

UNIVERSIDAD SAN FRANCISCO DE QUITO

**Modelos de optimización para las ventas de Novartis Ecuador
S.A. basados en algoritmos y análisis de regresión y simulación
Montecarlo.**

Michielle Zurita

Tesis de grado como requisito para la obtención de l título de Ingeniero Industrial

Quito Mayo 2010

© Derechos de autor
Michielle Zurita Vaca
2010

AGRADECIMIENTO

A todos quienes me enseñaron que el éxito no viene por si solo, sino que es un conjunto de metas, preparación, dedicación y mucho coraje. A Dios, mis padres, hermanos, amigos y profesores:

¡Muchas gracias!

RESUMEN

El presente trabajo de tesis presenta un estudio a la venta de medicinas de la farmacéutica Novartis Ecuador. Se analiza el comportamiento de cuatro variables: Fuerza de Ventas, Medios de difusión (Advertisement and Promoting), Descuentos y Muestras Médicas, las cuales inciden directamente en la compra del consumidor final. Por medio de una inferencia en sus distribuciones de probabilidad se realizó una simulación Montecarlo con el fin de generar datos aleatorios que simulen el comportamiento de las variables a través del tiempo. A partir de éstos, se estableció un modelo de regresión múltiple, con el cual se establezca la relación existente de estas variables con la función de rentabilidad generada por las ventas de algún producto. Finalmente, y mediante un modelo matemático se establecieron niveles óptimos de cada una de las variables, para distintos productos de la farmacéutica, con el fin de maximizar las ganancias obtenidas por la comercialización de las medicinas y a su vez generar ahorros sustanciales en la empresa.

ABSTRACT

This thesis presents a study on the sale of medicines from Novartis Ecuador. It comprehends an analysis of the behavior of four variables; Sales Force, Media (Advertisement and Promoting), Discounts, and Medical Samples, which directly affect the final consumer purchase. By an inference of probability distributions a Monte Carlo Simulation was performed to generate random data to simulate the behavior of variables over time. Since then, it was established a multiple regression model, with which it was determined the relationship of these variables with the function of return on investment, generated by sales of a product. Finally, using a mathematical model there were established optimal levels for each of the variables for various pharmaceutical products, in order to maximize profits from the sale of medicines and in turn, generate substantial savings for the company.

TABLA DE CONTENIDO

1 Capítulo 1- Introducción.....	1
1.1 Objetivo general.....	2
1.2 Objetivos específicos.....	2
1.3 Justificación del proyecto.....	2
1.4 Metas.....	3
1.5 Organización del documento.....	3
1.6 Descripción de la empresa Novartis Ecuador.....	4
1.6.1 Historia de la empresa.....	4
1.6.2 El negocio.....	6
1.6.3 Productos.....	6
2 Capítulo 2: Marco teórico.....	10
2.1 Análisis de regresión.....	10
2.1.1 Método de los mínimos cuadrados para la estimación de parámetros.....	11
2.1.2 Enfoque matricial a la regresión lineal múltiple.....	12
2.1.3 Propiedades de los estimadores de mínimos cuadrados y estimación de σ	14
2.1.4 Supuestos detrás del método de mínimos cuadrados.....	15
2.1.5 Pruebas de hipótesis en la regresión lineal múltiple.....	16
2.1.6 Intervalos de confianza para la regresión lineal múltiple.....	17
2.1.7 Medidas de la adecuación del modelo.....	18
2.1.8 Variables indicadoras.....	21
2.1.9 Selección de variables en regresión múltiple.....	22
2.2 Simulación Montecarlo.....	23
3 Capítulo 3: Modelo de regresión múltiple.....	26
3.1 Definición de variables y productos.....	26
3.2 Inferencia en las distribuciones de cada variable.....	29
3.3 Modelos de regresión.....	32
3.4 Análisis en productos.....	35
3.4.1 Producto VOLTAREN.....	35
3.4.1.1 Inferencia para la variable de salida Ventas de Voltaren.....	36
3.4.1.1.1 Inferencia para los datos generados de la variable de salida Ventas de VOLTAREN.....	38
3.4.1.2 Inferencia para la variable de entrada Advertisement & Promoting de VOLTAREN.....	39
3.4.1.2.1 Inferencia para los datos generados de la variable de entrada Advertisement & Promoting de VOLTAREN.....	41
3.4.1.3 Inferencia en la variable de entrada Descuentos en el producto Voltaren.....	42
3.4.1.3.1 Inferencia a los datos generados de la variable Descuentos del producto Voltaren.....	43
3.4.1.4 Inferencia en la variable de entrada Muestras Médicas del producto Voltaren.....	44

TABLA DE CONTENIDO (CONTINUACIÓN)

3.4.1.4.1 Inferencia a los datos generados de la variable Muestras Médicas del producto Voltaren.....	45
3.4.1.5 Modelo de regresión con los datos generados para el producto Voltaren.....	46
4. Capítulo 4: Modelos de optimización	54
4.1 Modelo de optimización con Solver (sin simulación Montecarlo)	54
4.2 Modelo de optimización con MC-Sim Solver (con simulación Montecarlo ..	57
5 Capítulo V- Conclusiones y recomendaciones	63
5.1 Conclusiones.....	63
5.2 Recomendaciones	66
ANEXO A: Análisis preliminar de los datos muestrales por variable y por producto	69
ANEXOS B: INFERENCIAS Y ANÁLISIS ESTADÍSTICOS PROBABILÍSTICOS PARA LOS PRODUCTOS RESTANTES	77
ANEXO B1: PRODUCTO GENTEAL.....	77
ANEXO B-2 PRODUCTO CATAFLAM	94
ANEXO B-3 PRODUCTO DIOVAN	109
ANEXO B-4 PRODUCTO TEGRETOL.....	126
ANEXO C: OPTIMIZACIÓN LINEAL PARA CADA PRODUCTO	144
ANEXO C-1 Modelo de optimización para el producto Genteal con Solver (sin simulación Montecarlo	144
ANEXO C-1 Modelo de optimización para el producto Genteal con Solver (sin simulación Montecarlo)	144
ANEXO C-2 Modelo de optimización para el producto Cataflam con Solver (sin simulación Montecarlo	150
ANEXO C-2 Modelo de optimización para el producto Cataflam con Solver (sin simulación Montecarlo)	150
ANEXO C-3 Modelo de optimización para el producto Diovan con Solver (sin simulación Montecarlo	156
ANEXO C-3 Modelo de optimización para el producto Diovan con Solver (sin simulación Montecarlo)	156
ANEXO C-4 Modelo de optimización para el producto Tegretol con Solver (sin simulación Montecarlo	162
ANEXO C-4 Modelo de optimización para el producto Tegretol con Solver (sin simulación Montecarlo)	162
Bibliografía.....	168

LISTA DE TABLAS

Tabla 1. Selección de productos y clasificación según línea, edad y política de mercado.....	29
Tabla 2. Niveles óptimos para cada variable del producto Voltaren	60
Tabla 3. Distribuciones para las variables del producto Voltaren	61
Tabla 4. Valores máximos de gastos por variable para el producto Voltaren	66
Tabla 5. Niveles óptimos para cada variable del producto Genteal.....	149
Tabla 6. Distribuciones para las variables del producto Genteal	150
Tabla 7. Valores máximos de gastos por variable para el producto Voltaren	155
Tabla 8. Niveles óptimos para cada variable del producto Cataflam	156
Tabla 9. Distribuciones para las variables del producto Cataflam	157
Tabla 10. Valores máximos de gastos por variable para el producto Cataflam ..	161
Tabla 11. Niveles óptimos para cada variable del producto Diovan	162
Tabla 12. Distribuciones para las variables del producto Diovan	163
Tabla 13. Valores máximos de gastos por variable para el producto Diovan	168
Tabla 14. Niveles óptimos para cada variable del producto Tegretol	169
Tabla 15. Distribuciones para las variables del producto Tegretol	170
Tabla 16. Valores máximos de gastos por variable para el producto Tegretol ...	174

LISTA DE FIGURAS

Figura 1. Gráfica de relación entre la variable Ventas vs. Tiempo del producto Voltaren.....	73
Figura 2. Gráfica de relación entre la variable Advertisement and Promoting vs. Tiempo del producto Voltaren	73
Figura 3. Gráfica de relación entre la variable Descuentos vs. Tiempo del producto Voltaren.....	74
Figura 5. Gráfica de relación entre la variable Ventas vs. Tiempo del producto Genteal.....	74
Figura 6. Gráfica de relación entre la variable Fuerza de Ventas vs. Tiempo del producto Genteal.....	74
Figura 7. Gráfica de relación entre la variable A&P vs. Tiempo del producto Genteal	75
Figura 8. Gráfica de relación entre la variable Descuentos vs. Tiempo del producto Genteal	75
Figura 9. Gráfica de relación entre la variable Muestras Médicas vs. Tiempo del producto Genteal.....	75
Figura 10. Gráfica de relación entre la variable Muestras Médicas vs. Tiempo del producto Cataflam.....	76
Figura 11. Gráfica de relación entre la variable A&P vs. Tiempo del producto Cataflam.....	76
Figura 12. Gráfica de relación entre la variable Descuentos vs. Tiempo del producto Cataflam.....	76
Figura 13. Gráfica de relación entre la variable Muestras Médicas vs. Tiempo del producto Cataflam.....	77
Figura 14. Gráfica de relación entre la variable Ventas vs. Tiempo del producto Diovan.....	77
Figura 15. Gráfica de relación entre la variable FF vs. Tiempo del producto Diovan	78
Figura 16. Gráfica de relación entre la variable A&P vs. Tiempo del producto Diovan.....	78
Figura 17. Gráfica de relación entre la variable Descuentos vs. Tiempo del producto Diovan.....	78
Figura 18. Gráfica de relación entre la variable Muestras Médicas vs. Tiempo del producto Diovan.....	78
Figura 19. Gráfica de relación entre la variable Ventas vs Tiempo del producto Tegretol.....	79
Figura 20. Gráfica de relación entre la variable FF vs. Tiempo del producto Tegretol.....	79
Figura 21. Gráfica de relación entre la variable A&P vs. Tiempo del producto Tegretol.....	79
Figura 22. Gráfica de relación entre la variable Descuentos vs. Tiempo del producto Tegretol.....	80
Figura 23. Gráfica de relación entre la variable Muestras Médicas vs. Tiempo del producto Tegretol.....	80
Figura 24. Distribución de las Ventas del producto Voltaren. Fuente: Arena.....	38

Figura 25. Gráfica de normalidad de los datos correspondientes a Ventas del producto Voltaren.....	39
Figura 26. Distribución para los datos generados de la variable Ventas del producto Voltaren.....	41
Figura 27. Distribución de A&P del producto Voltaren. Fuente: M.....	42
Figura 28. Distribución para los datos generados de la variable A&P del producto Voltaren.....	43
Figura 29. Distribución de Descuentos del producto Voltaren.....	45
Figura 30. Distribución para los datos generados de la variable Descuentos del producto Voltaren.....	46
Figura 31. Distribución de Muestras Médicas del producto Voltaren.....	47
Figura 32. Distribución para los datos generados de la variable Muestras Médicas del producto Voltaren.....	48
Figura 33. Residuales del modelo de regresión múltiple para el producto Voltaren	50
Figura 34. Prueba de normalidad de los residuales del modelo de regresión múltiple para el producto Voltaren	51
Figura 35. Transformación de Johnson de los datos generados de la variable Ventas de Voltaren.....	52
Figura 36. Residuales del modelo de regresión múltiple para el producto Voltaren con los datos transformados	53
Figura 37. Transformación de Johnson de los datos generados de la variable Ventas de Voltaren.....	82
Figura 38. Distribución para los datos generados de la variable Ventas del producto Genteal.....	83
Figura 39. Transformación de Johnson de los datos generados de la variable Ventas de Genteal	84
Figura 40. Distribución para los datos de FF del producto Genteal.....	85
Figura 41. Distribución para los datos generados de FF del producto Genteal	85
Figura 42. Transformación de Johnson de los datos de A&P de Genteal	87
Figura 43. Distribución para los datos de A&P del producto Genteal.....	87
Figura 44. Distribución para los datos generados de A&P del producto Genteal .	88
Figura 45. Transformación de Johnson de los datos de Descuentos de Genteal.	89
Figura 46. Distribución para los datos de Descuentos del producto Genteal	90
Figura 47. Distribución para los datos generados de Descuentos del producto Genteal	91
Figura 48. Distribución para los datos de MM del producto Genteal	92
Figura 49. Transformación de Johnson de los datos de MM de Genteal.....	93
Figura 50. Distribución para los datos generados de MM del producto Genteal ..	93
Figura 51. Residuales del modelo de regresión múltiple para el producto Genteal con los datos transformados	95
Figura 52. Prueba de normalidad de los residuales del modelo de regresión múltiple para el producto Genteal	96
Figura 53. Distribución para los datos de Ventas del producto Cataflam	100
Figura 54. Transformación de Johnson de los datos de Ventas de Cataflam.....	100
Figura 55. Distribución para los datos generados de Ventas del producto Cataflam	101
Figura 56. Transformación de Johnson de los datos de A&P de Cataflam.....	102
Figura 57. Distribución para los datos de A&P del producto Cataflam	103

Figura 58. Distribución para los datos generados de A&P del producto Cataflam	104
Figura 59. Distribución para los datos de Descuentos del producto Cataflam....	105
Figura 60. Transformación de Johnson de los datos de Descuentos de Cataflam	105
Figura 61. Distribución para los datos generados de Descuentos del producto Cataflam.....	106
Figura 62. Distribución para los datos de MM del producto Cataflam.	107
Figura 63. Distribución para los datos generados de MM del producto Cataflam	108
Figura 64. Distribución de probabilidad normal para el modelo de regresión de Cataflam.....	110
Figura 65. Residuales del modelo de regresión múltiple para el producto Cataflam	111
Figura 66. Distribución para los datos de Ventas del producto Diovan	115
Figura 67. Distribución para los datos generados de Ventas del producto Diovan	116
Figura 68. Transformación de Johnson de los datos de FF de Diovan	117
Figura 69. Distribución para los datos de FF del producto Diovan	117
Figura 70. Distribución para los datos generados de FF del producto Diovan ...	118
Figura 71. Distribución para los datos de A&P del producto Diovan	120
Figura 72. Distribución para los datos generados de A&P del producto Diovan.	121
Figura 73. Distribución para los datos de Descuentos del producto Diovan.....	122
Figura 74. Transformación de Johnson de los datos de Descuentos de Diovan	122
Figura 75. Distribución para los datos generados de Descuentos del producto Diovan.....	123
Figura 76. Distribución para los datos de MM del producto Diovan.....	124
Figura 77. Distribución para los datos generados de MM del producto Diovan..	125
Figura 78. Residuales del modelo de regresión múltiple para el producto Diovan	127
Figura 79. Prueba de normalidad de los residuales del modelo de regresión múltiple para el producto Diovan.....	128
Figura 80. Residuales del modelo de regresión múltiple para el producto Diovan con datos transformados.....	128
Figura 81. Distribución para los datos de Ventas del producto Tegretol	132
Figura 82. Transformación de Johnson de los datos de Ventas de Tegretol.....	132
Figura 83. Distribución para los datos generados de Ventas del producto Tegretol	133
Figura 84. Distribución para los datos de FF del producto Tegretol.	135
Figura 85. Transformación de Johnson de los datos de FF de Tegretol.....	135
Figura 86. Distribución para los datos generados de FF del producto Tegretol .	136
Figura 87. Distribución para los datos de A&P del producto Tegretol	137
Figura 88. Transformación de Johnson de los datos de A&P de Tegretol.....	138
Figura 89. Distribución para los datos generados de A&P del producto Tegretol	139
Figura 90. Distribución para los datos de Descuentos del producto Tegretol.....	140
Figura 91. Transformación de Johnson de los datos de Descuentos de Tegretol	141

Figura 92. Distribución para los datos generados de Descuentos del producto Tegretol.....	142
Figura 93. Distribución para los datos de MM del producto Tegretol.....	143
Figura 94. Distribución para los datos generados de MM del producto Tegretol	144
Figura 95. Residuales del modelo de regresión múltiple para el producto Tegretol con datos transformados.....	146
Figura 96: Hoja de cálculo para resolver el problema de optimización mediante Solver, Excel	60
Figura 97. Gráfica de correlación entre la variable A&P y Ventas de Voltaren.....	62
Figura 98. Gráfica de correlación entre la variable Descuentos y Ventas de Voltaren.....	63
Figura 99. Gráfica de correlación entre la variable MM y Ventas de Voltaren	64
Figura 100. Gráfico resumen de Ventas de Voltaren.....	65
Figura 101. Gráfica de correlación entre la variable FF y Ventas de Genteal. ...	150
Figura 102. Gráfica de correlación entre la variable A&P y Ventas de Voltaren.	151
Figura 103. Gráfica de correlación entre la variable Descuentos y Ventas de Genteal	152
Figura 104. Gráfica de correlación entre la variable MM y Ventas de Genteal...	153
Figura 105. Gráfico resumen de Ventas de Genteal	154
Figura 106. Gráfica de correlación entre la variable A&P y Ventas de Cataflam	157
Figura 107. Gráfica de correlación entre la variable Descuentos y Ventas de Cataflam.....	158
Figura 108. Gráfica de correlación entre la variable MM y Ventas de Cataflam.	159
Figura 109. Gráfico resumen de Ventas de Cataflam.....	160
Figura 110. Gráfica de correlación entre la variable FF y Ventas de Diovanl	163
Figura 111. Gráfica de correlación entre la variable A&P y Ventas de Diovan ...	164
Figura 112. Gráfica de correlación entre la variable Descuentos y Ventas de Diovan.....	165
Figura 113. Gráfica de correlación entre la variable MM y Ventas de Diovan	166
Figura 114. Gráfico resumen de Ventas de Diovan.....	167
Figura 115. Gráfica de correlación entre la variable FF y Ventas de Tegretol....	170
Figura 116. Gráfica de correlación entre la variable A&P y Ventas de Tegretol.	171
Figura 117. Gráfica de correlación entre la variable Descuentos y Ventas de Tegretol.....	171
Figura 118. Gráfica de correlación entre la variable MM y Ventas de Tegretol..	172
Figura 119. Gráfico resumen de Ventas de Tegretol.....	173

Capítulo 1- Introducción

En la actualidad, la continua lucha de las empresas por lograr la captación y retención de clientes ha forzado a que éstas desarrollen estrategias de negocio que les permita destacarse entre la competencia. Siendo Novartis Ecuador una empresa líder en el mercado nacional e internacional, requiere de dichas tácticas que le permitan seguir siendo la pionera en el mercado farmacéutico.

La siguiente tesis expone un estudio a algunas de las variables que influyen en la venta final de los productos medicinales de Novartis Ecuador con el objetivo de plantear una metodología de análisis robusto por la que se pueda, además de determinar el comportamiento de cada una de las variables influyentes, maximizar la rentabilidad generada por la venta de cualquier producto.

Se iniciará con la definición del problema que Novartis Ecuador quiere resolver y con la especificación de las variables de entrada a estudiarse. A continuación, se realiza una medición de la condición actual de la empresa mediante el análisis de sus registros históricos. Seguidamente, se propone una metodología basada en algoritmos de regresión y simulación Montecarlo que permitirán determinar el comportamiento de la variable de salida, Ventas, en función de las variables de entrada, Field Force, Advertisement & Promoting, Descuentos y Muestras Médicas.

Finalmente, a partir de una función de regresión se establecerá un modelo de optimización por el cual se pueda fijar el valor de cada variable de entrada, maximizando las ventas de Novartis.

1.1 Objetivo general

- Maximizar las ventas de Novartis Ecuador a partir de las variables de entrada: Field Force, Advertisement & Promoting, Descuentos, y Muestras Médicas.

1. 2 Objetivos específicos

- Desarrollar una estrategia competitiva frente al resto de farmacéuticas que permita a Novartis Ecuador generar la máxima rentabilidad por la venta de sus productos.
- Desarrollar un modelo de simulación por el cual se pueda determinar el comportamiento de las variables influyentes en la venta de medicinas,
- Desarrollar una herramienta en Excel por medio de la cual la metodología de simulación y optimización pueda ser aplicada a varios productos de toda la gama de medicinas que Novartis Ecuador presenta.

1. 3 Justificación del proyecto

Con la propuesta de un algoritmo de regresión y simulación Montecarlo Novartis Ecuador podrá determinar el comportamiento de cada variable de interés para la compañía, y con ello desarrollar un modelo de optimización de las mismas. Con esto sin duda, la empresa podrá obtener varias ventajas.

Una de ellas será el claro entendimiento de cada una de las variables a lo largo de la vida de un producto específico, ayudando a establecer estrategias tanto financieras como de mercadeo. Esto le ayudará a posicionarse como líder en la industria farmacéutica, al ser una de las primeras en contar con modelos

matemáticos que le ayuden a simular el comportamiento de las variables a través del tiempo, y optimizar sus canales y medios de venta de todos sus productos.

Complementariamente, Novartis aseguraría el reconocimiento adquirido en la actualidad y le permitirá lograr una retención de sus clientes existentes, así como además, la captación de otros nuevos.

1.4 Metas

- Poder ampliar la metodología propuesta a todos los productos de Novartis en el mercado ecuatoriano.
- Determinar la cantidad de recursos a ser utilizados en las diferentes etapas de vida de producto.
- Desarrollar estrategias de mercado que permitan aumentar las ventas en los productos estudiados.

1.5 Organización del documento

La siguiente tesis se enfoca en el estudio de las variables que influyen en la post venta de las medicinas Novartis. Con el fin de encontrar una función que relacione las variables y genere la máxima rentabilidad posible para la empresa.

El capítulo II hace referencia a todo el marco teórico utilizado para la elaboración de esta tesis, así se abarcan temas de estadística, regresión múltiple, econometría y simulación Montecarlo.

El capítulo III se enfoca en el estudio actual de la empresa Novartis, es decir un profundo análisis al comportamiento de cada variable y su relación con las ventas de los productos estudiados. Se incluyen las inferencias a distribuciones de cada

variable, simulación Montecarlo y los modelos de regresión múltiple ajustados para cada producto.

El capítulo IV finalmente, se centra en el diseño de un modelo matemático por medio del cual se logre una optimización de los modelos de regresión antes obtenidos. Se presentan las restricciones al modelo y los resultados obtenidos de éstos.

Por último el capítulo V expone las conclusiones del proyecto junto con las recomendaciones a considerar que resultaron de este estudio.

1.6 Descripción de la empresa Novartis Ecuador

1.6.1 Historia de la empresa

Novartis fue creada en 1996 con la fusión de Ciba- Geigy y Sandoz, con un alrededor de cien mil asociados, en 140 países. Su nombre significa “Nuevas artes en las ciencias de la vida”¹ y refleja el compromiso hacia la investigación y el desarrollo de productos innovadores en las comunidades donde están presentes. Su logo representado por una flor y una flecha, simbolizan la flor de la vida, así como también un mortero con su mano, (el mortero era usado en tiempos antiguos por farmacéuticos para la elaboración de medicinas). La flecha azul indica precisión, y sus colores rojo y amarillo, representan respectivamente, el calor del sol y la llama de un mechero de Bunsen, utilizado en experimentos científicos. (Manual)

Su historia se remonta al año 1758 con el nacimiento de Ciba- Geigy, empresa fundada por Johann Rudolf Geigy, dedicada a la comercialización de productos químicos, colorantes y medicamentos de distintas clases. Un siglo después, el 1859 Alexander Calvel inició la producción de colorantes de origen

¹ Manual de Bienvenida Novartis Ecuador. Pág. 3.

sintético y productos químicos para la industria textil. La compañía fue llamada “Company for Chemical Industry Basel” para luego ser conocida como “Ciba”. Para 1970 se produce la fusión de estas dos empresas, y se crea entonces, Ciba- Geigy Ltd. Años después, en 1992, la compañía cambia su nombre a solamente Ciba e introduce una nueva imagen corporativa. (Manual)

Por su parte en 1886, Alfred Kern y Edouard Sandoz fundan la compañía Chemical Company Kern & Sandoz, enfocada en la producción de colorantes. En los años siguientes se diversifica el negocio hacia el área farmacéutica, a la producción de sacarosa, pesticidas, y calcio, entre los productos más importantes. En 1967 ocurre la fusión con Wander Ltd., y se incursiona en el campo de productos dietéticos. Seguido a esto vino la compra de Delmark, Wasa y Gerber Baby Food. (Manual)

El 7 de marzo de 1996, Ciba y Sandoz, dos compañías farmacéuticas suizas, con 300 años de experiencia, anuncian su fusión para crear lo que es Novartis. Para diciembre de ese año, se legaliza en el ámbito mundial la creación de la nueva compañía, lo cual representó la fusión más grande de la historia.

En Ecuador, Novartis es una de las empresas más importantes en el mercado farmacéutico. Se constituyó en Febrero de 1997 y cuentan con más de 190 colaboradores en el ámbito nacional. Cuentan con una oficina principal en Quito y una sucursal en Guayaquil. (Manual)

1.6.2 El negocio

1.6.2.1 Misión:

“Descubrir, desarrollar y comercializar con éxito productos innovadores a curar y prevenir enfermedades, a aliviar el sufrimiento y a mejorar la calidad de vida”².

1.6.2.2 Estrategias de negocio:

- Centrarse en el cuidado de la salud.
- Innovar en beneficio de sus clientes.
- Creer en su personal.
- Operar de forma sostenible.

(Manual).

1.6.2.3 Aspiraciones:

- Ser conocidos por los beneficios que aportan a la sociedad.
- Conseguir una rentabilidad justa para aquellos que invierten dinero, tiempo y talento en la compañía.
- Disfrutar de un entorno laboral apasionante.
- Ser conocidos como una compañía responsable.

(Manual).

² Manual de Bienvenida Novartis Ecuador, pág. 6.

1.6.2.4 Cinco prioridades en Ecuador:

- Alcanzar los objetivos planteados de ventas, participación de mercado, resultados de Gerencia y el costo del personal.
- Maximizar lanzamientos y negocios actuales.
- Cultura generadora basada en un enfoque absoluto en el cliente.
- Mejorar productividad, alcanzar ahorros mejorando la productividad, eliminando burocracia y haciendo MÁS con MENOS.
- Invertir en su gente, desarrollar talentos, alcanzar los objetivos de diversidad e inclusión y lograr metas en contratación y desarrollo.

(Manual).

1.6.2 Productos:

En la actualidad, Novartis está organizada en cuatro divisiones: Farmacéutica (Pharma), Sandoz, Vacunas y diagnósticos, y Productos OTC. Dentro de éstas y en el Ecuador existen las siguientes líneas de negocio:

1.6.3.1 Mercado masivo: dentro del área de farmacéutica, se dedica al desarrollo innovador de medicinas por prescripción. Los productos están concentrados en las siguientes áreas terapéuticas:

- Cardiovascular y metabolismo, incluye productos para tratar problemas de presión alta y diabetes. Los productos más destacados son: Diovan, Exforge, y Rasilez.
- Gastro, para el tratamiento de enfermedades como la gastritis, úlcera, etc. Los productos incluyen a Glivec,

- Respiratorio, para el tratamiento de enfermedades como el Asma, enfisema pulmonar y bronquitis. Sus productos incluyen Xolair.
- Enfermedades inmunológicas e infecciosas, con productos como Neoral, Simulect. Myfortic, Certican, Aclasta y Tyzeka.

(Manual)

1.6.3.2 Especialidades y cuentas clave: mediante este canal se intenta llegar a principales instituciones del país como por ejemplo: IESS, Hospital Militar, SOLCA, etc. Tiene las siguientes áreas terapéuticas:

- Sistema Nervioso Central, con tratamientos para enfermedades como el Alzheimer, Parkinson, epilepsia, depresión, migraña, esquizofrenia, y el Trastorno por déficit de atención e hiperactividad, con productos como Exelon, Tegretol, Stalevo, Extavia.
- Opta, en el área de oftalmología con productos como Lucentis y Genteal y Visudyne.
- Transplantes.
- Oncología, con productos para el tratamiento del cáncer como Fermara, Zometa, Exjade y Afinitor,
- Vacunas, algunas son: contra la influenza, meningitis, rabia, encefalitis, polio, difteria, tétanos, entre las más importantes.
- Diagnóstico: con más de 20 productos para luchar contra las enfermedades virales y bacteriales.
- Anti- Infecciosos
- Osteoporosis: con productos como Aclasta.

(Manual)

1.6.3.3 Consumer Health: dedicada a productos OTC (por sus siglas en inglés, Over- The- Counter), es un líder en productos de cuidado de la salud, autorizando la automedicación de productos para el tratamiento y prevención de enfermedades comunes. Incluye productos CibaVision, Dermam y la franquicia Voltaren.

- **Ciba Vision** es uno de los líderes mundiales en investigación desarrollo y fabricación de lentes de contacto y de productos para el cuidado de éstos. Se incluyen: Air Optix, Solo Care Aqua.
- Derman: con productos como Lamisil.
- Franquicia Voltaren Emulgel.

Dentro de los productos OTC de mayor renombre están: Theraflu, Voltarén Emulgel, Comtrex, Tonopan, Alevian Duo, Mebocaina, Excedrin, Gas-X, entre muchos más.

Capítulo 2: Marco teórico

2.1 Análisis de regresión

Se llama a un análisis de regresión a la técnica estadística por medio de la cual se moldea o se investiga la relación existente entre dos o más variables. Un modelo de regresión puede también usarse para optimizar un proceso, por ejemplo al encontrar los niveles óptimos que minimicen o maximicen el valor de las variables para ya sea mejorar o controlar el proceso (Montgomery y Runger 431).

Un modelo de regresión lineal simple es aquel que cuenta con solamente una variable independiente o variable de regresión; cuando no se conoce acerca de la relación teórica entre las x y y de la función, se dice que el modelo es empírico (Montgomery y Runger 432).

Cuando se tienen situaciones en las que existe más de una variable de regresión, se las conoce como modelos de regresión múltiple. En este modelo, tal como el de regresión simple, la Y representa la variable de salida o respuesta del modelo, y las X representan las variables independientes que explican a Y . Todo modelo de regresión, ya sea lineal o múltiple van sujetos a un término de error aleatorio.

Entonces, un modelo de regresión múltiple adopta la siguiente relación:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

A β_k se la conoce como el coeficiente de regresión parcial, debido a que β_k mide el cambio esperado en Y , cuando ocurre un cambio unitario en x_k , siempre

que las demás variables se mantengan constantes. Al modelo se lo conoce como un modelo de regresión lineal múltiple con k variables regresoras. (Montgomery y Runger 485).

Con frecuencia se usa un modelo de regresión lineal múltiple como una función de aproximación, en donde no se conoce la relación directa entre las variables independientes y dependientes.

Los modelos que incluyen efectos de interacción pueden ser analizados también mediante un modelo de regresión lineal múltiple; la interacción entre las dos variables se representa por el producto cruzado de los de las variables independientes. En tal caso, el modelo puede ser acomodado a uno de la forma original, reemplazando las variables cruzadas por otra variable independiente.

Generalmente, “cualquier modelo de regresión cuyos parámetros, β_k son lineales es un modelo de regresión lineal, independientemente de la forma de la superficie que genera”³.

Un modelo con interacciones de las variables implica que el efecto que se produce por el cambio en una variable depende de la otra variable; estas interacciones son estudiadas de igual manera por medio de los modelos de regresión múltiple. (Montgomery y Runger 487).

2.1.1 Método de los mínimos cuadrados para la estimación de parámetros

Si se parte de un modelo de regresión lineal múltiple de la forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i$$

³ Montgomery y Runger, page 486.

Donde,

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i$$

De aquí, la ecuación de mínimos cuadrados es entonces:

$$L = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2$$

Como esta ecuación debe minimizarse con respecto a β_k , se deriva con respecto a estas variables, para obtener las estimaciones de mínimos cuadrados, que son:

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{ik} &= \sum_{i=1}^n y_i \\ \hat{\beta}_0 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1}x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{i1}x_{ik} &= \sum_{i=1}^n x_{i1}y_i \\ \hat{\beta}_0 \sum_{i=1}^n x_{ik} + \hat{\beta}_1 \sum_{i=1}^n x_{ik}x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{ik}x_{i2} + \hat{\beta}_k \sum_{i=1}^n x_{ik}^2 &= \sum_{i=1}^n x_{ik}y_i \end{aligned}$$

De este sistema de ecuaciones, en el que hay $p = k + 1$ ecuaciones, para cada uno de los regresores desconocidos, se pueden encontrar resolviendo el sistema de ecuaciones de cualquier método para resolver ecuaciones lineales.

(Montgomery y Runger 488).

2.1.2 Enfoque matricial a la regresión lineal múltiple

A veces puede resultar de gran facilidad presentar un modelo de regresión lineal múltiple utilizando notación matricial. A un sistema de n ecuaciones se lo puede representar de la siguiente manera:

$$y = X\beta + \varepsilon$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{1k} \\ 1 & x_{21} & x_{22} & x_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & x_{n2} & x_{nk} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

Con esta notación se tiene que \mathbf{y} es un vector de dimensión $(n \times 1)$ de las observaciones; \mathbf{X} una matriz de $(n \times p)$ correspondiente a los niveles de las variables independientes; β es un vector $(p \times 1)$ de los coeficientes de regresión, y finalmente el error es un vector de $(n \times 1)$ errores aleatorios. Para encontrarse los valores correspondientes a los mínimos cuadrados, se deriva la ecuación de L, con respecto a β , y se la iguala a cero. Las ecuaciones que se obtienen son iguales a las de la forma escalar, y para resolverlas se multiplica ambos miembros de las ecuaciones por la inversa de $X'X$, y se tiene que la estimación de mínimos cuadrados de β , es decir:

$$\hat{\beta} = (X'X)^{-1} X'y$$

(Montgomery y Runger 492).

Así, $X'X$ es una matriz simétrica de dimensiones $(p \times p)$ y está formada por las sumas de los cuadrados de los elementos de las columnas de X . $X'y$ es un vector columna $(p \times 1)$, en donde los elementos son las sumas de los productos cruzados de las columnas de X y las observaciones.

(Montgomery y Runger 493).

2.1.3 Propiedades de los estimadores de mínimos cuadrados y estimación de σ^2

Los coeficientes obtenidos de la regresión múltiple se conocen como los estimadores de mínimos cuadrados, los cuales cuentan con ciertas propiedades numéricas e estadísticas, que se describen a continuación: (Damodar 60).

- Los estimadores por medio del método de mínimos cuadrados están expresados en términos de cantidades observadas y por lo tanto pueden ser fácilmente calculados. (Damodar 60).
- Son estimadores puntuales, lo que significa que dada la muestra, cada estimador proporciona un solo valor del parámetro poblacional relevante. (Damodar 61).
- Bajo ciertos supuestos sobre los términos del error, se dice que éstos son estadísticamente independientes con media cero y varianza σ^2 . Con esto se puede establecer que los estimadores de los mínimos cuadrados $\hat{\beta}_k$ son estimadores insesgados de los coeficientes de regresión β_k . (Montgomery y Runger 504).
- Los residuos no están correlacionados ni con el valor predicho Y, ni con las variables regresoras, X. (Damodar 62).
- En cuanto a la varianza de las $\hat{\beta}_k$ están expresadas en términos de la inversa de la matriz de $X'X$. A ésta, multiplicándola por σ^2 se obtiene la matriz de covarianza de los coeficientes de regresión. Se

llama covarianza a la descripción de la forma en la que varían dos o más variables aleatorias en conjunto, es decir la relación que ambas guardan entre sí. (Montgomery y Runger 237).

Por lo general la matriz de la covarianza es simétrica de dimensión $(p \times p)$ en donde el elemento jj -ésimo es la varianza de $\hat{\beta}_j$ y cuyo elemento i, j -ésimo es la covarianza entre $\hat{\beta}_i$ y $\hat{\beta}_j$. . (Montgomery y Runger 504).

2.1.4 Supuestos detrás del método de mínimos cuadrados

- 2.1.4.1** El modelo de regresión es lineal en los parámetros. (Damodar 63).
- 2.1.4.2** Los valores de X son fijos en un muestreo repetido. (Damodar 64).
- 2.1.4.3** El valor medio de la perturbación es igual a cero, esto es, dado el valor de X , la media o el valor esperado del término aleatorio de perturbación es cero. (Damodar 64).
- 2.1.4.4** Homoscedasticidad o igual varianza de los errores; dado el valor de X , la varianza de su error es la misma para todas las observaciones, las varianzas son idénticas. (Damodar 65).
- 2.1.4.5** Dados valores cualesquiera para de X_i, X_j , la correlación entre sus dos errores es cero. No hay correlación entre las perturbaciones. (Damodar 67).
- 2.1.4.6** La covarianza entre el error y la variable regresora X , es cero. (Damodar 68).
- 2.1.4.7** El número de observaciones n debe ser mayor que el número de parámetros por estimar. (Damodar 69).

2.1.4.8 Variabilidad en los valores de X, es decir, no todos los valores de X en una muestra dada tienen que ser iguales, y técnicamente la varianza debe ser un número positivo finito. (Damodar 69)

2.1.4.9 El modelo de regresión está correctamente especificado. (Damodar 70).

2.1.5 Pruebas de hipótesis en la regresión lineal múltiple

Las pruebas de hipótesis son útiles para medir la adecuación del modelo; tal como en la regresión simple, es necesario que los términos del error del modelo de regresión tengan una distribución normal e independiente con media cero y varianza σ^2 .

La primera prueba importante de hipótesis para un modelo de regresión, es la prueba de significación. Esto es, el comprobar si existe una relación lineal entre la variable de respuesta, y , y las variables regresoras x . La hipótesis nula en este caso, corresponde a la igualdad de los coeficientes de regresión; la hipótesis alternativa es la desigualdad. Cuando se rechaza la hipótesis nula, se implica que al menos una de las variables regresoras contribuye de manera significativa al modelo. (Montgomery y Runger 506).

A partir del planteamiento de la prueba de hipótesis se procede a obtener los estadísticos de la prueba correspondientes, y a realizar un análisis de varianza para probar la significación de la regresión en una regresión múltiple (aquí servirán entonces la obtención de los mínimos cuadrados). (Montgomery y Runger 507).

En muchas ocasiones se requiere probar sobre los coeficientes de regresión de manera individual, para determinar el valor potencial de cada una de las variables regresoras. Para determinar dichos valores, a veces es necesario la

inclusión de variables adicionales o eliminación de alguna presente. Cuando se incluye una variable adicional al modelo la suma de los cuadrados aumenta, mientras que la suma de los cuadrados del error disminuye. Para decidir si el uso de la variable adicional en el modelo es adecuado, la suma de cuadrados deberá ser lo suficientemente grande. Caso contrario, cuando la variable nueva de inclusión, disminuye el desempeño del modelo, es posible que el cuadrado medio del error aumente. (Montgomery y Runger 509).

Existe otro método para probar la contribución de una variable individual en un modelo de regresión múltiple. Se lo conoce como la prueba general de la significación de la regresión o el método de suma de los cuadrados. Mediante este procedimiento lo que se intenta determinar es el incremento en la suma de los cuadrados de regresión. A este modelo también se lo puede usar para determinar la contribución de un subconjunto de variables regresoras deseadas. (Montgomery y Runger 511).

Las pruebas de hipótesis irán variando de acuerdo a lo que se requiere probar, y por lo tanto el estadístico usado también cambia. La validación de las hipótesis nulas o alternativas deberá tener el suficiente sustento estadístico para probar lo que se dice, y tener un respaldo en datos de dichas afirmaciones o negaciones.

2.1.6 Intervalos de confianza para la regresión lineal múltiple

Al igual que en los modelos de regresión simple, los intervalos de confianza en los modelos de regresión múltiple son útiles para estimar el grado de incertidumbre asociado con una estimación puntual. (Montgomery, Runger, 289).

Para obtener los intervalos de confianza de los coeficientes de regresión, también se asumen los supuestos de las pruebas de hipótesis, es decir, que los errores tengan una distribución normal, con media cero y varianza σ^2 . (Montgomery y Runger 515).

Un intervalo de confianza también es útil cuando se requiere hacer una predicción de nuevas observaciones a través de la regresión lineal múltiple. Esto es, el pronosticar observaciones futuras de la variable de respuesta Y que corresponda a valores particulares de las variables independientes X . Este tipo de intervalo de confianza será siempre de mayor anchura que el intervalo de confianza normal, esto se debe a que el intervalo de confianza de un coeficiente expresa el error al estimar la media de una distribución, mientras que el intervalo de predicción expresa el error al predecir una observación futura a partir de las distribuciones en el punto x_0 . (Montgomery y Runger 518).

2.1.7 Medidas de la adecuación del modelo

2.1.7.1 Coeficiente de determinación múltiple Ajustado (R^2 -adj)

R^2 -adj es una medida de la cantidad en que se reduce la variabilidad del valor de Y obtenido al usar las variables regresoras. Como en la regresión simple, debe tenerse un valor de $0 \leq R^2 \leq 1$ " (Montgomery y Runger 521). Mas un valor grande de R^2 -adj no asegura necesariamente que el modelo de regresión utilizado es el adecuado; cuando se agrega una variable al modelo, R^2 -adj aumentará aunque la variable adicional sea significativa o no. Por esto hay que tener especial atención con las conclusiones que se obtienen del coeficiente de

determinación múltiple, ya que pueden inferir en respuestas pobres o erróneas sobre la variable de respuesta.

Muchas veces se puede caer en el error de tratar de conseguir un R^2 -adj elevado, eligiendo aquel modelo de regresión que tenga un coeficiente de determinación múltiple más grande. Sin embargo, cuando un modelo cuenta con un R^2 -adj bajo, no significa que el modelo es necesariamente malo (Damodar 214). En el libro “Econometría”, de Damodar, se hace referencia a Goldberger quien afirmó lo siguiente:

“ R^2 -adj tiene un papel muy modesto en el análisis de regresión, y es una medida de la bondad del ajuste de una regresión lineal de mínimos cuadrados muestral en un cuerpo de datos. Nada en el modelo de regresión común, exige que R^2 -adj sea elevada. Por tanto, una R^2 -adj elevada no es evidencia a favor del modelo y una R^2 -adj baja no es evidencia en su contra”⁴.

Por otra parte, al extraer la raíz cuadrada a R^2 -adj, se obtiene el coeficiente de correlación múltiple R , entre y y el conjunto de variables regresoras. R representa una medida de asociación lineal entre la variable de salida con las variables de entrada. (Montgomery y Runger 521).

2.1.7.2 Análisis de residuales

Los residuales en un modelo de regresión múltiple están dados por $e_i = y_i - \hat{y}_i$ y son importantes para determinar la adecuación del modelo. (Montgomery y Runger 522).

⁴ Damodar, pág. 215.

Un análisis de residuales permite verificar si los supuestos sobre normalidad aproximada con varianza constante y media cero de los errores, se cumplen. De igual manera sirven para determinar si es conveniente o no, incluir términos adicionales en el modelo. El análisis de residuales se basa principalmente en la elaboración de gráficas por las que se determinan los supuestos. Así, son útiles por ejemplo, un histograma de frecuencias de los residuales o una gráfica de probabilidad normal de los residuales, siendo ésta última un tanto mejor. (Montgomery y Runger 461).

Muchas veces resulta útil graficar los residuales contra las variables que no están en el modelo pero que son posibles de ser añadidas, para ver entonces, si con la incorporación de éstas, el modelo podría mejorarse. (Montgomery y Runger 522).

Para complementar las decisiones obtenidas con las gráficas, se recurre a realizar una estandarización de los residuales, mediante:

$$d_i = \frac{e_i}{\sqrt{\hat{\sigma}^2}} \quad \text{Para } i= 1\dots n.$$

Cuando los errores siguen una distribución normal, un 95% aproximado de los residuales estandarizados deben estar incluidos en el intervalo (-2, +2). Mediante esto se pueden identificar los puntos atípicos del modelo, los cuales proporcionan importante información sobre el resultado obtenido. (Montgomery y Runger 462).

Es importante obtener tres graficas de los residuales: 1) en secuencias d tiempo 2) contra las \hat{y}_i y 3) contra las variables independientes x. Ninguna de estas gráficas deberá mostrar algún patrón específico para poder representar una situación ideal. Sin embargo, cuando se tiene un patrón, por ejemplo en forma de embudo, la varianza de los residuales puede aumentar con el tiempo o con la

magnitud de y_i o x_i . Para eliminar este problema, se usa una transformación de datos de la variable de salida y , con el fin de estabilizar la varianza. Entre las técnicas más usadas están: la utilización de \sqrt{y} , $\ln y$ o $1/y$ como variable de respuesta, habrá que buscar la transformación más adecuada al problema de regresión planteado. (Montgomery y Runger 463).

Es a veces conveniente graficar los residuales estandarizados, “ya que estos se escalan de tal modo que su desviación estándar es la unidad”⁵. Esto implica que los residuales grandes, que pueden indicar puntos anormales o atípicos, sean más notorios y evidentes de inspeccionar.

Con frecuencia los residuales estandarizados subestiman la verdadera magnitud de los residuales, y por ello se usan los determinados residuales studentizados, los cuales sirven de mejor ayuda al momento de examinar los residuales y sus puntos atípicos. (Montgomery y Runger 525).

2.1.8 Variables indicadoras

Los modelos de regresión múltiple pueden contener variables cuantitativas, es decir variables que se miden en una escala numérica, pero también pueden contener variables cualitativas. A éstas se las llama variables indicadoras y se las puede moldear de manera que si tiene t niveles, con $t-1$ variables indicadoras a las que se les asigna el valor de cero o uno dependiendo de la condición que adopte. A este tipo de variables también se las llama variables mudas. (Montgomery y Runger 534).

⁵ Montgomery y Runger page. 523

2.1.9 Selección de variables en regresión múltiple

2.1.9.1 Problema de la construcción del modelo

Para plantear un modelo de regresión es importante seleccionar adecuadamente las variables regresoras. En ocasiones dichas variables son escogidas sobre la base de la experiencia de aquellos que son expertos en el tema o bien de la teoría fundamental que el problema encierra. Sin embargo, lo que se desea es contar con el mejor subconjunto de variables regresoras, siendo las menores posibles para disminuir costos de calidad y tener un modelo de regresión no tan complejo. Mas nunca se encontrará un algoritmo en especial que produzca siempre una buena solución al problema de selección de las variables, se necesitará la ayuda de buenos criterios de analistas expertos. (Montgomery y Runger 539).

2.1.9.2 Procedimiento computacional para la selección de variables

Si se supone un número K de variables regresoras candidatas y una sola variable de respuesta Y , habrá un modelo con un elemento β_0 para la ordenada en el origen y por lo tanto $k + 1$ términos. Existen varios enfoques posibles para generar los modelos de regresión:

2.1.9.2.1 Todas las regresiones posibles

En este enfoque se analiza todas las ecuaciones de regresión que incluyen una variable candidata, todas las ecuaciones de regresión que incluyan dos variables candidatas y así sucesivamente. Entonces, si hay K variables

regresoras candidatas, es necesario examinar 2^k ecuaciones, esto implica que mientras el número de variables candidatas aumenta, lo hace también el número de ecuaciones posibles. (Montgomery y Runger 539).

Para determinar cuál de las ecuaciones es la más adecuada se usa frecuentemente, el coeficiente de determinación múltiple, denotado por R^2_p , con p términos, es decir $p-1$ variables regresoras y un término para la ordenada en el origen. Cuando R^2_p incrementa, p incrementa y es un máximo cuando $p=K+1$. Entonces se debe encontrar el punto en el que agregar más variables al modelo ya no es útil porque resulta en un incremento reducido de R^2_p .

2.11 Simulación Montecarlo

La regresión es una herramienta para inferir y conocer cómo ciertas variables se relacionan unas con otras. Mas esta inferencia se realiza sobre una muestra de la población entera, es decir en base sólo a la información que se observa. La meta real de inferir en las variables está en determinar qué tipo de evidencias brinda la muestra utilizada, sobre la relación de las variables. (Humberto, 215).

