

Capítulo 1

1. MARCO TEÓRICO

1.1 Histograma

La forma más común de representación gráfica de una distribución de frecuencias es el **histograma**. El histograma de una distribución de frecuencias se construye con rectángulos adyacentes, las alturas de los cuales representan las frecuencias de clase mientras que sus bases se extienden entre sucesivas fronteras de clase.

1.2 Medidas Descriptivas

Dado un conjunto de n medidas u observaciones, x_1, x_2, \dots, x_n , podemos describir su centro (medio o lugar central) de muchas maneras. Las más comunes son la **media aritmética** y la **mediana**.

La **media** o también conocida como **media aritmética** se define como:

$$\text{Media} \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

A veces es preferible usar la **mediana** como medida descriptiva del centro, o lugar, de un conjunto de datos. Esto es así, sobre todo, cuando se desea reducir los cálculos o eliminar el efecto de valores extremos (muy grandes o muy pequeños). La mediana de n observaciones, x_1, x_2, \dots, x_n , puede definirse como la mitad del conjunto de datos después de que las observaciones se han colocado en serie ordenada. Si las observaciones se organizan en serie ordenada y n es un número impar, la mediana es el valor de la observación con el número $\frac{n+1}{2}$; si n es un número par, la mediana se define como la media (promedio) de las observaciones con los números $\frac{n}{2}$ y $\frac{n+2}{2}$.

Ahora, con respecto a la **varianza** y la **desviación estándar** para una muestra representan medidas de dispersión alrededor de la media. Se calculan de manera parecida a aquellas para una población. La varianza s^2 es:

$$\text{Varianza} \quad s^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

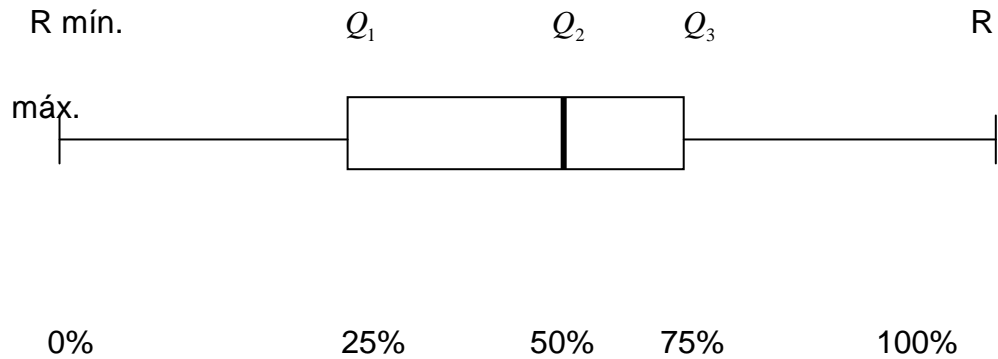
$$\text{Desviación estándar} \quad s = \sqrt{s^2}$$

1.2.1 Medidas de posición de datos

Con respecto a otras medidas de dispersión, además de la varianza y la desviación estándar que son las más útiles en análisis estadístico, existen otras técnicas con las cuales puede medirse la dispersión de un conjunto de datos. Estas medidas adicionales de dispersión son los denominados **cuartiles**. Cada conjunto de datos tiene **tres cuartiles** que lo dividen en cuatro partes iguales. **El primer cuartil** es ese valor debajo del cual clasifica el 25% de las observaciones, y sobre el cual puede encontrarse el 75% restante. **El segundo cuartil** es justo la mitad. La mitad de las observaciones están por debajo y la mitad por encima; en este sentido, es lo mismo que la mediana. **El tercer cuartil** es el valor debajo del cual está el 75% de las observaciones y encima del cual puede encontrarse el 25% restante.

1.3 Diagrama de Caja

La información resumida contenida en los cuartiles pone relieve en la representación gráfica llamada **diagrama de caja**. La mitad central de los datos, que va del primero al tercer cuartiles, se representa con un rectángulo. La mediana se identifica con una barra dentro de esta caja. Una línea se extiende del tercer cuartil al máximo y otra del primer cuartil al mínimo.



Dibujo 1. Representación de Cuartiles. Libro Estadística aplicada a los negocios y la economía.

Los diagramas de caja son especialmente eficaces para la descripción gráfica de comparaciones entre conjuntos de observaciones. Son fáciles de entender y ejercen un poderoso impacto visual.

1.4 Coeficientes de Variación

Como sabemos, un uso importante de la desviación estándar es servir como medida de dispersión. Sin embargo, se aplican ciertas limitaciones. Cuando se consideran 2 o más distribuciones que tienen medias significativamente diferentes, o que están medidas en

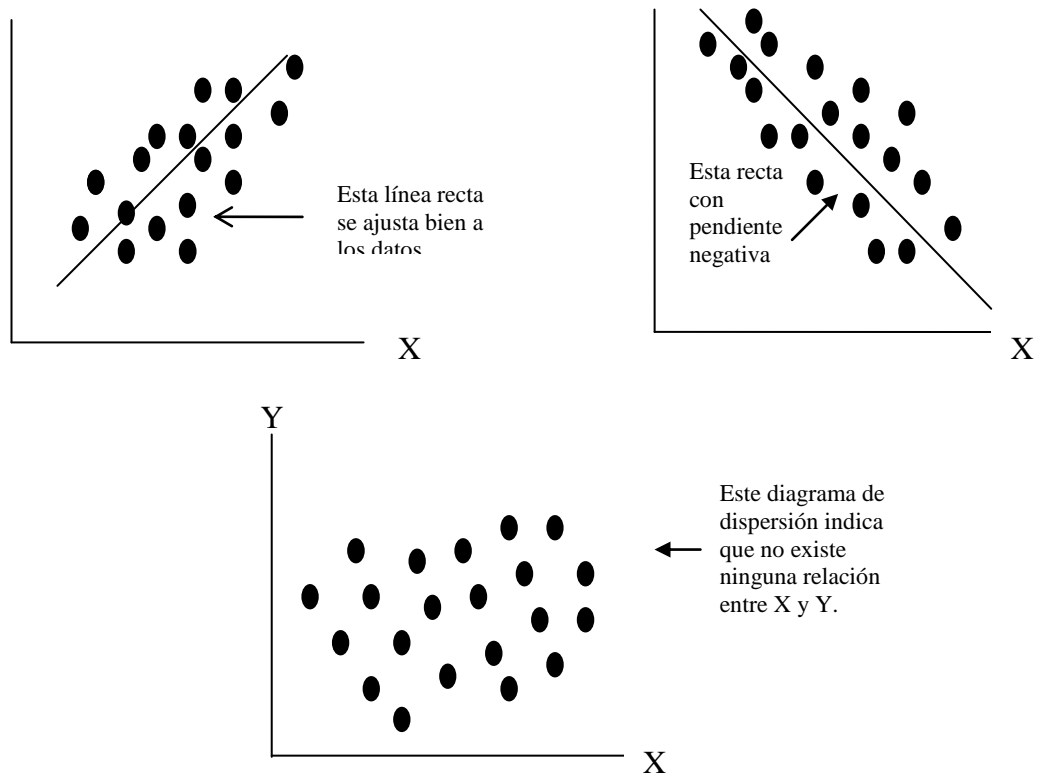
unidades distintas, es peligroso sacar conclusiones respecto a la dispersión sólo con base en la desviación estándar.

Por tanto, con frecuencia debemos considerar el **Coefficiente de Variación** (CV), el cual sirve como medida relativa de dispersión. El coeficiente de variación determina el grado de dispersión de un conjunto de datos relativo a su medida. Se calcula la desviación estándar de una distribución por su media y multiplicando por 100.

$$\text{Coeficiente de variación} \quad CV = \frac{s}{\bar{X}}(100)$$

1.5 Regresión Lineal

Es un modelo en el cual participan dos tipos de variables: La variable **X** que es la variable independiente (o variable explicativa), y la variable **Y** que es la dependiente (o variable de respuesta); éstas a su vez se relacionan y se presentan por medio de una línea recta.



Dibujo 2. Ejemplos de Diagrama de dispersión. Libro Estadística aplicada a los negocios y la economía.

1.5.1 Regresión Lineal Múltiple

En un modelo de regresión múltiple, Y es una función de dos o más variables independientes. Un modelo de regresión con k variables se puede expresar así:

$$Y = f(X_1, X_2, X_3, \dots, X_k)$$

1.5.2 Autocorrelación

La **autocorrelación** ocurre cuando los términos de error no son independientes. Existe la autocorrelación positiva cuando los signos iguales se agrupan y existe una autocorrelación negativa cuando cada error es seguido de un error de signo opuesto. Los diagramas residuales nunca son tan obvios o tan fáciles de leer, afortunadamente existe una forma más confiable para detectar la autocorrelación en base en la prueba de **Durbin-Watson**.

1.5.2.1 Estadístico de Durbin-Watson

El estadístico de Durbin-Watson se calcula así:

$$d = \frac{\sum (e_i - e_{i-1})^2}{\sum e_i^2}$$

En donde e_i es el error en el periodo de tiempo t , y e_{i-1} es el error en el periodo anterior. La fórmula requiere que el término de error $(Y_i - \hat{Y}_i)$ se calcula para cada periodo y es difícil calcularlo manualmente. Este valor se utiliza para probar la hipótesis de que no existe correlación entre términos de error sucesivos, así:

$$H_o : \rho_{e_i, e_{i-1}} = 0 \quad (\text{No existe autocorrelación})$$

$$H_a : \rho_{e_i, e_{i-1}} \neq 0 \quad (\text{Existe autocorrelación})$$

1.5.3 Análisis de Correlación

El coeficiente de correlación fue desarrollado por Carl Pearson a finales de siglo, y algunas veces se les llama el coeficiente de correlación producto-momento de Pearson. Representado con una r , el coeficiente de correlación puede asumir cualquier valor entre -1 y +1; es decir:

$$-1 \leq r \leq +1$$

Un valor de $r = -1$ indica una relación perfecta entre X y Y. Todas las observaciones quedan en una línea recta perfecta con una pendiente negativa. Por tanto, X y Y se moverán en direcciones opuestas, en cambio si $r = 1$ la relación entre X y Y será positiva perfecta. Por el contrario, si se muestra muy poca o ninguna relación entre X y Y, r se aproxima a cero. En general, entre mayor sea el valor absoluto de r , más fuerte será la relación entre X y Y.

