

Análisis de modelos de ecuaciones estructurales mediante el paquete lavaan

Structural equation modeling analysis using the lavaan package

Jose M Sallan¹, Vicenc Fernandez, Pep Simo, Oriol Lordan, David Gonzalez-Prieto

Abstract (English) Structural equation modeling is a technique to assess models with complex relationships between variables (path analysis), and models with latent variables. An especially interesting case is confirmatory factor analysis, a technique that allows assessing the construct validity of measurement scales. In this paper the lavaan package of R is presented, which easily allows testing structural equation models. Information about the model syntax is reported, and of the possibilities of the package.

Resumen (Castellano) El análisis de ecuaciones estructurales es una técnica para evaluar modelos con relaciones complejas entre variables (path analysis), así como modelos con variables latentes. Un caso especialmente interesante de aplicación es el análisis factorial confirmatorio, que permite evaluar la validez de constructo de escalas de medida. En este artículo se presenta el paquete de R lavaan, que permite evaluar con facilidad modelos de ecuaciones estructurales. Se reporta información sobre la sintaxis del modelo, así como de las posibilidades del paquete.

Keywords: structural equation models, confirmatory factor analysis, R, lavaan;
Palabras clave: modelos de ecuaciones estructurales, análisis factorial confirmatorio, R, lavaan.

¹ Jose M Sallan(✉)

Departamento de Organización de Empresas (DOE). Escola Tècnica Superior d'Enginyeries Industrial i Aeronàutica de Terrassa (UPC), C/ Colom, 08222 Terrassa, Spain
e-mail: jose.maria.sallan@upc.edu

1 Introducción

El análisis de ecuaciones estructurales es una técnica de gran utilidad en la evaluación de modelos con relaciones complejas entre variables, como los definidos en la técnica de *path analysis*. También es la técnica a utilizar cuando los modelos incluyen variables latentes, medidas a partir de dos o más variables observables. Un caso particular especialmente relevante de ecuaciones estructurales es el análisis factorial confirmatorio, utilizado para evaluar la validez de constructo de escalas de medida.

R es un entorno de software para análisis estadístico y de generación de gráficos, que también puede usarse para cálculo matricial (R core development team, 2011). Al ser software libre, R puede extenderse fácilmente mediante paquetes creados por la comunidad. Este hecho confiere a R un gran potencial para su uso en análisis de datos, pero en contrapartida tiene una interfaz de usuario poco amigable, y obliga a sus usuarios a escoger el paquete más adecuado de entre los diferentes desarrollados para una misma técnica.

En este artículo pretendemos presentar lavaan, un paquete de R para análisis de ecuaciones estructurales aún en desarrollo, pero que ya tiene funcionalidades suficientes para realizar los análisis requeridos en gran número de estudios que emplean esta técnica. Comenzaremos con una breve descripción de las técnicas de ecuaciones estructurales. Seguidamente describiremos la sintaxis con la que se entran modelos con lavaan, y las posibilidades de análisis del paquete. En las conclusiones se valora los pros y contras de usar lavaan frente a programas comerciales.

2 Modelos de Ecuaciones Estructurales

Los modelos de ecuaciones estructurales (SEM) nos permiten evaluar si un modelo que expresa una determinada relación entre las variables observables se ajusta a los datos empíricos (Schumacker y Lomax, 2004). La evaluación de un SEM requiere una primera fase de especificación del modelo, seguida de las fases de identificación y de estimación. En la fase de *especificación* se definen las relaciones a evaluar entre variables. Estas relaciones pueden ser más complejas que las definidas en los modelos lineales, en los que hay una única variable dependiente y pueden existir múltiples variables independientes. En SEM podemos tener tanto variables observables como variables latentes (no observables directamente). De hecho, con SEM podemos evaluar, entre otros modelos:

- *Path analysis*: Modelos que establecen relaciones entre variables observables. Estas relaciones pueden incluir relaciones recíprocas (ecuaciones simultáneas) y recursivas.

- *Análisis factorial confirmatorio*: modelo en que se establecen unas relaciones entre variables observables y unas variables latentes. Generalmente, se permite que las variables latentes correlacionen entre sí.
- *Ecuaciones estructurales*: modelo en el que se evalúa conjuntamente en modelo de medida (relación entre variables latentes y observables) y el modelo estructural (relaciones entre variables latentes).

En la fase de *identificación* se establece la relación entre los parámetros del modelo y las covarianzas de las variables observables. En la identificación se obtienen los *grados de libertad* del modelo, definidos como la diferencia entre el número de covarianzas diferentes entre variables observables, y el número de parámetros a estudiar en el modelo. Sólo se puede evaluar el ajuste del modelo a los datos en los modelos sobreidentificados, con valor de grados de libertad positivo.

