INSTITUTO UNIVERSITARIO
DE LA EMPRESA

LA PARADOJA DE RASCH: MEDIDAS Y ERRORES

JUAN RAMÓN OREJA-RODRÍGUEZ

SERIE ESTUDIOS 2008/ 68

SANTA CRUZ DE TENERIFE, NOVIEMBRE DE 2008

UNIVERSIDAD DE
LA LAGUNA

IUDE Documento de Trabajo. Serie Estudios Nº 2008/68 **ISSN 1139-4501**

Resumen
Las medidas logradas por el ajuste de datos al modelo de Rasch cumplen los requisitos de la medición fundamental exigida por Campbell. La obtención de estas medidas nos permite superar la limitación generalizada de la utilización de puntuaciones ordinales en administración de empresas como concepto métrico, que afecta a la interpretación de los resultados conseguidos en su tratamiento estadístico desde una perspectiva empresarial.
Junto a una introducción de la discusión conceptual de las mediciones en ciencias sociales, se presentan la tipología de medidas y errores del modelo de Rasch, concluyendo con la presentación de la Paradoja de Rasch.
Palabras claves: MEDICIÓN, MEDIDAS, ERRORES, MODELO DE RASCH

Abstract
The measures obtained by the fit of data al model of Rasch fulfil the requirements of the fundamental measurement demanded by Campbell. The obtaining of these measures allows us to surpass the generalized limitation of the use of ordinal scores in administration of companies like metric concept, which affects to the interpretation of the results obtained in its statistical treatment from an enterprise perspective. Next to an introduction of the conceptual discussion of the measurements in social sciences, they appear the typologies of measures and errors of the model of Rasch, concluding with the presentation of the Paradox of Rasch.
Key words: MEASUREMENT, MEASURES, ERRORS, RASCH MODEL

LA PARADOJA DE RASCH: MEDIDAS Y ERRORES+

JUAN RAMÓN OREJA-RODRÍGUEZ*

SERIE ESTUDIOS 2008/ 68

SANTA CRUZ DE TENERIFE, NOVIEMBRE 2008

+ Trabajo presentado al II Workshop de Modelos de Rasch en Administración de Empresas (Rasch Models on Business Administration) organizado por el IUDE de la Universidad de La Laguna. 10 de noviembre de 2008.

* joreja@ull.es
Departamento de Economía y Dirección de Empresas e Instituto Universitario de la Empresa.
Facultad de Ciencias Económicas y Empresariales. Campus de Guajara. Universidad de La Laguna.
Camino de La Hornera s/n 38071 La Laguna Tenerife Islas Canarias (España). Fax:+ 34 922 31 70 77

Introducción

El proceso de medición de conceptos en administración de empresas nos permite considerar tanto de constructos como de instrumentos de medida como elementos previos al mismo. La determinación de mediciones destaca la presencia de conceptos métricos que cumplen con los estrictos requisitos exigidos para obtener mediciones en escalas de intervalo o de razón. Para ello, siguiendo a Mosterin (1984) sería necesaria la presencia de un homomorfismo entre un sistema empírico y un sistema numérico.

El objetivo de este trabajo es destacar el cumplimiento de los requisitos de la medición fundamental exigida por Campbell de las mediciones obtenidas por la aplicación del modelo de Rasch que permite superar la limitación generalizada de la utilización de puntuaciones ordinales en administración de empresas como concepto métrico con las correspondientes salvedades interpretativas de los resultados obtenidos en su tratamiento estadístico.

Junto a una discusión conceptual de las mediciones en ciencias sociales, se presentan la tipología de medidas y errores del modelo de Rasch, concluyendo con la presentación de la Paradoja de Rasch.

1. Mediciones

El informe final del Comité Fergurson (Ferguson et al, 1940) convocado, para considerar e informar sobre la posibilidad de realizar estimaciones cuantitativas de los eventos sensoriales, por la Asociación Británica para el Avance de las Ciencias negó la posibilidad de mediciones cuantitativas en las ciencias sociales, al no cumplir con los requisitos científicos a los que las mediciones se deben ajustar.

Como indican Borsboom y Zand Scholten (2008) el físico Norman Campbell, uno de los miembros más activos de esta Comisión exigió una demostración empírica de que las propiedades psicométricas tienen estructura aditiva si querían ser consideradas como mensurables. Para Campbell (1921/1953) medir es la asignación de números a sistemas de acuerdo a leyes científicas. En su planteamiento metodológico, el principal aspecto que no cumplen las mediciones de los eventos sensoriales (y por extensión los hechos sociales) es la concatenación de las medidas, base del principio de aditividad.

Frente a este dictamen, algunos científicos sociales plantean distintas alternativas conceptuales sobre la medición. Stevens (1946) redefinió el concepto de medición reduciéndolo a la "asignación de números a observaciones de acuerdo con una regla". A partir de determinadas reglas para asignar números delimitó sus conocidas escalas de medición: nominal, ordinal, intervalo y razón.

Así, por ejemplo, delimitó la presencia de una escala nominal ante la regla de presencia de igualdades en las observaciones. La escala ordinal aparece al aplicar la regla de la desigualdad (mayor que) entre las observaciones.

