

ANALISIS FACTORIAL, LOGICA FUZZY Y SISTEMAS EXPERTOS: APLICACION AL ANALISIS COMPARATIVO DE ESTRUCTURAS FACTORIALES OBTENIDAS A PARTIR DE MUESTRAS DISTINTAS

F. J. Herrero

M. Cuesta

F.J. Grossi

Departamento de Psicología. UNIVERSIDAD DE OVIEDO.

RESUMEN

Una vez obtenidas las estructuras factoriales, sobre muestras diferentes, conviene comprobar por medios objetivos si las soluciones alcanzadas son equivalentes. El método estadístico tradicional consiste en calcular los coeficientes de congruencia que corresponden a los resultados obtenidos.

La problemática asociada a este tipo de análisis, consiste en la elección del criterio de rechazo que determine si dos estructuras son iguales. Normalmente depende de factores subjetivos, basados en la experiencia previa del investigador o de criterios aceptados tradicionalmente como correctos dentro de un área de investigación, y no de reglas estadísticas fijas.

El uso de bases de conocimientos dinámicos, donde un experto humano (psicólogo) expresa o refleja su experiencia previa o la de otros autores, se convierte en una herramienta fructífera a la hora de determinar los coeficientes de congruencia estimados como plausibles dentro de un campo de investigación.

Palabras Clave: *Psicología matemática; análisis factorial; ordenadores; inteligencia artificial.*

SUMMARY

Once the factorial structures are obtained from different samples, correspond to verify by objective ways if the solutions obtained are equivalent. The classical statistic method is based on the calculation of the congruence coefficients that correspond to the obtained results.

The associated problem to this type of analysis, it is to choose of a criterion which determine if two structures are similar. Normally it depends on subjective factors, based in the previous experience of the researcher or criteria accepted classically as corrects within a research area, and not only on fixed statistic rules.

The use of dynamic knowledge bases, where a expert human (psychologist) expresses previous experience and/or the knowledge of others authors, makes it a fruitful tool to determine the congruence coefficients estimated as available within a research area.

INTRODUCCION

El análisis factorial, como es de todos bien sabido, forma parte de la metodología clásica dentro de la Psicología. No obstante, existen ciertos aspectos que son objeto de controversia por parte de los autores: los métodos de extracción de las estructuras factoriales, tipos de rotación y la determinación de estructuras semejantes. Podemos elegir siempre varias alternativas posibles, en principio todas válidas, que están siempre muy condicionadas por factores externos al propio método matemático. De esta forma el método de extracción de los factores conocido como componentes principales, con rotación oblicua, suele ser el más utilizado dentro de la literatura psicológica. Sin embargo, siempre es posible encontrar autores con críticas a este método, enfocando los trabajos por otras vías (ej: no rotar las estructuras inicialmente obtenidas).

El problema básico y esencial de un ciencia es la replicación, o comparación de resultados sobre muestras distintas. Este hecho, dentro del análisis factorial, se convierte en una tarea difícil cuando llegamos al momento de analizar si las estructuras factoriales obtenidas son semejantes o bien deberemos aceptar que son distintas. Generalmente, la aceptación o rechazo de soluciones equivalentes, forma parte del análisis sustantivo de los investigadores, no existiendo métodos o criterios matemáticos que sean aceptados por la generalidad de los investigadores (Barret, 1986; González, Sancerni y Meliá, 1990).

De acuerdo a diferentes puntos de vista metodológicos existe siempre un abanico amplio de técnicas que el investigador puede utilizar, de acuerdo a las características particulares del estudio. De todas ellas, quizá las más utilizadas son las basadas en la obtención de coeficientes de similitud entre los factores, realizado a partir de los pesos factoriales calculados sobre las matrices factoriales que deseamos comparar (ver tabla 1).

COEFICIENTES DE SIMILITUD

Técnica	Objetivo	Inconvenientes
Correlación de Pearson	Mide la semejanza de acuerdo al patrón de los pesos factoriales.	Insensible a las magnitudes de los pesos factoriales.
Raíz Cuadrada de las desviaciones cuadráticas	Evalúa la similitud de acuerdo a la magnitud de los pesos.	Dificultad a la hora de interpretar los valores matemáticos intermedios obtenidos.
Coefficiente de Congruencia	Evalúa la equivalencia de patrones y magnitudes de los pesos.	Sobreestimación y ausencia de pruebas adecuadas de significación estadística.
Índice basado en las variables significativas	Evalúa la similitud de acuerdo a variables significativas (peso en valor absoluto >0.30).	Sobreestimación en los factores marginales.

