populares, frecuentemente utilizado en aquellos modelos estadísticos donde la variable respuesta es categórica. Tomando como referencia la matriz de respuestas presentada en la Tabla. 2.1, cada vector de respuestas  $\mathbf{y}_j$  presenta  $2^n$  posibles combinaciones. Por lo tanto, los modelos de TRI logísticos pueden pensarse como un modelo lineal generalizado, en particular, uno multinomial, donde existen  $C = 2^n$  categorías mutuamente excluyentes. En consecuencia, es posible utilizar los resultados existentes para realizar el test de hipótesis.

Bajo esta perspectiva, tendremos  $C$  probabilidades, donde cada una de ellas representa la probabilidad de pertenecer a una categoría, es decir,  $\hat{\mathbf{P}}(\mathbf{v}) = (\hat{P}_1(\mathbf{v}), \hat{P}_2(\mathbf{v}), \dots, \hat{P}_C(\mathbf{v}))$ , siendo  $\mathbf{v}$  el vector de parámetros cuya dimensión depende de  $\boldsymbol{\theta}$  y los parámetros del modelo involucrado (1PL o 2PL). Luego, el estadístico del test  $\chi^2$  (Ec. 3.41) se construye haciendo uso de dichas probabilidades y las frecuencias observadas en la matriz de respuesta.

$$X^2 = n \sum_{l=1}^C \frac{(\mathcal{F}_l - E(\mathcal{F}_l))^2}{E(\mathcal{F}_l)} = n \sum_{l=1}^C \frac{(\mathcal{F}_l - \hat{P}_l)^2}{\hat{P}_l} \quad (3.41)$$

En la expresión anterior hemos utilizado  $\mathcal{F}_l$  para representar la frecuencia absoluta observada, mientras que  $E(\mathcal{F}_l) = \hat{P}_l$  corresponde a la frecuencia esperada. La hipótesis nula sometida a test es: los datos se ajustan al modelo, es decir, el modelo es adecuado. Bajo  $H_0$ , asintóticamente  $X^2 \sim \chi^2$  con  $C - \dim(\mathbf{v}) - 1$  grados de libertad (Glas and Verhelst, 1989).

A pesar de su sencillez, el test no resulta de utilidad a medida que la cantidad de ítems aumenta. El problema radica en que el número de sumandos (patrones de respuestas posible) crece en forma exponencial, en consecuencia, observaremos muchos patrones de respuesta con frecuencia igual a cero, mientras que su valor esperado será muy pequeño (Rao and Sinharay, 2007, Chap. 6). Glas y Verhelst (1989) proponen un nuevo estadístico que permite sortear este problema. Éste se construye a partir de una extensión de la Ec. 3.41, la cual puede ser expresada de la siguiente manera:

$$X^2 = \mathbf{v}'\mathbf{D}^{-1}\mathbf{v} \quad (3.42)$$

Siendo  $\mathbf{v} = \sqrt{n}(\mathcal{F} - \mathbf{E}(\mathcal{F}))$  el vector de desviaciones, donde los vectores  $\mathcal{F}$  y  $\mathbf{E}(\mathcal{F})$  corresponden a las frecuencias observadas y esperadas respectivamente. Además,  $\mathbf{D}$  es una matriz diagonal formada por  $E(\mathcal{F}_l) = \hat{P}_l$ . El estadístico del test se basa en el vector  $\mathbf{d} = \mathbf{U}'\mathbf{v}$  de  $\tilde{c}$  combinaciones lineales, donde  $\mathbf{U}$  es la denominada matriz de contraste, que concentra las  $C$  desviaciones en una cantidad  $\tilde{c}$  ( $C \gg \tilde{c}$ ). Luego, definiendo<sup>15</sup>:

<sup>15</sup>Se ha utilizado la notación  $A^-$  para indicar a la matriz inversa generalizada.

$$Q = \mathbf{v}'\mathbf{U}(\mathbf{U}'\mathbf{D}^{-1}\mathbf{U})^{-1}\mathbf{U}'\mathbf{v} \quad (3.43)$$

Bajo  $H_0$ , asintóticamente  $Q \sim \chi^2$  y los grados de libertad son  $\text{rango}(\mathbf{U}'\mathbf{D}\mathbf{U}) - \dim(\mathbf{v}) - 1$ . Un aspecto crucial bajo este enfoque es cómo obtener la matriz  $\mathbf{U}$ , en este sentido un procedimiento detallado es desarrollado por Fischer and Molenaar (1995) y Glas and Verhelst (1989).

### 3.5.2. Test de Razón de Verosimilitudes

El test de razón de verosimilitud es utilizando tanto en modelos lineales generalizados, así como también en los modelos logísticos de la TRI (Andersen, 1973b). Dicho test permite comparar dos modelos anidados (completo y reducido), es decir, aquellos donde uno incluye un subconjunto de parámetros del otro.

En modelos lineales o logísticos, la hipótesis nula ( $H_0$ ) corresponde a evaluar si son igual a cero aquellos parámetros del modelo completo que no forman parte del modelo reducido. En el contexto del modelo de Rasch, se evalúa  $\theta_j = 1$ ,  $j = 1, 2, \dots, J$ , es decir, la cantidad de rasgos latentes que son igual a uno.

El estadístico utilizado en el test de razón de verosimilitud es:

$$D = 2 \log \frac{\hat{L}_C}{\hat{L}_R} = -2(\hat{l}_R - \hat{l}_C) \quad (3.44)$$

Donde  $\hat{L}_C$  y  $\hat{L}_R$  son las funciones de verosimilitud, utilizando los parámetros estimados, para el modelo completo y el modelo reducido<sup>16</sup>. Bajo la hipótesis nula, haciendo uso de resultados asintóticos, se tiene que  $D \sim \chi^2$ , siendo los grados de libertad igual a la cantidad de parámetros en que difieren uno y otro modelo, es decir, la cantidad de parámetros involucrados en la hipótesis nula.

### 3.5.3. Índices

Otra de las herramientas estadísticas para evaluar la bondad en el ajuste son los denominados índices. Existen distintas propuestas, cada una de ellas con sus características, ventajas y desventajas. En el área de la teoría de los test, su interpretación es análoga a la que se hace, por ejemplo, en los modelos de regresión. Es decir, se utilizan a la hora de evaluar la performance del modelo y son una herramienta valiosa a la hora de elegir entre un conjunto de modelos candidatos.

Uno de los más conocidos es el pseudo  $R^2$ , que se define como el cociente entre la diferencia de la verosimilitud entre los modelos a comparar y la verosimilitud del modelo de referencia (Bartolucci et al., 2015, Chap. 5), esto es:

$$\text{pseudo } R^2 = \frac{\hat{l}_R - \hat{l}_M}{\hat{l}_R} \quad (3.45)$$

---

<sup>16</sup>Hemos utilizado  $\hat{l}_C$  y  $\hat{l}_R$  para representar el logaritmo de las funciones de verosimilitud.

Donde  $\hat{l}_R$  y  $\hat{l}_M$  es el logaritmo de las verosimilitudes estimadas del modelo de referencia y el modelo candidato respectivamente. Éstas pueden obtenerse a través de algunos de los métodos presentados en la sección anterior.

Observar que el valor del pseudo  $R^2 \in [0, 1]$ , de modo que si el resultado es igual a cero, el modelo reducido se ajusta a los datos igual que el de referencia, por lo tanto, elegiríamos al modelo de referencia; mientras que si el resultado se aproxima a uno, entonces el modelo candidato se ajusta mejor a los datos.

Alternativas al pseudo  $R^2$  pueden ser el Akaike Information Criterion o Bayesian Information Criterion, que también pueden utilizarse a la hora de evaluar los modelos logísticos. Éstos, serán analizados con más detalle en el siguiente capítulo, sin embargo, no debemos de perder de vista que, cuando se haga referencia a la función de verosimilitud, los parámetros estimados se obtienen a través de los métodos JML, CML o MML.

### 3.6. Observaciones Finales

En el presente capítulo se han analizado tres enfoques en lo que refiere a la estimación de parámetros en los modelos logísticos de la TRI. En el primero de ellos, maximizando la función de verosimilitud, se procura la estimación de los parámetros: dificultad del ítem, discriminación y guessing, bajo la hipótesis que el rasgo latente de los individuos es conocido. En el segundo enfoque, nuevamente haciendo uso de la función de verosimilitud, se estima el rasgo latente de los participantes bajo la hipótesis que los parámetros del modelo, para cada ítem, son conocidos. Por último, el tercer enfoque consistió en presentar tres métodos (JML, CML, MML), los cuales permiten realizar la estimación de los parámetros y el rasgo latente en forma conjunta. Ha quedado fuera de nuestro alcance, entre otras cosas, analizar el caso en que existen datos faltantes<sup>17</sup>, el impacto en la convergencia o en las estimaciones cuando: aumenta la cantidad de ítems o en aquellos casos en que el test presenta una cantidad de ítems fija y aumenta la número de participantes.

Además, se ha presentado un grupo de herramientas estadísticas que permiten evaluar la bondad del modelo en el ajuste a los datos. Dos de éstas son test de hipótesis (test  $\chi^2$  y test de razón de verosimilitud) y la tercera el índice denominado pseudo  $R^2$ . No han sido abordados otro tipo de test, por ejemplo, los no paramétricos, así como también el análisis de la convergencia asintótica de los estadísticos. Todos estos son tópicos que, a nuestro juicio, demandarían al menos un capítulo completo.

---

<sup>17</sup>Corresponde al caso en que los participantes no responden a todos los ítems del test.

## Capítulo 4

# Modelos Lineales Generalizados Mixtos y Modelo de Rasch

## 4.1. Introducción

Hasta ahora analizamos un conjunto de modelos de respuesta al ítem y las técnicas empleadas para estimar sus parámetros bajo la óptica de la TRI. Sin embargo, no hemos considerado la estructura que presentan los datos provenientes de las respuestas a un test y el efecto que ésta puede tener en las estimaciones. En el presente capítulo estaremos abordando una perspectiva alternativa, que nos permitirá construir modelos cuando los datos presentan cierta estructura jerárquica. En particular, estaremos analizando el vínculo que existe entre los modelos logísticos de la TRI y los denominados GLMM.

Los GLMM ofrecen un mayor grado de flexibilidad ya que, a través de la incorporación de los denominados efectos fijos y aleatorios, es posible modelar de mejor forma la variabilidad entre ítems e individuos que forman parte de un mismo grupo, o entre grupos de individuos. Sin embargo, dicha flexibilidad no es gratuita, en contrapartida la inferencia requiere de algoritmos alternativos y suele ser más costosa desde el punto de vista computacional. En ocasiones cuando el problema no puede ser resuelto a través de algoritmos clásicos de optimización, es la inferencia bayesiana, y los algoritmos de Markov chain Monte Carlo (MCMC), los que ofrecen una solución alternativa.

Una vez que se ha elegido un conjunto de modelos candidatos dentro del abanico de posibilidades que ofrece GLMM, necesitamos disponer de criterios que permitan evaluar la bondad en el ajuste a los datos. En este sentido, algunos métodos provienen de la teoría de modelos lineales. A pesar de ello, se han desarrollado nuevas técnicas para los casos en que la inferencia es realizada bajo la perspectiva bayesiana.

A continuación estaremos abordando los modelos logit de la TRI desde la perspectiva de los GLMM, es decir, consideraremos la estructura de los datos en el modelado. También analizaremos las técnicas de inferencia y selección de modelos. El desarrollo que subsigue esta basado principalmente en los trabajos de McCulloch (2003), Rijmen, Tuerlinckx, De Boeck, and Kuppens (2003), De Boeck and Wilson (2004), Fox (2010), Levy and Mislevy (2016).

## 4.2. Modelo Lineales Generalizados Mixtos

Una familia de modelos que presenta gran flexibilidad, debido a sus características, es la denominada GLMM (Breslow and Clayton, 1993; McCullagh and Nelder, 1989; McCulloch and Searle, 2001). Éstos son una extensión<sup>1</sup> de los LMM los cuales permiten incorporar como predictores lineales efectos fijos y aleatorios. En particular, es la componente aleatoria quien posibilita representar la correlación entre observaciones, asumiendo que ésta proviene de cierta distribución de probabilidad.

Los modelos que involucran efectos aleatorios son frecuentemente utilizados en problemas donde, por ejemplo, el interés principal radica en cuantificar la variabilidad, en lugar de predecir nuevas observaciones (Mecklin, 2022, Chap. 25). También es posible encontrarlos en problemas donde los efectos aleatorios son parte de los efectos no deseados.

---

<sup>1</sup>En el Apéndice E puede encontrarse una introducción a los LMM.

La incorporación de una componente aleatoria en el modelo permite trabajar con datos que presentan cierta correlación. Por ejemplo, hace posible modelar el denominado *subject effects*, el cual está presente en aquellos problemas donde la variable respuesta puede ser explicada por la diferencia intrínseca entre individuos. Este es un efecto típico en datos que presentan cierta estructura de cluster o jerárquica. Otras aplicaciones donde se utilizan los GLMM es en experimentos longitudinales, es decir, cuando los datos provienen de una serie de muestras a las mismas unidades experimentales en un intervalo de tiempo.

Los GLMM se caracterizan, al igual que los Generalized Linear Models (GLM), por tener tres componentes, a saber: la distribución de la variable independiente, efectos fijos, efectos aleatorios y función de enlace. Comenzaremos por introducir la notación que utilizaremos para dichos componentes<sup>2</sup>:

- Distribución de la variable dependiente  $Y_i$ : asumiremos que pertenece a la familia de distribuciones exponenciales  $\forall i = 1, 2, \dots, n$ .
- Efectos fijos: serán modelados a través de una combinación lineal de predictores.

$$\beta_0 + \sum_{l_f=1}^{L_F} \beta_{l_f} \cdot x_{l_f}$$

- Efectos aleatorios: serán modelados a través de una combinación lineal de predictores.

$$u_0 + \sum_{l_r=1}^{L_R} u_{l_r} \cdot z_{l_r}$$

- Una función link  $g$  que enlaza a las variables a los efectos fijos y aleatorios con la variable dependiente.

Estamos asumiendo que  $L_F$  y  $L_R$  corresponde a la cantidad de variables que representan los efectos fijos y aleatorios respectivamente,  $\beta_{l_f}$  son los parámetros de los efectos fijos  $x_{il_f}$ , mientras que  $u_{l_r}$  son los parámetros de los efectos aleatorios y  $z_{il_R}$  son los niveles correspondientes a los efectos aleatorios.

En su forma general de los GLMM se representan de la siguiente manera:

$$\eta(\mathbf{x}_i, \mathbf{z}_i) = \beta_0 + \sum_{l_f=1}^{L_F} \beta_{l_f} \cdot x_{il_f} + u_0 + \sum_{l_r=1}^{L_R} u_{l_r} \cdot z_{il_r} \quad (4.1)$$

---

<sup>2</sup>Únicamente este caso el subíndice  $i$  es utilizado para indexar a las observaciones. En el resto del TFG se utiliza con otro fin y se hace explícito cuando es utilizado para indexar a los ítems.

En forma alternativa tenemos que la representación matricial es:

$$\eta(\mathbf{x}_i, \mathbf{z}_i) = \mathbf{x}_i' \cdot \boldsymbol{\beta} + \mathbf{z}_i' \cdot \mathbf{u} \quad (4.2)$$

Donde  $\mathbf{x}_i \in \mathbb{R}^{L_R}$  y  $\mathbf{z}_i \in \mathbb{R}^{L_R}$ ,  $\boldsymbol{\beta} \in \mathbb{R}^{L_F}$ ,  $\mathbf{u} \in \mathbb{R}^{L_R}$ , además,  $\eta(\cdot)$  es la función que contiene a los predictores lineales, la cual esta formada por dos componentes, es decir, efectos fijos y aleatorios. Finalmente, el vínculo entre la variable respuesta y la función  $\eta(\cdot)$  esta dado por:

$$\eta(\mathbf{x}_i, \mathbf{z}_i) = g(E(y_i)) \quad (4.3)$$

McCulloch (2003) propone una forma compacta para representar a los GLMM:

$$\begin{aligned} Y_i | \mathbf{u} &\sim f_{Y_i | \mathbf{u}}(y | \mathbf{u}) \in \mathcal{F}, \text{ indep.} \\ E(Y_i | \mathbf{u}) &= \mu_i \\ g(\mu_i) &= \eta(\mathbf{x}_i, \mathbf{z}_i) \\ \eta(\mathbf{x}_i, \mathbf{z}_i) &= \mathbf{x}_i' \cdot \boldsymbol{\beta} + \mathbf{z}_i' \cdot \mathbf{u} \\ \mathbf{u} &= f_U(\mathbf{u} | \boldsymbol{\Sigma}) \end{aligned} \quad (4.4)$$

Siendo  $\mathcal{F}$  la familia de distribuciones exponencial. La formulación anterior permite visualizar con mayor claridad cuatro aspectos importantes en este tipo de modelos. 1)  $Y_i$  pertenece a la familia exponencial y su distribución esta condicionada a los efectos aleatorios; 2) la función link ( $g$ ) se aplica al valor esperado de  $Y_i$ , condicionada también en los efectos aleatorios; 3) la componente de predictores lineales ( $\eta(\mathbf{x}_i, \mathbf{z}_i)$ ) se encuentra formada por la combinación lineal de efectos fijos y aleatorios; 4) se asume que  $\mathbf{u}$  proviene de cierta distribución, usualmente se asume normal<sup>3</sup>  $L_R$ -dimensional, con media nula y  $\boldsymbol{\Sigma}$  su matriz de covarianza.

Existen distintas funciones link que pueden ser utilizadas, sin embargo, dentro del área de psicometría son las funciones<sup>4</sup>  $\Phi$  y  $\Psi$  las función de enlace con las que se suele trabajar. En aquellos casos en que la función link es logit, al GLMM resultante se lo llama modelo logit mixto. Su representación algebraica esta dada por la Ec. 4.5, y al igual que en los GLM, se asume que  $y_i \in \{0, 1\}$ .

$$P(y_i | \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}, \mathbf{u}) = \frac{e^{\mathbf{x}_i' \cdot \boldsymbol{\beta} + \mathbf{z}_i' \cdot \mathbf{u}}}{1 + e^{\mathbf{x}_i' \cdot \boldsymbol{\beta} + \mathbf{z}_i' \cdot \mathbf{u}}} \quad (4.5)$$

Además, partiendo de la definición dada a través de la Ec. 4.4 y utilizando la Ec. 4.5, para un vector de observaciones  $\mathbf{y} = (y_1, y_2, \dots, y_n)$ , tenemos que la función de verosimilitud es:

$$L(\boldsymbol{\beta}, \boldsymbol{\Sigma} | \mathbf{y}, \mathbf{x}_i, \mathbf{z}_i) = \int \prod_{i=1}^n \frac{e^{\mathbf{x}_i' \cdot \boldsymbol{\beta} + \mathbf{z}_i' \cdot \mathbf{u}}}{1 + e^{\mathbf{x}_i' \cdot \boldsymbol{\beta} + \mathbf{z}_i' \cdot \mathbf{u}}} f_U(\mathbf{u} | \boldsymbol{\Sigma}) d\mathbf{u} \quad (4.6)$$

El hecho de incorporar efectos aleatorios al modelo trae consigo algunas consecuencias que se reflejan a través en el valor esperado, la varianza, covarianza y correlación marginales de la variable

<sup>3</sup>La hipótesis sobre media nula en los efectos aleatorios se debe al hecho que éstos representan la desviación respecto a la media modelada a través de los efectos fijos.

<sup>4</sup>En el present TFG la función  $\Phi$  corresponde a la función de distribución acumulativa para la  $N(0, 1)$ , mientras que  $\Psi$  es la función *logit*(0, 1).

de interés. Por ejemplo, McCulloch and Searle (2001) muestran que si se asume  $g(\cdot) = \log(\cdot)$  y  $u_i \sim N(0, \sigma_u^2)$ , entonces  $E(y_i) = \mathbf{x}'_i \cdot \boldsymbol{\beta} + \sigma_u^2/2$ , es decir, la esperanza se incrementa en función de la varianza del efecto aleatorio. Mientras que si  $y|u$  sigue una distribución de Poisson, utilizando la misma función link y distribución para  $u$ , entonces  $V(y_i) = E(y_i)(e^{\mathbf{x}'_i \cdot \boldsymbol{\beta}}[e^{3\sigma_u^2/2} - e^{\sigma_u^2/2}] + 1)$ , es decir, la varianza es mayor al valor esperado y crece de exponencialmente a medida que crece  $\mathbf{x}'_i \cdot \boldsymbol{\beta}$ .

Por último, cabe señalar que los GLMM nos permite estudiar el efecto medio a nivel poblacional al variar un efecto fijo, así como también a nivel de grupo o cluster. Esto se debe a que estamos incorporando en el modelo la posibilidad de modelar la variabilidad dentro del cluster, algo que al utilizar un GLM no es posible realizar.

### 4.3. Modelo Logit Mixto y Modelos de Respuesta al Ítem

En la Sec. 4.2 se introdujo los GLMM en forma genérica, a continuación estaremos presentando el vínculo con los modelos logísticos de la TRI. Como veremos, la expresión algebraica es la misma, sin embargo, algunos aspectos son necesarios considerar cuando los datos provienen de las respuestas a un test. Desde el punto de vista computacional, una de las principales ventajas que cuenta esta forma de representar los modelos de la TRI, radica en el hecho que existen muchos algoritmos implementados en software para los GLMM que pueden ser utilizados en el análisis de las respuestas a ítems.

Al igual que en los capítulos anteriores, supondremos que el test está formado por  $n$  ítems y son  $J$  los examinados que responden a cada ítem. Además, supondremos que las respuestas son dicotómicas, es decir,  $y_{ij} \in \{0, 1\}$ , donde  $y_{ij}$  corresponde a la respuesta del  $j$ -ésimo examinado al  $i$ -ésimo ítem, con  $i = 1, 2, \dots, n$  y  $j = 1, 2, \dots, J$ .

Partiendo de formulación dada por la Ec. 4.4, utilizando a la función logit como link y concibiendo a los vectores  $\mathbf{x}_{ij}, \mathbf{z}_{ij}$  como efectos fijos y aleatorios respectivamente; asumiendo que  $(\boldsymbol{\beta}, \boldsymbol{\theta}_j)$  son los parámetros<sup>5</sup> del modelo; entonces, bajo la hipótesis de independencia local<sup>6</sup>, la expresión dada por la Ec. 4.7 corresponde a el modelo logit mixto desde la perspectiva de la teoría de respuesta la ítem.

$$P(Y_{ij} = 1 | \mathbf{x}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\theta}_j) = \frac{e^{\mathbf{x}'_{ij} \cdot \boldsymbol{\beta} + \mathbf{z}'_{ij} \cdot \boldsymbol{\theta}_j}}{1 + e^{\mathbf{x}'_{ij} \cdot \boldsymbol{\beta} + \mathbf{z}'_{ij} \cdot \boldsymbol{\theta}_j}} \quad (4.7)$$

En la expresión anterior estamos considerando que  $\boldsymbol{\theta}_j \in \mathbb{R}^{L_R}$  y  $\boldsymbol{\beta} \in \mathbb{R}^{L_F}$ , son los vectores de parámetros del modelo. Asumiremos que la distribución de  $\boldsymbol{\theta}_j$  es normal multivariada con media nula y  $\boldsymbol{\Sigma}$  su matriz de covarianzas, es decir,  $\boldsymbol{\theta}_j \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ . Sumado a esto, supondremos que  $\boldsymbol{\theta}_j$  es independiente de  $\boldsymbol{\theta}_{j'}, \forall j \neq j'$ . El motivo por el cual los efectos aleatorios son considerados de media nula se debe al hecho que, éstos representan la desviación respecto a la media de los efectos fijos, por lo tanto, el interés radica en la matriz de covarianzas.

<sup>5</sup>Algunos autores lo denominan “efecto”.

<sup>6</sup>Observar que las respuestas  $(Y_{ij})$  de cada examinados no son mutuamente independientes, sin embargo, la hipótesis de independencia local permite asumir independencia en las observaciones condicionada a el nivel de habilidad.

A partir de la Ec. 4.7 y utilizando la distribución de  $\theta_j$ , para una matriz de respuestas  $\mathbf{y}$  cuyas componentes son los vectores de respuestas  $\mathbf{y}_j = (y_{1j}, y_{1j}, \dots, y_{nj})'$ , la distribución marginal se obtiene integrando sobre el vector de efectos aleatorios asociados al  $j$ -ésimo examinado. La expresión dada por la Ec. 4.8 corresponde a la verosimilitud marginal bajo las hipótesis antes mencionadas.

$$L(\boldsymbol{\beta}, \boldsymbol{\Sigma} | \mathbf{y}, \mathbf{X}, \mathbf{Z}) = L(\boldsymbol{\beta}, \boldsymbol{\Sigma} | (\mathbf{y}_1, \dots, \mathbf{y}_J), \mathbf{X}_j, \mathbf{Z}_j) = \int \prod_{j=1}^J P(\mathbf{y}_j | \mathbf{X}_j, \mathbf{Z}_j, \boldsymbol{\beta}, \boldsymbol{\Sigma}) N(\mathbf{0}, \boldsymbol{\Sigma}) d\boldsymbol{\theta}_j \quad (4.8)$$

En la expresión anterior  $\mathbf{X}_j \in \mathbb{R}^{n \times L_F}$  y  $\mathbf{Z}_j \in \mathbb{R}^{n \times L_R}$  representan las matrices cuya  $j$ -ésima fila corresponde a los vectores  $\mathbf{x}'_{ij}$  y  $\mathbf{z}'_{ij}$  asociados a los efectos fijos y aleatorios respectivamente.

Observar que el Modelo de Rasch, definido en la Ec. 2.28, en general se asume que  $\theta_j$  (rasgo latente) es una muestra de una población cuya distribución es normal de media nula y varianza  $\sigma_\theta^2$ , y  $b_i$  (dificultad del ítem) lo podemos considerar como invariante entre examinados. Entonces,  $\theta_j$  puede ser identificado en 4.7 como un efecto aleatorio, mientras que  $b_i$  es un efecto fijo. Luego, si definimos  $b_i = \beta_i$ , tenemos que el modelo de Rasch puede definirse a través de la Ec. 4.7, donde la matriz  $\mathbf{X}_j$  es la identidad, de dimensión  $n \times n$ , mientras que  $\mathbf{Z}_j$  es un vector de unos cuya dimensión es igual a  $n$ , es decir, la cantidad de ítems involucrados en el test. Esto nos conduce a un modelo logit mixto con efectos fijos y aleatorios únicamente en el término intercepto.

En forma análoga, si observamos la formulación del modelo LLTM (ver Sec. C.2), donde la dificultad del ítem se modela a través de la suma de funciones lineales. Entonces, siguiendo la misma línea de razonamiento, el modelo dado por la Ec. C.2 puede definirse a través de la Ec. 4.7, donde las matrices  $\mathbf{X}_j$  son todas iguales  $\forall j = 1, 2, \dots, J$ .

La misma idea puede aplicarse al Modelo de Rasch de Regresión Latente (ver Sec. C.4), el cual incorpora características propias de cada examinado a través de un modelo lineal. De modo que, si partimos de la Ec. C.5, la podemos transformar en la Ec. 4.7, ya que  $\theta_j = \mathbf{x}'_j \cdot \boldsymbol{\beta} + \epsilon_j$ , siendo  $\mathbf{x}_j$  el vector de covariables asociado a las características de los examinados y  $\epsilon_j \sim N(0, \sigma_\epsilon^2)$ ,  $Cov(\epsilon_j, \epsilon_{j'}) = 0, \forall j \neq j'$ . Es decir, la variable latente cuenta con una componente fija ( $\mathbf{x}'_j \cdot \boldsymbol{\beta}$ ) y una componente aleatoria ( $\epsilon_j$ ) modelada a través del error. Observar que la matriz  $\mathbf{X}_j$  estará formada por las covariables correspondientes a los ítems y a los examinados.

Al llegar a este punto vale la pena mencionar una aplicación que, de acuerdo con Rijmen et al. (2003), involucra la interacción entre covariables de ítems y covariables de examinados. Este tipo de formulación puede encontrarse en el análisis del Differential Item Functioning (DIF), y corresponde, por ejemplo, al caso en que dos grupos de examinados no se diferencian por el nivel habilidad, aunque sí se diferencian en el score total que obtienen. Esta clase de problemas es analizada en detalle por Holland and Wainer (1993).

De Boeck y Wilson (2004, Chap. 2) presentan una interesante perspectiva en lo que refiere al vínculo entre los GLMM y el modelo de Rasch, así como también a sus extensiones, clasificando los modelos (ver Tabla 4.1) en cuatro grupos: doblemente descriptivo, doblemente explicativo, explicativo por ítem y explicativo por examinado.

Tabla 4.1: Clasificación de modelos de respuesta la ítem. La función  $\eta(\cdot)$  corresponde a la inversa de la función link. Para cada modelo se presenta el valor de dicha función discriminando entre la componente asociada a el examinado y el ítem. Fuente: De Boeck and Wilson (2004)

Modelo	$\eta(\cdot) = g^{-1}(\cdot)$		Efectos Aleatorios	Grupo
	Componente del examinado	Componente del ítem		
Modelo de Rasch	$\theta_j$	$b_i$	$\theta_j$	Doblemente Descriptivo
Modelo de Rasch de Regresión Latente	$\mathbf{X}_j^E \cdot \boldsymbol{\beta} + \epsilon_j$	$b_i$	$\epsilon_j$	Explicativo por Examinado
Modelo de Test Lineal Logístico	$\theta_j$	$\mathbf{X}_i^I \cdot \mathbf{b}$	$\theta_j$	Explicativo por Ítem
Modelo de Test Lineal Logístico regresión latente	$\mathbf{X}_j^E \cdot \boldsymbol{\beta} + \epsilon_j$	$\mathbf{X}_i^I \cdot \mathbf{b}$	$\epsilon_j$	Doblemente Explicativo

En la Tabla 4.1 se encuentran los cuatro grupos de modelos que han sido presentados en el Cap. 2 y las covariables<sup>7</sup> que forman parte en cada uno de ellos. La diferencia entre grupos radica en el tipo de covariables involucradas, es decir, si éstas son covariables de ítem, examinado o ítem por examinado. Por lo tanto, partiendo de dicha clasificación es posible enlazar a cada grupo de modelos con su respectiva representación a través de un modelo de regresión logit mixto.

## 4.4. Modelos Jerárquicos y Modelos de Respuesta al Ítem

Los modelos lineales generalizados mixtos, por sus características, pueden ser aplicados en conjuntos de datos que presentan una estructura jerárquica o multinivel. Partiendo de la base que existen dos o más niveles de jerarquía, los GLMM permiten que los coeficientes puedan variar entre niveles. El concepto de jerarquía en los GLM es introducido por Lee and Nelder (1996) y luego es extendido por McCulloch (2003) a los GLMM.

Fox (2014) presenta algunos ejemplos de datos anidado: datos con dos niveles de jerarquía puede encontrarse cuando se analiza cierta característica de estudiantes que pertenecen (anidados) a una escuela, o individuos que pertenecen a un país; mientras que datos en tres niveles puede encontrarse en experimentos que involucran a estudiantes que pertenecen (anidados) a un aula (anidada) en la escuela en una localidad.

Aplicado este concepto a la teoría de los test, por ejemplo, es posible identificar una jerarquía de dos y tres niveles en los siguientes casos: 1) los ítems (nivel 1) se encuentran anidados en cada individuo (nivel 2); 2) los ítems (nivel 1) se encuentran anidados a los examinados (nivel 2), que a su vez, se encuentran anidados al centro donde se desarrolla dicha actividad (nivel 3), es decir, escuela, hospital, país, etc. En la literatura también se le suele llamar cluster a dicha estructura de grupos, ya que los elementos que forman parte del mismo cluster presentan características en común.

En cada nivel de la jerarquía tendremos componentes fijos y aleatorios que permiten modelar la variabilidad (heterogeneidad) dentro de cada grupo y entre los grupos. Por ejemplo, en el caso 2) antes mencionado, dentro del nivel 1 la dificultad del ítem puede formar parte de efectos fijos, es decir, representar la componente del modelo que no varía entre examinados o ser un efecto aleatorio si los ítems varían entre examinados. En el nivel 2, los efectos fijos podrían modelarse a

<sup>7</sup>Observar que aquí hemos utilizado los supraíndices  $E, I$  con el fin de diferenciar las covariables de examinados e ítems respectivamente.

través de una característica presente en los examinados (género), mientras que los efectos aleatorios representarán la componente que varía en el examinado, por ejemplo, el nivel de habilidad. Por último, en el nivel 3 de la jerarquía se encuentran los grupos de personas con alguna característica en común, que podrán ser toda la población objetivo (efecto fijo) o una muestra de ésta (efecto aleatorio).

A lo antes expuesto, como notamos en el Capítulo 2, es posible incorporar otro conjunto de variables que permitan describir la variabilidad en cada nivel. Fox y Glas (2016) presentan un interesante estudio sobre las pruebas PISA 2003 realizadas en Alemania dentro del área de matemáticas. Los autores utilizan un modelo de tres niveles (de idénticas características al caso 2), donde algunas de las variables explicativas que se incorporan al modelo son: 1) nivel de examinados, lugar de nacimiento y primera lengua; 2) nivel de escuela, se utiliza un índice que mide el nivel socio-económico y social de cada institución.

Esto nos conduce a pensar que existe una estructura jerárquica en los datos provenientes de las respuestas a los ítems en un test. Muchas veces dicha estructura es ignorada y se llevan a cabo investigaciones donde el análisis de los datos se realiza a través de, por ejemplo, los siguientes enfoques: proponer un modelo de respuesta al ítem y trabajar con los datos como un único grupo; proponer un modelo de respuesta al ítem para cada grupo; proponer un modelo de respuesta al ítem que contemple la existencia de grupos, aunque se ignora las posibles interacciones entre los niveles de jerarquía (Bartolucci et al., 2015).

Los planteamientos antes mencionados presentan distintas desventajas y el hecho de ignorar la existencia de una estructura jerárquica en los datos puede conducir a conclusiones equívocas y/o sesgo en los resultados. Es por tal motivo que, a la hora de proponer un modelo, resulta sumamente relevante poner en consideración las características intrínsecas que poseen las muestras a analizar.

En la TRI los modelos jerárquicos o multinivel fueron introducidos por Aitkin and Longford (1986), quienes mostraron su aplicación en el área de educación. Desde entonces han tomado cada vez más relevancia extendiéndose a diversos ámbitos, por ejemplo, en el diagnóstico de la enfermedad de Parkinson (Gottipati, Karlsson, and Plan, 2017).

Formalmente, adoptando la propuesta de Fox and Glas (2016), supongamos que un test con  $n$  ítems, será proporcionado a un conjunto de individuos que pertenecen a ciertos centros<sup>8</sup>. Además, consideremos que los centros,  $h = 1, 2, \dots, H$ , serán seleccionados a través de cierta estrategia de muestreo, y en cada centro elegido se toma una muestra de  $J$  individuos que responderán a los ítems del test. En nuestro enfoque asumiremos que los examinados responden a todos los ítems, sin embargo, el modelo jerárquico puede ser extendido a casos donde esta hipótesis no se cumple.