Finalmente, en la fase de *estimación*, se obtienen estimadores de los parámetros del modelo, e indicadores del ajuste del modelo a los datos. El método de estimación más empleado es el de máxima verosimilitud (*maximum likelihood*), que exige que las variables observables sigan una distribución normal multivariante. Para muestras no normales existen otros métodos, como el de mínimos cuadrados ordinarios o el asintótico libre de distribución.

3 El Paquete lavaan

En R existen diversos paquetes para análisis de ecuaciones estructurales. Los más conocidos son *sem* (Fox 2006, 2010), *OpenMx* (Bates, Mehta y Fox, 2011) y *lavaan* (Rosseel 2011, 2012a, 2012b). Aunque aún está en fase de desarrollo, *lavaan* permite escribir los modelos de medida y estructurales más sencillamente que los otros dos paquetes, y permite realizar análisis relativamente complejos, como el análisis de grupos de ecuaciones estructurales, o el uso de métodos de estimación alternativos al de máxima verosimilitud. Para instalar *lavaan*, hay que abrir una consola de R y escribir:

```
install.packages("lavaan")
```

Para comprobar que el paquete se ha instalado correctamente, y siempre que queramos usar el paquete en R, hay que escribir en la consola:

```
library(lavaan)
```

3.1 Especificación de Modelos SEM con lavaan

Las referencias para la sintaxis de especificación de modelos en *lavaan* son el texto introductorio de Rosseel (2011) y el manual de referencia (Rosseel, 2012b). Seguidamente mostraremos dos ejemplos de especificación de modelos: un análisis

factorial confirmatorio y un modelo estructural. En la figura 1 tenemos el diagrama de un modelo de análisis factorial confirmatorio obtenido a partir de nueve de las 26 variables de HolzingerSwineford1939, un conjunto de datos disponible en R y que se usa frecuentemente en documentos sobre ecuaciones estructurales. Se trata de los resultados de un test de capacidad mental, en el que se establece como hipótesis que las nueve variables miden tres factores:

- Las variables x_1 , x_2 y x_3 miden el factor representado por la variable latente *visual*
- Las variables x_4 , x_5 y x_6 miden el factor *textual*
- Las variables x_7 , x_8 y x_9 miden el factor *veloc*

En la propia figura 1 tenemos el modelo en sintaxis de lavaan, con los siguientes elementos por defecto:

- A cada una de las variables observables se le asocia una variable de error, cuya varianza es uno de los parámetros del modelo. Por defecto, estos errores no están correlacionados.
- Para fijar la escala de la variable latente, se fija el peso factorial del primero de sus indicadores igual a 1.
- Se hace que todas las correlaciones entre variables latentes sean diferentes de cero.

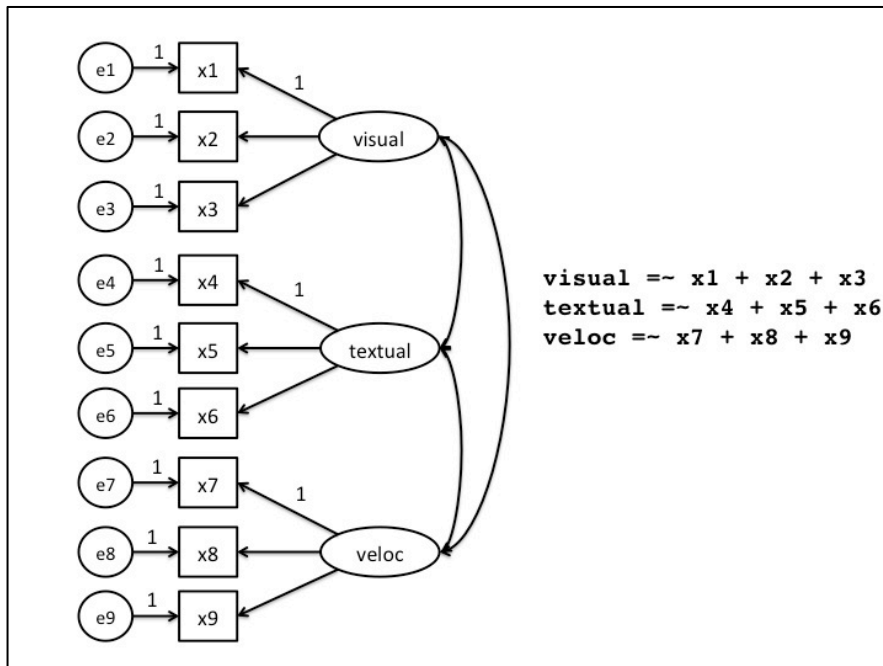


Fig 1. Un modelo de análisis factorial confirmatorio y su especificación en lavaan

Estas especificaciones por defecto son válidas para las funciones `cfa()` y `sem()`, las usadas más frecuentemente. En Rosseel (2011) puede encontrarse ejemplos e indicaciones para hacer las modificaciones necesarias en cada caso. Alternativamente, podemos optar por especificar el modelo completo, sin elementos por defecto, usando la función `lavaan` (Rosseel, 2012a).

