Psicothema, 1999. Vol. 11, nº 3, pp. 663-678
ISSN 0214 - 9915 CODEN PSOTEG

SOFTWARE, INSTRUMENTACIÓN Y METODOLOGÍA

DOS ENFOQUES ANALÍTICOS DEL DISEÑO CON GRUPO CONTROL NO EQUIVALENTE PARA VARIABLES DEPENDIENTES CATEGÓRICAS

Manuel Ato García, Rafael Rabadán Anta y Francisca Galindo Garre
Universidad de Murcia

El análisis estadístico del diseño con grupo control no equivalente (GCNE) con variables dependientes numéricas ha sido un tema de gran interés durante la última década. Mucho menos interés ha demandado el análisis estadístico con variables dependientes categóricas, aunque las mismas ideas se aplican en ambos casos. No obstante, los problemas generados por el error de medida de pretest (y postest) y las variables omitidas dificultan en mayor medida un análisis correcto. En este trabajo se aborda un análisis comparativo de dos enfoques recientemente propuestos para el análisis estadístico del diseño GCNE: el modelo lineal de crédito parcial (Fischer y Ponocny, 1994, 1995) y el modelo multivariante categórico de medidas repetidas (Agresti, 1996, 1997). La interpretación de los resultados con ambos enfoques se apoya en un ejemplo del campo clínico.

Two analytical approaches for the nonequivalent control group design with categorical dependent variables. The statistical analysis of nonequivalent control group design (NECGD) with numerical dependent variables has been a exciting topic of research during the last decade. Much less interest has been demanded for the use of categorical dependent variables in the statistical analysis of the same design, but the main ideas remain. Nevertheless, some dificulties mainly due to measurement error in variables and omitted variables preclude a correct analysis. In this work we compare two recent approaches that may be efficiently used for the statistical analysis of NECGD: the multivariate model with repeated measures (Agresti, 1996, 1997) and the linear partial credit model (Fischer and Ponocny, 1994, 1995). We use a clinical example commenting the interpretation of results with these two approaches.

Correspondencia: Manuel Ato García
Facultad de Psicología. Apdo. 4.021
Universidad de Murcia
30080 Murcia (Spain)
E-mail: matogar@fcu.um.es

Para ilustrar las ideas utilizaremos los datos de un estudio reportado por Hatzinger (1994:151) con dos grupos de sujetos: 64 de un Grupo de Tratamiento o Experimental (GE) que siguió una terapia psicológica y

70 de un Grupo de Control (GC) que no siguió ninguna terapia. Todos ellos respondieron antes y después de la aplicación de la terapia (pretest-postest) a dos preguntas de un cuestionario, cada una de las cuales evaluaba la gravedad percibida de un tipo de síntomas, registrándose sobre una escala ordinal con 4 categorías (de 1, menos grave, a 4, más grave). Para eludir el excesivo número de celdillas vacías se procedió a colapsar las categorías 1-2 en una escala con 3 categorías (0-1-2). En este ejemplo se consideran un total de cuatro ítems, de los cuales dos son variables o indicadores (los dos tipos de síntomas evaluados mediante cuestionario), y los otros dos son condiciones de medida u ocasiones (pretest y postest). Para simplificar la notación llamamos A y B a las dos variables medidas en el pretest; C y D, a las mismas variables medidas en el postest.

Los datos de la Tabla 1 representan 2x2=4 tablas de contingencia resultantes de cruzar cada una de las variables para los dos grupos en las dos ocasiones. Cada tabla tiene un total de 3x3=9 celdillas, constituyendo lo que llamamos *datos agregados*. El panel de la Figura 1 exhibe cuatro gráficos comparativos de la ejecución en cada uno de los ítems para el Grupo de Control y el Grupo Experimental medida en proporciones de respuesta. La primera columna de gráficos corresponde a la primera variable (A para el pretest y C para el postest); la segunda fila corresponde a la segunda variable (B para el pretest y D para el postest).

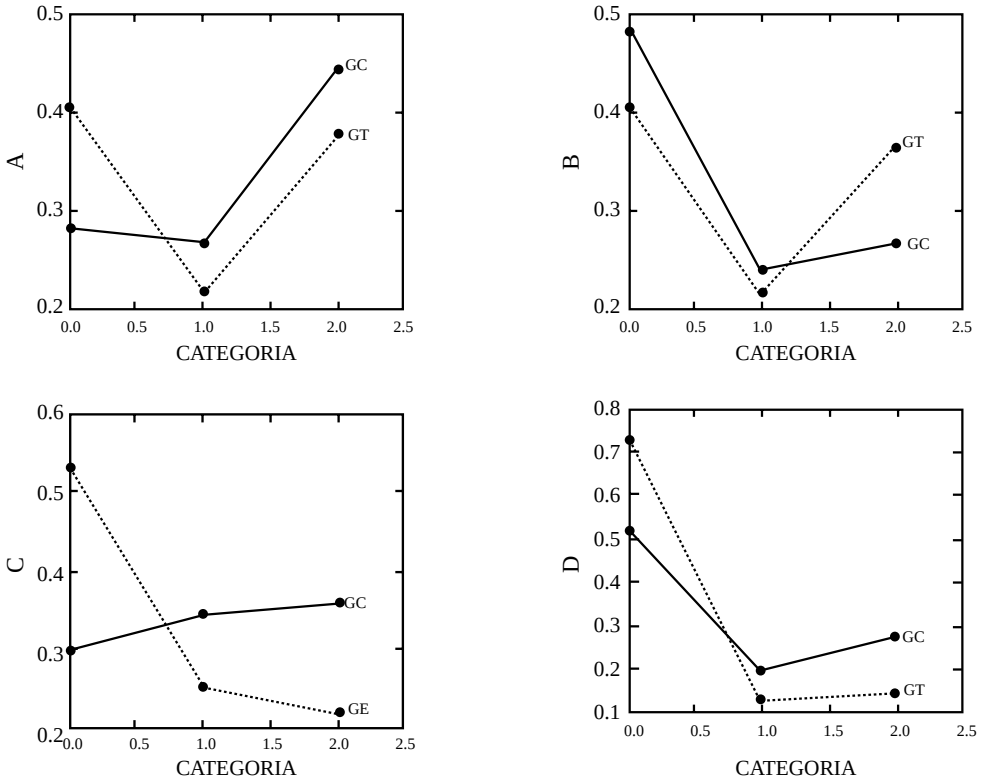

Figura 1

Tabla 1

Síntoma 1 (variables A y C)