Tabla 1 Coeficientes de similitud.

En este trabajo, abordaremos una nueva estrategia, basada inicialmente en la formulación los coeficientes de congruencia, donde los resultados son modificados tanto por formulaciones pertenecientes a la lógica borrosa como por mecanismos pertenecientes al campo de la Inteligencia Artificial (Bases de Conocimientos dinámicas).

2.FACTORES DE CONGRUENCIA

El coeficiente de congruencia ha sido desarrollado como un método estadístico que permite medir la similitud de dos soluciones factoriales, determinadas sobre muestras formadas con sujetos diferentes (Tucker, 1951; Broadbooks y Elmore, 1987). Su formulación se basa en la comparación de los pesos alcanzados en las dos soluciones, de acuerdo a la fórmula general:

$$CC_{pq} = \frac{\sum_{j=1}^n (a_{jp}) (b_{jq})}{\sqrt{\sum_{j=1}^n a_{jp}^2 \sum_{j=1}^n b_{jp}^2}} \quad (1)$$

donde...

n : número de variables comunes en las dos muestras

p : factor alcanzado en la primera muestra

q : factor de la segunda muestra

a : peso factorial de la variable j en el factor p

b : saturación o peso de la variable j en el factor q

CC_{pq} : factor de congruencia entre los dos factores estudiados (p, q) que se localiza en el rango $+1$, donde el 0 refleja una total ausencia de congruencia entre los dos factores estudiados, 1 relación o ajuste perfecto entre las soluciones alcanzadas en los dos estudios, y -1 similitud perfecta negativa.

Como podemos observar la fórmula anterior no es nada más que la formulación de la correlación producto momento, donde el numerador representa los productos cruzados de los dos factores analizados mientras el denominador de la formulación recoge las lambdas o autovalores (varianza explicada) por cada uno de los ejes matemáticos.

Una vez obtenido el factor de congruencia se contrasta el resultado a nivel estadístico, de acuerdo a las hipótesis estadísticas establecidas. En nuestro caso, generalmente estas hipótesis de congruencia son:

$$H_0 : p \geq .99 \quad ; \text{(congruente)}$$

$$H_1 : p < .99 \quad ; \text{(no congruente)}$$

que bajo ciertos campos de aplicación son menos estrictos si tenemos en cuenta el nivel estadístico exigido como indicador de soluciones equivalentes:

$$H_0 : p \geq .90 \quad ; \text{(congruente)}$$

$$H_1 : p < .90 \quad ; \text{(no congruente)}$$

es decir, inicialmente asumimos que los dos factores son idénticos (soluciones equivalentes), contrastando las hipótesis de acuerdo a la formulación clásica de la correlación producto momento (Amón, 1980), como la alternativa ofrecida por Glass y Stanley (1970):

$$Z_{C_{pq}} = \log_e \sqrt{(1 + C_{pq}) / (1 - C_{pq})} \quad (2)$$

donde...

\log_e : logaritmo natural.

CC_{pq} : factor de congruencia.

$Z_{C_{pq}}$: Z transformada de Fisher.

Como podemos ver en la tabla 2, este instrumento también puede servir bajo ciertas restricciones, para comparar las soluciones obtenidas por distintos métodos de extracción de la matriz factorial, funcionando como un mecanismo de parsimonia. En este caso, ayuda a determinar que las dos soluciones matemáticas obtenidas son similares, es decir dos factores con el mismo grado de explicación del modelo como de las variables empíricas estudiadas.

Componentes Principales				Factor Principal			
	FACTOR I	FACTOR II	h^2		FACTOR I	FACTOR II	h^2
V1	.62533	.76621	.97811		.6220	.7850	1.0
V2	.71417	-.55535	.81845		.7020	-.5240	.767
V3	.71414	.67949	.97170		.7010	.6810	.956
V4	.87979	-.15879	.79924		.8820	-.1450	.798
V5	.74107	-.57764	.88285		.7790	-.6040	.971
λ	2.734	1.716		λ	2.756	1.740	
$\%$	54.68	34.32		$\%$	55.12	34.80	
	$CC_{11} = .99968$	n. s.					
	$CC_{22} = .99938$	n. s.					

