

Tema 7: Teoría de la respuesta al ítem (1) :

Conceptos básicos

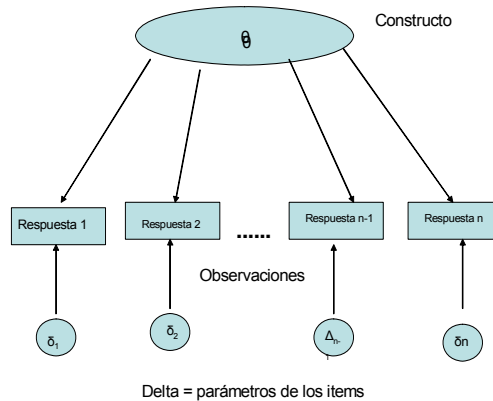


Índice



1. Introducción
2. Limitaciones de la TCT superadas por la TRI
3. Modelos y características comunes
4. Supuestos
5. La Curva Característica del Ítem
6. Modelos logísticos para ítems dicotómicos:
Forma y parámetros de los modelos:
 1. Modelo de 1PL
 2. Modelo 2PL
 3. Modelo 3PL
 4. Modelos de ojiva normal y su aproximación logística

1. 1. Modelo de la TRI



1.2. Breve historia

- Uso reciente, pero sus bases teóricas una larga historia
- Las CCI antecedentes en Binet y Simon
- Primeros desarrollos teóricos en los años 30 y 40 (Richardson, Lawley,...)
- Lazarsfeld (1950) en Sociología y Lord (1952) en Psicología
- Desarrollos a lo largo de dos líneas separadas:
 - Matemático danés G. Rasch (especialmente influyente en Europa y por el grupo de B.Wright en la Universidad de Chicago)
 - Desarrollos en USA, más guiados por la filosofía de Thurstone y desarrollados por Lord (Educational Testing Service) y Bock y sus colaboradores



2. 1. Limitaciones de la TCT que resuelve TRI


- Dependencia de los estadísticos del ítem de la muestra (población) en la que se obtienen
- Dependencia de los estadísticos del test de la muestra (población)
- Dependencia de las puntuaciones del sujeto del conjunto particular de ítems incluidos en el test
- Dependencia del concepto de paralelismo para la estimación de la fiabilidad
- El estimador de la puntuación verdadera V' depende del grupo de referencia
- La TCT no proporciona un modelo para las respuestas a los ítems
- Problemas con los supuestos sobre el error típico de medida (homoscedasticidad)
- Medidas no óptimas ni adaptadas a los sujetos particulares
- Dificultad de solucionar numerosos problemas encontrados en modernas aplicaciones de los tests (Tests adaptativos informatizados, evaluaciones a gran escala, equiparación,.....)



2.2. Para superar las limitaciones:

- Características de los ítems que no dependan del grupo en que se calculen
- Que las puntuaciones que describan el comportamiento de los sujetos examinados en el test no dependan de los contenidos particulares de éste ni de su dificultad.
- Que los modelos estén expresados a nivel del ítem, en vez de a nivel del test completo.
- Un modelo que no requiera del supuesto estricto de paralelismo u otras formas de equivalencia para establecer la fiabilidad del test.
- Modelo que proporcione medidas de precisión para cada nivel de aptitud, eliminando el supuesto de la homoscedasticidad de la varianza error.
- Modelo que permita relacionar el rasgo o constructo con el rendimiento del sujeto en el ítem.

2.3. Comparación TCT-TRI




Características	TCT	TRI
Modelo	Lineal	No lineal
Nivel	Test	Ítem
Relaciones ítem/rasgo	No especificadas	Función característica ítem
Supuestos	Débiles, fáciles de cumplir	Fuertes y difíciles cumplir
Invarianza de estadísticos	No, dependientes muestra	Sí, si el modelo se ajusta
Estadísticos de ítem	P, Δ, r_{bp}, r_b	a, b, c según modelos, funciones de información
Tamaño muestral para la calibración	200-500	Depende del modelo y procedimiento de estimación, en general $n > 500$
Estadísticos del test	ρ_{xx}, σ_e	Función de información, Error típico