Como un análisis de regresión, la simulación Montecarlo construye un modelo artificial sobre la base de uno real, para tratar de comprender su comportamiento. Además, su nombre recae en la aleatoriedad inherente que implica su análisis. Teóricamente, la simulación Montecarlo es un método de análisis artificial basado en la generación de números aleatorios que recrean un proceso, corriéndolo por varias veces y observando directamente los resultados que genera. (Humberto, 216).

Exactamente lo que hace la simulación Montecarlo es un proceso de generación de los datos del proceso, seguida de la generación de réplicas del proceso, para finalmente analizar los resultados del experimento. (Humberto, 225).

Dentro de las mayores ventajas de la simulación Montecarlo es que permite ver claramente la fuente de error y variación de un problema. De manera interactiva, con las réplicas del experimento que permite la simulación, es mucho más fácil notar en dónde está la variación y la fuente de dónde procede. (Humberto, 230).

Una simulación Montecarlo normalmente se desarrolla por medio de una simulación computacional en que se imita la situación real para hacer predicciones de su comportamiento futuro. En este caso se trata de realizar un modelo determinístico, es decir uno en el que sin importan cuántas veces se lo corra, los resultados siempre serán los mismos. Lo que se necesita para llevar a cabo un modelo de simulación Montecarlo son entradas o variables, una o varias ecuaciones que utilicen aquellas entradas para dar lugar a resultados finales o variables de respuesta (Monte Carlo).

Con el uso de entradas aleatorias lo que esencialmente se hace es transformar un modelo determinístico en uno estocástico. Esto significa que existe un alto grado de azar en el modelo pero que su análisis permite predecir el rendimiento de los sistemas y ayuda a la toma de decisiones, ya sea en el diseño de un sistema o bien en la comparación de políticas alternativas. Entiéndase como azar el hecho de que en el modelo intervienen factores que no se pueden pronosticar con precisión y por ello se hace uso de la probabilidad (Modelos).

Un modelo de simulación Montecarlo es comúnmente conocido como un método de muestreo ya que sus entradas son generadas aleatoriamente a partir de funciones de probabilidad inferidas en un muestreo de datos. (Montecarlo)

Son cinco los pasos que se requieren para llevar a cabo una simulación de este tipo:

1. Crear o definir un modelo matemático.
2. Generar una serie de entradas aleatorias.
3. Evaluar el modelo y almacenar los resultados de la variable de salida.
4. Repetir los pasos 2 y 3, por cuantas veces se haya determinado.
5. Analizar los resultados usando histogramas, estadística, intervalos de confianza, entre otras técnicas.

(Montecarlo).

Capítulo 3: Modelo de regresión múltiple

Con el fin de analizar la relación existente entre la variable de salida, con las variables de entrada, se generará un modelo de regresión múltiple. Éste determinará cuál es la influencia, si es que existe, de cada una de las variables propuestas con la variable de interés, que en este caso es la rentabilidad de los medicamentos Novartis.

A partir del modelo de regresión múltiple se podrán fijar valores óptimos correspondientes a cada variable de entrada, de manera que se ajuste el modelo y con él se pueda tener el mayor valor de rentabilidad posible por cada producto de la farmacéutica.

A continuación se presenta todo lo que concierne al diseño del modelo de regresión múltiple para la rentabilidad de la empresa Novartis.

3.1 Definición de variables y productos

La gama de productos Novartis comprende un alrededor de 300 productos; para efectos del estudio se analizarán solamente cinco de ellos. La selección de los mismos fue dada en base a conversaciones con expertos de la empresa, quienes diagnosticaron la importancia de estudiar a productos tanto de la división Pharma como de la división OTC. La razón de esto recae en el grado de influencia que tienen las variables seleccionadas, en la venta de los productos pertenecientes a esta rama. Así, la elección de los productos a ser estudiados comprendió los siguientes aspectos: la línea de comercialización, la edad del

producto, y su política de venta en el mercado. La directiva de Novartis decidió enfocar el estudio en productos que sean: uno que muestre actualmente ventas no tan elevadas, el de mayor rentabilidad para la empresa, un producto viejo con receta, uno viejo de tipo OTC y uno relativamente nuevo.

Así se seleccionó los siguientes productos (siguiendo el respectivo criterio anterior):

Nombre del producto	Línea a la que pertenece	Edad del producto	Política de mercado
Tegretol	Especialidad y cuentas clave/ Sistema Nervioso Central	Antiguo	Venta con receta
Diovan	Mercado masivo / Cardiovascular	Antiguo	Venta con receta
Cataflam	Mercado Masivo/ Productos maduros	Antiguo	Venta receta (aunque se lo puede conseguir sin mayor dificultad)
Voltaren	OTC / Productos maduros	Antiguo	Venta libre, fuerte
Genteal	Especialidad y cuentas clave/ Oftalmología	Relativamente nuevo	Venta receta (aunque se lo puede conseguir sin mayor dificultad)

Generación propia

Tabla 1. Selección de productos y clasificación según línea, edad y política de mercado.

La organización de este proyecto supone el estudio de cinco productos, mas a continuación se presenta el análisis de sólo uno de ellos, los restantes se muestran como ANEXOS al final del documento. Se lo hace de esta manera porque la metodología utilizada es iterativa para cada producto, por lo que lo que se hace para uno, se hace para todos. De esta manera se estudiará primero el caso de Voltaren, mientras que para los demás productos su estudio se encuentra en los ANEXOS B-1, B-2, B-3, B-4.

Una vez que se establecieron los productos a ser estudiados se debe definir las variables a considerar en sus modelos de regresión. Para esto, se mantuvieron reuniones con expertos en el tema y bajo el método Delphi se seleccionaron cuatro variables que inciden directamente en la venta final de alguna medicina Novartis. Hay que aclarar que la farmacéutica tiene dos tipos de clientes, el directo y el indirecto. Cliente directo son los centros de distribución y grandes cadenas farmacéuticas a las cuales vende sus productos, por ejemplo: Farcomed, Instituciones públicas, etc. Mas estos clientes no representan el destino final de los medicamentos; es el cliente indirecto, es decir pacientes o consumidores finales quienes deciden si comprar o no un producto Novartis, sin importar la cadena de abastecimiento en el que éste se encuentre.

Bajo este panorama, todos los medios de difusión que Novartis emplea, son para dar a conocer sus productos directamente al consumidor final, quien garantizan la rentabilidad de la empresa. Por ende las variables de interés en este estudio son aquellas que inciden directamente en la compra del producto por parte del cliente indirecto, el paciente. Así se definieron cuatro variables de entrada o regresoras independientes, junto con la variable de salida o variable dependiente. Estas son:

- Y = Venta en el mercado, conocida para Novartis como las ventas a grandes Distribuidores, Farmacias y Ventas institucionales o públicas, por medio de las cuales se llega al cliente final o indirecto. Es una cantidad dada en dólares y al ser la variable de respuesta, está sujeta, y varía en función del comportamiento de las variables regresoras, X .

- X_1 = dada por el dinero invertido en FF por sus siglas en inglés Field Force. X_1 viene dada por la cantidad anual de dinero que se emplea en Fuerza de ventas, es decir en los visitadores médicos y sus actividades que promueven la venta de cierto producto entre médicos e instituciones, con el fin de que lo receten más a sus pacientes.
- X_2 = representa el valor en dólares invertido en A&P, Advertisement and Promoting, es decir el marketing usado en todo medio de difusión para la venta de cierto producto. Esta variable incluye gastos de tipo anuncios publicitarios, congresos, eventos, y otros con el fin de promocionar e incrementar las ventas directas o indirectas de un producto.
- X_3 = denota los descuentos y bonificaciones por las ventas que realiza la empresa. Debido a que Novartis solamente vende a las cadenas de farmacias y centros de distribución grandes, los descuentos se aplican solamente a ellos, mas la importancia de analizar esta variable radica en conocer cómo el descuento que ofrece Novartis a su cliente directo es transmitido al cliente final, y la manera en que esta variable incide en la rentabilidad de la empresa.
- X_4 = denota la cantidad de dinero invertido en muestras médicas, como parte de promoción del producto entre médicos, farmacias, instituciones públicas, entre otros centros autorizados.

3.2 Inferencia en las distribuciones de cada variable

Con el fin de generar un modelo que se ajuste adecuadamente a las variables propuestas, se eligió un muestreo de datos para cada variable y para

cada producto, que consta de información desde el año 2006 hasta el 2009. Al ser importante el número de datos para la formulación y adecuación del modelo, se partirá de hacer en primera instancia, una inferencia en cada serie de datos, con el fin de determinar la distribución de probabilidad que siguen. Con ello, se podrán generar nuevos datos aleatorios, a partir de una simulación Montecarlo, y bajo los parámetros de la distribución encontrada; así la generación del modelo de regresión múltiple tendrá un error más bajo que si se tuviera un menor número de datos.

Como un análisis preliminar, se estudiaron los datos pertenecientes a cada variable por cada producto propuesto. El fin de esto, es tener una idea básica de cómo interactúan los datos entre sí a lo largo del tiempo, señalando los valores máximos, mínimos que alcanzaron en dichos años. Esto permitirá entender mejor los resultados que se obtendrán de la regresión y a su vez, facilitará al momento de plantear la optimización para cada variable. Estos se muestran en el Anexo A.

Para hacer la inferencia en cada serie de datos, se usaron dos software, Arena con el módulo Input Analizar, y Minitab. Ambos proveen estimaciones de los parámetros a los que la serie de datos se ajusta mejor, y dan una expresión de la función de probabilidad adecuada, llamada $F(p)$.

Como los nuevos datos serán generados a partir de históricos, en base a estimaciones de distribuciones dadas, se trabajará con variables aleatorias con una distribución de probabilidad específica. En el Input Analyzer de Arena se tiene la opción de ajustar los datos a dos tipos de distribuciones de probabilidad existentes, son: las distribuciones teóricas y las distribuciones empíricas. Las primeras son aquellas que se basan en una formulación matemática, tales como

la distribución gama o exponencial (Kelton, et all., 156). Mientras que las distribuciones de probabilidad empíricas son las que dividen a los datos en agrupaciones más pequeñas para calcular la proporción de los valores en cada grupo, haciendo una interpolación entre puntos para una mayor precisión (Kelton, et all., 156).

Para un mayor ajuste, se utiliza la opción de comparar los datos históricos a los dos tipos de distribuciones, teóricas y empíricas, con lo cual se podrá conseguir una mejor aproximación de la distribución a la que los datos pertenecen.

Para la selección de la distribución más apropiada Arena y Minitab tienen tres criterios numéricos, son:

1. Error medio cuadrado. Mientras mayor sea este valor, más alejados están los datos de la distribución usada (Kelton, et all., 160).
2. La prueba de bondad y ajuste Chi- cuadrada, en donde un valor de p “grande” indica que la distribución comparada es la adecuada. Un valor de p “grande” significa que éste sea mayor al nivel de significancia usado, que es igual a 0.05. (Kelton, et all., 161)
3. La prueba de bondad y ajuste de Kolmogorov- Smirnov, con el mismo concepto que la prueba Chi- cuadrada.

Las pruebas de hipótesis para el caso de las pruebas de bondad y ajuste usadas en los software son las siguientes:

H_0 = Los datos siguen una distribución _____ (cualquiera con la que se pruebe)

H_1 = Los datos no siguen una distribución _____ (cualquiera con la que se pruebe).

Con un valor de p menor a 0.05, la hipótesis nula se rechaza, por lo que no hay suficientes pruebas estadísticas para decir que los datos siguen cierta distribución específica.

3.2 Truncamiento de datos con distribuciones normales:

Siendo una de las propiedades de la distribución normal que puede ajustarse a cualquier valor dentro los números reales ($-\infty \leq X \leq \infty$), es decir tanto negativos, como positivos, habrá que tener precaución cuando los datos de alguna variable se ajusten a una distribución Normal. Para asegurarse de que los datos no tomen valores negativos, se debe truncar a la distribución para que la misma tome valores solo dentro de un rango permitido. (Hernández, 190).

Para truncar los datos de una distribución normal, para que éstos sean mayores o iguales que k se utiliza las siguientes ecuaciones:

$$\begin{aligned} E(X|X > k) &= \mu + \sigma\lambda(\alpha) \\ \text{Var}(X|X > k) &= \sigma^2[1 - \delta(\alpha)] \\ \alpha &= \frac{k-\mu}{\sigma} \end{aligned}$$

$$\begin{aligned} \phi(\alpha) &= \frac{\exp\left[-\frac{\alpha^2}{2}\right]}{\sqrt{2\pi}} \\ \Phi(\alpha) &= \int_{-\infty}^{\alpha} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \end{aligned}$$

$$\lambda(\alpha) = \frac{\phi(\alpha)}{[1-\Phi(\alpha)]}$$

$$\delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha]$$

En donde, si se reemplaza μ y σ^2 por los valores obtenidos de $E(X)$ y $\text{Var}(X)$, se asegura que la distribución normal solo tome valores mayores o iguales que k (Hernández y Garcés, 188).

Este procedimiento se aplicó para todas aquellas variables que presentaban normalidad en los datos.

3.3 Modelos de regresión:

3.3.1 Estimación de los parámetros por mínimos cuadrados

Con las variables de entrada y salida definidas, el modelo de regresión para cada producto está dado por:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

En donde $k=4$, denota cuatro variables aleatorias y se pretende determinar la relación existente entre Y y X , es decir entre la variable VENTAS con las variables regresoras (FF, A&P, DESCUENTOS, MM) para cada producto estudiado.

Se parte de la formulación de un sistema de ecuaciones, con $p= 4 + 1$ ecuaciones, para cada una de las variables regresoras desconocidas, lo cual permite realizar una estimación de los parámetros por mínimos cuadrados. Éste adopta la siguiente forma:

$$\begin{aligned}
n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{ik} &= \sum_{i=1}^n y_i \\
\hat{\beta}_0 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1}x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{i1}x_{ik} &= \sum_{i=1}^n x_{i1}y_i \\
\hat{\beta}_0 \sum_{i=1}^n x_{ik} + \hat{\beta}_1 \sum_{i=1}^n x_{ik}x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{ik}x_{i2} + \hat{\beta}_k \sum_{i=1}^n x_{ik}^2 &= \sum_{i=1}^n x_{ik}y_i
\end{aligned}$$

Se utiliza el método de los mínimos cuadrados, porque el software a utilizar Minitab, realiza la medición mediante este método.

Sin embargo, antes de plantear el modelo de regresión se parte de la definición de la prueba de hipótesis. Para todos los casos (5 productos) se probará dos combinaciones de hipótesis o pruebas; éstas serán:

- **Primera combinación: Prueba de significancia global de la regresión muestral.**

Con la prueba de significancia global lo que se quiere determinar es si Y, la variable de respuesta, está relacionada o no linealmente con las variables regresoras X_1, X_2, X_3, X_4 . (Damodar, 244).

No se realiza una prueba de significancia individual porque de hacerlo se trabajaría bajo el supuesto de que cada prueba de significancia estaba basada en una muestra diferente (Damodar, 245), mas esto no ocurre con los datos analizados, que provienen de una misma muestra de datos (pertenecientes a cada año estudiado). Por tanto las hipótesis planteadas serían las siguientes:

$$\begin{aligned}
H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 \\
H_0 : \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4
\end{aligned}$$

(Damodar, 244)

La técnica de estudiar a los coeficientes de manera conjunta se realiza mediante un análisis de varianza o también llamado ANOVA, el cual supone el siguiente planteamiento de hipótesis:

$$H_0 : \sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$$

$$H_0 : \sigma_1 \neq \sigma_2 \neq \sigma_3 \neq \sigma_4$$

Rechazando la hipótesis nula cuando el valor p del estudio es menor al nivel de significancia especificado (Damodar, 247). Hay que recalcar también que si la hipótesis nula es verdadera, se logra tener un verdadero valor de la varianza y con ello y la relación entre Y y X es trivial; la única fuente de variación en Y se debe a las fuerzas aleatorias representadas por el error del modelo. Por el contrario, si la hipótesis nula se rechaza, es decir las varianzas son distintas, implica que definitivamente las variables regresoras, X ejercen influencia en Y . (Damodar, 246).

Como se explicó antes, se estudiarán dos pruebas de hipótesis para cada producto; ANOVA, con el fin de determinar si las variables regresoras ejercen influencia en la variable de salida, y la otra prueba para determinar cuáles de las variables son la influyentes o no en el modelo, y en el caso de serlo cómo afecta cada una de ellas en el resultado. Para ambas pruebas de hipótesis el valor de α , o nivel de significancia es igual a 0.05.

Como los datos estudiados comprenden información mensual desde el año 2006 al 2009, y el cualquier modelo de regresión se ajusta mejor con un mayor número de datos se recurrió a una simulación Montecarlo para generar 5000 datos que se ajusten a la distribución específica de cada variable por cada producto. La razón por la cual se recurre a una simulación Montecarlo recae en el hecho de que se trabaja con datos continuos, es decir un sistema en el cual las

variables dependientes del modelo cambian continuamente a lo largo del tiempo (Banks, 12). Las variables aquí analizadas pueden tomar cualquier valor, pues se tratan todas de cantidades dadas en dólares, que pueden ser en un rango desde \$0 hasta infinito, cambiando continuamente su cantidad.

Con la ayuda de un complemento de Microsoft Excel diseñado específicamente para formar una simulación Montecarlo, se generaron los 5000 datos bajo la distribución adecuada de cada variable (inferidos con anterioridad) y con ellos se corrió el modelo de regresión.

La solución a la cual llega Minitab se resuelve por el método de los mínimos cuadrados, para el cual se deben probar todos los supuestos que se expusieron con anterioridad en este trabajo. Esto demostrará la validez del modelo usado y se podrán obtener mejores conclusiones al respecto.

Para asegurar la confidencialidad de las cifras que maneja Novartis para cada variable, los datos han sido multiplicados por un factor aleatorio; con esto no se presentan las cantidades reales de los gastos que la farmacéutica incurre en los valores aleatorio generados, más la simulación, regresión y optimización fueron realizados en función de las cantidades reales.

A continuación se muestran los resultados obtenidos tanto para la inferencia, generación y elaboración del modelo para cada producto estudiado.

3.4 Análisis en productos

3.4.1 Producto VOLTAREN

Voltaren es una droga anti- inflamatoria y analgésica, clasificado como un producto maduro que pertenece a la línea de productos Phrama dentro de la

familia de productos Novartis (Voltaren). Su posición en el mercado es fuerte, y se trata de un producto de tipo OTC, es decir de venta sin receta (en percha). Por esta razón Voltaren no cuenta con la variable Fuerza de Ventas, y a pesar de que el algún año lo tuvo, la recolección de estos datos es muy escasa para plantear una inferencia en ellos. Por tanto, las variables estudiadas en su modelo de regresión son: A&P, Descuentos y Muestras Médicas. La inferencia en la distribución de cada una se presenta a continuación:

3.4.1.1 Inferencia para la variable de salida Ventas de Voltaren

Se estudiaron los datos correspondientes a las ventas de Voltaren de los últimos cuatro años, es decir ventas mensuales desde el 2006 hasta el 2009. Con un $\alpha=0.05$, el resultado de Arena fue una distribución Normal con la siguiente gráfica y resultados:

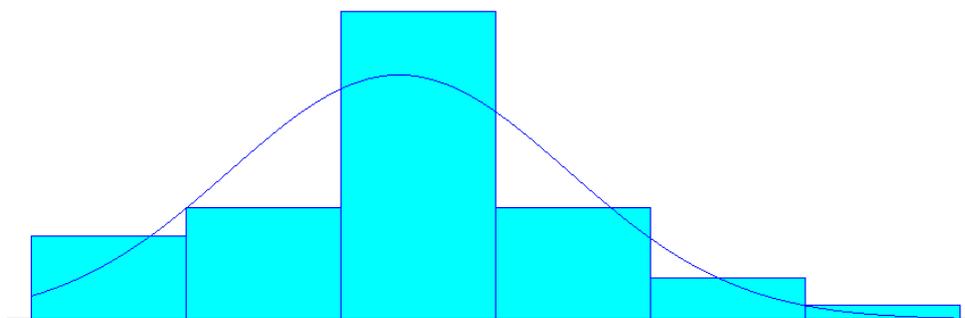


Figura 24. Distribución de las Ventas del producto Voltaren. Fuente: Arena.

Distribution Summary

Distribution: Normal

Expression: NORM(3.49e+005, 5.84e+004)

Square Error: 0.024496

Kolmogorov-Smirnov Test

Test Statistic = 0.102

Corresponding p-value > 0.15

Data Summary

Number of Data Points	= 48
Min Data Value	= 2.22e+005
Max Data Value	= 5.42e+005

Sample Mean = 3.49e+005
 Sample Std Dev = 5.9e+004

Histogram Summary

Histogram Range = 2.22e+005 to 5.42e+005
 Number of Intervals = 6

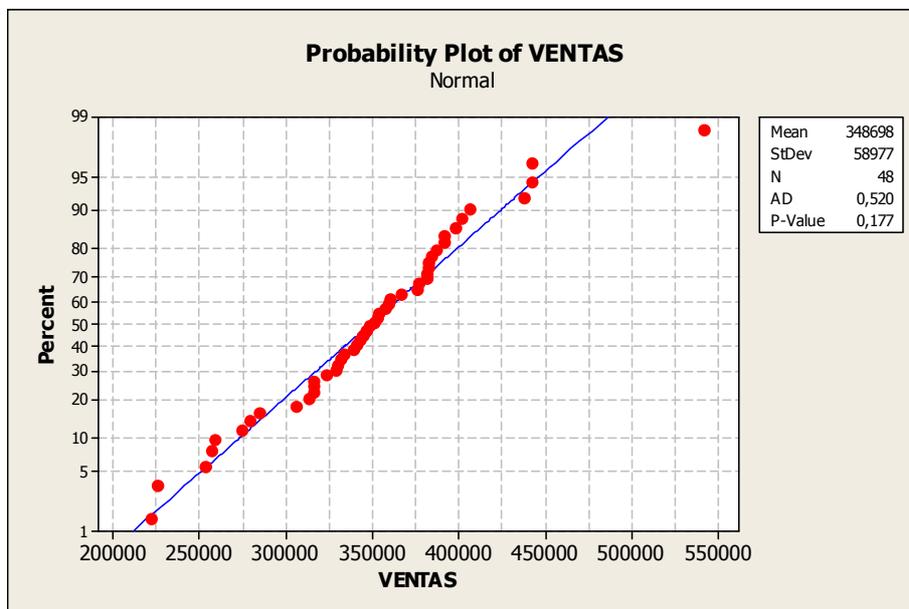


Figura 25. Gráfica de normalidad de los datos correspondientes a Ventas del producto Voltaren.
 Fuente: Minitab.

Tanto para la gráfica de normalidad de los datos, como para la prueba de Kolgomorov Smirnov, el valor p es mayor al nivel de significancia igual a 0.05, con lo que no se puede rechazar la hipótesis nula. Con ello se dice que los datos sí se ajustan a una distribución normal. De igual manera el error obtenido es bajo, lo que ratifica el resultado obtenido con la prueba de hipótesis y el gráfico presentado.

Se nota además que la única prueba de bondad usada es la prueba Kolgomorov Smirnov, ya que esta sirve para pruebas de ajuste con datos de una muestra pequeña. La prueba Chi cuadrado es normalmente usada cuando se tiene un número de datos grande (entiéndase grande como un $N > 50$ datos (Banks, Carson II et al, 125). La muestra con la que estudió la distribución consistía de 48 datos, por lo que la prueba Chi cuadrado no es aplicable. Sin

embargo la prueba Kolgomorov Smirnov es la más poderosa de las dos y tiene un grado de exactitud bastante bueno aunque el número de datos sea pequeño (Banks, Carson II et all, 125).

A partir de la distribución normal obtenida con parámetros de media y desviación estándar, se hace uso de la simulación Montecarlo para la generación aleatoria de 5000 datos con este tipo de distribución. Con ellos se generará un modelo de regresión y optimización que se verá más adelante.

3.4.1.1.1 Inferencia para los datos generados de la variable de salida

Ventas de VOLTAREN

Por medio de la simulación Montecarlo que se basa en la generación de datos aleatorios en base a una distribución dada, se generaron 5000 datos con distribución de probabilidad normal. Con un alfa de igualmente 0.05 se comprobó que los datos generados siguen dicha distribución. Se obtuvo lo siguiente:

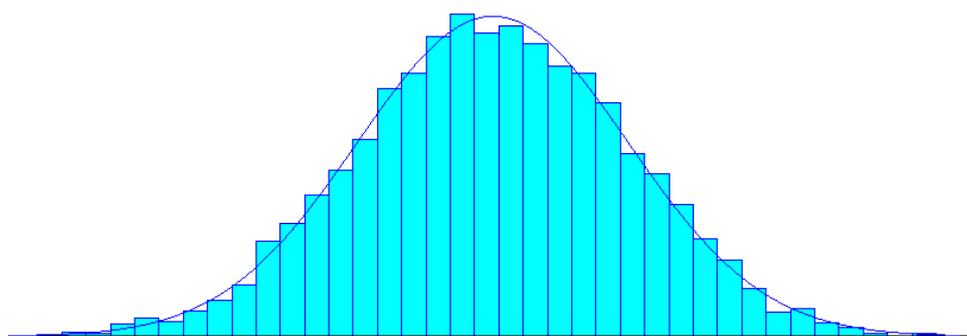


Figura 26. Distribución para los datos generados de la variable Ventas del producto Voltaren.
Fuente: Arena.

Distribution Summary

Distribution:Normal

Expression: NORM(3.5e+005, 5.8e+004)

Square Error:0.000099

Chi Square Test

Number of intervals = 32

Degrees of freedom = 29

Test Statistic = 31.3

Corresponding p-value = 0.36

Kolmogorov-Smirnov Test
 Test Statistic = 0.00811
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = 1.34e+005
 Max Data Value = 5.51e+005
 Sample Mean = 3.5e+005
 Sample Std Dev = 5.8e+004

Histogram Summary

Histogram Range = 1.34e+005 to 5.51e+005
 Number of Intervals = 40

Con un valor p, para ambas pruebas, pues ahora la Chi cuadrado sí es aplicable (N/5000), mayor al nivel de significancia no se rechaza la hipótesis nula y los datos aleatorios generados se ajustan a una distribución normal.

3.4.1.2 Inferencia para la variable de entrada Advertisement & Promoting de VOLTAREN

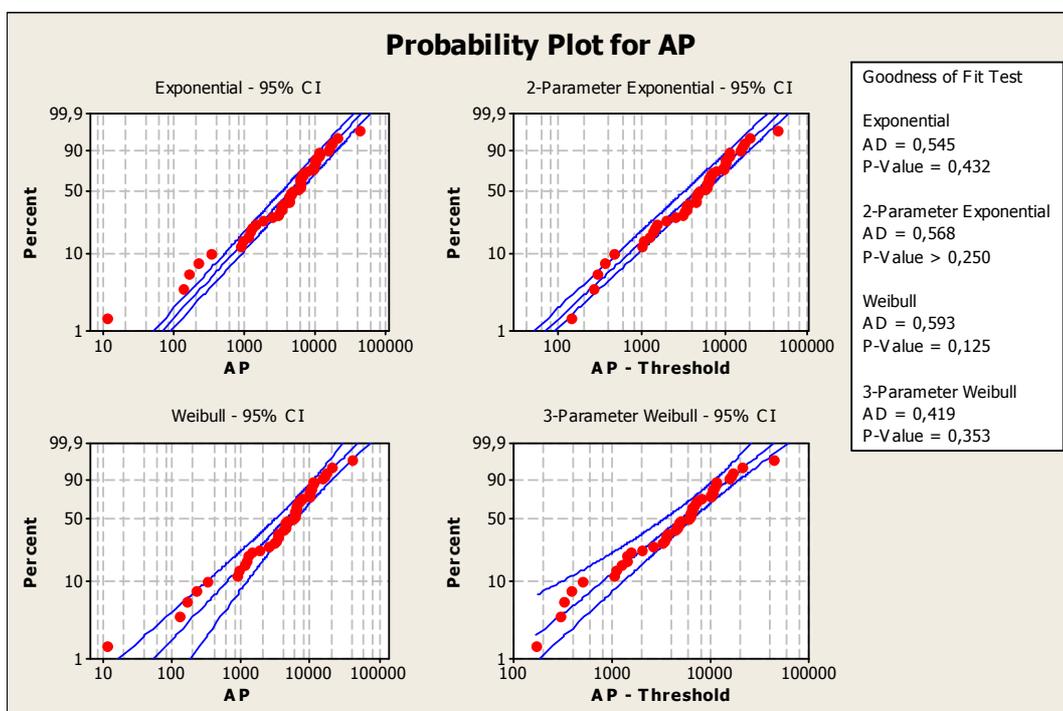


Figura 27. Distribución de A&P del producto Voltaren. Fuente: Minitab.

El valor p más grande de todas las distribuciones posibles a ser ajustadas, pertenece a una distribución Exponencial, con un valor $p = 0.432$. Éste no permite rechazar la hipótesis nula y por tanto se concluye que de todas las distribuciones de probabilidad posibles, la más adecuada es la Exponencial con una media igual a 6717,72396.

Para este caso la inferencia que mostró mejores resultados con un valor p mayor comparado al nivel de significancia de 0,05 fue el resultado generado por Minitab. El resultado de Arena mostraba un valor p más bajo que el mostrado y por ello se recurrió a determinar que la distribución más adecuada para la variable Ventas es una distribución exponencial. Al ajustar específicamente esta distribución en Arena, se obtiene lo siguiente:

```
Distribution Summary

Distribution: Exponential
Expression: 11 + EXPO(6.71e+003)
Square Error: 0.005568

Chi Square Test
Number of intervals      = 3
Degrees of freedom       = 1
Test Statistic           = 0.826
Corresponding p-value    = 0.393

Kolmogorov-Smirnov Test
Test Statistic           = 0.0976
Corresponding p-value    > 0.15
```

Donde si bien el valor p para las pruebas de ajuste difiere, ambos son mayores al nivel de significancia, por lo que la distribución Exponencial es la adecuada. Se recalca que en ambos software se mantiene el valor de su parámetro, la media, muy similares.

3.4.1.2.1 Inferencia para los datos generados de la variable de entrada

Advertisement & Promoting de VOLTAREN

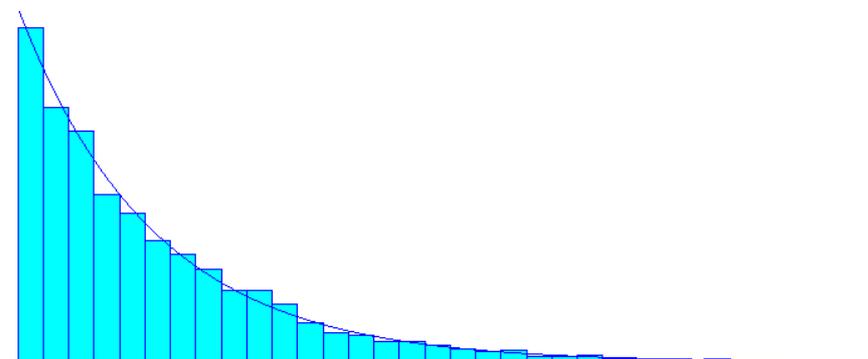


Figura 28. Distribución para los datos generados de la variable A&P del producto Voltaren. Fuente: Arena.

Distribution Summary

Distribution: Exponential
Expression: 2 + EXPO(6.77e+003)
 Square Error: 0.000284

Chi Square Test

Number of intervals = 28
 Degrees of freedom = 26
 Test Statistic = 37.5
Corresponding p-value = 0.0707

Kolmogorov-Smirnov Test

Test Statistic = 0.0107
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = 2.7
 Max Data Value = 5.07e+004
 Sample Mean = 6.77e+003
 Sample Std Dev = 6.71e+003

Histogram Summary

Histogram Range = 2 to 5.07e+004
 Number of Intervals = 40

Tanto la prueba Kolmogorov como la prueba Chi Cuadrado muestran que los datos generados se ajustan a la distribución requerida y sirven como variables de entrada para la regresión del modelo.

3.4.1.3 Inferencia en la variable de entrada Descuentos en el producto Voltaren

Con datos desde el 2006 hasta el 2009, se estudió la distribución de la variable Descuentos para el producto Voltaren. Igualmente con un alfa de 0.05 se obtuvo que los datos se ajusten a una distribución Normal.

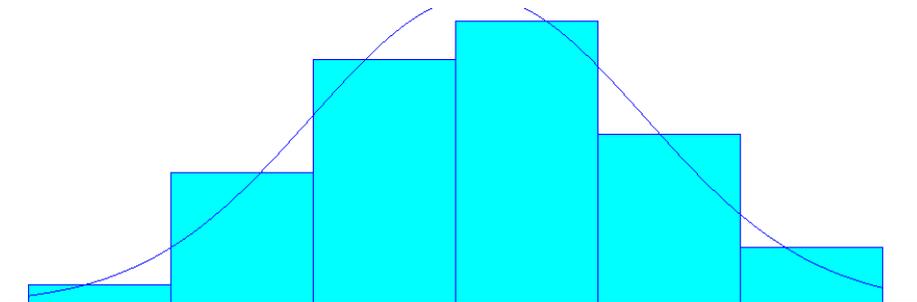


Figura 29. Distribución de Descuentos del producto Voltaren. Fuente: Arena

Distribution Summary

Distribution:Normal

Expression: NORM(1.15e+005, 2.67e+004)

Square Error:0.000807

Kolmogorov-Smirnov Test

Test Statistic = 0.0549

Corresponding p-value > 0.15

Data Summary

Number of Data Points	= 48
Min Data Value	= 4.26e+004
Max Data Value	= 1.79e+005
Sample Mean	= 1.15e+005
Sample Std Dev	= 2.7e+004

Histogram Summary

Histogram Range	= 4.26e+004 to 1.79e+005
Number of Intervals	= 6

Como en las otras variables, la prueba Kolmogorov es la necesaria para comprobar la distribución de los datos a una Normal. Los parámetros obtenidos, junto con la simulación Montecarlo serán usados para la generación de 5000 datos de la variable de entrada Descuentos. La inferencia a éstos para comprobar el ajuste a la distribución Normal, se muestra a continuación.

3.4.1.3.1 Inferencia a los datos generados de la variable Descuentos del producto Voltaren.

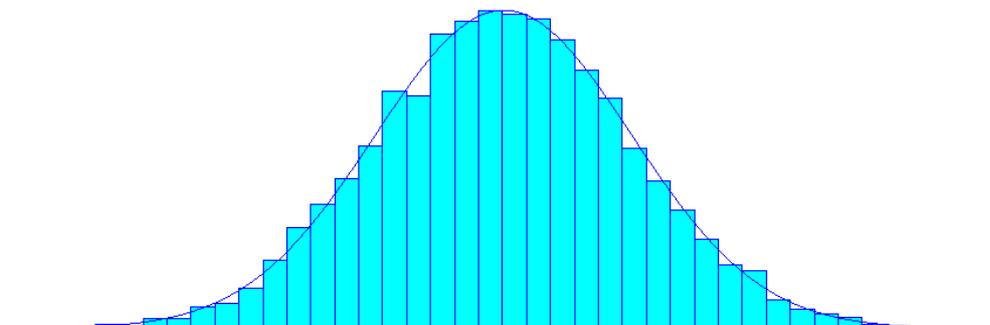


Figura 30. Distribución para los datos generados de la variable Descuentos del producto Voltaren.
Fuente: Arena.

Distribution Summary

Distribution: Normal

Expression: **NORM(1.15e+005, 2.65e+004)**

Square Error: 0.000079

Chi Square Test

Number of intervals = 30
 Degrees of freedom = 27
 Test Statistic = 14.9
Corresponding p-value > 0.75

Kolmogorov-Smirnov Test

Test Statistic = 0.00684
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = 1.59e+004
 Max Data Value = 2.14e+005
 Sample Mean = 1.15e+005
 Sample Std Dev = 2.65e+004

Histogram Summary

Histogram Range = 1.59e+004 to 2.14e+005
 Number of Intervals = 40

Un error cuadrado bajo obtenido, junto con los valores p de cada prueba de ajuste, Kolmogorov y Chi Cuadrado prueban que los datos generados siguen una distribución normal con los parámetros del ajuste de los datos primeramente inferidos.

3.4.1.4 Inferencia en la variable de entrada Muestras Médicas del producto Voltaren

Con datos desde el 2006 al 2009 se infirió en la distribución adecuada, obteniendo una distribución Exponencial. Los resultados generados por Arena, con un alfa de 0.05 se muestran a continuación:

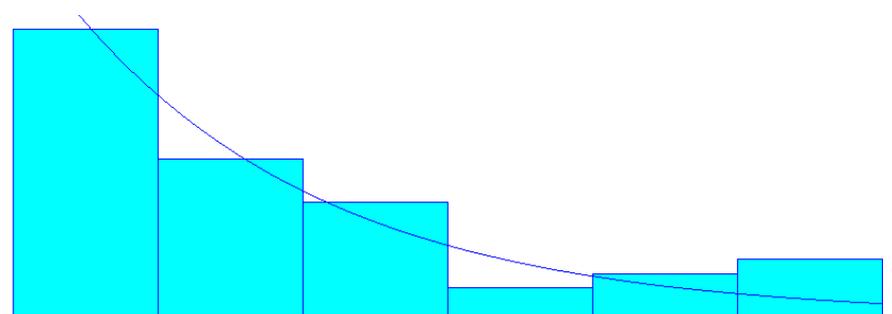


Figura 31. Distribución de Muestras Médicas del producto Voltaren. Fuente: Arena.

Distribution Summary

Distribution: Exponential
Expression: $-0.001 + \text{EXPO}(2.85e+004)$
 Square Error: 0.006330

Kolmogorov-Smirnov Test

Test Statistic = 0.959
Corresponding p-value > 0.632

Data Summary

Number of Data Points = 48
 Min Data Value = 0
 Max Data Value = $9.7e+004$
 Sample Mean = $2.85e+004$
 Sample Std Dev = $2.96e+004$

Histogram Summary

Histogram Range = -0.001 to $9.7e+004$
 Number of Intervals = 6

Un valor p igual mayor a 0.632 de la prueba Kolmogorov muestra que el ajuste de los datos a una distribución exponencial es válido. Junto con los parámetros correspondientes a esta distribución y una simulación Montecarlo, se

generaron 5000 datos. La inferencia en ellos para comprobar que se ajustan a una distribución exponencial se muestra a continuación.

3.4.1.4.1 Inferencia a los datos generados de la variable Muestras Médicas del producto Voltaren

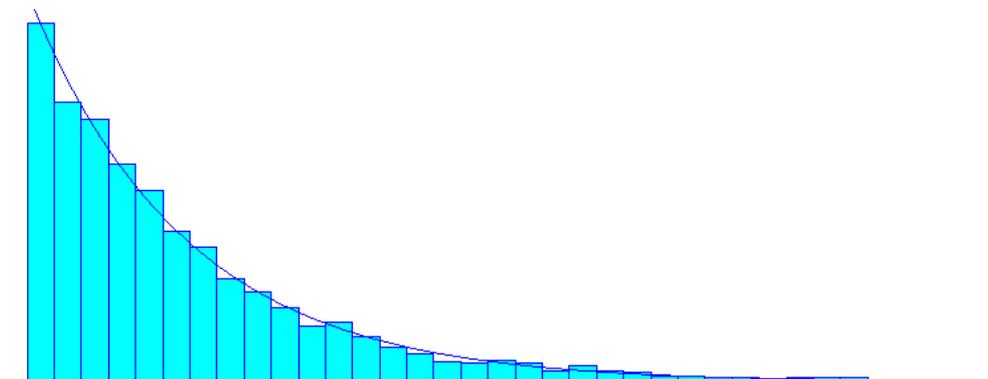


Figura 32. Distribución para los datos generados de la variable Muestras Médicas del producto Voltaren. Fuente: Arena.

Distribution Summary

Distribution: Exponential

Expression: **EXPO(2.83e+004)**

Square Error: 0.000159

Chi Square Test

Number of intervals = 28

Degrees of freedom = 26

Test Statistic = 22.7

Corresponding p-value = 0.649

Kolmogorov-Smirnov Test

Test Statistic = 0.00945

Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000

Min Data Value = 0.746

Max Data Value = 2.1e+005

Sample Mean = 2.83e+004

Sample Std Dev = 2.8e+004

Histogram Summary

Histogram Range = 0 to 2.1e+005

Number of Intervals = 40

Los valores p de ambas pruebas de ajuste hacen que la hipótesis nula no se rechace y que por tanto los datos sí se ajusten a una distribución exponencial.

3.4.1.5 Modelo de regresión con los datos generados para el producto Voltaren

En base a las distribuciones inferidas de cada variable para el producto Voltaren, y la generación Montecarlo de 5000 datos que se ajusten a dichas distribuciones, se procedió a correr un modelo de regresión múltiple. Por medio de éste se determinará cuál es el impacto (si es que existe) de las variables de entrada, sobre la variable de salida, Ventas.

Como se dijo antes, las hipótesis a probar son las siguientes:

$$H_0 : \sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$$

$$H_0 : \sigma_1 \neq \sigma_2 \neq \sigma_3 \neq \sigma_4$$

Mediante ese juego de hipótesis se sabrá si existe o no una variable de entrada que afecte de manera significativa en el modelo de regresión obtenido. Al rechazar la hipótesis nula, es decir cuando se tiene un valor p menor al alfa de 0,05 existirá una variable regresora que influye de manera importante en el modelo y que no puede ser eliminada del mismo.

Ahora bien para conocer cuál de las variables es la que causa una influencia significativa en la variable de respuesta, se prueba lo siguiente:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$$

$$H_0 : \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4$$

Con un valor p menor al valor alfa establecido, se rechaza la hipótesis nula que establece,

H₀ = El factor no influye de manera significativa en el modelo.

H₁ = El factor sí influye de manera significativa en el modelo.

Por tanto, con valores menores a 0,05 se dirá que el factor correspondiente afecta al modelo de manera no trivial. A continuación se muestran los resultados

obtenidos del modelo de regresión múltiple para el producto Voltaren con sus tres variables regresoras: Advertisement & Promoting, Descuentos y Muestras Médicas.

Primero se analizará el gráfico de residuales obtenido:

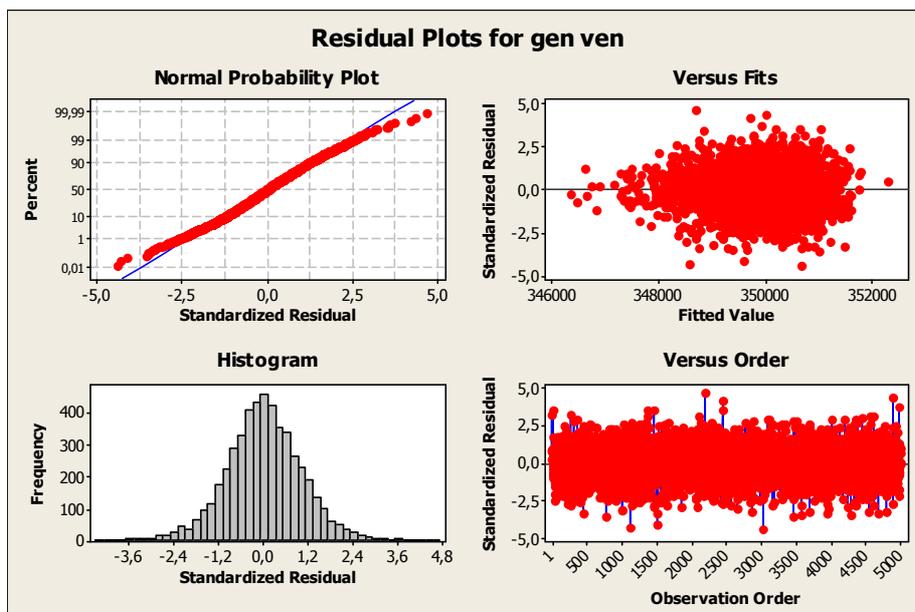


Figura 33. Residuales del modelo de regresión múltiple para el producto Voltaren. Fuente: Minitab.

Al parecer los residuales se acercan a una distribución normal, pero al graficarlos de manera individual se obtiene la gráfica a continuación:

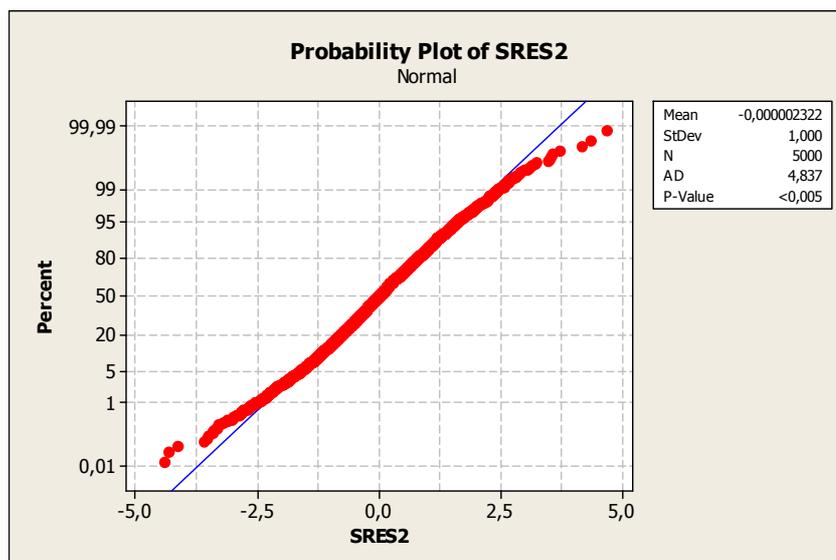


Figura 34. Prueba de normalidad de los residuales del modelo de regresión múltiple para el producto Voltaren. Fuente: Minitab.

Donde el valor p obtenido es menor al nivel de significancia por lo que no se puede aceptar que los residuales pertenecen a una distribución normal y es necesario realizar un ajuste a la variable de salida, con el fin de estabilizar su varianza y obtener resultados correctos de la regresión. En este caso la transformación adecuada para la transformación de la variable Ventas fue una transformación de Johnson. Este tipo de transformación requiere que los datos aleatorios generados de la variable de salida Ventas, sean transformados en unos nuevos, tales que, la distribución de sus residuales sea normal. Se usa entonces la siguiente fórmula:

$$-0.0490139 + 2.22046 * \operatorname{Arcsenh}\left(\frac{X - 347416}{116889}\right)$$

La gráfica de ajuste fue la siguiente, en la que se observa que el valor p del ajuste a una distribución normal es mayor que el nivel alfa, y por lo tanto la transformación de Johnson es la adecuada:

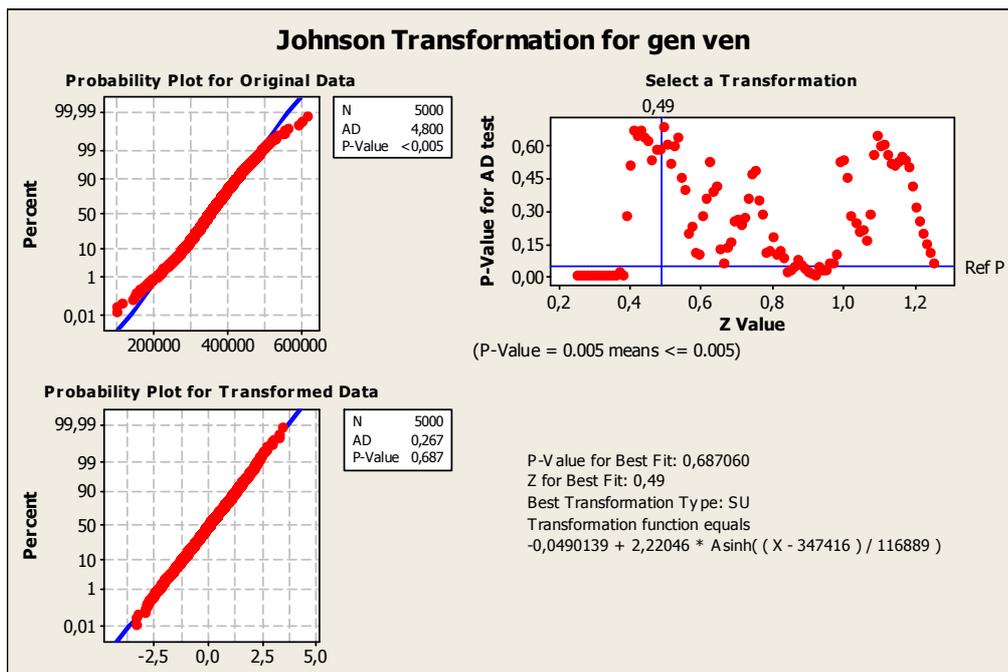


Figura 35. Transformación de Johnson de los datos generados de la variable Ventas de Voltaren. Fuente: Minitab.

