1993

structural equation modeling with lisrel

egiturazko ekuazioen moldaketa lisrelekin

modelado de ecuaciones estructurales con lisrel

KARL G. JÖRESKOG

29

nazioarteko estatistika mintegia euskadin

seminario internacional de estadística en euskadi

1993

structural equation modeling with lisrel

egiturazko ekuazioen moldaketa lisrelekin

modelado de ecuaciones estructurales con lisrel

KARL G. JÖRESKOG

29

EUSTAT

***Lanketa** / Elaboración:*

Euskal Estatistika-Erakundea /
Instituto Vasco de Estadística (EUSTAT)

***Argitalpena** / Edición:*

Euskal Estatistika-Erakundea /
Instituto Vasco de Estadística
C/ Dato, 14-16 - 01005 Vitoria-Gasteiz

***Botaldia** / Tirada*

1.000 ejemplares
XI-1993

Fotokonposaketarako tratamendu informatikoa:
Tratamiento informático de fotocomposición:

Fotocomposición DIDOT, S.A.
Nervión, 3-6.° Bilbao

***Inprimaketa eta koadernaketa** /*
Impresión y encuadernación:

ITXAROPENA, S.A.
Araba kalea, 45 - Zarautz (Gipuzkoa)

ISBN: 84-7749-155-0

***Lege-gordailua** / Depósito legal:* S.S.836/93

AURKEZPENA

Estatistikako Mintegi Internazionalak sustatzean, hainbat xederekin bete nahi luke Euskal Estatistika-Erakundeak, hala nola:

- Unibertsitatearekiko eta, Estatistika Sailarekiko lankidetza bultzatu.
- Funtzionari, irakasle, ikasle eta estatistikaren alorrean interesaturik leudekeen guztien birziklapen profesionala erraztu.
- Estatistikako alorrean eta mundu-mailan irakasle prestu eta abangoardiako ikerlari diren pertsonaiak Euskadira ekarri, guzti horrek zuzeneko harremanei eta esperientzien ezagupenei dagokienez suposatzen duen ondorio positiboarekin.

Iharduketa osagarri bezala eta interesaturik leudekeen ahalik eta pertsona eta Erakunde gehienetara iristearren, Ikastaro hauetako txostenak argitaratzea erabaki da horrela gure Herrian gai honi buruzko ezagutza zabaltzen laguntzeko asmoarekin.

Vitoria-Gasteiz, 1993ko Azaroa
JOSE MARIA AGIRRE ESKISABEL
Zuzendari Orokorra

PRESENTACION

Al promover los Seminarios Internacionales de Estadística, el Instituto Vasco de Estadística pretende cubrir varios objetivos:

- Fomentar la colaboración con la Universidad y en especial con los Departamentos de Estadística.
- Facilitar el reciclaje profesional de funcionarios, profesores, alumnos y cuantos puedan estar interesados en el campo estadístico.
- Traer a Euskadi a ilustres profesores e investigadores de vanguardia en materia estadística, a nivel mundial, con el consiguiente efecto positivo en cuanto a la relación directa y conocimiento de experiencias.

Como actuación complementaria y para llegar al mayor número posible de personas e Instituciones interesadas, se ha decidido publicar las ponencias de estos Cursos, para contribuir así a acrecentar el conocimiento sobre esta materia en nuestro País.

Vitoria-Gasteiz, Noviembre de 1993
JOSE MARIA AGIRRE ESKISABEL
Director General

OHAR BIOGRAFIKOAK

Karl G. Jöreskog Aldagai-anitzeko Estatistika Analisian irakaslea da Uppsalako Unibertsitatean (Suedia) eta beste erakunde hauetako kide ere bada, Amerikar Estatistika Elkarteko kidea, Estatistikako Errege-Elkarteko ohorezko kidea eta Zientzietako Errege-Akademia Suediarreko kidea.

Ondorengo gaiez anitz liburu eta artikulu idatzi ditu, analisi faktorialaren aplikazio eta teoriaz, egiturazko ekuazioen modeloez eta kobariantzazko egituren analisiaz.

Dag Sörbom irakaslearekin batera LISREL modeloa eta PRELIS eta LISREL programa informatikoak garatu ditu, eta hauek portaera eta gizarte zientzietan saiakuntzazko ikerketan erabilera zabala izan dute.

NOTAS BIOGRAFICAS

Karl G. Jöreskog es profesor de Análisis Estadístico Multivariante en la Universidad de Uppsala (Suecia), expresidente de la Sociedad de Psicometría, miembro de la Asociación Americana de Estadística, miembro honorario de la Sociedad Estadística Real y miembro de la Real Academia Sueca de Ciencias.

Ha publicado varios libros y numerosos artículos sobre teoría y aplicaciones de análisis factorial, análisis de estructuras de covarianza y modelos de ecuaciones estructurales.

Junto con el Profesor Dag Sörbom ha desarrollado el modelo LISREL y los programas informáticos PRELIS y LISREL, ampliamente utilizados en la investigación experimental de las ciencias sociales y del comportamiento.

INDICE

1.- MODELOS DE REGRESION 13

1.1. Una Ecuación Simple de Regresión 13

Ejemplo 1: Regresión del PIB (GNP) 14

1.2. Regresion Bivariada 18

Ejemplo 2: Predicción de Medias de Notas 19

2.- ANALISIS DE TRAYECTORIA 23

Ejemplo 3: Sentimiento Sindical de los Trabajadores Textiles 24

3.- MODELOS DE MEDICION 29

Ejemplo 4: Habilidad y Aspiración 30

4.- ANALISIS FACTORIAL CONFIRMATORIO 35

Ejemplo 5: Nueve Variables Psicológicas - Un Análisis Factorial Confirmatorio . 37

5.- ANALISIS DE TRAYECTORIA CON VARIABLES LATENTES 43

5.1. Sistema Recursivo 43

Ejemplo 6: Estabilidad de Enajenación 44

Ejemplo 7: Rendimiento y Satisfacción 48

5.2. Sistema No Recursivo 52

Ejemplo 8: Influencias de los Iguales sobre Ambición 52

6.- ANALISIS DE VARIABLES ORDINALES 59

Ejemplo 9: Modelo de Panel de Eficacia Política 60

7.- REFERENCIAS 67

SARRERA

Eredu kausal izenarekin ondorengoak izendatzen dira: eredu linealeko hipotesiak osatzen dituzten egiturazko ereduen parametroen estimazioa ahalbidetzen duten teknika eta kontzeptu multzoa.

Aipatu kontzeptuen garapen historikoaren jatorriarekin lotutako eredu kausalek, badute portaera eta gizarte zientzien pareko tradizioa. Baina une honetan, ezaugarri desberdinetako eredu anitzen kontzepzioa gainditu egiten du eredu orokor bakar bat martxan jartzeak. Eredu honek proposatu daitezkeen azpieredu bakoitza ondo errepresentatzeko beharrezko murrizketak onartzen ditu. Beraz, analisiaren garapen fase batean gaude, zeinak egitura linealdun ekuazioen analisiaren eredu orokor gisa izendatzea ahalbidetzen baitu.

Era berean saiakuntzazko ez den ikerketarako aplikazioa halabear historikotzat jo daiteke, esan nahi baita, aldagaien artean erlazio funtsezkoagoak aurkitzeko premiari esker sortu dela eredua, hain zuzen, analisi korrelazionialetik eratorriak baino funtsezkoagoak, honetan ezaguna da aldagaien artean noizbehinkako erlazioak detektatzeko zailtasuna. Hala ere, ereduaren potentzia ez da mugatzen saiakuntzazko ez den arlora, ezpabere ze saiakuntzazko taiuketen funtsezko parametroen estimazioak hartu ditzake oso ondo.

Horregatik gordetzen dute eredu kausalen izena, hurbilketa metodologikoa eta historikoa hartzeagatik, eta eurok dira eredu kausalarekin erlazioa duten apartatuak, hala erabiltzen baita saiakuntzazko kontrola posible ez denean. Eta nahiago dugu egiturazko ekuazioen ereduen izena, espezifikazio, identifikazio eta parametroen estimazioa apartatuengatik, zeren han aurkezten diren metodoa eta arrazonamendua askoz ere orokorragoak baitira, eta ekuazio anitzeko eredu linealen parametroak estimatzea ahalbidetzen du, edozein delarik ere ondorengo interpretazioak sostengatuko dituen taiuketa.

MODELADO DE ECUACIONES ESTRUCTURALES

1. Modelos de Regresión

1.1 Una Ecuación Simple de Regresión

Se usa una ecuación de regresión

$$y = \alpha + \gamma_1 x_1 + \gamma_2 x_2 + \ldots + \gamma_q x_q + z \qquad (1.1)$$

para describir la relación lineal entre una variable dependiente y y un grupo de variables independientes (explicatorias, causales, o predictores) $x_1, x_2, \ldots, x_q$. El término del error aleatorio z se asume estar no correlacionado con las variables independientes. Las variables x pueden ser fijas o aleatorias. Si son fijas, se asume que la distribución de z no dependa de los valores de $x_1, x_2, \ldots, x_q$.

El término de error z es un agregado de todas las variables que influyen en y pero no incluidas en la relación. Empleando otra terminología, z se llama término de perturbación estocástica en la ecuación.

La falta de correlación entre las variables z y x es una suposición crucial. Se deberían planificar y diseñar los estudios para que cumplan con esta suposición. No hacerlo así podría llevar a un sesgo considerable en los coeficientes γ estimados. A veces esto se llama *sesgo de variables omitidas*.

El análisis de regresión es una de las técnicas más ampliamente empleada en las ciencias sociales y otras ciencias; ver Johnston (1.972), Goldberger (1.964), y Draper y Smith (1.967). Normalmente se emplea para calcular la ecuación de regresión y la potencia explicativa relativa de cada variable *x* ó para determinar los mejores predictores de la variable dependiente *y*.

Consideramos el siguiente ejemplo de cómo usar LISREL para calcular una ecuación de regresión simple.

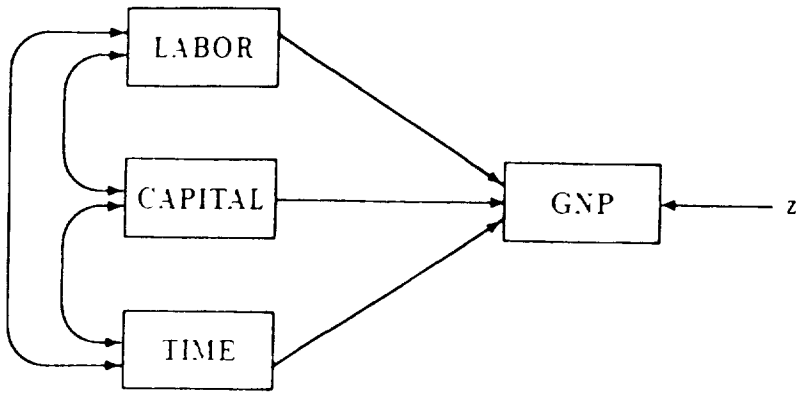

Figura 1.1: Diagrama de Trayectoria para la Regresión del PIB

Ejemplo 1: Regresión del PIB (GNP)

Goldberger (1.964, p. 187) presentó los datos originales sobre y = producto interior bruto en billones de dolares (PIB) (gross national product, GNP), x_1 = entradas laborales en millones de años-hombre (LABOR), x_2 = capital real en billones de dolares (CAPITAL), y x_3 = el período en años medido desde 1.928 (TIME). Un diagrama de trayectoria para la regresión de y sobre x_1, x_2, y x_3 se muestra en la figura 1.1. Los datos constan de 23 observaciones anuales para los Estados Unidos durante el período 1.929-1.941 y 1.946-1.955. La matriz de covarianza de las variables se da en la Tabla 1.1.

Tabla 1.1 Matriz de Covarianza para los Datos del PIB

	y	x_1	x_2	x_3
GNP	4256.530			
LABOR	449.016	52.984		
CAPITAL	1535.097	139.449	1114.447	
TIME	537.482	53.291	170.024	73.747
Means	180.435	15.565	50.087	13.739

Como indica el diagrama de trayectoria, se supone que LABOR, CAPITAL y TIME influyen en GNP. Se indica con las tres flechas de dirección única (unidirigidas) apuntadas hacia GNP. Las tres flechas de dos direcciones a la izquierda indican que las tres variables independientes podrían estar correlacionadas. La flecha de dirección única a la derecha representa el efecto del término de error *z*.

Se pueden computar las medias y la matriz de covarianza, usando el programa PRELIS que tiene en cuenta los valores omitidos y otros problemas en los datos.

El fichero de entrada SIMPLIS es **EX1A.SPL**:

```
Regression of GNP
Observed variables: GNP  LABOR  CAPITAL  TIME
Covariance Matrix
4256.530
 449.016      52.984
1535.097     139.449      1114.447
 537.482      53.291       170.024      73.747
Sample Size:  23
Equation:  GNP = LABOR  CAPITAL  TIME
End of Problem
```

La primera línea es el título. Se puede usar cualquier número de líneas de título y se puede escribir cualquier cosa en las líneas de titulo, con tal que no comiencen con las palabras *Observed Variables* o *Labels*.

La siguiente línea define los nombres de las variables. Los nombres de las variables deben corresponder en orden a las variables en la matriz de covarianza.

Las siguientes cinco líneas definen la matriz de covarianza. Sólo se da la parte inferior de la matriz de covarianza. Los elementos de la matriz de covarianza se dan en formato libre, es decir, con espacios en blanco entre ellos.

La siguiente línea especifica el tamaño de la muestra, es decir, el número de casos en que se basa la matriz de covarianza.

La línea

```
Equation:  GNP = LABOR  CAPITAL  TIME
```

especifica la ecuación de regresión a calcular y es interpretado como "GNP(PIB) depende de LABOR, CAPITAL, y TIME". También se puede especificar como

Paths: LABOR - TIME -> GNP

que se interpreta como "Hay una trayectoria de cada una de las variables LABOR, CAPITAL y TIME a GNP".

La línea End of Problem, que es opcional, se puede emplear para especificar el fin del problema. En este caso, también es el fin del fichero de entrada.

La ecuación de regresión estimada aparece en el fichero de salida (en el fichero de salida R2 aparece como R^2) como

GNP = 3.82*LABOR + .32*CAPITAL + 3.79*TIME, Errorvar.=12.47, R2 = 1

(.22)	(.031)	(.19)	(4.05)
17.7	10.54	20.35	3.08

Los coeficientes de regresión estimados aparecen delante de * antes de cada variable. El coeficiente de regresión parcial estimado de LABOR es 3.82. Se interpreta como sigue. Si LABOR se incrementa una unidad mientras CAPITAL y TIME se mantienen fijas, el incremento esperado de GNP es 3.82 unidades como media.

La varianza de error es 12.47 que es pequeña comparada con la varianza total de GNP 4256.53. Así que las tres variables independientes, LABOR, CAPITAL y TIME dan cuenta de casi toda la varianza de GNP.

Los números debajo de los coeficientes de regresión son los errores estándar de las estimaciones. Cada error estándar es una medición de la exactitud de la estimación del parámetro. Debajo de los errores estándar están los valores *t* . El valor *t* es el ratio entre la estimación y su error estándar. Si un valor *t* es alto, decimos que el parámetro correspondiente es *significante*, que quiere decir que se puede tener bastante confianza en que la variable correspondiente realmente influye a GNP. En nuestro ejemplo, todos los coeficientes de regresion son altamente significantes.

La correlación cuadrada múltiple, R^2 también se da para cada relación. Es una medición de la fuerza de la relación lineal. Un test formal del significado de toda la ecuación de regresión, es decir, un test de la hipótesis que todos los γ son cero, se puede obtener al computar

$$F = \frac{R^2/q}{(1 - R^2)/(N - q - 1)}. \qquad (1.2)$$

donde R^2 es la correlación cuadrada múltiple listado en el fichero de salida, N es el tamaño total de la muestra y q es el número de variables x reales. F se usa como una estadística F con q y N - q - 1 grados de libertad. Para nuestro ejemplo, R^2 = 0.997 y F = 2104.8 con 3 y 19 grados de libertad. (Porque el programa emplea dos puntos decimales por defecto, 0.997 es redondeado a

1.00 y se imprime como R^2 = 1 en el fichero de salida. Para obtener 3 puntos decimales en la salida, incluir la línea

Number of Decimals = 3

en el fichero de entrada.

