

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS
DEPARTAMENTO DE MATEMÁTICAS**

PROYECTO DE TITULACIÓN

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

**“MAGÍSTER EN ESTADÍSTICA CON MENCIÓN EN GESTIÓN DE
LA CALIDAD Y PRODUCTIVIDAD”**

TEMA:

**DESARROLLO DE UN MODELO ECONOMÉTRICO DE REGRESIÓN
MÚLTIPLE PARA PREDECIR LAS VENTAS DE UNA EMPRESA
COMERCIALIZADORA DE PRODUCTOS DE LIMPIEZA**

AUTOR:

JOSÉ DANNY BARONA GUADALUPE

Guayaquil - Ecuador

2019

RESUMEN

El presente trabajo desarrolla una ecuación de regresión múltiple para explicar las ventas anuales que realiza la empresa ABC comercializadora de productos de limpieza a sus clientes usando como variables explicativas las razones financieras de los clientes y dos variables dummy que son el sector económico al que pertenece el cliente y el tamaño del cliente; se ha desarrollado un modelo de corte transversal y se ha determinado que la razón corriente, la razón de deuda, el sector y el tamaño del cliente son significativos para explicar las ventas. Este modelo permite predecir las ventas para todos los clientes durante el año, y con esta información, elaborar la planificación financiera, específicamente el Estado de Resultados y el Flujo de Efectivo presupuestados; en el capítulo 1 se describe el problema y se plantean los objetivos del estudio, el capítulo 2 trata las definiciones y conceptos de la regresión múltiple, las razones financieras y los estados financieros; el capítulo 3 describe la estrategia de análisis y los supuestos a utilizar; el capítulo 4 muestra la estadística descriptiva univariante, bivariante, se construye el modelo estadístico de predicción y se elaboran los estados financieros; finalmente en el capítulo 5 se detallan las conclusiones del estudio y las recomendaciones para estudios posteriores.

Términos clave: planificación financiera, predicción, regresión múltiple, ventas

ABSTRACT

This paper develops a multiple regression equation to explain the annual sales made by the ABC marketing company of cleaning products to its customers using as explanatory variables the financial reasons of the clients and two dummy variables that are the economic sector to which the company belongs, and client size; a cross-sectional model has been developed and it has been determined that the current ratio, the debt ratio, the sector and the size of the client are significant in explaining sales. This model allows to predict the sales for all the clients during the year, and with this information, to elaborate the financial planning, specifically the Income Statement and the Budget Cash Flow; in chapter 1 the problem is described and the objectives of the study are stated, chapter 2 deals with definitions and concepts of multiple regression, financial ratios and financial statements; Chapter 3 describes the analysis strategy and the assumptions to be used; Chapter 4 shows the univariant, bivariate descriptive statistics, the statistical model of prediction is constructed and the financial statements are elaborated; finally in chapter 5 the conclusions of the study and the recommendations for later studies are detailed.

Key terms: financial planning, prediction, multiple regression, sales

DEDICATORIA

A mi familia por su incondicional apoyo y comprensión

AGRADECIMIENTO

Mis sinceros agradecimientos a todo el personal docente y administrativo de la FCNM por su calidad humana, y en especial a Sandra García, mi tutora.

DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación, me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo, corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.



JOSÉ DANNY BARONA GUADALUPE

TRIBUNAL DE GRADUACIÓN



María Nela Pastuzaca Fernández, Ph.D.
PRESIDENTE



Sandra García Bustos, Ph.D.
DIRECTOR



M.Sc. Francisco Moreira Villegas
VOCAL



Holger Cevallos Valdiviezo, Ph.D.
VOCAL

TABLA DE CONTENIDO

RESUMEN	i
ABSTRACT	ii
DEDICATORIA	iii
AGRADECIMIENTO	iv
DECLARACIÓN EXPRESA	v
TRIBUNAL DE GRADUACIÓN.....	vi
TABLA DE CONTENIDO	vii
LISTADO DE FIGURAS.....	ix
LISTADO DE TABLAS	xi
CAPÍTULO 1	1
1. INTRODUCCIÓN.....	1
1.1. Antecedentes	1
1.2. Descripción del problema.....	1
1.3. Objetivos	2
1.4. Hipótesis	3
1.5. Alcance	3
CAPÍTULO 2	4
2. MARCO TEÓRICO	4
2.1. Modelos estadísticos.....	4
2.2. El modelo econométrico.....	4
2.2.1. Definición del modelo estadístico de regresión múltiple	4
2.2.2. Supuestos del modelo de regresión	5
2.2.3. Estimación de los coeficientes de regresión.....	6
2.2.4. Validez del modelo de regresión estimado.....	8
2.2.5. Variables explicativas cualitativas o dummy.....	10
2.3. Las razones financieras	10
2.3.1. Razones de liquidez	10
2.3.2. Razones de endeudamiento.....	11

2.3.3. Razones de rentabilidad.....	11
2.4. Los estados financieros presupuestados.....	11
2.4.1. El estado de resultados presupuestado.....	12
2.4.2. Flujo de efectivo presupuestado.....	12
2.5. Estado del arte.....	13
CAPÍTULO 3	16
3. METODOLOGÍA.....	16
3.1. Estrategia de análisis.....	16
3.2. Datos, variables y estimaciones adicionales.....	18
3.2.1. Datos y variables del modelo econométrico.....	18
3.2.2. Estimaciones adicionales.....	19
CAPÍTULO 4	20
4. RESULTADOS.....	20
4.1. Análisis exploratorio de las variables del modelo econométrico.....	21
4.1.1. Análisis univariante.....	21
4.1.2. Análisis bivariante.....	26
4.2. Obtención del modelo econométrico.....	30
4.3. Estados financieros presupuestados.....	39
CAPÍTULO 5	41
5. CONCLUSIONES Y RECOMENDACIONES.....	41
5.1. Conclusiones.....	41
5.2. Recomendaciones.....	42
6. Referencias.....	6-1

LISTADO DE FIGURAS

Figura 2-1: Estado de Resultados Presupuestado.....	12
Figura 2-2: Ejemplo de Flujo de Efectivo Presupuestado	13
Figura 4-1: Encabezado de los datos.....	21
Figura 4-2: Resumen de estadísticos descriptivos.....	21
Figura 4-3: Medias, desviaciones estándar y varianzas.....	22
Figura 4-4: Porcentaje de empresas por sector	22
Figura 4-5: Porcentaje de empresas por tamaño	23
Figura 4-6: Histograma de ventas	23
Figura 4-7: Histograma de razón corriente.....	24
Figura 4-8: Histograma de prueba ácida.....	24
Figura 4-9: Histograma de la razón de deuda.....	25
Figura 4-10: Histograma del margen de utilidad	25
Figura 4-11: Histograma de las ventas por sector.....	26
Figura 4-12: Diagrama de cajas de las ventas por sector	27
Figura 4-13: Histograma de las ventas por tamaño de la empresa.....	28
Figura 4-14: Diagrama de cajas por tamaño de empresa	28
Figura 4-15: Ventas vs razón corriente	29
Figura 4-16: Ventas vs prueba ácida	29
Figura 4-17: Ventas vs razón de deuda	30
Figura 4-18: Ventas vs margen de utilidad.....	30
Figura 4-19: modelo lineal con todas las variables	31
Figura 4-20: Matriz de correlaciones de las razones financieras	32
Figura 4-21: Regresión por pasos en R	33
Figura 4-22: Selección de variables con random forest	34
Figura 4-23: Importancia de las variables	34
Figura 4-24: Modelo con las 4 variables explicativas	35
Figura 4-25: Factor de inflación de la varianza	36
Figura 4-26: Histograma de los residuos	36
Figura 4-27: Gráfico QQ-normal.....	37

Figura 4-28: Residuos vs valores ajustados	37
Figura 4-29: Función de autocorrelación.....	38
Figura 4-30: Estado de Resultados Presupuestado.....	40
Figura 4-31: Flujo de Efectivo Presupuestado	40

LISTADO DE TABLAS

Tabla 4-1: Ecuaciones de regresión.....	39
---	----

CAPÍTULO 1

1. INTRODUCCIÓN

1.1. Antecedentes

Las empresas comerciales, industriales o de servicios tienen que enfrentar problemas de distinta naturaleza en el corto, mediano y largo plazo, tales como problemas financieros, legales, de mercado, de producción, logísticos, con los clientes, con los proveedores, etc.; sin embargo, los problemas financieros suelen ser los que más preocupan a los gerentes y accionistas de la empresa.

En las pequeñas y medianas empresas (PYMES), uno de los problemas financieros más frecuentes son los desequilibrios del efectivo en el corto plazo, es decir, el exceso o déficit de efectivo en gran medida, lo que en ocasiones significa no poder cubrir sus deudas a corto plazo, y otras veces no saber qué hacer con el excedente de efectivo. Los gerentes o propietarios de estas empresas suelen tomar decisiones reactivas coyunturales, que solucionan momentáneamente el problema (por ejemplo, un sobregiro bancario) pero no atacan la raíz del problema.

Una de las causas principales de este problema es que en muchas PYMES no se realiza un presupuesto del flujo de efectivo anual, lo que impide que se realice una planificación financiera adecuada, lo que conlleva a los problemas anteriormente mencionados; sin embargo, para realizar un presupuesto de efectivo, es necesario realizar una proyección de las ventas mensuales o anuales de la empresa, que generalmente es un proceso no tan sencillo de llevar a cabo.

Existen distintos métodos de proyección de las ventas: subjetivos, contables, estadísticos y de series de tiempo. Uno de los más utilizados es el *modelo estadístico de regresión lineal múltiple*,

1.2. Descripción del problema

La empresa ABC se dedica a la comercialización de productos de limpieza de espacios físicos, y sus clientes son empresas pequeñas, medianas y grandes del sector industrial, comercial y de servicios. Sus ventas fundamentalmente se realizan a crédito y no tiene períodos estacionales de ventas durante el año, debido a la naturaleza de los productos que vende.

En los dos últimos años, la empresa ABC, ha tenido problemas en el área financiera tales como: déficit y superávits de efectivo, disminución de la rentabilidad, lo que ha complicado su desempeño en esta área. Esta empresa ha venido realizando sus proyecciones financieras en base a métodos subjetivos, y no se ha considerado un análisis de sensibilidad en base a posibles variaciones en los valores estimados de las ventas proyectadas.

Estos problemas han llevado a la empresa a solicitar créditos de corto plazo en ciertos meses para cubrir el déficit que se presenta, lo que implica un costo financiero que disminuye la rentabilidad neta.

El propósito de este trabajo es desarrollar y utilizar un modelo estadístico de REGRESIÓN LINEAL MÚLTIPLE para pronosticar las ventas anuales de la empresa ABC en función de variables explicativas cuantitativas y cualitativas DE SUS CLIENTES. Las variables cuantitativas serán las razones financieras obtenidas de los estados financieros de los clientes de la empresa ABC tales como la liquidez, rentabilidad y el nivel de endeudamiento; las variables cualitativas serán variables DUMMY tales como el sector económico de la actividad a la que pertenece el cliente (industrial, comercial o de servicios), y el tamaño del cliente (pequeño, mediano o grande).

Aplicando el modelo a todos los clientes de la empresa ABC se obtendrá una estimación para el total de ventas anual, y con este resultado elaborar el Presupuesto del Flujo de Efectivo y el Estado de Resultados

1.3. Objetivos

El **objetivo general** del presente trabajo es:

- Formular una estrategia basada en un modelo estadístico de Regresión Lineal Múltiple para la proyección de las ventas aplicable al área financiera de la empresa ABC, para facilitar la toma de decisiones de los ejecutivos.

Los **objetivos específicos** son:

- Realizar un análisis estadístico descriptivo de las variables financieras y cualitativas consideradas en el modelo a desarrollar, para conocer la situación actual de la empresa y sus clientes.

- Crear un modelo estadístico de regresión lineal múltiple para pronosticar las ventas en función de variables financieras y factores cualitativos de los clientes

- Elaborar el Presupuesto del Flujo de Efectivo y el Estado de Resultados con los resultados del pronóstico de ventas entregado por el modelo econométrico

1.4. Hipótesis

Las hipótesis a considerar son las siguientes:

- Las razones financieras de los clientes de la empresa ABC explican significativamente las ventas de la compañía.
- Los factores que representan el sector económico al que pertenece el cliente y el tamaño del cliente tienen un efecto significativo en las ventas.
- Las razones financieras, el sector y el tamaño del cliente son, en conjunto, significativos para explicar las ventas de la empresa ABC.

1.5. Alcance

El presente estudio es aplicable a la empresa ABC ubicada en la ciudad de Guayaquil, y temporalmente es aplicable al año 2018

CAPÍTULO 2

2. MARCO TEÓRICO

2.1. Modelos estadísticos

Un *modelo estadístico* es un conjunto de hipótesis que se establecen sobre el comportamiento de una o más *variables* que pueden estar o no relacionadas estadísticamente. Cuando se analiza una sola variable, los modelos estadísticos suelen referirse a la distribución probabilística que caracteriza el comportamiento de dicha variable; cuando se estudian dos o más variables, algunos modelos pretenden describir la relación que existe entre una variable denominada respuesta o explicada como una función de una o más variables llamadas explicativas o predictores por medio de una ecuación.

Entre los modelos estadísticos explicativos multivariados más utilizados se encuentran los modelos *de regresión lineal* y los modelos *de series de tiempo*. Los modelos de regresión lineal relacionan una variable explicada cuantitativa con una o más variables explicativas distintas a la explicada (cuantitativas o cualitativas) en forma de una ecuación lineal; los modelos de series de tiempo relacionan una variable explicada como una función de los valores pasados de la misma variable explicada.

En el presente trabajo se desarrollará un modelo estadístico de regresión múltiple.