2.4. Viejas y nuevas reglas de la medida - 1 (Embretson y Reise, 2000)

Viejas: TCT

Nuevas: TRI



El error típico de medida se aplica a todas las puntuaciones de una población particular (homosc.)	El error típico de medida difiere entre puntuaciones o patrones de respuesta, pero se generaliza a poblaciones
Tests más largos son más fiables que tests cortos	Aunque se mantienen las relaciones entre fiabilidad y longitud del test, tests cortos pueden ser más fiables que tests largos
Pueden establecerse comparaciones entre formas múltiples de tests cuando las formas son paralelas	Pueden establecerse comparaciones en general entre formas diferentes de tests, sin el requisito del paralelismo
Las propiedades de los estadísticos de ítem dependen de las muestras concretas en las que se calculan	Pueden obtenerse estimadores insesgados de los parámetros de los ítems
Las puntuaciones de los tests adquieren significado comparadas a las de un grupo normativo, que realizó una forma idéntica o paralela del test	Las puntuaciones de los tests tienen significado si se comparan por su distancia a ítems

Viejas y nuevas reglas de la medida - 2 (Embretson y Reise, 2000)

Viejas: TCT

Nuevas: TRI



Propiedades de escala de intervalo se obtienen obteniendo distribuciones de puntuaciones normales	Propiedades de escala de intervalo se obtienen aplicando modelos de medida justificables (Rasch)
Formatos mixtos de ítem llevan a desequilibrios de ponderación en las puntuaciones totales	Formatos mixtos pueden proporcionar puntuaciones óptimas
Problemas en las comparaciones para la medida del cambio	Puntuaciones de cambio son fácilmente establecidas
El AF de ítems binarios puede producir factores artefactuales	El AF de datos de ítem lleva a AF de información total
En el análisis de las propiedades psicométricas se enfatiza el test global como unidad de análisis	Las propiedades psicométricas se basan en las propiedades de los ítems

2.5. Principales innovaciones



- Fundamentación probabilística al problema de la medición de variables latentes
- El ítem es la unidad básica de análisis del test y no las puntuaciones totales como en TCT y TG
- Los modelos son funciones matemáticas que relacionan la probabilidad de un tipo de respuesta al ítem con la variables latente (rasgo, constructo, habilidad: θ)
- Permite definir una escala para la variable latente medida por los ítems del test, que no se obtiene como suma de puntuaciones como en TCT y TG
- Ítems e individuos pueden medirse en la misma escala
- El concepto de fiabilidad es superado por el de información del test



2.6. Principales aplicaciones

- Evaluaciones educativas a gran escala, ligadas a los diseños matrix-sampling
- Análisis del funcionamiento diferencial de los ítems (DIF)
- Equiparación de puntuaciones y comparabilidad
- Tests adaptativos e informatizados
- Modelos de tests basados en la psicología cognitiva



3. Modelos

- Bajo el mismo nombre genérico existen múltiples modelos, que exhiben un conjunto de características comunes:
 - Existencia de un rasgo latente responsable de la conducta o respuesta del sujeto ante un ítem (recientemente también existen modelos multidimensionales). Es preciso establecer la dimensionalidad
 - La relación entre el rasgo y la respuesta de un sujeto a un ítem puede describirse por medio de una función monótona no decreciente, denominada **Curva Característica del Ítem (CCI)** o Función de Respuesta al Ítem (FRI), que establece las probabilidades del tipo de respuesta
 - No se requieren los supuestos de paralelismo ni de homocedasticidad de los errores típicos
 - Los ajustes de los datos a los modelos pueden probarse
 - Cuando se cumplen los supuestos proporcionan estimadores con las siguientes propiedades:
 - Invarianza de los parámetros de los ítems
 - Invarianza de los parámetros de rasgo de los sujetos



