



U

P

T

UNIVERSIDAD POLITÉCNICA DE TULANCINGO

**DETECCIÓN DE MULTIMODALIDAD APLICANDO UN
ESTIMADOR DE DENSIDAD DE NÚCLEO EN
RESULTADOS DE LA PRUEBA DE
REY-OSTERRIETH**

Por

Audberto Hernández Cruz

Tesis Sometida como requisito parcial para obtener el grado de

Maestro en Desarrollo de Software

Asesor:

M.C.C. Luis Roberto Morales Manilla

Tulancingo de Bravo, Hidalgo

Agosto 2018



Dedicatoria

A mi mamá Cornelia, a mi hermana Lidia y tía Josefina que han creído y apoyado este proyecto de estudios, y con su amor que me inspiran cada día.

Agradecimientos

A la Universidad Politécnica de Tulancingo por aceptarme en la Maestría en Desarrollo de Software y brindarme la oportunidad para seguir preparandome.

Al Mtro. Luis Roberto Morales, por su confianza depositada en este proyecto, los conocimientos, consejos, orientación y ayuda brindada durante el proceso de estudios e investigación.

Contenido

Capítulo 1 . Introducción.....	1
1.1 Descripción del problema	3
1.2 Propuesta de solución.....	3
1.3 Objetivos de la tesis.....	5
1.3.1 Objetivo general	6
1.3.2 Objetivos específicos.....	6
1.4 Justificación	7
1.5 Organización del trabajo de tesis	8
Capítulo 2 . Marco teórico.....	11
2.1 Antecedentes de la teoría clásica de los test	13
2.2 Antecedentes de la teoría de respuesta al ítem	16
2.3 La teoría de respuesta al ítem.....	19
2.4 Curva característica del ítem	20
2.5 Supuestos de la IRT.....	22
2.5.1 Unidimensionalidad	22
2.5.2 Independencia local	23
2.6 La evaluación adaptable computarizada	23
2.7 Remuestreo de datos.....	26
2.7.1 Método Bootstrap.....	27

2.8	Estimación de densidad de núcleo	29
2.9	Prueba de hipótesis estadística de Kolmogorov-Smirnov	34
Capítulo 3 . Estado del arte		37
3.1	Prueba de la figura compleja de Rey-Osterrieth	39
3.1.1	Antecedentes	39
3.1.2	Elementos de evaluación de Rey-Osterrieth	40
3.1.3	Calificación de la prueba	42
3.2	Modelos psicométricos de IRT	44
3.2.1	Modelo logístico de un parámetro	45
3.2.3	Modelo logístico de dos parámetros	50
3.2.4	Modelo logístico de tres parámetros	53
3.2.5	Modelo logístico de cuatro parámetros	57
3.2.6	Selección de un modelo psicométrico de IRT	61
3.3	Puntuaciones que no se distribuyen de forma normal.....	65
3.4	Niveles de habilidad y dificultad de los ítems	66
3.5	La herramienta Solver de Excel.....	70
Capítulo 4 . Desarrollo del trabajo de investigación y resultados alcanzados		73
4.1	Datos de estudio	74
4.2	Experimentación.....	88
4.3	Resultados obtenidos.....	91
Conclusiones.....		103

Bibliografía.....	105
Anexos	115
A. Sistemas CAT existentes.....	117
B. Matriz de resultados de la prueba de Rey-Osterrieth	120
C. Matriz de habilidad en escala logit	122
D. Niveles de dificultad en escala logit	124
E. Matriz de resultados al discriminar filas y columnas.....	125
F. Matriz de clasificación de habilidades ordenadas	127
G. Niveles de habilidad para el ítem M1	128
H. Implementación del código de la solución.....	129
I. Tiempos de procesamiento.....	134
J. Valores del remuestreo.....	136

Índice de tablas

Tabla 2.1. Listado de sistemas CAT.....	26
Tabla 3.1. Modelos paramétricos de la IRT.	63
Tabla 3.2. Ejemplo de datos de un examen.	67
Tabla 3.3. Matriz de datos del nivel de habilidad de los examinados.	68
Tabla 3.4. Cálculo de niveles de habilidad de los examinados.	68
Tabla 3.5. Cálculo del nivel de dificultad de los ítems del examen.	69
Tabla 3.6. Matriz de resultado de los niveles de habilidad y dificultad.	70
Tabla 3.7. Matriz agrupada del nivel de habilidad de los sujetos evaluados.	70
Tabla 4.1. Fragmento de la matriz de resultados de la prueba de R-O.....	75
Tabla 4.2. Fragmento del nivel de habilidades en la escala de logit	76
Tabla 4.3. Nivel de dificultad de los ítems de la prueba de R-O.....	76
Tabla 4.4. Fragmento de habilidades ordenadas.....	78
Tabla 4.5. Niveles de habilidad por clase.	78
Tabla 4.6. Niveles de habilidad del ítem M1.....	80
Tabla 4.7. Resultados del ajuste por mínimos cuadrados.	81
Tabla 4.8. Características de software y hardware donde se implementó la solución.	87
Tabla 4.9. Tiempos de cálculos.....	87
Tabla 4.10. Fragmento de valores del remuestreo B de tamaño 500.	92

Tabla A. 1. Listado de sistemas CAT (Parte 1).	117
Tabla A. 2. Listado de sistemas CAT (Parte 2).	118
Tabla A. 3. Listado de Sistemas CAT (Parte 3).	119
Tabla B. 1. Matriz de resultados de la prueba de Rey-Osterrieth (parte 1).	120
Tabla B. 2. Matriz de resultados de la prueba de Rey-Osterrieth (parte 2).	121
Tabla C. 1. Matriz de niveles de habilidad en escala logit (Parte 1).	122
Tabla C. 2. Matriz de niveles de habilidad en escala logit (Parte 2).	123
Tabla D. 1. Niveles de dificultad en la escala logit.	124
Tabla E. 1. Matriz de resultados al discriminar las filas y columnas (Parte 1). ..	125
Tabla E. 2. Matriz de resultados al discriminar las filas y columnas (Parte 2). ..	126
Tabla F. 1. Habilidades ordenadas.	127
Tabla G. 1. Niveles de habilidad para el ítem M1.	128
Tabla I. 1. Tiempos de cálculo al usar las funciones tic y toc de Matlab.	134
Tabla I. 2. Tiempos de cálculo al arranque del S.O. usando tic y toc.	135
Tabla J. 1. Valores de un remuestreo de tamaño 500.	136

Índice de figuras

Figura 2.1. Curvas características de dos ítems aplicando el modelo 1PL.	22
Figura 2.2. Algoritmo de administración de un CAT.	25
Figura 2.3. Estimador de densidad mostrando los núcleos individuales.	32
Figura 3.1. Figura compleja de Rey-Osterrieth.	40
Figura 3.2. CCI con el parámetro $\mathbf{b} = \mathbf{0}$ aplicado al modelo 1PL.	47
Figura 3.3. CCI de 4 ítems aplicando el modelo 1PL.	48
Figura 3.4. CCI aplicando el modelo 2PL.	51
Figura 3.5. CCI con modelo 2PL con $\mathbf{b} = \mathbf{0}$ y distintos valores para \mathbf{a}	52
Figura 3.6. CCI aplicando el modelo 3PL.	55
Figura 3.7. CCI con modelo 3PL con distinto valor en los parámetros \mathbf{a} , \mathbf{b} y \mathbf{c} . .	56
Figura 3.8. CCI aplicando el modelo 4PL.	58
Figura 3.9. CCI del modelo 4PL con diferentes valores para el parámetro $\boldsymbol{\gamma}$	59
Figura 3.10. Representación de los modelos de la IRT.	64
Figura 4.1. Gráfico de dispersión para el ítem M1.	79
Figura 4.2. CCI del ítem M1 y gráfico de dispersión.	80
Figura 4.3. Diagrama de flujo para la detección de multimodalidad.	83
Figura 4.4. Diagrama de flujo para la detección de multimodalidad (continuación).	84
Figura 4.5. Caso de uso del sistema para la detección de multimodalidad.	85

Figura 4.6. GUI del sistema para la detección de multimodalidad.	86
Figura 4.7. Gráfico de dispersión, CCI antes y después de ajuste.	89
Figura 4.8. Estimadores de densidad usando la moda.	95
Figura 4.9. Estimadores de densidad usando la media.	98
Figura 4.10. Estimadores de densidad usando la mediana.	101

Capítulo 1 . Introducción

En el campo de las ciencias de la educación y de la psicometría existen métodos y técnicas para evaluar las habilidades de las personas en determinada área del conocimiento. La manera más común para evaluar es mediante la aplicación de un examen y los resultados obtenidos proveen un conjunto de datos que pueden ser estudiados desde diferentes enfoques. Algunos de estos enfoques recurren a las matemáticas aplicando técnicas estadísticas para dar una interpretación de los resultados obtenidos.

La mayoría de las interpretaciones de los resultados que se pueden obtener de los exámenes toman como métrica el promedio o el puntaje total de la prueba. Existen algunos casos donde la puntuación promedio presenta un comportamiento no estándar, es decir, su distribución no es normal y esto no es tan simple de interpretar. Dentro de la teoría de respuesta al ítem (IRT por sus siglas en inglés)

se propone el uso de un modelo psicométrico para dar una interpretación a este tipo de resultados no normales; sin embargo, no hay a la fecha información suficiente para poder decir qué modelo es el que da mejores resultados de interpretación.

Mediante la selección de un modelo psicométrico se puede realizar un análisis de los resultados de cada ítem de manera particular para conocer la relación que existe entre la habilidad del examinando y la dificultad del ítem. La IRT toma una misma escala para ubicar a la habilidad y a la dificultad, siendo esta la llamada escala logit.

Se sabe dentro de la literatura que en algunos casos las puntuaciones resultantes de un examen pueden no ser estándar o no mostrar el comportamiento de una distribución gaussiana y presentar más de una moda, es decir, se presenta más de un máximo local. Tal es el caso, de la prueba de la figura compleja de Rey-Osterrieth, la cual según (Mitrushina, Boone, Razini, & D'Elia, 2005) arroja resultados de naturaleza multimodal. Esta característica resulta particularmente interesante si se requiere dar una interpretación cognitiva más apegada al contexto dentro del cual se esté realizando la examinación.

La detección previa de la naturaleza multimodal de los datos provenientes de los resultados de un examen resulta ser para nada trivial y daría pauta a interpretaciones con base en una selección de modelos que mejor ajusten a los datos propiciando una mejor interpretación de los resultados.

1.1 Descripción del problema

Dada una muestra de datos obtenidos de la aplicación de la prueba de Rey-Osterrieth en un grupo de alumnos evaluados, se analizará si el puntaje obtenido presenta o no una distribución multimodal. Para ello se recurre al análisis de los resultados mediante una técnica estadística de remuestreo no paramétrico que permita hacer predicciones del comportamiento de los datos, teniendo en cuenta que de antemano no se conoce si los datos presentan multimodalidad.

Existe poca información en la literatura respecto a que una evaluación presente resultados de naturaleza multimodal, esta es una buena razón para analizar los resultados y experimentar con nuevas muestras de datos generadas a partir de una pequeña muestra de resultados.

Si los resultados obtenidos en las nuevas muestras presentan una distribución multimodal o al menos bimodal, esto permitirá elegir qué modelo psicométrico de la IRT se ajusta mejor al aplicar la prueba permitiendo medir el nivel de dificultad de un ítem en particular.

1.2 Propuesta de solución

Como se mencionó anteriormente, el problema que se desea atacar es la detección de multimodalidad en datos provenientes de pruebas o exámenes que en la literatura están reportadas como potenciales candidatos para arrojar resultados

que presentan más de una moda, es decir, no son distribuciones normales. Hasta la fecha no hay ninguna técnica reportada en la literatura que sea capaz de detectar esta multimodalidad con toda certeza, por lo tanto, en este trabajo se explora una alternativa para realizar dicha detección con la finalidad de aplicar los resultados dentro de la teoría de respuesta al ítem, donde la selección de un modelo psicométrico de varios parámetros pudiera modelar adecuadamente los datos multimodales y en tal caso dar una interpretación cognitiva más apegada al contexto donde se realice la prueba.

Para tal efecto, se analizarán los resultados de las puntuaciones obtenidas de la aplicación de la prueba de la figura compleja de Rey-Osterrieth, la cual fue administrada a tres grupos de alumnos de nivel licenciatura de las carreras de Ingeniería en Sistemas Computacionales y de la Licenciatura en Ciencias Computacionales pertenecientes a la Universidad Politécnica de Tulancingo y a la Universidad Autónoma del Estado de Hidalgo respectivamente.

Parte del análisis que se tiene que realizar consiste en obtener los niveles de habilidad de los examinados, así como los niveles de dificultad de los ítems que componen la prueba. Esto requiere la selección de un modelo psicométrico que determine los parámetros bajo los cuales se habrán de calcular tales niveles y establezcan la escala sobre la cual se habrán de medir. Una vez calculados los niveles se deberán agrupar haciendo una clasificación que permita identificar el número de grupos de habilidad por cada ítem. Es muy posible que el número de grupos o clases obtenidas sea muy pequeño por lo que se pretende utilizar la técnica de remuestreo llamada bootstrap no paramétrico. Esto permitirá generar una mayor cantidad de datos sobre los cuales se aplicará un estimador de densidad de núcleo utilizando un kernel gaussiano. Esta técnica según la

literatura permite calcular la distribución de densidad de un conjunto de datos para hallar multimodalidad para lo cual se utiliza un valor de tamaño del ancho de banda y un kernel gaussiano. Regularmente si el tamaño del ancho de banda es grande se obtiene unimodalidad, y si el tamaño del ancho de banda es pequeño se obtiene multimodalidad. Cuando se elige un tamaño del ancho de banda muy grande y se obtiene una distribución de datos multimodales pudiera ser que se esté forzando el cálculo del valor del ancho de banda, dado que este valor siempre estará en función de quien lo asigne se vuelve un tanto subjetiva su interpretación. Para evitar esta subjetividad se utiliza el estimador del ancho de banda propuesto por (Silverman, 1981), el cual permite obtener una distribución multimodal con un ancho de banda óptimo utilizando el tamaño de la muestra.

Para reforzar los resultados obtenidos usando el estimador de densidad de núcleo se aplicará la prueba de bondad de ajuste de Kolmogorov-Smirnov (conocida comúnmente como prueba K-S) al conjunto de resultados obtenidos después de aplicar la técnica bootstrap. Dicha prueba emite un criterio de existencia de unimodalidad y con esto se procederá a sacar conclusiones sobre la detección de multimodalidad en los resultados obtenidos en la prueba de la figura compleja de Rey-Osterrieth.

1.3 Objetivos de la tesis

A continuación se establecen los objetivos que se persiguen en este trabajo de investigación.

1.3.1 Objetivo general

Detectar multimodalidad en datos extraídos de los resultados de la prueba de la figura compleja de Rey-Osterrieth aplicada a grupos de alumnos de carreras profesionales orientadas al área de computación.

1.3.2 Objetivos específicos

Para garantizar el buen desempeño del trabajo de investigación y alcanzar el objetivo propuesto se requiere cumplir con los siguientes objetivos específicos:

- Realizar una revisión de las fuentes de información relacionadas con el tema de investigación para establecer el contexto actual de la problemática a resolver.
- Conocer el estado del arte y determinar si existen trabajos previos en el contexto del tema de investigación.
- Analizar los resultados que se obtuvieron de la aplicación de la prueba de la figura compleja en alumnos del área de computación.
- Obtener los niveles de habilidad de los examinados, así como los niveles de dificultad que componen la prueba de la figura compleja de Rey-Osterrieth.
- Identificar las técnicas reportadas en la literatura que permitan detectar multimodalidad en los datos y que sean adecuadas al contexto del problema que se está atacando.

- Implementar un algoritmo basado en las técnicas identificadas para resolver el problema que conlleva esta investigación.
- Validar los resultados obtenidos en la detección de multimodalidad.

1.4 Justificación

La importancia del presente trabajo de investigación radica principalmente en obtener distribuciones multimodales a partir de la muestra de datos con la que se cuenta. A la fecha no se encuentra reportado en la literatura algún trabajo relacionado con la predicción de multimodalidad en resultados de pruebas o exámenes similares a los de la figura compleja de Rey-Osterrieth, es decir, en exámenes cuyos ítems sean dicotómicos, donde solo tienen dos opciones de respuesta cierto o falso, correcto o incorrecto y la puntuación de las respuestas las consisten en unos y ceros. Lo relevante de las distribuciones multimodales es que al poder predecirse se creará la oportunidad de aplicar este conocimiento en aquellos resultados que provengan de exámenes contruidos con base en la teoría de respuesta al ítem, en donde la selección de un modelo psicométrico que mejor se adecue al contexto del examen es sustancial. Se sabe que los modelos más simples no pueden trabajar con datos multimodales, pero existen modelos más complejos que si lo pueden hacer y esto permite una mejor interpretación cognitivamente hablando de los resultados de un examen de este tipo.

En la literatura consultada se reportan varios estudios donde se hace mención de la existencia de diversas técnicas para obtener nuevas poblaciones mediante métodos de remuestreo de datos no paramétricos, una de estas técnicas es el

método de bootstrap que permite tomar una pequeña muestra de datos y generar nuevas poblaciones a partir de ella dando como resultado una distribución de datos empírica (Ramírez, Barrera, & Correa, 2013).

Los estimadores de densidad de datos no paramétricos son una herramienta valiosa que permite determinar la densidad de un conjunto de datos. Mediante un estimador de densidad y utilizando la función kernel de tipo gaussiano se podrá determinar si la población resultante del remuestreo con bootstrap presenta una distribución multimodal (Silverman, 1981).

Además de aplicar un estimador de densidad que permite establecer si la distribución presenta multimodalidad o no, se recurrirá a aplicar la prueba de Kolmogorov-Smirnov, la cual permite realizar un contraste de hipótesis estadística acerca de la existencia de normalidad para una distribución de datos.

En lo que se refiere a las tecnologías a utilizar para la propuesta de solución de la problemática planteada se decidió trabajar con Microsoft Excel y MathWorks Matlab para agilizar todos los cálculos.

1.5 Organización del trabajo de tesis

El presente trabajo está dividido en varios capítulos que deberán ser agradables al lector, los cuales se encuentran organizados de la siguiente forma:

En el capítulo uno se presenta la introducción, la problemática, la alternativa de solución, los objetivos y la justificación del por qué se está realizando el trabajo de investigación.

En el capítulo dos se hace una exposición general de los conceptos básicos teóricos que se requerirán para el desarrollo de este trabajo, donde se describen los antecedentes de la teoría clásica de los test, la teoría de respuesta al ítem, la técnica de bootstrap no paramétrico para el remuestreo de datos, y el estimador de densidad de núcleo que permite detectar multimodalidad en muestras de datos.

En el capítulo tres se menciona el estado del arte, donde se describen cuatro modelos paramétricos de la teoría de respuesta al ítem, la puntuación y cómo se obtienen los resultados de la prueba con copia y reproducción de la figura compleja de Rey-Osterrieth.

El capítulo cuatro se describe el desarrollo, los métodos utilizados y los resultados del análisis de la puntuación obtenida por un grupo de estudiantes universitarios a quienes se les aplicó la prueba de Rey-Osterrieth y de donde se parte para realizar la detección de multimodalidad.

Finalmente se discuten, se analizan y se validan los resultados obtenidos con base a la revisión de la literatura y los datos experimentales utilizados. Se hacen recomendaciones para la aplicación de los resultados obtenidos en este trabajo en la selección del modelo psicométrico en la interpretación de resultados de exámenes con base en la IRT. Se describen las conclusiones y el trabajo futuro.

Capítulo 2 . Marco teórico

Dada la importancia de la presente investigación se requiere conocer los antecedentes en los cuales se sustentan los procesos de examinación que se utilizan en este trabajo, tales procesos consideran la aplicación de algunas teorías muy utilizadas dentro de la literatura. Una de estas bases es la llamada teoría clásica de los test (CTT por sus siglas en inglés) la cual data desde principios del siglo XX y establece que existe una relación lineal entre el nivel de habilidad de los evaluados y la puntuación total obtenida en el examen. Las características del examen y las características del evaluado son dependientes, es decir, la habilidad del evaluado se mide mediante el número de respuestas acertadas en el examen realizado, si la dificultad de las preguntas es alta el nivel de habilidad de los sujetos resultará ser baja, si la prueba es fácil entonces la habilidad de responder correctamente será alta.

Al utilizar un modelo donde la prueba gire en torno al evaluado resulta muy complicado evaluar a distintos grupos de sujetos, es necesario un modelo donde no exista una relación directa entre los exámenes y los evaluados. De esta problemática surge un modelo denominado teoría de respuesta al ítem (IRT) el cual está centrado en evaluar a los ítems de manera particular y no a todo el examen como lo establece la CTT. La IRT mide el nivel de habilidad de los evaluados y el nivel de dificultad de los ítems situándolos en una misma escala de puntuación. Cada ítem se define mediante una función matemática conocida como modelo psicométrico y su representación gráfica se denomina curva característica del ítem (ICC), la cual establece la relación de medir el nivel de probabilidad de emitir una respuesta correcta y el nivel de habilidad de un evaluado frente a un ítem en específico.

Con los resultados de un examen basado en IRT se pueden obtener predicciones de los niveles de habilidades de un grupo de evaluados y los niveles de dificultad de los ítems. Cuando los resultados al medir los niveles de habilidad y dificultad obtenidos en el examen dan un número limitado de observaciones, es necesario incrementar el número de muestras que permitan tener un análisis más detallado de los resultados de las observaciones, para ello se toma como alternativa aplicar una técnica de remuestreo estadístico no paramétrico. Para el análisis de datos de este trabajo de investigación se recurre a aplicar el método de bootstrap no paramétrico, el cual consiste en un algoritmo iterativo que permite obtener nuevas muestras a partir de una porción de datos, es decir, se obtiene el remuestreo a partir de las muestras y los resultados de este permitirá analizar el comportamiento de la población de datos del objeto de estudio en cuestión (Solanas & Sierra, 1992).

Los resultados de un remuestreo no paramétrico arrojan una distribución de datos empírica, la cual se quisiera poder identificar de manera previa si es de naturaleza gaussiana o no, para ello se puede aplicar un estimador de densidad, el cual permite predecir la forma de la distribución ya sea unimodal o multimodal. Para llevar a cabo esta estimación se requiere del uso de una función kernel y del valor de un parámetro de ancho de banda. Un inconveniente importante es que el parámetro de ancho de banda se debe escoger *a priori* a la aplicación de la función kernel y esto afecta la estimación de densidad, por lo tanto, típicamente se recurre a utilizar un estimador de ancho de banda óptimo como el que se propone en (Silverman, 1981), el cual está centrado en la media y la desviación estándar de las observaciones para poder predecir si el comportamiento de una distribución es multimodal o no.

Para reforzar el resultado obtenido por el estimador de kernel, se realiza la prueba de hipótesis de normalidad como se establece en (Facchinetti, 2009), dicha prueba permite identificar si un grupo de observaciones se ajusta a una distribución normal permitiendo comprobar la hipótesis de normalidad de los datos (Pedrosa, Juarros Basterretxea, Robles Fernández, Basteiro, & Garcia Cueto, 2014).

2.1 Antecedentes de la teoría clásica de los test

Las evaluaciones psicométricas se centran en el uso del método científico el cual permite contrastar las mediciones y las hipótesis que se plantean al obtener los resultados de las evaluaciones, en este contexto las mediciones se centran en asignar valores a los objetos medidos mediante un conjunto de variables, de tal

manera que esos valores representan a los constructos de los objetos que se miden. A diferencia de los fenómenos físicos que establecen características de medición directa, las variables psicométricas tratan de enlazar las medidas a través de los constructos como indicadores empíricos que puedan ser observados de manera indirecta, por ejemplo, medir una conducta o el nivel de dominio de un tema en específico, por esta razón surgieron métodos para el análisis de resultados de las pruebas psicométricas mediante un sistema numérico (Prieto & Delgado, 2010). Las exámenes que se realizan se basan en aspectos tales como el nivel de conocimiento de un tema en particular, la inteligencia de la persona o sus rasgos de personalidad, por ejemplo, el comportamiento, el nivel de extroversión, el autoconcepto y la asertividad.

El modelo estadístico más antiguo para el análisis de instrumentos de medición de evaluaciones psicométricas es la denominada teoría clásica de la medida (TCM) la cual fue desarrollada por el psicólogo británico Charles Spearman, quien publicó una serie de argumentos matemáticos en los que expuso que los puntajes obtenidos en un examen son medidas inexactas de los rasgos humanos (Spearman, 1904). El objetivo de Spearman era encontrar un modelo estadístico que fundamentase adecuadamente las puntuaciones de los exámenes y permitiera realizar la estimación de errores asociados a todos los procesos de medición y el cual se basa en un modelo de regresión lineal aplicado en pruebas psicométricas. Spearman fue el primero en establecer una teoría de medición psicológica de inteligencia, por lo que es considerado el padre de la CTT.

La teoría clásica de los test fue desarrollada durante la década de los años veinte del siglo pasado, en esta teoría se sostiene el argumento de que los resultados de la medición de una variable latente dependen de la evaluación utilizada, lo que

determina la existencia de serios problemas para tratar de establecer una equivalencia entre los niveles de puntuación de dos exámenes diferentes que miden la misma variable, con lo que era poco probable poder comparar los resultados de distintos grupos para medir el rendimiento de la evaluación.

En la teoría clásica la medida de una variable es inseparable del instrumento utilizado para medirla, además se establecen los conceptos de validez, fiabilidad, sesgo del ítem y del examen, la invarianza de parámetros, dimensionalidad y objetividad de la medida, esto constituye una serie de limitaciones pues de manera operativa se define la variable que se utiliza para medirla (Muñiz, 1998). Las bases de la teoría de Spearman establecieron la presencia de errores de medida y la concepción del error de medida de una variable aleatoria, para ello surge la idea de correlación entre variables y un coeficiente para expresarla junto con la distribución normal bivariada.

Los supuestos de la teoría clásica de los test permiten enlazar las propiedades de distribución de los errores de medida y su independencia de las puntuaciones verdaderas, con la finalidad de obtener una cuantificación del error que se asocia a un proceso de medida o por el contrario el nivel de precisión del instrumento de medida (Hambleton & Van der Linden, 1982).

La idea fundamental del modelo de la CTT se basa en la ecuación 2.1:

$$Puntaje\ observado = Puntaje\ verdadero + error \quad (2.1)$$

Implica que el puntaje observado se obtiene a partir de la calificación del examen y está compuesto por dos elementos básicos: el puntaje verdadero y un error de medida (Hambleton, Swaminathan, & Rogers, 1991). Estos valores dependen de

varios factores, por ejemplo, si el evaluado repite la prueba varias veces el error tiende a desaparecer y es el puntaje verdadero del examinado el que permanece; sin embargo, en la práctica la aplicación repetida de una evaluación en un mismo sujeto resulta inviable.

La CTT ha avanzado y se ha adaptado a las necesidades actuales; sin embargo, sus principios básicos permanecen como la forma más popular de análisis de la evaluación moderna ya que es más fácil de utilizar. El análisis puede realizarse sin el uso de un software especializado, simplemente usando una calculadora o una hoja de cálculo, ya que por lo general basta con contar las respuestas correctas de la prueba.

Uno de los inconvenientes de la CTT es que el puntaje que obtiene el evaluado en la prueba es dependiente de los ítems que le tocan en un examen en particular. Mientras que los puntajes de los ítems dependen de los evaluados que contestaron en esa ocasión la evaluación, además supone que el error estándar de medida es igual a lo largo de toda la escala de habilidad.

2.2 Antecedentes de la teoría de respuesta al ítem

La teoría de respuesta al ítem surge como una alternativa que ofrece soluciones viables a algunas de las limitaciones teóricas y prácticas de la CTT, por lo cual se podría decir que llega a ser complementaria, además, no solo es una teoría de medida psicométrica y de educación, si no que ofrece múltiples y relevantes

aplicaciones de investigación en el campo de la elaboración de exámenes (Muñiz, 2010), (Martínez Arias, Hernández Lloreda, & Hernández Lloreda, 2014).

Los antecedentes que dan origen a esta teoría datan desde 1927 con las investigaciones de Louis Leon Thurstone sobre la percepción del estímulo, quien intentaba sintetizar la psicofísica y la teoría de test. Sus aportaciones ayudaron a comprender las diferencias intraindividuales observadas en el desempeño frente a pruebas de inteligencia general, permitiendo la construcción y mejora de las evaluaciones de inteligencia, de personalidad e intereses entre otros aspectos psicológicos. Thurstone afirmaba que el instrumento de medida no debe venir afectado por los objetos medidos y las mediciones deben ser independientes a los objetos medidos (Jones & Thissen, 2007).

Durante la década de los años cincuenta es cuando se da lugar al seguimiento de la teoría de respuesta al ítem, con la publicación de la tesis doctoral de Frederick Lord en 1952 titulada *A theory of score test*, en la cual presentó formalmente las bases teóricas del modelo de ojiva normal de dos parámetros (Lord, 1952). Por otro lado, Paul F. Lazarsfeld introduce un modelo de distancia latente y un modelo lineal, esta aportación describe el campo de actitudes latentes el cual no trascendió, pero contribuyeron a ampliar el marco teórico de esta importante teoría de test (Barbero García, 1999).

En la década de los años sesenta se publican los trabajos de (Rasch G., 1963) quien desarrolla dos modelos basados en la distribución de Poisson en pruebas de lectura, y un modelo basado en la función logística para la evaluación de inteligencia y rendimiento, este último fue denominado *Structural Model for Items in a Test*, que actualmente se conoce como modelo de Rasch (Rasch G., 1960), (Rasch G., 196x).

De manera independiente Allan Birnbaum trabajó los modelos logísticos de dos y tres parámetros exponiendo su investigación como capítulos en la publicación del libro titulado *Statistical Theories of Mental Test Scores* de Frederick M. Lord y Melvin R. Novick (Birnbaum, 1967).