1.5.3.1 Coeficiente de Determinación

El coeficiente de determinación r^2 tiene significado sólo para las relaciones lineales. Dos variables pueden tener un r^2 de cero y sin embargo, están relacionadas en sentido curvilíneo.

1.5.4 Análisis de Varianza (Anova)

Dado el modelo de regresión, puede realizarse el **análisis de varianza** (ANOVA). El procedimiento del **ANOVA** prueba si alguna de las variables independientes tiene una relación con la variable dependiente. Si una variable independiente no está relacionada con la variable Y, su coeficiente debería ser cero. El procedimiento ANOVA prueba la hipótesis nula de que todos los valores β son cero contra la hipótesis alternativa de que por lo menos un β no es cero.

Es decir:

$$H_o : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_A : \exists \beta_i \neq 0$$

Si no se rechaza la hipótesis nula, entonces no hay relación lineal entre Y y cualquiera de las variables independientes. Por otra parte, si la hipótesis nula se rechaza, por lo menos una variable independiente está relacionada linealmente con Y.

El proceso **ANOVA** establece una tabla y utiliza la prueba F, el formato general para una regresión múltiple es:

Fuente de Variación	Suma de Cuadrados	Grados de libertad	Cuadrado Medio	Valor F
Entre muestras (tratamiento)	SCR	K	SCR/k	F=CMR/CME
Dentro de las muestras (error)	SCE	n-k-1	SCE/n-k-1	
Variación total	SCT	n-1		

Tabla I. ANOVA. Libro Probabilidad y Estadística para Ingenieros.

Capítulo II

2. TRATAMIENTO DE LOS DATOS

2.1 Introducción

Durante los últimos años la empresa distribuidora de celulares y accesorios para los mismos, ha gozado de un crecimiento considerable en sus ingresos por ventas así como de sus gastos y costos de ventas, por lo que esta parte del trabajo permitirá realizar un análisis de sus variables. Este análisis permitirá detectar los movimientos de las variables de estudio.

Se realizará un estudio de las variables (Ventas, Gastos y Costo de Ventas), mediante gráficos de series de tiempo que nos permitirá ver la fluctuación de estas variables de forma diaria durante el año 2.006 y poder llegar a conclusiones sobre la rentabilidad de la empresa. El propósito de este análisis es predecir o proyectar los valores futuros de las variables a partir de observaciones anteriores; el método que utilizaremos en este trabajo para proyectar las variables de estudio es el método de regresión lineal múltiple que se lo analizará en el capítulo 3.

Además se analizará descriptivamente a las variables de estudio, es decir cálculos de medias, desviaciones estándares, máximos y mínimos y se llegará a conclusiones de gran importancia para el mejoramiento de la utilidad de la empresa.

2.2 Análisis de las ingresos por ventas diarias en el año 2.006

2.2.1 Análisis de los ingresos por venta, diarias en el año 2.006 por medio de un gráfico de serie de tiempo.

A continuación analizaremos la tendencia de los ingresos obtenidos por la empresa distribuidora de celulares para el año 2.006 de forma diaria, con la ayuda de un gráfico de serie de tiempo. El propósito es determinar las fluctuaciones del negocio durante el transcurso de los días y observar los meses donde la empresa llegó a tener mayores ingresos o ingresos bajos.

El gráfico de serie de tiempo a analizar es el siguiente:

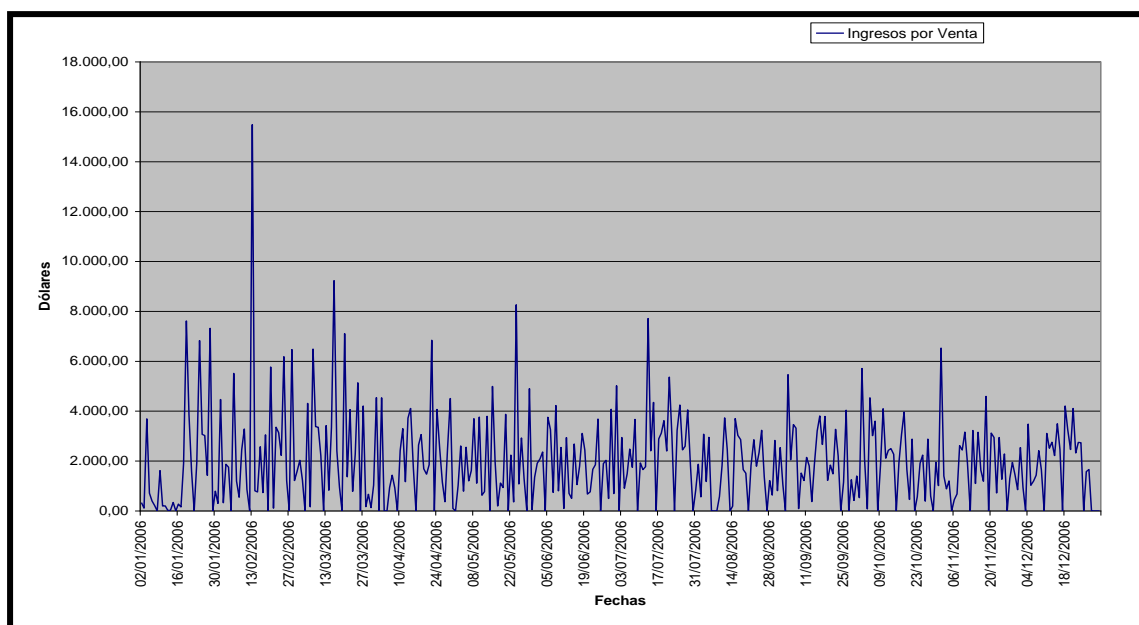


Gráfico 2.1. Diagrama de Serie de tiempo

Los ingresos obtenidos por la empresa distribuidora de celulares durante el año 2.006 tienen una fluctuación muy variada y no pareja. Se puede observar que la empresa ha logrado ingresos superiores a los \$6.500, por lo que se puede considerar de gran importancia mencionarlos, los días donde se logró estos ingresos fueron el 13-Febrero, 16-Marzo, 25-Abril, 13-Julio y el 1-Noviembre, con valores de \$15.479,05, \$9.228,32, \$8.263,57, \$7.708,67 y \$6.527,22 respectivamente. Estos valores se pueden visualizar mejor en el análisis que se hará al histograma de los ingresos por venta, donde se detallará la frecuencia de ocurrencia de determinados valores de ingresos; como por ejemplo en el caso de ingresos por venta de

\$15.479,05 posee una frecuencia de ocurrencia de uno, debido a que solo el 13 de Febrero ocurrió dicha transacción.

Además podemos observar que en la mayoría de los días existe ingresos inferiores o iguales a \$6.400, pero en muy determinados días existe ingresos por venta inferiores a los \$670, como es el caso del 14-Enero, 7-Abril, entre otros.

2.2.2 Análisis de los ingresos por venta, diarias en el año 2.006 por medio de histograma.

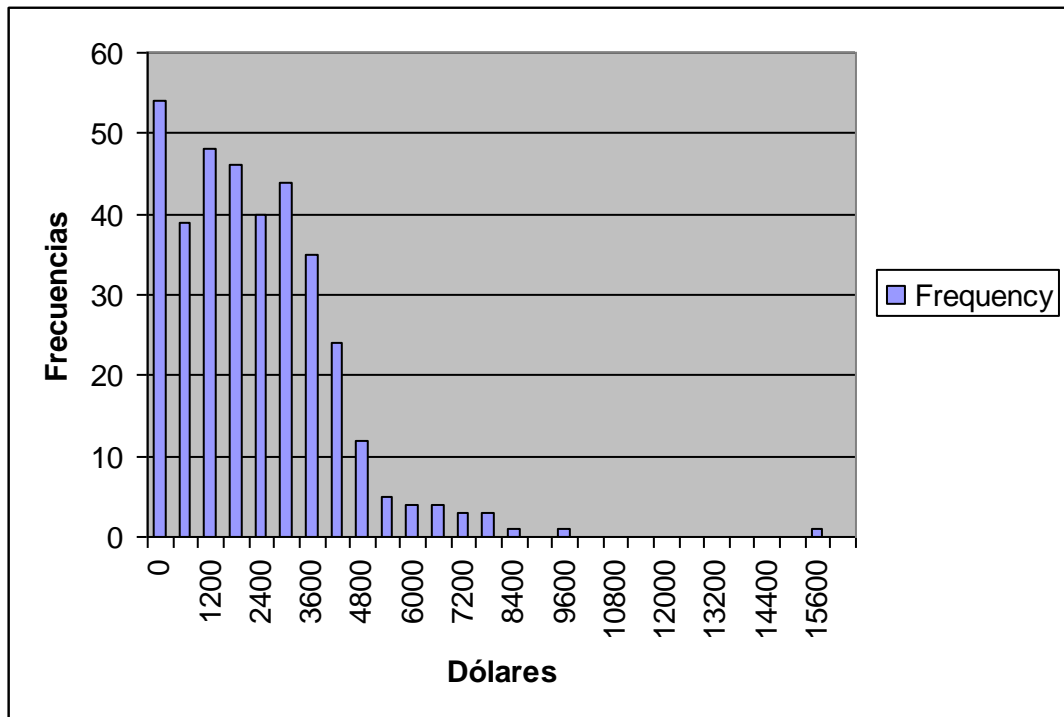


Gráfico 2.2 Histograma de los ingresos por venta diarios durante el año 2.006.

Como se puede apreciar en el gráfico 2.2, los ingresos diarios por venta podrían ser que se ajusten a una distribución exponencial, y no una distribución normal debido a la forma del gráfico.

