

La presente obra reúne una serie de ensayos elaborados por los estudiantes del doctorado de Ciencias de la Administración del Centro Universitario de Ciencias Económico Administrativas de la Universidad de Guadalajara, basados en lo aprendido en la asignatura de Investigación Cuantitativa I. Estos ensayos se orientan, en principio, a realizar un ejercicio de disertación que contribuya a la elaboración de su tesis o se configure como una contribución a la materia, resaltando la pertinencia de su redacción, conceptualizando y proponiendo los modelos revisados como desarrollo de su disertación. Lo anterior sirve como base para realizar la discusión que permite aclarar la contribución esperada, para finalmente concluir en puntos esenciales que sirvan al lector y al expositor, para estudios posteriores.

Es deseo de la coordinación del presente trabajo, que este contribuya al ánimo del lector por conocer los proyectos que se desarrollan e informar de las oportunidades que se muestran, con el fin de dar seguimiento a la evolución de los mismos en su estancia en el posgrado.

**ISBN: 978-607-98782-3-8**

**Centro Universitario de Ciencias  
Económico Administrativas**

**Juan Mejía Trejo**  
*Coordinador*

**ANÁLISIS MULTIVARIANTE CON ENFOQUE DEPENDIENTE EN LAS CIENCIAS  
DE LA ADMINISTRACIÓN COMO BASE PARA LA INNOVACIÓN**

**ENSAYOS 2018**

**ENSAYOS 2018**  
**ANÁLISIS MULTIVARIANTE CON**  
**ENFOQUE DEPENDIENTE**  
**EN LAS CIENCIAS DE LA**  
**ADMINISTRACIÓN COMO BASE**  
**PARA LA INNOVACIÓN**

**Juan Mejía Trejo**  
*Coordinador*

**UNIVERSIDAD DE GUADALAJARA**

# **ENSAYOS 2018**

## **ANÁLISIS MULTIVARIANTE CON ENFOQUE DEPENDIENTE EN LAS CIENCIAS DE LA ADMINISTRACIÓN COMO BASE PARA LA INNOVACIÓN**



**ENSAYOS 2018**

**ANÁLISIS MULTIVARIANTE  
CON ENFOQUE DEPENDIENTE  
EN LAS CIENCIAS DE LA  
ADMINISTRACIÓN COMO BASE  
PARA LA INNOVACIÓN**

**Juan Mejía Trejo**  
*Coordinador*

UNIVERSIDAD DE GUADALAJARA  
2019

Esta obra fue sometido a un proceso de dictamen por pares de acuerdo con las normas establecidas por el comité editorial del Centro Universitario de Ciencias Económico Administrativas de la Universidad de Guadalajara.

Primera edición 2019

D.R. © Universidad de Guadalajara  
Centro Universitario de Ciencias Económico Administrativas  
Periférico Nte. 799, núcleo universitario Los Belenes  
45100, Zapopan, Jalisco

**ISBN: 978-607-98782-3-8**

Impreso y hecho en México  
*Printed and made in Mexico*

# Contenido

Introducción .....	7
HERRAMIENTAS DE ANÁLISIS MULTIVARIANTE PREDICTO Y MINERÍA DE DATOS CON SPSS MODELER Y STATISTIC: ESTUDIO COMPARATIVO.	
Pascuala Josefina Cárdenas Salazar .....	11
DIRECTOR DE TESIS: <i>Dr. Juan Mejía Trejo</i>	
LA IMPORTANCIA DE LA TÉCNICA DE REGRESIÓN LINEAL SIMPLE EN EL ÁREA DE LAS CIENCIAS ECONÓMICO-ADMINISTRATIVAS.	
José Rosario Lara Salazar.....	41
DIRECTOR DE TESIS: <i>Dr. Alejandro Campos Sánchez</i>	
SOFTWARE COMO HERRAMIENTA PARA MÉTODOS DE PROSPECTIVA ESTRATÉGICA: MICMAC	
Alba Lucia Moreno Ortiz .....	51
DIRECTOR DE TESIS: <i>Dr. Ariel Vázquez Elorza</i>	
ANÁLISIS COMPARATIVO ENTRE REGRESIÓN LINEAL MÚLTIPLE – MÍNIMOS CUADRADOS PARCIALES Y SU APLICACIÓN EN LAS CIENCIAS ECONOMICO- ADMINISTRATIVAS	
Hugo César Enriquez García.....	61
DIRECTOR DE TESIS: <i>Dr. Ricardo Arechavala Vargas</i>	
REGRESIÓN LINEAL APLICADA: ANÁLISIS DE LA RESPONSABILIDAD SOCIAL EMPRESARIAL	
Itzel Alejandra Lara Manjarrez .....	73
DIRECTOR DE TESIS: <i>Dr. Rogelio Rivera Fernández</i>	

TÉCNICAS DE ANÁLISIS MULTIVARIANTE PARA LA VALIDACIÓN DE UN  
MODELO CONCEPTUAL DE TRANSFORMACIÓN DE ORGANIZACIÓN  
LINEAL A EXPONENCIAL.

Alfredo Aguilar Ruiz..... 85  
DIRECTOR DE TESIS: *Dr. Antonio de Jesús Vizcaino*

CORRELACIÓN DE VARIABLES DE LA COMPETITIVIDAD A PARTIR  
DE LA APLICACIÓN DE ANÁLISIS MULTIVARIABLES DE  
TÉCNICAS DEPENDIENTES (REGRESIÓN LINEAL MÚLTIPLE).

Jovanni Trinidad Saldaña ..... 99  
DIRECTOR DE TESIS: *Dr. Katia Magdalena Lozano Uvario*

EL ANÁLISIS MULTIVARIANTE COMO HERRAMIENTA PARA MEDIR LOS  
PROCESOS DE ADMINISTRACIÓN DE RECURSOS HUMANOS CON LA  
GESTIÓN DEL CONOCIMIENTO Y SU RELACIÓN CON LA INNOVACIÓN.

Julio Ceja Sainz ..... 117  
DIRECTOR DE TESIS: *Dr. Carlos Fong Reynoso*

EL USO DE TÉCNICAS ESTADÍSTICAS MULTIVARIANTES MEDIANTE  
EL ANÁLISIS DISCRIMINANTE, APLICADO EN LOS NEGOCIOS,  
LAS EMPRESAS Y ORGANIZACIONES EN GENERAL.

Luis Alberto Arroyo González ..... 127  
DIRECTOR DE TESIS: *Dr. Guillermo Vázquez Ávila*

INVESTIGACIONES CIENTÍFICAS EN RH: EL DESAFÍO QUE DEBEN  
ENCARAR LOS PROFESIONALES.

MIGUEL ANGEL HERNÁNDEZ GONZÁLEZ ..... 147  
DIRECTOR DE TESIS: *Dr. José Sánchez Gutiérrez*

# *Introducción*

La presente obra, Ensayos 2018, Análisis Multivariante con Enfoque Dependiente en las Ciencias de la Administración como base para la Innovación, pretende reunir una serie de ensayos elaborados por los estudiantes del Doctorado de Ciencias de la Administración (DCA) del Centro Universitario de Ciencias Económico Administrativas (CUCEA) de la Universidad de Guadalajara (UdeG), basados en lo aprendido en la asignatura de Investigación Cuantitativa I.

Dichos ensayos, se orientan en principio a realizar un ejercicio de disertación que refuerce ya sea su tesis o se configure como una contribución a la materia, resaltando la pertinencia de su redacción, conceptualizando y proponiendo los modelos revisados como desarrollo de su disertación. Lo anterior sirve como base para realizar la discusión que permite aclarar la contribución esperada para finalmente, concluir en puntos esenciales que sirvan al lector y al expositor, para estudios posteriores.

Es así, que esta obra se desglosa en diez ensayos, donde la primera obra, *Herramientas de análisis multivariante predictivo y minería de datos con SPSS Modeler y Statistic*, estudio comparativo, representa una excelente oportunidad al lector para que se familiarice con el software SPSS como la herramienta más utilizada en las ciencias sociales y de la salud en cuanto a las herramientas de estadística inferencial, siendo muy valiosa su aportación en cuanto a los alcances determinados de cada versión de SPSS.

La segunda contribución, *La importancia de la técnica de regresión lineal simple en el área de las ciencias económico-administrativas*, es un gran referente de la utilidad que representan las técnicas paramétricas de proyección basadas en la regresión lineal evidenciando su uso actual,



mediante un amplio estudio bibliométrico y su oportunidad de uso en las ciencias económico administrativas.

El tercer ensayo, *Software como herramientas para métodos de prospectiva estratégica: MICMAC*, presenta las justificaciones y bases pertinentes muy valiosas sobre cómo realizar estudios de prospectiva con este software, aclarando los alcances de herramientas informáticas adicionales, que permiten precisar los escenarios futuros posibles, probables y deseables mediante: MACTOR, MORPHOL, MULTIPOLAR y SMIC

Al respecto del trabajo número cuatro, *Análisis comparativo entre regresión lineal múltiple-mínimos cuadrados parciales y su aplicación en las ciencias económico-administrativas*, se resalta su valor en la contribución de realizar comparativos de alcance que permitan a los lectores, conocer de los alcances y los impactos de ambas técnicas al determinarse recomendaciones de uso en el campo, con la técnica de mínimos cuadrados que está cobrando de manera sostenida una mayor importancia.

El quinto ensayo, *Regresión lineal aplicada: análisis de la responsabilidad social empresarial*, muestra su pertinencia al realizar un análisis bibliométrico amplio y profundo de las contribuciones de la técnica de regresión lineal en el ámbito de la responsabilidad social empresarial comparando cuatro modelos relacionados y discutiendo el impacto de las variables descubiertas.

El sexto trabajo, *Técnicas de análisis multivariante para la validación de un modelo conceptual de transformación de organización lineal a exponencial*, revela la importancia de su aportación al verificar modelos previos de organización exponencial en el que se exponen la variables que se han discutido al momento para hacer una propuesta original de un nuevo modelo, basado en los constructos producto de la revisión del estado del arte.

La séptima obra, *Correlación de variables de la competitividad a partir de la aplicación de análisis multivariantes de técnicas dependientes (Regresión Lineal Múltiple)*, hace una disertación sobre los diversos modelos que abordan la competitividad mediante un análisis bibliométrico amplio y profundo que le permite comparar seis modelos con sus variables correspondientes y discutir la propuesta de un modelo original, como consecuencia de ello.

El octavo ensayo, *El análisis multivariante como herramienta para medir los procesos de administración de recursos humanos y con la gestión del conocimiento y su relación con la innovación*, está basado en un análisis bibliométrico extenso y profundo que muestra la oportunidad de aportar en las relaciones que se tienen de los diversos procesos de admi-

nistración de recursos humanos y la gestión del conocimiento en la innovación, mediante una propuesta original, que se basa en modelos previos publicados al respecto.

El noveno trabajo, *El uso de técnicas estadísticas multivariantes mediante el análisis discriminante, aplicado en los negocios, las empresas y organizaciones en general*, cumple con el reto de hacer la discusión de modelos y variables que explican las relaciones que se presentan en los negocios, empresas y organizaciones desde el punto de vista financiero, de ahí el valor que representa la propuesta de este ensayo.

El décimo ensayo, *Investigaciones científicas en RH: El desafío que los profesionales deben encarar*, informa de la oportunidad que se tiene en el área de recursos humanos de proponer modelos que permian descubrir los principales determinantes que expliquen las diversas relaciones a las que se someten los sujetos, a partir del análisis multivariante.

Es deseo de la coordinación del presente trabajo, que este contribuya al ánimo del lector por conocer los proyectos que se desarrollan e informar de las oportunidades que se muestran, con el fin de dar seguimiento a la evolución de los mismos en su estancia en el posgrado.

*Dr. Juan Mejía Trejo*  
Coordinador del DCA CUCEA  
Universidad de Guadalajara



# CAPÍTULO 1

## *Herramientas de análisis multivariante predictivo y minería de datos con SPSS Modeler y Statistics: Estudio comparativo*

*Pascuala Josefina Cárdenas Salazar*

DIRECTOR DE TESIS  
*Juan Mejía Trejo*

**Palabras clave:** IBM SPSS Modeler, IBM SPSS Statistics, análisis multivariante, minería de datos.

### **Introducción**

Una de las grandes limitaciones a la que los investigadores se enfrentan en la búsqueda de modelos predictores es el empleo del software adecuado para el análisis de datos. De acuerdo con León, Castellanos y Vargas (2006) uno de los problemas en la producción científica en su tarea de soportar decisiones empresariales que les permitan anticiparse a cambios, es que no se pone atención al software empleado y se terminan haciendo modelos que no predicen los fenómenos. El empleo del software posibilita a los investigadores de herramientas para la resolución de problemas organizacionales. No obstante, el tipo de programa empleado en la búsqueda de modelos predictivos puede relacionarse con la naturaleza del problema a tratar, por lo que aquí se analizan dos.

Uno de los planteamientos de la búsqueda de modelos predictivos es a partir de modelos estadísticos que previo a un análisis, se establecen

relaciones, es decir, cuando se parte de abstracciones de la realidad representada en modelos que representan los factores y sus relaciones para poder cuantificarlos. Dicho de otra manera, cuando la investigación parte de una hipótesis y se requiere de herramientas de análisis multivariante para cumplir con una serie de pruebas para la aceptación de los supuestos, la consolidación de la predicción y la posibilidad de la generalización a otras poblaciones a través del estudio en la muestra.

Otra propuesta para lograr modelos predictivos se enfoca partir de los datos existentes o históricos de las organizaciones. Por ejemplo, en varias partes del mundo se generan bases de datos pero no se hace uso óptimo de herramientas necesarias para presentar resultados que den solución a las problemáticas de los diferentes sectores de la sociedad (Castañeda, Cabrera, Navarro y de Bries, 2010). Por ello, es necesario extraer conocimiento a través de datos existentes, de manera específica se requiere de una minería de datos que permita conocer patrones antes ocultos en los datos (e.g. Aranda y Sotongo, 2013; Lobaina & Suárez, 2018). Por lo que se requiere de herramientas que permitan el análisis de diversas fuentes de datos y optimización de la información existente.

Con base en lo expuesto, es que se considera necesario revisar estas dos formas de construir o consolidar modelos de predicción así como analizar los softwares enfocados a los modelos de predicción. Para ello se toman como base los programas de la empresa IBM, se trata de SPSS Modeler y SPSS Statistics para descubrir si ¿Es posible el análisis multivariante predictivo a partir de un modelo estadístico con Modeler? o ¿Minería de datos a partir bases de datos preexistentes en Statistics?

De acuerdo a lo anterior es que el presente ensayo tiene como objetivo analizar las características de valor que ofrecen los programas IBM SPSS Modeler y Statistics para atender las incógnitas de investigación y ofrecer alternativas o soluciones óptimas en el establecimiento de modelos de predicción. Así también, conocer en qué circunstancias puede utilizarse uno para el estudio multivariante a partir de un modelo estadístico y, cuál otro, apto para la minería de datos. Para ello el documento está seccionado de la siguiente forma:

En la primera parte se presenta el antecedente de la búsqueda de conocimiento a partir de bases de datos existentes, minería de datos y metodología CRISP-DM, en la segunda, se trata el tema de modelo estadístico y análisis multivariante. En la tercera sección se presentan las características del software IBM SPSS Modeler y en la cuarta, lo relacionado al de Statistics. En la quinta parte se presenta un análisis comparativo de

las funciones que ofrecen los programas en comentario. Por último, se presentan la discusión y las ideas concluyentes del presente ensayo.

## 1. El conocimiento en bases de datos

El descubrimiento del conocimiento de las bases de datos se refiere a una serie de etapas que consiste en el descubrimiento de bases de datos, integración de los mismos, determinar los objetivos, preparación de datos (selección y transformación), minería de datos, evaluación, interpretación y toma de decisiones, siendo la minería de datos una de las partes más importantes dentro del proceso (Aranda y Sotolongo, 2013; Cabena, Hadjinian, Stadler, Verhees y Zanasi, 1998). La minería de datos es uno de los más necesarios en para la optimización de las bases de datos (Lobaina y Suárez, 2018). Este proceso integra análisis de datos y extracción de modelos (Fayyad, 1996). El proceso de la extracción de conocimiento en bases de datos se observa a continuación.

### *Proceso del descubrimiento de conocimiento en bases de datos.*

1. Integración y recopilación de los datos. Fase de extracción para decidir de donde se van a sacar los datos almacén de datos, data warehouse, (Hernández, Ramírez y Ferri, 2004).
2. Determinación de objetivos. Fase de extracción para determinar el problema a resolver y el objetivo a perseguir (Cabena, *et al.*, 1998).
3. Preparación de los datos. (Aranda y Sotolongo, 2013, Cabena, *et al.*, 1998).
  - Preparación de datos. Selección de fuentes de datos.
  - Pre-procesamiento respecto a calidad y determinación de los datos.
  - Transformación de datos: conversión en modelo analítico.
4. Minería de datos.
  - Técnicas de predicción,
  - Técnicas de clasificación (árboles de decisión y reglas de inducción),
  - Técnicas de asociación (correlación),
  - Técnicas de agrupamiento, (Aranda y Sotolongo, 2013).
5. Evaluación e interpretación de datos. Análisis y evaluación de los resultados (Cabena, *et al.*, 1998).
6. Toma de decisiones o difusión y uso del conocimiento (Hernández, Ramírez y Ferri, 2004. Asimilación y aplicación del conocimiento (Cabena, *et al.*, 1998).

## ***Minería de datos***

Como se mencionó párrafos arriba, la minería de datos es uno de los procesos más importantes en la extracción de conocimiento a partir de las bases de datos existentes. De acuerdo a Aranda y Sotongo (2013) es un proceso que permite establecer modelos predictivos con base a los datos existentes, mismos que pueden aplicarse en las diversas ramas del conocimiento (eg. Rama biológica, aplicaciones educativas y financieras, procesos industriales, policiales y políticos. Se trata de actividades mediante las cuales se identifican patrones ocultos (Fayyad, 1996; Lobaina & Suárez, 2018, Cabena, et al., 1998). Además de lo anterior, es una fase exploratoria de datos para identificar relaciones sin un objetivo particular (IBM, 2012). Por tal razón, la minería es el proceso que se toma como base para este análisis y no la extracción de conocimiento en general.

Además, dicho proceso reúne todas las características de la extracción de conocimiento a partir de bases de datos. Dado que se trata de extraer conocimiento no trivial de bases de datos disponibles de diversas fuentes y mediante diversas técnicas de extracción de conocimiento que resulta de las relaciones y dependencias entre los elementos constitutivos de los factores que se analizan ante un problema dado (e.g. Rodríguez *et. al* 2010). De ahí su importancia para la aplicación óptima de la información existente. Para ello se emplea una metodología o proceso estándar de seis fases para desarrollar e implementar análisis predictivos, Cross-Industry Standard Process for Data Mining, CRISP-DM por sus siglas en inglés y consiste en comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implantación, (Galán-Cortina, 2016). La metodología se aprecia en la tabla 1.

Como se puede observar, la minería de datos es un proceso mediante el cual se trata un problema organizacional a partir de los datos con que se cuenta en un momento dado, se debe tener conocimiento experto en el contexto organizacional así como de la organización misma para entender los datos con que se está trabajando. Es preciso destacar también que el análisis no se desprende de una hipótesis a probar ni de modelos preestablecidos, como se aprecia, parte de un planteamiento y análisis del contexto así como de la elección de las bases útiles para la identificación de conocimiento nuevo. Así también, se requiere de conocimiento en el tipo de datos que se tiene o debería tener para acceder a ellos, sí también para identificar si se trata de formatos diversos y variables distintas y verificar si es posible obtener nuevos datos con los ya existentes. Asimismo, en la modelación de los datos, se requiere de conocimiento de las técnicas

**Tabla 1. Minería de Datos a través de la metodología CRISP-DM**

Fase	Concepto	Actividades
Comprensión del negocio	Se trata de entender de forma profunda el contexto del problema.	Determinación de objetivos. Valoración de la situación. Determinación de los objetivos. Producción de un plan de proyecto
Comprensión de los datos	Se trata de entender a fondo los datos y su origen.	Recopilación de datos iniciales. Descripción de los datos Exploración de los datos Verificación de los datos.
Preparación de los datos.	Se unifica en un almacén de datos la diversidad de fuentes de datos internos (texto, html, bases de datos) y externos que se requieren para un análisis.	Selección de los datos. Limpieza de datos. Construcción de nuevos datos. Integración de datos. Formato de datos.
Extracción de modelos o modelado.	Se emplean métodos de análisis de datos para la extracción de información.	Selección de técnicas de modelado. Generación de un diseño de comprobación. Generación de modelos. Evaluación del modelo
Evaluación	Se evalúa el modelo de acuerdo al problema planteado.	Evaluación de los resultados. Proceso de revisión. Determinación de pasos siguientes.
Implementación	Se consideran los resultados de la evaluación y se define una estrategia. Técnicas de decisión (e.g. IBM Analytical Decision Management)	Planificación de la implementación. Planificación del control y mantenimiento. Creación de un informe final. Revisión final del proyecto. os resultados obtenidos, las predicciones (no son recomendaciones) se analiza ¿Cuál es la mejor decisión?

Fuente: Elaboración propia con base en IBM, 2016 y CRISP-DM, 2012.



de predicción, clasificación, segmentación y asociación para entender el proceso que se realiza. Por último, para la evaluación e implementación se precisa regresar al problema con que da origen el análisis para verificar si el modelo que se generó es el adecuado para entender el fenómeno organizacional y determinar las acciones a seguir así como seguimiento a su aplicación.

## **2. Conocimiento a partir de un modelo estadístico**

En la sección anterior, se analizó el caso mediante el cual el análisis de datos parte de bases de datos existentes, no se inicia con hipótesis de investigación ni se tiene un modelo en que se establecen las relaciones entre las variables. En esta sección se hace un análisis de datos que tiene su origen en la modelación previa, es decir, en el establecimiento de relaciones y se considera bases de datos que tienen su origen en ello y que fueron recopilados para el estudio específico para la que fue creada. Para explicar entonces este tipo de análisis se parte primero de qué es un modelo estadístico y cuál es el análisis multivariante que debería tener.

### ***Modelo estadístico***

El modelo estadístico es un tipo de modelo matemático que representa la realidad y sus aspectos; refiere la contribución de cada factor a un fenómeno determinado para poderlo cuantificar mediante estadística, esta última con el fin de inferir información de los resultados de una muestra al resto de la población. El modelado entonces es el proceso de desarrollar modelos estadísticos (IBM, 2018). En el modelado estadístico se especifica de forma previa el modelo en el que se representa una hipótesis, es decir, se especifican relaciones.

### ***Análisis multivariante***

Las técnicas de análisis multivariante son cada vez más aceptadas en la investigación científica pues se requiere de analizar la compleja realidad a partir de relaciones que incluyen más de tres variables, además los software incluyen en sus desarrollados procesamientos las herramientas que permiten realizar estos análisis (Mejía, 2018). De acuerdo con el autor, un

análisis multivariante es el análisis simultáneo de más de dos variables en las que éstas se relacionan y sus efectos no pueden interpretarse de forma independiente, su propósito es medir, explicar y predecir el grado de relación, por lo que aspectos como test de significación, escalas de medida y valor teórico deben estar considerados para su correcta aplicación.

El análisis multivariante es una herramienta para lograr entender la relación significativa y predictiva entre variables. De acuerdo a Mejía (2018) debe haber previo a éste, un constructo teórico conceptual muy sólido, por ello recomienda determinar la significación práctica y estadística, considerar un tamaño muestral adecuado, conocer los datos, verificar de forma constante la parsimonia del modelo, atender a los errores, validar los resultados y dar validez al modelo. Y aunque no exista una metodología como tal, existen procedimientos o fases que deberían considerarse en el diseño de modelos predictivos que facilitan la modelización y permitan identificar la significación de los resultados y modelos que describan mejor a la población en su conjunto.

Dichas fases se centran en un plan de investigación bien definido teniendo como base un modelo conceptual que establezca las relaciones a examinar, seguido de la investigación de campo o empírica, selección de la técnica multivariante y su ejecución. Luego la obtención de los resultados significativos y la interpretación, que se convierte en un asunto central, enfocándose al valor teórico. Por último, las medidas de diagnóstico del modelo posibilitan que el modelo no sea válido sólo para la muestra, sino, que sea generalizable. Las fases son las siguientes, se muestran en la tabla 2.

Como puede observarse y haciendo un comparativo de ambos planteamientos para encontrar modelos predictivos se tiene lo siguiente: en ambos casos es necesario revisar el contexto, su problemática y el objetivo. No obstante, en la minería de datos no se parte de una hipótesis ni de un modelo que establece las relaciones a probar. Así también, mientras que en el establecimiento del diseño (análisis multivariante) el investigador define cuales son las variables a incluir, su operacionalización, el tamaño óptimo de la muestra donde se recaban los datos, así como identifica qué tipos de datos se utilizarán en el análisis inferencial y descriptivo, mientras que en el entendimiento de los datos (minería de datos) se identifican los datos que participarán en el análisis de acuerdo al planteamiento del problema.

Así también, en la evaluación de los supuestos en el análisis multivariante, para cada técnica se debe revisar que los datos cumplan con ciertos requisitos de linealidad, normalidad, homocedasticidad, homogenei-

**Tabla 2**  
**Fases para la modelación de un problema multivariante**

Fase	Concepto	Actividades
1. Problema, objetivo y técnica multivariante.	Una vez planteado el problema, se definen los objetivos en términos conceptuales. Se determina el modelo conceptual con sus relaciones y por último se identifican las características de medida de las variables.	<p>Definir el problema de investigación y objetivos analíticos.</p> <p>Modelos conceptuales:</p> <p>relación de dependencia: especificar conceptos dependientes y conceptos independientes.</p> <p>Relación de interdependencia: determinar dimensiones de la estructura. Identificar técnica (medida a utilizar) útil para examinar la relación.</p> <p>1. Análisis dependiente.</p> <p>a) Una dependiente y escala métrica: regresión lineal múltiple y análisis de conjunto.</p> <p>b) Una dependiente y escala no métrica: análisis discriminante múltiple, análisis de regresión logística.</p> <p>c) Varias dependientes y escala métrica: análisis de correlación canónica.</p> <p>d) Varias dependientes y escala no métrica: análisis de la varianza multivariante (MANOVA).</p> <p>e) Múltiples dependientes e independientes: modelo de ecuaciones estructurales.</p> <p>2. Análisis interdependiente.</p> <p>a) Estructura en variables: análisis factorial.</p> <p>b) Estructura en casos: análisis cluster.</p> <p>Estructura en objetos métrica: análisis multidimensional.</p> <p>d) Estructura objetos no métrica: análisis de correspondencias.</p>

2. Diseño: desarrollo del proyecto de análisis.	Desarrollar un plan de análisis que dirijan el conjunto de supuestos que subyace a la técnica para finalizar la formulación del modelo y las especificaciones de la recolección de datos.	Tamaño de muestra óptimo. Tipos de datos métricos y no métricos, métodos de estimación del modelo.
3. Evaluación de los supuestos de la técnica multivariante.	Evaluar los supuestos subyacentes estadísticos como conceptuales para revisar su capacidad de representar relaciones multivariantes.	Estadística inferencial: Normalidad. Linealidad, independencia en los términos de error e igualdad de varianzas en una relación dependiente: Homocedasticidad. Homogeneidad de la muestra Nexos conceptuales. Cada técnica tiene sus propios supuestos conceptuales: formulación de modelos y representaciones.
4. Ejecución	Estimación efectiva del modelo. Valoración del ajuste del modelo o corroborar que se obtuvieron niveles aceptables sobre los criterios estadísticos para posibilitar la inferencia.	Estimación. Elección de características específicas de los datos: e.g. MANOVA. Maximizar ajuste de los datos: e.g. rotación de factores o funciones discriminantes. Valoración. E.g. nivel de significación.
5. Interpretación	Interpretación del valor teórico (combinaciones múltiples de variables) de los resultados multivariantes para conocer la naturaleza de las relaciones.	Examinar los valores de los coeficientes estimados (ponderaciones) para cada variable. En la interpretación se puede volver a formular el modelo re-especificando las variables.

Fuente: Elaboración propia con base en Mejía, 2017 y 2018.

dad mientras que en la preparación de los datos en la minería de datos se preparan los mismos para realizar el modelado, pero no se trata de cumplir con supuestos, se corre el riesgo de que el modelo predictivo no sea tan preciso (IBM, 2016). Ahora bien, en la ejecución del modelo y modelado, en ambos análisis es parte crucial para la predicción del fenómeno. Asimismo, en la interpretación y evaluación se requiere de una revisión analítica de los resultados por parte del investigador. Por último en la

parte de validación del análisis multivariante se quiere de garantizar el grado de generalización del modelo y en ambos, para la implementación, es preciso combinar los resultados con la experiencia del investigador en la toma de decisiones.

### 3. Características de IBM SPSS modeler versión 18

El uso de software especializado cada vez es más recurrente en el análisis de grandes bases de datos. Las herramientas de análisis de software son necesarias para analizar grandes cantidades de bases de datos (Aranda y Sotolongo, 2013). Pues cuando se analizan demasiados datos el proceso se convierte en algo engorroso (Soto-Jaramillo, 2009). No obstante, los programas tienen la cualidad de ser de difícil acceso pues son muy costosos (Aranda y Sotolongo, 2013). Al tener poco acceso a dichos programas reduce la posibilidad de realizar una extracción de conocimiento nuevo y de patrones o modelos predictivos

Es así que empresas como Microsoft, Oracle así como IBM crean las herramientas que permiten agilizar el tiempo de respuesta para procesar grandes bases y puedan ser transformados (Galán-Cortina, 2016). El Programa *SPSS Modeler de IBM* ha sido reconocido por su papel en brindar a sus clientes un enfoque integrado y holístico en la gestión empresarial y análisis de grandes bases de datos. Es por ello que en este apartado se analizan las principales características que posee el programa *IBM SPSS Modeler* en las herramientas que brinda en sus bases de datos.

*IBM SPSS Modeler* es una herramienta de software para la minería de datos desarrollada por IBM Corp., (IBM, 2016a). Originalmente se llamaba *SPSS Clementine* y en 2009 se llamó *PASW Modeler* (Galpan-Cortina, 2015). Se trata de un programa orientado las empresas con el objetivo de una mejor comprensión de los datos y obtener los mejores resultados. Esto, mediante una interfaz visual que conlleva a modelos de predicción (pronósticos, clasificaciones, segmentación y asociaciones) más eficaces y en menos tiempo.

En su diseño considera la metodología CRISP-DM, ésta como la minería e datos se muestran en los previos párrafos (IBM, 2012). Este software funciona con interfaz visual y las funciones basadas en nodos –íconos- que se van tomando y forman una ruta o stream que pueden archivarse (de manera individual o por proyectos, mismos a los que se puede acceder para reestructurarse y las bases de datos quedan independientes (IBM, 2016a). La ruta es llamada ruta de datos, consiste en una secuencia de

operaciones en donde los datos fluyen de registro en registro, se manipula un llegan a un destino, esto es, *IMB SPSS Modeler*, mediante el lienzo de rutas lee los datos, los ejecuta y por último los envía a un destino dibujando diagramas de operaciones (con nodos=íconos y rutas=flujo). Como resultado final visual se obtiene un nugget que contiene todas operaciones realizadas. En conjunto con Modeler Server ofrece un rendimiento mayor en el caso que se trabaje con grandes cantidades de datos. Las características del software se muestran en la siguiente tabla 3.

**Tabla 3.**  
**Características generales de SPSS Modeler**

<b>Características generales del software</b>	
Sistemas operativos	Windows, Linux, Sun Solaris, HP-UX o IBM-AIX, Mac OsX.
Lenguaje	Código abierto, Python (Jython), R. Pago: SPSS
Interfaz	Java
Licencia	Usuarios autorizados y concurrentes limitado en cantidad (de personas), plazo, mantenimiento, alquiler.
Competencia	Oracle Data Mining, WEKA, RapidMiner, Statistic Data Miner, IBM DB2 InfoSphere Warehouse, Microsoft Analysis Services.
Compatible con	Oracle Data Miner, IBM DB2 InfoSphere Warehouse y Microsoft Analysis Services, IBM Netezza Analytics
Compatibilidad	SQL (Structured Query Language), lenguaje de consulta estructurada
Estadísticos	Confirma relaciones sospechosas entre las variables. Los estadísticos de IBM SPSS Statistics pueden emplearse en IBM SPSS Modeler. Se puede acceder y ejecutar determinadas rutinas de IBM SPSS Statistics en IBM SPSS Modeler para generar y puntuar modelos.
Manipulación de datos	Construye nuevas bases de datos a partir de los existentes.
Exploración y visualización	En nodo auditoría de datos desarrolla auditoría inicial de los datos incluyendo gráficos y estadísticos.
Comprobación de hipótesis	Construye modelos que muestran la forma en que se comportan los datos, y verifica estos modelos.

Productos	IBM SPSS Modeler, IBM SPSS Modeler Server, IBM SPSS Modeler Administration Console, IBM SPSS Modeler Batch, IBM SPSS Modeler Solution Publisher, IBM SPSS Modeler Server adaptadores para IBM SPSS Collaboration and Deployment Services
Ediciones	IBM SPSS Modeler Profesional e IBM SPSS Modeler Premium conformado por IBM SPSS Modeler Entity Analytics, IBM SPSS Modeler Social Network Analysis e IBM SPSS Modeler Text Analytics.
Nodos	<p>Nodos de origen en la paleta de orígenes: lector de datos en forma de círculo</p> <p>Nodos de proceso en la paleta de operadores con registros y con campos: transformador de datos, en forma de hexágono.</p> <p>Nodos de resultados en la paleta de gráficos, informe, exportar: generador de documentos, en forma de triángulo (gráfico) y en forma de cuadrado (informe, exportar)</p> <p>Nodos de modelado en la paleta de modelado: generador de modelos en forma de pentágono.</p>
Preparación de datos	ADP nodo de preparación de datos automática Nodo auditoría de datos valores perdidos, atípicos, sesgos.
Métodos de modelado	<p>Clasificación. Los modelos usan el valor de uno o más campos de entrada para predecir el valor de uno o más resultados o campos de destino, incluyen: aprendizaje automático de las máquinas, inducción de reglas, identificación de subgrupos, métodos estadísticos y generación de varios modelos.</p> <p>Nodo clasificador, nodo auto-numérico, nodo de árbol de clasificación y regresión (C&amp;R), nodo QUEST, nodo CHAID, nodo C5.0, Nodo lista de decisiones, modelos de regresión lineal, nodo PCA/Factorial, Nodo Selección, análisis discriminante, regresión logística, modelo lineal generalizado, nodo de regresión Cox, nod máquina de vectores (SVM), nodo red bayesiana, nodo modelado de respuesta, nodo serial temporal, nodo K (KNN), nodo predicción espacio-temporal (STP).</p>

	Asociación. Modelos que identifican patrones en los datos en los que una o más entidades, se asocian con una o más entidades, útil cuando se desea predecir varios resultados. Nodo a priori, modelo CARMA, nodo secuencia, nodo reglas de asociación.
	Segmentación. Son modelos de agrupación en clústeres que dividen los datos en segmentos o clústeres de registros que tienen patrones similares de campos de entrada, cuando se desconoce el resultado específico. Nodo agrupación, nodo K-medias, nodo Kohonen, nodo Bietápico, nodo detección de anomalías.
Evaluación	Se pueden comprobar supuestos verificando el comportamiento de los datos y de los modelos.
Implementación	Los resultados obtenidos no son sugerencias, Modeler ofrece un componente central de Predictive Analytics de IBM en Linux para sistema Z que optimiza la decisión de acuerdo al planteamiento.
Productos IBM SPSS Modeler	IBM SPSS Modeler, IBM SPSS Modeler Server, IBM SPSS Modeler Administration Console, IBM SPSS Modeler Batch, IBM SPSS Modeler Solution Publisher, Adaptadores de IBM SPSS Modeler Server para IBM SPSS Collaboration and Deployment Services

Elaboración propia con base en IBM, 2016 y CRISP-DM, 2012.

Como puede observarse *Modeler* cuenta con una serie de algoritmos que proporcionan las herramientas para el análisis de datos cuando se pretende hacer uso óptimo la información con que se cuenta en el planteamiento de determinados problemas. Su interfaz es sencilla de emplear pues cuenta con una serie de íconos –nodos– que el investigador puede ir empleando en el desarrollo de sus modelos de acuerdo con la metodología CRISP-DM. Se pueden comprobar supuestos verificando el comportamiento de los datos y de los modelos.

#### 4. Estudios analizados con spss modeler

Se realizó una búsqueda de información respecto al uso de *software Modeler SPSS* que permita conocer la trascendencia que tiene en la minería de datos. Los estudios revelan que el programa se emplea en el análisis de datos que permite identificar aspectos que previamente, con otras técni-



cas, no ha sido posible. En el caso de Lotfnezhad, Ahmadi, Roudbari y Sadooghi (2015) se extrajeron y descubrieron patrones ocultos en las bases de datos recopilados y se analizaron diversos factores que generan el cáncer de mama; el modelo identificó diez variables predictoras. Asimismo, Shao, Liancheng y Han (2016) concluyeron que el método de predicción funciona de forma efectiva y precisa en la selección en la identificación de la falla de línea single-line-to-ground (SLG), analizando datos de cantidades eléctricas y no eléctricas.

Otros casos como el de Wang, Wen, Lu, Yao y Zhao (2016) desarrollaron modelos de predicción para predecir eventos relacionados con el esqueleto SER en pacientes con cáncer. Y el respecto al análisis de Di, Yang, Fu, Lin y Jiang (2013) desarrollaron un modelo de elección de puntos de acupuntura en la cefalea de tipo tensional, en la que ésta se materializa principalmente en la selección de puntos de acupuntura ubicados en la cabeza combinados con un meridiano distintivo. En éstos y los casos previos, se observa el uso del software en el análisis de datos y en la modelación de predicciones precisa y eficiente. Tal como IBM (2016) establece dentro de sus propósitos.

