

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

MEDICIÓN DEL NIVEL DE INNOVACIÓN EN LAS ORGANIZACIONES USANDO MODELOS DE RESPUESTA AL ÍTEM (IRT) EN EL CONTEXTO DE LA REALIDAD ECUATORIANA

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
MATEMÁTICO

PROYECTO DE INVESTIGACIÓN

ANA GABRIELA TAPE RÍOS

mat.ana.taipe@hotmail.com

Director: DR. CARLOS ALBERTO ALMEIDA RODRÍGUEZ

carlos.almeidar@epn.edu.ec

QUITO, OCTUBRE 2022

DECLARACIÓN

Yo ANA GABRIELA TAIPE RÍOS, declaro bajo juramento que el trabajo aquí escrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual, correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normatividad institucional vigente.



Ana Gabriela Taipe Ríos

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por ANA GABRIELA TAIPE RÍOS, bajo mi supervisión.

Dr. Carlos Alberto Almeida Rodriguez
Director del Proyecto

AGRADECIMIENTOS

A Dios por ser mi respaldo y confidente.

A mi familia por su guía y apoyo. Cada palabra y gesto fueron la energía necesaria para poder culminar esta etapa de mi vida.

A mi padre Marcelo por cultivar en mí la paciencia. A mi madre Dora por enseñarme a luchar y buscar los medios para conseguir lo que se desea. A mi hermana Paulina (sapiñ) por brindarme su cuidado y apoyo en todo sentido, por ser como una segunda madre para mí, y por último y no menos importante, a mi hermano Fernando por sus consejos, por sus charlas, por su apoyo y su compañía, gracias sope (saludo I.T.).

Familia, gracias por todo. Espero recompensar el apoyo y tiempo que me brindaron en los momentos que más necesité. Sin ustedes esta meta no hubiese sido posible.

A mi director Carlos Almeida, por el tiempo, por su dirección, por su enseñanza y conocimiento que permitieron el desarrollo de este trabajo de investigación.

A mis compañeros y amigos de facultad B.B., J.B., S.H., Juandy, Peluche y Leo por su amistad. Sebastián gracias por tus ánimos para continuar en este último peldaño “la tesis”, gracias en verdad por estar ahí.

Finalmente, gracias EPN, gracias carrera de Matemática, por formarme como profesional y persona. Por enseñarme la disciplina, paciencia, aprendizaje y valentía.

DEDICATORIA

A mi familia y a mis mascotas: Ángel y Conis por su compañía en esos días largos de estudio.

Índice general

Índice general

Índice de tablas

Índice de figuras I

1. Preliminares	1
1.1. Introducción	1
1.2. Teoría de Respuesta al Ítem	3
1.2.1. Modelo de ojiva normal de 2 parámetros	5
1.2.2. Modelo logístico de 1 parámetro	6
1.2.3. Modelo logístico de 2 parámetros	8
1.2.4. Modelo de Respuesta Graduada	9
1.3. Criterio de Información del Ítem	11
2. Estimación de parámetros	16
2.1. Estimación Bayesiana	16
2.1.1. Monte Carlo con cadenas de Markov	18
2.1.2. Modelización GRM con el método MCMC	24
2.2. Monte Carlo Hamiltoniano	26
2.3. Diagnóstico de Convergencia MCMC	31
2.3.1. Gráficos de Traza	31
2.3.2. Gráficos de autocorrelación	32
2.3.3. Gelman Rubin	33

3. Innovación	35
3.1. Índice Mundial de Innovación	35
3.2. Innovación Organizacional	38
3.3. Instrumento de Recolección de Datos	38
3.3.1. Cuestionario CRI2	40
3.3.2. Validación	44
3.4. Muestra	44
4. Simulación y Resultados	46
4.1. Hipótesis	46
4.2. MCMC en el caso de estudio	47
4.3. Simulación	47
4.4. Resultados	51
4.5. Clasificación de las Organizaciones	59
5. Conclusión y Recomendaciones	61
5.1. Conclusión	61
5.2. Recomendaciones	62
Anexos	64
A. Simulación del Modelo	65
A.1. Simulación	65
A.2. Estimación de parámetros	65
A.3. Gráficas de convergencia (traza y ACF)	67
A.4. Curvas Características de Categoría	85
A.5. Criterio de Información del ítem	95
A.6. Información del test	97
Bibliografía	I

Índice de tablas

3.1. Economías innovadoras por región	37
3.2. Actividades de Innovación	39
3.3. Niveles de abstracción del modelo CRI	41
3.4. Descripción del Modelo CRI	43
4.1. Frecuencia de respuesta para el subrasgo <i>Resultados</i>	53
4.2. Estimación del parámetro de discriminación del subrasgo <i>Resultados</i>	53
4.3. Estimación del parámetro de dificultad del subrasgo <i>Resultados</i>	53
4.4. Estimación del parámetro de habilidad del subrasgo <i>Resultados</i>	54
4.5. Estadístico PSRF para el subrasgo <i>Resultados</i>	59
4.6. Estimación promedio del potencial de innovación y las categorías CRI.	60
A.1. Estimación de parámetros MCMC	66

Índice de figuras

1.1. ICCs para dos ítems del test basados en el modelo 1PL	7
1.2. ICCs para dos ítems de test basados en el modelo 2PL	8
1.3. ICCs para tres ítems, en los distintos modelos 1PL, 2PL y 3PL	9
1.4. ICCs para el modelo GRM con una baja, media y alta discriminación α . . .	12
1.5. ICCs para el modelo GRM con baja, media y alta dificultad β	13
2.1. Problemas comunes en los gráficos de Trazas	31
2.2. Problemas comunes en los gráficos de Autocorrelación.	33
4.1. Estimación del parámetro de discriminación usando Stan	50
4.2. Criterio de Información del Ítem del subrasgo <i>Resultados</i>	51
4.3. CCCs del subrasgo <i>Resultados</i>	52
4.4. Gráficas de traza del parámetro de discriminación del subrasgo <i>Resultados</i> .	55
4.5. Gráficas de traza del parámetro de dificultad del subrasgo <i>Resultados</i>	55
4.6. Gráficos de traza de la habilidad del subrasgo <i>Resultados</i>	56
4.7. ACF del parámetro de discriminación del subrasgo <i>Resultados</i>	57
4.8. ACF del parámetro de dificultad del subrasgo <i>Resultados</i>	57
4.9. ACF de la habilidad del subrasgo <i>Resultados</i>	58
A.1. Gráficos de traza del parámetro de discriminación del subrasgo <i>Fuentes de Información</i>	67
A.2. Gráficos de traza del parámetro de dificultad del subrasgo <i>Fuentes de Información</i>	67
A.3. ACF del parámetro de discriminación del subrasgo <i>Fuentes de Información</i>	68
A.4. ACF del parámetro de dificultad del subrasgo <i>Fuentes de Información</i> . . .	68

A.5. Gráfico de traza y gráficos ACF de tres organizaciones de la muestra con el subrasgo <i>Fuentes de Información</i>	69
A.6. Gráficos de traza del parámetro de discriminación del subrasgo <i>Fuentes de Financiamiento</i>	70
A.7. Gráficos de traza del parámetro de dificultad del subrasgo <i>Fuentes de Financiamiento</i>	70
A.8. ACF del parámetro de discriminación del subrasgo <i>Fuentes de Financiamiento</i>	71
A.9. ACF del parámetro de dificultad del subrasgo <i>Fuentes de Financiamiento</i> .	71
A.10. Gráficos de traza y gráficos ACF de tres organizaciones de la muestra con el subrasgo <i>Fuentes de Financiamiento</i>	72
A.11. Gráficos de traza del parámetro de discriminación del subrasgo <i>Actividades</i>	73
A.12. Gráficos de traza del parámetro de dificultad del subrasgo <i>Actividades</i> . . .	73
A.13. ACF del parámetro de discriminación del subrasgo <i>Actividades</i>	74
A.14. ACF del parámetro de dificultad del subrasgo <i>Actividades</i>	74
A.15. Gráficos de traza y gráficos ACF de tres organizaciones de la muestra con el subrasgo <i>Actividades</i>	75
A.16. Gráficos de traza del parámetro de discriminación del subrasgo <i>Objetivos</i> .	76
A.17. Gráficos de traza del parámetro de dificultad del subrasgo <i>Objetivos</i>	76
A.18. ACF del parámetro de discriminación del subrasgo <i>Objetivos</i>	77
A.19. ACF del parámetro de dificultad del subrasgo <i>Objetivos</i>	77
A.20. Gráficos de traza y gráficos ACF de tres organizaciones de la muestra con el subrasgo <i>Objetivos</i>	78
A.21. Gráficos de traza del parámetro de discriminación del subrasgo <i>Resultados</i> .	79
A.22. Gráficos de traza del parámetro de dificultad del subrasgo <i>Resultados</i>	79
A.23. ACF del parámetro de discriminación del subrasgo <i>Resultados</i>	80
A.24. ACF del parámetro de dificultad del subrasgo <i>Resultados</i>	80
A.25. Gráficos de traza y gráficos ACF de tres organizaciones de la muestra con el subrasgo <i>Resultados</i>	81
A.26. Gráficos de traza del parámetro de discriminación del subrasgo <i>Impactos</i> . .	82
A.27. Gráficos de traza del parámetro de dificultad del subrasgo <i>Impactos</i>	82
A.28. ACF del parámetro de discriminación del subrasgo <i>Impactos</i>	83

A.29.ACF del parámetro de dificultad del subrasgo <i>Impactos</i>	83
A.30.Gráficos de traza y gráficos ACF de tres organizaciones de la muestra con el subrasgo <i>Impactos</i>	84
A.31.CCCs del subrasgo <i>Fuentes de Información</i> (parte 1)	85
A.32.CCCs del subrasgo <i>Fuentes de Información</i> (parte 2)	86
A.33.CCCs del subrasgo <i>Fuentes de Financiamiento</i>	87
A.34.CCCs del subrasgo <i>Actividades</i> (parte 1)	88
A.35.CCCs del subrasgo <i>Actividades</i> (parte 2)	89
A.36.CCCs del subrasgo <i>Objetivos</i> (parte 1)	90
A.37.CCCs del subrasgo <i>Objetivos</i> (parte 2)	91
A.38.CCCs del dominio <i>Resultados</i>	92
A.39.CCCs del subrasgo <i>Impactos</i> (parte 1)	93
A.40.CCCs del subrasgo <i>Impactos</i> (parte 2)	94
A.41.Criterio de Información del Ítem (parte 1)	95
A.42.Criterio de Información del Ítem (parte 2)	96
A.43.Información del test (parte 1)	97
A.44.Información del test (parte 2)	98

RESUMEN

El objetivo de este trabajo de investigación es construir una medición válida del potencial de innovación organizacional en relación con las respuestas de sujetos a un test realizado basándose en un modelo de medición CRI (Capacidades, Resultados e Impactos) de innovación. La solución propuesta se fundamenta en un enfoque de la Teoría de Respuesta al Ítem que utiliza un Modelo de Respuesta Gradual cuyos parámetros se estiman por inferencia bayesiana a través del método Monte Carlo con cadenas de Markov denominado Monte Carlo Hamiltoniano. Esto proporciona a las organizaciones del conjunto de datos una estimación de su potencial de innovación y de su competencia en las categorías CRI. La investigación que se presenta es muy flexible, ya que el procedimiento de análisis sigue siendo válido si se agregan más datos o se modifican los tests.

Palabras clave: Teoría de Respuesta al Ítem (IRT), Modelo de Respuesta Graduada (GRM), Monte Carlo Hamiltoniano (HMC), Innovación organizacional, Capacidades de innovación, Resultados de innovación, Impactos de innovación.

ABSTRACT

The objective of this research work is to construct a valid measurement of organizational innovation potential in relation to the responses of subjects to a test conducted based on a CRI (Capabilities, Results and Impacts) measurement model of innovation. The proposed solution is based on an Item Response Theory approach that uses a Gradual Response Model whose parameters are estimated by Bayesian inference through the Monte Carlo method with Markov chains called Monte Carlo Hamiltonian. This provides the organizations in the dataset with an estimate of their innovation potential and their competence in the CRI categories. The research presented is very flexible, since the analysis procedure remains valid if more data is added or the tests are modified.

Keywords: Item Response Theory (IRT), Graduated Response Model (GRM), Hamiltonian Monte Carlo (HMC), Organizational Innovation, Innovation Capabilities, Innovation Outcomes, Innovation Impacts.

Abreviaciones

IRT	Teoría de Respuesta al Ítem.
IRF	Función de Respuesta al Ítem.
ICC	Curva Característica del Ítem.
CRF	Función de Respuesta de Categoría o Función de Probabilidad de Categoría.
CCC	Curva Característica de Categoría o Curva de Probabilidad de Categoría.
IIC	Criterio de Información del Ítem.
IIF	Función de Información del Ítem
GRM	Modelo de Respuesta Gradual.
EM	Maximización de Expectativas.
MCMC	Monte Carlo con cadenas de Markov.
M-H	Metrópolis Hasting.
HMC	Monte Carlo Hamiltoniano.
ACF	Función de Autocorrelación.
PSRF	Factor de Reducción de Escala Potencial
CRI	Capacidades, Resultados e Impactos.
CVR	Razón de Validez de Contenido.

Capítulo 1

Preliminares

1.1. Introducción

El desarrollo de un país puede sustentarse en varios factores, uno de los más importantes es la estructura productiva que se basa en la ventaja competitiva creada a través del comportamiento innovador.

En Ecuador, en 2022, el nivel de innovación es muy bajo. Esto se debe a que el gobierno no invierte en investigación y desarrollo, y no hay incentivos para que las empresas innoven. A pesar de los esfuerzos de algunas de ellas para mejorar la innovación, el país sigue estancado en un modelo económico basado en la explotación de recursos naturales. Esto lleva a una situación en la que la mayoría de las empresas ecuatorianas son poco competitivas a nivel internacional, lo que a su vez limita el desarrollo económico del país.

Aunque el gobierno ecuatoriano haya tomado medidas para mejorar la innovación organizacional, creando en el 2010 la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (Senescyt)¹ y aprobando la Ley Orgánica de Emprendimiento e Innovación, en el 2020, ley que establece un marco legal para fomentar el emprendimiento, la innovación y el desarrollo tecnológico, y que promueve la cultura emprendedora e implementa nuevos modelos de negocio y financiamiento para fortalecer el ecosistema emprendedor («Ley Orgánica de Emprendimiento e Innovación», 2020). Muchas empresas ecuatorianas siguen siendo poco innovadoras, es más, el país ocupa el puesto 86 de 189 países en el Índice de Desarrollo Humano (IDH²; 2019) y el puesto 90 de 133 países en el Índice de Competitividad Global

¹Entidad que tiene como objetivo desarrollar, gestionar e implementar políticas públicas de investigación científica, innovación y transferencia de tecnología, articulando a nivel nacional e internacional los sistemas de educación superior, los sistemas nacionales de ciencia, tecnología, innovación y conocimientos ancestrales y los actores del sector productivo para promover la economía del conocimiento, tecnológica desarrollo e innovación.

²Es un indicador del desarrollo humano elaborado por las Naciones Unidas. A diferencia de otros indica-

(ICG³; 2019).

Para mejorar la situación, se necesita un cambio radical en la forma en que se piensa acerca de la innovación. En lugar de verla como un proceso técnico, se requiere entenderla como un proceso social y económico. Se precisa fomentar el espíritu emprendedor, apoyar a las pequeñas y medianas empresas, mejorar la educación y la formación profesional. Solo con un cambio radical de esta índole se podrá esperar que Ecuador comience a desarrollarse y convertirse en un país más innovador. Habrá innovaciones fundamentales y otras consecuencias menos obvias, pero siempre que den como resultado un producto, proceso o un servicio nuevo y mejorado, habrá innovación.

Por tanto, entender cómo medir el estado de los procesos de innovación en las organizaciones ecuatorianas tiene un propósito, por un lado, enriquecer el conocimiento competitivo de las empresas privadas al medir los procesos y así desarrollar una política empresarial que apoye la innovación, y, por otro lado, influir en el desarrollo económico y social de la población ecuatoriana al estructurar políticas públicas que puedan surgir de la información obtenida.

Este estudio tiene como propósito analizar la innovación en el entorno empresarial ecuatoriano, teniendo en cuenta los aspectos contextuales integrados por expertos académicos e investigadores con experiencia en el proceso de innovación, como Morales y col., 2019, quienes adaptaron un test (cuestionario CRI2) valorando ítems por su pertinencia y comprensión al modelo propuesto de Capacidad, Resultados e Impacto de innovación (CRI) que mide la innovación organizacional.

El modelo CRI permite comparar y describir los procesos de innovación dentro de una organización. Se basa en la construcción de un conjunto de capacidades que produzcan resultados, los mismos que reflejarán impactos palpables para la organización.

Para determinar el potencial de innovación en una organización se utiliza las observaciones resultantes del cuestionario CRI2 a través de un modelo de Teoría de Respuesta al Ítem denominado Modelo de Respuesta Graduada (GRM); cuyos parámetros son estimados mediante métodos estadísticos bayesianos como el algoritmo Monte Carlo con cadenas de Markov (MCMC) muestreado gracias al software de código abierto *Stan* que emplea el algoritmo Monte Carlo Hamiltoniano (HMC), el cual usa el principio de la dinámica Hamiltoniana para generar muestras de la distribución posterior. El algoritmo HMC requiere que se especifique una función de paso que determina cómo avanzar en el espacio de parámetros.

dores que miden el desarrollo de un país, el IDH analiza la salud, la educación e ingresos (Expansión, 2019).

³Es el estudio comparativo entre países más importante acerca de los factores que inciden en la productividad y el crecimiento económico y que ayudan a analizar las diferencias en el ingreso per cápita de los mismos (FUNDESA, 2022).

En cada paso del algoritmo, se genera un punto de la distribución posterior usando el hamiltoniano del modelo y luego se acepta o rechaza empleando un criterio de aceptación. Si se acepta el punto, se agrega a la muestra y si se rechaza, se descarta.

En resumen, este trabajo de investigación se estructura en cinco partes: La primera parte contiene la base teórica de la Teoría de Respuesta al Ítem y del Modelo de Respuesta Graduada. La segunda parte abarca el método de estimación MCMC para los parámetros del Modelo de Respuesta Graduada. La tercera parte incluye la interpretación de innovación a la cual haremos referencia en nuestro estudio. La cuarta parte comprende la aplicación del modelo a las 343 organizaciones ecuatorianas que participaron en la muestra y la obtención de resultados. Finalmente, en la quinta parte se encuentran las conclusiones y recomendaciones del trabajo.

1.2. Teoría de Respuesta al Ítem

Binet y Simon, 1904, publicaron *Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux*; donde se muestra el lanzamiento del primer test psicológico estandarizado. Esto formó parte de los primeros fundamentos para la Teoría de Respuesta al Ítem (IRT, con sus siglas en inglés).

Todo surge cuando la ciudad de París pide a Binet desarrollar un test que permitiera a las escuelas diferenciar entre los alumnos con retraso mental, con el fin de asignarlos a una educación especial, y los que simplemente estaban desmotivados.

Binet se dio cuenta de que tenía que medir una variable bastante compleja que podía suponerse estaba ahí, pero a la que no se tenía acceso directo. Un concepto que ahora se denomina *variable latente*.

La solución de Binet fue innovadora en varios aspectos. En primer lugar, diseñó una serie de tareas que indicaban las principales funciones mentales, como la memoria, el razonamiento, el juicio y la abstracción, que se creía que estaban presentes en la inteligencia. En segundo lugar, utilizó dichas tareas en un test estandarizado; todo en él, los materiales, la administración y las reglas de puntuación, estaban cuidadosamente protocolizados. Y, en tercer lugar, escaló los ítems de sus tests, pero se dio cuenta de que no existía una escala natural para medir la inteligencia. Su solución fue usar la edad cronológica de los alumnos para determinar los valores de la escala de sus ítems. Estos valores se utilizaron entonces para estimar la edad mental en la que cada alumno actuaba.

La idea de escalamiento de Binet asignó tanto a sus ítems como a los estudiantes en una sola escala, para ello, estimó los valores de la escala de sus ítems utilizando datos empíricos

recogidos en un pre-test, y luego calificó a los alumnos utilizando los valores de la escala de los ítems.

En 1925, Louis Thurstone introdujo un método de escalamiento para eliminar la necesidad de usar la edad como sustituto manifiesto de la variable de inteligencia latente. Lo hizo asumiendo una escala latente y desconocida para los ítems, y para las curvas empíricas de la proporción de respuestas correctas impuso la forma de la función de distribución normal acumulada utilizando sus parámetros de localización estimados como valores de escala para los ítems. Como resultado de eso, desvinculó la inteligencia de la edad (Thurstone, 1925).

Luego, en 1926, Thurstone amplió el dominio de la medición mostrando cómo podía medirse algo hasta entonces impreciso como una actitud. Para conseguirlo, necesitaba un nuevo método de escalado y un conjunto de respuestas *de acuerdo* y *desacuerdo* a una colección de afirmaciones que evaluaban el objeto de la actitud.

Todo parecía ir bien en los avances de Thurstone hasta que surgió una confusión entre el uso de la función ojiva normal como función de respuesta al ítem y como función de distribución para la puntuación observada. Thurstone concilió esta hipótesis debido a la psicofísica⁴, en la cual se establecía el empleo de la función ojiva normal como función de respuesta, solo que él acomodó esta hipótesis para una variable latente en lugar de manifiesta. En cualquier caso, se reconoce que el impacto del escalamiento psicofísico⁵ en la historia temprana de la teoría de los tests es inconfundible.

En definitiva, la IRT a través de modelos estadísticos, relaciona la variable latente o el puntaje de habilidad del sujeto que realiza el test con la probabilidad de responder a un ítem en una categoría de respuesta específica a través de una Función de Respuesta al Ítem (IRF, con sus siglas en inglés).

Todos estos modelos asumen el principio de independencia lineal y que los ítems del test miden una habilidad en común. La diferencia está en la forma matemática adoptada por la IRF y el sistema de puntuación dicotómico (cuando solo hay dos opciones de respuesta, por ejemplo; correcto o incorrecto, sí o no, etc.) y politómico (cuando hay más de dos opciones, siendo generalmente escalas de apreciación graduadas, por ejemplo; escala de respuesta estilo Likert, respuestas de opción múltiple o puntuación de crédito parcial).

En un modelo de respuesta al ítem con puntuación dicotómica, la IRF es la unidad esencial para modelar las respuestas del ítem y puede representarse mediante una Curva Carac-

⁴La psicofísica es una rama de la psicología que estudia la relación entre la magnitud de los estímulos físicos y la intensidad percibida por el observador. Su objetivo es poder encontrar la escala en la que se puede situar dicha relación.

⁵El escalamiento psicofísico implica establecer una escala de medición de dimensiones psicológicas con referencia física directa.

terística del Ítem (ICC, con sus siglas en inglés).

Por otro lado, en la modelización de respuestas politómicas, la unidad esencial está en el nivel de la categoría que es la Función de Respuesta de Categoría o Función de Probabilidad de Categoría (CRF, con sus siglas en inglés), y puede representarse gráficamente mediante una Curva Característica de Categoría o Curva de Probabilidad de Categoría (CCC, con sus siglas en inglés) (Paek y Cole, 2020).

Surge entonces, una amplia gama de modelos estadísticos clasificados según la base de sus respuestas, y para obtener una medida válida del rasgo latente a estudiar, los supuestos fundamentales de la IRT son:

1. Independencia experimental: las respuestas de cada sujeto s son independientes con respecto a las respuestas de los otros sujetos.
2. Independencia local: dada la habilidad ψ para un sujeto s , las respuestas a los ítems son independientes entre sí.
3. Invarianza de los parámetros: los parámetros de los ítems son invariantes sobre las muestras de los sujetos. El rasgo latente también es invariante con respecto a los ítems del test.

1.2.1. Modelo de ojiva normal de 2 parámetros

El psicómetra Frederic Lord y el matemático George Rasch no tuvieron la confusión entre las funciones de distribución y su uso como funciones de respuesta.

El modelo de ojiva normal de 2 parámetros (Lord, 1952), es un modelo unidimensional con naturaleza de data dicotómica; en el que la ICC tiene la forma de una función ojiva normal de dos parámetros. Esta función representa la probabilidad de una respuesta correcta dada la habilidad desconocida θ y está definida por

$$p_i(\theta) = \int_{-\infty}^{a_i(\theta - b_i)} \phi(z) dz$$

donde $i = 1, \dots, n$ para un test de n ítems, $p_i(\theta)$ es la probabilidad de que el sujeto seleccionado al azar con habilidad θ responda correctamente al ítem i , a_i representa la discriminación de los ítems (proporcional a la pendiente de $p_i(\theta)$ en el punto $\theta = b_i$), b_i representa el índice de dificultad del ítem i y el punto en el cual el sujeto tiene un 50% de probabilidad de responder bien al ítem, y $\phi(z)$ que simboliza la función de densidad de una variable z con distribución normal estándar.

En la IRT, los valores b positivos más altos reflejan ítems más difíciles, mientras que los valores negativos más bajos son indicativos de ítems más fáciles. Así mismo, los valores altos de a implican ICCs muy empinados y valores bajos conllevan ICCs crecientes en función de la habilidad.

Rasch por su lado, introdujo su modelo como un intento de cambiar el paradigma de la investigación social y conductual de manera más fundamental, y aunque era consciente que podía representar su modelo con una función cercana a la función ojiva normal utilizada por Lord mediante el uso de la transformada logística, apenas la usó.

En los años siguientes, los desarrollos del modelo ojiva normal se detuvieron debido a la simplicidad que ofrece la función logística

$$p(x) = \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-x}}$$

como IRF.

1.2.2. Modelo logístico de 1 parámetro

El modelo logístico de 1 parámetro (1PL; Rasch, 1960), involucró dos tipos diferentes de parámetros, los de habilidad del individuo que se evalúa y los de dificultad para los ítems del test.

En este modelo, la ICC es una función logística de un solo parámetro respecto a la característica del ítem

$$p_i(\theta) = \frac{e^{D\bar{a}(\theta-b_i)}}{1 + e^{D\bar{a}(\theta-b_i)}}$$

para $i = 1, \dots, n$, donde \bar{a} representa el nivel común de discriminación para todos los ítems (lo que la convertiría en una constante) y los elementos $p_i(\theta)$, b_i representan lo mismo que en el modelo de Lord. D es un factor de escala.

Con este modelo podemos estimar la probabilidad de que el sujeto responda correctamente a un ítem del test; dado un parámetro, en este caso la dificultad b , y una estimación de la habilidad θ .

La figura 1.1 muestra un ejemplo de ICCs para dos ítems de un test, basados en este modelo.

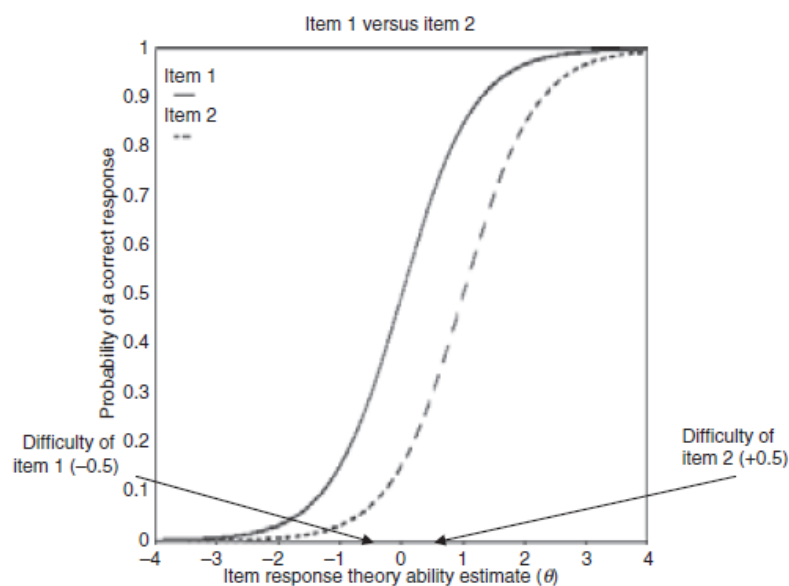


Figura 1.1: Ejemplo de ICCs para dos ítems de test basados en el modelo 1PL (Champlain, 2010).

El eje x corresponde a los valores de estimación de la habilidad. Los valores de θ más altos se asocian a candidatos más aptos. Y la probabilidad de una respuesta correcta al ítem se muestra a lo largo del eje y (que va de 0 a 1). Esto quiere decir, que cuánto más habilidad tenga el sujeto, más probable es que responda correctamente al ítem. Por tanto, en el ejemplo de la figura 1.1; el ítem 2 ($b = +0.5$) es más difícil que el ítem 1 ($b = -0.5$).

En 1960, Rasch introdujo otros dos modelos con una estructura de parámetros similar, un modelo Poisson para los errores de lectura y un modelo gamma para el tiempo de lectura.

En 1961, Rasch también propuso un modelo unidimensional general para ítems con múltiples categorías de respuesta que incluía su modelo dicotómico. Incluso sentó las bases estadísticas para la estimación de los parámetros y los tests estadísticos de la bondad del ajuste de sus modelos.

Cada uno de los modelos de Rasch fue formulado para pertenecer a la familia exponencial y su tratamiento estadístico se basó en el tipo de inferencia condicional que admite dicha familia.

En cuanto al modelo de Lord de 2 parámetros, el papel de Alan Birnbaum fue decisivo con el cambio de la función de respuesta normal a la función de respuesta logística. Birnbaum sugirió la misma transformación para el modelo de Lord.

1.2.3. Modelo logístico de 2 parámetros

El modelo logístico de 2 parámetros (2PL; Birnbaum, 1957), representa un modelo unidimensional de naturaleza dicotómica cuya ICC toma la forma de una función logística de dos parámetros (respecto a la característica del ítem); esta función está dada por

$$p_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}$$

para $i = 1, \dots, n$ donde $p_i(\theta)$, a_i , b_i y θ y D representan lo mismo que en el modelo 1PL.

En la Figura 1.2 se muestra ICCs basados en este modelo para dos ítems donde la dificultad $b = 0$ para ambos ítems es la misma, no obstante el parámetro del ítem 2 es más elevado que el del ítem 1 por ello su ICC es más pronunciada. Como se mencionó, la discriminación del ítem define la pendiente del ICC.

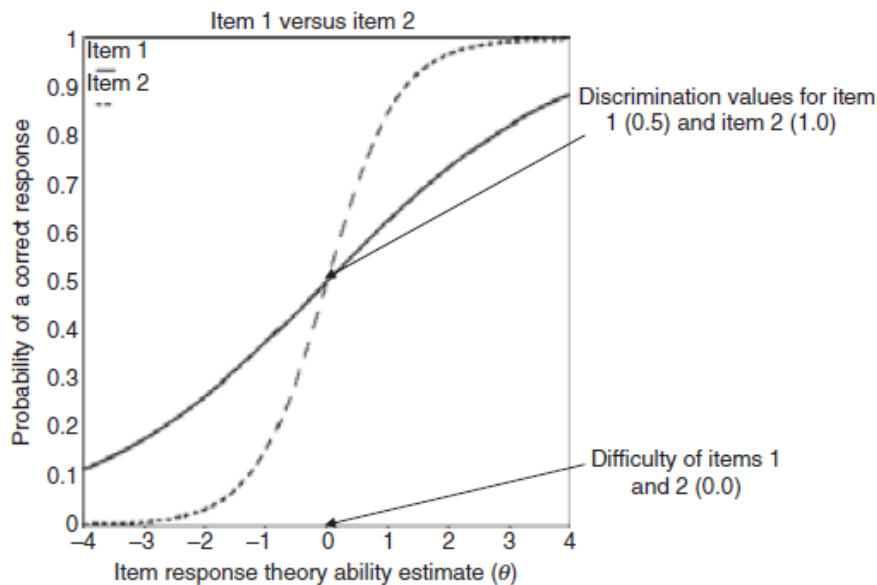


Figura 1.2: Ejemplo de ICCs para dos ítems de test basados en el modelo 2PL (Champlain, 2010).

Birnbaum además, sugirió la adición de un parámetro de adivinación al mismo y básicamente lanzó lo que se ha conocido como el *modelo logístico de tres parámetros (3PL)*.

Este modelo puede ser obtenido del modelo logístico de 2 parámetros, solo que añadiendo otro parámetro c_i que representa la asíntota inferior del ICC y supone la probabilidad de que el sujeto con baja habilidad responda correctamente a un ítem; debido a eso, al parámetro se lo denomina *parámetro pseudo-azar de nivel*.

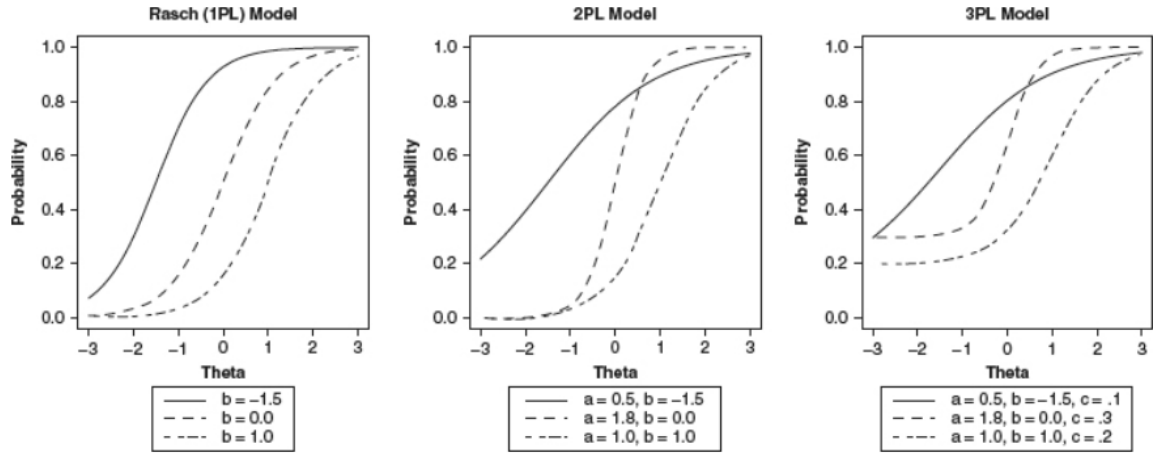


Figura 1.3: Ejemplo de ICCs para tres ítems de test en los distintos modelos 1PL, 2PL y 3PL (Champlain, 2010).

$$p_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

para $i = 1, \dots, n$ y con un factor de escala $D = 1.7$.

En la Figura 1.3 se muestra la diferencias de las ICCs para tres ítems en los distintos modelos 1PL, 2PL y 3PL, con valores variables del parámetro de discriminación, del parámetro de dificultad y del parámetro de adivinación en el modelo correspondiente.

Notemos que la ecuación logística de dos parámetros es un caso especial de la ecuación logística de tres parámetros en la que c_i se fija en 0.

Finalmente, los modelos logísticos y de ojiva normal vistos, solo pueden aplicarse a ítems de test cuya puntuación es dicotómica, pero existen modelos utilizables cuando los ítems se puntúan de forma politómica, uno de ellos es el siguiente.

1.2.4. Modelo de Respuesta Graduada

El modelo de respuesta graduada (GRM; Samejima, 1969), se ocupa de categorías politómicas ordenadas y es utilizado generalmente en encuestas de actitud o de crédito parcial otorgado de acuerdo con el grado de logro del sujeto en la resolución de un problema.

Entre estas categorías se incluye una calificación denominada *escala ordinal de Likert* que presenta al sujeto m respuestas posibles y asigna un número a cada una de ellas, representando así un conjunto de categorías politómicas ordenadas, por ejemplo, 0, 1, 2 y 3 (o 1, 2, 3 y 4) representa un conjunto de categorías politómicas ordenadas de cuatro niveles que simboliza una calificación de *totalmente en desacuerdo*, *en desacuerdo*, *de acuerdo* y *totalmente de acuerdo*.

En 1969, Samejima propuso dos tipos de modelos GRM para datos de respuesta graduada, basándose en un modelo ojiva normal y un modelo logístico.

Como ya se ha mencionado, los modelos logísticos son los que reciben más atención y uso por parte de los psicometristas, en razón de que, las curvas logísticas son más convenientes para trabajar que las curvas ojivas normales. (Ojala y Larsson, 2016).

Sea entonces θ el rasgo latente o la habilidad que representa un constructo hipotético subyacente a cierto comportamiento humano, como la habilidad de resolver problemas. Para un conjunto de I ítems (la unidad más pequeña para medir θ) y para cada sujeto s , existe un patrón de respuesta,

$$x_s = (x_{s_1}, x_{s_2}, \dots, x_{s_I}) \quad (1.1)$$

donde x_{s_i} denota la respuesta de individuos s al ítem i con $i = 1, \dots, I$.

La probabilidad con la que un sujeto con habilidad θ_s recibe una puntuación x_{s_i} , se denomina, *característica operativa* y no es más que la probabilidad de elegir una respuesta x_{s_i} para un ítem i dado un determinado nivel de rasgo latente θ_s , $P(X_{s_i} = x_{s_i} | \theta_s)$.

Por otro lado, si $k \in \{1, 2, \dots, m\}$ denota las posibles respuestas al ítem i , el modelo GRM utiliza una respuesta acumulativa para modelar la probabilidad acumulada de dar una respuesta al ítem $\geq k$,

$$P(X_{s_i} \geq k | \theta_s) = \frac{e^{a_i(\theta_s - b_{i_k})}}{1 + e^{a_i(\theta_s - b_{i_k})}} \quad (1.2)$$

donde x_{s_i} es la respuesta del sujeto s al ítem i en la categoría k , a_i es el parámetro de discriminación del ítem i , b_{i_k} es el parámetro de dificultad para la respuesta k al ítem i y θ_s es el rasgo latente o la puntuación de habilidad del sujeto s . Por comodidad, a la ecuación (1.2) la notaremos como $P_{i_k}^*$.

Observación 1 *Los parámetros de dificultad para la respuesta k al ítem i del modelo GRM siempre están ordenados, esto es, $b_{i_k} \geq b_{i_{k-1}}$.*

De esta manera, la probabilidad de responder específicamente en la categoría k , $x_{s_i} = k$, se obtiene restando la probabilidad acumulada para la respuesta de ese ítem k y la probabilidad acumulada de respuestas mayores que k ,

$$P_{i_k}(\theta) = P(X_{s_i} = k | \theta_s) = P(X_{s_i} \geq k - 1 | \theta_s) - P(X_{s_i} \geq k | \theta_s) \quad (1.3)$$

$$= P_{i_{k-1}}^*(\theta) - P_{i_k}^*(\theta) \quad (1.4)$$

Samejima define a las probabilidades $P_{i_0}^*$ y $P_{i_m}^*$ tal que $P_{i_0}^*(\theta) = 1$ y $P_{i_m}^*(\theta) = 0$.

A través de la ecuación 1.4, podemos obtener todas las probabilidades de respuesta al ítem,

$$P_{i_k}(\theta) = \begin{cases} 1 - \frac{e^{a_i(\theta_s - b_{i_k})}}{1 + e^{a_i(\theta_s - b_{i_k})}} & \text{si } k = 1 \\ \frac{e^{a_i(\theta_s - b_{i_{k-1}})}}{1 + e^{a_i(\theta_s - b_{i_{k-1}})}} - \frac{e^{a_i(\theta_s - b_{i_k})}}{1 + e^{a_i(\theta_s - b_{i_k})}} & \text{si } 1 < k < m \\ \frac{e^{a_i(\theta_s - b_{i_{k-1}})}}{1 + e^{a_i(\theta_s - b_{i_{k-1}})}} & \text{si } k = m \end{cases} \quad (1.5)$$

estas probabilidades son las CCCs del modelo GRM.

Observación 2 *El modelo GRM en la ecuación 1.2 modela las probabilidades de respuesta acumuladas, no las CCCs.*

Para cada ítem, se tendrá por tanto m CCCs, que definirán la probabilidad de la respuesta al ítem, dado el valor del rasgo latente del sujeto.

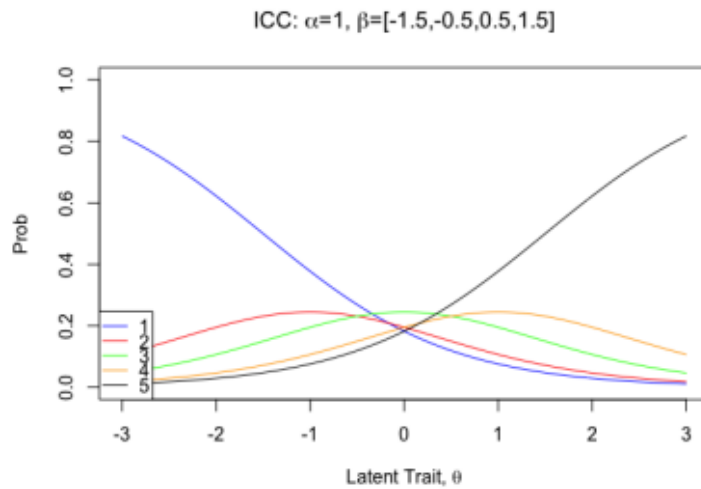
Observación 3 *Un α más alto da lugar a una curva de probabilidad más pronunciada y distinta para cada respuesta separada, como resultado, concatenan más información en comparación con los ítems con menor discriminación (véase Figura 1.4).*

Observación 4 *El parámetro de dificultad β desplaza la posición de la curva. Una dispersión más dispersa de los β s, permite CCCs más distintos y luego es más fácil identificar el nivel del rasgo latente de un sujeto dada la respuesta del ítem (véase Figura 1.5).*

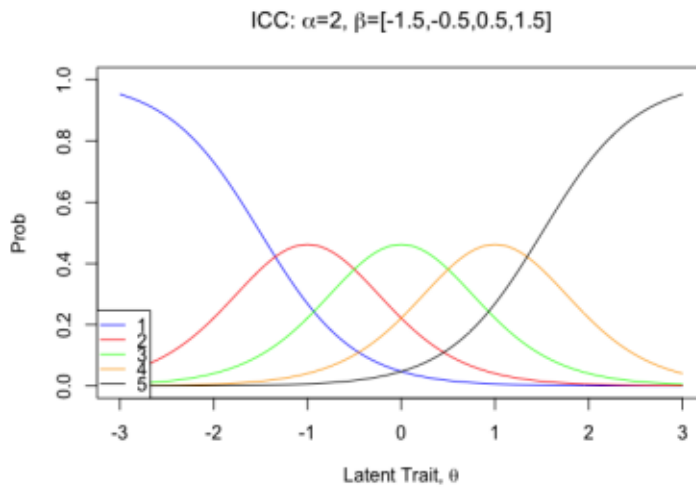
1.3. Criterio de Información del Ítem

Una de las medidas de confiabilidad que proporciona la IRT aparte de las puntuaciones de los tests, es el Criterio de Información del Ítem (IIC por sus siglas en inglés) que representa la información facilitada por un ítem específico sobre el continuo latente θ .