No obstante, Borsboom y Zand Scholten (2008) indican que rara vez los investigadores de la medida han destacado de forma explícita que Stevens reconocía que debe existir algún tipo de similitud entre las asignaciones numéricas y la propiedad medida para que el procedimiento que proponía tuviera sentido. Sin tal isomorfismo, estas definiciones quedarían incompletas, vacías, porque el mero hecho de que se asigne números de acuerdo a una regla es trivialmente satisfecho en casi todos los casos. Como indican estos autores, lo que realmente difícil es no poder ser medido de acuerdo a ese criterio.

El desarrollo de modelos de variables latentes a partir de los años 50 y 60 del siglo pasado ha incidido en el análisis empírico de realidades sociales. Los trabajos de medición iniciados por Rasch (1960/80, 1977) y continuados, entre otros, por Wright, Andrich, Masters y Linacre han tratado de cumplir con los requisitos de medición planteados por Campbell, evidenciando la posibilidad de mediciones objetivas en las ciencias sociales.

El modelo de Rasch lleva a cabo la transformación monótona de la variable dependiente (la probabilidad de respuesta al ítem) como función aditiva de dos variables independientes: características del sujeto y de los ítems de la variable latente (En el modelo dicotómico de Rasch se utilizó como variables independientes: habilidad personal y dificultad del ítem).

El modelo de Rasch se ha considerado que cumple los requisitos de medición conjunta delimitados por Luce y Tukey (1964), por lo que es un modelo de medición conjunta (Borsboomm, 2005), a la vez que al incorporar probabilidades podría ser calificado como un modelo de medición conjunta probabilístico (Perline, Wright & Wainer, 1979; Oreja-Rodríguez, 2006).

El hecho de incorporar probabilidades, con lo cual recoge las posibles imprecisiones en las mediciones, ha llevado a la conclusión de que el modelo de Rasch puede ser capaz de medir a nivel de intervalo, lo cual implicaría el cumplimiento de los requisitos de la medición fundamental expresada por Campbell.

Bond y Fox (2001) recogen esta posibilidad al subtitular su manual del modelo de Rasch como de medición fundamental en las ciencias humanas.

Rasch (1960/80) desarrolló un modelo dicotómico de medición, en donde las alternativas de respuestas sería 1: contestar correctamente un ítem y 0 contestar incorrectamente a dicho ítem. El desarrollo del modelo se puede apreciar a partir de la definición del ratio odds.

El ratio odds viene a relacionar dos probabilidades opuestas en la realización de un evento.

RO = $\frac{p}{1-p}$, en donde $p+(1-p)=1$

Con el modelo de Rasch, en el contexto de la teoría del rasgo latente, se trata de conocer la relación existente entre un sujeto y un ítem, modelizando la respuesta del sujeto n ante el ítem i. En el modelo de Rasch se plantea que el ratio odds que refleja dicha relación es el resultado de la interacción de dos aspectos. Por una parte la habilidad del sujeto n para enfrentarse al evento reflejado en el ítem i y, por otra, la facilidad del ítem i planteada. Esta relación se puede expresar en una métrica multiplicativa como:

$$\frac{p_{ni}}{1-p_{ni}} = B_n F_i$$

Utilizando una métrica logarítmica, esta relación se puede transformar en:

$$l\left(\frac{p_{ni}}{1-p_{ni}}\right) = lB_n + lF_i$$

Sustituyendo las expresiones lB_n por el símbolo β_n.

A lF_i lo reemplazamos por $l\frac{1}{D_i}$, siendo D_i la dificultad de la cuestión, contraria a la facilidad de la cuestión. Operando con los logaritmos se obtendría que $lF_i = -lD_i$, de ahí se puede llegar a la expresión $lF_i = -\delta_i$.

La relación con el logaritmo del ratio odds sería:

$l(\frac{p_{ni}}{1-p_{ni}}) = \beta_n - \delta_i$, en donde

p_{ni} = probabilidad de que la persona n conteste correctamente al ítem i

$1-p_{ni}$ = probabilidad de que la persona n no conteste correctamente al ítem i

β_n = parámetro de la habilidad del sujeto n

δ_i = parámetro de la dificultad del ítem i

El ratio odds quedaría formulado como:

$\frac{p_{ni}}{1-p_{ni}} = e^{(\beta_n - \delta_i)}$, de ahí despejando p_{ni}, quedaría:

$p_{ni} = \frac{e^{(\beta_n-\delta_i)}}{1-e^{(\beta_n-\delta_i)}}$, expresión del modelo dicotómico de Rasch, que recoge la probabilidad de que la persona n conteste correctamente al item i.

Esta probabilidad se puede formular como $p_{ni} = P[X_{ni} = 1]$, siendo

$X_{ni} = 1$, respuesta correcta del sujeto n al ítem i, dada la habilidad del sujeto n y la dificultad del ítem i.