	G. Control				G. Tratamiento			
	C (Postest)				C (Postest)			
A (Pretest)	(0)	(1)	(2)	Total	(0)	(1)	(2)	Total
(0)	11	5	4	20	18	6	2	26
(1)	4	9	6	19	8	3	3	14
(2)	6	10	15	31	8	7	9	24
Total	21	24	25	70	34	16	14	64

Síntoma 2 (variables B y D)

	G. Control				G. Tratamiento			
	D (Postest)				D (Postest)			
B (Pretest)	(0)	(1)	(2)	Total	(0)	(1)	(2)	Total
(0)	24	6	4	34	24	1	1	26
(1)	10	3	4	17	11	2	1	14
(2)	3	5	11	19	12	5	7	24
Total	37	14	19	70	47	8	9	64

Tabla 2

Grupo Control

			Postest									
Pretest		C		(0)			(1)			(2)		
A	B	D	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	Total
(0)			7	2	1	2	0	0	1	1	1	15
(1)			0	0	0	1	1	0	1	0	0	3
(2)			0	1	0	0	1	0	0	0	0	2
(0)			2	0	0	3	2	1	2	0	0	10
(1)	(1)		0	0	0	2	0	1	1	0	0	4
(2)			2	0	0	0	0	0	0	0	3	5
(0)			2	1	0	4	0	0	1	0	1	9
(1)			0	1	1	2	1	1	3	0	1	10
(2)			0	0	1	0	1	1	1	2	6	12
Total			13	5	3	14	6	4	10	3	12	70

Grupo Tratamiento

			Postest									
Pretest		C		(0)			(1)			(2)		
A	B	D	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	Total
(0)			13	0	0	3	0	0	0	0	0	16
(1)			3	0	0	0	1	0	1	0	0	5
(2)			2	0	0	2	0	0	0	0	1	5
(0)			4	0	1	1	0	0	0	0	0	6
(1)	(1)		0	1	0	2	0	0	1	0	0	4
(2)			1	1	0	0	0	0	1	1	0	4
(0)			2	0	0	1	0	0	0	1	0	4
(1)			2	0	0	2	0	0	0	0	1	5
(2)			2	0	2	1	1	2	3	2	2	15
Total			29	2	3	12	2	2	6	4	4	64

Una forma más completa de representar los datos consiste en cruzar las variables con las dos ocasiones de medida. En este caso, la tabla de contingencia resultante tiene un total de 3^{2x2}=81 celdillas por grupo y constituye lo que llamamos *datos completos* (Tabla 2).

El análisis clásico

Variables dependientes numéricas

Con variables dependientes numéricas, hay dos enfoques generales para el análisis de un diseño con grupo de control no equivalente, GCNE (Reichardt, 1979; Judd y Kenny, 1981; Stanek III, 1988; Ato, 1995; Trochim, 1996[1]). Ambos enfoques persiguen la eliminación de la variable de asignación, que se supone no aleatoria ni conocida.

Mediante el *enfoque condicional* —o enfoque ANCOVA— se pronostican primero las puntuaciones en el postest sobre la base de la relación entre pretest y postest y se estima el efecto de tratamiento como la diferencia entre las medias de ambos grupos en el postest menos la diferencia ponderada entre las medias de ambos grupos en el pretest. La magnitud de la ponderación es el efecto del pretest. Llamando respectivamente $\overline{Y}_1^C$ e $\overline{Y}_2^C$ a las medias pretest y postest del GC, $\overline{Y}_1^E$ e $\overline{Y}_2^E$ a las medias pretest y postest del GE, y siendo b_1 el efecto del pretest, la estimación del efecto de tratamiento b_2 es:

$$b_2 = \left(\overline{Y}_2^E - \overline{Y}_2^C\right) - b_1\left(\overline{Y}_1^E - \overline{Y}_1^C\right) \qquad (1)$$

Este enfoque se fundamenta en el *modelo mediacional* (Judd y Kenny, 1981:110; Ato, 1995:285), donde el efecto de la variable de asignación sobre el postest no es directo, sino mediado por el pretest y el «supuesto» efecto de tratamiento. Toda la varianza que la variable de asignación comparte con el postest se controla al utilizar como predictores el tratamiento y el pretest, no siendo preciso conocer el efecto de la variable de asignación.

Mediante el *enfoque no condicional* —o enfoque ANOVA— el efecto de tratamiento se estima sustrayendo las diferencias entre grupos en el postest de la diferencia entre grupos en el pretest, y la estimación del efecto de tratamiento b_2 es:

$$b_2 = \left(\overline{Y}_2^E - \overline{Y}_2^C\right) - \left(\overline{Y}_1^E - \overline{Y}_1^C\right) \qquad (2)$$

Este enfoque se basa en el *modelo de cambio* (Judd y Kenny, 1981:117; Ato, 1995:294), donde la variable de asignación se supone que afecta directamente a pretest y postest, pero el pretest no afecta al postest, y además sus efectos se presuponen de igual magnitud para ambos. Esto último, denominado *supuesto de estacionariedad*, conduce a prescindir también de la varianza que la variable de asignación comparte con el postest.

Mediante una simple inspección de la estimación de los efectos de tratamiento (ecuaciones 1 y 2), pueden producir conclusiones estadísticas diferentes sobre el efecto de tratamiento, a menos que el efecto del pretest sea b_1=1, en cuyo caso los resultados de ambos enfoques son iguales. Los dos modelos se han representado en la Figura 2.

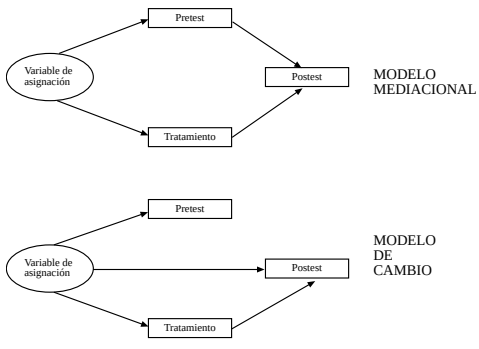

Figura 2

Variables dependientes categóricas

La extensión al caso de variables dependientes categóricas ha sido tratada, entre otros, por Plewis (1985) y Hagenaars (1990: 215-233).