El modelo de regresión (1.1) es de escala invariante en el siguiente sentido. Si se reemplaza *y* por $y^* = c_0 y$ y x_i se reemplaza por $x_i^* = cx_i, i = 1,2,...q.$ donde los *c* son constantes arbitrarias no de cero, el análisis de estas variables escaladas devolverá coeficientes de regresión

$$\hat{\gamma}_i^* = (c_0/c_i)\hat{\gamma}_i.$$

Similarmente, el error estándar cambiará pero los valores *t* son invariantes bajo las escalaciones de las variables.

El término constante interceptado α en (1.1) es la media de *y* cuando todas las variables *x* son de cero. Cuando sólo se da la matriz de covarianza en un fichero de entrada, LISREL 8 asume que todas las variables se miden por desviación de sus medias y que α es cero. Para estimar α sólo hace falta incluir las medias de las variables en el fichero de entrada.

El fichero de entrada para estimar α es

File EX1B.SPL

```
Regression of GNP
Observed variables: GNP  LABOR  CAPITAL  TIME
Means:  180.435  45.565  50.087  13.739
Covariance Matrix:
4256.530
 449.016      52.984
1535.097     139.449     1114.447
 537.482      53.291      170.024     73.747
Sample Size: 23
Equation:  GNP = LABOR  CAPITAL  TIME
End of Problem
```

Entonces, la ecuación de regresión estimada será

```
GNP =  - 61.73 + 3.82*LABOR + .32*CAPITAL + 3.79*TIME, Errorvar.= 12.47
         (7.77)   (.22)        (.031)       (.19)                  (4.05)
         -7.94    17.7         10.54        20.35                   3.08

         R2 = 1
```

Notar que los parámetros de la regresión estimada y sus errores estándar son iguales que en el ejemplo 1A. El término de intercepto se estima en -61.73 con un error estándar de 7.77. Esto indica que BNP (PIB) sería altamente negativa si LABOR, CAPITAL y TIME fuesen todos cero. Es una extrapolación que asume que la relación sea lineal sobre el rango entero de valores de las variables x.

También se puede realizar un análisis de varianza (ANOVA) y un análisis de covarianza (ANCOVA) por medio de un análisis de regresión al incluir variables postizas en la ecuación de regresión, ver Huitema (1.980) o Jöreskog y Sörbom (1.989, pp. 112-116).

1.2 Regresión Bivariada

El siguiente ejemplo ilustra el caso de dos variables dependientes y_1 y y_2 y tres variables explicativas x_1, x_2, y x_3.

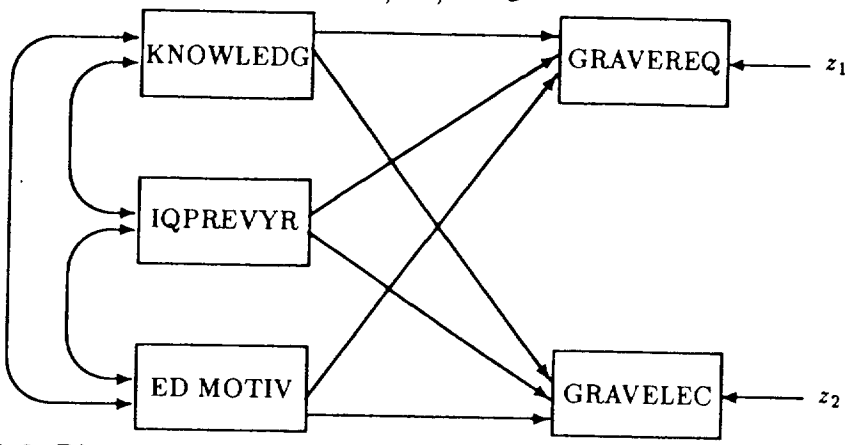

Figura 1.2: Diagrama de Trayectoria para la Predicción de Medias de Notas

Tabla 1.2:Puntuaciones para 15 Estudiantes de Primer Año Universitario sobre 5 Mediciones Educativas

Case	y_1	y_2	x_1	x_2	x_3
1	.8	2.0	72	114	17.3
2	2.2	2.2	78	117	17.6
3	1.6	2.0	84	117	15.0
4	2.6	3.7	95	120	18.0
5	2.7	3.2	88	117	18.7
6	2.1	3.2	83	123	17.9
7	3.1	3.7	92	118	17.3
8	3.0	3.1	86	114	18.1
9	3.2	2.6	88	114	16.0
10	2.6	3.2	80	115	16.4
11	2.7	2.8	87	114	17.6
12	3.0	2.4	94	112	19.5
13	1.6	1.4	73	115	12.7
14	.9	1.0	80	111	17.0
15	1.9	1.2	83	112	16.1

Ejemplo 2: Predicción de Medias de Notas

Finn (1.974) presenta los datos dados en la tabla 1.2. Estos datos representan las puntuaciones de 15 estudiantes de primer año en una gran universidad de Estados Unidos sobre cinco mediciones educativas. Las cinco mediciones son

- y_1 = *nota media de cursos obligatorios cursados (GRAVEREQ)*
- y_2 = *nota media de cursos opcionales cursados (GRAVELEC)*
- x_1 = *test de cultura general de bachillerato, hecho año anterior (KNOWLEDG)*
- x_2 = *coeficiente de inteligencia del año anterior (IQPREVYR)*
- x_3 = *nota de motivación educacional del año anterior (ED MOTIV)*

Examinamos el valor de predicción de x_1 x_2 *y* x_3 *en predecir las medias de nota* y_1 *y* y_2. *El diagrama de trayectoria del modelo se muestra en la figura 1.2.*

La entrada SIMPLIS para este modelo es muy sencilla, ver fichero **EX2A.SPL**:

```
Prediction of Grade Averages
Observed Variables: GRAVEREQ GRAVELEC KNOWLEDG IQPREVYR 'ED MOTIV'
Raw Data
0.8 2.0 72 114 17.3
2.2 2.2 78 117 17.6
1.6 2.0 84 117 15.0
2.6 3.7 95 120 18.0
2.7 3.2 88 117 18.7
2.1 3.2 83 123 17.9
3.1 3.7 92 118 17.3
3.0 3.1 86 114 18.1
3.2 2.6 88 114 16.0
2.6 3.2 80 115 16.4
2.7 2.8 87 114 17.6
3.0 2.4 94 112 19.5
1.6 1.4 73 115 12.7
0.9 1.0 80 111 17.0
1.9 1.2 83 112 16.1
Relationships
    GRAVEREQ = KNOWLEDG  IQPREVYR  'ED MOTIV'
    GRAVELEC = KNOWLEDG  IQPREVYR  'ED MOTIV'
End of Problem
```

Las variables se listan por sus etiquetas. Notar que como la etiqueta para la última variable incluye un espacio en blanco, se debe encerrar entre comillas.

Los datos para este ejemplo son datos originales que se leen en formato libre con una línea por caso. No hace falta especificar el tamaño de la muestra. El programa determina el tamaño de la muestra al contar el número de líneas de datos.

En la mayoría de las aplicaciones los datos originales contienen muchos casos y entonces se recomienda almacenar los datos en un fichero de datos en vez de en un fichero de entrada.

Las relaciones especifican que hay dos ecuaciones para calcular, una para cada variable dependiente. Cada ecuación se especifica listando primero la variable dependiente y luego las tres variables independientes.

Las tres variables a la derecha en cada ecuación son variables consecutivas, así que se puede reemplazar la segunda variable de la derecha por un guión (-) y escribir las dos relaciones:

```
GRAVEREQ = KNOWLEDG - 'ED MOTIV'
GRAVELEC = KNOWLEDG - 'ED MOTIV'
```

Las dos relaciones se pueden simplificar aún más al reconocer que las dos variables de la izquierda son también variables consecutivas y las variables a la derecha son las mismas en las dos relaciones. Las dos relaciones se pueden especificar en una sola línea:

```
GRAVEREQ - GRAVELEC = KNOWLEDG - 'ED MOTIV'
```

El programa interpreta esto como "Cada una de las variables de GRAVEREQ a GRAVELEC depende de todas las variables de KNOWLEDG a ED MOTIV".

Alternativamente, se puede especificar el modelo con trayectorias:

```
KNOWLEDG - 'ED MOTIV' -> GRAVEREQ - GRAVELEC
```

que significa

```
KNOWLEDG - 'ED MOTIV' -> GRAVEREQ
KNOWLEDG - 'ED MOTIV' -> GRAVELEC
```

Las dos relaciones se estiman como

```
GRAVEREQ =  - 5.62  + 0.085*KNOWLEDG + 0.0082*IQPREVYR - 0.015*ED MOTIV,
             (5.61)  (0.027)          (0.049)           (0.11)
             -1       3.17             0.17             -0.13

            Errorvar.= 0.26 , R2 = 0.57
                      (0.11)
                       2.35

GRAVELEC =  - 20.4 + 0.047*KNOWLEDG + 0.15*IQPREVYR + 0.13*ED MOTIV,
             (5.4)  (0.026)          (0.047)         (0.11)
             -3.78   1.82             3.12            1.17

            Errorvar.= 0.24, R2 = 0.69
                      (0.1)
                       2.35
```

Los valores *t* muestran que sólo KNOWLEDG es un predictor significante de GRAVEREQ, y sólo IQPREVYR es un predictor significante de GRAVELEC. ED MOTIV no es significante para ningún propósito. Sin embargo, hay que subrayar que el tamaño de la muestra es demasiado pequeño para sacar conclusiones con seguridad.

El fichero de salida también contiene una sección que se llama ESTADISTICAS DE BUEN AJUSTE, y la primera línea es

CHI-SQUARE WITH 1 DEGREE OF FREEDOM = 8.89 (P = .0029)

Por defecto, LISREL 8 asume que los dos términos de error z_1 y z_2 no son correlacionados. Es una asunción que significa que y_1 y y_2 serían no correlacionados después de quitar los efectos de x_1, x_2 y x_3. El valor de cuadrado chi 8.89, dado en el fichero de salida, es un test formal de esta asunción. Como es significante, podría ser interesante calcular la covarianza de error, es decir, la covarianza entre z_1 y z_2. Se puede hacer esto incluyendo la siguiente línea en el fichero de entrada, ver fichero EX2B.SPL,

Let the Error Terms of GRAVEREQ and GRAVELEC be Correlated

Alternativamente, podría especificarse como

Set the Error Covariance between GRAVEREQ and GRAVELEC Free

El fichero de salida para el modelo con términos de error correlacionado muestra las mismas relaciones estimadas pero también da la matriz de covarianza estimada de los dos términos de error z_1 y z_2 como

Error Covariance for GRAVELEC y GRAVEREQ = 0.17
(0.09)
1.88

el error de covarianza es 0.17. La correlación parcial entre z_1 y z_2. para x_1, x_2, x_3 dado es $0.17/\sqrt{(0.26x0.24)} = 0.68$.

2. Análisis de Trayectoria

El análisis de trayectoria, gracias a Wright (1.934), es una técnica para asesorar la contribución causal directa de una variable a otra en una situación no experimental. El problema, en general, es calcular los coeficientes de un grupo de *ecuaciones estructurales lineales*, representando las relaciones de causa y efecto según el investigador. El sistema incluye variables de dos tipos: variables independientes o de causa $x_1, x_2, \ldots x_q$ y variables dependientes ó de efecto $y_1, y_2, \ldots y_p$. La técnica clásica consiste en resolver primero las ecuaciones estructurales para las variables dependientes en términos de los términos de perturbación independiente y aleatoria $z_1, z_2, \ldots z_p$ para obtener las *ecuaciones de forma reducida,* calculando las regresiones de las variables dependientes sobre las variables independientes y luego resolviendo los parámetros estructurales en términos de los coeficientes de regresión. El último paso no es siempre posible . Modelos de este tipo y una variedad de técnicas de estimación han sido estudiados ampliamente por los econométricos: ver Theil (1.971), por los biométricos; ver Turner y Stevens (1.959) y las referencias incluidas, y por los sociológos; ver Blalock (1.985) y Duncan (1.975).

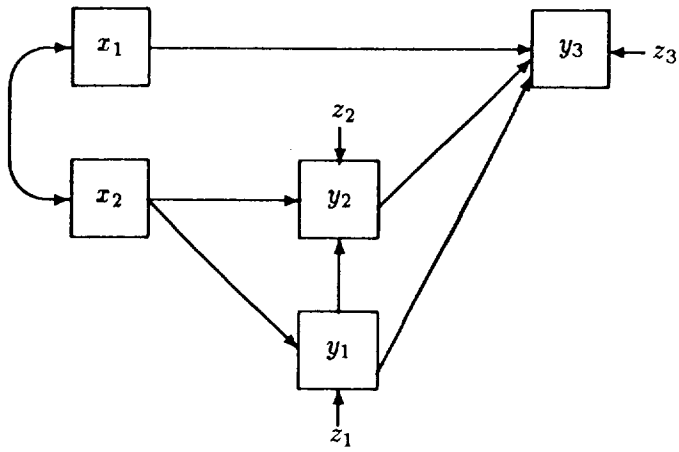

Figura 1.3: Diagrama de Trayectoria para el Modelo de Sentimiento Sindical

Algunos de estos modelos incluyen variables latentes; ver Duncan (1.966), Werts y Linn (1.970), y Hauser y Goldberger (1.971).

Para empezar consideramos los modelos de análisis de trayectoria para las variables observadas directamente. Calcular un modelo de análisis de trayectoria para las variables observadas directamente con LISREL 8 es sencillo. Antes que calcular cada ecuación por separado, LISREL 8 considera el modelo como un sistema de ecuaciones y calcula todos los coeficientes estructurales directamente. Se obtiene la forma reducida como un producto secundario.

La diferencia fundamental entre este tipo de modelo y los modelos de regresión, es que las variables dependientes aparecen también al lado derecho de las relaciones. El siguiente ejemplo es un *sistema recurrente* en el sentido de que se pueden ordenar las variables dependientes en tal secuencia que cada variable dependiente depende sólo de las variables x y de las variables dependientes anteriores.

Tabla 1.3: Matriz de Covarianza para las Variables de Sentimiento Sindical

y_1	y_2	y_3	x_1	x_2
14.610				
-5.250	11.017			
-8.057	11.087	31.971		
-0.482	0.677	1.559	1.021	
-18.857	17.861	28.250	7.139	215.662

Ejemplo 3: Sentimiento Sindical de los Trabajadores de Textiles

McDonald y Clelland (1.984) analizaron los datos sobre el sentimiento sindical en los trabajadores de textiles no sindicalistas en el sur de los Estados Unidos. Después de la transformación de una variable y el tratamiento de las remotas, Bollen (1.989a) reanalizó un subgrupo de las variables según el modelo presentado en la Figura 1.3. Las variables son

- y_1 = *deferencia hacia los gerentes*
- y_2 = *apoyo del activismo laboral*
- y_3 = *sentimiento hacia los sindicatos*
- x_1 = *logaritmo de años en la fabrica textil*
- x_2 = *edad*

La matriz de covarianza se da en la Tabla 1.3.

En los dos ejemplos anteriores, hemos utilizado nombres de variables como LABOR, CAPITAL y TIME en el primer ejemplo y GRAVEREQ y IQPREVYR en el segundo. Aunque es una práctica recomendada, algunos usuarios podrían preferir usar nombres genéricos tales como Y1, Y2, Y3, X1 y X2 como en el tercer ejemplo.

El modelo en la Figura 1.3 se puede especificar facilmente como relaciones (ecuaciones):

Relationships

Y1 = X2

Y2 = X2 Y1

Y3 = X1 Y1 Y2

Estas tres líneas expresan las siguientes tres declaraciones, respectivamente

- Y1 depende de X2
- Y2 depende de X2 y Y1
- Y3 depende de X1, Y1 y Y2

También es sencillo formular el modelo especificando sus trayectorias:

Paths

X1 -> Y3

X2 -> Y1 Y2

Y1 -> Y2 Y3

Y2 -> Y3

Estas líneas expresan las siguientes declaraciones

- Hay una trayectoria (path) de X1 a Y3
- Hay dos trayectorias de X2, una a Y1 y otra a Y2
- Hay dos trayectorias de Y1, una a Y2 y otra a Y3
- Hay una trayectoria de Y2 a Y3

En total hay seis trayectorias en el modelo, correspondiente a las seis flechas de dirección única (uni-dirigidas) en la Figura 1.3.

