
Análisis Multivariante: Teoría y práctica de las principales técnicas

Autores:

Richard Ramirez-Anormaliza
Fabricio Guevara-Viejo
Mayra D'Armas Regnault
Rubi Pena-Holguin
Ruth Farias-Lema
Freddy Bravo-Duarte
Jose Diaz-Montenegro
Juan Calderon-Cisneros
Omar Franco-Arias
Fabian Ramirez-Granda
Edwuin Carrasquero Rodriguez
David Vargas-Decimavilla
Roberto Basurto-Quilligana
Oscar Vargas-Ortiz
Javier Bermeo-Paucar
Jesmy Castelo-Gonzalez

Milagro, junio 2017

Título de la Obra: Análisis Multivariante: Teoría y práctica de las principales técnicas

Derecho de Propiedad Intelectual: GYE-008506

Depósito Legal: GYE-000268

ISBN: 978-9942-969-86-6

No existe responsabilidad por parte de los autores o editores si el lector actúa o deja de hacerlo como resultado del material expuesto en la presente publicación.

De esta primera edición. © Universidad Estatal de Milagro – UNEMI, 2017

Autores:

Richard Ramirez-Anormaliza

Fabricio Guevara-Viejo

Mayra D'Armas-Regnault

Rubi Pena-Holguín

Ruth Farias-Lema

Freddy Bravo-Duarte

Jose Diaz-Montenegro

Juan Calderon-Cisneros

Omar Franco-Arias

Fabian Ramirez-Granda

Edwuin Carrasquero-Rodriguez

David Vargas-Decimavilla

Roberto Basurto-Quilligana

Oscar Vargas-Ortiz

Javier Bermeo-Paucar

Jesmy Castelo-Gonzalez

Rector: Ing. Fabricio Guevara-Viejó, PhD.

Director del Proyecto: Ing. Richard Ramirez-Anormaliza, PhD.

Coordinadora del Proyecto: Ing. Mayra D'Armas-Regnault, PhD.

Revisores Pares: Ing. Danilo Holguín, MBA Dr. Heriberto Medina, MSc.

Ediciones Holguín S.A., equipo editorial:

Directora Editorial: Lic. Lucrecia Resabala Manosalvas, MSc.

Editor de Área: Ing. Jaime Samaniego, MSc.

Coordinador Editorial: Ing. Danilo Holguín Cabezas, MBA

Asistente Editorial: Ing. Johanna Coronel Vélez

Revisora de Ortografía y Estilo: Lic. Jenny Cabezas-Rodríguez, MSc.

Diseño Gráfico y Diagramación: Lic. Santiago Holguín

Universidad Estatal de Milagro – UNEMI

Cdla. Universitaria Km. 1.5 vía Milagro Km. 26

Teléfonos: (593) 04 2715081- 04 2715079

<http://www.unemi.edu.ec/>

Milagro – Ecuador

Todos los derechos reservados. Ninguna parte original de esta publicación puede ser reproducida, guardada en sistemas de archivo o transmitida, en ninguna forma o medio, sin previa autorización del Editor.

Autores



Richard Ramirez-Anormaliza

Doctor Cum Laude por la Universitat Politècnica de Catalunya (España). Profesor titular de la Facultad de Ciencias de la Ingeniería y Director del Departamento de Investigación, Desarrollo e Innovación Tecnológica de la Universidad Estatal de Milagro (Ecuador). Consultor de proceso de innovación tecnológica.



Fabricio Guevara-Viejo

Doctor por la Universidad Andina Simón Bolívar (Ecuador). Profesor titular de la Facultad de Ciencias de la Ingeniería y Rector de la Universidad Estatal de Milagro (Ecuador).



Mayra D'Armas Regnault

Doctora por la Universitat Politècnica de Catalunya (España). Profesora titular de la Facultad de Ciencias de la Ingeniería y Directora de la Revista CIENCIA UNEMI de la Universidad Estatal de Milagro (Ecuador).



Rubi Pena-Holguin

Profesora adscrita al Departamento de Investigación, Desarrollo e Innovación de la Universidad Estatal de Milagro (Ecuador).



Ruth Farias-Lema

Analista de Investigación del Departamento de Investigación de la Universidad Estatal de Milagro (Ecuador).



Freddy Bravo-Duarte

Profesor titular de la Cátedra de Redes de la Facultad de Ciencias de la Ingeniería de la Universidad Estatal de Milagro (Ecuador).

Autores



Jose Diaz-Montenegro

Profesor titular de asignaturas contables y tributarias de la Facultad de Ciencias Administrativas y Comerciales de la Universidad Estatal de Milagro (Ecuador).



Juan Calderón-Cisneros

Profesor titular de la Facultad de Ciencias Administrativas y Comerciales de la Universidad Estatal de Milagro (Ecuador).



Omar Franco-Arias

Profesor titular de la Facultad de Ciencias de la Ingeniería de la Universidad Estatal de Milagro (Ecuador).



Fabian Ramirez-Granda

Profesor titular de la Facultad de Ciencias Administrativas y Comerciales de la Universidad Estatal de Milagro (Ecuador).



Edwuin Carrasquero-Rodríguez

Doctor por la Université de Science et Technologies de Lille (Francia). Doctor por la Universidad Central de Venezuela (Venezuela). Profesor titular de la Facultad de Ciencias de la Ingeniería de la Universidad Estatal de Milagro (Ecuador). Director del Grupo de Investigación en Caracterización, Procesamiento y Protección de Materiales.

Autores



David Vargas-Decimavilla

Asesor del Vicerrectorado Académico y de Investigación de la Universidad Estatal de Milagro (Ecuador).



Roberto Basurto-Quilligana

Asesor del Vicerrectorado Académico y de Investigación de la Universidad Estatal de Milagro (Ecuador).



Oscar Vargas-Ortiz

Profesor de la Facultad de Ciencias Administrativas y Comerciales de la Universidad Estatal de Milagro (Ecuador).



Javier Bermeo-Paucar

Profesor de la Facultad de Ciencias de la Ingeniería de la Universidad Estatal de Milagro (Ecuador).



Jesmy Castelo-Gonzalez

Profesora de las cátedras Contabilidad e Ingeniería Económica en la Carrera de Ingeniería en Sistemas Computacionales de la Facultad de Ciencias de la Ingeniería de la Universidad Estatal de Milagro (Ecuador).

Dedicatoria

Richard Ramirez-Anormaliza

A mi familia, de forma especial a mi madre e hijos con todo mi amor y agradecimiento por su apoyo constante.

Fabricio Guevara-Viejo

A los que sueñan y luchan con un libro en la mano.

Mayra D'Armas Regnault

A mis padres por todo su amor y apoyo incondicional, a mis hijas fuente de inspiración y motivación.

Rubi Pena-Holguin

A mi madre y hermano que son mi pilar en la vida, y por quienes me esfuerzo día a día para darles lo mejor.

Ruth Farias-Lema

A Dios y a mi familia que siempre han sido mi fuente de inspiración.

Freddy Bravo-Duarte

A mi madre por darme la vida, a mi padre que me cuida desde el cielo, a mis hijos por darme la alegría de vivir y a mi esposa por regalarme su amor.

Jose Diaz-Montenegro

A Carmen, Elián...y una nueva ilusión que está por venir.

Juan Calderon-Cisneros

A Dios sobre todas las cosas, a mi madre, hermano que siempre me apoyan y mis hijos que son mi motor en todo.

Dedicatoria

Omar Franco-Arias

A Dios, que me ilumina para seguir logrando mis objetivos; a mis adorables padres Jorge Eladio que me guía desde el cielo y Ana que es mi fortaleza. A mi esposa Sonia con mucho amor. A mis hijos: Paula y Joel. A toda mi familia por su respaldo y dedicación.

Fabian Ramirez-Granda

A los estudiantes.

Edwuin Carrasquero-Rodriguez

A mi Lis, Nicoll, Daniel y Bebé por ser el centro de mi universo, mi todo, los amo.

David Vargas-Decimavilla

A mi familia y profesores, que a lo largo de mi vida contribuyeron a este profesional plasmar lecciones aprendidas. De manera especial a mi madre, que siempre me dijo: "Hijo nunca te rindas, si puedes hacerlo".

Roberto Basurto-Quilligana

A Dios, por haberme brindado salud; a mis padres, quienes en todo momento me alentaron de diferentes maneras y a mi esposa, cuyo apoyo incondicional durante este tiempo de fue de vital importancia.

Oscar Vargas-Ortiz

A mis hijos Nicolás y Martin, razón de mi lucha día tras día.

Jesmy Castelo-Gonzalez

A Dios, a mis hijos Romy, Emily y Richard; por ser impulso y protagonistas de cada uno de mis objetivos alcanzados.

Prólogo

El análisis multivariante es el conjunto de técnicas estadísticas y matemáticas que permiten el procesamiento simultáneo de las observaciones o datos de varias variables, encaminado a describir o interpretar un fenómeno bajo estudio.

Considerando la importancia de las técnicas multivariante para la investigación científica, se conformó un equipo de profesores para crear esta obra y dar conocer las principales técnicas, su utilización y aplicación práctica de dos paquetes informáticas que permiten procesar los datos en cada técnica descrita.

En este libro se aborda el análisis factorial, análisis de conglomerados, los modelos de ecuaciones estructurales y el análisis de clases latentes.

Nuestro agradecimiento especial a todos aquellos participantes en los ejercicios que se presentan en cada uno de los capítulos que ilustran el uso de las técnicas seleccionadas.

Índice de Contenido

Autores	3
Dedicatoria	5
Prólogo	7
Índice de Contenido	9

Capítulo 1:

Introducción al análisis multivariante	15
1.1. Introducción	16
1.2. Términos y conceptos en análisis multivariante	17
1.2.1. Variables	17
1.2.2. Variables latentes y observables	17
1.2.3. Variables independientes y variables dependientes	17
1.2.4. Indicadores reflectivos y formativos	18
1.2.5. Escalas de medición	18
1.2.6. Error de medición y medición multivariante	19
1.2.7. Validación y confiabilidad de los datos	20
1.2.8. Significancia estadística versus potencia estadística	20
1.3. Principales técnicas multivariante	21
1.3.1. Los componentes principales y análisis de factor común	23
1.3.2. La regresión múltiple	23
1.3.3. El análisis discriminante múltiple	23
1.3.4. La correlación canónica	24
1.3.5. El análisis multivariante de la varianza y covarianza (MANOVA)	24
1.3.6. El análisis conjunto	25
1.3.7. El análisis de conglomerados	26
1.3.8. Mapeo perceptual	26
1.3.9. El análisis de correspondencias	26
1.3.10. El análisis factorial confirmatorio (CFA)	26
1.3.11. Modelo de ecuaciones estructurales	27
1.4. Selección de la técnica de análisis multivariante	27
1.5. Análisis previo de los datos	27
1.6. Ejemplo de análisis previo de los datos	28
1.6.1. Examen de los datos	29
1.6.2. Datos ausentes	36
1.6.3. Aleatoriedad de los datos ausentes	41
1.6.4. Imputación de datos	41
1.7. Referencias	47

Capítulo 2:	
Análisis factorial	55
2.1. Introducción al análisis factorial	56
2.2. Campo de aplicación del análisis factorial	58
2.3. Software para análisis factorial	58
2.4. Ejemplos de análisis factorial	58
2.4.1. Utilización de SPSS en el análisis factorial	58
2.4.1.1. Introducción	58
2.4.1.2. Pasos	59
2.4.2. Utilización de R en análisis exploratorio de datos.	73
2.4.2.1. Definir directorio de trabajo	74
2.4.2.2. Carga de datos	74
2.4.2.3. Prueba Kaiser-Meyer-Olkin	74
2.4.2.4. Prueba de Barlett	75
2.4.2.5. Análisis Factorial - Componentes principales	75
2.4.2.6. Gráfico de sedimentación	77
2.4.2.7. Análisis final con rotación	78
2.5. Resumen del capítulo	79
2.6. Referencias	79

Capítulo 3:	
Análisis de conglomerados	83
3.1. Introducción	84
3.1.1. Propósito de los agrupamientos	84
3.1.2. Objetivo del análisis de conglomerados	85
3.2. Fundamento técnico	87
3.2.1. Selección de variables del análisis de conglomerados	88
3.2.2. Diseño de investigación mediante análisis de conglomerados	88
3.2.2.1. Detección de Atípicos	89
3.2.2.2. Medidas de Similitud	89
3.2.2.3. Medidas de Asociación	90
3.2.2.4. Estandarización de los datos	91
3.3. Clasificación de las técnicas	91
3.3.1. Métodos Jerárquicos	92
3.3.2. Métodos no Jerárquicos	94
3.4. Funcionamiento del Análisis de Conglomerados	95
3.5. Software para análisis de conglomerados	101
3.5.1. Ejemplos de Análisis de Conglomerados	102
3.5.1.1. WEKA	102
3.5.1.2. SPSS	111

3.5. Resumen del capítulo	116
3.6. Referencias	117

Capítulo 4:

Modelo de ecuaciones estructurales	119
4.1. Introducción	120
4.2. Características de los modelos de ecuaciones estructurales	121
4.3. Terminología y notación básica en la modelización de SEM	122
4.3. Enfoques de ecuaciones estructurales	123
4.6. Software para tratamiento de ecuaciones estructurales	127
4.7. Modelos de ecuaciones estructurales con enfoque PLS	127
4.7.1. Valoración del modelo global	128
4.7.2. Valoración del modelo de medida	129
4.7.2.1. Fiabilidad individual de los indicadores	129
4.7.2.3. Validez convergente	130
4.7.2.4. Validez discriminante	130
4.7.3. Valoración del modelo estructural	131
4.7.3.1. Valoración de problemas de colinealidad	131
4.7.3.2. Evaluación de los coeficientes path	131
4.7.3.3. Valoración del coeficiente de determinación	132
4.7.3.4. Valoración de los tamaños de los efectos	133
4.7.3.5. Valoración de la relevancia predictiva	133
4.7.4. Análisis multigrupo	134
4.7.4.1. Invarianza de medición	134
4.7.4.2. Significancia de las diferencias	135
4.8. Modelo de ecuaciones estructurales con enfoque en covarianzas	136
4.9. Ejemplo de SEM enfoque PLS	138
4.9.1. Introducción	138
4.9.2. Pasos	139
4.9.2.1. Instalación del software	139
4.9.2.2. Crear un proyecto	139
4.9.2.3. Crear un modelo	140
4.9.2.4. Crear variables latentes y conectarlas	140
4.9.2.5. Carga de datos	142
4.9.2.6. Cargar los indicadores	142
4.9.2.7. Calcular el modelo	144
4.9.2.8. Calcular significancia estadística (Remuestreo)	144
4.9.3. Valoración global del modelo	148
4.9.4. Valoración del modelo de medida	148
4.9.4.1. Fiabilidad individual de los indicadores	148
4.9.4.2. Fiabilidad del constructo	151

4.9.4.4. Validez Convergente	151
4.9.4.5. Validez Discriminante	152
4.9.5. Valoración del modelo estructural	155
4.9.5.1. Valoración de la colinealidad	155
4.9.5.2. Evaluación de los coeficientes path	155
4.9.5.3. Valoración del coeficiente de determinación	156
4.9.5.4. Valoración de los tamaños de los efectos	157
4.9.5.5. Valoración de la relevancia predictiva	157
4.10. Ejemplo de SEM enfoque basado en covarianzas	159
4.10.1. Introducción	159
4.10.2. Pasos	160
4.10.2.1. Establecer directorio de trabajo, carga de datos y paquete	160
4.10.2.2. Crea el modelo variable	160
4.10.2.3. Evaluar el modelo de media	161
4.10.2.4. Evaluar el modelo de estructural	165
4.10.3. Conclusiones	168
4.11. Resumen del Capítulo	169
4.12. Referencias	169

Capítulo 5:

Análisis de clases latentes	175
5.1. Introducción	176
5.2. Conceptos y referencia histórica	176
5.2.1. El análisis de clases latentes	177
5.2.2. Formulación matemática	178
5.2.2.1. El modelo de clase latente básico	178
5.2.2.2. Obtención de predicción de clases latentes	179
5.3. Las causas del análisis de variables latentes en el análisis multivariante	180
5.4. La relación entre la variable latente y sus indicadores	181
5.5. Software para el análisis LCA	181
5.5.1. Ajuste de un modelo LCA en software R	181
5.5.2. Instalación paquete polCA	182
5.5.3. Corriendo un modelo LCA en R	182
5.5.3.1. Elegir el número de grupos	182
5.5.3.2. Preparando los datos	183
5.5.3.3. Estimación del modelo	183
5.5.3.4. Ejemplo 1 (basado en Linzer y Lewis (2011))	185
5.6. Resumen del Capítulo	190
5.7. Referencias	190

Bibliografía	193
---------------------	------------

Anexos	207
---------------	------------

Listado de siglas

α	Alfa de cronbach
AMOS	Software estadístico para tratamiento de ecuaciones estructurales (http://www-03.ibm.com/software/products/es/spss-amos)
AVE	Varianza media extraída
CB	Enfoque de Covarianzas
CFA	Análisis factorial confirmatorio
EQS	Software estructural de modelado de ecuaciones
f^2	Valoración de los tamaños de los efectos
FIV	Factor de inflación de la varianza
HTMT	Relación reterotrait-monotrait
IEEE	Instituto de ingenieros eléctricos y electrónicos (https://www.ieee.org/)
IES	Instituciones de educación superior
IS	Sistema de información
ISO	Organización internacional de normalización (http://www.iso.org/)
LNR	Localización número de ruta
M	Media de la muestra
MANOVA	Análisis multivariante de la varianza y covarianza
MBC	Modelo basado en análisis las covarianzas
MICOM	Invarianza de medición mediante el procedimiento
MPLUS	Regresión y análisis de mediación uso
NFI	Índice de ajuste normado
NS	No significativo
O	Muestra original
OLS	Mínimos cuadrados ordinales
PATH	Modelos de trayectoria, senderos
ρ_A (rho_A)	Índice dijkstra-henseler's
ρ_c	Fiabilidad compuesta
PLS	Partial Least Squart (Mínimos Cuadrados Parciales)
Q^2	Stone-geisser, índice de relevancia predictiva de un modelo PLS
R^2	Valoración del coeficiente de determinación
RMS Theta	Medio de la raíz cuadrada de correlación de error
SEM	Modelo de ecuaciones estructurales
Sig.	Significancia estadística
SPSS	Paquete software estadístico para las ciencias sociales
SRMR	Normalización de raíz cuadrada media residual
STDEV	Desviación estándar
TIC	Tecnología de información y comunicación
URL	Localizador de recursos uniforme
www	World wide web

UNVENMI 2017



“No basta con alcanzar la sabiduría; es necesario saber utilizarla”

Marco Tullio Cicerón (106 a. C. – 43 a. C.)

Escritor, político y orador romano

Capítulo 1: Introducción al análisis multivariante

Richard Ramirez-Anormaliza
Freddy Bravo-Duarte
Edwain Carrasquero Rodríguez
Oscar Vargas-Ortíz

Objetivos de aprendizaje

En un estudio para cumplir los objetivos, responder a las preguntas de investigación y someter las hipótesis a prueba, se debe determinar su diseño (Hernández, Fernández-Collado, & Baptista, 2006). Como parte del diseño de una investigación es muy importante definir la técnica que se utilizará. Entre las opciones de técnicas las multivariante tienen gran acogida en los últimos años, por este motivo se planteó crear esta obra. Al finalizar este capítulo deberá ser capaz de lo siguiente:

- Conceptualizar las técnicas multivariantes y su aplicación.
- Distinguir las diferentes técnicas de análisis multivariante.
- Identificar los principales conceptos asociados a un estudio con técnica multivariante.
- Diferenciar entre variables latentes y variables observables.
- Establecer diferencias entre variables independientes y dependientes.
- Diferenciar entre indicadores reflectivos e indicadores formativos.
- Describir de manera general las técnicas multivariante y su aplicación.
- Seleccionar la técnica de análisis multivariante para un proceso investigo según las variables y objetivo que persigue.
- Analizar los datos previo a la aplicación de una técnica multivariante.

1.1. Introducción

El continuo incremento de los sistemas informáticos y la posibilidad de adquirir y procesar un gran número de datos en todas las disciplinas de las ciencias ha alentado el desarrollo y utilización del análisis estadístico de múltiples variables en todas las áreas del conocimiento. Por ejemplo en las ciencias económicas los métodos estadísticos multivariantes se utilizan para cuantificar el desarrollo de un país, o determinar las relaciones existentes entre ingresos y gastos de un grupo familiar, comprender el comportamiento de los consumidores y medir la calidad de productos y servicios (C.-C. Wu & Chiu, 2017)(Patton, 2012). En las ciencias de la ingeniería se emplean para diseñar equipos inteligentes que reconozcan objetos o caracteres, o para construir sistemas automatizados que aprendan de manera interactiva con su entorno y a su vez puedan establecer sistemas de control de proceso productivo para analizar la eficiencia energética (Corsini, De Propris, Feudo, & Stefanato, 2015), mientras que las ciencias de la informática para desarrollar sistemas de inteligencia artificial (Dobrzański, Trzaska, & Dobrzańska-Danikiewicz, 2014).

En el campo de las ciencias médicas para implementar procedimientos automáticos de ayuda al diagnóstico de enfermedades como el caso del cáncer de próstata o de páncreas tratado por radioterapia mediante un estudio multivariado (Albuisson et al., 1992) (Chang & Hsu, 2009). En ciencias de la tierra para la evaluación de los riesgos ambientales y económicos en los cambios climáticos (Fan, Shibata, & Chen, 2017). Interpretar y correlacionar los resultados de pruebas de aptitudes de pacientes en psicología. En el campo de la sociología y antropología para el análisis de encuestas de actitudes, opiniones sociales, estructura y composición de la población estudio sociocultural sobre la autopercepción, etc. En campo de la biología en el estudio y la clasificación de las mieles de las abejas adulteradas artificialmente (Amiry, Esmaili, & Alizadeh, 2017). En resumen el objetivo del estudio mediante la técnica del análisis multivariante en una población se basa en los siguientes aspectos:

- Organizar y agrupar todas las variables de un sistema en un conjunto de nuevas variables obtenidas mediante la transformación de las iniciales, sin una pérdida significativa de la información.
- Identificar la existencia de subconjuntos en los datos.
- Identificar y clasificar nuevas observaciones en grupos ya definidos.
- Relacionar dos o mas conjuntos de variables.

En el campo científico se analizan variables y cualquier estudio simultáneo de más de dos variables puede ser considerado como análisis multivariante. Por medio de las técnicas multivariante se puede crear conocimiento (J. Hair, Black, Babin, & Anderson, 2010). Un gran número de las técnicas multivariantes son

extensiones de las técnicas univariadas. Son ampliamente utilizadas, porque permiten a las organizaciones crear conocimiento y, con ello mejorar su toma de decisiones.

En la mayoría de las técnicas multivariantes se requiere de un trabajo cualitativo previo que consiste en seleccionar los constructos e indicadores para evaluar tales constructos. Para este proceso se pueden adaptar indicadores de estudios previos, o solicitar la contribución de expertos en el tema para lo cual se puede realizar un estudio Delphi, triangulación, etc.

La necesidad de información, en los últimos años es cada vez más fuerte, para ello se requiere procesar gran cantidad de datos, se realizan presentaciones de resúmenes, gráficos, es decir estadística simple. Pero para obtener información especializada y mejorar la toma de decisiones es necesario utilizar técnicas multivariantes.

1.2. Términos y conceptos en análisis multivariante

A pesar que el origen de las técnicas multivariantes se encuentra en las estadísticas univariadas y bivariadas, es necesario puntualizar ciertos términos y conceptos, que se describen más adelante.

1.2.1. Variables

Son el conjunto de constructos que se analizan, estos tienen pesos determinados empíricamente. El investigador selecciona las variables, las ponderaciones son determinadas por la técnica seleccionada, según el objetivo del estudio.

1.2.2. Variables latentes y observables

Las variables latentes se definen como aquellas que no son directamente observables o medibles, esto es, que no presentan una unidad de medida específica, por lo que cualquier estimación podría suponer un error asociado. Las variables observables son un conjunto de variables que se utilizan para definir o inferir la variable latente (Schumacker & Lomax, 2010).

1.2.3. Variables independientes y variables dependientes

Las variables independientes representan los procedimientos o escenarios que el investigador inspecciona para probar sus efectos sobre algún resultado, éstas también se denominan variables de tratamiento. Una variable independiente es manipulada en el curso de un experimento a fin de entender los efectos de tal manipulación sobre la variable dependiente.

En tanto que la variable dependiente es la que refleja los resultados de un estudio de investigación, se la considera como el resultado que podría obedecer al tratamiento experimental de lo que el investigador modifica o manipula (Namakforoosh, 2005).

1.2.4. Indicadores reflectivos y formativos

Los investigadores utilizan la teoría existente a la que llaman constructos y la miden por medio de datos a los que comúnmente denominan indicadores, estos constructos se relacionan entre ellos, formando las hipótesis; mientras que los indicadores se vinculan a los constructos. Se identifican dos tipos de vinculación. La primera donde se sostiene que los indicadores son el reflejo del constructo teórico no visualizado en el que están ligados, de forma que el constructo da paso a aquello que se observa. Por otro lado, es el que establece que los indicadores o medidas dan lugar al constructo. A los primeros se los define como indicadores reflectivos (efectos) mientras que a los segundos se los denomina indicadores formativos (causales) (J. Hair et al., 2010). En la **Tabla 1-1** se resumen las diferencias entre los indicadores reflectivos y formativos.

1.2.5. Escalas de medición

En términos sencillos, es la forma de evaluar a una variable o los valores permitidos para ella, estos pueden ser métricos (cuantitativos) y no métricos (cualitativos).

Las no métricas representan propiedades discretas, un ejemplo puede ser el sexo; aquí tenemos las escalas nominales, que consisten en asignar un número a cada grupo (por ejemplo en sexo 1=Mujer, 2=Hombre). Las escalas ordinales se puede ordenar y agrupar en función de las respuestas recibidas, pero dichos valores no se pueden considerar como cuantitativos (por ejemplo satisfacción del estudiante con las clases de sus profesores del semestre), por lo tanto no se pueden realizar más cálculos que los ya expuestos.

En las escalas métricas se utilizan para cantidad o grado, representan montos o magnitud; entre estas tenemos a las de intervalo y las de razón. En las escalas de intervalo no está incluido el cero arbitrario y no se pueden hacer operaciones entre variables de intervalo como por ejemplo la temperatura. Las medidas razón son el nivel más alto de precisión al medir y se pueden realizar todas las operaciones entre variables medidas en esta clase de escalas.

Tabla 1 - 1 Diferencias entre indicadores reflectivos y formativos

Característica	Indicativo de:	
	Reflexivo	Formativo
Causalidad del constructo	Los indicadores son causados por el constructo	El constructo es formado por los indicadores
Relación conceptual entre los indicadores	Todos los indicadores están relacionados conceptualmente porque tienen una causa común	No hay requisito de vinculación conceptual entre los indicadores
Dominio de los indicadores	Muestra representativa de los posibles indicadores	Inventario exhaustivo de todos los posibles indicadores
Covarianza entre los indicadores	Colinealidad esperada entre los indicadores	A priori no colinealidad esperada entre los indicadores
Consistencia interna	Requerida	No requerida
Formas de la validez de constructo	Interna y externa	Solo externa

Nota: Tomado del libro *Multivariante Data Analysis* por Hair, J., Black, W., Babin, B., & Anderson, R. (2010), p 751

1.2.6. Error de medición y medición multivariante

Es el grado en que las medidas no representan la realidad por múltiples factores, como por ejemplo, la percepción y disponibilidad para poder responder. Este error afecta los cálculos al momento de procesar la técnica y si el caso fuera de establecer una correlación entre dos variables esta se ve debilitada por la presencia del error de medición.

El error de medición refleja cierto ruido en la medición de las variables latentes (J. Hair et al., 2010). Por lo tanto, se debe asumir que todas las variables utilizadas en las técnicas multivariadas tienen cierto grado de error de medición. El error de las medidas añade un "ruido" a las variables observadas o medidas. Con ello, el valor que se obtiene representa tanto el nivel "verdadero" como el ruido.

Cuando se utiliza para calcular correlaciones, el efecto "verdadero" está parcialmente enmascarado por el error de medición, haciendo que las correlaciones se debiliten y los medios sean menos preciso.

1.2.7. Validación y confiabilidad de los datos

El objetivo del investigador es el de reducir el error de medición que se generan por varias fuentes (Demongeot, Hamie, Laksaci, & Rachdi, 2016). Al evaluar el grado del error de una medida presente en cualquiera de ellas, el investigador debe abordar dos características importantes de esta medición: primero su validez y luego su confiabilidad.

La validez de una medición es el grado en que una medida representa con precisión lo que se supone que es, por ejemplo, si queremos medir los ingresos discrecionales, no debemos preguntar acerca del ingreso total de un negocio.

Para garantizar la validez se debe partir de un conocimiento profundo de lo que se va a medir y luego hacer la medición como "correcta" y precisa como sea posible. Sin embargo, la precisión no garantiza la validez de la misma. En nuestro ejemplo de ingresos, el investigador podría definir con precisión el ingreso total del hogar, pero sería "incorrecto" (es decir, una medida inválida) al medir el ingreso discrecional porque no se estaba haciendo la pregunta "correcta".

Para que la validez de los datos esté garantizada, el investigador debe considerar la confiabilidad de las mediciones. La confiabilidad es el grado en que la variable observada mide el valor "verdadero" y es "libre de errores"; por lo tanto, es lo contrario del error de medición. La misma medida se hace repetidamente, por ejemplo, las medidas más fiables mostrarán una mayor coherencia que las medidas menos fiables.

El investigador siempre debe evaluar las variables que se están utilizando y, si existen otras medidas alternativas, elige la variable con mayor confiabilidad. La confiabilidad estadística es necesaria para garantizar la validez y precisión del análisis estadístico. Se relaciona con la capacidad de reproducir los resultados tantas veces como sea necesario. Esto es esencial, ya que genera confianza en el análisis estadístico y en los resultados obtenidos.

1.2.8. Significancia estadística versus potencia estadística

Las técnicas multivariantes están fundamentadas en la inferencia estadística entre los valores de una población o en la relación entre variables de una muestra escogida aleatoriamente de esa misma población, a excepción del análisis tipo cluster y el multidimensional.

La interpretación de las diferentes inferencias estadísticas está vinculado a los niveles de error aceptables en la investigación. Comúnmente se emplea el

nivel de error alfa (α) o conocido como tipo I, el cual consiste en rechazar la hipótesis nula cuando ésta es cierta (positivo falso).

Las ventajas de especificar un nivel alfa es que el investigador fija los márgenes admisibles de error especificando la probabilidad de concluir que la significación existe cuando en realidad no existe, así como también puede determina un error asociado. denominado el error de tipo II o beta (β) el cual consiste en la probabilidad de fallar en rechazar la hipótesis nula cuando es realmente falsa.

Una probabilidad más interesante es la tipo $1-\beta$, denominado “la potencia del test de inferencia estadística”, la cual es la probabilidad de rechazar correctamente la hipótesis nula cuando debe ser rechazada, por lo cual, la potencia es la probabilidad de que la inferencia estadística se indique cuando esté presente (Joseph F. Hair, Black, & Babin, 2010).

Cabe destacar, que aún cuando el error alfa establece el nivel de significación estadística aceptable, es solo el nivel de la potencia estadística el que indica la probabilidad de alcanzar el éxito en la búsqueda de las diferencias si en realidad existen. Ambos tipos de errores (alfa y beta) son inversamente proporcionales, en la medida que el tipo alfa se hace mas pequeño tendiendo a cero (siendo más restrictivo), el error de tipo beta aumenta (siendo más permisivo). Al mismo tiempo sucede que al disminuir el error de tipo alfa, también se reduce el poder de la prueba estadística, por tanto, es necesario conseguir un equilibrio entre el nivel de alfa y la potencia resultante (Joseph F. Hair et al., 2010).

1.3. Principales técnicas multivariante

Según Huir et., al (2010) la clasificación de las técnicas multivariantes se presenta en la **Figura 1 - 1**. Esta clasificación se basa en tres juicios que el analista debe hacer sobre el objeto a investigar y la naturaleza de los datos:

1. ¿Pueden dividirse las variables en dependientes o independientes basándose la clasificación en alguna teoría?
2. Si puede hacerse, ¿cuántas de estas variables son tratadas como dependientes en un análisis simple?
3. ¿Cómo son las variables medidas?

La selección de la técnica multivariante apropiada depende de las respuestas a estas tres cuestiones previas. Cuando consideramos la aplicación de las

técnicas estadísticas multivariantes, la primera cuestión que nos debemos preguntar es, ¿pueden dividirse las variables mediante la clasificación de dependiente e independiente? La respuesta a esta cuestión indica si se debería utilizar un análisis de dependencia o interdependencia.

Un análisis de dependencia puede definirse como aquel en el que una variable o conjunto de variables es identificado como la variable dependiente y que va a ser explicada por otras variables múltiples conocidas como variables independientes.

Como ejemplo de una dependencia técnica tenemos el análisis de regresión múltiple. Como contraste, un análisis de interdependencia es aquel en que ninguna variable o grupo de variables es definido como independiente o dependiente. Más bien, el procedimiento implica el análisis de todas las variables del conjunto simultáneamente. El análisis factorial es un ejemplo de un análisis de interdependencia (Joseph F. Hair et al., 2010).

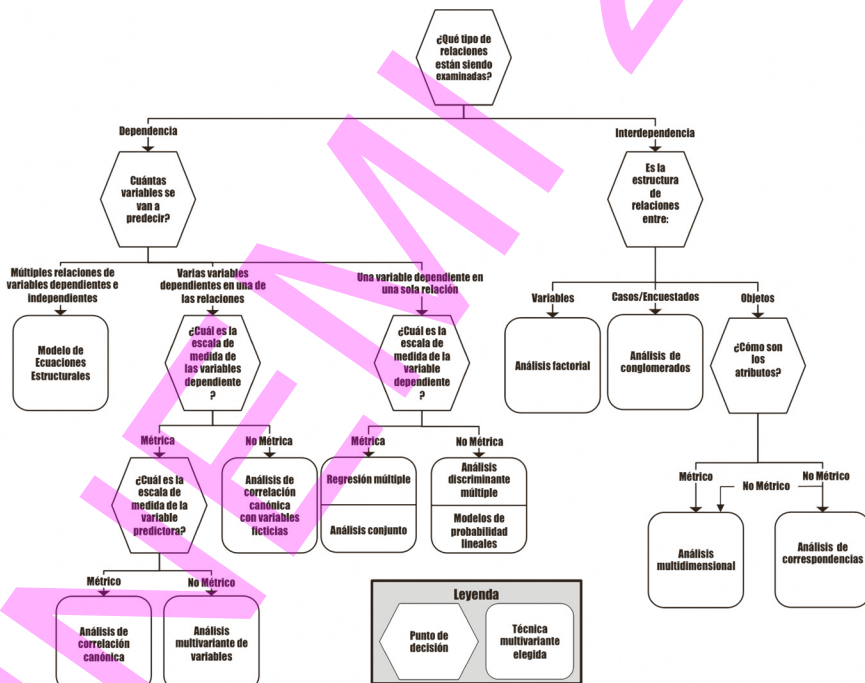


Figura 1 - 1 Selección de una técnica multivariante

Nota: Tomado del libro Multivariate Data Analysis por Hair, J., Black, W., Babin, B., & Anderson, R. (2010), p 31-32

1.3.1. Los componentes principales y análisis de factor común

Son un enfoque estadístico que se puede utilizar para estudiar las interrelaciones entre una gran cantidad de números de variables y exponer estas variables en términos de sus dimensiones subyacentes comunes. El objetivo es hallar una forma de concentrar la información almacenada en una serie de variables únicas en un conjunto más pequeño de variables circunstanciales (factores) con una pérdida mínima de información. Al proporcionar una apreciación empírica dentro de la distribución de las variables consideradas, el estudio factorial, se convierte en una base objetiva para la creación de escalas sumadas (J. Hair et al., 2010). En el capítulo 2 se explora en mayor profundidad esta técnica.

1.3.2. La regresión múltiple

Esta técnica evalúa como se relacionan las variables. Para hacerlo, se adicionan potencias a la ecuación de regresión y se observa si estas potencias mejoran de forma significativa la predicción (Pérez & Medrano, 2010).

Ye, Kang, Zuo y Zhong (2017) realizaron un estudio de los factores que afectan a la distancia de propagación del aire caliente en las salas de ventilación mediante un análisis de regresión múltiple. Estos investigadores reportaron la ecuación que vincula a los factores que influyen en la distancia de dispersión del chorro caliente en una ventilación siendo la distancia de propagación del aire caliente de un máximo y no mayor de 12 m para cualquier configuración de boquilla y condición de suministro. Además, la velocidad de flujo de aire de suministro debe ser mayor que un umbral y la altura y área de la boquilla también deben estar dentro de rangos particulares para satisfacer las necesidades especiales de la distancia de dispersión. Cuanto mayor sea la distancia de propagación requerida, más restricciones hay sobre las variables estudiadas. Con los modelos de regresión establecidos para la distancia de propagación del aire caliente, se determinó la combinación optimizada de estas variables para satisfacer el requisito de diseño.

1.3.3. El análisis discriminante múltiple

Es la técnica estadística apropiada cuando la variable dependiente es una variable categórica (nominal o no métrica). En muchos casos, la variable dependiente se compone de dos grupos o clasificaciones. El análisis discriminante es capaz de la manipulación de los dos grupos o múltiples (tres o más) grupos. Cuando dos clasificaciones están involucradas, la técnica se conoce como análisis discriminante de dos grupos. Cuando se identifican más de dos clasificaciones, la técnica se conoce como análisis discriminante

múltiple (J. Hair et al., 2010).

Como ejemplo: Shah, Patel, Nath y Menon (2016) llevaron a cabo la determinación del sexo mediante las dimensiones cefalo-faciales mediante el uso de ecuaciones de función discriminante y regresión logística. Basados en la antropometría, la cual es una herramienta esencial de la antropología biológica, que implica una serie de técnicas de medición estandarizadas, que expresan cuantitativamente las dimensiones del cuerpo humano como lo es la somatometría (disciplina de la antropometría) que trata de la medición del cuerpo vivo y del cadáver incluyendo la cabeza y la cara.

Estos investigadores validaron la función discriminante y los métodos de regresión logística que permitieron verificar el mejor nivel de precisión para la determinación del sexo. Reportando que los valores medios de las dimensiones cefalofaciales fueron mayores en los hombres que en las mujeres y que los mejores resultados se obtuvieron utilizando ecuaciones de regresión logística en varones (92%) y función discriminante en mujeres (80,9%).

1.3.4. La correlación canónica

Es una técnica estadística que se utiliza para analizar la relación entre múltiples variables dependientes (o endógenas) métricas y varias variables independientes (o exógenas) también métricas. El objetivo esencial del análisis de la correlación canónica es utilizar las variables independientes, las mismas que sus valores son conocidos, para predecir las variables criterios (dependientes) seleccionadas por el investigador. El análisis de la correlación canónica también puede extenderse al caso de variables dependientes no métricas y al caso de variables independientes no métricas (Alonso, 2008).

Fox y Hammond (2017) realizaron una investigación de la relación multivariada entre impulsividad y psicopatía mediante análisis de correlación canónica (ACC). Ellos concluyeron que los resultados obtenidos reflejan una compleja relación multidimensional entre la impulsividad y la psicopatía, y al mismo tiempo apoyan un modelo multifacético de psicopatía. Demostraron que el ACC es una técnica útil para explorar la relación compartida multivariante entre estos constructos.

1.3.5. El análisis multivariante de la varianza y covarianza (MANOVA)

Técnica estadística que se puede utilizar para explorar simultáneamente la relación entre varias variables independientes categóricas y dos o más variables métricas dependientes. Como tal, representa una extensión del análisis univariado de la varianza. El análisis multivariado de covarianza

se puede utilizar en conjunción con la covarianza para eliminar, una vez finalizado el experimento, el efecto de las variables independientes métricas no controladas (conocido como covariables) en la variable dependiente. El procedimiento es similar a la implicada en correlación parcial bivariada, en la que el efecto de una tercera variable se elimina de la correlación.

MANOVA es útil cuando el investigador esboza una situación experimental (empleo de distintas variables de tratamiento no métricas) para probar las hipótesis relativas a la variación en las respuestas del grupo en dos o más variables dependientes métricas (J. Hair et al., 2010).

La investigación realizada por Peruchi y colaboradores (Peruchi, Paiva, Balestrassi, Ferreira, & Sawhney, 2014) proponen nuevos índices basados en cuatro enfoques ponderados para el análisis multivariado de varianza en el análisis del sistema de medición. En la evaluación de la variación de un sistema de medición, la técnica más adecuada, una vez calibrado un instrumento, es la repetitividad y reproducibilidad de los indicadores (GR&R). Sin embargo, para evaluar los sistemas de medición multivariante, la discusión ha sido escasa.

Algunos investigadores han aplicado análisis de varianza multivariante para estimar los índices de evaluación; es por ello que la media geométrica se utiliza como una estrategia de aglutinación para los valores propios extraídos de matrices de varianza-covarianza.

Este enfoque, sin embargo, tiene algunas debilidades según estos autores los cuales reportaron que el análisis estadístico de los datos empíricos y de la literatura indica que la estrategia de ponderación más eficaz en los estudios (GR&R) multivariados se basa en una explicación de los porcentajes de los valores propios extraídos de la matriz de un sistema de medición. Al evaluar sistemas de medición con múltiples características correlacionadas, el método MANOVA ha demostrado algunos inconvenientes al estimar el índice de evaluación multivariante.

La principal contribución de esta investigación ha sido proponer nuevos índices multivariados basados en enfoques ponderados para superar tales inconvenientes (Peruchi et al., 2014).

1.3.6. El análisis conjunto

Técnica estadística utilizada para analizar la relación lineal o no lineal entre una variable dependiente (o endógena) generalmente ordinal (aunque también puede ser métrica) y varias variables independientes (o exógenas) no métricas (Alonso, 2008).

1.3.7. El análisis de conglomerados

Técnica estadística multivariante de clasificación automática de datos, que a partir de una tabla de casos-variables, trata de situar todos los casos en grupos similares (conglomerados o clústeres) desconocidos de antemano. Cabe tener en cuenta que los conglomerados se encuentran sugeridos por la propia esencia de los datos, de forma que individuos cuyas características puedan ser considerados similares, sean establecidos en un mismo clúster, mientras que individuos distintos se sitúen en clústeres diferentes (Alonso, 2008). El capítulo 3 aborda esta técnica en mayor profundidad.

1.3.8. Mapeo perceptual

Se refiere a un conjunto de técnicas que tratan de identificar la imagen percibida de un conjunto de objetos (empresas, productos, ideas, u otros artículos asociados con percepciones comúnmente aceptadas). El objetivo de cualquier método de asociación perceptiva es utilizar juicios de consumo de similitud o preferencia para representar los objetos o en un espacio multidimensional (J. Hair et al., 2010).

1.3.9. El análisis de correspondencias

Técnica descriptiva cuyo objetivo es resumir un gran número de datos en una cantidad reducida de dimensiones buscando la menor pérdida de información posible. Posee un objetivo de cierta forma similar a los métodos factoriales, sin embargo, en estos casos, el método se aplica a variables ordinales (De la Fuente, 2011).

1.3.10. El análisis factorial confirmatorio (CFA)

Es una técnica de análisis que se utiliza para contrastar los modelos de medida, dentro de un sistema de ecuaciones estructurales, que de forma hipotética suponen que una serie de variables observadas o indicadores miden una o más variables latentes (independientes o dependientes). Más adelante se expondrá en detalle en que consiste esta técnica. Para ello, se especifican las relaciones entre las variables observables o indicadores y las variables latentes a partir de unos coeficientes que en este caso se denominan cargas factoriales ("factor loading"), y que representan los efectos de las variables latentes o factores sobre los indicadores, es decir, demuestran en qué medida el factor (variable) se refleja en la puntuación del indicador (Cupani, 2012). El CFA también se utiliza para analizar la validez de constructo. El capítulo 3 aborda esta técnica con mayor detalle.

1.3.11. Modelo de ecuaciones estructurales

El modelo de ecuaciones estructurales (Structural Equation Modeling, SEM) se define como una técnica que combina la regresión múltiple y el análisis factorial. El investigador puede evaluar las complejas interrelaciones de dependencia, incluso, puede agregar los efectos causados por el error de medida que podría haber sobre los coeficientes estructurales simultáneamente (Verdugo, 2008). SEM basa su desarrollo en el modelo de regresión, análisis de senderos y análisis factorial. El modelo de regresión, basado en regresión lineal, utiliza un coeficiente de correlación y el criterio de mínimos cuadrados para calcular los pesos de regresión, Karl Pearson creó una fórmula para el coeficiente de correlación en 1896, que proporciona un índice de la relación entre dos variables (Schumacker & Lomax, 2010). En el capítulo 4 se detalla con ejemplos de aplicación esta técnica.

1.4. Selección de la técnica de análisis multivariante

La elección de la técnica multivariante depende de las respuestas a las tres preguntas expuestas en el apartado 1.3. Al considerar la aplicación de técnicas estadísticas multivariantes, la respuesta a la primera pregunta (¿Pueden las variables dividirse en categorías independientes y dependientes en base a alguna teoría?), indica si es una técnica de la dependencia o interdependencia la que debe ser utilizada. En la **Figura 1 - 1**, las técnicas de dependencia están en el lado izquierdo y las técnicas de la interdependencia están a la derecha.

Una técnica de dependencia se puede definir como una herramienta de análisis en la que una variable (o conjunto de variables) puede ser identificada como la variable dependiente que se predice o explica por otras variables conocidas como variables independientes. Un ejemplo de una técnica de dependencia es el análisis de regresión múltiple. Por el contrario, en una técnica de la interdependencia ninguna variable individual o grupo de variables se define como independiente o dependiente. El procedimiento consiste en el análisis simultáneo de todas las variables en el conjunto. El análisis factorial es un ejemplo de una técnica de interdependencia.

1.5. Análisis previo de los datos

Las técnicas multivariantes dan opciones potentes para el análisis al investigador, pero requieren de calidad significativa de los datos. Se debe revisar los datos previamente, para identificar faltantes y explorar relaciones entre las variables, es la primera e importante tarea. El análisis previo de los datos es un paso imprescindible antes del uso de una técnica multivariante, no hacerlo puede generar errores catastróficos (J. Hair et al., 2010).

Varias son las opciones de paquetes informáticos para el análisis previo de los datos, en la presente obra se realiza este proceso mediante los programas Microsoft Excel, IBM SPSS versión 21 para Mac y el programa R de licencia de uso libre.

El programa R, es ampliamente utilizado en investigación dado la gran cantidad de librerías que existen disponibles para tareas diversas en el procesamiento de los datos. En detalle al momento de realizar el análisis previo de los datos se realiza lo siguiente:

- Examen gráfico de la naturaleza de las variables a analizar y las relaciones que forman las bases del análisis,
- Proceso de evaluación para entender el impacto que pueden tener los datos ausentes sobre el análisis, y una serie de alternativas para casos reiterados de ausencia de datos en el análisis,
- Las técnicas que mejor se ajustan para la identificación de casos atípicos, aquellos casos que por su singularidad pueden distorsionar las relaciones sobre una o más variables analizadas, y
- Los métodos analíticos necesarios para evaluar adecuadamente la capacidad de los datos para cumplir los supuestos estadísticos específicos de muchas técnicas multivariantes.

1.6. Ejemplo de análisis previo de los datos

Con la finalidad de ilustrar los pasos que se dan en el análisis de datos previo a la aplicación de una técnica multivariante, se presenta un estudio realizado para evaluar la aceptación del sistema e-learning entre los profesores de una universidad estatal en Ecuador. La investigación se realizó entre junio y diciembre del 2015. La encuesta fue suministrada por correo electrónico. Para el efecto se utilizó una sistema de encuestas en línea, desarrollado por personal del departamento de Tecnologías de la Información y Comunicación de la misma institución.

El formulario aplicado en la investigación citada daba la oportunidad de responder cada pregunta de “1” total desacuerdo a “7” total acuerdo y se dejaba la opción de poner “0”, en caso de no conocer la respuesta. El instrumento de recolección de datos utilizado, fue el diseñado con la colaboración de expertos y se muestra en el **Anexo 1 - 1**. Se encuestó a 200 profesores, pero solo completaron satisfactoriamente el cuestionario un total de 145, con lo cual se obtuvo una tasa de respuesta del 73%. Para tener una idea de los profesores que integraron la muestra, en la **Tabla 1 - 2** se pueden apreciar las principales características demográficas.

Tabla 1 - 2 Datos demográficos de profesores encuestados

Características		Frecuencia	Porcentaje
Género	Masculino	83	57.24
	Femenino	62	42.76
Edad	De 20 a 30 años	17	11.72
	De 30 a 40 años	61	42.07
	Más de 40	67	46.21
Experiencia en internet	De 1 a 2 años	1	0.69
	Más de 2 años	144	99.31
Tiempo de dedicación	Medio	14	0.10
	Completo	131	0.90

1.6.1. Examen de los datos

Se realizó el examen gráfico de los datos con la finalidad de analizar la forma de la distribución, la relación entre variables, las diferencias entre grupos y los perfiles multivariantes.

Los datos recabados en la investigación descrita en el apartado anterior fueron almacenados en un archivo de texto delimitados por comas (formato CSV). El lector puede hacer uso de ellos con fines exclusivos de práctica, para el efecto debes descargar el archivo Ejercicio11_datos_docentes.csv desde el siguiente enlace <https://www.dropbox.com/sh/448ijhll1qovqj/AADYa8UTmaJFDNbpsSfvM7pZa?dl=0>. El procesamiento de estos datos terminó con una publicación en un congreso internacional (Ramirez-Anormaliza, Llinàs-Audet, & Sabaté, 2016).

Luego de haber descargado los datos, para el análisis previo de los mismos, ejecutar el software R-Studio, su instalación se expone en el **Anexo 1- 2**. En R-Studio ejecutar los siguientes pasos:

1. Menú **File**
2. Opción **New Project**
3. En el cuadro de diálogo dar clic sobre **Existing Directory**
4. Clic en **Browse** y elegir un directorio de trabajo
5. Clic en **Create Project**
6. Menú **File**
7. Clic **New File**
8. Clic **R Script**

Con los pasos anteriores se tiene acceso a un proyecto y archivo para ingresar sentencias R (script) en la interface de R-Studio, similar a la que se presenta en la **Figura 1 - 2**.

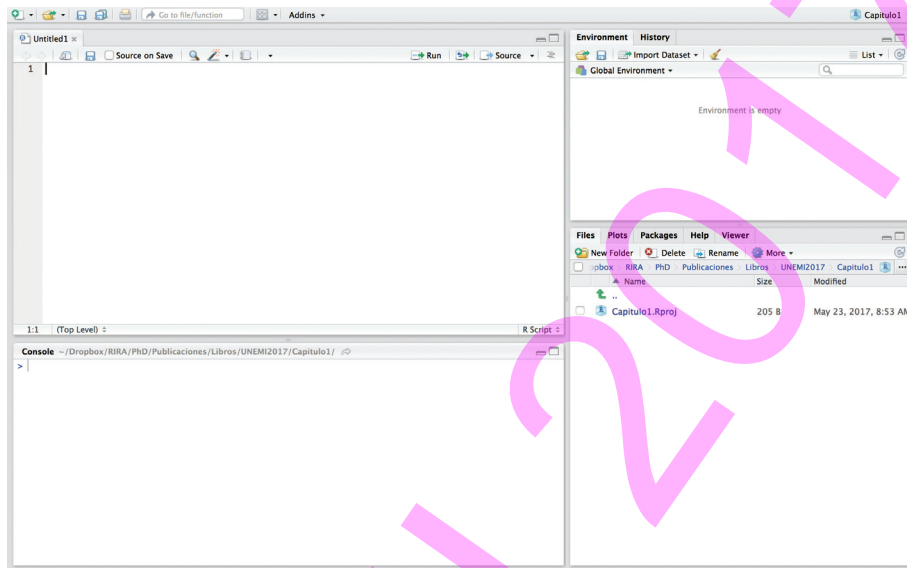


Figura 1 - 2 Interface de R-Studio

En el entorno de programación de R-Studio (ventana de la parte superior izquierda), se procede a ingresar las sentencias siguientes:

```
# (1) Establecer por defecto el directorio de trabajo donde está el archivo
de texto con los datos
setwd("/Users/imac/Dropbox/RIRA/PhD/Publicaciones/Libros/UNEMI2017/Datos")

# (2) Carga los datos de todos los profesores
Docentes <- read.csv("Ejercicio11_datos_docentes.csv", header = TRUE,
sep=";", dec=".", quote="\")

# (3) Toma solo los 52 items (no los datos demográficos)
Todos_Docentes <- Docentes [15:66]

# (4) Crea una función que convierte elementos 0 a NA
F_cerosToNA = funcion(x) { temp = NA; if (x != 0) temp=x; (temp)}

# (5) Convierte a NA las respuestas cuyo valor es 0 (sin respuesta) de esta
forma se dice a R los missing
DF_Docentes = data.frame(apply(Todos_Docentes, c(1,2), F_cerosToNA))

# (6) Instala el paquete para analizar los datos
install.packages(c("psych", "GPArotation", "MASS"))
library(psych)
```

```
# (7) La distribución, la relación entre variables, las diferencias entre
grupos
pairs.panels(DF_Docentes[1:6], jiggle=TRUE, factor=4) # Soporte tecnico
pairs.panels(DF_Docentes[7:15],jiggle=TRUE, factor=4) # Autoeficacia
computacional
pairs.panels(DF_Docentes[16:20],jiggle=TRUE, factor=4) # Influencia Social
pairs.panels(DF_Docentes[21:25],jiggle=TRUE, factor=4) # Disfrute percibido
pairs.panels(DF_Docentes[26:31],jiggle=TRUE, factor=4) # Satisfacción
pairs.panels(DF_Docentes[32:36],jiggle=TRUE, factor=4) # Utilidad percibida
pairs.panels(DF_Docentes[37:41],jiggle=TRUE, factor=4) # Facilidad de uso
percibida
pairs.panels(DF_Docentes[42:46],jiggle=TRUE, factor=4) # Intención de uso
pairs.panels(DF_Docentes[47:52],jiggle=TRUE, factor=4) # Uso del sistema
```

Se pueden ejecutar las sentencias línea por línea ó dando clic en la opción **Run**. Como resultado se obtienen las matrices de los gráficos de dispersión, en cada matriz se analizan los constructos con sus respectivos ítems o variables observables. En las **Figura 1 - 3** a la **Figura 1 - 12** se presentan las salidas respectivas.