El análisis condicional

El modelo mediacional implica la comparación de dos modelos loglineales: el más complejo postula que el pretest influye en el postest y el tratamiento y éste en el postest, y el más simple prescinde del efecto del pretest sobre el tratamiento. Esta situación es univariante.

Utilizando los datos del ejemplo, el análisis condicional del primer síntoma involucra las variables A (pretest) y C (postest); llamando G a la variable grupo (con dos niveles, Control y Tratamiento), el primer modelo loglineal es [GA,GC,AC] y el segundo [GC,AC]. Del mismo modo, el análisis condicional del segundo síntoma utiliza B (pretest) y D (postest) y los modelos loglineales respectivos son [GB,GD,BD] y [GD,BD]. La Tabla 3 presenta un resumen de los resultados del análisis comparativo de los dos modelos para cada uno de los síntomas, tanto si se consideran las variables ordinales como nominales.

Tomando en consideración la valoración del primer síntoma, los modelos [GA,GC] y [GA,GC,AC] presentan un ajuste aceptable, y el cambio producido por el componente GC es marginalmente significativo (P= .0226 con datos ordinales y P= .0517 con datos nominales). Se puede concluir por tanto que el tratamiento tiene una influencia moderada en el postest.

Tabla 3
Resultados del análisis condicional

Síntoma 1 (variables A y C ordinales)						
				Diferencia de modelos		
Modelo	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$
[GA,AC]	11.739	12	.4669			
[GA,GC,AC]	6.538	11	.8352	5.202	1	.0226
Síntoma 2 (variables B y D ordinales)						
Modelo	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$
[GB,AD]	23.740	12	.0221			
[GB,GD,BD]	10.586	11	.4786	13.146	1	.0003
Síntoma 1 (variables A y C nominales)						
Modelo	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$
[GA,AC]	7.891	6	.2462			
[GA,GC,AC]	1.966	4	.7421	5.926	2	.0517
Síntoma 2 (variables B y D nominales)						
Modelo	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$
[GB,AD]	12.802	6	.0463			
[GB,GD,BD]	0.906	4	.9238	11.896	2	.0026

Con respecto al segundo síntoma, el modelo [GB,AD] no alcanza un ajuste aceptable, pero el modelo [GB,GC,AC] sí, y el cambio producido por el componente GC es significativo (P=.0003 con datos ordinales y P=.0026 con datos nominales). La conclusión es la misma.

El análisis no condicional

La aplicación del modelo de cambio no involucra un modelo loglineal, sino un modelo de homogeneidad marginal (Hagenaars, 1990:224). Puede someterse a prueba de forma indirecta la hipótesis de homogeneidad marginal utilizando un procedimiento desarrollado por Caussinus (1966; véase Ato y López, 1996: 285-287) y basado en la ecuación

$$HM = QSP - QS$$

Homogeneidad (marginal) = Cuasisimetría (parcial) - Cuasisimetría

donde los de cuasi-simetría parcial y cuasi-simetría son modelos loglineales. El segundo de ellos contempla todos los ingredientes de un modelo de cuasi-simetría (incluyendo el efecto del tratamiento sobre pretest y postest) mientras que el primero (cuasi-simetría parcial) excluye el efecto del tratamiento. Si la diferencia entre ambos modelos es estadísticamente significativa, el modelo de homogeneidad marginal se rechaza; si no lo es, este modelo se acepta, siendo la conclusión resultante que no existe un efecto debido al tratamiento (Tabla 4).

En consecuencia, respecto al primer síntoma, la hipótesis de homogeneidad marginal se acepta (P=.2958 con datos ordinales y P=.4858 con datos nominales), conclu-

Tabla 4
Resultados del análisis no condicional

Síntoma 1 (variables A y C ordinales)						
				Diferencia de modelos		
Modelo	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$
QS. parcial	2.377	5	.7949			
Qsimetría	1.284	4	.8641	1.093	1	.2958
Síntoma 2 (variables B y D ordinales)						
Modelo	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$
QS. parcial	22.851	5	0.369			
Qsimetría	2.214	4	.6968	9.639	1	.0019
Síntoma 1 (variables A y C nominales)						
Modelo	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$
QS. parcial	1.502	4	.8263			
Qsimetría	0.058	2	.9715	1.444	2	.4858
Síntoma 2 (variables B y D nominales)						
Modelo	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$	Desvianza	*gl*	$P(D \geq \chi^2_{gl})$
QS. parcial	11.277	4	.0236			
Qsimetría	1.508	2	.4705	9.769	2	.0076

yendo que no hay efecto de tratamiento. Respecto al segundo síntoma, la hipótesis de homogeneidad marginal se rechaza (P= .0019 con datos ordinales y P= .0076 con datos ordinales), concluyendo ahora la existencia de un efecto de tratamiento.

El problema

Ambos análisis permiten constatar dos problemas generales en el análisis de datos del diseño con grupo de control no equivalente con variables dependientes categóricas:

– En primer lugar, las conclusiones obtenidas con un enfoque condicional pueden diferir de las obtenidas con un enfoque no condicional;
– En segundo lugar, con ambos enfoques sólo puede utilizarse una variable dependiente a un tiempo, a diferencia de lo que sucede con variables dependientes numéricas.

El primero de estos problemas es, ya se comentó anteriormente, común a la experiencia recogida en el análisis de datos con variables dependientes métricas. Los enfoques abordados (condicional y no condicional) tratan, por diferentes caminos, de corregir el confundido que las diferencias iniciales en el pretest pueden producir sobre las diferencias encontradas en el postest. Pero, ¿qué análisis es el correcto? Algunos recomiendan usar el análisis no condicional con cierta corrección de la fiabilidad (e.g. Trochim, 1986), pero el análisis condicional es todavía la estrategia analítica más recomendada. Judd y Kenny (1981; véase también Achen, 1986) argumentan que todo depende de la forma como opera el factor de selección y en función de ello recomiendan utilizar el modelo mediacional si se supone que el efecto de la variable de asignación sobre el postest está mediado por el pretest, o el modelo de cambio si se supone que el efecto de la variable de asignación es directo sobre pretest y postest. Pero resulta muy difícil decidir cuándo aplicar uno u otro modelo si no se posee ninguna información acerca de cómo actúa realmente la variable de asignación.