Durante los años setentas se centra el interés en el desarrollo y búsqueda de métodos para la estimación de los parámetros y su implementación en ordenadores, es cuando aparecen los primeros programas de software para la estimación de parámetros, estos fueron BICAL, LOGIST, BILOG, MULTILOG, MICROART, NOHARM, ANCILLES y OGIVA los cuales fueron fundamentales para la utilización de los modelos de la teoría de respuesta al ítem (Muñiz & Hambleton, 1992).

Las investigaciones realizadas en los años ochenta suponen la verdadera expansión y afianzamiento de la IRT y su masivo predominio en psicometría, Lord en 1980 publica su libro bajo el título de *Applications of Item Response Theory to Practical Testing Problems*, el cual recoge los desarrollos teóricos así como las aplicaciones de la teoría de respuesta al ítem que se habían desarrollado hasta entonces (Barbero García, 1999).

Los modelos actuales de IRT que se describen en la literatura son bastante amplios, y se caracterizan por ser más o menos restrictivos, teniendo en cuenta que existen modelos unidimensionales y modelos multidimensionales que difieren en función de la respuesta al ítem, que vienen dados por uno o más rasgos latentes y la curva característica está definida teóricamente en modelos paramétricos y no paramétricos, el formato de respuesta al ítem está clasificado en dicotómicos y politómicos (Martínez Arias, Hernández Lloreda, & Hernández Lloreda, 2014).

2.3 La teoría de respuesta al ítem

La denominación de teoría de respuesta al ítem agrupa varias líneas de investigación psicométricas independientes iniciadas por (Rasch G. , 1960) y (Birnbaum, 1967) como lo expresa (Muñiz, 2010). El factor común de estos desarrollos es que establecen una relación entre el comportamiento de un examinado frente a un ítem y el rasgo responsable de esta conducta o rasgo latente. Por ello, recurren a funciones matemáticas que describen la probabilidad de dar una determinada respuesta al nivel de rasgo medido por este (Attorresi, Lozzia, Abal, Galibert, & Aguerri, 2009).

La IRT considera al ítem como unidad de medición, lo cual no contradice los supuestos de la CTT, además hace aceptaciones adicionales que permiten superar las debilidades de la CTT y se enfoca en el análisis de cada uno de los ítems más que en el análisis de la evaluación en general (Muñiz, 2010).

Cuando se aplica la IRT se deben de usar los conceptos de función de información del ítem y de la función de información del test que son las herramientas de trabajo para la selección de ítems y la construcción de evaluaciones. El ajuste de un modelo de IRT proporciona estimaciones invariantes de las propiedades psicométricas de los ítems, así como las características de los sujetos, es decir, los parámetros que caracterizan al ítem no dependen de una muestra en particular de sujetos utilizados, y los parámetros que caracterizan al sujeto no dependen de las muestras en particular de los ítems utilizados (de Ayala, 2009).

El error típico de medida contrario al propuesto en la CTT se presenta condicionado a un nivel de habilidad específico de un evaluado, siendo factor

básico para describir la calidad de la evaluación y también permite obtener la interpretación individual de las puntuaciones obtenidas por el evaluado. Pueden encontrarse situaciones donde los errores típicos de la CTT y de la IRT serán similares cuando el nivel de dificultad del test coincida con el nivel medio de habilidad del grupo de sujetos y además de la distribución de habilidad para esa población cuando se cumpla el supuesto de normalidad (Muñiz, 2010).

2.4 Curva característica del ítem

La curva característica del ítem (CCI) indica la probabilidad de que un evaluado acierte a un ítem en un examen, esto depende naturalmente de cuál sea el nivel de habilidad del evaluado o lo que es lo mismo, la variable latente medida, es decir, para cada nivel de habilidad θ , existe una probabilidad $P(\theta)$ de que un examinado con dicha habilidad conteste correctamente al ítem y esta probabilidad crecerá en tanto aumente el nivel de habilidad (de Ayala, 2009).

Cada evaluado que responde a un ítem en un examen posee una cantidad de habilidad determinada o rasgo latente, es decir, se considera que tiene un valor de habilidad numérico con un puntaje que lo sitúa en una escala y se denota con la letra griega theta θ . Típicamente, esta escala trata de situar tanto a la habilidad del examinado como a la dificultad del ítem en la misma gráfica (Baker, 1985).

La IRT proporciona una descripción detallada del comportamiento individual de los ítems a lo largo de la escala de habilidad, cuya representación gráfica se denomina curva característica del ítem o función de respuesta del ítem (Hidalgo

Montesinos & French, 2016). La CCI tiene forma de "S" y se representa mediante una función sigmoidea de crecimiento exponencial.

Una manera de construir una CCI está dada por el uso de un modelo logístico o también llamado modelo psicométrico. La principal característica en los modelos logísticos radica en el número de parámetros que se deben calcular y los más conocidos son los siguientes: el modelo 1PL o modelo de Rasch, el 2PL, 3PL y 4PL. Las propiedades de los parámetros están en función del modelo elegido para describir el comportamiento de la CCI, en la figura 2.1 se ilustran dos ítems utilizando el modelo 1PL estimando el parámetro asociado con el nivel de dificultad (Attorresi, Lozzia, Abal, Galibert, & Agerri, 2009).

Si se considera la escala de puntuación en una recta horizontal donde los ítems de menor dificultad se sitúan a la izquierda y los de mayor dificultad a la derecha, tomando típicamente valores de $(-4,4)$ y con punto medio el cero. En consecuencia, los evaluados con menor habilidad se sitúan más a la izquierda que aquellos con una mayor habilidad, considerando el nivel de probabilidad de respuesta de 0.5, es decir, que responda correctamente o que falle al ítem, de ahí la relación fundamental que existe entre los valores de las variables que miden a los ítems y la probabilidad de emitir una respuesta correcta.

La CCI es un modelo de ajuste de la IRT con el cual se facilita la estimación de la precisión de cada nivel del rasgo y donde se admite la variación de la misma a lo largo de la escala de habilidad en una evaluación (Jiménez Alfaro & Montero Rojas, 2013).

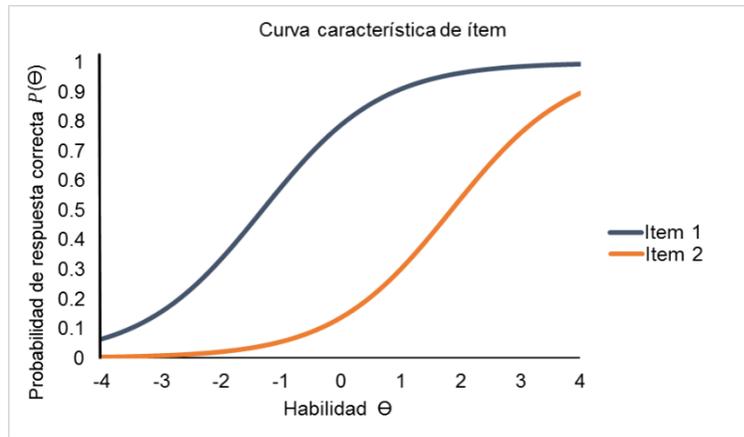


Figura 2.1. Curvas características de dos ítems aplicando el modelo 1PL.

2.5 Supuestos de la IRT

Los modelos logísticos paramétricos de la teoría de respuesta al ítem están basados en una familia de modelos matemáticos que están expresados mediante funciones logísticas y de probabilidad, para lo cual existen ciertas restricciones o condiciones que se deben considerar sobre la naturaleza de los parámetros del modelo elegido con respecto a los ítems.

2.5.1 Unidimensionalidad

No es válido aplicar algún modelo de IRT para analizar una prueba que incluya preguntas de varias asignaturas, para ello se deberá recurrir a analizar cada asignatura de manera independiente, es decir, cada ítem solo mide una habilidad.

Indica que el constructo que se requiere medir es una variable latente continua, es decir, el conjunto de ítems de una prueba está diseñada para medir una sola habilidad latente (Hambleton, Swaminathan, & Rogers, 1991).

2.5.2 Independencia local

Dado que las herramientas de la IRT se enfocan en el ítem y no en la evaluación, resulta sumamente importante que las respuestas de los evaluados a cualquier ítem sean independientes, que la probabilidad de responder correctamente a determinado ítem no esté influida por la probabilidad de respuesta correcta de otros ítems que componen la evaluación, es decir, cada ítem es independiente a todos los demás (Hambleton, Swaminathan, & Rogers, 1991).

2.6 La evaluación adaptable computarizada

En la actualidad la elaboración de los llamados test adaptativos computarizados (CAT por sus siglas en inglés) han sido una de las principales aplicaciones en las que la IRT ha tenido un gran impacto, mediante la elaboración de bancos de ítems que permiten la construcción de sistemas de evaluaciones automatizadas (Olea, Abad, & Barrada, 2010). El hecho de que una evaluación mediante un CAT se adapte a las características del evaluado facilita que de manera automática se apliquen exámenes que estén en función de las respuestas del evaluado (Herranz Torres, 2012). Por otro lado, se puede ver actualmente su aplicación en la

construcción de evaluaciones orientadas a la educación y a la psicometría (Simanca Herrera, Abuchar Porras, & Rivera Osorio, 2014).

Los exámenes se adaptan a la presentación de los ítems en función del nivel de atributo manifestado por el evaluado, a través de las repuestas y se busca seleccionar para cada examinado, el grupo de preguntas que estima con mayor precisión y eficiencia el nivel del atributo (López, Sanmartín, & Méndez, 2014). El mecanismo de trabajo de este tipo de instrumento de evaluación inicia con la presentación de un ítem que ha sido seleccionado según una estrategia de arranque definida previamente, luego con la respuesta dada por el evaluado a este primer ítem, se realiza una estimación provisional de su nivel de atributo θ , y con base en esta estimación se elige el siguiente ítem y se repite el proceso de estimación hasta que haya satisfecho cierto criterio de parada, culminando con una estimación definitiva del nivel de atributo θ del evaluado (Barrada, 2012), (Martínez Arias, Hernández Lloreda, & Hernández Lloreda, 2014). La figura 2.2 muestra el algoritmo de administración de evaluaciones mediante un sistema CAT basado en la IRT (Aybek & Demirtasli, 2017).

La construcción de sistemas CAT datan de comienzos de los años setenta y fue implementada por primera vez durante los ochenta a partir del sistema *Computerized Adaptive Screening Test* de la armada estadounidense, el precursor de lo que hoy en día se conoce como CAT-ASVAB, MICROPAT para la selección de pilotos y controladores aéreos, las pruebas de admisión a centros educativos como Law School Admission Test, para la evaluación educativa de niños y jóvenes como el COMPASS placement test, el NWEA o el sistema CARAT, la prueba holandesa MATHCAT para la evaluación de conocimientos matemáticos de personas adultas (Olea & Ponsoda, 2004).

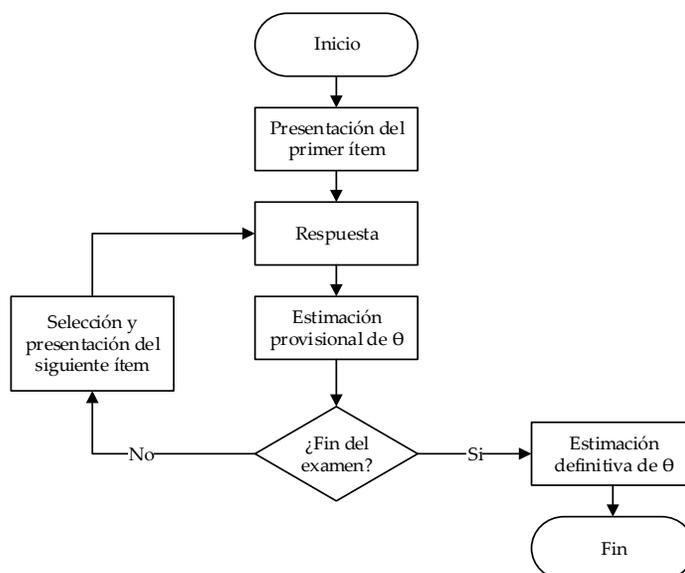


Figura 2.2. Algoritmo de administración de un CAT.

En la actualidad existen instituciones y programas estandarizados de evaluación que incluyen sistemas CAT basados en IRT, en los Estados Unidos existen evaluaciones adaptativas del GRE, que es un examen de conocimientos para acceder a estudios universitarios, el GMAT es una prueba para el ingreso a escuelas de negocios, el ASVAB que es un examen de aptitudes del ejército y de diversos exámenes para la acreditación profesional o de evaluación del nivel educativo de los estudiantes de primaria y secundaria (Olea, Abad, & Barrada, 2010). Las aplicaciones comerciales basadas en sistemas CAT como son FastTest, Pearson VUE, Prometic y McCann, además existen herramientas de código abierto como lo son Concerto, IRT-CAT y OSCAT (Aybek, 2016), (Aybek & Demirtasli, 2017), (Oppl, Reisinger, Eckmaier, & Helm, 2017). Como se observa en la tabla 2.1 que muestra 3 sistemas CAT basados en la IRT y en las tablas A.1, A.2 y A.3 del anexo A donde se describe un listado de los sistemas más conocidos que se reportan en la literatura.

Nombre	Descripción	Distribución
Concerto	Sistema CAT basado en web desarrollado en el lenguaje R.	The Psychometrics Center, Universidad de Cambridge
IRT-CAT	Utiliza el modelo 3PL, Desarrollado para web con PHP, JavaScript y PostgreSQL.	Lee, Park & Park
OSCATS	El desarrollo y simulaciones de sistemas CAT mediante los modelos 2PL y 3PL.	OpenSource

Tabla 2.1. Listado de sistemas CAT.

Para el desarrollo de un sistema CAT se necesitan los siguientes elementos básicos, un banco de ítems calibrado con parámetros estimados a partir de un modelo IRT, procedimientos para iniciar y finalizar el examen, escoger los ítems que se les presentara al examinado y el uso de un método estadístico de estimación del atributo (López, Sanmartín, & Méndez, 2014).

2.7 Remuestreo de datos

Para que la información sea más precisa y pueda ser utilizada mediante el análisis estadístico, es necesario caracterizar las variables medidas y después definir los parámetros de interés. Las distribuciones de estas variables se pueden calcular de manera sistemática mediante un algoritmo y esto brinda información que es relevante en un experimento. Un parámetro es la información contenida en la población y un estadístico resume el valor en una muestra. Los estadísticos varían de una muestra a otra en una misma población y pueden ser utilizados para generar inferencias, con lo que se puede emitir un juicio acerca de la probabilidad del valor del parámetro de la población de estudio (Ramírez, Barrera, & Correa, 2013).

Dentro de las técnicas de remuestreo de datos no paramétricos, el método bootstrap puede ser una alternativa para minimizar el error de medida y obtener los intervalos de confianza, su aplicación es aconsejable en los casos en los que no se conoce la distribución de los datos y no puede asumirse la hipótesis de normalidad (Badil, Castillo, Wong, & Landeros, 2006).

2.7.1 Método Bootstrap

En la actualidad se han desarrollado una serie de técnicas de cálculos intensivos por computadora, mediante la aplicación de algoritmos para la representación de muestreos, que están destinadas a obtener medidas y la precisión de las estimaciones del error típico, el sesgo y los intervalos de confianza, entre tales métodos se encuentra el método bootstrap introducido por Bradley Efron en el año de 1979 (Efron B. ,1980).

El método bootstrap consiste en una técnica de inferencia estadística denominada remuestreo, es decir, se utiliza un procedimiento que obtiene las observaciones a partir de la muestra y las reutiliza para para constituir una nueva población, la cual está formada por un conjunto de datos repetidos que se obtuvieron mediante un remuestreo con remplazo (Efron & Tibshirani, 1986).

Constituye la línea de investigación de los métodos de remuestreo más desarrollada tanto desde el punto de vista teórico como aplicado. Este método no paramétrico permite realizar una aproximación de la distribución de los estimadores estadísticos, por ejemplo, en la aplicación de un estimador de núcleo

para determinar la densidad de un conjunto de datos observados, y otras aplicaciones en modelos de regresión lineal, ajuste de curvas, contraste de hipótesis, etc. (Chernick M. R., 2008). Bootstrap es un método de experimentación probabilístico donde el resultado no se puede predecir con certeza, si no con cierto grado de probabilidad o riesgo. Con un tamaño pequeño para la muestra de datos se pueden obtener estimaciones confiables acerca de los parámetros poblacionales, y está indicado en los casos que no se sigue una distribución normal, es decir, es útil cuando las distribuciones de las muestras son desconocidas o que difícilmente se obtienen por métodos analíticos (Chernick & LaBudde, 2011). La idea central es que muchas veces se pueden extraer conclusiones sobre las características de la población a partir de la muestra que se maneja después de haber aplicado el método bootstrap, lo que permite hacer suposiciones más realistas del comportamiento de los datos en un experimento.

Dada una muestra con n observaciones se asigna una masa de probabilidad de $1/n$ a cada una de las observaciones. El método bootstrap implica realizar un remuestreo con remplazo de los datos obtenidos en una muestra y muchas veces se utiliza para generar una estimación empírica de la distribución (Hernández Abreu & Martínez Pérez, 2012).

Pasos básicos para la estimación con el método bootstrap:

1. Construir una distribución de probabilidad empírica $\hat{F}(x)$, a partir de la muestra, asignando una probabilidad de $1/n$ a cada punto X_1, X_2, \dots, X_n .
2. A partir de la función de distribución empírica $\hat{F}(x)$, se extrae una muestra aleatoria simple de tamaño n con reposición.

3. Se calcula el estadístico de interés mediante el valor de la variable aleatoria X , al iterar este proceso n veces, se obtiene un conjunto de datos que constituyen la denominada muestra bootstrap $\hat{\theta}(x_1^*, x_2^*, \dots, x_n^*)$. Para cada una de las muestras se obtiene el estadístico de interés y se denota por $\widehat{\theta}_i^* = \hat{\theta}(x_1^*, x_2^*, \dots, x_n^*)$ al resultado.
4. Se repiten los pasos 2 y 3 en B ocasiones, donde B tomaría un tamaño hasta 100000 como lo expone (Chernick & LaBudde, 2011) y está en función de la capacidad computacional.
5. Construir una distribución de probabilidad a partir de los B valores $\widehat{\theta}_i^*$, asignando una probabilidad de $1/B$ a cada punto $\widehat{\theta}_1^*, \widehat{\theta}_2^* \dots \widehat{\theta}_B^*$. Esta es la estimación bootstrap de la distribución muestral de $\hat{\theta}[F * (\tilde{\theta}^*)]$. (Hall, 1992)

La ventaja del método bootstrap parte del supuesto de que no es necesario conocer desde el principio la hipótesis acerca de la distribución de los datos, ya que se genera una distribución de probabilidad empírica del estimador y a partir de la cual se obtienen los intervalos de confianza, el sesgo y el error estándar de la muestra (Chernick & LaBudde, 2011).

2.8 Estimación de densidad de núcleo

Los estimadores de densidad de tipo núcleo propuestos por Murray Rosenblatt en 1956 y Emanuel Parzen (Parzen, 1962) también conocido como método Parzen-Rosenblatt, es el más popular y frecuentemente utilizado para la estimación no paramétrica de la función de densidad (Martínez Cambolor, 2008).

Una función de densidad describe la frecuencia con la que aparece una variable en un conjunto de datos y de la cual se obtiene el estimador, este es un método no paramétrico el cual no requiere de hacer supuestos sobre la naturaleza de la distribución (Normal, Binomial, de Poisson, etc.), la media o la desviación estándar, de una muestra finita X_1, \dots, X_n de n observaciones reales de un proceso estocástico, es decir, los datos deben de ser continuos y aleatorios (Minnotte, 1997).

Los estimadores de densidad de tipo kernel desde el punto de vista matemático poseen un cierto número de propiedades interesantes, los cuales se describen a continuación:

1. Son funciones simétricas centradas en cero y su integral equivale a la unidad.
2. El kernel es una función de densidad, el estimador \hat{f} también lo es.
3. En la mayoría de los casos se tratan de funciones continuas y diferenciables.
4. Los kernel siempre son positivos.

El estimador de densidad de tipo kernel tiene dos componentes básicos, el núcleo y el ancho de banda. El núcleo es la función de peso por el cual se estima la densidad y existen siete tipos comunes de funciones para datos no paramétricos, las cuales son: gaussiano o normal, Epanechnikov, uniforme, triangular, cuarto grado y coseno triponderado. Se puede elegir una de estas funciones de densidad para estimar la frecuencia de una variable en un conjunto de datos (Cheng & Ray, 2014). El ancho de banda o ventana suaviza los datos resultantes de la función kernel, la cual está representada por la letra h y el valor de esta debe ser positivo.

Se define la estimación de tipo núcleo K (Silverman, 1981), como se describe en la ecuación 2.2:

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (2.2)$$

Donde h es el ancho de ventana o parámetro de suavizado, x_i es el valor de la variable, x es el punto en el cual se trata de estimar la densidad, n representa el tamaño de la muestra y K representa la función kernel utilizada.

Un kernel es una función integrable, no negativa, con valores reales, simétrica y unimodal, denotada por K como se expresa en la ecuación 2.3, la cual se denomina función de núcleo y satisface (Parzen, 1962):

$$\int_{-\infty}^{\infty} K(u) du = 1 \quad (2.3)$$

La función $K(u)$ denominada kernel, núcleo o peso, descrita por la ecuación 2.4 debe satisfacer ciertas condiciones de regularidad, generalmente es una función de densidad simétrica como una distribución normal, y h es una secuencia de constantes positivas conocida como ancho de banda (Hall & York, 2001).

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \quad (2.4)$$

Donde $u = \left(\frac{x - x_i}{h}\right)$, x es el punto por el cual se le trata de estimar la densidad y h es el ancho de banda.

El estimador de la densidad de núcleo se representa como la suma de cada una de las curvas gaussianas y cada una está centrada en una observación, como se describe en la figura 2.2, la cual presenta la construcción de un estimador de densidad no paramétrico (línea sólida), el valor promedio de las funciones kernel (el kernel gaussiano está representado por las líneas punteadas) que muestra los núcleos individuales y los datos centrados en los puntos de cada una de las observaciones. Para un estimador de núcleo de tipo gaussiano el ancho de banda es el mismo que la desviación estándar $(x - x_i)$ (Bianchi, 1995).

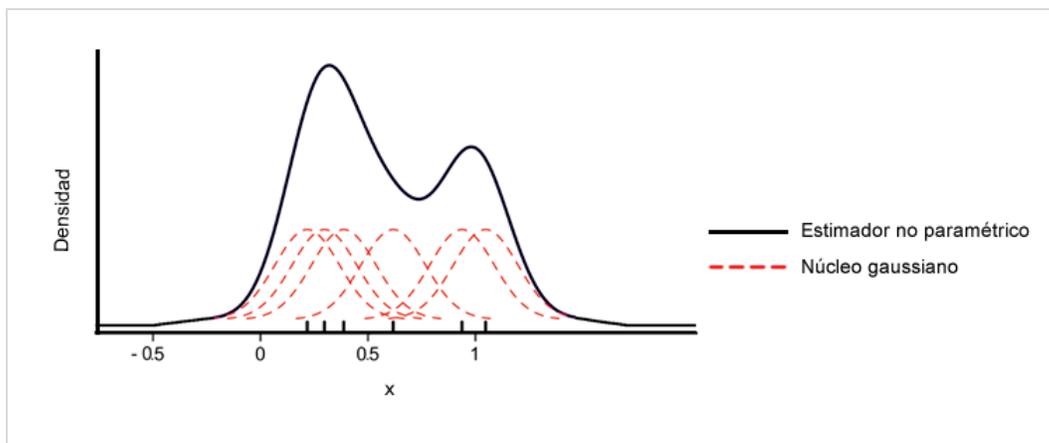


Figura 2.3. Estimador de densidad mostrando los núcleos individuales.

El cálculo del ancho de banda h no es una tarea sencilla ya que su valor óptimo depende en gran medida de la naturaleza de los datos de entrada y el número de patrones (Hall & York, 2001). Si se selecciona el ancho de banda de un estimador de núcleo, sería óptimo para una función de densidad de probabilidad (PDF) particular, y no se asume que $f(x)$ es normal, pero en lugar de seleccionar una h la cual podría ser óptima si la PDF fuera normal usando un núcleo gaussiano,

entonces el ancho de banda óptimo h está definido por la ecuación 2.5 (Bourel, 2011):

$$h = \sigma * \sqrt[5]{\frac{4}{3N}} \quad (2.5)$$

Dado que la distribución normal es unimodal y con forma de campana, se espera que h será demasiado grande para este tipo de distribuciones, generalmente con valores aproximados a la unidad, con media cero y varianza σ^2 , esta será adecuada si la población tiene una distribución normal. Si se trabaja con poblaciones multimodales se producirá una sobre suavización de la estimación del valor h , aunque sea pequeño resultaran algunos valores atípicos presentando unimodalidad en las observaciones con lo que se asume que el estimador gaussiano no es del todo confiable (Hall & Huang, 2002). Por lo anterior, se recurre al estimador de ancho de banda óptimo propuesto por (Silverman, 1981) el cual está descrito por la ecuación 2.6. Asumiendo que la distribución de los datos posee una densidad de probabilidad $\hat{f}(x)$, se propuso la regla general de carácter práctico para calcular el valor h , el cual aproxima el ancho de banda y se utiliza para contrastar la multimodalidad de una distribución de datos. Este procedimiento consiste en encontrar la ventana (h, k) más pequeña de forma que la estimación $f(h, k)$ para ese parámetro de ventana tenga k modas (Sheather, 2004).

$$h = 0.9 * \min \left(STD, \left(\frac{IRQ}{1.349} \right) \right) * \frac{1}{\sqrt[5]{n}} \quad (2.6)$$

Donde IRQ es el rango intercuartil, STD es la desviación estándar y n es el tamaño de la muestra.

Se asume que el estimador de densidad de los datos para valores pequeños de h presenta una distribución multimodal y para valores cercanos a la unidad h da como resultado una distribución unimodal.

2.9 Prueba de hipótesis estadística de Kolmogorov-Smirnov

Mediante el uso de pruebas estadísticas se puede demostrar el grado de confianza y la factibilidad de las hipótesis planteadas. La prueba de hipótesis permite decidir si un grupo de observaciones se ajustan a una distribución de probabilidad de un conjunto de valores de una muestra dada y una distribución específica (Facchinetti, 2009). El método no paramétrico sencillo para medir el grado de concordancia existente entre la distribución y una distribución teórica específica, es la prueba de bondad de ajuste de Kolmogorov-Smirnov que es conocida como prueba K-S (Pedrosa, Juarros Basterretxea, Robles Fernández, Basteiro, & García Cueto, 2014).

El contraste de K-S se basa en calcular una distancia entre la función de distribución empírica de los datos, F_n , y la función de F_0 postulada bajo H_0 , como se describe en la ecuación 2.7:

$$D = \max |F_n(x) - F_0(x)| \quad (2.7)$$

Siendo $F_n(x)$ la distribución muestral, $F_0(x)$ la función teórica o correspondiente a la población normal especificada en la hipótesis nula, x es el valor observado en

las muestras previamente ordenadas de menor a mayor (Stephens, 1970). Dando como resultado:

Hipótesis Nula H_0 : El conjunto de datos sigue una distribución normal.

Hipótesis Alternativa H_1 : El conjunto de datos no sigue una distribución normal.

En otras palabras, determina la existencia o no de unimodalidad en la muestra.

Capítulo 3 . Estado del arte

En la actualidad existen un conjunto de exámenes que permiten medir habilidades en los individuos para evaluar la presencia de posibles trastornos neurológicos y cognitivos relacionados con problemas de percepción, de memoria, motrices, de atención o de memoria visual inmediata (Mitrushina, Boone, Razini, & D'Elia, 2005). Por tal motivo, es necesario contar con instrumentos que ayuden a estimar el nivel del constructo a medir y que permitan dar una interpretación adecuada a los resultados obtenidos en los exámenes mediante un conjunto de valores numéricos. Los niveles de precisión de los instrumentos se deberán interpretar bajo ciertos parámetros de confiabilidad y validez para que sus interpretaciones sean acertadas y den un resultado más preciso. Para ello se requiere tener un conjunto de mediciones que permita hacer esto de manera adecuada, es decir, que el instrumento utilizado sea de tal forma que la estimación de las habilidades de un examinado no dependa del resto de

los examinados, ni de los ítems que componen al examen (Jiménez Alfaro & Montero Rojas, 2013).

La teoría de respuesta al ítem ayuda a estimar el nivel de habilidad de un examinado mediante la aplicación de un examen, el cual permite obtener un índice que describa un nivel de confiabilidad de cada una de las puntuaciones obtenidas en el examen realizado, lo que permite clasificar la calidad de cada ítem y por consiguiente de todo el examen en cuestión.

Mediante el uso de un conjunto de modelos matemáticos denominados modelos logísticos paramétricos que permiten relacionar el nivel de habilidad del examinado y el nivel de dificultad de un ítem se puede obtener el nivel de probabilidad de que un examinado responda correctamente a cada uno de los ítems de un examen y expresar esta probabilidad mediante su representación gráfica en la CCI (Cupani & Cortez, 2016).

La teoría de respuesta al ítem es muy extensa por lo que este trabajo está enfocado al análisis de resultados de una evaluación que mide habilidades cognitivas mediante ítems dicotómicos, es decir, a estos ítems se les asigna una calificación donde la respuesta correcta es uno y cero si la respuesta es incorrecta, utilizando la misma escala de puntuación, lo que ayuda a la elaboración de exámenes individuales que se ajusten a la habilidad de cada sujeto examinado y al nivel de dificultad de cada uno de los ítems que componen el examen.

