

**INSTITUTO TECNOLÓGICO Y DE ESTUDIOS
SUPERIORES DE MONTERREY**

CAMPUS MONTERREY

DIVISION DE INGENIERIA Y ARQUITECTURA

PROGRAMA DE GRADUADOS EN INGENIERIA



**TECNOLÓGICO
DE MONTERREY.**

**“Estudio del Comportamiento de la Productividad en
Departamentos Funcionales en una Empresa Detallista
de Autoservicio”**

T E S I S

**PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO ACADÉMICO DE:**

**MAESTRO EN CIENCIAS
CON ESPECIALIDAD EN CALIDAD Y PRODUCTIVIDAD**

POR

ING. KARINA FERNANDEZ QUINTERO

MONTERREY, NUEVO LEÓN

JULIO DE 2007

**INSTITUTO TECNOLÓGICO Y DE ESTUDIOS
SUPERIORES DE MONTERREY**

**CAMPUS MONTERREY
DIVISIÓN DE INGENIERIA Y ARQUITECTURA
PROGRAMA DE GRADUADOS EN INGENIERIA**



**TECNOLÓGICO
DE MONTERREY.®**

**“Estudio del Comportamiento de la Productividad en
Departamentos Funcionales en una Empresa Detallista
de Autoservicio”**

**TESIS
PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO ACADÉMICO DE**

**MAESTRO EN CIENCIAS
CON ESPECIALIDAD EN CALIDAD Y PRODUCTIVIDAD**

**POR:
ING. KARINA FERNÁNDEZ QUINTERO**

MONTERREY, N.L.

JULIO 2007

Dedicatoria

A Dios que me inspira a ser una mejor persona cada día.

A mis padres, Martha y Roberto por su amor incondicional.

A mis hermanos, Denise y Roberto por su apoyo, alegría y cariño.

Agradecimientos

A mi asesor de tesis el Dr. Neal Smith por su apoyo en la realización de este proyecto de investigación.

A la Ing. María del Carmen Temblador por guiarme en la elaboración de la presente tesis.

Al Dr. Jesús Velazco por confiar en mí para la realización de este estudio.

Al Dr. Guadalupe Ríos por ayudarme en el análisis estadístico e interpretación de los resultados.

A la Ing. Deyanira Trujillo por ser un ejemplo de vida y lucha.

A Blanca Gamez por su consejo y amistad.

A Concepción Lira y Roberto Infante por su ayuda en la recolección e interpretación de los datos.

Al Dr. Francisco Ángel Bello por su apoyo.

A Jorge Morales por su paciencia y amor.

A mis amigos que siempre están para escucharme, en especial a Marcia Ortega, Vanessa Huerta y Francisco Corona por su ayuda.

A mis compañeros y profesores del Departamento de Ingeniería Industrial y de Sistemas.

Contenido

Dedicatoria, i

Agradecimientos, ii

Contenido, iii

Lista de Tablas, viii

Lista de Figuras, x

Lista de Gráficas, xi

Resumen ejecutivo, xii

Capítulo 1. Contexto General

1.1 Introducción, 1

1.2 Planteamiento del problema, 1

1.3 Objetivos, 1

1.4 Hipótesis, 2

1.5. Preguntas de investigación, 3

1.6. Justificación, 4

1.7. Método de investigación, 4

1.7.1. Tipo de estudio, 4

1.7.2. Pasos para elaborar la investigación, 5

1.7.3. Selección de la muestra, 5

1.7.4. Recolección de datos, 6

1.7.5. Análisis de datos, 6

1.8. Alcance y limitaciones, 6

1.8.1. Características de la muestra, 6

1.8.2. Área geográfica para el estudio, 6

1.8.3. Limitaciones no definidas por el investigador, 7

Capítulo 2. Marco Teórico

2.1 Introducción, 8

2.2. Los cambios en las operaciones de la industria de los alimentos, 9

2.3. La función de la mercadotecnia y la calidad del servicio en el sector detallista, 10

2.4. Productividad laboral en la venta al detalle, 11

- 2.5. Empleados de tiempo parcial y completo en la industria detallista de alimentos, 15
- 2.6. Venta por hora hombre en el comercio detallista, 17
- 2.7. El uso de la tecnología en la venta al por menor, 17
- 2.8. Desempeño de las tiendas de alimentos, 19
- 2.9. La reinención de los supermercados, 22
- 2.10. Motivos de los consumidores en la selección de alimentos y tiendas de comestibles, 25
- 2.11. El efecto de la inflación y el índice de precios en las ventas minoristas de alimentos, 28

Capítulo 3. Metodología

- 3.1. Introducción, 30
- 3.2. Metodología, 30
 - 3.2.1. Definición de los objetivos del análisis, 30
 - 3.2.2. Determinación de variables, 30
 - 3.2.3. Selección de la muestra y recolección de datos, 31
 - 3.2.4. Modelo de regresión lineal múltiple, 31
 - 3.2.4.1. Etapa 1: Modelo general, 31
 - 3.2.4.2. Etapa 2: Segmentación, 32
 - 3.2.4.3. Etapa 3: Formación de grupos, 32
 - 3.2.4.4. Etapa 4: Eliminación de variables no significativas, 32
- 3.3. Conclusiones, 33

Capítulo 4. Resultados

- 4.1. Introducción, 34
- 4.2. Aplicación de la metodología, 34
 - 4.2.1. Definición de los objetivos del análisis, 34
 - 4.2.2. Determinación de variables, 34
 - 4.2.2.1. Ventas (y), 34
 - 4.2.2.2. Sucursal (x_1), 35
 - 4.2.2.3. Departamento (x_2), 36
 - 4.2.2.4. Quincenas (x_3), 37
 - 4.2.2.5. Cantidad de producto vendido (x_4), 38
 - 4.2.2.6. Cantidad de socios parciales (x_5), 38
 - 4.2.2.7. Cantidad de socios completos (x_6), 38
 - 4.2.2.8. Cantidad de bajas (x_7), 39
 - 4.2.2.9. Horas disponibles de trabajo efectivo (x_8), 39
 - 4.2.2.10. Nómina por departamento (x_9), 39
 - 4.2.2.11. *Add loss* (x_{10}), 40
 - 4.2.2.12. Margen de ganancia (x_{17}), 40

- 4.2.2.13. Venta por hora hombre (x_{18}), 40
- 4.2.3. Selección de la muestra y recolección de datos, 41
- 4.2.4. Modelo de regresión múltiple, 41
 - 4.2.4.1. Etapa 1: Modelo general, 41
 - 4.2.4.1.1. Estimación de los parámetros del modelo, 41
 - 4.2.4.1.2. Análisis de varianza (ANOVA), 45
 - 4.2.4.1.3. Gráficas de dispersión, 46
 - 4.2.4.1.4. Análisis de multicolinealidad, 49
 - 4.2.4.1.4.1. Matriz de correlaciones y covarianza, 49
 - 4.2.4.1.4.2. Factores de inflación de varianza, 50
 - 4.2.4.1.4.3. Análisis de eigensistema $X'X$, 51
 - 4.2.4.2. Etapa 2: Segmentación, 52
 - 4.2.4.2.1. Estimación de los parámetros del modelo, 53
 - 4.2.4.2.2. Gráficas de dispersión, 54
 - 4.2.4.3. Etapa 3: Formación de grupos, 57
 - 4.2.4.3.1. Modelo A, 57
 - 4.2.4.3.1.1. Método *Stepwise*, 59
 - 4.2.4.3.2. Modelo B, 60
 - 4.2.4.3.2.1. Método *Stepwise*, 62
 - 4.2.4.3.3. Modelo D, 64
 - 4.2.4.3.3.1. Método *Stepwise*, 65
 - 4.2.4.3.4. Modelo Combo, 67
 - 4.2.4.3.4.1. Método *Stepwise*, 68
 - 4.2.4.4. Etapa 4: Eliminación de variables no significativas, 70
 - 4.2.4.4.1. Modelo final A, 70
 - 4.2.4.4.1.1. Estimación de los parámetros del modelo, 71
 - 4.2.4.4.1.2. Supuestos del análisis de regresión, 72
 - 4.2.4.4.1.3. Prueba de hipótesis en regresión lineal múltiple, 78
 - 4.2.4.4.1.4. Análisis de varianza (ANOVA), 79
 - 4.2.4.4.1.5. Intervalos de confianza, 80
 - 4.2.4.4.1.6. Gráficas de dispersión, 81
 - 4.2.4.4.2. Modelo final B, 83
 - 4.2.4.4.2.1. Estimación de los parámetros del modelo, 84
 - 4.2.4.4.2.2. Supuestos del análisis de regresión, 85
 - 4.2.4.4.2.3. Prueba de hipótesis en regresión lineal múltiple, 90
 - 4.2.4.4.2.4. Análisis de varianza (ANOVA), 92
 - 4.2.4.4.2.5. Intervalos de confianza, 93
 - 4.2.4.4.2.6. Gráficas de dispersión, 95
 - 4.2.4.4.3. Modelo final D, 98
 - 4.2.4.4.3.1. Estimación de los parámetros del modelo, 99
 - 4.2.4.4.3.2. Supuestos del análisis de regresión, 100

4.2.4.4.3.3.	Prueba de hipótesis en regresión lineal múltiple,	104
4.2.4.4.3.4.	Análisis de varianza (ANOVA),	105
4.2.4.4.3.5.	Intervalos de confianza,	107
4.2.4.4.3.6.	Gráficas de dispersión,	108
4.2.4.4.4.	Modelo final Combo,	109
4.2.4.4.4.1.	Estimación de los parámetros del modelo,	110
4.2.4.4.4.2.	Supuestos del análisis de regresión,	111
4.2.4.4.4.3.	Prueba de hipótesis en regresión lineal múltiple,	116
4.2.4.4.4.4.	Análisis de varianza (ANOVA),	118
4.2.4.4.4.5.	Intervalos de confianza,	120
4.2.4.4.4.6.	Gráficas de dispersión,	121
4.2.5.	Análisis de sensibilidad,	123
4.3.	Conclusiones,	127

Capítulo 5. Conclusiones

5.1	Conclusiones,	128
5.1.1.	Del proceso,	128
5.1.2.	De la investigación,	128
5.1.3.	Del uso de la herramienta,	129
5.1.4.	De las pruebas de hipótesis,	130
5.1.5.	De los resultados,	131
5.2.	Recomendaciones para futuras investigaciones,	133

Referencias, 134

Anexos, 138

Anexo 1.	Índice Nacional de Precios al Consumidor 2003,	138
Anexo 2.	Índice Nacional de Precios al Consumidor 2004,	139
Anexo 3.	Índice Nacional de Precios al Consumidor 2005,	140
Anexo 4.	Índice Nacional de Precios al Consumidor 2006,	141
Anexo 5.	Índice Nacional de Precios al Consumidor 2007,	142
Anexo 6.	Gráficas de dispersión de la etapa 1: modelo general,	143
Anexo 7.	Matriz de correlaciones modelo general,	153
Anexo 8.	Matriz de covarianza modelo general,	154
Anexo 9.	Gráficas de dispersión de la etapa 2: segmentación,	155
Anexo 10.	Método <i>Stepwise</i> etapa 3: A,	164
Anexo 11.	Método <i>Stepwise</i> etapa 3: B,	167
Anexo 12.	Método <i>Stepwise</i> etapa 3: D,	170
Anexo 13.	Método <i>Stepwise</i> etapa 3: Combo,	173
Anexo 14.	Método <i>Stepwise</i> etapa 4: A,	176

Anexo 15. Método *Stepwise* etapa 4: B, 183
Anexo 16. Método *Stepwise* etapa 4: D, 189
Anexo 17. Método *Stepwise* etapa 4: Combo, 193

Lista de Tablas

Capítulo 2. Marco Teórico

- Tabla 2.1.** Ventas de supermercados 2004, 13
Tabla 2.2. Mediciones de desempeño de supermercados 2004, 22
Tabla 2.3. 45 razones para seleccionar un supermercado, 24
Tabla 2.4. Razones de sección de una tienda de alimentos entre jóvenes (menores a 55) y adultos (mayores a 55), 25
Tabla 2.5. Como gastar 100 dólares en un supermercado, 27

Capítulo 4. Resultados

- Tabla 4.1.** Variables indicadoras de sucursal, 35
Tabla 4.2. Variables indicadoras de departamento, 37
Tabla 4.3. Coeficientes del modelo etapa 1, 42
Tabla 4.4. Tabla ANOVA, 45
Tabla 4.5. Grados de libertad y suma al cuadrado secuencial, 46
Tabla 4.6. Correlaciones simples entre regresores, 50
Tabla 4.7. Factores de inflación de varianza, 51
Tabla 4.8. Análisis de eigensistema, 52
Tabla 4.9. Coeficientes del modelo etapa 2, 54
Tabla 4.10. Coeficientes del modelo etapa 3 - A, 58
Tabla 4.11. Coeficientes del modelo etapa 3 - B, 61
Tabla 4.12. Coeficientes del modelo etapa 3 - D, 64
Tabla 4.13. Coeficientes del modelo etapa 3 - Combo, 67
Tabla 4.14. Coeficientes del modelo etapa 4 - A, 72
Tabla 4.15. Valores T - A, 79
Tabla 4.16. Tabla ANOVA - A, 79
Tabla 4.17. Grados de libertad y suma al cuadrado secuencial - A, 80
Tabla 4.18. Intervalo de confianza - A, 81
Tabla 4.19. Coeficientes del modelo etapa 4 - B, 85
Tabla 4.20. Valores T - B, 92
Tabla 4.21. Tabla ANOVA - B, 92
Tabla 4.22. Grados de libertad y suma al cuadrado secuencial - B, 93
Tabla 4.23. Intervalo de confianza - B, 94
Tabla 4.24. Coeficientes del modelo etapa 4 - D, 99
Tabla 4.25. Valores T - D, 105
Tabla 4.26. Tabla ANOVA - D, 106
Tabla 4.27. Grados de libertad y suma al cuadrado secuencial - D, 107
Tabla 4.28. Intervalo de confianza - D, 107
Tabla 4.29. Coeficientes del modelo etapa 4 - Combo, 110
Tabla 4.30. Valores T - Combo, 132
Tabla 4.31. Tabla ANOVA - Combo, 118
Tabla 4.32. Grados de libertad y suma al cuadrado secuencial - Combo, 120
Tabla 4.33. Intervalo de confianza - Combo, 121
Tabla 4.34. Análisis de sensibilidad horas de trabajo efectivo, 124
Tabla 4.35. Análisis de sensibilidad horas de trabajo efectivo sin Depto. Carnes, 125
Tabla 4.36. Análisis de sensibilidad horas de trabajo efectivo con Depto. Carnes, 126

Capítulo 5. Conclusiones

Tabla 5.1. Resumen de los resultados de pruebas de hipótesis, 130

Lista de Figuras

- Figura 4.1.** Estimados del modelo de regresión etapa 1, 42
Figura 4.2. Estimados del modelo de regresión etapa 2, 54
Figura 4.3. Estimados del modelo de regresión etapa 3 - A, 58
Figura 4.4. Procedimiento *Stepwise* - A, 60
Figura 4.5. Estimados del modelo de regresión etapa 3 - B, 61
Figura 4.6. Procedimiento *Stepwise* - B, 63
Figura 4.7. Estimados del modelo de regresión etapa 3 - D, 65
Figura 4.8. Procedimiento *Stepwise* - D, 66
Figura 4.9. Estimados del modelo de regresión etapa 3 - Combo, 68
Figura 4.10. Procedimiento *Stepwise* - Combo, 69
Figura 4.11. Estimados del modelo de regresión etapa 4 - A, 72
Figura 4.12. Resultados de *Run Test* etapa 4 - A, 77
Figura 4.13. Estimados del modelo de regresión etapa 4 - B, 85
Figura 4.14. Resultados de *Run Test* etapa 4 - B, 90
Figura 4.15. Estimados del modelo de regresión etapa 4 - D, 100
Figura 4.16. Resultados de *Run Test* etapa 4 - D, 103
Figura 4.17. Estimados modelo de regresión etapa 4 - Combo, 111
Figura 4.18. Resultados de *Run Test* etapa 4 - Combo, 116

Lista de Gráficas

- Gráfica 4.1.** Histograma de residuales, 43
- Gráfica 4.2.** Gráfica de probabilidad normal de los residuales, 44
- Gráfica 4.3.** Gráfica de residuales contra valores ajustados, 45
- Gráfica 4.4.** Ventas versus cantidad de producto vendido etapa 1, 47
- Gráfica 4.5.** Ventas versus horas disponibles de trabajo efectivo etapa 1, 47
- Gráfica 4.6.** Ventas versus ventas por hora hombre etapa 1, 48
- Gráfica 4.7.** Ventas versus cantidad de producto vendido etapa 2, 55
- Gráfica 4.8.** Ventas versus horas disponibles de trabajo efectivo etapa 2, 56
- Gráfica 4.9.** Ventas versus ventas por hora hombre etapa 2, 56
- Gráfica 4.10.** Histograma de residuales etapa 4 - A, 73
- Gráfica 4.11.** Residuales versus los valores ajustados etapa 4 - A, 74
- Gráfica 4.12.** Residuales versus el orden de los datos etapa 4 - A, 75
- Gráfica 4.13.** Resumen para residuales etapa 4 - A, 76
- Gráfica 4.14.** Ventas versus cantidad de producto vendido etapa 4 - A, 82
- Gráfica 4.15.** Ventas versus horas disponibles de trabajo efectivo etapa 4 - A, 82
- Gráfica 4.16.** Ventas versus margen de ganancia etapa 4 - A, 83
- Gráfica 4.17.** Gráfica de residuales etapa 4 - B, 86
- Gráfica 4.18.** Gráfica de residuales versus valores ajustados etapa 4 - B, 87
- Gráfica 4.19.** Residuales versus el orden de los datos etapa 4 - B, 88
- Gráfica 4.20.** Gráfica para residuales etapa 4 - B, 89
- Gráfica 4.21.** Ventas versus cantidad de producto vendido etapa 4 - B, 95
- Gráfica 4.22.** Ventas versus nómina por departamento etapa 4 - B, 96
- Gráfica 4.23.** Ventas versus *add loss* etapa 4 - B, 96
- Gráfica 4.24.** Ventas versus margen de ganancia etapa 4 - B, 97
- Gráfica 4.25.** Histograma de residuales para etapa 4 - D, 100
- Gráfica 4.26.** Residuales versus los valores ajustados etapa 4 - D, 101
- Gráfica 4.27.** Residuales versus el orden de los datos etapa 4 - D, 102
- Gráfica 4.28.** Gráfica de normalidad de residuales etapa 4 - D, 103
- Gráfica 4.29.** Ventas versus margen de ganancia etapa 4 - D, 108
- Gráfica 4.30.** Histograma de residuales etapa 4 - Combo, 112
- Gráfica 4.31.** Residuales versus valores ajustados etapa 4 - Combo, 113
- Gráfica 4.32.** Residuales versus el orden de los datos etapa 4 - Combo, 114
- Gráfica 4.33.** Resumen para residuales etapa 4 - Combo, 115
- Gráfica 4.34.** Ventas versus cantidad de producto vendido etapa 4 - Combo, 122
- Gráfica 4.35.** Ventas versus venta por hora hombre etapa 4 - Combo, 122
- Gráfica 4.36.** Análisis de sensibilidad ventas vs. horas de trabajo efectivo, 124
- Gráfica 4.37.** Análisis de sensibilidad ventas vs. horas de trabajo efectivo sin Carnes, 125
- Gráfica 4.38.** Análisis de sensibilidad ventas vs. horas de trabajo efectivo con Carnes, 126

Resumen Ejecutivo

La intensa competencia a la que se enfrentan las empresas basadas en servicios, ha provocado la necesidad de desarrollar estrategias que mejoren los niveles de productividad, como un factor vital para alcanzar altos estándares de calidad que conduzcan el proceso de la venta de alimentos perecederos en un supermercado.

El remanente de proyecto de tesis plantea inicialmente la definición del problema, objetivos e hipótesis del estudio, así como la revisión de literatura existente sobre los factores que definen los indicadores de productividad del sector industrial referente a la venta al detalle de productos alimentarios. Posteriormente se describe la metodología para la definición del modelo matemático de regresión lineal múltiple, seguido de la presentación de los resultados. El presente documento concluye con la discusión de las implicaciones de los hallazgos obtenidos y las recomendaciones para futuras investigaciones.