Por otra parte, un problema bastante discutido en la literatura concierne al efecto del error de medida en el pretest y en el postest. Con variables dependientes métricas, es un hecho demostrado que el efecto del error de medida es más grave en el pretest que en el postest (Ato, 1995:285-90; Trochim, 1986), pero con variables dependientes categóricas, y particularmente si se utiliza un modelo loglineal (que no distingue variables explicativas de variables de respuesta), el problema del error de medida afecta de igual manera a pretest que a postest.

Un concienzudo análisis de Hagenaars (1990: 215-232) aporta algunas soluciones a esta cuestión. Ante todo, la no fiabilidad de las medidas pretest y postest y la exclusión de variables relevantes puede fácilmente ser manejada mediante la utilización de modelos loglineales con variables latentes. En este sentido, el análisis puede potenciarse sobremanera si se emplea más de un indicador de pretest y postest, como ocurre en el ejemplo.

Sin embargo, este enfoque tiene varios inconvenientes. El principal de ellos es que la interpretación de los parámetros no se basa en variables observables, sino en la relación entre variables observables y latentes o, más complejo todavía, en la relación entre variables latentes. Además, el *software* disponible para ajustar este tipo de modelos es poco asequible (por ejemplo, no se encuentra en ningún paquete estadístico profesional).

Un enfoque alternativo que utiliza también variables latentes, aunque considerándolas parámetros superfluos (que no son usualmente objeto de interpretación), toma como modelo analítico básico el modelo binario de Rasch (1960), el cual permite hete-

rogeneidad inter e intrasujeto en las distribuciones de la respuesta. Un conocido trabajo de Tjur (1982) demostró una conexión por la cual pueden estimarse los parámetros de ítem en el modelo de Rasch utilizando los estimadores ordinarios de un modelo log-lineal de cuasi-simetría equivalente. La extensión del modelo de Rasch al caso de medidas repetidas fue probada por Duncan (1979, 1984); Conaway (1989) la generalizó a datos politómicos. Involucra necesariamente el uso del enfoque no condicional, tal y como se presentó más arriba (véase Tabla 4).

Dos modelos específicos —que tienen en común el uso de variables latentes como parámetros superfluos— pueden resultar especialmente útiles en el análisis de datos para diseños GCNE con variables dependientes categóricas:

– el primero, propuesto desde el marco de la TRI, es el *modelo lineal de crédito parcial* (MLCP) con aplicaciones a la medida del cambio de Fischer y Ponocny (1994, 1995);

– el segundo, propuesto desde el marco de los modelos loglineales de cuasi-simetría, es el *modelo multivariante categórico de medidas repetidas* (MMCMR) de Agresti (1996, 1997).

En este trabajo se realiza un análisis comparativo de ambos modelos.

El modelo lineal de crédito parcial

En una serie de trabajos, Fischer y sus colaboradores (1972-1997) desarrollaron una familia de modelos para respuestas categóricas dicotómicas y politómicas, derivados del original de Rasch (1960), cuya característica principal es que pueden imponerse restricciones lineales y medidas repetidas, lo que los convierte en modelos apropiados para la evaluación del cambio. Los modelos básicos requieren que los ítems sean unidimensionales (Fischer, 1997); el más básico es el *modelo logístico lineal de test* (Fischer, 1995a,b). Los modelos más complejos no requieren el supuesto de unidimensionalidad de los ítems (Fischer, 1995c; Fischer y Seliger, 1997); es ejemplo el *modelo lineal de crédito parcial* (Fischer y Ponocny, 1994, 1995).

El modelo de crédito parcial fue originalmente propuesto por Masters (1982, 1985) para el análisis de datos politómicos ordenados. Se basa en la idea de que cuando las categorías ordenadas de un ítem se numeran de 0 a *H* hay *H* pasos que un individuo *i* puede dar en su respuesta al ítem *j*, definiendo ese número de pasos el crédito que el individuo recibe. Además, la probabilidad de que un individuo pase el umbral definido entre las categorías *h-1* y *h* depende de la habilidad o propensión latente del individuo y la localización del umbral. Puesto que ambos se localizan en el mismo continuo latente, el modelo es unidimensional (Thissen y Mooney, 1989; Hatzinger, 1994; Heinen, 1996).

A partir de los trabajos de Glas y Verhelst (1989), Fischer y Ponocny (1994, 1995) definieron una extensión de este modelo, que denominaron *Modelo Lineal de Crédito Parcial* (MLCP):

$$\phi_{hij=} \frac{exp(h\alpha_i + \beta_{hj})}{\sum_{h'=0}^{H} exp(h'\alpha_i + \beta_{h'j})} \qquad (3)$$

donde, para hacer el modelo identificable, los parámetros se normalizan haciendo β_{0j}= 0 para todo *j* y $\sum_h \sum_j \beta_{hj}$= 0.

Este modelo puede ser linealmente reparametrizado para permitir la inclusión de variables explicativas sustituyendo los parámetros β_{hj}. La reparametrización propuesta por Fischer y Ponocny (1994:179; 1995: 357) es

$$\beta_{hj} = \sum_{p} w_{hjp}\beta + hc \qquad (4)$$

donde β_p son los parámetros de interés, que miden los efectos de las condiciones experimentales sobre la respuesta, w_{hjp} son pesos de los parámetros β_p y *c* es una constante normalizadora.

El correspondiente modelo loglineal es una combinación de las ecuaciones 3 y 4, que se formula como (Hatzinger, 1994: 150)

$$log(m_y) = \sum_{j}\sum_{h}\lambda_{hj} + \sigma\left(\sum_{j} y_j\right) + \sum_{p}\beta_p u_{jp} \qquad (5)$$

donde $\sigma(\sum_j y_j)$ es un factor de simetría ordinal y u_{jp} es el estadístico suficiente para β_p y $\sigma(\sum_j y_j)$.

Hatzinger (1994:152-3) analizó los datos del ejemplo con este modelo. Los parámetros de interés (β_p) del análisis utilizan un efecto de tendencia (TEND), que representa el efecto diferencial conjunto entre pretest y postest para el grupo de control, y un efecto de tratamiento (TRAT), que se refiere al efecto diferencial conjunto entre pretest y postest para el grupo de tratamiento. Para la definición de ambos efectos se utilizan conjuntamente las dos variables registradas en el pretest (A, para síntoma 1 y B, para síntoma 2). Por su parte, los parámetros λ_{hj} se obtienen sumando las categorías de cada variable en los dos momentos de medida.

