

Técnicas Multivariantes

Modelos de Ecuaciones Estructurales

Juan C. Correa

Material de uso exclusivo para
INGENIO PANTALEON, S.A.
Diagonal 6, 10-31, Zona 10

Ciudad de Guatemala

- 1 Breve Introducción
- 2 Operaciones Básicas
- 3 Ejercicio Práctico

Breve Introducción

Los Modelos de Ecuaciones Estructurales (Structural Equation Modeling, “SEM”) son una familia de “herramientas de análisis” que permiten probar relaciones “causales” entre variables a través de técnicas de estadística multivariable.

Ωnyx

A graphical interface for Structural Equation Modeling

Ωnyx is a free software environment for Structural Equation Modeling. It runs on a wide variety of platforms, including UNIX, Mac, and Windows. Ωnyx is available **free of charge**.

Download

<http://onyx.brandmaier.de/>

El concepto de “relación causal” en el contexto de los SEM asume que las relaciones entre las **variables Observadas** son explicadas por **Variables Latentes** que no están presentes en nuestros datos.

Variables Latentes: Variables no presentes en nuestra base de datos.

Variables Observadas: Variables que sí están presentes como columnas en nuestras bases de datos.

Estos modelos se desarrollaron principalmente en el seno de la psicología y sociología a finales del siglo XIX.

Un ejemplo de variable latente es el concepto de inteligencia, pues a diferencia de la altura o el peso, es difícil afirmar que la inteligencia puede “verse” o “medirse” de manera semejante a la medición de variables de magnitudes físicas.

Lo que se hace para “medir” la inteligencia es un procedimiento de muestreo de respuestas a preguntas específicas de un test psicológico estandarizado.

Más allá de la psicología, en ciencias económicas también se han usado los SEM para abordar fenómenos complejos.



RESEARCH ARTICLE

Are average years of education losing predictive power for economic growth? An alternative measure through structural equations modeling

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0213651>

Desde el punto de vista algebraico, los SEM no son más que un sistema de ecuaciones lineales con igual número de incógnitas y ecuaciones, por ejemplo:

$$3x + 2y - 3z = 8$$

$$2x + 4y - z = 17$$

$$\frac{x}{2} + \frac{y}{4} + 2z = 3,5$$

Este sistema presenta tres variables (i.e., x , y , z) y tres ecuaciones. El sistema requiere conocer los valores de x , y y z que satisfacen las relaciones especificadas por las tres ecuaciones.

Para analizar las relaciones entre varias variables a través de modelos de ecuaciones estructurales se deben realizar las siguientes cuatro operaciones básicas:

- Especificación
- Identificación
- Estimación de Parámetros
- Evaluación de Bondad de Ajuste





La especificación de un modelo implica:

- Definir las relaciones entre variables observadas y latentes.
- Definir cuáles son variables “endógenas” (aquellas cuya varianza es objeto de explicación dentro del modelo) y “exógenas” (variables que no son explicables por las variables del modelo).
- Definir el tipo de relación entre variables (relaciones de covariación versus relaciones de co-dependencia).

Operaciones Básicas: Especificación

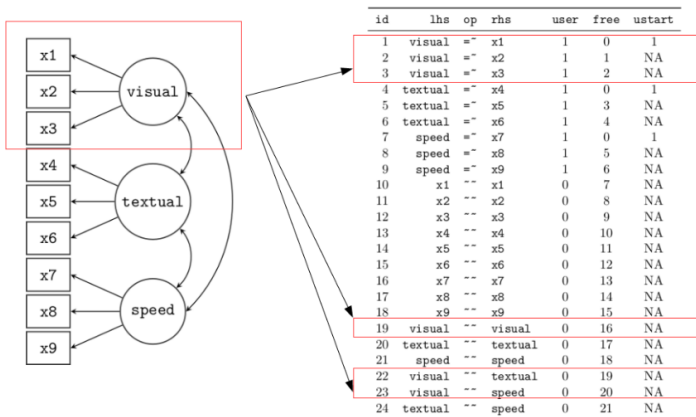
La especificación puede hacerse gráficamente (Ω nyx) o a través de sintaxis con los paquetes lavaan, rsem.