En este trabajo de investigación se recurre a analizar los resultados obtenidos de la prueba de copia y reproducción de la figura compleja de Rey-Osterrieth, la cual está compuesta por 18 ítems, la prueba consiste en copiar y después en reproducir

un dibujo geométrico complejo, esta prueba es aplicada de manera individual y de tiempos variables en niños a partir de los cuatro años y adultos.

3.1 Prueba de la figura compleja de Rey-Osterrieth

Para la evaluación de aspectos en la educación y neuropsicología en los individuos se recurre a un conjunto de pruebas estandarizadas que miden las habilidades cognitivas, motrices, de memoria visual, de organización gráfica de objetos y de la percepción del entorno. Una de ellas es la prueba de la figura compleja de Rey-Osterrieth, la cual consiste en copiar y después de un lapso de tiempo reproducir una figura geométrica compleja (Cortés S, Galindo y Villa M, & Salvador C, 1996).

3.1.1 Antecedentes

La prueba de copia y reproducción de la figura compleja de Rey fue diseñada inicialmente por André Rey en 1944, con el fin de evaluar la percepción y la memoria visual en individuos con lesión cerebral. El primer estudio psicométrico lo realizó Paul Alexandre Osterrieth quien aplicó la prueba a una muestra de 295 individuos de 4 años hasta la edad adulta (Cortés S, Galindo y Villa M, & Salvador C, 1996). Osterrieth propuso categorizar la figura en 18 elementos y anotarlos basándose en que tan compleja estaba y la colocación correcta, posteriormente se

ha utilizado para valorar la presencia de posibles trastornos neurológicos y cognitivos relacionados a problemas de percepción, memoria, visio-espaciales y motriz. Actualmente es una herramienta muy utilizada en la evaluación del trastorno por déficit de atención e hiperactividad.

3.1.2 Elementos de evaluación de Rey-Osterrieth

La prueba consiste en copiar y después reproducir un dibujo geométrico complejo, el cual está representado en la figura 3.1, por diversas formas, segmentos de líneas y otros elementos (Mitrushina, Boone, Razini, & D'Elia, 2005).

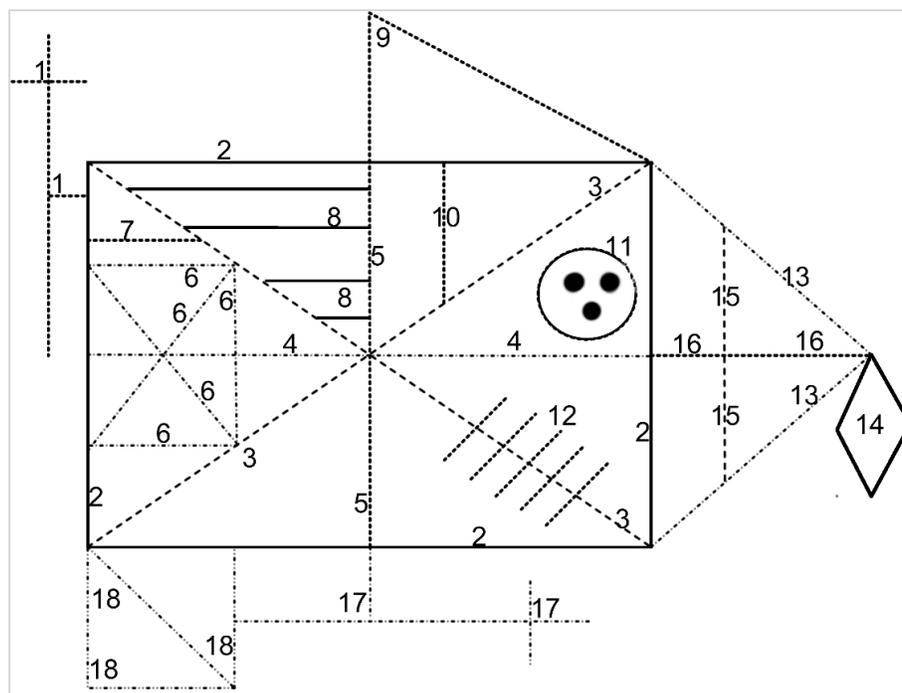


Figura 3.1. Figura compleja de Rey-Osterrieth.

La figura reúne las siguientes propiedades (Rey, 1987):

1. Ausencia de significado evidente.
2. Fácil realización gráfica.
3. Estructura de conjunto bastante complicada para exigir una actividad analítica y organizadora.

Los 18 elementos que se evalúan en la prueba de Rey-Osterrieth son:

1. Cruz exterior contigua al ángulo superior izquierdo del rectángulo grande.
2. Rectángulo grande, armazón de la figura.
3. Cruz de San Andrés, formada por las diagonales del rectángulo grande.
4. Mediana horizontal del rectángulo grande 2.
5. Mediana vertical del rectángulo grande 2.
6. Pequeño rectángulo interior (contiguo al lado izquierdo del rectángulo 2, limitado por las semidiagonales izquierdas de éste y cuyas diagonales se cortan sobre la mediana 4).
7. Pequeño segmento colocado sobre el lado horizontal superior del elemento 6.
8. Cuatro líneas paralelas situadas en el triángulo formado por la mitad superior de la diagonal izquierda del rectángulo 2.
9. Triángulo rectángulo formado por la mitad del lado superior del rectángulo 2, la prolongación hacia arriba de la mediana vertical 5 y el segmento que une el extremo de esta prolongación con el ángulo superior derecho del rectángulo 2.
10. Pequeña perpendicular al lado superior del rectángulo 2, situada debajo del elemento 9.

11. Círculo con tres puntos inscritos, situado en el sector superior derecho del rectángulo 2.
12. Cinco líneas paralelas entre sí y perpendiculares a la mitad inferior de la diagonal derecha del rectángulo 2.
13. Dos lados iguales que forman el triángulo isósceles construido sobre el lado derecho del rectángulo 2 por la cara exterior de éste.
14. Pequeño rombo situado en el vértice extremo del triángulo
15. Segmento situado en el triángulo 13 paralelamente al lado derecho del rectángulo 2.
16. Prolongación de la mediana horizontal y que constituye la altura del triángulo 13.
17. Cruz de la parte inferior comprendiendo en ella el brazo paralelo al lado inferior del rectángulo 2 y la pequeña prolongación de la mediana 5 que la une a este lado.
18. Cuadrado situado en el extremo inferior izquierdo del rectángulo 2, prolongación del lado izquierdo, comprendiendo también su diagonal.

3.1.3 Calificación de la prueba

Se trata de una prueba de aplicación individual y de tiempo variable, las unidades de puntuación en el área o detalles correctos de la figura han sido numeradas como se indicó en la figura 3.1 para facilitar su puntuación.

La aplicación de la prueba consiste en solicitar al sujeto examinado que copie la figura compleja a mano y sin límite de tiempo, posteriormente, sin previo aviso y

sin ayuda del modelo, el examinado debe reproducir inmediatamente de nuevo la misma figura con el fin de evaluar su capacidad de recuerdo material no verbal.

Puntuación:

- Cada unidad puede hacer que el evaluado gane hasta dos puntos.
- Se dan dos puntos cuando el dibujo esté correcto en la forma y ubicación.
- Se da un punto cuando falla en la ubicación o forma de la unidad.
- Se dan cero puntos cuando no logra colocar la ubicación ni la forma de la unidad.
- En cada aplicación de la prueba los examinados pueden obtener una puntuación entre 0 y 36 puntos. La puntuación media de un adulto es igual a 32 puntos y la sesión de memoria es puntuada de la misma manera.

Al realizar las comparaciones de las puntuaciones obtenidas en la sesión de copia y en la de reproducción ayudará a determinar la presencia de problemas relacionados a la percepción, déficit visio-gráfico y de memoria a corto plazo en los evaluados.

Al calificar las pruebas se cuenta con limitaciones para poder hacer un análisis cualitativo y cuantitativo confiable, pues admite una fuerte variación importante determinada a la subjetividad de la persona que califica la prueba (Rey, 1987).

3.2 Modelos psicométricos de IRT

Para medir el nivel de conocimientos, destrezas y habilidades se realiza mediante un conjunto de evaluaciones, la más común es mediante la aplicación de un examen con preguntas de opción de respuesta múltiple, es decir, solo una respuesta correcta y uno o varios distractores. Una vez resuelto el examen es necesario dar una interpretación a las puntuaciones a través de un promedio o calificación que relacione a los evaluados con las variables de interés a medir, ya que son difíciles de medir de manera directa en el contexto de la IRT (Barbero García, 1999).

Mediante un modelo matemático denominado función de respuesta del ítem es como la IRT se utiliza para el análisis de cada uno de los ítems que conforman al examen aplicando los modelos logísticos de uno, dos, tres y cuatro parámetros, ya que son los más conocidos y están basados en cuatro características diferentes las cuales están en función del número de parámetros que se pueden tener en cuenta o no.

El número de parámetros de los modelos logísticos está definido por el nivel de dificultad, el poder de discriminación del ítem, el acierto por azar y el error por descuido (de Ayala, 2009), donde cada uno describe la relación que existe entre el nivel de habilidad de un examinado y la probabilidad de emitir una respuesta correcta a cada uno de los ítem de un examen y se describen en los siguientes apartados (Simanca Herrera, Abuchar Porras, & Rivera Osorio, 2014).

3.2.1 Modelo logístico de un parámetro

El modelo propuesto por (Rasch G. , 1960) o modelo logístico de un parámetro (1PL) permite estimar el nivel de probabilidad de que un examinado conteste de manera correcta a un ítem en un examen. A partir de la interrelación que existe entre un conjunto de ítems y los examinados se pueden ubicar los resultados en un mismo nivel de puntuación en relación con el constructo que se desea medir y se basa en los siguientes supuestos (Hambleton, Swaminathan, & Rogers, 1991):

- El atributo que se desea medir se representa en una única dimensión en la que se sitúan conjuntamente a los examinados e ítems y está expresado en términos de la función logística o también llamada escala logit (Jiménez Alfaro & Montero Rojas, 2013).
- El nivel de habilidad del examinado en el rasgo latente y la dificultad del ítem determinan la probabilidad de que la respuesta sea correcta. Si el control de la situación es adecuado, esta expectativa es razonable y así se debe representar en el modelo matemático elegido (Prieto & Delgado, 2003).

El modelo 1PL establece la relación que existe entre el nivel de habilidad del examinado y la dificultad del ítem, solo tiene en cuenta como único parámetro el nivel de dificultad, es decir, se está midiendo al examinado y al ítem en una misma escala. La escala del modelo 1PL está representada por $(\theta - b)$, donde la probabilidad de respuesta correcta a un ítem es igual a la habilidad del examinado menos la dificultad del ítem, como se establece en la ecuación 3.1. Esta escala es llamada logit que se refiere al logaritmo del momio el cual expresa la

probabilidad de que el examinado cuente con un nivel de habilidad entre la probabilidad de que no se cuente con ella, es decir, de la probabilidad de éxito entre la probabilidad de fallo al ítem en cuestión, la cual puede tener valores de $(-\infty, \infty)$. En la práctica los valores obtenidos se encuentran en el rango de puntuación $(-4,4)$ con un valor central en cero (de Ayala, 2009), (Baker, 1985).

$$\text{logit}(P(\theta)) = \ln \frac{P(\theta)}{1 - P(\theta)} = \theta - b_i \quad (3.1)$$

La ecuación 3.2 describe el modelo de Rasch o 1PL:

$$P_{ij}(x = 1 | \theta_j, b_i) = \frac{1}{1 + e^{-(\theta_j - b_i)}} \quad i = 1, 2, \dots, n \text{ y } j = 1, 2, \dots, m \quad (3.2)$$

Donde $P_{ij}(x = 1 | \theta_j, b_i)$ es el valor de la probabilidad de respuesta correcta al ítem i del alumno j dados los parámetros: b_i que representa el parámetro de dificultad del ítem i , n es el número de ítems del examen, m es el número de alumnos examinados, el número e es la base del logaritmo neperiano 2.718.

El parámetro b_i está definido como la dificultad del ítem i e indica la posición en la CCI en la escala de habilidad y se define como el punto en donde la probabilidad de respuesta correcta y respuesta incorrecta es de 0.5 expresado en un nivel de habilidad media, es decir, cuando el nivel de dificultad del ítem es igual al nivel de habilidad del examinado la probabilidad de respuesta es igual a $1/2$.

El nivel de dificultad del ítem se representa en un punto de la escala de habilidad justo en que la probabilidad de éxito $P(\theta)$ es igual a la probabilidad de fallo $1 - P(\theta)$. La figura 3.2 representa a un evaluado con un nivel de habilidad θ y es igual

a la dificultad b del ítem i , el cual tendrá una probabilidad de 0.50 de contestar correctamente al ítem, por lo tanto, una probabilidad de 0.50 de contestarlo de manera incorrecta (Baker, 1985). El punto de la escala de habilidad $\theta = 0$ determina el umbral que divide a la escala en dos partes. Los evaluados con un nivel de habilidad menor a la dificultad del ítem tendrán una probabilidad de respuesta menor a 0.5, por lo tanto, una mayor probabilidad de fallo. Los evaluados con una mayor habilidad que la dificultad del ítem tendrán una probabilidad de respuesta mayor a 0.5 y en consecuencia una menor probabilidad de fallo, esta idea de umbral define la lógica al usar los modelos paramétricos logísticos de la IRT.

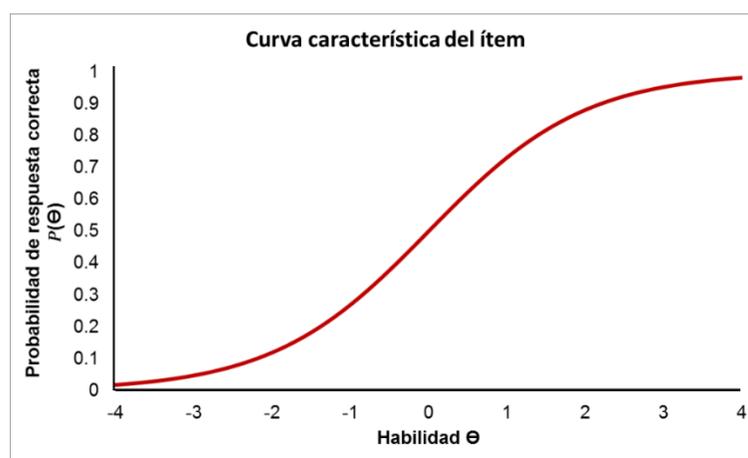


Figura 3.2. CCI con el parámetro $b = 0$ aplicado al modelo 1PL.

El modelo supone que la dificultad del ítem es la única característica que influye en la respuesta del examinado y el nivel de discriminación es igual a uno para todos los ítems de la prueba (Martínez Arias, Hernández Lloreda, & Hernández Lloreda, 2014).

Si se analizan cuatro ítems mediante el modelo 1PL y con distinto nivel de dificultad que se expresa con los siguientes valores para el ítem 1 $b_1 = -2.0$, ítem 2 $b_2 = 0$, ítem 3 $b_3 = 1.0$ y el ítem 4 $b_4 = 2.5$, cómo se observa en la figura 3.3, el nivel de dificultad del ítem 1 se encuentra ubicado al lado izquierdo de la escala con habilidad media $\theta = 0.88$, el ítem 2 se encuentra en medio y la CCI tiene un valor de probabilidad $\theta = 0.50$ de contestar correctamente, el ítem 3 indica que tiene un nivel de habilidad media $\theta = 0.26$ y el ítem 4 con nivel de dificultad $b_4 = 2.5$ se encuentra a la derecha con un nivel de habilidad media $\theta = 0.08$. Como se puede apreciar en las CCI de la figura 3.3 los examinados con niveles de habilidad tienen menor probabilidad de acertar a las respuestas de los ítems. El ítem 1 cuya dificultad $b_1 = -2.0$ es más fácil de responder que el ítem 4 con un nivel de dificultad $b_4 = 2.5$ en cualquier punto de la escala de habilidad. La CCI del ítem 1 presenta menor dificultad que el ítem 4 correspondiente ya que los evaluados con menor capacidad tienen mayor probabilidad de contestarlo correctamente.

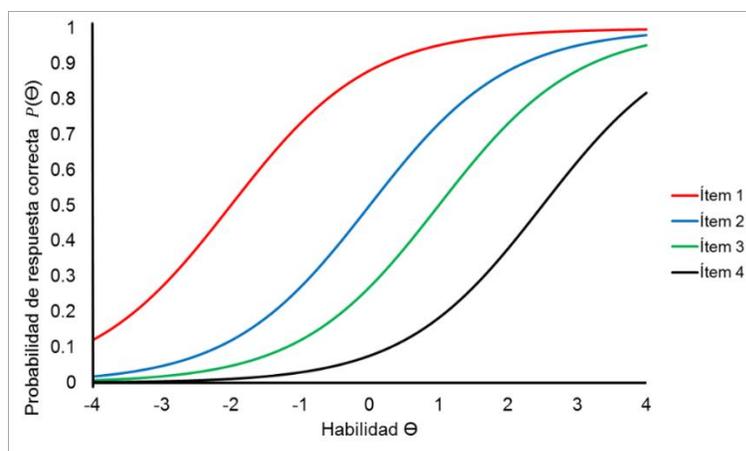


Figura 3.3. CCI de 4 ítems aplicando el modelo 1PL.

Las curvas solo difieren en su localización en la escala de habilidad y en este modelo las CCI de los cuatro ítems con niveles de dificultad diferentes no se cruzan en ningún punto.

Las ventajas del modelo de Rasch al aplicar un examen basado en IRT son las siguientes:

- La medición conjunta se refiere a que los parámetros de los examinados y de los ítems se expresan en las mismas unidades, es decir, no todos los ítems miden la cantidad del constructo, lo que permite analizar las interacciones entre los examinados y los ítems. La interpretación de las puntuaciones no se fundamenta necesariamente en grupos de sujetos evaluados, sino identificando a los ítems en que los examinados tienen una alta o baja probabilidad de responder correctamente en un examen (Attorresi, Lozzia, Abal, Galibert, & Agerri, 2009).
- Una medida puede ser considerada válida y generalizable si no depende de las condiciones específicas con que ha sido obtenida, esta propiedad fue denominada objetividad específica y se describe como la diferencia entre dos sustentantes en un atributo que no deben depender de los ítems específicos con los que se estiman (Hambleton, Swaminathan, & Rogers, 1991).
- La propiedad de intervalo establece que a diferencias iguales entre un individuo y un ítem le corresponden probabilidades idénticas de una respuesta correcta, es decir, diferencias iguales en el constructo están asociados a diferencias iguales en los puntajes (Muñiz, 2010).

El modelo de Rasch permite:

- a) Cuantificar la cantidad de información y también la cantidad de error con la que se mide en cada punto de la dimensión.
- b) Seleccionar aquellos ítems que permiten incrementar la información en regiones del constructo previamente especificadas, por ejemplo, para un examen de admisión a la universidad se desea seleccionar a los examinados con un alto nivel en el constructo, por lo que se pueden utilizar los ítems con mayor nivel de dificultad (Jiménez Alfaro & Montero Rojas, 2013).

3.2.3 Modelo logístico de dos parámetros

El modelo logístico de dos parámetros (2PL) o modelo de (Birnbaum, 1967), además del nivel de dificultad, supone un parámetro que representa el poder de discriminación del ítem, el cual hace referencia a la capacidad del ítem mediante la diferencia entre los examinados con distintos niveles de rasgos, el cual tendrá valores altos a mayor capacidad de discriminación del ítem (Hidalgo Montesinos & French, 2016).

El parámetro a_i es el poder de discriminación del ítem e indica la precisión para determinar la habilidad de los examinados, es decir, distingue a los evaluados con un nivel de habilidad baja y los evaluados con un nivel de habilidad alta. Este parámetro es proporcional a la pendiente de la CCI en el punto b_i de la escala de habilidad e indica la razón de cambio en la probabilidad de respuesta correcta cuando hay un cambio en el nivel de habilidad de los examinados (Brizuela Rodríguez, 2015).

La ecuación 3.3 define al modelo 2PL:

$$P_{ij}(x = 1 | \theta_j, b_i, a_i) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad i = 1, 2, \dots, n \quad j = 1, 2, \dots, m \quad (3.3)$$

Donde $P_{ij}(x = 1 | \theta_j, b_i, a_i)$ es el valor de la probabilidad de respuesta correcta al ítem i del alumno j dado los parámetros: b_i que representa la dificultad del ítem i , a_i es el valor de discriminación del ítem i , n es el número de ítems del examen, m es el número de alumnos examinados, el número e es la base del logaritmo neperiano 2.718.

Como se observa en la figura 3.4 dado un nivel de dificultad $b_i = 0$ y el nivel de discriminación que toma el valor de $a_i = 1.3$, si el valor de a_i incrementa la CCI será más inclinada, el rango teórico de valores de a_i comprende de $(-\infty, \infty)$, aunque suele estar comprendida entre $(-2.80, 2.80)$ para fines prácticos (Baker, 1985).

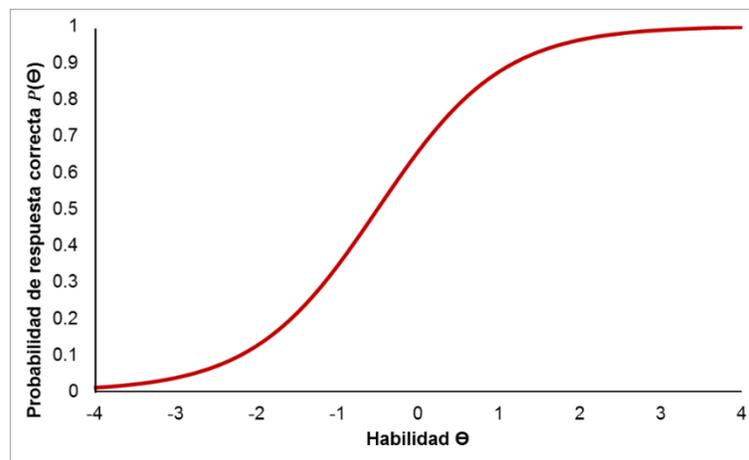


Figura 3.4. CCI aplicando el modelo 2PL.

En la figura 3.5 se muestran cuatro CCI usando un modelo 2PL, en los cuales se mantiene el nivel de dificultad constante $b_n = 0$, es decir, los examinados con un nivel de habilidad $\theta = 0$, donde tienen una probabilidad de responder de forma correcta de 0.5 y los valores del nivel de discriminación están comprendidos por los siguientes valores cada uno de los ítems representados por $a_1 = 0$, $a_2 = 1.0$, $a_3 = -2.0$ y $a_4 = 2.0$.

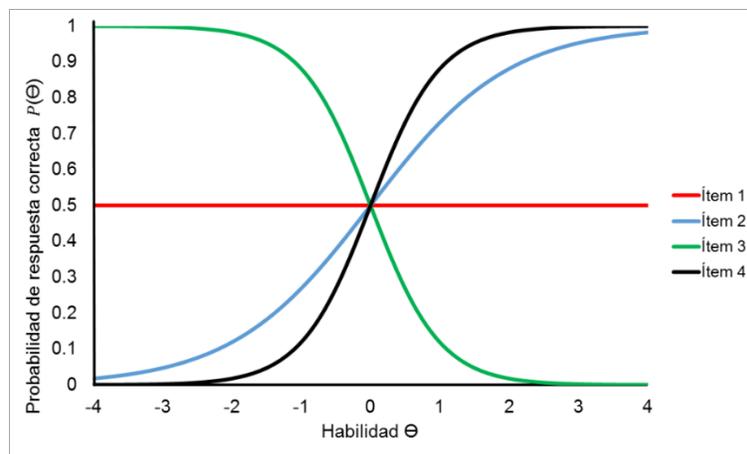


Figura 3.5. CCI con modelo 2PL con $b = 0$ y distintos valores para a .

Se observa en cada una de las CCI del modelo 2PL que tienen un comportamiento diferente si se presenta un valor del parámetro de discriminación distinto, para el valor de $a_2 = 1.0$ la curva tendrá una menor pendiente que cuando $a_4 = 2.0$, lo cual indica que en el primer caso se requiere un incremento mayor en el nivel de la habilidad para aumentar el poder de discriminación del ítem, mientras que en el caso de la curva del ítem 4 con valor $a_4 = 2.0$ presenta un cambio mucho más orientado hacia una mayor probabilidad de contestar en forma correcta. En el caso de la curva del ítem 1 con valor $a_1 = 0$ todos los individuos tendrían la misma

probabilidad de contestar al ítem independientemente de su nivel de habilidad, mientras que si se tiene un nivel de discriminación $a_3 = -2$ el ítem está presentando un nivel inverso al esperado en que los individuos con menor habilidad tienen más probabilidad de contestar correctamente que los examinados de mayor habilidad.

En general se busca que todos los ítems tengan valores de índice de discriminación positivos, este se puede calcular mediante la correlación biserial entre la calificación del ítem y la puntuación total del examen (Baker, 1985), (Muñiz, 2010).

El parámetro de discriminación a es un poco más difícil de interpretar que el nivel de dificultad especialmente cuando se trata de juzgar la calidad de los ítems que se van a retener en un banco de ítems para una evaluación definitiva (Martínez Arias, Hernández Lloreda, & Hernández Lloreda, 2014).

3.2.4 Modelo logístico de tres parámetros

Una de las características de los ítems es que tienen una probabilidad asociada a ser respondidos correctamente por azar, es decir, cuando se acierta sin que se tenga los conocimientos para responderlo de manera correcta (Barbero García, 1999), ejemplo: un ítem con tres opciones de respuesta tiene una probabilidad de 0.33 de ser contestado de manera correcta por azar mediante adivinación simple, un ítem con cuatro opciones de respuesta tendrá una probabilidad de 0.25 de responder correctamente.

El modelo logístico de tres parámetros incorpora la posibilidad de que la respuesta correcta haya sido acertada al azar, agregando el parámetro c al modelo 2PL de acuerdo a la ecuación 3.4 (Baker, 1985), (de Ayala, 2009):

$$P_{ij}(x = 1 | \theta_j, b_i, a_i, c_i) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad (3.4)$$

$$i = 1, 2, \dots, n \quad j = 1, 2, \dots, m$$

Donde $P_{ij}(x = 1 | \theta_j, b_i, a_i, c_i)$ es el valor de probabilidad de respuesta correcta del ítem i del alumno j dado los parámetros: b_i que representa la dificultad del ítem i , a_i es la discriminación del ítem i , c_i es el valor al acertar mediante azar al ítem i , n es el número de ítems del examen, m es el número de examinados y el número e es la base del logaritmo neperiano 2.718.

Si se indica que existe una probabilidad fija igual a c_i , que es independiente al nivel de habilidad del examinado, normalmente el parámetro c_i toma valores menores a la probabilidad estimada para el acierto por azar al ítem. Se le atribuye a que los examinados con habilidad baja normalmente son atraídos a seleccionar alguna de las respuestas incorrectas o distractores por lo que podrían obtener un puntaje mayor contestando aleatoriamente al ítem, es la razón por la que el parámetro c_i es denominado pseudo-azar y no simplemente acierto por azar ya que este valor se refleja en la asíntota inferior de la CCI.

En los modelos de uno y dos parámetros la probabilidad de respuesta correcta se aproximaba a cero cuando la habilidad de los examinados tiende a valores muy bajos ($-\infty$), para el modelo 3PL la probabilidad de respuesta correcta se aproximará al valor de c_i para valores de θ bajos (Brizuela Rodríguez, 2015). Como se aprecia en la figura 3.6 donde se representa un modelo 3PL, el valor del

parámetro de pseudo-azar está localizado en $c = 0.2$, el parámetro de dificultad $b = 1.3$ y con nivel de discriminación $a = 1.03$.

En el modelo 3PL la probabilidad de respuesta correcta para un nivel de habilidad igual a b_i ya no será $\theta = 0.50$ ya que refleja que existe un desplazamiento de la asíntota inferior de la CCI mediante un incremento en el nivel de probabilidad $P(\theta)$ dado por la expresión $(1 + c_i)/2$ (Brizuela Rodríguez, 2015).

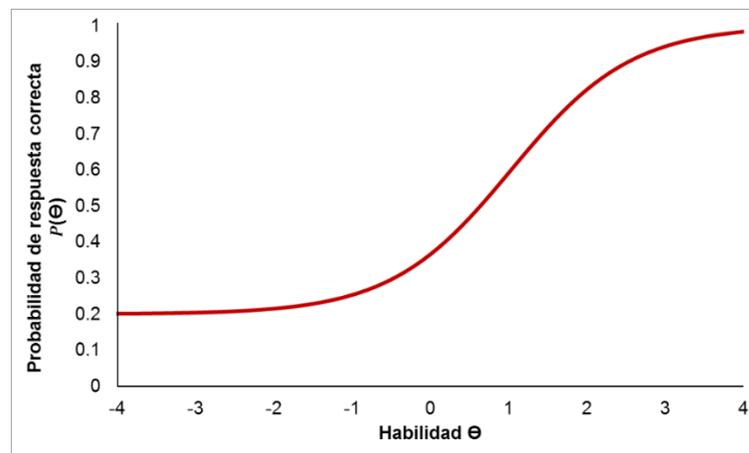


Figura 3.6. CCI aplicando el modelo 3PL.

En la figura 3.7 se muestran a cuatro ítems utilizando el modelo 3PL, en donde los valores de los parámetros $b_n = 0$ y $a_n = 1.0$ son constantes para cada uno de los ítems y con diferentes valores para el parámetro que refleja la probabilidad de tener una respuesta correcta cuando se acierta al azar, los valores están representados por $c_1 = 0.05$, $c_2 = 0.15$, $c_3 = 0.26$ y $c_4 = 0.35$. Se observa que los valores del parámetro c al estar en un rango comprendido entre 0.05 y 0.35 propician el aumento en el nivel de probabilidad de respuesta correcta aun

cuando se tienen niveles de habilidad bajos y se visualiza en la gráfica que la asíntota inferior ya no se aproximará a cero cuando se tienen niveles de habilidades θ cercanos a -4 .