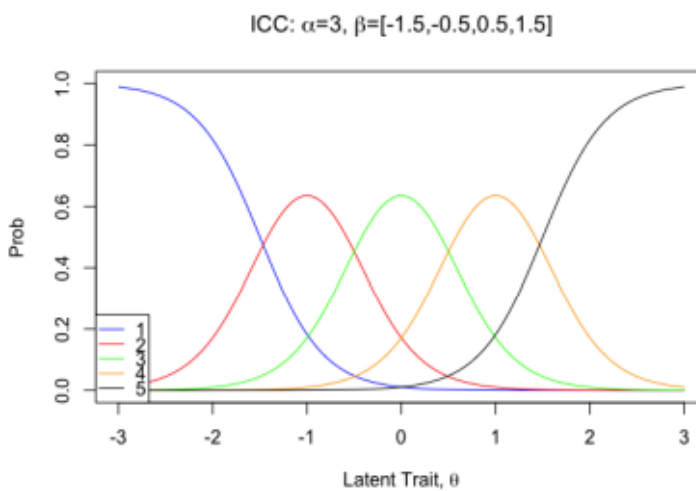
Aunque lo que interesa es la información del test como indicador de la precisión con la que se estima la habilidad. La información del test se define en términos de funciones de información del ítem. Por lo tanto, la discusión se centra en la función de información de los ítems en el contexto de la calificación de las respuestas.



(a) Se muestra una baja discriminación, por tanto, es difícil diferenciar las respuestas de los sujetos.

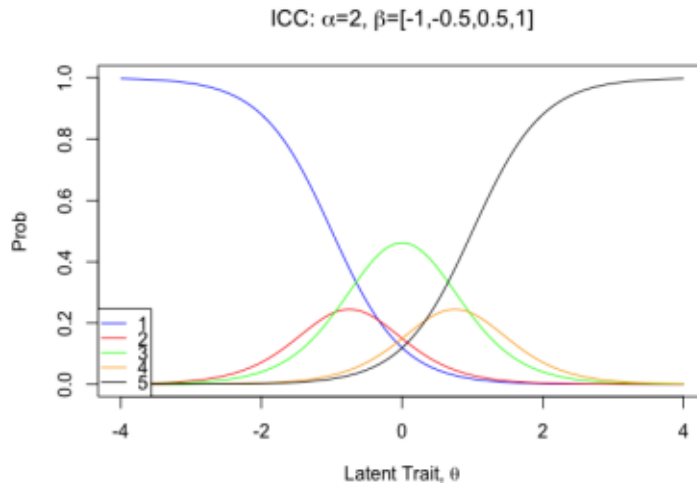


(b) Se muestra una discriminación media. Es posible diferenciar las respuestas de los sujetos.

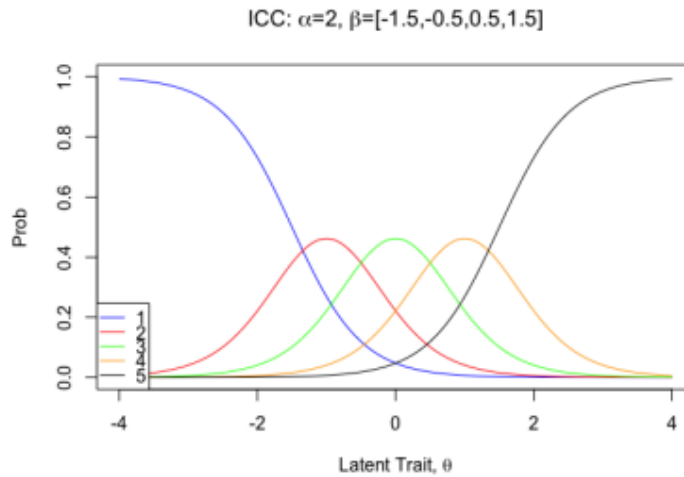


(c) Se muestra una alta discriminación, por tanto, es fácil diferenciar entre las respuestas de los sujetos.

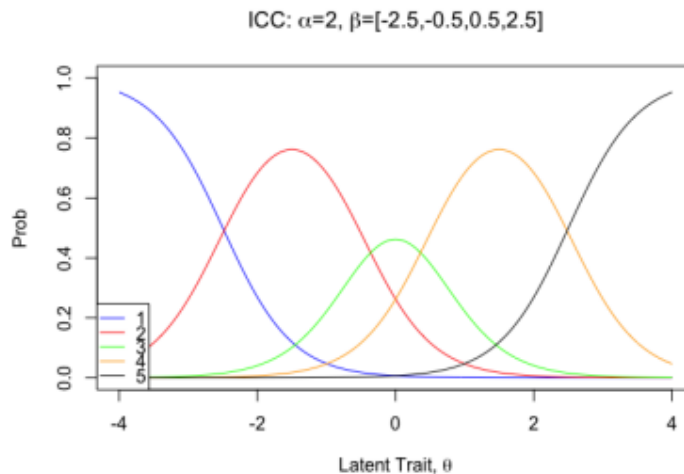
Figura 1.4: ICCs para el modelo GRM con baja, media y alta discriminación α (Ojala y Larsson, 2016).



(a) Se muestra una baja dispersión de los parámetros de dificultad, lo que impide la diferenciación de las ICCs con respuestas $x_{s_i} = 2, 3, 4$.



(b) Se muestra una dispersión media de los parámetros de dificultad, lo que facilita la diferenciación de las diferentes ICCs.



(c) Se muestra una alta dispersión de los parámetros de dificultad, lo que da como resultado ICCs definidas.

Figura 1.5: ICCs para el modelo GRM con baja, media y alta dificultad β (Ojala y Larsson, 2016).

Dado que hay información asociada con cada categoría de respuesta, la Función de Información del Ítem (IIF por sus siglas en inglés) en el nivel de habilidad θ se define como

$$I_i(\theta) = \sum_{k=1}^m I_{i_k}(\theta) P_{i_k}(\theta) \quad i = 1, \dots, n \quad (1.6)$$

donde $I_{i_k}(\theta) P_{i_k}(\theta)$ es la cantidad de información compartida de la categoría k .

Samejima define la IIC para modelos de respuesta al ítem politómicos, como

$$I_{i_k}(\theta) = -\frac{\partial^2 \log P_{i_k}(\theta)}{\partial \theta^2} = -\frac{\partial}{\partial \theta} \left[\frac{P'_{i_k}(\theta)}{P_{i_k}(\theta)} \right]$$

donde $P'_{i_k}(\theta) = \partial P_{i_k}(\theta) / \partial \theta$. Entonces,

$$I_{i_k}(\theta) = \frac{[P'_{i_k}(\theta)]^2 - P_{i_k}(\theta) P''_{i_k}(\theta)}{[P_{i_k}(\theta)]^2}$$

en el que $P''_{i_k}(\theta) = \partial P'_{i_k}(\theta) / \partial \theta$. Por tanto, la información compartida de la categoría de respuesta del ítem es

$$I_i(\theta) = \sum_{k=1}^m \frac{[P'_{i_k}(\theta)]^2 - P_{i_k}(\theta) P''_{i_k}(\theta)}{[P_{i_k}(\theta)]^2} P_{i_k}(\theta) \quad (1.7)$$

$$= \sum_{k=1}^m \left(\frac{[P'_{i_k}(\theta)]^2}{P_{i_k}(\theta)} - P''_{i_k}(\theta) \right). \quad (1.8)$$

En vista de que, $P_{i_k}(\theta) = P_{i_{k-1}}^*(\theta) - P_{i_k}^*(\theta)$, tenemos

$$\begin{aligned} P'_{i_k}(\theta) &= P_{i_{k-1}}^{*'}(\theta) - P_{i_k}^{*'}(\theta) \\ P''_{i_k}(\theta) &= P_{i_{k-1}}^{*''}(\theta) - P_{i_k}^{*''}(\theta). \end{aligned}$$

Sustituyendo estas relaciones en la ecuación 1.8, se obtiene a $I_i(\theta)$ como

$$I_i(\theta) = \sum_{k=1}^m \frac{[P_{i_{k-1}}^{*'}(\theta) - P_{i_k}^{*'}(\theta)]^2}{P_{i_{k-1}}^*(\theta) - P_{i_k}^*(\theta)} - \sum_{k=1}^m [P_{i_{k-1}}^{*''}(\theta) - P_{i_k}^{*''}(\theta)].$$

Samejima muestra además que el segundo término de la derecha desaparece y luego la función de información del ítem resulta ser

$$I_i(\theta) = \sum_{k=1}^m \frac{[P_{i_{k-1}}^{*'}(\theta) - P_{i_k}^{*'}(\theta)]^2}{P_{i_{k-1}}^*(\theta) - P_{i_k}^*(\theta)} \quad (1.9)$$

en la cual $P'_{i_k}(\theta) = P_{i_k}^*(\theta)(1 - P_{i_k}^*(\theta))$.

Debido a la suposición de independencia local de la IRT, la información del test es la información combinada de todos los ítems en un dominio, por tanto, las funciones de información del ítem son aditivas y la información del test se obtiene como

$$I(\theta) = \sum_{i=1}^n I_i(\theta). \quad (1.10)$$

Una característica importante de la definición dada en la ecuación (1.10) es que cuantos más ítems haya en el dominio, mayor será la cantidad de información.

Este capítulo, establece la modelización de los modelos principales de la IRT y al término de todo esto, dado que el test de nuestro interés mide la habilidad del potencial de innovación de una organización e involucra observaciones de respuesta politómica en una escala de respuesta estilo Likert, el modelo GRM es elegible para esta investigación.

Capítulo 2

Estimación de parámetros

Existen diversas técnicas que permiten dar un valor aproximado a los parámetros de un modelo IRT. La complejidad y el tiempo de estimación de estos dependen de varios factores como el número de ítems, el tamaño de la muestra, el número de parámetros, la parametrización del modelo y el número de dimensiones.

En virtud de ello, las técnicas se emplean en función de las características y propósitos del estudio en cuestión. Por ejemplo, están los enfoques estadísticos clásicos o frecuentistas como la estimación por Máxima Verosimilitud, intervalos de confianza o p-valores que se utilizan comúnmente en los tests operativos y enfoques estadísticos bayesianos como el método Monte Carlo con cadenas de Markov. Pese a que los métodos frecuentistas son los más disponibles, los parámetros de nuestro modelo se estimarán mediante métodos bayesianos que cada vez son más utilizables en la IRT.

2.1. Estimación Bayesiana

Los métodos inferenciales bayesianos requieren un modelo de muestreo que describa al fenómeno aleatorio que generó los datos observados. El modelo de muestreo suele definirse por un pequeño número de parámetros en relación con el tamaño de la muestra, y puede incluir información de variables explicativas, como las variables predictoras en un análisis de regresión. En la IRT, por ejemplo, si se tiene una matriz de respuestas binarias como datos observados; se puede asumir que el modelo logístico 2PL es el modelo de muestreo.

Uno de los principales intereses en la inferencia bayesiana es obtener información sobre la distribución a posteriori, que es la función de los datos recopilados en forma de función de verosimilitud definida por el modelo de muestreo.

Sea $\mathbf{x} = \{x_1, \dots, x_n\}$ una muestra recogida de un conjunto de datos de tamaño n y sea $\boldsymbol{\psi} = \{\psi_1, \dots, \psi_m\}$ los m parámetros asociados al modelo supuesto para \mathbf{x} , la distribución conjunta de los datos \mathbf{x} y el vector de parámetros $\boldsymbol{\psi}$ se obtiene multiplicando la función de verosimilitud y la distribución a priori

$$p(\mathbf{x}, \boldsymbol{\psi}) = p(\mathbf{x}|\boldsymbol{\psi})\pi(\boldsymbol{\psi})$$

cuya integración sobre el espacio de parámetros da la densidad de la distribución marginal de los datos observados, $p(\mathbf{x}) = \int p(\mathbf{x}, \boldsymbol{\psi})d\boldsymbol{\psi}$.

Observación 5 La notación $p(\cdot)$ representa varias funciones de probabilidad discretas y continuas, en la cual las integrales $\int g(x)f(dx)$ deben interpretarse como integrales de Stieltjes, es decir, si $p(\cdot)$ es una densidad continua, esta es una integral habitual, y si $p(\cdot)$ es una densidad discreta, esta es una suma.

La distribución a posteriori también definida *densidad posterior* o *distribución posterior* de $\boldsymbol{\psi}$ dado \mathbf{x} , puede darse utilizando el Teorema de Bayes,

$$\pi(\boldsymbol{\psi}|\mathbf{x}) = \frac{p(\mathbf{x}, \boldsymbol{\psi})}{p(\mathbf{x})} = \frac{p(\mathbf{x}|\boldsymbol{\psi})\pi(\boldsymbol{\psi})}{\int p(\mathbf{x}|\boldsymbol{\psi})\pi(\boldsymbol{\psi})d\boldsymbol{\psi}} \quad (2.1)$$

$$\propto p(\mathbf{x}|\boldsymbol{\psi})\pi(\boldsymbol{\psi}). \quad (2.2)$$

Nótese que en la ecuación (2.2) se ha omitido el término $p(\mathbf{x})$ y se ha sustituido la igualdad previa por una proporcionalidad. Esta es una forma útil del teorema de Bayes para los problemas de estimación de parámetros, ya que el denominador es solo una constante de normalización.

Observación 6 La función de verosimilitud $p(\mathbf{x}|\boldsymbol{\psi})$ está dominada por la medida de Conteo, ya que el patrón de respuestas con el que vamos a trabajar es contable. A su vez, la distribución a priori está dominada por la medida de Lebesgue, puesto que $\boldsymbol{\psi}$ es subconjunto de \mathbb{R}^m . Por tanto, la distribución posterior se encuentra dominada por el producto de ambas medidas. Para más información de las medidas producto, véase Chamorro, 2010, p.182.

Para muchos modelos estadísticos, incluidos los modelos IRT no siempre es fácil obtener la distribución posterior. Sin embargo, bajo ciertas condiciones de regularidad, la inferencia bayesiana puede utilizarse para construir una cadena de Markov que resuma las características de dicha distribución. La simulación de esta cadena logra aproximarse métodos Monte Carlo con cadenas de Markov.

2.1.1. Monte Carlo con cadenas de Markov

En Estadística, el éxito del método Monte Carlo mediante cadenas de Markov (MCMC), se impulsa por la medida en que hace que el cálculo y la estimación de modelos bayesianos novedosos y complejos sean manejables.

La IRT trata fundamentalmente con datos de respuestas codificadas a ítems, es decir, datos bidireccionales que consisten en respuestas, x_{si} , de los sujetos $s = 1, \dots, S$ a ítems $i = 1, \dots, I$ (tareas, estímulos, preguntas de test, etc.).

La respuesta de cada sujeto a los ítems constituye la matriz completa de respuesta, $\mathcal{X} = [x_{si}] \in \mathbb{N}^{S \times I}$, dando pie de uso a una familia de modelos probabilísticos

$$p(\mathcal{X}|\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\gamma}) \quad (2.3)$$

donde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$ es un conjunto de parámetros del sujeto, $\mathbf{B} = (\beta_1, \dots, \beta_p)$ es un conjunto de parámetros del ítem y $\boldsymbol{\gamma}$ un conjunto adicional de otros parámetros.

Observación 7 *Por facilidad de escritura, se recopila en un solo vector J -dimensional, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_J)$, a los parámetros $\boldsymbol{\theta}$, \mathbf{B} y $\boldsymbol{\gamma}$.*

Como se menciona en la sección 2.1, si proporcionamos una distribución a priori $\pi(\boldsymbol{\tau})$, sabemos que la distribución conjunta de los datos y los parámetros puede escribirse como $p(\mathcal{X}, \boldsymbol{\tau}) = p(\mathcal{X}|\boldsymbol{\tau})\pi(\boldsymbol{\tau})$. El problema está entonces en conocer la distribución posterior $\pi(\boldsymbol{\tau}|\mathcal{X})$ que se dificulta por la necesidad de calcular la integral en el denominador de la ecuación (2.1). En la mayoría de las aplicaciones, esta integral debe calcularse numéricamente en lugar de analíticamente y suele tener una dimensión bastante alta.

Observación 8 *La dimensión J del espacio de parámetros es al menos tan grande como el número de sujetos más el número de ítems, $S + I$.*

El MCMC ofrece una metodología directa para generar muestras de la distribución posterior $\pi(\boldsymbol{\tau}|\mathcal{X})$ en modelos bayesianos que se comportan bien, sin calcular directamente la integral de la ecuación (2.1).

La idea esencial es definir una cadena de Markov estacionaria M_0, M_1, M_2, \dots con estados $M_l = (\boldsymbol{\tau}^{(l)})$ y luego simular nuevas observaciones a partir de ella. Su comportamiento está determinado por un núcleo de transición

$$\tilde{k}(\mathbf{t}^{(0)}, \mathbf{t}^{(1)}) = f [M_l = (\mathbf{t}^{(1)}) | M_{l-1} = (\mathbf{t}^{(0)})] \quad \forall l$$

que es la función de densidad de transición de pasar al nuevo estado $\mathbf{t}^{(1)}$ dado el estado actual $\mathbf{t}^{(0)}$.

Si una densidad de probabilidad $\tilde{\pi}$ satisface

$$\int \tilde{\pi}(\mathbf{t}^{(0)}) \tilde{k}(\mathbf{t}^{(0)}, \mathbf{t}^{(1)}) dt = \tilde{\pi}(\mathbf{t}^{(1)}) \quad (2.4)$$

entonces $\tilde{\pi}(\mathbf{t})$ es la distribución estacionaria de la cadena de Markov,

$$M_l \sim \tilde{\pi} \implies M_{l+1} \sim \tilde{\pi}.$$

Observación 9 Una condición común para satisfacer la ecuación (2.4) es la reversibilidad, $\tilde{\pi}(\mathbf{t}^{(0)}) \tilde{k}(\mathbf{t}^{(0)}, \mathbf{t}^{(1)}) = \tilde{\pi}(\mathbf{t}^{(1)}) \tilde{k}(\mathbf{t}^{(1)}, \mathbf{t}^{(0)})$.

Las cadenas de Markov deben satisfacer las siguientes condiciones para el uso de los métodos MCMC. Véase la documentación Douc y col., 2019 para más detalle.

- *Homogeneidad*: La probabilidad de transición de un estado a otro es invariable.

Definición 1 Una distribución de probabilidad que permanece constante en una cadena de Markov a lo largo del tiempo se la denomina distribución estacionaria.

- *Irreducibilidad*: Cada estado puede ser alcanzado desde cualquier otro estado en un número finito de iteraciones.

Definición 2 Una cadena de Markov es $\tilde{\pi}$ -irreducible si tiene probabilidad positiva de entrar en cualquier conjunto A para el que

$$\tilde{\pi}(A) = \int_A \tilde{\pi}(d\mathbf{t}) > 0.$$

- *Aperiodicidad*: Sin estados absorbentes.

Definición 3 Una cadena de Markov es periódica si hay partes del espacio de estados que puede visitar solo a intervalos regularmente espaciados; de lo contrario, es aperiódico.

Así pues, si una cadena de Markov tiene una distribución estacionaria $\tilde{\pi}()$ como en la ecuación (2.4) y es $\tilde{\pi}$ -irreducible y aperiódica, entonces la distribución de M_l convergerá a $\tilde{\pi}()$ cuando $l \rightarrow \infty$ (Tierney, 1994).

Es factible definir el núcleo de transición $\tilde{k}(\mathbf{t}^{(0)}, \mathbf{t}^{(1)})$ de modo que $\tilde{\pi}(\mathbf{t}) = \pi(\mathbf{t}|\mathcal{X})$ en la ecuación (2.4) luego de los l primeros pasos de la cadena, pues debido a la naturaleza markoviana de la simulación, los primeros valores dependen del valor inicial M_0 y normalmente se eliminan de la muestra como *burn-in* o “warm-up”. Con ello, las observaciones restantes

$$(\boldsymbol{\tau}^{(1)}) = M_{l+1}, (\boldsymbol{\tau}^{(2)}) = M_{l+2}, \dots, (\boldsymbol{\tau}^{(L)}) = M_{l+L}$$

pueden tratarse como extracciones (dependientes) de la distribución posterior $\pi(\boldsymbol{\tau}|\mathcal{X})$ y así, brindar información sobre las propiedades de los parámetros.

A continuación se presentan los métodos computacionales más conocidos que utilizan los principios markovianos para la estimación de parámetros.

Muestreo de Gibbs

Construir el núcleo de transición $\tilde{k}(\mathbf{t}^{(0)}, \mathbf{t}^{(1)})$ para que la distribución estacionaria de la cadena de Markov sea la distribución posterior $\pi(\boldsymbol{\tau}|\mathcal{X})$, es simple. Sea $(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$ una partición disjunta del vector de parámetros $\boldsymbol{\tau}$. Un breve cálculo para verificar la ecuación (2.4) muestra que

$$\tilde{k}(\boldsymbol{\tau}^{(0)}, \boldsymbol{\tau}^{(1)}) = \pi(\boldsymbol{\tau}_1^{(1)}|\boldsymbol{\tau}_2^{(0)}, \mathcal{X}) \pi(\boldsymbol{\tau}_2^{(1)}|\boldsymbol{\tau}_1^{(1)}, \mathcal{X}) \quad (2.5)$$

tiene la distribución estacionaria $\pi(\boldsymbol{\tau}|\mathcal{X})$.

Generalizando la idea. Sea una partición disjunta $(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_H)$ cualquiera, pero fija, del vector de parámetros $\boldsymbol{\tau}$ en H bloques, tal que $H \leq J$. Se define el esquema de muestreo para pasar de $M_{l-1} = (\boldsymbol{\tau}_1^{(l-1)}, \boldsymbol{\tau}_2^{(l-1)}, \dots, \boldsymbol{\tau}_H^{(l-1)})$ a $M_l = (\boldsymbol{\tau}_1^{(l)}, \boldsymbol{\tau}_2^{(l)}, \dots, \boldsymbol{\tau}_H^{(l)})$ en la cadena de Markov

1. Muestra $\boldsymbol{\tau}_1^{(l)}$ de $\pi(\boldsymbol{\tau}_1|\boldsymbol{\tau}_2^{(l-1)}, \dots, \boldsymbol{\tau}_H^{(l-1)}, \mathcal{X})$
2. Muestra $\boldsymbol{\tau}_2^{(l)}$ de $\pi(\boldsymbol{\tau}_2|\boldsymbol{\tau}_1^{(l)}, \boldsymbol{\tau}_3^{(l-1)}, \dots, \boldsymbol{\tau}_H^{(l-1)}, \mathcal{X})$
3. Muestra $\boldsymbol{\tau}_3^{(l)}$ de $\pi(\boldsymbol{\tau}_3|\boldsymbol{\tau}_1^{(l)}, \boldsymbol{\tau}_2^{(l)}, \boldsymbol{\tau}_4^{(l-1)}, \dots, \boldsymbol{\tau}_H^{(l-1)}, \mathcal{X})$
- ⋮
- H. Muestra $\boldsymbol{\tau}_H^{(l)}$ de $\pi(\boldsymbol{\tau}_H|\boldsymbol{\tau}_1^{(l)}, \boldsymbol{\tau}_2^{(l)}, \boldsymbol{\tau}_3^{(l)}, \dots, \boldsymbol{\tau}_{H-1}^{(l)}, \mathcal{X})$.

Las densidades condicionales de la derecha se denominan *densidades condicionales completas*, porque expresan la distribución de cada elemento de partición $\boldsymbol{\tau}_h$ condicionada a

todos los demás parámetros y datos del modelo (van der Linden, 2016).

Observación 10 *Se abrevia las variables condicionantes como resto para facilitar la notación, así, por ejemplo, la última ecuación del esquema anterior se reduce a $\pi(\boldsymbol{\tau}_H | rest)$.*

En consecuencia, el núcleo de transición para la partición $(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_H)$ consiste en el producto de los condicionales completos

$$\begin{aligned}
k(\boldsymbol{\tau}^{(l-1)}, \boldsymbol{\tau}^{(l)}) &= \pi(\boldsymbol{\tau}_1^{(l)} | \boldsymbol{\tau}_2^{(l-1)}, \dots, \boldsymbol{\tau}_H^{(l-1)}, \mathcal{X}) \\
&\times \pi(\boldsymbol{\tau}_2^{(l)} | \boldsymbol{\tau}_1^{(l)}, \boldsymbol{\tau}_3^{(l-1)}, \dots, \boldsymbol{\tau}_H^{(l-1)}, \mathcal{X}) \\
&\times \pi(\boldsymbol{\tau}_3^{(l)} | \boldsymbol{\tau}_1^{(l)}, \boldsymbol{\tau}_2^{(l)}, \boldsymbol{\tau}_4^{(l-1)}, \dots, \boldsymbol{\tau}_H^{(l-1)}, \mathcal{X}) \\
&\times \dots \times \pi(\boldsymbol{\tau}_H^{(l)} | \boldsymbol{\tau}_1^{(l)}, \boldsymbol{\tau}_2^{(l)}, \boldsymbol{\tau}_3^{(l)}, \dots, \boldsymbol{\tau}_{H-1}^{(l)}, \mathcal{X}) \\
&= \pi(\boldsymbol{\tau}_1^{(l)} | rest) \times \pi(\boldsymbol{\tau}_2^{(l)} | rest) \times \dots \times \pi(\boldsymbol{\tau}_H^{(l)} | rest)
\end{aligned} \tag{2.7}$$

con $\pi(\boldsymbol{\tau} | \mathcal{X})$ como su distribución estacionaria (van der Linden, 2016).

Los pasos para derivar las densidades condicionales completas son análogos para todos los parámetros. Seguidamente se muestra la derivación de $\pi(\boldsymbol{\tau}_1 | rest)$ manteniendo fijos los otros bloques $\boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_H$ y los datos \mathcal{X}

$$\begin{aligned}
\pi(\boldsymbol{\tau}_1 | \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_H, \mathcal{X}) &= \frac{p(\mathcal{X} | \boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_H) \pi(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_H)}{\int_{\boldsymbol{\tau}_1} p(\mathcal{X} | d\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_H) \pi(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_H)} \\
&\propto p(\mathcal{X} | \boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_H) \pi(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_H).
\end{aligned} \tag{2.8}$$

Notemos que cada densidad condicional completa es proporcional a la distribución conjunta en función de su bloque de parámetros. Así, cuando la función de verosimilitud $p(\mathcal{X} | \boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_H)$ y la distribución a priori $\pi(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_H)$ se convierten en un producto de términos que incluyen bloques separados de la partición $(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_H)$, es sencillo escoger una función proporcional al condicional completo, solo manteniendo los términos en la densidad conjunta que dependen de $\boldsymbol{\tau}_1$.

Observación 11 *Cuando los condicionales completos en el esquema (2.6) pueden determinarse en un número finito de pasos y pueden muestrearse, el algoritmo MCMC se denomina Muestreador de Gibbs (Geman y Geman, 1984).*

En resumen, este algoritmo muestrea subconjuntos (o bloques) del vector de parámetros a partir de la distribución posterior condicional del bloque de parámetros para los datos dados y todos los demás valores de los parámetros. Luego de iterar a través de estos bloques

(durante varios miles de pasos) genera una cadena de Markov cuya distribución límite es la distribución posterior objetivo.

Dado que el muestreo de Gibbs requiere la capacidad de tomar muestras de la distribución posterior condicional, pero en muchas aplicaciones estas son intratables, se requiere de métodos Monte Carlo alternativos como el algoritmo Metrópolis Hastings.

Metrópolis Hastings

El algoritmo Metrópolis Hasting (M-H; Hastings, 1970) implica un núcleo de transición diferente al que se describe en la ecuación (2.7), pero con la misma distribución estacionaria. Este algoritmo asociado con una densidad objetivo π requiere la elección de una densidad condicional o candidato a núcleo g conocida como *densidad propuesta*.

Implementar un paso M-H para el condicional completo $\pi(\boldsymbol{\tau} | rest)$; consiste en obtener un valor candidato para el parámetro $\boldsymbol{\tau}^*$, simulando un valor de densidad propuesta, es decir, muestrear $\boldsymbol{\tau}^* \sim g(\boldsymbol{\tau} | \boldsymbol{\tau}^{(m-1)})$ con $m = 1, \dots, M$ pasos.

La densidad propuesta puede elegirse como cualquier densidad conveniente. Dos opciones comunes son las siguientes:

- *Independencia M-H*: Toma como densidad propuesta a $g(\boldsymbol{\tau}^m | \boldsymbol{\tau}^{(m-1)}) = g(\boldsymbol{\tau}^m)$ independiente de $\boldsymbol{\tau}^{(m-1)}$.
- *Paseo aleatorio normal M-H*: Toma como densidad propuesta a

$$g(\boldsymbol{\tau}^m | \boldsymbol{\tau}^{(m-1)}) = \mathcal{N}\left(\boldsymbol{\tau} | \boldsymbol{\mu} = \boldsymbol{\tau}^{(m-1)}, \boldsymbol{\Sigma}\right).$$

Una densidad normal con media $\boldsymbol{\tau}^{(m-1)}$ y matriz de varianza $\boldsymbol{\Sigma}$.

Luego de elegir la densidad propuesta (la más común es la primera opción) se calcula la probabilidad de aceptación del nuevo candidato

$$\alpha^* = \min \left\{ \frac{\pi(\boldsymbol{\tau}^* | rest) g(\boldsymbol{\tau}^{(m-1)} | \boldsymbol{\tau}^*)}{\pi(\boldsymbol{\tau}^{(m-1)} | rest) g(\boldsymbol{\tau}^* | \boldsymbol{\tau}^{(m-1)})}, 1 \right\}.$$

Cuando la densidad propuesta es simétrica en $\boldsymbol{\tau}$ y $\boldsymbol{\tau}^{(m-1)}$, la probabilidad de aceptación se reduce a la relación entre la densidad a posteriori evaluada en el valor candidato sobre la densidad a posteriori evaluada en el valor anterior del parámetro, es decir, $\frac{\pi(\boldsymbol{\tau}^* | rest)}{\pi(\boldsymbol{\tau}^{(m-1)} | rest)}$. Caso contrario, si la densidad propuesta es asimétrica se toma en cuenta el término $\frac{g(\boldsymbol{\tau}^{(m-1)} | \boldsymbol{\tau}^*)}{g(\boldsymbol{\tau}^* | \boldsymbol{\tau}^{(m-1)})}$ que indica la probabilidad de que se proponga un valor $\boldsymbol{\tau}^{(m-1)}$, ya que la cadena se encuentra en

τ^* . Por último, el valor generado de una distribución uniforme, $X \sim \text{Unif}(0, 1)$, se compara con la probabilidad de aceptación fijando $\tau^{(m)} = \tau^*$ si $X \leq \alpha^*$; caso contrario, $\tau^{(m)} = \tau^{(m-1)}$.

Observación 12 *El muestreo de Gibbs es en realidad un caso especial de M-H, ya que si $g(\tau | \tau^{(m-1)}) = \pi(\tau | \text{rest})$ se acepta τ^* , porque todos los términos de la probabilidad de aceptación se cancelan, dejando $\alpha^* = 1$.*

Hasta ahora, se ha asumido la elección de una distribución a priori, pero se ha brindado poca orientación sobre cómo elegirla. De la misma manera que a menudo se modela la distribución muestral con distribuciones simples como la distribución normal, la distribución binomial o la distribución de Poisson, el Bayesiano paramétrico usa una distribución conveniente para la distribución a priori del parámetro τ . Sin embargo, debe tenerse en cuenta que su elección no tiene que tomarse a la ligera, dado el impacto que los a priori pueden tener en las inferencias finales.

Selección de una distribución a priori

Hay varias distribuciones a priori convenientes para la distribución a priori del parámetro τ ; una de estas es la clase de distribución a priori conjugada.

Definición 4 *Cuando las funciones $p(x|\tau)$ y $\pi(\tau)$ se combinan de tal manera que la distribución posterior pertenece a la misma familia que la distribución a priori, entonces se dice que $\pi(\tau)$ es conjugada para $p(x|\tau)$ (Ortiz, 2018).*

Entonces, si las distribuciones a priori y a posteriori son miembros de la misma clase de distribuciones, la distribución a priori es conjugada para una distribución de muestreo en particular.

Pese a que en modelos multiparamétricos encontrar la distribución conjugada multivariante suele ser difícil, el enfoque bayesiano suele elegir distribuciones a priori condicionalmente conjugadas, es decir, la priori condicional, $\pi(\tau_1|\tau_2)$, y la posterior condicional, $\pi(\tau_1|\tau_2, x)$, son miembros de la misma clase de distribuciones. Esto es conveniente cuando se usa el método MCMC para la estimación.

Los ejemplos del uso de a priori condicionales conjugadas en la IRT incluyen el caso de la media y la varianza de la distribución de capacidad, que en muchas aplicaciones se supone es una variable aleatoria independiente normal $\theta_s \sim \mathcal{N}(\mu, \sigma^2)$.

El uso de a priori conjugados y de a priori paramétricos o convenientes a menudo tiene el potencial de producir resultados sesgados. Por este motivo, muchos autores prefieren utilizar a priori no informativos o planos. Estos prior no informativos a menudo se denominan *a priori objetivos*, ya que el resultado final rara vez depende del a priori.

En resumen, la elección de los a priori está motivada por la facilidad de cálculo y cualquiera sea la distribución a priori conjugada seleccionada, es importante realizar un análisis de sensibilidad y evaluar la solidez de la inferencia bayesiana a ligeras modificaciones de la distribución a priori (Gelman y col., 2003).

Observación 13 *Este trabajo de investigación toma el análisis de sensibilidad realizado en el estudio del Índice de Innovación de Berkeley, que evalúa la capacidad de innovación de un individuo, empleando modelos IRT (Ojala y Larsson, 2016). En un conjunto de datos simulados, ejecutaron el análisis cambiando una distribución a priori con respecto a las priorizaciones de partida para determinar cómo afectaría al resultado. Como conclusión del análisis obtuvieron que al comparar los resultados de las diferentes opciones de distribución a priori, la bondad de ajuste era indiferente a pequeños cambios. Determinado así que son los datos y no las distribuciones a priori elegidas los que influyen en las estimaciones.*

2.1.2. Modelización GRM con el método MCMC

El modelo genérico de la ecuación (2.3) en las aplicaciones IRT puede reescribirse como

$$p(\mathcal{X}|\tau) = \prod_{s=1}^S \prod_{i=1}^I p(x_{s_i} | \theta_s, \beta_i) \quad (2.9)$$

donde recordemos que el producto sobre todos los sujetos s se debe al supuesto de independencia experimental entre los sujetos y el producto sobre todos los ítems i se debe al supuesto de independencia local entre ítems.

Es típico especificar a priori independientes para todos los parámetros de modo que la distribución conjunta genérica en la ecuación $p(\mathcal{X}, \tau) = p(\mathcal{X}|\tau)\pi(\tau)$ tenga la forma

$$\begin{aligned} p(\mathcal{X}|\tau)\pi(\tau) &= \prod_{s=1}^S \prod_{i=1}^I p(x_{s_i} | \theta_s, \beta_i) \prod_{s=1}^S \pi_s(\theta_s | \tilde{\lambda}_\theta) \prod_{i=1}^I \pi_i(\beta_i | \tilde{\lambda}_\beta) f(\tilde{\lambda}_\theta) f(\tilde{\lambda}_\beta) \\ &= \prod_{s=1}^S \left\{ \prod_{i=1}^I p(x_{s_i} | \theta_s, \beta_i) \pi_i(\beta_i | \tilde{\lambda}_\beta) \right\} \pi_s(\theta_s | \tilde{\lambda}_\theta) f(\tilde{\lambda}_\theta) f(\tilde{\lambda}_\beta) \end{aligned} \quad (2.10)$$

donde $\tilde{\lambda}_\theta, \tilde{\lambda}_\beta$ son hiperparámetros¹ para θ_s y β respectivamente, y $f(\tilde{\lambda}_\theta)f(\tilde{\lambda}_\beta)$ son las distribuciones a priori de los hiperparámetros.

Observación 14 *En nuestro caso, $\theta_s \in \mathbb{R}$ y $\beta_i = (a_i, b_i)$ con las componentes de discriminación y dificultad del modelo GRM (véase sección 1.2.4).*

Por otro lado, como la selección de las distribuciones a priori influye en las estimaciones de los parámetros, es necesario importar algunos supuestos. La elección de las distribuciones prior para θ_s y b_i siguen las ideas de Patz y Junker, 1999. Para a_i se toma la misma distribución del parámetro de dificultad

$$\pi_s(\theta_s|\tilde{\lambda}_\theta) := \mathcal{N}(\theta_s|\mu_\theta, \sigma_\theta^2) \quad (2.11)$$

$$\pi_i(a_i|\tilde{\lambda}_a) := \mathcal{N}(a_i|\mu_a, \sigma_a^2) \quad (2.12)$$

$$\pi_i(b_i|\tilde{\lambda}_b) := \mathcal{N}(b_i|\mu_b, \sigma_b^2) \quad (2.13)$$

donde $\mathcal{N}(x|\mu, \sigma^2)$ es la densidad normal con media μ y varianza σ^2 . Más específicamente $\tilde{\lambda}_\theta = (\mu_\theta, \sigma_\theta^2)$, $\tilde{\lambda}_a = (\mu_a, \sigma_a^2)$ y $\tilde{\lambda}_b = (\mu_b, \sigma_b^2)$.

Los hiperparámetros a priori constantes para este estudio se eligen según los valores presentados en Sung y Kang, 2006

$$\theta_s \stackrel{iid}{\sim} \mathcal{N}(0, 1) \quad (2.14)$$

$$a_i \stackrel{iid}{\sim} \mathcal{N}(0, 1) \quad (2.15)$$

$$b_{iv} \stackrel{iid}{\sim} \mathcal{N}(0, 1). \quad (2.16)$$

La factorización mostrada en la ecuación (2.10) hace que los condicionales completos sean simples, en particular, la distribución condicional completa para un parámetro (digamos θ) es su función de verosimilitud $p(\mathcal{X}|\theta)$, multiplicada por su distribución a priori $\pi(\theta)$. En consecuencia, las densidades condicionales completas para los parámetros individuales serán

$$\pi(\theta_s|rest) \propto \prod_{i=1}^I p(x_{s_i}|\theta_s, a_i, b_i) \pi_s(\theta_s|\tilde{\lambda}_\theta) \quad \forall s = 1, \dots, S \quad (2.17)$$

$$\pi(a_i|rest) \propto \prod_{s=1}^S p(x_{s_i}|\theta_s, a_i, b_i) \pi_i(a_i|\tilde{\lambda}_a) \quad \forall i = 1, \dots, I \quad (2.18)$$

$$\pi(b_i|rest) \propto \prod_{s=1}^S p(x_{s_i}|\theta_s, a_i, b_i) \pi_i(b_i|\tilde{\lambda}_b) \quad \forall i = 1, \dots, I. \quad (2.19)$$

¹En la estadística bayesiana, un hiperparámetro es un parámetro de una distribución a priori. El término se utiliza para distinguirlos de los parámetros del modelo del sistema subyacente que se analiza.

Las dificultades con el algoritmo M-H surgen cuando la dimensionalidad del espacio de parámetros se vuelve grande y/o la distribución posterior tiene una forma analítica bastante compleja. En estos casos, a menudo se requieren largos tiempos de simulación para explorar todas las regiones del espacio de parámetros, lo que puede resultar en valores bastante pobres. Además, la efectividad del algoritmo depende de un buen esquema de actualización y una adecuada elección de la densidad propuesta, lo que no siempre es una tarea fácil. Debido a estas limitaciones, surge otro método de muestreo que supera las técnicas de Gibbs y M-H, denominado Monte Carlo Hamiltoniano.

2.2. Monte Carlo Hamiltoniano

Desde su introducción por Duane y col., 1987 y su desarrollo por Neal, 1995, el método Monte Carlo Hamiltoniano (HMC) denominado también Monte Carlo Híbrido, ha demostrado ser una poderosa metodología MCMC.

Al utilizar la dinámica hamiltoniana como un núcleo de transición de Markov, el HMC explora la distribución objetiva de manera más eficiente, lo que conduce a una convergencia más rápida. Esta dinámica depende de una función hamiltoniana concreta; una técnica general de la física teórica en la que la energía total de un sistema de partículas se describe mediante una función hamiltoniana (Betancourt y Stein, 2011).

Por lo cual, antes de hablar del algoritmo HMC, debemos familiarizarnos con la dinámica mencionada.

Dinámica hamiltoniana

La dinámica hamiltoniana es una de las formas en que los físicos describen cómo en un tiempo t en términos de su ubicación \boldsymbol{x} y su momento $\boldsymbol{\rho}$ (equivalente a la masa del objeto por su velocidad) se mueven los objetos en un sistema. Para cada ubicación que toma el objeto, hay una energía potencial asociada $U(\boldsymbol{x})$ y para cada momento una energía cinética $K(\boldsymbol{\rho})$.

A fin de obtener una intuición de lo que estudia la dinámica mencionada, consideremos el siguiente ejemplo en dos dimensiones. Sea un disco sin fricción que se desliza sobre una superficie de altura variable. El estado de este sistema consiste en la posición del disco, dado por el vector \boldsymbol{x} , y el momento del disco dado por el vector $\boldsymbol{\rho}$. La energía potencial del disco es proporcional a la altura de la superficie en la posición actual, y la energía cinética es igual a $\frac{\boldsymbol{\rho}'\boldsymbol{\rho}}{2m}$, donde m es la masa del disco. En una parte llana de la superficie, este se moverá a una velocidad constante igual a $\boldsymbol{\rho}/m$. Sin embargo, si se encuentra con una pendiente

ascendente, su energía cinética empezará a disminuir y su energía potencial a aumentar hasta que la energía cinética (y, por tanto, ρ) se anule. En ese momento, el disco iniciará también un movimiento descendente con aumento de la energía cinética y disminución de la energía potencial.

Tras esta introducción, el primer paso es definir la función hamiltoniana en términos de la distribución de probabilidad de la que deseamos tomar muestras.

Definición 5 *La función hamiltoniana representa la energía total constante de un sistema y es igual a la suma de las energías potencial y cinética*

$$H(\mathbf{x}, \boldsymbol{\rho}) = U(\mathbf{x}) + K(\boldsymbol{\rho})$$

donde $U(\mathbf{x})$ se establece como menos el logaritmo de la densidad de probabilidad de la distribución de \mathbf{x} que deseamos muestrear y $K(\boldsymbol{\rho})$ como

$$K(\boldsymbol{\rho}) = \frac{\boldsymbol{\rho}^T M^{-1} \boldsymbol{\rho}}{2}$$

con M una matriz de masa simétrica y definida positiva que suele ser diagonal y a menudo es múltiplo escalar de la matriz identidad (Neal, 2011).