Capítulo 1. Contexto General

1.1. Introducción

En el contexto actual de los negocios, cada vez más complejos, la gestión de una empresa de autoservicio minorista requiere de una visión estratégica que permita posicionar eficazmente el negocio ante la competencia.

Algunos de los retos más importantes a los que se enfrenta la industria detallista de alimentos incluyen la intensificación de la competencia global en el mercado, el incremento continuo de las expectativas y demandas de los clientes, así como la necesidad de alcanzar altos niveles de eficiencia y productividad.

El presente trabajo de investigación corresponde al desarrollo de modelos de regresión múltiple que estudian la relación existente entre variables que impactan la productividad de los departamentos de manufactura de alimentos perecederos de diversas tiendas que forman parte de una reconocida cadena de supermercados.

1.2. Planteamiento del problema

La obtención de información que pueda ser utilizada para la toma de decisiones, ha generado la necesidad de identificar como las variables de productividad tienen influencia sobre el comportamiento de las ventas de los departamentos de perecederos en una tienda detallista de autoservicio.

Por tanto, la problemática que se presenta en la investigación se define con la siguiente pregunta:

¿Cómo se relaciona el incremento en las ventas de una tienda detallista de autoservicio con las variables de productividad de los departamentos de alimentos de la misma?

1.3. Objetivos

El objetivo general del presente proyecto de tesis es plantear un modelo de regresión lineal múltiple para el análisis de variables de productividad que tienen influencia significativa en la cantidad de ventas generadas en empresas comercializadoras de autoservicio.

Los objetivos particulares de investigación que se buscan lograr se indican a continuación:

- (1) Adquirir información acerca de la variable dependiente al conocer los valores de las variables independientes.
- (2) Generar un modelo que permita pronosticar las ventas de una tienda en particular.
- (3) Obtener una medida cuantitativa de la magnitud con la que las variables significativas bajo estudio impactan a las ventas.
- (4) Estudiar la relación matemática determinista entre la variable dependiente e independiente del modelo de regresión aplicado a la predicción de ventas en función de datos tales como: sucursal, departamento, quincenas, cantidad de socios completos y parciales disponibles, cantidad de bajas y nómina por departamento, horas disponibles de trabajo efectivo, rebajas, margen de ganancia y venta por hora hombre.
- (5) Evaluar el coeficiente de correlación de cada una de las variables de interés para la medición del nivel de productividad alcanzado por una tienda minorista de autoservicio en un periodo de tiempo definido.

1.4. Hipótesis

De acuerdo a los objetivos descritos en la sección anterior, se identifica como el propósito principal de tesis a la generación de un modelo matemático de regresión múltiple que permita estudiar una relación útil entre la variable dependiente y las independientes.

Las hipótesis que se plantean a continuación representan la guía de investigación sobre el fenómeno que se intenta probar en el estudio.

- (1) No existe relación significativa entre las variables independientes del estudio y las ventas de una tienda de autoservicio detallista.**
- (2) No existe relación entre la variable quincenas y las ventas de una tienda de autoservicio detallista.**
- (3) No existe relación entre la cantidad de producto vendido de los departamentos de Carnes, Pescados y Mariscos, Delicatessen, Café, Panadería, y Tortillería, con las ventas de una tienda detallista de autoservicio.**

- (4) No existe relación entre la cantidad de socios parciales que laboran en una sucursal y las ventas de la misma.
- (5) No existe relación entre la cantidad de socios completos que laboran en un supermercado y las ventas del mismo.
- (6) No existe relación entre la cantidad de bajas semanales y las ventas de una tienda detallista de autoservicio.
- (7) No existe relación entre la variable de productividad de horas disponibles de trabajo efectivo y las ventas de una tienda detallista de autoservicio.
- (8) No existe relación entre la nómina por departamento y las ventas de una tienda detallista de autoservicio.
- (9) No existe relación entre las rebajas de precio (*add loss*) en los productos de los departamentos de Carnes, Pescados y Mariscos, Delicatessen, Café, Panadería, y Tortillería, y las ventas de una tienda detallista de autoservicio.
- (10) No existe relación entre el indicador de productividad de margen de ganancia y las ventas de una tienda detallista de autoservicio.
- (11) No existe relación entre las ventas por hora hombre y las ventas de una tienda detallista de autoservicio.

1.5. Preguntas de investigación

Para encontrar soluciones factibles a la situación problemática planteada y lograr alcanzar los objetivos establecidos, es de utilidad definir las siguientes preguntas de investigación:

- (1) ¿Qué es una empresa minorista de autoservicio?
- (2) ¿Cuáles son las características del comercio detallista?
- (3) ¿Qué operaciones se llevan a cabo en la industria de los alimentos?
- (4) ¿Qué impacto ha tenido sobre los clientes el desarrollo de la venta al por menor?
- (5) ¿Cuáles son los indicadores de productividad de una empresa minorista de alimentos?
- (6) ¿Qué variables de productividad impactan significativamente en los cambios de las ventas de una tienda de alimentos?
- (7) ¿Qué indicadores financieros se utilizan para analizar las ventas?
- (8) ¿Qué herramientas estadísticas pueden utilizarse para conocer la relación entre variables?
- (9) ¿Qué es regresión lineal múltiple?
- (10) ¿Cómo se aplica regresión lineal múltiple?

1.6. Justificación

En la actualidad, muchas organizaciones enfrentan los cambios acelerados que se presentan en los mercados internacionales; por esto, las empresas del sector de alimentos han trabajado en brindar valor agregado a sus clientes por medio de mejores prácticas en las operaciones clave de diferenciación.

Las tiendas detallistas de alimentos consisten en establecimientos donde se ofrece a los clientes un servicio de venta de productos comestibles para su preparación y consumo.

Existen diversos desafíos a los que se enfrentan actualmente los detallistas, entre ellos se encuentra el aumento de los costos de operación, la disminución en el ritmo de crecimiento, la fragmentación de los mercados, la existencia de mayor competencia por parte de grandes empresas, así como la presencia de clientes cada vez más exigentes (Adebanjo y Mahoney, 2004).

En consecuencia, la madurez del mercado de los alimentos ha iniciado una competencia intensa y una guerra de precios para mejorar la eficiencia organizacional de las empresas dedicadas a la venta al detalle de productos comestibles (Adebanjo y Mahoney, 2004).

Así, se ha considerado necesario estudiar el efecto de los indicadores de productividad sobre el comportamiento en las ventas de los departamentos de perecederos de las tiendas de una cadena de supermercados.

La investigación planteada en este proyecto contribuirá en la generación de modelos de carácter estadístico, que faciliten la predicción de los cambios en las ventas con respecto a la influencia de las variables de productividad definidas por la empresa.

1.7. Método de investigación

1.7.1. Tipo de estudio

El tipo de investigación para el presente proyecto corresponde a un estudio descriptivo. De acuerdo a Hernández et al. (2003), la investigación descriptiva busca medir y recolectar información sobre las propiedades, características y rasgos del fenómeno de interés.

Por tanto, la estrategia a seguir para obtener la información deseada y alcanzar los objetivos de la investigación, se llevará a cabo por medio de un estudio teórico donde se busque identificar las variables de productividad que tienen un impacto sobre los cambios en las ventas de una tienda minorista de alimentos.

Así mismo, la investigación tendrá un enfoque cuantitativo y práctico, ya que se busca obtener y evaluar con datos reales de una empresa del sector bajo estudio, los modelos de regresión múltiple desarrollados para la determinación de la relación entre la variable independiente y las variables dependientes.

1.7.2. Pasos para elaborar la investigación

Los pasos a seguir para la elaboración de la tesis se definen a continuación:

- (1) Llevar a cabo la definición de la metodología de investigación mostrada en el presente capítulo.
- (2) Realizar una investigación bibliográfica que permita revisar y documentar información relevante al problema bajo estudio.
- (3) Definir la herramienta estadística a utilizar para estudiar la problemática de esta tesis.
- (4) Recolectar los datos necesarios para generar los modelos de regresión múltiple.
- (5) Evaluar e integrar la información para desarrollar los modelos de regresión múltiple.
- (6) Elaborar el reporte de resultados y aplicación de la herramienta estadística utilizada.
- (7) Validar los resultados por medio de datos y aprobación del experto.
- (8) Elaborar conclusiones y recomendaciones para futuras investigaciones.

1.7.3. Selección de la muestra

Debido a la disponibilidad de datos para el análisis y desarrollo de los modelos de regresión, la información ha sido limitada a la selección de una muestra representativa de 6 sucursales y 6 departamentos de alimentos perecederos.

Los datos históricos que se revisan corresponden a 23 variables analizadas y han sido documentados para los años 2004, 2005, y 2006, medidos en

unidades de tiempo igual a 52 semanas. Por consiguiente, la cantidad de datos a ser evaluados es de 129,168.

1.7.4. Recolección de datos

La recolección de información se realiza por medio de la investigación bibliográfica que se muestra en el desarrollo del marco teórico de la presente tesis. Por otro lado, la recolección de datos numéricos a partir de registros históricos existentes y la determinación de las variables bajo estudio, se lleva a cabo por medio de la sugerencia del experto de la empresa donde se aplica el proyecto.

1.7.5. Análisis de datos

Con el objeto de eliminar el efecto de la inflación sobre los datos financieros que se manejan en el presente proyecto, se utiliza el Índice Nacional de Precios al Consumidor (INPC) para los años 2004, 2005 y 2006, tomando como base de transformación el mes de Noviembre de 2003 que corresponde al mes inicial del primer año de estudio según los calendarios fiscales de la empresa.

Así mismo, el análisis estadístico de los datos se realiza a través de la aplicación de la herramienta estadística de regresión lineal múltiple.

1.8. Alcance y limitaciones

1.8.1. Características de la muestra

La muestra fue seleccionada por medio de las recomendaciones del experto de la empresa donde se aplica el estudio, dado su nivel de experiencia y conocimiento sobre el tema.

1.8.2. Área geográfica para el estudio

El tema de tesis que se desarrolla en el presente documento ha sido inspirado en la problemática real de una cadena de supermercados localizados en la ciudad de Monterrey, Nuevo León, y su área metropolitana.

1.8.3. Limitaciones no definidas por el investigador

Por cuestiones de políticas internas de la empresa, fue necesario tomar en cuenta suposiciones para simplificar la definición de las variables que se incluyen en los modelos de regresión. Así mismo, el reglamento de confidencialidad de compañía indica que no es posible presentar en la tesis información financiera clave de sus procesos.

El tamaño de muestra seleccionada para el estudio fue limitado por la disponibilidad de datos en la empresa.

Los resultados discutidos en el Capítulo 4 permiten definir las posibles extensiones de investigación para futuros estudios en el tema.

Capítulo 2. Marco Teórico

2.1. Introducción

La intensa competencia a la que se enfrentan las empresas de autoservicio detallistas ha provocado la necesidad de desarrollar alternativas y estrategias que mejoren los niveles de productividad en las operaciones, de modo tal que sea posible satisfacer eficientemente las crecientes demandas de los clientes.

La industria de las tiendas de comestibles consiste en establecimientos que venden una amplia variedad de alimentos. Existen dos tipos principales de establecimientos: supermercados y tiendas de conveniencia, las cuales abarcan más de la mitad de las ventas de artículos comestibles (Budd y McCall, 2001).

El enfoque principal del presente capítulo es realizar una revisión de la literatura que corresponde al tema de la venta detallista de alimentos en cadenas de tiendas de autoservicio. La clasificación principal de este tipo de comercios en la ciudad de Monterrey y área metropolitana se define como: (1) tiendas de conveniencia - locales con menos de 1,524 metros cuadrados de superficie, operan durante largas horas de comercio, incluyendo domingos, ofrecen una amplia variedad de productos alimenticios y de abarrotos, (2) supermercados o tiendas grandes - tiendas de autoservicio de mercancías comestibles, entre 1,524 y 7,620 metros cuadrados, (3) supertienda - tiendas de autoservicio con más de 7,620 metros cuadrados, e (4) hipermercado - tiendas tipo mercado (*outlet*) con un área de venta mayor a 15,240 metros cuadrados, cuyo negocio central son usualmente los alimentos y bebidas (Browell e Ivers, 1998).

La venta detallista es un servicio basado en un proceso de compra donde un cliente que entra a una tienda no sólo procura buenos productos a un precio justo, sino que también busca un buen servicio (Koenes, 1996).

Un factor clave de éxito a largo plazo para una empresa detallista es el complementar las ventas con un sistema de servicios que logre satisfacer a sus clientes para que estos compren en repetidas ocasiones (Koenes, 1996).

2.2. Los cambios en las operaciones de la industria de los alimentos

La industria detallista de alimentos emplea aproximadamente tres millones de trabajadores en Estados Unidos. El mercado en dicho país se encuentra dominado principalmente por grandes cadenas de supermercados como lo es Walt-Mart, Kroger, Albertson's, Safeway y Ahold USA (Budd y McCall, 2001).

El alzamiento del trabajo de tiempo parcial en las tiendas de alimentos ocurrió durante 1950 y los primeros años de la década de 1960. Siendo el incremento en el nivel promedio de educación de los empleados un factor de relevancia en el desarrollo de la industria. Al inicio de 1970, las innovaciones tecnológicas permitieron el cambio en los empaques de algunos alimentos como lo es el caso de la carne empaquetada en cajas, el pollo y la carne de cerdo preempacada, hasta las recientes charolas de carne que son procesadas y envasadas desde la planta empacadora. Estas modificaciones ocasionaron la reducción significativa de la demanda de carniceros especializados en las tiendas de alimentos (Budd y McCall, 2001).

Durante las décadas de los 80's y 90's, la industria experimentó un crecimiento grande y diverso que hizo posible el manejo de los supermercados dentro de cadenas de operaciones. En el mismo periodo de tiempo se presentaron cambios significativos con respecto al incremento del área promedio total de la tienda de 23,829 a 34,735 pies cuadrados, el número de artículos en inventario (SKU) incrementó un 84%, las transacciones por semana aumentaron 34%, y número total de empleados de tiempo completo y parcial creció un 59% y 71% respectivamente (Budd y McCall, 2001).

Budd y McCall (2001), comentan que un cambio operacional que ha tomado lugar en los supermercados consiste en el incremento de las horas de servicio de la tienda, las cuales han aumentado de 95 a 125 horas de operación; en adición a la apertura de tiendas en domingo y a la fracción de sucursales abiertas 24 horas que han sido duplicadas.

La adopción de la tecnología de lector óptico (*scanner*) ha permitido la lectura de los símbolos de códigos universales del producto para registrar el precio de cada uno de ellos de forma automática. El uso de esta herramienta trajo consigo ahorros de 5% de los costos de operaciones, mientras que se incrementó la productividad y exactitud de los cajeros (Budd y McCall, 2001). Sin embargo, la utilización de tecnología provocó

que los sueldos bajaran, que se requirieran empleados con pocas habilidades, haciendo más fácil su reemplazo y por tanto reduciendo su poder de negociación.

2.3. La función de la mercadotecnia y la calidad del servicio en el sector detallista

La mercadotecnia tiene como objetivo el satisfacer las necesidades de los consumidores mientras genera beneficios para la empresa. Para lograrlo busca establecer vínculos de información que unen a la empresa con sus mercados y que resultan de utilidad para la toma de decisiones.

De acuerdo a Koenes (1996), la mercadotecnia es responsable de asegurar niveles óptimos de factores que van desde volúmenes de venta, frecuencia de las compras, lealtad de clientes, y participación en el mercado, hasta los niveles de rentabilidad y productividad de la organización.

Dentro de los mercados actuales, algunos de los retos más importantes a los que se enfrentan las organizaciones incluyen el incremento en las expectativas y demandas de los consumidores en términos de calidad en el servicio. Los esfuerzos por mantener dicha calidad ha permitido a las empresas líderes del sector ganar y mantener la lealtad de sus clientes (Wong y Sohal, 2003).

La calidad en el servicio puede definirse como el grado de discrepancia entre las expectativas de los consumidores ante un servicio, y sus percepciones con respecto al desempeño del mismo. Berry et al. (1985) afirman que los consumidores evalúan el nivel de calidad en el servicio que reciben en base a tres tipos de atributos: búsqueda, experiencia, y crédito.

Los atributos de búsqueda se refieren a la percepción de los clientes sobre las instalaciones físicas, la apariencia del personal y la imagen de los proveedores que ofrece la empresa. Los atributos de experiencia indican la capacidad para responder rápidamente a las solicitudes de asistencia, así como en el desempeño del servicio en el tiempo adecuado. Por último, los atributos de crédito tales como la seguridad de una inversión pueden no ser determinados aún después del uso repetido del servicio (Berry et al., 1985).

Parasuraman et al. (1988) sugieren una clasificación más detallada sobre calidad del servicio, las dimensiones que incluye son:

- (1) *Seguridad*: La habilidad para realizar el servicio prometido de forma confiable y precisa.
- (2) *Receptividad*: La buena voluntad para ayudar a los clientes y proveer un servicio rápido.
- (3) *Certidumbre*: La aptitud y credibilidad del sistema para proveer un servicio cortés y seguro.
- (4) *Empatía*: La aproximación, fácil acceso y el esfuerzo realizado para entender las necesidades de los clientes.
- (5) *Tangibles*: La apariencia física de las instalaciones, equipo, personas, y materiales de comunicación.

Para cualquier empresa que comercializa alimentos, la fidelidad de los clientes es un factor importante de éxito. La lealtad se refiere a una actitud favorable hacia la compra repetida de la marca de un producto o servicio definido (Wong y Sohal, 2003). Cuando un cliente tiene una experiencia negativa y se siente insatisfecho con el minorista, entonces cambiará de proveedor, en estos caso se puede concluir que los consumidores fieles son menos susceptibles a información negativa sobre un producto que los que no son leales (Xin-lu y Lin-du, 2005).

La fijación de los niveles de precios de venta al público, las rebajas y descuentos sobre los alimentos, son parte de la estrategia de mercadotecnia de la empresa (Minichiello, 1967). Las promociones basadas en el precio son predominantes en la actividad competitiva de los minoristas, las rebajas incrementan las ventas de la tienda a corto plazo debido a que los clientes compran los productos con descuento como extra a sus compras regulares, mientras que no lo adquirirían si el producto no estuviera en promoción (Walters y MacKenzie, 1988).

El nivel de servicio afecta la estructura de costos de un detallista, aquellos que ofrecen un buen servicio son menos propensos a presentar una imagen de bajos precios a través de promociones. Algunos vendedores de alimentos ofrecen descuentos para hacer frente a los cupones de los fabricantes mientras que otros minoristas tienden a utilizar cualquiera de las dos estrategias: bajos precios diariamente o promociones frecuentes (Reinartz et al., 2000).

2.4. Productividad laboral en la venta al detalle

El aumento en la productividad es un factor vital para lograr alcanzar altos estándares de vida en cualquier sociedad en desarrollo. Históricamente, ha ocurrido un repartimiento desproporcionado del crecimiento de la productividad hacia el sector productivo (Ingene, 1982). A medida que las economías se vuelven mayormente basadas en

servicios, es necesario que la productividad del mercado crezca rápidamente.

La productividad de la fuerza laboral en la industria detallista es de las más representativas en el sector de mercado debido al contacto directo que existe con los clientes. El personal de la organización tiene como propósito prestar un servicio de alta calidad que conduzca el proceso de la venta (Koenes, 1996). Los empleados tienen la responsabilidad de mostrar disposición de ayuda, así como su capacidad y actitud para brindar información cierta a los clientes.

Se puede definir en cualquier ámbito de la economía el concepto de productividad como la proporción entre el total de productos obtenidos en relación con el uso de los factores de producción (Ingene, 1982) Así pues, podemos decir que productividad se refiere a la capacidad y habilidad para lograr determinados objetivos mediante el uso eficiente de tiempo, esfuerzo y recursos (Koenes, 1996).

Los niveles de productos que se generan de la productividad pueden medirse en términos de transacciones realizadas, unidades físicas y ventas (Ingene, 1982).

Existen dificultades prácticas con respecto al uso de las transacciones como unidad de medición, dado que en una tienda de abarrotes pueden realizarse diversas mezclas de adquisiciones de productos, las cuales representan transacciones independientes de menor o mayor magnitud según la capacidad de la tienda de conveniencia o supermercado, respectivamente (Ingene, 1982).

Carey y Otto (1977) han escrito que el mejor indicador de producción para las industrias de comercio minorista se obtiene de datos de unidades físicas; es decir, sobre las cantidades de los bienes y servicios vendidos por la industria, cada uno de ellos ponderado por medio de las horas de trabajo efectivo requeridos para vender una unidad de producto en un periodo de tiempo base específico.