Los datos provenientes del test, cuya estructura admite tres niveles, es decir, los ítems se encuentran anidados a personas y las personas se encuentran anidadas a centros, pueden ser modelados a través de las Ecs. 4.9, 4.10 y 4.11. El primer nivel está asociado a las habilidades de los individuos y se define a través del modelo que representa la probabilidad de responder correctamente a los ítems<sup>9</sup>. Mientras que el segundo nivel corresponde al modelo que refleja variabilidad del rasgo latente de los individuos en cada centro, y en el tercer nivel se modela la variabilidad del rasgo latente entre centros.

---

<sup>8</sup>Por ejemplo, hospitales o escuelas.

<sup>9</sup>Por ejemplo, el modelo de Rasch, un modelo 2PL u ojiva normal.

$$P_i(\theta_{jh}) = \frac{e^{(\theta_{jh}-b_j)}}{1 + e^{(\theta_{ih}-b_i)}} \quad (4.9)$$

$$\theta_{jh} = \beta_{0j} + \sum_{l=1}^{l=L_J} x_{ljh} \beta_{lh} + \epsilon_{jh} = \mathbf{x}'_{jh} \cdot \boldsymbol{\beta}_h + \epsilon_{jh}^{(2)}, \quad \epsilon_{jh}^{(2)} \sim N(0, \sigma_\theta^2) \quad (4.10)$$

$$\beta_{lh} = \gamma_{l0} + \sum_{l'=1}^{l'=L_H} \gamma_{ll'} w_{l'lh} + \epsilon_{lj}^{(3)} = \mathbf{w}'_{lh} \cdot \boldsymbol{\gamma} + \epsilon_{lh}^{(3)}, \quad \epsilon_h^{(3)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\beta) \quad (4.11)$$

Donde  $P_i(\theta_{jh})$  es la probabilidad que el  $j$ -ésimo examinado (con un nivel de habilidad  $\theta_{jh}$ ), en el centro  $h$ , responde correctamente al  $i$ -ésimo ítem. Hemos supuesto que  $\theta_{jh}$  es explicado por un conjunto de variables que describen las características del examinado  $\mathbf{x}_{jh} = (1, x_{1jh}, x_{2jh}, \dots, x_{L_H jh})'$ , a través de un modelo lineal, cuyos errores<sup>10</sup> son iid y no correlacionados. Además, se asume que los parámetros de dicho modelo lineal varían entre centros y para ello se utiliza otro modelo lineal  $\mathbf{w}_{lh} = (1, w_{1lh}, w_{2lh}, \dots, w_{L_H lh})'$ , donde el vector de errores<sup>11</sup> se considera proveniente de una distribución normal multivariada con media cero y  $\boldsymbol{\Sigma}_\beta$  su matriz de covarianzas, tal que sus elementos son  $\sigma_{kk'}^2 \forall k, k' = 0, 1, 2, \dots, L_H$  y  $\sigma_k^2$  cuando  $k = k'$ .

En forma equivalente, el modelo representado por las Ecs. 4.10, 4.11 puede ser expresado en forma matricial de la siguiente forma:

$$\boldsymbol{\theta}_h = \mathbf{X}_h \cdot \boldsymbol{\beta}_h + \boldsymbol{\epsilon}_h^{(2)}, \quad \boldsymbol{\epsilon}_h^{(2)} \sim N(0, \sigma_\theta^2 \mathbf{I}_J) \quad (4.12)$$

$$\boldsymbol{\beta}_h = \mathbf{W}_h \cdot \boldsymbol{\gamma} + \boldsymbol{\epsilon}_h^{(3)}, \quad \boldsymbol{\epsilon}_h^{(3)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\beta) \quad (4.13)$$

$$\boldsymbol{\theta}_h = \begin{pmatrix} \theta_{1h} \\ \theta_{2h} \\ \vdots \\ \theta_{jh} \\ \vdots \\ \theta_{Jh} \end{pmatrix}, \quad \mathbf{X}_h = \begin{pmatrix} 1 & x_{11h} & \cdots & x_{L_J 1h} \\ 1 & x_{12h} & \cdots & x_{L_J 2h} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1jh} & \cdots & x_{L_J jh} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1Jh} & \cdots & x_{L_J Jh} \end{pmatrix}, \quad \boldsymbol{\beta}_h = \begin{pmatrix} \beta_{0h} \\ \beta_{1hj} \\ \vdots \\ \beta_{lh} \\ \vdots \\ \beta_{L_J h} \end{pmatrix}, \quad \boldsymbol{\epsilon}_h^{(2)} = \begin{pmatrix} \epsilon_{1h}^{(2)} \\ \epsilon_{2h}^{(2)} \\ \vdots \\ \epsilon_{jh}^{(2)} \\ \vdots \\ \epsilon_{Jh}^{(2)} \end{pmatrix}$$

<sup>10</sup>Se ha utilizado el supraíndice (2) para indicar que corresponden a los errores de nivel 2.

<sup>11</sup>Se ha utilizado el supraíndice (3) para indicar que corresponden a los errores de nivel 3.

$$\mathbf{W}_h = \begin{pmatrix} \mathbf{w}'_{0h} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \mathbf{w}'_{1h} & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \mathbf{w}'_{lh} & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \mathbf{w}'_{L_J h} \end{pmatrix}, \boldsymbol{\gamma} = \begin{pmatrix} \gamma_{00} \\ \gamma_{01} \\ \vdots \\ \gamma_{0L_H} \\ \gamma_{10} \\ \vdots \\ \gamma_{L_J L_H} \end{pmatrix}, \boldsymbol{\epsilon}_h^{(3)} = \begin{pmatrix} \epsilon_{0h}^{(3)} \\ \epsilon_{1h}^{(3)} \\ \vdots \\ \epsilon_{lh}^{(3)} \\ \vdots \\ \epsilon_{L_J h}^{(3)} \end{pmatrix}$$

Sustituyendo la Ec. 4.13 en 4.12 se tiene que:

$$\begin{aligned} \boldsymbol{\theta}_h &= \mathbf{X}_h \cdot \mathbf{W}_h \cdot \boldsymbol{\gamma} + \mathbf{X}_h \cdot \boldsymbol{\epsilon}_h^{(3)} + \boldsymbol{\epsilon}_j^{(2)} \\ &= \tilde{\mathbf{X}}_h \cdot \boldsymbol{\gamma} + \mathbf{X}_h \cdot \boldsymbol{\epsilon}_h^{(3)} + \boldsymbol{\epsilon}_h^{(2)} \end{aligned} \quad (4.14)$$

Observar que  $\tilde{\mathbf{X}}_h$  incluye información a nivel de examinado, centro, e interacción entre ellos (también llamada cross-level effect), es decir,  $\tilde{\mathbf{X}}_h$  refleja cambios en el efecto asociado a las variables en el nivel 1 debido a las características del nivel 2 y viceversa.

A partir del modelo propuesto a través de las Ecs. 4.9, 4.10, 4.11, utilizando las hipótesis antes mencionadas, tenemos que los dos primeros momentos para  $\boldsymbol{\theta}_h$  son:

$$E(\boldsymbol{\theta}_h | \mathbf{X}_h, \mathbf{W}_h) = \mathbf{X}_h \cdot \boldsymbol{\beta}_h \quad (4.15)$$

$$\begin{aligned} Var(\boldsymbol{\theta}_h | \mathbf{X}_h, \mathbf{W}_h) &= Var(\mathbf{X}_h \cdot \boldsymbol{\beta}_h) + \sigma_\theta^2 \mathbf{I}_J \\ &= \mathbf{X}_h \boldsymbol{\Sigma}_\beta \mathbf{X}_h^T + \sigma_\theta^2 \mathbf{I}_J \end{aligned} \quad (4.16)$$

Por último, vale la pena subrayar que esta forma de conceptualizar los modelos de TRI como modelos no lineales jerárquicos, es de aplicabilidad restringida, es decir, sólo es válido para los modelos Rasch o 1PL y 2PL, no así para otros modelos en la TRI como los modelos 3PL, ya que este último involucra el producto del parámetro de adivinación (Bartolucci et al., 2015). Un análisis detallado sobre los modelos jerárquicos y el impacto en las estimaciones al ignorar datos con estructura de cluster pueden encontrarse en Snijders and Bosker (2011).

## 4.5. Estimación de los parámetros del Modelo

El camino natural hacia la estimación de los parámetros, desde la óptica de la estadística clásica, es hacer uso de la función de verosimilitud del mismo modo que fue presentado en el Cap. 3. Por lo tanto, partiendo de la Ec. 4.6, una vez definida la distribución de los efectos aleatorios, estaremos en condiciones de buscar el valor óptimo de los parámetros que maximizan dicha función.

Supongamos que el modelo de respuesta al ítem es un logit mixto, entonces la Ec. 4.7 representa la probabilidad de responder en forma correcta un ítem. Por lo tanto, la Ec. 4.8 debe ser maximizada

en los parámetros  $(\beta, \Sigma)$  del modelo. Un aspecto clave radica en el cálculo de la integral involucrada en la función de verosimilitud. En general, dicha integral es multidimensional y no es posible calcularla en forma cerrada (Rijmen et al., 2003).

En este sentido se han propuesto distintos algoritmos, sin embargo, no siempre se logra convergencia. Además, el problema se hace más complejo cuando en el modelo se incorporan variables explicativas, como en caso de los modelos jerárquicos, lo cual trae consigo mayor cantidad de parámetros a estimar (Levy and Mislevy, 2016, Chap. 11). En consecuencia, es necesario recurrir a métodos alternativos.

La búsqueda de soluciones al problema de computar una integral en un espacio de dimensión finita, en particular cuando se trabaja con la función de verosimilitud en los GLMM, es un campo de investigación que lleva décadas y aún continúa activo. Algunas de las propuestas más relevantes, de acuerdo a McCulloch and Searle (2001), son: el algoritmo de cuadratura Gauss-Hermite, ampliamente utilizado cuando los efectos aleatorios presentan una distribución normal; integración Laguerre, la cual extiende el algoritmo de cuadratura Gauss-Hermite para casos en que la distribución de los efectos aleatorios pertenece a la familia exponencial; el algoritmo de expectation-maximization.

Alternativamente puede optarse por una perspectiva bayesiana, donde Albert (1992), Fox and Glas (2001) son pioneros en el desarrollo de estimaciones para modelos jerárquicos cuando el modelo de respuesta al ítem es ojiva normal de dos parámetros. Posteriormente, Fox (2010) extiende estas ideas para el caso de los modelos logísticos y es este último enfoque el que estaremos analizando en la sección que subsigue. Cabe señalar que, recientemente han surgido nuevas propuestas en la inferencia bayesiana, en particular mejoras a los algoritmos utilizados en la estimación de parámetros, por ejemplo, Kuo (2015), Sheng and Headrick (2012), o Su (2012).

## 4.6. Estimación bayesiana de los parámetros en el Modelo Jerárquico

En la presente sección abordaremos el enfoque bayesiano, bajo el alcance de los modelos jerárquicos, con el fin de estimar los parámetros involucrados en el modelo de TRI. Para ello, seguiremos la propuesta de Fox (2010), quien ha desarrollado un algoritmo que permite estimar en forma simultánea dichos parámetros haciendo uso<sup>12</sup> de MCMC, en particular combinando M-H y el muestreo de Gibbs.

El objetivo es estimar la distribución conjunta a posteriori de todos los parámetros involucrados en el modelo. En esta dirección el primer paso consiste en obtener la función de verosimilitud haciendo uso del modelo jerárquico. Supondremos, tal como vimos antes<sup>13</sup>, que los datos son obtenidos a partir de las respuestas de los examinados a un test y éstos son representados a través de un modelo jerárquico de tres niveles. Es decir, las respuestas a los ítems se encuentran anidadas a los examinados, y éstos se encuentran anidados a centros<sup>14</sup>.

---

<sup>12</sup>En el Apéndice D puede encontrarse una descripción de MCMC, así como también de los algoritmos Metropolis Hastings (M-H) y muestreo de Gibbs.

<sup>13</sup>Utilizaremos la misma notación que en la Sec. 4.4.

<sup>14</sup>Por ejemplo, escuelas u hospitales.

Si el modelo es logístico, las Ecs. 4.9, 4.10 y 4.11 son las expresiones analíticas para dicho modelo. De lo contrario alcanza con reemplazar la Ec. 4.9 por el modelo elegido, es decir, uno en donde se ha utilizado una función link normal, probit, etc.

Llamaremos<sup>15</sup>  $\mathbf{d}_i = (\mathbf{a}_i, \mathbf{b}_i)$  al vector de parámetros asociados a los ítems<sup>16</sup>, y asumimos que  $\mathbf{y}$  es la matriz de respuestas. Entonces, considerando la estructura jerárquica que presentan los datos, la función de verosimilitud<sup>17</sup> es:

$$p(\mathbf{y}|\mathbf{d}, \sigma_\theta^2, \gamma, \Sigma_\beta) = \prod_{h=1}^H \prod_{j=1|_h}^J p(\mathbf{y}_{jh}|\theta_{jh}, \mathbf{d}_i) p(\theta_{jh}|\mathbf{x}_{jh}, \beta_h, \sigma_\theta^2) p(\beta_h|\mathbf{W}_h, \gamma, \Sigma_\beta) p(\gamma|\Sigma_\beta) p(\mathbf{d}) p(\sigma_\theta^2) p(\Sigma_\beta) \quad (4.17)$$

Observar que en la expresión anterior estamos asumiendo independencia en: las respuestas de cada centro y entre los examinados. Además, la estructura de la Ec. 4.17 permite identificar cómo interviene cada nivel en la función de verosimilitud: el término  $p(\mathbf{y}_{jh}|\theta_{jh}, \mathbf{d}_i)$  corresponde a la distribución de las respuestas del j-ésimo examinado en el h-ésimo centro (nivel 1); el siguiente término,  $p(\theta_{jh}|\mathbf{x}_{jh}, \beta_h, \sigma_\theta^2)$ , corresponde a la distribución condicional del parámetro que modela la habilidad del j-ésimo examinado en el centro  $h$  (nivel 2); por último, el término  $p(\beta_h|\mathbf{W}_h, \gamma, \Sigma_\beta)$  y  $p(\gamma|\Sigma_\beta)$  corresponde a la distribución asociada a los efectos aleatorios entre centros (nivel 3).

El siguiente paso consiste en definir las distribuciones a priori (por ejemplo no informativas) para las componentes aleatorias involucradas en el modelo. En este sentido, siguiendo la misma línea de lo visto en el Cap. 3, J. Fox propone:

- Varianza de los errores  $\sigma_\theta^2$  (nivel 2) es generada por una distribución inversa gamma de parámetros  $g_1$  y  $g_2$

$$p(\sigma_\theta^2) \propto (\sigma_\theta^2)^{-(g_1+1)} e^{-\frac{g_2}{\sigma_\theta^2}} \quad (4.18)$$

Dado que, en general, los hiperparámetros son desconocidos, se recomienda utilizar  $g_1 = 1$  y  $g_2$  con un valor pequeño.

- Efectos fijos:

$$\gamma_{ll'} \stackrel{iid}{\sim} N(0, \sigma_\gamma^2), \forall ll' \quad (4.19)$$

El hiperparámetro  $\sigma_\gamma^2$  suele elegirse con un valor grande.

- Matriz de covarianzas  $\Sigma_\beta$  es generadas por una distribución inversa Wishart.

$$p(\Sigma_\beta|J, S_T) \propto |\Sigma_\beta|^{-\frac{J+Q+2}{2}} e^{-\frac{tr(S_T \Sigma_\beta^{-1})}{2}} \quad (4.20)$$

Se recomienda que  $S_T$  sea una matriz diagonal de unos, y  $J \geq Q + 1$ .

<sup>15</sup>En el Capítulo 3 definimos  $\mathbf{d} = (\mathbf{a}, \mathbf{b}, \mathbf{c})$ . sin embargo, como mencionamos en la sección anterior, al trabajar con estructuras jerárquicas sólo tienen sentido los modelos 1PL y 2PL, lo que equivale a decir  $\mathbf{c} = \mathbf{0}$ . Por tal motivo, en el presente capítulo se considera  $\mathbf{d} = (\mathbf{a}, \mathbf{b})$ .

<sup>16</sup>Parámetros de discriminación y dificultad de los ítems respectivamente.

<sup>17</sup>Tal como vimos antes en la Ec. 4.8.

En lo que sigue utilizaremos la notación del Cap. 2 donde, de acuerdo a las Ecs. 2.3 y 2.8, tenemos que  $\Upsilon_{ijh} \sim N(a_i(\theta_{jh} - b_i), 1)$  e  $Y_{ijh} \sim \text{logit}(a_i(\theta_{jh} - b_i), 1)$ . Además, como vimos a través de la Ec. 2.20, el factor de escala 1,7 es el que permite establecer el vínculo entre ambos enfoques.

A continuación se presenta la secuencia de pasos en el algoritmo MCMC propuesto por Fox 2010, para el caso de ítems binarios, que permite estimar los parámetros en forma conjunta:

1. Obtener la muestra de las realizaciones  $\mathbf{z}^{(m+1)}$ .

Esto es, la matriz de cuyos elementos son la probabilidad de responder correctamente a los ítems del test (o la matriz de respuestas binarias).

Se asume que son generadas por la distribución dado por la Ec. 2.3 (o 2.8), es decir,  $\Upsilon_{ijh}|Y_{ijh}, \theta_{jh}, \mathbf{d}_i \sim N(a_i(\theta_{jh} - b_i), 1)$  (o  $Y_{ijh}|\theta_{jh}, \mathbf{d}_i \sim \text{logit}(a_i(\theta_{jh} - b_i), 1)$ )

2. Obtener la muestra del vector de parámetros asociados a los ítems  $\mathbf{d}_i^{(m+1)}$  utilizando la distribución condicional.

Se asume que la distribución a priori de  $\mathbf{d}$  es normal multivariada intercambiable<sup>18</sup>, con media  $\boldsymbol{\mu}_d$  y matriz de covarianzas  $\boldsymbol{\Sigma}_d$ .

Sea  $\mathbf{H} = (1, 7\boldsymbol{\theta}, -1, 7\mathbf{1}_{\dim(\boldsymbol{\theta})})$  y  $\mathbf{d}_i^{(m+1)}$ , entonces, la muestra es generada a partir de:

$$\mathbf{d}_i | \mathbf{z}_i^{(m+1)}, \boldsymbol{\theta}^{(m)}, \boldsymbol{\mu}_d^{(m)}, \boldsymbol{\Sigma}_d^{(m)} \sim N(\boldsymbol{\mu}_d^*, \boldsymbol{\Omega}_d)$$

$$\boldsymbol{\mu}_d^* = \boldsymbol{\Omega}_d(\mathbf{H}'\mathbf{z}_i + \boldsymbol{\mu}_d\boldsymbol{\Sigma}_d^{-1})$$

$$\boldsymbol{\Omega}_d^{-1} = \varphi^{-1}(\mathbf{H}'\mathbf{H}) + \boldsymbol{\Sigma}_d^{-1}$$

$$\varphi = \begin{cases} 1 & \text{si } \Upsilon_i \text{ tiene una distribución normal.} \\ \pi^2/3 & \text{si } \Upsilon_i \text{ tiene una distribución logística.} \end{cases}$$

Observación: la distribución condicional de  $\mathbf{d}_i$  se utiliza para generar muestras, a través del algoritmo M-H, cuando  $\Upsilon_i \sim \text{logit}$ .

3. Obtener las muestras  $\boldsymbol{\mu}_d^{(m+1)}$  y  $\boldsymbol{\Sigma}_d^{(m+1)}$  a partir de las distribuciones condicional:

---

<sup>18</sup>Intercambiable en el sentido que la probabilidad conjunta de los parámetros del ítem es invariante con respecto a las permutaciones de los subíndices.

$$\boldsymbol{\mu}_d \mid \boldsymbol{\Sigma}_d^{(m)}, \mathbf{d}^{(m+1)} \sim N\left(\frac{K_0}{K_0 + I}\boldsymbol{\mu}_0 + \frac{I}{K_0 + I}\bar{\mathbf{d}}, \frac{1}{K_0 + I}\boldsymbol{\Sigma}_d\right)$$

$$\boldsymbol{\Sigma}_d \mid \mathbf{d}^{(m+1)} \sim IW(I + n_{df}, \boldsymbol{\Sigma}^*)$$

$$\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_0 + I\mathbf{S} + \frac{K_0 I}{K_0 + I}(\bar{\mathbf{d}} - \boldsymbol{\mu}_0)(\bar{\mathbf{d}} - \boldsymbol{\mu}_0)'$$

$$S = \sum_i \frac{1}{I}(\bar{\mathbf{d}} - \boldsymbol{\mu}_0)(\bar{\mathbf{d}} - \boldsymbol{\mu}_0)'$$

$$\bar{\mathbf{d}} = \sum_i \frac{1}{I}\mathbf{d}_i$$

Observar que en el paso 2, la distribución condicional de  $\mathbf{d}_i$  es utilizada para generar un valor candidato haciendo uso del algoritmo M-H. En consecuencia, ésta puede ser normal o logística dependiendo del modelo elegido. En el paso 3,  $J + n_{df}$  corresponde a los grados de libertad de la distribución inversa Wishart, donde  $n_{df} \geq 2$  es usualmente asumido. Además,  $K_0$  es un factor de escala, habitualmente utilizado cuando se trabaja con prioris conjugadas, como es el caso de la distribución normal (Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin, 2021, Chap. 3).

De esta forma el algoritmo propuesto por Fox (2010) permite, en cada paso de iteración, obtener las muestras de  $\boldsymbol{\zeta}, \boldsymbol{\theta}, \mathbf{d}, \boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d, \sigma_\theta^2, \boldsymbol{\gamma}, \boldsymbol{\Sigma}_\beta$ . Cuando las muestras son suficientemente grande, a partir de éstas es posible estimar las distribuciones a posteriori involucradas, sus momentos u otros estadísticos. Una de las principales ventajas, a diferencia de la propuesta de Fox y Glas 2001, radica en que el algoritmo puede ser utilizado cuando el modelo propuesto es ojiva normal o logit (1PL o 2PL). Además, puede ser extendido para el caso de estructuras más complejas que involucren mayor cantidad de niveles en la jerarquía.

## 4.7. Selección del Modelo

En las secciones anteriores presentamos una forma de abordar el problema distinta a la que se pueden encontrar en los textos tradicionales de la TRI. Es decir, haciendo uso de las características que presentan los datos en un test (jerarquía), fue posible incorporar efectos fijos y aleatorios en el modelo. De esta forma, desde una perspectiva de precisión en las estimaciones, es posible disminuir el sesgo. Sumado a esto, debido a la complejidad que trae consigo optimizar la función de verosimilitud, analizamos la aproximación bayesiana al problema, que nos permite estimar los parámetros involucrados en el modelo.

El siguiente paso, una vez que se tiene un conjunto de modelos candidatos, consiste en elegir aquel modelo que mejor se ajuste a los parámetros. El proceso de selección se torna complejo a medida que la cantidad de variables involucradas (efectos fijos y aleatorios) crece. Además, bajo

el enfoque bayesiano, no debe perderse de vista el costo computacional en la estimación de todos los parámetros involucrados, así como también la sensibilidad en los resultados al cambiar las densidades a priori.

Algunas de las técnicas empleadas dentro de la TRI en lo que refiere a la bondad en el ajuste del modelo son: Bayes Factor, Bayesian Information Criterion y Deviance Information Criterion, Widely Applicable Information Criterion y Leave-One-Out Cross-Validation. A continuación estaremos presentando cada uno de estos criterios, así como también sus propiedades.

#### 4.7.1. Factor de Bayes - BF

En estadística bayesiana el Bayes Factor (BF), propuesto por Jeffreys (1961), es comúnmente utilizado cuando se requiere comparar dos modelos. La idea detrás de dicho factor es similar a la del test de razón de verosimilitudes en la estadística clásica, aunque a diferencia de esta última, no se requiere que los modelos se encuentren anidados.

El BF se define como el cociente de la razón de la probabilidad a posteriori bajo la hipótesis nula y la hipótesis alternativa sobre la razón de la probabilidad a priori bajo la hipótesis nula y la hipótesis alternativa. Esto es:

$$BF = \frac{P(H_0|\mathbf{y})/P(H_1|\mathbf{y})}{P(H_0)/P(H_1)} \quad (4.21)$$

La hipótesis nula ( $H_0$ ) y alternativa ( $H_1$ ) corresponde a que el modelo  $M_1$  y  $M_2$  es el que mejor se ajusta a los datos respectivamente. Por lo tanto, podemos reescribir la Ec. 4.21 de la siguiente manera:

$$BF = \frac{P(M_1|\mathbf{y})}{P(M_2|\mathbf{y})} \cdot \frac{P(M_2)}{P(M_1)} = \frac{P(\mathbf{y}|M_1)}{P(\mathbf{y}|M_2)} \quad (4.22)$$

En la expresión anterior, para la segunda igualdad, se ha utilizado el Teorema de Bayes. Observar que  $BF \geq 0$  siempre, por lo tanto, se necesita algún criterio que permita decir cuándo seleccionar un modelo u otro. Jeffreys (Jeffreys, 1961, Chap. 5) obtiene una aproximación general que permite establecer un criterio de decisión, éste consiste en: si  $BF \geq 1$  entonces existe suficiente evidencia para afirmar que  $M_1$  se ajusta mejor a los datos, mientras que si  $BF < 0,01$ , entonces la evidencia es a favor  $M_2$  en el ajuste a los datos.

Una de las desventajas del BF se encuentra en el hecho que, en general, su valor está influenciado por las distribuciones a priori elegidas (Sinharay and Stern, 2012). Esto conduce a que sea necesario realizar un análisis de sensibilidad una vez obtenido el valor de dicho coeficiente (Kass, 1993; Weiss, 1996). Además, no siempre es sencillo obtener la distribución marginal de los datos, por lo tanto, su aplicación suele estar limitada.

### 4.7.2. Criterio de Información Bayesiana - BIC

El Bayesian Information Criterion (BIC), propuesto por Schwarz (1978), surge con motivo de abordar el problema de inconsistencia que presenta Akaike Information Criterion (AIC), es decir, no considerar el tamaño de la muestra en el término de penalización. BIC esta diseñado para ser consistente, para ello hace uso del Teorema de Bayes y se aplica una penalización asociada al número de parámetros utilizados en el modelo. Dicha penalización se incorpora a través de un sumando que depende del tamaño muestral.

El coeficiente BIC para el modelo  $M$  se define de la siguiente manera:

$$BIC(M) = -2 \ln(L(\mathbf{y}|\hat{\mathbf{d}}, \hat{\boldsymbol{\theta}})) + n_M \ln(n_T) \quad (4.23)$$

Siendo  $L(\mathbf{y}|\hat{\mathbf{d}}, \hat{\boldsymbol{\theta}})$  la función de verosimilitud evaluada en los parámetros estimados,  $n_M$  es el número de parámetros utilizados en el modelo y  $n_T$  es el tamaño de la muestra. Con el fin de simplificar la notación hemos utilizado  $(\hat{\mathbf{d}}, \hat{\boldsymbol{\theta}})$ , sin embargo, no debemos perder de vista que estos parámetros pueden depender de otros parámetros que forman parte del modelo, como es el caso de los modelos TRI jerárquicos.

El BIC resulta de gran utilidad cuando es difícil de obtener el coeficiente BF. Se puede observar que puede ser aplicado a modelos anidados, así como también a modelos no anidados. además, a medida que se añaden más variables al modelo, el ajuste mejora, sin embargo, el coeficiente de penalización ( $n_M \ln(n_T)$ ) se incrementa. De esta forma existe una compensación que penaliza el sobreajuste del modelo. Sumado a esto, otra de las ventajas presente en este criterio es el hecho que, a la hora de calcularlo, no involucra a las distribuciones a priori que se hayan utilizado en los parámetros.

Si  $M_1$  y  $M_2$  son los dos modelos a comparar tenemos que:

$$\Delta BIC(M_1, M_2) = -2 \ln(L_1 - L_2) + (n_{M_1} - n_{M_2}) \ln(n_T) \quad (4.24)$$

Donde  $L_1$  y  $L_2$  son las funciones de verosimilitud de cada modelo evaluadas en las estimaciones de los parámetros involucrados, mientras que  $n_{M_1}$  y  $n_{M_2}$  es la cantidad total de parámetros en cada caso. La regla de decisión consiste en elegir el modelo que presente el valor de BIC más bajo.

Vale la pena señalar que a partir de la Ec. 4.24, utilizando un resultado asintótico, se tiene que  $\Delta BIC(M_1, M_2)$  es aproximadamente igual a  $2 \log(BF)$  (Fox, 2010, Chap. 4). De allí el vínculo entre ambos criterios, el cual es válido cuando el número de parámetros del modelo es constante a medida que aumenta el número de observaciones.

Esto último es algo que no se cumple en los modelos de TRI, ya que un aumento en la número de muestras trae consigo un aumento en la cantidad de parámetros. Por tal motivo, no es recomendable utilizar  $\Delta BIC$  a la hora de estimar el BF.

Por último señalar que cuando el modelo en consideración presenta una estructura jerárquica, es decir, incorpora efectos fijos y aleatorios, BIC presenta algunas desventajas adicionales. La primera de ellas esta asociada al tamaño de la muestra. Definir  $n_T$  puede resultar ambiguo debido a que

en cada nivel de la jerarquía hay un tamaño de la muestra involucrado<sup>19</sup> (Hox, 2010, Chap. 3). En segundo lugar, no siempre es sencillo definir el número efectivo de parámetros que forman parte del modelo. En los modelos con efectos aleatorios dicho valor depende de la varianza en los niveles superiores, que a priori es desconocida su estructura, y que puede variar entre cero (todos los efectos aleatorios son iguales) e infinito (Fox, 2010, Chap. 4).

### 4.7.3. Criterio de Información Devianza - DIC

El criterio Deviance Information Criterion (DIC), propuesto por Spiegelhalter, Best, Carlin, and van der Linde (2002), es una métrica comúnmente utilizada a la hora de comparar modelos bayesianos. Está basado en elegir un modelo tal que el valor esperado de la distancia de Kullback-Leibler, entre el proceso que generó los datos y el modelo que representa a los datos, sea mínima.

Bajo la perspectiva de Spiegelhalter, DIC surge con el fin de ofrecer una alternativa a AIC o BIC cuando se trabaja con modelos jerárquicos y el número de parámetros involucrados no está claramente definido. En este sentido, DIC es una medida que permite establecer el número efectivo de parámetro en un modelo. Además, al igual que BIC, incorpora un término de penalización, de modo que el modelo parsimonioso sea el elegido.

Formalmente DIC se define de la siguiente manera:

$$DIC(\mathbf{d}, \boldsymbol{\theta}) = 2 \log(L(\mathbf{y}|\hat{\mathbf{d}}, \hat{\boldsymbol{\theta}})) + 2p_{DIC} \quad (4.25)$$

$$\begin{aligned} p_{DIC} &= E_{\mathbf{d}, \boldsymbol{\theta}|\mathbf{y}}(-2 \log L(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta})) + 2(\log(L(\mathbf{y}|\hat{\mathbf{d}}, \hat{\boldsymbol{\theta}}))) \\ &= \overline{D(\mathbf{d}, \boldsymbol{\theta})} - D(\hat{\mathbf{d}}, \hat{\boldsymbol{\theta}}) \end{aligned} \quad (4.26)$$

Siendo  $L(\cdot)$  la función de verosimilitud,  $E_{\mathbf{d}, \boldsymbol{\theta}|\mathbf{y}}$  el valor esperado con respecto a  $p(\mathbf{d}, \boldsymbol{\theta}|\mathbf{y})$  y  $(\hat{\mathbf{d}}, \hat{\boldsymbol{\theta}})$  son los parámetros estimados a partir de la muestra. El término  $p_{DIC}$  corresponde a la propuesta de Spiegelhalter en lo que refiere al número óptimo de parámetros en el modelo y es igual a la diferencia de devianza media y estimada.

Reescribiendo la Ec. 4.26 tenemos que  $\overline{D(\mathbf{d}, \boldsymbol{\theta})} = D(\hat{\mathbf{d}}, \hat{\boldsymbol{\theta}}) + p_{DIC}$ , de modo que  $p_{DIC}$  puede ser interpretado como una medida de complejidad cuando se está trabajando con modelos desde una perspectiva bayesiana. Además, en la Ec. 4.25 corresponde al término de penalización a la hora de seleccionar el un modelo.

A partir de  $\overline{D(\mathbf{d}, \boldsymbol{\theta})}$ ,  $D(\hat{\mathbf{d}}, \hat{\boldsymbol{\theta}})$  podemos reescribir la Ec. 4.25 de la siguiente manera:

$$\begin{aligned} DIC(\mathbf{d}, \boldsymbol{\theta}) &= D(\hat{\mathbf{d}}, \hat{\boldsymbol{\theta}}) + 2p_{DIC} \\ &= \overline{D(\mathbf{d}, \boldsymbol{\theta})} + p_{DIC} \end{aligned} \quad (4.27)$$

Al igual que antes, el modelo que presente menor valor de  $DIC$  será seleccionado. El número óptimo de parámetros  $p_{DIC}$  puede obtenerse a través de MCMC o utilizando la mediana a posteriori. Sin

---

<sup>19</sup>Este aspecto puede resultar controversial, ya que otros autores como Morales Correa and Salazar Uribe (2016) o Pinheiro and Bates (2000) definen  $n_M$  y  $n_T$  como el número total de parámetros estimados y el número de observaciones (no missing) respectivamente utilizadas en el proceso de estimación.

embargo, este último estimador puede tomar valores negativos, lo cual indicaría que no es adecuado utilizarlo en estos casos Lock (2022); Spiegelhalter et al. (2002). Un análisis completo sobre DIC puede encontrarse en Li and Wang (2022); LI, YU, , and ZENG (2017); Spiegelhalter, Best, Carlin, and van der Linde (2014)

#### 4.7.4. WAIC

En estadística bayesiana también podemos encontrar el Widely Applicable Information Criterion (WAIC), propuesto por Watanabe (2010), el cual es considerado como un criterio completamente bayesiano que permite estimar el valor esperado de una observación que no pertenece a la muestra. El criterio consiste en utilizar nuevos datos para evaluar la performance del modelo ajustado y para ello se incluye una corrección debido al sesgo introducido por la muestra de entrenamiento en el ajuste.