2.2. El modelo estadístico de regresión múltiple

2.2.1. Definición del modelo estadístico de regresión múltiple

En economía y finanzas, en muchas ocasiones, se tiene la necesidad de predecir el valor de alguna variable financiera tal como las ventas de un producto o el PIB de un país en función de otras variables, que se supone guardan relación con la variable a predecir, y la manera de hacerlo, es mediante una ecuación que relacione a la variable respuesta (explicada o dependiente), con las variables que hacen el papel de predictoras (explicativas o independientes). El modelo estadístico de regresión lineal es el siguiente (Fernández, Econometría, 2005):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i ; i = 1, 2, \dots, n \quad (2.1)$$

en donde:

Y_i : observación i – ésima de la variable dependiente

X_{ki} : observación i – ésima de la variable X_k

$\beta_0, \beta_1, \dots, \beta_p$: coeficientes de la ecuación de regresión

ε_i : error aleatorio de la observación i -ésima

Este modelo relaciona a la variable dependiente Y con las variables independientes X_1, \dots, X_p y el término ε se conoce como el término de error aleatorio; la linealidad se refiere a la relación lineal expresada entre la variable explicada y los coeficientes beta de la regresión. Esta ecuación de regresión también se puede escribir de forma matricial de la siguiente manera (Court & Rengifo, 2011):

$$Y = X\beta + \varepsilon \quad (2.2)$$

donde Y , X , β y ε son la matriz y vectores siguientes:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & X_{11} & \dots & X_{p1} \\ 1 & X_{12} & \dots & X_{p2} \\ \vdots & \vdots & \dots & \vdots \\ 1 & X_{1n} & \dots & X_{pn} \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}; \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Para que el modelo dado por la ecuación (2-1) o (2-2) sea considerado un modelo válido de regresión lineal se deben cumplir algunos supuestos, los cuales se describen en la siguiente sección.

2.2.2. Supuestos del modelo de regresión

Los supuestos que debe cumplir una ecuación de regresión lineal son (Court & Rengifo, 2011), (Pérez, 2002):

1) Especificación: El modelo supone que la relación entre la variable dependiente Y y los coeficientes de regresión betas es lineal; además se supone que el modelo está correctamente especificado, es decir, que no se incluyen variables irrelevantes, y se han incluido las variables relevantes.

2) Incorrelación de las variables explicativas: Se supone que las variables explicativas no están correlacionadas entre sí.

3) Valor esperado del término de error es cero, $E(\varepsilon_i) = 0$

4) Homocedasticidad: La varianza del término de error es constante para todas las observaciones, $\text{Var}(\varepsilon_i) = \sigma^2$

5) No autocorrelación: La correlación (o covarianza) entre los errores de observaciones diferentes es cero, $\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$

6) Normalidad: Aunque no es un supuesto obligatorio de un modelo de regresión, usualmente se asume que el término de error sigue una distribución normal multivariada.

En forma matricial, los supuestos 2 al 6 se expresan como:

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

en donde $\mathbf{0}$ es un vector de n ceros, e \mathbf{I}_n es la matriz identidad de orden n .

2.2.3. Estimación de los coeficientes de regresión

Las ecuaciones 2.1 o 2.2 se refieren a la ecuación de regresión poblacional, que se supone existe; sin embargo, debido a que sólo se dispone de una muestra de los datos de la población, la ecuación de regresión poblacional se estima mediante la *ecuación de regresión muestral*, cuya expresión es la siguiente (Novales, 1993):

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_p X_{pi} ; \quad i = 1, 2, \dots, n \quad (2.3)$$

O, en forma matricial:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \quad (2.4)$$

En la ecuación (2.4) $\hat{\boldsymbol{\beta}}$ representa el vector de coeficientes estimados de regresión a partir de los datos muestrales, y $\hat{\mathbf{Y}}$ es la estimación obtenida para la variable explicada \mathbf{Y} a partir de los datos muestrales y los coeficientes de regresión estimados. Básicamente existen tres métodos para estimar los coeficientes beta de la ecuación de regresión: el método de mínimos cuadrados, el método de máxima verosimilitud y el método generalizado de momentos (Hansen, 2018).

El método de mínimos cuadrados minimiza la suma de los cuadrados de los errores de estimación, es decir, los coeficientes $\hat{\boldsymbol{\beta}}$ se obtienen a partir de la expresión (Fernández, Econometría, 2005):

$$\min_{\boldsymbol{\beta}}(\mathbf{u}'\mathbf{u}) \quad \text{donde} \quad \mathbf{u} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$$

u es el error entre el valor observado de Y y la estimación $X\hat{\beta}$. A partir de esta expresión se obtiene que los coeficientes de regresión vienen dados por la fórmula (2.5) (Hansen, 2018):

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (2.5)$$

Una medida de bondad de ajuste global del modelo es el coeficiente de determinación R^2 que se define como (Gujarati & Porter, 2010):

$$R^2 = \frac{(\hat{Y} - \bar{Y})'(\hat{Y} - \bar{Y})}{(Y - \bar{Y})'(Y - \bar{Y})} \quad (2.6)$$

En la expresión (2.6) el numerador representa la variabilidad de Y explicada por la ecuación de regresión estimada, y el denominador es la variabilidad total de la variable Y , de manera que, el coeficiente de determinación es la proporción de la variabilidad de la variable Y explicada por la ecuación de regresión estimada. R^2 es un valor que está entre 0 y 1, y mientras más se acerque a 1, el modelo estimado explica mejor a la variable dependiente Y .

El coeficiente de determinación aumenta al aumentar el número de variables independientes del modelo, por lo que se define el coeficiente de determinación corregido que ajusta el valor por la inclusión de más variables dependientes al modelo, su expresión es (Fernández, Econometría, 2005):

$$\bar{R}^2 = 1 - \frac{(n-1)}{(n-p-1)} (1 - R^2) \quad (2.7)$$

donde n es el tamaño de la muestra y p es el número de variables explicativas.

Una de las pruebas más importantes es determinar *la significación de los coeficientes de regresión individuales* β_k cuando el término de error sigue una distribución normal. En particular, debe probarse la hipótesis nula (Hansen, 2018):

$$H_0: \beta_k = 0 \quad vs \quad H_1: \beta_k \neq 0; \quad k = 0, 1, 2, \dots, p$$

en donde p es el número de variables explicativas del modelo. La teoría econométrica ha demostrado que, bajo la hipótesis nula, el estadístico:

$$t = \frac{\hat{\beta}_k}{\hat{S}_{\beta_k}} \quad (2.8)$$

donde \hat{S}_{β_k} es el error estándar de $\hat{\beta}_k$, sigue una distribución t de Student con $(n - p - 1)$ grados de libertad, a un nivel de significancia α , se rechaza H_0 si $|t| > t_{\alpha/2, n-p-1}$.

$p-1$, de lo contrario no se rechaza H_0 . Utilizando el criterio del p-value, se rechaza H_0 si p-value es menor que α , de lo contrario no se rechaza H_0 . Software estadístico como R, SPSS o Minitab devuelven los valores del estadístico t y el p-value cuando se utiliza la opción de regresión lineal.

2.2.4. Validez del modelo de regresión estimado

Para que un modelo de regresión sea válido debe examinarse si existen problemas de naturaleza muestral o de incumplimiento de los supuestos del modelo. A continuación, se describen cuáles son estos problemas y sus posibles soluciones:

Multicolinealidad: Este es un problema de naturaleza muestral y se presenta cuando existe correlación fuerte entre las variables explicativas del modelo (Maddala, 1977) y puede detectarse mediante la matriz de correlación de las variables explicativas, el Factor de Inflación de la Varianza (VIF) o el Número de Condición. Si en la matriz de correlación de las variables explicativas del modelo aparecen valores de correlación altos (cerca de 1) es un indicativo de un posible problema de multicolinealidad.

El Número de Condición relaciona el mayor valor propio con el menor valor propio de la matriz de correlación de las variables explicativas (Gujarati & Porter, 2010), si su valor se encuentra entre 100 y 1000 existiría una multicolinealidad suave, pero si el valor es mayor que 1000 la multicolinealidad es un problema grave; el VIF se calcula para cada variable explicativa y si su valor es mayor que 10, se considera que esa variable está muy correlacionada con las demás.

Las soluciones al problema de multicolinealidad son muy diversas y dependen de cada caso en particular, y éstas podrían ser (Fernández, Econometría, 2005) (Gujarati & Porter, 2010): transformación de variables, eliminación de variables, obtención de una muestra diferente, regresión sobre componentes principales, entre otras.

No normalidad de los residuos: Puesto que las inferencias relacionadas con los coeficientes de regresión están relacionadas con la normalidad el término de error del modelo, debe realizarse un examen de los residuos ($Y - \hat{Y}$) y probar la normalidad de éstos. Puede utilizarse un gráfico QQ-Normal y las pruebas formales de normalidad de Kolmogorov-Smirnov, Shapiro-Wilk, Jarque-Bera o Anderson-FCNM

Darling, la prueba de Shapiro-Wilk se usa cuando el tamaño de la muestra es grande (mayor que 30) y compara dos estimaciones diferentes de la varianza mediante una división, una estimación no paramétrica con la estimación de la varianza muestral.

Si los residuos no son normales, puede intentarse realizar alguna transformación de la variable respuesta y realizar la regresión con dicha variable transformada, para posteriormente probar la normalidad de los residuos.

Autocorrelación de los residuos: Uno de los supuestos del modelo de regresión es que los errores no están correlacionados. Una manera de detectar problemas de autocorrelación es mediante el gráfico de la función de autocorrelación de los residuos, y mediante las pruebas de Durbin-Watson (Maddala, 1977) o de Breusch-Godfrey (Fernández, Econometría, 2005). El test de Breusch-Godfrey básicamente realiza una regresión auxiliar de los residuos sobre las variables explicativas y los retardos de los errores hasta un orden dado, y luego prueba la significación de los coeficientes de los retardos de los errores

Posibles soluciones al problema de autocorrelación es realizar la regresión en diferencias generalizadas o utilizar modelos de series de tiempo autorregresivo condicional heterocedástico (ARCH) o autorregresivo condicional heterocedástico generalizado (GARCH) (Gujarati & Porter, 2010).

Heterocedasticidad: La heterocedasticidad se presenta cuando la varianza del término de error no es constante que es uno de los supuestos del modelo de regresión. De forma gráfica la heterocedasticidad puede detectarse realizando el gráfico de residuos vs valores ajustados y determinar si existe algún patrón no aleatorio, o mediante pruebas formales como la de Breusch-Pagan o el test de Bartlett-White. Se puede contrarrestar la heterocedasticidad mediante (Novales, 1993): la transformación de Box-Cox o mediante un modelo ARCH para modelar el término de la varianza del error. El test de Breusch-Pagan realiza una regresión del cuadrado de los residuales sobre las variables explicativas y prueba la significación de los coeficientes de las variables explicativas para determinar si la varianza depende de las variables predictoras.

Una vez que se han detectado y solucionado los posibles problemas que se presenten en el modelo de regresión estimado inicialmente, entonces el modelo

será válido para predicción, interpretación de los coeficientes, y obtener conclusiones para el modelo que se está estudiando.

2.2.5. Variables explicativas cualitativas o dummy

En el modelo de regresión múltiple se suponen que las variables explicativas son cuantitativas; sin embargo, en ocasiones, es posible incluir factores cualitativos en el análisis mediante la utilización de variables numéricas ficticias o dummy. Esto es posible debido a que las categorías del factor se codifican utilizando ceros y unos, en donde un cero significa la ausencia de una categoría y el uno indica la presencia de esa categoría. Por ejemplo, a un factor con 3 categorías se les asignaría los códigos 00, 01, y 10 para cada categoría, para un factor con 4 categorías la codificación sería 000, 001, 010 y 100 para las categorías.

Cuando se utilizan variables dummy y ésta resulta ser significativa, en realidad se obtiene un conjunto de ecuaciones de regresión una para cada categoría de la variable cualitativa, utilizando los códigos para los niveles del factor. La mayoría del software estadístico devuelve la significación de las categorías del factor y los coeficientes de regresión respectivos.

2.3. Las razones financieras

Las razones financieras son valores numéricos adimensionales resultantes de una división entre dos rubros extraídos del Estado de Resultados o el Balance General de una empresa. Estas razones reflejan la posición financiera y económica de la entidad en aspectos tales como: la liquidez, la rentabilidad, el nivel de deudas, su posición dentro del mercado o la velocidad de sus operaciones comerciales.

Estos indicadores se interpretan ya sea como índices o porcentajes, dependiendo del caso, y se utilizan fundamentalmente para tener una idea preliminar de posibles problemas de naturaleza financiera; así como también, realizar comparaciones con otras empresas similares o con el promedio de la industria en el mercado.

A continuación, se detallan las razones que serán utilizadas en este trabajo.

2.3.1. Razones de liquidez

La liquidez de una empresa es una medida de su capacidad para cumplir con sus deudas a corto plazo. La razón corriente se calcula mediante la siguiente fórmula (Gitman & Zutter, 2012):

$$\text{Razón corriente} = \frac{\text{Activo corriente}}{\text{Pasivo corriente}} \quad (2.9)$$

Esta razón representa cuántos dólares de activos corrientes están disponibles para cubrir cada dólar de deuda a corto plazo. La prueba ácida se calcula mediante la fórmula (Ross, Westerfield, & Jordan, 2010):

$$\text{Prueba ácida} = \frac{\text{Activo corriente} - \text{Inventario}}{\text{Pasivo corriente}} \quad (2.10)$$

Este indicador es una medida de la capacidad de la empresa para hacer frente a sus obligaciones a corto plazo sin considerar los inventarios.

2.3.2. Razones de endeudamiento

Las razones de endeudamiento miden el nivel de deudas de la empresa, tanto en el corto como en el largo plazo. La razón de deuda total tiene la siguiente expresión (Gitman & Zutter, 2012):

$$\text{Razón de deuda total} = \frac{\text{Pasivo total}}{\text{Activo total}} \quad (2.11)$$

Esta razón es muy importante ya que refleja el nivel total de deudas, tanto a corto como a largo plazo, como un porcentaje del total de los activos de la empresa.

2.3.3. Razones de rentabilidad

La utilidad es la diferencia entre los ingresos y los costos y gastos relacionados con la operación y financiamiento de la empresa. La rentabilidad es la utilidad expresada como un porcentaje con respecto a las ventas, activos, patrimonio o algún otro rubro de referencia.

La rentabilidad neta o margen de utilidad neta tiene la siguiente fórmula de cálculo (Ross, Westerfield, & Jordan, 2010):

$$\text{Margen de utilidad neta} = \frac{\text{Utilidad neta}}{\text{Ventas}} \quad (2.12)$$

Esta razón refleja las utilidades como un porcentaje de las ventas y es un indicador de la capacidad de la empresa para generar ganancias, si su valor fuese negativo significa que la empresa ha tenido pérdidas.