3. Modelos

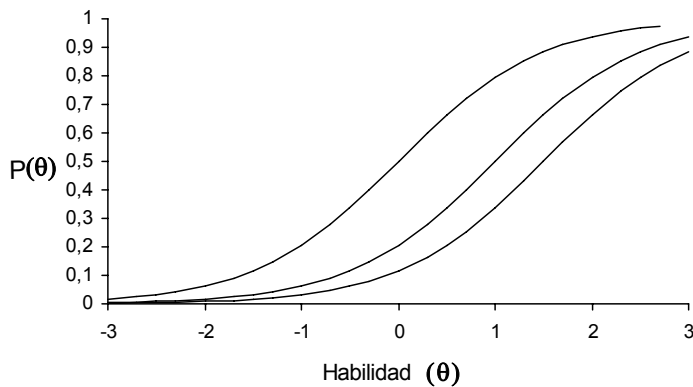
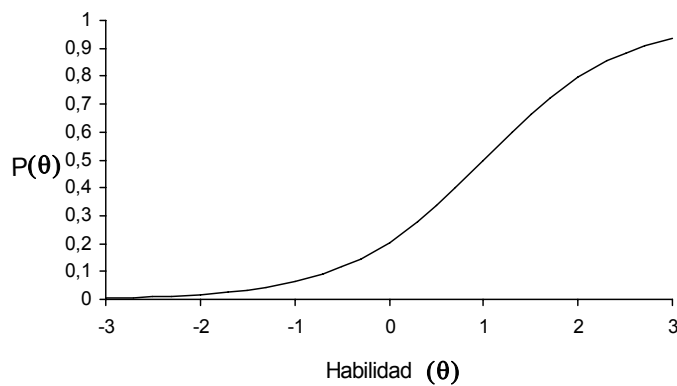
- Items y sujetos pueden representarse en el espacio definido por el (los) rasgo(s) latente(s)
- El papel que en la TCT asumía V , lo tiene ahora θ o nivel de habilidad del sujeto
- Las unidades básicas son los ítems del test y los modelos se establecen en el nivel de los ítems
- El modelo que se supone siguen los ítems del test se establece antes de puntuar, y permiten transformar las respuestas en puntuaciones



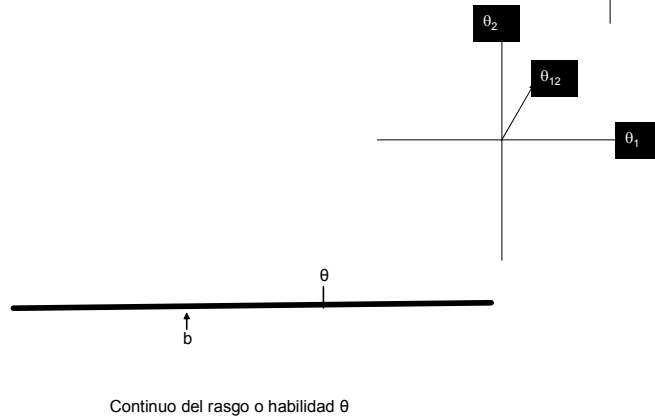
4. Supuestos

- La probabilidad de acertar un ítem es una función monótona no decreciente de la habilidad
- Unidimensionalidad (o en general, dimensionalidad conocida)
- Independencia local: si se mantiene constante la habilidad, las respuestas de los examinados a un par de ítems cualesquiera son estadísticamente independientes
- (Ausencia de velocidad)

Función monótona no decreciente



Dimensionalidad conocida



Independencia local

- Puede resumirse en la siguiente expresión

$$P(U_1, U_2, \dots, U_n | \theta) = P(U_1 | \theta) P(U_2 | \theta) \dots P(U_n | \theta) = \prod_{i=1}^n P(U_i | \theta)$$

Patrón	Probabilidad
000	qqq = .2 x .5 x .6 = .06
100	pqq = .8 x .5 x .6 = .24
010	qpq = .2 x .5 x .6 = .06
001	qqp = .2 x .5 x .4 = .04
110	ppq = .8 x .5 x .6 = .24
101	pqp = .8 x .5 x .4 = .16
011	qpp = .2 x .5 x .4 = .04
111	ppp = .8 x .5 x .4 = .16

$$p_1 = 0,80$$

$$p_2 = 0,50$$

$$p_3 = 0,40$$

5. Curvas características de ítem



- Funciones monótonas no decrecientes de la habilidad que describen la probabilidad de dar una respuesta determinada (correcta o incorrecta en los ítems dicotómicos)
- Existen diversas funciones que permiten expresar dichas probabilidades, difiriendo por la forma especificada y/o por los parámetros que las describen. En cuanto a la **forma**, suelen utilizarse la Función de Distribución Logística y la Función de Distribución Normal.
- Bajo cualquiera de las dos formas, las Funciones o Curvas Características del Ítem pueden diferir en cuanto al número de parámetros del ítem que contemplan.