Con la fórmula obtenida se transformaron todos los datos generados de la variable Ventas de Voltaren, para ahora sí poder correr un modelo de regresión adecuado que se ajuste a todos los supuestos del Método de los mínimos cuadrados. En él se probará qué variables son significativas para las Ventas de Novartis, qué variables no lo son y si existe correlación entre alguna de ellas. Con el modelo de regresión obtenido se tendrá la función objetivo del modelo lineal que se quiere optimizar, convirtiendo los valores obtenidos, a la inversa de la ecuación de Johnson obtenida, para asegurar que los datos pertenezcan al intervalo de valores originales.

Con la ayuda del software Minitab se realizó una regresión múltiple y se obtuvo lo siguiente:

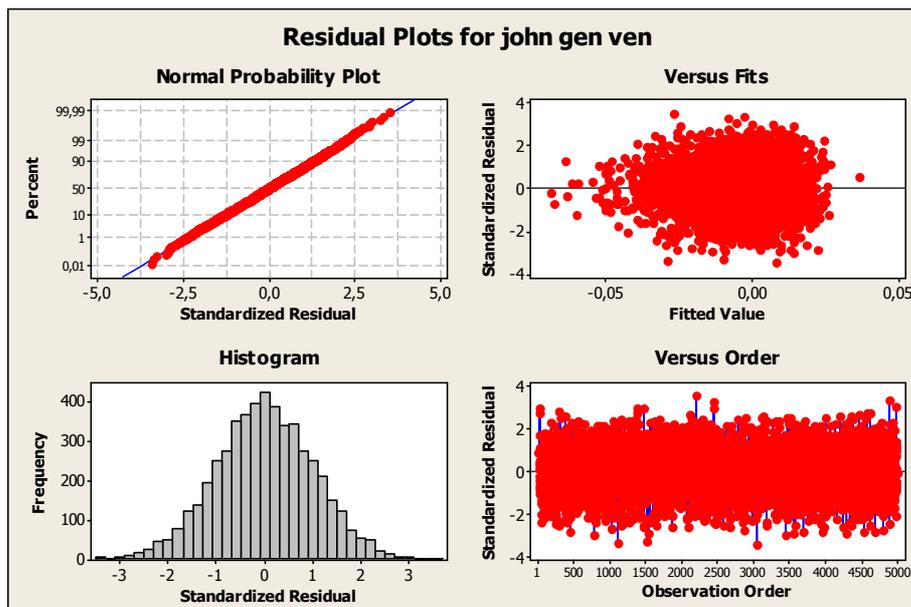


Figura 36. Residuales del modelo de regresión múltiple para el producto Voltaren con los datos transformados. Fuente: Minitab.

Regression Analysis: john gen ven versus gen ap. gen des. gen mm

The regression equation is

$$\text{john gen ven} = 183252 + 0,921 \text{ gen ap} + 1,38 \text{ gen des} + 0,050 \text{ gen mm}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	183252	31921	5,74	0,000	
gen ap	0,9211	0,9666	0,95	0,346	1,008
gen des	1,3770	0,2562	5,37	0,000	1,009
gen mm	0,0498	0,2332	0,21	0,832	1,001

S = 47246,1 R-Sq = 39,9% R-Sq(adj) = 35,8%

PRESS = 112129221528 R-Sq(pred) = 31,41%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	65263533663	21754511221	9,75	0,000
Residual Error	44	98216390249	2232190687		
Total	47	1,63480E+11			

Source	DF	Seq SS
gen ap	1	505575037
gen des	1	64656312173
gen mm	1	101646452

En donde:

John gen ven: Representa a los datos generados de Ventas y transformados mediante el algoritmo de Johnson.

Gen ap: representa la generación de datos de la variable A&P.

Gen des: representa la generación de datos de la variable Descuentos.

Gen mm: representa la generación de datos de la variable Muestras Médicas.

Analizando en primera instancia a la gráfica de residuales como era de esperarse, éstos siguen una distribución normal. Se ve claramente que los residuos se ajustan perfectamente a la recta y de igual manera y mediante el histograma, se comprueba la existencia de dicha distribución con la forma de campana que siguen los datos en el histograma. Se concluye que éstos siguen una distribución normal y de esta manera el supuesto 11 se satisface. En cuanto a las dos gráficas restantes, Residuos vs. Ajustes y Orden, respectivamente, no se distingue ningún patrón que llame la atención en los residuos resultantes, además que no se acercan a la línea de referencia lo cual indica parcialmente que sí tienen una varianza constante. Este último hecho hace que el supuesto 4, de la varianza del residuo constante sea verdadero.

Para la aceptación de los demás supuestos se recurrió a la obtención de un par de estadísticos que explican si el supuesto se cumple o no. Para el supuesto 3, la media de los errores es igual a cero, Damodar en su libro “Econometría” dice que cuando este supuesto no se cumple no existe la constante de la pendiente, es decir β_1 es igual a cero; en el modelo obtenido sí se lo obtiene a β_1 . Por tanto el supuesto 3 sí se cumple.

Para analizar el supuesto de no multicolinealidad entre las variables regresoras, se recurrió a obtener el Factor inflador de la varianza (FIV). Éste indica la velocidad con la cual las varianzas y covarianzas se incrementan, es decir muestra como la varianza de un estimador es inflada por la presencia de la multicolinealidad. (Damodar, 337). Cuando el grado de colinealidad aumenta, la varianza de un estimado aumenta y el límite puede volverse infinito. Cuando no

existe colinealidad entre las variables regresoras participantes, el FIV de cada una de ellas será igual a 1. (Damodar 337). Los resultados obtenidos muestran que el VIF de cada variable regresora es igual o con una mínima desviación a 1, por lo que se concluye que no existe multicolinealidad en los datos y el supuesto 5 y por lo tanto el 6 quedan satisfechos.

Por su lado, el supuesto 1, que dice que el modelo de regresión es lineal en los parámetros, se ve que todos ellos mantienen coeficientes lineales y el supuesto se cumple.

Haciendo la regresión sobre 5000 datos sobre 3 variables de interés, el supuesto 7 queda satisfecho. Los demás supuestos 8 y 9, pueden presumirse como correctos con los resultados expuestos por Minitab.

Una vez que los supuestos del Método de Mínimos cuadrados se cumplen, se debe analizar los resultados obtenidos de la regresión. Analizando primeramente los resultados del análisis de varianza, el que indica si existe o no una variable de importante influencia en el modelo, el valor p obtenido, igual a cero, muestra la existencia de al menos una variable regresora no trivial en el modelo.

Si se observa la ecuación de la regresión obtenida, ésta muestra un coeficiente más grande para la variable Descuentos, y el más pequeño para la variable Muestras Médicas. Analizando la tabla de valores obtenidos para cada coeficiente, se comprueba que para la variable Descuentos el valor p resultante es cero, que al ser menor que 0.05 (nivel de significancia) permite rechazar la hipótesis nula y establecer que esta variable es la única representativa para la ecuación de regresión de Voltaren. Además mediante el signo obtenido (positivo)

se determina que la relación entre Ventas y Descuentos es directamente proporcional.

R^2 -adj muestra el grado de ajuste del modelo a los datos, que en este caso corresponde a un 35,8%. Si bien no es un valor “elevado”, tal como se expuso en el Marco teórico de esta tesis, este estadístico no es la única prueba de una correcta validación del modelo. Lo que se puede destacar de esta observación es que quizá con la implementación de más variables en el análisis, el modelo pueda ajustarse de mejor manera y tener unos resultados que se acerquen mejor al comportamiento de cada variable estudiada.

Ahora bien, siendo la variable Descuentos la de mayor importancia en el producto Voltaren, quizá sea importante aclarar que éste es un producto que pertenece a la línea de OTC, es decir un producto que se vende sin receta médica. ¿Qué implicaciones tiene esto sobre el modelo? Se podría decir que Voltaren es un medicamento con una larga trayectoria en el mercado ecuatoriano, en donde la marca es altamente reconocida. Si bien las demás variables, Advertisement & Promoting y Muestras Médicas, son importantes para las ventas de Voltaren, al ser un producto que mantiene una fuerte influencia en el mercado, el modelo permite declarar que el aspecto más importante en la variable de salida son los descuentos ofrecidos en el producto.

Más adelante se podrá determinar qué valor es el óptimo, no sólo para esta variable, sino para todas las demás, fijando especial atención en el resultado que el modelo de optimización adopte para la variable Descuentos.

Capítulo 4: Modelos de optimización

Ahora que se han obtenido la regresión para el producto Voltaren y todos los demás presentados en los Anexos, se procede a calcular los valores óptimos de cada variable regresora con el fin de generar ventas máximas para cada producto. De igual manera a continuación se calculan los niveles óptimos solo para el producto Voltaren, y en forma de ANEXOS se presenta el estudio para los otros cuatro productos. Éstos están en los ANEXOS C-1, C2, C-3, C-4.

Se decidió correr dos modelos de optimización, uno mediante el complemento de Excel Solver, y el siguiente a partir de una simulación Montecarlo y el complemento MCSim- Solver que presenta también las hojas de cálculo de Excel. El propósito de realizar dos modelos es el de encontrar cuál de los dos se ajusta de mejor manera a cada variable y producto y a su vez, cuál de estos genera mejores y más razonables resultados.

4.1 Modelo de optimización con Solver (sin simulación Montecarlo)

El modelo que se utilizará es uno de optimización lineal, igual para todos los productos, en los que sólo variará la función objetivo y el número de variables participantes, de acuerdo a lo antes presentado.

Se parte entonces de una función objetivo, la cual será la ecuación obtenida en la regresión a partir de los 5000 datos generados. Tiene la siguiente estructura:

$$MaxY = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4$$

En donde:

Y = Representa a las ventas generadas en función de las variables X s de cada producto.

β_i = Es el coeficiente obtenido para cada variable. Obtenidos mediante la regresión múltiple aplicada.

X_i = Representa al gasto incurrido por cada variable regresora en cada producto.

Son las celdas cambiantes a determinar en el modelo.

Toda la función objetivo de todos los productos es la resultante de la regresión lineal múltiple obtenidas en el capítulo anterior.

Como todo modelo de optimización lineal, junto a la función objetivo deben constar restricciones del modelo que permitan fijar los valores de X en cantidades adecuadas y razonables para la función Ventas. En el modelo a desarrollarse se presentan solamente dos restricciones; son:

1. No negatividad: los valores de X no pueden tomar valores menores a cero.

Si el valor de la variable resulta en cero se entenderá que no se debe incurrir en ningún gasto para esa variable, mientras que si toma un valor mayor a cero, será el gasto adecuado. Este valor nunca podrá ser menor a cero y por ende una cantidad negativa.

2. El modelo de optimización lineal que se plantea comprende variables continuas, es decir variables que pueden tomar cualquier valor a lo largo del tiempo. Mas como el objetivo de esta tesis es producir más ventas de las medicinas, aumentando la cantidad vendida o bien reduciendo los costos de las variables influyentes, se han fijado valores máximos a los que cada variable por producto, puede alcanzar. Estos valores fueron tomados en base a históricos desde el año 2006 al 2009, en donde se determinó la

cantidad de Ventas mayor generada, y con ello los gastos en cada variable regresora incurridos. A estos valores se los fijo como máximos, de manera que se sepa que si con ellos se alcanzaron Ventas máximas, los valores de las variables regresoras pueden ser menores y alcanzar los mismos o muy parecidos niveles de ventas.

En base a esto, se planteará en Solver el modelo de optimización lineal que se ajuste a las restricciones de cada producto. De manera general en una hoja de cálculo, adopta la siguiente forma:

Función Objetivo

Ecuación de la regresión

0

Variables

X1	gastos en A&P	
X2	gastos en Descuentos	
X3	gastos en Muestras Médicas	

Coefficientes

Constante	
X1	
X2	
X3	

Celdas Cambiantes:

x1	x2	x3

Restricciones

0	<=	máximo
0	<=	máximo
0	<=	máximo
0	>=	0
0	>=	0
0	>=	0

Generación propia.

Figura 96: Hoja de cálculo para resolver el problema de optimización mediante Solver, Excel.

A continuación los resultados obtenidos para Voltaren:

Ya los datos obtenidos de la regresión se procede a llenar el formato de resolución antes mostrado; es decir se fija la Función Objetivo, los coeficientes de cada variable regresora y el de la constante, y los máximos posibles por cada variable. Definiendo la celda objetivo en el Solver, junto con las respectivas restricciones de no negatividad y máximos límites, se elige adoptar un modelo lineal para que el complemento Solver halle la solución más adecuada.

Para el caso de Voltaren se obtuvieron los siguientes resultados:

Gastos en A&P	\$ 117725,9
Gastos en Descuentos	\$ 1485734,33
Gastos en Muestras Médicas	\$ 3535019,5

Generación propia.

Tabla 2. Niveles óptimos para cada variable del producto Voltaren

Reemplazando estos niveles en la función objetivo se obtiene unas Ventas de \$ 2353741,9.

4.1 Modelo de optimización con MC-Sim Solver (con simulación Montecarlo)

Tal como se explicó antes, son cinco los pasos requeridos para llevar a cabo una simulación Montecarlo. Con la regresión obtenida en el apartado anterior, se ha definido ya la función matemática o modelo matemático necesario. Ahora y mediante las hojas de cálculo de Excel es fácil generar variables de entrada aleatorias que se ajusten a las distribuciones de datos antes inferidos. Así para el caso de Voltaren, las 3 variables de entrada deberán ajustarse a las siguientes distribuciones:

Variable	AP	DESCUENTOS	MM
Distribución	EXPO(6.77e+003)	NORM(1.15e+005, 2.65e+004)	EXPO(2.83e+004)

Generación propia.

Tabla 3. Distribuciones para las variables del producto Voltaren

A partir de estas distribuciones se crearon 1000 datos aleatorios correspondientes a cada variable.

El tercer paso de la simulación Montecarlo comprende la evaluación del modelo, guardando las variables de salida resultantes. Para esto se reemplazó los valores de cada variable de entrada en la función de rentabilidad obtenida de la regresión. El cuarto paso es el análisis de los datos. Para esto se realizó gráficos para observar el comportamiento de cada variable de manera individual con la variable de salida Ventas.

- **A&P con Ventas:**

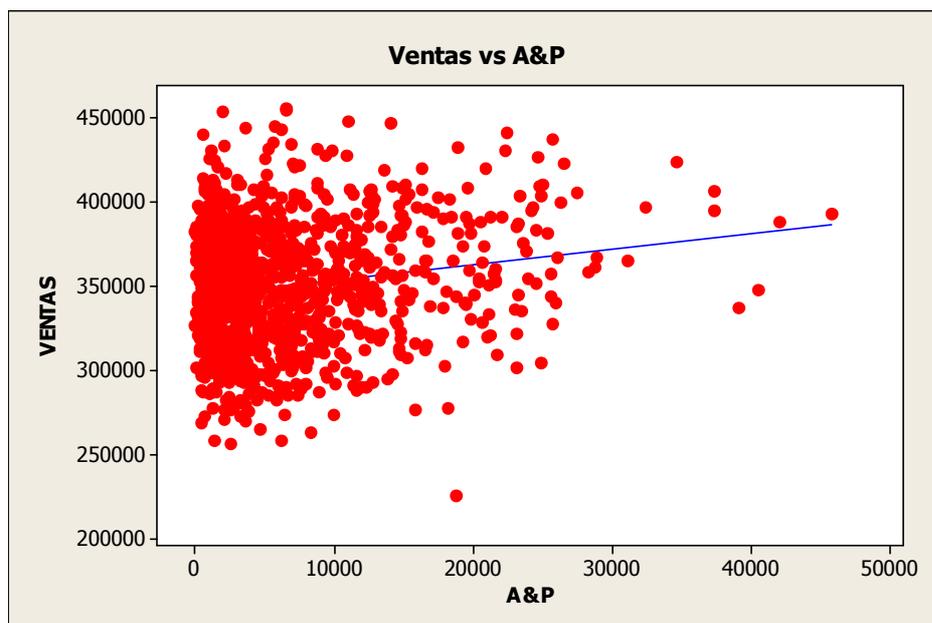


Figura 97. Gráfica de correlación entre la variable A&P y Ventas de Voltaren. Fuente: Minitab.

La gráfica de correlación indica que existe cierta relación entre los valores bajos de A&P y las ventas de Voltaren. Esto puede resultar importante ya que indica que un gasto excesivo en Medios de Difusión para este producto no tiene efectos favorables a las ventas, información que se complementa con los

resultados obtenidos de la regresión múltiple en donde A&P no resultó ser una variable significativa.

- **Descuentos vs. Ventas:**

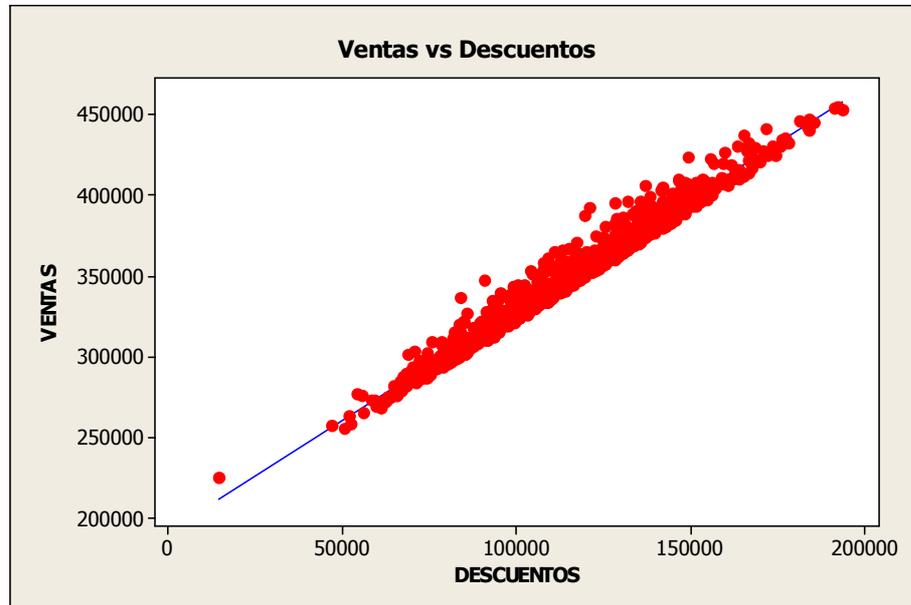


Figura 98. Gráfica de correlación entre la variable Descuentos y Ventas de Voltaren. Fuente: Minitab.

Claramente se ve una relación entre estas dos variables, tal como se obtuvo en la ecuación de regresión ya que las observaciones siguen un patrón. Además se ve también que la relación es directamente proporcional, es decir mientras más descuentos se ofrecen por Voltaren, más Ventas se generan.

- **Muestras Médicas vs. Ventas:**

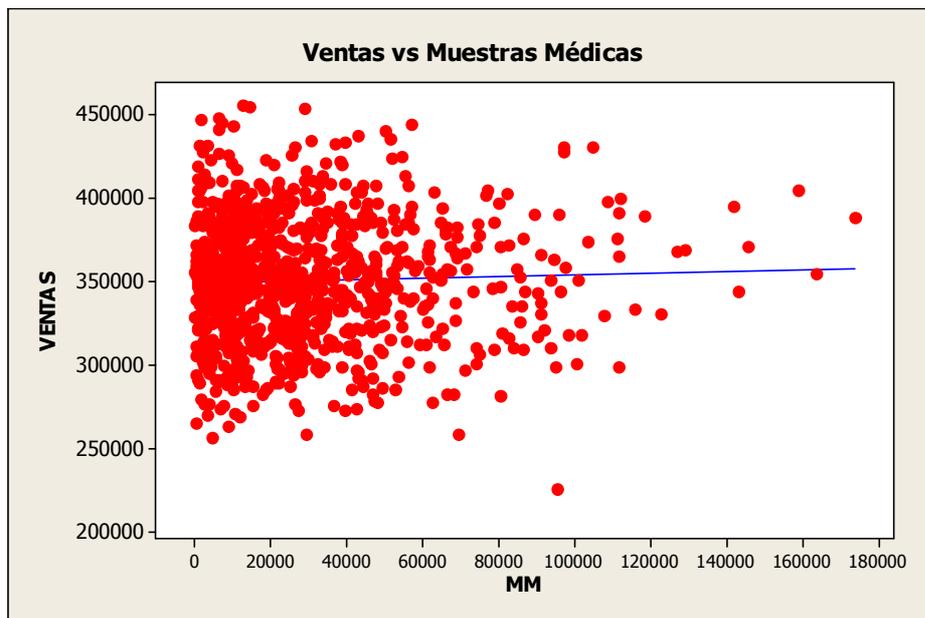


Figura 99. Gráfica de correlación entre la variable MM y Ventas de Voltaren. Fuente: Minitab.

Se observa un comportamiento similar al de la variable A&P, que los datos guardan cierta similitud en niveles bajos de Muestras Médicas. Lo que a su vez generan mayores ventas para el producto. Mas la relación de regresión no es obvia entre estas dos variables, información que se corrobora con el modelo de regresión antes obtenido.

- **Gráfico resumen de Ventas de Voltaren:**

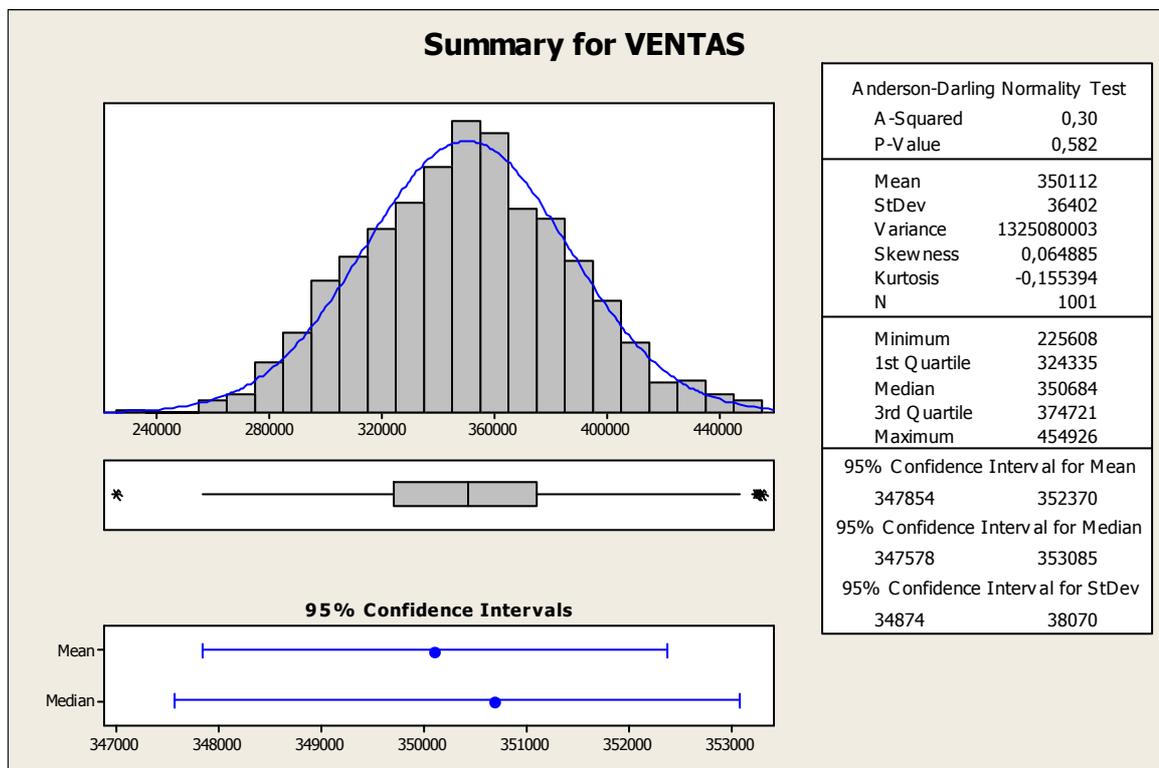


Figura 100. Gráfico resumen de Ventas de Voltaren. Fuente: Minitab.

Se ve que los datos se ajustan a una distribución Normal tal como se infirió inicialmente, además que su valor p es mayor al nivel de significancia establecido. Un dato importante de este cuadro resumen son la media, y valores mínimos y máximos determinados a partir de la simulación Montecarlo. Con el dato de las ventas máximas generadas se podrá determinar los niveles óptimos de las demás variables y poder compararlos con los obtenidos en el primer modelo de optimización.

De esta manera para el caso de Voltaren las ventas máximas que se generan son de \$454926, con los siguientes gastos en cada una de sus variables:

Variable	Gasto máximo en el que se debe incurrir
A&P	\$ 6569,06
Descuentos	\$ 192026
Muestras Médicas	\$ 12568,1
VENTAS GENERADAS	\$ 454926

Generación propia.

Tabla 4. Valores máximos de gastos por variable para el producto Voltaren

Comparando ambos modelos de optimización y de acuerdo a los resultados obtenidos de la regresión múltiple en el que la variable Descuentos resultó ser la de mayor significancia y Muestras médicas la de menos, el método de optimización que se ajusta mejor es el generado con simulación Montecarlo ya que del total del gasto en las tres variables, el 91% es asignado a la variable de mayor significancia en el modelo Descuentos, seguido de un 6% a la variable A&P y finalmente un 3% a la variable Muestras Médicas. (Los números generados y las variables encontradas se muestran en el ANEXO C, todos los datos han sido multiplicados por un factor para mantener la confidencialidad de Novartis Ecuador)

Todo este análisis, ambas optimizaciones, para los otros 4 productos y sus resultados se muestran en el ANEXO D-1, D-2, D-3, D-4.

CAPÍTULO V- CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

- Los análisis de regresión mostraron que dentro de toda la gama de productos seleccionados, la variable Descuentos es la única significativa para la función de rentabilidad por la venta de los productos. Mientras que las demás mostraban ser no significativas.
- El único producto que presentó, además de Descuentos, a la variable Advertisement and Promoting como significativa, fue Cataflam, que a pesar de pertenecer al mismo grupo que Diovan en cuanto a la Línea de Mercado masivo, éste último no cuenta con la variable Advertisement and Promoting como importa. Esto muestra que dos productos dentro de una misma línea no tienen el mismo comportamiento en el mercado.
- Independientemente de la línea a la que pertenece el producto, o de su edad o de su política de mercado, la variable Descuentos es la única significativa en todos ellos (como se ve en el Anexo B). Esto señala que el tipo de Descuentos que Novartis Ecuador realiza a sus clientes directos, sí se ve direccionado hacia el cliente final y por ello su aporte en la función Ventas.
- Al tener a esta variable como significativa, Novartis debe prestar especial énfasis en determinar los niveles óptimos de ésta con el fin de generar mayores ventas de sus productos y por lo tanto mayor rentabilidad para la empresa.
- Los dos productos pertenecientes a la línea de Especialidades y cuentas clave, Tegretol y Genteal, que cuentan con la variable Fuerza de Venta, no

señalan mayor indicativo de que esta variable aporte positivamente en sus modelos de regresión. Esto puede mostrar el indicio de que posiblemente sean productos que ya no lo necesiten.

- La inferencia en los datos de Muestras Médicas y Fuerza de Ventas resultaron difíciles de analizar, puesto que las distribuciones ajustadas lo hicieron con un valor p ligeramente más alto que el nivel de significancia tomado, igual a 0,05.
- Todas las regresiones mostraron un R^2 -adj elevado, es decir mayor al 50%. Esto significa que la toma de datos fue adecuada y que los modelos obtenidos contaban con un grado de ajuste conveniente para poder concluir sobre cada uno de ellos.
- Para facilidad de cálculo y análisis la mayoría de las distribuciones fueron ajustadas por medio de Minitab a distribuciones Normales por medio de la transformación de Johnson. Esto permitió generar números aleatorios de manera más rápida y confiable.
- Los modelos de optimización para cada producto mostraron los niveles óptimos de cada variable con el fin de maximizar la función objetivo de sus ventas. En algunos casos hubo variables que presentaron valores de cero, que complementariamente a los resultados de la regresión muestran ser no significativas para el producto.
- Cuando se resolvió la optimización por medio del complemento Solver de Excel se obtuvieron ventas mayores que cuando el modelo fue resuelto por medio de Simulación Montecarlo. Sin embargo con este último método se asignó valores a todas las variables regresoras.

- La simulación Montecarlo fue de gran utilidad para generar variables aleatorias continuas que se ajusten a la distribución de probabilidad específica de cada producto. Con la generación de muchos más datos, sin duda el sistema puede entenderse mejor y encontrar niveles adecuados para cada variable en función de las Ventas deseadas de cada producto.
- De entre los dos métodos de optimización usados, se pudo obtener respuestas más detalladas de la optimización con Simulación Montecarlo, ya que con la generación de 5000 datos aleatorios, se puede entender mejor el comportamiento de cada variable y a su vez, poder encontrar valores en un intervalo de datos bastante extenso.
- La Simulación Montecarlo puede resultar en grandes beneficios para la empresa, pues es un método de pronóstico del nivel de Ventas y gracias a las regresiones obtenidas, de los niveles de sus variables independientes. Novartis podrá hacer uso de esta herramienta las veces que desee, y para cualquier producto de su extensa gama con el fin de generar ahorros significativos en las cuatro variables descritas: Fuerza e Venta, Medios de Difusión, Descuentos y Muestras Médicas.
- Si se comparan con datos históricos anuales de las ventas de Voltaren se nota que con reduciendo el gasto A&P en casi un 50%, se logra las mismas ventas para el producto. Si bien cuando hay un incremento de A&P, existe también uno en Ventas, dicho aumento no es justificado. Es decir con un monto reducido en A&P, se logra tener Ventas altas de Voltaren sin la necesidad de invertir tanto en esta variable.
- Para el caso de Genteal lo que más resalta la atención es el resultado obtenido por Solver para el óptimo de Fuerza de Ventas, que dio un valor

de cero. Esto puede indicar que Genteal ya no es un producto que necesita de este soporte y o que su inversión no está generando tan buenos resultados como se espera. Con un gasto nulo en Fuerza de Ventas para Genteal se generan Ventas de \$ 1078172,38, mientras que con un gasto de FF de por ejemplo \$73913,91 se generaron Ventas de \$ 756664,09. Es decir \$321508,29 menos a pesar de que se invirtió una gran cantidad de dinero en FF. Esto es un hecho muy importante para Genteal, que un colirio medicado para el ojo seco, que quizá en el pasado necesitaba de Fuerza de Ventas (en calidad de visitantes médicos) mas ahora ya no en un nivel como antes.

5.2 Recomendaciones

- Considerar a la variable Descuentos como un tema de análisis especial y como el medio principal para optimizar las ventas de los productos.
- Reevaluar la cantidad de dinero invertido en Fuerza de Ventas para cada producto, pues si bien no es necesario eliminarlo, la cantidad que se utiliza para esta variable no está correspondiendo positivamente a las ventas de Novartis.
- Estudiar la posibilidad de cambiar productos que actualmente utilizan a la Fuerza de Venta como apoyo, a productos que se manejan sin ella, es decir productos de tipo OTC. Tal puede ser el caso para el producto Diovan.
- Diagnosticar el efecto que produce la variable Advertisement and Promoting especialmente en la rama de productos maduros como lo es

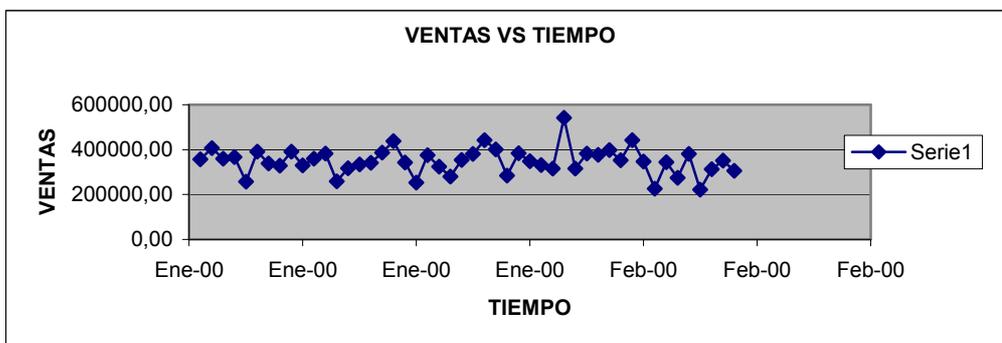
Cataflam y Voltaren. Mientras que para el primero resultó significativa, para el segundo no.

- Ampliar este análisis a más productos de la marca Novartis con el fin de conocer la relación que existe entre las variables estudiadas y la función a maximizar Ventas, con el fin de generar ahorros significativos para toda la empresa al igual que retener clientes existentes y crear otros nuevos.

ANEXOS

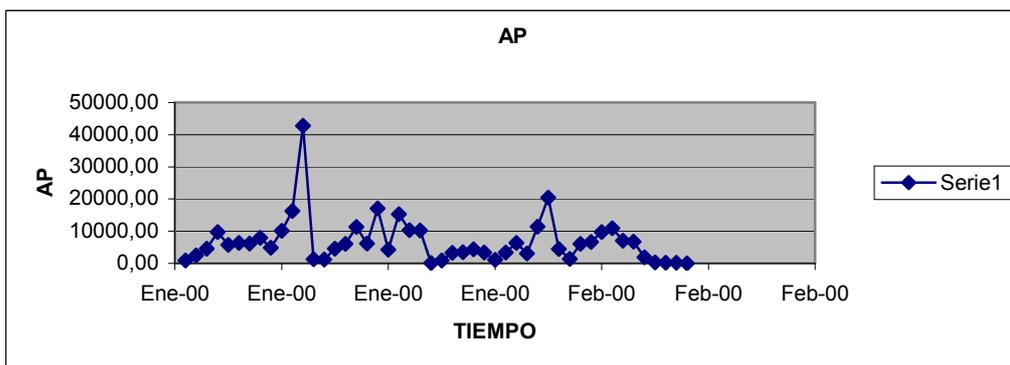
ANEXO A: Análisis preliminar de los datos muestrales por variable y por producto.

* Voltaren:



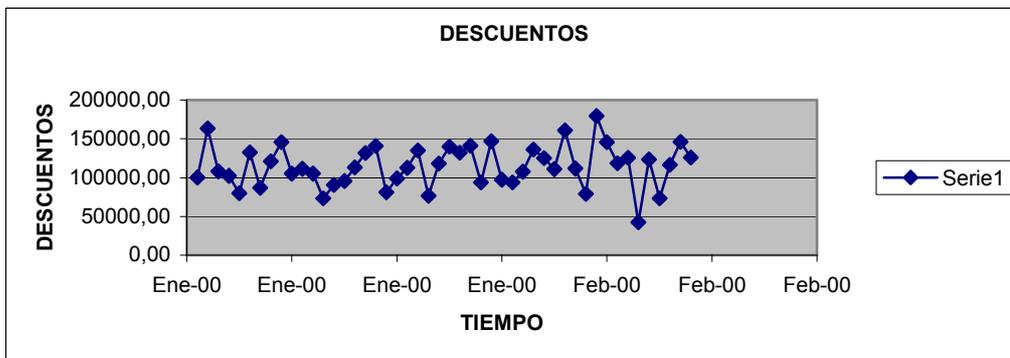
Generación propia

Figura 1. Gráfica de relación entre la variable Ventas vs. Tiempo del producto Voltaren



Generación propia

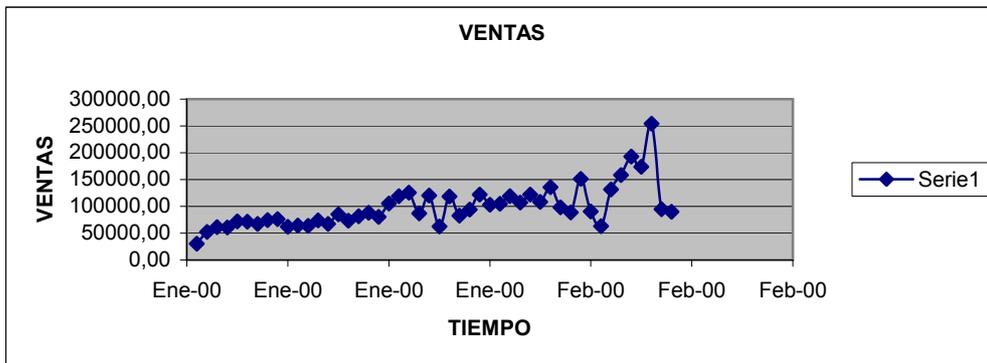
Figura 2. Gráfica de relación entre la variable Advertisement and Promoting vs. Tiempo del producto Voltaren



Generación propia

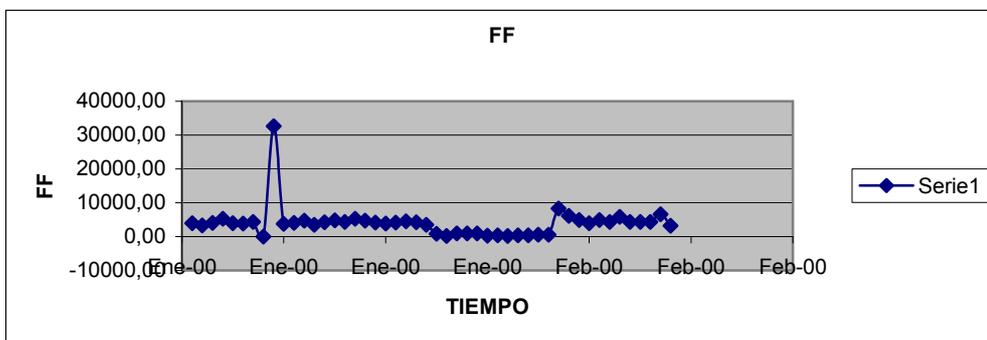
Figura 3. Gráfica de relación entre la variable Descuentos vs. Tiempo del producto Voltaren

* **Genteal:**



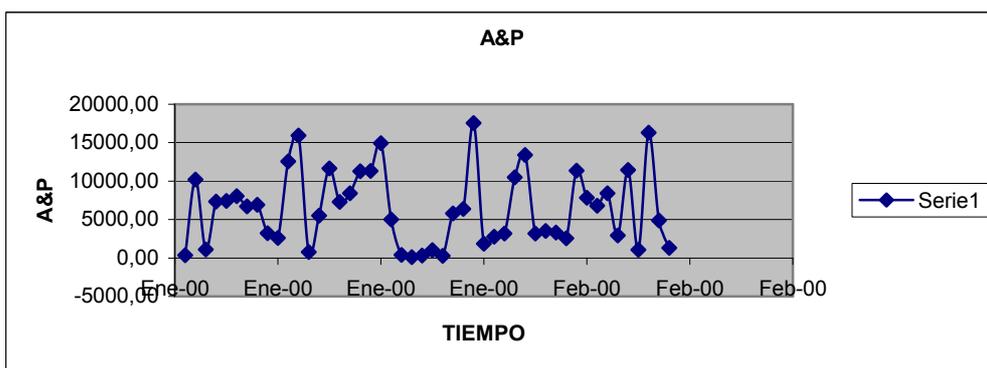
Generación propia

Figura 5. Gráfica de relación entre la variable Ventas vs. Tiempo del producto Genteal.



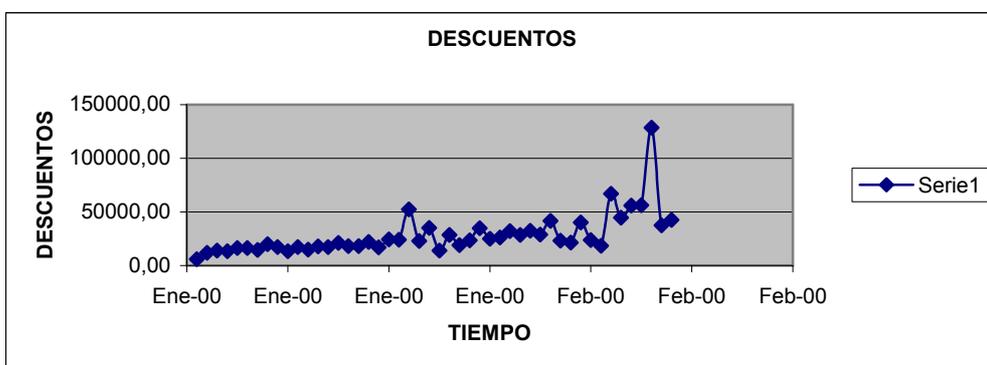
Generación propia

Figura 6. Gráfica de relación entre la variable Fuerza de Ventas vs. Tiempo del producto Genteal.



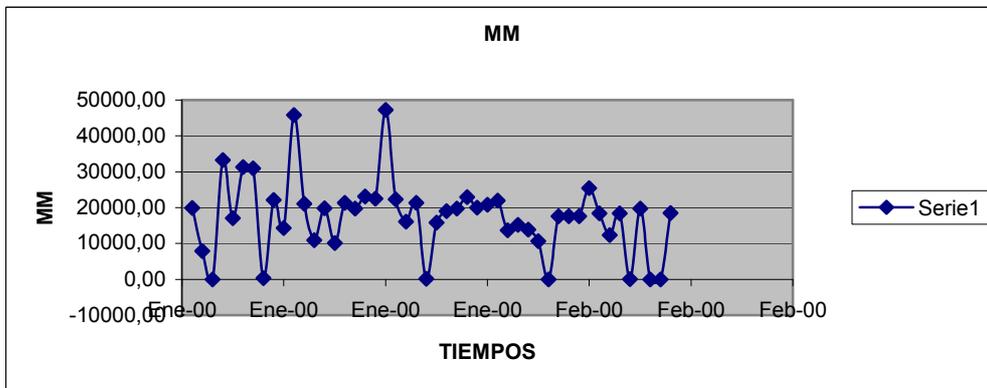
Generación propia

Figura 7. Gráfica de relación entre la variable A&P vs. Tiempo del producto Genteal.



Generación propia

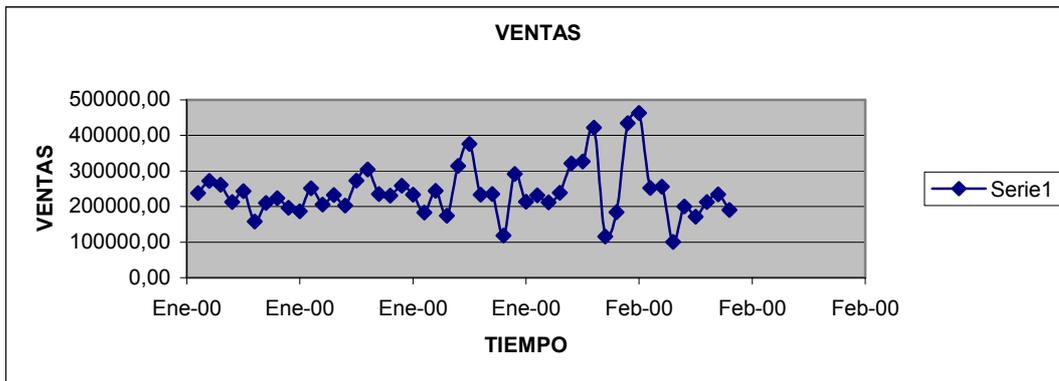
Figura 8. Gráfica de relación entre la variable Descuentos vs. Tiempo del producto Genteal.



Generación propia

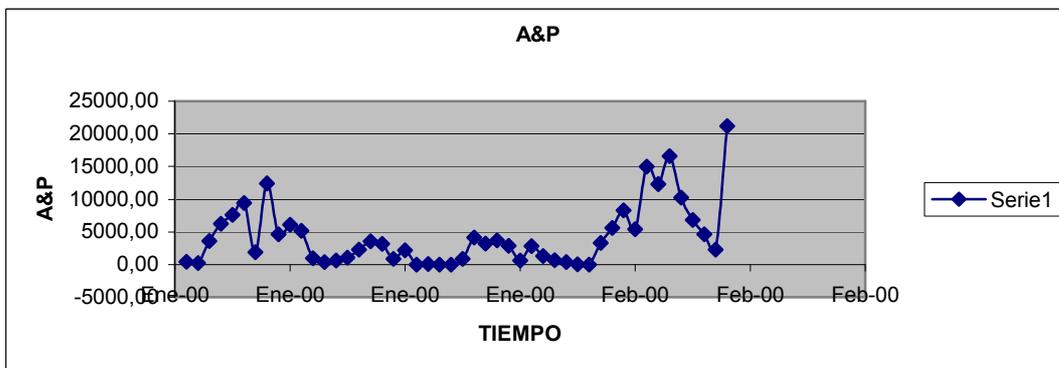
Figura 9. Gráfica de relación entre la variable Muestras Médicas vs. Tiempo del producto Genteal.

*** Cataflam**



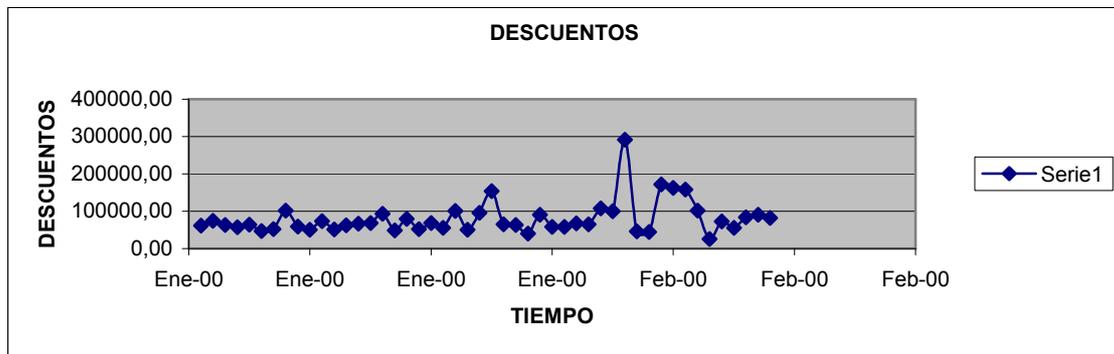
Generación propia

Figura 10. Gráfica de relación entre la variable Muestras Médicas vs. Tiempo del producto Cataflam.