Cuando se tienen los niveles de habilidades y la probabilidad de respuesta correcta, lo que se busca es modelar la probabilidad del parámetro de pseudo-azar aun cuando se conoce la respuesta ya que no se requiere diferenciar el nivel de probabilidad de la respuesta y de los distractores de un mismo ítem, solo se analiza la información dicotómica de la respuesta lo que conlleva a una mejor estimación del nivel de habilidad de los evaluados.

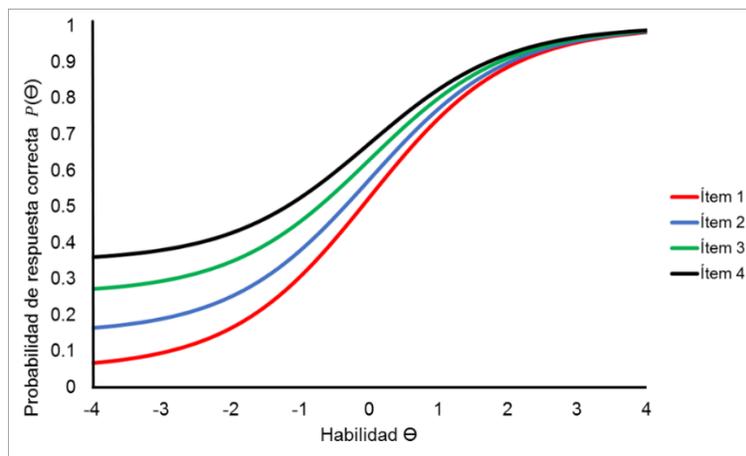


Figura 3.7. CCI con modelo 3PL con distinto valor en los parámetros a , b y c .

3.2.5 Modelo logístico de cuatro parámetros

El modelo logístico de cuatro parámetros es una extensión del modelo 3PL el cual pretende explicar aquellos casos en los que los examinados tienen un nivel de aptitud alto y tienen una probabilidad de no responder correctamente al ítem, es denominado error por descuido, es decir, cuando la probabilidad de que un examinado con un nivel de habilidad alto responde mal al ítem, este fallará así una respuesta que conoce ya sea por falta de interés, cansancio, descuido o por cualquier otra razón, es posible estimar esta falta de cuidado mediante un modelo 4PL el cual describe los parámetros en la ecuación 3.5 y se representa por:

$$P_{ij}(x = 1 | \theta_j, b_i, a_i, c_i, \gamma_i) = c_i + (\gamma_i - c_i) \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad (3.5)$$
$$i = 1, 2, \dots, n \quad j = 1, 2, \dots, m$$

Donde $P_{ij}(x = 1 | \theta_j, b_i, a_i, c_i, \gamma_i)$ es el valor de la probabilidad de respuesta correcta de al ítem i del alumno j dado los parámetros: b_i que representa la dificultad al ítem i , a_i es la discriminación del ítem i , c_i representa el valor al acertar mediante pseudo-azar al ítem i , γ_i es el error por descuido o pseudo-fallo, n es el número de ítems del examen, m es el número de alumnos examinados y el número e es la base del logaritmo neperiano 2.718.

Se observa en la figura 3.8 que el parámetro $\gamma = 0.9$ cuando los examinados con un nivel alto de habilidad no llegan a contestar correctamente a los ítems del examen el valor de la asíntota superior es menor a 1.

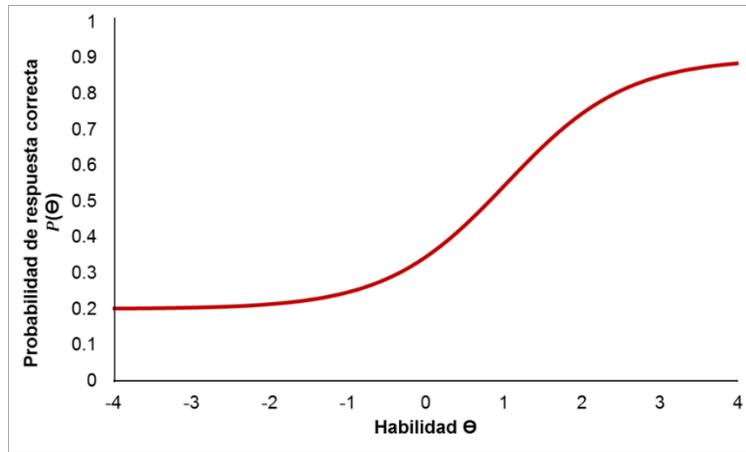


Figura 3.8. CCI aplicando el modelo 4PL.

En la figura 3.9 se observa a cuatro ítems utilizando el modelo 4PL, en donde los valores de los parámetros $b_n = 0$, $a_n = 1.0$ y $c_n = 0.20$ son constantes para cada uno de los ítems y con diferentes valores para el parámetro que refleja la probabilidad de tener una respuesta correcta cuando se falla a los ítems que de antemano se conoce la respuesta y los valores están representados por $\gamma_1 = 0.95$, $\gamma_2 = 0.90$, $\gamma_3 = 0.80$ y $\gamma_4 = 0.75$.

Se observa que para un nivel de habilidad de $b = 0$, con discriminante $a = 1.0$ y acierto por pseudo-azar $c = 0.20$, la probabilidad de respuesta correcta aumenta cuando se aproxima a uno la asíntota superior de los valores de γ_n .

Este modelo se desarrolló para mejorar la estimación del nivel de habilidad del evaluado, si un sujeto con un nivel muy alto tiene un error en un ítem fácil, entonces su estimación se haría drásticamente usando un modelo con una asíntota superior en la CCI con valor igual a uno, que cuando esta asíntota sea menor a uno, el modelo 4PL es el más completo al considerar cuatro parámetros;

el nivel de dificultad, el poder de discriminación del ítem, el acierto por azar y el error por descuido (de Ayala, 2009).

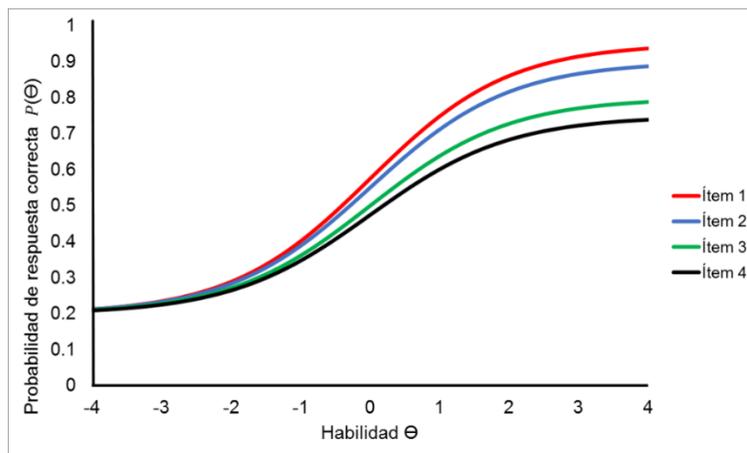


Figura 3.9. CCI del modelo 4PL con diferentes valores para el parámetro γ .

Las primeras investigaciones realizadas de manera empírica del modelo 4PL fueron realizadas por (Barton & Lord, 1981), al agregar un valor menor a uno en la asíntota superior para la estimación de las habilidades en exámenes estandarizados. Su principal motivación fue que el modelo de tres parámetros podría ser excesivamente restrictivo cuando los estudiantes con alto nivel de habilidad obtenían errores en ítems fáciles, el modelo 3PL puede ubicar a un examinado con habilidad baja al adivinar correctamente un ítem difícil, pero en la asíntota superior el valor uno en el modelo de tres parámetros asigna una probabilidad de cero a un examinado con un nivel de habilidad alto que no responda correctamente a un ítem fácil.

A medida que aumenta el número de parámetros en los modelos de IRT el ajuste tiende a ser mejor, pero a costa de las interacciones entre los ítems, es más difícil hacer inferencias a partir de una respuesta de un ítem si la probabilidad de respuesta correcta aumenta. El modelo 4PL ofrece un mecanismo más robusto contra las respuestas erróneas ya que el modelo 3PL causa una estimación con alto valor de sesgo especialmente si los errores ocurren cuando se comienza la prueba en un sistema CAT; sin embargo, ha recibido poca atención en la literatura hasta la fecha. En los modelos 1PL, 2PL y 3PL la capacidad del examinado puede ser subestimada si comete errores por descuido en los primeros ítems de la prueba, el modelo 4PL se establece como un mecanismo de corrección de errores comparándolo con el rendimiento en un examen aplicando un modelo 3PL, por ejemplo, si se tiene la administración de un CAT con arranque donde los dos primeros ítems son intencionalmente evaluados como incorrectos, el modelo 4PL ayudará al examinado a recuperarse de los errores iniciales sin disminuir la precisión y la eficiencia de la medición del examen (Loken & Rulison, 2010).

Este modelo es raramente utilizado en la práctica hasta hace poco como lo expresan (Liao, Ho, & Yen, 2012) se descartaba por ofrecer poco beneficio práctico y que es difícil de estimar, pero las investigaciones más recientes están superando los inconvenientes conceptuales del modelo 4PL, la razón es la mejora de los recursos computacionales para los cálculos junto con el desarrollo de software de modelado estadístico preciso. El trabajo realizado por (Loken & Rulison, 2010) donde desarrollaron un marco bayesiano para calibrar ítems en un entorno CAT mediante un modelo 4PL utilizando el enfoque de Markov Chain Monte Carlo (MCMC) y el software WinBUGS, constituye un gran avance hacia la aplicación del modelo 4PL para fines prácticos.

El parámetro de la asíntota superior que es objeto de estudio en el modelo 4PL permite tener una probabilidad distinta a cero al responder incorrectamente al ítem para los examinados con un nivel de habilidad alto (Magis, 2013). Las pruebas de Eric Loken y Kelly L. Rulison demostraron que el impacto de los errores iniciales cometidos por los evaluados con niveles de habilidad altos pueden reducirse utilizando el modelo 4PL y que se necesitan administrar menos ítems para cancelar el sesgo en la estimación del nivel de habilidad con respecto a la aplicación del modelo 3PL. Los estudios posteriores realizados sobre las ventajas de aplicar el modelo 4PL en los sistemas CAT fueron realizados por (Green, 2011) y (Liao, Ho, & Yen, 2012) y se introduce como línea base del modelo de IRT en el paquete *catR* para la generación de CAT en el lenguaje R desarrollados por (Magis & Raiche, 2012).

3.2.6 Selección de un modelo psicométrico de IRT

La elección de un modelo logístico de la IRT que mejor se ajuste al medir las habilidades de los examinados y el nivel de dificultad de los ítems en un examen es un problema que lleva a pensar cuál es el más adecuado de utilizar. Cualquier elección que se haga *a priori* es válida para quien realiza la interpretación de los resultados, pero será necesario que los datos se ajusten al modelo.

Los aspectos que se analizan al realizar el ajuste de los datos en un modelo paramétrico de la IRT son los siguientes, hay que comprobar los supuestos de unidimensionalidad e independencia local y los parámetros (el nivel de dificultad, el poder de discriminación del ítem, el acierto por azar y el error por

descuido) que son específicos para cada modelo, la estimación del ajuste entre los datos que fueron predichos por el modelo y los datos obtenidos en la muestra, por último hay que comprobar que las ventajas del modelo seleccionado no producen invarianza en los parámetros de los examinados a través de diferentes muestras de ítems y la invarianza de los parámetros de los ítems en diferentes muestras de examinados (Hambleton, Swaminathan, & Rogers, 1991), (Hidalgo Montesinos & French, 2016).

Las características de los ítems ayudan a proporcionar claves que mejoren la elección de un modelo paramétrico de IRT, es poco razonable intentar ajustar un modelo 1PL si se toma en cuenta que los índices de discriminación no son iguales, o si existe un nivel alto de probabilidad de acertarlos por azar donde se asume que los parámetros $a = 1$ y $c = 0$ dado que en este modelo son constantes. Si el valor de acierto por pseudo-azar es $c = 0$ en un modelo 2PL, entonces será más complicado de calcular *a priori*. El modelo 3PL podría ser el ideal a comparación del 1PL y 2PL pues constituyen casos particulares; sin embargo, para el valor del parámetro c su cálculo e interpretación no resulta sencillo, por lo que en la práctica es preferible aceptar que las respuestas del examinado solo dependen de su nivel de habilidad y de la dificultad del ítem.

Después de comprobar que el modelo de IRT elegido presenta un ajuste a los datos se procede a realizar las estimaciones de parámetros de los examinados que fueron producidas por el modelo y que se ve reflejada en la medida de habilidad del rasgo latente y la representación de esta medida se realiza mediante la curva característica del test (Simanca Herrera, Abuchar Porras, & Rivera Osorio, 2014).

Los modelos logísticos se basan en el supuesto de unidimensionalidad, donde se conoce el parámetro del nivel de habilidad θ de un examinado y es suficiente para

conocer el nivel de habilidad ante el ítem que se le presenta, considerando fijo el valor de θ se valora igual la probabilidad de que el examinado acierte al ítem independientemente de que haya contestado de manera correcta o no. Los parámetros estimados deben de ser los mismos en diferentes poblaciones y la dificultad del ítem es invariante ante las transformaciones lineales entre las muestras (Prieto & Delgado, 2010). Un modelo logístico permite predecir la probabilidad de respuesta correcta al ítem para sujetos con características similares pero que no hayan presentado la evaluación (Olea, Abad, & Barrada, 2010).

El modelo 3PL es el más general descrito en la literatura y a partir de este se desprenden los modelos 1PL y 2PL, cuando el parámetro de pseudo-azar se fija en un valor de cero, el modelo de tres parámetros se convierte en modelo 2PL y si se establece el valor de uno al nivel de discriminación a todos los ítems, el modelo de dos parámetros se convierte en el modelo 1PL (Muñiz, 2010).

En la figura 3.10 se puede observar la forma en que pueden interactuar los parámetros de cada uno de los modelos logísticos y los valores de los parámetros a , b , c y γ son constantes dependiendo del modelo elegido como se visualiza en la tabla 3.1.

Parámetros	b	a	c	γ
1PL	0	-	-	-
2PL	0	1.3	-	-
3PL	0	1.3	0.21	-
4PL	0	1.3	0.21	0.91

Tabla 3.1. Modelos paramétricos de la IRT.

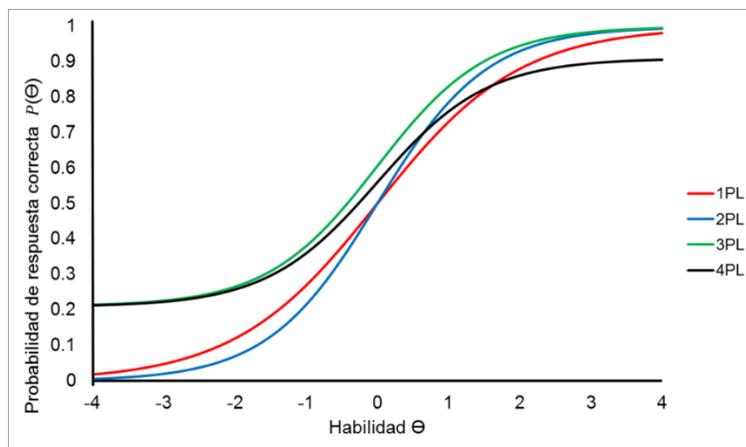


Figura 3.10. Representación de los modelos de la IRT.

Como se observa el modelo de Rasch o 1PL es un caso especial del modelo 2PL, el cual se obtiene a partir del nivel de discriminación constante igual a uno para todos los ítems que componen la prueba. El modelo 2PL se obtiene por la restricción de $c = 0$ para cada ítem. El modelo 4PL es más flexible y generalmente tiene mejor ajuste en los datos mientras que los modelos 1PL y 2PL son más exigentes y fácilmente se rechaza el pseudo-azar y el error por descuido. Los modelos 3PL y 4PL son más flexibles dado que la interpretación es menos clara en los valores de los parámetros c y γ .

Es importante que los datos sean los que provean el mejor juicio y se debe elegir aquel modelo que dé cuenta de ellos. Para ajustes similares se debe de escoger el más sencillo para la estimación de los parámetros y de la habilidad de los examinados.

Una vez que se seleccione el modelo el paso siguiente será estimar los parámetros de cada ítem y el valor de la variable medida para cada examinado, a partir de los datos obtenidos al aplicar los ítems a una muestra amplia de examinados.

La estimación se va realizando mediante un método iterativo y su cálculo es muy laborioso por lo que es necesario realizarlos mediante computadora. El proceso de las iteraciones se detiene cuando los valores estimados en los parámetros convergen, es decir, cuando tras una iteración no se producen cambios significativos en los valores estimados. Una vez estimados los parámetros del modelo se debe comprobar que los resultados pronosticados coinciden con los obtenidos, es decir, es necesario comprobar el ajuste de los datos al modelo. Tal ajuste se produce cuando los valores de $P(\theta)$ pronosticados por el modelo no difieran estadísticamente de los obtenidos de manera empírica, es decir, de la proporción de examinados que realmente acierten al ítem.

3.3 Puntuaciones que no se distribuyen de forma normal

Existen exámenes que miden habilidades cognitivas en los individuos aplicadas en pruebas de investigación y de práctica clínica en neuropsicología, debido a que son fáciles de administrar y con alta sensibilidad para detectar diversas patologías ya que no se requiere de una preparación especial o experiencia previa para sustentarlas, lo que permite dar una magnitud individual de las funciones cognitivas y como diagnóstico rápido de las habilidades medidas, así como sus fallas.

Para evitar errores al interpretar los puntajes de las evaluaciones donde el resultado se obtiene de manera individual, dado que los niveles de habilidades pueden ser muy bajos o muy altos, es necesario dar una interpretación que este en el contexto de la evaluación que se esté llevando a cabo. Esta situación lleva a

una inapropiada conversión de los datos en las pruebas debido a que los resultados obtenidos en la población de examinados presentan niveles de habilidades que varían drásticamente obteniendo puntajes muy bajos hasta los puntajes más altos de la prueba y pocos datos centrados en la media, presentando una distribución de los datos de naturaleza asimétrica, es decir, que presentan gran dispersión entre las puntuaciones (Aranciva, y otros, 2012).

Una distribución estandarizada resultado del puntaje de la prueba, el cual tiene la misma forma que la distribución original ya que dependen de las características individuales que componen al examen y está diseñado de manera que la mayoría de examinados puedan tener un buen desempeño en todos los ítems que componen la prueba.

Existen pruebas de práctica clínica, neuropsicológica y de educación que miden habilidades cognitivas en los individuos, por ejemplo, la prueba de la figura compleja de Rey-Osterrieth, las Matrices Progresivas de Raven y el Boston Naming Test (BNT), estas pruebas tienen un mayor poder de discriminación en sujetos con niveles de habilidades muy bajos (Mitrushina, Boone, Razini, & D'Elia, 2005). En los exámenes donde se tienen los resultados con niveles de rendimiento o bajos la distribución del puntaje esta sesgada positivamente.

3.4 Niveles de habilidad y dificultad de los ítems

En este apartado se describe a manera de ejemplo el proceso para el cálculo de los niveles de habilidad y dificultad de una muestra de 13 sujetos y una prueba

compuesta de 10 ítems. Como se observa en la tabla 3.2, se construyó una matriz de datos con respuesta a ítems dicotómicos.

	Ítem1	Ítem2	Ítem3	Ítem4	Ítem5	Ítem6	Ítem7	Ítem8	Ítem9	Ítem10
Sujeto1	0	0	0	1	1	1	0	0	1	0
Sujeto2	1	0	0	0	1	1	0	0	0	1
Sujeto3	1	1	0	1	0	0	1	1	0	1
Sujeto4	0	1	1	0	1	1	1	0	1	0
Sujeto5	0	0	1	1	0	1	0	0	1	1
Sujeto6	1	1	1	1	1	0	1	1	1	1
Sujeto7	1	1	1	1	1	1	0	1	1	1
Sujeto8	0	1	1	1	0	0	1	0	1	1
Sujeto9	1	0	1	1	0	1	1	1	0	1
Sujeto10	1	1	0	1	1	1	0	1	1	0
Sujeto11	0	1	1	0	1	0	0	1	1	1
Sujeto12	1	1	0	1	0	1	1	1	1	0
Sujeto13	0	1	1	1	1	0	1	1	0	1

Tabla 3.2. Ejemplo de datos de un examen.

Se calculan los niveles de habilidad de cada uno de los sujetos que contestaron la prueba, se calcula el promedio de cada uno de los ítems representados por I_n , y se denota por la letra P (Promedio obtenido en la prueba) de toda la fila mediante la ecuación 3.6.

$$P = \frac{1}{n} \sum_{j=1}^n I_j = \frac{I_1 + I_2 + I_3 \dots I_n}{n} \quad n = 10 \quad (3.6)$$

Se calcula el valor de Q expresado por $Q = 1 - P$, una vez obtenidos los valores de P y Q , se calcula $\log\left(\frac{P}{Q}\right)$ el cuál es el valor de la habilidad Θ en la escala logit, como se describe en la tablas 3.3.

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10
Sujeto 1	0	0	0	1	1	1	0	0	1	0
Sujeto 2	1	0	0	0	1	1	0	0	0	1
Sujeto 3	1	1	0	1	0	0	1	1	0	1
Sujeto 4	0	1	1	0	1	1	1	0	1	0
Sujeto 5	0	0	1	1	0	1	0	0	1	1
Sujeto 6	1	1	1	1	1	0	1	1	1	1
Sujeto 7	1	1	1	1	1	1	0	1	1	1
Sujeto 8	0	1	1	1	0	0	1	0	1	1
Sujeto 9	1	0	1	1	0	1	1	1	0	1
Sujeto 10	1	1	0	1	1	1	0	1	1	0
Sujeto 11	0	1	1	0	1	0	0	1	1	1
Sujeto 12	1	1	0	1	0	1	1	1	1	0
Sujeto 13	0	1	1	1	1	0	1	1	0	1

Tabla 3.3. Matriz de datos del nivel de habilidad de los examinados.

La tabla 3.4 describe el cálculo de los niveles de habilidad obtenidos para la muestra de sujetos evaluados que se describen en el ejemplo.

	$P = \frac{1}{n} \sum_{j=1}^{n=10} I_j$	$Q = (1 - P)$	$\frac{P}{Q}$	$\theta = LN\left(\frac{P}{Q}\right)$
Sujeto 1	0.4	0.6	0.66667	-0.40547
Sujeto 2	0.4	0.6	0.66667	-0.40547
Sujeto 3	0.6	0.4	1.5	0.40547
Sujeto 4	0.6	0.4	1.5	0.40547
Sujeto 5	0.5	0.5	1	0
Sujeto 6	0.9	0.1	9	2.19722
Sujeto 7	0.9	0.1	9	2.19722
Sujeto 8	0.6	0.4	1.5	0.40547
Sujeto 9	0.7	0.3	2.33333	0.84730
Sujeto 10	0.7	0.3	2.33333	0.84730
Sujeto 11	0.6	0.4	1.5	0.40547
Sujeto 12	0.7	0.3	2.33333	0.84730
Sujeto 13	0.7	0.3	2.33333	0.84730

Tabla 3.4. Cálculo de niveles de habilidad de los examinados.

El nivel de dificultad de los ítems que componen la prueba se denota por b , el procedimiento se realiza al obtener el promedio de cada ítem y se representa por el valor de μ , se calcula el valor de $1 - \mu$, una vez obtenidos los valores de μ y

$(1 - \mu)$, se calcula el valor de la dificultad $b = \log\left(\frac{\mu}{1-\mu}\right)$ como se observa en la tabla

3.5.

	Ítem1	Ítem2	Ítem3	Ítem4	Ítem5	Ítem6	Ítem7	Ítem8	Ítem9	Ítem10
Sujeto 1	0	0	0	1	1	1	0	0	1	0
Sujeto 2	1	0	0	0	1	1	0	0	0	1
Sujeto 3	1	1	0	1	0	0	1	1	0	1
Sujeto 4	0	1	1	0	1	1	1	0	1	0
Sujeto 5	0	0	1	1	0	1	0	0	1	1
Sujeto 6	1	1	1	1	1	0	1	1	1	1
Sujeto 7	1	1	1	1	1	1	0	1	1	1
Sujeto 8	0	1	1	1	0	0	1	0	1	1
Sujeto 9	1	0	1	1	0	1	1	1	0	1
Sujeto 10	1	1	0	1	1	1	0	1	1	0
Sujeto 11	0	1	1	0	1	0	0	1	1	1
Sujeto 12	1	1	0	1	0	1	1	1	1	0
Sujeto 13	0	1	1	1	1	0	1	1	0	1
$\mu = \frac{1}{n} \sum_{j=1}^{n=13} S_j$	0.5385	0.6923	0.6154	0.7692	0.6154	0.6154	0.5385	0.6154	0.6923	0.6923
$1 - \mu$	0.4615	0.3077	0.3846	0.2308	0.3846	0.3846	0.4615	0.3846	0.3077	0.3077
$\frac{\mu}{1 - \mu}$	1.1667	2.25	1.6	3.3333	1.6	1.6	1.1667	1.6	2.25	2.25
$b = \ln\left(\frac{\mu}{1-\mu}\right)$	0.1542	0.8109	0.4700	1.2040	0.4700	0.4700	0.1542	0.4700	0.8109	0.8109

Tabla 3.5. Cálculo del nivel de dificultad de los ítems del examen.

Se eliminan filas y columnas en donde los niveles de habilidad de los evaluados y dificultad de los ítems sean 1 o 0 en el promedio, ya que estos valores no son de interés en el contexto de la IRT y para la muestra de datos de este ejemplo no se presenta este tipo de resultados.

Se ordenan los niveles de habilidad de los evaluados de menor a mayor como se muestra en la tabla 3.6, se agrupan los niveles de habilidad, es decir se promedian los mismos valores de habilidad de los examinados mediante el promedio de ítems correcto entre el número total de sujetos con mismo nivel de habilidad.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	θ
Sujeto 1	0	0	0	1	1	1	0	0	1	0	-0.4054
Sujeto 2	1	0	0	0	1	1	0	0	0	1	-0.4054
Sujeto 5	0	0	1	1	0	1	0	0	1	1	0
Sujeto 3	1	1	0	1	0	0	1	1	0	1	0.4054
Sujeto 4	0	1	1	0	1	1	1	0	1	0	0.4054
Sujeto 8	0	1	1	1	0	0	1	0	1	1	0.4054
Sujeto 11	0	1	1	0	1	0	0	1	1	1	0.4054
Sujeto 9	1	0	1	1	0	1	1	1	0	1	0.8473
Sujeto 10	1	1	0	1	1	1	0	1	1	0	0.8473
Sujeto 12	1	1	0	1	0	1	1	1	1	0	0.8473
Sujeto 13	0	1	1	1	1	0	1	1	0	1	0.8473
Sujeto 6	1	1	1	1	1	0	1	1	1	1	2.19722
Sujeto 7	1	1	1	1	1	1	0	1	1	1	2.19722
b	0.1542	0.8109	0.47	1.204	0.47	0.47	0.1542	0.47	0.8109	0.8109	

Tabla 3.6. Matriz de resultado de los niveles de habilidad y dificultad.

Se construye una matriz con los resultados obtenidos de los promedios donde se obtuvo cinco clases como los muestra la tabla 3.7.

Clase θ	Ítem1	Ítem2	Ítem3	Ítem4	Ítem5	Ítem6	Ítem7	Ítem8	Ítem9	Ítem10
-0.4054	0.5	0	0	0.5	1	1	0	0	0.5	0.5
0	0	0	1	1	0	1	0	0	1	1
0.4054	0.25	1	0.75	0.5	0.5	0.25	0.75	0.5	0.75	0.75
0.8473	0.75	0.75	0.5	1	0.5	0.75	0.75	1	0.5	0.5
2.1972	1	1	1	1	1	0.5	0.5	1	1	1
b	0.154	0.811	0.47	1.204	0.47	0.47	0.154	0.47	0.811	0.8109

Tabla 3.7. Matriz agrupada del nivel de habilidad de los sujetos evaluados.

3.5 La herramienta Solver de Excel

La herramienta Solver de Microsoft Excel permite encontrar un valor óptimo (máximo, mínimo o exacto) para una formula en una celda, que es denominada celda objetivo, en donde se encuentra el problema en cuestión y está sujeta a restricciones o limitaciones, con valores de otras celdas en una hoja de cálculo.

Solver trabaja con un grupo de celdas que cambian al participar en el cálculo de la fórmula de las celdas objetivo y/o restricción.

Se ajustan los valores en las celdas de las variables de decisión para cumplir con los límites en las celdas de restricción y producen el resultado deseado para la celda objetivo.

Solver utiliza diversos métodos de solución dependiendo de la opción que se seleccione, para la resolución de problemas no lineales utilizando el algoritmo del gradiente generalizado reducido (Generalized Reduced Gradient o GRG2 por sus siglas en inglés) de optimización no lineal desarrollado por la Universidad Leon Lasdon de Austin, Texas y la Universidad Allan Waren en Cleveland, Ohio. Este algoritmo parte de una solución factible conocida como punto inicial, e intenta moverse a partir de este punto en una dirección a través de la región factible, de tal forma que el valor de la función objetivo mejore (Winston, 2014).

Condiciones para un ajuste no lineal con GRG2:

- Tiempo máximo, limita el tiempo que tarda el proceso de solución, puede introducirse un valor de hasta 32,367 segundos, donde el valor predeterminado es de 100 segundos el cual es adecuado para la mayoría de los problemas.
- Iteraciones, limita el tiempo que tarda el proceso de solución reduciendo el número de cálculos provisionales. El algoritmo permite introducir un valor hasta 32,767, el valor predeterminado es de 100 iteraciones, lo cual es adecuado para la mayoría de los problemas pequeños.
- Precisión, permite controlar la precisión de las soluciones utilizando el número que se introduce para averiguar si el valor de una restricción cumple

un objetivo o satisface un límite inferior o superior. Se deberá indicar la precisión mediante una fracción entre cero y uno. Si se tienen más decimales en el número que se introduzca, mayor será la precisión, donde 0.000001 indica que tiene mayor precisión que 0.01. Cuanto mayor sea la precisión, más tiempo se tardará en encontrar una solución.