2.4. Los estados financieros presupuestados

Las empresas deben realizar su planificación financiera para establecer sus resultados probables, sus necesidades de efectivo, sus desembolsos por gastos

operacionales y financieros, así como sus necesidades de inversión y financiamiento. Para esto deben elaborarse los estados financieros presupuestados, entre los cuales están: el Estado de Resultados y el Presupuesto de Flujo de Efectivo.

2.4.1. El estado de resultados presupuestado

Dentro de la planificación financiera el Estado de Resultados presupuestado muestra los ingresos, costos y gastos de la empresa para determinar la utilidad o pérdida proyectada. Uno de los métodos que con mayor frecuencia se utiliza es realizar una estimación de las ventas para el año siguiente, y a partir de este dato, construir el Estado de Resultados con información adicional proporcionada por los distintos departamentos de la empresa. Un ejemplo simplificado del Estado de Resultados se muestra en la figura 2.1:

Figura 2-1: Estado de Resultados Presupuestado

LA UNIÓN S.A.	
Estado de Resultados Presupuestado	
Del 1 de enero al 31 de diciembre de 20XX	
(en dólares)	
Ventas	2.000,00
(-) Costo de Ventas	-500,00
(=) Utilidad Bruta en Ventas	2.500,00
(-) Gastos de Administración	-400,00
(-) Gastos de Ventas	-200,00
(-) Gastos Financieros	-100,00
(=) Utilidad antes de Impuestos	1.800,00
(-) Impuestos (36,25%)	-652,50
(=) Utilidad Neta	1.147,50

Elaborado por: José Barona G.

2.4.2. Flujo de efectivo presupuestado

El Presupuesto de Flujo de Efectivo muestra los ingresos y salidas de efectivo para cada período de planificación (mensual, trimestral) y permite determinar el excedente o déficit de efectivo por período. Dentro de la planificación financiera, el flujo de efectivo es muy importante ya que permitirá prever los posibles préstamos a corto plazo en el caso de déficit, o las probables inversiones a corto

plazo en el caso de excedentes. Un ejemplo de flujo de efectivo presupuestado se muestra en la figura 2.2:

Figura 2-2: Ejemplo de Flujo de Efectivo Presupuestado

LA UNIÓN S.A.				
Flujo de Efectivo Presupuestado				
Trimestres del año 20XX				
(en dólares)				
	Trim. 1	Trim. 2	Trim. 3	Trim. 4
Saldo inicial de efectivo	3.000,00	800,00	1.300,00	300,00
(+) Entradas de efectivo	7.000,00	5.000,00	6.000,00	5.300,00
(=) Efectivo disponible	10.000,00	5.800,00	7.300,00	5.600,00
(-) Salidas de efectivo	-9.200,00	-4.500,00	-7.000,00	-4.800,00
(=) Saldo final de efectivo	800,00	1.300,00	300,00	800,00
(-) Saldo mínimo de efectivo	500,00	500,00	500,00	500,00
(=) Excedente / Déficit de efectivo	300,00	800,00	-200,00	300,00

Elaborado por: José Barona G.

2.5. Estado del arte

Los modelos estadísticos de regresión múltiple deben cumplir algunos supuestos para determinar su validez, la no multicolinealidad, normalidad de los residuos, no autocorrelación y homocedasticidad. Estos supuestos se analizan detalladamente en textos clásicos como (Gujarati & Porter, 2010) y (Maddala, 1977), enfoques más modernos se encuentran en (Fernández, Econometría, 2005).y (Wooldridge, 2010). Un tratamiento de la econometría financiera se encuentra en (Court & Rengifo, 2011)

En lo relativo a estudios recientes, (Khalid & Khan, 2017) realizan una investigación utilizando un modelo de regresión lineal múltiple para analizar el efecto sobre las ventas de una empresa productora de fertilizantes (variable dependiente) que tienen los gastos de ventas y financieros (variables explicativas). Utilizan datos temporales del 2002 al 2015 y encuentran que las variables explicativas en conjunto son significativas para explicar las ventas mediante la prueba F, de la misma forma encuentran que los coeficientes individuales son significativos mediante el test de Wald.

Realizan las pruebas de Jarque-Bera para la normalidad, el test de Breusch-Godfrey para la homocedasticidad y el test ARCH para la autocorrelación y encuentran que los supuestos de la regresión se cumplen adecuadamente. Con los valores encontrados para los coeficientes de regresión estudian el impacto de estas variables calculando las elasticidades de los gastos operativos y financieros con respecto a las ventas, y determinan que los gastos operativos son elásticos ($e=1,14$) y los financieros son inelásticos ($e=0,08$), lo que significa que las ventas son más sensibles a la variación de los gastos operativos que a los financieros.

En (Wang, Li, Yin, & He, 2018), se utiliza un modelo de regresión lineal con múltiples factores acoplados para proyectar las ventas de electricidad y con esto, garantizar la seguridad del sistema de energía. Utilizan como variables explicativas: el promedio del crecimiento económico mensual de la escala industrial y el promedio de temperatura mensual de los años 2010 a 2016. Los autores encuentran que ambas variables son significativas y realizan todas las pruebas para los supuestos del modelo de regresión. Como resultado obtienen que el modelo encontrado explica el 99,7% de la variabilidad en las ventas de electricidad en la provincia de Henan (China), y esto les permite la proyección para el año 2017 y la planificación operativa y financiera de la compañía.

En (Agus, Hassan, & Ahmad, 2017) se estudia el efecto que tiene sobre las ventas de las empresas de manufactura de Malasia la administración de la cadena de suministro (SCM). El análisis lo realizan desde tres perspectivas: la relación con los clientes, las tecnologías de información y la información compartida con los proveedores. Para esto realizan un estudio de corte transversal mediante una encuesta en la que se definen 7 variables explicativas que son: estrategia de acompañamiento con proveedores, programas de relación con clientes, información compartida, calidad del intercambio de información, sistemas de apoyo, concepto de aplazamiento, tecnología e innovación.

Luego de elaborar un modelo de regresión lineal múltiple mediante el método de regresión por pasos, se obtiene que las variables que resultan significativas son: programas de relación con clientes, tecnología e innovación e información compartida. Los autores realizan todas las pruebas de validación del modelo y se encuentran con problemas de multicolinealidad que resuelven mediante la regresión por pasos. Los resultados les permite concluir que una implementación

de administración de la cadena de suministro tiene un impacto positivo en las ventas.

Los estudios presentados muestran que el investigador debe escoger adecuadamente las variables explicativas del modelo basándose en el conocimiento del problema a investigar, ya sea para realizar proyecciones de ventas o analizar el impacto de estas variables sobre las ventas, para realizar análisis de elasticidades o implementar políticas administrativas. (Hanke, 2010) indica que, en muchas ocasiones, es mejor un modelo sencillo que explique muy bien a la variable explicada, que utilizar modelos sofisticados y complejos.

CAPÍTULO 3

3. METODOLOGÍA

3.1. Estrategia de análisis

La estrategia a seguir para cumplir con los objetivos planteados es la siguiente:

Se recopilará información de los estados financieros anuales de todos los clientes de la empresa ABC almacenada en las bases de datos de la empresa con relación a las variables consideradas en el modelo estadístico a desarrollar. Se calcularán los indicadores financieros que se consideran en el estudio. A continuación, con las variables cuantitativas y cualitativas se realizará un análisis exploratorio descriptivo univariante y bivariante de las variables incluidas en el modelo, mediante histogramas y diagramas de barras, además de los estadísticos descriptivos

Se utilizará dos métodos de selección de variables y se compararán los resultados, el primer método es regresión por pasos hacia atrás y el segundo método la técnica de random forest.

Para seleccionar variables predictoras importantes para predecir las ventas, se aplicará el **método de regresión por pasos** hacia atrás que consiste en construir un primer modelo aditivo sin interacción, que incluye a todas las variables y factores, y se identifica cuáles son significativos y cuáles no. Luego se aplica un proceso de construcción de varios modelos, eliminando e incluyendo variables en cada paso, hasta llegar al mejor modelo posible basando esta elección utilizando el Criterio de Información de Akaike (AIC). La regresión por pasos se realiza automáticamente con la función step incluida en el software R que será el programa estadístico que se utilizará en el proyecto. En este trabajo no se ha considerado efectos de interacción, sólo efectos principales debido a que se ha aplicado el principio de parsimonia, que prefiere modelos sencillos con un alto grado de ajuste a modelos más complicados, que es lo que prefieren los gerentes de la empresa ABC.

Adicionalmente a la regresión por pasos se escogerá el mejor conjunto de variables utilizando el algoritmo de bosques aleatorios (**random forest**) que

consiste en construir modelos de regresión escogiendo aleatoriamente las variables a incluir en el modelo y utilizando validación cruzada (cross validation) escoger el mejor conjunto de variables explicativas utilizando como criterio de ajuste la menor raíz cuadrática media del error (RMSE). De forma simplificada este algoritmo lo que hace básicamente es construir muchos modelos de regresión con dos variables explicativas elegidas al azar en cada modelo y calcula la predicción para la variable explicada como el promedio de todas las predicciones de los modelos de dos variables que se han creado; luego repite el proceso con tres, cuatro, etc. variables explicativas, finalmente se escoge el mejor número de variables explicativas basándose en aquel conjunto que tenga la menor raíz cuadrática media del error (RMSE). Adicionalmente el algoritmo asigna un nivel de importancia a cada variable explicativa en su aporte a la predicción, lo que permite seleccionar aquellas variables que mejor explican a la dependiente (Liaw & Wiener, 2002)

Al mejor modelo obtenido mediante la regresión por pasos se le aplicará la prueba de F de Fisher para probar la hipótesis global de que el conjunto de coeficientes obtenidos es significativo para explicar a la variable dependiente ventas. Luego se aplica la prueba T a cada coeficiente por separado y mediante el valor p se establece la significatividad de cada coeficiente; también se calculará el coeficiente de determinación R^2 para medir el nivel de ajuste del modelo a los datos

Luego, se probarán los supuestos del modelo: no multicolinealidad, normalidad de los residuos, homocedasticidad y no autocorrelación. Para la detección de multicolinealidad se obtendrá la matriz de correlaciones de las variables cuantitativas, el factor de inflación de la varianza y el número de condición, y en caso de existir multicolinealidad se procederá a eliminar alguna variable con correlación muy alta con otra o realizar alguna transformación de las variables explicativas para obtener un nuevo modelo, y volver a aplicar todas las pruebas anteriores.

Posteriormente se analizarán los residuos y se aplicarán las pruebas de Shapiro-Wilk para la normalidad; el test de Breusch-Godfrey para la autocorrelación y el test de Breusch-Pagan para la homocedasticidad. Si algún supuesto no se cumple, se realizará transformaciones a las variables para que el modelo cumpla los supuestos.

Finalmente se pronosticarán las ventas anuales y trimestrales del año, y con esta información y la información contable-financiera adicional estimada, se elaborará el Estado de Resultados y el Flujo de Efectivo Presupuestados

3.2. Datos, variables y estimaciones adicionales

3.2.1. Datos y variables del modelo econométrico

Los datos que se han recolectado para construir el modelo econométrico corresponden a la información incluida en los Estados de Resultados y Balances Generales anuales de los 263 clientes que tiene la empresa correspondiente al período comprendido entre el 1 de enero de 2018 al 31 de diciembre de 2018. Específicamente se ha obtenido para cada uno de ellos los siguientes datos: activos corrientes totales, pasivos corrientes totales, activos totales, inventarios, pasivos totales, utilidad neta y ventas. Con estos datos se han calculado las razones de liquidez, prueba ácida, razón de deuda total y margen de utilidad neta.

De las bases de datos de la empresa se ha obtenido el monto de ventas durante el año 2018 que ha realizado la empresa ABC a cada uno de los 263 clientes; se ha obtenido el sector al que pertenecen: comercial, industrial o de servicios; y el tamaño del cliente, que se ha categorizado en función del número de empleados que tenga el cliente, siendo pequeña si tiene menos de 50 empleados; mediana, si tiene entre 50 y 200 empleados; y grande, si tiene más de 200 empleados.

Las variables que se consideran para el desarrollo del modelo de regresión múltiple son las siguientes:

ventas: Ventas realizadas a los clientes en miles de dólares

razon corriente: La razón corriente del cliente

prueba acida: La prueba ácida del cliente

razon deuda: La razón de deuda del cliente

margen util: El margen de utilidad del cliente

sector: El sector empresarial al que pertenece el cliente codificado como 0 = comercial, 1 = servicios, 2 = industrial

tamano: El tamaño del cliente codificado como 0 = pequeño, 1 = mediano, 2 = grande

3.2.2. Estimaciones adicionales

Para la elaboración del Estado de Resultados y el Flujo de Efectivo presupuestados trimestral de la empresa ABC se tomará en cuenta las siguientes estimaciones establecidas mediante análisis por el equipo directivo de la empresa ABC:

- Las ventas anuales proyectadas se distribuyen uniformemente durante los 4 trimestres.
- El costo de ventas corresponde al 50% de las ventas del trimestre
- Las cuentas por cobrar a clientes al inicio del año son \$25.670
- La política de crédito a clientes es de 30 días
- Las cuentas por pagar a proveedores al inicio del año son \$30.200
- La política de crédito de proveedores es de 30 días
- Las compras de cada trimestre corresponden al 60% de las ventas del trimestre siguiente
- Los gastos de administración se supondrán en \$42.000 anuales y se distribuyen uniformemente durante los 4 trimestres.
- Los gastos de ventas se estiman en \$35.000 anuales y se distribuyen uniformemente durante los 4 trimestres
- Los gastos financieros se estiman en \$ 11.200 anuales y se distribuyen uniformemente durante los 4 trimestres
- La tasa de impuestos es del 36,25%
- El saldo inicial de efectivo es de \$3.200
- El saldo mínimo de efectivo requerido es de \$5.000
- En el segundo y cuarto trimestre se debe pagar cuotas de \$15.200 correspondientes a la adquisición de un camión para distribución.
- En el primer trimestre del año se pagarán dividendos a accionistas por \$12.000
- Los gastos de administración, ventas y financieros se pagan de contado

CAPÍTULO 4

4. RESULTADOS

En este capítulo se muestran los resultados obtenidos correspondientes al tratamiento estadístico de los datos de modo univariante, bivariante y multivariante.