En los párrafos anteriores se reconoce la función del software *IBM SPSS Modeler* en la minería de datos, es decir en el análisis de datos con el propósito de encontrar nuevos patrones o patrones ocultos y las relaciones que tienen los datos, esto es, cuando el análisis no inicia con un modelo de relaciones de variables. Así también, no se inicia con una hipótesis a probar y los datos con los que se cuenta fueron recopilados para fines diversos en las diferentes áreas y no con un objetivo en particular, además se hizo uso de múltiples bases y diversos formatos y generalmente se reutilizaron. Dicho software ha sido útil en las decisiones que requieren de una modelación rápida con las diversas bases existentes. Se encuentra en todos ellos la eficiencia de los modelos en modelación predictiva, la que ya se encuentra en la fase de implementación.

## 5. Características de IBM SPSS Statistics versión 24

SPSS corresponde al nombre de Statistical Package for the Social Sciences, software estadístico que ha sido empleado por las ciencias exactas, sociales y aplicadas, incluso las de investigación de mercado (Mejía, 2018). De acuerdo a este autor la primera versión data de 1968 y la última, la versión 24 en 2016. El software *IBM SPSS Statistics* emplea bases casi de

cualquier tipo para convertirlos en informes a manera de tablas, gráficos, diagramas de distribuciones, así también empleado para el análisis de datos estadísticos descriptivos y complejos (IBM, 2011).

*Statistics* emplea un lenguaje en comandos, menús y cuadros de diálogo. Entre sus principales opciones, este programa cuenta con *Statistics base* que proporciona una gama de procedimientos estadísticos como tablas de contingencia, estadísticas descriptivas, así como variedad de operaciones como reducción de dimensiones, clasificación, segmentación (análisis factorial, análisis de conglomerados, discriminante). Y sobre todo, ofrece una serie de algoritmos para comparar medias y técnicas predictivas, t, ANOVA, Regresión lineal y regresión ordinal. Otra opción que ofrece es estadísticas avanzadas como análisis de supervivencia Kaplan-Meyer, otra como lo es *Bootstrapping* (método para obtener estimaciones más precisas, intervalos de confianza para estimar por ejemplo coeficientes de correlación y regresión; otras como *conjoint*, tablas personalizadas, preparación de datos, árboles de decisión, predicciones (ajustes de curvas), valores perdidos. En la tabla 4 se observan algunas características.

**Tabla 4.**  
**Características generales del software, IBM SPSS STATISTICS 24**

<b>Características generales del software</b>	
Sistemas operativos	Windows, Linux, Mac OSX
Lenguaje o sintaxis	Código abierto, Python (2,7), R Pago: SPSS
Interfaz	Gráfica
Licencia y costo	Usuarios autorizados y concurrentes limitado en cantidad (de personas), plazo, mantenimiento, alquiler
Competencia	SAS, MATLAB, Statistica, Stata, R

**Nuevas funciones de la versión 24**

Nuevas funciones	Acceso a extensiones, IBM SPSS Extensión Hub, Custom Dialog Builder y Módulo tablas personalizadas
Acceso	A más de 100 extensiones lo que permite acceso a bibliotecas gratuitas (Ver lenguaje)
IBM SPSS Extensión Hub	Explorar, descargar, actualizar eliminar y administrar extensiones

Custom Dialog Builder	Actualización para construir e instalar extensiones.
Módulo tablas personalizadas	Para importar y exportar datos a SPSS
Ediciones	SPSS Statistics Estándar, SPSS Statistic Profesional, SPSS Statistics Premium.
SPSS Statistics Estándar,	Procedimientos estadísticos para modelos estadísticos lineales, no lineales, de simulación predictiva.
SPSS Statistics Profesional	Preparación de datos, valores perdidos, validez de los datos, árboles de decisión y pronóstico.
SPSS Statistics Premium	Permite modelos de ecuaciones estructurales y pruebas de maestro. Paquetes de marketing directo, tablas y gráficos de alta gama.

Elaboración propia con base en IBM, 2016.

## 6. Análisis comparativo de IBM modeler y statistics

En la siguiente tabla se clasificaron las funciones de los dos programas con el objetivo de identificar si se puede hacer minería de datos con *Statistics* o se puede realizar un análisis multivariante a partir de un modelo estadístico con *Statistics*.

Como puede observarse y haciendo un comparativo del software en el análisis de datos bajo los planteamientos de análisis multivariante y minería de datos en el establecimiento de modelos predictivos se tiene lo siguiente: en las primeras fases de establecer objetivo y entender el negocio-entendimiento de los datos en ambos casos es preciso que el investigador revise el contexto, su problemática y el objetivo, independientemente del software. No obstante, *Modeler* cuenta con un espacio en su plataforma que permite dar un seguimiento a las fases de la minería de datos, empezando con entendimiento del negocio y entendimiento de los datos, espacio en que se pueden guardar archivos o datos que el investigador precise para el entendimiento de los datos como del contexto. En el caso del diseño en el análisis multivariante, también se realiza de forma independiente al software que se emplee para su análisis, ya que el investigador define cuales variables elegir, cómo medirlas, así como determinar el tamaño muestra y conocer los datos (métricos o no métricos).

Así también, en la evaluación de los supuestos de normalidad, homocedasticidad, homogeneidad en el análisis multivariante, pueden ejecutarse con comandos a través de cuadros de diálogo en el programa SPSS Statistics, asimismo en *Modeler* los supuestos y preparación de datos ta-

**Tabla 6. Análisis comparativo de las herramientas IBM Modeler y Statistics**

Fase	Concepto	Actividades	IBM Modeler	IBM Statistics
Objetivo AM /entendimiento del negocio MD	Se trata de entender de forma profunda el contexto del problema	Analizar el contexto Seleccionar objetivos	En su plataforma cuenta con un espacio para que el investigador guarde, archive y visualice la información útil para el entendimiento del negocio y de los datos.	El investigador revisa de forma independiente al software el planteamiento del problema, así como el objetivo.
Diseño AM/Entendimiento de los datos MD	<b>AM</b> Se trata de establecer el modelo conceptual solido en el cual se establezcan las relaciones a medir, las hipótesis a probar y se cuente con datos representativos de la población	Definir variables. Establecer cómo medir variables. Determinar tamaño muestral. identificar datos métricos y no métricos		De forma independiente el investigador diseña el modelo en que establece las relaciones con las variables a incluir, la operacionalización, el tamaño de la muestra donde se recaban los datos y revisa los datos (métricos y no métricos) para elegir la técnica multivariante.
<b>MD</b> Se trata de entender a fondo los datos y su origen	Recopilación de datos iniciales. Descripción de los datos. Exploración de los datos. Verificación de los datos	En su plataforma cuenta con un espacio para que el investigador guarde, archive y visualice la información útil para el entendimiento del negocio y de los datos. El investigador puede formular hipótesis si explora los datos, esto último útil en dar forma a la transformación de los datos.		

<p><b>AM</b> Evaluar los supuestos subyacentes estadísticos como conceptuales para revisar su capacidad de representar relaciones multivariantes.</p>	<p>Estadística inferencial: Normalidad. Linealidad, independencia en los términos de error e igualdad de varianzas en una relación dependiente: Homocedasticidad. Homogeneidad de la muestra Nexos conceptuales. Cada técnica tiene sus propios supuestos conceptuales: formulación de modelos y representaciones</p>	<p>Acceso a Excel y archivos SPSS. Rápida instantánea visual de los datos. Permite aplicar reglas de validación que identifiquen valores de los datos no válidos, fuera de rango, perdidos o en blanco. Homocedasticidad: e.g. analizar-comparar medias-ANOVA de un factor-seleccionar en lista de dependientes variable métrica-selección de variable Factor-opciones-prueba de homogeneidad de las varianzas.</p>
	<p>Selección de los datos. Limpieza de datos. Construcción de nuevos datos. Integración de datos. Formato de datos.</p>	<p>Fuentes de datos ODBC, tablas Excel, archivos planos ASCII y archivos SPSS.  pick &amp; mix, muestreo, particiones, reordenación de campos, nuevas estrategias para la fusión de tablas, nuevas técnicas para recodificar intervalos numéricos, etc Auditoría de datos (perdidos, sesgos, atípicos) Linealidad (eg. Nodo Tipo o GLE)</p>
<p><b>MD</b> Se unifica en un almacén de datos la diversidad de fuentes de datos internos (texto, html, bases de datos) y externos que se requieren para un análisis.</p>	<p>Acceso de datos  Pre-procesado de datos</p>	
<p>Supuestos AM/Preparación de datos MD</p>		

Fase	Concepto	Actividades	IBM Modeler	IBM Statistic	IBM Statistic
Ejecución AM Extracción de modelos o modelado CRISP -DM.	<p>AM</p> <p>Estimación efectiva del modelo.</p> <p>Valoración del ajuste del modelo o corroborar que se obtuvieron niveles aceptables sobre los criterios estadísticos para posibilitar la inferencia</p>	<p>Estimación. Elección de características específicas de los datos: e.g. MANOVA.</p> <p>Maximizar ajuste de los datos: e.g. rotación de factores o funciones discriminantes.</p> <p>Valoración. E.g. nivel de significación</p>	Técnicas de aprendizaje. Modelado predictivo; de asociación y secuencia y; de segmentación	e.g. nodos de árboles de decisión, redes neuronales, agrupamiento, reglas de asociación, regresión lineal y logística, combinación de modelos.	<p>Árboles de decisión, redes neuronales, agrupamiento, asociación, regresión lineal, combinación de modelos.</p> <p>Técnicas de asociación.</p> <p>Análisis de conjunto</p>
	<p>MD</p> <p>Se emplean métodos de análisis de datos para la extracción de información.</p>	<p>Selección de técnicas de modelado.</p> <p>Generación de un diseño de comprobación.</p> <p>Generación de modelos.</p> <p>Evaluación del modelo</p>		<p>Statistic permite el modelado a partir de cuadros de diálogo y comandos</p>	<p>e.g. analizar - regresión - lineal, selección de variable dependiente y de variable independiente.</p>
<p>Interpretación AM/Evaluación CRISP -DM.</p>	<p>AM</p> <p>Interpretación del valor teórico (combinaciones múltiples de variables) de los resultados multivariantes para conocer la naturaleza de las relaciones.</p>	<p>Examinar los valores de los coeficientes estimados (ponderaciones) para cada variable. En la interpretación se puede volver a formular el modelo re-especificando las variables.</p>		<p>Los resultados pueden desplegarse fácilmente en bases de datos como IBM Statistics y otras aplicaciones.</p> <p>Los modelos generados incluyen los niveles de significación, pruebas t, F</p>	<p>Los modelos que se generan pueden probarse los niveles de significación, pruebas t, F</p> <p>Proceso analítico en la interpretación de resultados multivariantes.</p> <p>Revisión del valor teórico (ponderaciones de las variables)</p> <p>Significación práctica</p>

Implementación MD	<p>MD</p> <p>Se evalúa el modelo de acuerdo al problema planteado.</p>	<p>Evaluación de los resultados. Proceso de revisión. Determinación de pasos siguientes.</p>	<p>Técnicas para la evaluación</p>	<p>Modelos guiadas por las condiciones especificadas por el experto</p>				
		<p>Planificación de la implementación. Planificación del control y mantenimiento. Creación de un informe final. Revisión final del proyecto. Los resultados obtenidos, las predicciones (no son recomendaciones) se analiza ¿Cuál es la mejor decisión?</p>	<p>Visualización de resultados</p> <p>Informes</p>	<p>Ofrece un potente soporte gráfico que permite al usuario tener una visión global de todo el proceso, que comprende desde el análisis del problema hasta la imagen del modelo aprendido.</p> <p>en HTML y texto, volcar los resultados de la minería de datos y exportar los modelos a distintos lenguajes como C, SPSS, HTML, PMML, SQL, etc.</p>	<p>Para la implementación los resultados no son recomendaciones, Modeler cuenta con un componente central PREDICTIVE Analytics de IBM que coadyuva a identificar la mejor opción para el problema planteado.</p>			
Validez del modelo AM	<p>Del modelo para garantizar cierto grado de generalidad del modelo obtenido.</p>	<p>Demostrar la generalidad del modelo obtenido de los datos de una muestra al resto de la población. bootstrapping</p>	<p>Boosting Para generar modelos más precisos, estables</p>	<p>En cada modelo generado, Modeler obtiene el nivel de significación para las pruebas de hipótesis, incluyendo pruebas de bondad de ajuste, F.</p>	<p>Para garantizar cierto grado de generalidad. Validación cruzada Jacknife Bootstrap</p>			
				<p>En cada modelo generado, Statistic permite revisar el nivel de significación, incluyendo pruebas F, t, ajustes de bondad.</p>				

Otras características adicionales a la metodología o fases de AM y MD		
Diseño	Minería de datos en la construcción de modelos predictivos y análisis de datos	Paquete de productos para consolidación de modelos predictivos y análisis de datos profesionales.
Usuarios	Poca habilidad en programación por el diseño de la interfaz. Mucha habilidad en técnicas.	Mucha habilidad por tratarse de comandos, códigos o cuadros de diálogo. Mucha habilidad en técnicas.
Usuarios	Poca habilidad en programación por el diseño de la interfaz. Mucha habilidad en técnicas.	Mucha habilidad por tratarse de comandos, códigos o cuadros de diálogo. Mucha habilidad en técnicas.
Objetivo del modelo	Fuerte para decisiones operativas, débil para investigación y comprensión.	Fuerte para investigación y comprensión, apropiado para procedimientos multivariados clásicos.
Precisión de la predicción	Construye, compara, prueba y combina rápidamente modelos, pero no se comprende la significación, ni se demuestra variación.	Comprende variación desde el punto de vista estadístico y significación.
Origen de los datos	Datos existentes que se recopilaron	Una base de datos recopilada para un cierto análisis, con un proceso analítico
Fuentes de datos	Diversas fuentes y múltiples formatos. Consolida la diversidad de datos y formatos. Reutilizan datos	Una sola fuente, archivos un solo formato. Base nueva, se analiza cuando aparece.
Informes analíticos	Detección de patrones ocultos	Capacidad de tabulación de datos, creación de gráficos.
Planteamiento del problema	Minería de datos. Para encontrar nuevos patrones y relaciones entre los datos, resultado modelo de predicción. No se especifican relaciones Recorrido exploratorio Sin un objetivo particular No requiere hipótesis Debe cumplir con metodología CRISP-DM	Modelo estadístico. El modelo se especifica de antemano. Se especifican relaciones Se busca una respuesta a una pregunta y un objetivo Requiere de una hipótesis a comprobar Método científico



Entify Analytics	Detección de duplicación para identificar los que tienen una gran probabilidad de pertenecer a la misma entidad.	No tiene la herramienta.
Pronóstico	Identifica el método de series de tiempo que mejor se ajusta a los datos históricos	El paquete SPSS Statistics Premium permite entre otras cosas el pronóstico.
Datos	Big data, ágil con grandes volúmenes de datos. Se puede tener acceso a ellos sin sacarlos del repositorio.	2 millones de registros y 250,000 variables Debe extraerse la base de datos al software para su análisis.
Expresión del modelo de predicción.	Es un objeto físico y portátil que se puede convertir a PMML y adjuntarlo a nuevos nodos.	Se expresa en forma de coeficientes y estadísticos. El analista analiza estos resultados, le da sentido
Optimización de la predicción	La predicción no es una recomendación. Modeler ofrece IBM Analytical Decision Management para que con los resultados predictivos obtenidos, se tome la mejor decisión. Más allá de la predicción, ofrece escenarios futuros para la toma de decisiones.	La predicción no es una recomendación. El analista recomienda de acuerdo a los resultados, estudios previos y teoría con que se analiza el fenómeno.
$Y=mx + c$	LINEAL, versión mejorada de regresión lineal	Regresión lineal
Conversión a otros lenguajes.	Los algoritmos de inducción de reglas/árboles de decisión (clasificación) pueden convertirse en SQL y se pueden comparar con otros métodos.	No tiene la herramienta
Exportación de modelos a otros lenguajes	SAS,SPSS	
Exportación de datos	XLS	
Lenguaje	Éxito en analíticas de código abierto. Interfaz directa a R y puede combinarse con otros procedimientos.	La versión 24 ofrece acceso a más de 100 extensiones para aprovechar bibliotecas gratuitas en s intaxis R, Python y SPSS
$Y=mx + c$	LINEAL, versión mejorada de regresión lineal	Regresión lineal
Conversión a otros lenguajes.	Los algoritmos de inducción de reglas/árboles de decisión (clasificación) pueden convertirse en SQL y se pueden comparar con otros métodos.	No tiene la herramienta

Exportación de modelos a otros lenguajes	SAS,SPSS	
Exportación de datos	XLS	
Lenguaje	Éxito en analíticas de código abierto. Interfaz directo a R y puede combinarse con otros procedimientos.	La versión 24 ofrece acceso a más de 100 extensiones para aprovechar bibliotecas gratuitas en sintaxis R, Python y SPSS
Análisis de Texto	Text Analytics de Modeler Premium permite transformar datos no estructurados a datos estructurados.	No tiene la herramienta de procesamiento de texto.
Interfaz	Visual, basada en procesos, flujos de datos (streams), iconos, fáciles de usar. Programación visual que ilustra.	Códigos, comandos o cuadros de diálogo.
Indicadores	Índice de Victoria Huwiz. Modeler permite el análisis holístico para la administración y análisis de grandes bases de datos.	

Elaboración propia con base a IBM modeler 2016 e IBM statistics 2016.

les como atípicos, sesgo y perdidos (e.g. auditoría de datos) y de linealidad en nodo Tipo o nodo GLE en el que se determina condiciones del modelo.

Ahora bien, en la ejecución del modelo y modelado, en ambos análisis es parte crucial para la predicción del fenómeno, para el análisis multivariante, una vez seleccionada la técnica indicada para el tipo de problema y variables a emplear, así como los datos recolectados en una muestra representativa, se ejecuta en el programa SPSS Statistics mediante comandos y cuadros de diálogo, (e.g. regresión lineal simple: analizar, regresión, lineal, selección de variable dependiente, selección de variables). Mientras que el modelado en SPSS se da mediante diversos nodos, sin tener que buscar los cuadros de diálogo o ejecutar comandos, sino en forma automática empleando los nodos que cuenta con una gran variedad que facilita el trabajo al investigador y optimiza las bases de datos resultando en cada uno de los modelos el más adecuado de acuerdo al objetivo. En ambos se puede hacer uso de datos métricos como no paramétricos (e.g. nodo C&R, CHAID en *modeler*).

Asimismo, en la interpretación y evaluación para ambos casos la revisión analítica del investigador es necesaria y ningún programa la sustituye. Para ambos casos la modelación permite analizar los valores de las ponderaciones, en las cuales se revisa el nivel de significancia en la prueba de hipótesis, las de ajuste de bondad, F y t. Así también, en la parte de validación del análisis multivariante se quiere garantizar el grado de generalización del modelo con algunas técnicas como validación cruzada, *jackknife* y *bootstrap*, mismas que el programa SPSS *Statistics* ofrece mediante comandos y cuadros de diálogo. Así también, *Modeler* ofrece las herramientas de *boosting* y *bootstrap* para generar modelos más precisos más estables y más consistentes.

En ambos, para la implementación, es preciso combinar los resultados con la experiencia del investigador en la toma de decisiones, no obstante, y considerando que los resultados no son recomendaciones, *Modeler* ofrece una alternativa para evaluar los resultados y someterlos a otro proceso para delinear las técnicas de decisión; *IBM Analytical decision management*, herramienta que apoya en proporcionar la mejor alternativa posible para el problema de que se trate.

## Discusión

Atendiendo a la pregunta y tratando de contestar si ¿Se puede hacer minería de datos con *Statistics*? y con base en todo lo anteriormente expues-

tos podría decir que no, que la minería de datos es adecuada con software *Modeler* ya que dicho proceso consiste en un análisis de datos con el objetivo de encontrar nuevos patrones o patrones ocultos y las relaciones que tienen los datos, esto es, cuando el análisis no inicia con un modelo de relaciones de variables, ni con un modelo estadístico previo, sino con una metodología estructurada CRISP-DM.

Así también, no se inicia necesariamente con una hipótesis a probar y los datos con los que se cuenta fueron recopilados para fines diversos a lo largo del tiempo y no con un objetivo preciso, además son múltiples bases y diversos formatos y generalmente se reutilizan. *Modeler* puede ser óptimo en las decisiones operativas en la modelación rápida. Asimismo, las características del software permiten llevar a cabo el proceso de describir conocimiento y crear modelos predictivos ofreciendo una serie de facilitadores a los investigadores que aunque no requieren gran experiencia de programación (e.g. IMB, 2016a), sí requiere de conocer las técnicas y la plataforma que se emplea, ya que aunque el sistema ofrece automatización en sus operaciones, debe entenderse todo lo relacionado a los datos y sus resultados. Este programa tiene la opción de tener en la interfaz, la herramienta de *Statistics*, para que, poder realizar algunas pruebas en *Modeler*, por ejemplo modelado.

En caso del cuestionamiento de si ¿Se puede realizar un análisis multivariante a partir de un modelo estadístico con *Modeler*? No necesariamente, si se requiere de analizar una base de datos para probar una hipótesis, es el software *Statistics* el apropiado cuando se pretende realizar un análisis de datos con el objetivo de encontrar relaciones con un modelo estadístico previo. Así también, el análisis parte con un objetivo claro y una hipótesis de investigación a probar. Cabe destacar que los datos con los que se cuenta fueron recopilados para ese objetivo, además corresponden a una sola base, un solo formato y se trata de una base nueva. Dicho software puede ser óptimo en la investigación y comprensión de un fenómeno en la modelación precisa, pues contiene las herramientas para comprender la variación de los datos estadísticamente así como su significación, aspecto necesario para la demostración de la probabilidad del evento y su precisión. Por lo tanto, para realizar análisis multivariantes a partir de hipótesis y con pruebas de significación el software *IBM Statistics* es el adecuado para hacerlo.

Las herramientas de SPSS *Statistics* permiten el análisis multivariante de técnicas dependientes de la forma convencional, en la que la investigación encuentra las bases para explorar la variación de los datos, es decir el error de la probabilidad en que se puede dar el evento estudiado,

así como la significación que es de vital importancia en la demostración del potencial explicativo de un modelo estadístico. No obstante, *Modeler* también ofrece el resultado de las pruebas de significancia que requiere un modelo para poder probar su nivel de predicción, su precisión y estabilidad, permitiendo por otro lado, una interfaz más sencilla, pero más completa en el establecimiento de modelos predictivos.

## Conclusiones

El propósito de este ensayo fue esclarecer las características de *Modeler* y *Statistics* en la búsqueda de modelos predictivos a partir de dos enfoques, la extracción de conocimiento a través de la minería de datos de bases existentes por un lado y por otro, el análisis multivariante a partir de modelos estadísticos. Y a través de su desarrollo se expusieron las diferentes herramientas que los softwares ofrecen para ambos enfoques.

Respecto a la minería de datos, se apreció que se trata de una de las principales fases del proceso de extracción de datos en la que se identifican datos así como patrones antes ocultos, se modelan relaciones, así como su nivel de predicción a partir de bases existentes. Por lo que el software *Modeler* es más adecuado cuando se pretende llegar a modelos predictivos a partir ésta, pues entre otras cosas puede ser óptimo en las decisiones operativas en la modelación rápida identificando patrones a través de bases de múltiples fuentes y diversos formatos, cuando no se tiene un objetivo claro. Asimismo, las características del software permiten llevar a cabo el proceso de describir conocimiento, y además posibilita las investigaciones científicas que se proponen responder a variaciones y test de significancia de los datos.

Así también, en lo que respecta a los modelos estadísticos, se observó que representan de forma previa los aspectos y relaciones de un evento el cual señala supuestos (hipótesis) para posibilitar su cuantificación mediante estadística y poder inferir los resultados al resto de la población. Así también las técnicas de análisis multivariante permiten realizar estas mediciones de las relaciones entre variables (tres o más) y que requieren de pasar las pruebas de significación, escalas de medida y valor teórico para su correcta aplicación.

Por lo que el software *Statistics* puede ser óptimo en la investigación y comprensión de un fenómeno en la modelación precisa a partir de una hipótesis y una base de datos, pues contiene las herramientas para comprender la variación de los datos estadísticamente así como su significa-

ción, aspecto necesario para la demostración de la probabilidad del evento y su precisión. Por lo tanto, para realizar análisis multivariantes a partir de hipótesis y con pruebas de significación el software *IBM Statistics* es el adecuado para hacerlo. Sus herramientas permiten el análisis multivariante de técnicas dependientes de la forma convencional, en la que la investigación encuentra las bases para explorar la variación de los datos, es decir el error de la probabilidad en que se puede dar el evento estudiado, así como la significación que es de vital importancia en la demostración del potencial explicativo de un modelo estadístico.

Así también, los modelos predictivos que surgen del análisis en *Modeler*, ofrecen precisión, estabilidad y consistencia. Además de ofrecer el plus de la automatización, ya que esta plataforma hace más sencillo el trabajo complejo de investigadores, académicos y empresarios. En cambio *Statistics* no ofrece la interfaz tan gráfica como la primera, pero sus comandos, cuadros de diálogo y el paquete de programas estadísticos permite al investigador experto obtener el análisis estadístico de la los datos. Por tanto, el análisis multivariante tiene lugar en ambos paquetes de IMB, pues los datos pueden explotarse y emplearse para la identificación de patrones desconocidos y conocimiento que no se considera en los modelos previos. No obstante, la minería de datos, puede hacerse sólo con *Modeler* o programas afines, ya que tiene las herramientas para tratar con varias bases de datos y sobre todo de optimizar los datos, sin una hipótesis u objetivo precisos, pero sí con un planteamiento que guíe el análisis.

Por último, se puede concluir que los softwares analizados proporcionan la oportunidad a investigadores de diversas áreas del conocimiento de explorar ambas plataformas en las que se emplean diversas herramientas en la búsqueda de modelos de predicción. Las herramientas podrían ser complementarias, ya que las investigaciones que inician con una base de datos creada a través de indicadores que tienen su origen en la operacionalización de las variables que se considera tienen dependencia en su relación podrían considerarse para analizarlas a través de *Modeler* en una nueva búsqueda de conocimiento, en conjunto con otras bases de datos relacionadas con el contexto que se analiza.

## Referencias

Aranda, Y. R., & Sotolongo, A. R. (2013). Integración de los algoritmos de minería de datos 1R, Prism E ID3 a PostgreSQL. *Journal of Informa-*

- tion Systems and Technology Management: JISTEM*, 10(2), 389-406.
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. Zanasi, A. (1998). *Discovering Data Mining: From Concept to Implementation*.
- Castañeda, M. B., Cabrera A., Navarro, Y. & de Bries, W. (2010). *Procesamiento de datos y análisis estadísticos utilizando SPSS: Un libro práctico para investigadores y administradores educativos*. Brasil: Edipucrs.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*.
- Di, Z., Yang, Y., Fu, Q., Lin, X., Jiang, S. (2013). Exploiting machine learning for predicting skeletal-related events in cancer patients with bone metastases. *Proceedings - 2013 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2013* 6732633, pp. 31-35
- Galán-Cortina V. (2016). *Aplicación de la metodología CRISP-DM a un proyecto de minería de datos en el entorno Universitario*. (Bachelor's thesis) Universidad Carlos III de Madrid Escuela Politécnica Superior Ingeniería en informática.
- Hernández J, Ramírez M.J. y Ferri, C. (2004). *Introducción a la Minería de Datos*. España: Ed. Pearson Educación, S.A.
- IBM. (2012). *Manual CRISP-DM de IBM SPSS Modeler*. Recuperado de <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>
- IBM. (2013). *IBM SPSS Statistics 22 Core System guía del Usuario*. Recuperado de [ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/22.0/es/client/Manuals/IBM\\_SPSS\\_Statistics\\_Core\\_System\\_User\\_Guide.pdf](ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/22.0/es/client/Manuals/IBM_SPSS_Statistics_Core_System_User_Guide.pdf)
- IBM. (2013). *Manual del usuario del sistema básico de IBM SPSS Statistics 20*. Recuperado de [ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/20.0/es/client/Manuals/IBM\\_SPSS\\_Statistics\\_Core\\_System\\_Users\\_Guide.pdf](ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/20.0/es/client/Manuals/IBM_SPSS_Statistics_Core_System_Users_Guide.pdf)
- IBM. (2016a). *Guía del usuario de IBM SPSS Modeler 18.0*. Recuperado de <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.0/es/ModelerUsersGuide.pdf>
- IBM. (2016b). *Guía del usuario de IBM SPSS Statistics 24 Core System*. Recuperado de [ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/24.0/es/client/Manuals/IBM\\_SPSS\\_Statistics\\_Core\\_System\\_User\\_Guide.pdf](ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/24.0/es/client/Manuals/IBM_SPSS_Statistics_Core_System_User_Guide.pdf)

- Joshig. (2015). El poder que tienen los datos. *Portafolio*.
- León, A., & Castellanos, O., & Vargas, F. (2006). Valoración, selección y pertinencia de herramientas de software utilizadas en vigilancia tecnológica. *Ingeniería e Investigación*, 26 (1), 92-102.
- Lobaina, E. M. R., & Suárez, C.P.R. (2018). Resultados obtenidos en un proceso de minería de datos aplicado a una base de datos que contiene información bibliográfica referida a cuatro segmentos de la ciencia. *Journal of Information Systems and Technology Management : JISTEM*, 15, 1-11.
- Lotfnezhad, H., Ahmadi, M., Roudbari, M., Sadoughi, F. (2015). Prediction of breast cancer survival through knowledge discovery in databases. *Global journal of health science* 7(4), 392-398.
- Mejía, J. (2017). Las ciencias de la administración y el análisis multivariante. Proyectos de investigación, análisis y discusión de resultados. Tomo II. Las técnicas interdependientes. (1ra. Ed.). México: Universidad de Guadalajara.
- Mejía, J. (2018). Análisis estadístico multivariante con SPSS para las Ciencias Económico Administrativas. *Teoría y Práctica de las Técnicas Dependientes*. México: D.R. Cloudbook.
- Molina López, J. M., & García Herrero, J. (2006). *Técnicas de análisis de datos*. Universidad Carlos: Madrid.
- Rodríguez Suárez, Y., Díaz Amador, A. (2011). Herramientas de minería de datos. *Revista Cubana de Ciencias Informáticas*, 3 ( 3-4)
- Rodríguez, D., Pollo-Cattaneo, M. F., Britos, P. V., & García-Martínez, R. (2010). Estimación Empírica de Carga de Trabajo en Proyectos de Explotación de Información. *In XVI Congreso Argentino de Ciencias de la Computación*.
- Shao, Z., Liancheng, W., Han, Z. (2016). A fault line selection method for small current grounding system based on big data. *Asia-Pacific Power and Energy Engineering Conference*, pp. 2470-2474.
- Soto Jaramillo, C. M. (2009). *Incorporación de técnicas multivariantes en un sistema gestor de bases de datos*. Universidad Nacional de Colombia.
- U Fayyad, G. P.-S. (1996). *Data mining and knowledge discovery in databases: an overview, communications of acm*
- Wang, Z., Wen, X., Lu, Y., Yao, Y., Zhao, H. (2016). Exploiting machine learning for predicting skeletal-related events in cancer patients with bone metastases. *Oncotarget* 7(11) 12612-12622.





## CAPÍTULO 2

# *La importancia de la técnica de regresión lineal simple en el área de las ciencias económico-administrativas*

*José Rosario Lara Salazar*

DIRECTOR DE TESIS

*Alejandro Campos Sánchez*

### **Resumen:**

La regresión lineal simple (RLS) es una técnica de predicción utilizada en investigaciones en las que se destacan de carácter cuantitativo. Dicha técnica yace de muchos años de existencia y por ello existe una crítica de que su efectividad no sea la más adecuada para obtener resultados por su antigüedad y obsolescencia. Sin embargo, existe la contraparte que afirma que la RLS funciona efectivamente y que si ha sido una técnica con tanto tiempo de uso se debe a los resultados precisos que esta puede dar. Por ello, se destaca que para este estudio se analiza la RLS y la predicción que esta pueda dar en investigaciones y estudios en el área de las ciencias económico-administrativas, así como la relevancia de su uso.

**Palabras clave:** Regresión lineal, ventajas, desventajas, ciencias económico-administrativas.

### **Abstract:**

Simple linear regression (SLR) is a prediction technique used in investigations in which they stand out in a quantitative way. This technique lies many years of existence and therefore there is a criticism that its effec-

tiveness is not the most appropriate to obtain results due to its age and obsolescence. However, there is a counterpart that states that the SLR works effectively and that if it has been a technique with so much time of use it is due to the precise results that it can give. For this reason, it is highlighted that for this study the SLR is analyzed and the prediction that it can give in research and studies in the area of economic-administrative sciences, as well as the relevance of its use.

**Key words:** Lineal regression, advantages, disadvantages, management sciences.

## Introducción

Entre las técnicas de proyección de un resultado en las ciencias económico administrativas están las regresiones, las cuales pueden ser simples o múltiples.

La regresión es un modelo que permite describir cómo influye una variable sobre la otra (Carollo, 2012). En un modelo de regresión lineal simple, se pretende ver y explicar la relación existente entre la variable de respuesta  $Y$  y una única variable explicativa  $X$ .

Mediante las técnicas de regresión de una variable  $Y$  sobre una variable  $X$ , se busca una función que sea una buena aproximación de una nube de puntos, mediante una curva.

Asimismo, según Carrasquilla-Batista *et al* (2016) la regresión lineal simple es definida como el cálculo de la ecuación correspondiente a la línea que mejor describe entre la respuesta y la variable que la explica, así como en esta técnica se tiene una única variable predictora.

Por ejemplo, en el tema de ventas, puede predecir qué parte de dichas ventas anuales de un vendedor (variable dependiente) es influida por variables independientes como la experiencia o la formación (IBM, 2018).

Es por ello que la regresión lineal en las ciencias económico-administrativas tiene suma relevancia por los resultados que estos pueden dar de ser aplicados en el área de los negocios como posible resultado de utilidades de las organizaciones tal cual como Cardona *et al* (2013) establecen que en la administración pública utiliza un estudio con variables sobre mejora de la efectividad en el clima organizacional de un ente público da una predicción sobre estas variables y con esto la toma de decisiones la cual puede ser si la efectividad es baja, como aumentarla a través de programas, capacitación o mejoras.

## **Desarrollo**

En este apartado se analiza con mayor profundidad la técnica de regresión lineal simple, la cual para un mayor entendimiento se hará una división en un breve análisis histórico, características, análisis bibliométrico así como las ventajas y desventajas de esta técnica.

### ***Historia***

El origen de la regresión lineal, el cual de acuerdo a Palacios-Cruz *et al* (2013) data desde 1805 cuando fue publicada por Lendré en el método de mínimos cuadrados con el que abordaba una versión del teorema de Gauss-Márkov, no obstante, fue el médico sir Francis Galton, primo de Charles Darwin, quien introdujo el término de regresión en 1886.

Así pues, esta técnica ya posee muchos años desde su creación y aplicación, la cual sin duda a través del tiempo ha tenido diversos cambios para irse perfeccionando.

### ***Características***

El modelo de regresión lineal simple, de acuerdo a Cortés *et al* (2015) cuenta con una variable predictora y su relación en forma de línea recta no porque se pretenda que esta relación lineal sea cierta, sino porque se quiere averiguar qué capacidad de anticipación tiene el modelo.

En teoría, una variable pueda tomar valores entre menos y más infinito. Pero en la práctica, los datos proceden de un rango determinado que conviene especificar con claridad.

Para poder hacer un análisis de regresión, se necesita clasificar a las variables en cuáles son las independientes y las dependientes. A su vez, de acuerdo a Mejía-Trejo (2018) es de suma importancia contar con datos métricos para el análisis para un mejor entendimiento de los resultados del estudio.

### ***Análisis bibliométrico***

A continuación, se hace muestra de un análisis bibliométrico que fue realizado con el programa “VOSviewer” utilizando como fuente primaria la base de datos de “Scopus” de los últimos 5 años a la fecha (2013-2018), todo con el objetivo de ver qué autores, en qué países se trabaja el estudio de

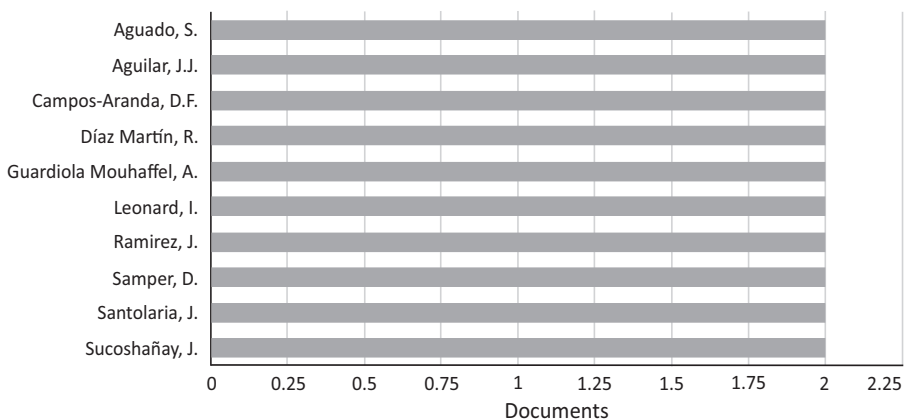
regresión lineal, qué países trabajan en conjunto, qué tipo de documentos son publicados, qué áreas de conocimiento estudian la técnica y el número de artículos publicados por año, con el objetivo de hacer un análisis a profundidad sobre lo que se discute sobre la regresión lineal simple para en este caso, en el área de las ciencias económico-administrativas.