La empresa tuvo ventas de cero dólares durante 54 de los 364 días laborables, es decir que el 14,84% de los ingresos fueron por ese valor.

Los ingresos por venta durante el año 2.006 tiene mayor concentración entre 0 y \$4.200; es decir que la empresa obtuvo durante el año 2.006 ingresos por venta inferiores o iguales a este valor; representando este rango el 90,66% de los ingresos.

Es verdad que ciertos ingresos por venta que la empresa tuvo durante el año 2.006 fueron superiores a \$4.200, esto es el caso de un ingreso por venta de \$15.479,07 que ocurrió el 13-Febrero; como se puede observar en el gráfico 2.1.2 este valor tuvo una frecuencia de ocurrencia del 0,27%. Además la frecuencia de que la empresa obtenga de ingresos por venta superiores a los \$9.000 es prácticamente de 0,00%, salvo el caso mencionado.

2.2.3 Análisis descriptivos de los ingresos por medio de diagramas de caja

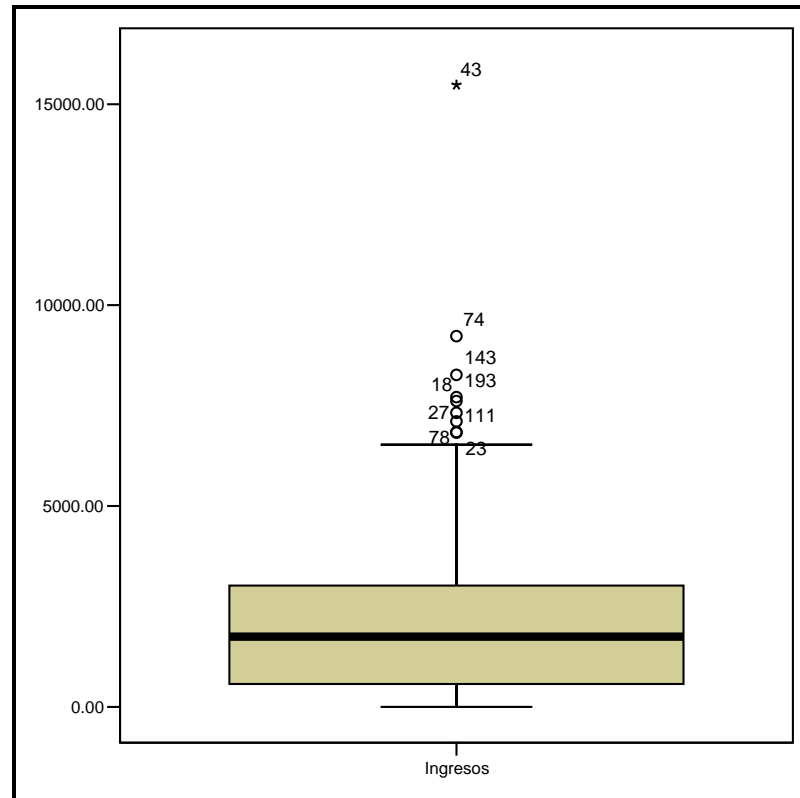


Gráfico 2.3 Diagrama de Caja de Ingresos por ventas para el año 2.006

Mediana		1.749,98
Cuartiles	25	570,56
	50	1.749,98
	75	3.022,17

Tabla II. Tabla de Cuartiles y mediana de los ingresos

	N	Mínimo	Máximo	Media	Desv. típ.
Ingresos	364	0,00	15.479,07	2.001,194	1.852,56

Tabla III. Estadísticos descriptivos diarios de los Ingresos por venta

El gráfico 2.3 y la tabla II nos ayuda a ver la concentración de los datos, mediante la utilización de los cuartiles.

El primer cuartil nos indica que el 25% de los ingresos por ventas son inferiores a \$570,56; es decir que el 75% de los ingresos son superiores a dicho valor.

El segundo cuartil nos indica que la mitad de las observaciones están por debajo de \$ 1.749,98.

El tercer cuartil en cambio nos indica que el 75% de los ingresos por venta son inferiores a \$3.022,17; es decir que el 25% de los ingresos son superiores a ese valor.

Además se puede observar en el gráfico 2.3 que existen valores de ingresos atípicos, es decir valores que se dieron en el transcurso del año pero su frecuencia es muy esporádica.

Mediante la tabla III podemos observar el promedio diario de ingresos por venta para el año 2.006, siendo de \$2.001,94. En cuanto a la desviación estándar podemos concluir que existe mucha dispersión entre los ingresos diarios por venta y la media obtenida.

Por otro lado tenemos los ingresos mínimos diarios que la empresa tuvo por un valor de \$0, el ingreso máximo que la empresa tuvo durante el año 2.006 por \$15.479,07.

	Error típ. de la media
Ingresos	97,10

95% Intervalo de confianza para la diferencia	
Inferior	Superior
1810,24	2192,14

Tabla IV. Intervalos de confianza para la media diaria de los ingresos por venta para el año 2.006.

Se puede decir que la empresa en promedio tiene ingresos por venta en un intervalo de \$1.810,24 a \$2.192,14 con un 95% de confianza. Además el error de la media obtenida es de 97,10, un valor no tan significativo.

2.3 Análisis de los Gastos diarios en el año 2.006.

2.3.1 Análisis de los Gastos diarios en el año 2.006 por medio de un gráfico de serie de tiempo.

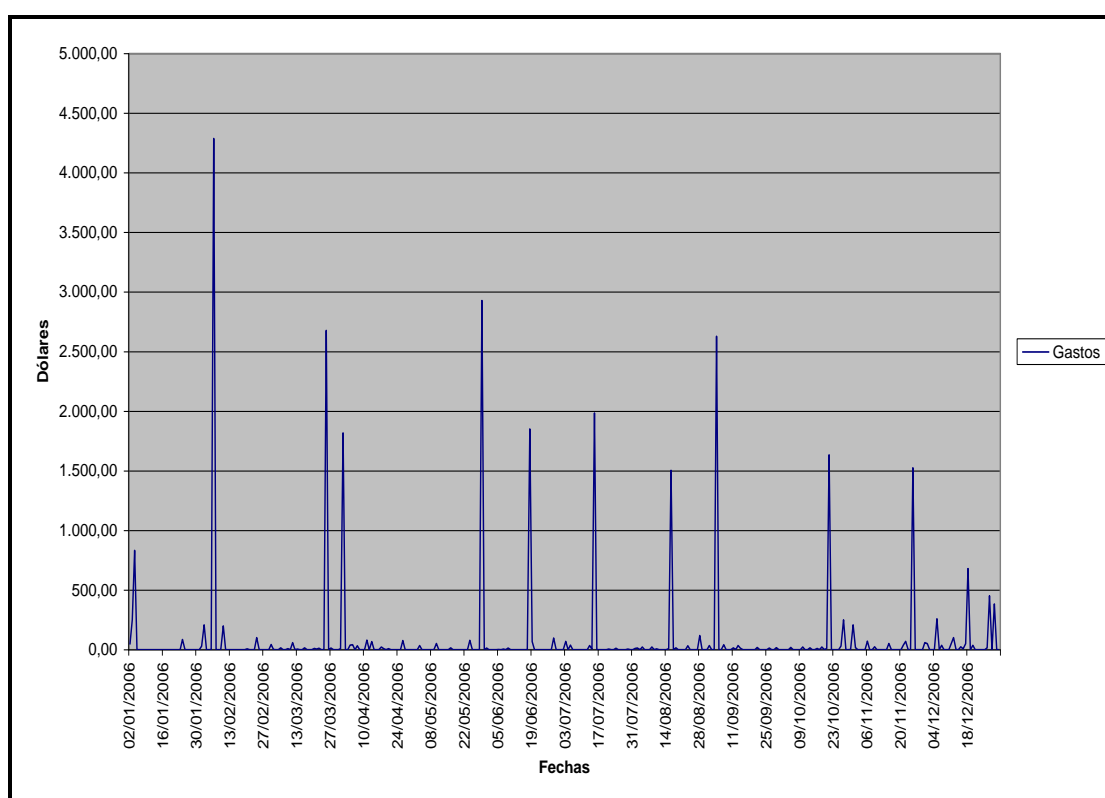


Gráfico 2.4. Gráfico de Serie de tiempo de los Gastos diarios para el año 2.006

Mediante el gráfico 2.4 podemos observar claramente los grandes saltos de los gastos, como por ejemplo el 6-Febrero, 25-Marzo, 29-Mayo y el 4-Septiembre con valor de \$4.287,33, \$2.677,77, \$2.929,22 y \$2.628,68 respectivamente.

Claro está que existen gastos diarios inferiores a \$1.818,31 y superiores a los \$1.000 pero con una frecuencia no tan alta, este punto lo podemos entender mejor cuando se tope el análisis del histograma de los gastos diarios. El gráfico 2.4 nos permite ver que en la mayoría de los días los gastos fueron inferiores a \$300, incluso hay días donde los gastos fueron de \$0.