Una alternativa es el uso de ventas como una medición de las salidas de productividad. Las ventas son equivalentes a las unidades físicas pero contabilizadas por medio de precios. Según Ingene (1982), la industria minorista es uno de los sectores de la economía que refleja la existencia de equilibrio a largo plazo, por tanto el uso de pesos como medida no genera distorsiones significativas en la medición de la productividad laboral cuando los costos de nómina son altos en relación con los costos

totales. En el área de minoristas de alimentos, aproximadamente el 60% del total corresponde a costos de operación.

Las ventas como unidad de medición permiten observar el estado financiero de la empresa. En la tabla 2.1. se muestra que las ventas de los supermercados incrementaron un 5.7%, siendo la tasa de crecimiento más alta desde 1989. Las ventas reales experimentaron una inflación en los alimentos de un 3.8%. Por otro lado, las ventas de las cadenas subieron 6.6%, mientras que el surtido limitado saltó a 15.3%. Las tiendas de conveniencia mostraron un comportamiento plano, y el número de supermercados creció de 411 al récord de 34,252 (McTaggart y Heller, 2005).

	No. de tiendas	% del total	Ventas (\$ millones)	% del total
Supermercados (\$2 millones o más)	34,252	100%	\$457,414	100%
Cadena (11 o más tiendas)	22,253	65.6	386,356	84.5
\$2-4 millones	1,237	3.6	3,587	0.8
\$4-8 millones	4,418	12.9	26,804	5.9
\$8-12 millones	3,845	11.2	37,297	8.2
\$12-20 millones	5,531	16.1	85,169	18.6
\$20-30 millones	4,215	12.3	99,848	21.8
\$30-40 millones	1,458	4.3	49,754	10.9
\$40-50 millones	925	2.7	40,850	8.9
\$50 millones en adelante	824	2.4	43,048	9.4
Independiente (10 o menos tiendas)	11,799	34.4	71,058	15.5
\$2-4 millones	3,644	10.6	10,568	2.3
\$4-8 millones	6,372	18.6	35,925	7.9
\$8-12 millones	1,027	3	9,962	2.2
\$12-20 millones	533	1.6	7,761	1.7
\$20-30 millones	132	0.4	3,112	0.7
\$30-40 millones	48	0.1	1,637	0.4
\$40-50 millones	28	0.1	1,230	0.3
\$50 millones en adelante	15	0	864	0.2
Tiendas de conveniencia	138,205		114,009	
Supertienda	25,205		13,205	

Tabla 2.1. Ventas de supermercados 2004 (McTaggart y Heller, 2005).

La medida para las entradas más adecuada en productividad laboral son las horas de trabajo efectivo utilizadas. Dado que los empleados trabajan a diferentes niveles de eficiencia, las horas de labor por trabajador comúnmente se clasifican como parciales o completos, es decir que se desempeñan turnos de trabajo de 24 ó 48 horas semanales respectivamente. Cuando se presenta el caso de ausencia de información, el mejor sustituto de medición es considerar el número de empleados activos (Browell e Ivers, 1998).

Como se ha mencionado, la creación de salidas requiere de la interacción de un conjunto de factores controlados, los cuatro mayores componentes que afectan la conversión de entradas en salidas son: el nivel de tecnología, la cantidad de capital relativo a la nómina, la eficiencia alcanzada y el grado de las economías a escala (Browell e Ivers, 1998).

Las innovaciones de tecnología así como mejoras en la distribución y manejos de productos ha contribuido en la productividad del sector minorista (Bucklin, 1980). A medida que la cantidad de nómina incrementa, la productividad laboral es generalmente realizada debido al desempeño del capital como parte del trabajo.

Bucklin (1980) comenta que la introducción de cajas registradoras electrónicas en conjunto con el uso de códigos de barras universales ha mejorado la productividad laboral de muchas empresas de venta al detalle mediante la agilización del registro de mercancías, control de inventarios y disminución en el tiempo requerido para etiquetar los productos con el precio correspondiente. Por otro lado, la eficiencia alcanzada hace referencia a la habilidad de utilizar la cantidad correcta de trabajo, capital y otros factores de mercadotecnia.

Según Ingene (1982), la productividad en las tiendas de comestibles ha crecido rápidamente en las últimas décadas debido a seis factores: (1) el incremento en el tamaño de las transacciones medidas por número y valor de las mercancías, (2) el cambio al autoservicio, (3) la utilización completa de la capacidad existente, (4) el incremento en el abastecimiento, (5) al crecimiento histórico del tamaño de la tienda que cede por las economías a escala, y (6) el incremento en la calidad del servicio ofrecido por los trabajadores.

El éxito en la industria de los alimentos con estándares de productividad esta impulsando a otras industrias detallistas y manufactureras a la adopción de la misma estrategia. Esto debido a que cuentan con características similares a las de la industria de comestibles como lo son: altos volúmenes, gran cantidad de embarques, y pequeños márgenes de operación (Cooke, 2003).

2.5. Empleados de tiempo parcial y completo en la industria detallista de alimentos

La flexibilidad demandada en el contexto de productividad y competitividad de los negocios actuales, ha incrementado la necesidad de emplear personas capaces de enfrentar asertivamente los cambios continuos del mercado.

Los negocios detallistas de alimentos se enfrentan diariamente ante demandas inciertas, largas horas de servicio, y cambios drásticos en el comportamiento de compra de los clientes. Dichas circunstancias apuntan a la búsqueda de mayor flexibilidad, a la reducción de costos, al aumento de la productividad, y al incremento de la demanda de trabajadores de tiempo parcial (Browell e Ivers, 1998).

Un trabajador de tiempo parcial es aquel que labora menos de 30 horas por semana, y un empleado de tiempo completo trabaja mínimo 48 horas a la semana. Browell e Ivers (1998) comentan que las razones por las cuales se emplean trabajadores parciales son las siguientes:

- (1) carencia de mano de obra;
- (2) flexibilidad;
- (3) el hecho que el trabajo no requiera atención de tiempo completo;
- (4) demanda de empleados;
- (5) horas de operación extendidas.

Actualmente, la mano de obra flexible de los empleados de medio tiempo es crucial para responder a la creciente demanda de los clientes. La flexibilidad es por tanto vista como uno de los elementos más significativos en términos de competitividad en las industrias. Se considera que una compañía es competitiva cuando es capaz de responder rápidamente a las alteraciones en la demanda de los productos y servicios que ofrece (Browell e Ivers, 1998).

Según Browell e Ivers (1998), dentro de la venta al detalle la proporción de empleados de tiempo parcial en relación al personal de tiempo completo es aproximadamente de 72 a 28, mientras que más del 60% de los empleados en el sector minorista de alimentos son empleados de tiempo parcial.

La eficiencia en el comercio minorista se considera como la utilización efectiva de la fuerza laboral para satisfacer el comportamiento irregular de la demanda de los clientes a diversos periodos de tiempo. Se

considera que dos tercios de las ventas de alimentos semanales toman lugar entre los días jueves y sábado, con picos diarios entre las 12 p.m. y las 2 p.m. La importancia del comercio en el día domingo se ha incrementado rápidamente, así como las ventas siete días a la semana las 24 horas del día. Por esto, las prácticas laborales orientadas a la flexibilidad hacen posible que los empleados sean capaces de alcanzar el nivel máximo de la demanda (Browell e Ivers, 1998).

Los trabajadores de tiempo completo disfrutan de beneficios y mejores puestos, particularmente en posiciones gerenciales y de supervisión. Por otro lado, los empleados parciales por lo general no son acreedores de diversas ventajas que puede ofrecer una compañía, entre ellos se encuentran los descuentos a trabajadores, el derecho a descansar en días festivos, pagos por incapacidad o enfermedad, entre otros (Browell e Ivers, 1998).

En cuanto a términos de productividad, los empleados de tiempo completo ofrecen mayor eficiencia, energía y entusiasmo. Sin embargo, de acuerdo a Browell e Ivers (1998) los empleados de medio tiempo son más productivos, generan nuevas ideas, e incrementan la productividad. Por otro lado, Geber (1994) comenta lo contrario al decir que la productividad es menor en los empleados parciales debido a que no ofrecen la misma dedicación, habilidades y compromiso como los trabajadores de tiempo completo; por consiguiente, la calidad de su trabajo es mala debido a su escasa experiencia o entrenamiento.

Koenes (1996) argumenta que para lograr que el personal de las empresas minoristas establezca contacto con la clientela, deben ser entrenados en los conocimientos sobre la organización, sus productos y las técnicas para potenciar la capacidad para la venta. Por otro lado, es importante tomar en consideración la motivación del personal, la cual se logra creando un ambiente laboral positivo en que se establezcan objetivos y metas que enriquezcan el trabajo que realizan los vendedores.

Dubinsky et al. (2004), afirman que la mala conducta de los empleados dedicados a las ventas ocurre debido a factores tales como la deficiencia de los programas de capacitación, las constantes presiones a los que son sometidos para alcanzar los objetivos de la empresa, a la supervisión constante, y a causa de un inadecuado sistema de soporte laboral.

Los negocios comercializadores de alimentos emplean predominantemente trabajadores de medio tiempo no calificados y con habilidades parciales. La investigación conducida por Browell e Ivers

(1998), muestra que los empleados parciales constituyen a lo más el 82% del total de empleados a pesar de que la proporción promedio era un 63% de empleados parciales y 37% de empleados completos.

Del mismo estudio se concluye que las principales razones por las cuales se emplean en los supermercados trabajadores de medio tiempo son: la flexibilidad que ofrecen para cubrir los patrones de demanda a la alza, para satisfacer el suministro y demanda de mano de obra, para generar mayor productividad, y beneficiar a la empresa con respecto a la reducción de costos de nómina (Browell e Ivers, 1998).

Finalmente, es posible concluir que las tiendas de comercialización de comestibles no emplean trabajadores de medio tiempo debido a escasez de empleados de tiempo completo o de personal experto, ni para reducir los gastos organizacionales; sino por razones comerciales relacionadas con la flexibilidad requerida dentro de la organización.

2.6. Venta por hora hombre en el comercio detallista

La medición con mayor utilidad en el estudio de la productividad en la venta al detalle es el indicador de ventas por hora hombre. En las tiendas de comestibles de distintos tamaños, la ventas por persona pueden medirse de acuerdo a la clasificación de empleados de tiempo parcial y tiempo completo (McClelland, 1958). De acuerdo a Ingene (1982), siempre que sea posible el número de empleados debe expresarse en términos de trabajadores de tiempo completo.

Los estudios realizados por McClelland (1958) muestran que las tiendas pequeñas tienen una menor cantidad de ventas por hora hombre en comparación con las grandes. El promedio de ventas por trabajador es mayor para productos alimenticios que para productos como ropa, muebles, hardware, entre otros.

2.7. El uso de tecnología en la venta al por menor

En décadas pasadas, la industria de las tiendas de alimentos experimentó un dramático incremento en el uso de tecnología, así como el crecimiento en el tamaño y rango de productos y servicios que ofrecen.

La tecnología de información tiene un impacto en la eficiencia total del negocio de los minoristas, y afecta las operaciones que van desde la compra en línea, hasta la obtención del producto y la administración de

cadena de suministro. Además, a medida que los minoristas globalizan sus operaciones, la ventaja competitiva en el futuro no será entre compañías, si no a través de la capacidad de las redes de abastecimiento (Yan et al., 2004).

Los vendedores al detalle de comestibles utilizan la tecnología para mejorar la eficiencia y el servicio al cliente. Más de 4 de 10 minoristas ofrecen compras en línea, cerca de 91.5% utilizan datos escaneados para medir el impacto de las promociones y 8 de 10 manejan datos de compra frecuentes para enviar correos electrónicos. Además, 55% usa computadoras para programar las actividades de trabajo (Parks, 2002).

De acuerdo a Parks (2002), las contribuciones en tecnología y eficiencia en la industria de los alimentos en Estados Unidos, ha incrementado las utilidades después de impuestos en un 1.25% de las ventas para el 2001. Por otro lado, dichas iniciativas han impulsado el aumento en dos indicadores de productividad claves: ventas semanales por metro cuadrado y ventas por hora hombre.

Por otro lado, el uso de tecnología en el sector ha permitido mejorar la seguridad de los alimentos mediante la aplicación de métodos de conservación como lo es la irradiación. También se monitorean las temperaturas de los alimentos perecederos para asegurar la seguridad del producto. Más de 7 de 10 detallistas controlan el contacto de los comestibles con las superficies en la tienda (Parks, 2002).

El entrenamiento de los empleados en el manejo de los alimentos es fundamental para asegurar la calidad de los mismos. Entre 6 y 10 minoristas educan a los responsables del manejo de la comida sobre el uso de la metodología basada en el Análisis de Riesgos y Puntos Críticos de Control (HACCP por sus siglas en inglés) (Parks, 2002). El HACCP es un sistema de seguridad para analizar, identificar, evaluar y controlar los riesgos en los alimentos.

El garantizar la seguridad de los alimentos es clave para promover un nivel de servicio adecuado que satisfaga las expectativas de los clientes. Xin-lu y Lin-du (2005) sugieren el uso de un Sistema de Crédito de Seguridad Alimenticia (CSFS por sus siglas en inglés), el cual consiste en una base de datos cuyos principios se basan en: (1) introducir y editar cambios en la descripción, pesos y datos de seguridad de los alimentos en una forma confiable y eficiente; (2) organizar tareas complicadas y verificar los aspectos de control de calidad que afectan la base de datos en el sistema en línea; (3) proveer a los administradores acceso rápido a la base de datos para obtener información estadística e histórica que

permita tomar decisiones; (4) tener acceso y fácil uso de los datos por parte de los consumidores; y (5) tener un formato modular de modo que el sistema pueda mejorarse continuamente sobre el tiempo.

El desarrollo de los estándares de operaciones en centros de distribución permite el uso de software para medir la productividad de la mano de obra. La introducción de tecnología como sistemas de radio frecuencia (RF) y códigos de barras hacen posible el rastreo de las actividades de trabajo (Cooke, 2003), permitiendo con ello que los operadores identifiquen los productos de forma automatizada con el uso de computadoras, haciendo posible el rastreo de los alimentos desde el almacén hasta la mesa del consumidor final.

2.8. Desempeño de las tiendas de alimentos

Las operaciones de la industria se encuentran en un proceso de expansión hacia mercados internacionales como una estrategia de crecimiento. La administración estratégica permite a las compañías minoristas obtener ventajas competitivas con respecto al concepto de distribución al detalle y logística. El interés en la expansión internacional indica que el 41% de 200 minoristas de los Estados Unidos llevan a cabo negocios internacionalmente en comparación con Japón y más de la mitad de detallistas europeos (Vida et al. 2000).

Tradicionalmente, se han identificado cuatro factores que tienen influencia sobre el desempeño de un comercio de autoservicio: la tienda, el mercado, las características del consumidor y la competencia (Reinartz et al., 2000).

Por otro lado, Silva y Cardoso (2005) consideran que existen incontables factores que pueden relacionarse con el desempeño de una tienda. Éstos varían desde características intrínsecas de la sucursal, así como las características de la zona geográfica de influencia hasta las relaciones sociales entre el consumidor y la marca de la cadena de supermercados. En cuanto a los factores concernientes a la influencia por zona de la tienda, la presencia de competidores puede indicar una locación con un alto potencial económico y poder de compra que afecta positivamente el rendimiento de la tienda.

Otros factores como el grado de urbanización de la zona, cantidad de habitantes, antecedentes étnicos, e ingresos familiares, tienen una influencia importante sobre el comportamiento de compra (Silva y Cardoso, 2005).

En relación con las características intrínsecas, se observa una relación entre el rendimiento de la tienda y su ambiente interno, nivel de servicio, nivel de variedad, y las políticas de promociones del supermercado (Silva y Cardoso, 2005).

La relación social entre el consumidor y la imagen de la marca tienen influencia sobre las ventas y el comportamiento de compra de los clientes. Otras variables como visibilidad y accesibilidad son también importantes para predecir ventas (Silva y Cardoso, 2005).

La nueva tendencia de las tiendas de alimentos consiste en agregar una mezcla de productos en su inventario. Las tiendas están agregando servicios adicionales, principalmente servicios alimenticios, así como nuevas categorías de mercancías de abarrotes. Ejemplos de las mezclas que realizan son la incursión de venta de dulces, farmacias, renta de videos, o barras de desayunos (Reinartz et al., 2000).

El rendimiento de una tienda puede medirse en ventas, porción del mercado, clientela, patrocinio, tráfico de la tienda, y utilidades. Reinartz et al. (2000) diferencian entre tres distintas mediciones para el desempeño de una tienda: (1) desempeño basado en el mercado como ventas o repartición del mercado, (2) desempeño basado en productividad como ventas por metro cuadrado, y (3) desempeño basado en utilidades como el margen de ganancia o el retorno sobre los activos.

El posicionamiento de un detallista se comunica a los clientes a través de los productos y servicios, ofreciendo estrategia y características de la tienda. Los vendedores tienen que tomar decisiones en precio, promoción, nivel de servicio, variedad de productos, tipo de mercancías y conveniencia (Reinartz et al., 2000).

Evidencia de estudios recientes sugiere que las características socioeconómicas y raciales de una comunidad tienen un efecto importante en la salud de su población. La segregación residencial medida en ingresos per capita y las diferencias étnicas, producen que el consumo de alimentos saludables como frutas y vegetales frescos, productos lácteos bajos en grasa, y cereales, sean relativamente más costosos y estén a menor alcance de las personas que viven en localidades de pobreza (Moore y Diez, 2006).

Moore y Diez (2006), condujeron un estudio donde se incluyeron un total de 3,337 tiendas de alimentos y licores que fueron clasificadas como: tiendas de comestibles y supermercados, tiendas de conveniencia,

mercados de carnes y pescados, mercados de frutas y vegetales, panaderías, tiendas de alimentos naturales, comercios de alimentos especiales, y sucursales de venta de licores.

Los supermercados se diferencian de las tiendas de conveniencia en base al reconocimiento del nombre de la cadena al que pertenece, además de una nómina anual mayor a 50 empleados (Moore y Diez, 2006).

A pesar de las similitudes en el número total de tiendas por población, la distribución de los tipos de tienda varía de acuerdo al nivel económico de las regiones bajo estudio. Las tiendas de comestibles se encuentran en mayor cantidad en los sectores pobres, en cambio los supermercados se localizan predominantemente en comunidades de niveles socioeconómicos medio a superior. Las tiendas de alimentos naturales y especiales, así como las tiendas de conveniencia se encuentran comúnmente en sitios donde viven personas estables económicamente hablando. En contraste, los mercados de carnes, pescados, frutas y vegetales, y panaderías son más comunes en vecindarios donde predominan diversos grupos étnicos o raciales de bajos ingresos (Moore y Diez, 2006).

Los resultados del análisis realizado por Moore y Diez (2006), muestran que las comunidades con ingresos bajos tienen cuatro veces más tiendas de alimentos por población, y la mitad de supermercados en comparación con los vecindarios acaudalados. Las tiendas de licores son más comunes en áreas pobres que en comunidades con altos ingresos por persona.

La extensa variedad de alimentos que se ofrecen en tiendas de comestibles y supermercados juegan un papel importante en la dieta de la población. Actualmente existe una demanda de productos que funcionen como sustitutos de la carne, este fenómeno ha ocasionado un incremento considerable en el consumo de pescados y mariscos como una alternativa viable (Duff, 2001).

Diversas cadenas de supermercados como lo es Wal-Mart han llevado a cabo campañas de información donde se comunica a los clientes sobre las ventajas a la salud derivadas del consumo de productos del mar. Los vendedores del departamento de pescados y mariscos son entrenados sobre los factores nutrimentales de estos productos, de modo tal que sea posible hacer recomendaciones a los consumidores sobre los beneficios a la salud y la preparación de estos alimentos (Duff, 2001).

Como se ha mencionado anteriormente, la productividad de un supermercado puede medirse en base a indicadores de desempeño como los que se muestran en la tabla 2.2. (McTaggart y Heller, 2005).

	2002	2003	2004
Supermercado promedio			
Área de venta (ft ²)	28,838	30,194	31,245
Volumen (\$ millones)	12.49	12.79	13.35
Número de transacciones	9.2	8.9	9
Ventas promedio semanales			
\$ por tienda	240,000	245,943	256,730
\$ por transacción	26,140	27,505	28,414
\$ por hora hombre	86.32	91.23	93.25
\$ por familia	73.49	77.25	80
\$ per capita	27.61	29	30.02
\$ por pie cuadrado	8.33	8.15	8.22

Tabla 2.2. Mediciones de desempeño de supermercados 2004.

