

Visualización de alto nivel para el análisis e interpretación de Modelos de Ecuaciones Estructurales *

Carlos A. Pérez Moncada *Estudiante de Maestría, Ingeniería de Sistemas - Universidad de Antioquia*

John Freddy Duitama Muñoz *Director - Universidad de Antioquia*

Juan Delgado Lastra *Co-director - Universidad de Antioquia*

Resumen: El presente documento tiene como finalidad describir de manera general la solución de software SEMVIZ. La cual es producto del proyecto de investigación asociado con una tesis de maestría en ingeniería, que tiene como objetivo principal proporcionar un método de análisis por medio de técnicas de visualización de alto nivel para los resultados de un modelo de ecuaciones estructurales (Structural Equation Model, SEM por sus siglas en inglés).

Keywords: SEMVIZ, modelos SEM, visualización, grámatica de los gráficos, técnicas de interacción gráfica, r-studio, shiny, dashboard

Introducción

Un Modelo de Ecuaciones Estructurales (Structural Equation Model, SEM por sus siglas en inglés) es una técnica de análisis estadístico de alto nivel que combina análisis factorial, análisis de rutas y modelos de regresión lineal. Donde resalta la definición de asociaciones entre variables medibles (observadas) y latentes o constructos (no observadas).

Dichos modelos son de mucha utilidad en contextos de investigación y a nivel empresarial; resaltando en especial: psicometría, econometría, bioingeniería y marketing.

En particular, tradicionalmente los entornos de software para el análisis estadístico de SEM se han basado en aplicaciones de escritorio (SPSS, STATA, LISREL, LAVAAN, entre otros). Adicionalmente, sus interfaces de presentación de información se basan en esquemas tabulares y son apoyadas por gráficas estáticas de enfoque netamente técnico y sin interacciones gráficas.

Lograr una técnica (proyecto de investigación: *Graphical Analysis Driven SEM*) que permita unir el análisis estadístico de los modelos SEM, con el manejo de una visualización enriquecida, adaptable y con capacidad de interacción para usuarios finales (no expertos en estadística de alto nivel), es claro que puede llegar a tener un aporte (demanda de uso) potencialmente muy importante para la áreas de aplicación mencionadas anteriormente.

i) Objetivo

Proporcionar una aplicación tipo dashboard en R-Shiny que permita una visualización interactiva de alto nivel para el análisis e interpretación de los resultados de un modelo de ecuaciones estructurales (SEM).

*Contacto: carlos.perezoft@gmail.com

ii) Metodología

Ya que la solución de software SEMVIZ se enmarca dentro de un proyecto de investigación para una maestría, la metodología de trabajo se enfocó en dos líneas:

1. Referenciación tecnológica sobre técnicas de visualización a nivel estadístico y de científico de datos

Se analizó el estado del arte en técnicas y plataformas tecnológicas que permitieran un enfoque de análisis interactivo a nivel de interfaz de usuario, que propocionaran un conjunto de elementos gráficos que facilitaran de forma efectiva la comprensión de los resultados obtenidos luego de estimar un modelo de ecuaciones estructurales.

2. Ciclo de desarrollo de software para SEMVIZ

Se realizó un proceso de ciclo de desarrollo de software, combinado con un proceso de análisis y definición de gráficos interactivos para visualización de alto nivel de los resultados obtenidos de un modelo de ecuaciones estructurales (SEM).

Análisis y Diseño: Se analizaron varios casos de estudio para ser utilizados como base en la implementación de la aplicación SEMVIZ y diseñar el conjunto de gráficos que permitieran el análisis de los resultados del modelo SEM estimado.

Se definen las opciones de menú relacionadas con el análisis de bondad de ajuste, analisis de pruebas de hipótesis, análisis del modelo de medición (factores) y análisis del modelo estructural (regresión).

Implementación: Basándose en una arquitectura de referencia (ver figura 1) que usa R como motor base, sobre la cual se usa el enfoque de una solución WEB, según las capas propocionadas por una aplicación tipo R-Shiny. Lo cual fundamentalmente implica una serie de elementos en el frontend (ui.R) que interactúan por medio de un esquema de programación reactiva (*orientada a eventos*) con el lado backend (server.R), que permite la generación de los componentes gráficos con capacidad de interacción dinámica.

