ANALISIS DE RASCH PARA TODOS

Una guía simplificada para evaluadores educativos

PRIMERA EDICION (Oct.1998)
REVISIÓN 2013

Instituto de Evaluación e Ingeniería Avanzada

Agustín Tristán López

Doctor en Ingeniería
Profesor Investigador del Instituto Tecnológico de San Luis Potosí, México
Consultor Tecnológico, CONACYT
Consultor del CENEVAL, México (1998)
Director del Instituto de Evaluación e Ingeniería Avanzada, S.C.

Instituto de Evaluación e Ingeniería Avanzada
Cordillera Occidental 635
Col. Lomas 4ª sección
78216 San Luis Potosí, MEXICO

Análisis de Rasch Para Todos

TABLA DE MATERIAS

1.INTRODUCCION

1.1 Propósito de este libro

Me he encontrado con numerosos amigos profesores o profesionales que se dedican a la evaluación (pero cuya especialidad no es propiamente la evaluación educativa), que me preguntan si tengo disponible un libro de divulgación como complemento a mis cursos de Evaluación y de Análisis de Rasch. Su motivación es generalmente la misma: asistieron a una reunión donde un especialista de la evaluación menciona el análisis de Rasch y sus ventajas en contraposición con el modelo clásico. No solamente hace la mención de Rasch, sino que habla del "logit" , del "infit" y del "outfit" (sin traducir para que resulte más impresionante), hace que toda la audiencia se quede admirada de sus conocimientos, que los asistentes no sepan qué decir, cómo interpretar sus sabias palabras, cómo manejar los maravillosos resultados; en suma, hace que los asistentes se sientan infelices y sin ningún sustento para el tema de la evaluación.

Este no es un libro de matemáticas, de estadística o de evaluación, aunque emplea elementos de estas áreas del conocimiento. Está pensado para que cualquier profesor o evaluador, interesado en el análisis de Rasch y de la Teoría de la Respuesta al Reactivo, pueda emplear estas técnicas, interpretarlas y obtener el máximo provecho en su trabajo cotidiano.

Este libro pretende que los evaluadores que se inician en el tema puedan tener un lenguaje básico de estas teorías "no clásicas", en particular del análisis de Rasch, junto con una comprensión intuitiva y gráfica que le apoyará en el estudio de los datos provenientes de exámenes objetivos. Por ello, es muy probable que algunos evaluadores experimentados y de alto nivel consideren que el lenguaje y la forma de presentar los conceptos son muy poco ortodoxos, inclusive comentarán que lo que se presenta está plagado de errores. Bueno, en realidad el libro no está pensado para este tipo de evaluadores.

El propósito es que la gran mayoría de profesionales de cualquier rama dedicados a la evaluación, así como los evaluadores noveles, conozcan los fundamentos de esta técnica, que puedan comprender los principios básicos y el lenguaje, de tal modo que durante los congresos, coloquios o reuniones de trabajo no sean engañados cuando alguien diga una mentira en este campo, ni se sientan apabullados cuando un "evaluador experimentado" hable sobre el análisis de Rasch, imponiendo un respeto mal entendido entre la concurrencia.

No hay que olvidar que la evaluación es una especialidad. Si bien es importante que los profesores o profesionales dedicados a la evaluación comprendan los fundamentos del método, no se pretende hacer especialistas a todos. El trabajo de los especialistas espero que sea dejado a cargo de los especialistas. Pero sí es necesario disponer de un lenguaje común, conocer las implicaciones que tiene esta técnica y su ámbito de aplicación.

Se tiene que hacer algo de tratamiento matemático, pero he tratado que sea lo mínimo posible. En anexos se tienen justificaciones y demostraciones, a las cuales podrá acudir el lector interesado. Se incluyen numerosos ejemplos y todo lector que disponga de una hoja de cálculo (por ejemplo EXCEL) podrá realizar directamente algunos ejercicios que le apoyarán en la comprensión de los conceptos presentados.

El orden de los temas no sigue el enfoque tradicional que se tiene en los libros de Análisis de Rasch, de Teoría de la Respuesta al Reactivo o Teoría del Rasgo Latente. Se presentan los temas de acuerdo con la secuencia seguida en mi curso de análisis de Rasch y espero que permita al lector comprender ordenadamente las ideas de base de esta interesante técnica.

Es claro que en un libro de estas dimensiones no es posible presentar los variados aspectos relacionados con el Análisis de Rasch, por lo que el lector queda invitado para profundizar posteriormente en libros más serios aunque, probablemente, menos didácticos.

Por último, deseo expresar que ninguna técnica relacionada con la evaluación merece que el evaluador se vuelva maniqueo. El análisis de Rasch tiene aciertos, pero también puntos débiles. La Teoría de la Respuesta al Item tiene otras ventajas y varios puntos obscuros y dudosos... del mismo modo que los tiene la Teoría Clásica. Este libro presenta, por lo tanto, los elementos de estas técnicas sin apasionamiento y de manera objetiva, para que el mismo lector tome sus propias decisiones y pueda utilizar lo que tienen de bueno, así como dejar de lado lo que considere menos razonable.

Ojalá que este libro permita que haya una mejor difusión del análisis de Rasch en nuestro idioma y en su justa medida.

1.2 Breve definición del Análisis de Rasch.

Cuando se habla de Análisis de Rasch no se hace referencia a "hojuelas crujientes" o "polvos para rascarse", como me hicieron creer la primera vez que me hablaron del tema. Se hace referencia a Georg Rasch, matemático danés, quien desarrolló su modelo alrededor de los años cincuenta del presente siglo. A pesar de que se trata de un modelo maduro, ha pasado por diversos estados de aceptación y rechazo, por grupos que lo idolatran y otros que lo repudian, además de pasar por momentos de ignorancia e incomprensión.

Podemos dar una primera definición:

El Análisis de Rasch es un modelo que establece la probabilidad de respuesta de una persona ante un estímulo dado, en términos de la diferencia entre la medida del rasgo de una persona y la medida del estímulo utilizado.

Se trata de un modelo estocástico (no determinista) donde la medida del rasgo de la persona y la medida del estímulo aplicado, quedan ubicadas en una misma escala lineal con un origen común. La variable de interés es la diferencia de ambas medidas.

El modelo de Rasch establece que la medida del rasgo de la persona es independiente del conjunto de estímulos aplicados. Asimismo se dice que la medida de cada estímulo es independiente del conjunto de personas a las que se somete. En rigor es la diferencia de medida de rasgo y medida de estímulo la que es independiente del instrumento o de la población.

Por último, el modelo de Rasch requiere que la variable sea unidimensional, ordenada e inclusiva.

Este libro presenta los fundamentos necesarios para comprender los elementos que integran a esta definición, según se explica más adelante.

1.3 Nomenclatura

Para los fines que perseguimos, usaremos como sinónimas las palabras estímulo, reactivo, ítem o pregunta. A las personas no las medimos directamente, en cambio lo que medimos es algún rasgo, habilidad, capacidad o aspecto que pueda presentar la persona (lo que se denomina “Rasgo Latente”) y que es explorado por medio de una muestra de reactivos tomados del dominio en estudio.

En este libro no hacemos distinción entre sujeto, persona, alumno, sustentante, individuo. El lector podrá ajustar esta nomenclatura a sus propios fines

2. EL LOGITO. UNA NUEVA UNIDAD DE MEDIDA

2.1 La escala de medida

Cuando se habla de evaluación de conocimientos es común emplear cuestionarios que se aplican a los estudiantes para conocer el número de aciertos al conjunto de preguntas, reactivos o ítemes utilizados. Se acostumbra reportar el "grado de dominio" de cada estudiante en términos de "número de aciertos" o de "porcentajes de aciertos". Independientemente de la conveniencia o no de usar estos indicadores, se afirma que la escala que se genera tiene varios defectos:

a) Cuando se dice que un estudiante obtuvo 0 aciertos, ¿quiere decirse que tiene cero conocimientos ? ¿es un ignorante total?
b) Cuando un estudiante obtiene todos los aciertos ¿se puede afirmar que conoce TODO? ¿Qué es "TODO"? ¿Está listo para el Premio Nobel?
c) Un estudiante obtiene 80 aciertos y otro obtiene 40 aciertos, ¿Puede afirmarse que un estudiante "sabe" el doble del otro ?

A continuación discutiremos estos aspectos que son aclarados con el concepto de la escala de medida.

2.1.1 La extensión

Veamos primero el aspecto de la extensión de la escala. Es evidente que la escala hace referencia exclusivamente al rasgo evaluado, por ejemplo: conocimientos de Historia del 2o grado de Enseñanza Secundaria, habilidad para interpretar datos en el 4o grado de Primaria, capacidad analítica en la solución de problemas del 3er curso de Matemáticas Universitarias, etc. Un alumno que obtiene 100 aciertos en Matemáticas no necesariamente obtiene 100 aciertos en Historia. La escala hace referencia, exclusivamente, al rasgo medido, que en la “jerga” de la evaluación se le denomina “rasgo latente”. Se dice que no es posible determinar con precisión el grado de dominio en Historia, pero sí se puede explorar (por medio de algún muestreo) el conjunto de conocimientos de Historia a través de preguntas, que el evaluador supone que solamente tratan de Historia, a un nivel estipulado por medio de una Tabla de Validez de Contenido.

Una pregunta de Historia mide (o cuando menos trata de medir) un cierto conocimiento de la materia, pero la pregunta tiene influencia de la forma de redactar que tenga el evaluador, de la presentación en el cuestionario, del entorno del alumno, de la forma de aplicar el examen, etc. Por ello se dice que se está tratando de medir el “rasgo latente en el estudiante” que también es un “rasgo latente en el reactivo”. La medida de este rasgo tiene error y es de interés emitir un dictamen de la medida y del error de la medición para que la evaluación que se hace al estudiante sea lo más justa y precisa posible.

Por lo anterior, cuando se habla de 0 aciertos o de 100% de aciertos se tiene un desconocimiento preciso del rasgo latente, aunque sí se puede decir que, bajo ciertas condiciones y premisas, se trata de una aproximación o estimación de la medida del dominio del sujeto en el rasgo. La medida debe, en consecuencia, referirse, señalarse o ubicarse en una escala para poder comparar el dominio de un sujeto contra otros, o el dominio de un sujeto respecto a un criterio de aceptación o de competencia previamente establecido. En este momento estramos, por lo tanto, en el problema de definir una escala.

Es importante que la escala sea tan extensa que permita medir a todos los sujetos, pero no tanto que resulte extremadamente amplia y, por lo tanto, se vuelva inútil. Por ejemplo, los límites naturales de la escala Celsius de temperaturas se definen en 0 y 100 ^{o}C. Un termómetro en esta escala es conveniente para medir los puntos de fusión del hielo y de ebullición del agua, pero se vuelve muy poco útil si se desea emplear para medir la temperatura corporal de una persona, ya que en este caso solamente se requiere un intervalo menor, entre 35 y 42 ^{o}C , como el de los termómetros comerciales que pueden encontrarse en cualquier clínica médica.

Junto con la extensión, que define el rango de medida, se tiene que pensar en la precisión. Es conveniente que la escala tenga sus divisiones igualmente espaciadas, con objeto de facilitar la medición y conseguir una mejor precisión en la medida. En evaluación de conocimientos cada “división” de la escala correspondería con un reactivo, dosificado en términos de la dificultad. La idea es cubrir toda la extensión de la escala por medio de un número suficiente de reactivos en toda la gama de dificultades.

Supóngase que se va a medir la estatura de un conjunto de personas y para ello se traza una marca en un muro a una altura aproximada de 1.5 m. ¿Qué puede decirse de esta “escala”? Resulta evidente que no tiene extensión. Cuando se haga la medición sólo podrá identificarse que una persona está por arriba, por abajo o sobre la marca, pero su medida es indefinida.

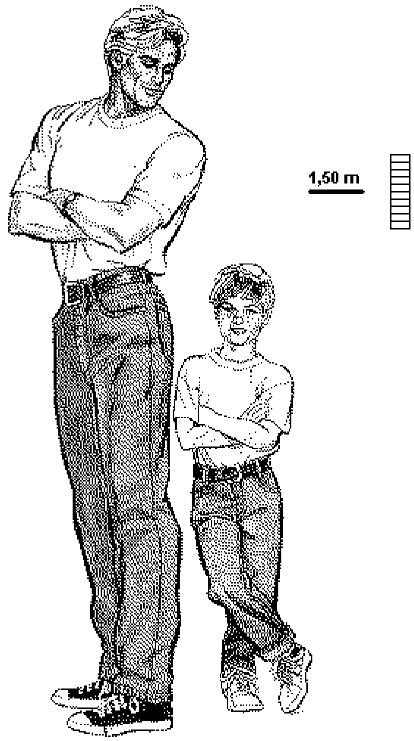


Fig.2.1 Escalas sin extensión para medir a todos los sujetos

Para mejorar la medida se trazan cinco marcas por arriba y por abajo de la inicial, separadas 1 cm cada una. Ahora se tiene una escala de 10 cm en total, centrada alrededor de la marca de 1.5 m. ¿Puede decirse que se ha mejorado la escala? En cierta forma sí, porque ya se dispone de mayor número de marcas, pero sigue careciendo de extensión suficiente y no será factible medir a las personas por arriba de 1.55 ni abajo de 1.45 m.

Obsérvese que una escala mal diseñada sólo permitirá identificar que un sujeto está por arriba de la marca establecida, o por abajo, pero no faculta al evaluador para estimar qué tan arriba o qué tan abajo de la marca se encuentra el sujeto.

Si se traduce el caso anterior a lo que ocurre con un examen, se debería seguir exactamente la misma lógica. Es conveniente disponer de reactivos en toda la gama de dificultades y no solamente reactivos centrados al 50% de dificultad, con objeto de poder medir el dominio de cada persona con precisión. Con una escala completa sí se puede ubicar a las personas y estimar qué tanto sabe o qué tanto no sabe.

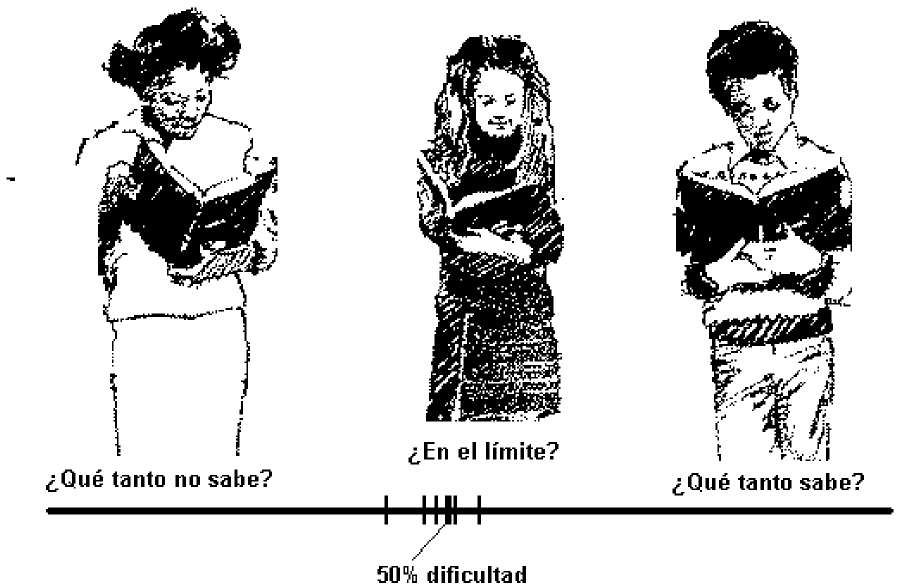


Fig.2.2 Una escala de conocimientos sin extensión

El tema de la escala se plantea independientemente de que se trate de un cuestionario referido a norma o referido a criterio. Dependiendo del propósito de la evaluación, se puede plantear un cuestionario referido a norma (donde la población funciona como patrón de referencia) o referido a criterio (donde se tiene que demostrar un grado de dominio y el patrón es fijado de antemano por juicio de expertos en la materia, sin influencia de los resultados de la población). Se dice que en la evaluación referida a norma se requiere una gran varianza en los resultados de los sujetos, mientras que en la evaluación referida a criterio la varianza no tiene por qué ser amplia. El tamaño de la varianza está relacionada con los puntajes obtenidos por las personas, lo cual es, en principio, independiente de la extensión de la escala, aunque muchos evaluadores suelen confundir una cosa con la otra.

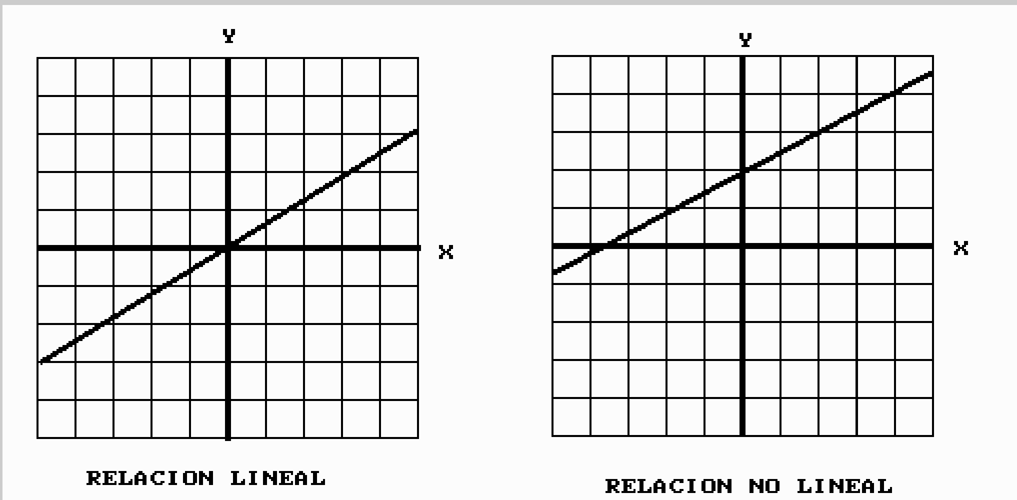
2.1.2 La medición

Ahora analicemos la forma de "medir" en la escala. En la escala reportamos, por ejemplo, el número de aciertos desde 0 aciertos hasta 126 en un examen de 126 reactivos. Es evidente que podemos intuir que un alumno que obtiene 80 aciertos es probablemente mejor que otro que obtuvo 40 aciertos, pero ¿80 es el doble de conocimientos que 40? La respuestas es no. En esta escala no puede garantizarse que 4 unidades representen el doble del rasgo que 2 unidades. El problema es que las divisiones de la escala no están ubicadas a distancias constantes, ni las cantidades son directamente proporcionales y calculables con una simple regla de tres. En este sentido no puede afirmarse que 80 sea el doble de conocimientos que 40. Se dice por ello que la escala no es lineal.

Una relación funcional es "lineal" cuando se cumple la proporcionalidad directa entre la variable independiente y la dependiente. Se escribe en la forma simbólica:

$$y = kx$$

donde x es la variable independiente, y es la variable dependiente, k es el parámetro de proporcionalidad.



La expresión y=kx+c NO ES LINEAL, aunque corresponda con la ecuación de una recta. En esta expresión el doble de x no conduce al doble de y.

Las funcionesdel tipo y=sen(x) , y=log(x), y=exp(x), y=x^2, etc. son claramente no lineales.

Hay varios ejemplos muy claros de la vida cotidiana que muestran una relación no lineal. Por ejemplo, en la carrera de velocidad de los 100 metros planos, se tiene que desarrollar un cierto esfuerzo para recorrer la distancia en 20 segundos, pudiera ser que con un doble de esfuerzo se pueda hacer la carrera en 10 segundos. Pero, en general, no puede afirmarse que el desempeño es proporcional al esfuerzo: basta con imaginar el terrible esfuerzo que se requiere para pasar de 10 segundos a 9 y sin duda no hay ser humano que pueda batir la marca lléndose a 8.1 segundos con un "pequeño esfuerzo" de un 10%... lo más probable es que el esfuerzo sea infinito y sin posibilidades de ser efectuado por un atleta, cuando menos en las condiciones actuales.

A partir de ahora el lector deberá tomar en cuenta que una medida de los conocimientos de una persona no corresponde con un valor preciso y determinado de los conocimientos que atesora en su cerebro (cosa que no podemos medir), sino con una "probabilidad de respuesta". Un alumno con una "medida alta" tendrá una mayor probabilidad de respuesta correcta a un reactivo. También conviene en este momento anotar que la probabilidad de respuesta la vamos a calcular y a estimar. Hablaremos de "cálculo" cuando la determinemos por métodos analíticos correspondientes a un modelo matemático. Hablaremos de "estimar" cuando hagamos una operación muy sencilla que consiste en obtener la frecuencia relativa de respuestas en un examen.

Y al entrar al terreno de la probabilidad debemos entonces prepararnos para saber que las puntuaciones de una persona, así como las respuestas a un reactivo, no son valores deterministas ni fijos. Se trata de valores estocásticos y, como tales, pueden ocurrir con mayor o menor probabilidad. Los modelos que vamos a ver en este libro son modelos estocásticos (modelos probabilistas) y no se puede garantizar nunca que una persona tiene una medida X, sino se habla de que una persona de medida X tiene una probabilidad p de respuesta correcta y otra probabilidad q de respuesta incorrecta.

La medida X es una estimación de su medida de dominio real T. Se trata de una aproximación que contiene un error de medida. Si el instrumento utilizado para la medición es altamente preciso, la medida X y el valor real T serán prácticamente coincidentes. En muchas ramas de la ciencia y la técnica se han desarrollado instrumentos de medida de altísima precisión, desgraciadamente en el terreno educativo ningún instrumento es perfecto. Es común que nuestros exámenes contengan un error importante, por lo que el estimado X contendrá igualmente un error y podría no parecerse al valor real T. De ahí la gran responsabilidad del evaluador de conocer las características del instrumento, su calidad, su error, su confiabilidad, para poder emitir las mejores estimaciones posibles de la medida real de los sujetos, así como la medida real del instrumento. El evaluador debe estar consciente de estas limitaciones y reconocer, asimismo, que su instrumento no es perfecto y deberá tomar las medidas correctivas que sean necesarias para no perjudicar o beneficiar injustamente a los sujetos, informándoles de manera errónea la posición que ocupan en la escala. Por ello la estimación de la probabilidad de respuesta es muy importante.

Los “cálculos” de las probabilidades son objeto de otro capítulo. En éste revisemos la “estimación” de la probabilidad. Por ejemplo, si a una persona se le administran 20 preguntas y contesta 15 preguntas, podemos estimar que la probabilidad de respuesta correcta es:

p = 15/20 = 0.75

La fórmula es muy simple y se basa en el cálculo de las frecuencias relativas como una aproximación de la probabilidad:

p = Respuestas correctas/Total de reactivos (2.1)

El valor p es estimado porque depende de los reactivos aplicados. Este estimador sólo toma en cuenta el éxito del alumno.

Es muy importante mencionar que, solamente bajo ciertas condiciones, p es el estimador más probable del nivel de dominio de una persona (Ver anexo 3). En principio se requiere que se haga un muestreo al azar y sin reemplazo de los contenidos u objetivos educacionales con los que se va a estimar el dominio del alumo. El evaluador no debe tomar a la ligera estas condiciones, ya que puede hacer estimaciones falsas a partir de un instrumento defectuoso, conteniendo un gran error de medida.

El fracaso o las fallas que puede tener el alumno se denomina q y es el complemento de p a la unidad:

q = 1-p (2.2)

Para este ejemplo q = 1-0.75 = 0.25, lo cual quiere decir que la persona tiene un 0.25 de probabilidad de contestar erróneamente el examen, o que podrá fallar 5 preguntas (Fallas=q x número de reactivos = 0.25 x 20 = 5).

Una duda que se presenta con mucha frecuencia se relaciona con el carácter estocástico de la estimación, cuando se trata de inferir el resultado de cada experimento.

El caso más claro es el de los lanzamientos de una moneda. Si la moneda no está "cargada", se espera que la probabilidad de "cara" sea igual a la de "cruz" y ambas valen, por lo tanto, 0.5. Cuando se hace el experimento de lanzamientos de una moneda es claro que a mayor número de lanzamientos la frecuencia relativa de caras y cruces tiende a 0.5 en una moneda no "cargada". Pero aunque se sepa que la probabilidad de obtener cara es de 0.5, no es posible predecir cuál va a ser el resultado en el siguiente lanzamiento de la moneda.

Bajo este mismo esquema, puede estimarse con relativa facilidad la probabilidad de respuesta de un sujeto a un conjunto de reactivos, pero es imposible predecir cuál va ser la respuesta que va a proporcionar al siguiente reactivo que se le administre. Muchos evaluadores están esperando la "receta mágica" que permita predecir si el sujeto debe o no contestar al reactivo que se le va a proporcionar, tal receta no puede existir. Sería igual que querer disponer de un método para saber si el resultado del siguiente lanzamiento de la moneda será "cara".

En función del nivel de dominio de la persona y de la dificultad del reactivo, sólo se puede disponer de una probabilidad de respuesta. Esto se estudia más adelante.

2.2 El lógito y la medida de una persona

Si se combinan el éxito p y el fracaso q se mejora la manera de expresar la estimación de la "Medida". Para ello usaremos el "momio". El momio o apuesta es una manera de expresar la expectativa que se tiene de éxito con relación al fracaso. El momio o apuesta es la cantidad que se acostumbra emplear en los encuentros de box, las carreras de caballos y otros eventos semejantes, donde se cruzan apuestas con puntos a favor y puntos en contra de que un boxeador (o un caballo) pueda ganar.

Con objeto de mejorar el concepto de la escala, se propone el uso de una nueva unidad de medida lineal (en este caso 4 unidades serían el doble de 2 unidades) y que facilite cubrir todo el rasgo medido. Esta unidad es el logaritmo natural del momio, que se denominará "lógito". Esta medida tiene la ventaja de que limita los valores de medida a intervalos razonables y permite tomar en cuenta tanto el éxito como el fracaso en una sola cantidad.

Lógito es una traducción libre de la unidad de medida definida en inglés "logit". La palabra "logit" es una forma abreviada de "log odd ratio" que se traduce por "logaritmo del momio".

Vamos a explorar esta nueva unidad de medida y a hacer algunos ejercicios sencillos. Supóngase que hay una materia M formada por 100 objetivos específicos y que un alumno tiene un dominio de un 60% en los conocimientos de dicha materia. Si se le administra un examen de 100 reactivos (un reactivo por cada objetivo), esperaríamos que el alumno contestara correctamente a 60 de las preguntas y no pudiera contestar 40 de ellas. Esto puede interpretarse como que el alumno tiene una probabilidad 0.6 de respuestas correctas y una probabilidad 0.4 de fallas.

Recuérdese que el modelo permite hacer un buen estimador del dominio del estudiante siempre y cuando el muestreo sea efectuado al azar y sin reemplazo (Ver anexo 3). En este momento de la medición no se consideran errores ocasionados por respuestas inesperadas de este tipo:

a) el alumno adivina o acierta <u>por azar</u> una pregunta de un tema que no conoce

b) el alumno contesta erróneamente un reactivo dentro del 60% de aspectos que conoce, por haber interpretado mal la pregunta o porque el evaluador la hizo confusa

Pero dejemos de lado estos aspectos (importantes pero más complicados) para regresar al caso del alumno y calculemos su medida en lógitos. La operación es muy sencilla. La medida del alumno se denominará B, su probabilidad de respuestas correctas es p=0.6 y de fallas es q=0.4.

Primeramente calculemos su momio (aciertos dividido entre errores):

momio(p)=p/q (2.3)

momio(p)= 0.6/0.4 = 1.5

Podemos decir que si corremos las apuestas para ver cómo va a salir este alumno en el examen, las apuestas estarían de 1.5 a 1 (también puede decirse que la apuesta sería de 0.6 a 0.4, aunque no es muy cómoda esta forma de expresarse), lo cual quiere decir que le damos un 50% más de posibilidades de salir bien que de salir mal.

El logaritmo natural del momio es la medida:

B(p) = ln(p/q) (2.4)

En esta notación B(p) indica "Medida ante una probabilidad p de respuesta", también se puede leer "B de p", simplemente "B" o llamarle "Medida". La letra p entre paréntesis nos recuerda a qué valor hace referencia la medida (probabilidad de aciertos).

Regresando a nuestro alumno, su medida en lógitos es:

B(0.4) = ln(p/q) = ln(1.5) = 0.4055 lógitos

Antes de interpretar esta medida, calculemos algunos otros casos más. Véase que basta con disponer de los valores p y q para calcular la medida en lógitos.

Ejercicio 2.1 ¿Cuál es la medida de un alumno que tiene un 50% de dominio en una materia?

<u>Solución</u>

De acuerdo con el dato: p= 0.5 es la cantidad de respuestas correctas esperadas en el examen. El complemento es q=0.5 , que indica la cantidad de respuestas que se espera sean incorrectas en su examen.

El momio es:

momio(0.5) = 0.5/0.5 = 1

La medida es:

B(0.5) = ln(0.5/0.5) = ln(1) = 0

Un alumno que está en el 50% de dominio tiene una medida de 0.

Una medida de 0 simplemente indica que su posición está centrada en la escala, igual que lo está una moneda no "cargada".

Ejercicio 2.2 Obtener la medida de una persona de 30% de dominio.

<u>*Solución*</u>

De manera similar a la anterior:

p=0.3 , q=0.7

El momio es:

momio=0.3/0.7

B(0.3) = ln(0.3/0.7) = -.8473

Ya vimos que un alumno que está en el 50% de dominio tiene una medida que corresponde con el 0 de la escala. El primer alumno tiene un dominio superior al 50% por lo que su medida es mayor que cero (0.4055 lógitos). El tercer estudiante al tener un dominio inferior al 50% tiene una medida menor que cero (-0.8473 lógitos).

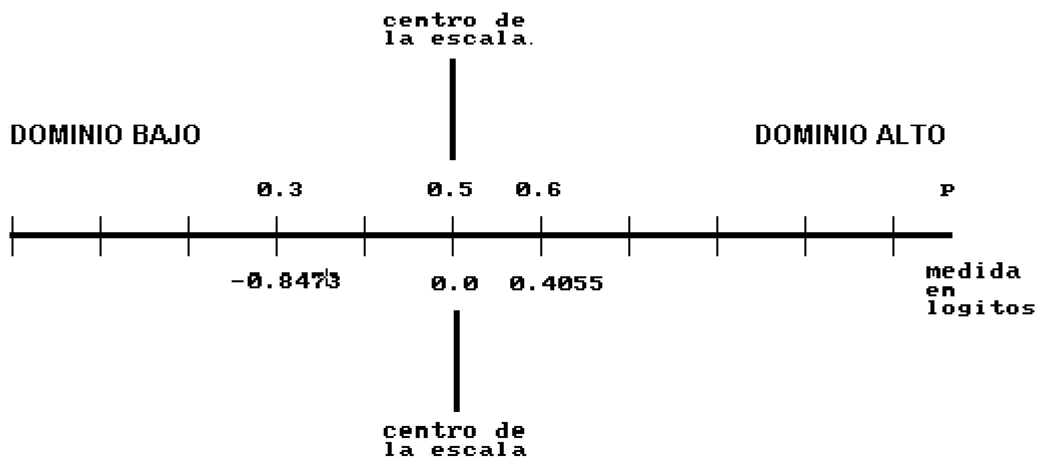


fig.2.3 La escala de valores estimados de probabilidad y la medida en lógitos

La Medida de una persona está relacionada, pues, con la posibilidad de que conteste correctamente a un conjunto dado de reactivos en un cuestionario. Por ello está asociada a la probabllidad p (aciertos).

Queda pendiente, por el momento, buscar un método que permita estimar el dominio de la persona. Entretanto, puede darse una primera interpretación de la unidad en lógitos: la medida en lógitos hace referencia a la probabilidad de respuesta correcta en un examen dado. Cuando un estudiante obtiene una medida de 0, se dice que tiene una probabilidad 0.5 de responder correctamente el cuestionario. Cuando un estudiante tiene una medida de 0.4055 lógitos, tiene una probabilidad de responder correctamente de 0.6.

¿Es posible, entonces, obtener la probabilidad dada una medida en lógitos? Desde luego. Cuando se tiene una medida en lógitos puede obtenerse la probabilidad de respuesta resolviendo la fórmula dada en la definición de la medida.

Recordemos la fórmula de la medida:

$B(p) = \ln(p/q)$ (2.4)

Haciendo algo de álgebra se puede resolver para p:

$p(B) = e^B/(1+e^B)$ (2.5)

Donde e es la base de los logaritmos naturales (e=2.7182818...).

Ahora se lee “probabilidad para una medida B”, o también “p de B” o simplemente “p”. El argumento entre paréntesis B nos recuerda que la probabilidad ha sido calculada para una medida dada B.

Para fines de programación y uso en hojas de cálculo, se puede obtener el valor de e^B empleando la función exponencial exp(B).

Ejercicio 2.3 ¿Cuál es la probabilidad p de respuesta de una persona que tiene una medida B de 1.1 lógitos?

Solución

Usaremos la fórmula 2.5:

$p(1.1) = e^{1.1}/(1+e^{1.1})$

Con ayuda de una calculadora de escritorio o la computadora se puede verificar que $e^{1.1} = \exp(1.1) = 3.004...$

$p(1.1) = 3.004/(1+3.004) = 0.75$

Comprobemos este resultado haciendo el cálculo en forma inversa:

si $p=0.75$, $q=0.25$

$B(0.75) = \ln(p/q) = \ln(0.75/0.25) = 1.0986 = 1.1$

En todos estos cálculos el lector siempre encontrará algunas imprecisiones numéricas que dependen del número de dígitos que conserve en la calculadora. Para fines prácticos 1.0986 es igual a 1.1.

En algunas calculadoras no se tiene directamente la función “exp”, sino que se debe solicitar INV(LN X). Si su calculadora no tiene LN X le sugiero trate de comprar otra.

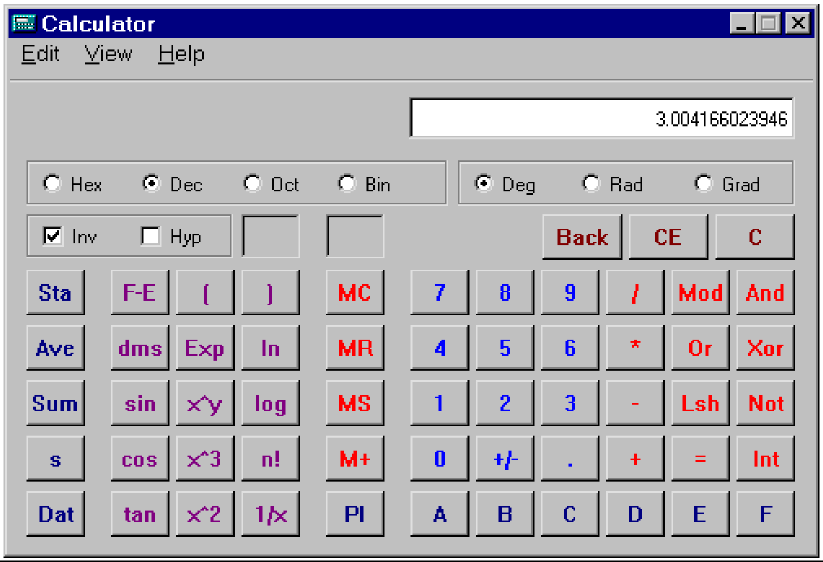
Si usa la calculadora que proporciona Windows, deberá seguir estos pasos:

a) Introduzca el número deseado, por ejemplo 1.1

b) marque el casillero Inv

c) pida la función ln

d) El resultado queda indicado en la carátula de la calculadora, en este caso "3.004166023946"



Ejercicio 2.4 ¿Cuál es el rango de valores que toma la medida en lógitos?

<u>*Solución*</u>

A partir de la fórmula 2.4, se tiene que los límites van a ser fijados por los valores extremos de p, que varía entre 0 y 1:

si p=0, q=1 , se tiene: B=ln(0/1) = - ∞

si p=1, q=0 , se tiene: B=ln(1/0) = + ∞

Observación: Aunque el intervalo cubre todos los números reales, desde menos infinito hasta más infinito, se pude decir que para fines prácticos la medida varía entre -5 y +5 lógitos, muy exageradamente se tendría el rango de -8 a +8 lógitos.

Como ya se anotó previamente, está pendiente el saber cómo garantizar que p (valor que indica la probabilidad de respuesta correcta a un examen) corresponde con la medida del dominio de una persona, dada en lógitos. Esto lo estudiaremos más adelante.

2.3 El lógito y la medida de un reactivo

Hasta aquí hemos trabajado la escala en lógitos para las personas. La misma medición se puede plantear para los reactivos, por lo que los reactivos también se miden en lógitos.

En diversos aspectos de la vida utilizamos balanzas de precisión, termómetros, manómetros u otros aparatos especiales para medir un objeto o un fenómeno dado. El instrumento de medida que empleamos en evaluación del aprendizaje es el reactivo o el examen. A la medida de los reactivos se le denomina "calibración". Así se dice que se está "calibrando" un reactivo y se está "midiendo" a las personas. Esta nomenclatura es convencional y se justifica en el sentido de que el instrumento que sirve para medir, por ejemplo una balanza de precisión, se tiene que "calibrar" en un laboratorio especializado; una vez calibrado sirve para "medir" el peso de las personas.

Supóngase que en un examen se incluyen tres reactivos. Uno es de 40% de dificultad, otro de 50% y otro más de 70% de dificultad. Aquí debe observarse que conforme se hace más grande el valor de la dificultad se dice que el reactivo es más difícil: un reactivo de 0 de dificultad quiere decir que no es difícil (su dificultad es nula), mientras que otro del 100% de dificultad es totalmente difícil o imposible de ser respondido. La dificultad para esta metodología es el inverso del "Grado de Dificultad" clásico, que se debería de llamar más propiamente "Grado de Facilidad". En lo sucesivo se hablará de "Dificultad", que indica qué tan difícil es un reactivo y se dejará "Grado de Dificultad" para el concepto clásico que indica qué tan fácil es un reactivo.

Recuérdese que el Grado de Dificultad se define como el cociente de las respuestas correctas entre el número de personas que contestan un reactivo. Mientras más personas contestan el reactivo se hace más fácil y el Grado de Dificultad aumenta; si pocas personas contestan se dice que el reactivo es difícil, sin embargo el Grado de Dificultad disminuye... por ello el Grado de Dificultad clásico es un índice inverso.

Calculemos ahora la "calibración" de los reactivos. La calibración D se define por la expresión:

$$D(q) = \ln(q/p) \qquad (2.6)$$

A partir de esta definición se puede ver que el momio empleado para la calibración del reactivo es el cociente de fallas entre aciertos.

$$\text{momio}(D) = q/p$$

En la calibración el momio o apuesta se refiere a las fallas, en la medición de las personas se "apostaba" al éxito o acierto.

Ejercicio 2.5 Determinar la calibración de los reactivos que tienen 40%, 50% y 70% de dificultad respectivamente.

Solución

Para el primer reactivo, q=0.4, p=0.6 y su calibración es:

D(0.4) = ln(0.4/0.6) = ln(0.6666) = -.4055

Para el segundo reactivo, p=0.5, q=0.5 y su calibración es:

D(0.5) = ln(0.5/0.5) = ln(1)= 0

Para el tercer reactivo, q=0.7, p=0.3 y su calibración es:

D(0.7) = ln(0.7/0.3) = ln(2.3333) = 0.8473

Observamos que un reactivo que se encuentra en el 50% de dificultad tiene una calibración de 0 lógitos. Un reactivo difícil tendrá una calibración mayor que cero, mientras que un reactivo fácil tendrá una calibración menor que cero. Mientras más difícil sea el reactivo más "positivo" y grande será, en cambio los reactivos más fáciles tendrán valores más chicos y "negativos".