2.9. La reinención de los supermercados

Una empresa exitosa es aquella capaz de responder con mayor eficacia que sus competidores ante las exigencias, necesidades, y expectativas del mercado en el que operan (Koenes, 1996).

Las tiendas de comestibles deben ser revolucionarias y estar a la vanguardia de los cambios en el mercado. La razón es simple, para la mayoría de los compradores, los alimentos son artículos básicos que pueden ser adquiridos en cualquier lugar, por tanto lo que realmente importa es la conveniencia y el precio (Mullin, 2005).

Un elemento del que deben disponer los supermercados para crear el posicionamiento que desea se relaciona con el producto o servicio que ofrece, los cuales deben ser vendidos en buen estado de conservación, organizados para facilitar su elección y agrupados en categorías (Koenes, 1996).

Las cadenas más grandes de supermercados se encuentran trabajando en la generación de productos y programas innovadores. Muchos supermercados mezclan una gran variedad de mercancías en sus pasillos, mientras que otros ofrecen precios bajos por medio de marcas propias (Mullin, 2005). La oferta de marcas propias permite incrementar los márgenes de ganancia de la empresa mientras aumenta la lealtad de los clientes.

El precio como factor crítico es un determinante importante sobre el sitio donde los clientes compran sus alimentos. Los departamentos de alimentos frescos y perecederos representan el centro de las operaciones de los supermercados debido a que estos manejan la mayoría de las ventas y tráfico totales de la tienda (McTaggart y Heller, 2005).

McTaggart y Heller (2005), comentan que la industria de los alimentos es madura, utiliza el mismo formato bajo el cual fue diseñado hace 30 ó 40 años. Por esto uno de los mayores retos y oportunidades corresponde a la reinversión del modelo de negocio para hacer más eficientes los departamentos centrales de un supermercado. Una estrategia involucra el recorte del número total de unidades en almacén, para enfocarse en la promoción de las marcas líderes y en las privadas, y venderlas a bajos precios para aumentar las ventas por metro cuadrado.

El cambio de formato de los supermercados se enfoca en el ofrecimiento de alimentos frescos, mejor presentación de los mismos y servicio personalizado. A medida que la nutrición y el consumo de alimentos saludables se vuelve más popular, los minoristas han enfatizado la importancia de incluir dentro de los pasillos de las tiendas alimentos naturales y orgánicos, así como mercancías no comestibles (McTaggart y Heller, 2005).

McTaggart y Heller (2005), muestran un reporte de las 45 razones por las que las personas seleccionan un supermercado para llevar a cabo sus compras, el índice fue calculado en base a una escala de 0 a 100, donde 0 significa que no es importante y 100 es extremadamente importante. Los resultados indican que la limpieza y los bajos costos son las prioridades de los clientes. Así mismo, el énfasis en productos perecederos ha sido la táctica defensiva de los supermercados para mantener a sus clientes. La tabla 2.3. muestra como el contar con un buen departamento de carnes se ubica en el lugar 13, seguido por tener un buen departamento de delicatessen en el 29, tener dentro de la tienda panadería en el sitio 33, y ofrecer un buen departamento de pescados y mariscos en la posición 38.

Ranking	Atributo	Índice
1	Limpieza	93
2	Bajos precios	90
3	Preciso escaneo de precios	90
4	Todos los precios claramente etiquetados	90
5	Empleados de inspección agradables y precisos	88
6	Productos marcados con fecha de caducidad	86
7	Buen departamento de productos agrícolas	86
8	Localización conveniente de la tienda	85
9	Anaqueles abastecidos adecuadamente	84
10	Layout adecuado para la compra rápida y fácil	83
11	Tiempo de fila corto en cajas registradoras	83
12	Ventas de especiales frecuentes	83
13	Buen departamento de carnes	83
14	Personal servicial en el departamento de servicios	82
15	Tener mercancías en existencia disponibles en especiales	82
16	Señalización en anaqueles del precio del producto	82
17	Buen departamento de lácteos	79
18	Habilidad para realizar todas las compras en la misma tienda	77
19	Los nuevos productos promocionados están disponibles	76
20	Selección de productos de marca propia a bajo costo	75
21	La tienda acepta tarjetas de crédito y débito	74
22	La tienda usualmente no se encuentra saturada	73
23	Buen departamento de alimentos congelados	72
24	Ofrecen cupones de doble valor	71
25	Selección de marcas nacionales anunciadas	71
26	El gerente es amigable y servicial	70
27	Ambiente y decoración agradable	70
28	Empacadores disponibles	69
29	Buen departamento de delicatessen	67
30	La tienda tiene un programa de cliente frecuente	67
31	Información nutricional disponible	67
32	Abierto hasta horas tardías	66
33	Tienen panadería dentro de la tienda	64
34	Selección de genéricos (productos sin marca)	62
35	Productos amigables con el medio ambiente	59
36	Surtido de productos no alimenticios	59
37	Sección de artículos de tocador y medicamentos	58
38	Pescados y mariscos frescos	52
39	Servicio de registro y cobranza	48
40	Tiene farmacia	45
41	Ofrecen quesos de especialidad/gourmet	36
42	Disponibilidad de alimentos naturales/orgánicos	36
43	Disponibilidad de alimentos étnicos	36
44	Venta de alimentos preparados	33
45	Renta de videos disponible	13

Tabla 2.3. 45 razones para seleccionar un supermercado.

2.10. Motivos de los consumidores en la selección de alimentos y tiendas de comestibles

La sociedad actual vive una etapa de madurez, en el que el envejecimiento de la población causará cambios en diversas áreas de nuestras vidas. Por consiguiente, es importante para los negocios fomentar un entendimiento sólido del mercado. La industria de las tiendas de alimentos no es la excepción, dado que los consumidores adultos tienden a gastar más dinero en promedio en las tiendas de comida que las personas jóvenes (Moschis et al., 2004).

La lealtad hacia un supermercado en particular es alta entre los grupos de adultos entre 45 y 64 años de edad, ya que de acuerdo a estudios realizados por Moschis et al. (2004), corresponden al conjunto de personas que pasan mayor tiempo en las tiendas de alimentos y que gastan 10% más en comida que el promedio nacional de los Estados Unidos.

Existen diversos factores y eventos que afectan las preferencias o motivos por los cuales se selecciona una tienda de comestibles. La investigación conducida por Moschis et al. (2004) comparó las razones por la que individuos jóvenes menores a 55 años de edad y personas de 55 o más años, prefieren ser clientes de una específica tienda de conveniencia.

El estudio tomó una muestra aleatoria de 1,950 personas contempladas en siete categorías de edades: 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, y mayores de 85. Los motivos por los que son clientes de una determinada tienda de alimentos fueron enlistados en 14 razones de selección. Los resultados se muestran en la tabla 2.4.

Razones	Menor a 55 (%)	Mayor a 55 (%)
Facilidad para localizar mercancías/productos	85.66	87.71
Políticas de devolución de mercancías	40.33	39.74
Localización cercana al lugar de trabajo o vivienda	87.81	84.31
Frecuencia de promociones, descuentos y rebajas	70.74	76.67
Ofertas especiales a clientes de cierta edad	9.87	28.65
Tener productos apropiados para el cuidado de la salud	37.11	40.32
Tener personal capacitado para asistir a los clientes	32.4	40.79
Preferencia por un método de pago	10.09	10.25
Tener cajas registradoras rápidas	78.57	80.21
Ofrecer servicios de atención especial	26.01	38.55
Recomendado por personas de su edad	13.7	18.27
Venta de marcas o productos familiares	77.37	80.37
Cercanía con otros establecimientos de preferencia	59.05	64.44
Lugar confortable para socializar	52.26	54.75

Tabla 2.4. Razones de sección de una tienda de alimentos entre jóvenes (menores a 55) y adultos (mayores a 55).

Diversos factores socioeconómicos y demográficos afectan la percepción de los clientes en relación a la comida y las tiendas de alimentos, entre ellos encontramos (Moschis et al., 2004):

- (1) *Edad*: La percepción de los 14 factores en la selección de tiendas de alimentos cambia ligeramente con respecto a la edad. La tabla 2.4. muestra que la importancia de la localización de la tienda en relación con el hogar o el lugar de trabajo permanece siendo un motivo relevante en la selección de la tienda; además los descuentos empiezan a ser importantes a medida que incrementa la edad, y finalmente la influencia de las recomendaciones de tienda por parte de personas de la misma edad tiende a aumentar a medida que las personas crecen.
- (2) *Sexo*: Muchos factores examinados son de igual importancia para hombres y mujeres; sin embargo, las mujeres perciben varios factores como más importantes que los hombres, como lo es el caso de la flexibilidad de las políticas de devolución de mercancías (42% contra 37%), descuentos especiales (79.3% contra 73.4%), habilidad de los empleados para brindar ayuda (43.3 % contra 37.9%), localización cerca de sitios de donde se es cliente (64.7% contra 59.6%), así como que la tienda de comestibles sea un lugar confortable para realizar las compras y socializar (57.2% contra 51.4%).
- (3) *Socioeconómico*: A medida que el ingreso de la persona incrementa, la importancia que le brinda a los factores disminuye. Solamente tres de los factores parecen seguir siendo de interés: localización de la tienda cerca de la vivienda, disponibilidad de productos que satisfagan las necesidades de salud, y que existan cajas registradoras rápidas.
- (4) *Localización*: Las personas que viven en áreas rurales consideran que los descuentos son muy importantes en la selección de la tienda donde realizarán las compras, así como la asistencia del personal. Los clientes de áreas urbanas en comparación con los de zonas rurales consideran que lo más relevante es la cercanía de la tienda a su casa o trabajo.

Cada persona tiene necesidades individuales diferentes, por esta razón es que los supermercados procuran ofrecer a sus clientes todo tipo de productos y servicios bajo un mismo techo. A continuación se presentan los resultados de una encuesta realizada en el 2004 donde se muestran

como una persona gasta 100 dólares americanos en un supermercado (Moschis et al., 2004).

ABARROTES (TOTAL)	\$39.45
Comestibles, comida	31.08
Bebidas, alcohol	3.62
Artículos para la comida principal	7.58
Refrigerios, dulces	6.33
Condimentos	5.42
Productos no alimenticios	8.37
PERECEDEROS (TOTAL)	50
Panadería	2.15
Productos horneados (frescos)	3.26
Delicatessen, en tienda (servicio)	3.47
Delicatessen (autoservicio)	1.12
Alimentos congelados	7.29
Carne, pescados, y aves (frescos)	13.33
Frutas y vegetales	9.86
Flores	9.86
MERCANCÍA GENERAL (TOTAL)	3.81
MG (rastreada)	1.52
MG (no rastreada)	2.3
SALUD Y BELLEZA	3.83
FARMACIA	3.83
GRAN TOTAL	\$100

Tabla 2.5. Como gastar 100 dólares en un supermercado.

Las recomendaciones para las tiendas de alimentos consisten en desarrollar estrategias de mercado que busquen fomentar la lealtad de los clientes al ofrecer marcas y productos conocidos, así como brindar mercancías apropiadas para cubrir las necesidades de salud y nutrición. Por otro lado, la asistencia especial de servicios como *valet parking*, asistencia en entregas y apoyo para cargar los paquetes sería de agrado para muchos clientes de tiendas de abarrotes y supermercados. Los minoristas de alimentos deben entrenar a sus trabajadores para ser corteses y capaces de memorizar las necesidades y preferencias de productos de sus clientes más frecuentes (Moschis et al., 2004).

2.11. El efecto de la inflación y el índice de precios en las ventas minoristas de alimentos

Thuesen y Fabrycky (2001) definen que inflación es un término que describe los cambios en los niveles de precios en una economía. El Banco de México (2002) indica que se debe entender por inflación a la tasa de crecimiento promedio de los precios de la canasta de bienes y servicios de un periodo a otro. La medición de los cambios históricos de nivel de precio para mercancías se calcula por medio de un índice de precios.

Un índice de precios es una proporción de los precios históricos de algún bien o servicio en un punto en el tiempo en relación al precio en algún punto más temprano conocido como el año base (Thuesen y Fabrycky, 2001).

Según el Banco de México (2002), el Índice Nacional de Precios al Consumidor (INPC) es un indicador económico cuya finalidad es medir a través del tiempo la variación de los precios de una canasta de bienes y servicios representativa del consumo de los hogares mexicanos. El INPC se considera un indicador oficial de la inflación en México.

El INPC brinda información al gobierno, empresas, sindicatos y ciudadanos sobre los cambios que tiene el costo de vida en el país. Las variaciones del indicador son importantes para la revisión de los precios de diversos bienes y servicios de la economía (Banco de México, 2002).

La inflación en los precios de los alimentos ha causado mayor conciencia de los clientes sobre el consumo de comestibles. En términos de ventas, las tiendas de abarrotes corresponden al segundo mayor minorista, promediando alrededor del 13% de la fuerza laboral de dicho sector (Ingene, 1982).

Ingene (1982) argumenta que el crecimiento en la productividad de una economía es una fuente impulsora que permite alcanzar altos estándares de vida mediante el incremento en los salarios de los trabajadores y por medio de la resistencia ante el aumento de los precios de los alimentos debido a la inflación.

El abastecimiento de alimentos ocurre relativamente a corto plazo, cualquier expansión de la demanda genera presiones en los precios de las mercancías en las tiendas de comestibles. La curva teórica de la demanda no es perfectamente elástica, un cambio en la demanda puede incrementar el nivel de precio del producto en cuestión, así como

la cantidad que se toma del mercado. El impacto del precio relativo o la cantidad de producto dependerá de: (1) el cambio en la demanda, (2) la elasticidad de la demanda, y (3) la elasticidad del abastecimiento (Belongia, 1979).

Hoch et al. (1995) encontraron que la demografía del consumidor y los factores competitivos estiman el 67% de la varianza explicada en la elasticidad de precio a nivel de una tienda. Sin embargo, no incluyen factores específicos de la tienda tales como: tamaño, surtido, grado de mezcla, y nivel de servicio.

El estudio conducido por Belongia (1979) sobre el impacto de los precios en los alimentos básicos muestra que los mayores impactos en el costo se presentan en productos de cereales, panadería y carnes.

Capítulo 3. Metodología

3.1. Introducción

En el capítulo 3 se describen los pasos a seguir para el estudio de la información disponible y la aplicación e interpretación de la herramienta estadística seleccionada.

3.2. Metodología

Se presenta a continuación la metodología para la aplicación del análisis de regresión lineal múltiple.

3.2.1. Definición de los objetivos del análisis

El primer paso consiste en decidir sobre el área de aplicación del modelo de regresión. Esta decisión tiene un impacto importante en el diseño de la investigación y en la forma en que los modelos serán desarrollados.

Asumiendo que los objetivos del análisis de regresión consisten en la combinación de objetivos particulares de predicción (pronóstico de ventas) y explicativos (desempeño actual de indicadores de productividad de la tienda), se determina cuáles son los factores que pueden influir en el comportamiento de la variable dependiente que se busca analizar.

3.2.2. Determinación de variables

El segundo paso consiste en el desarrollo de ideas sobre las variables que probablemente pueden tener influencia sobre las ventas de la tienda. Las ideas se traducen en variables para su medición y la formación de una muestra de datos. En este paso, la omisión de alguna variable importante, puede resultar en la obtención de un modelo de predicción de ventas poco exacto.

La selección de los factores que han de ser incluidos en el análisis de regresión puede ser determinada por medio de la aportación de un experto que conozca a detalle el comportamiento del proceso o fenómeno que se observa. Asimismo, el analista puede elegir a su criterio los regresores que sean de interés para el estudio, de modo tal que sea posible conocer el efecto de las variables independientes sobre la variable de respuesta.

3.2.3. Selección de la muestra y recolección de datos

La selección de una muestra representativa de las tiendas existentes es el tercer paso. La muestra debe reflejar el objetivo del análisis. Por ejemplo, Rogers (1992) define que si el objetivo es el desarrollo de pronósticos de ventas para supermercados cuyo tamaño sobrepase los 40,000 pies cuadrados, la muestra no debe incluir supermercados pequeños de 20,000 pies cuadrados o menos.

Una vez que las variables y la muestra han sido seleccionadas, una base de datos debe ser establecida. La base de datos debe reunirse con gran cuidado y esmero, ya que de ello depende la estabilidad del modelo que se desea generar.

La base de datos puede generarse en una hoja de cálculo de Excel, este programa facilita el manejo de datos numéricos para su manipulación posterior en Minitab, un paquete de cómputo para PC de uso frecuente para el ajuste de modelos de regresión.

3.2.4. Modelo de regresión lineal múltiple

El procedimiento para encontrar los modelos finales de regresión lineal múltiple se comenta a continuación.

3.2.4.1. Etapa 1: Modelo general

Las variables predictoras candidatas se introducen para el análisis de regresión múltiple y se genera el primer modelo. En esta etapa se conducen pruebas estadísticas por medio de Minitab, como lo son el estudio de residuales, análisis de varianza, y la evaluación de los estadísticos VIF, R^2 , y S.

Se observa la presencia del efecto de multicolinealidad en el modelo, debido a los valores del estadístico VIF superiores a 10 en algunas de las variables. Para comprobar la presencia de este problema, se lleva a cabo el análisis de eigenvalores, número de condición, e índice de condición, así como el estudio de la matriz de correlaciones.

Las relaciones entre las variables independientes y la variable dependiente (ventas) se grafican y se evalúan para identificar las correlaciones. Las gráficas de dispersión permiten observar el comportamiento de cada una de las variables x con respecto a y , de tal modo que es posible apreciar la linealidad de los datos de la muestra.

En este caso, los resultados de las gráficas indican dispersión en los datos y un comportamiento no lineal de las variables regresoras con respecto a la variable de respuesta debido a la presencia de la sucursal A. Por esta razón, se decide estudiar la información de dicha tienda por separado.

3.2.4.2. Etapa 2: Segmentación

Para modelar la segunda ecuación de regresión, se segmenta la base de datos al utilizar en conjunto la información de las tiendas B, C, D, E, y F.

Las gráficas de dispersión generadas permiten identificar un comportamiento afín entre los datos de las sucursales C, E y F.

Por otro lado, las tiendas de B y D muestran un comportamiento distinto al resto de las sucursales, lo que indica la necesidad de separar esas tiendas en grupos independientes.

3.2.4.3. Etapa 3: Formación de grupos

A partir de los resultados de la segmentación, se procede a la formación de grupos de datos para generar cuatro modelos de regresión.

Los grupos se presentan como las siguientes sucursales:

- (1) A
- (2) B
- (3) D
- (4) Combo (C, E, y F)

La información se introduce en las hojas de trabajo de Minitab y se obtienen las ecuaciones de regresión correspondientes a cada conjunto de datos.

3.2.4.4. Etapa 4: Eliminación de variables no significativas

A cada uno de los cuatro modelos resultantes se le aplica el método *stepwise*. Los resultados arrojados y los valores-p de cada una de las variables bajo estudio permite identificar y eliminar las variables que no son satisfactorias en la ecuación de regresión.

Una vez que se anulan las variables no significativas, se obtienen los modelos finales de regresión múltiple. En este paso es necesario aplicar pruebas estadísticas a cada modelo para verificar normalidad, homocedasticidad, e independencia de los datos.

3.3. Conclusiones

Tan pronto como los modelos hayan sido elaborados, un paso crucial es la descripción de la forma en que las variables independientes clave han sido medidas y recolectadas. La documentación es importante, ya que permite que personas que no estuvieron involucradas en el desarrollo inicial del modelo de regresión puedan aplicarlo y reproducirlo.

Un modelo de regresión puede operar dentro de los límites de la información incluida en la base de datos empleada para su desarrollo. Por esto, es importante conocer el contexto del área de aplicación del modelo para su correcto uso e interpretación.

Capítulo 4. Resultados

4.1. Introducción

Este capítulo tiene como objetivo la aplicación del análisis de regresión lineal múltiple en el caso real de una empresa dedicada a la comercialización de productos alimenticios.

Por razones de confidencialidad no se hace mención del nombre de la empresa; Así mismo, se omite la presentación de los datos originales proporcionados para el estudio. El análisis estadístico se realiza por medio del uso del software Minitab.