2.3.2 Análisis de los Gastos diarias en el año 2.006 por medio de Histograma.

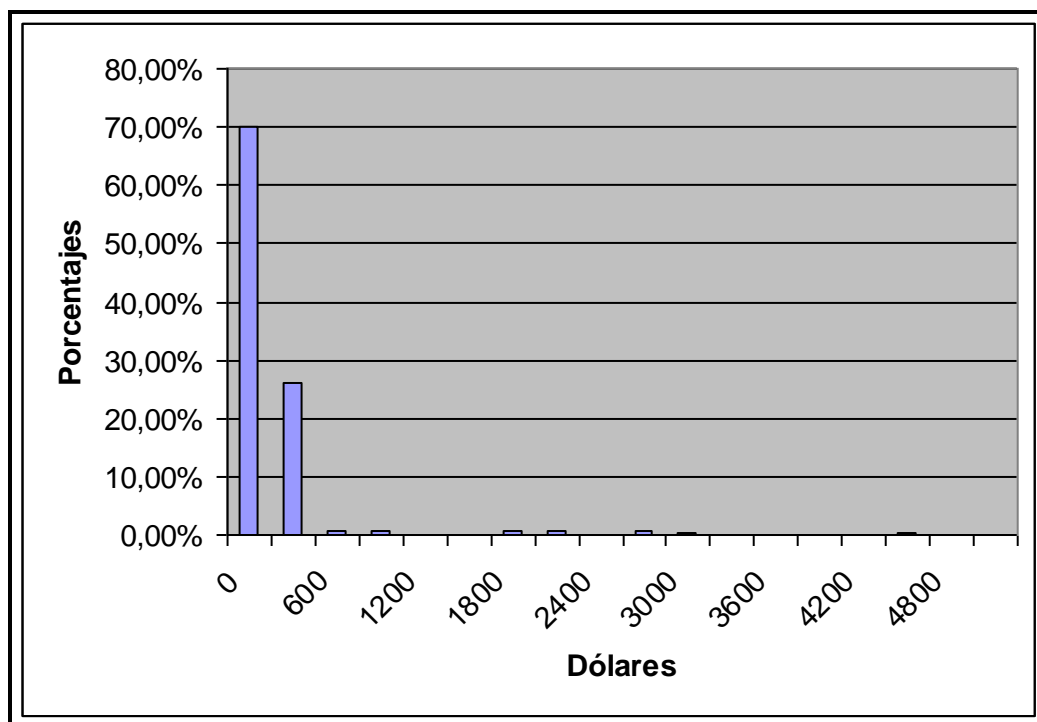


Gráfico 2.5. Histograma de los gastos diarios durante el año 2.006

El gráfico 2.5 nos muestra que los gastos podrían ser que se ajusten a una distribución exponencial, y no una distribución normal debido a la forma del gráfico.

El 70% de los gastos representan gastos nulos, es decir de \$0, y sólo el 25,97% representan gastos por un valor de \$300, es decir que el complemento 4,03% representan gastos por \$600, \$900, \$1.800, \$2.100, \$2.700, \$3.000 y \$4.500 siendo estos valores muy esporádicos en el ejercicio económico del año 2.006; pero tan solo el 1,38%

representan gastos entre \$1.000 y \$1.800. Por lo que se puede concluir que los gastos que mayormente incurrió esta empresa son valores menores o iguales a \$300, por lo que se esperaría una utilidad representativa para la empresa; este supuesto se lo afirmará o se lo negará cuando se analice el histograma de la utilidad/pérdida, así como las medidas descriptivas.

2.3.3 Análisis de los Gastos por medio de Diagramas de Caja

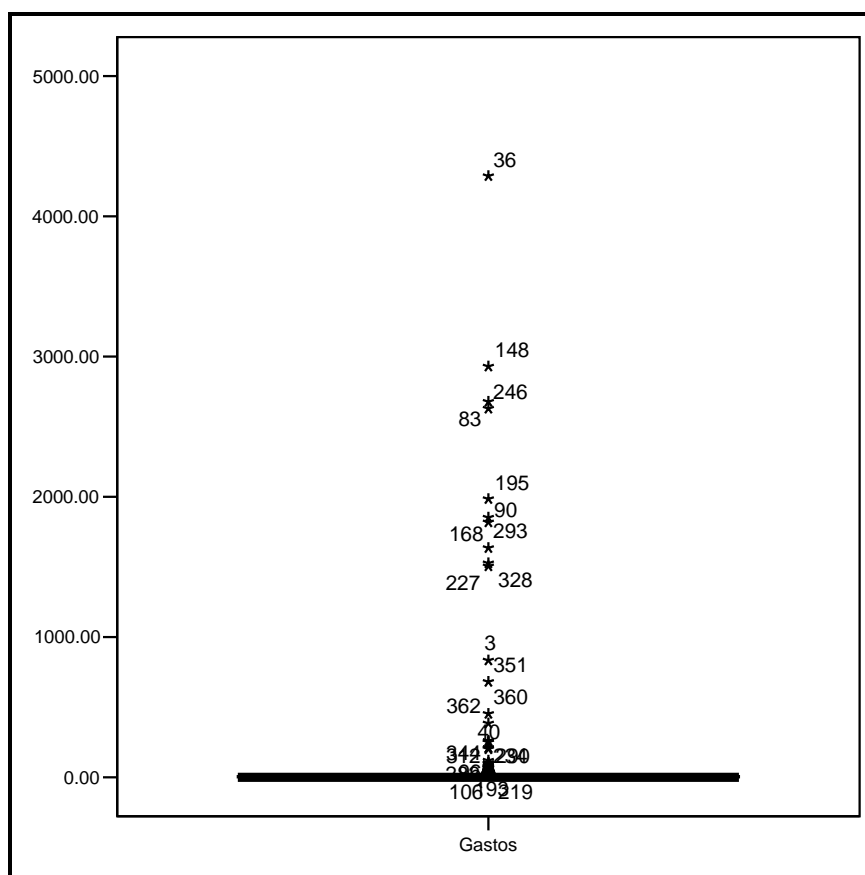


Gráfico 2.6. Diagrama de caja de los gastos para el año 2.006

Mediana		0,0
Cuartiles	25	0,0
	50	0,0
	75	8,04

Tabla V. Tabla de cuartiles y medianas de los gastos

	N	Mínimo	Máximo	Media	Desv. típ.
Gastos	364	0,00	4.287,23	80,85	402,89

Tabla VI Estadísticos descriptivos de los gastos diarios

Mediante el gráfico 2.6 y la tabla V podemos decir, que el primer cuartil nos indica que el 25% de los gastos son inferiores a \$0.

El segundo cuartil nos indica que la mitad de las observaciones están por debajo de \$0.

El tercer cuartil en cambio nos indica que el 75% de los gastos son inferiores a \$8; es decir que el 25% de los gastos tan solo son superiores a \$8.

Algo muy importante que se debe recalcar es la existencia de muchos valores de gastos atípicos, eso se debe en gran parte a que los gastos no inciden mucho en transacciones económicas de la empresa, habiendo incluso en la mayoría de los días gastos de \$0, gastos muy inferiores y esporádicamente gastos superiores a \$200.

Mediante la tabla VI podemos observar el promedio diario de los Gastos para el año 2.006 es de \$80,85.

En cuanto a la desviación estándar se puede concluir que existe mucha dispersión entre los datos diarios de los gastos con la media obtenida, por un valor de \$402,89, dicha dispersión la pudimos observar mejor en los gráficos 2.4 y 2.5. (Ver gráfico).

Los gastos mínimos que la empresa tuvo diariamente es de \$0 y un máximo de gastos de \$4.287,23, como se lo analizó en el gráfico 2.5 este valor tuvo una frecuencia de 0,28% durante el año 2.006.

	Error típ. de la media
Gastos	21,18

95% Intervalo de confianza para la diferencia	
Inferior	Superior
39,21	122,49

Tabla VII. Intervalos de confianza para la media diaria de los gastos para el año 2.006

Se puede decir que la empresa en promedio tiene gastos en un intervalo de \$39,21 a \$122,49 con un 95% de confianza. Además el error de la media de los gastos obtenida es de 21,18, un valor no tan significativo.

2.4. Análisis de los Costos de Venta diarios en el año 2.006.

2.4.1 Análisis de los Costos de Ventas diarias en el año 2.006 por medio de un gráfico de serie de tiempo.

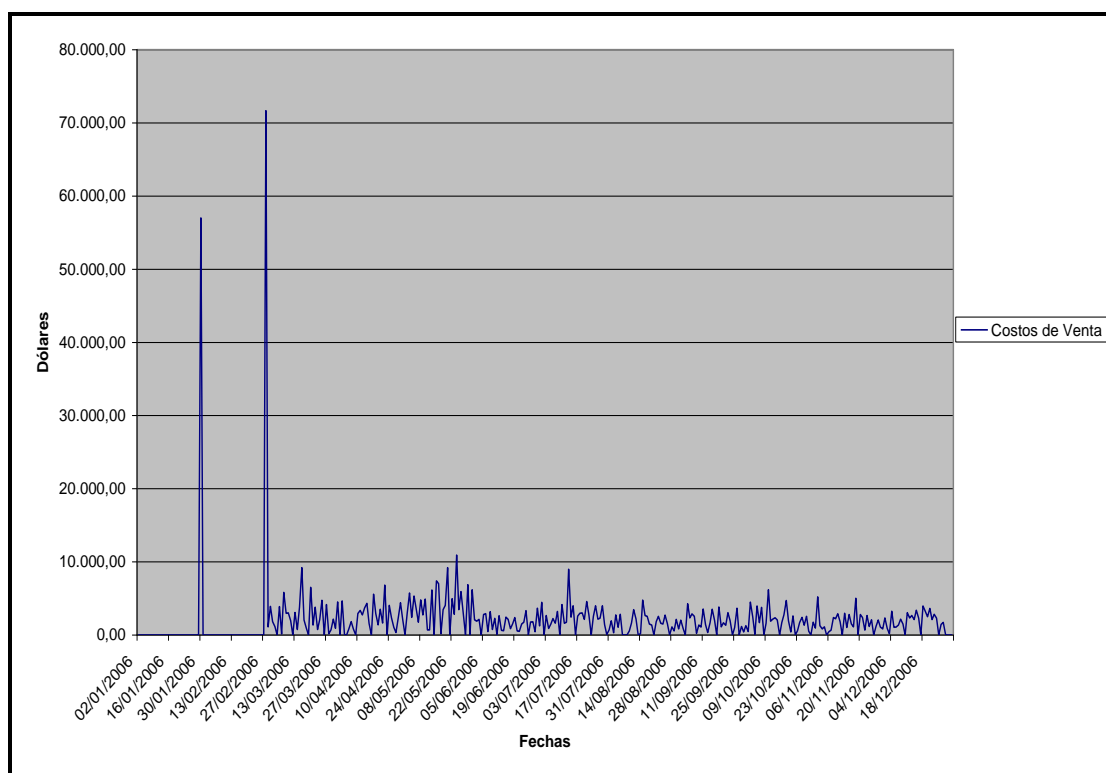


Gráfico 2.7. Gráfico de serie de tiempo de los costos de ventas diarios para el año 2.006

La gráfica 2.7 nos muestra que hubo mayores costos de venta el 30 de Enero y el 28 de Febrero del año 2.006, siendo sus valores de

\$57.013,25 y \$71.658,76 respectivamente, y por supuesto muy superiores a los ingresos donde el ingreso máximo fue de \$15.479,07.