El proceso de modelado se realizó utilizando el programa LEM de Vermunt (1997) y se resume en la Tabla 5. Los parámetros λ_{hj} son aquí L12, que representa una restricción de igualdad para la última categoría de las variables A y C, y L21 y L22, que representan sendas restricciones de igualdad para las penúltimas categorías y para las últimas categorías de B y D, respectivamente. El efecto de tendencia (TEND) se define mediante una restricción de igualdad entre las variables A y B; representa el efecto diferencial conjunto entre pretest y postest, tomando como referencia el pretest. El efecto de tratamiento (TRAT) compara el efecto de tendencia conjunto entre los dos grupos, tomando el grupo

Tabla 5
Aplicación del MLCP de Fischer-Ponocny

Estrategia de modelado

Modelo	BIC	AIC	ID	Desvianza	*gl*	*P*
G*SIM	-546.03	-128.74	37.06	159.258	144	.1819
+(L12+L21+L22)	-522.69	-144.10	33.89	137.903	141	.5580
+TEND	-564.39	-158.70	31.01	121.304	140	.8710
+TRAT	-571.20	-168.40	28.61	109.602	139	.9689

Análisis de los estimadores del modelo

Efecto	β	E.típico	exp(β)	Wald	*gl*	*P*
SIM (G=1)				34.70	8	.000
SIM (G=2)				74.81	8	.000
L12	-0.2762	0.3205	0.7587	0.74	1	.389
L21	-0.9142	0.2454	0.4008	13.88	1	.000
L22	-1.3550	0.4405	0.2580	9.46	1	.002
TEND	0.1481	0.1726	0.8623	0.74	1	.391
TRAT	0.9363	0.2821	0.3922	11.02	1	.001

control como referencia. El factor de simetría ordinal se representa con SIM.

El modelo lineal de crédito parcial ajusta óptimamente (D= 109.602 con 139 grados de libertad y P= .9689). La interpretación de estos resultados con el estadístico de Wald destaca una tendencia consistente de reducción conforme se acentúa la gravedad, que para el primer síntoma no es significativa (L12= .74; P=. 389) mientras que para el segundo sí lo es (L21= 13.88; P= .000, y L22= 9.46; P= .002). Además, no existe un efecto de tendencia conjunto entre pretest y postest para el grupo de control (W_{TEND}= .74; P= .391), pero sí se aprecia un efecto de tendencia conjunto para el grupo de tratamiento (W_{TRAT}= 11.02; P= .001).

El empleo de las razones de productos cruzados (*odds ratios*) permite valorar apropiadamente los efectos intrasujeto. Para un determinado sujeto del grupo de control, la razón estimada de encontrar mayor gravedad percibida de ambos síntomas en el postest es exp(.1481)= 1.16 veces la del pretest. Del mismo modo, para un determinado sujeto del grupo de tratamiento, la razón estimada de encontrar mayor gravedad de los síntomas en el postest es exp(.9363)= 2.55 veces la del pretest. Nótese, en primer lugar, que la razón estimada de la diferencia postest-pretest es más de dos veces mayor en el GT que en el GC; además, ambas se refieren conjuntamente a las dos variables que evalúan la gravedad percibida, y por tanto la interpretación adquiere una dimensión multivariante.

El modelo multivariante categórico de medidas repetidas

Como una extensión multivariante de la prueba de McNemar, Agresti (1996,1997) propone un modelo logit, con un vector de efectos aleatorios para cada sujeto, para explicar la correlación entre las medidas repetidas (pretest-postest), que puede ser especialmente útil en el análisis de datos de un diseño GCNE. Un enfoque no paramétrico con los efectos aleatorios involucra un modelo marginal multivariante que, para cada variable, resulta ser un modelo loglineal de cuasi-simetría (Conaway, 1989).

El modelo de Agresti presupone que una muestra de N sujetos responden a J variables (para j=1,…,J), cada una de las cuales puede medirse en K condiciones (para k=1,…,K), y utilizar una escala con H categorías de respuesta (para h=0,…H)[2]. Las condiciones tienen carácter longitudinal y pueden ser medidas diferentes en el tiempo (pretest-postest), diferentes cuestiones de un mismo cuestionario o diferentes tratamientos. Además, pueden incorporarse también efectos intersujetos, por ejemplo, G grupos (para g=1,…,G). Suponiendo variables binarias, las probabilidades específicas de sujeto resultantes configuran el modelo general:

$$logit(\phi_{1i(g)jk}) = \alpha_{i(g)j} + \beta_{jkg} \qquad (6)$$

Es importante notar que, considerando un solo grupo, una sola variable y dos categorías de respuesta para cada condición, este modelo tiene la forma de un modelo de Rasch clásico para las respuestas bajo las diferentes condiciones (Duncan, 1984):

$$logit(\phi_{1ik}) = \alpha_i + \beta_k \qquad (6)$$

donde las k condiciones se corresponden con los ítems del modelo de Rasch clásico. La principal novedad del modelo de Agresti reside en emplear dos o más variables (indicadores) para la misma condición de medida, de ahí su condición de multivariante.

Los N sujetos constituyen una muestra multinomial con probabilidades π_y que puede ser expresada como un modelo log-lineal para las frecuencias esperadas m_y en una tabla de contingencia H^{JxK} para las JK com-

binaciones de las respuestas. El correspondiente modelo loglineal del modelo logit de la ecuación (6) es, para variables binarias,

$$log(m_y) = \sum_j \sum_k \lambda_{gik} y_{jk} + \sigma_g \left(\sum_k y_{1k,\ldots,} \sum_k y_{Jk} \right) \quad (8)$$

donde λ_{gjk} son los parámetros β_{gjk}, y σ_g $(\sum_k y_{1k},\ldots, \sum_k y_{Jk})$ representa un factor de simetría que se define mediante la suma de las diferentes condiciones para cada variable. Puesto que se refiere a un conjunto de variables, el factor de simetría es *multivariante* (Agresti, 1997; Rabadán, Ato, López y Galindo, 1997; Rabadán, 1998); se representa con SIMM (Tabla 6).