Una de las alternativas para computar el WAIC es siguiendo un razonamiento análogo al que hicimos para DIC, es decir, en lugar de construir  $p_{DIC}$ , se construye  $p_{WAIC}$ :

$$p_{WAIC} = 2 \sum_{i=1}^n \left( \log(E_{post}(p(y_i|\boldsymbol{\theta}))) - E_{post}(\log(p(y_i|\boldsymbol{\theta}))) \right) \quad (4.28)$$

Siendo  $E_{post}(\cdot)$  el valor esperado utilizando la distribución a posteriori, que puede ser computado a través de algoritmos de simulación, y  $n$  es la cantidad de muestras. Luego,  $p_{WAIC}$  es utilizado como factor de corrección del sesgo:

$$\widehat{\text{elppd}}_{WAIC} = \text{lppd} - p_{WAIC} \quad (4.29)$$

Luego, se define:

$$\text{WAIC} = -2(\text{lppd} - p_{WAIC}) \quad (4.30)$$

En la Ec. 4.29 se ha utilizado la denominada log pointwise predictive density (lppd) para un nuevo dato, que es una forma de resumir la precisión predictiva del modelo ajustado a los datos, mientras que expected log pointwise predictive density (elppd) es su valor esperado cuando son consideradas  $n$  nuevas observaciones, esto es:

$$\text{lppd} = \log \prod_{i=1}^n p_{post}(y_i) = \sum_{i=1}^n \log \int_{\boldsymbol{\Theta}} p(y_i|\boldsymbol{\theta}) p_{post}(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (4.31)$$

$$\text{elppd} = \sum_{i=1}^n \log E_f(p_{post}(y_i)) \quad (4.32)$$

La forma de computar las expresiones 4.28, 4.31 es haciendo uso del método de Monte Carlo, es decir, a partir de las  $s = 1, 2, \dots, S$  muestras tomar el promedio:

$$p_{WAIC} \approx 2 \sum_{i=1}^n \left( \log \left( \frac{1}{S} \sum_{s=1}^S p(y_i|\boldsymbol{\theta}^s) \right) - \frac{1}{S} \sum_{s=1}^S \log \left( p(y_i|\boldsymbol{\theta}^s) \right) \right) \quad (4.33)$$

$$\text{lppd} \approx \sum_{i=1}^n \log \left( \frac{1}{S} \sum_{s=1}^S p(y_i | \boldsymbol{\theta}^s) \right) \quad (4.34)$$

En modelos lineales, con errores iid provenientes de una distribución normal, y con un tamaño de muestra grande, se puede probar que  $p_{WAIC}$  es aproximadamente igual al número de parámetros en el modelo. Por lo tanto, al igual que sucedía en DIC con  $p_{DIC}$ , el término  $p_{WAIC}$ , bajo esta hipótesis, puede ser visto como una medida de complejidad del modelo.

Una ventaja de WAIC respecto a DIC es que cuenta con la propiedad de promediar sobre la distribución a posteriori, en lugar de condicionar en una estimación puntual (Gelman, Hwang, and Vehtari, 2014). Esta es una propiedad relevante cuando lo que se busca es el mejor modelo con capacidad predictiva. A pesar de sus bondades, en muchos problemas reales no se utiliza debido a la demanda computacional necesaria que posee.

#### 4.7.5. LOOCV

Los criterios AIC, BIC, DIC y WAIC, pueden ser vistos como distintas aproximaciones o versiones de la técnica cross-validation (Gelman et al., 2014; Stone, 1997). Un problema intrínseco de esta última es el aspecto computacional, ya que es necesario ajustar el modelo reiteradas veces para distintos conjuntos de datos. Además, Gelman et al. (2014) muestra que, en ciertos escenarios, AIC, DIC y WAIC pueden fallar, conduciendo a decisiones equívocas.

Recientemente, a partir de la imposibilidad en garantizar efectividad mediante los métodos de selección conocidos, se gestó el criterio Leave-One-Out Cross-Validation (LOOCV) en modelos bayesianos. El costo computacional involucrado es reducido haciendo uso de la técnica Importance Sampling (IS). Ésta es definida por Tokdar and Kass (2009) como una colección de métodos Monte Carlo, donde el valor esperado, respecto a la distribución objetivo, es aproximado por el promedio ponderado de muestras aleatorias tomadas de otras distribuciones. Geweke (1989) demuestra que en la aplicación de IS es esencial asumir varianza finita en la distribución a posteriori de los parámetros, esto permite asegurar convergencia en distribución al verdadero parámetro.

En LOOCV, la idea es dividir el conjunto de observaciones en dos grupos, donde uno de éstos (denominado conjunto de entrenamiento) será utilizado para ajustar el modelo, y de esta forma se obtiene la distribución a posteriori  $p_{\text{post}}(\boldsymbol{\theta} | \mathbf{y}_{\text{train}})$  de los parámetros. Luego, el modelo obtenido es validado haciendo uso de el segundo subconjunto de datos (denominado datos de validación), esto es,  $\log(p_{\text{fit}}(\mathbf{y}_{\text{valid}})) = \log \int p_{\text{fit}}(\mathbf{y}_{\text{valid}} | \boldsymbol{\theta}) p_{\text{post}}(\boldsymbol{\theta} | \mathbf{y}_{\text{train}}) d\boldsymbol{\theta}$ .

El algoritmo, de acuerdo a Gelman et al. (2021); Stern (2019), lo podemos resumir en los siguientes pasos:

- Generar simulaciones de  $p_{\text{post}-i}(\boldsymbol{\theta}^s | \mathbf{y}_{-i})$ , s-ésima distribución a posteriori sin la i-ésima muestra, con  $s = 1, 2, \dots, S$ ,  $i = 1, 2, \dots, n$ .

- Computar lppd:

$$\begin{aligned}
\text{lppd}_{\text{LOOCV}} &= \log \prod_{i=1}^n p_{\text{post}(-i)}(y_i) = \sum_{i=1}^n \log p_{\text{post}(-i)}(y_i) \\
&= \sum_{i=1}^n \log \int_{\Theta} p_{\text{post}(-i)}(y_i | \boldsymbol{\theta}) p_{\text{post}}(\boldsymbol{\theta} | \mathbf{y}_{\text{train}}) d\boldsymbol{\theta} \\
&\approx \sum_{i=1}^n \log \left( \frac{1}{S} \sum_{s=1}^S p(y_i | \boldsymbol{\theta}^s) \right)
\end{aligned} \tag{4.35}$$

Se asume que  $S$  es lo suficientemente grande, de forma tal que permite capturar adecuadamente la forma de la distribución a posteriori.

Observación: en la estimación de  $p_{\text{post}(-i)}(y_i)$  existe sesgo, ya que el modelo utilizado ha sido ajustado utilizando  $n - 1$  muestras.

- Las comparaciones entre modelos se realizan al computar el número efectivo de parámetros:

$$p_{\text{LOOCV}} = \text{lppd} - \text{lppd}_{\text{LOOCV}} \tag{4.36}$$

En este desarrollo estamos utilizando la definición dada en la Ec. 4.31, que puede ser aproximada por la Ec. 4.34. Además, hemos omitido un factor de ajuste por sesgo que habitualmente no es aplicado. Resulta importante destacar que el algoritmo puede ser extendido dejando  $k$  muestras fuera del subconjunto de datos de entrenamiento.

Finalmente, señalar lo siguiente: algunos autores, por ejemplo Epifani, MacEachern, and Peruggia (2008), prueban que en cierto conjunto de problemas, los resultados obtenidos en la aplicación de Leave-One-Out (LOO) haciendo uso de IS conducen a varianza no finita en la distribución a posteriori de los parámetros. Es Vehtari, Gelman, and Gabry (2017) quien propone una solución a este problema introduciendo Pareto Smoothed Importance Sampling Leave-One-Out (PSIS). Es decir, al igual que antes, aplicar LOO e incorporar la distribución de Pareto en el ajuste de la distribución de los pesos de IS. De esta forma se logra mejorar la confiabilidad, estabilidad y precisión de las estimaciones de la varianza.

## 4.8. Observaciones Finales

En el presente capítulo hemos realizado una introducción a los modelos lineales generalizados e introdujimos la estructura jerárquica presente en los datos que provienen de un test. Señalamos el potencial problema de sesgo que puede introducirse en las estimaciones de los parámetros, en un modelo de TRI, al no considerar dicha estructura. Estos conceptos nos abrieron paso para enlazar los GLMM y el modelo logit en la TRI.

Otro aspecto relevante que analizamos corresponde a las técnicas que se pueden utilizar a la hora de estimar los parámetros del modelo. En particular, la dificultad presente tanto en el modelo ojiva normal, así como también en el logit. Esto nos condujo al enfoque bayesiano y el uso del algoritmo MCMC. Éste introduce un gran desafío, desde el punto de vista computacional, debido a la cantidad de parámetros a estimar en un modelo con efectos fijos y aleatorios.

Por último, fueron abordadas las técnicas más utilizadas en la selección de modelos GLMM bajo el enfoque bayesiano. Cada una de éstas cuenta con una sólida base teórica que no hemos profundizado, ya que requeriría un TFG en sí mismo, sin embargo, han sido presentados los principales resultados y las referencias a los autores que realizan una línea de investigación activa en esta línea.

Esta página ha sido intencionalmente dejada en blanco.

## Capítulo 5

# Aplicación del Modelo de Rasch a un Test utilizado en la identificación del TEA

## 5.1. Introducción

El TEA es considerado una afección que se encuentra relacionada con el desarrollo del cerebro, y afecta la manera en que una persona percibe y socializa con otras personas (Mayo Clinic, 2021). Dicho trastorno presenta un alto grado de complejidad y sus causas pueden deberse a distintos factores. Recientes investigaciones muestran que una posible causa estaría asociada al desarrollo neurológico, factores genéticos y ambientales (Sandin, Lichtenstein, Kuja-Halkola, Larsson, Hultman, and Reichenberg, 2014).

Según la Organización Mundial para la Salud (2023), uno de cada cien niños en el mundo presenta TEA. En 2014, la Asamblea Mundial de la Salud adoptó la resolución “Medidas integrales y coordinadas para gestionar los trastornos del espectro autista”, donde, entre otras cosas, se busca fortalecer la capacidad del personal de salud, mejorar la atención de las personas con autismo, y fomentar los entornos inclusivos.

Un aspecto crucial en esta dirección es el proceso de diagnóstico, sin embargo, en la actualidad los costos de atención médica asociados al TEA, a nivel mundial, son significativos. Un diagnóstico temprano de dicho trastorno puede contribuir a reducirlos, así como también en mejorar la calidad de vida e interacción social de las personas.

En Uruguay, de acuerdo a información de Presidencia<sup>1</sup>, en 2015 aproximadamente entre 30,000 y 35,000 uruguayos presentaban TEA. Al día de hoy no se cuenta con información precisa por parte de las oficinas gubernamentales o asociaciones civiles. Será sobre el final de 2023, gracias al censo que se llevará a cabo, donde dispondremos de datos oficiales. En dicho censo, conforme a la comunicación<sup>23</sup> del INE a través del sitio web, por primera vez serán registradas las personas que han sido diagnosticadas con autismo.

La relevancia que ha tomado la investigación en torno al TEA y las características de los modelos que hemos presentado, nos motivan a ponerlos a prueba con este tipo de datos. Cabe señalar que el Modelo de Rasch es una herramienta que se ha utilizado en test de diagnóstico del TEA. Por ejemplo, Lundqvist and Lindner (2017) han diseñado un test y lo han aplicado a pacientes en Suecia.

Desafortunadamente, a la hora de llevar adelante investigación en esta área, uno de los problemas que podemos encontrar está asociado a la disponibilidad de datos. Existen organizaciones en USA y Europa que disponen de este tipo de bases de datos, las cuales comercializan el acceso, y muchas veces únicamente proyectos académicos (que cuenten con un aval) pueden obtenerlos. Esta realidad restringe la posibilidad de contrastar modelos que han sido propuestos con respecto a nuestro acercamiento al problema.

---

<sup>1</sup><https://www.gub.uy/presidencia/comunicacion/noticias/30000-uruguayos-presentan-trastorno-del-espectro-autista>.

<sup>2</sup><https://censo2023.uy/nombre-tiene- diagnostico-de-autismo-o-trastorno-del-espectro-autista-tea/>.

<sup>3</sup><https://preprod.censo2023.uy/wp-content/uploads/2023/02/Cuestionario-Censo-personas-viviendas-hogares-2011.pdf>

El acceso a datos fue un obstáculo a sortear en el presente TFG. Luego de establecer distintos contactos con investigadores en Argentina, Australia, Suecia y USA, sin éxito en el establecimiento de acuerdos, encontramos una base de datos abierta que era el resultado de un proyecto del Department of Digital Technology, en el Manukau Institute of Technology de Nueva Zelanda.

El proyecto de investigación consistió en el desarrollo de un test de comportamiento, así como también una página web y una aplicación para teléfonos celulares, con acceso a personas de cualquier parte del planeta. El objetivo principal es la generación de datos abiertos que pudieran ser utilizados, con fines académicos, en la identificación del TEA en niños, adolescentes y adultos Thabtah (2017a,1,1,2,1,1).

La importancia en el diagnóstico a edades tempranas nos condujo a pensar en datos que provienen de niños. En este sentido, el conjunto de datos seleccionado corresponde al resultado del test de comportamiento realizado a niños menores de 5 años (Thabtah, 2020). A través de dichos datos evaluaremos un conjunto de modelos y analizaremos su performance.

En las siguientes secciones presentaremos el conjunto de datos y realizaremos un análisis descriptivo. Además, un conjunto de modelos son propuestos, y se aplicarán las técnicas antes vistas en lo que refiere a selección. Luego compararemos la bondad en el ajuste, y por último, expondremos las conclusiones y trabajos a futuro. Los resultados aquí presentados se han implementado en R (R Core Team, 2022, versión 4.1.3) y son complementados en el Apéndice G. El código utilizado se encuentra en el Apéndice H.

## 5.2. Descripción de los Datos

El conjunto de datos elegido es el denominado “Autistic Spectrum Disorder Screening Data for Toddlers” Thabtah (2020). La forma en que se han obtenido las respuestas al test de comportamiento para la identificación del TEA, es a través de la plataforma ASDTests (Thabtah, 2017b). Esto hace que no exista un diseño de muestreo en la selección de participantes, sino que aquellos que cuenten con acceso a internet pueden realizar el test.

El test cuenta con diez características (o ítems) de comportamiento (llamadas Q-Chat, ver Tabla 5.1) y una serie de variables que son proporcionadas por la persona que realiza el test. Dichas características individuales, de acuerdo Thabtah (2020), han demostrado ser efectivas para detectar los casos de TEA en la ciencia del comportamiento.

Tabla 5.1: Ítems del test “Autistic Spectrum Disorder Screening Data for Toddlers” Thabtah (2020).

Ítem	Pregunta
A1	¿El niño/niña te mira cuando lo llamas por su nombre?
A2	¿Qué tan fácil es hacer contacto visual con el niño/niña?
A3	¿El niño/niña señala indicando que el/ella quiere algo? (ejemplo, un juguete que está fuera de su alcance)
A4	¿El niño/niña señala algo de interés para compartir con usted? (ejemplo, señalar una vista interesante)
A5	¿El niño/niña simula acciones? (ejemplo, cuidar a las muñecas, hablar por teléfono)
A6	¿El niño/niña sigue donde usted está mirando?
A7	¿Si usted o alguien en la familia esta visiblemente decepcionado, el niño/niña muestra signos de querer reconfortarlos? (ejemplo, acariciándoles el cabello, abrazándolos)
A8	Describiría las primeras palabras del niño/niña (la app. mostrará una lista desplegable)
A9	¿El niño/niña utiliza gestos simples? (ejemplo, mover la mano para decir adiós)
A10	¿El niño/niña mira fijamente a la nada sin un propósito aparente?

En la Tabla 5.2 se ha incluido el nombre de los campos en el conjunto de datos, su tipo, y una descripción de las características, así como también de las variables relevadas.

Tabla 5.2: Descripción de las variables que forman parte del conjunto de datos publicado por Thabtah (2020).

Variable	Tipo	Descripción
ID	Numérica	Identificador de participante
A1, . . . , A10	Binaria	Código de respuesta a la pregunta según el método de selección utilizado {0, 1}
Age.Mons	Numérica	Edad de los niños en meses
Score	Numérica	Score total en el test 1, . . . , 10 Si Score $\leq 3 \Rightarrow$ no requiera un tratamiento de TEA. Si Score $> 3 \Rightarrow$ requiera un tratamiento de TEA.
Sex	Caracter	Sexo: F (Femenino), M (Masculino)
Ethnicity	String	Etnia: lista de etnias en formato texto
Jaundice	Booleano	Nacida con ictericia: Y (Yes), N (No)
Family_mem.with.ASD	Booleano	Familiar cercano con antecedentes de ASD: Y (Yes), N (No)
Who.completed.the.test	String	Quién ha completado el Test, Padres, paciente, tutor, personal médico, médico, etc.
Class.ASD.Traits.	String	Rasgos ASD (Yes) o Rasgos No ASD (No) (asignados automáticamente por la aplicación ASDTests)

Además de las variables disponibles en el conjunto de datos, decidimos crear una nueva variable categórica llamada Age\_Cat, donde se agrupan los niños de acuerdo a si su edad en meses es menor a 24 meses (categoría L1) o mayor a 24 meses (categoría L2).

### 5.3. Descripción de los Datos

El conjunto de datos cuenta con 1054 participantes, de los cuales 29 se han identificado como profesionales de la salud, el resto se han identificado como miembros de la familia u otros. Dado que las respuestas provenientes de profesionales de la salud pueden corresponder a pruebas realizadas a través de la plataforma, hemos decidido eliminar dichas respuestas.

Por tanto, el conjunto de datos a analizar presenta 1025 participantes, 314 son de sexo femenino, mientras que 711 son de sexo masculino. Del total de participantes, 861 han indicado que un familiar ha sido diagnosticado con TEA, 283 han nacido con ictericia (Jaundice) y se han registrado etnias en donde han predominado europeos y asiáticos.

Una de las columnas proporcionadas es el resultado del test, de los 1025 participantes, 317 presentan rasgos de TEA. Es decir, estos 317 participantes han obtenido un score total en el test mayor a tres y los 708 restantes han obtenido un score total menor a tres.

En las Tablas 5.3 y 5.4 se muestran las medidas de resumen y cantidad de participantes para las variables en el conjunto de datos. Mientras que en la Tabla 5.5 se presenta la media y el desvío estándar del Score total obtenido por parte de los participantes que han sido identificados con rasgos de TEA. Los valores se han discriminado de acuerdo a las variables Sexo, ictericia (Juandice), familiar diagnosticado con TEA (Family\_mem\_with\_ASD) y Categoría de edad (Age\_cat).

Tabla 5.3: Medidas de resumen para las variables Age\_Mons y Score.

	Mínimo	1er. Cuartil	Mediana	Media	3er. Cuartil	Máximo
Age_Mons	12	23	30	27	36	36
Score	0	3	5	5.2	8	10

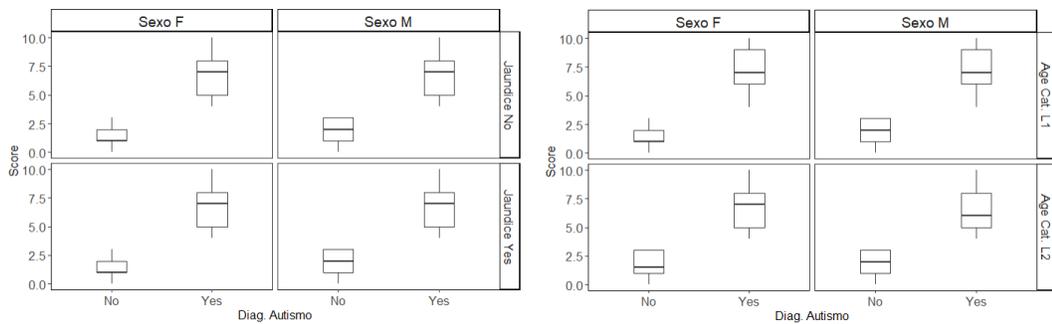
Tabla 5.4: Cantidad de participantes en cada categoría para las variables Sex, Jaundice, Class.ASD.Traits., Family\_mem\_with\_ASD, Age\_Cat y Who.completed.the.test.

	F	M	No	Yes	L1	L2	Family Member	Others
Sex	314	711						
Jaundice			742	283				
Class.ASD.Traits.			317	708				
Family_mem_with_ASD			861	164				
Age_Cat					276	749		
Who.completed.the.test							1018	3

Tabla 5.5: Promedio y desvío estándar (SD) del Score total obtenido en el test, agrupados según el diagnóstico (Class.ASD.Traits.), desglosado por: Todos los participantes, Sexo, ictericia (Jaundice), familiar diagnosticado con TEA (Family\_mem\_with\_ASD), Categoría de edad (Age\_cat).

	Class.ASD.Traits.: Yes		Class.ASD.Traits.: No	
	Media	DS	Media	DS
Todos Part.	6.78	1.91	1.69	1.07
Sexo F	6.82	1.83	1.43	1.07
Sexo M	6.76	1.94	1.86	1.05
Jaundice: No	6.85	1.88	1.70	1.09
Jaundice: Yes	6.61	1.99	1.67	1.03
Family_mem_with_ASD: No	6.74	1.92	1.71	1.07
Family_mem_with_ASD: Yes	6.99	1.87	1.61	1.09
Age_cat	7.26	1.81	1.53	1.03
Age_cat	6.62	1.92	1.77	1.08

En la Figura 5.1 se presenta la variación del Score total, observado en los participantes, agrupados por Sexo, según si presentan ictericia (Jaundice) y la categoría de edad.



(a) Participantes agrupados por Sexo e ictericia (Jaundice).

(b) Participantes agrupados por Sexo y categoría de edad.

Figura 5.1: Boxplot: variación del Score total observado en los participantes.

A partir de los resultados de la Figura 5.1 podemos señalar:

- No existen diferencias notorias en el rango de variación del Score total en participantes con rasgos de TEA, en el grupo que ha indicado presentar ictericia, así como también en las categorías de edad. Lo opuesto sucede cuando el grupo corresponde a los participantes que no presentan rasgos de TEA. En ambos casos sí observan diferencias en las medianas entre grupos.
- Existen diferencias en el rango de variación del Score total y la mediana en los participantes que no presentan rasgos de TEA. Esto se observa en el grupo que ha indicado presentar ictericia, así como también en las categorías de edad.

Además de lo presentando en esta sección, hemos incluido otras medidas de centralidad y gráficos en el Apéndice G.

## 5.4. Elección de Variables

Al comienzo del capítulo se han presentado las variables disponibles en el conjunto de datos. Básicamente tenemos dos grupos de variables, en uno de ellos se encuentran los ítems que deben responder los examinados, mientras que en otro se encuentran las variables que se han relevado asociadas a las características propias de los examinados y de su entorno.

En el Capítulo 2 y 4 fue presentado el Modelo de Rasch y sus extensiones, así como también los GLMM. Ambas perspectivas incorporan variables con el fin de explicar la variabilidad en la probabilidad de responder correctamente a un ítem. Esto nos conduce a la necesidad de tomar una decisión sobre cuáles y cuántas variables incorporar en los modelos a ser propuestos.

Dado que no contamos con conocimiento previo en el área de investigación del TEA, así como la posibilidad de consultar a expertos, decidimos dejar de lado las variables asociadas a aspectos étnicos y antecedentes familiares con TEA. Sabemos que esta decisión es arbitraria y por ende cuestionable. Sin embargo, puede considerarse como una primera aproximación al problema.

En resumen, consideraremos en el modelado las variables: ID, A1, . . . , A10, Sex, Jaundice y la nueva variable que construimos Age.Cat. Los modelos que propondremos en el presente TFG pueden ser extendidos, incorporando el resto de las variables, e incluso hasta podría incorporarse interacción entre variables.

## 5.5. Modelos Propuestos

En el marco de la aplicación de los modelos antes vistos al conjunto de datos seleccionado, proponemos comparar el desempeño del Modelo de Rasch con respecto al Modelo de Rasch Regresión Latente representado como un GLMM. Esta aproximación al problema nos permitirá construir modelos explicativos de la probabilidad a cada ítem correctamente.

Bajo nuestro abordaje del problema, nos proponemos utilizar como referencia un modelo sencillo, como lo es el Modelo de Rasch (representado a través de la Ec. 2.28), e incorporar complejidad al modelo a través de una de sus extensiones, esto es, el Modelo de Rasch de Regresión Latente (introducido en la Sec. C.4). En particular, este último será representado través de un modelo logit mixto, el cual fue analizado en la Sec. 4.3. Por sencillez, lo llamaremos Modelo de Rasch de Regresión Latente - GLMM (MRRL-GLMM).

El Modelo de Rasch de Regresión Latente permite considerar el efecto de individuo, es decir, la variabilidad propia inherente en cada examinado. Por lo tanto, supondremos que el rasgo latente  $\theta_j \in \mathbb{R}$ ,  $\theta_j = \mathbf{x}'_j \cdot \boldsymbol{\beta} + \epsilon_j$ , siendo  $\mathbf{x}_j$  el vector de covariables asociado a las características de los examinados y  $\epsilon_j \sim N(0, \sigma_{\epsilon_j}^2)$ ,  $Cov(\epsilon_j, \epsilon_{j'}) = 0, \forall j \neq j'$ . Es decir, la variable latente cuenta con una componente fija ( $\mathbf{x}'_j \cdot \boldsymbol{\beta}$ ) y una componente aleatoria modelada a través del error. Las covariables a utilizar son las que hemos seleccionado en la sección anterior.

Los parámetros en el MR serán estimados utilizando CML a través de la librería eRm (Mair and Hatzinger, 2007), mientras que los parámetros del MRRL-GLMM serán estimados mediante ML (a través de la aproximación de Laplace), e inferencia bayesiana. Para ello, haremos uso de las librerías lmer (Bates, Mächler, Bolker, and Walker, 2015) y brms (Bürkner, 2017) respectivamente.

La Tabla 5.6 resume las variables involucradas en cada modelo, el efecto que representa (fijo o aleatorio) y el tipo de inferencia utilizada. Observar que en el caso de la variable ID se encuentra como efecto fijo y aleatorio, esto se debe a que en el caso del Modelo Rasch es considerada como un efecto fijo, mientras que en los MRRL-GLMM es un efecto aleatorio.

Tabla 5.6: Modelos propuestos, variables involucradas y tipo de inferencia utilizada en la estimación de los parámetros

		Efecto		Modelo		
		Fijo	Aleatorio	Rasch	MRRL-GLMM	MRRL-GLMM
Variables Involucradas	ID	✓	✓	✓	✓	✓
	$A1, \dots, A10$	✓		✓	✓	✓
	Sex	✓			✓	✓
	Age.Cat	✓			✓	✓
	Jaundice	✓			✓	✓
Tipo de Inferencia	Clásica			✓	✓	
	Bayesiana					✓

La construcción de los MRRL-GLMM se realiza bajo el siguiente esquema: se parte de la misma estructura que el Modelo de Rasch, es decir, las únicas variables involucradas son ID (utilizada para modelar la variabilidad en cada examinado) y los ítems  $A1, \dots, A10$ ; luego se van incorporando variables explicativas, en forma secuencial, comenzado por Sex, seguido por Age\_cat y finalmente Jaundice. De esta forma se llega a los cuatro modelos candidatos, cuyos parámetros son inferidos por dos caminos distintos.

Haciendo abuso de la notación admitida en R para representar los modelos, y con el fin de fijar ideas, la Tabla 5.7 muestra los modelos que han sido evaluados y el nombre que le hemos asignado a cada uno. La variable  $Y_{ij} \in \{0, 1\}$  indica la respuesta del j-ésimo examinado al i-ésimo ítem, el resto de las variables corresponden a las antes descritas, y en todo los casos logit es la función link.

Al igual que nos sucedió a la hora de seleccionar las variables involucradas en los modelos, la incorporación secuencial en el orden antes dicho no obedece a ninguna recomendación o al seguimiento de resultados antes publicados. Esto hace que sea cuestionable dicho criterio. La falta de herramientas automáticas del estilo stepwise o una recomendación por parte de expertos, nos condujo a esta elección.

En lo que refiere a las distribuciones a priori para los parámetros  $\beta_p$  y  $\sigma_{\epsilon_j}^2$ , seguimos el criterio propuesto por Fox (2010), es decir, prioris no informativas<sup>4</sup>:  $\beta_p \sim N(0, 10)$  y  $\sigma_{\epsilon_j} \sim N_+(0, 20)$ . En el muestreo se utilizaron 4 cadenas, con 2000 iteraciones, 1000 de éstas son de warm-up. En la implementación en R, a través de la biblioteca brms, hemos seguido como referencia a Bürkner (2019).

En todos los casos se ha obtenido convergencia en los algoritmos utilizados para hacer la estimación de los parámetros. El Apéndice G incluye la salida completa de los modelos y el análisis de convergencia para el caso bayesiano. Asociado a este último punto, la validación en el desempeño del modelo a la hora de predecir la variable respuesta, a partir de las prioris utilizadas, se hizo utilizando la proporción de ceros, proporción de unos y evaluando la función entropía.

<sup>4</sup>En el caso de  $\sigma_{\epsilon_j}$  de ha utilizado una normal trucada, con el fin de admitir únicamente valores positivos.

Tabla 5.7: Modelos propuestos, variables involucradas y tipo de inferencia utilizada en la estimación de los parámetros

Nombre	Modelo	Tipo de Inferencia	
m_rasch	M Rasch $Y_{ij} \sim Item + ID$	Clásica	
m1_rasch_glmm	MRRL-GLMM	Clásica	
m2_rasch_glmm			$Y_{ij} \sim -1 + Item_i + (1 ID_j)$
m3_rasch_glmm			$Y_{ij} \sim -1 + Item_i + Sex_j + (1 ID_j)$
m4_rasch_glmm			$Y_{ij} \sim -1 + Item_i + Sex_j + Age.Cat_j + (1 ID_j)$
m1_rasch_B_glmm	MRRL-GLMM	Bayesiana	
m2_rasch_B_glmm			$Y_{ij} \sim -1 + Item_i + (1 ID_j)$
m3_rasch_B_glmm			$Y_{ij} \sim -1 + Item_i + Sex_j + (1 ID_j)$
m4_rasch_B_glmm			$Y_{ij} \sim -1 + Item_i + Sex_j + Age.Cat_j + (1 ID_j)$

## 5.6. Selección de Modelos y Diagnóstico

En la selección de modelos únicamente se encuentran involucrados los modelos denominados MRRL-GLMM, ya que como hemos mencionado al comienzo del presente capítulo, el objetivo es comparar el Modelo de Rasch respecto a sus extensiones representados como un GLMM.

La selección del modelo que mejor se ajuste a los datos se ha realizado, para el caso de inferencia clásica, utilizando los criterios AIC, BIC y el test de razón de verosimilitudes. Estos tres han sido implementados en R haciendo uso de la función `anova` disponible en la biblioteca `base`. En la Tabla 5.8 se puede observar los resultados obtenidos, indicando que, de los cuatro modelos, `m2_rasch_glmm` es el que mejor se ajusta a los datos. Observar que el modelo `m1_rasch_glmm` es tomado como referencia, ya que éste es el más parsimonioso en variables entre los modelos a analizar.

Tabla 5.8: Valores obtenidos de AIC, BIC y Test de Razón de Verosimilitudes correspondientes a los modelos MRRL-GLMM cuyos parámetros han sido estimados utilizando ML a través de la aproximación de Laplace.

Modelo	AIC	BIC	$\chi^2_{df=1}$	$Pr(> Chisq)$
m1_rasch_glmm	12436	12516		
m2_rasch_glmm	12424	12511	13.87	< 0,001
m3_rasch_glmm	12426	12521	0.032	0,857
m4_rasch_glmm	12428	12529	0.86	0,353

En el caso de los modelos bayesianos, se ha utilizado DIC y LOOCV en la selección de modelos. La aplicación del primer criterio se ha realizado a través de una función en R que hemos implementado, mientras que en el segundo se ha utilizado la librería `loo` (Vehtari, Gabry, Magnusson, Yao, Bürkner, Paananen, and Gelman, 2023). La Tabla 5.9 muestra los resultados obtenidos, donde podemos observar que ambos criterios arrojan distintos modelos a elegir, es decir, `m1_rasch_B_glmm` y `m4_rasch_B_glmm` respectivamente. Siguiendo las recomendaciones (Gelman et al., 2021, Chap. 7), optamos por LOOCV para definir cuál es el modelo que mejor se ajusta a los datos.

Tabla 5.9: Valores obtenidos de DIC y LOOCV correspondientes a los modelos MRRL-GLMM cuyos parámetros han sido estimados utilizando inferencia bayesiana.

Modelo	DIC	LOOCV
m1_rasch_B_glmm	17481.83	2.0
m2_rasch_B_glmm	17497.59	0.9
m3_rasch_B_glmm	17502.06	0.9
m4_rasch_B_glmm	17512.17	0.0

La validación en lo que refiere a la significancia de los efectos fijos en el modelo m2\_rasch\_glmm, de acuerdo a las recomendaciones de Faraway (Faraway, 2016, Chap 13) y Fox (Fox and Weisberg, 2019, Sec. 7.3.2), la hemos realizado utilizando el Test de Wald. La Tabla 5.10 muestra los resultados del test, implementado a través de la función Anova disponible en biblioteca car (Fox and Weisberg, 2019). La hipótesis nula corresponde a que el parámetro  $\beta$  asociado a cada variable es igual a cero. Podemos observar que para ambas variables se rechaza la hipótesis nula y en consecuencia son significativas en el modelo.

Tabla 5.10: Resultados del Test de Wald aplicado al modelo m2\_rasch\_glmm.

Modelo	Variable	$\chi^2$	gr. libertad	$Pr(> Chisq)$
m2_rasch_glmm	Item	284.41	10	< 0,01
	Sex	14.03	1	< 0,01

Nos resta validar la hipótesis de normalidad en los residuos para el modelo m2\_rasch\_glmm. En esta dirección, siguiendo a Faraway (Faraway, 2016, Chap 13), realizamos un análisis gráfico (ver resultados en Apéndice G) y además los test de Kolmogorov Smirnov y Jarque Bera. El primer test asume como hipótesis nula que la distribución de los residuos es normal, mientras que el segundo asume que la asimetría y kurtosis de los residuos no son significativamente diferentes a los correspondientes en una distribución normal.

Tabla 5.11: Resultados del Test de Kolmogorov Smirnov y Jarque Bera aplicado a los residuos del modelo m2\_rasch\_glmm.

Modelo	T. Kolmogorov Smirnov		T. Jarque Bera	
	D	$p - valor$	$\chi^2_{df=2}$	$p - valor$
m2_rasch_glmm	0.163	< 0,001	867.72	< 0,01

La Tabla 5.11 muestra los resultados arrojados por el test de Kolmogorov Smirnov y Jarque Bera. En ambos casos se rechaza la hipótesis nula, por lo tanto, los residuos no cumplen la hipótesis de normalidad.

Bates et al. (2015) advierte acerca de los objetos disponibles, al utilizar la librería lmer, a la hora de realizar el diagnóstico de modelos. En este sentido, describe una alternativa a las herramientas gráficas, que corresponde a una adaptación del método posterior predictive simulation. La idea

consiste en computar alguna medida de resumen para un conjunto de simulaciones y analizar qué modelo representa mejor los datos.

Siguiendo dicha recomendación, realizamos 1000 simulaciones de los modelos  $m\{1,2,3,4\}$ \_rasch\_glmm y computamos para cada una de éstas la proporción de ceros en las respuestas estimadas. La elección de dicho estadístico se debe a que existe menor cantidad de ceros en las observaciones. Además, teniendo en cuenta que dicha proporción puede variar según si existe (o no) diagnóstico de TEA, decidimos analizar el resultado de la simulación dentro de cada clase de la variable Class.ASD.Traits.

La Figura 5.2 muestra los resultados obtenidos subdivididos de acuerdo al resultado del test acerca del rasgo de TEA. En ésta se puede observar la distribución de la proporción de ceros para los distintos modelos. Una inspección visual nos indicaría que los modelos  $m1$ \_rasch\_glmm y  $m2$ \_rasch\_glmm podrían ser buenos candidatos para representar los datos. Nuestra decisión ha sido seleccionar el modelo  $m2$ \_rasch\_glmm en sintonía con el test de razón de verosimilitudes.