Esta probabilidad varía entre 0 y 1: $0 \leq P[X_{ni} = 1] \leq 1$

El rango de variación de la diferencia entre la habilidad del sujeto n y la dificultad del ítem i es: $-\infty \leq (\beta_n - \delta_i) \leq +\infty$

Si esta distancia actúa como exponente de la base e, el rango de variación es {0,+∞):

$0 \leq e^{(\beta_n - \delta_i)} \leq +\infty$, con lo que la expresión de la probabilidad p_{ni}, tendría un rango de variación {0,1}: $0 \leq \left(\dfrac{e^{(\beta_n - \delta_i)}}{1 - e^{(\beta_n - \delta_i)}} \right) \leq 1$

El modelo dicotómico de Rasch permite la estimación de la probabilidad de una respuesta correcta a un ítem i por el sujeto n, dada la habilidad del sujeto β_n y la dificultad del ítem δ_i:

$$P[X_{ni} = 1 / \beta_n, \delta_i] = \frac{e^{(\beta_n - \delta_i)}}{1 - e^{(\beta_n - \delta_i)}}$$

La probabilidad de respuesta afirmativa es una función de la distancia entre los parámetros de habilidad del sujeto β_n y de dificultad del ítem δ_i.

Si la diferencia entre ellos es positiva, la probabilidad de respuesta correcta es superior al 50%.

$$(\beta_n - \delta_i) > 0 \Rightarrow P[X_{ni} = 1] \geq 50\%$$

Si la diferencia es nula, la probabilidad de respuesta correcta es del 50%.

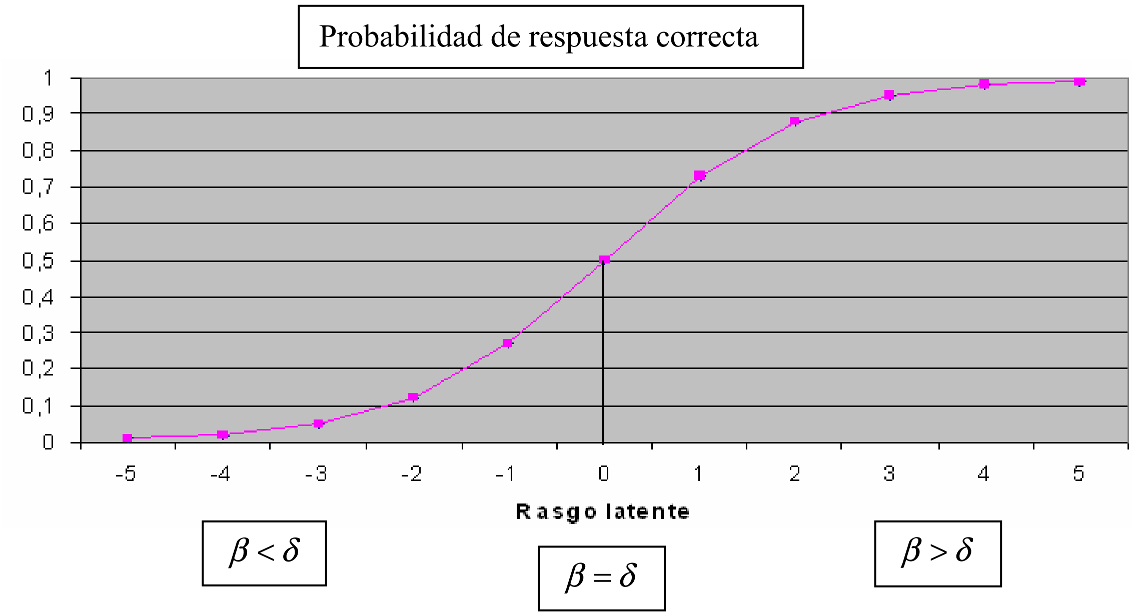
$$(\beta_n - \delta_i) = 0 \Rightarrow P[X_{ni} = 1] = 50\%$$

Si la diferencia es negativa, la probabilidad de respuesta correcta es inferior al 50%

$$(\beta_n - \delta_i) < 0 \Rightarrow P[X_{ni} = 1] < 50\%$$

Se puede representar la curva característica del ítem i, relacionando la distancia entre los parámetros de habilidad del sujeto β_n y de dificultad del ítem δ_i y la probabilidad de respuesta correcta.

$\beta_n - \delta_i$	$P[X_{ni} = 1 / \beta_n, \delta_i]$
5	.99
4	.98
3	.95
2	.88
1	.73
0	.50
-1	.27
-2	.12
-3	.05
-4	.02
-5	.01



2. Medidas

Observaciones.

Las puntuaciones obtenidas en la aplicación de las escalas de los instrumentos de medida de los conceptos son ambiguas respecto a los procesos de generalización al menos que puedan transformarse en medidas. La aplicación de las escalas de Stevens (1946), que ignoró que para que un atributo sea medible debe poseer una estructura aditiva (Michell, 1997), ha llevado a los investigadores a considerar que asignar valores numéricos a objetos (percepciones) es suficiente como si fuera una medición científica (Michell, 1997). Es decir, si en una encuesta se utiliza una escala Likert y se trata la puntuación obtenida en cada percepción de los sujetos como si fuera una medición (con su posterior tratamiento como si fuera una variable intervalo mediante procesos estadísticos específicos para dicha tipología de variables, ya que se está ignorando que las puntuaciones ordinales no tienen estructuras aditivas) se estaría faltando al rigor científico exigido para una medición fundamental de la variable latente analizada (Bond y Fox, 2001).