4.2. Aplicación de la metodología

La metodología propuesta en el capítulo 3, se emplea con el propósito de conocer el comportamiento de las ventas de las tiendas de una cadena de supermercados en base a la influencia de las variables independientes seleccionadas.

4.2.1. Definición de los objetivos del análisis

Como se menciona en el primer capítulo de la presente tesis, los objetivos del análisis corresponden al estudio del desempeño actual de las variables de productividad que forman parte de los departamentos de perecederos de las tiendas de una cadena de supermercados, así como el desarrollo de una serie de modelos de pronóstico que permitan predecir el comportamiento de las ventas a futuro.

4.2.2. Determinación de variables

Las variables que se describen a continuación, han sido determinadas por medio de la opinión del experto de la empresa bajo estudio.

4.2.2.1. Ventas (y)

Una venta es la acción y efecto de vender un producto o servicio a un precio pactado. En un supermercado, las ventas representan un indicador clave de desempeño que permite conocer el estado financiero y operacional de la empresa.

Capítulo 4. Resultados

4.1. Introducción

Este capítulo tiene como objetivo la aplicación del análisis de regresión lineal múltiple en el caso real de una empresa dedicada a la comercialización de productos alimenticios.

Por razones de confidencialidad no se hace mención del nombre de la empresa; Así mismo, se omite la presentación de los datos originales proporcionados para el estudio. El análisis estadístico se realiza por medio del uso del software Minitab.

4.2. Aplicación de la metodología

La metodología propuesta en el capítulo 3, se emplea con el propósito de conocer el comportamiento de las ventas de las tiendas de una cadena de supermercados en base a la influencia de las variables independientes seleccionadas.

4.2.1. Definición de los objetivos del análisis

Como se menciona en el primer capítulo de la presente tesis, los objetivos del análisis corresponden al estudio del desempeño actual de las variables de productividad que forman parte de los departamentos de perecederos de las tiendas de una cadena de supermercados, así como el desarrollo de una serie de modelos de pronóstico que permitan predecir el comportamiento de las ventas a futuro.

4.2.2. Determinación de variables

Las variables que se describen a continuación, han sido determinadas por medio de la opinión del experto de la empresa bajo estudio.

4.2.2.1. Ventas (y)

Una venta es la acción y efecto de vender un producto o servicio a un precio pactado. En un supermercado, las ventas representan un indicador clave de desempeño que permite conocer el estado financiero y operacional de la empresa.

Capítulo 4. Resultados

4.1. Introducción

Este capítulo tiene como objetivo la aplicación del análisis de regresión lineal múltiple en el caso real de una empresa dedicada a la comercialización de productos alimenticios.

Por razones de confidencialidad no se hace mención del nombre de la empresa; Así mismo, se omite la presentación de los datos originales proporcionados para el estudio. El análisis estadístico se realiza por medio del uso del software Minitab.

4.2. Aplicación de la metodología

La metodología propuesta en el capítulo 3, se emplea con el propósito de conocer el comportamiento de las ventas de las tiendas de una cadena de supermercados en base a la influencia de las variables independientes seleccionadas.

4.2.1. Definición de los objetivos del análisis

Como se menciona en el primer capítulo de la presente tesis, los objetivos del análisis corresponden al estudio del desempeño actual de las variables de productividad que forman parte de los departamentos de perecederos de las tiendas de una cadena de supermercados, así como el desarrollo de una serie de modelos de pronóstico que permitan predecir el comportamiento de las ventas a futuro.

4.2.2. Determinación de variables

Las variables que se describen a continuación, han sido determinadas por medio de la opinión del experto de la empresa bajo estudio.

4.2.2.1. Ventas (y)

Una venta es la acción y efecto de vender un producto o servicio a un precio pactado. En un supermercado, las ventas representan un indicador clave de desempeño que permite conocer el estado financiero y operacional de la empresa.

Capítulo 4. Resultados

4.1. Introducción

Este capítulo tiene como objetivo la aplicación del análisis de regresión lineal múltiple en el caso real de una empresa dedicada a la comercialización de productos alimenticios.

Por razones de confidencialidad no se hace mención del nombre de la empresa; Así mismo, se omite la presentación de los datos originales proporcionados para el estudio. El análisis estadístico se realiza por medio del uso del software Minitab.

4.2. Aplicación de la metodología

La metodología propuesta en el capítulo 3, se emplea con el propósito de conocer el comportamiento de las ventas de las tiendas de una cadena de supermercados en base a la influencia de las variables independientes seleccionadas.

4.2.1. Definición de los objetivos del análisis

Como se menciona en el primer capítulo de la presente tesis, los objetivos del análisis corresponden al estudio del desempeño actual de las variables de productividad que forman parte de los departamentos de perecederos de las tiendas de una cadena de supermercados, así como el desarrollo de una serie de modelos de pronóstico que permitan predecir el comportamiento de las ventas a futuro.

4.2.2. Determinación de variables

Las variables que se describen a continuación, han sido determinadas por medio de la opinión del experto de la empresa bajo estudio.

4.2.2.1. Ventas (y)

Una venta es la acción y efecto de vender un producto o servicio a un precio pactado. En un supermercado, las ventas representan un indicador clave de desempeño que permite conocer el estado financiero y operacional de la empresa.

Para fines del proyecto, las ventas de las sucursales exploradas se expresan en pesos.

Para eliminar el efecto de la inflación sobre las ventas, se utiliza el Índice Nacional de Precios al Consumidor (INPC) tomando como base de transformación el mes de Noviembre de 2003, que corresponde al mes inicial del primer año de estudio según los calendarios fiscales de la empresa (Ver anexos 1-5).

4.2.2.2. Sucursal (x_1)

La sucursal corresponde a la tienda donde se realiza la venta al por menor de productos perecederos y alimentos preparados.

Se consideran un total de 6 sucursales que son analizadas como variables indicadoras. Las tiendas pertenecen a distintos clusters según el tipo de consumidores a los que se dirige.

El cluster A incluye las sucursales de A, B y D, las cuales ofrecen productos y servicios orientados a consumidores de clase económica media a superior con ingresos per capita entre los \$420 y \$600 pesos diarios.

La sucursal C pertenece al cluster B, en donde se atiende a consumidores con poder de compra entre los \$210 y \$420 pesos diarios.

El cluster C comprende las tiendas de E y F, las cuales se ubican en zonas geográficas donde predominan clientes de ingresos per capita entre \$120 y \$210 pesos diarios.

Los niveles de las variables indicadoras se muestran en la tabla 4.1.

X11	X12	X13	X14	X15	X16	
1	0	0	0	0	0	Si la observación procede de la tienda A
0	1	0	0	0	0	Si la observación procede de la tienda B
0	0	1	0	0	0	Si la observación procede de la tienda C
0	0	0	1	0	0	Si la observación procede de la tienda D
0	0	0	0	1	0	Si la observación procede de la tienda E
0	0	0	0	0	1	Si la observación procede de la tienda F

Tabla 4.1. Variables indicadoras de sucursal.

4.2.2.3. Departamento (x₂)

Los departamentos representan las áreas de la empresa dedicadas a la venta de mercancías y productos preparados. En esta tesis se hace énfasis en los departamentos de alimentos como lo son: Carnes, Pescados y Mariscos, Delicatessen, Café, Panadería, y Tortillería.

En el departamento de Carnes se ofrece a los clientes una amplia variedad de cortes de carne de tipo bovino, porcino y aviar. Los productos se venden en charolas de carne precortada y empacada por la cadena de supermercados o por medio de los proveedores, también se brinda el servicio de venta de kilos de carne a solicitud del consumidor.

Los Pescados y Mariscos se comercializan en forma similar que las carnes, en este caso se ofrecen diversos productos del mar que son vendidos en kilo o fracción de producto fresco o congelado.

Los departamentos dedicados a la venta de alimentos preparados corresponden a Delicatessen, Café, Panadería y Tortillería.

El área de Delicatessen, comúnmente conocida como Salchichonería, se encarga de la oferta de productos alimenticios como carnes frías, embutidos, y quesos; las mercancías son abastecidas a las tiendas por medio de proveedores de marcas reconocidas de alimentos.

En Café se ofrece el servicio de venta de bebidas como jugos, agua embotellada, refrescos, y café. Además se venden alimentos preparados como sándwiches y rebanadas de pastel.

El departamento de Panadería se encarga de la elaboración de pan estilo europeo, pan dulce, y pastelería. El producto de panadería se vende en unidades de pan o pasteles.

Finalmente, se considera el departamento de Tortillería, el cual produce tortillas de maíz y de harina, así como tostadas y totopos. Las tortillas se venden en kilos de producto que son empacados en bolsas de plástico o papel estraza según las preferencias del cliente.

Para el desarrollo del análisis estadístico de los datos, las sucursales se estudian como variables indicadoras.

Los niveles de las variables indicadoras se muestran en la tabla 4.2.

X21	X22	X23	X24	X25	X26	
1	0	0	0	0	0	Si la observación procede del departamento Carnes
0	1	0	0	0	0	Si la observación procede del departamento Pescados y Mariscos
0	0	1	0	0	0	Si la observación procede del departamento Delicatessen
0	0	0	1	0	0	Si la observación procede del departamento Café
0	0	0	0	1	0	Si la observación procede del departamento Panadería
0	0	0	0	0	1	Si la observación procede del departamento Tortillería

Tabla 4.2. Variables indicadoras de departamento.

4.2.2.4. Quincenas (x₃)

El sistema de pagos de nómina en la mayoría de las empresas se maneja en periodos de tiempo quincenales. Para este proyecto se considera importante conocer el impacto que tienen sobre las ventas las semanas que corresponden o no a una quincena. Se analizan los años 2004, 2005, y 2006, considerando que cada año tiene un total de 52 semanas, las cuales son divididas en 13 periodos fiscales que inician en el mes de Noviembre del año anterior al que se esta analizado, y terminan en el mes de Octubre del año en cuestión.

Las quincenas se manejan como una causa de variación debido a que corresponde a un dato de solamente dos opciones. Se define entonces la variable indicadora x₃ como:

$$x_3 = \begin{cases} 0 & \text{La observación no corresponde a quincena} \\ 1 & \text{La observación corresponde a quincena} \end{cases}$$

Para definir qué semanas son o no una quincena, fue necesario establecer los siguientes supuestos:

- (1) Cada mes del año tiene dos días de pago de nómina que se consideran como quincena.
- (2) En caso de que el día 15 de cada mes sea sábado o domingo, la quincena se asigna al día viernes anterior a tal fecha.
- (3) Al finalizar el mes en el día 30 ó 31, siendo sábado o domingo, la quincena se asigna al día viernes último del mes. Para el caso específico de Febrero, el último día del mes puede ser 28 ó 29.
- (4) La excepción al supuesto número dos se aplica en la selección de la segunda quincena de Octubre para los años 2005 y 2006. En este caso, se considera como la quincena el último viernes del mes.

4.2.2.5. Cantidad de producto vendido (x_4)

La cantidad de producto vendido corresponde a las unidades de alimentos que compran los clientes. Dichas unidades varían según el departamento al que pertenecen.

En el departamento de Carnes, Pescados y Mariscos, Delicatessen y Tortillería, las unidades se miden en kilos de producto vendido. En Café, se consideran el número de bebidas que han sido escaneadas en la caja registradora. Por último, en el departamento de Panadería, la cantidad vendida se contabiliza en unidades de pan o en cantidad de pasteles que salen de la tienda.

4.2.2.6. Cantidad de socios parciales (x_5)

Los trabajadores de tiempo parcial son aquellos que laboran una jornada de trabajo de 4 horas diarias. En la empresa bajo estudio se aplica una clasificación distinta, ya que se consideran dos tipos de socios parciales.

- (1) Parcial/Diario: Son empleados que trabajan 24 horas a la semana, tienen un contrato de planta en la compañía, y verifican tarjeta de entrada y salida.
- (2) Eventual/Parcial: Son empleados que trabajan 24 horas a la semana, no tienen un contrato de planta en la compañía, y verifican tarjeta de entrada y salida.

En el análisis se toman en cuenta la cantidad de socios de medio tiempo que laboran semanalmente en las sucursales y departamentos antes mencionados.

4.2.2.7. Cantidad de socios completos (x_6)

Los trabajadores de tiempo completo son aquellos que laboran una jornada de trabajo de 8 horas diarias. Al igual que en el caso anterior, la empresa aplica su propia clasificación para definir diferencias entre los socios completos.

- (1) Completo/Mensual: Son empleados que trabajan 48 horas semanales, tienen contrato de planta, y ejercen puestos administrativos.
- (2) Eventual/Mensual: Son empleados que trabajan 48 horas semanales, no tienen contrato de planta, y ejercen puestos administrativos.

- (3) Completo/Diario: Son empleados que trabajan 48 horas semanales, tienen contrato de planta, y verifican tarjeta de entrada y salida.
- (4) Eventual/Diario: Son empleados que trabajan 48 horas semanales, no tienen contrato de planta, y verifican tarjeta de entrada y salida.

En el estudio se toman en cuenta la cantidad de socios de tiempo completo que laboran semanalmente en las sucursales y departamentos antes mencionados.

4.2.2.8. Cantidad de bajas (x_7)

Esta variable considera la cantidad de personal que renuncia a su puesto de trabajo en cada una de las sucursales y departamentos bajo análisis.

Las bajas se contabilizan en la semana del año en que ocurrieron, cabe mencionar que no se toma en cuenta si el socio laboraba como empleado de tiempo completo o parcial.

4.2.2.9. Horas disponibles de trabajo efectivo (x_8)

Las horas disponibles de trabajo efectivo corresponden al tiempo que el socio dedica a la elaboración y venta de las mercancías que se ofrecen en los departamentos de una tienda, omitiendo tiempos de ocio y actividades que no agregan valor.

Las horas disponibles se definen por periodos fiscales de operación de la empresa. Para fines prácticos se considera que cada uno de los 13 periodos esta compuesto por 4 semanas laborales. Las horas por periodo se dividen entre 4 y el resultado se aplica igual para cada una de las 4 semanas del mes al que corresponde dicho periodo.

4.2.2.10. Nómina por departamento (x_9)

La nómina corresponde a la lista de personal que labora para un departamento en específico y que percibe un sueldo por la realización de las actividades correspondientes al puesto de trabajo.

La nómina por departamento se define en 13 periodos fiscales comprendidos por 4 semanas cada uno. Los datos utilizados en el estudio se aplican al dividir cada uno de los periodos entre 4 para obtener el valor semanal de los salarios.

Al igual que en las ventas, la nómina es afectada por la inflación, por tanto se aplica el INPC para eliminar el efecto de desvalorización sobre los datos.

4.2.2.11. Add loss (x_{10})

Add loss por su nombre en el idioma inglés, representa las rebajas en precio que se aplica en los productos que se ofrecen a los consumidores en los diversos departamentos de un supermercado.

Se pretende evaluar el efecto del decremento en los costos del producto sobre las ventas del supermercado. Las rebajas se miden en pesos. Al igual que las ventas, los valores de *add loss* son transformados para eliminar el efecto de la inflación al utilizar el INPC tomando como base el mes de Noviembre de 2003.

4.2.2.12. Margen de ganancia (x_{17})

El margen de ganancia revela las ganancias obtenidas por la compañía, tomando en consideración los costos incurridos para producir bienes y/o servicios. En otras palabras, el margen de ganancia es igual a los ingresos brutos divididos entre las ventas netas, y puede expresarse como un porcentaje (Guajardo, 2002).

El margen de ganancia es un indicador de la rentabilidad de la empresa, entre más grande sea el margen de ganancia mayor es la cantidad de dinero disponible para la realización de operaciones alternas del negocio (Guajardo, 2002).

Así mismo, resulta fácil trabajar con el margen de ganancia, debido a que indica directamente cuánto dinero de la venta del producto corresponde a las ganancias o utilidades para la empresa (Guajardo, 2002).

Para usos prácticos, los datos se manejan en pesos, eliminando el efecto de la inflación como en los casos anteriormente descritos.

4.2.2.13. Venta por hora hombre (x_{18})

Las ventas por hora hombre miden la cantidad de dinero que se obtiene por hora de trabajo efectivo. En el análisis de los datos se manejan las ventas en pesos transformados sobre la base del mes de noviembre de 2003 del INPC.

4.2.3. Selección de la muestra y recolección de datos

Una vez que se ha determinado cuales son las variables que serán analizadas en la regresión lineal múltiple, se procede a la recolección de los datos.

Se realiza un estudio retrospectivo donde se utilizan datos históricos de las variables antes descritas.

En este caso se determina el uso de una muestra representativa de un total de 129,168 datos, los cuales provienen de la información recopilada para 3 años de operaciones de la empresa, siendo cada año compuesto por 52 semanas.

Los datos se organizan de modo tal que se dispone de información de cada una de las 6 sucursales en relación con los 6 departamentos seleccionados, para cada una de las 23 variables estudiadas.

4.2.4. Modelo de regresión múltiple

Con el propósito de encontrar el mejor modelo de regresión para pronosticar ventas y conocer el desempeño de las variables de productividad seleccionadas, se procede al análisis de la base de datos por medio del uso de paquetes estadísticos de cómputo como lo son Minitab y SPSS.

4.2.4.1. Etapa 1: Modelo general

En la primera etapa de modelación, se evalúan los datos de las variables en conjunto considerando en el estudio las 6 sucursales y los 6 departamentos.

La ecuación de regresión arrojada por Minitab es la siguiente:

$$Y = - 56611 + 24455 X_{11} + 32025 X_{12} + 10463 X_{13} + 8487 X_{14} - 3590 X_{15} + 268227 X_{20} + 59165 X_{21} + 105697 X_{22} - 24562 X_{23} - 156321 X_{24} + 16282 X_3 + 0.669 X_4 - 748 X_5 + 14175 X_6 - 7241 X_7 + 90.0 X_8 - 1.09 X_9 + 0.725 X_{10} + 0.770 X_{17} + 36.4 X_{18}$$

4.2.4.1.1. Estimación de los parámetros del modelo

Los coeficientes del modelo son:

4.2.3. Selección de la muestra y recolección de datos

Una vez que se ha determinado cuales son las variables que serán analizadas en la regresión lineal múltiple, se procede a la recolección de los datos.

Se realiza un estudio retrospectivo donde se utilizan datos históricos de las variables antes descritas.

En este caso se determina el uso de una muestra representativa de un total de 129,168 datos, los cuales provienen de la información recopilada para 3 años de operaciones de la empresa, siendo cada año compuesto por 52 semanas.

Los datos se organizan de modo tal que se dispone de información de cada una de las 6 sucursales en relación con los 6 departamentos seleccionados, para cada una de las 23 variables estudiadas.

4.2.4. Modelo de regresión múltiple

Con el propósito de encontrar el mejor modelo de regresión para pronosticar ventas y conocer el desempeño de las variables de productividad seleccionadas, se procede al análisis de la base de datos por medio del uso de paquetes estadísticos de cómputo como lo son Minitab y SPSS.

4.2.4.1. Etapa 1: Modelo general

En la primera etapa de modelación, se evalúan los datos de las variables en conjunto considerando en el estudio las 6 sucursales y los 6 departamentos.

La ecuación de regresión arrojada por Minitab es la siguiente:

$$Y = - 56611 + 24455 X_{11} + 32025 X_{12} + 10463 X_{13} + 8487 X_{14} - 3590 X_{15} + 268227 X_{20} + 59165 X_{21} + 105697 X_{22} - 24562 X_{23} - 156321 X_{24} + 16282 X_3 + 0.669 X_4 - 748 X_5 + 14175 X_6 - 7241 X_7 + 90.0 X_8 - 1.09 X_9 + 0.725 X_{10} + 0.770 X_{17} + 36.4 X_{18}$$

4.2.4.1.1. Estimación de los parámetros del modelo

Los coeficientes del modelo son:

4.2.3. Selección de la muestra y recolección de datos

Una vez que se ha determinado cuales son las variables que serán analizadas en la regresión lineal múltiple, se procede a la recolección de los datos.

Se realiza un estudio retrospectivo donde se utilizan datos históricos de las variables antes descritas.

En este caso se determina el uso de una muestra representativa de un total de 129,168 datos, los cuales provienen de la información recopilada para 3 años de operaciones de la empresa, siendo cada año compuesto por 52 semanas.

Los datos se organizan de modo tal que se dispone de información de cada una de las 6 sucursales en relación con los 6 departamentos seleccionados, para cada una de las 23 variables estudiadas.