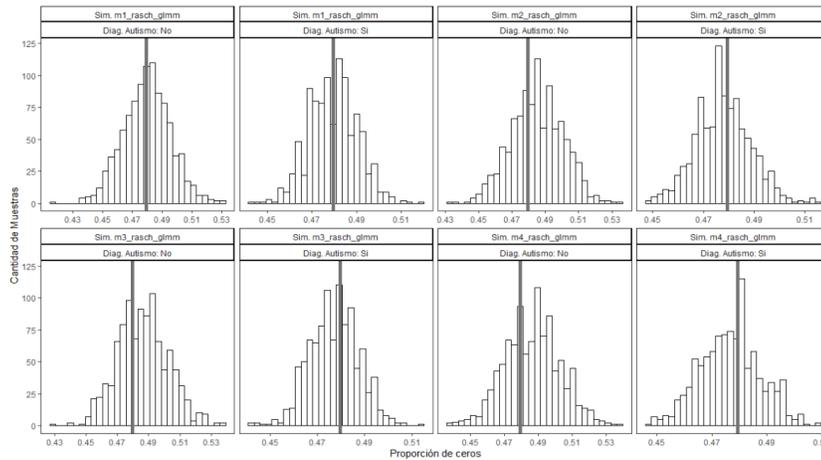


Figura 5.2: Adaptación del método posterior predictive simulation aplicado a los modelos  $m\{1,2,3,4\}$ \_rasch\_glmm.

En resumen, al utilizar inferencia clásica e inferencia bayesiana en los modelos MRRL-GLMM, los que mejor se ajustan a los datos son  $m2$ \_rasch\_glmm y  $m4$ \_rasch\_B\_glmm respectivamente. En el caso del modelo  $m2$ \_rasch\_glmm existe un incumplimiento en una de las hipótesis sobre las cuales se construye la inferencia. Sin embargo, los resultados obtenidos a partir de la adaptación del método posterior predictive simulation, nos muestran que puede ser un buen candidato a incluir en el análisis.

## 5.7. Resultados

A continuación estaremos presentando los resultados obtenidos en las estimaciones de los parámetros correspondientes a los modelos: `m_rasch`, `m2_rasch_glm` y `m4_rasch_B_glm`, es decir, aquellos que mejor se han ajustado a los datos. Como hemos mencionado, en los dos primeros modelos se ha utilizado inferencia clásica, en particular los algoritmos CML y ML a través de la aproximación de Laplace, mientras que en el tercero se ha utilizado inferencia bayesiana.

En la Tabla 5.12 se pueden observar las estimaciones para cada modelo, así como también el p-valor en el caso de `m_rasch`, `m2_rasch_glm`. Podemos señalar que en el caso de los ítems, algunos de éstos no resultaron significativos a nivel 5 % para el modelo. En el caso del modelo `m4_rasch_B_glm`, se observa que el intervalo de credibilidad al 95 % asociado a algunos ítems y la variable `Age_cat L2` incluyen el valor cero, es decir, su estimación podría tomar el valor nulo.

Tabla 5.12: Estimación de los parámetros en los modelos seleccionados.

	Modelo <code>m_rasch</code>			Modelo <code>m2_rasch_glm</code>			Modelo <code>m4_rasch_B_glm</code>		
	Estimación	z	p - valor	Estimación	z	Pr(>  z )	Estimación	Est. Error	$\hat{R}$
$\beta_{A1}$	0.217	0.204	0.838	0.049	0.433	0.665	0.07	0.14	1
$\beta_{A2}$	-0.405	0.607	0.544	-0.573	-5.039	< 0,01	-0.56	0.14	1
$\beta_{A3}$	-0.663	-0.190	0.850	-0.831	-7.26	< 0,01	-0.82	0.14	1
$\beta_{A4}$	-0.038	-1.587	0.113	-0.206	-1.816	0.069	-0.19	0.14	1
$\beta_{A5}$	0.021	-1.628	0.103	-0.146	-1.29	0.197	-0.13	0.14	1
$\beta_{A6}$	0.294	-3.339	< 0,01	0.126	1.107	0.268	0.14	0.14	1
$\beta_{A7}$	0.759	-1.806	0.071	0.589	5.119	< 0,01	0.61	0.14	1
$\beta_{A8}$	-0.361	2.325	< 0,05	-0.529	-4.661	< 0,01	-0.51	0.14	1
$\beta_{A9}$	-0.173	-3.356	< 0,01	-0.340	-3.004	< 0,01	-0.32	0.14	1
$\beta_{A10}$	0.349	10.691	< 0,01	0.181	1.591	0.111	0.20	0.14	1
$\beta_{\text{Sex M}}$				0.404	3.745	< 0,01	0.41	0.11	1
$\beta_{\text{Age.cat L2}}$							-0.02	0.11	1
$\beta_{\text{Jaundice Yes}}$							0.11	0.11	1

En el gráfico de la Figura 5.3 se muestran las estimaciones y los respectivos intervalos de confianza y credibilidad al 95 % para los tres modelos. También hemos incluido, siguiendo la recomendación de Bates et al. (2015), el intervalo de confianza para el modelo `m2_rasch_glm` utilizando la técnica de bootstrap.

En los casos donde los modelos cuentan con variables en común, existe un solapamiento en los intervalos, en consecuencia no se obtiene una mejora en la precisión de las estimaciones. Sumado a esto, a pesar que el modelo `m2_rasch_glm` viola la hipótesis de normalidad, se observa que los intervalos de confianza obtenidos se aproximan a los respectivos mediante estimación bayesiana.

Respecto al rasgo latente de los examinados, la Figura 5.4 muestra las estimaciones y los respectivos intervalos de confianza y credibilidad al 95 %. Observar que en el modelo `m_rasch` se tiene menos valores estimados, esto es consecuencia de las características de CML, donde el score total obtenido por los examinados (estadístico suficiente) es utilizado en el proceso de estimación. Mientras que en los modelos `m2_rasch_glm` y `m4_rasch_B_glm`, a diferencia de éste, las técnicas utilizadas en la estimación de parámetros permite lograr mayor diferenciación entre participantes.

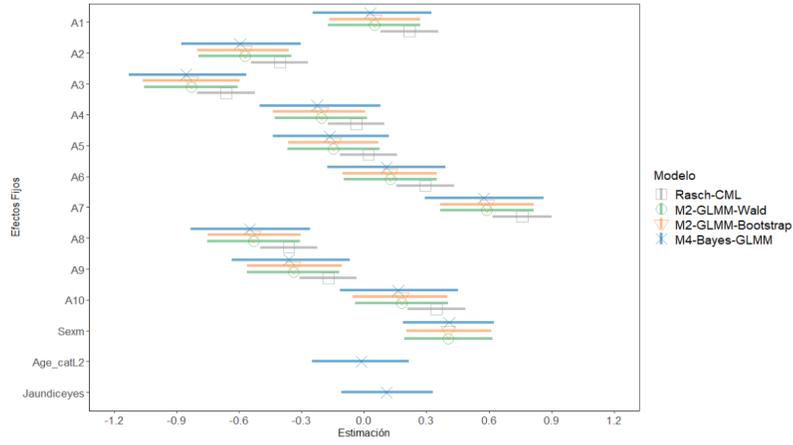
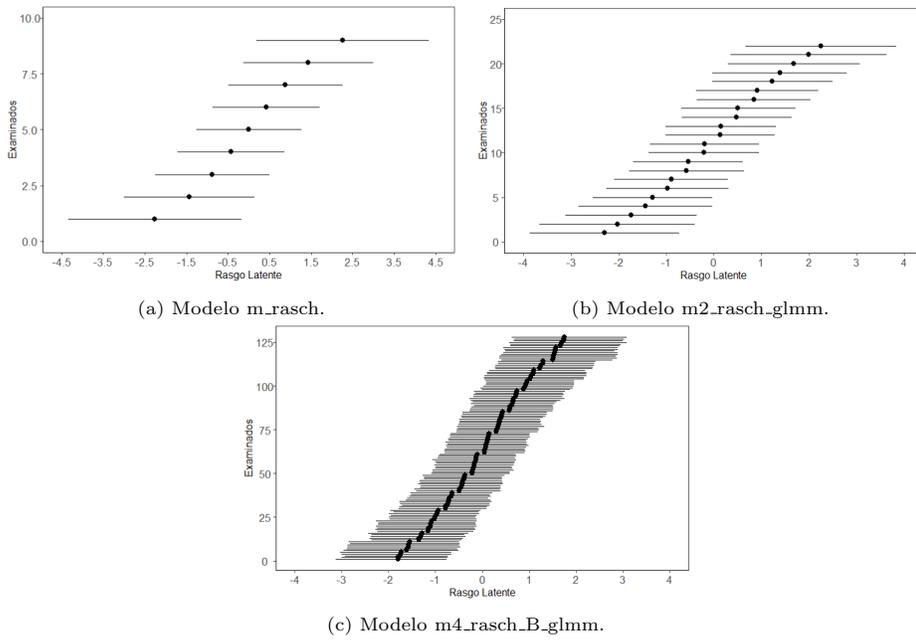


Figura 5.3: Cantidad de respuestas correctas, según el diagnóstico recibido, agrupadas por Sexo y categoría de edad.



(a) Modelo m\_rasch.

(b) Modelo m2\_rasch\_glmm.

(c) Modelo m4\_rasch\_B\_glmm.

Figura 5.4: Estimación de los parámetros asociado a los examinados.

## 5.8. Evaluación de los Modelos

La bondad en el ajuste de los modelos es evaluada como es habitual en los problemas de clasificación, es decir, computando<sup>5</sup> precisión, sensibilidad, especificidad y el área bajo la curva ROC (AUC). Éstas resultan ser herramientas de gran utilidad a la hora de comprar la predicción del modelo respecto a lo que se observa a partir de los datos.

Tabla 5.13: Evaluación de la capacidad predictiva de los modelos.

Modelo	Precisión	Sensibilidad	Especificidad	AUC
m_rasch	0.728	0.748	0.708	0.802
m2_rasch_glmm	0.764	0.762	0.765	0.846
m4_rasch_B_glmm	0.763	0.763	0.765	0.846

En la tabla 5.13 se muestran los resultados obtenidos. A partir de éstos se puede observar una mejora en la performance al utilizar los modelos MRRL-GLMM. Sin embargo, no existe diferencias notorias entre los modelos m2\_rasch\_glmm y m4\_rasch\_B\_glmm. Por último, señalar que en el Apéndice se han incluido las curvas ROC para cada modelo.

## 5.9. Observaciones finales

En el presente capítulo hemos aplicado a un problema real los resultados presentados a lo largo del TFG. El modelo de Rasch, desde una perspectiva clásica y bayesina, ha sido analizado, arrojando resultados que muestran su potencial en la Teoría de los Test.

Un aspecto a subrayar es el siguiente, durante el estudio de las bibliotecas disponibles en R para llevar adelante la implementación de los modelos, nos encontramos que en la mayoría de los casos era necesario realizar un desarrollo adicional que permitiera extraer los resultados finales. Sumado a esto, al analizar cómo han sido implementadas las funciones, en particular en la biblioteca eRm, encontramos escasa flexibilidad cuando se quiere integrar, por ejemplo, con Tydeverse.

El siguiente capítulo ha sido dedicado a exponer las conclusiones a partir de los resultados aquí obtenidos, así como también los trabajos a futuro identificados en el TFG.

---

<sup>5</sup>Ver definición en Florkowski (2008).

## Capítulo 6

# Conclusiones

## 6.1. Conclusiones

A partir de los resultados obtenidos, en primer lugar, podemos concluir que el abordaje del Modelo de Rasch a través de los GLMM permite incorporar la variabilidad inherente entre individuos. Además, es posible utilizar un conjunto de herramientas, teóricas y de software, a la hora de realizar el análisis que ofrece gran flexibilidad.

En segundo lugar, el abordaje desde la óptica de la TRI, a los datos de “Autistic Spectrum Disorder Screening Data for Toddlers”, permite un análisis e interpretación de resultados que no se logra a través de los algoritmos de machine learning. Estos últimos han sido, hasta ahora, el conjunto de herramientas utilizadas en el estudio de las respuestas al test.

En tercer lugar, al utilizar los GLMM se obtiene una mejora en la performance, respecto al Modelo de Rasch, con un costo computacional que no ha resultado significativo. Además, abre las puertas a incorporar mayor complejidad al modelo, por ejemplo, a través de los modelados multinivel.

En cuarto lugar destacamos las ventajas que presentan los GLMM a nivel de software. En la actualidad, existe un importante número de bibliotecas pensadas con integración al nuevo paradigma de desarrollo en R, como lo es `tydiverse` y `shiny`. De acuerdo a nuestro análisis, las bibliotecas originalmente desarrolladas para el Modelo de Rasch y sus extensiones, carecen de integración sencilla con dichas bibliotecas e incluso con `rstan`. Sin embargo, las bibliotecas `lmer` y `brms` cuentan con compatibilidad al mundo de `tydiverse`, donde hay una comunidad muy activa desarrollando nuevas funciones. En esta dirección destacamos la biblioteca `ShinyItemAnalysis` Martinkova and Drabinova (2018).

Finalmente, debemos señalar con respecto a la precisión en la estimación de los parámetros del modelo (efectos fijos), no se logra una ventajas sustancial a través de los algoritmos utilizados en los GLMM. A pesar de ello, no debemos perder de vista la flexibilidad que estos últimos presentan en el modelado, principalmente a la hora de introducir efectos asociados a la autocorrelación, por ejemplo, entre observaciones o entre ítems.

Una debilidad que queremos señalar está asociada al tipo de datos con los que estamos trabajando. De acuerdo a la descripción aportada por Thabtah (2020), no se ha realizado un diseño de muestreo para seleccionar a los participantes del test. Esto hace que, a partir de los resultados obtenidos, no sea posible realizar afirmaciones sobre los ítems o las habilidades de los examinados, debido al sesgo de selección.

Otra de las debilidades que visualizamos está asociada a la predicción de nuevas observaciones. Los GLMM funcionan muy bien cuando la aproximación al problema es mediante un modelo explicativo. Sin embargo, si lo que se busca es un modelado predictivo, existe la restricción a la hora de predecir efectos aleatorios. Rizopoulos (2018a; 2018b) describe este problema y expone algunas alternativas.

A lo anterior sumamos aspectos de performance. Si comparamos los modelos aquí propuestos con respecto a los algoritmos de machine learning Al-diabat (2018); Bala et al. (2022); Erkan and Thanh (2019); Jannat and Canavan (2021); Jayakumaran and Sweetlin (2020); Raj and Masood (2020); Saihi and Alshraideh (2021); Singh et al. (2021); Thabtah (2017c); Thabtah et al. (2018,1), en algunos casos obtenemos un resultado similar y en otros casos una performance menor.

## 6.2. Trabajos a Futuro

En lo que refiere a mejoras y trabajos a futuro, creemos que el primer paso debería ser abordar el análisis de las respuestas al test con la colaboración de investigadores en el área del TEA. Esto permitirá tener un mejor entendimiento del experimento y sus características, así como también mejorar los modelos que han sido propuestos.

Respecto a la capacidad de predecir nuevas observaciones, creemos que resta trabajo por realizar, y abre un abanico de posibilidades a nivel teórico y en el desarrollo de bibliotecas en R.

El siguiente paso natural es extender los modelos aquí propuestos a través de modelos 2PL, es decir, tomar en consideración la dificultad del ítem en el modelado. En este sentido, también se puede recurrir a modelar dicho parámetro a través de una variable latente que involucre alguna característica del test y/o del examinado.

Asociado a los puntos anteriores, y teniendo en cuenta la realización del censo 2023 en Uruguay, el cual por primera vez se realizará un relevamiento de la población diagnosticada con TEA, se abren nuevas oportunidades para investigar el TEA en nuestro país. Por ejemplo, haciendo uso de la herramienta desarrollada por Thabtah (2017b), en una muestra probabilística y aplicando los modelos aquí desarrollados, así como también extensiones a éstos o nuevas propuestas.

Finalmente una pregunta que nos realizamos es la siguiente: ¿pueden combinarse modelos de la TRI y algoritmos de machine learning, por ejemplo, mediante la técnica de stacking? ¿qué beneficios traería consigo?

Esta página ha sido intencionalmente dejada en blanco.

# Apéndices

Esta página ha sido intencionalmente dejada en blanco.

# Apéndice A

## Resultados Generales

En el desarrollo de las definiciones y los resultados que se presentan a continuación se ha tomado como referencia a los siguientes autores: Barndorff-Nielsen (2014), Boente and Yohai, Casella and Berger (2002), Gómez and Delicado (2006), Gómez Villegas (2005), Notas del curso de Inferencia Estadística IESTA (2020), Martin (2017), Mukhopadhyay (2000), Olive (2014), Olsson (2002), Perera (2011), Pintarelli (2010), Rice (2007), Wasserman (2004).

### A.1. Familia de Distribuciones

**Definición 1.** Se define una familia de distribuciones como el conjunto de funciones de distribución definidas sobre el mismo espacio probabilístico:

$$\mathcal{F} = \{F_x(\cdot|\theta) : \theta \in \Theta \subset \mathbb{R}^k\} \quad (\text{A.1})$$

A  $\Theta$  se lo denomina espacio paramétrico.

### A.2. Familia Exponencial

**Definición 2.** Una familia de distribuciones se dice que es una familia exponencial (o familia exponencial natural) de  $k$  parámetros si su función de densidad puede factorizarse de la siguiente manera:

$$f(x|\theta) = h(x) \cdot c(\theta) \cdot e^{\sum_{i=1}^k w_i(\theta) \cdot t_i(x)}, \quad \theta \in \Theta, \quad \forall c(\theta) \geq 0, \quad h(x) \geq 0 \quad (\text{A.2})$$

Siendo  $c(\theta)$ ,  $w_i(\theta)$ ,  $h(x)$ ,  $t_i(x)$  funciones reales que dependen únicamente de  $\theta$  y  $x$  respectivamente. Observar que la parametrización no es única, ya que multiplicando a  $w_i$  por una constante y dividiendo  $t_i$  por la misma constante se llega a la misma factorización.

Muchas de las distribuciones comúnmente utilizadas pertenecen a la familia exponencial, algunos ejemplos son: beta, binomial, gamma, normal. Además, un resultado importante que se puede

probar en esta familia es acerca de la suficiencia, ya que  $[t_1(x), t_2(x), \dots, t_k(x)]^T$  son estadísticos suficientes<sup>1</sup>.

Si  $X_1, X_2, \dots, X_n$  son una muestra aleatoria de  $X \sim f_X(x|\theta)$  e independientes entre sí y  $f_X(x|\theta)$  pertenece a la familia exponencial, entonces la función se puede escribir de la siguiente manera:

$$L(\theta|\vec{x}) = \left( \prod_{i=1}^n h(x_i) \right) \cdot (c(\theta))^n \cdot e^{\sum_{j=1}^k w_j(\theta) \cdot (\sum_{i=1}^n t_j(x_i))} \quad (\text{A.3})$$

Supongamos que  $k = 1$ , en forma alternativa  $f(x|\theta)$  puede reescribirse de la siguiente manera:

$$f(x|\theta) = e^{\frac{x\theta - b(\theta)}{a(\phi)} + c(x, \phi)}, \quad \theta \in \Theta. \quad (\text{A.4})$$

Donde al parámetro  $\theta$  se lo denomina parámetro natural, mientras que a  $\phi$  se lo llama parámetro de dispersión. Observar que cuando  $a(\phi) = 1$  y  $c(x, \phi)$  se tiene que  $f(x|\theta) = h(x)e^{x\theta - b(\theta)}$ , la cual corresponde a la familia exponencial natural.

### A.3. Función de Verosimilitud

**Definición 3.** Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de  $X \sim f_X(x|\theta)$ ,  $\theta = (\theta_1, \theta_2, \dots, \theta_k)^T \in \Theta \subseteq \mathbb{R}^k$ ,  $\mathbf{x} = \vec{x} = (x_1, x_2, \dots, x_n) \in \mathcal{X}$  (espacio muestral). La función de verosimilitud (likelihood function)  $L(\theta|\vec{x}) : \Theta \rightarrow \mathbb{R}^+$  se define como:

$$L(\theta|\vec{x}) = f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n|\theta) \quad (\text{A.5})$$

La función  $L(\theta|\vec{x})$  mide la evidencia contenida en la muestra ( $\vec{x}$ ) sobre los posibles valores que el parámetro  $\theta$  puede tomar.

Observar que cuando  $X_1, X_2, \dots, X_n$  son independientes, la función de verosimilitud es:

$$L(\theta|\vec{x}) = \prod_{i=1}^n f_{X_i}(x_i|\theta) \quad (\text{A.6})$$

En muchas oportunidades resulta de utilidad trabajar con el logaritmo de la función de verosimilitud, que bajo el supuesto de independencia puede expresarse de la siguiente manera:

$$l(\vec{\theta}|\vec{x}) = \log(L(\vec{\theta}|\vec{x})) = \sum_{i=1}^n f_{X_i}(x_i|\theta) \quad (\text{A.7})$$

Esta expresión suele ser muy utilizada a la hora de realizar la estimación de los parámetros ( $\hat{\theta}$ ) de  $f(x|\theta)$  mediante el método de máxima verosimilitud, ya que la derivada de  $l(\theta|\vec{x})$  presenta una expresión algebraica más sencilla respecto a  $L(\theta|\vec{x})$ . Observemos que:

<sup>1</sup>Los estadísticos suficientes son definidos más adelante en el presente Apéndice.

$$\frac{\partial}{\partial \theta_j} l(\boldsymbol{\theta}|\vec{x}) = \sum_{i=1}^n \frac{\partial}{\partial \theta_j} f_{x_i}(x_i|\boldsymbol{\theta}), \forall j = 1, 2, \dots, k. \quad (\text{A.8})$$

De esta forma se obtiene el sistema de ecuaciones que permite obtener  $\hat{\boldsymbol{\theta}}$ , es decir, la estimación del vector de parámetros de interés.

## A.4. Condiciones de Regularidad FI

**Definición 4.** Sea  $X$  una VA tal que  $f_x(x|\boldsymbol{\theta})$  es su función de densidad,  $\boldsymbol{\theta} \in \mathbb{R}^k$ , entonces se define las condiciones de regularidad de la Información de Fisher (FI regularity conditions) de la siguiente manera:

R1.  $\frac{\partial}{\partial \theta_j} f(x|\boldsymbol{\theta}) < \infty \forall x \in \mathcal{X}, \forall \theta_j, j = 1, 2, \dots, k.$

R2.  $\int f(x|\boldsymbol{\theta}) d\mu(x)$  es derivable respecto a  $\theta_j$  ( $j = 1, 2, \dots, k$ ) y los operadores  $\int, d/d\theta_j$  pueden intercambiarse.

R3. Soporte fijo:  $\{x : f(x|\boldsymbol{\theta}) > 0\} \forall \boldsymbol{\theta} \in \Theta$ , es decir, el recorrido de  $X$  no depende de  $\boldsymbol{\theta}$ .

## A.5. Función Score de Fisher

**Definición 5.** Sea  $\boldsymbol{\theta} \in \mathbb{R}^k$  y  $X$  una VA cuya función de densidad es  $f_x(x|\boldsymbol{\theta})$ , la función Score de Fisher se define como:

$$S(\boldsymbol{\theta}) = \left( \frac{\partial}{\partial \theta_1} \log(f(\boldsymbol{\theta}|X)), \frac{\partial}{\partial \theta_2} \log(f(\boldsymbol{\theta}|X)), \dots, \frac{\partial}{\partial \theta_k} \log(f(\boldsymbol{\theta}|X)) \right) \quad (\text{A.9})$$

Tomando en cuenta que el logaritmo es una función monótona creciente, entonces cada componente de la función  $S(\boldsymbol{\theta})$  proporciona una medida de la sensibilidad de  $f_x(x|\boldsymbol{\theta})$  a cambios en  $\theta_j, j = 1, 2, \dots, k$ . En el caso unidimensional, evaluada  $S(\theta)$  en un punto  $\theta_0$ , su valor corresponde a la pendiente de la tangente a la curva en dicho punto.

Por comodidad reescribiremos el vector  $S(\boldsymbol{\theta})$  de la siguiente manera:

$$S(\boldsymbol{\theta}) = (s_1(\boldsymbol{\theta}), s_2(\boldsymbol{\theta}), \dots, s_k(\boldsymbol{\theta})) \quad (\text{A.10})$$

Una característica interesante, que resultará de gran utilidad, es el hecho que el valor esperado de la función score es igual a cero.

## A.6. Función Información de Fisher

**Definición 6.** La matriz de Información de Fisher acerca de  $\Theta$  basada en  $X$  ( $I_{\bar{x}}(\theta)$  de  $k \times k$ ), bajo las condiciones de regularidad R1, R2, R3, se define como:

$$I_x(\theta) = E_{\theta}[\nabla_{\theta} \log(f_x(X|\theta)) \cdot \nabla_{\theta}^T \log(f_x(X|\theta))] = Cov(\nabla_{\theta} \log(f_x(X|\theta))) \quad (\text{A.11})$$

donde  $E_{\theta}$  denota el valor esperado tomado con respecto a  $f(X|\theta)$ , es decir,  $\theta$  es un parámetro;  $\nabla_{\theta}$  es el gradiente respecto a las componentes de  $\theta$ ; la matriz de covarianza es simétrica y semidefinida positiva.

A partir de esta definición tenemos que los elementos  $i, j = 1, 2, \dots, k$ , de la matriz de Información de Fisher son:

$$(I_x(\theta))_{i,j} = Cov\left(\frac{\partial}{\partial \theta_i} \log(f_x(X|\theta)), \frac{\partial}{\partial \theta_j} \log(f_x(X|\theta))\right)$$

Alternativamente, debido a que se cumplen las condiciones de regularidad, tenemos que una expresión alternativa es:

$$I_x(\theta) = E(S(\theta) \cdot S^T(\theta)) = -E\left[\frac{\partial}{\partial \theta \partial \theta^T} S^T(\theta)\right] = -E\left[\frac{\partial^2}{\partial \theta \partial \theta^T} \log(f_x(X|\theta))\right] \quad (\text{A.12})$$

O su equivalente en los elementos  $i, j$  de la matriz de Información de Fisher:

$$(I_x(\theta))_{i,j} = -E_{\theta}\left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(f_x(X|\theta))\right] \quad (\text{A.13})$$

El estimador máximo verosímil  $\hat{\theta}$  del parámetro  $\theta$  presenta una distribución asintóticamente normal multivariada, cuya media es  $\theta$  y matriz de varianza-covarianza  $(I_{\bar{x}}(\theta))^{-1} = n(I_x(\theta))^{-1}$ .

En el caso particular en que  $\theta \in \mathbb{R}$  deja de tener sentido hablar de matriz, sino que simplemente se define la Información de Fisher de la siguiente manera:

$$I_{\bar{x}}(\theta) = E_{\theta}\left[\frac{\partial}{\partial \theta^2} \log(f_x(X|\theta))\right] \quad (\text{A.14})$$

En el desarrollo hemos utilizado  $E_{\theta}$  para indicar que el valor esperado se toma con respecto a  $f_x(\theta|X)$ , es decir,  $\theta$  es la VA en consideración.

## A.7. Métodos Numéricos

En muchos casos maximizar la función de verosimilitud, lo que equivale a resolver el sistema de ecuaciones dado por las funciones Score de Fisher, implica resolver un sistema de ecuaciones no lineal. Ante esta situación es necesario recurrir a métodos numéricos desarrollados para resolver el problema de optimización bajo ciertas condiciones de regularidad.

A continuación se describe brevemente dos de los métodos frecuentemente utilizados, estos son el método de Newton-Raphson y el Método de Socring de Fisher.

### A.7.0.1. Algoritmo de Newton-Raphson

Supongamos que el objetivo es maximizar la función  $l(\boldsymbol{\theta}|\mathbf{x})$  respecto al vector de parámetros  $\boldsymbol{\theta}$ , es decir, buscamos  $\hat{\boldsymbol{\theta}}$  tal que el logaritmo de la función de verosimilitud sea máxima. Sea  $l'$  el vector cuyas componentes corresponde a la primer derivada en las componente del vector de parámetros. Además, llamaremos  $\mathbf{H}$  a la matriz Hessiana.

El método de Newton-Raphson consiste en realizar el desarrollo de Tylor para la función  $l'(\boldsymbol{\theta}|\mathbf{x})$ . Partiendo de un valor inicial  $\boldsymbol{\theta}_0$  tenemos que:

$$l'(\boldsymbol{\theta}_1|\mathbf{x}) = l'(\boldsymbol{\theta}_0|\mathbf{x}) + (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0) \cdot \mathbf{H}(\boldsymbol{\theta}_0) \quad (\text{A.15})$$

Si  $\boldsymbol{\theta}_1$  el el vector que maximiza a la función  $l(\boldsymbol{\theta}|\mathbf{x})$ , entonces  $l'(\boldsymbol{\theta}_1|\mathbf{x}) = 0$ , por lo tanto, a partir de la expresión anterior se obtiene:

$$\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0 - l'(\boldsymbol{\theta}_0|\mathbf{x}) \cdot \mathbf{H}^{-1}(\boldsymbol{\theta}_0) \quad (\text{A.16})$$

Si  $|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0| < \epsilon$ , entonces  $\boldsymbol{\theta}_1$  es el valor que optimiza a la función de verosimilitud, de lo contrario se repite el proceso utilizando a  $\boldsymbol{\theta}_1$  como valor inicial. En el paso  $k + 1$  tendremos que:

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - l'(\boldsymbol{\theta}_k|\mathbf{x}) \cdot \mathbf{H}^{-1}(\boldsymbol{\theta}_k) \quad (\text{A.17})$$

De forma iterativa continúa el proceso hasta converger a la solución, la cual esta garantizada cuando se parte de un valor inicial próximo a  $\hat{\boldsymbol{\theta}}$ .

Si  $f(\mathbf{x}|\boldsymbol{\theta})$  es suficientemente regular, entonces  $l(\boldsymbol{\theta}|\mathbf{x})$  será aproximadamente cuadrático a medida que crece el tamaño de la muestra, por lo tanto, el algoritmo convergerá. Mientras que en otros casos el algoritmo podría no converger.

### A.7.0.2. Método de Scoring de Fisher

El método de Scoring de Fisher es una variante a Newton-Raphson, donde la matriz Hessiana  $\mathbf{H}(\boldsymbol{\theta})$  es sustituida por su valor esperado. Dicha esperanza coincide con la matriz de información de Fisher a menos de un signo, esto es:

$$E(\mathbf{H}(\boldsymbol{\theta})) = -\mathbf{I}(\boldsymbol{\theta}) \quad (\text{A.18})$$

Utilizar el valor esperado en lugar de la matriz Hessiana presenta la ventaja que sólo es necesario calcular derivadas primeras. Esto se debe a que es posible probar el siguiente resultado:

$$E\left(\frac{\partial^2 l(\boldsymbol{\theta}|\mathbf{x})}{\partial\theta_i\partial\theta_j}\right) = -E\left(\frac{\partial l(\boldsymbol{\theta}|\mathbf{x})}{\partial\theta_i} \frac{\partial l(\boldsymbol{\theta}|\mathbf{x})}{\partial\theta_j}\right) \quad (\text{A.19})$$

Otra ventaja radica en el hecho que el valor esperado de la matriz Hessiana es siempre una matriz definida positiva, por lo tanto, no está presente el problema de no convergencia como en el caso del método de Newton-Raphson. Sin embargo, las bondades del método de Scoring de Fisher no son gratuitas, y el costo que hay que pagar a cambio se encuentra en mayor tiempo de convergencia.

Por último, es oportuno señalar que cuando la función de densidad pertenece a la familia exponencial, el método de Scoring de Fisher y el método de Newton-Raphson son equivalentes.

## A.8. Estadísticos Suficientes

A continuación se describe algunos de los resultados más relevantes en lo que refiere a los estadísticos suficientes.

### Teorema 2. Teorema de Factorización.

Sea  $L(\boldsymbol{\theta}|\mathbf{x})$  un función de verosimilitud,  $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ . Un estadístico real  $T = T(X_1, X_2, \dots, X_n)$  es suficiente para  $\boldsymbol{\theta}$  si y sólo si:

$$L(\boldsymbol{\theta}|\mathbf{x}) = g(T(\mathbf{x}); \boldsymbol{\theta}) \cdot h(\mathbf{x}), \forall \mathbf{x} \in \mathcal{X}. \quad (\text{A.20})$$

*Demostración.* ver Chap. 6, Teorema 6.2.6 Casella and Berger (2002).

### Teorema 3. Sea $L(\boldsymbol{\theta}|\mathbf{x})$ una función de verosimilitud.

Un estadístico  $T = \vec{T}(X_1, X_2, \dots, X_n) = (T_1(X_1, X_2, \dots, X_n), T_2(X_1, X_2, \dots, X_n), \dots, T_k(X_1, X_2, \dots, X_n))$  de componentes reales es conjuntamente suficiente para  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k) \in \Theta \subset \mathbb{R}^k$  si y sólo si

$$L(\boldsymbol{\theta}|\mathbf{x}) = g(\vec{T}(X_1, X_2, \dots, X_n); \boldsymbol{\theta}) \cdot h(\mathbf{x}), \forall \mathbf{x} \in \mathcal{X}. \quad (\text{A.21})$$

Ver. Cap. 3, Teorema 3.3.3 IESTA (2020) y Lecture 12 Panchenko (2003).

Observación: si  $T_1(X_1, X_2, \dots, X_n), T_2(X_1, X_2, \dots, X_n), \dots, T_k(X_1, X_2, \dots, X_n)$  es conjuntamente suficiente para  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ , esto no implica que  $T_2(X_1, X_2, \dots, X_n)$  es suficiente para  $\theta_i \forall i = 1, 2, \dots, k$ .

**Teorema 4. Cota inferior de Cramer-Rao.** Sea  $\vec{X} = (X_1, X_2, \dots, X_n)$  un vector aleatorio cuya función de densidad conjunta  $f_{\vec{X}}(\mathbf{x}|\theta)$ ,  $\theta \in \Theta \subseteq \mathbb{R}$ . Sea  $T(X_1, X_2, \dots, X_n)$  un estimador insesgado de  $\tau(\theta)$  tal que  $E(T(X_1, X_2, \dots, X_n)) = \tau(\theta)$  con  $\tau(\theta)$  una función diferenciable en  $\theta$ .

Si  $f_{\vec{X}}(\mathbf{x}|\theta)$  satisface las condiciones de Regularidad FI, entonces:

$$\text{Var}(T(X_1, X_2, \dots, X_n)) \geq \frac{d^2\tau(\theta)}{d\theta^2} \frac{1}{I_{\vec{X}}(\theta)} = \frac{d^2\tau(\theta)}{d\theta^2} \frac{1}{nI_{X_1}(\theta)} \quad (\text{A.22})$$

*Demostración.* ver Chap. 2, Theorem 2.9 Martin (2017).

Observación: la cota de Cramer-Rao es alcanzable si y sólo si  $f_{\vec{X}}(\mathbf{x}|\theta)$  pertenece a la familia exponencial.

Esta página ha sido intencionalmente dejada en blanco.

## Apéndice B

# Modelos Logísticos Dicotómicos - Estimación de los Parámetros

### B.1. Método de Máxima Verosimilitud - Estimación de los Parámetros de Discriminación y Dificultad en el Modelo 2PL

En el presente Apéndice analizaremos el método de Máxima Verosimilitud (Maximum Likelihood, ML), aplicado al modelo 2PL, con el objetivo de estimar el parámetro de discriminación y dificultad de cada ítem en el test. El desarrollo que sigue a continuación está basado principalmente en los trabajos de Baker and Kim (2004), Bartolucci et al. (2015), Cai and Thissen (2014), Fischer and Molenaar (1995), Melià (2003), Santisteban Requena and Alvarado (2001).