Los costos de venta presentan muchos saltos en determinados meses, empiezan con costos de ventas de \$0 de los primeros días de Enero y luego aumentan enormemente en las fechas citadas en el primer párrafo. Aunque a partir del 8-Marzo los costos de venta siguen una distribución un poco más uniforme, igual existen días en que los costos de venta son de \$0, pero en la mayoría de los casos los costos de venta son muy superiores a los ingresos, obteniendo así una pérdida representativa para la empresa.

Estos registros de costos de venta de \$0 en los dos primeros meses y los registros de costos de ventas de forma diaria en los meses siguientes, incumple el principio de contabilidad de consistencia, ya que el procedimiento de cuantificación de los costos de ventas no ha permanecido en el tiempo.

Además al tener costos de venta superiores a los ingresos por venta, podemos suponer que se están dando descuentos muy especiales a los clientes, teniendo un margen bruto negativo. Otra hipótesis es que se está registrando como costo de venta salidas de inventarios que no necesariamente fueron originadas por las ventas, como por ejemplo robo de mercadería. Esta situación afecta la confiabilidad de la información

contable, presentando información distorsionada, y evitando la buena toma de decisiones financieras, basadas en ella.

2.4.2 Análisis de los Costos de Venta diarias en el año 2.006 por medio de un Histograma.

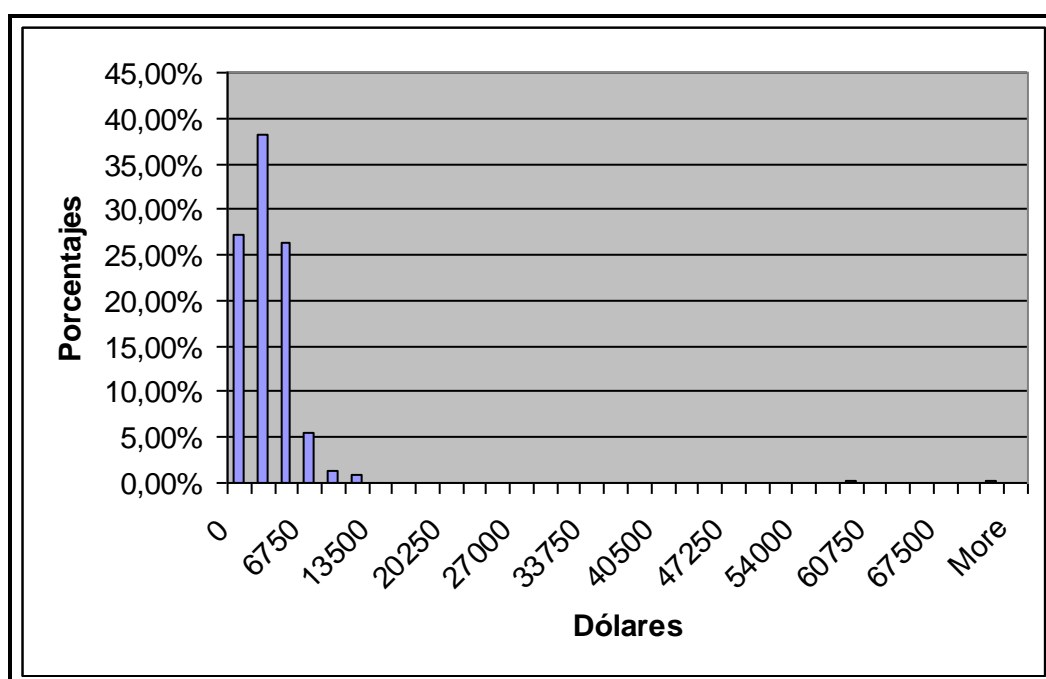


Gráfico 2.8. Histograma de los costos de venta para el año 2.006

Podemos observar que los costos de venta podrían ser que se ajusten a una distribución exponencial, y no una distribución normal debido a la forma del gráfico.

El 27,20% de los datos, representan costos de venta de \$0. Los costos de venta tienen mayor concentración en valores inferiores e iguales a \$4.500 durante el año 2.006.

Como valores muy esporádicos, podemos mencionar costos de venta de \$6.750, \$9.000, \$11.250, \$58.500, \$72.000 que representan el 7,97% de los datos.

En el año 2006, hubo también costos de venta de \$0 con una frecuencia del 27,20%. El rango de los costos de venta que mayormente se dio en el año 2.006 fue inferior a \$4.500 con una frecuencia de \$91,76%.

2.4.3 Análisis de los Costos de venta diarios en el año 2.06 por medio de Diagrama de caja

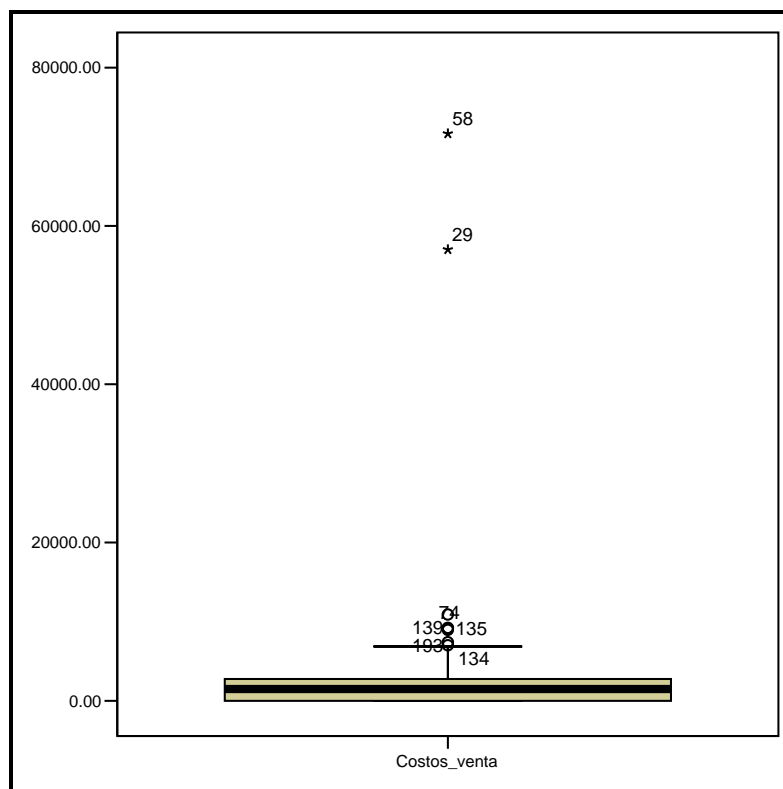


Gráfico 2.9. Diagrama de caja de los costos de venta para el año 2.006

Mediana		1.494,71
Cuartiles	25	0,0
	50	1.494,71
	75	2.755,74

Tabla VIII. Tabla de Cuartiles y mediana para los costos de venta.

	N	Mínimo	Máximo	Media	Desv. típ.
Costos Venta	364	0,00	71.658,76	2.117,23	5.012,14

Tabla IX. Estadísticos descriptivos diarios de los costos de venta

Mediante el gráfico 2.9 y la tabla VIII podemos decir que el 75% de los costos de venta son superiores \$0, este valor lo deducimos por medio del primer cuartil.

El segundo cuartil nos indica que la mitad de los costos de venta están por debajo \$ 1.494,71.

El tercer cuartil en cambio nos indica que el 75% de costos de venta son inferiores a \$2.755,74; es decir que el 25% de los costos de venta son superiores a dicho valor.

Además podemos observar que existen ciertos datos atípico pero no son tantos como en el caso de los gastos.

Mediante la tabla IX podemos observar que el promedio diario del Costo de Ventas es de \$2.177,23; en cuanto a la desviación estándar podemos concluir que existe mucha dispersión entre los datos diarios de los costos de venta con su media. Vemos que la dispersión es de \$5.012,14,

siendo los costos de ventas los que tiene mayor variabilidad que los ingresos y los gastos.

Por otro lado que los costos de venta diarios mínimos para el año 2.006 fue de \$0 igual que en las otras variables y un valor máximo de \$71.658,76.

	Error típ. de la media
Costos_venta	262,71

95% Intervalo de confianza para la diferencia	
Inferior	Superior
1.600,61	2.633,84

Tabla X. Intervalo de confianza para la media diaria de los costos de venta para el año 2.006

Se puede decir que la empresa en promedio tiene costos de venta en un intervalo de \$1.600,61 a \$2.633,84 con un 95% de confianza. Además el error de la media de los costos de venta obtenida es de 262,71, siendo un valor muy significativo y el mayor error obtenido de las variables de estudio.