Esta fórmula permite la extensión a modelos más generales donde algunas o todas las variables son nominales u ordinales. Dada una variable *j* y una condición *k* para un determinado sujeto *i*, sea y_{hjk}=1 si la respuesta cae en la categoría *h*, y 0 en cualquier otro caso. Siendo además ϕ_{ghijk} la probabilidad de respuesta a la categoría *h* en la variable *j* para el sujeto *i* bajo la condición *k* y en el grupo *g*, la formulación más general del modelo (6) es:

$$\phi_{ghijk} = \frac{exp(\alpha_{hi(g)j} + \beta_{ghijk})}{\sum_{h'} exp(\alpha_{h'i(g)j} + \beta_{gh'jk})} \quad (9)$$

donde los parámetros son cero para la categoría de referencia (en general, *h*=0) de cada variable. El correspondiente modelo loglineal tiene la forma siguiente:

$$log(m_y) = \sum_h \sum_j \sum_k \lambda_{ghij} y_{hjk}$$
$$+\sigma_g \left(\sum_k y_{11k}, \ldots, \sum_k y_{H-1,1k}, \ldots, \sum_k y_{HJk} \right) \quad (10)$$

donde se ha tomado la primera categoría (*h*=0) como referencia. Para una sola variable o indicador *j*, es el modelo nominal de cuasi-simetría para ítems politómicos (Conaway, 1989).

El modelo multivariante categórico de Agresti permite estimar los efectos de condición intrasujeto para cada variable con un enfoque no paramétrico y estimación por máxima verosimilitud condicional (Tjur, 1982; Agresti, 1996). Como apunta Agresti (1993b), los modelos con términos de efectos aleatorios pueden controlar desviaciones de los supuestos y representar efectos de variables omitidas o error de medida en las variables explicativas.

La aplicación del modelo multivariante categórico de medidas repetidas (MMCMR) se resume en la Tabla 6, suponiendo datos ordinales, y en la Tabla 7, con datos nominales.

La correcta interpretación de los resultados se concentra en los parámetros significativos, que básicamente son los efectos intrasujeto para grupo de control y grupo de tratamiento, y requiere la utilización de logaritmos de razones de productos cruzados (*log-odds ratios*).

El parámetro para el efecto A es $\lambda_{C\text{-}A}$ = .2178 y representa la diferencia en la gravedad percibida del primer síntoma (variables A y C) entre pretest y postest para el grupo de control. La razón estimada de la gravedad percibida del primer síntoma en el postest es exp($\lambda_{C\text{-}A}$)= 1.24 veces la del pretest. Esta diferencia no es estadísticamente significativa (W_1^{GC}= .75; *P*= .388). Del mismo modo, el efecto $\lambda_{D\text{-}B}$ representa la diferencia en la gravedad percibida del segundo síntoma (variables B y D) entre postest y pretest para el grupo de control. La razón estimada de la gravedad percibida del segundo síntoma en el postest es exp($\lambda_{D\text{-}B}$)= 1.12 veces la del pretest, una diferencia que tampoco resulta significativa (W_2^{GC}== .17; *P*= .681).

La situación es algo diferente para el grupo de tratamiento. La razón estimada de la gravedad percibida del primer síntoma en el postest es exp($\lambda_{C\text{-}A}$)= 1.48 veces la del pretest, diferencia que no es significativa (W_1^{GT}= 1.08; *P*= .300). Sin embargo, la razón estimada de la gravedad percibida del

segundo síntoma en el postest es exp($\lambda_{D\text{-}B}$)= 4.46 veces la del pretest, y la diferencia es estadísticamente significativa (W_2^{GT}=7.41; *P*=.006).

La interpretación es similar en el supuesto de considerar ítems nominales, pero se concentra sobre las categorías del factor bajo consideración. La conclusión estadística

Tabla 6
Aplicación del MMCMR de Agresti con datos ordinales

Estrategia de modelado

Modelo	BIC	AIC	ID	Desvianza	*gl*	*P*
G*SIMM	-346.11	-85.30	26.12	94.697	90	.3469
+(A+B)	-354.30	-99.29	21.55	76.715	88	.7993
+G. (A+B)	-355.23	-106.02	19.62	65.982	86	.9465

Análisis de los estimadores del modelo

Efecto	λ	E. típico	exp(λ)	Wald	*gl*	*P*
SIMM (G=1)				62.01	35	.003
SIMM (G=2)				62.91	35	.003
A	0.2178	0.2522	1.2433	0.75	1	.388
B	0.1135	0.2759	1.1202	0.17	1	.681
G.A	0.3929	0.3789	1.4812	1.08	1	.300
G.B	1.4959	0.5497	4.4634	7.41	1	.006

Tabla 7
Aplicación del MMCMR de Agresti con datos nominales

Estrategia de modelado

Modelo	BIC	AIC	ID	Desvianza	*gl*	*P*
G*SIMM	-346.11	-85.30	26.12	94.697	90	.3469
+(A+B)	-345.95	-96.73	20.95	75.265	86	.7893
+G. (A+B)	-337.57	-99.95	18.50	64.052	82	.9287

Análisis de los estimadores del modelo

Efecto	λ	E. típico	exp(λ)	Wald	*gl*	*P*
SIMM (G=1)				55.14	35	.016
SIMM (G=2)				54.04	35	.021
A (2)	-0.1764	0.5199	0.8383			
A (3)	0.3623	0.5117	1.4366	1.47	2	.480
B (2)	0.3029	0.4523	1.3538			
B (3)	0.1705	0.5618	1.1858	0.45	2	.799
G.A (22)	0.5173	0.7081	1.6775			
G.A (23)	0.9141	0.7734	2.4944	1.41	2	.494
G.B (22)	1.6798	0.9040	5.3644			
G.B (23)	2.9536	1.1010	19.1757	7.45	2	.024

es la misma, pero un atento examen de las RPC que se presentan en la columna exp(λ) manifiesta claramente el profundo cambio en la gravedad percibida de los síntomas que ha sufrido el grupo de tratamiento respecto al grupo de control.