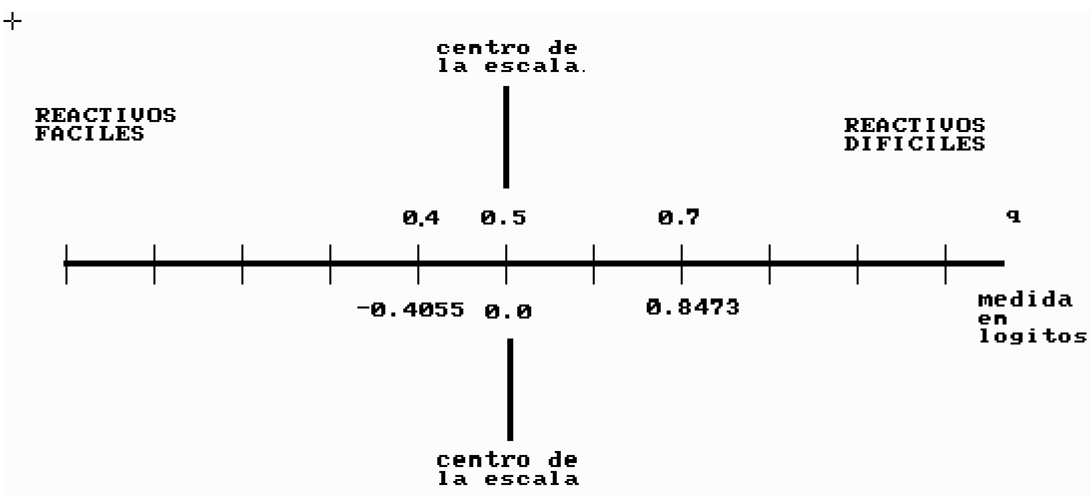


Fig.2.4 Dificultad y escala en lógitos

La Dificultad, como su nombre lo indica, tiene que ver con la posibilidad de que los alumnos fallen o no respondan correctamente, por ello la medida D es función de q (falla) y no de p. Si se hubiera planteado D en función de p, entonces hablaríamos de la "facilidad" del reactivo y tendríamos un valor equivalente al Grado de Dificultad clásico. Recordemos, entonces, que dificultad corresponde con probabilidad de fallar...las fallas las indica el parámetro q.

Del mismo modo que se resolvió la ecuación para la medida de la persona, se puede tener la expresión explícita para la probabilidad de respuesta dada una calibración. Así se tiene:

$$p(D)= e^{-D}/(1+e^{-D}) \qquad (2.7)$$

que se parece a la anteriormente empleada para B, pero se observa que ahora los exponentes del número e son negativos.

Esta última expresión puede escribirse en una forma equivalente aprovechando las propiedades algebráicas de la función exponencial:

$$p(D) = 1/(e^{D}+1) \qquad (2.8)$$

Ejercicio 2.6 Si se sabe que un reactivo tiene una calibración de -0.3 lógitos, ¿ cuál es su probabilidad de respuesta ?

Solución

$p(D)= 1/(e^{-0.3}+1) = 1/(0.740818+1) = 0.5744$

El reactivo es ligeramente más fácil que el “reactivo medio” al 50% de dificultad, por ello tiene una calibración menor que cero (-0.3) y una probabilidad de respuesta mayor que 0.5 (0.5744)

(El lector podrá verificar fácilmente con ayuda de la computadora o de una calculadora que $e^{-0.3} = 0.740818$)

Notará el lector que, de acuerdo con las propiedades de los logaritmos (ver Anexo 1), se tiene que:

$$\ln(p/q) = - \ln(q/p) \qquad (2.9)$$

Ejercicio 2.7 Verificar que D(0.6)=-D(0.4)

Solución

a) Para D(0.4) se tiene q=0.4, p=0.6

momio: p/q = 0.66666

logaritmo del momio: ln(p/q)= -0.4055

D(0.4) = -0.4055

b) Para D(0.6) se tiene q=0.6, p=0.4

momio: q/p = 1.5

logaritmo del momio: ln(q/p)= 0.4055

D(0.6) = 0.4055

Lo cual verifica la igualdad pedida

Queda por resolver algo que no hemos tratado hasta este momento: ¿Cómo se determina la "medida" de una persona? ¿Cómo se "calibra" un reactivo? En los siguientes capítulos se tratan estos asuntos.

2.4 Ejercicios propuestos

1. Determinar el momio de un sujeto que tiene una probabilidad de falla de 0.72

2. Determinar el momio de un reactivo que es contestado correctamente por el 24% de personas

3. Calcular la medida de un alumno que tiene un 45% de dominio en una materia

4. Obtener la probabilidad p de respuesta de una persona que tiene una medida B de 0.45 lógitos

5. Determinar la calibración de los reactivos que tienen 37%, 62% y 74% de dificultad respectivamente.

6. Obtener la probabilidad de respuesta de los reactivos cuya calibración es -0.16, 0.16 y 2.3, respectivamente.

7. Completar la tabla de medidas en lógitos y probabilidades de respuesta.

p	q	p/q	B(p)=ln(p/q)
0.25			
0.4			
0.6			
0.7			
0.9			

8. Completar la tabla de probabilidades y medidas en lógitos:

B	e^B	$1+e^B$	$p(B)=e^B/(1+e^B)$
-1.5			
-1.0			
1.2			
2.3			

9. Completar las tablas de dificultades y calibraciones que se indican.

D	e^D	e^D+1	$p(D)=1/(e^D+1)$
-1.0			
-0.5			
0.5			
1.2			
2.3			

q	p	q/p	D=ln(q/p)
	0.2		
	0.4		
0.3			
	0.8		
0.05			

10. Indique por lo menos dos características deseables de una escala de medida

11. ¿Por qué es conveniente que la escala tenga gran extensión?

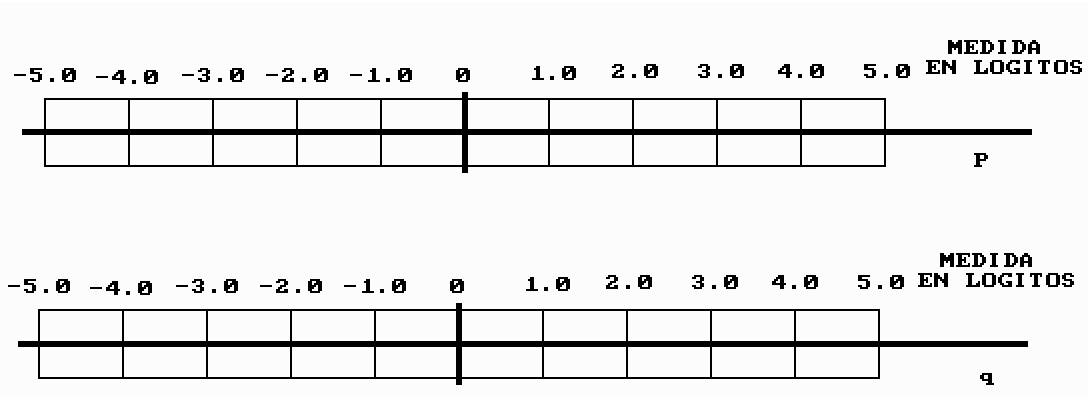
12. En muchos tratados de evaluación se afirma (erróneamente) que los reactivos óptimos se tienen al 50% de dificultad, por ello recomiendan (también erróneamente) que los cuestionarios deben ser diseñados con reactivos de 50% de dificultad ¿Qué invonvenientes se tienen en la escala cuando se aplica un examen formado por reactivos cercanos al 50% de dificultad?

13. Indique la diferencia entre “medir” y “calibrar”, o entre “medida” y “calibración”.

14. ¿Por qué se introduce el concepto de lógito? ¿Qué ventajas reporta?

15. Se tienen dos rectas en el plano, dadas por (a) $y=3x$; (b) $y=3x+1$. Verifique que (a) es lineal, mientras que (b) no es lineal.
Ayuda. Calcule el valor de “y” en los casos $x=5$, $x=10$. Recuerde que una relación es lineal si se conserva la proporcionalidad en “y” cuando cambia “x”.

16. En las figuras 2.3 y 2.4 se presentó una escala con algunos valores de p y q contra su equivalente en lógitos. Complete ahora estas escalas dadas en lógitos para tener su equivalente en valores de p y q.



Verifique que la escala en lógitos es lineal, mientras que las de p y q no lo son.

2.5 Puntos claves del capítulo

1. Escala de medida: Recta numérica lineal que sirve para ubicar y ordenar un rasgo latente. La escala debe ser unidimensional (permitir la medida de un solo rasgo) y disponer de suficiente extensión para poder medir a todos los sujetos o reactivos. En la misma escala se pueden identificar las medidas de las personas y las calibraciones de los reactivos.

2. Medida de una persona: Posición en la escala del rasgo latente de una persona, medido con un conjunto de reactivos.

3. Calibración de un reactivo: Posición en la escala del rasgo latente que mide un reactivo sobre un conjunto de personas.

4. Momio: Relación de aciertos a fallas (para el caso de personas) o de fallas a aciertos (para el caso de reactivos). Cuando se habla del momio de una persona, se indica qué tanto se puede arriesgar al predecir que la persona puede contestar correctamente un reactivo. El momio de un reactivo indica qué tanto se puede arriesgar al predecir que el reactivo puede ser difícil para un conjunto de sujetos.

5. Lógito: Logartimo natural del momio. Sirve como unidad de medida en una escala lineal entre rasgo y posición en la recta numérica.

6. Momio de una persona: $ms = p/q$
donde p es la probabilidad de respuesta correcta ante el conjunto de reactivos y $q=1-p$
(Se señala con ms, donde "s" indica "sujeto")

7. Momio de un reactivo: $mr = q/p$
donde q es la probabilidad de falla del conjunto de sustentantes y $p=1-q$
(Se señala con mr, donde "r" indica "reactivo")

8. Medida de una persona: $B(p) = \ln(ms) = \ln(p/q)$
Se relaciona con el nivel de dominio o posibilidad acierto de una persona.
La medida está centrada en 0. Una medida negativa indica dominio bajo, una medida positiva indica dominio alto.

9. Calibración de un reactivo: $D(q) = \ln(mr) = \ln(q/p)$
Se relaciona con la dificultad o posibilidad de falla en el reactivo.
La Calibración está centrada en 0. Una calibración negativa indica reactivo fácil, una calibración positiva indica reactivo difícil.

10. Probabilidad de respuesta de una persona: $p(B) = e^B/(1+e^B)$

11. Probabilildad de respuesta en un reactivo: $p(D) = 1/(e^D+1)$

3. LA CURVA CARACTERISTICA

3.1 El concepto de curva característica de un reactivo

Cuando se aplica un reactivo a una población de estudiantes siempre ocurre que hay alumnos que responden correctamente y, complementariamente, hay alumnos que fallan. Como ya dijimos anteriormente se supone, para simplificar el problema, que los alumnos responden lo que conocen y que, en principio, nadie responde por azar o tratando de adivinar. Más adelante analizaremos a detalle el asunto de la adivinación.

La hipótesis que estamos haciendo (sólo se responde lo que se conoce) puede ser catalogada de “hipótesis débil”, porque no hay manera de poderla demostrar o justificar. Otras vías de análisis pueden catalogarla como “hipótesis fuerte”, porque permite ser planteada sin ninguna relación con otra parte del modelo y no se puede demostrar de manera recurrente usando otras hipótesis o evidencias posteriores. Para fines prácticos reconozcamos que se trata de una hipótesis (a secas) que permite desarrollar el modelo.

Si aceptamos que los estudiantes responden en función de lo que conocen o dominan, se puede hacer una tabla de las diferentes respuestas que tienen los alumnos a los reactivos del examen. Para fines del ejemplo, supongamos que hay M alumnos y N reactivos aplicados. Se tendría, por lo tanto, una tabla o matriz de M renglones (alumnos) y N columnas (reactivos). Algunos reactivos serán difíciles para los alumnos, otros serán fáciles y algunos reactivos más serán de dificultad media. Unos alumnos obtendrán las mejores calificaciones (mayor número de aciertos), otros tendrán las peores calificaciones (menor número de aciertos) y algunos resultarán con calificaciones intermedias. Podemos construir los diagramas de frecuencias de calificaciones de los alumnos y de calificaciones de los reactivos (figura 3.1).

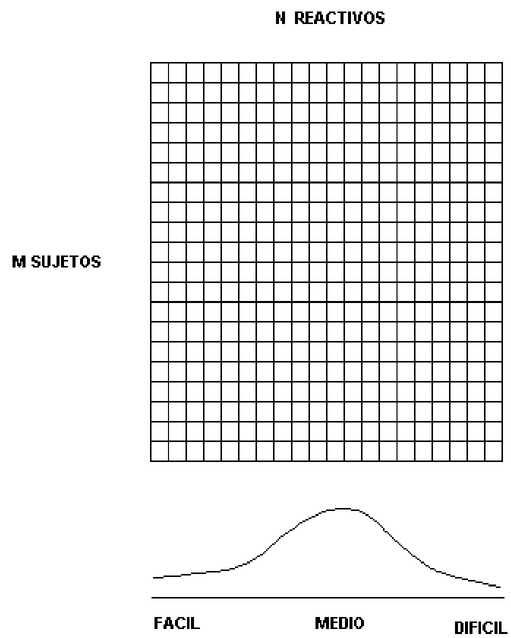


Fig.3.1 Tabla de respuestas de un examen

Cada reactivo se comporta de manera especial con cada sustentante y, a su vez, los sustentantes responderán de cierta manera al conjunto de reactivos. Lo que nos interesa es disponer de información para estimar una probabilidad de respuesta de un alumno ante un reactivo dado, algo así como saber si una moneda está “cargada” y cuál es la probabilidad de obtener una “cara” en el siguiente lanzamiento.

Supongamos que se aplicó el cuestionario descrito en la fig.3.1 a un conjunto de 100 alumnos y elijamos el reactivo número 1. Dividamos al grupo en deciles (subgrupos que contienen al 10% de alumnos) y anotemos el número de respuestas correctas que tiene cada decil en el reactivo 1, con ello podemos hacer un diagrama de frecuencias como el que se muestra en la figura 3.2

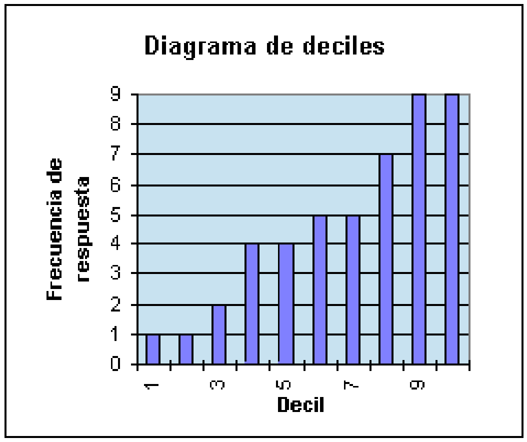


Fig. 3.2 Diagrama de frecuencia de respuestas por decil

Observemos que los extremos superiores de las barritas describen una “curva”. Sería interesante contar con la función que describe a esa curva, porque permitiría hacer estimados de frecuencia de respuestas, por ejemplo, podríamos estimar cual es la posible respuesta de los sujetos del decil 6, sin tener que contarlos en la tabla.

Para dibujar más correctamente la curva evitaremos el uso de barritas, en su lugar trazaremos un punto correspondiente a la medida en lógitos del punto medio del decil y de altura igual a la proporción de respuestas correctas que se tienen en el decil. Esta proporción de respuestas correctas es la frecuencia relativa, que hemos dicho que es un buen estimador de la probabilidad de respuesta. Con ello llegamos a la figura 3.3 (medida vs probabilidad) que es una aproximación a la Curva Característica del Reactivo.

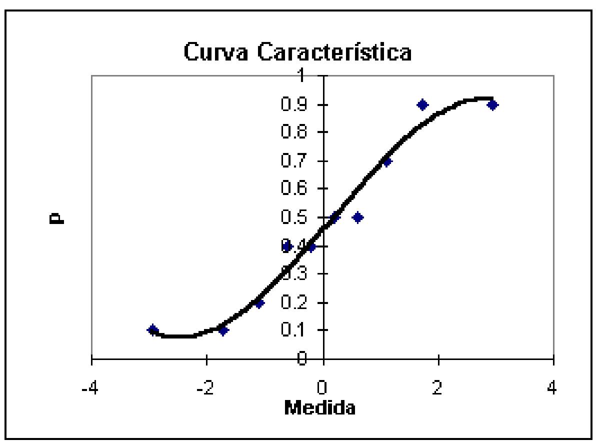


Fig.3.3 Trazo aproximado de la Curva Característica del Reactivo

La curva de la figura 3.3 se obtuvo a partir de los datos de las barritas de la figura 3.2, con la ventaja de que se habla de medidas en lógitos en lugar de barritas o de deciles.

El comportamiento de los reactivos y de las personas puede representarse por medio de una curva denominada Curva Característica del Reactivo (CCR), que relaciona la medida de las personas y la probabilidad que tienen de responder al reactivo. Para construir esta curva necesitamos, por lo tanto, la medida y la probabilidad. El cálculo de estos valores involucra una serie complicada de pasos, pero en este capítulo vamos a hacer una aproximación para entender la idea inherente a la Curva Característica de un Reactivo, por un procedimiento muy simple. El método que se presenta aquí es aproximado, inclusive burdo, pero suficientemente didáctico como para aclarar las ideas.

3.2 Construcción aproximada de una curva característica

Supongamos que se aplica un cuestionario de 5 preguntas a un grupo de 10 personas. Construyamos la tabla 3.1, de acuerdo con lo que se presentó en la figura 3.1, conteniendo las respuestas de cada persona ante cada reactivo.

Tabla 3.1

	REACTIVOS						DOMINIO
Persona	1	2	3	4	5	p	
a	1	1	1	1	1	1.0	ALTO
b	1	1	1	1	1	1.0	
c	1	1	1	0	1	0.8	
d	1	1	1	1	0	0.8	
e	1	1	0	1	0	0.6	MEDIO
f	1	1	1	0	0	0.6	
g	1	0	1	0	0	0.4	
h	1	1	0	0	0	0.4	
i	1	0	0	0	0	0.2	
j	0	0	0	0	1	0.2	BAJO
DIFICULTAD	0.1	0.3	0.4	0.6	0.6		
Clasificación	FACIL		MEDIO		DIFICIL		

La Tabla 3.1 tiene estas características:

a) Los reactivos están ordenados del más fácil al más difícil

b) Las personas están ordenadas de la de mejor resultado (ALTO DOMINIO) a la de peor resultado (BAJO DOMINIO)

c) Se tienen 2 personas por cada resultado: 2 personas que contestaron todos los reactivos, 2 que fallaron en uno, otros dos que fallaron en 2 reactivos, hasta llegar a 2 personas que contestaron bien a un solo reactivo. Esto es un caso muy particular y no debe generalizarse por parte del lector a otros exámenes.

d) La última columna indica el valor de p, estimador del grado de dominio de cada persona obtenido dividiendo el total de aciertos entre el número de reactivos.

e) El último renglón indica el estimador de la dificultad, que se obtiene dividiendo el total de fallas entre el número de personas.

A partir de la tabla se pueden hacer muy variados comentarios, como por ejemplo estos:

- Es probable que las personas (a) y (b) contesten bien a otro grupo de reactivos del mismo tipo y nivel que se les administre, ya que su probabilidad de respuesta es 1

- Es probable que las personas (i) y (j) contesten bien solamente a un 20% de las preguntas que se les administre (del mismo tipo y nivel, desde luego), ya que su probabilidad de respuesta es 0.2

Pero lo que deseamos no es hacer comentarios obtenidos por observación, sino poder sistematizar la información y disponer de mejores elementos de juicio. Nuestro objetivo va a resumirse en resolver el siguiente problema:

Obtener una función que permita estimar la probabilidad esperada de aciertos para un sujeto cuya medida se conoce ante un reactivo propuesto.

En la tabla propuesta se tienen muy pocos sujetos y pocos reactivos, por lo que no debemos esperar que la función será muy precisa, pero por el momento vamos a seguir este procedimiento simplificado:

1) Dividir al grupo en 5 subgrupos, por lo menos, conteniendo un mínimo de 2 sujetos cada uno

2) Obtener la medida de cada subgrupo

Para cada reactivo:

3) Obtener la probabilidad de respuesta de cada subgrupo

4) Representar en forma gráfica la medida contra la probabilidad de respuesta de cada subgrupo para cada reactivo

5) Ajustar una curva a los puntos obtenidos en el paso (4)

Repetir los pasos 3, 4 y 5 para todos los reactivos

Realicemos paso a paso el procedimiento indicado.

1) Dividir al grupo en subgrupos.

Por la forma que tienen los datos de la tabla, conviene agrupar en 5 subgrupos de 2 personas cada uno (a estos subgrupos se denomina “quintiles”, si se hubiera dividido el grupo en diez subgrupos se denominarían deciles).

Este paso es muy simple, ya que se puede ver que hay 2 personas para cada puntuación y por lo tanto para cada quintil. Se tiene esta tabla que agrupa los datos:

Tabla 3.2

(1) Quintil	(2) DOMINIO	(3) Núm. de personas	(4) Límite inferior del subgrupo	(5) Límite superior del subgrupo PUNTUACION (p)	(6) Puntuación media
1	BAJO	2	0.0	0.2	0.1
2		2	0.2	0.4	0.3
3	MEDIO	2	0.4	0.6	0.5
4		2	0.6	0.8	0.7
5	ALTO	2	0.8	1.0	0.9

En esta tabla los subgrupos (quintiles) se ordenan del resultado más bajo al más alto. Cada quintil agrupa los sujetos cuyas puntuaciones se encuentran entre los valores inferior y superior (columnas (4) y (5)); por ejemplo, el primer quintil tiene a los sujetos cuya calificación se encuentra entre 0 y 0.2, el segundo quintil tiene a los sujetos con calificaciones superiores a 0.2 y hasta 0.4.

La columna (6) de “puntuación media” corresponde al valor promedio de los límites de cada quintil. Por ejemplo en el quintil 4 caen las calificaciones teóricas que pudiésemos tener por arriba de 0.6 y hasta 0.8, el valor medio será, entonces:

puntuación media = (0.6+0.8)/2 = 0.7

2) Obtener la medida de cada quintil

Completamos la tabla con la columna correspondiente a la medida, como se estudió en el capítulo 2. Esta medida la calculamos sobre la “puntuación media”.

Tabla 3.3

(1) Quintil	(3) Núm. de personas	(5) Puntuación p	(6) Puntuación media	(7) Medida de la puntuación media
1	2	0.2	0.1	-2.197
2	2	0.4	0.3	-0.847
3	2	0.6	0.5	0
4	2	0.8	0.7	0.847
5	2	1.0	0.9	2.197

En esta tabla se eliminaron las columnas (2) y (4) para reducir un poco su contenido.

Los siguientes pasos se realizan para cada reactivo.

3) Obtener la probabilidad de respuesta de cada quintil.

En principio se debería hacer este cálculo con el reactivo número 1, pero para fijar ideas tomaremos un reactivo medio. Elijamos el reactivo número 3, por ejemplo, cuyas respuestas de los sujetos están sombreadas en la tabla 3.1.

Tabla 3.4

(1) Quintil	(3) Núm. de personas	(5) Puntuación p	(6) Puntuación media	(7) Medida de la puntuación media	(8) Aciertos por quintil	(9) p aciertos por quintil
1	2	0.2	0.1	-2.197	0	0
2	2	0.4	0.3	-0.847	1	.5
3	2	0.6	0.5	0	1	.5
4	2	0.8	0.7	0.847	2	1
5	2	1.0	0.9	2.197	2	1

La columna (8) aciertos por quintil se llena contando el número de casos que contestaron correctamente el reactivo en el subgrupo y la columna (9) se obtiene dividiendo el número de aciertos entre el número de personas del subgrupo.

Del primer quintil no hay respuestas correctas en el reactivo 3, se anota 0 en la columna (8). Dividiendo 0 entre 2 sujetos se tiene 0 en la columna (9).

Del segundo quintil se tiene que hay una respuesta correcta, se anota 1 en la columna (8) y al dividir 1 entre 2 sujetos se tiene 0.5 en la columna (9).

4) Representar en forma gráfica la medida contra la probabilidad de respuesta

Con las columnas (7) y (9) de la tabla 3.4 se forma la tabla 3.5, que contiene los resultados que vamos a representar en forma gráfica:

Tabla 3.5

(7) Medida de la puntuación media	(9) p por quintil
-2.197	0
-0.847	.5
0	.5
0.847	1
2.197	1

Con ayuda de un programa que permita dibujar los puntos se tiene esta gráfica. En este caso usamos EXCEL con el procedimiento indicado en el Anexo 2.

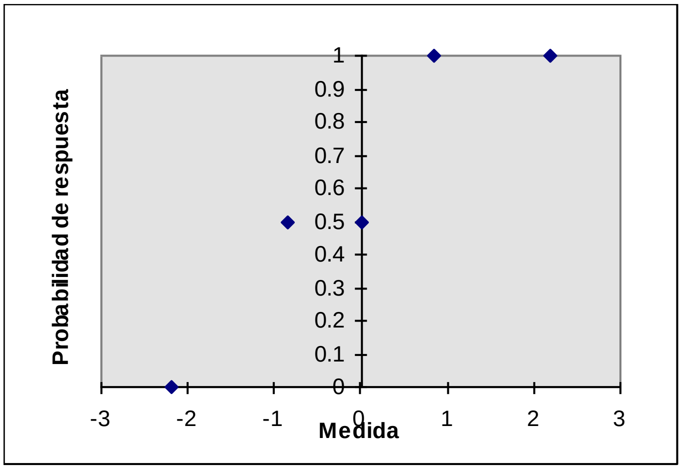


Fig.3.4 Puntos descriptivos del reactivo 3

5) Ajustar una curva a los puntos obtenidos en el paso (4)

Con ayuda del mismo EXCEL es posible ajustar una curva a los puntos obtenidos. La curva que vamos a elegir es una cúbica.

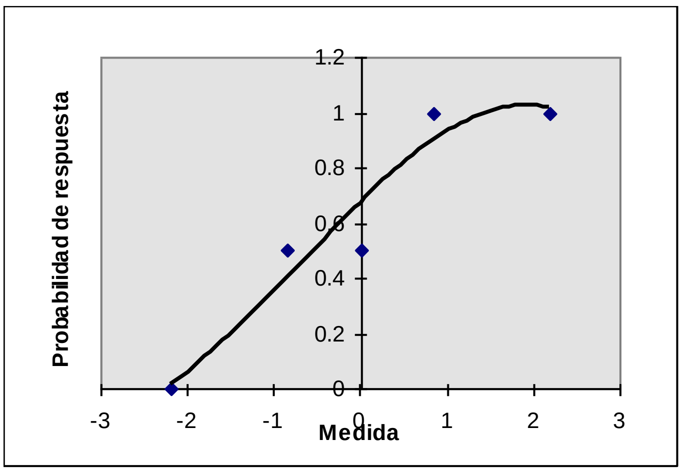


Fig.3.5 Curva de ajuste para el reactivo 3

Esta curva es una aproximación a la Curva Característica del Reactivo (CCR), que permite interpretar de manera sistemática los resultados obtenidos en el reactivo 3.

Ahora se puede afirmar con mejor fundamento, a partir de la curva que:

- para un sujeto de medida 0 se tiene una probabilidad de respuesta p=0.67
- un sujeto de medida 3, tiene una alta probabilidad de respuesta, p=1
- un alumno de medida -2 (bajo dominio), tiene baja probabilidad de respuesta p=0.08

La CCR permite disponer de un modelo que se enfoca a estimar la probabilidad de respuesta de cada persona ante un reactivo dado. Con ello se pueden hacer estimaciones de aciertos esperados en una persona.

Ejercicio 3.1

Completar la tabla mostrada para un examen aplicado a 100 alumnos

(1) Decil	(2) Núm. de alumnos	(3) Límite inferior decilar	(4) Límite superior decilar	(5) Puntuación media decilar	(6) Medida puntuación media decilar
1	10	0.000	0.10		
2	10	0.101	0.20		
3	10	0.201	0.30		
4	10	0.301	0.40		
5	10	0.401	0.50		
6	10	0.501	0.60		
7	10	0.601	0.70		
8	10	0.701	0.80		
9	10	0.801	0.90		
10	10	0.901	1.00		

<u>*Solución:*</u>

a) Cálculo del punto medio decilar, columna (5)

A partir de los datos de las columnas (3) y (4) se calcula el punto medio como el promedio de los límites extremos de las puntuaciones de cada decil. Así por ejemplo para el cuarto decil se tiene:

puntuación media =(0.3+0.4)/2 = 0.35

La figura ilustra la forma de determinar el punto medio decilar.

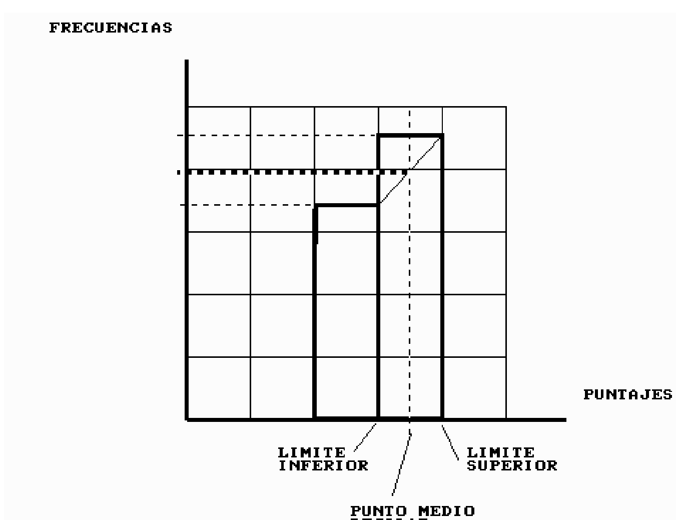


Fig.3.6

(1) Decil	(2) Núm. de alumnos	(3) Límite inferior decilar	(4) Límite superior decilar	(5) Puntuación media decilar	(6) Medida puntuación media decilar
1	10	0.000	0.10	0.05	
2	10	0.101	0.20	0.15	
3	10	0.201	0.30	0.25	
4	10	0.301	0.40	0.35	
5	10	0.401	0.50	0.45	
6	10	0.501	0.60	0.55	
7	10	0.601	0.70	0.65	
8	10	0.701	0.80	0.75	
9	10	0.801	0.90	0.85	
10	10	0.901	1.00	0.95	

b) Cálculo de la medida de la puntuación media decilar.

Para cada punto medio decilar estimamos su medida. Por ejemplo, en el tercer decil el punto medio es 25% de dominio, o sea una probabilidad p=0.25. De acuerdo con la fórmula 2.4 se tiene:

$B(0.25)= \ln(p/q) = \ln(0.25/0.75) = -1.0986$

Procediendo del mismo modo se completa la tabla.

(1) Decil	(2) Núm. de alumnos	(3) Límite inferior decilar	(4) Límite superior decilar	(5) Puntuación media decilar	(6) Medida puntuación media decilar
1	10	0.000	0.10	0.05	-2.9944
2	10	0.101	0.20	0.15	-1.7346
3	10	0.201	0.30	0.25	-1.0986
4	10	0.301	0.40	0.35	-0.6190
5	10	0.401	0.50	0.45	-0.2007
6	10	0.501	0.60	0.55	0.2007
7	10	0.601	0.70	0.65	0.6190
8	10	0.701	0.80	0.75	1.0986
9	10	0.801	0.90	0.85	1.7346
10	10	0.901	1.00	0.95	2.9944

Ejercicio 3.2

Se aplicó un cuestionario a un grupo de 100 personas. Las respuestas se organizaron en una tabla similar a la descrita en la tabla 3.1 y se agruparon los resultados del reactivo 1 en la tabla que se muestra a continuación.

(1) Decil	(2) Núm. de alumnos	(3) Límite inferior decilar	(4) Límite superior decilar	(5) Puntuación media por decil	(6) Medida de la puntuación media por decil	(7) Número de aciertos	(8) Probabilidad de respuesta por decil
1	10	0.000	0.10	0.05	-2.9944	1	
2	10	0.101	0.20	0.15	-1.7346	1	
3	10	0.201	0.30	0.25	-1.0986	2	
4	10	0.301	0.40	0.35	-0.6190	4	
5	10	0.401	0.50	0.45	-0.2007	4	
6	10	0.501	0.60	0.55	0.2007	5	
7	10	0.601	0.70	0.65	0.6190	5	
8	10	0.701	0.80	0.75	1.0986	7	
9	10	0.801	0.90	0.85	1.7346	8	
10	10	0.901	1.00	0.95	2.9944	8	

.

a) Calcular el Grado de Dificultad y estimar la calibración de la Dificultad del reactivo

b) Completar la tabla

<u>Solución:</u>

Se observa que se trata de una tabla de puntuaciones decilares donde ya se calculó el punto medio y su respectiva medida.

Por un procedimiento similar al que se usó en esta sección se obtuvieron los aciertos de cada uno de los deciles. Esto quiere decir que se encontró de la tabla de respuestas, por ejemplo, que de las 10 personas que constituyen al decil 7, hubo solamente 5 personas que contestaron correctamente el reactivo.

a) El Grado de Dificultad es la forma clásica de expresar el porcentaje de aciertos que se consiguieron en un reactivo. En este caso se tiene que el número de aciertos es 45, sumando los aciertos de la columna (7)

El Grado de Dificultad se expresa:

GD=Aciertos/Total de sujetos X 100 = 45/100 x 100 = 45%

Un estimado de la calibración del reactivo se obtiene con el número de fallas que se tienen en un reactivo y calculando el logaritmo del momio, como se vio en el capítulo anterior.

q = proporción de fallas = fallas / Total de sujetos = 55/100

q= 0.55

su complemento es p=1-q = 0.45

La dificultad en lógitos es:

D=ln(q/p) = ln(0.55/0.45) = ln(1.2222) = 0.20

b) Para completar la tabla basta con calcular la proporción de aciertos por cada decil, lo cual se obtiene dividiendo el número de aciertos entre el número de personas del decil (corresponde a dividir las columnas (7) y (2) de la tabla). Por ejemplo, para el decil 7 se tiene p(decil 7) = 5/10 = 0.5

Del mismo modo se completa la tabla

(1) Decil	(2) Núm. de alumnos	(3) Límite inferior decilar	(4) Límite superior decilar	(5) Puntuación media por decil	(6) Medida de la puntuación media por decil	(7) Número de aciertos	(8) Probabilidad de respuesta por decil
1	10	0.000	0.10	0.05	-2.9944	1	0.1
2	10	0.101	0.20	0.15	-1.7346	1	0.1
3	10	0.201	0.30	0.25	-1.0986	2	0.2
4	10	0.301	0.40	0.35	-0.6190	4	0.4
5	10	0.401	0.50	0.45	-0.2007	4	0.4
6	10	0.501	0.60	0.55	0.2007	5	0.5
7	10	0.601	0.70	0.65	0.6190	5	0.5
8	10	0.701	0.80	0.75	1.0986	7	0.7
9	10	0.801	0.90	0.85	1.7346	8	0.8
10	10	0.901	1.00	0.95	2.9944	8	0.8

Ejercicio 3.3

Con ayuda de EXCEL o de algún programa similar, dibujar los puntos de las columnas (6) y (8) de la tabla del ejercicio 3.2 y hacer el ajuste a una cúbica (polinomio de grado 3).

Solución:

En el eje horizontal dibujamos la medida en lógitos (columna (6)), en el eje vertical dibujamos la probabilidad por cada decil (columna (8)).

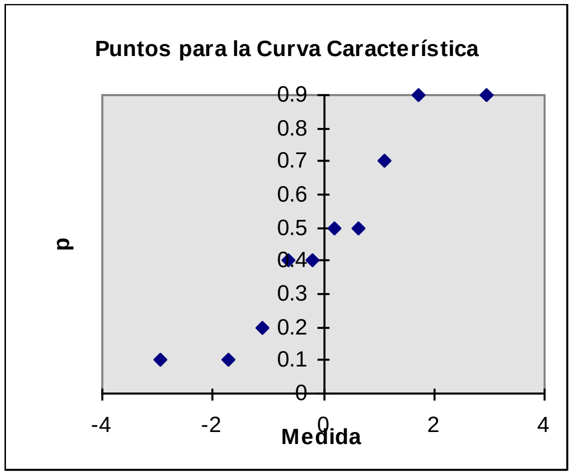


Fig.3.7

Una vez dibujados los puntos se hace pasar una curva “suave” cerca de ellos. En EXCEL puede solicitarse el ajuste a una cúbica como se indica en el anexo 2.

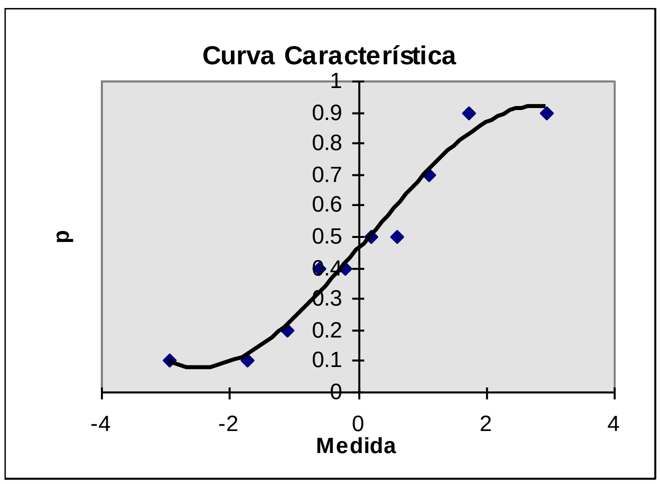


Fig. 3.8

El trabajo realizado hasta ahora nos ha permitido obtener una aproximación a la CCR (Curva Característica del Reactivo). Para los datos de los ejercicios 3.1 a 3.3, la probabilidad es creciente con la medida de las personas (a mayor medida se tiene mayor probabilidad de respuesta). Esto es muy deseable, ya que permite decir que el reactivo mide en la dirección esperada: los alumnos de mayor desempeño responden mejor que los de bajo desempeño. Dicho de otra manera: un alumno de los deciles más altos correlaciona con probabilidades de respuesta más altas. Se dice que un reactivo como éste discrimina correctamente a los alumnos, o que tiene una alta discriminación positiva. Sería motivo de preocupación que el reactivo se comportara al revés.

La curva 3.9 corresponde a un reactivo que discrimina de manera inversa y su curva es decreciente con la medida.

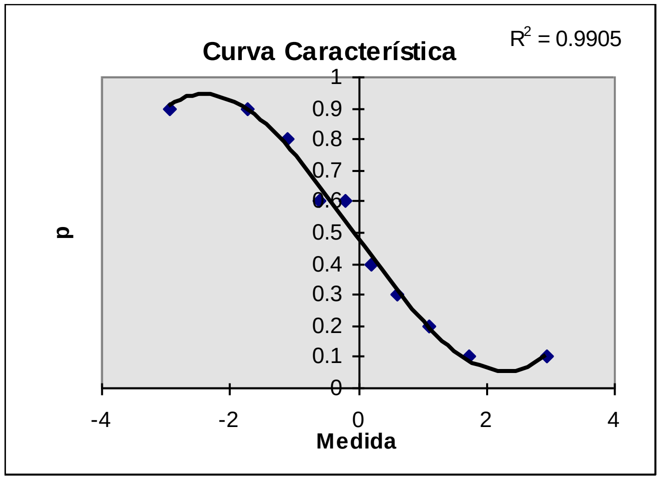


Fig. 3.9 Curva característica

Este reactivo se comporta de manera anormal, porque los alumnos de los deciles más bajos lo contestan bien, mientras que los mejores alumnos tienen una probabilidad muy baja respuesta. Si un reactivo hiciera esto (y hay muchos casos de reactivos que así se comportan) entonces podríamos pensar que hay problemas con él: confunde a los alumnos, la respuesta correcta tal vez esté equivocada, o quizá el reactivo mide otra cosa y no el rasgo deseado. Este reactivo discrimina en forma inversa (su poder de discriminación es negativo) y debería ser desechado del cuestionario.

La curva 3.10 muestra un caso en el cual los puntos no se parecen mucho a una curva, no obstante EXCEL permite hacer el ajuste a una cúbica.

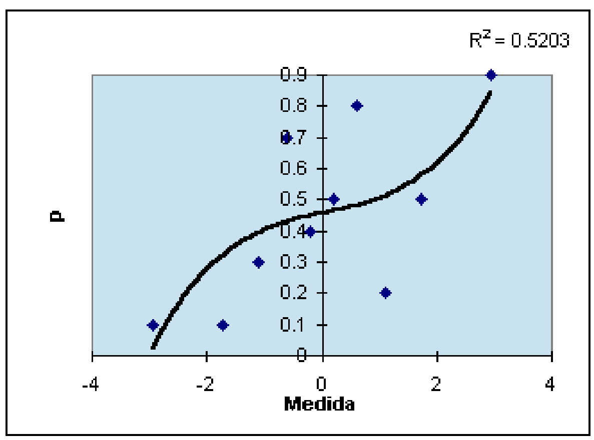


fig. 3.10 Curva característica

En las figuras 3.8 y 3.9 vemos que los puntos pasan cerca de una curva de aspecto muy elegante, misma que estudiaremos en el capítulo 4. En general puede afirmarse que los puntos no pasan exactamente por una curva, pero es posible buscar alguna curva que se acerque o que “ajuste” a los datos encontrados. Este ajuste tendrá mayor o menor error dependiendo de qué tan cerca pase de los puntos.