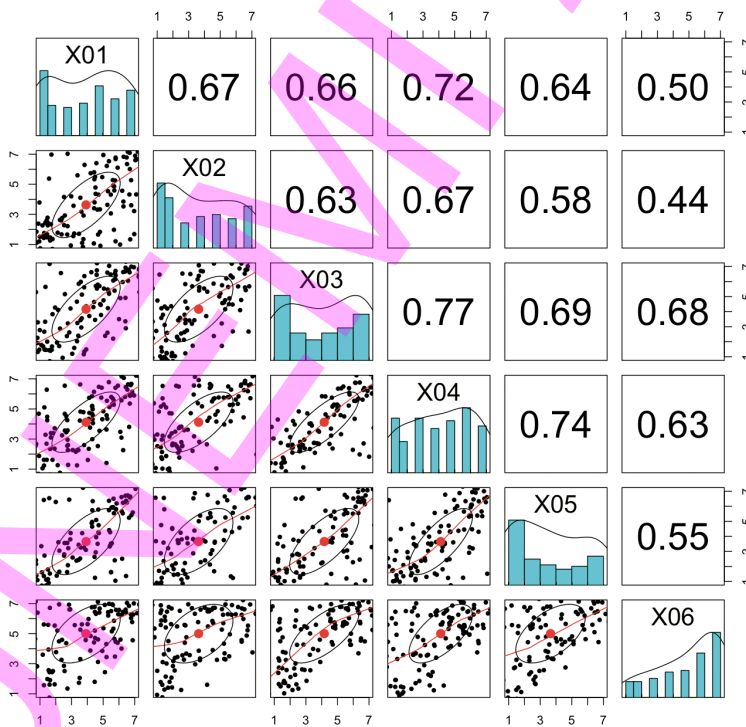


Figura 1 - 3 Matriz del gráfico de dispersión de Soporte Técnico

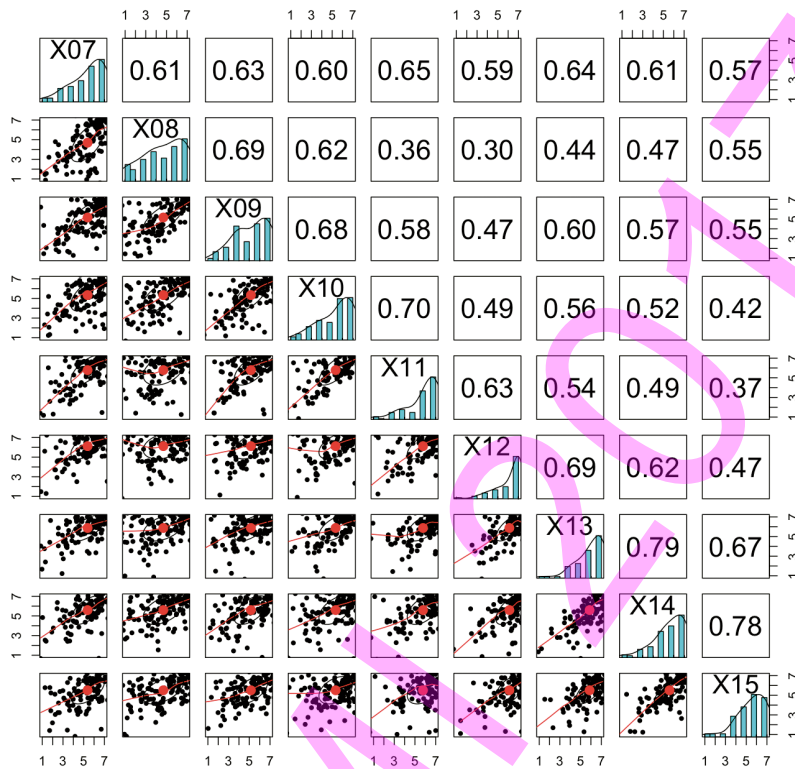


Figura 1 - 4 Matriz del gráfico de dispersión de Autoeficacia computacional

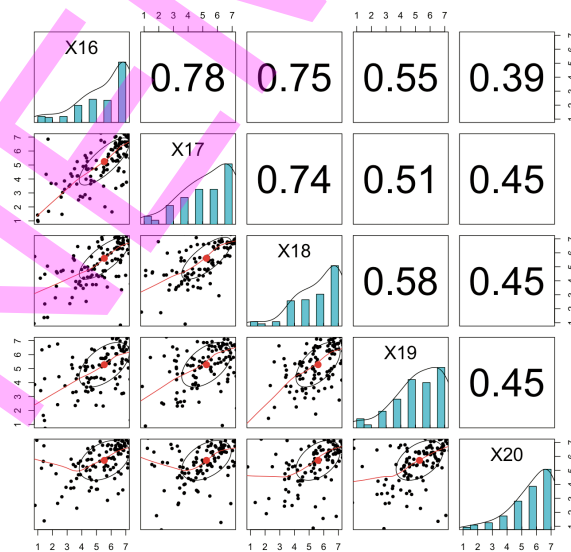


Figura 1 - 5 Matriz del gráfico de dispersión de Influencia Social

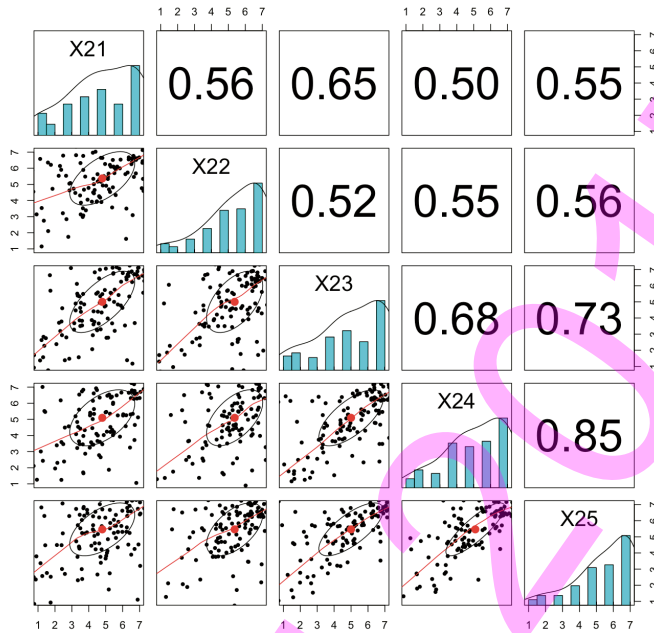


Figura 1 - 6 Matriz del gráfico de dispersión de disfrute percibido

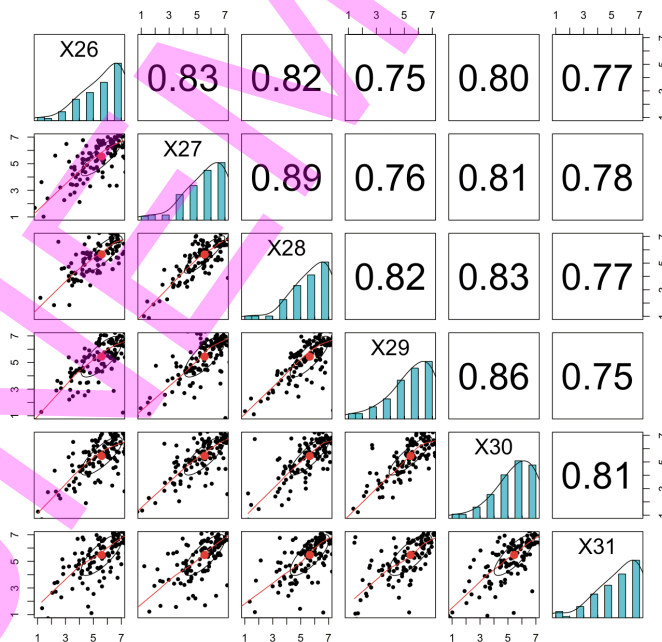


Figura 1 - 7 Matriz del gráfico de dispersión de satisfacción

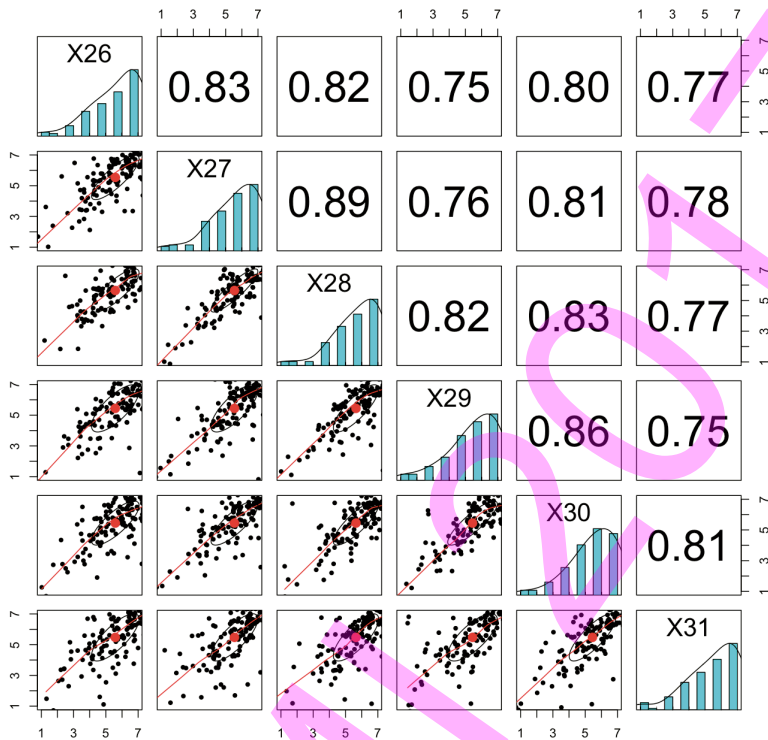


Figura 1 - 8 Matriz del gráfico de dispersión de Satisfacción

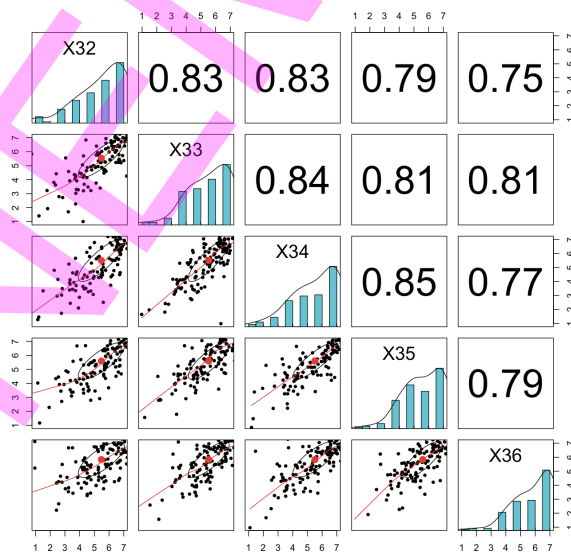


Figura 1 - 9 Matriz del gráfico de dispersión de Utilidad Percibida

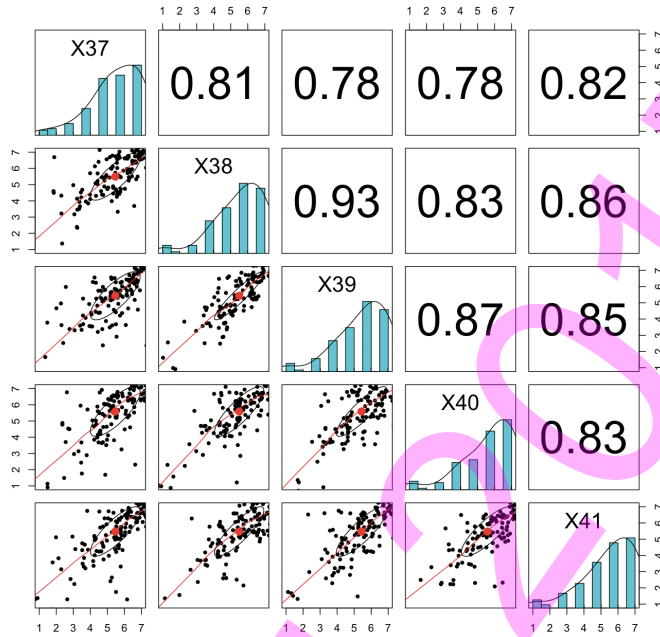


Figura 1 - 10 Matriz del gráfico de dispersión de facilidad de uso percibida

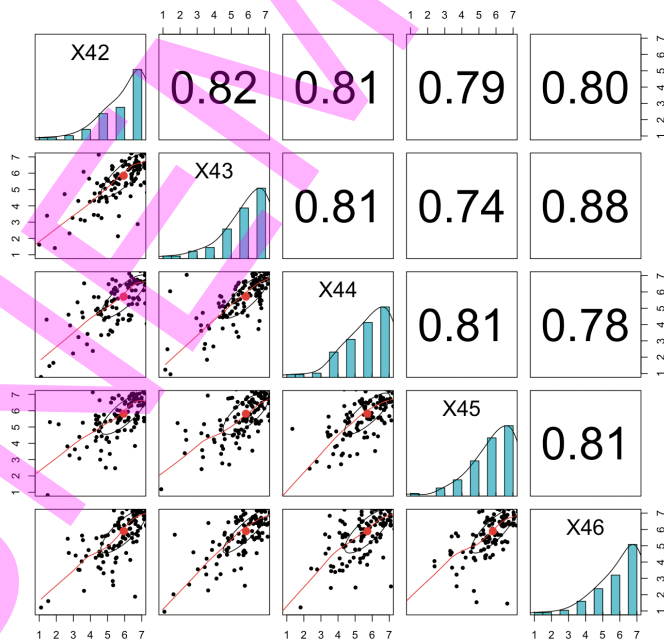


Figura 1 - 11 Matriz del gráfico de dispersión de intención de uso

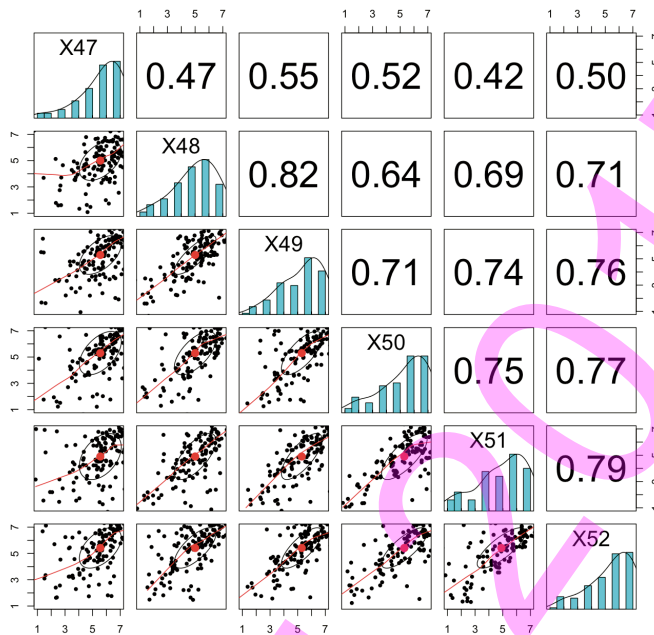


Figura 1 - 11 Matriz del gráfico de dispersión de intención de uso

1.6.2. Datos ausentes

Por la opción dada de colocar “0” en caso de no conocer la respuesta, la muestra de los profesores presentó ausencia de datos. Para diagnosticar las pautas de los datos ausentes se elaboró la **Tabla 1 - 3**, que contiene estadísticas descriptivas de los casos válidos, los porcentajes con datos ausentes en cada variable. Se obtuvo una ausencia global del 4.20% y 105 casos sin datos ausentes.

Tabla 1 - 3 Estadísticas resumen de los datos previos

Variables	Casos con datos válidos	Media	Desviación estándar	Datos ausentes	
				Número	Porcentaje
X01	133	3.92	2.164	12	8.3
X02	132	3.63	2.177	13	9.0
X03	128	4.16	2.191	17	11.7
X04	134	4.12	2.011	11	7.6
X05	121	3.65	2.201	24	16.6
X06	134	5.00	1.951	11	7.6

Continúa

Tabla 1 - 3 Continuación

Variables	Casos con datos válidos	Media	Desviación estándar	Datos ausentes	
				Número	Porcentaje
X07	141	5.38	1.615	4	2.8
X08	140	4.69	1.945	5	3.4
X09	141	5.16	1.614	4	2.8
X10	139	5.34	1.618	6	4.1
X11	141	5.78	1.526	4	2.8
X12	139	6.12	1.305	6	4.1
X13	138	5.87	1.323	7	4.8
X14	137	5.59	1.458	8	5.5
X15	137	5.51	1.394	8	5.5
X16	139	5.53	1.712	6	4.1
X17	140	5.25	1.688	5	3.4
X18	141	5.61	1.506	4	2.8
X19	140	5.28	1.601	5	3.4
X20	140	5.71	1.426	5	3.4
X21	136	4.79	1.921	9	6.2
X22	138	5.37	1.661	7	4.8
X23	136	5.00	1.905	9	6.2
X24	136	5.10	1.738	9	6.2
X25	138	5.46	1.640	7	4.8
X26	142	5.58	1.474	3	2.1
X27	142	5.54	1.452	3	2.1
X28	142	5.65	1.405	3	2.1
X29	142	5.44	1.537	3	2.1
X30	141	5.45	1.436	4	2.8
X31	141	5.48	1.538	4	2.8
X32	141	5.48	1.570	4	2.8
X33	140	5.54	1.391	5	3.4
X34	140	5.51	1.510	5	3.4
X35	139	5.60	1.317	6	4.1
X36	142	5.83	1.283	3	2.1
X37	142	5.46	1.462	3	2.1
X38	142	5.49	1.453	3	2.1
X39	141	5.43	1.485	4	2.8

Continúa

Tabla 1 - 3 Continuación

Variables	Casos con datos válidos	Media	Desviación estándar	Datos ausentes	
				Número	Porcentaje
X40	142	5.58	1.536	3	2.1
X41	142	5.47	1.528	3	2.1
X42	141	5.94	1.335	4	2.8
X43	141	5.85	1.336	4	2.8
X44	141	5.72	1.327	4	2.8
X45	141	5.82	1.268	4	2.8
X46	140	5.89	1.339	5	3.4
X47	142	5.56	1.518	3	2.1
X48	141	5.00	1.507	4	2.8
X49	140	5.29	1.472	5	3.4
X50	141	5.30	1.651	4	2.8
X51	136	4.90	1.755	9	6.2
X52	139	5.42	1.537	6	4.1

De los 145 profesores que llenaron la encuesta, 14 casos presentaron datos ausentes en porcentajes del 12% hasta el 90%. Por lo tanto, dicho casos fueron omitidos del estudio. Un resumen de estos casos se presenta en la **Tabla 1 - 4**.

Tabla 1 - 4 Resumen de casos con datos ausentes mayor 10%

Caso	Variables con datos ausentes	Porcentaje de datos ausentes
52	28	54%
56	7	13%
58	47	90%
59	33	63%
63	44	85%
72	16	31%
74	12	23%
101	9	17%
115	6	12%
117	11	21%
120	8	15%
121	6	12%
128	22	42%
132	7	13%

Respecto a los ítems (variables), luego de eliminados los casos que presentaron datos ausentes en un nivel mayor al 10%, aún quedaron ítems con datos ausentes en un porcentaje moderado, los mismos se reflejan en la **Tabla 1 - 5**.

Tabla 1 - 5 Constructos con datos ausentes menor al 10%

Constructo	Variable	Casos	Porcentaje
Soporte Técnico	X01	3	2.29
	X02	7	5.34
	X03	8	6.11
	X04	2	1.53
	X05	13	9.92
	X06	3	2.29
Autoeficacia computacional	X08	1	0.76
	X12	2	1.53
	X13	1	0.76
	X14	2	1.53
Influencia social	X16	4	3.05
	X17	2	1.53
	X18	2	1.53
	X19	3	2.29
	X20	2	1.53
Disfrute percibido	X21	2	1.53
	X24	1	0.76
	X25	1	0.76
Intención de uso	X46	1	0.76
Uso del sistema	X51	1	0.76

Se procedió a realizar un análisis exhaustivo de la **Tabla 1 - 5** por constructo. En el caso del Soporte Técnico de los 6 ítems, uno tiene el nivel de datos ausentes del 9%, es el ítem X05, corresponde a *“Se pueden hacer consultas en línea tipo chat con personas de soporte técnico cuando hay algún problema con el sistema e-learning”*. Por el porcentaje de datos ausentes y considerando que hay 5 ítems más en este constructo, se decidió eliminar este indicador del cuestionario final.

Para el caso de la Autoeficacia Computacional se tenían 9 ítems, de los cuales 4 resultaron con 1% de datos ausentes, dado que estos ítems ya estaban siendo evaluados en los que no reportaron ausencia de datos, se decidió eliminar los ítems X08, X12, X13, X14 en formulario final.

En los constructos Influencia Social, Disfrute percibido e Intención de uso; se identificaron 5, 3 y 1 ítems con datos ausentes respectivamente. Se decidió mantener a todos ítems al proceso siguiente por su bajo nivel de ausencia de datos y para no disminuir el número de variables observables de los respectivos constructos.

En el uso del sistema, de los 6 ítems considerados en este constructo, el X51 presentó ausencia de datos del 1%, no es alta la ausencia, pero considerando que el aspecto que evalúa el ítem X51 ya se contempla en otros. Se decidió eliminarlo de la encuesta final.

Con los ajustes antes citados se obtuvo un formulario de 9 constructos con 5 ítems cada uno, a excepción del constructo Satisfacción que hasta este punto permanece con 6 indicadores. Por lo tanto, hasta este punto se tenían 46 ítems. Para dejar 5 variables en el constructo Satisfacción y considerando, que dicho constructo está constituido por ítems reflectivos; luego de realizar análisis de fiabilidad, se determina que si se elimina el ítem X31, esto nos daba un alfa de Cronbach de 0.957 como lo muestra la **Tabla 1 - 6**, lo cual es muy aceptable, con esta acción se llegó a un formulario de encuesta de 45 ítems agrupados en 9 constructos. El análisis de fiabilidad expuesto se realizó en el software SPSS.

Tabla 1 - 6 Análisis de fiabilidad del constructo satisfacción

Ítem	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
X26	27.51	46.180	0.864	0.955
X27	27.56	46.075	0.891	0.952
X28	27.46	46.394	0.907	0.950
X29	27.64	45.512	0.855	0.956
X30	27.65	46.186	0.897	0.951
X31	27.60	45.738	0.848	0.957

1.6.3. Aleatoriedad de los datos ausentes

Se debe verificar la aleatoriedad de los datos ausentes para decidir si se realiza la imputación de datos en función de los ítems que si se contestaron.

Con la finalidad de tener mayor elemento de juicio respecto a los datos ausentes, se realizó el test MCAR (Missing completely at random) (Little, 1988) mediante al paquete BaylorEdPsych (Beaujean, 2015) del programa estadístico R.

El análisis se realizó a dos muestras, primero a los 145 casos y luego al subconjunto de este, donde se dejaron solo aquellos casos que tenía menos de 10% de datos ausentes que fueron 131 casos de los 145 iniciales. Estos resultados se ilustran en la **Tabla 1 - 7**.

Tabla 1 - 7 Prueba multivariante ausencia completamente aleatoria (MCAR)

Muestra	N	Chi-cuadrado	GI	P
Toda la muestra	145	1980.2440	1611	0.000000
Datos ausentes < 10%	131	887.4677	747	0.000285

El nivel de significancia de la prueba MCAR para cada una de las muestras fue de 0.000000 y 0.000285; al ser menores a 0.05, se concluye que los datos ausentes no se producen completamente aleatoria.

Pero como el nivel global de datos ausentes es menor al 10%, la técnica sugiere que se puede aplicar cualquier método de imputación de datos ausentes (J. Hair et al., 2010).

1.6.4. Imputación de datos

Se realizó la corrección de los datos ausentes a los 131 casos que conforman la muestra que resultó del proceso descrito hasta apartado anterior, para el efecto se evaluaron tres métodos; el método de regresión e imputación (EM), el de reemplazo por media de los datos presentes y el de imputación por regresión.

Tabla 1 - 8 Estadísticos univariantes, datos sin imputar e imputados por EM, media y regresión

Vars.	Sin imputar					Por EM				Por media		Regresión			
	N	Mín	Máx	Media	Desv. típ.	Mín	Máx	Media	Desv. típ.	Media	Desv. típ.	Mín	Máx	Media	Desv. típ.
TS1	128	1	7	3.93	2.163	0*	7	3.89	2.166	3.93	2.138	1	7	3.92	2.143
TS2	124	1	7	3.62	2.147	-1*	7	3.60	2.180	3.62	2.089	-2**	7	3.60	2.179
TS3	123	1	7	4.23	2.165	1	8*	4.23	2.156	4.23	2.097	1	7	4.21	2.143
TS4	129	1	7	4.14	2.011	1	7	4.13	2.012	4.14	1.995	1	7	4.14	1.999
TS5	128	1	7	4.98	1.930	1	8*	4.97	1.940	4.98	1.907	1	9**	4.97	1.964
CSE1	131	1	7	5.40	1.568	1	7	5.40	1.568	5.40	1.568	1	7	5.40	1.568
CSE2	131	1	7	5.16	1.616	1	7	5.16	1.616	5.16	1.616	1	7	5.16	1.616
CSE3	131	1	7	5.29	1.634	1	7	5.29	1.634	5.29	1.634	1	7	5.29	1.634
CSE4	131	1	7	5.74	1.552	1	7	5.74	1.552	5.74	1.552	1	7	5.74	1.552
CSE5	131	1	7	5.54	1.343	1	7	5.54	1.343	5.54	1.343	1	7	5.54	1.343
SI1	127	1	7	5.44	1.749	1	8*	5.49	1.744	5.44	1.722	1	8**	5.48	1.746
SI2	129	1	7	5.20	1.693	1	8*	5.21	1.697	5.20	1.680	1	7	5.20	1.685
SI3	129	1	7	5.60	1.486	1	7	5.62	1.479	5.60	1.475	1	7	5.61	1.478
SI4	128	1	7	5.20	1.623	1	7	5.20	1.634	5.20	1.605	1	7	5.23	1.613
SI5	129	1	7	5.64	1.451	1	7	5.65	1.442	5.64	1.440	1	7	5.64	1.445
PE1	129	1	7	4.85	1.854	1	7	4.86	1.843	4.85	1.840	1	7	4.86	1.843
PE2	131	1	7	5.40	1.635	1	7	5.40	1.635	5.40	1.635	1	7	5.40	1.635
PE3	131	1	7	5.00	1.905	1	7	5.00	1.905	5.00	1.905	1	7	5.00	1.905
PE4	130	1	7	5.12	1.728	1	7	5.09	1.756	5.12	1.722	0**	7	5.08	1.769
PE5	130	1	7	5.48	1.648	1	7	5.45	1.679	5.48	1.642	1	7	5.45	1.661
S1	131	1	7	5.60	1.435	1	7	5.60	1.435	5.60	1.435	1	7	5.60	1.435
S2	131	1	7	5.56	1.421	1	7	5.56	1.421	5.56	1.421	1	7	5.56	1.421
S3	131	1	7	5.67	1.367	1	7	5.67	1.367	5.67	1.367	1	7	5.67	1.367
S4	131	1	7	5.47	1.510	1	7	5.47	1.510	5.47	1.510	1	7	5.47	1.510
S5	131	1	7	5.47	1.383	1	7	5.47	1.383	5.47	1.383	1	7	5.47	1.383
PU1	131	1	7	5.52	1.521	1	7	5.52	1.521	5.52	1.521	1	7	5.52	1.521
PU2	131	1	7	5.59	1.324	1	7	5.59	1.324	5.59	1.324	1	7	5.59	1.324
PU3	131	1	7	5.55	1.479	1	7	5.55	1.479	5.55	1.479	1	7	5.55	1.479
PU4	131	1	7	5.62	1.304	1	7	5.62	1.304	5.62	1.304	1	7	5.62	1.304
PU5	131	1	7	5.82	1.288	1	7	5.82	1.288	5.82	1.288	1	7	5.82	1.288
PEOU1	131	1	7	5.50	1.405	1	7	5.50	1.405	5.50	1.405	1	7	5.50	1.405
PEOU2	131	1	7	5.50	1.438	1	7	5.50	1.438	5.50	1.438	1	7	5.50	1.438
PEOU3	131	1	7	5.46	1.448	1	7	5.46	1.448	5.46	1.448	1	7	5.46	1.448
PEOU4	131	1	7	5.63	1.458	1	7	5.63	1.458	5.63	1.458	1	7	5.63	1.458
PEOU5	131	1	7	5.45	1.505	1	7	5.45	1.505	5.45	1.505	1	7	5.45	1.505
BI1	131	1	7	5.92	1.339	1	7	5.92	1.339	5.92	1.339	1	7	5.92	1.339
BI2	131	1	7	5.82	1.350	1	7	5.82	1.350	5.82	1.350	1	7	5.82	1.350
BI3	131	1	7	5.72	1.337	1	7	5.72	1.337	5.72	1.337	1	7	5.72	1.337
BI4	131	1	7	5.79	1.257	1	7	5.79	1.257	5.79	1.257	1	7	5.79	1.257
BI5	130	1	7	5.87	1.343	1	7	5.87	1.338	5.87	1.338	1	7	5.87	1.338
SU1	131	1	7	5.49	1.521	1	7	5.49	1.521	5.49	1.521	1	7	5.49	1.521
SU2	131	1	7	4.98	1.483	1	7	4.98	1.483	4.98	1.483	1	7	4.98	1.483
SU3	131	1	7	5.25	1.485	1	7	5.25	1.485	5.25	1.485	1	7	5.25	1.485
SU4	131	1	7	5.24	1.669	1	7	5.24	1.669	5.24	1.669	1	7	5.24	1.669
SU5	131	1	7	5.37	1.561	1	7	5.37	1.561	5.37	1.561	1	7	5.37	1.561

Nota: N válidos 110, *Datos imputados fuera de rango por EM, **Datos imputados fuera de rango por regresión

En la **Tabla 1 - 8** se resumen los estadísticos descriptivos univariados que resultaron de los tres métodos de imputación aplicados, en ella se puede identificar que en los métodos EM y de regresión calcularon valores fuera de los rangos esperados (1 a 7). Seis variables presentaron este inconveniente en la imputación por EM, mientras que por regresión fueron cuatro las variables que afectada por la misma situación.

También nos podemos dar cuenta que los estadísticos descriptivos de los datos de los tres métodos de imputación, en algunos son los mismos o tienen una ligera variación entre ellos o respecto a los datos sin imputar. Por lo tanto, luego de corregir los valores fuera de rango que resultaron de las imputaciones, cualquiera de los datos imputados según Hair et al., (2010) puede ser utilizado.

Para finalizar este capítulo, se consideró importante dar a conocer trabajos de investigación publicados en revistas indexadas donde se aplicaron varias de las técnicas multivariantes brevemente descritas. Estas publicaciones se resumen en la **Tabla 1 - 9**.

Tabla 1 - 9 Publicaciones con aplicación de técnicas multivariante

Técnica	Título	Fuente
Regresión Múltiple	<ul style="list-style-type: none"> • <i>Substituting random forest for multiple linear regression improves binding affinity prediction of scoring functions: Cyscore as a case study</i> 	(Li, Leung, Wong, & Ballester, 2014)
	<ul style="list-style-type: none"> • <i>A prediction method of ice breaking resistance using a multiple regression analysis</i> 	(Cho & Lee, 2015)
	<ul style="list-style-type: none"> • <i>QSAR models for prediction study of HIV protease inhibitors using support vector machines, neural networks and multiple linear regression</i> 	(Darnag, Minaoui, & Fakir, 2017)
	<ul style="list-style-type: none"> • <i>Model selection with multiple regression on distance matrices leads to incorrect inferences</i> 	(Franckowiak et al., 2017)
Modelo de ecuaciones estructurales	<ul style="list-style-type: none"> • <i>Relationships between structure, process and outcome to assess quality of integrated chronic disease management in a rural South African setting: applying a structural equation model</i> 	(Ameh, Gómez-Olivé, Kahn, Tollman, & Klipstein-Grobusch, 2017)
	<ul style="list-style-type: none"> • <i>Comparing Indirect Effects in Different Groups in Single-Group and Multi-Group Structural Equation Models</i> 	(Ryu & Cheong, 2017)

Continúa

Tabla 1 - 9 Continuación

Técnica	Título	Fuente
Modelo de ecuaciones estructurales	<i>Parenting Style, Parental Feeding Practices and Children's nutritional Status in Authoritative Parenting Style Model: A Structural Equation Modeling</i>	(Morowatisharifabad et al., 2017)
	<i>Urban Competitiveness Measurement of Chinese Cities Based on a Structural Equation Model</i>	(Yuan, Zheng, Zhang, & Zhao, 2017)
	<i>Accentance and use e-learning systems by ungraduate students of Ecuador: The case of a state University</i>	(Ramirez-Anormaliza, Sabaté, Llinàs-Audet, & Lordan, 2017)
	<i>The acceptance and use of the e-learning systems among the university teachers in Ecuador</i>	(Ramirez-Anormaliza et al., 2016)
Análisis factorial confirmatorio	<i>Exploratory and confirmatory factor analysis of the Roteiro de Avaliação da Consciência Fonológica, a phonological awareness test</i>	(Aparecida & Thatiana, 2017)
	<i>Psychometric properties of the SDM-Q-9 questionnaire for shared decision-making in multiple sclerosis: item response theory modelling and confirmatory factor analysis</i>	(Ballesteros et al., 2017)
	<i>Erratum to: Confirmatory factor analysis of Clinical Outcomes in Routine Evaluation (CORE-OM) used as a measure of emotional distress in people with tinnitus</i>	(Handscomb, Hall, Hoare, & Shorter, 2016)
	<i>Measuring implementation behaviour of menu guidelines in the childcare setting: confirmatory factor analysis of a theoretical domains framework questionnaire (TDFQ)</i>	(Seward et al., 2017)
	<i>Confirmatory Factor Analysis of the Combined Social Phobia Scale and Social Interaction Anxiety Scale: Support for a Bi factor Model</i>	(Gomez & Watson, 2017)
Análisis de correspondencia	<i>The Use of Correspondence Analysis in Assessing the Antecedents of Innovativeness in Public Hospitals</i>	(Jończyk & Olszewska, 2016)
	<i>Association knowledge for fatal run-off-road crashes by Multiple Correspondence Analysis</i>	(Das & Sun, 2016)
	<i>Identification of areas of intervention for public safety policies using multiple correspondence analysis</i>	(Carrillo, Rubio, & ..., 2016)
	<i>Analysis of information disclosure for chilean case using multiple correspondence methodology</i>	(Palavecinos, Espinoza, Mella, & Rosso, 2016)

Continúa

Tabla 1 - 9 Continuación

Técnica	Título	Fuente
Análisis de correspondencia	<i>Genome Data Exploration Using Correspondence Analysis</i>	(Tekai, 2014)
	<i>Study on homophobic bullying in Portugal using Multiple Correspondence Analysis (MCA)</i>	(Rodrigues, Grave, de Oliveira, & Nogueira, 2016)
Mapeo perceptual	<i>A model of perceptual distance for group delays based on ellipsoidal mapping</i>	(Banno, Takeda, & Itakura, 2002)
	<i>Understanding help-seeking intentions in male military cadets: An application of perceptual mapping</i>	(Bass, Muñoz, Gordon, Maurer, & Patterson, 2016)
Análisis de Conglomerado	<i>Combinación de métodos factoriales y de análisis de conglomerados en R: el paquete FactoClass</i>	(Pardo & del Campo, 2007)
	<i>Génesis: cluster analysis of microarray data</i>	(Sturn, Quackenbush, & Trajanoski, 2002)
	<i>Cluster analysis: a simple, versatile, and robust algorithm for endocrine mlse detection</i>	(Veldhuis & Johnson, 1986)
	<i>Análisis de Conglomerados</i>	(Aldás Manzano, 2011)
Análisis discriminante múltiple	<i>A Complete Subspace Analysis of Linear Discriminant Analysis and Its Robust Implementation</i>	(Lu & Liang, 2016)
	<i>Potential of classification of cáncer by multiple discriminant analysis for relationship between cáncer and expression of human cellular phosphoprotein</i>	(Motofuji et al., 2012)
	<i>Semisupervised Tangent Space Discriminant Analysis</i>	(Zhou & Sun, 2015)
	<i>Distributed Face Recognition Using Multiple Kernel Discriminant Analysis in Wireless Sensor Networks</i>	(Liu & Yang, 2014)
	<i>Assessment applicability of selected models of multiple discriminant analyses to forecast financial situation of Polish wood sector enterprises.</i>	(Adamowicz & Noga, 2017)
Correlación canónica	<i>Sparse canonical correlation analysis for identifying, connecting and completing gene-expression networks</i>	(Waaaijenborg & Zwinderman, 2009),
	<i>Identification of cell cycle-related regulatory motifs using a kernel canonical correlation analysis</i>	(Rhee, Joung, Chang, Fei, & Zhang, 2009)

Continúa

Tabla 1 - 9 Continuación

Técnica	Título	Fuente
Correlación canónica	<i>Sparse kernel canonical correlation analysis for discovery of nonlinear interactions in high-dimensional data</i>	(Yoshida, Yoshimoto, & Doya, 2017)
	<i>Non-linear canonical correlation for joint analysis of MEG signals from two subjects</i>	(Campi, Parkkonen, Hari, & Hyvarinen, 2013)
	<i>Time-series alignment by non-negative multiple generalized canonical correlation analysis</i>	(Fischer, Roth, & Buhmann, 2007)
	<i>Canonical correlations between agronomic traits and seed physiological quality in segregating soybean populations</i>	(Pereira et al., 2017)
Análisis multivariante de la varianza y covarianza (MANOVA)	<i>Identification of differentially expressed subnetworks based on multivariate ANOVA</i>	(Hwang & Park, 2009)
	<i>Multivariate variance-components analysis of longitudinal blood pressure measurements from the Framingham Heart Study</i>	(Kraft, Bauman, Yuan, Horvath, & Study, 2003)
	<i>On generalized multivariate analysis of variance</i>	(Díaz-García, 2011)
	<i>MANOVA for Nested Designs with Unequal Cell Sizes and Unequal Cell Covariance Matrices</i>	(Xu, 2014)
Análisis conjunto	<i>Robust Joint Analysis with Data Fusion in Two-Stage Quantitative Trait Genome-Wide Association Studies</i>	(Pan et al., 2013)
	<i>Comparison of bivariate and multivariate joint analyses on the selection loss of beef cattle</i>	(Pedrosa, Groeneveld, Eler, & Ferraz, 2014)
	<i>Joint Analysis for Genome-Wide Association Studies in Family-Based Designs</i>	(Sha, Zhang, & Zhang, 2011)
	<i>Joint analysis of multiple high-dimensional data types using sparse matrix approximations of rank-1 with applications</i>	(Okimoto et al., 2016)
	<i>A joint analysis of influenza-associated hospitalizations and mortality In Hong Kong, 1998–2013</i>	(P. Wu et al., 2017)

1.7. Referencias

- Albuissou, E., Allain, Y. M., Rozan, R., Giraud, B., Bolla, M., Brune, D., ... Richaud, P. (1992). Prostate cancer treated by radiotherapy: a multivariate study. *Radiotherapy and Oncology*, 25(2), 89–96. [http://doi.org/10.1016/0167-8140\(92\)90014-L](http://doi.org/10.1016/0167-8140(92)90014-L)
- Aldás Manzano, J. (2011). *Análisis De Conglomerados*. Universitat de València.
- Alonso, A. (2008). *Técnicas de análisis multivariante de datos*. (I. Capella & M. Diego, Eds.) (Primera). Madrid-España: PEARSON PRENTICE HALL.
- Ameh, S., Gómez-Olivé, F. X., Kahn, K., Tollman, S. M., & Klipstein-Grobusch, K. (2017). Relationships between structure, process and outcome to assess quality of integrated chronic disease management in a rural South African setting: applying a structural equation model. *BMC Health Services Research*, 17(1), 229. <http://doi.org/10.1186/s12913-017-2177-4>
- Amiry, S., Esmaili, M., & Alizadeh, M. (2017). Classification of adulterated honeys by multivariate analysis. *Food Chemistry*, 224, 390–397. <http://doi.org/10.1016/j.foodchem.2016.12.025>
- Aparecida, A., & Thatiana, H. (2017). Fonológica, a phonological awareness test Análise fatorial exploratória e confirmatória do Roteiro de Avaliação da Consciência Fonológica. *Estudos de Psicologia Campina*, 34(2), 211–218.
- Arteaga, R., & Duarte, A. (2010). Motivational factors that influence the acceptance of Moodle using TAM. *Computers in Human Behavior*, 26(6), 1632–1640. <http://doi.org/10.1016/j.chb.2010.06.011>
- Ballesteros, J., Moral, E., Brieva, L., Ruiz-Beato, E., Prefasi, D., & Maurino, J. (2017). Psychometric properties of the SDM-Q-9 questionnaire for shared decision-making in multiple sclerosis: item response theory modelling and confirmatory factor analysis. *Health and Quality of Life Outcomes*, 15(1), 79. <http://doi.org/10.1186/s12955-017-0656-2>
- Banno, H., Takeda, K., & Itakura, F. (2002). A model of perceptual distance for group delays based on ellipsoidal mapping. *Acoust. Sci. & Tech*, 2, 110–112.
- Bass, S. B., Muñoz, J., Gordon, T. F., Maurer, L., & Patterson, F. (2016). Understanding help-seeking intentions in male military cadets: An application of perceptual mapping. *BMC Public Health*, 16(1), 413. <http://doi.org/10.1186/s12889-016-3092-z>
- Beaujean, A. A. (2015). Package “BaylorEdPsych.” CRAN.
- Campi, C., Parkkonen, L., Hari, R., & Hyvarinen, A. (2013). Non-linear canonical correlation for joint analysis of MEG signals from two subjects. *Frontiers in Neuroscience*, 7(7 JUN), 1–7. <http://doi.org/10.3389/fnins.2013.00107>
- Carrillo, J., Rubio, J., & ... (2016). Identification of areas of intervention

- for public safety policies using multiple correspondence analysis. *Dyna*, 83(196), 31–38. <http://doi.org/10.15446/dyna.v83n196.56606>
- Chang, C.-L., & Hsu, M.-Y. (2009). The study that applies artificial intelligence and logistic regression for assistance in differential diagnostic of pancreatic cancer. *Expert Systems with Applications*, 36(7), 10663–10672. <http://doi.org/10.1016/j.eswa.2009.02.046>
 - Cho, S.-R., & Lee, S. (2015). A prediction method of ice breaking resistance using a multiple regression analysis. *International Journal of Naval Architecture and Ocean Engineering*, 7(4), 708–719. <http://doi.org/10.1515/ijnaoe-2015-0050>
 - Corsini, A., De Propriis, L., Feudo, S., & Stefanato, M. (2015). Assessment of a Diagnostic Procedure for the Monitoring and Control of Industrial Processes. *Energy Procedia*, 75, 1772–1778. <http://doi.org/10.1016/j.egypro.2015.07.467>
 - Cupani, M. (2012). Análisis de Ecuaciones Estructurales: conceptos, etapas de desarrollo y un ejemplo de aplicación. *Revista Tesis*.
 - Das, S., & Sun, X. (2016). Association knowledge for fatal run-off-road crashes by Multiple Correspondence Analysis. *IATSS Research*, 39(2), 146–155. <http://doi.org/10.1016/j.iatssr.2015.07.001>
 - De la Fuente, F. . (2011). Análisis de Correspondencias Santiago de la Fuente Fernández. *Fac. Ciencias Económicas Y Empresariales*, 58.
 - Demongeot, J., Hamie, A., Laksaci, A., & Rachdi, M. (2016). Relative-error prediction in nonparametric functional statistics: Theory and practice. *Journal of Multivariate Analysis*, 146, 261–268. <http://doi.org/10.1016/j.jmva.2015.09.019>
 - Díaz-García, J. A. (2011). On generalized multivariate analysis of variance. *Brazilian Journal of Probability and Statistics*, 25(1), 1–13. <http://doi.org/10.1214/09-BJPS107>
 - Dobrzański, L. A., Trzaska, J., & Dobrzańska-Danikiewicz, A. D. (2014). 2.09 - Use of Neural Networks and Artificial Intelligence Tools for Modeling, Characterization, and Forecasting in Material Engineering. In S. Hashmi, G. F. Batalha, C. J. Van Tyne, & B. Yilbas (Eds.), *Comprehensive Materials Processing* (pp. 161–198). Oxford: Elsevier.
 - Fan, M., Shibata, H., & Chen, L. (2017). Environmental and economic risks assessment under climate changes for three land uses scenarios analysis across Teshio watershed, northernmost of Japan. *Science of The Total Environment*, 599–600, 451–463. <http://doi.org/10.1016/j.scitotenv.2017.05.010>
 - Findik, D., & Kunçay, C. O. Ş. (2009). A model for instructors adoption of learning management systems : empirical validation in higher education context. *Middle East Technical University*.
 - Fischer, B., Roth, V., & Buhmann, J. M. (2007). Time-series alignment by non-negative multiple generalized canonical correlation analysis. *BMC*

Bioinformatics, 8 Suppl 10, S4. <http://doi.org/10.1186/1471-2105-8-S10-S4>

- Fox, S., & Hammond, S. (2017). Investigating the multivariate relationship between impulsivity and psychopathy using canonical correlation analysis. *Personality and Individual Differences*, 111, 187–192. <http://doi.org/10.1016/j.paid.2017.02.025>
- Gomez, R., & Watson, S. D. (2017). Confirmatory Factor Analysis of the Combined Social Phobia Scale and Social Interaction Anxiety Scale: Support for a Bifactor Model. *Frontiers in Psychology*, 8(February), 1–13. <http://doi.org/10.3389/fpsyg.2017.00070>
- Hair, J., Black, W., Babin, B., & Anderson, R. (2010). *Multivariate Data Analysis*. (Pearson, Ed.) (Seventh).
- Hair, J. F., Black, W. C., & Babin, B. J. (2010). *Multivariate Data Analysis: A Global Perspective*. Pearson Education.
- Handscomb, L., Hall, D. A., Hoare, D. J., & Shorter, G. W. (2016). Confirmatory factor analysis of Clinical Outcomes in Routine Evaluation (CORE-OM) used as a measure of emotional distress in people with tinnitus. *Health and Quality of Life Outcomes*, 14(1), 124. <http://doi.org/10.1186/s12955-016-0524-5>
- Hernández, R., Fernández-Collado, C., & Baptista, P. (2006). *Metodología de la Investigación*. (M. A. Castellanos Toledo, Ed.) (Cuarta). México - México: McGraw-Hill.
- Hwang, T., & Park, T. (2009). Identification of differentially expressed sub-networks based on multivariate ANOVA scoring method. *BMC Bioinformatics*, 10(1), 128. <http://doi.org/10.1186/1471-2105-10-128>
- Jończyk, J. A., & Olszewska, A. M. (2016). The Use of Correspondence Analysis in Assessing the Antecedents of Innovativeness in Public Hospitals. *Studies in Logic, Grammar and Rhetoric*, 47(1), 85–101. <http://doi.org/10.1515/slgr-2016-0048>
- Kraft, P., Bauman, L., Yuan, J. Y., Horvath, S., & Study, F. H. (2003). Multivariate variance-components analysis of longitudinal blood pressure measurements from the Framingham Heart Study. *BMC Genetics*, 4 Suppl 1, S55. <http://doi.org/10.1186/1471-2156-4-S1-S55>
- Lay, J.-G., Chen, Y.-W., & Chi, Y.-L. (2013). GIS Adoption Among Senior High School Geography Teachers in Taiwan. *Journal of Geography*, 112(3), 120–130. <http://doi.org/10.1080/00221341.2012.682226>
- Lin, W.-S. (2012). Perceived fit and satisfaction on web learning performance: IS continuance intention and task-technology fit perspectives. *International Journal of Human-Computer Studies*, 70(7), 498–507. <http://doi.org/10.1016/j.ijhcs.2012.01.006>
- Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data With Missing Values. *Journal of the American Statistical Association*, 83(404), 1198–1202.

- Lu, Z., & Liang, Z. (2016). A Complete Subspace Analysis of Linear Discriminant Analysis and Its Robust Implementation, 2016(Mmc).
- Mbarek, R., & Zaddem, F. (2013). The examination of factors affecting e-learning effectiveness. *International Journal of Innovation and Applied Studies*, 2(4), 423–435. <http://doi.org/2028-9324>
- Morowatisharifabad, M. A., Khankolabi, M., Mozaffari-Khosravi, H., Fallahzade, H., Gerami, M. H., & Salehi-Abargouei, A. (2017). Parenting style, parental feeding practices and children's nutritional status in authoritative parenting style model: A structural equation modeling. *Iranian Red Crescent Medical Journal*, 19(3). <http://doi.org/10.5812/ircmj.41401>
- Motofuji, Y., Saito, A., Koike, M., Kodera, Y., Maeda, T., & Komatsu, H. (2012). Potential of classification of cancer by multiple discriminant analysis for relationship between cancer and expression of human cellular phosphoprotein. *Biomedical Research (Tokyo, Japan)*, 33(2), 139–43. <http://doi.org/10.2220/biomedres.33.139>
- Namakforoosh, M. N. (2005). Metodología de la investigación. (GRUPO NORIEGA, Ed.) (LIMUSA S.A). Mexico. Retrieved from https://books.google.com.ec/books?id=ZEJ7-0hmvhwC&pg=PA67&dq=variables+dependientes&hl=es-419&sa=X&redir_esc=y#v=onepage&q&f=false
- Okimoto, G., Zeinalzadeh, A., Wenska, T., Loomis, M., Nation, J. B., Fabre, T., ... Kwee, S. (2016). Joint analysis of multiple high-dimensional data types using sparse matrix approximations of rank-1 with applications to ovarian and liver cancer. *BioData Mining*, 9(1), 24. <http://doi.org/10.1186/s13040-016-0103-7>
- Palavecinos, B. S., Espinoza, R. C., Mella, H. de la F., & Rosso, D. C. (2016). Analysis of information disclosure for chilean case using multiple correspondence methodology. *Análisis de Revelación de La Información Para El Caso Chileno Usandola Metodología de Correspondencias Múltiples.*, 24(2), 327–336. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=114708779&site=eds-live&authtype=uid>
- Pan, D.-D., Xiong, W.-J., Zhou, J.-Y., Pan, Y., Zhou, G.-L., & Fung, W.-K. (2013). Robust Joint Analysis with Data Fusion in Two-Stage Quantitative Trait Genome-Wide Association Studies. *Computational and Mathematical Methods in Medicine*, 843563. <http://doi.org/10.1155/2013/843563>
- Pardo, C. E., & del Campo, P. C. (2007). Combinación de métodos factoriales y de análisis de conglomerados en R: El paquete factoclass. *Revista Colombiana de Estadística*, 30(2), 231–245.
- Patton, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110, 4–18. <http://doi.org/10.1016/j.jmva.2012.02.021>
- Pedrosa, V. B., Groeneveld, E., Eler, J. P., & Ferraz, J. B. S. (2014). Comparison

of bivariate and multivariate joint analyses on the selection loss of beef cattle. *Genetics and Molecular Research*, 13(2), 4036–4045. <http://doi.org/10.4238/2014.May.23.14>

- Pérez, E. R., & Medrano, L. (2010). Análisis factorial exploratorio: Bases conceptuales y metodológicas. *Revista Argentina de Ciencias Del Comportamiento*, 2(1889), 58–66. Retrieved from http://www.academia.edu/12607324/Revista_Argentina_de_Ciencias_del_Comportamiento_RACC_Análisis_Factorial_Exploratorio_Bases_Conceptuales_y_Metodológicas_Artículo_de_Revisión
- Peruchi, R. S., Paiva, A. P., Balestrassi, P. P., Ferreira, J. R., & Sawhney, R. (2014). Weighted approach for multivariate analysis of variance in measurement system analysis. *Precision Engineering*, 38(3), 651–658. <http://doi.org/10.1016/j.precisioneng.2014.03.001>
- Ramirez-Anormaliza, R., Llinàs-Audet, X., & Sabaté, F. (2016). THE ACCEPTANCE AND USE OF THE E-LEARNING SYSTEMS AMONG THE UNIVERSITY TEACHERS IN ECUADOR. In IATED (Ed.), *EDULEARN16 Conference 4th-6th July 2016* (pp. 3666–3674). Barcelona, Spain. Retrieved from <https://library.iated.org/view/RAMIREZANORMALIZA2016ACC>
- Ramirez-Anormaliza, R., Sabaté, F., Llinàs-Audet, X., & Lordan, O. (2017). Aceptación y uso de los sistemas e-learning por estudiantes de Grado de Ecuador: El caso de una universidad estatal. *Intangible Capital*, 13(3), 548. <http://doi.org/10.3926/ic.820>
- Rhee, J.-K., Joung, J.-G., Chang, J.-H., Fei, Z., & Zhang, B.-T. (2009). Identification of cell cycle-related regulatory motifs using a kernel canonical correlation analysis. *BMC Genomics*, 10 Suppl 3, S29. <http://doi.org/10.1186/1471-2164-10-S3-S29>
- Rodrigues, L., Grave, R., de Oliveira, J. M., & Nogueira, C. (2016). Study on homophobic bullying in Portugal using Multiple Correspondence Analysis (MCA). *Revista Latinoamericana de Psicología*, 48(3), 191–200. <http://doi.org/10.1016/j.rlp.2016.04.001>
- Ryu, E., & Cheong, J. (2017). Comparing Indirect Effects in Different Groups in Single-Group and Multi-Group Structural Equation Models. *Frontiers in Psychology*, 8(May), 1–14. <http://doi.org/10.3389/fpsyg.2017.00747>
- Sanjuán Suárez, P., Pérez García, A. M., & Bermúdez Moreno, J. (2000). Escala de autoeficacia general: Datos psicométricos de la adaptación para población española. *Psicothema*, 12(SUPPL. 2), 509–513. <http://doi.org/ISSN 0214-9915>
- Schumacker, R. E., & Lomax, R. G. (2010). *A Beginner's Guide to Structural Equation Modeling: Third Edition*. (Taylor & Francis Group, Ed.) (3rd ed.). New York - USA: Routledge. Retrieved from <http://books.google.com/books?id=58pWPxWPC90C&pgis=1>
- Seward, K., Wolfenden, L., Wiggers, J., Finch, M., Wyse, R., Oldmeadow, C., ... Yoong, S. L. (2017). Measuring implementation behaviour of menu

guidelines in the childcare setting: confirmatory factor analysis of a theoretical domains framework questionnaire (TDFQ). *International Journal of Behavioral Nutrition and Physical Activity*, 14(1), 45. <http://doi.org/10.1186/s12966-017-0499-6>

- Sha, Q., Zhang, Z., & Zhang, S. (2011). Joint analysis for genome-wide association studies in family-based designs. *PLoS ONE*, 6(7). <http://doi.org/10.1371/journal.pone.0021957>
- Shah, T., Patel, M. N., Nath, S., & Menon, S. K. (2016). Determination of sex using cephalo-facial dimensions by discriminant function and logistic regression equations. *Egyptian Journal of Forensic Sciences*, 6(2), 114–119. <http://doi.org/10.1016/j.ejfs.2015.03.004>
- Sturn, A., Quackenbush, J., & Trajanoski, Z. (2002). Genesis: cluster analysis of microarray data. *Bioinformatics (Oxford, England)*, 18(1), 207–8. <http://doi.org/10.1093/bioinformatics/18.1.207>
- Tarhini, A., Hone, K., & Liu, X. (2013). Factors Affecting Students' Acceptance of e-Learning Environments in Developing Countries : A Structural Equation Modeling Approach. *International Journal of Information and Education Technology*, 3(1), 54–59. <http://doi.org/10.7763/IJET.2013.V3.233>
- Tekaiia, F. (2014). Genome Data Exploration Using Correspondence Analysis. *Bioinformatics and Biology Insights*, 1(7), 7523. <http://doi.org/10.4137/BBI.S39614.TYPE>
- Veldhuis, J. D., & Johnson, M. L. (1986). Cluster analysis: a simple, versatile, and robust algorithm for endocrine pulse detection. *The American Journal of Physiology*, 250(4 Pt 1), E486-93.
- Verdugo, M. Á. (2008). Metodología en la investigación sobre discapacidad. Introducción al uso de las ecuaciones estructurales.
- Waaijenborg, S., & Zwinderman, A. H. (2009). Sparse canonical correlation analysis for identifying, connecting and completing gene-expression networks. *BMC Bioinformatics*, 10, 315. <http://doi.org/10.1186/1471-2105-10-315>
- Weng, C., & Tsai, C. (2015). Social support as a neglected e-learning motivator affecting trainee 's decisions of continuous intentions of usage. *Australasian Journal of Educational Technology*, 31(2), 177–192.
- Wu, C.-C., & Chiu, J. (2017). Economic evaluation of asymmetric and price range information in gold and general financial markets. *Journal of International Money and Finance*, 74, 53–68. <http://doi.org/10.1016/j.jimonfin.2017.03.001>
- Wu, P., Presanis, A. M., Bond, H. S., Lau, E. H. Y., Fang, V. J., & Cowling, B. J. (2017). A joint analysis of influenza-associated hospitalizations and mortality in Hong Kong, 1998-2013. *Scientific Reports*, 7(1), 929. <http://doi.org/10.1038/s41598-017-01021-x>
- Xu, L. W. (2014). MANOVA for Nested Designs with Unequal Cell Sizes and Unequal Cell Covariance Matrices. *Journal of Applied Mathematics*, 2014.

<http://doi.org/10.1155/2014/649202>

- Ye, X., Kang, Y., Zuo, B., & Zhong, K. (2017). Study of factors affecting warm air spreading distance in impinging jet ventilation rooms using multiple regression analysis. *Building and Environment*, 120, 1–12. <http://doi.org/10.1016/j.buildenv.2017.03.044>
- Yoshida, K., Yoshimoto, J., & Doya, K. (2017). Sparse kernel canonical correlation analysis for discovery of nonlinear interactions in high-dimensional data. *BMC Bioinformatics*, 18(1), 108. <http://doi.org/10.1186/s12859-017-1543-x>
- Yuan, Z., Zheng, X., Zhang, L., & Zhao, G. (2017). Urban Competitiveness Measurement of Chinese Cities Based on a Structural Equation Model. *Sustainability*, 9(4), 666. <http://doi.org/10.3390/su9040666>
- Zhou, Y., & Sun, S. (2015). *Semisupervised Tangent Space Discriminant Analysis*, 2015.

UNENMI 2017

*“Tan sólo por la educación puede el hombre llegar a ser hombre.
El hombre no es más que lo que la educación hace de él”*

Immanuel Kant (1724- 1804)
Filósofo prusiano

Capítulo 2: Análisis factorial

Richard Ramirez-Anormaliza
Juan Calderon-Cisneros
Javier Bermeo-Paucar
Jesmy Castelo-Gonzalez

Objetivos de aprendizaje

En un estudio para cumplir los objetivos, el cuestionario es una de las herramientas que más se utiliza en la valoración de la certeza científica. Los investigadores sustentan sus trabajos basados en la modelación y valoración de sus resultados efectuados por un cuestionario, el cual no ha pasado por el juicio de validación. Validar un cuestionario se establece por un conjunto de argumentos basados en contrastes de hipótesis correctamente formuladas. El método Multivariante de Análisis Factorial puede ser Exploratorio o Confirmatorio, este modelo estadístico nos permite evaluar la validez y la fiabilidad de cada ítem con el cual está conformado el cuestionario y se ejecuta un contraste de hipótesis particular. Al finalizar este capítulo deberá ser capaz de lo siguiente:

- Constituir una correspondencia entre variables y atributos que inicialmente no son supuestamente relacionados.
- Proporcionar la estructura interna, consintiendo crear cómo unos factores afectan a otros y permite modelar matemáticamente la estructura.
- Estudiar la técnica que aprueba reducir un conjunto de variables a un número de factores representativos.
- Analizar los datos previos a la aplicación de una técnica Multivariante análisis factorial.

2.1. Introducción al análisis factorial

El Análisis Factorial como tratamiento estadístico representa todas las relaciones que existe entre un conjunto de variables, según (Cuadras, 2014), menciona “es un método Multivariante que pretende expresar variables observables como una combinación lineal de m variables hipotéticas o latentes, denominadas factores.”, como un concepto basado en variables no observables denominadas latentes e interpretadas como factores, siendo fundamental el número de factores sustancialmente menor que el de las variables.

En pocas palabras el tratamiento estadístico se obtiene con más de un factor común y se verifica su estructura simple mediante modelación matemática de regresión lineal y correlación parcial para las rotaciones de los factores. El ajuste del modelo a los datos es por la presencia de varias dimensiones latentes derivadas de la regresión lineal y de las correlaciones se derivan los valores claves para el ajuste del modelo a los datos.

Manuel Batista-Foguet, Coenders & Alonso (2004) exponen que “el análisis de la varianza, con mediciones pre-tratamiento y pos-tratamiento, permite evaluar la sensibilidad al cambio o discriminar entre distintas poblaciones. En el mejor de los casos, un análisis factorial exploratorio (AFE) de la matriz de correlaciones «dictará» las dimensiones latentes y sus resultados se utilizarán como indicación de validez convergente y discriminante.”, ya después de varias modelaciones matemáticas en los modelos de ajuste de los datos para extraer los factores que permitirán decidir cuál es el mejor modelo a explicar.

Para Pérez & Medrano (2010), señala que para el análisis factorial “el hecho de que dos variables se encuentran relacionadas entre sí proviene del hecho que ambas variables poseen algo en común y algo que las diferencia. De esta manera, la varianza total de una variable se debe a factores que comparte con las otras variables (comunalidad) y a factores específicos de la variable (especificidad).”.

Para Lloret-Segura y colaboradores (2014), señalan que “El análisis factorial es una técnica de reducción de datos que sirve para encontrar grupos homogéneos de variables a partir de un conjunto numeroso de variables. Esos grupos homogéneos se forman con las variables que correlacionan mucho entre sí y procurando, inicialmente, que unos grupos sean independientes de otros”.

Martinez (2014) señala que “El Análisis Factorial es, por tanto, una técnica de reducción de la dimensionalidad de los datos. Su propósito último consiste en buscar el número mínimo de dimensiones capaces de explicar el máximo

de información contenida en los datos. A diferencia de lo que ocurre en otras técnicas como el análisis de varianza o el de regresión, en el análisis factorial todas las variables del análisis cumplen el mismo papel”, las mismas tiene un independencia de las variables sobre otras variables lo que se pretende con el Análisis Factorial, es simplificar la información que nos da una matriz de correlaciones para hacerla más fácilmente interpretable.

Según Morales Vallejo (2011), los análisis indicados en la **Figura 2 - 1** son los típicos en la construcción de escalas de actitudes (análisis de ítems, fiabilidad, validez) y la fiabilidad (precisión) van a ser dos consideraciones importantes en todo el proceso de construcción de una escala se encarga de analizar la varianza común a todas las variables.