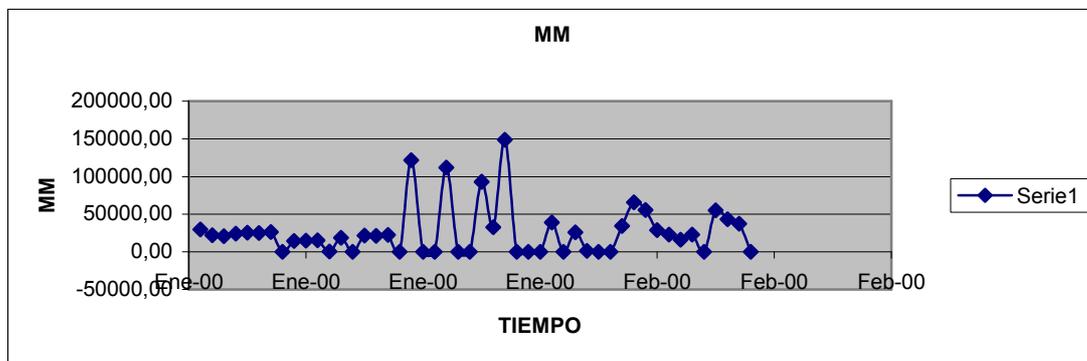
Generación propia

Figura 11. Gráfica de relación entre la variable A&P vs. Tiempo del producto Cataflam.



Generación propia

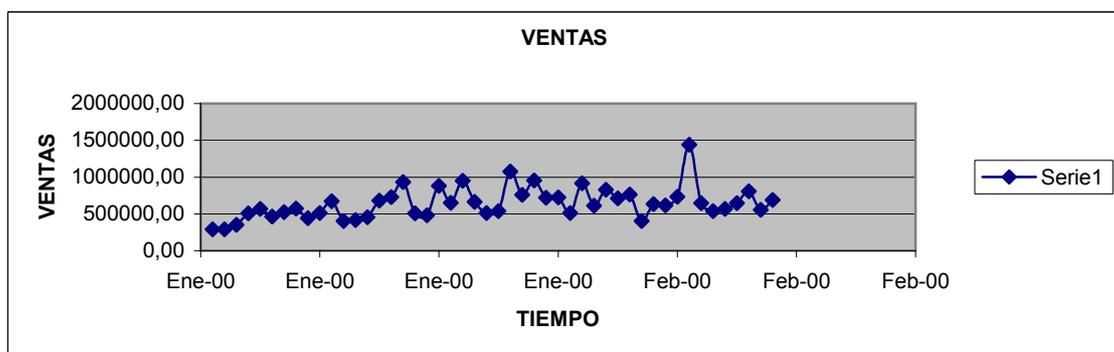
Figura 12. Gráfica de relación entre la variable Descuentos vs. Tiempo del producto Cataflam.



Generación propia

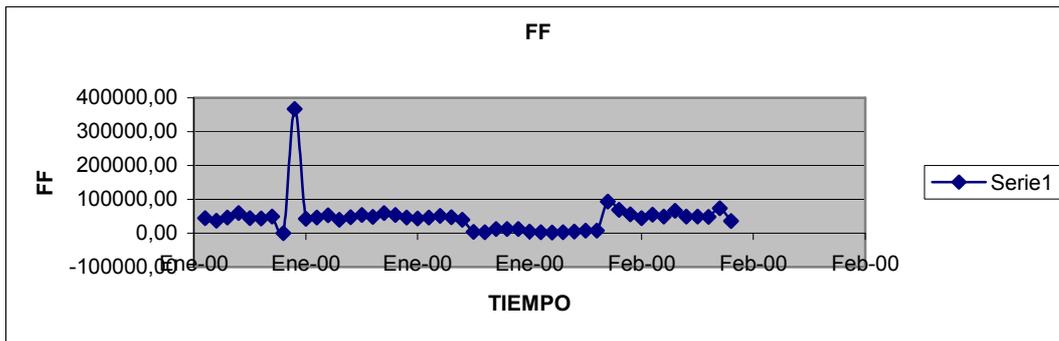
Figura 13. Gráfica de relación entre la variable Muestras Médicas vs. Tiempo del producto Cataflam.

*** Diovan**

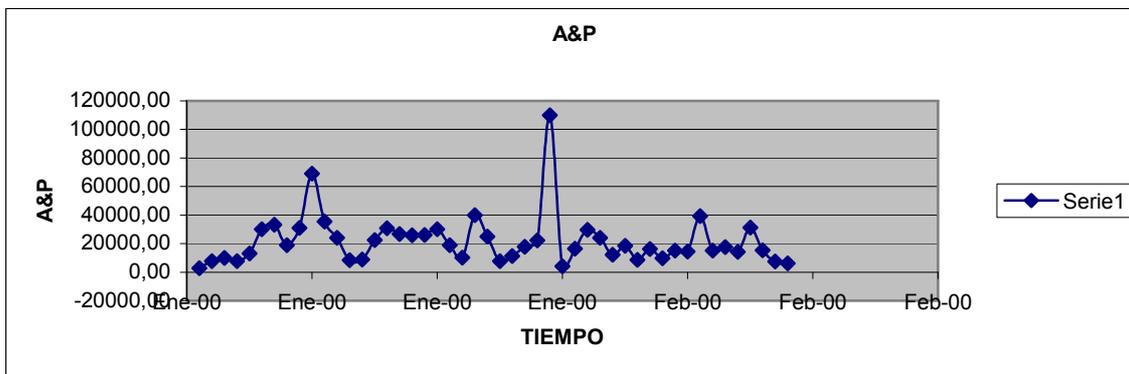


Generación propia

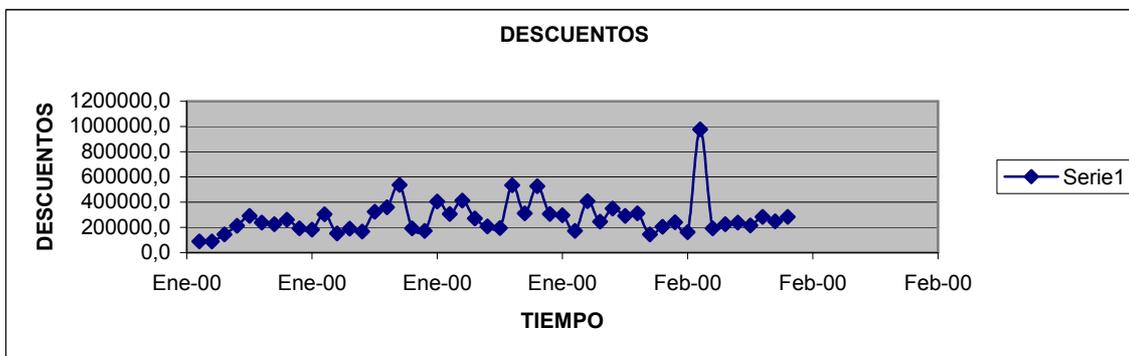
Figura 14. Gráfica de relación entre la variable Ventas vs. Tiempo del producto Diovan.



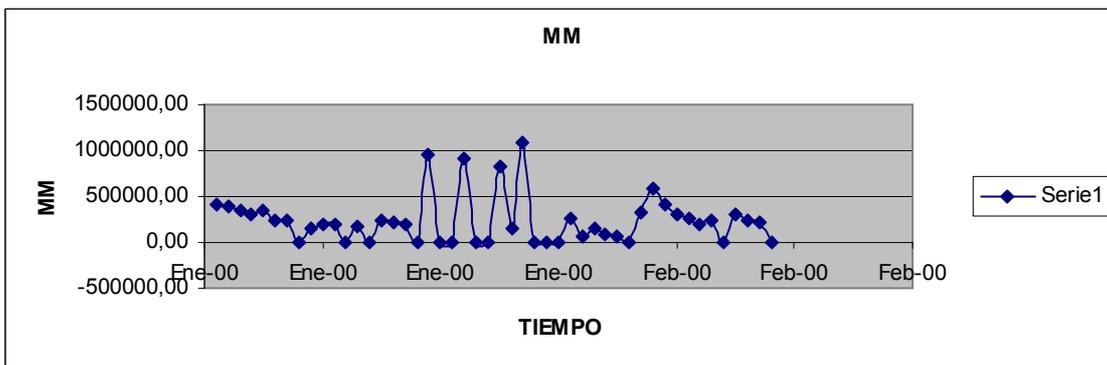
Generación propia
 Figura 15. Gráfica de relación entre la variable FF vs. Tiempo del producto Diovan.



Generación propia
 Figura 16. Gráfica de relación entre la variable A&P vs. Tiempo del producto Diovan.



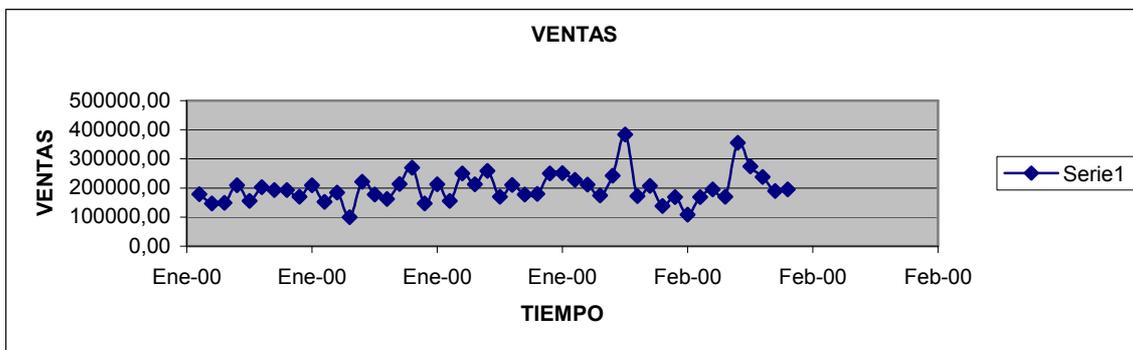
Generación propia
 Figura 17. Gráfica de relación entre la variable Descuentos vs. Tiempo del producto Diovan.



Generación propia

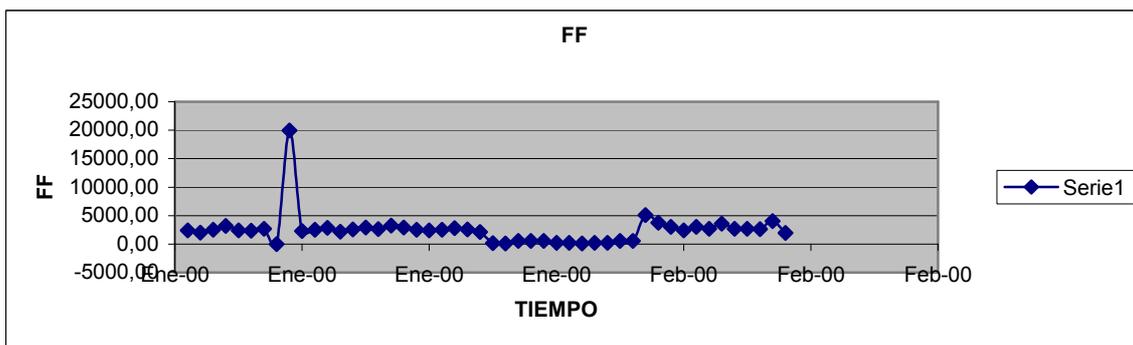
Figura 18. Gráfica de relación entre la variable Muestras Médicas vs. Tiempo del producto Diovan.

*** Tegretol**



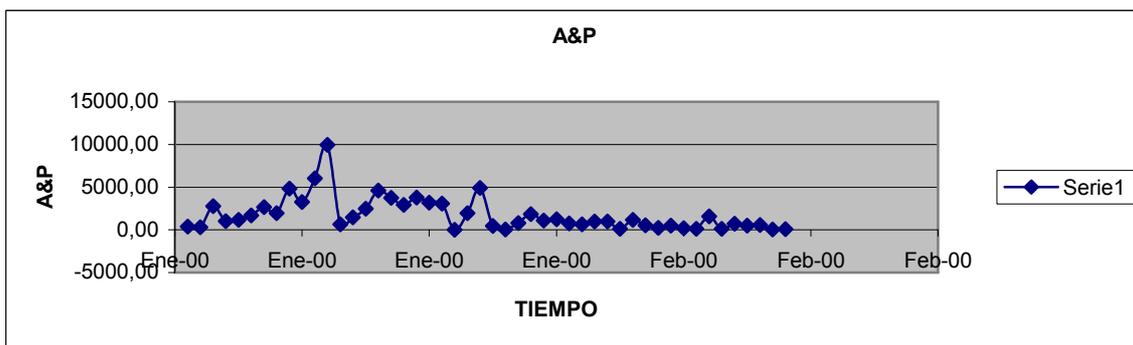
Generación propia

Figura 19. Gráfica de relación entre la variable Ventas vs Tiempo del producto Tegretol



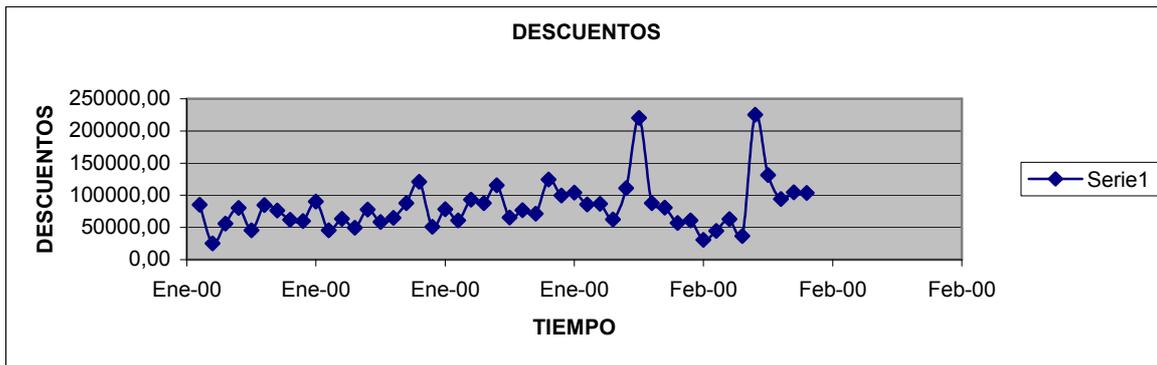
Generación propia

Figura 20. Gráfica de relación entre la variable FF vs. Tiempo del producto Tegretol.

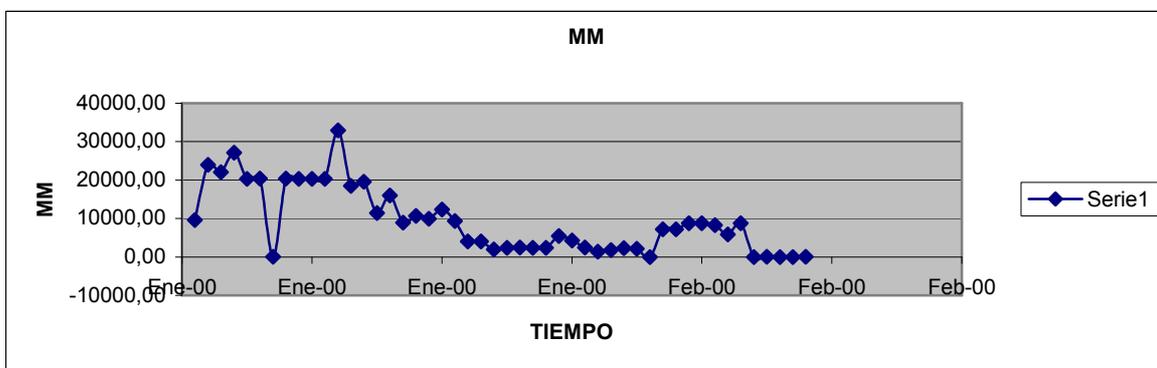


Generación propia

Figura 21. Gráfica de relación entre la variable A&P vs. Tiempo del producto Tegretol.



Generación propia
 Figura 22. Gráfica de relación entre la variable Descuentos vs. Tiempo del producto Tegretol.



Generación propia
 Figura 23. Gráfica de relación entre la variable Muestras Médicas vs. Tiempo del producto Tegretol.

ANEXOS B: INFERENCIAS Y ANÁLISIS ESTADÍSTICOS PROBABILÍSTICOS PARA LOS PRODUCTOS RESTANTES

ANEXO B1: PRODUCTO GENTEAL

Genteal es un producto para el alivio temporal de las molestias por irritaciones leves del ojo, provocadas por la exposición el viento, sol y demás irritaciones. Es especialmente medicado para aliviar la sequedad del ojo. (Genteal). Es un producto que pertenece a la línea de Opta dentro de la familia de productos Novartis, de la línea de Especialidades.

A diferencia de Voltaren, Genteal es un producto que necesita prescripción médica y que por ello cuenta con la variable Fuerza de Ventas, quienes tienen por objetivo promover las ventas del producto a través de médicos para que ellos receten Genteal a un gran número de sus pacientes. Por tanto cuenta con cuatro variables regresoras: FF, A&P, Descuentos, y MM.

Inferencia para la variable de salida Ventas de Genteal

Al igual que para el anterior producto, para Genteal se estudiaron datos de sus ventas desde el año 2006 al 2009. Con un alfa de 0.05 se infirió en los datos de la variable de salida, Ventas para conocer el tipo de distribución a la que se ajustan. La respuesta dada por el software Minitab mostró que para Ventas, la distribución de probabilidad más adecuada es una distribución Normal, luego de una transformación de Johnson. La ecuación y gráfica obtenidas son los siguientes:

$$-1.57561 + 1.52618 * \operatorname{Arcsinh}\left(\frac{X - 51356.4}{32575.0}\right)$$

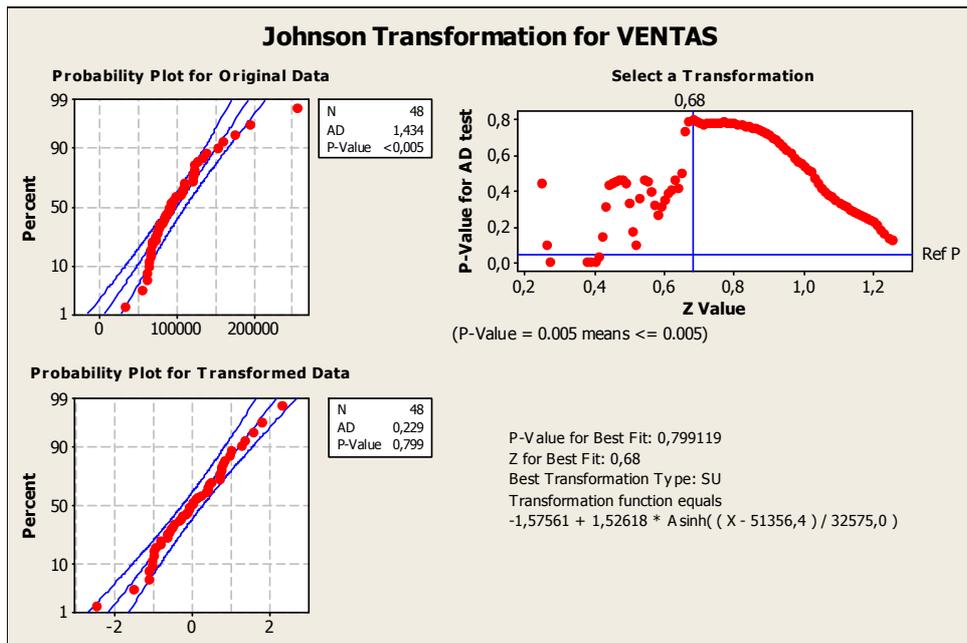


Figura 37. Transformación de Johnson de los datos generados de la variable Ventas de Voltaren.
Fuente: Minitab.

Un valor igual a 0.799, muestra que la hipótesis nula de que los datos se acercan a una distribución normal, no puede ser rechazada.

En base a los parámetros obtenidos y a la Simulación Montecarlo, se generaron 5000 datos con este tipo de distribución que servirán posteriormente para elaborar el modelo de regresión que se ajuste a este producto.

Inferencia en los datos generados de la variable Ventas para el producto Genteal

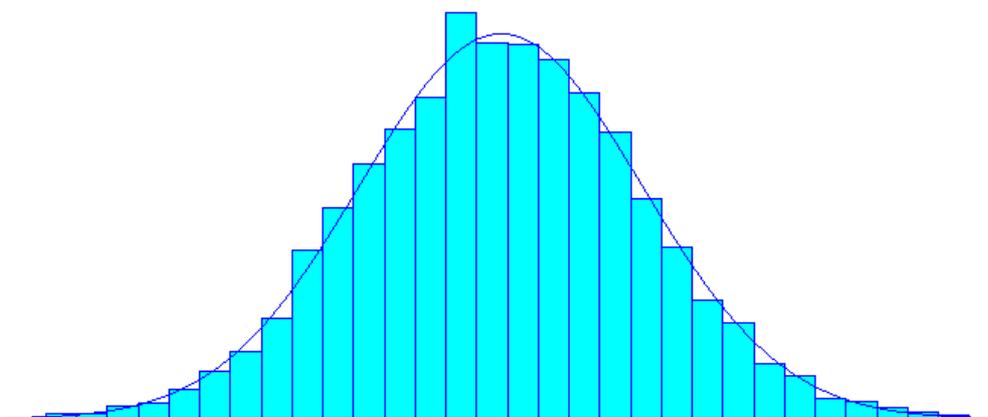


Figura 38. Distribución para los datos generados de la variable Ventas del producto Genteal.
Fuente: Arena.

Distribution Summary

Distribution: Normal
Expression: NORM(-0.0156, 0.91)
 Square Error: 0.000149

Chi Square Test

Number of intervals = 27
 Degrees of freedom = 24
 Test Statistic = 23.6
Corresponding p-value = 0.488

Kolmogorov-Smirnov Test

Test Statistic = 0.00937
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -3.22
 Max Data Value = 3.89
 Sample Mean = -0.0156
 Sample Std Dev = 0.91

Histogram Summary

Histogram Range = -3.94 to 4
 Number of Intervals = 40

Los estadísticos, valores p tanto para la prueba Kolgomorov y Chi Cuadrado son mayores a 0.05 por lo que se prueba que los datos generados a partir de una transformación de Johnson para la variable Ventas es la adecuada.

Inferencia en la variable de entrada Field Force del producto Genteal

Con los datos desde el año 2006 al 2009 correspondientes a la Fuerza de Ventas destinados al producto Genteal, se infirió en ellos para ver el tipo de distribución a la que se ajustan. Usando un alfa de 0.05 se obtuvo que la distribución más adecuada sea una Normal, de igual manera luego de una transformación Johnson. La ecuación óptima en este caso es:

$$-6,46302 + 0,990081 * \text{LN}(X - 230,996)$$

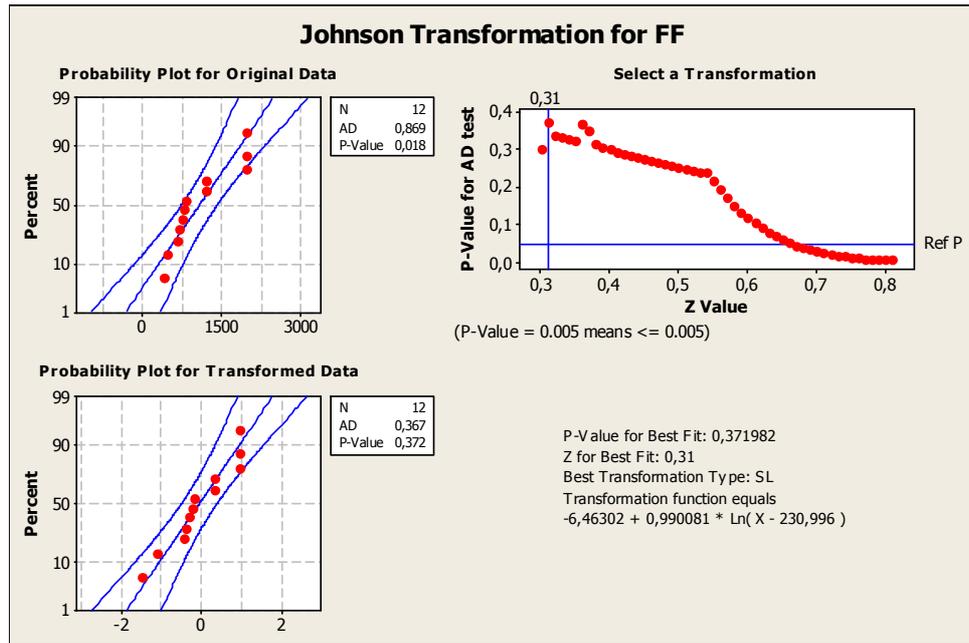


Figura 39. Transformación de Johnson de los datos generados de la variable Ventas de Genteal. Fuente: Minitab.

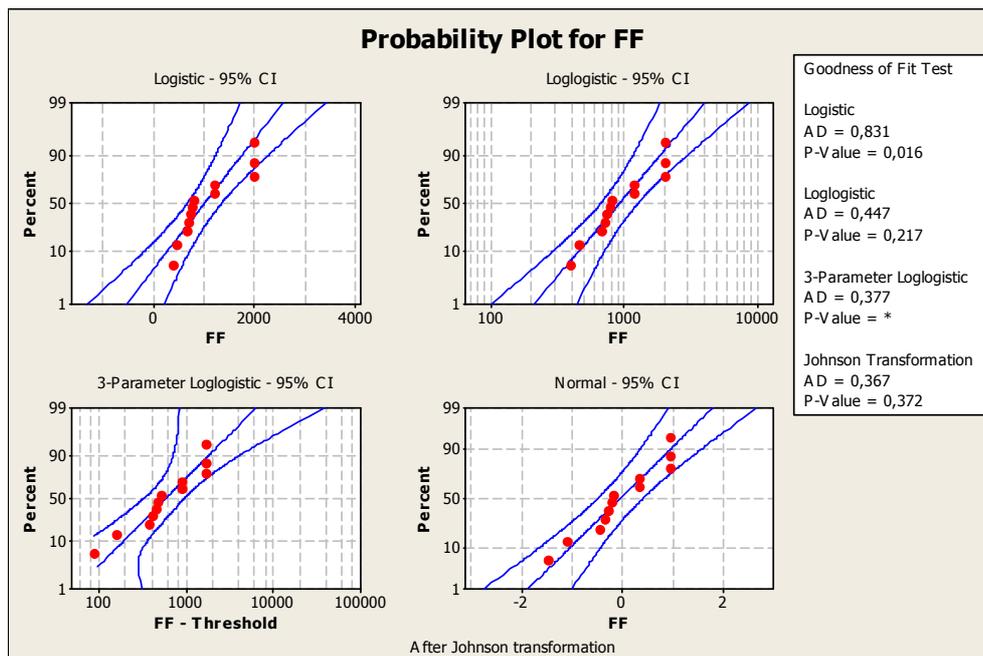


Figura 40. Distribución para los datos de FF del producto Genteal. Fuente: Minitab.

Con un valor p de la transformación de Johnson igual a 0.372, se acepta la hipótesis nula y los nuevos datos transformados siguen una distribución normal.

Con los parámetros correspondientes, y la simulación Montecarlo se generaron 5000 datos y se infirió comprobando que es una distribución normal.

Inferencia en los datos generados de la variable Field Force del producto Genteal

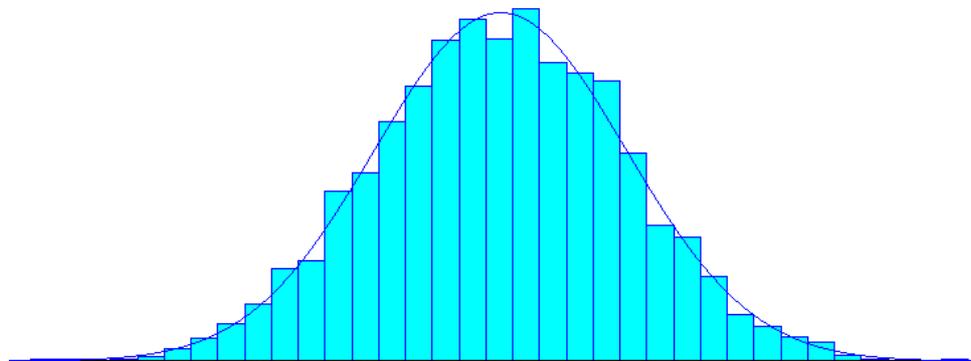


Figura 41. Distribución para los datos generados de FF del producto Genteal. Fuente: Arena.

Distribution Summary

Distribution: Normal
Expression: NORM(-0.052, 0.776)
 Square Error: 0.000238

Chi Square Test

Number of intervals = 27
 Degrees of freedom = 24
 Test Statistic = 31.2
Corresponding p-value = 0.16

Kolmogorov-Smirnov Test

Test Statistic = 0.0089
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -3.03
 Max Data Value = 2.76
 Sample Mean = -0.052
 Sample Std Dev = 0.776

Histogram Summary

Histogram Range = -3.61 to 3
 Number of Intervals = 40

En este caso, cualquier de las dos pruebas de bondad de ajuste sirve para comprobar la normalidad de los datos. Ambas cuentan con valores p mayores al nivel de significancia por lo que la hipótesis nula no puede ser rechazada.

Inferencia en la variable de entrada Advertisement & Promoting del producto Genteal

Con información que data desde el 2006 al 2009 se infirió en los datos de Advertisement & Promoting de Genteal. Igualmente con un alfa de 0.05 y probando el mismo juego de hipótesis antes planteado, se determinó que la distribución adecuada es nuevamente una distribución Normal con transformación de Johnson bajo la siguiente ecuación:

$$0,491666 + 0,655127 * \text{Ln}((X + 211,076) / (18427,9 - X))$$

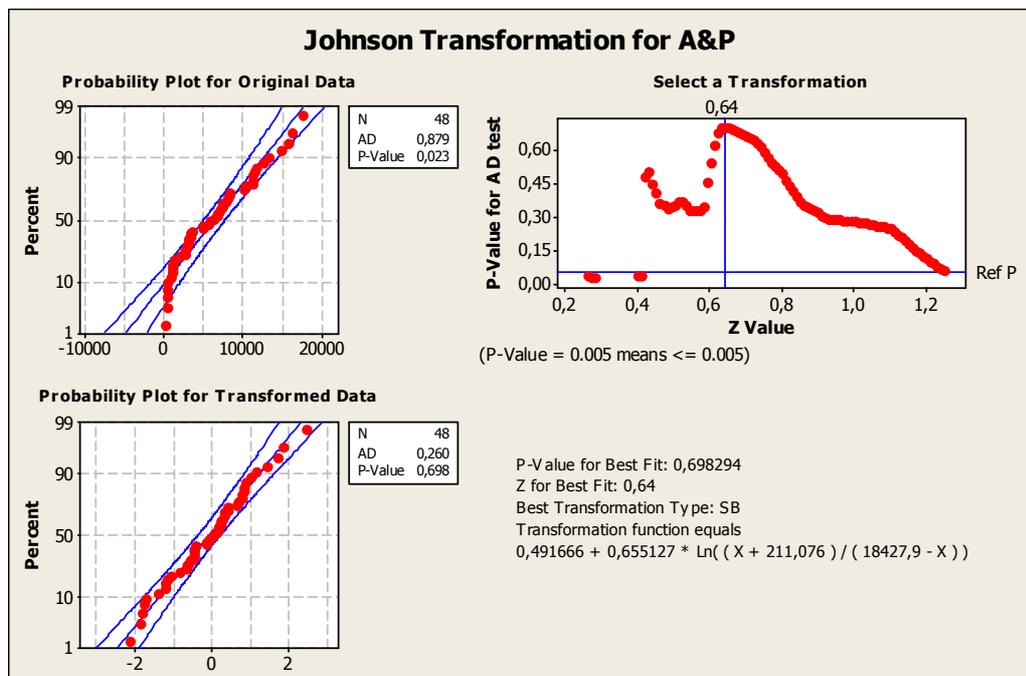


Figura 42. Transformación de Johnson de los datos de A&P de Genteal. Fuente: Minitab.

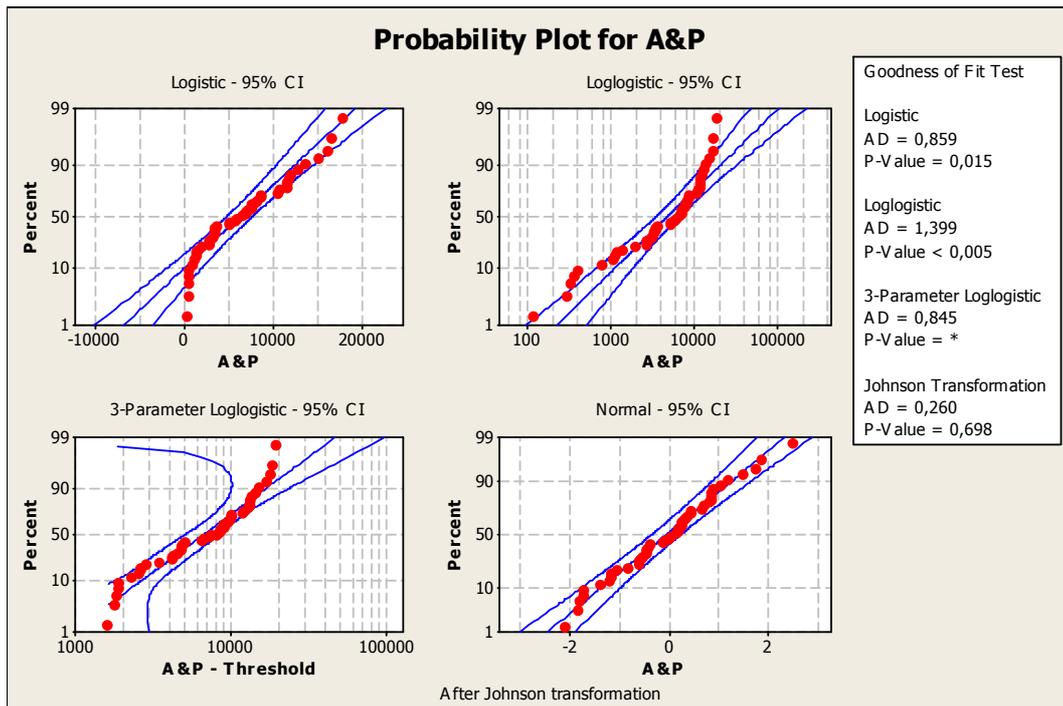


Figura 43. Distribución para los datos de A&P del producto Genteal. Fuente: Minitab.

El valor p de la prueba de ajuste permite aceptar la hipótesis nula y se concluye que los datos siguen una distribución Normal. Con los parámetros de este tipo de distribución, y en base a una simulación Montecarlo se generaron 5000 datos y se infirió nuevamente en ellos para comprobar que siguen la distribución adecuada. El resultado fue el siguiente.

Inferencia en los datos generados de la variable **Advertisement & Promoting** del producto Genteal

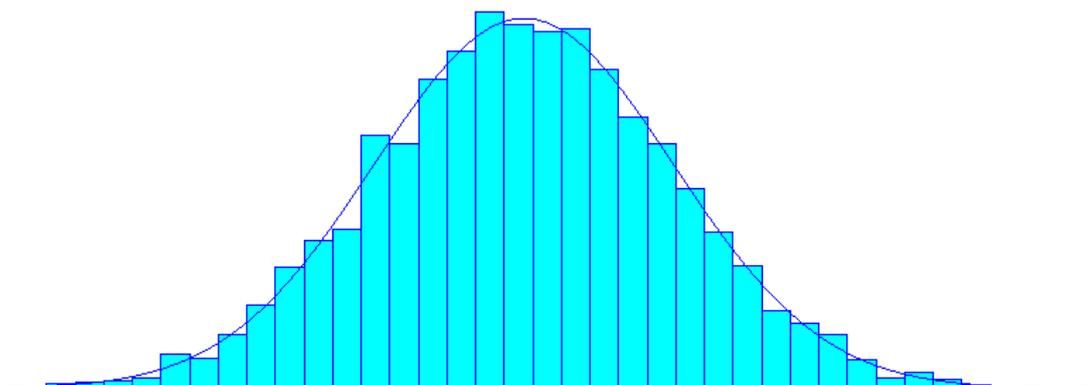


Figura 44. Distribución para los datos generados de A&P del producto Genteal. Fuente: Minitab

Distribution Summary

Distribution: Normal
Expression: NORM(-0.0661, 1.04)
 Square Error: 0.000141

Chi Square Test

Number of intervals = 30
 Degrees of freedom = 27
 Test Statistic = 31
Corresponding p-value = 0.275

Kolmogorov-Smirnov Test

Test Statistic = 0.00721
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -3.87
 Max Data Value = 3.7
 Sample Mean = -0.0661
 Sample Std Dev = 1.04

Histogram Summary

Histogram Range = -4 to 4
 Number of Intervals = 40

La prueba Chi cuadrado y Kolgomorov muestran valores p adecuados, mayores al 0.05 que aseguran que los datos transformados siguen una distribución Normal y están listos para ser usados en el modelo de regresión múltiple.

Inferencia en la variable de entrada Descuentos en el producto Genteal

Para la variable Descuentos del producto Genteal, se usó un alfa de 0.05, obteniendo que los datos, tomados desde el 2006 al 2009, se ajusten de mejor manera a una distribución Normal mediante una transformación Johnson. Su gráfica y resultados se ven de la siguiente manera:

$$-1,41563 + 1,01057 * \text{Asinh}((X - 12920,7) / 5281,94)$$

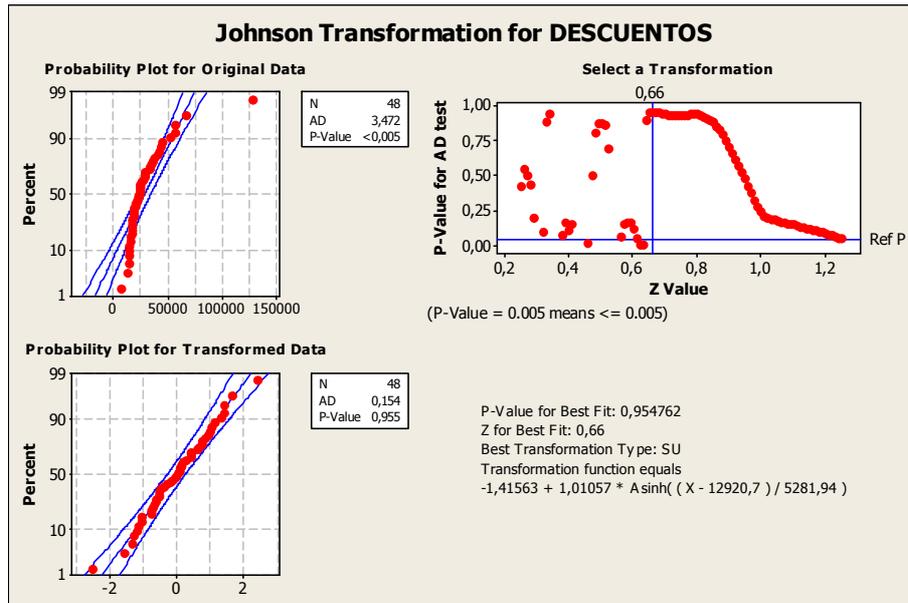


Figura 45. Transformación de Johnson de los datos de Descuentos de Genteal. Fuente: Minitab.

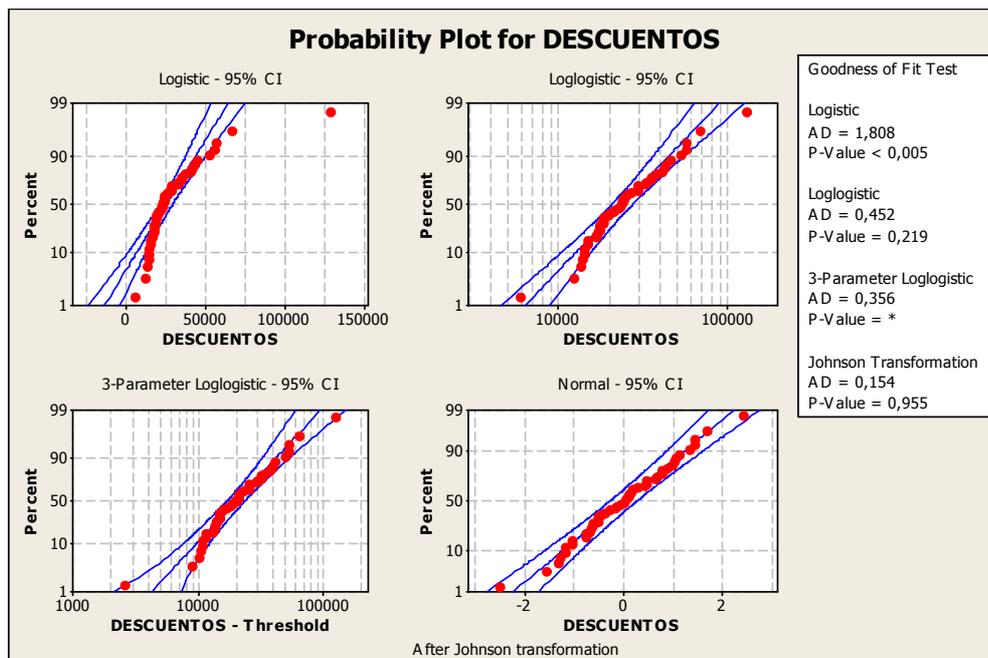


Figura 46. Distribución para los datos de Descuentos del producto Genteal. Fuente: Minitab

Mediante la ecuación de transformación adecuada, los datos pertenecientes a la variable Descuentos se pueden ajustar a una distribución Normal. Esto se comprueba en el valor p que brinda la aproximación de Johnson, el cual es muy bueno para mostrar la normalidad de los datos.

Con sus parámetros y la simulación Montecarlo se generaron 5000 datos. Su normalidad se prueba en el siguiente apartado.

Inferencia en los datos generados de la variable Descuentos del producto Genteal

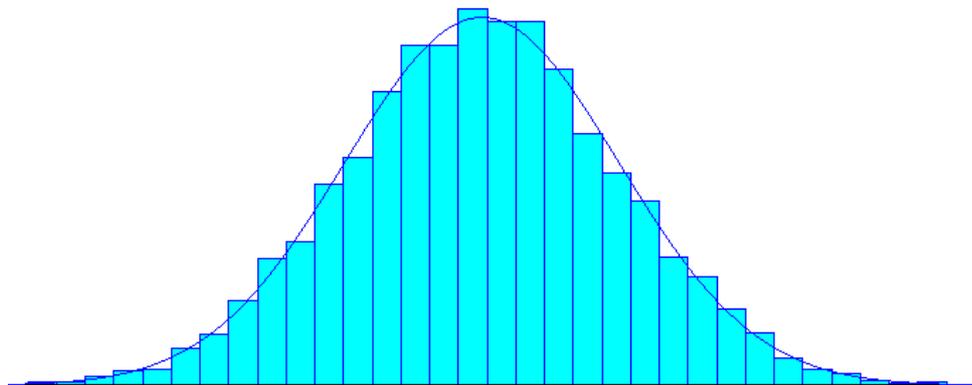


Figura 47. Distribución para los datos generados de Descuentos del producto Genteal. Fuente: Arena.

Distribution Summary

Distribution: Normal
Expression: NORM(-0.0394, 0.954)
 Square Error: 0.000123

Chi Square Test

Number of intervals = 28
 Degrees of freedom = 25
 Test Statistic = 16.6
Corresponding p-value > 0.75

Kolmogorov-Smirnov Test

Test Statistic = 0.01
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -3.35
 Max Data Value = 3.54
 Sample Mean = -0.0394
 Sample Std Dev = 0.954

Histogram Summary

Histogram Range = -4 to 4
 Number of Intervals = 40

El valor p mayor a 0.75 y 0.15 para las pruebas Chi Cuadrado y Kolgomorov respectivamente obtenidos, ambos mayores al nivel de significancia,

comprueba que la distribución de los datos es normal con parámetros iguales o muy parecidos a la distribución original.

Inferencia en la variable de entrada Muestras Médicas del producto Genteal

Finalmente se infiere en la variable Muestras Médicas del producto Genteal. Curiosamente y con un alfa igual a 0.05, la misma cantidad de datos, para esta variable la distribución más adecuada es igualmente una Normal, posterior de una transformación de Johnson. Su ecuación, gráficos y resultados se presentan a continuación:

$$0,138751 + 0,633417 * \text{Asinh}((X - 18989,0) / 3480,45)$$

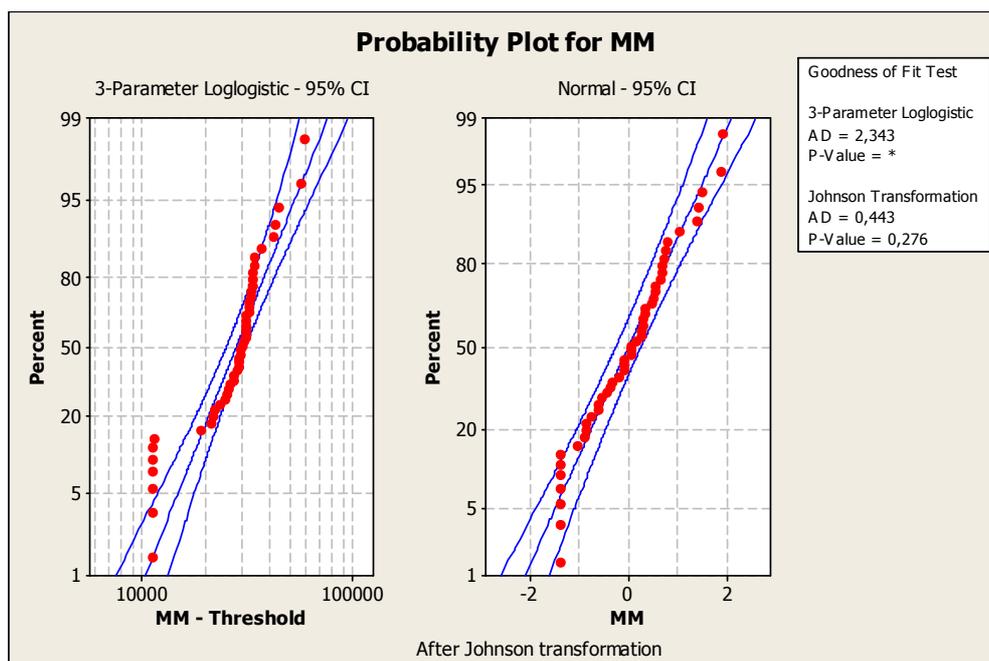


Figura 48. Distribución para los datos de MM del producto Genteal. Fuente: Minitab.

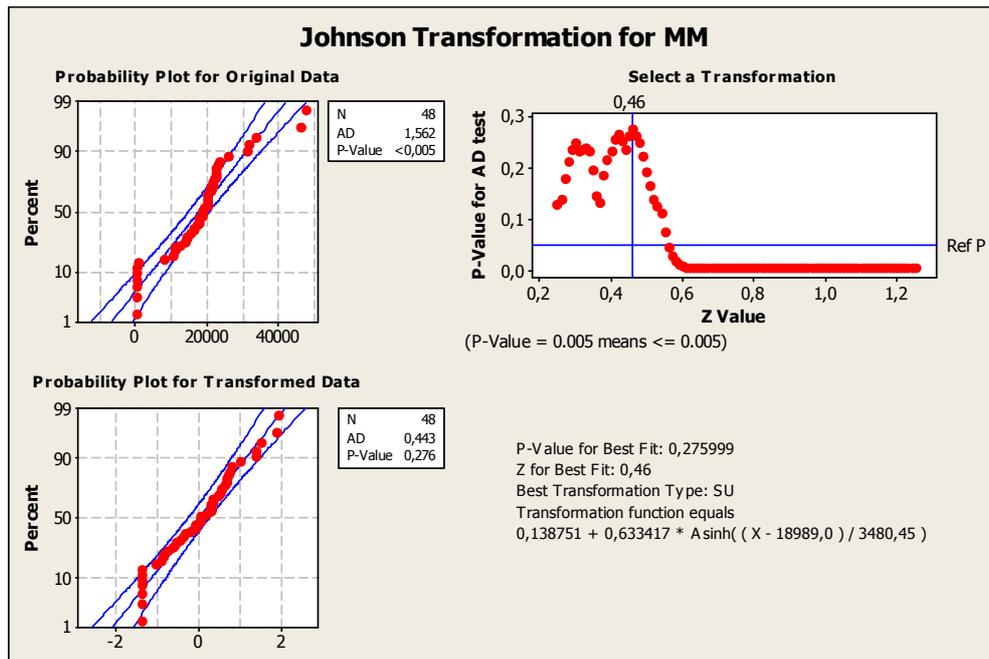


Figura 49. Transformación de Johnson de los datos de MM de Genteal. Fuente: Minitab.