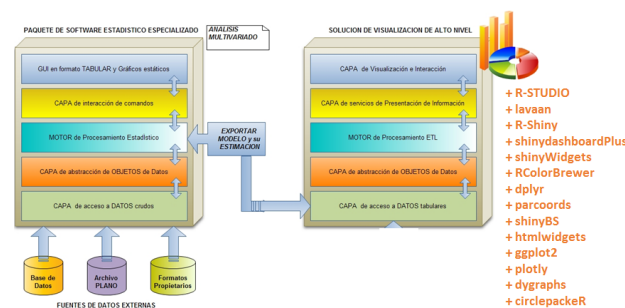


Figure 1: Modelo de Capas Operativas para SEMVIZ.

Control de Calidad (Pruebas de Software): Luego de tener la aplicación *R-Shiny dashboard SEMVIZ* con sus módulos funcionales implementados, se realizó un ciclo de control de calidad de software,

validando para cada gráfico la presentación de la información y las opciones dinámicas de ajuste para el gráfico respectivo.

Despliegue: Para uso de la aplicación SEMVIZ se procedió a realizar un despliegue en R-Studio Desktop Local y luego en un servidor Linux, con una instancia activa de **R-Studio Shiny Server** para acceso público desde un dominio propio del laboratorio de ingeniería de sistemas en la U. de A.

iii) Aportes principales

Enfocandonos en la solución de software SEMVIZ, como resultado del proyecto de investigación se tienen como aportes principales los siguientes:

Usar R-Lavaan y R-Shiny en conjunto

Debido al análisis de modelos de ecuaciones estructurales se hace uso del paquete **lavaan** para realizar la estimación del modelo SEM, y luego de esto usar todo el conjunto de operaciones y estructuras de datos (*data frame*) que permite **lavaan** para la extracción de información; que luego es adaptada de forma adecuada para ser procesada y presentada en los gráficos especializados propocionados por la plataforma R-Shiny y sus implementaciones tipo **htmlwidgets**.

Análisis de bondad de ajuste

Al obtener los resultados de un modelo de SEM, los paquetes estadísticos propocionan una serie de índices de bondad de ajustes absolutos, comparativos y parsimoniosos (ver figura 2). Estos índices tienen una serie de validaciones que indican al analista si el modelo estimado tiene una buena bondad de ajuste y es útil para posteriores análisis de hipótesis e inferencias sobre las variables del modelo.

En este aspecto la aplicación SEMVIZ propociona una sección tipo dashboard con elementos tipo semáforo que dan de forma directa un conjunto de componentes visuales asociados con la bondad de ajuste del modelo y los niveles de aceptación de los índices obtenidos.

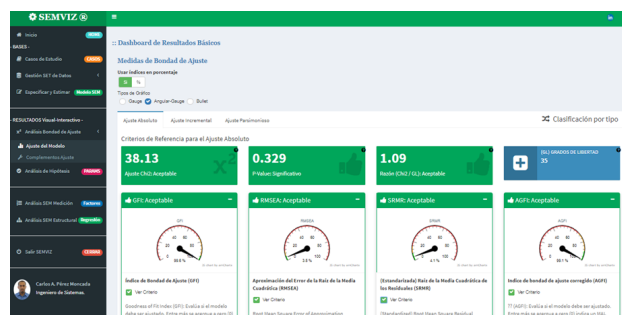


Figure 2: SEMVIZ dashboard de Bondad de Ajuste.