Frías-Navarro & Soler (2012) expone que “Si se desea analizar exclusivamente la varianza compartida habrá que eliminar los unos de la matriz de correlaciones y poner en su lugar la proporción de varianza que cada ítem tiene en común con todos los demás”.

El análisis factorial exploratorio, para Arias y Justicia (2003), señalan que “se usa el AFE para tratar de descubrir la estructura interna de un número relativamente grande de variables. La hipótesis a priori del investigador es que pueden existir una serie de factores asociados a grupos de variables. Las cargas de los distintos factores se utilizan para intuir la relación de éstos con las distintas variables. Es el tipo de análisis factorial más común”.

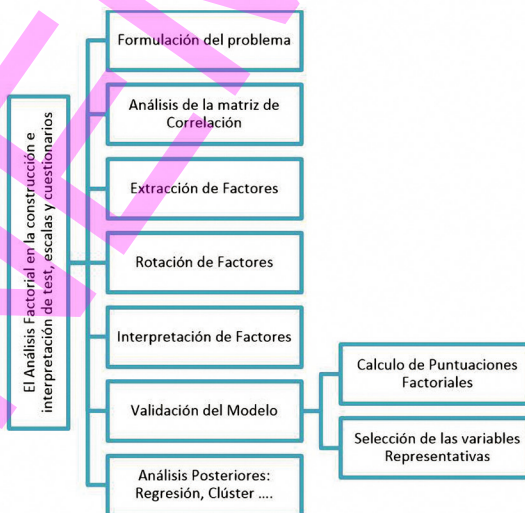


Figura 2 - 1. Esquema de un Análisis Factorial

Nota: Tomado de la Guía para construir cuestionarios y escalas de actitudes. Vallejo (2011)

Arias (2008) señala que “el AFC trata de determinar si el número de factores obtenidos y sus cargas corresponden con los que cabría esperar a la luz de una teoría previa acerca de los datos. La hipótesis a priori es que existen unos determinados factores preestablecidos y que cada uno de ellos está asociado con un determinado subconjunto de las variables. El análisis factorial confirmatorio entonces arroja un nivel de confianza para poder aceptar o rechazar dicha hipótesis”

2.2. Campo de aplicación del análisis factorial

El AF pueden explicarse en cualquier campo que se desee a partir de una serie de variables no observables (latentes) denominadas factores, expresarlas en un número menor de variables. Se asume explícitamente que la combinación de variables que resulta es tan sólo una aproximación, y que una parte de la puntuación del criterio no podrá ser predicha o explicada desde otro conjunto de variables bajo análisis.

2.3. Software para análisis factorial

Los últimos avances en análisis factorial en psicometría se encuentran ligados en buena medida al desarrollo de tecnologías informáticas, especialmente en relación al acceso masivo a programas personales para cálculos estadísticos específicos. Dentro de los paquetes más difundidos puede mencionarse el IBM SPSS Statistics 23, el SYSTAT 13.1, el IBM AMOS 22, el EQS, LISREL 8 y MPLUS 7, entre otros, de los resultados surgen constantes alternativas tendientes a mejorar la calidad de los análisis y de los soportes estadísticos.

2.4. Ejemplos de análisis factorial

2.4.1. Utilización de SPSS en el análisis factorial

2.4.1.1. Introducción

Con la finalidad de ilustrar los pasos que se dan en el análisis de datos previo a la aplicación de una técnica multivariante AFE, según la estrategia sugerida Ramírez Anormaliza (2013), se siguió para la búsqueda de información bibliográfica. Se basó en documentos publicados por los organismos estatales que intervienen en el área de turismo y desarrollo local como es el caso del Ministerio de Turismo a través del PLANDETUR y Marketing Turístico; Senplades con el Plan Nacional del Buen Vivir; Ministerio Coordinador de la Producción, Empleo y Competitividad; Universidad Estatal de Milagro; Gobierno Autónomo Descentralizado del cantón Milagro, Instituto Nacional de Estadísticas y Censos; adicionalmente se consideró información relevante

de libros, estudios de casos, y proyectos desarrollados por estudiantes de Universidades Nacionales.

En este estudio, se realizó una encuesta a la población económicamente activa (PEA) de la ciudad de Milagro en Ecuador, se distribuyó a 401 habitantes del sector. En total, 360 personas respondieron correctamente a la encuesta, lo que resulta en una tasa de respuesta del 89,8% válida para los análisis.

Con el fin de analizar las percepciones de la PEA sobre la importancia de los principales tipos de negocios en el sector, preguntas sobre comercio, turismo, manufactura, impresión, agricultura y ganadería también se incluyeron en el cuestionario (Arteaga-arcentales, Peralta-carpio, Castelo-gonzalez, & Echeverría-vásquez, 2016).

Se consideró la validez del instrumento de recolección de datos (es decir, la cara de validez y la validez de contenido) antes de su dimensión estadística (es decir, la fiabilidad, la validez de constructo) para Hernández el primer borrador del instrumento fue revisado en base a las opiniones de los expertos, un master en turismo y un master en administración.

Con la finalidad de que el instrumento sea lo menos técnico posible y de fácil llenado, dado la población consultada. Se acordó en 14 ítems de aprobación – desaprobarción, la escala de Likert de 5 puntos (de 1 “muy bajo” a 5 “muy de alto”) fue empleada.

2.4.1.2. Pasos

En el análisis de los datos, SPSS versión 21.0 fue utilizado tanto para la fiabilidad y el AFE. Los datos de investigación se almacenaron en Microsoft Excel y luego con ellos se generó el archivo de texto delimitado por comas.

El mismo que tiene el nombre Ejercicio21_turismo.csv y se encuentra disponible en el link <https://www.dropbox.com/sh/448ijhl11qovqj/AADYa8UTmaJFDNbpsSfvM7pZa?dl=0>. Para que el lector pueda replicar este ejercicio.

Las acciones específicas realizadas fueron las siguientes:

1. Abrimos el icono de SPSS y nos presenta una pantalla en la que encontramos diferentes opciones, donde se pueden abrir archivos recientes, ejecutar tutorial o abrir archivos de otro sitio, en nuestro caso se abrirá un archivo de datos. Esta pantalla se muestra en la **Figura 2 - 2**.

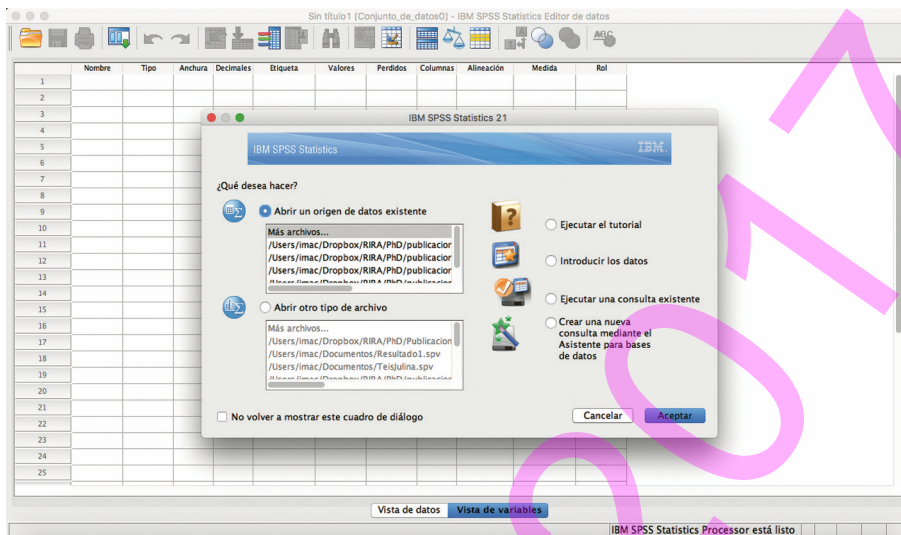


Figura 2 - 2. Pantalla de inicio de SPSS

2. En Abrir un origen de datos existente, dar clic sobre Más archivos y ubicar la ruta donde descargo el archivo de datos Ejercicio21_turismo.csv, ante lo cual se presenta el cuadro de diálogo Abrir datos que se visualiza en la Figura 2 - 3.

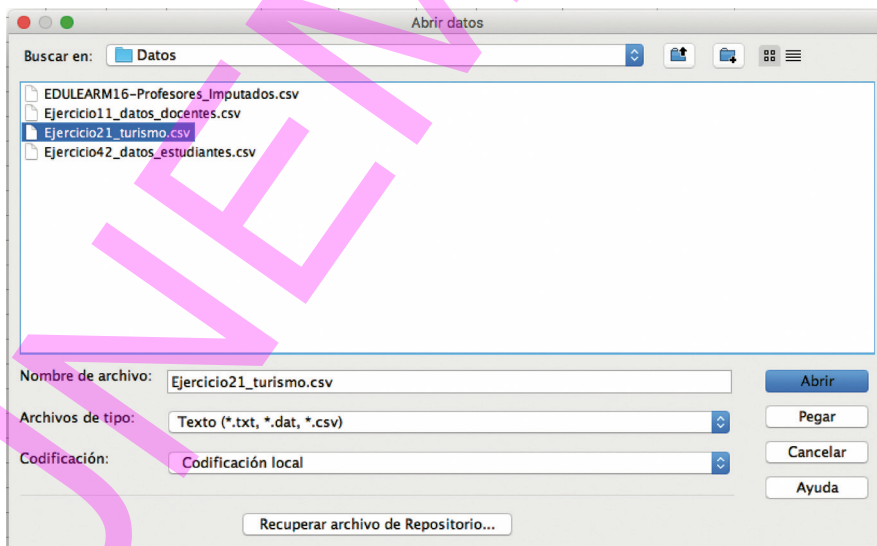


Figura 2 - 3. Cuadro de diálogo Abrir datos, SPSS

- Se inicia el asistente para importación de datos de SPSS y aparecen cuadros de diálogos sucesivos de seis pasos que se deben contestar tal como se muestran en las **Figura 2 - 4** a **Figura 2 - 9**.

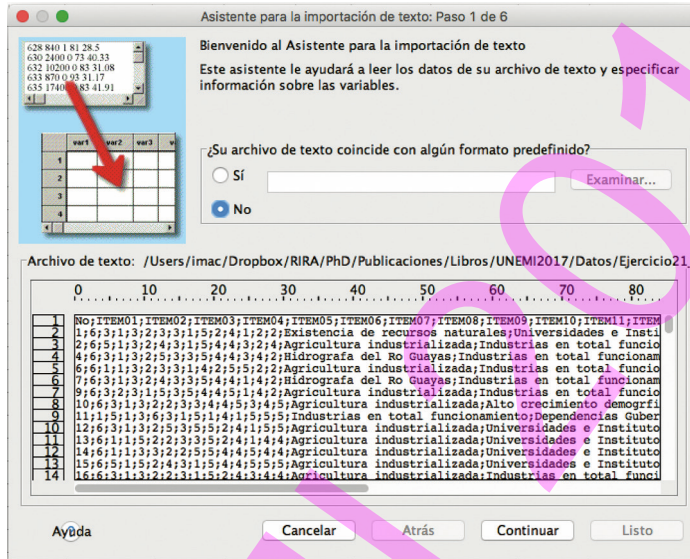


Figura 2 - 4. Asistente para importación de texto: Paso 1 de 6

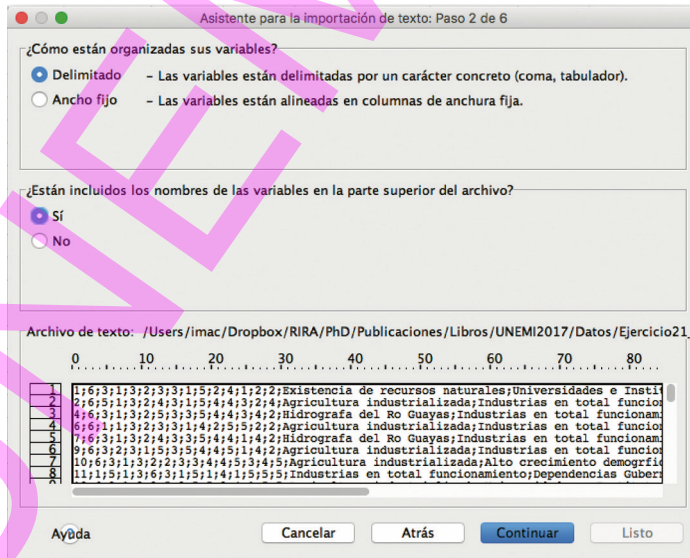


Figura 2 - 5. Asistente para importación de texto: Paso 2 de 6

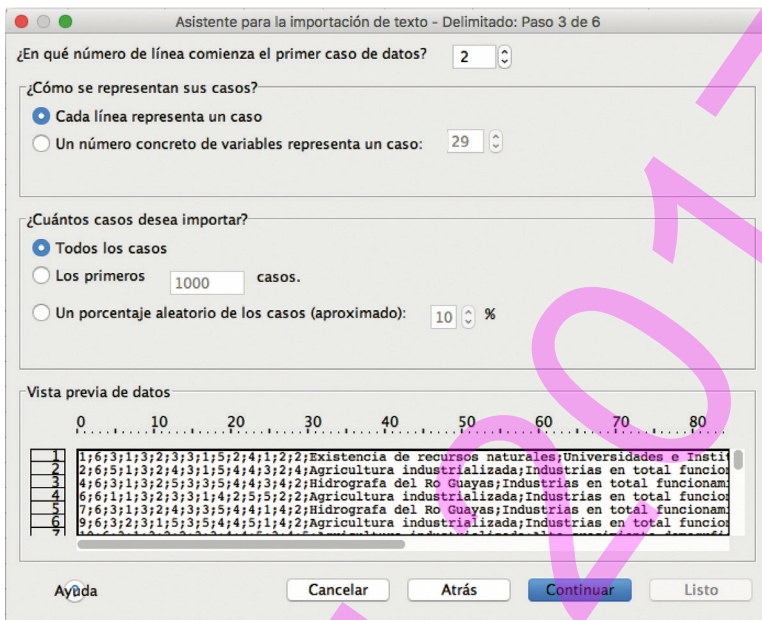


Figura 2 - 6. Asistente para importación de texto: Paso 3 de 6

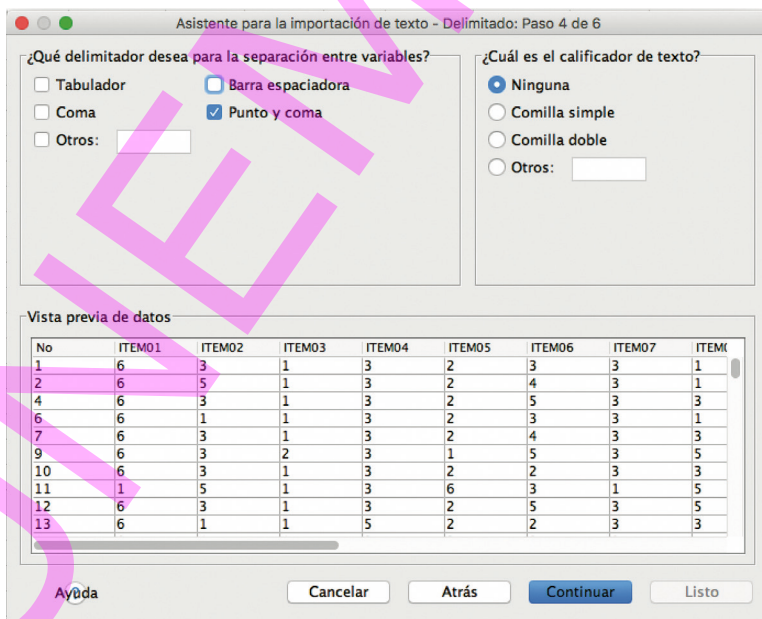


Figura 2 - 7. Asistente para importación de texto: Paso 4 de 6

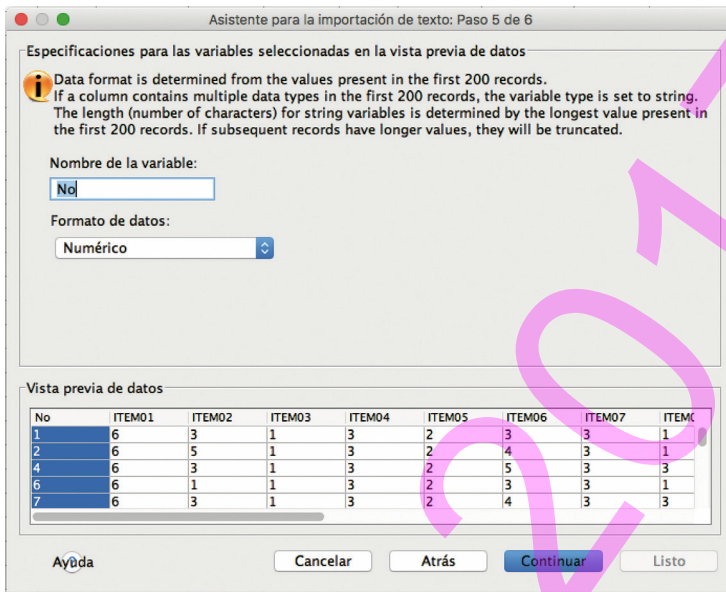


Figura 2 - 8. Asistente para importación de texto: Paso 5 de 6

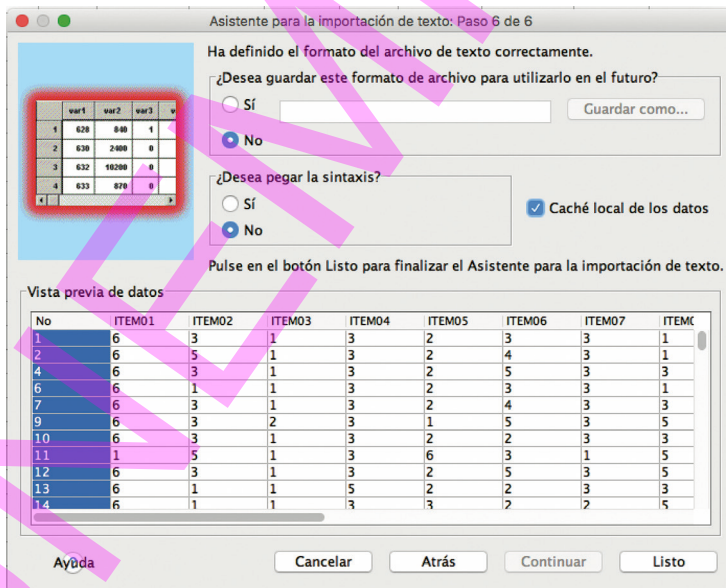


Figura 2 - 9. Asistente para importación de texto: Paso 6 de 6

Una vez cargados los datos, SPSS muestra las variables cargadas y su tipo de dato. Tal como se puede observar en la **Figura 2 - 10** donde consta el editor de datos.

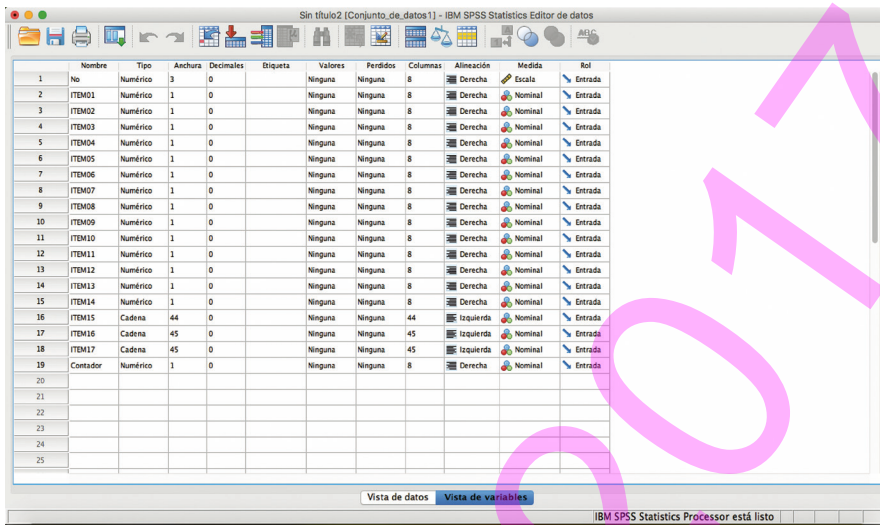


Figura 2 - 10. Editor de datos

- Se ejecuta el análisis factorial, para ello se debe dar clic en Analizar, Reducción de dimensiones. Luego en Factor. Como se puede observar en la Figura 2 - 11.

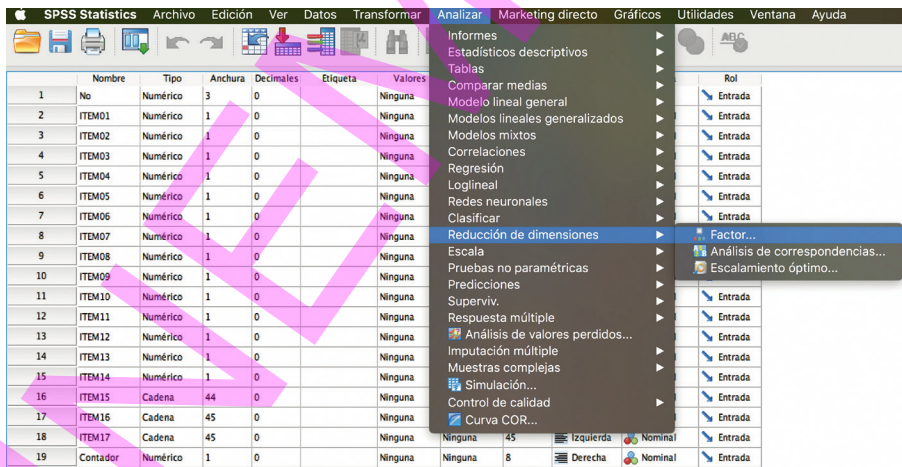


Figura 2 - 11. Opción Factor

- Se solicita a SPSS que nos haga un análisis factorial para determinar si existe una estructura factorial subyacente a estas variables que hemos observado. Ver Figura 2 - 12. En el cuadro de diálogo de Análisis Factorial, hay una serie de opciones para configurar como vamos a hacer ese análisis

factorial, lo primero que vamos a hacer es indicar al SPSS que variables observadas son las que forman parte del análisis, en nuestro caso los ITEM02, ITEM04, ITEM06, ITEM08, ITEM10, ITEM12, ITEM13 e ITEM14, los marcamos y hacemos click en esta flecha que está a la derecha, pasamos las variables que van a ser analizadas a esta caja de opciones, en la derecha tenemos diferentes opciones para indicar distintas opciones que nos lleva a ejecutar el análisis factorial. Ver **Figura 2 - 13**.

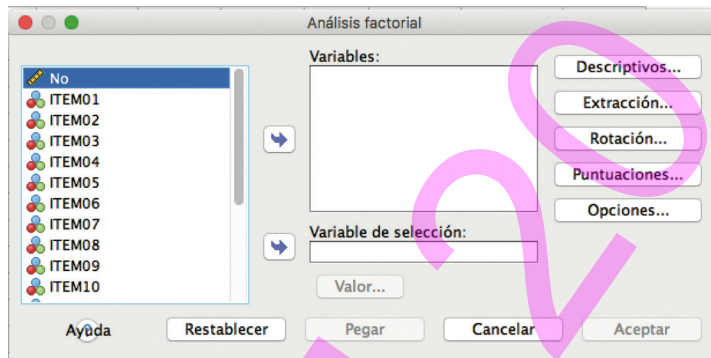


Figura 2 - 12. Diálogo de Análisis Factorial

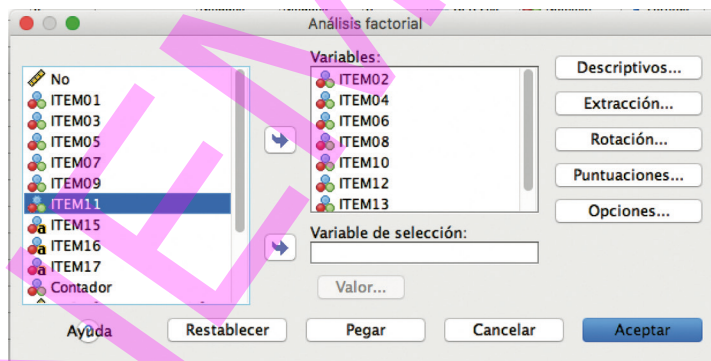


Figura 2 - 13. Selección de variables

6. En la opción descriptivos que se muestra en la **Figura 2 - 14**, nos muestra la posibilidad de solicitar:
 - a) Descriptivos univariados, es decir, que nos muestre la salida parámetros con respecto a una de las variables observadas, de manera unidimensional, por ejemplo la media, desviación típica de cada una de las variables.

- b) La solución inicial.
- c) Todas las opciones relativas a la matriz de correlaciones. Podemos indicarle que nos muestre los coeficientes de la propia matriz de correlaciones, los niveles de significación de esos coeficientes, el determinante lo cual es importante también conocer. Podemos solicitarle que nos haga la prueba KMO y la prueba de esfericidad de Bartlett para comprobar la adecuación muestral. El inverso no se suele utilizar mucho, vamos a omitirlo en este caso, la matriz de correlación reproducir a partir de los parámetros estimados y la matriz Anti-imagen.

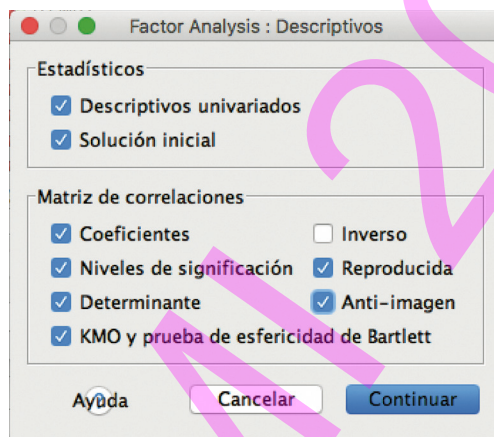


Figura 2 - 14. Análisis de Factor- Descriptivos

- 7. En el botón Extracción, podemos indicar el método de identificación o de extracción de los factores. Podemos elegir componentes principales, en el que se asume que toda la varianza a explicar por los factores es toda, no es la compartida o elegir uno de los otros métodos. Podemos también partir del análisis de la matriz de correlaciones o desde la matriz de covarianza, en este caso vamos a partir de la matriz de correlaciones, con respecto a la visualización nos puede dar la solución factorial antes de hacer cualquier tipo de rotación, es decir, los que son los coeficientes, las saturaciones, podemos solicitar que nos haga un gráfico de segmentación, el script plot el cual también es importante para decidir el número de factores a extraer, y con respecto a la extracción podemos basarnos en el criterio de Kaiser, es decir, vamos a retener únicamente los factores cuyo valor propio o autovalor sean mayores que 1. Podríamos indicar y ser más restrictivos o poner cualquier otro valor, en principio y para un primer análisis es aconsejable dejar el criterio de Kaiser. Se puede determinar números fijos de factores también. Ver **Figura 2 - 15**.

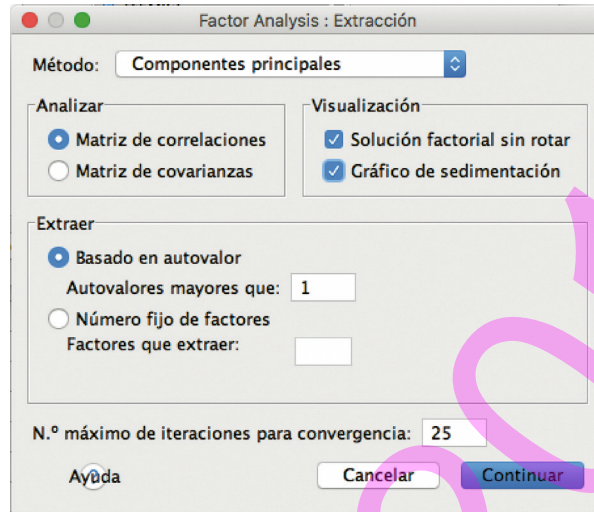


Figura 2 - 15. Análisis de Factor- Extracción

8. En cuanto a la rotación, si consideramos que es necesario podemos indicar alguno de los tipos de rotación ortogonal como es: quartimax, equamax o varimax, que quizás es la más recomendada. Si vamos a considerar que los factores después de la rotación sigan siendo ortogonales o algunos de los dos métodos de rotación oblicua como Oblimin directo o Promax, también podemos pedir que proporcione la solución rotada y un gráfico de la saturaciones con los pesos, este gráfico, si tiene previsto que pueden salir dos factores, tiene sentido ya que nos permite interpretar mejor la solución, si no realmente no va a portar mucha información. Ver **Figura 2 - 16.**

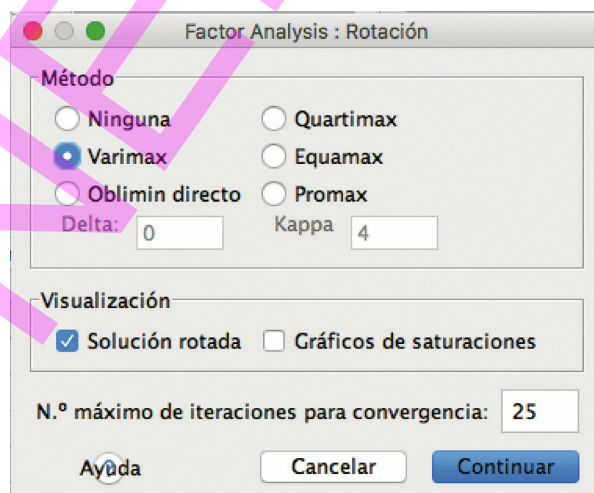


Figura 2 - 16. Análisis de Factor- Rotación

9. En cuanto a las puntuaciones factoriales, ver si los valores que estimamos que va tener cada individuo en los diferentes factores se puede también calcular y guardar como variables, aquí indicamos que método de estimación de las puntuaciones factoriales queremos, vamos a elegir de inicio el de regresión y a decirle que nos muestre una vez calculada esa matriz de los coeficientes, a partir de los cuales se obtiene las puntuaciones factoriales. Ver **Figura 2 - 17**.

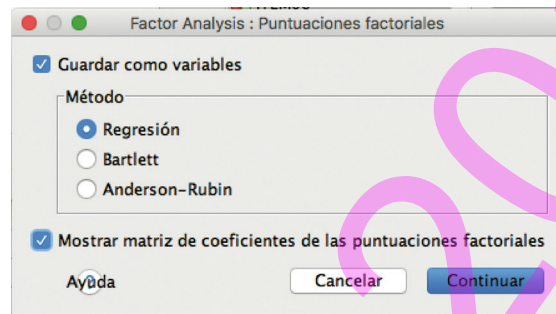


Figura 2 - 17. Análisis de Factor- Puntuaciones Factoriales

10. Por ultimo en el botón opciones tenemos distintas sub opciones para tratar los valores perdidos y también el formato de la presentación de los coeficientes, los pesos, las saturaciones finales, vamos a pedir que nos ordene por tamaños lo que nos clarifica la importancia de cada uno de los factores, se puede indicar para que la interpretación sea más clara, suprima aquellos pesos que son más bajos en valor absoluto que un determinado valor, en principio se suele poner el valor 04 aunque esto dependerá del tamaño de la muestra. Ver **Figura 2 - 18**.

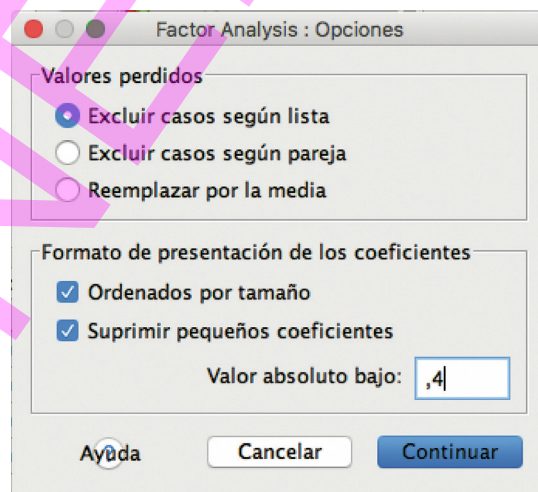


Figura 2 - 18. Análisis de Factor- Opciones

Finalmente, una vez definidos todos los parámetros en el cuadro de diálogo Análisis Factorial, damos clic en botón aceptar y se tiene el resultado. La pantalla que presente esta acción se muestra en la **Figura 2 - 19**.

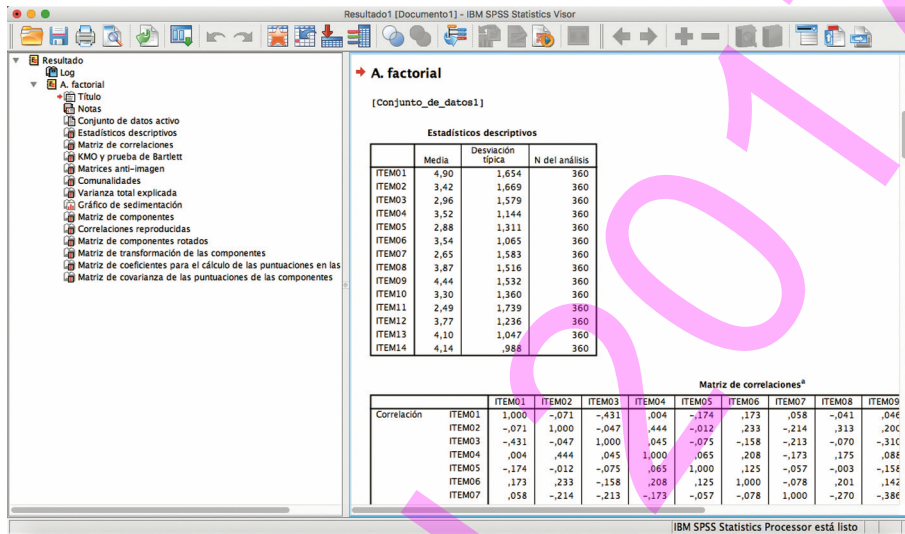


Figura 2 - 19. Visor de resultados de SPSS

Para finalizar el proceso, corresponde luego analizar e interpretar los resultados que entrega SPSS, respecto al ejercicio que se describe, en los párrafos siguientes se procede con dicha tarea. Para lo cual se resumirán ciertas tablas que permiten clarificar las decisiones tomadas.

Una cuestión fundamental que afecta la fiabilidad y la validez de la medición y determinación de la idoneidad de los datos para el factor de análisis es un tamaño de muestra adecuado y hay varias opiniones con respecto a este tema. En este estudio se apegó al criterio que un número entre 100 y 400 puede ser considerado como adecuado Hair, Black, & Babin (2010). Por lo tanto, el tamaño de la muestra del estudio (n = 360) se puede considerar como apropiado.

Para el examen de la distribución de cada una de las respuestas de los ítems y de las relaciones de cada ítem con los demás y toda la encuesta, se analizaron los ítems.

En este contexto, las correlaciones ítem-total de cada elemento se examinaron junto con el valor de alfa de Cronbach nuevo si se elimina el elemento. A continuación, los ítems cuyos correlaciones ítem-total fueron por debajo del criterio de corte de 0.20 (Vázquez-Cano & García Sevillano, 2015) fueron examinados y excluidos de la encuesta.

Posteriormente, las pruebas de medida de adecuación del muestreo Kaiser-Meyer-Olkin (KMO) y la prueba de esfericidad de Bartlett (BTS) fueron examinados con el fin de entender si el conjunto de datos es apropiado para el análisis factorial. Es considerado como el valor perfecto donde la KMO se acerca a 1.0 (> 0.60 es adecuada, > 0.80 es grande) mientras que la BTS debe ser significativa al nivel 0.05 Hair, Anderson, Tatham, & Black (2008), Al ver que los resultados de estas pruebas cumplieron las condiciones, se iniciaron los procedimientos de Análisis factorial exploratorio.

En este punto, las extracción de factores se realizó mediante componentes principales y rotación Varimax. Este procedimiento es la forma más utilizada en la literatura, da la mejor solución para el AFE incluso si los datos no tienen distribución normal (Fabrigar, Wegener, MacCallum, & Strahan, 1999).

Para finalizar se procedió a tabular las respuestas sobre la percepción de las que pudieran ser consideradas como fortalezas a la hora de impulsar el turismo en el sector, para el efecto se utilizó Microsoft Excel.

En la primera fase del análisis se revisaron los ítems, las correlaciones elemento-total de los ítems fueron examinadas con el fin de determinar las coherencias de las respuestas de cada ítem, en relación a los otros ítems y toda la encuesta. Además, seis ítems cuyos coeficientes de correlación fueron menor a 0.20 (ítem 1, 3, 5, 7, 9 y 11) fueron eliminados. Como resultado, 8 ítems se dejaron en la encuesta para los siguientes análisis.

En el AFE, se analizaron la KMO y BTS. Se encontró que el valor de KMO fue 0.750, que puede decirse que es un valor bueno y BTS mostró un valor aproximadamente de chi-cuadrado de 422.421 ($p < 0,000$), lo que significaba que la matriz de correlación de los datos para el análisis factorial es apropiada (Alonso, 2008), estos datos se muestran en la **Tabla 2 - 1**.

Tabla 2 - 1. Resultados de las pruebas KMO y BTS

Medida de adecuación muestral de Kaiser-Meyer-Olkin:		.750
Prueba de esfericidad de Bartlett:	Chi-cuadrado aproximado	422.421
	gl	28
	Sig.	.000

Nota: Tomado del artículo Principales Factores para Desarrollo Turístico: Estudio Empírico a la Percepción de la Población Económicamente Activa Arteaga-Arcentales y colaboradores (2016) p 5

Por último, de acuerdo con los resultados del AFE, el porcentaje acumulativo de la varianza explicada que se observó fue de 59.675%. Además, el primer factor tiene un autovalor de 2.626 y explica 32.828% de la varianza total, mientras que el segundo factor tiene un autovalor de 1.146 y explicó el 14.322% de la varianza total.

El autovalor y la varianza para el tercer factor se encontraron que eran de 1,002 y 12.525%, respectivamente. Los detalles relativos a los autovalores y la varianza explicada de los componentes se presentan en la **Tabla 2 - 2**.

Las cargas factoriales calculadas fueron entre 0.50 y 0.76. Además de las cargas factoriales, los valores de la media y la desviación estándar para cada ítem de cada componente se presentan en la **Tabla 2 - 3**.

Tabla 2 - 2. Autovalores y varianza explicada para cada componente

Componente	Auto valores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	2.626	32.828	32.828	2.626	32.828	32.828	1.966	24.575	24.575
2	1.146	14.322	47.151	1.146	14.322	47.151	1.638	20.477	45.052
3	1.002	12.525	59.675	1.002	12.525	59.675	1.170	14.623	59.675
4	.798	9.978	69.653						
5	.750	9.376	79.030						
6	.629	7.863	86.892						
7	.605	7.564	94.456						
8	.444	5.544	100.000						

Nota: Tomado del artículo Principales Factores para Desarrollo Turístico: Estudio Empírico a la Percepción de la Población Económicamente Activa Arteaga-Arcentales y colaboradores (2016) p 6

Tabla 2 - 3. Medias, desviaciones estándar, ítems, y cargas factoriales

Ítem y factor	Media	Desviación estándar	Carga factorial		
			Factor 1	Factor 2	Factor 3
Factor 1: Imagen turística percibida (ITP)					
ITEM13: En qué medida se estima que la afluencia de personas a la ciudad depende de su información turística disponible	4.10	1.047	.729	.265	-.038
ITEM08: La socialización de los beneficios del turismo en qué nivel ayudaría que la población mejore su trato al visitante	3.87	1.516	.717	.147	.085
ITEM12: En qué grado la aplicación de estrategias publicitarias ayudaría a mejorar la imagen turística del sector	3.77	1.236	.702	.203	-.050
Factor 2: Percepción de la importancia del turismo (PIT)					
ITEM02: El conocimiento de las potencialidades económicas de la zona en qué grado ayudaría a mejorar su economía	3.42	1.669	.214	.756	.116
ITEM04: EL conocimiento de la importancia del turismo en qué grado ayudaría a su fortalecimiento	3.52	1.144	.175	.745	.145
ITEM06: La organización del casco comercial en qué grado considera influye en el desarrollo del turismo	3.54	1.065	.290	.503	-.377
Factor 3: Percepción del apoyo gubernamental al turismo (PAGT)					
ITEM10: En qué grado estima que los programas gubernamentales deben abordar el desarrollo turístico	3.30	1.360	-.146	.336	.748
ITEM14: En qué grado usted considera que Milagro puede alcanzar la competitividad en turismo con apoyo gubernamental	4.14	.988	.495	-.118	.650
Alfa de Cronbach total (a) = 0.70					

Nota: Tomado del artículo Principales Factores para Desarrollo Turístico: Estudio Empírico a la Percepción de la Población Económicamente Activa Arteaga-Arcentales y colaboradores (2016) p 6

Después de examinar cada ítem en función de su significación teórica y coherencia con los otros ítems en el factor, los tres factores fueron nombrados de la siguiente manera: La imagen turística percibida (IPT) (3 ítems), Percepción de la importancia del turismo (PIT) (tres ítems), y Percepción del apoyo gubernamental al turismo (PAGT) (dos ítems).

Los ítems en el Factor 1 (IPT) se refieren principalmente a la construcción mental que existe sobre un lugar determinado, haciendo referencia al conocimiento, impresiones y creencias que el turista posee sobre un destino

turístico (Camprubí, Guía, & Comas, 2009) y PIT en el factor 2 los ítems evalúan la creencia de que el turismo representa un eslabón importante para el crecimiento económico de un país, generando empleo, incrementando la capacidad nacional para producir un servicio y reduciendo la pobreza. (Naciones Unidas, 1998). Por otro lado, el factor 3 incluye ítems considerando al turismo como sector prioritario, como impulsador para atraer inversión nacional o extranjera, qué el gobierno debe crear proyectos de desarrollo en este campo (Ministerio de Turismo, 2015).

Los resultados finales de la prueba de fiabilidad revelaron un aceptable coeficiente alfa de Cronbach (0.70) para toda la encuesta.

La percepción sobre las que puede ser consideradas como fortalezas para impulsar el turismo se presentan en la **Tabla 2 - 4**, donde destacan la existencia de recursos naturales y agricultura industrializada.

Tabla 2 - 4. Fortalezas percibidas para apoyar el turismo en Milagro

Fortalezas percibidas	Casos	Porcentajes
Existencia de recursos naturales	164	46%
Agricultura industrializada	87	24%
Hidrografía del Río Guayas	36	10%
Industrias en total funcionamiento	30	8%
Universidades e Institutos Superiores	22	6%
Dependencias Gubernamentales de servicio	10	3%
Presencia significativa del sistema bancario	5	1%
Alto crecimiento demográfico	4	1%
Accesos viales para la entrada a la ciudad	1	0%
Cadenas de Comisariatos y centros comerciales	1	0%
Total general	360	100%

Nota: Tomado del artículo Principales Factores para Desarrollo Turístico: Estudio Empírico a la Percepción de la Población Económicamente Activa Arteaga-Arcentales y colaboradores (2016) p 7

2.4.2. Utilización de R en análisis exploratorio de datos.

Con la finalidad de que el lector tenga la opción de trabajar con análisis factorial, utilizando un programa informático de uso libre o código abierto (no pago). En este apartado se describen los principales pasos para poder realizar el mismo ejercicio desarrollado en SPSS en este capítulo, pero ahora utilizando el programa R.

En el capítulo 1 se expuso como instalar los programas R y R-Studio. Por lo tanto, a continuación se presentan las sentencias que se deben ingresar en el entorno de programación y a continuación la salida respectiva, que genera R.

2.4.2.1. Definir directorio de trabajo

Sentencias:

```
# (1) Establecer por defecto el directorio de trabajo donde está el archivo
de texto con los datos
setwd("~/Users/imac/Dropbox/RIRA/PhD/Publicaciones/Libros/UNEMI2017/Datos")
```

2.4.2.2. Carga de datos

Sentencias:

```
# Carga los datos
Datos <- read.csv("Ejercicio21_turismo.csv", header = TRUE, sep=";",
dec="," ,quote="\")

# toma solo los datos cuantitativos ITEM02, ITEM04, ITEM06,
# ITEM08, ITEM10, ITEM12, ITEM13 e ITEM14
Datos1 <- data.frame(Datos [1:360,c(3, 5, 7, 9, 11, 13, 14,15)])
```

2.4.2.3. Prueba Kaiser-Meyer-Olkin

Sentencias:

```
# Matriz de correlaciones (Rcor) y determinante de la matriz
Rcor <- cor(Datos1)
det (Rcor)

# Crea función para test KMO
kmo <- function(x)
{
  x <- subset(x,complete.cases(x)) # Omitir datos ausentes
  r <- cor(x) # Matriz de correlaciones
  r2 <- r^2 # Cuadrado de los coeficientes de correlación
  i <- solve(r) # Matriz inversa de la matriz de correlaciones
  d <- diag(i) # Elementos de la diagonal de la matriz inversa
  p2 <- (-i/sqrt(outer(d,d)))^2 # Parciales cuadrados de los coeficientes
  de correlación
  diag(r2) <-diag(p2) <- 0 # Borra los elementos de la diagonal
  KMO <- sum(r2)/(sum(r2)+sum(p2))
  MSA <- colSums(r2) / (colSums(r2)+colSums(p2))
  return(list(KMO=KMO, MSA=MSA))
}
kmo(Datos1)
```

Salida:

```
> kmo(Datos1)
$KMO
[1] 0.7502238

$MSA
  ITEM02  ITEM04  ITEM06  ITEM08  ITEM10  ITEM12  ITEM13  ITEM14
0.7370308 0.7053798 0.8264252 0.7489521 0.5936371 0.8145010 0.7498843 0.7457018
```

2.4.2.4. Prueba de Barlett

Sentencias:

```
# Test de esfericidad de Bartlett
install.packages("psych")
library(psych)
print(cortest.bartlett(Rcor, n=nrow(Datos1)))
```

Salida:

```
> print(cortest.bartlett(Rcor, n=nrow(Datos1)))
$chisq
[1] 422.4214

$p.value
[1] 5.331233e-72

$df
[1] 28
```

2.4.2.5. Análisis Factorial - Componentes principales

Sentencias:

```
# Análisis de los componentes principales (ACP)
install.packages("GPArotation")
library(GPArotation)

# ACP sin rotación nº1 (explorando numero de componentes)
CP_NoRotado1 <- principal(Datos1, nfactors = 8, rotate = "none", use=pairwise)
CP_NoRotado1
```

Salida:

```
> CP_NoRotado1 <- principal(Datos1, nfactors = 8, rotate = "none",
use=pairwise)
> CP_NoRotado1
Principal Components Analysis
Call: principal(r = Datos1, nfactors = 8, rotate = "none", use = pairwise)
Standardized loadings (pattern matrix) based upon correlation matrix
```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	h2	u2	com
ITEM02	0.65	0.14	0.43	-0.16	-0.21	-0.09	-0.45	-0.29	1	3.3e-16	3.7
ITEM04	0.62	0.18	0.43	-0.44	0.18	0.05	0.21	0.35	1	0.0e+00	4.2
ITEM06	0.48	-0.38	0.33	0.55	0.43	0.18	-0.07	0.01	1	6.7e-16	4.8
ITEM08	0.65	-0.11	-0.32	0.16	-0.48	0.27	-0.17	0.31	1	0.0e+00	3.8
ITEM10	0.21	0.79	0.14	0.46	-0.15	-0.06	0.26	-0.04	1	3.3e-16	2.2
ITEM12	0.66	-0.23	-0.23	0.09	0.03	-0.67	0.07	0.07	1	-2.2e-16	2.6
ITEM13	0.72	-0.22	-0.20	-0.15	-0.06	0.23	0.43	-0.36	1	1.1e-16	3.1
ITEM14	0.41	0.46	-0.55	-0.11	0.48	0.11	-0.25	-0.03	1	-2.2e-16	4.5

```
SS loadings
```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Proportion Var	2.63	1.15	1.00	0.8	0.75	0.63	0.61	0.44
Cumulative Var	0.33	0.14	0.13	0.1	0.09	0.08	0.08	0.06
Proportion Explained	0.33	0.14	0.13	0.1	0.09	0.08	0.08	0.06
Cumulative Proportion	0.33	0.47	0.60	0.7	0.79	0.87	0.94	1.00

```
Mean item complexity = 3.6
Test of the hypothesis that 8 components are sufficient.

The root mean square of the residuals (RMSR) is 0
with the empirical chi square 0 with prob < NA

Fit based upon off diagonal values = 1
```

Sentencias:

```
# ACP sin rotation nº2 (con 3 componentes decididos, para sacar matriz no
rotada)
CP_NoRotado2 <- principal(Datos1, nfactors = 3, rotate = "none", use=pairwise)
CP_NoRotado2
```

Salida:

```
> CP_NoRotado2 <- principal(Datos1, nfactors = 3, rotate = "none",
use=pairwise)
> CP_NoRotado2
Principal Components Analysis
Call: principal(r = Datos1, nfactors = 3, rotate = "none", use = pairwise)
Standardized loadings (pattern matrix) based upon correlation matrix
```

	PC1	PC2	PC3	h2	u2	com
ITEM02	0.65	0.14	0.43	0.63	0.37	1.8
ITEM04	0.62	0.18	0.43	0.61	0.39	2.0
ITEM06	0.48	-0.38	0.33	0.48	0.52	2.7
ITEM08	0.65	-0.11	-0.32	0.54	0.46	1.5

```
ITEM10 0.21 0.79 0.14 0.69 0.31 1.2
ITEM12 0.66 -0.23 -0.23 0.54 0.46 1.5
ITEM13 0.72 -0.22 -0.20 0.60 0.40 1.3
ITEM14 0.41 0.46 -0.55 0.68 0.32 2.8
```

```

                PC1  PC2  PC3
SS loadings      2.63  1.15  1.00
Proportion Var   0.33  0.14  0.13
Cumulative Var   0.33  0.47  0.60
Proportion Explained 0.55 0.24 0.21
Cumulative Proportion 0.55 0.79 1.00

```

Mean item complexity = 1.9

Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is 0.12
with the empirical chi square 294.96 with prob < 7.2e-60

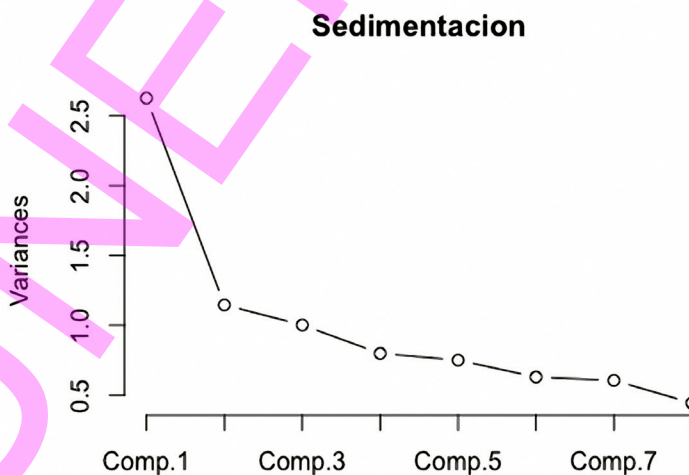
Fit based upon off diagonal values = 0.76

2.4.2.6. Gráfico de sedimentación

Sentencias:

```
# Otro elemento para decidir el número de factores, gráfico de sedimentación
Sedimentacion <- princomp(Datos1, scores=TRUE, cor=TRUE)
plot(Sedimentacion, type="lines")
```

Salida:



2.4.2.7. Análisis final con rotación

Sentencia:

```
# ACP con rotación varimax (con 3 componentes decididos)
CP_Rotado <- principal(Datos1, nfactors = 3, rotate = "varimax")
CP_Rotado
```

Salida:

```
> CP_Rotado <- principal(Datos1, nfactors = 3, rotate = "varimax")
> CP_Rotado
Principal Components Analysis
Call: principal(r = Datos1, nfactors = 3, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
```

	RC1	RC3	RC2	h2	u2	com
ITEM02	0.21	0.76	0.12	0.63	0.37	1.2
ITEM04	0.17	0.74	0.15	0.61	0.39	1.2
ITEM06	0.29	0.50	-0.38	0.48	0.52	2.5
ITEM08	0.72	0.15	0.08	0.54	0.46	1.1
ITEM10	-0.15	0.33	0.75	0.69	0.31	1.5
ITEM12	0.70	0.20	-0.05	0.54	0.46	1.2
ITEM13	0.73	0.27	-0.04	0.60	0.40	1.3
ITEM14	0.50	-0.12	0.65	0.68	0.32	1.9

	RC1	RC3	RC2
SS loadings	1.97	1.64	1.17
Proportion Var	0.25	0.20	0.15
Cumulative Var	0.25	0.45	0.60
Proportion Explained	0.41	0.34	0.25
Cumulative Proportion	0.41	0.75	1.00

Mean item complexity = 1.5
Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is 0.12
with the empirical chi square 294.96 with prob < 7.2e-60

Fit based upon off diagonal values = 0.76

Luego de procesados los datos, los pasos que se siguen a partir de este punto son los mismos que se siguieron al utilizar el programa informático SPSS y que ya fue descrito con anterioridad en este capítulo.

2.5. Resumen del capítulo

En este capítulo se conceptualiza la técnica multivariante Análisis Factorial, que en esencia se utiliza para reducir factores. Luego para ilustrar el uso de técnica se describe una investigación donde se aplicó un análisis factorial exploratorio.

Análisis factorial, se puede aplicar en cualquier campo, cuando el estudio lo que requiere es reducir factores o identificar los factores principales. Para ello en este capítulo se exponen los principales programas informáticos que permiten utilizar esta técnica.

Se termina el capítulo con la descripción de una investigación realizada sobre turismo, se describe los pasos seguidos utilizando el programa informático SPP y posteriormente con los mismo datos, se describe el uso del software R. Elemento importante al momento de aplicar análisis factorial es evaluar los resultados de las pruebas KMO y de Barlett, ya que estas determinan si se puede o no aplicar la técnica.

2.6. Referencias

- Abridged ACRA scale of learning strategies for Escala de estrategias de aprendizaje ACRA-Abreviada para alumnos universitarios Jesús de la Fuente Arias *, Fernando Justicia Justicia **. (2003), (January).
- Alonso, A. (2008). *Técnicas de análisis multivariante de datos*. (I. Capella & M. Diego, Eds.) (Primera). Madrid-España: PEARSON PRENTICE HALL.
- Arias, B. (2008). SEM, Structural Equation Models),.
- Arteaga-arcentales, E., Peralta-carpio, F., Castelo-gonzalez, J., & Echeverría-vásquez, H. (2016). Principales Factores para Desarrollo Turístico : Estudio Empírico a la Percepción de la Población Económicamente Activa. PODIUM.
- Arteaga, R., & Duarte, A. (2010). Motivational factors that influence the acceptance of Moodle using TAM. *Computers in Human Behavior*, 26(6), 1632–1640. <http://doi.org/10.1016/j.chb.2010.06.011>
- Camprubí, R., Guia, J., & Comas, J. (2009). La formación de la imagen turística inducida: Un modelo conceptual. *Pasos, Revista de Turismo Y Patrimonio Cultural*, 7, 255–270.
- Cuadras, C. M. (2014). *Nuevos Métodos de Análisis Multivariante*. (CMC Editions, Ed.) CMC Editions (Primera). Barcelona- España: CMC Editions. <http://doi.org/10.1017/CBO9781107415324.004>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. <http://doi.org/10.1037/1082->

989X.4.3.272

- Findik, D., & Kunçay, C. O. Ş. (2009). A model for instructors adoption of learning management systems : empirical validation in higher education context. Middle East Technical University.
- Frías-Navarro, D., & Soler, M. P. (2012). Prácticas Del Análisis Factorial Exploratorio (Afe) En La Investigación Sobre Conducta Del Consumidor Y Marketing Exploratory Factor Analysis (Efa) in Consumer Behavior and Marketing Research. *Suma Psicológica*, 19, 47–58. <http://doi.org/ISSN 0121-4381>
- Hair, J. F., Black, W. C., & Babin, B. J. (2010). *Multivariate Data Analysis: A Global Perspective*. Pearson Education.
- Hair, Anderson, R., Tatham, R., & Black, W. (2008). *Análisis Multivariante*. (A. Otero, Ed.) (Quinta). Madrid-España: Prentice Hall Iberia.
- Lay, J.-G., Chen, Y.-W., & Chi, Y.-L. (2013). GIS Adoption Among Senior High School Geography Teachers in Taiwan. *Journal of Geography*, 112(3), 120–130. <http://doi.org/10.1080/00221341.2012.682226>
- Lin, W.-S. (2012). Perceived fit and satisfaction on web learning performance: IS continuance intention and task-technology fit perspectives. *International Journal of Human-Computer Studies*, 70(7), 498–507. <http://doi.org/10.1016/j.ijhcs.2012.01.006>
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151–1169. <http://doi.org/10.6018/analesps.30.3.199361>
- Manuel Batista-Foguet, J., Coenders, G., & Alonso, J. (2004). Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud. *Medicina Clínica*, 122(Supl.1), 21–27. <http://doi.org/10.1157/13057542>
- Martínez, M. Á. (2014). Bioestadística amigable (Díaz De Sa).
- Mbarek, R., & Zaddem, F. (2013). The examination of factors affecting e-learning effectiveness. *International Journal of Innovation and Applied Studies*, 2(4), 423–435. <http://doi.org/2028-9324>
- Ministerio de Turismo. (2015). Proyecto Ecuador potencia turística. Quito - Ecuador.
- Morales Vallejo, P. (2011). Guía para construir cuestionarios y escalas de actitudes. *Revista Chilena de Obstetricia Y Ginecología*, 9(6), 80. <http://doi.org/10.4067/S0717-75262010000600009>
- Naciones Unidas. (1998). Turismo sostenible: contribución del turismo al crecimiento económico y al desarrollo sostenible (Vol. 51048).
- Pérez, E. R., & Medrano, L. (2010). Análisis factorial exploratorio : Bases conceptuales y metodológicas. *Revista Argentina de Ciencias Del Comportamiento*, 2(1889), 58–66. Retrieved from http://www.academia.edu/12607324/Revista_Argentina_de_Ciencias_del_Comportamiento_

RACC_Análisis_Factorial_Exploratorio_Bases_Conceptuales_y_Metodológicas_Artículo_de_Revisión

- Ramírez Anormaliza, R. (2013). Evaluación de los sistemas e-learning: estudio de las publicaciones realizadas en la web of knowledge, 1–20.
- Sanjuán Suárez, P., Pérez García, A. M., & Bermúdez Moreno, J. (2000). Escala de autoeficacia general: Datos psicométricos de la adaptación para población española. *Psicothema*, 12(SUPPL. 2), 509–513. <http://doi.org/ISSN 0214-9915>
- Tarhini, A., Hone, K., & Liu, X. (2013). Factors Affecting Students' Acceptance of e-Learning Environments in Developing Countries : A Structural Equation Modeling Approach. *International Journal of Information and Education Technology*, 3(1), 54–59. <http://doi.org/10.7763/IJJET.2013.V3.233>
- Vázquez-Cano, E., & García Sevillano, M. L. (2015). Analysis of risks in a Learning Management System : A case study in the Spanish National University of Distance Education (UNED). *Journal of New Approaches in Educational Research*, 4(1), 62–68. <http://doi.org/10.7821/naer.2015.1.107>
- Weng, C., & Tsai, C. (2015). Social support as a neglected e-learning motivator affecting trainee ' s decisions of continuous intentions of usage. *Australasian Journal of Educational Technology*, 31(2), 177–192.