En la gráfica 1 se puede ver que varios de los autores que publican sobre la técnica de regresión lineal son de países habla hispana, dato que se verá posteriormente sobre los países que más publican sobre esta técnica

Continuando con la afirmación dicha, en la gráfica 2 se muestra que los países de habla hispana son los que más publican sobre esta técnica encontrándose España, México, Colombia, Argentina, Chile y Cuba en los lugares 1, 2, 3 ,4 ,5 y 7 respectivamente, lo que indica que hay muchos estudios en español sobre la regresión lineal.

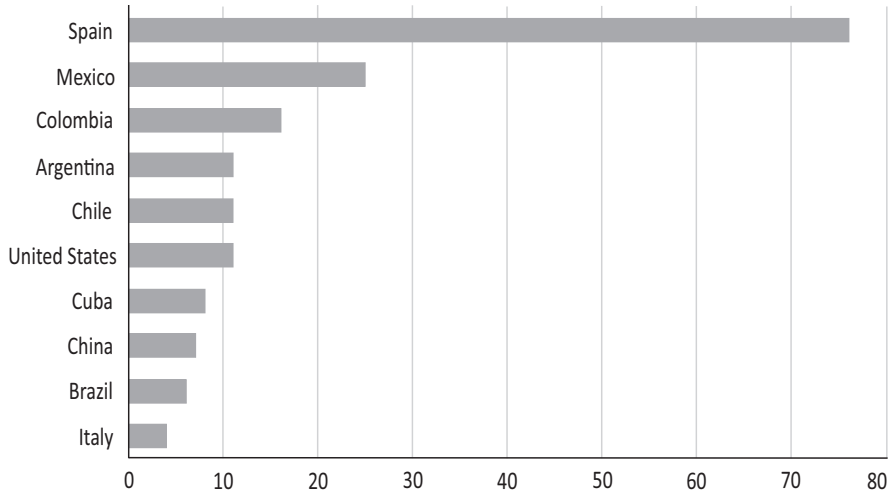
Asimismo, se puede observar que en la gráfica 3 se muestra que todos los países tienen un trabajo con España en lo que respecta a la regresión lineal, sin embargo, hay pequeños clústeres que trabajan entre ellos, por ejemplo, México, Estados Unidos, Reino Unido y Cuba es uno; Brasil, Chile y Alemania; Colombia, Argentina y Australia; y, Perú, Italia y Portugal. Al final todos esos pequeños clústeres trabajan como se mencionó anteriormente con España.

**Gráfica 1. Autores que trabajan regresión lineal**



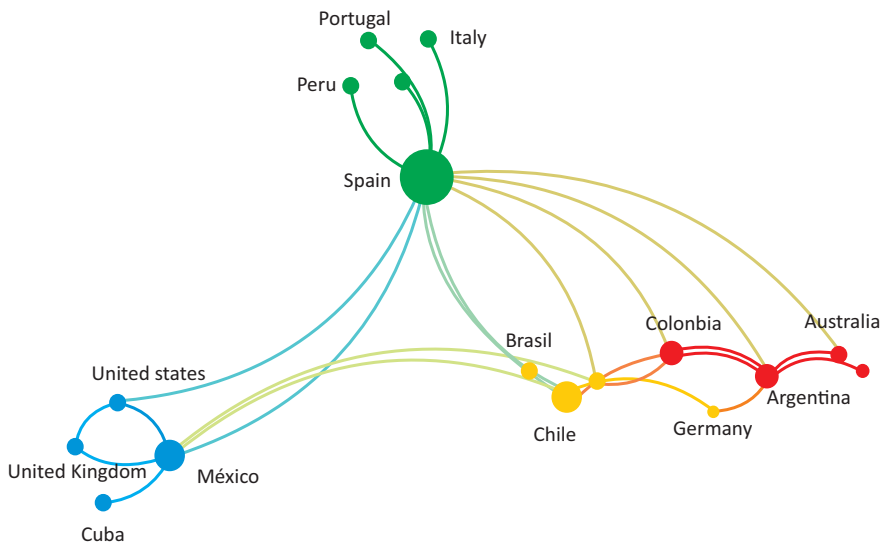
Fuente: Scopus, a través del programa VOSviewer.

**Gráfica 2. Países que trabajan regresión lineal**



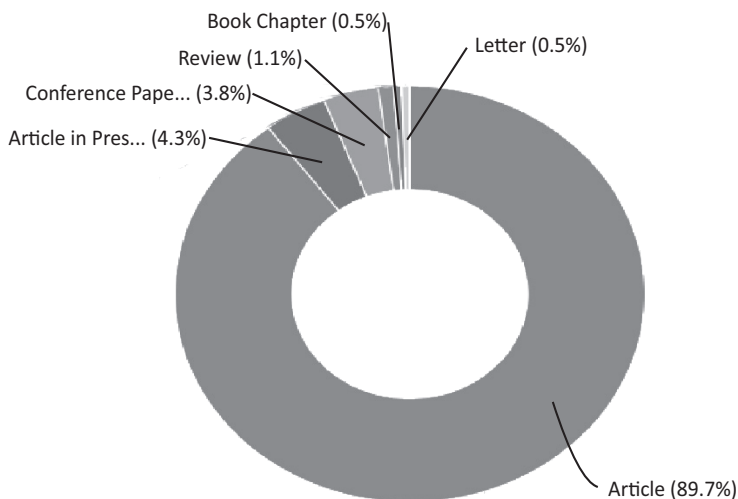
Fuente: Scopus, a través del programa VOSviewer.

**Gráfica 3. Países que trabajan en conjunto con regresión lineal**



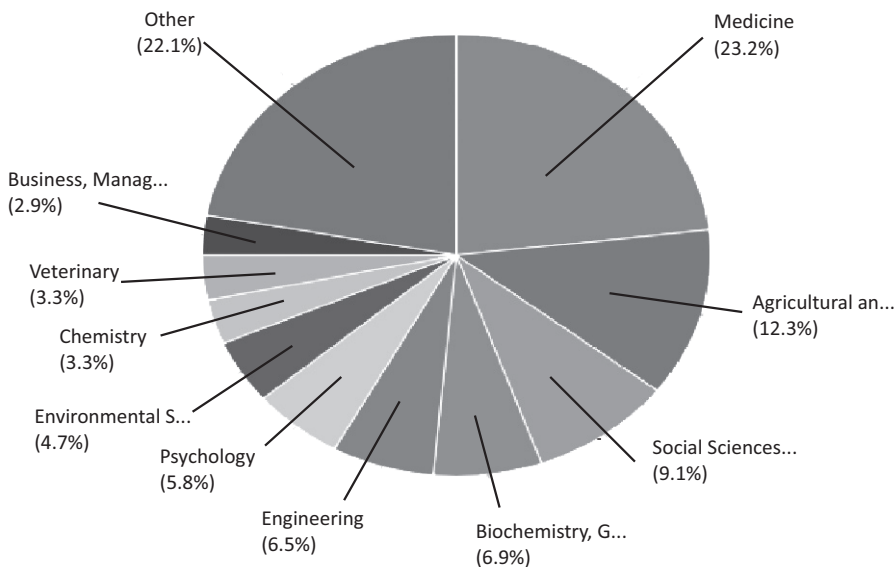
Fuente: Scopus, a través del programa VOSviewer.

**Gráfica 4. Tipo de documento en el se publican trabajos de regresión lineal**



Fuente: Scopus, a través del programa VOSviewer.

**Gráfica 5. Documentos publicados sobre regresión lineal en distintas áreas del conocimiento**



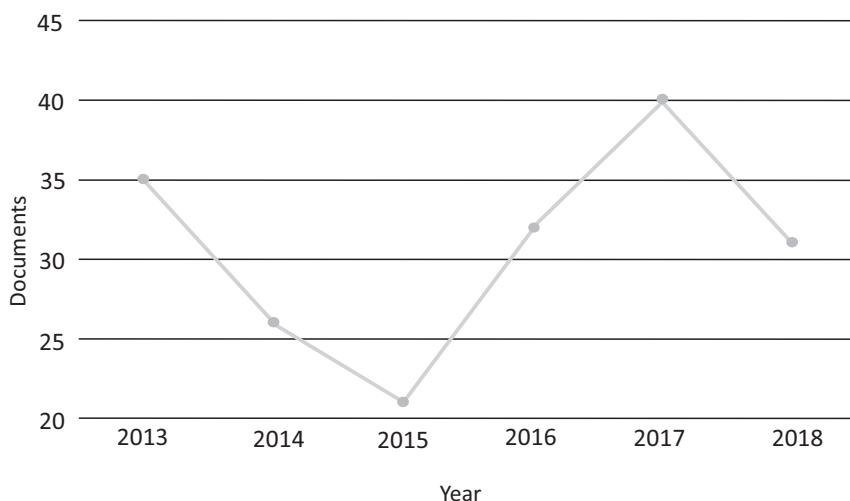
Fuente: Scopus, a través del programa VOSviewer.

Como pudo ser observado en la gráfica 4, la mayoría de los tipos de documentos donde se publica algo relacionado a la regresión lineal es en un 89.7% en artículos, mientras para importancia de este trabajo, solo en un 0.5% en capítulos de libros.

Como fue mostrado en la gráfica 5, el área de negocios y administración solo abarca un 2.9% de toda la publicación de regresión lineal existente en base de datos. Hay un liderazgo en el área de la medicina (23.2%) y agricultura (12.3%). Sin embargo, se ha visto un incremento paulatino, pero existe tal incremento en publicaciones en el área de los negocios y administración.

Finalmente, en la gráfica 6 se observará que en 2015 hubo una baja considerable de artículos publicados sobre regresión lineal, pero a partir de 2016 comenzó a incrementar hasta llegar al pináculo de publicaciones en 2017. No obstante, en el presente 2018 ha habido un decremento de publicaciones a pesar de ser aún varias para este tema.

**Gráfica 6.**  
**Número de artículos que se publican por año de regresión lineal**



Fuente: Scopus, a través del programa VOSviewer.



## ***Ventajas y desventajas***

Como fue mencionado anteriormente, la regresión lineal permite describir cómo influye una variable sobre otra. Sin embargo, es necesario analizar además del uso, la conveniencia de la utilización de dicha técnica. Para ello se hace un contraste sobre los puntos a favor y los puntos en contra de la técnica de regresión lineal simple.

En ese tenor, Pockels-Díaz (2012) afirma que alguna de las ventajas de utilizar la regresión lineal simple son la revelación de información acerca de las estructuras de costos y la distinción entre los roles de las diferentes variables; además que se trata de una herramienta útil para estudiar e identificar las posibles relaciones entre los cambios observados en dos conjuntos diferentes de variables.

Asimismo, es importante destacar que este método así como ofrece ventajas, también lo hacen las desventajas, puesto que existe una identificación de un rango lineal y las variaciones en la linealidad (López-Briceño, 2011); es decir, la misma linealidad no siempre da resultados certeros debido a que muchas cosas durante el proceso llegan a tener variaciones las cuales no son observables en la regresión lineal simple.

Sin embargo, algunos autores difieren sobre dichas desventajas, por ejemplo Rojo (2007) establece que la regresión lineal simple ofrece ventajas que proporcionan más información en la construcción del modelo y esto da como consecuencia a obtener resultados más precisos.

En contraparte, Fallas (2012) establece que la regresión lineal es el tiempo que el investigador requiere para el análisis ya que cuando existen diversas variables predictoras, pueden existir un gran número posibles regresiones así como no siempre las variables predictoras no siempre reflejan su importancia práctica.

Por último, es importante mencionar que de acuerdo a Montero-Granados (2016) la técnica de regresión lineal simple ofrece como ventaja de brindar información sobre la magnitud de la correlación, el incremento marginal, el valor de una de ellas cuando la otra es cero y si dicha relación puede considerarse significativa o fuerte.

## **Conclusión**

La regresión lineal tiene avistamientos desde el siglo XIX y es una técnica que sigue siendo utilizada hasta la fecha. Es notorio que al ser una técnica con muchos años, a lo largo del tiempo han surgido nuevas y diversas

necesidades a las cuales ha llevado al cuestionamiento de la veracidad de los resultados que pueda dar la regresión lineal simple.

Algunos autores defienden la técnica, estableciendo diversas ventajas que pueden obtenerse de su aplicación, pero también existen otros más que afirman que la técnica de regresión lineal al ser antigua y no haber tenido las adaptaciones necesarias para los requerimientos de la actualidad, esta técnica no brinda un resultado confiable de su aplicación.

En conclusión, la utilización de la regresión lineal simple es buena y contrario a ser una técnica que debería de dejarse de utilizar, es una herramienta disponible para cualquier investigador, sin embargo, se puede hacer uso junto con la regresión lineal simple otras técnicas para que estas brinden un plus y a su vez una confiabilidad aún mayor del resultado.

## Referencias

- Cardona-Madariaga, D., González-Rodríguez, J., Rivera-Lozano, M., & Cárdenas-Vallejo, E. (2013). Aplicación de la regresión lineal en un problema de pobreza. *Revista Interacción*, 12, 73-84.
- Carollo, C. (2012). *Regresión lineal simple*. En C. Carollo, Estadística (págs. 1-31). Santiago: Universidad de Santiago de Compostela.
- Carrasquilla-Batista, A., Chacón-Rodríguez, A., Núñez-Montero, K., Gómez-Espinoza, O., Valverde, J., & Guerrero-Barrantes, M. (2016). *Regresión lineal simple y múltiple: aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal*. Tecnología en Marcha, 33-45.
- Cortés, J., Bielsa, N., & Cobo, E. (2015). *Regresión lineal simple*. Barcelona: Universitat Politècnica de Catalunya.
- Fallas, J. (2012). *Regresión lineal simple y múltiple*. Heredia: Universidad Nacional de Costa Rica.
- IBM. (22 de octubre de 2018). Uso de la regresión lineal. Recuperado el 22 de octubre de 2018, de ibm.com: [https://www.ibm.com/support/knowledgecenter/es/SSEP7J\\_10.1.1/com.ibm.swg.ba.cognos.ug\\_cr\\_rptstd.10.1.1.doc/t\\_id\\_task\\_linear.html](https://www.ibm.com/support/knowledgecenter/es/SSEP7J_10.1.1/com.ibm.swg.ba.cognos.ug_cr_rptstd.10.1.1.doc/t_id_task_linear.html)
- López-Briceño, E. (2011). Análisis de regresión lineal para correlacionar datos del valor b en catálogos de sismicidad, obtenidos con dos técnicas. Monterrey: Universidad Autónoma de Nuevo León.
- Mejía-Trejo, J. (2018). *Análisis estadístico multivariante con SPSS para las Ciencias Económico-Administrativas*. México: Cloudbook.

- Montero-Granados, R. (2016). *Modelos de regresión lineal múltiple*. España: Universidad de Granada.
- Palacios-Cruz, L., Pérez, M., Rivas-Ruiz, R., & Talavera, J. (2013). Investigación clínica XVIII del juicio clínico al modelo de regresión lineal. *Revista Médica Instituto Mexicano del Seguro Social*, 656-661.
- Pockels-Díaz, C. (17 de diciembre de 2012). Regresión lineal como técnica más eficiente para la previsión de la demanda. Recuperado el 21 de octubre de 2018, de eoi.es: <http://www.eoi.es/blogs/scm/2012/12/17/regresion-lineal-como-tecnica-mas-eficiente-para-le-prevision-de-la-demanda/>
- Rojo, J. (2007). *Regresión lineal simple*. Madrid: Instituto de Economía y Geografía.

## CAPÍTULO 3

# *Software como herramienta para métodos de prospectiva estratégica: MICMAC*

*Alba Lucia Moreno Ortiz*

DIRECTOR DE TESIS  
*Ariel Vázquez Elorza*

**Palabras clave:** prospectiva, software de análisis para prospectiva, MICMAC.

### **Introducción**

El acelerado movimiento y dinamismo que se experimenta en la vida actual obliga a toda una sociedad a implementar en sus distintos campos, herramientas o métodos para planear estrategias y diferentes alternativas en busca de un mejor porvenir y, tener idea hacia dónde dirigir u orientar todos los esfuerzos.

*“La prospectiva, sea cual sea, constituye una anticipación (preactiva y proactiva) para iluminar las acciones presentes con la luz de los futuros posibles y deseables”, Godet (2007).*

El deseo y anhelo colectivo del hombre por conocer el futuro es un tema que lo ha inquietado e impulsado a buscar caminos para esclarecer lo inesperado y disminuir la incertidumbre ante los sucesos futuros. Los caminos que el hombre ha intentado diseñar e implementar a través de los tiempos son los modelos, estrategias, herramientas, software, técnicas, formulas, entre otros, generando diferentes escenarios, dónde los resultados de cada uno de ellos permite su mejoramiento o exclusión como alternativa para transformar y construir un futuro mejor.

La importancia de los software’s para generar prospectiva, ayudan positivamente ante los desafíos que atraviesa el mundo entero, en sec-

tores económicos, políticos, orden social, ambientales, internacionales, entre otros, como consecuencia de la acelerada evolución de la tecnología. Igualmente ayudan a impulsar a los países con economías en desarrollo, o sectores en crecimiento, generando una visión a futuro, basada en el presente, en la realidad que vive el sujeto o la situación a prospectar.

En este ensayo se pretende explicar el funcionamiento de un software utilizado en la prospectiva como herramienta para visualizar o tener una óptica del futuro, *MICMAC* (Matriz de impactos cruzados multiplicación aplicada a una clasificación).

## Desarrollo

La prospectiva<sup>1</sup> aparece con el filósofo francés Gaston Berger, a mitad de los años cincuenta. Este filósofo desde 1955, tuvo como objetivo principal introducir en su explicación, el futuro para que formara parte de las decisiones del hombre. En los métodos clásicos, su base eran las experiencias y el pasado, lo que no daba lugar a administrar, y los resultados de muchas decisiones son concretados o aplicadas en un lugar diferente a donde se tomaron estas decisiones. Berger no se aparta de la historia ya que están relacionadas la prospectiva y la historia por ser ambos hechos potenciales. Para Berger, la función de la ciencia es que los hombres tengan una visión de diferentes escenarios en un futuro, donde se planteen diferentes situaciones y pueda enfrentarlas. “El pasado, pasado es; el futuro aún no llega (Berger, 1959)” (Godet y Durance, 2009, p.13).

Michel Godet dirige el laboratorio de investigación en Prospectiva, Estrategia y Organización (LIPSOR), su actividad consiste en un grupo de investigadores, profesores, doctorantes, dedicados a las cátedras de prospectiva industrial y desarrollos de sistemas de organización en el Conservatorio Nacional de Artes y Oficios, ubicado en Francia (CNAM), institución pública dedicada a la enseñanza superior, con presencia en más de 150 ciudades de Francia y el extranjero.

La prospectiva, es una herramienta que busca crear futuros deseables y hoy toma relevancia en varios campos de la tecnología, ciencia y

---

1 La prospectiva constituye el arte de la conjetura por antonomasia y es una disciplina que tiene como propósito fundamental la exploración del futuro en el contexto de las ciencias sociales. Para ello se parte de la situación actual de los sistemas analizados y se estudian las circunstancias económicas, sociales, científicas, tecnológicas que ocurren en los mismos, con el propósito de prever e imaginar situaciones que podrían derivarse de influencias conjugadas de tales circunstancias (Gómez, Gómez y Sánchez. 2014. p. 2)

generación de nuevos estudios. La OCDE es citada por Garza y Cortez (2011) para definir la prospectiva como “tentativas sistemáticas para observar a largo plazo el futuro de la ciencia, la tecnología, la economía y la sociedad con el propósito de identificar las tecnologías o métodos emergentes que probablemente produzcan los mayores beneficios económicos y/o sociales” (p. 336).

Actualmente, en las economías del mundo industrializado, los software’s se han convertido en un factor dominante, donde la tecnología requiere de un software específico para su aplicación (Pressman, 2010). Los software’s como programas integrados en un hardware, tienen como objetivo realizar determinadas tareas y hoy vienen revolucionando el mundo entero. Se conocen otros software’s de prospección, LIPSOR (Laboratorio de Investigación en Prospectiva, Estrategia y Organización) y EPITA (Escuela de Informática y Técnicas Avanzadas) y, para el análisis estructural prospectivo.

El profesor Michel Godet recomienda no tener un uso mecanicista con el software o programas ya que son solo herramientas para generar prospectiva estratégica. Prospectivo (2010). Arango y Cuevas (2014), comentan: La caja de herramientas diseñada por Godet, consiste en instrumentos, métodos y técnicas de software para el análisis y la creación de escenarios en el campo del futuro, algunos de ellos son:

- **TALLERES DE PROSPECTIVA** – selección de preguntas y método.
- **MICMAC** – análisis estructural para identificar variable clave.
- **MACTOR** – diferentes actores y sus estrategias, posición, fuerzas, convergencias y divergencias para ser analizadas.
- **MORPHOL** – análisis morfológico.
- **MULTIPOLAR** – evaluar y escoger opción estratégica.
- **SMIC Prob-Expert** – disminuye incertidumbre (entrevistas con expertos).

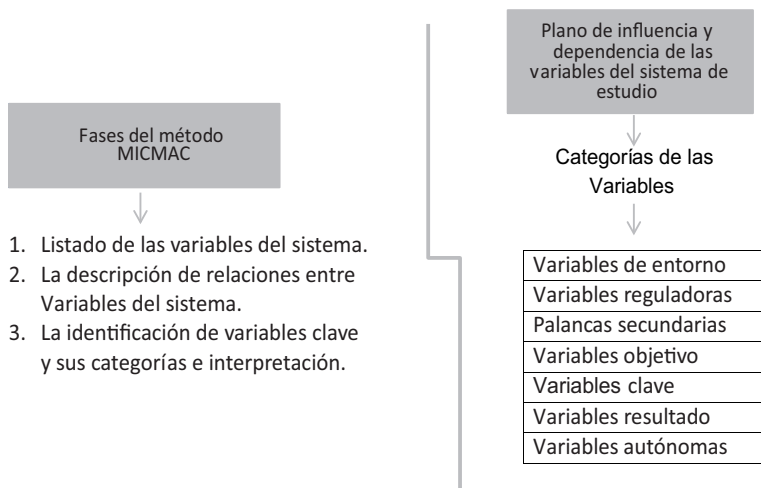
*MICMAC (Análisis estructural prospectivo)* y *MACTOR (análisis del juego de actores)*, son alguno de los programas informáticos de la prospectiva más utilizados.

*MICMAC*, es una herramienta conocida como análisis estructural y su estructuración es de reflexión colectiva. Su análisis es cualitativo, porque las conclusiones finales son dirigidas por expertos, actores y conocedores capaces de ofrecer planteamientos de un sistema en específico.

En cuanto al objetivo consiste en investigar de forma cualitativa una lista de variables estructuradas y una matriz que relaciona todas las variables, donde por medio de cuadros y gráficos, extrayendo las variables claves, genera la construcción de un modelo a futuro del problema de estudio.

Éste software maneja tres fases, igualmente genera una matriz y un esquema con el nombre de plano de influencia y dependencia de las variables y las categoriza. Se muestra en la figura 1.

**Figura 1. Fases del método MICMAC y categoría del Plano de Influencia y dependencia de las variables**



Fuente: Fuente Jairo César Laverde R., adaptación propia.

La siguiente explicación sobre las fases del método *MICMAC*, se apoya en el manual de Aplicación del análisis estructural con *MICMAC*, del docente investigador Jairo César Laverde R.

### ***Listado de las variables del sistema***

Enumerar el total de las variables que definen el entorno y el sistema de estudio, se aconseja ser lo más cuidadoso posible y no excluir ninguna variable para no afectar la investigación. Estas variables son enlistadas por los actores o personas que intervienen en el sistema de estudio para obtener un listado homogéneo tanto de variables internas como externas. No exceder de 70 a 80 variables y es necesario que estén bien detalladas,

para ayudar a una buena relación de las variables, base para la reflexión prospectiva.

**Descripción de relaciones entre las variables**

Las variables deben estar relacionadas entre sí y parte del análisis estructural consiste en generar una relación directa de todas las variables, tarea realizada por el grupo de personas que anteriormente estuvo en la enumeración del listado de variables.

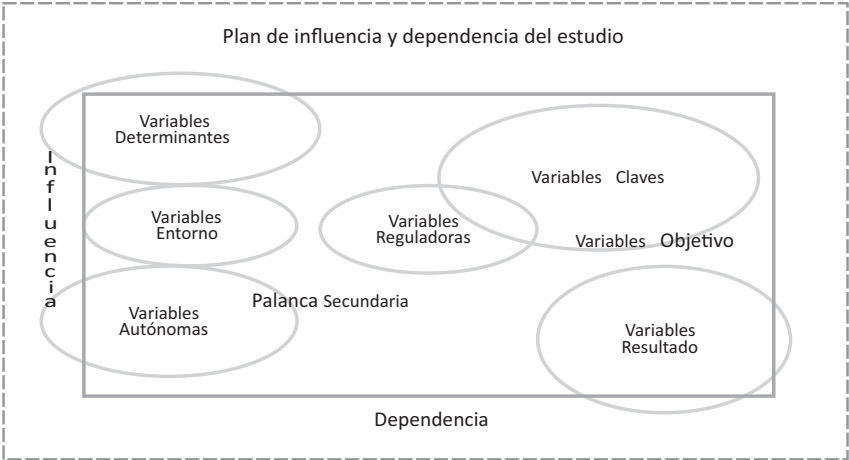
Se realiza un procedimiento de interrogación, para llenar la matriz con diferentes planteamientos, con el objetivo de evitar errores, ordenar y clasificar las ideas, generando un lenguaje común. Esta tarea de llenado es cualitativo.

**Identificación de las variables clave con el MICMAC**

La última parte o fase es identificar la variable clave. Primero se hace clasificación directa y luego una indirecta (**MICMAC**); esta última es resultado de la elevación en potencia de la matriz. La clasificación directa, indirecta y potencial, permite identificar y confirmar la importancia de las variables .

En la figura número 2, se muestra las direcciones que toman las variables en el diagrama de Plan de Influencia y Dependencia del Estudio.

**Figura 2. Dirección de las variables en diagrama de Plan de Influencia y Dependencia del Estudio**



Fuente: Jairo César Laverde R., adaptación propia.



**Tabla 1. Características del plano de influencia y dependencia del estudio**

<b>Variables</b>	<b>Características</b>
VARIABLES DETERMINANTES	Variables que frenan o impulsan su desarrollo a largo plazo, durante la evolución del sistema. Impulsan o inhiben el sistema.
VARIABLES ENTORNO	Escasa dependencia del sistema (decorado del sistema). Estas variables, complementan el valor agregado del sistema.
VARIABLES REGULADORAS	Llave de paso para alcanzar el cumplimiento de las variables claves. En condiciones normales calculan el buen funcionamiento del sistema.
PALANCAS SECUNDARIAS	Son variables complementarias a las variables reguladoras. Estas impulsan la evolución de las variables reguladores y al mismo tiempo afectan la evolución de las variables claves.
VARIABLES OBJETIVO	Son variables dependientes y su impulso es mediano. Sobre ella se puede influir para lograr lo que se desea.
VARIABLES CLAVE	Conocidas como variables reto del sistema. Son variables muy propulsoras y dependientes, afectando el funcionamiento normal del sistema. Mantienen el sistema, son inestables y corresponden a los retos.
VARIABLES RESULTADO	Baja motricidad y alta dependencia. Son variables de indicadores descriptivos de la evolución del sistema, junto a las variables objetivo. Se abordan a través de las variables que dependen en el sistema. Estricto seguimiento para su verificación.
VARIABLES AUTÓNOMAS	Variables poco dependientes e influyentes. Refiere a tendencias pasadas o desconectadas, del estudio. No son determinantes para el futuro y orientan al reto.

Fuente: Elaboración propia, datos de Garza y Cortez (2011).

## **Algunas aportaciones de MICMAC**

En la tabla 2, se presentan cuatro casos, donde ha sido usado el método *MICMAC*, como herramienta para generar escenarios con todas las variables que interfieren en un proceso de estudio. La intervención de expertos en cada tema de investigación, permite que este software, logre abarcar todas las variables que pudieran estar presentes en el futuro, detallando

**Tabla 2. Aportaciones de MICMAC**

<b>Autor</b>	<b>País</b>	<b>Objetivo</b>	<b>Aportación MICMAC</b>
Gómez, Gómez y Sánchez (2014)	España (Castilla León)	Explicar las causas y efectos de los procesos en la agricultura y analizar el diseño y las implementaciones políticas aplicadas en el futuro para el sector (2020).	Búsqueda de las variables claves, para identificar las variables motrices y relevantes. Los resultados fueron 11 variables motrices claves. Los escenarios presentados, permiten analizar las ventajas y desventajas de cada uno de ellos.
Guerrero, Rojas, Torres y Bourdon (2014)	Colombia	Análisis prospectivo para el sector agrícola, a partir de los acuerdos de Paz con la guerrilla y el gobierno (2025)	Se uso el método Delphi y como software para el análisis de información se utilizó MICMAC. Las variables claves de influencia arrojadas fueron: políticas estatales, recursos para inversión, subsidios a los campesinos, capacitación, nuevos modelos de desarrollo, recursos naturales, proceso de paz, entre otros.
Arango, Garza, Cuevas, Leyva y Estrada (2013)	México	Identificar las variables más impactantes en la prestación de servicio en la administración pública municipal en Monterrey, México, que alteran el sistema y con mayor recurrencia de quejas en ciudadanos.	Con el ánimo de mejorar el servicio en la administración pública en Monterrey, las variables motrices o de impacto en las quejas por los ciudadanos fueron: falta de actitud de servicio, tiempo de respuesta, prepotencia, esperar, instalaciones, conflicto de interés, falta de respeto e información de requisitos.
Mendoza, Quintero y Sarmiento (2011)	Colombia	Direccionamiento estratégico, para disminuir la incertidumbre de gestión en las empresas y la participación de los actores de manera proactiva para el futuro.	Se utiliza el método de matriz MICMAC, para validar primeramente las variables de jerarquía en grupos (de poder, enlace, resultados y autónomas). En los resultados, MICMAC detecta variables ocultas que algunas veces ejercen influencia en el problema de estudio.

Fuente: Elaboración propia, datos de Garza y Cortez (2011).

su grado de influencia en el proceso y permitiendo la toma de decisiones estratégicas, partiendo de los diferentes escenarios como resultado de la aplicación de *MICMAC*.

## Conclusiones

El método MICMAC (análisis estructural prospectivo), permite describir y relacionar todos los componentes de una estructura u organismo, a través de una matriz. Una vez se tiene todo el panorama descriptivo de la estructura o tema de estudio, MICMAC, clasifica las variables en las principales o de mayor influencia en el proceso, y las dependientes. La estructura que la matriz genera, se compone de tres pasos: listado de variables, descripción entre variables y la identificación de la variable clave. Dentro de sus resultados, los gráficos y cuadros arrojados, permiten a los expertos crear un modelo para el problema o tema de estudio.

El uso de este software como herramienta para la creación de escenarios en un futuro deseable y posible, es cada día más utilizado por diferentes disciplinas del conocimiento. La confianza se genera ante el fallido intento del método “Planeación a largo plazo”, originado después de la Segunda Guerra Mundial, argumentando que el futuro era predecible; paralelamente en Francia, se empieza a desarrollar la prospección como alternativa para visualizar escenarios a futuro deseables y posibles, generando una matriz donde las variables se estratifican, permitiendo a los expertos diseñar modelos que faciliten la toma de decisiones estratégicas, apuntando al futuro que se planteó desde el inicio.

Se considera que la manipulación de expertos en cada tema, amplía la visión para diseñar modelos, ya que se intenta abarcar e integrar todos los elementos que conforman un sistema. Estos modelos son creados una vez *MICMAC*, arroja una clasificación de variables, indispensables para la generación de estrategias y toma de decisiones, con el objetivo de trabajar en ese escenario a futuro.

La implementación y aplicación de este software, puede ayudar positivamente en diferentes campos de estudio, pues en cualquier sistema u organización la proactividad detecta con tiempo, posibles problemas u obstáculos que pueden ser reparados, corregidos o detenidos, evitando daños mayores y de gran impacto para todo un cuerpo.

## Referencias

- Arango Morales, X. A., Garza Villegas, J. B., Cuevas Pérez, V. A., Leyva Cordero, O., y Estrada Camargo, M. (2013). Estudio exploratorio de variables a través de análisis estructural MIC MAC en la prestación del servicio en la administración pública municipal de Monterrey, México. *Revista internacional administración y finanzas*, 73-88.
- Arango Morales, X., y Cuevas Pérez, V. (2015). *Método de análisis estructural: matriz de impactos cruzados multiplicación aplicada a una clasificación (MICMAC)*. Nuevo León: UANL.
- Garza Villegas, J. B., y Cortez Alejandro, D. V. (2011). El uso del método MICMAC y MACTOR análisis prospectivo en un área operativa para la búsqueda de la excelencia operativa a través del Lean Manufacturing. *Innovaciones de Negocios*, 335-256.
- Godet, M. (2007). *Prospectiva estratégica: problemas y métodos*. Paris: CNAM.
- Godet, M., y Durance, P. (2010). *La prospectiva estratégica para las empresas y los territorios*. Paris: Lipsor.
- Guerrero Sanchez, E., Rojas Peña, A. A., Torres, M. Y., y Bourdon Rojas, N. A. (2014). Plan prospectivo para el desarrollo agrario en las regiones colombianas a partir de posconflicto al año 2025. *Revista Scielo Colombia*, 397-417.
- Gómez Limón, J. A., Gómez Ramos, A., y Sanchez Fernandez, G. (2014). Análisis prospectivo de la agricultura en Castilla y León. *ResearchGate*, 1-18.
- Laverde R., J. C. (s.f.). *Aplicación del análisis estructural con MICMAC*.
- Mendoza Mendoza, A., Quintero Hoyos, I., y Sarmiento Coronado, E. (2011). Aplicación de técnicas prospectivas. *Ingeniare*, 25-36.
- Pressman, R. S. (2010). *Ingeniería del software*. México: Mc Graw Hill.
- Prospectivo, A. e. (2010). Prospectiva blog. Obtenido de Prospectiva.eu: [prospectiva.eu/blog/606](http://prospectiva.eu/blog/606)



## CAPÍTULO 4

# *Análisis comparativo entre regresión lineal múltiple- mínimos cuadrados parciales y su aplicación en las ciencias económico administrativas*

*Hugo César Enríquez García*

DIRECTOR DE TESIS  
*Ricardo Arechavala Vargas*

### **Resumen**

El objetivo de este artículo es el de hacer un comparativo entre la *RLM* (Regresión Lineal Múltiple) y *PLS* (Mínimos Cuadrados Parciales, por sus siglas en inglés), estas son dos herramientas estadísticas multi-variantes que son ampliamente usadas a partir de un conjunto de datos obtenidos a fin de predecir una o varias variables, se hará un breve pasaje histórico de cada una destacando sus características, continuando con una discusión o desarrollo en varios ámbitos de las ciencias pero sobretodo en ciencias económico administrativas, terminando con unas conclusiones que lleven al lector y/o al investigador a entender en que momentos puede hacer uso de estas técnicas sofisticadas que han ayudado a explicar un sinnúmero de fenómenos en ciencias duras y ciencias blandas.

### **Introducción**

En el constructo de un modelo de regresión pudieran presentarse dos cuestiones: uno de multi-colinealidad y otro de alta dimensionalidad en-

tre sus variables explicativas. En este capítulo se revisan dos técnicas multi-variantes relativamente similares y utilizadas en la solución de estos problemas: La regresión de mínimos cuadrados y la *RLM*. Los dos procedimientos convierten a las variables explicativas en componentes ortogonales, mismos que representan la solución al problema de multi-colinealidad y permiten hacer una disminución de la dimensionalidad del espacio de variables explicativas.

La *RLM* es una técnica multi- variante muy utilizada para medir el impacto, interrelación y correlación entre un conjunto de variables predictores ( $\mathbf{X}$ ) y algunas predichas ( $\mathbf{Y}$ ), es muy utilizada en distintos campos de la ciencia para poder explicar un sinnúmero de fenómenos, teorías e hipótesis que los investigadores argumentan. Para dar solución a algunos de los problemas que se presentan en esta herramienta como la multi-colinealidad existen herramientas adyacentes como la *PLS* y que tienen múltiples ventajas y desventajas contra la *RLM* bajo ciertas condiciones, eventos y fenómenos. La *PLS* es una extensión de regresión del análisis de componentes principales, donde se especifica una matriz  $\mathbf{X}$ , que incluye las variables del descriptor y una matriz  $\mathbf{Y}$  (o vector  $\mathbf{Y}$ ) que incluye la variable de respuesta.

Es importante hacer una distinción entre ambas técnicas con el objetivo de entender los beneficios que cada una obtiene al solucionar problemas de microdatos y metadatos en las ciencias económico administrativas, con ello se obtienen respuestas con prontitud al momento de elegir cualquiera de las dos.