2.5 Análisis de las Utilidades/Pérdidas diarias en el año 2.006.

2.5.1. Análisis de las Utilidades/Pérdidas diarias en el año 2.006 por medio de un gráfico de serie de tiempo.

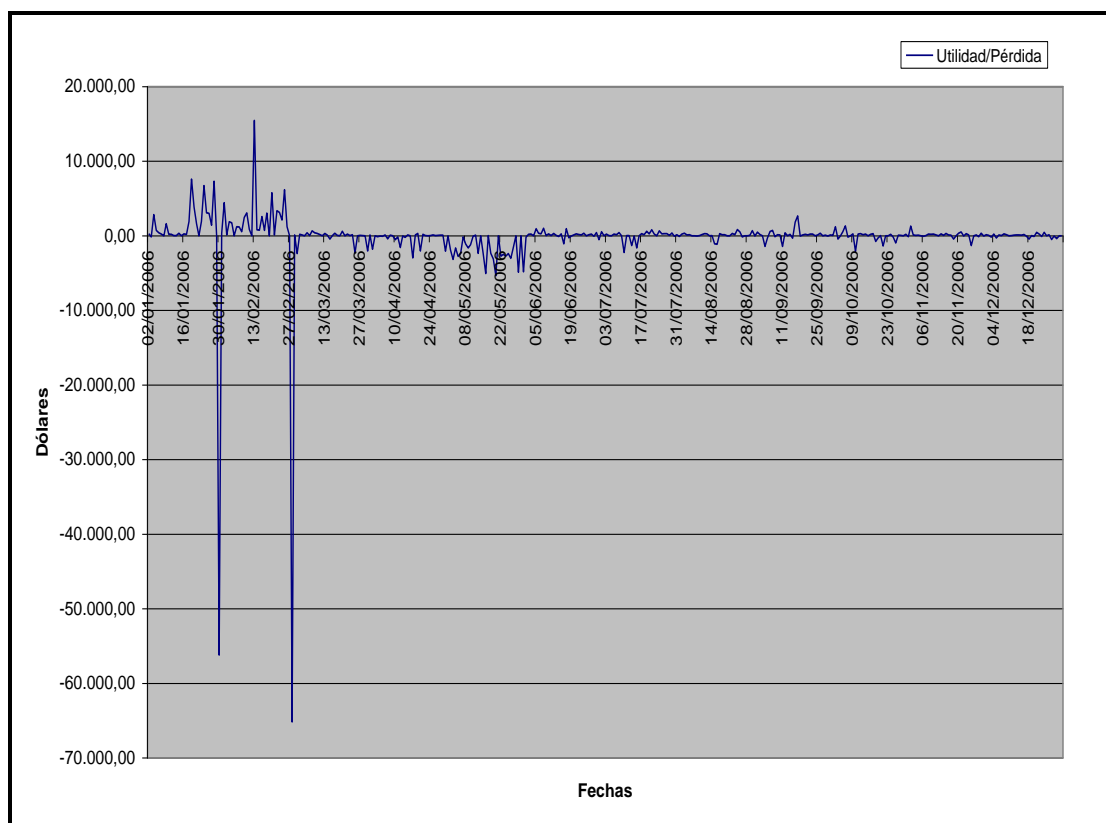


Gráfico 2.10. Gráfica Serie de tiempo de las utilidades/pérdidas diarias durante el año 2006

El gráfico 2.10 nos muestra las fluctuaciones de la utilidad/pérdida durante el año 2.006. Podemos apreciar que la empresa para el día 30 de

Enero presenta una pérdida muy significativa de \$ - 56.224,11; el mismo caso podemos apreciar para el día 28 de Febrero, con una pérdida de \$ - 65.192, 68, siendo estas pérdidas las más significativas durante el año 2.006.

En cuanto a las utilidades que la empresa obtuvo durante el año 2.006, se puede considerar como importantes las obtenidas en los días 19, 24 y 28 de Enero con un valor de \$7.612,18, \$6.739,69 y \$7.320,27 respectivamente.

Como se puede apreciar en el gráfico 2.10, los mejores meses donde la empresa obtuvo utilidad fueron desde Junio hasta Septiembre del 2.006 con una utilidad de \$ 12.839,07 acumulados. La misma situación se repite para los meses de Noviembre y Diciembre con una utilidad de \$3.767,68.

A partir del 5 de Marzo del 2.006, las utilidades/pérdidas se armonizan, es decir que no se aprecia en el gráfico 2.10 picos ya sea de utilidades o de pérdidas.

2.5.2 Análisis de las Utilidades/Pérdidas diarias en el año 2.006 por medio de un Histograma

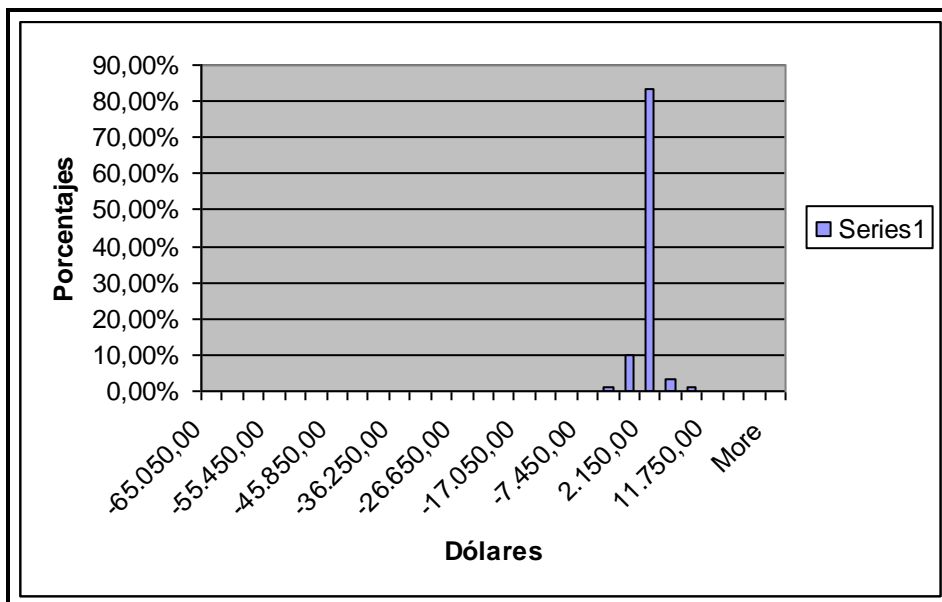


Gráfico 2.11. Histograma de la Utilidad/Pérdida para el año 2.006

En el gráfico 2.10 se puede observar que la utilidad podría ajustarse a una distribución normal. El 1,10% de los datos se concentran en pérdidas inferiores a \$ - 4.250, mientras que el 10,16% se concentra en pérdidas inferiores a \$ - 1.050.

El 83,24% de los datos se concentra en utilidades de \$2.150, es decir que hubo mayor probabilidad de que se dé este rango como utilidad.

Mientras que utilidades superiores a \$5.350 posee una frecuencia de 4,67% de ocurrencia durante el año 2.006.

2.5.3 Análisis de las Utilidades/Pérdidas por medio de Diagrama de Caja

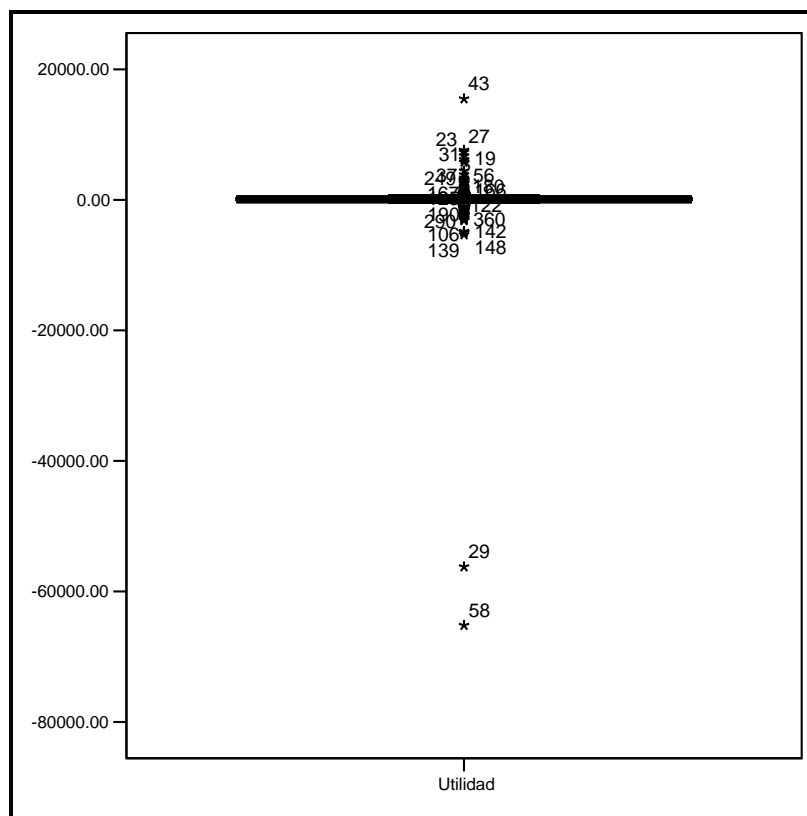


Gráfico 2.11 Diagrama de Caja de las utilidades/pérdidas diarias para el año 2.006

Mediana		79,94
Cuartiles	25	-20,15
	50	79,94
	75	270,83

Tabla XI. Tabla de Cuartiles y Mediana de la utilidad/pérdida

	N	Mínimo	Máximo	Media	Desv. típ.
Utilidad	364	-65.192,68	15.479,07	-196,44	4.784,84

Tabla XII. Estadísticos descriptivos de la utilidad/pérdida diaria para el año 2.006

Mediante el gráfico 2.11 y la tabla XI podemos decir que el 75% de las pérdidas son superiores \$-20,15, este valor lo deducimos por medio del primer cuartil.

El segundo cuartil nos indica que la mitad de los datos diarios de la utilidad están por debajo \$ 79,94.

El tercer cuartil en cambio nos indica que el 75% de las utilidades diarias son inferiores a \$270,83; es decir que el 25% de las utilidades son superiores a dicho valor.

Además podemos observar que existen una gran cantidad de datos atípicos y esto lo podemos corroborar mediante el análisis de la variabilidad de los datos (desviación estándar).

Mediante la tabla XII podemos observar el promedio diario de pérdida en la empresa para el año 2.006 fue de \$ - 196,44; en cuanto a la desviación estándar podemos concluir que existe mucha dispersión entre los datos

diarios de las utilidades/pérdidas con su media. Vemos que la dispersión es de \$4.784,84 siendo casi tan disperso como en el caso de los costos de venta.

Por otro lado que las pérdidas diarios mínimos para el año 2.006 fue de \$ - 65.192,68 y un valor máximo de utilidad de \$15.479,07.