4.2.4. Modelo de regresión múltiple

Con el propósito de encontrar el mejor modelo de regresión para pronosticar ventas y conocer el desempeño de las variables de productividad seleccionadas, se procede al análisis de la base de datos por medio del uso de paquetes estadísticos de cómputo como lo son Minitab y SPSS.

4.2.4.1. Etapa 1: Modelo general

En la primera etapa de modelación, se evalúan los datos de las variables en conjunto considerando en el estudio las 6 sucursales y los 6 departamentos.

La ecuación de regresión arrojada por Minitab es la siguiente:

$$Y = - 56611 + 24455 X_{11} + 32025 X_{12} + 10463 X_{13} + 8487 X_{14} - 3590 X_{15} + 268227 X_{20} + 59165 X_{21} + 105697 X_{22} - 24562 X_{23} - 156321 X_{24} + 16282 X_3 + 0.669 X_4 - 748 X_5 + 14175 X_6 - 7241 X_7 + 90.0 X_8 - 1.09 X_9 + 0.725 X_{10} + 0.770 X_{17} + 36.4 X_{18}$$

4.2.4.1.1. Estimación de los parámetros del modelo

Los coeficientes del modelo son:

Predictor	Coef	SE Coef	T	P
Constant	-56611	3977	-14.24	0.000
X ₁₁	24455	4694	5.21	0.000
X ₁₂	32025	4547	7.04	0.000
X ₁₃	10463	3333	3.14	0.002
X ₁₄	8487	3919	2.17	0.030
X ₁₅	-3590	3568	-1.01	0.314
X ₂₀	268227	8422	31.85	0.000
X ₂₁	59165	4147	14.27	0.000
X ₂₁	105697	7963	13.27	0.000
X ₂₃	-24562	3854	-6.37	0.000
X ₂₄	-156321	12438	-12.57	0.000
X ₃	16282	1862	8.75	0.000
X ₄	0.6693	0.1846	3.63	0.000
X ₅	-748.1	493.0	-1.52	0.129
X ₆	14175	1531	9.26	0.000
X ₇	-7241	4514	-1.60	0.109
X ₈	90.00	19.36	4.65	0.000
X ₉	-1.0941	0.3327	-3.29	0.001
X ₁₀	0.72457	0.02224	32.59	0.000
X ₁₇	0.770027	0.008873	86.78	0.000
X ₁₈	36.362	7.656	4.75	0.000

Tabla 4.3. Coeficientes del modelo etapa 1.

Los valores-p de los coeficientes estimados de la sucursal E (x₁₅), cantidad de socios parciales (x₅) y cantidad de bajas (x₇) son 0.314, 0.129 y 0.109, respectivamente, indicando que no están relacionados con las ventas a un nivel de significancia α de 0.05.

Los estimados del modelo de regresión arrojados por Minitab se muestran en la figura 4.1.

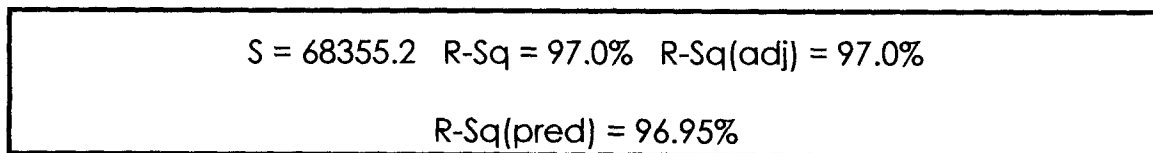
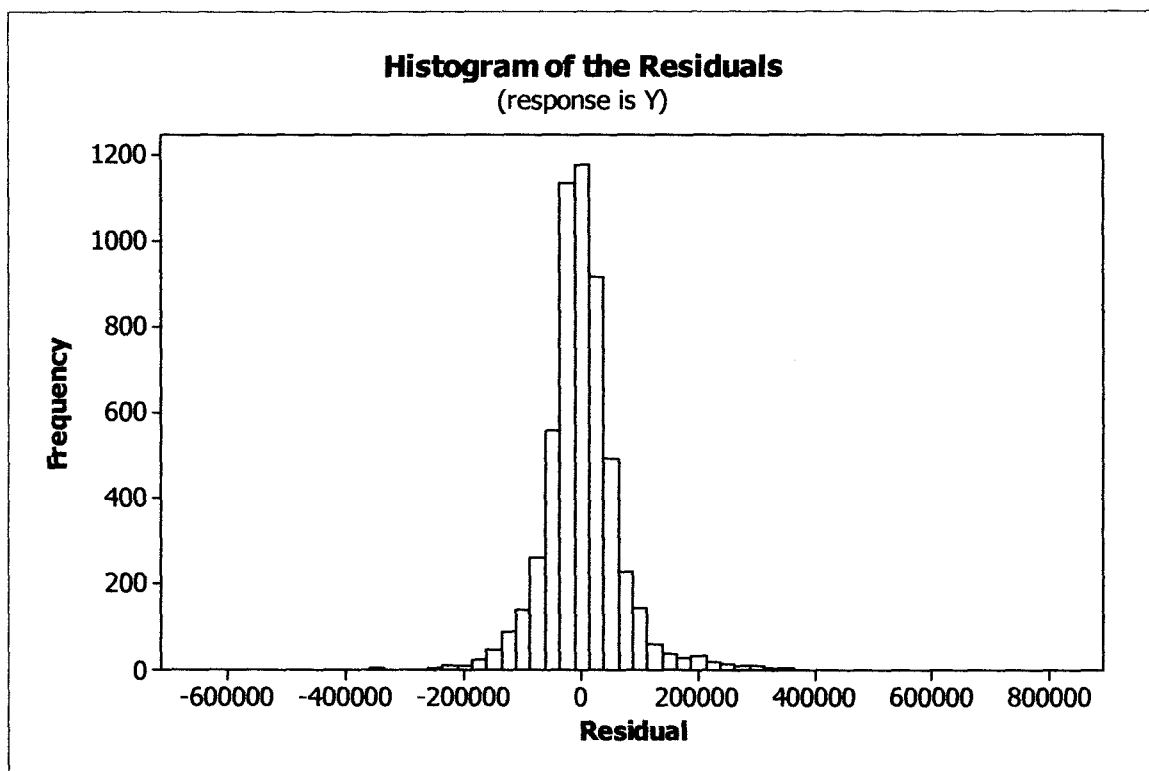


Figura 4.1. Estimados del modelo de regresión etapa 1.

La desviación estándar (S) indica que la variación esperada de las ventas con respecto de la media aritmética de los datos es de \$68,355.2 pesos.

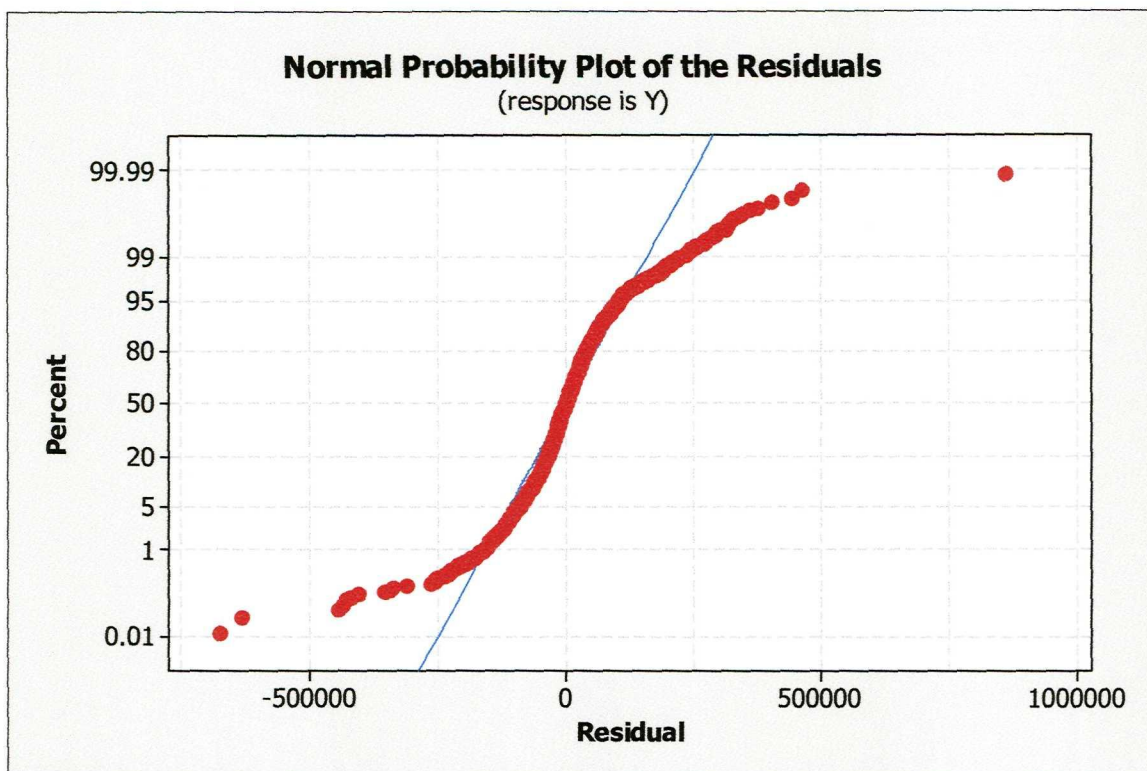
El coeficiente de determinación (R^2) expresa que el 97% de la variación total de las ventas puede ser explicada por el modelo cuantitativo. El coeficiente de determinación ajustado ($R^2_{ajustado}$) toma en cuenta los grados de libertad del estudio, siendo un valor más realista que R^2 . En este caso, $R^2_{ajustado}$ indica que la proporción de cambio de las ventas que se aplica al modelo es de 97%.

El valor R^2 de predicción se calcula a partir del estadístico PRESS e indica que tan bueno es el modelo para pronosticar ventas para nuevas observaciones. El $R^2_{predicción}$ de 0.9695 es cercano a los valores de R^2 y $R^2_{ajustado}$, se sugiere que el modelo es bueno para predecir nuevos datos. Las gráficas de análisis de residuales de la regresión se presentan a continuación:



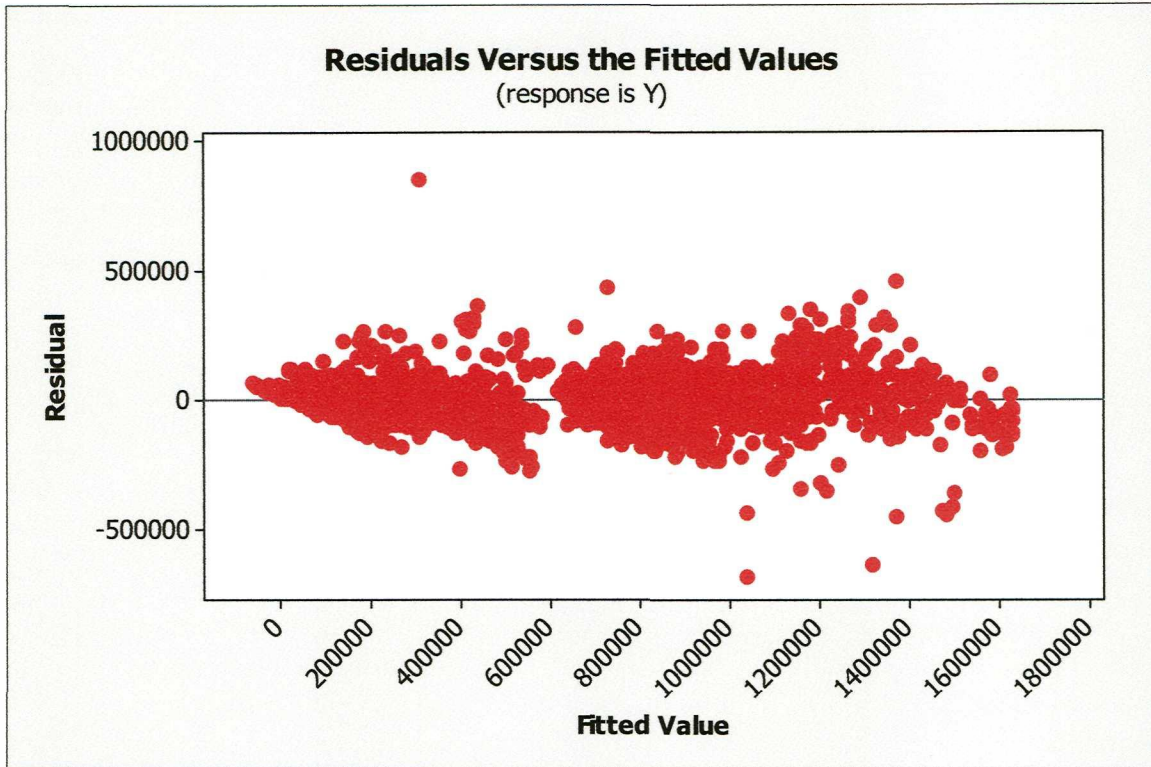
Gráfica 4.1. Histograma de residuales.

El histograma indica la presencia de datos atípicos en los residuales debido principalmente a la barra del lado izquierdo de la gráfica.



Gráfica 4.2. Gráfica de probabilidad normal de los residuales.

La gráfica de probabilidad normal muestra un patrón aproximadamente lineal consistente con una distribución normal con colas delgadas. Los puntos en las esquina inferior izquierda 3691 y 3796, así como el punto del lado superior derecho 3744, son datos atípicos de las observaciones.



Gráfica 4.3. Gráfica de residuales contra valores ajustados.

La gráfica de residuales versus los valores ajustados muestran la existencia de puntos que no se distribuyen con cierta uniformidad, esto indica que los residuales tienen una varianza no constante.

4.2.4.1.2. Análisis de varianza (ANOVA)

La técnica estadística del ANOVA se presenta en la tabla 4.4., de acuerdo a los resultados de salida de Minitab.

Source	DF	SS	MS	F	P
Regression	20	8.15421E+14	4.07711E+13	8725.87	0.000
Residual Error	5435	2.53947E+13	4672434830		
Lack of Fit	5434	2.53947E+13	4673293888	1083.27	0.024
Pure Error	1	4314071	4314071		
Total	5455	8.40816E+14			

Tabla 4.4. Tabla ANOVA.

El valor-p (0.000) de la tabla de análisis de varianza muestra que el modelo estimado por el procedimiento de regresión es significativo a un

nivel α de 0.05, indicando que al menos un coeficiente es diferente de cero.

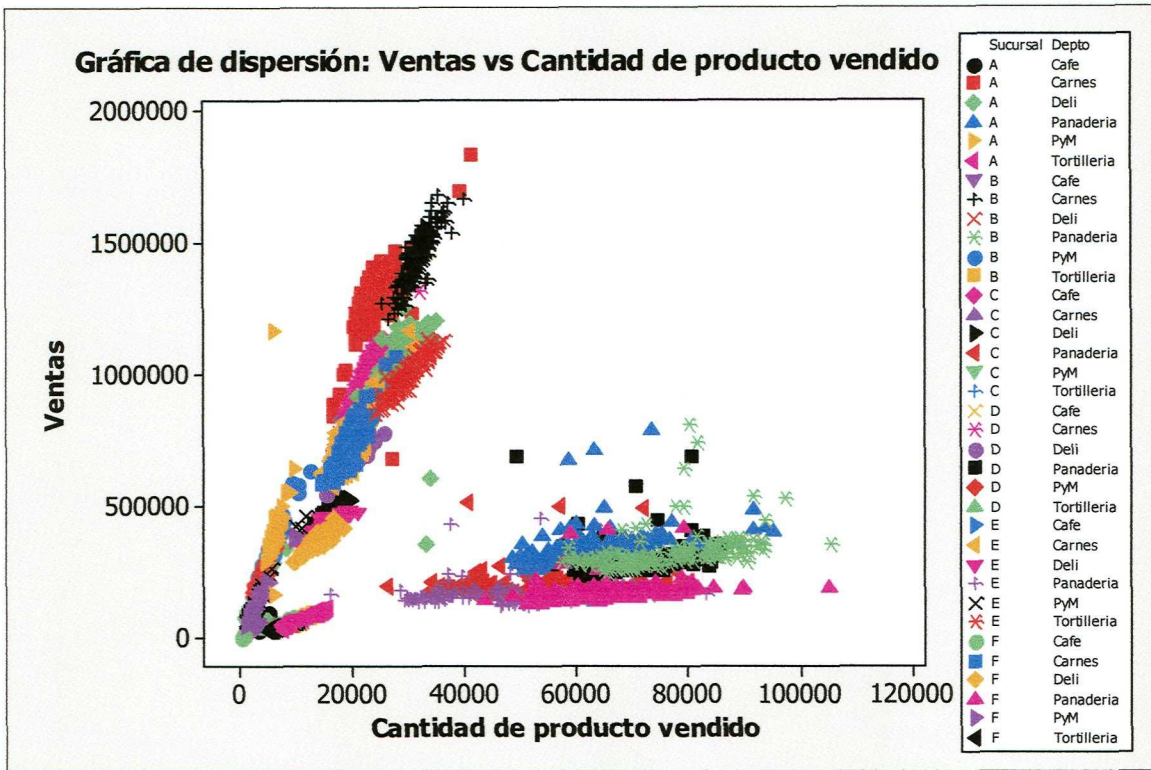
La tabla 4.5., referente a los grados de libertad y a la suma al cuadrado secuencial (Seq SS) de las variables independientes, indica que el regresor nómina por departamento (x_9) no explica una cantidad substancial de varianza única, lo cual sugiere la posibilidad de eliminar dicha variable del modelo.

Source	DF	Seq SS
X ₁₁	1	2.90492E+13
X ₁₂	1	4.32997E+13
X ₁₃	1	1.01998E+12
X ₁₄	1	8.60259E+12
X ₁₅	1	1.40826E+12
X ₂₀	1	4.33051E+14
X ₂₁	1	2.19969E+12
X ₂₂	1	2.09551E+14
X ₂₃	1	1.05057E+13
X ₂₄	1	1.62420E+13
X ₃	1	4.06003E+11
X ₄	1	9.04163E+12
X ₅	1	6.96821E+11
X ₆	1	1.79310E+12
X ₇	1	2.84646E+11
X ₈	1	2.65464E+12
X ₉	1	13431572214
X ₁₀	1	9.81437E+12
X ₁₇	1	3.56826E+13
X ₁₈	1	1.05391E+11

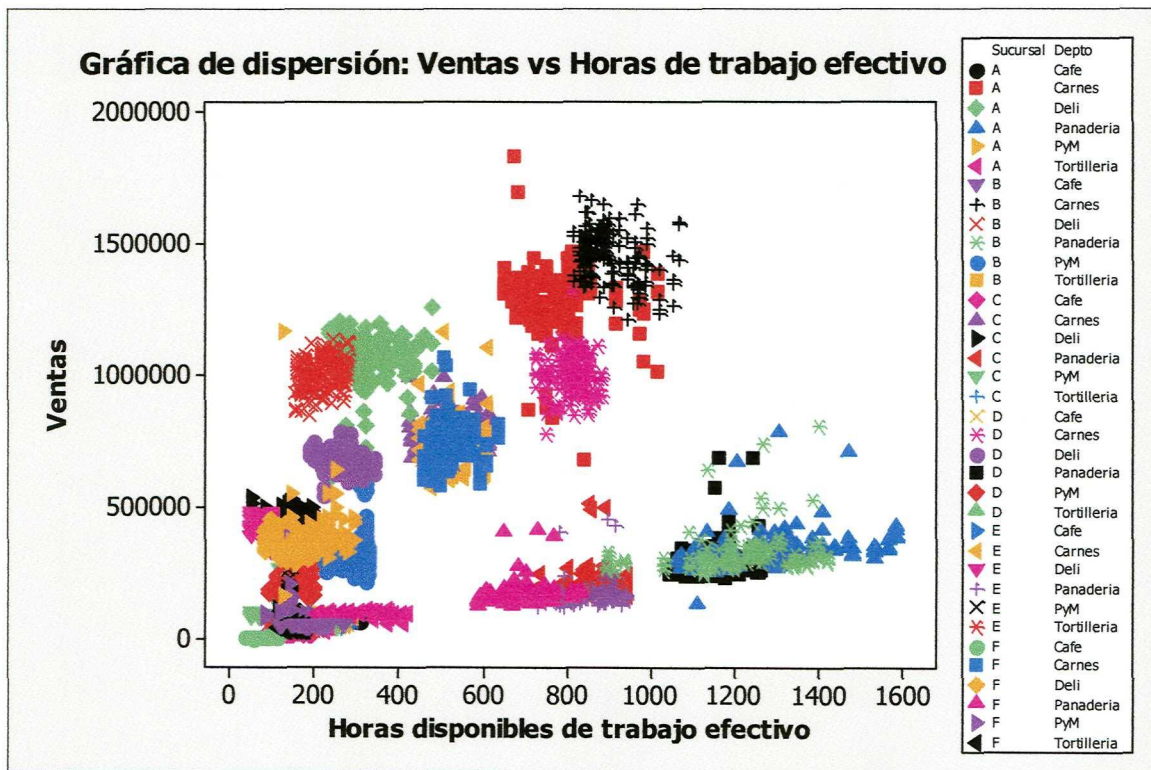
Tabla 4.5. Grados de libertad y suma al cuadrado secuencial.