Siguiendo a estos autores (Bond y Fox, 2001), se puede destacar que tradicionalmente se utilizan las escalas de medidas admitiendo dos asunciones:

1. Cada ítem contribuye de igual forma en la medida del constructo. La realidad es que cada ítem tiene diferente peso en el constructo dado, por lo que los datos deben analizarse de tal forma que la puntuación total resultante de la aplicación de la escala refleje esa diferencia de contribución de los distintos ítem al valor total.
2. Cada intervalo se mide en la misma escala de intervalo. En una escala Likert 1-5 (por ej.) se admite que la distancia entre cada categoría de la escala es uniforme entre los ítems. Esto no es cierto desde la perspectiva representada por las cuestiones y la posición del encuestado. Esta falta de linealidad entre los ítems se manifiesta a lo largo de los ítems.

La dificultad de un ítem es el punto en la variable latente (continuo lineal unidimensional) en el que ambas categorías (acierto/error) en un modelo dicotómico tienen un 50% de probabilidad de ser observada.

Es necesario para el proceso de inferencia que las medidas estimadas por la aplicación del modelo de Rasch a partir de las puntuaciones obtenidas a partir del instrumento de medida del cuestionario sean independientes tanto de la muestra de personas utilizadas como de lo ítems que representan en el cuestionario la variable latente que se pretende medir (Rasch, 1968).

La determinación de las medidas se efectúa a partir de las puntuaciones totales de cada parámetro que son los estadísticos suficiente (Fisher, 1922) para su obtención (Rasch, 1960/1980).

Siguiendo a Andrich y Marais (2005) podemos establecer las matrices de las observaciones X_{ni} de acuerdo a la respuesta al ítem con dificultad δ_i del sujeto con la habilidad β_n

Las respuestas (valores observados / puntuaciones) de n personas, con una habilidad personal β_n (para todo n €{1 a N (tamaño de la muestra)} a i ítems, con una dificultad en cada ítem δ_i (para todo i €{1 a I (longitud del cuestionario)}, se recoge en la siguiente matriz:

X_{ni}	δ_1	Δ_2		δ_i		δ_I	$r_n = \sum_{i=1}^{I} X_{ni}$
β_1	X_{11}	X_{12}		X_{1i}		X_{1I}	R_1
β_2	X_{21}	X_{22}		X_{2i}		X_{2I}	R_2
β_n	X_{n1}	X_{n2}		X_{ni}		X_{nI}	r_n
β_N	X_{N1}	X_{N2}		X_{ni}		X_{NI}	r_N
$s_i = \sum_{n=1}^{N} X_{ni}$	s_1	S_2		s_i		s_I	

La puntuación total del ítem i $s_i = \sum_{n=1}^{N} X_{ni}$ es el estadístico suficiente (contiene toda la información necesaria) para la estimación de la dificultad de dicho ítem. Todos los ítems con la misma puntuación total tendrá la mima estimación del parámetro dificultad.

La puntuación total del sujeto n $r_n = \sum_{i=1}^{I} X_{ni}$ es el estadístico suficiente para la estimación de la habilidad de dicho sujeto. Todos los sujetos con la misma puntuación total tendrán la misma estimación del parámetro habilidad

El modelo de Rasch usa las puntuaciones totales como punto de partida para la estimación de las probabilidades de respuestas. (Bond y Fox, 2001). La probabilidad de respuesta (estimadas como medias) que cualquier persona puede obtener respuestas correcta a cada ítem se recoge en la siguiente matriz:

P_{ni}	δ_1	δ_2		δ_i		δ_I	$r_n = \sum_{i=1}^{I} P_{ni}$
β_1	P_{11}	P_{12}		P_{1i}		P_{1I}	r_1
β_2	P_{21}	P_{22}		P_{2i}		P_{2I}	r_2
β_n	P_{n1}	P_{n2}		P_{ni}		P_{nI}	r_n
β_N	P_{N1}	P_{N2}		P_{ni}		P_{NI}	r_N
$s_i = \sum_{n=1}^{N} P_{ni}$	s_1	s_2		s_i		s_I	

Las sumas de las probabilidades por cada ítem (columnas) se obtienen agregando el número de sujetos que ha contestado correctamente el ítem. Este valor es equivalente a la suma de respuestas correctas de cada sujeto.

De donde:

$$s_i = \sum_{n=1}^{N} X_{ni} = \sum_{n=1}^{N} P_{ni} \qquad \text{ecuación que debe satisfacerse para cualquier ítem}$$

De idéntica forma se puede expresar las sumas de probabilidades (o medias) obtenidas por cada ítem correcto, que debe ser igual al número correcto de respuestas obtenidas.

$$r_n = \sum_{i=1}^{I} X_{ni} = \sum_{i=1}^{I} P_{ni} \qquad \text{ecuación que debe satisfacerse para cualquier sujeto}$$

De las anteriores ecuaciones se puede deducir que:

$$r_n = \sum_{i=1}^{I} \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i}})$$

$$s_i = \sum_{n=1}^{N} \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}}$$

3. Errores

La construcción de sistemas de reproducción de mediciones está siempre en revisión Bond y Fox (2001) nos recuerdan que no hay un modelo verdadero y no hay un modelo que no genere imprecisión. Para trabajar con la imprecisión (errores) es necesario asumirla y proceder a construir medidas que la tenga en cuenta, lo cual es crítico para el avance científico, dado que estos modelos, cuando están basados en estructuras aditivas, son muy útiles para resolver los problemas científicos actuales.

