

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO
ESCUELA DE ESTADÍSTICA

TÍTULO IMPRESIONANTE

CÉSAR ANDRÉS GAMBOA SANABRIA B12672
ANDRÉS ESTEBAN ARGUEDAS LEIVA B40535

Ciudad Universitaria Rodrigo Facio, Costa Rica

2020

RESUMEN

El Resumen

*Palabras clave:*kkk

ABSTRACT

El resumen en inglés

*Palabras clave:*kkk

Índice

1	INTRODUCCIÓN	1
1.1	Antecedentes	1
1.2	El problema	1
1.3	Objetivos del estudio	2
1.3.1	Objetivo general	2
1.3.2	Objetivos específicos	2
1.4	Metodología de la investigación	2
1.5	Organización del estudio	2
2	METODOLOGÍA	3
2.1	Casos de simulación	3
2.2	Generación de datos con kurtosis	3
2.3	Modelo teórico a estimar	4
2.4	Medidas de bondad de ajuste	4
2.5	Software y paquetes	4
3	RESULTADOS	5
3.1	Introducción	5
4	CONCLUSIONES Y RECOMENDACIONES	6
4.1	Introducción	6
4.2	Conclusiones	6
4.3	Recomendaciones	6
5	ANEXOS	7
5.1	La función <code>funcion_1</code>	7
6	REFERENCIAS	7

1 INTRODUCCIÓN

1.1 Antecedentes

Los Modelos de Ecuaciones Estructurales (en adelante SEM, por sus siglas en inglés) representan un compendio de métodos estadísticos que buscan estimar y examinar las relaciones existentes entre varias mediciones fácilmente observables con conceptos más abstractos, denominados constructos, que no pueden ser medidos ni analizados de manera directa. Los SEM trabajan de una manera similar a los modelos de regresión más clásicos, pero representan una mejora pues analizan las relaciones causales lineales entre las variables involucradas al mismo tiempo que los errores de medición (Beran & Violato, 2010).

Los SEM están presentes en multitud de campos de investigación. Según Beran y Violato (2010), la cantidad de referencias a SEM en 1994 fueron de 164, aumentaron a 343 en el 2000 y llegaron a 742 en el 2009, lo cual es una señal de que muchos investigadores alrededor del mundo están mostrando cada vez más interés en este tipo de estudios, pues representan una potente herramienta para la investigación partiendo de la teoría sustantiva que poseen los diversos estudios.

Uno de los principales campos de aplicación de los SEM son las ciencias sociales, pues se busca explicar y/o predecir con un grado de validez el comportamiento específico de una o varias personas en grupo. Teniendo siempre en consideración (aunque de forma limitada) las condiciones que afectan a cada individuo involucrado en el estudio, así como las características propias de su entorno, los grupos de investigación pueden definir factores y relaciones latentes que se encuentran implícitas en el comportamiento humano. Este tipo de investigaciones permite entender los fenómenos no solo de forma descriptiva, sino que es posible también determinar relaciones de causalidad (Tarka, 2018).

Las variables indicadoras que sirven para construir las relaciones implícitas en cuestión, llamadas comúnmente constructos, pueden llegar a comportarse de manera muy diversa. Las ciencias sociales, al trabajar con seres humanos, es común trabajar con variables cuyo comportamiento es particularmente irregular, presentando valores muy distintos entre los sujetos de estudio, generando de esta manera que los indicadores de manera multivariada no sigan una distribución normal, lo cual representa un supuesto fundamental al trabajar con SEM (Sura-Fonseca, 2020). El no cumplimiento de este supuesto puede deberse, entre otras cosas, a medidas particularmente altas o bajas de una medida estadística en específico: La kurtosis.