Al igual que para las demás variables, el estadístico p muestra que la distribución Normal luego de una transformación de Johnson es la adecuada. Generando 5000 datos con simulación Montecarlo y los parámetros adecuados de la distribución normal, se infirió en ellos. Sus resultados a continuación:

Inferencia en los datos generados de la variable Muestras Médicas

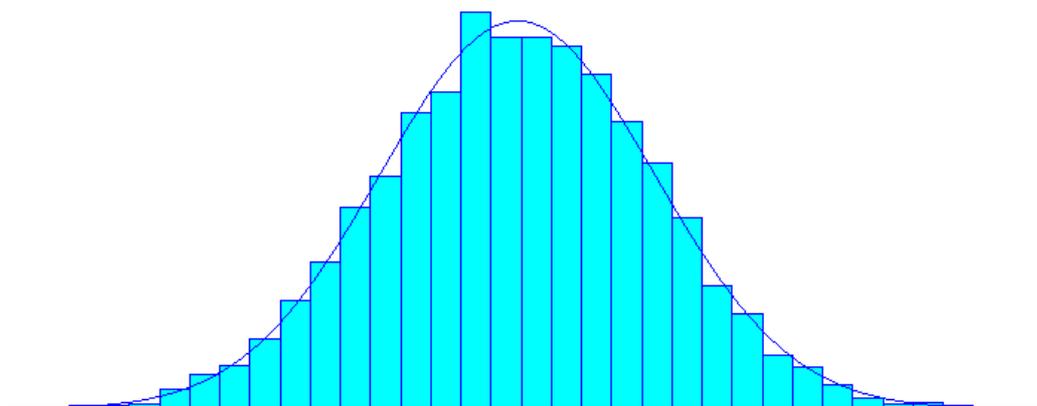


Figura 50. Distribución para los datos generados de MM del producto Genteal. Fuente: Arena

Distribution Summary
 Distribution: Normal
Expression: NORM(-0.0282, 0.909)
 Square Error: 0.000126

Chi Square Test
 Number of intervals = 27
 Degrees of freedom = 24
 Test Statistic = 16.2
Corresponding p-value > 0.75

Kolmogorov-Smirnov Test
 Test Statistic = 0.00925
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -3.32
 Max Data Value = 3.68
 Sample Mean = -0.0282
 Sample Std Dev = 0.909

Histogram Summary

Histogram Range = -4 to 4
 Number of Intervals = 40

La prueba de ajuste adecuada es la prueba Chi cuadrado y Kolgomorov con valores p mayores a 0,05 indica que se acepta la hipótesis nula de que los datos se ajustan a una distribución normal.

Modelo de regresión con los datos generados para el producto Genteal

En base a las distribuciones inferidas de cada variable para el producto Genteal, que todas resultaron en ser transformadas para dar lugar a distribuciones Normales, y la generación Montecarlo de 5000 datos que se ajusten a dichas distribuciones, se procedió a correr un modelo de regresión múltiple. Tal como para el anterior producto se pretende saber cuál es el impacto de las variables de entrada, sobre la variable de salida, Ventas de Genteal.

Las hipótesis a probar son las siguientes:

$$H_0 : \sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$$

$$H_0 : \sigma_1 \neq \sigma_2 \neq \sigma_3 \neq \sigma_4$$

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$$

$$H_1 : \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4$$

Con un valor p menor al valor alfa establecido, se rechaza la hipótesis nula que establece,

H_0 = El factor no influye de manera significativa en el modelo.

H_1 = El factor sí influye de manera significativa en el modelo.

A continuación se muestran los resultados obtenidos del modelo de regresión múltiple para el producto Genteal con sus cuatro variables regresoras: Field Force, Advertisement & Promoting, Descuentos y Muestras Médicas.

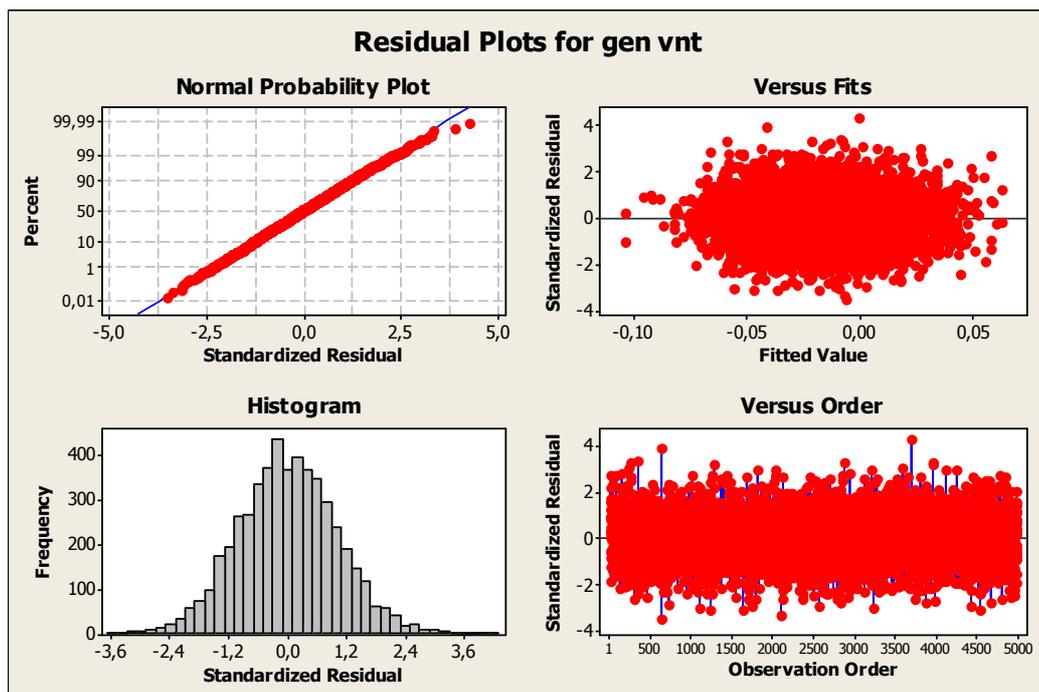


Figura 51. Residuales del modelo de regresión múltiple para el producto Genteal con los datos transformados. Fuente: Minitab.

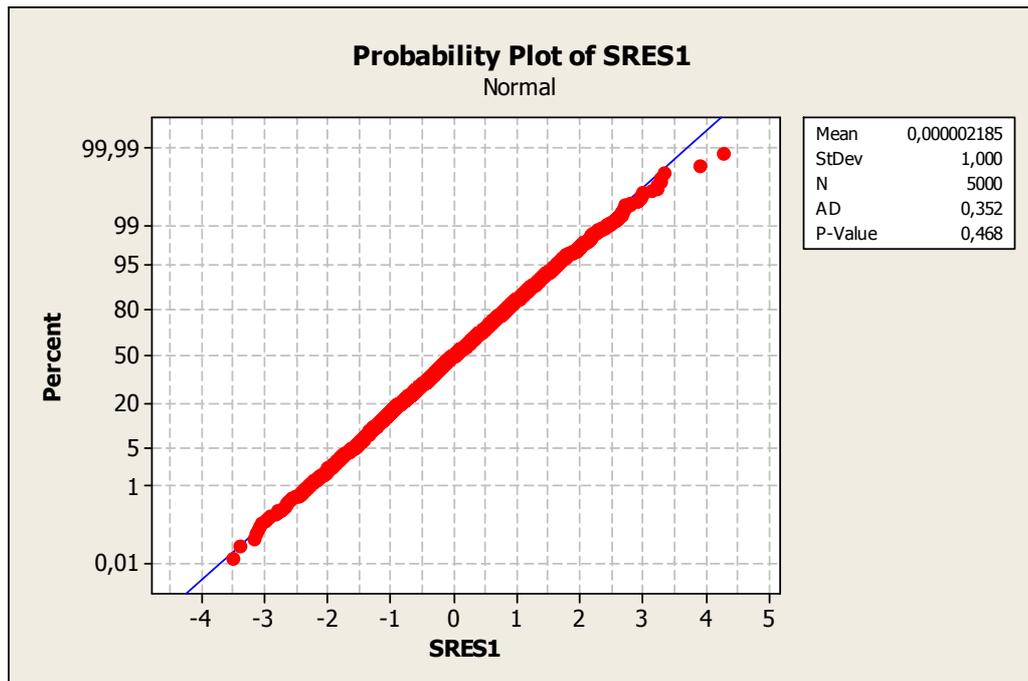


Figura 52. Prueba de normalidad de los residuales del modelo de regresión múltiple para el producto Genteal. Fuente: Minitab.

Regression Analysis: gen vnt versus ff. gen ap. gen des. gen mm

The regression equation is

$$\text{gen vnt} = 54033 - 6,28 \text{ gen ff} + 0,131 \text{ gen ap} + 1,81 \text{ gen des} - 0,030 \text{ gen mm}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	54033	9512	5,68	0,000	
gen ff	-6,283	4,304	-1,46	0,152	1,019
gen ap	0,1311	0,5547	0,24	0,814	1,143
gen des	1,8130	0,1449	12,52	0,000	1,289
gen mm	-0,0296	0,2776	-0,11	0,916	1,306

S = 17251,2 R-Sq = 82,8% R-Sq(adj) = 81,2%

PRESS = 19764991062 R-Sq(pred) = 73,47%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	61713458758	15428364689	51,84	0,000
Residual Error	4995	12797017342	297605054		
Total	4999	74510476099			

Source	DF	Seq SS
gen ff	1	0,1976
gen ap	1	1,0717
gen des	1	0,1069
gen mm	1	1,4091

Primero se analizará el gráfico de residuales obtenido:

Analizando en primera instancia a la gráfica de residuales como era de esperarse, éstos siguen una distribución normal. Se ve claramente que los residuos se ajustan perfectamente a la recta y de igual manera con la forma de campana que siguen los datos en el histograma. Se concluye que éstos siguen una distribución normal y de esta manera el supuesto 11 se satisface. Las gráficas restantes, Residuos vs. Ajustes y Orden, respectivamente, no presentan patrón alguno que llame la atención en los residuos resultantes, además que no se acercan a la línea de referencia lo cual indica parcialmente que sí tienen una varianza constante. Con esto el supuesto 4, de la varianza del residuo constante sea verdadero.

Para la aceptación del supuesto 3, al existir un valor para β_1 queda satisfecho. Para analizar el supuesto de no multicolinealidad entre las variables regresoras, se analiza el FIV. Los resultados obtenidos muestran que el VIF de cada variable regresora es igual o con una mínima desviación a 1, por lo que se concluye que no existe multicolinealidad en los datos y el supuesto 5 y por lo tanto el 6 quedan satisfechos.

Por su lado, el supuesto 1, que dice que el modelo de regresión es lineal en los parámetros, se ve que todos ellos mantienen coeficientes lineales y el supuesto se cumple. Haciendo la regresión sobre 5000 datos sobre 3 variables de interés, el supuesto 7 queda satisfecho. Los demás supuestos 8 y 9, pueden presumirse como correctos con los resultados expuestos por Minitab.

Una vez que los supuestos del Método de Mínimos cuadrados se cumplen, se debe analizar los resultados obtenidos de la regresión.

Enfocándose en el análisis de varianza, que indica si existe o no una variable de importante influencia en el modelo, el valor p obtenido, igual a cero, muestra la existencia de al menos una variable regresora no trivial en el modelo.

Si se observa la ecuación de la regresión obtenida, ésta muestra un coeficiente más grande para la variable Descuentos, y el más pequeño para la variable Muestras Médicas. Analizando la tabla de valores obtenidos para cada coeficiente, se comprueba que para la variable Descuentos el valor p resultante es cero, que al ser menor que 0.05 (nivel de significancia) permite rechazar la hipótesis nula y que esta variable es la única representativa para la ecuación de regresión de Voltaren.

R^2 -adj muestra el grado de ajuste del modelo a los datos, que en este caso corresponde a un 81.2%. Porcentaje bastante bueno para concluir que el modelo se ajusta de manera correcta a los datos generados de Genteal.

Al igual que Voltaren la variable Descuentos es la de mayor importancia en el producto Genteal, será importante analizar la relación que existe entre esta variable y su influencia en las distintas líneas a las que pertenece el producto.

En el modelo de optimización correspondiente al producto Genteal se podrán fijar valores para cada variable en juego, considerando a Descuentos como la más importante del modelo.

ANEXO B-2 PRODUCTO CATAFLAM

Cataflam es un producto para el tratamiento de las afecciones agudas como inflamación, afecciones dolorosas, reumatismo no articular, ataques de migraña, entre muchos más (Cataflam). Es una medicina que a pesar de que se la vende libremente y cualquier persona puede acceder a ella, no se la categoriza dentro de un producto OTC. Pertenece a la rama de productos maduros, al igual que Voltaren, pero a diferente área terapéutica.

Guardando similitud con Voltaren, Cataflam no cuenta con Fuerza de Ventas para su promoción, y es precisamente esta similitud entre ambos productos maduros sin Fuerza de Ventas, que se pretende analizar. La diferencia la marca el hecho de que mientras Voltaren se vende en percha (OTC), Cataflam no lo hace. Se requiere incurrir en la posible relación existente entre estos dos productos y su comportamiento individual en cuanto a las similitudes y la diferencia que existe al momento de su venta.

Inferencia para la variable de salida Ventas de Cataflam

Para Cataflam, se estudiaron datos desde el año 2006 al 2009. Con el mismo nivel de significancia que para los demás productos, es decir un alfa de 0.05 se infirió en los datos de la variable de salida, Ventas para conocer el tipo de distribución a la que se ajustan.

El resultado de esto, mediante Minitab, fue el siguiente. La variable Ventas se ajusta a una distribución Normal, antes realizada una transformación mediante el algoritmo de Johnson. La ecuación para dicha conversión es:

$$-10.387335 + 0.983885 * \operatorname{Arcsinh}\left(\frac{X - 212520}{36807.5}\right)$$

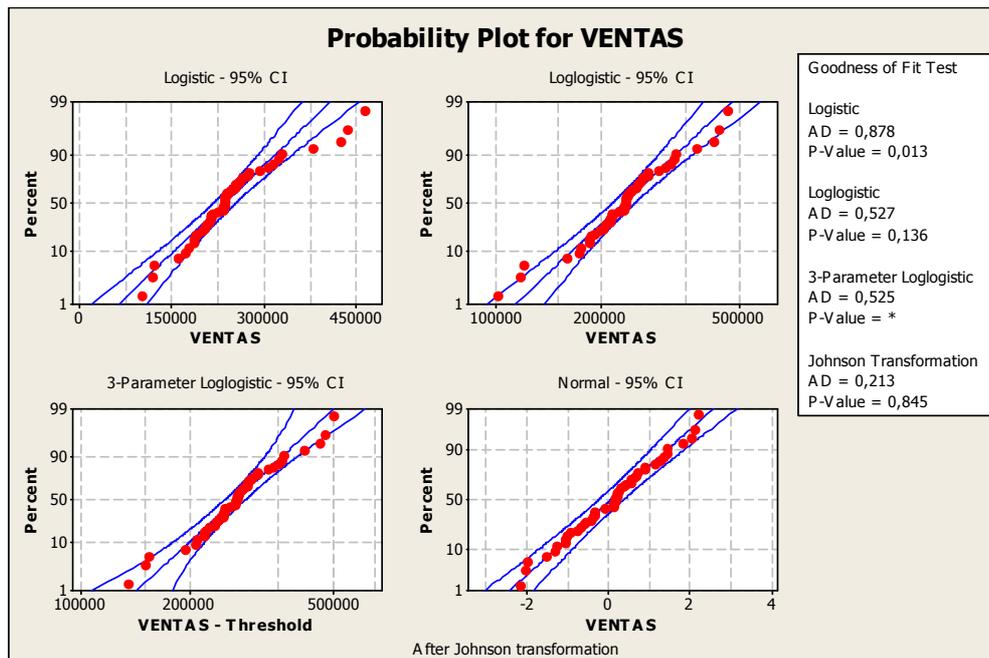


Figura 53. Distribución para los datos de Ventas del producto Cataflam. Fuente: Minitab.

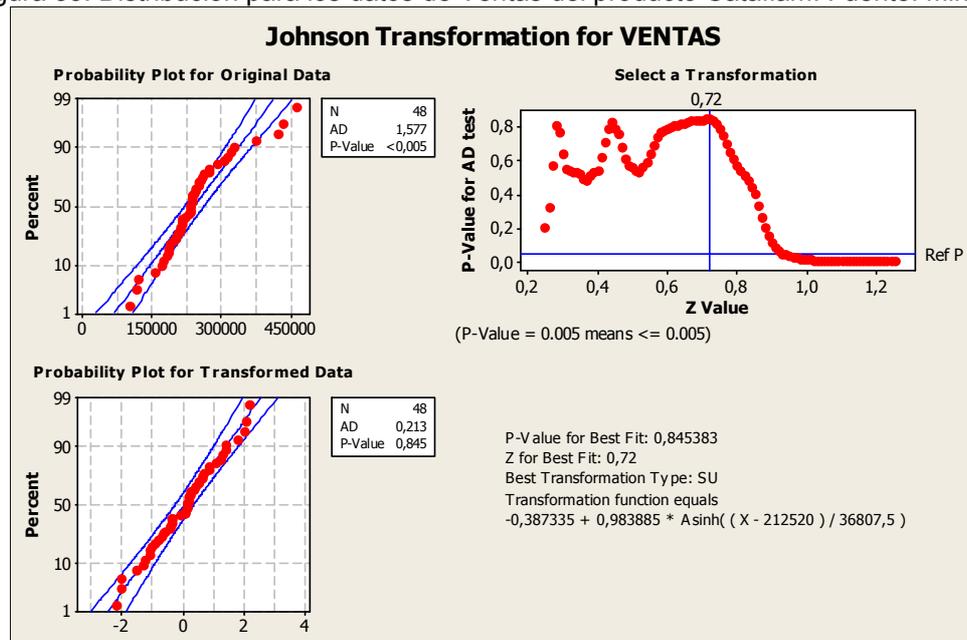


Figura 54. Transformación de Johnson de los datos de Ventas de Cataflam. Fuente: Minitab.

Un valor p igual a 0.945, muestra que la hipótesis nula de que los datos se acercan a una distribución normal, no puede ser rechazada.

En base a los parámetros obtenidos y a la Simulación Montecarlo, se generaron 5000 datos con este tipo de distribución que servirán posteriormente para elaborar el modelo de regresión que se ajuste a las ventas de Cataflam.

Inferencia en los datos generados de la variable Ventas para el producto Cataflam

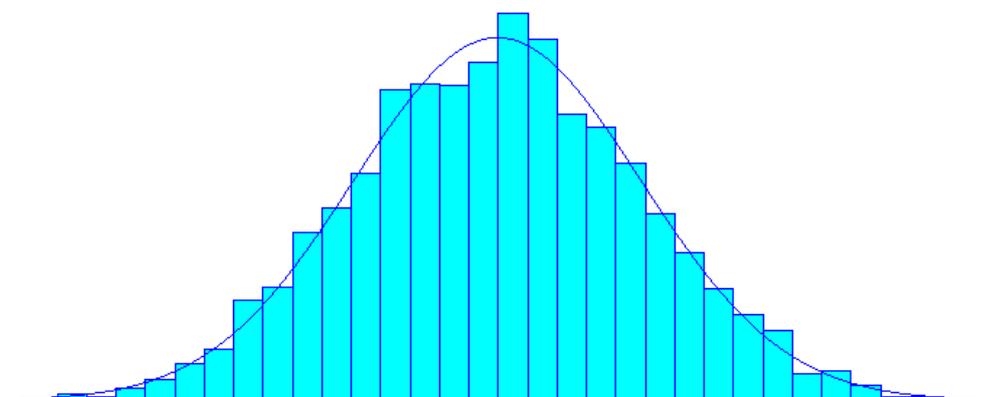


Figura 55. Distribución para los datos generados de Ventas del producto Cataflam. Fuente: Arena.

Distribution Summary

Distribution: Normal
Expression: NORM(0.213, 0.828)
 Square Error: 0.000279

Chi Square Test

Number of intervals = 28
 Degrees of freedom = 25
 Test Statistic = 43
Corresponding p-value = 0.0555

Kolmogorov-Smirnov Test

Test Statistic = 0.00905
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -2.59
 Max Data Value = 3.2
 Sample Mean = 0.213
 Sample Std Dev = 0.828

Histogram Summary

Histogram Range = -3 to 3.78
 Number of Intervals = 40

Ambos estadísticos, tanto de la prueba Chi cuadrado, como de la de Kolgomorov dan un valor mayor a 0.05, por lo que se acepta la hipótesis nula de que los datos se acercan a una distribución normal, claro está luego de una transformación de Johnson de ajuste de los datos a una distribución precisamente, normal.

Inferencia en la variable de entrada Advertisement & Promoting del producto Cataflam

Se estudiaron los datos correspondientes a gastos en Advertisement and Promoting de la marca Cataflam desde el año 2006 al 2009. La inferencia en ellos dio como resultado que mediante una transformación de Johnson, los datos se asemejan a una distribución normal. La ecuación de transformación es entonces:

$$1.22662 + 0.585830 * \text{Ln}\left(\frac{X + 157,960}{22559.1 - X}\right)$$

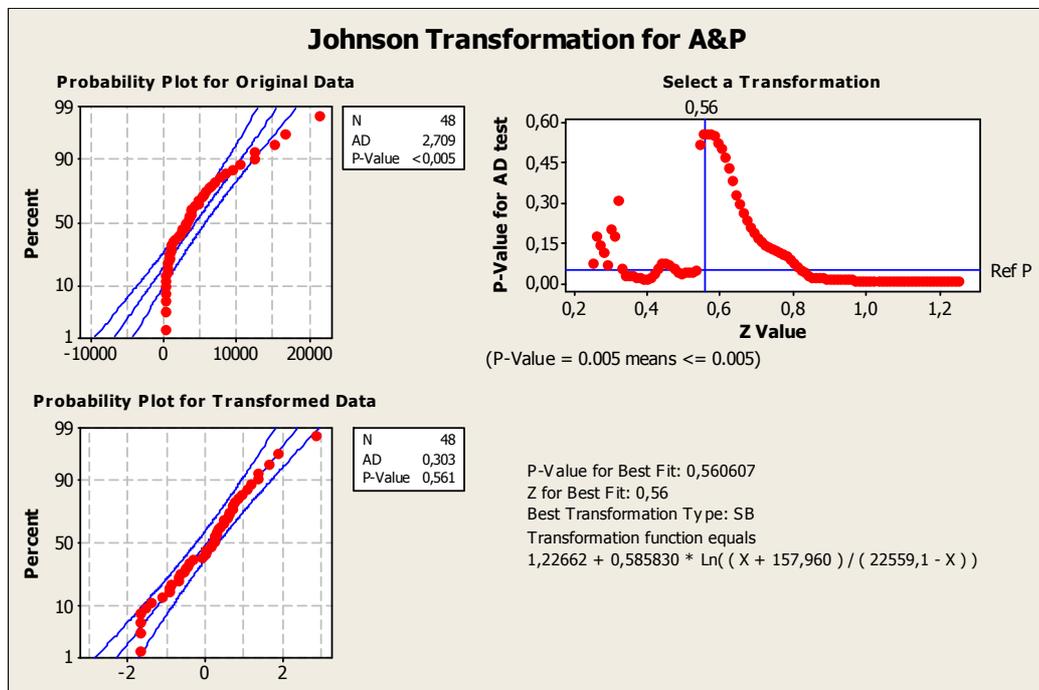


Figura 56. Transformación de Johnson de los datos de A&P de Cataflam. Fuente: Minitab.

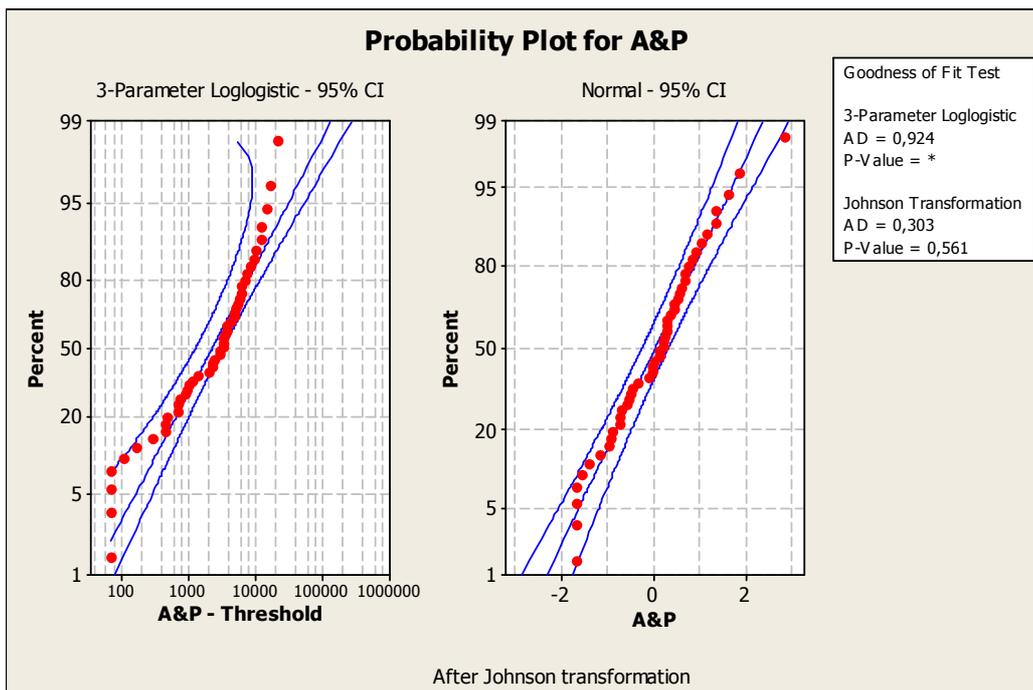


Figura 57. Distribución para los datos de A&P del producto Cataflam. Fuente: Minitab.

Un valor p de 0.561 afirma la aceptación de la hipótesis nula y por tanto se concluye que los datos de A&P siguen una distribución normal. En base a los parámetros obtenidos para este tipo de distribución, y con una simulación Montecarlo se generaron 5000 datos y se infirió nuevamente en ellos para comprobar que siguen la distribución adecuada. El resultado fue el siguiente.

Inferencia en los datos generados de la variable **Advertisement & Promoting** del producto **Cataflam**

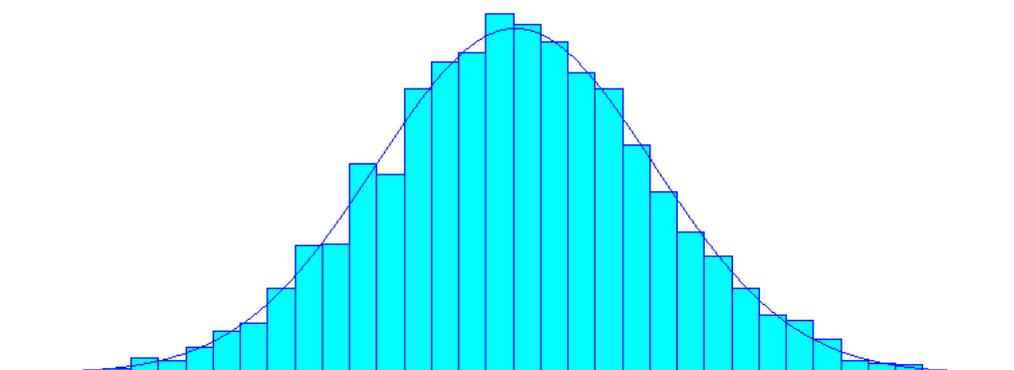


Figura 58. Distribución para los datos generados de A&P del producto Cataflam. Fuente: Arena

Distribution Summary

Distribution: Normal
Expression: NORM(0.011, 1.02)
 Square Error: 0.000186

Chi Square Test
 Number of intervals = 30
 Degrees of freedom = 27
 Test Statistic = 32.1
Corresponding p-value = 0.233

Kolmogorov-Smirnov Test
 Test Statistic = 0.00857
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -3.67
 Max Data Value = 3.46
 Sample Mean = 0.011
 Sample Std Dev = 1.02

Histogram Summary

Histogram Range = -4 to 4
 Number of Intervals = 40

Tanto el valor p de la prueba Chi Cuadrado, como el de la prueba de Kolgomorov dan resultados positivos en cuanto a la correcta asignación de una distribución normal a los datos generados.

Inferencia en la variable de entrada Descuentos en el producto Cataflam

Para inferir en la variable Descuentos de Cataflam se usaron datos desde el 2006 al 2009. Con un alfa de 0.05, la inferencia dio como resultado una distribución Normal, luego de una transformación de los datos por medio del algoritmo de Johnson. Los resultados de esta inferencia se presentan a continuación:

$$-1.14845 + 0.883945 * \operatorname{Asenh}\left(\frac{X - 48918.1}{10863.2}\right)$$

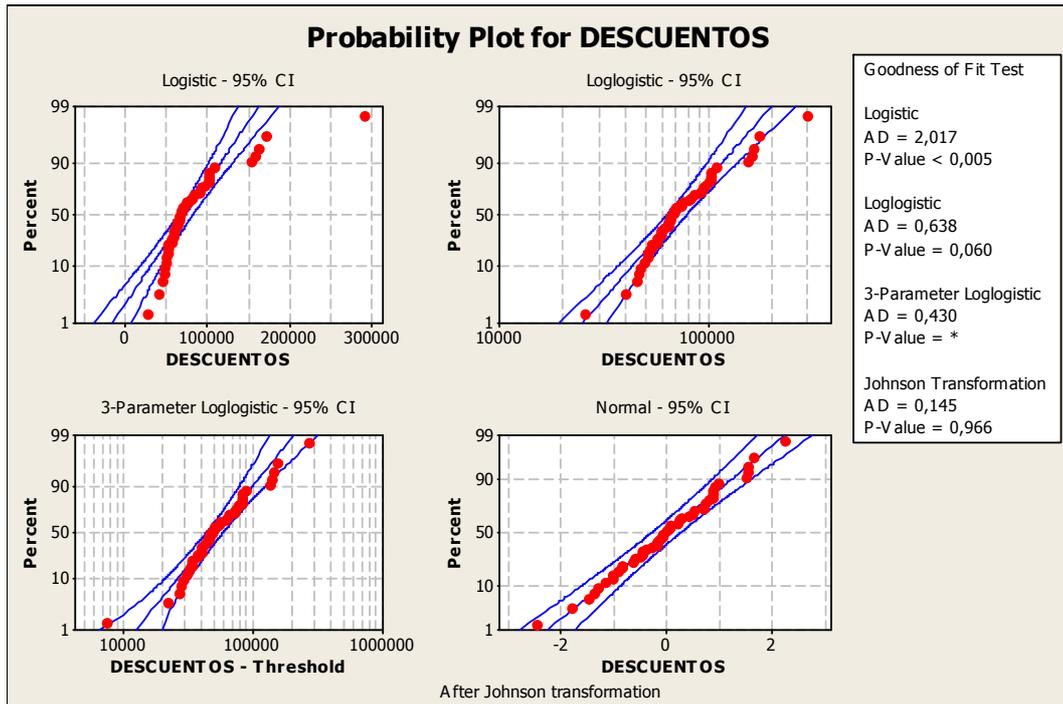


Figura 59. Distribución para los datos de Descuentos del producto Cataflam. Fuente: Minitab.

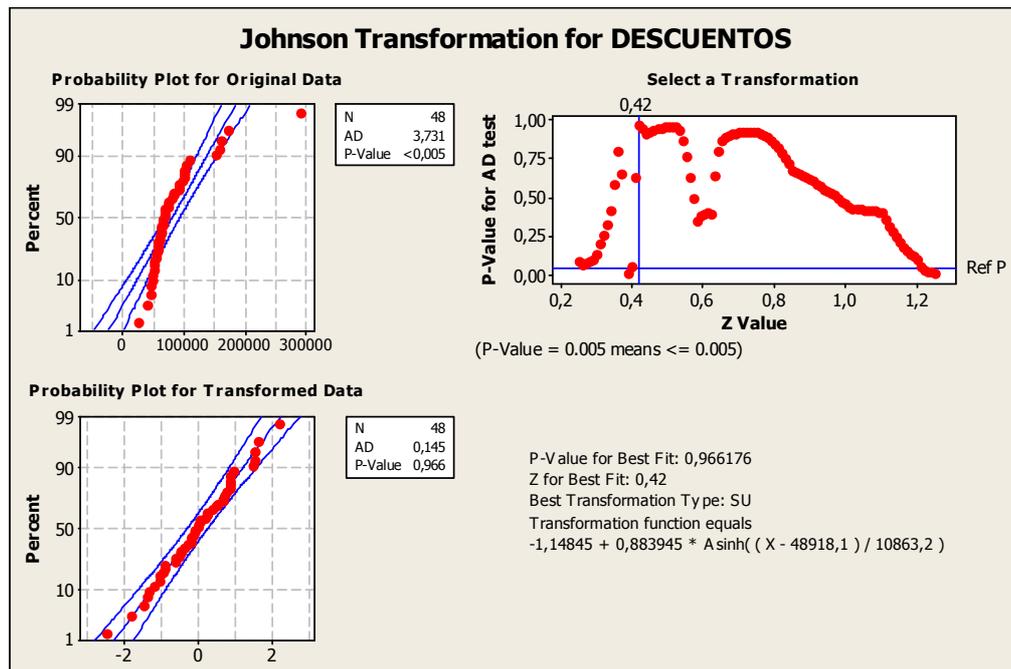


Figura 60. Transformación de Johnson de los datos de Descuentos de Cataflam. Fuente: Minitab.

Para la variable Descuentos del producto Cataflam el valor p del ajuste a una distribución normal es igual a 0.966, por lo que la hipótesis nula de que los datos se ajustan a este tipo de distribución no será rechazada.

Mediante a la Simulación Montecarlo, y a los parámetros obtenidos de la distribución normal, se generaron 5000 datos aleatorios. Su inferencia se muestra a continuación.

Inferencia en los datos generados de la variable Descuentos del producto Cataflam

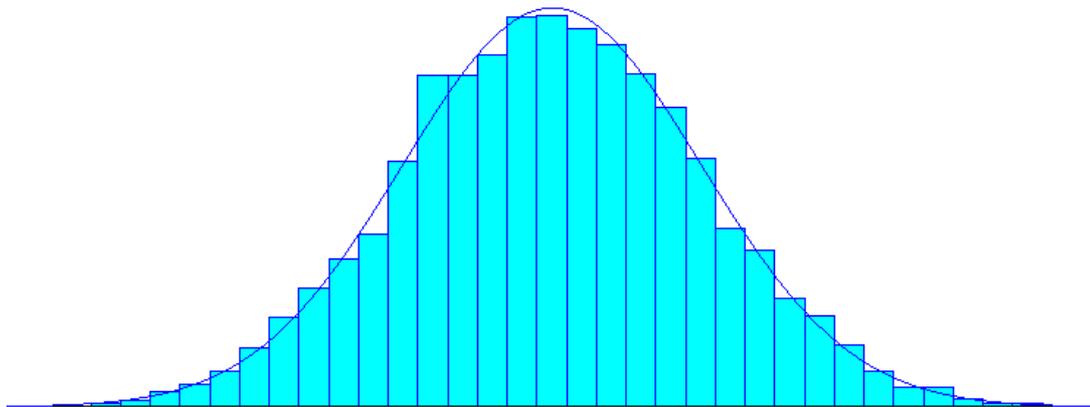


Figura 61. Distribución para los datos generados de Descuentos del producto Cataflam. Fuente: Arena

Distribution Summary

Distribution: Normal
Expression: NORM(-0.00295, 0.98)
 Square Error: 0.000140

Chi Square Test

Number of intervals = 29
 Degrees of freedom = 26
 Test Statistic = 21.1
Corresponding p-value = 0.735

Kolmogorov-Smirnov Test

Test Statistic = 0.00911
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -3.13
 Max Data Value = 3.57
 Sample Mean = -0.00295
 Sample Std Dev = 0.98

Histogram Summary

Histogram Range = -3.81 to 4
 Number of Intervals = 40

El valor p mayor a 0.735 y 0.15 para las pruebas Chi Cuadrado y Kolgomorov respectivamente obtenidos, ambos mayores al nivel de significancia, comprueba que la distribución de los datos es normal con parámetros iguales o muy parecidos a la distribución original.

Inferencia en la variable de entrada Muestras Médicas del producto Cataflam

Finalmente se infiere en la variable Muestras Médicas del producto Cataflam. Con un alfa igual a 0.05 y la misma cantidad de datos, para esta variable la distribución más adecuada es una distribución exponencial con la siguiente forma y parámetros:

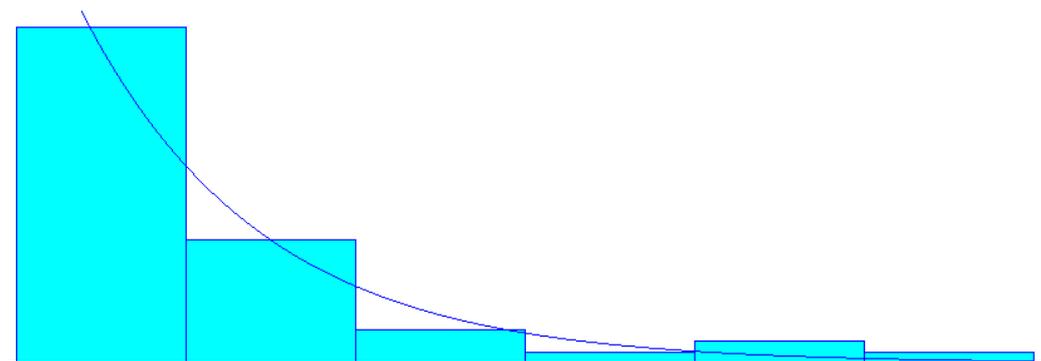


Figura 62. Distribución para los datos de MM del producto Cataflam. Fuente: Arena

Distribution Summary

Distribution: Exponential
Expression: $-0.001 + EXPO(2.63e+004)$
 Square Error: 0.002424

Chi Square Test
 Number of intervals = 3
 Degrees of freedom = 1
 Test Statistic = 0.802
Corresponding p-value = 0.4

Kolmogorov-Smirnov Test
 Test Statistic = 0.313
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 48
 Min Data Value = 0
 Max Data Value = 1.49e+005
 Sample Mean = 2.63e+004
 Sample Std Dev = 3.33e+004

Histogram Summary

Histogram Range = -0.001 to 1.49e+005
 Number of Intervals = 6

Al igual que para las demás variables, el estadístico p muestra que la distribución Exponencial es la adecuada para las variable Muestras Médicas. Generando 5000 datos con simulación Montecarlo y los parámetros adecuados, se infirió en ellos. Sus resultados a continuación:

Inferencia en los datos generados de la variable Muestras Médicas

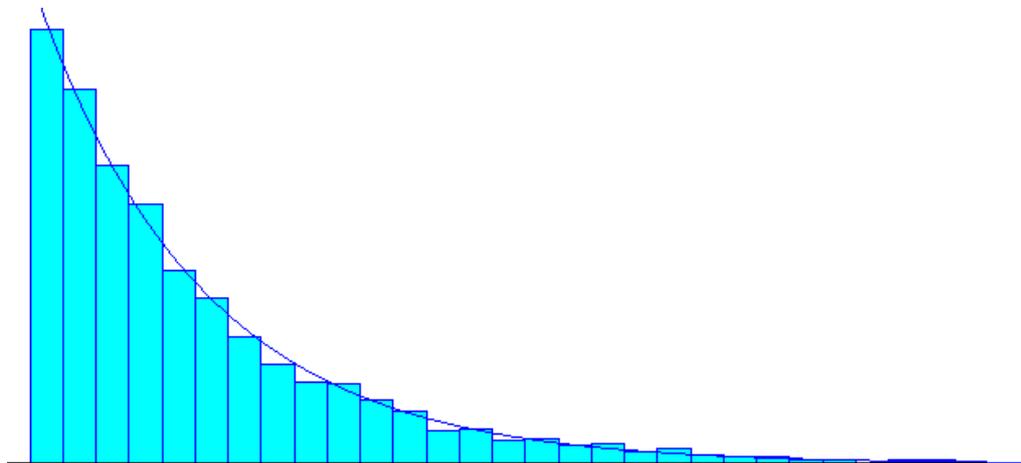


Figura 63. Distribución para los datos generados de MM del producto Cataflam. Fuente: Arena

Distribution Summary

Distribution: Exponential
Expression: $-558 + \text{EXPO}(2.68e+004)$
 Square Error: 0.000161

Chi Square Test

Number of intervals = 27
 Degrees of freedom = 25
 Test Statistic = 23.4
Corresponding p-value = 0.553

Kolmogorov-Smirnov Test

Test Statistic = 0.0101
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -558
 Max Data Value = 2.11e+005
 Sample Mean = 2.62e+004
 Sample Std Dev = 2.71e+004

Histogram Summary

Histogram Range = -558 to 2.11e+005
 Number of Intervals = 40

Ambos estadísticos prueban que la distribución exponencial es la adecuada para los datos generados de la variable Muestras Médicas.

Modelo de regresión con los datos generados para el producto Cataflam

En base a las distribuciones inferidas de cada variable para el producto Cataflam, se prosiguió el mismo proceso que para los anteriores dos productos, y mediante una simulación Montecarlo de 5000 datos que se ajusten a dichas distribuciones, se procedió a correr un modelo de regresión múltiple. Igualmente se desea conocer el impacto de las variables de entrada, sobre la variable de salida, Ventas de Cataflam.

Las hipótesis a probar siguen siendo las mismas que antes y son las siguientes:

$$H_0 : \sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$$

$$H_0 : \sigma_1 \neq \sigma_2 \neq \sigma_3 \neq \sigma_4$$

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$$

$$H_0 : \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4$$

Con un valor p menor al valor alfa establecido, se rechaza la hipótesis nula que establece,

H_0 = El factor no influye de manera significativa en el modelo.

H_1 = El factor sí influye de manera significativa en el modelo.

La regresión múltiple para los datos de producto Cataflam, con solamente tres variables de entrada: Advertisement & Promoting, Descuentos y Muestras Médicas, es el siguiente:

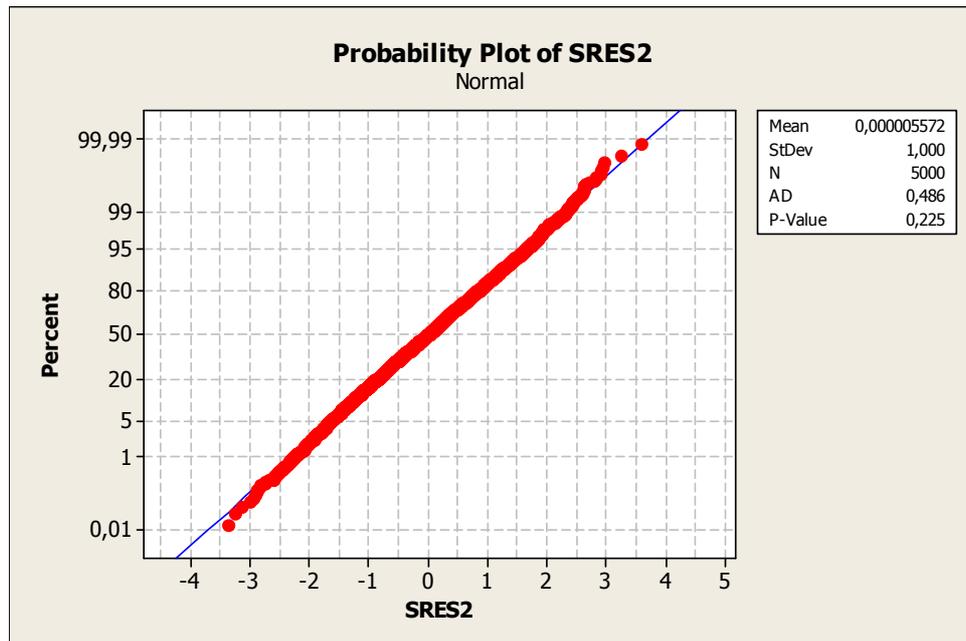


Figura 64. Distribución de probabilidad normal para el modelo de regresión de Cataflam. Fuente: Minitab.

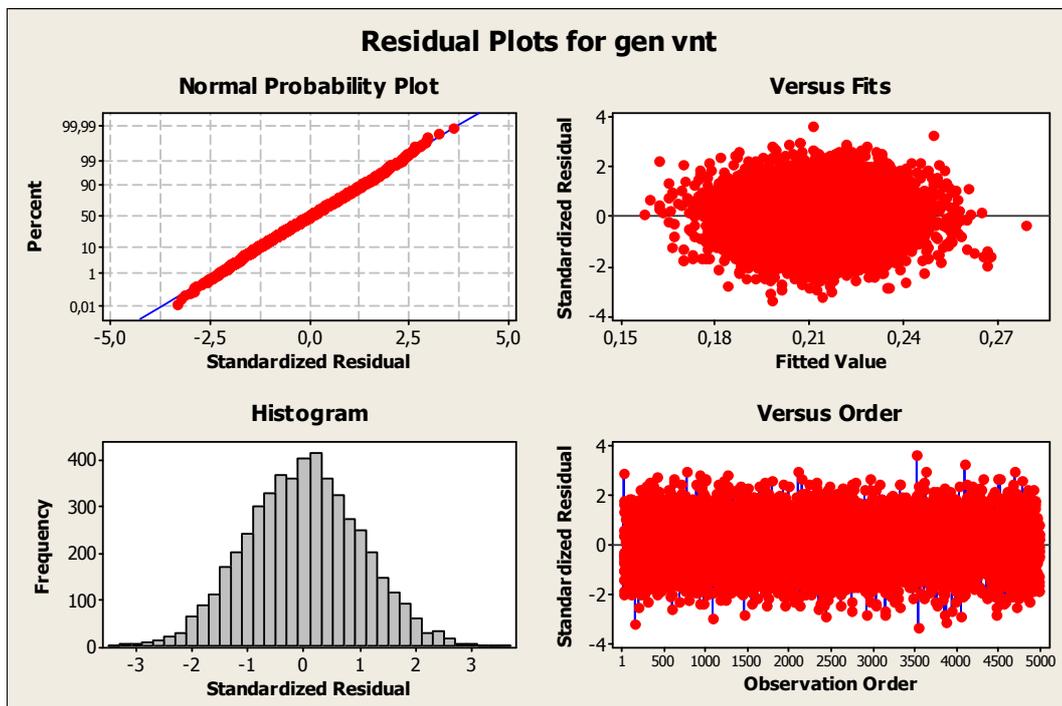


Figura 65. Residuales del modelo de regresión múltiple para el producto Cataflam. Fuente: Minitab

Regression Analysis: gen vnt versus gen ap. gen des. gen mm

The regression equation is
 $gen\ vnt = 144597 - 3,55\ gen\ ap + 1,33\ gen\ des + 0,194\ gen\ mm$

Predictor	Coef	SE Coef	T	P	VIF
Constant	144597	14452	10,01	0,000	
gen ap	-3,547	1,262	-2,81	0,007	1,007
gen des	1,3295	0,1349	9,86	0,000	1,000
gen mm	0,1939		0,1816	1,07	0,291 1,007

S = 41310,0 R-Sq = 70,8% R-Sq(adj) = 68,8%

PRESS = 131056235277 R-Sq(pred) = 49,01%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression		3 1,81959E+11	60652927308	35,54	0,000
Residual Error	4996	75086692864	1706515747		
Total	4999	2,57045E+11			

Source	DF	Seq SS
gen ap	1	14245687524
gen des	1	1,65767E+11
gen mm	1	1945953745

Se analizará primero el gráfico de residuales obtenido. Se ve que los residuales de la regresión obtenida se ajustan perfectamente a la recta y de igual manera y mediante el histograma, donde se comprueba la existencia de una forma de campana, que es prueba de la normalidad de los datos. No hay necesidad de una transformación como en los otros casos. Esto satisface el supuesto 11 del método de los Mínimos Cuadrados. Las dos gráficas adicionales muestran que no hay ningún patrón que señale la evidencia de un comportamiento anormal de los residuales, y tal como en el anterior producto se puede asumir que la varianza si llega a ser constante. Con esto se cumple el supuesto número 4.