El fichero de entrada para este problema es sencillo:

File EX3A.SPL

```
Title: Union Sentiment of Textile Workers

       Variables: Y1 = deference (submissiveness) to managers
                  Y2 = support for labor activism
                  Y3 = sentiment towards unions

                  X1 = Years in textile mill
                  X2 = age

Observed Variables: Y1 - Y3 X1 X2
Covariance Matrix:
   14.610
   -5.250  11.017
   -8.057  11.087  31.971
   -0.482   0.677   1.559   1.021
  -18.857  17.861  28.250   7.139 215.662
Sample Size 173
Relationships
   Y1 = X2
   Y2 = X2 Y1
   Y3 = X1 Y1 Y2
End of Problem
```

Las primeras 10 líneas son líneas titulares que definen el problema y las variables. Para indicar el principio de las líneas titulares, se puede usar la palabra Title pero no es necesario. La primera línea de comando real comienza con las palabras Observed Variables. Notar como Y1, Y2 y Y3 se pueden etiquetar conjuntamente como Y1-Y3.

Se definen las variables observadas, la matriz de covarianza y el tamaño de la muestra al igual que en el Ejemplo 1. El modelo se especifica como relaciones.

El modelo hace una distinción fundamental entre dos tipos de variable: dependiente e independiente. Se supone que el modelo explique las *variables dependientes*, es decir, se supone que la variación y covariación en las variables dependientes sean explicadas por las *variables independientes*. Las variables dependientes están al lado izquierdo del símbolo igual (=). Corresponden a aquellas variables en el diagrama de trayectoria señaladas con flechas de dirección única. Las variables que no son dependientes se llaman independientes.

La distinción entre las variables dependientes e independientes es ya inherente en las etiquetas: las variables X son independientes, y las variables Y son dependientes. Notar que las variables Y pueden aparecer a la derecha de las ecuaciones.

El fichero de salida da las siguientes relaciones estimadas

```
Y1 =  - 0.087*X2, Errorvar.= 12.96, R2 = 0.11
       (0.019)              (1.41)
       -4.65                9.22

Y2 =  - 0.28*Y1 + 0.058*X2, Errorvar.= 8.49 , R2 = 0.23
       (0.062)   (0.016)              (0.92)
       -4.58      3.59                9.22

Y3 =  - 0.22*Y1 + 0.85*Y2 + 0.86*X1, Errorvar.= 19.45, R2 = 0.39
       (0.098)   (0.11)    (0.34)              (2.11)
       -2.23      7.53     2.52                9.22
```

3. Modelos de Medición

Generalmente, hay dos importantes problemas básicos en las ciencias sociales y de comportamiento. El primer problema concierne a las propiedades de medición - valideces y fiabilidades - de los instrumentos de medición. El segundo problema concierne a las relaciones causales entre las variables y su relativa potencia explicativa.

La mayoría de las teorías y los modelos en las ciencias sociales y de comportamiento se formulan en términos de conceptos o constructos teóricos o hipotéticos, o variables latentes, que no son observables ni medibles directamente. Sin embargo, a menudo un número de indicadores o síntomas de estas variables se pueden usar para representar las variables latentes, más o menos bien. El propósito de un modelo de medición es describir hasta qué punto los indicadores observados sirven como instrumento de medición para las variables latentes. Los conceptos claves aquí son medición, fiabilidad, y validez. Los modelos de medición a menudo sugieren formas en que se puedan mejorar las mediciones observadas. Los modelos de medición son importantes en las ciencias sociales y de comportamiento, al intentar medir tales abstracciones como los comportamientos, actitudes, sentimientos y motivaciones de las personas. La mayoría de las mediciones empleadas para tal fin contiene considerables errores de medición y los modelos de medición nos permiten tener en cuenta estos errores.

Tabla 1.4: Correlaciones entre Mediciones de Habilidad y Aspiración

	x_1	x_2	x_3	x_4	x_5	x_6
S-C ABIL	1.00					
PPAREVAL	0.73	1.00				
PTEAEVAL	0.70	0.68	1.00			
PFRIEVAL	0.58	0.61	0.57	1.00		
EDUC ASP	0.46	0.43	0.40	0.37	1.00	
COL PLAN	0.56	0.52	0.48	0.41	0.72	1.00

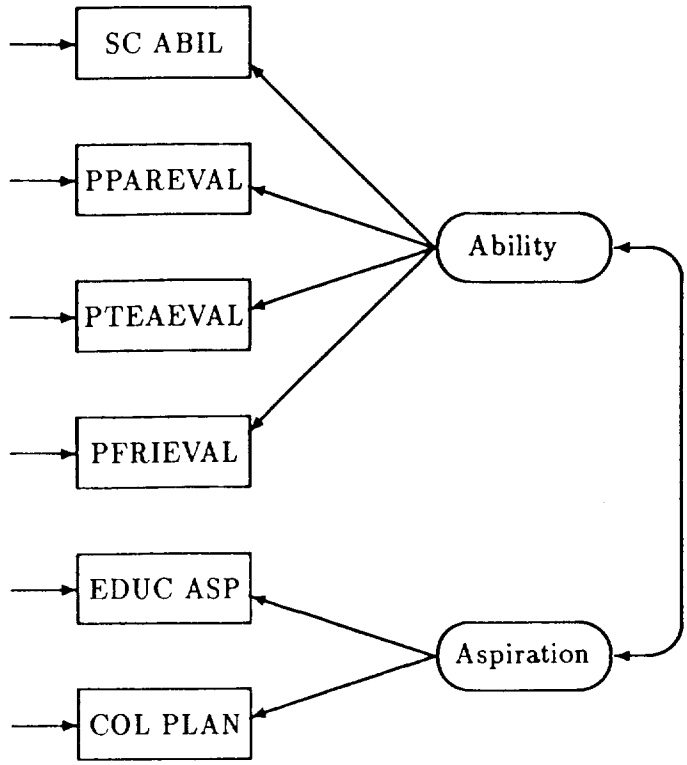

Figura 1.4: Diagrama de Trayectoria para Habilidad y Aspiración

Ejemplo 4: Habilidad y Aspiración

Calsyn y Kenny (1.977) presentaron la matriz de correlación en la Tabla 1.4 basada en 556 estudiantes blancos del octavo curso. Las mediciones son

- x_1 = *autoconcepto de habilidad (S-C ABIL)*
- x_2 = *evaluación parental percibida (PPAREVAL)*
- x_3 = *evaluación docente percibida (PTEAEVAL)*
- x_4 = *evaluación de amigos percibida (PFRIEVAL)*
- x_5 = *aspiración educacional (EDUC ASP)*
- x_6 = *planes del colegio (COL PLAN)*

Analizamos un modelo donde se asumen que x_1 , x_2 , x_3, y x_4 sean indicadores de "habilidad" y x_5 y x_6 sean indicadores de "aspiración" . Principalmente nos interesa calcular la correlación entre la habilidad real y la aspiración real. El diagrama de trayectoria se da en la Figura 1.4. Para distinguir las variables latentes de las variables observadas, en el diagrama las primeras están encerradas en cajas ovales y las otras en cajas rectangulares.

En este ejemplo se demuestra el uso de un fichero de datos externo. El fichero de datos contiene la matriz de correlación de las variables. El nombre del fichero es **EX4.COR.** El fichero es así:

```
1.00
0.73 1.00
0.70 0.68 1.00
0.58 0.61 0.57 1.00
0.46 0.43 0.40 0.37 1.00
0.56 0.52 0.48 0.41 0.72 1.00
```

El fichero de entrada ilustra, particularmente, cómo especificar el modelo por trayectorias en vez de por relaciones (fichero bf EX4A.SPL).

```
Ability and Aspiration
Observed Variables
'S-C ABIL' PPAREVAL PTEAEVAL PFRIEVAL 'EDUC ASP' 'COL PLAN'
Correlation Matrix From File: EX4.COR
Sample Size: 556
Latent Variables: Ability Aspiratn
Paths
Ability -> 'S-C ABIL' PPAREVAL PTEAEVAL PFRIEVAL
Aspiratn 'EDUC ASP' 'COL PLAN'
Print Residuals
End of Problem
```

Las líneas 2 y 3 definen las variables observadas. Línea 4 especifica que hay que leer la matriz de correlación del fichero EX4.COR y no del fichero de entrada. Línea 5 da el tamaño de la muestra. Este modelo incluye tanto las variables latentes (no observables) como las variables observadas, así que hay que distinguir entre estos dos tipos de variable en el fichero de entrada. Se definen a las variables latentes en la línea 6. En este libro nombramos a las variables latentes empleando una mayúscula para la primera letra para diferenciarlas de las etiquetas de variables observadas que se escriben totalmente con mayúsculas. El modelo se especifica en términos de trayectorias. La primera línea de trayectorias (línea 8) especifica que hay una trayectoria de Habilidad a cada una de las variables desde S-C ABIL hasta PFRIEVAL. Esto significa 4 trayectorias. La segunda línea de trayectorias (línea 9) especifica una trayectoria desde Aspiración hasta cada una de EDUC ASP y COL PLAN. Esto significa dos más trayectorias. La línea Print Residuals es una petición de imprimir ciertas matrices de residuales. Esto se explica más adelante.

El fichero de salida revela los siguientes resultados.

```
S-C ABIL =    0.86*Ability, Errorvar =  0.25   ,  R2 = 0.75
               (0.035)                 (0.023)
                24.56                   10.91
```

PPAREVAL = 0.85*Ability, Errorvar.= 0.28 , R2 = 0.72
(0.035) (0.024)
23.96 11.55

PTEAEVAL = 0.81*Ability, Errorvar.= 0.35 , R2 = 0.65
(0.036) (0.027)
22.11 13.07

PFRIEVAL = 0.7*Ability, Errorvar.= 0.52 , R2 = 0.48
(0.039) (0.035)
18 14.88

EDUC ASP = 0.78*Aspiratn, Errorvar.= 0.4 , R2 = 0.6
(0.04) (0.038)
19.21 10.45

COL PLAN = 0.93*Aspiratn, Errorvar.= 0.14 , R2 = 0.86
(0.039) (0.044)
23.57 3.15

CORRELATION MATRIX OF INDEPENDENT VARIABLES

	Ability	Aspiratn
Ability	1,00	
Aspiratn	.67	1.00
	(0.03)	
	21.53	

Las variables latentes se estandarizan por defecto, a menos que el usuario especifique otra unidad de medición. Dado que en este ejemplo también se estandarizan las variables observadas, significa que las cargas factoriales que aparecen delante de la variable latente en el ejemplo arriba, son cargas factoriales estandarizadas o coeficientes de validez estandarizados, ver Bollen (1.989a).

La correlación entre Habilidad y Aspiración se calcula como .67 con un error estándar de .03 y un valor *t* de 21.53. El valor alto de *t* indica que la correlación no es cero. En otro contexto podría ser más interesante comprobar si la correlación es 1. Se puede hacer esto al formar un intervalo de confianza aproximado para la correlación verdadera, empleando el error estándar. En este caso, el intervalo de confianza será (.604.728). Dado que este intervalo no incluye el valor 1, concluimos que la correlación es menor que 1.

No es una comprobación rigurosa, pero da un metodo rápido para comprobar si un modelo unidimensional, antes que uno de dos dimensiones, sería suficiente para explicar las intercorrelaciones entre las variables observadas.

La correlación entre Habilidad y Aspiración es una correlación desatenuante en el sentido que los efectos de los errores de medición en las variables observadas han sido eliminados. Las correlaciones desatenuantes de .67 entre Habilidad y Aspiración podrían compararse con las correlaciones atenuantes entre las mediciones observadas de habilidad y aspiración que tienen un rango entre .37 y .61. Por lo tanto la correlación desatenuante es mayor.

Para las relaciones de medición, hay un término de error en cada variable observada como indican las flechas de dirección única al lado izquierdo de el diagrama de trayectoria. Estos términos de error representan *errores en las variables* más que *errores en las ecuaciones* como los términos de error en los ejemplos anteriores. Los términos de error normalmente se interpretan como errores de medición (o errores de observación) en las variables observadas, aunque también pueden contener componentes sistemáticos específicos. Se da la varianza de error estimada junta con cada relación de medición. La correlación múltiple cuadrática R^2 también se da para cada ecuación. Es una medición de la fuerza de la relación lineal. En este contexto, normalmente R^2 se interpreta como la fiabilidad de la medición observada a la izquierda. Se ve que S C ABIL es el más fiable de los indicadores de Habilidad y que COL PLAN es el indicador más fiable de Aspiración. Dado que el modelo se ajusta bien a los datos, también podemos interpretar las cargas delante de las variables latentes como coeficientes de validez e interpretar S C ABIL como el indicador más válido de Habilidad y COL PLAN como el indicador más válido de Aspiración.

El ajuste del modelo es bastante bueno como demuestra el cuadrado chi de 9.26 con 8 grados de libertad. El ajuste se puede examinar aún más al inspeccionar las secciones del fichero de salida con la etiqueta **FITTED COVARIANCE MATRIX, FITTED RESIDUALS** y **STANDARDIZED RESIDUALS**. Tienen la siguiente apariencia:

FITTED COVARIANCE MATRIX

	S-C ABIL	PPAREVAL	PTEAEVAL	PFRIEVAL	EDUC ASP	COL PLAN
S-C ABIL	1.00					
PPAREVAL	.73	1.00				
PTEAEVAL	.69	.68	1.00			
PFRIEVAL	.60	.59	.56	1.00		
EDUC ASP	.45	.44	.42	.36	1.00	
COL PLAN	.53	.53	.50	.43	.72	1.00

FITTED RESIDUALS

	S-C ABIL	PPAREVAL	PTEAEVAL	PFRIEVAL	EDUC ASP	COL PLAN
S-C ABIL	.00					
PPAREVAL	.00	.00				
PTEAEVAL	.01	.00	.00			
PFRIEVAL	-.02	.02	.01	.00		
EDUC ASP	.01	-.01	-.02	.01	.00	
COL PLAN	.03	-.01	-.02	-.02	.00	.00

STANDARDIZED RESIDUALS

	S-C ABIL	PPAREVAL	PTEAEVAL	PFRIEVAL	EDUC ASP	COL PLAN
S-C ABIL	.00					
PPAREVAL	-.61	.00				
PTEAEVAL	.73	-.50	.00			
PFRIEVAL	-1.91	1.69	.72	.00		
EDUC ASP	1.01	-.57	-.87	.46	.00	
COL PLAN	2.09	-.43	-1.13	-.95	.00	.00

Si la línea

Print Residuals

no se incluye en el fichero de entrada, la información sobre los residuales sólo se da de forma resumida como

```
SUMMARY STATISTICS FOR FITTED RESIDUALS
SMALLEST FITTED RESIDUAL =    -.02
  MEDIAN FITTED RESIDUAL =     .00
 LARGEST FITTED RESIDUAL =     .03



STEMLEAF PLOT
- 2|00
- 1|86
- 0|96430000000
  0|5
  1|0149
  2|6

SUMMARY STATISTICS FOR STANDARDIZED RESIDUALS
SMALLEST STANDARDIZED RESIDUAL =   -1.91
  MEDIAN STANDARDIZED RESIDUAL =     .00
 LARGEST STANDARDIZED RESIDUAL =    2.09

STEMLEAF PLOT
- 1|91
- 0|9966540000000
  0|577
  1|07
  2|1
```

4. Análisis Factorial Confirmatorio

Es importante distinguir entre un análisis explorador y confirmatorio. En un análisis explorador, uno quiere explorar los datos empíricos para descubrir y detectar características y relaciones interesantes sin imponer ningún modelo definitivo en los datos. Un análisis explorador puede ser generador de estructura, generador de modelo y generador de hipótesis. Por otra parte, en un análisis confirmatorio uno construye un modelo que se asume describa, explique o de razón de los datos empíricos en términos de relativamente pocos parámetros. El modelo se basa en información *a priori* sobre la estructura de los datos en forma de una teoría o hipótesis específica, un diseño clasificatorio dado para artículos o subtests según las características objetivas de contenido y formato, condiciones experimentales conocidas, o conocimientos de estudios anteriores basados en datos extensos.