Análisis gráfico interactivo

SEMVIZ propociona una serie de gráficos interactivos para el análisis descriptivo (*exploratorio*) de los resultados obtenidos luego de estimar el modelo SEM (ver figura 3). Debido a que es posible

realizar una amplia variedad de validaciones y/o inferencias con dichos resultados, la aplicación SEMVIZ realiza una agrupación de los gráficos por medio de una serie de preguntas contextuales que se enfocan en los siguientes tipos:

- Distribución
- Correlación
- Barras
- Jerárquicos
- Redes
- Evolución
- Series
- Circularizar

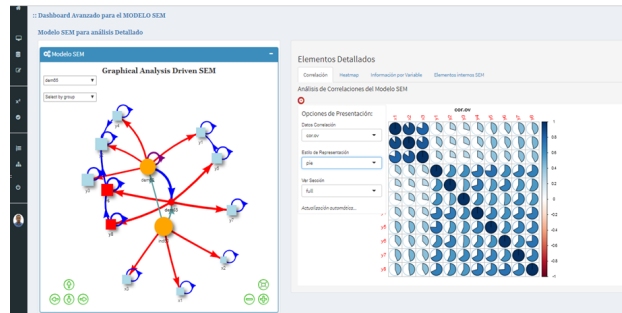


Figure 3: SEMVIZ Análisis gráfico interactivo.

Enfoque principal en los modelos de medición y estructural

Debido a que los resultados luego de la estimación de un modelo SEM son fundamentalmente una serie de valores numéricos (*score*) asociados con los elementos del modelo de medición (*factores*) y del modelo estructural (*regresión entre constructos*), es allí donde principalmente puede estar interesado el analista de negocio o investigador. Luego entonces, la aplicación SEMVIZ da un aporte con su galería de gráficos interactivos enfocados hacia dicha división funcional del modelo SEM (ver figura 4).

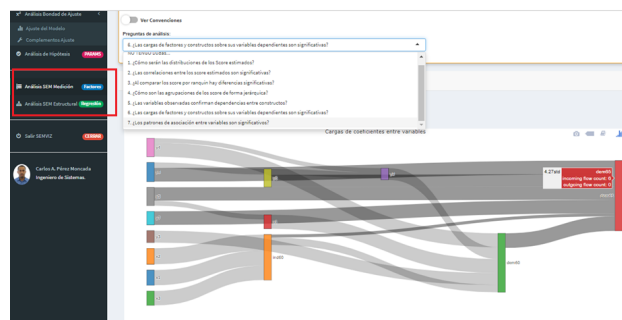


Figure 4: SEMVIZ análisis de modelo SEM.

iv) Implicaciones prácticas

SEMVIZ se deriva de un proyecto de investigación basado en la definición formal de una técnica de análisis científico de datos (Graphical Analysis Driven SEM - **GADS**) que permita realizar un análisis (*bondad de ajuste*) e interpretación (*inferencias*) de los resultados de un modelo SEM.

Se realiza la implementación de la aplicación SEMVIZ tipo dashboard interactivo basando en tecnologías de punta para visualización de información, con un entorno WEB y soportado sobre una plataforma altamente difundida en entornos académicos y empresariales: **R-STUDIO**.

La aplicación SEMVIZ presenta de manera efectiva el potencial tan amplio que actualmente tiene la plataforma R-Shiny con sus paquetes gráficos disponibles en su enorme galería de componentes tipo **htmlwidget**. Es allí donde se encuentra su aporte proporcionando un entorno integrado para el análisis científico de los datos obtenidos como resultado de un modelo SEM, esto articulando la información extraída con visualización interactiva de alto nivel.

Adicionalmente, se hace uso de los conceptos fundamentales de la especificación de gráficos en estadística establecidos inicialmente por Edward [Tufte \(2001\)](#), y que luego de forma significativa se potenció por medio de todas las definiciones establecidas en la gramática de los gráficos ([Wilkinson, 2005](#)); que en [R Core Team \(2019\)](#) fue llevado a la práctica de una manera muy robusta y confiable por medio del paquete ggplot2 ([Wickham, 2009](#)), que combinado con todo el potencial de R-Shiny y Shiny Dashboard logran soluciones académicas y empresariales de un alto nivel competitivo.

References

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

URL: <https://www.R-project.org/>

Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. 2nd ed. Graphics Press.

Wickham, Hadley. 2009. *ggplot2: Elegant graphics for data analysis*. Springer.

Wilkinson, Leland. 2005. *The Grammar of Graphics*. 2nd ed. Springer.