Como se explicó antes, para el cumplimiento del supuesto 3, debe existir la constante de la pendiente β_1 , valor que el modelo sí presenta.

De igual manera se procedió a obtener el Factor Inflador de la Varianza (FIV) para analizar el supuesto de no multicolinealidad entre las variables regresoras. Todos ellos se acercan al valor de 1, lo que muestra que las variables

no guardan relación de multicolinealidad entre ellas; por tanto el supuesto 5 y 6 quedan satisfechos.

El modelo de regresión obtenido es lineal, con lo que el supuesto 1 se cumple. El supuesto 7 queda satisfecho al usar 5000 datos, muy superior al número de variables usadas, tres. Los demás supuestos 8 y 9, pueden presumirse como correctos con los resultados expuestos por Minitab.

Con los supuestos del Método de los Mínimos Cuadrados satisfechos, se debe analizar los resultados obtenidos en cuanto a la influencia o no de las variables regresoras. El análisis de varianza determina que sí existe al menos una variable regresora no trivial en el modelo. En este caso, para el producto Cataflam, tenemos dos variables de influencia significativa en el modelo de regresión. Tanto la variable Descuentos como la variable Advertisement & Promoting, presentan valores p menores al nivel de significancia. Esto indica que ambas variables no pueden ser excluidas del modelo y su aporte es importante para las Ventas de Cataflam. Esto se comprueba con los coeficientes obtenidos de la función de regresión, en la cual ambas variables llevan los valores más altos.

El R^2 -adj obtenido muestra el grado de ajuste del modelo a los datos, que en este caso corresponde a un 68.8%. Este valor es mayor al obtenido con el modelo del producto Voltaren, lo que puede implicar que la recolección de datos para Cataflam es más certera y su comportamiento es más fácil de predecir. Además al tener dos variables de tres, significativas en el modelo, hace que el R^2 sea más grande.

Se han obtenido entonces dos variables importantes para las Ventas de Cataflam, Advertisement and Promoting y Descuentos. Es curioso notar que

comparado al Voltaren, en el cual no aparecía la variable A&P como significativa, a pesar de que ambos productos pertenecen a la misma línea de productos maduros.

Más adelante se podrá determinar qué valor es el óptimo, de todas las variables influyentes, prestando especial atención en las dos regresoras significativas obtenidas.

ANEXO B-3 PRODUCTO DIOVAN

Cataflam es un producto para el tratamiento de la hipertensión arterial esencial. Es una medicina que se vende bajo prescripción médica y que pertenece a la línea de productos Cardiovascular. Es actualmente el producto que genera mayores ventas en Novartis en el ámbito nacional.

A diferencia del anterior producto, Diovan si mantiene un gasto en Fuerza de Ventas, por lo que está variable tiene que ser añadida en el modelo. Diovan es un producto interesante de ser analizado ya que actualmente, y como se mencionó antes, es el producto que genera mayores ventas para toda la compañía y se recibe mayores ingresos por la venta de éste.

Inferencia para la variable de salida Ventas de Diovan

Para Diovan, se estudiaron datos desde el año 2006 al 2009. Con el mismo nivel de significancia que para los demás productos, es decir un alfa de 0.05 se infirió en los datos de la variable de salida, Ventas para conocer el tipo de distribución a la que se ajustan.

El resultado de esto, mediante Minitab, fue el siguiente. La variable Ventas se ajusta a una distribución Weibull, con la siguiente gráfica de ajuste:

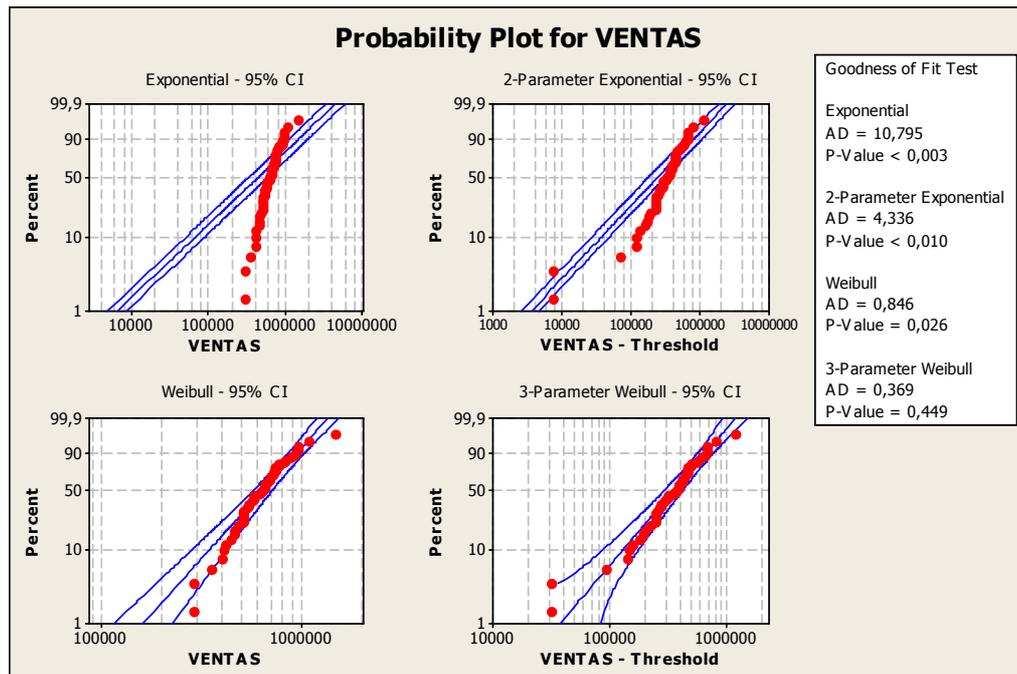


Figura 66. Distribución para los datos de Ventas del producto Diovan. Fuente: Minitab.

Un valor p igual a 0.883 asegura que los datos correspondientes a Ventas de Diovan se ajustan a una distribución Weibull de tres parámetros.

En base a los parámetros obtenidos y a la Simulación Montecarlo, se generaron 5000 datos con este tipo de distribución que servirán posteriormente para elaborar el modelo de regresión que se ajuste a las ventas de Diovan.

Inferencia en los datos generados de la variable Ventas para el producto Diovan

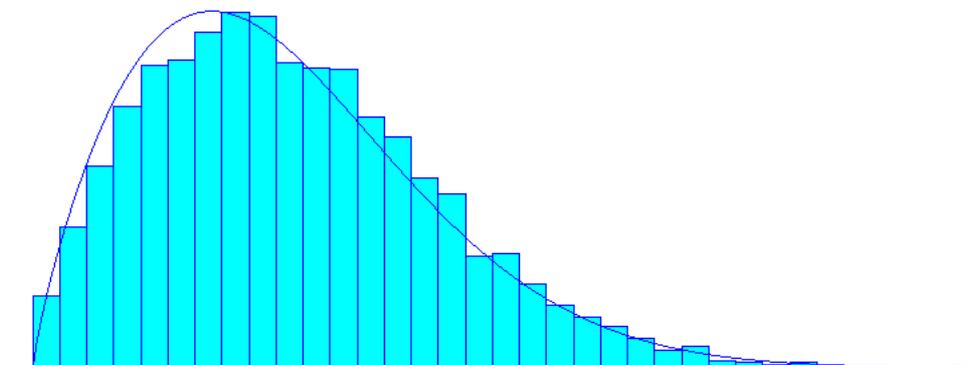


Figura 67. Distribución para los datos generados de Ventas del producto Diovan. Fuente: Arena.

Distribution Summary

Distribution: Weibull
Expression: $2.62e+005 + WEIB(4.09e+005, 1.79)$
 Square Error: 0.000493

Chi Square Test

Number of intervals = 28
 Degrees of freedom = 25
 Test Statistic = 58.1
 Corresponding p-value > 0.15

Kolmogorov-Smirnov Test

Test Statistic = 0.0403
 Corresponding p-value > 0.10

Data Summary

Number of Data Points = 5000
 Min Data Value = $2.62e+005$
 Max Data Value = $1.83e+006$
 Sample Mean = $6.4e+005$
 Sample Std Dev = $2.09e+005$

Histogram Summary

Histogram Range = $2.62e+005$ to $1.83e+006$
 Number of Intervals = 40

Ambos estadísticos, tanto de la prueba Chi cuadrado, como de la de Kolmogorov dan un valor mayor a 0.05, por lo que se acepta la hipótesis nula de que los datos se acercan a una distribución Weibull de tres parámetros.

Inferencia en la variable de entrada Fuerza de Ventas del producto Diovan

Se estudiaron los datos correspondientes a gastos en Fuerza de Ventas del producto Diovan. Información que data desde el año 2006 al 2009. La inferencia en ellos dio como resultado que mediante una transformación de Johnson, los datos se asemejan a una distribución normal. La ecuación de transformación es entonces:

$$-12.9409 + 1.54205 * Ln(X - 853.966)$$

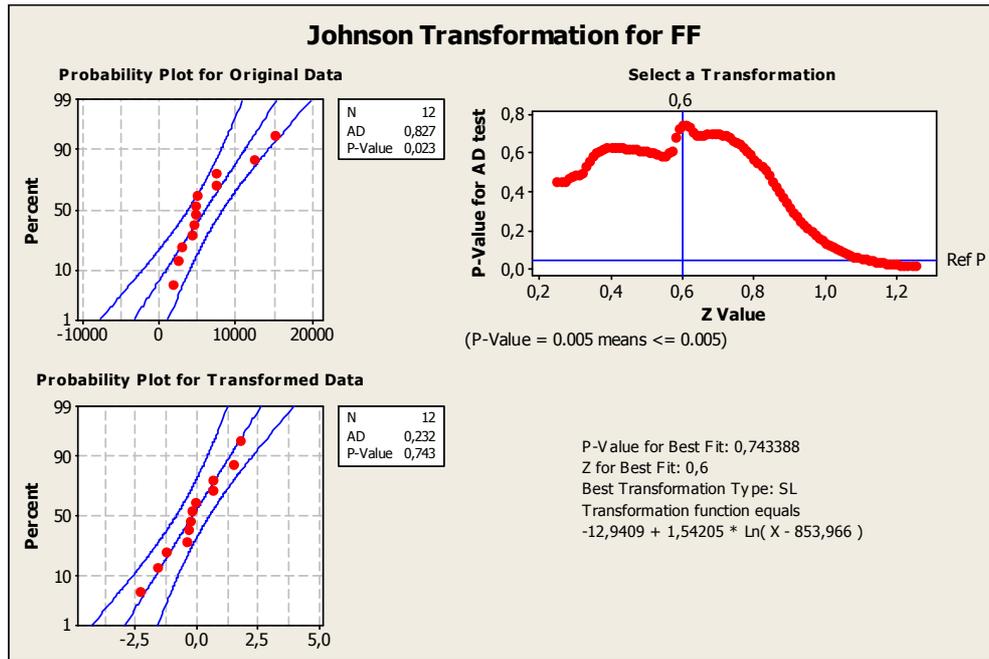


Figura 68. Transformación de Johnson de los datos de FF de Diovan. Fuente: Minitab.

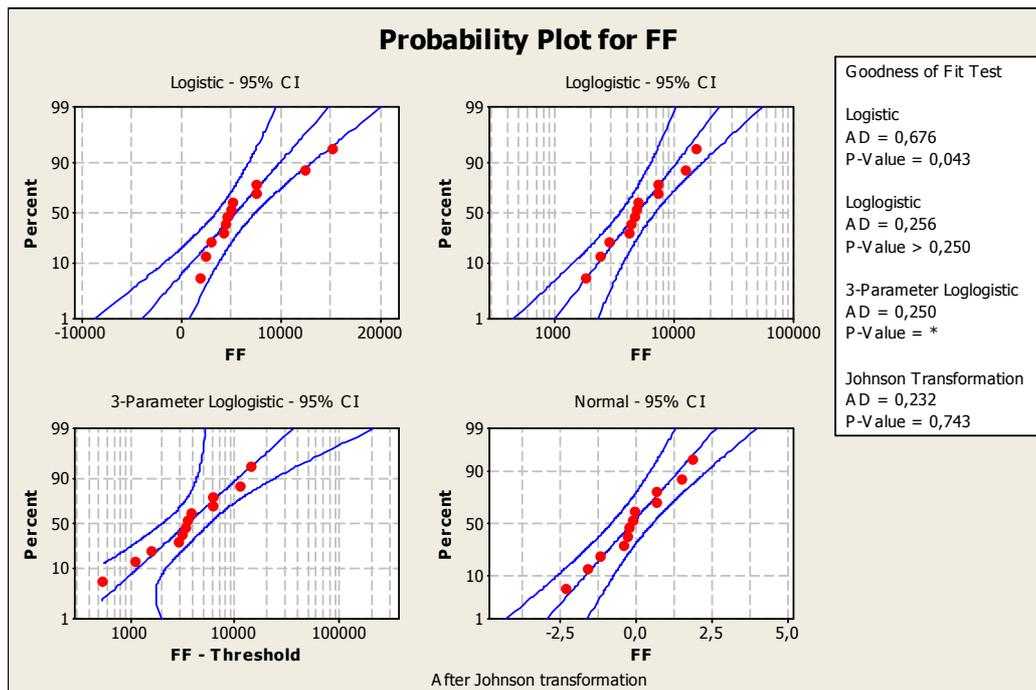


Figura 69. Distribución para los datos de FF del producto Diovan. Fuente: Minitab.

Un valor p de 0.743 afirma la aceptación de la hipótesis nula y por tanto se concluye que los datos de Fuerza de Ventas siguen una distribución normal. En base a los parámetros obtenidos para este tipo de distribución, y en base a una

simulación Montecarlo se generaron 5000 datos y se infirió nuevamente en ellos para comprobar que siguen la distribución adecuada. El resultado fue el siguiente.

Inferencia en los datos generados de la variable Fuerza de Ventas del producto Diovan

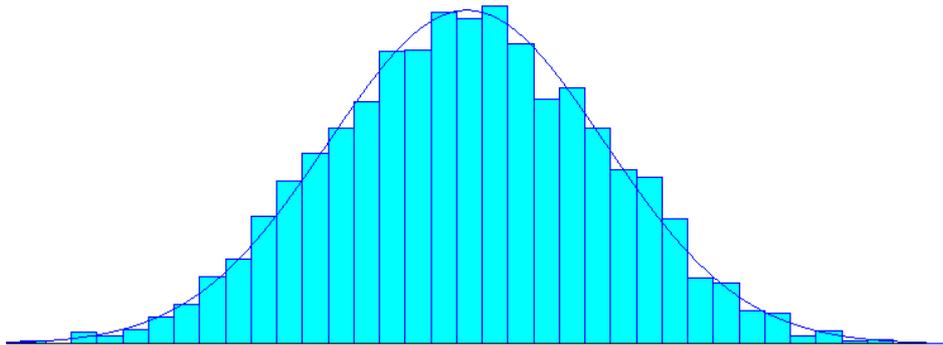


Figura 70. Distribución para los datos generados de FF del producto Diovan. Fuente: Arena.

Distribution Summary

Distribution: Normal
Expression: NORM(-0.158, 1.2)
 Square Error: 0.000222

Chi Square Test

Number of intervals = 31
 Degrees of freedom = 28
 Test Statistic = 38.8
Corresponding p-value = 0.0864

Kolmogorov-Smirnov Test

Test Statistic = 0.0102
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -4.13
 Max Data Value = 3.83
 Sample Mean = -0.158
 Sample Std Dev = 1.2

Histogram Summary

Histogram Range = -4.93 to 4
 Number of Intervals = 40

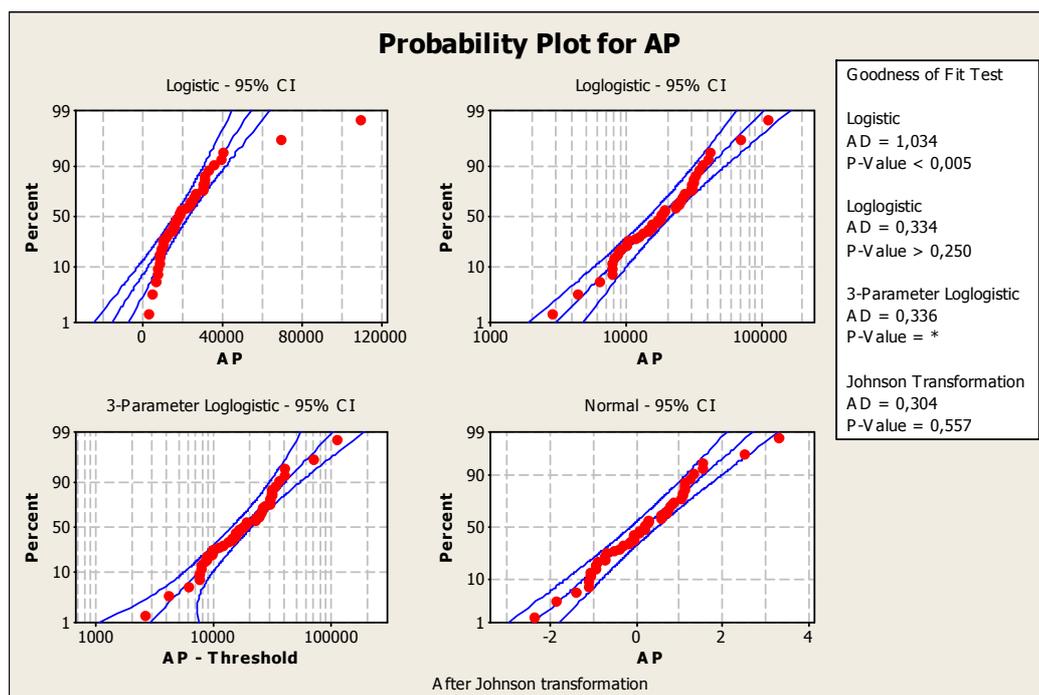
Tanto el valor p de la prueba Chi Cuadrado, como el de la prueba de Kolgomorov dan resultados positivos en cuanto a la correcta asignación de una

distribución normal a los datos generados tal como presentaron los datos originales.

Inferencia en la variable de entrada Advertisement and Promoting del producto Diovan

Con data desde el 2006 al 2009, se infirieron en todos los gastos incurridos en Advertisement and Promoting del producto Diovan. Con un alfa de 0.05, se obtuvo que la distribución de probabilidad más adecuada a los datos sea una distribución Normal luego de una transformación de Johnson, con ecuación:

$$-17,7803 + 1,81322 * \text{LN}(X + 2073,78)$$



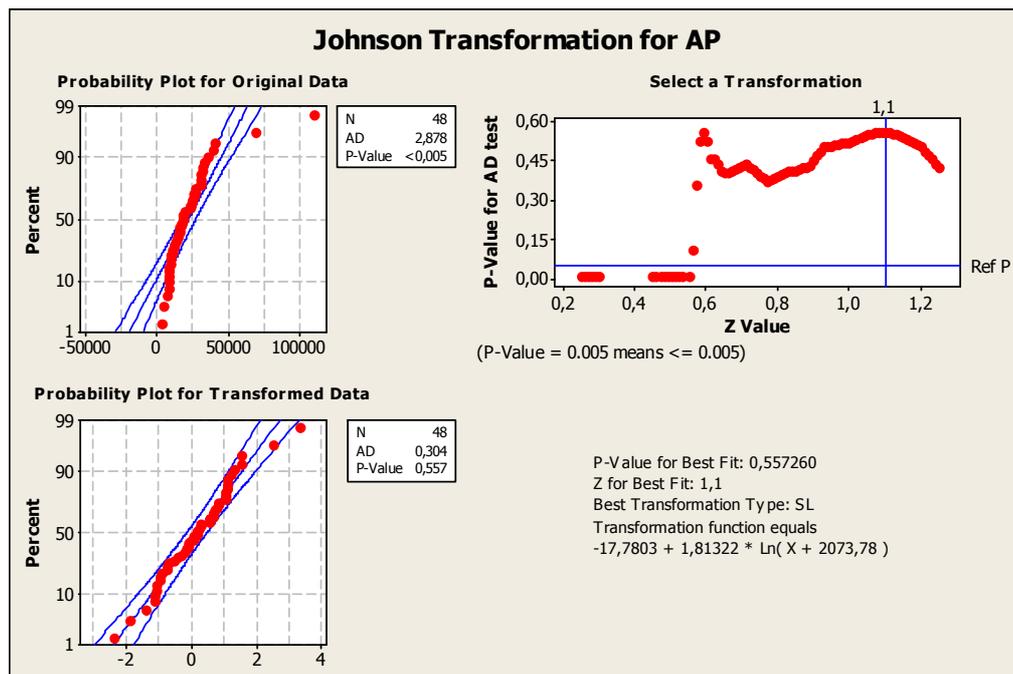


Figura 71. Distribución para los datos de A&P del producto Diovan. Fuente: Minitab.

Con un valor p igual a 0.557 no se rechaza la hipótesis nula y se concluye que los datos, luego de una transformación de Johnson adecuada, siguen una distribución normal.

Para comprobar que esta distribución sea la misma en los 5000 datos generados a partir de una simulación Montecarlo, se realizó la misma prueba de hipótesis. Se obtuvo lo siguiente:

Inferencia en los datos generados de la variable Advertisement and Promoting del producto Diovan

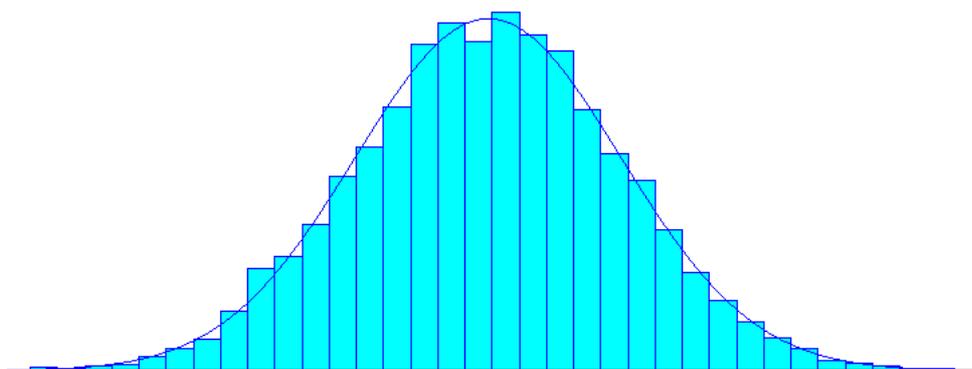


Figura 72. Distribución para los datos generados de A&P del producto Diovan. Fuente: Arena

Distribution Summary

Distribution: Normal
Expression: NORM(0.139, 1.1)
 Square Error: 0.000119

Chi Square Test

Number of intervals = 28
 Degrees of freedom = 25
 Test Statistic = 16.4
Corresponding p-value > 0.75

Kolmogorov-Smirnov Test

Test Statistic = 0.00692
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -4.65
 Max Data Value = 3.91
 Sample Mean = 0.139
 Sample Std Dev = 1.1

Histogram Summary

Histogram Range = -5 to 4
 Number of Intervals = 40

Los valores p para ambas pruebas de bondad de ajuste muestran que la distribución Normal es la adecuada para los datos generados a partir de los datos muestrales pertenecientes a la variable Advertisement and Promoting.

Inferencia en la variable de entrada Descuentos en el producto Diovan

Para inferir en la variable Descuentos de Diovan se usaron datos desde el 2006 al 2009. Con un alfa de 0.05, la inferencia dio como resultado una distribución Normal, luego de una transformación de los datos por medio del algoritmo de Johnson. Los resultados de esta inferencia se presentan a continuación:

$$-0,744537 + 1,07734 * \text{Asinh}((X - 184534) / 73518,5)$$

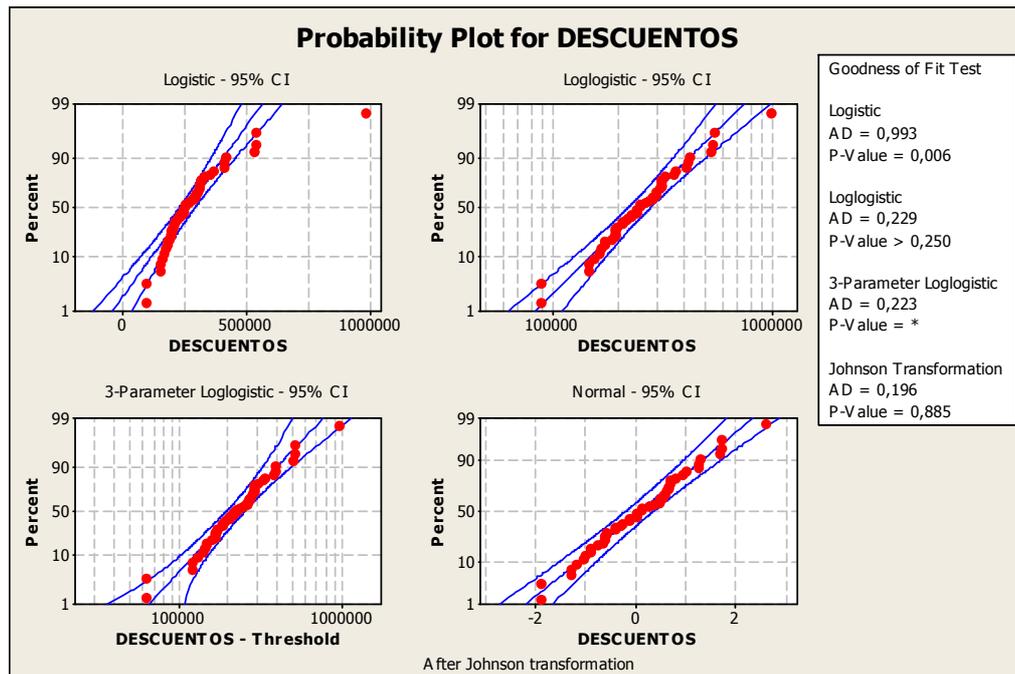


Figura 73. Distribución para los datos de Descuentos del producto Diovan. Fuente: Minitab.

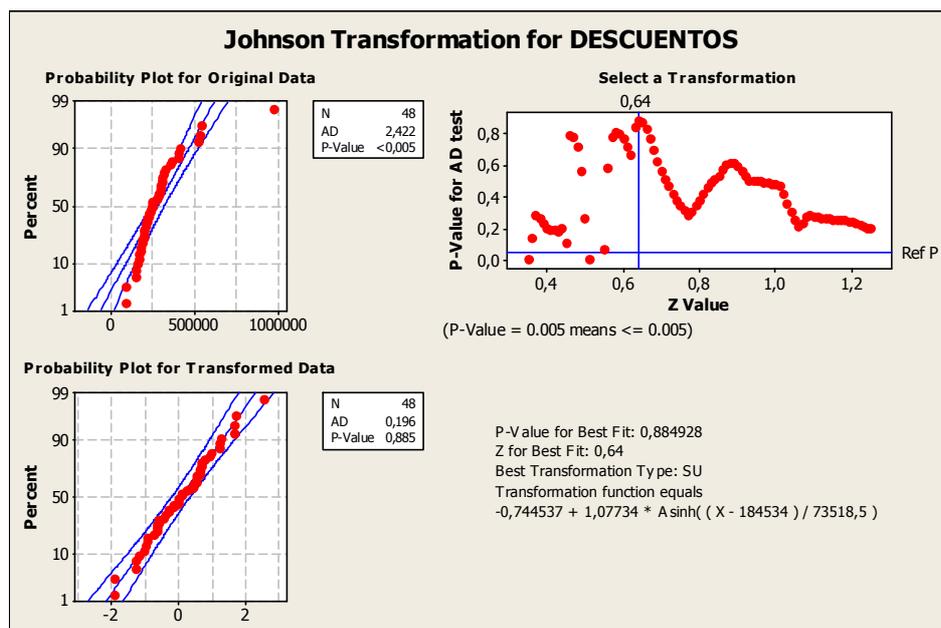


Figura 74. Transformación de Johnson de los datos de Descuentos de Diovan. Fuente: Minitab.

Para la variable Descuentos del producto Diovan el valor p del ajuste a una distribución normal es igual a 0.885, por lo que se acepta la hipótesis nula de que los datos se ajustan a este tipo de distribución.

Ahora, con la simulación Montecarlo y tal como en los otros casos, se generarán 5000 datos aleatorios. Su inferencia se muestra a continuación.

Inferencia en los datos generados de la variable Descuentos del producto

Diovan

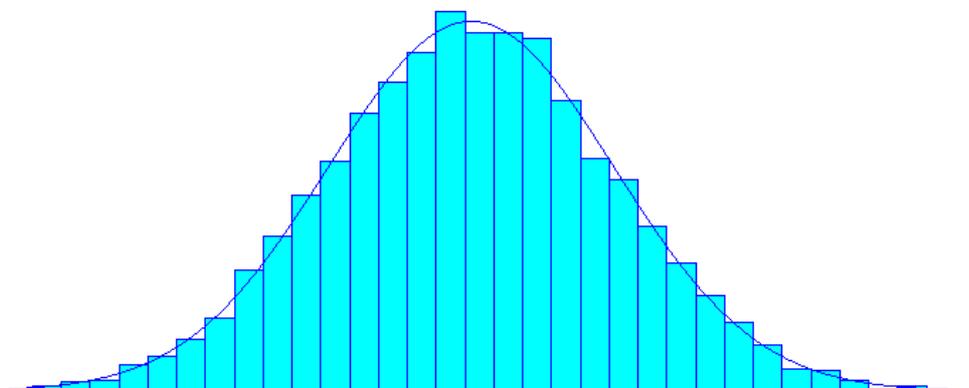


Figura 75. Distribución para los datos generados de Descuentos del producto Diovan. Fuente: Arena

Distribution Summary

Distribution: Normal
Expression: NORM(0.0291, 0.97)
 Square Error: 0.000087

Chi Square Test

Number of intervals = 29
 Degrees of freedom = 26
 Test Statistic = 13.1
Corresponding p-value > 0.75

Kolmogorov-Smirnov Test

Test Statistic = 0.00572

Data Summary

Number of Data Points = 5000
 Min Data Value = -3.27
 Max Data Value = 3.25
 Sample Mean = 0.0291
 Sample Std Dev = 0.97

Histogram Summary

Histogram Range = -3.93 to 3.91
 Number of Intervals = 40

Tanto el valor p para la prueba de bondad de ajuste de Chi Cuadrado, como para la prueba de Kolmogorov, es mayor al nivel de significancia escogido de 0.05. Por lo tanto la hipótesis nula de que los datos se ajustan a una distribución normal no puede ser rechazada.

Inferencia en la variable de entrada Muestra Médicas del producto Diovan

Para terminar el estudio de las variables de entrada del producto Diovan, se analizan los datos que corresponden a los gastos por Muestras Médicas incurridos en los años del 2006 al 2009. Un ajuste en el software Minitab, muestra que la distribución más cercana es una de tipo Exponencial, con la siguiente forma y parámetros:

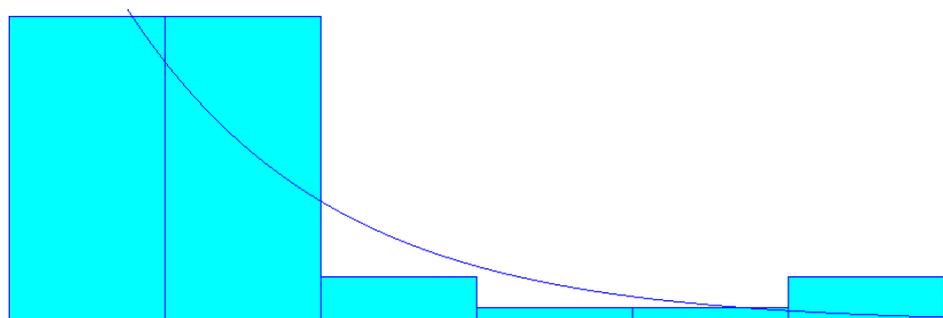


Figura 76. Distribución para los datos de MM del producto Diovan. Fuente: Arena

Distribution Summary

Distribution: Exponential
Expression: $-0.001 + \text{EXPO}(2.37\text{e}+005)$
 Square Error: 0.048398

Chi Square Test

Number of intervals = 3
 Degrees of freedom = 1
 Test Statistic = 8.23
Corresponding p-value > 0.05

Kolmogorov-Smirnov Test

Test Statistic = 0.269
Corresponding p-value > 0.05

Data Summary

Number of Data Points = 48
 Min Data Value = 0
 Max Data Value = $1.08\text{e}+006$
 Sample Mean = $2.37\text{e}+005$
 Sample Std Dev = $2.6\text{e}+005$

Histogram Summary

Histogram Range = -0.001 to $1.08\text{e}+006$
 Number of Intervals = 6

Si bien ninguno de los dos estadísticos sobrepasa con una fuerte diferencia al nivel de significancia, se puede aceptar que la variable Muestras Médicas se ajusta a este tipo de distribución.

En base a los parámetros ofrecidos se genera 5000 datos y se comprueba su distribución en el siguiente apartado.

Inferencia en los datos generados de la variable Muestras Médicas de Diovan

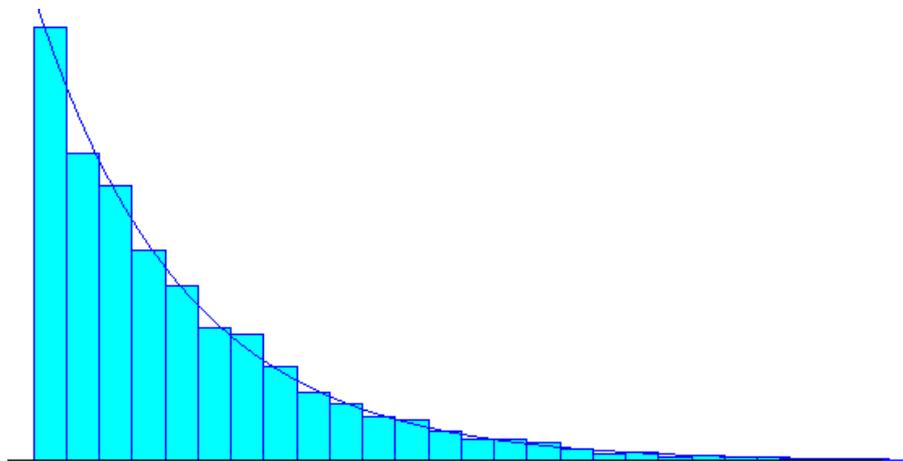


Figura 77. Distribución para los datos generados de MM del producto Diovan. Fuente: Minitab.

Distribution Summary

Distribution: Exponential
Expression: $-4.87e+003 + EXPO(2.42e+005)$
 Square Error: 0.000320

Chi Square Test

Number of intervals = 24
 Degrees of freedom = 22
 Test Statistic = 22
Corresponding p-value = 0.465

Kolmogorov-Smirnov Test

Test Statistic = 0.0109
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = $-4.87e+003$
 Max Data Value = $2.12e+006$
 Sample Mean = $2.37e+005$
 Sample Std Dev = $2.39e+005$

Histogram Summary

Histogram Range = -4.87e+003 to 2.12e+006
 Number of Intervals = 40

Ambos estadísticos prueban que la distribución exponencial es la adecuada para los datos generados de la variable Muestras Médicas.

Modelo de regresión con los datos generados para el producto Diovan

En base a las distribuciones inferidas de cada variable para el producto Diovan, se prosiguió el mismo proceso que para los anteriores productos, y mediante una simulación Montecarlo de 5000 datos que se ajusten a dichas distribuciones, se procedió a correr un modelo de regresión múltiple.

Las hipótesis a probar siguen siendo las mismas que antes y son las siguientes:

$$H_0 : \sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$$

$$H_0 : \sigma_1 \neq \sigma_2 \neq \sigma_3 \neq \sigma_4$$

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$$

$$H_0 : \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4$$

Con un valor p menor al valor alfa establecido, se rechaza la hipótesis nula que establece,

H_0 = El factor no influye de manera significativa en el modelo.

H_1 = El factor sí influye de manera significativa en el modelo.

La regresión múltiple para los datos de producto Diovan, con todas las variables analizadas: Fuerza de Ventas, Advertisement & Promoting, Descuentos y Muestras Médicas, es el siguiente:

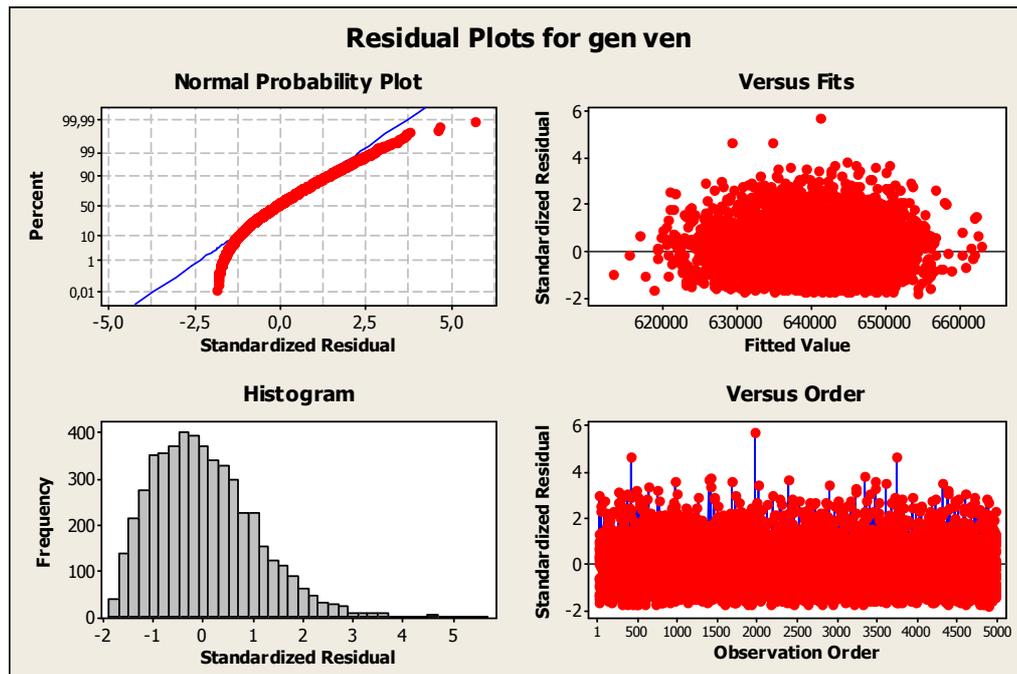


Figura 78. Residuales del modelo de regresión múltiple para el producto Diovan. Fuente: Minitab.

Claramente se ve que los residuales no se ajustan a una distribución normal, lo que se comprueba con el valor p obtenido es menor al nivel de significancia. Éste se observa en la gráfica siguiente:

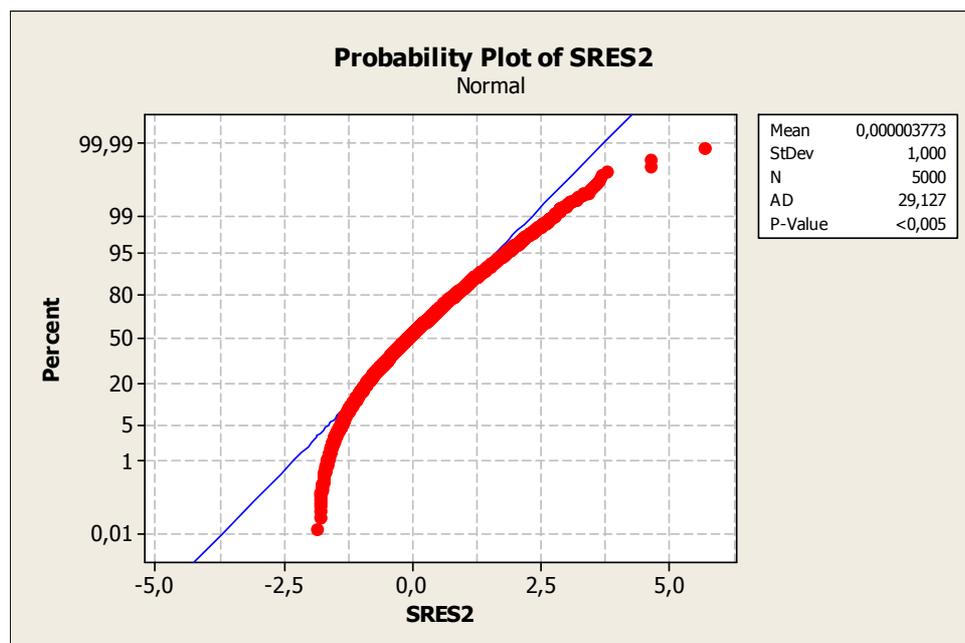


Figura 79. Prueba de normalidad de los residuales del modelo de regresión múltiple para el producto Diovan. Fuente: Minitab.

Con un valor p menor que 0.05, se necesita de una transformación de los datos de salida a una que permita obtener residuales con distribución normal. De todas las posibles transformaciones, la adecuada fue del tipo Box Cox, con un lambda de -2 , es decir una transformación de tipo $1/x^2$. Con los datos transformados a Box Cox se obtiene lo siguiente:

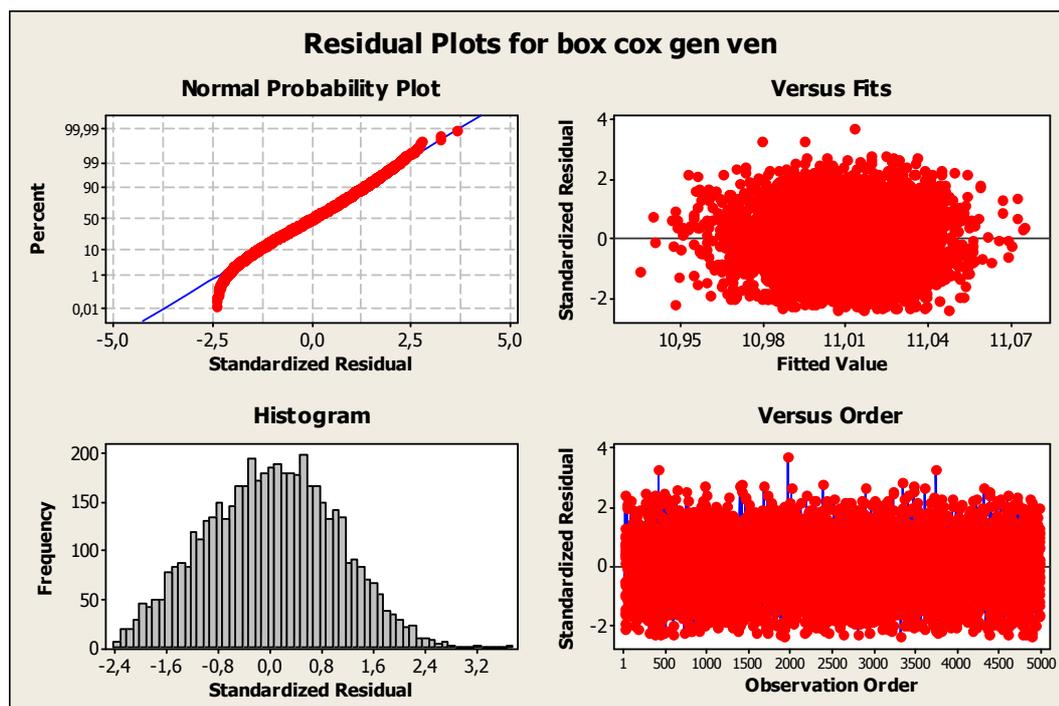


Figura 80. Residuales del modelo de regresión múltiple para el producto Diovan con datos transformados. Fuente: Minitab.

Con la transformación los residuales ahora se puede obtener la regresión y las variables influyentes en el modelo. Se tiene que:

Regression Analysis: box cox gen versus gen ff. gen ap. gen des. gen mm

The regression equation is

$$\text{box cox gen ven} = 285120 - 3,28 \text{ gen ff} - 0,213 \text{ gen ap} - 1,36 \text{ gen des} + 0,0254 \text{ gen mm}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	285120	37708	7,56	0,000	
gen ff	-3,282	2,833	-1,16	0,253	1,024
gen ap	-0,2128	0,7111	-0,30	0,766	1,096
gen des	1,36475	0,08424	16,20	0,000	1,042
gen mm	0,02538	0,04743	0,54	0,595	1,048

S = 82589,8 R-Sq = 86,3% R-Sq(adj) = 85,0%

PRESS = 429715502209 R-Sq(pred) = 79,90%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	1,84466E+12	4,61166E+11	67,61	0,000
Residual Error	4995	2,93306E+11	6821074155		
Total	4999	2,13797E+12			

Source	DF	Seq SS
gen ff	1	2666852968
gen ap	1	50070740931
gen des	1	1,78997E+12
gen mm	1	1952650531

Tal como se dijo antes la gráfica de los residuales se acerca a una distribución normal, por lo que el supuesto 11 queda satisfecho. El modelo cuenta con un valor para la constante β_1 , y por tanto el supuesto 3 también se satisface. Observando el estadístico VIF, se nota que todos los valores son muy cercanos a 1, por lo que la multicolinealidad entre variables es nula (supuesto 5 y 6 quedan satisfechos).

Por las mismas razones que para los otros modelos, los supuestos 1, 7, 8 y 9, pueden presumirse como correctos con los resultados expuestos por Minitab.

Con todos los supuestos analizados, se procede al análisis de varianza para la verificación de factores relevantes en el modelo. Al tener un valor o menor al nivel de significancia establecido como 0.05, se comprueba la existencia de al menos una variable importante en el modelo de regresión. Esta resulta ser la variable la Descuentos, con un alfa igual a 0.

El R^2 -adj obtenido muestra el grado de ajuste del modelo a los datos, que en este caso corresponde a un 85%, lo cual muestra un buen grado de adecuación de los datos y las variables en el modelo.

Más adelante se podrá determinar qué valor es el óptimo, de todas las variables influyentes, prestando especial atención en la regresora significativa obtenida.

ANEXO B-4 PRODUCTO TEGRETOL

Tegretol es un medicamento exclusivamente para tratar la enfermedad de Epilepsia, aunque puede ser recetado para la polineuritis, o tratamientos para el alcoholismo. Pertenece a la línea de productos Phrama dentro de la familia de productos Sistema Nervioso Central, en la línea de Especialidades. Es una medicina que se vende sólo bajo receta médica por lo que dentro de su modelo de regresión se incluye la variable Fuerza de Ventas. Por tanto, las variables estudiadas en su modelo son: FF, A&P, Descuentos y Muestras Médicas. La inferencia en la distribución de cada una se presenta a continuación.