1.2 El problema

Si al trabajar con un SEM no se cumple el supuesto de normalidad multivariada y además el modelo se estima vía máxima verosimilitud podría cometerse el error de sobreestimar el estadístico chi-cuadrado, el cual sirve de referencia para conocer la magnitud de la diferencia entre la matriz de covariancias estimadas por el modelo con la obtenida en la muestra. Lo anterior suele llevar al rechazar modelos que en realidad resumen bien la realidad y además a la subestimación de los errores asociados a los parámetros, lo cual genera interpretaciones inadecuadas en lo referente a la significancia de las relaciones planteadas por el modelo teórico. Considerar distintos niveles de kurtosis permite conocer el impacto que esta medida tiene sobre las estimaciones de un SEM dependiendo del tamaño de muestra utilizado (Muthen & Kaplan, 1992).

1.3 Objetivos del estudio

La presente investigación busca estudiar el efecto que tienen distintos niveles de kurtosis en varios tamaños de muestra sobre las estimaciones de un SEM. Para ello, tomando como base un estudio de la Universidad de California (Gao, Mokhtarian, & Johnston, 2008) se plantean los siguientes objetivos:

1.3.1 Objetivo general

Comparar mediante un estudio de simulación las estimaciones de modelos de ecuaciones estructurales en presencia de variables observadas con niveles de kurtosis de 0, 0.62, 6.65, 21.41 y 13.92 en tamaños de muestra de 50, 100, 200, 400 y 800.

1.3.2 Objetivos específicos

- 1) Definir como modelo poblacional el obtenido por Sura-Fonseca (2020) con dos variables exógenas y una endógena con tres variables indicadoras cada uno (**página 99 de la tesis**) como modelo de referencia teórico cuyas cargas factoriales se utilizarán para la generación de los datos simulados.
- 2) Medir el posible sesgo causado en la estimación de los modelos mediante el estadístico chi-cuadrado del modelo y la raíz del cuadrado medio de error de aproximación (RMSEA), la raíz de residuos de cuadrado medio estandarizado (SRMR) y el índice de bondad de ajuste (GFI).
- 3) Comparar los valores poblacionales de las cargas factoriales con los obtenidos en las simulaciones.
- 4) Publicar en una revista científica con revisión por pares el manuscrito final, en forma de un artículo científico.

1.4 Metodología de la investigación

AQUÍ VA UN RESUMEN DE LO QUE PONEMOS EN LA SECCIÓN DE METODOLOGÍA

1.5 Organización del estudio

AQUÍ VA UNA DESCRIPCIÓN BREVE DE CADA ETAPA DEL TRABAJO

2 METODOLOGÍA

2.1 Casos de simulación

2.2 Generación de datos con kurtosis

Los datos fueron simulados mediante la función `simulateData` del paquete `lavaan` (Rosseel, 2012), el cual utiliza el método propuesto por Vale y Maurelli (Vale & Maurelli, 1983) para la simulación de datos no normales multivariados. Este método, comúnmente conocido como VM, se basa en el método propuesto por Fleishman (Fleishman, 1978), el cual, con base en una variable aleatoria distribuida como una normal estándar, permite simular una variable con un promedio, variancia, asimetría y kurtosis dada. El método VM permite especificar, adicionalmente, correlaciones entre las variables a estimar. Para utilizar el método de Fleishman, para generar una cierta variable aleatoria Y , se utiliza la siguiente ecuación:

$$Y = a + bX + cX^2 + dX^3 \quad (1)$$

donde $X \sim \mathcal{N}(0, 1)$. Es decir, se puede generar una variable no normal Y , con sus primeros cuatro momentos iguales a valores especificados, con base en los valores a , b , c y d de la ecuación 1, con base en una variable normal estándar X hasta su tercer potencia. Luego, para poder obtener los valores de a , b , c y d , se necesitan resolver las siguientes ecuaciones de forma simultánea:

$$b^2 + 6bd + 2c^2 + 15d^2 - 1 = 0 \quad (2)$$

$$2c(b^2 + 24bd + 105d^2 + 2) - \gamma_1 = 0 \quad (3)$$

$$24(bd + c^2(1 + b^2 + 28bd) + d^2(12 + 48bd + 141c^2 + 225d^2)) - \gamma_2 = 0 \quad (4)$$

