

01821010

(4898)
V/m



Fecha recibida: 10 NOV. 1977

ARCHIVO de DOCUMENTOS

Original NO SALE de la oficina

CENTRO LATINOAMERICANO DE DEMOGRAFIA
SECTOR DE FECUNDIDAD

EXPLICACION BREVE DEL ANALISIS DE TRAYECTORIA (PATH ANALYSIS)
CON UN EJEMPLO QUE EMPLEA UN PROGRAMA DISPONIBLE EN EL
TERMINAL APL DE CELADE

Johanna de Jong
Arthur M. Conning

S.151/11/74
(Junio de 1974)

I. ANÁLISIS DE TRAYECTORIA (PATH ANALYSIS)

1. Con el análisis de trayectoria se analizan los efectos directos e indirectos lineales de unas variables sobre otras dentro de un sistema cerrado, basado en la teoría del investigador. Es una forma de análisis de regresión parcial en el que se usan los puntajes estandarizados. La ventaja de esto radica en que los coeficientes de todas las relaciones en el sistema son directamente comparables.

La relación entre el coeficiente de regresión parcial (no estandarizado) y el coeficiente de trayectoria (estandarizado) es:

$$P_{ij} = \frac{s_j}{s_i} \cdot c_{ij}$$

i = variable dependiente

j = variable independiente (A)

s = desviación estándar

c = coeficiente de regresión parcial

P = coeficiente de trayectoria de la variable j a la variable i.

2. En el modelo se distinguen varios tipos de variables:

- a) VARIABLES EXÓGENAS: las variables que se supone son independientes de las demás variables dentro del modelo. Entre las variables exógenas pueden existir correlaciones, pero éstas quedan sin analizar. En la figura 1, Z_1 y Z_2 son variables exógenas.
- b) VARIABLES ENDÓGENAS: las que son dependientes de otras variables dentro del sistema. Se supone que toda la varianza de las variables endógenas está determinada por una combinación lineal de las variables en el sistema, más las variables residuales. Las variables Z_3 y Z_4 en la figura 1 son endógenas.
- c) VARIABLES RESIDUALES: las que no se miden sino que se introducen para explicar los residuos de la varianza de las variables endógenas no explicada por las demás variables dentro del sistema. El valor del coeficiente de trayectoria de la variable residuo r actuando sobre la variable i es por lo tanto $P_{ir} = \sqrt{1 - R^2}$ donde R es el coeficiente de correlación múltiple de las variables hasta e inclusive i.

en que: Z_1 y Z_2 : variables exógenas
 Z_3, Z_4, Z_5 : variables endógenas
 R_a, R_b, R_c : variables residuales
 $P_{51}, P_{52}, P_{53}, P_{54}, P_{41}, P_{42}, P_{43}, P_{32}, P_{31}, P_{3a}, P_{4b}, P_{5c}$: coeficientes de trayectoria
 r_{12} : coeficiente de correlación.

5. Entre las variables exógenas se mantienen los coeficientes de correlación porque se las considera como causantes y no influidas por ninguna de las variables endógenas. Las demás relaciones originales medidas con coeficientes de correlación, se supone cambian porque están recalculadas tomando en cuenta los demás caminos por los cuales la variable endógena está influida. Ellas se expresan en los coeficientes de trayectoria mencionados.
6. Se puede definir al coeficiente de trayectoria (Reichwein, 1971, citando a Wright, 1934) como el número que indica la fracción de la desviación estándar de una variable dependiente que está explicada en forma directa por la variación de la variable independiente. Esta fracción se encontraría si el factor variara en la misma medida de los datos observados, manteniéndose constantes las demás variables relevantes, incluso las variables residuales. El método normalmente usado para solucionar las ecuaciones del modelo da coeficientes de trayectoria que son iguales a los coeficientes beta en una regresión, es decir, a coeficientes de regresión parcial estandarizados.
7. Esta definición se ve reflejada en el teorema básico del análisis de trayectoria, que se puede expresar en la siguiente fórmula:

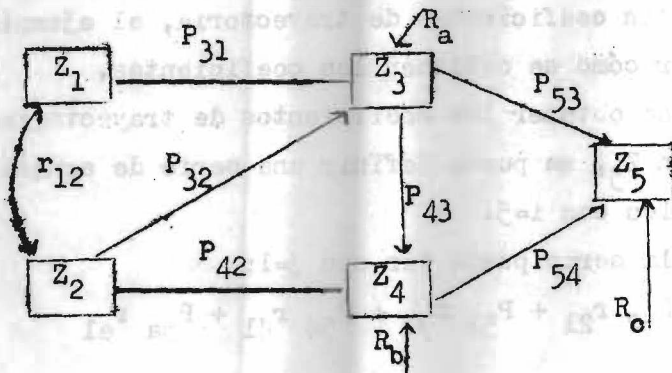
$$r_{ij} = \sum_{q=1}^n P_{iq} \cdot r_{jq} \quad \text{en que:}$$

II. UN EJEMPLO CALCULADO CON EL PROGRAMA DE CELADE

9. CELADE tiene un programa escrito en el lenguaje APL para su terminal de computación. Este programa requiere como entrada la matriz de correlación de orden cero entre todas las variables especificadas en el modelo, y usa el teorema básico para determinar una serie de ecuaciones requeridas por el coeficiente de trayectoria.
10. Para ilustrar el uso del programa consideramos el modelo de la Figura 2; (el ejemplo y el programa de computación están basados en gran parte en un artículo de Nygreen, 1971).