Convenciones de representación gráfica para la especificación de modelos.

Representación Gráfica		Significación
		Variable latente o no directamente observada (e.g., factores, errores).
		Variable observada o directamente medida.
		Ruta unidireccional que sugiere la influencia “causal” de una variable sobre otra, expresada en términos de pesos de regresión.
		Ruta bidireccional que sugiere la relación “no causal” entre variables, generalmente expresada en términos de covarianza o correlación.

Operaciones Básicas: Especificación

En lavaan o rsem las sintaxis siguen las convenciones gráficas, aunque estas se reflejan a través de operadores específicos tales como “==” o “=~”.



La identificación comprende la designación de los parámetros (estadísticos) que deben estimarse dentro del conjunto de relaciones especificadas en el modelo, siguiendo ciertas reglas para lograr que el sistema de ecuaciones sea determinable:

- Asignación de una métrica a los términos de error y variables latentes (i.e., $\sigma^2 = 1$).
- Escalamiento de las cargas factoriales del primer ítem de una variable latente (i.e., $\lambda^2 = 1$).

Consiste en efectuar los cálculos requeridos para conocer el valor de los parámetros involucrados en el sistema de ecuaciones del modelo. La siguiente expresión presenta la estructura algebraica de un análisis factorial confirmatorio.

$$\Sigma(\Theta) = \Lambda_X \Phi \Lambda_X' + \Theta_\delta$$

donde:

- Λ_X Matriz con la asociación entre variables observadas y latentes.
- Φ Matriz de varianza-covarianza entre variables latentes
- Θ_δ Matriz de varianza-covarianza de términos de error

Un modelo de análisis factorial confirmatorio de una variable latente con tres variables observadas, tendría la siguiente estructura.

$$\Lambda_X = \begin{bmatrix} \lambda_{1,1} \\ \lambda_{2,1} \\ \lambda_{3,1} \end{bmatrix}, \Phi = [\phi_{1,1}], \Lambda'_X = [\lambda_{1,1} \lambda_{1,2} \lambda_{1,3}] + \Theta_\delta = \begin{bmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_2 & 0 \\ 0 & 0 & \epsilon_3 \end{bmatrix}$$

La estimación de parámetros se reduce al cálculo inferencial de los valores poblacionales de estadísticos como la varianza de variables latentes, covarianzas entre latentes y co-dependencia entre latentes y observadas.

En los SEM, además de la evaluación que se hace a cada parámetro, también se hace una evaluación global (omnibus) de todas las relaciones planteadas en el modelo, a través de tres tipos de índice de bondad de ajuste:

- Índices de ajuste absoluto (GFI, NCP, SNCP, SRMR, RMSEA)
- Índices de ajuste incremental (NFI, TLI, IFI, RNI, RFI, CFI).
- Índices de parsimonia (AGFI, PNFI, PGFI, AIC).

La mayoría de los índices de ajuste se expresan en el rango $[0, 1]$. Algunas excepciones son los índices de χ^2 y sus derivados y algunos índices como AIC.

Algunos índices (e.g., GFI, CFI) deben mostrar valores igual o superiores a 0,9 para decir que el modelo contrastado es plausible. Otros índices (e.g., SRMR, RMSEA) deben mostrar valores menor o igual a 0,1 para sostener la plausibilidad de un modelo contrastado.