Al construir las curvas con EXCEL se puede pedir que indique la correlación obtenida a la curva de aproximación. En estos casos hemos elegido una cúbica y podemos apreciar que la curva de la figura 3.10 tiene una correlación muy baja (0.7211), mientras que las curvas de las figuras 3.8 y 3.9 tienen correlaciones bastante altas (0.9842 y 0.9952). Aunque poco usual entre los evaluadores, la correlación es una medida correcta del error de ajuste de la curva característica del reactivo.

Nota: En el caso de las curvas obtenidas en EXCEL, se obtiene el valor de R^2, la correlación es la raíz cuadrada de este valor.

Antes de ver otros ejemplos, supóngase que se tienen estos resultados:

Tabla 3.6

(1) Decil	(2) Núm. de sujetos	(3) Calificación decilar	(4) Punto medio decilar	(5) Medida del punto medio decilar B	(6) Probabilidad de respuesta $p(B) = e^B/(1+e^B)$
1	10	10	5	-2.9944	0.05
2	10	20	15	-1.7346	0.15
3	10	30	25	-1.0986	0.25
4	10	40	35	-0.6190	0.35
5	10	50	45	-0.2007	0.45
6	10	60	55	0.2007	0.55
7	10	70	65	0.6190	0.65
8	10	80	75	1.0986	0.75
9	10	90	85	1.7346	0.85
10	10	100	95	2.9944	0.95

La columna (6) no presenta datos observados, sino valores calculados utilizando la fórmula de la probabilidad teórica utilizando directamente las fórmulas del capítulo 2.

Por ejemplo, dada la medida -0.6190 para el decil 4, es fácil calcular la probabilidad de 0.35. Recuérdese que

$p(B) = e^B/(1+e^B) = e^{-0.6190}/(1+e^{-0.6190}) = 0.35$

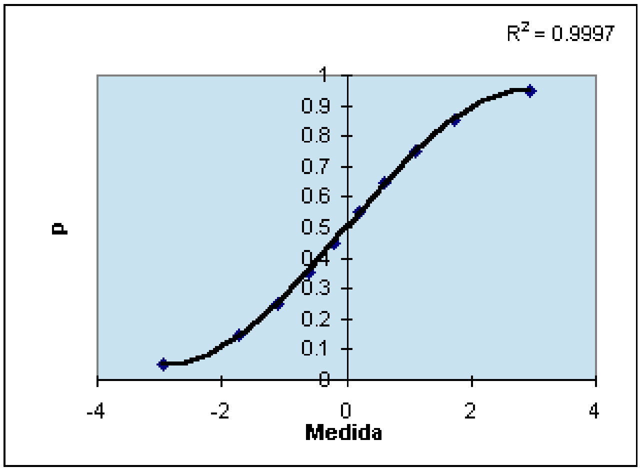


Fig.3.11 Curva característica ajustada al modelo de medida

Si se hace la gráfica para esta tabla se obtiene la curva 3.11, que corresponde a la curva característica de un reactivo que ajusta perfectamente al modelo de medida dado por p (probabilidad de respuesta) y B (medida), la correlación obtenida en el caso de una cúbica es 0.9997 (prácticamente perfecta).

Las curvas que hemos dibujado **NO SON**, rigurosamente hablando, las curvas características de los reactivos, pero se parecen mucho. Para fines didácticos lo importante es conocer la idea básica de las curvas características de los reactivos, aunque no sea una construcción muy exacta. Desde el punto de vista práctico, lo que hemos hecho hasta ahora cumple el objetivo que busca este libro. Ya tendremos ocasión de mejorar el modelo más adelante.

Rigurosamente hablando, desde el punto de vista de las teorías que vamos a ver posteriormente, no son las curvas características de los reactivos por varias razones:

a) las estamos obteniendo directamente de la tabla de frecuencias por deciles
b) las construimos usando las frecuencias relativas a cada decil, como una aproximación a la probabilidad
c) no "separan" o distinguen la medida de la persona y la calibración del reactivo. Sobre este punto tendremos oportunidad de hablar más adelante.

Para disponer de las verdaderas curvas características de un reactivo necesitamos conocer la verdadera medida de las personas (no solamente su posición en cada decil) y requerimos una probabilidad más certera para cada valor de medida. Hemos dicho desde el capítulo 2 que nos falta información para poder conocer la medida de las personas y la calibración de los reactivos. Esto lo veremos en capítulos siguientes.

Ejercicio 3.4.

Se aplica un examen de 30 preguntas a un grupo de 20 alumnos. Determine las calificaciones decilares para el siguiente conjunto de calificaciones.

0,6,0,0,4,10,28,15,11,23,19,19,14,3,5,13,12,21,22,8

Solución

Primero debemos ordenar las calificaciones en forma ascendente. Tenemos, por lo tanto la siguiente lista ordenada:

0,0,0,3,4,5,6,8,10,11,12,13,14,15,19,19,21,22,23,28

Se desea dividir al grupo en deciles. Cada decil tiene al 10% de personas, en este caso:

Núm.de personas/decil = 20 personas x 0.1 = 2 personas /decil

Primer método.

Tomamos subgrupos de 2 personas y los acomodamos en una tabla por deciles:

Decil	Núm. de personas	calificaciones	calificación decilar
1	2	0	0
2	2	0	3
3	2	4	5
4	2	6	8
5	2	10	11
6	2	12	13
7	2	14	15
8	2	15	19
9	2	21	22
10	2	23	28

Segundo método.

Hay varias formas de obtener las calificaciones decilares, proporcionando mayor o menor error en función de la precisión que se busque. Para fines prácticos la construcción propuesta es suficiente, aunque se tiene el problema del decil 7 que tiene una calificación asignada de 15. Es costumbre , por lo tanto, utilizar de preferencia una tabla de frecuencias e interpolar para obtener los valores de los deciles.

Hagamos la tabla de frecuencias con intervalos de 4 en 4. Las frecuencias relativas se calculan dividiendo la frecuencia de cada intervalo entre el número de alumnos, expresado en por ciento. La tabla de frecuencias relativas acumuladas se obtienen simplemente sumando las frecuencias relativas.

Intervalo de calificaciones	Frecuencia	Frecuencias relativas %	Frecuencias. Relativas acumuladas
0-3.9	4	20%	20%
4-7.9	3	15%	35%
8-9.9	1	5%	40%
10-13.9	4	20%	60%
14-17.9	3	15%	75%
18-21.9	2	10%	85%
22-25.9	2	10%	95%
26-28	1	5%	100%
TOTAL	20	100%	

3.2 Ejercicios propuestos

1. Completar la tabla siguiente.

1	2	3	4	5	6	7
Decil	Núm. de alumnos	Calificación decilar	Punto medio decilar	Medida	Núm de aciertos	Probabilidad de respuesta por decil
1	10	10			10	
2	10	20			8	
3	10	30			7	
4	10	40			5	
5	10	50			4	
6	10	60			4	
7	10	70			3	
8	10	80			3	
9	10	90			2	
10	10	100			2	
Total	100				48	

2. Complete la tabla siguiente y dibuje los puntos para ajustar a una curva

1	2	3	4	5	6	7
Decil	Núm de alumnos	Calificación decilar	Punto medio decilar	Medida	Núm. de aciertos	Probabilidad de respuesta por decil
1	10	15			0	
2	10	22			1	
3	10	34			3	
4	10	56			3	
5	10	69			4	
6	10	74			6	
7	10	77			6	
8	10	80			8	
9	10	82			9	
10	10	87			9	
TOTAL	100					

3. Use una hoja de EXCEL para cada juego de datos de los ejercicios de este capítulo y trace los puntos y las curvas. En particular complete el trabajo para los datos de la Tabla 3.1 para los reactivos 1 a 5.

4. La tabla presenta los resultados de 20 alumnos a un examen de 8 reactivos. Ordene los datos para dibujar la curva característica de cada uno de los reactivos. las respuestas correctas están identificadas por 1 y las incorrectas por 0. Utilice quintiles.

ALUMNO	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	0	1
2	1	1	1	1	1	0	1	0
3	1	1	0	1	0	1	1	1
4	1	1	1	0	0	1	1	1
5	1	0	1	1	1	0	1	1
6	1	1	0	1	1	1	1	1
7	1	1	1	0	1	0	1	0
8	0	1	1	1	1	1	1	1
9	0	1	0	1	0	1	1	1
10	1	0	1	0	0	1	0	1
11	0	1	1	1	1	1	0	1
12	1	1	1	0	0	0	0	1
13	1	1	1	0	0	0	0	0
14	1	0	0	1	1	0	0	0
15	1	0	0	0	1	1	0	0
16	1	0	1	0	0	0	0	0
17	1	1	0	0	0	1	0	0
18	1	1	0	1	1	0	0	0
19	0	1	1	0	0	1	0	0
20	0	1	0	0	0	1	0	1

5. Explique para qué sirve la CCR. ¿Cómo se define el plano (eje horizontal, eje vertical) en que se dibuja la CCR?

6. ¿Por qué se ha insistido en este capítulo que la construcción de la CCR es solamente una aproximación?

3.3 Puntos clave del capítulo

Curva Característica del Reactivo: Es la curva que se traza en el plano Medida en lógitos vs probabilidad de respuesta.
Permite estimar la probabilidad de respuesta de una persona cuya medida se conoce; inversamente, permite estimar la medida de una persona que tiene una probabilidad de respuesta dada.

Trazo aproximado de la CCR: En este capítulo se sugiere un método que sigue estos pasos:

1) Dividir al grupo en 5 subgrupos, por lo menos, conteniendo un mínimo de 2 sujetos cada uno

2) Obtener la medida de cada subgrupo

Para cada reactivo:

3) Obtener la probabilidad de respuesta de cada subgrupo

4) Representar en forma gráfica la medida contra la probabilidad de respuesta de cada subgrupo para cada reactivo

5) Ajustar una curva a los puntos obtenidos en el paso (4), por el momento usar una cúbica

Repetir los pasos 3, 4 y 5 para todos los reactivos

4. MODELOS PARA LA CURVA CARACTERISTICA

Ahora estudiemos la curva característica para ver de qué manera nos puede ayudar en el análisis de los reactivos y veremos algunos modelos matemáticos para definirla.

4.1 Propiedades de la curva característica

Dibujemos la curva de un reactivo ideal, como se muestra en la figura 4.1.

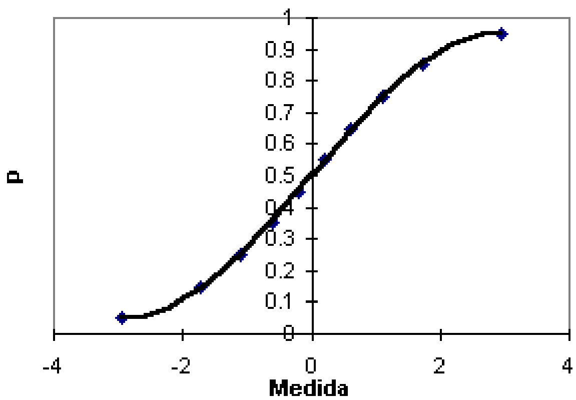


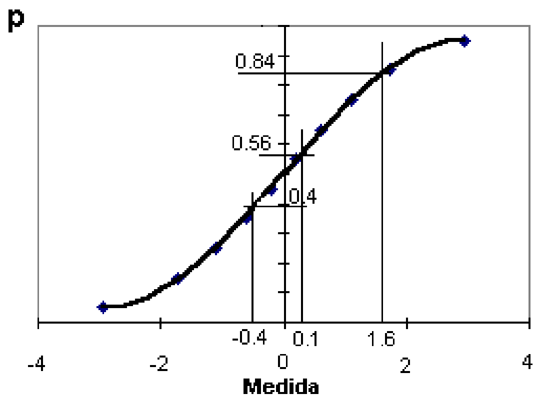
Fig.4.1 Curva característica de un reactivo

Véase que un alumno de medida baja , por ejemplo -3 (dominio muy bajo) tiene una probabilidad prácticamente nula de contestar correctamente al reactivo. Contrariamente, un alumno de medida alta, por ejemplo +3 (dominio alto) tiene una probabilidad prácticamente de 1 de contestar el reactivo. Entre estos dos extremos hay alumnos con diferentes medidas y por tanto diversas probabilidades de responder correctamente al reactivo.

Ejercicio 4.1

Determinar de manera aproximada, con ayuda de la curva, las probabilidades de respuesta correcta de tres alumnos cuyas medidas son: -0.4, 0.1 y 1.6

Solución:



A partir de los valores que leemos directamente sobre la curva, se tiene:

- para la medida de -0.4 levantamos una vertical hasta cortar a la curva. La altura del punto de intersección con la curva es la probabilidad pedida: 0.4.
- para la medida 0.1 la vertical corta a la curva aproximadamente en una probabilidad de respuesta correcta de 0.56.
- para la medida de 1.6 se tiene una altura del punto de intersección en 0.84

Retomando la figura 4.1 se pueden hacer sobre ella estas anotaciones:

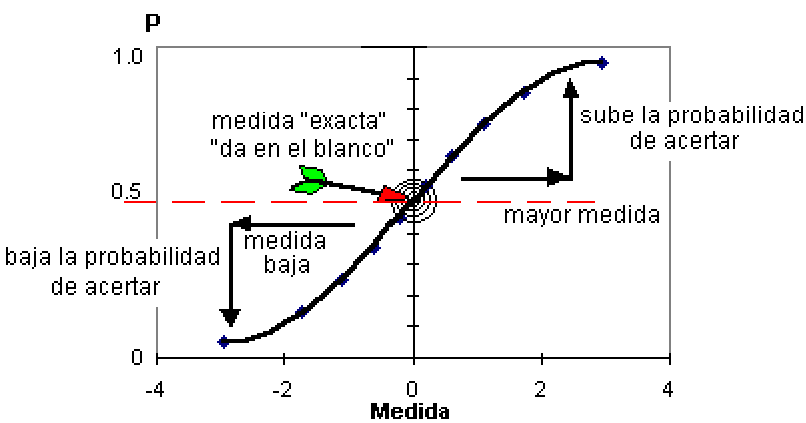


Fig. 4.1 bis

Recojamos estas anotaciones pongámoslas en forma simbólica. Denotemos por p(B,D) la probabilidad de respuesta de la persona de medida B a un reactivo de calibración D. A partir de la definición de la p se tienen estas propiedades para la curva de la figura 4.1:

p(B,D) □ [0.5,1]　　　　si B>D　　　　(4.1a)

esto se lee: p(B,D) toma valores en el intervalo de 0.5 a 1 si B es mayor que D

p(B,D) □ [0,0.5]　　　　si B<D　　　　(4.1b)

Léase: p(B,D) toma valores en el intervalo de 0 a 0.5, si B es menor que D

p(B,D) = 0.5　　　　si B=D　　　　(4.1c)

Léase: p(B,D) es igual a 0.5 cuando el reactivo mide exactamente a la persona.

Cuando el reactivo “apunta” a la medida exacta de la persona se dice que “da en el blanco” (“TARGET” en inglés), en el reactivo de la figura 4.1 esta probabilidad correspondiente a la capacidad de la personas vale 0.5. Al aplicar el reactivo se podrá tener respuesta acertada o errónea si la calibración fuera incorrecta o si la persona tuviera otra medida. Pero también basta con un descuido del alumno al contestar, con que tenga una confusión al interpretar la pregunta, con que haya funcionado una buena opción distractora, con que el día del examen se encuentre enfermo o indispuesto, para que la respuesta del alumno se modifique. Esto forma parte del error de la medición.

El punto donde el reactivo “da en el blanco” (la Dificultad) corresponde con el punto de inflexión de la Curva Característica del Reactivo. El punto de inflexión identifica el lugar donde la curva cambia de concavidad: a la derecha del punto de inflexión la curva tiene la concavidad hacia abajo, a la izquierda del punto de inflexión la concavidad es hacia arriba. En la curva que aparece en la figura 4.1 este punto de inflexión coincide con p=0.5, pero puede estar en otro sitio.

Aquí tenemos una primera definición muy importante:

Se denomina “Dificultad” del reactivo a la medida del punto de inflexión de la Curva Característica del Reactivo.

Se acostumbra denominar a la Dificultad con la letra “b”.

El reactivo de la figura 4.1 tiene dificultad o calibración de 0. Se puede afirmar que un alumno cuya medida es 0 tiene una probabilidad de respuesta de 0.5. Observamos algo muy interesante en la curva característica: La escala se aplica tanto para la medida de las personas como para la calibración de los reactivos.

Analicemos el caso en el cual se aplican reactivos más fáciles o más difíciles. La figura 4.2 muestra ahora tres curvas:

- La curva central (1) es la misma presentada en la figura 4.1, correspondiendo a un reactivo de Dificultad 0

- La curva de la izquierda (2) es una curva para un reactivo más fácil

- La curva de la derecha (3) es una curva para un reactivo más difícil.

En un reactivo fácil, la probabilidad de respuesta de cualquier alumno de medida B aumenta con relación a la curva 1. Las curvas fáciles se dibujan hacia la izquierda del diagrama. En el reactivo difícil la probabilidad de respuesta para una medida B es, como podríamos esperar, menor que la de la curva 1. Esto se escribe, en forma simbólica:

- si B1>B2, p(B1,D)>p(B2.D), para toda D

- si B1<B2, p(B1,D) < p(B2,D) , para toda D

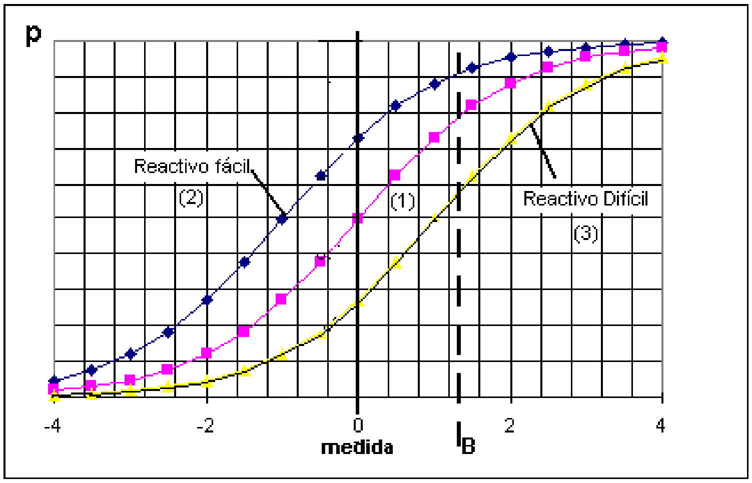


Fig 4.2 Curvas características de diversos reactivos

En los tres casos dibujados hasta ahora los puntos de inflexión están en p=0.5 y la Dificultad es la medida en dichos puntos, de acuerdo con esto, la Dificultad de la curva 2 es -1 y la de la curva 3 es 1 (la Dificultad tiene por unidades a los lógitos).

Las tres curvas dibujadas en la figura 4.2 se obtuvieron por una traslación sobre el eje horizontal (en lenguaje coloquial, aunque matemáticamente incorrecto, se dice que estas curvas son “paralelas”). Pero en el capítulo anterior vimos que las curvas no necesariamente tienen la misma forma, de hecho pueden tener formas muy diversas. Dibujemos ahora un juego de curvas de igual dificultad pero “inclinación” diferente.

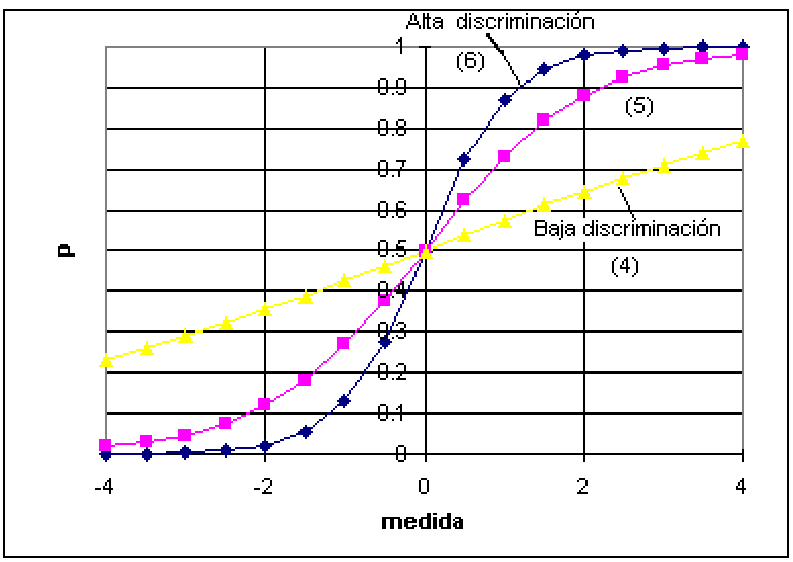


Fig.4.4 Curvas características de diferente pendiente

En la figura 4.4 podemos apreciar que la curva 6 es de mayor pendiente que las otras (más “inclinada” dirán algunos, más “vertical” le llamarán otros), mientras que la curva 4 es la de menor pendiente. Las tres curvas son “positivas”, lo que quiere decir que cada curva es monótonamente creciente (conforme crece la medida en lógitos crece la probabilidad de responder correctamente al reactivo).

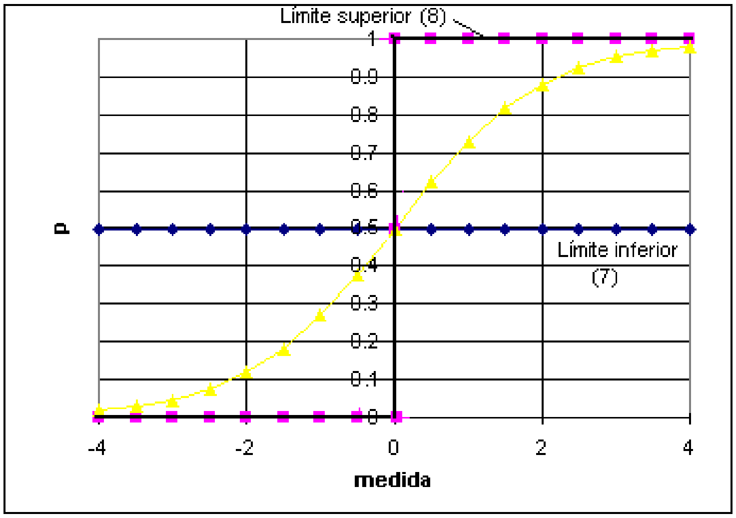


fig. 4.5 Límites de las curvas “positivas”

Hay un límite inferior para las curvas “positivas”, se trata de la curva que coincide con una horizontal como se muestra en la curva 7 de la figura 4.5. En este caso se tiene que la probabilidad es una constante (0.5 en este caso) para todos los valores de medida. Un reactivo con estas características puede ser contestado de la misma manera, es decir con igual probabilidad, tanto por los alumnos de menor medida como los de mayor medida. Es por ello que el reactivo “no discrimina” entre los alumnos de alto desempeño y los de bajo nivel; se trataría de un reactivo de “discriminación nula”. Un reactivo de este tipo no es interesante para ser utilizado en un examen donde queremos identificar a los alumnos de alto desempeño y separarlos de los de bajo desempeño.

Hay un límite superior para estas curvas. Se trata del caso en que todos los alumnos por arriba de la medida del reactivo contestan correctamente, mientras que los de abajo de la medida fallan siempre. Esto se representa en la curva 8 (figura 4.5). Por arriba de la medida 0 la probabilidad de respuesta es 1 (todos contestan correctamente), mientras que por abajo la probabilidad es 0 (nadie contesta correctamente). El reactivo de la curva (8) “discrimina perfectamente” a los alumnos de desempeño superior de los alumnos de niveles inferiores. Los reactivos de este tipo son preferibles, aunque dudosamente podremos tener muchos reactivos tan “perfectos” en nuestro examen.

Entre los extremos citados (discriminaciones nula y máxima) los reactivos tienen diversas inclinaciones. Son deseables los reactivos de mayor discriminación posible, para un grado de dificultad dado. El segundo parámetro que interesa en este modelo es, entonces, la Discriminación, con lo que tenemos la segunda definición importante:

Se define como “Discriminación” a un valor que permite distinguir a los alumnos de alto desempeño de los de bajo desempeño y es un función de la pendiente en el punto de inflexión de la Curva Característica del Reactivo.

En la Teoría de la Respuesta al Reactivo la Discriminación se relaciona con la pendiente de la curva y que se denota con la letra “a”.

En este modelo no interesan los reactivos “negativos” o monótonamente decrecientes, como el caso que se vio en el capítulo anterior, o como se muestran en la figura 4.6. Los reactivos “negativos” son indeseables ya que se comportan de manera contraria a lo esperado y, por lo tanto, se deben desechar de un cuestionario.

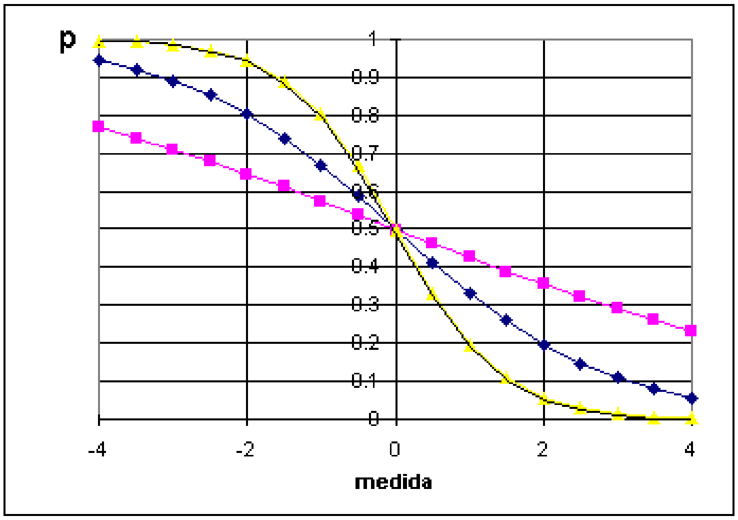


Fig. 4.6 Reactivos de comportamiento “negativo”

Las curvas características tienen, como podemos ver, una forma de S extendida en el plano medida-probabilidad. Esta curva se parece mucho a la curva logística empleada en diversas áreas del conocimiento: demografía, desarrollo industrial, planeación, estudios de mercado, etc. La curva logística “positiva” se caracteriza por ser creciente y tener tres zonas más o menos claras: una zona baja a la izquierda, una zona central aproximadamente recta y una zona alta a la derecha.

Introduzcamos una tercera definición:

Se define como “Extensión” al rango en el que se encuentra la zona central de la curva logística.

Ahora veamos un aspecto adicional de las curvas características. La figura 4.7 muestra dos curvas que tienen la misma dificultad, la misma pendiente y diferentes extensiones.

La curva (9) tiene una menor extensión (e1) y corta al eje vertical para el valor más bajo de la medida en una altura c1. El valor c1 es un valor de probabilidad más alto que el que ocurre con la curva (11) de mayor extensión. El punto de corte con el eje vertical para el valor más bajo de la medida se denota con la letra “c”.

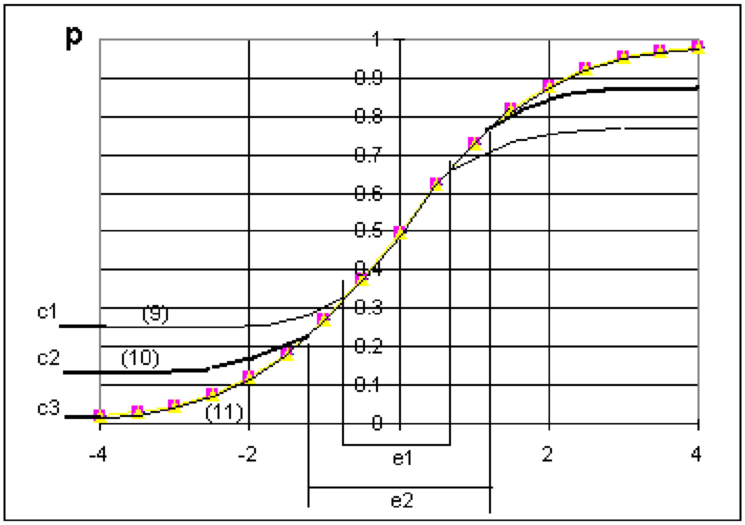


Fig. 4.7 Curvas de diversas extensiones

La extensión sería un parámetro de mucha importancia para la descripción de un reactivo, pero los evaluadores generalmente no reparan en la extensión sino en la ordenada para el valor de medida más baja, que será el tercer parámetro de interés para el modelo. Introducimos la cuarta definición importante:

Se denomina "adivinación sistemática" al valor de probabilidad de acierto para las personas de medida más baja. También puede decirse que es la probabilidad mínima esperada de respuesta en un reactivo.

Un alumno que no sepa nada del objetivo o contenido del reactivo, podría tener por lo menos esta probabilidad de respuesta, si tratara de contestar al azar. Se considera que es la "carga de azar" que contiene de manera inherente el reactivo. Este parámetro se denota con la letra "c". Al parámetro "c" también se le denomina "parámetro de adivinación" y "nivel de puntaje por pseudo-adivinación" entre otros nombres.

Se dice que un reactivo tiene "adivinación sistemática" si el valor "c" es mayor que 0. Muchos evaluadores aceptan que c representa la probabilidad que tiene una persona de medida inferior para contestar un reactivo por azar o por adivinación. Este parámetro se aplica por igual a TODOS los alumnos, por ello se le llama adivinación sistemática.

Es claro que el parámetro "c" puede ser interesante para describir a un reactivo; pero de ahí a aceptar que el reactivo es respondido por todos los alumnos con el mismo patrón de adivinación hay un gran trecho. En realidad es incorrecto suponer que "c" representa adivinación sistemática, ya que no es posible demostrar que todos los alumnos contestan con el mismo patrón de azar. No hay duda que en un grupo de personas que contestan un cuestionario puede haber quienes intenten contestar los reactivos al azar, pero seguramente también habrá alumnos que contestarán muy pocas preguntas por adivinación y muchos más que responderán sin adivinar.

Hasta aquí hemos visto que la curva característica puede describirse por medio de tres parámetros geométricos: a (discriminación), b (dificultad) y c (adivinación sistemática), como se muestra en la figura 4.8.

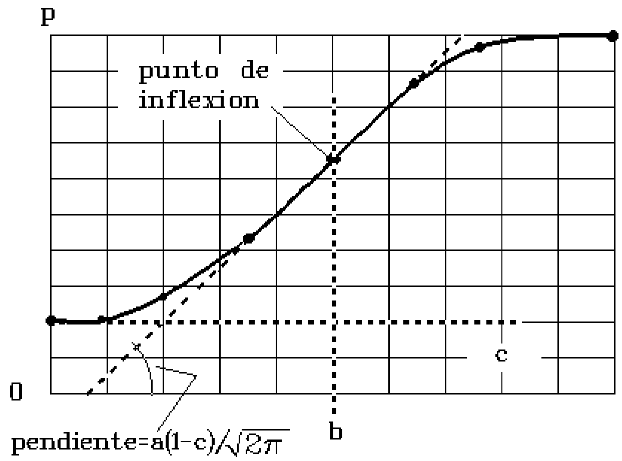


Fig. 4.8 Parámetros de la Curva Característica del Reactivo

El punto en el que se encuentra el valor de b, corresponde con el “punto de inflexión”. Cuando c es igual a 0 el punto de inflexión coincide con p=0.5, en caso contrario el punto se encuentra al centro del intervalo entre c y 1. Puede establecerse esta fórmula:

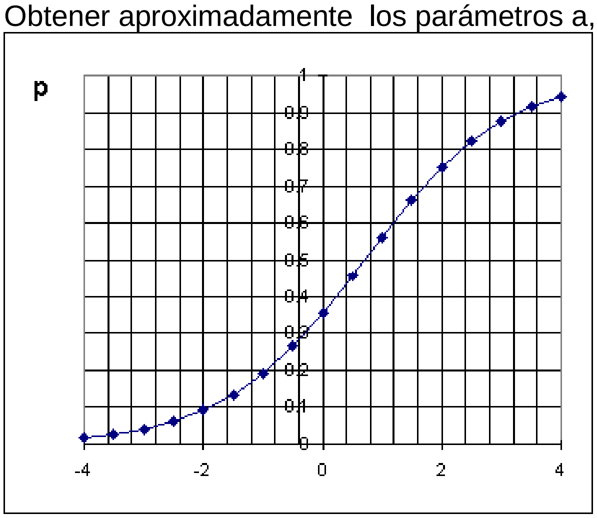
Punto de inflexión = (c+1)/2 (4.2)

En la misma figura 4.8 se señala la relación entre la pendiente de la curva logística sobre el punto de inflexión y los parámetros a y c. A su vez, la pendiente se calcula con la tangente del ángulo que forma la “parte recta” de la curva logística y el eje horizontal.

a= Pendiente $\sqrt{2\pi}$ /(1-c) (4.3)

Ejercicio 4.2

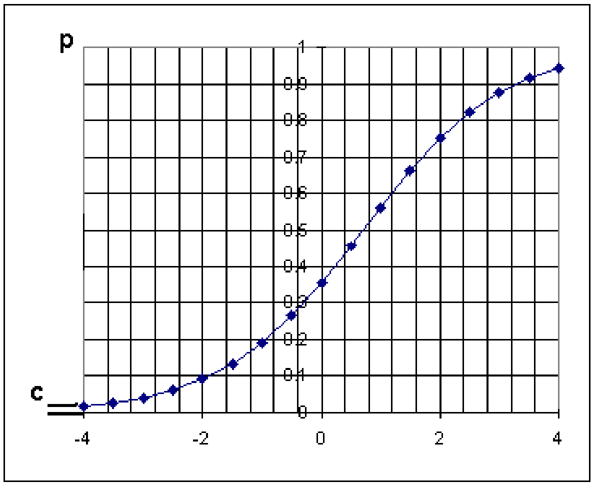
Obtener aproximadamente los parámetros a, b y c de la curva mostrada



<u>*Solución*</u>

1) Parámetro “c”. Adivinación Sistemática.

No hay un valor preciso para el cual debe identificarse al parámetro c, en principio habría que buscarlo en la medida de menos infinito (dominio nulo), pero para fines prácticos es común cortar la CCR por abajo de 2.5 a 3 lógitos. En este caso se usarán los 4 lógitos mostrados en la figura del problema.



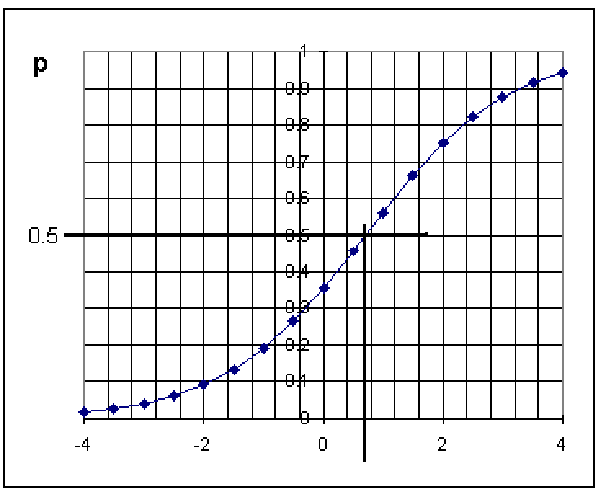
Se tiene que prácticamente c=0 en la medida de -4.

2) Punto de inflexión

Este punto ocurre en p=(c+1)/2, en este caso, como c=0 se tiene p=0.5

3) Parámetro "b"

Ahora determinemos la Dificultad, que es la medida en el punto de inflexión. En este caso basta con identificar el punto para el cual p=0.5

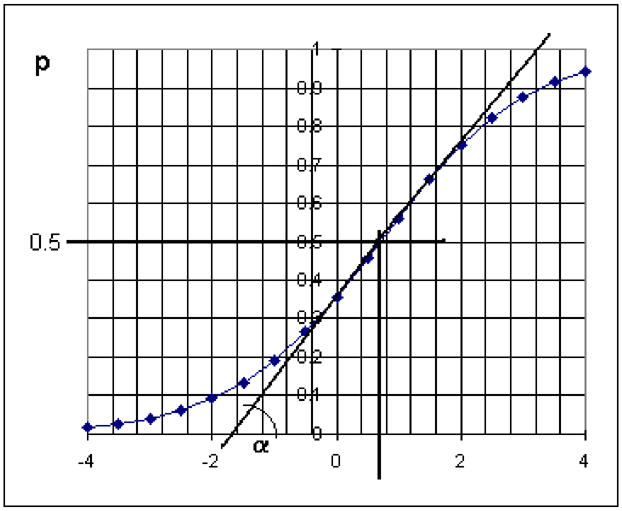


Trácese una recta horizontal que pasa por p=0.5 y se tiene una intersección aproximadamente en 0.7, que es el valor de Dificultad del reactivo.

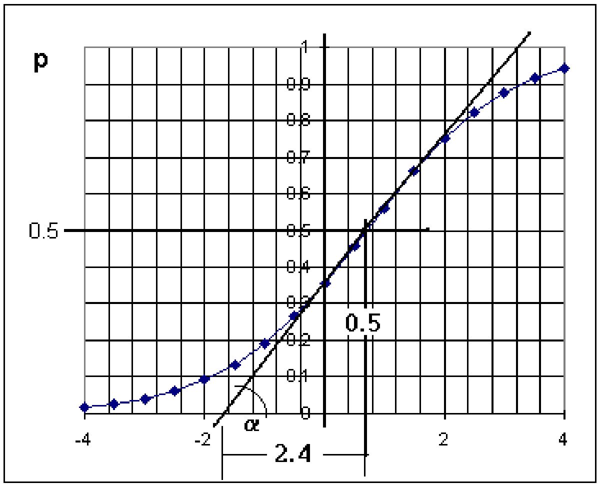
4) Parámetro "a"

Queda por último determinar la Discriminación. A partir del punto de inflexión se traza la recta tangente a la CCR y se determina su pendiente. Aquí usaremos un método simplificado para determinar esta pendiente.

Primeramente se traza la recta que pasa por el punto de inflexión tangente a la CCR, como se muestra en la figura:



La recta trazada es la hipotenusa de un triángulo, conjuntamente con el eje horizontal y el eje vertical. Tomemos las medidas de los dos catetos:



La pendiente se obtiene calculando la tangente del ángulo α, que es igual a dividir la altura entre la base del triángulo:

Pendiente = tan α = cateto vertical/cateto horizontal

Pendiente = 0.5/2.4 = 0.2083

Por último la Discriminación se obtiene con la expresión:

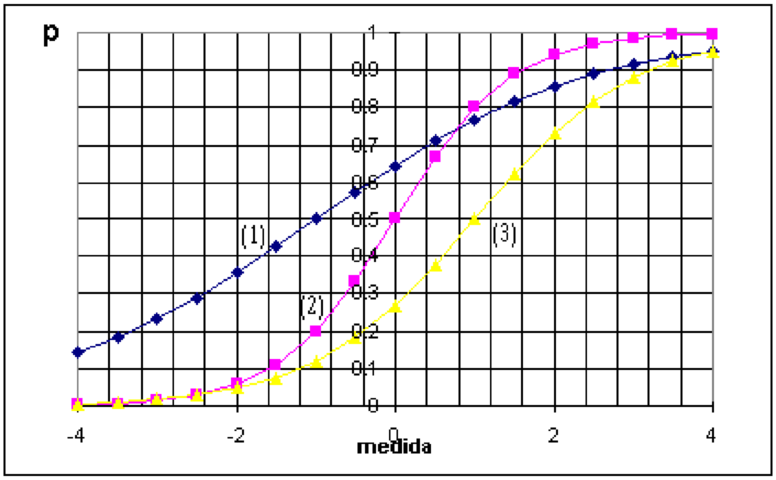
a= Pendiente $\sqrt{2\pi}$ /(1-c) = 0.2083 x 2.5 / (1-0) = 0.52

En algunas referencias se maneja esta otra expresión aproximada:
a = Pendiente x 2.353 / (1-c)

que tiene una diferencia de menos del 6% con respecto a la expresión utilizada en la figura 3.8.

Ejercicio 4.3

Clasificar y ordenar los reactivos de la figura de acuerdo con los parámetros a, b y c, y dar un dictamen sobre la dificultad, la discriminación y la adivinación sistemática.



<u>*Solución:*</u>

A partir de las curvas se pueden obtener sus parámetros, siguiendo el procedimiento del ejercicio 4.2, mismos que se organizan en la tabla:

CURVA NO.	a	b	c
(1)	0.18	-0.6	0.15
(2)	0.31	0	0
(3)	0.23	1	0

Puede clasificarse a los reactivos de esta forma:
a) Por Dificultad: El reactivo más difícil es el (3) y el más fácil es el reactivo (1)
b) Por Discriminación: El reactivo que mejor discrimina es el (2) y el reactivo (1) discrimina mal
c) Por Adivinación Sistemática: Los reactivos (2) y (3) no propician la respuesta por adivinación, el (1) tiene una probabilidad de 0.15 de acierto por azar

En conclusión, y a reserva de conocer los propósitos de la prueba, puede pensarse que los reactivos (2) y (3) son mejores que el reactivo (1).