Fortalezas y debilidades de ambos enfoques

Aunque son en esencia modelos de análisis no condicional, existen diferencias sustanciales entre el modelo lineal de crédito parcial (MLCP) de Fischer y Ponocny y el modelo multivariante categórico de medidas repetidas (MMCMR) de Agresti. Sin embargo, si se consideran las variables como ordinales, ambos producen resultados muy similares. La Tabla 8 resume los parámetros y las razones de productos cruzados para los dos modelos que se comparan. Nótese que tanto los parámetros como las RPC del análisis conjunto (modelo MLCP) son, aproximadamente, el promedio de los obtenidos para el análisis individualizado de cada tipo de síntoma (modelo MMCMR).

El modelo MLCP de Fischer-Ponocny, propuesto desde el marco de la TRI, presenta ciertas ventajas importantes sobre el modelo MMCMR de Agresti, que es representativo de los modelos loglineales de cuasisimetría, en particular las siguientes:

* Es más parsimonioso, es decir, presenta un ajuste apropiado con menor número de grados de libertad. El ajuste del modelo MLCP se produce con una desvianza de D=109.6 para gl=139 y P=.9689; el ajuste del modelo MMCMR se produce con una desvianza de D=65.982 para gl=84 y P=.9036. El mayor grado de parsimonia se obtiene como consecuencia de utilizar un factor de simetría unidimensional independiente con 8 vectores por grupo; por el contrario, el modelo MMCMR requiere definir un factor de simetría multidimensional dependiente con 35 vectores por grupo.

* Es más económico en recursos computacionales, es decir, requiere menos gasto de computador para obtener el resultado. El ajuste del ejemplo con el modelo MLCP con el programa LEM requirió la mitad del tiempo que consumió el ajuste con el modelo MMCMR. Como consecuencia, además, el modelo MLCP puede practicarse con un número mayor de variables y mayor número de categorías que el modelo MMCMR equivalente, que es impracticable con más de 5 variables y 2 condiciones de medida.

* Es más simple de interpretar conceptualmente. La interpretación conjunta del efecto intrasujeto descansa sobre el supuesto de independencia local de las variables (ítems virtuales). En consecuencia, una de las

Tabla 8
Comparación de los modelos MMCMR vs MLCP

Modelos	Parámetros MMCMR	Parámetros MLCP	Razones *odds* MMCMR	Razones *odds* MLCP
G. Control				
Síntoma 1	.2178		1.24	
Síntoma 2	.1135		1.12	
Conjunto		.1481		1.16
G. Tratam.				
Síntoma 1	.3929		1.48	
Síntoma 2	1.4959		4.46	
Conjunto		.9363		2.55

razones que pueden explicar la falta de ajuste de este modelo es, precisamente, la violación del supuesto de independencia local.

El modelo MLCP es por tanto el modelo de elección en aquellas situaciones de investigación pretest-postest donde se requiere utilizar muchas variables, por su simplicidad conceptual y su economía de recursos. El modelo MMCMR, propuesto desde el marco de los modelos loglineales de cuasi-simetría, presenta también las siguientes ventajas sobre el modelo MLCP (véase también Agresti, 1997:319):

* La naturaleza estrictamente nominal de las variables. El modelo MMCMR puede operar con variables nominales u ordinales. Por contra, la estructura del modelo MLCP sólo admite variables ordinales.

* La riqueza interpretativa de los parámetros, que se realiza de forma individualizada para cada variable (síntoma, en el ejemplo que hemos tratado), aunque el modelo corrige la correlación entre ocasiones de medida para un mismo sujeto al utilizar un factor de simetría multidimensional politómico. La magnitud de los parámetros no cambia tanto si se efectúa un análisis particularizado de ambos grupos como si se efectúa un análisis individualizado de ambas variables.

* La comparación con modelos más simples. A diferencia del modelo MLCP, que es un modelo único, el modelo MMCMR permite una cierta simplificación de sus elementos. La simplificación puede afectar tanto a los parámetros incidentales (factor de simetría) como a los parámetros estructurales de cada una de las variables consideradas.

* La versatilidad del modelo MMCMR. Aunque los dos modelos pueden emplearse con cualquier número de variables de agrupamiento, y con dos o más condiciones de medida, en el modelo MMCMR las variables pueden incluso tener diferentes niveles de respuesta (siempre que se mantengan para todas las condiciones de medida).

El modelo MMCMR es, por tanto, el modelo de elección cuando se desea abordar una interpretación sustantiva precisa de los resultados, y compararlos con modelos más parsimoniosos aplicando alguna estrategia de modelado, siempre que se disponga de un número relativamente manejable de variables (por ejemplo, no más de 5 variables dicotómicas). En el caso de variables politómicas, por el contrario, la elección del modelo MMCMR es forzosa si las variables tienen un nivel de medición estrictamente nominal; el modelo MLCP no puede aplicarse en estos casos.

Por otra parte, ambos modelos pueden ser ajustados y sus parámetros estimados mediante cualquier programa con capacidad para ajustar modelos lineales generalizados. Las alternativas más apropiadas son el procedimiento GENMOD del paquete SAS (SAS Institute, 1993), el módulo *glm* del paquete S-PLUS (Mathsoft Inc., 1997), el más básico paquete GLIM (Francis, Green, Payne *et al.*, 1993) y el programa LEM (Vermunt, 1997). Más complicado resultaría emplear el procedimiento GENLOG del SPSS (SPSS Inc., 1996), puesto que este programa requiere definir manualmente los 36 vectores que componen el factor de simetría. Por su versatilidad y las amplias posibilidades de modelado loglineal que ofrece, la mejor opción es probablemente LEM, que hemos utilizado aquí y que además tiene la ventaja de ser *software* gratuito. Un listado de todos los comandos necesarios para ajustar los modelos tratados en este trabajo puede obtenerse solicitándolo a los autores.

Notas

1 Esta cita se refiere a una página WEB sobre Métodos de Investigación Social. La dirección de Internet donde puede consultarse es <http: trochim.human.cornell.edu>.

2 En su formulación original, Agresti desarrolla su modelo con variables binarias.

Referencias

Achen, C.H. (1986): *The Statistical Analysis of Quasiexperiments*. Berkeley, CA: University of California Press.

Agresti, A. (1993): Distribution-free fitting of logit models with random effects for repeated categorical responses. *Statistics in Medicine, 12*, 1969-87.