El error del modelo "Model S. E." es la precisión de la medida que indica la borrosidad en la localización de la medida de las facetas (elementos) en la variable latente. Una mejor precisión de las mediciones se logra por medio de:

1. Más observaciones de los elementos, tanto referidos a la muestra de sujetos como de ítems.

2. Mejor ajuste de los elementos, dependiendo de la dificultad del cuestionario

3. Más categorías en las escalas de puntuaciones

Siguiendo a Linacre (2008) se denomina RMSE a la raíz del error cuadrático medio determinado tanto para las personas e ítems.

RMSE del modelo se determina a partir de los datos que se ajustan al modelo, considerando que todo desajuste en los datos es el reflejo de la naturaleza estocástica del modelo. Es el “mejor de los casos” de fiabilidad al informar del límite superior de la fiabilidad de las medidas basadas en este conjunto de ítems para la muestra utilizada. En cambio RMSE real se determina a partir de la consideración de que el desajuste de los datos se debe a diferencias en los mismas con respecto a especificaciones del modelo. Sería “el peor caso” a nivel de fiabilidad, informándonos de un límite inferior de la fiabilidad de las medidas de acuerdo a este conjunto de ítems para la muestra utilizada.

La precisión de la medida indica la reproducidad de las localizaciones de las medidas en la variable latente con datos referidos al mismo colectivo poblacional. Es como la graduación de la escala de medida y se cuantifica con el error estándar de la medición.

Desviación estándar ajustada (Verdadera): D. E. (Verdadera) Ajustada significa desviación estándar de las medidas, ajustadas en el error de medición.

$(\text{D. E. Ajustada})^2 = (\text{D. E. de la medida})^2 – (\text{RMSE})^2$

La desviación estándar ajustada es una estimación de la desviación estándar muestral “verdadera” de la cual se eliminan las desviaciones causadas por el error de medición.

Fiabilidad y separación

El índice de Separación (sea de los sujetos o de los ítems) indica cuántos niveles de medición se pueden distinguir estadísticamente entre las medidas.

Separación= D. E. (Verdadera) Ajustada / RMSE

Índice de Fiabilidad se define como la relación entre la varianza verdadera de la medida en relación con la varianza observada.

La relación entre los índices de Separación y Fiabilidad es:

$\text{Fiabilidad} = \text{Separación}^{2} / (1 + \text{Separación}^2)$

$\text{Separación} = \text{Fiabilidad} / (1- \text{Fiabilidad}^{½})$

Error RMSE	Desviación Estándar Verdadera	Varianza Verdadera	Varianza Observada	Separación	Fiabilidad
1	0	0	1	0	0
1	1	1	2	1	0,5
1	2	4	5	2	0,8
1	3	9	10	3	0,9
1	4	16	17	4	0,94

Adaptación propia a partir de Linacre (2008)

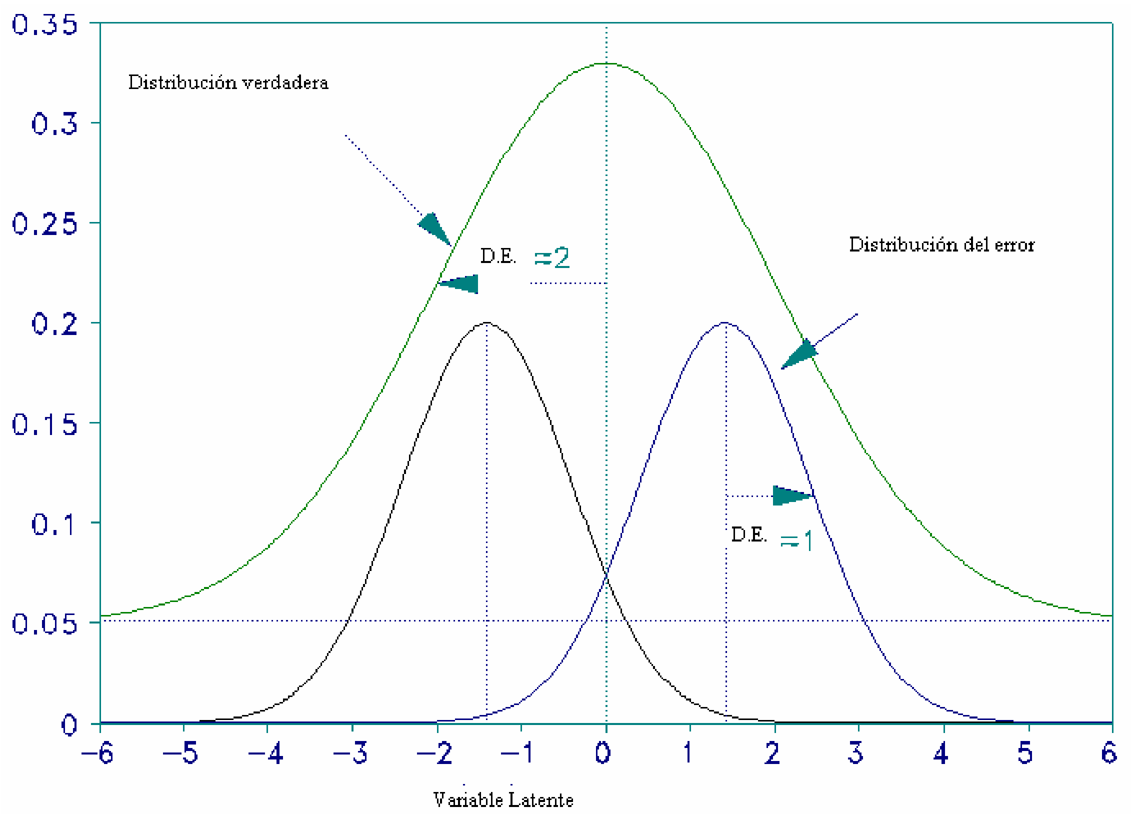
Siguiendo a Wright (1966) se presentan dos ejemplos de gráficos que reflejan la relación entre Separación y Fiabilidad,