4.2.4.1.3. Gráficas de dispersión

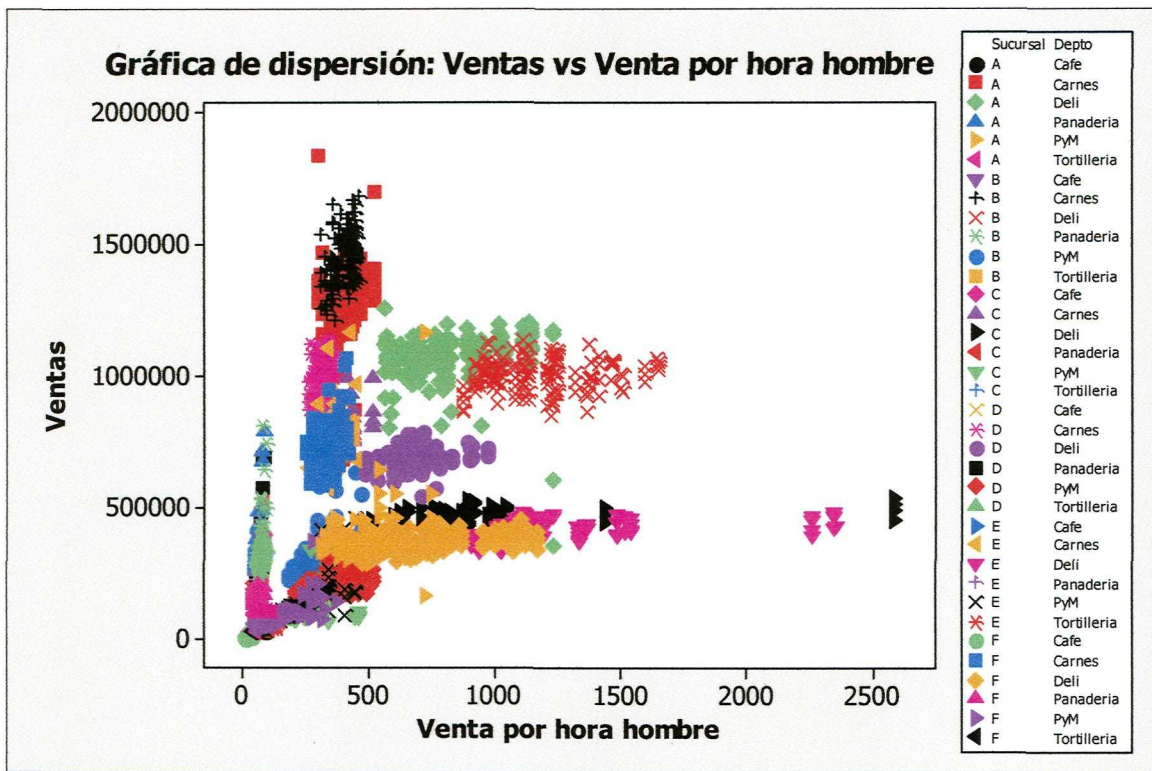
Las gráficas de dispersión son muy útiles para ajustar modelos de regresión, ya que cada una de ellas trata de arrojar señales sobre la relación entre un par de variables. Los indicadores de productividad que se modelan en la regresión se analizan en gráficas de dispersión, con el objeto de identificar el sentido de linealidad o de no linealidad de cada una de ellas con respecto a las ventas (y). Se presentan solamente algunas gráficas representativas para explicar el comportamiento general de los datos, el resto de las variables se encuentran en el anexo 6.



Gráfica 4.4. Ventas versus cantidad de producto vendido etapa 1.



Gráfica 4.5. Ventas versus horas disponibles de trabajo efectivo etapa 1.



Gráfica 4.6. Ventas versus ventas por hora hombre etapa 1.

De la gráfica 4.4., es posible apreciar patrones de comportamiento similares según el departamento analizado con respecto a las ventas de cada sucursal. Las pendientes de los datos indican ventas grandes aún cuando la cantidad de producto vendido en las áreas de Carnes, Pescados y Mariscos, y Delicatessen sea menor a 40,000 unidades. Mientras que en los departamentos de Café, Panadería y Tortillería, la pendiente muestra que las ventas de producto son grandes, sin embargo las ventas totales generadas son pequeñas en comparación con el resto de los departamentos.

El diagrama de dispersión 4.5., permite comparar la relación lineal entre la variable horas disponibles de trabajo efectivo y las ventas. Es importante notar las tendencias de agrupación de los datos, en la parte inferior derecha se observa que los puntos de las sucursales C, E y F representando al departamento Panadería, se unen para seguir una tendencia semejante. Así mismo, ocurre con A, B y D, las cuales se agrupan para mostrar que los datos de Panadería son similares. Las pendientes indican un incremento en las ventas de los departamentos de Carnes, Pescados y Mariscos, y Delicatessen, mientras más sean las horas disponibles de trabajo efectivo. Por otro lado, Café, Panadería y Tortillería

muestran una pendiente de regresión que demuestra que entre más sean las horas de trabajo efectivo, el incremento en las ventas no es grande.

En la gráfica 4.6. es difícil apreciar las relaciones lineales entre la variable de ventas por hora hombre y las ventas totales. Se aprecian dos pendientes distintas que representan al departamento de Delicatessen para diferentes sucursales. Además, se observan puntos empalmados, lo cual complica la interpretación del comportamiento de los datos.

Existen problemas importantes con respecto a la dispersión y a la falta de linealidad de los datos que se aprecian en las gráficas 4.4. hasta la 4.6. Se presenta la necesidad de separar los datos en segmentos con el fin de observar con mayor facilidad la relación de las variables bajo estudio con respecto a las ventas.

4.2.4.1.4. Análisis de multicolinealidad

En un estudio de regresión múltiple, resulta de interés conocer el significado de las relaciones entre las variables explicativas y la variable de respuesta del modelo.

Por esta razón, el análisis de multicolinealidad se realiza con el propósito de identificar en el modelo general, la posibilidad de que las variables predictoras estén correlacionadas entre ellas.

4.2.4.1.4.1. Matriz de correlaciones y covarianza

La matriz de correlaciones presentada en el Anexo 7, permite inspeccionar los elementos r_{ij} no diagonales en $X'X$, con el objeto de encontrar un indicio de multicolinealidad que ocurre cuando el valor $|r_{ij}|$ de los regresores x_i y x_j es próximo a la unidad.

Al examinar las correlaciones simples r_{ij} entre los regresores para detectar la dependencia casi lineal, se observa que existen correlaciones apareadas r_{ij} sospechosamente grandes entre las variables indicadoras siguientes:

X_i	X_j	r_{ij}
X10	X17	0.657
X10	X18	0.610
X24	X8	0.769
X24	X9	0.808
X24	X4	0.883
X18	X22	0.835
X9	X4	0.882
X9	X5	0.889
X8	X4	0.855
X8	X5	0.898
X8	X9	0.982

Tabla 4.6. Correlaciones simples entre regresores.

En general, la inspección de los r_{ij} no es suficiente para detectar cosas más complejas que la multicolinealidad por pares de las variables ficticias.

La matriz de covarianza muestra una medida de la asociación lineal que existe entre las variables regresoras del estudio. Los valores de correlación positivos indican una dependencia directa o positiva, mientras que los valores iguales a 0 muestran que no existe relación lineal entre las dos variables, finalmente un valor negativo en la matriz señala la existencia de una dependencia inversa o negativa (Ver anexo 8).

4.2.4.1.4.2. Factores de inflación de varianza

En la tabla 4.7. se observan factores de inflación de varianza mayores a 10 en las variables X_{20} , X_{22} , X_{24} , X_{25} , X_4 , X_8 , y X_9 , lo cual es un indicio de que los coeficientes de regresión asociados están mal estimados debido a la presencia de intercorrelación.

	VIF
(Constante)	
X ₁₁	3.656963
X ₁₂	3.430516
X ₁₃	1.844132
X ₁₄	2.548566
X ₁₅	1.813783
X ₂₀	11.77261
X ₂₁	2.854251
X ₂₂	10.48747
X ₂₃	2.125104
X ₂₄	25.67512
X ₃	1.003566
X ₄	18.72093
X ₅	7.263129
X ₆	1.455611
X ₇	1.094429
X ₈	53.6501
X ₉	50.80066
X ₁₀	4.702562
X ₁₇	8.801714
X ₁₈	7.640034

Tabla 4.7. Factores de inflación de varianza.

4.2.4.1.4.3. Análisis de eigensistema X'X

Según los resultados de la tabla 4.8., existen varios eigenvalores muy pequeños a partir de la dimensión número 15, este efecto es síntoma de la presencia de datos deteriorados en grado extremo.

Dimensión	Eigenvalores λ_i	Índice de condición $ic(\lambda_i)$
1	8.269791	1
2	2.258397	1.913582
3	1.546177	2.312691
4	1.287333	2.534556
5	1.211534	2.61264
6	1.046419	2.811218
7	1.030918	2.832273
8	1.004173	2.869743
9	0.935422	2.973332
10	0.835923	3.145316
11	0.539433	3.915421
12	0.486535	4.122781
13	0.19413	6.526815
14	0.100021	9.092877
15	0.071701	10.73949
16	0.056438	12.10491
17	0.047595	13.18158
18	0.03557	15.24774
19	0.021686	19.52816
20	0.016054	22.69599
21	0.00475	41.7264

Tabla 4.8. Análisis de eigensistema.

El número de condición es una medida de la dispersión del espectro de eigenvalores, para calcularlo es necesario aplicar la ecuación 2.12.14., con la cual se obtiene:

$$\kappa = \frac{8,269791}{0.00475} = 1741.008 > 1000 \text{ Indica fuerte multicolinealidad}$$

Para evaluar un problema de intercorrelación por medio del índice de condición, es necesario observar si los valores de $ic(\lambda_i)$ oscilan entre 20 y 30. La dimensión número 20 sobrepasa el límite inferior de 20, mientras que en la dimensión número 21, el indicador supera el valor de 30 mostrando un problema de multicolinealidad grave.

4.2.4.2. Etapa 2: Segmentación

Los resultados de las gráficas de dispersión del modelo general, permiten observar un comportamiento no lineal de las variables independientes con respecto a la variable dependiente. En particular los datos de la

Dimensión	Eigenvalores λ_i	Índice de condición $ic(\lambda_i)$
1	8.269791	1
2	2.258397	1.913582
3	1.546177	2.312691
4	1.287333	2.534556
5	1.211534	2.61264
6	1.046419	2.811218
7	1.030918	2.832273
8	1.004173	2.869743
9	0.935422	2.973332
10	0.835923	3.145316
11	0.539433	3.915421
12	0.486535	4.122781
13	0.19413	6.526815
14	0.100021	9.092877
15	0.071701	10.73949
16	0.056438	12.10491
17	0.047595	13.18158
18	0.03557	15.24774
19	0.021686	19.52816
20	0.016054	22.69599
21	0.00475	41.7264

Tabla 4.8. Análisis de eigensistema.

El número de condición es una medida de la dispersión del espectro de eigenvalores, para calcularlo es necesario aplicar la ecuación 2.12.14., con la cual se obtiene:

$$\kappa = \frac{8,269791}{0.00475} = 1741.008 > 1000 \text{ Indica fuerte multicolinealidad}$$

Para evaluar un problema de intercorrelación por medio del índice de condición, es necesario observar si los valores de $ic(\lambda_i)$ oscilan entre 20 y 30. La dimensión número 20 sobrepasa el límite inferior de 20, mientras que en la dimensión número 21, el indicador supera el valor de 30 mostrando un problema de multicolinealidad grave.

4.2.4.2. Etapa 2: Segmentación

Los resultados de las gráficas de dispersión del modelo general, permiten observar un comportamiento no lineal de las variables independientes con respecto a la variable dependiente. En particular los datos de la

sucursal A muestran una tendencia irregular importante en comparación con el resto de las sucursales.

El proceso de segmentación de los datos se determina en dos grupos: (1) A y (2) las tiendas B, C, D, E y F.

En la etapa 2 de modelación se evalúa el segundo grupo de datos. La ecuación calculada por el software Minitab es:

$$Y = - 59418 + 35508 X_{12} + 10546 X_{13} + 9776 X_{14} - 2205 X_{15} + 246873 X_{20} + 46967 X_{21} + 98087 X_{22} - 24792 X_{23} - 171656 X_{24} + 17993 X_3 + 0.503 X_4 - 4913 X_5 + 16488 X_6 + 2927 X_7 + 223 X_8 - 1.22 X_9 + 0.810 X_{10} + 0.706 X_{17} + 47.7 X_{18}$$

De acuerdo a los resultados del paquete de cómputo, las variables X_{16} y x_{25} se encuentran altamente correlacionadas con otra variable x ; por tanto, dicha variable ha sido eliminada de la ecuación. Así mismo, la variable de la sucursal A (x_{11}) fue removida del modelo debido a que las columnas se conforman de valores iguales a cero.

4.2.4.2.1. Estimación de los parámetros del modelo

La tabla 4.9., indica que los coeficientes estimados de las variables de sucursal E (x_{25}) y cantidad de bajas (x_7), indican que dichos regresores no están relacionados con las ventas a un nivel de significancia α de 0.05.

Los estimados mostrados en la pantalla de resultados de Minitab se muestran en la figura 4.2.

El estimado S indica la existencia de variación en las ventas de 61491.7 pesos, comparado con la media de la distribución de los datos.

El coeficiente de determinación (R^2) indica que el 97.2% de la proporción de la variación total de las ventas puede ser explicada por el modelo de regresión.

El coeficiente de determinación ajustado ($R^2_{ajustado}$) indica que la proporción de cambio de las ventas que se aplica al modelo es de 97.2%.

El valor R^2 de predicción indica que tan bueno es el modelo para pronosticar ventas para nuevas observaciones. El $R^2_{predicción}$ de 0.9713 es cercano a la unidad, por tanto se sugiere que el modelo tiene una habilidad de predicción grande.

Predictor	Coef	SE Coef	T	P
Constant	-59418	3792	-15.67	0.000
X ₁₂	35508	4309	8.24	0.000
X ₁₃	10546	3025	3.49	0.000
X ₁₄	9776	3652	2.68	0.007
X ₁₅	-2205	3234	-0.68	0.495
X ₂₀	246873	8156	30.27	0.000
X ₂₁	46967	3936	11.93	0.000
X ₂₂	98087	7610	12.89	0.000
X ₂₃	-24792	3922	-6.32	0.000
X ₂₄	-171656	12178	-14.10	0.000
X ₃	17993	1841	9.77	0.000
X ₄	0.5031	0.1801	2.79	0.005
X ₅	-4913.2	585.0	-8.40	0.000
X ₆	16488	1552	10.62	0.000
X ₇	2927	4656	0.63	0.530
X ₈	223.28	21.33	10.47	0.000
X ₉	-1.2219	0.3369	-3.63	0.000
X ₁₀	0.80972	0.02485	32.58	0.000
X ₁₇	0.70605	0.01014	69.61	0.000
X ₁₈	47.714	7.682	6.21	0.000

Tabla 4.9. Coeficientes del modelo etapa 2.

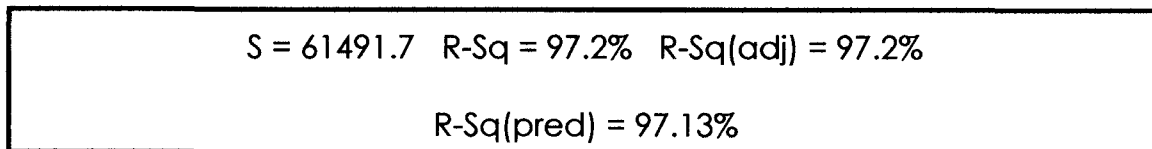
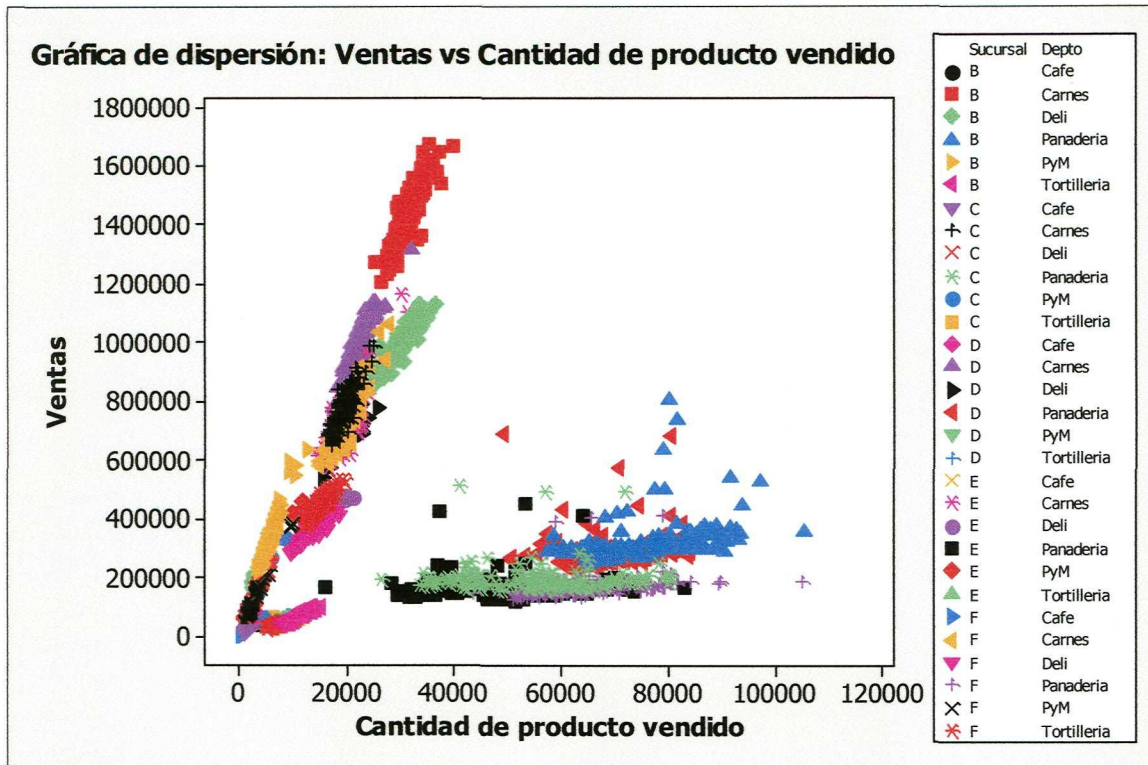


Figura 4.2. Estimados del modelo de regresión etapa 2.

4.2.4.2.2. Gráficas de dispersión

Al graficar el conjunto de datos que fueron segmentados, es posible observar con mayor claridad la relación entre las variables regresoras versus las ventas.