El análisis factorial explorador es una técnica que se emplea a menudo para detectar y asesorar las fuentes latentes de variación y covariación en las mediciones observadas. Es ampliamente reconocido que el análisis factorial explorador es bastante útil en las primeras etapas de experimentación o desarrollo de tests. Las habilidades mentales primarias de Thurston (1.938), factores de los tests de aptitud y logro de French (1.951) y la estructura de inteligencia de Guilford (1.956) son buenos ejemplos. Los resultados de un análisis explorador pueden tener un valor heurístico y sugestivo y pueden generar hipótesis que son capaces de una comprobación más objetiva por otros métodos multivariados. Sin embargo, mientras se ganan más conocimientos sobre la naturaleza de las mediciones sociales y psicológicas, puede que no valga el análisis factorial explorador como una herramienta e incluso puede llegar a ser un estorbo.

La mayoría de los estudios son en alguna medida tanto exploradores como confirmatorios dado que incluyen algunas variables tanto de composición conocida como no conocida.

Las primeras se deben elegir con mucho cuidado para poder extraer tanta información como sea posible sobre las últimas. Es altamente deseable que una hipótesis sugerida principalmente por procedimientos exploradores sea posteriormente confirmada, o rechazada por la obtención de datos nuevos sujetos a una técnica estadística más rigurosa. Aunque LISREL es más útil en los estudios confirmatorios, también se puede emplear para hacer un análisis explorador por medio de una secuencia de análisis confirmatorios. Sin embargo, debemos subrayar que hay que tener al menos una teoría o hipótesis experimental para poder empezar. La idea básica del análisis factorial es la siguiente. Para un grupo dado de variables de respuesta $x_1,...,x_q$ queremos encontrar un grupo de factores latentes subyacentes $\varepsilon_1,...\varepsilon_n$. menor en número que las variables observadas. Se supone que estos factores latentes expliquen las intercorrelaciones de las variables de respuesta en el sentido de que cuando los factores estén apartados de las variables observadas, ya no debería quedar ninguna correlación entre las mismas. Lleva al modelo, ver Jöreskog (1.979a),

$$x_i = \lambda_{i1}\xi_1 + \lambda_{i2}\xi_2 + \cdots + \lambda_{in}\xi_n + \delta_i \,, \qquad (1.3)$$

donde δ_i la parte única de x_i.se asume no correlacionado con $\varepsilon_1.\varepsilon_2.....\varepsilon_n$ y con δ_j para $j \neq i$. La parte única de δ_i consiste de dos componentes: un factor especifico S_i y un error de medición puramente aleatorio ε_i. Son indistinguibles el uno del otro a menos que las mediciones x_i. se diseñen de tal forma que se puedan ser identificados por separado (diseños de panel y diseños multitrato, multimétodo). El término δ_i menudo se llama *error de medición* en x_i. aunque está ampliamente reconocido que este término también puede contener un factor específico como se explica arriba. Continuaremos con esta tradición y usaremos este término en este libro.

En un análisis factorial confirmatorio, el investigador tiene suficiente conocimiento sobre la naturaleza factorial de las variables para ser capaz de especificar que cada medición x_i. depende sólo de unos pocos de los factores ε_j Si x_i. no depende de ε_j . $\lambda_{ij} = 0$ en (1.3). En un diagrama de trayectoria, esto significa que no hay una flecha de dirección única deε_j .a x_i. . En muchas aplicaciones, el factor latente ε_j representa una construcción teórica y las mediciones observadas x_i. son diseñadas para ser indicadores de esta construcción. En este caso sólo hay un λ_{ij} no de cero en cada ecuación (1.3). Era el caso en el ejemplo anterior.

Vamos a ilustrar un análisis factorial confirmatorio por medio de un ejemplo detallado. Particularmente, este ejemplo ilustra el asesoramiento del ajuste del modelo y el uso del índice de modificación de modelo.

Ejemplo 5: Nueve Variables Psicológicas - Un Análisis Factorial Confirmatorio

Holzinger y Swineford (1.939) recogían datos sobre veintiseis tests psicológicos administrados a 145 niños del séptimo y octavo curso en el colegio Grant-White de Chicago. Se seleccionaban nueve de los tests y para este ejemplo se aplicó la hipótesis que éstos miden tres factores comunes: percepción visual (P), habilidad verbal (V) y velocidad (S) , de tal forma que las primeras tres variables miden P, las tres siguientes miden V, y las últimas tres miden S. Las nueve variables seleccionadas y sus intercorrelaciones se dan en la Tabla 1.5. Se presenta una digrama de trayectoria en la Figura 1.5.

Queremos examinar el ajuste del modelo supuesto por la hipótesis declarada. Si el modelo no se ajusta bien a los datos, queremos sugerir un modelo alternativo que se ajuste mejor a los datos.

Tabla 1.5: Matriz de Correlación para Nueve Variables Psicológicas

VIS PERC	1.000								
CUBES	0.318	1.000							
LOZENGES	0.436	0.419	1.000						
PAR COMP	0.335	0.234	0.323	1.000					
SEN COMP	0.304	0.157	0.283	0.722	1.000				
WORDMEAN	0.326	0.195	0.350	0.714	0.685	1.000			
ADDITION	0.116	0.057	0.056	0.203	0.246	0.170	1.000		
COUNTDOT	0.314	0.145	0.229	0.095	0.181	0.113	0.585	1.000	
S-C CAPS	0.489	0.239	0.361	0.309	0.345	0.280	0.408	0.512	1.000

El modelo es muy parecido al modelo del ejemplo anterior. Las únicas diferencias son que hay 9 variables observadas en vez de 6 y que hay 3 factores en vez de 2. Sin embargo, como veremos, el ejemplo mostrará cómo se puede evaluar y modificar el modelo inicial cuando no se ajuste suficientemente bien a los datos.

El fichero de entrada SIMPLIS es casi igual que el fichero de entrada para el ejemplo 4, ver fichero **EX5A.SPL**.

```
Nine Psychological Variables - A Confirmatoru Factor Analysis
Observed Variables
'VIS  PERC' CUBES LOZENGES 'PAR COMP' 'SEN COMP' WORDMEAN
ADDITION COUNTDOT 'S - C CAPS'
Correlation Matrix From File EX5.COR
Sample Size 145
Latent Variables: Visual Verbal Speed
```

La primera línea especifica que queremos tener los resultados en el fichero de salida con tres puntos decimales. Dado que la mayoría de los usuarios sólo pueden interpretar hasta dos puntos decimales, LISREL 8 emplea dos puntos decimales por defecto. Este ejemplo ilustra como se puede cambiar este valor por defecto. La segunda línea se usa para especificar un fichero de salida con una anchura de 132 caracteres por línea. Si no, habrá 80 caracteres por línea.

Mirar al fichero de salida obtenido para este modelo. Aparte de las relaciones estimadas y la matriz factorial de correlación, que parecen razonables, se dan muchas estadísticas de buen ajuste. Por el momento sólo usamos el cuadrado chi en la primera línea:

CHI-SQUARE WITH 24 DEGREES OF FREEDOM = 52.626 (P = .000648)

Un cuadrado chi de 52.63 con 24 grados de libertad indica que el modelo no se ajusta bien a los datos. ¿Cómo se debería modificar el modelo para que se ajuste mejor a los datos? Una herramienta muy potente para contestar esta pregunta es el índice de modificación. Hay un índice de modificación para cada parámetro fijo en el modelo, es decir, para cada trayectoria que está omitida en el diagrama de trayectoria. Para cada trayectoria omitida, el índice de modificación es una estimación o predicción de la reducción en el cuadrado chi que se obtendrá si se introduce esa trayectoria en particular en el modelo. LISREL 8 lista todos los índices de modificación altos como sigue:

THE MODIFICATION INDICES SUGGEST TO ADD THE

PATH TO	FROM	DECREASE IN CHI-SQUARE	NEW ESTIMATE
ADDITION	Visual	10.5	-.37
COUNTDOT	Verbal	10.1	-.28
S-C CAPS	Visual	24.7	.57
S-C CAPS	Verbal	10.0	.26

El índice de modificación más alto es 24.74 para la trayectoria de Visual a S-C CAPS. Indica que podemos esperar una gran reducción en cuadrado chi si incluimos esta trayectoria en el modelo. Por lo tanto, si podemos interpretar esta trayectoria sustantivamente, ver abajo, podemos modificar el modelo al añadir esta trayectoria y ejecutar el modelo modificado. También se predice que la nueva trayectoria será 0.57.

El hecho de que el modelo está mal especificado se puede ver también en los residuales estandarizados en el fichero de salida:

Relationships:
'VIS PERC' - LOZENGES = Visual
'PAR COMP' - WORDMEAN = Verbal
ADDITION - 'S - C CAPS' = Speed
Number of Decimals = 3
Wirde Print
Print Residuals
End of Problem

Esta vez especificamos el modelo de medición como relaciones antes que trayectorias, tal y como hicimos en el ejemplo anterior. Los dos nuevos elementos en este fichero de entrada son

Number of Decimals = 3
Wide Print

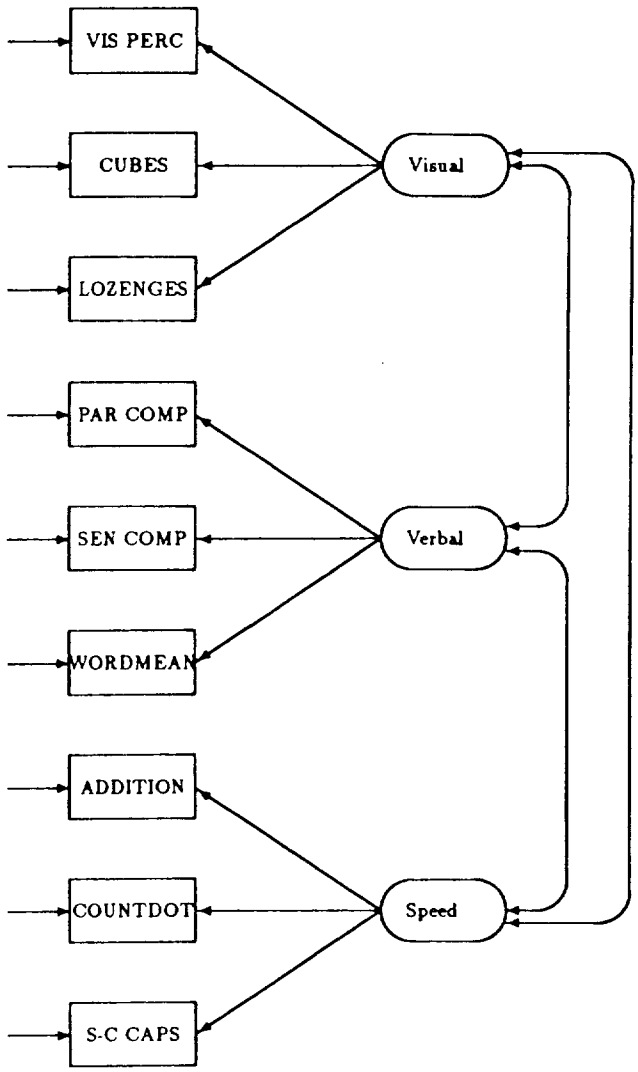

Figura 1.5: Modelo de Análisis Factorial Confirmatorio para Nueve Variables Psicológicas

STANDARDIZED RESIDUALS

	VIS PERC	CUBES	LOZENGES	PAR COMP	SEN COMP	WORDMEAN	ADDITION	COUNTDOT	S-C CAPS
VIS PERC	.000								
CUBES	-.793	.000							
LOZENGES	-1.353	2.087	.000						
PAR COMP	.438	-.139	.020	.000					
SEN COMP	.020	-1.294	-.565	.416	.000				
WORDMEAN	.529	-.609	.950	-.356	-.054	.000			
ADDITION	-1.946	-1.762	-3.117	.375	1.225	-.056	.000		
COUNTDOT	.884	-1.102	-1.141	-3.013	-.655	-2.084	5.007	.000	
S-C CAPS	4.650	.958	2.294	2.241	2.904	1.776	-2.007	-2.886	.000

Muestra un residual estandarizado alto de 4.65 entre VIS PERC y S-C CAPS, indicando que estas dos variables correlacionan más que explica el modelo. Aunque muestra donde se encuentra la falta de ajuste, no nos dice cómo se debería modificar el modelo para ajustarse mejor a los datos. Desde este punto de vista, los índices de modificación a menudo son más utiles que los residuales estandarizados para detectar los errores de especificación en el modelo.

Para ejecutar el modelo modificado, cambiar la línea, ver fichero **EX5B.SPL**.

'VIS PERC' - LOZENGES = Visual

en el fichero de entrada a

'VIS PERC' - LOZENGES 'S-C CAPS' = Visual

El cuadrado chi para el modelo modificado es

CHI-SQUARE WITH 23 DEGREES OF FREEDOM = 28.862 (P = .185)

Esto indica que el ajuste del modelo modificado es aceptable. Notar que la reducción en cuadrado chi es 52.63 - 28.86 = 23.77 que es aproximadamente igual que la predicción del índice de modificación. Notar también que la carga estimada de S-C CAPS sobre Visual es .46 que es un poco más baja que en la predicción. El valor *t* es 5.16. Así que esta carga es significante. Por lo tanto, S-C CAPS no es una medición pura de SPEED, si no una medición compuesta de Visual y de Speed.

La interpretación sustantiva de los resultados de estos análisis puede ser la siguiente. El primer factor es "percepción visual" representado por las primeras tres variables que contienen problemas espaciales con configuraciones geométricas. El tercer factor es un factor de "velocidad" (speed) que se supone mida la habilidad de realizar tareas muy sencillas rapidamente y con exactitud. Sin embargo, a diferencia que las dos mediciones "Adición" y "Contando puntitos" que son puramente númericas, la novena variable, "Mayúsculas rectas y curvadas" requiere la habilidad de

distinguir entre las mayúsculas que contienen partes curvadas (como la P) y las que tienen sólo líneas recta (como L) y hacerlo rapidamente y con exactitud. Por lo tanto, es concebible que "Mayúsculas rectas y curvadas" contiene un componente que se correlaciona con "percepción visual" tal y como se representa en estos datos, y también contiene un componente de "velocidad". Así que la variable "Mayúsculas rectas y curvadas" es una medición compuesta, a diferencia de todas las demás mediciones en este ejemplo, que son mediciones puras.

5. Análisis de Trayectoria con Variables Latentes

Discutimos el análisis de trayectoria con variables observadas directamente en la Sección 1.2. También es posible considerar un análisis de trayectoria para las variables latentes. En su forma más generalizada hay un sistema de ecuación estructural para un grupo de variables latentes clasificadas como dependientes o independientes. En la mayoría de las aplicaciones, el sistema es recursivo pero también se han propuesto modelos con sistemas no recursivos. Los sistemas recursivos se consideran en los Ejemplos 6 y 7 y un sistema no recursivo para variables latentes se considera en el Ejemplo 8.

Tabla 1.6: Matriz de Covarianza para la Estabilidad de Enajenación

	y_1	y_2	y_3	y_4	x_1	x_2
ANOMIA67	11.834					
POWERL67	6.947	9.364				
ANOMIA71	6.819	5.091	12.532			
POWERL71	4.783	5.028	7.495	9.986		
EDUC	-3.839	-3.889	-3.841	-3.625	9.610	
SEI*	-2.190	-1.883	-2.175	-1.878	3.552	4.503

* Se ha reducido a escala la variable SEI por un factor 10.

5.1 Sistema Recursivo

Los modelos recursivos son particularmente útiles en el análisis de datos de estudios longitudinales sobre psicología, educación y sociología. Ha habido varios artículos en la literatura sociológica sobre la especificación de los modelos, incorporando errores de causación y de medición, y análisis de datos de estudios de panel; ver Bohrnstedt (1.969), Heise (1.969, 1.970) y Duncan

(1.969, 1.972). Jöreskog y Sörbom (1.976, 1.977, 1.985), Jöreskog (1.979b), Jagodzinski y Kühnel (1.988), entre otros, discuten los modelos estadísticos y los métodos de análisis de datos longitudinales.