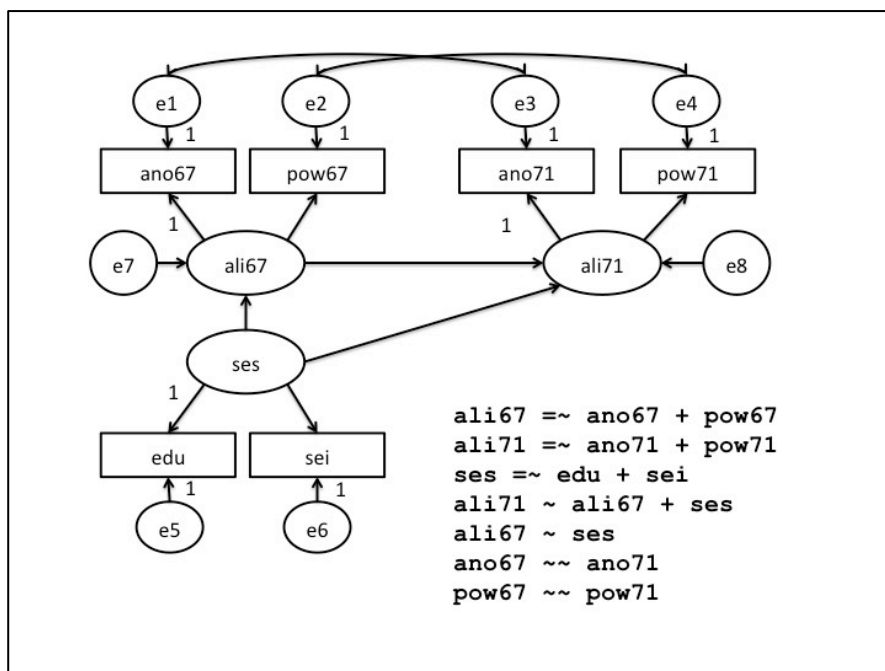


Fig 2. Un modelo de ecuaciones estructurales y su especificación en `lavaan`

En la figura 2 tenemos otro modelo estructural clásico en la literatura de ecuaciones estructurales, el modelo de estabilidad de alienación de Wheaton, Muthén, Alwin y Summers (1977). Este modelo incluye un *modelo de medida* (relación entre variables latentes y observables) y un *modelo estructural* (relaciones entre variables latentes). En Fox (2006) o en Rosseel (2012a) puede encontrarse los datos originales, en este caso el número de observaciones y la matriz de covarianzas entre variables observables. En la propia figura 2 tenemos el modelo en sintaxis `lavaan`. Además de la especificación del modelo de medida, similar a la de la figura 1, este modelo incluye la especificación del modelo estructural, y el establecimiento de correlaciones entre varianzas de variables observables.

3.2 Estimación de Modelos con lavaan

En lavaan disponemos de tres funciones para estimar modelos de ecuaciones estructurales. La función `cfa()` está dedicada a modelos de análisis factorial confirmatorio, que pueden ser estimados con una función más sencilla que los modelos estructurales genéricos que son estimados con la función `sem()` (cf. Neale et al., 2003). La función `lavaan()` también permite estimar modelos genéricos de ecuaciones estructurales, en los que se especifica el modelo completo. En `sem()` y `cfa()` especificamos primero el modelo, y después la fuente de los datos (podemos utilizar los datos brutos, o bien la matriz de covarianzas y el número de observaciones). Haciendo `fit.measures=TRUE`, podemos obtener un listado de indicadores de ajuste del modelo.

Así, para estimar el modelo de la figura 1 haremos:

```
library(lavaan)
modelo1 <- 'visual =~ x1 + x2 + x3
textual =~ x4 + x5 + x6
veloc =~ x7 + x8 + x9'
fit <- cfa(modelo1, data = HolzingerSwineford1939)
summary(fit)
```

Para especificar en R el modelo de la figura 2, haremos:

```
library(lavaan)
wheaton.cov <- matrix(c(
11.834,      0,      0,      0,      0,      0,
6.947,      9.364,      0,      0,      0,      0,
6.819,      5.091, 12.532,      0,      0,      0,
4.783,      5.028,  7.495,  9.986,      0,      0,
-3.839,     -3.889, -3.841, -3.625,  9.610,      0,
-21.899,   -18.831, -21.748, -18.775, 35.522, 450.288),
6, 6, byrow=TRUE)
colnames(wheaton.cov) <- rownames(wheaton.cov) <- c("ano67", "pow67",
"ano71", "pow71", "edu", "sei")
wheaton.modelo <- '
ali67 =~ ano67 + pow67
ali71 =~ ano71 + pow71
ses =~ edu + sei
ali71 ~ ali67 + ses
ali67 ~ ses
ano67 ~~ ano71
pow67 ~~ pow71'
fit <- sem(wheaton.modelo, sample.cov=wheaton.cov, sample.nobs=932)
summary(fit, standardized=TRUE, fit.measures=TRUE)
```