Separación = 2

En el caso de que Separación = 2, la desviación estándar de la distribución verdadera = 2 y la desviación estándar de la distribución de error = 1.

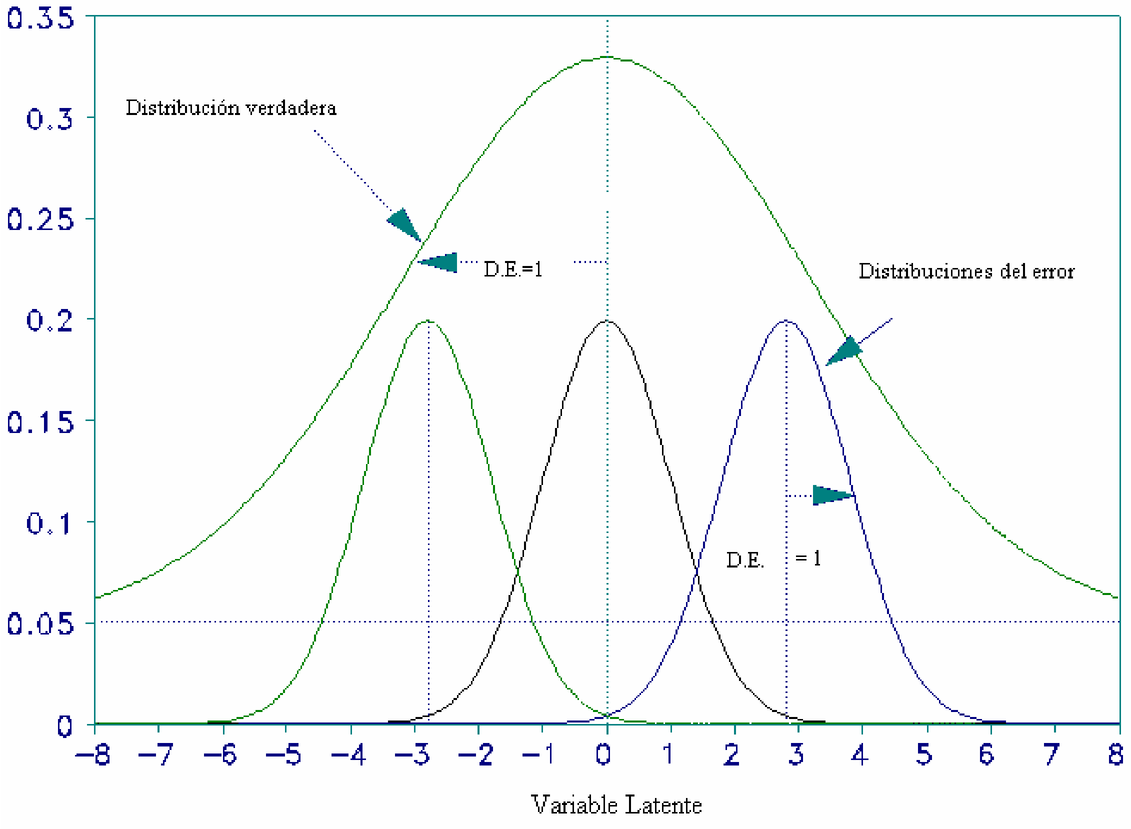
Fiabilidad = (2*2) / (2*2 + 1*1) = 0.8

Separación = 2, indica que hay dos niveles de errores



Separación = 3

Separación = 3, indica que hay tres niveles de errores



En el caso de que Separación = 3, la desviación estándar de la distribución verdadera = 3 y la desviación estándar de la distribución de error = 1.

Fiabilidad = (3*3) / (3*3+ 1*1) = 0.9

4. La Paradoja de Rasch.

El análisis de los conceptos en la actividad científica tiene su razón de ser, siguiendo a Castro et al (2008), en su capacidad representativa – su dimensión ontoepistemológica- y su funcionalidad metodológica. Para estos autores el conocimiento es una compleja conjunción de dos factores: las condiciones que la realidad impone al sujeto cognoscente y ciertas características de nuestro aparato cognitivo. Ello lleva a considerar la presencia de dos posiciones extremas en la filosofía de la ciencia: Realismo ingenuo vs. Constructivismo radical.