En el análisis exploratorio univariante, se describen los principales estadísticos descriptivos de la variable cuantitativa ventas anuales (en miles de dólares) que es la variable explicada, y de las variables explicativas cuantitativas razón corriente, prueba ácida, razón de deuda y margen de utilidad. Se obtienen además un histograma y se incluye el p-value de la prueba de normalidad de Shapiro-Wilk en el histograma. Para las variables explicativas categóricas sector y tamaño se incluye un diagrama de barras porcentual que muestra la proporción de empresas para cada categoría del sector y tamaño en la muestra analizada.

En el análisis exploratorio bivariante, se ha realizado un diagrama de dispersión y se ha calculado el coeficiente de correlación entre la variable explicada ventas con cada una de las variables explicativas cuantitativas: razón corriente, prueba ácida, razón de deuda y margen de utilidad; además se ha realizado un histograma y un diagrama de cajas de la variable ventas por cada una de las categorías de las variables cualitativas sector y tamaño.

Para la obtención del modelo econométrico de regresión lineal múltiple se ha construido un primer modelo que está compuesto de la variable explicada ventas anuales y las 6 variables explicativas (cuantitativas y cualitativas) y se ha evaluado la significación global e individual de los coeficientes de regresión. Luego se ha escogido a las mejores variables explicativas utilizando el método de regresión por pasos y por el método random forest Finalmente se ha construido el mejor modelo con las variables anteriormente seleccionadas, se ha probado su significación global e individual y se han probado los supuestos de normalidad, homocedasticidad y no autocorrelación.

Con el mejor modelo encontrado se ha realizado la predicción de ventas anuales para cada cliente y se ha obtenido la suma de dichas proyecciones para obtener el total de ventas del año 2019 de la empresa ABC y con esta información,

y la información adicional descrita en el capítulo 3 se ha elaborado el Estado de Resultados y el presupuesto trimestral de Flujo de Efectivo del año 2019.

4.1. Análisis exploratorio de las variables del modelo econométrico

4.1.1. Análisis univariante

En la figura 4.1 se muestra un encabezado de las razones financieras, el sector y el tamaño de las 263 clientes de la empresa ABC, correspondiente a los 6 primeros registros:

Figura 4-1: Encabezado de los datos

```
> head(datos)
```

	razon_corriente	prueba_acida	razon_deuda	margen_util	sector	tamano	ventas
1	0.814	0.723	0.629	0.101	Servicios	Pequeño	0.99236
2	0.938	0.817	0.553	0.077	Servicios	Pequeño	1.63350
3	0.919	0.487	0.660	0.034	Industrial	Grande	4.13540
4	0.746	0.412	0.725	0.066	Industrial	Mediana	3.20240
5	1.316	0.967	0.514	0.137	Industrial	Mediana	3.97529
6	0.532	0.087	0.566	0.076	Industrial	Mediana	3.90618

Elaborado por: José Barona G.

Fuente: Empresa ABC

En relación a los datos de la variable ventas debe recordarse que están expresados en miles de dólares, por ejemplo, el primer valor es \$992, 36 Un resumen de los principales estadísticos descriptivos de los datos se muestra en la figura 4.2:

Figura 4-2: Resumen de estadísticos descriptivos

```
> summary(datos)
```

razon_corriente	prueba_acida	razon_deuda	margen_util	sector
Min. :0.4140	Min. :0.0300	Min. :0.1310	Min. :0.0220	Comercial :87
1st Qu.:0.6645	1st Qu.:0.4080	1st Qu.:0.4410	1st Qu.:0.0595	Servicios :79
Median :0.7990	Median :0.5340	Median :0.5310	Median :0.0860	Industrial:97
Mean :0.8268	Mean :0.5539	Mean :0.5264	Mean :0.0885	
3rd Qu.:0.9445	3rd Qu.:0.6980	3rd Qu.:0.6215	3rd Qu.:0.1060	
Max. :1.4580	Max. :1.1640	Max. :0.8730	Max. :0.1800	
tamano	ventas			
Pequeño: 40	Min. :0.5831			
Mediana:112	1st Qu.:1.8663			
Grande :111	Median :2.5518			
	Mean :2.5762			
	3rd Qu.:3.4167			
	Max. :4.2423			

Elaborado por: José Barona G.

Fuente: Empresa ABC

De la figura anterior se puede observar que existen 87 empresas del sector comercial, 79 del sector servicios y 97 del sector industrial; también se observa que hay 40 empresas de tamaño pequeño, 112 empresas medianas y 111 empresas grandes. En relación a las ventas anuales (variable explicada), el mínimo valor es de \$583,14, el mayor es \$4.242,25 y el promedio es de \$2.576,17.

La figura 4.3 resume las medias, desviaciones estándar y varianzas de las variables cuantitativas:

Figura 4-3: Medias, desviaciones estándar y varianzas

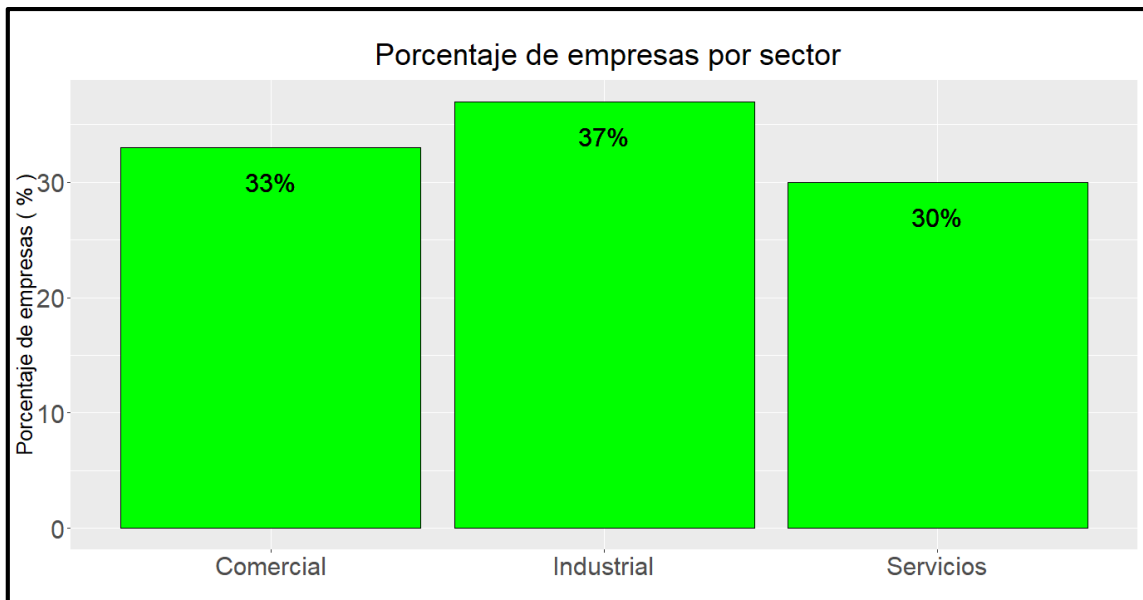
> resumen			
	media	desv_est	varianza
razon_corriente	0.8268365	0.20312199	0.041258542
prueba_acida	0.5538707	0.22026652	0.048517342
razon_deuda	0.5263650	0.12644507	0.015988355
margen_util	0.0884981	0.03723639	0.001386549
ventas	2.5761743	0.91570405	0.838513899

Elaborado por: José Barona G.

Fuente: Empresa ABC

En las figuras 4.4 y 4.5 se muestra la distribución de empresas por sector y tamaño:

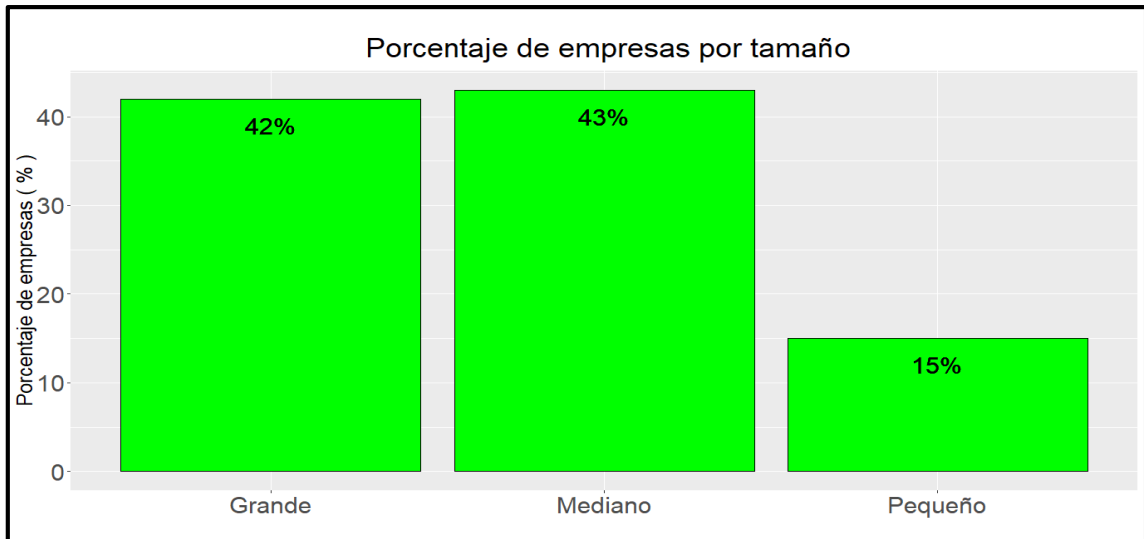
Figura 4-4: Porcentaje de empresas por sector



Elaborado por: José Barona G.

Fuente: Empresa ABC

Figura 4-5: Porcentaje de empresas por tamaño

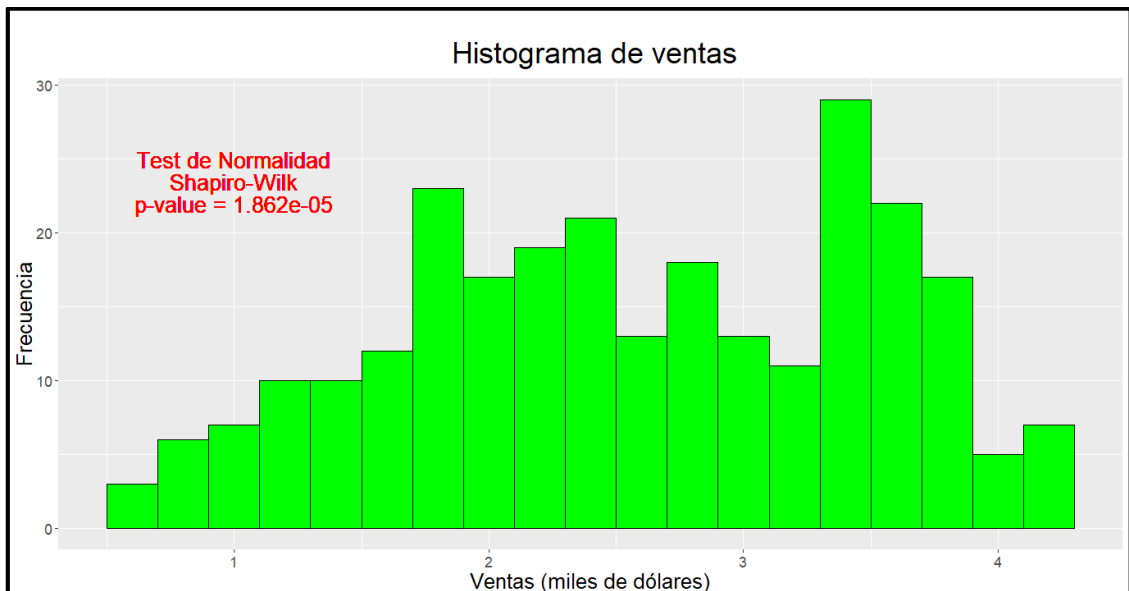


Elaborado por: José Barona G.

Fuente: Empresa ABC

En la figura 4.6 se muestra un histograma de la variable explicada ventas y la prueba de normalidad de Shapiro-Wilk cuyo p-value es de 0,00001862, lo que permite concluir que la variable explicada ventas no sigue una distribución normal

Figura 4-6: Histograma de ventas

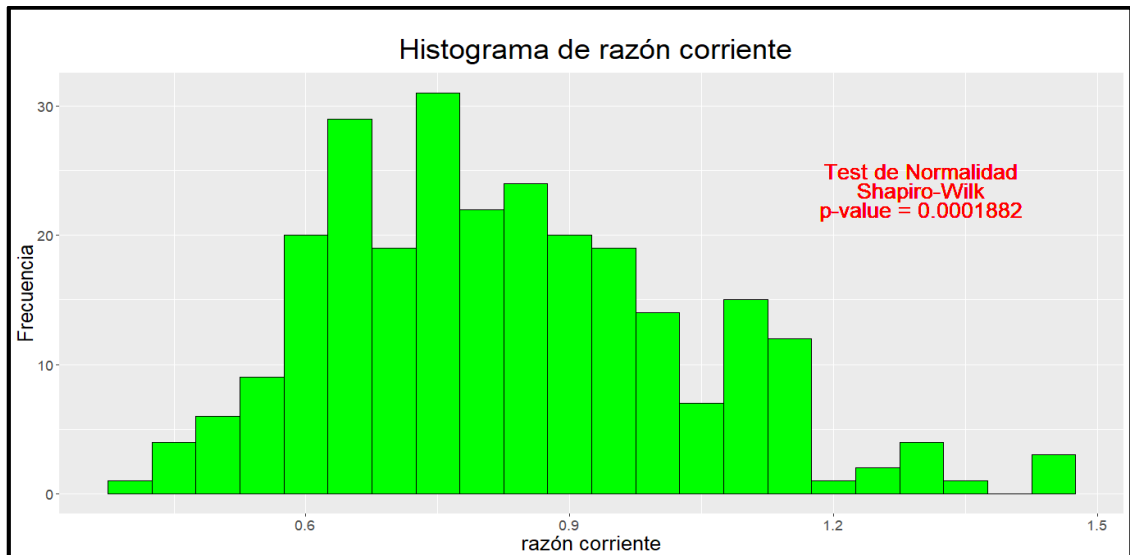


Elaborado por: José Barona G.