Tabla 2: Coeficiente de congruencia sobre dos métodos de extracción factorial.

Por supuesto, el algoritmo es extensible a muestras distintas siempre que las variables sean las mismas, y hubiesemos tenido cuidado de expresar el mismo orden dentro de las bases de datos. Permitiendo de esta forma la obtención de un valor estadístico sobre la congruencia de las matrices factoriales, de forma tal que facilite la tarea de interpretación sustantiva del psicólogo.

3. COEFICIENTES DE CONGRUENCIA Y LÓGICA FUZZY

Los coeficientes de congruencia, por si mismos son engañosos para evaluar la congruencia de las distintas soluciones factoriales. En realidad, no deja de ser una reformulación de la correlación de Pearson, indicando básicamente relación lineal, y no semejanza entre las medidas. Cuando el valor es bajo no suele haber problemas, pero es posible que un coeficiente de congruencia sea muy alto y sin embargo, lo pares de valores de cada saturación factorial no se parezcan en nada (excepto tras un cambio de escala), como podemos observar en la tabla 3.

Muestra 1			Muestra 2		
	FACTOR I	h^2		FACTOR I'	h^2
V1	.950	.90		.650	.42
V2	.830	.69		.530	.28
V3	.765	.59		.465	.22
V4	.800	.64		.500	.25
V5	.777	.60		.477	.23
λ	3.420		λ	1.397	
%	68.40		%	27.94	
CC11=1.000		n. s.			

Tabla 3: Coeficiente de congruencia alto con estructuras cualitativamente distintas.

Esta problemática de los coeficientes de congruencia, en la medida que se pueden obtener valores altos cuando los factores son distintos pero se aproximan a combinaciones lineales, se intenta evitar en este trabajo con una formulación complementaria basada en la lógica "fuzzy".

La nueva solución asumida parte de la formulación sigmoideal siguiente:

$$m(x) = \frac{1}{1 + e^{-a(x-b)}} \quad (3)$$

y de la definición explícita de dos Estructuras como Equivalentes (m) si cumplen la regla general:

Un Coeficiente de Congruencia Alto en los dos factores estudiados (CCA), y un as varianza explicada del mismo tipo (VET) en los factores analizados.

Que podamos formalizar en formato "fuzzy" como:

$$m_{EE}(x) = \text{MIN}(m_{CCA}, m_{VET}).$$

donde ...

$$m_{CCA}(x) = \frac{1}{1 + e^{-20(CC-.70)}} \quad (4)$$

representando CC el coeficiente de congruencia obtenido por medio de la fórmula 3, y m el valor de posibilidad asociado al coeficiente de congruencia para la proposición "coeficiente de congruencia alto".

y donde ...

$$m_{VET}(x) = \frac{1}{1 + e^{90(D-0.10)}} \quad (5)$$

representando m_{VET} el valor de la posibilidad asociado a la proposición "varianza explicada del mismo tipo", mientras D es el valor obtenido de aplicar la fórmula siguiente:

$$D = \left| \frac{\lambda_p}{n} - \frac{\lambda_q}{n} \right| \quad (6)$$

donde ...

λ_i : autovalor del factor i .

n : número de variables comunes.

El valor final obtenido en m nos da el grado de posibilidad acerca de la posible igualdad entre las dos soluciones factoriales analizadas (FC_{pq}), de acuerdo al rango de valores (0-1):

≈ 1 : Certeza absoluta.

.8 : Alta posibilidad.

La problemática de todo este proceso gira en torno a la toma de decisiones sobre los parámetros de las fórmulas 2, 4 y 5. Es decir sobre la decisión del nivel aceptable de los factores de certeza como indicadores de estructuras semejantes, o lo que es lo mismo soluciones factoriales equivalentes (2). Y en segundo lugar, sobre los parámetros a y b de las funciones sigmoideas que nos permiten

determinar la posibilidad de varianzas explicadas equivalentes (5) y posibilidad de estructuras semejantes (4). Estos parámetros no son fijos, sino que dependen del tipo de aplicación estudiada, reflejando de este modo las divergencias existentes en los propios expertos humanos a la hora de dar una interpretación sustantiva sobre unos resultados numéricos.

No obstante, la ejecución del algoritmo anterior permite solventar la problemática detectada en la tabla 3, siempre que hayamos solucionado el problema de los parámetros, como podemos comprobar en la tabla 4. La determinación de dos estructuras semejantes no sólo por la forma de relacionarse linealmente sino por su capacidad de explicación sobre las variables estudiadas, determina como resultado final el grado de congruencia final de las dos soluciones estudiadas.