Inferencia para la variable de salida Ventas de Tegretol

Se estudiaron los datos correspondientes a las ventas de Tegretol de los últimos cuatro últimos años, es decir ventas mensuales desde el 2006 hasta el 2009. Con un $\alpha=0.05$, el resultado fue una distribución Normal, mediante una transformación de Johnson, se tiene la siguiente gráfica y resultados:

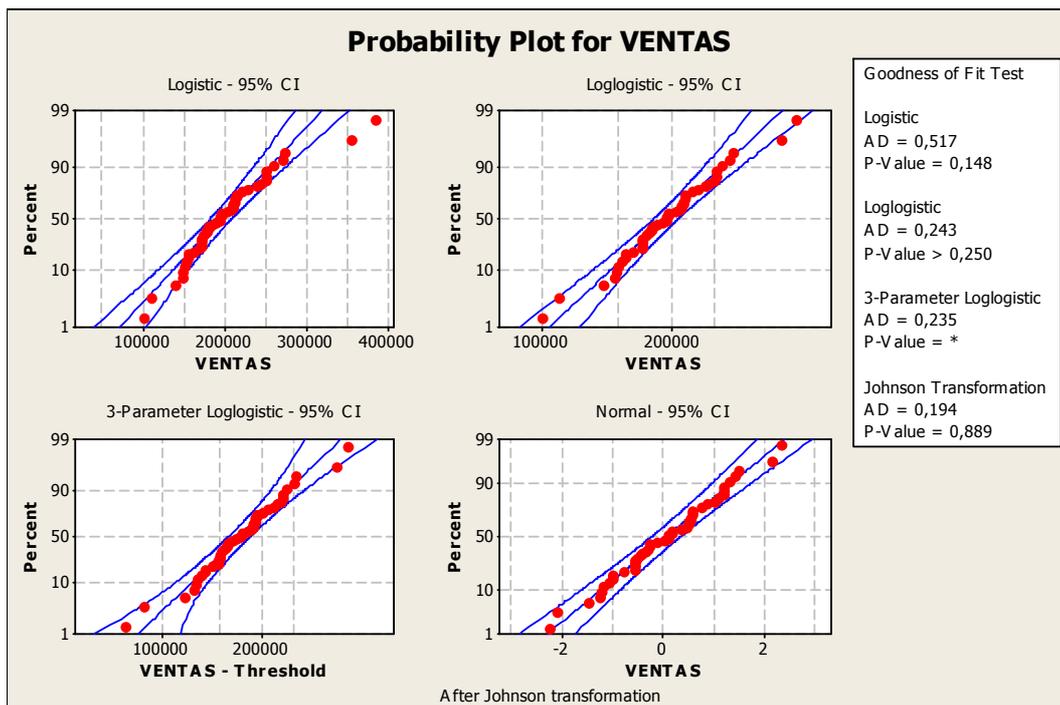


Figura 81. Distribución para los datos de Ventas del producto Tegretol. Fuente: Minitab.

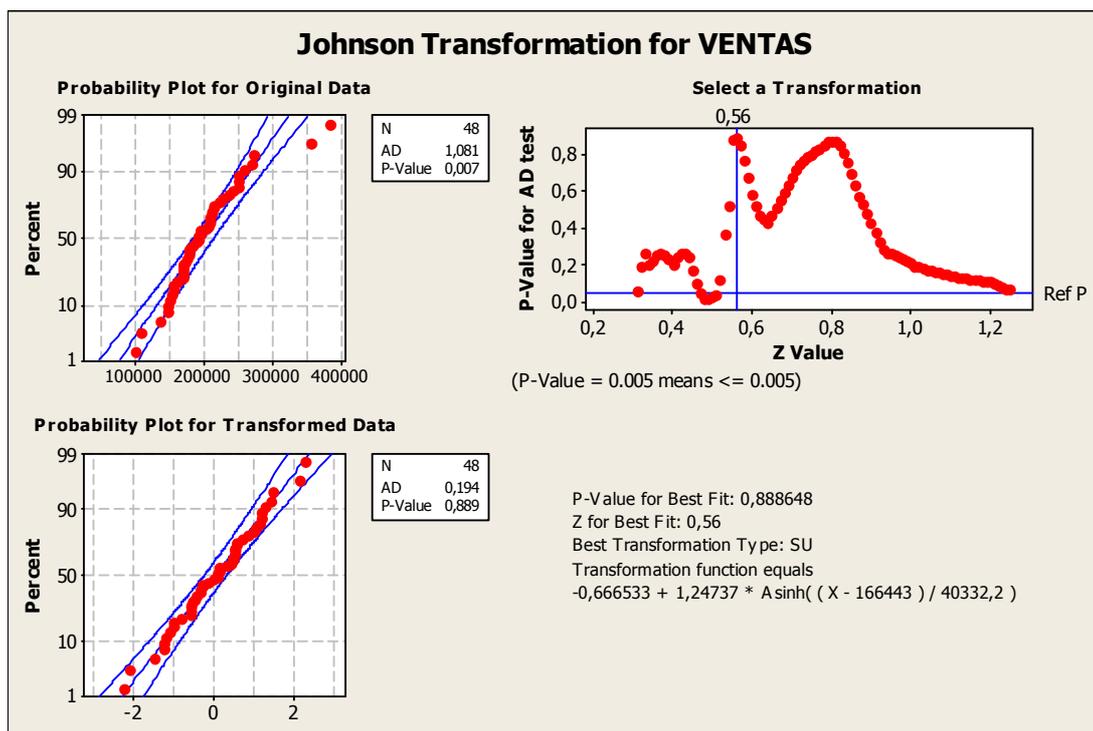


Figura 82. Transformación de Johnson de los datos de Ventas de Tegretol. Fuente: Minitab.

Con un valor p igual a 0,889 no se rechaza la hipótesis nula y por tanto se afirma que los datos pertenecen a una distribución normal, luego de la transformada de Johnson.

Mediante una simulación Montecarlo se generaron 5000 datos con los parámetros específicos y su inferencia se comprueba a continuación.

Inferencia para los datos generados de la variable de salida Ventas de Tegretol

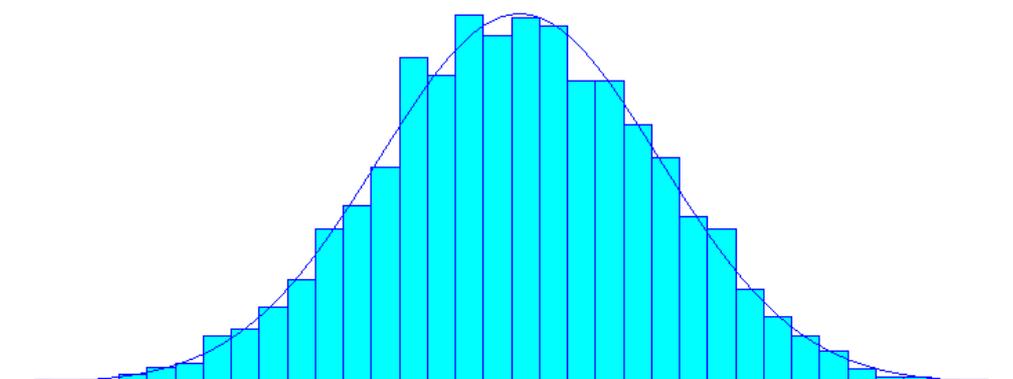


Figura 83. Distribución para los datos generados de Ventas del producto Tegretol. Fuente: Arena

Distribution Summary

Distribution: Normal
Expression: NORM(0.0527, 1)
 Square Error: 0.000258

Chi Square Test

Number of intervals = 29
 Degrees of freedom = 26
 Test Statistic = 31
Corresponding p-value = 0.235

Kolmogorov-Smirnov Test

Test Statistic = 0.00914
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -3.45
 Max Data Value = 3.74
 Sample Mean = 0.0527
 Sample Std Dev = 1

Histogram Summary

Histogram Range = -4 to 4
Number of Intervals = 40

Los dos valores p obtenidos para ambas pruebas de bondad de ajuste son mayores que alfa (0,05) y por tanto se comprueba la normalidad de los datos generados.

Inferencia para la variable de entrada Field Force de Tegretol

Al ser Tegretol un producto que se vende bajo receta médica necesita la ayuda de la fuerza de ventas para su promoción. En este caso se infirió en los gastos correspondientes a esta variable y se obtuvo que éstos siguen una distribución Normal luego de la transformación de Johnson. Mediante la siguiente ecuación:

$$-6,44670 + 0,989899 * \ln(X - 227,704)$$

Las gráficas generadas son:

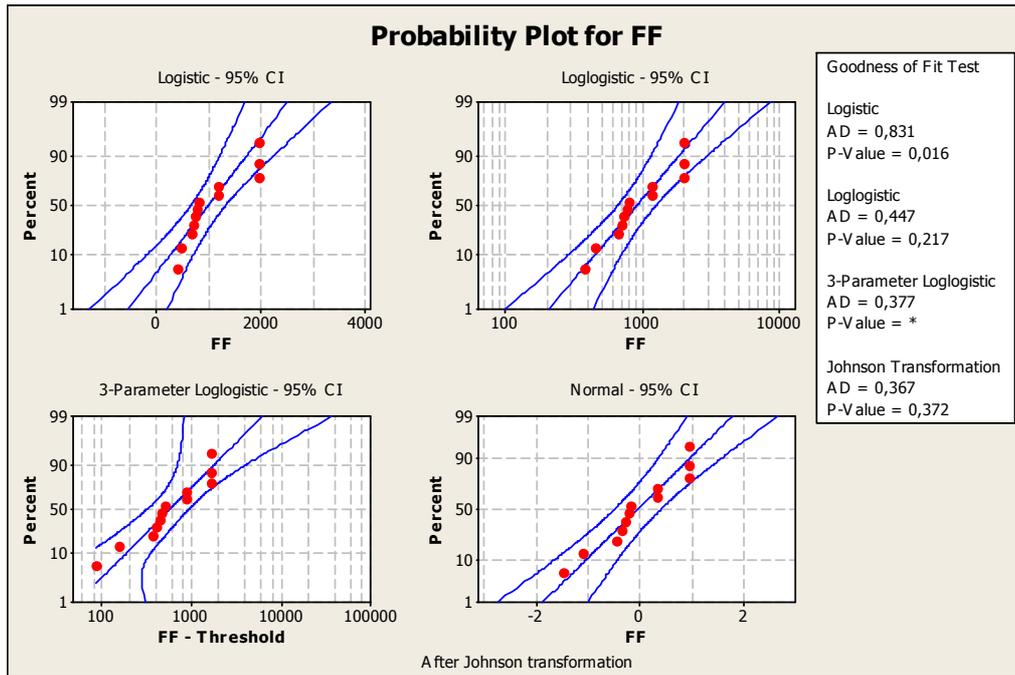


Figura 84. Distribución para los datos de FF del producto Tegretol. Fuente: Minitab.

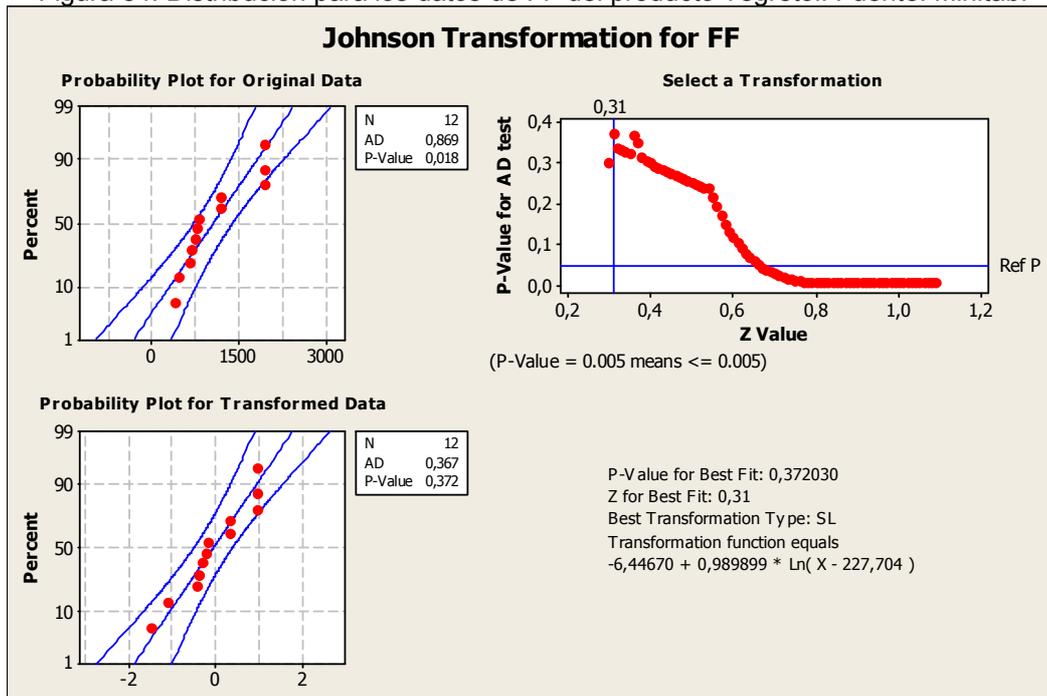


Figura 85. Transformación de Johnson de los datos de FF de Tegretol. Fuente: Minitab.

El valor p más grande de todas las distribuciones posibles a ser ajustas, pertenece a una distribución Normal, con un valor $p = 0.372$. Éste no permite rechazar la hipótesis nula y por tanto se concluye que de todas las distribuciones de probabilidad posibles, la más adecuada es la Normal.

Con la generación mediante simulación Montecarlo, se puede inferir en los datos para ratificar la presencia de una distribución normal en ellos. A continuación los resultados.

Inferencia para los datos generados de la variable de entrada Field Force de Tegretol

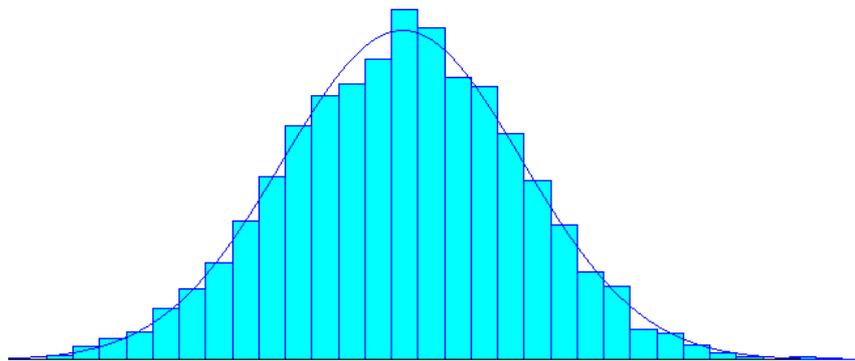


Figura 86. Distribución para los datos generados de FF del producto Tegretol. Fuente: Arena

Distribution Summary

Distribution: Normal
Expression: NORM(-0.0792, 0.788)
 Square Error: 0.000187

Chi Square Test

Number of intervals = 27
 Degrees of freedom = 24
 Test Statistic = 23.9
Corresponding p-value = 0.473

Kolmogorov-Smirnov Test

Test Statistic = 0.0116
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -3.72
 Max Data Value = 2.98
 Sample Mean = -0.0792
 Sample Std Dev = 0.789

Histogram Summary

Histogram Range = -4 to 3
 Number of Intervals = 40

Tanto los valores p de las pruebas de bondad de ajuste como los parámetros de la distribución normal generadas muestran que los datos sí siguen una distribución normal.

Inferencia en la variable de entrada **Advertisement and Promoting** en el producto **Tegretol**

Con datos desde el 2006 hasta el 2009, se estudió la distribución de la variable A&P para el producto Tegretol. Igualmente con un alfa de 0.05 se obtuvo que los datos se ajusten a una distribución Normal a partir nuevamente de una distribución de Johnson, con la siguiente fórmula:

$$1,57886 + 0,727811 * \text{LN} ((X + 58,5504) / (12309,8 - X))$$

La gráfica de adecuación de la distribución se muestra a continuación:

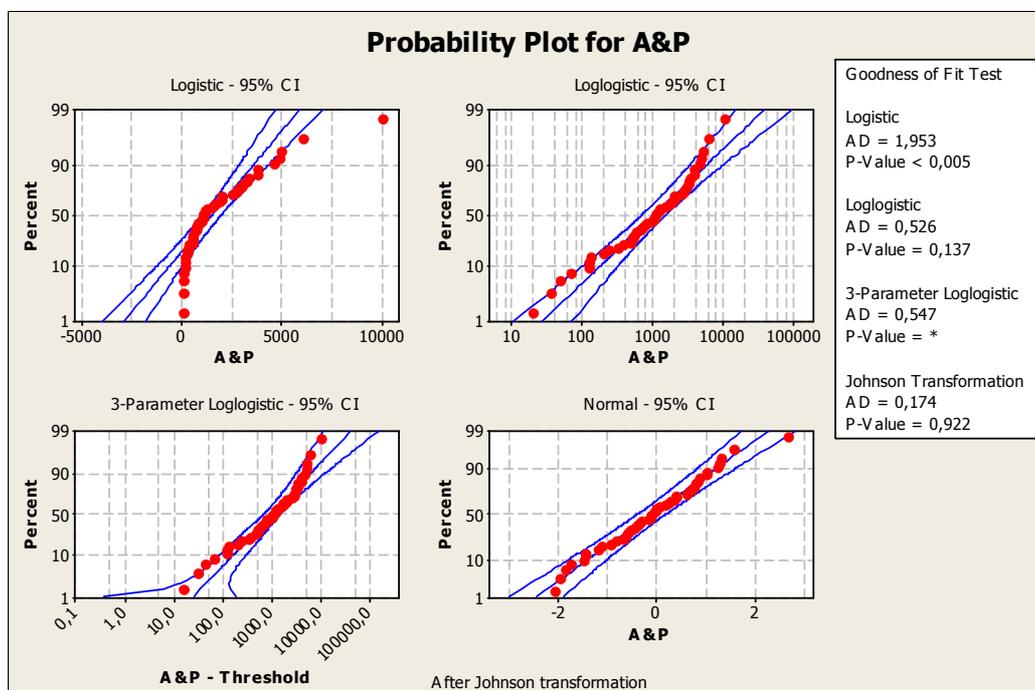


Figura 87. Distribución para los datos de A&P del producto Tegretol. Fuente: Minitab.

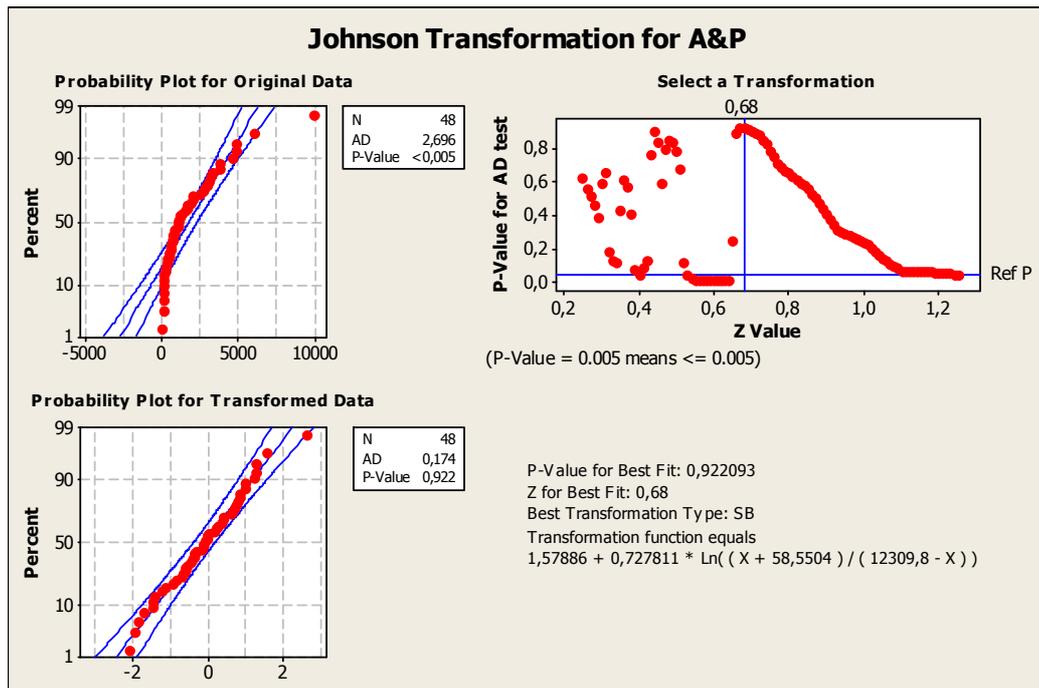


Figura 88. Transformación de Johnson de los datos de A&P de Tegretol. Fuente: Minitab.

El valor p obtenido igual a 0.992 asegura que la distribución de los datos es normal. Esto también se lo puede ver en la grafica de normalidad en la cual los datos se acercan a la línea de manera muy próxima.

Los parámetros obtenidos, junto con la simulación Montecarlo serán usados para la generación de 5000 datos de la variable de entrada A&P. La inferencia a éstos para comprobar el ajuste a la distribución Normal, se muestra a continuación.

Inferencia a los datos generados de la variable Advertisement and Promoting del producto Tegretol.

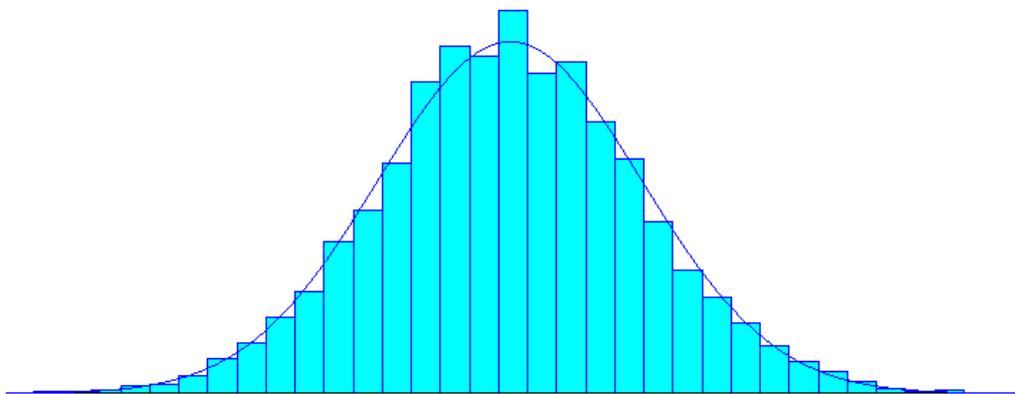


Figura 89. Distribución para los datos generados de A&P del producto Tegretol. Fuente: Arena.

Distribution Summary

Distribution: Normal
Expression: **NORM(-0.111, 1)**
 Square Error: 0.000253

Chi Square Test

Number of intervals = 27
 Degrees of freedom = 24
 Test Statistic = 24.7
 Corresponding p-value = 0.432

Kolmogorov-Smirnov Test

Test Statistic = 0.0118
 Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -4.03
 Max Data Value = 3.83
 Sample Mean = -0.111
 Sample Std Dev = 1

Histogram Summary

Histogram Range = -4.83 to 4
 Number of Intervals = 40

Un error cuadrado bajo obtenido, junto con los valores p de cada prueba de ajuste, Kolgomorov y Chi Cuadrado prueban que los datos generados siguen una distribución normal con los parámetros del ajuste de los datos primeramente inferidos.

Inferencia en la variable de entrada Descuentos del producto Tegretol

Con datos desde el 2006 al 2009 se infirió en la distribución adecuada, obteniendo una distribución Normal por medio de una transformación de Johnson.

Con un alfa de 0.05 se obtuvo los siguientes resultados:

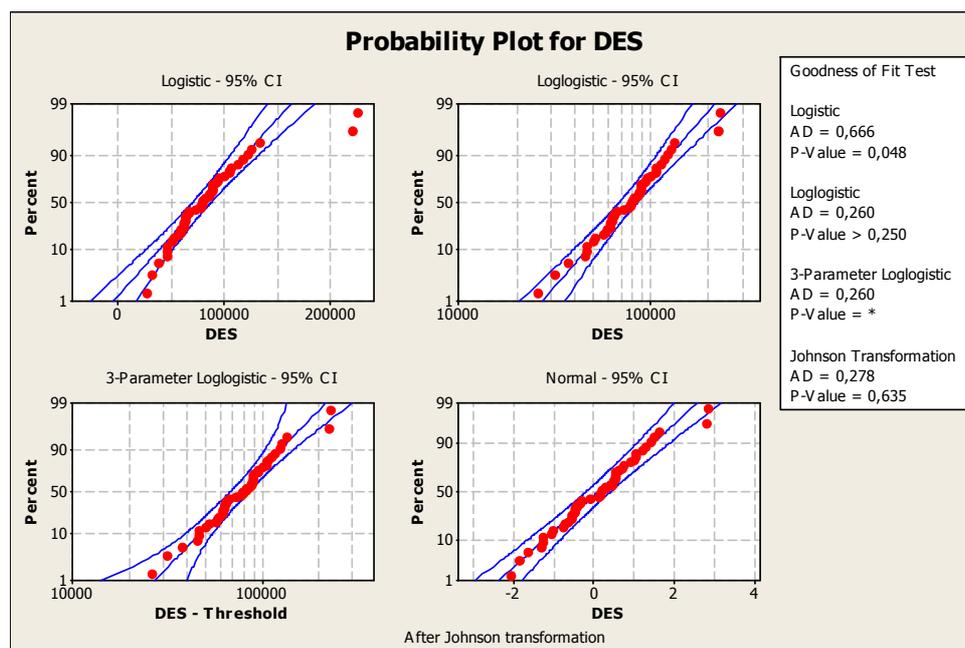


Figura 90. Distribución para los datos de Descuentos del producto Tegretol. Fuente: Minitab.

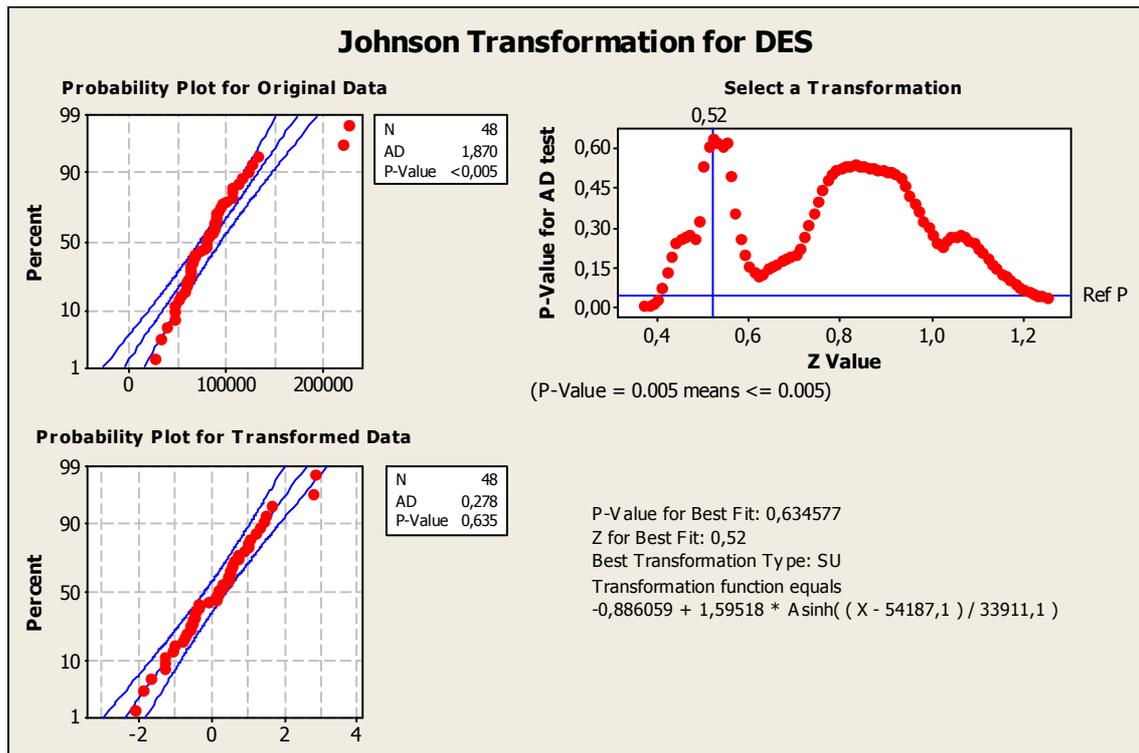


Figura 91. Transformación de Johnson de los datos de Descuentos de Tegretol. Fuente: Minitab

La ecuación utilizada para la transformación fue:

$$-0,886059 + 1,59518 * \text{Asinh}((X - 54187,1) / 33911,1)$$

Con un valor p igual mayor a 0.635 se corrobora que los datos se ajustan a una distribución normal. Junto con los parámetros correspondientes a esta distribución y una simulación Montecarlo, se generaron 5000 datos. La inferencia en ellos para comprobar que se ajustan a una distribución exponencial se muestra a continuación.

Inferencia a los datos generados de la variable Descuentos del producto Tegretol

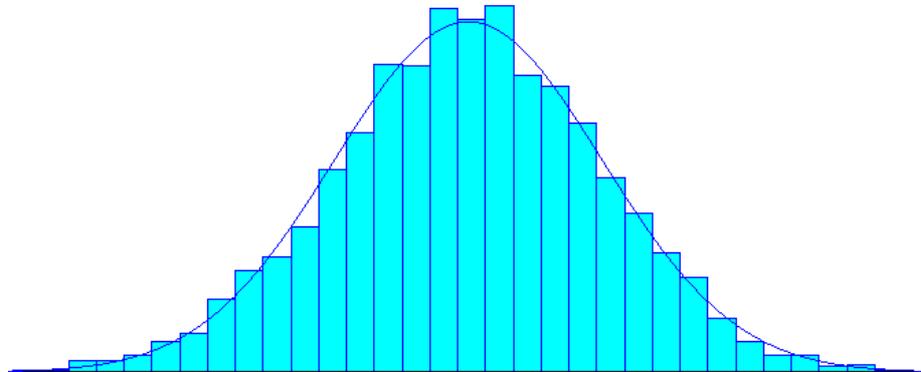


Figura 92. Distribución para los datos generados de Descuentos del producto Tegretol. Fuente: Arena.

Distribution Summary

Distribution: Normal
Expression: NORM(0.115, 1.08)
 Square Error: 0.000193

Chi Square Test

Number of intervals = 29
 Degrees of freedom = 26
 Test Statistic = 26.8
Corresponding p-value = 0.43

Kolmogorov-Smirnov Test

Test Statistic = 0.0114
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -3.49
 Max Data Value = 4.19
 Sample Mean = 0.115
 Sample Std Dev = 1.08

Histogram Summary

Histogram Range = -4 to 4.96
 Number of Intervals = 40

Los valores p de ambas pruebas de ajuste hacen que la hipótesis nula no se rechace y que por tanto los datos sí se ajusten a una distribución normal.

Inferencia en la variable de entrada Muestras Médicas del producto Tegretol

Con datos desde el 2006 al 2009, se infirió en los datos correspondientes a los gastos por Muestras Médicas. Con un alfa de 0.05, y un juego de hipótesis igual al de las anteriores variables, se tiene que los datos se ajustan de mejor manera a una distribución Exponencial de dos parámetros. Se aprecia en la siguiente figura:

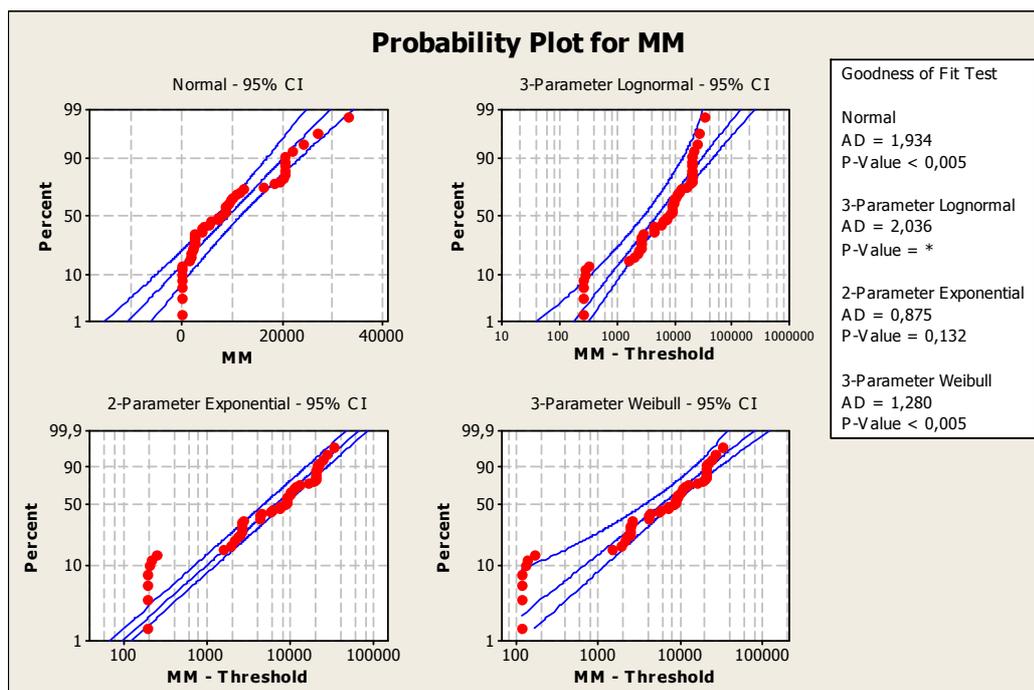


Figura 93. Distribución para los datos de MM del producto Tegretol. Fuente: Minitab.

Con un valor p igual a 0.132 no se puede rechazar la hipótesis nula y por tanto se dice que los datos correspondientes a Muestras Médicas pertenecen a una distribución exponencial de dos parámetros.

Con la simulación Montecarlo se generaron 5000 datos en base a los parámetros obtenidos. Para comprobar su distribución, la inferencia en ellos se presenta a continuación:

Inferencia a los datos generados de la variable Muestras Médicas del producto Tegretol

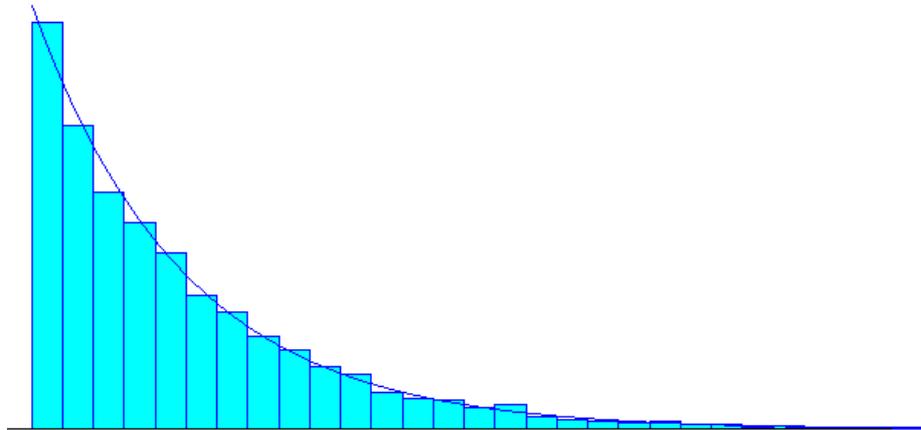


Figura 94. Distribución para los datos generados de MM del producto Tegretol. Fuente: Minitab.

Distribution Summary

Distribution: Exponential
Expression: $-199 + \text{EXPO}(9.54e+003)$
 Square Error: 0.000272

Chi Square Test

Number of intervals = 26
 Degrees of freedom = 24
 Test Statistic = 25.9
Corresponding p-value = 0.368

Kolmogorov-Smirnov Test

Test Statistic = 0.0121
Corresponding p-value > 0.15

Data Summary

Number of Data Points = 5000
 Min Data Value = -198
 Max Data Value = $7.71e+004$
 Sample Mean = $9.34e+003$
 Sample Std Dev = $9.52e+003$

Histogram Summary

Histogram Range = -199 to $7.71e+004$
 Number of Intervals = 40

Ambos estadísticos mayores a 0.05 prueban que los datos generados pertenecen a una distribución exponencial.

Modelo de regresión con los datos generados para el producto Tegretol

En base a las distribuciones inferidas de cada variable para el producto Tegretol, y la generación Montecarlo de 5000 datos que se ajusten a dichas distribuciones, se procedió a correr un modelo de regresión múltiple. Por medio de éste se determinará cuál es el impacto (si es que existe) de las variables de entrada, sobre la variable de salida, Ventas.

Tal como se dijo antes, se prueba que:

$$H_0 : \sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$$

$$H_0 : \sigma_1 \neq \sigma_2 \neq \sigma_3 \neq \sigma_4$$

Con esto se determinará si existe o no una variable de entrada que afecte de manera significativa en el modelo de regresión obtenido. Para conocer cuál de las variables es la que causa una influencia significativa en la variable de respuesta, se prueba lo siguiente:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$$

$$H_0 : \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4$$

Con un valor p menor al valor alfa establecido, se rechaza la hipótesis nula que establece,

H_0 = El factor no influye de manera significativa en el modelo.

H_1 = El factor sí influye de manera significativa en el modelo.

Los resultados de la regresión múltiple con las variables pertenecientes al producto Tegretol son los siguientes:

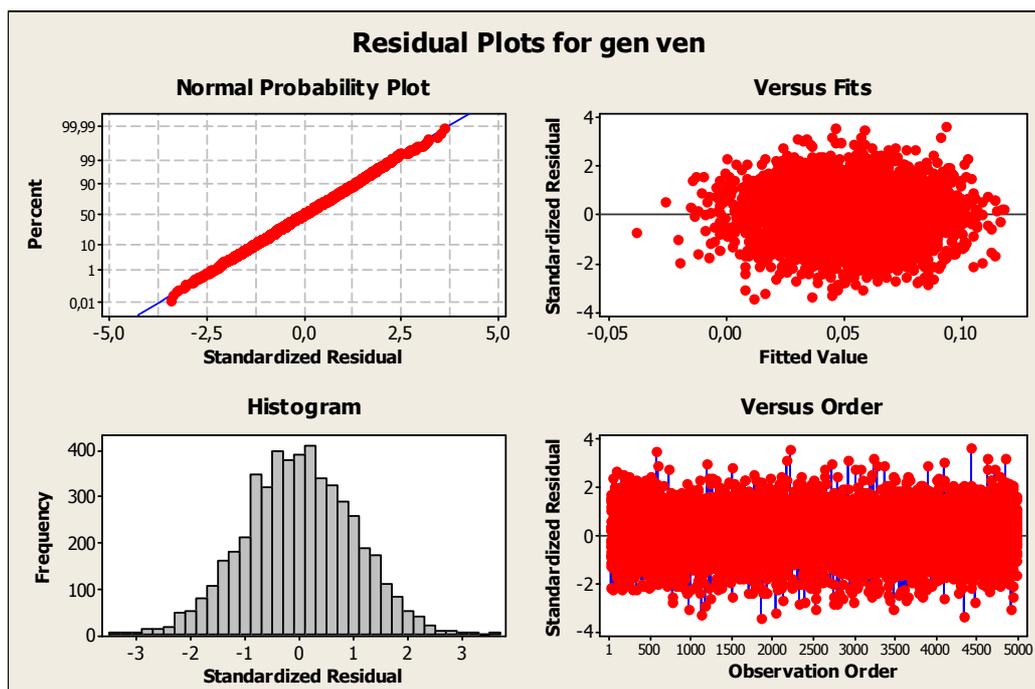


Figura 95. Residuales del modelo de regresión múltiple para el producto Tegretol con datos transformados. Fuente: Minitab.

Regression Analysis: gen ven versus gen ff. gen ap. gen des. gen mm

The regression equation is

$$\text{gen ven} = -1,65 - 0,000008 \text{ gen ff} + 0,000027 \text{ gen ap} + 0,000021 \text{ gen des} - 0,000006 \text{ gen mm}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	-1,6529	0,3259	-5,07	0,000	
gen ff	-0,0000078	0,0001549	-0,05	0,960	1,000
gen ap	0,00002687	0,00005589	0,48	0,633	1,001
gen des	0,00002083	0,00000274	7,59	0,000	1,000
gen mm	-0,00000572	0,00001397	-0,41	0,684	1,001

S = 0,628317 R-Sq = 64,7% R-Sq(adj) = 61,4%

PRESS = 21,6380 R-Sq(pred) = 54,97%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	31,0779	7,7695	19,68	0,000
Residual Error	4995	16,9756	0,3948		
Total	4999	48,0535			

Source	DF	Seq SS
gen ff	1	0,8801
gen ap	1	0,4487
gen des	1	29,6828
gen mm	1	0,363

En donde:

John gen ven: Representa a los datos generados de Ventas y transformados mediante el algoritmo de Johnson.

Gen ap: representa la generación de datos de la variable A&P.

Gen des: representa la generación de datos de la variable Descuentos.

Gen mm: representa la generación de datos de la variable Muestras Médicas.

En la gráfica de los residuales se obtiene un valor p mayor al nivel de significancia por lo que se puede aceptar que los residuales pertenecen a una distribución normal. Los residuos se ajustan perfectamente a la recta y de igual manera y mediante el histograma, el supuesto 11 se satisface.

Con el estadístico VIF, el cual es igual a 1 para todas las variables, se demuestra que no hay correlación o multicolinealidad entre ellas y el supuesto 5 y por lo tanto el 6 quedan satisfechos.

De igual manera, y por lo antes expuesto para los demás productos, el supuesto 1, 3, 7, 8 y 9 pueden presumirse como correctos con los resultados expuestos por Minitab. Una vez que los supuestos del Método de Mínimos cuadrados se cumplen, se debe analizar los resultados obtenidos de la regresión.

Analizando los resultados del análisis de varianza, éstos indican que sí existe una variable de importante influencia en el modelo, por el valor p obtenido, igual a cero. Si se observa la ecuación de la regresión obtenida, ésta muestra un coeficiente más grande para la variable Descuentos, y el más pequeño para la variable Fuerza de Ventas. Analizando la tabla de valores obtenidos para cada coeficiente, se comprueba que para la variable Descuentos el valor p resultante es cero, que al ser menor que 0.05 (nivel de significancia) permite rechazar la hipótesis nula y que esta variable es la única representativa para la ecuación de

regresión de Tegretol. Además mediante el signo obtenido (positivo) se determina que la relación entre Ventas y Descuentos es directamente proporcional.

R^2 -adj muestra el grado de ajuste del modelo a los datos, que en este caso corresponde a un 61.4%, lo cual muestra un buen ajuste de los datos a la regresión.

Ahora bien, siendo la variable Descuentos la de mayor importancia en el producto Tegretol y si bien las demás variables, Advertisement & Promoting y Muestras Médicas, también son importantes para las ventas de este medicamento, se prestará especial atención al momento de encontrar el óptimo para esta variable.

ANEXO C- OPTIMIZACIÓN LINEAL PARA CADA PRODUCTO

ANEXO C-1 Modelo de optimización para el producto Genteal con Solver (sin simulación Montecarlo)

Al igual que para Voltaren, con los resultados generados a partir de obtención del modelo de regresión múltiple, se procedió a llenar el formato de resolución de la optimización para el caso de Genteal. Bajo las mismas condiciones de no negatividad y niveles máximos como límites de las variables, se ajustó un modelo lineal. Se consiguió las siguientes respuestas:

Gastos en FF	\$ 0
Gastos en A&P	\$ 82447,57
Gastos en Descuentos	\$ 559855,66
Gastos en Muestras Médicas	\$ 0

Generación propia.

Tabla 5. Niveles óptimos para cada variable del producto Genteal.

Esto genera Ventas de Genteal de \$ 1078172,38

Modelo de optimización con MC-Sim Solver (con simulación Montecarlo) para Genteal

De igual manera se siguieron los cinco pasos que se hicieron para el producto Genteal. Con el modelo matemático se definió la función objetivo, se generaron 1000 números aleatorios pertenecientes a la distribución inferida de cada variable aleatoria. Para el caso de Genteal se contaban con 4 variables regresoras, con las siguientes distribuciones de probabilidad:

Variable	FF	AP	DESCUENTOS	MM
Distribución	NORM(-0.052, 0.776)	NORM(-0.0661, 1.04)	NORM(-0.0394, 0.954)	NORM(-0.0282, 0.909)

Generación propia.

Tabla 6. Distribuciones para las variables del producto Genteal.

A partir de estas distribuciones se crearon 1000 datos aleatorios correspondientes a cada variable. A continuación se reemplazó los valores de cada variable de entrada en la función de rentabilidad obtenida de la regresión. A continuación el análisis de los datos obtenidos:

- **FF con Ventas:**

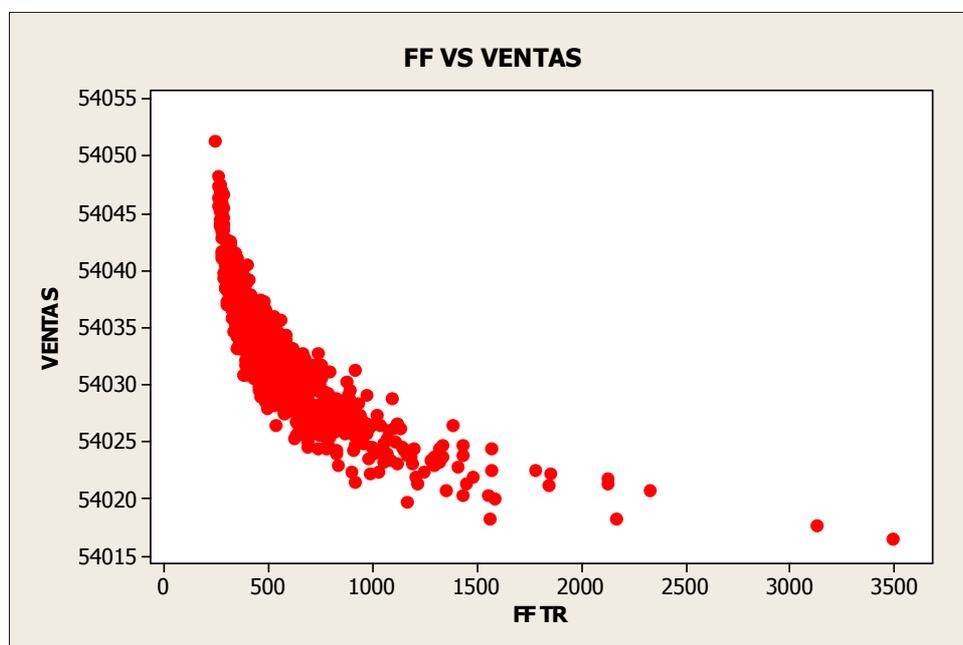


Figura 101. Gráfica de correlación entre la variable FF y Ventas de Genteal. Fuente: Minitab.

Se puede notar que si hay una relación entre la FF y Ventas, de manera exponencial. Esto significa que la inversión en FF grande no genera necesariamente mayores ventas del producto. A pesar de que esta variable no resultó significativa en el modelo, fue la segunda con un valor p más cercano a cero y por ello se puede ver que se guarda cierta relación. Será importante considerar gastos excesivos en la Fuerza de Ventas ya que no estarían aportando a las Ventas de Genteal de manera en la que se espera.

- **A&P con Ventas:**

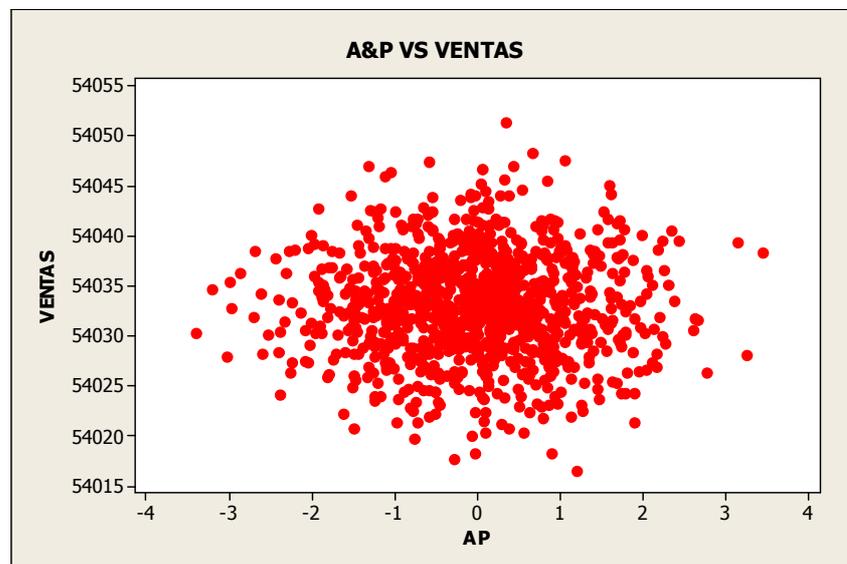


Figura 102. Gráfica de correlación entre la variable A&P y Ventas de Voltaren. Fuente: Minitab.