	Error típ. de la media
Utilidad	250,79

95% Intervalo de confianza para la diferencia	
Inferior	Superior
-689,63	296,76

Tabla XIII. Intervalo de confianza para la media diaria de la utilidad/pérdida para el año 2.006

Se puede decir que la empresa en promedio tiene utilidad/pérdidas en un intervalo de \$ - 689,63 a \$296,76 con un 95% de confianza. Además el error de la media de la pérdida obtenida es de 250,79, siendo un valor muy significativo, al igual que en el caso de los costos de venta.

CAPITULO 3.

3. MODELO DE REGRESIÓN MÚLTIPLE PARA LA DETERMINACIÓN DE LA UTILIDAD.

3.1 Introducción

La regresión lineal y la correlación son las dos herramientas estadísticas más poderosas y versátiles que se puede utilizar para solucionar problemas, por este motivo la utilizaremos en esta tesis.

Para plantear el modelo de regresión lineal múltiple para proyectar la utilidad de la empresa distribuidora de celulares y accesorios, es necesario definir la variable dependiente y las variables independientes.

Debido a que la utilidad depende de los ingresos, gastos y costos de venta; se ha definido a la *Utilidad* como la *Variable Dependiente* y a los *Ingresos, Gastos y Costos de Venta* como las *Variables Independientes*, quedando el modelo planteado de la siguiente manera:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$$

Donde :

$\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_3 =$ Constantes.

$X_1 =$ Variable Ingresos

$X_2 =$ Variable Gastos

$X_3 =$ Variable Costos de Ventas

$\hat{Y} =$ Variable Utilidad Estimada.

3.2 Modelo de Regresión Lineal Múltiple para la determinación de la Utilidad Diaria

Debido a que tanto los costos de venta como los gastos originados diariamente son valores de \$0 como se indicó en el análisis de los datos en el Capítulo II, no se pudo realizar el modelo de regresión lineal múltiple para la determinación de la utilidad diaria, debido a que el programa estadístico utilizado SPSS considera estos datos como constantes impidiendo implementar el modelo, debido a que ciertos factores necesarios para validarlo no aparecen en las tablas que son utilizadas para aceptar o rechazar el modelo. Por este motivo se ha considerado implementar un modelo de regresión lineal múltiple para la determinación de la utilidad pero de manera mensual.

También podemos corroborar los resultados obtenidos en SPSS para el desarrollo del modelo de regresión lineal múltiple de la utilidad diaria

mediante el gráfico 3.1 de dispersión entre las variables de estudios, llegando a la conclusión de que no existe una relación entre las variables. A continuación se muestra el gráfico de dispersión:

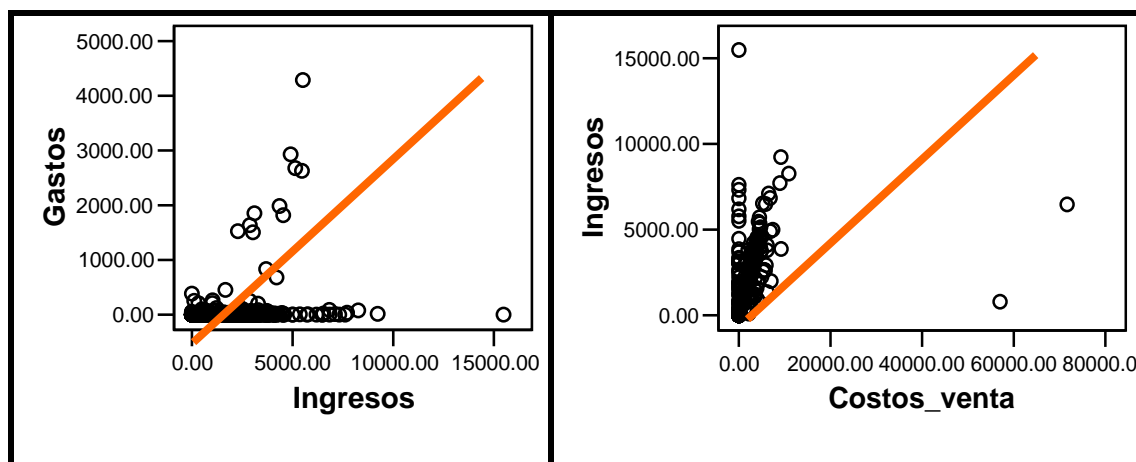


Gráfico 3.1 Gráfico de Dispersión entre las variables de estudio (Diaria)

Luego de ver que no existe una relación entre las variables de forma diaria, a continuación se propone un modelo de regresión de la utilidad de manera mensual.

3.3 Modelo de Regresión Múltiple para la determinación de la Utilidad Mensual.

3.3.1 Análisis de Diagramas de dispersión de las variables de estudio

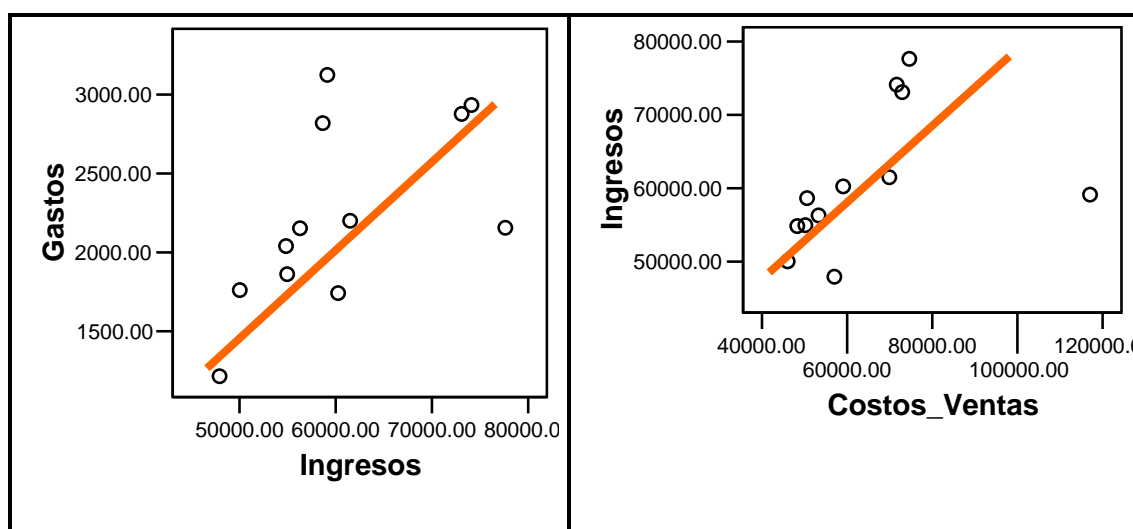


Gráfico 3.2. Diagrama de dispersión de las variables de estudio de (Mensual)

Como se puede apreciar en la gráfico 3.2. las variables de estudio llegan a tener una relación lineal. Este caso no se lo pudo apreciar en el gráfico 3.1 donde su diagrama de dispersión nos mostraba que no existía ninguna relación entre las variables.

Por tal motivo se ha llegado a proponer un modelo de regresión lineal múltiple de manera mensual, que a continuación se detallará.

3.3.2 Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Durbin-Watson
1	1,00	,999	,999	575,26892	2,277

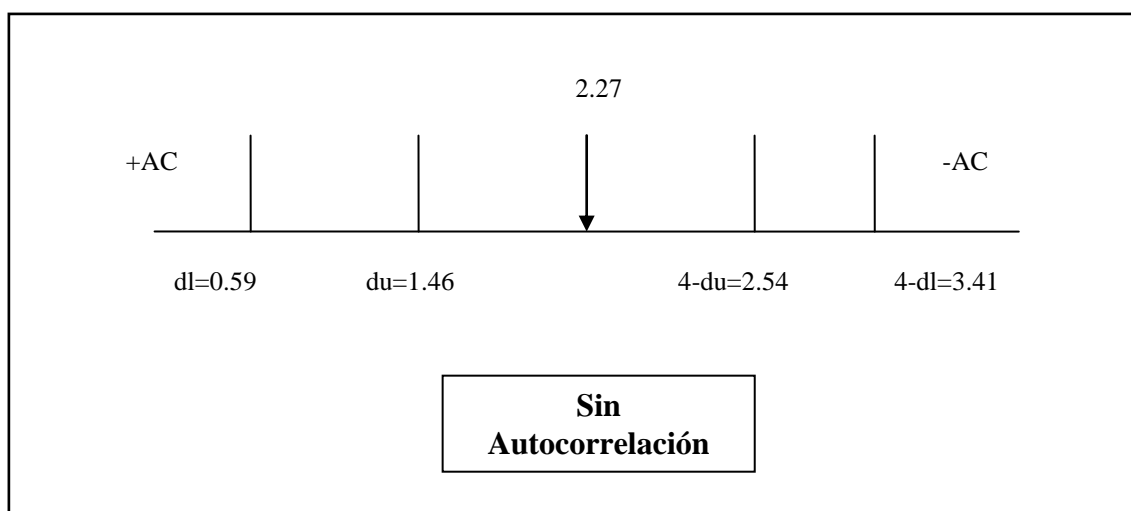
Tabla XIV. Resumen del Modelo planteado.

3.3.2.1 Autocorrelación

Los diagramas residuales nunca son tan obvios o tan fáciles de leer, por lo que se ha calculado mediante la Prueba de Durbin-Watson un valor para detectar una autocorrelación. Para esto he supuesto dos hipótesis para probar si existe o no correlación entre términos de errores sucesivos:

$$H_o = \rho_{e_i, e_{i-1}} = 0 \quad (\text{No existe autocorrelación})$$

$$H_A = \rho_{e_i, e_{i-1}} \neq 0 \quad (\text{Existe autocorrelación})$$



Dibujo 3. Prueba de Durbin-Watson

En este caso como Durbin-Watson está entre d_u y $4-d_u$, no se rechaza la hipótesis nula, por lo tanto concluyo que la correlación no existe. Este tema de autocorrelación lo podemos clarificar cuando se analice la matriz de covarianza y varianza.

3.3.3 Análisis de Correlación

3.3.3.1 Coeficiente de Correlación

Para el caso de estudio el coeficiente de correlación obtenido en la Tabla 3.1, salió +1, lo que nos indica una relación positiva perfecta entre la variable dependiente (Utilidad) y las variables independientes (Ventas, Gastos y Costo de Venta).