Muestra 1			Muestra 2		
	FACTOR I	h^2		FACTOR I'	h^2
V1	.950	.90		.650	.42
V2	.830	.69		.530	.28
V3	.765	.59		.465	.22
V4	.800	.64		.500	.25
V5	.777	.60		.477	.23
λ	3.420		λ	1.397	
	%		68.40	%	
	%		27.94	%	
	CC ₁₁ =1.000	n.s.			
	CF ₁₁ =0.000	Impo.			

Ho: .99
D%: .10

Tabla 4: Resultados del sistema "FUZZY" con los datos de la tabla 3.

Por supuesto, en ciertos casos, puede ser considerada como solución semejante los estadísticos ofrecidos anteriormente en la tabla 3 (misma estructura pero con menor grado de importancia). En este caso la solución dentro de este sistema informático pasa por delimitar un criterio más "relajado" en torno a las diferencias alcanzadas en cada una de las soluciones a la hora de considerar dos estructuras como equivalentes (ver tabla 5).

Muestra 1			Muestra 2		
	FACTOR I	h^2		FACTOR I'	h^2
V1	.950	.90		.650	.42
V2	.830	.69		.530	.28
V3	.765	.59		.465	.22
V4	.800	.64		.500	.25
V5	.777	.60		.477	.23
λ	3.420		λ	1.397	
%	68.40		%	27.94	
CC ₁₁	1.000	n.s.			
CF ₁₁	0.993	Cert.			

Ho: .99
D%: .50

Tabla 5: Resultados del sistema "FUZZY" con los datos de la tabla 2, con criterio "laxo" de diferencias explicativas.

4. INTELIGENCIA ARTIFICIAL Y FACTORES DE CONGRUENCIA

Como hemos podido fácilmente comprobar, dependiendo del criterio tomado, el método tradicional nos conducirá a resultados contradictorios, a no ser que tengamos soluciones completamente distintas. Por otra parte, estos criterios dependen de criterios bastante circunstanciales, como es el tipo de área (aprendizaje, personalidad, experimental, etc.) e incluso pueden variar de unos autores a otros los criterios psicológicos estimadores de estructuras equivalentes (ej: solución rotada ortogonal vs oblicua). Estas particularidades deben ser tenidas en cuenta cuando un experto humano analiza las diferentes soluciones obtenidas, de forma tal que sus conclusiones puedan coincidir con la metodología imperante en un campo determinado, so pena de introducir una nueva controversia que se deberá obligado a defender frente a los otros especialistas.

Esta falta de criterios objetivos, debe ser tenida en cuenta en los sistemas de análisis de datos actuales, de forma que contengan elementos referenciales más allá que los meros algoritmos numéricos (Herman, 1982). De forma tal que la experiencia del campo puede ser implementada de manera que facilite la toma de decisiones sobre el tipo de constantes a utilizar en las formulaciones anteriores.

El resultado final es un sistema informático, creado bajo la filosofía "back-end" (Hand, 1986), encargado de cubrir una etapa concreta del análisis factorial (análisis sustantivo de resultados). Es un módulo, dentro de un sistema más amplio

encargado de ayudar a los investigadores en el análisis estadístico de sus datos. Como podemos observar en la figura 1, es una combinación de técnicas informáticas tradicionales (paquetes estadísticos y lenguajes de alto nivel) junto a las nuevas herramientas provenientes del campo de la IA (bases de datos dinámicas y Sistemas Expertos).

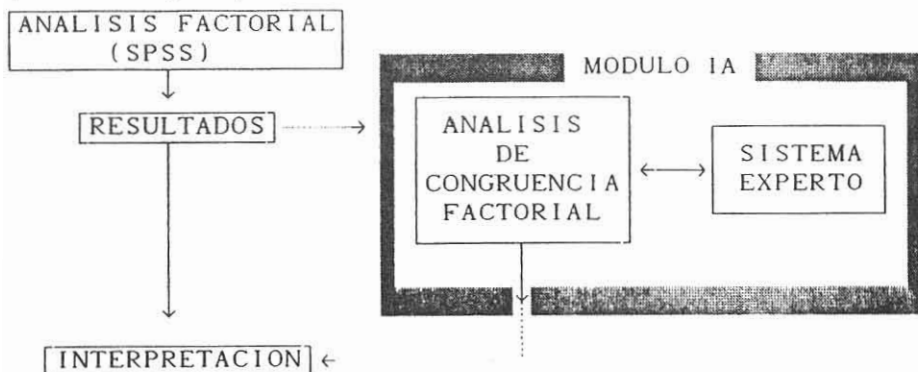


Figura 1: Diagrama general del Sistema.