- Convergencia, se aplica únicamente a problemas no lineales entre 0 y 1 cuanto menor sea el valor de convergencia más tiempo tardará en encontrar la solución.

Capítulo 4 . Desarrollo del trabajo de investigación y resultados alcanzados

En este capítulo se describe el desarrollo del trabajo de investigación que se realizó, así como los métodos y procedimientos utilizados en el análisis efectuado sobre el conjunto de estudio, el cual contiene los resultados de una aplicación de la prueba de la figura compleja de Rey-Osterrieth a un grupo de alumnos. Dicha prueba se aplicó a una población de estudiantes afines a la carrera de Ingeniería en Sistemas Computacionales, pertenecientes a dos universidades dentro del Estado de Hidalgo. Cabe mencionar que, para efectos de este trabajo de investigación, solo se ocupó el conjunto de resultados obtenidos, utilizando la interpretación dicotómica de los mismos; es decir, se tienen valores de cero y uno, donde cero implica fallo y uno implica acierto en cada uno de los ítems que contempla la prueba. Por otro lado, es pertinente aclarar que la aplicación de la

prueba se realizó dentro del contexto de otro trabajo de investigación y por tal motivo las condiciones y metodologías utilizadas en dicha aplicación no formaron parte de este trabajo, solo es de interés el conjunto de resultados que se obtuvo ya que sobre estos datos se centra el actual trabajo de investigación.

4.1 Datos de estudio

La mayoría de las interpretaciones que se dan a los resultados de los exámenes se basan en tomar el puntaje total de la prueba, el cual, está representado generalmente por la media y con una distribución normal. Sin embargo, existen evaluaciones donde los puntajes obtenidos presentan comportamientos diferentes a los tradicionales, esto pudiera ser debido a alguna de las siguientes situaciones:

- Primero, porque los examinados tienen niveles de habilidades o muy altos o muy bajos, lo cual genera una distribución sesgada.
- Segundo, porque la interpretación de los resultados obedece a criterios diferentes a los establecidos comúnmente y por lo tanto son pocas las puntuaciones que están centradas en la media, y esto puede producir distribuciones con más de una moda.
- Tercero, la interpretación está sujeta a un rasgo latente diferente al de la habilidad del examinado, ya que este rasgo es el que comúnmente se busca, pero no es el único y por lo tanto cuando se elige otro rasgo los resultados pueden ser muy variados. Para el caso de este trabajo se tomará en cuenta solamente la habilidad del examinado como rasgo latente.

El análisis realizado en este trabajo comenzó con el estudio de la muestra de resultados obtenidos de la prueba de Rey-Osterrieth, cabe destacar, que tal prueba permite evaluar habilidades cognitivas, de percepción y de memoria a corto plazo. Para esto, la prueba está compuesta por 18 ítems y sus respectivas respuestas correctas.

Con los resultados que se tienen de la prueba se construyó una matriz de datos que se obtuvieron en la etapa de reproducción, la cual está compuesta por 71 filas que representan a los examinados y 18 columnas que representan a los ítems como lo marca la prueba (ver figura 3.1) y como se puede observar en la tabla 4.1, la cual contiene los primeros 10 resultados. Asimismo, en las tablas B.1 y B.2 del anexo B se pueden encontrar todos los resultados obtenidos.

Ítem	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18
Sujeto 1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0	0	1
Sujeto 2	0	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	0	0
Sujeto 3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Sujeto 4	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	0	0	1
Sujeto 5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Sujeto 6	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0
Sujeto 7	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	0
Sujeto 8	0	1	1	1	1	1	0	1	0	0	1	1	1	1	0	1	0	0
Sujeto 9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Sujeto 10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Tabla 4.1. Fragmento de la matriz de resultados de la prueba de R-O.

Para representar el nivel de habilidad de cada uno de los examinados se utiliza la escala logit, la misma se ocupa para representar la dificultad de cada uno de los ítems que componen la prueba. Con los resultados obtenidos de los estudiantes que presentaron la prueba de Rey-Osterrieth, se calcula el nivel de habilidad de los estudiantes en los 18 ítems utilizando la ecuación 3.1 para hallar el nivel de

habilidad en la escala de logit como se observa en la tabla 4.2, la cual contiene los 10 primeros resultados y en las tablas C.1 y C.2 del anexo C se encuentran todos los resultados obtenidos.

	P	1-P	P/Q	LN(P/Q)
Sujeto 1	0.76470588	0.23529412	3.25	1.178655
Sujeto 2	0.64705882	0.35294118	1.83333333	0.6061358
Sujeto 3	0.94117647	0.05882353	16	2.77258872
Sujeto 4	0.70588235	0.29411765	2.4	0.87546874
Sujeto 5	0.94117647	0.05882353	16	2.77258872
Sujeto 6	0.88235294	0.11764706	7.5	2.01490302
Sujeto 7	0.76470588	0.23529412	3.25	1.178655
Sujeto 8	0.58823529	0.41176471	1.42857143	0.35667494
Sujeto 9	0.94117647	0.05882353	16	2.77258872
Sujeto 10	1	0	-	-

Tabla 4.2. Fragmento del nivel de habilidades en la escala de logit .

Para hallar el nivel de dificultad en la escala logit de los ítems que componen la prueba se utiliza la ecuación 3.1, teniendo un rango de nivel de dificultad b comprendido entre -0.69 y 3.51, como se aprecia en los 10 primeros ítems de la tabla 4.3 y en la tabla D.1 del anexo D se hayan todos los resultados obtenidos.

Ítem	M1	M2	M3	M4	M6	M7	M8	M9	M10	M11
μ	0.77	0.97	0.94	0.99	0.81	0.81	0.86	0.77	0.33	0.93
$1 - \mu$	0.23	0.03	0.06	0.01	0.19	0.19	0.15	0.23	0.67	0.07
$\frac{\mu}{1 - \mu}$	3.31	33.50	16.25	68.00	4.31	4.31	5.90	3.31	0.50	12.80
$b = \ln\left(\frac{\mu}{1 - \mu}\right)$	1.20	3.51	2.79	4.22	1.46	1.46	1.78	1.20	-0.69	2.55

Tabla 4.3. Nivel de dificultad de los ítems de la prueba de R-O.

Tras haber efectuado los cálculos de los niveles de habilidad, se pueden encontrar columnas o filas que contengan valores cero o uno, estos valores no aportan información y por lo tanto pueden ser removidos de la matriz de valores. En la matriz de trabajo que se muestra en la tabla 4.1, ocurre este fenómeno y para ello se eliminó la columna asociada al ítem 5, ya que esta contenía solo valores 1. Después de quitar esta columna se obtuvieron dos filas que presentaban el mismo fenómeno, estas correspondiendo a dos estudiantes con el nivel de habilidad igual a 1, los cuales fueron el número 10 y 59 como se observa en las tablas E.1 y E.2 del anexo E.

Al obtener los resultados de los niveles de habilidad de los estudiantes evaluados mismos que se muestran en la tabla 4.2, fue necesario agrupar en clases, donde cada clase agrupa a estudiantes con un mismo nivel de habilidad. Las clases encontradas contemplan valores de habilidad θ en un rango comprendido entre -0.8754 y 2.772 en la escala logit como se observa en la tabla 4.4. Se ordenaron los resultados de las habilidades de forma ascendente, es decir, de menor a mayor y se calculó el promedio de las habilidades por cada clase.

Tomando en cuenta los niveles de habilidad de la tabla 4.2 cuyo valor es 0.3566, se puede apreciar que los examinados pertenecientes a esta clase son 8, 21, 31, 35, 37, 58 y 61, quienes obtuvieron puntuaciones con 0, 0, 1, 0, 1, 1 y 0 respectivamente. Se realiza la suma de las puntuaciones y se calcula el promedio cuyo valor es 0.429. Esto representa el promedio de la dificultad para la clase 0.3566 para el ítem M1, este proceso se realiza para todos los niveles de habilidad de los evaluados, lo cual da como resultado 11 filas que están agrupadas en clases de acuerdo al nivel de habilidad. Los resultados anteriormente mencionados se pueden observar en la tabla 4.4, mientras que un fragmento de la matriz de los

resultados de los promedios por clase se muestra en la tabla 4.5 para los primeros 10 ítems. La matriz completa se puede encontrar en la tabla F.1 del anexo F.

	M1	M2	M3	M4	M6	M7	M8	M9	M10	M11	...	θ
Sujeto 56	1	0	0	0	1	0	0	0	0	1	...	-0.875468737
Sujeto 36	0	1	1	1	0	0	1	0	0	1	...	-0.356674944
Sujeto 25	0	1	1	1	1	1	0	1	0	1	...	-0.117783036
Sujeto 33	1	0	0	1	1	1	1	1	0	0	...	-0.117783036
Sujeto 22	1	1	1	1	1	0	0	1	0	1	...	0.117783036
Sujeto 21	0	1	1	1	0	0	1	1	0	1	...	0.356674944
Sujeto 31	0	1	0	1	1	1	1	0	0	0	...	0.356674944
Sujeto 35	1	1	1	1	1	0	0	1	0	0	...	0.356674944
Sujeto 37	0	1	1	1	1	1	0	0	0	1	...	0.356674944
Sujeto 58	1	1	1	1	0	1	1	0	0	1	...	0.356674944
Sujeto 61	1	1	0	1	1	0	0	0	0	1	...	0.356674944
Sujeto 8	0	1	1	1	1	0	1	0	0	1	...	0.356674944

Tabla 4.4. Fragmento de habilidades ordenadas.

Clase	θ	M1	M2	M3	M4	M6	M7	M8	M9	M10
1	-0.875	1	0	0	0	1	0	0	0	0
2	-0.357	0	1	1	1	0	0	1	0	0
3	-0.118	0.5	0.5	0.5	1	1	1	0.5	1	0
4	0.118	1	1	1	1	1	0	0	1	0
5	0.357	0.429	1	0.714	1	0.714	0.429	0.571	0.286	0
6	0.606	0.556	1	1	1	0.556	0.556	0.778	0.667	0.222
7	0.875	0.833	1	1	1	0.833	1	0.667	0.667	0.167
8	1.179	0.7	1	1	1	0.7	0.9	1	0.9	0.4
9	1.540	0.857	1	1	1	0.857	0.929	1	0.857	0.357
10	2.015	1	1	1	1	1	1	1	0.889	0.444
11	2.773	1	1	1	1	1	1	1	1	0.778

Tabla 4.5. Niveles de habilidad por clase.

Los datos obtenidos proveerán un mejor juicio de la muestra de sujetos para medir los niveles de habilidad y dificultad del ítem, para ello es necesario elegir

un modelo logístico que permita realizar un ajuste a los datos para aproximar el nivel de habilidad θ a la CCI, en este ajuste se valida que los datos obtenidos de manera empírica y los calculados al aplicar el modelo 1PL usando la ecuación 3.2 no difieran para los niveles de θ y b .

Para ello se tomaron los valores de las puntuaciones del nivel de habilidad θ del ítem M1 de la tabla 4.5, los cuales están representados mediante un gráfico de dispersión en la figura 4.1. Este gráfico indica la posición de los niveles de habilidad de las clases y permite visualizar el comportamiento de las puntuaciones obtenidas en la muestra del ítem M1. Con los valores de θ y b se obtiene $P(\theta)$, donde con un valor de $b = 0.99240$ y un nivel de habilidad θ comprendido entre -4 y 4 se obtiene un resultado de habilidad con valor promedio $\theta = 0.284$.

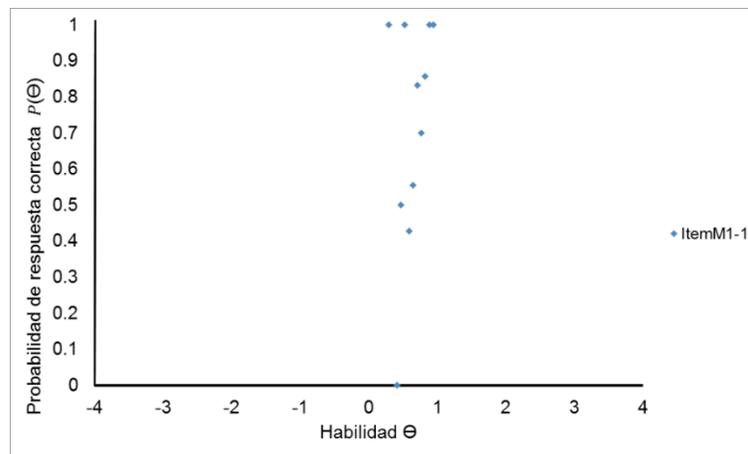


Figura 4.1. Gráfico de dispersión para el ítem M1.

Cabe destacar, que los valores de θ en la tabla 4.6 presentan incrementos unitarios para efectos de su representación gráfica, pero no se debe olvidar que el nivel de

habilidad θ debe ser continuo. La CCI permite visualizar que las habilidades obtenidas al aplicar el modelo 1PL para el ítem M1 se aproximan a las puntuaciones obtenidas de manera empírica. Para visualizar todos los datos continuos de la CCI véase la tabla G.1 del anexo G.

θ	M1
-4	0.00721694
-3	0.01937738
-2	0.05097589
-1	0.12740711
0	0.28412698
1	0.51897051
2	0.74572081
3	0.88854045
4	0.9558884

Tabla 4.6. Niveles de habilidad del ítem M1.

En la figura 4.2 se observa que las puntuaciones obtenidas de manera empírica (puntos azules) se encuentran alejados del nivel de habilidad media en la CCI.

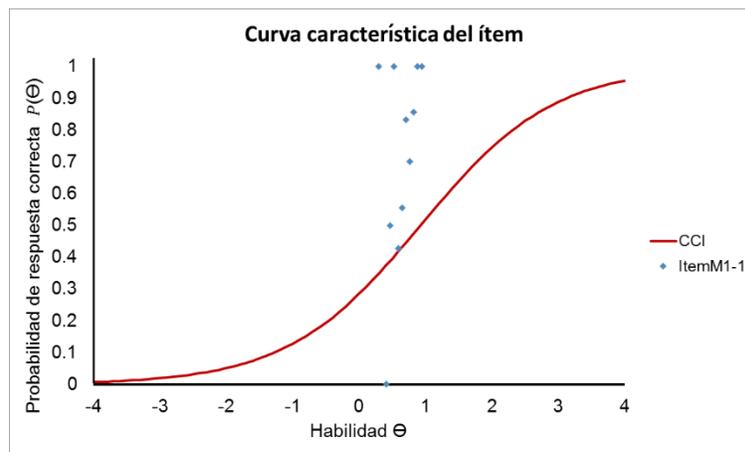


Figura 4.2. CCI del ítem M1 y gráfico de dispersión.

Al realizar el análisis se observa que las puntuaciones obtenidas de manera empírica no se aproximan a los obtenidos con el modelo 1PL en la CCI, entonces se procede a realizar un ajuste al resultado de b mediante un análisis de regresión aplicando el método de los mínimos cuadrados, para ello se recurre a utilizar la ecuación 4.1 que expresa como se obtiene el error cuadrático del nivel de habilidad para cada una de las clases.

$$(\theta - P(\theta))^2 \quad (4.1)$$

Donde θ toma el valor de la habilidad de cada clase del ítem M1 y $b = 0.9241$, cuyo valor se obtuvo al aplicar el modelo 1PL al nivel de dificultad del ítem M1. Después se calcula el error cuadrático al elevar al cuadrado la diferencia del nivel de habilidad θ de cada una de las clases y su correspondiente valor $P(\theta)$, para obtener el error cuadrático de cada clase y al final se calcula la sumatoria del error cuadrático resultando el valor de 1.3731 como se muestra en la tabla 4.7.

Clase	Ítem M1	$P(\theta)$	Error cuadrático
0.2941	1	0.5189	0.2313
0.4117	0	0.2841	0.0807
0.4705	0.5	0.3955	0.0109
0.5294	1	0.5189	0.2313
0.5882	0.4285	0.3785	0.0024
0.6470	0.5555	0.4088	0.0215
0.7058	0.8333	0.4773	0.1267
0.7647	0.7	0.4442	0.0654
0.8235	0.8571	0.4832	0.1397
0.8823	1	0.5189	0.2313
0.9411	1	0.5189	0.2313
b	0.9240	Suma error cuadrático	1.3731

Tabla 4.7. Resultados del ajuste por mínimos cuadrados.

Después de haber obtenido la suma del error cuadrático se procede a realizar el ajuste de los datos mediante la función de optimización no lineal con el algoritmo del Gradiente Reducido Generalizado (GRG2) de la herramienta Solver de Microsoft Excel (Winston, 2014), para obtener el mejor ajuste al minimizar la suma del error cuadrático para encontrar el valor de b y que la CCI se ajuste a los datos con modelo 1PL.

Las condiciones para el cálculo de la regresión del ajuste de los datos fueron: establecer la celda objetivo con un valor de 1.4829 resultado de la sumatoria del error cuadrático aplicando la ecuación 4.1, combinar los valores de la variable $b = 0.9241$ con un valor mínimo, y establecer un tiempo de 100 segundos, con 100 iteraciones y una precisión de 0.000001. Con los resultados generados en Solver al minimizar la suma del error cuadrático a 0.6356 se obtiene el mejor ajuste para $b = 0$.

Una vez obtenidos los niveles de habilidad de los evaluados, la dificultad de los ítems y efectuado el ajuste a los datos mediante el modelo 1PL a los resultados de la prueba de la figura compleja de Rey-Osterrieth, es necesario seguir un proceso para efectuar los cálculos y llegar a los resultados esperados.

En la figuras 4.3 y 4.4 se describe la metodología utilizada para la detección de multimodalidad, mediante un diagrama de flujo donde se describe de manera gráfica los pasos que se siguieron.

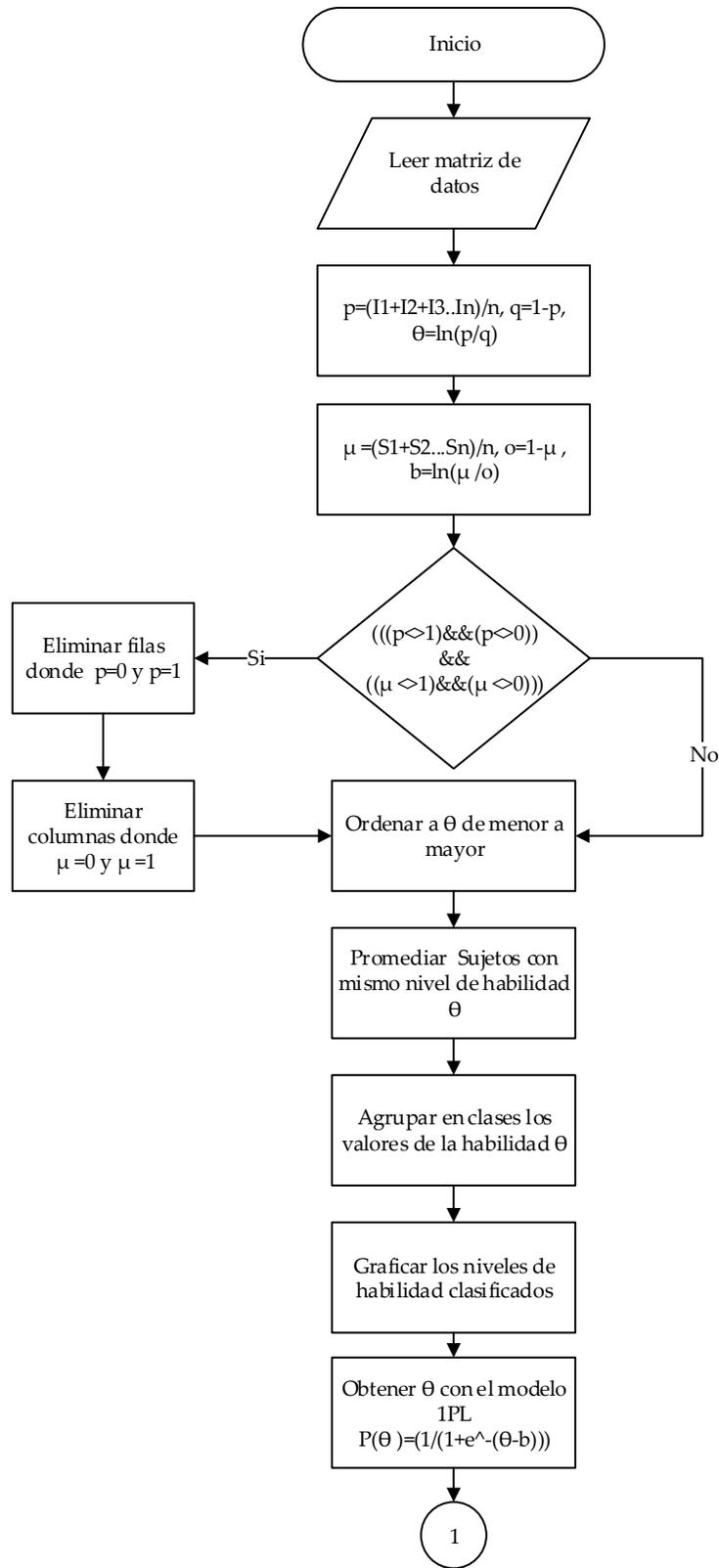


Figura 4.3. Diagrama de flujo para la detección de multimodalidad.

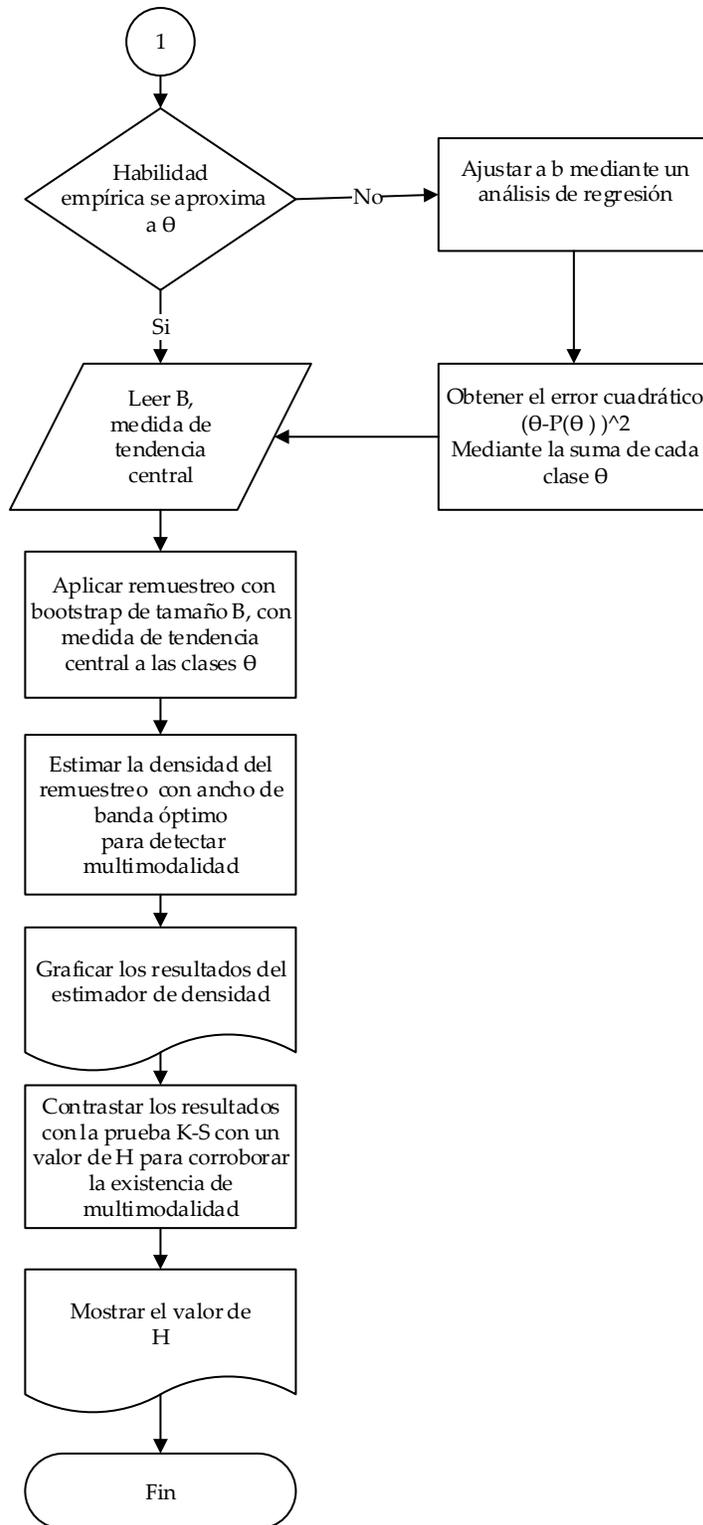


Figura 4.4. Diagrama de flujo para la detección de multimodalidad (continuación).

En la figura 4.5 se describe el caso de uso que se ha de realizar para la implementación del sistema para la detección de multimodalidad, donde se especifica la funcionalidad y el comportamiento del sistema mediante su interacción con el usuario.

Caso de uso	Detectar multimodalidad
Actores	Usuario
Tipo	Primario
Propósito	Detección de multimodalidad a partir de una muestra de datos.
Resumen	Este caso de uso se inicia por el usuario, valida la selección del tamaño de la muestra de datos, la selección de una función de medida de tendencia central y la carga del vector de datos de entrada en una hoja de Excel. Dando como resultado la CDF, el estimador de densidad y el valor del contraste de hipótesis H con K-S test.
Precondiciones	Se requiere tener la clasificación de los niveles de habilidad de los examinados y capturados previamente en una fila de una hoja de cálculo de Excel.
Flujo principal	Se presenta la pantalla principal al usuario. Se selecciona función de medida de tendencia central, el tamaño del remuestreo y el botón para cargar la muestra de datos en un libro de Excel. Si el libro de Excel contiene datos en una fila se ejecutará el subflujo y mostrará los resultados. De lo contrario no se ejecutarán los cálculos y no se mostrarán los resultados como se describe en las excepciones.
Subflujo	<ol style="list-style-type: none"> 1. El usuario selecciona la función de medida de tendencia central, puede seleccionar entre las siguientes opciones: "Moda", "Media" y "Mediana". 2. El usuario selecciona el tamaño del remuestreo entre las siguientes opciones: "500", "1000", "2000", "3000", "5000", "10000" y "100000". 3. El usuario al hacer clic en el botón "Cargar datos", el sistema desplegará un cuadro de dialogo que permitirá buscar un libro de Excel con la extensión ".xlsx", y cargar en memoria el libro con los datos de la muestra. 4. El usuario deberá dar clic en el botón "Calcular multimodalidad", el sistema ejecutará los cálculos de las funciones y mostrará el resultado del cálculo del ancho de banda óptimo, el área bajo la curva, el resultado de H del contraste de hipótesis estadística y el tiempo de cálculo en segundos. 5. El sistema graficará la función de distribución acumulada y el estimador de densidad de núcleo.
Excepciones	<ul style="list-style-type: none"> • Información incompleta: Falta cargar información indispensable para ejecutar los cálculos. El usuario cargará nuevamente la información al sistema. • Información incorrecta: La información contenida en la carga de datos del libro de Excel es incorrecta.

Figura 4.5. Caso de uso del sistema para la detección de multimodalidad.

En la figura 4.6 se muestra la interfaz gráfica de usuario (GUI) del sistema para la detección de multimodalidad implementada en el lenguaje Matlab, en el anexo H se muestra el código fuente del mismo.

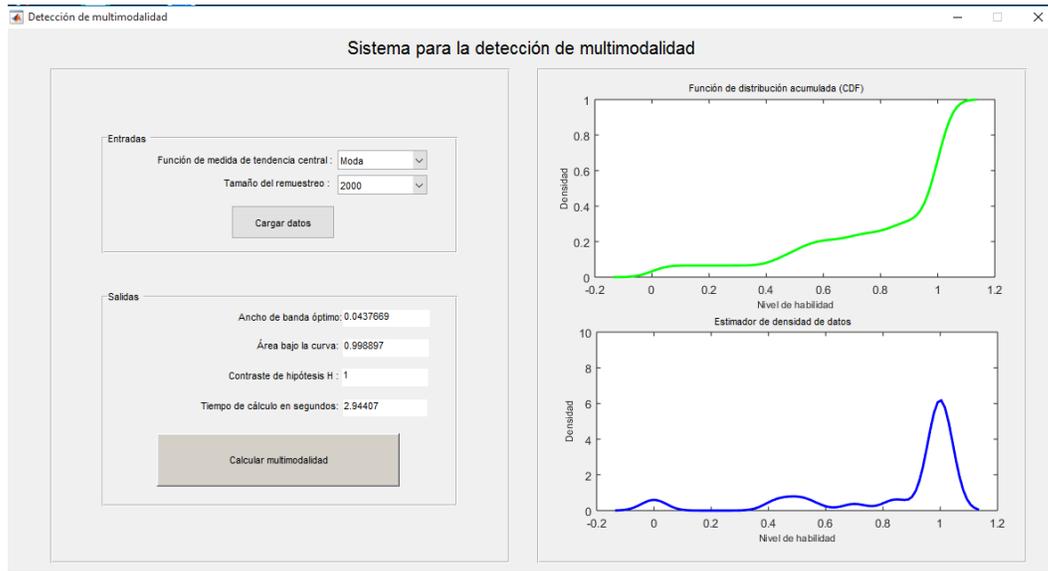


Figura 4.6. GUI del sistema para la detección de multimodalidad.