La característica de un diseño de investigación longitudinal es que las mismas herramientas de medición se usan en las mismas personas en dos o más ocasiones. El propósito de un estudio de panel o estudio longitudinal es asesorar los cambios que ocurren entre una ocasión y otra y atribuir estos cambios a ciertas características y sucesos históricos que existían o ocurrían antes de la primera ocasión y/o a varios tratamientos y desarrollos que ocurrían después de la primera ocasión. A menudo, cuando se usan las mismas variables repetidamente, existe una tendencia de que los errores de medición en estas variables se correlacionen sobre el período a causa de factores específicos u otros efectos de repetición de los tests. Por lo tanto, es preciso considerar a los modelos con errores de medición correlacionados.

Ejemplo 6: Estabilidad de Enajenación

Wheaton et al. (1.977) informa de un estudio de la estabilidad, sobre un período, de actitudes como la enajenación y la relación con las variables históricas como educación y ocupación. Se recogían datos sobre escalas de actitud de 932 personas en dos regiones rurales de Illinois en tres ocasiones: 1.966, 1.967 y 1.971. Las variables que se usan en este ejemplo son la subescala Anomia y la subescala Powerlessness (Impotencia), que se toman como indicadores de enajenación. El ejemplo solamente emplea datos de 1.967 y 1.971. Las variables históricas son la educación del encuestado (años de enseñanza completados) y el Indice Socioeconómico de Duncan (SEI). Se toman como indicadores del nivel socioeconómico del encuestado (Ses). La matriz de covarianza de la muestra de las 6 variables observadas se da en Tabla 1.6.

El modelo considerado aquí se presenta en la Figura 1.6. Especificamos los términos de

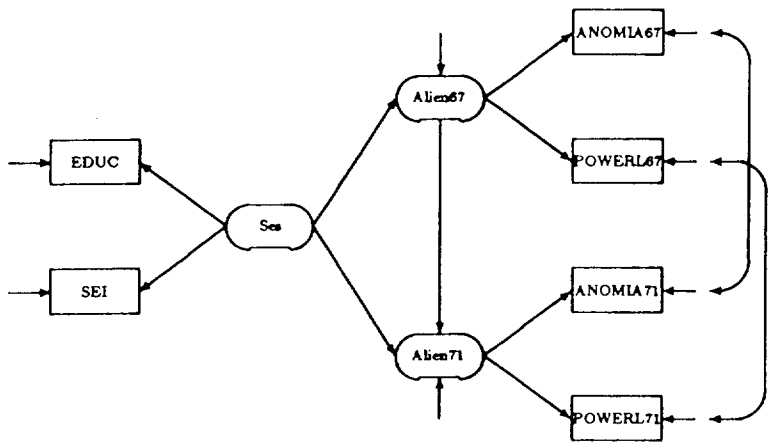

Figura 1.6: Modelo para la Estabilidad de Enajenación

error de ANOMIA y POWERL para que sean correlacionados sobre el período para tener en cuenta cualquier factor específico. La covarianza entre los dos términos de error para cada variable se puede interpretar como una varianza de err0or específico. Para otros modelos con los mismos datos, ver Jöreskog y Sörbom (1.989, pp. 170-171).

La configuración de este modelo para SIMPLIS es muy sencillo como demuestra el siguiente fichero de entrada, ver fichero **EX6A.SPL**.

```
Stability of Alienation
Observed Variables
   ANOMIA67   POWERL67   ANOMIA71   POWERL71   EDUC  SEI
Covariance Matrix
   11.834
    6.947      9.364
    6.819      5.091     12.532
    4.783      5.028      7.495      9.986
   -3.839     -3.889     -3.841     -3.625      9.610
   -2.190     -1.883     -2.175     -1.878      3.552    4.503
Sample Size 932
Latent Variables  Alien67 Alien71 Ses

Relationships
   ANOMIA67 POWERL67 = Alien67
   ANOMIA71 POWERL71 = Alien71
   EDUC SEI = Ses

   Alien67 = Ses
   Alien71 = Alien67 Ses

Let the Errors of ANOMIA67 and ANOMIA71 Correlate
Let the Errors of POWERL67 and POWERL71 Correlate

End of Problem
```

El modelo se especifica en términos de relaciones. Las primeras tres líneas especifican las relaciones entre las variables latentes y observadas, Las dos últimas líneas especifican las relaciones estructurales. Por ejemplo;

ANOMIA71 POWERL71 = Alien71

significa que las variables observadas ANOMIA71 y POWERL71 dependen de la variable latente Alien71, es decir, que ANOMIA71 y POWERL71 son indicadores de Alien71. La línea

Alien71 = Alien67 Ses

significa que la variable latente Alien71 depende de las dos variables latentes Alien67 y Ses. Es una de las dos relaciones estructurales.

Se puede especificar el modelo en términos de trayectorias en vez de relaciones:

```
Paths
        Alien67 -> ANOMIA67 POWERL67
        Alien71 -> ANOMIA71 POWERL71
        Ses -> EDUC SEI

        Alien67 -> Alien71
        Ses -> Alien67 Alien71
```

El fichero de salida revela que el modelo se ajusta muy bien. El cuadrado chi es 4.73 con 4 grados de libertad. Las dos ecuaciones estructurales se calculan como

```
Alien67 =  - 0.56*Ses, Errorvar.= 0.68, R2 = 0.32
            (0.051)
            -11.11

Alien71 = 0.57*Alien67 - 0.21*Ses, Errorvar.= 0.5, R2 = 0.5
          (0.055)        (0.046)
           10.27         -4.51
```

La matriz de covarianza estimada de todas las variables latentes también aparece en el fichero de salida:

```
        COVARIANCE MATRIX OF LATENT VARIABLES
            Alien67    Alien71        Ses
Alien67       1.00
Alien71        .68       1.00
    Ses       -.56       -.53       1.00
```

Dado que las variables latentes se estandarizan en esta solución, es una matriz de correlación. Las covarianzas de error aparecen en el fichero de salida como

```
Error Covariance for ANOMIA71 and ANOMIA67 = 1.62
                                            (0.31)
                                             5.17

Error Covariance for POWERL71 and POWERL67 = 0.34
                                            (0.26)
                                              1.3
```

La solución que se acaba de presentar es en términos de variables latentes estandarizadas. LISREL 8 estandariza automáticamente todas las variables latentes, a menos que se especifiquen otras unidades de medición. En este ejemplo, donde se analiza una matriz de covarianza, y las unidades de medición son las mismas en las dos ocasiones, sería más significativo asignar unidades de medición a las variables latentes con relación a las variables observadas. Haría dos trayectorias de Ses directamente comparables.

Por esta razón, se deberían especificar las relaciones como (ver fichero **EX6B.SPL**):

```
ANOMIA67 = 1*Alien67
POWERL67 = Alien67
ANOMIA71 = 1*Alien71
POWERL71 = Alien71
EDUC = 1*Ses
SEI = Ses

Alien67 = Ses
Alien71 = Alien67 Ses
```

1* en la primera relación de medición especifica un coeficiente fijo de 1 en la relación ANOMIA 67 y Alien67. El efecto es de fijar la unidad de medición en Alien67 con relación a la unidad en la variable observada ANOMIA67. Similarmente, en la tercera relación, la unidad de medición en Alien71 es fija con relación a la unidad en la variable observada ANOMIA71. Dado que ANOMIA67 y ANOMIA71 se miden por las mismas unidades, se ponen Alien67 y Alien71 en la misma escala. La quinta relación especifica que Ses tiene la misma escala que EDUC.

Cuando se calcula el modelo con el nuevo grupo de escalas, los resultados son:

```
Alien67 =  - 0.58*Ses, Errorvar.= 4.85 , R2 = 0.32
            (0.056)              (0.47)
            -10.19                10.35

Alien71 = 0.61*Alien67 - 0.23*Ses, Errorvar.= 4.09, R2 = 0.5
          (0.051)        (0.052)              (0.4)
           11.89         -4.33                10.1

        COVARIANCE MATRIX OF LATENT VARIABLE
            Alien67    Alien71        Ses
Alien67       7.10
Alien71       5.20       8.13
    Ses      -3.91      -3.92        6.80
```

Deberíamos subrayar que las dos soluciones presentadas aquí sólo representan al mismo modelo con las variables latentes en unidades diferentes. Las dos soluciones son equivalentes en el sentido de ajuste a los datos. También notar que los valores *t* para aquellos parámetros que se calculan en ambas soluciones son los mismos.

El efecto de Ses sobre Alienation (Enajenación) es negativo y mayor en 1.967 que en 1.971, tal y como se debería esperar. La autocovarianza en la medición de POWERLESSNESS (IMPOTENCIA) no es significante. Así que, mientras la varianza específica en la medición de ANOMIA es bastante alta, no hay evidencia de una varianza específica en la medición de POWERLESSNESS.

Ejemplo 7: Rendimiento y Satisfacción

Bagozzi (1.980c) formuló un modelo de ecuación estructural para estudiar la relación entre rendimiento y satisfacción en un cuerpo de ventas industriales. Se diseño el modelo especificamente para contestar preguntas como "¿Es mito o realidad el enlace entre el rendimiento y la satisfacción en el trabajo ?, ¿Influye el rendimiento en la satisfacción, o la satisfacción influye en el rendimiento?" (Bagozzi, 1.980, p. 65). El modelo que consideramos aquí es una modificación del modelo de Bagozzi, presentado en Figura 1.7. El modelo modificado se basa en un reanálisis de los datos de Bagozzi por parte de Jöreskog y Sörbom (1.982), ver también Jöreskog y Sörbom (1.989, pp 151-156).

Las variables observadas son:

- medición de rendimiento (PERFORM)
- medición de satisfacción laboral 1 (JBSATIS1)
- medición de satisfacción laboral 2 (JBSATIS2)
- medición de motivación de éxito 1 (ACHMOT1)
- medición de motivación de éxito 2 (ACHMOT2)
- medición de autoestimación específica a la tarea 1 (T-S S-E1)
- medición de autoestimación específica a la tarea 2 (T-S S-E2)
- medición de inteligencia verbal (VERBINTM)

y las variables latentes son

- rendimiento (Perform)
- satisfacción laboral (Jobsatis)

- motivación de éxito (Achmot)
- autoestimación especifica a la tarea (T-s s-e)
- inteligencia verbal (Verbint)

En el artículo de Bagozzi se da información detallada sobre las variables observadas. La Tabla 1.7 proporciona las medias, desviaciones estándar, y la correlación del momento de producto de las variables observadas, basadas en una muestra de $N = 122$.

Tabla 1.7: Medias, Desviaciones Estándar, y Correlaciones para las Variables Observadas en el Modelo de Bagozzi

variable	Correlations							
PERFORMM	1.000							
JBSATIS1	.418	1.000						
JBSATIS2	.394	.627	1.000					
ACHMOT1	.129	.202	.266	1.000				
ACHMOT2	.189	.284	.208	.365	1.000			
T-S S-E1	.544	.281	.324	.201	.161	1.000		
T-S S-E2	.507	.225	.314	.172	.174	.546	1.000	
VERBINTM	-.357	-.156	-.038	-.199	-.277	-.294	-.174	1.000
Means	720.86	15.54	18.46	14.90	14.35	19.57	24.16	21.36
St. Dev.	2.09	3.43	2.81	1.95	2.06	2.16	2.06	3.65

El modelo en la Figura 1.7 tiene la misma forma que el modelo de Figura 1.6. Difiere sólo en un aspecto fundamental, es decir, sólo hay un indicador de cada una de las variables latentes Verbint y Perform. Como consecuencia, no podemos estimar el error de medición en los correspondientes indicadores observados VERBINTM y PERFORMM. Las varianzas de la medición de error en estas variables no son parámetros identificados. La forma más sencilla para tratar a este problema es suponer que VERBINTM sea igual a Verbint y que PERFORMM sea igual a Perform, o lo que es lo mismo, decir que las varianzas de error de VERBINTM y PERFORMM son cero.

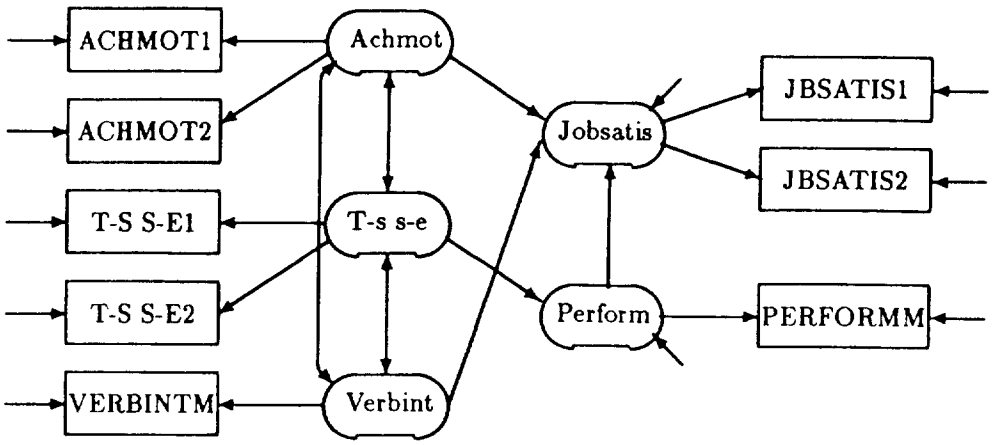

Figura 1.7: Modelo Modificado de Rendimiento y Satisfacción

Primero, ejecutamos el ejemplo, suponiendo que no sepamos los problemas. El fichero de entrada es

File EX7A.SPL

```
Modified Model for Performance and Satisfaction
Observed Variables: PERFORMM JBSATIS1 JBSATIS2 ACHMOT1 ACHMOT2
                    'T-S S-E1' 'T-S S-E2' VERBINTM
Correlation Matrix from File EX7.DAT
Standard Deviations from File EX7.DAT
Sample Size = 122
Latent Variables: Perform Jobsatis Achmot 'T-s s-e' Verbint
Relationships:
   PERFORMM = 1*Perform
   JBSATIS1 = 1*Jobsatis
   JBSATIS2 = Jobsatis
   ACHMOT1  = 1*Achmot
   ACHMOT2  = Achmot
   'T-S S-E1' = 1*'T-s s-e'
   'T-S S-E2' = 'T-s s-e'
   VERBINTM = 1*Verbint

   Perform = 'T-s s-e'
   Jobsatis = Perform Achmot Verbint

End of Problem
```

La primera línea es el título. Las siguientes 6 líneas definen las variables observadas y latentes y leen los datos y el tamaño de la muestra. Las primeras 8 líneas de las relaciones definen la ecuación de medición para cada una de las variables observadas. 1* son coeficientes fijos que definen la unidad de medición de las variables latentes. Las dos líneas siguientes definen las dos relaciones estructurales.

El fichero de datos **EX7.DAT** contiene la matriz de correlación y las desviaciones estándar, así

```
1.000
 .418 1.000
 .394  .627 1.000
 .129  .202  .266 1.000
 .189  .284  .208  .365 1.000
 .544  .281  .324  .201  .161 1.000
 .507  .225  .314  .172  .174  .546 1.000
-.357 -.156 -.038 -.199 -.277 -.294 -.174 1.000
2.09 3.43 2.81 1.95 2.06 2.16 2.06 3.65
```

El fichero de salida finaliza con el mensaje

F-A-T-A-L E-R-R-O-R: THE ERROR VARIANCE FOR PERFORMM MAY NOT BE IDENTIFIED.

El problema se resuelve al añadir la línea

Set the Error Variance of PERFORMM to O

Si sólo se añade esta línea, el nuevo fichero de salida finaliza con el mensaje

F-A-T-A-L E-R-R-O-R: THE ERROR VARIANCE FOR VERBINTM MAY NOT BE IDENTIFIED.

que tiene sentido porque la varianza de error de VERBINTM tampoco está identificada. LISREL 8 para en el primer parámetro no identificado que encuentra.