Ejercicio Práctico

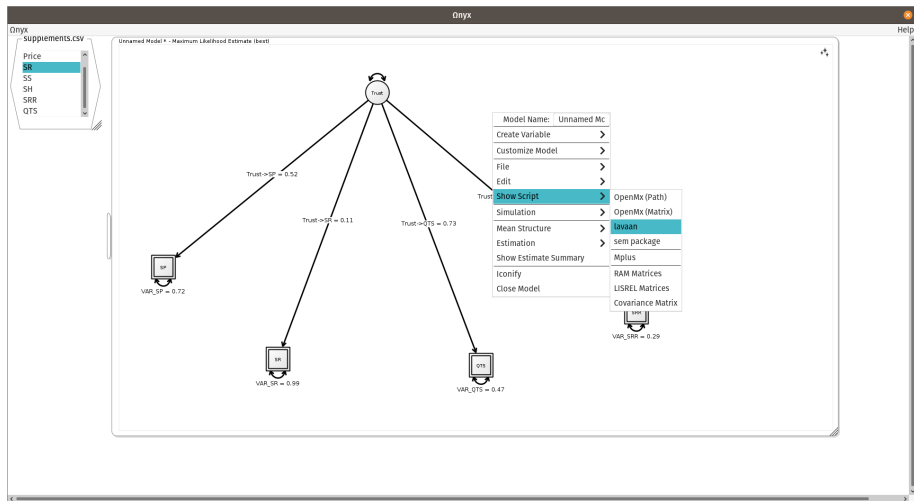
Vamos a realizar un ejercicio sencillo de análisis factorial confirmatorio, siguiendo las recomendaciones de este video-tutorial

<https://youtu.be/dFN6TBpbf3E> con los datos disponible acá:

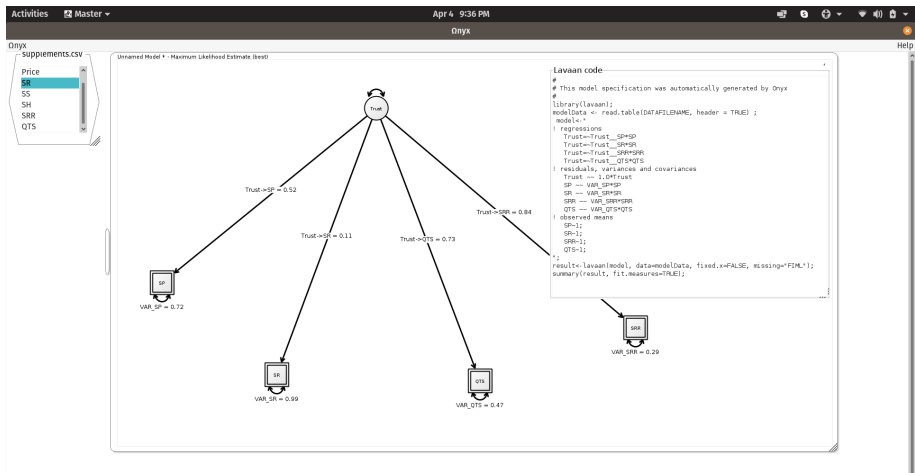
https://drive.google.com/file/d/1YDWNc9TYlyuUwE2hwKKkgqS-UT_xAA4y/view?usp=sharing

Ejercicio Práctico

Clic derecho —> Show Script —> lavaan



Ejercicio Práctico



Ejercicio Práctico

```
1 library(lavaan);
2 modelData <- read.table(DATAFILENAME, header = TRUE) ;
3 model<-"
4 ! regressions
5   Trust=~Trust__SP*SP
6   Trust=~Trust__SR*SR
7   Trust=~Trust__SRR*SRR
8   Trust=~Trust__QTS*QTS
9 ! residuals, variances and covariances
10  Trust ~~ 1.0*Trust
11  SP  ~~ VAR_SP*SP
12  SR  ~~ VAR_SR*SR
13  SRR ~~ VAR_SRR*SRR
14  QTS ~~ VAR_QTS*QTS
15 ! observed means
16  SP~1;
17  SR~1;
18  SRR~1;
19  QTS~1;
20 ";
21 result<-lavaan(model, data=modelData, fixed.x=FALSE, missing="FIML");
22 summary(result, fit.measures=TRUE);
```

En la línea 2, hay que reemplazar "DATAFILENAME" por el nombre del archivo que contiene los datos.