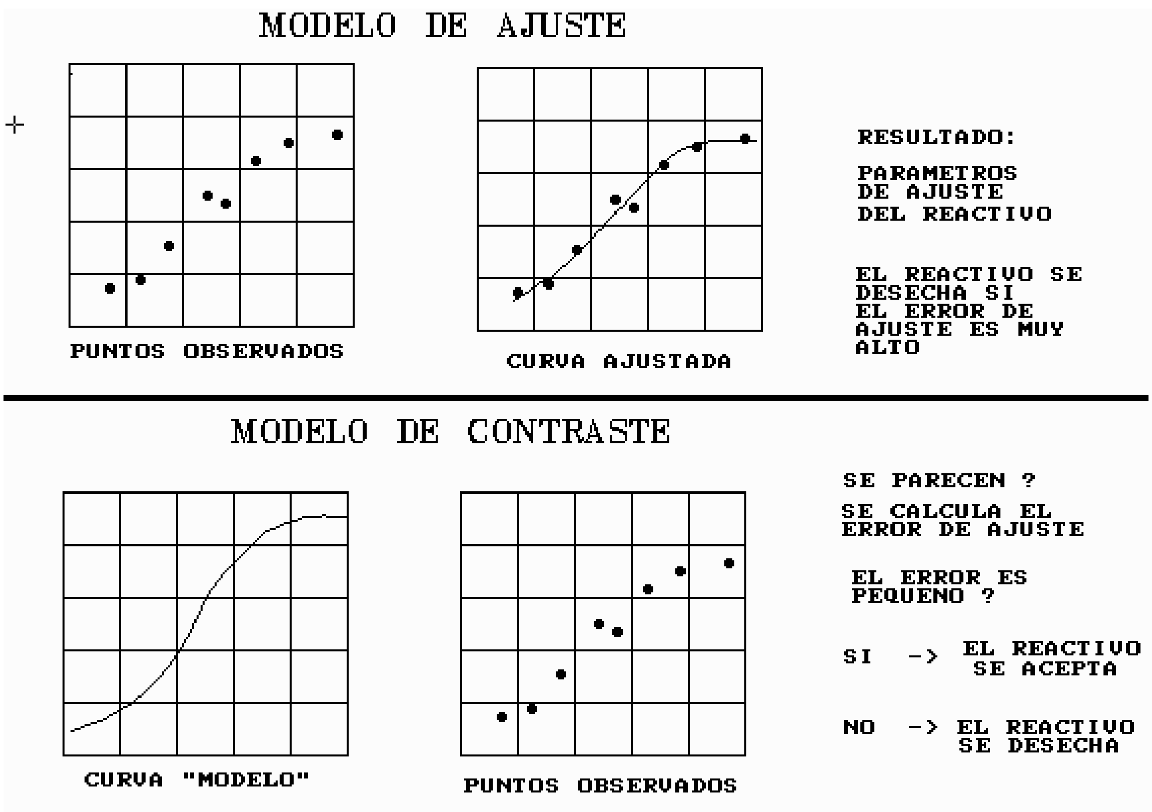
4.2 Modelos para la curva característica.

No hay uno sino varios modelos para aproximarse a la curva característica del reactivo. De estos modelos se deben distinguir dos esquemas básicos:

a) **MODELOS DE AJUSTE.** Modelos que buscan una función que pase lo más cercanamente posible por el conjunto de puntos observados

b) **MODELOS DE CONTRASTE.** Modelos que definen una curva de propiedades y características deseables para los reactivos, contra la cual se contrastan los puntos de la curva.

Es importante ver la diferencia esencial entre ambos modelos porque esto dará una primera luz sobre la calidad de cada uno de ellos.



En el caso de los MODELOS DE AJUSTE se tiene un conjunto de puntos y lo que se pretende es construir una curva que ajuste lo mejor posible a dichos puntos. Es el método que se siguió en el procedimiento aproximado del capítulo 3. Bajo este procedimiento SIEMPRE será posible obtener una curva que “ajuste” con los puntos dados. El ajuste será bueno o malo dependiendo de qué tan alineados estén los puntos, pero siempre podrá obtenerse una curva de ajuste. A este respecto baste con recordar la curva de la Figura 3.10 que se aproxima a unos puntos francamente azarosos.

En este caso resulta imprescindible obtener, además de los parámetros de ajuste, el orden de error (“FIT” en inglés) de la curva de ajuste. Puesto que siempre hay parámetros de ajuste, es tarea del evaluador decidir qué tanto error acepta. La costumbre es emplear un criterio amplio y por regla general se tiende a aceptar que el ajuste es “bueno”, bajo el argumento de que la “evidencia empírica” dicta que así es el comportamiento de los sujetos en el reactivo. En el capítulo 3 vimos tres curvas que se obtuvieron por medio de un ajuste a un polinomio cúbico y se indicó el error en cada caso, inclusive en los puntos azarosos de la figura 3.10 ya mencionada. Este enfoque es el que sigue la Teoría de la Respuesta al Reactivo con el llamado “Modelo de 3 parámetros”. Sólo debe advertirse que esta Teoría no emplea en general una cúbica sino otra función que se presentará más adelante.

Los MODELOS DE CONTRASTE especifican una forma “deseable” para la curva característica de los reactivos. Esta forma deseable es teórica y se basa en hipótesis que establecen los evaluadores tratando de representar el comportamiento “ideal” de un reactivo. Una vez aceptada la curva teórica se trata de ver si los puntos se parecen a ella. Para poder decidir si hay ajuste o no, se calcula un error de aproximación o error de ajuste. En función del error de ajuste se toma la decisión de aceptar el reactivo cuando “se ajusta al modelo” o rechazarlo “cuando no se ajusta al modelo”.

El enfoque (b), como vemos, es más riguroso que el (a) en el sentido de que no busca ajustar una curva a cualquier precio; en cambio trata de averiguar si los puntos obtenidos se ajustan al modelo. Queda a criterio de cada evaluador aceptar o rechazar el modelo teórico. Este enfoque es el que sigue el Análisis de Rasch.
El modelo de curva que usted elija puede ser tan correcto como cualquiera otro. El problema no reside en la función elegida sino en querer buscar “ajuste a cualquier precio”. Un modelo describe la realidad dentro de ciertos límites, por lo que es normal (y hasta razonable) que en varias ocasiones el modelo y la realidad no se parezcan. El investigador que elige un modelo debe conocer sus limitaciones y alcances, de manera de poder saber cuándo aceptar y cuándo rechazar las hipótesis de trabajo. Si el reactivo se ajusta a una curva diseñada cuidadosamente, se podrá concluir algo sobre el reactivo, en función de las hipótesis de trabajo. Si el reactivo no se ajusta entonces no podrá decirse nada sobre el mismo. Dije bien: si el reactivo se ajusta al modelo... **No se trata, por lo tanto, de que el modelo se ajuste al reactivo.**

¡ATENCION! Espero que no cometa el lector el error de obtener conclusiones sobre cosas que no se han dicho aquí. Este es el razonamiento que podemos hacer:

> Premisa 1: Si el error es pequeño se puede decir que la curva ajusta a los puntos
> Premisa 2: Después de construir la curva se encontró que el error es pequeño
> Conclusión: Por lo tanto, la curva ajusta a los puntos

Lo que no es lo mismo que esta conclusión:

> Por lo tanto el reactivo se comporta bien y es aceptable.

Las premisas dadas no establecen nada con relación a la bondad del reactivo ni a su calidad. ¡No es posible llegar a esta conclusión a partir de las premisas dadas! En ningún punto de las premisas se dice: “si en un reactivo se encuentra una curva de ajuste a los puntos observados entonces el reactivo es bueno y aceptable”. Espero que el lector vea la falacia de este razonamiento tan común en el ambiente de le evaluación educativa.

A continuación veremos diversos modelos para determinar la función que represente a la Curva Característica del Reactivo.

4.3 La logística normal. Un modelo de contraste.

Hay varios modelos de contraste, en este libro estudiaremos dos: La logística normal (también llamada ojiva normal) y el Modelo de Rasch. En esta sección estudiaremos solamente el primer modelo para fijar alguna ideas útiles para uso posterior.

Este modelo ajusta una curva normal acumulada a los datos reportados por la aplicación del cuestionario. La curva normal acumulada se denomina “ojiva” (porque se parece al arco ojival de las catedrales góticas) o “logística normal”. En rigor no se está trabajando con una curva normal acumulada ya que no hay una distribución de referencia, pero matemáticamente se trata de la misma ecuación.

No vamos a dar aquí la ecuación de la logística acumulada porque requiere de un poco más de matemáticas y hemos decidido usar las matemáticas lo menos posible. Pero sí podemos ver en forma gráfica la curva (fig. 4.9).

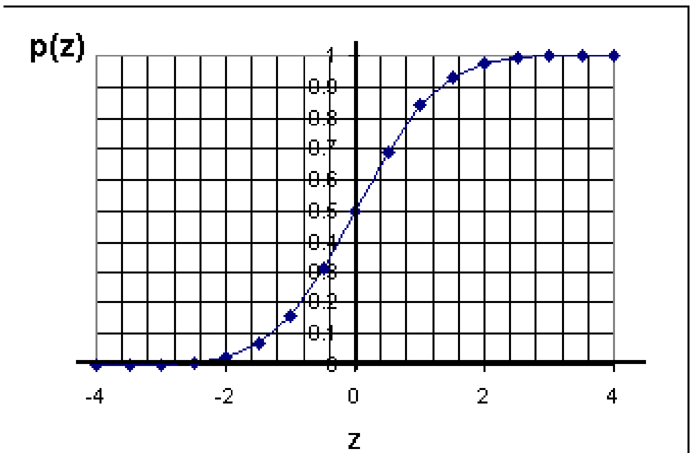


Fig.4.9 La ojiva normal

Esta es una ojiva “perfecta” en el sentido de que se obtiene a partir de los datos de la propia distribución Z que es normal con media 0 y desviación estándar 1. Esta curva tiene dos parámetros: la media y la desviación estándar. Pero es posible ajustar los parámetros característicos: la Adivinación sistemática “c”, la pendiente de la zona central, relacionada con la Discriminación “a” y la Dificultad “b” en el punto de inflexión, que para esta curva siempre estará a la altura p=0.5. Hay que insistir que esta curva tiene su punto de inflexión en p=0.5 (es un valor invariante que corresponde con la media).

Se dibuja la curva a partir de los datos de la probabilidad para un valor de Z dado. Esta probabilidad se encuentra tabulada en cualquier libro de estadística.

La tabla que vamos a construir contiene en una columna los valores de la variable normalizada Z desde -4 hasta +4, en intervalos de 0.5. La segunda columna tiene el área bajo la curva normal (que se encuentra en las tablas de los libros de estadística).

Tabla 4.1

z	Área bajo la curva normal
-4	0.0000001
-3.5	0.0002
-3	0.0013
-2.5	0.0062
-2	0.0228
-1.5	0.0668
-1	0.1587
-0.5	0.3085
0	0.5
0.5	0.6915
1	0.8413
1.5	0.9332
2	0.9772
2.5	0.9938
3	0.9987
3.5	0.9998
4	0.999999

Al trazar la gráfica de Z (eje horizontal) contra p (eje vertical) se obtiene la elegante y bella curva de la figura 4.9.

Ejercicio 4.4

Determinar la logística normal que mejor ajuste a los datos del ejercicio 3.3

Solución:

Se reproduce la tabla con los datos de las columnas (6) y (8) del ejercicio 3.3.

celda	(6) Medida de la puntuación media por decil	(8) Probabilidad de respuesta por decil
1	-2.9944	0.1
2	-1.7346	0.1
3	-1.0986	0.2
4	-0.619	0.4
5	-0.2007	0.4
6	0.2007	0.5
7	0.619	0.5
8	1.0986	0.7
9	1.7346	0.8
10	2.9944	0.8

Se resuelve el problema completando la tabla con dos columnas adicionales, para el valor de la variable normalizada Z y para el área bajo la curva normal en el valor de Z.

La variable Z se calcula con la expresión clásica:

$Z = (X-M)/\sigma$

donde M es la media de las puntuaciones y σ es la desviación estándar.

> Con ayuda de EXCEL se pueden obtener la media y la desviación de los datos de la columna (6) de la tabla. Pídanse las funciones “Promedio” y “DesvEst” con los datos de las celdas de la columna (6).
>
> M=Promedio(celda1..celda10) = 0
>
> σ =desvest(celda1..celda10) = 1.7388139

Los valores de Z se obtienen para los valores dados en la columna (6)

(6) Medida de la puntuación media por decil	(8) Probabilidad de respuesta por decil	(9) Z	(10) Área bajo la curva normal en Z
-2.9944	0.1	-1.7220	0.04252626
-1.7346	0.1	-0.9975	0.15924237
-1.0986	0.2	-0.6318	0.26375546
-0.619	0.4	-0.3559	0.36092416
-0.2007	0.4	-0.1154	0.45405468
0.2007	0.5	0.1154	0.54594532
0.619	0.5	0.3559	0.63907584
1.0986	0.7	0.6318	0.73624454
1.7346	0.8	0.9975	0.84075763
2.9944	0.8	1.7220	0.95747374

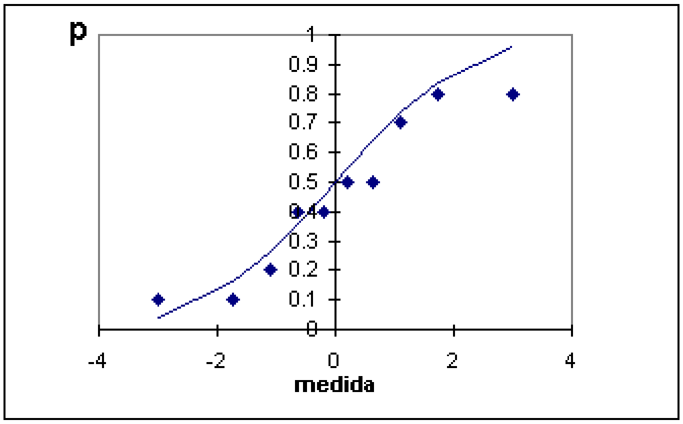
La columna (10) es el área bajo la curva normal hasta el valor Z dado en la columna (9). Este valor se encuentra en cualquier libro de estadística.

> Puede usarse EXCEL para calcular automáticamente la columna (10). En este caso se usa la función “DISTR.NORM.ESTAND(Z)”, donde Z está dado en la columna (9).

Una vez calculados los datos se dibujan con EXCEL dos conjuntos de datos:

a) Columna (6) vs columna (8) - PUNTOS OBSERVADOS

b) Columna (6) vs columna (10) - LOGISTICA NORMAL DE AJUSTE



Se puede decir que la logística normal pasa razonablemente cerca de los puntos observados.

Con lo que hemos hecho hasta ahora corroboramos las observaciones del capítulo 3:

para Z=0, p=0.5 (4.4a)

(a dominio igual a la dificultad del ítem, la probabilidad de respuesta correcta igual a 0.5)

para Z>0, p ε(0.5,1) (4.4b)

(a un dominio grande corresponde una probabilidad de respuesta correcta mayor a 0.5)

para Z<0, p ε(0,0.5) (4.4c)

(a dominio bajo la probabilidad de respuesta correcta es menor que 0.5)

La logística normal se parece a la logística descrita por la medida en lógitos (figura 3.11). De hecho hay un rango en las zonas centrales en que son proporcionales. Por ello se habla siempre de una aproximación entre la logística normal y la logística de la medida. Para poder visualizar esta aproximación tomemos los datos de la tabla 4.1 y agreguemos una columna con los valores de la medida en lógitos correspondientes a p:

tabla 4.2

(1) z	(2) p	(3) medida ln(p/q)
-4	0.000001	-13.8155096
-3.5	0.0002	-8.51699317
-3	0.0013	-6.64409017
-2.5	0.0062	-5.07698669
-2	0.0228	-3.7579308
-1.5	0.0668	-2.63691646
-1	0.1587	-1.66793269
-0.5	0.3085	-0.80714131
0	0.5	0
0.5	0.6915	0.80714131
1	0.8413	1.66793269
1.5	0.9332	2.63691646

2	0.9772	3.7579308
2.5	0.9938	5.07698669
3	0.9987	6.64409017
3.5	0.9998	8.51699317
4	0.9999999	16.1180956

Se puede ver aunque las curvas de las figuras 4.9 y 3.11 se parezcan, difieren en extensión (Z varía de -4 a +4, mientras que la medida fluctúa entre -16 y + 16).

Al revisar la tabla se aprecia que hay cierta "proporcionalidad" entre unos y otros valores, sobre todo en la zona intermedia. Si trazamos los puntos Z y ln(p/q) (columnas (1) y (3)) para Z desde -1.5 a +1.5, veremos que los puntos se alojan en una recta.

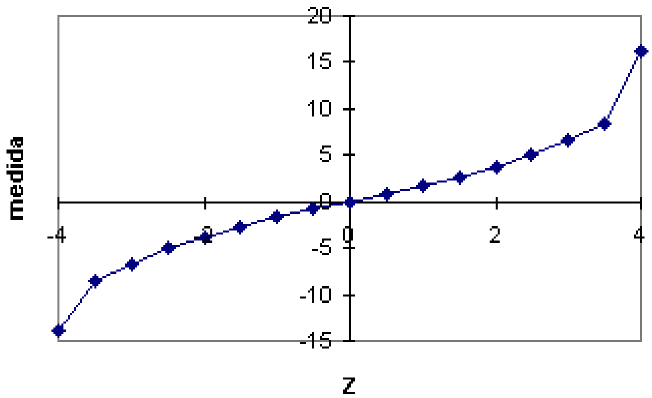


Fig.4.10 Relación Z vs Medida

La ecuación para dicho intervalo es aproximadamente:

$$\ln(p/q) \cong 1.7\,Z \qquad (4.5)$$

Esto quiere decir que la logística es muy aproximadamente 1.7 veces los valores de Z. El error de esta aproximación es del orden de 0.01. Para valores fuera del intervalo de -1.5 a +1.5 la correlación no es tan buena, como puede verse en los extremos de la figura 4.10. El factor 1.7 se convierte en un número misterioso en varios textos de análisis de Rasch y de Teoría de la Respuesta al Ítem, donde no se justifica su origen.

4.4 Modelos de función de ajuste

Dentro de este tipo de aproximación hay una oferta muy amplia de expresiones para las funciones de ajuste. Aquí veremos solamente estos dos tipos de funciones: el modelo de tres parámetros y una curva cúbica.

4.4.1. Ajuste con una logística de tres parámetros.

Ahora vamos a estudiar la fórmula que describe al modelo de tres parámetros.

De acuerdo con este modelo, la curva que describe a la probabilidad de respuesta de un reactivo que mide el rasgo H está dada por la fórmula:

$$p(H)= c + \frac{1-c}{1+e^{-1.7a(H-b)}} \qquad (4.6)$$

Donde a,b y c han sido descritos ampliamente en la sección anterior. En rigor no hay nada que justifique a esta curva. Se trata de un modelo teórico de una curva logística, tan válida como la que usted o yo pudiésemos plantear. Pero es ampliamente aceptada y se dan algunos argumentos matemáticos, estadísticos y hasta psicológicos para justificar que la función es buena y corresponde con la realidad. Desde luego que no es un mal modelo; el valor 1.7 se introduce para hacer que difiera poco de la logística normal. La figura 4.11 muestra varias logísticas ajustadas con este modelo.

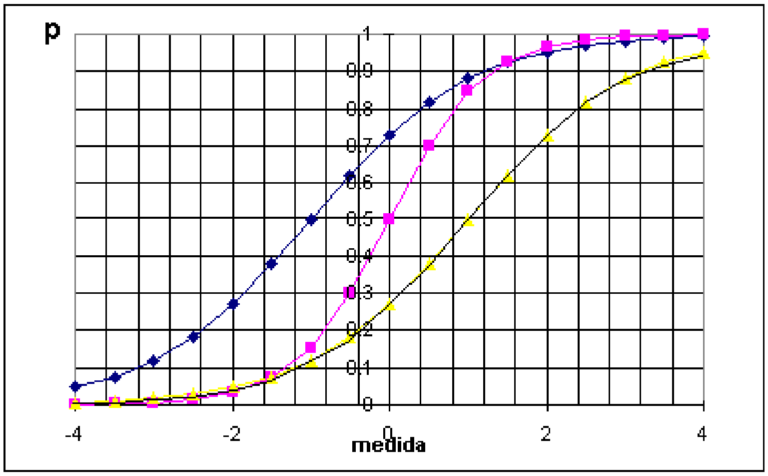


Fig. 4.11 Curvas características, modelo de tres parámetros

Por diversos métodos y aproximaciones numéricas se puede obtener el conjunto de los tres parámetros para la curva de un reactivo. En términos generales puede afirmarse que SIEMPRE se encuentran los valores a, b y c. Una vez determinada la curva de ajuste debe obtenerse su error o valor de ajuste (el "FIT" de la curva).

Ejercicio 4.5

Determinar la función del modelo de tres parámetros que mejor ajuste a los datos del ejercicio 3.3

Solución:

Véase la tabla 3.3 o su forma reproducida en el ejercicio 4.4

Complétese la tabla con una columna donde se calcula p(H) de acuerdo con la expresión (4.6). No calcularemos el error de ajuste ("FIT"), pues seguiremos un procedimiento intuitivo, asignando valores a los parámetros "a", "b" y "c" hasta encontrar una curva que ajuste razonablemente a los datos. El procedimiento que se sigue en la Teoría de la Respuesta al Reactivo es diferente y más preciso.

(6) Medida de la	(8) Probabilidad	(9) p(H)

puntuación media por decil	de respuesta por decil	(fórmula 4.6)
-2.9944	0.1	0.0676
-1.7346	0.1	0.1169
-1.0986	0.2	0.1761
-0.619	0.4	0.2460
-0.2007	0.4	0.3276
0.2007	0.5	0.4218
0.619	0.5	0.5299
1.0986	0.7	0.6526
1.7346	0.8	0.7890
2.9944	0.8	0.9370

Los valores de la columna (9) de esta tabla se obtuvieron para los parámetros

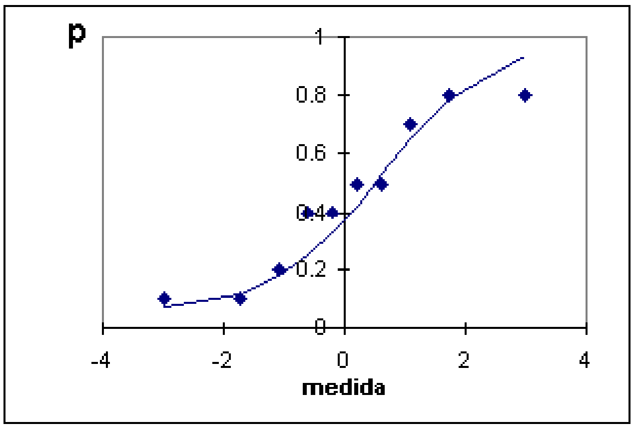
a= 0.65, b= 0.6 , c= 0.05

Una vez calculados los datos se dibujan con EXCEL dos conjuntos de datos:

a) Columna (6) vs columna (8) - PUNTOS OBSERVADOS

b) Columna (6) vs columna (9) - CURVA DE AJUSTE

La curva de ajuste que se encuentra con estos valores es la siguiente:



De nueva cuenta vemos que la curva pasa razonablemente bien por los puntos observados. El lector podrá idear otros valores de los tres parámetros hasta encontrar otra curva más precisa o de mejor aspecto.

Ejercicio 4.6

Con ayuda de una tabla de EXCEL trazar la curva del modelo de tres parámetros para:

a=0.5, b= 1.0 y c=0.18

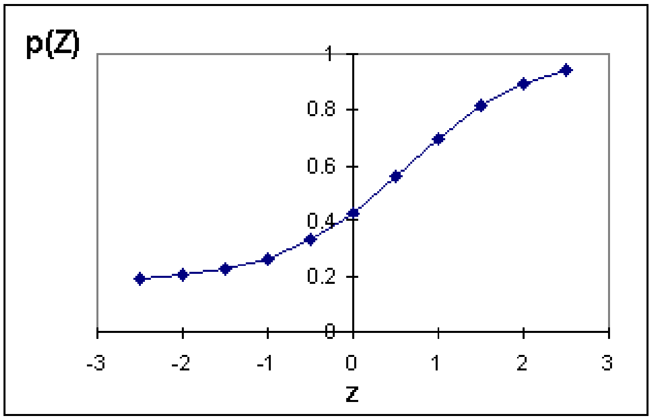
Solución

Se debe construir esta tabla:

Z	p(Z) (fórmula 4.6)

-2.5	0.19192521
-2	0.20321018
-1.5	0.22458504
-1	0.26357973
-0.5	0.33007875
0	0.43142299
0.5	0.56216289
1	0.69884879
1.5	0.8136687
2	0.89368173
2.5	0.94245627

donde p(Z) se calcula con la fórmula 4.6, con los parámetros indicados. Una vez dispuesta la tabla se obtiene esta curva con ayuda de EXCEL:



Ejercicio 4.7

Determinar la probabilidad de respuesta de tres sujetos cuyas medidas son -0.4, 0.5 y 1.6 lógitos, para el reactivo del ejercicio 4.5

<u>*Solución:*</u>

La expresión que brinda la probabilidad está dada por la fórmula (4.6) a la que se introducen los parámetros solicitados.

$$p(H)= 0.18 + \frac{1-0.18}{1+e^{-1.7x0.5(H-1.0)}}$$

Con ayuda del mismo EXCEL o con una calculadora se obtienen los valores al substituir H por las medidas proporcionadas:

p(-0.4) = 0.371
p(0.5) = 0.504
p(1.6) = 0.692

Obsérvese la conveniencia de disponer de una función como (4.6) que permite dibujar la curva y obtener valores de probabilidad de respuesta correcta, así como diversos estimadores. Esto es mucho mejor que los métodos intuitivos utilizados en los ejemplos 4.1 y 4.2 resueltos anteriormente.

4.4.2. Ajuste con una cúbica.

En algunas ocasiones los evaluadores se sienten mal porque la logística francamente no aproxima a los puntos experimentales, como ocurrió con los puntos mostrados en la figura 3.10. Decir que el ajuste es malo (“low fit”) obliga al evaluador a rechazar al reactivo. A muchos evaluadores esto les produce serios problemas, porque no están dispuestos a desechar ninguno de sus reactivos.

Es por ello que se da la preocupación de disponer de otro modelo que ajuste a los puntos cuando el modelo de tres parámetros falla. Este proceder puede parecernos poco científico, pero debemos ser conscientes de que es algo que ocurre en la práctica profesional y, por ello, vale la pena estar advertidos para no dejarnos llevar por la corriente.

Recientemente se ha buscado ajustar un polinomio de tercer grado a los puntos, porque “se ha visto” que ajustan mejor a los datos. Independientemente de esta afirmación, ¿qué razón puede haber para elegir una cúbica?. Una primera razón es que, en efecto, una cúbica se parece mucho al comportamiento descrito por la curva característica del reactivo (recuérdense las gráficas mostradas en el capítulo 2, trazadas con ayuda de EXCEL que obtiene directamente una cúbica de ajuste). Pero no quiero partir del hecho de que “la evidencia empírica” dicta que la cúbica puede ser un buen modelo. No creo en la evidencia empírica si no está soportada por un modelo. Hay otra razón mucho más sensata y de origen matemático.

Es posible que algunos de los lectores estén familiarizados con la Serie de Taylor, pero para los que no lo estén se hará una breve presentación. Baste con saber que toda función continua puede desarrollarse en Serie de Taylor, lo cual quiere decir que dicha función se transforma en un polinomio de X, es decir, la Serie de Taylor transforma a una función trascendente en otra de tipo algebraica. El lector interesado en detalles sobre esta herramienta deberá consultar un libro especializado de cálculo diferencial.

Recordemos, del primer modelo que hemos hecho hasta ahora, que la curva que relaciona la medida con la probabilidad es:

$$B=\ln(p/q)$$

Lo interesante en este momento es que la función logaritmo que aparece en el segundo miembro de la medida B, tiene un equivalente con una función polinómica. En efecto. Si desarrollamos ln(p/q) con ayuda de la Serie de Taylor (o de la Serie de MacLaurin que para el caso brinda el mismo resultado), se tiene que:

$$\ln(p/q) = 2\,(X + X^3/3 + X^5/5 + X^7/7 + ...+ X^{(2n-1)}/(2n-1)+...) \quad (4.7)$$

donde la variable $X = 2p-1$

o bien: $p=(X+1)/2$

La serie de Taylor indica que la función logaritmo es igual a la serie de potencias de X dada en el segundo miembro. En estos desarrollos se puede cortar la serie en el término que se desee y se tendrá lo que se llama “error de truncamiento” que es más o menos grande dependiendo del valor de la variable X en este caso.

No es motivo de este libro discutir a detalle las implicaciones de esta equivalencia, pero es suficiente conocer esta interesante propiedad:

$\ln(p/q) \cong 2X + 2x^3/3$ (4.8)

(Léase: el logaritmo natural de p/q es aproximadamente igual a la función $2X + 2x^3/3$)

con X=2p-1

Esta fórmula brinda un error del orden de $2X^5/5$

La Serie de Taylor dice que hay una cúbica que aproxima razonablemente a la función logaritmo, con un error mayor que $2\ X^5/5$ ¿Qué tan grande es este error? No muy grande y lo podemos estudiar rápidamente.

Si X es pequeño, pongamos por caso X=0.1 (lo cual ocurre para p=(X+1)/2=0.505) , el error que se comete es del orden de $2(0.1)^5/5$ (esto es aproximadamente 0.000004), que es un número muy pequeño realmente y la diferencia entre el logaritmo y la cúbica es despreciable.

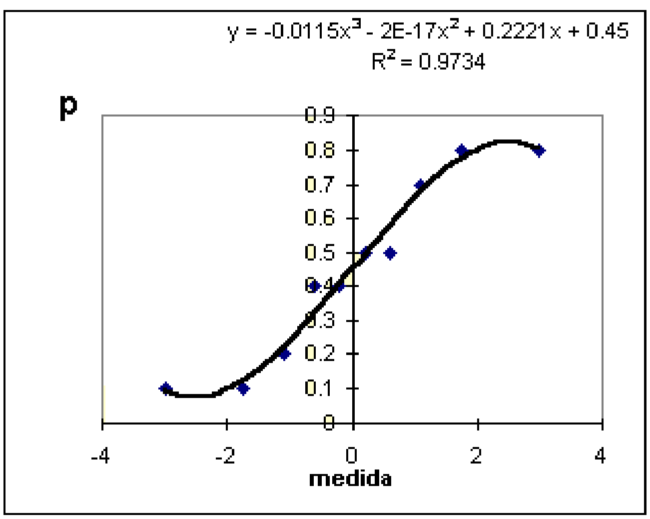
La justificación matemática queda ahí. Una cúbica es una buena aproximación a la logística. Claro que, según vemos, no cualquier cúbica ni bajo cualquier circunstancia. La cúbica que proporciona EXCEL como aproximación a los datos no es necesariamente la mejor cúbica que se parezca al logaritmo del momio. Hay muchas cosas que no trataré en este capítulo respecto a la cúbica y a los otros ajustes vistos en esta sección pero, para los fines de este libro, baste con saber que puede obtenerse una muy buena aproximación a la curva característica de un reactivo con ayuda de una cúbica.

Ejercicio 4.8

Determinar la curva cúbica que mejor ajuste a los datos del ejercicio 3.3

Solución:

Tómense los datos de la tabla 3.3 o como se reprodujeron en el ejercicio 4.4. Aquí bastará con trazar los puntos de las columnas (6) y (8) y pedir a EXCEL que encuentre el polinomio de grado 3 que mejor ajuste a los puntos. La curva que se obtiene con EXCEL es la siguiente.



Puede decirse que este polinomio pasa razonablemente cerca de los puntos observados.

Otras curvas no tratadas en esta sección pueden ser encontradas en libros especializados. Se trata del modelo de dos parámetros, las curvas Spline, B-Spline, Bézier, etc.

De nueva cuenta hay que insistir que el hecho de que podamos obtener una curva que aproxime a los puntos no es garantía de que el reactivo se va a comportar bien.

4.5 El Modelo de Rasch. Un modelo de contraste.

En el Modelo de Contraste hemos dicho que se hace la hipótesis de que los reactivos se "deben" comportar como establece el modelo para poderlos aceptar. Si no se ajustan al modelo los desechamos. Esto es porque hacemos una hipótesis de "como deben ser las curvas típicas de los reactivos de buena calidad".

Hay razones para decir cuál es la curva deseable. Muchos evaluadores discuten estas razones a favor y en contra, pero básicamente se trata de justificaciones de carácter lógico, matemático y de tendencias filosóficas respecto a la evaluación educativa. En este momento podría yo decir al lector que puede tomar o rechazar las hipótesis del modelo de contraste en el terreno de la tendencia evaluativa, pero se requiere de un gran esfuerzo en el terreno matemático para poder rechazar el planteamiento de estos modelos.

Los modelos de contraste no buscan ajustar una curva a los puntos de manera indiscriminada. Si los puntos se ajusta al modelo ¡qué bueno! porque el reactivo se comportará como establece dicho modelo y conoceremos todas sus propiedades sin lugar a dudas. En cambio, si los puntos no se ajustan a la curva teórica, entonces se debe rechazar el reactivo porque no se comporta como dice el modelo. No necesariamente el comportamiento que dice el modelo es el que nos pueda gustar, pero es un modelo que finca algunas limitaciones y alcances. Es claro que bajo este enfoque habrá muchos reactivos que "no ajusten" al modelo, lo cual implicará que habrá reactivos rechazados. Esto puede no gustarle a los evaluadores que quisieran que sus exámenes salieran evaluados lo mejor posible. Tal vez por ello este enfoque es tan discutido y tiene tantos detractores.

Hay varios modelos, ya vimos el de la logística normal en la sección 4.3 y ahora haremos referencia es el modelo de Rasch, cuya expresión es:

$$p(H) = \frac{1}{1+e^{-(H-b)}} \qquad (4.9)$$

Observamos que este modelo es muy parecido al de la expresión 4.4 del modelo de tres parámetros, cuando c=0 y cuando a=1/1.7

El modelo de Rasch establece que solamente puede obtenerse b a partir de la información disponible. La curva de Rasch es una logística que tiene una pendiente única y que se dibuja trasladándose sobre el eje horizontal. El parámetro b corresponde con la medida de la persona o la calibración del reactivo y se obtiene para el punto de inflexión que para está curva siempre está en p=0.5.

Si representamos la curva de Rasch para diversos valores de b se tiene una gráfica como la de la figura.

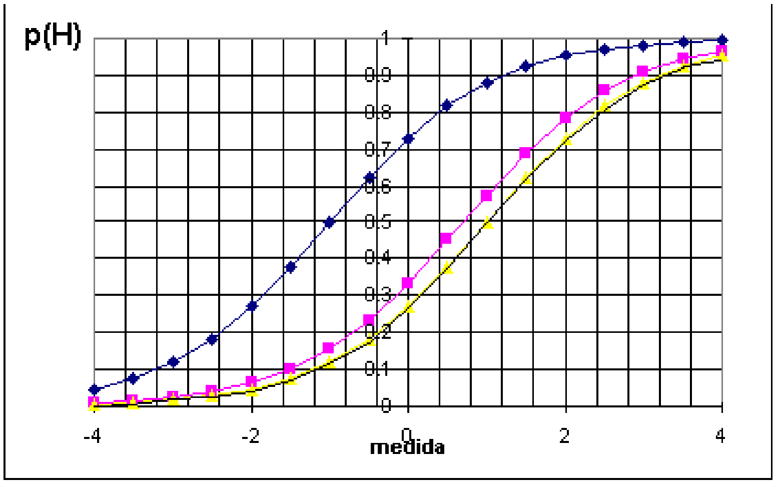


Fig.4.12 Modelo de curvas de Rasch

Este modelo cancela, como podemos ver, el uso de estos parámetros

a) *Adivinación sistemática.* Para el modelo de Rasch puede haber adivinación individual, pero no hay adivinación sistemática. Esto coincide con lo que comentamos en la sección 4.1.
b) *Discriminación* relacionada con la inclinación de la logística. Todas las curvas tiene la misma inclinación en la zona central, por lo que se dice que en el modelo de Rasch todos los reactivos discriminan igual.

Comentaremos acerca de estos aspectos en el capítulo 6 donde se deduce el modelo de Rasch y se presentan sus implicaciones más a detalle.

Ejercicio 4.9

Determinar la logística normal que mejor ajuste a los datos del ejercicio 3.3

Solución:

Tómense los datos de la tabla 3.3 o su presentación en el ejercicio 4.4

Se completa la tabla con una columna donde se calcula p(H) de acuerdo con la expresión (4.9). El cálculo que haremos aquí será intuitivo, asignando valores al parámetro "b" hasta encontrar una curva que ajuste lo mejor posible a los datos.

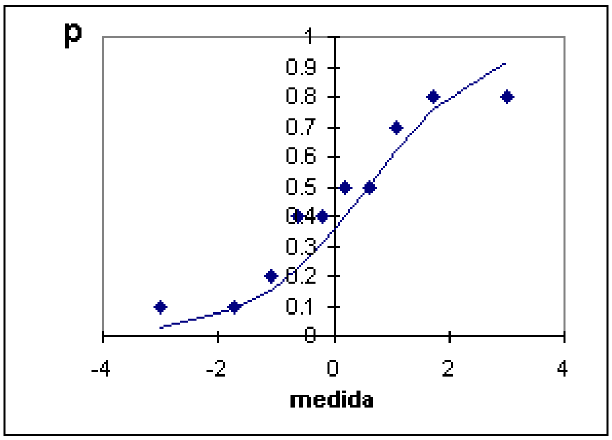
El procedimiento establecido en los libros de la Teoría de Rasch es diferente, mucho más preciso, pero para los fines didácticos de este libro podemos aceptar que el procedimiento intuitivo es suficiente.

(6) Medida de la puntuación media por decil	(8) Probabilidad de respuesta por decil	(9) p(H) (fórmula 4.9)
-2.9944	0.1	0.0267
-1.7346	0.1	0.0883
-1.0986	0.2	0.1546
-0.619	0.4	0.2281
-0.2007	0.4	0.3099
0.2007	0.5	0.4015
0.619	0.5	0.5047
1.0986	0.7	0.6221
1.7346	0.8	0.7567
2.9944	0.8	0.9164

Una vez calculados los datos se dibujan con EXCEL dos conjuntos de datos:

a) Columna (6) vs columna (8) - PUNTOS OBSERVADOS

b) Columna (6) vs columna (9) - CURVA DEL MODELO DE RASCH



Vemos ahora que los puntos pasan muy cerca de la curva de Rasch (en cambio en los problemas anteriores se decía que las curvas pasaban muy cerca de los puntos).

El parámetro elegido para calcular los puntos de la curva de Rasch es b=0.6. El lector podrá utilizar la misma tabla de EXCEL y buscar otros valores de b para los cuales los puntos observados estén más cercanos a la curva.

Ejercicio 4.10

Determinar la probabilidad de respuesta de tres sujetos cuyas medidas son -0.4, 0.5 y 1.6 lógitos, para el reactivo del ejercicio 4.5, usando el Modelo de Rasch.

Solución:

Se usa la fórmula dada por (4.9), para la Dificultad = 1.0

p(H) = $\frac{1}{1+e^{-(H-1.0)}}$

Con ayuda de la calculadora o de una tabla de EXCEL se encuentran estos resultados:

p(-0.4)=0.198

p(0.5)=0.378

p(1.6)=0.646

Obsérvese la diferencia de estimación respecto a los valores obtenidos con el modelo de 3 parámetros en el ejercicio 4.7.