Supongamos que un test está formado por  $n$  ítems dicotómicos y nos interesa estimar los parámetros del modelo 2PL para el  $i$ -ésimo ítem,  $i = 1, \dots, n$ . Además, vamos a suponer que los participantes están agrupados en  $J$  grupos de acuerdo al nivel de rasgo latente (conocido)  $\theta_j$ ,  $j = 1, \dots, J$ . En cada grupo hay  $f_j$  individuos y denotaremos con  $r_j$  a la cantidad de respuestas correctas que han proporcionado, por lo tanto,  $f_j - r_j$  es la cantidad de respuestas incorrectas. Por último, llamaremos  $R = (r_1, r_2, \dots, r_n)$  al vector de respuestas correctas observadas.

Sea  $p_j$  la proporción de respuestas correctas dado por:

$$p_j = p(\theta_j) = \frac{r_j}{f_j} \quad (\text{B.1})$$

Por lo tanto, la proporción de respuestas incorrectas es:

$$q_j = q(\theta_j) = \frac{f_j - r_j}{f_j} \quad (\text{B.2})$$

Bajo la hipótesis que las respuestas de los examinados son independientes, podemos suponer que  $r_j \sim \text{Bin}(f_j, p_j)$ . Haciendo uso de Ec. 2.18 tenemos que la probabilidad de observar en el  $i$ -ésimo ítem,  $i \in \{1, 2, \dots, n\}$ , el vector  $R$  de respuestas correctas es:

$$P(Y = R) = L = \prod_{j=1}^J \binom{f_j}{r_j} P_j^{r_j} Q_j^{f_j - r_j} \quad (\text{B.3})$$

Donde, de acuerdo a la notación utilizada en el Capítulo 2,  $P_j = P_i(\theta_j)$ ,  $Q_j = Q_i(\theta_j)$  y  $L$  corresponde a la distribución conjunta o función de verosimilitud, a partir de ésta buscamos los parámetros  $a_i, b_i$  que maximicen  $L$ , lo que equivale a maximizar su logaritmo.

$$l = \ln(L) = \sum_{j=1}^J \binom{f_j}{r_j} + \sum_{j=1}^J r_j \ln(P_j) + \sum_{j=1}^J (f_j - r_j) \ln(Q_j) \quad (\text{B.4})$$

En la práctica se suele realizar una reparametrización de la Ec. 2.18 definiendo  $\nu_i = -a_i b_i$ , por lo tanto  $P_i(\theta_j)$  en el modelo 2PL puede expresarse de la siguiente manera:

$$\begin{aligned} P_j = P_i(\theta_j) = P(Y_{ij} = 1 | \theta_j, a_i, b_i) &= \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \\ &= \frac{1}{1 + e^{-(a_i \theta_j + \nu_i)}} \end{aligned} \quad (\text{B.5})$$

$$\begin{aligned} Q_j = Q_i(\theta_j) = P(Y_{ij} = 0 | \theta_j, a_i, b_i) &= \frac{1}{1 + e^{a_i(\theta_j - b_i)}} \\ &= \frac{1}{1 + e^{(a_i \theta_j + \nu_i)}} \end{aligned} \quad (\text{B.6})$$

Maximizar B.4 implica buscar  $\hat{a}_i = \hat{a}_i(\theta_1, \dots, \theta_J)$ ,  $\hat{\nu}_i = \hat{\nu}_i(\theta_1, \dots, \theta_J)$  tal que  $\partial l / \partial a_i = 0$ ,  $\partial l / \partial \nu_i = 0$  y se debe de cumplir el criterio del Hessiano. Esto nos lleva a resolver el sistema de ecuaciones:

$$\frac{\partial l}{\partial a_i} \Big|_{(\hat{a}_i, \hat{\nu}_i)} = 0 \quad , \quad \frac{\partial l}{\partial \nu_i} \Big|_{(\hat{a}_i, \hat{\nu}_i)} = 0 \quad (\text{B.7})$$

Derivando B.5, B.6, definiendo  $W_j = P_j Q_j$  (pesos de Urban-Müller), y sustituyendo en B.7 tenemos que:

$$\frac{\partial l}{\partial a_i} = \sum_{j=1}^J \theta_j f_j W_j \left( \frac{p_j - P_j}{P_j Q_j} \right) \quad (\text{B.8})$$

$$\frac{\partial l}{\partial \nu_i} = \sum_{j=1}^J f_j W_j \left( \frac{p_j - P_j}{P_j Q_j} \right)$$

$$\frac{\partial^2 l}{\partial a_i^2} = \sum_{j=1}^J f_j W_j$$

$$\frac{\partial^2 l}{\partial \nu_i^2} = \sum_{j=1}^J \theta_j^2 f_j W_j \quad (\text{B.9})$$

$$\frac{\partial^2 l}{\partial \nu_i \partial a_i} = \frac{\partial^2 l}{\partial a_i \partial \nu_i} = \sum_{j=1}^J \theta_j f_j W_j$$

La búsqueda de soluciones a el sistema de ecuaciones B.7 (no lineal), que maximicen  $l$ , se realiza a través de algoritmos de aproximaciones sucesivas. Uno de los procedimientos utilizados es el método de Newton-Raphson<sup>1</sup>, que consiste en iterar haciendo uso de la serie de Taylor de primer orden en torno a  $(\hat{a}_i, \hat{\nu}_i)$ , puntos que maximizan  $l = l(a_i, \nu_i)$ .

El proceso de iteración consiste en asumir que  $\hat{a}_i^{(1)} = \hat{a}_i^{(0)} + \Delta \hat{a}_i^{(0)}$  y  $\hat{\nu}_i^{(1)} = \hat{\nu}_i^{(0)} + \Delta \hat{\nu}_i^{(0)}$ , donde  $\hat{a}_i^{(0)}$ ,  $\hat{\nu}_i^{(0)}$  son las aproximaciones (solución semilla o primer estimación) a  $\hat{a}_i^{(1)}$  y  $\hat{\nu}_i^{(1)}$  (potenciales soluciones a B.8), mientras que  $\Delta \hat{a}_i^{(0)}$  y  $\Delta \hat{\nu}_i^{(0)}$  son los errores en la aproximación. Por lo tanto, la serie de Taylor en torno a  $(\hat{a}_i, \hat{\nu}_i)$  de las funciones  $\partial l / \partial a_i$  y  $\partial l / \partial \nu_i$  es:

$$\frac{\partial l}{\partial a_i}(\hat{a}_i^{(1)}, \hat{\nu}_i^{(1)}) \cong \frac{\partial l}{\partial a_i}(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) + \frac{\partial}{\partial a_i} \left( \frac{\partial l}{\partial a_i} \right) (\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) \cdot \Delta \hat{a}_i^{(0)} + \frac{\partial}{\partial \nu_i} \left( \frac{\partial l}{\partial a_i} \right) (\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) \cdot \Delta \hat{\nu}_i^{(0)} \quad (\text{B.10})$$

$$\frac{\partial l}{\partial \nu_i}(\hat{a}_i^{(1)}, \hat{\nu}_i^{(1)}) \cong \frac{\partial l}{\partial \nu_i}(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) + \frac{\partial}{\partial a_i} \left( \frac{\partial l}{\partial \nu_i} \right) (\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) \cdot \Delta \hat{a}_i^{(0)} + \frac{\partial}{\partial \nu_i} \left( \frac{\partial l}{\partial \nu_i} \right) (\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) \cdot \Delta \hat{\nu}_i^{(0)}$$

El sistema de ecuaciones B.10 puede reescribirse así:

$$\frac{\partial l}{\partial a_i}(\hat{a}_i^{(1)}, \hat{\nu}_i^{(1)}) \cong \frac{\partial l}{\partial a_i}(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) + \frac{\partial^2 l}{\partial a_i^2}(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) \cdot \Delta \hat{a}_i^{(0)} + \frac{\partial^2 l}{\partial \nu_i \partial a_i}(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) \cdot \Delta \hat{\nu}_i^{(0)} \quad (\text{B.11})$$

$$\frac{\partial l}{\partial \nu_i}(\hat{a}_i^{(1)}, \hat{\nu}_i^{(1)}) \cong \frac{\partial l}{\partial \nu_i}(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) + \frac{\partial l}{\partial a_i \partial \nu_i}(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) \cdot \Delta \hat{a}_i^{(0)} + \frac{\partial^2 l}{\partial \nu_i^2}(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) \cdot \Delta \hat{\nu}_i^{(0)}$$

En la expresión anterior estamos utilizando la siguiente notación:

$$\frac{\partial l}{\partial a_i}(\hat{a}_i^{(1)}, \hat{\nu}_i^{(1)}) : \text{la función derivada parcial es evaluada en el punto } (\hat{a}_i^{(1)}, \hat{\nu}_i^{(1)}),$$

análogo para la derivada segunda.

---

<sup>1</sup>Una descripción a alto nivel de este método puede encontrarse en A.

Teniendo en cuenta que  $(\hat{a}_i^{(1)}, \hat{\nu}_i^{(1)})$  debe satisfacer B.7, podemos reescribir el sistema de ecuaciones de la siguiente manera:

$$\begin{aligned}
-\frac{\partial l}{\partial a_i}(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) &= \frac{\partial^2 l}{\partial \hat{a}_i^2}(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) \cdot \Delta \hat{a}_i + \frac{\partial^2 l}{\partial \nu_i \partial a_i}(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) \cdot \Delta \hat{\nu}_i \\
-\frac{\partial l}{\partial \nu_i}(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) &= \frac{\partial l}{\partial a_i \partial \nu_i}(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) \cdot \Delta \hat{a}_i + \frac{\partial^2 l}{\partial \nu_i^2}(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)}) \cdot \Delta \hat{\nu}_i
\end{aligned} \tag{B.12}$$

Dado que  $\partial^2 l / \partial a_i \partial \nu_i$  y  $\partial^2 l / \partial \nu_i \partial a_i$  existen y son continuas, se cumple la propiedad:  $\partial^2 l / \partial a_i \partial \nu_i = \partial^2 l / \partial \nu_i \partial a_i$ , por lo tanto, el sistema de ecuaciones puede escribirse en forma matricial:

$$\begin{bmatrix} \frac{\partial l}{\partial a_i} \\ \frac{\partial l}{\partial \nu_i} \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 l}{\partial \hat{a}_i^2} & \frac{\partial^2 l}{\partial \nu_i \partial \hat{a}_i} \\ \frac{\partial^2 l}{\partial \nu_i \partial \hat{a}_i} & \frac{\partial^2 l}{\partial \hat{\nu}_i^2} \end{bmatrix} \cdot \begin{bmatrix} \Delta \hat{a}_i^{(0)} \\ \Delta \hat{\nu}_i^{(0)} \end{bmatrix} \tag{B.13}$$

En la expresión anterior hemos omito explicitar que las derivadas parciales son evaluadas en  $(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)})$ . Observar que B.13 es un sistema lineal en  $\Delta \hat{a}_i^{(0)}, \Delta \hat{\nu}_i^{(0)}$  y su solución es:

$$\begin{bmatrix} \Delta \hat{a}_i^{(0)} \\ \Delta \hat{\nu}_i^{(0)} \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 l}{\partial \hat{a}_i^2} & \frac{\partial^2 l}{\partial \nu_i \partial \hat{a}_i} \\ \frac{\partial^2 l}{\partial \nu_i \partial \hat{a}_i} & \frac{\partial^2 l}{\partial \hat{\nu}_i^2} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \frac{\partial l}{\partial a_i} \\ \frac{\partial l}{\partial \nu_i} \end{bmatrix} \tag{B.14}$$

De esta forma obtenemos una estimación para  $\Delta \hat{a}_i^{(0)}, \Delta \hat{\nu}_i^{(0)}$  en base  $(\hat{a}_i^{(0)}, \hat{\nu}_i^{(0)})$ , es decir, primer aproximación a  $(\hat{a}_i^{(1)}, \hat{\nu}_i^{(1)})$  que habíamos propuesto. Si dicha primer aproximación es buena, entonces  $\Delta \hat{a}_i^{(0)} < \epsilon_{a_i}$  y  $\Delta \hat{\nu}_i^{(0)} < \epsilon_{\nu_i}$ , para cualquier  $\epsilon_{a_i}, \epsilon_{\nu_i}$  arbitrarios pequeños y en consecuencia hemos encontrado la solución. Cuando esta desigualdad no se cumple, a partir del valor obtenido para  $(\hat{a}_i^{(1)}, \hat{\nu}_i^{(1)})$  se construye una nueva aproximación a la solución de la siguiente manera:  $\hat{a}_i^{(2)} = \hat{a}_i^{(1)} + \Delta \hat{a}_i^{(1)}$  y  $\hat{\nu}_i^{(2)} = \hat{\nu}_i^{(1)} + \Delta \hat{\nu}_i^{(1)}$  y se repetimos el proceso. En forma matricial tenemos que:

$$\begin{bmatrix} \hat{a}_i^{(2)} \\ \hat{\nu}_i^{(2)} \end{bmatrix} = \begin{bmatrix} \hat{a}_i^{(1)} \\ \hat{\nu}_i^{(1)} \end{bmatrix} + \begin{bmatrix} \Delta \hat{a}_i^{(1)} \\ \Delta \hat{\nu}_i^{(1)} \end{bmatrix} \tag{B.15}$$

La solución al sistema lineal es:

$$\begin{bmatrix} \Delta \hat{a}_i^{(1)} \\ \Delta \hat{\nu}_i^{(1)} \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 l}{\partial \hat{a}_i^2} & \frac{\partial^2 l}{\partial \hat{a}_i \partial \nu_i} \\ \frac{\partial^2 l}{\partial \nu_i \partial \hat{a}_i} & \frac{\partial^2 l}{\partial \hat{\nu}_i^2} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \frac{\partial l}{\partial a_i} \\ \frac{\partial l}{\partial \nu_i} \end{bmatrix} \tag{B.16}$$

En B.16 hemos omitido incluir el punto en donde son evaluadas las derivadas parciales  $(\hat{a}_i^{(1)}, \hat{\nu}_i^{(1)})$ . Utilizando el mismo criterio, si la aproximación es buena, entonces  $\Delta \hat{a}_i^{(1)} < \epsilon_{a_i}$  y  $\Delta \hat{\nu}_i^{(1)} < \epsilon_{\nu_i}$ . En caso contrario se continua iterando, donde la expresión genérica de la solución propuesta en el paso  $k + 1$  es:

$$\begin{bmatrix} \hat{a}_i^{(k+1)} \\ \hat{\nu}_i^{(k+1)} \end{bmatrix} = \begin{bmatrix} \hat{a}_i^{(k)} \\ \hat{\nu}_i^{(k)} \end{bmatrix} + \begin{bmatrix} \Delta \hat{a}_i^{(k)} \\ \Delta \hat{\nu}_i^{(k)} \end{bmatrix} \quad (\text{B.17})$$

Mientras que su solución puede expresarse, utilizando B.8 y B.9, de la siguiente manera:

$$\begin{aligned} \begin{bmatrix} \Delta \hat{a}_i^{(k)} \\ \Delta \hat{\nu}_i^{(k)} \end{bmatrix} &= - \begin{bmatrix} \frac{\partial^2 l}{\partial \hat{a}_i^2} & \frac{\partial^2 l}{\partial \hat{\nu}_i \partial \hat{a}_i} \\ \frac{\partial^2 l}{\partial \hat{\nu}_i \partial \hat{a}_i} & \frac{\partial^2 l}{\partial \hat{\nu}_i^2} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \frac{\partial l}{\partial \hat{a}_i} \\ \frac{\partial l}{\partial \hat{\nu}_i} \end{bmatrix} \\ &= - \begin{bmatrix} -\sum_{j=1}^J \theta_j^2 f_j W_j & -\sum_{j=1}^J \theta_j f_j W_j \\ -\sum_{j=1}^J \theta_j f_j W_j & -\sum_{j=1}^J f_j W_j \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_{j=1}^J \theta_j f_j W_j \left( \frac{p_j - P_j}{P_j Q_j} \right) \\ \sum_{j=1}^J f_j W_j \left( \frac{p_j - P_j}{P_j Q_j} \right) \end{bmatrix} \end{aligned} \quad (\text{B.18})$$

Teniendo en cuenta que las derivadas de B.18 son evaluadas en  $(\hat{a}_i^{(k)}, \hat{\nu}_i^{(k)})$  y sustituyendo en B.17 tenemos que:

$$\begin{bmatrix} \hat{a}_i^{(k+1)} \\ \hat{\nu}_i^{(k+1)} \end{bmatrix} = \begin{bmatrix} \hat{a}_i^{(k)} \\ \hat{\nu}_i^{(k)} \end{bmatrix} - \begin{bmatrix} -\sum_{j=1}^J \theta_j^2 f_j W_j & -\sum_{j=1}^J \theta_j f_j W_j \\ -\sum_{j=1}^J \theta_j f_j W_j & -\sum_{j=1}^J f_j W_j \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_{j=1}^J \theta_j f_j W_j \left( \frac{p_j - P_j}{P_j Q_j} \right) \\ \sum_{j=1}^J f_j W_j \left( \frac{p_j - P_j}{P_j Q_j} \right) \end{bmatrix} \quad (\text{B.19})$$

Una vez que se ha obtenido la estimación para  $\hat{a}_i$  y  $\hat{\nu}_i$ , tenemos que la estimación para el parámetro de dificultad es  $\hat{b}_i = -\hat{\nu}_i / \hat{a}_i$ .

Esta página ha sido intencionalmente dejada en blanco.

## Apéndice C

# Extensiones al Modelo de Rasch

### C.1. Introducción

En las últimas décadas diversos autores han propuesto una serie de extensiones al MR, para ello han utilizando distintos enfoques. En algunos casos se ha perseguido el objetivo de resolver un problema específico. Mientras que en otros casos, se busca flexibilizar las hipótesis sobre las cuales Rasch construyó su propuesta en el análisis de la CCI, es decir, la independencia local y homogeneidad en la dificultad del ítem.

Un posible abordaje a la hora de modelar el problema de extensión del MR es descrito por von Davier (2007, Chap. 2), quien extiende los criterios propuestos Junker (1999). A éste lo podemos resumir en tres casos:

- Caso 1: asumir una estrategia común entre personas.
- Caso 2: la estrategia puede ser distinta entre personas.
- Caso 3: la estrategia puede ser distinta entre ítems dentro de un mismo grupo de personas.

Por ejemplo, en el primer caso podemos encontrar el denominado modelo de test lineal logístico, mientras que en el segundo caso se encuentran mezclas de éste o del modelo de Rasch, por último en el tercer caso se encuentran los modelos politómicos.

A continuación estaremos presentando tres extensiones al modelo propuesto por Rasch, en particular, aquellos donde los ítems son considerados como dicotómicos, puesto que éstos son el objeto de estudio en nuestro TFG.

## C.2. Modelo de Test Lineal Logístico - LLTM

El modelo de test lineal logístico o LLTM es una extensión al modelo de Rasch, propuesto por Fischer (1972,7), donde se impone un conjunto de restricciones, a través de funciones lineales, a los parámetros  $b_i$ , es decir, aquellos que son utilizados para modelar la dificultad del ítem. La propuesta nace a partir de la idea de modelar la dificultad del ítem, a través de una función que depende de ciertas operaciones cognitivas (o pasos) involucradas en el proceso cuyo resultado es la respuesta a un ítem.

A diferencia del modelo propuesto por Rasch, en LLTM la dificultad del ítem se modela a través de la suma de funciones lineales, cada una de las cuales contiene un parámetro cuyo fin es modelar la dificultad cognitiva en el paso correspondiente. Es decir, supongamos que dos ítems  $i, i'$  involucran dos y tres pasos en el proceso cognitivo respectivamente, Además, supongamos que ambos ítems coinciden en dos de estos procesos, por lo tanto, asumiendo un modelo lineal aditivo, la dificultad del ítem puede expresarse de la siguiente manera:

$$b_i = b_0 + \alpha_1 + \alpha_2 \tag{C.1}$$

$$b_{i'} = b_0 + \alpha_1 + \alpha_2 + \alpha_3$$

Donde  $\alpha_1, \alpha_2, \alpha_3$  se utilizan para modelar la dificultad en cada operación cognitiva y  $b_0$  es una constante. Observar que el sumando utilizado para modelar la dificultad en la tercer operación ( $\alpha_3$ ), bajo este esquema, coincide con la diferencia de la dificultad en los ítems involucrados, lo que equivale a  $b_{i'} - b_i = \alpha_3$ . De esta forma se logra un modelo más flexible para un conjunto de problemas que así lo requieran.

Estas ideas pueden ser extendidas al caso en que se utiliza un conjunto de  $L_I$  características, que notaremos como  $X_{il_I}$ ,  $l_I = 1, 2, \dots, L_I$ , para explicar la diferencia entre ítems (Fischer and Molenaar, 1995). Por lo tanto, si partimos del modelo presentado en la Ec. 2.28, éste es extendido a un LLTM, bajo las mismas condiciones antes mencionadas, de la siguiente manera:

$$P_i(\theta_j) = P(Y_{ij} = 1 | \theta_j, b_i) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}} \tag{C.2}$$

$$b_i = b_0 + \sum_{l_I=1}^{L_I} b_{l_I} X_{il_I}$$

$$l_I = 1, 2, \dots, L_I, i = 1, 2, \dots, n., j = 1, 2, \dots, J.$$

Observar que  $X_{il_I}$  corresponde al valor en la  $l_I$ -ésima operación cognitiva para el  $i$ -ésimo ítem, considerada como un efecto fijo, mientras que  $b_{l_I}$  es el peso de dicha operación. Además, es la habilidad del examinado la componente aleatoria en el modelo y en general se asume  $\theta_j \sim N(0, \sigma_\theta^2)$ , siendo independiente entre participantes.

En lo que refiere a la estimación de los parámetros  $\alpha_h$ , los enfoques utilizados son los mismos que se emplean en la estimación de los parámetros del modelo de Rasch. Estos son: Joint Maximum Likelihood (JML), Conditional Maximum Likelihood (CML), Marginal Likelihood Maximum (MML), y junto con evaluación de la bondad en el ajuste son abordados en el Capítulo 3.

El ajuste de LLTM a un conjunto de datos demanda, por un lado, obtener resultados satisfactorios del ajuste de los datos a un MR, y por otro, que se cumplan las ecuaciones lineales sobre los parámetros básicos. Por lo tanto, este tipo de modelo presenta una característica más restrictiva respecto al propuesto por Rasch. Al igual que con los métodos de estimación, en el Capítulo refchap-est-param se analizan los distintos enfoques que pueden utilizarse a la hora de analizar el ajuste del modelo. Además, advertir que existen diversas extensiones al LLTM, algunos de éstos recientemente publicados Abbakumov, Desmet, and Van den Noortgate (2019); Brandt, Moulton, and Duckor (2015); Hoijtink and Vollema (2003); van der Linden (2016).

### C.3. Modelo de Test Lineal Logístico para Cambios - LLTMC

El LLTM ofrece un enfoque alternativo al propuesto inicialmente por Rasch, brindando un mayor grado de flexibilidad en el modelado a través del vínculo entre la dificultad del ítem y las operaciones cognitivas en el proceso de respuesta. Sin embargo, conlleva un acabado conocimiento, o un instinto muy agudo, sobre cuáles son dichas operaciones. Algo que, dependiendo del tipo de rasgo latente en estudio, no siempre es posible conocer.

Estas restricciones motivaron la propuesta de conferir a LLTM una interpretación alternativa, la cual consiste en considerarlo como un modelo para el análisis de cambios que permite medir los efectos de un tratamiento. A esta nueva interpretación se la denomina modelo de test lineal logístico para cambios o Linear Logistic Test Model for Change (LLTMC), y fue desarrollada por Fischer (1972,8).

Bajo el paraguas del nuevo enfoque, una aplicación inmediata puede encontrarse en los estudios cognitivos longitudinales. Pensemos en el caso donde los examinados responden el mismo test en una cantidad finita de oportunidades, cada cierto período de tiempo. Además, imaginemos el escenario en que a lo largo de dicho período los examinados pueden ser sometidos a algún tipo de tratamiento. Una vez finalizado el estudio, es posible construir una secuencia en el tiempo de la matriz de respuestas<sup>1</sup>. Lo cual permitirá estudiar el cambio en  $\theta_j$ , es decir, el rasgo latente de cada examinado.

Supongamos que el test es proporcionado a los participantes en  $H$  oportunidades durante un período de tiempo, entonces, podemos pensar que por cada ítem  $i$ , ahora se disponen de  $H$  ítems virtuales ( $i_h, h = 1, 2, \dots, H$ ). De modo que, los cambios que se puedan producir en  $\theta_j$ , a lo largo de las distintas instancias que el examinado responde a los ítems del test, pueden ser vistos como cambios en la parametrización ( $b_{i_h}$ ) del modelo.

Un supuesto sobre el cual se construye el LLTMC esta vinculado al efecto del tratamiento, que notaremos como  $\tilde{b}_h$ , sobre  $b_{i_h}$  a lo largo del tiempo. Dicho efecto se asume constante entre dos oportunidades consecutivas ( $h - 1, h$ ) para todos los participantes (Fischer, 1989, Chap 9). En otras palabras, tendremos que:

$$b_{i_h} - b_{i_{h-1}} = \tilde{b}_h = \text{cte} , \quad \forall h > 1 \tag{C.3}$$

En forma equivalente, si realizamos la recursión, se tiene que:

---

<sup>1</sup>Nos referimos a una tabla de idénticas características a la Tabla 2.1.

$$b_{i_h} = b_{i_1} + \sum_{h=2}^h \tilde{b}_h \quad (\text{C.4})$$

La expresión obtenida en la Ec. C.4 refleja que el parámetro de dificultad es modelado a través de un modelo lineal en  $\tilde{b}_h$ . Además podemos observar que su forma es análoga a la Ec. C.2. Lo cual refleja, tal como indicamos antes, que LLTMC representa una interpretación de LLTM.

La inferencia sobre los parámetros puede realizar con en mismo abordaje que en LLTM, como hemos mencionado, éstas son analizadas en el Capítulo 3.

## C.4. Modelo de Rasch Regresión Latente

La tercer extensión que presentaremos corresponde a la que De Boeck and Wilson (2004) denominan modelo de Rasch Regresión Latente. Dicho modelo contempla la posibilidad de incluir características propias de cada examinado como variables regresoras, por ejemplo, sexo, raza, período de tiempo desde que se diagnosticó cierta enfermedad de afectación cognitiva, etc.

Bajo este nuevo enfoque la variable latente  $\theta_j$  es modelada a través de una combinación lineal de un conjunto de variables predictoras, consideradas como efectos fijos, asociadas al  $j$ -ésimo examinado. A éstas las representaremos como  $z_{jl_j}$ , que corresponde a la  $l_j$ -ésima variable regresora en el  $j$ -ésimo examinado. De modo que el modelo es de Rasch de regresión latente es:

$$P_i(\theta_j) = P(Y_{ij} = 1 | \theta_j, b_i) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}} \quad (\text{C.5})$$

$$\theta_j = \vartheta_0 + \sum_{l_j=1}^{L_j} \vartheta_{l_j} \cdot z_{jl_j} + \epsilon_j$$

$$l_j = 1, 2, \dots, L_j, \quad i = 1, 2, \dots, n., \quad j = 1, 2, \dots, J.$$

Siendo  $L_j$  la cantidad de variables que representan los efectos fijos involucradas en el modelo y  $\epsilon_j$  corresponde a los errores aleatorios, no correlacionados, que asumiremos con distribución  $N(0, \sigma_\epsilon^2)$ . Observar que  $z_{jl_j}$  y  $\vartheta_{l_j}$  corresponden a el valor del predictor  $l_j$  y su peso respectivamente para el  $j$ -ésimo examinado.

Distintos autores proponen este enfoque con el fin de analizar el conjunto de respuestas a un test, cuando es posible conocer o medir ciertas características en los examinados. Zwinderman (1991,9) ha presentado un completo desarrollo en esta dirección. Además, como es de esperar, también se han realizado propuestas que combinan LLTM y el modelo de Rasch regresión latente, donde las características del ítem y los examinados forman parte de los efectos fijos, mientras que la componente aleatoria esta dada en el error al estimar la habilidad de los individuos.

Finalmente mencionar que los métodos utilizados en la estimación de los parámetros son: Joint Maximum Likelihood (JML), Conditional Maximum Likelihood (CML), Marginal Likelihood Maximum (MML) y junto con evaluación de la bondad en el ajuste son analizados en el Capítulo 3.

## Apéndice D

# Métodos de Cadenas de Markov Monte Carlo

### D.1. Introducción

Los métodos MCMC son una útil herramienta dentro del área de la estadística bayesiana computacional, así como también en otras disciplinas. En la presente sección nos proponemos dar una visión a alto nivel de este método, así como también del algoritmo M-H y muestreo de Gibbs. Para ello utilizamos como referencia a los siguientes autores: Alvarez-Castro (2021), Ausín (2012), Fox (2010), Hoff (2009), Robert and Casella (2010).

Los MCMC son métodos de simulación que permite generar muestras de distribuciones de probabilidad y estimar estadísticos de interés, por ejemplo, media, mediana desviación estándar cuantiles, etc. En la estadística bayesiana son utilizados para generar muestras de la distribución a posteriori, principalmente cuando la forma analítica es desconocida o difícil de obtener.

Una cadena de Markov se define como la secuencia de Variables Aleatorias (VAs)  $\Theta^{(0)}, \Theta^{(1)}, \Theta^{(2)}, \dots, \Theta^{(t)}, \dots$  tal que la distribución de  $\Theta^t$  sólo depende de  $\Theta^{(t-1)}$ , es decir, el futuro sólo depende del pasado a través del presente. En forma analítica esto es:

$$P(\Theta^{(t)} \in \mathcal{A} | \Theta^{(0)}, \Theta^{(1)}, \Theta^{(2)}, \dots, \Theta^{(t-1)}) = P(\Theta^{(t)} \in \mathcal{A} | \Theta^{(t-1)} = \theta^{(t-1)}) \quad (\text{D.1})$$

La cadena de Markov se dice que es homogénea en el tiempo cuando esta definida por su valor inicial  $\theta^0$  en  $t = 0$ . Por lo tanto, tenemos que el núcleo de transición es  $P(\Theta^t = \theta^t | \Theta^{(t-1)} = \theta^{(t-1)})$ . Bajo ciertas condiciones de regularidad, la cadena de Markov converge a su distribución estacionaria, es decir:

$$\lim_{t \rightarrow \infty} P(\Theta^{(t)} = \theta | \Theta_0 = \theta_0) = \pi(\theta) \quad (\text{D.2})$$

La idea de los métodos MCMC consiste en obtener la distribución estacionaria  $\pi(\theta | \mathbf{x})$  a partir de las muestras  $(\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots)$  que simulan la cadena de Markov. El vector  $\mathbf{x}$  corresponde a las observaciones de una VA  $X$  tal que asumimos  $Y \sim f(x|\theta)$ . Además, se trabaja bajo la hipótesis que la cadena es estacionaria, por lo tanto,  $\theta^{(t)}$  sólo depende de  $\theta^{(t-1)}$ .

Las primeras muestras de la cadena son descartadas (se les denomina iteraciones de burn-in), y la cantidad a descartar debe de ser tal que la cadena logre llegar a su estado estacionario. De modo que, es necesario simular un número grande de iteraciones. Cuando el algoritmo esta bien implementado la convergencia de la cadena está garantizada independientemente del valor inicial (semilla) elegido.

Una vez que se ha logrado la convergencia de la cadena, resulta de interés, por ejemplo, obtener la media a posteriori. En este sentido un estimador insesgado de la media y la varianza es:

$$\hat{\theta} = \frac{1}{M} \sum_{t=T_b+1}^{M+T_b} \theta^{(t)} \quad (\text{D.3})$$

$$\hat{\sigma}_\theta^2 = \frac{1}{M-1} \sum_{t=T_b+1}^{M+T_b} (\theta^{(t)} - \hat{\theta})^2 \quad (\text{D.4})$$

En la expresión anterior estamos asumiendo que las primeras muestras ( $T_b$ ) han sido descartadas en la cadena y son tomadas en cuenta las siguiente  $M$  candidatas. Bajo las hipótesis del Teorema Central del Límite, la convergencia de  $E(\theta) \rightarrow \hat{\theta}$  esta garantizada cuando el numero de observaciones ( $M$ ) es grande. Sumado a esto, construyendo el histograma de las observaciones podemos obtener la forma, dispersión, simetría, etc.

## D.2. Algoritmo Metropolis Hasting

El algoritmo M-H permite simular cadenas de Markov y es una técnica utilizada en la estadística bayesiana a la hora de obtener la densidad a posteriori de una parámetro. Supongamos que  $\theta$  es el parámetro que queremos estimar, cuya distribución estacionaria es  $\pi(\theta|\mathbf{x})$ . El algoritmo consiste en los siguientes pasos:

1. Fijar un valor inicial  $\theta^{(0)}$ .
2. Definir la distribución candidata  $q(\tilde{\theta}|\theta^{(t)})$  que será utilizada para generar las muestras.
3. Simular un valor candidato  $\tilde{\theta}$  utilizando  $q(\tilde{\theta}|\theta^{(t)}) \forall t = 1, 2, 3 \dots$
5. Computar la probabilidad de aceptar el  $\tilde{\theta}$ :

$$\alpha = \min \left\{ 1, \frac{\pi(\tilde{\theta}|\mathbf{x})/q(\tilde{\theta}|\theta^{(t)})}{\pi(\theta^{(t)}|\mathbf{x})/q(\theta^{(t)}|\tilde{\theta})} \right\} = \min \left\{ 1, \frac{\pi(\tilde{\theta}|\mathbf{x})q(\theta^{(t)}|\tilde{\theta})}{\pi(\theta^{(t)}|\mathbf{x})q(\tilde{\theta}|\theta^{(t)})} \right\} = \min \left\{ 1, \frac{f(\mathbf{x}|\tilde{\theta})\pi(\tilde{\theta})q(\theta^{(t)}|\tilde{\theta})}{f(\mathbf{x}|\theta^{(t)})\pi(\theta^{(t)})q(\tilde{\theta}|\theta^{(t)})} \right\}$$

6. Obtener una realización  $u$  a partir de  $U \sim \mathcal{U}(0, 1)$ .
7. Si  $u < \alpha$ , entonces tomar  $\theta^{(t+1)} = \tilde{\theta}$ , de lo contrario  $\theta^{(t+1)} = \theta^{(t)}$ .
8. Volver al paso 2.

Observar que los pasos 2 y 3 se repiten un cantidad elevada ( $T_b$ ) de veces hasta tener certeza que la cadena ha alcanzado convergencia. Una vez que se han descartado las iteraciones de burn-in, el valor candidato  $\tilde{\theta}$  con probabilidad  $\alpha$  será aceptado.