## Antecedentes de las técnicas

Los orígenes de la regresión se dan en el año de 1875 por Galton, él había distribuido paquetes de semillas de arvejas dulces a siete amigos; cada amigo recibió semillas uniformes, pero hubo una variación sustancial entre los diferentes paquetes. Los amigos de Galton cosecharon semillas de las nuevas generaciones de plantas y se las devolvieron. Galton trazó los pesos de las semillas hijas contra los pesos de las semillas madre. Galton se dio cuenta de que los pesos medianos de semillas hijas de un tamaño particular de semilla madre describían aproximadamente una línea recta con pendiente positiva menor que 1.0:

*“Por lo tanto, naturalmente, alcanzó una línea de regresión recta, y la variabilidad constante para todas las matrices de un carácter se die-*

*ron por un segundo carácter dado. Quizás fue mejor para el progreso del cálculo correlacional que este simple caso especial se promulgará primero”.*

Stanton (2001) describe que el tamaño de la arveja madre en el eje  $X$  y el tamaño de la arveja en el eje  $Y$  tenían una variabilidad aproximadamente igual. Por lo tanto, la pendiente de la línea que conecta las medias de las diferentes columnas de puntos es equivalente tanto a la pendiente de regresión como al coeficiente de correlación. Para los propósitos de Galton, cualquier pendiente menor de 1.0 indicaba una regresión a la media para esa generación de semillas.

El primer estudio documentado de Galton de este tipo sugirió una pendiente de 0.33 (obtenida a través de una inspección cuidadosa de sus diagramas de dispersión), lo que le indicó que las “semillas madre” extremadamente grandes o pequeñas generaban semillas hijas sustancialmente menos extremas. Este hallazgo es, por supuesto, prototípico de regresión a la media: para muchas variables, los procesos naturales trabajan para “amortiguar” los valores extremos y acercarlos a sus respectivos medios (Stanton, 2001).

Asimismo Madariaga, Rodríguez, Lozano & Vallejo (2013) mencionan que el término regresión fue utilizado por primera vez como un concepto estadístico en 1877 por sir Francis Galton, quien llevó a cabo un estudio que mostró que la estatura de los niños nacidos de padres altos tiende a retroceder o “regresar” hacia la estatura media de la población. Designó la palabra regresión como el nombre del proceso general de predecir una variable (la estatura de los niños) a partir de otra (la estatura del padre o de la madre).

Por otra parte y considerando el otro método de interés, que es la técnica de mínimos cuadrados parciales. Wold, Sjöström & Eriksson (2001) explican que esta técnica se originó alrededor de 1975 por Herman Wold para el modelado de conjuntos de datos complicados en términos de cadenas de matrices (bloques), los llamados modelos de trayectoria. Esto incluía una manera simple pero eficiente de estimar los parámetros en estos modelos llamados *NIPALS* (mínimos cuadrados parciales iterativos no lineales). Esto condujo, a su vez, al acrónimo *PLS* para estos modelos.

La regresión de *PLS* es particularmente útil cuando se necesita predecir un conjunto de variables dependientes a partir de un gran conjunto de variables independientes, es decir, predictores (Abdi, 2003). Se originó en las ciencias sociales (específicamente en economía, Herman Wold 1966) pero se hizo popular por primera vez en quimiometría (es decir,



química computacional) debido en parte al hijo de Herman, Svante (Wold, 2001) y en evaluación sensorial (Martens y Naes, 1989). Pero la regresión *PLS* también se está convirtiendo en una herramienta de elección en las ciencias sociales como una técnica multi- variante para datos no experimentales y experimentales.

Además, hubo una evaluación más reciente que muestra que se utiliza también en las ciencias económico- administrativas, en este respecto Henseler, Ringle y Sinkovics (2009) revisaron a través de importantes bases de datos de publicaciones académicas (por ejemplo, ABI / Inform, Elsevier, Science Direct, Emerald Insight, Google Scholar, Psyc INFO, Swetswise) reveló que más de 30 artículos académicos en el dominio del marketing internacional utilizó el modelado de trayectoria *PLS* como análisis estadístico.

## Características de las técnicas

Mejía (2018) concluye que la RLM es una técnica adecuada cuando el investigador incluye solo una variable métrica dependiente que se asume que está relacionada con una o más variables métricas independientes. Su objetivo es la predicción de cambios en la variable dependiente en respuesta a cambios en las variables independientes.

Weibserg (2005) establece que la regresión es el estudio de la dependencia. Se usa para responder preguntas como: ¿Cambiar el tamaño de la clase afecta el éxito de los estudiantes? ¿Podemos predecir el momento de la próxima erupción del géiser Old Faithful a partir de la erupción más reciente? ¿Los cambios en la dieta dan como resultado cambios en el nivel de colesterol y, de ser así, los resultados dependen de otras características como la edad, el sexo y la cantidad de ejercicio? ¿Los países con ingresos más altos por persona tienen tasas de natalidad más bajas que los países con ingresos más bajos? El análisis de regresión es una parte central de muchos proyectos de investigación.

Mientras que la variable  $Y$  sea predicha por las variables  $X$ , entonces el modelo para las  $K$  variables de la ecuación (1) representa la relación entre las variables, la cual puede ser simple (si  $k=1$ ) o múltiple (si  $k>1$ ), además de lineal o no lineal de acuerdo con los exponentes de las variables intervinientes en el modelo (Uriel, 2013)

Este modelo matemático, tiene una forma general algebraicamente para la estimación de los parámetros y se expresa de la siguiente manera:

Sea la muestra aleatoria:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + U \quad (i=1,2,\dots, n)$$

Donde:

- Los coeficientes beta representan la pendiente de la línea de regresión e indica el número de unidades que aumentará la variable dependiente o criterio por cada unidad que aumente la variable independiente.
- $X_1, X_2 \dots X_n$ . Son las variables independientes o explicativas.
- Y es la variable dependiente.
- El parámetro  $\beta_0$  es la ordenada al origen del modelo (punto de corte con el eje Y, es de dónde parte la línea de regresión).
- “U” representa un segmento de factores sin control pertinentes al modelo que se enmarcan con el nombre de perturbación o error aleatorio.

A su vez Baeza & Vázquez (2014) comenta que esta técnica en proceso con  $X$  variables independientes [ $x_1, x_2, \dots, x_k$ ] que afecten o puedan afectar a una variable dependiente Y, puede ser representado por modelos estadísticos de regresión que permiten explicar y predecir su comportamiento, asumiendo el supuesto de causalidad en las variables analizadas, sin poder demostrarlo; sabiendo esto, entonces, con el modelo adecuado, se pueden realizar estimaciones sobre la variable Y a partir de las variables X.

Sobre esta técnica multi- variante Yakubu (2018) afirma que la importancia de los coeficientes de regresión se prueba con un estadístico “T”, mientras que la bondad de ajuste de la regresión se evalúa utilizando el coeficiente de determinación ( $R^2$ ) y el  $R^2$  ajustado. También debe considerarse que el uso de variables explicativas interdependientes se debe tratar con precaución, ya que se ha demostrado que la multi-colinealidad está asociada con estimaciones inestables de los coeficientes de regresión (Ibe, 1989; Yakubu, 2009).

Por otro lado, la técnica multi- variante *PLS* tiene como objetivo el predecir o analizar un conjunto de variables dependientes a partir de un gran conjunto de variables independientes o predictores. Esta predicción se logra al extraer de los predictores un conjunto de factores ortogonales denominados variables latentes que tienen el mejor poder predictivo.

Dijkstra (2010) alude a esta técnica indicando que su enfoque básico es construir “proxies” para las variables latentes, en forma de compuestos

lineales, por medio de una secuencia de algoritmos de mínimos cuadrados alternos, cada vez que resuelven un problema lineal local, con el objetivo de extraer la información predictiva en la muestra. Una vez que se construyen los compuestos, los parámetros de la forma estructural y reducida se estiman con los sustitutos que reemplazan las variables latentes. La información particular incorporada en la forma estructural no se usa explícitamente en la determinación de los proxies.

Hay dos tipos básicos de algoritmos, llamados modo A y modo B, y un tercer tipo, el modo C, que mezcla estos dos. Cada modo genera un vector de peso estimado  $\mathbf{bw}$ , con un  $\mathbf{bwi}$  sub-vector típico del mismo orden que  $\mathbf{Y}_i$ . Estos vectores de peso son puntos fijos de mapeos definidos algorítmicamente (Dijkstra, 2010).

Renteria & Milidiú (2004) describen que el métodos *PLS* es un proceso de modelado que simultáneamente estima factores subyacentes en ambos tipos de variables  $\mathbf{X}$ ,  $\mathbf{Y}$ . Estos factores son usados para definir o encontrar sub espacios en la variable  $\mathbf{X}$  que son más adecuados de modelar en la variable  $\mathbf{Y}$ .

Ya se ha presentado la metodología estadística de Regresión Lineal Múltiple y *PLS*. En resumen, *PLS* es una extensión de regresión del análisis de componentes principales, donde se especifica una matriz  $\mathbf{X}$ , que incluye las variables del descriptor y una matriz  $\mathbf{Y}$  (o vector  $\mathbf{Y}$ ) que incluye la variable de respuesta. El modelo *PLS* establece la relación entre las matrices  $\mathbf{X}$  e  $\mathbf{Y}$  y se expresa como un conjunto de vectores de puntuación  $\mathbf{X}$ , vectores de puntuación  $\mathbf{Y}$ , vectores de peso  $\mathbf{X}$  y peso  $\mathbf{Y}$ , que definen las dimensiones del modelo *PLS*. Cada dimensión (índice  $a$ ) expresa una relación lineal entre un vector de puntuación  $\mathbf{X}$  y un vector de puntuación  $\mathbf{Y}$ .

Henseler, Ringle y Sinkovics (2009) declaran así que “se recomienda el modelado de rutas *PLS* en una etapa temprana de desarrollo teórico para probar y validar modelos exploratorios”.

## Discusión

En sus estudios en las quimioterapia Wold, Sjöström, & Eriksson (2001) mencionan que si el número de variables latentes es realmente igual al número de variables  $\mathbf{X}$ ,  $\mathbf{K}$ , entonces las variables  $\mathbf{X}$  son independientes, aquí *PLS* y la regresión lineal dan resultados idénticos.

Wold, Sjöström, & Eriksson (2001) vieron el *PLS* como una generalización de la regresión lineal múltiple, que contiene este último como un caso especial en situaciones en las que existe la soluciones de la *RLM*

es decir, cuando el número de variables  $X$  y  $Y$  es bastante pequeño en comparación con el número de observaciones  $N$ . En la mayoría de los casos prácticos, excepto cuando  $X$  se genera de acuerdo con un diseño experimental, las variables  $X$  no son independientes. Entonces llamamos  $X$  rango deficiente. Luego el *PLS* proporciona una solución “AshrunkB” que es estadísticamente más robusta que la solución de la *RLM*, y por lo tanto ofrece mejores predicciones.

*PLS* proporciona un enfoque para el modelado cuantitativo de las relaciones que a menudo son complicadas entre los predictores  $X$  y las respuestas  $Y$ , con problemas complejos casi siempre es más realista que la *RLM*, incluidas las variantes de selección por pasos. Esto debido a las suposiciones que subyacen a *PLS* (correlaciones entre las  $X$ , ruido en  $X$ , errores del modelo) es más realistas que las suposiciones de *RLM* de  $X$  independientes y sin errores.

Los diagnósticos de *PLS*, en particular los gráficos de validación cruzada y puntuación (u, t y t, t) con los correspondientes gráficos de carga, proporcionan información sobre la estructura de correlación de  $X$  que no se obtiene mediante una *RLM* ordinaria. En particular, los resultados de *PLS* que muestran que los datos no son homogéneos (como el ejemplo de AA que se ve aquí), son difíciles de obtener por la *RLM*. En sistemas complicados, las no linealidades son tan fuertes que no se puede construir un solo modelo polinomial, pues parecen ser bastante comunes. Por lo tanto, un enfoque flexible para el modelado frecuentemente se justifica con modelos separados para diferentes clases de tópicos. Y no hay pérdida de información con este enfoque en comparación con el enfoque de modelo único (Wold, Sjöström & Eriksson, 2001).

Fragkaki, Farmaki, Thomaidis, Tsantili, Angelis, *et al* (2012) resuelven que debido al problema de co-linealidad en el análisis de *RLM*, uno debe eliminar los descriptores colineales antes del desarrollo del modelo de *RLM*. Por otro lado, la regresión *PLS* puede manejar los descriptores colineales, por lo tanto, utilizando la información máxima y en consecuencia se pueden obtener mejores modelos predictivos mediante el método *PLS*.

La capacidad de *PLS* para analizar los perfiles de las respuestas, facilita la medición de las medidas de respuesta del dispositivo que son relevantes para el objetivo declarado de la investigación; es más fácil capturar el comportamiento de un sistema complicado con una batería de mediciones que con una sola variable.

Los autores creen que la flexibilidad del enfoque *PLS*, su orientación gráfica y su capacidad inherente para manejar datos incompletos y rui-

dosos con muchas variables (y observaciones) hace de PLS un enfoque simple pero poderoso para el análisis de datos de problemas complicados.

Garson (2016), describe que las ventajas de *PLS* incluyen la capacidad de modelar múltiples dependientes, así como múltiples independientes; capacidad de manejar multi-colinealidad entre las independientes; robustez frente al ruido de datos y datos faltantes; y la creación de variables latentes independientes directamente sobre la base de productos cruzados que involucran la (s) variable (s) de respuesta, haciendo predicciones fuertes. Las desventajas de *PLS* incluyen una mayor dificultad para interpretar las cargas de las variables latentes independientes (que se basan en relaciones entre productos con las variables de respuesta, no se basan en el análisis de factores comunes sobre covarianzas entre las manifestaciones de las independientes) y debido a las propiedades de distribución de las estimaciones no se conocen, el investigador no puede evaluar la importancia, excepto a través de la inducción “*bootstrap*”.

Garson (2016) comenta que en general, la combinación de ventajas y desventajas significa que *PLS* se favorece como técnica predictiva y no como técnica interpretativa, excepto para el análisis exploratorio como preludeo a una técnica interpretativa como la regresión lineal múltiple o el modelado de ecuaciones estructurales basadas en covarianza.

*PLS* tiene la capacidad de modelar y analizar varias *Y* juntas, lo que tiene la ventaja de ofrecer una imagen general más simple que un modelo separado para cada variable *Y*. Por lo tanto, cuando las *Y* están correlacionadas, deben analizarse juntas. Si la *Y* realmente mide diferentes cosas, y por lo tanto son bastante independientes, un solo modelo *PLS* tiende a tener muchos componentes y es difícil de interpretar.

La popularidad de *PLS* Parece que ha aumentado en gran medida en el nuevo milenio. Hay Journals que han publicado aplicaciones *PLS* en el dominio de las ciencias de la administración, pues los autores Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009) presentan un resumen de dichas publicaciones (Abreviatura de la revista y número de estudios entre paréntesis):

- ▶ Avances en Marketing Internacional (AIM; 1)
- ▶ Teoría y práctica del emprendimiento (ETP; 1)
- ▶ Revista Europea de Marketing (EJM; 2)
- ▶ Dirección de Marketing Industrial (IMM; 1)
- ▶ Revisión de Negocios Internacionales (IBR; 2)
- ▶ Revista Internacional de Comercio Electrónico (IJEC; 1)

- ▶ Revista Internacional de Organización Industrial (IJIO; 1)
- ▶ Revista Internacional de Marketing de Investigación (IJRM; 1)
- ▶ Revisión de Marketing Internacional (IMR; 2)
- ▶ Revista de Investigación de Negocios (JBR; 2)
- ▶ Diario de comportamiento del consumidor (JCR; 1)
- ▶ Revista de psicología económica (JEP; 1)
- ▶ Revista de investigación de comercio electrónico (JECR; 1)
- ▶ Revista de Estudios de Negocios Internacionales (JIBS; 3)
- ▶ Revista de Marketing Internacional (JIM; 2)
- ▶ Revista de Gestión de la Innovación del Producto (JPIM; 1)
- ▶ Revista de la Academia de Ciencias de Marketing (JAMS; 1)
- ▶ Diario de Negocios del Mundo (JWB; 2)
- ▶ Revisión Internacional de Gestión (MIR; 2)
- ▶ Ciencias de la Gestión (Mgmt.Sc.; 1)
- ▶ Schmalen bach Business Review (SBR; 1)
- ▶ Revista de Gestión Estratégica (SMJ; 3)

En las investigaciones de Marketing Internacional, el método *PLS* proporciona un marco poderoso para estimar modelos causales con variables latentes y sistemas de ecuaciones simultáneas con errores de medición.

Henseler, Ringle, & Sinkovics, (2009) hicieron una revisión crítica de la aplicación *PLS* en el marketing internacional, donde revelan que esta metodología ha aumentado en popularidad, especialmente para los análisis *PLS* de resultados multi-grupo para diferentes naciones. Además, introdujeron varios avances en la evaluación de estimaciones de modelos *PLS*, como un enfoque novedoso para la comparación de múltiples grupos (*PLS-MGA*).

Por su parte, los autores Calantone, Graham & Mintu-Wimsatt (1998) observan que las estimaciones de los parámetros de *PLS* revelan una mejor fuerza y dirección (por ejemplo en valores positivos vs. negativos) de las relaciones entre las variables en comparación con los coeficientes de correlación, asimismo destacan que la *PLS* evita los sesgos de estimación de parámetros comunes en el análisis de regresión.

También, es de resaltar lo que explican los autores Birkinshaw, Morrison & Hulland, (1995), versan que la regresión *PLS* es más apropiada cuando los tamaños de muestra son pequeños, cuando no se pueden hacer

suposiciones de normalidad multivariada y los datos de escala de intervalo, además cuando el investigador está más interesado principalmente en la predicción de la variable dependiente.

## Conclusiones

Ambos tipos de metodologías de regresión son muy utilizadas en múltiples ámbitos de las ciencias, hoy en día en las ciencias económicas y administrativas son dos herramientas fundamentales para la realización de estudios de diferente complejidad, aunque la regresión lineal múltiple es más común, debido probablemente a que sea un poco más antigua, sin embargo, la regresión *PLS* cada vez está siendo más utilizada en función de la efectividad de los resultados obtenidos, es por ello que cada vez tiene más auge y como se mostró con anterioridad, hay una cantidad interesante de revistas de las ciencias económico y administrativas que cada vez más muestran publicaciones, por tal razón se espera que siga siendo una técnica usada con mayor frecuencia.

La regresión lineal múltiple es más recomendada para los tipos de estudios descriptivos, explicativos y/o correlacionales donde la interrelación de variables, dimensiones y mediciones cuantitativas sean más conocidas o establecidas en estudios previos, donde quizás sean modelos en los cuales ya existen relaciones semejantes entre las variables dependientes e independientes y que se le pueden agregar algunas variables más pero digamos que no son estudios exploratorios o totalmente nuevos.

En virtud de la información revisada y analizada se puede decir que la utilización de la regresión *PLS* es especialmente más adecuada para estudios o investigaciones exploratorias donde las medidas son nuevas y las relaciones no han sido demostradas anteriormente. Es decir que su uso o empleo se justifica cuando el marco teórico no está suficientemente argumentado y las variables o dimensiones no se ajustan a un modelo de medición rigurosamente especificado, o bien no se ajustan a una distribución determinada.

Es de llamar la atención también, que los autores mencionados se inclinan más por la utilización del modelo de regresión *PLS*, pues consideran que resuelve algunos de los problemas o adversidades que ya fueron explicados y que pueden presentarse en la utilización de la *RLM*. Esto la convierte en una metodología atractiva y que debería incluirse más a menudo en las paqueterías para poder correr los modelos.

## Referencias

- Abdi, H. (2003). Partial least square regression (PLS regression). *Encyclopedia for research methods for the social sciences*, 6(4), 792-795.
- Alberto, P., & González, F. (2012). Partial Least Squares regression on symmetric positive-definite matrices. *Rev Col Estad*, 36(1), 177-192.
- Birkinshaw, J. Morrison, A & Hulland, J. (1995). Structural and competitive determinants of a global integration strategy. *Strategic Management Journal*, 16(8), 637-655.
- Calantone, R.J. Graham, J.L & Mintu- Wimsatt, A. (1998). Problem-solving approach in an international context: Antecedents and outcomes. *International Journal of Researching. Marketing*, 15(1), 19-35.
- Dijkstra, T. K. (2010). Latent variables and indices: Herman Wold's basic design and partial least squares. In *Handbook of partial least squares* (pp. 23-46). Springer, Berlin, Heidelberg.
- Fragkaki, A. G., Farmaki, E., Thomaidis, N., Tsantili-Kakoulidou, A., Angelis, Y. S., Koupparis, M., & Georgakopoulos, C. (2012). Comparison of multiple linear regression, partial least squares and artificial neural networks for prediction of gas chromatographic relative retention times of trimethylsilylated anabolic androgenic steroids. *Journal of Chromatography A*, 1256, 232-239.
- Garson, D. (2016). Partial least squares (PLS-SEM).
- Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In *New challenges to international marketing* (pp. 277-319). Emerald Group Publishing Limited.
- Ibe, S. N. 1989. Measures of size and conformation in commercial broilers. *J Anim Breed Genet* 106: 461- 469.
- Madariaga, D. F. C., Rodríguez, J. L. G., Lozano, M. R., & Vallejo, E. H. C. (2013). "Aplicación de la regresión lineal en un problema de pobreza". *Interacción*, 12, 73-84.
- Martens, H, & Naes, T. (1989). *Multivariate Calibration*. London: Wiley
- Mejía, J (2018). "Análisis estadístico multivalente con SPSS para las ciencias económico administrativas. Teoría y práctica de las técnicas dependientes". ISBN en trámite.
- Baeza-Serrato, R., & Vázquez-López, J. A. (2014). "Transición de un modelo de regresión lineal múltiple predictivo, a un modelo de regresión no lineal simple explicativo con mejor nivel de predicción: Un enfoque de dinámica de sistemas". *Revista Facultad de Ingeniería*, 71(71), 59-71.



- Hwang, J. G., & Nettleton, D. (2003). Principal components regression with data chosen components and related methods. *Technometrics*, 45(1), 70-79.
- Renteria, R & Milidiú, R (2004). PLS based regression algorithms and their use in multi agent systems. MCC18/04 June, 2004.
- Stanton, J. M. (2001). Galton, Pearson, and the peas: A brief history of linear regression for statistics instructors. *Journal of Statistics Education*, 9(3).
- Uriel, E. (2013). *Regresión lineal múltiple: estimación y propiedades*. Universidad de Valencia Versión, 09-2013.
- Vega-Vilca, J. C., & Guzmán, J. (2011). Regresión PLS y PCA como solución al problema de multicolinealidad en regresión múltiple. *Revista de Matemática Teoría y Aplicaciones*, 18(1), 09-20.
- Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2), 109-130.
- Yakubu, A. (2009). Fixing collinearity instability in the estimation of body weight from morpho-biometrical traits of West African Dwarf goats. *Trakia Journal of Sciences* 7: 61-66.
- Yakubu, A., & Idahor, K. O. (2018). Using factor scores in multiple linear regression model for predicting the carcass weight of broiler chickens using body measurements | . *UDO Agrícola*, 9(4).
- Weisberg, S. (2005). *Applied linear regression* (Vol. 528). John Wiley & Sons.

## CAPÍTULO 5

# *Regresión lineal aplicada: análisis de la responsabilidad social empresarial*

*Itzel Alejandra Lara Manjarrez*

DIRECTOR DE TESIS  
*Rogelio Rivera Fernández*

### **Resumen**

El método de regresión lineal es una de las técnicas de análisis multivariante con enfoque dependiente más usadas debido a la versatilidad de la misma y su fácil aplicación en diferentes ciencias. Debido a esto se hace un análisis descriptivo de diferentes casos en donde se utiliza la regresión lineal como técnica para analizar diversas variables de la responsabilidad social empresarial, así como los resultados obtenidos con el método cuantitativo.

**Palabras clave:** Regresión lineal, responsabilidad social empresarial.

**Abstract:** The linear regression method is one of the multivariate analysis techniques with dependent approach most used due to the versatility of it and its easy application in different sciences. Due to its easy application, a descriptive analysis is made of different cases in which linear regression is used as a technique to analyze various variables of corporate social responsibility, as well as the results obtained with the quantitative

method.

**Keywords:**Lineal regression, corporate social responsibility

## Introducción

Dentro de las técnicas de análisis multivariante con enfoque dependiente el método de regresión lineal, es el mayormente utilizado por su adaptabilidad y versatilidad en el tema de aplicación (Mejía Trejo, 2018).

La técnica de la regresión lineal consta de una conjunción de métodos estadísticos que son utilizados cuando la variable de respuesta y la variable predictiva son continuas y se quiere pronosticar valores de la variable de respuesta en relación con los valores observados en la variable predictiva.

En este sentido se encuentra que el objetivo principal de la regresión lineal es el establecer las distribuciones bidimensionales, es decir el conocer la influencia de una variable sobre la otra, analizando la relación causa y efecto que se da entre ambas (Enciclopedia virtual, 2018).

En el presente ensayo se mostrará un análisis bibliométrico, el cual consta de un método documental que busca mostrar la producción científica, así como conocer la dinámica y la estructura de los mismos (González de Dios, Moya, & Mateos Hernández, 1997). Este fue realizado a través del programa estadístico Vosviewer, con información extraída de la plataforma de bases de datos de Scopus, sobre todos los artículos referentes al tema en los últimos 5 años a fin de analizar la afluencia de la técnica en la actualidad y ver en qué forma lo están utilizando los autores en sus trabajos de investigación.

A través de dicho análisis se puede observar la gran cantidad de trabajos y autores que están manejando el tema de regresión lineal como un elemento estadístico en sus investigaciones a nivel mundial, así como cuáles son las áreas en dónde más se utiliza el método y cuáles son los países que lo han trabajado más.

Del mismo modo, se analiza la técnica de la regresión lineal utilizada en investigaciones realizadas en el tema de la responsabilidad social empresarial, así como la manera en que los autores han tratado dicha técnica en función de conocer la relación de las variables que se quieran analizar, así como los resultados que se quieren obtener.

La importancia del ensayo recae en conocer la importancia de la adopción de la técnica de la regresión lineal en el estudio de la responsabilidad social empresarial, ya que se ha detectado que dicho tema se ha abordado

más preferentemente como un tema cualitativo que cuantitativo. Se busca conocer los alcances a los que se puede llegar al hacer uso de la técnica en un tema poco abordado de forma cuantitativa.

### ***La técnica***

La técnica de la regresión lineal consta de una conjunción de métodos estadísticos que son utilizados cuando la variable de respuesta y la variable predictiva son continuas y se quiere pronosticar valores de la variable de respuesta en relación con los valores observados en la variable predictiva. Ésta técnica consiste en adaptar un modelo a los datos que ya se tienen, haciendo una estimación de coeficientes en función de las observaciones, con el objetivo de pronosticar valores de la variable de respuesta a partir de una variable, la cual puede ser predictiva o explicativa. Al ser esto mismo pero con más de una variable resulta lo que es la regresión múltiple (Vinuesa, 2016).

Asimismo la regresión lineal se utiliza en la estadística para diferentes aspectos, tales como reconocer las variables predictivas relacionadas con la variable de respuesta; detallar la forma de la relación que se suscita entre las variables y para originar una función matemática óptima que modele dicho vínculo; pronosticar la variable de respuesta en función de las variables explicativas o predictoras (Montgomery, Peck, & Vining, 2006).

Dentro de las técnicas de análisis multivariante con enfoque dependiente el método de regresión lineal, es el mayormente utilizado por su adaptabilidad y versatilidad en el tema de aplicación (Mejía Trejo, 2018).

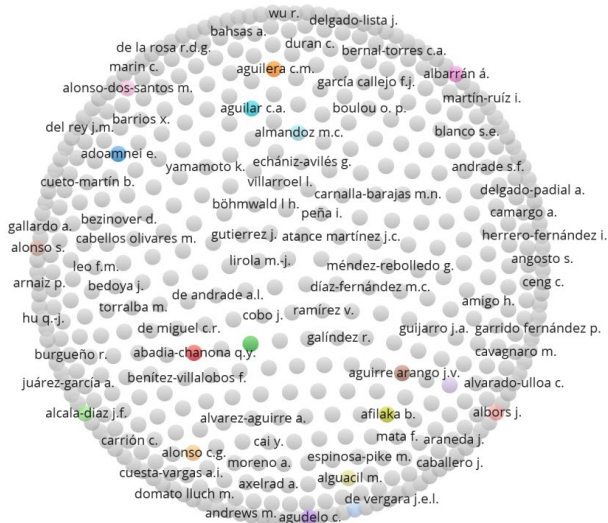
### ***Análisis bibliométrico***

A fin de visualizar de manera más gráfica la información, se realizó un análisis bibliométrico en el programa estadístico Vosviewer, a través de la plataforma de bases de datos de Scopus de todos los artículos referentes al tema en los últimos 5 años a fin de analizar la afluencia de la técnica en la actualidad y ver en qué forma lo están utilizando los autores en sus trabajos de investigación.

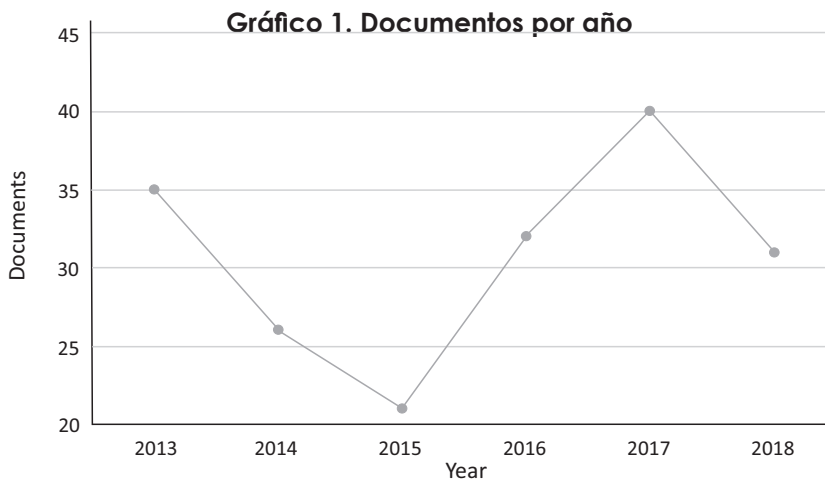
A través de dicho análisis se puede observar la gran cantidad de trabajos y autores que están manejando el tema de regresión lineal como un elemento estadístico en sus investigaciones a nivel mundial. La imagen 1 muestra de manera muy visual los autores que están introduciendo

regresión lineal del año 2013 al 2018.

**Imagen 1. Autores que están trabajando con regresión lineal**



Asimismo se muestran los artículos que han sido publicados según el año, donde se puede observar que la gráfica muestra niveles altos de publicaciones en el año 2013; posteriormente desciende a nivel considerable en el 2014, para terminar en el nivel más bajo de publicaciones en el año 2015 con aproximadamente 20 documentos.



Fuente: Scopus.

Otro aspecto importante a considerar es cómo se ha estado presentando la técnica de la regresión lineal a nivel mundial, haciendo un análisis de los trabajos realizados según la zona geográfica y qué países son los que han trabajado más ese tópico.

En la imagen 2, se puede apreciar a España con el círculo de mayor tamaño, lo cual indica que es el país con mayor número de autores que han trabajado o se encuentran trabajando con el método de la regresión lineal, seguido por México y países de América Latina como Colombia, Argentina, Chile.

**Imagen 2. Análisis bibliométrico por países**

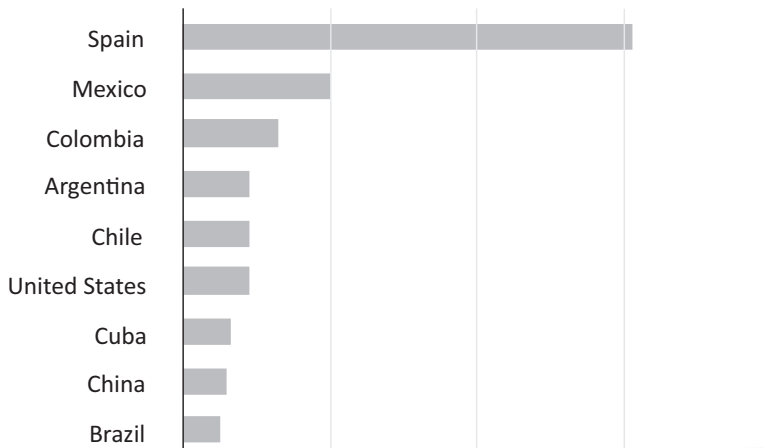


Fuente: Scopus.

La información presentada igualmente se puede ver mayormente desglosada en el gráfico 2.

Del mismo modo se analiza el tema desde las distintas áreas de conocimiento desde las que se ha abordado el tema de la regresión lineal como estadístico en los temas de investigación. En el gráfico 4 se puede observar que el área en el que se maneja mayormente la regresión lineal es en medicina, seguido por agricultura y en tercer lugar se encuentran las ciencias sociales con 9.1% del total de los trabajos realizados con dicha

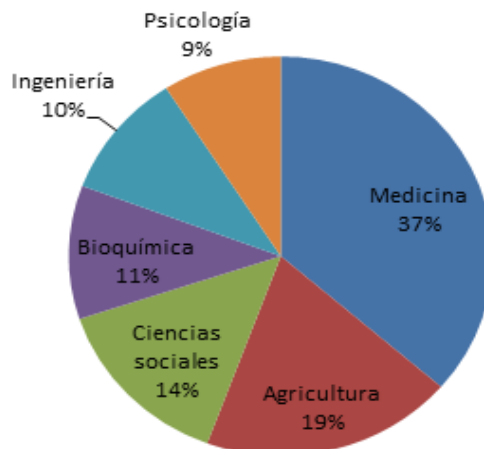
**Gráfico 2. Documentos por países**



Fuente: Scopus

Viendo cada uno de estos análisis y gráficos que muestran estadísticamente la utilización del método de la regresión lineal se puede apreciar la versatilidad mencionada con anterioridad del método, ya que como lo muestra el gráfico 3, es una técnica que puede utilizarse en las diferentes áreas del conocimiento sin distinción al querer realizar un análisis estadístico de sus instrumentos de medición, o bien cualquier análisis de tipo numérico que se emplee en el estudio.

**Gráfico 3. Documentos por área de conocimiento**



Fuente: Scopus

herramienta.

### ***La regresión lineal como herramienta de análisis de datos en la responsabilidad social***

Haciendo alusión a la versatilidad antes mencionada del empleo de la técnica de análisis multivalente con enfoque dependiente de la regresión lineal, se puede hacer notar la carencia de trabajos cuantitativos en el tema, ya que la responsabilidad social empresarial ha sido tratada de manera más notable como un elemento cualitativo ya que por lo regular se hace análisis en torno a la percepción del sujeto de investigación. Dicho lo anterior, se analiza cómo se ha utilizado la técnica de la regresión lineal en la rama de la responsabilidad social empresarial, así como cuales han sido las herramientas más utilizadas dentro de la misma técnica, a fin de poder determinar los resultados posibles que se pueden generar con dicho análisis.

### ***Casos de estudio seleccionados:***

A continuación se muestran algunos casos de estudios en las cuales se utiliza el método de regresión lineal para probar sus hipótesis y teorías, obteniendo con eso información valiosa de representatividad para la investigación. Los casos fueron seleccionados tomando en consideración las variables utilizadas, así como el tratamiento que se le dio a las mismas, obteniendo con estas una visión más amplia en la que se puede destacar los diferentes usos de la técnica en relación a un mismo tema.

En este sentido, se analiza la investigación realizada por Prado-Lorenzo, Gallego-Álvarez, García-Sánchez, y Rodríguez-Domínguez, (2008), la cual tuvo como objetivo el analizar las estrategias de Responsabilidad Social Empresarial (*ESR*) implementadas por compañías españolas y establecer una relación entre estas y el desempeño financiero corporativo.

Para encontrar esta relación propuesta desde un inicio, se llevó a cabo el procesamiento de la información en dos etapas. En un primer momento se realizó un análisis de contenido y la utilización de diversas pruebas de medias. Posteriormente se procedió a realizar una análisis de las motivaciones que llevan a la implementación de *RSE* por medio de la técnica de modelo dependiente de regresión lineal múltiple a fin de encontrar el efecto existente de dichas prácticas en el desempeño de la organización.

Del mismo modo se analiza la investigación realizada por Sánchez y Benito-Hernández (2013), la cual parte del objetivo de desarrollar nuevos



enfoques de sensibilidad hacia el trabajo desde una perspectiva de ética, valores y responsabilidad social empresarial vinculado con las buenas relaciones con los empleados.

Para en análisis de la información, se realizó una regresión lineal de dos modelos, los cuales fueron probados para muestras paralelas y complementarias, a fin de estudiar los efectos de diferentes estrategias de responsabilidad social empresarial en los empleados de micro y pequeñas empresas de España.

Los modelos mencionados anteriormente se encuentran constituidos de la siguiente manera:

- **Modelo 1:** Producción del empleado como medida de productividad laboral, llevado a cabo con pequeñas empresas, es decir con menos de 49 empleados cada una.
- **Modelo 2:** Ventas por empleado como medida de productividad laboral, realizado en microempresas, es decir las que cuentan con menos de 10 empleados cada una.

En el análisis de la regresión se obtuvo que ambos modelos muestran los mismos resultados independientemente de la manera en que se tomó la variable dependiente.

Para el estudio se obtuvo como resultado que las políticas de responsabilidad social empresarial enfocadas a los aspectos externos, tales como la relación de las partes interesadas o stakeholders y la preocupación por el medio ambiente no tuvieron mayor relación, por lo cual en la investigación se catalogaron como que poseían una relación negativa. Sin embargo se pronóstica que dicha relación se pueda volver positiva en el largo plazo, por lo cual la investigación deja esta línea abierta a futuros estudios que puedan comprobar dicha correlación entre las variables ya descritas.

Del mismo modo se encuentra la investigación realizada por Valenzuela Fernández, Jara-Bertin, y Villegas Pineaur (2014), la cual analiza la adopción de estrategias de divulgación voluntaria de *RSE* y la incidencia en el desempeño financiero y su reputación corporativa. Para dicho análisis se hizo uso de la técnica de regresión lineal a fin de encontrar correlaciones entre las variables ya expuestas.