Ejercicio 4.11

Con ayuda de una tabla de EXCEL trazar la curva del Modelo de Rasch para b=1.5

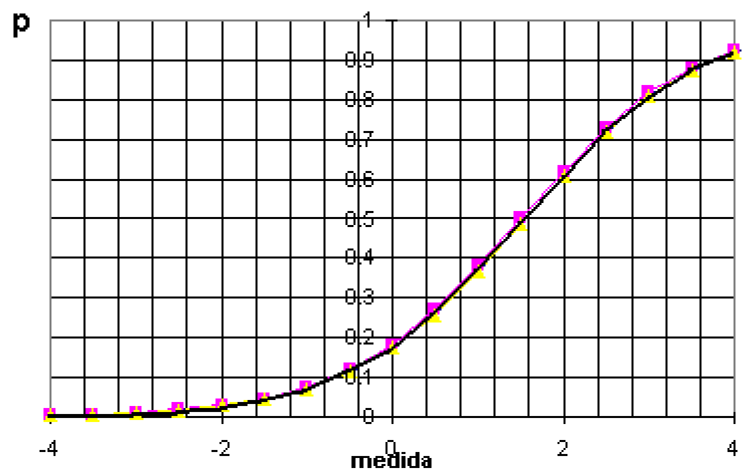
Solución:

Se prepara esta tabla:

H	p(H) (fórmula 4.9)
-4	0.004070138
-3.5	0.006692851
-3	0.010986943
-2.5	0.01798621
-2	0.029312231
-1.5	0.047425873
-1	0.07585818
-0.5	0.119202922
0	0.182425524
0.5	0.268941421
1	0.377540669
1.5	0.5
2	0.622459331
2.5	0.731058579
3	0.817574476
3.5	0.880797078
4	0.92414182

donde H son valores de -4 a +4 lógitos y p(H) se obtiene de la fórmula (4.9)

Una vez arreglada la tabla, se dibuja la curva con ayuda de EXCEL:



Conviene insistir que en el Modelo de Rasch, el punto de inflexión siempre corresponde con p=0.5 , en el valor de la calibración del reactivo, en este caso b=1.5

4.6 Comentarios finales. Una paradoja

Hemos visto que a un conjunto dado de respuestas puede darse diferente tratamiento, según sea el modelo que se desea seguir. Se han presentado los modelos de una forma muy simplificada y se hizo la construcción aproximada de los cuatro modelos con los datos del reactivo del ejemplo 3.3 (Ejercicios 4.4, 4.5, 4.8 y 4.9).

Conviene observar que las curvas de los diferentes modelos se parecen, pero no coinciden necesariamente. En general no son iguales ni las curvas ni los modelos. Cada modelo se enfoca a diferentes aspectos. Recordemos qué buscan todos estos modelos:

Disponer de una función (una Curva Característica del Reactivo) que permita estimar la probabilidad de respuesta de un sujeto al enfrentarse a un reactivo dado.

La tabla 4.3 muestra algunas diferencias y coincidencias entre modelos.

Tabla 4.3

Modelo	Punto de inflexión	Dificultad (b)	Discriminación (a)	Adivinación sistemática (c)	Otros aspectos
	Lugar donde cambia la concavidad de la curva. p=(1+c)/2	Medida en el punto de inflexión de la curva	Relacionada con la pendiente en el punto de inflexión	Ordenada de la curva para la medida más baja de los sujetos	
Tres parámetros	Diferente para cada reactivo	Diferente para cada reactivo	Diferente para cada reactivo	Diferente para cada reactivo	(fórmula 4.6) Generalmente ajusta en la mayoría de reactivos de un examen
Curva cúbica	No se considera. Diferente para cada reactivo	No se considera. Diferente para cada reactivo	No se considera. Diferente para cada reactivo	No se considera. Diferente para cada reactivo	SIEMPRE habrá una cúbica que ajuste a cada reactivo de un examen
Logística normal	Ubicado en p=0.5 en todos los reactivos	Función de media de respuestas del cuestionario	Función de media y desviación estándar de puntuaciones. Todos los reactivos del examen tienen la misma pendiente	c=0 en todos los reactivos	Curva única para un examen y una población. La curva cambia para otros exámenes o poblaciones. No siempre ajusta a los reactivos de un examen
Rasch	Ubicado en p=0.5 en todos los reactivos	Diferente para cada reactivo. Independiente de los sujetos	a=1/1.7 en todos los reactivos de cualquier examen	c=0 en todos los reactivos	(fórmula 4.9) No siempre ajustan los reactivos al modelo

El modelo de tres parámetros usa directamente los valores de “a”, “b” y “c”, para encontrar una curva que mejor aproxime a los puntos observados. El procedimiento riguroso no es como el que se siguió aquí de tipo intuitivo, ya que debe reportar no solamente la curva sino también el error de ajuste (FIT). El ajuste generalmente es bueno en la mayoría de los reactivos de un cuestionario. El modelo busca reportar una curva que sea independiente de los sujetos y de los reactivos aplicados, lo cual no es rigurosamente cierto para todos los reactivos de un cuestionario.

El ajuste con una cúbica trata de obtener una curva que pase lo más cercanamente por los puntos observados. Tal vez sea el modelo que mejor se vea en comparación con la evidencia empírica y se encontrará SIEMPRE una curva que ajuste a los puntos observados en todos los reactivos de un cuestionario.

El modelo de la logística normal usa dos parámetros (media y desviación estándar de las respuestas) para ajustar una curva tipo ojiva normal acumulada, correspondiente al grupo de sustentantes en particular. En este modelo todos los reactivos tendrán un mismo aspecto para un conjunto dado de sujetos. La curva depende de la población y de los reactivos aplicados.

El modelo de Rasch dispone de una curva tipo, que solamente "está autorizada" a trasladarse horizontalmente, hasta encontrar la dificultad del reactivo. Se buscará la mejor posición de acuerdo con los puntos observados. En algunos casos la curva y los puntos pueden verse muy cercanos entre sí (tendrían un buen ajuste o FIT), pero esto no necesariamente ocurre con todos reactivos de un cuestionario, por lo que el modelo nos reportará un valor inconveniente del error de ajuste.

Para terminar este capítulo, haremos referencia a una situación paradójica que se presenta en los reactivos cuando se trabaja con la Curva Característica del Reactivo, en especial bajo el modelo de 3 parámetros. Esta situación paradójica ocurre cuando un reactivo que discrimina de manera positiva, se comporta así:

a) ***Difícil*** para los sujetos de **<u>mejor</u>** desempeño	$\Rightarrow$	Los sujetos de mejor desempeño posiblemente no respondan al reactivo
b) ***Fácil*** para los sujetos de **<u>bajo</u>** desempeño.	$\Rightarrow$	Los sujetos de bajo desempeño tienen más posibilidad de acertar

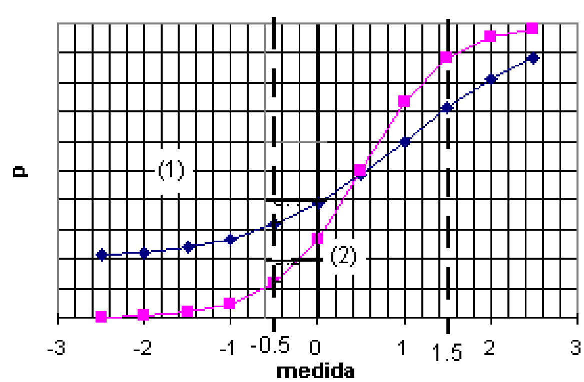
Veamos un ejemplo.

Ejercicio 4.12

Se tienen dos reactivos cuyas CCR se muestran en la figura. Sus parámetros son:

Curva (1): a=0.7, b=1.0 , c=0.2

Curva (2): a= 1.2 , b= 0.5, c=0.0



Clasifique y dictamine a los reactivos. Determine la probabilidad de respuesta en ambos casos para dos sujetos de medida -0.5 y 1.5 y explique por qué se tiene una situación paradójica.

Solución:

a) Clasificación y dictamen de los reactivos

A partir de los datos puede hacerse esta tabla:

Reactivo	Clasificación en Dificultad	Discriminación	Adivinación sistemática
(1)	Reactivo Difícil (b=1.0)	Baja (a=0.7)	Propicia en un 20% (c=0.2)
(2)	Reactivo Fácil (b=0.5)	Alta (a=1.2)	No tiene (c=0.0)

b) Con ayuda de una tabla de EXCEL se obtienen los valores pedidos:

B Medida del sujeto	p(B) Curva (1)	p(B) Curva (2)
-0.5	0.31494947	0.11506673
1.5	0.71560924	0.88493327

c) Se tiene una situación paradójica, porque la Dificultad de cada reactivo es relativa a la posición de la medida del sujeto, en lugar de ser intrínseca del reactivo.

En efecto, dado un reactivo de dificultad D y dos personas de medida B1 y B2, tales que B1>B2, se debe cumplir:

p(B1,D) > p(B2,D) (condición 1)

(Una persona de mayor dominio debe contestar con mayor probabilidad a un mismo reactivo)

y, en forma complementaria, dada una persona de medida B y dos reactivos de Dificultades D1 y D2, tales que D1>D2, se debe cumplir:

p(B,D1) < p(B,D2) (condición 2)

(Un reactivo de mayor dificultad debe tener menor probabilidad de respuesta correcta para una misma persona).

La condición 1 sí se cumple (esto es obligatorio en un reactivo de pendiente positiva), pero curiosamente la condición 2 no se verifica a lo largo de toda la curva.

B Medida de la persona	p(B) Curva (1) DIFICIL	p(B) Curva (2) FACIL	Condición entre p(B,D1) y p(B,D2)	¿Cumple la condición 2 ?
-0.5	0.31494947	0.11506673	>	NO
1.5	0.71560924	0.88493327	<	SI

De acuerdo con estos datos, hay una parte de las CCR donde la condición sí se cumple y otra donde la condición no se verifica. Nótese que el comportamiento es correcto hacia la derecha de la intersección de las dos CCR, mientras que es absurdo a la izquierda de dicho punto. El punto de intersección se encuentra aproximadamente en una medida B=0.5.

Clasifiquemos a los alumnos en “Grupo alto”, por arriba del punto de intersección y en “Grupo bajo” por abajo del mismo punto. Observamos que a la derecha del punto de intersección una persona de medida B=1.5 en el reactivo (1) tiene menos probabilidad de respuesta correcta que en el (2), esto es lógico, ya que el reactivo (1) es más difícil.

Sin embargo, ¿cómo es posible que por abajo de la intersección, para B=0.5, el reactivo difícil (1) obtenga mayor número de respuestas posibles en los sujetos del grupo bajo que el reactivo fácil (2)? ¡ Curiosamente, a los alumnos del grupo bajo el reactivo “difícil” les resulta fácil y el “fácil” se les hace más difícil !

Esta es la situación paradójica a que se hace referencia. Esta paradoja no ocurre con el modelo de Rasch, ya que las curvas nunca se cortan, en cambio en el modelo de 3 parámetros y los otros modelos de ajuste las intersecciones entre curvas sí ocurren y hacen que la interpretación de los reactivos se vuelva muy turbia e inconsistente. Este es uno de los argumentos que esgrimen los seguidores del modelo de Rasch para desechar el de 3 parámetros, mientras que los que emplean este último modelo asumen que están reportando la evidencia empírica.

4.7 Ejercicios propuestos

1. Con ayuda de una tabla de EXCEL trazar la curva del modelo de tres parámetros para: a=0.8, b=-0.6, c= 0.23

2. Con ayuda de una tabla de EXCEL trazar las curvas del modelo de tres parámetros y del modelo de Rasch para estos datos:

Curva 1: a= 1.4 , b=1.0, c=0.0

Curva 2: a= 0.6 , b=0 , c=0.2

Curva 3: a= 1.1 , b= -1.0 , c= 0.0

Compárense las curvas obtenidas en cada caso.

3. Determinar el valor de la probabilidad de respuesta de las personas que tienen medidas B=1.0 y B=-1.0 a los reactivos de los ejercicios propuestos 1 y 2.

4. ¿Por qué se dice que un reactivo de discriminación negativa se comporta contrariamente a lo esperado?

5. Catalogar los reactivos de los ejercicios propuestos 1 y 2 en cuanto a dificultad, discriminación y adivinación sistemática. Hacer un dictamen siguiendo el procedimientos seguido en el ejercicio 4.4

6. Usar una tabla de EXCEL para dibujar la curva cúbica de acuerdo con la fórmula dada en (4.8) y comparar contra una logística, en los datos de los ejercicios 1 y 2.

7. Dados los dos reactivos del ejercicio 4.12, dibujar las curvas del modelo de Rasch y verificar que no se intersectan, teniéndose siempre una interpretación consistente de la dificultad del reactivo. Verificar que en este caso no se presenta la paradoja de los reactivos que se intersectan.

8. Construya un par de reactivos donde se presente la paradoja de la dificultad, explique de nuevo el comportamiento de los reactivos y explique por qué se considera que este comportamiento paradójico no es razonable.

9. Verifique con ayuda de una tabla de EXCEL que las curvas del modelo de 3 parámetros y del modelo de Rasch pueden ser iguales, bajo ciertos valores de los parámetros a, b y c.

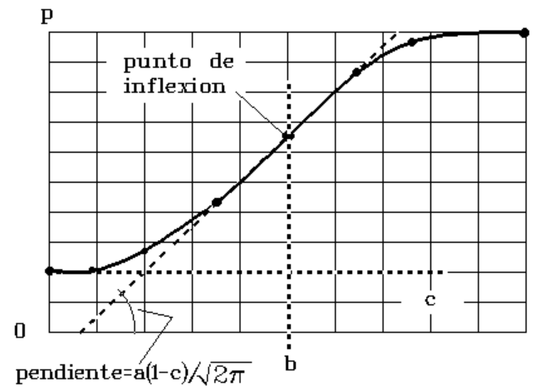
10. Se presentan los conteos de respuestas correctas por decil de un grupo de sujetos ante 3 reactivos.

Decil	Número de sujetos	Límite inferior	Límite superior	Proporción de respuestas correctas en el decil		
				Reactivo 1	Reactivo 2	Reactivo 3
1	20	0	.13	0.00	0.30	0.30
2	20	.13	.17	0.00	0.20	0.40
3	20	.17	.25	0.20	0.15	0.45
4	20	.25	.34	0.25	0.50	0.60
5	20	.34	.46	0.45	0.70	0.60
6	20	.46	.48	0.60	0.60	0.65
7	20	.48	.57	0.75	0.40	0.80
8	20	.57	.68	0.70	0.65	0.75
9	20	.68	.74	0.75	0.80	0.95
10	20	.74	.88	0.90	0.40	0.90

a) Calcule las curvas características de los reactivos siguiendo los procedimientos descritos en este capítulo, para los cuatro modelos presentados: logística normal, Modelo de Rasch, modelo de tres parámetros, cúbica proporcionada por el programa EXCEL.
b) Identifique cual modelo se ajusta mejor a los datos
c) Identifique cual reactivo se comporta mejor en relación a los que dictan los modelos de contraste
d) Clasifique a los reactivos en cuanto a dificultad, discriminación y adivinación sistemática

4.8 Puntos clave del capítulo

1. Parámetros y elementos de la Curva Característica del Reactivo



1.1 Punto de inflexión. Punto donde cambia de concavidad la CCR.

Punto de inflexión = (c+1)/2

1.2 Dificultad (b). Medida en el punto de inflexión.

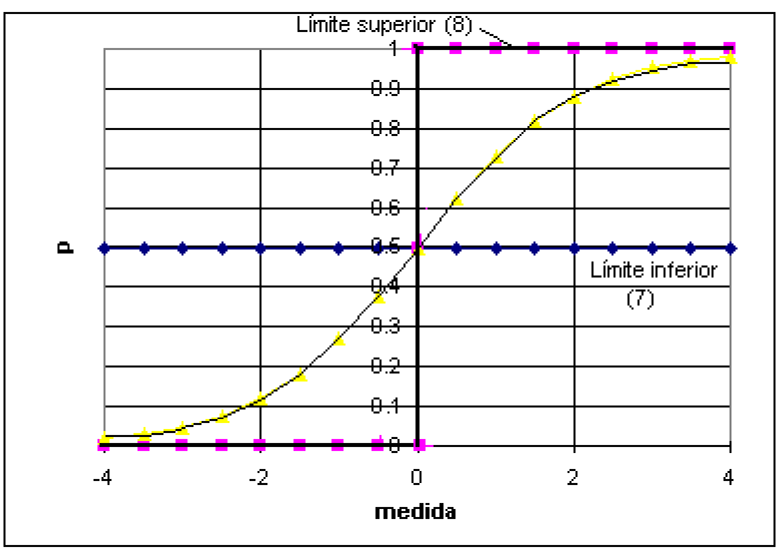
1.3 Extensión. Ámbito de la parte recta central de la CCR

1.4 Discriminación (a). Se define por “Discriminación” a un valor que permite distinguir a los alumnos de alto desempeño de los de bajo desempeño y es un función de la pendiente en el punto de inflexión de la Curva Característica del Reactivo.

a= Pendiente $\sqrt{2\pi}$ /(1-c)

1.5 Adivinación sistemática (c). Se denomina “adivinación sistemática” al valor de probabilidad de acierto para las personas de medida más baja. También puede decirse que es la probabilidad mínima esperada de respuesta en un reactivo, se encuentra en la ordenada para el valor de medida más bajo esperado.

2. Límites de la CCR. Toda CCR, para una Dificultad dada, tiene un límite superior de máxima discriminación y un límite inferior de discriminación nula.



3. Modelos de CCR.

3.1 MODELOS DE AJUSTE. Modelos que ajustan una función al conjunto de puntos de la curva característica

- ojiva normal
- modelo de tres parámetros
- curva cúbica
- Splines, B-Splines, Bézier, etc.

3.2 MODELOS DE CONTRASTE. Modelos que definen una curva de propiedades y características deseables para los reactivos, contra la cual se contrastan los puntos de la curva.

- modelo de Rasch
- modelos de 2 parámetros

3.3 Ajuste de un modelo. Proceso de comparar los puntos experimentales contra las curvas teóricas. Requieren de un cálculo de parámetros y del valor de error de ajuste (FIT)

4. Modelo de 3 parámetros. Usa la fórmula:

$$p(H)= c + \frac{1-c}{1+e^{-1.7a(H-b)}}$$

5. Modelo de Rasch. Se define con la fórmula:

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

5. EL ESCALOGRAMA DE GUTTMAN Y EL PROBLEMA DEL ERROR EN LA ESCALA

5.1 Algunos aspectos relativos a la escala.

Los modelos que hemos visto tienen un supuesto inherente: tratan de medir una sola variable. Se dice que son modelos unidimensionales o que la escala y la medición hacen referencia a una sola dimensión.

Cuando hablamos de dimensión queremos decir que el reactivo o conjunto de reactivos que forman una prueba, hacen referencia a un rasgo exclusivamente: Habilidad Verbal, por ejemplo. La cosa se complica cuando en un cuestionario tenemos mezcla de varios rasgos o áreas de conocimiento (habilidad verbal, matemáticas, civismo, inglés y habilidades para la cocina). En este caso algunos reactivos no necesariamente van a correlacionar con todo el cuestionario, ni es obligatorio que unos reactivos correlacionen con otros.

Pero si aceptamos que estamos midiendo un solo rasgo, entonces un reactivo mide una sola cosa que es factible de ubicarse en la escala. Así podemos dibujar la escala de medida (entre -4 y +4 por ejemplo) y aceptar que el resultado obtenido con ayuda de un reactivo nos permite ubicar una "raya" de la escala (figura 5.1).

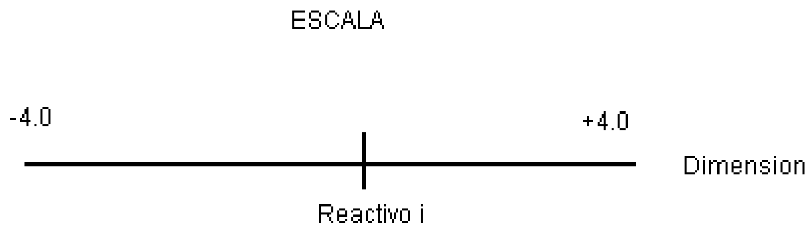


Fig.5.1 Construcción de un punto en la escala unidimensional

A la escala que empleamos le denominamos métrica y debe ser lo más uniforme posible y, en especial lineal para la variable en estudio.

Si construimos varios reactivos que van a medir el rasgo dentro de la misma escala, queremos decir que los reactivos van a hacer referencia a diferentes niveles de un mismo rasgo latente. El ejemplo más simple consiste en pensar que queremos medir habilidad para realizar saltos de altura y utilizamos un juego de vallas a diferentes alturas: 20 cm, 30 cm, 40 cm , etc. En este caso se trata de vallas en orden creciente de dificultad en rangos de 10 en 10 cm. Mientras más vallas tengamos podremos hacer más precisa la medición, podríamos disponer de vallas en rangos de 5 en 5 cm o 1 cm en 1 cm.

Si llevamos el ejemplo de las vallas a un examen, los reactivos corresponden con las vallas. Convendría tener un buen número de reactivos (vallas) en orden creciente de dificultad (alturas) (fig.5.2).

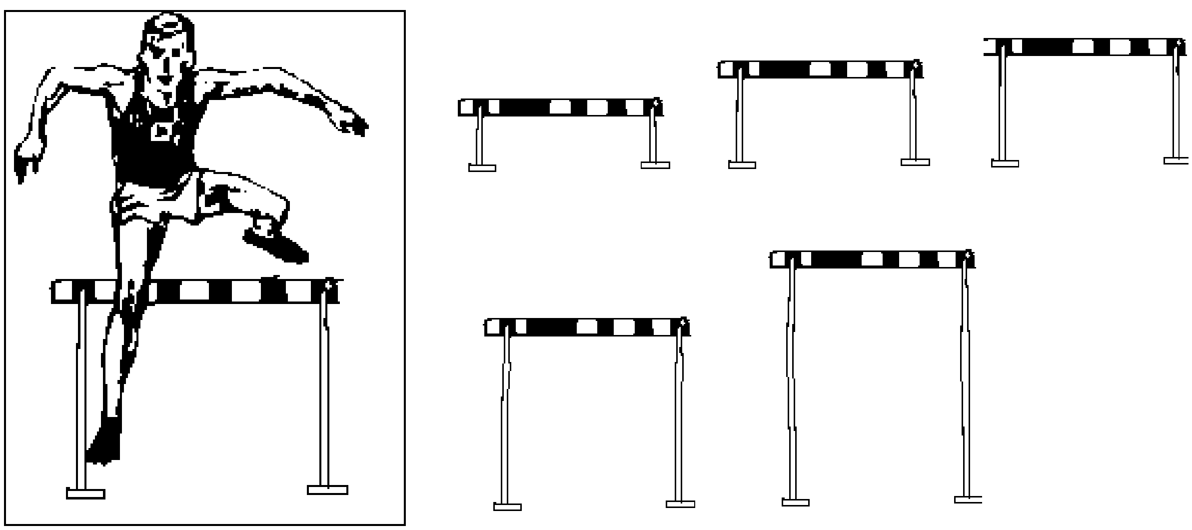


Fig.5.2 Una escala graduada de vallas

En este ejemplo pensamos que la persona que salta la valla de 50 cm con toda seguridad debe saltar las vallas inferiores: 40, 30, 20 cms. También pensamos que el salto de 50 cm es más difícil que el de 40 y a su vez éste es más difícil que el de 30.

Si tenemos en mente el ejemplo de las vallas para el salto de altura, es claro que nos conviene poner más reactivos espaciados de manera más fina para poder disponer de una medida más precisa del rasgo que queremos medir. La figura 5.3 muestra una escala más fina que la de la figura 5.1.

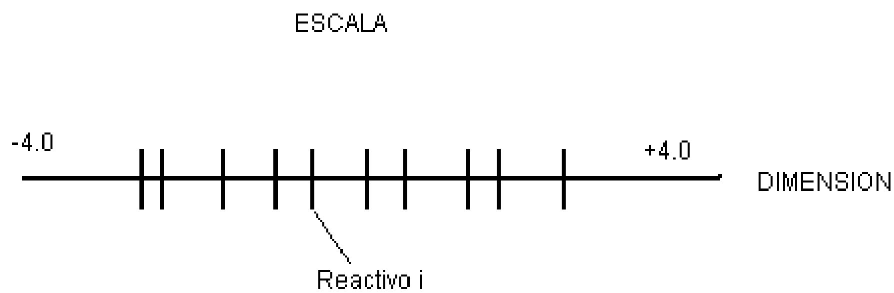


Fig.5.3 Una escala graduada de reactivos

No hay que abusar, sin embargo, poniendo un número exagerado de vallas (reactivos) porque podemos generar cansancio a las personas y hacer que el tiempo de solución de la prueba se alargue.

También es claro que no nos conviene colocar dos vallas a la misma altura, porque sería tanto como "repetir" el nivel de dificultad del salto. Esto equivale a decir que no nos conviene poner reactivos de igual dificultad para medir el mismo rasgo, porque estaríamos desperdiciando un reactivo a un mismo nivel de medida.

A este respecto conviene recordar que los reactivos tal vez no son tan perfectos como imaginamos a las vallas del salto de altura. Pero ¿ qué pasaría si nuestras vallas fueran de estos tipos tan variados?

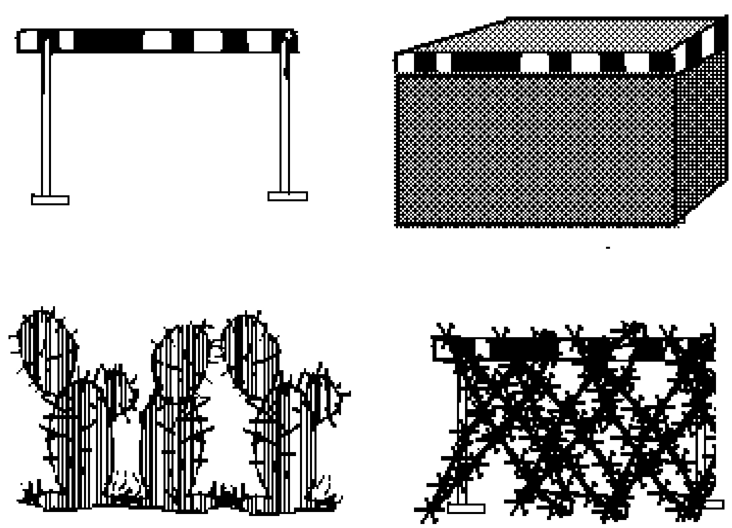


Fig.5.4 Diferentes vallas a una misma altura

Todas las vallas mostradas en la figura 5.4 intentan medir el salto a una misma altura. Pero es claro que hay de diferencias entre las vallas. Podríamos decir que una de ellas es la más "pura" pero que hay otras que además de altura seguramente miden otras cosas... tal vez todas miden la misma habilidad del salto de altura pero tienen "errores" de medida diferentes; muy probablemente uno lo piense mucho antes de saltar sobre los cactus o sobre la malla con púas, aunque estén a la misma altura. Debido a estos errores es seguro que aunque la dificultad se parezca (entendiendo la dificultad como la altura, que es el rasgo que estamos midiendo), habrá diferencia de capacidad de salto entre las personas, por lo que cada valla "discriminará" de una manera diferente. Igual pasa con los reactivos.

5.2 El Escalograma de Guttman

Estamos pensando que un rasgo es una dimensión que se va a medir con reactivos de dificultades y discriminaciones muy diversas. Una forma de representar las respuestas de las personas a nuestro examen consiste en hacer una tabla donde anotaremos en cada renglón las personas y en las columnas las respuestas a cada reactivo. Tomemos el caso de un grupo de 14 personas que contestan un cuestionario de 8 preguntas.

tabla 5.1

persona	1	2	3	4	5	6	7	8
vicente	1	1	0	1	0	1	0	0
nina	1	1	0	1	1	0	0	0
lupe	1	1	0	1	0	0	0	0
magda	1	1	0	0	0	0	0	0
lorena	1	0	0	0	0	0	0	0
juan	1	1	1	1	1	1	1	1
pepe	1	1	1	1	1	1	1	0
luisa	1	1	1	1	1	0	1	1
diana	1	1	1	1	0	0	0	0
samuel	1	0	0	0	0	1	0	0
rebe	1	1	1	0	0	0	0	0
ruperto	1	0	1	1	0	0	0	0
toña	1	1	1	1	1	1	0	0
memo	1	1	1	1	0	1	1	0

Las respuestas que aparecen en esta tabla son obtenidas de acuerdo con el nivel de dominio o de habilidad que tiene cada persona al enfrentarse a cada uno de los reactivos. Esperamos que las respuestas sigan algún "patrón" del tipo: a mayor dominio debe haber mayor número de respuestas. Sin embargo un alumno de cualquier nivel de dominio puede fallar en un reactivo "fácil" por diversos motivos (error de lectura, distracción, malas instrucciones en la pregunta, etc.). Se dice entonces que los valores de esta tabla son estocásticos.

Lo que vamos a construir ahora es el Escalograma de Guttman. Se trata de ordenar a las personas en orden descendente de aciertos (de la que tiene más aciertos a la que contestó menos) y a los reactivos en forma ascendente de aciertos (del reactivo que tiene más respuestas o más fácil al que tiene menos respuestas o más difícil). Este ordenamiento no es difícil pero sí laborioso. Después de ordenar a las personas y a los reactivos tenemos ya el Escalograma de Guttman, como sigue:

tabla 5.2

persona	7	2	5	4	3	8	6	1
juan	1	1	1	1	1	1	1	1
pepe	1	1	1	1	1	1	1	0
luisa	1	1	1	1	1	0	1	1
toña	1	1	1	1	1	1	0	0
memo	1	1	1	1	0	1	1	0
vicente	1	1	0	1	0	1	0	0
nina	1	1	0	1	1	0	0	0
diana	1	1	1	1	0	0	0	0
rebe	1	1	1	0	0	0	0	0
ruperto	1	0	1	1	0	0	0	0
lupe	1	1	0	1	0	0	0	0
magda	1	1	0	0	0	0	0	0
samuel	1	0	0	0	0	0	1	0
lorena	1	0	0	0	0	0	0	0

Como estamos midiendo una sola variable o dimensión, el Escalograma es una buen herramienta que nos permite ver algunas cosas interesantes con relación al reactivo y con relación a las respuestas de las personas.

Podemos apreciar que hay una persona que contestó todo el cuestionario (Juan) y podemos localizar a la persona que contestó menos (Lorena). La construcción ayuda, por lo tanto a ubicar a los alumnos de manera rápida.

Vemos que algunos alumnos contestan los reactivos fáciles y a partir de un cierto reactivo dejan de contestar. Esto lo señalamos en color gris dentro de la tabla 5.2. Tal es el caso de Pepe, Toña, Diana, Rebe, etc. Se puede afirmar que estas personas contestan de acuerdo con un "patrón lógico", a saber: solamente contestan las preguntas abajo de su nivel de dominio, una vez alcanzado dicho nivel las preguntas son más difíciles de lo que pueden contestar y a partir de ahí no pueden responder correctamente.

Pero también vemos algunos casos de personas que no contestan con el "patrón lógico", por ejemplo Luisa, Memo, Vicente están en este caso. Podemos ver una diferencia substancial entre Magda y Samuel: ambos tienen 2 aciertos, pero Magda está en el "patrón lógico" y Samuel en cambio responde una pregunta que es muy difícil. Podríamos conjeturar que Samuel contestó el reactivo 6 por adivinación (o tal vez se le preguntó acerca de una cosa que sabía).

El patrón que muestra el Escalograma de Guttman permite hacer conjeturas respecto a la adivinación personal. Es probable que Samuel adivinó el reactivo 6, pero seguramente Magda no adivinó en sus respuestas. Por ello hemos dicho que no estamos de acuerdo con la adivinación sistemática y que es factible "detectar" la adivinación personal.

Ahora revisemos el Escalograma en sentido vertical. Observamos que el reactivo 7 es contestado por todas las personas, es muy fácil y no discrimina entre los alumnos. Mientras que el reactivo 1 es muy difícil y solamente algunos alumnos pudieron contestarlo. El Escalograma nos ayuda a ordenar los datos y revisarlos rápidamente.

Se repite aquí la tabla 5.2 para presentar los resultados de los reactivos.

tabla 5.2b

persona	7	2	5	4	3	8	6	1
juan	1	1	1	1	1	1	1	1
pepe	1	1	1	1	1	1	1	0
luisa	1	1	1	1	1	0	1	1
toña	1	1	1	1	1	1	0	0
memo	1	1	1	1	0	1	1	0
vicente	1	1	0	1	0	1	0	0
nina	1	1	0	1	1	0	0	0
diana	1	1	1	1	0	0	0	0
rebe	1	1	1	0	0	0	0	0
ruperto	1	0	1	1	0	0	0	0
lupe	1	1	0	1	0	0	0	0
magda	1	1	0	0	0	0	0	0
samuel	1	0	0	0	0	0	1	0
lorena	1	0	0	0	0	0	0	0

Si comparamos el reactivo 4 con el 3 podemos ver que este último favorece respuestas de alumnos del grupo inferior. Puede decirse que el reactivo 4 tiene un patrón más lógico que el reactivo 3. Asimismo el reactivo 6 (uno de los más difíciles) tuvo una respuesta inesperada de Samuel. El patrón del reactivo 6 no es lógico y podría atribuirse a una respuesta al azar del alumno, no tanto a que el reactivo favorezca las respuestas del grupo inferior.

Conviene aclarar las hipótesis involucradas en el Escalograma de Guttman.

a) **La variable es unidimensional -** Mide un solo rasgo. No se permite juntar o analizar rasgos de diversas dimensiones en una misma tabla. Dicho de otro modo: no se deben analizar en el mismo Escalograma las respuestas de Habilidad Verbal y las de Matemáticas.

b) **La variable es ordenada -** El rasgo tiene una relación directa con la dificultad del reactivo. Por ello el conjunto de reactivos puede ordenarse en dificultad, de tal forma que un reactivo donde contestan 10 personas es más difícil que el reactivo que es contestado por 20 personas y, a su vez, más difícil que el reactivo contestado por 80 personas.

c) **La variable es inclusiva -** Si una persona contesta un reactivo de valor "30" entonces sería de esperarse que posee el dominio de los reactivos de valores iguales o menores que 30.

De lo anterior se observa que dada una habilidad B y un reactivo de dificultad D,

si B>D -> el reactivo se le hará "fácil" a la persona y por lo tanto tiene probabilidad de acertar

si B<D -> el reactivo se le hará difícil a la persona y por lo tanto tiene probabilidad de fallar.

Podemos considerar que el escalograma identifica a los reactivos y a las personas de tal modo que se tiene una configuración como la que muestra la figura 5.5

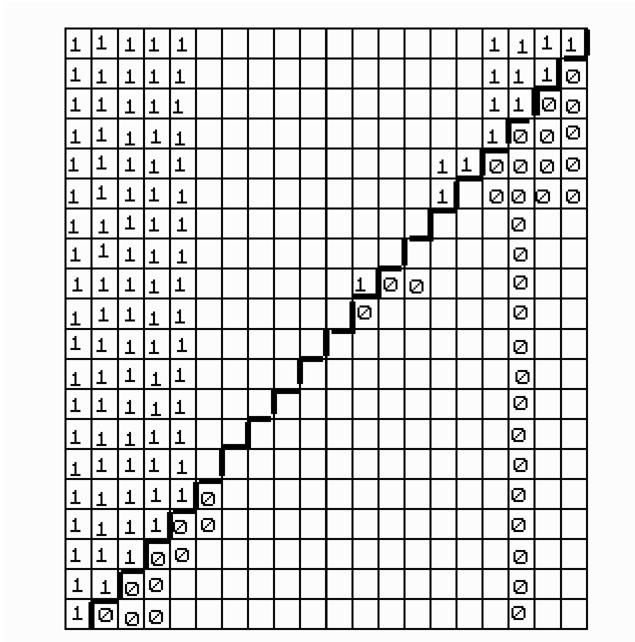


Fig 5.5 Escalograma de Guttman “ideal”

Llamaremos “diagonal” del Escalograma a la línea que se traza de abajo a la izquierda hacia arriba a la derecha y que divide al Escalograma aproximadamente en dos zonas triangulares. Observamos que arriba de la diagonal se esperan aciertos y que abajo se esperan errores.

Es claro que las respuestas son estocásticas, por lo que no todos los valores arriba de la diagonal serán 1, habrá necesariamente algunos errores normales, atribuibles a fallas de comprensión de la pregunta, buenas opciones distractoras en un reactivo de opción múltiple, etc. Igualmente se puede esperar que no todas las respuestas abajo de la diagonal serán 0. Habrá presencia de aciertos y fallas en forma estocástica.

Definimos el Escalograma Perfecto al que cumple la no estocasticidad, es decir que garantiza que arriba de la diagonal se tienen siempre aciertos y abajo de la diagonal se tienen siempre errores, como el que se muestra en la figura 5.5.

Es prácticamente imposible encontrar un Escalograma Perfecto. Por lo mismo, aunque es el modelo previsto, no debemos forzar los cuestionarios a que se vuelvan escalogramas perfectos ni esperar que se nos presenten. Es más: si el lector llega a encontrar un Escalograma Perfecto al analizar los datos de su examen, entonces se verá en serios problemas para usar esta herramienta dentro del análisis de Rasch, como veremos en la sección 5.3. En este caso deberá recurrirse a los procedimientos clásicos de análisis de reactivos: grado de dificultad, poder de discriminación, etc.

Ejercicio 5.1

Construir el Escalograma de Guttman para el conjunto de respuestas dado por la tabla.

persona	Reac.1	Reac.2	Reac.3	Reac.4	Reac.5
Alberto	1	1	1	0	0
Gonzalo	1	0	1	1	1
Luisa	0	0	1	1	1
Laura	0	0	0	0	1
Verónica	1	0	0	0	1

<u>*Solución:*</u>

Paso a) Se toma la tabla de respuestas y se ordena a los reactivos del más fácil al más difícil

persona	Reac.5	Reac.1	Reac.3	Reac.4	Reac.2
Alberto	0	1	1	0	1
Gonzalo	1	1	1	1	0
Luisa	1	0	1	1	0
Laura	1	0	0	0	0
Verónica	1	1	0	0	0

En este punto sólo hay la duda si se debe colocar primero al reactivo 1 y luego al 3 o viceversa, ya que ambos tienen 3 aciertos. Esto se revisa al ordenar a los alumnos.

Paso b) Se ordena la nueva tabla por los renglones, del alumno de mejor resultado al del resultado más bajo.

persona	Reac.5	Reac.1	Reac.3	Reac.4	Reac.2
Gonzalo	1	1	1	1	0
Alberto	0	1	1	0	1
Luisa	1	0	1	1	0
Verónica	1	1	0	0	0
Laura	1	0	0	0	0

Queda la duda quien debe ir primero de Alberto y de Luisa, ya que ambos tienen 3 aciertos. Esto se ajusta en el último paso.

Paso c) Acomodo final de la tabla, en reactivos y en personas, cuando hay empate, buscando la mejor configuración del escalograma.

persona	Reac.5	Reac.1	Reac.3	Reac.4	Reac.2
Gonzalo	1	1	1	1	0
Alberto	0	1	1	0	1
Luisa	1	0	1	1	0
Verónica	1	1	0	0	0
Laura	1	0	0	0	0

Los empates se tuvieron en los reactivos 1 y 3, así como entre Alberto y Luisa. Si se intercambian las columnas y los renglones se puede ver que no se mejora la configuración del escalograma. Se hace el cambio entre Alberto y Luisa porque nos parece que el patrón de respuestas de Luisa es más aceptable.

persona	Reac.5	Reac.1	Reac.3	Reac.4	Reac.2
Gonzalo	1	1	1	1	0
Luisa	1	0	1	1	0
Alberto	0	1	1	0	1
Verónica	1	1	0	0	0
Laura	1	0	0	0	0

Ejercicio 5.2

Identificar las respuestas inesperadas en las personas y en los reactivos del ejercicio 5.1

Solución:

A partir del escalograma se tienen estos casos para las personas:
1) Gonzalo, Verónica y Laura contestan en patrón lógico
2) Alberto no contestó el reactivo más fácil (reactivo 5) y en cambio acertó el más difícil (reactivo 2), tiene un patrón "no razonable"
3) Luisa tiene una respuesta errónea inesperada en el reactivo 1 o, complementariamente, un acierto inesperado en el reactivo 4, pero su patrón puede considerarse estocástico "razonable"

Con relación a los reactivos se puede decir:
Todos los reactivos tienen un patrón estocástico plausible, con excepción del reactivo 2 que aparenta una respuesta correcta inesperada del alumno Alberto.

5.3 El "Pre-proceso" del Escalograma de Guttman

Ya vimos en capítulos anteriores que un punto en la escala de dificultades corresponde con una probabilidad de respuesta. Así recordemos que un reactivo que es contestado por un 50% de alumnos corresponde con una Dificultad de 0 (sería el centro de la escala). Al trabajar en lógitos se sabe que un reactivo de calibración 1 está una unidad por arriba del origen y tiene una distancia de 2 unidades por arriba de otro reactivo de dificultad -1. Si encontramos un reactivo que contestan muy pocas personas podríamos tener una medida de -2.5, por ejemplo.