En el desarrollo que antecede partimos de la hipótesis que  $f(\mathbf{x}|\theta)$  es conocida y proponemos una densidad candidata  $q(\tilde{\theta}|\theta^{(t)})$ . Por la regla de Bayes tenemos que  $\pi(\tilde{\theta}|x) = f(\mathbf{x}|\tilde{\theta})\pi(\tilde{\theta})/f(\mathbf{x})$  y  $\pi(\theta^{(t)}|x) = f(\mathbf{x}|\theta^{(t)})\pi(\theta^{(t)})/f(\mathbf{x})$ . En general, la constante de proporcionalidad  $f(\mathbf{x})$  es desconocida. Sin embargo, la probabilidad de aceptación  $\alpha$  es independiente de dicha densidad marginal, esto se ve reflejado en la última igualdad.

### D.3. Muestreo de Gibbs

El muestreo de Gibbs es un caso particular del algoritmo M-H donde la probabilidad ( $\alpha$ ) de aceptar el valor candidato ( $\tilde{\theta}$ ) es siempre igual a uno. En otras palabras, se asume que las densidades propuestas coinciden con las densidades a posteriori condicionadas en lo datos observados

Supongamos que  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$  el vector de parámetros involucrados en el modelo y deseamos obtener para cada  $\theta_i$  la distribución condicional (en el resto de los parámetros y datos observados) a posteriori. Una de las dificultades bajo el enfoque del muestreo de Gibbs se encuentra en el hecho que no siempre es posible obtener, para algunos de los parámetros involucrados, el valor candidato. Es decir, durante el proceso donde se genera la muestra del valor candidato, por ejemplo,  $\theta_i^{(t+1)} \sim P(\theta_i|\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{i-1}^{(t)}, \theta_{i+1}^{(t)}, \dots, \theta_p^{(t)})$ , no siempre es posible calcular dicha densidad condicional.

En estos casos se utiliza el algoritmo denominado Hastings-within-Gibbs, donde a través de M-H son generados aquellos parámetros que son difíciles de generar una muestra. De esta forma, es posible obtener, respecto a el muestro de Gibbs, mejores resultados en lo que refiere a performance en la convergencia en aquellos casos en los que existe alta correlación entre muestras.

Vale la pena señalar que tanto el método M-H, el muestreo de Gibbs o una combinación de éstos pueden resultar menos eficiente, respecto a un algoritmo que haya sido diseñado especialmente para resolver el problema de estimación de la densidad a posteriori. Sin embargo, suelen ser una excelente herramienta que puede ser utilizada, por ejemplo, para compara performance, tiempo de convergencia y precisión en las estimaciones.

## D.4. Análisis de la Convergencia

Un aspecto importante a la hora de implementar los algoritmos M-H, el muestreo de Gibbs o una combinación de éstos, esta asociada a la convergencia de la cadena de Markov, así como también la correlación entre muestras. En este sentido se han realizado distintas propuestas para abordar dicha problemática. A continuación se describen algunas de las más relevantes.

Una vez que se ha logrado obtener una potencial distribución a posteriori, es recomendable realizar:

- Una inspección visual en lo que refiere a la convergencia de las cadenas al valor estacionario. Éstas deben ser generadas a partir de distintas semillas.
- Verificar que la muestra obtenida, sin las iteraciones de burn-in, sea (aproximadamente) una muestra independiente e idénticamente distribuida.
- Inspeccionar las funciones de autocorrelación de los valores generados en la muestra. Dado que la muestra proviene de una cada de Markov, debería de observarse una correlación positiva.
- Verificar la convergencia de la cadena a algún estadístico predefinido, por ejemplo, la media.

Un riguroso abordaje sobre análisis de la convergencia, así como de los algoritmos antes presentados es presentados, puede encontrarse en Brooks, Gelman, Jones, and Meng (2011), Gelman et al. (2021).

# Apéndice E

## Modelo Lineales Mixtos

### E.1. Introducción

En los modelos lineales se asume que los regresores  $(\beta_0, \beta_1, \dots, \beta_p)$  son parámetros fijos y desconocidos, éstos corresponden a los llamados efectos fijos o fixed effects. Un nivel más de flexibilidad puede alcanzarse al recurrir a los llamados LMM, donde además de los efectos fijos están presentes los llamados efectos aleatorios o random effects.

Al trabajar con LMM, los regresores asociados a los efectos aleatorios son considerados como variables aleatorias, donde los parámetros de su función de densidad pueden o no ser conocidos. Laird y Ware 1982 fueron los pioneros en desarrollar estos conceptos para el caso de datos longitudinales.

La idea central detrás de LMM radica en el hecho que, los efectos fijos modelan la forma en que las medidas difieren en una muestra, y los efectos aleatorios modelan la variabilidad general entre cierta unidad característica o estructura de asociación.

Desde esta nueva perspectiva se da paso a lo que se conoce como modelos jerárquicos, los cuales pueden ser pensados como una extensión de los modelos de regresión. A pesar de las similitudes entre ambos modelos, a través de LMM es posible incluir en el análisis subunidades específicas, de acuerdo a las características de la población en estudio, lo que se traduce en agrupamientos de acuerdo a algún criterio o características intrínseca.

Una de las principales características de LMM radica en la posibilidad de trabajar con conjuntos de datos donde las observaciones no son independientes, es decir, existe un grado de correlación entre ellas, convirtiéndolos en modelos con mayor grado de flexibilidad.

En la práctica es altamente probable encontrar datos correlacionados, un ejemplo de ello es la repetición de una medida de un mismo sujeto (o cluster) que forma parte de la muestra. Otro ejemplo puede encontrarse al analizar los resultados que obtienen estudiantes de un mismo nivel, podría ser el mismo grado en primaria, en sus pruebas finales a nivel nacional. Si tomamos muestras en distintas instituciones educativas, es de orden pensar que el puntaje final de cada estudiante esta correlacionado con la institución donde se formó, entre otros factores. Bajo estas condiciones estamos frente a una estructura jerárquica de dos niveles: escuelas y alumnos.

Más formalmente, los modelos lineales mixtos son una clase de modelos utilizados cuando los datos presentan algunas de las siguientes características: tienen una estructura de cluster (agrupados y relacionados), son longitudinales (evolución de un fenómeno a lo largo del tiempo), provienen de medidas repetidas o estructuras jerárquicas Morales Correa and Salazar Uribe (2016); West, Welch, and ecki (2007). Algunas de sus ventajas radica en el hecho que es posible obtener mejores estimaciones de los efectos fijos y disminuir el rango de variación de los errores, por tanto mejora la performance del modelo.

Comencemos por presentar la estructura del modelo de regresión lineal mixto:

$$Y = \underbrace{X\beta}_{\text{Efectos Fijos}} + \underbrace{ZU + \epsilon}_{\text{Efectos Aleatorios}} \quad (\text{E.1})$$

Siendo  $Y$  el vector de respuestas,  $X\beta$  representa los efectos fijos,  $ZU$  y  $\epsilon$  modelan los efectos aleatorios. A continuación analizaremos las hipótesis sobre  $U$  y  $\epsilon$  que permiten obtener las estimaciones sobre los parámetros del modelo.

Consideremos el caso en que los datos provienen de un conjunto de  $i$  sujetos  $i \in \{1, 2, 3, \dots, m\}$  (o clusters<sup>1</sup>), donde para cada uno de ellos se han realizado (seleccionado)  $j \in \{1, 2, \dots, n_i\}$  medidas (integrantes del cluster). Además se han tomado en cuenta de  $p$  efectos fijos covariantes y  $q$  efectos aleatorios. Por lo tanto, el modelo de efectos mixtos para un sujeto  $i$  de acuerdo a Helwig (2017) es:

$$y_{ij} = \beta_0 + \sum_{k=1}^{p-1} \beta_k x_{ijk} + u_{i0} + \sum_{k=1}^{q-1} u_{ik} z_{ijk} + e_{ij} \quad (\text{E.2})$$

Donde  $y_{ij} \in \mathbb{R}$ ,  $X_{ij} \in \mathbb{R}^p$ ,  $\beta \in \mathbb{R}^p$ ,  $Z_{ij} \in \mathbb{R}^q$ ,  $U_i \in \mathbb{R}^q$ ,  $\epsilon_{ij} \in \mathbb{R}$ ,  $\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma_e^2)$ ,  $j = 1, 2, \dots, n_i$ . De acuerdo a Fahrmeir, Kneib, Lang, and Marx (2013); West et al. (2007) las hipótesis para los efectos aleatorios del modelo son:  $U_i \stackrel{iid}{\sim} N(0, \Sigma_r)$ ,  $\Sigma_r \in \mathbb{R}^{q \times q}$  matriz de covarianza,  $\epsilon_i \stackrel{iid}{\sim} N(0, \Sigma_e)$ ,  $\Sigma_e \in \mathbb{R}^{n_i \times n_i}$  matriz de covarianza,  $i = 1, 2, \dots, m$ . Además  $U_i \in (U_1, U_2, \dots, U_m)$ ,  $\epsilon_i \in (\epsilon_1, \epsilon_2, \dots, \epsilon_m)$  independientes entre sí.

En forma matricial, de acuerdo a Brunner (2012); Czado (2018), la representación es la siguiente:

$$Y_i = X_i \beta + Z_i U_i + \epsilon_i \quad (\text{E.3})$$

$$i = 1, 2, \dots, m.$$

$$Y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ \vdots \\ y_{in_i} \end{pmatrix} \in \mathbb{R}^{n_i}; \quad X_i = \begin{pmatrix} X_{i1}^T \\ X_{i2}^T \\ \vdots \\ X_{in_i}^T \end{pmatrix} \in \mathbb{R}^{n_i \times p}; \quad Z_i = \begin{pmatrix} Z_{i1}^T \\ X_{i2}^T \\ \vdots \\ Z_{in_i}^T \end{pmatrix} \in \mathbb{R}^{n_i \times q}$$

<sup>1</sup>La descripción dentro del paréntesis corresponde al caso en que los datos están agrupados en clusters.

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_{p-1} \end{pmatrix} \in \mathbb{R}^p; U_i = \begin{pmatrix} u_{i0} \\ u_{i2} \\ u_{i3} \\ \vdots \\ u_{iq-1} \end{pmatrix} \in \mathbb{R}^q; \epsilon_i = \begin{pmatrix} \epsilon_{i0} \\ \epsilon_{i2} \\ \epsilon_{i3} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix} \in \mathbb{R}^{n_i}$$

Si tomamos en cuenta todos los sujetos  $i = 1, 2, \dots, m$ , el total de muestras esta dado por  $n = \sum_{i=1}^m n_i$ , entonces el modelo general puede ser expresado de la siguiente manera:

$$Y = X\beta + ZU + \epsilon$$

$$\begin{pmatrix} U \\ \epsilon \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Gamma_r & 0_{mq \times n} \\ 0_{n \times mq} & \Sigma_e \end{pmatrix} \right) \quad (\text{E.4})$$

Donde  $Y \in \mathcal{M}_{n \times 1}$  es el vector de respuestas,  $X \in \mathcal{M}_{n \times p}$  matriz de constantes conocidas,  $\beta \in \mathcal{M}_{p \times 1}$  vector de parámetros fijos desconocidos,  $Z \in \mathcal{M}_{n \times mq}$  matriz de constantes conocidas,  $U \in \mathcal{M}_{mq \times 1}$  matriz de efectos aleatorios<sup>2</sup>,  $\epsilon \in \mathcal{M}_{n \times 1}$  residuos. Con  $\epsilon$  y  $U$  variables aleatorias, independientes entre sí, además  $\Gamma_r$ ,  $\Sigma_e = \sigma_e I_{n_i}$  son matrices diagonal cuya estructura es:

$$\Gamma_r = \begin{pmatrix} \Sigma_r & 0 & \cdots & 0 \\ 0 & \Sigma_r & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_r \end{pmatrix} \in \mathbb{R}^{mq \times mq}; \Sigma_e = \begin{pmatrix} \Sigma_{e_1} & 0 & \cdots & 0 \\ 0 & \Sigma_{e_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{e_m} \end{pmatrix} \in \mathbb{R}^{n \times n}$$

$$\Sigma_r = \begin{pmatrix} \sigma_{r_1}^2 & \sigma_{r_{12}} & \cdots & \sigma_{r_{1q}} \\ \sigma_{r_{12}} & \sigma_{r_2}^2 & \cdots & \sigma_{r_{2q}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{r_{1q}} & \sigma_{r_{2q}} & \cdots & \sigma_{r_q}^2 \end{pmatrix}; \Sigma_e = \begin{pmatrix} \sigma_e^2 & 0 & \cdots & 0 \\ 0 & \sigma_e^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_e^2 \end{pmatrix}$$

Bajo estas condiciones, de acuerdo a Baumgartner (2013); Czado (2018), el modelo representando a través de E.4 puede ser expresado de la siguiente manera:

<sup>2</sup>En la literatura Fahrmeir et al. (2013); West et al. (2007) puede encontrarse a las matrices  $X$ ,  $Z$  con el nombre matriz de diseño.

$$\begin{aligned} Y|U &\sim N(X\beta + ZU, \Sigma_e) \\ U &\sim N(0, \Gamma_r) \end{aligned} \tag{E.5}$$

La distribución marginal<sup>3</sup> de  $Y$  esta dada por:

$$\begin{aligned} Y &\sim N(X\beta, \Sigma) \\ \Sigma &= Z\Gamma_r Z^T + \Sigma_e \end{aligned} \tag{E.6}$$

En la siguiente sección introduciremos algunos conceptos que serán el soporte para la estimación y predicción de los parámetros del modelo. Vale la pena aclarar que utilizaremos el término estimación cuando nos reframos al valor que se obtiene, sujeto a un criterio, de los parámetros correspondientes a los efectos fijos, por hipótesis éstos son constantes y a priori desconocidos. Mientras que, con el fin de diferenciar la naturaleza de los parámetros correspondientes a los efectos aleatorios, sobre los cuales hemos hecho el supuesto que son desconocidos y que además provienen de cierta distribución de probabilidad, a los valores obtenidos a través de algún criterio (que explicaremos más adelante) los llamaremos predicción.

### E.1.1. REML

A la hora de estimar la varianza de los parámetros del modelo, en general, se utiliza la función de máxima verosimilitud. Un camino alternativo, propuesto por Patterson and Thompson (1971), es Restricted Maximum Likelihood (REML), el cual permite estimar la varianza cuando, por ejemplo, las estimaciones por ML son sesgadas. La idea central detrás de REML es maximizar aquella parte de la función de verosimilitud invariante a los efectos fijos.

Un posible abordaje en el análisis es el siguiente: analizar los residuos sin tomar en cuenta los efectos fijos, por ejemplo maximizando el logaritmo de la función de verosimilitud de  $K^T Y$  en lugar de la correspondiente a  $Y$ , siendo  $K$  una matriz cuyas componentes son constantes, de rango completo por filas, tal que  $K^T X = 0$ .

Teniendo en cuenta las hipótesis de normalidad e independencia de los efectos fijos y aleatorios (E.4, E.6) se llega a que  $Y \sim N(0, K^T \Sigma K)$ :

$$\begin{aligned} E(K^T Y) &= E(K^T (X\beta + ZU + \epsilon)) = 0 \\ Var(K^T Y) &= K^T \Sigma K \end{aligned} \tag{E.7}$$

Por lo tanto, el logaritmo de la función de máxima verosimilitud, a menos de una constante, esta dado por:

$$l_{REML}(\Sigma) = -\frac{1}{2} \log |K^T \Sigma K| - \frac{1}{2} \mathbf{y}^T K (K^T \Sigma K)^{-1} K^T \mathbf{y} \tag{E.8}$$

---

<sup>3</sup>Ver Chap. 4 en Balsells (2009) o theorem 17.2 Rencher and Schaalje (2008)

Al utilizar REML se busca obtener  $\Sigma_r$  y  $\Sigma_e$  maximizando E.8.

Una característica importante de REML, cuando los datos son balanceados<sup>4</sup>, esta vinculada al hecho que sus estimaciones coinciden con las que se obtienen a través de ANOVA (Searle, McCulloch, and Casella, 2006, Chap. 2 y Chap. 6). Además es una potente herramienta para datos no balanceados en el análisis de los modelos mixtos (Chap 2. West et al. (2007)), aunque hay que tener cierto recaudo. Un completo desarrollo en la derivación de la función REML puede encontrarse en LaMotte (2007).

No debemos perder de vista que la aplicación de ambos métodos (ML y REML) requiere optimizar una función objetivo, lo cual se hace mediante algoritmos como Newton-Raphson o Expectation-Maximization. De modo que ML y REML, bajo ciertas condiciones, pueden verse afectados debido a problemas de convergencia.

En particular REML esta muy ligado a los grados de libertad de los efectos fijos, ya que la dimensión de  $K$  depende de la dimensión de  $X$ . Por lo tanto, su aplicación requiere especial cuidado cuando, el rango de la matriz  $X$  es grande en comparación al tamaño de la muestra.

Acerca de los métodos computacionales en el análisis de los modelos mixtos, una excelente guía son los artículos de Bates (March 27, 2018); Bates et al. (2015) o el artículo de Laird and Ware (1982). Mientras que acerca de la discusión ML vs REML existe variada literatura, algunos ejemplos son McCulloch and Searle (2001); Pinheiro and Bates (2000); West et al. (2007).

A partir de este momento ya estamos en condiciones de dar un paso más en el análisis de los modelos mixtos en el sentido de estimar y predecir los parámetros del modelo.

### E.1.2. Estimación de los parámetros del modelo

A la hora de estimar y predecir los parámetros asociados a los efectos fijos y aleatorios, es necesario realizar algunas consideraciones adicionales respecto a las vistas anteriormente en los modelos de regresión lineal. Vale la pena advertir que el desarrollo de estos resultados puede resultar tedioso.

En la literatura sobre LMM puede encontrarse que, distintos autores han utilizado distintas hipótesis o aproximaciones, en desarrollo de soluciones al problema de estimar y predecir los parámetros del modelo. El análisis puede ser dividido en dos casos, dependiendo si  $\Sigma$  es conocido o desconocido.

Asumir que  $\Sigma$  es conocido equivale a decir que  $\Gamma_r$  y  $\Sigma_e$  son conocidos, de modo que las hipótesis del modelo dadas por E.4 o E.5 implican E.6, sin embargo el camino inverso no siempre es válido.

Teniendo en cuenta este aspecto, una alternativa en el análisis es la siguiente: cuando sólo se requiere realizar la estimación de  $\beta$ , podemos partir del modelo marginal (asumiendo  $\Sigma$  conocido); mientras que cuando es de interés realizar la estimación de  $\beta$  y predicción de  $U$  partimos de E.5 Czado (2018).

Las herramientas utilizadas en este tipo de problemas son el principio de máxima verosimilitud, su variante REML y las ecuaciones de Henderson. A continuación presentaremos un esbozo de los resultados más relevantes, siendo nuestro soporte las publicaciones Davison (2003); Hobmaier (2011); McCulloch (2003); McCulloch and Searle (2001); Pinheiro (1994); Pinheiro and Bates (2000); Rencher and Schaalje (2008); Searle et al. (2006); West et al. (2007).

---

<sup>4</sup>Se tienen datos balanceados cuando, por ejemplo, la cantidad de muestras repetidas para cada individuo es la misma.

**E.1.2.1. Caso 1 -  $\Sigma$  conocido**

Partiendo de la hipótesis que  $\Sigma$  es conocida, a continuación desarrollamos un bosquejo de las alternativas para llegar a la estimación de  $\beta$  y  $U$ .

Comencemos por estimar los efectos fijos del modelo, para este propósito utilizaremos la distribución marginal de  $Y$  dada por E.6.

Por lo tanto, tenemos que la función de densidad para un conjunto de observaciones  $\mathbf{y}$  esta dada por la normal multivariante:

$$f(\mathbf{y}; \beta) = 2\pi^{-n/2} |\Sigma|^{-1/2} e^{-\{\frac{1}{2}(\mathbf{y}-X\beta)^T \Sigma^{-1}(\mathbf{y}-X\beta)\}} \quad (\text{E.9})$$

A partir de este punto la estimación de  $\beta$  puede realizarse utilizando:

- i. Principio de máxima verosimilitud: minimizando la función logaritmo de la verosimilitud, obteniéndose así  $\hat{\beta}_{ML}$ .
- ii. Mínimos cuadrados generalizados: aplicando el mismo proceso utilizado en el modelo de regresión, tomado en cuenta la simetría de  $\Sigma$  se obtiene el estimador  $\hat{\beta}_{GLS}$ .

Ambos caminos nos conducen al mismo resultado en la estimación de los efectos fijos. La tabla E.1 resume la estimación  $(\hat{\beta}_{ML}, \hat{\beta}_{GLS})^5$ .

El desarrollo antes descrito se basa en el hecho que  $Y$  tiene una distribución normal, sin embargo, aunque esta hipótesis no se cumpla se puede probar que la estimación obtenida para  $X\beta$  coincide con el mejor estimador lineal insesgado Best Linear Unbiased Estimator (BLUE), es decir, el mejor en el contexto de todos los estimadores lineales en  $\mathbf{y}$  con menor varianza (Hayes and Haslett, 1999).

	Efectos Fijos	Efectos Aleatorios
	$\hat{\beta}_{ML} = \hat{\beta}_{GLS} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \mathbf{y}$	$BP(U) = \tilde{u}_{bp} = \Gamma_r Z^T \Sigma^{-1} (\mathbf{y} - X \hat{\beta})$
$\Sigma$ Conocido		$BLP(U) = \tilde{u}_{blp} = \mu_Y + C \Sigma^{-1} (\mathbf{y} - \mu_Y)$
	$BLUE(X\beta) = X \hat{\beta}_{ML}$	$BLUP(U) = \tilde{u}_{blup} = \mu_Y + C \Sigma^{-1} (\mathbf{y} - \mu_Y)$

Tabla E.1: Estimación de efectos fijos y predicción de efectos aleatorios en LMM cuando  $\Sigma$  es conocida.

Podemos observar que  $\hat{\beta}_{ML}$  depende de la inversa generalizada  $(X^T \Sigma^{-1} X)^{-}$ , la cual en general no tiene una solución única. Sin embargo, si reescribimos la expresión para el estimador de la siguiente manera:

$$X \hat{\beta}_{ML} = X (X^T \Sigma^{-1} X)^{-} X^T \Sigma^{-1} \mathbf{y} \quad (\text{E.10})$$

<sup>5</sup>Utilizaremos la notación  $(A)^{-}$  para referirnos a la inversa generalizada de  $A$ .

Se tiene que  $X\widehat{\beta}_{ML}$  es invariante debido a que  $X(X^T\Sigma^{-1}X)^{-1}X^T$  también lo es (McCulloch and Searle, 2001, Chap. 6.3).

La predicción de los efectos aleatorios para el caso de  $\Sigma$  conocido, como mencionamos antes, la realizaremos a partir de E.4 y E.5, donde la función de densidad distribución conjunta la podemos expresar de la siguiente manera:

$$f(y, u; \beta) = f_{y|u} \cdot f_u \quad (\text{E.11})$$

Los tres métodos frecuentemente utilizados en la predicción de los efectos aleatorios son: Best Predictor (BP), Best Linear Predictor (BLP) y Best Linear Unbiased Predictor (BLUP). El primero de ellos es utilizado cuando se tiene conocimiento de todos los parámetros de la distribución conjunta entre  $U$  e  $Y$ , mientras que los otros dos son utilizados cuando sólo se tiene conocimiento de algunos parámetros de  $f(y, u)$  (Chap. 7 Searle et al. (2006)). A continuación estaremos desarrollando estos conceptos, la Tabla E.1 resume los resultados que subsiguen.

El mejor predictor de  $U$  se obtiene al minimizar el error cuadrático medio. Recordemos que éste no necesariamente es igual al predictor de mínima varianza. Supongamos que la función de densidad conjunta esta dada por  $f(y, u)$ , entonces el error cuadrático medio en la predicción de  $U$  es:

$$E((\tilde{u} - U)^T A (\tilde{u} - U)) = \int \int (\tilde{u} - u)^T A (\tilde{u} - u) f(y, u) dy du \quad (\text{E.12})$$

Donde  $A$  es una matriz definida positiva y simétrica. Al minimizar E.12 tenemos que:

$$BP(U) = \tilde{u}_{bp} = E(U|Y) \quad (\text{E.13})$$

Este resultado es válido cualquiera sea  $f(y, u)$  y no depende de las características de la matriz  $A$  (McCulloch and Searle, 2001, Sec. 9.8.a). Por otro lado el valor esperado y la varianza de  $BP(U)$  son (McCulloch and Searle, 2001, Appendix S1):

$$E_y(\tilde{u}_{bp}) = E_y(E_y(U|Y)) = E(U) \quad (\text{E.14})$$

$$Var(\tilde{u}_{bp} - U) = E_y(Var(U|Y))$$

De modo que  $\tilde{u}_{bp}$  es una VA cuyos momentos están dados por E.14, y su función de densidad depende de  $y$ , notar además que  $\tilde{u}_{bp}$  es insesgado sobre  $y$ .

A la hora de obtener el mejor predictor de  $\tilde{u}$  se requiere el conocimiento de la densidad conjunta ( $f(u, y) = f(u|y)f(y)$ ), recordemos que  $Y$  tiene una distribución normal, de acuerdo a E.6, entonces bajo estas condiciones se puede probar (Searle et al., 2006, App. S2) que en su forma general:

$$\begin{pmatrix} U \\ Y \end{pmatrix} \sim N \left( \begin{pmatrix} \mu_U \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Gamma_r & C \\ C^T & \Sigma \end{pmatrix} \right) \quad (\text{E.15})$$

Dada esta condición de normalidad genérica se puede probar (McCulloch and Searle, 2001, App. S2.b.) que la distribución condicional de  $U|Y$  esta dada por:

$$U|Y \sim N(\mu_U + C\Sigma^{-1}(Y - \mu_Y), \Gamma_r - C\Sigma^{-1}C^T) \quad (\text{E.16})$$

En particular bajo las hipótesis dadas por E.4 y E.6:

$$\begin{aligned}
E(U) &= \mu_U = 0 \\
E(Y) &= \mu_Y = X\beta \\
Var(U) &= \Gamma_r \\
C^T &= Cov(Y, U^T) = Cov(X\beta + ZU + \epsilon, U^T) = ZCov(U, U^T) = Z\Gamma_r^T
\end{aligned} \tag{E.17}$$

Por lo tanto, el primer y segundo momento de BP estará dado por:

$$\begin{aligned}
BP(U) &= \tilde{u}_{bp} = E(U|Y) = \mu_Y + C\Sigma^{-1}(Y - \mu_Y) \\
&= \Gamma_r Z^T \Sigma^{-1}(Y - X\hat{\beta})
\end{aligned} \tag{E.18}$$

$$Var(U|Y) = \Gamma_r - C\Sigma^{-1}C^T \tag{E.19}$$

A partir del desarrollo anterior podemos observar que  $BP(u)$  no necesariamente es lineal en  $y$ , cuando esto ocurre, es decir  $\tilde{u} = a + By$ , siendo  $a$  un vector y  $B$  una matriz cuyas componentes son constantes, estamos frente a lo que se define como BLP. Se puede probar (McCulloch and Searle, 2001, Chap. 9) que al minimizar E.12, considerando  $\tilde{u}$  lineal en  $y$  y teniendo en cuenta las definiciones dadas en E.17, se obtiene:

$$\begin{aligned}
BLP(u) &= \tilde{u}_{blp} = \mu_U + C\Sigma^{-1}(Y - \mu_Y) \\
&= \Gamma_r Z^T \Sigma^{-1}(Y - X\hat{\beta}_{ML})
\end{aligned} \tag{E.20}$$

Es importantes resaltar en este punto lo siguiente,  $BLP(u)$  es un estimador insesgado en el sentido que  $E_y(u) = E(\tilde{u})$ .

Hasta ahora desarrollamos los conceptos de BP y BLP para dos VA cuya distribución es  $f(y, u)$ . Cuando trabajamos con un modelo mixto, definido por E.1, existe una variación de estos predictores llamado BLUP, se define como aquel que minimiza el Mean Squared Error (MSE) entre predictores lineales insesgados. Partiendo de E.17 se puede demostrar (Cap 2. McCulloch (2003)):

$$\begin{aligned}
BLUP(U) &= \tilde{u}_{blup} = \mu_U + C\Sigma^{-1}(Y - \mu_Y) \\
&= \Gamma_r Z^T \Sigma^{-1}(Y - X\hat{\beta})
\end{aligned} \tag{E.21}$$

El resultado anterior se obtiene cuando se cumple E.17, tomando el estimador  $X\hat{\beta}_{ML}$ , es decir, el de máxima verosimilitud (Searle et al., 2006, Chap. 7.5).

Por último, queremos hacer notar las diferencias que existen en las hipótesis a la hora de utilizar predictores BP, BLP BLP, BLUP. En en caso de BP exige asumir condiciones más exigentes que BLP y a su vez BLP requiere de más exigencias que BLUP. En el primer caso necesitamos del conocimiento de la distribución conjunta  $f(u|y)$ , mientras que para utilizar BLP alcanza con conocer de  $Y, U$  la esperanza y varianza, en tanto para BLUP es necesario el conocimiento de  $Var(y)$  y  $Cov(U, Y^T)$ . A pesar que hemos hecho uso de la hipótesis de normalidad, ninguno de los tres predictores demanda esta condición.



Esto nos lleva a que la solución  $\hat{\theta}_{REML}$  esta dada por:

$$\hat{\theta}_{REML} = \arg \max_{\theta \in \Theta} l_{REML}(\theta) \quad (\text{E.25})$$

Una solución a este problema de optimización se obtiene utilizando el Teorema de Khatri (Searle et al., 2006, App. M.4f)), el cual establece que si  $K$  es de rango completo,  $KX = 0$  y  $\Sigma$  es definida positiva, entonces:  $P = K(K^T \Sigma K)^{-1} K^T$  para  $P = \Sigma^{-1} - \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}$ . Aplicando este resultado y bajo la hipótesis  $\Sigma = \sum_i \sigma_r^2 Z_i Z_i^T + \sigma_e I_n$ , se llega al sistema de ecuaciones:

$$\left( \text{tr}(\hat{P} Z_i Z_i^T) \right)_{i=1, \dots, q} = \left( \mathbf{y}^T \hat{P}^T Z_i Z_i^T \hat{P} \mathbf{y} \right)_{i=1, \dots, q} \quad (\text{E.26})$$

El proceso antes descrito nos permite llegar a estimaciones con menor sesgo respecto a las que se podrían obtener a través de ML, aunque la complejidad de cómputo es de similares características (McCulloch and Searle, 2001, Chap. 6.9, 6.12), (McCulloch, 2003, Chap. 8).

Llegados a este punto, conociendo  $\hat{\theta}$ , entonces es posible obtener  $\hat{\Gamma}_r$  y  $\hat{\Sigma}$ , por lo tanto, la predicción de  $\hat{U}$  puede realizarse al igual como lo hicimos para  $\Sigma$  conocido, por ejemplo utilizando BLUP, teniendo en cuenta que ahora dicha predicción dependerá de  $\hat{\beta}$  y  $\hat{\Sigma}$ , es decir,  $\hat{U} = \hat{\Gamma}_r Z^T \hat{\Sigma}^{-1} (\mathbf{y} - X \hat{\beta})$ .

Efectos Fijos	Efectos Aleatorios
$\hat{\beta} = (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1} \mathbf{y}$	$\hat{U} = \hat{\Gamma}_r Z^T \hat{\Sigma}^{-1} (\mathbf{y} - X \hat{\beta})$
$\Sigma$ Desconocido	$\hat{\Sigma} = \begin{matrix} \hat{\Sigma}_{ML} \\ \hat{\Sigma}_{REML} \end{matrix}$

Tabla E.2: Estimación de efectos fijos y predicción de efectos aleatorios en LMM cuando  $\Sigma$  es desconocida.

Resumiendo, la Tabla E.2 incluye los resultados antes analizados para de  $\hat{\beta}, \hat{U}$  cuando  $\Sigma$  es desconocido.

### E.1.3. Ecuaciones de Henderson

Un resultado ampliamente difundido en la literatura de LMM corresponde a las llamadas ecuaciones del modelo mixto, también conocidas como ecuaciones de Henderson (Henderson, 1953; Vleck, 1998), las cuales permiten obtener BLUP y ML para  $\beta$  y  $\mathbf{u}$  en forma simultánea.

Hemos visto que partiendo de E.4 se llega a E.5, la distribución de  $Y|U \sim N(X\beta + ZU, \Sigma_e)$  y  $U \sim N(0, \Gamma_r)$ . Con esta información el enfoque de Henderson fue el siguiente: a partir de la distribución conjunta  $f(\mathbf{y}, \mathbf{u})$ , maximizar la función de verosimilitud  $L(\beta, \mathbf{u}; \mathbf{y})$  y obtener  $(\hat{\beta}, \hat{\mathbf{u}})$ .

El desarrollo que permite llegar a las ecuaciones de Henderson se encuentra en Czado (2018); Davison (2003); McCulloch (2003); McCulloch and Searle (2001); Wu (2009). A continuación presentaremos un breve esbozo.

Conociendo  $f(\mathbf{y}|\mathbf{u})$  y  $f(\mathbf{u})$  se obtiene  $f(\mathbf{y}, \mathbf{u})$ , a partir de ésta llegamos a la función de verosimilitud  $L(\beta, \mathbf{u}; \mathbf{y})$ :

$$f(\mathbf{y}, \mathbf{u}) = f(\mathbf{y}|\mathbf{u}) \cdot f(\mathbf{u}) \quad (\text{E.27})$$

$$L(\beta, \mathbf{u}; \mathbf{y}) \propto e^{\{-\frac{1}{2}(\mathbf{y}-X\beta-Z\mathbf{u})^T \Sigma_e^{-1}(\mathbf{y}-X\beta-Z\mathbf{u})\}} \cdot e^{\{-\frac{1}{2}\mathbf{u}^T \Gamma_r^{-1} \mathbf{u}\}}$$

Buscamos maximizar  $L(\beta, \mathbf{u}; \mathbf{y})$  que nos permita obtener  $(\hat{\beta}, \hat{\mathbf{u}})$ , para ello utilizamos el logaritmo de la función de verosimilitud  $l(\beta, \mathbf{u}; \mathbf{y})$ , esto nos conduce a:

$$l(\beta, \mathbf{u}; \mathbf{y}) \propto -\frac{1}{2} \left( (\mathbf{y} - X\beta - Z\mathbf{u})^T \Sigma_e^{-1} (\mathbf{y} - X\beta - Z\mathbf{u}) + (\mathbf{u}^T \Gamma_r^{-1} \mathbf{u}) \right) \quad (\text{E.28})$$

$$(\hat{\beta}, \hat{\mathbf{u}}) = \arg \min_{\substack{\beta \in \mathbb{R}^p \\ \mathbf{u} \in \mathbb{R}^q}} l(\beta, \mathbf{u}; \mathbf{y}) \quad (\text{E.29})$$

Esto nos conduce a las ecuaciones del modelo mixto o ecuaciones de Henderson:

$$\begin{pmatrix} X^T \Sigma_e^{-1} X & X^T \Sigma_e^{-1} Z \\ Z^T \Sigma_e^{-1} X & Z^T \Sigma_e^{-1} Z + \Gamma_r \end{pmatrix} \cdot \begin{pmatrix} \hat{\beta} \\ \hat{\mathbf{u}}_{blup} \end{pmatrix} = \begin{pmatrix} X^T \Sigma_e^{-1} \mathbf{y} \\ Z^T \Sigma_e^{-1} \mathbf{y} \end{pmatrix} \quad (\text{E.30})$$

El sistema de ecuaciones puede ser resuelto cuando se tiene conocimiento de  $\Sigma_e$  y  $\Gamma_r$ , como hemos mencionado antes en la práctica esto no sucede, por lo tanto, métodos de iteración son utilizados en la búsqueda de soluciones.