Para la ejecución se hizo una combinación de series de tiempo por seis años y haciendo uso de datos de sección cruzada para 55 empresas. Esto obtuvo como resultado la estimación de variables inobservables lo cual se convierte en un elemento aleatorio del modelo estimado. Lo anterior se convierte en un asunto de importancia ya que se encontraría una relación

entre la variable independiente y el modelo aleatorio.

De igual manera Testera Fuentes y Cabeza García (2012) cuyo objetivo es el indagar en los determinantes de la divulgación de información referente a la *RSE* de las empresas. Para su análisis se llevó a cabo una regresión lineal a fin de encontrar una relación entre la identidad del accionista principal de la organización y la transparencia en la responsabilidad social empresarial.

Como resultado se obtuvo una relación negativa entre la identidad del accionista principal de la organización y la transparencia informativa en la *RSE*. Sin embargo se encontró que el tamaño de la empresa, así como el nivel de endeudamiento y el nivel de transparencia en general se encuentran relacionados directamente con la transparencia de las acciones de responsabilidad social.

A continuación cuadro 1, muestra el comparativo en el que se muestran los casos planteados con anterioridad, así como la forma en la que fueron tratadas las variables y las conclusiones obtenidas a través de la aplicación de la técnica de regresión lineal.

**Cuadro 1. Comparativo de casos**

Autor:	Variables:	Tipo de relación:	Conclusiones:
Prado-Lorenzo, Gallego-Álvarez, García-Sánchez, y Rodríguez-Do-mínguez, (2008)	<ul style="list-style-type: none"> <li>• Responsabilidad social empresarial</li> <li>• Comportamiento financiero empresarial</li> </ul>	Positiva	El uso de RSE dentro de las organizaciones como promover los derechos de los trabajadores, el invertir en espacios publicos, etc. Trajo consigo un impacto positivo en la productividad de la empresa

<p>Sánchez y Benito-Hernández (2013)</p>	<ul style="list-style-type: none"> <li>• RSE</li> <li>• Relaciones con los empleados de micro y pequeñas empresas</li> </ul>	<p>Negativa</p>	<p>Los stakeholders y la preocupación por el medio ambiente no presentaron mayor relación. Sin embargo se pronóstica que dicha relación se pueda volver positiva en el largo plazo.</p>
<p>Valenzuela Fernández, Jara-Bertin, y Villegas Pineaur (2014)</p>	<ul style="list-style-type: none"> <li>• Divulgación voluntaria de RSE</li> <li>• Desempeño financiero</li> <li>• Reputación corporativa</li> </ul>	<p>Positiva</p>	<p>Se obtuvo como resultado que la divulgación voluntaria de RSE trajo consigo un aumento en el desempeño financiero de la empresa así como una mejora en la reputación corporativa</p>
<p>Testera Fuentes y Cabeza García (2012)</p>	<ul style="list-style-type: none"> <li>• Identidad del accionista principal de la organización</li> <li>• Transparencia en la responsabilidad social empresarial</li> </ul>	<p>Negativa</p>	<p>Se obtuvo una relación negativa entre la identidad del accionista principal de la organización y la transparencia informativa en la RSE. Sin embargo se encontró que el tamaño de la empresa, así como el nivel de endeudamiento y el nivel de transparencia en general se encuentran relacionados directamente con la transparencia de las acciones de responsabilidad social.</p>

## Conclusiones

El análisis de regresión lineal es una técnica de gran utilidad debido a la versatilidad que presenta en relación al área de estudio a la cual se puede dirigir. Por tal motivo es una de las técnicas de análisis multivariante de mayor uso en la actualidad.

En este tenor se analiza la forma en como diversos autores han hecho uso del análisis de regresión lineal para tratar temas de responsabilidad social empresarial, permitiendo ver como un mismo tema puede ser abordado desde diferentes perspectivas y, que siendo la misma técnica se pueden encontrar diversas formas de ser aplicada, dependiendo de los resultados que se pretenda obtener.

Dicho análisis fija la pauta para futuras investigaciones en las cuales se planea hacer uso de la técnica y poder adaptarla según sean las características propias de la investigación.

## Referencias

- Downie, M. N. (1971). *Métodos estadísticos aplicados*. Madrid: Ediciones del Castillo.
- Enciclopedia virtual . (2018). *Enciclopedia virtual* . Recuperado el 2018, de Introducción a la regresión lineal: <http://www.eumed.net/curso-con/medir/introd.htm>
- González de Dios, J., Moya, M., & Mateos Hernández, M. A. (1997). *Indicadores bibliométricos: Características y limitaciones en el análisis de la actividad científica*. Anales españoles de pediatría, 235-244.
- Mejía Trejo, J. (2018). *Análisis Estadístico Multivariante con SPSS para las Ciencias Económico-Administrativas*. México: Cloudbook.
- Montgomery, D. C., Peck, E. A., & Vining, G. (2006). *Introducción al análisis de regresión lineal*. México: Continental.
- Prado-Lorenzo, J. M., Gallego-Álvarez, I., García-Sánchez, I. M., & Rodríguez-Domínguez, L. (2008). Social responsibility in Spain: Practices and motivations in firms. *Management Decision*, 46(8), 1247-1271.
- Sánchez, P. E., & Benito-Hernández, S. (2013). CSR Policies: *Effects on Labour Productivity in Spanish Micro and Small Manufacturing Companies*. Springer Science , 705-724.
- Testera Fuertes, A., & Cabeza García, L. (2012). *Análisis de los factores determinantes de la transparencia en RSC en las empresas españolas cotizadas*. Omnia science, 9(1), 225-261.

- Valenzuela Fernández, L., Jara-Bertin, M., & Villegas Pineaur, F. (2014). Prácticas de reponsabilidad social, reputación corporativa y desempeño financiero. *Revista de Administración de Empresas*, 55(3), 329-344.
- Vinuesa, P. (2016). CCG-UNAM. Recuperado el 23 de Septiembre de 2018, de *Regresión líneal Simple y Polinomial: teoría y práctica*: [http://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema9\\_regression.html#introduccion-el-concepto-de-regresion](http://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema9_regression.html#introduccion-el-concepto-de-regresion)

## CAPÍTULO 6

# *Técnicas de análisis multivariante para la validación de un modelo conceptual de transformación de organización lineal a exponencial*

*Alfredo Aguilar Ruiz*

DIRECTOR DE TESIS

*Antonio de Jesús Vizcaino*

**Palabras Clave:** Organizaciones exponenciales, organizaciones lineales, modelo de transformación, técnicas de análisis multivariante.

**Abstract:** This document is a descriptive qualitative research that highlights the importance of multivariate analysis techniques for the validation of conceptual models. It focuses mainly on the proposal of a research project for the creation of a transformation model from linear to exponential organization for medium-sized companies in Mexico, as well as the key elements involved in this transition. It also links the theory and the methodological proposal with the multivariate analysis as a means to answer the research question.

### **Introducción**

El concepto de organizaciones exponenciales (ExOs) define a organizaciones cuyo impacto o resultado es desproporcionadamente grande, al menos

diez veces superior, al compararla con sus iguales gracias al uso de nuevas técnicas organizativas que se sirven de las tecnologías aceleradoras (Ismail, Malone & Van Geest, 2014). Algunos ejemplos de empresas que operan bajo el modelo exponencial son Uber, Netflix, Tesla, Airbnb, Waze, Snapchat, BlaBlaCar, entre otras; compañías que en su mayoría tienen menos de diez años de haberse fundando y que algunas de ellas se posicionan ya dentro de las 100 marcas más valiosas del mundo de acuerdo a cifras publicadas por la firma Brandz (2018).

De acuerdo a Ruffini (2016), las organizaciones exponenciales dominarán en el futuro los mercados y las organizaciones exitosas de ayer no podrán continuar teniendo éxito si no cambian la forma en que operan, ya que están siendo expuestas desde hace años a una obsolescencia prematura. De acuerdo a lo publicado en el portal de ExO Works (2018), el modelo exponencial está comenzando a ser replicado por grandes empresas de clase mundial, poniendo en desventaja a la mediana empresa, quienes tardarán años en que puedan adoptar estas nuevas prácticas.

Con base en lo anterior, el propósito de esta investigación será determinar las técnicas de análisis multivariante que se emplearán para la validación de un modelo conceptual que forma parte de la investigación denominada “Modelo de transformación de organización lineal a exponencial para la mediana empresa del canal de TI en México”. A su vez, buscará detallar cómo a través de las técnicas multivariantes se pueden medir las variables que integran dicho modelo, con el fin de conocer el grado en que se relacionan unas con otras y evaluar su relevancia dentro del modelo conceptual propuesto.

## Desarrollo

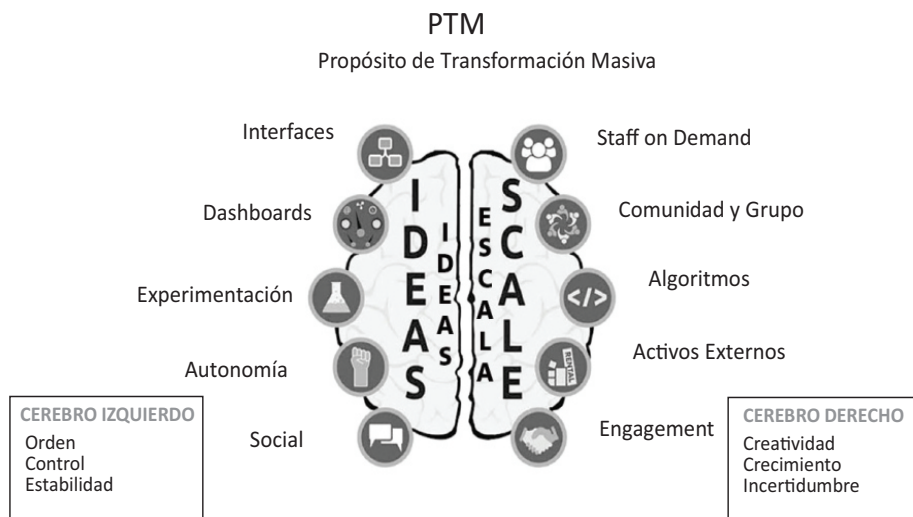
La tesis de organizaciones exponenciales surge a raíz de los resultados del análisis a las cien startups de mayor crecimiento mundial, entre 2008 y 2014, y cómo éstas tuvieron un crecimiento acelerado a menos de seis años de su aparición. La gran característica de las ExOs es que, en lugar de contar con una nómina robusta o grandes instalaciones físicas, las organizaciones exponenciales se construyen sobre tecnologías de la información que toman lo que una vez fue físico o tangible y lo desmaterializan en el mundo digital a demanda (Charania, 2015). En contraste, una empresa lineal es aquella que opera bajo un modelo tradicional, que presenta características como: el ser organizaciones jerárquica, poca to-

lencia al riesgo, gran número de empleados, planificación estratégica basada en el pasado, procesos inflexibles, entre otras (Ismail *et al.*, 2014).

De acuerdo a Charania (2015), más que una explicación, las organizaciones exponenciales son un *mindset*, es decir, una elección que las compañías toman, no sólo para volverse más competitivas sino para sobrevivir a largo plazo. Además, mientras que las organizaciones tradicionales tienden a ser jerárquicas, centralizadas y cerradas, a la vez que operan en torno a un modelo de propiedad basado en la escasez (de personas, recursos, activos, plataformas, etc.), las organizaciones exponenciales abrazan y aprovechan la apertura, la transparencia y la abundancia; las ExO se enfocan hacia afuera y no hacia adentro, lo cual les da una ventaja sobre las demás empresas (Dresner, 2015).

Entre las características de una ExO, destaca que mantienen un núcleo muy pequeño de empleados e instalaciones físicas, lo que permite una gran flexibilidad mientras los márgenes repuntan; reclutan a sus usuarios y aprovechan las comunidades virtuales y físicas para todo, desde el diseño del producto hasta desarrollo de aplicaciones; se desplazan sobre las infraestructuras existentes y emergentes en lugar de tratar de poseer las suyas propias; y crecen a un ritmo increíblemente rápido debido en gran medida a que no están enfocadas a tomar propiedad de su mercado (Ismail *et al.*, 2014).

**Figura 1. Modelo de una Organización Exponencial**



Fuente: Ismail, Malone y Van Geest (2014).



Con relación a las dimensiones internas y externas que se conjugan para alcanzar el crecimiento exponencial, se emplea el acrónimo SCALE para reflejar los cinco atributos externos, y el acrónimo IDEAS para los cinco atributos internos. Cabe destacar que no todas las ExOs presentan los diez atributos, pero cuantos más reúnan, más escalable suelen ser. De acuerdo a los autores (Ismail *et al.*, 2014), el tener un mínimo de cuatro atributos implementados convierte a la organización en una ExO contribuyendo a la aceleración de la misma.

De acuerdo a lo antes expuesto, es importante destacar que ya existen los principios y criterios a considerar para crear una ExO a partir de cero, pero aún no existe un modelo publicado a la fecha que contemple todos los elementos hacia el interior de una organización que intervienen en el proceso de transformación de organización lineal a exponencial; específicamente no existe un modelo con una orientación para la mediana empresa sobre cómo realizar la conversión de lineal a exponencial, ni mucho menos un ejemplo documentado de esta transición en alguna organización a nivel México, por lo que el objetivo de la investigación buscará responder a esta pregunta: ¿Qué elementos requiere un modelo de transformación de organización lineal a exponencial orientado a la mediana empresa en México?

### ***Transformación de Organización Lineal a Exponencial***

De acuerdo a sus creadores, el modelo ExO no es exclusivo de emprendedores ni de startups, y afirman que es posible partir de una empresa ya establecida (mediana o grande) y fortalecerla hasta alcanzar un crecimiento exponencial. A diferencia de una startup en donde se puede diseñar y construir el modelo de negocio y las operaciones internas desde cero, en las empresas ya establecidas la solución tiene que ser personalizada: se debe partir de lo que ya existe y comenzar a construirse a partir de ahí; en pocas palabras, no existe un modelo universal para convertirse en una organización exponencial (Ismail *et al.*, 2014).

De entre los ejemplos de empresas establecidas que migraron de un modelo lineal a uno exponencial, destacan el caso de: TED, GitHub, Coyote Logistics, Studio Roosegaarde y GoPro; todas empresas de tamaño mediano que para potenciar su competitividad y lograr un crecimiento acelerado integraron a su modelo de negocio estrategias y principios ExO

(propósito de transformación masiva, staff on demand, comunidad, uso de algoritmos, activos externos, compromiso, interfaces, uso de dashboards, experimentación e innovación, autonomía de equipos de trabajo y uso de tecnologías sociales). Ver Tabla 1

**Tabla 1. Empresas medianas que se convirtieron a organizaciones exponenciales**

Empresa	Fundación	País	Giro
TED	1984	Estados unidos	Medios
GoPro	2002	Estados Unidos	Fotografía
Coyote logistics	2006	Estados Unidos	Logística
Stdio roosegaarde	2007	Holanda	Arte y diseño
GitHub	2008	Estados Unidos	Software

Fuente: Ismail *et al.* (2014).

Con base en lo anterior y a través de una revisión exhaustiva a la literatura publicada sobre el fenómeno de las ExOs, se han identificado diez criterios a partir de los cuales se desarrollará la propuesta de modelo de transformación de organización lineal a exponencial. Estos conceptos clave se mencionan a lo largo de la publicación de Ismail *et al.* (2014), dispersos entre los diversos ejemplos que ofrecen sobre empresas que migraron hacia un modelo exponencial e indudablemente son los factores internos a considerar para realizar una transición de organización lineal a exponencial exitosa. Los atributos a considerar son:

1. *Liderazgo*. Se refiere a la capacidad de influir en un grupo para que se logren las metas (Robbins & Judge, 1999); es un proceso altamente interactivo y compartido, en el cual los miembros de todos los equipos desarrollan habilidades en un mismo proceso; implica establecer una dirección, visión y estrategias para llegar a una meta, alineando a las personas y al mismo tiempo motivándolas (French & Bell, 1996). Un liderazgo efectivo garantizará cumplimiento y consistencia en el proyecto; de acuerdo a Ismail *et al.* (2014), el primer requisito para transformar una empresa lineal a ExO es contar con un líder visionario que cuente con el respaldo total de la junta directiva y la alta gerencia.
2. *Adaptación al Cambio*. Se refiere al cambio para la actualización y la modernización; mientras que la estabilidad preserva la identidad de la organización; una organización sobrevive y crece en la medida

- en que pueda combinar la estabilidad con la adaptación y el cambio (Chiavenato, 2009). De acuerdo a Ismail *et al.* (2014), conforme la empresa escala y su actividad se transforma, lo mismo deben hacer sus gestores. Asimismo, el segundo elemento esencial para transformar una empresa existente en una ExO es la adaptación rápida al cambio.
3. *Motivación*. Se define como el impulso para actuar para lograr satisfacciones, los cuales responden a los tres ámbitos de las necesidades humanas: materiales, cognoscitivas y afectivas (Ferreiro & Alcázar, 2001). En relación a las ExO, es la base para empoderar a las personas, y que sientan que están realizando algo que les importa; la motivación ayuda a disminuir costos, mejora la efectividad y aceleran el aprendizaje del individuo.
  4. *Compromiso*. Se define como la fuerza relativa a la identificación individual e implicación con una organización en particular (Mowday, Steers & Porter, 1979). De acuerdo a Ismail *et al.* (2014), la verdadera comunidad surge cuando se produce un compromiso entre compañeros de igual a igual; debidamente implementado, el compromiso crea efectos en red y bucles de feedback positivo de alcance extraordinario.
  5. *Cultura Organizacional*. Se refiere a un sistema de significado compartido por los todos los miembros, el cual distingue a una organización de las demás (Robbins & Judge, 2009), y está formada por las normas informales y no escritas que orientan el comportamiento cotidiano de los miembros de una organización quienes dirigen sus acciones a la realización de los objetivos de ésta (Chiavenato, 2009). En una organización itinerante y que escala rápido, la cultura es el pegamento que mantiene al equipo unido a través de los saltos cuánticos de crecimiento de una ExO (Ismail *et al.*, 2014).
  6. *Perfil Innovador*. Se refiere a aquellos que de manera cuidadosa, intencional y constante, encuentran pequeños detalles de comportamiento en las actividades de los clientes, proveedores y otras empresas, con el fin de obtener percepciones sobre nuevas formas de hacer las cosas (Dyer, Gregersen & Christensen, 2009). De acuerdo a Ismail *et al.* (2014), el equipo responsable de la transición debe contar con la habilidad de generar ideas (asociar, preguntar, observar, establecer redes y experimentar), así como habilidades ejecutarlas (analizar, planear, implementar, seguir adelante y orientarse a los detalles).
  7. *Capital Intelectual*. Suma de todos los conocimientos que poseen los empleados y que otorgan a la empresa ventaja competitiva (Stewart, 1998). De acuerdo a los autores, el capital intelectual es una de las

- ventajas clave para adaptarse estructuralmente a la nueva era ExO.
8. *Diversidad*. Consiste en pasar de la igualdad de trato y evitar cualquier tipo de discriminación, ya sea racial, nacional, de género, de clase o de discapacidad, a promover la igualdad y, en última instancia, a cultivar la diversidad y la diferencia (Golembiewski, 1995); la diversidad supone una nueva cultura de trabajo que dinamiza la innovación y la creatividad hacia el logro de la excelencia y la calidad total (Barberá *et al.*, 2004). Asimismo, se ha demostrado repetidamente que la diversidad en términos de género, edad y experiencia produce los mejores resultados; la heterogeneidad en el equipo de transición es fundamental para el resultado exitoso del proyecto (Ismail *et al.*, 2014).
  9. *Autonomía*. Es el grado en el cual el puesto proporciona libertad, independencia y discrecionalidad sustanciales para que el individuo programe el trabajo y determine los procedimientos que deberán ser utilizados para llevarlo a cabo (Amorós, 2007); también se relaciona con la responsabilidad, independencia y poder de decisión de que goza el sujeto en su organización (Silva, 1996). Para la transición efectiva, se requiere apoyarse en equipos pequeños, independientes, multidisciplinarios con el objetivo de construir nuevos negocios desde la fase de la idea hasta la comercialización; este estilo organizativo también crea una cultura sociable, abierta y de confianza que resulta en una plantilla muy satisfecha Ismail *et al.* (2014).
  10. *Disrupción*. Describe un proceso por el cual una empresa más pequeña con menos recursos puede desafiar con éxito a empresas establecidas; las innovaciones disruptivas se originan en los puntos bajos del mercado o en nuevos mercados, es decir, los disruptores crean un mercado donde no existía uno (Christensen, Raynor & McDonald, 2015). De acuerdo a Ismail *et al.* (2014), es mucho más difícil para las grandes organizaciones el aprovechar las tecnologías disruptivas, puesto que las estructuras organizativas de las compañías establecidas existen para suprimir influencias disruptivas.

## Discusión

Se propone realizar esta investigación bajo un enfoque cuantitativo a través del cual se medirá la relación entre las variables independientes—liderazgo, adaptación al cambio, motivación, compromiso, cultura orga-

nizacional, perfil innovador, capital intelectual, autonomía, diversidad y disrupción— y la dependiente —transformación de organización lineal a exponencial— que conforman el modelo conceptual propuesto en la Figura 2, a través de la evaluación a una organización de tamaño mediano en México que haya realizado la transición de organización lineal a exponencial.

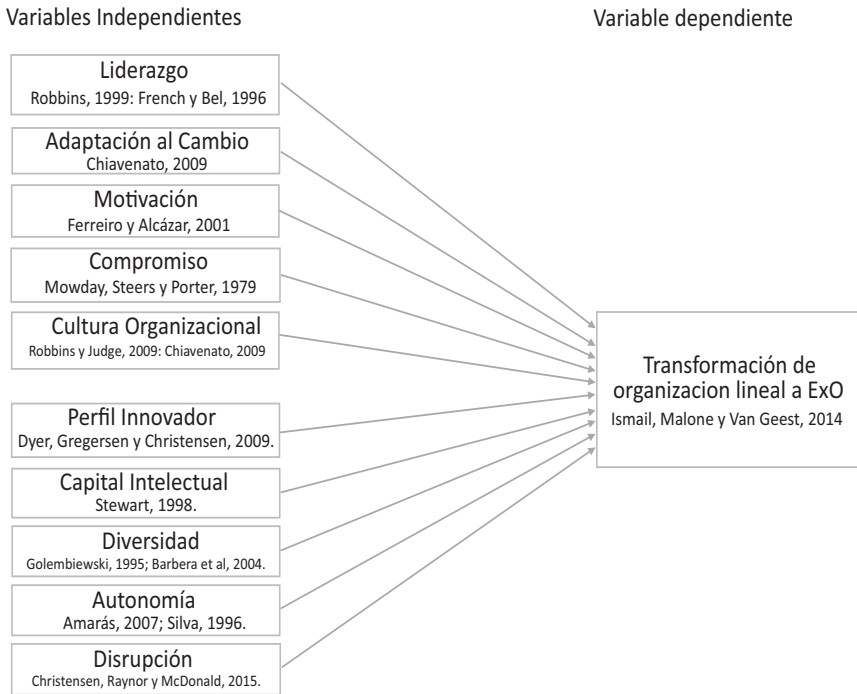
## **Metodología**

El instrumento por utilizar será un cuestionario diseñado en una escala Likert de cinco puntos y las dimensiones a evaluar serán: el liderazgo, la adaptación al cambio, la motivación, el compromiso, la cultura organizacional, el perfil innovador, el capital intelectual, la autonomía, la diversidad y la disrupción. Los sujetos participantes en la evaluación serán todo el equipo (hombres y mujeres) de los distintos departamentos de la empresa que hayan participado de manera directa en la transición del modelo lineal a exponencial.

Para el tratamiento de datos perdidos o faltantes, se aplicará un proceso de imputación múltiple, el cual consiste en reemplazar los datos perdidos por estimaciones; en relación a los datos atípicos, que pueden definirse como aquel punto que se encuentra lejos del centro de los datos (Peña, 2002), se emplearán la prueba de Grubbs, el criterio de Pierce y la prueba Q de Dixon. Para evitar obtener datos atípicos se realizará un análisis preliminar para detectar valores extremos mediante los residuos del modelo; en caso de detectarse, se evaluará si procede de un error humano o del instrumento de medida para ser corregido (Abellana & Farran, 2015).

**Figura 2.**

**Modelo Ex Ante con variables a someter al análisis multivariante**



Fuente: Elaboración propia

Es importante el considerar el error de medida en el análisis, que se refiere al grado en que los valores observados no son representativos de los valores verdaderos. Para reducir el error de medida en las variables analizadas y obtener mejor precisión en las medias, se desarrollarán mediciones multivariantes, donde diversas variables se unirán en una medida compuesta para representar un concepto y obtener así una perspectiva más completa (Hair *et al.*, 1999).

### **Análisis Multivariante para Validar el Modelo**

El análisis multivariante se refiere a todos los métodos estadísticos que analizan simultáneamente medidas múltiples de cada individuo u objeto sometido a investigación. Cabe destacar que para ser considerado verdaderamente multivariante todas las variables deben ser aleatorias y estar

interrelacionadas de tal forma que sus diferentes efectos no puedan ser interpretados separadamente con algún sentido. Algunos autores afirman que el propósito del análisis multivariante es medir, explicar y predecir el grado de relación de los valores teóricos (Hair, Anderson, Tatham & Black, 1999).

Para responder a la pregunta de investigación y conocer de qué forma se relaciona el liderazgo, la adaptación, la motivación, el compromiso, la cultura organizacional, el perfil innovador, el capital intelectual, la autonomía, la diversidad y la disrupción con la transformación de organización lineal a exponencial, se emplearán las siguientes técnicas multivariantes:

1. *Componentes principales y análisis factorial común.* Es una aproximación estadística que puede usarse para analizar interrelaciones entre un gran número de variables y explicar estas variables en términos de sus dimensiones subyacentes comunes (Hair *et al.*, 1999).

El análisis factorial es una técnica de interdependencia en la que se consideran todas las variables simultáneamente, cada una relacionada con todas las demás y empleando todavía el concepto del valor teórico, el compuesto lineal de las variables (Hair *et al.*, 1999); su objetivo es reducir los datos obtenidos en las variables originales hacia un conjunto más pequeño de factores con mínima pérdida de información (Mejía, 2018).

La respuesta que se busca a través de este procedimiento estadístico, es identificar qué grupos de variables están estrechamente relacionadas entre sí, postulando que esa estrecha asociación entre ellas responde a la existencia de un factor (dimensión o variable latente) que no es observable directamente (Alaminos, Francés, Penalva & Santacreu, 2015).

2. *Regresión múltiple.* Es una técnica estadística que puede utilizarse para analizar la relación entre una única variable dependiente (criterio) y varias variables independientes (predictores). El objetivo del análisis de regresión múltiple es usar las variables independientes cuyos valores son conocidos para predecir la única variable criterio; cada variable predictor es ponderada, de forma tal que las ponderaciones indicarán su contribución relativa a la predicción conjunta (Hair *et al.*, 1999). Para Pérez y Medrano (2010), esta técnica evalúa cómo 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.

Al conjunto de variables independientes ponderadas se le conoce también como valor teórico de la regresión, una combinación lineal de las variables independientes que predice mejor la variable criterio. La ecuación de regresión, también denominada como el valor teórico de la regresión, es el ejemplo de valor teórico más ampliamente reconocido entre todas las técnicas multivariantes (Hair *et al.*, 1999).

3. *Análisis de la varianza (ANOVA)*. Es una técnica estadística de dependencia que puede ser usada simultáneamente para explorar las relaciones entre diversas categorías de variables independientes (usualmente denominadas como tratamientos) y una variable métrica dependiente (Hair *et al.*, 1999). El ANOVA se denomina como proceso univariante debido a que se emplea para valorar las diferencias entre grupos utilizando una única variable dependiente métrica; es particularmente útil con diseños de experimentos en donde el investigador controla o manipula directamente una o más variables independientes para determinar el efecto sobre una variable dependiente ya que proporciona las herramientas necesarias para juzgar la fiabilidad de cualquier efecto observado (Hair *et al.*, 1999).

## Conclusiones

El objetivo central de esta investigación fue establecer las técnicas de análisis multivariante que se emplearán para la validación del modelo de transformación de organización lineal a exponencial para la mediana empresa en México, con el fin de medir los elementos clave que intervienen en dicha transición y poder determinar la efectividad de la conversión. A su vez, vincular la teoría y la propuesta metodológica con el análisis multivariante como medio para responder a la pregunta de investigación, mediante lo cual es posible concluir que:

Con el análisis multivariante será posible medir y conocer la relación entre los elementos que intervienen en el proceso de transformación de organización lineal a exponencial. Para identificar qué grupos de variables están estrechamente relacionadas entre sí y reducir los datos obtenidos en las variables originales, se aplicará la técnica de análisis factorial a los resultados del cuestionario; para evaluar los determinantes de la efectividad del modelo y con el objetivo de predecir la variable dependiente, se realizará el análisis de regresión múltiple; además, se realizará el análisis de la varianza (ANOVA) con el fin de juzgar la fiabilidad y determinar



la probabilidad de que las diferencias en las medias entre los grupos sean debidas meramente al error muestral.

## Referencias

- Abellana, R. & Farran, A. (2015). Identificación, impacto y tratamiento de datos perdidos y atípicos en epidemiología nutricional. *Revista Española de Nutrición Comunitaria*, 21(1), pp. 188-194. Recuperado de: [http://www.renc.es/imagenes/auxiliar/files/NUTR.%20COMUN.%20SUPL.%201-2015\\_Tratamiento%20atipicos.pdf](http://www.renc.es/imagenes/auxiliar/files/NUTR.%20COMUN.%20SUPL.%201-2015_Tratamiento%20atipicos.pdf)
- Alaminos, A., Francés, F., Penalva, C. & Santacreu, O. (2015). *Análisis multivariante para las ciencias sociales I*. Índices de distancia, conglomerados y análisis factorial. Cuenca, Ecuador: Pydlos Ediciones.
- Amorós, E. (2007). *Comportamiento Organizacional: En Busca del Desarrollo de Ventajas Competitivas*. Recuperado de: <http://www.eumed.net/libros-gratis/2007a/231/indice.htm>
- Barberá, E. (2004). La diversidad de género como estrategia favorecedora de la igualdad de oportunidades en los entornos laborales. *Revista de Economía Pública, Social y Cooperativa*, nº 50, noviembre 2004, pp. 37-53. Recuperado de: <http://www.redalyc.org/pdf/174/17405003.pdf>
- BrandZ (2018). *BrandZ Top 100 Most Valuable Global Brands 2018*. Recuperado de: [http://brandz.com/admin/uploads/files/BZ\\_Global\\_2018\\_DL.pdf](http://brandz.com/admin/uploads/files/BZ_Global_2018_DL.pdf)
- Charania, N. (2015). Exponential Organizations are the future of global business and innovation. Recuperado de: <https://techcrunch.com/2015/07/05/exponential-organizations-are-the-future-of-global-business-and-innovation/>
- Chiavenato, I. (2009) *Comportamiento Organizacional. La dinámica del éxito en las organizaciones*. México: Mc Graw Hill.
- Christensen, C., Raynor, M. & McDonald, R. (2015). What is disruptive Innovation? *Harvard Business Review*, December 2015 Issue, pp. 44-53. Recuperado de: <https://hbr.org/2015/12/what-is-disruptive-innovation>
- Dresner, M. (2015). *Why an Exponential Organization May Bury Your Company?* Recuperado de: <https://knect365.com/innovation/article/2aa6cbac-725c-46b89686-ce7863fd4a2f/why-an-exponential-organization-may-bury-your-Company>
- Dyer, J., Gregersen, H. & Christensen, C. (2009). The innovator's DNA. *Harvard Business Review*, December 2009 Issue. Recuperado de: ht-

- tps://hbr.org/2009/12/the-innovatorsdna
- ExoWorks (2018). *An Exponential Journey*. Recuperado de: <https://www.exo.works/become-an-exponential-organization>
- Ferreiro, P. & Alcázar, M. (2002). *Gobierno de Personas en la Empresa*. Barcelona, España: Editorial Ariel.
- French, W. & Bell, C. (1995). *Desarrollo Organizacional*. México: Prentice Hall.
- Golembiewski, R. (1995). *Managing diversity in organizations*. Alabama, Estados Unidos: The University of Alabama Press.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1999). *Análisis multivariante*. Madrid, España: Prentice Hall.
- Ismail, S., Malone, M. S., & Van Geest, Y. (2014). *Exponential organizations. Why New Organizations Are Ten Times Better, Faster, and Cheaper Than Yours (and What to Do About It)*. Nueva York, Estados Unidos: Diversion Books
- Mejía, J. (2018). *Análisis Estadístico Multivariante con SPSS para las Ciencias. Económico Administrativas*. Recuperado de: <https://www.bibliotecaebook.com/product/analisis-estadistico-multivariante-con-spss-para-las-ciencias-economicoadministrativas>
- Mowday, R.T., Steers, R.M. & Porter, L.W. (1979). The measurement of Organizational Commitment. *Journal of Vocational Behavior*, 14, pp. 224-247. Recuperado de: <http://www.dtic.mil/dtic/tr/fulltext/u2/a057377.pdf>
- Peña, D. (2002). *Análisis de datos multivariantes*. Madrid, España: McGraw-Hill.
- Pérez, E. R., & Medrano, L. (2010). Análisis factorial exploratorio: Bases conceptuales y metodológicas. *Revista Argentina de Ciencias Del Comportamiento*, 2(1889), pp. 58–66. Recuperado de: [http://www.academia.edu/12607324/Revista\\_Argentina\\_de\\_Ciencias\\_del\\_Comportamiento\\_RACC\\_Análisis\\_Factorial\\_Exploratorio\\_Bases\\_Conceptuales\\_y\\_Metodológicas\\_Artículo\\_de\\_Revisión](http://www.academia.edu/12607324/Revista_Argentina_de_Ciencias_del_Comportamiento_RACC_Análisis_Factorial_Exploratorio_Bases_Conceptuales_y_Metodológicas_Artículo_de_Revisión)
- Robbins, S. & Judge, T. (2009). *Comportamiento Organizacional*. México: Pearson Educación.
- Ruffini, F. (2016). *Exponential Organizations* (tesis de grado). Universidad Guido Carli, Roma, Italia.
- Silva, M. (1996). *El clima en las organizaciones: teoría, método e intervención*. Barcelona, España: Editorial P.P.U.
- Stewart, T. (1998): *La nueva riqueza de las organizaciones: el capital intelectual*. Buenos Aires, Argentina: Granica.



## CAPÍTULO 7

# *Correlación de variables de la competitividad a partir de la aplicación de análisis multivariables de técnicas dependientes (Regresión Lineal Múltiple)*

*Jovanni Trinidad Saldaña*

DIRECTOR DE TESIS

*Katia Magdalena Lozano Uvario*

### **Resumen**

Para el abordaje de del presente capítulo se pretende desde una primera instancia tener un acercamiento sobre el análisis multivariable de técnicas dependientes enfocado en la regresión lineal, con la finalidad de realizar una correlación con la variable de competitividad y mostrar la importancia que se tiene la aplicación de este método matemático en el estudio de las ciencias sociales. Por último, se realizó un análisis bibliométrico con el objeto de establecer directrices con referencia a los autores que generan conocimiento con estas variables.

**Palabras clave:** Regresión lineal múltiple, competitividad empresarial, PYMES.

### **Abstract**

In this chapter is intended from a first instance to have an approach on the multivariable analysis of dependent techniques focused on linear re-

gression, in order to make a correlation with the competitiveness variable and show the importance of the application of this mathematical method in the study of social sciences. Finally, a bibliometric analysis was carried out in order to establish guidelines with reference to authors who generate knowledge with these variables.

**Keywords:** Multiple linear regression, business competitiveness, SMEs.

## **Introducción**

La presente investigación pretende llevar a cabo el análisis de técnicas multivariantes de dependencia utilizadas en la competitividad empresarial de las PYMES en México con la finalidad de determinar cuáles son aquellos factores que tienen correlación o que afectan directamente en las acciones competitivas. Se pretende investigar la técnica de regresión lineal múltiple debido a que la misma tiene como objeto el analizar la interrelación entre variables y así establecer una dependencia de las mismas.

Dentro de las ciencias sociales cada día es más común la necesidad de realizar análisis numérico de la información utilizando distintos cálculos estadísticos. En la actualidad, hay métodos aplicables dentro de estas ciencias que tienen como objetivo otorgar posibilidades que den el tratamiento cuantitativo que se requiere para su análisis, además, cabe mencionar que esto no sería posible realizarlo con los procedimientos tradicionales estadísticos univariantes o bivariantes. Estos métodos, se consideran a aquellos que integrados por una serie de técnicas de análisis de datos son capaces de formar parte de la rama de la Estadística conocida como análisis multivariante (Closas, Arriola, Kuc, Amarilla, & Jovanovich, 2013), por tal motivo, son considerados de gran utilidad en la realización de estudios de dependencia entre variables.

Por otra parte, se menciona que el análisis multivariante tiene sus inicios a partir de la utilización de la regresión lineal por parte de Gauss en 1809 y, subsecuentemente, algunos otros estadísticos como Markov en 1900; considerando que las técnicas más recientes tuvieron origen en los años treinta.

### ***Análisis multivariante de técnicas dependientes en la competitividad***

El análisis de la competitividad empresarial de acuerdo a lo que menciona Bonales, Zamora y Ortiz (2015) considera que la misma, está confor-

mada por tres elementos principales: factores internos, factores externos e índices. La participación en su conjunto de estos elementos definen a la competitiva de las empresas en una posición, además, se considera que poseen las características necesarias con las cuales se puede generar un análisis para mejorar el desempeño de las organizaciones.