UNENMI 2017

“La tecnología, como el arte, es un ejercicio altísimo de la imaginación humana”

Daniel Bell (1919-2011)
Sociólogo estadounidense

Capítulo 3: Análisis de conglomerados

Fabrizio Guevara-Viejo
Mayra D'Armas Regnault
David Vargas-Decimavilla
Roberto Basurto- Quilligana

Objetivos de aprendizaje

El Análisis de Conglomerados, también conocido con Análisis Cluster es una técnica multivariante, exploratoria y descriptiva, que tiene por objeto clasificar una muestra de variables en grupos que sean muy similares entre sí y a la vez que sean distintos del resto. A partir de una tabla de casos-variables, este análisis trata de situar todos los casos en grupos homogéneos (conglomerados o clusters) no conocidos de antemano pero sugeridos por la propia esencia de los datos, de manera que individuos que puedan ser considerados similares sean asignados a un mismo cluster, mientras que individuos diferentes (disimilares) se sitúen en clusters distintos (Alonso, 2008).

El Análisis de Conglomerados se diferencia del análisis discriminante en que el establecimiento de los grupos se realiza a posteriori, ya que el investigador no tiene conocimiento de la existencia de los subgrupos o conglomerados, ni del número resultante, ni mucho menos de las características que los definen (Vilà-baños, Rubio-Hurtado, Berlanga-Silvente, & Torrado-Fonseca, 2014).

Al finalizar este capítulo deberá ser capaz de lo siguiente:

- Entender la importancia de clasificar objetos.
- Comprender el proceso y la medición de la similitud entre objetos.
- Conocer las diferentes técnicas de análisis de conglomerados.
- Entender la diferencia entre un método jerárquico y no jerárquico.
- Conocer el funcionamiento del análisis de conglomerados.
- Conocer algunas herramientas informáticas que permiten realizar análisis de conglomerados.
- Observar lineamientos principales de validación del análisis de conglomerados.

3.1. Introducción

3.1.1. Propósito de los agrupamientos

En las investigaciones realizadas por profesionales en diferentes ramas del estudio, se encuentran a menudo que para resolver los problemas se parte de una definición de grupos de tratamiento homogéneos, sean estos productos, personas o empresas. Partir de una definición de grupos de estudios que sean diferentes entre ellos para buscar soluciones a las situaciones que se encuentren en ámbitos tanto académicos como de investigación, puede conllevar a resultados equivocados. La interrogante sería ¿dónde ocurrió el error que conllevó a estos resultados? La respuesta es simple, se asumieron grupos de estudios que no fueron definidos técnicamente.

Pensando en el mundo del marketing, la identificación en la población, la segmentación y el posicionamiento de los productos no fueran posibles sin el uso de una metodología objetiva. Así como en la Psicología cuando se desea determinar con claridad los perfiles psicológicos o analizarlos, en la Biología para la clasificación de alguna especie animal; y así entre otras áreas.

El académico o investigador en las situaciones expuestas, se encuentra en la búsqueda de una composición fundamentada entre los datos basados en múltiples variables; es decir, de un análisis multivariante. A continuación, se muestran en la **Tabla 3 - 1** algunos propósitos del proceso de realizar agrupamientos por campo o área de estudio.

Tabla 3 - 1 Propósitos del proceso de realizar agrupamientos

Campo o Área de estudio	Propósito
Marketing	Segmentación del mercado para identificar perfiles de consumidores. Identificación de productos similares o clientes con características similares.
Psicología	Identificación de perfiles psicológicos. Incidencia de varios factores en el comportamiento de un conjunto de individuos.
Biología	Clasificación de animales y plantas (taxonomía numérica).
Medicina	Identificación de síndromes a partir de los síntomas. Detección precoz de posibles focos epidémicos.
Arqueología	Clasificación de los ítems arqueológicos y construcción de tipologías precisas.

3.1.2. Objetivo del análisis de conglomerados

Una vez entendido que los investigadores, indistinto de su campo de estudio, buscan agrupar para luego analizar. Lo primero que hay que entender, es que se trata de una técnica multivariada (más de dos variables) cuyo objetivo principal es agrupar objetos (productos o entidades) basándose en las características que estos poseen; el análisis en sí, clasifica.

Clasificar los elementos de un conjunto finito de datos consiste en realizar una partición del conjunto en subconjuntos homogéneos, siguiendo un determinado criterio de clasificación, donde cada elemento pertenece a un único subconjunto que a menudo tiene un nombre que lo caracteriza. Como ejemplos de nuestra vida diaria podemos mencionar que clasificamos: las personas en hombres y mujeres; los animales en especies, géneros, familias y órdenes (Cuadras, 2014).

La agrupación se centra, en clasificar un objeto muy similar con otro en función a un criterio de selección, este criterio que en muchos casos se conocerá como una medida, es una medida establecida previamente. Los conglomerados de aquellos objetos seleccionados entre sí con similares características deberían mostrar un alto grado de homogeneidad entre ellos; y un alto grado de heterogeneidad de manera externa con otros conglomerados. Es decir, la técnica busca formar grupos cuyos objetos integrantes sean similares y al mismo tiempo cada grupo sea diferente sus miembros que las componen. Gráficamente, podemos observarlo de manera esquemática en la **Figuras 3 - 1 y 3 - 2**.

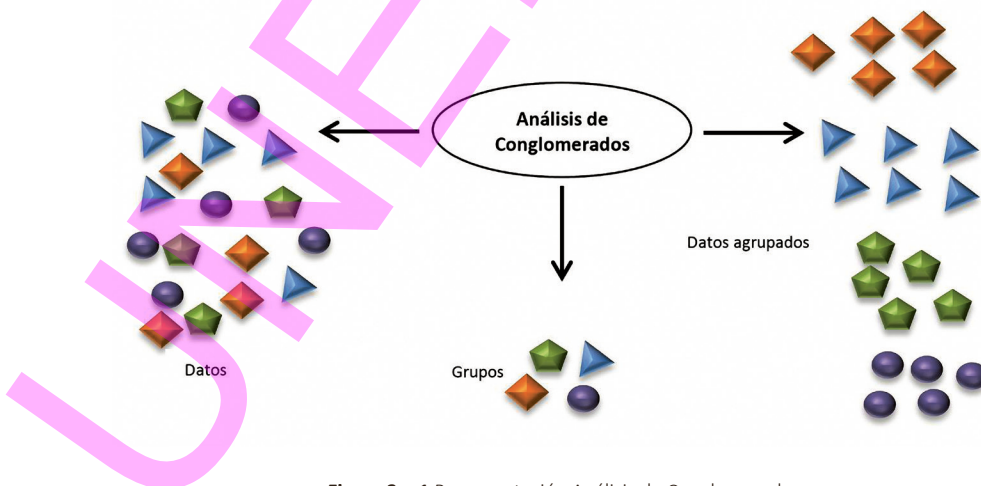


Figura 3 - 1 Representación Análisis de Conglomerados

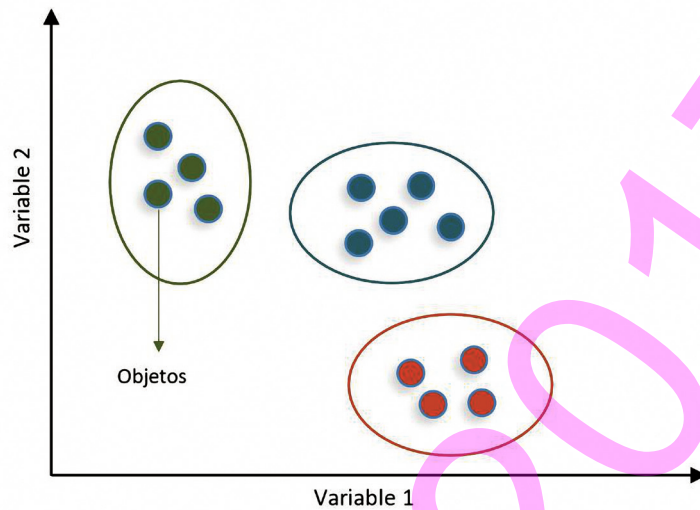


Figura 3 - 2 Representación gráfica del análisis de conglomerado (tres grupos)

El análisis de conglomerados constituye una importante técnica de análisis de datos (Leyva & Olague, 2014), ampliamente utilizada en distintas áreas de conocimiento (Biología, Psicología, Arqueología, Sociología, etc.) con el propósito de identificar entidades similares a partir de las características que poseen (Esquivel Guerrero, Contreras Cortes, Molina González, & Capel Martínez, 1991). Este análisis también es conocido bajo otras denominaciones como análisis clúster, análisis Q, análisis de clasificación y taxonomía numérica. Estas terminologías dependen del campo de estudio o disciplinas diversas donde se desarrolle, tales como Psicología, Sociología, Biología, Economía, Ingeniería y los negocios (Gómez Suárez, 2002). En fin, aunque los nombres o términos sean diferentes, la metodología tiene la misma misión que se centra en clasificar y agrupar objetos.

El fundamento del análisis de conglomerado se centra en la clasificación de los datos. De las técnicas de análisis multivariante es la única que no estima el valor teórico sino que utiliza el valor especificado por el investigador. El objetivo de la técnica es la comparación de los objetos basándose en el valor teórico, no es su estimación en si misma. Este análisis agrupa objetos (observaciones, eventos) basados en la información encontrada en los datos que describen los objetos o sus relaciones. El objetivo es que los objetos dentro de un grupo sean similares (o relacionados) y distintos (o no relacionados) de los objetos de otros grupos. Cuanto mayor sea la semejanza (u homogeneidad) dentro de un grupo y cuanto mayor sea la disparidad entre grupos, es mejor el agrupamiento (Sharma, Bajpai, & Litoriya, 2012).

El análisis de conglomerados es realmente una herramienta útil, que se compara con el análisis factorial con el objetivo de evaluar la composición o estructura. Sin embargo, estos dos análisis difieren entre sí dado que el análisis de conglomerados agrupa objetos y el análisis factorial agrupa variables. Si buscamos un ejemplo del concepto fundamental de la técnica, se puede describir como hecho histórico, al esfuerzo de Aristóteles que planteó un sistema de clasificación de especies del reino animal, clasificándolos en dos grupos principales: animales que tienen y que no tienen sangre roja; es decir, vertebrados versus invertebrados. Subdividiendo además esos dos grupos en ovíparos y vivíparos, de acuerdo con la forma en la que los descendientes vienen al mundo.

Finalmente, el análisis de conglomerados se caracteriza por ser una técnica descriptiva, ateorica y no inferencial. No tiene una base estadística de deducir inferencias para una población a partir de una muestra, su fundamento es el uso como técnica exploratoria. Dada esta caracterización, el análisis de conglomerado no muestra una solución única, pueden ser varias y depende mucho del investigador.

3.2. Fundamento técnico

El diseño del análisis por conglomerados, como una de las tantas técnicas estadísticas multivariantes, abarca los procesos de partición de un conjunto de datos para formar conglomerados, la interpretación de estos conglomerados y la validación de los resultados. El primer proceso (partición de los datos) tiene que ver cómo se pueden desarrollar los conglomerados; el segundo proceso se relaciona con entender las características de cada conglomerado y definir un nombre que represente su naturaleza. El proceso final que es la validación de los resultados tiene que ver con determinar su estabilidad y generalidad, junto con las características de cada conglomerado.

El objetivo principal del análisis de conglomerados es obtener un conjunto de objetos en dos o más grupos basándose en características comunes (homogéneos). Cuando se forman estos grupos logran varios objetivos (J. Hair et al., 2010):

1. Ser capaz de describir una taxonomía. El análisis de conglomerado aparte de servir para la formulación de una taxonomía (clasificar objetos de manera empírica), también puede ser utilizado para argumentar una hipótesis relacionada con la estructura de los objetos. Por lo tanto, si una estructura se define en un conjunto de objetos, se puede utilizar para comparar entre una tipología propuesta y la derivada del análisis de conglomerados.

2. Simplificar los datos. Si se posee una estructura definida, las observaciones pueden agruparse para poder realizar un análisis ulterior. A diferencia del análisis factorial donde se precisa proporcionar dimensiones, en el análisis de conglomerados hacen los mismos, pero para las observaciones, entonces, en lugar de ver todas las observaciones como únicas, se pueden considerar como miembros de un conglomerado.
3. Identificación de relación. Una vez que se tienen definidos los conglomerados y la estructura de los datos presente en los conglomerados, se posee un medio para descubrir las relaciones entre observaciones que no se pudieran realizar con las observaciones individuales. A diferencia de otros análisis como el de discriminante, el análisis de conglomerados muchas veces puede mostrar las similitudes y diferencias que no se habían descubierto previamente.

3.2.1. Selección de variables del análisis de conglomerados

La selección de las variables para el análisis de conglomerados debe basarse en consideraciones teóricas, conceptuales y prácticas, por lo tanto cualquier aplicación de éste análisis se debe basar o apoyar en la lógica en función de la cual van a ser seleccionadas las variables. Se debe tener en cuenta que si la lógica se basa tanto en una teoría como en una investigación pasada o en una suposición, se deben incluir sólo aquellas variables que caractericen los objetos que se están agrupando y que se refieran únicamente a los objetivos del análisis de conglomerados.

Es de resaltar que el análisis de conglomerados no posee un medio para diferenciar las variables relevantes de las irrelevantes, por tal motivo éste análisis puede verse muy afectado cuando se incluyan dos o más variables inapropiadas.

3.2.2. Diseño de investigación mediante análisis de conglomerados

Una vez que se han definido los objetivos y se han seleccionado todas las variables, se deben tener en cuenta tres condiciones antes de iniciar el proceso de partición. Estas tres condiciones son:

- Identificación y descarte de los atípicos
- Medición de la similitud de los objetos
- Estandarización de los datos

Para responder estas interrogantes existen muchos enfoques, pero ninguno ha sido evaluado suficientemente, por lo tanto se puede utilizar una combinación

alternante entre el análisis de conglomerados y el análisis factorial.

3.2.2.1. Detección de Atípicos

En contexto, el análisis de conglomerados es sensible a la inclusión de variables irrelevantes y también a los atípicos, que no son otra cosa más que objetos que son diferentes del resto. Los atípicos pueden representar:

Observaciones verdaderamente absurdas que no son representativas de la población general.

Una muestra reducida del grupo (grupos) de la población que provoca una mala representación del grupo (grupos) de la muestra.

En ambos casos, los atípicos distorsionan la estructura y hacen que los conglomerados obtenidos no sean representativos de la población. Debido a esto se sugiere que estos atípicos sean representados gráficamente mediante un diagrama de perfil, en donde las variables se encontrarán en el eje de las "x" y los valores de las variables en el eje de las "y". Obviamente el proceso de visualización del diagrama de perfil se hará mucho más engorroso cuando se tenga una gran cantidad de variables.

3.2.2.2. Medidas de Similitud

La creación de grupos basados en similitud de casos exige una definición de similitud, es decir de distancia entre individuos. Estas distancias pueden ser medidas de muchas formas ya que existen diferentes reglas matemáticas para asignar los individuos a distintos grupos, dependiendo del fenómeno estudiado y del conocimiento previo de posible agrupamiento que se tenga (Alonso, 2008).

La similitud entre objetos es una medida de correspondencia o parecido entre los objetos que van a ser agrupados. Casi al igual que en el análisis factorial donde se crea una matriz de correlación entre las variables que fueron utilizadas para agrupar estas variables en factores, en el análisis de conglomerados las características que definen similitud se especifican en primer lugar y luego se combinan las características en una medida de similitud calculada para todos los objetos, al igual que se utilizan las correlaciones en el análisis factorial. Existen varias medidas de similitud, siendo las principales las siguientes:

- Medidas de Correlación
- Medidas de Distancia
- Medidas de Asociación

Medidas de Correlación. El coeficiente de correlación es una de las medidas de similitud con la que más fácilmente se puede hacer el análisis de conglomerados. En este caso se invierte la matriz de todas las x variables que se tengan de tal manera que las columnas representan los objetos y las filas represente las variables. Con esta configuración se tiene la correlación entre las dos columnas de números es la correlación (o similitud) entre los perfiles de los dos objetos. Así se tiene que altas correlaciones indican similitud y bajas correlaciones indican la falta de ella.

Medidas de Distancia. Las medidas de similitud de distancia, representan la medida de similitud en forma de la proximidad de las observaciones respecto a las otras para las variables del valor teórico del análisis por conglomerados. Las medidas de similitud son en realidad medidas de diferencia, donde elevados valores indican una menor similitud.

La medida de distancia más utilizada es la distancia Euclídea. Para entender que es la distancia Euclídea, supongamos que tenemos dos puntos de dos dimensiones que tienen coordenadas (x_1, y_1) y (x_2, y_2) . La distancia Euclídea entre los puntos es la longitud de la hipotenusa de un triángulo rectángulo (**Figura 3 – 3**). Adicionalmente, existe la distancia Euclídea cuadrada (absoluta) que es la suma de diferencias al cuadrado sin tomar en cuenta la raíz cuadrada, la cual tiene la ventaja de no tomar la raíz cuadrada lo que acelera notablemente los cálculos.

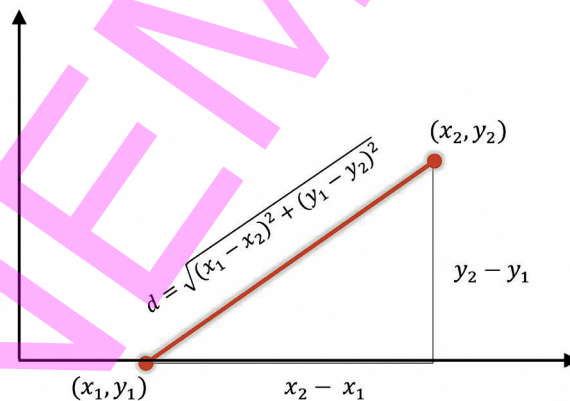


Figura 3 -3 Ilustración de la fórmula distancia Euclídea

3.2.2.3. Medidas de Asociación

Esta medida de similitud se utiliza para comparar objetos cuyas características se pueden medir en términos no métricos. Un ejemplo que se puede utilizar para entender puede ser el caso de los encuestados que respondan sí o no

a cierto número de preguntas. La medida de asociación podría evaluar el grado de acuerdo entre cada par de encuestados. Según esto una medida de asociación sería el porcentaje que hubo de acuerdos para el mismo conjunto de preguntas.

3.2.2.4. Estandarización de los datos

Una vez que se ha escogido la medida de similitud, sólo se debe conocer una cuestión más: ¿deberían tipificarse los datos antes de calcular las similitudes? Se debe tener en cuenta si la mayoría de las medidas son bastantes sensibles a las diferentes escalas o magnitudes de las variables. Se recomienda trabajar con datos que estén estandarizados y de esta forma eliminar las distorsiones por efectos de las diferencias de escalas de medida. Es decir, el análisis se debe realizar sobre variables que presenten similares valores medios y desviaciones estándar de forma tal que se facilite la interpretación.

Al aplicar el análisis de conglomerados, se debe tener muy en cuenta que la muestra tomada debe ser representativa de la población. Como se mencionó anteriormente, los atípicos pueden ser resultados de una muestra de grupos divergentes que pueden introducir sesgo en la estimación de la estructura.

3.3. Clasificación de las técnicas

Las clasificaciones pueden ser jerárquicas o no jerárquicas. La **Figura 3 - 4** muestra un resumen de los diferentes métodos de análisis de conglomerados.



Figura 3 - 4 Descripción de las diferentes métodos de análisis de conglomerados

3.3.1. Métodos Jerárquicos

Los métodos jerárquicos tiene como funcionalidad agrupar los conglomerados para formar uno nuevo o bien separar a los existentes para que den origen a otros, con el objetivo de minimizar o maximizar alguna medida de distancia (regla o función). Como su nombre lo indica, en base a una jerarquía el método construye o destruye conglomerados dependiendo del objetivo final de minimizar o maximizar una función de similitud.

Según Cuadras (2014), una clasificación jerárquica consiste de una sucesión de conglomerados tal que cada conglomerado se obtiene agrupando conglomerados. Por ejemplo, si tenemos un conjunto de datos formado $\Omega = \{a, b, c, d, e\}$, una clasificación jerárquica es:

$$\begin{aligned}\Omega &= \{a\} + \{b\} + \{c\} + \{d\} + \{e\} \\ \Omega &= \{a, b\} + \{c, d\} + \{e\} \\ \Omega &= \{a, b\} + \{c, d, e\} \\ \Omega &= \Omega\end{aligned}$$

Los métodos de análisis de conglomerados jerárquicos se dividen en: (1) Métodos aglomerativos; y (2) Métodos disociativos. En la **Tabla 3 - 2** se muestra de manera resumida el propósito de cada una de las técnicas y en donde el investigador en función del método de cada uno de ellos y del objetivo de la investigación podrá determinar su uso.

Los métodos aglomerativos tienen como función principal formar conglomerados a partir de una solución inicial de grupos que equivalen al mismo número de objetos (individuos u observaciones), hasta que al finalizar el proceso todos pertenezcan a un mismo grupo. En la trayectoria del proceso, se forman grupos y subgrupos (conglomerados que se pertenecen hasta llegar a uno solo). Dentro de los métodos aglomerativos, se puede indicar principalmente que la estrategia se basará en minimizar distancias o maximizar similitudes. Las técnicas más utilizadas son las siguientes: (a) Método del amalgamamiento simple; (b) Método del amalgamamiento completo; (c) Método del promedio entre grupos; (d) Método del centroide; (e) Método de la mediana; y (f) Método de Ward.

Los métodos disociativos en comparación a los aglomerativos funcionan de manera contraria, se trata de un proceso divisor que inicia con un conglomerado que encierra a todos los objetos o individuos y que a través de sucesivas divisiones van formando conglomerados cada vez mas pequeños hasta que éstos sean igual al total de observaciones que se estén estudiando. Dentro de los métodos disociativos aparte de los métodos aglomerativos pero

con la estrategia de maximizar distancias o minimizar similitudes. En adición a las mencionadas, se tiene las siguientes: (a) El análisis de asociación; (b) El detector automático de interacción.

Tabla 3 - 2 Descripción de los métodos jerárquicos

Método Jerárquico	Finalidad	Metodología
Método del amalgamamiento simple	Aglomerativo	En su terminología en inglés single linkage. La técnica se basa en que considera a la distancia o similitud entre dos conglomerados viene representada por la mínima distancia o máxima similitud. El cálculo es determinado a partir de la matriz de distancia inicial.
Método del amalgamamiento completo	Aglomerativo	En su terminología en inglés complete linkage. La técnica se basa en que considera a la distancia o similitud entre dos conglomerados hay que medirla por sus elementos más dispares; es decir, viene dada por la máxima distancia o mínima similitud. El cálculo es determinado a partir de la matriz de distancia inicial.
Método del promedio entre grupo	Aglomerativo	Existen dos técnicas una es usando el promedio no ponderado (unweighted arithmetic average) y la segunda es el promedio ponderado (Weighted arithmetic average). En la primera, la distancia o similitud se obtiene como la media aritmética entre la distancia o similitud de las componentes del conglomerado. El método no toma importancia el tamaño de elementos que contenga los conglomerados y por ende todos tienen igual importancia en la distancia. En la segunda, la distancia o similitud se obtiene como la media aritmética ponderada entre la distancia o similitud de las componentes del conglomerado.
Método del centroide	Aglomerativo	En este método, la técnica trata de la semejanza en los centroides entre dos conglomerados. La centroide es la media del vector de medidas de los objetos en cada conglomerado. También es conocido como método de centroide ponderado.
Método de la mediana	Aglomerativo	Es el mismo principio del método de centroide siendo la diferencia que no se realiza una ponderación en el cálculo de la media de los objetos en el conglomerado.
Método de Ward	Aglomerativo	Este método utiliza la distancia Euclídea. Su procedimiento es que en cada etapa se unen dos conglomerados que serían los que tengan el mínimo incremento en el valor total de la suma cuadrática de la diferencias de cada grupo, de cada objeto a la centroide.

Continúa

Tabla 3 - 2 Continuación

Método Jerárquico	Finalidad	Metodología
Análisis de asociación	Disociativo	Técnica utilizada para dividir grupos en objetos cuyas variables son nominales. Aplican medidas como el coseno del ángulo de dos vectores o medida de parejas simples, etc.
Detector automático de interacción	Disociativo	Se trata de una técnica para determinar relaciones entre variables explicativas y variables explicadas, de manera muy similar a los modelos de regresión lineal. El uso de la varianza toma también un papel importante (razón de verosimilitud), para seleccionar la variable que permite separar de mayor manera los grupos. En resumen, divide a la población en dos o más grupos distintos basados en las categorías de la variable dependiente que es un mejor predictor.

3.3.2. Métodos no Jerárquicos

Hemos observado que los métodos jerárquicos parten de un funcionamiento de agrupar conglomerados a uno solo o dividir un conglomerado en varios, pero manteniendo una secuencia dada. En el caso de los métodos no jerárquicos, no contemplan esa funcionalidad.

Los métodos no jerárquicos, tiene como función realizar solo una partición de los objetos en K grupos o conglomerados y no de manera secuencial como los métodos jerárquicos. Lo crucial en el uso de estas técnicas, viene dado por la elección del investigador en determinar el número de conglomerados que a priori desea conformar en su análisis. En esto radica la diferencia fundamental de su uso. El proceso de asignación de los objetos se hace mediante una optimización del criterio de selección y otra diferencia principal es que no se trabaja con una matriz de distancias o de similitudes sino con los datos originales.

Los métodos no jerárquicos pueden clasificarse en los grupos siguientes: (1) Métodos de reasignación; (2) Métodos de búsqueda de la densidad; (3) Métodos directos; y (4) Métodos de reductivos. De los métodos de reasignación se dispone: (a) K –Medias; (b) Quick-Cluster análisis; (c) Forgy; y (d) Nubes dinámicas. De los métodos de búsqueda de la densidad se dispone: (a) Análisis modal de Wishart; (b) Taxmap; y (c) De Fortin. De los métodos directos el más conocido es el Block-Clustering; y de los métodos de reductivos el más comúnmente usado es el Análisis Factorial tipo Q.

En la **Tabla 3 - 3** se muestra de manera resumida el propósito de algunos de los métodos no jerárquicos.

Tabla 3 - 3 Descripción de los métodos no jerárquicos

Método no Jerárquico	Finalidad	Metodología
K –Medias	Reasignación	Se trata de la aplicación de un algoritmo que previo a K conglomerados prefijados, el proceso de asignar cada objeto al conglomerado (grupo o clúster) con el centroide más próximo. En este método propuesto por MacQueen en el año 1972, la clave radica en que el centroide se calcula a partir de los miembros del clúster tras cada asignación y no al final de cada ciclo.
Forgy	Reasignación	El método Forgy elige aleatoriamente k objetos del conjunto de datos y las utiliza como centroides iniciales.
Nubes dinámicas	Reasignación	Es un método de clasificación de objetos difuso, que permite a cada objeto o individuo pertenecer a diferentes clases al mismo tiempo en base a una función de pertenencia muy similar a las utilizadas los conjuntos difusos.
Análisis modal de Wishart	Búsqueda de la densidad	Es un método que utiliza la distribución de Wishart para encontrar los grupos de alta densidad a un nivel de densidad x elegido por el investigador.
Block-Clustering	Directo	Es un técnica que se fundamenta en el uso de la varianza, en donde la agrupación es en bloques, la misma que trabaja simultáneamente la agrupación con las filas y la columnas de una matriz.
Análisis Factorial tipo Q	Reductivo	Se trata de una técnica que no fundamenta la similitud o distancia de objetos, sino la estructura de la red o la topología de manera independiente. Muchos autores consideran la técnica como un análisis conglomerado estructural.

3.4. Funcionamiento del Análisis de Conglomerados

El funcionamiento del análisis de conglomerados, depende de un proceso que primero necesita identificar las observaciones mas parecidas en grupos. Este proceso responde a: (1) Medir la similitud de las observaciones; (2) Formar los conglomerados; y (3) Cuantificar los conglomerados.

Medir la similitud, requiere de un método que mida simultáneamente las observaciones en todas las variables. La técnica que se utilice, sea correlación entre objetos, medida de asociación o alguna medida de proximidad utilizada. No importa cómo se mida la similitud, lo importante es medir previo a la

agrupación. La formación de los conglomerados, será un procedimiento de agrupar aquellas observaciones que son más similares dentro de un conglomerado, este proceso determinará la pertenencia al grupo de cada observación. La determinación del número de conglomerados o grupos, es una tarea que el investigador debe enfrentar, ya que debe considerar que si el número de grupos es menor necesariamente la homogeneidad dentro del grupo disminuye. Por lo tanto, se debe buscar un punto medio o equilibrio de grupos en donde se permita conseguir una mayor similitud de los objetos dentro del grupo (homogeneidad) y mayor diferencia entre ellos (heterogeneidad).

El análisis de conglomerados comienza con el cálculo de la matriz de distancias entre los elementos de la muestra. Esa matriz contiene las distancias existentes entre cada elemento y todos los restantes de la muestra. Posteriormente se buscan los 2 elementos más próximos (es decir, los 2 más similares en términos de distancia) y se agrupan en un conglomerado. El conglomerado resultante es indivisible a partir de ese momento (de ahí el nombre de jerárquico asignado al procedimiento). De esta manera, se van agrupando los elementos en conglomerados cada vez más grandes y más heterogéneos entre ellos, hasta llegar al último paso, en el que todos los elementos muestrales quedan agrupados en un único conglomerado global (Pérez Hernández, Lara Gómez, & Gómez Hernández, 2017).

Tabla 3 - 4 Conjunto de datos observados

Variables	Observaciones (repuestas)						
	A	B	C	D	E	F	G
Variable 1	3	4	4	2	6	7	6
Variable 2	2	5	7	7	6	7	4

Nota: Tomado del libro *Multivariate Data Analysis* por Hair, J., Black, W., Babin, B., & Anderson, R. (2010), p 511

Para ilustrar el funcionamiento de análisis de conglomerados, se utilizará un ejemplo de 6 observaciones en dos variables de estudio. Este ejercicio es tomado de Hair et al., (2010), el cual consiste en una investigación para determinar los segmentos de mercado en función de los patrones de preferencias que tienen los compradores a las marcas y a las tiendas. La muestra estuvo constituida por siete encuestados, representando la Variable 1 la preferencia a la tienda y la Variable 2 la preferencia a la marca, las cuales fueron medidas en una escala de Likert de 0 al 10.

Se considera la **Tabla 3-4** que muestra las respuesta de los encuestados. Además, la representación gráfica de los valores de cada encuestado en las dos variables se muestra en la **Figura 3-5**.

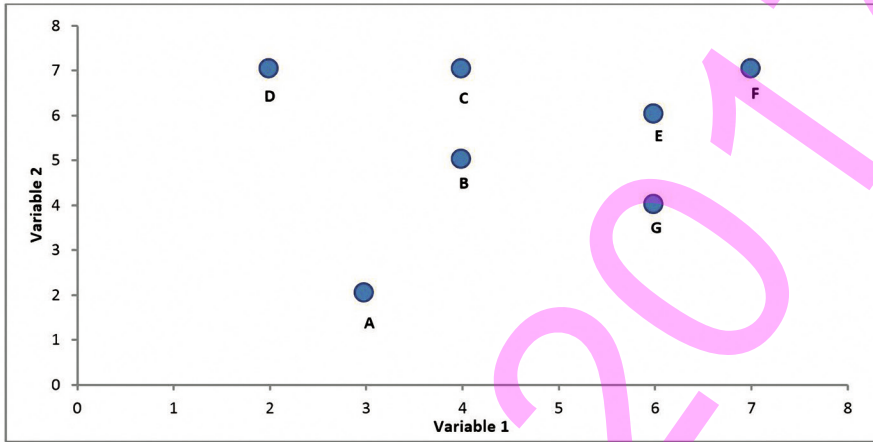


Figura 3 - 5 Diagrama de dispersión (variable 1 vs variable 2)

Para el ejemplo se utilizará la distancia Euclídea (una línea recta) entre un par de observaciones (desde la A hasta la G). Es decir, la medida de similitud.

Tabla 3 - 5 Matriz de proximidad (Distancia Euclídea)

Observación	Observación						
	A	B	C	D	E	F	G
A	-						
B	3,162	-					
C	5,099	2,000	-				
D	5,099	2,828	2,000	-			
E	5,000	2,236	2,236	4,123	-		
F	6,403	3,606	3,000	5,000	1,414	-	
G	3,606	2,236	3,606	5,000	2,000	3,162	-

Nota: Tomado del libro Multivariate Data Analysis por Hair, J., Black, W., Babin, B., & Anderson, R. (2010), p 512

En la aplicación de uno de los métodos por conglomerados existente (Aglomerativo jerárquico) se obtiene el siguiente cuadro resultante del como se conforman todos los conglomerados:

Tabla 3 - 6 Tabla descriptiva de la aplicación del Método Aglomerativo Jerárquico

Paso	Proceso de agrupación		Solución del conglomerado	
	Distancia euclídea	Par de observaciones	Pertenencia al conglomerado	Número de conglomerados
0	Solución Inicial	-	A – B – C – D – E – F – G	7
1	1,414	E-F	A – B – C – D – EF – G	6
2	2,000	E-G	A – B – C – D – EFG	5
3	2,000	C-D	A – B – CD – EFG	4
4	2,000	B-C	A – BCD – EFG	3
5	2,236	B-E	A – BCDEFG	2
6	3,162	A-B	ABCDEFG	1

Nota: Tomado del libro *Multivariate Data Analysis* por Hair, J., Black, W., Babin, B., & Anderson, R. (2010), p 513

Sin entrar en profundizar en el método, se puede describir de manera general como el mismo va agrupando y creando los grupos (conglomerados). Si revisamos la **Tabla 3 - 5** matriz de proximidad, la menor distancia es 1,414 que corresponde a la relación entre la observación E-F (ver **Figura 3 - 6** de relación E-F de color amarillo “conglomerado #1”). Como se trabaja con el método de similitud distancia Euclídea y se busca la similitud entre grupo; la mejor medida es el valor menor de todas las aproximaciones (1,414).

Los siguientes pares de observaciones más cercanos, es decir con menor distancia, son las observaciones E-G, CD y B-C con igual distancia de 2,000. La siguiente menor distancia es 2,236 que corresponde a las observaciones B-E, B-G, C-E. Utilizamos sólo una de estas distancias B-E ya que cada pareja de observaciones contiene un miembro que ya está incluido en los conglomerados existentes. Finalmente, encontramos que la menor distancia de la observación A es 3,162 para la pareja de observaciones A-B.

Si observamos simultáneamente entre la columna “Pertenencia al conglomerado” de la **Tabla 3 - 6** y la **Figura 3 - 6** se observa en esta última como el método va seleccionando y agregando observaciones a cada grupo. Pensando en el investigador, él puede partir de un análisis por conglomerado desde un (1) conglomerado (la agrupación de toda las observaciones) hasta seis (6) conglomerados. Como se trató anteriormente, se debe buscar un equilibrio que permita disponer del principio de homogeneidad dentro del grupo y la heterogeneidad entre los grupos.

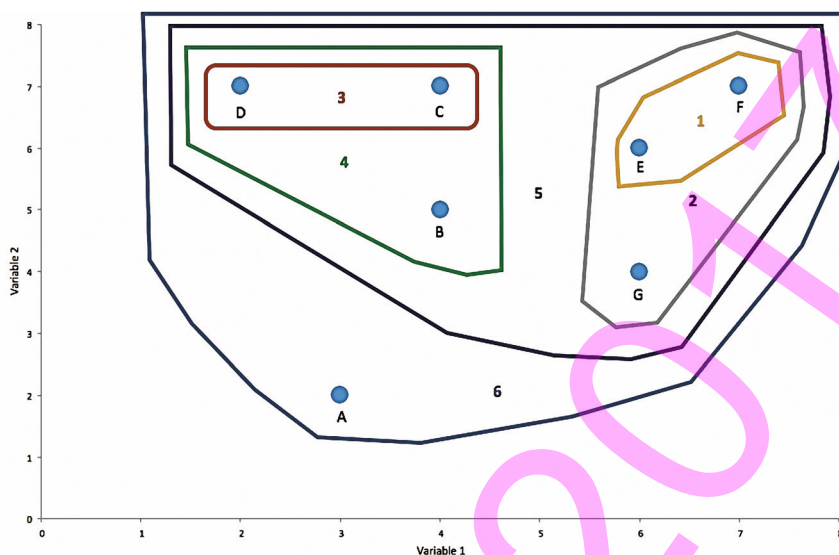


Figura 3 - 6 Representación gráfico de pertenencia y conformación de conglomerados

Nota: Tomado del libro *Multivariate Data Analysis* por Hair, J., Black, W., Babin, B., & Anderson, R. (2010), p 514

Una vez observado, el como utilizar una medida de proximidad y la forma como éstos se pueden agrupar; es decir, luego del paso 1 – la medida de similitud y el paso 2 – la formación de los conglomerados. Se debe llegar al paso 3 – cuantificar los conglomerados. En fin, determinar el número adecuado de grupos con que el investigador concluirá su análisis o el principio de estudios posteriores partiendo de estos conglomerados.

Para determinar el número de conglomerados, se debe concretar una estructura que permita definir la homogeneidad. Siguiendo con el ejemplo (método Aglomerativo jerárquico) para determinar la medida de similitud se toma la distancia media de todas las observaciones en el conglomerado (ver **Tabla 3 - 7**).

Ahora si, sin entender aún el método conglomerado utilizado, lo que se desea es comprender que en el paso 1, la solución inicial con 7 conglomerados (igual a las observaciones) la medida homogeneidad es 0, lo que significa que ninguna observación se encuentra emparejada con otra. Para la solución con 6 conglomerados, la medida resulta ser de 1,414 la que corresponde a la distancia entre las observaciones E y F; y así podemos observar en la tabla anterior que esta medida parte de 0 a 3,420.

Tabla 3 - 7 Medida de similitud dentro de los conglomerados

Paso	Pertenencia al conglomerado	Número de conglomerados	Medida de similitud
Solución inicial	A – B – C – D – E – F – G	7	0
1	A – B – C – D – EF – G	6	1,414
2	A – B – C – D – EFG	5	2,192
3	A – B – CD – EFG	4	2,144
4	A – BCD – EFG	3	2,234
5	A – BCDEFG	2	2,896
6	ABCDEFG	1	3,420

Ahora bien, ¿cómo debemos utilizar dichas medidas? Al observar el valor que toma la medida en cada paso debemos prestar especial atención a los cambios o aumentos “significativos”. Es decir, podemos analizarlos de la manera siguiente:

- **Del paso inicial al paso 1:** La medida aumenta de 0 a 1,414. Cambio significativo. Es decir, solución con 6 conglomerados.
- **Del paso 1 al paso 2:** La medida aumenta de 1,414 a 2,192. Cambio significativo al unir un conglomerado con tres miembros. Es decir, solución con 5 conglomerados.
- **Del paso 2 al paso 3:** La medida disminuye ligeramente de 2,192 a 2,141. Es decir, la solución con 4 conglomerados prácticamente tiene la misma homogeneidad que los conglomerados ya conformados en el paso anterior.
- **Del paso 3 al paso 4:** La medida aumenta ligeramente de 2,141 a 2,234. Lo que indica que la homogeneidad se mantiene prácticamente igual al integrar un tercer miembro en uno de los conglomerados. Es decir, solución con 3 conglomerados.
- **Del paso 4 al paso 5:** La medida aumenta sustancialmente de 2,234 a 2,896. Es decir, la solución con 2 conglomerados implica que al combinar los dos conglomerados de tres miembros se obtiene un grupo marcadamente menos homogéneo.
- **Del paso 5 al paso 6:** La medida pasa de 2,896 a 3,420. Lo que indica que cambia la homogeneidad al unir el último miembro. Es decir, solución con 1 conglomerado.

En el ejemplo, la medida aumenta cuando se unen dos pares de observaciones, como sucedió en el paso 1 (E-F) y disminuye el número de conglomerados en este caso a 6. Siguiendo con el criterio, la medida vuelve aumentar en el paso 2 y los conglomerados disminuyen a 5. Pero en los cambios del paso 3 y paso 4, la medida no cambia sustancialmente. Esto no quiere decir que se está formando conglomerados que mantienen la misma homogeneidad. Si analizamos que sucede en el paso 6, se determina que existe un aumento importante de la medida, y en contraste con la **Figura 3 - 6** (observación A) podemos determinar un grupo entropía (observación atípica).

En definitiva, el investigador en función de su estudio y de las medidas calculadas como lo muestra el ejemplo, es el encargado de seleccionar el número adecuado de conglomerados. Para este ejercicio, podría ser una buena solución la del paso 4, en donde se obtienen dos conglomerados del mismo tamaño (3 observaciones cada uno) y un tercero aislado. En total 3 grupos, como se aprecia en la **Figura 3 - 7**.

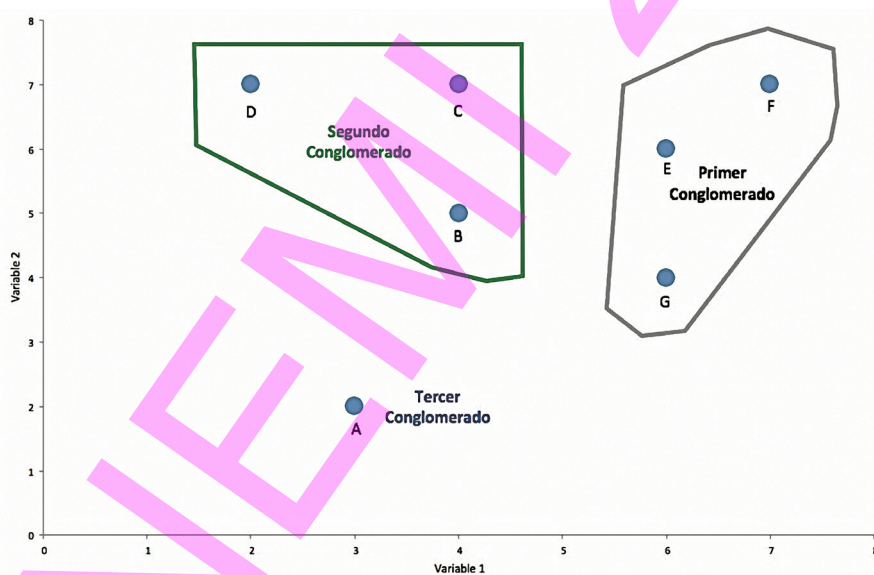


Figura 3 - 7 Solución final de tres conglomerados

3.5. Software para análisis de conglomerados

El desarrollo computacional en la actualidad es cada vez más determinante en el uso de los algoritmos y diferentes técnicas. En el análisis por conglomerado, cada vez es más utilizado mediante el apoyo de diferentes aplicaciones informáticas en donde disponen de está técnicas. A continuación se presenta

en la **Tabla 3 - 8** los sistemas informáticos (software) más utilizados para el análisis de conglomerados o análisis clúster como se le denomina en algunos de estos software.

Tabla 3 – 8 Software que disponen técnicas para Análisis por conglomerado (Análisis Cluster)

Software	Tipo	Enlaces	Otros enlaces
R	Código abierto	https://www.r-project.org/	http://www.statmethods.net/advstats/cluster.html
APACHE SPARK	Código abierto	https://cran.r-project.org/web/packages/cluster/cluster.pdf	
WEKA	Código abierto	http://spark.apache.org/	https://spark.apache.org/docs/2.1.0/mllib-clustering.html
KNIME	Código abierto	http://www.cs.waikato.ac.nz/ml/index.html	https://weka.wikispaces.com/Using+cluster+algorithms
SPSS	Comercial (licenciado)	http://www.knime.org/	https://tech.knime.org/document-clustering-example
ORACLE	Comercial (licenciado)	https://www.ibm.com/support/knowledgecenter/en/SSLVMB_21.0.0/com.ibm.spss.statistics.help/alg_cluster.htm	http://docs.rapidminer.com/studio/operators/modeling/segmentation/agglomerative_clustering.html
Rapidminer	Comercial (licenciado) Versión educativo	https://www.ibm.com/analytics/us/en/technology/spss/	https://www.ibm.com/support/knowledgecenter/en/SSLVMB_22.0.0/com.ibm.spss.statistics.help/spss/base/cluster_choosing.htm

3.5.1. Ejemplos de Análisis de Conglomerados

3.5.1.1. WEKA

En este apartado se presenta un ejemplo con el fin de conocer algunos aspectos:

1) El uso del software WEKA para realizar un análisis por conglomerado o clúster, 2) Como funciona el análisis de conglomerados, y 3) La interpretación de los resultados. En primer lugar se debe descargar el aplicativo del sitio web <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>, para lo cual se visualizará una pantalla como la mostrada en la **Figura 3 – 8**.

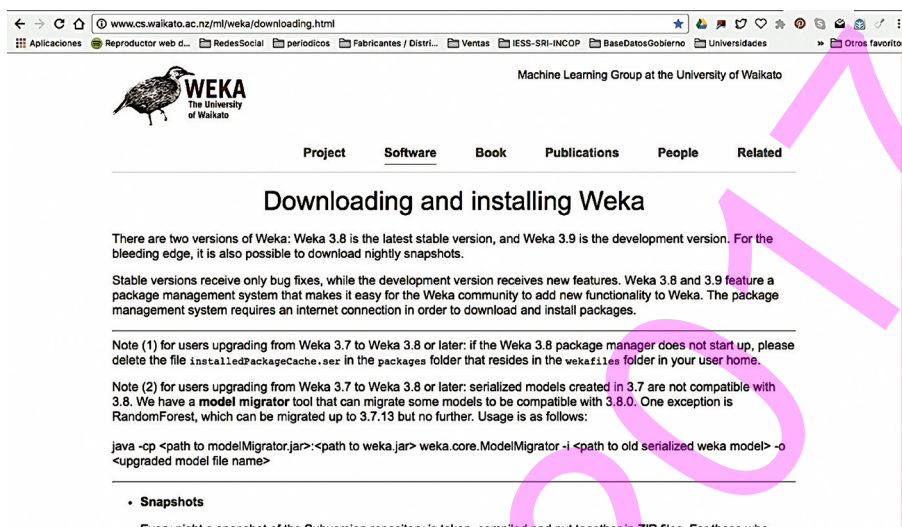


Figura 3 - 8 Sitio para descargar WEKA

De acuerdo al sistema operativo de su equipo, deberán seleccionar de la mejor manera el aplicativo a descargar en su ordenador. Una vez que haya sido instalado el aplicativo en su equipo, podrá ejecutarlo.

Al iniciar la aplicación se encontrará con una pantalla de apariencia similar a la mostrada en la **Figura 3 – 9**, la cual dependerá del sistema operativo y la versión descargada.

En las siguientes direcciones electrónicas propias del sitio podrán consultar cada una de las funciones de WEKA:

- <http://www.cs.waikato.ac.nz/ml/weka/documentation.html>
- <http://www.cs.waikato.ac.nz/ml/weka/book.html>

Con el fin de poder describir el análisis de conglomerados mediante el software WEKA, consideramos como un buen ejemplo el uso de un archivo que reposa en los ejemplos del mismo software. Este archivo se llama “iris.arff”, y se encuentra en la carpeta data.

Como podrán observar la extensión del archivo que maneja el software es “.arff”. En la **Figura 3 - 10** podrá observar como acceder al archivo mencionado, solo realizando un clic en el botón “Explorer”.

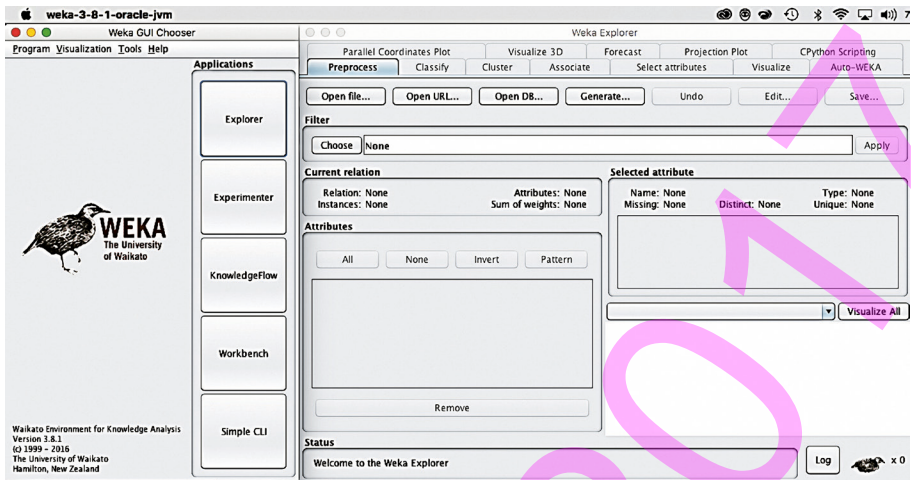


Figura 3 - 9 Pantalla de inicio WEKA

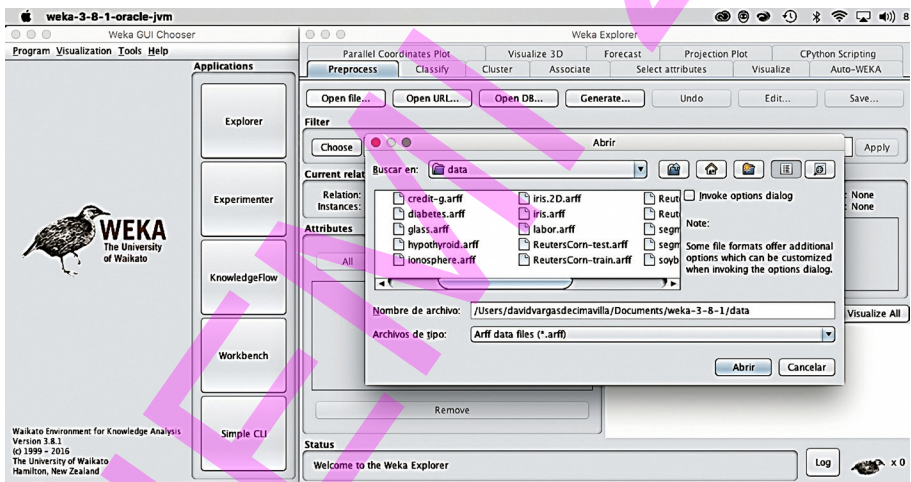


Figura 3 - 10 Acceso al archivo de datos

El archivo "iris.arff", es un conjunto de datos multivariados relacionado a las tres especies de la Flor Iris, distribuidas en 50 muestras por cada una de ellas distribuidas en Iris-setosa, Iris-virginica y Iris-versicolor (total 150 instancias o individuos). El archivo muestra originalmente 5 atributos los cuales son los siguientes (ver **Figura 3 – 11**):

- Sepallength (largo del sépalo).
- Sepalwidth (ancho del sépalo).
- Petallength (largo del pétalo).

- Petalwidth (ancho del pétalo).
- Class (clase de la especie de la Flor Iris).

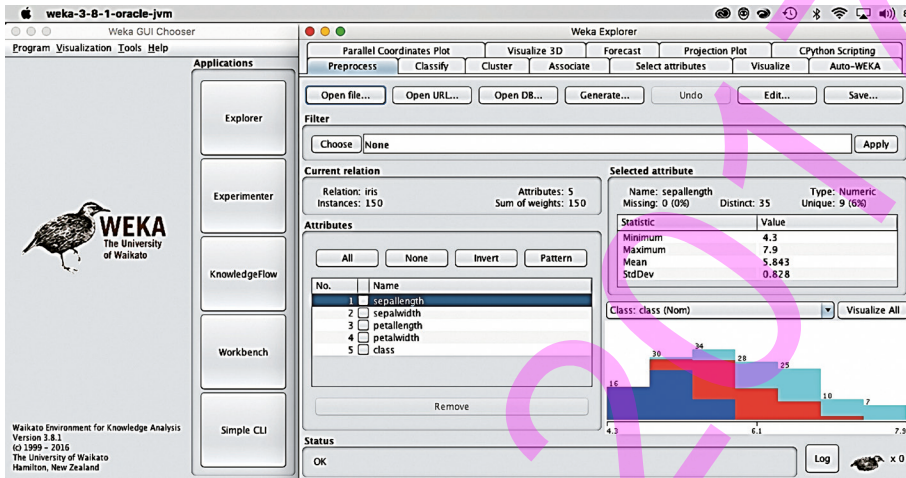


Figura 3 - 11 Atributos que muestra el archivo "iris.arff"

Haciendo uso en menú "Preprocess" (menú superior) y en la sección "Atributtes" se removerá el atributo "Class" (número 5), el mismo que deberá quedar de la siguiente manera mostrada en la **Figura 3 – 12**.

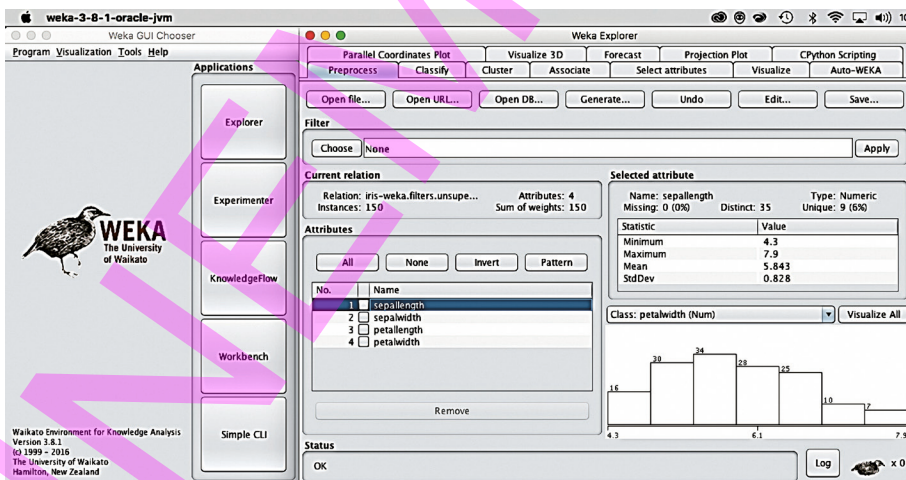


Figura 3 - 12 Eliminación del atributo "Class"

Una vez eliminado o removido el atributo "Class", el siguiente paso es poder ejecutar técnicas o modelos para el análisis de conglomerados, esto se obtiene presionando un clic en el menú principales la opción "Cluster", como se puede observar en la **Figura 3 – 13**.

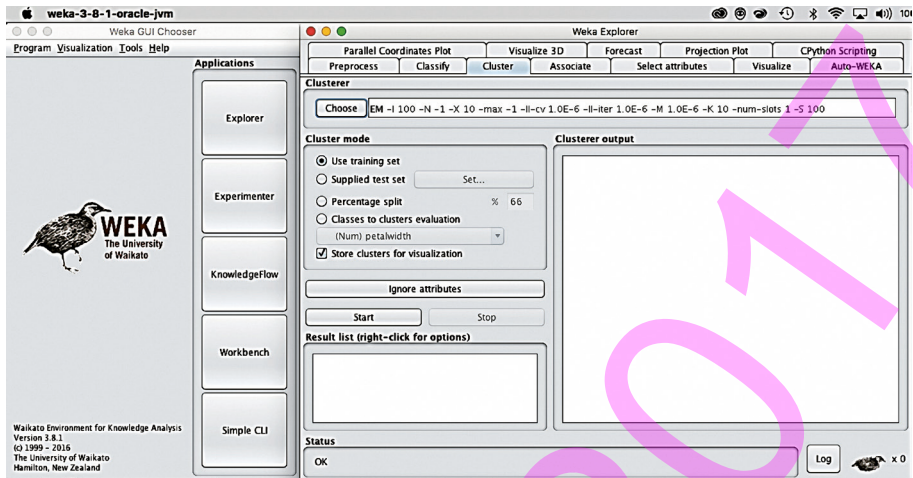


Figura 3 - 13 Opción “Cluster” del Menú Principal

En el botón “Choose”, se desplegarán una serie de técnicas de análisis de conglomerados que el software dispone. Para efectos de este ejercicio se utilizará el algoritmo SimpleKMeans. En el siguiente figura se puede observar lo indicado (Figura 3 – 14).

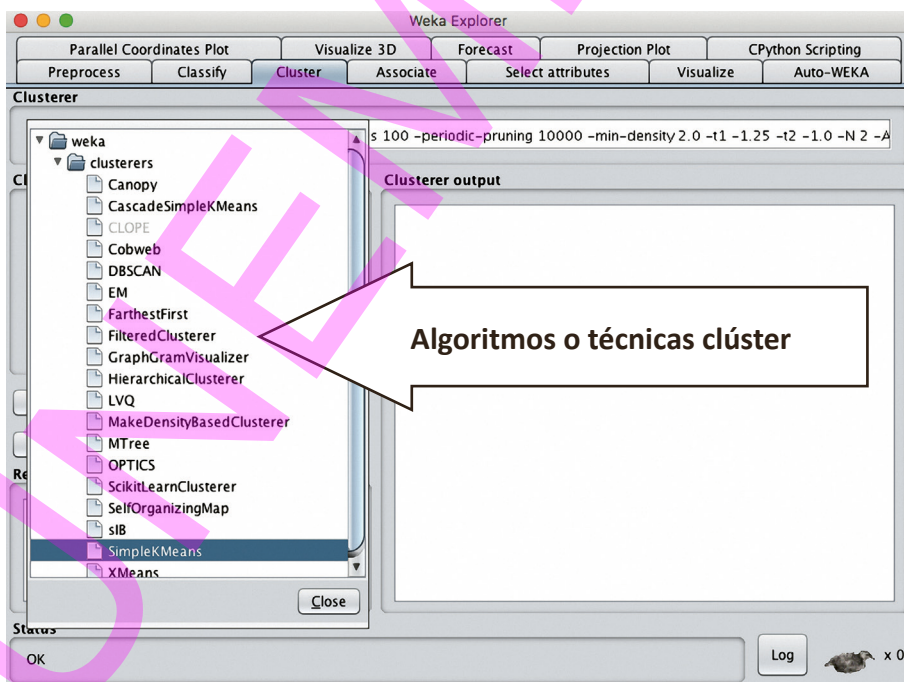


Figura 3 - 14 Opciones de análisis de conglomerados que dispone WEKA

Al seleccionar el algoritmo, que en el estudio se trata de la aplicación de análisis de conglomerado no jerarquizado, por el método de K-means. En WEKA seleccionamos el modelo con un número a priori de 3 grupos o conglomerados de la siguiente manera (**Figura 3 – 15**).