Agresti, A. (1996): Logit models with random effects and quasi-symmetric loglinear models. En A.Forcina *et al.*: *Statistical Modelling. Proceedings of the 11th International Workshop on Statistical Modelling*, pp. 3-12. Orvieto, IT: Graphos.

Agresti, A. (1997): A model for repeated measurements of a multivariate binary response. *Journal of the American Statistical Association, 92*, 315-21.

Ato, M. (1995): Análisis estadístico I: Diseños con variable de asignación no conocida. En M.T.Anguera y otros: *Métodos de Investigación en Psicología*, pp. 271-303. Madrid: Síntesis.

Ato, M. y López, J.J. (1996): *Análisis Estadístico para Datos Categóricos.* Madrid: Síntesis.

Caussinus, H. (1966). Contribution à l'analyse statistique des tableaux de correlation. *Annales de la Faculté des Sciences de l{'Université de Toulouse, 29*, 77-182.

Conaway, M.R. (1989): Analysis of repeated categorical measurements with conditional likelihood methods. *Journal of the American Statistical Association, 84*, 53-62.

Duncan, O.D. (1979): Testing key hypothesis in panel analysis. In K.F. Schuessler, ed.: *Sociological Methodology, 1980*, pp. 279-89. San Francisco: CA: Jossey-Bass.

Duncan, O.D. (1984): Rasch measurement in survey research: further examples and discussion. In C.F.Turner and E.Martin, eds.: *Surveying Subjective Phenomena, vol. 2*, pp. 367-403. New York, NY: Russell Sage Foundation.

Fischer, G.H. (1995a): The linear logistic test model. En G.H.Fischer e I.W.Molenaar, eds.: *Rasch Models, Foundations, Recent Developments and Applications*, pp. 131-55. New York, NY: Springer-Verlag.

Fischer, G.H. (1995b): Linear logistic models for change. En G.H.Fischer e I.W.Molenaar, eds.: *Rasch Models, Foundations, Recent Developments and Applications*, pp. 156-80. New York, NY: Springer-Verlag.

Fischer, G.H. (1995c): The derivation of polytomous Rasch models. En G.H.Fischer e I.W.Molenaar, eds.: *Rasch Models, Foundations, Recent Developments and Applications*, pp. 293-305. New York, NY: Springer-Verlag.

Fischer, G.H. y Ponocny, I. (1994): An extension of the partial credit model with an application to the measurement of change. *Psychometrika, 59*, 177-92.

Fischer, G.H. y Ponocny, I. (1995). Extended rating scale and partial credit models for assessing change. En G.H.Fischer e I.W.Molenaar, eds.: *Rasch Models, Foundations, Recent Developments and Applications*, pp. 353-70. New York, NY: Springer-Verlag.

Fischer, G.H. (1997). Unidimensional linear logistic Rasch models. En W.J.van der Linden y R.K.Hambleton, eds.: *Handbook of Modern Item Response Theory*, pp. 225-43. New York, NY: Springer-Verlag.

Fischer, G.H. y Seliger, E. (1997). Multidimensional linear logistic models for change. En W.J.van der Linden y R.K.Hambleton, eds.: *Handbook of Modern Item Response Theory*, pp. 323-46. New York, NY: Springer-Verlag.

Francis, B.; Green, M.; Payne, C. *et al.* -eds.- (1993). *The GLIM System. Release 4 Manual*. Oxford, UK: Clarendon Press.

Glas, C.A.W. y Verhelst, N.D. (1989): Extensions of partial credit model. *Psychometrika, 54*, 635-59.

Hagenaars, J. (1990): *Categorical Longitudinal Data*. Newbury Park, CA: Sage.

Hatzinger, R. (1994): *A GLM Framework for Item Response Theory Models*. Unpublished manuscript.

Heinen, T. (1996): *Latent Class and Discrete Latent Trait Models*. Newbury Park, CA: Sage.

Judd, C.M. y Kenny, D. (1981): *Estimating the Effects of Social Interventions*. Cambridge, MA: Cambridge University Press.

Masters, G.N. (1982): A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-73.

Masters, G.N. (1985): A comparison of latent trait and latent class analysis of Likert-type data. *Psychometrika, 49*, 529-44

Mathsoft, Inc. (1997). *S-PLUS 4 Guide to Statistics*. Data Analysis Products Division. Seattle: Mathsoft.

Plewis, I. (1980): *Analyzing change: Measurement and Explanation Using Longitudinal Data*. New York, NY: Wiley.

Rabadán, R. (1998). *El modelo loglineal de Rasch y sus extensiones: Aplicaciones para el análisis de respuestas dependientes con datos categóricos.* Tesis doctoral no publicada. Universidad de Murcia.

Rabadán, R.; Ato, M.; López, J.J. y Galindo, P. (1997). *Standardization of symmetry factors in generalized loglinear Rasch models*. Poster presented at the 10th European Meeting of the Psychometric Society. Santiago de Compostela (Spain).

Rasch, G. (1960): *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen, DE: Denmarks Paedagogiske Institute.

Reichardt, C.A. (1979): The statistical analysis from non-equivalent group designs. En T.D.Cook y D.T.Campbell: Quasi-experimentation: *Design and Analysis Issues for Field Settings*, pp. 147-205. Chicago, IL: Rand McNally.

SAS Institute (1993*). SAS Software: Changes and enhancements 6.08.* (SAS Technical Report P252). Cary, NC: Sas Institute.

SPSS Inc. (1996). *SPSS Advanced (version 7.5)*. Chicago, Il: SPSS Inc.

Stanek III, E.J. (1988): Choosing a pretest-postest analysis. *The American Statistician, 41*, 178-83.

Thissen, D. y Mooney, J.A. (1989): Loglinear item response models, with applications to data from social surveys. En C.C. Clogg, ed.: *Sociological Methodology 1989*, pp. 299-330. Oxford, UK: Basil Blackwell.

Tjur, T. (1982): A connection between Rasch's item analysis model and a multiplicative Poisson Model. *Scandinavian Journal of Statistics, 9*, 23-30.

Trochim, W.M.K. (1986): *Advances in Quasi-experimental Design and Analysis: New Directions for Program Evaluation.* San Francisco, CA: Jossey-Bass.

Vermunt, J. (1997). *LEM: A General Program for the Analysis of Categorical Data*. Unpublished report. Tilburg University.

Aceptado el 25 de noviembre de 1998