El realismo ingenuo tiende a considerar sólo la presencia del primer factor (las condiciones que la realidad impone al sujeto cognoscente) desconociendo la importancia del segundo. Para ellos el conocimiento es como un registro especular de la realidad y la disparidad como un error causado por diferentes factores en último término detectables y corregibles.

Para el constructivismo más radical tiende a considerar el segundo factor (ciertas características de nuestro aparato cognoscente) omitiendo la dimensión realista del conocimiento. Considerando irrelevante la investigación sobre la posible conexión ontológica entre conocimiento y realidad.

En esta disyuntiva se debe adoptar una posición ecléctica entre ambos extremos, derivado de los presupuestos filosóficos que se han adoptado en la actividad científica, que en nuestro caso partiría de la admisión de la importancia de los dos factores de creación del conocimiento, pero admitiendo la prevalencia construccionista.

Desde esa perspectiva los errores provenientes de ambos factores deben ser dectectados y analizados, no eliminados, dada la importancia que en la construcción social de la realidad (Berger y Luckmann, 1968) tendría la información que de ellos se deriva.

Al analizar la importancia del error en la medición de las observaciones y en último extremo en la construcción social de la realidad, se pueden distinguir entre dos tipos de modelos de medición desde la perspectiva de los datos utilizados:

Modelos explicativos de los datos y

Modelos teóricos a los que se ajustan los datos.

En la primera de las aproximaciones se trata de buscar el modelo que mejor se ajuste a los datos disponibles mediante la parametrización del rasgo latente y de las propiedades de los ítemes. Se les podrían considerar modelos de contraste. Como indican Thissen y Orlando (2001) los modelos de esta teoría miden lo que son los ítemes no lo que debería ser. Entre los modelos de variable latente sería un ejemplo los englobados en la Teoría de la Respuesta al Ítem.

La aproximación alternativa se parte de un modelo teórico que dispone de unas propiedades específicas a las que los datos obtenidos se ajustan o no. Estos modelos se denominan de ajuste. Del análisis de los desajustes se obtiene la información necesaria para determinar las calibraciones de ítemes y las medidas convenientes en cada caso. La familia de modelos de Rasch estaría englobada en este caso. En estos modelos no se considera que todos los datos reflejen la realidad analizada y que el modelo teórico representa. En este caso sería posible que algunos datos no reflejasen esa realidad teórica, por lo que no se ajustan al modelo. Sería posible desde esta perspectiva la supresión de datos para lograr un mejor ajuste al modelo.

Siguiendo a Michell (2008) se podría indicar que se usa el Modelo de Rasch como un criterio para construcción de test. El test que se busca nos debería permitir la medición de la habilidad del sujeto y la dificultad del ítem de forma independiente entre ellos, lo que el modelo de Rasch dado que utiliza una función no interactiva promete lograr bajo condiciones óptimas, considerando que permanecen constantes el resto de los atributos que pueden incidir en la presencia de respuestas diferentes de los encuestados.

El modelo de Rasch proporciona mediciones cuantitativas si los atributos relevantes son cuantitativos y no solo ordinales. Podemos advertir esta situación si comparamos el modelo de Rasch con un viejo modelo ordinal como es el modelo de Guttman (1944). Según este modelo una persona contesta correctamente un ítem si y solo si su nivel de habilidad es superior al de dificultad del ítem. Esto es, una persona está en lo:

Correcto si y solo si $X \geq Y$; si no sería incorrecto

El modelo de Rasch es una versión confusa del de Guttman. Se considera que el proceso no es más cercano a la realidad que el modelo de Guttman. A veces a causa de factores de error incontrolados hay personas que falla un ítem aún cuando su habilidad personal excede a la dificultad del ítem e incluso esos errores incontrolados pueden llevar al sujeto a acertar un ítem superior a su habilidad personal. Ello ha llevado a la consideración de que una persona está en lo:

Correcto si y solo si $X + e \geq Y$; si no sería incorrecto.

En donde “e” sería la expresión de la contribución de los factores de error.

En psicometría (Michell, 2008) el error es un concepto no observable, que se asume que su contribución a los resultados se valora siempre de acuerdo a algún estándar preconcebido. No obstante, si se investigara el proceso psicológico real que genera los resultados, la presencia o ausencia del error actual podría identificarse. Michell (2008) enfatiza la importancia del error debido a que éste concepto juega un papel básico en el modelo de Rasch.

La diferencia entre el modelo de Rasch (cuantitativo) y de Guttman (ordinal) es el error "e". El contenido cuantitativo del modelo de Rasch se deriva precisamente de la presencia del error "e". El error es el reflejo de la estructura cuantitativa. No obstante, esto nos lleva a la Paradoja de Rasch. En las ciencias de la naturaleza, si se eliminan los errores la medición cuantitativa sería (por definición) perfecta. En el modelo de Rasch, si se eliminan los errores no sería posible lograr la medición cuantitativa: si las condiciones de encuestación pueden mejorarse hasta el nivel de eliminar permanentemente los errores (e = 0) la medición de la habilidad no mejoraría. El modelo de Rasch se reduciría al modelo de Guttman y sólo se podrían ordenar las habilidades. Pero, por definición, la eliminación de los errores deberían mejorar las observaciones, lo que nos lleva a una conclusión contradictoria.