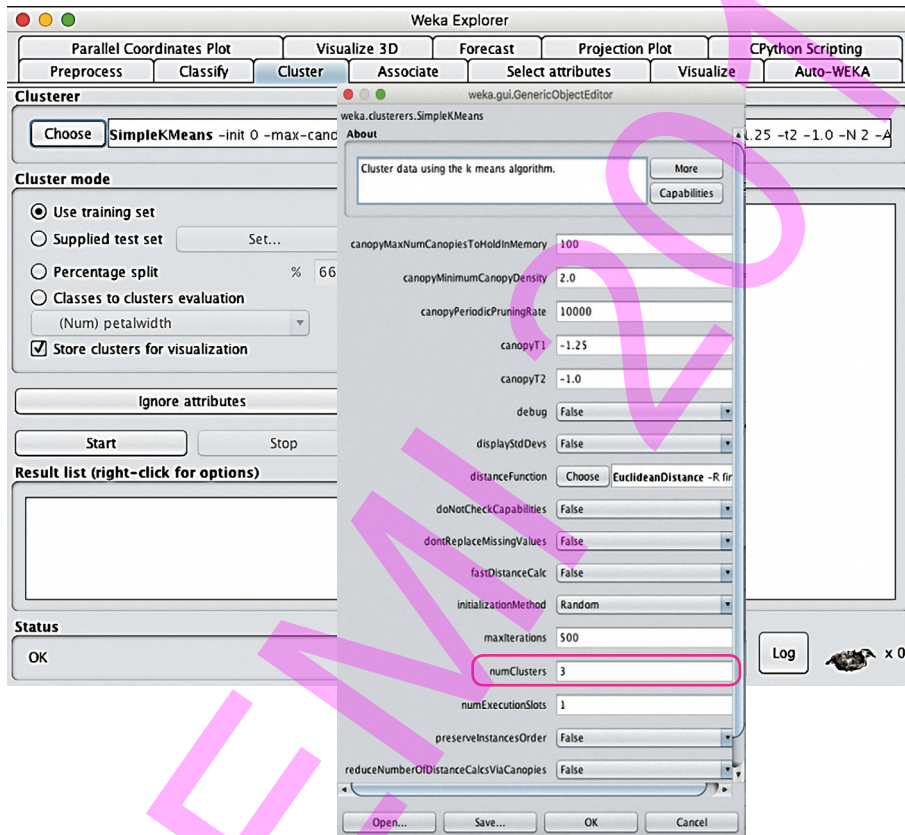


Figura 3 - 15 Selección del número de grupos o conglomerados

Para ejecutar el modelo en WEKA, se presiona en el botón “Start” y obtiene el siguiente resultado tal como se observa en la **Figura 3 – 16**.

En la **Figura 3 - 16**, se pueden observar cuatro grupos encerrados en recuadros de colores. El primero (recuadro de color rojo) muestra la información del algoritmo que se ejecutó con los debidos parámetros (k = 3 como ejemplo), el segundo (recuadro de color azul) muestra principalmente las iteraciones (6) y la suma cuadrática de error sin agrupar, como tercero (recuadro de color verde) muestra información de las centroides de los 3 conglomerados, y cuarto (recuadro de color púrpura) muestra el porcentaje y cantidad de individuos pertenecientes a cada uno de los 3 conglomerados o clúster.

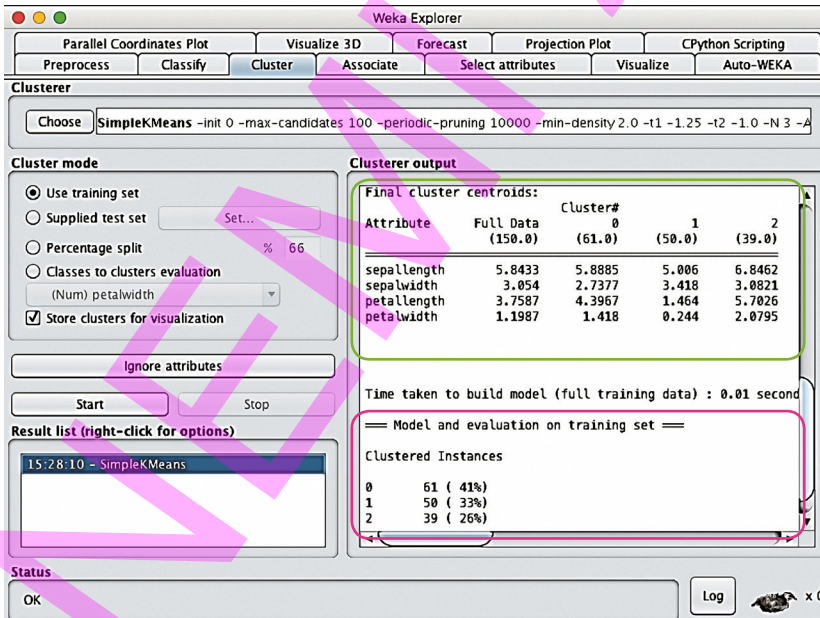
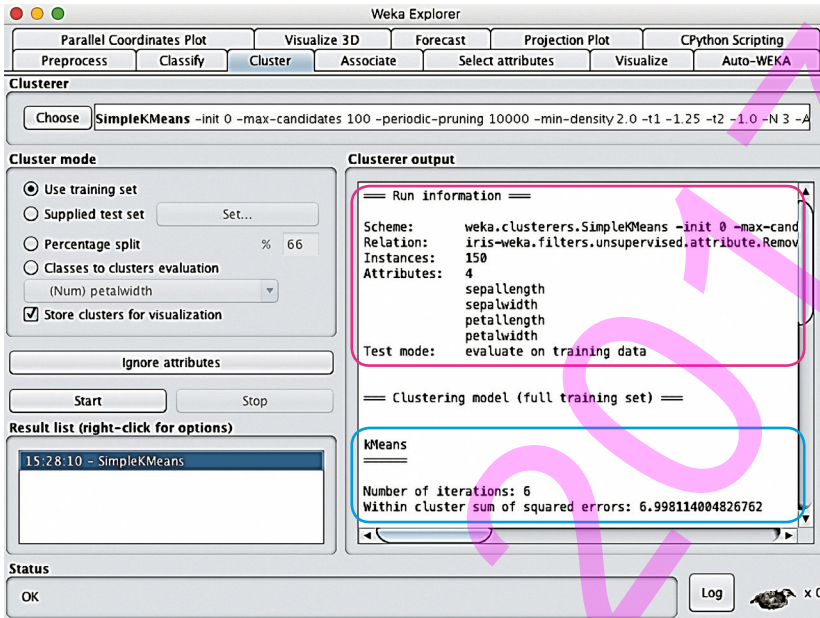


Figura 3 - 16 Salidas del análisis de conglomerados WEKA

Si se ejecuta clic derecho sobre el nombre del algoritmo empleado, esto es debajo del botón "Star". Se puede disponer de acceso a diferentes funciones como el visualizar los conglomerados o clúster determinados por el algoritmo en WEKA. La siguientes **Figuras 3 - 17** y **3 -18** muestran lo indicado.

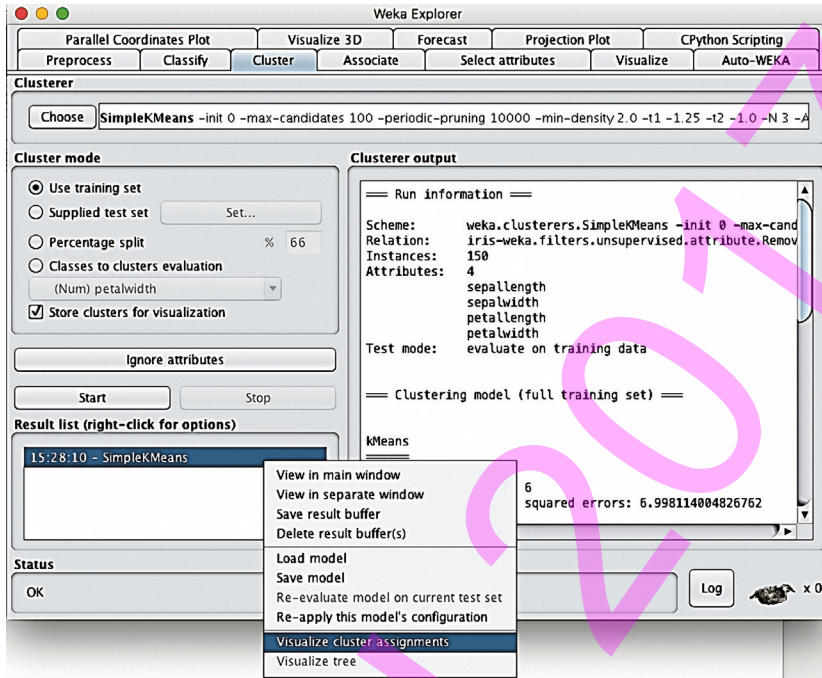


Figura 3 - 17 Opciones de visualización de los conglomerados

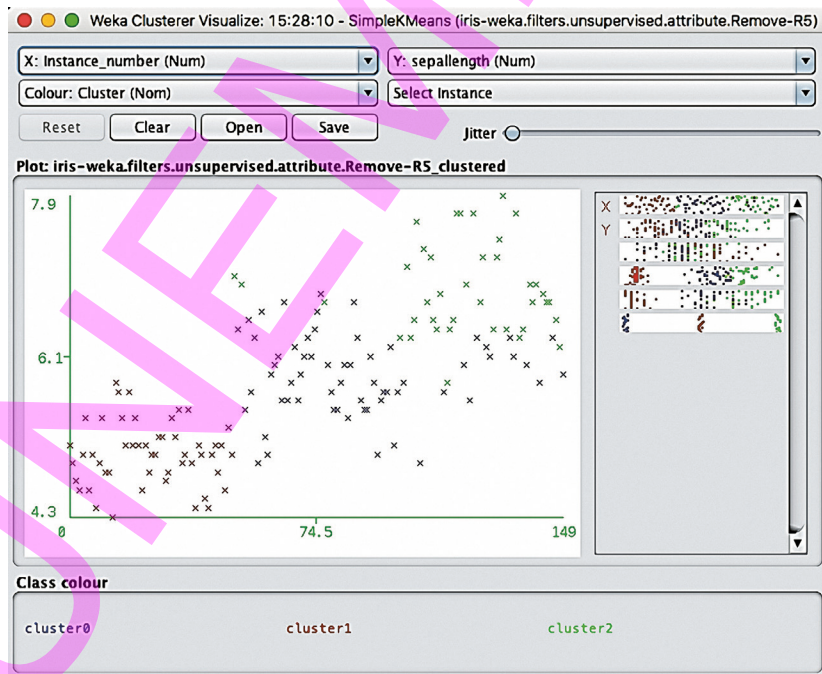


Figura 3 - 18 Visualización opción "cluster assignments"

En la **Figura 3 - 19** podemos observar como WEKA mediante el algoritmo agrupa en tres conglomerados diferenciándolos con diferentes colores. El programa permite visualizar de diferentes maneras las relaciones en los distintos atributos que la data disponga. También es posible disponer de una “matriz” de gráficas, para lo cual deberá seleccionar en el menú principal de WEKA la opción “Visualize”.

Regresando a la solución y los resultados de la ejecución del algoritmo, se obtiene el agrupamiento en tres conglomerados o clúster mostrados en la **Tabla 3 – 9**. Es decir, de los 150 muestras se obtiene que 61 se encuentran agrupadas en el clúster 0 (41%), 50 en el clúster 1 (33%) y finalmente 39 en el clúster 2 (26%).

Tabla 3 – 9 Modelo de Conglomerados obtenidos con WEKA

Conglomerado	Instancias	Porcentaje
0	61	41 %
1	50	33 %
2	39	26 %

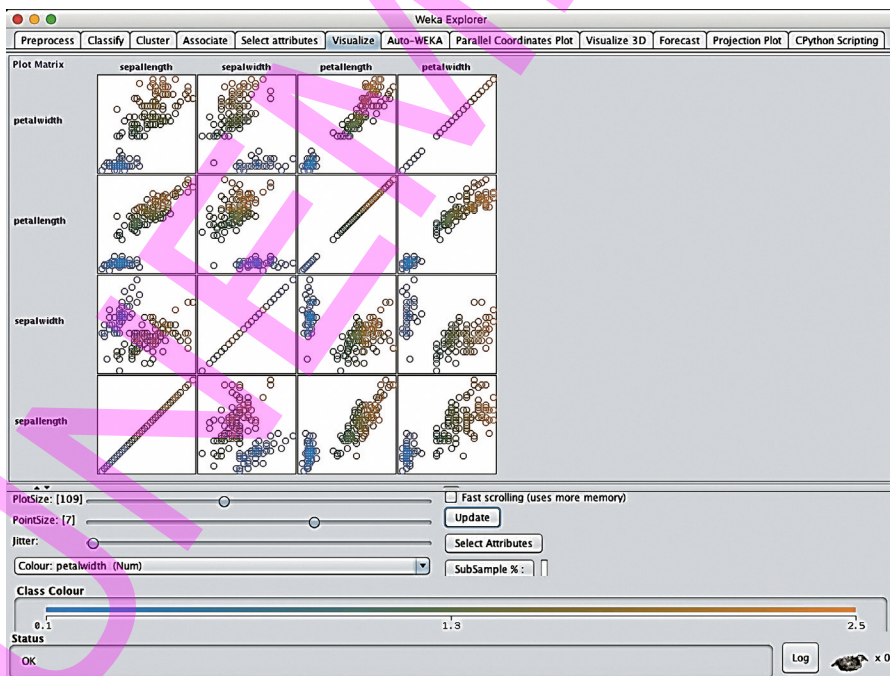


Figura 3 – 19 Visualización de los conglomerado

Para finalizar este ejemplo, se puede concluir que el uso de WEKA como un software para realizar análisis de conglomerados es muy sencillo y rápido de aplicar. WEKA ofrece diversos algoritmos, los cuales el investigador podrá comprender con la práctica debida.

Es importante resaltar que la herramienta es de libre acceso y comparada con otras plataformas pueden ser mejor utilizar este software.

3.5.1.2. SPSS

En este segundo apartado se presenta un ejemplo con el fin de conocer el uso del software SPSS para realizar un análisis de conglomerado o clúster. El SPSS es un software licenciado (debe pagarse) que contiene herramientas para realizar análisis estadístico de toda índole.

El programa SPSS dispone de tres tipos de análisis de conglomerados: el análisis de conglomerados jerárquico, bietápico y de K medias (Vilà-baños et al., 2014). A continuación se ilustra cómo funciona el análisis de conglomerados en el SPSS, y la interpretación de los resultados a partir de un ejemplo “divorcios” tomado del Instituto Nacional de Estadística y Censos (INEC) - Secretaría Nacional de Planificación y Desarrollo (SENPLADES) de Ecuador. Aplicaremos el método clustes k-medias ya que es el más idóneo para determinar el contenido óptimo de conglomerados existentes en nuestro caso práctico.

El archivo “divorcios”, es una base de datos relacionada con los divorcios ocurridos en Ecuador en el año 2012 e inscritos a nivel nacional. En ese año se registraron 20.299 divorcios, presentándose información clasificada de acuerdo con el mes de ocurrencia, regiones y provincias de residencia habitual de los divorciados, lugar geográfico del divorcio, tiempo de duración del matrimonio, los datos de los divorciados: edad, número de hijos a cargo de los divorciados, reconocimientos, adopciones y nivel de instrucción.

Al ejecutar el Software SPSS, se muestra una ventana por defecto (**Figura 3 – 20**), en la cual se permite al usuario escoger el archivo, el cual contiene la información con la que se va a trabajar.

En la opción “Nuevo conjunto de datos” exploramos el buscador para seleccionar el archivo de datos. En nuestro ejemplo usaremos el archivo en formato Excel denominado “divorcios.xlsx”.

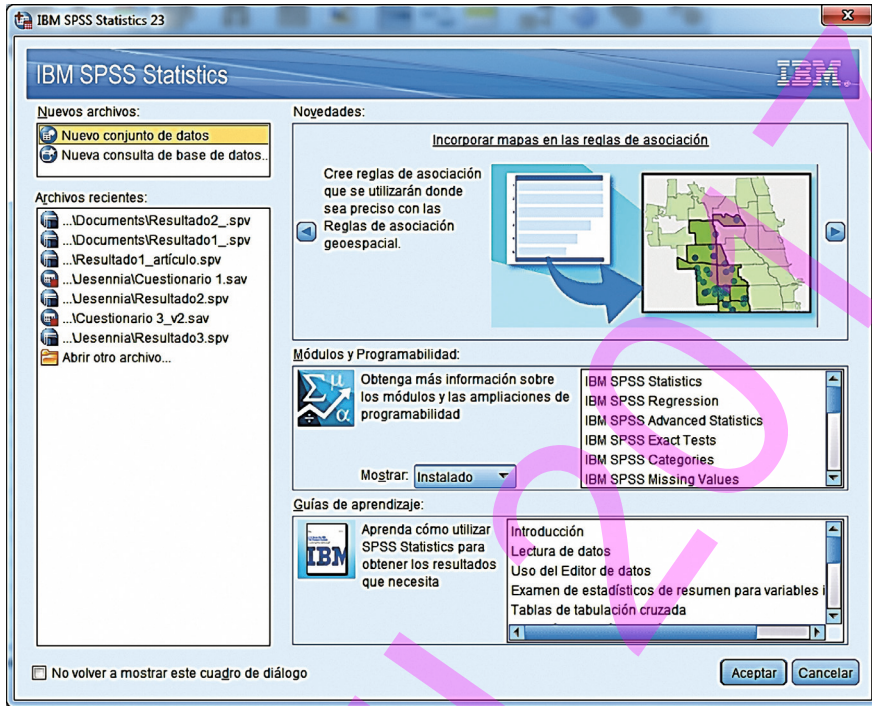


Figura 3 – 20 Pantalla de inicio por defecto del SPSS

Una vez que se ha seleccionado el archivo de datos con el cual se va a trabajar (divorcios.xlsx) se presenta la pantalla mostrada en la **Figura 3 – 21** en donde debemos indicar “Leer el nombre de las variables de la primera fila de datos”. Posteriormente, al “Aceptar” visualizaremos el Visor de Resultados y el Editor de Datos en background.

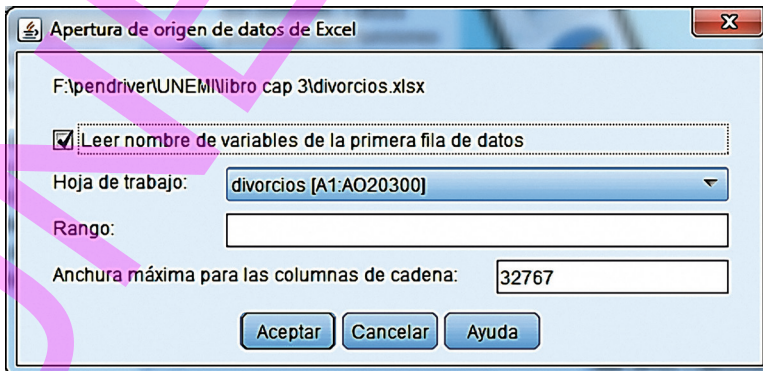


Figura 3 – 21 Cuadro de diálogo apertura de origen de datos de Excel

	of_insc	prov_insc	cant_insc	par_insc	anio_insc	mes_insc	acta_insc	anio_dv	mes_dv	anio_mat	mes_mat	mcap_bie	dur_mat	ran_durm	edad_hom1	ran_edadhom1	edad_hom2	ran_edadhom2	hijos_h
1	1	8	801	80150	2012	11	274	1980	6	1979	10	2	0	0	55	10	87	13	
2	37	9	901	90150	2012	10	2723	1985	9	1985	8	9	0	0	47	8	74	13	
3	5	9	901	90150	2012	8	1	1990	9	1989	10	9	0	0	45	8	66	12	
4	37	9	901	90150	2012	10	2849	1997	12	1997	12	2	0	0	36	6	50	9	
5	37	9	901	90150	2012	10	2907	2000	7	2000	4	2	0	0	36	6	48	8	
6	1	8	801	80150	2012	2	30	2002	3	2002	3	9	0	0	47	8	99	14	
7	1	7	702	70250	2012	3	17	2005	8	2005	8	9	0	0	31	5	37	6	
8	1	18	1801	180150	2012	9	351	2006	1	2006	1	2	0	0	30	5	36	6	
9	37	9	901	90150	2012	10	2703	2007	2	2007	2	2	0	0	40	7	45	8	
10	25	9	901	90150	2012	11	280	2007	11	2007	1	2	0	0	29	4	34	5	
11	1	1	101	10150	2012	6	448	2008	10	2007	12	2	0	0	18	2	21	3	
12	37	9	901	90150	2012	8	2234	2008	5	2007	11	9	0	0	34	5	38	6	
13	37	9	901	90150	2012	10	2691	2008	11	2008	2	2	0	0	31	5	34	5	
14	37	9	901	90150	2012	11	3107	2008	8	2008	8	2	0	0	28	4	32	5	
15	1	16	1601	160150	2012	2	24	2009	10	2008	10	9	0	0	40	7	99	14	
16	1	23	2301	230150	2012	3	97	2009	6	2008	6	9	0	0	21	3	99	14	
17	37	9	901	90150	2012	4	977	2009	8	2008	12	9	0	0	42	7	44	7	
18	1	17	1701	170150	2012	4	1336	2009	7	2008	6	9	0	0	32	5	99	14	
19	1	3	301	30150	2012	8	94	2009	1	2008	8	2	0	0	27	4	30	5	
20	1	12	1205	120550	2012	1	11	2010	10	2010	3	9	0	0	44	7	99	14	
21	3	9	901	90150	2012	3	11	2010	2	2009	5	9	0	0	52	9	99	14	
22	1	17	1701	170150	2012	3	958	2010	5	2009	4	9	0	0	27	4	98	14	
23	1	17	1701	170150	2012	5	1607	2010	7	2010	1	9	0	0	30	5	99	14	

Figura 3 – 22 Visualización de la matriz de datos

Para ejecutar el análisis de conglomerados (Análisis Clúster) con el Software SPSS, en la matriz de datos se seleccionan los menús: “Analizar” – “Clasificar” – “Cluster de k-medias”, como se muestra en la Figura 3 – 23.

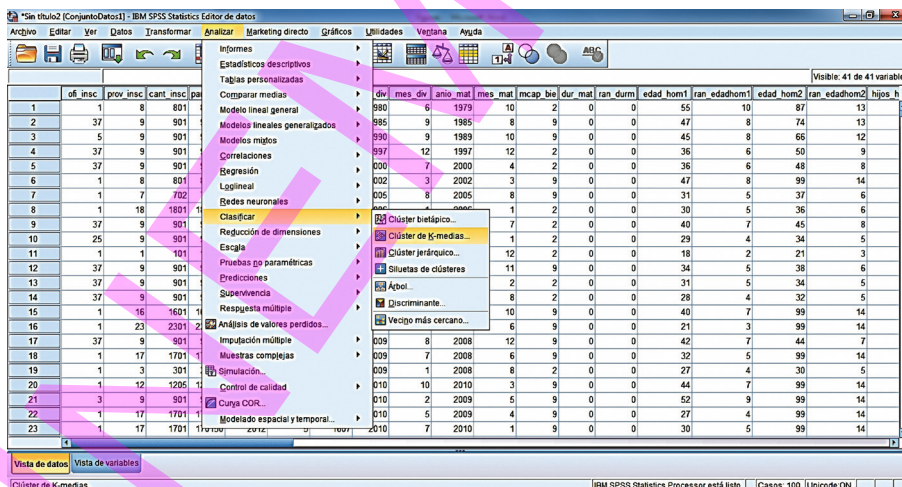


Figura 3 – 23 Cuadro de diálogo análisis de conglomerados en SPSS

Una vez que se ha seleccionado la opción de “Cluster de k-medias”, introduciremos las variables de las que se desean formar los conglomerados (clusters), así como también el número conglomerados (clusters). En nuestro caso introduciremos la variable “edad_hom2”, “hijos_hom”, “niv_insth” y “p_

etnica_hom”, y definiremos como número de conglomerados: 3 (ver **Figura 3 – 24**). Luego “Aceptar”.

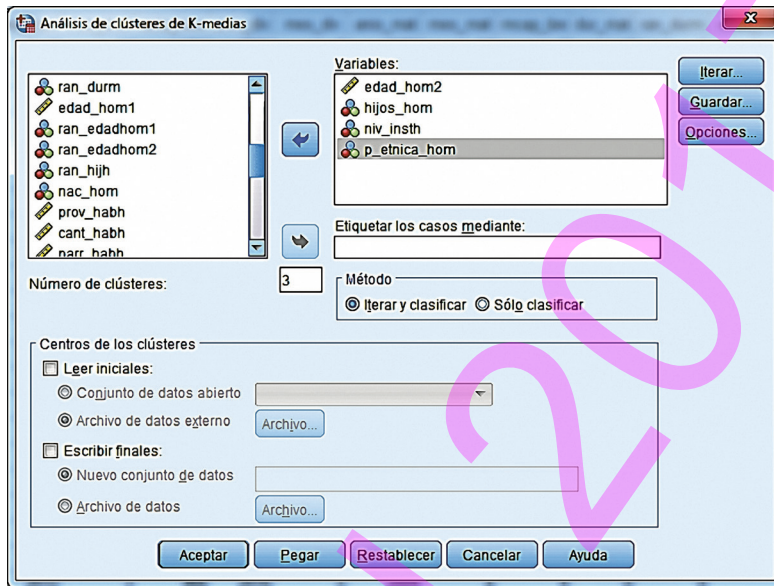


Figura 3 – 24 Ventana del análisis SPSS

Una vez ejecutada la acción, el software SPSS comienza la ejecución del proceso, y arroja una serie de salidas mostrando en la parte superior datos informativos sobre el proceso que se ejecuta, tal como localización de la base de datos, el tipo de procesos, sus parámetros, así como los resultados tabulados que se detallan en las figuras mostradas a continuación.

Primero se presentan los centros de clústeres iniciales, es decir, como inicialmente se crean los grupos que tiene posibles características comunes (**Figura 3 – 25**).

Centros de clústeres iniciales

	Clúster		
	1	2	3
edad_hom2	101	59	18
hijos_hom	0	0	0
niv_insth	9	9	2
p_etnica_hom	6	9	1

Figura 3 – 25 Centros iniciales de los conglomerados

A continuación, se muestra el número de iteraciones que el software SPSS ha realizado hasta lograr la convergencia. En este caso realizó 6 iteraciones, hasta encontrar la distancia mínima entre los centros de conglomerados iniciales que fue de 42,107.

Seguidamente, se muestra una tabla con los centros de conglomerados finales de la partición de los datos (**Figura 3 – 27**), agrupados en 3 conglomerados y las cuatro variables seleccionadas “edad_hom2”, “hijos_hom”, “niv_insth” y “p_etnica_hom”, las cuales representan : “Edad del divorciado a la fecha de inscripción”, “Número de hijos del divorciado”, “Nivel de instrucción del divorciado” y “Auto identificación étnica del divorciado”, respectivamente.

Historial de iteraciones ^a

Iteración	Cambiar en centros de clústeres		
	1	2	3
1	6,739	10,882	14,992
2	,476	1,144	1,094
3	,000	,689	,496
4	,005	,726	,499
5	,131	,086	,000
6	,000	,000	,000

a. Convergencia conseguida debido a que no hay ningún cambio en los centros de clústeres o un cambio pequeño. El cambio de la coordenada máxima absoluta para cualquier centro es ,000. La iteración actual es 6. La distancia mínimo entre los centros iniciales es 42,107.

Figura 3 – 26 Historial de iteraciones

Centros de clústeres finales

	Clúster		
	1	2	3
edad_hom2	98	52	33
hijos_hom	0	0	0
niv_insth	3	3	4
p_etnica_hom	9	8	8

Figura 3 – 27 Centro de conglomerados finales

El último paso del análisis es la tabla resumen de la clasificación de los casos, número de casos (individuos) de cada conglomerado (**Figura 3 – 28**). En nuestro caso, vemos que se ha clasificado correctamente el 100%. Es decir, del total de divorcios en el año 2012 (20.299), obtiene que 4.470 se encuentran agrupadas en el clúster 1 (22%), 6.643 en el clúster 2 (33%) y finalmente 9.186 en el clúster 3 (45%).

Número de casos en cada clúster

Clúster	1	4470,000
	2	6643,000
	3	9186,000
Válidos		20299,000
Perdidos		,000

Figura 3 – 28 Número de casos en cada conglomerado

Como conclusión se puede determinar que el uso de SPSS como un software para realizar análisis por conglomerados es muy sencillo de utilizar y el tiempo de ejecución es muy rápido.

3.5. Resumen del capítulo

En muchas áreas del conocimiento, las investigaciones realizadas necesitan identificar cuáles son las características que diferencian unos grupos de sujetos u objetos respecto de otros, para así poder realizar proyecciones (Torrado-Fonseca & Berlanga-Silvente, 2013). Básicamente hay dos técnicas que permiten clasificar sujetos y objetos: el análisis discriminante y el análisis de conglomerados, las cuales se diferencian entre sí por el momento del establecimiento del grupo. En el caso del análisis de conglomerados el momento es posteriori, ya que los grupos se determinan y configuran después de haber sido estudiadas y analizadas las agrupaciones.

El análisis de conglomerados puede clasificarse, en cuanto al procedimiento, en métodos jerárquicos y métodos no jerárquicos. El procedimiento jerárquico es más adecuado para muestras pequeñas en las que los grupos se configuran por agrupaciones sucesivas de individuo a individuo, o de individuo a grupo configurando una estructura arborescente con niveles, que desemboca en una jerarquización de conglomerados. El procedimiento no jerárquico parte, de entrada, de la determinación de un número de grupos y asigna los casos a grupos diferenciados sin que unos dependan de otros (Vilà-bañes et al., 2014). El funcionamiento del análisis de conglomerados, depende de un proceso que

primero necesita identificar las observaciones más parecidas en grupos. Este proceso responde a: (1) Medir la similitud de las observaciones; (2) Formar los conglomerados; y (3) Cuantificar los conglomerados. Medir la similitud, requiere de un método que mida simultáneamente las observaciones en todas las variables, pudiendo utilizar técnicas como correlación entre objetos, medida de asociación o alguna medida de proximidad utilizada. La formación de los conglomerados, consiste en agrupar aquellas observaciones que son más similares dentro de un conglomerado, este proceso determinará la pertenencia al grupo de cada observación. La determinación del número de conglomerados o grupos, es una tarea que el investigador debe enfrentar, ya que debe considerar que si el número de grupos es menor necesariamente la homogeneidad dentro del grupo disminuye.

Existen diferentes aplicaciones informáticas en donde disponen de esta técnica. Siendo los software más utilizados para el análisis de conglomerados o análisis clúster los siguientes : R (código abierto), APACHE (código abierto), SPARK (código abierto), WEKA (código abierto), KNIME (código abierto), SPSS (comercial licenciado), ORACLE (comercial licenciado), Rapidminer (comercial licenciado)

3.6. Referencias

- Alonso, A. (2008). Técnicas de análisis multivariante de datos. (I. Capella & M. Diego, Eds.) (Primera). Madrid-España: PEARSON PRENTICE HALL.
- Arteaga, R., & Duarte, A. (2010). Motivational factors that influence the acceptance of Moodle using TAM. *Computers in Human Behavior*, 26(6), 1632–1640. <http://doi.org/10.1016/j.chb.2010.06.011>
- Cuadras, C. M. (2014). Nuevos Métodos de Análisis Multivariante. (CMC Editions, Ed.) CMC Editions (Primera). Barcelona - España: CMC Editions. <http://doi.org/10.1017/CBO9781107415324.004>
- Esquivel Guerrero, J., Contreras Cortes, F., Molina González, F., & Capel Martínez, J. (1991). Una aplicación de la Teoría de Información al análisis de datos definidos mediante variables cualitativas multi-estado: medidas de similitud y análisis cluster. *Complutum*, (1), 53–64. Retrieved from <http://www.bcin.ca/Interface/openbcin.cgi?submit=submit&Chinkey=161889>
- Findik, D., & Kunçay, C. O. Ş. (2009). A model for instructors adoption of learning management systems : empirical validation in higher education context. Middle East Technical University.
- Gómez Suárez, M. (2002). EL ANALISIS CLUSTER EN INVESTIGACIÓN DE MARKETING: METODOLOGIA Y CRITIC. *ESIC Market*, (113), 9–34.
- Hair, J., Black, W., Babin, B., & Anderson, R. (2010). *Multivariate Data Analysis*. (Pearson, Ed.) (Seventh).
- Lay, J.-G., Chen, Y.-W., & Chi, Y.-L. (2013). GIS Adoption Among Senior High

- School Geography Teachers in Taiwan. *Journal of Geography*, 112(3), 120–130. <http://doi.org/10.1080/00221341.2012.682226>
- Leyva, O., & Olague, J. (2014). Métodos y Técnicas Cualitativas y Cuantitativas Aplicables a la Investigación en Ciencias Sociales. In TIRANT HUMANIDADES MÉXICO (Ed.), *Métodos y Técnicas Cualitativas y Cuantitativas Aplicables a la Investigación en Ciencias Sociales* (pp. 480–497). Monterrey- México: TIRANT HUMANIDADES MÉXICO.
 - Lin, W.-S. (2012). Perceived fit and satisfaction on web learning performance: IS continuance intention and task-technology fit perspectives. *International Journal of Human-Computer Studies*, 70(7), 498–507. <http://doi.org/10.1016/j.ijhcs.2012.01.006>
 - Mbarek, R., & Zaddem, F. (2013). The examination of factors affecting e-learning effectiveness. *International Journal of Innovation and Applied Studies*, 2(4), 423–435. <http://doi.org/2028-9324>
 - Pérez Hernández, C. C., Lara Gómez, G., & Gómez Hernández, D. (2017). Evolución de la capacidad tecnológica en México. Aplicación del análisis estadístico multivariante de cluster. *Contaduría Y Administración*, 62(2), 505–527. <http://doi.org/10.1016/j.cya.2017.01.002>
 - Sanjuán Suárez, P., Pérez García, A. M., & Bermúdez Moreno, J. (2000). Escala de autoeficacia general: Datos psicométricos de la adaptación para población española. *Psicothema*, 12(SUPPL. 2), 509–513. <http://doi.org/ISSN 0214-9915>
 - Sharma, N., Bajpai, A., & Litoriya, R. (2012). Comparison the various clustering algorithms of weka tools. *International Journal of Emerging Technology and Advanced Engineering*, 2(5), 73–80.
 - Tarhini, A., Hone, K., & Liu, X. (2013). Factors Affecting Students' Acceptance of e-Learning Environments in Developing Countries : A Structural Equation Modeling Approach. *International Journal of Information and Education Technology*, 3(1), 54–59. <http://doi.org/10.7763/IJiet.2013.V3.233>
 - Torrado-Fonseca, M., & Berlanga-Silvente, V. (2013). Análisis Discriminante mediante SPSS. *REIRE Revista d'Innovació I Recerca En Educació*, 6(2), 150–166. <http://doi.org/10.1344/reire2013.6.26210> //
 - Vilà-baños, R., Rubio-Hurtado, M.-J., Berlanga-Silvente, V., & Torrado-Fonseca, M. (2014). Cómo aplicar un cluster jerárquico en SPSS. *REIRE Revista d'Innovació I Recerca En Educació*, 7(1), 113–127.
 - Weng, C., & Tsai, C. (2015). Social support as a neglected e-learning motivator affecting trainee ' s decisions of continuous intentions of usage. *Australasian Journal of Educational Technology*, 31(2), 177–192.

"La Estadística es uno de los escalones a subir para alcanzar el conocimiento"

Richard Ramirez-Anormaliza (2016)

Doctor Cum Laude

Capítulo 4: Modelo de ecuaciones estructurales

Richard Ramirez-Anormaliza

Freddy Bravo-Duarte

Ruth Farias-Lema

Rubi Pena-Holguin

Objetivos de aprendizaje

Para este capítulo se plantea dar a conocer en profundidad los modelos de ecuaciones estructurales (SEM) como técnica multivariante para investigar en varios campos. De manera específica los objetivos planteados son los siguientes:

- Conocer las características de los modelos de ecuaciones estructurales su terminología y notación básica.
- Distinguir los enfoques de ecuaciones estructurales y los criterios para su selección a la hora de realizar una investigación.
- Conocer los diferentes programas informáticos para tratamiento de ecuaciones estructurales.
- Aplicar el procesamiento de datos por medio de modelos de ecuaciones estructurales con enfoque en mínimos cuadrados parciales.
- Conocer el análisis multigrupo mediante ecuaciones estructurales con enfoque en mínimos cuadrados parciales.
- Aplicar el procesamiento de datos por medio de modelos de ecuaciones estructurales con enfoque en covarianzas.

4.1. Introducción

Los modelos de ecuaciones estructurales (structural equations modelling, SEM) tienen como objetivo evaluar si un modelo que formula una determinada analogía entre las variables observables concuerda con los datos empíricos. La evaluación de un SEM requiere de una primera etapa para especificar el modelo, seguido de las etapas de identificación y de estimación. En la etapa de especificación se delimitan las relaciones a evaluar entre variables. Se puede dar el caso de que estas relaciones sean más complicadas que las precisadas en los modelos lineales, en los que sólo existe una variable dependiente y múltiples variables independientes. En SEM podemos tener tanto variables observables como variables latentes (Sallan, Fernandez, Simo, Lordan, & Gonzalez-Prieto, 2012).

La historia de las ecuaciones estructurales empieza con la unión de diferentes materias, siendo una de las primeras la biométrica. Los principales trabajos de esta fueron aquellos publicados por Sewall Green Wright en la década de los 30, en el siglo pasado. Otra corriente fue la psicometría, que tuvo en Spearman (1904) a su principal actor con su aporte, el modelo factorial. Desde la econometría, la definición de ecuaciones simultáneas también será notable en la historia del SEM. En la década de 1970, Karl Gustav Jöreskog junto a Arthur Goldberger acoplan las nociones de estas corrientes e integrando el trabajo desarrollado por Dag Sörbom, crean el primer programa para la modelización de ecuaciones estructurales, LISREL.

De acuerdo con su estructura y con la naturaleza de las variables que contienen, hay diferentes tipos de modelos de ecuaciones estructurales: de trayectoria o senderos, factorial confirmatoria, factorial de segundo orden, de regresión estructural, mimic (multiple indicators and multiple causes of a single latent variable, Múltiples indicadores y múltiples causas de una sola variable latente), de crecimiento, etc. (Manzano & Zamora, 2009).

Los modelos de trayectoria o senderos (Path) es basan en un conjunto de variables dependientes, independientes e intermedias que se relacionan entre sí y pueden ser presentadas mediante un diagrama. Las fechas de un único sentido indican la causalidad entre las variables exógenas o intermedias y las dependientes. Además, existen flechas que conectan los términos de error con sus variables endógenas y las flechas de doble sentido reflejan la correlación entre los pares de variables exógenas (Lévy Mangin & Varela Mallou, 2006).

Este modelo es el más simple, en tanto que involucra únicamente variables observadas, y puede valorar la consecuencia indirecta que tiene una variable

sobre otra. Hay dos tipos de modelos de trayectoria, los recursivos y los no recursivos. No es posible que haya causalidad recíproca en los modelos recursivos, ciclos, ni incluso correlación entre los errores; en un modelo no recursivo (Manzano & Zamora, 2009).

4.2. Características de los modelos de ecuaciones estructurales

SEM utiliza varios tipos de modelos para describir las relaciones entre las variables observadas, con el objetivo básico de proporcionar una prueba cuantitativa de un modelo teórico planteado por el investigador.

Como ejemplos de aplicación de SEM, un investigador educativo podría plantear la hipótesis de que el ambiente familiar de un estudiante influye más tarde en su rendimiento en la escuela. Un investigador de marketing puede plantear la hipótesis de que la confianza del consumidor en una sociedad conduce a un aumento de las ventas de productos de esa corporación. Un profesional de la salud puede creer que una adecuada dieta y ejercicio regular disminuye la posibilidad de un ataque cardíaco.

En cada ejemplo citado en el párrafo anterior, el investigador cree, basado en la teoría y la investigación empírica, que conjuntos de variables definen (miden) constructos, que relacionados de una manera determinada, constituyen las hipótesis. SEM busca determinar el grado en que el modelo teórico es apoyado por datos de la muestra. En consecuencia, SEM pone a prueba modelos teóricos utilizando el método científico de la prueba de hipótesis para avanzar en la comprensión de las complejas relaciones entre los constructos.

Los SEM valoran en un análisis único, sistemático e integrador, el modelo de medida, que consiste en las relaciones entre las variables latentes y sus indicadores; y el modelo estructural, que es la parte del modelo completo que propone relaciones entre las variables latentes. Estas relaciones reflejan hipótesis sustantivas basadas en consideraciones teóricas (Gefen, Straub, & Boudreau, 2000).

El modelo factorial confirmatorio permite exponer la correlación entre variables latentes y la agrupación entre cada variable latente y sus respectivas variables observadas, está orientado a confirmar la estructura sugerida mediante el modelo (Manzano & Zamora, 2009).

SEM son ampliamente utilizados en la administración de empresa y en la psicología, por las facilidades que ofrecen de poder involucrar numerosas variables y sus relaciones en un mismo estudio (Cepeda Carrión & Roldan Salgueiro, 2004).

4.3. Terminología y notación básica en la modelización de SEM

La clave un diagrama SEM es la notación utilizada para etiquetar, los indicadores, los constructos y las relaciones entre ellos. Cada software utiliza su propio enfoque, aunque una convención estándar se ha asociado con LISREL.

Existen tres tipos de relaciones: relaciones de medición entre indicadores/ ítems y constructos, relaciones estructurales entre constructos y relaciones correlacionales entre constructos. También hay dos tipos de términos de error, uno relacionado con los indicadores individuales y el otro con los constructos. Las flechas muestran las relaciones entre constructos, así como entre constructos y sus respectivos indicadores; flechas unidireccionales son consideradas como relaciones predictivas y, con el apoyo de un fuerte soporte teórico, pueden ser interpretadas como relaciones causales.

Tabla 4 - 1 Notación básica de SEM

Variable	Descripción	Ejemplo
X	Variable observada (Indicador) dependiente	X_1
Y	Variable observada (Indicador) independiente	Y_1
λ	Coefficiente entre una variable observada o entre una variable latente y una observada	λ_1
ε	Error asociado a Y	ε_1
δ	Error asociado a X	δ_1
ξ	Variable latente independiente	ξ_1
η	Variable latente dependiente	η_1
ζ	Error asociado a η	ζ_1
β	Coefficiente entre variables latentes dependientes	β_1
γ	Coefficiente entre una variable latente independiente y una dependiente	γ_1

Nota: Tomado de (Manzano & Zamora, 2009)

En la **Figura 4 - 1** se presenta la terminología básica de SEM; se muestran con diferentes colores: el modelo estructural (interno), los modelos de medidas (externos), las variables latentes (endógenas y exógenas) y las variables manifiestas o indicadores.

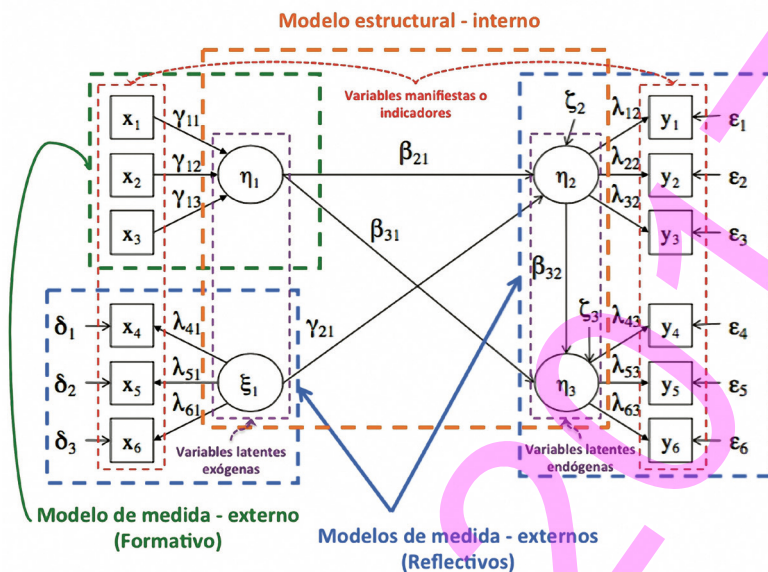


Figura 4 - 1. Terminologías básicas de SEM

Nota: Adaptado de (Cepeda Carrión & Roldan Salgueiro, 2004)

4.3. Enfoques de ecuaciones estructurales

Se tienen dos enfoques principales para la estimación de los parámetros en SEM. El modelo basado en análisis las covarianzas (MBC), que se orienta a la explicación y el de modelo de análisis de senderos mínimos cuadrados parciales (Partial Least Squares, PLSPM o simplemente PLS), que se orienta a estudios predictivos y que trabaja con las varianzas (Domínguez & Jesús, 2006).

PLS se puede utilizar para modelar investigaciones de variables latentes del comportamiento (como las actitudes o rasgos de personalidad), los compuestos se pueden aplicar para formar conceptos fuertes (Höök & Löwgren, 2012), es decir, la producción de artefactos tales como los instrumentos de gestión, los métodos de información o innovaciones. El modelo de senderos PLS es un instrumento estadístico favorito para estudios de factores de éxito (Albers, 2010).

El núcleo del PLS es una familia de mínimos cuadrados que emula y extiende el análisis de componentes principales, así como el análisis de correlación canónica. El método, que fue inventado por Wold (1974) para el análisis de datos de alta dimensión en un ambiente de baja estructura, ha sido objeto de varias ampliaciones y modificaciones.

4.5. Criterios para selección del enfoque de ecuaciones estructurales

Efectuar la evaluación de un modelo de ecuaciones estructurales ya sea ajustando la matriz de covarianzas o mediante mínimos cuadrados parciales, dependerá de los objetivos de investigación, primordialmente de la calidad del marco teórico precedente y de los fines confirmatorios o exploratorios del estudio.

Joe F. Hair et al. (2011) SEM is equivalent to carrying out covariance-based SEM (CB-SEM emiten una serie de criterios para guiar a los investigadores en su decisión de utilizar PLS o un análisis de modelo basado en covarianzas (MCB). Esto se resumen en la **Tabla 4 - 2**.

Tabla 4 - 2 Reglas generales para la selección de MCB-SEM ó PLS-SEM

Objetivos de investigación
<ul style="list-style-type: none"> • Si el objetivo es predecir constructos fundamentales o la identificación de constructos clave, elegir PLS.
<ul style="list-style-type: none"> • Si el objetivo es poner a prueba una teoría, conformación de una teoría o comparación entre teorías alternativas, seleccionar MBC.
<ul style="list-style-type: none"> • Si la investigación es exploratoria o una extensión de una teoría estructural existente, elegir PLS.
Especificación del modelo de medición
<ul style="list-style-type: none"> • Si hay constructos formativos como parte del modelo estructural, escoger PLS. (También se pueden utilizar constructos formativos con MBC, pero requiere tener en cuenta complicadas reglas de especificación)
<ul style="list-style-type: none"> • Si los términos de error requieren especificación adicional, como covarianza, escoger MBC.
Modelo estructural
<ul style="list-style-type: none"> • Si el modelo estructural es complejo (varios constructos con varios indicadores), escoger PLS.
<ul style="list-style-type: none"> • Si el modelo es no recursivo, elegir MBC.
<ul style="list-style-type: none"> • Si no se requiere experimentar la invariabilidad del modelo de medición, usar BC
Características de los datos y el algoritmo
<ul style="list-style-type: none"> • Si los datos reúnen los supuestos necesarios para BC de forma exacta (tamaño de la muestra, normalidad, etc.) escoger MBC; de otra forma, PLS.
<ul style="list-style-type: none"> • Consideraciones del tamaño de la muestra

Continúa

Tabla 4 - 2 Continuación

<ul style="list-style-type: none"> ○ Si la dimensión de la muestra es muy pequeña, elegir PLS. Con una gran cantidad de conjuntos de datos los resultados conseguidos con PLS y MBC son similares, siempre y cuando se utilicen un gran número de indicadores para medir los constructos latentes.
<ul style="list-style-type: none"> ○ La magnitud de muestra mínima para PLS debe ser igual al mayor de los siguientes dos aspectos: 1) Diez veces el número mayor de indicadores formativos utilizados para medir un constructo; ó, 2) Diez veces el mayor número de trayectorias estructurales dirigidas hacia un constructo latente en particular en el modelo estructural.
<ul style="list-style-type: none"> ● Si los datos son anormales en cualquier magnitud, usar PLS; de otra manera, bajo condiciones de normalidad en los datos, los resultados en PLS y en MBC son similares, con MBC arrojando resultados más precisos en la estimación del modelo.
<ul style="list-style-type: none"> ● Si no se cumplen los supuestos exigidos por MBC (especificación del modelo, identificación, no convergencia, supuestos en la distribución de los datos) puede usarse PLS como aproximación a los resultados de MBC.
<ul style="list-style-type: none"> ● Los resultados con PLS y con MBC deben ser semejantes; de no ser así debe revisarse la especificación del modelo para ratificar una correcta aplicación de MBC, si no los resultados de PLS son buena proximidad a los de MBC.
Evaluación del modelo
<ul style="list-style-type: none"> ● Si se requiere utilizar las puntuaciones de las variables latentes en análisis posteriores, PLS es la mejor técnica.
<ul style="list-style-type: none"> ● Si la investigación demanda un criterio de bondad de ajuste global, MBC es el mejor procedimiento.
<ul style="list-style-type: none"> ● Si no se requiere experimentar la invariabilidad del modelo de medición, usar MBC

Nota: Tomado de Joe F. Hair et al. (2011)

Los SEM son ampliamente utilizados en la administración de empresa y psicología, por la facilidad de poder involucrar varias variables y sus relaciones en un mismo estudio (Cepeda Carrión & Roldan Salgueiro, 2004).

Para sustentar mejor la decisión de utilizar enfoque en PLS, en concreto se puede sugerir los argumentos siguientes:

- Orientación del estudio a la predicción.
- Tamaño de la muestra no muy grande.
- Distribución de los datos no demuestra normalidad.
- Utilización de indicadores reflexivos.
- El modelo no posee términos de error que requieran covariar.
- Se requiere experimentar la invariabilidad del modelo de medición, pues se realiza un análisis multigrupo.

Sobre el tema específico de la normalidad de los datos. Varios son los estudios que dan argumentos a favor de la no normalidad requerida al momento de aplicar SEM con enfoque PLS. La **Tabla 4 - 3** resume los principales.

Tabla 4 - 3 Estudios con argumento a favor de la no normalidad en PLS

Argumento	Fuente
<ul style="list-style-type: none"> Respecto a la distribución de las variables, no se requiere que los datos correspondan a distribuciones normales o conocidas. 	(Falk & Miller, 1992)
<ul style="list-style-type: none"> En análisis con enfoque PLS, no se puede suponer que se cuenta con una distribución normal multivariada y observaciones independientes (paramétricas). 	(Components & Variables, 2014)
<ul style="list-style-type: none"> El método PLS no atribuye grandes restricciones al modelo. Conjuntamente, la hipótesis de normalidad de dichos fundamentos raramente se asimilan realmente, aunque se pueda obviar esta restricción, los resultados y decisiones apoyadas en ellos quedan visiblemente comprometidas. 	(Domínguez & Jesús, 2006)
<ul style="list-style-type: none"> El análisis de senderos utilizando PLS es particularmente apropiado cuando hay muchas variables manifiestas y latentes, en los que todas o algunas de las variables manifiestas son categóricas, donde las distribuciones no son normales y en condiciones de heterocedasticidad. 	(Ziersch, Baum, Macdougall, & Putland, 2005)
<ul style="list-style-type: none"> PLS es adecuado cuando la normalidad no se puede demostrar. 	(Ayodeji & Al-lawati, 2010)
<ul style="list-style-type: none"> El procedimiento PLS ha ido ganando interés debido a su capacidad para modelar constructos latentes en condiciones de no normalidad y para tamaños de muestras pequeños y medianos. 	(Chin, Marcolin, & Newted, 2003)
<ul style="list-style-type: none"> PLS es capaz de manejar las deficiencias de datos, tales como datos que no son normales y tiene capacidad para medidas formativas de los constructos. 	(Coltman, Devinney, Midgley, & Venaik, 2008)
<ul style="list-style-type: none"> Entre las razones más frecuentemente para el uso de PLS están, tamaño pequeño de la muestra, los datos no normales, el uso de constructos formativos. 	(Kaufmann & Gaeckler, 2015)

PLS ha disfrutado de popularidad creciente en la investigación por su capacidad para modelar constructos latentes con la condición de no normalidad como el presente estudio (Chin, 1998a). Las variables recogidas en las encuestas no eran normales, por lo que se prefirió PLS en lugar de SEM a base de covarianza. Fundamentalmente porque esta técnica es adecuada para la evaluación de modelos orientados a la predicción (Sarstedt, Ringle, Smith, Reams, & Hair, 2014), como es el caso del presente estudio. La herramienta utilizada para este fin fue el software SmartPLS versión 3.2.4 (Garson, 2016).

4.6. Software para tratamiento de ecuaciones estructurales

LISREL, por entonces un programa basado en comandos, fue el primer programa informático para resolver SEM. Un segundo programa es EQS, que ofrece una especificación modelo de libre matriz y que acepta algunos tipos de variables: variables de medidas, variables latentes, y variables residuales o error. Por su parte, AMOS aporta una interfaz gráfica que permite a los usuarios especificar su modelo dibujando los diagramas de ruta. Mplus permite especificar el modelo sin ser necesario enumerar las varianzas residuales que son parte del modelo, ya que este programa adiciona los parámetros automáticamente conservando la sintaxis (Rosseel, 2012).

Para el caso específico de modelado PLS los principales programas informáticos disponibles se muestran en la **Tabla 4 - 4**, siendo elegido SmartPLS para el presente estudio, por su facilidad de uso y la forma en que presenta los resultados.

Tabla 4 - 4 Programas informáticos para tratamiento de PLS

Programa	Sitio Web	Tipo de licencia
ADANCO	http://www.composite-modeling.com	De pago
LVPL	http://www2.kuas.edu.tw/prof/fred/vpls/aboutPLSPC.htm	Abierta
matrixpls	https://cran.r-project.org/web/packages/matrixpls/matrixpls.pdf	Abierta
PLS-Graph	http://www.plsgraph.com/	De pago
plspm	http://cran.r-project.org/web/packages/plspm/plspm.pdf	Abierta
PLS-GUI	http://pls-gui.com	De pago
semPLS	http://cran.r-project.org/web/packages/semPLS/index.html	Abierta
SmartPLS	http://www.smartpls.com/	De pago
VisualPLS	http://fs.mis.kuas.edu.tw/~fred/vpls/index.html	Abierta
WarpPLS	http://www.scriptwarp.com/warppls/	De pago
XLSTAT-PLSPM	http://www.xlstat.com/en/products/xlstat-plspm/	De pago

4.7. Modelos de ecuaciones estructurales con enfoque PLS

La apreciación de los parámetros del modelo PLS tiene cuatro pasos: primero, un algoritmo iterativo que establece las puntuaciones compuestas para cada constructo; segundo, una corrección para la atenuación de esos constructos que se modelan como factores (Dijkstra & Henseler, 2015a); tercero, la estimación de parámetros; y por último, bootstrapping (se define más adelante) para las pruebas de significancia estadística.

En las dos últimas décadas el modelo de medida (también llamado modelo externo) se ha basado en la utilización de varias metodologías encargadas de evaluar variables latentes a partir de las variables observadas. Algunas de las áreas en las que se ha aplicado este tipo de técnica son: sociología, psicología, mercadotecnia, economía, entre otras. Es necesario mencionar que hay varios modelos de medida, y su aplicación obedece a un planteamiento teórico particular.

El modelo estructural está compuesto de dos tipos de constructos: exógenos y endógenos relacionados entre sí. Los constructos exógenos se hallan afuera del modelo. Por lo tanto, las variables exógenas no reciben influencia de otros constructos en el modelo, es decir, no hay ninguna flecha en el modelo estructural que apunte a los constructos exógenos. En contraste, los constructos endógenos se explican, al menos parcialmente por otros constructos en el modelo. Cada constructo endógeno debe tener al menos una flecha del modelo estructural que apunta a la misma. Las relaciones entre los constructos por lo general suponen ser lineales. El tamaño y la importancia de relaciones de rutas es generalmente el foco de los esfuerzos científicos que se persiguen en la investigación empírica.

4.7.1. Valoración del modelo global

La bondad de ajuste global del modelo es el punto de partida de la valoración del mismo. Si el modelo no se ajusta a los datos, los datos contienen más información que la que el modelo proporciona; las estimaciones obtenidas podrían carecer de sentido, y las conclusiones que se alcanzaran podrían ser cuestionables (Henseler, Hubona, & Ray, 2016).

Según Henseler, Hubona & Ray (2016) el ajuste del modelo global se puede llevar a cabo por medio de dos vías no excluyentes: por medio de estadísticos inferenciales, pruebas de ajustes exactos basados en bootstrap; por medio de índices de ajuste, que proporcionan una valoración aproximada del ajuste del modelo. Para este estudio se utilizó la segunda opción.

Entre las medidas de ajuste aproximados tenemos:

- Al Standardized root mean square residual (SRMR), que mide la diferencia entre la matriz de correlaciones observada y la matriz de correlaciones implicada por el modelo. SRMR refleja la magnitud media de tales diferencias: cuanto más bajo sea el SRMR, mejor ajuste. Un modelo tiene un buen ajuste cuando $SRMR < 0.08$ (Hu & Bentler, 1998). Ringle propone una opción más flexible de $SRMR < 0.10$.

- Otro índice es Normed Fit Index (NFI) o índice Bentler & Bonett. Cuanto más cerca de 1, mejor. Valores por encima de 0,9 representan un ajuste aceptable.
- El índice root mean square error correlation (RMS Theta), también se utilizar para evaluar el ajuste del modelo; para este índice los umbrales aún están por determinar en PLS; inicialmente, valores por debajo de 0.12, se pueden considerar como buen ajuste.

En resumen para este estudio al momento de evaluar el ajuste global de modelo se utilizó SRMR, NFI y RMS Theta.

4.7.2. Valoración del modelo de medida

4.7.2.1. Fiabilidad individual de los indicadores

La fiabilidad individual de los indicadores reflectivos es valorada examinando las cargas factoriales (λ) o correlaciones simples, de las medidas o indicadores con su respectivo constructo. Los indicadores son fiables si $\lambda \geq 0.707$ (Carmines & Zeller, 1979). Varios investigadores sostienen que esta regla heurística no debería ser tan rígida en etapas iniciales de desarrollo de escalas (Chin, 1998b) y cuando las escalas se aplican a contextos diferentes (Barclay, Higgins, & Thompson, 1995).