Los análisis multivariantes de técnicas dependientes pueden ayudar a generar un modelo conceptual que lleve a visualizar las interrelaciones de competitividad con la finalidad de obtener una perspectiva diferente en el análisis de las variables que permitan dar forma a la práctica de la competitividad de las empresas, que permita tener un mejor entendimiento con respecto a las relaciones entre variables (Bonales Valencia, Zamora Torres, & Ortíz Paniagua, 2015).

De acuerdo por lo mencionado por Vázquez y Bernard (1988) existen dentro de los modelos estadísticos aplicables en el análisis multivariante se encuentran los siguientes respecto a sus fines particulares que persiguen:

#### Modelos descriptivos

- Análisis de los Componentes Principales.
- Análisis Factorial Clásico.
- Análisis Factorial de las Correspondencias.
- Clasificación Automática (Análisis de Clúster)

#### Modelos decisionales

- Análisis Discriminante.
- Prueba de Hipótesis sobre Vectores de Medias.
- Regresión Lineal Múltiple.

La naturaleza y distribución de las variables incluidas en el estudio, dentro de los aspectos multivariantes se deben de considerar su naturaleza y la distribución de variables que presenta con la finalidad de llegar a tener una observación, además, considerar realizar el examen de los datos ausentes y casos atípicos, por otra parte es necesario pensar en las posibles soluciones para cada caso, así como la verificación de los supuestos de normalidad, linealidad y homocedasticidad tomando en cuenta el estudio y análisis de las posibles alternativas para intentar solucionar los problemas encontrados de forma general y los aspectos para generar un análisis multivariante (Álvarez Suárez, Caballero, & Pérez Lechuga, 2006).

Los métodos multivariantes consideran algunos factores a tomar en cuenta al momento de realizar sus análisis como: la dependencia o no

entre las variables, las escalas de medición utilizadas y el objetivo que se persigue (Lebart, Morineau, & Fénelon, 1981); (Dagnelie, 1981); (Hair, Anderson, Tatham, & Black, 1999).

Análisis multivariante tiene la finalidad de permitir la selección de un modelo para lograr resolver los problemas que se les presenten a los investigadores hasta el momento de llegar al diagnóstico crítico de los resultados. Hair, Anderson, Tatham, & Black, (1999) utilizan seis pasos para llevar a cabo el análisis multivariante, considerando que los primeros tres se representan el análisis que antepone los datos, el cuarto al análisis y selección de del modelo a desarrollar y los dos últimos van dirigidas a la interpretación de los resultados obtenidos, con respecto a lo anterior se muestra lo siguiente:

1. Definición del problema de investigación, objetivos y técnica multivariante conveniente.
2. Desarrollo del plan de análisis (tamaños de muestra mínimos, tipos de variables permitidas y métodos de estimación).
3. Evaluación de los supuestos básicos de la técnica propuesta.
4. Estimación del modelo multivariante y valoración del ajuste del modelo.
5. Interpretación del valor teórico.
6. Validación del modelo multivariante.

### ***Competitividad empresarial***

La competitividad se define según Martínez Fierro (1999) como “la capacidad que ésta manifiesta para producir bienes y servicios destinados a los distintos mercados donde compite, manteniendo o incrementando su cuota de participación relativa en ellos y obteniendo una renta con la que se retribuye a los propietarios de todos los recursos implicados” esto les permite mantenerse dentro de la competencia con las demás empresas, a través de la inversión de las utilidades generadas.

Michael Porter (1998) en su libro “ser competitivo” menciona que la competitividad está determinada por la productividad, la cual está definida como el valor del producto generado por una unidad de trabajo o de capital. La productividad es función de la calidad de los productos (de la que a su vez depende el precio) y de la eficiencia productiva.

Porter a su vez en su libro “la ventaja competitiva de las naciones” (Porter, 1991), menciona que la competitividad de las empresas es resultado de la competitividad de las naciones, por tal motivo se considera que

esto depende de la interacción de cuatro condiciones que permiten esta acción y a su vez, también pueden impedir el obtener una ventaja competitiva, estas se consideran de la siguiente manera:

1. Las condiciones de los factores, significa que cada país tiene en su haber, ciertas características económicas o tecnológicas, que lo hacen desarrollar más y mejores condiciones para el surgimiento de nuevas empresas y demás actividades productivas en relación a los demás países.
2. Las condiciones de la demanda, esto es que, el cliente determina los niveles de producción y los métodos de fabricación de un producto.
3. Los sectores de apoyo industrias relacionadas, es decir, cuando empresas cierto rango se interrelacionan para poder obtener beneficios mutuos.
4. Implementar estrategias estructurales para obtener una ventaja competitiva tomando en cuenta el actuar de las empresas rivales.

Desde un punto de vista sectorial, se define la competitividad como “la capacidad de un sector para aumentar, en condiciones de libre competencia, su participación en los mercados interior y exterior, a la vez que mantiene un crecimiento satisfactorio de las rentas reales generadas por su actividad” (Cohen, Teece, Tyson, & Zysman, 1984). Considerando esta definición permite resaltar como principal función la de incrementar la utilidad en base a la mejora competitiva.

A partir de 1991 Porter dentro de unos de sus importantes estudios de competitividad argumenta que las ventajas competitivas han desplazado de manera definitiva la concepción clásica de las ventajas comparativas en las cuales las capacidades en la dotación de recursos naturales de un país determinaban el nivel competitivo de los países.

La importancia de la competitividad ha impulsado a las empresas a buscar nuevas estrategias que le permitan entrar en el papel de crecimiento competitivo con la finalidad de cubrir sus necesidades financieras, comerciales, operacionales, desarrollo humano y capital social.

Con base en lo anterior, se encuentra una necesidad preponderante de las micro, pequeñas y medianas empresas de una región determinada sean apoyadas con el fin de incrementar su competitividad, a través de aumentar su productividad, apoyando la sustitución de las importaciones e incrementando sus exportaciones integrándose en cadenas productivas (Simón & Rueda, 2004). Mediante las cuales permitan a los empresarios llegar a sus objetivos y ser más competitivos frente a otras empresas.



Dentro del contexto de la competitividad cabe mencionar enfoques que se le da a la competitividad de acuerdo con diferentes autores y las variantes que estas definiciones pueden tener. La literatura tiene un enfoque de competitividad empresarial la cual se trabaja en torno a la organización, tiene un enfoque de competitividad sistémica por la forma de trabajo y las habilidades que las empresas implementan.

Por otra parte, haciendo referencia a lo que menciona Morales & Beach (2000) sobre la competitividad en donde se describe está en su surgimiento en forma paralela dentro de los ámbitos macroeconómico, es decir, la administración gubernamental y los instrumentos de política económica con el objeto de crear un medio ambiente favorable para el desempeño de las empresas en la región; y microeconómico, en el que las empresas se proponen incrementar su eficiencia, productividad, calidad, con el propósito de tener un desempeño mayor al de sus competidores.

La competitividad empresarial busca que el sector logre identificar sistemas de calidad internacionales y procesos de innovación, esto con la finalidad de que el mueble obtenga un valor agregado. La competitividad en una empresa por Morales & Beach (2000), define como “la capacidad para disponer de algún atributo que le permiten un desempeño empresarial superior que le otorga cierto tipo de ventaja sobre sus competidores, gracias a la creación de productos de valor”, a su vez, Porter (1996) la considera como una capacidad para operar rentablemente y competir exitosamente en los mercados mundiales.

Para que una empresa pueda ser competitiva se considera que estas dependen tanto de los factores internos (empleados, directivos, accionistas, dueños de empresas, entre otros) como de los factores externos (proveedores, clientes, inversionistas, gobierno, entre otros) ya que estos deben de incentivarse directamente por la empresa realizando unas buenas relaciones públicas, debido a que este es considerado como otro punto estratégico de competitividad, para que estas obtengan su competitividad empresarial en el momento en que estas estén en condiciones para competir con otras del mismo giro en similares o iguales mercados (Aragón & Rubio, 2005).

Este tipo de competitividad se considera como la capacidad que las empresas tienen para suministrar productos o prestar servicios con la calidad deseada, con las exigencias de sus clientes al mejor costo en el mercado para que esta permita sustentar (Porter, 1998). “la empresa se mantenga o acreciente su cuota en el mercado”.

La competitividad sistémica puede verse implementada por las empresas a través de la generación de un patrón en conjunto organizativo de

las empresas. Los parámetros de relevancia competitiva en todos los niveles del sistema y la interacción entre ellos es lo que genera ventajas competitivas (Esser, Hillebrand, Messner, & Meyer-Stamer, 1994), por tal motivo la competitividad sistémica es realizada cuando la organización ha logrado la implementación de los llamados “6 círculos de la competitividad” en los cuales se menciona como: la microeconomía, lo meso-económico, macroeconómico, internacional, Institucional y el político-social, lo que ha llevado a las empresas ha cualidades como el ser inteligentes, flexibles y ágiles; integradas por un capital organizacional, logístico e intelectual, competitivo y eficiente tanto a nivel nacional como internacional.

El término sistémico está dirigido a la especialización en la economía, la innovación tecnológica, la calidad de las redes de distribución y los factores de localización, lo que constituye el estado de suministro de bienes y servicios. (Hatzichronoglou, 1997), esto permite generar mejores análisis, planteamientos e implementación de los sistemas con los que se dirige y desempeña cualquier organización.

Por otra parte en el libro “Procesos de clusterización en Jalisco” se menciona que “cuando las empresas interactúan y colaboran complementando sus recursos y estrategias, estos pueden generar capacidades colectivas, eficiencia y sinergia que les permiten competir exitosamente con empresas de otras regiones y países” (Arechavala, 2014), a su vez, se considera que los conglomerados de empresas y organizaciones complementan actividades para fortalecer su competitividad colectiva, además, señala elementos básicos para posibilitar el hablar de una integración potencial y de la eventualidad de un clúster competitivo, donde considera a estos como procesos de interés en colaboración, coordinación, aprendizaje colectivo, sinergia y capital social.

### ***Regresión lineal múltiple aplicada a las ciencias sociales***

Para el análisis de la competitividad se considera dentro de este capítulo el modelo de la regresión lineal múltiple que es considerada como la técnica adecuada siempre y cuando el análisis tenga una variable dependiente de la cual varias variables independientes dependan de ella, además, de que pueden ser tanto métricas como no métricas. Dentro de las investigaciones se ha determinado que esta técnica tiene un papel relevante dado que su aplicación permite, entre otras cosas, la observación de las variables independientes y la manera en que logran predecir la variable dependiente. Dicho de otra manera, se considera que a partir del análisis

sis de regresión puede ser posible inferir acerca de la existencia o no de relaciones significativas entre las variables independientes y la variable dependiente, siempre considerando el marco científico que se estableció para la investigación (Lévin & Valeta, 2003). Para la regresión lineal múltiple se muestra el siguiente es un esquema simplificado:

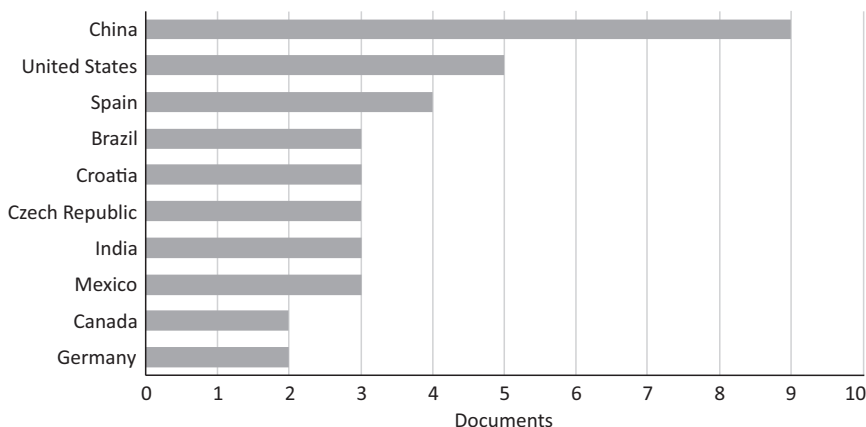
$$\underbrace{Y_1}_{\text{(métrica)}} \leftarrow \underbrace{(X_1, X_2, X_3, \dots, X_m)}_{\text{(métricas, no métricas)}}$$

### ***Análisis bibliométrico***

Con respecto al análisis bibliométrico realizado en la base de datos de Scopus sobre el tema de competitividad y la aplicación del modelo multivariante de regresión lineal múltiple es importante mencionar que se encontraron 50 artículos con la correlación entre estos dos variables de los cuales alrededor del 18% (Véase en imagen 1) son generados en el área de conocimiento del país asiático con mayor crecimiento, el país al cual nos referimos en China, por otra parte, podemos encontrar que los países de Norte América desarrollan aproximadamente un 20% del total de los estudios.

**Imagen 1**  
**Documents by coutry of territory**

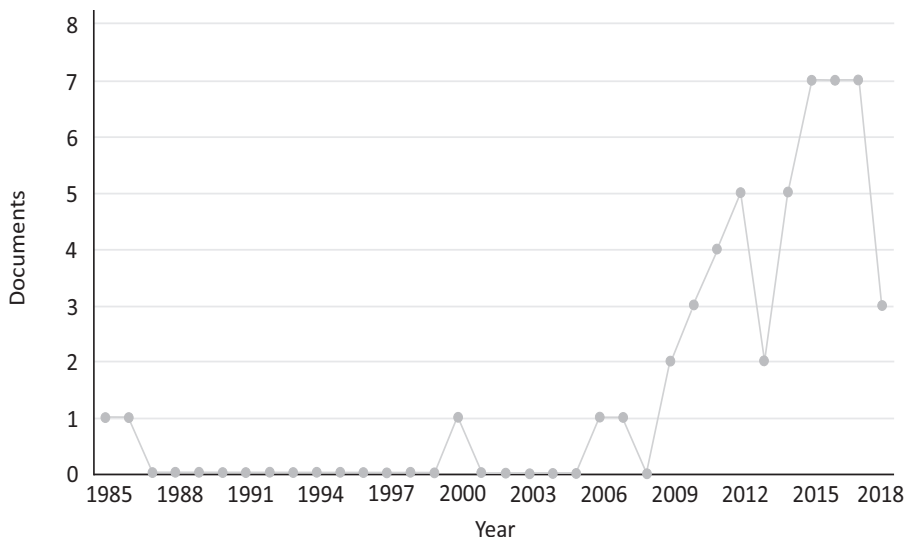
Compare the document counts for up to 15 countries/territories



Fuente: Base de datos Scopus.

A su vez, se puede argumentar que desde el 2010 a la fecha la aplicación del modelo de regresión múltiple en el estudio de la competitividad se ha aplicado con mayor frecuencia, debido a que solo en este tiempo se ha generado el 90% (Véase en imagen 2) con respecto a la primera fecha que se localiza que fue en 1985.

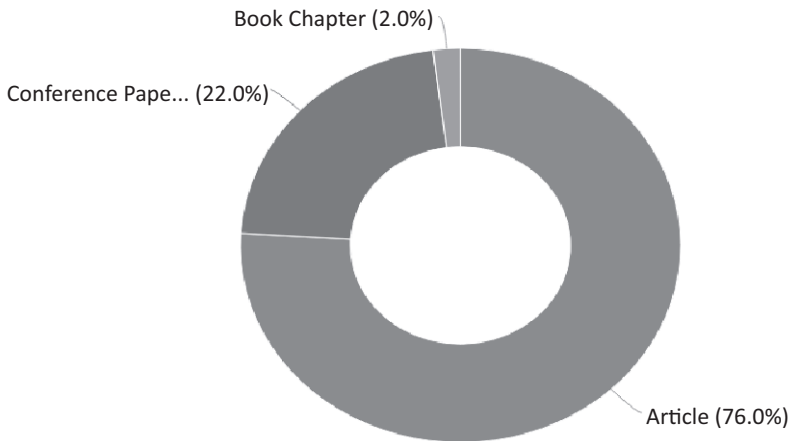
**Imagen 2**  
**Documents by year**



Fuente: Base de datos Scopus

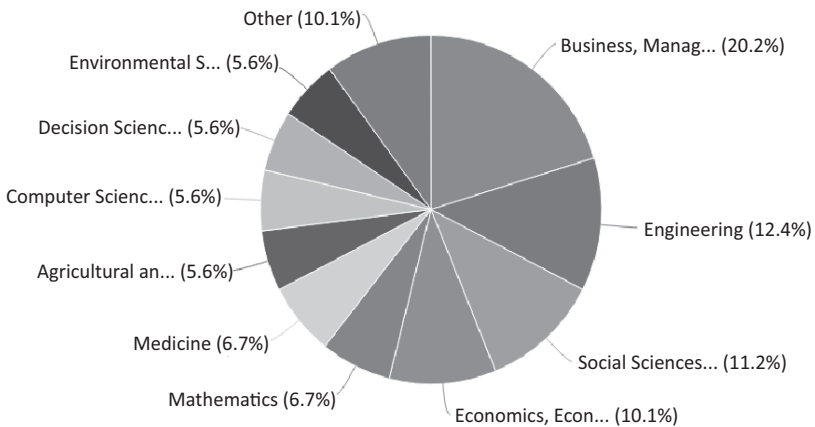
Por otra parte, la generación del conocimiento sobre el tema tratado, corresponde a que los investigadores consideran en un 76% la realización de artículos (Véase en imagen 3) para la implementación del modelo de regresión lineal en la competitividad, además, que el área del conocimiento que más considera eficiente es la de Negocios y Administración (Véase en imagen 4).

**Imagen 3**  
**Documents by type**



Fuente: Base de datos Scopus.

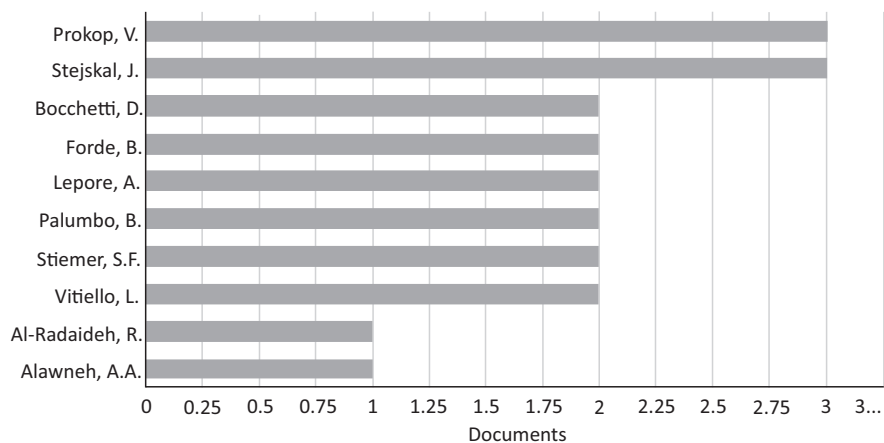
**Imagen 4**  
**Documents by subject area**



Fuente: Base de datos Scopus.

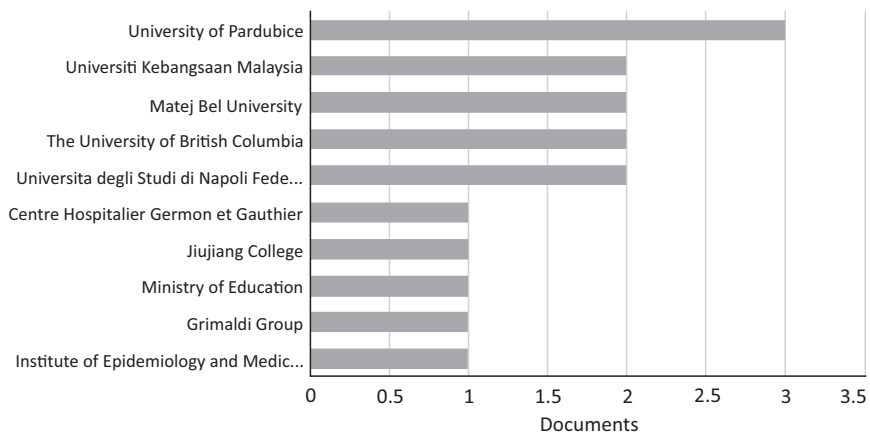
Por último, es relevante mencionar que los principales autores en la aplicación de este método multivariante son Prokop y Stejskal con 3 artículos cada uno (Véase en imagen 5) y la Universidad donde se lleva a cabo la generación de este conocimiento es la University of Pardubice con 3 producciones (Véase en imagen 6).

**Imagen 5**  
**Documents by author**



Fuente: Base de datos Scopus.

**Imagen 6**  
**Documents by subject area**



Fuente: Base de datos Scopus.

## Discusión

El análisis bibliométrico en la actualidad es considerado como una herramienta que permite medir y analizar el impacto que tiene algunos au-

tores y sus publicaciones dentro de ámbito científico (Escorcía Otalora, 2008), por tal motivo, su uso e implementación lleva a cabo un papel relevante dentro de la difusión y divulgación del nuevo conocimiento con la finalidad de obtener explicaciones concretas y solidas de cada uno de los temas científicos tratados.

Partiendo de lo anterior se llevó a cabo una investigación dentro del tema para generar una análisis que permita revisar la literatura de distintos autores y como cada uno de ellos abordan el análisis multivariante dentro de sus investigaciones, además, de visualizar en el cuadro 1 los factores y componentes que cada autor llega a considerar para el estudio de la competitividad, considerando a estos autores como los principales involucrados en estas áreas de acuerdo a la base de datos Scopus.

Oxelheim argumenta que en las regresiones utilizadas en su estudio se tomaron encienta variables dependiente las cuales se analizaron a partir del desarrollo de la información, midiendo primeramente su calidad, por otra parte, considera dentro de su marco de razonamiento las siguientes categorías: variables macroeconómicas, impacto, estrategias y total (Oxelheim, 2003).

Por otra parte, Aryal, Mann, Loveridge, y Joshi (2018) consideran que los modelos de regresión lineal no son apropiados en su estudio, dado que ellos, toman como variable dependiente la innovación de la empresa y argumentan que al tomar esta variable de esa manera, es necesario partir del hecho de que para su análisis es necesario partir del conteo de patentes de modelos, debido a que es una variable de recuento que toma solo valores enteros no negativos.

A su vez, mencionan que el utilizar la regresión lineal con respecto a la homocedasticidad y la distribución normal de los residuos llevara a los supuestos a ser atípica y obtener estimaciones de coeficientes sesgados e inconsistentes (Greene, 2003), sin embargo, creen que para su investigación, tiene mayor factibilidad el utilizar modelos de conteo como Poisson y binomial negativo, ya que estos pueden llegar a ser más apropiados para analizar datos de conteo, como el número de solicitudes de patente (Allison & Waterman, 2002); (Greene, 2003); (Hall, Griliches, & Hausman, 1986).

Yil, Han y Cha (2018) generaron para el análisis de sus variables una prueba de Cronbach, con el propósito de obtener la consistencia interna con base a la medición de múltiples ítems, por otra parte, se realizó un análisis factorial confirmatorio para examinar la confiabilidad y la validez de las medidas utilizadas en su estudio. Para finalizar se llevó a cabo la prueba de hipótesis, en la cual se logró analizar la relación entre

las variables, esto fue junto con un análisis de correlación con la que pudo verificar la validez discriminante y encontrar que las correlaciones entre las variables deducidas de la hipótesis eran consistentes con la hipótesis general y el valor de la fracción era mayor que el cuadrado del valor de correlación en todas las variables (Fornell & Larcker, 1981).

Sadaf, Oláh, Popp y Máté (2018) para el análisis de su estudio considera un modelo de regresión lineal con errores estándar consistentes de heterocedasticidad, a través del cual toma en cuenta que dentro de las organizaciones están numerosos problemas económicos y por tal motivo, pone en riesgos su éxito a través de diferentes maneras y en diversos grados, por lo cual, enfoca su análisis a partir de la utilización de variables como fraude, abuso, entre otros, con la finalidad de encontrar la correlación entre estas acciones y encontrar la manera de contrarrestarlo a través de la educación.

Nyga-Łukaszewska y Chilimoniuk-Przeździecka (2017) utilizaron un modelo de regresión de efectos fijos con análisis de datos de panel a través de la cual se estimó la influencia de cada variable exógena en la variable endógena dentro de la función de regresión, además, dentro de los cálculos se esperaba detectar el efecto fijo en el modelo tomando en cuenta las características individuales que pueden influir o no en las variables predictorias. Continuando con lo anterior, implementaron una regresión artificial con el objeto de generar una ecuación de efectos aleatorios que pueda estimar a partir del aumento de variables adicionales para obtener desviaciones de la media.

Por último, el documento hecho por Yurievna y Aleksandrovich (2016) fue a partir de la utilización adecuada de un aparato matemático que permitió implementar el pronóstico de puntos e intervalos de los indicadores acotados, en particular, el índice de competitividad regional. Este se llevó a cabo para a través de la utilización de la regresión logística, en la cual los parámetros se determinaron mediante métodos convencionales que utilizan la transformación logarítmica inversa de la variable dependiente. A su vez, se hizo la distribución de errores para los modelos de regresión en su clase y un algoritmo con la finalidad de poder calcular adecuadamente los intervalos de confianza que llevaran a los pronósticos puntuales. A partir de lo anterior, se consideró que los indicadores acotados generan serie de ventajas, entre las que se encuentran la posibilidad de traducirlos a los índices de cadena y de referencia como resultado de una falta de negatividad estricta y la informatividad.

El análisis de la importancia de los métodos multivariantes lleva a argumentar sobre los factores considerados en cada uno de los estudios



con el objeto de entender cuáles fueron aquellos componentes que se suspusieron a partir de cada uno de los métodos, para así, tener el conocimiento que llevo a reflexionar a los anteriores autores sobre cuál es el proceso adecuado a utilizar y como se muestra en el cuadro 1, que se consideró para tomar la decisión sobre el método multivariable a elegir.

**Cuadro 1**

<b>Autores</b>	<b>Aportaciones</b>
Oxelheim, L. (2018).	Argumenta que la asimetría de información potencial puede tener muchas explicaciones, tales como evitar un suministro de información que pueda llegar a competir los indicadores, lo que lleva a un impacto negativo en la propia competitividad de la empresa.
Aryal, G., Mann, J., Loveridge, S. & Joshi, S. (2018).	Consideran que hay varios factores que sugieren que las empresas urbanas son más competitivas que las rurales, por ejemplo, debido a su proximidad con otras empresas innovadoras o en función del grado o intensidad de acceso a mercados más amplios (como las exportaciones y el comercio electrónico).
Yil, H. T.; Han, C. N.; Cha, Y. B. (2018).	Dicen que es necesario tener los procesos de coordinación, integración y redistribución, en la empresa, ya que sin ellos, no puede mantener una ventaja competitiva en entornos externos complejos y altamente volátiles. Por lo tanto, considera que no se puede inferir dentro de esas capacidades internas específicas de la empresa, como marketing, I + D y capacidad de operación, afectan la capacidad dinámica.
Sadaf, Oláh, Popp y Máté (2018)	Mencionan que la competitividad comienza a partir de la influencia positiva de un grupo de instituciones, políticas y factores, que llegan a determinarse considerando la deuda, la gestión de riesgo, la cultura organizativa y a confianza.

<p>Nyga-Łukaszewska, H. &amp; Chilimoniuk-Przeździecka, E. (2017)</p>	<p>Comentan que algunos de los factores importantes para tener mayor competitividad respecto a otras empresas, tiene influencia los componentes que proporciona el estado como facilidad para la exportación, transporte, logística, energía tecnología, construcción, ingeniería mecánica y servicios.</p>
<p>Yurievna, G. G., &amp; Aleksandrovich, M. N. (2016)</p>	<p>Introducen dentro del estudio de la competitividad algunos índices macroeconómicos como: competitividad actual, competitividad industrial, desarrollo de infraestructura y comunicaciones, desarrollo regional innovador y factores externos que repercutan en la competitividad interna.</p>

Fuente: Elaboración propia

## Metodología

La metodología considerada para el desarrollo del presente trabajo inicia a partir de la revisión documental que lleva a la exploración exhaustiva de la información y búsqueda bibliográfica como: libros, artículos, periódicos, entre otros, para lograr obtener lo que tenga mayor relevancia.

Por otra parte, se considera de carácter empírico y su alcance es descriptivo, ya que, tiene como objetivo el analizar y correlacionar las distintas teorías que abordan la cooperación empresarial, con la finalidad de generar una perspectiva simplificada del estudio de los distintos aportes sobre la temática (Hernández, Fernández, & Baptista, 2010).

## Conclusiones

Para concluir el presente capítulo considero pertinente el hecho de mencionar que en la actualidad los análisis multivariantes de técnicas dependientes, han sido de gran utilidad en la resolución de las problemáticas actuales, además de que permiten llevar a cabo el estudio sistemático de las variables para dar una posible solución más en concreto, por otra parte, el estudio de la competitividad en muchas de las áreas se conside-

ra completamente cualitativo y al implementar un modelo de regresión lineal múltiple permite el análisis de las distintas variables desde una perspectiva numérica lo que le puede dar una mayor certidumbre a los estudios generados.

Por último, señalar que los análisis multivariantes de técnicas dependientes parte de la primicia que a través de ellos se puede generar un modelo conceptual y pragmático, capaz de establecer aquellas relaciones en los factores que influyen en la generación de la competitividad y así, lograr obtener una perspectiva distinta en el análisis de las variables, para lograr una mejor comprensión de la interrelación entre los elementos que intervienen directa o indirectamente entre variables de la competitividad.

## Referencias

- Allison, P., & Waterman, R. (2002). Fixed-effects negative binomial regression models. *Sociological Methodology*, 32(1), 247-265.
- Álvarez Suárez, M. M., Caballero, A., & Pérez Lechuga, G. (2006). Análisis multivariante: clasificación, organización y validación de resultados. In *Fourth LACCEI International Latin American and Caribbean Conference for Engineering and Technology*, NA.
- Aragón, A., & Rubio, A. (2005). Factores asociados con el éxito competitivo de las pymes industriales en España. *Universia Business Review*, 36.
- Arechavala, R. (2014). Procesos de clusterización en Jalisco, retos del aprendizaje y la colaboración interempresarial. Editorial Universitaria, NA.
- Aryal, G., Mann, J., Loveridge, S., & Joshi, S. (2018). Exploring innovation creation across rural and urban firms: Analysis of the National Survey of Business Competitiveness. *Journal of Entrepreneurship and Public Policy*, 7(4), 357-376.
- Bonales Valencia, J., Zamora Torres, A. I., & Ortíz Paniagua, C. F. (2015). Variables e Índices de Competitividad de las Empresas Exportadoras, utilizando el PLS. *CIMEXUS*, 10(2), 14 - 32.
- Closas, A. H., Arriola, E. A., Kuc, C. I., Amarilla, M. R., & Jovanovich, E. C. (2013). Análisis multivariante, conceptos y aplicaciones en Psicología Educativa y Psicometría. *Enfoques*, 25(1), 65 - 92.
- Cohen, S., Teece, D. J., Tyson, L., & Zysman, J. (1984). Global competition: the new reality. *President Commission on Competitiveness*, 3.

- Dagnelie, P. (1981). Principes d'expérimentation. *Les Presse Agronomique de Gembloux*, NA.
- Escorcía Otalora, T. A. (2008). Análisis bibliométrico como herramienta para el seguimiento de publicaciones científicas, tesis y trabajos de grado . *Bachelor's thesis*.
- Esser, K., Hillebrand, W., Messner, D., & Meyer Stamer, J. (1996). Competitividad Sistémica: nuevo desafío par a las empresas y la política. *CEPAL*(59), NA.
- Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, 18(3), 382-388.
- Gauss, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. *Perthes et Besser*, NA.
- Greene, W. (2003). *Econometric Analysis*. New Jersey: Prentice Hall.
- Hair, J., Anderson, R., Tatham, R., & Black, W. (1999). *Análisis Multivariante*. Madrid, España: Prentice Hall, IBERIA.
- Hall, B., Griliches, Z., & Hausman, J. (1986). Patents and R and D: is there a lag? *International Economic Review*, 27(2), 265-283.
- Hatzichronoglou, T. (1997). Technology and Industry Working Papers Series. *Revision of the High-Technology Sector and Product Classification*, 1 - 25.
- Hernández, R., Fernández, C., & Baptista, M. (2010). *Metodología de la investigación*. México: McGraw Hill.
- Lebart, L., Morineau, A., & Fénelon, J. (1981). *Traitement des données statistiques*. DUNOD, NA.
- Lévin , J., & Valeta, J. (2003). *Analisis multivariante para las ciencias sociales*. Madrid: Pearson.
- Martínez Fierro, S. (2001). Aproximación teórica a los acuerdo de cooperación empresarial. Universidad de Cádiz, *Documento de trabajo de investigación*, 2 - 134.
- Morales, M. A., & Beach, J. L. (2000). Competitividad y estrategia: El enfoque basado en las competencias esenciales y el enfoque basado en los recursos. *Revista Contaduría y Administración* , 197.
- Nyga-Łukaszewska, H., & Chilimoniuk-Przeździecka, E. (2017). Modelling Energy Security and International Competitiveness: The Export Perspective. *Entrepreneurial Business and Economics Review*, 5(2), 71.
- Oxelheim, L. (2003). Macroeconomic variables and corporate performance. *Financial Analysts Journal*, 59(4), 36 - 50.

- Oxelheim, L. (2018). Optimal vs satisfactory transparency: The impact of global macroeconomic fluctuations on corporate competitiveness. *International Business Review*, NA.
- Porter, M. (1991). La ventaja competitiva de las naciones. *Vergara*, NA.
- Porter, M. (1996). Competitive Advantage, Agglomeration Economies, and Regional Policy. *International Regional Science Review*, NA.
- Porter, M. E. (1998). Location, clusters and new microeconomic of competition. *Business Economics*, 33(1), NA.
- Sadaf, R., Oláh, J., Popp, J., & Máté, D. (2018). An Investigation of the Influence of the Worldwide Governance and Competitiveness on Accounting Fraud Cases: A Cross-Country Perspective. *Sustainability*, 10(3), 588.
- Simón, N., & Rueda, I. (2004). Changes in the Mexican steel industry alter privatization: The case of Altos Hornos de México. *Mexico and the World Web Journal*, 9.
- Vázquez Villazón, M., & Bernard, M. E. (1988). Técnicas del análisis estadístico multivariado: objetivo y aplicación. *Revista Cubana de Psicología*, 5(1), 65 - 73.
- Yil, H. T., Han, C. N., & Cha, Y. B. (2018). The Effect of Entrepreneurship of SMEs on Corporate Capabilities, Dynamic Capability and Technical Performances in South Korea. *The Journal of Asian Finance, Economics and Business (JAFEB)*, 5(4), 135-147.
- Yurievna, G. G., & Aleksandrovich, M. N. (2016). Estimation and forecast of regional competitiveness level. *Экономика региона*, 12(4), NA.

## CAPÍTULO 8

# *El análisis multivariante como herramienta para medir los procesos de administración de recursos humanos con la gestión del conocimiento y su relación con la innovación*

*Julio Ceja Sainz*

DIRECTOR DE TESIS

*Carlos Fong Reynoso*

**Palabras clave:** Gestión de recursos humanos, gestión del conocimiento, innovación, regresión lineal, mínimos cuadrados

### **Introducción**

La intención de este trabajo es adentrarse un poco más en la forma como las herramientas de análisis multivalente que pueden contribuir en la gestión de recursos humanos, la gestión del conocimiento y la innovación a través de comparar un par de modelos que proponen como la relación con gestión de recursos humanos y la innovación.

Primero se hablara de un estudio bibliográfico que se realizó para identificar como la gestión de recursos humanos se encuentra ligada con otras teorías como la de gestión del conocimiento y la innovación. Para llevar a cabo este proceso se realizó con la ayuda de *Scopus* una descarga de una base de datos donde aparece la relación de artículos científicos que relacionan el tema de gestión de recursos humanos con temas de gestión del conocimiento como son la teoría de recursos y capacidades, teoría de

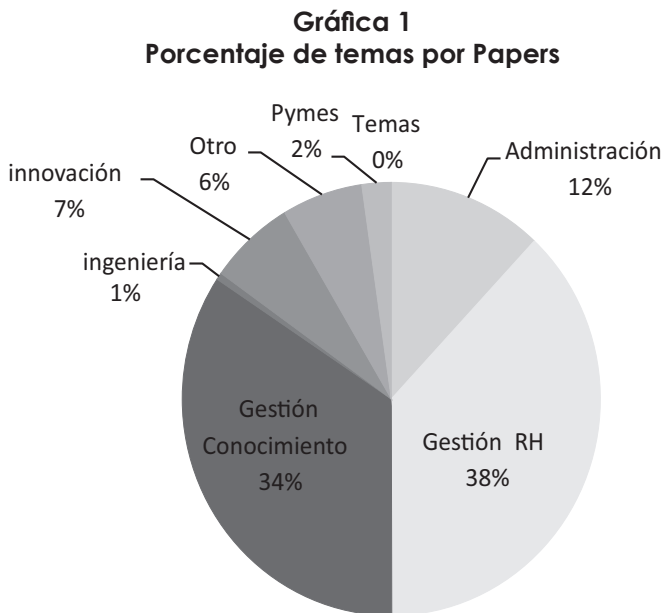
capacidad de absorción, capacidades dinámicas, pequeña y mediana empresa, análisis multivariante. Limitando las áreas de estudio de administración y ciencias sociales.

## Desarrollo

Con estas limitantes *Scopus* arrojo como resultado 339 artículos que oscilan en fecha de publicación de 1986 a hasta 2018. Lo que habla de que se han escrito pocos artículos sobre cómo se relacionan estos temas.