Figura 2

MODELO CAUSAL EN QUE SOLAMENTE SE CONSIDERAN LAS RELACIONES POSTULADAS



11. Este modelo es similar a la Figura 1, pero el investigador ha eliminado ciertas posibles trayectorias tales como la que va entre Z_1 y Z_4 porque cree que no existen razones teóricas para incluirlas. En efecto, esto implica que el coeficiente de trayectoria P_{41} (etc.), es considerado como igual a cero.
12. La matriz de correlación que se usa aquí, calculada tal vez usando la rutina de correlación del programa SPSS (también disponible en CELADE), se muestra en la Tabla 1.

De manera similar las otras ecuaciones se pueden definir para Z_5 dando una serie completa:

$$\begin{aligned} r_{51} &= P_{53} r_{31} + P_{54} r_{41} \\ r_{52} &= P_{53} r_{32} + P_{54} r_{42} \\ r_{53} &= P_{53} r_{33} + P_{54} r_{43} \\ r_{54} &= P_{53} r_{34} + P_{54} r_{44} \end{aligned} \quad \text{Por supuesto, } r_{33} \text{ y } r_{44} \text{ son iguales a 1.} \quad (E)$$

15. Esta serie de ecuaciones (E) debe ser usada para resolver los coeficientes de trayectoria P_{53} y P_{54} . Pero esto requiere solamente dos ecuaciones y nosotros tenemos cuatro. Esto se conoce como "sobreidentificación", ya que cada uno de los posibles seis pares de ecuaciones pueden proporcionar valores únicos de estos dos coeficientes de trayectoria y las soluciones individuales pueden no ser las mismas.
16. La solución al problema de sobreidentificación que da los coeficientes de trayectoria como definidos en el párrafo 1 (coeficientes de regresión estandarizados), está en usar las dos ecuaciones en las cuales los coeficientes de trayectoria tienen los mismos subíndices que los coeficientes de correlación. En el ejemplo:

$$\begin{aligned} r_{53} &= P_{53} r_{33} + P_{54} r_{43} \\ r_{54} &= P_{53} r_{34} + P_{54} r_{44} \end{aligned} \quad (F)$$

o con las correlaciones obtenidas de la matriz de correlación en Tabla 1:

$$\begin{aligned} 0,52 &= P_{53} + 0,56 (P_{54}) \\ 0,49 &= 0,56 (P_{53}) + P_{54} \end{aligned} \quad (G)$$

17. Estas dos ecuaciones se pueden solucionar fácilmente en forma algebraica para P_{53} y P_{54} pero naturalmente, en éste y en casos más complejos, el programa de computación dará los resultados de inmediato y calculado en forma más precisa. Los valores de estos dos coeficientes de correlación, como todos los demás en el modelo, se dan en la Tabla 2 que muestra la entrada y la salida del programa.

Tabla 2

ENTRADA Y SALIDA DEL PROGRAMA CON DATOS Y EJEMPLO DEL TEXTO

JCOPY CELAEJT IPA

M ← 5 5p.01 x100 12 41 42 29 12 100 35 15 31
 41 35 100 5b 52 42 15 56 100 49 29 31 52 49
 100

M

1	0.12	0.41	0.42	0.29
0.12	1	0.35	0.15	0.31
0.41	0.35	1	0.56	0.52
0.42	0.15	0.56	1	0.49
0.29	0.31	0.52	0.49	1

IPA M

INGRESE N° DE ORDEN DE:

VARIABLES EXOGENAS

□:

1 2

VARIABLES CAUSALES DE:

VARIABLE 3

□:

1 2

VARIABLE 4

□:

2 3

VARIABLE 5

□:

3 4

COEFICIENTES DE INTERACCION

(↓) CAUSAS/EFFECTOS (→)

0	3	4	5
1	0.373377	0	0
2	0.305195	- 0.0524217	0
3	0	0.578348	0.357809
4	0	0	0.289627
5	0	0	0
0	0.860289	0.827036	0.81977

BIBLIOGRAFIA

- Heise, David R.,
1969 "Problems in Path Analysis and Causal Inference", en Sociological Methodology, 1969, editado por E.F. Borgatta, Jossey-Bass (San Francisco), pp. 38-73.
- Land, Kenneth C.,
1969 "Principles of Path Analysis", en Sociological Methodology, 1969, editado por E.F. Borgatta, Jossey-Bass (San Francisco) pp. 3-37.
- Nygreen, G.I.,
1971 "Interactive Path Analysis", The American Sociologist, Vol. 6 (Feb.) pp. 37-43.
- Reichwein, Ch.L.M.,
1971 Pad analyse in de sociale wetenschappen. Utrecht, Holanda.
- Wright, Si,
1934 "The Method of Path Coefficient", Annals of Mathematical Statistics 5 (Sept.) 161-215.