La inclusión de ítems débiles ayudará a extraer la información útil que está disponible en el indicador para crear una mejor puntuación de la variable latente.

Los indicadores débiles en ocasiones son retenidos sobre la base de su contribución a la validez de contenido, pero indicadores con cargas muy bajas, menor que 0.4, deberían ser eliminados (Joe F. Hair et al., 2011) SEM is equivalent to carrying out covariance-based SEM (CB-SEM). En este caso, el modelo de medida necesita ser ajustado y el algoritmo PLS ejecutado de nuevo para obtener nuevos resultados (Urbach & Ahlemann, 2010).

4.7.2.2. Fiabilidad del constructo

La fiabilidad de constructo, de escala o consistencia interna determina si los ítems que miden un constructo son similares en sus puntuaciones. Además evalúan la rigurosidad con la que se están midiendo las variables manifiestas a la misma variable latente. Para este propósito las medidas son: Coeficiente alfa de Cronbach (α); Fiabilidad compuesta (ρ_c) (Garson, 2016).

La fiabilidad compuesta es más adecuada que el alfa de Cronbach para PLS, ya que no asume que todos los indicadores reciben la misma ponderación. Se sugiere el valor de 0.7 como un nivel adecuado para una fiabilidad 'modesta' en etapas tempranas de investigación, y un más estricto 0.8 o 0.9 para etapas más avanzadas de investigación. El índice Dijkstra-Henseler's (ρ_A) (rho_A) también fue evaluado y se considera como una medida de fiabilidad consistente (Dijkstra & Henseler, 2015b).

4.7.2.3. Validez convergente

La validez convergente implica que un conjunto de indicadores representa un único constructo subyacente, pudiendo ser demostrado esto por medio de su unidimensionalidad (Henseler, Ringle, & Sinkovics, 2009). La varianza extraída media (average variance extracted- AVE) proporciona la cantidad de varianza que un constructo obtiene de sus indicadores con relación a la cantidad de varianza debida al error de medida (Fornell & Larcker, 1981). AVE es una medida más conservadora que la fiabilidad compuesta, y se recomienda un $AVE \geq 0.50$, que significa que cada constructo explica al menos el 50% de la varianza de los indicadores asignados.

4.7.2.4. Validez discriminante

La validez discriminante indica en qué medida un constructo dado es diferente de otros constructos. Existían dos métodos clásicos para su valoración:

- El análisis de cargas cruzadas. Ningún ítem debería cargar más fuertemente sobre otro constructo que sobre aquel que trate de medir (Barclay et al., 1995). A su vez, cada constructo debería cargar más sobre sus indicadores asignados que sobre otros ítems. El análisis de cargas cruzadas se lleva a cabo calculando las correlaciones entre las puntuaciones de los constructos y los datos estandarizados de los indicadores (Gefen & Straub, 2005).
- El criterio de Fornell y Larcker (1981). La cantidad de varianza que un constructo captura de sus indicadores (AVE) debería ser mayor que la varianza que dicho constructo comparte con otros constructos en el modelo (la correlación al cuadrado entre los dos constructos). Con la intención de facilitar esta valoración, la raíz cuadrada del AVE de cada variable latente debería ser mayor que las correlaciones que tiene con el resto de variables latentes del modelo. Los elementos en la diagonal son la raíz cuadrada de la varianza compartida entre el constructo y sus medidas (AVE); los elementos fuera de la diagonal son las correlaciones

entre constructos; Para lograr la validez discriminante, la raíz cuadrada del AVE de un constructo debería ser mayor que la correlación que este tenga con cualquier otro constructo.

El examen de las cargas cruzadas y del criterio Fornell-Larcker presenta deficiencias. Ambos no son lo suficientemente sensibles para detectar problemas de validez discriminante. Ambos funcionan bien con tamaños muestrales altos y con patrones de cargas muy heterogéneos.

Henseler, Ringle, & Sarstedt (2016) desarrollaron estudios de simulación para demostrar que la falta de validez discriminante es detectada mejor por medio del ratio heterotrait-monotrait (HTMT) que ellos desarrollaron. La ratio heterotrait-monotrait (HTMT), representa el promedio de las correlaciones heterotrait-heteromethod en relación al promedio de las correlaciones monotrait-heteromethod. Correlaciones Monotrait-heteromethod, son las correlaciones entre los indicadores que miden el mismo constructo. Correlaciones Heterotrait-heteromethod, son las correlaciones entre indicadores de constructos distintos que miden fenómenos diferentes. En un modelo bien ajustado, las correlaciones heterotrait deberían ser más pequeñas que las correlaciones monotrait, lo que implica que la ratio HTMT debería estar por debajo de 1. Los umbrales aceptables son: 0.85 (Kline, 2011) HTMT 0.85 y 0.90 (Gold, Malthora, & Segars, 2001) HTMT 0.90.

4.7.3. Valoración del modelo estructural

4.7.3.1. Valoración de problemas de colinealidad

Dado que la estimación de los coeficientes de trayectoria (path) se hace con base en regresiones de mínimos cuadrados ordinales (OLS), al igual que en una regresión múltiple, debemos de evitar la presencia de multicolinealidad entre las variables antecedentes de cada uno de los constructos endógenos. De acuerdo con Garson (2016), existirán indicios de multicolinealidad cuando $FIV > 5$, y con niveles de tolerancia < 0.20 .

4.7.3.2. Evaluación de los coeficientes path

Los coeficientes path (coeficientes de regresión estandarizados) muestran las estimaciones de las relaciones del modelo estructural, es decir, de las relaciones hipotetizadas entre constructos. Se analizará el signo algebraico, la magnitud y la significación estadística.

Aquel path cuyo signo sea contrario al signo postulado en la hipótesis, conducirá a inferir que ésta no sea soportada. La magnitud de los coeficientes

path se muestran como valores estandarizados entre +1 y -1. Mayores valores absolutos denotan mayores relaciones (predictivas) entre constructos; cuanto más cercano a cero es el valor, más débil es la relación. Valores muy bajos, cercanos a cero, son habitualmente no significativos, es decir, no significativamente diferente de cero.

Para la valoración de la significación de los efectos se utiliza la técnica del Bootstrapping, que es una técnica de remuestreo no paramétrica, que consiste en un muestreo repetido, aleatorio, con reposición de la muestra original, que genera un número de muestras bootstrap, lo que permite responder a la pregunta: ¿Son las relaciones significativamente diferentes de cero? Se sugiere un mínimo de 5000 muestras y el número de casos debe ser igual al número de observaciones en la muestra original. Como resultado de este proceso, se obtienen los errores estándar, los estadísticos t y los intervalos de confianza de los parámetros, lo cual permite las pruebas de hipótesis (Joe F. Hair et al., 2011).

En la valoración de la significación estadística por medio de bootstrapping, en esta investigación se considera la distribución t de student, considerando que emplea hipótesis que especifican la dirección de la relación (+ ó -) de las variables del tipo, la variable A tendrá una influencia positiva sobre la variable B ($A \rightarrow B$). Por lo tanto, se usa una distribución t de student de 1 cola con $n-1$ grados de libertad, donde n es el número de sub-muestras. Para $n = 5000$ sub-muestras: * $p < .05$; ** $p < .01$; *** $p < .001$ (basado en una distribución t (4999) de Student de una cola). $t(0.05; 4999) = 1,645$; $t(0.01; 4999) = 2,327$; $t(0.001; 4999) = 3,092$. Si el valor empírico de t es mayor que el valor crítico de t para un nivel de significación seleccionado, entonces el coeficiente es significativamente distinto de cero.

4.7.3.3. Valoración del coeficiente de determinación

El coeficiente de determinación (R^2) representa una medida de poder predictivo, en tanto que indica la cantidad de varianza de un constructo que es explicada por las variables predictoras de dicho constructo endógeno en el modelo. Los valores de R^2 oscilan entre 0 a 1; cuanto más alto es el valor, más capacidad predictiva tiene el modelo para dicha variable.

Respecto a la descomposición del valor R^2 , la varianza explicada en un constructo endógeno por otra variable latente viene dado por el valor absoluto del resultado de multiplicar el coeficiente path por el correspondiente coeficiente de correlación entre ambas variables.

Los valores de R^2 deberían ser suficientemente altos para que el modelo

alcance un nivel mínimo de poder explicativo. Falk & Miller (1992) sugieren como mínimo ≥ 0.10 ; Chin (1998b) expone que 0.67 es sustancial, 0.33 es moderado y 0.19 es débil; Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt (2014)(Marketing) manifiestan que 0.75 es sustancial, 0.5 es moderado y 0.25 se puede considera débil.

Un problema asociado al R^2 , es que este se incrementa cuando constructos predictores adicionales se incluyen en el modelo. Ante tal situación se usa R^2 ajustado, lo cual permite controlar la complejidad del modelo cuando se comparan diferentes configuraciones de modelos (diferentes números de variables exógenos y/o diferentes muestras).

4.7.3.4. Valoración de los tamaños de los efectos

El tamaño del efecto (f^2) valora el grado con el que un constructo exógeno contribuye a explicar un determinado constructo endógeno en términos de R^2 (Cohen, 1988).

El cálculo es, $f^2 = (R^2_{\text{incluida}} - R^2_{\text{excluida}}) / (1 - R^2_{\text{incluida}})$, donde R^2_{incluida} y R^2_{excluida} son los valores R^2 de la variable latente endógena cuando una variable latente exógena es incluida o excluida del modelo. El cambio en los valores de R^2 se calculan estimando el modelo PLS dos veces: una con la variable latente exógena incluida (generando R^2_{incluida}); y, la segunda vez, con la variable latente exógena excluida (generando R^2_{excluida}).

Una regla heurística de Cohen (1988) para evaluar f^2 sostiene que: $0.02 \leq f^2 < 0.15$, es un efecto pequeño; $0.15 \leq f^2 < 0.35$, es un efecto moderado y $f^2 \geq 0.35$ es un efecto grande. Un modelo anidado debería ser rechazado si no produce un f^2 significativo.

4.7.3.5. Valoración de la relevancia predictiva

Para medir la relevancia predictiva de los constructos dependientes reflectivos, se utiliza el test de Stone-Geisser (Q^2). Sigue un procedimiento de blindfolding, donde se omiten parte de los datos para un determinado constructo durante la estimación de parámetros, para a continuación intentar estimar lo que se ha omitido, usando los parámetros estimados (Chin, 1998b).

Se pueden estimar diferentes Q^2 según la forma de la predicción: Cross-validated communality Q^2 , se obtiene si la predicción de los datos se hace a partir de las puntuaciones (scores) de la variable latente subyacente; y, Cross-validated redundancy Q^2 , se obtiene si la predicción es realizada por aquellas variables latentes que predicen la variable endógena en cuestión. Éste último

es el que se debería usar si queremos examinar la relevancia predictiva del modelo teórico o estructural. Hay una regla general que expresa que si $Q^2 > 0$ tiene relevancia predictiva, caso contrario no.

Siguiendo la sugerencia de Cohen (1988). Se puede valorar el impacto relativo sobre la relevancia predictiva por medio del indicador $q^2 = (Q^2_{\text{incluido}} - Q^2_{\text{excluido}}) / (1 - Q^2_{\text{incluido}})$. Donde 0.02 representa un tamaño del efecto pequeño, 0.15 representa un tamaño del efecto medio, y 0.35 representa un tamaño del efecto de altamente predictivo.

4.7.4. Análisis multigrupo

Se consideró importante realizar el análisis multigrupo para aumentar la relevancia de la investigación y determinar si la categoría de universidad modera las relaciones de los modelos. Las variables moderadoras pueden ser variables categóricas (género, clase social o categoría de universidad – como en esta investigación) y también pueden ser constructos medidos por escalas tipo Likert. Se denominan así por su influencia en la fuerza y/o dirección de una relación entre una variable exógena y una endógena (Baron & Kenny, 1986).

Para determinar la influencia señalada, se deben formar dos grupos o submuestras basadas en la variable categórica y compararlas; siendo utilizadas en este estudio la categoría de universidad. A continuación, el modelo propuesto se calcula por separado para cada grupo de observaciones. Finalmente, las diferencias estadísticamente significativas entre los coeficientes de trayectoria de los grupos se interpretan como efectos moderadores (Qureshi & Compeau, 2009).

En la presente investigación al tener cuatro categorías de universidad en Ecuador (A, B, C y D) y al aplicar lo escrito en el párrafo anterior, resultaron 6 pares de comprobaciones para el modelo de profesores y 6 para el modelo de estudiantes. Los pares fueron: A-B, A-C, A-D, B-C, B-D y C-D.

Es importante destacar que antes de realizar las comprobaciones multigrupo, se debe verificar la invarianza de medida.

4.7.4.1. Invarianza de medición

La comparación de los modelos, tal y como se expuso en el apartado anterior, tiene un sentido lógico solamente si los constructos se miden de la misma manera en cada grupo y, por lo tanto, tienen el mismo significado. Es decir, el establecimiento de la invarianza de medición es un requisito previo necesario

para realizar el análisis multigrupo (Garson, 2016). Para evaluar la invarianza de medición se utilizó el procedimiento MICOM (*measurement invariance of composite models*) en SmartPLS (Henseler, Ringle, et al., 2016). MICOM es un procedimiento de tres pasos:

- **Paso 1:** Invarianza de configuración (Configural invariance). Mismos indicadores en el modelo de medida, igual tratamiento en los datos, igual ajustes en el algoritmo o en los criterios de optimización e igual naturaleza constructos (Modo A, Modo B).
- **Paso 2:** Invarianza de compuesto, que ocurre cuando las puntuaciones de un compuesto usando los pesos del grupo 1 no difieren de aquellas creadas usando los pesos del grupo 2. En concreto, el algoritmo comprueba la hipótesis nula de que la correlación (c), c , en los datos originales de los grupos es igual a uno. Si c es menor al 5% de la distribución de c en las permutaciones de los datos agrupados, se rechaza la hipótesis de la invarianza de compuesto. Un hallazgo de medias no-significativas implica que la invarianza de compuesto puede ser asumida. Esto ocurrirá cuando las correlaciones no son significativamente inferiores a 1.0.
- **Paso 3:** Invarianza escalar, esto implica: igualdad de compuestos de 3A) medias y 3B) varianzas. Se calculan de manera análoga al paso 2. ¿Las observaciones del primer grupo tienen la misma media y varianza en los valores de las puntuaciones del constructo que los del segundo grupo?

El paso 1, no se calcula en SmartPLS. Si los pasos 2 y 3, no tienen significación, es “Invarianza de medición completa”. Si solamente paso 2 no tiene significación, hay “Invarianza de medición parcial” (Garson, 2016).

4.7.4.2. Significancia de las diferencias

Para confirmar la importancia de la diferencia entre los coeficientes de trayectoria de los grupos que se evalúan, en esta investigación se utiliza el método basado en permutaciones. Además, como ya se expuso, el algoritmo puede ser utilizado para implementar el procedimiento de evaluación de la invarianza de medición en PLS-SEM (MICOM) (Reguera-Alvarado, Blanco-Oliver, & Martín-Ruiz, 2016).

En MICOM las permutaciones se crean con observaciones generadas aleatoriamente del conjunto inicial de datos (sin reposición). Más concretamente, las primeras n observaciones se generan sin reposición y se asignan al grupo A.

El tamaño es igual al número de observaciones que originariamente tenía el grupo A. Los datos restantes se asignan al grupo B que también tienen el mismo número de observaciones que el grupo B. Consecuentemente, en cada permutación realizada, el tamaño de la muestra de cada grupo específico permanece constante y de igual tamaño que en el grupo en origen. Para asegurar la estabilidad de los resultados, el número de permutaciones debería ser grande de 500-100 en una rápida valoración inicial, y para resultados finales 5000 (Garson, 2016).

Para MICOM no se establecen suposiciones de la distribución, se considera el test más fiable y de uso recomendado por encima otros. Para su procesamiento se utilizó el software SmartPLS, que facilita la ejecución de los pasos 2 y 3 (Schubring, Lorscheid, Meyer, & Ringle, 2016). El nivel de significación que se considera por defecto es 0.05.

Una vez revisado el proceso metodológico general seguido, es importante señalar que con la finalidad de tener mejores instrumentos de recolección de datos, se realizaron dos captaciones previas a la recogida general de los datos a nivel nacional.

4.8. Modelo de ecuaciones estructurales con enfoque en covarianzas

4.8.1. Introducción

El MBC es un enfoque utilizado para la confirmación de teorías y requiere de cumplimiento más estricto del comportamiento de los datos, por ejemplo se requiere normalidad y una cantidad mayor de elementos en la muestra que un enfoque PLS.

4.8.2. Especificación del modelo

La especificación del modelo consiste en describir las relaciones causales entre los constructos planteados para la investigación, se identifican los parámetros del modelo y se detallan los indicadores (ítems) que se utilizarán para medir las variables no observables. El número de ecuaciones será igual al número de constructos latentes (variables endógenas) que se explican por medio de variables exógenas (latentes u observadas). Se definirá el modelo de medida y el modelo estructural.

4.8.3. Identificación del modelo

En esta parte del proceso se verifica si los parámetros del modelo completo pueden calcularse con los elementos de la matriz de covarianzas de las

variables observables. Se cuenta el número total de variables endógenas y exógenas (s). Luego, el número de elementos no redundantes en el modelo, que es igual a: $\frac{1}{2} s (s + 1)$.

Además, se identifica al número total de parámetros a ser estimados en el modelo como (t). Finalmente para saber la identificación del modelo se debe tener la siguiente condición necesaria $t \leq \frac{1}{2} s(s+1)$. De manera que si se tiene la igualdad, el modelo está identificado; si t es menor que $\frac{1}{2} s(s+1)$, el modelo está sobre identificado y si t es mayor que $\frac{1}{2} s(s+1)$, el modelo no está identificado.

4.8.4. Estimación del modelo

Primero se identificarán, corregirán inconsistencias en las estimaciones de los coeficientes de los modelos de medida y estructurales, tales como varianzas de error negativas o varianzas de error no significativas para cualquiera de los constructos; coeficientes estandarizados cuyos valores sobrepasen 1.000; errores estándar muy elevados asociados con cualquier coeficiente estimado, etc. Se continuará con el ajuste de los modelos de medida, modelo estructural y el modelo global.

En SEM se aplican diferentes métodos de estimación de los parámetros, para determinar cuál de ellos presenta un mejor ajuste: máxima verosimilitud o ML (*maximum likelihood*), mínimos cuadrados ponderados o WLS (*weighted-least squares*) y mínimos cuadrados generalizados o GLS (*generalized least squares*).

En función de la normalidad o no de las variables y de otros factores, puede decidirse cuál es el mejor método de estimación. A priori, en vista del número de elementos que se estima captar, es de esperar que se cumpla la condición de normalidad, por lo que es probable que se opte por utilizar la técnica ML (García, 2011). En cualquier caso, la técnica finalmente aplicada, será la adecuada según las necesidades impuestas por los datos obtenidos.

4.8.5. Evaluación del modelo

La evaluación consiste en determinar si con los cálculos obtenidos hasta este punto se puede determinar si el modelo representa o no la realidad que se estudia. Para ello se tienen un conjunto de medias que sirven para evaluar el modelo de medida, el modelo estructural y al modelo global.

Para evaluar y ajustar el modelo de medida se aplica análisis factorial, que

relaciona los indicadores con variables latentes. Para la presente investigación se utilizará el análisis factorial confirmatorio (CFA). Para evaluar y ajustar el modelo estructural se emplea el análisis path o de senderos, es el que representa las relaciones entre las variables latentes. (García, 2011). Para una evaluación global se tienen: medidas absolutas de ajuste, medidas incrementales de ajuste y medidas de ajuste de parsimonia. Se revisarán para determinar si hay una correspondencia perfecta entre la matriz reproducida por el modelo y la matriz de observaciones.

4.8.6. Modificación del modelo

Modificar el modelo para mejorar tanto en su ajuste como en la complejidad, se estiman los parámetros del modelo y se calculan los índices de ajuste global. Al modelo modificado también se le aplicará el análisis de la consistencia interna, alfa de Cronbach, para asegurar que todo sigue en orden.

Se revisará cada constructo y se replantearán los indicadores o variables observables con la finalidad de obtener un modelo que mejor se ajuste. En resumen con la evaluación de la fase anterior se define un nuevo modelo que puede omitir indicadores o incluso constructos.

4.9. Ejemplo de SEM enfoque PLS

4.9.1. Introducción

Para ilustrar la aplicación de SEM con enfoque PLS, exponemos un estudio que consistió en la construcción de un modelo conceptual para predecir el nivel de aceptación y uso del sistema de e-learning entre los profesores.

Este modelo extiende el Modelo de Aceptación de la Tecnología 2, e incluye soporte técnico, autoeficacia computacional, influencia social, disfrute percibido y satisfacción. Se recolectaron datos de 145 profesores de una Universidad del Ecuador.

Para la recolección de datos, un formulario en escala Likert fue enviado por correo electrónico. Con la finalidad de tener una idea de los profesores que integraron la muestra, en la **Tabla 4 - 5** se pueden apreciar las principales características demográficas.

El procesamiento de los datos se realizó a través del software estadístico SmartPLS. Este trabajo, realizó un proceso de imputación de datos, que se describió en el capítulo 1 y fue presentado en un congreso internacional (Ramirez-Anormaliza et al., 2016).

Tabla 4 - 5 Datos demográficos de profesores en el estudio

Características		Frecuencia	Porcentaje
<i>Género</i>	Masculino	83	57.24
	Femenino	62	42.76
<i>Edad</i>	De 20 a 30 años	17	11.72
	De 30 a 40 años	61	42.07
	Más de 40	67	46.21
<i>Experiencia en internet</i>	De 1 a 2 años	1	0.69
	Más de 2 años	144	99.31
<i>Tiempo de Dedicación</i>	Medio	14	0.10
	CompletoW	131	0.90

4.9.2. Pasos

4.9.2.1. Instalación del software

Para iniciar el procesamiento de los datos, lo primero que se debe realizar es la instalación del software SmartPLS, para el efecto se sugiere acceder a la siguiente dirección en internet <https://www.smartpls.com/> (Ringle, Wende, & Becker, 2015). El software elegido es de uso comercial, pero ofrece la opción de uso a estudiantes que realicen proyectos con hasta 100 observaciones o registros.

4.9.2.2. Crear un proyecto

Una vez instalado es paquete informáticos SmartPLS, su interface de usuario es muy sencilla e intuitiva, se procedió a crear un proyecto, dando clic en: **Fichero** → **Crear nuevo proyecto**, como lo muestra **Figura 4 - 2**. También se puede utilizar la barra de herramientas de la aplicación (**Crear nuevo proyecto**).

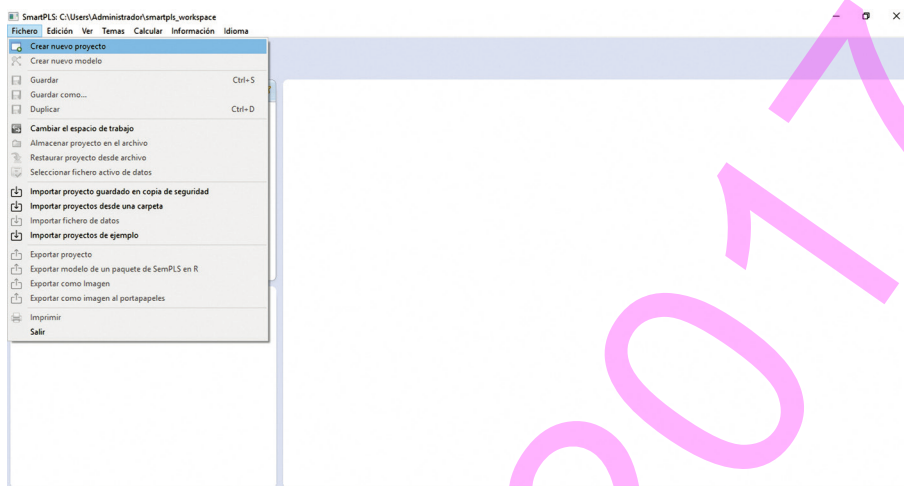


Figura 4 - 2. Crear nuevo proyecto en SmartPLS

4.9.2.3. Crear un modelo

Luego de seleccionar con un clic el proyecto creado en el paso anterior, se procede a crear un modelo, para el efecto puede dar doble clic sobre el elemento que tiene el ícono de un modelo o dar clic en la barra de herramientas de SmartPLS en la opción **(Nuevo Modelo)**. La Figura 4 - 3 muestra esta acción.

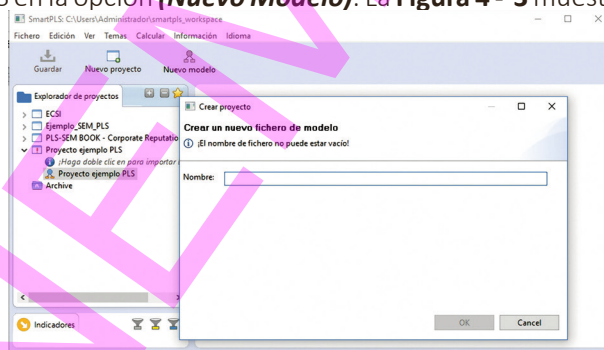


Figura 4 - 3. Crear nuevo modelo

4.9.2.4. Crear variables latentes y conectarlas

Con el modelo creado se procede a crear las variables latentes, para el efecto se utiliza la opción **(Variable latente)** de la barra de herramientas. Al momento de crear las variables latentes, estas aparecen un nombre genérico, el mismo que debe ser cambiado, presionado **(F2)**, mientras se tiene seleccionada la

variable latente.

Para crear la relación entre las variables latentes, se debe dar clic sobre la opción **(Conectar)** de la barra de herramientas, luego dar clic sobre la variable latente donde empieza la relación, acto seguido, dar clic en la variable latente donde termina la relación. Esto se puede observar en las **Figura 4 - 4** y **Figura 4 - 5**.

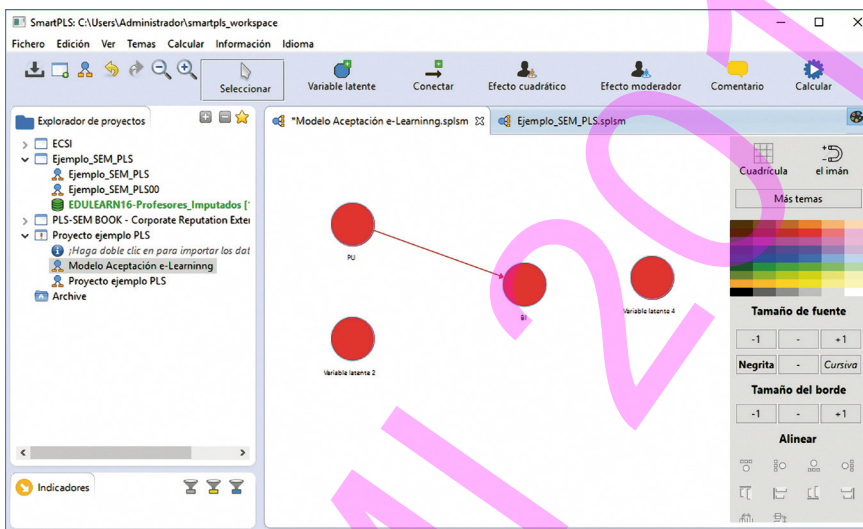


Figura 4 - 4. Crear variables latentes

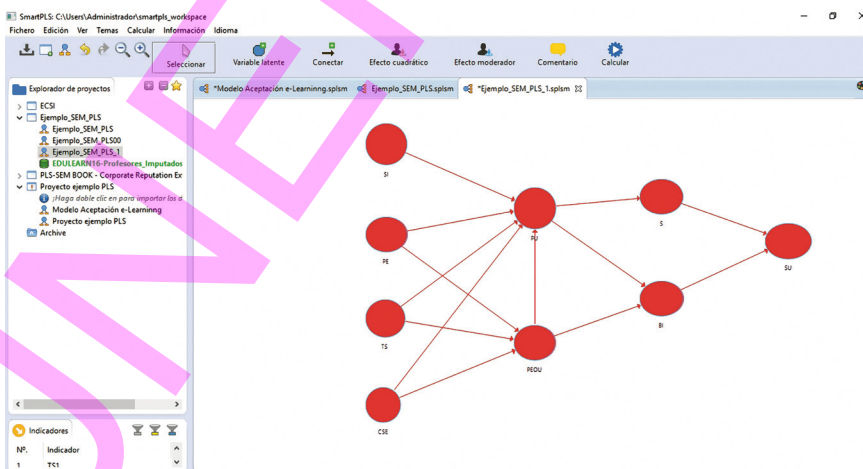


Figura 4 - 5. Conectar variables latentes

4.9.2.5 Carga de datos

Se procede con el modelo creado a cargar los datos dando doble clic en la opción del menú (**Modelo Aceptación e-Learning**) luego cargamos la información, importándola desde la (**carpeta estipulada como almacenamiento en la unidad del computador**), en los cuales los tenemos almacenados, se cargan los indicadores según los datos que se importaron. La **Figura 4 - 6** ilustra lo descrito.

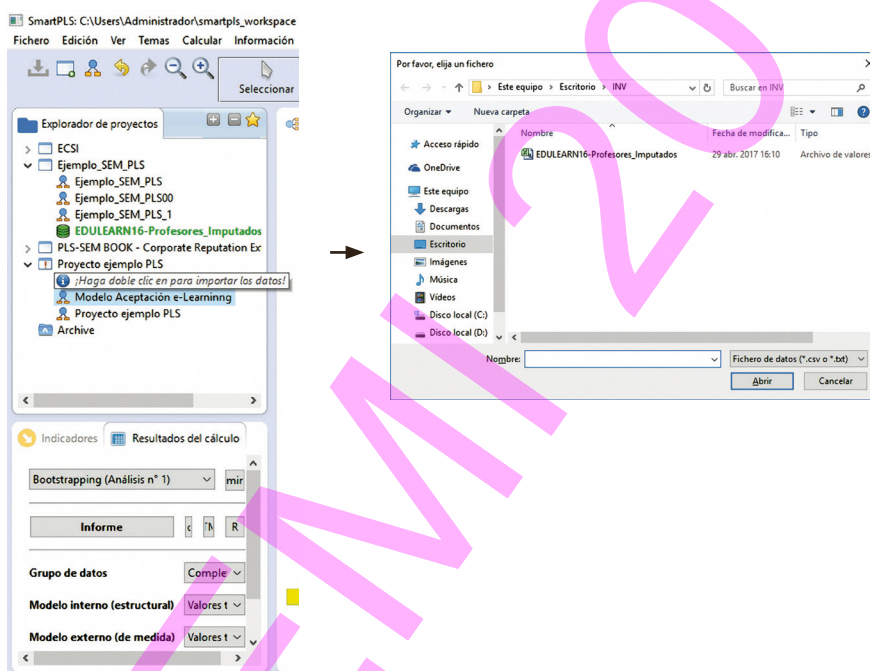


Figura 4 - 6. Carga de datos en SmatPLS

4.9.2.6. Cargar los indicadores

Cargados los datos en el programa SmartPLS, se procede a asignar los indicadores (**seleccionando el grupo de indicadores**) a cada constructo, esto se consigue arrastrando los indicadores con el mouse a cada variable latente según correspondan los datos. La **Figura 4 - 7** muestra al constructo TS con sus respectivos indicadores.

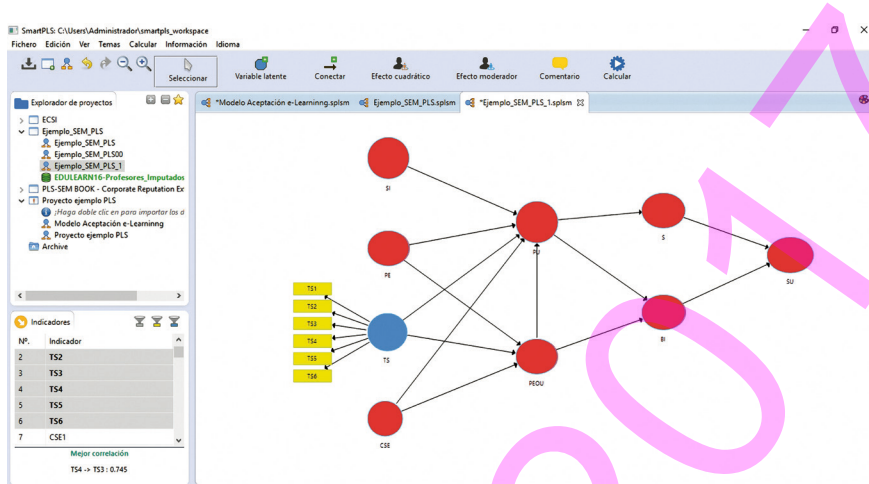


Figura 4 - 7. Asignar indicadores a variables latentes en SmartPLS

Al terminar el proceso de asignar a cada variable latente del modelo su respectivo conjunto de indicadores que permiten evaluar cada constructo, se tiene lo que muestra la **Figura 4 - 8**. Es recomendable como buena practica, codificar los indicadores, de manera que se pueda de forma fácil asociar a su respetivo constructo.

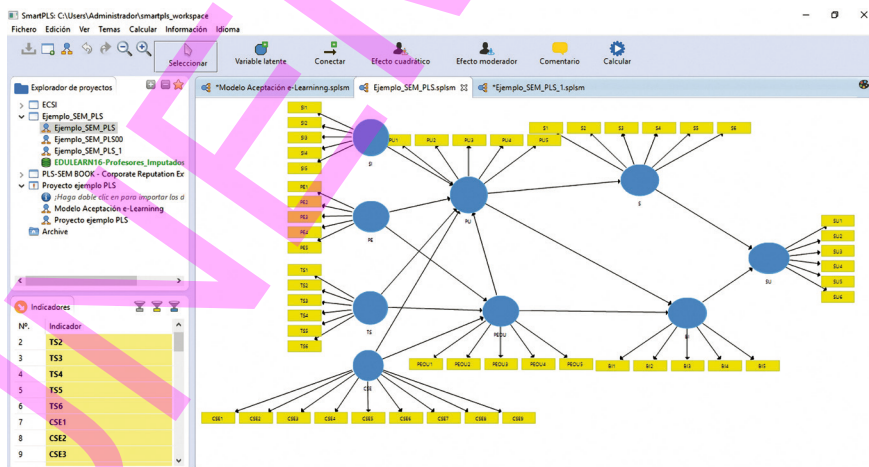


Figura 4 - 8. Modelo completo con sus indicadores

4.9.2.7. Calcular el modelo

Una vez asignados todos los indicadores a las variables latentes, se calcula el modelo. Para realizar el cálculo del modelo se debe dar clic sobre la opción **(Calcular)** luego seleccionar la opción **(Algoritmo PLS)** de la barra de herramientas. A continuación nos aparece la pantalla **(Partial Least Squares Algorithm)** en la cual seleccionamos la opción **(Iniciar cálculos)**. A continuación muestra el cuadro de diálogo para calcular el modelo.

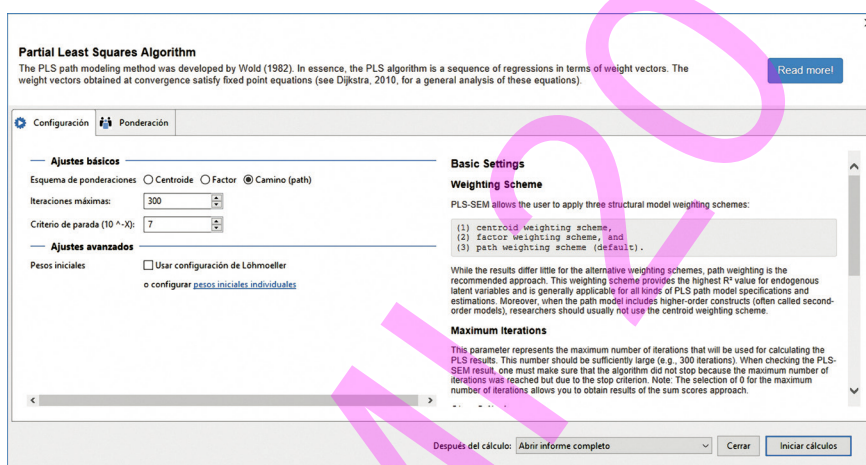


Figura 4 - 9. Calcular modelo en SmartPLS

4.9.2.8. Calcular significancia estadística (Remuestreo)

Luego de obtener el cálculo de los datos, el programa SmartPLS nos permite realizar un remuestreo que consiste en identificar los indicadores de intensidad de relación alta es decir que sean mayor que 0,80 los mismos que una vez identificados se procede a suprimirlos del modelo marcándolos y seleccionándolos, luego realizamos una copia del modelo y enviamos a realizar la operación de (Calcular- Bootstrapping) ver **Figura 4 - 10**.

Una vez seleccionada la opción **(Calcular- Bootstrapping)** se presenta un cuadro de diálogo en el cual en el menú de herramientas seleccionamos **(Configuración)** **(Ajustes básicos- Submuestras se establece el valor de 5000)** luego **(Ajustes avanzados- Nivel de significación se establece el valor de 0,05)** a continuación de que se establecen los valores y se selecciona **(Iniciar Cálculos)** ver **Figura 4 - 11**.

Aparece otro cuadro de diálogo de (**Bootstrapping**) en el cual se muestra como avanza el proceso de remuestreo ver **Figura 4 - 12**. Luego realizado el remuestreo se presenta (**Coefficientes Path.con P-valores de nivel de significación menores de 0,05**); valor aceptable en el proyecto ver **Figura 4 - 13**.

A continuación para obtener el (**Informe Final**) se exportan los datos obtenidos a Excel seleccionando la opción (**Exportar a Excel**) de la barra de herramientas ver **Figura 4 - 14**, luego nos aparece una lista de ítems en el cual se da clic en la opción (**si**) una vez seleccionados se da clic en (**Iniciar exportación**) ver **Figura 4 - 15 y Figura 4 - 16**.

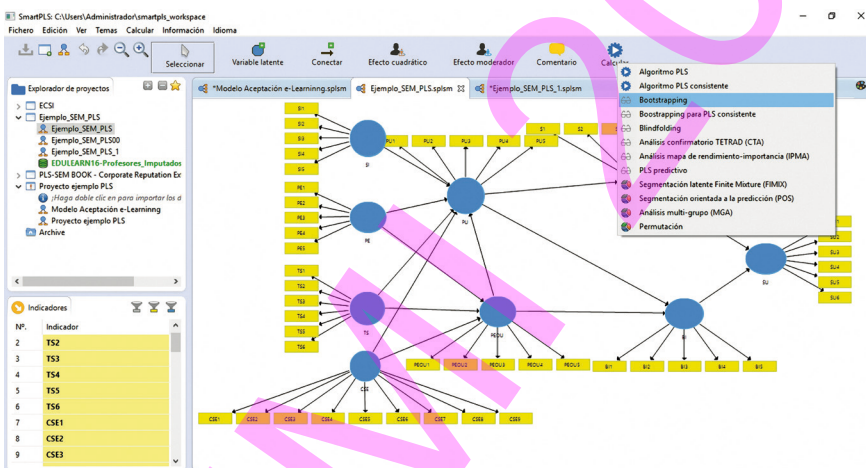
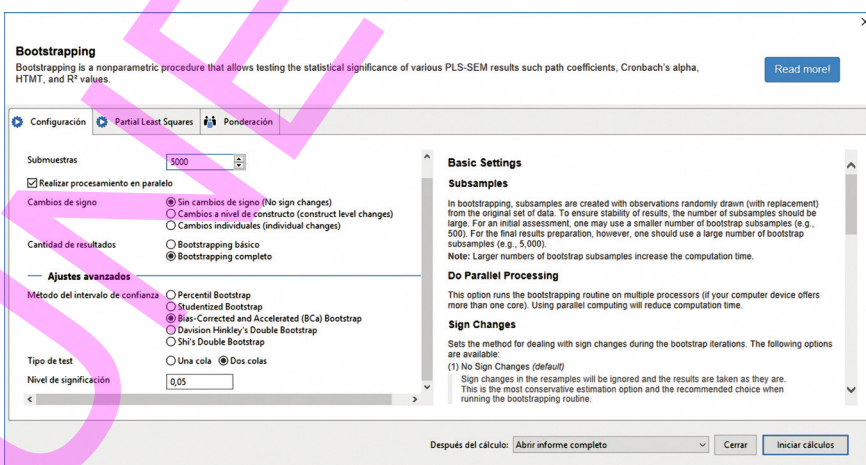


Figura 4 - 10. Calcular bootstrapping



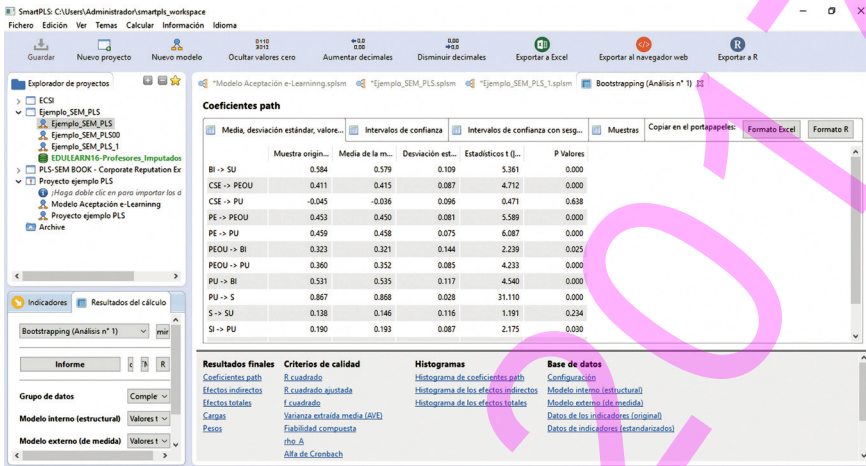


Figura 4 - 11. Ajustes para calculos de bootstrapping

Figura 4 - 12. Proceso de remuestreo

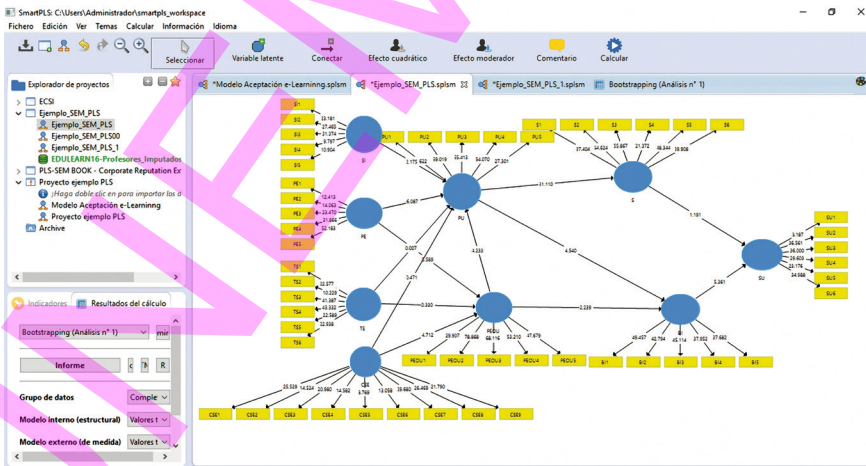


Figura 4 - 13. Coeficientes Path.con P-valores de nivel de significación menores de 0.05

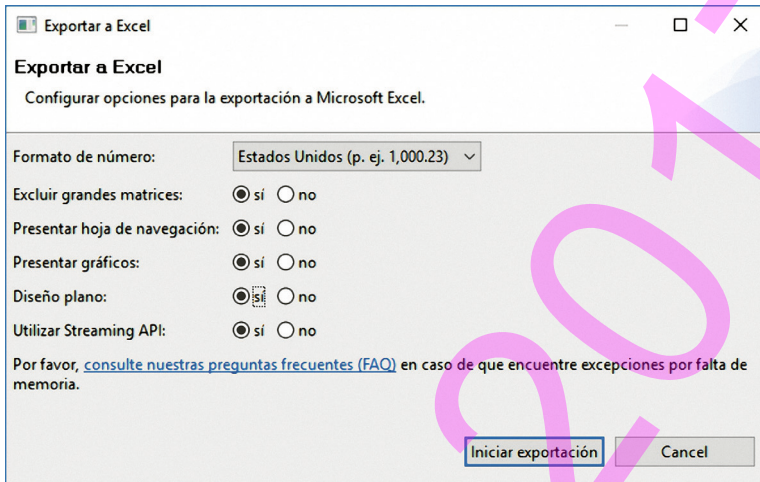


Figura 4 - 14. Proceso de exportación de datos a Excel

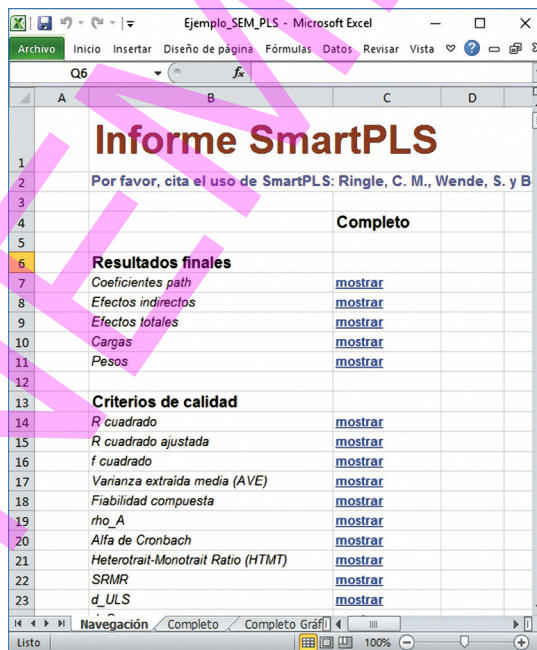


Figura 4 - 15. Informe final de datos exportados a Excel

	Muestra original (O)	Media de la muestra (M)	Desviación estándar (STDEV)	Estadísticos t (O/STDEV)	p Valores
BI → SU	0.584	0.579	0.109	5.361	0.000
CSE → PEOU	0.411	0.415	0.087	4.712	0.000
CSE → PU	-0.046	-0.036	0.096	0.471	0.638
PE → PEOU	0.453	0.450	0.081	5.589	0.000
PE → PU	0.459	0.458	0.075	6.087	0.000
PEOU → BI	0.323	0.321	0.144	2.239	0.025
PEOU → PU	0.360	0.352	0.085	4.233	0.000
PU → BI	0.531	0.535	0.117	4.540	0.000
PU → S	0.867	0.868	0.028	31.110	0.000
S → SU	0.138	0.146	0.116	1.191	0.234
SI → PU	0.190	0.193	0.087	2.175	0.030
TS → PEOU	0.024	0.022	0.074	0.330	0.741
TS → PU	0.001	0.002	0.071	0.007	0.994

Figura 4 - 16. Resultados de coeficientes path

A continuación se exponen las acciones realizada para llegar a los resultados finales, los datos fueron tomados del archivo en Excel que generó el software SmartPLS.

4.9.3. Valoración global del modelo

En esta evaluación preliminar al modelo de profesores, considerando lo definido en la metodología que se siguió, se determina que el modelo se ajusta a los datos. Así lo demuestran los índices que se exponen en la **Tabla 4 - 6**. El SRMR es menor que 0.08, el NFI tiende a 1 y el RMS Theta, a pesar de no ser inferior a 0.12, su valor es aceptable por lo favorable de los otros índices.

Tabla 4 - 6 Índices de ajuste global, modelo profesores

Índice de ajuste	Valor
SRMR	0.066
NFI	0.712
RMS Theta	0.132

4.9.4. Valoración del modelo de medida

4.9.4.1. Fiabilidad individual de los indicadores

Para evaluar la confiabilidad individual de cada indicador, el factor de carga fue observado a partir del análisis factorial confirmatorio (CFA), según lo definido en la metodología hay dos indicadores que no se pueden considerar fiables

CSE2 y SU1, su carga factorial es menor a 0.707, en la **Tabla 4 - 7** podemos ver las cargas factoriales de todos los indicadores.

Tabla 4 - 7 Fiabilidad individual de indicadores

Constructos ítems	Media	Desviación estándar	Carga factorial
<i>Soporte Técnico (TS)</i>			
TS1	3.918	2.065	0.810
TS2	3.629	2.069	0.724
TS3	4.164	2.050	0.880
TS4	4.119	1.926	0.895
TS5	3.652	2.002	0.805
TS6	5.000	1.868	0.771
<i>Autoeficacia computacional (CSE)</i>			
CSE1	5.376	1.587	0.809
CSE2	4.693	1.904	0.681
CSE3	5.156	1.586	0.799
CSE4	5.338	1.578	0.761
CSE5	5.780	1.500	0.732
CSE6	6.122	1.273	0.750
CSE7	5.870	1.286	0.855
CSE8	5.591	1.412	0.848
CSE9	5.511	1.350	0.778
<i>Influencia Social (SI)</i>			
SI1	5.532	1.670	0.837
SI2	5.250	1.653	0.864
SI3	5.610	1.480	0.854
SI4	5.279	1.567	0.739
SI5	5.714	1.396	0.716
<i>Entretenimiento Percibido (PE)</i>			
PE1	4.794	1.854	0.741
PE2	5.370	1.615	0.763
PE3	5.000	1.838	0.847
PE4	5.096	1.677	0.873
PE5	5.464	1.594	0.891

Continua

Tabla 4 - 7 Continuación

Constructos ítems	Media	Desviación estándar	Carga factorial
Satisfacción (S)			
S1	5.584	1.454	0.907
S2	5.542	1.432	0.926
S3	5.648	1.386	0.936
S4	5.444	1.516	0.898
S5	5.447	1.411	0.928
S6	5.475	1.511	0.888
Utilidad Percibida (PU)			
PU1	5.475	1.543	0.897
PU2	5.536	1.362	0.931
PU3	5.507	1.479	0.929
PU4	5.597	1.285	0.920
PU5	5.831	1.265	0.890
Facilidad de Uso Percibido (PEOU)			
PEOU1	5.465	1.442	0.893
PEOU2	5.493	1.432	0.952
PEOU3	5.433	1.459	0.953
PEOU4	5.584	1.514	0.927
PEOU5	5.472	1.507	0.934
Intención hacia el uso (BI)			
BI1	5.943	1.312	0.917
BI2	5.851	1.313	0.929
BI3	5.716	1.304	0.917
BI4	5.816	1.246	0.900
BI5	5.893	1.311	0.929
Uso del Sistema(SU)			
SU1	5.556	1.497	0.677
SU2	5.000	1.481	0.850
SU3	5.293	1.441	0.907
SU4	5.298	1.622	0.866
SU5	4.904	1.693	0.846
SU6	5.417	1.499	0.888

Tabla 4 - 8 Fiabilidad y validez de constructo, modelo profesores

Constructos	α	ρ_A	ρ_c
BI	0.954	0.954	0.964
CSE	0.920	0.927	0.933
PE	0.882	0.893	0.914
PEOU	0.962	0.964	0.971
PU	0.950	0.951	0.962
S	0.960	0.961	0.968
SI	0.863	0.872	0.901
SU	0.916	0.921	0.936
TS	0.900	0.915	0.923

4.9.4.2. Fiabilidad del constructo

La consistencia interna de los indicadores que miden los constructos, se evaluó como estaba definido, se revisó la fiabilidad del constructo por medio del Alfa de Cronbach (α), Fiabilidad Compuesta (ρ_c) y el índice Dijkstra-Henseler's (ρ_A) (ρ_A). Todos los constructos satisfacen el requerimiento de la fiabilidad de constructo ya que los tres índices evaluados (α , ρ_A , ρ_c) obtuvieron valores mayores que 0.8. Ver **Tabla 4 - 8**.

4.9.4.4. Validez Convergente

La validez convergente se evaluó mediante la varianza extraída media (average variance extracted - AVE). Todas las variables latentes alcanzaron la validez convergente, ya que sus medidas AVE superaron el nivel mínimo de 0.50. Esto se puede ver en la **Tabla 4 - 9**.

Tabla 4 - 9 Validez convergente, modelo profesores

Constructos	AVE
BI	0.844
CSE	0.610
PE	0.681
PEOU	0.869
PU	0.835
S	0.835
SI	0.647
SU	0.710
TS	0.667

4.9.4.5. Validez Discriminante

Para evaluar la validez discriminante se utilizó el análisis de cargas cruzadas; el criterio de Fornell y Larcker (1981); y, el ratio heterotrait-monotrait (HTMT).

Por el análisis de las cargas cruzadas, todos los constructos alcanzan la validez discriminante, pues ningún ítem carga más fuertemente sobre otro constructo que sobre aquel constructo que trata de medir; a su vez, cada constructo carga más sobre sus indicadores asignados que sobre otros ítems. Lo expuesto se puede ver en la **Tabla 4 - 10**.

Tabla 4 - 10 Cargas cruzadas, modelo profesores

Ítems	BI	CSE	PE	PEOU	PU	S	SI	SU	TS
BI1	0.917	0.594	0.645	0.612	0.679	0.623	0.506	0.627	0.357
BI2	0.929	0.568	0.684	0.658	0.752	0.752	0.510	0.633	0.363
BI3	0.917	0.624	0.640	0.624	0.671	0.652	0.491	0.685	0.381
BI4	0.900	0.619	0.578	0.604	0.630	0.649	0.511	0.622	0.372
BI5	0.929	0.570	0.604	0.630	0.734	0.730	0.535	0.586	0.383
CSE1	0.474	0.809	0.363	0.457	0.359	0.380	0.499	0.452	0.521
CSE2	0.333	0.681	0.325	0.431	0.326	0.340	0.274	0.383	0.360
CSE3	0.415	0.799	0.388	0.453	0.353	0.378	0.426	0.401	0.420
CSE4	0.409	0.761	0.370	0.455	0.318	0.354	0.283	0.387	0.443
CSE5	0.495	0.732	0.351	0.384	0.328	0.396	0.375	0.338	0.424
CSE6	0.630	0.750	0.397	0.508	0.495	0.495	0.478	0.414	0.435
CSE7	0.610	0.855	0.410	0.557	0.443	0.467	0.449	0.380	0.429
CSE8	0.583	0.848	0.481	0.597	0.507	0.520	0.464	0.496	0.486
CSE9	0.523	0.778	0.447	0.572	0.467	0.440	0.356	0.494	0.413
PE1	0.493	0.366	0.741	0.462	0.456	0.474	0.212	0.449	0.349
PE2	0.551	0.494	0.763	0.544	0.656	0.577	0.402	0.566	0.449
PE3	0.543	0.363	0.847	0.465	0.553	0.536	0.240	0.522	0.302
PE4	0.565	0.416	0.873	0.577	0.717	0.654	0.379	0.551	0.431
PE5	0.666	0.443	0.891	0.592	0.672	0.673	0.338	0.596	0.375
PEOU1	0.591	0.598	0.527	0.893	0.560	0.528	0.294	0.601	0.347
PEOU2	0.643	0.569	0.651	0.952	0.703	0.726	0.368	0.606	0.409
PEOU3	0.637	0.566	0.620	0.953	0.726	0.728	0.409	0.593	0.455
PEOU4	0.642	0.622	0.584	0.927	0.690	0.662	0.376	0.595	0.457
PEOU5	0.662	0.624	0.625	0.934	0.625	0.657	0.353	0.680	0.408

Continua

Tabla 4 - 10 Continuación

Ítems	BI	CSE	PE	PEOU	PU	S	SI	SU	TS
PU1	0.727	0.426	0.729	0.669	0.897	0.817	0.421	0.584	0.389
PU2	0.660	0.483	0.647	0.668	0.931	0.745	0.413	0.538	0.406
PU3	0.682	0.454	0.706	0.635	0.929	0.714	0.423	0.546	0.413
PU4	0.680	0.518	0.698	0.652	0.920	0.749	0.459	0.526	0.478
PU5	0.699	0.511	0.654	0.620	0.890	0.749	0.496	0.490	0.407
S1	0.767	0.581	0.728	0.668	0.776	0.907	0.439	0.535	0.426
S2	0.682	0.499	0.639	0.611	0.775	0.926	0.460	0.512	0.408
S3	0.726	0.498	0.656	0.643	0.755	0.936	0.419	0.490	0.432
S4	0.612	0.453	0.642	0.672	0.713	0.898	0.409	0.522	0.459
S5	0.655	0.474	0.647	0.697	0.752	0.928	0.452	0.483	0.444
S6	0.623	0.478	0.606	0.612	0.763	0.888	0.430	0.476	0.454
SI1	0.400	0.476	0.206	0.317	0.329	0.365	0.837	0.325	0.473
SI2	0.396	0.454	0.287	0.354	0.432	0.411	0.864	0.399	0.446
SI3	0.406	0.450	0.233	0.290	0.333	0.350	0.854	0.317	0.407
SI4	0.383	0.351	0.307	0.252	0.325	0.291	0.739	0.335	0.412
SI5	0.589	0.351	0.469	0.319	0.468	0.447	0.716	0.422	0.284
SU1	0.530	0.427	0.551	0.556	0.590	0.502	0.388	0.677	0.391
SU2	0.583	0.409	0.589	0.550	0.510	0.435	0.343	0.850	0.301
SU3	0.656	0.458	0.622	0.610	0.569	0.510	0.417	0.907	0.393
SU4	0.599	0.499	0.490	0.531	0.460	0.468	0.400	0.866	0.349
SU5	0.491	0.403	0.496	0.509	0.386	0.418	0.344	0.846	0.374
SU6	0.586	0.509	0.551	0.563	0.441	0.440	0.414	0.888	0.389
TS1	0.344	0.447	0.417	0.354	0.382	0.426	0.354	0.320	0.810
TS2	0.216	0.416	0.238	0.190	0.220	0.231	0.319	0.241	0.724
TS3	0.362	0.501	0.428	0.436	0.424	0.411	0.444	0.439	0.880
TS4	0.356	0.473	0.393	0.384	0.396	0.419	0.428	0.348	0.895
TS5	0.302	0.367	0.414	0.321	0.347	0.372	0.383	0.341	0.805
TS6	0.353	0.515	0.350	0.421	0.409	0.422	0.473	0.389	0.771

El criterio de Fornell-Larcker se puede ver la AI revisar el ratio Heterotrait-Monotrait (HTMT), encontramos que todos los valores se encuentran por debajo de 0.90 que es el umbral aceptable según la metodología que se sigue. Lo expuesto se puede apreciar en la **Tabla 4 - 12**. Todos los valores cumplen con la validez discriminante del constructo.

Tabla 4 - 11. Donde los elementos en la diagonal (en negrita) son la raíz cuadrada de la varianza compartida entre el constructo y sus medidas (AVE), los elementos fuera de la diagonal son las correlaciones entre constructos. Todos los constructos alcanzan validez discriminante ya que, la raíz cuadrada del AVE de los constructos es mayor que la correlación que estos tiene con los otros constructos.