El test de inteligencia verbal VERBINTM es una medición falible y por lo tanto, no sería razonable suponer que su varianza de error sea cero. En vez de esto, suponemos que la fiabilidad de VERBINTM sea 0.85. Se discute que un valor arbitrario de 0.85 sea una mejor suposición que el valor 1.00, igualmente arbitrario. El supuesto valor de la fiabilidad afectará tanto a las estimaciones de parámetros como a los errores estándar. Una fiabilidad de 0.85 para VERBINTM es equivalente a una varianza de error 0.15 veces la varianza de VERBINTM. Así que suponemos que el error de varianza de VERBINTM sea $0.15 \times 3.65^2 = 1.998$. En el fichero de entrada se especifica como, ver fichero **EX7B.SPL**

Set the Error Variance of VERBINTM to 1.998

Las dos ecuaciones estructurales se calculan como

```
Perform = 0.92*T-s s-e, Errorvar.= 2.04
          (0.14)                  (0.4)
           6.4                    5.15

Jobsatis = 0.59*Perform + 1.23*Achmot + 0.21*Verbint, Errorvar.= 3.87
           (0.14)         (0.48)        (0.11)                  (1.22)
            4.24           2.57          2                       3.16
```

Tabla 1.8: Estimaciones de Parámetros con y sin Error de Medición en VERBINTM

Parameter	Without measurement error	With measurement error
T-s s-e ⟶ Perform	0.92 (0.14)	0.92 (0.14)
Perform ⟶ Jobsatis	0.58 (0.15)	0.60 (0.14)
Achmot ⟶ Jobsatis	1.18 (0.46)	1.23 (0.48)
Verbint ⟶ Jobsatis	0.18 (0.10)	0.21 (0.11)

Nuestro análisis confirma los resultados de Bagozzi con una modificación. El modelo original de Bagozzi incluía la hipótesis que la inteligencia verbal debería tener un efecto positivo sobre el rendimiento, pero encontraba que este efecto era negativo y insignificante. En nuestro análisis la inteligencia verbal tiene un efecto ligeramente positivo sobre satisfacción. La diferencia en los resultados se debe a nuestra inclusión de un error de medición de 15% en la medición de inteligencia verbal. El efecto de este error de medición en general se puede ver en la Tabla 1.8, donde comparamos los resultados obtenidos con y sin el error de medición en VERBINTM. Los efectos del error de medición en VERBINTM son más bien pequeños.

5.2 El Sistema No Recursivo

El siguiente ejemplo parece ser el primer modelo no recursivo con variables latentes en la literatura. Además de su no recursividad, incluye una característica que no hemos visto en el modelo anterior: efectos directos de las variables observadas sobre las variables latentes. Anteriormente, todas las trayectorias de las variables latentes han ido a las variables observadas, o a otras variables latentes. En este ejemplo, tenemos "causas" observadas y indicadores observados de las variables latentes.

Ejemplo 8: Influencias de los Iguales sobre Ambición

A menudo los sociólogos han llamado la atención a la forma en que los iguales de una persona , por ejemplo, amigos íntimos, influyen en las decisiones de la persona, por ejemplo, elección de ocupación. Han reconocido

que la relación tiene que ser recíproca - si mi mejor amigo influye en mi elección, debo influir en la suya. Duncan, Haller y Portes (1.968) presentan un modelo de ecuación simultánea de las influencias de los iguales en la elección de ocupación, empleando una muestra de estudiantes de escuela secundaria de Michigan, emparejados con sus mejores amigos. Los autores interpretan la elección educacional y ocupacional como dos indicadores de una variable latente "ambición", y especifican las elecciones. Este modelo con simultaneidad y errores de medición se presenta en la Figura 1.8. Notar que las variables en esta figura son simétricas respecto a una línea horizontal que pasa por el medio.

Tabla 1.9: Correlaciones y Desviaciones Estándar para Mediciones de Familia y Aspiración para 329 encuestados y sus Mejores Amigos

Respondent										
REINTGCE	1.0000									
REPARASP	.1839	1.0000								
RESOCIEC	.2220	.0489	1.0000							
REOCCASP	.4105	.2137	.3240	1.0000						
RE EDASP	.4043	.2742	.4047	.6247	1.0000					
Best Friend										
BFINTGCE	.3355	.0782	.2302	.2995	.2863	1.0000				
BFPARASP	.1021	.1147	.0931	.0760	.0702	.2087	1.0000			
BFSOCIEC	.1861	.0186	.2707	.2930	.2407	.2950	-.0438	1.0000		
BFOCCASP	.2598	.0839	.2786	.4216	.3275	.5007	.1988	.3607	1.0000	
BF EDASP	.2903	.1124	.3054	.3269	.3669	.5191	.2784	.4105	.6404	1.0000
St. Dev.	1.095	1.071	1.030	.978	1.001	1.036	1.021	1.002	1.055	1.022

En la Figura 1.8 hemos omitido todas las flechas de doble dirección entre las variables x, es decir, las 6 variables a la izquierda de la figura. Deja que

- x_1 = aspiración parental del encuestado (REPARASP)
- x_2 = inteligencia del encuestado (REINTGCE)
- x_3 = nivel socioeconómico del encuestado (RESOCIEC)
- x_4 = nivel socioeconómico del mejor amigo (BFSOCIEC)
- x_5 = inteligencia del mejor amigo (BFINTGCE)
- x6 = aspiración parental del mejor amigo (BFPARASP)
- y_1 = aspiración ocupacional del encuestado(REOCCASP)
- y_2 = aspiración educacional del encuestado (REEDASP)

- y_3 = aspiración educacional del mejor amigo(BFEDASP)
- y_4 = aspiración ocupacional del mejor amigo (BFOCCASP)

Las dos variables latentes son la ambición del encuestado (Reambitn) y la ambición de su mejor amigo (Bfambitn).

Los datos para este ejemplo se dan en la Tabla 1.9. Las desviaciones estándar no se dan en Duncan, Haller y Portes (1.968). Aquí se usan desviaciones estándar ficticias para enfatizar que hay que analizar una matriz de covarianza para obtener los errores estándar correctos de las estimaciones de parámetro, ver Cudeck (1.989).

Para empezar, preparamos tres ficheros **EX8.LAB, EX8.COR** y **EX8.STD**

EX8.LAB contiene los nombres de las variables en formato libre, como sigue

```
REINTGCE  REPARASP  RESOCIEC  REOCCASP  'RE EDASP'
BFINTGCE  BFPARASP  BFSOCIEC  BFOCCASP  'BF EDASP'
```

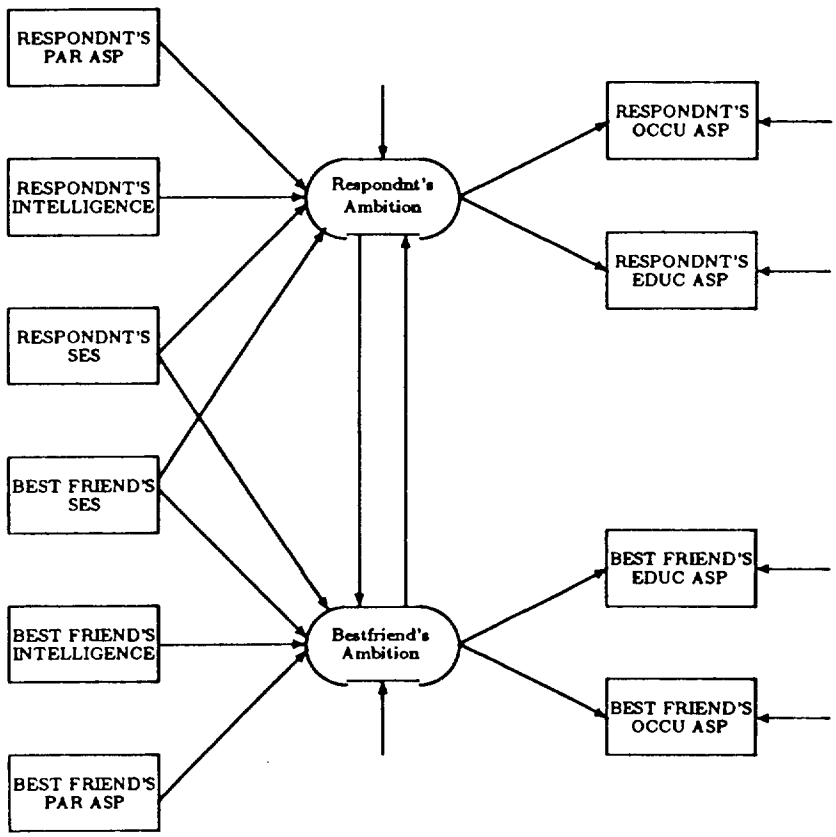

Figura 1.8: Diagrama de Trayectoria para Influencias de los Iguales sobre Ambición

Notar que, aunque se usan nombres largos en el diagrama de trayectoria, sólo se usan nombres de ocho caracteres en el fichero porque el programa sólo acepta etiquetas de ocho letras como máximo. También notar que el orden de las etiquetas en el fichero corresponde al orden de las variables en la matriz de correlación en la Tabla 1.7, no al orden de las variables en el diagrama de trayectoria.

El fichero **EX8.COR** contiene la matriz de correlación en formato libre.

```
1.0000
 .1839 1.0000
 .2220  .0489 1.0000
 .4105  .2137  .3240 1.0000
 .4043  .2742  .4047  .6247 1.0000
 .3355  .0782  .2302  .2995  .2863 1.0000
 .1021  .1147  .0931  .0760  .0702  .2087 1.0000
 .1861  .0186  .2707  .2930  .2407  .2950 -.0438 1.0000
 .2598  .0839  .2786  .4216  .3275  .5007  .1988  .3607 1.0000
 .2903  .1124  .3054  .3269  .3669  .5191  .2784  .4105  .6404 1.0000
```

El fichero **EX8.STD** contiene las desviaciones estándar de las ariables:

```
.095 1.071 1.030 .978 1.001 1.036 1.021 1.002 1.055 1.022
```

El fichero de entrada para el modelo en la Figura 1.8 es el siguiente

File EX8A.SPL

```
Peer Influences on Ambition
---------------------------
Observed Variables from File EX8.LAB
Correlation Matrix from File EX8.COR
Standard Deviations from File  EX8.STD
Reorder Variables: REPARASP REINTGCE RESOCIEC BFSOCIEC BFINTGCE
                   BFPARASP REOCCASP 'RE EDASP' 'BF EDASP' BFOCCASP
Sample Size 329
Latent Variables: Reambitn  Bfambitn
Relationships
   REOCCASP  = 1*Reambitn
   'RE EDASP'= Reambitn
   'BF EDASP'= Bfambitn
   BFOCCASP  = 1*Bfambitn
   Reambitn  = Bfambitn REPARASP - BFSOCIEC
   Bfambitn  = Reambitn RESOCIEC - BFPARASP

End of Problem
```

Es posible reordenar las variables observadas para que correspondan al orden en que aparecen en el diagrama de trayectoria. Se hace con las dos líneas

```
Reorder Variables: REPARASP REINTGCE RESOCIEC BFSOCIEC BFINTGCE
                   BFPARASP REOCCASP 'RE EDASP' 'BF EDASP' BFOCCASP
```

No es necesario, pero es cómodo porque hace que el fichero de salida sea más facilmente legible. También produce un diagrama de trayectoria que parece similar a la de la figura 1.8.

El modelo se especifica en la forma de relaciones. Las primeras cuatro relaciones especifican cómo las variables y, es decir, las cuatro variables a la derecha de el diagrama de trayectoria, dependen de las dos variables latentes. 1* en la primera y última de las cuatro relaciones significa que se deberían fijar las trayectorias a cero. Esto define que la unidad de medición de las dos variables latentes Rembitn y Bfambitn sea igual que la de REOCCASP y BFOCCASP, respectivamente. Como las variables observadas REOCCASP y BFOCCASP se miden en la misma unidad, implica que también las variables latentes Reambitn y Bfambitn tienen las mismas unidades, algo necesario para poder comparar los coeficientes de trayectoria entre el encuestado y su mejor amigo.

Las dos últimas relaciones especifican cómo las dos variables latentes Reambitn y Bfambitn dependen de las variables x.

En este modelo nos interesa comprobar la hipótesis que el efecto de Reambitn sobre Bfambitn es igual al efecto de Bfambitn sobre Reambitn. Esto quiere decir que debemos calcular estas dos trayectorias bajo la condición de su igualdad. LISREL maneja este tipo de problema bajo el título de *condiciones de igualdad.* Para este modelo se logra al especificar

Set Path from Reambitn to Bfambitn egual to Path from Bfambitn to Reambitn

ó

Let Path from Reambitn to Bfambitn = Path from Bfambitn to Reambitn

en el fichero de entrada. La declaración "trayectoria de A a B" también se expresa como "trayectoria a B de A". Hay varias formas alternativas para especificar lo mismo.Se puede usar casi cualquier forma libre con tal de que

- La línea empiece con Set o Let.
- Hay dos pares de variables mencionadas en la línea.

Quizás la forma más sencilla es

Set Reambitn -> Bfambitn = Bfambitn -> Reambitn

donde, al igual que antes, el símbolo -> se usa para denotar una trayectoria.

El modelo sin la condición de igualdad impuesta da un cuadrado chi de 26.89 con 16 grados de libertad. Cuando se impone la condición de igualdad, el cuadrado chi es 26.96 con 17 grados de libertad. La diferencia del cuadrado chi, 0.07, se puede usar como una comprobación de la estadística cuadrado chi con 1 grado de libertad para comprobar la hipótesis de que las trayectorias recíprocas entre las variables latentes sean iguales. Por lo tanto, existe evidencia que las dos trayectorias recíprocas sean iguales.

También nos podría interesar la hipótesis de simetría total entre el mejor amigo y el encuestado, es decir, que el modelo sea totalmente simétrico por encima y por debajo de la línea horizontal en medio del diagrama de trayectoria. Para comprobar esta hipótesis, hay que incluir todas las siguientes condiciones de igualdad en el fichero de entrada, ver fichero **EX8C.SPL**.

```
Set Reambitn -> Bfambitn = Bfambitn -> Reambitn
Set REPARASP -> Reambitn = BFPARASP -> Bfambitn
Set REINTGCE -> Reambitn = BFINTGCE -> Bfambitn
Set RESOCIEC -> Reambitn = BFSOCIEC -> Bfambitn
Set BFSOCIEC -> Reambitn = RESOCIEC -> Bfambitn
Set Reambitn -> 'RE EDASP' = Bfambitn -> 'BF EDASP'
```

El cuadrado chi generalizado para este modelo es 32.03 con 22 grados de libertad. Empleando el modelo inicial como modelo base, la diferencia en cuadrado chi es 5.14 con 6 grados de libertad, así que no se puede negar la hipótesis de simetría total.

Este análisis del Modelo 8C especifica simetría total al poner las condiciones de igualdad en las correspondientes trayectorias de el diagrama de trayectoria. También es razonable suponer que las varianzas de error sean iguales entre el encuestado y su mejor amigo para cada variable observada y latente.

Para especificar varianzas de error iguales para las dos variables latentes, incluir la línea

Set the Error Variances of Reambitn and Bfambitn Equal

Similarmente, configurar las varianzas de error iguales para cada una de las variables observadas por

Set the Error Variances of REOCCASP and BFOCCASP Equal
Set the Error Variances of 'RE EDASP' and 'BF EDASP'
(ver fichero EX8D.SPL.)

El cuadrado chi generalizado para el Modelo 8D es 33.44 con 25 grados de libertad. La diferencia del modelo anterior es 1.41 con 3 grados de libertad. Así que existe evidencia que también son iguales las varianzas de error.

6 Análisis de Variables Ordinales

En muchos casos, especialmente cuando los datos se recogen por medio de entrevistas o cuestionarios, las variables observadas son ordinales, es decir, las respuestas se clasifican en diferentes categorías ordenadas.

Una variable ordinal *z* (z puede ser una variable y ó x en el sentido de LISREL) se puede considerar como una medición bruta de una variable z^* subyacente, no observada o no observable. Por ejemplo, una escala ordinal de cuatro puntos se puede concebir como

- Si $z^* \leq \tau_1$, z tiene puntuación 1
- Si $\tau_1 < z^* \leq \tau_2, z$ tiene puntuación 2
- Si $\tau_2 < z^* \leq \tau_3, z$ tiene puntuación 3
- Si $\tau_3 < z^*, z$ tiene puntuación 4

donde $\tau_1 < \tau_2 < \tau_3$ son los valores mínimos para z^*. A menudo se supone que z^* tiene una distribución estándar normal, en cuyo caso se pueden estimar los mínimos a partir del inverso de la función normal de distribución.