El módulo principal, escrito en BASIC utiliza un Sistema Experto a la hora de ser fijados los parámetros por parte del investigador. Este, deduce los valores más adecuados de acuerdo a la información suministrada por el usuario, previa consulta de la base de conocimientos. La explicitación de los conocimientos tiene su origen en la información aportada por la propia experiencia de los expertos humanos, o bien por los datos suministrados por la literatura pertinente de cada uno de los campos de investigación contemplados dentro del sistema (ver figura 2).

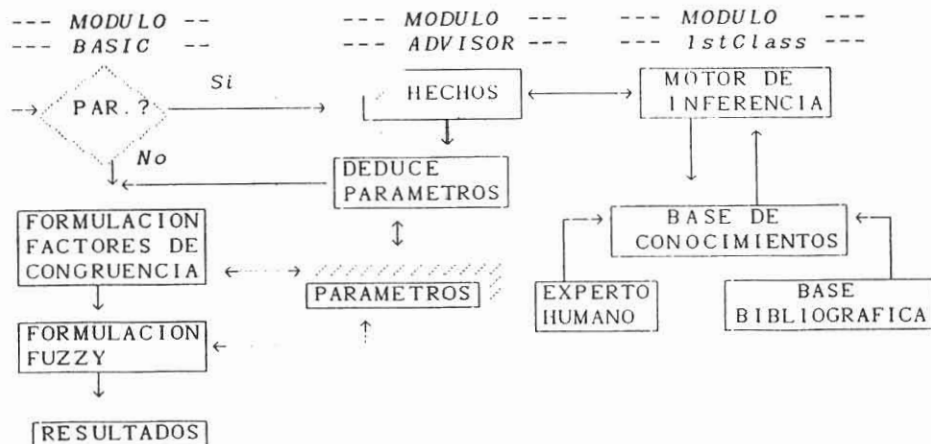


Figura 2: Diagrama funcional del SE.

El sistema de trabajo consta de dos etapas bien diferenciadas:

a) Análisis estadístico tradicional (paquete SPSS), dónde es posible obtener las soluciones factoriales de acuerdo a las restricciones deseadas por el investigador (método, número de factores, tipo de rotación, etc.).

b) Análisis sustantivo de resultados, en la cual intervienen dos módulos de programación bien diferenciados:

b.1) Obtención de los parámetros de referencia gracias a las bases de datos (1stClass)¹. El sistema pregunta al usuario por las condiciones de su investigación: escala de medida, tipo de variables, área de investigación, etc. (ver figura 3), y a continuación determina los parámetros más adecuados de contraste (basándose en la base de conocimientos suministrada previamente al sistema).

b.2) Determinación del grado de coincidencia de las soluciones alcanzadas (formulación de los factores de congruencia y reformulación "fuzzy") gracias al programa CONGRU². El programa se limita a determinar el valor estadístico alcanzado a partir de las soluciones factoriales alcanzadas previamente durante la ejecución del paquete estadístico (SPSS).

Software Estadístico(SPSSX:factor)



Figura 3: Fijación de los parámetros en 1stClass.

1 1stClass ADVISOR (release 3.0), es un "shell" comercial de programación que facilita la obtención de un SE sin tener grandes conocimientos informáticos.

2 Esta implementado en Turbo-Basic, dialecto del lenguaje BASIC desarrollado por la casa comercial BORLAND. La decisión de utilizar este tipo de lenguaje se ha sustentado en la facilidad que tiene este lenguaje para enlazar con lo base de conocimientos (ADVISOR), de forma tal que el usuario puede acceder a ella sin tener un nivel informático elevado.

5. EVALUACIÓN DEL SISTEMA

El algoritmo descrito ha sido evaluado en trabajos pertenecientes al campo del diagnóstico psicológico, mediante la comparación de los resultados ofrecidos por el sistema frente a los aportados por los expertos humanos (Rodríguez et al., 1990), así como suministrando a la máquina información bibliográfica del campo y contrastando si las soluciones ofrecidas por el sistema informático son coincidentes con las interpretaciones humanas. En el ejemplo ofrecido de validación del sistema (ver tabla 6), observamos las soluciones factoriales pertenecientes al test ESPQ (TEA, 1981), para dos muestras pertenecientes al curso de segundo (hombres vs mujeres), tomadas como semejantes en el manual correspondiente.