Una vez definida la función hamiltoniana o el *Hamiltoniano*, es necesario determinar cómo se mueven \mathbf{x} y $\boldsymbol{\rho}$ durante el tiempo t . Esta descripción se logra cuantitativamente a través de un conjunto de ecuaciones diferenciales denominadas *ecuaciones hamiltonianas*

$$\frac{\partial \mathbf{x}}{\partial t} = \frac{\partial H}{\partial \boldsymbol{\rho}} = \frac{\partial K(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}} \quad (2.20)$$

$$\frac{\partial \boldsymbol{\rho}}{\partial t} = -\frac{\partial H}{\partial \mathbf{x}} = -\frac{\partial K(\boldsymbol{\rho})}{\partial \mathbf{x}} - \frac{\partial U(\mathbf{x})}{\partial \mathbf{x}}. \quad (2.21)$$

Para cualquier intervalo de tiempo de duración s , estas ecuaciones definen un mapeo T_s , del estado en cualquier momento t al estado en el tiempo $t + s$ (Neal, 2011).

Por otra parte, dado que la densidad del impulso es independiente de la densidad objetiva, $p(\boldsymbol{\rho}|\mathbf{x}) = p(\boldsymbol{\rho})$. El primer término de la derivada temporal se anula, $\frac{\partial K}{\partial \mathbf{x}} = 0$, en consecuencia

$$\frac{\partial \mathbf{x}}{\partial t} = +\frac{\partial K}{\partial \boldsymbol{\rho}} = [M^{-1}p] \quad (2.22)$$

$$\frac{\partial \boldsymbol{\rho}}{\partial t} = -\frac{\partial U}{\partial \mathbf{x}}. \quad (2.23)$$

Así que, si tenemos expresiones para $\frac{\partial U(\mathbf{x})}{\partial \mathbf{x}}$ y $\frac{\partial K(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}}$ y un conjunto de condiciones iniciales (es decir, una posición inicial \mathbf{x}_0 un momento inicial $\boldsymbol{\rho}_0$ el tiempo t_0) es posible predecir la posición y el momento de un objeto en cualquier punto del tiempo $t = t_0 + s$ simulando la dinámica para una duración s .

Para simular la dinámica hamiltoniana en un ordenador, las ecuaciones de Hamilton deben aproximarse discretizando el tiempo. Esto se hace dividiendo el intervalo s en intervalos más pequeños de longitud δ . Cuanto más pequeño sea el valor de δ , más cercana será la aproximación a la dinámica en tiempo continuo. Hay una serie de procedimientos que se han desarrollado para discretizar el tiempo, el más utilizado es el *método de Leapfrog* o método del salto.

Método Leapfrog

Este método es un algoritmo de integración numérica que se adapta para proporcionar resultados estables para los sistemas de ecuaciones hamiltonianos, actualizando las variables de momento y posición secuencialmente. Comienza por simular la dinámica del momento en un pequeño intervalo de tiempo $\delta/2$, luego simula la dinámica de la posición en un intervalo de tiempo un poco más largo δ , y luego completa la simulación del momento en otro pequeño intervalo de tiempo $\delta/2$ para que \mathbf{x} y $\boldsymbol{\rho}$ existan en el mismo punto en el tiempo. En concreto, el método Leapfrog hace lo siguiente:

1. Dar un medio paso en el tiempo para actualizar la variable de impulso

$$\rho_i(t + \delta/2) = \rho_i(t) - (\delta/2) \frac{\partial U}{\partial x_i(t)}.$$

2. Dar un paso completo en el tiempo para actualizar la variable de posición

$$x_i(t + \delta) = x_i(t) + \delta \frac{\partial K}{\partial \rho_i(t + \delta/2)}.$$

3. Tomar el medio paso restante en el tiempo para terminar de actualizar la variable de momento

$$\rho_i(t + \delta) = \rho_i(t + \delta/2) - (\delta/2) \frac{\partial U}{\partial x_i(t + \delta)}.$$

Observación 15 El método Leapfrog se ejecuta durante L pasos de salto simulando la dinámica un total de $L \times \delta$ unidades de tiempo. El estado resultante al final se denota $(\boldsymbol{\rho}^*, \mathbf{x}^*)$.

Ahora que tenemos una mejor comprensión de lo que es la dinámica hamiltoniana y de cómo se puede simular, vamos a discutir cómo podemos utilizarla para un análisis MCMC.

La idea principal detrás del HMC, es desarrollar una función hamiltoniana $H(\mathbf{x}, \boldsymbol{\rho})$ tal que la dinámica resultante nos permita explorar alguna distribución objetiva $p(\mathbf{x})$ de manera eficiente. ¿Cómo podemos elegir la función hamiltoniana? Utilizando un concepto de *distribución canónica* de la mecánica estadística que relaciona $H(\mathbf{x}, \boldsymbol{\rho})$ con $p(\mathbf{x})$.

Definición 6 *Dada una función de energía $E(\gamma)$, para el estado γ de un sistema físico. La distribución canónica sobre los estados tiene una función de densidad de probabilidad dada por*

$$p(\gamma) = \frac{1}{Z} \exp(-E(\gamma))$$

donde Z es la constante de normalización necesaria para que la función sume o se integre en uno, creando así una distribución de probabilidad válida.

Por tanto, la distribución canónica para la función de energía de la dinámica hamiltoniana viene dada por

$$\begin{aligned} p(\mathbf{x}, \boldsymbol{\rho}) &= \frac{1}{Z} \exp(-H(\mathbf{x}, \boldsymbol{\rho})) \\ &= \frac{1}{Z} \exp(-U(\mathbf{x})) \exp(-K(\boldsymbol{\rho})). \end{aligned}$$

Observación 16 *Vemos que \mathbf{x} y $\boldsymbol{\rho}$ son independientes y cada una tiene distribuciones canónicas con funciones de energía $U(\mathbf{x})$ y $K(\boldsymbol{\rho})$.*

Criterio de aceptación

Partiendo de un estado inicial $[x_0, p_0]$, el estado propuesto se acepta utilizando un paso de aceptación análogo al criterio de aceptación de Metrópolis; en el cual, si la probabilidad de mantener la propuesta $(\mathbf{x}^*, \boldsymbol{\rho}^*)$ después de la dinámica hamiltoniana

$$p(\mathbf{x}^*, \boldsymbol{\rho}^*) \propto \exp-[U(\mathbf{x}^*) + K(\boldsymbol{\rho}^*)]$$

es mayor que la probabilidad del estado anterior a la dinámica hamiltoniana

$$p(\mathbf{x}_0, \boldsymbol{\rho}_0) \propto \exp-[U(\mathbf{x}^{(t-1)}) + K(\boldsymbol{\rho}^{(t-1)})]$$

entonces el estado propuesto es aceptado. Si la propuesta no es aceptada, se devuelve el valor del parámetro anterior para el siguiente sorteo y se utiliza para inicializar la siguiente iteración. Por consiguiente, la probabilidad de mantener la propuesta $(\mathbf{x}^*, \boldsymbol{\rho}^*)$ generada por la transición de $(\mathbf{x}, \boldsymbol{\rho})$ es

$$\min[1, \exp(H(\mathbf{x}, \boldsymbol{\rho}) - H(\mathbf{x}^*, \boldsymbol{\rho}^*))].$$

La combinación de estos pasos, el muestreo del momento aleatorio, seguido de la dinámica hamiltoniana y el criterio de aceptación, define el algoritmo HMC para extraer N muestras de una distribución objetiva.

En resumen, el algoritmo comienza con un conjunto inicial de parámetros $\mathbf{x} \in \mathbb{R}^d$ distribuido según una densidad a posteriori $\pi(\mathbf{x})$ y una variable de momento inicial con la distribución canónica de momento $p(\boldsymbol{\rho})$. Luego, para un número determinado de iteraciones, se muestrea un nuevo vector de momento y se actualiza el valor actual del parámetro \mathbf{x} mediante el método Leapfrog con tiempo de discretización δ y número de pasos L . Por último, aplica un paso de aceptación y se decide si se actualiza al nuevo estado $(\mathbf{x}^*, \boldsymbol{\rho}^*)$ o se mantiene el estado existente.

Algoritmo Monte Carlo Hamiltoniano - HMC

```

1: Sea una posición inicial  $\mathbf{x}_{(0)} \sim \pi(\mathbf{x})$ 
2: Sea  $\boldsymbol{\rho} \sim p(\boldsymbol{\rho})$  y  $u \sim \mathcal{U}(0, 1)$ 
3: for  $i = 1 : N$  do
4:    $(\mathbf{x}_{(0)}^*, \boldsymbol{\rho}_{(0)}^*) = (\mathbf{x}_{(i-1)}, \boldsymbol{\rho})$ 
5:   for  $j = 1 : N$  do
6:     hacer un movimiento Leapfrog  $(\mathbf{x}_{(j-1)}^*, \boldsymbol{\rho}_{(j-1)}^*) \longrightarrow (\mathbf{x}_{(j)}^*, \boldsymbol{\rho}_{(j)}^*)$ 
7:   end for
8:    $(\mathbf{x}^*, \boldsymbol{\rho}^*) = (\mathbf{x}_{(N)}, \boldsymbol{\rho}_{(N)})$ 
9:   if  $u < \min\{1, e^{(H(\mathbf{x}, \boldsymbol{\rho}) - H(\mathbf{x}^*, \boldsymbol{\rho}^*))}\}$  then
10:     $\mathbf{x}_{(i)} = \mathbf{x}^*$ 
11:   else
12:     $\mathbf{x}_{(i)} = \mathbf{x}_{(i-1)}$ 
13:   end if
14: end for

```

Finalmente, dado que nuestro interés está en hallar la distribución posterior para los parámetros del modelo. En el HMC, estos parámetros se representan en la posición \mathbf{x} y se introduce $\boldsymbol{\rho}$ a fin de permitir que opere la dinámica hamiltoniana (Luo y Jiao, 2017).

2.3. Diagnóstico de Convergencia MCMC

No hay forma de garantizar cuándo el algoritmo MCMC convergerá a su distribución estacionaria, pero existen varios métodos heurísticos para evaluar la convergencia. Unos de los métodos son el gráfico de trazas (*trace plots*), el gráfico de autocorrelación (ACF) y el diagnóstico de Gelman-Rubin.

2.3.1. Gráficos de Traza

Un gráfico ideal se vería como un ruido blanco (sin patrón) centrado en la media posterior. En tipo de gráficos, existen cuatro problemas comunes:

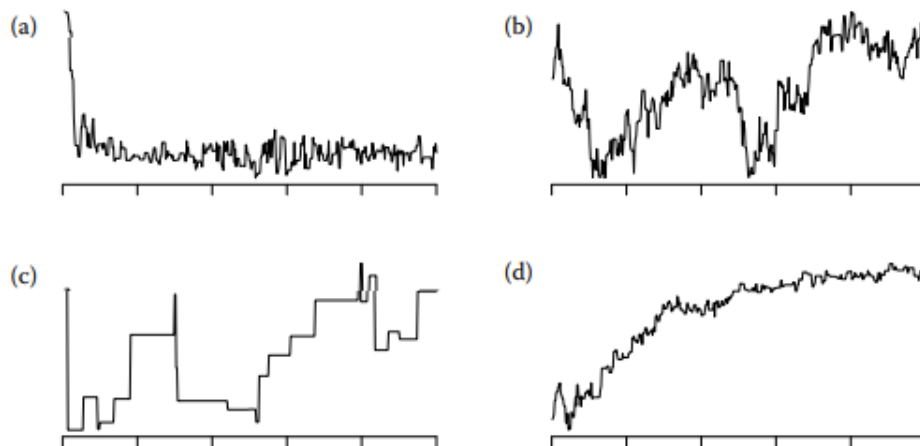


Figura 2.1: Problemas comunes en los gráficos de Trazas (a) Drift en el segmento inicial (burn-in) (b) Oscilación u otro patrón no constante (c) Cadena pegajosa y (d) Drift en toda la cadena (van der Linden, 2016).

a) Drift en el segmento inicial.

Sucede cuando la parte inicial de la cadena no está muy cerca de la distribución estacionaria. La solución habitual es descartar la parte inicial y usar la parte estable restante para la inferencia.

b) Oscilación u otro patrón no constante.

Sucede cuando existe una dependencia excesiva entre los pasos del algoritmo MCMC. La solución es rediseñar la cadena para reducir la dependencia.

c) Cadena pegajosa.

El gráfico permanece igual durante varios pasos antes de ir al nuevo valor. Suele ocurrir cuando la tasa de aceptación es demasiado baja. La solución es ajustar o cambiar la propuesta de densidad.

d) Drift en toda la cadena.

Puede ser un caso especial del problema (a) o (b) que se descubre ejecutando la cadena durante más tiempo para ver qué ocurre. Si la drift persiste incluso después de ejecutar la cadena durante mucho tiempo, puede ser evidencia de un problema con el modelo estadístico subyacente.

2.3.2. Gráficos de autocorrelación

Los gráficos de autocorrelación son útiles para evaluar la dependencia entre los pasos de la cadena. Sea $\tau^{(m)}$ el resultado del algoritmo MCMC y sean las funciones $g_1(\tau)$ y $g_2(\tau)$. La ACF se define para $g_1(\tau)$ de la siguiente manera

$$\rho_k^{g_1} = \frac{\text{cov}(g_1(\tau^{(m)}), g_1(\tau^{(m+k)}))}{\text{var}(g_1(\tau^{(m)}))} \quad (2.24)$$

y de forma similar para $g_2(\tau)$. Estas funciones estiman el movimiento $\hat{\rho}_k^{g_1}$, como una función de k y son útiles si k es pequeño en comparación con M .

Observación 17 $\hat{\rho}_0^{g_1}$ es forzado a ser 1 por lo que a medida que aumenta k el valor absoluto de $\hat{\rho}_k^{g_1}$ debería caer a 0.

Para ayudar a interpretar el gráfico es útil agregar una línea horizontal de correlación cero y también líneas para indicar el área de rechazo de un test de hipótesis $\rho_k^{g_1} = 0$.

De igual manera que hay problemas en los trace plots, el comportamiento problemático en una ACF, incluye

- a) $\hat{\rho}_k^{g_1}$ permanece distinto de cero para todos los valores observables de k .
- b) $\hat{\rho}_k^{g_1}$ oscila entre valores cercanos a cero y valores alejados de cero.

En la siguiente figura, ambos problemas (a) y (b) reflejan una autocorrelación excesiva en la cadena de Markov. Estos problemas, especialmente la autocorrelación positiva, están relacionados con los problemas (a), (b) o (c) de la figura 2.1.

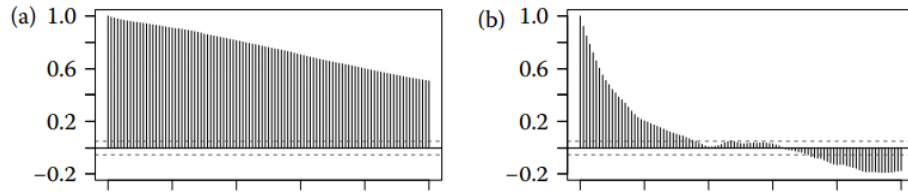


Figura 2.2: Problemas comunes en gráficos de autocorrelación: (a) Autocorrelación alta, (b) Autocorrelación oscilan entre valores cercanos y lejanos a 0 (van der Linden, 2016).

La solución es rediseñar la cadena para reducir esta dependencia.

Por tanto, si los gráficos de trazas se ven como ruido blanco centrado en la media posterior y el diagrama de autocorrelación cae rápidamente a 0 y se mantiene, se puede decir que la cadena ha alcanzado su distribución estacionaria.

2.3.3. Gelman Rubin

El diagnóstico de Gelman-Rubin (Gelman y Rubin, 1992) evalúa la convergencia MCMC analizando las diferencias entre varias cadenas de Markov. La convergencia se evalúa comparando las varianzas estimadas entre cadenas y dentro de la cadena para cada parámetro del modelo.

Suponga que hay N cadenas, todas de igual longitud M . El parámetro de interés del modelo es θ y $\{\theta_n^m\}_{m=1}^M$ es la n -ésima cadena simulada con $n = 1, \dots, N$.

Sean, además, la media de la muestra para la cadena $\hat{\theta}_n = \frac{1}{M} \sum_{m=1}^M \theta_n^m$ y la media posterior de la muestra *overall* $\hat{\theta} = \frac{1}{N} \sum_{n=1}^N \hat{\theta}_n$. La convergencia se evalúa comparando las siguientes varianzas para cada parámetro del modelo,

1. la varianza entre cadenas

$$B = \frac{M}{N-1} \sum_{n=1}^N (\hat{\theta}_n - \hat{\theta})^2$$

2. la varianza dentro de la cadena

$$W = \frac{1}{N} \sum_{n=1}^N \hat{\sigma}_n^2$$

donde $\hat{\sigma}_n^2 = \frac{1}{M-1} \sum_{m=1}^M (\theta_n^m - \hat{\theta}_n)^2$ es la varianza de la n -ésima cadena.

Grandes diferencias entre estas varianzas indican falta de convergencia (Brooks y Gelman, 1998). Por otro lado, si todas las N cadenas convergen a sus distribuciones objetivas, la estimación insesgada de la varianza posterior marginal de θ

$$\widehat{Var}(\theta) = \frac{M-1}{M} W + \frac{N+1}{NB} B$$

debería ser cercana a W .

Una forma “fácil” de determinar la convergencia o no del algoritmo MCMC usando los criterios de Gelman-Rubin, es calcular el *factor de reducción de escala potencial* (PSRF, por sus siglas en inglés de *Potential Scale Reduction Factor*)

$$\hat{R} = \sqrt{\frac{\hat{d} + 3}{\hat{d} + 1} \frac{\widehat{Var}(\theta)}{W}} \quad (2.25)$$

donde \hat{d} es la estimación de los grados de libertad de una *t-distribución*. En consecuencia,

- Si la medida de convergencia \hat{R} es grande, entonces $B > W$, lo que indica que las cadenas no han convergido.
- Si $\hat{R} \leq 1.1$, la convergencia se alcanza, si no, se itera hasta conseguir ese valor para todos los parámetros del modelo (Brooks y Gelman, 1998).

Capítulo 3

Innovación

Hoy en día el mundo busca recuperarse de la crisis de la pandemia, COVID-19. Siendo un desafío aún más para las economías emergentes poder mejorar sistemáticamente su desempeño y sus sistemas de innovación para estar a la altura de las economías más prósperas y de altos ingresos que se caracterizan por tener sistemas de innovación equilibrados y de alto rendimiento, altamente dinámicos, combinando la eficacia a la hora de traducir los recursos de innovación en resultados con desempeño equilibrado (Tang, 2021).

Dado que la innovación es un proceso difícil de contextualizar, la trataremos con base en lo que menciona Kahn, 2021. Para Kahn, la innovación debe conceptualizarse en tres puntos de vista al mismo tiempo: un resultado, un proceso y una mentalidad. En cuanto a los resultados, la innovación no se limita a simples modificaciones de las funciones de producción, como las invenciones, sino que también introduce novedades en el mercado, aumentando así la competitividad. Kahn también señala que permitir la innovación requiere apoyo cultural, así como, personas comprometidas con el pensamiento de diseño y otros rasgos de pensamiento empático asociados con la creatividad.

En un marco general, Mella y col., 2009, definen la innovación como el proceso de proponer, adoptar, desarrollar e implementar nuevas ideas, interna o externamente, relacionadas con productos, procesos, políticas, prácticas o comportamientos. Por tanto, la innovación se refiere a cualquier cambio basado en algún tipo de conocimiento, siempre y cuando genere valor y tenga consecuencias económicas directas.

3.1. Índice Mundial de Innovación

Anualmente, se mide la innovación a través del Índice Mundial de Innovación (GII, por su acrónimo en inglés de *Global Innovation Index*), que clasifica los resultados de 132 eco-

nomías de distintas regiones del mundo, sobre la base de dos subíndices: el subíndice de insumos de innovación y el subíndice de resultados de innovación (Dutta y col., 2021).

El primer subíndice se basa en cinco pilares: Instituciones, Capital humano e investigación, Infraestructura, Sofisticación del mercado y Sofisticación empresarial. El segundo subíndice se basa en dos pilares: Resultados del conocimiento y la tecnología, y Resultados creativos. El objetivo de este índice es proporcionar datos relevantes en materia de innovación y, a su vez, ayudar a los encargados de la formulación de políticas a comprender y cuantificar mejor el grado de innovación de sus actividades a fin de tomar decisiones en este ámbito.

La puntuación de este índice se asocia a la realidad que se vive al interior de cada economía. En el ranking los países desarrollados son los mayores puntuados, ya sea por el crecimiento de su economía, tecnología y/o desarrollo científico. Muestran en sí, la relación positiva que existe entre innovación y desarrollo económico.

En 2021, el Índice Mundial de Innovación, muestra que pese a la enorme repercusión que ha tenido la pandemia de COVID-19, muchos sectores han mostrado resiliencia. En particular, los que se han adherido a la digitalización, la tecnología y la innovación.

En materia de innovación es claro pensar que existe discrepancias con respecto a los resultados nacionales entre unas y otras regiones del mundo. América del Norte es la región más innovadora; obtiene los mejores resultados en todos los pilares del índice en comparación con las demás. En esa misma línea, Europa, a pesar de ser la segunda región más innovadora del mundo, en desarrollo de mercado no lo es, es más, se encuentra por detrás de la media de las regiones del Sudeste Asiático, Asia Oriental y Oceanía. A mayor distancia, en materia de innovación, prosigue las regiones del Norte de África y Asia Occidental, luego América Latina y el Caribe, y por último las regiones de Asia Central y Meridional y el África Subsahariana (Tang, 2021).

En América Latina y el Caribe, ninguna economía se encuentra entre las 50 principales y con la excepción de México, pocas economías de la región han logrado ascender en la clasificación de manera constante durante la última década.

La Tabla 3.1 muestra los países más representativos por su índice de innovación; en sus respectivas regiones.

Innovación en Ecuador

Según el GII; el Ecuador ocupa el puesto 91 en el ranking de innovación global de 132 países, con un puntaje de 25.4 puntos (sobre 100). A nivel regional, el Ecuador se ubica en

Región	País	Ranking	Índice Mundial de Innovación
Europa	Suiza	1	65,5
	Suecia	2	63,1
	Reino Unido	4	59,8
America del Norte	EE.UU.	3	61,3
	Canadá	16	53,1
América Latina y el Caribe	Chile	53	35,1
	México	55	34,5
	Costa Rica	56	34,5
Asia Central y Asia Meridional	India	46	36,4
	Irán (República Islámica)	60	32,9
	Kazajstán	79	28,6
Sudeste Asiático, Asia Oriental, y Oceanía	República de Corea	5	59,3
	Singapur	8	57,8
	China	12	54,8
África del Norte y Asia Occidental	Israel	15	53,4
	Emiratos Árabes Unidos	33	43,0
	Turquía	41	38,3
África Subsahariana	Sudáfrica	61	32,7
	Kenya	85	27,5
	República Unida de Tanzania	90	25,6

Tabla 3.1: Principales economías innovadoras por región (Dutta y col., 2021)

el puesto 12 en América Latina y el Caribe.

En cuanto al desempeño en las dimensiones clave de la innovación, nuestro país obtuvo los siguientes resultados: Entorno institucional: 44.1 puntos (puesto 126 a nivel mundial), Capital humano e investigación: 20.5 puntos (puesto 97 a nivel mundial), Infraestructura: 39.6 puntos (puesto 74 a nivel mundial), Sofisticación del mercado: 50.3 puntos (puesto 44 a nivel mundial), Sofisticación de negocios: 19.9 puntos (puesto 97 a nivel mundial), Productos de conocimiento y tecnología: 13.2 puntos (puesto 97 a nivel mundial) y Salidas creativas: 18.5 puntos (puesto 86 a nivel mundial).

Esto significa que el país está muy por debajo de la media mundial en términos de innovación.¹ Esto es preocupante, ya que muestra que el país no está invirtiendo lo suficiente en áreas elementales que fomenten innovación.

¹En el período 2011-2020, la puntuación más alta de innovación global de Ecuador fue de 32,83 en 2013.

3.2. Innovación Organizacional

Para poder generar redes de innovación, la organización debe realizar al menos algún acto de innovación, como los que se menciona en la tabla 3.2. Estas actividades dependen, en gran parte, de la variedad y estructura del sistema de innovación de la empresa, y se definen como todas las tareas científicas, técnicas, organizativas, financieras y comerciales que conducen real o potencialmente a la implementación de la innovación (OECD y EUROSTAT, 2005).

Con ello queda claro que la innovación no es una simple modificación hecha a voluntad por la organización (desuso, sustitución, producción a medida, etc.), sino que actividades como las mencionadas inician el desarrollo en una empresa.

Proceso de Innovación

El proceso de innovación se define como el conjunto de las etapas técnicas, industriales y comerciales que conducen al lanzamiento con éxito en el mercado de nuevos productos y servicios o a la utilización comercial de nuevos procesos técnicos (Hidalgo y col., 2013).

La utilidad e importancia de medir estos procesos está relacionada con el vínculo entre la innovación, los aumentos genuinos de la competitividad, el crecimiento económico y los niveles de bienestar social (Lugones, 2009). Estas medidas requieren un enfoque multidisciplinario para obtener datos que se utilizan en la toma de decisiones, la formulación de políticas públicas y en la definición estratégica que fomenta la innovación en organizaciones públicas y privadas (Tang, 2021).

3.3. Instrumento de Recolección de Datos

Este trabajo de investigación utiliza el cuestionario como instrumento de recolección de datos. Su desarrollo se sustenta en el proyecto PIS 19-05,² el cual valida el cuestionario CRI2,³ que plantea la sugerencia metodológica de Morales y col., 2019 proponiendo medir la innovación desde un constructo principal denominado *potencial de innovación* en función de tres categorías o constructos secundarios llamados *Capacidades, Resultados e Impactos* de innovación.

El acrónimo de dichas categorías se resume en *CRI* y su enfoque es el siguiente:

²Reporte técnico de validación del cuestionario de Capacidades Resultados e Impactos CRI2 para medir el Potencial de Innovación de las organizaciones (Morales y col., 2022).

³La primera versión del cuestionario CRI fue desarrollado por Robalino y col., 2017b y proponía analizar el nivel de innovación de las organizaciones a partir de sus Capacidades, Resultados e Impactos.

Investigación y desarrollo experimental

Incluye el trabajo creativo/sistemático destinado a aumentar el conocimiento existente y utilizarlo para encontrar nuevas aplicaciones.	
I+D intramuros	Intramuros I+D incluye toda la I+D realizada dentro de la empresa. También envuelve adquisiciones de bienes de capital directamente relacionados con la I+D.
I+D extramuros	La I+D extramuros incluye la adquisición de servicios de I+D de unidades extranjeras de corporaciones multinacionales.

Actividades para innovaciones de producto y de proceso

Adquisición de conocimiento externo.	Se pueden adquirir tecnología y conocimiento de diferentes maneras y de distintas fuentes para desarrollar e introducir innovaciones. Esta adquisición puede incluir: servicios informáticos, servicios científicos y técnicos para actividades de productos y procesos innovadores.
Adquisición de maquinaria, equipo y otros bienes de capital.	Incluye las adquisiciones de bienes de capital para la innovación que no se incluyen en I+D (adquisición de terrenos y edificios, adquisición de maquinaria, instrumentos y equipos, y la adquisición de software de computadoras).
Otros preparativos para innovaciones de producto y de proceso.	Incluyen actividades intramuros relacionadas con el desarrollo e implementación de innovaciones de productos y procesos no consideradas en I+D (diseño industrial, ingeniería y montaje, pruebas de producción, gestiones de patentes y licencias, producción inicial y pruebas.)
Preparativos comerciales para innovaciones de producto.	Incluyen la investigación y pruebas de mercado, Además, la publicidad de bienes y servicios nuevos y mejorados.
Formación	Actividades innovadoras de productos o procesos, necesarias para implementar la innovación.

Actividades para innovaciones comerciales y organizativas

Preparativos para las innovaciones comerciales	Actividades de desarrollo y nuevos sistemas de comercialización no empleados antes en la empresa.
Preparativos para innovaciones organizativas	Incluyen el desarrollo y planificación de nuevos sistemas organizativos y su introducción.

Tabla 3.2: Actividades de Innovación (OECD y EUROSTAT, 2005)

- a) Capacidades de Innovación: Son el resultado de un proceso de aprendizaje consciente e intencional, costoso en términos de tiempo y de recursos, y no lineal, ya que depende del pasado de la empresa y es acumulativo (OECD y EUROSTAT, 2005).
- b) Resultados de Innovación: Son el tipo de innovaciones que produce una organización durante un determinado periodo de tiempo (OECD y EUROSTAT, 2005). Están asociadas a los objetivos de la innovación y suelen reflejarse en los efectos de las actividades de innovación.
- c) Impactos de Innovación: Son aquellos que muestran en un periodo de tiempo determinado que efectivamente se está teniendo éxito en los objetivos organizacionales establecidos, a través de las medidas de desempeño detectadas (OECD y EUROSTAT, 2005). Dicho esto, este constructo es el componente que mejor muestra que una organización está desarrollando e implementando con éxito la innovación.

Estos constructos están directamente relacionados, ya que sin capacidad, que es la fuente de la innovación, es imposible generar resultados de innovación (cambios en los productos, procesos, marketing u organización de una empresa) y por ende es complicado medir la innovación en impacto o efecto durante un período de tiempo.

El número de factores incluidos en el cuestionario CRI2 empieza con cuatro subcategorías pertenecientes al primer constructo C que resume las *Fuentes de información y financiamiento* de la innovación, además de las *Actividades y Objetivos* necesarios para implementar procesos en la organización. Luego continúa con una subcategoría del constructo R llamada *Resultados* y una subcategoría más del constructo I llamada *Impactos*. De modo que el modelamiento CRI contempla cuatro niveles de abstracción que se detallan en la siguiente tabla.

3.3.1. Cuestionario CRI2

El cuestionario CRI2 es de clase estructurada y consta de dos tipos de preguntas, el primero recopila información general de las organizaciones y del personal que completan el cuestionario; suelen ser textos libres o categorías. El segundo recopila datos sobre los factores que componen cada constructo secundario CRI con preguntas de elección (ítems) valoradas en una puntuación politómica de escala Likert que presenta al sujeto 4 respuestas posibles $k = 1, 2, 3, 4$ para cada ítem, donde 4 representa *Totalmente en acuerdo*; 3 representa *En acuerdo*; 2 representa *En desacuerdo* y 1 representa *Totalmente en desacuerdo*.

El segundo tipo de preguntas es el que vamos a estudiar en el modelo GRM y consta de 46 ítems o variables repartidas en 6 secciones: i) fuentes de información de la innovación (8

Nivel de Abstracción	Componente	Explicación del componente
Nivel 1	Potencial de Innovación	Medición del grado o nivel de innovación de una organización.
Nivel 2	Capacidades, Resultados, Impactos	Categorías claves del modelo CRI para la medición de la innovación.
Nivel 3	subcategorías	Cada categoría se divide en varios factores o componentes principales interrelacionados.
Nivel 4	ítem de cada factor	Son las preguntas del instrumento construido para validar el modelo CRI.

Tabla 3.3: Niveles de abstracción del modelo CRI.

ítems), consiste en las ideas y datos obtenidos por la organización de forma externa o interna para generar innovación; ii) fuentes de financiamiento de la innovación (5 ítems), consiste en el inicio de la financiación de los gastos de innovación; iii) actividades de innovación (9 ítems), consiste en el campo científico, tecnológico, organizativo, financiero y comercial para la iniciación de la innovación; iv) objetivos de innovación (9 ítems), consiste en los objetivos propuestos por la organización para generar innovación; v) resultados de innovación (6 ítems), consiste en la generación de los diversos tipos de innovación (productos, procesos, métodos comerciales o métodos organizacionales) a través de las actividades de innovación promovidas por la organización. Finalmente; vi) impactos de innovación (9 ítems), consiste en los cambios económicos y comerciales de la organización debido a la aplicación de la innovación (Robalino y col., 2017a).

Esto quiere decir, que en total la categoría C tiene 31 ítems, la categoría R tiene 6 ítems y la categoría I tiene 9 ítems. La Tabla 3.4 detalla las secciones mencionadas.

Categoría	Subcategoría	Descripción	Ítems
Capacidades			C
	Fuentes de información para la Innovación		C1
		Departamento o unidad interna de I+D o de Diseño.	C101
		Departamento o unidad de Ventas y Marketing.	C102
		Directivos de la Organización.	C103
		Casa Matriz (del país de origen).	C104
		Clientes.	C105
		Competidores.	C106

		Departamentos de producción, Logística, Distribución o similares.	C107
		Entidades externas (Consultores, Universidades u otros Proveedores de conocimiento).	C108
	Fuentes de financiamiento de la Innovación.		C2
		Recursos Propios.	C201
		Recurso de la Casa Matriz (del país de origen).	C202
		Banca Comercial.	C203
		Grupo Empresarial al que pertenece la organización.	C204
		Recursos Gubernamentales o Mixtos.	C205
	Actividades de Innovación.		C3
		Actividades de I+D con su respectiva asignación de recursos (personal, equipos, insumos).	C301
		Inversión en licencias o acuerdos de propiedad intelectual (patentes, marcas, etc.).	C302
		Actividades de diseño u otras actividades creativas.	C303
		Implementación de programas de modernización y gestión de procesos de producción.	C304
		Implementación de programas de capacitación orientados a la innovación y mejoramiento de procesos productivos.	C305
		Diseño del portafolio de negocio y/o de procesos.	C306
		Implementación de nuevas formas de distribución y mercadeo.	C307
		Comercialización de productos innovados.	C308
		Implementaciones orientadas a la transformación digital y el teletrabajo.	C309
	Objetivos de la Innovación.		C4
		Ampliar el mercado actual.	C401
		Reducir costos laborales unitarios, de consumo de materiales y/o consumo de energía.	C402
		Mejorar las condiciones de trabajo.	C403
		Flexibilizar la producción y/o desarrollar soluciones para clientes específicos.	C404

		Reducir los tiempos muertos. Aprovechar los nuevos conocimientos científicos - tecnológicos. Aprovechar los nuevos materiales o insumos existentes. Aprovechar la capacidad organizacional de producción. Mejorar la transformación digital de la organización.	C405 C406 C407 C408 C409
Resultados			R
	Resultados de la Innovación		R1
		Ha introducido al mercado productos nuevos o significativamente mejorados.	R101
		Ha introducido al mercado procesos nuevos o significativamente mejorados.	R102
		Ha introducido métodos organizacionales nuevos o significativamente mejorados.	R103
		Ha introducido métodos organizacionales de toma de decisiones nuevos o significativamente mejorados.	R104
		Ha introducido métodos/modelos /prácticas comerciales nuevos o significativamente mejorados.	R105
		Ha introducido métodos de distribución o colocación de productos en el mercado nuevos o significativamente mejorados.	R106
Impactos			I
	Impactos de la Innovación		I1
		Impacto positivo en rentabilidad.	I101
		Impacto positivo en la utilidad bruta, utilidad operacional y/o utilidad antes de impuestos.	I102
		Impacto positivo en la competitividad.	I103
		Impacto positivo en la productividad.	I104
		Impacto positivo en las relaciones laborales.	I105
		Se han incrementado las ventas.	I106
		Se han reducido los costos de producción.	I107
		Se ha reducido el impacto ambiental negativo.	I108
		Se ha mejorado la salud, calidad de vida y bienestar en el entorno de la organización.	I109

Tabla 3.4: Descripción de las categorías, factores e ítems componentes del Modelo CRI

3.3.2. Validación

Dado que el cuestionario CRI2 es un instrumento de medición, este necesita ser validado. Morales y col., 2022 lo hacen a través de dos categorías, la validación externa⁴ y la validación interna de contenido⁵.

En la validación externa, siguen la metodología empleada por ellos en 2019 (Morales y col., 2019), en la cual establecen un modelo de macromedición que relaciona los constructos CRI y conduce a la identificación de variables que son medidas por el cuestionario CRI2. Por su parte, en la validación interna, los métodos de Lawshe, 2006 y Tristán-López, 2008 son la base para validar el contenido del cuestionario. Estos métodos obtienen una Razón de Validez de Contenido (CVR) que indica si un ítem es aceptable o no.

El método de Tristán-López es una modificación del modelo CVR de Lawshe. Este modelo, toma en cuenta el valor mínimo de ítems aceptables y consigue una nueva razón de validez de contenido denominado CVR'

$$CVR' = \frac{CVR + 1}{2} \quad (3.1)$$

donde $CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}}$ con n_e el número de expertos que consideran el ítem y N el número total de expertos.

Un ítem se considera aceptable si el CRV' no es menor a 0.5 o si el CRV no es menor a 0.2. Al aplicar la metodología de Tristán-López en los ítems del cuestionario CRI2, Morales, Robalino y Almeida, obtuvieron los siguientes resultados: Los ítems que conforman el constructo C tienen aceptabilidad, ya que todos los valores de CRV son superiores a 0.2 y los valores de CVR' son superiores a 0.5, en cuanto a los ítems de los constructos R e I, también tuvieron aceptabilidad con valores CRV y CRV' iguales a 1.

3.4. Muestra

El cuestionario CRI2 fue aplicado a una muestra heterogénea de 350 organizaciones de diferentes sectores productivos a nivel nacional, entre ellas, empresas de transporte, comercializadoras, fábricas de plástico, productos lácteos, etc. De esta forma, las respuestas obtenidas nos permiten realizar un análisis multisectorial. La toma de la muestra fue efectuada en el último trimestre del año 2021.

⁴La validación externa emplea métodos o procedimientos para generalizar los resultados del instrumento a situaciones similares, puede ser: poblaciones, situaciones o entornos parecidos (Ursachi y col., 2015).

⁵La validación interna verifica si el contenido de un ítem está relacionado con la variable a medir (Kempf, 2005).

Considerando que una respuesta sesgada da lugar a una diferencia en la interpretación de su puntuación y muchas respuestas falsas afectan todo el modelo, se descartaron 7 muestras defectuosas del conjunto de datos, aproximadamente el 2 % del número total de muestra.

En conclusión, dado que Morales y col., 2022, adecuan el contenido del cuestionario CRI2, valorando los ítems y su pertinencia en el modelamiento de los constructos CRI que miden el potencial de innovación de la organización. Esto nos garantiza un punto de partida fiable para nuestro análisis.

Capítulo 4

Simulación y Resultados

4.1. Hipótesis

Para determinar una medición válida del potencial de innovación de una organización, existen hipótesis del concepto “Innovación” e hipótesis de la IRT que forman una parte fundamental en el análisis de esta investigación.

Las hipótesis de la IRT aluden los supuestos fundamentales que son: la independencia experimental, la independencia local y la invarianza de los parámetros, para obtener una medida válida del rasgo latente (véase más información en 1.2).

En cuanto a las hipótesis de concepto, surgen con base en el modelamiento de los constructos CRI; donde se agrupan 6 secciones diferentes y cada una estima una habilidad diferente. Las 6 secciones son: 1) Fuentes de información para la Innovación, 2) Fuentes de financiamiento de la Innovación, 3) Actividades de Innovación, 4) Objetivos de la Innovación, 5) Resultados de la Innovación y 6) Impactos de la Innovación; el número de ítems de estas secciones conforman respectivamente el conjunto $I' \in \{8, 5, 9, 9, 6, 9\}$ como se observa en la tabla 3.4. Esto significa que el primer subconjunto (8 ítems) junto con el análisis GRM estiman la habilidad de *Fuentes de información* para el potencial de innovación de una organización. Lo mismo prosigue con los siguientes subconjuntos hasta tener los valores de las 6 habilidades.

Otra hipótesis a tomar en cuenta, es suponer que los sujetos responden a las preguntas con sinceridad e imparcialidad. Sin embargo, dado que el cuestionario se construye como una “autoevaluación” no se garantiza que siempre sea así.

4.2. MCMC en el caso de estudio

En breves rasgos, tenemos que el conjunto de datos consta de 343 entradas en las que cada sujeto $s = 1, \dots, 343$ ha respondido a los ítems $i = 1, \dots, 46$ del cuestionario CRI2. Sabemos que para cada ítem la respuesta se da en una escala de Likert con 4 niveles; siendo $\{1, 2, 3, 4\}$ el conjunto de números enteros con las únicas y posibles respuestas.

Por otro lado, se tiene la matriz de respuesta al ítem, $\mathcal{X} \in \mathbb{N}^{343,46}$, dada por

$$\mathcal{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,46} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,46} \\ \vdots & \vdots & \ddots & \vdots \\ x_{343,1} & x_{343,2} & \cdots & x_{343,46} \end{bmatrix}.$$

Dado que la derivación del método MCMC para los modelos IRT comienza especificando la función de verosimilitud conjunta, la ecuación 2.9 vendría a estar dada de la siguiente manera

$$p(\mathcal{X} | \theta_s, b, a) = \prod_{s=1}^{343} \prod_{i=1}^{I'} p(x_{s_i} | \theta_s, b_i, a_i) \quad (4.1)$$

donde I' es un subconjunto del conjunto de ítems I . De esta manera, las densidades condicionales completas se expresan como:

$$\begin{aligned} \pi(\theta_s | rest) &\propto \prod_{i=1}^{I'} p(x_{s_i} | \theta_s, a_i, b_i) \pi_s(\theta_s | \tilde{\lambda}_\theta) & \forall s = 1, \dots, 343 \\ \pi(a_i | rest) &\propto \prod_{s=1}^{343} p(x_{s_i} | \theta_s, a_i, b_i) \pi_i(a_i | \tilde{\lambda}_a) & \forall i = 1, \dots, I' \\ \pi(b_i | rest) &\propto \prod_{s=1}^{343} p(x_{s_i} | \theta_s, a_i, b_i) \pi_i(b_i | \tilde{\lambda}_b) & \forall i = 1, \dots, I'. \end{aligned}$$

Por último, las CCCs se determinan usando el sistema de ecuaciones 1.5 tomando $m = 4$ y la información del test se precisa usando $P_{i_0}^* = 1$ y $P_{i_4}^* = 0$ pues $k \in [1, 4]$.

4.3. Simulación

Para implementar el modelo GRM se utiliza el software de programación bayesiano Stan (Carpenter y col., 2017). Este programa de código abierto implementa el método Monte Carlo Hamiltoniano para generar simulaciones de distribuciones posteriores MCMC.

Stan es una plataforma de vanguardia para el modelado estadístico. Interactúa con los lenguajes de análisis de datos más populares como RStudio a través del paquete `rstan`. La función principal del paquete `rstan` es `stan()` que llama al programa Stan para estimar un modelo estadístico específico.