Esta página ha sido intencionalmente dejada en blanco.

## Apéndice F

# Test de Comportamiento - Identificación de TEA

### F.1. El Test

El conjunto de datos elegido y utilizado en el Cap. 5 es el denominado “Autistic Spectrum Disorder Screening Data for Toddlers” Thabtah (2020). La forma en que se han obtenido las respuestas al test de comportamiento para la identificación del TEA, es a través de la plataforma ASDTests (Thabtah, 2017b).

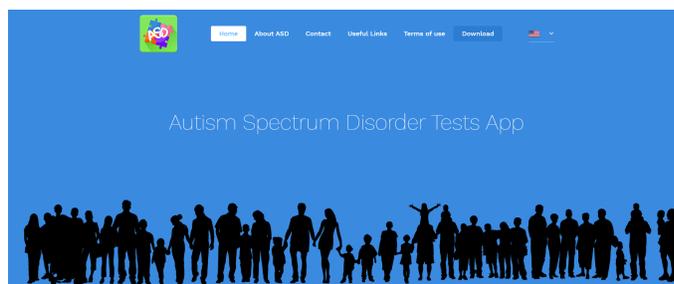


Figura F.1: Aplicación ASDTests disponible en <https://www.asdtests.com/#>.

Complementando la información presentada en el Cap. 5, la Tabla F.1 muestra las preguntas, en su idioma de origen, realizadas en el test.

Tabla F.1: Preguntas en el Test.

Variable in Dataset	Type	Corresponding Q-chat-10-Toddler Features
A1	Binary {0, 1}	Does your child look at you when you call his/her name?
A2	Binary {0, 1}	How easy is it for you to get eye contact with your child?
A3	Binary {0, 1}	Does your child point to indicate that s/he wants something? (e.g. a toy that is out of reach)
A4	Binary {0, 1}	Does your child point to share interest with you? (e.g. pointing at an interesting sight)
A5	Binary {0, 1}	Does your child pretend? (e.g. care for dolls, talk on a toy phone)
A6	Binary {0, 1}	Does your child follow where you're looking?
A7	Binary {0, 1}	If you or someone else in the family is visibly upset, does your child show signs of wanting to comfort them? (e.g. stroking hair, hugging them)
A8	Binary {0, 1}	Would you describe your child's first words as:
A9	Binary {0, 1}	Does your child use simple gestures? (e.g. wave goodbye)
A10	Binary {0, 1}	Does your child stare at nothing with no apparent purpose?

# Apéndice G

## Resultados Complementarios

### G.1. Análisis Descriptivo

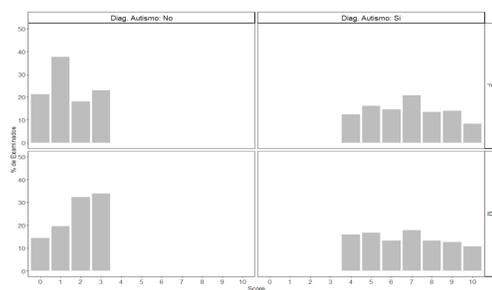


Figura G.1: Porcentaje de examinados diagnosticados, a través del test, con TEA agrupados por sexo.

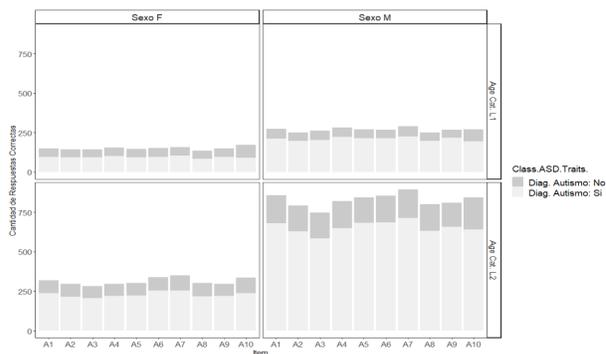


Figura G.2: Cantidad de respuestas correctas, según el diagnóstico recibido, agrupadas por Sexo y categoría de edad.

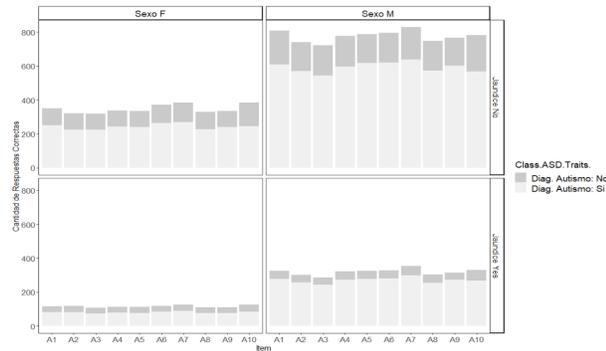


Figura G.3: Cantidad de respuestas correctas, según el diagnóstico recibido, agrupadas por Sexo e ictericia (Jaundice).

Medidas de resumen de las variables involucradas en el análisis.

```
> summary(df_data_all)
```

```

      ID          A1          A2          A3
Min.   : 1.0    Min.   :0.00    Min.   :0.0000    Min.   :0.000
1st Qu.: 267.0  1st Qu.:0.00    1st Qu.:0.0000    1st Qu.:0.000
Median : 531.0  Median :1.00    Median :0.0000    Median :0.000
Mean   : 528.8  Mean   :0.56    Mean   :0.4478    Mean   :0.402
3rd Qu.: 791.0  3rd Qu.:1.00    3rd Qu.:1.0000    3rd Qu.:1.000
Max.   :1054.0  Max.   :1.00    Max.   :1.0000    Max.   :1.000

      A4          A5          A6          A7
Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:0.0000
Median :1.0000    Median :1.0000    Median :1.0000    Median :1.0000
Mean   :0.5141    Mean   :0.5249    Mean   :0.5737    Mean   :0.6537
3rd Qu.:1.0000    3rd Qu.:1.0000    3rd Qu.:1.0000    3rd Qu.:1.0000
Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   :1.0000

      A8          A9          A10         Age_Mons
Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :12.00
1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:23.00
Median :0.0000    Median :0.0000    Median :1.0000    Median :30.00
Mean   :0.4556    Mean   :0.4898    Mean   :0.5834    Mean   :27.88
3rd Qu.:1.0000    3rd Qu.:1.0000    3rd Qu.:1.0000    3rd Qu.:36.00
Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   :36.00

      Score      Sex      Ethnicity  Jaundice
Min.   : 0.000  f:314  White European:329  no :742
1st Qu.: 3.000  m:711  asian                :291  yes:283
Median : 5.000                middle eastern:181
Mean   : 5.205                south asian   : 57
3rd Qu.: 8.000                black         : 51
Max.   :10.000                Hispanic     : 39
(Other)      : 77

Family_mem_with_ASD  Who.completed.the.test  Class.ASD.Traits.  Age_cat
no :861              family member:1018      No :317            L1:276
yes:164              Others          : 3      Yes:708           L2:749
Self                : 4

```

## G.2. Modelo de Rasch

A continuación se presentan resultados complementarios sobre el denominado modelo m\_rasch.

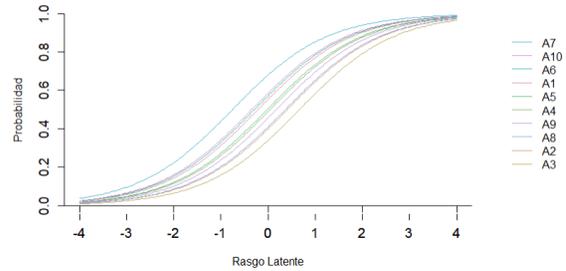


Figura G.4: Curvas características de los ítems.

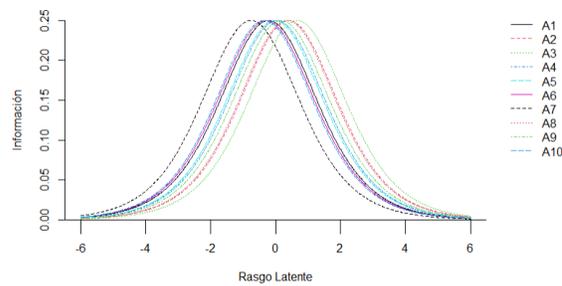


Figura G.5: Curvas de información vs Rasgo Latente.

La tabla G.1 muestra el resultado del Test de Wald aplicado al modelo de Rasch. Observar que cuando  $p - value < \alpha$ , entonces, existe indicios que los ítems difieren significativamente entre los examinados con score por encima y por debajo del score medio.

Tabla G.1: Resultados del Test de Wald aplicado al modelo m\_rasch.

	$z$	$Pr(> Chisq)$
$\beta_{A1}$	0.204	0,838
$\beta_{A2}$	0.607	0,544
$\beta_{A3}$	-0.190	0,850
$\beta_{A4}$	-1.587	0,1131
$\beta_{A5}$	-1.628	0,103
$\beta_{A6}$	-3.339	< 0,01
$\beta_{A7}$	-1.806	< 0,01
$\beta_{A8}$	2.325	0,071
$\beta_{A9}$	-3.356	< 0,01
$\beta_{A10}$	10.691	< 0,01

```

> summary(m_rasch)

Results of RM estimation:

Call:  eRm::RM(X = df_data_all %>% select(A1:A10))

Conditional log-likelihood: -3752.322
Number of iterations: 8
Number of parameters: 9

Item (Category) Difficulty Parameters (eta): with 0.95 CI:

Item Easiness Parameters (beta) with 0.95 CI:
Estimate Std. Error lower CI upper CI
beta A1    0.217      0.070    0.079    0.354
beta A2   -0.405      0.070   -0.542   -0.267
beta A3   -0.663      0.071   -0.802   -0.524
beta A4   -0.038      0.070   -0.175    0.099
beta A5    0.021      0.070   -0.115    0.158
beta A6    0.294      0.070    0.156    0.432
beta A7    0.759      0.073    0.615    0.902
beta A8   -0.361      0.070   -0.499   -0.224
beta A9   -0.173      0.070   -0.309   -0.036
beta A10   0.349      0.071    0.210    0.487

Goodness-of-Fit Results:
Collapsed Deviance = 417.909 (df = 90, p-value = 0)
Pearson R2: 0.275
Area Under ROC: 0.802

```

## G.3. Modelo de Rasch de Regresión Latente - Inferencia Clásica

```
> summary(m2_rasch_glm)

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Resp ~ -1 + Item + Sex + (1 | ID)
Data: df_data_all_long
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+06))

AIC      BIC      logLik deviance df.resid
12424.5  12511.4  -6200.3  12400.5   10238

Scaled residuals:
Min      1Q  Median      3Q      Max
-3.2892 -0.6994  0.2948  0.6882  3.1090

Random effects:
Groups Name      Variance Std.Dev.
ID      (Intercept) 1.931    1.39
Number of obs: 10250, groups: ID, 1025

Fixed effects:
Estimate Std. Error z value Pr(>|z|)
ItemA1    0.04904    0.11338    0.433  0.66533
ItemA2   -0.57264    0.11364   -5.039  4.68e-07 ***
ItemA3   -0.83089    0.11445   -7.260  3.88e-13 ***
ItemA4   -0.20552    0.11319   -1.816  0.06942 .
ItemA5   -0.14618    0.11319   -1.291  0.19654
ItemA6    0.12567    0.11351    1.107  0.26825
ItemA7    0.58951    0.11516    5.119  3.07e-07 ***
ItemA8   -0.52921    0.11353   -4.661  3.14e-06 ***
ItemA9   -0.34019    0.11326   -3.004  0.00267 **
ItemA10  0.18073    0.11363    1.591  0.11171
Sexm     0.40421    0.10792    3.745  0.00018 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
ItemA1 ItemA2 ItemA3 ItemA4 ItemA5 ItemA6 ItemA7 ItemA8 ItemA9 ItemA10
ItemA2    0.587
ItemA3    0.583  0.587
ItemA4    0.589  0.590  0.586
ItemA5    0.589  0.590  0.586  0.591
ItemA6    0.586  0.586  0.581  0.588  0.588
ItemA7    0.576  0.573  0.568  0.577  0.577  0.576
ItemA8    0.588  0.590  0.588  0.590  0.590  0.586  0.574
ItemA9    0.589  0.591  0.587  0.591  0.591  0.588  0.576  0.591
ItemA10   0.586  0.585  0.580  0.587  0.587  0.585  0.575  0.585  0.587
Sexm     -0.661 -0.663 -0.660 -0.663 -0.663 -0.659 -0.646 -0.664 -0.664 -0.658
```

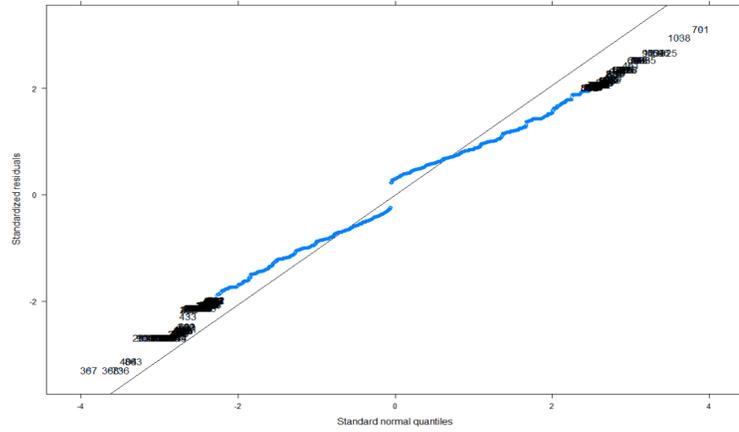


Figura G.6: Residuos del modelo `m2_rasch_glm`.

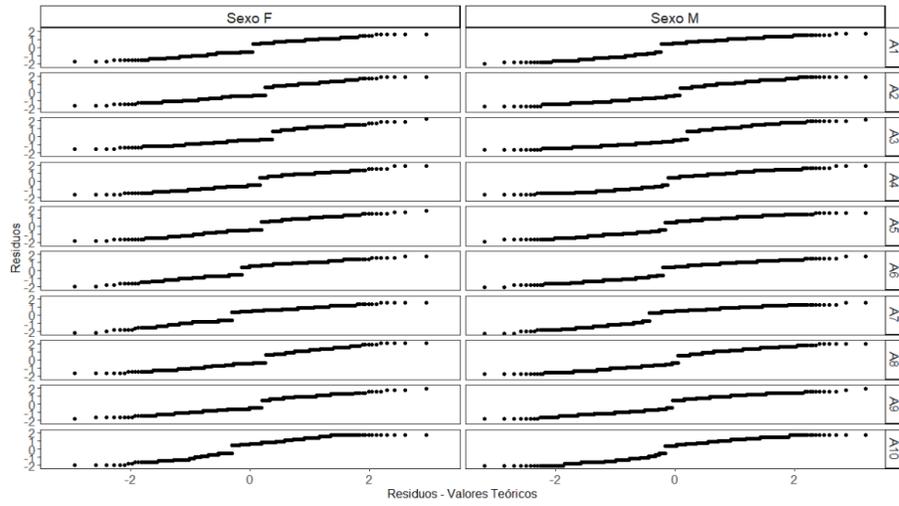


Figura G.7: Residuos del modelo `m2_rasch_glm` agrupados por Sexo.

## G.4. Modelo de Rasch de Regresión Latente - Inferencia Bayesiana

```
> summary(m4_rasch_B_glm)

Family: bernoulli
Links: mu = logit
Formula: Resp ~ -1 + Item + Sex + Age_cat + Jaundice + (1 | ID)
Data: df_data_all_long (Number of observations: 10250)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
total post-warmup draws = 4000

Group-Level Effects:
~ID (Number of levels: 1025)
Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)      1.43      0.05      1.34      1.53 1.00      1431      2107

Population-Level Effects:
Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
ItemA1      0.03      0.14     -0.25      0.32 1.00      801      1820
ItemA2     -0.59      0.15     -0.88     -0.30 1.00      819      1823
ItemA3     -0.86      0.14     -1.13     -0.56 1.00      824      1957
ItemA4     -0.23      0.15     -0.50      0.08 1.00      797      1876
ItemA5     -0.17      0.15     -0.44      0.12 1.00      798      1777
ItemA6      0.11      0.15     -0.18      0.39 1.00      799      1790
ItemA7      0.57      0.15      0.29      0.86 1.00      792      1987
ItemA8     -0.55      0.15     -0.83     -0.26 1.01      764      1963
ItemA9     -0.36      0.14     -0.64     -0.07 1.01      720      1866
ItemA10     0.16      0.15     -0.12      0.45 1.00      814      1769
Sexm        0.41      0.11      0.19      0.62 1.00     1137     2169
Age_catL2   -0.01      0.12     -0.25      0.22 1.00     1170     2103
Jaundiceyes 0.11      0.11     -0.11      0.33 1.00     1694     2310

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

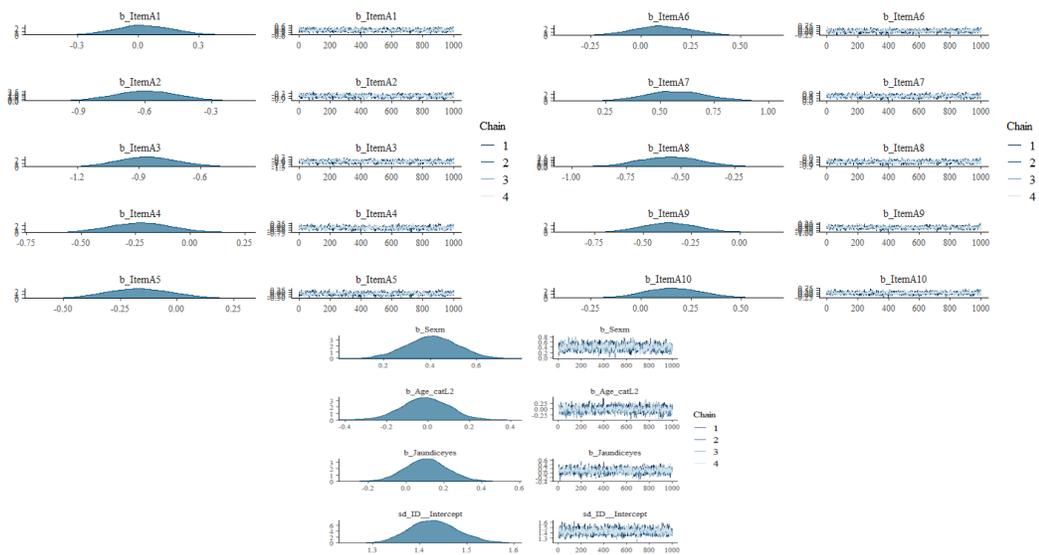


Figura G.8: Distribución a posteriori de los parámetros y convergencia de las cadenas.

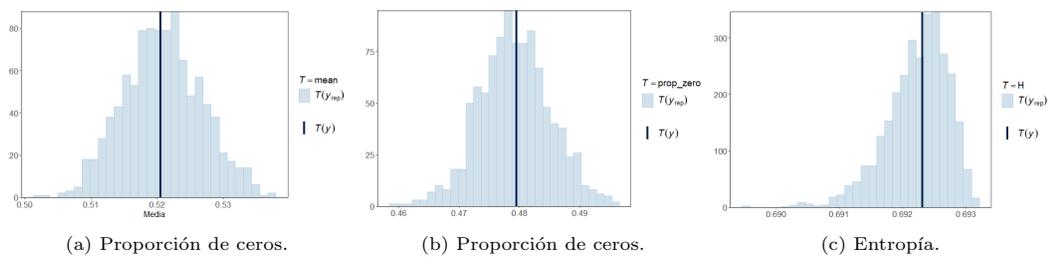
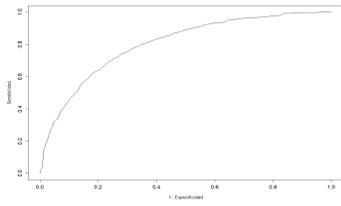
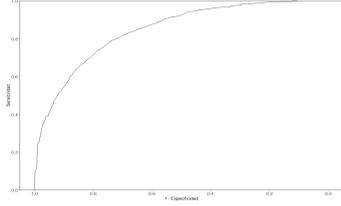


Figura G.9: Predictivas a posteriori.

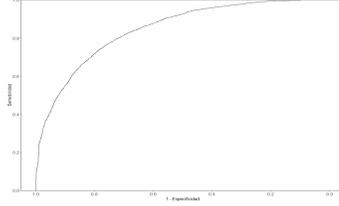
## G.5. Curva ROC



(a) Mod. de Rasch.



(b) Mod. de Rasch GLMM.



(c) Mod. de Rasch bayesiano GLMM.

Figura G.10: Curva ROC.

Esta página ha sido intencionalmente dejada en blanco.

# Apéndice H

## Código en R

```
#-----  
# Setup  
#-----  
  
# Semilla  
set.seed(13579)  
  
# Bibliotecas  
library(tidyverse) # data processing  
library(ggplot2)   # plots  
library(lme4)      # GLMM  
library(eRm)       # Rasch Model  
library(brms)      # Bayesian GLMM  
library(loo)       # loo  
library(car)       # Anova  
library(caret)     # confusion matrix, model performance  
library(pROC)      # plot ROC  
  
#-----  
# Funciones  
#-----  
  
## Proporción de Ceros  
prop_zero <- function(x) mean(as.numeric(x) == 0)  
  
## Entropía  
H <- function(y){  
H <- (sum(y)/length(y))*log((sum(y)/length(y))^{-1}) +  
(1-sum(y)/length(y))*log((1-sum(y)/length(y))^{-1})  
return(H)  
}
```

```

## DIC
f_dic <- function(y_data, y_pbinom_m, log_lik_m){

logLikelihood <- function(y, prob) sum(dbinom(y, 1, prob, log=T))

pDIC <- 2*(logLikelihood(y_data, y_pbinom_m) - mean(log_lik_m))

dic <- -2*logLikelihood(y_data, y_pbinom_m) + 2*pDIC

return(dic)
}

#-----
# Dataset
#-----

# Los datos se encuentran disponibles en:
# https://www.kaggle.com/datasets/fabdelja/autism-screening-for-toddlers

df_data_all <- read.csv("Toddler_Autism_dataset_July_2018.csv",
sep = ",", header = TRUE)

df_data_all <- as_tibble(df_data_all)

## Transformación de variables

df_data_all <- df_data_all %>%
mutate(Sex = as.factor(Sex),
Ethnicity = as.factor(Ethnicity),
Jaundice = as.factor(Jaundice),
Family_mem_with_ASD = as.factor(Family_mem_with_ASD),
Who.completed.the.test = as.factor(Who.completed.the.test),
Class.ASD.Traits. = as.factor(Class.ASD.Traits.))

f_data_all <- df_data_all %>%
rename(Score = Qchat.10.Score) %>%
rename(ID = Case_No)

df_data_all <-
df_data_all %>%
mutate(Age_Mons_Norm = (Age_Mons - min(Age_Mons))/(max(Age_Mons) - min(Age_Mons)))

df_data_all <-

```

```

df_data_all %>%
mutate(Age_cat = ifelse(Age_Mons < 24 , "L1", "L2")) %>%
mutate(Age_cat = as.factor(Age_cat))

## Limpieza de datos

df_data_all <- df_data_all %>%
filter(!Who.completed.the.test %in% c("Health care professional", "Health Care Professional"))

df_data_all$Who.completed.the.test <- droplevels(df_data_all$Who.completed.the.test)

## Datos 'Long'

df_data_all_long <-
df_data_all %>%
select(ID, A1:A10, Jaundice, Sex, Class.ASD.Traits., Age_Mons_Norm, Age_cat) %>%
# Transformo a formato 'long'
pivot_longer(!c(ID, Jaundice, Sex, Class.ASD.Traits., Age_Mons_Norm, Age_cat),
names_to = "Item", values_to = "Resp")

df_data_all_long <- df_data_all_long %>%
mutate(Item = factor(Item,
levels = c("A1", "A2", "A3", "A4", "A5",
"A6", "A7", "A8", "A9", "A10")),
Resp = as.factor(Resp))

#-----
# Rasch Model
#-----

m_rasch_eRm <- eRm::RM(df_data_all %>% select(A1:A10))

m_rasch_eRm_gofit <- eRm::gofIRT(eRm::person.parameter(m_rasch_eRm))

eRm::Waldtest(m_rasch_eRm, splitcr = "mean")

## M. Confusión & AUC

summary(m_rasch_eRm_gofit)

#-----
# Rasch Model - GLMM
#-----

## Modelos

```

```

m1_rasch_glmm <- glmer(Resp ~ -1 + Item + (1 | ID),
family = binomial,
data = df_data_all_long,
control = glmerControl(optimizer = "bobyqa",
optCtrl = list(maxfun = 100000)))

m2_rasch_glmm <- glmer(Resp ~ -1 + Item + Sex + (1 | ID),
family = binomial,
data = df_data_all_long,
control = glmerControl(optimizer = "bobyqa",
optCtrl = list(maxfun = 1e6)))

m3_rasch_glmm <- glmer(Resp ~ -1 + Item + Sex + Age_cat + (1 | ID),
family = binomial,
data = df_data_all_long,
control = glmerControl(optimizer = "bobyqa",
optCtrl = list(maxfun = 1e6)))

m4_rasch_glmm <- glmer(Resp ~ -1 + Item + Sex + Age_cat + Jaundice + (1 | ID),
family = binomial,
data = df_data_all_long,
control = glmerControl(optimizer = "bobyqa",
# optimizer = "Nelder_Mead",
optCtrl = list(maxfun = 1e6)))

## Check Efectos fijos

car::Anova(m2_rasch_glmm, type="II")

## Comparación de Modelos

anova(m1_rasch_glmm,
m2_rasch_glmm,
m3_rasch_glmm,
m4_rasch_glmm,
test="Chi")

## IdC y Bootstrapping

m2_rasch_glmm_idc <-
confint(m2_rasch_glmm, method = "Wald") %>%
as.data.frame() %>%

```

```

rownames_to_column() %>%
rename(Item = rowname,
Q2.5 = '2.5 %',
Q97.5 = '97.5 %') %>%
slice(2:n()) %>%
mutate(Est = item_par_m2_rasch_glmm$Est) %>%
mutate(Metodo = "Wald") %>%
select(Item, Est, Q2.5,Q97.5,Metodo)

## Test Normalidad

ks.test(residuals(m2_rasch_glmm), "pnorm")
tseries::jarque.bera.test(residuals(m2_rasch_glmm))

## M. Confusión & AUC

m2_rasch_glmm_roc <-
pROC::roc(response = df_data_all_long$Resp,
# response = df_valid_long$Resp,
predictor = predict(m2_rasch_glmm,
df_data_all_long,
type = "response"))

#-----
# Rasch Model - GLMM Bayesian
#-----

## Modelos

prior1_lv_glmm <- prior("normal(0, 10)", class = "b") +
prior("normal(0, 20)", class = "sd", group = "ID")

bf_y1 <- bf( Resp ~ -1 + Item + (1 | ID) )

m1_rasch_lv_glmm <- brm(formula = bf_y1,
family = brmsfamily("bernoulli", link = "logit"),
prior = prior1_lv_glmm,
data = df_data_all_long,
file = "Models_fulldataset/m1_rasch_lv_glmm",
chains = 4,
iter = 2000,
warmup = 1000,
)

prior2_lv_glmm <- prior("normal(0, 10)", class = "b") +

```

```

prior("normal(0, 20)", class = "sd", group = "ID")

bf_y2 <- bf( Resp ~ -1 + Item + Sex + (1 | ID) )

m2_rasch_lv_glmm <- brm(formula = bf_y2,
family = brmsfamily("bernoulli", link = "logit"),
prior = prior2_lv_glmm,
data = df_data_all_long,
file = "Models_fullldataset/m2_rasch_lv_glmm",
chains = 4,
iter = 2000,
warmup = 1000
)

prior3_lv_glmm <- prior("normal(0, 10)", class = "b") +
prior("normal(0, 20)", class = "sd", group = "ID")

bf_y3 <- bf( Resp ~ -1 + Item + Sex + Age_cat + (1 | ID) )

m3_rasch_lv_glmm <- brm(formula = bf_y3,
family = brmsfamily("bernoulli", link = "logit"),
prior = prior3_lv_glmm,
data = df_data_all_long,
file = "Models_fullldataset/m3_rasch_lv_glmm",
chains = 4,
iter = 2000,
warmup = 1000
)

prior4_lv_glmm <- prior("normal(0, 10)", class = "b") +
prior("normal(0, 20)", class = "sd", group = "ID")

bf_y4 <- bf( Resp ~ -1 + Item + Sex + Age_cat + Jaundice + (1 | ID) )

m4_rasch_lv_glmm <- brm(formula = bf_y4,
family = brmsfamily("bernoulli", link = "logit"),
prior = prior4_lv_glmm,
data = df_data_all_long,
file = "Models_fullldataset/m4_rasch_lv_glmm",
chains = 4,
iter = 2000,
warmup = 1000
)

```

```

## Plot Estimación y Cadenas, ppcheck

plot(m2_rasch_lv_glmm, ask=FALSE)

pp_check(m4_rasch_lv_glmm, type = "stat",
stat = "mean", ndraws = 1000) +
xlab("Media") +
theme_bw()

pp_check(m4_rasch_lv_glmm, type = "stat",
stat = "prop_zero", ndraws = 1000) +
theme_bw()

pp_check(m4_rasch_lv_glmm, type = "stat",
stat = "H", ndraws = 3000) +
theme_bw()

## Selección de Modelo

y_pbinom_m4 <-predict(m4_rasch_lv_glmm) %>%
as.data.frame() %>%
select(estimate) %>%
pull()

log_lik_m4 <-
as_draws_matrix(m4_rasch_lv_glmm) %>%
as.data.frame() %>%
select(lp__) %>%
pull()

dic_m4 <- f_dic(as.numeric(df_data_all_long$Resp) -1,
y_pbinom_m4,
log_lik_m4)

loo_models <- loo::loo_compare(loo::loo(m1_rasch_lv_glmm),
loo::loo(m2_rasch_lv_glmm),
loo::loo(m3_rasch_lv_glmm),
loo::loo(m4_rasch_lv_glmm))