Al revisar el ratio Heterotrait-Monotrait (HTMT), encontramos que todos los valores se encuentran por debajo de 0.90 que es el umbral aceptable según la metodología que se sigue. Lo expuesto se puede apreciar en la **Tabla 4 - 12**. Todos los valores cumplen con la validez discriminante del constructo.

Tabla 4 - 11 Validez discriminante, modelo profesores

Constructos	BI	CSE	PE	PEOU	PU	S	SI	SU	TS
BI	0.918								
CSE	0.648	0.781							
PE	0.687	0.509	0.825						
PEOU	0.682	0.638	0.646	0.932					
PU	0.756	0.523	0.753	0.711	0.914				
S	0.743	0.545	0.715	0.712	0.828	0.914			
SI	0.556	0.518	0.390	0.388	0.482	0.476	0.804		
SU	0.687	0.538	0.656	0.659	0.588	0.551	0.458	0.842	
TS	0.404	0.559	0.467	0.447	0.458	0.478	0.498	0.435	0.817

Tabla 4 - 12 Ratio HTMT, modelo profesores

Constructo	BI	CSE	PE	PEOU	PU	S	SI	SU	TS
BI									
CSE	0.680								
PE	0.745	0.555							
PEOU	0.711	0.669	0.695						
PU	0.792	0.548	0.809	0.741					
S	0.774	0.570	0.768	0.737	0.864				
SI	0.597	0.576	0.418	0.417	0.519	0.509			
SU	0.732	0.580	0.724	0.703	0.628	0.587	0.503		
TS	0.426	0.611	0.510	0.460	0.481	0.501	0.562	0.468	

4.9.5. Valoración del modelo estructural

4.9.5.1. Valoración de la colinealidad

Para evaluar posibles problemas de colinealidad en el modelo estructural, se revisaron los valores FIV internos (estructural). Luego de esta revisión se concluye que no hay problemas de colinealidad, pues todos los FIV resultaron menores a 5. Tal afirmación se puede apreciar en la **Tabla 4 - 13**.

Tabla 4 - 13 Valores FIV del modelo estructural, modelo profesores

Constructo	BI	CSE	PE	PEOU	PU	S	SI	SU	TS
BI								2.229	
CSE				1.642	2.161				
PE				1.443	1.867				
PEOU	2.023				2.209				
PU	2.023					1			
S								2.229	
SI					1.512				
SU									
TS				1.556	1.671				

4.9.5.2. Evaluación de los coeficientes path

La evaluación del signo algebraico, magnitud y significación estadística de los coeficientes path se puede ver en la **Tabla 4 - 14**. Con este propósito se ejecutó bootstrapping para n = 5000 submuestras: * p < .05; ** p < .01; ***p < .001 (basado en una distribución t (4999) de Student de una cola). t (0.05; 4999) = 1.645 ; t (0.01; 4999) = 2.327; t (0.001; 4999) = 3.092. En la tabla se muestra si el valor empírico de t es mayor que el valor crítico de t para un nivel de significación seleccionado.

Tabla 4 - 14 Coeficientes path, modelo profesores

Constructos	Muestra Original(O)	Media de la Muestra (M)	Desviación estándar (STDEV)	T Valor	P Valor	Sig.
BI → SU	0.620	0.618	0.093	6.691	0.000	***
CSE → PEOU	0.410	0.417	0.079	5.171	0.000	***
CSE → PU	-0.047	-0.041	0.087	0.543	0.587	NS
PE → PEOU	0.430	0.428	0.074	5.801	0.000	***
PE → PU	0.464	0.462	0.073	6.35	0.000	***
PEOU → BI	0.292	0.296	0.126	2.318	0.020	**
PEOU → PU	0.365	0.359	0.072	5.093	0.000	***
PU → BI	0.548	0.548	0.101	5.456	0.000	***
PU → S	0.828	0.829	0.036	23.199	0.000	***
S → SU	0.091	0.097	0.096	0.946	0.344	NS
SI → PU	0.178	0.181	0.079	2.259	0.024	**
TS → PEOU	0.017	0.014	0.07	0.247	0.805	NS
TS → PU	0.016	0.016	0.063	0.26	0.795	NS

Se puede apreciar que la relación CSE → PU, resultó con signo contrario al planteado; de igual manera se puede notar que resultaron no significativas estadísticamente las relaciones: CSE → PU, S → SU, TS → PEOU y TS → PU.

4.9.5.3. Valoración del coeficiente de determinación

La evaluación del modelo estructural también se realizó mediante el coeficiente de determinación, R cuadrado (R^2) de las variables manifiestas, las cuales se muestran en la **Tabla 4 - 15**.

Tabla 4 - 15 R^2 Variables latentes endógenas, modelo profesores

Constructos	Muestra Original(O)	Media de la Muestra (M)	Desviación estándar (STDEV)	T Valor	P Valor
BI	0.613	0.628	0.069	8.916	0.000
PEOU	0.547	0.562	0.071	7.725	0.000
PU	0.678	0.693	0.049	13.729	0.000
S	0.685	0.688	0.059	11.680	0.000
SU	0.476	0.488	0.074	6.400	0.000

El modelo tiene un poder explicativo moderado, ya que los valores de R^2 de las variables endógenas llegan y superan el 0.5 según lo sugieren Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt (2014).

4.9.5.4. Valoración de los tamaños de los efectos

Para evaluar el grado con el que un constructo exógeno contribuye a explicar un determinado constructo endógeno en términos de R^2 , se revisó el f^2 .

Los valores de f^2 se pueden observar en la **Tabla 4 - 16**. En dicha tabla se aprecia que PE sobre PU, PU sobre BI y PU sobre S; tienen un efecto alto ya que su f^2 es ≥ 0.35 . Continuando con el análisis; BI sobre SU, CSE sobre PEOU, PE sobre PEOU y PEOU sobre PU; tienen un efecto moderado, ya que $0.15 \leq f^2 < 0.35$. Por otra parte PEOU sobre BI y SI sobre PU tiene un efecto pequeño, debido a que $0.02 \leq f^2 < 0.15$. Mientras que CSE sobre PU, TS sobre PEOU, TS sobre PU y S sobre SU; no producen un efecto significativo.

Tabla 4 - 16 Valoración de f^2 , modelo profesores

Constructos	BI	CSE	PE	PEOU	PU	S	SI	SU	TS
BI								0.328	
CSE				0.226	0.003				
PE				0.283	0.358				
PEOU	0.109				0.187				
PU	0.384					2.173			
S								0.007	
SI					0.065				
SU									
TS				0.000	0.001				

4.9.5.5. Valoración de la relevancia predictiva

Para valorar la relevancia predictiva se realizó la prueba de Stone-Geisser (Q^2), para esta investigación se optó evaluar el calculado mediante validación cruzada de la redundancia del constructo (Constructo Crossvalidated Redundancy). El resultado evidencia validez predictiva del modelo ya que todos sus constructos exógenos evidencian un Q^2 mayor que cero. Tal como se observa en la **Tabla 4 - 17**.

Tabla 4 - 17 Índice Q2, modelo de profesores

Constructos	Q ²
BI	0.509
PEOU	0.470
PU	0.555
S	0.563
SU	0.330

A continuación se muestra cada uno de los resultados del modelo completo, los valores que tienen asteriscos son los que muestran los impactos estadísticamente significativos, mientras que los NS muestran impactos no significativos estadísticamente. Ver la **Figura 4 - 17**.

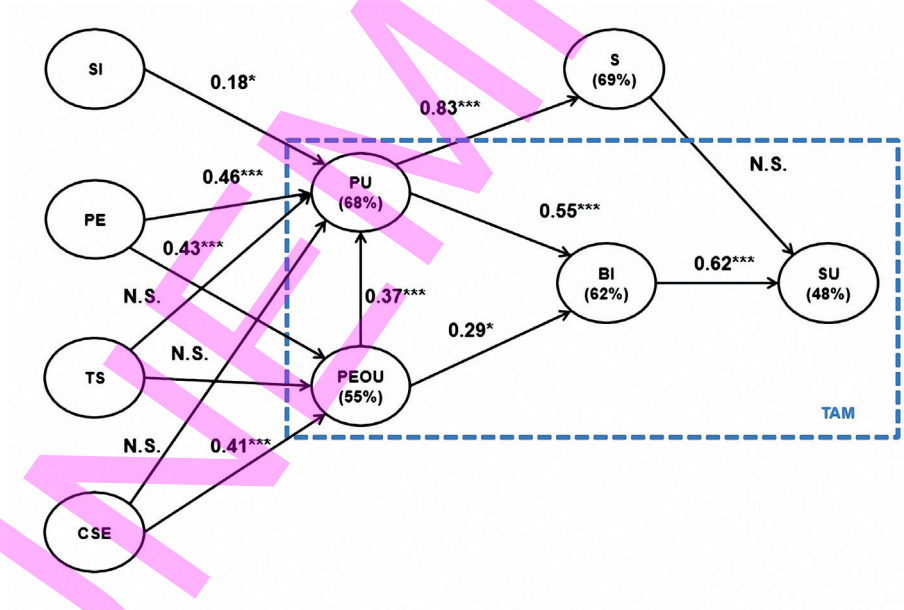


Figura 4 - 17. Modelo estructural

Notas: Varianza explicada (R²). * Coeficiente de ruta significativa al nivel de 0.05; ** Al nivel 0.01; *** al nivel 0.001

4.10. Ejemplo de SEM enfoque basado en covarianzas

4.10.1. Introducción

Para ilustrar el uso del software R para manejo de SEM se utiliza un estudio que tenía como objetivo adaptar un modelo de investigación que describe el comportamiento de uso y el uso previsto de los sistemas de aprendizaje electrónico en los estudiantes Ecuador, basado en el modelo de aceptación de Tecnología (TAM) desarrollado por Fred Davis. El modelo de investigación se presenta en la **Figura 4 - 18**.

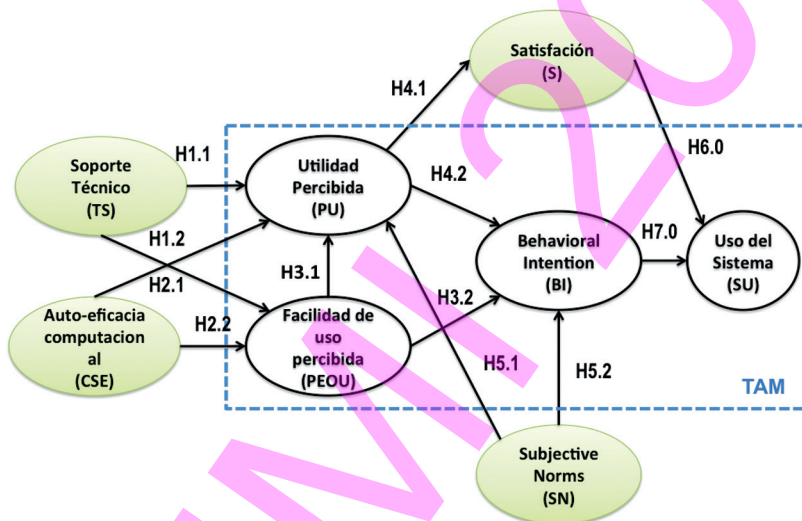


Figura 4 - 18. Modelo de investigación

Nota: Los constructos en fondo verde son propuesto por el estudio

En la investigación, un cuestionario auto reporte fue diseñado para examinar la aceptación del estudiante y la actitud hacia el sistema de e-learning. Los datos de 263 estudiantes de pregrado y posgrado se recogieron en una universidad de Ecuador. El modelo se compone de ocho construcciones, a saber, soporte técnico, autoeficacia equipo, normas subjetivas, de satisfacción, utilidad percibida, facilidad de uso percibida, la intención de utilizar el sistema y utilizar el sistema.

El modelo se estimó utilizando modelado estructural ecuación (SEM), para lo cual se utilizó el software R. Se encontró que el apoyo y el ordenador autoeficacia técnica sirvió como los dos anclajes de la percepción significativa de constructos fundamentales en TAM. Satisfacción convirtió en el único factor determinante en la predicción de uso, mientras que las normas subjetivas

fueron significativas a la predicción de la intención de utilizar el sistema de e-learning. En general, las normas subjetivas, autoeficacia equipo y facilidad de uso percibida fueron capaces de explicar la mayor variación observada en la intención de los estudiantes a utilizar el sistema de e-learning.

4.10.2. Pasos

Los datos de la investigación citada fueron colocados en un archivo de texto delimitado por comas (formato CSV). El archivo tiene el nombre Ejercicio42_datos_docentes.csv y puede ser descargado desde el siguiente enlace <https://www.dropbox.com/sh/448ijhlll1govqj/AADYa8UTmaJFDNbpsSfvM7pZa?dl=0>.

Luego de instalar R y R-Studio, para este paso puede recurrir al **Anexo 1- 2**; es necesario instalar el paquete lavaan para el análisis factorial confirmatorio y el análisis de ecuaciones estructurales.

A continuación se crea un proyecto y un script en R-Studio, tal como lo describe el apartado **1.6.1**. Después se procede a insertar en la ventana de R-Studio las sentencias que se detalla en los apartados siguientes.

4.10.2.1. Establecer directorio de trabajo, carga de datos y paquete

```
# (1) Establecer por defecto el directorio de trabajo donde
está el archivo de texto con los datos
setwd("/Users/imac/Dropbox/RIRA/PhD/Publicaciones/Libros/
UNEMI2017/Datos")

# (2) Carga los datos de todos los profesores
Docentes <- read.csv("Ejercicio11_datos_docentes.csv",
header = TRUE, sep=";", dec=",", quote="\")

# (3) Instala y carga el paquete para análisis SEM
install.packages ("lavaan")
library(lavaan)
```

4.10.2.2. Crea el modelo variable

```
# (4) Crea el modelo SEM para evaluar aceptación de e-learning
en estudiante
modelo <- ‘
# Modelo de medidas
```



```

SU   =~ SU1 + SU2 + SU3
BI   =~ BI1 + BI2 + BI3
SN   =~ SN1 + SN2 + SN3 + SN4
S    =~ S1 + S2 + S3 + S4
PU   =~ PU1 + PU2 + PU3 + PU4 + PU5
PEOU =~ PEOU1 + PEOU2 + PEOU3 + PEOU4 + PEOU5
CSE  =~ CSE1 + CSE2 + CSE3 + CSE4 + CSE5
TS   =~ TS1 + TS2 + TS3 + TS4
# regresiones
SU   ~ S + BI
BI   ~ SN + PEOU + PU
S    ~ PU
PEOU ~ CSE + TS
PU   ~ PEOU + CSE + TS
# covarianzas residuales (NO HAY)
'
```

4.10.2.3. Evaluar el modelo de media

```

# (5) Se evalúa el modelo de medidas
fit <- cfa(modelo, data = Datos)
summary(fit, fit.measures=TRUE)
```

Con la ejecución de las sentencias de este apartado, el software R arroja la siguiente salida:

```

> fit <- cfa(modelo, data = Datos)
> summary(fit, fit.measures=TRUE)
lavaan (0.5-23.1097) converged normally after 79 iterations

Number of observations                263

Estimator                             ML
Minimum Function Test Statistic       1259.213
Degrees of freedom                     481
P-value (Chi-square)                  0.000

Model test baseline model:

Minimum Function Test Statistic       12613.579
Degrees of freedom                     528
P-value                                0.000
```

User model versus baseline model:

Comparative Fit Index (CFI)	0.936
Tucker-Lewis Index (TLI)	0.929

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-10661.870
Loglikelihood unrestricted model (H1)	-10032.264
Number of free parameters	80
Akaike (AIC)	21483.741
Bayesian (BIC)	21769.513
Sample-size adjusted Bayesian (BIC)	21515.875

Root Mean Square Error of Approximation:

RMSEA	0.078
90 Percent Confidence Interval	0.073 0.084
P-value RMSEA \leq 0.05	0.000

Standardized Root Mean Square Residual:

SRMR	0.068
------	-------

Parameter Estimates:

Information	Expected
Standard Errors	Standard

Latent Variables:

	Estimate	Std.Err	z-value	P(> z)
SU \approx				
SU1	1.000			
SU2	1.145	0.055	20.795	0.000
SU3	1.183	0.054	21.818	0.000
BI \approx				
BI1	1.000			
BI2	0.982	0.021	46.165	0.000
BI3	0.884	0.034	26.159	0.000
SN \approx				
SN1	1.000			
SN2	1.065	0.067	15.865	0.000
SN3	1.136	0.060	18.839	0.000
SN4	0.964	0.067	14.472	0.000

S =~				
S1	1.000			
S2	0.981	0.030	32.384	0.000
S3	0.989	0.031	32.249	0.000
S4	0.959	0.029	32.910	0.000
PU =~				
PU1	1.000			
PU2	1.014	0.024	41.610	0.000
PU3	1.016	0.028	36.883	0.000
PU4	0.982	0.027	36.783	0.000
PU5	0.975	0.030	32.885	0.000
PEOU =~				
PEOU1	1.000			
PEOU2	1.068	0.042	25.504	0.000
PEOU3	1.081	0.039	27.950	0.000
PEOU4	0.989	0.045	21.983	0.000
PEOU5	1.049	0.039	26.816	0.000
CSE =~				
CSE1	1.000			
CSE2	1.031	0.057	17.981	0.000
CSE3	1.009	0.053	18.927	0.000
CSE4	0.961	0.056	17.081	0.000
CSE5	0.882	0.061	14.575	0.000
TS =~				
TS1	1.000			
TS2	1.236	0.070	17.597	0.000
TS3	1.202	0.077	15.597	0.000
TS4	1.194	0.071	16.908	0.000
Regressions:				
	Estimate	Std.Err	z-value	P(> z)
SU ~				
S	0.218	0.060	3.636	0.000
BI	0.512	0.064	8.009	0.000
BI ~				
SN	0.324	0.057	5.630	0.000
PEOU	0.454	0.089	5.116	0.000
PU	0.277	0.078	3.543	0.000
S ~				
PU	0.954	0.041	23.090	0.000
PEOU ~				
CSE	0.826	0.068	12.190	0.000
TS	0.063	0.054	1.171	0.242
PU ~				
PEOU	0.713	0.076	9.377	0.000
CSE	0.132	0.083	1.595	0.111

TS	0.191	0.048	3.949	0.000
Covariances:				
	Estimate	Std.Err	z-value	P(> z)
SN ~				
CSE	1.232	0.149	8.251	0.000
TS	1.035	0.146	7.076	0.000
CSE ~				
TS	1.245	0.159	7.819	0.000
Variances:				
	Estimate	Std.Err	z-value	P(> z)
.SU1	0.767	0.077	10.009	0.000
.SU2	0.399	0.056	7.127	0.000
.SU3	0.237	0.050	4.687	0.000
.BI1	0.147	0.023	6.449	0.000
.BI2	0.088	0.020	4.480	0.000
.BI3	0.531	0.050	10.614	0.000
.SN1	0.659	0.073	8.992	0.000
.SN2	0.942	0.099	9.529	0.000
.SN3	0.431	0.065	6.669	0.000
.SN4	1.068	0.106	10.084	0.000
.S1	0.391	0.039	10.137	0.000
.S2	0.165	0.020	8.362	0.000
.S3	0.173	0.020	8.448	0.000
.S4	0.141	0.018	7.987	0.000
.PU1	0.204	0.021	9.678	0.000
.PU2	0.108	0.014	7.984	0.000
.PU3	0.198	0.021	9.554	0.000
.PU4	0.187	0.020	9.576	0.000
.PU5	0.280	0.027	10.233	0.000
.PEOU1	0.385	0.039	9.926	0.000
.PEOU2	0.356	0.037	9.558	0.000
.PEOU3	0.225	0.027	8.357	0.000
.PEOU4	0.545	0.052	10.407	0.000
.PEOU5	0.268	0.030	9.016	0.000
.CSE1	0.675	0.070	9.648	0.000
.CSE2	0.657	0.069	9.476	0.000
.CSE3	0.494	0.055	8.925	0.000
.CSE4	0.704	0.071	9.860	0.000
.CSE5	1.023	0.097	10.541	0.000
.TS1	0.900	0.096	9.347	0.000
.TS2	0.780	0.101	7.695	0.000
.TS3	1.368	0.145	9.455	0.000
.TS4	0.926	0.109	8.495	0.000
.SU	0.841	0.105	8.043	0.000

.BI	0.576	0.059	9.711	0.000
SN	1.612	0.195	8.270	0.000
.S	0.404	0.046	8.698	0.000
.PU	0.425	0.046	9.251	0.000
.PEOU	0.475	0.060	7.891	0.000
CSE	1.732	0.205	8.449	0.000
TS	1.928	0.240	8.021	0.000

4.10.2.4. Evaluar el modelo de estructural

```
# (6) Se evalua el modelo de estructural
fit1 <- sem(modelo, data=Datos)
summary(fit1, standardized = TRUE)
```

La salida que muestra el programa informático R, luego de ejecutar las sentencias para evaluar el modelo estructural es la siguiente:

```
> fit1 <- sem(modelo, data=Datos)
> summary(fit1, standardized = TRUE)
lavaan (0.5-23.1097) converged normally after 79 iterations

Number of observations                    263

Estimator                                ML
Minimum Function Test Statistic          1259.213
Degrees of freedom                        481
P-value (Chi-square)                     0.000

Parameter Estimates:

Information                                Expected
Standard Errors                            Standard

Latent Variables:
Estimate Std.Err z-value P(>|z|) Std.lv Std.all
SU =~
SU1      1.000
SU2      1.145  0.055  20.795  0.000  1.578  0.928
```

SU3	1.183	0.054	21.818	0.000	1.630	0.958
BI =~						
BI1	1.000				1.490	0.968
BI2	0.982	0.021	46.165	0.000	1.464	0.980
BI3	0.884	0.034	26.159	0.000	1.318	0.875
SN =~						
SN1	1.000				1.270	0.842
SN2	1.065	0.067	15.865	0.000	1.352	0.812
SN3	1.136	0.060	18.839	0.000	1.442	0.910
SN4	0.964	0.067	14.472	0.000	1.224	0.764
S =~						
S1	1.000				1.520	0.925
S2	0.981	0.030	32.384	0.000	1.492	0.965
S3	0.989	0.031	32.249	0.000	1.504	0.964
S4	0.959	0.029	32.910	0.000	1.458	0.968
PU =~						
PU1	1.000				1.447	0.954
PU2	1.014	0.024	41.610	0.000	1.467	0.976
PU3	1.016	0.028	36.883	0.000	1.471	0.957
PU4	0.982	0.027	36.783	0.000	1.421	0.957
PU5	0.975	0.030	32.885	0.000	1.410	0.936
PEOU =~						
PEOU1	1.000				1.339	0.907
PEOU2	1.068	0.042	25.504	0.000	1.430	0.923
PEOU3	1.081	0.039	27.950	0.000	1.447	0.950
PEOU4	0.989	0.045	21.983	0.000	1.325	0.874
PEOU5	1.049	0.039	26.816	0.000	1.404	0.938
CSE =~						
CSE1	1.000				1.316	0.848
CSE2	1.031	0.057	17.981	0.000	1.357	0.859
CSE3	1.009	0.053	18.927	0.000	1.328	0.884
CSE4	0.961	0.056	17.081	0.000	1.265	0.833
CSE5	0.882	0.061	14.575	0.000	1.161	0.754
TS =~						
TS1	1.000				1.389	0.826
TS2	1.236	0.070	17.597	0.000	1.716	0.889
TS3	1.202	0.077	15.597	0.000	1.668	0.819
TS4	1.194	0.071	16.908	0.000	1.658	0.865
Regressions:						
	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all

SU ~						
S	0.218	0.060	3.636	0.000	0.241	0.241
BI	0.512	0.064	8.009	0.000	0.554	0.554
BI ~						
SN	0.324	0.057	5.630	0.000	0.276	0.276
PEOU	0.454	0.089	5.116	0.000	0.407	0.407
PU	0.277	0.078	3.543	0.000	0.269	0.269
S ~						
PU	0.954	0.041	23.090	0.000	0.908	0.908
PEOU ~						
CSE	0.826	0.068	12.190	0.000	0.812	0.812
TS	0.063	0.054	1.171	0.242	0.065	0.065
PU ~						
PEOU	0.713	0.076	9.377	0.000	0.660	0.660
CSE	0.132	0.083	1.595	0.111	0.120	0.120
TS	0.191	0.048	3.949	0.000	0.183	0.183
Covariances:						
	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
SN ~~						
CSE	1.232	0.149	8.251	0.000	0.738	0.738
TS	1.035	0.146	7.076	0.000	0.587	0.587
CSE ~~						
TS	1.245	0.159	7.819	0.000	0.682	0.682
Variances:						
	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.SU1	0.767	0.077	10.009	0.000	0.767	0.288
.SU2	0.399	0.056	7.127	0.000	0.399	0.138
.SU3	0.237	0.050	4.687	0.000	0.237	0.082
.BI1	0.147	0.023	6.449	0.000	0.147	0.062
.BI2	0.088	0.020	4.480	0.000	0.088	0.039
.BI3	0.531	0.050	10.614	0.000	0.531	0.234
.SN1	0.659	0.073	8.992	0.000	0.659	0.290
.SN2	0.942	0.099	9.529	0.000	0.942	0.340
.SN3	0.431	0.065	6.669	0.000	0.431	0.172
.SN4	1.068	0.106	10.084	0.000	1.068	0.416
.S1	0.391	0.039	10.137	0.000	0.391	0.145
.S2	0.165	0.020	8.362	0.000	0.165	0.069
.S3	0.173	0.020	8.448	0.000	0.173	0.071
.S4	0.141	0.018	7.987	0.000	0.141	0.062

.PU1	0.204	0.021	9.678	0.000	0.204	0.089
.PU2	0.108	0.014	7.984	0.000	0.108	0.048
.PU3	0.198	0.021	9.554	0.000	0.198	0.084
.PU4	0.187	0.020	9.576	0.000	0.187	0.085
.PU5	0.280	0.027	10.233	0.000	0.280	0.123
.PEOU1	0.385	0.039	9.926	0.000	0.385	0.177
.PEOU2	0.356	0.037	9.558	0.000	0.356	0.148
.PEOU3	0.225	0.027	8.357	0.000	0.225	0.097
.PEOU4	0.545	0.052	10.407	0.000	0.545	0.237
.PEOU5	0.268	0.030	9.016	0.000	0.268	0.120
.CSE1	0.675	0.070	9.648	0.000	0.675	0.280
.CSE2	0.657	0.069	9.476	0.000	0.657	0.263
.CSE3	0.494	0.055	8.925	0.000	0.494	0.219
.CSE4	0.704	0.071	9.860	0.000	0.704	0.306
.CSE5	1.023	0.097	10.541	0.000	1.023	0.432
.TS1	0.900	0.096	9.347	0.000	0.900	0.318
.TS2	0.780	0.101	7.695	0.000	0.780	0.209
.TS3	1.368	0.145	9.455	0.000	1.368	0.329
.TS4	0.926	0.109	8.495	0.000	0.926	0.252
.SU	0.841	0.105	8.043	0.000	0.443	0.443
.BI	0.576	0.059	9.711	0.000	0.259	0.259
.SN	1.612	0.195	8.270	0.000	1.000	1.000
.S	0.404	0.046	8.698	0.000	0.175	0.175
.PU	0.425	0.046	9.251	0.000	0.203	0.203
.PEOU	0.475	0.060	7.891	0.000	0.265	0.265
CSE	1.732	0.205	8.449	0.000	1.000	1.000
TS	1.928	0.240	8.021	0.000	1.000	1.000

4.10.3. Conclusiones

En este estudio, la fiabilidad y la validez de escala se evaluaron mediante análisis factorial confirmatorio (CFA). Usado para determinar si un conjunto de variables observadas explica un factor latente. El estudio se realizó con el paquete lavaan del software R, y el procedimiento de estimación SEM utilizado fue la estimación de máxima verosimilitud. Se utilizó un conjunto de índices de ajuste para examinar el modelo estructural. Se llevaron a cabo estadísticos de bondad de ajuste para demostrar la aceptabilidad del modelo propuesto. Comparativas de los índices de ajuste con sus correspondientes valores recomendados, proporcionaron evidencia de un ajuste aceptable del modelo ($\chi^2 / df = 2.618$, CFI = 0.94, TLI = 0.93, RMSEA = 0.078 y SRMR = 0.068).

4.11. Resumen del Capítulo

En este capítulo se abordó los modelos de ecuaciones estructurales, que es una técnica que combina la regresión múltiple y el análisis factorial. Con ella se evalúan complejas interrelaciones de dependencia, incluso, puede agregar los efectos causados por el error de medida que habría sobre los coeficientes estructurales simultáneamente. Conforman los modelos de ecuaciones estructurales: el modelo estructural o modelo interno, que consiste en las relaciones entre los diferentes constructos que integran el modelo; y, el modelo de medida, que es la relación entre los diferentes indicadores y los constructos que se evalúan en el modelo.

Se identifican dos principales enfoques en los modelos de ecuaciones estructurales, el análisis basado en covarianzas (SEM-MBC) y el que trabaja con las varianzas (SEM-PLS), este último también conocido como mínimos cuadrados parciales o, simplemente por sus siglas en inglés, PLS.

Para definir el enfoque a utilizar al momento de evaluar un modelo con ecuaciones estructurales, se encuentran: el objetivo de la investigación, la configuración del modelo y las características de los datos. El primer software para tratamiento de modelos de ecuaciones estructurales se llama LISREL, en la actualidad existen varias opciones, tanto de código abierto como de licencias de pago. Para el caso de tratamiento de PLS, existe el software SmartPLS y el plspm, que es un paquete del software de código abierto R.

4.12. Referencias

- Albers, S. (2010). PLS and success factor studies in marketing. (V. Vinzi Esposito, W. . Chin, J. Henseler, & H. (Eds) Wang, Eds.)Springer. Berlin: Handbook of Partial Least Squares.
- Arteaga, R., & Duarte, A. (2010). Motivational factors that influence the acceptance of Moodle using TAM. *Computers in Human Behavior*, 26(6), 1632–1640. <http://doi.org/10.1016/j.chb.2010.06.011>
- Ayodeji, A., & Al-lawati, A. M. (2010). Automation in Construction Using PLS-SEM technique to model construction organizations' willingness to participate in e-bidding. *Automation in Construction*, 19(6), 714–724. <http://doi.org/10.1016/j.autcon.2010.02.016>
- Barclay, D., Higgins, C., & Thompson, R. (1995). The partial least square (PLS) approach to casual modeling: Personal computer adoption and use as an illustration. *Technology Studies (Special Issue of Research Methodology)*.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and

Methodology).

- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/3806354>
- Carmines, E. G., & Zeller, R. A. (1979). Reliability and validity assessment. (J. L. Sullivan & R. G. Niemi, Eds.) *Quantitative Applications in the Social Sciences* (Vol. 17). Beverly Hill - USA: McCune, Sara. <http://doi.org/10.1037/018269>
- Cepeda Carrión, G., & Roldan Salgueiro, J. L. (2004). Aplicando en la práctica la técnica PLS en administración de empresas. In Congreso de la ACEDE (pp. 1–30). Murcia-España: ciberconta.unizar.es. Retrieved from <http://ciberconta.unizar.es/doctorado/PLSGabrielCepeda.pdf>
- Chin, W. W. (1998a). Issues and Opinion on Structural Equation Modeling. *MIS Quarterly*, 22(1), vii–xvi. <http://doi.org/Editorial>
- Chin, W. W. (1998b). The Partial Least Squares Approach to Structural Modeling. In *Modern Methods for Business Research* (pp. 295–336).
- Chin, W. W., Marcolin, B. L., & Newted, P. R. (2003). A Partial least Squares Latent Variable Modeling Approach For Measuring Interaction Effects: Results From a Monte Carlo Simulation Study and Voice Mail Emotion/Adoption Study. In J. I. DeGross, S. Jarvenpaa, & A. Srinivasan (Eds.), *Proceedings of the Seventeenth International Conference on Information Systems* (pp. 21–41). Cleveland- USA. <http://doi.org/10.1287/isre.14.2.189.16018>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Second). Lawrence Erlbaum.
- Coltman, T., Devinney, T. M., Midgley, D. F., & Venaiik, S. (2008). Formative versus reflective measurement models : Two applications of formative measurement. *Journal of Business Research*, 61(12), 1250–1262. <http://doi.org/10.1016/j.jbusres.2008.01.013>
- Components, O., & Variables, L. (2014). PLS-SEM: Looking Back and Moving Forward. *Long Range Planning*, 47, 132–137. <http://doi.org/10.1016/j.lrp.2014.02.008>
- Dijkstra, T. K., & Henseler, J. (2015a). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics and Data Analysis*, 81, 10–23. <http://doi.org/10.1016/j.csda.2014.07.008>
- Dijkstra, T. K., & Henseler, J. (2015b). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics and Data Analysis*, 81, 10–23. <http://doi.org/10.1016/j.csda.2014.07.008>
- Domínguez, C., & Jesús, A. (2006). SEM vs . PLS : UN ENFOQUE BASADO EN LA PRACTICA. In IV Congreso de Metodología de Encuestas (pp. 57–66). Pamplona- España.

- Falk, R. F., & Miller, N. B. (1992). A Primer for Soft Modeling.pdf.
- Findik, D., & Kunçay, C. O. Ş. (2009). A model for instructors adoption of learning management systems : empirical validation in higher education context. Middle East Technical University.
- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39–50. <http://doi.org/10.2307/3151312>
- García, M. (2011). Análisis causal con ecuaciones estructurales de la satisfacción ciudadana con los servicios municipales. Universidad Santiago de Compostela.
- Garson, G. D. (2016). *Partial Least Squares: Regression & Structural Equation Models* (2016th ed.). Asheboro - USA: Statistical Associates Publishing.
- Gefen, D., & Straub, D. (2005). A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. *Communications of the Association for Information Systems*, 16(5), 91–109. <http://doi.org/Article>
- Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural Equation Modeling and Regression : Guidelines for Research Practice. *Structural Equation Modeling*, 4(August), 7. <http://doi.org/10.1.1.25.781>
- Gold, A. H., Malthora, A., & Segars, A. H. (2001). Knowledge Management: An Organizational Capabilities Perspective. *Journal of Management Information Systems*, 18(1), 185–214. <http://doi.org/10.1002/ceat.201000522>
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). Partial least squares structural equation modeling (PLS-SEM). *Sage Publisher*, 26(2), 106–121. <http://doi.org/10.1108/EBR-10-2013-0128>
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. *The Journal of Marketing Theory and Practice*, 19(2), 139–152. <http://doi.org/10.2753/MTP1069-6679190202>
- Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS Path Modeling in New Technology Research : Updated Guidelines. *Industrial Management & Data Systems*, 116(1), 2–20. <http://doi.org/10.1108/IMDS-09-2015-0382>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review*, 33(3), 405–431. <http://doi.org/10.1108/IMR-09-2014-0304>
- Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, 20(2009), 277–319. [http://doi.org/10.1016/0167-8116\(92\)90003-4](http://doi.org/10.1016/0167-8116(92)90003-4)
- Höök, K., & Löwgren, J. (2012). Strong concepts. *ACM Transactions on Computer-Human Interaction*, 19(3), 1–18. <http://doi.org/10.1145/2094441>

org/10.1145/2362364.2362371

- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424–453.
- Kaufmann, L., & Gaeckler, J. (2015). A structured review of partial least squares in supply chain management research. *Journal of Purchasing and Supply Management*, 21(4), 259–272. <http://doi.org/10.1016/j.pursup.2015.04.005>
- Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling. Analysis* (Vol. 77). New York- USA. <http://doi.org/10.1038/156278a0>
- Lay, J.-G., Chen, Y.-W., & Chi, Y.-L. (2013). GIS Adoption Among Senior High School Geography Teachers in Taiwan. *Journal of Geography*, 112(3), 120–130. <http://doi.org/10.1080/00221341.2012.682226>
- Lévy Mangin, J. pierre, & Varela Mallou, J. (2006). *Modelización con Estructuras de Covarianzas en Ciencias Sociales*. (C. Seco, Ed.) (Gesbiblo S). España. Retrieved from <https://books.google.com.ec/books?id=WEfC1TG VJBgC&pg=PA190&lpg=PA190&dq=que+son+los+diagramas+path&source=bl&ots=fpF6XR8JXr&sig=-zYxhHaOZGJGoMpRgkQOoTtxXvI&hl=es-419&sa=X&ved=0ahUKewjD4ei1o4TLAhXBQiYKHboaDYUQ6AEIjJAC#v=onepage&q&f=false>
- Lin, W.-S. (2012). Perceived fit and satisfaction on web learning performance: IS continuance intention and task-technology fit perspectives. *International Journal of Human-Computer Studies*, 70(7), 498–507. <http://doi.org/10.1016/j.ijhcs.2012.01.006>
- Manzano, A., & Zamora, S. (2009). *Sistema de ecuaciones estructurales: una herramienta de investigación*. México- México: CENEVAL.
- Mbarek, R., & Zaddem, F. (2013). The examination of factors affecting e-learning effectiveness. *International Journal of Innovation and Applied Studies*, 2(4), 423–435. <http://doi.org/2028-9324>
- Qureshi, I., & Compeau, D. (2009). Assessing between-group differences in information systems research: A comparison of covariance- and component-based SEM. *Management Information Systems Quarterly*, 33(1), 197–214.
- Ramirez-Anormaliza, R., Llinàs-Audet, X., & Sabaté, F. (2016). THE ACCEPTANCE AND USE OF THE E-LEARNING SYSTEMS AMONG THE UNIVERSITY TEACHERS IN ECUADOR. In Iated (Ed.), *EDULEARN16 Conference 4th-6th July 2016* (pp. 3666–3674). Barcelona, Spain. Retrieved from <https://library.iated.org/view/RAMIREZANORMALIZA2016ACC>
- Reguera-Alvarado, N., Blanco-Oliver, A., & Martín-Ruiz, D. (2016). Testing the predictive power of PLS through cross-validation in banking. *Journal of Business Research*, 69(10), 4685–4693. <http://doi.org/10.1016/j.jbusres.2016.04.016>
- Ringle, C. M., Wende, S., & Becker, J. M. (Eds.). (2015). *SmartPLS 3*.

Boenningstedt: SmartPLS GmbH. Retrieved from <http://www.smartpls.com>.

- Rosseel, Y. (2012). lavaan : An R Package for Structural Equation. *Journal of Statistical S*, 48(2).
- Sallan, J. M., Fernandez, V., Simo, P., Lordan, O., & Gonzalez-Prieto, D. (2012). Análisis de modelos de ecuaciones estructurales mediante el paquete lavaan. In 6th International Conference on Industrial Engineering and Industrial Management. XVI Congreso de Ingeniería de Organización. (pp. 951–958). Vigo- España. Retrieved from <http://www.adingor.es/ocs/index.php?conference=cio2012&schedConf=CIO2012>
- Sanjuán Suárez, P., Pérez García, A. M., & Bermúdez Moreno, J. (2000). Escala de autoeficacia general: Datos psicométricos de la adaptación para población española. *Psicothema*, 12(SUPPL. 2), 509–513. <http://doi.org/ISSN 0214-9915>
- Sarstedt, M., Ringle, C. M., Smith, D., Reams, R., & Hair, J. F. (2014). Partial least squares structural equation modeling (PLS-SEM): A useful tool for family business researchers. *Journal of Family Business Strategy*, 5(1), 105–115. <http://doi.org/10.1016/j.jfbs.2014.01.002>
- Schubring, S., Lorscheid, I., Meyer, M., & Ringle, C. M. (2016). The PLSagent : Predictive modeling with PLS-SEM and agent-based simulation. *Journal of Business Research*, 69(10), 4604–4612. <http://doi.org/10.1016/j.jbusres.2016.03.052>
- Tarhini, A., Hone, K., & Liu, X. (2013). Factors Affecting Students' Acceptance of e-Learning Environments in Developing Countries : A Structural Equation Modeling Approach. *International Journal of Information and Education Technology*, 3(1), 54–59. <http://doi.org/10.7763/IJiet.2013.V3.233>
- Urbach, N., & Ahlemann, F. (2010). Structural Equation Modeling in Information Systems Research Using Partial Least Squares. *Journal of Information Technology Theory and...*, 11(2), 5–40. Retrieved from [http://iris.ebs.edu/accessdb/WWW/iris_pub_web.nsf/wwwPublAuthorSingleEng/A58520D5E3C484DCC1257C50007F8697/\\$file/Urbach et al. 2010 Structural Equation Modeling in IS Research Using PLS.pdf](http://iris.ebs.edu/accessdb/WWW/iris_pub_web.nsf/wwwPublAuthorSingleEng/A58520D5E3C484DCC1257C50007F8697/$file/Urbach et al. 2010 Structural Equation Modeling in IS Research Using PLS.pdf)
- Weng, C., & Tsai, C. (2015). Social support as a neglected e-learning motivator affecting trainee ' s decisions of continuous intentions of usage. *Australasian Journal of Educational Technology*, 31(2), 177–192.
- Wold, H. (1974). Causal flows with latent variables. *European Economic Review*, 5(1), 67–86. [http://doi.org/10.1016/0014-2921\(74\)90008-7](http://doi.org/10.1016/0014-2921(74)90008-7)
- Ziersch, A. M., Baum, F. E., Macdougall, C., & Putland, C. (2005). Neighbourhood life and social capital: the implications for health. *Social Science & Medicine* 60, 60, 71–86. <http://doi.org/10.1016/j.socscimed.2004.04.027>

UNENMI 2017

“La mente que se abre a una nueva idea, jamás volverá a su tamaño original”

Albert Einstein (1879-1955)

Físico alemán

Capítulo 5: Análisis de clases latentes

Jose Diaz-Montenegro

Fabian Ramirez-Granda

Omar Franco-Arias

Objetivos de aprendizaje

Para este capítulo se plantea dar a conocer en profundidad los modelos de ecuaciones estructurales (SEM) como técnica multivariante para investigar en varios campos. De manera específica los objetivos planteados son los siguientes:

- Conocer las características de los modelos de ecuaciones estructurales su terminología y notación básica.
- Distinguir los enfoques de ecuaciones estructurales y los criterios para su selección a la hora de realizar una investigación.
- Conocer los diferentes programas informáticos para tratamiento de ecuaciones estructurales.
- Aplicar el procesamiento de datos por medio de modelos de ecuaciones estructurales con enfoque en mínimos cuadrados parciales.
- Conocer el análisis multigrupo mediante ecuaciones estructurales con enfoque en mínimos cuadrados parciales.
- Aplicar el procesamiento de datos por medio de modelos de ecuaciones estructurales con enfoque en covarianzas.

5.1. Introducción

En los procesos de investigación es común encontrarse con variables no observables, pero que son explicadas por los movimientos de otras variables que si lo son e incluso estas pueden constituir conjunto de variables y también es cierto que en muchas ocasiones una de estas variables es la que mayor incidencia tiene en la explicación de las condiciones de la variable no observable, a estas variables se les ha dado en llamar variables latentes, en tal sentido es totalmente relevante considerar el error de medida que se genera en la especificación de los modelos estadísticos según (Collins & Lanza, 2010) Bajo estas referencias el error de medida ha dado lugar a diferentes modelos de estructuras latentes basados en la idea de independencia local, lo que significa que las variables observadas son independientes para un valor particular de la variable no observada o latente, esta estructura va a generar distintos modelos.

Podría pensarse que las variables latentes no corresponden a la realidad física pues no es así, corresponden a conceptos abstractos de esa realidad, como categorías, estados del comportamiento, estructuras de datos. En ocasiones se usan como denominaciones alternativas los términos variables hipotéticas o constructos hipotéticos. La inferencia que se realiza a través de modelos matemáticos que usan el análisis de las variables observables en términos de las variables latentes, y se les denomina modelos de variables latentes.

Estos modelos se utilizan en muchas disciplinas, como la psicología, la economía, la medicina, la física, el aprendizaje máquina, la inteligencia artificial, bioinformática, procesamiento del lenguaje natural, la econometría, la gestión y las ciencias sociales. Un ejemplo de variables latentes muy obvio es: la economía, que como tal tiene otras variables latentes que son la calidad de vida, la confianza empresarial, la moral, la felicidad y el conservadurismo, si consideramos las variables observables que pueden atribuirse a la determinación de la calidad de vida estas son la riqueza, el empleo, el medio ambiente, la salud física y mental, educación, recreación, tiempo libre, y la pertenencia social.

5.2. Conceptos y referencia histórica

El análisis de clases latentes (LCA) es un enfoque de naturaleza probabilística y centrado en la persona, utilizado para identificar grupos de individuos en una población heterogénea. Este enfoque de la modelización permite observar la heterogeneidad de la población organizando individuos en grupos que no eran observables ni conocidos antes del análisis. La pertenencia a un

grupo se deduce de los datos y se identifican las subpoblaciones. LCA agrupa a los individuos creando una variable latente categórica, que representa la membresía no observable (probabilística) en un perfil.

Uno de los primeros trabajos sobre el análisis de la clase latente fue el libro de Lazarsfeld y Henry (1968). No fueron los primeros en sugerir la idea de una variable latente categórica, pero su libro representa el primer tratamiento matemático y conceptual completo y detallado del tema. Aunque Lazarsfeld y Henry demostraron convincentemente el potencial del análisis de clase latente en las ciencias sociales y del comportamiento, la falta de un método general y confiable para obtener estimaciones de parámetros fue una barrera importante que impidió la implementación generalizada de sus ideas por un tiempo. (Miglioretti, 2003)

En la siguiente década Goodman desarrolló un método sencillo y fácilmente de implementar para obtener estimaciones de máxima verosimilitud de los parámetros del modelo de clase latente. El enfoque de Goodman para la estimación se demostró más tarde que está estrechamente relacionado con el algoritmo de expectativa-maximización (EM). El algoritmo EM se utiliza hoy en día en muchos programas de LCA. (Miglioretti, 2003)

El modelo de clase latente se hizo mucho más general cuando se colocó dentro del marco de modelos log-lineales. Esto abrió una serie de posibilidades para la adaptación de modelos y sentó las bases para algunos nuevos desarrollos importantes. Un desarrollo fue la incorporación de covariables en modelos de clases latentes. Otro modelo fueron los modelos que identifican las clases latentes basadas en trayectorias individuales de crecimiento en datos longitudinales. (Miglioretti, 2003)

5.2.1. El análisis de clases latentes

Un modelo de clase latente utiliza los diferentes patrones de respuesta en los datos para encontrar grupos similares. Intenta asignar grupos que sean "independientes condicionales". Esto significa que dentro de un grupo las correlaciones entre las variables se convierten en cero, porque la pertenencia al grupo explica cualquier relación entre las variables.

El análisis de clases latentes es diferente del análisis de perfil latente, ya que este último utiliza datos continuos y el primero puede utilizarse con datos categóricos. Otro aspecto importante del análisis de la clase latente es que sus elementos (personas, observaciones) no están asignados absolutamente, sino en la probabilidad. De este modo obtendrá un valor de probabilidad para cada persona asignada al grupo 1, grupo 2 [...] grupo k.

Antes de estimar su modelo LCA usted tiene que elegir el número de grupos que desea tener. Se apunta a un pequeño número de clases, por lo que el modelo sigue siendo adecuado para los datos, pero también parsimonioso. Es muy común utilizar los denominados “criterios de información” para determinar el grupo que mejor se ajusta a los datos. Los principales criterios utilizados son: el Criterio de Información Bayesiano (BIC), el Criterio de Información Akaike (AIC) y el Criterio de Información Constante de Akaike (CAIC)

Comúnmente, LCA se lleva a cabo especificando una serie de modelos con diferentes números de clases latentes. Los cambios en las estadísticas de ajuste relativas, tales como el Criterio de Información de Akaike [AIC], el Criterio de Información Bayesiano [BIC] y el BIC ajustado por tamaño de muestra [aBIC]) Estos criterios de información ponderan el ajuste y la parsimonia de un modelo: los menores valores de BIC inferior, AIC, o CAIC representan el mejor modelo.

Los modelos de clases latentes pertenecen a la familia de modelos de mezcla (finita). Los parámetros son estimados por el algoritmo EM. Se llama EM, porque tiene dos pasos: un paso de estimulación “E” y un paso de aximización “M”. En el primero se estiman las probabilidades de pertenecer a la clase (la primera vez con algunos valores iniciales) y en el segundo se modifican esas estimaciones para maximizar la función de verosimilitud. Ambos pasos son iterativos y repetidos hasta que el algoritmo encuentra el máximo global. Esta es la solución con la mayor probabilidad posible.

5.2.2. Formulación matemática

5.2.2.1. El modelo de clase latente básico

Siguiendo a Bakk et al. (2013), denotemos la variable latente categórica por X , una clase latente particular por t , y el número de clases por T ; Como tal tenemos $X \in \{1, 2, \dots, T\}$. Sea Y una de las variables indicadoras manifiestas, donde $Y_j \in \{0, 1\}$. Sea \mathbf{y} un vector que contenga un patrón de respuesta completo y \mathbf{y} su realización. Un modelo de clase latente para la probabilidad de observar el patrón de respuesta \mathbf{y} se puede definir como sigue:

$$P(\mathbf{Y} = \mathbf{y}) = \sum_t P(X = t)P(\mathbf{Y} = \mathbf{y} | X = t), \quad (1)$$

Donde $P(X=t)$ representa la probabilidad de pertenecer a la clase t y $P(\mathbf{Y}=\mathbf{y}|X=t)$ la probabilidad de que el patrón de respuesta \mathbf{y} condicional pertenezca a la clase t . Como podemos ver en la ecuación 1, se supone que

la probabilidad marginal de obtener el patrón de respuesta \mathbf{y} es una media ponderada de las t probabilidades específicas de clase.

En un LCA clásico asumimos la independencia local, lo que significa que las variables del indicador K se supone que son mutuamente independientes dentro de cada clase t . Esto implica que la probabilidad conjunta de un patrón de respuesta específico en el vector de las variables del indicador es el producto de las probabilidades específicas de cada ítem:

$$P(\mathbf{Y} = \mathbf{y} | X = t) = \prod_k P(Y_k = y_k | X = t). \quad (2)$$

Combinando las ecuaciones 1 y 2, obtenemos lo siguiente:

$$P(\mathbf{Y} = \mathbf{y}) = \sum_t P(X = t) \prod_k P(Y_k = y_k | X = t). \quad (3)$$

Los parámetros del modelo de interés son las proporciones de clase $P(X = t)$ y las probabilidades de respuesta de clase $P(Y_k = y_k | X = t)$. Estos parámetros son usualmente estimados por máxima verosimilitud (maximum likelihood, ML).

5.2.2.2. Obtención de predicción de clases latentes

Siguiendo a Bakk et al. (2013), mientras que las verdaderas membresías de clase no pueden ser observadas, los parámetros del modelo de medición descrito en las ecuaciones 1 a 3 pueden usarse para derivar procedimientos para estimar estas pertenencias a clases, es decir, para asignar individuos a clases. La predicción se basa en la probabilidad posterior de pertenecer a la clase dado un patrón de respuesta observado \mathbf{y} , $P(X = t | \mathbf{Y} = \mathbf{y})$, que puede obtenerse aplicando el teorema de Bayes, es decir,

$$P(X = t | \mathbf{Y} = \mathbf{y}) = \frac{P(X = t) P(\mathbf{Y} = \mathbf{y} | X = t)}{P(\mathbf{Y} = \mathbf{y})}. \quad (4)$$

Estas probabilidades de pertenencia a clases posteriores proporcionan información sobre la distribución de las clases T entre individuos con patrón de respuesta \mathbf{y} , lo que refleja que las personas que tienen el mismo patrón de respuesta pueden pertenecer a diferentes clases. Es importante señalar que cada individuo pertenece a una sola clase pero que no sabemos a qué. Utilizando las probabilidades posteriores de pertenencia a clases, se pueden

k

utilizar diferentes tipos de reglas para asignar sujetos a clases, las más populares son asignación modal y proporcional.

Cuando se utiliza la asignación modal, cada individuo se asigna a la clase para la cual su probabilidad de pertenencia posterior es la más grande. Denotando la clase predicha por W y el patrón de respuesta del sujeto i por \mathbf{y}_i , el particionamiento duro correspondiente a la asignación modal puede expresarse como el siguiente:

$$P(W = s | \mathbf{Y} = \mathbf{y}_i) = 1 \text{ si } P(X = s | \mathbf{Y} = \mathbf{y}_i) > P(X = t | \mathbf{Y} = \mathbf{y}_i) \quad \forall t \neq s; 0 \text{ en otros casos (5)}$$

Se asigna un individuo con probabilidad o peso igual a 1 a la clase con mayor probabilidad posterior y con peso 0 a las otras clases.

Una alternativa a la asignación modal es la asignación proporcional, que en el contexto de la agrupación basada en modelos se denomina método de partición suave. Un individuo con el patrón de respuesta \mathbf{y}_i será asignado a cada clase s con un peso $P(W = s | \mathbf{Y} = \mathbf{y}_i) = P(X = s | \mathbf{Y} = \mathbf{y}_i)$, es decir, con un peso igual a la probabilidad de pertenencia posterior (Bakk, Tekle, & Vermunt, 2013).

5.3. Las causas del análisis de variables latentes en el análisis multivariante

En la revisión realizada se define como la razón más importante para facilitar el aumento de la popularidad de los análisis de clases latentes como herramienta estadística para el análisis de conglomerados, el que los ordenadores de alta velocidad hacen ahora estos métodos computacionalmente intensivos prácticamente aplicables. Varios paquetes de software están disponibles para la estimación de modelos de clúster de clases latentes. (Magidson & Vermunt, 2002) y a la aplicación práctica de los resultados que se obtienen que están estrechamente vinculados a la realidad.

En lo específicamente técnico el aumento del interés en los modelos LC se atribuye al desarrollo de algoritmos informáticos extendidos, que permiten a los ordenadores de hoy en día para llevar a cabo el análisis de clases latentes en los datos que contienen algo más que unas pocas variables. (Magidson & Vermunt, 2005)

Dentro de las áreas y modelos LC propuestas por Magidson and Vermunt las clasifica primero en tres áreas de aplicación estadísticas comunes de análisis de LC, son aquellos que implican:

1. agrupación de los casos,
2. reducción variable y construcción escala, y
3. predicción de una variable dependiente.

Y luego en tres principales tipos de modelos LC:

1. Modelos de racimo LC,
2. Los modelos de factores LC,
3. LC regresión y Modelos de elección (Magidson & Vermunt, 2005).

5.4. La relación entre la variable latente y sus indicadores

Esta relación queda definida en la relación sea reflectiva o sea formativa dependiendo de su naturaleza de causa y efecto, y la intensidad de la relación definida para las mismas, podría decirse que no existe ninguna novedad, más la novedad está en la definición de las variables con nivel de error para generar una variable latente que no admite error si no por consecuencia de sus componentes.

Bajo este antecedente y cuando hablamos de la economía del conocimiento, todas las variables se orientan a la categorización abstracta de elementos en distintos ámbitos del quehacer humano, el desarrollo de constructos, que se expresan como consecuencia nos genera espacios más reales para el análisis bajo la condición de la objetividad en el análisis y en la información obtenida.

5.5. Software para el análisis LCA

5.5.1. Ajuste de un modelo LCA en software R

En esta sección, vamos a utilizar la función `poLCA` un paquete del software libre R que utiliza la maximización de la expectativa (*expectation-maximization*) y los algoritmos de Newton-Raphson para hallar la máxima verosimilitud (*maximum likelihood*) de los parámetros. El paquete `poLCA` tiene su nombre de *Polytomous Latent Class Analysis*.

A continuación, desarrollaremos un ejercicio de LCA utilizando algunas funciones del paquete `poLCA` de R. Para un completo detalle sobre los diferentes contenidos de este paquete recomendamos ver (Linzer & Lewis, 2011)

5.5.2. Instalación paquete polCA

Primero debe tener instalada una versión de R en su computadora <https://cran.r-project.org/mirrors.html>. Luego debe abrir R y en la consola de códigos ingresar:

```
install.packages("polCA") (1)
```

Una vez finalizado el proceso de instalación, para cargar el paquete en la memoria escriba:

```
library(polCA) (2)
```

PoLCA también viene con 5 conjuntos de datos **preinstalados** de muestra, que son útiles para explorar diferentes aspectos de la clase latente y los modelos de regresión de clases latentes: "carcinoma", "cheating", "election", "gss82", y "values". Se puede acceder a cualquiera de ellos mediante el comando `data` (nombre), por ejemplo:

```
data(carcinoma) (3)
```

Por otro lado, el comando:

```
polCA.simdata (4)
```

generará conjuntos de **datos simulados** que se pueden utilizar para examinar las propiedades de la clase latente y los estimadores del modelo de regresión de clases latentes. Ejemplos de usos posibles de los datos de `polCA.simdata` se incluyen en la documentación interna de polCA a la cual se puede acceder ingresando:

```
?polCA.simdata (5)
```

Para conocer con mayor detalle el uso de estas opciones de conjuntos de datos **preinstalados** y **simulados** ver Linzer y Lewis (2011).

5.5.3. Corriendo un modelo LCA en R

5.5.3.1. Elegir el número de grupos

Antes de estimar su modelo LCA usted tiene que elegir el número de grupos que desea tener. Se apunta a un pequeño número de clases, por lo que el modelo sigue siendo adecuado para los datos, pero también parsimonioso.

La definición del número de grupos se lo puede hacer basado en dos consideraciones: i) En base a una teoría o estudios anteriores, en cuyo caso se especificará que solo se muestre esta solución; ii) un enfoque más exploratorio a través del cual se desea varios modelos -uno con 2 grupos, otro con 3 grupos y otro con 4 grupos, por ejemplo- y comparar estos modelos entre sí, comparándose las diferentes soluciones con varios criterios de información como BIC, AIC, CAIC, etc.

5.5.3.2. Preparando los datos

Como se ha explicado en la sección teórica de este capítulo, los modelos de clases latentes asumen que las variables son categóricas, por lo tanto, no se permite que las variables contengan ceros, valores negativos o decimales. Se requiere, entonces codificar las variables como factores o como valores enteros comenzando en uno (1) para la primera categoría de resultados y aumentando hasta el número máximo de resultados para cada variable, de lo contrario poLCA producirá un mensaje de error y terminará sin estimar el modelo. Los datos de entrada si pueden contener valores faltantes.

5.5.3.3. Estimación del modelo

Una vez instalado el paquete poLCA, realizado los cambios en los datos –si fuera necesario- para que las variables sean categóricas, y cargado el archivo de datos en R, estamos listos para estimar un modelo LCA.

Para estimar un modelo LCA con dos clases latentes (el valor predeterminado), la instrucción simplemente es:

```
ejem1lca <- poLCA(f,data) (6)
```

f: denominado muchas veces fórmula. Es un objeto creado en R en el cual las variables observadas (ítems) se combinan y se ponen igual a 1. Así, por ejemplo, dado un archivo de datos cargado en R con dos variables observadas X1, X2, las almacenaríamos en el objeto f (puede recibir cualquier nombre) con la siguiente instrucción:

```
f <- cbind(X1,X2) ~1 (7)
```

data: es el nombre del archivo de datos con el cual vamos a trabajar el modelo LCA.