Supone que z_1 y z_2 son dos variables ordinales con variables subyacentes continuas z_1^* y z_2^* respectivamente. Asumiendo que z_1^* y z_2^* tienen una distribución normal bivariada, se llama su correlación el *coeficiente de correlación policórico*. Un caso especial es el *coeficiente de correlación tetracórico* cuando z_1 y z_2 son dicotomosas. Vamos más lejos, suponiendo que z_3 es una variable continua, medida en una escala de intervalos. La correlación entre z_1^* y z_3 se llama el *coeficiente de correlación poliserial*, asumiendo que z_1^* y z_3 tienen una distribución normal bivariada. Un caso especial es la *correlación biserial* cuando z_1 es dicótoma.

Una variable ordinal *z* no tiene una escala métrica. Para usar tal variable en una relación lineal, empleamos en vez de ella, la correspondiente variable subyacente z*. Las correlaciones policóricas y poliseriales no son correlaciones computadas a partir de las puntuaciones reales, más bien son correlaciones teóricas de las variables subyacentes z*. Se calculan estas correlaciones de las tablas de contingencia observada por parejas de las variables ordinales. Ver Jöreskog y Sörbom (1.988, 1.993a) y refencias incluidas para la teoría en que se basan las correlaciones policóricas y poliseriales, y Jöreskog y Aish (1.993) para más ejemplos de análisis con variables ordinales.

Cuando las variables observadas en LISREL son todas ordinales, o son de tipo de escala mixta (ordinales y de intervalos), no se recomienda el uso de correlaciones de momento de producto ordinarias basadas en puntuaciones brutas. En vez de esto, se sugiere que se computen las estimaciones de las correlaciones policóricas y poliseriales y que la matriz de tales correlaciones sea analizada por el método WLS.

La matriz de ponderación que se requiere para este tipo de análisis es la inversa de la matriz **W** de covarianza asintótica estimada de las correlaciones policóricas y poliseriales. Tanto la matriz de covarianza asintótica como la matriz de correlaciones policóricas y poliseriales se obtienen con PRELIS.

Los pasos de este análisis se describirán en el siguiente ejemplo que incluye solamente variables ordinales y solamente correlaciones policóricas. Para otros ejemplos que incluyen también variables continuas y/o censuradas y otros tipos de correlaciones, ver PRELIS User's Guide (Jöreskog y Sörbom, 1.988).

Ejemplo 9: Modelo de Panel de Eficacia Política

Aish y Jöreskog (1.990) analizan datos sobre actitudes políticas. Los datos constan de 16 variables ordinales medidas en las mismas personas en dos ocasiones. Seis de las 16 variables se consideraban como indicadores de Eficacia Política y Sensibilidad hacia el Sistema. Las preguntas de actitud correspondientes a las seis variables son

- *Personas como yo no tenemos voz referente a lo que hace el gobierno (NOSAY (NODECIR))*
- *Votar es la única manera en que personas como yo podemos opinar sobre la manera de gobernar (VOTING (VOTAR))*
- *A veces la política y el gobierno parecen tan complicados que las personas como yo no podemos entender lo que pasa (COMPLEX (COMPLEJO))*

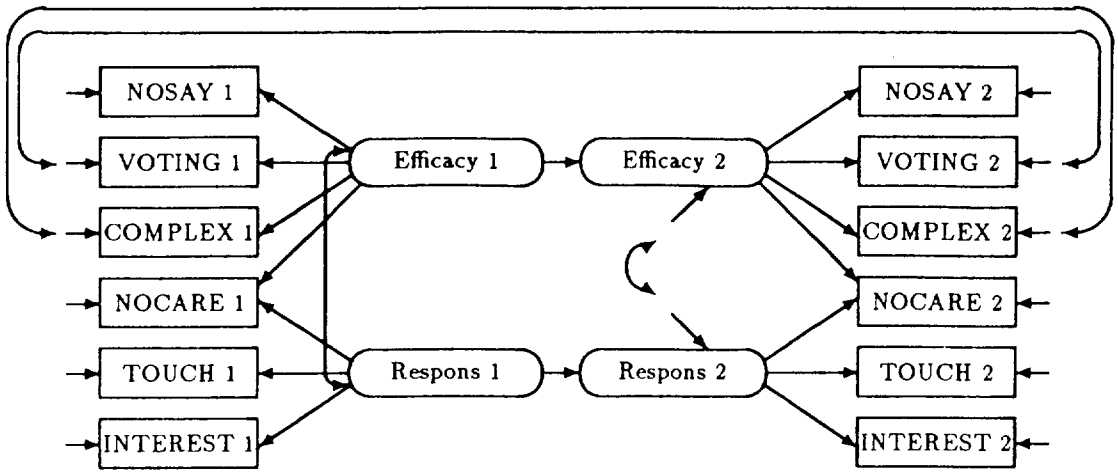

Figura 1.9: Modelo de Panel de Eficiacia Política

- *No creo que a los dirigentes públicos importe mucho lo que las personas como yo pensamos (NOCARE (NOIMPORTA))*

- *Generalmente, aquellas personas que elegimos para el Parlamento pierden el contacto rápidamente con las personas (TOUCH (CONTACTO))*

- *A los partidos sólo les interesan los votos de las personas, pero no sus opiniones (INTEREST (INTERES))*

Las respuestas permitidas a estas preguntas eran fuerte acuerdo, acuerdo, desacuerdo, fuerte desacuerdo, no sabe y no contesta.

Aish y Jöreskog (1.990) consideraban muchos modelos diferentes. Sólo uno de ellos se emplea aquí, presentado en la Figura 1.9. Incluye la idea de errores de medición correlacionados. Aish y Jöreskog (1.990) dividieron los datos originales en dos submuestras aleatorias, una muestra de exploración de tamaño 410 y una muestra de confirmación (validación) de tamaño 395. La modelación que lleva el modelo en la Figura 1.9 se presenta en Jöreskog y Aish (1.993).

Jöreskog y Sörbom (1.989) ilustran cómo emplear PRELIS para obtener la matriz de correlaciones policóricas para la muestra de exploración y cómo calcular la matriz de covarianza asintótica de las correlaciones policóricas del total de la muestra. Para nuestros propósitos suponemos que se haya almacenado la matriz de correlaciones policóricas en el fichero **PANELUSA.PME** y que la correspondiente matriz de covarianza asintótica se haya almacenado en el fichero **PANELUSA.ACP**. Los nombres de las variables están en el fichero **PANEL.LAB**.

Primero ejecutamos el modelo sin los errores de medición correlacionados. El fichero de entrada es

File EX9A.SPL

Panel Model for Political Efficacy
See Aish and Jöreskog, Quality and Quantity (1990)
Observed Variables from File PANEL.LAB
Correlation Matrix from file PANELUSA.PME
Asymptotic Covariance Matrix from file PANELUSA.ACP

Sample Size: 410
Latent Variables: Efficac1 Respons1 Efficac2 Respons2
Relationships:
NOSAY1 - NOCARE1 = Efficac1
NOCARE1 - INTERES1 = Respons1

NOSAY2 - NOCARE2 = Efficac2
NOCARE2 - INTERES2 = Respons2

Efficac2 = Efficac1
Respons2 = Respons1

Let the Errors for Efficac2 and Respons2 correlate

End of Problem

Aquí el elemento nuevo es la línea

Asymptotic Covariance Matrix from File PANELUSA.ACP

Se necesita el fichero **PANELUSA.ACP** producido por PRELIS para analizar las correlaciones policóricas. Esta línea instruye a LISREL 8 que lea la matriz de covarianza asintótica de este fichero y que emplee **WLS** (Weighted Least Squares(Cuadrados Mínimos Ponderados)) para calcular los parámetros del modelo. Si no se lee ninguna matriz de covarianza asintótica, LISREL 8 usará **ML** (maximum likelihood(máxima probabilidad)) por defecto para calcular los parámetros, al igual que en todos los ejemplos anteriores. Usar ML para ajustar el modelo a la matriz de correlaciones policóricas viola la teoría estadística en que se basan la estimación del cuadrado chi y los errores estándar.

Dado que las variables observadas son ordinales y no tienen unidades de medición, no es una buen idea emplearlas como variables de referencia para asignar unidades de medición a las variables latentes. La cosa más razonable sería suponer que todas las variables, tanto observadas como latentes, sean estandarizadas.

Las primeras dos líneas de relaciones definen el modelo de medición en la ocasión 1 y las dos líneas siguientes definen un modelo de medición similar en la ocasión 2. Las dos líneas siguientes definen las relaciones estructurales entre las variables latentes. Finalmente, la última línea de relaciones especifica que los dos términos de perturbación estructural deberían estar correlacionados como indica la pequeña flecha de dos direcciones en medio de el diagrama de trayectoria.

El fichero de salida revela que el modelo no se ajusta a los datos. El cuadrado chi es 115.89 con 48 grados de libertad. El valor *P* para esto es menor que 0.0000002. Si el modelo es "verdadero" y se sostienen todas las asumiciones del análisis, las posibilidades de obtener tal cuadrado chi son menos que 2 en 1000000.

Los índices de modificación son

THE MODIFICATION INDICES SUGGEST TO ADD AN ERROR COVARIANCE

BETWEEN	AND	DECREASE IN CHI-SQUARE	NEW ESTIMATE
TOUCH2	VOTING2	8.2	.12
NOSAY1	VOTING2	13.8	-.15
VOTING1	VOTING2	19.5	.24
VOTING1	NOSAY1	9.5	.16
COMPLEX1	COMPLEX2	45.1	.36

que sugiere que existan dos grandes índices de modificación, es decir, para la covarianza de los errores de medición en VOTING1 y VOTING2, COMPLEX1 y COMPLEX2. Sugiere que haya dos grandes factores específicos en VOTING y COMPLEX.

Por lo tanto, en la siguiente ejecución especificamos el modelo con dos grupos de términos de errores correlacionados como se presentan en la Figura 1.9. Se hace al añadir dos líneas, ver fichero EX9B.SPL.

Let the errors for VOTING1 and VOTING2 correlate
Let the errors for COMPLEX1 and COMPLEX2 correlate

en el fichero de entrada.

El cuadrado chi para este modelo es 50.62 con 46 grados de libertad que tiene un valor *P* =.296. Entonces este modelo se ajusta bien a los datos. Las autocovarianzas en VOTING y COMPLEX se dan en el fichero de salida como

Error Covariance for VOTING1 and VOTING2 = 0,24
(0.051)
4.66

Error Covariance for COMPLEX1 and COMPLEX2 = 0.36
(0.05)
7.04

Ambas autocovarianzas son altamente significantes. Entonces hay fuerte evidencia de que hay grandes factores específicos en VOTING y COMPLEX. Los resultados del Modelo 9B se resumen en las tablas 1.10-1.13. Para más análisis de estos datos e interpretación de los resultados, ver Jöreskog y Aish (1.993).

Tabla 9: Cargas y sus Errores Estándar

	Time 1		Time 2	
	Efficacy	Respons	Efficacy	Respons
NOSAY	.85 (.04)		.80 (.08)	
VOTING	.59 (.06)		.71 (.07)	
COMPLEX	.57 (.05)		.47 (.06)	
NOCARE	.37 (.11)	.50 (.11)	.42 (.12)	.49 (.12)
TOUCH		.80 (.04)		.84 (.05)
INTEREST		.93 (.03)		.89 (.06)

Tabla 10: Varianzas de Error, Fiabilidades y Autocovarianzas

	Error variances		Reliabilities		Autocovariances
	Time 1	Time 2	Time 1	Time 2	Specific Variances
NOSAY	.27	.37	.73	.63	
VOTING	.42	.25	.58	.74	.24(.05)
COMPLEX	.32	.42	.68	.58	.36(.05)
NOCARE	.32	.25	.68	.75	
TOUCH	.35	.29	.65	.71	
INTEREST	.13	.20	.87	.80	

Tabla 11: Correlaciones entre Eficacia y Sensibilidad

Time 1	Time 2
.78	.81

Tabla 12: Coeficientes de Estabilidad y Matriz de Covarianza Residual

	Stability Coefficients	Residual Covariance Matrix	
Efficacy	.72(.09)	.48	
Respons	.64(.07)	.45	.59

Este texto está basado en el libro
:LISREL 8:Structural Equation Modeling with the SIMPLIS Command Language

Publicado y Distribuido por

Scientific Software International
1525 East 53rd Street, Suite 906
Chicago, IL
U.S.A.
Tel: (312) 684-4920
Fax: (312) 684-4979

iec ProGamma
P.O. Box 841
9700 AV Groningen
The Netherlands
Tel: +31 50 636900
Fax: +31 50 636687

References

Akaike, H. (1974) A new look at statistical model identification. *IEEE transactions on Automatic Control*, 19, 716-723.

Akaike, H. (1987) Factor analysis and AIC. *Psychometrika*, 52, 317-332.

Aish, A.M., and Jöreskog, K.G. (1990) A panel model for political efficacy and responsiveness: An application of LISREL7 with weighted least squares. *Quality and Quantity, 24*, 405 - 426.

Alwin, D.F., and Jackson, D.J. (1981) Applications of simultaneous factor analysis to issues of factorial invariance. Pp 249-279 in D. Jackson and E. Borgatta (Eds.): *Factor analysis and measurement in sociological research: A Multi-Dimensional Perspective*. Beverly Hills: Sage.

Bagozzi, R.P. (1980c) Performance and satisfaction in an industrial sales force: An examination of their antecedents and simultaneity. *Journal of Marketing*, 44, 65-77.

Bentler, P.M. (1990) Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.

Bentler, P.M., and Bonett, D.G. (1980) Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.

Bentler, P.M., and Woodward, J.A. (1978) A Head Start reevaluation: Positive effects are not yet demonstrable. *Evaluation Quarterly*, 2, 493-510.

Blalock, H.M., Jr., (Ed.) (1985) *Causal models in the social sciences*. Second Edition. New York: Aldine Publishing Co.

Bohrnstedt, G.W. (1969) Observations on the measurement of change. In E.F. Borgatta (Ed.): *Sociological Methodology 1969*. San Francisco: Jossey-Bass, 113-133.

Bollen, K.A. (1986) Sample size and Bentler and Bonett's nonnormed fit index. *Psychometrika*, 51, 375-377.

Bollen, K.A. (1987) Total, direct, and indirect effects in structural equation models. In C. Clogg (Ed.): *Sociological Methodology 1987*. San Francisco: Jossey Bass.

Bollen, K.A. (1989a) *Structural equations with latent variables*. New York: Wiley.

Bollen, K.A. (1989b) A new incremental fit index for general structural equation models. *Sociological Methods and Research*, 17, 303-316.

Bollen, K.A., and Liang, J. (1988) Some properties of Hoelter's CN. *Sociological Methods and Research*, 16, 492-503.

Bozdogan, H. (1987) Model selection and Akaike's information criteria (AIC). *Psychometrika*, 52, 345-370.

Browne, M.W. (1984) Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.

Browne, M.W., and Cudeck, R. (1989) Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, 24, 445-455.

Browne, M. W. and Cudeck, R. (1992) Alternative ways of assessing model fit. In K. A. Bollen and J. S. Long (Editors): *Testing Structural Equation Models*, Sage Publications, in press.

Calsyn, J.R., and Kenny, D.A. (1977) Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement? *Journal of Educational Psychology*, 69, 136-145.

Chou, C-P., and Bentler, P.M. (1990) Model modification in covariance structure modeling: A comparison among likelihood ratio, Lagrange multiplier, and Wald tests. *Multivariate Behavioral Research*, 25, 115-136.

Costner, H.L. (1969) Theory, deduction, and rules of correspondence. *American Journal of Sociology*, 75, 245-263.

Cudeck, R. (1989) Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, 105, 317 - 327.

Cudeck, R., and Browne, M.W. (1983) Cross-validation of covariance structures. *Multivariate Behavioral Research*, 18, 147-157.

Draper, N.S., and Smith, H. (1967) *Applied regression analysis*. New York: Wiley

Duncan, O.D. (1966) Path analysis: Sociological examples. *American Journal of Sociology*, 72, 1-16.