5. Curvas características de ítem



- Hay un parámetro básico que contienen todos los modelos: la **dificultad del ítem** (b_i)
- Algunos modelos incluyen además la **discriminación del ítem** (a_i)
- En los ítems de elección de respuesta o elección múltiple, suele ser conveniente la inclusión del parámetro de **acierto por adivinación o conjetura** (c_i)
- El modelo que incluye solamente la dificultad es conocido como modelo de 1P. Cuando el modelo incluye dificultad y discriminación, nos encontramos con el modelo de 2P. Finalmente, cuando el modelo incluye además de los parámetros anteriores, el de probabilidad de acierto por adivinación, tendremos el modelo de 3P
- Los tres modelos anteriores pueden aparecer con la función de distribución normal (1PN, 2PN, 3PN) o con la logística (1PL, 2PL, 3PL)

6.1. Modelo TRI logístico de 1P (modelo de Rasch)



- Función de distribución logística:

$$\Psi(x) = \frac{\exp(x)}{1 + \exp(x)} = \frac{1}{1 + \exp(-x)}$$

- Cálculo de probabilidades acumuladas:

$$P = \frac{e^x}{1 + e^x}$$

$$P = \frac{1}{1 + e^{-x}}$$

- El modelo 1PL:
- El exponente de e se denomina *logit*

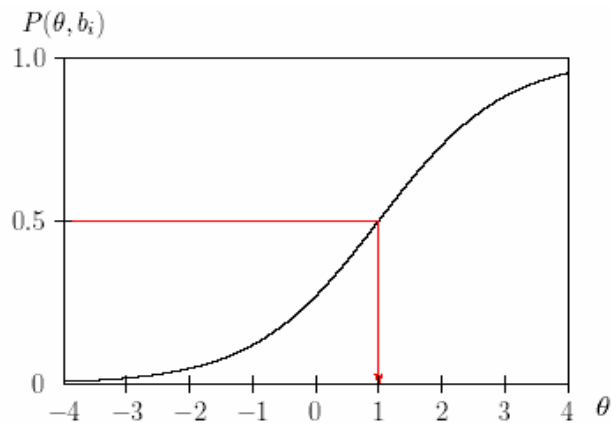
$$P(u_i = 1 | \theta_s, b_i) = P(\theta_s) = \frac{e^{(\theta_s - b_i)}}{1 + e^{(\theta_s - b_i)}} = \frac{1}{1 + e^{-(\theta_s - b_i)}}$$

6.1. El modelo logístico de 1P, 1PL: b_i

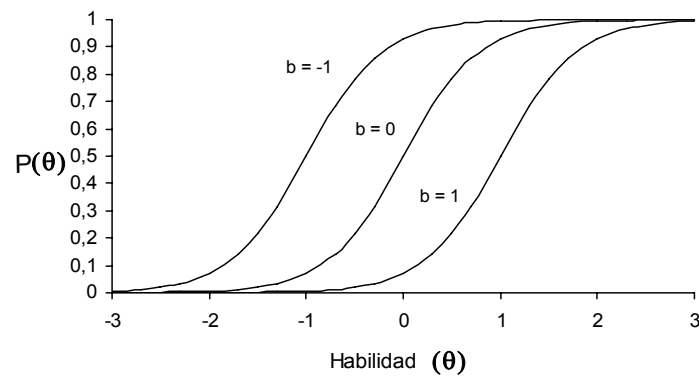


- La probabilidad de acertar el ítem depende únicamente de la habilidad del sujeto (θ_s) y de la dificultad del ítem (b_i)
- Describe dónde está situado el ítem en la escala de habilidad, o la cantidad de ésta que el ítem requiere para ser resuelto con éxito. Es un índice de posición del ítem en la escala de medida del rasgo o habilidad
- Técnicamente es el nivel de habilidad θ en el punto de inflexión de la curva (las probabilidades cambian de menos de 0.50 a más de 0.50 y coincide con $p=.50$, ya que habilidad y dificultad son iguales en este punto)