La estimación de nuestro interés consta de los siguientes pasos. Primero, se toma la data `Data1_Items.xlsx` concerniente al primer subconjunto de ítems $I' = 8$ (Fuentes de Información) y se comprueba que se hayan utilizado todas las categorías para cada ítem.

```
library(openxlsx)
data<-read.xlsx("Data1_Items.xlsx")
apply(data,2,table)
```

Luego, se carga el paquete `rstan` y se especifica que al ejecutarlo se almacene en el disco duro para que no se vuelva a compilar. Además, se solicita que Stan se realice en paralelo, es decir, en un ordenador multinúcleo local para evitar que la computadora se cuelgue.

```
library(rstan)
options(mc.cores = parallel::detectCores())
rstan_options(auto_write = TRUE)
```

Seguido de eso, se configura las dimensiones a utilizar en el modelo, indicando el número de sujetos que rinden el test, el número de ítems del conjunto de datos y la longitud del vector que contiene las categorías de respuesta existentes en la data.

```
J=dim(data)[1] #numero de sujetos
I=dim(data)[2] #numero de items
K=length(c(1,2,3,4))
```

Con la información extraída, se crea una lista de datos que incluye el número de sujetos, el número de ítems, la matriz de respuestas y la longitud del vector de categorías.

```
data_irt=list(n_student=J,n_item=I,Y=data, K=K)
```

Observación 18 *El nombre asignado en la lista de datos debe coincidir con lo que se proporciona en el bloque de datos del programa Stan.*

El algoritmo de código Stan consta de tres componentes necesarios. El primero de ellos, es el bloque de datos; en el que se distingue la información relevante de los datos.

```
data {
  int <lower=1, upper=4> K;
  int <lower=0> n_student;
  int <lower=0> n_item;
```

```

int<lower=1,upper=K> Y[n_student , n_item ];
}

```

Para el modelo politómico GRM, se especifica el número de categorías k para cada ítem, el número de sujetos y de ítems como enteros (int), y la matriz de respuestas como una matriz de enteros, seguida de su dimensión.

El segundo componente es el bloque de parámetros en el que se detalla los parámetros del modelo GRM (lado derecho de la ecuación 1.2) y los hiperparámetros correspondientes.

```

parameters {
  vector[n_student] theta ;
  real<lower=0> a [n_item ];
  ordered[K-1] b[n_item ];
}

```

Es necesario especificar el tipo de datos de todos los parámetros del primer bloque y si uno de ellos se define como un vector o una matriz, también es necesario fijar su dimensión. En este caso, el parámetro de capacidad θ se especifica como un vector de longitud igual al número de sujetos, así mismo, el parámetro de discriminación a se especifica como un real de longitud igual al número de ítems, y dado que el modelo GRM plantea el requisito de que las respuestas estén ordenadas según su nivel de dificultad, se especifica el orden para parámetro b .

El tercer componente es el bloque del modelo; en él se precisan los factores a priori de todos los parámetros del primer bloque y el modelo en sí. Este bloque es el componente más esencial de un programa Stan.

```

model{
  a ~ normal(0,1);
  theta ~ normal(0,1);
  for (i in 1:n_item){
    for (k in 1:(K-1)){
      b[i,k] ~ normal(0,1);
    }
  }
  for (i in 1:n_student){
    for (j in 1:n_item){
      Y[i,j] ~ ordered_logistic(theta[i]*a[j],b[j]);
    }
  }
}

```

Para la elección de la distribución a priori de los parámetros θ , a y b del modelo GRM se utiliza la distribución normal estándar. Luego de ello, el modelo se especifica en la declaración de muestreo a través de una función *ordered_logistic*.

La función *ordered.logistic* en el GRM, combinada con el parámetro de dificultad b especificado como el tipo de dato ordenado en el bloque de parámetros, conduce a una ecuación matemática que es idéntica a la ecuación 1.2, con el parámetro de dificultad de la categoría b reemplazado por $b[j]$.

Finalmente, se ejecuta el programa Stan. La función `stan()` requiere de cuatro argumentos. El primero es el archivo de programa Stan previamente guardado. El segundo es la lista de datos creada anteriormente. El tercer y cuarto, es el número de cadenas y el número de iteraciones en cada cadena, respectivamente. Para este modelo, se ejecutaron tres cadenas con 2000 iteraciones cada una.

```
GRM <- stan(model_code = stanmodelcode, data = data_irt, iter = 6000,
           chains = 3)
```

Observación 19 Como el número de iteraciones de burn-in no se especifica a través del argumento *warmup*, por defecto las primeras 1000 iteraciones se descartan automáticamente como iteraciones de burn-in.