Pero esta construcción de la escala tiene un problema. ¿Cuánto mide el reactivo que no es contestado por ningún alumno ? Podemos afirmar que está por abajo del reactivo cuya calibración es -2.5, pero ¿qué tan abajo? ¿podría ser -3? ¿tal vez llegue a ser -4?, ¿-4.5? Nótese que cualquier reactivo no contestado implica carecer de pistas para ubicarlo en la escala. Desde el punto de vista del análisis clásico de reactivos estos reactivos de respuesta nula son desechables, porque no permiten hacer una medida precisa y, además, no discriminan entre los sujetos. Desde el punto de vista de lo que estamos construyendo en este libro, un reactivo de estas características también es desechable, porque un reactivo de nula respuesta resulta “imposible” de medirse y ubicarse en la escala.

El mismo problema lo tenemos con los reactivos que son contestados por la totalidad de las personas. ¿qué tan arriba están ubicados? Estos reactivos también son desechables. Los reactivos descritos hasta ahora se denominarán “reactivos extremos”.

Algo similar ocurre con las personas y sus medidas. No es posible ubicar a la persona que contesta todo el cuestionario ni a la que no contesta nada. ¿qué tan alto está la primera? ¿qué tan bajo queda la segunda? Las personas a que hacemos referencia se denominarán “personas extremas”.

¿Qué debemos hacer en estos casos? Hay varias opciones, pero la que mencionaremos aquí es la de hacer un “pre-proceso” de los datos, que consiste en eliminar del análisis a los reactivos extremos así como a las personas extremas. Con este “pre-proceso” se eliminan los casos que presentan problemas de ubicación en la escala. El conjunto de datos que resta ya puede ser analizado y ubicado en su correcto valor de medida.

Retomemos el caso de la tabla que estamos analizando y que volvemos a presentar aquí tal y como quedó ordenada en la tabla 5.2b.

Tabla 5.3

persona	7	2	5	4	3	8	6	1
juan	1	1	1	1	1	1	1	1
pepe	1	1	1	1	1	1	1	0
luisa	1	1	1	1	1	0	1	1
toña	1	1	1	1	1	1	0	0
memo	1	1	1	1	0	1	1	0
vicente	1	1	0	1	0	1	0	0
nina	1	1	0	1	1	0	0	0
diana	1	1	1	1	0	0	0	0
rebe	1	1	1	0	0	0	0	0
ruperto	1	0	1	1	0	0	0	0
lupe	1	1	0	1	0	0	0	0
magda	1	1	0	0	0	0	0	0
samuel	1	0	0	0	0	0	1	0
lorena	1	0	0	0	0	0	0	0

Hagamos el pre-proceso para eliminar reactivos extremos y personas extremas. Estos son los pasos:

1) Quitamos a Juan (alumno extremo que contestó todo el cuestionario)
2) Quitamos el reactivo 7 (reactivo extremo que contestaron todos los sustentantes)

con ello se obtiene esta nueva tabla:

Tabla 5.4

persona	2	5	4	3	8	6	1
pepe	1	1	1	1	1	1	0
luisa	1	1	1	1	0	1	1
toña	1	1	1	1	1	0	0
memo	1	1	1	0	1	1	0
vicente	1	0	1	0	1	0	0
nina	1	0	1	1	0	0	0
diana	1	1	1	0	0	0	0
rebe	1	1	0	0	0	0	0
ruperto	0	1	1	0	0	0	0
lupe	1	0	1	0	0	0	0
magda	1	0	0	0	0	0	0
samuel	0	0	0	0	0	1	0
lorena	0	0	0	0	0	0	0

Repetimos el pre-proceso con la tabla 5.4

3) Quitamos a Lorena (alumna extrema que no contestó nada en la nueva tabla)

Tabla 5.5

persona	2	5	4	3	8	6	1
pepe	1	1	1	1	1	1	0
luisa	1	1	1	1	0	1	1
toña	1	1	1	1	1	0	0
memo	1	1	1	0	1	1	0
vicente	1	0	1	0	1	0	0
nina	1	0	1	1	0	0	0
diana	1	1	1	0	0	0	0
rebe	1	1	0	0	0	0	0
ruperto	0	1	1	0	0	0	0
lupe	1	0	1	0	0	0	0
magda	1	0	0	0	0	0	0
samuel	0	0	0	0	0	1	0

Esta última tabla ya es “estable”, en el sentido de que no tiene casos extremos de personas o reactivos, por lo que se puede proceder al análisis. A partir de estos datos puede calcularse la medida de las personas y la calibración de los reactivos.

Sin embargo el “pre-proceso” puede tener problemas. En particular se trata de un problema muy serio relacionado con el Escalograma Perfecto. Esto lo trataremos en el siguiente ejercicio.

Ejercicio 5.3

Hacer el pre-proceso del cuestionario cuya tabla de respuestas conduce a este Escalograma de Guttman perfecto:

	REACTIVOS					
persona	1	2	3	4	5	6
juan	1	1	1	1	1	1
luis	1	1	1	1	1	0
memo	1	1	1	1	0	0
lola	1	1	1	0	0	0
marta	1	1	0	0	0	0

Solución:

Para el pre-proceso se harán estos pasos:

1) Eliminar reactivos extremos 1 y 2

2) Eliminar persona extrema "Juan"

Se tiene esta nueva tabla:

persona	3	4	5	6
luis	1	1	1	0
memo	1	1	0	0
lola	1	0	0	0
marta	0	0	0	0

Repetimos los pasos con esta nueva tabla:

3) Eliminar reactivo extremo 6

4) Eliminar persona extrema "Marta"

Se tiene esta nueva tabla:

persona	3	4	5
luis	1	1	1
memo	1	1	0
lola	1	0	0

Se repiten los pasos con esta nueva tabla:

5) Eliminar reactivo extremo 3

6) Eliminar persona extrema "Luis"

Se tiene esta nueva tabla:

persona	4	5
memo	1	0
lola	0	0

Esta última tabla obligará a eliminar a Lola y al reactivo 5, con lo cual se deberá quitar finalmente a Memo

Resulta curioso que en un Escalograma Perfecto el “pre-proceso” obliga a deshacerse de todos los datos. Realmente se trata de un problema muy serio. Por ello el Escalograma Perfecto no puede analizarse con la técnica que se presenta en este libro, a menos que se cancele el “pre-proceso”. Por ejemplo, el programa BIGSTEPS para análisis de Rasch falla durante el pre-proceso cuando se enfrenta a un Escalograma Perfecto.

5.4 El concepto de “Error”

A partir del Escalograma podemos ver que hay alumnos que responden con lo que hemos denominado un “patrón lógico” y otros que tienen respuestas inesperadas. Entendemos como respuestas inesperadas los aciertos que ocurren en reactivos más difíciles que la medida de la persona o bien respuestas incorrectas en preguntas más fáciles que la medida de la persona.

Hay autores que sugieren hacer un conteo simple de respuestas correctas inesperadas y de respuestas incorrectas inesperadas, a este conteo le denominan “error”. Desde luego este enfoque es deficiente: si regresamos a la Tabla 5.2b, apreciamos que no es igual el patrón de respuestas de Rebe que el de Ruperto, aunque tengan exactamente el mismo número de respuestas inesperadas. Asimismo el patrón de respuestas de Samuel tiene, seguramente, un error muy grande.

No es el número de respuestas inesperadas, sino la posición relativa respecto a la medida la que indica el error. Una respuesta inesperada alejada de la medida de la persona conduce a un error grande. Una respuesta inesperada cerca de la medida de la persona indica un error pequeño. El error debe medir, por lo tanto LA DISTANCIA ENTRE LA RESPUESTA Y LA MEDIDA DE LA PERSONA.

5.4.1 El error y la distancia

¿Cómo medimos la distancia entre la respuesta y la medida? El modo de hacerlo es relativamente simple. Hemos trabajado desde el inicio de este libro con el concepto de la medida en lógitos. Hemos dicho que tanto la calibración de los reactivos como la medida de las personas se representan en el mismo eje. La distancia entre la respuesta y la medida se obtiene simplemente por la diferencia de la calibración y de la medida en lógitos:

$$\text{Distancia} = B - D \qquad (5.1)$$

donde recordamos que B indica medida en lógitos del dominio de una persona y D indica la calibración en lógitos de la Dificultad de un reactivo.

Por lo anterior concluimos que lo importante no es la calibración por sí misma o la medida por sí misma, sino la diferencia entre ellas. En función de esa diferencia se tendrá una probabilidad de respuesta. Dicho de manera simbólica:

Si $B=D \rightarrow p=p(D)$ (5.2a)
Si $B>D \rightarrow p>p(D)$ (5.2b)
Si $B<D \rightarrow p<p(D)$ (5.2c)

Esto se parece mucho a expresiones que vimos en el capítulo 3. La expresión que relaciona la probabilidad de respuesta con la distancia B-D tiene, pues, esta forma:

$$\ln(p/q) = B-D \qquad (5.3)$$

Esta forma es una condición suficiente para conocer la probabilidad de respuesta de un alumno de medida B ante un reactivo de calibración D. Decimos que es suficiente porque no se requiere de más parámetros para estimar la probabilidad de respuesta; más aún, resulta claro que a partir de la información disponible en el Escalograma de Guttman (que son todas las respuestas de los alumnos en el examen) no es factible estimar otros parámetros en forma clara y fidedigna. La fórmula 5.3 no es otra cosa que el mismo modelo de Rasch, presentado en la expresión 4.9. El modelo de Rasch resulta de manera natural al analizar el error a partir del Escalograma de Guttman. Debido a que se obtiene el modelo de Rasch, las expresiones 5.2 se pueden escribir así:

Si $B=D \rightarrow p=0.5$ (5.2d)
Si $B>D \rightarrow p>0.5$ (5.2e)
Si $B<D \rightarrow p<0.5$ (5.2f)

Si conocemos la probabilidad a partir de B-D, habrá una esperanza de respuesta correcta de cada alumno frente a cada uno de los reactivos que se le apliquen. Estamos en posibilidad de estimar el “error” en función de qué tan lejos se encuentre la respuesta correcta del sujeto respecto de la probabilidad estimada.

Para los propósitos de este capítulo diremos que el error se calculará (provisionalmente cuando menos) por medio de la diferencia de respuesta observada respecto a la probabilidad calculada por el modelo de Rasch.

$$\text{error} = \text{respuesta real} - p(B-D) \qquad (5.4)$$

En esta fórmula, la “respuesta real” es dicótoma (correcta-incorrecta, 0-1, etc.), mientras que la probabilidad p(B-D) es un número real comprendido entre 0 y 1, por lo que el error será igualmente un número real.

Observemos que el “error” tiene signo. En efecto, podemos tener respuestas inesperadas en varios casos distintos:

a) Si B=D sabemos que p=0.5. Esto corresponde al lanzamiento de una moneda, por lo que la respuesta pudiera ser tanto correcta como incorrecta. El error puede ser:

error = 1 - 0.5 = 0.5 -> si responde correctamente
error = 0 - 0.5 = -0.5 -> si responde incorrectamente

El "error por respuesta inesperada" que denominaremos aquí simplemente "Error" es nulo, porque ambas respuestas son factibles cuando el reactivo y el dominio del alumno coinciden.

b) Si $B>D$ sabemos que $p>0.5$. El alumno tiene mayores probabilidades de contestar, por lo que las respuestas incorrectas son inesperadas. Esto se escribe

$p_{observada} < p(B-D)$ para $B>D$

en este caso se tendría un error negativo.

$\text{Error} = 0 - p(B-D) < 0$

si $p_{observada} > p(B-D)$ entonces consideraremos que el error es nulo, porque la respuesta correcta es factible cuando el alumno tiene un dominio por arriba de la calibración del reactivo.

c) Si $B<D$ sabemos que $p<0.5$. El alumno tiene una medida por abajo de las dificultades que se le presentan en los reactivos, por lo que no debería contestar correctamente al reactivo. Una respuesta correcta es inesperada.

$p_{observada} > p(B-D)$ para $B<D$

en este caso tendríamos un error positivo.

$\text{error} = 1 - p(B-D) > 0$

Si $p_{observada}<p(B-D)$ entonces no se considera como error, ya que es normal que el alumno falle en preguntas más difíciles que su nivel de dominio.

Podemos hacer esta tabla resumen.

Tabla 5.10

Caso	B-D	probabilidad	Respuesta	Clasificación		Observación	Error
				Esperada	Inesperada		
1	$B>D$	>0.5	1	X		como $p>0.5$ es natural que resp=1	0
2	$B>D$	>0.5	0		X	como $p>0.5$ es inesperado que resp=0	$0-p(B-D)$
3	$B<D$	<0.5	1		X	como $p<0.5$ es inesperado que resp=1	$1-p(B-D)$
4	$B<D$	<0.5	0	X		como $p<0.5$ es natural que resp=0	0
5	$B=D$	0.5	cualquiera			como $p=0.5$ podemos esperar cualquier respuesta	0

Con objeto de contabilizar los errores que tienen diferente signo disponemos de varios esquemas algebraicos. Una forma es tomar el error en valor y hacer la suma algebraica de los errores (positivos todos en este caso); otra forma es tomar elevar los errores al cuadrado y obtener el error cuadrático o el error cuadrático medio, que es el promedio de errores cuadráticos de los reactivos.

En resumen se pueden establecer estas fórmulas para el error:

error absoluto = $\sum$ error$_i$ (5.5a)

error absoluto medio = [$\sum$ error$_i$] / N (5.5b)

error cuadrático = $\sum$ (error$_i$)^2 (5.5c)

error cuadrático medio = [$\sum$ (error$_i$)^2]/N (5.5d)

donde N es el número de reactivos.

Ejercicio 5.4

Se aplica un examen de 4 reactivos, cuya calibración está indicada por la tabla:

reactivo	D
1	-1.1
2	-0.3
3	+0.9
4	+1.3

Estimar los errores absolutos, cuadráticos y medios que tienen las respuestas de un alumno de medida B=1.0:

reactivo	Respuesta del alumno
1	1
2	0
3	1
4	0

Solución:

Para estimar el error se usa la fórmula:
error = respuesta - p(B-D)

Constrúyase la tabla, recordando que B=1:

reactivo	D	B-D	p(B-D)	Respuesta del alumno	error	error absoluto	error cuadrático
1	-1.1	-2.1	0.1091	1	0.8909	0.8909	0.7937
2	-0.3	1.3	0.2142	0	-0.2142	0.2142	0.0459
3	+0.9	0.1	0.4750	1	0.525	0.525	0.2756
4	+1.3	-0.3	0.5744	0	-0.5744	0.5744	0.3299
					TOTAL	2.2045	1.4451

Así se tiene que el error absoluto es 2.2045 y el error absoluto medio será:

error absoluto medio = error absoluto/No. de reactivos = 2.2045/4 = 0.5511

De igual modo, el error cuadrático es 1.4451 y el error cuadrático medio es:

error cuadrático medio = error cuadrático/No. de reactivos = 1.4451/4 = 0.3613

5.4.2 El residuo estandarizado

El último modelo de error que veremos involucra a la variable normalizada Z. Para usar este modelo recordemos que la variable Z se define por:

$$Z = (X-E)/s \qquad (5.6)$$

donde X es el valor observado, E es la media de calibraciones de los reactivos, s es la desviación estándar de las calibraciones. Esta expresión es idéntica a la de la variable normalizada Z que se estudia en cualquier curso de estadística básica. Cuando se emplea en el análisis del error de los reactivos a este valor de Z se le denomina “Residuo estandarizado”.

Adicionalmente se define el error cuadrático medio en Z o residuo cuadrático medio (Mean square) por medio de la expresión:

$$\text{Residuo Cuadrático Medio} = \sum (Z^2)/N \qquad (5.7)$$

También se usa la raíz del error cuadrático medio o raíz cuadrada del residuo cuadrático medio (Root mean square)

$$\text{Raiz del Residuo Cuadrático Medio} = \sqrt{\sum (Z^2)/N} \qquad (5.8)$$

Ejercicio 5.5

Determinar el Residuo Estandarizado y el Residuo Cuadrático Medio para el cuestionario del ejercicio 5.4

<u>Solución:</u>

El examen que estamos empleando es muy pequeño, ya que solamente tiene 4 reactivos, por lo que el estudio del error puede carecer de utilidad real para medir una variable, pero haremos abstracción de esta deficiencia y calcularemos E y s.

Así tenemos:

a) media del examen:

E = [(-1.1)+(-0.3)+(+0.9)+(+1.3)]/4 = 0.2

b) desviación estándar

$$s = \sqrt{[(-1.1-0.2)^\wedge 2 + (-0.3-0.2)^\wedge 2 + (+0.9-0.2)^\wedge 2 + (+1.3-0.2)^\wedge 2]/4} = 0.9912$$

Una vez calculados estos valores podemos determinar Z para cada persona, de acuerdo con las respuestas que dieron a cada reactivo.

Reactivo	Respuestas del sujeto	Residuo estandarizado Z
1	1	0.8071025
2	0	-0.20177563
3	1	0.8071025
4	0	-0.20177563
	SUMA	1.21065375

El Residuo estandarizado es igual a la suma de los residuos de los diferentes reactivos o, mejor aún, el residuo medio es la suma entre el número de reactivos:

Residuo estandarizado medio = 1.2106/4 = 0.3027

El residuo medio cuadrático se obtiene elevando al cuadrado los residuos

Reactivo	Respuestas del sujeto	Residuo estandarizado Z	Residuo cuadrático Z^2
1	1	0.8071025	0.8071025
2	0	-0.20177563	0.20177563
3	1	0.8071025	0.8071025
4	0	-0.20177563	0.20177563
	SUMA	1.21065375	2.01775626
	MEDIO	0.30266344	0.50443906

El residuo cuadrático medio se obtiene de acuerdo con la fórmula 5.7:

RCM= 2.0177/4 = 0.5044

y la raíz del RCM:

RCM = 0.7102

El uso del residuo estandarizado tiene algunas ventajas. En particular se calcula haciendo la suma en todos los reactivos, por lo que se vuelve independiente de si clasificamos a la respuesta como “esperada” o “inesperada”. Se trata de un estimador sistemático que puede compararse fácilmente contra algún valor de cotejo que se elija.

El cálculo de errores y residuos que hemos hecho aquí se sujeta a la posibilidad de que la tabla de respuestas no conduzca a un Escalograma perfecto. Antes de hacer estos cálculos debe, por lo tanto, hacerse el pre-proceso, de tal manera de que la tabla resultante no tenga “error” por sí misma.

Tanto el error, como el residuo estandarizado y sus valores medios, se emplearán en el siguiente capítulo cuando hablemos de la calidad del ajuste con el Modelo de Rasch.

5.5 Ejercicios propuestos

1. Repetir el ejercicio 5.4 para las personas de medida B=0.2 y B=1.5

2. Determinar el error absoluto, error absoluto medio, error cuadrático y error cuadrático medio para los resultados del ejercicio 1.

3. Calcular el residuo cuadrático medio y la raíz cuadrada del residuo cuadrático medio para los resultados del ejercicio 1.

4. Hacer el pre-proceso para los datos siguientes:

Sujeto	Reactivos								
	1	2	3	4	5	6	7	8	9
Sandra	1	0	1	1	1	0	1	1	1
Ramiro	1	0	1	1	1	0	1	1	1
Oscar	0	0	0	0	1	1	1	0	0
Alex	0	0	0	1	1	1	0	0	0
Lucy	0	1	0	0	0	1	0	0	0
Anita	0	0	1	0	1	0	1	0	0
Melisa	0	1	1	1	1	0	1	1	1
Aracely	0	1	0	0	1	1	1	1	0
Miguel	1	1	1	0	1	1	1	0	1
Angel	0	0	0	0	1	1	0	0	0
Lili	1	0	0	0	1	0	0	0	1
Lizeth	0	1	0	0	0	0	0	1	1
Carola	1	1	0	0	1	0	0	0	1

5. Calcule los errores y residuos estandarizados para los datos de la tabla del ejercicio 4.

6. Los investigadores en la materia afirman que el Escalograma de Guttman “perfecto” no es posible. Justifique esta afirmación.

7. Escriba las tres hipótesis sobre la variable que involucra el Escalograma de Guttman.

8. Justifique por qué la distancia entre medida y dificultad tiene interpretación en la métrica.

9. Escriba por lo menos tres características deseables de una escala

10. Se construye un cuestionario de 10 reactivos y todos tienen una Dificultad de 0 (igual a una probabilidad de respuestas de 0.5). En términos clásicos se diría que todos los reactivos tienen 50% de grado de dificultad. ¿Por qué decimos que el cuestionario es defectuoso? ¿Cómo afecta a la escala? ¿Se tiene una buena métrica?

5. 6 Puntos clave del capítulo

1. Escalograma de Guttman. Es un acomodo matricial de las puntuaciones de los sujetos en los reactivos de un examen, ordenando los reactivos en orden ascendente de Dificultad y a los sujetos en orden descendente de dominio.

2. Hipótesis sobre la variable en el Escalograma de Guttman:
a) La variable es unidimensional - Mide un solo rasgo.

b) La variable es ordenada - El rasgo tiene una relación directa con la dificultad del reactivo.

c) La variable es inclusiva - Un rasgo de dificultad Di engloba a los de dificultad menor

3. Error en la medida de una variable.

error = distancia entre medida y dificultad = B-D

error absoluto = $\sum$ $error_i$

error absoluto medio = [$\sum$ $error_i$] / N

error cuadrático = $\sum$ $(error_i)^2$

error cuadrático medio = [$\sum$ $(error_i)^2$]/N

4. Residuo estandarizado

Z= (X-E)/S

donde X es la respuesta, E es la media y S la desviación estándar de respuestas, medidas o dificultades según el caso.

Residuo Cuadrático Medio = $\sum$ (Z^2)/N

Raiz del Residuo Cuadrático Medio = $\sqrt{\sum(Z^2)/N}$

6. EL MODELO DE RASCH. MEDIDA Y AJUSTE

6.1 Una deducción del modelo de Rasch

Ahora tenemos todo preparado para iniciar la discusión del modelo de Rasch. Se trata de un modelo uniparamétrico, es decir, un solo parámetro de medición ligado al número de respuestas. Este parámetro corresponde, de acuerdo con Rasch, con una sola dimensión relativa a una sola escala para medir tanto el desempeño de la persona (habilidad, conocimientos, etc.), como la calidad de un reactivo (relacionada con la Dificultad). Existe una ventaja aparente en usar una sola escala para los reactivos y para las personas, porque ambos se pueden comparar directamente en dicha escala única.

De acuerdo con lo que puede esperarse de una buena medición, el instrumento de medida no debería verse influido por los objetos que mide. Dice Thurstone: "Si la regla de medida brinda diferentes valores cuando se usa en una superficie rugosa, en una pintura o en una hoja de papel, entonces se puede dudar de la calidad de la regla. Dentro del rango de objetos para el cual se usa un instrumento de medida, su función debe ser independiente del objeto medido".

El modelo de Rasch trata de responder a este problema usando este razonamiento: Supónganse las preguntas i,j que se van a analizar y que se espera que sean independientes de los sujetos medidos. Si se comparan las probabilidades de aciertos y errores que tiene la persona n en estos reactivos se tiene:

$$\frac{p(i = si, j = no)}{p(i = no, j = si)} = \frac{pni\,qnj}{qnipnj} \qquad (6.1)$$

donde pni es la probabilidad de aciertos de la persona n en la pregunta i, qni es la probabilidad de falla en la misma pregunta. Ya sabemos que qni=1-pni.

La expresión (6.1) debe ser válida independientemente de qué personas se emplean para su cálculo. Hágase la misma fórmula para la persona m. Entonces:

$$\frac{p(i = si, j = no)}{p(i = no, j = si)} = \frac{pmi\,qmj}{qmipmj} \qquad (6.2)$$

Y, forzosamente, (6.2) debería ser igual a (6.1) porque hemos dicho que las medidas de los reactivos son independientes de los sujetos. Queda por lo tanto:

$$\frac{pniqnj}{qnipnj} = \frac{pmi\,qmj}{qmipmj} \qquad (6.3)$$

Ahora establezcamos como "orígenes" de referencia de las preguntas j=k (pregunta cualquiera) y m=k (persona cualquiera). Con estos "orígenes", la pregunta i se calibrará respecto a k y la persona n se medirá igualmente respecto a k. Puede usarse el mismo valor como origen por la hipótesis básica del modelo que consiste en usar una sola escala tanto para personas como para reactivos. El origen además se va a ubicar, por comodidad, al centro de la escala en una medida de 0, que corresponde con un valor de p=0.5. Con ello se ubica el centro de los reactivos en el 50% de dificultad y el de personas en el 50% de dominio o de habilidad.

Esto simplifica a (6.3), ya que pmj=qmj=pkk=qkk=0.5. La expresión (6.3) se escribe:

$$\frac{pniqnk}{qnkpnk} = \frac{pki}{qki} \qquad (6.4)$$

A su vez:

$$\frac{pni}{qni} = \frac{pnk\,pki}{qnkqki} \qquad (6.5)$$

Se aprecia que el segundo miembro define dos cocientes: uno sobre la persona n referida al origen k, y otro sobre la pregunta i referida al origen k, con lo cual se estima un valor del primer miembro que es la interacción entre la persona n y la pregunta i.

La expresión (6.5) puede linealizarse tomando logaritmos en ambos miembros, quedando:

$$\ln\left(\frac{pni}{qni}\right) = \ln\left(\frac{pnk}{qnk}\right) + \ln\left(\frac{pki}{qki}\right) \qquad (6.6)$$

(recuérdese que por las propiedades de los logaritmos:

ln(ab) = ln(a)+ln(b))

Revisemos los términos de (6.6) y recordemos las cosas que hemos visto en los capítulos anteriores.

El primer miembro ln($\frac{pni}{qni}$) corresponde con la medida en lógitos de la persona n al responder el reactivo i.

El primer término del segundo miembro ln($\frac{pnk}{qnk}$) es la medida en lógitos de la persona n referida al origen k. Esto lo denominamos Bn.

El segundo término del segundo miembro ln($\frac{pki}{qki}$) se puede escribir también (aprovechando las propiedades de los logaritmos):

$$\ln\left(\frac{pki}{qki}\right) = -\ln\left(\frac{qki}{pki}\right)$$

Y ya sabemos que ln($\frac{qki}{pki}$) es la calibración en lógitos del reactivo i referida al origen k, que se denomina Di.

Entonces (6.6) se puede escribir:

$$\ln\left(\frac{pni}{qni}\right) = Bn - Di \qquad (6.7)$$

Si resolvemos para pni usando antilogaritmos, se llega a:

$$pni = e^{(Bn-Di)} / (1+e^{(Bn-Di)}) \qquad (6.8)$$

Esta función se representa como muestra la figura 6.1.

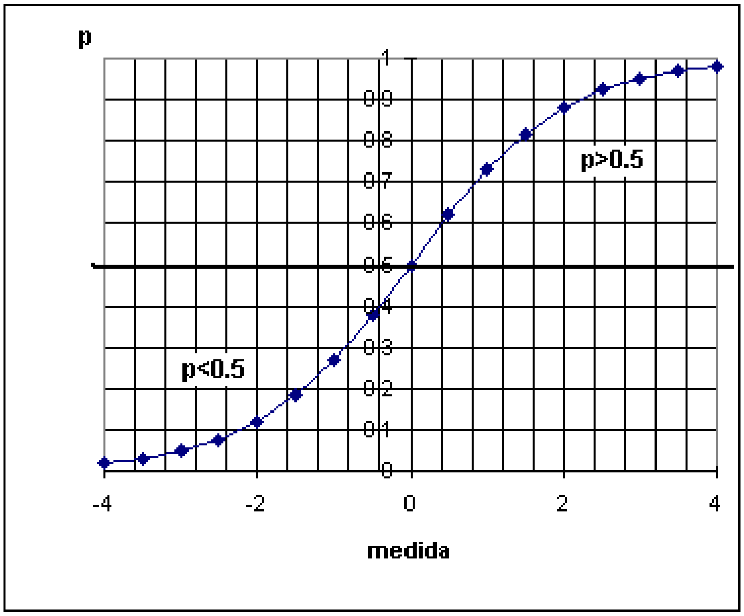


Fig. 6.1 Curva del Modelo de Rasch

La expresión (6.7) es el modelo de Rasch, mientras que (6.8) es el modelo de Thurstone de 1 parámetro. Se demuestra de esta manera que el modelo de Rasch es el que corresponde con la exigencia de independencia que propone Thurstone (se trata por lo tanto de una condición suficiente).

De acuerdo con esta deducción, el modelo de 3 parámetros no es suficiente sino redundante. Más concretamente podríamos decir que no es posible encontrar más de un parámetro a partir de la información disponible. Siguiendo con la deducción propuesta por Wright, la expresión (6.7) no solamente es suficiente sino necesaria para la independencia de la medida.

6.2 Algunas características del Modelo de Rasch

Podemos apreciar algunas características del modelo (6.7), que convergen con aspectos tratados anteriormente:

a) la medida en lógitos de una persona ante un reactivo dado es función de la diferencia Bn-Di
b) El valor Bn es independiente del examen aplicado, se entiende por Bn la medida de una persona referida al origen común k. Se trata de la medida "real" de una persona que debe poderse estimar independientemente de cualquier instrumento de medida que se desee.
c) El valor Di es la calibración del reactivo, referida a un origen común k. Esta calibración es independiente del conjunto de personas a quienes se aplicó el instrumento y representa la Dificultad "real" del reactivo.

El método propuesto por Rasch busca eliminar el problema de la métrica que se tiene en las medidas tradicionales o clásicas (porcentajes de aciertos, número bruto de reactivos correctos, etc.) Se trata de una métrica lineal, donde una distancia de 1 lógito indica una unidad, independientemente de la posición en la cual se encuentre una persona o un reactivo.

En la métrica de Rasch se trata de detectar y cuantificar una "estructura" entre los datos: personas-reactivos, por ejemplo, siempre que esta "estructura" exista. La estructura es un concepto de tipo predictivo, donde se puede afirmar que una persona de bajo rendimiento tiene una probabilidad baja de acertar a las preguntas difíciles, o que una persona de alto rendimiento tiene una alta probabilidad de acertar a las preguntas fáciles.

Debido a que el método de Rasch es probabilista, se establece de antemano que van a escasear los datos, o que los datos son insuficientes e, inclusive, que se tengan datos procedentes de un problema mal planteado. Bajo su enfoque se estudia la probabilidad de que los datos -insuficientes o escasos - establezcan una estructura, independientemente además de las posibles fluctuaciones que pueden tenerse en la toma de datos (errores en la aplicación, diferencias regionales, diferencias raciales, problemas personales de las personas, etc.).

La teoría de Rasch, por lo tanto, puede catalogarse entre las "teorías fuertes", ya que permite identificar estructuras entre los datos a pesar de todos los errores inherentes al proceso de medición. Con el método propuesto por Rasch se pueden hacer análisis para dos o más variables, dicótomas o polítomas, etc.

Ahora ya sabemos por qué el método de Rasch propone un modelo de comportamiento deseable para un reactivo. Por lo tanto el análisis de Rasch estudia qué tanto se acercan los datos medidos al modelo.

Si hacemos una recapitulación, podemos anotar estas hipótesis generales para el análisis de Rasch:

a) La escala que se maneja debe ser lineal
b) La escala debe permitir realizar medidas de reactivos y personas referidas a un origen común
c) La variable que se mide representa la diferencia entre la habilidad de la persona y la dificultad del instrumento
d) La variable es unidimensional, ordenada e inclusiva, de acuerdo con el modelo del Escalograma de Guttman
e) Las medidas son estocásticas, con una probabilidad de ocurrencia tanto en aciertos como en errores.

El análisis de Rasch incluye estos dos trabajos:

a) Un estudio de la medida de una persona al tomar una prueba o de un reactivo al aplicarse a una población (Análisis de medida).
b) Un estudio de qué tan bien ajustan los datos a las hipótesis (análisis de ajuste).

La medida de una persona es independiente del instrumento y de la población medida. El ajuste, en cambio, depende de la muestra y por lo tanto no se puede garantizar que los valores podrán ajustar en cualquiera otra aplicación o situación.

Por lo anterior, en este modelo se analiza:

- la plausibilidad de las respuestas de las personas en su conjunto
- la plausibilidad de las respuestas de cada persona al conjunto de reactivos de una prueba. Esto se hace con una revisión detallada de las respuestas de cada persona para ver si corresponden con el patrón general de respuestas.

Recuérdese que el patrón hipotético de Rasch, de acuerdo con el modelo de Guttman es que la probabilidad p es una función de B y D, tal que:

si Bi>Bj => P(Bi,D)>P(Bj,D)

si Di<Dj => P(B,Di)>P(B,Dj)

6.3 El error estándar

Una vez que se calcula una medida debe determinarse su error estándar de medida, que en el modelo de Rasch es mínimo cuando el reactivo está cerca de p=0.5 y tiene al infinito para p=0 y p=1. Esto indica que las medidas son más precisas cuando el reactivo se acerca a valores "centrales" y conduce a medidas imprecisas en los extremos. De hecho se puede decir que se conoce mejor a los reactivos centrales y no se conoce nada (o muy poco) sobre los reactivos donde todos contestan o nadie contesta.

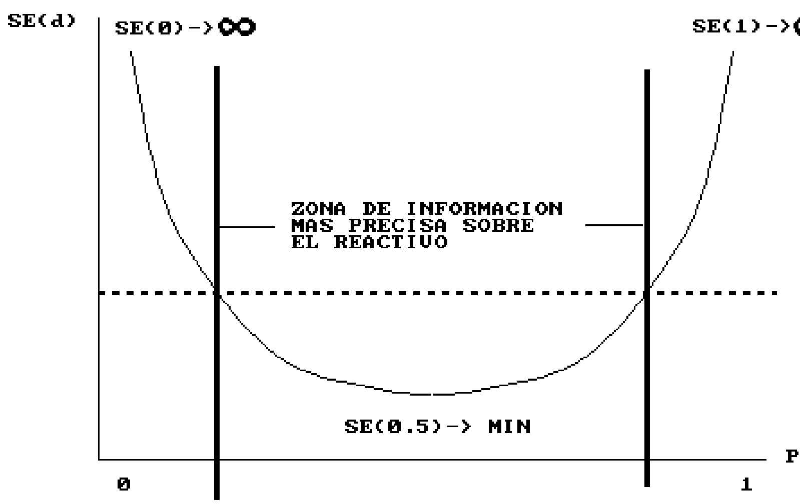


Fig. 6.2 Error estándar

¿Qué quiere decir este error estándar ? Que la medida nunca es precisa, sino que hay error debido a los diversos patrones que tienen las personas al contestar el cuestionario. Ya hemos hablado del error anteriormente: tiene que ver con la diferencia de la respuesta observada y la probabilidad de respuesta esperada. Debido a lo estocástico de las respuestas la diferencia se hace más grande en los extremos y más pequeña al centro.

6.4 La independencia de la medida

Para garantizar que la medida es independiente de la muestra e independiente del instrumento, se modifican los valores de acuerdo con la medida obtenida en el cuestionario por las personas. Esta modificación consiste de un “corrimiento” hacia la media, de tal modo que se tiene un lógito de 0 en la media de reactivos y de personas.

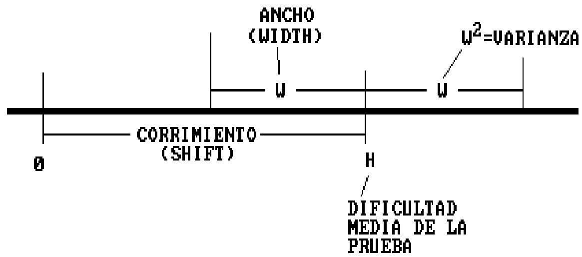


Fig. 6.3 Corrimiento en la medida

Es gracias a este procedimiento que se pueden representar en una misma escala las medidas de las personas y las calibraciones de los reactivos.

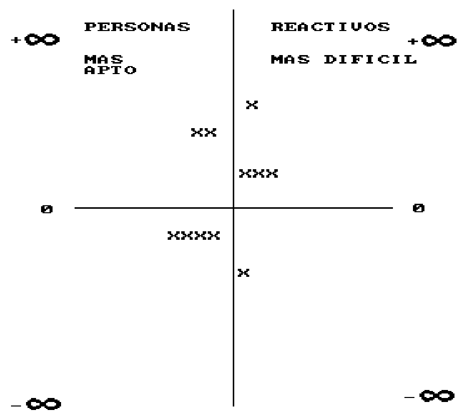


Fig 6.4 Métrica común para reactivos y sujetos

El corrimiento (“SHIFT” en inglés) es un valor en términos de varianza que hace que la medida de un alumno se refiera a la dificultad del cuestionario. Por ejemplo pensemos en un alumno de nivel medio que resuelve un cuestionario difícil. Debido a la dificultad del examen sus resultados pueden ser bajos y estaríamos tentados a decir que su medida es baja. De modo contrario le ocurriría al mismo alumno ante un examen fácil, creeríamos que su medida es mayor simplemente porque el instrumento le permite tener mejores resultados.

Si hacemos referencia a la dificultad del instrumento, es claro que en el primer caso deberemos “aumentar” la calificación del alumno para que refleje que le tocó resolver un cuestionario difícil. En el segundo caso deberemos “reducir” la calificación del alumno porque su examen resultó fácil. El resultado de aumentar una calificación baja o el de reducir una calificación alta debe conducir a la misma medida de la persona.

Como estamos trabajando en una escala lineal, está permitido hacer el corrimiento (hacia arriba o hacia abajo), con objeto de obtener la medida “real” del alumno.

6.5 El problema del “ajuste” o control de calidad del modelo.

Para poder estudiar la presencia o ausencia de “estructura” entre los datos se emplea el concepto de “ajuste” (“FIT” en inglés), que ya comentamos anteriormente, de modo de poder responder a la pregunta: ¿qué tan bien ajustan los resultados obtenidos en el experimento o medición con el modelo de Rasch?

El “ajuste” es un proceso de cálculo que permite estimar la calidad de los resultados comparados contra el modelo. Por medio de unos parámetros de ajuste se hace el “control de calidad”, que indica el grado en el cual las variables dadas (o las categoría de dichas variables) permiten identificar y definir una “estructura” entre los datos. Puede decirse que el “ajuste” permite medir la manera en que se “coordinan” o “correlacionan” los datos en la estructura.

Hay varios parámetros de “ajuste”. Se sugiere el uso de los residuos medios estandarizados y más específicamente, al cuadrado de los residuos medios estandarizados, que vimos en la sección 5.4.

Las fórmulas dadas en la sección 5.4 sólo requieren de una pequeña modificación, con objeto de que se emplee un valor de referencia de fácil interpretación. En este caso se sugiere utilizar el valor 1 como referencia. Al hacer esta modificación se tiene el valor cuadrático medio de “ajuste” con esperanza de 1. Los valores superiores a 1 indican que los patrones de respuesta no están de acuerdo con la hipótesis de que el cuestionario permite identificar una estructura general entre los datos.

La primera medida de ajuste se denomina “OUTFIT” (del inglés “Outlier-sensitive fit statistics”) y que podríamos traducir muy libremente como “ajuste externo”. Se trata de un valor sensible al comportamiento inesperado que afecta a los reactivos LEJOS del nivel de habilidad de una persona.

El OUTFIT se calcula con la fórmula:

$$\text{OUTFIT} = \sum Z^2/N \qquad (6.9)$$

esta fórmula es la misma expresión 5.7 correspondiente al residuo cuadrático medio.

La otra forma de calcular el ajuste es con el “INFIT” (del inglés “Information-weighted fit statistics”) y que podríamos traducir muy libremente como “ajuste interno”. Se trata de un valor sensible al comportamiento inesperado que afecta a los reactivos CERCA del nivel de habilidad de ina persona.

Se calcula con esta expresión que hace un promedio ponderado (“pesado”) de los residuos estandarizados:

$$\text{INFIT} = \frac{\sum z^2 s^2}{\sum s^2} = \frac{\sum (x-E)^2}{\sum s^2} \qquad (6.10)$$

El Ajuste Interno (INFIT) es aplicable a las variables que se encuentran en la zona intermedia de la escala. Permite identificar los patrones inesperados entre observaciones que corresponden a la zona en estudio.

En forma esquemática, si se tiene la escala de calibración que va de “fácil” a “difícil”, el ajuste interno se enfoca a la zona media o de transición entre los extremos.

Bajo este parámetro es de esperar que en la zona fácil y en la zona difícil los patrones correspondan con el modelo (muchas respuestas en la zona fácil y pocas respuestas en la zona difícil). En la zona de transición se esperan mezclas estocásticas de “éxitos y fracasos” o “aciertos y errores”. Si los datos corresponden con el modelo, el ajuste interno corresponderá con lo esperado, es decir con un valor cercano a 1.

Los puntos correspondientes a las calibraciones de los reactivos o a las medidas de las personas se representan en un plano de medida vs valor del residuo cuadrático medio o residuo estandarizado. En este caso se obtiene una nube de puntos que se distribuyen de manera normal respecto al origen.