Fuente: Empresa ABC

En las figuras 4.7 a 4.10 se muestran los histogramas de las razones financieras

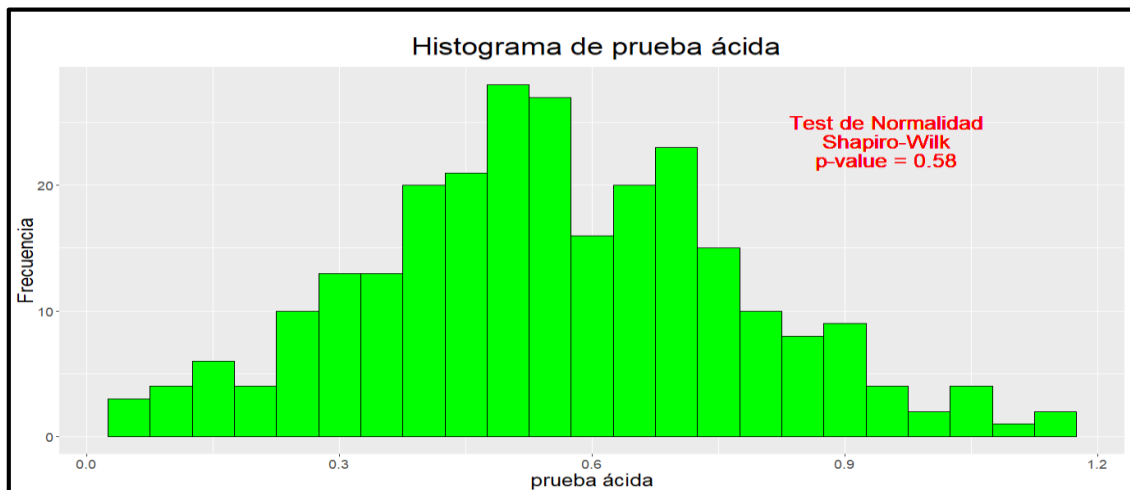
Figura 4-7: Histograma de razón corriente



Elaborado por: José Barona G.

Fuente: Empresa ABC

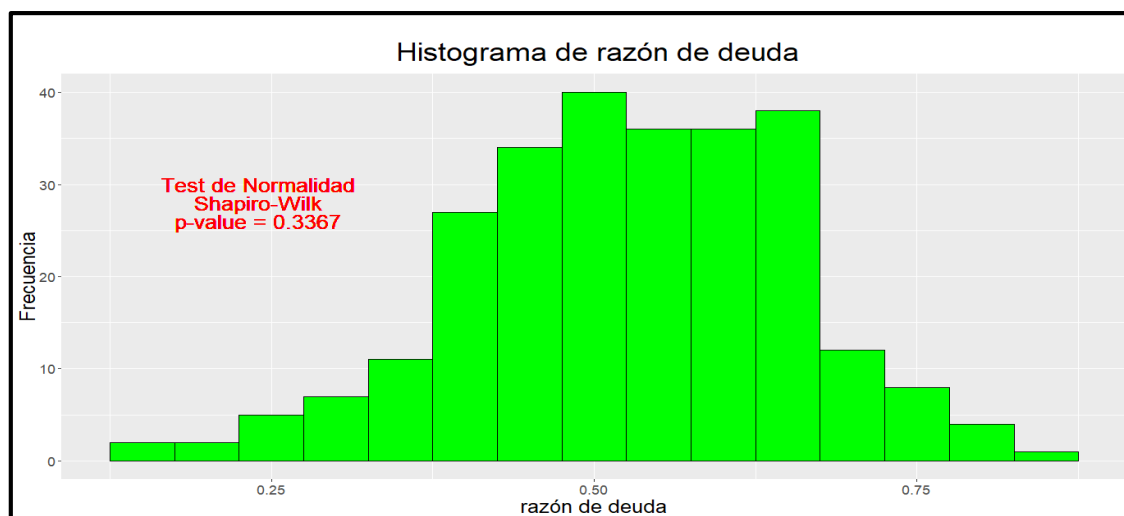
Figura 4-8: Histograma de prueba ácida



Elaborado por: José Barona G.

Fuente: Empresa ABC

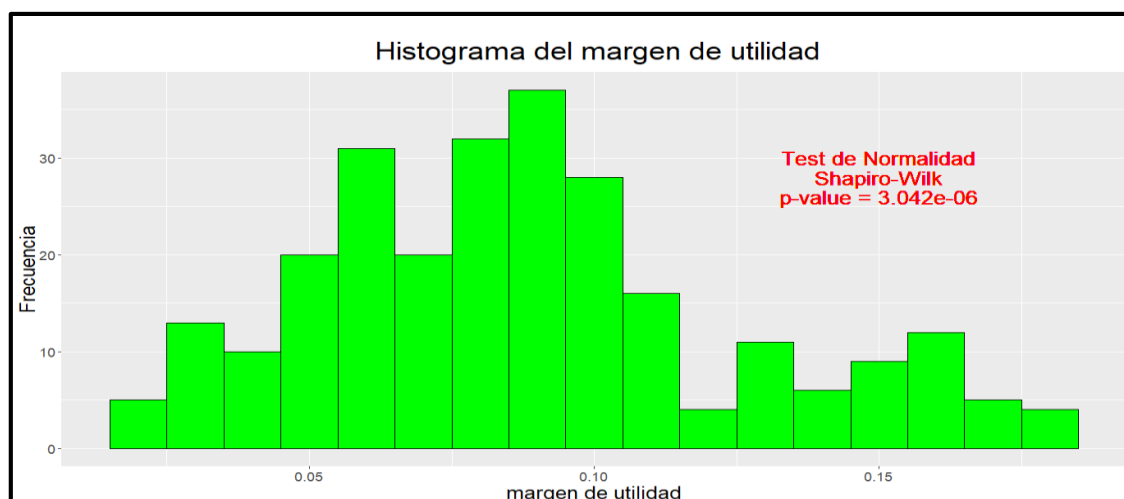
Figura 4-9: Histograma de la razón de deuda



Elaborado por: José Barona G.

Fuente: Empresa ABC

Figura 4-10: Histograma del margen de utilidad



Elaborado por: José Barona G.

Fuente: Empresa ABC

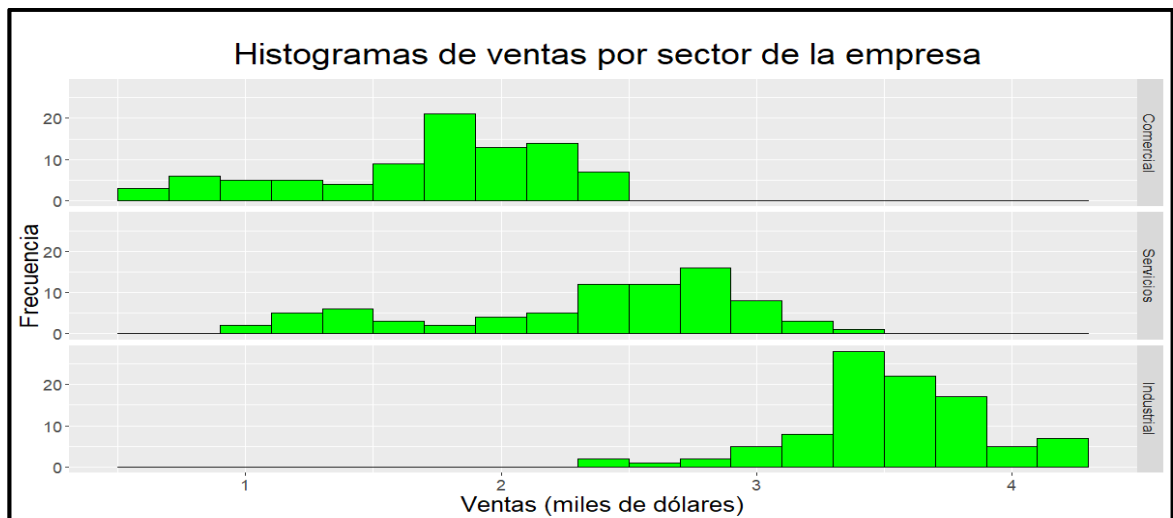
De los histogramas anteriores y la prueba de normalidad de Shapiro-Wilk, se puede concluir que las variables razón corriente y margen de utilidad no siguen una distribución normal ya que sus p-value son 0,0001882 y 3,042e-6 respectivamente por lo que se rechaza la hipótesis nula de normalidad; las variables prueba ácida y razón de deuda siguen una distribución normal debido a que sus p-value son 0,58 y 0,3367 por lo que no se rechaza la hipótesis nula de normalidad.

4.1.2. Análisis bivariante

En esta sección se analizará las ventas en función de cada variable cualitativa (sector y tamaño), y posteriormente, las ventas en función de cada razón financiera.

En las figuras 4.11 y 4.12 se muestran el histograma y el diagrama de cajas de las ventas para cada categoría del sector empresarial del cliente, y se puede apreciar que existe un efecto del sector sobre las ventas, ya que los gráficos se desplazan a medida que cambia la categoría del sector, aunque el efecto no parece ser muy fuerte. En el diagrama de cajas se puede apreciar que existen algunos valores atípicos en la parte inferior del diagrama de cajas del sector industrial, y corresponden a empresas metalmecánicas pequeñas y otras de producción a pequeña escala.

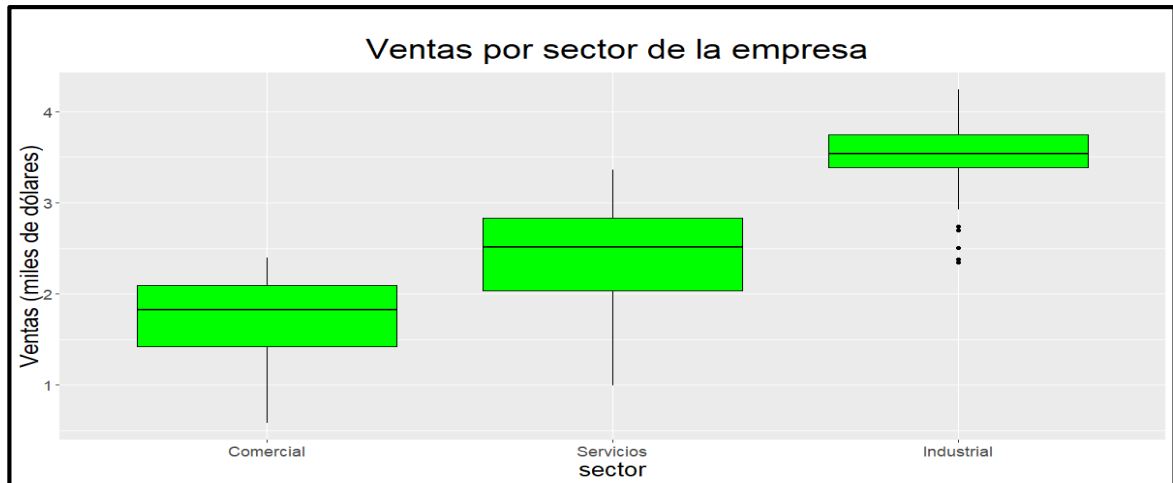
Figura 4-11: Histograma de las ventas por sector



Elaborado por: José Barona G.

Fuente: Empresa ABC

Figura 4-12: Diagrama de cajas de las ventas por sector

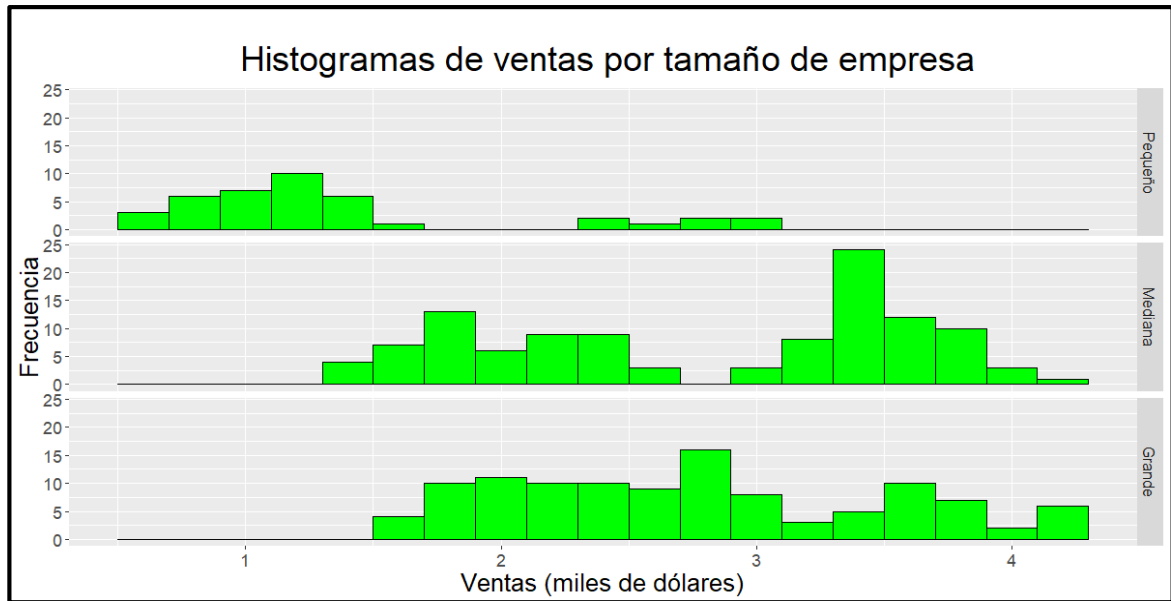


Elaborado por: José Barona G.

Fuente: Empresa ABC

En las figuras 4.13 y 4.14 se muestran el histograma y el diagrama de cajas de las ventas para cada categoría del tamaño del cliente, se puede observar que existe un efecto en las ventas cuando el tamaño de la empresa cambia de pequeña a mediana, sin embargo, no se nota un efecto claro en las ventas cuando el tamaño de la empresa cambia de mediana a grande. En el diagrama de cajas se pueden observar algunos valores atípicos en la parte superior del diagrama de cajas de las empresas pequeñas, realizando las respectivas verificaciones, estas empresas son pequeñas empresas dedicadas a prestar servicios de limpieza a distintos tipos de clientes, por lo que sus compras son mayores que las que se encuentran dentro de la caja de este grupo.