Para probar el correcto funcionamiento de la aplicación, se realizaron diversas pruebas. Esta etapa de pruebas es independiente de la metodología de desarrollo que se utilizó para la construcción del sistema. En las pruebas de sistema se verifica la funcionalidad de la aplicación desde el punto de vista del usuario final, con base a los requerimientos especificados. El sistema se somete a prueba como una unidad, es decir, se ejecutan varias pruebas con diferentes cargas de datos, esto involucra pruebas de operación, funcionalidad, rendimiento y velocidad de procesamiento.

Para efectos de este trabajo de investigación se revisa la funcionalidad de los comandos tic y toc para los tiempos de cálculo y la ejecución de los ToolBox de

Matlab, las características técnicas donde se implementó y probó el sistema están descritas en la tabla 4.8.

Descripción	Características de software y hardware
Modelo del equipo	HP Pavilion 14-al0011a.
Microprocesador	Intel® Core™ i5-6200U (2,3 GHz, hasta 2.8 GHz, 3 MB de caché, 2 núcleos).
Disco duro	SATA de 1 TB, 5400 rpm.
Sistema operativo	Windows 10 Enterprise 64x.
Versión de Matlab	Matlab R2015A.

Tabla 4.8. Características de software y hardware donde se implementó la solución.

Los comandos tic y toc de Matlab permiten calcular el tiempo de la ejecución de un programa, su cálculo está expresado en segundos y funciona como cronometro. El comando tic se utiliza al inicio del código que se va a evaluar, el comando toc imprime el tiempo transcurrido desde que el comando tic se utilizó. En la tabla 4.9 se observan los promedios de los tiempos de cálculo y en las tablas I.1 e I.2 del anexo I se muestran las puntuaciones de 10 pruebas realizadas con los tiempos de cálculo para la detección de multimodalidad en diferentes ejecuciones.

Remuestreo	Promedio de cálculo con la función Moda	Promedio de cálculo con la función Media	Promedio de cálculo con la función Mediana
500	2.3122	1.0853	1.2126
1000	2.0722	1.0558	1.6353
2000	2.2977	1.2759	1.3317
3000	2.7958	1.6647	1.6065
5000	3.0298	2.2202	2.1018
10000	4.1642	3.4021	4.2299
100000	25.0916	17.9539	15.9383

Tabla 4.9. Tiempos de cálculos.

4.2 Experimentación

Como se observa en la figura 4.7, la CCI resultado del ajuste por el método de los mínimos cuadrados se encuentra alejada de las puntuaciones de los niveles de habilidad más altos obtenidos de manera empírica. Por ello se recurre a construir una muestra de datos con más valores a partir de los 11 valores iniciales aplicando una técnica de remuestreo de datos. Para ello se tomaron los valores del nivel de habilidad del ítem M1 de la tabla 4.7, cuyos valores son 1, 0, 0.5, 1, 0.4286, 0.5556, 0.8333, 0.7, 0.8571, 1 y 1.

Se aplicó la técnica bootstrap para obtener nuevas muestras de datos las cuales se obtienen a partir de un remuestreo con remplazo, donde el valor B representa el tamaño de la población, es decir, en cada remuestreo se obtendrá una población de $B = 500, 1000, 2000, 3000, 5000$ y 10000 respectivamente.

Para obtener el remuestreo de los niveles de habilidad se utilizará la función *bootstrap* que viene incluida en el Toolbox de estadística y aprendizaje máquina de Matlab, cuyo comando o instrucción está descrito por los parámetros de la función $[boots] = bootstrap(B, m_tend, habilidad)$ y recibe como argumentos a B , m_tend y $habilidad$, donde B indica el tamaño del remuestreo a obtener. El parámetro m_tend es la medida de tendencia central sobre la cual se va a obtener el remuestreo y para ello se eligieron los parámetros de *mean*, *median* y *mode*. Y $habilidad$ representa el vector de datos de las 11 clases de los niveles de habilidad, el valor $[boots]$ es el vector resultante de tamaño B que contiene las muestras representativas a las que se les calculará el estimador de densidad de núcleo.

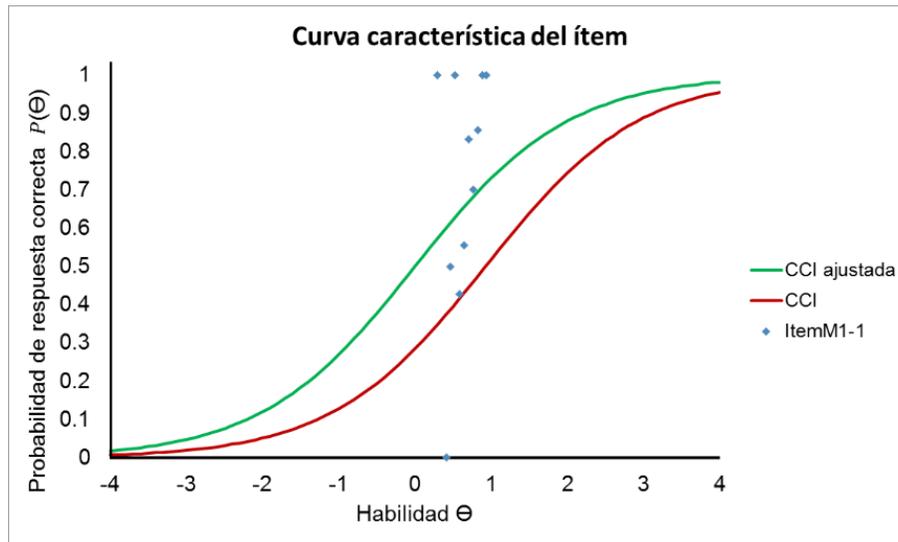


Figura 4.7. Gráfico de dispersión, CCI antes y después de ajuste.

Para calcular el estimador de densidad de núcleo se procede a obtener el valor de h como se aprecia en la ecuación 2.6 mediante la técnica de Silverman, la cual contempla el cálculo del ancho de banda óptimo a partir de los valores del remuestreo resultante en el vector $[boots]$, el procedimiento para el cálculo de h está denotado por $h = 0.9 * B^{(-0.2)} * \min(std(boots), iqr(boots)/1.349)$, donde $B^{(-0.2)}$ es el valor de la raíz quinta del tamaño de muestras B , min representa el valor mínimo del conjunto de datos contenidos en el vector $[boots]$ y el valor de iqr es el rango intercuartil de la serie de los valores de $[boots]$.

Una vez obtenido el valor de h se procede a utilizar la función $ksdensity$ de Matlab la cual devuelve la estimación de densidad de probabilidad, es decir, permite describir la frecuencia con la que aparece un conjunto de observaciones en un estimador no paramétrico de densidad y está descrita por la función

$[f, xi, h] = ksdensity (boots, 'width', h)$, donde $[boots]$ es un vector de datos de tamaño B al cual se le calculará el estimador de densidad. El argumento *'width'* indica que la función recibirá un valor h o valor de ancho de banda ingresado por el usuario, el cual se utilizará como valor de amplitud en el *ksdensity*, donde el valor de h puede tomar por default el valor de la estimación normal de densidad y para una mejor estimación se toman valores altos o pequeños. El valor f es el valor estimado a partir del vector de entrada $[boots]$ y es devuelto en la misma dimensión que xi . El valor de xi representa la evaluación en que *ksdensity* calcula a f y devuelve un vector cubriendo un rango de datos en el eje x . Los valores obtenidos xi en este estudio están comprendidos de -0.2 a 1.2 para los niveles de habilidad a razón de que las observaciones de los núcleos individuales se encuentran comprendidos de cero a uno.

La función de densidad empírica (CDF) es una función que permite calcular la ocurrencia de cada observación para obtener la distribución acumulativa de los puntos de una muestra de datos. Al obtener muestras de tamaño B en cada experimentación aplicando la función *ksdensity*, el valor de la CDF se deberá aproximar o tener un valor igual a la unidad. Para comprobar el valor de la CDF se calcula el área bajo la curva aplicando la integral definida mediante la función $int_definida = trapz(xi, f)$ en Matlab, donde el valor de *int_definida* devuelve el valor de la integral aproximada a través del método del trapecio mediante un espaciado entre unidades con tamaño xi y el conjunto de valores f a integrar, donde xi y f se obtuvieron previamente con la función *ksdensity*. Por defecto, *trapz* opera en la dimensión de f cuyo tamaño se aproxima a la unidad. La longitud xi debe ser igual al tamaño de esta dimensión, por lo que en este estudio el valor del área bajo la curva de la CDF siempre se aproxima a la unidad.

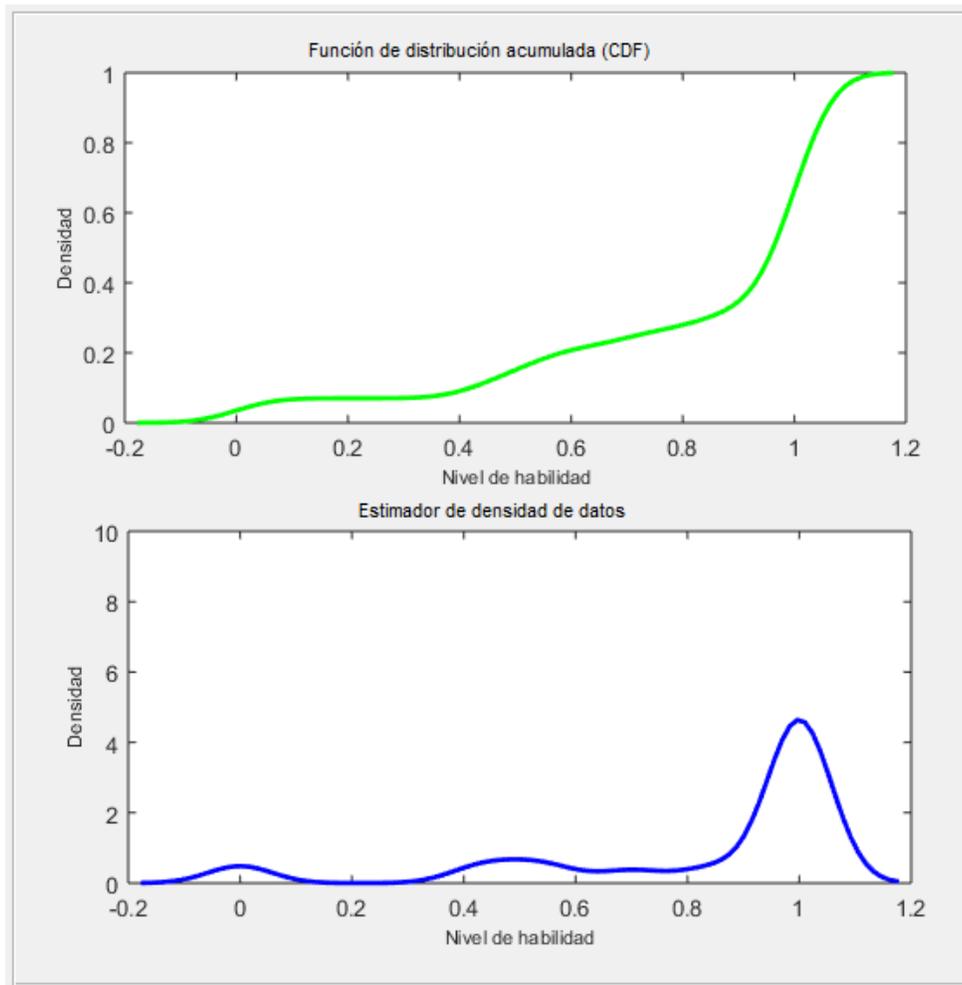
Para reforzar los resultados obtenidos mediante el estimador de núcleo donde las observaciones tienen comportamiento multimodal, se recurre a realizar la prueba de contraste de hipótesis estadística y validar si los datos proceden de una distribución normal o no. Con los resultados obtenidos en *boots* se realizará la prueba de bondad de ajuste de Kolmogorov-Smirnov como se establece en la ecuación 2.7 del capítulo 2, mediante la función de *kstest(boots)* incluida en el Toolbox de estadística y aprendizaje máquina de Matlab, esta función permitirá rechazar la hipótesis de normalidad dando como resultado 1 al pasar como parámetro los valores del vector *boots*, esto validará que los resultados obtenidos por el estimador de kernel tienen un comportamiento multimodal, es decir, se acepta la hipótesis alternativa que indica que no existe unimodalidad en los niveles de habilidad.

4.3 Resultados obtenidos

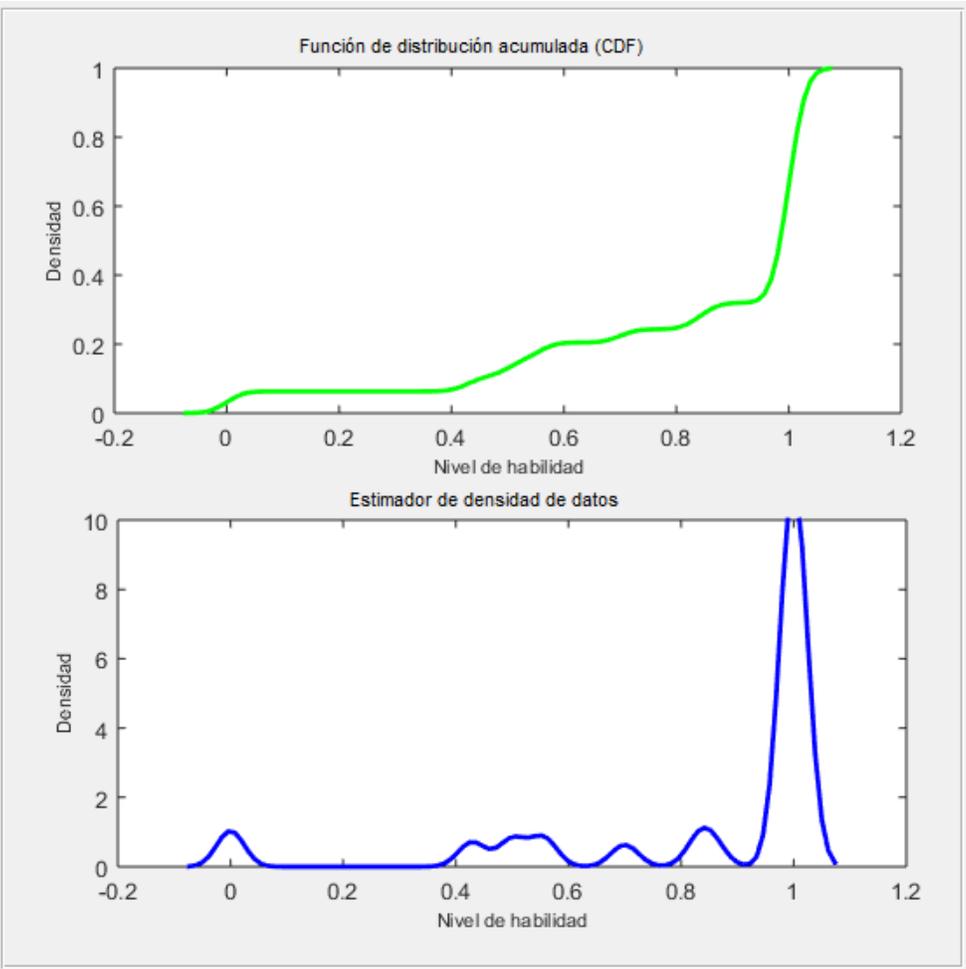
Después de haber realizado la experimentación necesaria para encontrar la multimodalidad en los datos con los que se trabajó, se puede decir, que los resultados obtenidos al aplicar los instrumentos y técnicas seleccionadas para el análisis de datos fueron satisfactorios.

Haciendo un recapitulado, se puede mencionar que se procedió a aplicar la técnica bootstrap a los 11 niveles de habilidad del ítem M1, para ello se eligieron los tamaños de B igual a 500, 1000, 2000, 3000, 5000 y 10000 respectivamente aplicando la media, la mediana y la moda en cada experimentación.

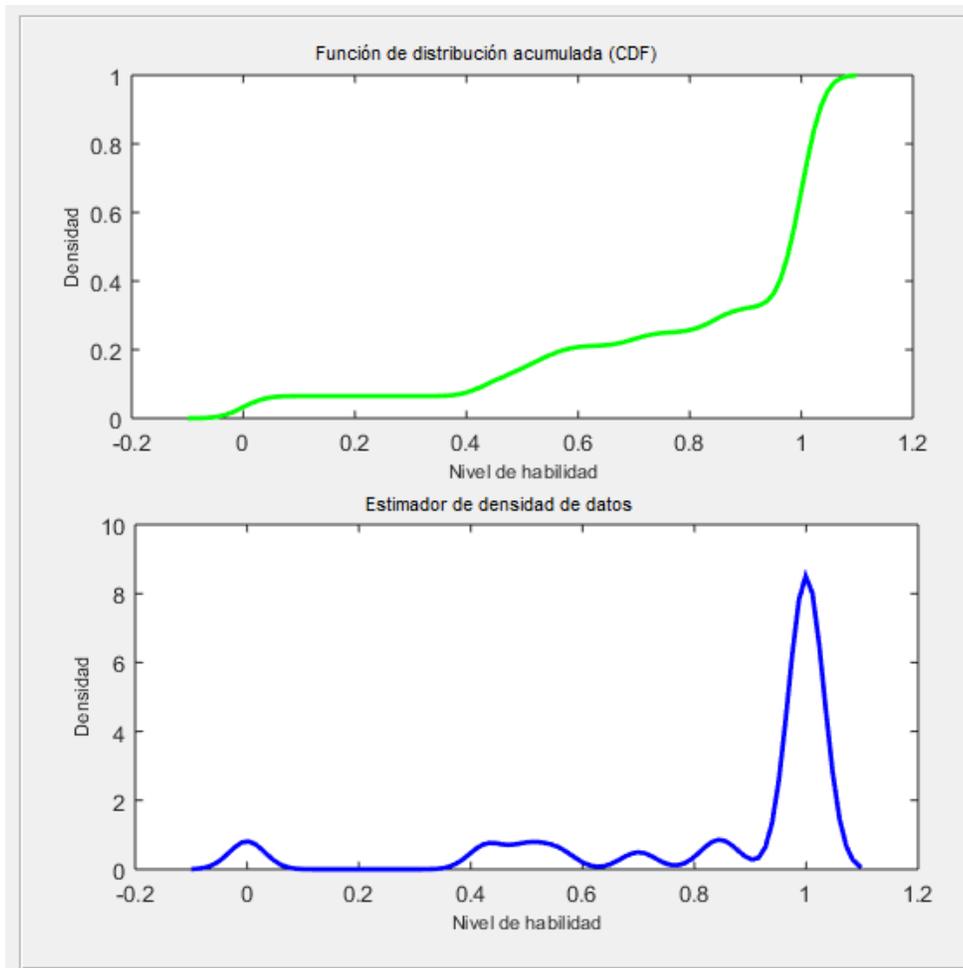
habilidad obtenidos de manera empírica, en la figura 4.8(a) se representa una estimación de densidad de tamaño $B = 500$ para la muestra de datos con valor del ancho de banda óptimo $h = 0.0557$, en la figura 4.8(b) de tamaño $B = 2000$ con valor de $h = 0.0243$ y en la figura 4.8(c) muestra el estimador de densidad con tamaño $B = 10000$ y $h = 0.0317$, donde se observa que los núcleos individuales se encuentran en los niveles de habilidad 0 y 1, en el rango de 0.4 a 0.8 se encuentran limitadas las observaciones cuando se incrementa el tamaño de B .



(a) $B=500, h=0.0557$.



(b) $B=2000$, $h=0.0243$.

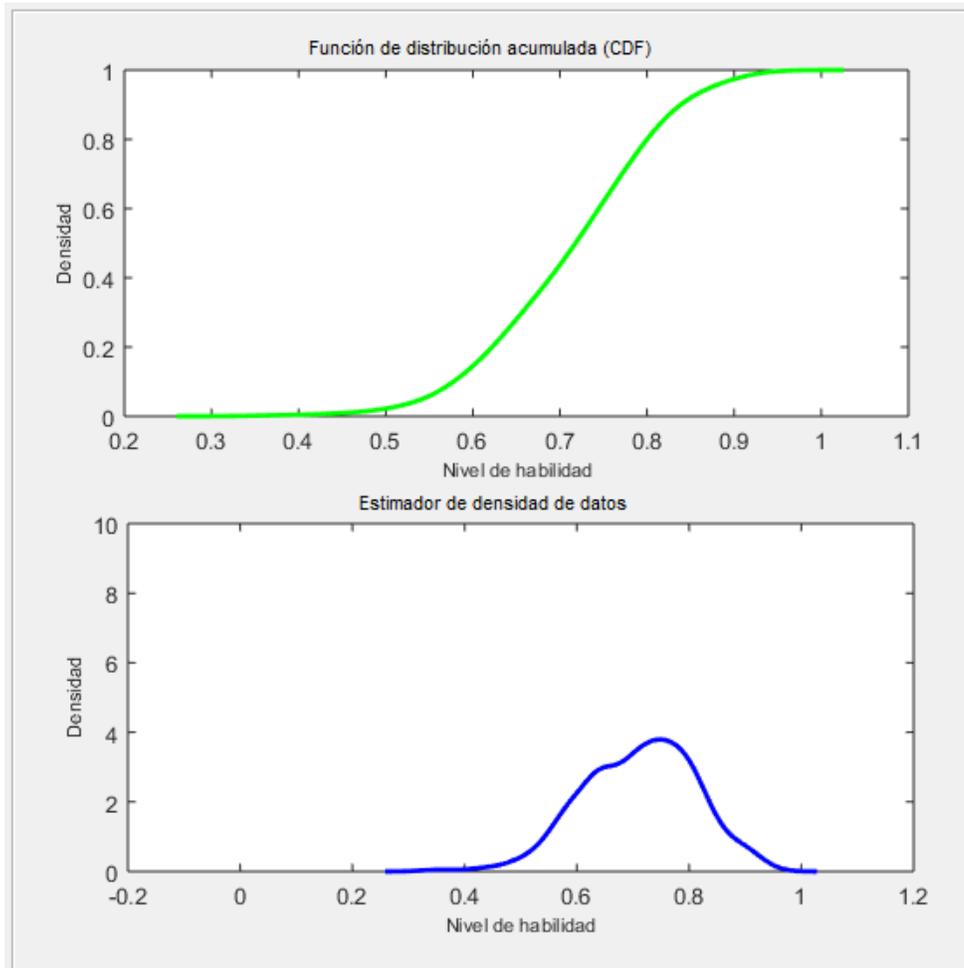


(c) $B=10000$, $h=0.0317$.

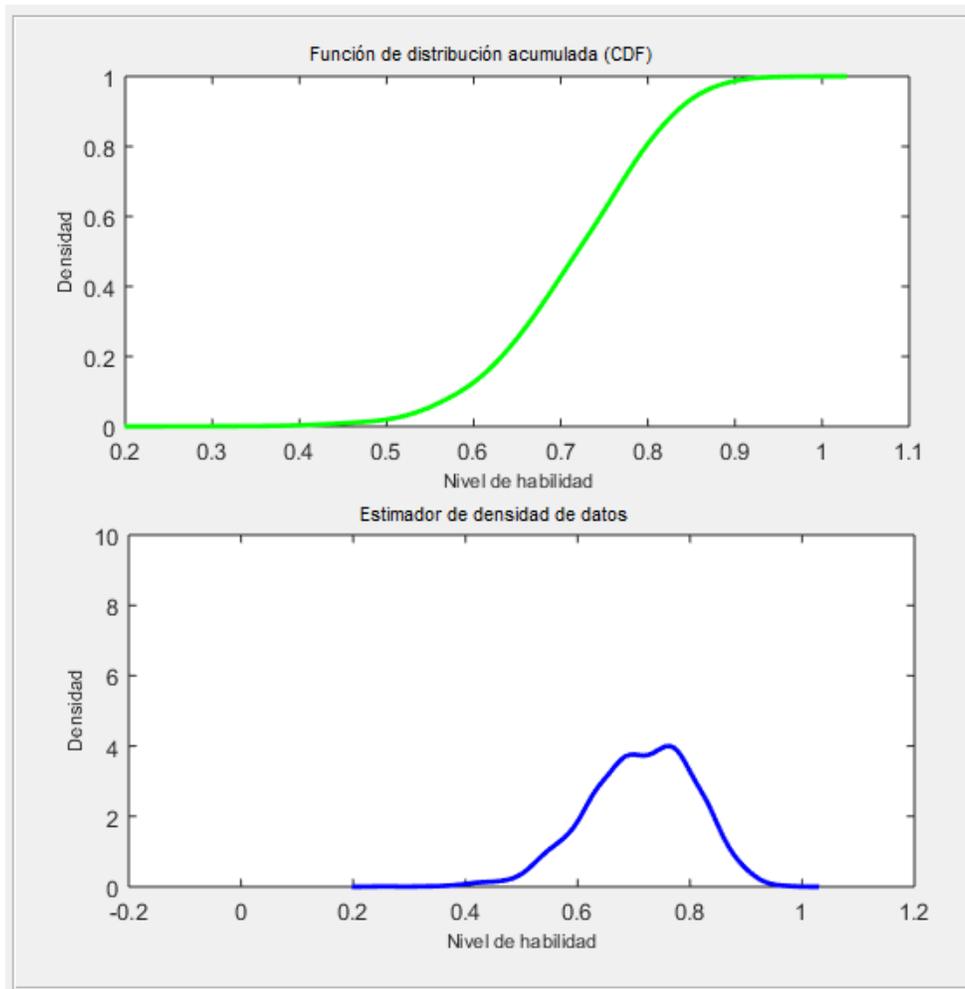
Figura 4.8. Estimadores de densidad usando la moda.

En la figura 4.9 se muestran tres distribuciones de datos donde se usó la media como parámetro de la función *bootstrp* para generar el remuestreo de los niveles de habilidad con la finalidad de obtener poblaciones de mayor tamaño resultando distribuciones empíricas de datos, a estas distribuciones se les calcula el estimador de densidad de tipo kernel usando el ancho de banda óptimo. En la figura 4.9(a) se observa una estimación de la densidad de tamaño $B = 500$ con valor para $h = 0.0258$, en la figura 4.9(b) se ilustra el estimador de densidad de tamaño $B = 2000$ y el valor para el ancho de banda $h = 0.0188$, en la figura 4.9(c)

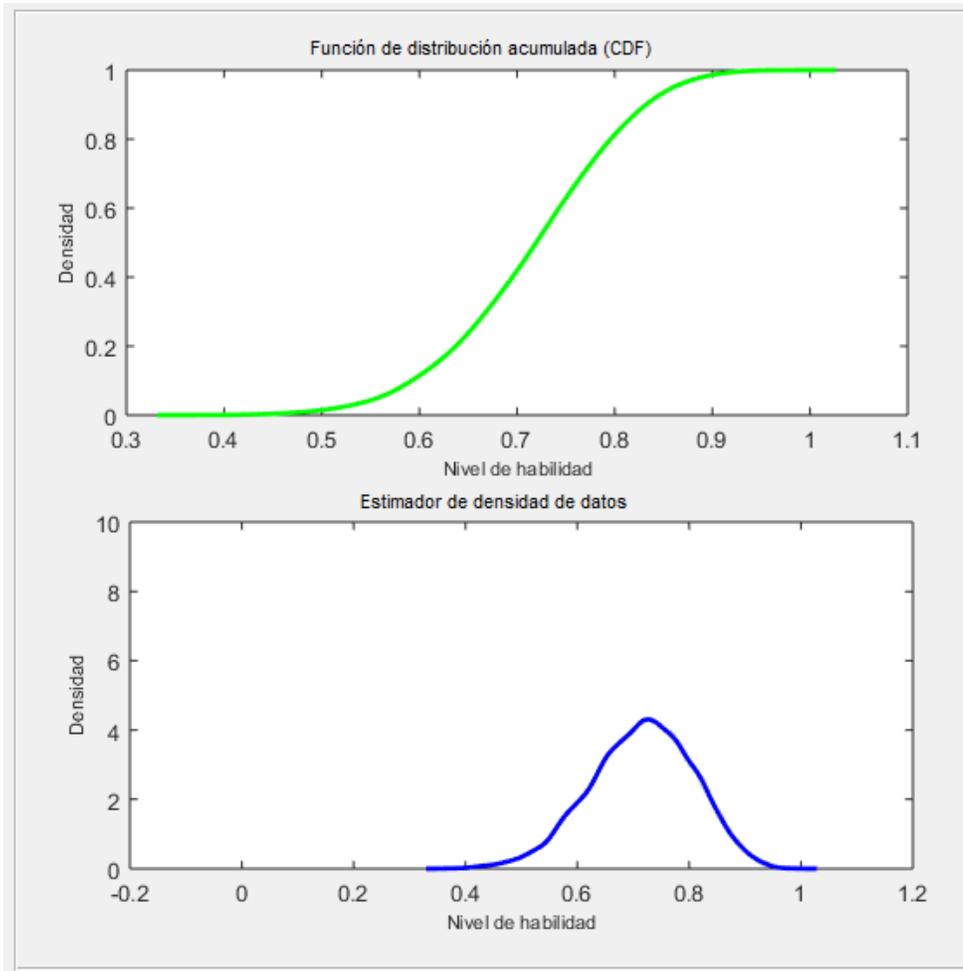
se observa el estimador de densidad de tamaño $B = 10000$ y con valor para $h = 0.0130$. Las observaciones de los núcleos individuales se encuentran en el rango de 0.4 a 0.9 de los niveles de habilidades.



(a) $B=500, h=0.0258$.



(b) $B=2000$, $h=0.0188$.