Ejecutar `rstan` con la función `stan()` produce un objeto de clase `S4`. Este objeto contiene información sobre la distribución posterior de los parámetros del modelo. En nuestro caso, guardamos el objeto de ajuste del modelo GRM como `GRM`. La impresión directa de `GRM` en la consola de R proporciona información resumida sobre las distribuciones posteriores de todos los parámetros. Por ejemplo, la Figura 4.1 es una captura de pantalla de la consola de R después de ejecutar la siguiente línea de código:

```
print(GRM, par='a')
```

```
Inference for Stan model: 73a00a5d745e46857c5d614ca064da09.
3 chains, each with iter=6000; warmup=3000; thin=1;
post-warmup draws per chain=3000, total post-warmup draws=9000.

   mean se_mean   sd 2.5% 25% 50% 75% 97.5% n_eff Rhat
a[1]  1.47     0 0.22  1.07  1.31  1.45  1.61   1.94  4315   1
a[2]  1.75     0 0.26  1.29  1.57  1.74  1.91   2.29  3390   1
a[3]  1.24     0 0.20  0.88  1.11  1.24  1.37   1.64  4121   1
a[4]  1.79     0 0.25  1.33  1.62  1.77  1.94   2.33  3946   1
a[5]  1.08     0 0.19  0.73  0.95  1.06  1.19   1.47  3672   1
a[6]  1.19     0 0.19  0.86  1.06  1.19  1.31   1.58  3680   1
a[7]  1.12     0 0.17  0.78  1.00  1.11  1.23   1.47  5303   1
a[8]  0.88     0 0.16  0.57  0.76  0.87  0.98   1.20  5650   1

Samples were drawn using NUTS(diag_e) at Sun Jun 19 12:45:59 2022.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

Figura 4.1: Estimación del parámetro de discriminación s del ítem en el modelo GRM.

En ella podemos ver, la media posterior, la desviación estándar y el error estándar que pueden utilizarse como estimaciones puntuales de los parámetros de discriminación. Se muestra, también, los percentiles 2.5 al 97.5, y la columna `n_eff` que representa el número efectivo de extracciones de la simulación y puede considerarse como la muestra efectiva para una distribución posterior en la que se basan las inferencias. Por último, se tiene la columna `Rhat` que representa el índice del diagnóstico de convergencia de Gelman-Rubin.

4.4. Resultados

Todos los parámetros de los ítems de cada subcategoría o subrasgo del caso de estudio, se presentan en la tabla A.1 del Apéndice A.2 junto con las gráficas de convergencia, Apéndice A.3, las curvas características de los ítems (ICC), Apéndice A.4, y las funciones de información de los ítems (IIC), Apéndice A.5.

Con el fin de tener una visión general de la modelización resultante del MCMC, se eligió el subrasgo *Resultados* porque contiene ítems con alta y baja información de ítems (Figura 4.2) y, por tanto, presenta los CCCs que mejor representan los resultados típicos (Figura 4.3).

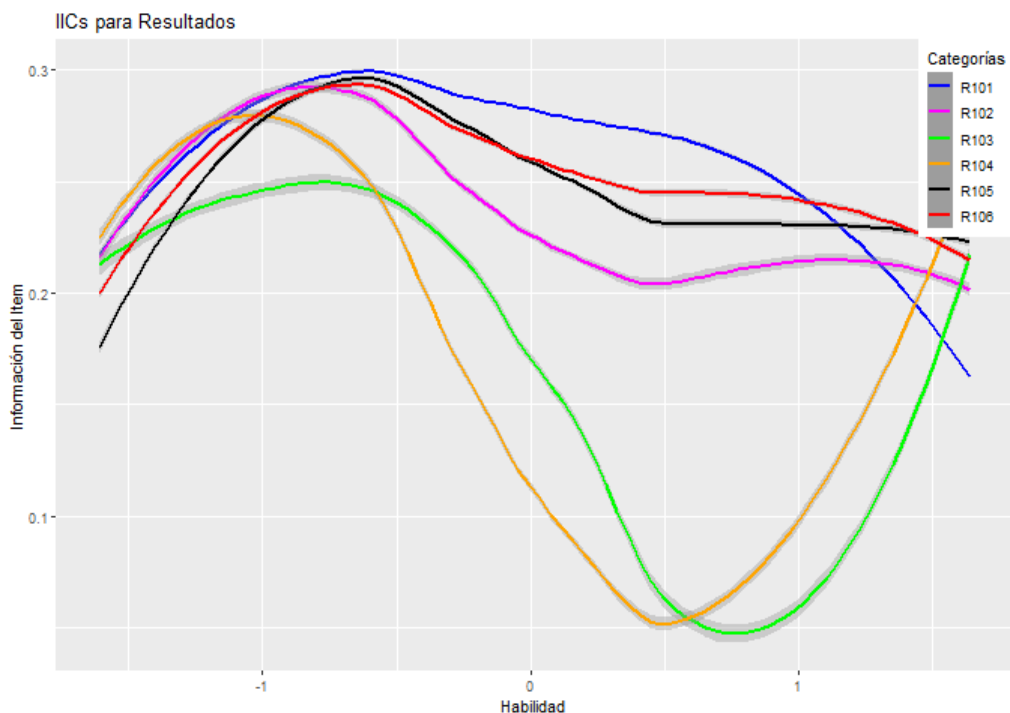
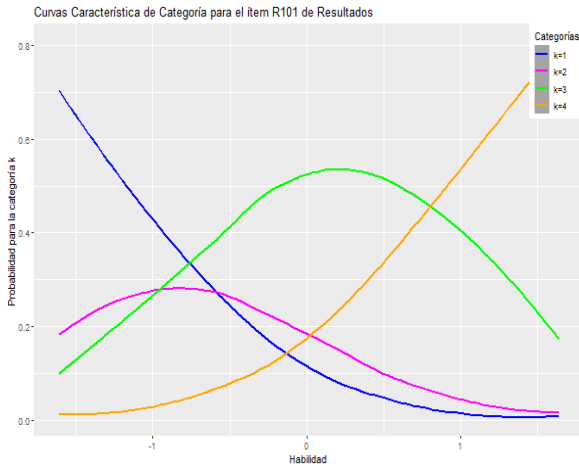
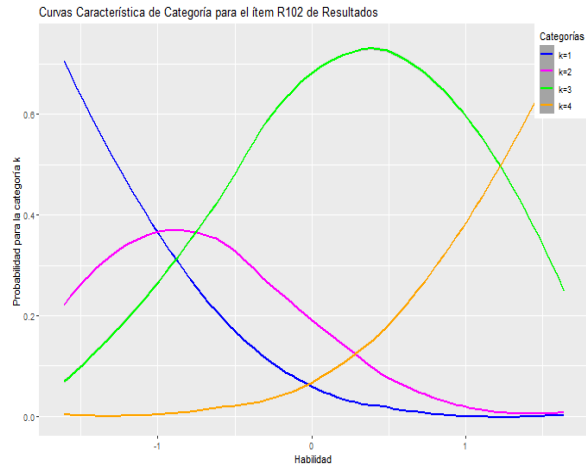


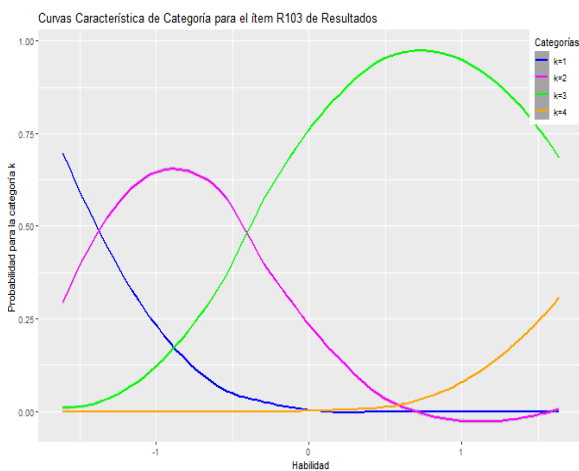
Figura 4.2: Criterio de Información del Ítem del subrasgo *Resultados*.



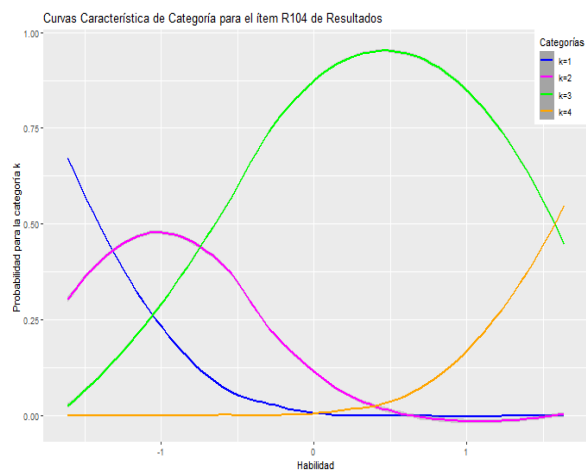
(a) CCCs del ítem R101.



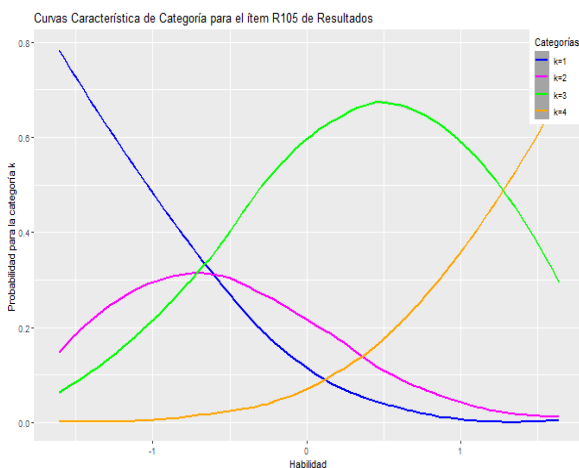
(b) CCCs del ítem R102.



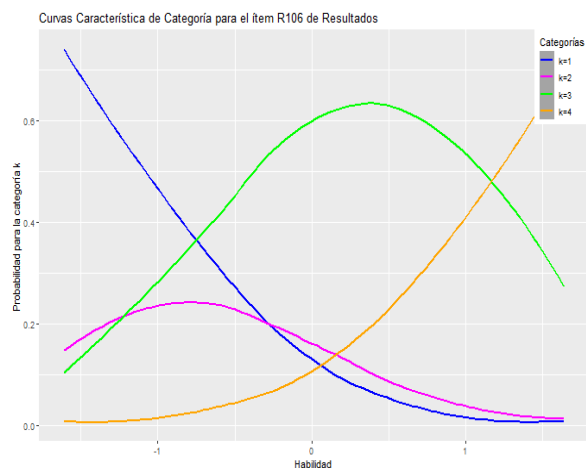
(c) CCCs del ítem R103.



(d) CCCs del ítem R104.



(e) CCCs del ítem R105.



(f) CCCs del ítem R106.

Figura 4.3: Curvas Características de Categoría del subrasgo *Resultados*.

La frecuencia de respuestas en el subrasgo *Resultados* dentro de cada categoría k se refleja en la Tabla 4.1 y en efecto se comprueba que se utilizaron todas las categorías para cada ítem.

	R101	R102	R103	R104	R105	R106
1	109	113	123	120	121	116
2	36	36	38	29	34	31
3	76	82	80	85	82	91
4	122	112	102	109	106	105

Tabla 4.1: Frecuencia de respuesta para el subrasgo *Resultados*

Al compilar el objeto GRM, el proceso de estimación para este subrasgo tardó aproximadamente 10 minutos en un ordenador con procesador Intel Core i5 8th generación, alcanzando la convergencia en las primeras 3000 iteraciones.

La impresión directa del objeto GRM en la consola de R proporciona las siguientes estimaciones (véase las Tablas 4.2, 4.3, 4.4).

	mean	se_mean	sd	2.5 %	25 %	50 %	75 %	97.5 %	n_eff	Rhat
a[1]	1.80	0.00	0.21	1.41	1.65	1.79	1.93	2.24	5455	1.0000
a[2]	2.24	0.00	0.25	1.78	2.06	2.23	2.40	2.74	4647	0.9998
a[3]	3.47	0.01	0.39	2.76	3.29	3.45	3.72	4.28	3623	1.0010
a[4]	3.26	0.01	0.37	2.58	3.01	3.24	3.49	4.03	3640	1.0003
a[5]	2.04	0.00	0.22	1.63	1.90	2.04	2.18	2.51	5991	1.0005
a[6]	1.82	0.00	0.20	1.44	1.68	1.81	1.95	2.25	5555	0.9998

Tabla 4.2: Estimación del parámetro de discriminación del subrasgo *Resultados*.

	mean	se_mean	sd	2.5 %	25 %	50 %	75 %	97.5 %	n_eff	Rhat
b[1,1]	-1.14	0.00	0.17	-1.47	-1.24	-1.14	-1.02	-0.82	4061	1.0007
b[1,2]	-0.48	0.00	0.15	-0.79	-0.58	-0.48	-0.38	-0.19	3866	1.0006
b[1,3]	0.87	0.00	0.16	0.56	0.76	0.87	0.98	1.19	4299	1.0004
b[2,1]	-1.23	0.00	0.19	-1.61	-1.35	-1.23	-1.10	-0.87	3171	1.0010
b[2,2]	-0.50	0.00	0.17	-0.84	-0.62	-0.50	-0.39	-0.17	3210	1.0011
b[2,3]	1.18	0.00	0.19	0.83	1.06	1.18	1.31	1.56	3550	1.0012
b[3,1]	-1.37	0.00	0.25	-1.88	-1.53	-1.36	-1.20	-0.91	2566	1.0017
b[3,2]	-0.37	0.00	0.22	-0.82	-0.52	-0.36	-0.21	0.07	2622	1.0010
b[3,3]	1.86	0.00	0.27	1.35	1.67	1.85	2.04	2.42	3135	1.0001
b[4,1]	-1.39	0.00	0.24	-1.88	-1.55	-1.39	-1.23	-0.94	2837	1.0007
b[4,2]	-0.63	0.00	0.22	-1.07	-0.78	-0.63	-0.49	-0.22	2711	1.0012
b[4,3]	1.57	0.00	0.25	1.10	1.40	1.57	1.74	2.08	3109	1.0004
b[5,1]	-1.00	0.00	0.17	-1.34	-1.11	-1.00	-0.88	-0.67	3537	1.0001
b[5,2]	-0.35	0.00	0.16	-0.68	-0.46	-0.35	-0.24	-0.04	3437	1.0006
b[5,3]	1.26	0.00	0.18	0.92	1.14	1.26	1.38	1.63	3989	1.0004
b[6,1]	-1.05	0.00	0.17	-1.39	-1.16	-1.04	-0.93	-0.73	3637	1.0005
b[6,2]	-0.49	0.00	0.16	-0.81	-0.60	-0.49	-0.39	-0.19	3643	1.0005
b[6,3]	1.17	0.00	0.17	0.84	1.05	1.17	1.29	1.51	3867	1.0013

Tabla 4.3: Estimación del parámetro de dificultad del subrasgo *Resultados*.

	mean	se_mean	sd	2.5 %	25 %	50 %	75 %	97.5 %	n_eff	Rhat
$\theta[1]$	1.62	0.00	0.54	0.71	1.24	1.56	1.94	2.80	12416	0.9998
$\theta[2]$	0.81	0.00	0.41	0.06	0.53	0.79	1.07	1.68	13256	0.9998
$\theta[3]$	-0.20	0.00	0.30	-0.81	-0.39	-0.19	0.00	0.36	11923	0.9998
$\theta[4]$	-0.02	0.00	0.39	-0.80	-0.28	-0.02	0.24	0.74	14339	1.0000
$\theta[5]$	1.12	0.00	0.44	0.36	0.82	1.09	1.39	2.09	12538	0.9997
$\theta[6]$	1.62	0.00	0.54	0.71	1.25	1.57	1.96	2.80	11647	0.9999
$\theta[7]$	0.39	0.00	0.39	-0.35	0.13	0.38	0.64	1.16	13327	0.9999
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$\theta[16]$	0.04	0.00	0.28	-0.50	-0.14	0.03	0.22	0.57	13131	1.0000
$\theta[17]$	1.63	0.01	0.56	0.69	1.23	1.58	1.97	2.91	11027	1.0000
$\theta[18]$	1.63	0.01	0.56	0.69	1.23	1.57	1.96	2.87	11192	1.0000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$\theta[337]$	1.62	0.01	0.55	0.70	1.24	1.56	1.95	2.87	10961	0.9998
$\theta[338]$	-0.03	0.00	0.26	-0.54	-0.20	-0.03	0.14	0.47	12534	0.9997
$\theta[339]$	1.62	0.01	0.54	0.71	1.23	1.57	1.97	2.83	11430	0.9998
$\theta[340]$	1.62	0.01	0.55	0.70	1.23	1.57	1.96	2.85	11416	0.9998
$\theta[341]$	1.17	0.00	0.44	0.42	0.87	1.13	1.43	2.15	11567	0.9999
$\theta[342]$	1.10	0.00	0.42	0.36	0.80	1.06	1.37	2.02	12105	0.9999
$\theta[343]$	0.22	0.00	0.29	-0.33	0.03	0.21	0.41	0.80	12759	0.9998

Tabla 4.4: Estimación del parámetro de habilidad del subrasgo *Resultados*.

La convergencia en este subrasgo se evidencia en la estacionariedad que existe en las cadenas tomadas. Para obtener los gráficos de traza se utilizó la función `traceplot()` del paquete `rstan`. Como se observa, cada trazo es único y los valores extraídos varían alrededor de una media específica.

Las Figuras 4.4 y 4.5 muestran los gráficos de traza para los 8 parámetros de discriminación a y los 24 parámetros de dificultad b por los $length(k) - 1 = 3$ grados de libertad en la estimación.

La Figura 4.6 muestra 3 gráficos al azar que representan la habilidad de innovación para las organizaciones 1, 25 y 50.

Se debe tomar en cuenta que los gráficos solo muestran las últimas 3000 iteraciones, ya que `inc_warmup = FALSE`. Por lo general, la aparición de una “oruga gorda y peluda” en los gráficos de traza indica convergencia (Lunn y col., 2012). Tal como lo hacen ver las Figuras 4.4, 4.5 y 4.6.

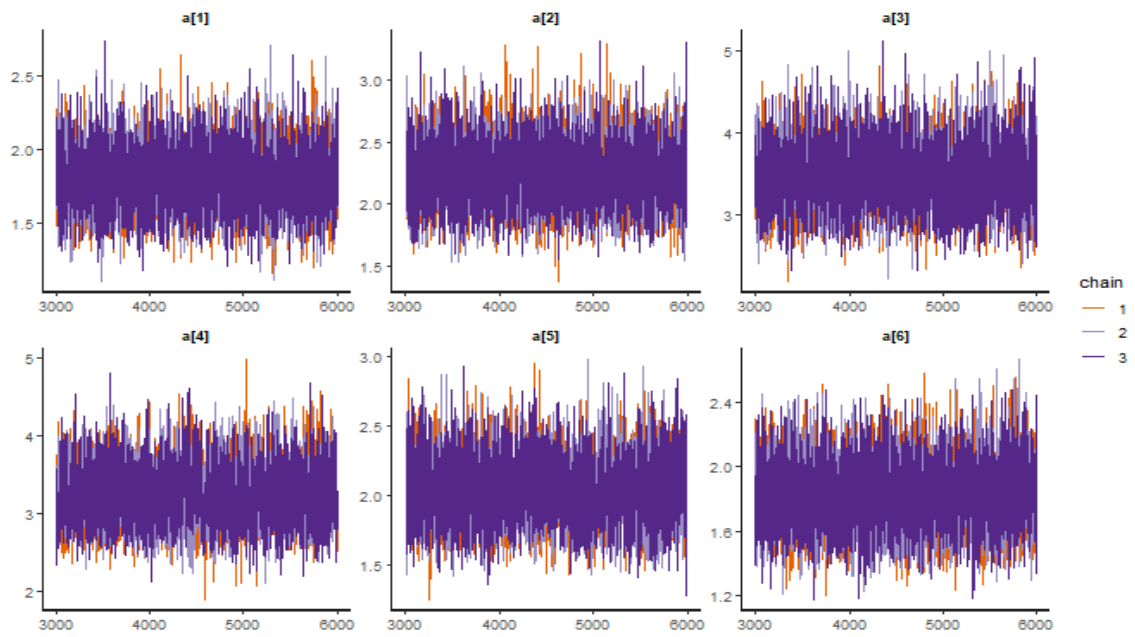


Figura 4.4: Gráfica de trazas de los parámetros de discriminación para el subrasgo *Resultados*.

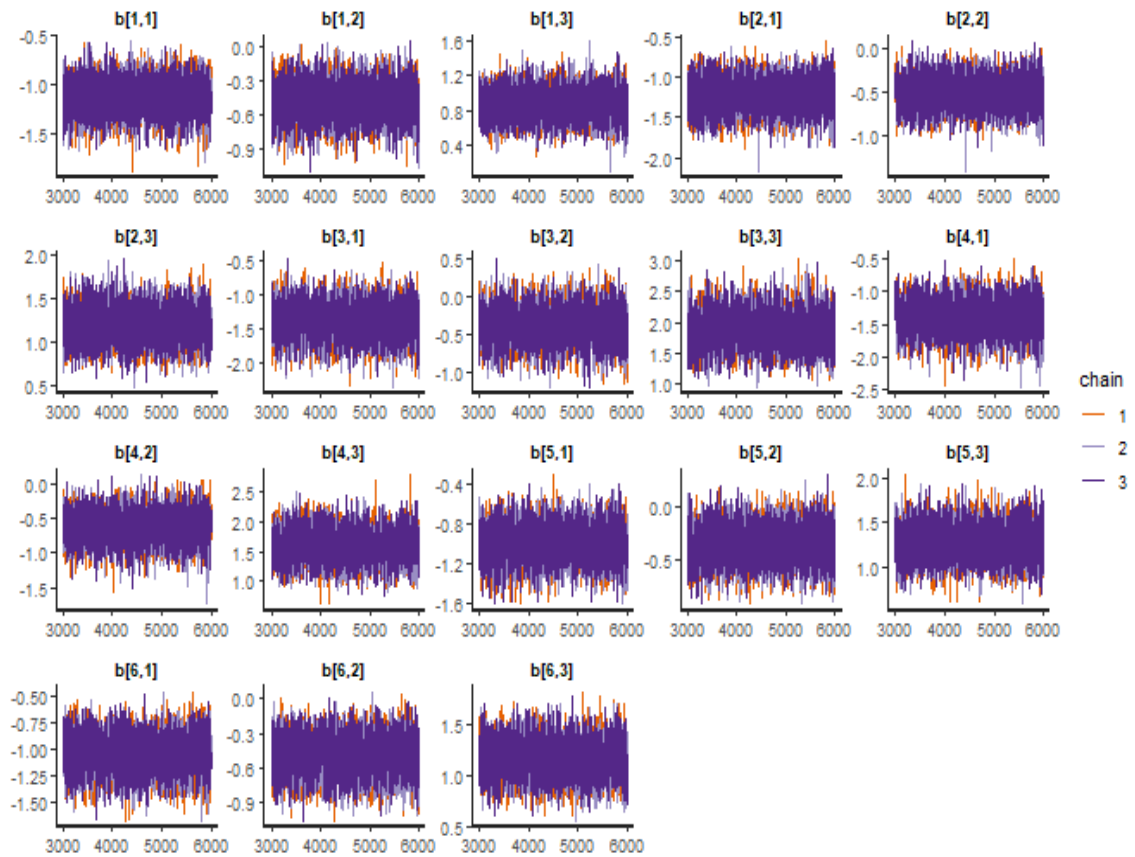
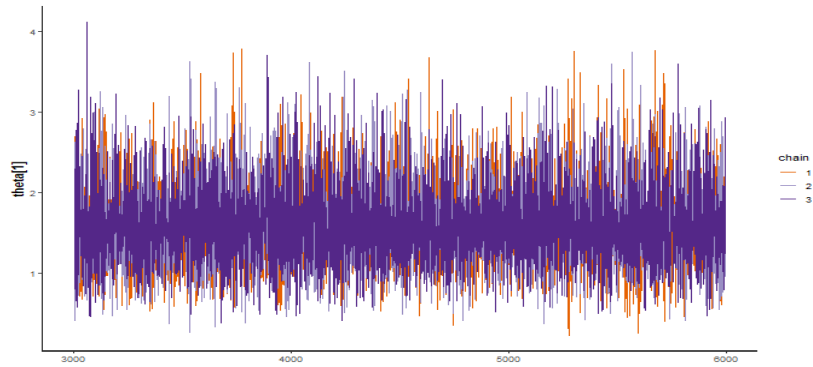
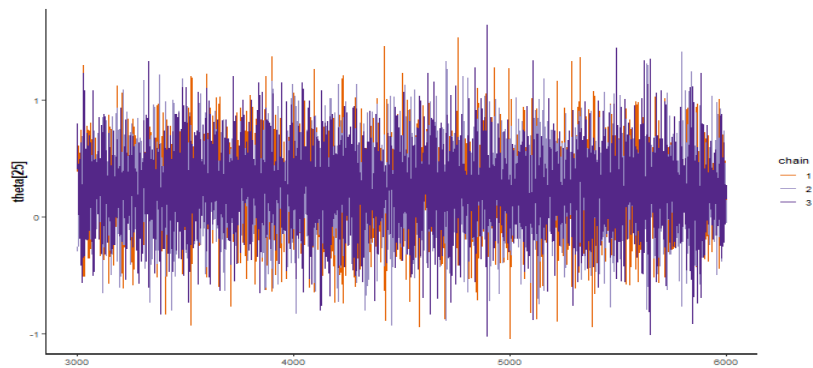


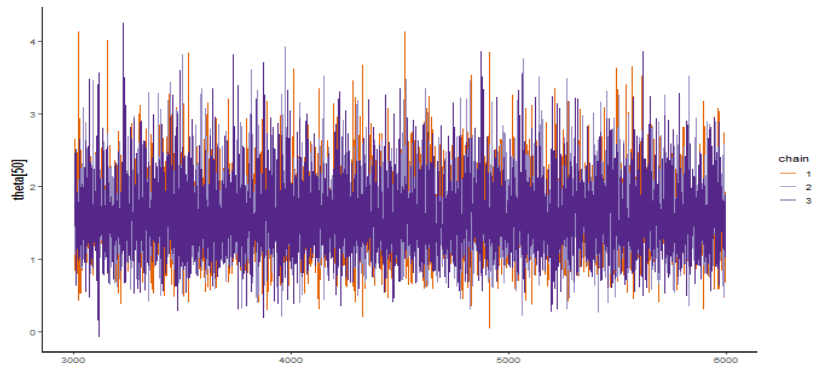
Figura 4.5: Gráfica de trazas de los parámetros de dificultad para el subrasgo *Resultados*.



(a) Gráfico de traza de la habilidad del subrasgo *Resultados* de la organización 1.



(b) Gráfico de traza de la habilidad del subrasgo *Resultados* de la organización 25.



(c) Gráfico de traza de la habilidad del subrasgo *Resultados* de la organización 50.

Figura 4.6: Gráficos de traza de la habilidad del subrasgo *Resultados*.

Notemos que en todas las figuras 4.4, 4.5, 4.6 las trazas de las tres cadenas paralelas se superponen entre sí, descartando el periodo burn-in, significando así, una correcta mezcla de las cadenas MCMC.

Por otro lado, la función de autocorrelación (ACF) de todos los parámetros correspondientes indican que hay poca o ninguna correlación entre extracciones secuenciales de la distribución (véase figuras 4.7, 4.8 y 4.9). Por tanto, las cadenas se mezclan bien y podemos

decir que las extracciones sucesivas de la distribución son (casi) independientes entre sí.

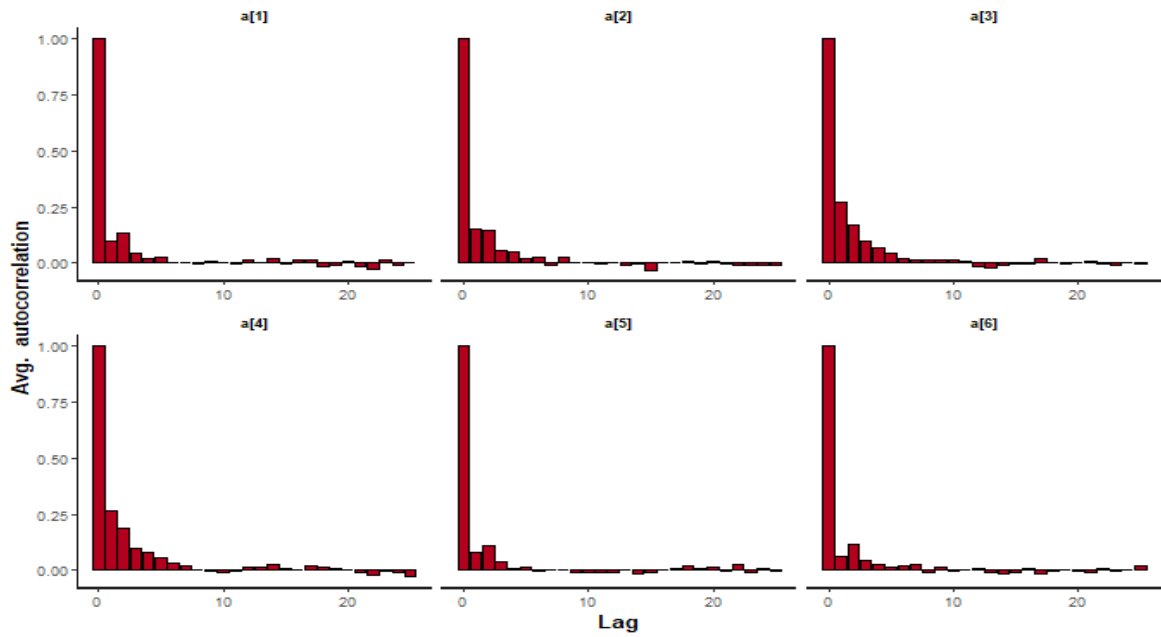


Figura 4.7: ACF del parámetro de discriminación para el subrasgo *Resultados*.

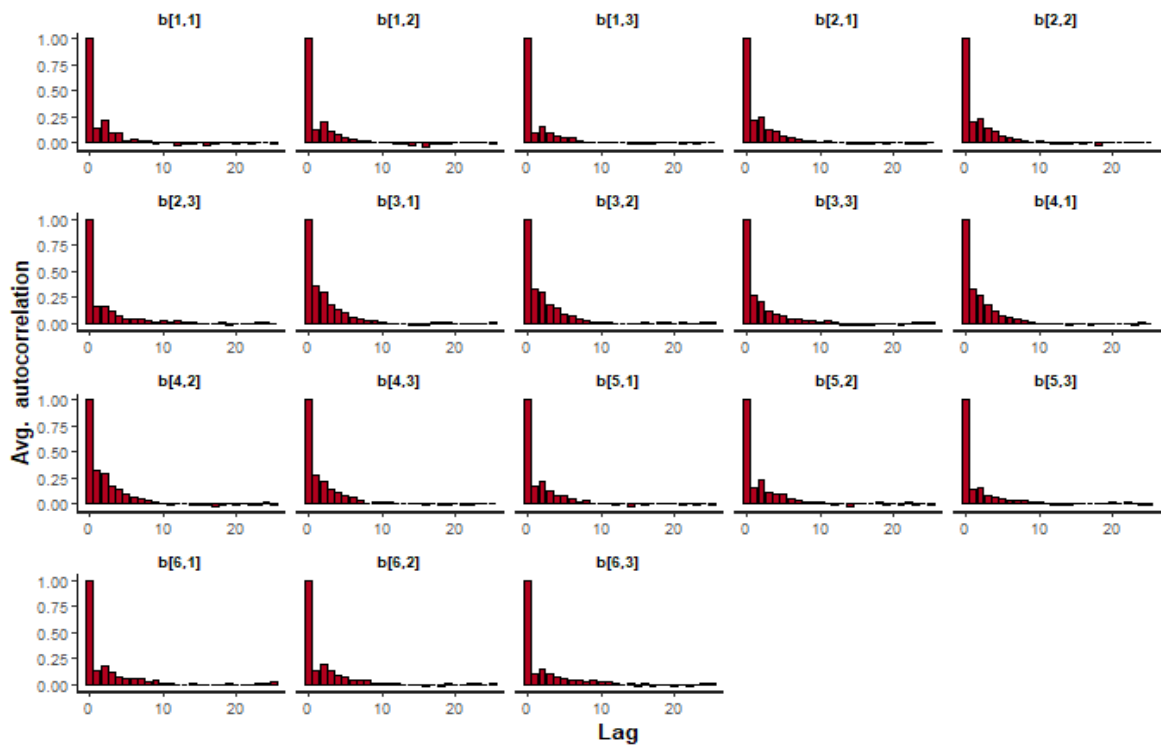
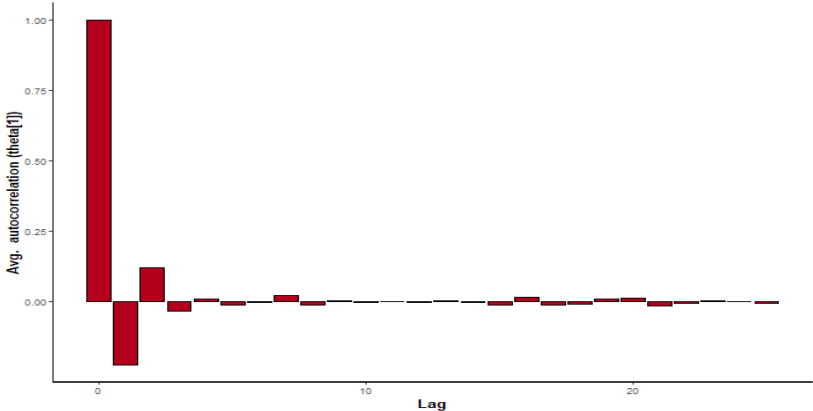


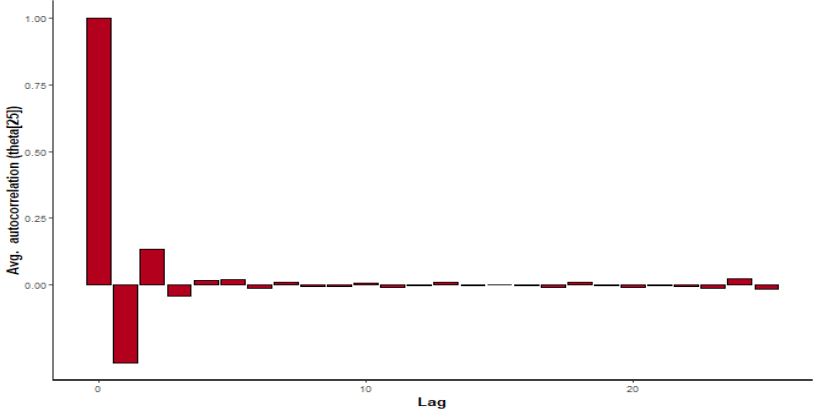
Figura 4.8: ACF del parámetro de dificultad para el subrasgo *Resultados*.

En resumen, el análisis gráfico de la simulación no mostró signos de falta de convergencia para ningún parámetro. Además, dado que las cadenas convergieron relativamente rápido y

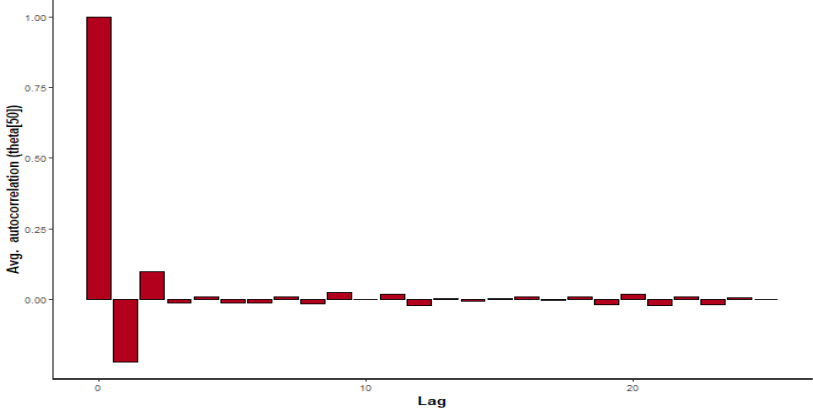
visto que la autocorrelación de las mismas fue baja, las muestras de las múltiples cadenas se agrupan para hacer inferencias válidas sobre la distribución posterior.



(a) ACF de la habilidad del subrasgo *Resultados* de la organización 1.



(b) ACF de la habilidad del subrasgo *Resultados* de la organización 25.



(c) ACF de la habilidad del subrasgo *Resultados* de la organización 50.

Figura 4.9: ACF de la habilidad del subrasgo *Resultados*.

Como todas las cadenas alcanzaron la estacionariedad, la varianza entre cadenas (B) y la varianza dentro de la cadena (W) son comparables, resultando un valor de umbral del estadístico PSRF inferior a 1.1 (como vemos en la Tabla 4.5); lo que confirma aún más la

convergencia del modelo.

Parámetros	máx(\hat{R})
a	1.0010
b1	1.0017
b2	1.0012
b3	1.0013
θ	1.0005

Tabla 4.5: Estadístico PSRF para el subrasgo *Resultados*

Resta definir la cantidad de pasos para inferir la distribución posterior. Una regla general es tomar aproximadamente 1000 pasos de la cadena o cadenas, después del burn-in. Por ejemplo, si se desea calcular los percentiles 2.5 y 97.5 para informar intervalos creíbles posteriores al 95 % de colas iguales, 1000 es un tamaño de muestra mínimo; lo que se cumple en la columna n_{eff} de las Tablas 4.2, 4.3, 4.4; en donde ninguno de los valores es inferior a 1000.