		Muestra 1		Muestra 2	
		FACTOR I	FACTOR II	FACTOR I	FACTOR II
	V01	-.15	.42	-.08	.51
	V02	-.24	.05	-.03	.36
	V03	-.66	.07	-.60	.16
	V04	.62	-.05	.57	-.01
	V05	-.18	-.64	-.05	-.59
Ho: .90	V06	-.36	-.39	-.44	-.43
D%: .15	V07	-.50	.25	-.34	.24
	V08	-.61	.02	-.63	.13
	V09	-.09	.68	-.05	.58
	V10	.35	.20	.32	.05
	V11	.10	-.50	-.02	-.56
	V12	.69	.14	.66	-.01
	V13	.46	-.41	.38	-.44
λ		2.51	1.75	2.09	1.87
%		0.19	0.13	0.16	0.14
CC ₁₁ =0.978		n. s.			
CF ₁₁ =0.990		Cert.			
CC ₂₂ =0.948		n. s.			
CF ₂₂ =0.981		A. Po.			

Tabla 6: Soluciones factoriales de Test ESPQ.

Como puede ser visto el sistema aporta la misma solución que el experto humano, tanto si valoramos el criterio estadístico de los factores de congruencia apuntado en la tabla 6 por los valores $CC_{11} = 0.978$ y $CC_{22} = 0.948$, como el determinado por la lógica "fuzzy" $CF_{11} = 0.990$ (certeza) y $CF_{22} = 0.981$ (posibilidad alta).

CONCLUSION

En este trabajo hemos visto como la combinación de técnicas estadísticas clásicas (coeficientes de congruencia) y técnicas provenientes de la formulación de la lógica "fuzzy", junto a las herramientas más modernas aportadas por el campo de la Inteligencia Artificial (Sistemas Expertos), facilitan la labor de análisis dentro del campo de la Psicología (interpretación de las soluciones factoriales obtenidas).

Concluyendo, podemos afirmar que estos nuevos sistemas pueden ser considerados una vía fructífera en el procesamiento de datos, al dotar a máquinas con mecanismos más sofisticados que los tradicionales algoritmos numéricos (software numérico clásico). El control cualitativo que pueden realizar sobre los análisis de datos, mediante la incorporación en las bases de conocimientos de los elementos sustantivos utilizados por un experto humano, convierten a estos programas en una "segunda opinión", que es muy interesante en la medida que controlan el análisis de resultados dentro de campos de aplicación concretos.

BIBLIOGRAFIA

- AMON, J. (1980): *Estadística para Psicólogos*. (Tomo II) Madrid: Pirámide.
- BARRET, P. (1986): Factor comparison: an examination of three methods. *Personality and Individual Differences*, 7(3), 327-340.
- BROADBOOKS, W.J. y ELMORE, P.B. (1987): A monte carlo study of the sampling distribution of the congruence coefficient. *Educational and Psychological Measurement*, 47 (1), 1-11.

- GLASS, G.V. y STANLEY, J.C. (1970): **Statistical Methods in Education and Psychology**. Englewood Cliffs, New Jersey: Prentice Hall.
- GONZÁLEZ, V.; SANCERNI, M.D. y MELIÁ, J.L. (1990): Comparación de los factores del cuestionario de satisfacción laboral S20/23 a través de dos muestras. *Psicológica*, **11**, 9-23.
- HERMAN, J. (1982): *Analyse de Donées*. Paris: Masson.
- RODRÍGUEZ, F.J.; HERRERO, F.J.; GROSSI, F.J. y CUESTA, M. (1990): Differential Analysis of the gender in the achievements of subjects at the beginning of formal learning. En *European Perspectives in Psychology* (vol. I). Amsterdam: John Wiley (en prensa).
- TUCKER, L.R. (1951): A method of synthesis of factor analysis studies. (*Personnel Research Section Report No. 984*). Washington, DC: Department of the Army.
- VELDMAN, D.J. (1967): *Fortran Programming for the Behavioral Sciences*. New York: Holt, Rinehart & Winston, Inc.