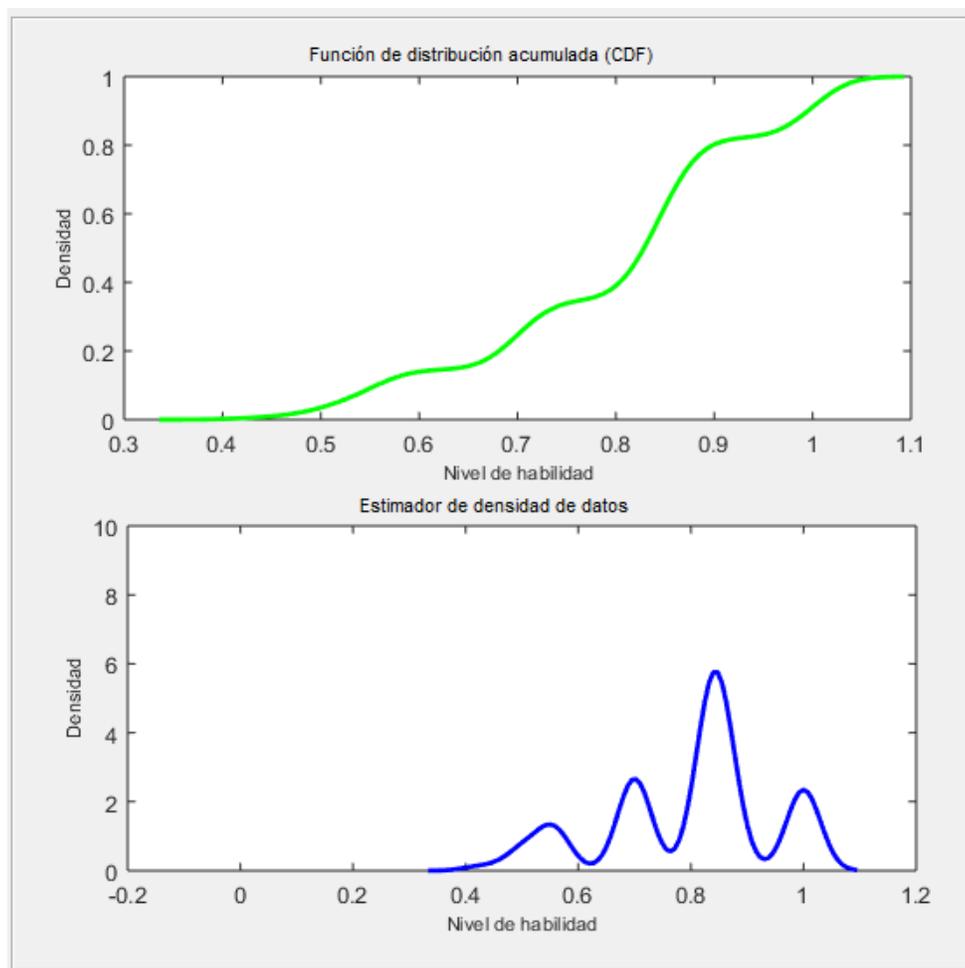


(c) $B=10000, h=0.0130$.

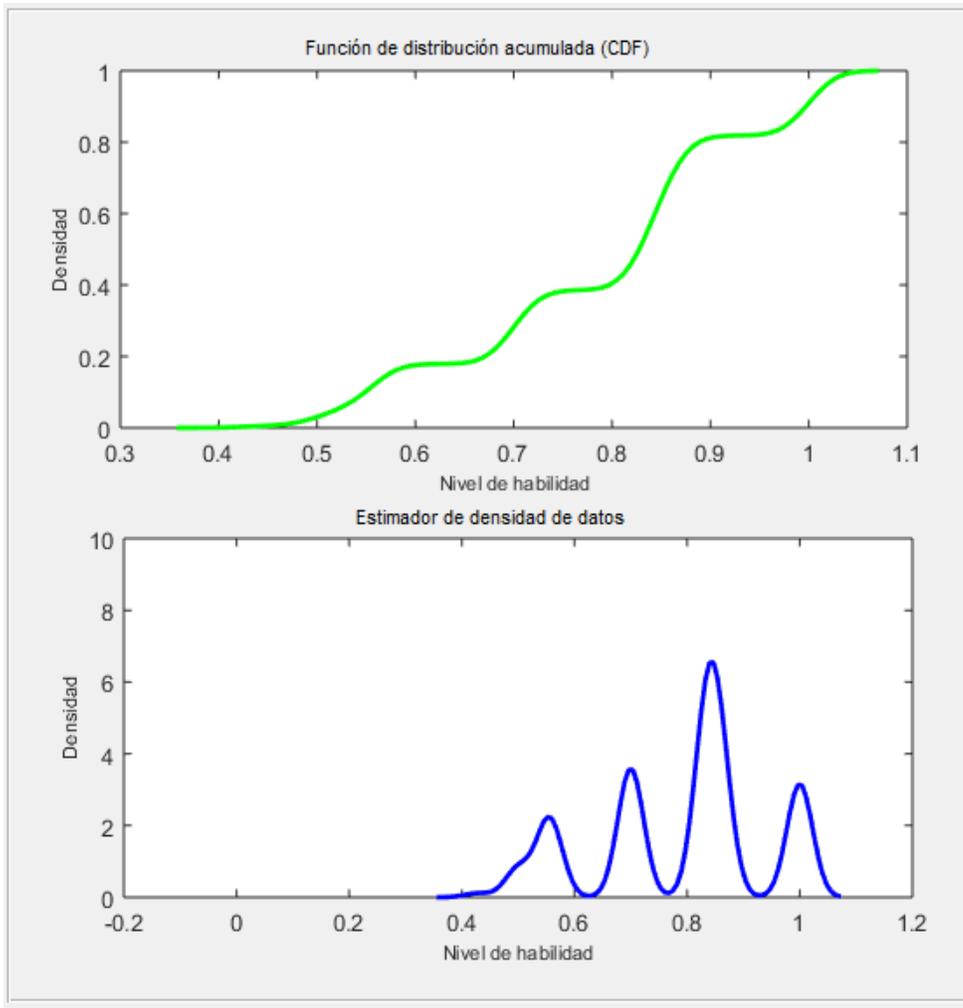
Figura 4.9. Estimadores de densidad usando la media.

En la figura 4.10 se muestran los resultados de los estimadores de densidad al usar la mediana como parámetro de la función *bootstrp*, en la figura 4.10(a) se observa el estimador de densidad de una muestra de datos para niveles habilidad de tamaño $B = 500$ con un ancho de banda óptimo $h = 0.0302$, en la figura 4.10(b) se aprecia el estimador de densidad de tamaño $B = 2000$ y con valor para $h = 0.0229$, en la figura 4.10(c) se representa un estimador de densidad de tamaño $B = 10000$ y un ancho de banda óptimo $h = 0.0166$, en donde los niveles de

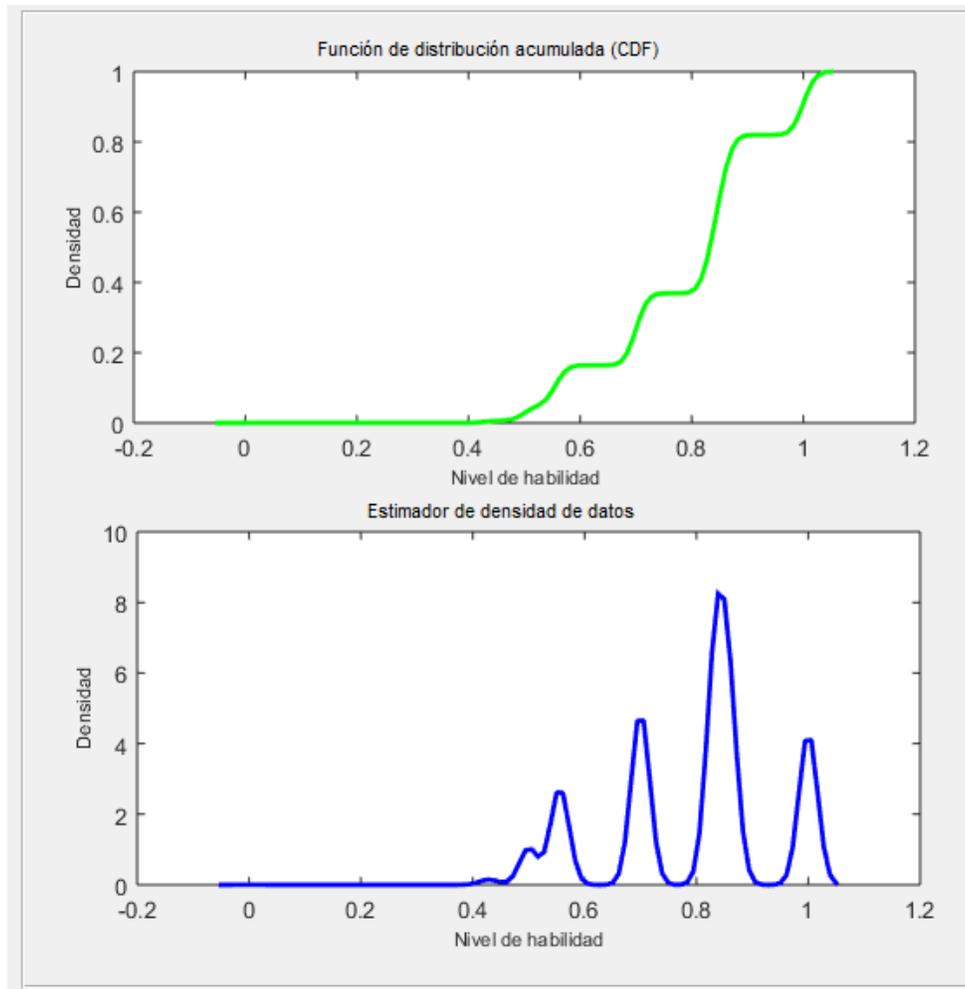
habilidad de las observaciones de los núcleos individuales se encuentran en rango de 0.4 a 1.



(a) $B=500, h=0.5775$.



(b) $B=2000, h=0437$.



(c) $B=10000, h=0.317$.

Figura 4.10. Estimadores de densidad usando la mediana.

En los resultados obtenidos en las distribuciones empíricas de las muestras experimentales para los niveles de habilidad, el valor de la integral definida del área bajo la curva de la CDF se aproxima a la unidad. El valor de la CDF cuando se utiliza la moda como parámetro de la función *bootstrap* se tiene un valor = 0.9996, al usar la media como parámetro en el remuestreo = 0.9999 y el valor de la CDF al utilizar la mediana como parámetro de la función *bootstrap* se obtiene = 0.9996.

Los resultados de las pruebas de hipótesis estadística usando la función *kstest* arrojaron valores para $H = 1$, con lo cual se puede decir que se rechaza la hipótesis de normalidad de las distribuciones empíricas, es decir, se acepta que las distribuciones obtenidas con la función *bootstrp* tienen comportamiento multimodal.

Conclusiones

Una vez obtenidos los resultados de las experimentaciones y después de haber validado los mismos, se puede concluir que las técnicas aplicadas en este trabajo son eficientes para detectar multimodalidad sobre la muestra de datos de resultados de la prueba de la figura compleja de Rey-Osterrieth. Utilizando el método bootstrap no paramétrico se generaron distribuciones empíricas cuya naturaleza no está sujeta al supuesto de normalidad, el estimador de densidad de núcleo y la prueba de contraste de hipótesis estadística de Kolmogorov-Smirnov fueron muy útiles para corroborar que los resultados obtenidos a partir del remuestreo permiten obtener distribuciones multimodales.

Como trabajo futuro se pueden tomar los resultados que se obtuvieron del uso de las técnicas tanto de remuestreo como de estimación de densidad, para aplicarlos de manera previa en la selección de un modelo psicométrico en evaluaciones

basadas en IRT. De esta manera se podría identificar la presencia de multimodalidad en los resultados de este tipo de pruebas y en consecuencia se podría elegir un modelo psicométrico que pueda trabajar con datos de esta naturaleza. Tal modelo, seguramente, tendría más de tres parámetros, ya que a la fecha no existe en la literatura ninguna referencia que mencione la capacidad del modelo 3PL para tratar datos multimodales. Además, lo anterior sería de gran ayuda en la implementación de sistemas CAT, ya que supondría una mejora a los ya existentes. Por otro lado, la implementación de técnicas de reconocimiento de patrones, principalmente aquellas orientadas a imágenes, podrían ser de utilidad para una posible automatización en el proceso de evaluación de los dibujos obtenidos en la aplicación de la prueba de la figura compleja de Rey-Osterrieth.

Bibliografía

Abal, F. J., Lozzia, G. S., Aguerri, M. E., Galibert, M. S., & Attorresi, H. F. (2010). La escasa aplicación de la teoría de respuesta al ítem en test de ejecución típica. *Revista colombiana de psicología*, 111-122.

Aranciva, F., Casals, C. M., Sánchez Benavides, G., Quintana, M., Manero, R. M., Rognoni, T., . . . Peña Casanova, J. (2012). Estudios normativos españoles en población adulta joven (Proyecto NEURONORMA jóvenes): normas para el Boston Naming Test y el Token Test. *Neurología*, 394-399.

Attorresi, H. F., Lozzia, G. S., Abal, F. J., Galibert, M. S., & Aguerri, M. E. (2009). Teoría de respuesta al ítem. conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de clínica psicológica*, 179-188.

Aybek, E. C. (2016). Concerto: A Framework for Computerized Adaptive Test Applications. *Journal of Measurement and Evaluation in Education and Psychology*, 251-271.

Aybek, E. C., & Demirtasli, R. N. (2017). Computerized adaptive test (CAT) applications and item response theory models for polytomous items. *International Journal of Research in Education and Science (IJRES)*, 475-487.

Badil, M. H., Castillo, J., Wong, A., & Landeros, J. (2006). Precisión de los índices estadísticas: Técnicas jackknife & bootstrap. *Innovaciones de Negocios*, 63-78.

- Baker, F. (2001). *The Basic of Item Response Theory*. Wiconsin, United States of America: Eric Clearnghouse on Asseement and Evaluation.
- Baker, F. B. (1985). *The basic of item response theory*. Boston: ERIC Clearinghouse on Assessment and Evaluation.
- Barbero García, M. I. (1999). Desarrollos recientes de los modelos psicométricos de la teoría de respuesta a los ítems. *Psicothema*, 195-210.
- Barrada, J. R. (2012). Tests adaptativos informatizados: una perspectiva general. *Anales de Psicología*, 289-302.
- Barton, M. A., & Lord, F. M. (1981). *An upper asymptote for the three-parameter logistic item-response model*. Princeton: Educational Testing Service.
- Bianchi, M. (1995). *Testing for convergence: evidence from nonparametric multimodality test*. Londres: Bank of England.
- Birnbaum, A. (1967). *Statistical theory for logistic mental test models whith a prior distribution of ability*. Princenton: Educational Testing Service.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Raschb Mode: Fundamental measurement in human sciences*. New York: Routledge Taylor & Francis Group.
- Boomsma, A., van Duijn, M. A., & Snijders, T. A. (2001). *Essays on item response theory*. New York: Springer Science + Business Media New York.

- Bourel, M. (2011). Comparación en la elección de una ventana óptima para algunos estimadores de densidad. *Memoria Investigaciones en Ingeniería*, 1-16.
- Brizuela Rodríguez, A. (2015). Los modelos de mezcla de la teoría de respuesta al ítem. *Actualidades en Psicología*, 79-90.
- Cheng, Y., & Ray, S. (2014). Multivariate Modality Inference Using Gaussian Kernel. *Open Journal of Statistics*, 419-434.
- Chernick, M. R. (2008). *Bootstrap Methods: A Guide for Practitioners and Researches*. Hoboken: John Wiley & Sons, Inc.
- Chernick, M. R., & LaBudde, R. A. (2011). *An introduction to bootstrap methods with applications to R*. Hoboken: John Wiley & Sons, Inc.
- Choi, S. (2009). Firestar: Computerized Adaptive Testing (CAT) Simulation Program for Polytomous IRT Models. *Applied psychological measurement*, 644-645.
- Cortés S, J. F., Galindo y Villa M, G., & Salvador C, J. (1996). La figura compleja de Rey: propiedades psicométricas. *Salud Mental*, 42-48.
- Cupani, M., & Cortez, F. D. (2016). Análisis psicométricos del subtest de razonamiento numérico utilizando el modelo de Rasch. *Revista de Psicología*, 1-16.
- de Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York: The Guilford.

- DeMars, C. (2010). *Item response theory*. New York: Oxford University Press Inc.
- Duarte Pedroza, L., Espitia, A., & Montañés, P. (2016). Aportes y limitaciones del Boston Naming Test: evidencia a partir de controles colombianos. *Acta Neurológica Colombiana*, 290-296.
- Efron, B. (1980). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, 589-599.
- Efron, B., & Tibshirani, R. J. (1986). Bootstrap Methods for Standards Errors, Confidence Interval, and Other Measures of Statical Accuracy. *Statistical Science*, 54-77.
- Facchinetti, S. (2009). A procedure to find exact critical values of Kolmogorov-Smirnov test. *Statistica Applicata - Italian Journal of Applied Statistics*, 337-359.
- Green, B. F. (2011). A Comment on Early Student Blunders on Computer-Based Adaptive Tests. *Applied Psychological Measurement*, 165.174.
- Hall, P. (1992). *The Bootstrap and Edgewort Expansion*. New York: Springer.
- Hall, P., & Huang, L.-S. (2002). Unimodal density estimation using kernel methods. *Statistica Sinica*, 956-990.
- Hall, P., & York, M. (2001). On the calibration of Silverman's test for multimodality. *Statistica Sinica*, 515-536.

- Hambleton, R. K., & Van der Linden, W. J. (1982). Advances in item response theory and applications: an introduction. *Applied psychological measurement*, 373-378.
- Hambleton, R. K., Robin, F., & Xinh, D. (2000). Item response models for the analysis of education and psychological test data. *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, 553-581.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of ítem response theory*. Newbury Park: Sage Publications, Inc.
- Hernández Abreu, E., & Martínez Pérez, M. (2012). *El método bootstrap en la estimación de incertidumbres*. La Habana: Boletín científico tecnico INIMENT.
- Herranz Torres, N. (2012). La medida de la inteligencia. Primeros acercamientos, desarrollo y procedimientos actuales. *Revista de investigación y divulgación en psicología y logopedia*, 21-25.
- Hidalgo Montesinos , M. D., & French, B. F. (2016). Una introducción didáctica a la Teoría de Respuesta al Ítem para comprender la construcción de escalas. *Revista de Psicología Clínica con Niños y Adolescentes*, 13-21.
- Jiménez Alfaro, K., & Montero Rojas, E. (2013). Aplicación del modelo de Rasch, en el análisis psicométrico de una prueba de diagnóstico de matemáticas. *Revista digital Matemática, Educación e Internet*, 1-24.
- Jones, L. V., & Thissen, D. (2007). A history and overview of psychometrics. *Handbook of Statistics*, 1-27.

- Joyanes Aguilar, L. (2003). *Fundamentos de programación: Algoritmos, estructura de datos y objetos*. Madrid: Mc Graw Hill.
- Kalender, I., & Berberoglu, G. (2017). Can Computerized Adaptive Testing Work in Students' Admission to Higher Education Programs in Turkey? *Educational Sciences: Theory & Practice*, 573–596.
- Liao, W.-W., Ho, R.-G., & Yen, Y.-C. (2012). The four-parameter logistic item response theory model as a robust method of estimating ability despite aberrant responses. *Social behavior and personality*, 1679-1694.
- Loken, E., & Rulison, K. L. (2010). Estimation of a 4-parameter Item Response Theory model. *The British Journal of Mathematical and Statistical Psychology*, 509-525.
- López Cuadrado, J., & Sánchez Fernández, J. M. (2005). GenTAI: generador de tests adaptativos informatizados. *Revista Iberoamericana de Informática Educativa*, 9-24.
- López, R., Sanmartín, P., & Méndez, F. (2014). Revisión de las evaluaciones adaptativas computarizadas (CAT). *Educación y Humanismo*, 27-40.
- Lord, F. (1952). *A theory of test Scores*. Richmond: Psychometric Corporation.
- Lu, P., & Xiao, C. (2016). The Research on Computerized Adaptive Testing. *Journal of Physics: Conference Series*, 1-10.
- Magis, D. (2013). A Note on the Item Information Function of the Four-Parameter Logistic Model. *Applied Psychological Measurement*, 304–315.

- Magis, D., & Raiche, G. (2012). Random generation of response patterns under computerized adaptive testing with the R package catR. *Journal of Statistical Software*, 1-19.
- Martín Guaregua, N., Díaz Torres, C., & Picaquart, M. (2011). Calibración de una prueba de química por el modelo de Rasch. *Revista electrónica de Investigación Educativa*, 132-148.
- Martínez Arias, M. R., Hernández Lloreda, M. V., & Hernández Lloreda, M. J. (2014). *Psicometría*. Madrid: Alianza Editorial S. A.
- Martínez Cambor, P. (2008). Estudio sobre los efectos del parámetro de suavizado en contrastes no paramétricos para k-muestras. *Revista Colombiana de Estadística*, 157-168.
- Minnotte, M. C. (1997). Nonparametric testing of existence of modes. *The Annals of Statistics*, 1646-1660.
- Mitrushina, M., Boone, K. B., Razini, J., & D'Elia, L. F. (2005). *Handbook of Normative Data for Neuropsychological Assessment*. New York: Oxford University Press, Inc.
- Muñiz, J. (1998). La medición de lo psicológico. *Psicothema*, 1-21.
- Muñiz, J. (2010). Las teorías de los test: teoría clásica y teoría de respuesta a los ítems. *Papeles del psicólogo*, 57-66.
- Muñiz, J., & Hambleton, R. K. (1992). Medio siglo de teoría de respuesta al ítem. *Anuario de psicología*, 44-66.

- Olea, J., & Ponsoda, V. (2004). *Test Adaptativos Informatizados*. Madrid: UNED.
- Olea, J., Abad, F. J., & Barrada, J. R. (2010). Test informatizados y otros nuevos tipos de tests. *Papeles del Psicólogo*, 94-107.
- Oppl, S., Reisinger, F., Eckmaier, A., & Helm, C. (2017). A flexible online platform for computerized adaptive testing. *International Journal of Educational Technology in Higher Education*, 1-21.
- Parzen, E. (1962). On estimation of probability density function and mode. *Annals of mathematical statistics*, 1065-1076.
- Pedrosa, I., Juarros Basterretxea, J., Robles Fernández, A., Basteiro, J., & García Cueto, E. (2014). Pruebas de bondad de ajuste en distribuciones simétricas, ¿qué estadístico utilizar? *Universitas Psychologica*, 245-254.
- Pressman, R. S. (2010). *Ingeniería del Software un enfoque práctico*. New York: McGraw Hill.
- Prieto, G., & Delgado, A. D. (2010). Fiabilidad y Validez. *Papeles del Psicólogo*, 67-74.
- Prieto, G., & Delgado, A. R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 94-100.
- Ramírez, I. C., Barrera, C. J., & Correa, J. C. (2013). Efecto de tamaño de la muestra y el número de réplicas bootstrap. *Ingeniería y Competitividad*, 93-101.
- Rasch, G. (1960). *Probabilistic model for some attainment and intelligence test*. Copenhagen.

- Rasch, G. (1963). *The Poisson Process as a Model for a Diversity of Behavioral Phenomena*. Washington.
- Rasch, G. (196x). *On Objectivity and Specific of the probabilistic basis for testing*. Copenhagen.
- Rey, A. (1987). *REY: Test de copia y de reproducción de memoria de figuras geométricas complejas* (8 ed.). Madrid: TEA Ediciones.
- Sheather, S. J. (2004). Density Estimation. *Statistical Science*, 588-597.
- Silverman, B. W. (1981). Using Kernel density estimates to investigate multimodality. *J.R. Statist. Soc*, 97-99.
- Simanca Herrera, F., Abuchar Porras, A., & Rivera Osorio, P. J. (2014). ODISEO: Algoritmo de evaluación inteligente en e-learning. *Congreso Iberoamericano de Ciencia, Tecnología, Innovación y Educación*, 2-20.
- Solanas, A., & Sierra, V. (1992). Bootstrap: fundamentos e introducción a sus aplicaciones. *Anuario de psicología*, 143-154.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American journal of psychology*, 72-101.
- Stephens, M. A. (1970). Use of the Kolmogorov-Smirnov, Cramér-Von Mises and Related Statistics without Extensive Tables. *Journal of the Royal Statistical Society*, 115-122.
- Valdés Veloz, H. (2007). *Docimología: de la teoría clásica del test a la teoría de respuesta al ítem*. La Habana: Organo editor educación Cubana.

van Buuren, N., & Eggen, T. H. (2017). Latent-Class-Based Item Selection for Computerized. *Journal of Computerized Adaptive Testing*, 22-43.

Weitzenfeld, A. (2005). *Ingeniería de Software Orientada a Objetos*. México: Cengage Learning Editores,S.A. de C.V.

Winston, W. (2014). *Microsoft Excel 2013: Data analysis and bussines modeling*. Microsoft Press.

Anexos

A. Sistemas CAT existentes

Nombre	Descripción	Distribución
Open Source	Es un sistema implementado basado en la teoría de respuesta al ítem (IRT) y modelos de diagnóstico cognitivo y algoritmos para la selección de ítems.	Open Source
Computerized Adaptive Testing System (OSCATS)	El desarrollo de CAT y simulaciones mediante los modelos 2PL y 3PL. Bibliotecas escritas en lenguaje C orientadas a objetos utilizando GObject y tiene enlaces a lenguajes de programación Python, Perl, PHP, Java y Matlab mediante Java.	https://code.google.com/archive/p/oscats/wikis/Features/wiki
IRT-CAT	El proyecto de código fuente abierto IRT-CAT proporciona un código de CAT basado en Internet utiliza los modelos de la teoría de respuesta al ítem 3PL. Desarrollado para web con PHP, JavaScript y PostgreSQL.	Lee, Park & Park
Concerto: R-Based Online Adaptive Testing Platform	Es un sistema CAT basado en web que es gratuito. Está desarrollado en el lenguaje R, de código abierto. Admite elementos en HTML, multimedia integrada y la producción de informes, gráficos y comentarios.	The Psychometrics Center, Universidad de Cambridge https://www.psychometrics.cam.ac.uk/newconcerto
Web Examiner FastCat IRT	Es un sistema CAT basado en web que presenta herramientas de análisis y desarrollo de pruebas IRT.	Web Examiner Limited http://www.webexaminer.com/our-products/fastcat-assessment-operations
CAT-ASVAB	Primera evaluación adaptativa computarizada a gran escala en EE.UU. Usada para el alistamiento de personal en las fuerzas armadas, y es obligatoria para todo estudiante de último año de bachillerato. Esta prueba está basada en el modelo 3PL de la IRT.	http://official-asvab.com/

Tabla A. 1. Listado de sistemas CAT (Parte 1).

Nombre	Descripción	Distribución
GMAT Graduate Management Admission Test	GMAT es usado en el mundo para medir las aptitudes de los recién graduados en escuelas de Negocios y Finanzas.	Graduate Management Admission Council https://www.mba.com/global
Graduate Record Examination (GRE)	El examen de entrada a los programas de posgrado en las áreas de Ciencias y Humanidades que exigen numerosas universidades de los Estados Unidos, Canadá y algunas instituciones educativas de la Unión Europea que mide los conocimientos generales.	https://www.ets.org/gre
WLE Theta Estimator	Sistema de código abierto implementado en PHP que toma matrices de umbrales de dificultad de ítems, y puntajes de personas para esos ítems, y estima un valor de theta a través de una estimación de verosimilitud ponderada. Puede ser útil para personas y organizaciones que tienen un banco de ítems calibrados con el modelo de Rasch y les gustaría experimentar con la creación de su propio sistema CAT en línea rudimentario.	Ablemetrics
Assessment Center	Es una herramienta gratuita en línea que proporciona funcionalidad para crear y entregar pruebas adaptativas.	Universidad de Northwestern http://assessmentcenter.net
FIRESTAR: CAT Simulator	Es adecuado para su uso con modelos de respuesta a ítems poltómicos. (Choi, 2009)	Patient Reported Outcomes Northwestern University's Feinberg School of Medicine.
FastTEST Web	Proporciona un entorno fácil de usar para el ciclo completo de evaluaciones basadas en CAT.	ASC, 4ROI http://www.assess.com/fasttest/
Web eXaminer FastCat IRT	Es un sistema CAT basado en web que presenta herramientas de análisis y desarrollo de pruebas IRT.	WebExaminer Limited http://www.webexaminer.com/our-products/fastcat-assessment-operations

Tabla A. 2. Listado de sistemas CAT (Parte 2).

Nombre	Descripción	Distribución
Smart Test Technology®	<p>Es la primera plataforma multidimensional de pruebas adaptativas en el mundo.</p> <p>Es capaz de administrar pruebas adaptativas de computadora basadas en Internet sobre la base de una o más habilidades.</p> <p>Varias funciones de seguridad incorporadas incluyen opciones de preguntas aleatorias, deshabilitar la impresión de pantalla y navegación controlada.</p> <p>Smart Test Technology® se ha utilizado para pruebas de alto riesgo en universidades y departamentos de policía en los Estados Unidos.</p>	<p>Adaptive Assessment Services (AAS), Inc.</p> <p>http://www.assess.com/adaptive-testing/</p>
CATSim	<p>Ejecuta tres tipos de simulación de datos para las etapas clave del desarrollo de CAT.</p> <p>Trabaja con ítems dicotómicos y cinco modelos de ítems politómicos.</p>	<p>ASC</p> <p>http://www.assess.com/cat-sim/</p>
LeaderAmp	<p>CAT móvil y sistema experto de eCoaching calibrado psicométricamente.</p>	<p>LeaderAmp</p>

Tabla A. 3. Listado de Sistemas CAT (Parte 3).