Respecto a los intervalos de confianza, tomamos como ejemplo la estimación $\theta_{[17]}$ de la Tabla 4.4. Dado que $\theta_{[17]}$ tiene una media posterior de 1.63 con un intervalo de credibilidad del 95 % de (0.69, 2.91); esta estimación tiene un efecto de estimación significativamente superior a la media y, por tanto, se clasifica como eficaz.

Se concluye entonces que todas las cadenas de MCMC en el modelo GRM convergen y las estimaciones de las distribuciones posteriores de los parámetros correspondientes son eficaces. Con estas estimaciones y gracias al sistema 1.5 obtenemos las probabilidades de respuesta al ítem y las CCCs del subrasgo *Resultados* para cada ítem.

De igual forma, gracias a estas estimaciones y a la ecuación 1.9 se obtiene el criterio de información del ítem que indica la información proporcionada por un ítem particular sobre θ . Además, de la información del test dada por la ecuación 1.10 que indica la precisión con la que un test estima la habilidad del subrasgo (Apéndice A.6).

4.5. Clasificación de las Organizaciones

Dado que el subrasgo no es ordinal con respecto a otros sujetos y se lo usa para comparaciones dentro de la persona, y el rasgo latente o habilidad se utiliza para comparaciones entre personas. El potencial de innovación de la organización puede clasificarse u ordenarse acorde a su competencia en los diferentes subrasgos pertenecientes al modelo CRI. La idea entonces es sintetizar los resultados de manera inversa para sacar conclusiones en cada nivel hasta alcanzar el de mayor abstracción, es decir, el potencial buscado.

La forma en que se mide dicho potencial es la siguiente: Primero, se toma las habilidades estimadas de la categoría *Capacidades (C)*, es decir, las habilidades de los subrasgos *Fuentes de Información, Fuentes de Financiamiento, Actividades y Objetivos*, y se lo divide para cuatro, concluyendo un promedio general de la categoría C. Segundo, al constar las categorías *Resultados (R)* e *Impactos (I)* con un solo subrasgo cada uno, no fue necesario realizar otro proceso adicional. Con los tres resultados de las categorías CRI se hace una suma de las mismas, teniendo como resultado una medición ordinal del potencial de innovación de la organización, donde una puntuación más alta, refleja un mayor nivel de competencia.

Con estos resultados, se crea tres niveles de medición del potencial de innovación: Alto, Medio y Bajo, con el fin de tener un análisis más general y sintetizado. La siguiente tabla muestra la estimación promedio del potencial y las categorías CRI, en los tres niveles.

Niveles	# Empresas	Medición	Capacidades	Resultados	Impactos
ALTO	86	3.1592	0.7556	1.1430	1.2606
MEDIO	86	0.9841	0.2611	0.0631	0.6599
BAJO	171	-1.4394	-0.3397	-0.5770	-0.5226

Tabla 4.6: Estimación promedio del potencial de innovación y las categorías CRI.

Toda organización necesita saber que los esfuerzos y recursos invertidos en innovación realmente genera crecimiento. Este análisis de medición del potencial de innovación nos permite responder esa necesidad.

Los resultados de la Tabla 4.6 evidencian que 86 empresas ocupan una posición alta en innovación, con un promedio del 3.16 en medición del potencial. En este nivel, la categoría C tiene un promedio de puntuación del 0.76, la categoría R tiene un promedio del 1.14 y la categoría I tiene un promedio del 1.26. Del mismo modo, 86 empresas ocupan una posición media en innovación, con un promedio del 0.98 en medición del potencial. En este nivel, la categoría C tiene un promedio de puntuación del 0.26, la categoría R tiene un promedio del 0.06 y la categoría I tiene un promedio del 0.66. Finalmente, 171 empresas ocupan una posición baja en innovación, con un promedio del -1.44 en medición del potencial. En este nivel, la categoría C tiene un promedio de puntuación del -0.34, la categoría R tiene un promedio del -0.58 y la categoría I tiene un promedio del -0.52.

En consecuencia, el 50% de la muestra tiene un bajo nivel de innovación, y, por tanto, las organizaciones que representan este porcentaje presentan el reto y la necesidad de implementar procesos innovadores y la capacidad de aplicarlos a nuevos productos para que puedan posicionarse competitivamente en el mercado.

Capítulo 5

Conclusión y Recomendaciones

5.1. Conclusión

El objetivo de este trabajo de investigación era utilizar los modelos IRT para entender el concepto latente de la Innovación y los subrasgos o dominios que comprenden su estudio.

La solución propuesta se basa en un enfoque de la Teoría de Respuesta al Ítem que utiliza un Modelo de Respuesta Gradual. Para proporcionar a cada empresa que participó en el cuestionario CRI2 del conjunto de datos, una estimación de su potencial de innovación, así como de su competencia en cada subrasgo, se empleó un método de Monte Carlo con cadena de Markov llamado Monte Carlo Hamiltoniano. El enfoque de inferencia bayesiana permitió estimar simultáneamente la gran cantidad de parámetros del modelo y para validar la convergencia MCMC se empleó varios criterios, dos de ellos gráficos (trace plots, ACF) y uno estadístico (Gelman-Rubin).

Para simplificar el enfoque multidisciplinario y racionalizar el procedimiento de puntuación, se empleó el Modelo de Respuesta Gradual unidimensional para cada subrasgo de las categorías CRI y luego se determinó el rasgo latente de mayor abstracción, es decir, el potencial de innovación de la empresa.

De la muestra, 86 empresas están por encima de la media de medición, donde el máximo puntaje es 4.90 y el mínimo -5.55. Esto nos dice que más de la mitad de la muestra, 257, no tiene un buen potencial de innovación y dado que la baja innovación organizacional es un problema estructural que requiere un cambio profundo en la forma en que la empresa se organiza y funciona, se necesita de políticas y programas más ambiciosos para que estas empresas mejoren su puntuación en innovación.

La aparición de valores atípicos es muy improbable, esto se debe a que el espacio muestral de las posibles respuestas es pequeño y discreto $x_{s,i} \in \{1, \dots, 4\}$.

El método de estimación utilizado en este trabajo de investigación es escalable en el sentido de que se puede aplicar a cualquier conjunto de datos que consista en respuestas politómicas en escala de Likert, siempre que el conjunto de ítems que conforme el test mida solo un subrasgo. Por tanto, el modelo y los pasos de construcción también son válidos si se agregan más datos o se cambia el test.

La pandemia del COVID-19 ha provocado una crisis económica muy severa en Ecuador, reduciendo un 7,8 % la economía ecuatoriana en 2020 y aunque esta se recuperó un 4,2 % en 2021, es la más baja de Sudamérica (GESTION-DIGITAL, 2022). Si bien muchos predicen que nuestra recuperación económica llevará tiempo, es importante aprovechar el potencial innovador que presentan las empresas y trabajar juntos para crear una recuperación fluida, dinámica y sostenible. Este trabajo de investigación es precisamente uno de los pasos que pueden efectuarse, ya que ofrece obtener una estimación del potencial de innovación de las organizaciones de una muestra y de su destreza en cada subrasgo. Creemos además que la participación de un panel de expertos en materia de innovación daría conclusiones más propicias respecto a que mejoras pueden realizar estas empresas para aumentar su competitividad en el mercado.

5.2. Recomendaciones

- El modelo estima de acuerdo al contexto de las preguntas que han producido los datos, entonces es importante señalar que la data utilizada refleje los subrasgos que conforman el objeto de estudio. De la misma manera, que se ha realizado en este trabajo de investigación, al emplear la información proveniente del cuestionario CRI2 validado por expertos en medición de innovación como Morales y col., 2022.
- Para un sujeto, lo mejor sería ofrecerle un test más accesible con un menor número de ítems, reduciendo así el tiempo que este tardaría en completarlo, pero se recomienda no hacerlo, ya que un mayor número de ítems por dominio produce una estimación más precisa de los subrasgos y un mayor número de dominios arroja un resultado más preciso para estimar el rasgo latente.
- Para determinar con mayor precisión el impacto de cada dominio en la puntuación general se recomienda incluir ítems de dificultad variable en cada dominio y podría ser útil añadir también preguntas neutrales en las que sea beneficioso responder “No sé”,

en lugar de “Totalmente de acuerdo”, “de acuerdo”, “en desacuerdo” o “Totalmente en desacuerdo”. Esto podría aumentar la dispersión de las respuestas, lo que a su vez aumentaría la variabilidad entre sujetos. Esta combinación de niveles de dificultad entre preguntas puede dar como resultado valores más bajos para las puntuaciones de los subrasgos, por lo tanto, creemos que esto conduciría a una estimación más precisa de la importancia de cada dominio por separado.

- Un análisis interesante a realizar sería validar la elección de modelos IRT avanzados para expresar de manera lineal la relación que existe entre las habilidades de dominio (que en nuestro caso los llamamos subrasgos) y la habilidad general (que en nuestro caso lo llamamos rasgo latente o habilidad). Utilizando la idea de de la Torre y Song, 2009, esta relación se expresaría como $\theta_{sq}^{(1)} = \lambda_q \theta_s^{(2)} + \epsilon_{sq}$, donde $\theta_{sq}^{(1)}$ es el rasgo de primer orden del dominio q para el sujeto s , $\theta_s^{(2)}$ es el rasgo de segundo orden o habilidad general para el sujeto s , λ_q es la capacidad de regresión del coeficiente latente q sobre la habilidad general, tal que $0 \leq \lambda_q \leq 1$ y ϵ_{sq} es un término de error que evita la sobreestimación del rasgo latente de segundo orden y es independiente de todos los demás términos, excepto de λ .

Entonces, dado que el nivel de habilidad general (potencial de innovación) se vería directamente afectado por las habilidades del rasgo latente de primer orden (Fuentes de Financiamiento, Fuentes de Información, Actividades, Objetivos, Resultados e Impactos) la elección del modelo IRT de Orden Superior (HO-IRT) sería el modelo GRM de Orden Superior (HO-GRM) para estructura simple con $q = 1, 2, \dots, Q$ dimensiones.

Aunque nuestra recomendación es analizar de manera individual las categorías CRI y posterior a ello tener una medición general del potencial de innovación, este tipo de modelamiento podría funcionar como un análisis complementario al nuestro.

Anexos

Apéndice A

Simulación del Modelo

A.1. Simulación

Los subrasgos *Fuentes de Información*, *Fuentes de Financiamiento*, *Actividades*, *Objetivos*, *Resultados* e *Impactos*, siguen el mismo diseño de código. Puede revisar la simulación dirigiéndose al repositorio de GitHub con el siguiente link <https://github.com/Ana-Gabriela-Taipe/IRT-Innovacion-Ana-Taipe.git>

A.2. Estimación de parámetros

Ítem	a	b1	b2	b3
C101	1.4670	0.3348	0.9069	1.7150
C102	1.7504	-0.4453	-0.1061	1.1468
C103	1.2426	-1.1633	-0.9678	0.3848
C104	1.7890	-0.4758	-0.1139	1.4508
C105	1.0752	-1.3641	-0.9932	0.2360
C106	1.1946	-1.0447	-0.3859	0.9999
C107	1.1184	0.0014	0.3405	1.2906
C108	0.8759	0.1471	0.6748	1.4722
C201	0.0667	-2.6288	-2.2286	-0.5129
C202	1.6028	-0.4042	0.0477	1.5189
C203	1.4914	0.0189	0.5727	1.4775
C204	1.0695	1.2694	1.8462	2.2006
C205	1.0078	1.6671	2.2529	2.658
C301	1.7268	-1.7246	-1.2856	0.1613
C302	1.2077	-0.4216	-0.0143	0.8996
C303	1.4628	-1.433	-1.0333	0.4068
C304	1.7285	-0.9991	-0.6205	0.6387
C305	1.6402	-1.2502	-0.8631	0.3059