3.3.3.2 Coeficiente de Determinación

Para el caso de estudio, se obtuvo un coeficiente de determinación R^2 de 0.99, lo que nos indica que el modelo se encuentra bien planteado, ya que se considera que un R^2 superior al 75% es aceptable.

3.3.4 Anova

	Suma de cuadrados	gl	Media cuadrática	F	F_α
Regresión	3.594.253.606,29	3	1.198.084.535,43	3.620,31	4,07
Residual	2.647.474,61	8	330.934,33		
Total	3.596.901.080,90	11			

Tabla XV. ANOVA

Mediante el ANOVA vamos a probar las siguientes hipótesis:

$$H_0 : \beta_0 = \beta_1 = \beta_3 = \beta_4 = 0$$

$$H_1 : \exists i, \beta_i \neq 0$$

En este caso el estadístico de prueba $F_{\text{obs}}=3620,309$ y $F_\alpha(3,8)$ donde $\alpha=0,05$ es igual a 4,07. Para rechazar la hipótesis nula, el estadístico de prueba debe ser mayor a F_α . En este caso de estudio se rechaza la

hipótesis nula, es decir que al menos una de las variables X_i contribuye significativamente al modelo.

Al nivel del 5% parece existir una relación entre ventas, gastos, costo de ventas con las utilidades de la empresa, por lo que se puede decir que el modelo planteado cumpliría las futuras expectativas de la empresa para el año 2007.

3.3.5 Coeficientes.

	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
	B	Error típ.	Beta		
Cte.	1.298,16	1.128,81		1,15	0,28
Ventas	0,976	0,023	0,512	43,31	0,0
Gastos	-1,166	0,422	-0,038	-2,76	0,03
Costos_Ventas	-0,995	0,011	-1,079	-88,79	0,0

Tabla XVI. Coeficientes

En la Tabla XVI se puede observar las estimaciones de los $\beta_0, \beta_1, \beta_2, \beta_3$ con su respectivos errores estándares.

El modelo de regresión lineal múltiple, está dado por la siguiente ecuación:

$$Y = 1298,156 + 0,976x_1 - 1,166x_2 - 0,995x_3$$

Los signos negativos para las variables: gastos y costos de ventas, son correctos, ya que como sabemos, estas dos variables disminuyen los ingresos, y consecuentemente se obtendrá la utilidad esperada.

De acuerdo a la tabla XVI nos podemos dar cuenta que β_0 no es significativa para el modelo, ya que con 8 grado de libertad y con un alfa del 5%, no se rechaza la hipótesis nula que es que $\beta_0 = 0$, debido a que $t_{8,0.05} = 2,306$ y este valor es mayor a 1. Cabe recalcar que podemos confirmar lo expuesto en el párrafo anterior, por medio del valor p, ya que como se muestra en la tabla 3.2.5 el valor p de β_0 es 0,28 y este valor es mayor que el valor alfa 0,05, por lo tanto se concluye también que no se rechaza la hipótesis de que β_0 no es significativo para el modelo, es decir que es igual a cero. Pero para nuestro vamos a considerar el valor de β_0 en nuestra ecuación.

3.3.6 Correlaciones de los Coeficientes

	Variables	Costos_Ventas	Ventas	Gastos
Correlaciones	Costos_Ventas	1,000	-,090	-,499
	Ventas	-,090	1,000	-,456
	Gastos	-,499	-,456	1,000
Covarianzas	Costos_Ventas	,000	,000	-,002
	Ventas	,000	,001	-,004
	Gastos	-,002	-,004	,178

Tabla XVII. Correlación de Coeficientes.

Mediante la Tabla XVII, podemos decir que la variable Costo de Venta y la variable Ventas poseen una relación inversa, con un coeficiente de correlación de -0,090. Lo mismo podemos concluir de la variable Costos de Ventas con la variable Gastos, con un coeficiente de correlación de -0,499.

Ahora, entre las variables ventas y gastos en relación con los costos de ventas, la variable Gastos tiene una relación lineal más fuerte con los costos de ventas ya que el valor de -0,499 es más cercano a -1.

La variable Ventas posee una relación inversa también con variable Gastos, con un coeficiente de correlación de -0,456, además entre

estas 2 variables existe un relación lineal ya que el valor de correlación está cerca de -0,456.

Claro está que, las variables que poseen una relación lineal fuerte son los costos de ventas con los gastos.

Capítulo 4

1. CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES

1.1. Mediante el trabajo realizado se pudo obtener una ecuación que nos permitirá obtener la utilidad/pérdida de la empresa para el año 2007.

$$Y = 1.298,156 + 0,976x_1 - 1,166x_2 - 0,995x_3$$

Este modelo se basó en ecuaciones matemáticas, relacionando una variable dependiente (Utilidad) con tres variables independientes (Ventas, Gastos y Costo de Venta).

1.2. No se puede decir que el modelo de regresión lineal múltiple para proyectar la utilidad/pérdida que se obtuvo, reflejaría una información financiera confiable, debido a que el registro de los costos de venta no siguió uno de los principios de contabilidad que es la consistencia de la información, es decir que no siguió un procedimiento de cuantificación que permanezca en el tiempo, debido a que en los dos primeros meses

se registró el costo de venta a fin de mes y a partir del tercer mes se registró de manera diaria, pero en ciertos casos se observó que no se seguía este procedimiento de forma secuencial.

1.3. En la mayoría de los días del año 2.006 se observó que los costos de venta son muy superiores a los ingresos, dando como resultado un margen bruto negativo para la empresa, por este motivo se puede concluir que la empresa tal vez dio descuentos muy especiales a sus clientes o que se está registrando en la cuenta costos de venta, salidas de inventarios que no necesariamente se hayan dado por la venta. Esta conclusión no pudo ser constatada físicamente por la falta de información de la empresa, tan solo se la puede considerar como una posible hipótesis para poder explicar el motivo de que los costos de venta son muy superior a los ingresos.

1.4. La pérdida operacional que se espera para la empresa para el año 2.007 es de **(\$155.459,21)** según los datos supuestos de ingresos, gastos y costos de venta, incrementando en un 117,42% las pérdidas en relación al ejercicio económico anterior. El desglose de las utilidades/pérdidas mensuales durante el año 2.007 se lo puede observar en la sección de Apéndice, apéndice A.

1.5. Del modelo planteado para proyectar las utilidades, se pudo concluir que todas las variables que integran el modelo son significativas, esto se pudo comprobar al realizar el análisis del ANOVA, donde se rechazó la hipótesis nula, a favor de que las constantes que acompañan a las variables independientes son distintas de cero.

1.6. Gracias al análisis de la prueba de Durbin-Watson, se pudo concluir que no existe autocorrelación.

1.7. Mediante la Tabla XVII, podemos decir que la variable Costo de Venta y la variable Ventas poseen una relación inversa, lo mismo podemos concluir de la variable Costos de Ventas con la variable Gastos. La variable Gastos tiene una relación lineal más fuerte con los costos de ventas. La variable Ventas posee una relación inversa también con variable Gastos, existiendo entre estas dos variables una relación lineal.

1.8. Por medio de los análisis descriptivos de los datos de la empresa se pudo concluir que los ingresos, gastos y costos de venta se comportaron en el año 2.006 con una distribución que podrían ser modelo de una

exponencial diaria y salvo el caso de la utilidad que podría tener una distribución normal.

1.9. Por medio de los gráficos de series de tiempo de las variables de estudio se pudo observar mucha variabilidad de los datos diarios, debido a que no poseen un patrón de las fluctuaciones ya sea de ingresos como de egresos. Esta conclusión la pudimos confirmar mediante las desviaciones estándares y los diagramas de caja donde se pudo observar datos atípicos.

1.10. Se puede decir que la media diaria de los ingresos que recibió la empresa durante el año 2.006 fue de \$2.001,194; la media de los gastos diarios fue de \$80,85; la media de los costos de venta diarios fue de \$2.117,23 y la media de la utilidad/pérdida de la empresa diaria fue de \$ -196,44.

RECOMENDACIONES

2.1 Se sugiere a la empresa replantear sus metas y su gestión para obtener mayores ingresos, debido a que con los datos supuestos de ingresos, gastos y costos de venta utilizados en la ecuación de proyección de la utilidad/pérdida se concluyó que al finalizar el año 2.007 ésta obtendría una pérdida considerable. Adicionalmente, se debe tener cuidado de no utilizar el modelo de regresión para predecir valores de que estén fuera del rango del conjunto original de datos.

2.2 Se recomienda a la empresa seguir un mismo procedimiento de cuantificación para el caso de los costos de venta que permanezca en el tiempo, debido a que el efecto de dicho cambio produce cifras contables no confiables, sin embargo si se desea realizar cualquier cambio en el registro se recomienda dar a conocer la situación para evitar tener errores.

2.3 Un análisis descriptivo de sus variables de interés del negocio, les ayudaría bastante a la empresa a mejorar su rentabilidad, por lo que se recomienda realizar un estudio parecido al que se está presentando.

2.4 Se recomienda tomar en cuenta que un análisis de regresión no puede determinar relaciones causa-efecto aunque con frecuencia se ha demostrado ser de utilidad en la toma de decisiones para una gran variedad de negocios y de asuntos económicos, la correlación no implica causalidad.

2.5 No se ha abarcado la correlación espúrea, que es la correlación que ocurre simplemente por suerte. No hay sustituto para el sentido común en el análisis de correlación ni en el de regresión.

BIBLIOGRAFIA

1. Jonson Richard A, 2002, Probabilidad y Estadística para Ingenieros de Millar y Freud, Pearson Educación, Quinta edición, páginas 56-96.
2. Webster Allen L, 1998, Estadística aplicada a los negocios y la economía, Irwin McGraw-Hill, Tercera edición, páginas 21-400.
3. Instituto Mexicano de Contadores Públicos, 1986, Principios de Contabilidad Generalmente Aceptados, páginas 9-15.