B. Matriz de resultados de la prueba de Rey-Osterrieth

Matriz de resultados de la prueba de la figura compleja de Rey Osterrieth en la fase de reproducción, la cual está compuesta por 18 columnas y 71 filas codificadas de manera dicotómica.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18
Sujeto 1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0	0	1
Sujeto 2	0	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	0	0
Sujeto 3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Sujeto 4	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	0	0	1
Sujeto 5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Sujeto 6	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0
Sujeto 7	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	0
Sujeto 8	0	1	1	1	1	1	0	1	0	0	1	1	1	1	0	1	0	0
Sujeto 9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Sujeto 10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Sujeto 11	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
Sujeto 12	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0
Sujeto 13	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0
Sujeto 14	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1
Sujeto 15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Sujeto 16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Sujeto 17	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1
Sujeto 18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Sujeto 19	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1
Sujeto 20	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0
Sujeto 21	0	1	1	1	1	0	0	1	1	0	1	1	1	0	0	1	0	1
Sujeto 22	1	1	1	1	1	1	0	0	1	0	1	0	1	0	0	1	0	0
Sujeto 23	1	1	1	1	1	1	0	1	0	0	1	1	1	0	0	1	0	1
Sujeto 24	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1
Sujeto 25	0	1	1	1	1	1	1	0	1	0	1	0	0	0	0	1	0	0
Sujeto 26	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
Sujeto 27	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	1
Sujeto 28	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0
Sujeto 29	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	0
Sujeto 30	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1
Sujeto 31	0	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	0	0
Sujeto 32	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0
Sujeto 33	1	0	0	1	1	1	1	1	1	0	0	0	0	1	0	1	0	0
Sujeto 34	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
Sujeto 35	1	1	1	1	1	1	0	0	1	0	0	0	1	1	1	1	0	0
Sujeto 36	0	1	1	1	1	0	0	1	0	0	1	0	1	0	0	0	0	1

Tabla B. 1. Matriz de resultados de la prueba de Rey-Osterrieth (parte 1).

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18
Sujeto 37	0	1	1	1	1	1	1	0	0	0	1	1	1	0	1	1	0	0
Sujeto 38	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0
Sujeto 39	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Sujeto 40	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0
Sujeto 41	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0
Sujeto 42	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	0
Sujeto 43	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0
Sujeto 44	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	0	0
Sujeto 45	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	0	0
Sujeto 46	1	1	1	1	1	1	1	1	0	0	1	1	1	0	1	0	0	0
Sujeto 47	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	1	0	0
Sujeto 48	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	0	0
Sujeto 49	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1
Sujeto 50	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0
Sujeto 51	0	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	0
Sujeto 52	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	0
Sujeto 53	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0
Sujeto 54	0	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	0	0
Sujeto 55	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0
Sujeto 56	1	0	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0
Sujeto 57	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	0	0	0
Sujeto 58	1	1	1	1	1	0	1	1	0	0	1	1	1	0	0	1	0	0
Sujeto 59	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Sujeto 60	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
Sujeto 61	1	1	0	1	1	1	0	0	0	0	1	1	1	1	1	1	0	0
Sujeto 62	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1
Sujeto 63	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1
Sujeto 64	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1
Sujeto 65	1	1	1	1	1	1	0	1	1	0	1	1	1	0	0	1	0	0
Sujeto 66	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
Sujeto 67	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	0
Sujeto 68	1	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	0
Sujeto 69	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Sujeto 70	1	1	1	1	1	0	1	1	1	0	0	1	1	1	1	0	0	0
Sujeto 71	0	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	0	0

Tabla B. 2. Matriz de resultados de la prueba de Rey-Osterrieth (parte 2).

C. Matriz de habilidad en escala logit

Matriz de resultados de los niveles de habilidades en la escala logit para la muestra de examinados en la fase de reproducción.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	P	Q	P/Q	Ln0
Sujeto 1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0	0	1	0.7778	0.2222	3.5	1.2528
Sujeto 2	0	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	0	0	0.6667	0.3333	2	0.6931
Sujeto 3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0.9444	0.0556	17	2.8332
Sujeto 4	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	0	0	1	0.7222	0.2778	2.6	0.9555
Sujeto 5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0.9444	0.0556	17	2.8332
Sujeto 6	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0.8889	0.1111	8	2.0794
Sujeto 7	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	0	0.7778	0.2222	3.5	1.2528
Sujeto 8	0	1	1	1	1	1	0	1	0	0	1	1	1	1	0	1	0	0	0.6111	0.3889	1.5714	0.4520
Sujeto 9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0.9444	0.0556	17	2.8332
Sujeto 10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	-	-
Sujeto 11	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0.9444	0.0556	17	2.8332
Sujeto 12	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0.8333	0.1667	5	1.6094
Sujeto 13	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	0.8333	0.1667	5	1.6094
Sujeto 14	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0.8889	0.1111	8	2.0794
Sujeto 15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0.9444	0.0556	17	2.8332
Sujeto 16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0.8333	0.1667	5	1.6094
Sujeto 17	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0.8889	0.1111	8	2.0794
Sujeto 18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0.9444	0.0556	17	2.8332
Sujeto 19	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	0.7778	0.2222	3.5	1.2528
Sujeto 20	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	0.8333	0.1667	5	1.6094
Sujeto 21	0	1	1	1	1	0	0	1	1	0	1	1	1	0	0	1	0	1	0.6111	0.3889	1.5714	0.4520
Sujeto 22	1	1	1	1	1	1	0	0	1	0	1	0	1	0	0	1	0	0	0.5556	0.4444	1.25	0.2231
Sujeto 23	1	1	1	1	1	0	1	0	0	1	1	1	0	0	0	1	0	1	0.6667	0.3333	2	0.6931
Sujeto 24	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	0.8333	0.1667	5	1.6094
Sujeto 25	0	1	1	1	1	1	1	0	1	0	1	0	0	0	0	1	0	0	0.5000	0.5000	1	0.0000
Sujeto 26	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0.7778	0.2222	3.5	1.2528
Sujeto 27	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	1	0.8333	0.1667	5	1.6094
Sujeto 28	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	0.7778	0.2222	3.5	1.2528
Sujeto 29	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0.7778	0.2222	3.5	1.2528
Sujeto 30	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	0.8889	0.1111	8	2.0794
Sujeto 31	0	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	0	0	0.6111	0.3889	1.5714	0.4520
Sujeto 32	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0.8889	0.1111	8	2.0794
Sujeto 33	1	0	0	1	1	1	1	1	1	0	0	0	0	1	0	1	0	0	0.5000	0.5000	1	0.0000
Sujeto 34	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0.9444	0.0556	17	2.8332
Sujeto 35	1	1	1	1	1	1	0	0	1	0	0	0	1	1	1	1	0	0	0.6111	0.3889	1.5714	0.4520
Sujeto 36	0	1	1	1	1	0	0	1	0	0	1	0	1	0	0	0	0	1	0.4444	0.5556	0.8	-0.2231

Tabla C. 1. Matriz de niveles de habilidad en escala logit (Parte 1).

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	P	Q	P/Q	Ln0
Sujeto 37	0	1	1	1	1	1	1	0	0	0	1	1	1	0	1	1	0	0	0.6111	0.3889	1.5714	0.4520
Sujeto 38	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	0.8333	0.1667	5	1.6094
Sujeto 39	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0.8333	0.1667	5	1.6094
Sujeto 40	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	0.8333	0.1667	5	1.6094
Sujeto 41	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0.8889	0.1111	8	2.0794
Sujeto 42	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	0	0.8333	0.1667	5	1.6094
Sujeto 43	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	0.8333	0.1667	5	1.6094
Sujeto 44	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	0	0	0.7222	0.2778	2.6	0.9555
Sujeto 45	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	0	0	0.7222	0.2778	2.6	0.9555
Sujeto 46	1	1	1	1	1	1	1	1	0	0	1	1	1	0	1	0	0	0	0.6667	0.3333	2	0.6931
Sujeto 47	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	1	0	0	0.7222	0.2778	2.6	0.9555
Sujeto 48	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	0	0	0.7222	0.2778	2.6	0.9555
Sujeto 49	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0.8889	0.1111	8	2.0794
Sujeto 50	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	0.7222	0.2778	2.6	0.9555
Sujeto 51	0	1	1	1	1	0	0	0	1	1	1	1	1	0	0	1	0	0	0.6667	0.3333	2	0.6931
Sujeto 52	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	0	0.7778	0.2222	3.5	1.2528
Sujeto 53	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	0.7778	0.2222	3.5	1.2528
Sujeto 54	0	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	0	0	0.6667	0.3333	2	0.6931
Sujeto 55	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	0.7778	0.2222	3.5	1.2528
Sujeto 56	1	0	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0.3333	0.6667	0.5	-0.6931
Sujeto 57	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0.6667	0.3333	2	0.6931
Sujeto 58	1	1	1	1	1	0	1	1	0	0	1	1	1	0	0	1	0	0	0.6111	0.3889	1.5714	0.4520
Sujeto 59	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	-	-
Sujeto 60	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0.8889	0.1111	8	2.0794
Sujeto 61	1	1	0	1	1	1	0	0	0	0	1	1	1	1	1	1	0	0	0.6111	0.3889	1.5714	0.4520
Sujeto 62	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0.8889	0.1111	8	2.0794
Sujeto 63	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1	0.8333	0.1667	5	1.6094
Sujeto 64	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0.8333	0.1667	5	1.6094
Sujeto 65	1	1	1	1	1	0	1	1	0	1	1	1	1	0	0	1	0	0	0.6667	0.3333	2	0.6931
Sujeto 66	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0.9444	0.0556	17	2.8332
Sujeto 67	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	0	0.8333	0.1667	5	1.6094
Sujeto 68	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	0	0	0.7778	0.2222	3.5	1.2528
Sujeto 69	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0.9444	0.0556	17	2.8332
Sujeto 70	1	1	1	1	1	0	1	1	1	0	0	1	1	1	1	0	0	0	0.6667	0.3333	2	0.6931
Sujeto 71	0	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	0	0	0.6667	0.3333	2	0.6931

Tabla C. 2. Matriz de niveles de habilidad en escala logit (Parte 2).

D. Niveles de dificultad en escala logit

Niveles de dificultad calculados de manera experimental sobre el conjunto de datos de la prueba de la figura compleja de Rey-Osterrieth.

Ítem	μ	$1 - \mu$	$\frac{\mu}{1 - \mu}$	$b = \ln\left(\frac{\mu}{1 - \mu}\right)$
M1	0.7681	0.2318	3.3125	1.1977
M2	0.9710	0.0289	33.5	3.5115
M3	0.9420	0.0579	16.25	2.7880
M4	0.9855	0.0144	68	4.2195
M6	0.8115	0.1884	4.3076	1.4604
M7	0.8115	0.1884	4.3076	1.4604
M8	0.8550	0.1449	5.9	1.7749
M9	0.7681	0.2318	3.3125	1.1977
M10	0.3333	0.6666	0.5	-0.6931
M11	0.9275	0.0724	12.8	2.5494
M12	0.8550	0.1449	5.9	1.7749
M13	0.9710	0.0289	33.5	3.5115
M14	0.7826	0.2173	3.6	1.2809
M15	0.5072	0.4927	1.0294	0.0289
M16	0.8260	0.1739	4.75	1.5581
M17	0.3768	0.6231	0.6046	-0.5031
M18	0.3768	0.6231	0.6046	-0.5031

Tabla D. 1. Niveles de dificultad en la escala logit.

E. Matriz de resultados al discriminar filas y columnas

Matriz de datos que contiene las puntuaciones después de haber removido los valores que contenían cero o uno en filas y columnas, donde se eliminó al ítem M5 ya que los resultados obtenidos por los 71 estudiantes eran de valor uno para toda la columna, al eliminar esta columna, las filas 10 y 59 presentaron el mismo fenómeno ya que los estudiantes tenían respuesta a todos los ítems igual a uno.

	M1	M2	M3	M4	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18
Sujeto 1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0	0	1
Sujeto 2	0	1	1	1	0	1	1	1	0	1	1	1	1	0	1	0	0
Sujeto 3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Sujeto 4	1	1	1	1	1	1	0	1	0	1	1	1	1	0	0	0	1
Sujeto 5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Sujeto 6	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0
Sujeto 7	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	0
Sujeto 8	0	1	1	1	1	0	1	0	0	1	1	1	1	0	1	0	0
Sujeto 9	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Sujeto 11	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Sujeto 12	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0
Sujeto 13	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0
Sujeto 14	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1
Sujeto 15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Sujeto 16	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Sujeto 17	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1
Sujeto 18	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Sujeto 19	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1
Sujeto 20	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0
Sujeto 21	0	1	1	1	0	0	1	1	0	1	1	1	0	0	1	0	1
Sujeto 22	1	1	1	1	1	0	0	1	0	1	0	1	0	0	1	0	0
Sujeto 23	1	1	1	1	1	0	1	0	0	1	1	1	0	0	1	0	1
Sujeto 24	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1
Sujeto 25	0	1	1	1	1	1	0	1	0	1	0	0	0	0	1	0	0
Sujeto 26	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
Sujeto 27	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	1
Sujeto 28	0	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0
Sujeto 29	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	0
Sujeto 30	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1
Sujeto 31	0	1	0	1	1	1	1	0	0	0	1	1	1	1	1	0	0
Sujeto 32	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0
Sujeto 33	1	0	0	1	1	1	1	1	0	0	0	0	1	0	1	0	0
Sujeto 34	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
Sujeto 35	1	1	1	1	1	0	0	1	0	0	0	1	1	1	1	0	0
Sujeto 36	0	1	1	1	0	0	1	0	0	1	0	1	0	0	0	0	1

Tabla E. 1. Matriz de resultados al discriminar las filas y columnas (Parte 1).

	M1	M2	M3	M4	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18
Sujeto 37	0	1	1	1	1	1	0	0	0	1	1	1	0	1	1	0	0
Sujeto 38	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0
Sujeto 39	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Sujeto 40	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0
Sujeto 41	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0
Sujeto 42	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	0
Sujeto 43	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0
Sujeto 44	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	0	0
Sujeto 45	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	0	0
Sujeto 46	1	1	1	1	1	1	1	0	0	1	1	1	0	1	0	0	0
Sujeto 47	1	1	1	1	1	1	0	1	0	1	1	1	1	0	1	0	0
Sujeto 48	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	0	0
Sujeto 49	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1
Sujeto 50	0	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0
Sujeto 51	0	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	0
Sujeto 52	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	0
Sujeto 53	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0
Sujeto 54	0	1	1	1	0	1	1	1	0	1	1	1	1	0	1	0	0
Sujeto 55	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0
Sujeto 56	1	0	0	0	1	0	0	0	0	1	0	1	0	0	1	0	0
Sujeto 57	1	1	1	1	1	0	0	0	1	1	1	1	1	1	0	0	0
Sujeto 58	1	1	1	1	0	1	1	0	0	1	1	1	0	0	1	0	0
Sujeto 60	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
Sujeto 61	1	1	0	1	1	0	0	0	0	1	1	1	1	1	1	0	0
Sujeto 62	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1
Sujeto 63	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1
Sujeto 64	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1
Sujeto 65	1	1	1	1	1	0	1	1	0	1	1	1	0	0	1	0	0
Sujeto 66	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
Sujeto 67	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	0
Sujeto 68	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	0
Sujeto 69	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Sujeto 70	1	1	1	1	0	1	1	1	0	0	1	1	1	1	0	0	0
Sujeto 71	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	0	0

Tabla E. 2. Matriz de resultados al discriminar las filas y columnas (Parte 2).

F. Matriz de clasificación de habilidades ordenadas

Matriz de datos resultado de la clasificación de los niveles de habilidad, donde se pueden observar 11 niveles de habilidades diferentes obtenidos de manera empírica.

Habilidad	-0.8755	-0.3567	-0.1178	0.1178	0.3567	0.6061	0.8755	1.1787	1.5404	2.0149	2.7726
M1	1	0	0.5	1	0.4286	0.5556	0.8333	0.7000	0.8571	1	1
M2	0	1	0.5	1	1	1	1	1	1	1	1
M3	0	1	0.5	1	0.7143	1	1	1	1	1	1
M4	0	1	1	1	1	1	1	1	1	1	1
M6	1	0	1	1	0.71429	0.55556	0.8333	0.7	0.8571	0.9	1
M7	0	0	1	0	0.42857	0.55556	1	0.9	0.9286	1	1
M8	0	1	0.5	0	0.57143	0.77778	0.6667	1	1	1	1
M9	0	0	1	1	0.28571	0.66667	0.6667	0.9	0.8571	0.9	1
M10	0	0	0	0	0	0.22222	0.1667	0.4	0.3571	0.4	0.7778
M11	1	1	0.5	1	0.71429	0.88889	0.8333	1	1	1	1
M12	0	0	0	0	0.85714	1	1	0.9	1	0.8	0.8889
M13	1	1	0	1	1	1	1	1	1	1	1
M14	0	0	0.5	0	0.57143	0.55556	1	0.8	0.8571	1	1
M15	0	0	0	0	0.57143	0.44444	0.3333	0.4	0.5714	0.8	0.6667
M16	1	0	1	1	1	0.55556	0.5	0.8	0.8571	1	1
M17	0	0	0	0	0	0.11111	0	0.3	0.5714	0.6	1
M18	0	1	0	0	0.14286	0.11111	0.1667	0.2	0.2857	0.5	0.6667

Tabla F. 1. Habilidades ordenadas.

G. Niveles de habilidad para el ítem M1

Valores del ítem M1 obtenidos al aplicar el modelo 1PL, con valor de $b = 0.9924$, con un rango de -4 a 4 para niveles de habilidad, ocupados para construir la gráfica de la CCI en la figura 4.2.

θ	Ítem M1	θ	Ítem M1
-4	0.007216937	0.1	0.304897951
-3.9	0.0079699	0.2	0.326494835
-3.8	0.008800725	0.3	0.34885374
-3.7	0.009717311	0.4	0.371898331
-3.6	0.010728325	0.5	0.395540489
-3.5	0.01184327	0.6	0.41968137
-3.4	0.013072555	0.7	0.444212856
-3.3	0.014427572	0.8	0.469019355
-3.2	0.015920775	0.9	0.493979908
-3.1	0.017565766	1	0.518970506
-3	0.019377375	1.1	0.543866546
-2.9	0.021371757	1.2	0.568545299
-2.8	0.023566475	1.3	0.592888311
-2.7	0.025980589	1.4	0.616783618
-2.6	0.02863475	1.5	0.640127706
-2.5	0.031551275	1.6	0.662827133
-2.4	0.034754228	1.7	0.684799784
-2.3	0.038269482	1.8	0.705975714
-2.2	0.042124773	1.9	0.726297601
-2.1	0.04634973	2	0.745720814
-2	0.050975885	2.1	0.76421314
-1.9	0.056036643	2.2	0.781754208
-1.8	0.061567226	2.3	0.798334688
-1.7	0.067604563	2.4	0.813955297
-1.6	0.074187124	2.5	0.828625696
-1.5	0.081354696	2.6	0.842363312
-1.4	0.08914808	2.7	0.855192147
-1.3	0.097608707	2.8	0.867141609
-1.2	0.106778162	2.9	0.878245395
-1.1	0.116697611	3	0.888540454
-1	0.127407113	3.1	0.898066041
-0.9	0.138944842	3.2	0.906862882
-0.8	0.151346183	3.3	0.914972437
-0.7	0.164642746	3.4	0.922436277
-0.6	0.178861279	3.5	0.92929556
-0.5	0.194022525	3.6	0.935590604
-0.4	0.210140034	3.7	0.941360546
-0.3	0.227218974	3.8	0.946643081
-0.2	0.245254986	3.9	0.95147427
-0.1	0.26423313	4	0.955888403
0	0.284126984		

Tabla G. 1. Niveles de habilidad para el ítem M1.

H. Implementación del código de la solución

Implementación del código de la solución en Matlab para detectar multimodalidad, se describen los pasos seguidos mediante el uso de funciones propias del lenguaje.

```
function varargout = IRTRO(varargin)

gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                  'gui_Singleton',  gui_Singleton, ...
                  'gui_OpeningFcn', @IRTRO_OpeningFcn, ...
                  'gui_OutputFcn',  @IRTRO_OutputFcn, ...
                  'gui_LayoutFcn',  [] , ...
                  'gui_Callback',   []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end

if nargout
    [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});
end

function IRTRO_OpeningFcn(hObject, eventdata, handles, varargin)

handles.output = hObject;

guidata(hObject, handles);

function varargout = IRTRO_OutputFcn(hObject, eventdata, handles)
varargout{1} = handles.output;
```

```

function lblVentana_Callback(hObject, eventdata, handles)

function lblVentana_CreateFcn(hObject, eventdata, handles)
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUiControlBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

% Funciones para la construcción de la lista desplegable para la
selección de medidas de tendencia central con CreateFcn y Callback
function lstFuncion_Callback(hObject, eventdata, handles)

function lstFuncion_CreateFcn(hObject, eventdata, handles)

if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUiControlBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

% --- Botón que permite cargar el archivo en formato xlsx.
function btnCargar_Callback(hObject, eventdata, handles)
[filename, pathname] = uigetfile({'*.xlsx'}, 'Seleccione un archivo en
formato Excel');
if isequal(filename,0) || isequal(pathname,0)
    return;
else
    archivo=fullfile(filename);
end

x=xlsread(archivo)
handles.x=x
guidata(hObject, handles);

%Función que realiza los cálculos para la detección de multimodalidad.
function btnCalcular_Callback(hObject, eventdata, handles)

```

```
    habilidad=handles.x %Recibe los valores que previamente se
    cargaron desde el archivo con extensión ".xlsx", es decir, el vector de
    datos de las habilidades clasificadas.
```

```
    tamMtra = get(handles.lstTamanoMuestra, 'Value') % Toma el valor para
    el tamaño del remuestreo 500,1000,2000,3000,5000,10000 y 100000.
```

```
%bootstrap número de muestras a obtener.
```

```
    if(tamMtra == 1)
        bootstrap=500;
    elseif(tamMtra == 2)
        bootstrap=1000;
    elseif(tamMtra == 3)
        bootstrap=2000;
    elseif(tamMtra == 4)
        bootstrap=3000;
    elseif(tamMtra == 5)
        bootstrap=5000;
    elseif(tamMtra == 6)
        bootstrap=10000;
    elseif(tamMtra == 7)
        bootstrap=100000;
    else
        bootstrap=0
    end
```

```
%m_tend Indica la función de la medida de tendencia central que se está
%utilizando.
```

```
    salida=get(handles.lstFuncion, 'Value') %Toma el valor para la
    selección de la función de medida de tendencia central.
```

```
    if(salida==1)
        m_tend='mode';
    elseif(salida==2)
        m_tend='mean';
    elseif(salida==3)
        m_tend='median';
    else
        set(handles.text12, 'String', 'none')
    end
```

```

tic %Inicio del comando tic

[boots]=bootstrp(bootstrap,m_tend,habilidad);%Parámetros que recibe
la función bootstrp.

h_opt=0.9*bootstrap^(-0.2)*min(std(boots),iqr(boots)/1.349); %
Estimación del ancho de banda óptimo por el método de Silverman.

set(handles.lblVentana,'String',h_opt) % Muestra el resultado del
ancho de banda óptimo.

[f,xi,h_opt]=ksdensity(boots,'width',h_opt) % Mediante la función
ksdensity se calcula el estimador de densidad de núcleo.

int_definida=trapz(xi,f)% Área bajo la curva por el método de los
trapecios

set(handles.lblCurva,'String',int_definida) % Muestra en pantalla el
resultado del área bajo la curva.

hip_estd=kstest(boots) % Prueba de contraste de hipótesis de
multimodalidad con la función kstest

set(handles.lblHipotesis,'String',int2str(hip_estd)) %Muestra en
pantalla el resultado del valor de kstest con valor H

axes(handles.axes1)% Carga el gráfico de la CDF.

plot(xi,f,'blue','LineWidth',2) %Imprime el resultado en el gráfico
de xi y f

xlabel('Nivel de habilidad','FontName','Arial','FontSize',8);
%Muestra la leyenda Nivel de habilidad en el eje X para el gráfico de
la CDF.

ylabel('Densidad','FontName','Arial','FontSize',8);%Muestra la
leyenda Densidad para el gráfico de la CDF.

xlim([-0.2 1.2]); % Límite de valores para el eje X de la gráfica de
la CDF.

ylim([0 10.0]); %Límite de valores para el eje y de la gráfica del
estimador de densidad de núcleo.

[f,xi,h_opt]=ksdensity(boots,'function','cdf','width',h_opt);
%Gráfica los valores del estimador de densidad de núcleo

axes(handles.axes2) %Muestra el gráfico resultado de la estimación de
densidad de nucleo para mostrar K modas.

plot(xi,f,'g.-','LineWidth',2); %Muestra en pantalla el gráfico de
color azul.

xlabel('Nivel de habilidad','FontName','Arial','FontSize',8);
%Leyenda del eje X Nivel de habilidad

ylabel('Densidad','FontName','Arial','FontSize',8);% Leyenda del
eje Y Densidad

set(handles.text12,'String',toc)

% el comando toc indica el tiempo transcurrido desde el inicio de tic

```

```
% Funciones para la construcción de la lista desplegable tamaño del  
remuestreo con CreateFcn y Callback
```

```
function lstTamanoMuestra_Callback(hObject, eventdata, handles)
```

```
function lstTamanoMuestra_CreateFcn(hObject, eventdata, handles)
```

```
if ispc && isequal(get(hObject, 'BackgroundColor'),  
get(0, 'defaultUicontrolBackgroundColor'))
```

```
    set(hObject, 'BackgroundColor', 'white');
```

```
end
```

I. Tiempos de procesamiento

Se muestran los tiempos de procesamiento de los remuestreos con la función *bootstrap* que viene incluido en el ToolBox de estadística y aprendizaje máquina de Matlab.

Remuestreo	Moda	Media	Mediana
500	0.4539	0.3366	0.3319
1000	0.5806	0.3423	0.2859
2000	0.7292	0.5144	0.5569
3000	1.0616	0.7419	0.6853
5000	1.6275	1.0939	1.1095
10000	2.5828	1.9040	1.9461
100000	21.9907	16.5046	13.0686

(a). Prueba 1.

Remuestreo	Moda	Media	Mediana
500	0.2926	0.2462	0.2118
1000	0.4388	0.3446	0.3063
2000	0.7651	0.5356	0.4886
3000	1.0528	0.7373	0.7425
5000	1.4546	1.3688	1.9294
10000	2.6676	1.8964	2.3276
100000	21.6539	13.3869	13.9898

(b). Prueba 2.

Remuestreo	Moda	Media	Mediana
500	0.3030	0.2239	0.2059
1000	0.4511	0.3359	0.3125
2000	0.6458	0.5394	0.5041
3000	1.3554	0.7292	0.7211
5000	1.4545	1.0741	1.7989
10000	2.7665	1.9782	2.3833
100000	20.8509	13.4198	13.6413

(c). Prueba 3.

Remuestreo	Moda	Media	Mediana
500	0.2662	0.2250	0.2204
1000	0.4415	0.3207	0.3289
2000	0.6728	0.5236	0.4949
3000	1.0616	0.7110	0.7234
5000	1.4979	1.0977	1.1233
10000	2.9417	1.8496	2.3827
100000	21.8524	13.6727	13.1041

(d). Prueba 4.

Remuestreo	Moda	Media	Mediana
500	0.3880	0.2184	0.2262
1000	0.4450	0.3287	0.2960
2000	0.7254	0.5411	0.5243
3000	0.9635	0.7193	0.5961
5000	1.6923	1.2223	1.0970
10000	2.6979	1.9256	2.4564
100000	29.0977	14.0909	12.9048

(e). Prueba 5.

Tabla I. 1. Tiempos de cálculo al usar las funciones *tic* y *toc* de Matlab.

Pruebas individuales de los tiempos de ejecución al arranque del sistema operativo de los remuestreos para la detección de multimodalidad.

Remuestreo	Moda	Media	Mediana
500	3.5248	1.6670	2.6631
1000	5.9242	1.4701	3.8675
2000	4.5926	2.5942	2.1635
3000	3.4586	2.1737	2.4256
5000	2.5163	2.6455	3.3667
10000	5.9608	6.5953	6.5315
100000	33.6289	23.9168	17.3499

(a). Grupo de 21 pruebas individuales.

Remuestreo	Moda	Media	Mediana
500	4.5961	2.0656	1.6614
1000	2.1259	1.6771	1.7660
2000	2.6945	1.8323	2.5003
3000	3.2884	3.1943	2.9547
5000	4.3778	2.0744	2.1114
10000	5.4515	5.3201	6.2167
100000	24.6675	23.3810	18.3862

(b). Grupo de 21 pruebas individuales.

Remuestreo	Moda	Media	Mediana
500	5.3876	1.8792	1.5775
1000	3.8777	1.9289	3.5758
2000	3.5442	1.8365	1.9190
3000	5.9411	2.3826	2.1067
5000	4.9504	2.5838	2.4471
10000	5.5389	4.2294	6.5234
100000	24.7674	18.0692	16.9341

(c). Grupo de 21 pruebas individuales.

Remuestreo	Moda	Media	Mediana
500	3.3576	2.0998	3.5757
1000	2.7148	1.7592	2.3549
2000	4.8856	1.9024	2.1725
3000	3.4820	3.0574	2.0209
5000	6.2929	2.5587	3.3810
10000	5.4777	3.8384	5.3948
100000	24.9214	20.0653	22.6787

(d). Grupo de 21 pruebas individuales.

Remuestreo	Moda	Media	Mediana
500	4.5519	1.8912	1.4521
1000	3.7219	2.0509	3.2595
2000	3.7219	1.9399	1.9928
3000	6.2929	2.2003	3.0887
5000	4.4336	6.4829	2.6540
10000	5.5562	4.4840	6.1361
100000	27.4850	23.0322	17.3251

(e). Grupo de 21 pruebas individuales.

Tabla I. 2. Tiempos de cálculo al arranque del S.O. usando tic y toc.