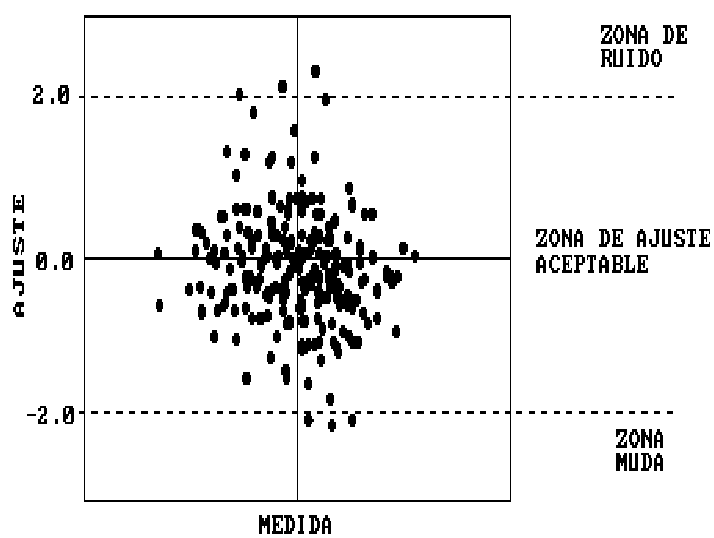


Fig 6.6 Ubicación de puntos en la nube de ajuste

En una representación tipo “histograma” de los valores de ajuste, se puede apreciar que se distribuyen con una media de 0.

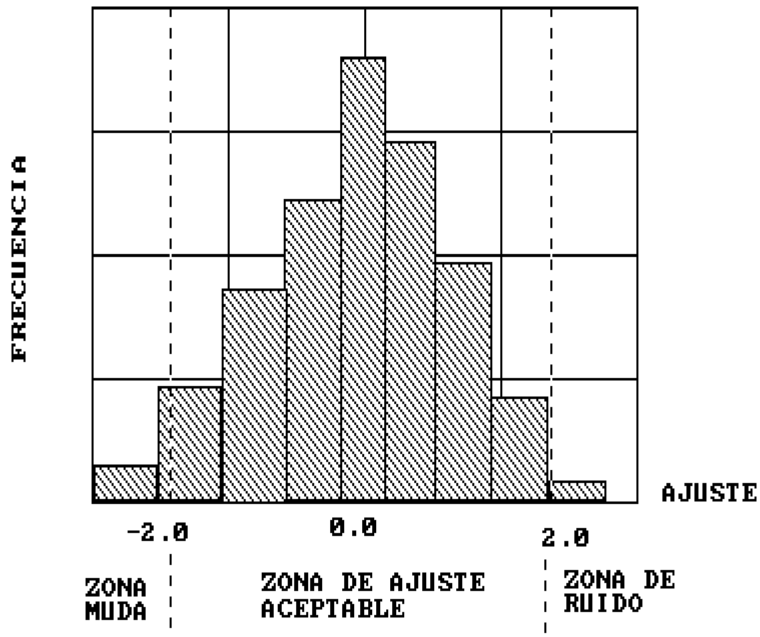


Fig. 6.7 Diagrama de frecuencias para el ajuste

La interpretación del “ajuste” (“FIT) es como sigue:

a) Valores cercanos a 0, o dentro del intervalo (-1, +1) indican un ajuste razonable al modelo estocástico
b) Valores inferiores a -2 indican demasiado “determinismo” en el patrón de respuestas o “poca estocasticidad”, digamos que “no hay errores” en el patrón de respuestas. A esto se le llama “patrón mudo” (“MUTE” en inglés).
c) Valores superiores a +2 indican demasiada estocasticidad, mucha posibilidad de azar (más allá del azar “razonable”), o demasiadas respuestas inesperadas. A este caso se le denomina “patrón ruidoso” o “patrón de ruido” (“NOISY” en inglés).

El Modelo de Rasch no se compromete a rangos específicos de valores, pero los anteriores son muy utilizados en la práctica. Disponemos entonces de un conjunto de valores de aceptación de los reactivos: aceptaremos los reactivos cuyo ajuste (INFIT o OUTFIT) esté entre -2 y +2 lógitos.

6.6 ¿Qué tan conveniente es el modelo de Rasch?

Esta pregunta se hace siempre en términos de qué tanto conviene en comparación con el modelo clásico (no visto en este libro) o con el modelo de tres parámetros (presentado en secciones anteriores).

En primer lugar conviene aclarar que el modelo clásico mide a los reactivos con un esquema diferente. En el modelo clásico no se interpreta la curva característica (nada lo impide por cierto), por lo que la discriminación no es función de la pendiente de la curva sino de otra forma de medirla: consisten generalmente de pruebas de diferencias de medias entre dos subgrupos de la población definidos de alguna manera. Por ello no se puede garantizar que los criterios empleados en el análisis clásico brindan resultados similares a los que emite el modelo de Rasch o los que se obtienen con el modelo de tres parámetros. El análisis clásico involucra a los distractores y obliga a estudiarlos, en los otros esquemas el análisis de distractores no está contemplado (aunque nada impediría hacerlo), ya que se trata de un estudio de comportamiento “global” del reactivo.

Si comparamos ahora el modelo de Rasch con los otros presentados para la curva característica, podemos ver que todos los reactivos que se parezcan suficientemente al modelo de Rasch serán aceptados también por el modelo de tres parámetros. En cambio, el modelo de tres parámetros acepta inclusive aquellos reactivos que Rasch rechaza.

¿Esto quiere decir que el modelo de tres parámetros es más benévolo? En realidad sí. Al evaluador le conviene más el modelo de tres parámetros porque “ajustará” en la gran mayoría de los casos, mientras que el modelo de Rasch no necesariamente ajustará y obligará al evaluador a revisar o desechar sus reactivos.

Pero también ocurren cosas muy curiosas con el modelo de tres parámetros, como la paradoja presentada en el capítulo 4. Esta paradoja, como ya se dijo, no se presenta en el Modelo de Rasch ya que las curvas jamás se cruzan. Recuérdese que las curvas del modelo de tres parámetros provienen de datos de la realidad y por ello las acepta cuando se consigue un buen ajuste de la ecuación al conjunto de puntos observados y la “evidencia empírica” muestra que este es el comportamiento del reactivo.

Cuando un reactivo se comporta como muestran las curvas de la figura se sabe que no se cumplen las hipótesis del modelo:

- unidimensionalidad (posiblemente esté midiendo más de un solo rasgo)
- orden e inclusividad (posiblemente el reactivo no tiene validez de contenido)
- posibilidad de respuestas al azar
- influencias externas en las respuestas

6.7 Cálculo del Modelo de Rasch.

El cálculo de la medida, de la calibración y de los valores de ajuste se realizan por medio de la computadora siguiendo los esquemas presentados hasta aquí. Se requiere de una computadora ya que la cantidad de operaciones se vuelve excesiva muy rápidamente. Estos cálculos no se facilitan por medio de tablas de EXCEL, por lo que no los haremos aquí.

Existen varios algoritmos que hacen el cálculo de los parámetros descriptivos de los reactivos y de las personas. Tratándose de diferentes métodos, ocurre que se puede llegar a valores diferentes para la calibración y para la medida. Aunque no hay una normalización de estos métodos, se prefieren estos dos:

1) Algoritmo PROX
2) Algoritmo UCON

El algoritmo PROX permite obtener estimadores iniciales bajo la hipótesis de que los reactivos y las personas se distribuyen normalmente. Este algoritmo hace una linealización de la escala latente que incluye un ajuste de los efectos locales de la distribución de habilidades de la muestra.

El algoritmo UCON se usa, generalmente, para refinar o mejorar los resultados obtenidos con PROX. En este algoritmo no se hace hipótesis alguna sobre la distribución de los reactivos o de las personas, por lo que pueden distribuirse en forma no normal.

Tanto PROX como UCON pueden utilizarse de manera exclusiva. Por ejemplo nada impide que sólo se trabaje con PROX y no solamente se emplee para obtener estimadores iniciales. Algunos autores postulan que uno es mejor que el otro, de nueva cuenta se trata de algoritmos basados en métodos iterativos que pueden ser convenientes para una muestra o un cuestionario en particular, pero no ser apropiado para otros casos. No hay justificación a priori respecto a la calidad de cada método, con excepción de la plausibilidad de las hipótesis que cada uno involucra.

No es propósito de este libro estudiar a detalle estos algoritmos, pero sí conviene saber que existen y que hay varios programas que se pueden comprar para realizar análisis de Rasch (BIGSTEPS, RASCAL, etc.)

El esquema de aplicación del modelo de Rasch en cualquier programa sigue estos pasos:

PREPROCESO

Construcción del Escalograma de Guttman y eliminación de reactivos extremos y personas extremas.

Para los reactivos: cálculo de

Puntajes

Medida en lógitos

Calibración inicial

Para las personas: cálculo de

Puntajes

Medida inicial en lógitos

Hacer cambio de escala y corrimiento

para ítems y personas

Calcular cantidades primarias:

Dificultad de los reactivos

Medida de las personas

Cálculo de cantidades secundarias:

INFIT OUTFIT

Error estándar Confiabilidad Separación

Estos pasos se efectúan en forma iterativa y el tiempo de cálculo depende de la velocidad de la computadora, de la eficiencia del algoritmo así como del tamaño del cuestionario y del número de personas que lo contestan.

6.8 Presentación de algunos resultados de BIGSTEPS

Sin entrar a detalles sobre el uso de un programa, se presentan a continuación parte de los reportes de salida del programa BIGSTEPS (de la Universidad de Chicago), sobre un cuestionario de 20 reactivos aplicado a un grupo de 46 personas. Los reactivos tratan sobre zoología y se identifican por medio de una descripción breve relativa al animal sobre el cual trata el reactivo.

Menciono que es una visión parcial de los resultados, ya que BIGSTEPS proporciona un gran número de reportes. Aquí sólo se van a presentar unos cuantos y se dará una interpretación general de los resultados.

6.8.1 Información del algoritmo

Un reporte ilustrativo de lo que se ha presentado aquí es el informe del algoritmo, donde BIGSTEPS muestra las iteraciones que realizó y datos sobre el pre-proceso.

TABLE 0.2 EJEMPLO DE EXAMEN NUMERO 1 Sep 19 14:55 1996
46 PERSONS 20 REACTIS

CONVERGENCE TABLE

PROX ITERATION	ACTIVE COUNT PERSONS	REACTIS	CATS	EXTREME 5 RANGE PERSONS	REACTIS	MAX LOGIT CHANGE MEASURES	STEPS
1	46	20	2	2.43	4.48	3.8067	
2	46	17	2	4.65	4.82	-2.4631	
3	46	17	2	4.85	5.62	.7460	
4	46	17	2	5.35	5.71	-.4479	
5	46	17	2	5.40	5.94	.2766	
6	46	17	2	5.55	5.97	-.1340	
7	46	17	2	5.57	6.04	.1460	

UCON ITERATION	MAX SCORE RESIDUAL*	MAX LOGIT CHANGE	LEAST CONVERGED PERSON	REACTI	CAT	CATEGORY RESIDUAL	STEP CHANGE
1	.84	-.4975	7*	11			
2	.60	.1088	15	10*			
3	.40	.0748	15	8*			
4	.30	-.0540	15	8*			
5	.23	-.0448	15	8*			
6	.18	-.0392	1	8*			
7	.15	-.0330	1	8*			
8	.12	-.0274	1	8*			
9	.10	-.0226	1	8*			
10	.08	-.0186	1	8*			
11	.06	-.0153	1	8*			
12	.05	-.0126	1	8*			
13	.04	-.0104	1	8*			
14	.04	-.0086	1	8*			

Obsérvese que como resultado del pre-proceso BIGSTEPS sólo conservó 17 de los 20 reactivos iniciales. Esto indica que hay 3 reactivos que no se pueden calibrar correctamente. El reporte indica asimismo el residuo máximo que se obtiene en los puntajes; nótese cómo este residuo es decreciente, lo cual hace que el proceso iterativo sea convergente. El máximo cambio en lógitos que se obtuvo en la última iteración es del orden de 8 milésimas, lo cual indica que los valores tienen una alta precisión.

6.8.2 Mapa de personas y reactivos

De acuerdo con la medida y la calibración, sabemos que se pueden referir los resultados a una misma métrica o escala.

```
TABLE 1.1 EJEMPLO DE EXAMEN NUMERO 1                          Sep 19 14:55 1996
    46 PERSONS  20 REACTIS ANALYZED:    46 PERSONS  17 REACTIS   2 CATEGORIES
--------------------------------------------------------------------------------
                         MAP OF PERSONS AND REACTIS
MEASURE                                                                MEASURE
          -------------------- PERSONS-+- REACTIS------------------
   5.0                                 + X                              5.0
                                       | X
   4.0                               X +                                4.0
                                   XXX | X
   3.0                                 + XX                             3.0
                                    XX | X
                                    XX |
   2.0                                 + X                              2.0
                                    XX | XX
   1.0                           XXXXX +                                1.0
                                       | X
                                 XXXXX |
    .0                                 +                                 .0
                         XXXXXXXXXXXXX |
  -1.0                            XXXX + X                             -1.0
                                   XXX |
  -2.0                                 +                               -2.0
                                    XX | XXX
                                    XX | X
  -3.0                                 + X                             -3.0
                                     X | X
  -4.0                                 +                               -4.0
  -5.0                                 +                               -5.0
                                     X | X
  -6.0                                 + XX                            -6.0
          -------------------- PERSONS-+- REACTIS------------------
```

En este mapa puede observarse que la distribución de personas se parece a una normal, mientras que los reactivos se ubican de manera uniforme.

Observamos también que hay zonas donde faltan reactivos (por ejemplo entre -0.5 y + 0.5) , así como otras donde hay acumulación de reactivos (por ejemplo entre -2 y -3.5). La escala tiene, por lo tanto, pequeños defectos que deberán depurarse en futuros exámenes.

6.8.3 Curva característica del examen

Al igual que la curva característica del reactivo se puede dibujar una para el cuestionario completo, que resulta de dibujar la medida contra la probabilidad de respuesta asociada a la Dificultad. BIGSTEPS prepara dos curvas del examen, la primera brinda las probabilidades de respuestas de una persona y la otra indica los reactivos que podrían ser contestados por un sustentante.

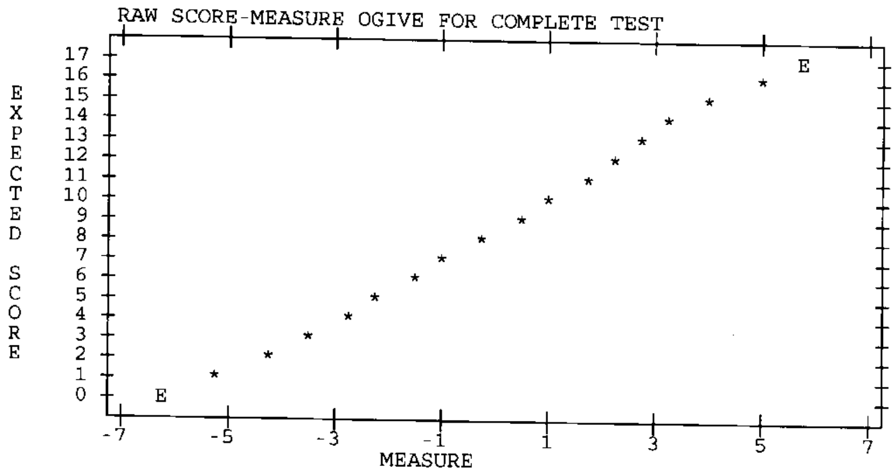
La primera curva presenta los posibles puntajes (desde 0 hasta 17) que pueden ser medidos por el cuestionario, junto con su medida en lógitos y el error estándar que se determinó en cada caso. Obsérvese que el error es más grande en los extremos que al centro, como ya se apuntó previamente.

Enseguida se tiene la curva con el puntaje esperado de las personas y las calibraciones posibles de los reactivos. Las letras M indican la media, la S una desviación estándar a cada lado de la media, la Q señala dos desviaciones estándar a cada lado de la media.

TABLE 20.1 EJEMPLO DE EXAMEN NUMERO 1 Sep 19 14:55 1996
46 PERSONS 20 REACTIS ANALYZED: 46 PERSONS 17 REACTIS 2 CATEGORIES

TABLE OF MEASURES ON COMPLETE TEST

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-6.20E	1.58	6	-1.59	.79	12	2.25	.73
1	-5.26	1.21	7	-.94	.83	13	2.78	.74
2	-4.14	.94	8	-.23	.85	14	3.36	.78
3	-3.39	.82	9	.48	.83	15	4.04	.88
4	-2.77	.77	10	1.12	.78	16	5.00	1.13
5	-2.19	.76	11	1.70	.75	17	5.84E	1.51




```
PERSON          1      1  2 2  3 4  1    5 5  2 2 2 3  1
                       Q       S    3    M    S        Q
                                    M
REACTI      2  1       11112     1     1   111 1 21    1     1
               Q       S             M           S            Q
```

La segunda curva se refiere a los reactivos. Esta es reportada por BIGSTEPS marcando en el extremo izquierdo valores de respuesta nula “0” y en el extremo derecho valores de acierto “1”. Entre ambos extremos se ubican los reactivos del cuestionario y marcan la zona que divide a la falla del acierto.

```
TABLE 2.1 EJEMPLO DE EXAMEN NUMERO 1                          Sep 19 14:55 1996
    46 PERSONS  20 REACTIS ANALYZED:    46 PERSONS  17 REACTIS   2 CATEGORIES
--------------------------------------------------------------------------------

MOST PROBABLE RESPONSE: MODE  (BETWEEN "0" AND "1" IS "0", ETC.)
 NUM REACTI      -6        -4        -2         0         2         4         6
                  |---------+---------+---------+---------+---------+---------|
 19 SAPO          0           .                  .                  .   1     1
 18 RANA          0           .                  .                1 .         1
 16 TIGRE         0           .                  .              1   .         1
 17 CEBRA         0           .                  .              1   .         1
 15 TITI          0           .                  .            1     .         1
 14 MONO          0           .                  .          1       .         1
 12 PUMA          0           .                  .        1         .         1
 13 CHOFER        0           .                  .       1          .         1
 11 RENO          0           .                  .  1               .         1
 10 DODO          0           .             1    .                  .         1
  6 CHACAL        0           .       1          .                  .         1
  8 YAK           0           .       1          .                  .         1
  9 PLATIPUS      0           .      1           .                  .         1
  7 ELEFANTE      0           .    1             .                  .         1
  5 POLLO         0           .   1              .                  .         1
  4 LINCE         0           .1                 .                  .         1
  3 CABALLO       0  1        .                  .                  .         1
                  |---------+---------+---------+---------+---------+---------|
                 -6        -4        -2         0         2         4         6

                                               1
PERSON                1        1  2  2  3  4   3  5   5 2  2  2  3  1
                              Q        S         M        S        Q
```

Esta curva es interesante porque permite estimar la probabilidad de respuesta de una persona de medida B ante el examen aplicado. Por ejemplo, para B=2 se tiene una probabilidad de contestar el reactivo asociado al “Puma”, esto quiere decir que el sujeto debería contestar correctamente las preguntas que están por abajo del “Puma”. Si no las contesta se tratará de una falla inesperada. Por el contrario, no debería poder contestar a las preguntas por arriba de “Puma”, a riesgo de decir que se trata de una cierto inesperado.

6.8.4 Estadísticas de las personas

Se trata de una tabla donde se presentan las 46 personas que contestaron el cuestionario ordenadas alfabéticamente, por medida o por ajuste (INFIT o OUTFIT). En este caso se presenta la salida ordenada por OUTFIT porque nos interesa saber qué alumnos están mejor medidos, cuáles tienen mayor “ruido” o, por el contrario, son “demasiado deterministas” en su patrón de respuestas.

TABLE 6.1 EJEMPLO DE EXAMEN NUMERO 1 Sep 19 14:55 1996
46 PERSONS 20 REACTIS ANALYZED: 46 PERSONS 17 REACTIS 2 CATEGORIES

PERSON STATISTICS: OUTFIT ORDER

ENTRY NUM	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT STD	OUTFIT MNSQ	OUTFIT STD	PTBIS	PERSON
43	7	17	-.93	.83	1.64	1.0	7.08	1.9	A .61	DONALD
44	13	17	2.78	.74	2.43	3.4	9.90	1.5	B-.01	RAMIRO
45	14	17	3.35	.78	2.28	2.7	9.90	1.1	C-.11	MAGDA
46	10	17	1.11	.78	2.64	2.5	4.76	1.1	D .33	ALICIA
38	15	17	4.03	.88	1.11	.2	9.90	1.0	E .03	GOMI
36	6	17	-1.58	.79	2.34	2.3	4.29	1.0	F .35	NABUCO
41	14	17	3.35	.78	1.52	1.3	9.90	.8	G .15	FEDERICO
29	8	17	-.23	.85	2.23	1.5	1.94	.6	H .58	ROMULO
13	8	17	-.23	.85	2.18	1.5	1.75	.5	I .59	MELISA
39	14	17	3.35	.78	1.28	.7	.76	.0	J .42	GUMER
33	4	17	-2.76	.77	1.11	.4	.64	-.1	K .55	MARTITA
35	1	17	-5.25	1.21	.38	-1.0	.08	-.1	L .33	DANIEL
25	3	17	-3.38	.81	.96	-.1	.42	-.2	M .51	DAMIAN
24	12	17	2.24	.73	1.08	.2	.54	-.2	N .64	LORENZA
37	13	17	2.78	.74	.54	-1.8	.26	-.3	O .64	WALTER
4	4	17	-2.76	.77	.78	-.8	.34	-.3	P .63	ESTER
17	7	17	-.93	.83	.84	-.3	.62	-.3	Q .83	OCTAVIO
7	12	17	2.24	.73	.62	-1.4	.31	-.3	R .70	AGUSTIN
11	6	17	-1.58	.79	.81	-.5	.52	-.3	S .78	TOBIAS
3	8	17	-.23	.85	1.02	.0	.63	-.3	T .80	LUISA
34	10	17	1.11	.78	.99	.0	.47	-.4	U .77	ELSA
23	10	17	1.11	.78	.94	-.1	.45	-.4	V .78	EDUARDO
1	5	17	-2.18	.76	.68	-1.1	.30	-.4	W .72	PEPE
27	5	17	-2.18	.76	.68	-1.1	.30	-.4	X .72	ALFREDO
12	6	17	-1.58	.79	.92	-.2	.41	-.4	Y .76	SANDRA
15	11	17	1.69	.75	.51	-1.7	.25	-.4	Z .79	CESAR
42	11	17	1.69	.75	.51	-1.7	.25	-.4	t .79	DIEGO
22	10	17	1.11	.78	.53	-1.2	.26	-.6	s .85	CARLITOS
21	10	17	1.11	.78	.48	-1.4	.23	-.6	r .86	PEDRO
18	9	17	.47	.83	.65	-.7	.31	-.7	q .87	LUPITA
20	9	17	.47	.83	.65	-.7	.31	-.7	p .87	COCO
16	8	17	-.23	.85	.74	-.5	.32	-.7	o .87	ELOY
28	8	17	-.23	.85	.74	-.5	.32	-.7	n .87	ELEUTERIO
10	9	17	.47	.83	.33	-1.7	.16	-.9	m .93	NINO
14	9	17	.47	.83	.33	-1.7	.16	-.9	l .93	ARACELI
32	9	17	.47	.83	.33	-1.7	.16	-.9	k .93	ELENA
19	7	17	-.93	.83	.33	-1.7	.16	-1.0	j .91	ROSITA
30	7	17	-.93	.83	.33	-1.7	.16	-1.0	i .91	LISA
2	8	17	-.23	.85	.22	-2.0	.12	-1.2	h .96	MARCOS
5	8	17	-.23	.85	.22	-2.0	.12	-1.2	g .96	PACO
6	8	17	-.23	.85	.22	-2.0	.12	-1.2	f .96	MEMO
8	8	17	-.23	.85	.22	-2.0	.12	-1.2	e .96	SAMUEL
9	8	17	-.23	.85	.22	-2.0	.12	-1.2	d .96	REBE
26	8	17	-.23	.85	.22	-2.0	.12	-1.2	c .96	BENITO
31	8	17	-.23	.85	.22	-2.0	.12	-1.2	b .96	CHAGO
40	8	17	-.23	.85	.22	-2.0	.12	-1.2	a .96	LUGO
MEAN	9.	17.	.12	.82	.87	-.5	1.53	-.3		
S.D.	3.	0.	1.88	.07	.67	1.4	2.90	.8		

En esta tabla aparecen estas columnas:

- Número del sujeto
- Puntaje bruto (número de aciertos bruto)
- Número de reactivos sobre el cual se trabajó (quitando los reactivos extremos que no pueden calibrarse)
- Medida en lógitos (Es el parámetro B del modelo)
- Error
- Valores de INFIT y OUTFIT (Media cuadrática y residuo estandarizado)

- Correlación punto-biserial (no usada normalmente para dictaminar a las personas, pero muy utilizada en el dictamen de los reactivos para estimar una posible discriminación)

- Nombre de la persona

El reporte se pidió ordenado por OUTFIT, para ver fácilmente si hay personas con mucho o poco azar en su patrón de respuestas. Los valores de cotejo +2 y -2. Todas las personas están dentro del intervalo de aceptación.

No ocurre lo mismo con el INFIT, donde podemos apreciar que hay varias personas cuyo INFIT es superior a +2 y otras que están en el límite de -2. Por ejemplo Ramiro (ubicado en el segundo lugar de la tabla) tiene una medida alta (2.78 lógitos), su OUTFIT es 1.5, mientras que el INFIT es de 3.4.

Recuérdese que:

- el INFIT es sensible al comportamiento inesperado que afecta a los reactivos cercanos al nivel de habilidad de una persona.

- el OUTFIT es sensible al comportamiento inesperado que afecta a los reactivos lejanos del nivel de habilidad de una persona

En todo caso Ramiro tiene comportamiento con ruido desde el punto de vista del INFIT, posiblemente tiene muchas respuestas inesperadas en los reactivos difíciles, que se supone debe poder contestar ya que su medida es alta.

6.8.5 Respuestas inesperadas

Un reporte interesante es el de las respuestas inesperadas, ya que puede apoyar las conclusiones que se preparan sobre los reactivos o las personas. Esto sólo lo presenta en los casos de personas que tuvieron problemas de desajuste y señala por medio de una X las respuestas inesperadas. Aquí volvemos a encontrar a Ramiro. Sus respuestas están señaladas en un primer renglón y se señalan abajo los valores de residuo de cada respuesta respecto a su medida. El programa señala que hay un reactivo por arriba de 2 lógitos que la persona no contestó.

```
TABLE 7.1 EJEMPLO DE EXAMEN NUMERO 1                              Sep 19 14:55 1996
    46 PERSONS  20 REACTIS ANALYZED:    46 PERSONS  17 REACTIS   2 CATEGORIES
------------------------------------------------------------------------------------
TABLE OF POORLY FITTING PERSONS   (REACTIS IN ENTRY ORDER)
NUMBER - NAME -- POSITION ------- MEASURE - INFIT (STD) OUTFIT - MISFIT OVER 2.0

   44 RAMIRO                         2.78     3.4   B     1.5
       RESPONSE:    1:  1 1 1 0 1  1 1 1 1 0   1 1 0 1 1  1 1 0 1 0
       RESIDUAL:        X X  -9              -6                    2 X

   45 MAGDA                          3.35     2.7   C     1.1
       RESPONSE:    1:  1 1 1 1 1  1 0 1 1 0   1 1 1 1 0  1 1 1 1 0
       RESIDUAL:        X X                -9    -8                   X

   46 ALICIA                         1.11     2.5   D     1.1
       RESPONSE:    1:  1 1 1 1 1  0 1 1 0 1   1 0 0 1 1  0 0 1 0 0
       RESIDUAL:        X X          -5    -5               2      3  X
```

6.8.6 Resultados por reactivo

Del mismo modo que el reporte de las personas, se tiene uno especial para los reactivos. Los elementos de esta tabla son similares a los descritos en 6.8.4.

TABLE 10.1 EJEMPLO DE EXAMEN NUMERO 1 Sep 19 14:55 1996
46 PERSONS 20 REACTIS ANALYZED: 46 PERSONS 17 REACTIS 2 CATEGORIES

REACTI STATISTICS: OUTFIT ORDER

ENTRY NUM	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT STD	OUTFIT MNSQ	OUTFIT STD	PTBIS	REACTI
7	40	46	-2.71	.53	1.38	1.0	9.90	4.8	A-.06	ELEFANTE
6	38	46	-2.21	.47	1.20	.7	6.74	2.7	B .18	CHACAL
10	31	46	-.98	.38	1.06	.3	3.16	2.5	C .28	DODO
4	42	46	-3.37	.63	.96	-.1	9.90	2.0	D .21	LINCE
14	10	46	2.10	.44	1.10	.4	1.52	.4	E .36	MONO
3	45	46	-5.39	1.15	1.60	.6	1.90	.1	F .03	CABALLO
9	39	46	-2.45	.50	.77	-.8	1.01	.0	H .44	PLATIPUS
	BETTER	FITTING	OMITTED							
19	2	46	4.56	.78	.99	.0	.25	-.3	M .29	SAPO
12	12	46	1.73	.42	.90	-.4	.73	-.3	N .49	PUMA
18	5	46	3.31	.55	.66	-1.1	.34	-.4	O .49	RANA
5	41	46	-3.01	.57	.90	-.3	.42	-.5	P .41	POLLO
16	6	46	3.02	.52	.52	-1.9	.21	-.7	Q .59	TIGRE
17	6	46	3.02	.52	.52	-1.9	.21	-.7	R .59	CEBRA
8	38	46	-2.21	.47	.60	-1.7	.37	-.8	S .55	YAK
11	20	46	.52	.37	.77	-1.5	.56	-1.0	T .57	RENO
MEAN	23.	46.	.00	.54	.93	-.4	2.31	.5		
S.D.	16.	0.	2.87	.18	.28	.9	3.17	1.5		

Los reactivos cuyo comportamiento se considera correcto se omiten del reporte. Aquí sólo se presentan los reactivos que tiene “ruido” o los que son “mudos”. El reactivo que hace referencia al Elefante tiene un OUTFIT excesivo (4.8 lógitos), su medida es -2.71, se trata por lo tanto de un reactivo muy fácil que no ajusta bien al modelo de Rasch. Puede cotejarse la facilidad revisando que lo contestaron 40 personas del total de 46 alumnos.

Desde el punto de vista de la correlación punto-biserial, el reporte de BIGSTEPS intenta dar un estimado de la discriminación del reactivo (siempre y cuando el usuario esté de acuerdo en que la correlación punto-biserial es un buen estimador de la discriminación). El estudio de este tema cae fuera del propósito de este libro, pero sí podemos observar por el momento que el reactivo del Elefante tiene una correlación negativa (-0.06), lo cual indica que el reactivo se comporta de manera contraria a lo esperado (ver capítulo 3).

Este vistazo sobre los resultados de BIGSTEPS no son más que una breve parte de lo que puede analizarse con ayuda del Modelo de Rasch.

7. CONCLUSIONES

Hasta aquí se llega en este libro sobre aspectos introductorios del análisis de Rasch. El propósito del libro ha sido presentar los fundamentos sobre los que se basa el modelo de Rasch y que sirva de preámbulo a estudios más específicos sobre el tema, donde los temas vistos aquí se manejan como antecedentes.

Queda mucho por conocer de este modelo de evaluación. Los temas de confiabilidad, separación, error estándar, validez, igualación de formas, uso de anclas, normalización y diseño de pruebas, son tan sólo algunos de los aspectos que no hemos explorado aquí, pero que pueden ser encontrados en las referencias.

Espero haber despertado en el lector el interés de saber cómo se presentan los resultados en un programa, cómo se deben interpretar, de qué manera emplear los resultados para retroalimentar a una población o a los que construyen el cuestionario. Todos ellos son aspectos medulares que requieren de más detalle y que pueden ser estudiados en los textos de las referencias, principalmente en “Best Test Design” (Wright y Stone), así como en el manual de Bigsteps (Mesa Press) o en los apuntes del Seminario de Análisis de Rasch (Tristán).

En el capítulo 6 se presentó un pequeño ejemplo de los resultados que se obtienen con el Análisis de Rasch aprovechando el programa BIGSTEPS. El uso de la computadora es muy necesario en este tipo de análisis y por esta razón no se incluyeron ejercicios propuestos para el lector. Es muy probable que desee adquirir pronto una herramienta para hacer sus propios trabajos exploratorios, que le resultarán de mucha utilidad para emprender posteriormente estudios más serios sobre evaluación utilizando el **Modelo de Rasch**. Este libro, introductorio para los fundamentos del Análisis de Rasch para todos los evaluadores educativos, fue destinado a poner punto y aparte en el lugar donde otros autores empiezan. Es así que queda mucho por conocer de este modelo de evaluación.

El lector ya habrá podido comprobar que el propósito del libro ha sido presentar los fundamentos sobre los que se basa el Modelo de Rasch y, de paso, comentar algunos puntos importantes acerca de evaluación en general y de la **Teoría de la Respuesta al Reactivo**. Otros tratados de Rasch y de medición consideran estos temas como antecedentes, por lo que es muy común que el evaluador que se inicia en este asunto se enfrente a ciertas carencias y a lagunas informativas que son a veces muy complicadas para ser resueltas al paso.

Deseo recordar aquí que buena parte de los temas tratados fueron atacados en forma aproximada, más con intención didáctica que con afán de formalizar una teoría. Las referencias se encargarán de fundamentar apropiadamente lo que pueda considerarse que ha sido presentado en forma deficiente o superficial en este libro.

Algunos puntos son medulares y conviene recordarlos aquí:

- El modelo de Rasch es un **modelo probabilista** que permite estimar la probabilidad de respuesta de un sujeto ante un estímulo o reactivo dado.

- El modelo considera que se mide una variable con características de **inclusividad, ordenamiento y unidimensionalidad**.

- El modelo considera que, a partir de la información disponible, sólo es posible disponer de un estimador de la distancia entre la **medida** de una persona y la **calibración** de un reactivo. Por ello se acostumbra hablar de **modelo de un parámetro**.

- El modelo es una función representada como una **curva logística** que involucra a los mismos tres parámetros de otros modelos, con estas salvedades:
  - ◊ Sólo se determina la **Dificultad** del reactivo (en términos del **momio** de falla o respuestas incorrectas), correspondiendo al **punto de inflexión** donde la probabilidad de respuesta es igual a 0.5
  - ◊ La **Discriminación** en los modelos logísticos se mide en términos de la **pendiente de la curva** en el punto de inflexión. La curva de Rasch tiene una pendiente de valor constante en dicho punto de inflexión, para cualquier valor de la Dificultad. Se habla por ello que todos los reactivos discriminan de la misma manera.
  - ◊ La curva de Rasch involucra una ordenada nula para los sujetos de dominio mínimo o nulo. Este modelo implica que no hay **adivinación sistemática**

- El modelo no hace hipótesis globales sobre el patrón de respuestas de un grupo de sujetos ni sobre la adivinación sistemática. Por el contrario, postula que la **adivinación es un patrón personal** y puede ser detectada en función del ajuste o desajuste (**FIT**) de las respuestas de una persona al modelo de la curva de Rasch. Igualmente se hace la hipótesis para los reactivos.

- En la Teoría de la Respuesta al Reactivo y en el Modelo de Rasch se plantea la necesidad de conocer el **error de la medida** de las personas y de los reactivos, con relación a una curva teórica o ajustada. Este error se mide por medio de un valor de **ajuste** y puede expresarse en términos de **error absoluto**, **residuo estandarizado**, **residuo cuadrático medio** , también llamado **OUTFIT** y el **INFIT**

- Los modelos de tres parámetros y de Rasch suponen **una independencia de la medida**. Esto quiere decir que la medida de una persona debe ser independiente del instrumento empleado (cuestionario) y la calibración del reactivo debe ser independiente de la población que lo resuelve.

- El modelo de Rasch contempla la necesidad de hacer un **corrimiento de las medidas** respecto de las calibraciones, con objeto de que la independencia de medida se alcance y pueda dictaminarse el punto de la escala donde se encuentra un sujeto, independientemente del instrumento aplicado.

- El Modelo de Rasch es un **modelo de contraste**, contrariamente al modelo de tres parámetros y otros más que son **modelos de ajuste**. En el caso del Modelo de Rasch se propone un modelo teórico de comportamiento deseable de un reactivo y contra él se analiza el ajuste de los puntos. En los modelos de ajuste se trata de buscar una curva que pase lo más cercanamente a los puntos experimentales.

Algunos otros aspectos no son privativos del Análisis de Rasch, pero han sido fundamentales; tal es el caso de la **escala**, la cual debe llenar ciertos requisitos para ser útil. Vimos la conveniencia de que la escala tenga suficiente **extensión** para poder medir a todos los sujetos, las "marcas" deben estar ubicadas con un **espaciamiento uniforme** y en una base de **medida lineal**. Con esto se pretende que los exámenes no midan un mismo rasgo con varios reactivos, sino que se aprovechen los estímulos a diferentes dificultades para poder emitir un mejor dictamen de la ubicación de los sujetos. Estas características deseables de la escala no deben confundirse con el Modelo de Rasch, ya que se trata de recomendaciones que deberíamos de seguir independientemente del modelo elegido.

Uno de los tópicos tratados con mayor interés es el concepto de **lógito o logaritmo natural del momio**, como una nueva unidad de medida (momio de aciertos para la medida de los sujetos, momio de fallas para la calibración de los reactivos). Es cierto que el lógito no es cómodo para ser transmitido a los profesionales que se inician en este tema de la evaluación, pero resulta todavía más incómodo cuando se trata de presentar a los mismos alumnos o sustentantes, a los padres de familia o a la sociedad en general. Hay un esfuerzo muy importante por traducir esta unidad a valores más "humanos", por lo que se han inventado nuevas unidades y escalas, como los NITS y los CHIPS que se presentan en la referencia de Wright y Stone y que no fueron estudiados en este libro. Creo que todo es cuestión de más tiempo y de que nos acostumbremos todos a una nueva forma de expresión. El lógito no es una unidad exclusiva del Modelo de Rasch, puesto que se emplea en otras aplicaciones de la evaluación, pero sí forma parte indisociable del concepto definido por este modelo.

No debe confundirse el Modelo de Rasch con el **Escalograma de Guttman**, que es una herramienta para ordenar a los sujetos y a los reactivos en un intento de buscar "patrones" de respuesta. El Escalograma se utiliza en general para "arreglar" los datos en el **pre-proceso**, esto es aprovechado por el Modelo de Rasch para eliminar a los **sujetos y reactivos extremos**.

Es usual que los evaluadores contrapongan el Análisis de Rasch **al Modelo de tres parámetros**, así como lo hacen antagónico del **Modelo Clásico** (grado de dificultad, poder de discriminación, análisis de distractores). Considero que es un error tomar un partido extremista. Puede obtenerse mucho provecho con los aspectos positivos que tiene cada uno de los modelos, sin pretender encontrar coincidencias, ya que cada uno de ellos estudia diferentes aspectos de la evaluación. Más bien deberíamos verlos como modelos complementarios. Se dice que el modelo clásico está basado en **hipótesis débiles**, pero puede fundamentarse suficientemente bien y servir para los propósitos de la evaluación con **hipótesis fuertes** como las del modelo de Rasch. El modelo de KALT, empleado en varios ambientes educativos de México, es una prueba del buen uso que puede hacerse con el Modelo Clásico.

Quedan, sin embargo, muchos puntos pendientes, que son tratados por otros autores y que se presentarán en "Construcción de Cuestionarios con el Modelo de Rasch", que es la continuación de este libro. Los temas de confiabilidad, separación, error estándar, validez, corrimiento, igualación de formas, construcción de escalas, uso de anclas, normalización y diseño de pruebas referidas a norma o referidas a criterio, son tan sólo algunos de los aspectos sobre los que no he hecho mención aquí y que se estudian en el siguiente libro. En compendios más especializados podrá el lector interesado estudiar estos temas, así como profundizar en los algoritmos que sólo fueron citados en el capítulo 6: **PROX** y **UCON**, además de otros algoritmos disponibles en la literatura especializada o en programas comerciales.

Puede verse que el Análisis de Rasch es un vehículo muy útil para la retroalimentación y el mejoramiento de la calidad de un instrumento que debe usar el especialista de la evaluación, sin tener que estar vedado a otros profesionales.

He tratado de que el lenguaje fuera simple y el tratamiento de los temas fuera lo más didáctico posible, con ello espero haber conseguido que muchos colegas se sientan atraídos por la herramienta; además espero haberles quitado el miedo de enfrentarse a un análisis de Rasch para su trabajo cotidiano y, sobre todo, el pavor de enfrentarse a otros evaluadores más experimentados.