## M. Confusión & AUC

m4_rasch_lv_glmm_roc <-
pROC::roc(response = df_data_all_long$Resp,
# response = df_valid_long$Resp,
predictor = predict(m4_rasch_lv_glmm,
type = "response")[ , "Estimate"])

```

Esta página ha sido intencionalmente dejada en blanco.

# Bibliografía

- D. Abbakumov, P. Desmet, and W. Van den Noortgate. Measuring growth in students' proficiency in MOOCs: Two component dynamic extensions for the Rasch model. *Behavior Research Methods*, 51:332–341, 2019. 10.3758/s13428-018-1129-1.
- M. Aitkin and N. Longford. Statistical Modelling Issues in School Effectiveness Studies. *Journal of the Royal Statistical Society Series A*, 149(1):1–43, 1986. 10.2307/2981882.
- M. Al-diabat. Fuzzy Data Mining for Autism Classification of Children. *International Journal of Advanced Computer Science and Applications*, 9(7):11–17, 2018.
- J. H. Albert. Bayesian estimation of normal ogive item response curves using gibbs sampling. *Journal of Educational Statistics*, 17(3):251–269, 1992.
- I. Alvarez-Castro. Notas del Curso Inferencia II - Bayesiana, Instituto de Estadística (IESTA). Facultad de Ciencias Económicas y Administración (FCEA). UdelaR., 2021.
- F. Alvarez-Vaz, R. and Massa. Evaluación de la Salud Bucal a través de la Teoría de la Respuesta al Ítem en un Estudio Poblacional en Uruguay. XI Semana Internacional de la Estadística y la Probabilidad, Septiembre 2020.
- R. Alvarez-Vaz, F. Massa, S. Lorenzo, and A. Fabruccini. International Association for Dental Research. VII Congreso de la Región Latinoamericana. Montevideo, Uruguay, 08 2018.
- E. B. Andersen. Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society Series B (Methodological)*, 32(2):283–301, 1970.
- E. B. Andersen. Conditional Inference for Multiple-Choice Questionnaires. *British Journal of Mathematical and Statistical Psychology*, 26:31–44, 1973a.
- E. B. Andersen. A goodness of fit test for the Rasch model. *Psychometrika*, 38:123–140, 1973b.
- E. B. Andersen. Sufficient Statistics and Latent Trait Models. *Psychometrika*, 42:221–227, March 1977.
- E. B. Andersen and L. Wøhlk. *Essays on Item Response Theory*, volume 157 of *Lecture Notes in Statistics*, chapter The Life of Georg Rasch as a Mathematician and as a Statistician, pages 3–24. Springer, New York, NY, 2001. 10.1007/978-1-4613-0169-1\_1.
- D. Andrich. Controversy and the Rasch model: a characteristic of incompatible paradigms? *Medical Care*, 42(1):I7–I16, Jan 2004. 10.1097/01.mlr.0000103528.48582.7c.
- M. Ard, D. Galasko, and S. Edland. Improved Statistical Power of Alzheimer Clinical Trials by Item-Response Theory: Proof of Concept by Application to the Activities of Daily Living Scale. *Alzheimer Dis Assoc Disord.*, 27(2):187–191, 2013. 10.1097/WAD.0b013e318265bcc1.
- C. Ausín. Notas del Curso Simulación Monte Carlo y Métodos Bayesianos, 2012. Universidad Carlos III. Madrid.
- F. B. Baker. *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation, 2 edition, 2001.

- F. B. Baker and S.-H. Kim. *Item Response Theory. Parameter Estimation Techniques*. Taylor & Francis Group, Boca Raton, FL, 2 edition, 2004.
- M. Bala, M. H. Ali, M. S. Satu, K. F. Hasan, and M. A. Moni. Efficient Machine Learning Models for Early Stage Detection of Autism Spectrum Disorder. *Algorithms*, 15(5):166, 2022.
- M. F. Balsells. Análisis longitudinal de los factores asociados a la pérdida de densidad mineral ósea en pacientes infectados con VIH. Master’s thesis, Departamento de Estadística e Investigación Operativa. Universitat Politècnica de Catalunya, Barcelona, España, 2009.
- S. Balsis, A. A. Unger, J. F. Bengue, L. Geraci, and R. S. Doody. Gaining precision on the Alzheimer’s Disease Assessment Scale-cognitive: A comparison of item response theory-based scores and total scores. *Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association*, 8(4): 288–294, Jul. 2012.
- S. Balsis, T. K. Choudhury, J. F. Geraci, L. Bengue, and C. J. Patrick. Alzheimer’s Disease Assessment: A Review and Illustrations Focusing on Item Response Theory Techniques. *Assessment*, 25(3):360–373, 2018.
- M. I. Barbero García, E. García Cueto, F. P. Holgado Tello, and E. Vila Abad. *Psicometría*. Sanz y Torres, S.L., Alcorcón, Madrid, 1 edition, 2015.
- O. Barndorff-Nielsen. *Information and Exponential Families In Statistical Theory*. John Wiley & Sons, 2014.
- F. Bartolucci, S. Bacci, and M. Gnaldi. *Statistical Analysis of Questionnaires. A Unified Approach Based on R and Stata*. Mathematics & Statistics. Chapman & Hall/CRC, 1 edition, 2015.
- D. Bates. Computational methods for mixed models, March 27, 2018. <https://cran.r-project.org/web/packages/lme4/vignettes/Theory.pdf> – last visited 15<sup>th</sup> of Sep, 2019.
- D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. 10.18637/jss.v067.i01.
- C. Baumgartner. Comparison of Multivariate Mixed and Copula Models for the Estimation of Losses. Master’s thesis, Department of Mathematics, Technische Universität München, München, Germany, 2013.
- J. Berkson. Why I Prefer Logits to Probits. *Biometrics*, 7(4):327–339, 1951.
- J. Berkson. A Statistically Precise and Relatively Simple Method of Estimating the Bioassay with Quantal Response, Based on the Logistic Function. *Journal of the American Statistical Association*, 48(263):565–599, 1953.
- J. Berkson. Tables for the Maximum Likelihood Estimate of the Logistic Function. *Biometrics*, 13(1):28–34, 1957.
- G. Birbeck, K. Sehyun, R. Hays, and B. Vickrey. Quality of life measures in epilepsy: How well can they detect change over time? *Neurology.*, 54(9):1822–1827, 2000.
- A. Birnbaum. Efficient design and use of tests of a mental ability for various decision making problems. Series Report No. 58-16, Project No. 7755-23, USAF School of Aviation Medicine, Randolph Air Force Base, Texas, January 1957.
- A. Birnbaum. On the estimation of mental ability. Series Report No. 15. Project No. 7755-23, USAF School of Aviation Medicine, Randolph Air Force Base, Texas, 1958a.
- A. Birnbaum. Further considerations of efficiency in tests of a mental ability. Technical Report No. 17. Project No. 7755-23, USAF School of Aviation Medicine, Randolph Air Force Base, Texas, 1958b.
- A. Birnbaum. Some Latent Trait Models. In *Statistical Theories of Mental Test Scores, with contributions by A. Birnbaum*, chapter 17, pages 397–423. Addison-Wesley, 1968.
- R. D. Bock and M. Aitkin. Marginal Maximum Likelihood estimation of item parameters: Appli-

- cation of an EM algorithm. *Psychometrika*, 46(4):443–459, 1981.
- R. D. Bock and M. Lieberman. Fitting a response model for  $n$  dichotomously scored items. *Psychometrika*, 35:179–197, 1970. 10.1007/BF02291262.
- G. Boente and V. Yohai. Notas de Estadística. Universidad de Buenos Aires. <http://cms.dm.uba.ar/academico/materias/2docuat2018/estadistica-teorica/> – last visited 15<sup>th</sup> of March, 2022.
- S. Brandt, M. Moulton, and B. Duckor. Advances in Rasch modeling: New applications and directions. *Psychological Test and Assessment Modeling*, 57:338–341, 01 2015.
- N. E. Breslow and D. G. Clayton. Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 88(421):9–25, 1993. 10.2307/2290687.
- S. Brooks, G. Gelman, G. L. Jones, and X.-L. Meng, editors. *Handbook of Modern Statistical Methods*. Mathematics & Statistics. Chapman & Hall/CRC, 2011.
- J. Brunner. Lectures Notes. STA 2101: Methods of Applied Statistics I. Department of Statistics. University of Toronto, 2012.
- P.-C. Bürkner. brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1):1–28, 2017. 10.18637/jss.v080.i01.
- P.-C. Bürkner. Bayesian item response modeling in r with brms and stan. *arXiv preprint arXiv:1905.09501*, 2019.
- L. Cai and D. Thissen. Modern Approaches to Parameter Estimation in Item Response Theory. In S. P. Reise and D. A. Revicki, editors, *Handbook of Item Response Theory Modeling*, chapter 3, pages 41–59. Routledge, 2014. 10.4324/9781315736013.
- J. M. Casabianca and C. Lewis. IRT Item Parameter Recovery With Marginal Maximum Likelihood Estimation Using Loglinear Smoothing Models. *Journal of Educational and Behavioral Statistics*, 40(6):547–578, 2015. 10.3102/1076998615606112.
- G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Advanced Series. Duxbury. Thomson Learning Inc., 2 edition, 2002.
- J. M. Chambers, W. S. Cleveland, K. B., and P. A. Tukey. *Graphical Methods for Data Analysis*. Chapman and Hall/CRC, 1 edition, 2017. 10.1201/9781351072304.
- . Columbia Public Health. Population Health Methods. Item Response Theory. Columbia University Irving Medical Center. <https://www.publichealth.columbia.edu/research/population-health-methods/item-response-theory> – last visited 15<sup>th</sup> of March, 2022.
- R. Cordier, N. Munro, S. Wilkes-Gillan, S. R, L. Parsons, and A. Joosten. Applying Item Response Theory (IRT) Modeling to an Observational Measure of Childhood Pragmatics: The Pragmatics Observational Measure-2. *Front Psychol.*, 10(408), Feb. 2019. 10.3389/fpsyg.2019.00408.
- C. Czado. MA3403: Generalized Linear Models. Department of Mathematics. Technical University of Munich, 2018. <https://www.groups.ma.tum.de/en/statistics/home/> – last visited 15<sup>th</sup> of September, 2022.
- A. C. Davison. *Statistical Models*. Cambridge University Press, The Edinburgh Building, Cambridge CB2 8RU, UK, 1 edition, 2003.
- R. J. De Ayala. *The Theory and Practice of Item Response Theory*. Methodology in the Social Sciences. The Guilford Publications, 370 Seventh Avenue, Suite 1200, New York, 2009.
- R. J. De Ayala. *The Reviewer’s Guide to Quantitative Methods in the Social Sciences*, chapter 11, pages 145–163. Taylor & Francis Group, 2 edition, 2019.
- P. De Boeck and M. Wilson. *Explanatory item response models. A generalized linear and nonlinear approach*. Statistics for Social and Behavioral Sciences. 2004. 10.1007/978-1-4757-3990-9.
- L. Debera and L. Nalbarte. Pruebas diagnósticas: una aplicación a la teoría de respuesta alitem,

- aproximación clásica y bayesiana. Serie documentos de trabajo DT (06/01), Octubre 2006.
- C. DeMars. *Item response theory*. Understanding Statistics. Oxford University Press, Inc., 2010.
- C. D. Desjardins and B. Bulut. *Handbook of Educational Measurement and Psychometrics Using R*. The R Series. Taylor & Francis Group, Boca Raton, FL, 2018.
- P. Dunn-Rankin, G. A. Knezek, S. R. Wallace, and S. Zhang. *Scaling Methods*. Psychology Press, New York, 2 edition, 2004.
- D. P. Ellis. *Item-Analysis Methods and Their Implications for the ILTA Guidelines for Practice: A Comparison of the Effects of Classical Test Theory and Item Response Theory Models on the Outcome of a High-Stakes Entrance Exam*. PhD thesis, Faculty of the Graduate School. University of Maryland, 2011.
- I. Epifani, S. N. MacEachern, and M. Peruggia. Case-deletion importance sampling estimators: Central limit theorems and related results. *Electronic Journal of Statistics*, 2:774–806, 2008.
- U. Erkan and D. N. Thanh. Autism spectrum disorder detection with machine learning methods. *Current Psychiatry Research and Reviews Formerly: Current Psychiatry Reviews*, 15(4):297–308, 2019.
- H. F. and O. Karimi. Approximate composite marginal likelihood inference in spatial generalized linear mixed models. *Journal of Applied Statistics*, 46(3):542–558, 2019. 10.1080/02664763.2018.1506020.
- L. Fahrmeir, T. Kneib, S. Lang, and B. Marx. *Regression Models, Methods and Applications*. Springer, Berlin, Heidelberg, 2013. 10.1007/978-3-642-34333-9.
- J. J. Faraway. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Non-parametric Regression Models*. Texts in Statistical Science. Chapman & Hall/CRC Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL, 2 edition, 2016.
- G. Fischer. Applying the principles of specific objectivity and of generalizability to the measurement of change. *Psychometrika*, 52:565–587, December 1987. 10.1007/BF02294820.
- G. H. Fischer. A measurement model for the effect of mass-media. *Acta Psychologica*, 36(3):207–220, 1972. 10.1016/0001-6918(72)90005-4.
- G. H. Fischer. The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37(6):359–374, 1973. ISSN 0001-6918. 10.1016/0001-6918(73)90003-6.
- G. H. Fischer. An irt-based model for dichotomous longitudinal data. *Psychometrika*, 54:pages 599–624, 1989. 10.1007/BF02296399.
- G. H. Fischer and I. W. Molenaar, editors. *Rasch Models Foundations, Recent Developments, and Applications*. Springer-Verlag, New York, 1995.
- C. M. Florkowski. Sensitivity, specificity, receiver-operating characteristic (roc) curves and likelihood ratios: communicating the performance of diagnostic tests. *The Clinical biochemist*, 2008.
- J. Fox. *Bayesian Item Response Modeling: Theory and Applications*. Statistics for Social and Behavioral Sciences. Springer New York, 2010.
- J. Fox. Sociology 761. Lecture Notes. Introduction to Mixed-Effects Models for Hierarchical and Longitudinal Data, 2014. <https://socialsciences.mcmaster.ca/jfox/Courses/soc761/mixed-models.pdf> – last visited 15<sup>th</sup> of July, 2022.
- J. Fox and S. Weisberg. *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, third edition, 2019. URL <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- J. P. Fox and C. A. W. Glas. BAYESIAN ESTIMATION OF A MULTILEVEL IRT MODEL USING GIBBS SAMPLING. *Psychometrika*, 66(2), 2001. 10.1007/BF02294839.

- J. P. Fox and C. A. W. Glas. Multilevel Response Models with Covariates and Multiple Groups. In W. J. van der Linden, editor, *Handbook of Item Response Theory. Volume One Models*, chapter 24, pages 407–419. Taylor & Francis Group, Boca Raton, FL, 2016.
- A. Gelman, J. Hwang, and A. Vehtari. Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24:997–1016, 2014. 10.1007/s11222-013-9416-2.
- A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. Rubin. *Bayesian Data Analysis*. Behavioral Sciences, Mathematics & Statistics. Chapman and Hall/CRC, 3 edition, 2021. 10.1201/b16018.
- J. Geweke. Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica*, 57(6):1317–1339, 1989. 10.2307/1913710.
- C. A. W. Glas and N. D. Verhelst. Extensions of the partial credit model. *Psychometrika*, 54: 635–659, 1989. 10.1007/BF02296401.
- G. Gómez and P. Delicado. *Curso de Inferencia y Decisión*, 2006. Departament d’Estadística i Investigació Operativa, Universitat Politècnica de Catalunya.
- M. A. Gómez Villegas. *Inferencia Estadística*. Díaz de Santos, 2005.
- S. W. Goodwin, M. A. Ferro, and K. N. Speechley. Development and assessment of the Quality of Life in Childhood Epilepsy Questionnaire. *Epilepsia.*, 59(3):668–678, Mar. 2018.
- G. Gottipati, M. O. Karlsson, and E. L. Plan. Modeling a Composite Score in Parkinson’s Disease Using Item Response Theory. *AAPS*, 19(3):837–845, May 2017. 10.1208/s12248-017-0058-8.
- D. C. Haley. Estimation of the dosage mortality relationship when the dose is subject to error. Technical Report 15, Stanford: Applied Mathematics and Statistics Laboratory, Stanford University, 1952.
- R. K. Hambleton, H. Swaminathan, and H. J. Rogers. *Fundamental of Item Response Theory*. SAGE Publications, Inc., 2455 Teller Road, Newbury, California, 1991.
- K. Hayes and J. Haslett. Simplifying general least squares. *The American Statistician*, 53(4): 376–381, 1999. 10.1080/00031305.1999.10474493.
- N. E. Helwig. Applied Statistics. Department of Psychology. School of Statistics. University of Minnesota, 2017. <http://users.stat.umn.edu/~helwig/teaching.html> – last visited 15<sup>th</sup> of March, 2022.
- C. R. Henderson. Estimation of Variance and Covariance Components. *Biometrics*, 9(2):226–252, 1953. URL <http://www.jstor.org/stable/3001853>.
- E. Hobmaier. Influences on the loss of bone density in perimenopausal women - applying linear mixed models. Diplomarbeit, Department of Mathematics, Technische Universität München, München, Germany, 2011.
- P. D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer Text in Statistics. Springer, 1 edition, 2009.
- H. Hoijtink and M. Vollema. Contemporary Extensions of the Rasch Model. *Quality & Quantity*, 37:263–276, 2003. 10.1023/A:1024497124563.
- P. W. Holland and H. Wainer. *Differential Item Functioning*. New York, 1 edition, 1993.
- J. J. Hox. *Multilevel Analysis: techniques and applications*. Routledge Taylor & Francis, 270 Madison Avenue, New York, 2 edition, 2010.
- IESTA. Notas del Curso Introducción a la Inferencia Estadística, 2020. FCEA. UdelaR.
- H. Irtel. An extension of the concept of specific objectivity. *Psychometrika*, 60:115–118, March 1995. 10.1007/BF02294432.
- S. R. Jannat and S. Canavan. Classification of Autism Spectrum Disorder Across Age Using Questionnaire and Demographic Information. In A. Del Bimbo, R. Cucchiara, S. Sclaroff,

- G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, and R. Vezzani, editors, *Pattern Recognition. ICPR International Workshops and Challenges*, pages 52–61, Cham, 2021. Springer International Publishing.
- C. Jayakumaran and J. D. Sweetlin. An Optimized Neural Network with Inertia Weight Variation of PSO for the detection of Autism. In *International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, pages 439–445. IEEE, 2020.
- H. Jeffreys. *Theory of Probability*. Oxford University Press, New York, 1961.
- B. W. Junker. Some statistical models and computational methods that may be useful for cognitively-relevant assessment, 1999. Paper prepared for the Committee on the Foundations of Assessment, National Research Council.
- R. E. Kass. Bayes Factors in Practice. *Journal of the Royal Statistical Society Series D (The Statistician)*, 42(5):551–560, 1993. 10.2307/2348679.
- M. Kuhfeld and S. A. An examination of the precision of the Autism Diagnostic Observation Schedule using item response theory. *Psychol Assess.*, 30(5):656–668, May. 2018. 10.1037/pas0000512.
- T.-C. Kuo. *Bayesian estimation of multi-unidimensional graded response IRT models*. PhD thesis, Department of Counseling, Quantitative Methods, and Special Education. Southern Illinois University Carbondale, Carbondale, Illinois, December 2015.
- N. M. Laird and J. H. Ware. Random-effects Models for Longitudinal Data. *Biometrics*, 38(4): 963–974, 1982.
- L. R. LaMotte. A direct derivation of the reml likelihood function. *Statistical Papers*, 48(2): 321–327, 2007.
- X. Le, I. Lancashire, G. Hirst, and R. Jokel. Longitudinal detection of dementia through lexical and syntactic changes in writing: A case study of three British novelists. *LLC.*, 26:435–461, 10 2011. 10.1093/lc/fqr013.
- Y. Lee and J. A. Nelder. Hierarchical Generalized Linear Models. *Journal of the Royal Statistical Society Series B (Methodological)*, 4:619–678, 1996.
- R. Levy and R. Mislevy. *Bayesian Psychometric Modeling*. Chapman and Hall/CRC, New York, 1 edition, 2016. 10.1201/9781315374604.
- Y. Li and N. Wang. Deviance Information Criterion for Model Selection: Theoretical Justification and Applications, 2022. <http://www.mysmu.edu/faculty/yujun/Research/DIC.Theory27.pdf> – last visited 15<sup>th</sup> of November, 2022.
- Y. LI, J. YU, , and T. ZENG. Deviance information criterion for Bayesian model selection: Justification and variation. *Research Collection School Of Economics*, pages 1–40, 2017. URL [https://ink.library.smu.edu.sg/soe\\_research/1927](https://ink.library.smu.edu.sg/soe_research/1927).
- Q. Liu and D. A. Pierce. A Note on Gauss-Hermite Quadrature. *Biometrika*, 81(3):624–629, 1994. URL <http://www.jstor.org/stable/2337136>.
- E. F. Lock. PUBH 8442. Lecture Notes. Bayes Decision Theory and Data Analysis, 2022.
- F. Lord. *Applications of Item Response Theory To Practical Testing Problems*. Routledge, New York, 1980.
- F. M. Lord. *A theory of test scores*. Number 7 in Psychometric Monograph. Psychometric Corporation, Richmond, VA, 1952.
- F. M. Lord and M. Novick. *Statistical Theories of Mental Test Scores, with contributions by A. Birnbaum*. Behavioral Science: Quantitative Methods. Addison-Wesley, 1968.
- A. Lundin, K. Kosidou, and C. Dalman. Measuring Autism Traits in the Adult General Population with the Brief Autism-Spectrum Quotient, AQ-10: Findings from the Stockholm Public

- Health Cohort. *J Autism Dev Disord.*, 2:773–780, Feb. 2019. 10.1007/s10803-018-3749-9.
- L. Lundqvist and H. Lindner. Is the Autism-Spectrum Quotient a Valid Measure of Traits Associated with the Autism Spectrum? A Rasch Validation in Adults with and Without Autism Spectrum Disorders. *J Autism Dev Disord*, 47:2080–2091, 2017. 10.1007/s10803-017-3128-y.
- M. Luzardo. *Contribuciones a la Teoría de Respuesta al Ítem no Paramétrica*. PhD thesis, Universidad Autónoma de Madrid, Madrid, 2017.
- P. Mair. *Modern Psychometrics With R*. Springer, New York, 2018.
- P. Mair and R. Hatzinger. Extended rasch modeling: The erm package for the application of irt models in r. *Journal of Statistical Software*, 20(9):1–20, 2007. URL <https://www.jstatsoft.org/v20/i09>.
- R. Martin. Lecture Notes on Advanced Statistical Theory. Department of Statistics. North Carolina State University, 2017.
- P. Martinkova and A. Drabinova. ShinyItemAnalysis for teaching psychometrics and to enforce routine analysis of educational tests. *The R Journal*, 10(2):503–515, 2018. 10.32614/RJ-2018-074.
- M. R. Martínez Arias, M. V. Hernández Lloreda, and M. J. Hernández Lloreda. *Psicometría*. Alianza Editorial, Madrid, 2014.
- F. Massa. Librería Raschreg. Modelos de la teoría de respuesta al ítem con efectos de regresión. Seminario SIESTA, Septiembre 2020.
- F. Massa, G. Camaño, and R. Álvarez Vaz. Item response theory modelling assessment of oral health in a uruguayan population study. In *Proceedings of the 33rd International Workshop on Statistical Modelling*, volume 2, pages 100–104, Bristol, UK, 07 2018.
- F. Massa, N. Berberian, and R. Álvarez-Vaz. Modeling in an oral health study through two statistical methods in uruguay. *bioRxiv*, 2019. 10.1101/611921. URL <https://www.biorxiv.org/content/early/2019/04/19/611921>.
- A. E. Maxwell. Maximum likelihood estimates of item parameters using the logistic function. *Psychometrika*, 24:221–227, 1959.
- A. Maydeu-Olivares and R. Montaña. How Should We Assess the Fit of Rasch-Type Models? Approximating the Power of Goodness-of-Fit Statistics in Categorical Data Analysis. *Psychometrika*, 78:116–133, January 2013. 10.1007/s11336-012-9293-1.
- . Mayo Clinic. Trastorno del Espectro Autista, 2021. <https://www.mayoclinic.org/es-es/diseases-conditions/autism-spectrum-disorder/symptoms-causes/syc-20352928#:~:text=El%20trastorno%20del%20espectro%20autista,interacci%C3%B3n%20social%20y%20la%20comunicaci%C3%B3n.> – last visited 15<sup>th</sup> of March, 2023.
- C. A. Mazefsky, L. Yu, S. W. White, M. Siegel, and P. A. Pilkonis. The emotion dysregulation inventory: Psychometric properties and item response theory calibration in an autism spectrum disorder sample. *Autism Res.*, 11(6):928–941, Jun. 2018. 10.1002/aur.1947.
- P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC. Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300. Boca Raton, FL, 2 edition, 1989.
- C. E. McCulloch. Generalized Linear Mixed Models. volume 7 of *Probability and Statistics*. Institute of Mathematical Statistics Beachwood, Ohio, NSF-CBMS Regional Conference, 2003.
- C. E. McCulloch and S. R. Searle. *Generalized, Linear, and Mixed Models*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 605 Third Avenue, New York, USA, 2001.
- R. P. McDonald. *Test Theory: A Unified Treatment*. Taylor & Francis Group, 2011.

- C. Mecklin. STA 565/STA 665 Notes. Murray State University, 2022. [http://campus.murraystate.edu/academic/faculty/cmecklin/STA565/\\_book/](http://campus.murraystate.edu/academic/faculty/cmecklin/STA565/_book/) – last visited 15<sup>th</sup> of July, 2022.
- J. L. Melià. Métodos de Estimación en Teoría de la Respuesta al Ítem, 2003. Universidad de València. [https://www.uv.es/~meliajl/Research/EstimationWeb/Fundamentos\\_archivos/frame.htm](https://www.uv.es/~meliajl/Research/EstimationWeb/Fundamentos_archivos/frame.htm) – last visited 15<sup>th</sup> of March, 2022.
- J. Meneses, M. Barrios, A. Bonillo, A. Cosculluela, L. M. Lozano, J. Turbany, and S. Valero. *Psicometría*. Editorial UOC, 1 edition, 2013.
- J. C. Morales Correa and J. C. Salazar Uribe. *Introducción a los modelos mixtos*. Universidad Nacional de Colombia, Medellín, Colombia, 2 edition, 2016.
- J. Muñiz. *Teoría Clásica de los Test*. Ediciones Pirámide, Madrid, 2000.
- J. Muñiz. Las Teorías de los Tests: Teoría Clásica y Teoría de Respuesta a los Ítems. *Papeles del Psicólogo*, 31(1):57–66, 2010.
- J. Muñiz. *Introducción a la Psicometría. Teoría clásica y TRI*. Ediciones Pirámide, Madrid, 2018.
- N. Mukhopadhyay. *Probability and Statistical Inference*, volume 162 of *STATISTICS: Textbooks and Monographs*. Marcel Dekker, Inc., 2000.
- A. L. Murray, T. Booth, K. McKenzie, and R. Kuenssberg. What range of trait levels can the Autism-Spectrum Quotient (AQ) measure reliably? An item response theory analysis. *Psychol Assess.*, 28(6):673–83, Aug. 2016. 10.1037/pas0000215.
- V. P. Nia. Gauss-hermite quadrature: Numerical or statistical method? *Proc. Iranian Stat. Conf*, pages 209–215, 2006. URL <https://datawisdom.ca/paper/2006-GHQ.pdf>.
- J. C. Nunnally and I. H. Bernstein. *Psychometric Theory*. McGraw-Hill series in psychology. McGraw-Hill Inc., 3 edition, 1994.
- D. J. Olive. *Statistical Theory and Inference*. Springer, 2014.
- U. Olsson. *Generalized Linear Models An Applied Approach*. Studentlitteratur, Studentlitteratur, Lund, 2002.
- . Organización Mundial para la Salud. Autismo, 2023. <https://www.who.int/es/news-room/fact-sheets/detail/autism-spectrum-disorders#:~:text=A1%20igual%20que%20e1%20resto,de%20enfermedades%20agudas%20y%20cr%C3%B3nicas> – last visited 15<sup>th</sup> of March, 2023.
- D. Panchenko. Lecture Notes on Statistics For Applications. Massachusetts Institute of Technology: MIT OpenCourseWare. License: Creative Commons BY-NC-SA. <https://ocw.mit.edu/courses/18-443-statistics-for-applications-fall-2003/> – last visited 15<sup>th</sup> of July, 2022, 2003.
- H. D. Patterson and R. Thompson. Recovery of Inter-Block Information when Block Sizes are Unequal. *Biometrika*, 58(3):545–554, 1971.
- G. Perera. *Probabilidad y Estadística*. Fin de Siglo, Convención 1537, Montevideo, Uruguay, 2011.
- J. Pfanzagl. *On Item Parameter Estimation in Certain Latent Trait Models*, pages 249–264. Springer-Verlag New York, Inc., 1994.
- J. C. Pinheiro. *Topics in Mixed Effects Models*. PhD thesis, University of Wisconsin, Madison, 1994.
- J. C. Pinheiro and D. M. Bates. *Mixed-effects models in S and S-PLUS*. Statistics and Computing. Springer-Verlag New York, Inc., New York, 2000.
- M. B. Pintarelli. *Matemática 3*, 2010. Universidad Nacional de la Plata.
- L. R. Price. *Psychometric Methods: Theory into Practice*. Methodology in the Social Sciences. The Guilford Press, 1 edition, 2016.

- . Psychometric Society. What is psychometrics? [https://www.psychometricsociety.org/what-  
-psychometrics](https://www.psychometricsociety.org/what-psychometrics) – last visited 15<sup>th</sup> of Feb, 2022.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL <https://www.R-project.org/>.
- S. Raj and S. Masood. Analysis and detection of autism spectrum disorder using machine learning techniques. *Procedia Computer Science*, 167:994–1004, 2020. 10.1016/j.procs.2020.03.399. International Conference on Computational Intelligence and Data Science.
- C. R. Rao and S. Sinharay, editors. *Handbook of Statistics 26: Psychometrics*. North Holland, Elsevier edition, 2007.
- G. Rasch. *Probabilistic Models for Some Intelligence and Attainment Tests*. Studies in mathematical psychology. Danmarks Paedagogiske Institut, 1960.
- A. C. Rencher and G. B. Schaalje. *LINEAR MODELS IN STATISTICS*. John Wiley & Sons, Inc, Hoboken, New Jersey, 2 edition, 2008.
- J. A. Rice. *Mathematical Statistics and Data Analysis*. Duxbury Advanced Series. Thomson Brooks/Cole, 3 edition, 2007.
- F. Rijmen, F. Tuerlinckx, P. De Boeck, and P. Kuppens. A Nonlinear Mixed Model Framework for Item Response Theory. *Psychological Methods*, 8(2):185–205, 2003.
- D. Rizopoulos. Interpretation of fixed effects from-mixed effect logistic regression, 2018a. <https://stats.stackexchange.com/questions/365907/interpretation-of-fixed-effects-from-mixed-effect-logistic-regression> – last visited 15<sup>th</sup> of Feb., 2023.
- D. Rizopoulos. Marginal effects from glmer, 2018b. <https://stackoverflow.com/questions/52005435/marginal-effects-from-glmer> – last visited 15<sup>th</sup> of Feb., 2023.
- C. P. Robert and G. Casella. *Introducing Monte Carlo Methods with R*. Use R! Springer New York, NY, 1 edition, 2010. 10.1007/978-1-4419-1576-4,.
- O. R. Rodríguez Jiménez, R. F. Rosero Burbano, M. L. Botia Sanabria, and L. H. Duarte Mateus. Producción de Conocimiento en Psicometría en Instituciones de Educación Superior de Bogotá y Chía. *Revista Colombiana de Psicología*, 20(1):9–25, Enero-Junio 2011.
- A. Saihi and H. Alshraideh. Development of an autism screening classification model for toddlers. In *Advances in Machine Learning*. Academy and Industry Research Collaboration Center (AIRCC), sep 2021. 10.5121/csit.2021.111508. URL <https://arxiv.org/abs/2110.01410>.
- S. Sandin, P. Lichtenstein, R. Kuja-Halkola, H. Larsson, C. M. Hultman, and A. Reichenberg. The familial risk of autism. *Jama*, 311(17):1770–1777, 2014.
- C. Santisteban Requena. *Principios Psicometría*. Editorial Síntesis, Madrid, España, 2009.
- C. Santisteban Requena and J. M. Alvarado. *Modelos Psicométricos*. UNED, Madrid, España, 2001.
- S. Schilling and R. D. Bock. High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, 70:533–555, 2005. 10.1007/s11336-003-1141-x.
- G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, March 1978.
- S. R. Searle, C. E. McCulloch, and C. Casella. *Variance Components*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics Section. John Wiley & Sons, Inc., Hoboken, New Jersey, USA, 2006.
- H. Seung, J. Ji, S. J. Kim, I. Sung, Y. A. Youn, G. Hong, L. H., Y. H. Lee, L. H., and H. K. Youm. Examination of the Korean Modified Checklist of Autism in Toddlers: Item Response Theory. *J Autism Dev Disord.*, 45(9):2744–57, Sep. 2015. 10.1007/s10803-015-2439-0.
- Y. Sheng and T. C. Headrick. A Gibbs Sampler for the Multidimensional Item Response Model.

- Applied Mathematics*, pages 1–13, Jan 2012.
- A. Singh, Z. Farooqui, B. Sattler, U. Usua, and M. Helde. Using Machine Learning Optimization to Predict Autism in Toddlers. In *Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management*. Singapore, 2021.
- S. Sinharay and H. S. Stern. On the Sensitivity of Bayes Factors to the Prior Distributions. *The American Statistician*, 56(3):196–201, 2012. 10.1198/000313002137.
- T. A. B. Snijders and R. J. Bosker. *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*. Sage, London, 2 edition, 2011.
- D. Spiegelhalter, N. Best, B. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Methodological)*, 64(4):583–639, 2002. 10.1111/1467-9868.00353.
- D. Spiegelhalter, N. Best, B. Carlin, and A. van der Linde. The Deviance Information Criterion: 12 Years On. *Journal of the Royal Statistical Society. Series B (Methodological)*, 76(3):485–493, 2014.
- H. Stern. , 2019. Lectures Notes. Statistics 225: Bayesian Statistical Analysis. <https://www.ics.uci.edu/~sternh/courses/225/> – last visited 15<sup>th</sup> of March, 2023.
- M. Stone. An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike’s Criterion. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):44–47, 1997.
- X. Su. *Bayesian estimation of multilevel item response model with missing data*. PhD thesis, The University of Texas School of Public Health, Houston, Texas, August 2012.
- R. Sundberg. *Statistical Modelling by Exponential Families*. IMS Textbooks. Cambridge University Press, Cambridge CB2 8BS, United Kingdom, 2019.
- F. Thabtah. Autistic spectrum disorder screening data for adult, 2017a. Department of Digital Technology. Manukau Institute of Technology.
- F. Thabtah, 2017b. [www.asdtests.com](http://www.asdtests.com) – last visited 15<sup>th</sup> of March, 2023.
- F. Thabtah. Autism spectrum disorder screening: machine learning adaptation and dsm-5 fulfillment. In *Proceedings of the 1st International Conference on Medical and health Informatics 2017*, pages 1–6, 2017c.
- F. Thabtah. Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. *Informatics for Health and Social Care*, 44(3):278–297, 2019a.
- F. Thabtah. An accessible and efficient autism screening method for behavioural data and predictive analyses. *Health Informatics Journal*, 25(4):1739–1755, 2019b. 10.1177/1460458218796636.
- F. Thabtah. Autism screening data for toddlers, 2020. <https://www.kaggle.com/datasets/fabdelja/autism-screening-for-toddlers> – last visited 15<sup>th</sup> of Feb., 2023.
- F. Thabtah, F. Kamalov, and K. Rajab. A new computational intelligence approach to detect autistic features for autism screening. *International journal of medical informatics*, 117: 112–124, 2018.
- F. Thabtah, N. Abdelhamid, and D. Peebles. A machine learning autism classification based on logistic regression analysis. *Health information science and systems*, 7:1–11, 2019.
- F. F. Thabtah. Autistic spectrum disorder screening data for adolescent, 2017d. Department of Digital Technology. Manukau Institute of Technology.
- F. F. Thabtah. Autistic spectrum disorder screening data for children, 2017e. Department of Digital Technology. Manukau Institute of Technology.
- L. L. Thurstone. A Method of Scaling Psychological and Educational Tests. *Journal of Educational Psychology*, 16:433– 451, 1925.

- L. L. Thurstone. Attitudes Can Be Measured. *American Journal of Sociology*, 33:529–554, 1928.
- S. Tokdar and R. Kass. Importance sampling: a review. *Wiley Interdisciplinary Reviews. Computational Statistics*, 2009. 10.1002/wics.56.
- L. R. Tucker. Maximum validity of a test with equivalent items. *Psychometrika*, 11:1–13, 1946. 10.1007/BF02288894.
- Y. Umeda-Kameyama, T. Mori, K. Wada-Isoe, T. Kikuchi, S. Kojima, T. Kagimura, A. Ueki, T. Watabe, C. Kudoh, M. Akishita, and Y. Nakamura. Development of a novel convenient alzheimer’s disease assessment scale, using item response theory. *Geriatrics & Gerontology International*, 19, 11 2018. 10.1111/ggi.13552.
- W. van der Linden, editor. *Handbook of Item Response Theory*, volume 1. Chapman and Hall/CRC, New York, 2016.
- W. van der Linden and R. K. Hambleton. *Handbook of Modern Item Response Theory*, volume 1. Springer, New York, 1997.
- A. Vehtari, A. Gelman, and J. Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27:1413–1432, 2017. 10.1007/s11222-016-9696-4.
- A. Vehtari, J. Gabry, M. Magnusson, Y. Yao, P.-C. Bürkner, T. Paananen, and A. Gelman. loo: Efficient leave-one-out cross-validation and waic for bayesian models, 2023. URL <https://mc-stan.org/loo/>. R package version 2.6.0.
- N. Verma and M. K. Markey. Item response analysis of Alzheimer’s disease assessment scale. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2476–2479, 2014. 10.1109/EMBC.2014.6944124.
- L. D. V. Vleck. Charles Roy Henderson 1911-1989 a Biographical Memoir, 1998. National Academies Press Washington D.C., 1998.
- M. von Davier and C. H. Rost, J. Carstensen. *Multivariate and Mixture Distribution Rasch Models. Extensions and Applications*. Statistics for Social and Behavioral Sciences. Springer New York, NY, 2007. 10.1007/978-0-387-49839-3.
- L. Wasserman. *All of Statistics. A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer, 2004.
- S. Watanabe. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11:3571–3594, 2010.
- R. Weiss. An Approach to Bayesian Sensitivity Analysis. *Journal of the Royal Statistical Society Series B (Methodological)*, 58(4):739–750, 1996. 10.1111/j.2517-6161.1996.tb02112.x.
- B. T. West, K. B. Welch, and A. T. G. ecki. *LINEAR MIXED MODELS. A Practical Guide Using Statistical Software*. Chapman & Hall/CRC. Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300. Boca Raton, FL, 2007.
- P. O. White. A note on keats’ generalization of the rasch model. *Psychometrika*. *Psychometrika*, 41:405–407, Spetember 1976. 10.1007/BF02293564.
- B. Wright and G. Masters. *Rating scale analysis*. MESA Press, Chicago, IL, 1982.
- L. Wu. *Mixed Effects Models for Complex Data*, 2009. University of British Columbia.
- M. Wu, R. L. Davis, B. W. Domingue, C. Piech, and N. Goodman. Variational Item Response Theory:Fast, Accurate, and Expressive. In A. N. Rafferty, J. Whitehill, V. Cavalli-Sforza, and C. Romero, editors, *Proceedings of The 13th International Conference on Educational Data Mining*, pages 257–268, 2020.
- M. Yela. Introducción a la teoría de los tests. *Psicothema*, 8(1):249–263, 1996.

- A. Zwinderman. A generalized rasch model for manifest predictors. *Psychometrika*, 56:589–600, 1991. 10.1007/BF02294492.
- A. Zwinderman. Response Models with Manifest Predictors. In W. van der Linden and R. K. Hambleton, editors, *Handbook of Modern Item Response Theory*, volume 1, chapter 14, pages 245–256. Springer, New York, 1997.