La gráfica de correlación indica que no existe relación entre los valores bajos de A&P y las ventas de Voltaren. Esto puede resultar importante ya que indica que un gasto excesivo en Medios de Difusión para este producto no tiene efectos favorables a las ventas, información que se complementa con los resultados obtenidos de la regresión múltiple en donde A&P no resultó ser una variable significativa.

- **Descuentos vs. Ventas:**

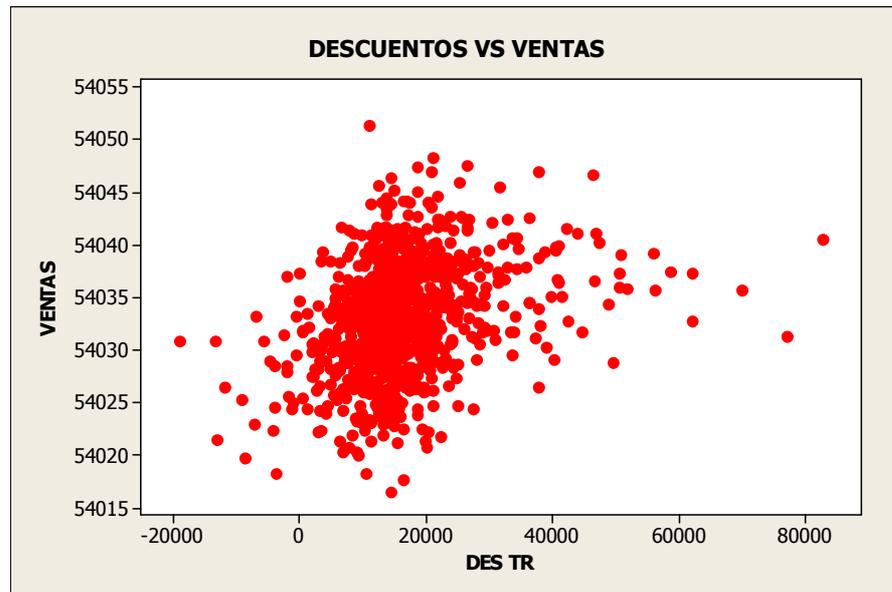


Figura 103. Gráfica de correlación entre la variable Descuentos y Ventas de Genteal. Fuente: Minitab.

Se ve que existe una relación entre estas dos variables pues los puntos se concentran alrededor de un mismo punto. Tal como se obtuvo en la regresión múltiple, la variable Descuentos mostró ser significativa y esto se comprueba con esta gráfica de correlación.

- **Muestras Médicas vs Ventas:**

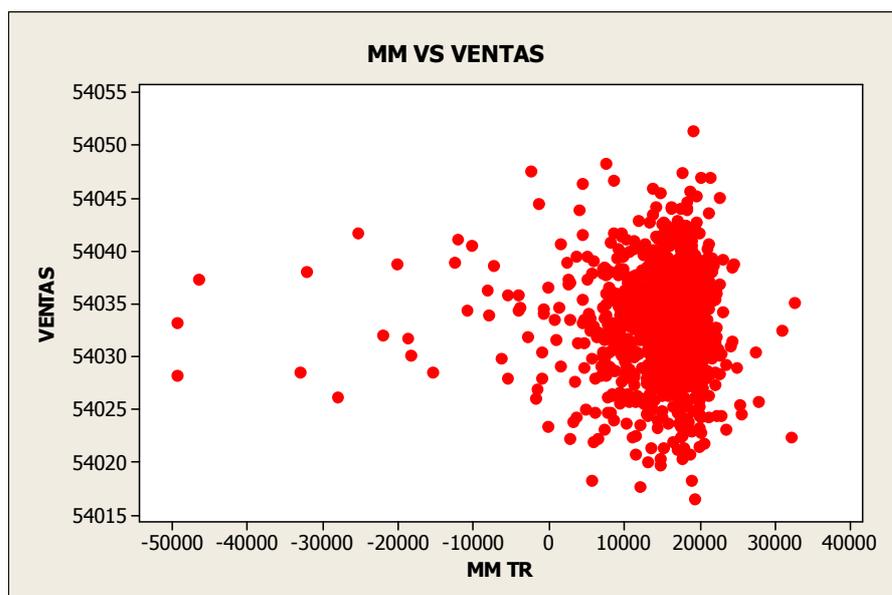


Figura 104. Gráfica de correlación entre la variable MM y Ventas de Genteal . Fuente: Minitab.

Se observa un comportamiento de la variable MM con Ventas en la cual existe una agrupación de los datos alrededor del nivel \$20000 en el eje de Muestras Médicas. Sin embargo se observa que bajo ese nivel se pueden obtener tanto Ventas bajas como altas de Genteal, por lo que no guardan mucha relación entre una variable y la otra.

- **Gráfico resumen de Ventas de Genteal:**

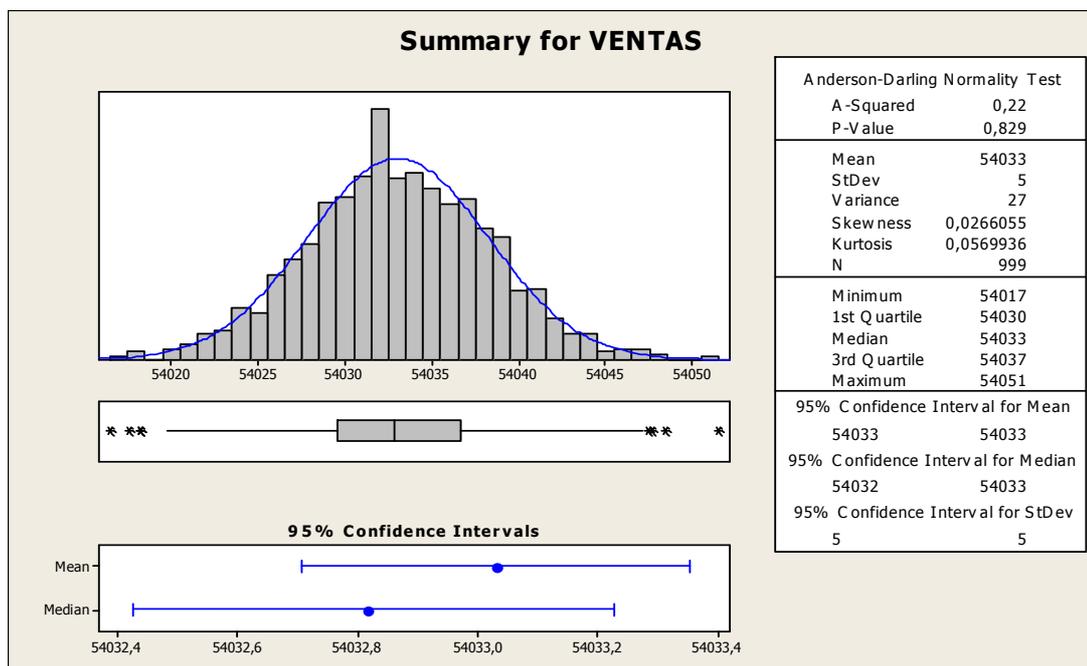


Figura 105. Gráfico resumen de Ventas de Genteal. Fuente: Minitab.

Se ve que los datos se ajustan a una distribución Normal tal como se infirió inicialmente, además que su valor p es mayor al nivel de significancia establecido. Un dato importante de este cuadro resumen son la media, y valores mínimos y máximos determinados a partir de la simulación Montecarlo.

\$54051, con los siguientes gastos en cada una de sus variables:

Variable	Gasto máximo en el que se debe incurrir
FF	241,253
A&P	\$ 4852
Descuentos	\$ 10949,9
Muestras Médicas	\$ 19068,8
VENTAS GENERADAS	\$ 54051

Generación propia.

Tabla 7. Valores máximos de gastos por variable para el producto Voltaren

Comparando ambos modelos de optimización y de acuerdo a los resultados obtenidos de la regresión múltiple en el que la variable Descuentos se observa que ambos modelos asignan un valor bajo a la Fuerza de Ventas, por lo que se comprueba que su aporte no es muy significativo para las Ventas del producto. Mientras que para la variable Descuentos en ambos se le da el valor ponderado de acuerdo a los resultados obtenidos de la regresión. Es decir, para este tipo de producto ambos modelos se ajustan de manera adecuada, quizá con mayor detalle el de la simulación Montecarlo.

ANEXO C-2 Modelo de optimización para el producto Cataflam con Solver (sin simulación Montecarlo)

Con los resultados generados a partir de obtención del modelo de regresión múltiple, se procedió a llenar el formato de resolución de la optimización para el caso de Cataflam. Bajo las mismas condiciones se ajustó un modelo lineal con no negatividad y se consiguió las siguientes respuestas:

Gastos en A&P	\$ 0
Gastos en Descuentos	\$ 1160857,43
Gastos en Muestras Médicas	\$ 383682,18

Generación propia.

Tabla 8. Niveles óptimos para cada variable del producto Cataflam.

Esto genera Ventas de Cataflam de \$ 1762971,72

Modelo de optimización con MC-Sim Solver (con simulación Montecarlo) para Cataflam

Para el caso de Cataflam se contaban con 3 variables regresoras, con las siguientes distribuciones de probabilidad:

Variable	AP	DESCUENTOS	MM
Distribución	NORM(0.011, 1.02)	NORM(-0.00295, 0.98)	-0.001 + EXPO(2.63e+004)

Generación propia.

Tabla 9. Distribuciones para las variables del producto Cataflam.

A partir de estas distribuciones se crearon 1000 datos aleatorios correspondientes a cada variable. A continuación se reemplazó los valores de cada variable de entrada en la función de rentabilidad obtenida de la regresión. A continuación el análisis de los datos obtenidos:

- **A&P con Ventas:**

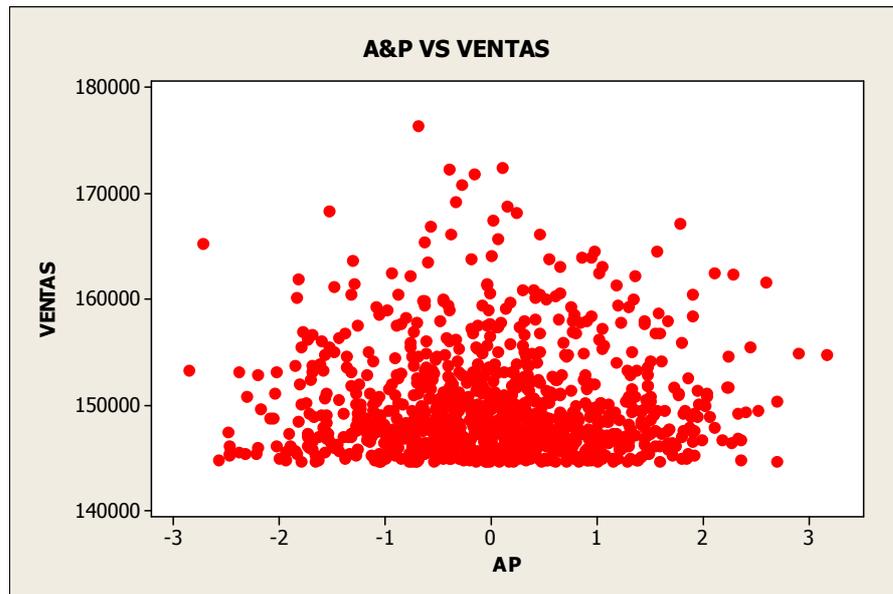


Figura 106. Gráfica de correlación entre la variable A&P y Ventas de Cataflam. Fuente: Minitab.

La gráfica de correlación indica que esta variable tampoco guarda relación entre los valores bajos de A&P y las ventas de Cataflam, pues no se identifica un patrón de comportamiento notorio.

- **Descuentos vs. Ventas:**

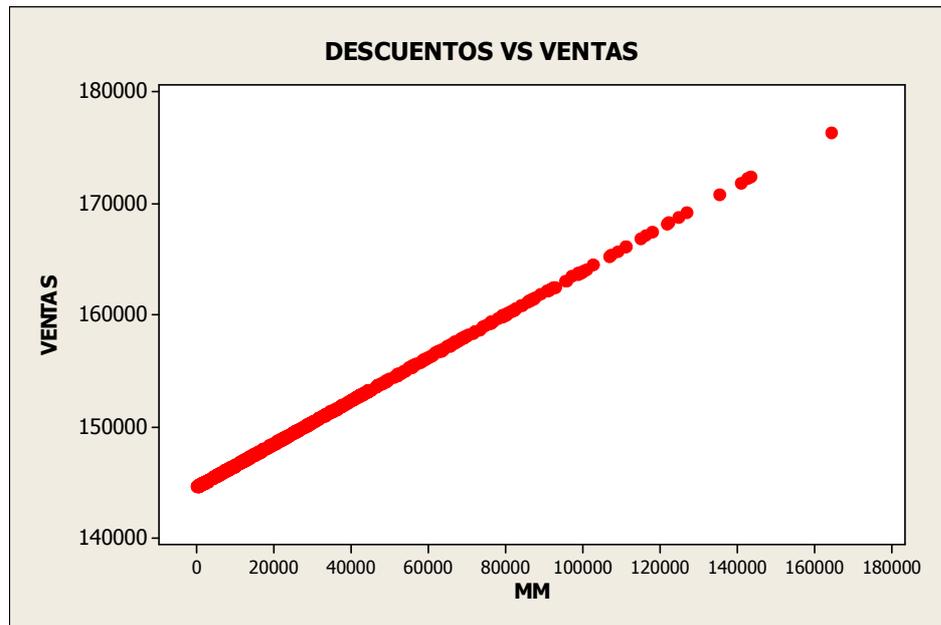


Figura 107. Gráfica de correlación entre la variable Descuentos y Ventas de Cataflam. Fuente: Minitab.

Se ve que existe una relación entre estas dos variables pues los puntos siguen el mismo patrón de comportamiento. Tal como se obtuvo en la regresión múltiple, la variable Descuentos mostró ser significativa y esto se comprueba con esta gráfica de correlación.

- **Muestras Médicas vs Ventas:**

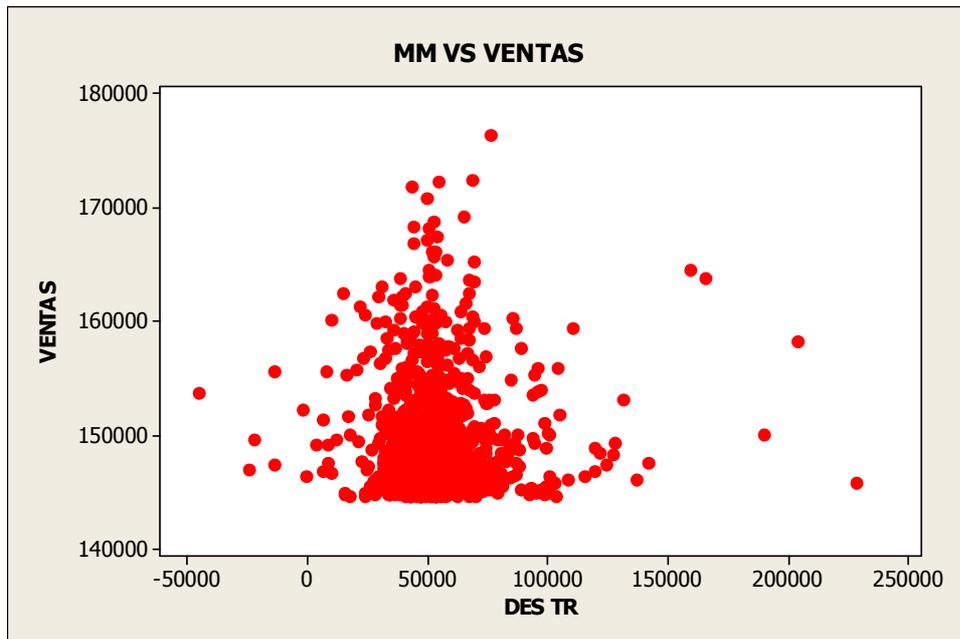


Figura 108. Gráfica de correlación entre la variable MM y Ventas de Cataflam. Fuente: Minitab.

Se observa un comportamiento de la variable MM con Ventas en la cual existe una agrupación de los datos alrededor del nivel \$50000 en el eje de Muestras Médicas.

observa que ambos modelos asignan un valor bajo a la variable Medios de difusión, por lo que se comprueba que su aporte no es muy significativo para las Ventas del producto. Mientras que para la variable Descuentos en ambos se le da el valor ponderado de acuerdo a los resultados obtenidos de la regresión.

ANEXO C-3 Modelo de optimización para el producto Diovan con Solver (sin simulación Montecarlo)

Resolviendo el Solver con la función objetivo planteada y las restricciones establecidas se obtuvo que para Diovan, los valores óptimos para las variables: FF, A&P, Descuentos y Muestras Médicas son:

Gastos en FF	\$0
Gastos en A&P	\$ 0
Gastos en Descuentos	\$ 3943694,64
gastos en Muestras Médicas	\$ 3111995,12

Generación propia.

Tabla 11. Niveles óptimos para cada variable del producto Diovan.

Esto genera Ventas de Diovan de \$ 5727589,39

Modelo de optimización con MC-Sim Solver (con simulación Montecarlo) para Diovan

De igual manera se siguieron los cinco pasos que se hicieron para el producto Voltaren. Con el modelo matemático se definió la función objetivo, se generaron 1000 números aleatorios pertenecientes a la distribución inferida de cada variable aleatoria. Par el caso de Diovan se contaba con 4 variables regresoras, con las siguientes distribuciones de probabilidad:

Variable	FF	AP	DESCUENTOS	MM
Distribución	NORM(-0.158, 1.2)	NORM(0.139, 1.1)	NORM(0.0291, 0.97)	-0.001 + EXPO(2.37e+005)

Generación propia.

Tabla 12. Distribuciones para las variables del producto Diovan.

A partir de estas distribuciones se crearon 1000 datos aleatorios correspondientes a cada variable. A continuación se reemplazó los valores de

cada variable de entrada en la función de rentabilidad obtenida de la regresión. A continuación el análisis de los datos obtenidos:

- **FF con Ventas:**

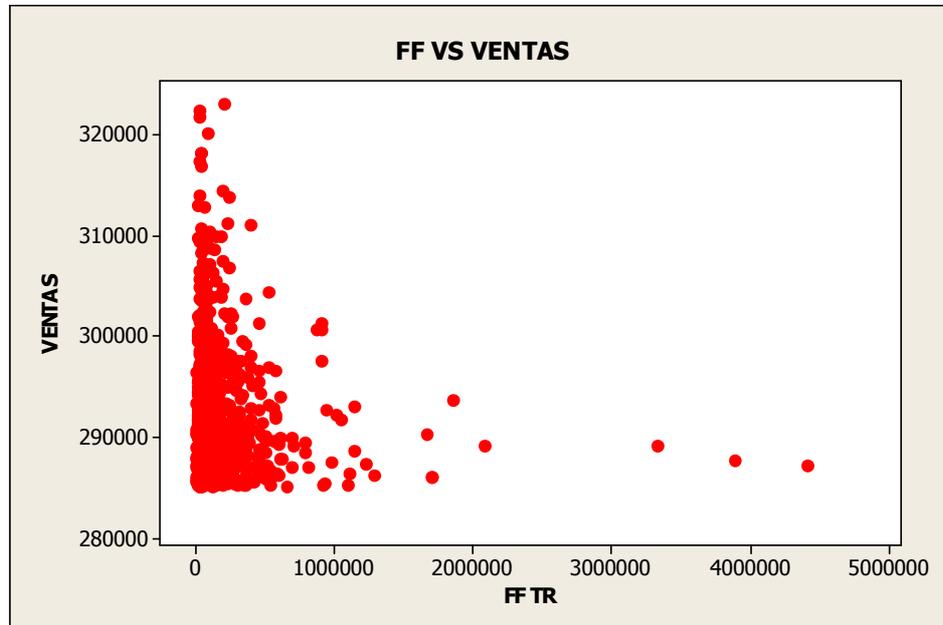


Figura 110. Gráfica de correlación entre la variable FF y Ventas de Diovan. Fuente: Minitab.

Se puede notar que si hay una relación entre la FF y Ventas, de manera exponencial. Esto significa que la inversión en FF grande no genera necesariamente mayores ventas del producto. A pesar de que esta variable no resultó significativa en el modelo, fue la segunda con un valor p más cercano a cero y por ello se puede ver que se guarda cierta relación. Será importante considerar gastos excesivos en la Fuerza de Ventas ya que no estarían aportando a las Ventas de Genteal de manera en la que se espera.

- **A&P con Ventas:**

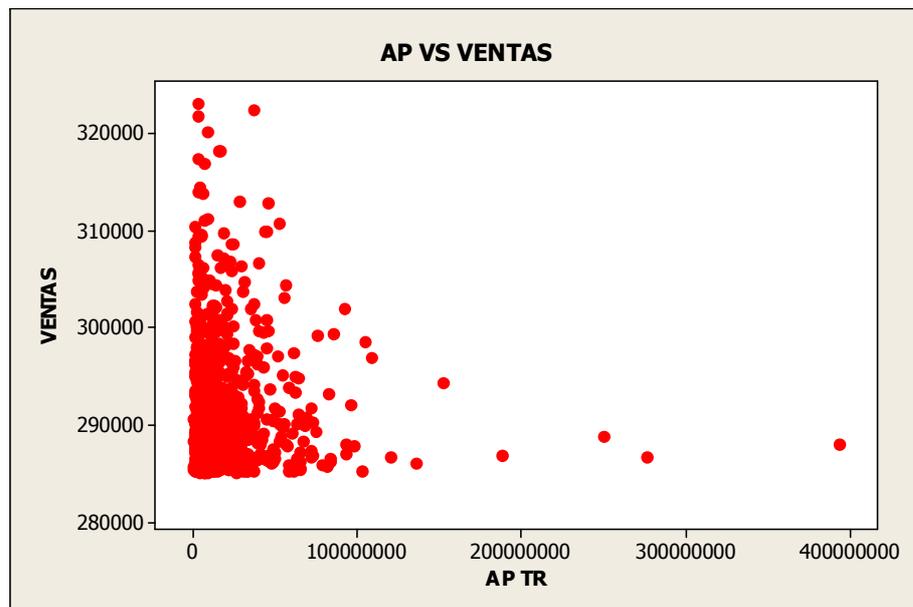


Figura 111. Gráfica de correlación entre la variable A&P y Ventas de Diovan. Fuente: Minitab.

La gráfica de correlación indica que existe cierta relación entre los valores bajos de A&P y las ventas de Voltaren. Esto puede resultar importante ya que indica que un gasto excesivo en Medios de Difusión para este producto no tiene efectos favorables a las ventas, información que se complementa con los resultados obtenidos de la regresión múltiple en donde A&P no resultó ser una variable significativa.

- **Descuentos vs Ventas:**

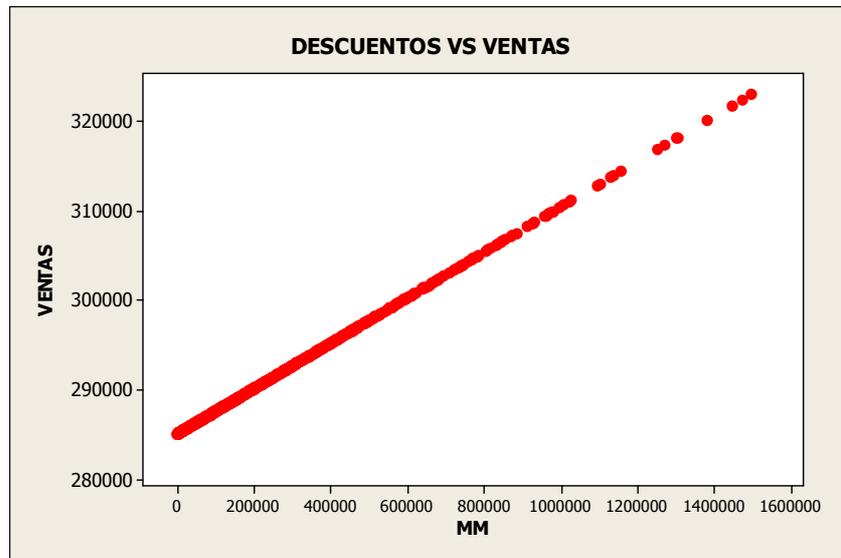


Figura 112. Gráfica de correlación entre la variable Descuentos y Ventas de Diován. Fuente: Minitab.

Se ve que existe una relación entre estas dos variables pues los puntos se concentran alrededor de una misma tendencia. Tal como se obtuvo en la regresión múltiple, la variable Descuentos mostró ser significativa y esto se comprueba con esta gráfica de correlación.

- **Muestras Médicas vs Ventas:**

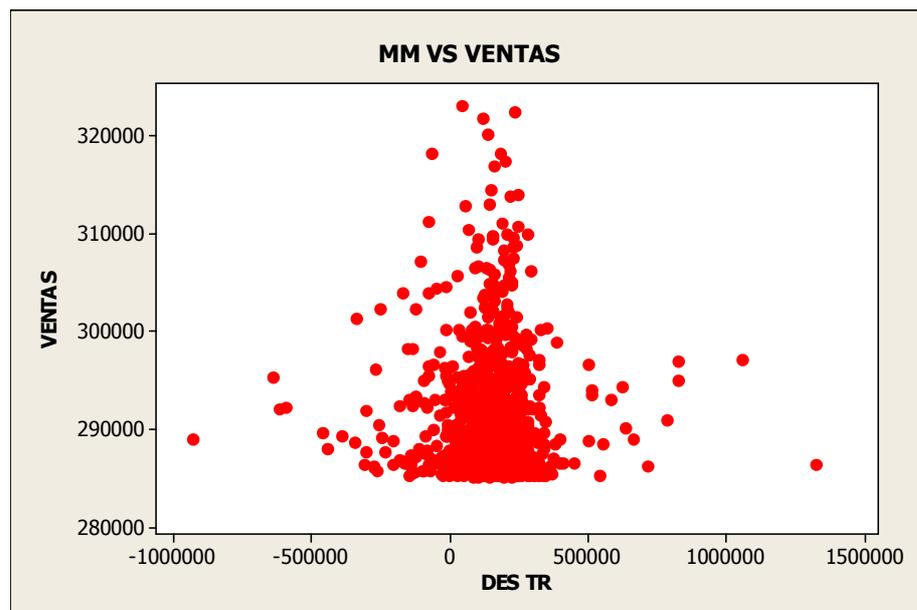


Figura 113. Gráfica de correlación entre la variable MM y Ventas de Diován. Fuente: Minitab.

Se observa un comportamiento la variable MM con Ventas en la cual no existe un patrón definido. Sin embargo se observa que bajo ese nivel se pueden obtener tanto Ventas bajas como altas de Diovan, por lo que no guardan mucha relación entre una variable y la otra.

- **Gráfico resumen de Ventas de Diovan:**

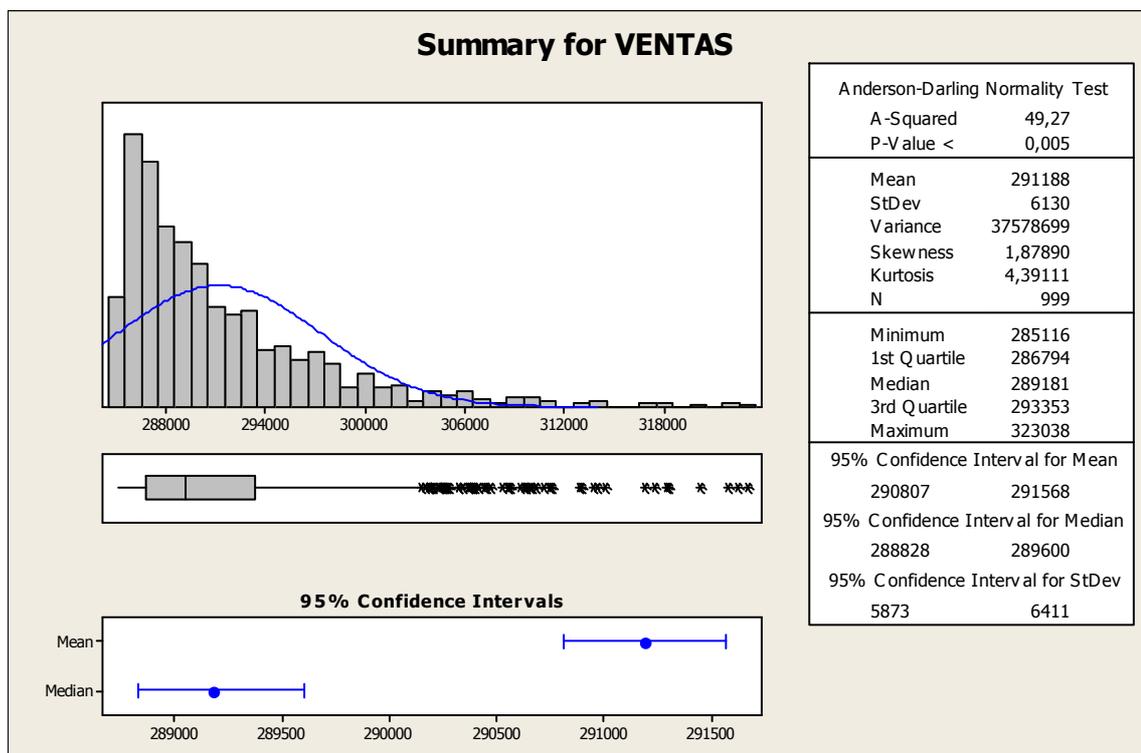


Figura 114. Gráfico resumen de Ventas de Diovan. Fuente: Minitab.

Se ve que los datos se ajustan a una distribución Normal tal como se infirió inicialmente, además que su valor p es mayor al nivel de significancia establecido. Un dato importante de este cuadro resumen son la media, y valores mínimos y máximos determinados a partir de la simulación Montecarlo.

\$323038, con los siguientes gastos en cada una de sus variables:

Variable	Gasto máximo en el que se debe incurrir
FF	\$ 2000,48
A&P	\$ 26860,63
Descuentos	\$ 45800,2
Muestras Médicas	\$ 14928,55
VENTAS GENERADAS	\$323038

Generación propia.

Tabla 13. Valores máximos de gastos por variable para el producto Diovan

Comparando ambos modelos de optimización y de acuerdo a los resultados obtenidos de la regresión múltiple en el que la variable Descuentos se observa que ambos modelos asignan un valor bajo a la Fuerza de Ventas, por lo que se comprueba que su aporte no es muy significativo para las Ventas del producto. Mientras que para la variable Descuentos en ambos se le da el valor ponderado de acuerdo a los resultados obtenidos de la regresión. Con simulación Montecarlo se incluyeron valores en la variable A&P mientras que en Solver no, por lo que el primero puede ser más específico que el segundo.

ANEXO C-4 Modelo de optimización para el producto Tegretol con Solver (sin simulación Montecarlo)

Resolviendo el Solver con la función objetivo planteada y las restricciones establecidas se obtuvo que para Tegretol, los valores óptimos para las variables: FF, A&P, Descuentos y Muestras Médicas son:

Gastos en FF	\$ 4046,14
Gastos en A&P	\$ 36024,4
Gastos en Descuentos	\$ 1196591,88
Gastos en Muestras Médicas	\$ 0

Generación propia.

Tabla 14. Niveles óptimos para cada variable del producto Tegretol.

Esto genera Ventas de Tegretol de \$ 244510,883

Modelo de optimización con MC-Sim Solver (con simulación Montecarlo) para Tegretol

De igual manera se siguieron los cinco pasos que se hicieron para el producto Voltaren. Con el modelo matemático se definió la función objetivo, se generaron 1000 números aleatorios pertenecientes a la distribución inferida de cada variable aleatoria. Par el caso de Tegretol se contaba con 4 variables regresoras, con las siguientes distribuciones de probabilidad:

Variable	FF	AP	DESCUENTOS	MM
Distribución	NORM(-0.0792, 0.788)	NORM(-0.111, 1)	NORM(0.115, 1.08)	-199 + EXPO(9.54e+003)

Generación propia.

Tabla 15. Distribuciones para las variables del producto Tegretol.

A partir de estas distribuciones se crearon 1000 datos aleatorios correspondientes a cada variable. A continuación se reemplazó los valores de

cada variable de entrada en la función de rentabilidad obtenida de la regresión. A continuación el análisis de los datos obtenidos:

- **FF con Ventas:**

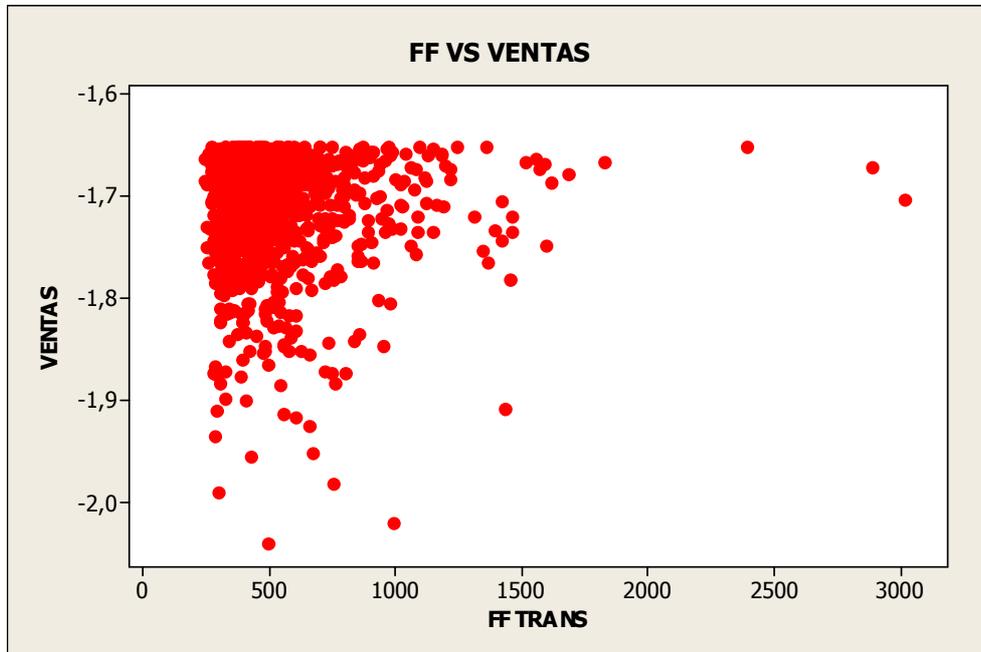


Figura 115. Gráfica de correlación entre la variable FF y Ventas de Tegretol. Fuente: Minitab.

Se puede notar que no hay una relación entre la FF y Ventas, ya que no se distingue un patrón en especial.

- **A&P con Ventas:**

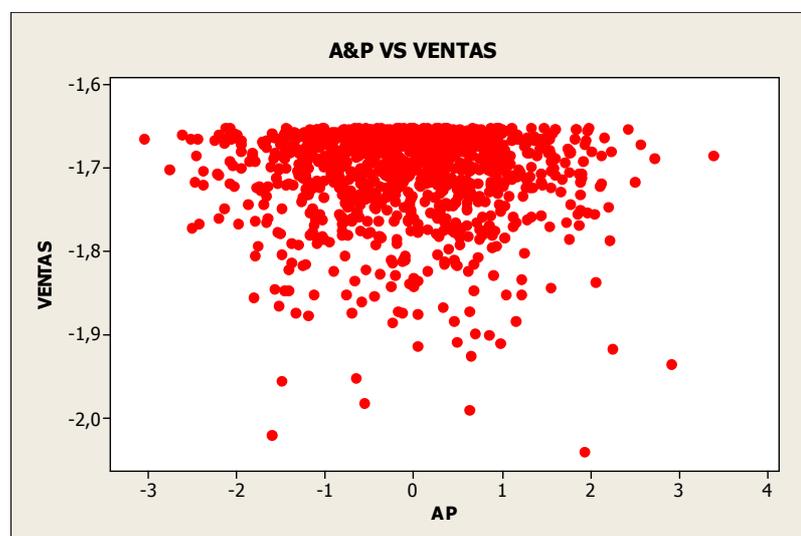


Figura 116. Gráfica de correlación entre la variable A&P y Ventas de Tegretol. Fuente: Minitab.

La gráfica de correlación indica que tampoco existe relación entre los valores bajos de A&P y las ventas de Voltaren. Esto puede resultar importante ya que indica que un gasto excesivo en Medios de Difusión para este producto no tiene efectos favorables a las ventas, información que se complementa con los resultados obtenidos de la regresión múltiple en donde A&P no resultó ser una variable significativa.

- **Descuentos vs Ventas:**

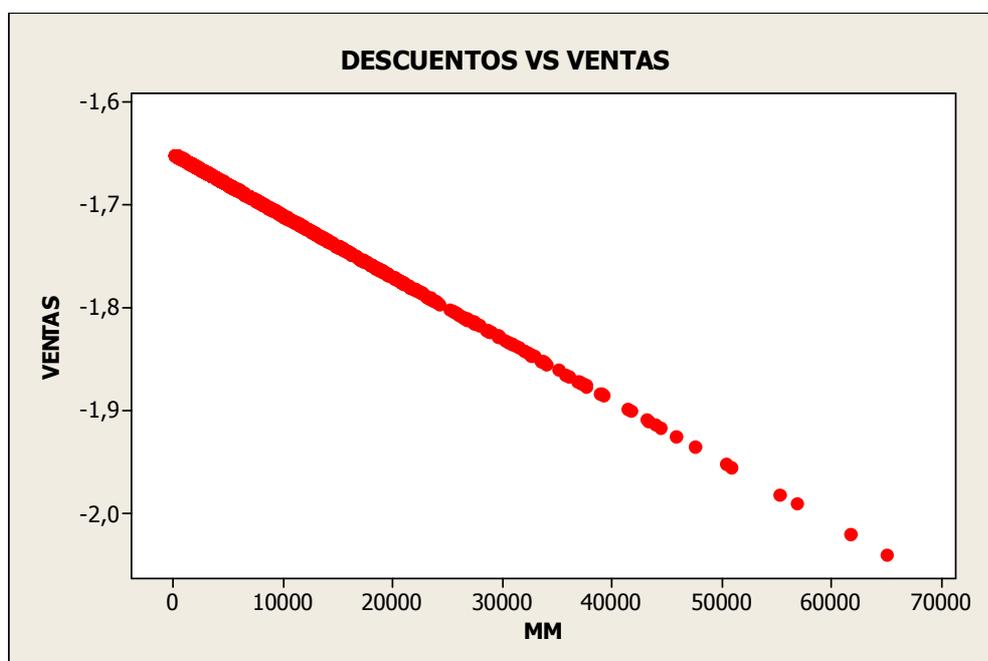


Figura 117. Gráfica de correlación entre la variable Descuentos y Ventas de Tegretol. Fuente: Minitab.

Se ve que existe una relación entre estas dos variables pues los puntos se concentran alrededor de un mismo punto. Tal como se obtuvo en la regresión múltiple, la variable Descuentos mostró ser significativa y esto se comprueba con esta gráfica de correlación.

- **Muestras Médicas vs Ventas:**

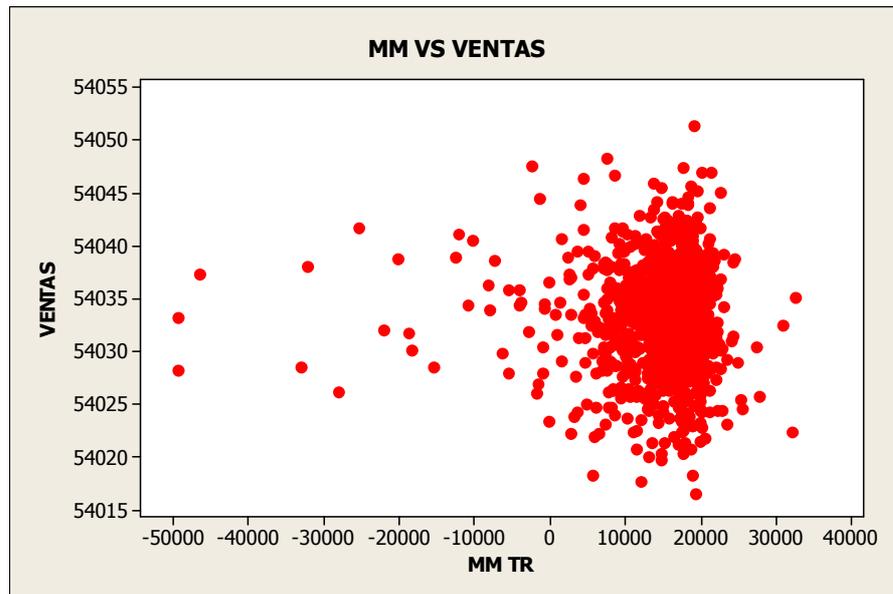


Figura 118. Gráfica de correlación entre la variable MM y Ventas de Tegretol. Fuente: Minitab.

Se observa un comportamiento de la variable MM con Ventas en la cual existe una agrupación de los datos alrededor del nivel \$20000 en el eje de Muestras Médicas. Sin embargo se observa que bajo ese nivel se pueden obtener tanto Ventas bajas como altas de Tegretol, por lo que no guardan mucha relación entre una variable y la otra.

- **Gráfico resumen de Ventas de Genteal:**

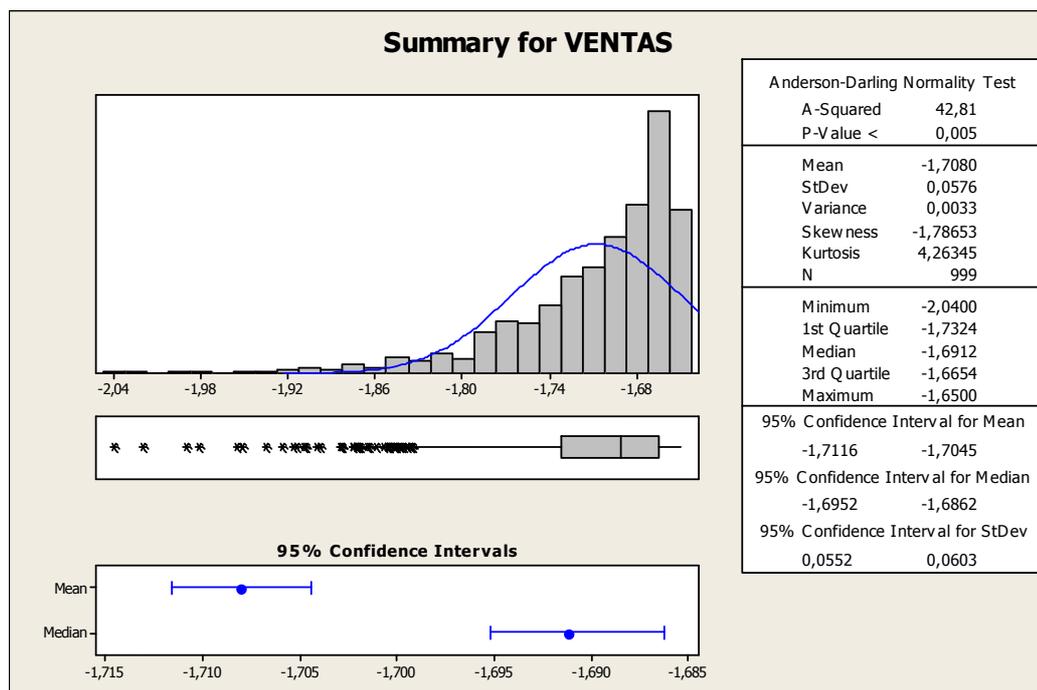


Figura 119. Gráfico resumen de Ventas de Tegretol. Fuente: Minitab.

Se ve que los datos se ajustan a una distribución Normal tal como se infirió inicialmente, además que su valor p es mayor al nivel de significancia establecido. Un dato importante de este cuadro resumen son la media, y valores mínimos y máximos determinados a partir de la simulación Montecarlo.

\$165463, con los siguientes gastos en cada una de sus variables:

Variable	Gasto máximo en el que se debe incurrir
FF	\$ 379,069
A&P	\$ 5940
Descuentos	\$ 52459,91
Muestras Médicas	\$ 196
VENTAS GENERADAS	\$165463

Generación propia.

Tabla 16. Valores máximos de gastos por variable para el producto Tegretol

Comparando ambos modelos de optimización se observa que mediante la simulación Montecarlo sí se asigna un valor a la variable FF mientras que con Solver no. Esto puede ser curioso para generar más datos aleatorios y ver qué sucede si en la simulación Montecarlo el valor de FF es cero. Por su parte en la variable Descuentos es donde se asigna mayor presupuesto por su relación de proporcionalidad obtenida en la regresión.

BIBLIOGRAFÍA

Banks, Jerry. Discrete- Event Systems Simulation. Estados Unidos: Pearson, 2005.

“Cataflam”. 24 de Abril. <<http://www.facmed.unam.mx/bmnd/plm/34731.htm>>

Córdoba, Pablo. BPA NOVARTIS. Jara, Carlos. BPA NOVARTIS. Entrevista Personal. (Michielle Zurita). Febrero- Mayo 2010.

Damodar N. Gujarati. Econometría. Cuarta edición. México: Mc Graw Hill, 2003.

“Diovan”. 26 de Abril. <<http://www.diovan.com>>

“Genteal”. 24 Abril. <<http://www.genteal.com>>

Hernandez, Blenda, and Tobias Garces. Aplicación de los modelos de regresión tobit en la modelización de variables epidemiológicas censuradas, 2006, pag. Web. 10 May 2010.

Humberto, Barreto. Frank. Howland. Introductory Econometrics, Montecarlo Simulation with Microsoft Excel. New York: Cambridge University Press, 2006.

Jara, Carlos. BPA NOVARTIS. Entrevista Personal. (Michielle Zurita). Febrero-Mayo 2010.

Kelton, David, Randal Sadowski, Sturrock, David. Simulation with Arena. 3era edición. Nueva York: Mc Graw- Hill, 2004.

Manual De Bienvenida, Novartis Ecuador. Edición 2009-2010.

“Modelos estocasticos en ingeniería”. 13 Marzo 2010. <<http://metodosestadisticos.unizar.es/asignaturas/12072/MaterialClase/guia.pdf>>

“Montecarlo Simulation Basic”. 25 Marzo 2010. <<http://www.vertex42.com/ExcelArticles/mc/MonteCarloSimulation.html>>

Montgomery. Diseño de experimentos. 2nda edición. México: Limusa Wiley, 2006.

Montgomery, Douglas C. Control Estadístico de la calidad. 3era Edición. México: Limusa S.A., 2006.

Montgomery, Runger. Probabilidad y estadística aplicada a la ingeniería. 2nda Edición. México: Limusa Wiley, 2005.

“Tegretol”. 28 de Abril. <http://www.vademecum.es/medicamento-tegretol_3858.htm>

Vergara, Héctor Andrés. M.SC. Entrevista Personal. (Michielle Zurita). Abril 2010.

“Voltaren”. 24 Abril. <<http://www.voltaren.com>>