Si tenemos alguna evidencia independiente de que la medición de que las habilidades personales son cuantitativas (y no simplemente ordinales) deberíamos ser capaces de utilizar el modelo de Rasch para determinar los errores, que tienen un papel fundamental en la determinación de la calibración de los atributos cuantitativo bajo condiciones específicas. Si no tenemos tales evidencias, el ajuste de los datos al modelo de Rasch por si solo no supone evidencia de estructura cuantitativa.

La aceptación de las definiciones de Stevens (1946) ha generado un escaso nivel de investigación en la cualidad cuantitativa de las observaciones usadas en los métodos científicos (Michell 1997, 1999, 2000), lo que ha llevado a otra falacia psicométrica (Michell, 2006) que es considerar que un atributo debe ser cuantitativo simplemente porque la observación sea ordinal.

Bibliografía

Andrich, D. e I. Marais (2005): *Instrument Design with Rasch IRT and Data Analysis I.* Unit Material EDU 435/635. School of Education. Murdoch University

Berger, P. L. y Th. Luckmann (1968): *La construcción social de la realidad*. Buenos Aires: Amorrortu

Bond, T. G. y C.M. Fox (2001): *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*. Mahwah, N.J. : Erlbaum

Borsboom, D. (2005): *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge: Cambridge University Press

Borsboom, D y A. Zand Scholten (2008): "The Rasch Model and Conjoint Measurement Theory from the Perspective of Psychometrics" *Theory & Psychology*. Vol 18(1):111-117

Campbell, N. (1953): *What is Science?* Dover Publications, INC. New York. (Originalmente publicado en 1921, por Methuen & Co. Ltd)

Castro, L; M. A. Castro y J. Morales (2008): *Metodología de las Ciencias Sociales. Una introducción crítica*. Madrid: Tecnos. 2ª Edición.

Fergurson, A.; Myers, C.S.; Bartlett, R. J.; Banister, H, Bartlett, F.C.; Brown, W. et al (1940): "Quantitative estimates of sensory events: Final Report of the Committee Appointed to Consider and report upon Possibility of Quantitative Estimates of Sensory Events". *Advancement of Science*, 1, 331-349 (citado por Borsboom & Zand WScholten, 2008).

Fisher, R. A. (1922): "On the mathematical foundations of theoretical statistics". *Proc. Roy. Soc.* Vol CCXXII p. 309 – 368

Guttman, L. (1944): " A basis for scaling qualitative data" *American Sociological Review*, 9, 139-150

Linacre,M. (2008): "Reliability and separation of measures". *A User's Guide to Winsteps / Ministep Rasch – Model Computer Programs. Program Manual 3.65.0* www.winsteps.com

Luce, R. D. y J.W. Tukey (1964): "Simultaneous conjoint measurement: A new scale type of fundamental measurement" *Journal of Mathematical Psychology*, I, 1-27

Michell, J. (1997): "Quantitative science and the definition of measurement in psychology". *British Journal of Psychology*, 88, 355-383 (citado por Michell, 2008).

Michell, J. (1999): *Measurement in psychology: A critical history of a methodological concept*. Cambridge: Cambridge University Press (citado por Michell, 2008).

Michell, J. (2000): "Normal science, pathological science y psichometics". *Theory & Psychology*. Vol 10:639-667 (citado por Michell, 2008).

Michell, J. (2006): "Psychophysics, intensive magnitudes, and the psichometricians' fallacy" *Studies in History & Philosophy of Biological and Biomedical Sciences*, 17, 414-431 (citado por Michell, 2008).

Michell, J. (2008): "Conjoint Measurement and the Rasch Paradox. A Response to Kyngdon" *Theory & Psychology*. Vol 18(1):119-124

Mosterin, J (1984): *Conceptos y teorías en la ciencia*, Madrid: Alianza

Oreja-Rodríguez, J. R (2006): "Modelos de medición conjunta en administración de empresas: Del análisis conjunto al modelo de Rasch". En Febles, J. y J.R. Oreja [Coordinadores]: *Modelos de Rasch en administración de e*mpresas. Santa Cruz de Tenerife: FYDE-CajaCanarias. Colección E-Book nº 1, pp. 146-166

Perline, R.; Wright, B.D. y Wainer, H. (1979): "The Rasch model as additive conjoint measurement" *Applied Psychological Measurement*, 3, 237-255.

Rasch, G. (1960/1980): *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research. Expanded edition (1980) with foreword and afterword por B.D. Wright. Chicago: The University of Chicago Press.

Rasch, G. (1968) 1968 *A Mathematical Theory of Objectivity and Its Consequences for Model Construction*. Invited paper at the Institute of Mathematical Statistics, European Branch, 6th Sept. 1968. http://www.rasch.org/memo1968.pdf

Stevens, S.S. (1946): On the theory of scales of measurement. *Science*, 103, 677-680 (citado por Michell, 2008)

Thissen, D. y M. Orlando (2001): “Item Response Theory for Items Scored in Two Categories”, en Thissen, D. y W. Waimer [Eds.] : *Test Scoring*. Hillsdale, N.J. : Erlbaum

Wright BD. (1996): “Reliability and Separation” *Rasch Measurement Transactions*, 9:4 p.472