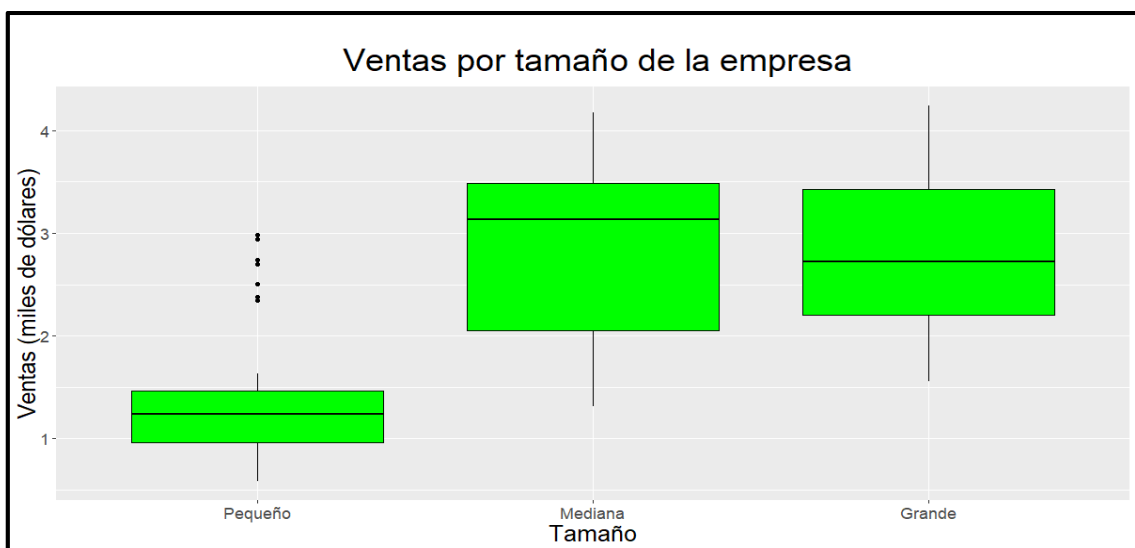
Figura 4-13: Histograma de las ventas por tamaño de la empresa



Elaborado por: José Barona G.

Fuente: Empresa ABC

Figura 4-14: Diagrama de cajas por tamaño de empresa



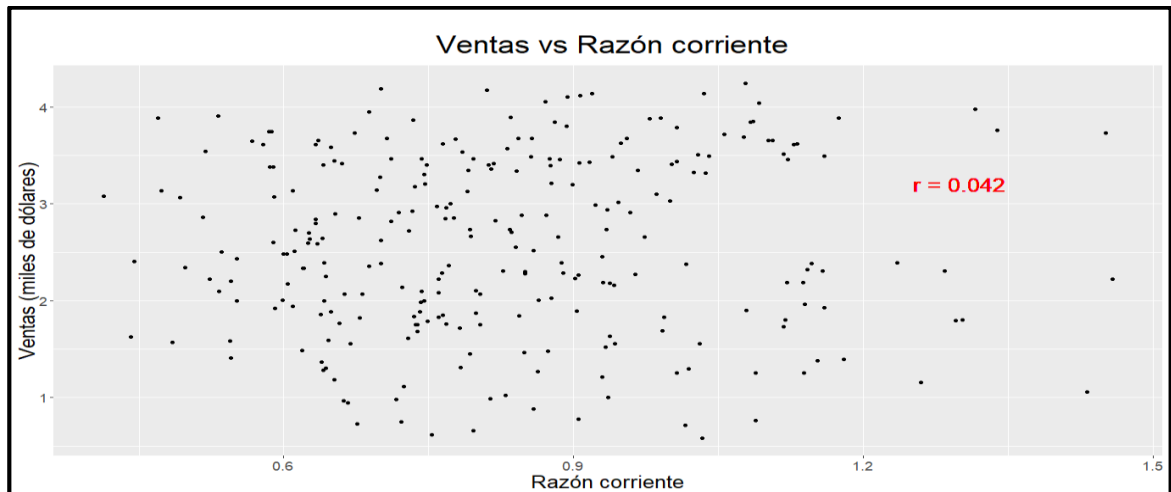
Elaborado por: José Barona G.

Fuente: Empresa ABC

En las figuras 4.15 a 4.18 se muestra el diagrama de dispersión de las ventas con cada una de las razones financieras, y además se muestra el coeficiente de correlación respectivo. En los 4 gráficos el coeficiente de correlación entre las ventas y las razones financieras es muy bajo y prácticamente se puede decir que no existe una relación bivalente directa entre las ventas y ninguna razón financiera;

pero esto no implica necesariamente que no exista una relación multivariante entre las ventas y varias razones financieras en conjunto.

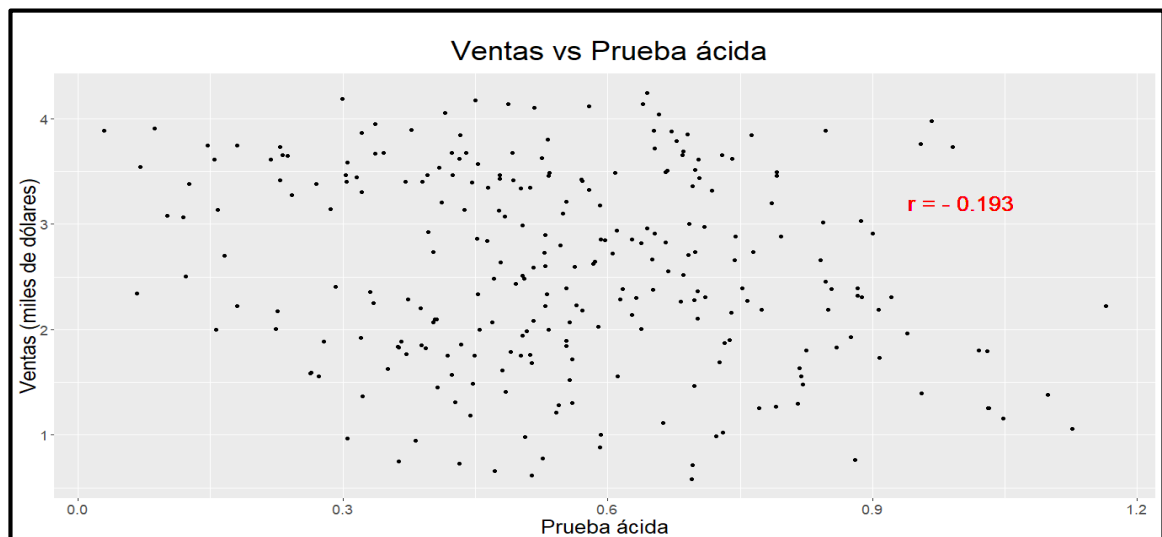
Figura 4-15: Ventas vs razón corriente



Elaborado por: José Barona G.

Fuente: Empresa ABC

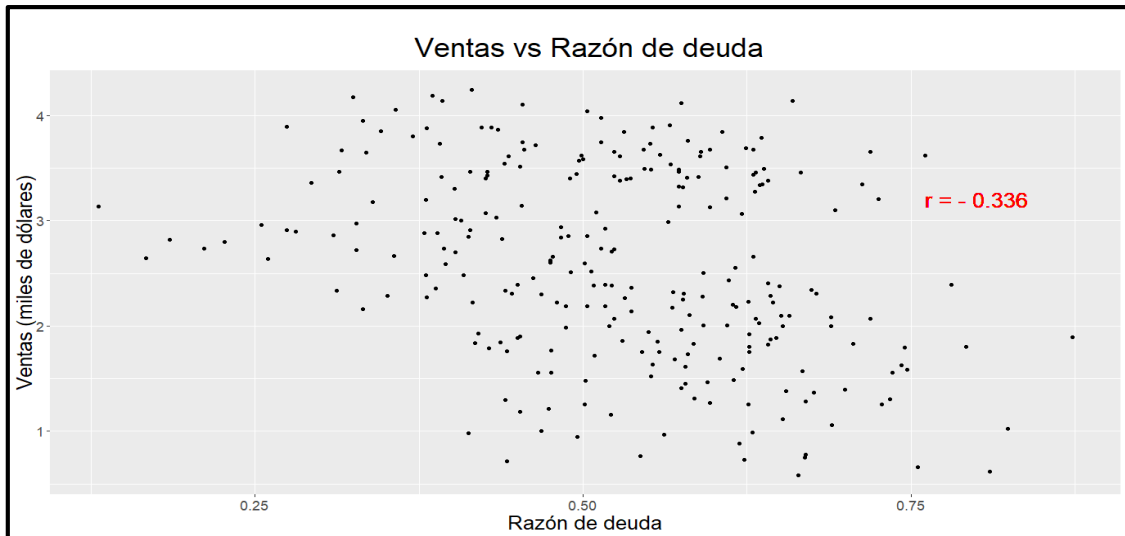
Figura 4-16: Ventas vs prueba ácida



Elaborado por: José Barona G.

Fuente: Empresa ABC

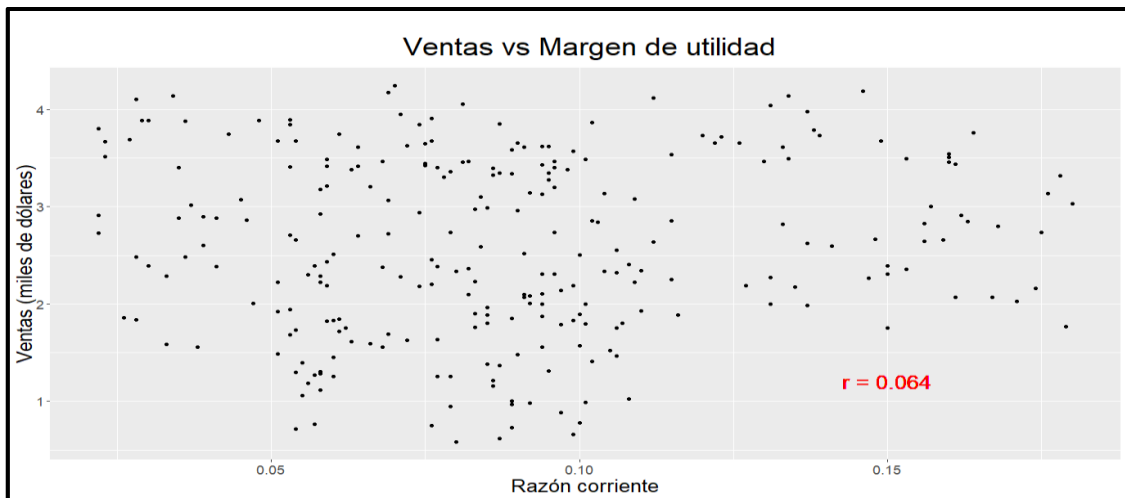
Figura 4-17: Ventas vs razón de deuda



Elaborado por: José Barona G.

Fuente: Empresa ABC

Figura 4-18: Ventas vs margen de utilidad



Elaborado por: José Barona G.

Fuente: Empresa ABC

4.2. Obtención del modelo econométrico

Como primer paso en la construcción del modelo econométrico se obtiene un modelo lineal siendo *ventas* la variable explicada y como predictores las 4 razones financieras y las dos variables dummy (sector y tamaño). En la figura 4.19 se muestra el código y la salida en R para este modelo:

Figura 4-19: modelo lineal con todas las variables

```

> modelo<-lm(ventas ~ razon_corriente + prueba_acida + razon_deuda +
margen_util + sector + tamaño, data = datos)
> summary(modelo)

Call:
lm(formula = ventas ~ razon_corriente + prueba_acida + razon_deuda +
  margen_util + sector + tamaño, data = datos)

Residuals:
    Min       1Q   Median       3Q      Max
-0.51731 -0.15038 -0.00665  0.15647  0.62954

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.78015    0.13713   5.689 3.51e-08 ***
razon_corriente  0.99529    0.28753   3.461 0.00063 ***
prueba_acida  -0.36218    0.27208  -1.331 0.18433
razon_deuda   -0.95698    0.12409  -7.712 2.81e-13 ***
margen_util    0.09508    0.37043   0.257 0.79764
sectorServicios  0.62919    0.06307   9.976 < 2e-16 ***
sectorIndustrial 1.66551    0.04254  39.151 < 2e-16 ***
tamañoMediana  0.86741    0.04216  20.575 < 2e-16 ***
tamañoGrande   1.17597    0.04470  26.307 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2195 on 254 degrees of freedom
Multiple R-squared:  0.9443, Adjusted R-squared:  0.9425
F-statistic: 538.2 on 8 and 254 DF, p-value: < 2.2e-16

```

Elaborado por: José Barona G.

Fuente: Empresa ABC

De la salida en R se puede observar que las variables explicativas en conjunto tienen un R^2 de 0,9443 y un R^2 ajustado de 0,9425 que son valores altos, también se observa un p-value de 2,2e-16 para la prueba F de significación del modelo global que implica que las variables explicativas en conjunto son significativas para explicar las ventas; sin embargo, se puede observar que hay 2 variables que individualmente no son significativas: prueba ácida y margen de utilidad, ya que sus p-value son menores que 0,05.

El hecho de que estas 2 variables sean no significativas puede representar uno o varios problemas, multicolinealidad, heterocedasticidad o autocorrelación, o simplemente es evidencia que estas variables no tienen efecto en las ventas. Se examina, en primer lugar, la posible existencia de multicolinealidad y en la figura 4.20 se muestra la matriz de correlaciones de las 4 razones financieras:

Figura 4-20: Matriz de correlaciones de las razones financieras

```
> cor(datos[, c(1,2,3,4)])
```

	razon_corriente	prueba_acida	razon_deuda	margen_util
razon_corriente	1.00000000	0.82835594	0.17955090	0.08681497
prueba_acida	0.82835594	1.00000000	0.09324488	0.08813320
razon_deuda	0.17955090	0.09324488	1.00000000	-0.01047092
margen_util	0.08681497	0.08813320	-0.01047092	1.00000000

Elaborado por: José Barona G.