Duncan, O.D. (1969) Some linear models for two-wave, two-variable panel analysis. *Psychological Bulletin*, 72, 177-182.

Duncan, O.D. (1972) Unmeasured variables in linear models for panel analysis. In H.L. Costner (Ed.): *Sociological Methodology 1972*. San Francisco: Jossey-Bass, 36-82.

Duncan, O.D. (1975) *Introduction to structural equation models*. New York: Academic Press.

Duncan, O.D., Haller, A.O., and Portes, A. (1968) Peer influence on aspiration: A reinterpretation. *American Journal of Sociology*, 74, 119-137.

Finn, J.D. (1974) *A general model for multivariate analysis*. New York: Holt, Reinhart and Winston.

French, J.V. (1951) The description of aptitude and achievement tests in terms of rotated factors. *Psychometric Monographs*, 5.

Goldberger, A.S. (1964) *Econometric theory*. New York: Wiley.

Guilford, J.P. (1956) The structure of intellect. *Psychological Bulletin*, 53, 267-293.

Hoelter, J.W. (1983) The analysis of covariance structures: Goodness-of- fit indices. *Sociological Methods and Research*, 11, 325-344.

Huitema, B.H. (1980) *The analysis of covariance and alternatives*. New York: Wiley.

Hauser, R.M., and Goldberger, A.S. (1971) The treatment of unobservable variables in path analysis. Pp. 81-117 in Costner, H.L.: *Sociological Methodology*.

Heise, D.R. (1969) Separating reliability and stability in test-retest correlation. *American Sociological Review*, 34, 93-101.

Heise, D.R. (1970) Causal inference from panel data. In E.F. Borgatta and G.W. Bohrnstedt (Eds.): *Sociological Methodology 1970*. San Francisco: Jossey-Bass,3-27.

Holzinger, K., and Swineford, F. (1939) *A study in factor analysis: The stability of a bi-factor solution*. Supplementary Educational Monograph no. 48. Chicago: University of Chicago Press.

Jagodzinski, W., and Kühnel, S.M. (1988) Estimation of reliability and stability in single-indicator multiple-wave models. *Sociological Methods and Research*, 15, 219-258.

James, L.R., Mulaik, S.A., and Brett, J.M. (1982) *Causal analysis: Assumptions, models, and data*. Beverly Hills: Sage.

Johnston, J. (1972) *Econometric methods*. New York: McGraw- Hill.

Jöreskog, K.G. (1971) Simultaneous factor analysis in several populations. *Psychometrika*, 57, 409-426.

Jöreskog, K.G. (1979a) Basic ideas of factor and component analysis. In K.G. Jöreskog and D. Sörbom: *Advances in factor analysis and structural equation models*. Cambridge, Mass.: Abt Books, 5-20.

Jöreskog, K.G. (1979b) Statistical estimation of structural models in longitudinal developmental investigations. In J.R. Nesselroade and P.B. Baltes (Eds.): *Longitudinal research in the study of behavior and development*. New York: Academic Press.

Jöreskog, K.G. (1992) Testing structural equation models. In K.A. Bollen and J.S. Long (Eds), *Testing Structural Equation Models*. Sage Publications, in press.

Jöreskog, K.G., and Aish, A.M. (1993) *Structural Equation Modeling with Ordinal Variables*. Book Manuscript in Preparation.

Jöreskog, K.G., and Sörbom, D. (1976) Statistical models and methods for test-retest situations. In D.N.M. deGruijter and L.J.Th. van der Kamp (Eds.): *Advances in psychological and educational measurement*. New York: Wiley, 285-325.

Jöreskog, K.G., and Sörbom, D. (1977) Statistical models and methods for analysis of longitudinal data. In D.J. Aigner and A.S. Goldberger (Eds.): *Latent variables in socio-economic models*. Amsterdam: North- Holland Publishing Co.,285-325.

Jöreskog, K.G., and Sörbom, D. (1985) Simultaneous analysis of longitudinal data from several cohorts. Pp. 323-341 in W.M. Mason and S.E. Fienberg (Eds.): *Cohort analysis in social research: Beyond the identification problem*. New York: Springer-Verlag.

Jöreskog, K.G., and Sörbom, D. (1982) Recent developments in structural equation modeling. *Journal of Marketing Research*, 19, 404 -416.

Jöreskog, K.G., and Sörbom, D. (1988) *PRELIS - A program for multivariate data screening and data summarization. A preprocessor for* LISREL. Second Edition. Chicago, Illinois: Scientific Software, Inc.

Jöreskog, K.G., and Sörbom, D. (1989) *LISREL 7 - A guide to the program and applications*. Second Edition. Chicago: SPSS Publications.

Jöreskog, K.G., and Sörbom, D. (1990) Model search with TETRAD II and LISREL. *Sociological Methods and Research*, 19, 93-106.

Jöreskog, K.G., and Sörbom, D. (1992a) New features in PRELIS 2. Chicago: Scientific Software.

Jöreskog, K.G., and Sörbom, D. (1992b) New features in LISREL 8. Chicago: Scientific Software.

Kaplan, D. (1989) Model modification in covariance structure analysis: Application of the expected parameter change statistic. *Multivariate Behavioral Research*, 24, 285-305.

Kenny, D.A., and Judd, C.M. (1984) Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, 96, 201-210.

Lee, S. and Hershberger, S. (1990) A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research*, 25, 313-334.

Lomax, R.G. (1983) A guide to multiple-sample structural equation modeling. *Behavior Research Methods and Instrumentation*, 15, 580- 584.

Magidson, J. (1977) Toward a causal model approach for adjusting for pre-existing differences in the non-equivalent control group situation. *Evaluation Quarterly*, 1, 399-420.

Maiti, S.S., and Mukherjee, B.N. (1990) A note on distributional properties of the Jöreskog-Sörbom fit indices. *Psychometrika*, 55, 721-726.

Mare, R.D., and Mason, W.M. (1981) Children's report of parental socioeconomic status. In G.W. Bohrnstedt and E.F. Borgatta (Eds.): *Social Measurement: Current Issues*. Beverly Hills: Sage Publications.

McDonald, R.P. (1989) An index of goodness of fit based on non-centrality. *Journal of Classification*, 6, 97-103.

McDonald, J.A., and Clelland, D.A. (1984) Textile workers and union sentiment. *Social Forces*, 63, 502-521.

McGaw, B., and Jöreskog, K.G. (1971) Factorial invariance of ability measures in groups differing in intelligence and socio-economic status. *British Journal of Mathematical and Statistical Psychology*, 24, 154-168.

Matsueda, R.L., and Bielby, W.T. (1986) Statistical power in covariance structure models. Pp. 120-158 in N.B. Tuma (Ed.): *Sociological Methodology 1986*. San Francisco: Jossey Bass.

Mulaik, S., James, L., Van Alstine, J., Bennett, N., Lind, S., and Stilwell, C. (1989) Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105, 430-445.

Muthén, B. (1984) A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115-132.

Olsson, U. (1979) Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44, 443-460.

Saris, W.E., Satorra, A. and Sörbom, D. (1987) The detection and correction of specification errors in structural equation models. In C. Clogg (Ed.): *Sociological Methodology 1987*. San Francisco: Jossey. Bass.

Satorra, A., and Saris, W.E. (1985) Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, 50, 83-90.

Silvey, S.D. (1970) *Statistical inference*. Middlesex, U.K.: Penguin Books.

Sobel, M. (1982) Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.): *Sociological Methodology 1982*. San Francisco: Jossey Bass.

Steiger, J.H. (1990) Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.

Stelzl, I. (1986) Changing causal relationships without changing the fit: Some rules for generating equivalent LISREL models. *Multivariate Behavioral Research*, 21, 309-331.

Sörbom, D. (1974) A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology*, 27, 229-239.

Sörbom, D. (1975) Detection of correlated errors in longitudinal data. *British Journal of Mathematical and Statistical Psychology*, 28, 138- 151.

Sörbom, D. (1976) A statistical model for the measurement of change in true scores. In D.N.M. de Gruijter and J.L.Th. van der Kamp (Eds.): *Advances in psychological and educational measurement*. New York: Wiley. 159-169.

Sörbom, D. (1978) An alternative to the methodology for analysis of covariance. *Psychometrika*, 43, 381-396.

Sörbom, D. (1981) Stuctural equation models with structured means. In K.G. Jöreskog and H. Wold (Eds.): *Systems under indirect observation: Causality, structure and prediction*. Amsterdam: North- Holland Publishing Co.

Sörbom, D. (1989) Model modification. *Psychometrika*, 54, 371-384.

Sörbom, D., and Jöreskog, K.G. (1981) The use of LISREL in sociological model building. In E. Borgatta and D.J. Jackson (Eds.) *Factor analysis and measurement in sociolgical research: A multidimensional perspective*. Beverly Hills: Sage.

Theil, H. (1971) *Principles of econometrics*. New York: Wiley.

Thurstone, L.L. (1938) Primary mental abilities. *Psychometric Monographs*, 1.

Tucker, L.R., and Lewis, C. (1973) A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.

Turner, M.E., and Stevens, C.D. (1959) The regression analysis of causal paths. *Biometrics*, 15, 236-258.

Werts, C.E., and Linn, R.L. (1970) Path analysis: Psychological examples. *Psychological Bulletin*, 67, 193-212.

Werts, C.E., Rock, D.A., Linn, R.L., and Jöreskog, K.G. (1976) A comparison of correlations, variances, covariances and regression weights with or without measurement errors. *Psychological Bulletin*, 83, 1007-1013.

Werts, C.E., Rock, D.A., Linn, R.L., and Jöreskog, K.G. (1977) Validating psychometric assumptions within and between populations. *Educational and Psychological Measurement*, 37, 863-871.

Wheaton, B., Muthén, B., Alwin, D., and Summers, G. (1977) Assessing reliability and stability in panel models. In D.R. Heise (Ed.): *Sociological Methodology 1977*. San Francisco: Jossey-Bass.

Wright, S. (1934) The method of path coefficients. *Annals of Mathematical Statistics*, 5, 161-215.

NAZIOARTEKO ESTATISTIKA MINTEGIA
SEMINARIO INTERNACIONAL DE ESTADISTICA

ORAIN ARTE ARGITARATUTAKO GAIAK
TEXTOS PUBLICADOS HASTA LA FECHA

N.º 1. Zbkia.: LINEAL STATISTICAL INFERENCE
(Inglés, Euskera, Español)
Autor/Egilea: C.R. RAO
Año/Urtea: 1983

N.º 2. Zbkia.: MUESTREO Y APLICACIONES
(Español, Euskera)
Autor/Egilea: E. CANSADO
Año/Urtea: 1983

N.º 3. Zbkia.: STATISTICAL EDUCATION
(Inglés, Euskera, Español)
Autor/Egilea: V. BARNETT
Año/Urtea: 1983

N.º 4. Zbkia.: ANALYSE DES DONNEES
(Francés, Euskera, Español)
Autor/Egilea: P. CLAPIER
Año/Urtea: 1983

N.º 5. Zbkia.: DESIGN OF EXPERIMENTS
(Inglés, Euskera, Español)
Autor/Egilea: D.J. FINNEY
Año/Urtea: 1984

N.º 6. Zbkia.: ASPECTOS DE TEORIA Y APLICACIONES EN EL MUESTREO
(Español, Euskera)
Autor/Egilea: F. AZORIN POCH
Año/Urtea: 1984

N.º 7. Zbkia.: CURSO BASICO INTENSIVO DE MUESTREO
(Español, Euskera)
Autor/Egilea: J.L. SANCHEZ-CRESPO
Año/Urtea: 1985

N.º 8. Zbkia.: ANALYSE DES SERIES CHRONOLOGIQUES: LES INDICES STATISTIQUES
(Francés, Euskera, Español)
Autor/Egilea: J. FOURASTIE
Año/Urtea: 1985

N.º 10. Zbkia.: METHODOLOGY AND TREATMENT FOR NON-RESPONSE
(Inglés, Euskera, Español)
Autor/Egilea: R. PLATEK
Año/Urtea: 1986

N.º 11. Zbkia.: STATISTICAL OPERATIONS BY SAMPLING
(Inglés, Euskera, Español)
Autor/Egilea: L. KISH
Año/Urtea: 1986

N.º 12. Zbkia.: ANALISIS DE SERIES TEMPORALES: ALGUNAS TECNICAS DE PREDICCION
(Español, Euskera)
Autor/Egilea: I. GALLASTEGI
Año/Urtea: 1986

N.º 13. Zbkia.: BASES DE DATOS
(Español, Euskera)
Autor/Egilea: F. SALTOR
Año/Urtea: 1987

N.º 14. Zbkia.: METODOS ESTADISTICOS PARA LA INVESTIGACION EPIDEMIOLOGICA
(Español)
Autor/Egilea: L.C. SILVA
Año/Urtea: 1987

N.º 15. Zbkia.: SAMPLING AND NON-SAMPLING ERRORS IN SURVEYS
(Inglés)
Autor/Egilea: A. MARTON
Año/Urtea: 1988

N.º 16. Zbkia.: LES ENQUÊTES TELEPHONIQUES
(Francés)
Autor/Egilea: V. SALVY
Año/Urtea: 1988

N.º 17. Zbkia.: GENERALIZED LINEAR MODELS IN EPIDEMIOLOGY
(Inglés)
Autor/Egilea: J.C. DUFFY
Año/Urtea: 1989

N.º 18. Zbkia.: NEW TECHNOLOGIES IN COMPUTER ASSITED SURVEY PROCESSING
(Inglés)
Autor/Egilea: J.G. BETHLEHEM and W.J. KELLER
Año/Urtea: 1989

N.º 19. Zbkia.: EVALUATION OF QUESTIONNAIRE DESIGN EFFECTS
(Inglés)
Autor/Egilea: G. NATHAN
Año/Urtea: 1990

N.º 20. Zbkia.: PROCEDIMIENTO DE DEPURACION DE DATOS ESTADISTICOS
(Español)
Autor/Egilea: I. VILLAN CRIADO / M.S. BRAVO CABRIA
Año/Urtea: 1990

N.º 21. Zbkia.: THE X11ARIMA / 88 SEASONAL ADJUSTMENT METHOD
(Inglés)
Autor/Egilea: E. BEE DAGUM
Año/Urtea: 1990

N.º 22. Zbkia.: RENTA Y DISTRIBUCION DE LA RIQUEZA, DESIGUALDAD Y POBREZA: TEORIA, MODELOS Y APLICACIONES
(Español)
Autor/Egilea: C. DAGUM
Año/Urtea: 1991

N.º 23. Zbkia.: LA CONTABILIDAD NACIONAL COMO MARCO DE LAS ESTIMACIONES DE VARIABLES ECONOMICAS
(Español)
Autor/Egilea: V. ANTON VALERO
Año/Urtea: 1991

N.º 24. Zbkia.: MACRO-EDITING. METHODS FOR RATIONALIZING THE EDITING OF QUANTITATIVE DATA.
(Inglés)
Autor/Egilea: L. GRANQUIST
Año/Urtea: 1991

N.º 25. Zbkia.: METHODOLOGICAL ISSUE IN FAMILY EXPENDITURE SURVEYS
(Inglés, Euskera, Español)
Autor/Egilea: M. KANTOROWITZ
Año/Urtea: 1992

N.º 26. Zbkia.: QUALITY CONTROL IN STATISTICS FROM ADMINISTRATIVE REGISTERS AND RECORDS
(Inglés, Euskera, Español)
Autor/Egilea: HANS PETTERSSON
Año/Urtea: 1992

N.º 27. Zbkia.: ANALYSE DES DONNEES ET CLASSIFICATION AUTOMATIQUE NUMERIQUE ET SYMBOLIQUE
(Francés)
Autor/Egilea: E. DIDAY
Año/Urtea: 1992

N.º 28. Zbkia.: METHODOLOGIE DE LA RECHERCHE EXPERIMENTALE
(Francés, Euskera, Español)
Autor/Egilea: ROGER PHAN TAN LUU
Año/Urtea: 1993

N.º 29. Zbkia.: STRUCTURAL EQUATION MODELING WITH LISREL
(Español)
Autor/Egilea: KARL G. JÖRESKOG
Año/Urtea: 1993