donde γ_1 es la asimetría deseada y γ_2 es la kurtosis deseada, además se define $a = -c$. Con base en las constantes calculadas a , b , c y d , además de una variable normal estándar, se puede simular variables no normales. Para poder generalizar el método de Fleishman a variables aleatorias multivariantes, Vale y Maurelli proponen una generalización. Esta se basa, para el caso bivariado, en la generación de dos variables aleatorias independientes, $X_1, X_2 \sim \mathcal{N}(0, 1)$, para la cuales se obtienen las constantes a , b , c y d , para cada una de dichas variables, como se describe en el método de Fleishman, obteniendo así el vector $w'_1 = (a_1, b_1, c_1, d_1)$, para el caso de X_1 , y el vector $w'_2 = (a_2, b_2, c_2, d_2)$, para el caso de X_2 . Además, se definen los vectores $x'_1 = (1, X_1, X_1^2, X_1^3)$ y $x'_2 = (1, X_2, X_2^2, X_2^3)$. Por lo tanto, se pueden crear variables no normales, Y_1 y Y_2 , como:

$$Y_1 = w'_1 x_1$$

$$Y_2 = w'_2 x_2$$

donde se puede verificar que:

$$\begin{aligned} r_{Y_1, Y_2} = & \rho_{X_1, X_2} (b_1 b_2 + 3b_1 d_2 + 3d_1 b_2 + 9d_1 d_2) \\ & + \rho_{X_1, X_2}^2 (2c_1 c_2) + \rho_{X_1, X_2}^3 (6d_1 d_2) \end{aligned}$$

Y resolviendo esta ecuación en términos de ρ_{X_1, X_2} se puede obtener una matriz de correlaciones para generar datos normales multivariados, que pueden ser transformados en variables no normales mediante el método de Fleishman.

2.3 Modelo teórico a estimar

2.4 Medidas de bondad de ajuste

2.5 Software y paquetes

La simulación de los datos, junto con la estimación de los modelos, se realizó mediante el paquete **lavaan** (Rosseel, [2012](#)) usando el software R (R Core Team, [2020](#)) mediante la interfaz gráfica de RStudio (RStudio Team, [2015](#)).

3 RESULTADOS

3.1 Introducción

4 CONCLUSIONES Y RECOMENDACIONES

4.1 Introducción

4.2 Conclusiones

4.3 Recomendaciones

5 ANEXOS

5.1 La función `funcion_1`

Código 1: Una función

```
funcion_1 <- function(x,y){
  x+y
}
```

6 REFERENCIAS

- Beran, T. N., & Violato, C. (2010). Structural equation modeling in medical research: a primer. *BMC Research Notes*, 3, 267-267. Recuperado de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2987867/#>
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43(4), 521-532. <https://doi.org/10.1007/BF02293811>
- Gao, S., Mokhtarian, P. L., & Johnston, R. A. (2008). Nonnormality of Data in Structural Equation Models. *Transportation Research Record*, 2082(1), 116-124. Recuperado de <https://doi.org/10.3141/2082-14>
- Muthen, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45(1), 19-30. Recuperado de <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2044-8317.1992.tb00975.x>
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Recuperado de <https://www.R-project.org/>
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. Recuperado de <http://www.jstatsoft.org/v48/i02/>
- RStudio Team. (2015). *RStudio: Integrated Development Environment for R*. Recuperado de <http://www.rstudio.com/>
- Sura-Fonseca, R. (2020). *Modelos de ecuaciones estructurales: consecuencias de la asimetría positiva en los indicadores endógenos sobre las estimaciones puntuales de sus coeficientes y la bondad de ajuste*. Recuperado de <http://www.kerwa.ucr.ac.cr/handle/10669/80716>
- Tarka, P. (2018). An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences. *Quality & Quantity: International Journal of Methodology*, 52(1), 313-354. Recuperado de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5794813/>
- Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48(3), 465-471. <https://doi.org/10.1007/BF02293687>