6.1. Modelo logístico de 1P: el parámetro b_i



6.1. Modelo de 1P: CCI con diferentes niveles de dificultad

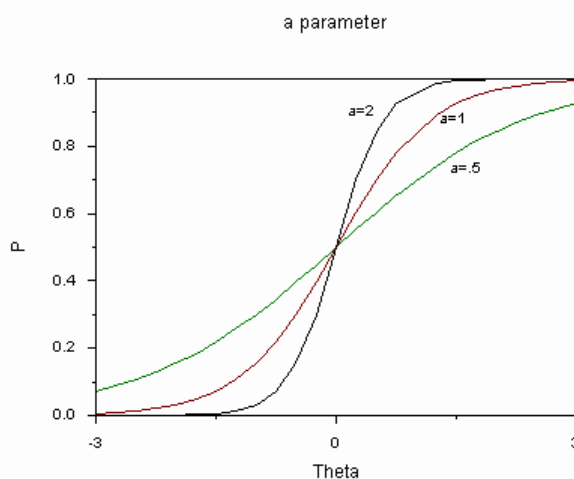


6.2. Modelo logístico de 2P, 2PL: b_i , a_i

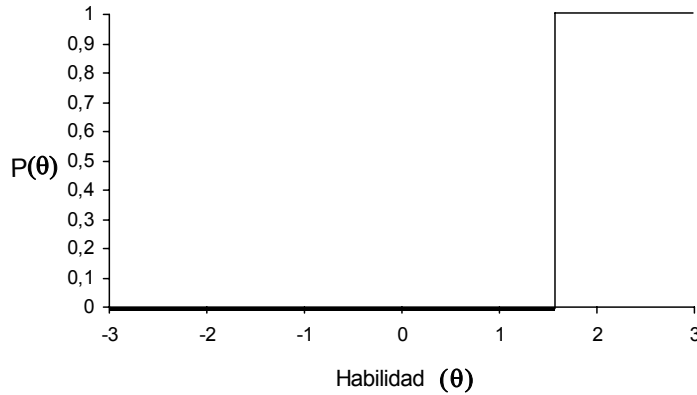


- El parámetro a_i indica hasta qué punto el ítem permite diferenciar entre sujetos que tienen una habilidad inferior a la posición del ítem y los que la tienen superior
- Refleja la tasa de cambio en la probabilidad de éxito según aumenta la habilidad
- Es función de la pendiente de la curva en el punto de inflexión ($\theta_s = b_i$)
- Curvas con mayor pendiente discriminarán mejor

6.2. Modelo 2P: Curvas con la misma dificultad y diferente discriminación



Modelo 2P: Curva con discriminación perfecta



6.2. Modelo logístico 2P: Función de respuesta al ítem



- Probabilidad de acierto del ítem:

$$P_i(u_i=1 | \theta_s, a_i, b_i) = P_i(\theta_s) = \frac{e^{a_i(\theta_s - b_i)}}{1 + e^{a_i(\theta_s - b_i)}} = \frac{1}{1 + e^{-a_i(\theta_s - b_i)}}$$



El logit: exponente de e

- Sea $P_i(\theta_s)$ la probabilidad de acertar el ítem i dado el nivel de habilidad θ_s
- Sea $Q_i(\theta_s) = 1 - P_i(\theta_s)$ la probabilidad de no acertar el ítem en el nivel θ_s
- La cantidad $\frac{P(\theta)}{1 - P(\theta)} = e^{a(\theta - b)}$
- El ln de la anterior cantidad, nos da la expresión denominada **logit**, denotado como L .

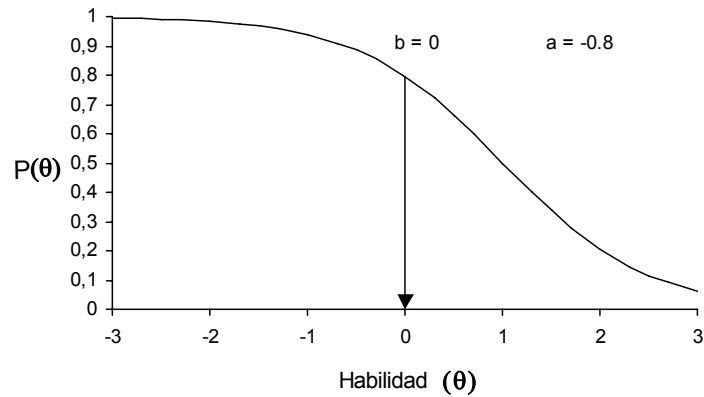
$$L = \ln \frac{P(\theta)}{1 - P(\theta)} = a(\theta - b)$$



La probabilidad de respuesta incorrecta

- Un ítem tiene tantas CCI como opciones de respuesta. Así, un ítem dicotómico tiene realmente dos CCI, una para la respuesta correcta (las que hemos presentado en las figuras anteriores) y otra para la respuesta incorrecta, $Q_i(\theta_s)$, cuya probabilidad también puede representarse como función de la aptitud. Esta curva es la imagen en espejo de la de $P_i(\theta)$; su parámetro de dificultad b_i tendrá el mismo valor y su parámetro de discriminación a_i el mismo valor numérico, pero con signo opuesto. Por esta razón Samejima (1969) prefiere hablar de curvas características de las categorías de respuesta del ítem.