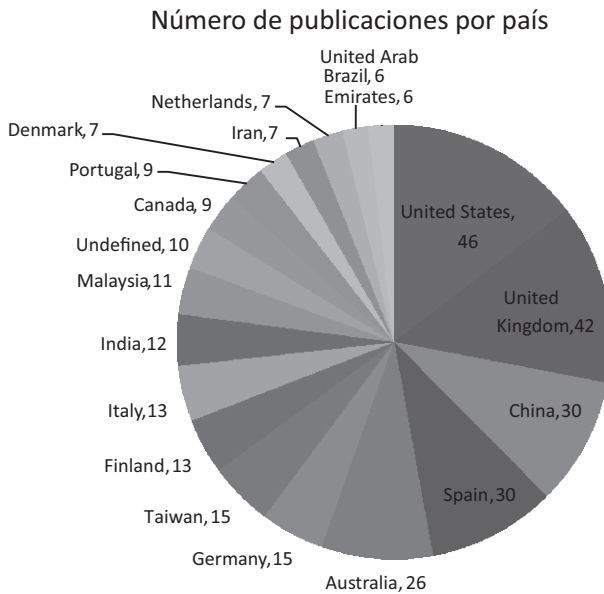
En la siguiente grafica como se puede observar la cantidad documentos que hablan de gestión del conocimiento es casi igual a los que también tocan el tema de gestión de recursos humanos, y es por esto que se podría inferir que esta relación es un tema que tiene bastante interés en el medio de la investigación de la gestión del conocimiento.

En cuanto a la producción de estudios realizados por país, se puede observar el top 20 de países productores de artículos que hablan de las de cómo se relacionan estos temas, vemos que la gran mayoría se divide en 4 países que son Estados Unidos, Reino Unido, China y España.



Fuente: Elaboración propia.

**Grafica 2**  
**Número de publicaciones por país**



Fuente: Elaboración propia.

Es por esto que se señala la importancia de este trabajo es contribuir al conocimiento generado en México. En donde se toma como base modelos realizados con anterioridad y con ellos se sugiere un nuevo modelo que permita profundizar un poco más en estudio de esta relación de gestión del conocimiento, gestión de recursos humanos y la innovación y como este nuevo modelo puede apoyarse en la técnica multivariante de la regresión múltiple para que pueda llevarse esto a cabo.

## Sobre el Análisis Multivariante

A continuación se describirán las generalidades de la regresión lineal.

En la actualidad, nos comenta Mejía (2018) que existe la necesidad de conocer las técnicas estadísticas multivariantes en los campos de las ciencias administrativas.

Según Anormaliza *et al.* (2017) el análisis multivariante es el conjunto de técnicas estadísticas y matemáticas que permiten el procesamiento simultáneo de datos de varias variables y nos comentan que para la elección de técnicas multivariantes dependerá de hacerse la pregunta de ¿qué



si es posible dividir las variables en independientes y dependientes con base en alguna teoría?

La fórmula de la regresión lineal simple se muestra de la siguiente manera:

$$Y = \alpha + \beta X + \varepsilon$$

Dónde:

Y=Es la variable dependiente.

$\alpha$ =Es la ordenada al origen.

$\beta$ =Es la pendiente de la recta e indica cómo cambia Y al incrementar X en una unidad.

$\varepsilon$ = error.

Con el análisis de regresión lineal simple, De la Puente (2009) plantea de que con este método se inicia la estadística multivariante considerada de dependencia y exploratoria; plantea que la regresión lineal es de carácter explicativo-predictivo porque se trata de explicar y/o predecir la variable dependiente considerando su relación con una variable independiente y supone admitir una relación de causa-efecto entre las variables.

Carollo (2012) explica que el análisis multivariante se utiliza cuando se trata de explicar la relación que existe entre variable dependiente (variable de respuesta, "Y") y un conjunto de variables independientes (variables explicativas, "X"). Franco, Gutiérrez y Jiménez (2014) complementan esta idea al afirmar que la relación entre estas variantes pueden ser de dos tipos: Precisa y determinística donde en la primera forma, la variable independiente se determina de forma única por el valor especificado de la variable dependiente las cuales se utilizan mucho en ciencias exactas. Mientras que en la relación imprecisa y estocástica es en la que existen diversos posibles valores de "Y" que pueden estar asociados con cualquier valor de "X". la representación gráfica de ésta relación entre la variable X e Y estará dada por una línea de regresión donde se reducirá al mínimo los errores cometidos al utilizarse para estimar Y a partir de X; siendo las ciencias sociales donde se dan este tipo de relación con mayor frecuencia.

La fórmula de la regresión lineal múltiple se muestra de la siguiente manera:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i$$

Donde

Y = Es la variable dependiente.

$\beta_0$  = Es la pendiente 0 de la recta e indica cómo cambia  $Y$  al incrementar  $X_0$  en una unidad.  
 $\varepsilon_i$  = error.

Según señalan Gutiérrez y De la Vara (2008) la regresión lineal múltiple se usa cuando existen situaciones con varias variables independientes que se cree que influyen o están relacionadas con una variable de respuesta o variable  $Y$ . Mientras que Franco *et al.* (2014) señalan que la conveniencia de su uso cuando resulta insuficiente usar solo una variable dependiente y una independiente para explicar un fenómeno a estudiar. Por su parte Escalante Cortin, Mogollon, Vergara y Schmalbach (2010) comentan que el objetivo de la regresión lineal múltiple es buscar una función de regresión poblacional a partir de una función de la regresión de una muestra y que la idea de esta regresión es mostrar un análisis cuantitativo los fenómenos financieros y económicos que se combinan con la inferencia de la variable explicada.

También estos autores hablan de 4 etapas en la realización de la regresión múltiple:

- 1- Plantear una hipótesis para estudiar un problema basado en una teoría financiera o económica.
- 2- Especificar el modelo matemático de la teoría económica.
- 3- Especificar el modelo econométrico de la teoría económica.
- 4- Obtención de los datos.

Algunos de los indicadores que se analizaran para evaluar el modelo será:

- Prueba de linealidad.
- Prueba de Multicolinealidad.
- Cálculo de Press para obtener una medida de ajuste predictivo.

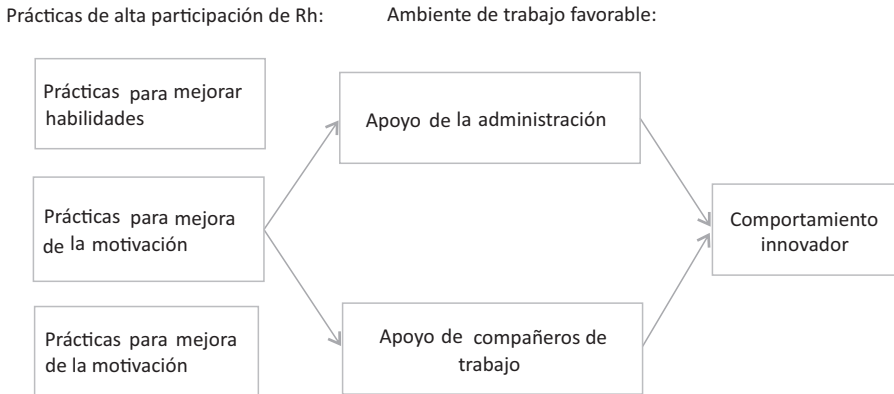
### ***Aplicación en la Gestión de recursos humanos***

Prieto y Pérez-Santana (2014) tratan en su trabajo que se basa en estudiar el entorno, la idea de que las prácticas de recursos humanos se relacionan positivamente con los comportamientos de trabajo innovadores por medio de dos variables del entorno de trabajo: el apoyo de la administración y el apoyo de los compañeros de trabajo y su interrelación que se denominó prácticas de alta participación en la administración de recursos humanos

Para este caso su modelo se realizó de la siguiente forma en 2 familias de variables principales.

Prácticas de alta participación de Rh: prácticas para mejorar habilidades, prácticas para mejora de la motivación, prácticas para mejora de oportunidades. Y por el lado del medio ambiente de trabajo favorable: Apoyo de la administración, apoyo de compañeros de trabajo.

**Ilustración 1.**  
**Modelo de relación entre GRH con el comportamiento innovador**  
**Prieto y Pérez-Santana (2014)**



Fuente: Prieto Pérez-Santana (2014).

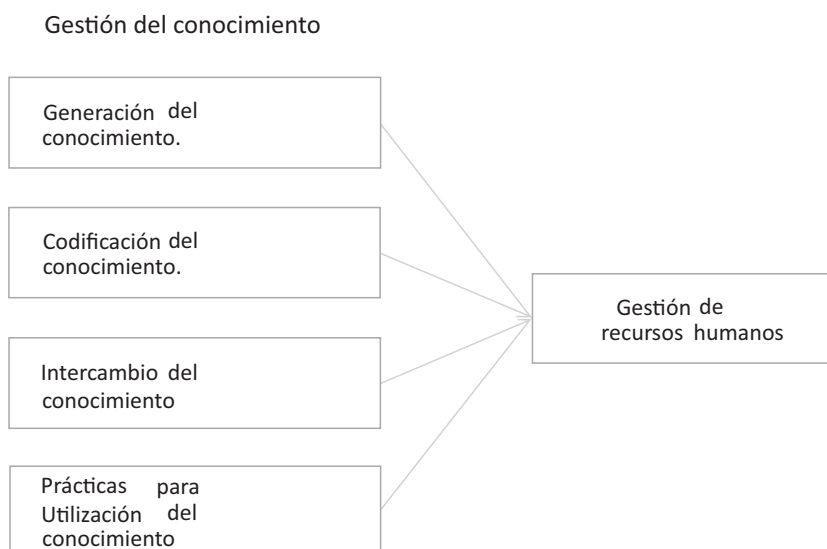
De los resultados obtenidos a través de su estudio sugieren la idea que para que en las organizaciones los empleados tengan una mayor disposición hacia la innovación las empresas deberían de enfocarse en prácticas que propicien un ambiente de apoyo y reconcomiendo a los grupos de trabajo que a los individuos mismos.

Este es un punto interesante a estudiar ya que un entorno que generalice un bienestar común y propicio para la innovación debería de tener más impacto que políticas que se enfoquen a los individuos de una organización de manera individual.

Por su parte que Zaim (2016) trata el tema de que la gestión de conocimiento como la GRH son necesarias para obtener ventajas competitivas sostenibles, al estudiar la relación entre ambas encontró que lo procesos de gestión del conocimiento como la generación, la codificación y la utilización del conocimiento tienen un impacto directo en las prácticas de GRH pero no así en el intercambio de conocimiento que tiene una influencia indirecta.

En el modelo que este autor plantea, toma la gestión de recursos humanos como la variable dependiente y la variable independiente es gestión del conocimiento dividida en 4 sub áreas: Generación del conocimiento, codificación del conocimiento, intercambio del conocimiento y prácticas para mejorar utilización del conocimiento.

**Ilustración 2.**  
**Modelo de relación de gestión del conocimiento con GRH Zaim (2016)**



Fuente: Zaim (2016).

Al final de su trabajo Zaim (2016) observa que todas las sub variables propuestas variables de la gestión del conocimiento tienen una relación directa con la gestión de recursos humanos a excepción del intercambio de conocimiento.

## Discusión

### *Planteamiento de nuevo modelo*

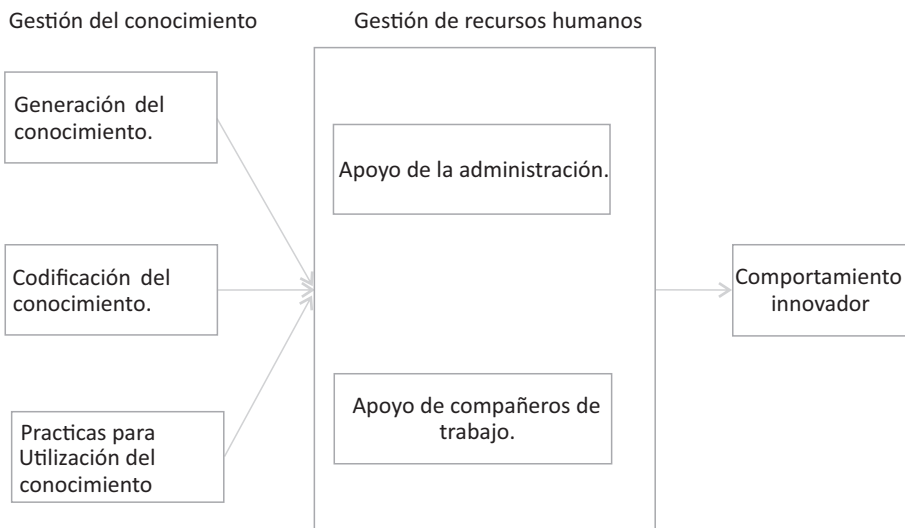
En este trabajo se plantea estudiar la relación de la gestión de recursos humanos y gestión del conocimiento y como esta afecta la innovación to-

mando algunas partes de ambos modelos. No porque se consideren malos sino por el contrario. Ambos tienen sus aportes que pueden ser aprovechados en el caso del modelo de Zaim (2016), se sugiere estudiar solo las 3 variables que muestran una relación significativa con la gestión de recursos humanos.

Solo que en vez de ver a la gestión de recursos humanos como una variable dependiente general, el modelo nuevo se enfocara solo en los aspectos de esta que intervienen en el ambiente propicio para la innovación que es lo que propusieron Prieto y Pérez-Santana (2014).

Así el nuevo modelo que se sugiere ahora consta de 3 variables por un lado en la gestión del conocimiento se estudiara la relación de la generación, la codificación y la utilización de prácticas de la gestión del conocimiento, su relación con la gestión de recursos humanos en los puntos de desarrollo al habiente de innovación que son el apoyo por parte de la administración y el apoyo de compañeros de trabajo, para ver como es la relación de estas y que tanto es la relación positiva entre todas y cual influye más que las otras ya una vez aplicado.

**Ilustración 3.**  
**Modelo sugerido para estudiar la relación de la**  
**gestión de conocimiento con la GRH y comportamiento innovador**



Fuente: Elaboración propia.

La relación de estas variables pueden ser estudiadas con el análisis multivariante ya que como se vio con anterioridad se plantea el hecho de que una variable en este caso la innovación, depende de la gestión de recursos humanos y esta a su vez de la gestión de conocimiento. Por lo que el método de mínimos cuadrados o regresión lineal es por de mas factible.

## Conclusiones

Es importante señalar que la relación de la gestión de recursos humanos y la gestión del conocimiento y la innovación es un fenómeno que debería de estudiarse a profundidad en las empresas ya que este en teoría propicia la innovación dentro de las empresas.

Y es para poder medir que tanta influencia existe entre estas variables que la regresión lineal múltiple juega un papel importante, esto sin querer llegar a decir que se trate de la única forma de medirlo, sin embargo al ser una que ya se probó y tomando en cuenta los aportes de los dos trabajos antes mencionados se podrían realizar estudios a mayor profundidad que nos digan que la magnitud de la relación entre estas y si existiera algún vacío que indique que exista otra variable que no se esté contemplando actualmente.

Al final este trabajo sugiere un nuevo modelo que puede aplicarse para ver con mayor detalle las variables que influyen en el campo de la gestión de recursos humanos y la herramienta analítica para interpretar los datos. Y contribuir en el trabajo hecho en latinoamérica sobre este tema que resulta importante estudiar.

## Referencias

- Anormaliza, R. R., Viejo, F. G., Regnault, M., Holguin, R. P., Lema, R. F., Duarte, F. B., ... Gonzales, J. C. (2017). *Análisis Multivariante: Teoría Y práctica de las principales técnicas*.
- Carollo C. (2012). Regresión Lineal Simple. Estadística FBA I Departamento de Estadística e Investigación Operativa. *Regresion Lineal Simple*. Retrieved from [http://eio.usc.es/eipc1/BASE/BASEMASTER/FORMULARIOS-PHP-DPTO/MATERIALES/Mat\\_50140116\\_Regr\\_simple\\_2011\\_12.pdf](http://eio.usc.es/eipc1/BASE/BASEMASTER/FORMULARIOS-PHP-DPTO/MATERIALES/Mat_50140116_Regr_simple_2011_12.pdf)
- Escalante Cortin, R. D., Mogollon, M., & Vergara Schmalbach, J. C. (2010). *Manual de aplicacion del modelo de regresion lineal multiple*

- con correcciones de especificacion* usos de STATA 9.0, STATA 10.0, EVIEWS 5.0, SSPS 11.0. e-libro, Corp.
- Franco, J. R., Rodríguez, A. I. P., & Jiménez, E. C. R. (2014). *Estadística Aplicada II: Estadística en Administración para Toma de Decisiones*. Grupo Editorial Patria.
- Gutiérrez, Humberto; De la vara, R. (2008). *Análisis y Diseño de Experimentos* (Segunda Ed). México DF: The McGraw-Hill Companies, Inc.
- Mejía, J. (2018). *Análisis Estadístico Multivariante para las Ciencias Económico-Administrativas Teoría y Práctica de las Técnicas Dependientes* (1st ed.). CloudBook.
- Prieto, I. M., & Pérez-Santana, M. P. (2014). Managing innovative work behavior: The role of human resource practices. *Personnel Review*, 43(2), 184–208. <http://doi.org/10.1108/PR-11-2012-0199>
- Puente Viedma, C. de la. (2009). *Estadística descriptiva e inferencial y una introducción al método científico*. e-libro, Corp.
- Zaim, H. (2016). Analysing the effects of knowledge management processes on human resource management practices: a case study on an oil company in the Gulf region. *Middle East Journal of Management*, 3(3), 230–243. <http://doi.org/10.1504/MEJM.2016.079749>

## CAPÍTULO 9

# *El uso de técnicas estadísticas multivariantes mediante el análisis discriminante, aplicado en los negocios, las empresas y organizaciones en general*

*Luis Alberto Arroyo González*

DIRECTOR DE TESIS  
*Guillermo Vázquez Ávila*

**Palabras clave:** Análisis discriminante múltiple, organizaciones, negocios, empresas.

### **Introducción**

El análisis discriminante múltiple es una técnica basada en el análisis estadístico multivariante. En la actualidad su uso se ha hecho muy común en los estudios aplicados a los negocios y las empresas cuando se busca analizar conductas tanto al interior de las organizaciones como al exterior de las mismas, por ejemplo, en los estudios que se ocupan de los comportamientos y preferencias de los consumidores, o en los relacionados con la solvencia financiera de las empresas. Las llamadas escalas de likert son la herramienta ideal cuando se trata de categorizar conductas, tendencias y preferencias en diferentes ámbitos de los negocios, empresas y organizaciones. Su aplicación ha sido ampliamente utilizada en distintas



y muy diversas áreas. El modelo Z de Altman emplea ratios financieros con la finalidad de conocer la situación financiera de las organizaciones.

Este artículo expone algunos trabajos en los cuales se ha utilizado el análisis discriminante múltiple, con la finalidad de mostrar su aplicabilidad y uso. El presente trabajo tiene el objetivo de mostrar “El uso del análisis discriminante múltiple, aplicado en los negocios, las empresas y organizaciones”.

En este ensayo se abordarán diferentes tipos de trabajos académicos, informes de industrias y organizaciones en las que se hace uso de la técnica estadística multivariante, análisis discriminante, la intención es dar a conocer el uso de esta herramienta y su aplicabilidad en distintas y muy diversas áreas. Se comienza por la conceptualización de la herramienta y el método estadístico, también se hace referencia a las escalas likert. El análisis discriminante es utilizado con diferentes métodos como se muestran en los trabajos que se analizan, como es el caso de los métodos de stepwise, *lambda* de wilkins y “Z” Altman. Este trabajo servirá para dar a conocer el uso de la herramienta estadística con la idea de que el lector cuente con un catalogo de trabajos que muestran su aplicabilidad. Al final se presenta un cuadro resumen en orden cronológico por fecha con la referencia bibliográfica de los autores y el aporte de sus trabajos. La recomendación es que sea un primer paso para el lector que se interesa en un trabajo en específico y que pueda consultar mas tarde, el trabajo original a detalle y profundidad. Esperamos encarecidamente que el mismo sea también una guía que pueda conducir, introducir y motivar al lector para el uso y conocimiento de la “Técnica estadística multivariante del análisis discriminante”.

## Desarrollo y conceptualización

Para iniciar, tenemos que comenzar por definir lo que es el análisis discriminante múltiple. Según Mejía (2018) el análisis discriminante múltiple es:

Es la técnica de dependencia multivariante utilizada más extensamente. Su popularidad se basa en su capacidad para predecir y explicar las variables métricas. El propósito básico del análisis discriminante es estimar la relación entre una única variable dependiente no métrica (categórica) y un conjunto de variables independientes métricas, en esta forma general:  $Y_1 = X_1 + X_2 + X_3 + \dots + X_n$ . (Mejía, 2018, p.279)

De la Fuente (2011) acerca del análisis discriminante menciona que:

El análisis discriminante es una técnica estadística multivariante cuya finalidad es analizar si existen diferencias significativas entre grupos de objetos respecto a un conjunto de variables medidas sobre los mismos para, en el caso de que existan, explicar en qué sentido se dan y facilitar procedimientos de clasificación sistemática de nuevas observaciones de origen desconocido en uno de los grupos analizados. (de la Fuente, 2011, p.1)

Las escalas likert, son muy usuales en diversos campos de análisis, de acuerdo a Artigas (2018) una escala de tipo Likert es:

Una escala de Likert es una escala psicométrica utilizada comúnmente en cuestionarios y es la escala de uso más amplio en encuestas de investigación en diversos campos de la industria y la salud. Las escalas de Likert se usan para medir actitudes y opiniones que un nivel más cercano, que una pregunta binaria de sí/no. Las escalas de Likert normalmente usan tres o cinco elementos de opinión. Las de tres elementos usan un máximo, un mínimo y un punto medio o neutral. Las de cinco elementos suelen ser las más usadas en nuestra disciplina, tienen las categorías: Muy satisfecho, parcialmente satisfecho, normal, parcialmente insatisfecho, muy insatisfecho. (Artigas, 2018)

Matas (2018) en su trabajo referente a un estado de la cuestión, menciona que las escalas Likert son un instrumento que se usa mucho en las ciencias sociales y estudios de mercado, aunque existe un debate sobre el método o formato usado para la obtención de datos de calidad.

Por su parte Bertram, citado en (Matas, 2018) argumenta que las escalas likert son pruebas psicométricas, donde la persona objeto de la encuesta debe de seleccionar su acuerdo o desacuerdo, referente a una afirmación, “*ítem* o reactivo, lo que se realiza a través de una escala ordenada y unidimensional” (p.39).

### ***Trabajos de análisis discriminante múltiple***

En éste apartado continuamos con la presentación de trabajos en forma cronológica de acuerdo a la fecha comenzando por el más actual, de los que se presentan y analizan en este ensayo.

Uno de los trabajos que utiliza razones financieras o ratios es el trabajo de, Chávez., Córdova & Alvarado (2017), en este trabajo también se hace uso del análisis discriminante, técnica estadística multivariante.

En su documento se estudian 32 empresas que se denominan “empresas fuertes” y 39 denominadas “empresas débiles”, además se utilizan 12 razones financieras.

Así mismo el objetivo de éste trabajo es el de;

encontrar funciones discriminantes que a través de la caracterización a priori de las compañías y de los resultados de las razones o ratios financiero, muestren el riesgo de la gestión financiera y la probabilidad de desaparición de las compañías industriales manufactureras de la región 7 del Ecuador. (Chavez *et al*, 2017, p.94)

En éste trabajo se utilizaron los ratios financieros o indicadores, los cuáles según (Vallejo, 2012) citado en (Chavez *et al*, 2017) conforman las variables del estudio para la “aplicación de técnicas estadísticas multivariadas como el análisis discriminante” con la finalidad de establecer la pertinencia probabilística y establecer diferencias entre ellos.

Con los ratios financieros se procedió a ordenarlos comenzando por el de mayor valor para las razones financieras y se les dio un valor en orden ascendente comenzando desde el uno para el más alto y así sucesivamente. Para los indicadores de deuda, los valores más elevados correspondieron a valores más bajos, es decir en orden inverso.

Para la aplicación de la técnica de estadística multivariante se utilizó el paquete estadístico del SPSS y para a la clasificación y segmentación se hizo con el método por pasos o *stepwise* y la estimación de *lambda* de Wilkins. (De la Hoz tal.,2013)

El estudio de Chavez *et al*. (2017) denominado, medición del riesgo de la gestión financiera de las compañías con la utilización del análisis discriminante: el caso de las industrias de la región 7 de Ecuador, arroja como resultado de la investigación;

que los ratios, corriente o de liquidez, de rentabilidad del activo total, de apalancamiento total y de rentabilidad del patrimonio son los índices o ratios más importantes a la hora de determinar el riesgo de las industrias. Se concluye que la técnica multivariante del análisis discriminante constituye una de las herramientas gerenciales más importantes al cuantificar el riesgo de la gestión financiera de las compañías. (Chavez *et al*, 2017,p.90)

Herrera (2014) en su documento de investigación intitulado, “Aplicación de análisis discriminante para evaluar la productividad como resultado de la certificación BASC en las empresas de la ciudad de Cartagena” presenta una evolución de la productividad en las empresas de la ciudad

de Cartagena. En este trabajo se busca a través de los indicadores de productividad de 23 empresas pertenecientes a la “Coalición empresarial anticontrabando” (CEAC), por medio del análisis discriminante, explicar la pertenencia y discriminación de las organizaciones para cada grupo de los indicadores de productividad evaluados.

Con la función discriminante obtenida en el trabajo de Herrera y los datos de estadística que se analizaron se puede encontrar que la certificación de la norma “Business Anti Simuling Coalition” (BASC), no representa diferencias significativas.

El trabajo de Herrera, es un referente:

Para desarrollar otras investigaciones que evalúen no sólo variables e indicadores asociados con la productividad, sino otras variables e indicadores asociados con otras variables internas y externas en diferentes sectores que permitan medir eficiencia, productividad y competitividad. De igual manera la investigación permitió establecer una función objetivo para las empresas evaluadas en la ciudad de Cartagena, con lo que se puede estudiar y analizar que indicadores discriminan mejor y se pueden así tomar acciones teniendo en cuenta el cálculo y estudio de los indicadores que presentaron una buena discriminación. (Herrera, 2014,p. 58-59)

El trabajo de Herrera encontró una disminución en la eficacia respecto a la selección de las dos poblaciones estudiadas. Respecto a la muestra de origen el modelo presentó que en 2008 fue de 39.1% mientras que para el 2010 presentó incremento del 56.5%, siendo el promedio total de clasificación del 52% en las empresas, lo anterior de acuerdo al estudio denota una confiabilidad aceptable, para aquellos indicadores que muestran diferencias significativas o que discriminan bien en los periodos de estudio.

Los indicadores o razones financieras utilizadas por el estudio de Herrera se presentan en el siguiente recuadro, como indicadores de productividad.

IP1	(Utilidad bruta) (Valor agregado (ventas - pagos a proveedores + inventarios)) x 100
IP2	(Utilidad operacional) (Valor agregado (ventas - pagos a proveedores + inventarios)) x 100
IP3	(Utilidad neta) (Valor agregado (ventas - pagos a proveedores + inventarios)) x 100

IP4	(Valor agregado (ventas - pagos a proveedores + inventarios) (capital operativo (activos corrientes y fijos)) x 100
IP5	(Utilidad operacional) (capital operativo (activos corrientes y fijos)) x100
IP6	(Utilidad neta) (capital operativo (activos corrientes y fijo)) x100

Fuente: elaboración propia con datos de Herrera (2014).

Contreras & Vera (2014) en su trabajo intitulado “Análisis discriminante aplicado a modelos de predicción de quiebra”, utilizan el análisis discriminante múltiple, con la finalidad de establecer parámetros acerca de la solvencia financiera y el riesgo de quiebra en las organizaciones que tratan con clientes y proveedores, utilizando el modelo Z de Altman.

Altman (1968) hace uso del análisis discriminante múltiple planteando un modelo de predicción de quiebra, que ha sido muy utilizado para estudios que pretenden este fin. Los modelos de análisis discriminante múltiple basados en el modelo Z de Altman emplean para su análisis las llamadas razones financieras o ratios para conocer la situación financiera de la empresa.

Según Contreras, *et al*, (2014) en el empleo de los modelos de análisis discriminante múltiple, se emplean las razones financieras o ratios como variables independientes, porque, permiten realizar comparaciones entre organizaciones de diferentes tamaños. Es importante señalar que existe una diferencia entre empresas de distintos sectores económicos para el análisis de los ratios como variables independientes.

En su trabajo los autores de “Análisis discriminante aplicado a modelos de predicción de quiebra”, emplean el modelo de Altman, para determinar la solvencia financiera de la, “Caja Municipal de Ahorro y Crédito Pisco” ya que se incumplió con los compromisos que se tenían en el plan de recuperación financiera. La misma institución poseía una alta tasa de morosidad además de falta de liquidez.

El trabajo concluye argumentando en los resultados que:

El modelo no predice, cuándo una empresa se manifestará en quiebra, pero si mide el desempeño financiero y lo que puede ocasionar la quiebra de las organizaciones. La herramienta según los autores del artículo no debe ser tomada o considerada una prueba absoluta de solvencia de una empresa, pero sí como una herramienta importante para su evaluación. Además según los autores se debe de tomar en cuenta que Altman, se valió de datos financieros pertenecientes

a un número reducido de empresas y en un contexto de más de 45 años atrás, para el diseño de su modelo, pero que a la fecha sigue siendo utilizado, sin embargo el modelo de Z Altman, ha resultado ser una herramienta estadística de importancia, con la que se puede diagnosticar financieramente a las empresas en lo relacionado a las posibilidades de quiebra y la solvencia de las mismas (Contreras & Vera, 2014).

Los autores encuentran que el modelo Z Altman no es un indicador perfecto. Por las siguientes razones:

- No considera quiebras causadas por los factores aparte de aquellos que aparecen sobre el balance.
- Se ve afectada por empresas que manipulan su información contable para conservar su imagen corporativa.
- No es útil para las nuevas compañías con poco o ningún ingreso.
- No es adecuado para evaluar empresas pequeñas. (Contreras, *et al* 2014, p.58)

Las razones o ratios financieros, que son utilizadas como variables independientes usados en el modelo Z Altman, están basadas en las transacciones con las que se generan los estados financieros registrados por la contabilidad de las empresas y son las siguientes: Como primer paso se determinan los dos grupos, el grupo a) Insolvencia, y el grupo B) Solvencia. Para éste caso y como una de las ventajas del análisis discriminante múltiple (ADM) planteado por Altman, se caracteriza por la disminución de la cantidad de variables dependientes, para éste caso en particular se usó sólo la variable dependiente Insolvencia.

Así la función discriminante tiene la forma:

$$Z = V_1X_1 + V_2X_2 + \dots + V_nX_n.$$

Según los valores obtenidos Z se pueden obtener 3 posibles resultados.

1. Saludable; no presentará insolvencia en el futuro.
2. Zona gris; tiene alto riesgo de presentar problemas de insolvencia los siguientes 2 años.
3. Enferma; la insolvencia será muy alta y quebrará probablemente en 2 años.

Para empresas manufactureras que cotizan en bolsa la función discriminante según en estudio de Contreras fue:

$$Z = 1.23X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 0.99X_5$$

$X_1$  = Capital de trabajo / activo total.

$X_2$  = Utilidades retenidas / activo total.

$X_3$  = Utilidades antes de intereses e impuestos / activo total.

$X_4$  = Valor de mercado de las acciones / pasivo total.

$X_5$  = Ventas / activo total.

Fuente: elaboración propia con datos de Contreras & Vera (2014).

En la cual:

El resultado de Z se debe de interpretar de la siguiente forma

$Z \geq 2.99$ : Saludable

$2.99 < Z < 1.81$ : Zona gris

$Z < 1.81$  : Enferma

Alonso (2010) en su trabajo sobre, el análisis y medición del liderazgo, trata de averiguar el estado del fenómeno del liderazgo, proponiendo algunas hipótesis que permitan en lo subsecuente ser contrastadas por medio del software SPSS1. Se utiliza la escala de medición Likert, con cinco categorías; siempre, casi siempre, normalmente, a veces y casi nunca.

Con la obtención de los datos proporcionados por las encuestas se comprueba la validez de la escala utilizada, así como los indicadores utilizados por medio de la aplicación SPSS.

En éste trabajo, para la evaluación del modelo de medida se analizan los conceptos teóricos por medio de las variables observadas. Se comprueba si los indicadores miden lo que deben de medir.

La validez discriminante, indica en qué medida un constructo dado es diferente de los otros constructos en el modelo. Para que exista dicha validez en un constructo han de existir correlaciones débiles entre éste y otras variables latentes que midan fenómenos diferentes. Esto supone que la raíz cuadrada de AVE<sup>2</sup> sea mayor que la correlación existente entre los constructos. (Alonso, 2010, p.27)

El trabajo de Alonso (2010) concluye, en que las características del líder eficaz según las encuestas con las escalas Likert y el análisis discrimi-

---

<sup>1</sup> IBM® SPSS Statistics® es el principal software estadístico que ofrece técnicas de recolección de datos y analítica predictiva para solucionar múltiples problemas empresariales y de investigación. Brinda varias técnicas, que incluyen pruebas de hipótesis lo que facilita la gestión de los datos, la selección y la ejecución de análisis y el intercambio de resultados, así como análisis multivariados y de sobrevida.

<sup>2</sup> AVE=Varianza Promedio Extraída

minante, es una persona con carácter sincero, pero ambicioso, entregado, comprometido con su trabajo y que transmite positividad, lo que significa que se necesita de una forma en específico, que permite el logro de objetivos planteados.

El trabajo de Alonso propone propuestas de mejora, para el desarrollo del liderazgo, estas propuestas son el reflejo de los hallazgos del trabajo de investigación, por tal motivo y de acuerdo al análisis con la escala Likert, se puede observar que existe una mayor tendencia de variabilidad entre los valores 3 y 4 de la escala. Esto quiere decir que se muestran en la empresa analizada y se encuentran entre los aspectos “normalmente” y “casi siempre”. Siempre se aspira a lo más alto, realmente los factores “compromiso, sinceridad, ambición y conocimiento de sí mismo” se encuentran en un buen lugar, por lo que las propuestas de desarrollo para el liderazgo van referidas al resto de los factores.

Propuestas:

- Cultivar relaciones con los empleados, mostrarles atención, orientación.
- Organizar breves reuniones semanales de los empleados con sus responsables para comunicarse acerca de la situación real de la empresa.
- Establecer reuniones periódicas con el objeto de revisar la dirección que sea de seguir en la organización.
- Fomentar la realización de cursos de formación para empleados y directivos.
- Coaching empresarial, con el objetivo de desarrollar habilidades específicas como la toma de decisiones, la administración del cambio y la capacidad organizativa.
- Promover desde la dirección un ambiente de entusiasmo y positividad. Los entornos crispados no favorecen el rendimiento laboral. (Alonso, 2010, p.39)

Nava y Marbelis (2009), en su artículo denominado; *Análisis financiero*: una herramienta clave para una gestión financiera eficiente, argumentan que el análisis discriminante es un técnica estadística, que tiene la predominancia de ser utilizada para analizar los aspectos financieros, los mismos señalan que Edwar Altman, la utilizó desde 1968, con la finalidad de seleccionar indicadores financieros, que brindaran información sobre aspectos de las empresas que tenían riesgo de quiebra y las que presentaban éxito financiero.

El análisis discriminante según Elizondo y Altman (2003), citados en Nava y Marbelis (2009), sirve básicamente para:



Clasificar observaciones previamente obtenidas, cuya clasificación tiene como punto de partida un conjunto de variables que caracterizan los individuos u objetos que se pretenden estudiar. Desde el punto de vista financiero, esta técnica se fundamenta en la combinación y estudio de los indicadores o razones financieras de una empresa o de un grupo de empresas. (Nava y Marbelis, 2009, p.18)

Nava y Marbelis, mencionan que Altman diseño un modelo con la intención de predecir la quiebra de las empresas y el mismo lo llamo “Z de Altman”, el modelo representa el resultado de aplicar el análisis discriminante a un grupo de indicadores financieros con el objetivo de seleccionar a las organizaciones en dos grupos: bancarota y no bancarota.

El resultado del análisis discriminante, muestra la correlación entre las variables individuales seleccionando los que contribuyen más al valor discriminante denominado valor de Z, el cuál ha tenido una gran aceptación.

La función discriminante que resultó del valor de Z de Altman es la siguiente:

$$Z = 6.56X1 + 3.26X2 + 6.72X3 + 1.05X4$$

donde X = Indices seleccionados para confirmar el valor Z

valores X	Relación
X1	Capital de Trabajo/ Activo Total
X2	Utilidades Retenidas/Activo total
X3	Utilidad antes de impuestos/ Activo total
X4	Total patrimonio/Pasivo Total

Fuente: Elaboración propia, cuadro tomado de Nava y Marbelis (2009).

### ***Limtes de Referencia***

Z >= 2.60 Baja probabilidad de quiebra; puede tratarse de una empresa financieramente bien gerenciada que no muestra posibilidad de llegar a la quiebra en un futuro cercano; es decir, implica que es una empresa saludable desde la perspectiva financiera.
Z <= 1.10 Alta probabilidad de quiebra.

1.10 < Z < 2.60 Zona gris o de incertidumbre; donde puede ocurrir que las empresas con valores Z ubicados entre estos límites, sean muy buenas pero se encuentre gerenciadas en forma ineficiente, o empresas malas pero gerenciadas con la mayor eficiencia.

Fuente: Elaboración propia, con datos de Nava y Marbelis (2009).