C306	3.0631	-1.363	-0.7067	0.8097
C307	1.9731	-1.0785	-0.7164	0.8416
C308	2.143	-0.9897	-0.5764	0.8976
C309	1.2303	-1.1364	-0.5309	0.7116
C401	1.8417	-2.9361	-2.0755	-0.2722
C402	1.6299	-1.4594	-1.1876	0.3703
C403	2.7088	-3.4448	-3.2328	-0.1702
C404	2.3348	-2.6686	-2.2996	0.2177
C405	2.7995	-2.9355	-2.5624	0.2173
C406	2.8658	-3.6995	-3.0943	-0.1286
C407	2.7671	-3.6743	-3.3463	-0.0692
C408	1.9298	-2.8406	-2.1232	0.0041
C409	2.4374	-3.0449	-2.5211	0.0704
R101	1.7983	-1.1363	-0.4802	0.8695
R102	2.238	-1.2281	-0.5031	1.1831
R103	3.4687	-1.3719	-0.3661	1.8624
R104	3.2569	-1.3916	-0.635	1.5746
R105	2.0448	-0.9994	-0.3493	1.2649
R106	1.8201	-1.0467	-0.4942	1.1707
I101	2.9339	-3.4187	-2.7086	0.3591
I102	2.6267	-2.4363	-1.5962	0.9736
I103	2.5877	-2.8116	-2.201	0.3774
I104	2.3476	-2.614	-1.8479	0.5324
I105	1.5944	-2.3128	-2.0003	0.1507
I106	2.3115	-2.4995	-1.7411	0.7414
I107	1.8754	-1.149	-0.6039	0.9352
I108	1.6788	-1.2477	-0.8423	0.658
I109	2.301	-2.7792	-2.2661	0.1314

Tabla A.1: Estimación de parámetros de los 6 subrasgos: Fuentes de Información, Fuentes de Financiamiento, Actividades, Objetivos, Resultados e Impactos

A.3. Gráficas de convergencia (traza y ACF)

Fuentes de Información

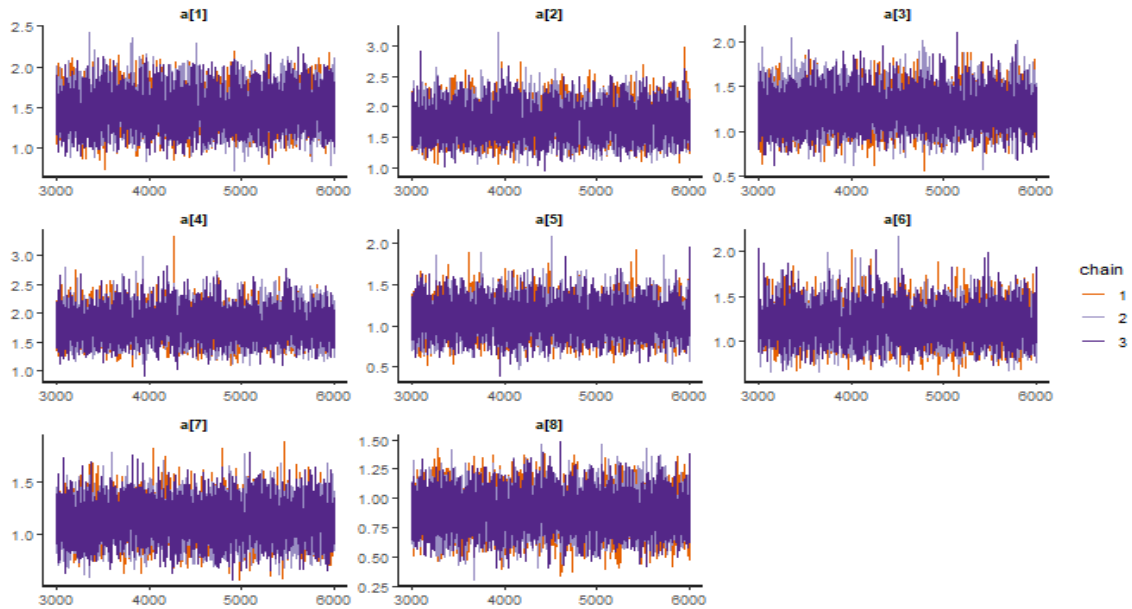


Figura A.1: Gráficos de traza del parámetro de discriminación del subrasgo *Fuentes de Información*.

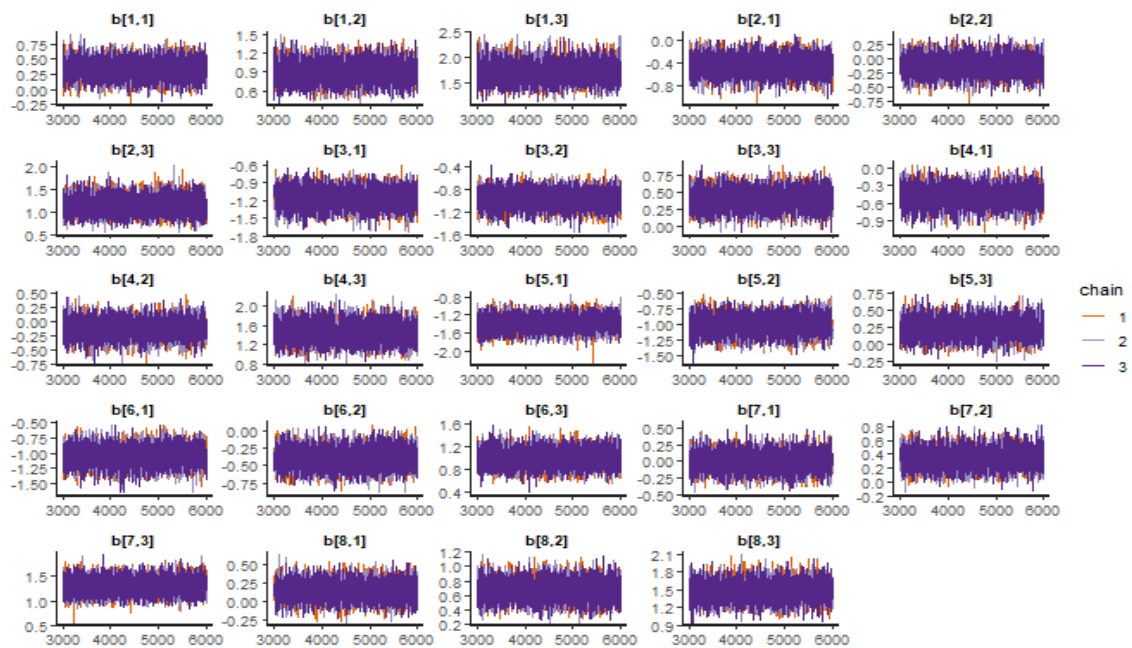


Figura A.2: Gráficos de traza del parámetro de dificultad del subrasgo *Fuentes de Información* con tres grados de libertad.

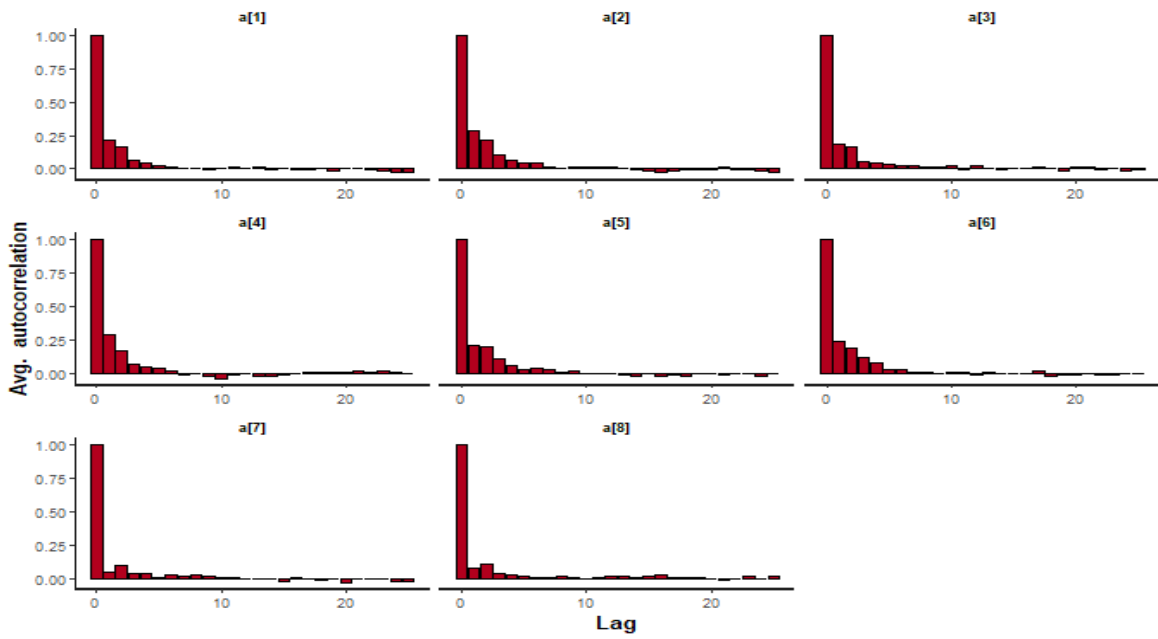


Figura A.3: Gráfica ACF del parámetro de discriminación del subrasgo *Fuentes de Información*.

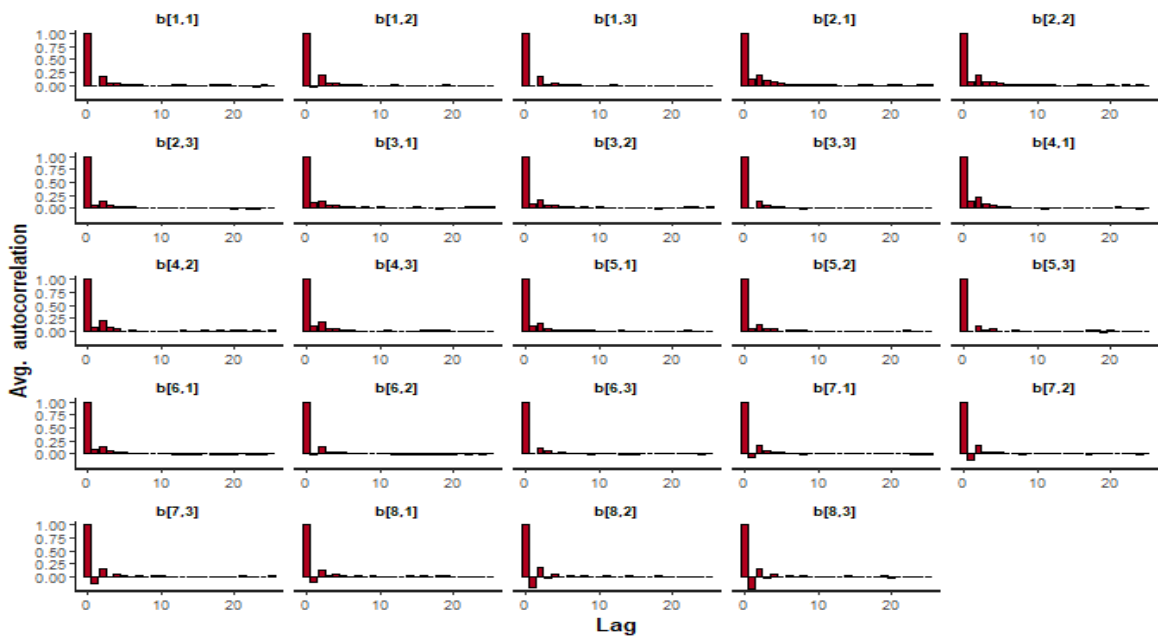
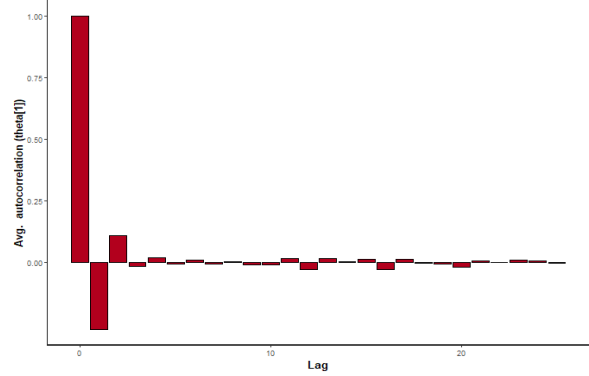
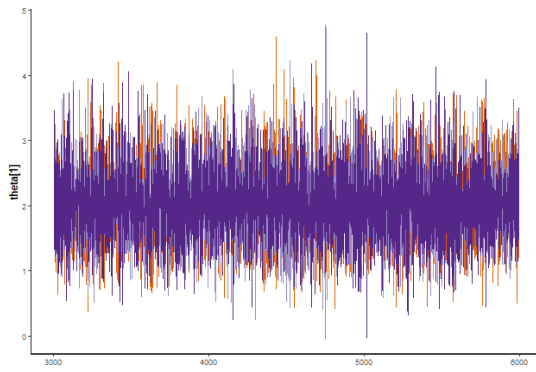
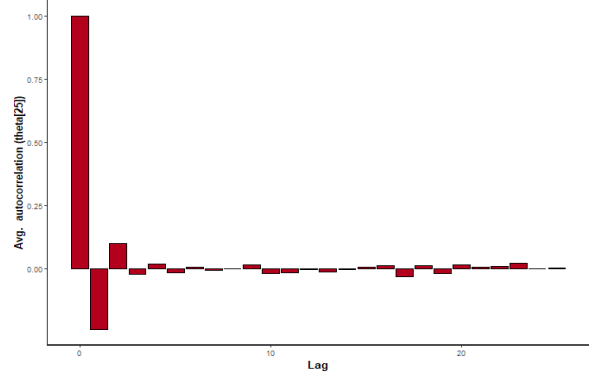
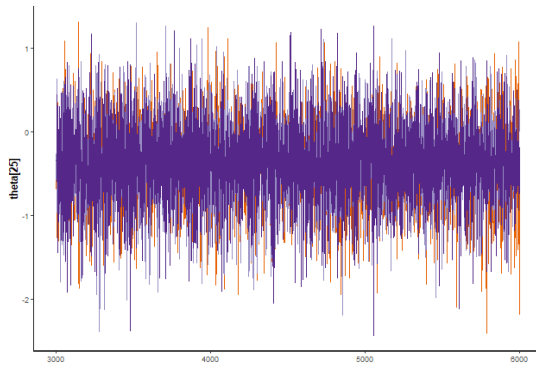


Figura A.4: Gráfica ACF del parámetro de dificultad del subrasgo *Fuentes de Información*.



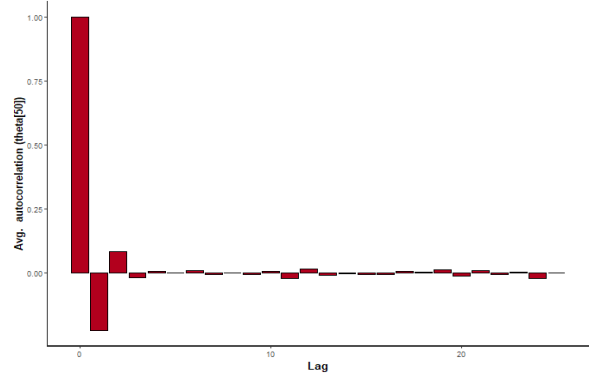
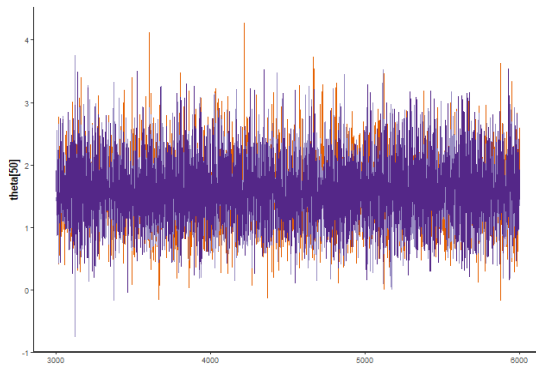
(a) Gráfico de traza para la habilidad del subrasgo *Fuentes de Información* en la organización 1.

(b) Gráfica ACF para la habilidad del subrasgo *Fuentes de Información* en la organización 1.



(c) Gráfico de traza para la habilidad del subrasgo *Fuentes de Información* en la organización 25.

(d) Gráfica ACF para la habilidad del subrasgo *Fuentes de Información* en la organización 25.



(e) Gráficas de traza para la habilidad del subrasgo *Fuentes de Información* en la organización 50.

(f) Gráfica ACF para la habilidad del subrasgo *Fuentes de Información* en la organización 50.

Figura A.5: Gráfico de traza y gráficos ACF de tres organizaciones de la muestra con el subrasgo *Fuentes de Información*.

Fuentes de Financiamiento

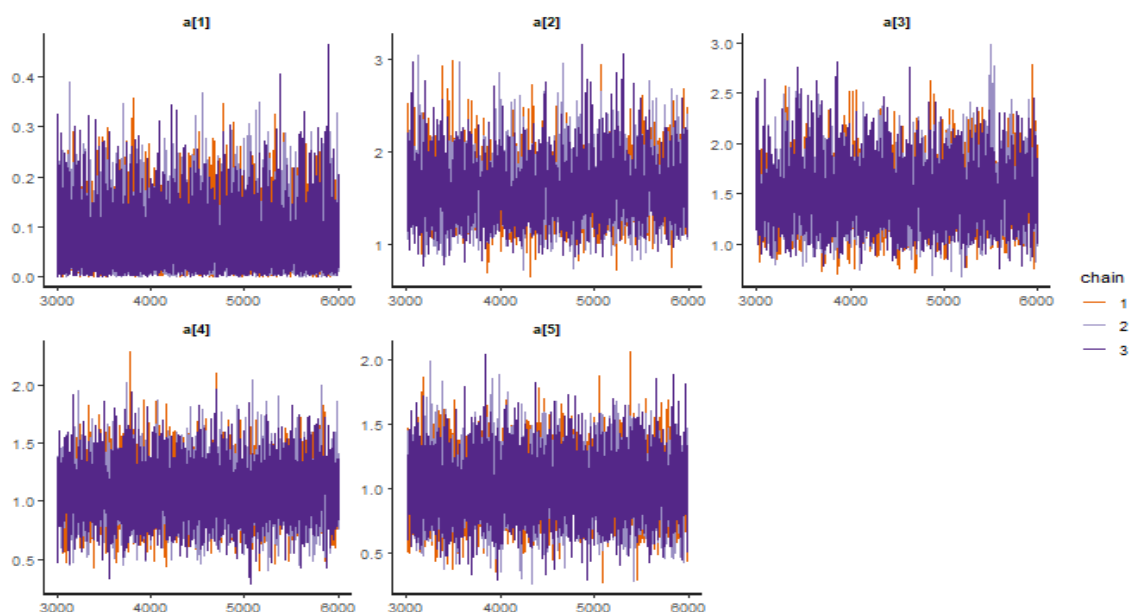


Figura A.6: Gráficos de traza del parámetro de discriminación del subrasgo *Fuentes de Financiamiento*.

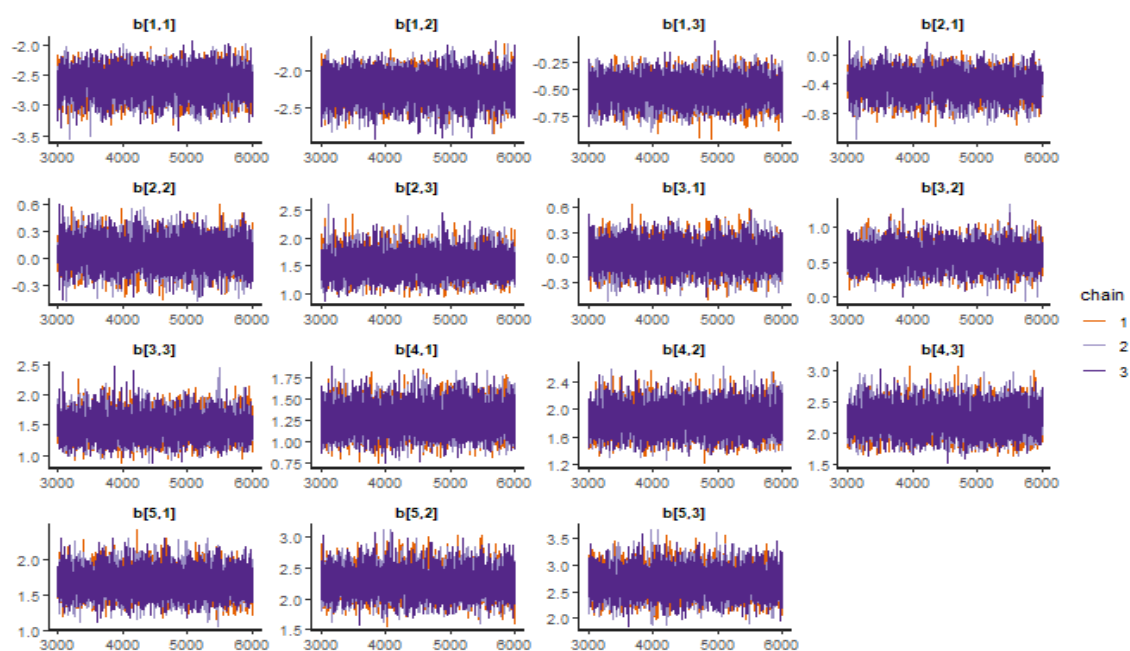


Figura A.7: Gráficos de traza del parámetro de dificultad del subrasgo *Fuentes de Financiamiento* con tres grados de libertad.

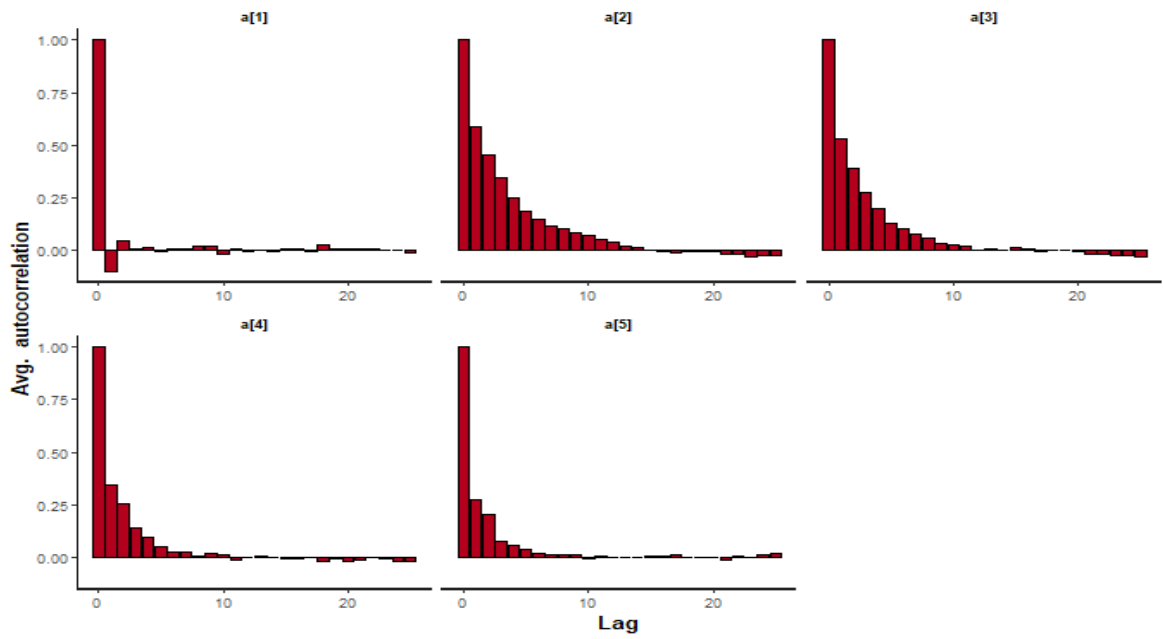


Figura A.8: Gráfica ACF del parámetro de discriminación del subrasgo *Fuentes de Financiamiento*.

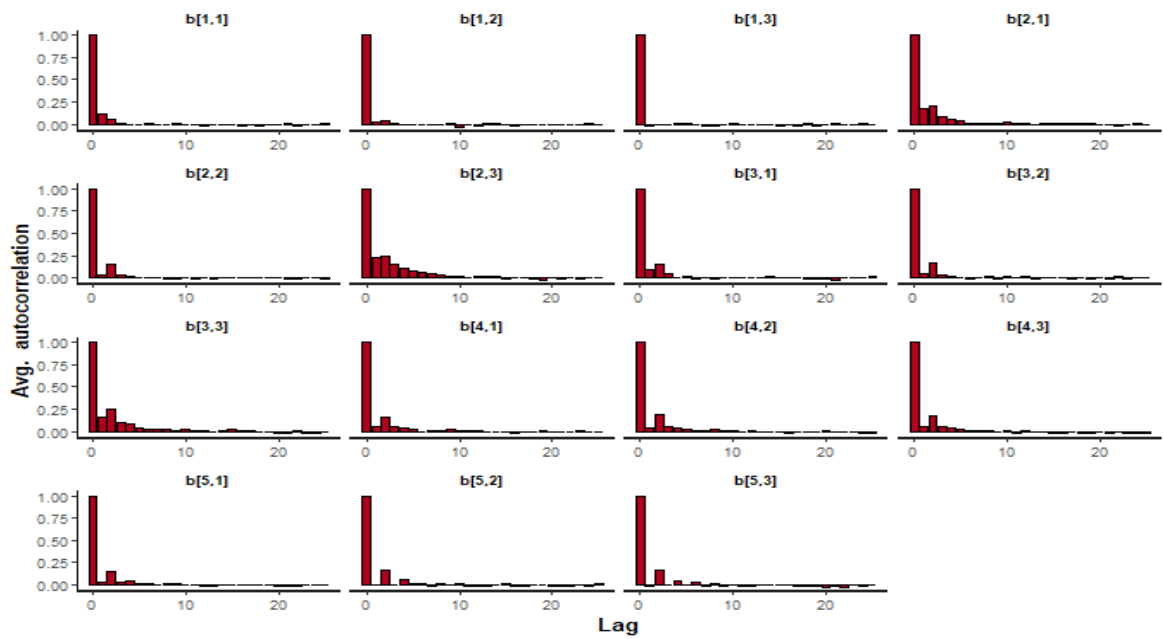
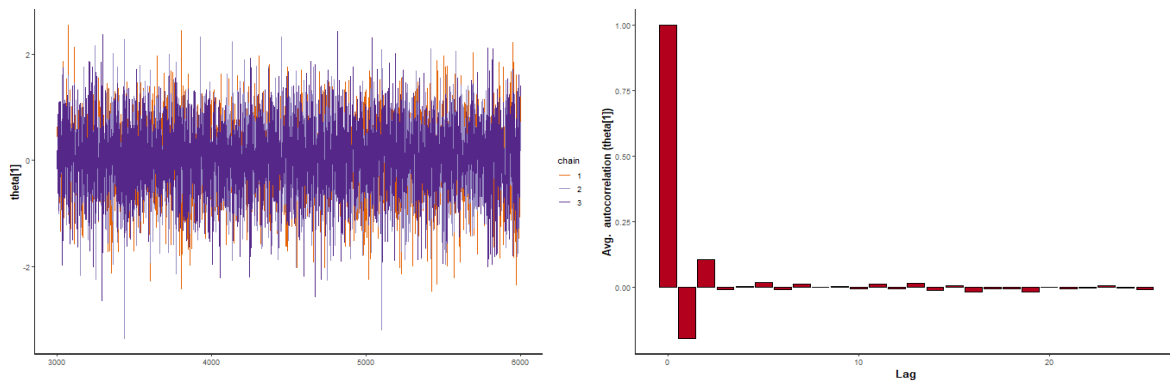
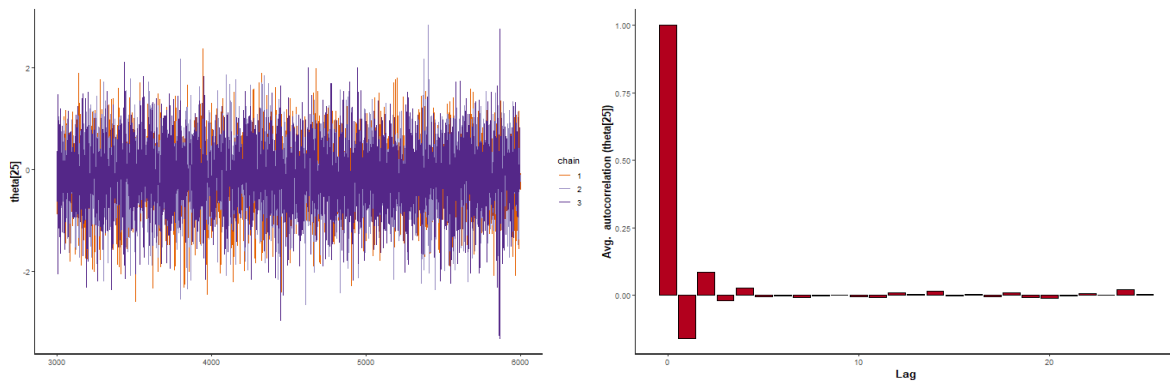


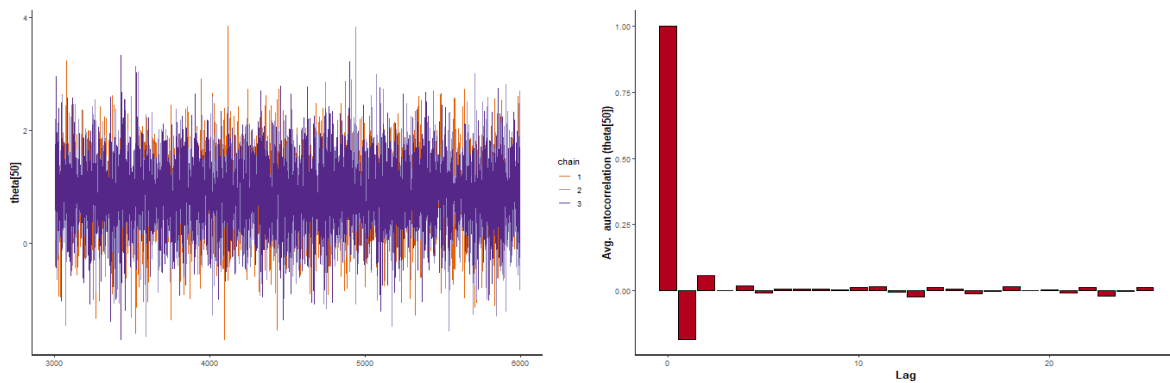
Figura A.9: Gráfica ACF del parámetro de dificultad del subrasgo *Fuentes de Financiamiento*.



(a) Gráfico de traza para la habilidad del subrasgo *Fuentes de Financiamiento* en la organización 1. (b) Gráfica ACF para la habilidad del subrasgo *Fuentes de Financiamiento* en la organización 1.



(c) Gráfico de traza para la habilidad del subrasgo *Fuentes de Financiamiento* en la organización 25. (d) Gráfica ACF para la habilidad del subrasgo *Fuentes de Financiamiento* en la organización 25.



(e) Gráfico de traza para la habilidad del subrasgo *Fuentes de Financiamiento* en la organización 50. (f) Gráfica ACF para la habilidad del subrasgo *Fuentes de Financiamiento* en la organización 50.

Figura A.10: Gráficos de traza y gráficos ACF de tres organizaciones de la muestra con el subrasgo *Fuentes de Financiamiento*.

Actividades

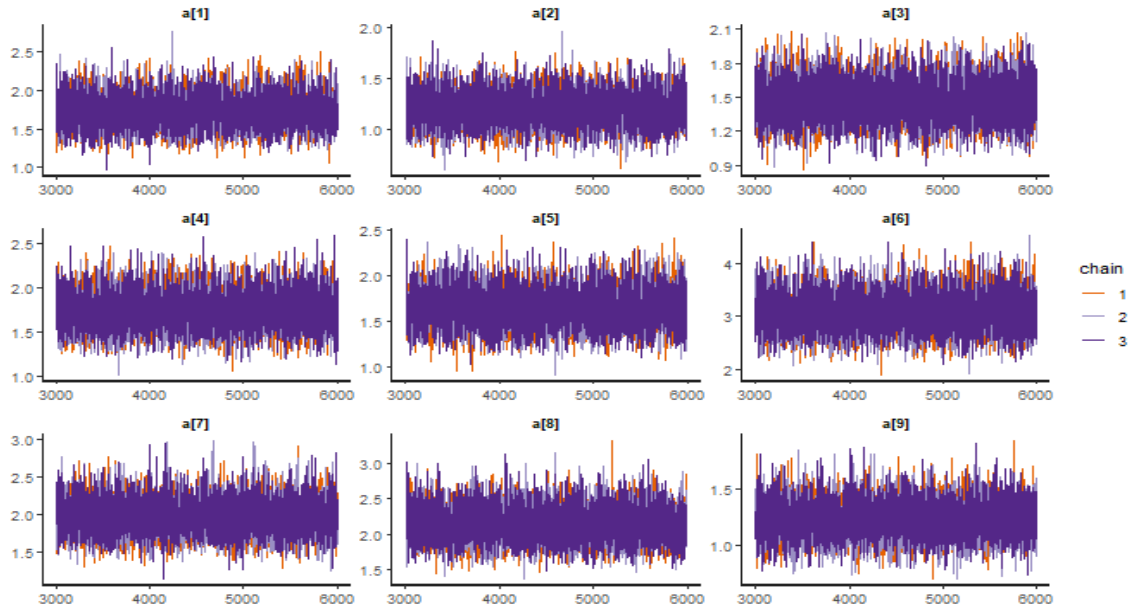


Figura A.11: Gráficos de traza del parámetro de discriminación del subrasgo *Actividades*.

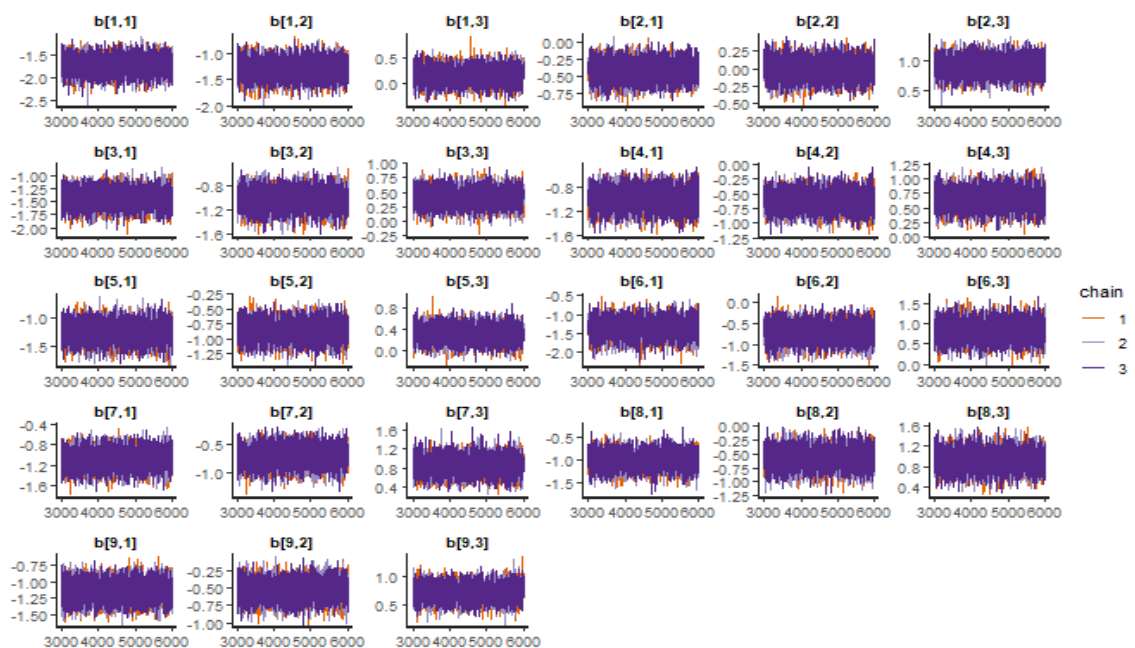


Figura A.12: Gráficos de traza del parámetro de dificultad del subrasgo *Actividades* con tres grados de libertad.

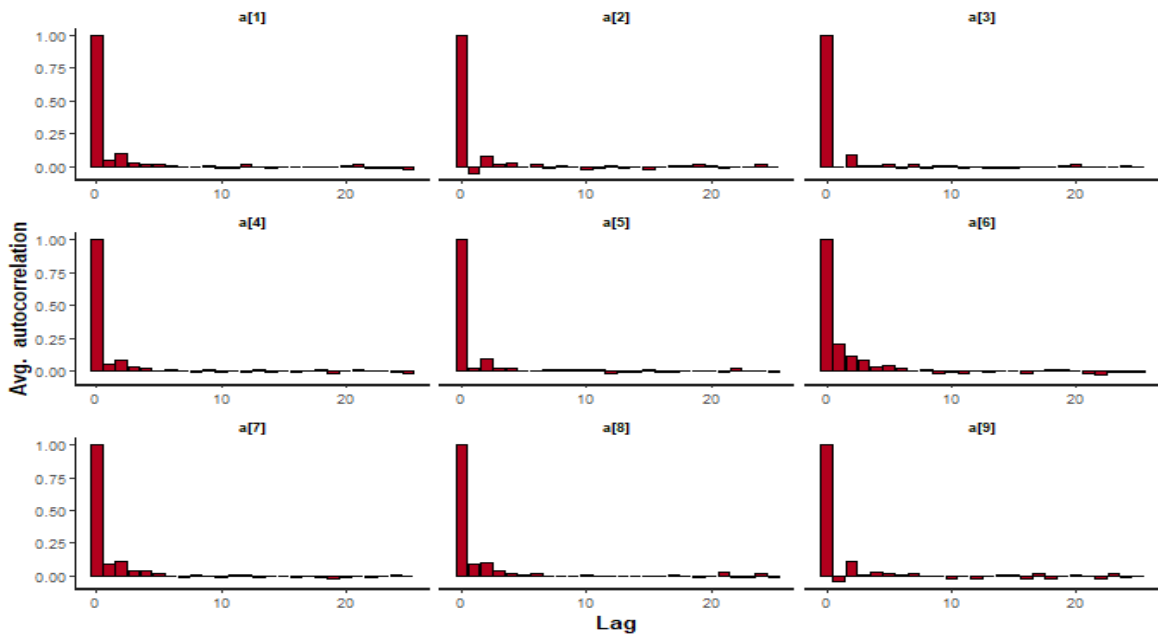


Figura A.13: Gráfica ACF del parámetro de discriminación del subrasgo *Actividades*.

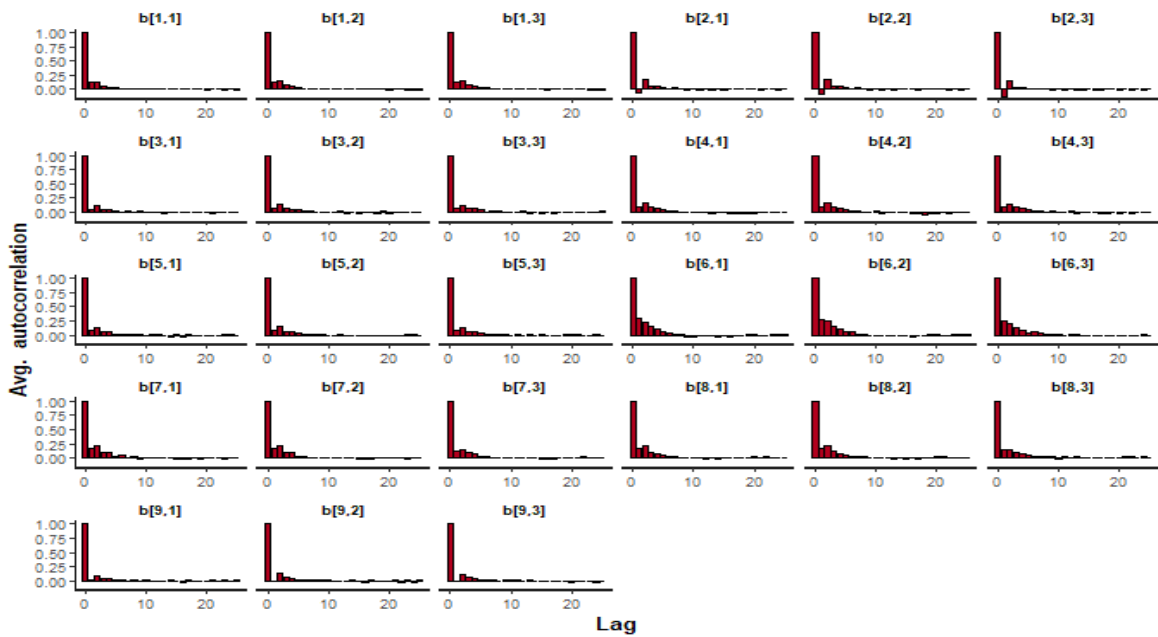
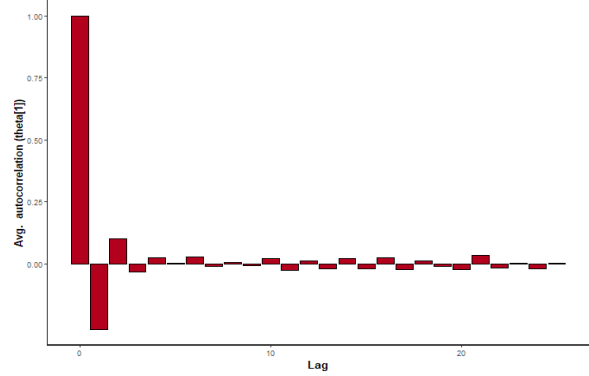
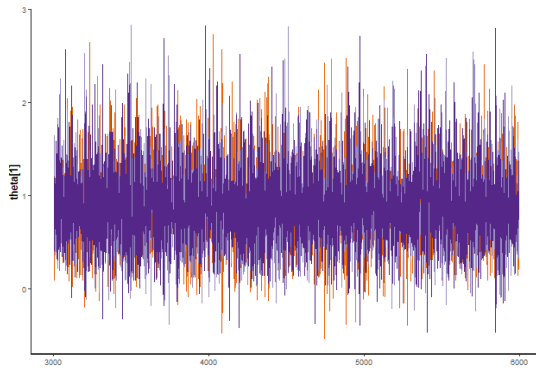
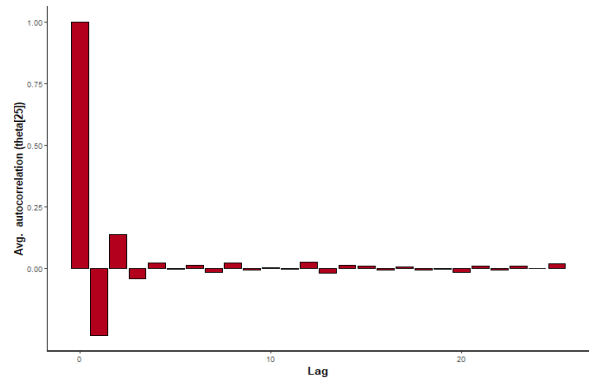
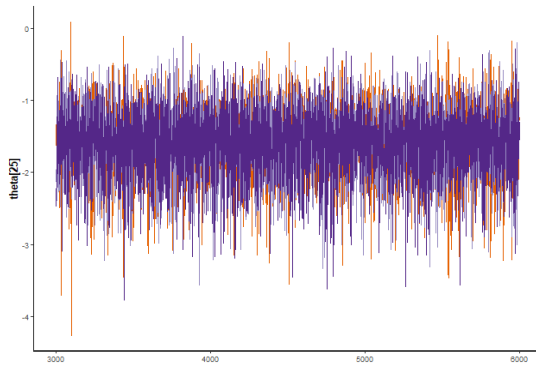


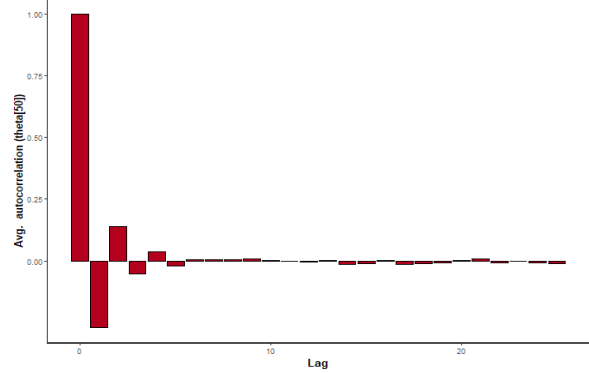
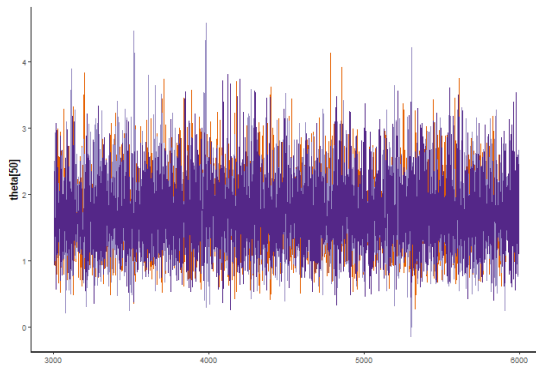
Figura A.14: Gráfica ACF del parámetro de dificultad del subrasgo *Actividades*.



(a) Gráfico de traza para la habilidad del subrasgo *Actividades* en la organización 1. (b) Gráfica ACF para la habilidad del subrasgo *Actividades* en la organización 1.



(c) Gráfico de traza para la habilidad del subrasgo *Actividades* en la organización 25. (d) Gráfica ACF para la habilidad del subrasgo *Actividades* en la organización 25.



(e) Gráfico de traza para la habilidad del subrasgo *Actividades* en la organización 50. (f) Gráfica ACF para la habilidad del subrasgo *Actividades* en la organización 50.

Figura A.15: Gráficos de traza y gráficos ACF de tres organizaciones de la muestra con el subrasgo *Actividades*.

Objetivos

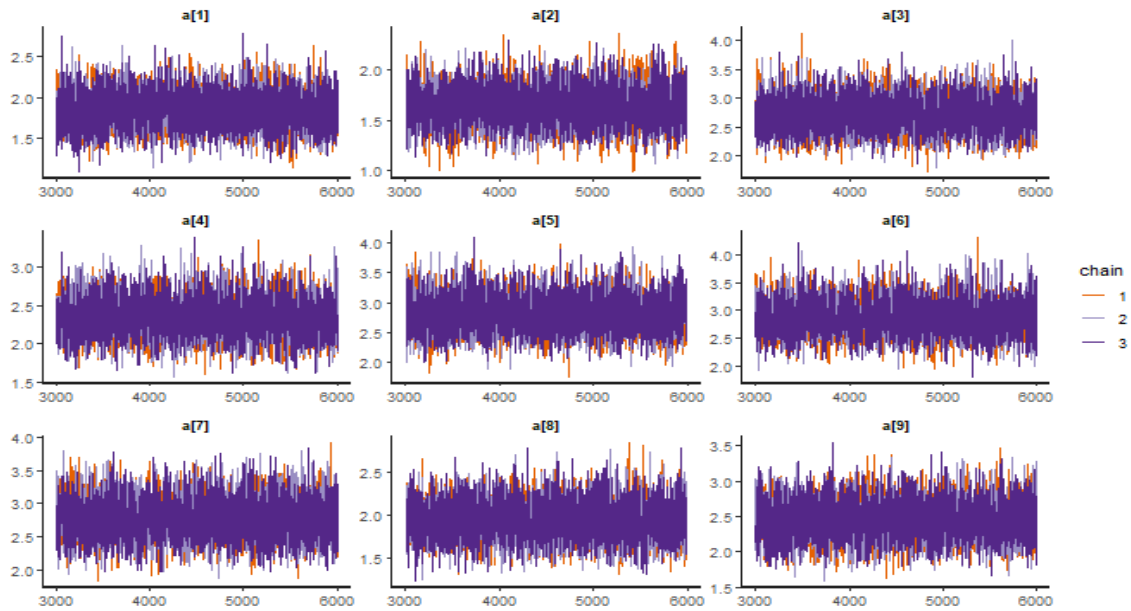


Figura A.16: Gráficos de traza del parámetro de discriminación del subrasgo *Objetivos*.

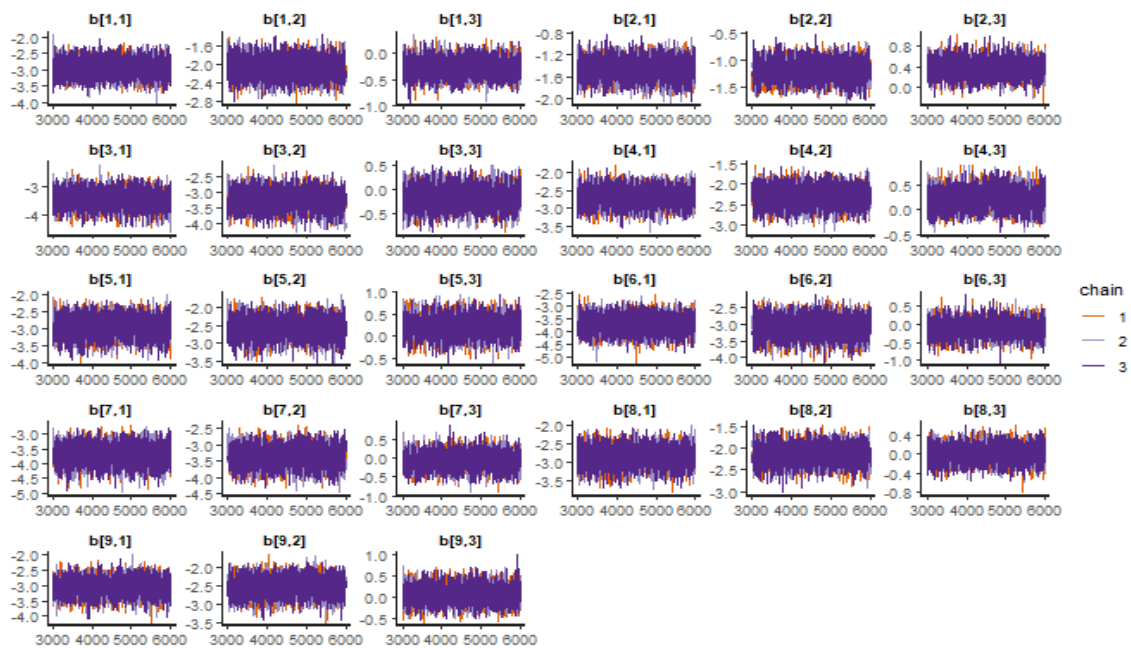


Figura A.17: Gráficos de traza del parámetro de dificultad del subrasgo *Objetivos* con tres grados de libertad.

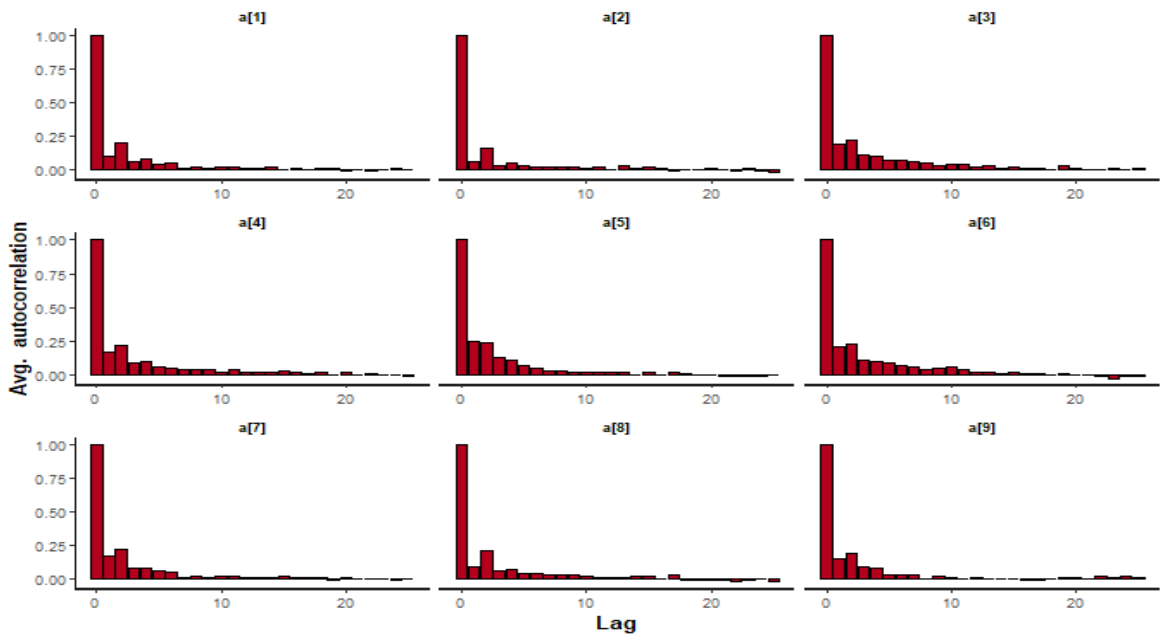


Figura A.18: Gráfica ACF del parámetro de discriminación del subrasgo *Objetivos*.

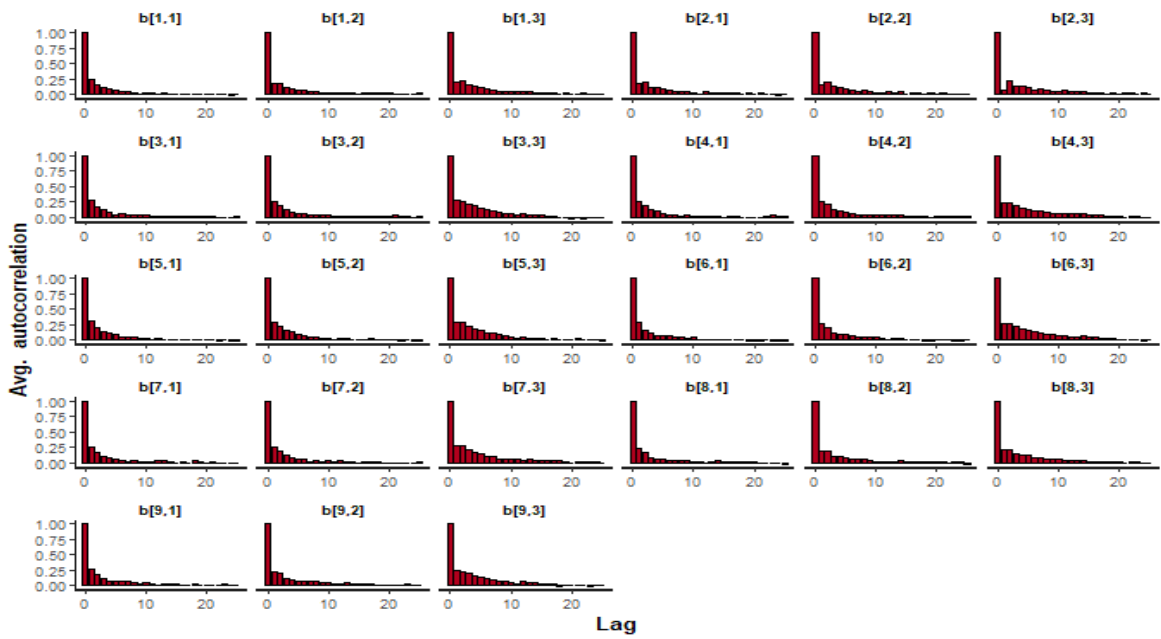
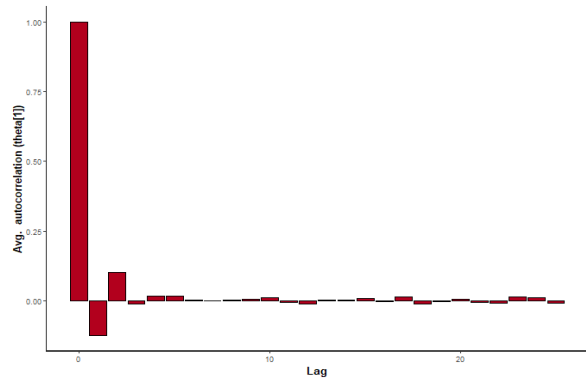
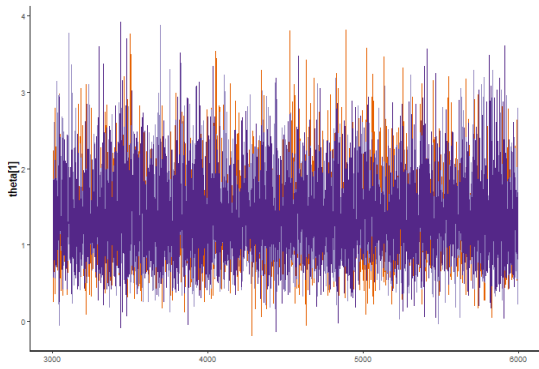
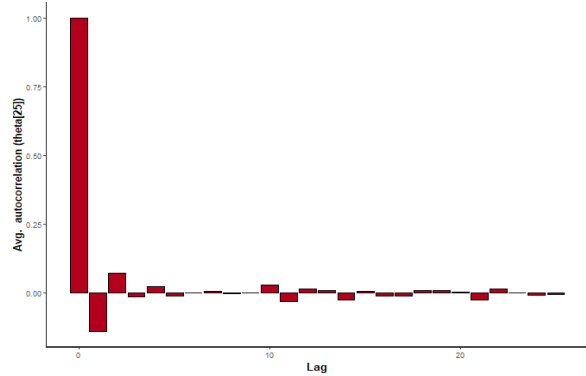
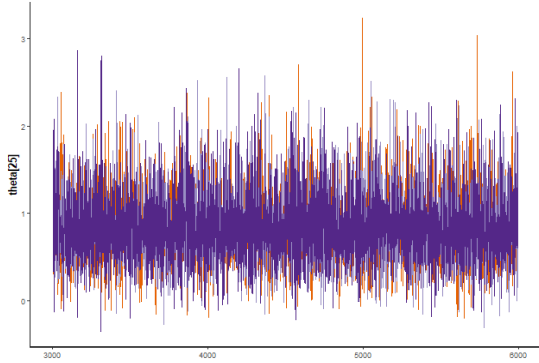


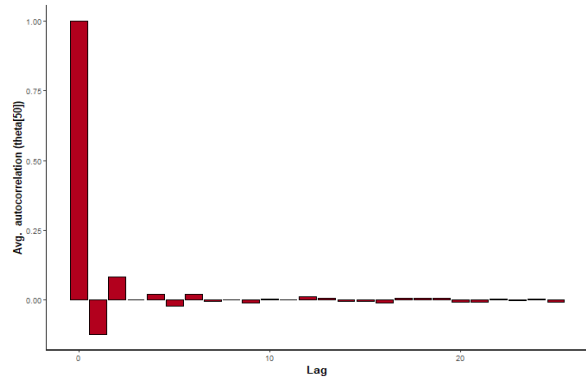
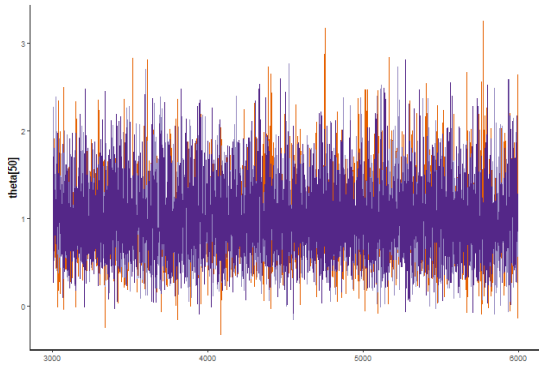
Figura A.19: Gráfica ACF del parámetro de dificultad del subrasgo *Objetivos*.



(a) Gráfico de traza para la habilidad del subrasgo *Objetivos* en la organización 1. (b) Gráfica ACF para la habilidad del subrasgo *Objetivos* en la organización 1.



(c) Gráfico de traza para la habilidad del subrasgo *Objetivos* en la organización 25. (d) Gráfica ACF para la habilidad del subrasgo *Objetivos* en la organización 25.



(e) Gráfico de traza para la habilidad del subrasgo *Objetivos* en la organización 50. (f) Gráfica ACF para la habilidad del subrasgo *Objetivos* en la organización 50.

Figura A.20: Gráficos de traza y gráficos ACF de tres organizaciones de la muestra con el subrasgo *Objetivos*.

Resultados

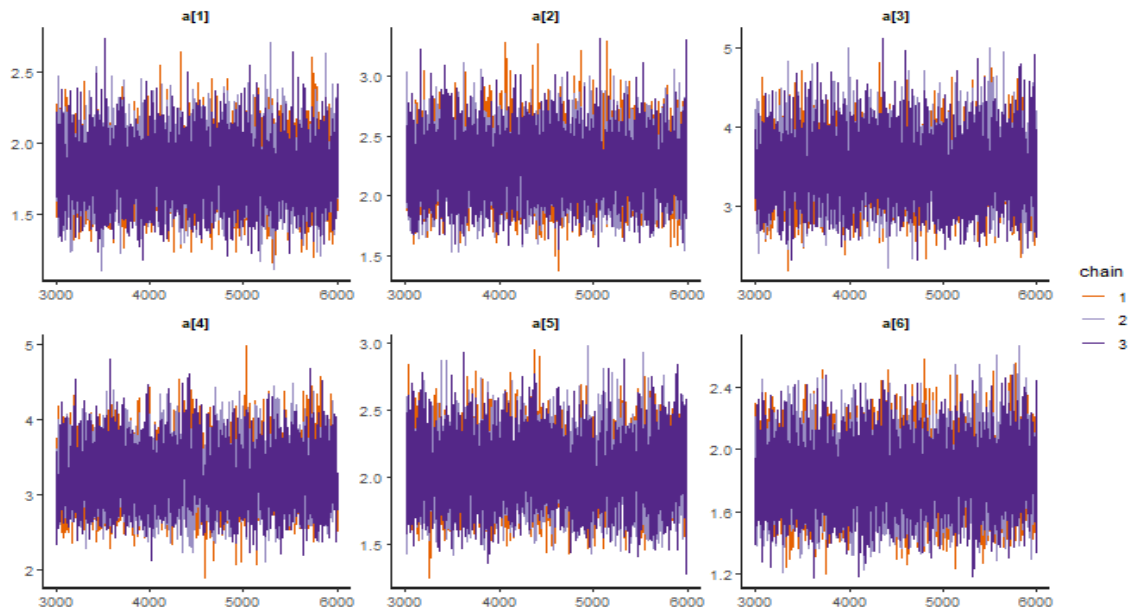


Figura A.21: Gráficos de traza del parámetro de discriminación del subrasgo *Resultados*.

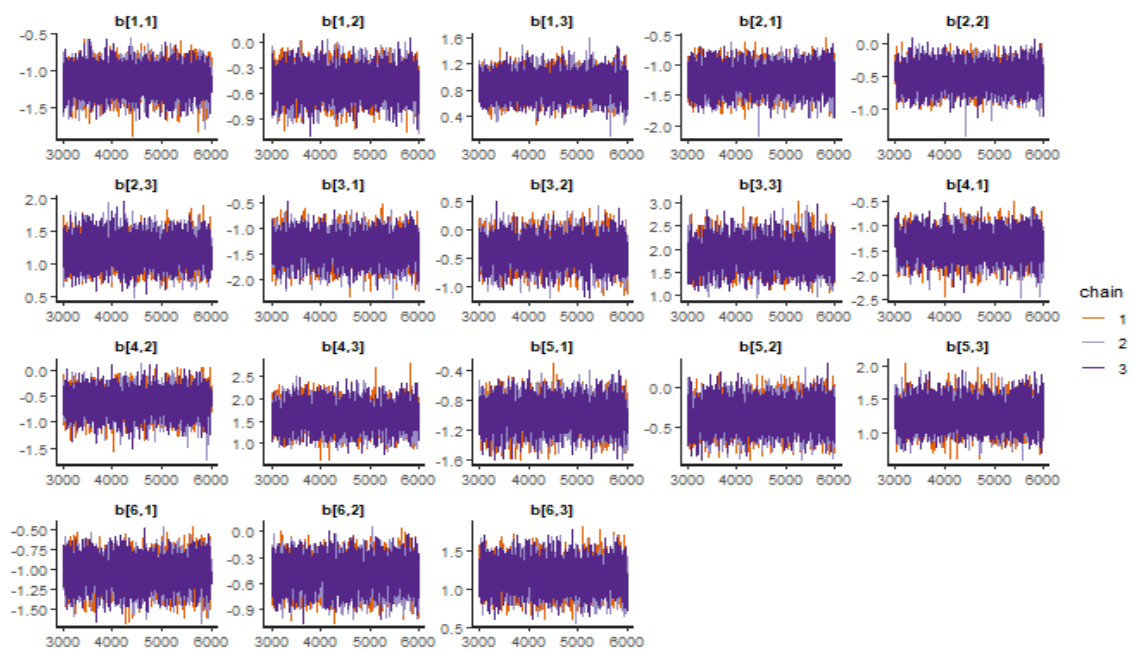


Figura A.22: Gráficos de traza del parámetro de dificultad del subrasgo *Resultados* con tres grados de libertad.

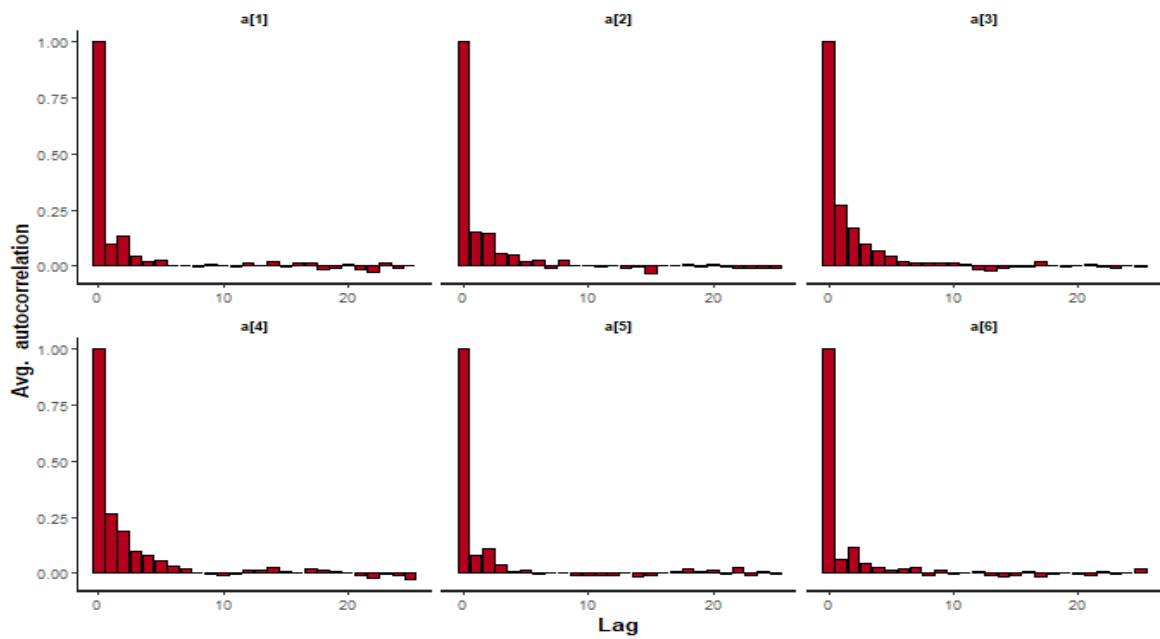


Figura A.23: Gráfica ACF del parámetro de discriminación del subrasgo *Resultados*.

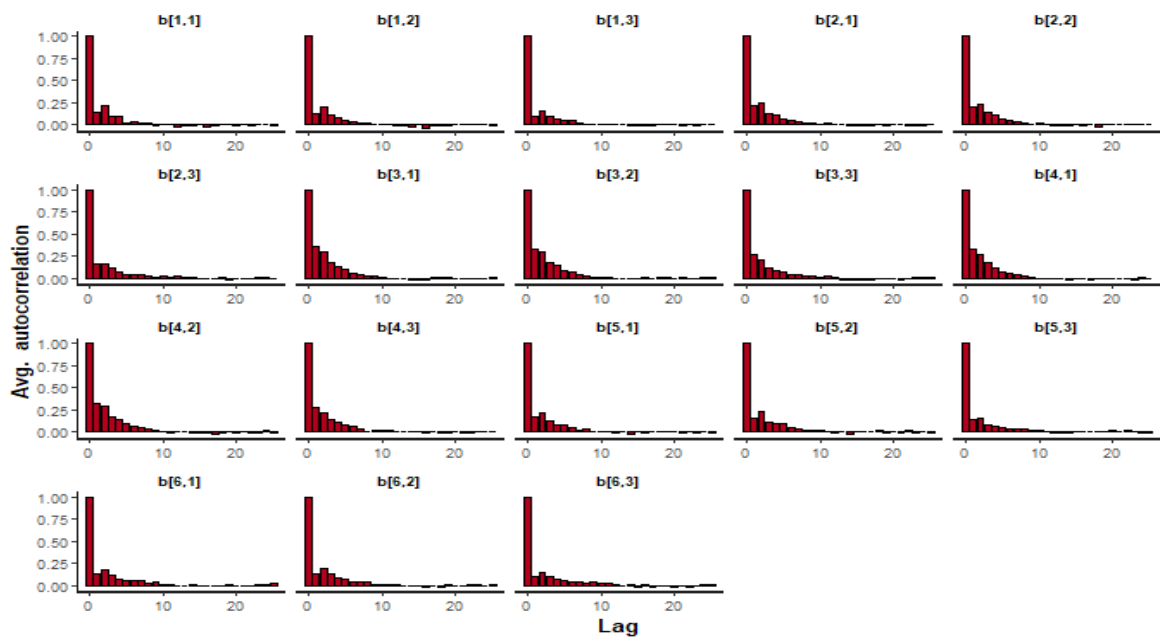
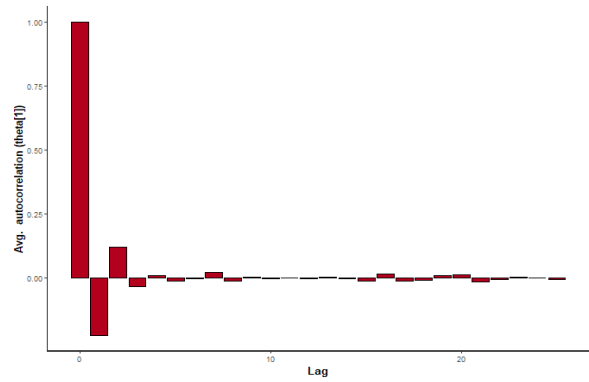
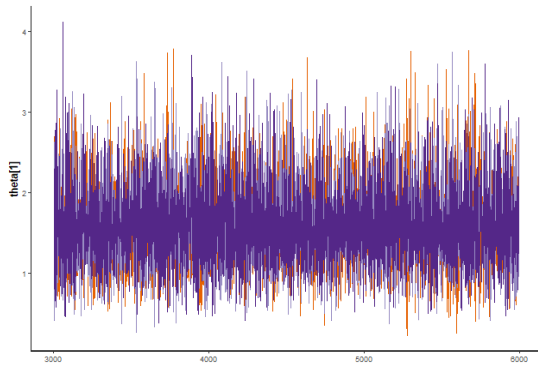
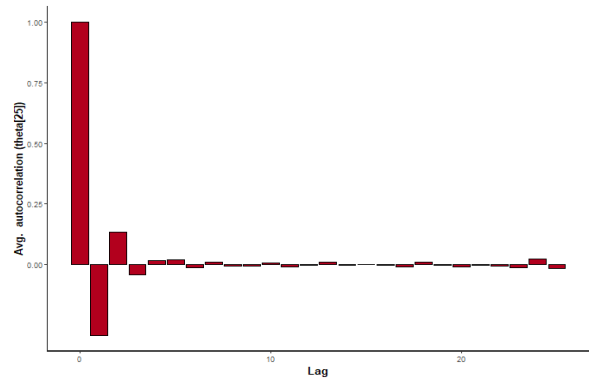
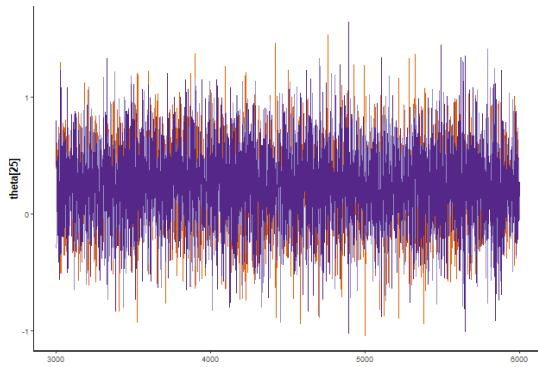


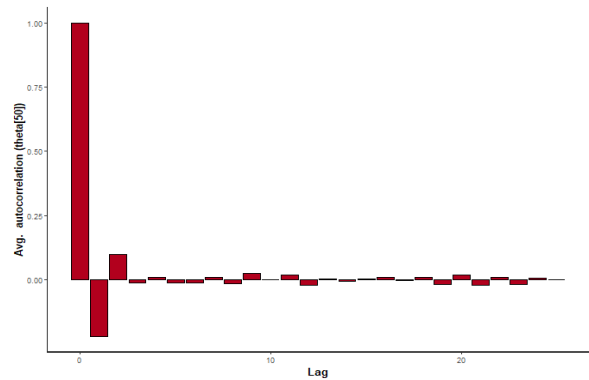
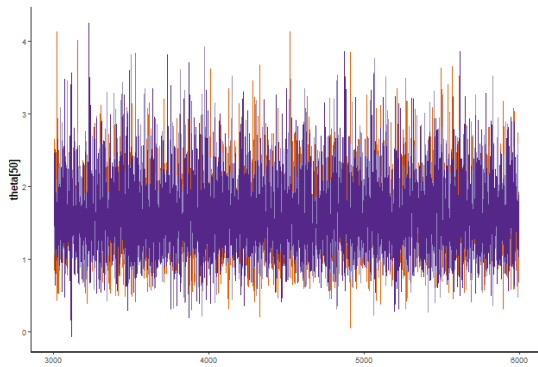
Figura A.24: Gráfica ACF del parámetro de dificultad del subrasgo *Resultados*.



(a) Gráfico de traza para la habilidad del subrasgo *Resultados* en la organización 1. (b) Gráfica ACF para la habilidad del subrasgo *Resultados* en la organización 1.



(c) Gráfico de traza para la habilidad del subrasgo *Resultados* en la organización 25. (d) Gráfica ACF para la habilidad del subrasgo *Resultados* en la organización 25.



(e) Gráfico de traza para la habilidad del subrasgo *Resultados* en la organización 50. (f) Gráfica ACF para la habilidad del subrasgo *Resultados* en la organización 50.

Figura A.25: Gráficos de traza y gráficos ACF de tres organizaciones de la muestra con el subrasgo *Resultados*.

Impactos

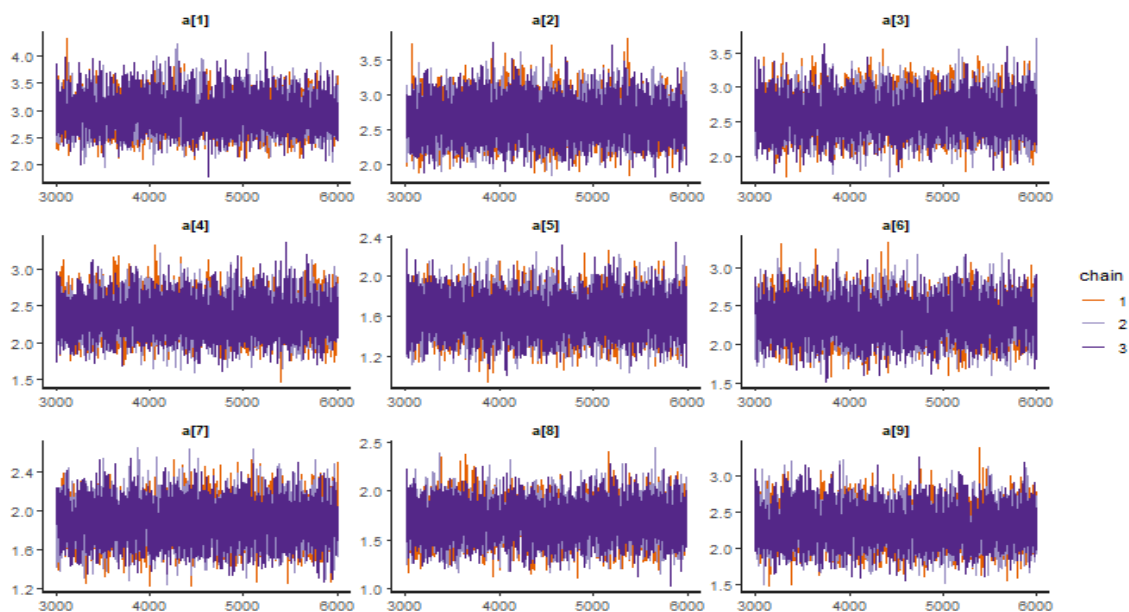


Figura A.26: Gráficos de traza del parámetro de discriminación del subrasgo *Impactos*.

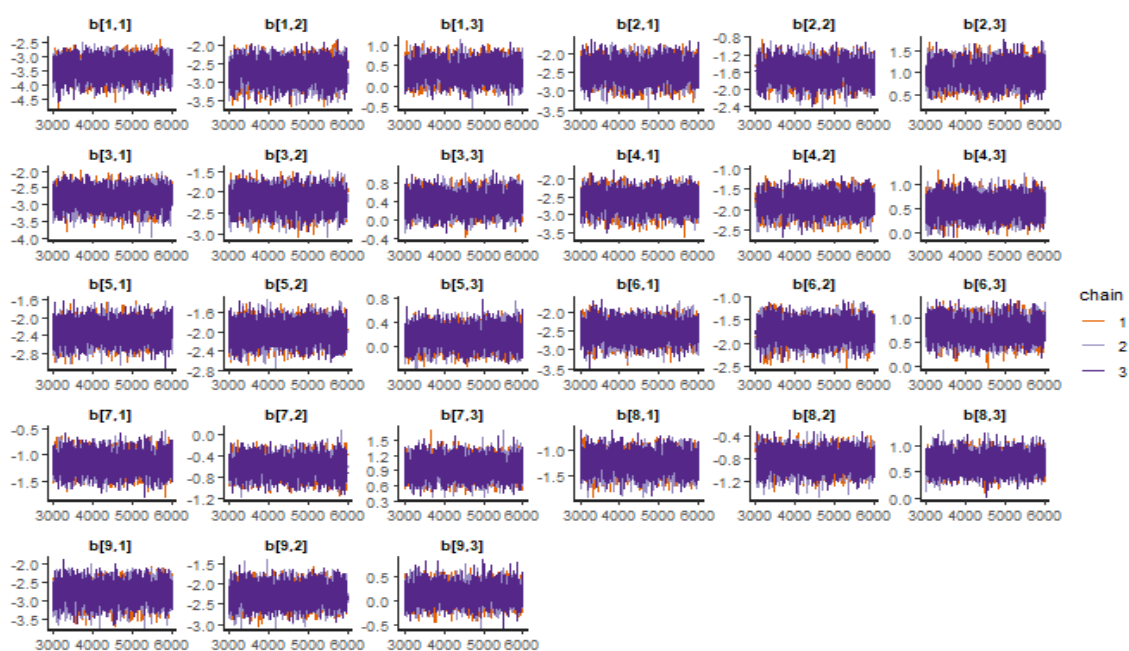


Figura A.27: Gráficos de traza del parámetro de dificultad del subrasgo *Impactos* con tres grados de libertad.

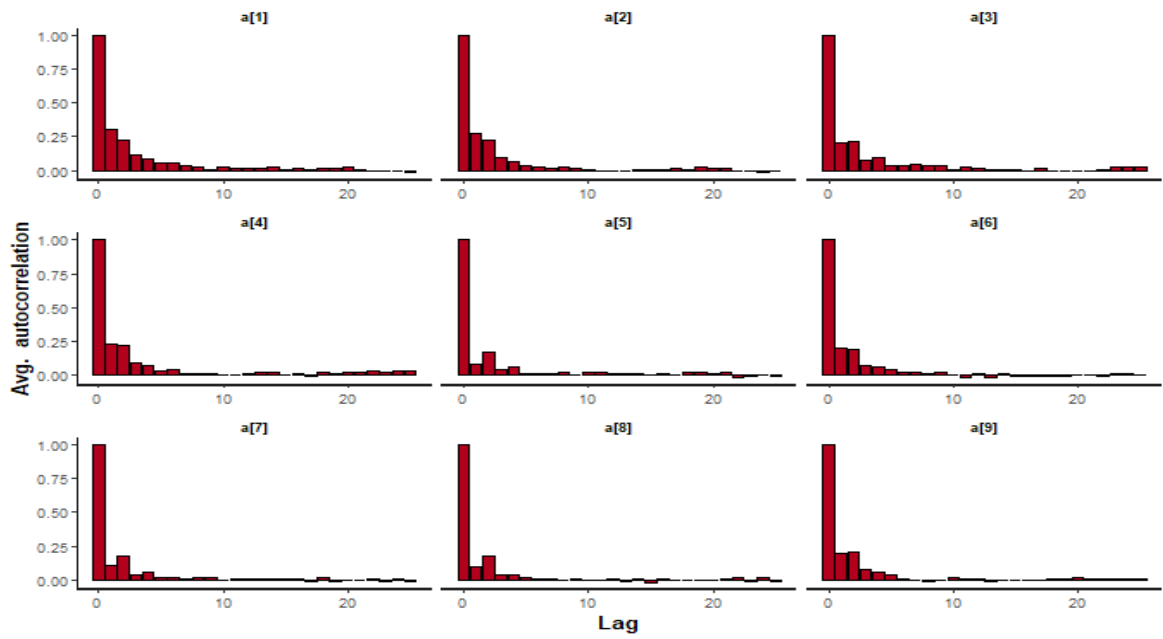


Figura A.28: Gráfica ACF del parámetro de discriminación del subrasgo *Impactos*.

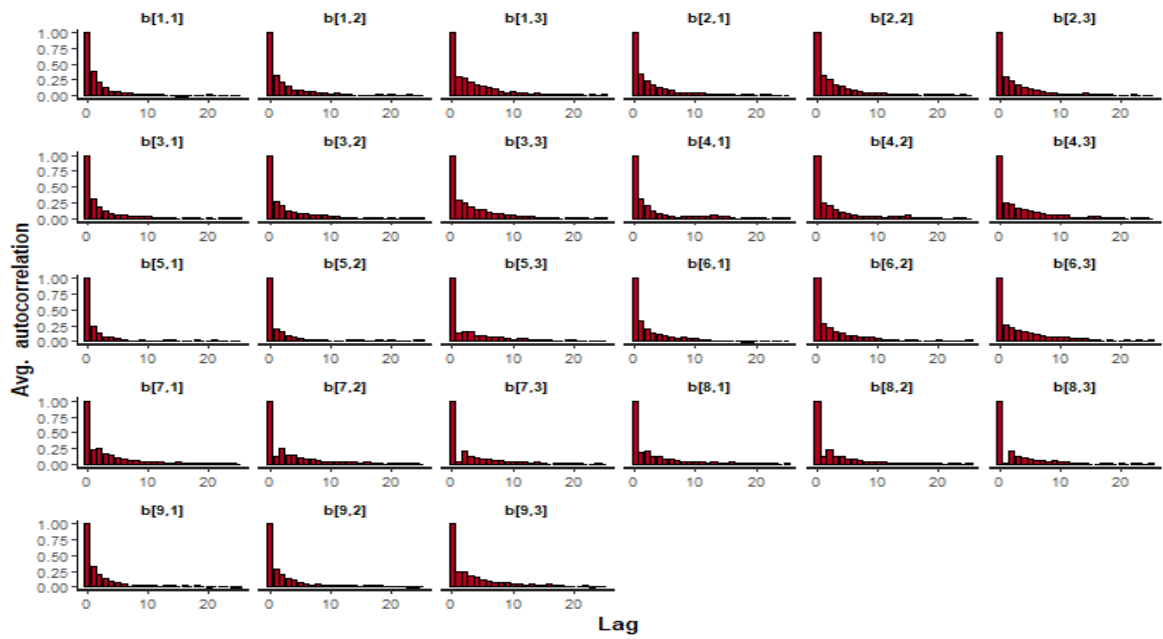
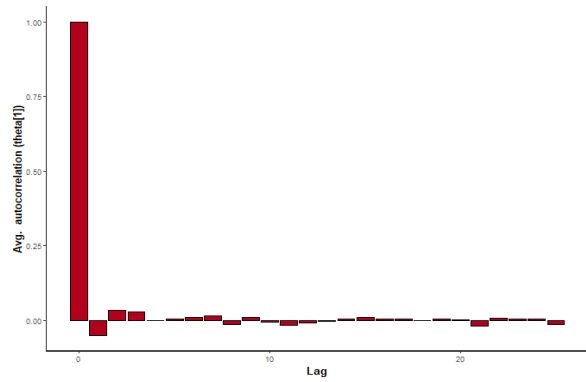
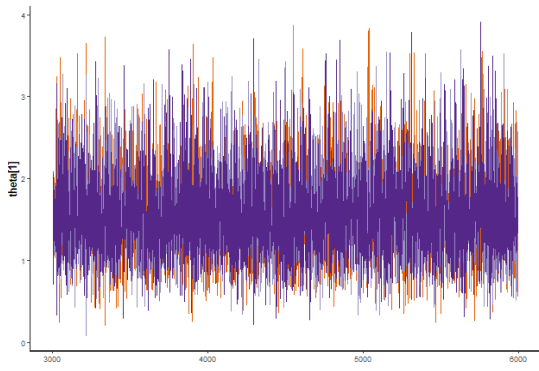
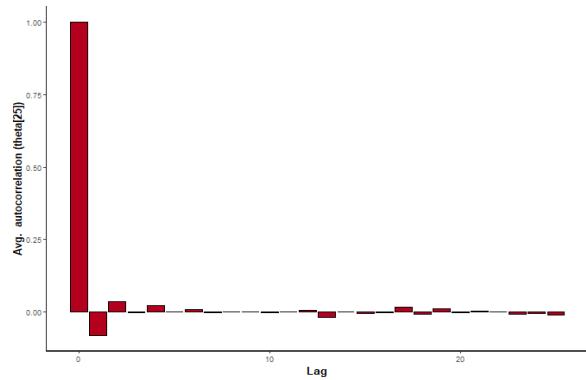
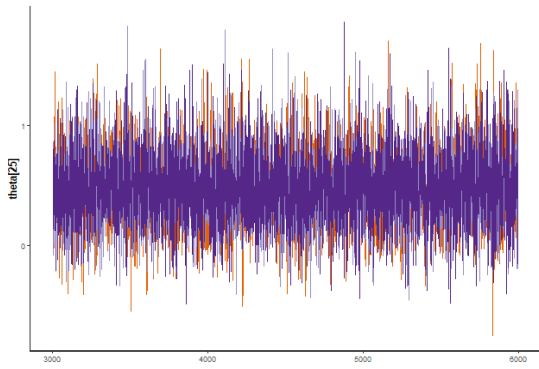


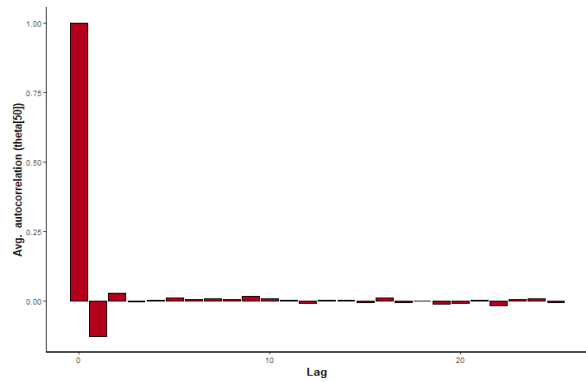
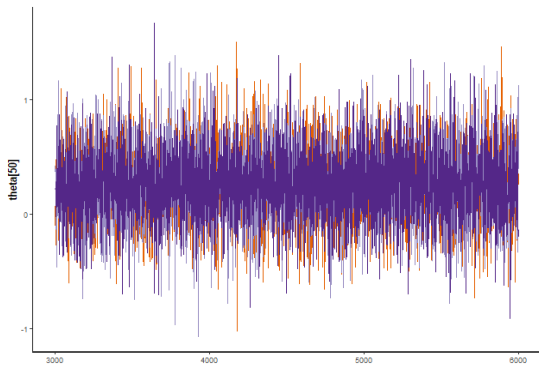
Figura A.29: Gráfica ACF del parámetro de dificultad del subrasgo *Impactos*.



(a) Gráfico de traza para la habilidad del subrasgo *Impactos* en la organización 1. (b) Gráfica ACF para la habilidad del subrasgo *Impactos* en la organización 1.



(c) Gráfico de traza para la habilidad del subrasgo *Impactos* en la organización 25. (d) Gráfica ACF para la habilidad del subrasgo *Impactos* en la organización 25.



(e) Gráfico de traza para la habilidad del subrasgo *Impactos* en la organización 50. (f) Gráfica ACF para la habilidad del subrasgo *Impactos* en la organización 50.

Figura A.30: Gráficos de traza y gráficos ACF de tres organizaciones de la muestra con el subrasgo *Impactos*.

A.4. Curvas Características de Categoría

Fuentes de Información

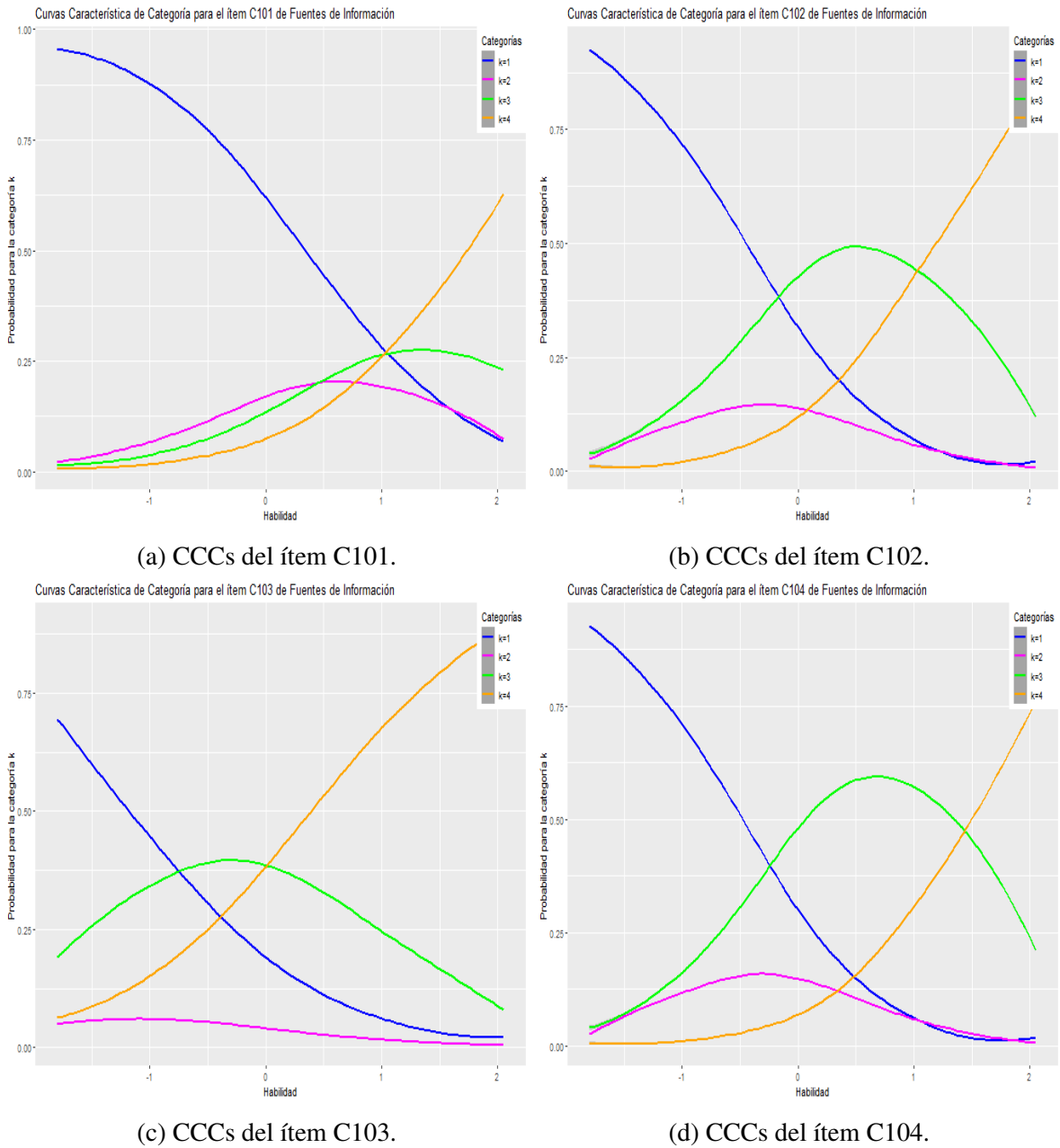
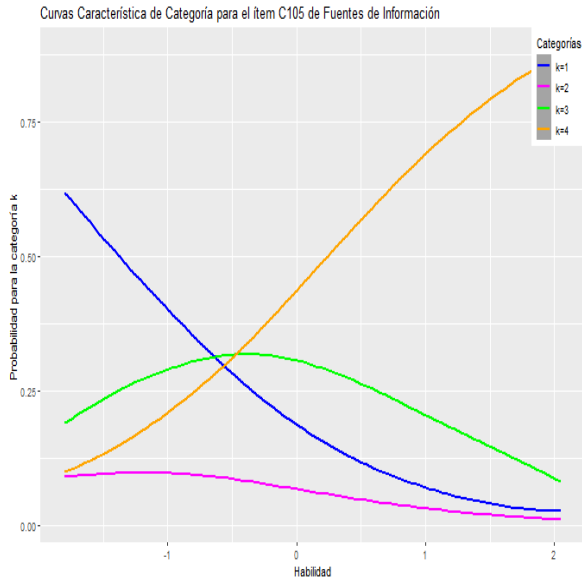
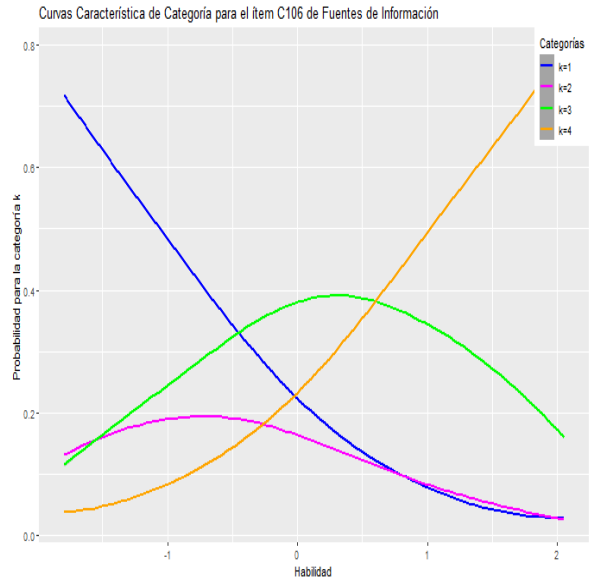


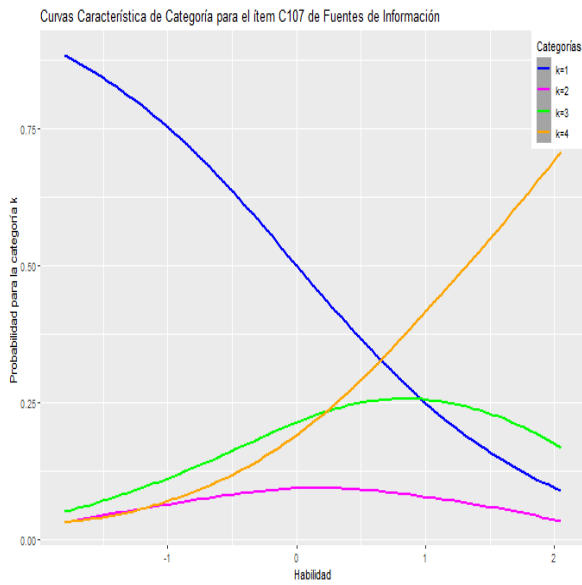
Figura A.31: Curvas Características de Categoría para los ítems del subrasgo *Fuentes de Información*.



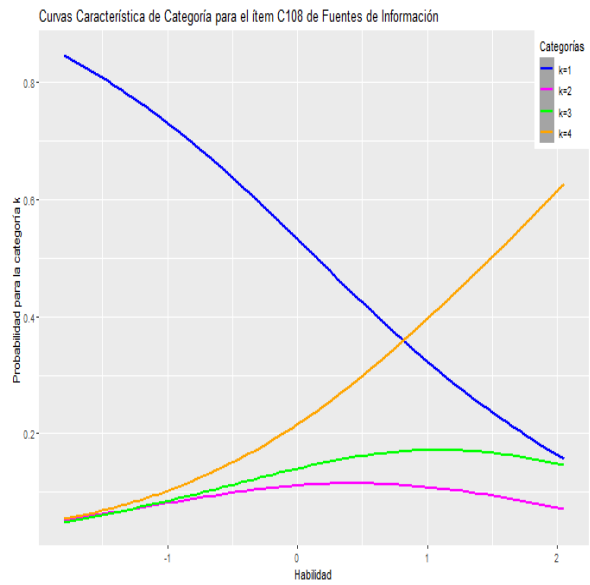
(a) CCCs del ítem C105.



(b) CCCs del ítem C106.



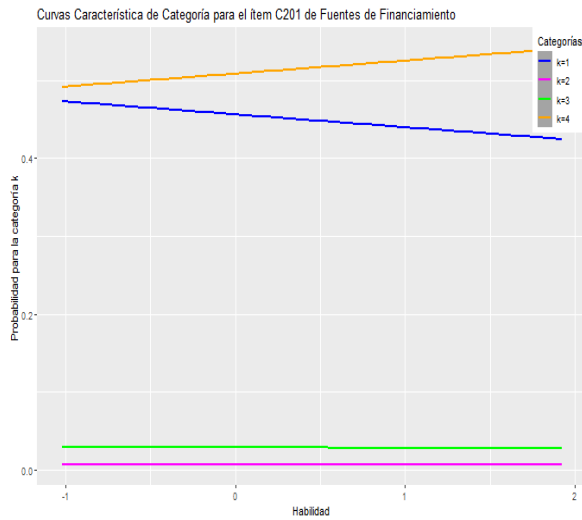
(c) CCCs del ítem C107.



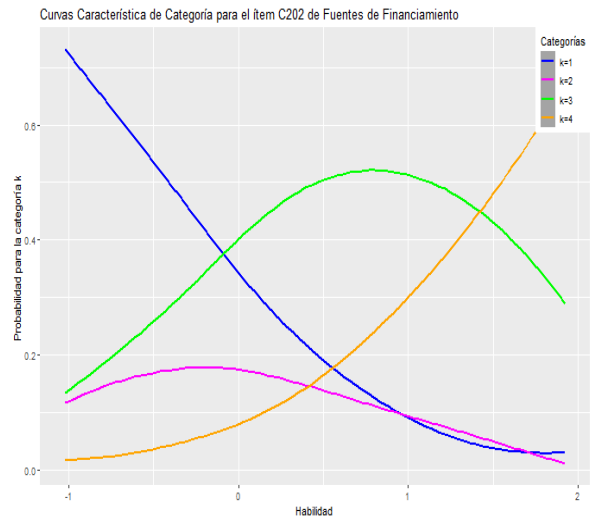
(d) CCCs del ítem C108.

Figura A.32: Curvas Características de Categoría para los ítems del subrasgo *Fuentes de Información*.

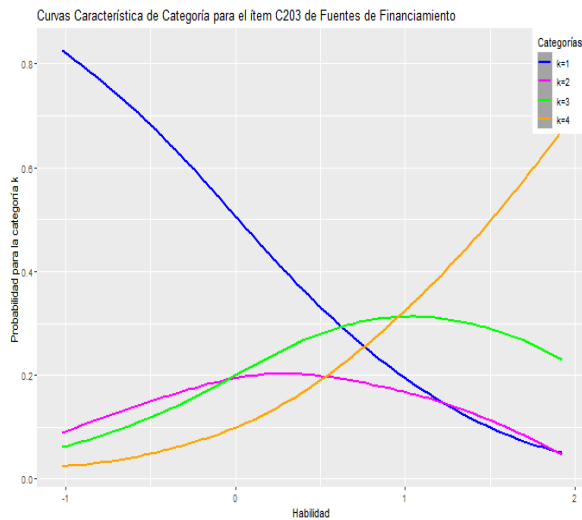
Fuentes de Financiamiento



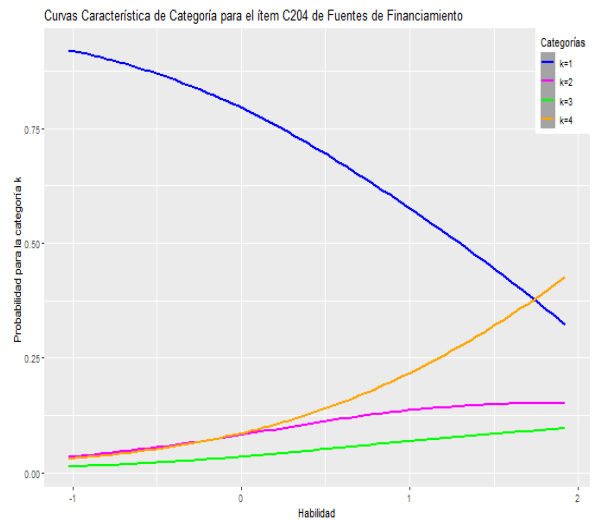
(a) CCCs del ítem C201.



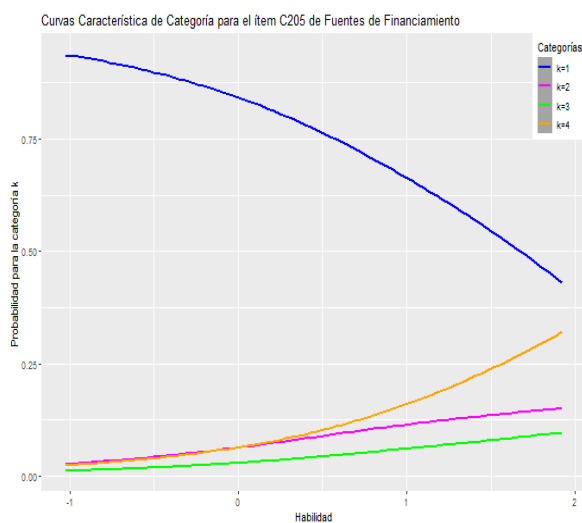
(b) CCCs del ítem C202.



(c) CCCs del ítem C203.



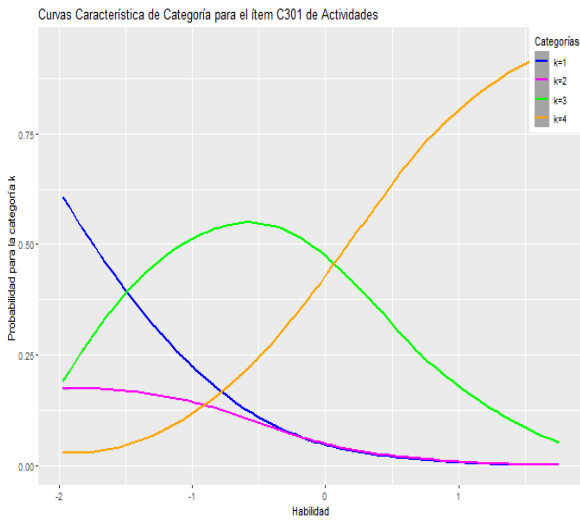
(d) CCCs del ítem C204.



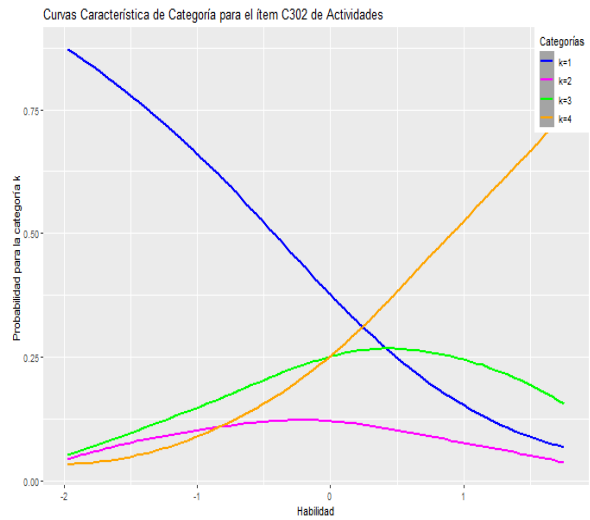
(e) CCCs del ítem C205.

Figura A.33: Curvas Características de Categoría para los ítems del subrasgo *Fuentes de Financiamiento*.

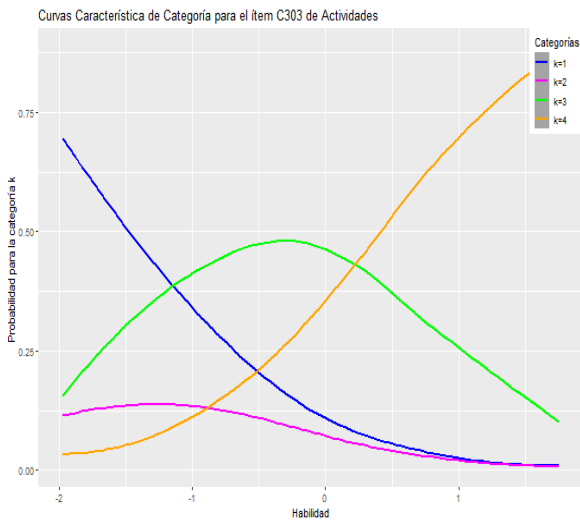
Actividades



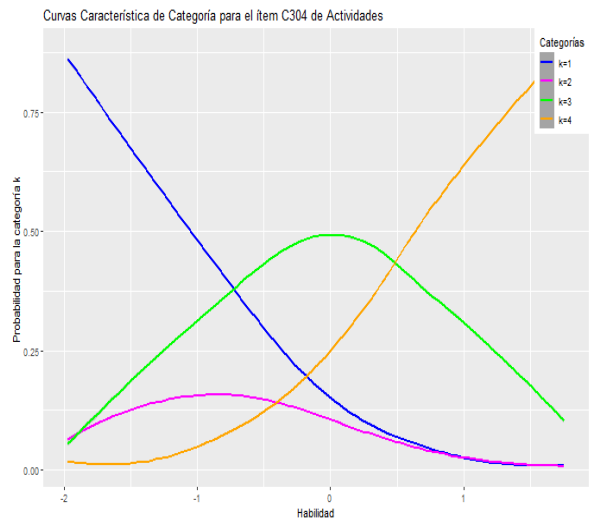
(a) CCCs del ítem C301.



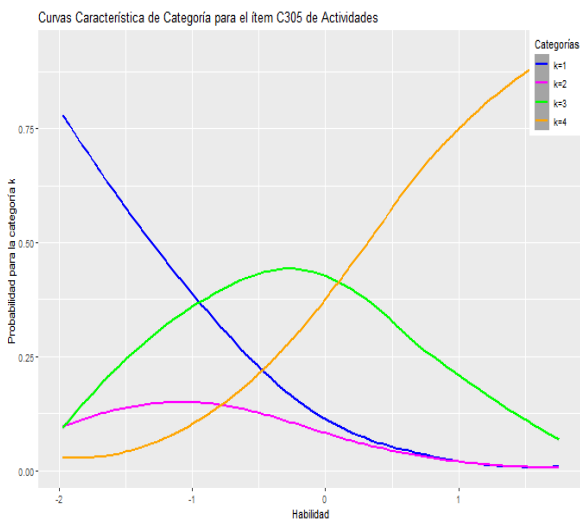
(b) CCCs del ítem C302.



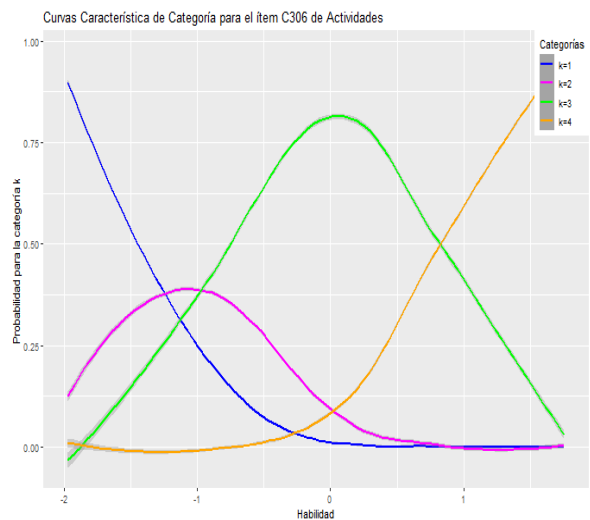
(c) CCCs del ítem C303.