CCI de la respuesta incorrecta

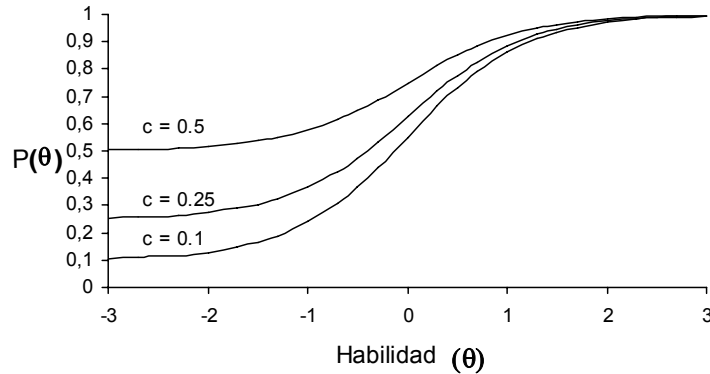


6.3. Modelo logístico 3P, 3PL: Parámetro c_i : acierto por conjetura o adivinación



$$P(u_i = 1 | \theta_s, a_i, b_i, c_i) = P_i(\theta_s) = c_i + (1 - c_i) \frac{e^{a_i(\theta_s - b_i)}}{1 + e^{a_i(\theta_s - b_i)}} = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_s - b_i)}}$$

6.3. Modelo logístico 3P, 3PL: Parámetro c_i : acierto por conjetura o adivinación



6.4. Modelos de ojiva normal



- Los modelos de 1P, 2P y 3P, también pueden formularse utilizando la función de distribución normal

$$P_i(\theta_s) = \Phi(z) = \int_{-\infty}^{(\theta_s - b_i)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

$$P_i(\theta_s) = \Phi(z) = \int_{-\infty}^{a_i(\theta_s - b_i)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

$$z_{is} = a_i(\theta_s - b_i)$$

$$P_i(\theta_s) = c_i + (1 - c_i) \int_{-\infty}^{a_i(\theta_s - b_i)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

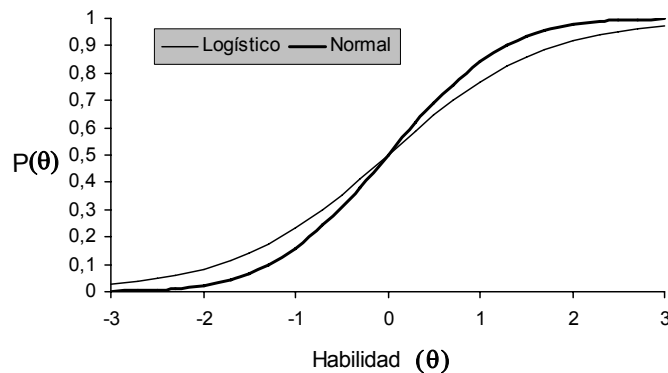


6.4. Modelos de ojiva normal

- El límite superior de la integral es como una puntuación estandarizada; de modo que la proporción acumulada de casos hasta dicho valor proporciona la probabilidad de una respuesta correcta. Por ejemplo si $a_i = 1$, $b_i = 1.0$ y $\theta = 2.0$, calculamos la z correspondiente que será 1.0. El área de una distribución normal estandarizada hasta $z=1$, es 0.8413.
- En la práctica, la distribución normal, aunque es en la que se basaban los primeros modelos tiene el problema de la integral, por lo que suelen utilizarse los modelos logísticos, como tales o con un factor de escalamiento $D = 1,7$, que los aproxima a las probabilidades de una función de distribución normal