La salida de la línea de comando (6) se guardará en el objeto ejem1lca.

Si las clases a estimar fueran más de dos, por ejemplo 4, la línea de comando

(6) se escribiría como:

```
ejem1lca <- poLCA(f,data,nclass=4) (8)
```

Por tanto,

nclass: especifica el número de clases latentes asumidas en el modelo, que es 2 por defecto, y se debe señalar puntualmente en caso contrario.

Una línea de comando mucha más completa está dada por:

```
> ejem1lca <- poLCA(formula, data, nclass=2, maxiter=1000,  
graphs=FALSE, tol=1e-10, na.rm=TRUE, probs.start=NULL,  
nrep=1, verbose=TRUE, calc.se=TRUE)
```

En la cual:

- **maxiter:** el número máximo de iteraciones a través de las cuales el algoritmo estimación el ciclo de repeticiones. Si no se alcanza la convergencia antes de alcanzar este número de iteraciones, poLCA termina e informa un mensaje de error. El valor predeterminado es 1000, pero esto será insuficiente para ciertos modelos.
- **graphs:** es un comando de tipo lógico que señala si poLCA debe mostrar gráficamente las estimaciones de parámetros en la finalización del algoritmo de estimación. El valor predeterminado es falso.
- **tol:** un valor de tolerancia para juzgar cuándo se ha alcanzado la convergencia. Cuando el cambio de iteración en la probabilidad de log estimada es menor que tol, el algoritmo de estimación deja de actualizarse y considera que se ha encontrado la máxima probabilidad de logaritmos. El valor predeterminado es 1×10^{-10} , que es un valor estándar. Esta opción raramente tendrá que ser invocada.
- **na.rm:** es un comando de tipo lógico, a través del cual poLCA maneja casos con valores faltantes en las variables manifiestas. Si es VERDADERO, estos casos se eliminan (suprimidos en la lista) antes de estimar el modelo. Si FALSO, se mantienen los casos con valores faltantes. El valor predeterminado es TRUE.
- **probs.start:** una lista de matrices de probabilidades de respuesta condicional-clase, π_{jrk} que se utilizarán como los valores iniciales para el algoritmo de estimación EM. Cada matriz de la lista corresponde a una variable de manifiesto, con una fila para cada clase latente y una columna para cada resultado posible. El valor predeterminado es NULL, lo que significa que los valores iniciales se generan al azar. Se debe tener en cuenta que si $nrep > 1$, todos los valores probs.start especificados por el

usuario sólo se utilizan en el primer intento de nrep.

- **nrep:** número de veces que se debe estimar el modelo, usando diferentes valores de probs.start. El valor predeterminado es 1. El establecimiento de nrep > 1 automatiza la búsqueda de la función de log-verosimilitud en lugar de sólo un máximo local. PolCA devuelve sólo las estimaciones de parámetros correspondientes al modelo que produce la mayor log-verosimilitud.

5.5.3.4. Ejemplo 1 (basado en Linzer y Lewis (2011))

Vamos a trabajar con los datos del archivo carcinoma que viene incorporado en R.

```
#cargar el archivo carcinoma en R
data(carcinoma) (9)

#ver el contenido del archivo
head(carcinoma) (10)
```

Se mostrará la siguiente información:

```
  A B C D E F G
1 1 1 1 1 1 1 1
2 1 1 1 1 1 1 1
3 1 1 1 1 1 1 1
4 1 1 1 1 1 1 1
5 1 1 1 1 1 1 1
6 1 1 1 1 1 1 1
```

Este archivo consta de 7 variables dicotómicas que representan las calificaciones de siete patólogos, en base a 118 muestras, sobre la presencia (1) o ausencia (2) de carcinoma. El propósito de estudiar estos datos es modelar el “acuerdo interobservador” examinando cómo los sujetos podrían ser divididos en grupos dependiendo de la consistencia de sus diagnósticos.

```
#creamos un objeto en el que las 7 variables observadas
#(ítems) se combinan y se fijan igual a 1
f <- cbind(A,B,C,D,E,F,G)~1 (11)
```

La corrida de los diferentes modelos puede darse por separado. Por ejemplo, si deseamos correr un modelo con 3 clases, bastaría la instrucción:

```
lc3 <- poLCA(f,carcinoma,nclass=3) (12)
```

Luego para los siguientes modelos, nclass pasaría a ser igual a 3, 4, etc. Finalmente tendríamos que comparar en cuál de estos modelos el BIC y el AIC es más bajo, para escoger el mejor.

Sin embargo, para este ejemplo vamos a dar una instrucción a fin de que automáticamente corra una secuencia de modelos determinada y que al final nos muestre cuál de ellos tiene el BIC y el AIC más bajo. Concretamente vamos a pedirle que corra modelos de 2 a 10 clases y cada modelo se ejecute 10 veces, para ello escribiremos en R la siguiente secuencia de comandos:

```
max_II <- -100000
min_bic <- 10000
for(i in 2:10){lc <- poLCA(f,carcinoma,nclass=i,maxiter=3000,
  tol=1e-5, na.rm=FALSE, nrep=10, verbose=TRUE, calc.se=TRUE)
  if(lc$bic < min_bic){min_bic <- lc$bic
  LCA_best_model<-lc}} (13)
```

Omitimos la explicación de algunas instrucciones por cuanto entran en el ámbito de la programación en R en general. Puede se puede observar en la última línea de instrucciones, hemos pedido que el mejor modelo se guarde LCA_best_model. A continuación, pediremos se nos muestre su contenido:

```
LCA_best_model (14)
```

Con lo cual se desplegará la siguiente información:

```
Conditional item response (column) probabilities,
  by outcome variable, for each class (row)
$A
      Pr(1) Pr(2)
class 1: 0.4870 0.5130
class 2: 0.0000 1.0000
class 3: 0.9427 0.0573
$B
      Pr(1) Pr(2)
class 1: 0.0000 1.0000
class 2: 0.0191 0.9809
class 3: 0.8619 0.1381
```

```

$C
      Pr(1) Pr(2)
class 1: 1.0000 0.0000
class 2: 0.1425 0.8575
class 3: 1.0000 0.0000
$D
      Pr(1) Pr(2)
class 1: 0.9424 0.0576
class 2: 0.4137 0.5863
class 3: 1.0000 0.0000
$E
      Pr(1) Pr(2)
class 1: 0.2492 0.7508
class 2: 0.0000 1.0000
class 3: 0.9449 0.0551
$F
      Pr(1) Pr(2)
class 1: 1.0000 0.0000
class 2: 0.5236 0.4764
class 3: 1.0000 0.0000
$G
      Pr(1) Pr(2)
class 1: 0.3691 0.6309
class 2: 0.0000 1.0000
class 3: 1.0000 0.0000
Estimated class population shares
 0.1817 0.4447 0.3736
Predicted class memberships (by modal posterior prob.)
 0.1949 0.4322 0.3729

=====
Fit for 3 latent classes:
=====
number of observations: 118
number of estimated parameters: 23
residual degrees of freedom: 95
maximum log-likelihood: -293.705
AIC(3): 633.41
BIC(3): 697.1357
G^2(3): 15.26172 (Likelihood ratio/deviance statistic)
X^2(3): 20.50183 (Chi-square goodness of fit)

```

Esta salida es la que realmente nos interesaría para realizar los análisis pertinentes. Explicamos a continuación sus diferentes secciones, teniendo presente que los datos se refieren a las calificaciones de 7 patólogos (variables A a G), en base a 118 muestras sobre la presencia (1-diagnóstico negativo) o ausencia (2-diagnóstico positivo) de carcinoma.

- La primera sección de la salida incluye las **probabilidades de respuesta** para cada clase latente en cada elemento, es decir, las probabilidades de respuesta para los patólogos A a G, con cada fila correspondiente a una clase latente, y cada columna correspondiente a un diagnóstico: negativo en la primera columna, y positivo en la segunda. Por ejemplo, una muestra perteneciente a la primera clase (“negativa”) tiene una probabilidad del 49% de ser clasificada como libre de carcinoma por el evaluador A, una probabilidad del 0% de la misma del evaluador B, una probabilidad del 100% del evaluador C, y así sucesivamente.
- La segunda sección de salida consta de la **proporción de muestras clasificadas en cada clase latente**, las cuales aparecen después de las probabilidades de respuesta del ítem *-Estimated class population shares-*. Podemos ver que la primera clase constaba de 0,1817 de las 118 muestras de la población, mientras que la clase 2 constaba de 0.4447 de las muestras, y la clase 3 de 0.3736.
- En una tercera sección, las **probabilidades posteriores modales de la pertenencia a la clase** aparecen a continuación *-Predicted class memberships (by modal posterior prob.)-*. Estos valores son muy similares a las poblaciones de clase estimadas de la tabla anterior, pero representan la media de las probabilidades de la clase modal para cada muestra de la población.
- Finalmente, se nos presenta la información de **ajuste del modelo** para la solución de tres clases latentes *-Fit for 3 latent classes-*. Esta tabla incluye el tamaño total de la muestra utilizado para ajustar el modelo, seguido por el número de parámetros estimados, los grados de libertad del modelo (tamaño de la muestra - parámetros estimados) y el valor máximo de log-verosimilitud para la solución. El ajuste relativo del modelo se puede evaluar utilizando el AIC y el BIC que son las posibilidades de evaluación que por el momento ofrece poLCA en R.

Conociendo ya que el modelo de 3 clases es el de mejor ajuste podemos pedir un gráfico mediante la siguiente instrucción:

```
lc <- poLCA(f,carcinoma, graphs = TRUE, nclass=3,maxiter=3000) (15)
```

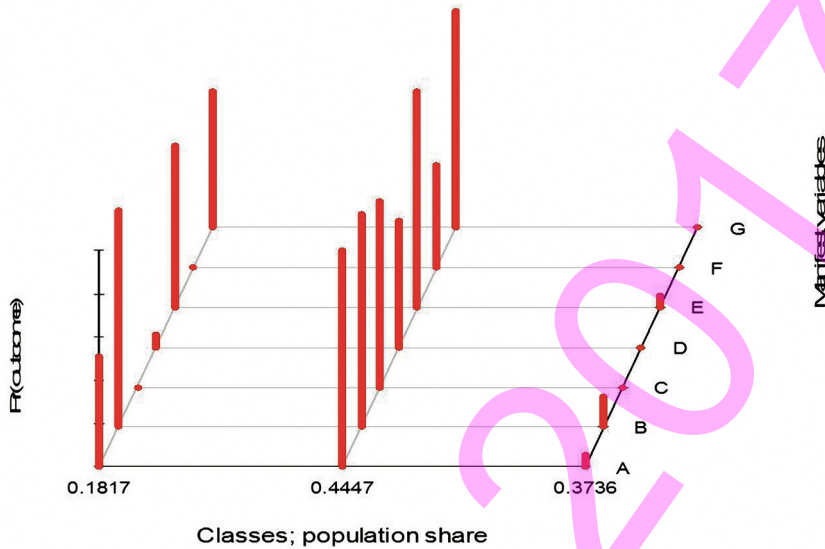


Figura 5 - 1. Modelo de solución de 3 clases

Cada grupo de barras rojas representa las probabilidades condicionales, por clase latente, de ser valoradas positivamente por cada uno de los siete patólogos (etiquetados de A a G). Las barras más altas corresponden a probabilidades condicionales más cercanas a dar un diagnóstico positivo, en tanto que las barras más bajas indican mayores probabilidades condicionales de dar un diagnóstico negativo.

Tal y como nos mostró la segunda sección de la salida comentada anteriormente, en la primera clase se concentra el 18% de las muestras, en la segunda clase el 44% y en la tercera clase el 37%. Analizamos cada una de las clases.

- La primera clase podría ser denominada “problemática” por cuanto no hay una tendencia a diagnósticos positivos o negativos por parte de los patólogos. Mientras los patólogos B, E y G tienden a diagnosticar positivos; C, D y F tienden a diagnosticar negativos; y A es aproximadamente 50/50.
- La segunda clase podría denominarse “positivos”, por cuanto la mayoría de sus barras son altas.
- La tercera clase podría denominarse “negativos”, por cuanto la mayoría de sus barras son bajas.

5.6. Resumen del Capítulo

En el inicio del capítulo se ha presentado el concepto básico de clases latentes entendidas como constructos o variables no observados de error cero que se miden a través de otras variables observables cuyo referente de medición se denomina indicador y es producto de un proceso para la obtención de información, a continuación se presenta una breve reseña histórica con la referencia de los autores que iniciaron el tratamiento científicamente riguroso de las variables latentes, se muestra el modelo matemático y cómo funciona el mismo mediante la presentación de ecuaciones. También se menciona que el desarrollo tecnológico a través de software que se aplica a variables latentes ha permitido el desarrollo de la técnica por la cantidad de cálculos que debe realizarse para lograr las estimaciones y la sintonía de estas con la realidad que finalmente se evidencia, para finalizar se presenta el desarrollo de ejercicios para encontrar resultados de variables latentes a través de la corrida de un modelo LCA en R.

5.7. Referencias

- Arteaga, R., & Duarte, A. (2010). Motivational factors that influence the acceptance of Moodle using TAM. *Computers in Human Behavior*, 26(6), 1632–1640. <http://doi.org/10.1016/j.chb.2010.06.011>
- Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the Association between Latent Class Membership and External Variables Using Bias-adjusted Three-step Approaches. *Sociological Methodology* (Vol. 43). <http://doi.org/10.1177/0081175012470644>
- Collins, L. M., & Lanza, S. T. (2010). LATENT CLASS AND LATENT.
- Findik, D., & Kunçay, C. O. Ş. (2009). A model for instructors adoption of learning management systems : empirical validation in higher education context. Middle East Technical University.
- Lay, J.-G., Chen, Y.-W., & Chi, Y.-L. (2013). GIS Adoption Among Senior High School Geography Teachers in Taiwan. *Journal of Geography*, 112(3), 120–130. <http://doi.org/10.1080/00221341.2012.682226>
- Lin, W.-S. (2012). Perceived fit and satisfaction on web learning performance: IS continuance intention and task-technology fit perspectives. *International Journal of Human-Computer Studies*, 70(7), 498–507. <http://doi.org/10.1016/j.ijhcs.2012.01.006>
- Linzer, D. A., & Lewis, J. B. (2011). polCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software*, 42(10), 1–29.
- Magidson, J., & Vermunt, J. K. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research*, 20.
- Magidson, J., & Vermunt, J. K. (2005). A nontechnical introduction to latent class models. *DMA Research Council Journal*, 1–15.

- Mbarek, R., & Zaddem, F. (2013). The examination of factors affecting e-learning effectiveness. *International Journal of Innovation and Applied Studies*, 2(4), 423–435. <http://doi.org/2028-9324>
- Miglioretti, D. L. (2003). Latent transition regression for mixed outcomes. *Biometrics*, 59(3), 710–720. <http://doi.org/10.1111/1541-0420.00082>
- Sanjuán Suárez, P., Pérez García, A. M., & Bermúdez Moreno, J. (2000). Escala de autoeficacia general: Datos psicométricos de la adaptación para población española. *Psicothema*, 12(SUPPL. 2), 509–513. <http://doi.org/ISSN 0214-9915>
- Tarhini, A., Hone, K., & Liu, X. (2013). Factors Affecting Students' Acceptance of e-Learning Environments in Developing Countries : A Structural Equation Modeling Approach. *International Journal of Information and Education Technology*, 3(1), 54–59. <http://doi.org/10.7763/IJiet.2013.V3.233>
- Weng, C., & Tsai, C. (2015). Social support as a neglected e-learning motivator affecting trainee 's decisions of continuous intentions of usage. *Australasian Journal of Educational Technology*, 31(2), 177–192

UNVENMI 2017



“La evidencia es la más decisiva demostración”

Marco Tulio Cicerón (106 a. C. – 43 a. C.)

Escritor, político y orador romano

BIBLIOGRAFÍA

- Abridged ACRA scale of learning strategies for Escala de estrategias de aprendizaje ACRA-Abreviada para alumnos universitarios Jesús de la Fuente Arias *, Fernando Justicia Justicia **. (2003), (January).
- Albers, S. (2010). PLS and success factor studies in marketing. (V. Vinzi Esposito, W. . Chin, J. Henseler, & H. (Eds) Wang, Eds.)Springer. Berlin: Handbook of Partial Least Squares.
- Albuissou, E., Allain, Y. M., Rozan, R., Giraud, B., Bolla, M., Brune, D., ... Richaud, P. (1992). Prostate cancer treated by radiotherapy: a multivariate study. *Radiotherapy and Oncology*, 25(2), 89–96. [http://doi.org/10.1016/0167-8140\(92\)90014-L](http://doi.org/10.1016/0167-8140(92)90014-L)
- Aldás Manzano, J. (2011). Análisis De Conglomerados. Universitat de València.
- Alonso, A. (2008). Técnicas de análisis multivariante de datos. (I. Capella & M. Diego, Eds.) (Primera). Madrid-España: PEARSON PRENTICE HALL.
- Ameh, S., Gómez-Olivé, F. X., Kahn, K., Tollman, S. M., & Klipstein-Grobusch, K. (2017). Relationships between structure, process and outcome to assess quality of integrated chronic disease management in a rural South African setting: applying a structural equation model. *BMC Health Services Research*, 17(1), 229. <http://doi.org/10.1186/s12913-017-2177-4>
- Amiry, S., Esmaili, M., & Alizadeh, M. (2017). Classification of adulterated honeys by multivariate analysis. *Food Chemistry*, 224, 390–397. <http://doi.org/10.1016/j.foodchem.2016.12.025>
- Aparecida, A., & Thatiana, H. (2017). Fonológica, a phonological awareness test Análise fatorial exploratória e confirmatória do Roteiro de Avaliação da Consciência Fonológica. *Estudos de Psicologia Campina*, 34(2), 211–218.
- Arias, B. (2008). SEM, Structural Equation Models),.
- Arteaga-arcentales, E., Peralta-carpio, F., Castelo-gonzalez, J., & Echeverría-vásquez, H. (2016). Principales Factores para Desarrollo Turístico : Estudio Empírico a la Percepción de la Población Económicamente Activa. PODIUM.
- Arteaga, R., & Duarte, A. (2010). Motivational factors that influence the acceptance of Moodle using TAM. *Computers in Human Behavior*, 26(6), 1632–1640. <http://doi.org/10.1016/j.chb.2010.06.011>
- Ayodeji, A., & Al-lawati, A. M. (2010). Automation in Construction Using PLS-SEM technique to model construction organizations' willingness to participate in e-bidding. *Automation in Construction*, 19(6), 714–724. <http://doi.org/10.1016/j.autcon.2010.02.016>
- Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the Association between Latent Class Membership and External Variables Using Bias-

- adjusted Three-step Approaches. *Sociological Methodology* (Vol. 43). <http://doi.org/10.1177/0081175012470644>
- Ballesteros, J., Moral, E., Brieva, L., Ruiz-Beato, E., Prefasi, D., & Maurino, J. (2017). Psychometric properties of the SDM-Q-9 questionnaire for shared decision-making in multiple sclerosis: item response theory modelling and confirmatory factor analysis. *Health and Quality of Life Outcomes*, 15(1), 79. <http://doi.org/10.1186/s12955-017-0656-2>
 - Banno, H., Takeda, K., & Itakura, F. (2002). A model of perceptual distance for group delays based on ellipsoidal mapping. *Acoust. Sci. & Tech*, 2, 110–112.
 - Barclay, D., Higgins, C., & Thompson, R. (1995). The partial least square (PLS) approach to casual modeling: Personal computer adoption and use as an illustration. *Technology Studies (Special Issue of Research Methodology)*.
 - Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/3806354>
 - Bass, S. B., Muñiz, J., Gordon, T. F., Maurer, L., & Patterson, F. (2016). Understanding help-seeking intentions in male military cadets: An application of perceptual mapping. *BMC Public Health*, 16(1), 413. <http://doi.org/10.1186/s12889-016-3092-z>
 - Beaujean, A. A. (2015). Package “BaylorEdPsych.” CRAN.
 - Campi, C., Parkkonen, L., Hari, R., & Hyvarinen, A. (2013). Non-linear canonical correlation for joint analysis of MEG signals from two subjects. *Frontiers in Neuroscience*, 7(7 JUN), 1–7. <http://doi.org/10.3389/fnins.2013.00107>
 - Camprubí, R., Guia, J., & Comas, J. (2009). La formación de la imagen turística inducida: Un modelo conceptual. *Pasos, Revista de Turismo Y Patrimonio Cultural*, 7, 255–270.
 - Carmines, E. G., & Zeller, R. A. (1979). Reliability and validity assessment. (J. L. Sullivan & R. G. Niemi, Eds.) *Quantitative Applications in the Social Sciences* (Vol. 17). Beverly Hill - USA: McCune, Sara. <http://doi.org/10.1037/018269>
 - Carrillo, J., Rubio, J., & ... (2016). Identification of areas of intervention for public safety policies using multiple correspondence analysis. *Dyna*, 83(196), 31–38. <http://doi.org/10.15446/dyna.v83n196.56606>
 - Cepeda Carrión, G., & Roldan Salgueiro, J. L. (2004). Aplicando en la práctica la técnica PLS en administración de empresas. In *Congreso de*

- la ACEDE (pp. 1–30). Murcia-España: ciberconta.unizar.es. Retrieved from <http://ciberconta.unizar.es/doctorado/PLSGabrielCepeda.pdf>
- Chang, C.-L., & Hsu, M.-Y. (2009). The study that applies artificial intelligence and logistic regression for assistance in differential diagnostic of pancreatic cancer. *Expert Systems with Applications*, 36(7), 10663–10672. <http://doi.org/10.1016/j.eswa.2009.02.046>
 - Chin, W. W. (1998a). Issues and Opinion on Structural Equation Modeling. *MIS Quarterly*, 22(1), vii–xvi. <http://doi.org/Editorial>
 - Chin, W. W. (1998b). The Partial Least Squares Approach to Structural Modeling. In *Modern Methods for Business Research* (pp. 295–336).
 - Chin, W. W., Marcolin, B. L., & Newted, P. R. (2003). A Partial least Squares Latent Variable Modeling Approach For Measuring Interaction Effects: Results From a Monte Carlo Simulation Study and Voice Mail Emotion/Adoption Study. In J. I. DeGross, S. Jarvenpaa, & A. Srinivasan (Eds.), *Proceedings of the Seventeenth International Conference on Information Systems* (pp. 21–41). Cleveland- USA. <http://doi.org/10.1287/isre.14.2.189.16018>
 - Cho, S.-R., & Lee, S. (2015). A prediction method of ice breaking resistance using a multiple regression analysis. *International Journal of Naval Architecture and Ocean Engineering*, 7(4), 708–719. <http://doi.org/10.1515/ijnaoe-2015-0050>
 - Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Second). Lawrence Erlbaum.
 - Collins, L. M., & Lanza, S. T. (2010). LATENT CLASS AND LATENT.
 - Coltman, T., Devinney, T. M., Midgley, D. F., & Venaiik, S. (2008). Formative versus reflective measurement models : Two applications of formative measurement. *Journal of Business Research*, 61(12), 1250–1262. <http://doi.org/10.1016/j.jbusres.2008.01.013>
 - Components, O., & Variables, L. (2014). PLS-SEM: Looking Back and Moving Forward. *Long Range Planning*, 47, 132–137. <http://doi.org/10.1016/j.lrp.2014.02.008>
 - Corsini, A., De Propris, L., Feudo, S., & Stefanato, M. (2015). Assessment of a Diagnostic Procedure for the Monitoring and Control of Industrial Processes. *Energy Procedia*, 75, 1772–1778. <http://doi.org/10.1016/j.egypro.2015.07.467>
 - Cuadras, C. M. (2014). *Nuevos Métodos de Análisis Multivariante*. (CMC Editions, Ed.) CMC Editions (Primera). Barcelona - España: CMC Editions. <http://doi.org/10.1017/CBO9781107415324.004>
 - Cupani, M. (2012). Análisis de Ecuaciones Estructurales: conceptos, etapas de desarrollo y un ejemplo de aplicación. *Revista Tesis*.

- Das, S., & Sun, X. (2016). Association knowledge for fatal run-off-road crashes by Multiple Correspondence Analysis. *IATSS Research*, 39(2), 146–155. <http://doi.org/10.1016/j.iatssr.2015.07.001>
- De la Fuente, F. . (2011). Análisis de Correspondencias Santiago de la Fuente Fernández. *Fac. Ciencias Económicas Y Empresariales*, 58.
- Demongeot, J., Hamie, A., Laksaci, A., & Rachdi, M. (2016). Relative-error prediction in nonparametric functional statistics: Theory and practice. *Journal of Multivariate Analysis*, 146, 261–268. <http://doi.org/10.1016/j.jmva.2015.09.019>
- Díaz-García, J. A. (2011). On generalized multivariate analysis of variance. *Brazilian Journal of Probability and Statistics*, 25(1), 1–13. <http://doi.org/10.1214/09-BJPS107>
- Dijkstra, T. K., & Henseler, J. (2015a). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics and Data Analysis*, 81, 10–23. <http://doi.org/10.1016/j.csda.2014.07.008>
- Dijkstra, T. K., & Henseler, J. (2015b). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics and Data Analysis*, 81, 10–23. <http://doi.org/10.1016/j.csda.2014.07.008>
- Dobrzański, L. A., Trzaska, J., & Dobrzańska-Danikiewicz, A. D. (2014). 2.09 - Use of Neural Networks and Artificial Intelligence Tools for Modeling, Characterization, and Forecasting in Material Engineering. In S. Hashmi, G. F. Batalha, C. J. Van Tyne, & B. Yilbas (Eds.), *Comprehensive Materials Processing* (pp. 161–198). Oxford: Elsevier.
- Domínguez, C., & Jesús, A. (2006). SEM vs . PLS : UN ENFOQUE BASADO EN LA PRACTICA. In *IV Congreso de Metodología de Encuestas* (pp. 57–66). Pamplona- España.
- Esquivel Guerrero, J., Contreras Cortes, F., Molina González, F., & Capel Martínez, J. (1991). Una aplicación de la Teoría de Información al análisis de datos definidos mediante variables cualitativas multi-estado: medidas de similaridad y análisis cluster. *Complutum*, (1), 53–64. Retrieved from <http://www.bcin.ca/Interface/openbcin.cgi?submit=submit&Chinkey=161889>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. <http://doi.org/10.1037/1082-989X.4.3.272>
- Falk, R. F., & Miller, N. B. (1992). *A Primer for Soft Modeling.pdf*.
- Fan, M., Shibata, H., & Chen, L. (2017). Environmental and economic risks assessment under climate changes for three land uses scenarios analysis across Teshio watershed, northernmost of Japan. *Science of The Total Environment*, 599–600, 451–463. <http://doi.org/10.1016/j>

scitotenv.2017.05.010

- Findik, D., & Kunçay, C. O. Ş. (2009). A model for instructors adoption of learning management systems : empirical validation in higher education context. Middle East Technical University.
- Fischer, B., Roth, V., & Buhmann, J. M. (2007). Time-series alignment by non-negative multiple generalized canonical correlation analysis. *BMC Bioinformatics*, 8 Suppl 10, S4. <http://doi.org/10.1186/1471-2105-8-S10-S4>
- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39–50. <http://doi.org/10.2307/3151312>
- Fox, S., & Hammond, S. (2017). Investigating the multivariate relationship between impulsivity and psychopathy using canonical correlation analysis. *Personality and Individual Differences*, 111, 187–192. <http://doi.org/10.1016/j.paid.2017.02.025>
- Frías-Navarro, D., & Soler, M. P. (2012). Prácticas Del Análisis Factorial Exploratorio (Afe) En La Investigación Sobre Conducta Del Consumidor Y Marketing Exploratory Factor Analysis (Efa) in Consumer Behavior and Marketing Research. *Suma Psicológica*, 19, 47–58. <http://doi.org/ISSN 0121-4381>
- García, M. (2011). Análisis causal con ecuaciones estructurales de la satisfacción ciudadana con los servicios municipales. Universidad Santiago de Compostela.
- Garson, G. D. (2016). *Partial Least Squares: Regression & Structural Equation Models* (2016th ed.). Asheboro - USA: Statistical Associates Publishing.
- Gefen, D., & Straub, D. (2005). A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. *Communications of the Association for Information Systems*, 16(5), 91–109. <http://doi.org/Article>
- Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural Equation Modeling and Regression : Guidelines for Research Practice. *Structural Equation Modeling*, 4(August), 7. <http://doi.org/10.1.1.25.781>
- Gold, A. H., Malthora, A., & Segars, A. H. (2001). Knowledge Management: An Organizational Capabilities Perspective. *Journal of Management Information Systems*, 18(1), 185–214. <http://doi.org/10.1002/ceat.201000522>
- Gomez, R., & Watson, S. D. (2017). Confirmatory Factor Analysis of the Combined Social Phobia Scale and Social Interaction Anxiety Scale: Support for a Bifactor Model. *Frontiers in Psychology*, 8(February), 1–13. <http://doi.org/10.3389/fpsyg.2017.00070>

- Gómez Suárez, M. (2002). EL ANALISIS CLUSTER EN INVESTIGACIÓN DE MARKETING: METODOLOGIA Y CRITIC. *ESIC Market*, (113), 9–34.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). Partial least squares structural equation modeling (PLS-SEM). *Sage Publisher*, 26(2), 106–121. <http://doi.org/10.1108/EBR-10-2013-0128>
- Hair, J., Black, W., Babin, B., & Anderson, R. (2010). *Multivariate Data Analysis*. (Pearson, Ed.) (Seventh).
- Hair, J. F., Black, W. C., & Babin, B. J. (2010). *Multivariate Data Analysis: A Global Perspective*. Pearson Education.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. *The Journal of Marketing Theory and Practice*, 19(2), 139–152. <http://doi.org/10.2753/MTP1069-6679190202>
- Hair, Anderson, R., Tatham, R., & Black, W. (2008). *Análisis Multivariante*. (A. Otero, Ed.) (Quinta). Madrid-España: Prentice Hall Iberia.
- Handscomb, L., Hall, D. A., Hoare, D. J., & Shorter, G. W. (2016). Confirmatory factor analysis of Clinical Outcomes in Routine Evaluation (CORE-OM) used as a measure of emotional distress in people with tinnitus. *Health and Quality of Life Outcomes*, 14(1), 124. <http://doi.org/10.1186/s12955-016-0524-5>
- Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS Path Modeling in New Technology Research : Updated Guidelines. *Industrial Management & Data Systems*, 116(1), 2–20. <http://doi.org/10.1108/IMDS-09-2015-0382>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review*, 33(3), 405–431. <http://doi.org/10.1108/IMR-09-2014-0304>
- Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, 20(2009), 277–319. [http://doi.org/10.1016/0167-8116\(92\)90003-4](http://doi.org/10.1016/0167-8116(92)90003-4)
- Hernández, R., Fernández-Collado, C., & Baptista, P. (2006). *Metodología de la Investigación*. (M. A. Castellanos Toledo, Ed.) (Cuarta). México - México: McGraw-Hill.
- Höök, K., & Löwgren, J. (2012). Strong concepts. *ACM Transactions on Computer-Human Interaction*, 19(3), 1–18. <http://doi.org/10.1145/2362364.2362371>
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424–453.

- Hwang, T., & Park, T. (2009). Identification of differentially expressed sub-networks based on multivariate ANOVA scoring method. *BMC Bioinformatics*, 10(1), 128. <http://doi.org/10.1186/1471-2105-10-128>
- Jończyk, J. A., & Olszewska, A. M. (2016). The Use of Correspondence Analysis in Assessing the Antecedents of Innovativeness in Public Hospitals. *Studies in Logic, Grammar and Rhetoric*, 47(1), 85–101. <http://doi.org/10.1515/slgr-2016-0048>
- Kaufmann, L., & Gaeckler, J. (2015). A structured review of partial least squares in supply chain management research. *Journal of Purchasing and Supply Management*, 21(4), 259–272. <http://doi.org/10.1016/j.pursup.2015.04.005>
- Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling. Analysis* (Vol. 77). New York- USA. <http://doi.org/10.1038/156278a0>
- Kraft, P., Bauman, L., Yuan, J. Y., Horvath, S., & Study, F. H. (2003). Multivariate variance-components analysis of longitudinal blood pressure measurements from the Framingham Heart Study. *BMC Genetics*, 4 Suppl 1, S55. <http://doi.org/10.1186/1471-2156-4-S1-S55>
- Lay, J.-G., Chen, Y.-W., & Chi, Y.-L. (2013). GIS Adoption Among Senior High School Geography Teachers in Taiwan. *Journal of Geography*, 112(3), 120–130. <http://doi.org/10.1080/00221341.2012.682226>
- LévyMangin, J. pierre, & Varela Mallou, J. (2006). *Modelización con Estructuras de Covarianzas en Ciencias Sociales*. (C. Seco, Ed.) (Gesbiblo S). España. Retrieved from <https://books.google.com.ec/books?id=WfC1TGVJBgC&pg=PA190&lpg=PA190&dq=que+son+los+diagramas+path&source=bl&ots=fpF6XR8JXr&sig=-zYxhHa0ZGJGoMpRgkQ0oTtxXvI&hl=es-419&sa=X&ved=0ahUKEwjD4ei1o4TLAhXBQiyKHboaDYUQ6AEIJAC#v=onepage&q&f=false>
- Leyva, O., & Olague, J. (2014). *Métodos y Técnicas Cualitativas y Cuantitativas Aplicables a la Investigación en Ciencias Sociales*. In TIRANT HUMANIDADES MÉXICO (Ed.), *Métodos y Técnicas Cualitativas y Cuantitativas Aplicables a la Investigación en Ciencias Sociales* (pp. 480–497). Monterrey- México: TIRANT HUMANIDADES MÉXICO.
- Lin, W.-S. (2012). Perceived fit and satisfaction on web learning performance: IS continuance intention and task-technology fit perspectives. *International Journal of Human-Computer Studies*, 70(7), 498–507. <http://doi.org/10.1016/j.ijhcs.2012.01.006>
- Linzer, D. A., & Lewis, J. B. (2011). polCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software*, 42(10), 1–29.
- Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data With Missing Values. *Journal of the American Statistical*

- Association, 83(404), 1198–1202.
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151–1169. <http://doi.org/10.6018/analesps.30.3.199361>
 - Lu, Z., & Liang, Z. (2016). *A Complete Subspace Analysis of Linear Discriminant Analysis and Its Robust Implementation*, 2016(Mmc).
 - Magidson, J., & Vermunt, J. K. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research*, 20.
 - Magidson, J., & Vermunt, J. K. (2005). A nontechnical introduction to latent class models. *DMA Research Council Journal*, 1–15.
 - Manuel Batista-Foguet, J., Coenders, G., & Alonso, J. (2004). Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud. *Medicina Clínica*, 122(Supl.1), 21–27. <http://doi.org/10.1157/13057542>
 - Manzano, A., & Zamora, S. (2009). Sistema de ecuaciones estructurales: una herramienta de investigación. México- México: CENEVAL.
 - Martínez, M. Á. (2014). *Bioestadística amigable* (Díaz De Sa).
 - Mbarek, R., & Zadem, F. (2013). The examination of factors affecting e-learning effectiveness. *International Journal of Innovation and Applied Studies*, 2(4), 423–435. <http://doi.org/2028-9324>
 - Miglioretti, D. L. (2003). Latent transition regression for mixed outcomes. *Biometrics*, 59(3), 710–720. <http://doi.org/10.1111/1541-0420.00082>
 - Ministerio de Turismo. (2015). *Proyecto Ecuador potencia turística*. Quito - Ecuador.
 - Morales Vallejo, P. (2011). Guía para construir cuestionarios y escalas de actitudes. *Revista Chilena de Obstetricia Y Ginecología*, 9(6), 80. <http://doi.org/10.4067/S0717-75262010000600009>
 - Morowatisharifabad, M. A., Khankolabi, M., Mozaffari-Khosravi, H., Fallahzade, H., Gerami, M. H., & Salehi-Abargouei, A. (2017). Parenting style, parental feeding practices and children's nutritional status in authoritative parenting style model: A structural equation modeling. *Iranian Red Crescent Medical Journal*, 19(3). <http://doi.org/10.5812/ircmj.41401>
 - Motofuji, Y., Saito, A., Koike, M., Kodera, Y., Maeda, T., & Komatsu, H. (2012). Potential of classification of cancer by multiple discriminant analysis for relationship between cancer and expression of human cellular phosphoprotein. *Biomedical Research (Tokyo, Japan)*, 33(2), 139–43. <http://doi.org/10.2220/biomedres.33.139>
 - Naciones Unidas. (1998). *Turismo sostenible: contribución del turismo al*

crecimiento económico y al desarrollo sostenible (Vol. 51048).

- Namakforoosh, M. N. (2005). Metodología de la investigación. (GRUPO NORIEGA, Ed.) (LIMUSA S.A). Mexico. Retrieved from https://books.google.com/ec/books?id=ZEJ7-0hmvhwC&pg=PA67&dq=variables+dependientes&hl=es-419&sa=X&redir_esc=y#v=onepage&q&f=false
- Okimoto, G., Zeinalzadeh, A., Wenska, T., Loomis, M., Nation, J. B., Fabre, T., ... Kwee, S. (2016). Joint analysis of multiple high-dimensional data types using sparse matrix approximations of rank-1 with applications to ovarian and liver cancer. *BioData Mining*, 9(1), 24. <http://doi.org/10.1186/s13040-016-0103-7>
- Palavecinos, B. S., Espinoza, R. C., Mella, H. de la F., & Rosso, D. C. (2016). Analysis of information disclosure for chilean case using multiple correspondence methodology. *Análisis de Revelación de La Información Para El Caso Chileno Usandola Metodología de Correspondencias Múltiples.*, 24(2), 327–336. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=114708779&site=eds-live&authtype=uid>
- Pan, D.-D., Xiong, W.-J., Zhou, J.-Y., Pan, Y., Zhou, G.-L., & Fung, W.-K. (2013). Robust Joint Analysis with Data Fusion in Two-Stage Quantitative Trait Genome-Wide Association Studies. *Computational and Mathematical Methods in Medicine*, 843563. <http://doi.org/10.1155/2013/843563>
- Pardo, C. E., & del Campo, P. C. (2007). Combinación de métodos factoriales y de análisis de conglomerados en R: El paquete factoclass. *Revista Colombiana de Estadística*, 30(2), 231–245.
- Patton, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110, 4–18. <http://doi.org/10.1016/j.jmva.2012.02.021>
- Pedrosa, V. B., Groeneveld, E., Eler, J. P., & Ferraz, J. B. S. (2014). Comparison of bivariate and multivariate joint analyses on the selection loss of beef cattle. *Genetics and Molecular Research*, 13(2), 4036–4045. <http://doi.org/10.4238/2014.May.23.14>
- Pérez, E. R., & Medrano, L. (2010). Análisis factorial exploratorio: Bases conceptuales y metodológicas. *Revista Argentina de Ciencias Del Comportamiento*, 2(1889), 58–66. Retrieved from http://www.academia.edu/12607324/Revista_Argentina_de_Ciencias_del_Comportamiento_RACC_Análisis_Factorial_Exploratorio_Bases_Conceptuales_y_Metodológicas_Artículo_de_Revisión
- Pérez Hernández, C. C., Lara Gómez, G., & Gómez Hernández, D. (2017). Evolución de la capacidad tecnológica en México. Aplicación del análisis estadístico multivariante de cluster. *Contaduría Y Administración*, 62(2),

- 505–527. <http://doi.org/10.1016/j.cya.2017.01.002>
- Peruchi, R. S., Paiva, A. P., Balestrassi, P. P., Ferreira, J. R., & Sawhney, R. (2014). Weighted approach for multivariate analysis of variance in measurement system analysis. *Precision Engineering*, 38(3), 651–658. <http://doi.org/10.1016/j.precisioneng.2014.03.001>
 - Qureshi, I., & Compeau, D. (2009). Assessing between-group differences in information systems research: A comparison of covariance- and component-based SEM. *Management Information Systems Quarterly*, 33(1), 197–214.
 - Ramirez-Anormaliza, R., Llinàs-Audet, X., & Sabaté, F. (2016). THE ACCEPTANCE AND USE OF THE E-LEARNING SYSTEMS AMONG THE UNIVERSITY TEACHERS IN ECUADOR. In Iated (Ed.), *EDULEARN16 Conference 4th-6th July 2016* (pp. 3666–3674). Barcelona, Spain. Retrieved from <https://library.iated.org/view/RAMIREZANORMALIZA2016ACC>
 - Ramirez-Anormaliza, R., Sabaté, F., Llinàs-Audet, X., & Lordan, O. (2017). Acceptance and use e-learning systems by ungraduate students of Ecuador: The case of a state University. *Intangible Capital*.
 - Ramirez Anormaliza, R. (2013). Evaluación de los sistemas e-learning: estudio de las publicaciones realizadas en la web of knowledge, 1–20.
 - Reguera-Alvarado, N., Blanco-Oliver, A., & Martín-Ruiz, D. (2016). Testing the predictive power of PLS through cross-validation in banking. *Journal of Business Research*, 69(10), 4685–4693. <http://doi.org/10.1016/j.jbusres.2016.04.016>
 - Rhee, J.-K., Joung, J.-G., Chang, J.-H., Fei, Z., & Zhang, B.-T. (2009). Identification of cell cycle-related regulatory motifs using a kernel canonical correlation analysis. *BMC Genomics*, 10 Suppl 3, S29. <http://doi.org/10.1186/1471-2164-10-S3-S29>
 - Ringle, C. M., Wende, S., & Becker, J. M. (Eds.). (2015). *SmartPLS 3*. Boenningstedt: SmartPLS GmbH. Retrieved from <http://www.smartpls.com>.
 - Rodrigues, L., Grave, R., de Oliveira, J. M., & Nogueira, C. (2016). Study on homophobic bullying in Portugal using Multiple Correspondence Analysis (MCA). *Revista Latinoamericana de Psicología*, 48(3), 191–200. <http://doi.org/10.1016/j.rlp.2016.04.001>
 - Rosseel, Y. (2012). lavaan : An R Package for Structural Equation. *Journal of Statistical Software*, 48(2).
 - Ryu, E., & Cheong, J. (2017). Comparing Indirect Effects in Different Groups in Single-Group and Multi-Group Structural Equation Models. *Frontiers in Psychology*, 8(May), 1–14. <http://doi.org/10.3389/fpsyg.2017.00747>
 - Sallan, J. M., Fernandez, V., Simo, P., Lordan, O., & Gonzalez-Prieto, D.

(2012). Análisis de modelos de ecuaciones estructurales mediante el paquete lavaan. In 6th International Conference on Industrial Engineering and Industrial Management. XVI Congreso de Ingeniería de Organización. (pp. 951–958). Vigo- España. Retrieved from <http://www.adingor.es/ocs/index.php?conference=cio2012&schedConf=CIO2012>

- Sanjuán Suárez, P., Pérez García, A. M., & Bermúdez Moreno, J. (2000). Escala de autoeficacia general: Datos psicométricos de la adaptación para población española. *Psicothema*, 12(SUPPL. 2), 509–513. <http://doi.org/ISSN 0214-9915>
- Sarstedt, M., Ringle, C. M., Smith, D., Reams, R., & Hair, J. F. (2014). Partial least squares structural equation modeling (PLS-SEM): A useful tool for family business researchers. *Journal of Family Business Strategy*, 5(1), 105–115. <http://doi.org/10.1016/j.jfbs.2014.01.002>
- Schubring, S., Lorscheid, I., Meyer, M., & Ringle, C. M. (2016). The PLS agent : Predictive modeling with PLS-SEM and agent-based simulation. *Journal of Business Research*, 69(10), 4604–4612. <http://doi.org/10.1016/j.jbusres.2016.03.052>
- Schumacker, R. E., & Lomax, R. G. (2010). *A Beginner's Guide to Structural Equation Modeling: Third Edition.* (Taylor & Francis Group, Ed.) (3rd ed.). New York - USA: Routledge. Retrieved from <http://books.google.com/books?id=58pWPxWPC90C&pgis=1>
- Seward, K., Wolfenden, L., Wiggers, J., Finch, M., Wyse, R., Oldmeadow, C., ... Yoong, S. L. (2017). Measuring implementation behaviour of menu guidelines in the childcare setting: confirmatory factor analysis of a theoretical domains framework questionnaire (TDFQ). *International Journal of Behavioral Nutrition and Physical Activity*, 14(1), 45. <http://doi.org/10.1186/s12966-017-0499-6>
- Sha, Q., Zhang, Z., & Zhang, S. (2011). Joint analysis for genome-wide association studies in family-based designs. *PLoS ONE*, 6(7). <http://doi.org/10.1371/journal.pone.0021957>
- Shah, T., Patel, M. N., Nath, S., & Menon, S. K. (2016). Determination of sex using cephalo-facial dimensions by discriminant function and logistic regression equations. *Egyptian Journal of Forensic Sciences*, 6(2), 114–119. <http://doi.org/10.1016/j.ejfs.2015.03.004>
- Sharma, N., Bajpai, A., & Litoriya, R. (2012). Comparison the various clustering algorithms of weka tools. *International Journal of Emerging Technology and Advanced Engineering*, 2(5), 73–80.
- Sturn, A., Quackenbush, J., & Trajanoski, Z. (2002). Genesis: cluster analysis of microarray data. *Bioinformatics (Oxford, England)*, 18(1), 207–8. <http://doi.org/10.1093/bioinformatics/18.1.207>

- Tarhini, A., Hone, K., & Liu, X. (2013). Factors Affecting Students' Acceptance of e-Learning Environments in Developing Countries : A Structural Equation Modeling Approach. *International Journal of Information and Education Technology*, 3(1), 54–59. <http://doi.org/10.7763/IJJET.2013.V3.233>
- Tekai, F. (2014). Genome Data Exploration Using Correspondence Analysis. *Bioinformatics and Biology Insights*, 1(7), 7523. <http://doi.org/10.4137/BBI.S39614.TYPE>
- Torrado-Fonseca, M., & Berlanga-Silvente, V. (2013). Análisis Discriminante mediante SPSS. *REIRE Revista d'Innovació i Recerca En Educació*, 6(2), 150–166. <http://doi.org/10.1344/reire2013.6.26210//>
- Urbach, N., & Ahlemann, F. (2010). Structural Equation Modeling in Information Systems Research Using Partial Least Squares. *Journal of Information Technology Theory and...*, 11(2), 5–40. Retrieved from [http://iris.ebs.edu/accessdb/WWW/iris_pub_web.nsf/wwwPubIAuthorSingleEng/A58520D5E3C484DCC1257C50007F8697/\\$file/Urbach et al. 2010 Structural Equation Modeling in IS Research Using PLS.pdf](http://iris.ebs.edu/accessdb/WWW/iris_pub_web.nsf/wwwPubIAuthorSingleEng/A58520D5E3C484DCC1257C50007F8697/$file/Urbach_et_al_2010_Structural_Equation_Modeling_in_IS_Research_Using_PLS.pdf)
- Vázquez-Cano, E., & García Sevillano, M. L. (2015). Analysis of risks in a Learning Management System : A case study in the Spanish National University of Distance Education (UNED). *Journal of New Approaches in Educational Research*, 4(1), 62–68. <http://doi.org/10.7821/naer.2015.1.107>
- Veldhuis, J. D., & Johnson, M. L. (1986). Cluster analysis: a simple, versatile, and robust algorithm for endocrine pulse detection. *The American Journal of Physiology*, 250(4 Pt 1), E486-93.
- Verdugo, M. Á. (2008). Metodología en la investigación sobre discapacidad. Introducción al uso de las ecuaciones estructurales.
- Vilà-baños, R., Rubio-Hurtado, M.-J., Berlanga-Silvente, V., & Torrado-Fonseca, M. (2014). Cómo aplicar un cluster jerárquico en SPSS. *REIRE Revista d'Innovació i Recerca En Educació*, 7(1), 113–127.
- Waaijenborg, S., & Zwinderman, A. H. (2009). Sparse canonical correlation analysis for identifying, connecting and completing gene-expression networks. *BMC Bioinformatics*, 10, 315. <http://doi.org/10.1186/1471-2105-10-315>
- Weng, C., & Tsai, C. (2015). Social support as a neglected e-learning motivator affecting trainee ' s decisions of continuous intentions of usage. *Australasian Journal of Educational Technology*, 31(2), 177–192.
- Wold, H. (1974). Causal flows with latent variables. *European Economic Review*, 5(1), 67–86. [http://doi.org/10.1016/0014-2921\(74\)90008-7](http://doi.org/10.1016/0014-2921(74)90008-7)
- Wu, C.-C., & Chiu, J. (2017). Economic evaluation of asymmetric and price range information in gold and general financial markets. *Journal of*

International Money and Finance, 74, 53–68. <http://doi.org/10.1016/j.jimonfin.2017.03.001>

- Wu, P., Presanis, A. M., Bond, H. S., Lau, E. H. Y., Fang, V. J., & Cowling, B. J. (2017). A joint analysis of influenza-associated hospitalizations and mortality in Hong Kong, 1998-2013. *Scientific Reports*, 7(1), 929. <http://doi.org/10.1038/s41598-017-01021-x>
- Xu, L. W. (2014). MANOVA for Nested Designs with Unequal Cell Sizes and Unequal Cell Covariance Matrices. *Journal of Applied Mathematics*, 2014. <http://doi.org/10.1155/2014/649202>
- Ye, X., Kang, Y., Zuo, B., & Zhong, K. (2017). Study of factors affecting warm air spreading distance in impinging jet ventilation rooms using multiple regression analysis. *Building and Environment*, 120, 1–12. <http://doi.org/10.1016/j.buildenv.2017.03.044>
- Yoshida, K., Yoshimoto, J., & Doya, K. (2017). Sparse kernel canonical correlation analysis for discovery of nonlinear interactions in high-dimensional data. *BMC Bioinformatics*, 18(1), 108. <http://doi.org/10.1186/s12859-017-1543-x>
- Yuan, Z., Zheng, X., Zhang, L., & Zhao, G. (2017). Urban Competitiveness Measurement of Chinese Cities Based on a Structural Equation Model. *Sustainability*, 9(4), 666. <http://doi.org/10.3390/su9040666>
- Zhou, Y., & Sun, S. (2015). *Semisupervised Tangent Space Discriminant Analysis*, 2015.
- Ziersch, A. M., Baum, F. E., Macdougall, C., & Putland, C. (2005). Neighbourhood life and social capital: the implications for health. *Social Science & Medicine* 60, 60, 71–86. <http://doi.org/10.1016/j.socscimed.2004.04.027>

“No hay nada repartido de modo más equitativo que la razón: todo el mundo está convencido de tener suficiente”

René Descartes (1596-1650)
Filósofo y matemático francés.

ANEXOS

Anexo 1- 1 Instrumento para evaluar la aceptación de e-learning en profesores

Constructos – ítems		Fuente
<i>Soporte técnico (TS)</i>		
TS1	El sistema e-learning proporciona ayuda cuando hay un problema técnico.	(Arteaga & Duarte, 2010)
TS2	Una línea telefónica está disponible en cualquier momento para atender problemas técnicos con el sistema e-learning.	
TS3	Se pueden hacer preguntas por correo electrónico cuando se presenta un problema técnico con en el sistema e-learning.	
TS4	El sistema e-learning ofrece respuestas a preguntas frecuentes sobre su uso.	
TS5	Se pueden hacer consultas en línea tipo chat con personas de soporte técnico cuando hay algún problema con el sistema e-learning.	
TS6	El personal de soporte del sistema e-learning tiene una predisposición positiva o actitud de ayuda cuando es consultado.	
<i>La autoeficacia computacional (CSE)</i>		
CSE1	Puedo completar las tareas de enseñanza en el sistema e-learning si no hay nadie alrededor para decirme qué hacer a medida que avanzo.	(Mbarek & Zaddem, 2013),
CSE2	Puedo completar las tareas de enseñanza en el sistema e-learning si yo nunca había usado un sistema informático como este antes.	
CSE3	Puedo completar las tareas de enseñanza en el sistema e-learning si tuviera sólo los manuales de referencia del sistema.	(Sanjuán Suárez, Pérez García, & Bermúdez Moreno, 2000)
CSE4	Puedo completar las tareas de enseñanza en el sistema e-learning si he visto a alguien más usar el sistema antes que yo.	
CSE5	Puedo completar las tareas de enseñanza en el sistema e-learning si alguien más me ayuda con una inducción en el sistema.	(Sanjuán Suárez, Pérez García, & Bermúdez Moreno, 2000)

Continua

Anexo 1- 1 Continuación

CSE6	Una formación adicional en manejo de e-learning mejoría mi autoeficacia computacional y con ello el manejo del sistema e-learning.	(Sanjuán Suárez, Pérez García, & Bermúdez Moreno, 2000)
CSE7	Gracias a mis cualidades y recursos puedo superar situaciones imprevistas dentro del manejo del sistema e-learning.	
CSE8	Puedo resolver la mayoría de los problemas que se presentan en el uso del sistema e-learning si me esfuerzo lo necesario.	
CSE9	Si me encuentro en una situación difícil en el manejo del sistema e-learning, generalmente se me ocurre qué debo hacer.	
<i>Influencia social (SI)</i>		
SI1	Los directivos de mi carrera y facultad piensan que debería participar en las actividades de enseñanza basadas en el sistema e-learning.	(Tarhini et al., 2013 ^a)
SI2	Mis colegas profesores de la facultad piensan que yo debería participar en las actividades de enseñanza basadas en el sistema e-learning.	
SI3	Los directivos de mi universidad consideran que debo realizar actividades de enseñanza basadas en el sistema e-learning.	
SI4	En términos generales, me gustaría hacer lo que mi decano cree que debo hacer.	
SI5	Las personas cuyas opiniones valoro, piensan que debo utilizar el sistema e-learning.	
<i>Entretenimiento Percibido (PE)</i>		
PE1	El sistema e-learning ha conseguido que mi estrés por calificar las tareas baje, pues el lo hace de manera inmediata.	(Weng & Tsai, 2015)
PE2	Pienso que usando el sistema e-learning al calificar, la nota se otorga con la misma rigurosidad a todos.	(Weng & Tsai, 2015)
PE3	Me agrada más calificar por el sistema e-learning, que de manera manual.	
PE4	Encuentro el sistema e-learning entretenido.	
PE5	Me gusta usar el sistema e-learning.	

Continua

Anexo 1- 1 Continuación

<i>Satisfacción (S)</i>		
S1	Estoy satisfecho de utilizar el sistema e-learning como herramienta para de aprendizaje.	(Lin, 2012)
S2	El sistema e-learning es eficiente para la construcción del conocimiento.	
S3	El sistema e-learning es eficiente para el intercambio de conocimiento.	
S4	Estoy satisfecho con la interactividad de los cursos que utilizan el sistema e-learning.	
S5	Estoy satisfecho con las diversas opciones de evaluación que ofrece el sistema e-learning.	
S6	Me satisface el uso de los sistemas e-learning, porque contribuye al desarrollo de las competencias transversales, como el trabajo en equipo.	
<i>Utilidad Percibida (PU)</i>		
PU1	En los cursos que imparto con el soporte del sistema e-learning, mejora mi rendimiento.	(Lay, Chen, & Chi, 2013)
PU2	En los cursos que imparto con el soporte del sistema e-learning, el aprendizaje es eficaz.	
PU3	En los cursos que imparto con el soporte del sistema e-learning, se hace más fácil la enseñanza.	
PU4	En los cursos que imparto con la ayuda del sistema e-learning, el sistema es de utilidad para la enseñanza.	
PU5	Es conveniente en la enseñanza, el uso del sistema e-learning.	
<i>Facilidad de Uso Percibido (PEOU)</i>		
PEOU1	Aprender a operar el sistema e-learning resulta muy fácil.	(Lay et al., 2013)
PEOU2	La interacción con el sistema e-learning es clara y comprensible.	
PEOU3	Encuentro que el sistema e-learning es flexible para interactuar con el.	
PEOU4	Sería fácil ser competente en el uso del sistema e-learning.	
PEOU5	Encuentro que el sistema e-learning es fácil de usar.	

Continua

Anexo 1- 1 Continuación

<i>Intención hacia el uso (BI)</i>	
BI1	Voy a utilizar el sistema e-learning en todas mis asignaturas que permitan dictar clases con este soporte.
BI2	Es importante usar el sistema e-learning, recomendaría su utilización.
BI3	Voy a modificar las actividades docentes de mis asignaturas para aprovechar las capacidades del sistema e-learning.
BI4	Motivaré mi participación y la de los estudiantes en el sistema e-learning, ofreciendo alguna actividad basada en participar en el foro de la asignatura.
BI5	Me gustaría utilizar el sistema e-learning en el futuro si tuviera la oportunidad.
<i>Uso del Sistema (SU)</i>	
SU1	En término de horas semanales, tiendo a utilizar el sistema e-learning el tiempo necesario para gestionar las actividades que propongo en mis asignaturas.
SU2	Me paso tiempo interactuando dentro del sistema e-learning.
SU3	Me involucro con el sistema e-learning.
SU4	Ingreso al sistema e-learning para interactuar con los curso que dicto, al menos una vez al día.
SU5	Me conecto con frecuencia para participar en aquellas actividades interactivas (foros) que he propuesto en el sistema e-learning.
SU6	Me conecto con frecuencia al sistema e-learning para visualizar el grado de participación y avance de los estudiantes.

Nota: El instrumento fue elaborado adaptando los ítems utilizados en estudios definidos en la columna fuente.

Anexo 1- 2 Instalación de R y R-Studio

R es un software libre de GNU y se podría definir desde dos puntos de vista, por una parte es un lenguaje de programación y por otra un entorno de trabajo, estando ambos orientados al cálculo estadístico y a la generación de gráficas.

R como lenguaje de programación proporciona una amplia variedad de técnicas y recursos para el trabajo con gráficas y análisis estadístico y, a su vez, es altamente ampliable. Se distribuye con licencia GNU GPL v2 y está disponible para distintos sistemas operativos de tipo Unix y similares (FreeBSD y Linux), Windows y Mac OS.

R como entorno de trabajo se entiende como un sistema totalmente planificado y coherente y no una acumulación incremental de herramientas muy específicas y poco flexibles, como es frecuentemente el caso con otro software de análisis de datos. En este caso el entorno de trabajo R nos proporciona una serie de utilidades para manipulación de datos, cálculo y representación gráfica.

RStudio es un entorno de desarrollo integrado (IDE) para R. Es un software libre con licencia GPLv3 y se puede ejecutar sobre distintas plataformas (Windows, Mac, or Linux) o incluso desde la web usando RStudio Server.

Descargar R

<https://cran.r-project.org/bin/windows/base/>

Descargar r studio

<https://www.rstudio.com/products/rstudio/download2/>

Primero debe instalar R, luego de aquello deberá instalar RStudio