(d) CCCs del ítem C304.

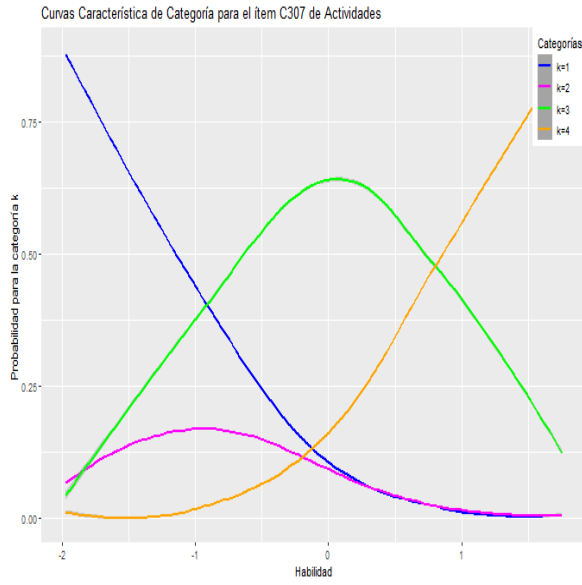


(e) CCCs del ítem C305.

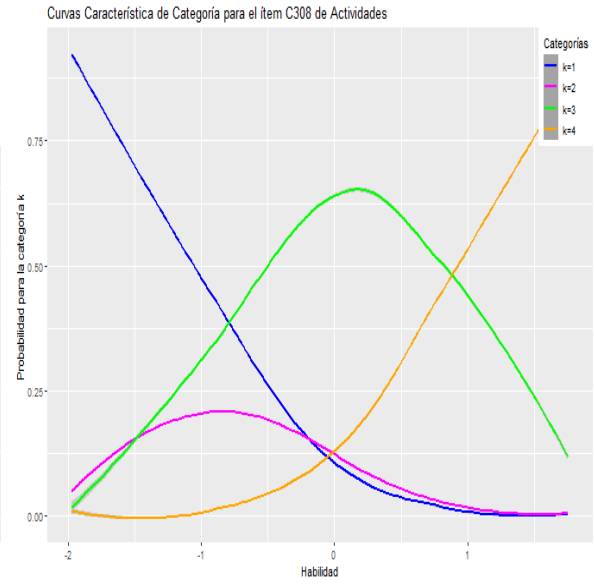


(f) CCCs del ítem C306.

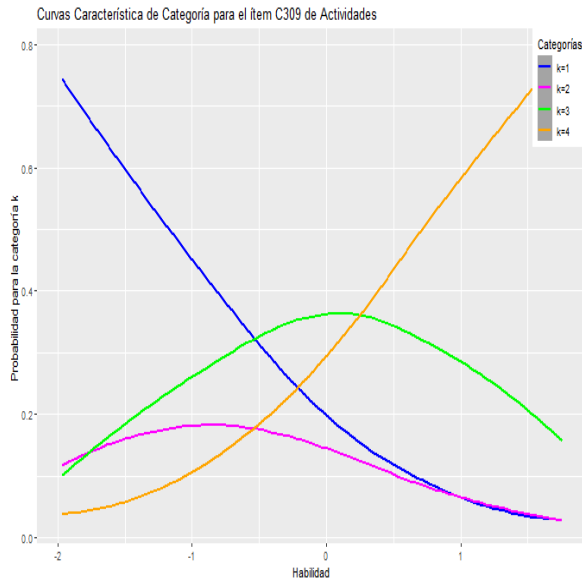
Figura A.34: Curvas Características de Categoría para los ítems del subrasgo *Actividades*.



(a) CCCs del ítem C307.



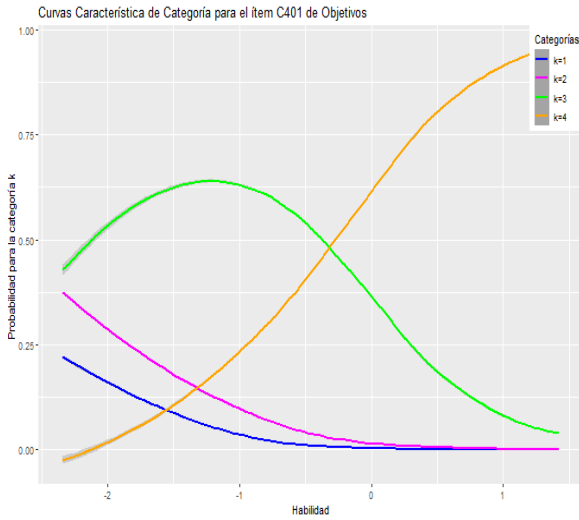
(b) CCCs del ítem C308.



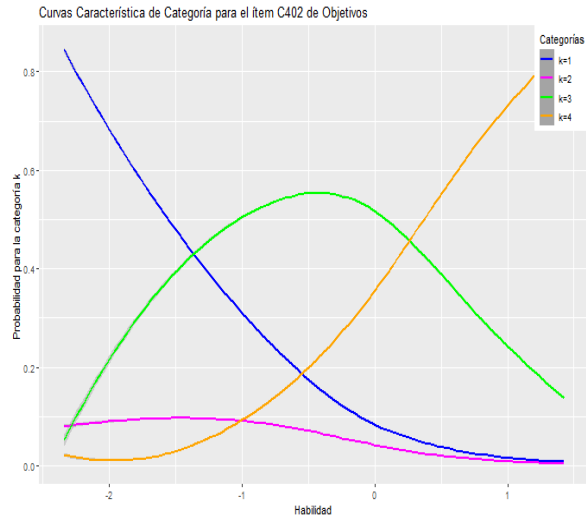
(c) CCCs del ítem C309.

Figura A.35: Curvas Características de Categoría para los ítems del subrasgo *Actividades*.

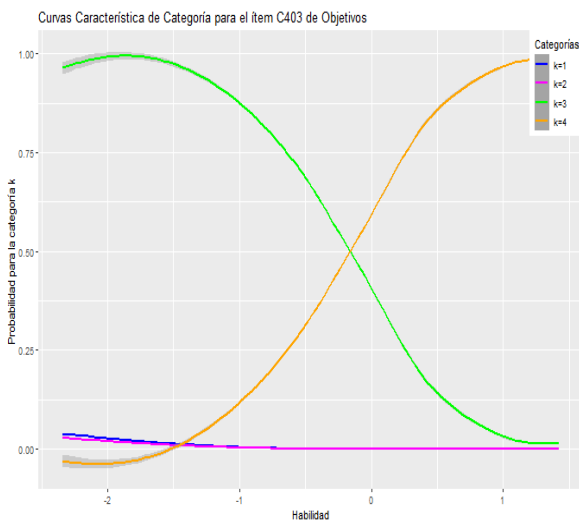
Objetivos



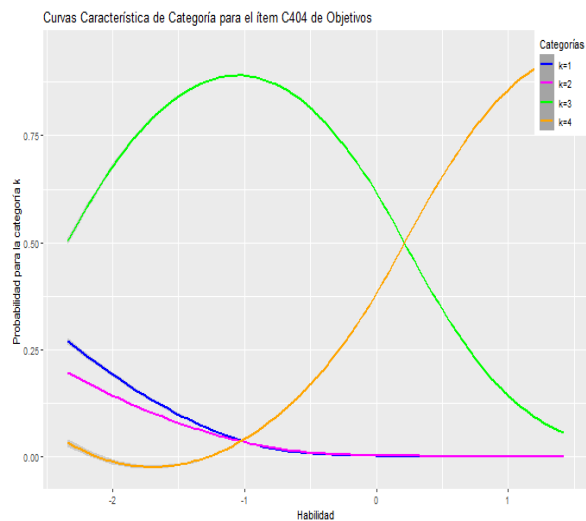
(a) CCCs del ítem C401.



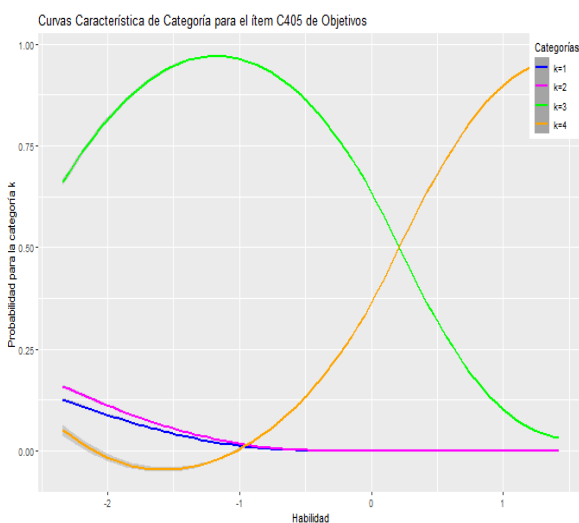
(b) CCCs del ítem C402.



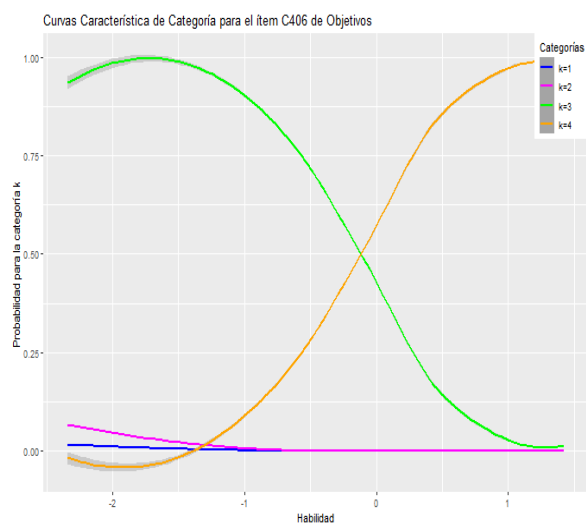
(c) CCCs del ítem C403.



(d) CCCs del ítem C404.

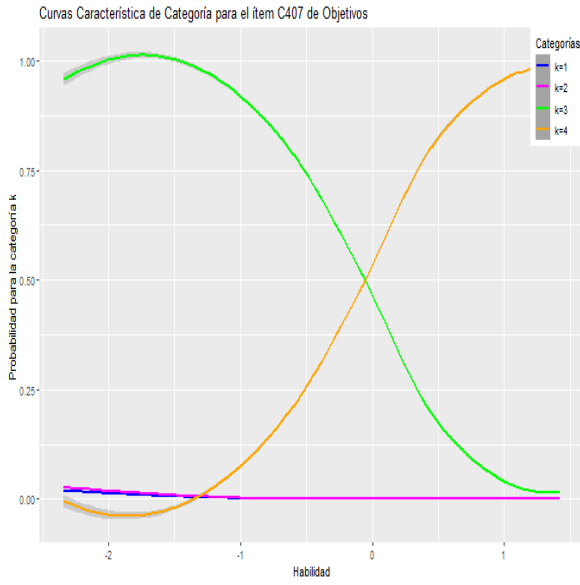


(e) CCCs del ítem C405.

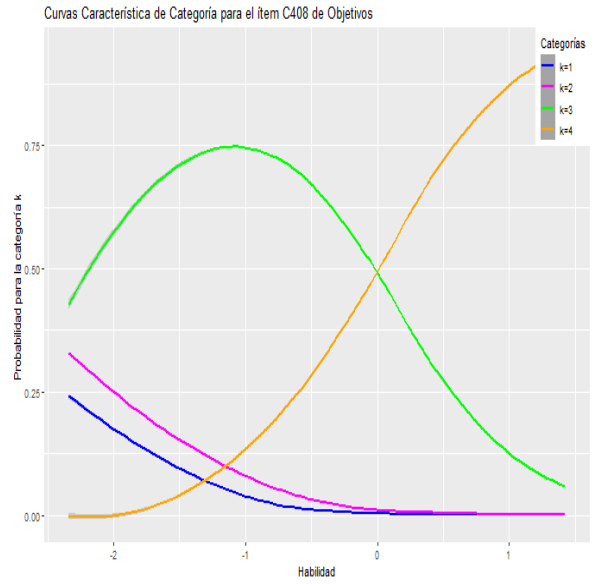


(f) CCCs del ítem C406.

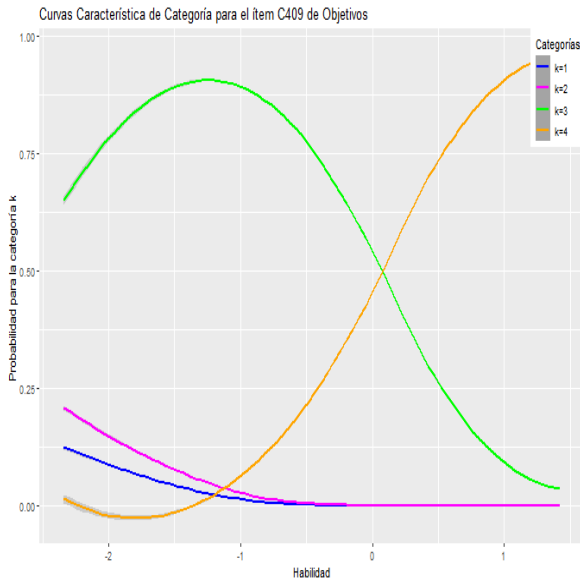
Figura A.36: Curvas Características de Categoría para los ítems del subrasgo *Objetivos*.



(a) CCCs del ítem C407.



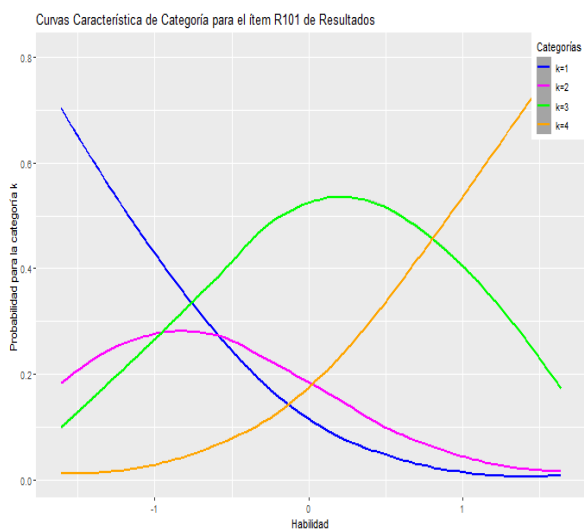
(b) CCCs del ítem C408.



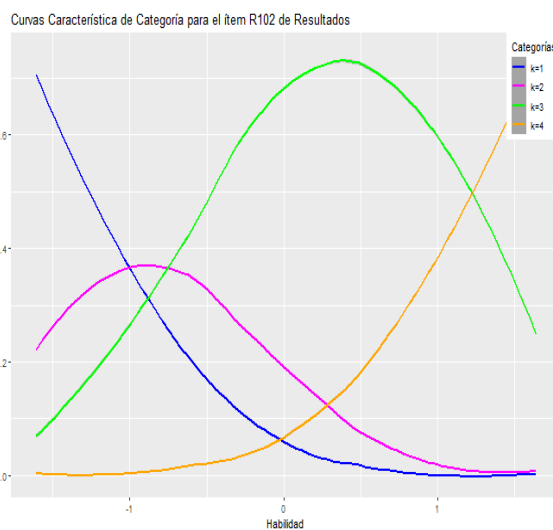
(c) CCCs del ítem C409.

Figura A.37: Curvas Características de Categoría para los ítems del subrasgo *Objetivos*.

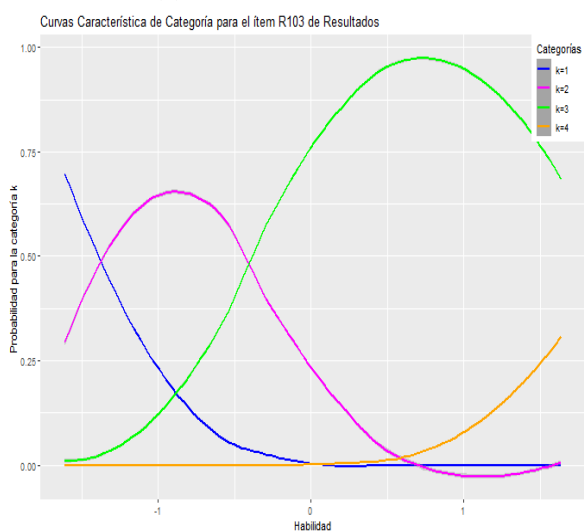
Resultados



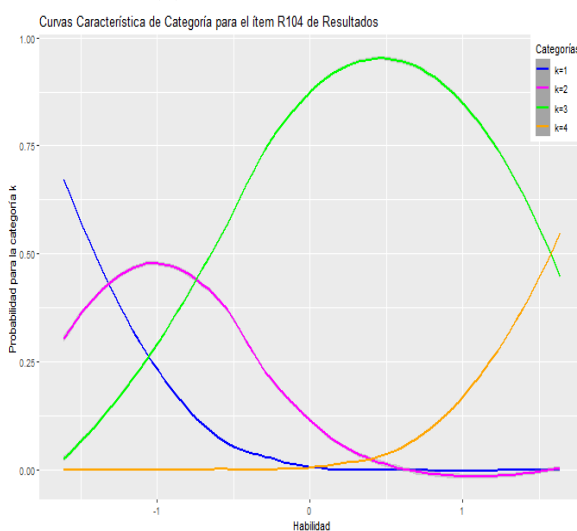
(a) CCCs del ítem R101.



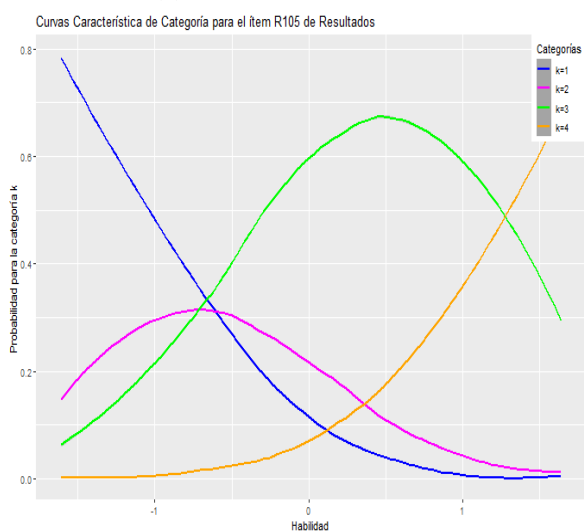
(b) CCCs del ítem R102.



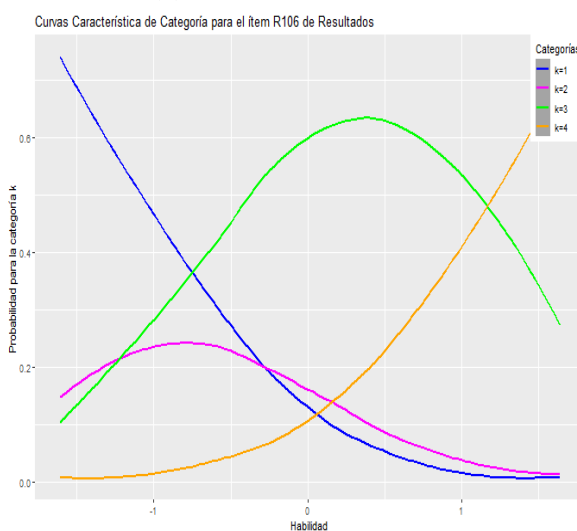
(c) CCCs del ítem R103.



(d) CCCs del ítem R104.



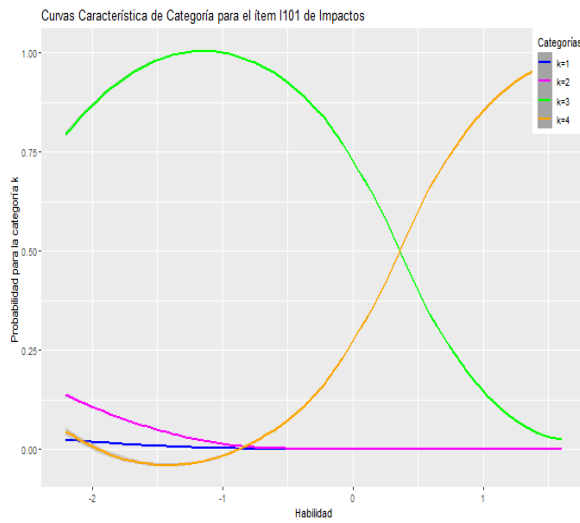
(e) CCCs del ítem R105.



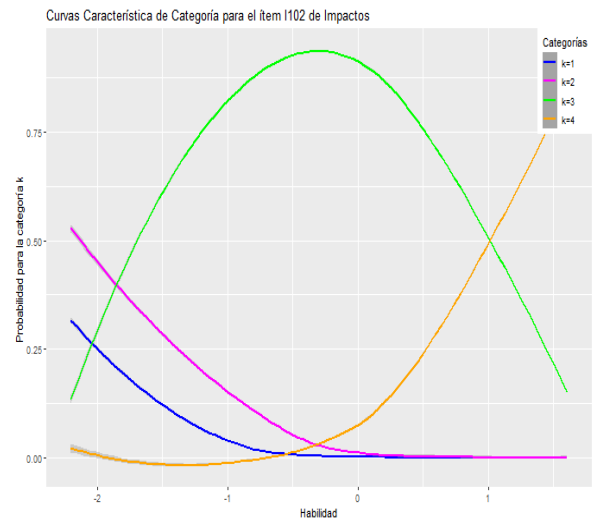
(f) CCCs del ítem R106.

Figura A.38: Curvas Características de Categoría para los ítems del dominio *Resultados*.

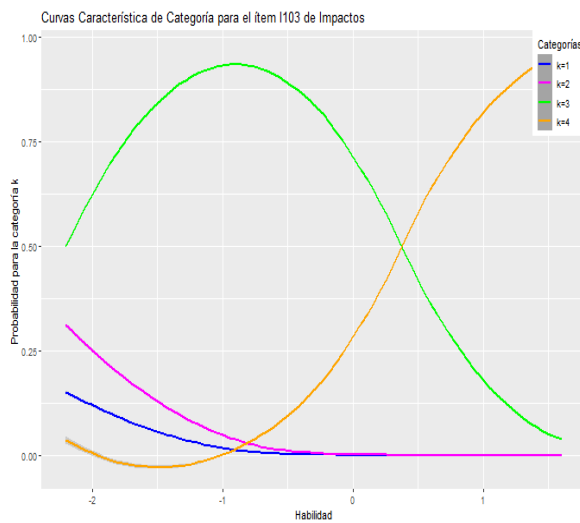
Impactos



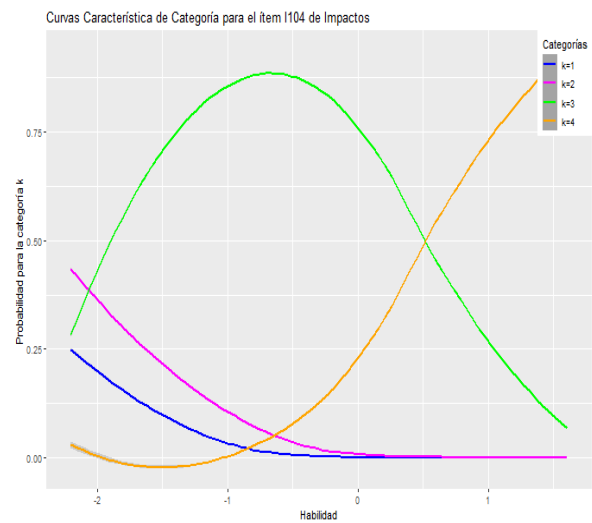
(a) CCCs del ítem I101.



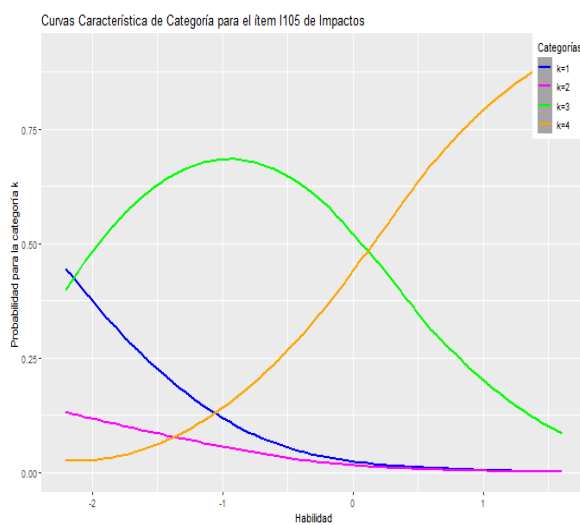
(b) CCCs del ítem I102.



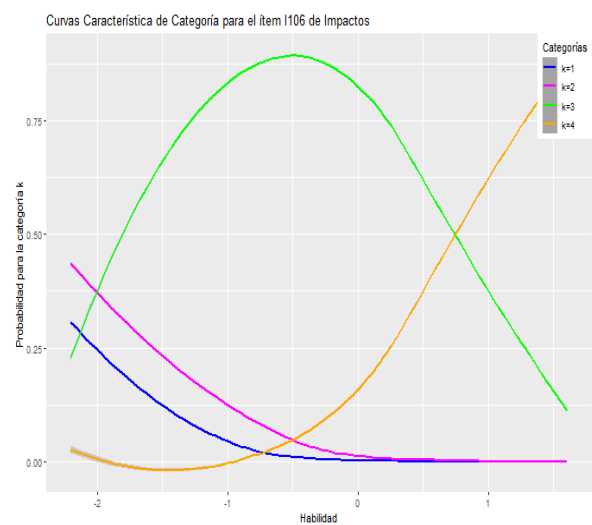
(c) CCCs del ítem I103.



(d) CCCs del ítem I104.

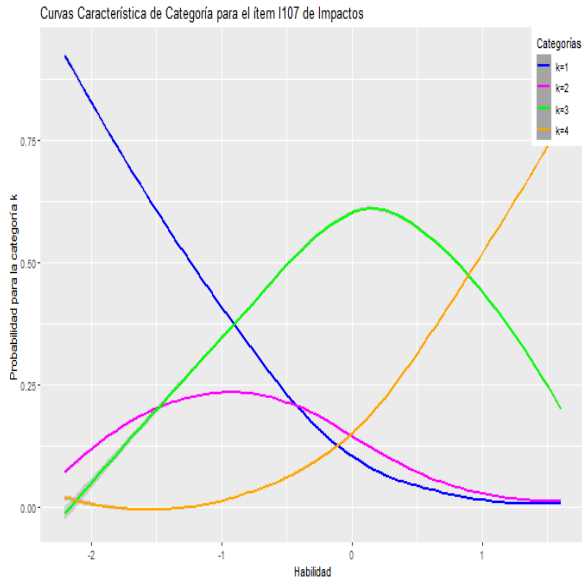


(e) CCCs del ítem I105.

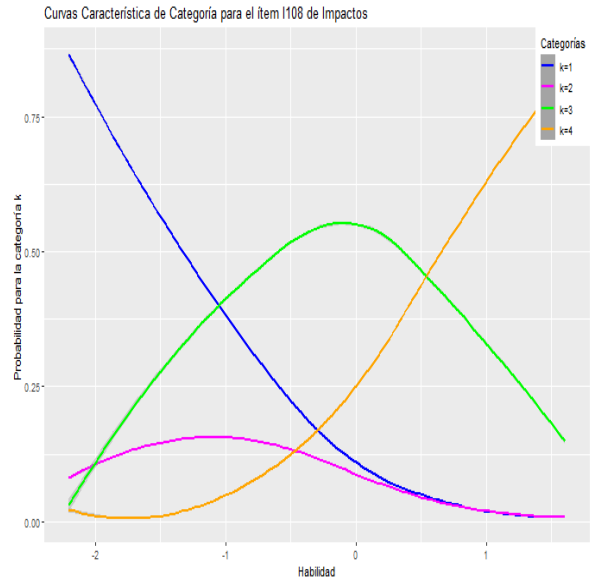


(f) CCCs del ítem I106.

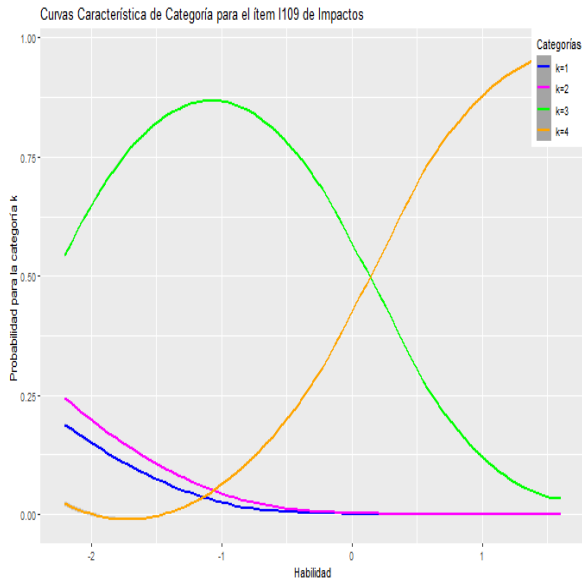
Figura A.39: Curvas Características de Categoría para los ítems del subrasgo *Impactos*.



(a) CCCs del ítem I107.



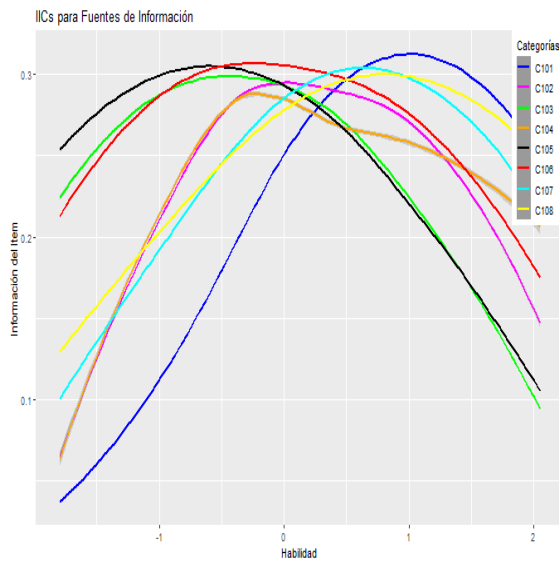
(b) CCCs del ítem I108.



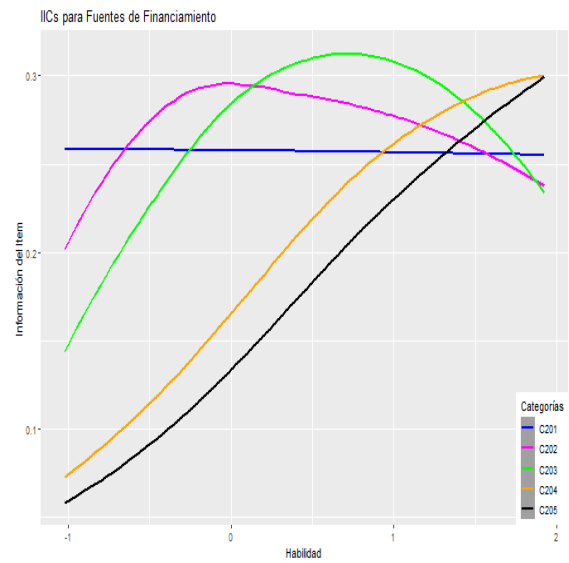
(c) CCCs del ítem I109.

Figura A.40: Curvas Características de Categoría para los ítems del subrasgo *Impactos*.

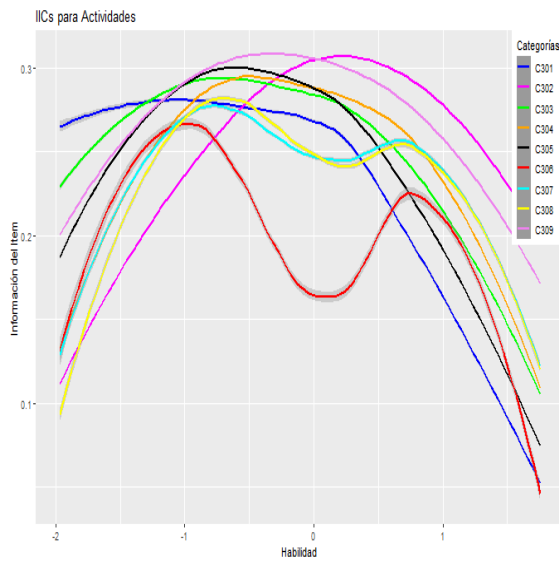
A.5. Criterio de Información del ítem



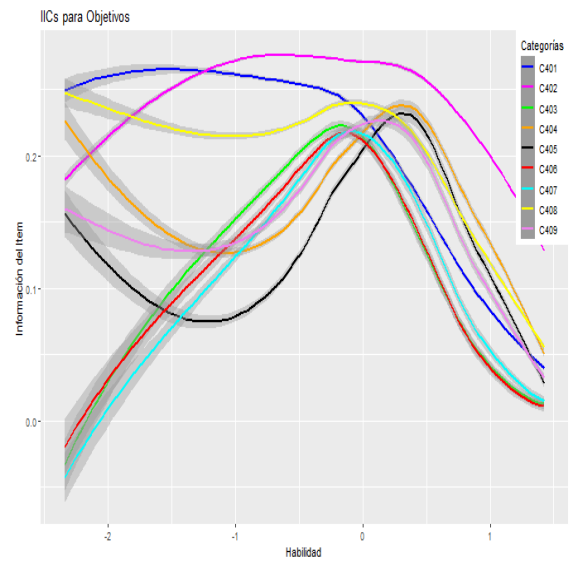
(a) IICs de Fuentes de Información.



(b) IICs de Fuentes de Financiamiento.

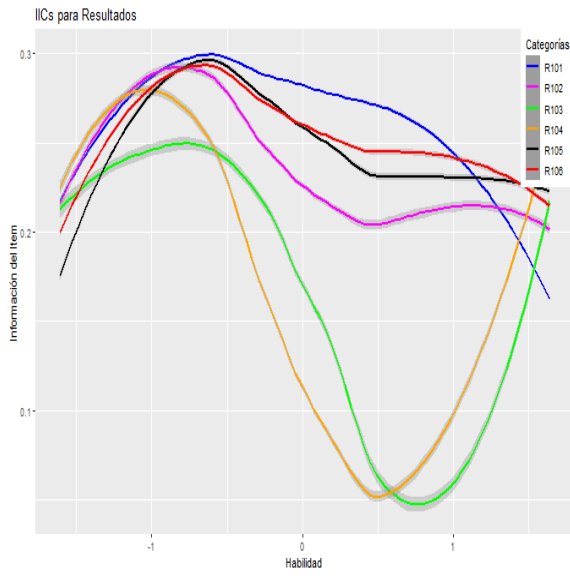


(c) IICs de Actividades.

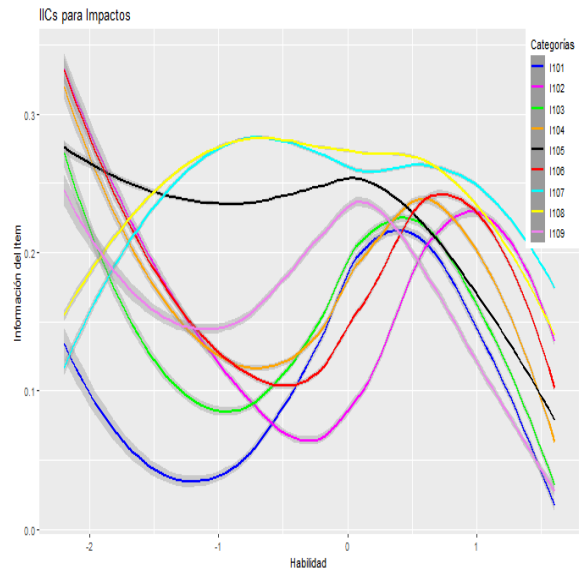


(d) IICs de Objetivos.

Figura A.41: Criterio de Información del Ítem para Fuentes de Información, Fuentes de Financiamiento, Actividades y Objetivos



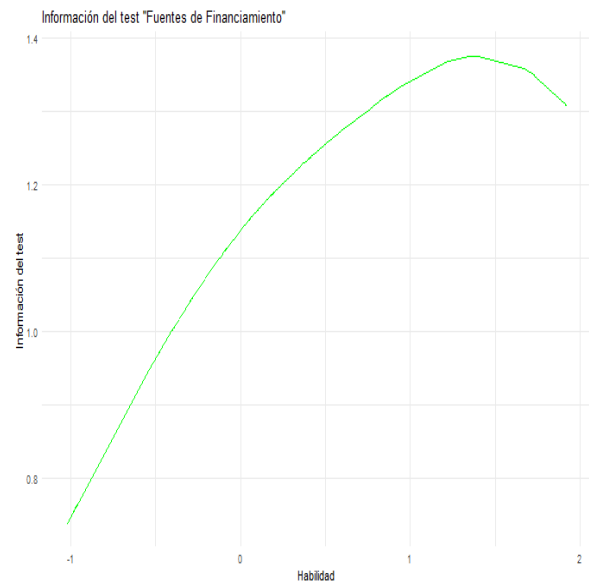
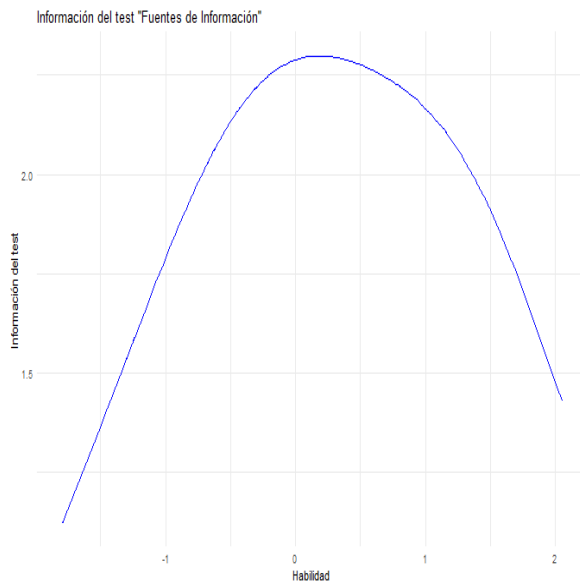
(a) IICs de Resultados.



(b) IICs de Impactos.

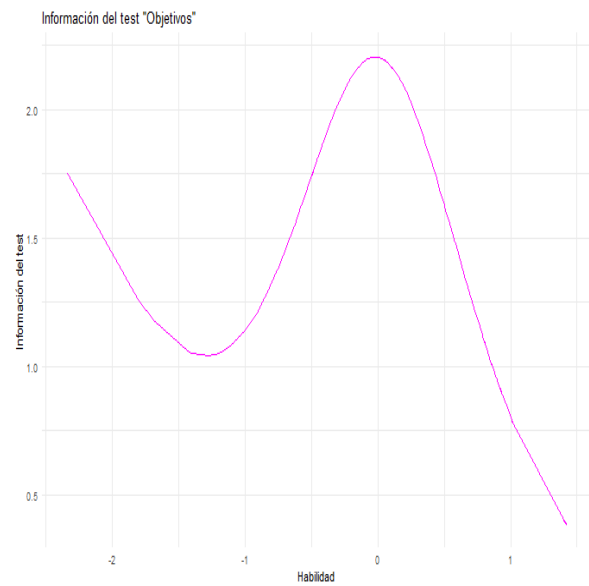
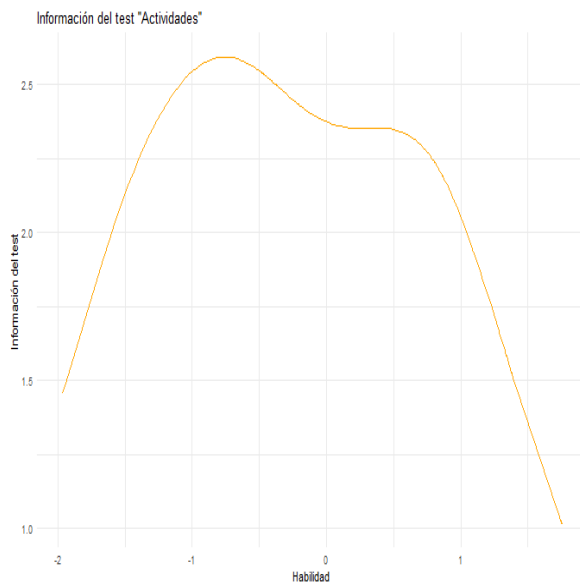
Figura A.42: Criterio de Información del Ítem para Resultados e Impactos

A.6. Información del test



(a) Información del test de Fuentes de Información.

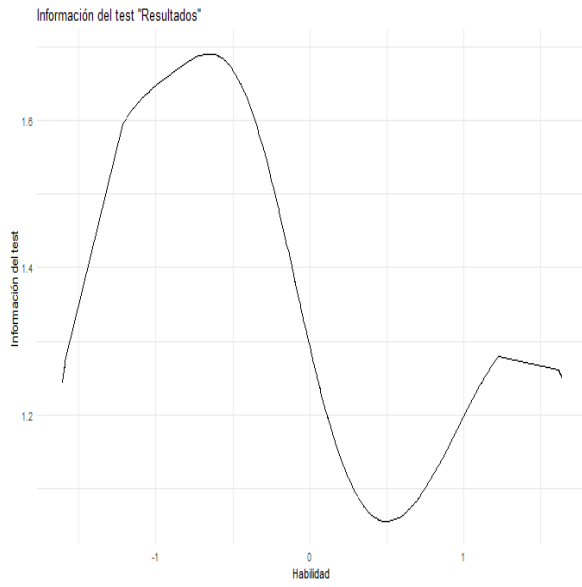
(b) Información del test de Fuentes de Financiamiento.



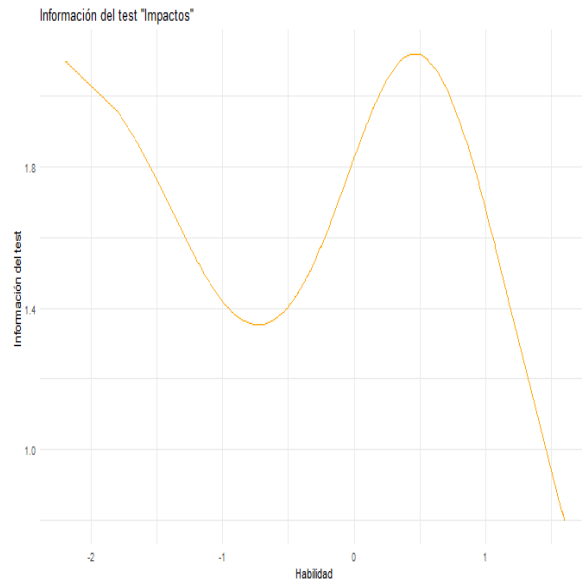
(c) Información del test de Actividades.

(d) Información del test de Objetivos.

Figura A.43: Información del test para *Fuentes de Información*, *Fuentes de Financiamiento*, *Actividades* y *Objetivos*



(a) Información del test de Resultados.



(b) Información del test de Impactos.

Figura A.44: Información del test para *Resultados* e *Impactos*

Bibliografía

- Betancourt, M. & Stein, L. C. (2011). The Geometry of Hamiltonian Monte Carlo. *arXiv: Methodology*.
- Binet, A. & Simon, T. (1904). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année psychologique*, 191-244. https://www.persee.fr/doc/psy_0003-5033_1904_num_11_1_3675
- Birnbaum, A. (1957). Efficient design and use of tests of ability for various decision-making problems Series Report N° 58-16, Project N° 7755-23. *Universal City, TX: US Air Force, Randolph Air Force Base, School of Aviation Medicine*.
- Brooks, S. & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 434-457.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P. & Riddell, A. B. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, 76, 1-32.
- Chamorro, D. (2010). *Espacios de Lebesgue y de Lorentz* (Vol. 1). AMARUN. <https://amarun.org/index.php/editorial/coleccion-de-matematicas-universitarias>
- Champlain, A. F. D. (2010). A primer on classical test theory and item response theory for assessment in medical education. *Medical Education*, 8, 109-117.
- de la Torre & Song. (2009). Simultaneous estimation of overall and domain abilities: A higher-order IRT model approach. *Applied Psychological Measurement*, 620-639.
- Douc, R., Moulines, E., Priouret, P. & Soulier, P. (2019). *Markov Chains*. Springer International Publishing. <https://books.google.com.ec/books?id=eTYnuQEACAAJ>
- Duane, S., Kennedy, A., Pendleton, B. J. & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B. ISSN 0370-2693*, 195, 216-222. [https://doi.org/10.1016/0370-2693\(87\)91197-X](https://doi.org/10.1016/0370-2693(87)91197-X).
- Dutta, S., Lanvin, B., León, L. R. & Wunsch-Vincent, S. (2021). Global Innovation Index 2021. *WIPO: World Intellectual Property Organization*.
- Expansión. (2019). Índice de Desarrollo Humano. <https://datosmacro.expansion.com/idh>

- FUNDESA. (2022). Índice de Competitividad Global [Foro Económico Mundial]. https://fundesa.org.gt/content/files/publicaciones/Documento_Tcnico_ICG_2022.pdf
- Gelman, A., Carlin, J., Stern, H. & Rubin, D. (2003). *Bayesian Data Analysis*.
- Gelman, A. & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* 7, 457-472.
- Geman, S. & Geman, D. (1984). *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*. (Vol. 6). IEEE Transactions on Pattern Analysis; Machine Intelligence.
- GESTION-DIGITAL. (2022). Ecuador debe crecer al menos 4 % en 2022 para volver a niveles prepandemia [Economía y Finanzas]. <https://xurl.es/1qb9j>
- Hastings, W. (1970). *Monte Carlo sampling methods using Markov chains and their applications*. Biometrika.
- Hidalgo, A., León, G. & Pavón, J. (2013). *La gestión de la innovación y la tecnología en las organizaciones* (1ra edición). Ediciones Pirámide.
- Kahn, K. (2021). Understanding Innovation. *Business Horizons*, 453-460. <https://doi.org/10.1016/j.bushor.2018.01.011>
- Kempf, K. (2005). *Encyclopedia of Social Measurement*. Academic. <https://books.google.com.ec/books?id=3ywOAQAAMAAJ>
- Lawshe, C. (2006). A Quantitative Approach to Content Validity. *Personnel Psychology*, 28, 563-575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Ley Orgánica de Emprendimiento e Innovación [Suplemento – Registro Oficial N° 151]. (2020). *ASAMBLEA NACIONAL REPÚBLICA DEL ECUADOR*.
- Lord, F. (1952). A THEORY OF TEST SCORES. *Educational Testing Service*, 7. <https://www.psychometricsociety.org/sites/main/files/file-attachments/mn07.pdf?1576607452>
- Lugones, G. (2009). Módulo de capacitación para la recolección y el análisis de indicadores de innovación. *redes BID, Working paper 8*.
- Lunn, D., Jackson, C., Best, N., Thomas, A. & Spiegelhalter, D. (2012). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. <https://doi.org/10.1201/b13613>
- Luo, Y. & Jiao, H. (2017). Using the Stan Program for Bayesian Item Response Theory. *Educational and Psychological Measurement*, 78, 001316441769366. <https://doi.org/10.1177/0013164417693666>
- Mella, R. S., d.Rosa, L. & Jiménez, B. (2009). *El desafío de la Innovación*. Editorial Universitaria.

- Morales, V., Robalino, A. & Almeida, C. (2019). Propuesta metodológica para la medición del potencial de innovación en las organizaciones ecuatorianas. *Debates Sobre Innovación*. <https://hdl.handle.net/20.500.13048/1811>
- Morales, V., Robalino, A. & Almeida, C. (2022). Validación del cuestionario de Capacidades Resultados e Impactos CRI2 para medir el Potencial de Innovación de las organizaciones (Technical Report No. PIS19-05–003)- Working paper. *Escuela Politécnica Nacional*. https://www.researchgate.net/publication/363069586_Validacion_del_cuestionario_de_Capacidades_Resultados_e_Impactos_CRI2_para_medir_el_Potencial_de_Innovacion_de_las_organizaciones
- Neal, R. M. (1995). Bayesian Learning for Neural Networks. *University of Toronto, Toronto*.
- Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics. *Handbook of Markov Chain Monte Carlo* (p. 115). <http://www.mcmchandbook.net/HandbookChapter5.pdf>
- OECD & EUROSTAT. (2005). Manual de OSLO: Directrices para la recogida e interpretación de información relativa a Innovación. *Comunidad de Madrid, 3rd Edición*.
- Ojala, A. F. & Larsson, J. E. (2016). *Construction of the Berkeley Innovation Index: A Higher-Order Item Response Theory Model Approach*. Lund University; University of California Berkeley.
- Ortiz, M. T. (2018). Notas del curso Estadística Computacional. <https://tereom.github.io/est-computacional-2018/>
- Paek, I. & Cole, K. (2020). *Using R for Item Response Theory Model Applications*. Routledge.
- Patz, R. & Junker, B. (1999). A straightforward approach to markov chain monte carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, 24(2), 146-178.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. *Copenhagen: Danish Institute for Educational Research*.
- Robalino, A., Ramos, V., Unda, X. & Franco, A. (2017a). Diseño de un modelo-herramienta para la medición de la innovación en la industria ecuatoriana. *CienciAmérica (2017)*, 6.
- Robalino, A., Ramos, V., Unda, X. & Franco, A. (2017b). University's contribution to industries in the creation of a tool to diagnose innovation management processes. *Proceedings of INTED2017*, 2351-2360.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Society*. <https://doi.org/https://doi.org/10.1007/BF03372160>

- Sung, H. & Kang, T. (2006). *Choosing a polytomous IRT model using bayesian model selection methods*. In Annual Meeting of the National Council on Measurement in Education.
- Tang, D. (2021). Índice Mundial de Innovación 2021. *OMPI Organización Mundial de la Propiedad Intelectual, Resumen 14*.
- Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology*, 433-451. <https://doi.org/10.1037/h0073357>
- Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4), 1701-1728.
- Tristán-López, A. (2008). Modificación al modelo de Lawshe para el dictamen cuantitativo de la validez de contenido de un instrumento objetivo.
- Ursachi, G., Horodnic, I. & Zait, A. (2015). How Reliable are Measurement Scales? External Factors with Indirect Influence on Reliability Estimators. *Procedia Economics and Finance*, 20, 679-686. [https://doi.org/https://doi.org/10.1016/S2212-5671\(15\)00123-9](https://doi.org/https://doi.org/10.1016/S2212-5671(15)00123-9)
- van der Linden, W. J. (2016). *Handbook of Item Response Theory, Statistical Tools*. CRC Press.