Para este modelo, los parámetros indicadores del ajuste son:

Lavaan (0.4-9) converged normally after 82 iterations

Number of observations	932
Estimator	ML
Minimum Function Chi-square	4.735
Degrees of freedom	4
P-value	0.316
Chi-square test baseline model:	
Minimum Function Chi-square	2133.722
Degrees of freedom	15
P-value	0.000
Full model versus baseline model:	
Comparative Fit Index (CFI)	1.000
Tucker-Lewis Index (TLI)	0.999
Loglikelihood and Information Criteria:	
Loglikelihood user model (H0)	-15213.274
Loglikelihood unrestricted model (H1)	-15210.906
Number of free parameters	17
Akaike (AIC)	30460.548
Bayesian (BIC)	30542.783
Sample-size adjusted Bayesian (BIC)	30488.792
Root Mean Square Error of Approximation:	
RMSEA	0.014
90 Percent Confidence Interval	0.000 0.053
P-value RMSEA <= 0.05	0.930
Standardized Root Mean Square Residual:	
SRMR	0.007

Además de los indicadores de ajuste del modelo indicados, obtenemos información de los estimadores de los parámetros del modelo, junto con las pruebas de hipótesis nula para cada uno de ellos (no se muestra por razones de espacio). También podemos obtener los *índices de modificación*, que estiman la variación que puede producirse en el modelo cuando introducimos una nueva relación entre variables. En Rosseel (2012) podemos encontrar un listado completo de las posibilidades de lavaan en su actual estado de desarrollo, que incluyen modelos con grupos múltiples.

4 Conclusiones

En los últimos años, R se está convirtiendo en una alternativa válida a los paquetes estadísticos comúnmente usados en diversas áreas científicas, entre ellas las cien-

cias sociales (Culpepper y Aguinis, 2010). R permite aplicar la mayoría de técnicas estadísticas usando un único programa informático, y al ser software en código abierto permite la modificación del código a los investigadores, y beneficiarse de las aportaciones de la comunidad. El desarrollo de lavaan es un ejemplo de cómo pueden resolverse dos de los mayores problemas del software en código abierto: la mala documentación y los costes de aprendizaje.

A pesar de que está en las fases iniciales de su desarrollo, lavaan cuenta ya con una buena documentación (Rosseel, 2011, 2012b), y la forma de especificar el modelo reduce, a nuestro entender, los costes de aprendizaje para aquellos usuarios familiarizados con los modelos de ecuaciones estructurales. Cabe esperar que, en el futuro próximo, dispongamos de un paquete completo y estable en código abierto para la aplicación de una de las técnicas más útiles en ciencias sociales.

5 Referencias

- Bates T, Mehta P, Fox J (2011) OpenMx: an Open Source Extended Structural Equation Modeling Framework. *Psychometrika*, Vol. 76, pp. 306–317.
- Culpepper, S. A, Aguinis, H, (2010) R is for revolution: a cutting-edge, free, open source statistical package. *Organizational research methods*, Vol. 14, No. 4, pp. 735–740.
- Fox J (2006) Structural Equation Modeling with the sem Package in R. *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 13, pp. 465–486.
- Fox J (2010). sem: Structural Equation Models. R package version 0.9-21, disponible en <http://CRAN.R-project.org/package=sem>.
- Neale, M. C., Boker, S. M., Xie, G., Maes, H. H. (2003). *Mx: Statistical Modeling*. VCU Box 900126, Richmond, VA 23298: Department of Psychiatry. 6th Edition.
- Rosseel, Y. (2012a). lavaan: an R package for structural equation modeling. *Journal of Statistical Software* (in press).
- Rosseel, Y. (2012b). Package ‘lavaan’. Disponible en: <http://cran.r-project.org/web/packages/lavaan/lavaan.pdf>
- Rosseel, Y. (2011). Lavaan: an R package for structural equation modelling and more. Version 0.4-9 (BETA). Disponible en <http://users.ugent.be/~yrosseel/lavaan/lavaanIntroduction.pdf>
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, disponible en <http://www.R-project.org>.
- Schumacker, R. E., Lomax, R. G. (2004) *A beginner’s guide to structural equation modeling*. Lawrence Erlbaum
- Wheaton, B., Muthén, B., Alwin, D. F., & Summers, G. F. (1977) Assessing reliability and stability in panel models. In D. R. Heise (Ed.), *Sociological methodology 1977* (pp. 84–136). San Francisco: Jossey-Bass.