F.D. Normal y logística



6.4. Aproximaciones logísticas a los modelos de ojiva normal: el factor D



- Las dos funciones de distribución, normal y logística, tienen formas muy similares, con ligeras diferencias en las probabilidades o áreas, especialmente en los extremos
- Haley (1954) demostró que:
- Multiplicando el exponente de e de la función logística por D = 1,7, pueden aproximarse las probabilidades normales
- Es frecuente representar los modelos logísticos con D en el exponente

$$|\Phi(x) - \Psi(1.7x)| < .01$$

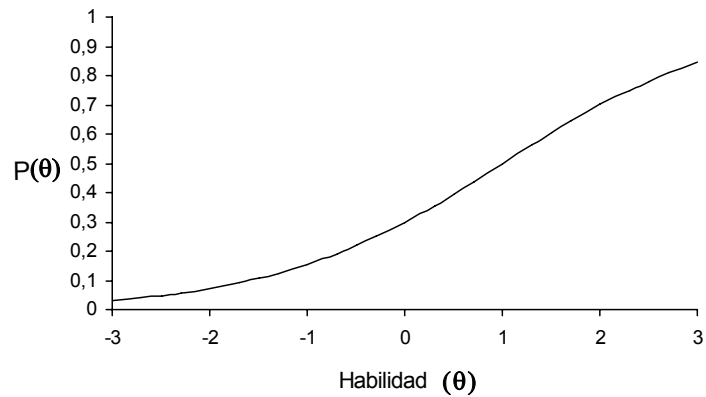
Ejemplo 2P (con D = 1,7)

a = 0,5
b = 1



Aptitud	Logit (L)	Exp (-DL)	1+exp(-DL)	Probabilidad
-3	-2.0	29.96	30.96	0.03
-2	-1.5	12.81	13.81	0.07
-1	-1.0	5.47	6.47	0.15
0	-0.5	2.34	3.34	0.30
1	0.0	2.00	2.00	0.50
2	0.5	1.43	1.43	0.70
3	1.0	1.18	1.82	0.85

CCI para los datos del ejemplo



Rango de valores de a_i



Etiqueta verbal	Rango de valores
Ninguna	0
Muy baja	0.01-0.34
Baja	0.35-0.64
Moderada	0.65-1.34
Alta	1.35-1.69
Muy alta	> 1.70
Perfecta	

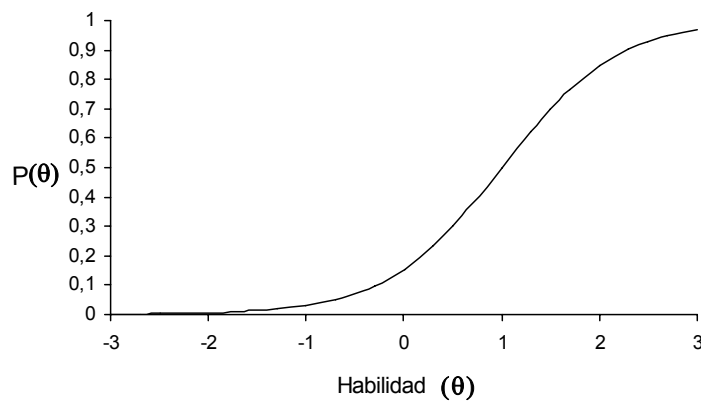
Ejemplo: Modelo de 1P

$b = 1$, con
 $D = 1,7$



Aptitud	Logit (L)	Exp (-DL)	1+exp(-DL)	Probabilidad
-3	-4	897.85	898.85	0.001
-2	-3	164.02	165.02	0.006
-1	-2	29.96	30.96	0.032
0	-1	5.47	6.47	0.154
1	0	1.00	2.00	0.500
2	1	0.18	1.18	0.645
3	2	0.03	1.03	0.968

CCI del ejemplo de 1P



Ejemplo modelo 3P

$$a = 1,3$$

$$b = 1,5$$

$$c = 0,20 ;$$

$$\text{con } D = 1,7$$



Aptitud	Logit (L)	Exp (-DL)	1+exp(-DL)	Probabilidad
-3	-5.85	20847.72	0.00 ¹	0.20 ¹
-2	-4.55	2287.01	0.00 ¹	0.20 ¹
-1	-3.25	250.89	0.00 ¹	0.20 ¹
0	-1.95	27.52	0.04	0.23
1	-0.65	3.10	0.25	0.40
2	0.65	0.33	0.75	0.90
3	1.95	0.04	0.96	0.97

CCI del ejemplo de 3P