Fuente: Empresa ABC

De la matriz de correlaciones se observa que la correlación entre la razón corriente y la prueba ácida es 0,82835594 que es un valor alto para estas dos variables, lo que podría evidenciar la presencia de multicolinealidad; las demás correlaciones tienen valores bajos.

Existen distintas maneras de enfrentar el problema de la multicolinealidad, siendo la más simple de ellas, eliminar una de las variables correlacionada fuertemente con otra; y la más compleja, determinar las componentes principales de las variables cuantitativas predictores, y realizar la regresión sobre dichas componentes principales. Si se opta por eliminar la variable *prueba ácida*, aún persistiría el problema de la no significación de la variable *margen de utilidad*, que no es de origen multicolineal.

Una manera de resolver el problema de la no significatividad de las 2 variables simultáneamente es escoger las variables que sean más significativas individual y globalmente al modelo, se hará esto utilizando dos métodos: mediante regresión por pasos usando la función **step()** de R que establece el mejor conjunto de variables explicativas utilizando el Criterio de Información de Akaike (AIC), y con el algoritmo de bosques aleatorios que utiliza la raíz del error cuadrático medio (RMSE) como criterio de selección del mejor conjunto de variables explicativas, este algoritmo se encuentra en el paquete **randomForest** de R y puede ser llamado desde el paquete **caret**.

En la figura 4.21 se muestra la salida en R de la función **step()**, y se puede observar que cuando se incluyen todas las variables el AIC es 788,78, luego, en un primer paso elimina la variable margen de utilidad y el AIC aumenta a 790,71, luego elimina la variable prueba ácida y el AIC aumenta a 790,87 que es el máximo valor posible para el AIC. De esta manera, las variables explicativas que deben incluirse en el modelo son: razón corriente, razón deuda, sector y tamaño.

Figura 4-21: Regresión por pasos en R

```

1. > step(modelo)
2. Start: AIC=-788.78
3. ventas ~ razon_corriente + prueba_acida + razon_deuda +
margen_util + sector + tamano
4.
5.           Df Sum of Sq      RSS      AIC
6. - margen_util    1    0.003  12.241 -790.71
7. - prueba_acida   1    0.085  12.323 -788.95
8. <none>                12.238 -788.78
9. - razon_corriente  1    0.577  12.815 -778.66
10. - razon_deuda     1    2.866  15.104 -735.45
11. - tamano          2   34.150  46.388 -442.34
12. - sector         2   93.816 106.054 -224.86
13.
14. Step: AIC=-790.71
15. ventas ~ razon_corriente + prueba_acida + razon_deuda + sector
+ tamano
16.
17.           Df Sum of Sq      RSS      AIC
18. - prueba_acida    1    0.086  12.327 -790.87
19. <none>                12.241 -790.71
20. - razon_corriente  1    0.583  12.824 -780.47
21. - razon_deuda     1    2.863  15.104 -737.44
22. - tamano          2   34.912  47.153 -440.03
23. - sector         2   93.879 106.120 -226.69
24.
25. Step: AIC=-790.87
26. ventas ~ razon_corriente + razon_deuda + sector + tamano
27.
28.           Df Sum of Sq      RSS      AIC
29. <none>                12.327 -790.87
30. - razon_deuda     1    2.834  15.161 -738.44
31. - razon_corriente  1    3.532  15.859 -726.62
32. - tamano          2   35.339  47.666 -439.19
33. - sector         2  122.859 135.186 -165.03
34.
35. Call:
36. lm(formula = ventas ~ razon_corriente + razon_deuda + sector +
37. tamano, data = datos)
38.
39. Coefficients:
40. (Intercept)  razon_corriente      razon_deuda  sectors
41. 0.8927          0.6281          -0.9503
0.5609
42. sectorIndustrial  tamanoMediana  tamanoGrande
43. 1.6989          0.8697          1.1822

```

Elaborado por: José Barona G.

Fuente: Empresa ABC

Las figuras 4.22 y 4.23 muestran el proceso de selección de random forest con 1.000 árboles. En la figura 4.22 se puede observar que el menor valor del RMSE es 0.2426648 que corresponde a la selección de 4 variables y en la figura 4.23 se muestra el nivel de importancia de cada variable en el modelo, el menor

valor es para margen de utilidad por lo que se la descarta, y los niveles de importancia de razón corriente y prueba ácida son muy cercanos, aunque razón corriente es el menor de los dos, aunque como se determinó anteriormente, estas variables están muy correlacionadas, por lo que se podría eliminar cualquiera de las dos.

En conclusión, de acuerdo a los resultados que muestran los métodos de regresión por pasos y bosques aleatorios, las variables explicativas que deben seleccionarse son: razón corriente, razón de deuda, sector y tamaño, y estas son las variables que se utilizarán para obtener el mejor modelo de regresión.

Figura 4-22: Selección de variables con random forest

```
> modelo2<-train(x=subset(datos, select = -ventas),
+               y=datos$ventas,
+               method = "rf",
+               ntree = 1000,
+               trControl = trainControl(method = "cv", number = 10))
> modelo2
Random Forest

263 samples
  6 predictor

No pre-processing
Resampling: Cross-validated (10 fold)
Summary of sample sizes: 237, 235, 235, 238, 238, 239, ...
Resampling results across tuning parameters:
```

mtry	RMSE	Rsquared	MAE
2	0.2577230	0.9296215	0.2090633
4	0.2426648	0.9316925	0.1976608
6	0.2456667	0.9295954	0.2013696

```
RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 4.
```

Elaborado por: José Barona G.

Fuente: Empresa ABC

Figura 4-23: Importancia de las variables

```
> modelo2$finalModel$importance
              IncNodePurity
razon_corriente      4.585354
prueba_acida        4.941647
razon_deuda         15.808349
margen_util         3.733168
sector              139.178436
tamano              49.283398
```

Elaborado por: José Barona G.

Fuente: Empresa ABC

En la figura 4.24 se muestra la salida en R correspondiente al modelo con las 4 variables explicativas: razón corriente, razón de deuda, sector y tamaño,

Figura 4-24: Modelo con las 4 variables explicativas

```

> modelo3<-lm(ventas ~ razon_corriente + razon_deuda + sector + tamano
, data = datos)
> summary(modelo3)

Call:
lm(formula = ventas ~ razon_corriente + razon_deuda + sector +
tamano, data = datos)

Residuals:
    Min       1Q   Median       3Q      Max
-0.50769 -0.15453 -0.00343  0.15919  0.64863

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.89266    0.10935   8.163 1.49e-14 ***
razon_corriente 0.62811    0.07334   8.564 1.03e-15 ***
razon_deuda   -0.95031    0.12387  -7.672 3.55e-13 ***
sectorServicios 0.56091    0.03613  15.527 < 2e-16 ***
sectorIndustrial 1.69892    0.03441  49.376 < 2e-16 ***
tamanoMediana  0.86970    0.04203  20.694 < 2e-16 ***
tamanoGrande   1.18217    0.04408  26.818 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2194 on 256 degrees of freedom
Multiple R-squared:  0.9439, Adjusted R-squared:  0.9426
F-statistic: 717.7 on 6 and 256 DF, p-value: < 2.2e-16

```

Elaborado por: José Barona G.

Fuente: Empresa ABC

De la figura 4.24 se puede observar que las variables explicativas en conjunto tienen un R^2 de 0,9439 y un R^2 ajustado de 0,9426 que son valores altos, también se observa un p-value de 2,2e-16 para la prueba F de significación del modelo global que implica que las 4 variables explicativas en conjunto son significativas para explicar las ventas; además los p-value de todas las variables explicativas son significativas, con lo que el modelo cumple adecuadamente con la significación global e individual de las predictoras.

A continuación, se prueban los supuestos de no multicolinealidad, normalidad de los residuos, homocedasticidad y no autocorrelación.

Multicolinealidad

En la figura 4.25 se muestra el Factor de Inflación de la Varianza (VIF) para cada variable del modelo y se observa que todos son mucho menores que 10, adicionalmente el Número de Condición (K) es de 24,36 que es menor que 100 con lo que se puede concluir que no existen problemas de multicolinealidad.

Figura 4-25: Factor de inflación de la varianza

```
> vif(modelo3)
```

razon_corriente	razon_deuda	sectorServicios	sectorIndustrial	tamanoMediana
1.207596	1.334786	1.497916	1.505231	2.358619
tamanoGrande				
2.588849				

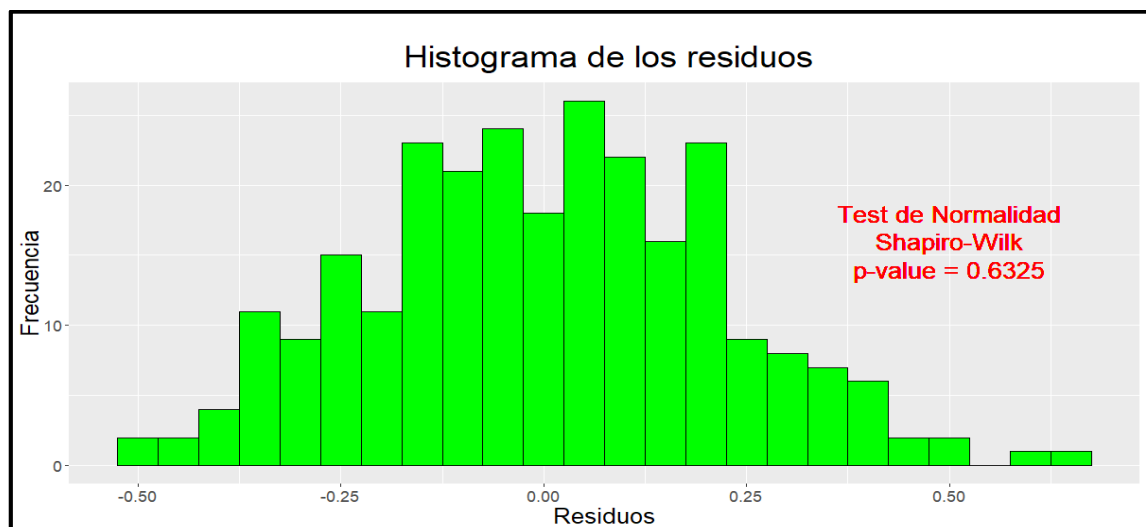
Elaborado por: José Barona G.

Fuente: Empresa ABC

Normalidad de los residuos

Las figuras 4.26 y 4.27 muestran el histograma y el gráfico QQ-normal de los residuos de la regresión; también se muestra el p-value de 0,6325 para la prueba de normalidad de Shapiro-Wilk, lo que permite concluir la normalidad de los residuos

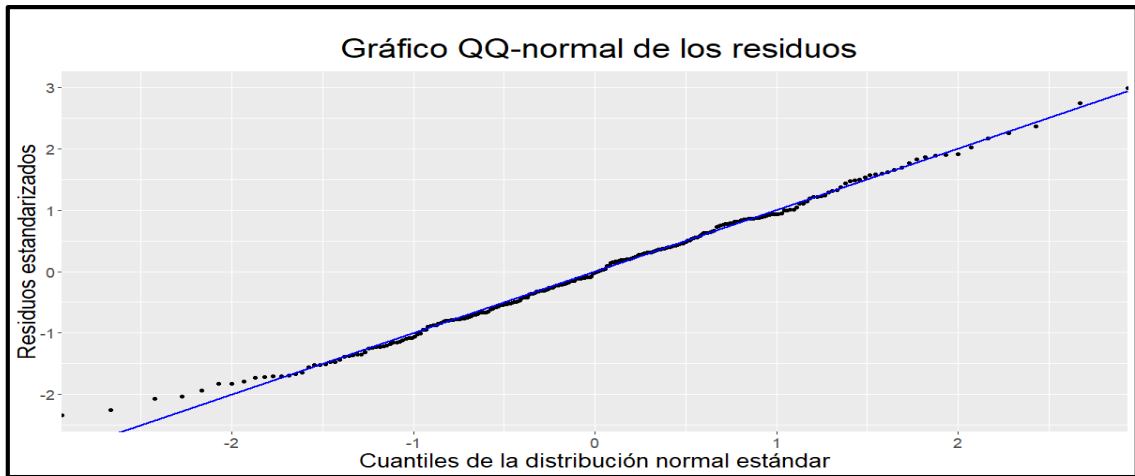
Figura 4-26: Histograma de los residuos



Elaborado por: José Barona G.

Fuente: Empresa ABC

Figura 4-27: Gráfico QQ-normal



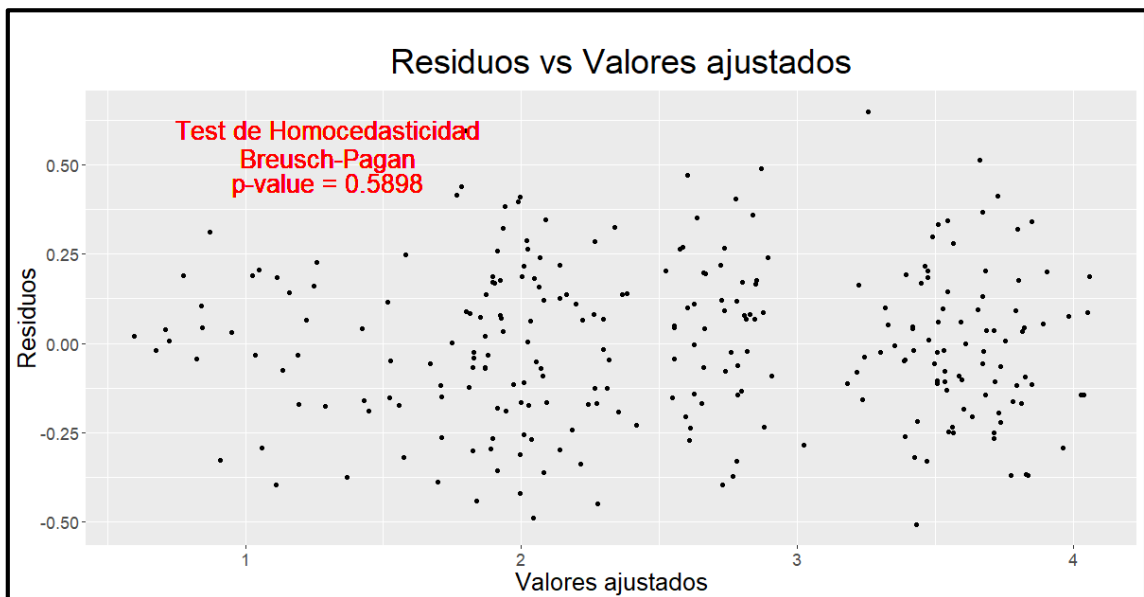
Elaborado por: José Barona G.