Estos parámetros indican que mientras mayor sea el valor de Z, menor es la probabilidad de que la empresa se vaya a la quiebra. Resulta importante destacar que este modelo puede ser aplicado para un grupo de empresas o para una empresa individual. Para un grupo de empresas, como un primer paso se debe calcular cada uno de los indicadores que contiene el modelo para cada empresa de la muestra estudiada; posteriormente, se calcula el valor promedio de cada índice en el grupo de empresas; y por último, se sustituye en la función discriminante para obtener el valor de Z. (Nava y Marbelis, 2009,p.622)

Para aplicar el modelo en una sola organización, se lleva a cabo el mismo procedimiento, se realizan los cálculos de las variables del modelo y el resultado obtenido se sustituye en la función discriminante para obtener el dato de Z. Por último se realiza un comparativo con los datos de referencia y se realiza el diagnostico de las organizaciones estudiadas.

Andrés (2000) en su artículo sobre “Los parámetros característicos de las empresas manufactureras de alta rentabilidad. Una aplicación del análisis discriminante” utiliza el análisis discriminante para determinar las características que hacen una marcada diferencia entre organizaciones manufactureras que logran obtener una alta rentabilidad con respecto a su sector de actividad económica versus las de menor rentabilidad relativa.

El trabajo de Andrés (2000) se llega a la conclusión de que por medio del análisis discriminante se puede determinar que características economico financieras, diferencian a las organizaciones del sector manufacturero de alta y baja rentabilidad económica, siguiendo la metodología propuesta en el trabajo, se encuentra que las cuestiones de rentabilidad que se han tratado con métodos de clasificación, muestran algunos problemas para encontrar su verdadero nivel de rentabilidad, por medio de funciones matemáticas.

En su trabajo Andrés (2000) realiza la medición de la rentabilidad, de la siguiente manera:

Para medir la rentabilidad se ha utilizado la rentabilidad económica (RE), en detrimento de los indicadores de rentabilidad financiera. Ello se debe a que se ha querido eliminar el efecto que sobre la rentabilidad de la empresa tiene la forma en que la misma está financiada. La especificación concreta es la siguiente:

$$\text{RE} = \text{Cif: Negocios} - \text{G. Explotación} / \text{I. Material} + \text{I. Inmaterial} + \text{Act. Circulante} - \text{Inv. Fin. Temp.}$$

**Cuadro de términos utilizados en la razón financiera RE (rentabilidad) según, Andrés (2000)**

CONCEPTO	DEFINICION
Cif: Negocios	Cifra de negocios; importe de las ventas y prestaciones de servicios realizados en el ejercicio, netas de rappels, devoluciones y operaciones similares.
G. Explotación	Gastos de explotación, suma de los consumos de explotación, los gastos de personal, las dotaciones a la amortización del inmovilizado, la variación de provisiones de tráfico, las pérdidas de créditos incobrables y otros gastos de explotación.
I. Material	Inmovilizado material neto de amortizaciones y provisiones
I. Inmaterial	Inmovilizado inmaterial, neto de amortizaciones y provisiones.
Act. Circulante	Activo circulante, total de partidas de activo circulante del balance, a excepción de los derechos frente a accionistas por desembolsos exigidos y las acciones propias a corto plazo.
Inv. Fin. Temp.	Inversiones financieras temporales, apartado D.IV del activo del balance abreviado, que incluye todas las inversiones financieras temporales, tanto en empresas del grupo y asociadas como las realizadas en otras compañías.

Fuente: elaboración propia con datos de Andrés (2000)

Muñoz (1998) en su trabajo denominado “La técnica de análisis discriminante: una aplicación para el área bancaria” define que el análisis discriminante, busca identificar, a partir de un conjunto de variables, si existe la posibilidad de discriminar, si un dato pertenece a un grupo en específico entre los grupos existentes, determinar a que grupo de indica-

dores suman más al proceso de discriminación y además facilita calcular funciones para la determinación de la ubicación de casos nuevos.

La autora en su trabajo pretende demostrar mediante un ejemplo práctico la aplicación del método de análisis discriminante para saber que indicadores ayudan a discriminar entre bancos de altas utilidades y bajas utilidades, utilizando las variables independientes; obligaciones con el público, cartera de crédito vigente, ingresos financieros e ingresos por comisiones (servicios).

Muñoz (1998), sugiere que los métodos comunes en los temas financieros, en específico el de razones financieras, hacen imperante el considerar la advertencia de el empleo del análisis discriminante para la visualización de organizaciones o comportamientos atípicos.

Los hallazgos de Muñoz, en este estudio son los siguientes:

De las variables consideradas la que más aporta a la discriminación entre bancos de altas y bajas utilidades es ingresos por concepto de intermediación financiera, seguido de la cartera de créditos vigente, variables que en el fondo están relacionadas una con otra, los ingresos por comisiones por servicios aportan un poco a la discriminación, pero en lo que respecta a las obligaciones que los bancos mantienen con el público no hay aporte importante a la discriminación. (Muñoz, 1998,p.24)

Las variables utilizadas en el trabajo de Muñoz, fueron las siguientes:

Como variables dependientes tenemos a los bancos de altas utilidades y bancos de bajas utilidades, es decir se establecen dos grupos.

Las variables independientes fueron; Obligaciones con el público, cartera de crédito vigente, ingresos financieros e ingresos por comisiones (servicios) y la información que se utilizó corresponde al año de 1995.

## **Discusión**

El presente ensayo presenta varias propuestas de modelos, en los cuales se utiliza el método de análisis multivariante discriminante en diversas áreas, el lector será el que tenga la última palabra en la elección de la herramienta que más se ajuste a su estudio según sea el caso. Es difícil inclinarse por uno u otro ya que cada uno maneja sus particularidades y depende de la utilidad que se le quiera dar o del tipo de información que se quiera obtener. La recomendación sería en todo caso que si se quieren

utilizar el método para saber la rentabilidad de una empresa, el riesgo de quiebra, el liderazgo dentro de una organización siempre se deberá buscar que el grupo de empresas sujetas del análisis del del mismo giro, es decir, no se deben mezclar en un mismo análisis por ejemplo de riesgo de quiebra a un empresa del sector calzado y analizar con el mismo modelo y en el mismo estudio una empresa del sector restauranero, en todo caso se deberá de utilizar un análisis sólo para empresas de un mismo ramo, por ejemplo en el ramo de la construcción incluso es diferente una empresa que se dedica a construir vivienda que una constructora que se dedica a construir puentes y carreteras, así siempre y cuando se cuiden estos aspectos los hallazgos de los estudios y análisis de las industrias serán más confiables e incluso el proceso de recolección de datos será siempre mas homogéneo facilitando así la tarea del investigador.

## **Conclusiones**

La técnica estadística multivariante a través del análisis discriminante es una herramienta que como ya se vio es bastante utilizada en los negocios, empresas y organizaciones, sin duda en el área de la consultoría, otorgará beneficios tanto a los empresarios como a los consultores. Es importante que desde el área de las ciencias de la administración se aprenda bien a utilizar este tipo de herramientas y se utilice para el análisis de la industria en general ya sea en trabajos académicos o privados. Dependiendo del análisis que se quiera llevar a cabo, será el tipo de herramienta de estadística multivariante que se utilice. Los trabajos de análisis, consultoría, académicos y privados en los que se puede implementar el método en mención es muy grande. El que domine este tipo de instrumentos, en cualquier ámbito, podrá ofrecer servicios a las organizaciones a cambio de un buen pago, convirtiéndose en un “modus vivendi” para el oferente de servicios de consultoría a empresas, negocios y organizaciones.

**Cuadro de autores y trabajos realizados con la técnica estadística multivariante, análisis discriminante.**

<b>Autor</b>	<b>Título del trabajo</b>	<b>Técnica estadística multivariante</b>
Chávez, N., Córdova, C., & Alvarado, P. (2017)	Medición del riesgo de la gestión financiera de las compañías con la utilización del análisis discriminante el caso de la región 7 de Ecuador.	Análisis Discriminate. Stepwise. Lambda de Wilkins.
Herrera, T. J. F. (2014)	Aplicación de análisis discriminante para evaluar la productividad como resultado de la certificación BASC en las empresas de la ciudad de Cartagena.	Análisis discriminante.
Contreras, J. C. A., & Vera, A. F. N. (2014)	Análisis discriminante aplicado a modelos de predicción de quiebra.	Análisis discriminante Altman
Alonso Felix, A. (2010)	Análisis y medición del liderazgo en una empresa del sector automoción y propuesta de desarrollo (Proyecto fin de pregrado).	Análisis discriminante. Escalas Likert.
Nava Rosillón, Marbelis Alejandra. (2009).	Análisis financiero: una herramienta clave para una gestión financiera eficiente.	Análisis discriminante.
Andrés Suárez, J. D. (2000)	Los parámetros característicos de las empresas manufactureras de alta rentabilidad. Una aplicación del análisis discriminante.	Análisis discriminante.
Muñoz Salas, E. (1998).	La técnica de análisis discriminante: una aplicación para el área bancaria.	Análisis discriminante.

Fuente: Elaboración propia, con datos de las referencias bibliográficas de este mismo documento.

## Cuadro de resultados de los trabajos presentados

Autores	Titulo del trabajo	Resultados encontrados.
Chávez, N., Córdova, C., & Alvarado, P. (2017)	Medición del riesgo de la gestión financiera de las compañías con la utilización del análisis discriminante el caso de la región 7 de Ecuador.	Los ratios corriente o de liquidez, de rentabilidad del activo total, de total de apalancamiento total y de rentabilidad del patrimonio son los índices o ratios más importantes a la hora de determinar el riesgo de las industrias. Se concluye que la técnica multivariante del análisis discriminante constituye una de las herramientas gerenciales más importantes al cuantificar el riesgo de la gestión financiera de las compañías. (Chavez <i>et al</i> , 2017, p.90)
Herrera, T. J. F. (2014)	Aplicación de análisis discriminante para evaluar la productividad como resultado de la certificación BASC en las empresas de la ciudad de Cartagena.	El trabajo de Herrera encontró una disminución en los eficaz respecto a la selección de de las dos poblaciones estudiadas. Respecto a la muestra de origen el modelo presentó que en 2008 fue de 39.1% mientras que para el 2010 presentó incremento del 56.5%, siendo el promedio total de clasificación del 52% en las empresas, lo anterior de acuerdo al estudio denota una confiabilidad aceptable, para aquellos indicadores que muestran diferencias significativas o que discriminan bien en los periodos de estudio.

<p>Contreras, J. C. A., &amp; Vera, A. F. N. (2014)</p>	<p>Análisis discriminante aplicado a modelos de predicción de quiebra.</p>	<p>El modelo no predice, cuándo una empresa se manifestará en quiebra, pero si mide el desempeño financiero y lo que puede ocasionar la quiebra de las organizaciones. La herramienta según los autores del artículo no debe ser tomada considerada una prueba absoluta de solvencia de una empresa, pero sí como una herramienta importante para su evaluación. Además según los autores se debe de tomar en cuenta que Altman, se valió de datos financieros pertenecientes a un numero reducido de empresas y en un contexto de más de 45 años atrás, para el diseño de su modelo, pero que al fecha sigue siendo utilizado, sin embargo el modelo de Z Altman, ha resultado ser una herramienta estadística de importancia, con la que se puede diagnosticar financieramente a las empresas en lo relacionado a las posibilidad de quiebra y la solvencia de las mismas.</p>
<p>Alonso Felix, A. (2010).</p>	<p>Análisis y medición del liderazgo en una empresa del sector automoción y propuesta de desarrollo (Proyecto fin de pregrado).</p>	<p>El trabajo de Alonso, concluye en que las características del líder eficaz según las encuestas con las escalas Likert y el análisis discriminante, es una persona con carácter sincero, pero ambicioso, entregado, comprometido con su trabajo y que transmite positividad, lo que significa que se necesita de una forma en específico, que permite el logro de objetivos planteados.</p>



<p>Nava Rosillón, Marbelis Alejandra. (2009).</p>	<p>Análisis financiero: una herramienta clave para una gestión financiera eficiente.</p>	<p>El resultado del análisis discriminante, muestra la correlación entre las variables individuales seleccionando los que contribuyen más al valor discriminante denominado valor de Z, el cuál ha tenido una gran aceptación.</p>
<p>La función discriminante que resultó del valor de Z de Altman es la siguiente:  <math>Z = 6.56X_1 + 3.26X_2 + 6.72X_3 + 1.05X_4</math>; donde X = Índices seleccionados para confirmar el valor Z          Andrés Suárez, J. D. (2000)</p>	<p>Los parámetros característicos de las empresas manufactureras de alta rentabilidad. Una aplicación del análisis discriminante.</p>	<p>Llega a la conclusión de que por medio del análisis discriminante se puede determinar que características económico financieras, diferencian a las organizaciones del sector manufacturero de alta y baja rentabilidad económica, siguiendo la metodología propuesta en el trabajo, se encuentra que las cuestiones de rentabilidad que se han tratado con métodos de clasificación, muestran algunos problemas para encontrar su verdadero nivel de rentabilidad, por medio de funciones matemáticas.</p>
<p>Muñoz Salas, E. (1998).</p>	<p>La técnica de análisis discriminante: una aplicación para el área bancaria.</p>	<p>De las variables consideradas la que más aporta a la discriminación entre bancos de altas y bajas utilidades en ingresos por concepto de intermediación financiera, seguido de la cartera de créditos pero en lo que respecta a las obligaciones que los bancos mantienen con el público no hay aporte importante a la discriminación.</p>

Fuente: Elaboración propia, con datos de las publicaciones presentadas en este mismo documento.

## Referencias

- Altman, E. I. (1968) Financial ratios, discriminate analysis and the prediction of corporate bankruptcy, *Journal of Finance*, September 1968, pp. 589-609.
- Alonso Felix, A. (2010). *Análisis y medición del liderazgo en una empresa del sector automoción y propuesta de desarrollo* (Proyecto fin de pregrado). Centro Politécnico Superior Universidad de Zaragoza, España.
- Andrés Suárez, J. D. (2000). Los parámetros característicos de las empresas manufactureras de alta rentabilidad. Una aplicación del análisis discriminante. *Revista Española de financiación y contabilidad*, 29(104), 443-481.
- Artigas, S.G. (2018). *Escala de Likert*. Recuperado de <http://www.torresburriel.com/weblog/2018/06/12/escala-de-likert/>
- Bertram, D. (2008). *Likert Scales... are the meaning of life*. Topic report: Recuperado de <http://poincare.matf.bg.ac.rs/~kristina/topic-dane-likert.pdf>
- Contreras, J. C. A., & Vera, A. F. N. (2014). Análisis discriminante aplicado a modelos de predicción de quiebra. *Facultad de Ciencias Contables*, 22, 53.
- Elizondo, Alan y Altman, Edward (2003). *Medición Integral del Riesgo de Crédito*. Limusa. 200 pp.
- Chávez, N., Córdova, C., & Alvarado, P. (2017). Medición del riesgo de la gestión financiera de las compañías con la utilización del análisis discriminante: el caso de las industrias de la región 7 del Ecuador. *Revista Publicando*, 4(13), 90-107.
- De la Fuente Fernández, S. (2011). *Análisis discriminante*. Fac de ciencias económico y empresariales. Universidad Autónoma de Madrid. Recuperado de <http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/DISCRIMINANTE/analisis-discriminante.pdf>
- Herrera, T. J. F. (2014). Aplicación de análisis discriminante para evaluar la productividad como resultado de la certificación BASC en las empresas de la ciudad de Cartagena. *Contaduría y administración*, 59(1), 43-62.
- Matas, A. (2018). Diseño del formato de escalas tipo Likert: un estado de la cuestión. *Revista Electrónica de Investigación Educativa*. 20(1), 38-47. <https://doi.org/10.24320/redie.2018.20.1.1347>

- Mejía Trejo, J. (2018). *Análisis estadístico multivariante con SPSS para las ciencias económico-administrativas teoría y práctica de las técnicas dependientes*. Jalisco, México: D.R. 2018 CLOUDBOOK
- Muñoz Salas, E. (1998). La técnica de análisis discriminante: una aplicación para el área bancaria. San José: Banco Central de Costa Rica, División Económica, Departamento de Investigaciones Económicas.
- Nava Rosillón, Marbelis Alejandra. (2009). Análisis financiero: una herramienta clave para una gestión financiera eficiente. *Revista Venezolana de Gerencia*. 14(48), 606-628. Recuperado en 25 de noviembre de 2018, de [http://www.scielo.org/ve/scielo.php?script=sci\\_arttext&pid=S1315-99842009000400009&lng=es&tlng=es](http://www.scielo.org/ve/scielo.php?script=sci_arttext&pid=S1315-99842009000400009&lng=es&tlng=es).
- Vallejo. G (2012), *Análisis multivariantes aplicados a las ciencias comportamentales*. Servicio de publicaciones de la Universidad de Oviedo, 14(1), 1-20.

# CAPÍTULO 10

## *Investigaciones científicas en RH: El desafío que deben encarar los profesionales*

*Miguel Angel Hernández González*

DIRECTOR DE TESIS

*José Sánchez Gutiérrez*

**Resumen:** El acelerado ritmo de cambios derivados principalmente de la innovación tecnológica, la globalización de los negocios, la mentalidad de las nuevas generaciones de trabajadores, y la continua variación de escenarios en los mercados, son unos de los principales desafíos que las empresas enfrentan hoy en día. Dichas situaciones las obligan a generar nuevo conocimiento y a buscar mecanismos que les permitan contar con información de calidad para tomar mejores decisiones e implementarlas a mayor velocidad que sus competidores.

En este sentido, los profesionales de recursos humanos deben asumir el desafío de realizar investigaciones pertinentes de carácter científico que agreguen valor a su organización, y les permita diseñar e implementar estrategias eficaces para adaptarse al entorno y gestionar el cambio de tal manera que juegue en beneficio de la empresa al representar, más que una amenaza, una oportunidad.

**Palabras clave:** Conocimiento científico, investigación, factor humano, técnicas de análisis de información.

**Abstract:** The accelerated pace of changes derived mainly from technological innovation, the globalization of business, the mentality of the

new generations of workers, and the continuous variation of scenarios in the markets, are some of the main challenges that companies face today. These situations force them to generate new knowledge and to look for mechanisms that allow them to have quality information to make better decisions and implement them at a higher speed than their competitors.

In this sense, human resources professionals must take on the challenge of carrying out relevant scientific research that adds value to their organization, and allows them to design and implement effective strategies to adapt to the environment and manage change in such a way that they benefit of the company to represent, more than a threat, an opportunity.

**Key words:** Scientific knowledge, research, human factor, information analysis techniques.

## Introducción

La investigación científica que genera nuevo conocimiento, durante mucho tiempo ha sido desarrollada casi en su totalidad dentro del ámbito académico u universitario (Calvo, 2008). Sin embargo, las empresas que no sólo generan riqueza material y que no se limitan exclusivamente a producir beneficios “tangibles”, sino que se esfuerzan por crear una amplia gama de bienes “intangibles” sin los que la vida humana pierde buena parte de su calidad (Álvarez de Mon, 1998), también tienen la necesidad de crear conocimiento y estudiar su propia realidad de manera profesional, es decir, con apoyo del método científico.

Especialmente, los líderes que dirigen el área estratégica de recursos humanos tienen el desafío de realizar investigaciones que les permitan tener una administración eficiente y eficaz del factor más importante y complejo de la organización, a saber, el factor humano. Ya que las empresas exitosas perciben que sólo pueden crecer, prosperar y mantener su continuidad si son capaces de optimizar el retorno sobre las inversiones de todos los socios, en especial, de los colaboradores (Werther, 2001).

Dentro del ámbito empresarial, para poder tomar decisiones más precisas en beneficio de la organización y de su personal, los directivos del área de recursos humanos deben contar con las herramientas necesarias para crear sistemas de información que tengan un alto grado de confiabilidad. Un sistema de información sobre los recursos humanos según Herrera citado en Calvo (2008) es el conjunto de datos, recursos mate-

riales e inmateriales; y personas, que dentro de la empresa se dedican a conseguir, almacenar, tratar y distribuir información relevante para tomar decisiones sobre alguna de las funciones de la dirección, la gestión o la administración de personas.

Las alteraciones del medio, asociadas principalmente a la revolución científica y tecnológica, demandan en el personal de las organizaciones una gran cantidad de energía y habilidades que antes no eran tenidas en cuenta, como saber trabajar con algoritmos o la adaptabilidad para trabajar con la digitalización o los continuos cambios (Carey, 2017). Dichas habilidades se deben desarrollar para afrontar exitosamente las exigencias y amenazas de un entorno en el que, como decía Heráclito, lo único constante es el cambio. Una de las virtudes importantes a cultivar en este contexto, es la capacidad de aprendizaje que Senge considera la principal ventaja competitiva de las empresas (1990).

En el mundo actual que Druker ha denominado la era de las organizaciones, los grandes desafíos que enfrentan las empresas las obligan a estar a la vanguardia en todas sus líneas, y generar un conocimiento científico que constituye el pilar donde se sostiene su ventaja competitiva y se gesta el potencial para crecer e incursionar en nuevos mercados.

Por tanto, el presente ensayo busca argumentar el por qué los profesionales de recursos humanos deben investigar la realidad de sus organizaciones apoyándose en el método científico que integre no sólo la parte cualitativa, sino también la parte cuantitativa mediante las técnicas de análisis multivariante para que el fruto de sus investigaciones tenga una mayor veracidad, sea más confiable y permita hacer predicciones que faciliten la conducción de la empresa y la toma de decisiones por parte de la alta dirección.

## **Desarrollo**

Las investigaciones de carácter científico llevadas a cabo dentro de las organizaciones con el objetivo de estudiar al factor humano, pueden ser consideradas como estudios sociales en el sentido de que analizan a un grupo de seres humanos en un entorno especial y que tiene ciertas características. Las empresas durante mucho tiempo no repararon en la importancia que tiene investigar con seriedad, es decir, con un método científico los fenómenos sociales que se suscitaban en sus estructuras.

Sin embargo, en este nuevo contexto marcado por el cambio, las organizaciones empresariales empiezan a ser conscientes de que investigar

sobre la realidad humana que las coforma supone ir un paso más allá. Esto conlleva implicarse en una actividad que antes se hacía de modo no riguroso, y que con el paso del tiempo han descubierto que genera un valor agregado para ellas al entrañar las claves del crecimiento (Calvo, 2008).

Para realizar un trabajo de investigación con estricto apego al método científico lo primero que debe de quedar claro es el concepto de ciencia. Gortari afirma que la ciencia “es la explicación objetiva y racional del universo” (1974, p. 14). Por tanto, para poder brindar a la sociedad conocimiento científico es fundamental conocer y seguir el método que estipula la metodología adecuada para realizar cada tipo de estudio.

Según Caballero (2014), metodología es la “ciencia cuya especialidad o campo de estudio son las orientaciones racionales que requerimos para resolver problemas nuevos, y para adquirir o descubrir nuevos conocimientos a partir de los provisoriamente establecidos y sistematizados por la humanidad” (p. 78).

El compromiso asumido al investigar un fenómeno, implica un estricto apego al método científico que debe fungir como un vector hacia el cual se dirijan todos los esfuerzos. Pues “lo que garantiza la objetividad de la ciencia es la corroboración de su discurso con el contenido, con la realidad que explica, esto es, con el universo” (Camero, 2010, p. 158).

El seguimiento puntual y preciso de la metodología establecida en la investigación asegura que los resultados obtenidos tengan un alto índice de confiabilidad. Resulta infértil hacer aportes en un campo del conocimiento sobre un tema que suscite interés si no se cuenta con una sólida metodología que dé sustento a los resultados.

Así, el objetivo de toda investigación debe ser el contribuir a la ciencia concebida como “un sistema de ideas establecidas provisionalmente (conocimiento científico), y como actividad productora de nuevas ideas (investigación científica)” (Bunge, 1991, p. 9). Dicho conocimiento se debe generar mediante procedimientos transparentes y claramente definidos.

Obviamente, los profesionales de recursos humanos deben tener las bases para hacer una investigación documental que, según Hernández Sampieri (2014), consiste en un análisis de la información escrita sobre un determinado tema, con el propósito de establecer relaciones, diferencias, etapas, posturas o estado actual del conocimiento respecto al tema objeto de estudio. La investigación documental debe ser muy sólida ya que constituye el andamiaje donde se construirá el conocimiento.

En el campo de la investigación científica la incursión de la tecnología ha cobrado un auge importante en las últimas décadas y ha favorecido

que el conocimiento generado sea más preciso, veraz y confiable para ayudar a comprender, explicar y predecir fenómenos concretos de la realidad circundante. El avance tecnológico ha permitido la incorporación en los estudios sociales de programas estadísticos como el Statistical Package for the Social Sciences (SPSS) para el procesamiento y análisis de información (Mejía, 2018).

En esta línea, el presente trabajo también busca analizar cuál es el impacto de las técnicas de análisis multivariable en las investigaciones científicas llevadas a cabo en el área de recursos humanos de las empresas. Pues, generalmente, este tipo de trabajos de investigación analizan fenómenos complejos y que tienen numerosos factores que los impactan directa o indirectamente. Por tanto, se deben elegir las herramientas más pertinentes según la naturaleza el estudio.

Las técnicas estadísticas multivariantes persiguen el descubrimiento automático del conocimiento contenido en la información almacenada de modo ordenado en grandes bases de datos. Estas técnicas tienen como objetivo descubrir patrones, perfiles y tendencias a través del análisis de los datos utilizando técnicas estadísticas avanzadas de análisis multivariante de datos (Pérez, 2009, p.1).

Las investigaciones científicas en el campo de la dirección de recursos humanos en las empresas iniciaron en los años setenta, con una serie de estudios que replantearon las formas tradicionales de administrar a las personas en las organizaciones, algunos de los pioneros en esta línea fueron (Drucker, 1968; Foulkes, 1975; Burack y Smith, 1977; Watson, 1977), etc.

La difícil tarea de estar al frente de otras personas en el ámbito organizacional, ha desafiado a los profesionales de recursos humanos a realizar investigaciones mediante técnicas cada vez más sofisticadas que les permitan tener información relevante para poder tomar decisiones más acertadas.

Pero no sólo el método científico se considera suficiente a la hora de conocer la realidad del factor humano en el ámbito organizacional, pues si realmente se quiere obtener información cualitativa y cuantitativa para mejorar la toma de decisiones en la conducción del personal, la aplicación de conocimientos y teorías de las diferentes ciencias sociales es necesaria (Calvo, 2008). Ni qué decir de las técnicas de análisis multivariable que representan una gran ayuda a la hora de generar conocimiento científico.

Por poner un ejemplo concreto, “El Análisis de Varianza (ANOVA) analiza la posible influencia de una o varias variables independientes



(o Factores) sobre una variable dependiente. Las primeras son variables cualitativas, es decir, variables discretas o categoriales mientras que la variable dependiente debe ser cuantitativa” (Calvo, 1993, p. 17). Por esta razón, en los trabajos de investigación relacionados con el factor humano se considera pertinente usar este tipo de instrumentos.

Los modelos de ANOVA (ANalysis Of VAriance) son técnicas de Análisis Multivariante de dependencia, que se utilizan para analizar datos procedentes de diseños con una o más variables independientes cualitativas (medidas en escalas nominales u ordinales) y una variable dependiente cuantitativa (medida con una escala de intervalo o de razón). En este contexto, las variables independientes se suelen denominar factores (y sus diferentes estados posibles o valores son niveles o tratamientos) y la variable dependiente se conoce como respuesta (Ordaz Sanz, Melgar Hiraldo, & Rubio Castaño, sin año, p.2)

La técnica ANOVA tiene como objetivo primordial la comparación de las medias de más de dos poblaciones que hayan sido previamente seleccionadas. Esta comparación se hace a través de un contraste de hipótesis donde se analiza la varianza y las diferencias de medias que hay entre las muestras, y, también se analiza el tamaño de muestra. Es importante considerar que técnicas de comparación siempre analizan los siguientes elementos: dispersión, diferencias de medias y tamaño muestral.

Para lograr una mayor comprensión de esta técnica es necesario familiarizarse con el lenguaje entorno al ANOVA. Por ejemplo, factor es una variable cualitativa que contempla una serie de poblaciones a comparar. Es importante señalar que puede haber factor fijo o factor aleatorio. Generalmente esta técnica trabaja con pocos factores, sin embargo, hay ocasiones en que los factores pueden estar cruzados, es decir, cuando todos los niveles de un factor se combinan con todos los niveles del otro factor.

Esta técnica comparativa (ANOVA), busca contrastar la Hipótesis nula de igualdad de niveles de un factor con la Hipótesis alternativa de no igualdad de esos niveles. Esto se hace para cada factor contemplado en el ANOVA. La interacción entre factores indica que la variable estudiada, en este caso, la variable dependiente, se comporta, ante niveles de un factor, dependiendo de cuáles sean los niveles del otro factor.

La complejidad de los fenómenos sociales que se pueden estudiar en las empresas tiene muchas aristas y variables debido a todas las implicaciones que tiene el acercarse a investigar la condición humana. Por esta razón, los profesionales de recursos humanos dedicados al estudio cien-

tífico del personal necesariamente deben incorporar técnicas de análisis multivariable, por ejemplo, ANOVA, si quieren que sus aportaciones estén mejor fundamentadas al contar con un soporte basado en técnicas sofisticadas de procesamiento de información como el SPSS.

## Discusión

En un mundo caracterizado por el constante cambio donde rápidamente las tecnologías y el conocimiento mismo tiende a la obsolescencia, resulta indispensable para los profesionales que lideran las organizaciones estar al día e investigar, muchas veces de forma autodidacta, las nuevas tendencias y saberes que están boga y que coadyuvan en la praxis humana.

La investigación que deben emprender los profesionales del área de recursos humanos para apoyar en las decisiones a la alta dirección a través del suministro de datos confiables, los obliga a incorporar nuevas técnicas para recabar información pues

...en la medida en que los avances tecnológicos se desarrollan velozmente, los administradores toman conciencia de la alta responsabilidad de desarrollar la administración de recursos humanos, tomando en cuenta los nuevos enfoques que se presentan y que necesariamente, la aplicación del proceso administrativo se torna más compleja, para lo cual cada una de sus fases requiere una especialización, ya que va adquiriendo, cada vez más, una característica propia, delimitada, única, como es la coordinación de las diferentes fases y su aplicación en las diversas áreas funcionales que componen la empresa, es importante establecer los fundamentos epistemológicos y las diversas metodologías que deberán aplicarse para, por un lado, delimitar bien el campo de la administración de recursos humanos y por el otro, continuar desarrollándola con miras a estructurarla cada vez más como un campo independiente de las demás ciencias sociales (Jensen Castañeda, 2007, p. 20).

En esta línea, los responsables de las áreas de recursos humanos deben encarar un gran desafío al no conformarse con generar conocimiento a través de la investigación científica, sino también a hacer contribuciones para que la administración de recursos humanos tenga su propio campo y sea reconocida con el carácter de ciencia.

## Conclusiones

Ante la necesidad de saber y conocer la realidad que los circunda, especialmente los directivos del área más importante y compleja, a saber, de recursos humanos, deben asumir el desafío de realizar investigaciones apoyadas en un método enfocado a la generación de conocimiento científico que incorpore técnicas de análisis multivariable, y que considere el uso de la tecnología como el programa SPSS a fin de que sus aportaciones sean más consistentes y confiables.

Actualmente, el área de recursos humanos en buena parte de las grandes empresas ha evolucionado y ha logrado encumbrarse en las altas esferas organizacionales gracias al valor que agrega a la alta dirección. Es decir, gracias al arduo trabajo de muchos directivos de recursos humanos el departamento como tal, ha pasado a ser de una simple instancia administrativa y un aparato burocrático, a un socio estratégico que juega un papel fundamental en la empresa porque se encarga de la gestión humana que representa, finalmente, el factor más importante y delicado pues de él depende el éxito o el fracaso de cualquier institución y/o empresa.

Por tanto, dada la enorme responsabilidad que con mucho esfuerzo han logrado asumir los directivos de recursos humanos, es necesario que estén dotados con más y mejores herramientas de búsqueda y procesamiento de información, a fin de que cuenten con los elementos confiables suficientes a la hora de asesorar a la alta dirección en cuanto a la toma de decisiones de gran trascendencia para la vida organizacional y para las personas que la conforman.

Qué mejor manera de contribuir a la sociedad desde el rol estratégico de recursos humanos en una empresa que realizando investigaciones profesionales basadas en el método científico para dar luz a los complejos fenómenos sociales que se dan al interior de las organizaciones. Asumir este desafío no es tarea fácil y muy pocos directivos de recursos humanos cuentan con la formación profesional para generar conocimiento científico.

Sin embargo, los desafíos que trae un mundo globalizado pueden ser un detonante para que los profesionales de recursos humanos salgan de su zona de confort y enfoquen también su liderazgo para contribuir a la ciencia pues, como dice Tasselli, es posible que las personas cambien positivamente si se les invita a colaborar en proyectos que las hagan sentir importantes, que disfruten y que estén alineados con sus valores (2018). Así pues, los directivos de recursos humanos deben incursionar en la in-

investigación científica respaldados por sus empresas si quieren continuar jugando un rol estratégico que contribuya al desarrollo humano, social y organizacional. Pues las organizaciones, como campos de desarrollo de las competencias de sus miembros, pueden contribuir y convertirse en una oportunidad para lograr la renovación de las sociedades (Conforti, 2018).

## Referencias

- Álvarez de Mon, S. (1998). *La empresa Humanista y competitiva*. Bilbao: Deusto.
- Caballero Romero, A. (2014). *Metodología integral innovadora para planes y tesis*. México: Cengage Learning.
- Calvo Gómez, F. (1993). *Técnicas estadísticas Multivariantes*. Bilbao: Universidad de Deusto.
- Calvo, R., & Gómez, J. (2008). *La investigación social en recursos humanos*. La necesidad de seguir una metodología. *Athenea Digital*, 14, 181-189. Recuperado el 10 de diciembre de 2018 de <http://psicologia-social.uab.es/athenea/index.php/atheneaDigital/article/view/517>
- Bunge, M. (1991). *La ciencia, su método y su filosofía*. México: Ediciones siglo Veinte-Nueva imagen.
- Burack, E. H.; Smith, R. D. (1977). *Personnel Management*. New York: Wiley.
- Camero Rodríguez, F. (2004). *La investigación científica: Filosofía, teoría y método*. México: Fontamara.
- Carey, D., Barton, D., & Charan, R. (2017). *Las personas antes que la estrategia: un nuevo papel para la dirección de RRHH*. *Harvard Business Review*. Recuperado el 10 de diciembre de 2018, de <https://hbr.es/gesti-n-de-recursos-humanos/429/las-personas-antes-que-la-estrategia-un-nuevo-papel-para-la-direcci>
- Conforti, F. (2018). *Integridad. El aporte de organizaciones y empresas a la construcción de paz*. Madrid: Dykinson.
- Drucker, P. (1968). *The Practice of Management*. London: Pan.
- Foulkes, F. K. (1975). "The Expanding Role of the Personnel Function", *Harvard Business Review*, (marzo-abril), pp. 146-156.
- Gortari, E. (1974). *Iniciación a la lógica*. México: Grijalbo.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la Investigación*. México: Mc Graw Hill.
- Jensen Castañeda, R. (2007). *Fundamentos epistemológicos de la línea de investigación en administración de recursos humanos*. Recuperado el

- 12 de diciembre de 2018, de <https://www.uv.mx/iiesca/files/2013/01/fundamentos2007-2.pdf>
- Mejía, J. (2018). *Análisis estadístico multivalente con SPSS para las ciencias económico administrativas. Teoría y práctica de las técnicas dependientes*. ISBN en trámite.
- Ordaz Sanz, J., Melgar Hiraldo, M., & Rubio Castaño, C. *Introducción a las técnicas de análisis Multivariable*. Métodos estadísticos y econométricos en las empresas y para finanzas. Departamento de Economía, Métodos Cuantitativos e Historia Económica: Universidad Pablo de Olavide. Recuperado el 29 de octubre de 2018, de [https://www.upo.es/export/portal/com/bin/portal/upo/profesores/jaordsan/profesor/1311101268463\\_mxtodos\\_estadisticos\\_y\\_economxtricos\\_en\\_la\\_empresa\\_y\\_para\\_finanzas.pdf](https://www.upo.es/export/portal/com/bin/portal/upo/profesores/jaordsan/profesor/1311101268463_mxtodos_estadisticos_y_economxtricos_en_la_empresa_y_para_finanzas.pdf)
- Pérez López, C. (2009). *Técnicas estadísticas multivariantes con SPSS*. Madrid: Garceta.
- Senge, P. (1990). *La quinta disciplina*. Buenos Aires: Editorial Granica.
- Watson, J. (1977). *The Personnel Manager: A study in the Sociology of Work and Employment*. London: Routledge and Kegan Paul.
- Tasselli, S., Kilduff, M., & Landis, B. (2018). Los beneficios de cambiar su forma de ser en el trabajo. Harvard Business Review. Recuperado el 10 de diciembre de 2018, de <https://hbr.es/gesti-n-del-desempe-o/1261/los-beneficios-de-cambiar-su-forma-de-ser-en-el-trabajo>
- Werther, W. (2001). *Administración de personal y recursos humanos*. México: Mc Graw Hill.

### Referencia de sitios WEB

La estadística, una orquesta hecha instrumento. Recuperado el 29 de octubre de 2018, de <https://estadisticaorquestainstrumento.wordpress.com/2012/12/20/tema-15-anova/>

***Ensayos 2018.  
Análisis Multivariante  
con enfoque dependiente en las  
de las Ciencias de la Administración  
como base para la Innovación***

se terminó de imprimir  
noviembre de 2019  
en los Talleres Gráficos  
de Imprelibros BM  
Brillante No. 916  
Col. Alcalde Barranquitas  
Guadalajara, Jal.  
Tel.: 3613-8426

[imprelibrosbm@gmail.com](mailto:imprelibrosbm@gmail.com)