Por último espero que el libro haya resuelto algunas dudas sobre el uso y alcances de este tipo de análisis. El Análisis de Rasch es una buena herramienta, pero no la única. Ojalá el lector pueda tomar sus propias conclusiones y adoptar la mejor postura ante este tipo de técnicas y usarlas con propiedad, en su justa medida.

APENDICES

APÉNDICE 1 RECORDATORIO DE ALGUNAS PROPIEDADES DE LOS LOGARITMOS

Esta sección tiene por objeto apoyar a lector que no acostumbra el uso de los logaritmos. Puede ser omitida sin problema por las personas que dominen el tema.

1.1 Definición de logaritmo

Antes de ofrecer una definición, revisemos estos datos que todos conocemos:

$10^1 = 10$

$10^2 = 100$

$10^3 = 1000$

El exponente al cual estamos elevando el número 10 se llama "logaritmo". Tan sencillo como eso. Siempre debe indicarse la base de referencia, el número 10 en este caso.

Así decimos que el logaritmo (base 10) de 100 es 2. El logaritmo (base 10) de 1000 es 3. El logaritmo no es más que el exponente al que hay que elevar una base para obtener algún número deseado.

Cuando el número que estamos buscando es una potencia de 10 (por ejemplo 10, 100, 1000, etc.), el logaritmo es un número natural.

¿Cuál es el logaritmo de 200?

De acuerdo con lo que estamos presentando y sabiendo que 200 es un número que está entre 100 y 1000, podemos suponer que el logaritmo será un número mayor que 2 y menor que 3. Antes se calculaban los logaritmos con ayuda de tablas, pero ahora se pueden usar las calculadoras y podemos pedir log(200). El número que vamos a obtener es:

$\log(200)=2.3010$

Se dice que 2.3010 es el logaritmo (base 10) de 200. Esto quiere decir que si elevamos el número 10 a la potencia 2.3010 obtendremos 200. Verifíquelo.

1.2 La base de los logaritmos

Es común trabajar con los logaritmos base 10 (llamados logaritmos decimales) y también con los logaritmos naturales cuya base es el número e=2.718182... Se acostumbra denotar el logaritmo decimal como "log" y el logaritmo natural como "ln"

No vamos a justificar en este libro la presencia del número e, pero baste con saber que es otra base usual y que permite trabajar ciertas funciones útiles en evaluación (denominadas funciones exponenciales).

Se puede solicitar a la calculadora el logaritmo natural de 200.

$\ln(200)= 5.2983...$

Si elevamos el número e a la potencia 5.2983 obtendremos 200.

1.3 Algunas propiedades de los logaritmos

En este libro usaremos algunas propiedades que no serán demostradas (para ver las demostraciones, que se fundamentan simplemente en las leyes de los exponentes, sugerimos al lector que consulte libros especializados de álgebra). Estas propiedades son igualmente válidas para logaritmos decimales y para logaritmos naturales. Los ejemplos se harán exclusivamente con logaritmos naturales.

a) ln(XY)=ln(X) + ln(Y)

(el logaritmo de un producto es la suma de los logaritmos)

Ejemplo:

Verificar que ln(6)=ln(3x2)=ln(3)+ln(2)

con ayuda de una calculadora obtenemos:

ln(6)= 1.791759...
ln(3)= 1.098612...
ln(2)= 0.693147...

por lo tanto ln(3)+ln(2) = 1.098612+0.693147 = 1.791759... = ln(6)

como queríamos verificar

b) ln(X/Y)=ln(X)-ln(Y)

(el logaritmo de un cociente es la diferencia de los logaritmos)

Ejemplo.

Verificar que ln(6)=ln(24/4)=ln(24)-ln(4)

con ayuda de una calculadora obtenemos:

ln(6) =1.791759...
ln(24)=3.178053...
ln(4) =1.386294...

por lo tanto ln(24)-ln(4) = 3.178053-1.386294 = 1.791759... = ln(6)

como queríamos verificar.

c) ln(X/Y)=-ln(Y/X)

(el logaritmo de un número es igual al logaritmo del recíproco con signo contrario)

Esta propiedad se demuestra a partir de la propiedad (b):

ln(X/Y)=ln(X)-ln(Y)= -(ln(Y)-ln(X)) = - ln(Y/X)

Ejemplo

Verificar que $\ln(0.7/0.2) = -\ln(0.2/0.7)$

$\ln(0.7/0.2)=\ln(3.5)=1.2527629...$

$\ln(0.2/0.7)=\ln(0.285714...) = -1.2527629...$

lo que verifica la propiedad.

d) ln(1)=0

Use una calculadora para verificar esta propiedad

e) El logaritmo corresponde con la función inversa de la exponencial. La función inversa también se llama tradicionalmente "antilogaritmo".

Dado

$X=\ln(Y)$

$Y=e^X$

por lo que

$X=\ln(e^X)$

Ejemplo:

Verificar que $3= \ln(e^3)$

Primero calculemos e^3 con ayuda de una calculadora.

$e^3= 20.08553692319$

ahora calculemos

$\ln(20.08553692319) = 3$

lo que verifica la propiedad pedida.

Ejemplo:

Resolver para X, si se sabe que:

$\ln(X)=B-D$

tomando la función exponencial de ambos lados de la igualdad vemos:

$e^{(\ln(X))} = e^{(B-D)}$

por la propiedad (e) se tiene:

$X = e^{(B-D)}$

que es la expresión buscada.

APÉNDICE 2 USO DE EXCEL PARA APLICACIONES EN ESTE LIBRO

Algunos ejercicios sugieren el uso de EXCEL para poder hacer cálculos y representar los valores obtenidos en forma gráfica. No se pretende en este libro dar un curso de EXCEL, por lo que recomiendo al lector acudir al manual o a la ayuda de EXCEL para conocer el funcionamiento del programa. Aquí sólo se comentará el uso de un par de funciones para agilizar el trabajo.

2.1 Suma de celdas.

En algunos casos necesitamos sumar el contenido de las celdas (renglones o columnas) de una tabla. EXCEL hace este cálculo en forma automática. Para ello basta con hacer estos pasos:

a) Seleccione el renglón (o columna) que se va a sumar. Para ello presiónese el botón izquierdo del ratón para iluminar las celdas deseadas (quedan marcadas en color negro).
b) Presione el botón S en la barra de herramientas

Si se selecciona un renglón el resultado de la suma se coloca a la derecha de las celdas elegidas. Si se selecciona una columna, el resultado aparecerá abajo del último renglón.

2.2 Ordenamiento de celdas

EXCEL permite hacer ordenamientos simples de celdas (renglones o columnas). Háganse estos pasos:

a) Seleccione todas las celdas a ordenar con ayuda del botón izquierdo del ratón.
b) Elegir Datos en el Menú Principal (línea superior de EXCEL)
c) Dentro del submenú de Datos, elegir “Ordenar”
d) Elegir si se desean ordenar renglones o columnas (por omisión EXCEL ordena filas o renglones).

   presionando “Opciones” dentro de la ventana de “Ordenar”, y elegir “ordenar filas” u “ordenar columnas”
e) Elegir criterio de ordenamiento. Por ejemplo COLUMNA F o la columna que sirva de referencia para el ordenamiento.

   Indicar si se desea el ordenamiento ascendente o descendente.
f) Terminar el comando con el botón “Aceptar”

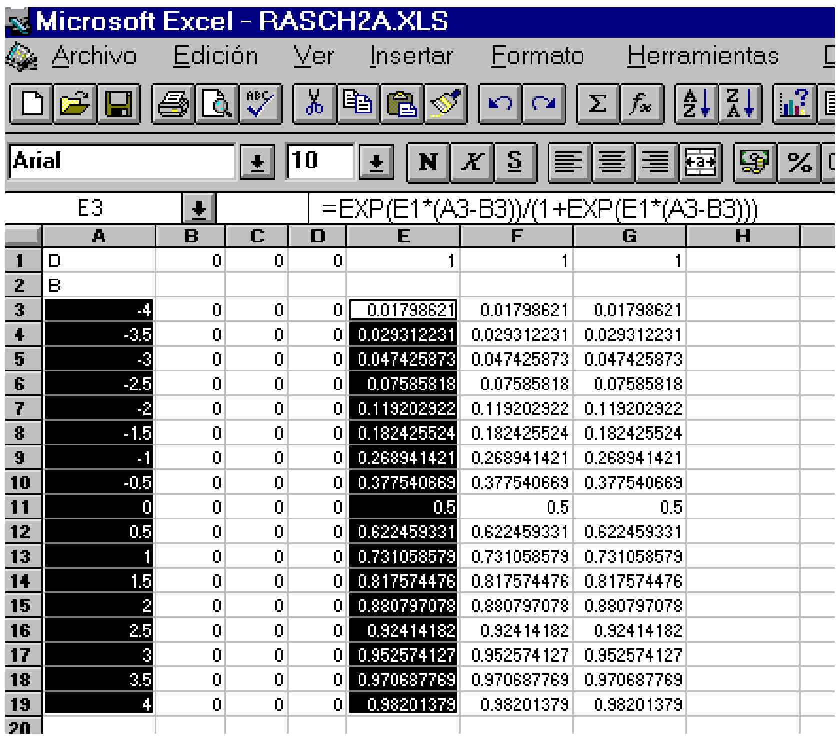
2.3 Presentación gráfica

Es posible hacer algunas gráficas con ayuda de EXCEL. Siga estos pasos:

a) Dada una tabla elija los datos que corresponden con el eje X, seleccionando todas las celdas con ayuda del botón izquierdo del ratón

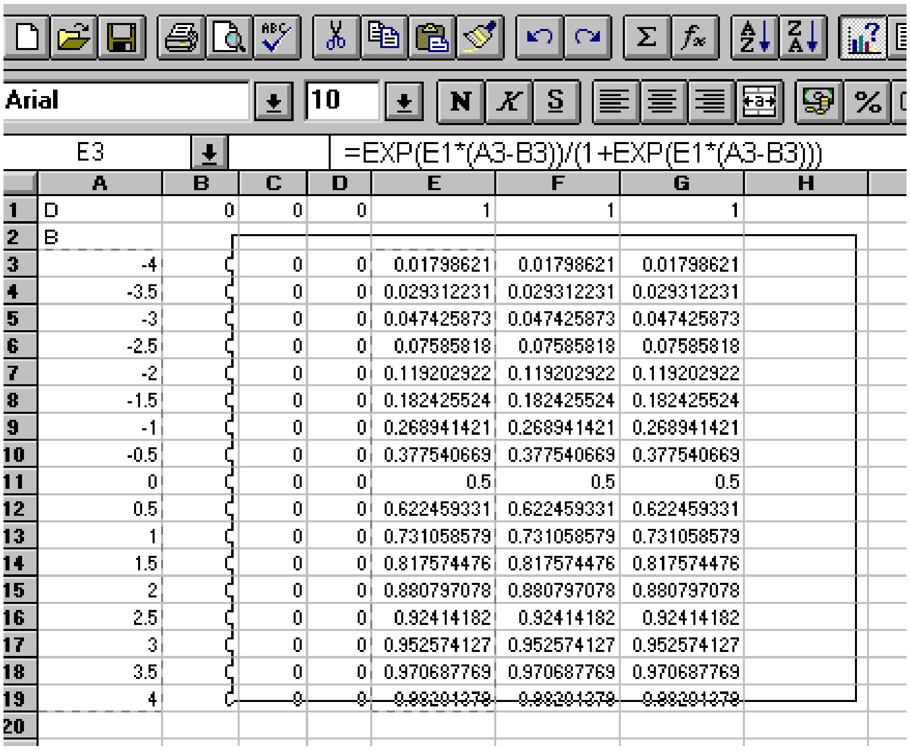
b) Elija los datos del eje Y con el mismo botón izquierdo.

Nota: si los datos de las variables X,Y están en columnas contiguas (por ejemplo las columnas C y D), basta con hacer la selección de todas las celdas con ayuda del ratón. En cambio si se trata de dos columnas separadas (por ejemplo las columnas A y E), debe seleccionarse primero la columna del eje X y una vez iluminada, presionar la tecla “Control” en el teclado y elegir la siguiente columna con el ratón, manteniendo presionada dicha tecla “Control”



	A	B	C	D	E	F	G	H
1	D	0	0	0	1	1	1	
2	B							
3	-4	0	0	0	0.01798621	0.01798621	0.01798621	
4	-3.5	0	0	0	0.029312231	0.029312231	0.029312231	
5	-3	0	0	0	0.047425873	0.047425873	0.047425873	
6	-2.5	0	0	0	0.07585818	0.07585818	0.07585818	
7	-2	0	0	0	0.119202922	0.119202922	0.119202922	
8	-1.5	0	0	0	0.182425524	0.182425524	0.182425524	
9	-1	0	0	0	0.268941421	0.268941421	0.268941421	
10	-0.5	0	0	0	0.377540669	0.377540669	0.377540669	
11	0	0	0	0	0.5	0.5	0.5	
12	0.5	0	0	0	0.622459331	0.622459331	0.622459331	
13	1	0	0	0	0.731058579	0.731058579	0.731058579	
14	1.5	0	0	0	0.817574476	0.817574476	0.817574476	
15	2	0	0	0	0.880797078	0.880797078	0.880797078	
16	2.5	0	0	0	0.92414182	0.92414182	0.92414182	
17	3	0	0	0	0.952574127	0.952574127	0.952574127	
18	3.5	0	0	0	0.970687769	0.970687769	0.970687769	
19	4	0	0	0	0.98201379	0.98201379	0.98201379	
20								

c) Una vez elegidas las columnas a graficar, elegir el botón “Asistente para gráficos” en el menú general de EXCEL, que hace que aparezcan unas pequeñas barras a un costado del cursor en la pantalla.

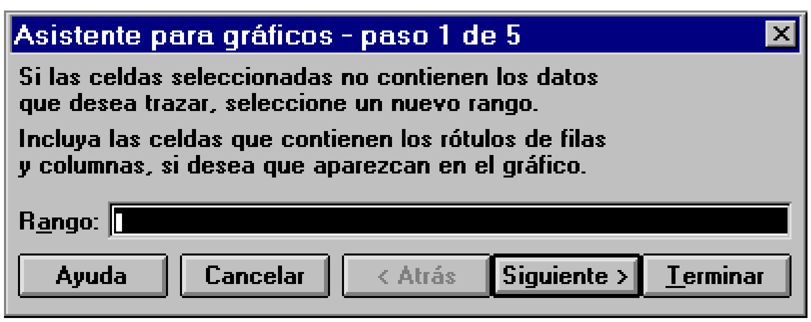


	A	B	C	D	E	F	G	H
1	D	0	0	0	1	1	1	
2	B							
3	-4	0	0	0	0.01798621	0.01798621	0.01798621	
4	-3.5	0	0	0	0.029312231	0.029312231	0.029312231	
5	-3	0	0	0	0.047425873	0.047425873	0.047425873	
6	-2.5	0	0	0	0.07585818	0.07585818	0.07585818	
7	-2	0	0	0	0.119202922	0.119202922	0.119202922	
8	-1.5	0	0	0	0.182425524	0.182425524	0.182425524	
9	-1	0	0	0	0.268941421	0.268941421	0.268941421	
10	-0.5	0	0	0	0.377540669	0.377540669	0.377540669	
11	0	0	0	0	0.5	0.5	0.5	
12	0.5	0	0	0	0.622459331	0.622459331	0.622459331	
13	1	0	0	0	0.731058579	0.731058579	0.731058579	
14	1.5	0	0	0	0.817574476	0.817574476	0.817574476	
15	2	0	0	0	0.880797078	0.880797078	0.880797078	
16	2.5	0	0	0	0.92414182	0.92414182	0.92414182	
17	3	0	0	0	0.952574127	0.952574127	0.952574127	
18	3.5	0	0	0	0.970687769	0.970687769	0.970687769	
19	4	0	0	0	0.98201379	0.98201379	0.98201379	
20								

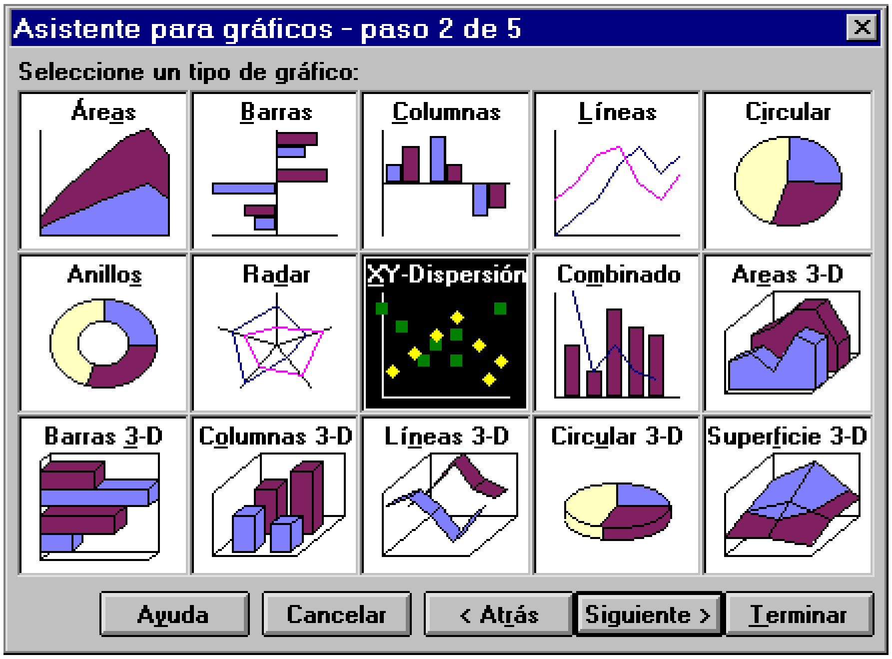
d) Marque una zona en la pantalla donde se van a dibujar los datos. Presione el botón izquierdo del ratón y manténgalo presionado de modo de que se señale el límite de un rectángulo que es la zona de dibujo.

Una vez definido el tamaño deseado suelte el botón izquierdo del ratón.

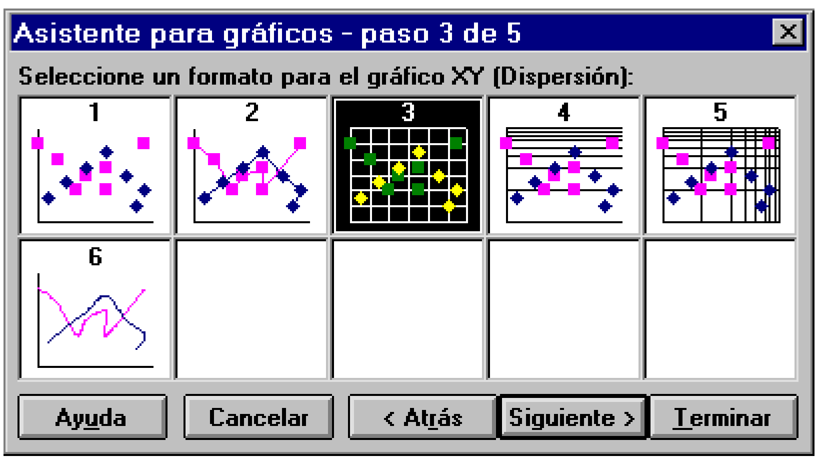
e) Aparece una ventana de diálogo del Asistente para gráficos. Presione el botón "Siguiente"



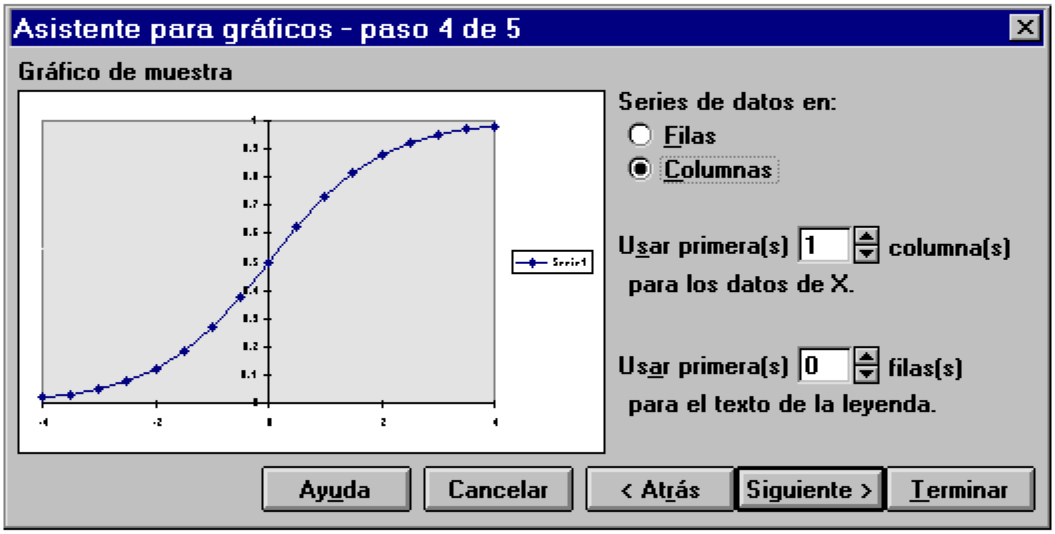
f) Elija el tipo de gráfico "Dispersión" y "Siguiente"



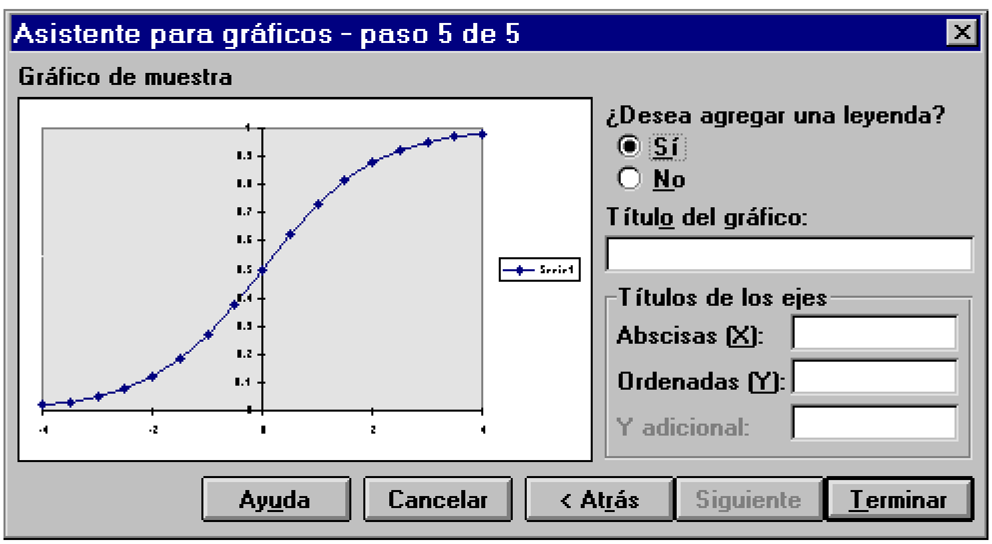
g) Seleccione el formato del gráfico con el tipo 2 (o 3 como lo desee) y “Siguiente”



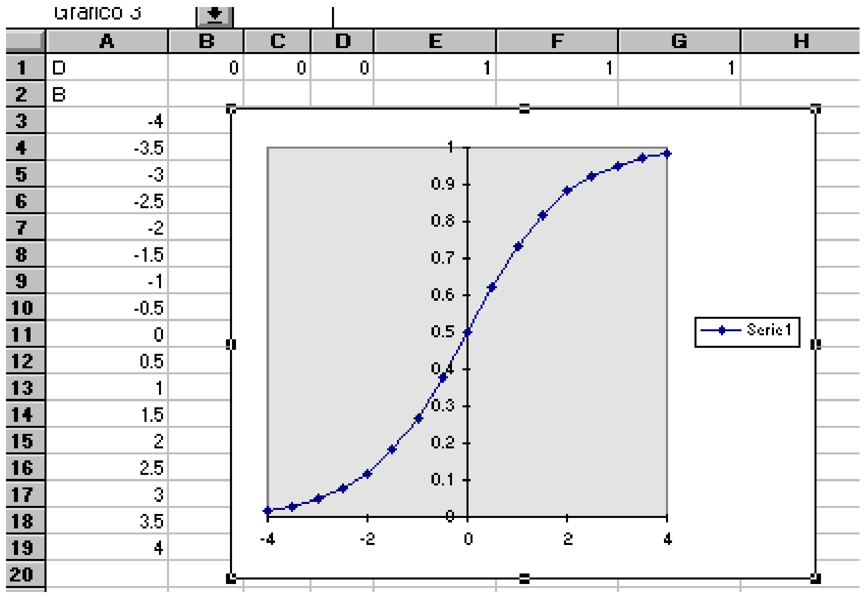
h) Aparece el gráfico de muestra, pida “Siguiente”



i) Anote los títulos que desee (Título del gráfico, abscisas y ordenadas) y “Terminar”



Con ello dispondrá del gráfico deseado en la pantalla.

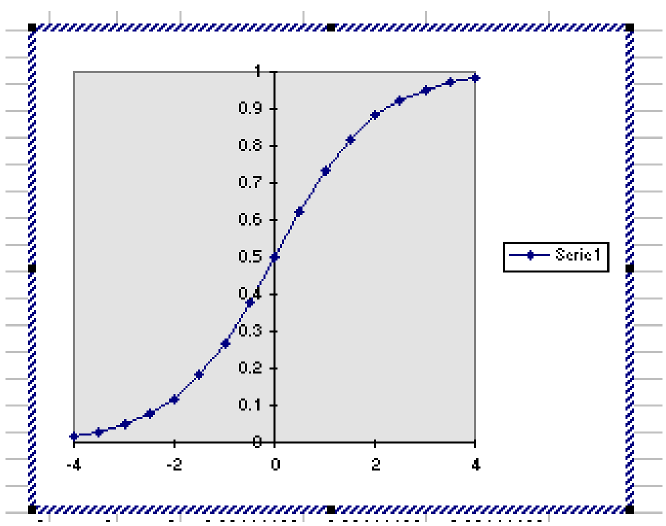


2.4 Determinación de curvas de interpolación

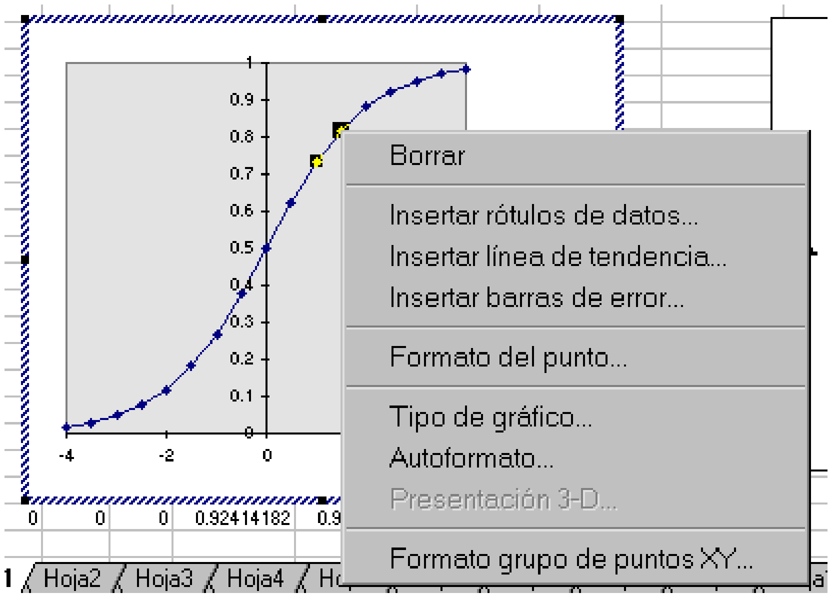
Es posible encontrar funciones de interpolación (no necesariamente precisas) con ayuda de EXCEL.

a) Siga los pasos (a) hasta (i) descritos en la sección 2.3 de este anexo.

b) Haga “doble clic” sobre el gráfico, para que muestre el “marco”

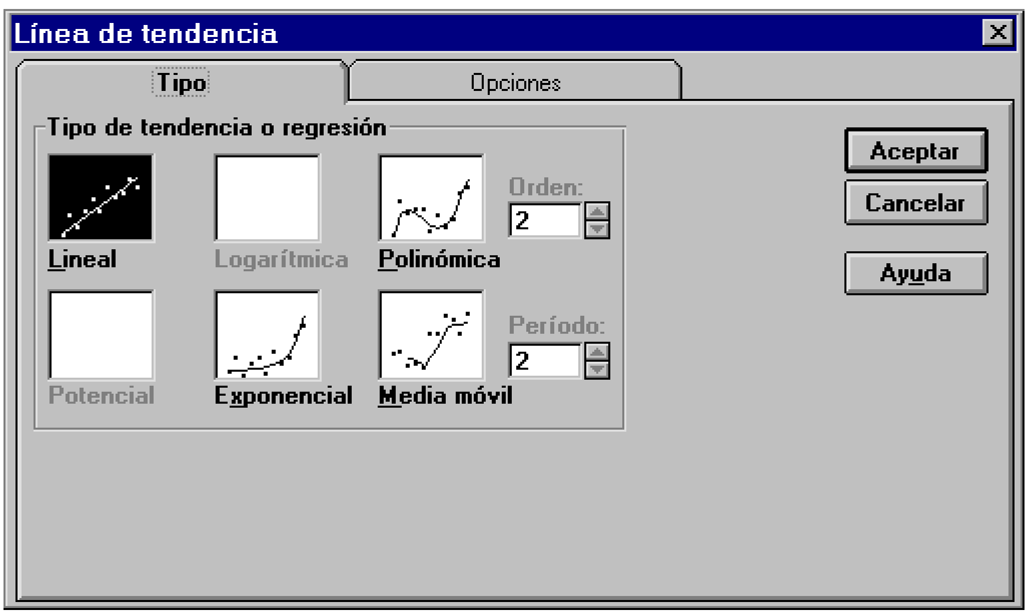


c) Ubíquese sobre uno de los puntos de la gráfica y presione el botón de la derecha, aparecerá un submenú.

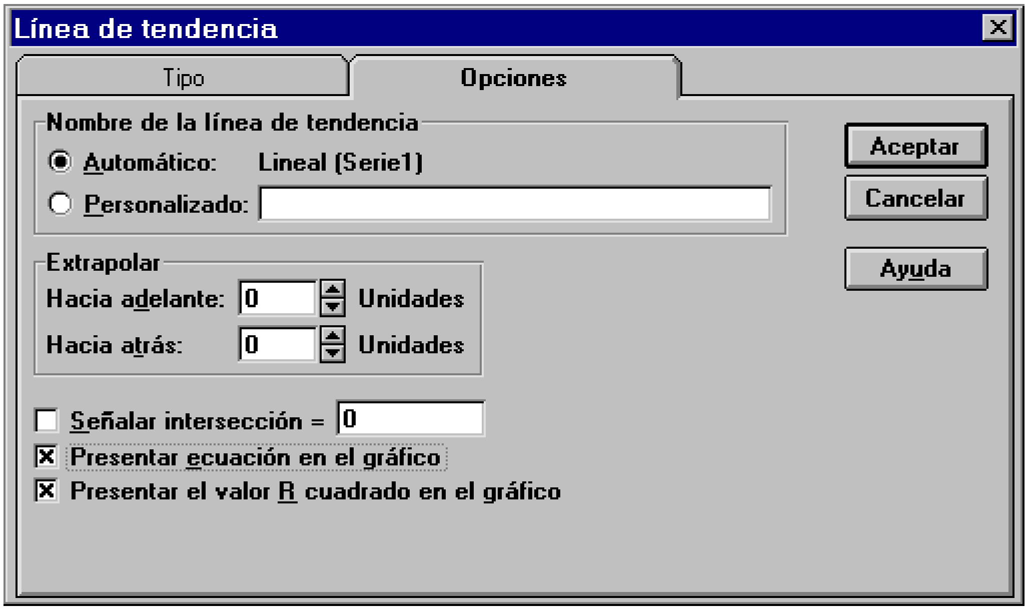


d) Elija la opción “Insertar línea de tendencia”

e) Aparece la carpeta de la opción con dos elecciones: “Tipo” y “Opciones”

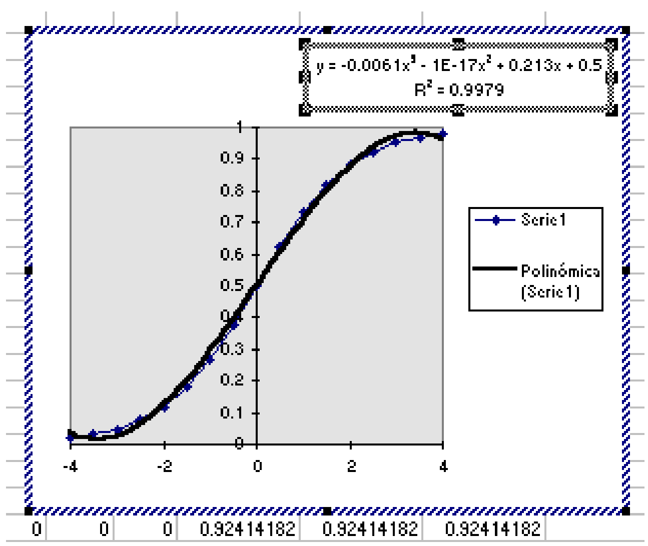


f) Elija “Opciones” y marque “Presentar ecuación”. Puede elegir también “Presentar el valor de R”



g) En “Tipo” seleccione la curva de su interés. En este libro se usa con mucha frecuencia la interpolación polinómica a una expresión cúbica (elija orden 3) o también la interpolación lineal.

h) Presione el botón “Aceptar” para tener el trazo de la curva de interpolación y su ecuación.



La fórmula de interpolación puede quedar “encimada” sobre el gráfico. Es factible mover la fórmula. Para ello deberá con ubicarse sobre ella y presionar el botón izquierdo del ratón. Esto produce un marco alrededor de la fórmula y con un movimiento del ratón se podrá mover el recuadro a la posición que se desee.

APENDICE 3. EL MODELO DE CALIFICACION

Es común en la práctica de la evaluación educativa asignar calificaciones de acuerdo con el número de aciertos o, preferentemente, como cociente de aciertos sobre total de reactivos, que brinda una estimación de la probabilidad de respuesta del sujeto.

La calificación hace referencia a un estimador del dominio que tiene un sujeto en función de los resultados que se obtienen sobre un muestreo de reactivos relativos a un conjunto de objetivos, contenidos o temas. Este muestreo es siempre imperfecto debido a que no es posible explorar todos y cada uno de los contenidos u objetivos a evaluar, ya que se requeriría un tiempo excesivo y un costo muy alto. Por ello se debe hacer la estimación basándose exclusivamente en una muestra de aciertos sobre una muestra de reactivos.

El modelo de calificación trata de explicar la estimación citada. Para ello supóngase que los conocimientos a evaluar se ponen en correspondencia con un conjunto de "N" canicas rojas que se colocan dentro de una urna. Supongamos ahora que pudiéramos saber de antemano todos los conocimientos que tiene el sujeto y marcamos con color blanco las "B" canicas que corresponden con dichos conocimientos que realmente posee el sujeto y se vuelven a colocar las canicas en la urna. El problema es el siguiente:

> **Si se toma una muestra de "n" canicas y de ellas se encuentra que hay "b" canicas blancas, ¿cuál es la cantidad "B" de canicas blancas que había en la urna?**

Es claro que al tomar el muestreo no se va a hacer el conteo de todas las canicas de la urna, sino "estimarlas" por medio del resultado obtenido: "b" canicas blancas en una muestra de tamaño "n".

El problema no es trivial. Belaunzarán demostró que bajo ciertas hipótesis es factible demostrar que el cociente b/n es el estimador más probable de B/(N+1), o dicho de otro modo, el número de canicas blancas que había en la urna puede estimarse en:

$$B = b(N+1)/n$$

Esto es muy bueno para el evaluador, ya que permite decir que el dominio más probable de un sujeto puede obtenerse por medio del cociente p/n, que es la frecuencia relativa de respuestas correctas de un sujeto ante un examen.

Para que esta estimación sea buena, se deben cumplir estas hipótesis básicas:

a) La extracción de las canicas debe hacerse al azar (el muestreo no debe ser dirigido)

b) Todas las canicas deben ser igualmente probables de ser extraídas (no debe haber canicas grandes, chicas, etc.)

c) Toda canica extraída no puede ser regresada a la urna

En términos de evaluación se puede decir que las hipótesis serían:

a) El cuestionario debe ser construido utilizando un muestreo al azar

b) Los reactivos deben ser del mismo valor (o peso)

c) El muestreo debe ser sin reemplazo

El modelo de calificación descarta, desde luego, cualquier agente de origen psicológico, fatiga en el sujeto, conocimiento previo del examen, fraudes, etc.

Finalmente, para que el estimador sea eficiente, el tamaño de muestra debe ser suficiente. Esto facilita la estimación en cuestionarios amplios, pero puede ser un mal estimador en el caso de “subtemas” cuando se tienen pocos reactivos. El evaluador deberá hacer un análisis de calidad del muestreo para los exámenes muy cortos.

REFERENCIAS

Sugiero estos materiales de referencia para continuar con el estudio del Modelo de Rasch y otros temas vistos en este libro:

1. Linacre J.M. y Wright B.D. (1994) “A users’s guide to BIGSTEPS”, MESA Press, Chicago EUA, 96 pp

2. Lord F.M. y Novick M.R. (1968) “Statistical theories of mental test scores”, Addison-Wesley Publishing Co. Reading Massachusetts, EUA, 568 pp

3. Wright B.D. y Stone M.H. (1979) “Best test design”, MESA Press, Chicago EUA , 222 pp. (también existe versión en español: “Diseño de mejores pruebas”, Ceneval, 1998.)

BIBLIOGRAFIA

La bibliografía es muy grande, por lo que sólo se incluyen aquí unos cuantos documentos que pueden ser de utilidad para reforzar el conocimiento de estos temas.

Belaunzarán E. (1975) "El modelo de calificación". Reimpreso como Noticias ICI E-34, 6 de marzo de 1996. Ingeniería Computarizada Integral, México.

Draba R.E. (1977) "The identification and interpretation of item bias", Education Statistics Laboratory, Memorandum 25. Mesa Press, Chicago EUA, 11 pp

Fischer G.H. y Molenaar I.W. eds. (1995) "Rasch models. Foundations, recent developments and applications". Springer-Verlag, Nueva York EUA. 436 pp.

Lord F.M. (1980) "Applications of item response theory to practical testing problems". Lawrence Erlbaum Associates, Publ. Hillsdale NJ, EUA, 272 pp.

Rasch G. (1980) "Probabilistic models for some intelligence and attainment tests", MESA Press. Chicago. EUA. 199 pp.

Thurstone L.L. (1929) "The scoring of individual performance", Journal of Educational Psychology, APA Washington EUA, Vol.17, pp 446-457

Tristán L.A. (1996) "Seminario de Análisis de Rasch con el Programa BIGSTEPS¨, en Técnicas no clásicas de calificación de cuestionarios objetivos (teoría y práctica). 2o. Foro de Evaluación, Octubre, Ceneval, 60 pp

Tristán L.A. (1995) "Breves notas sobre el modelo de Rasch (1)", Noticias ICI E-08, 24 febrero. Ingeniería Computarizada Integral, México

Tristán L.A. (1996) "Breves notas sobre el modelo de Rasch (2)", Noticias ICI E-29, 26 febrero. Ingeniería Computarizada Integral, México

Wright B.D. (1967) "Sample-free test calibration and person measurement", Invitational conference on testing problems, Educational Testing Service, Princeton NJ. EUA, pp 85-101

Wright B.D. y Stone M.H. (1988) "identification of item bias using Rasch measurement", Research memorandum N. 55. MESA Press 11 pp

Wright B.D., Mead R. y Draba R. (1976) “Detecting and correcting test item bias with a logistic response model”, Research Memorandum N. 22, Statistical Laboratory Department of Education, The University of Chicago, EUA. 24 pp

Wright B.D. y Douglas G.A. (1977) “Best procedures for sample-free item analysis”, Applied Psychological Measurement, Vol. 1, N. 2, Primavera 1977, pp 281-295

Wright B.D. y Masters G.N. (1982) “Rating scale analysis”, MESA Press. Chicago EUA. 206 pp

Wright B.D. y Stone M.H. (1979) “Best test design”, MESA Press, Chicago EUA , 222 pp

Wright B.D. y Stone M.H. (1988) “Validity in Rasch measurement”, Research Memorandum N. 54, MESA Press, Chicago EUA, 11 pp.

Wright B.D. y Stone M.H. (1988) “Reliability in Rasch measurement”, Research Memorandum N. 53, MESA Press, Chicago EUA, 21 pp.

INDICE

O

P

R

S

T

U

V