Fuente: Empresa ABC

Homocedasticidad

La figura 4.28 muestra la gráfica de los residuos vs valores ajustados y se observa que no existe algún patrón en los puntos que pudiera sugerir la existencia de heterocedasticidad; también se muestra el p-value de 0,5898 correspondiente al test de homocedasticidad de Breusch-Pagan, lo que permite concluir que los residuos son homocedásticos.

Figura 4-28: Residuos vs valores ajustados



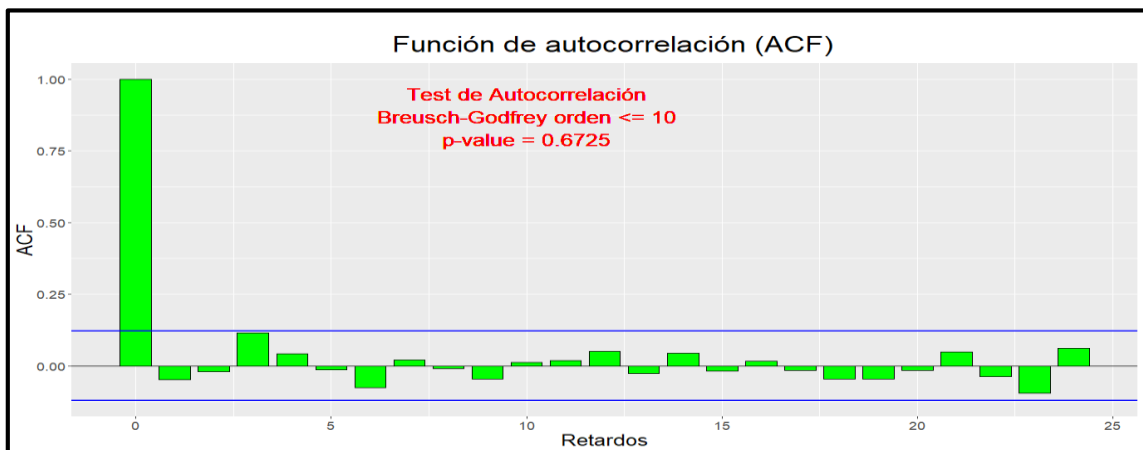
Elaborado por: José Barona G.

Fuente: Empresa ABC

Autocorrelación

La figura 4.29 muestra la función de autocorrelación ACF de los residuos del modelo y el p-value 0,6725 del test de autocorrelación de Breusch-Godfrey para un orden menor o igual que 10, por lo que se puede concluir que no existe autocorrelación de los residuos.

Figura 4-29: Función de autocorrelación



Elaborado por: José Barona G.

Fuente: Empresa ABC

En resumen, el modelo que incluye razón corriente, razón de deuda, sector y tamaño, cumple con todos los supuestos del modelo de regresión y con la significación de los coeficientes individuales y en conjunto.

En la Tabla 4.1 se muestran las 9 ecuaciones de regresión que estiman el promedio de las ventas \hat{V} correspondientes a la combinación de niveles de los factores *sector* y *tamaño* del cliente que se utilizarían en la interpretación y predicción (RC = razón corriente, RD = razón de deuda):

Tabla 4-1: Ecuaciones de regresión

Sector	Tamaño	Ecuación
Comercial	Pequeña	$\hat{V} = 0,89266 + 0.62811(RC) - 0.95031(RD)$
Comercial	Mediana	$\hat{V} = 1,76236 + 0.62811(RC) - 0.95031(RD)$
Comercial	Grande	$\hat{V} = 2,07483 + 0.62811(RC) - 0.95031(RD)$
Servicios	Pequeña	$\hat{V} = 1,45357 + 0.62811(RC) - 0.95031(RD)$
Servicios	Mediana	$\hat{V} = 2,32327 + 0.62811(RC) - 0.95031(RD)$
Servicios	Grande	$\hat{V} = 2,63574 + 0.62811(RC) - 0.95031(RD)$
Industrial	Pequeña	$\hat{V} = 2,59158 + 0.62811(RC) - 0.95031(RD)$
Industrial	Mediana	$\hat{V} = 3,46128 + 0.62811(RC) - 0.95031(RD)$
Industrial	Grande	$\hat{V} = 3,77375 + 0.62811(RC) - 0.95031(RD)$

Elaborado por: José Barona G.

Fuente: Empresa ABC

Hay que recordar que las ventas están expresadas en miles de dólares, por lo tanto, la interpretación de los coeficientes de RC y RD es la siguiente:

- Un incremento de una unidad en la razón corriente produce un incremento de \$628,11 en el promedio de las ventas, manteniendo el sector, tamaño y razón de deuda constantes.
- Un incremento de una unidad en la razón de deuda produce una disminución de \$950,31 en el promedio de las ventas, manteniendo el sector, tamaño y razón corriente constantes.

4.3. Estados financieros presupuestados

Con las suposiciones establecidas anteriormente en el capítulo 3 y con la proyección de ventas anual de la regresión para todos los clientes de la empresa ABC que resulta en un valor de \$677.533,90, se puede elaborar el Estado de Resultados Presupuestado que se muestra en la figura 4.30. Se observa que se proyecta una utilidad anual de \$159.736.

Figura 4-30: Estado de Resultados Presupuestado

EMPRESA ABC	
Estado de Resultados Presupuestado	
Del 1 de enero al 31 de diciembre de 2019	
(en dólares)	
Ventas	677.534
(-) Costo de Ventas	<u>-338.767</u>
= Utilidad Bruta en Ventas	338.767
(-) Gastos de Administración	-42.000
(-) Gastos de Ventas	-35.000
(-) Gastos Financieros	<u>-11.200</u>
= Utilidad Antes de Impuestos	250.567
(-) Impuestos (36,25%)	<u>-90.831</u>
= Utilidad Neta	<u>159.736</u>

Elaborado por: José Barona G.

Fuente: Empresa ABC

Con el supuesto de la uniformidad trimestral en las ventas que resulta en un valor de \$169.383,50 y los supuestos adicionales del capítulo 3 se elabora el Presupuesto de Flujo de Efectivo que se muestra en la figura 4.31

Figura 4-31: Flujo de Efectivo Presupuestado

EMPRESA ABC				
Presupuesto de Flujo de Efectivo				
trimestres del año 2019				
(en dólares)				
	Trim. 1	Trim. 2	Trim. 3	Trim. 4
Saldo inicial de efectivo	3.200	9789	40292	85996
+ Entradas de Efectivo	138.592	169.383	169.383	169.383
= Efectivo Disponible	<u>141.792</u>	<u>179.172</u>	<u>209.676</u>	<u>255.379</u>
(-) Salidas de Efectivo:				
Pago a proveedores	-97.953	-101.630	-101.630	-101.630
Gastso Administrativos	-10.500	-10.500	-10.500	-10.500
Gastso de Ventas	-8.750	-8.750	-8.750	-8.750
Gastos Financieros	-2.800	-2.800	-2.800	-2.800
Cuotas de camión		-15.200		-15.200
Duidendos a accionistas	-12.000			
= Saldo Final de Efectivo	<u>9.789</u>	<u>40.292</u>	<u>85.996</u>	<u>116.499</u>
(-) Saldo Mínimo de Efectivo	<u>-5.000</u>	<u>-5.000</u>	<u>-5.000</u>	<u>-5.000</u>
= Excedente / Déficit Acumulado de Efectivo	<u>4.789</u>	<u>35.292</u>	<u>80.996</u>	<u>111.499</u>

Elaborado por: José Barona G.

Fuente: Empresa ABC

CAPÍTULO 5

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

Como conclusiones del presente trabajo se pueden anotar las siguientes relacionadas con los objetivos planteados:

En el análisis descriptivo univariante se pudo establecer que la variable respuesta *ventas* no sigue una distribución normal; las variables explicativas *prueba ácida* y *margen de utilidad* siguen una distribución normal, mientras que *razón corriente* y *razón de deuda* no siguen una distribución normal.

En el análisis descriptivo bivariante se pudo determinar gráficamente que el *sector* y el *tamaño* de la empresa tienen una influencia positiva en el nivel de ventas de la empresa ABC, es decir al cambiar el nivel o el tamaño de la empresa se produce un efecto sobre el nivel de ventas. Los coeficientes de correlación bivariantes entre las ventas y cada una de las razones financieras resultan cercanos a cero, lo que permite concluir que, de manera individual, ninguna razón financiera tiene un efecto sobre el nivel de ventas

En el estudio multivariante de regresión múltiple, se ha podido establecer que la razón corriente y la razón de deuda son significativas para explicar las ventas en el modelo de regresión lineal, mientras que la prueba ácida y el margen de utilidad no resultaron significativas, esto significa que la liquidez del cliente y su estructura financiera general influyen significativamente en las ventas que se realicen a dicho cliente; de igual forma, el sector y el tamaño son factores significativos para explicar las ventas, sin embargo, estas conclusiones no se pueden generalizar para otros datos.

Por lo tanto, se puede concluir que las variables financieras razón corriente y razón de deuda, así como los factores sector y tamaño son significativos individual y globalmente cuando se analizan de forma multivariante, y producen un coeficiente de determinación R^2 de 0,9439 y un R^2 ajustado de 0,9426 que corresponden a la variabilidad en las ventas explicada por el modelo que se ha desarrollado.

Al utilizar el modelo obtenido y proyectar las ventas para el 2019 se obtiene que el Estado de Resultados muestra una utilidad neta de \$159.736 en el año, FCNM

mientras que el Flujo de Efectivo trimestral muestra que en los 4 trimestres se obtiene superávit de efectivo, lo que permitiría prever que no se requerirán préstamos de corto plazo para cubrir déficit de efectivo.

5.2. Recomendaciones

Como recomendaciones para estudios posteriores se anotan las siguientes: Desarrollar un modelo que incluya datos de panel, es decir, que recoja información a lo largo de varios años sobre los indicadores financieros, sector y tamaño de los clientes y compararlo con el modelo de corte transversal que se ha desarrollado en el presente trabajo.

Construir un modelo de corte transversal, pero considerando interacción entre variables y factores, esto abarca innumerable cantidad de posibilidades, lo que permitiría determinar si existe algún modelo que explique de mejor manera al modelo obtenido.

Para empresas medianas y pequeñas, se recomienda el uso de este tipo de modelos de corte transversal para la proyección de ventas, ya que estas empresas si disponen de información de sus clientes cuando venden a crédito, y permitirá una mejor planificación financiera en todas las áreas de la empresa y la fijación de objetivos y metas anuales.

Considerar la posibilidad de incluir otras variables que pudieran estar influyendo en las ventas a los clientes, y construir modelos como los descritos en los párrafos anteriores para determinar su bondad de ajuste y capacidad predictiva.

Finalmente realizar comparación entre los distintos modelos creados para obtener el mejor de acuerdo a las características deseadas del modelo, ya sea para predicción o interpretación de las variables explicativas.

6. Referencias

- Agus, A., Hassan, Z., & Ahmad, S. (17 de Junio de 2017). *The Significant Impact of Customer Relations Practices (CRP), Information Technology (IT) and Information Sharing between Supply Chain Partners (IS) on Product Sales*. Obtenido de Gading Journal for the Social Sciences: <http://www.gading.learningdistance.org/index.php/gading/article/view/61>
- Court, E., & Rengifo, E. (2011). *Estadísticas y Econometría Financiera*. Buenos Aires: Cengage Learning.
- Fernández, A. (2005). *Econometría*. Madrid: Pearson Educación S.A.
- Fernández, A. (2005). *Econometría*. Madrid: Pearson Educación S.A.
- Gitman, L., & Zutter, C. (2012). *Principios de Administración Financiera - Décimosegunda edición*. México: Pearson Educación.
- Gujarati, D., & Porter, D. (2010). *Econometría*. México D.F.: MCGRAW-HILL/INTERAMERICANA EDITORES, S.A. DE C.V.
- Hanke, J. (2010). *Pronóstico en los Negocios, 9na. edición*. México: PEARSON EDUCACIÓN.
- Hansen, B. (Enero de 2018). *University of Wisconsin*. Obtenido de ECONOMETRICS: <https://www.ssc.wisc.edu/~bhansen/econometrics/>
- Khalid, W., & Khan, S. (5 de Julio de 2017). *Impact of Operating and Financial Expenses on Sales Revenue: The Case of Fauji Fertilizer Company Limited*. Obtenido de International Journal of Business and Economics Research: <http://www.sciencepublishinggroup.com/journal/paperinfo?journalid=178&doi=10.11648/j.ijber.20170603.12>
- Liaw, A., & Wiener, M. (Diciembre de 2002). *Classification and Regression by randomForest*. Obtenido de ResearchGate: https://www.researchgate.net/publication/228451484_Classification_and_Regression_by_RandomForest
- Maddala, G. S. (1977). *Econometrics*. New York: McGraw Hill.
- Novales, A. (1993). *ECONOMETRÍA, segunda edición*. Madrid: MCGRAW-HILL INTERAMERICANA DE ESPAÑA S.A.U.
- Pérez, C. (2002). *Estadística aplicada a través de Excel*. Madrid: Pearson Educación S.A.

- Ross, S., Westerfield, R., & Jordan, B. (2010). *Fundamentos de Finanzas Corporativas - Novena Edición*. México: McGraw-Hill Interamericana S.A. de C.V.
- Wang, J. L., Li, T., Yin, S., & He, X. (30 de Enero de 2018). *The Research of Regression Method for Forecasting Monthly Electricity Sales Considering Coupled Multi-factor*. Obtenido de IOP Science: <http://iopscience.iop.org/article/10.1088/1755-1315/108/5/052105/meta>
- Wooldridge, J. (2010). *Introducción a la Econometría un Enfoque Moderno 4a edición*. México: Cengage Learning Editores S.A.