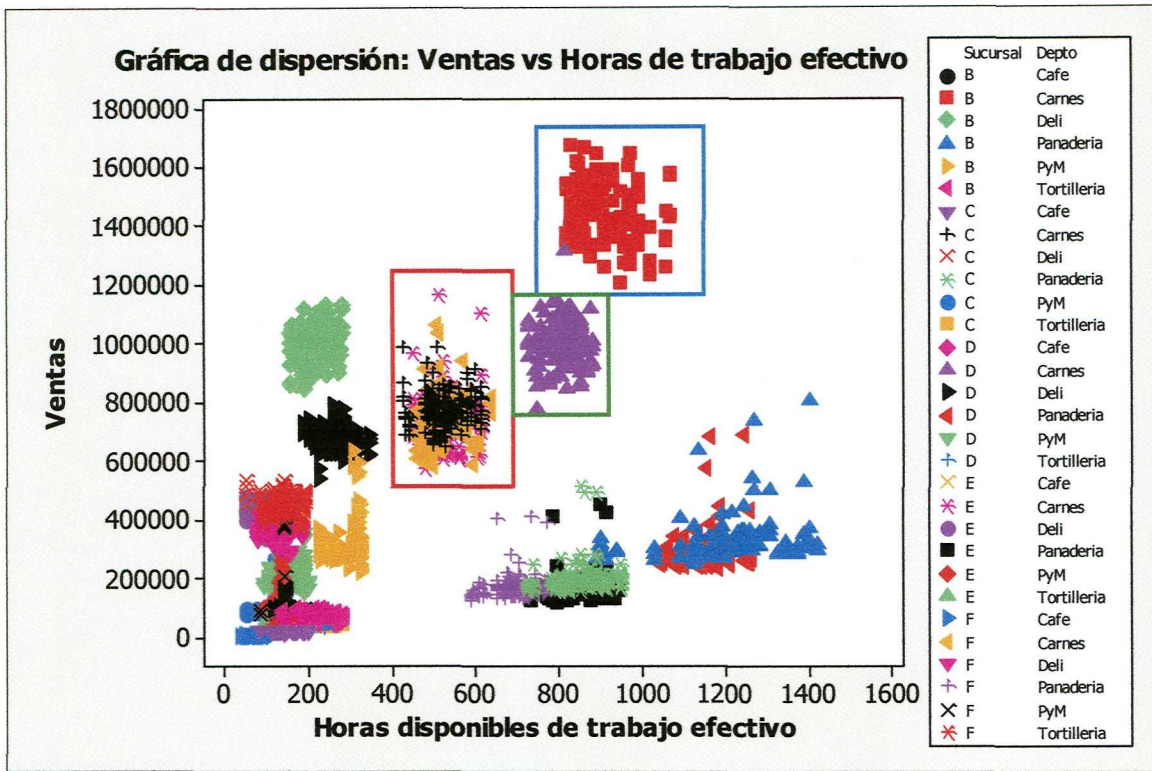
Para facilitar la comparación de las gráficas resultantes de la etapa 1 y 2 se muestran como representación las mismas tres variables antes estudiadas, el resto de las gráficas se encuentran en el anexo 9.



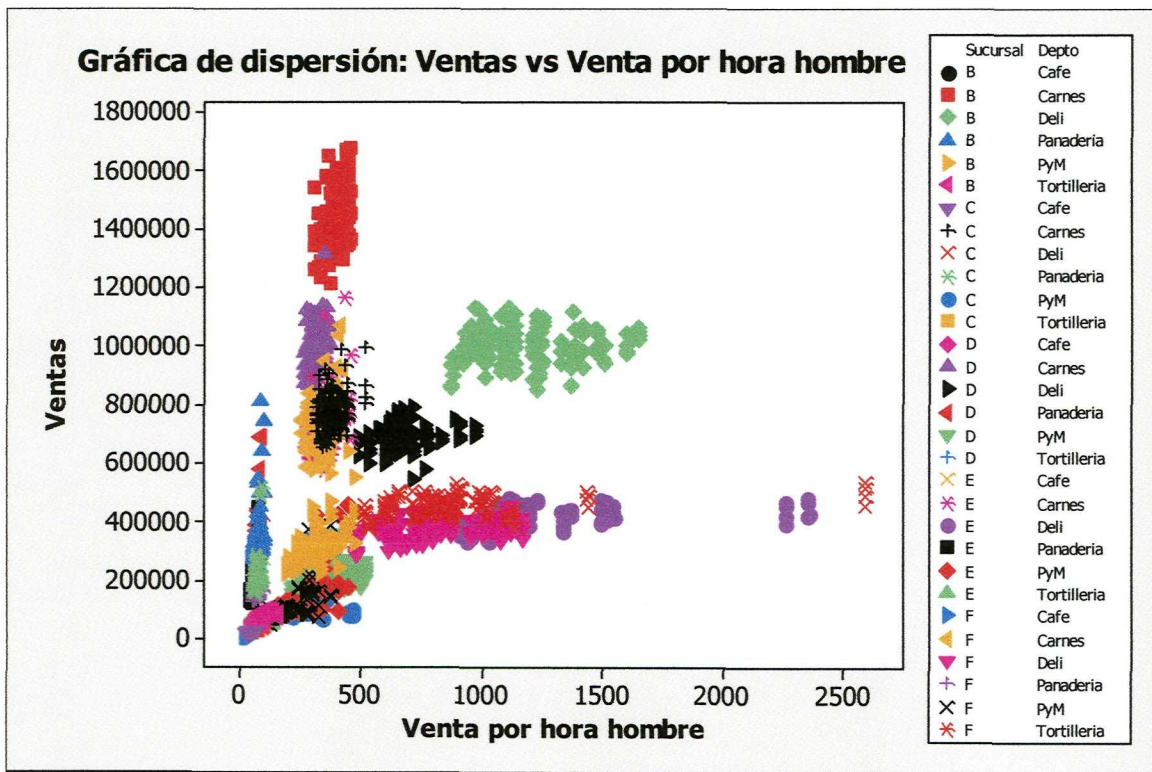
Gráfica 4.7. Ventas versus cantidad de producto vendido etapa 2.

De la gráfica 4.7., se observa que las pendientes del departamento de Panadería para cada una las cinco sucursales indican que entre mayor sea la cantidad de producto vendido, las ventas tienden a ser pequeñas. Un comportamiento similar se aprecia con el departamento de Tortillería y Café. Por otro lado, las pendientes de Carnes, Pescados y Mariscos, y Delicatessen muestran que las ventas tienden ser grandes aunque la cantidad de producto que se vende en cada uno de ellos sea pequeña.

En el diagrama 4.9., es posible identificar que el incremento en las ventas por hora hombre en el departamento de Delicatessen, no necesariamente significan mayores ventas totales, este comportamiento puede deberse a que los empleados trabajan en parte a base de comisión de los proveedores a los que representan. El resto de los departamentos muestran ventas por hora hombre inferiores a los \$500 pesos, mientras que las ventas totales tienden a incrementarse significativamente.



Gráfica 4.8. Ventas versus horas disponibles de trabajo efectivo etapa 2.



Gráfica 4.9. Ventas versus ventas por hora hombre etapa 2.

Por otro lado, las gráficas 4.7. a la 4.9., facilitan la identificación de grupos de información que siguen un patrón de comportamiento similar, ya que se observa que los datos de las tiendas C, E y F se distribuyen de forma muy parecida según el departamento al que representan. Para ilustrar este fenómeno se muestra en el recuadro rojo de la gráfica 4.8. como los puntos correspondientes al área de carnes de las sucursales antes mencionadas siguen una tendencia semejante. Así mismo ocurre con el resto de los departamentos estudiados.

De nuevo en la gráfica 4.8. se ejemplifica con un recuadro azul como la sucursal B representando el departamento de carnes, sigue un comportamiento independiente con respecto al resto de las tiendas, mientras que lo mismo ocurre con el supermercado D identificado con el recuadro verde.

Como conclusión se determina la necesidad de formar modelos por grupos, con el objeto de generar ecuaciones de regresión confiables para la predicción de nuevas observaciones de ventas.

4.2.4.3. Etapa 3: Formación de grupos

La etapa 3 de la metodología corresponde a la formación de grupos de datos para la creación de cuatro modelos de regresión. De acuerdo a los resultados observados en las gráficas de dispersión de la etapa 2, se llega a la conclusión de distribuir las sucursales de la siguiente manera:

- (1) A
- (2) B
- (3) D
- (4) Combo (C, E, y F)

Las ecuaciones de regresión resultantes se indican en las secciones mostradas a continuación.

4.2.4.3.1. Modelo A

La pantalla de salida de Minitab indica que la ecuación de regresión para la tienda A es:

$$Y = - 33643 + 899358 X_{20} + 271430 X_{21} + 645122 X_{22} + 37991 X_{23} - 173130 X_{24} + 7107 X_3 + 5.92 X_4 - 2808 X_5 + 4597 X_6 + 16685 X_7 + 96.1 X_8 - 0.66 X_9 + 0.249 X_{10} + 0.177 X_{17} + 58.8 X_{18}$$

Por otro lado, las gráficas 4.7. a la 4.9., facilitan la identificación de grupos de información que siguen un patrón de comportamiento similar, ya que se observa que los datos de las tiendas C, E y F se distribuyen de forma muy parecida según el departamento al que representan. Para ilustrar este fenómeno se muestra en el recuadro rojo de la gráfica 4.8. como los puntos correspondientes al área de carnes de las sucursales antes mencionadas siguen una tendencia semejante. Así mismo ocurre con el resto de los departamentos estudiados.

De nuevo en la gráfica 4.8. se ejemplifica con un recuadro azul como la sucursal B representando el departamento de carnes, sigue un comportamiento independiente con respecto al resto de las tiendas, mientras que lo mismo ocurre con el supermercado D identificado con el recuadro verde.

Como conclusión se determina la necesidad de formar modelos por grupos, con el objeto de generar ecuaciones de regresión confiables para la predicción de nuevas observaciones de ventas.

4.2.4.3. Etapa 3: Formación de grupos

La etapa 3 de la metodología corresponde a la formación de grupos de datos para la creación de cuatro modelos de regresión. De acuerdo a los resultados observados en las gráficas de dispersión de la etapa 2, se llega a la conclusión de distribuir las sucursales de la siguiente manera:

- (1) A
- (2) B
- (3) D
- (4) Combo (C, E, y F)

Las ecuaciones de regresión resultantes se indican en las secciones mostradas a continuación.

4.2.4.3.1. Modelo A

La pantalla de salida de Minitab indica que la ecuación de regresión para la tienda A es:

$$Y = - 33643 + 899358 X_{20} + 271430 X_{21} + 645122 X_{22} + 37991 X_{23} - 173130 X_{24} + 7107 X_3 + 5.92 X_4 - 2808 X_5 + 4597 X_6 + 16685 X_7 + 96.1 X_8 - 0.66 X_9 + 0.249 X_{10} + 0.177 X_{17} + 58.8 X_{18}$$

La variable x_{25} que corresponde a tortillería, fue removida de la ecuación de regresión debido a que se encuentra altamente correlacionada con otras variables independientes.

Los coeficientes y estimados del modelo de regresión se encuentran en la tabla 4.10. y en la figura 4.3., respectivamente.

Predictor	Coef	SE Coef	T	P
Constant	-33643	15078	-2.23	0.026
X ₂₁	899358	46654	19.28	0.000
X ₂₂	271430	19606	13.84	0.000
X ₂₃	645122	43611	14.79	0.000
X ₂₄	37991	11161	3.40	0.001
X ₂₆	-173130	50475	-3.43	0.001
X ₃	7107	4873	1.46	0.145
X ₄	5.9180	0.6416	9.22	0.000
X ₅	-2808	2306	-1.22	0.224
X ₆	4597	6622	0.69	0.488
X ₇	16685	10670	1.56	0.118
X ₈	96.07	56.23	1.71	0.088
X ₉	-0.663	1.049	-0.63	0.527
X ₁₀	0.24881	0.04912	5.06	0.000
X ₁₆	0.17718	0.04042	4.38	0.000
X ₁₇	58.85	30.71	1.92	0.056

Tabla 4.10. Coeficientes del modelo etapa 3 - A.

Los coeficientes estimados muestran varios regresores con valor-p mayor a un nivel de significancia α de 0.05, las variables que no se relacionan con las ventas según este estimado son: quincenas (x_3), cantidad de socios parciales (x_5), cantidad de socios completos (x_6), cantidad de bajas (x_7), horas disponibles de trabajo efectivo (x_8), y nómina por departamento (x_9).

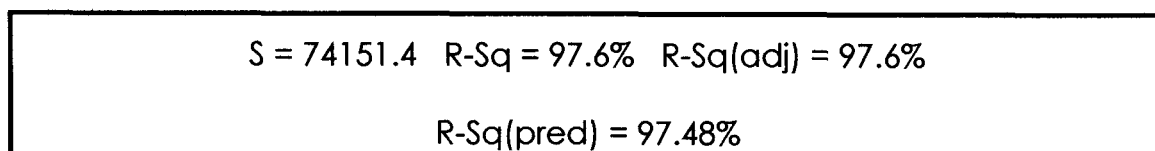


Figura 4.3. Estimados del modelo de regresión etapa 3 - A.

La variación en las ventas es de 74151.4 con respecto a la media aritmética de los datos.

El valor de R^2 muestra que el modelo de regresión se ajusta en un 97.6% a los datos. Mientras que R^2_{ajustado} de 97.6% es la proporción de los cambios en las ventas que pueden ser explicados por el modelo. Así mismo, el R^2 de predicción igual a 97.48% sugiere que el modelo tiene una amplia habilidad predictiva.

4.2.4.3.1.1. Método *Stepwise*

Con el objeto de identificar las variables que son significativas para la generación de cada uno de los modelos de regresión, se aplica el procedimiento *stepwise* con la modalidad "hacia delante y hacia atrás". Los valores de entrada utilizados son $\alpha = 0.15$ y $F = 4$, y los de salida son $\alpha = 0.15$ y $F = 4$.

La pantalla de resultados de Minitab arroja los resultados de todos los pasos del método *stepwise*, para fines prácticos se presenta solamente los valores del paso que representa al mejor modelo, el procedimiento completo se presenta en el anexo 10.

La muestra de A que fue analizada corresponde a un total de 936 datos con una variable de respuesta y sobre 15 variables predictoras (se ignoran los regresores x_{11} hasta x_{16} debido a que corresponden a los niveles de la variable indicadora de sucursal).

De la figura 4.4. se determinan las variables independientes que deben ser eliminadas de la ecuación de regresión final. Para el caso de A se remueven los datos de las variables: quincenas (x_3), cantidad de socios parciales (x_5), cantidad de socios completos (x_6), cantidad de bajas (x_7), horas disponibles de trabajo efectivo (x_8), nómina por departamento (x_9), y ventas por hora hombre (x_{18}).

Al aplicar el método *stepwise* se calculan los estimados del modelo de regresión, se observa que la variación de las ventas con respecto a la media de la distribución de los datos es de \$74,363 pesos. El coeficiente de determinación múltiple de 97.58% indica el porcentaje de los datos de la muestra que se ajustan al modelo. Así mismo, el R^2_{ajustado} sigue siendo similar a los resultados de la regresión, explicando que un 97.56% es la proporción de los cambios en las ventas que pueden ser explicados por el modelo.

El C_p -Mallow del paso 14 es cercano al número de parámetros del modelo, ya que $p+1=9$ y $C_p=14.3$. Los 8 predictores que se incluyen son: las variables indicadoras de sucursal (x_{20} hasta x_{24}), cantidad de producto vendido (x_4), *add loss* (x_{10}), y margen de ganancia (x_{17}).

La ecuación de regresión que resulta de aplicar el procedimiento *Regression* en Minitab con los datos sin las variables que deben ser removidas se encuentra en la etapa 4 del presente capítulo.

Stepwise Regression: Y versus X21, X22, ...			
Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15			
Response is Y on 15 predictors, with N = 936			
Step	14	Step	14
Constant	-17910		
X17	0.176	X21	290270
T-Value	4.48	T-Value	21.87
P-Value	0.000	P-Value	0.000
X24	-146032	X4	6.23
T-Value	-4.65	T-Value	10.08
P-Value	0.000	P-Value	0.000
X10	0.241	X23	36605
T-Value	5.02	T-Value	3.56
P-Value	0.000	P-Value	0.000
X9		X7	
T-Value		T-Value	
P-Value		P-Value	
X18		X3	
T-Value		T-Value	
P-Value		P-Value	
X20	918723	S	74363
T-Value	23.42	R-Sq	97.58
P-Value	0.000	R-Sq(adj)	97.56
X22	687501	Mallows C-p	14.3
T-Value	19.19		
P-Value	0.000		

Figura 4.4. Procedimiento *Stepwise* - A.

4.2.4.3.2. Modelo B

La ecuación de regresión resultante de la muestra de datos de la sucursal B es la siguiente:

$$Y = - 28699 + 895395 X_{20} + 192914 X_{21} + 522846 X_{22} + 3166 X_{23} - 202873 X_{24} + 17366 X_3 + 4.26 X_4 - 3372 X_5 + 9407 X_6 - 14483 X_7 + 125 X_8 - 0.304 X_9 + 0.286 X_{10} + 0.242 X_{17} + 76.2 X_{18}$$

La variable x_{25} que corresponde a tortillería, fue removida de la ecuación de regresión debido a que se encuentra altamente correlacionada con otras variables independientes.

Los estimados del modelo se indican en la tabla 4.11.

Predictor	Coef	SE Coef	T	P
Constant	-28699	9691	-2.96	0.003
X ₂₀	895395	30073	29.77	0.000
X ₂₁	192914	9506	20.29	0.000
X ₂₂	522846	26342	19.85	0.000
X ₂₃	3166	6997	0.45	0.651
X ₂₄	-202873	39582	-5.13	0.000
X ₃	17366	3162	5.49	0.000
X ₄	4.2555	0.4017	10.59	0.000
X ₅	-3372	1613	-2.09	0.037
X ₆	9407	2421	3.89	0.000
X ₇	-14483	5922	-2.45	0.015
X ₈	124.75	30.49	4.09	0.000
X ₉	-0.3041	0.4961	-0.61	0.540
X ₁₀	0.28635	0.03605	7.94	0.000
X ₁₆	0.24236	0.01950	12.43	0.000
X ₁₇	76.25	22.09	3.45	0.001

Tabla 4.11. Coeficientes del modelo etapa 3 - B.

Los coeficientes estimados del modelo indican que solamente la variable indicadora del departamento de café (x_{23}) y nómina por departamento (x_9) no son significativas a un nivel α de 0.05.

Los estimados del modelo de regresión para B se presentan en la figura 4.5.

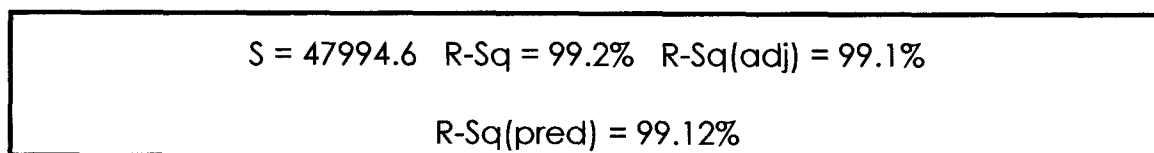


Figura 4.5. Estimados del modelo de regresión etapa 3 - B.

La variación de las ventas con respecto a la media de la distribución de los datos es de 47,994.6 pesos.

El R^2 indica que el 99.2% de la variabilidad de las ventas es explicada por su relación lineal con las variables explicativas del modelo. Mientras que $R^2_{ajustado}$ tienen en cuenta el tamaño del conjunto de datos, su valor es ligeramente inferior al de su correspondiente R^2 con un 99.1%. El valor de $R^2_{predicción}$ indica que el modelo es bueno en un 99.12% para predecir nuevas observaciones.

4.2.4.3.2.1. Método Stepwise

El procedimiento *stepwise* con el método de eliminación paso a paso, se utiliza en los datos de la sucursal B con valores de entrada de $\alpha = 0.15$ y $F = 4$, y de salida de $\alpha = 0.15$ y $F = 4$.

La pantalla de resultados de Minitab arroja los resultados de todos los pasos del método *stepwise*, para fines prácticos se presenta solamente el valor del paso para el mejor modelo, el procedimiento completo se presenta en el anexo 11.

Para la sucursal B, se utiliza una muestra de datos de un total de 936 datos con una la variable de respuesta y (ventas) sobre 15 variables explicativas, al igual que en el procedimiento de la tienda A, no se consideran los regresores x_{11} hasta x_{16} debido a que corresponden a los niveles de la variable indicadora de sucursal denotadas por valores de 0 y 1.

De los resultados del método de eliminación paso a paso presentado en la figura 4.6., se concluye la necesidad de remover de la ecuación de regresión las variables correspondientes a departamento de café (x_{23}), cantidad de socios parciales (x_5), cantidad de bajas (x_7), nómina por departamento (x_9), y ventas por hora hombre (x_{18}).

Los estimados del modelo indican que la variación de las ventas con respecto a la media de la distribución de los datos es de \$48,373 pesos, este valor disminuye en comparación con el valor de la desviación estándar (S) resultante de la regresión. El coeficiente de determinación múltiple indica que el 99.14% de varianza en las ventas se justifica por las con las variables independientes del modelo.

Así mismo, el $R^2_{ajustado}$ sigue siendo similar a los resultados de la regresión, explicando que un 99.13% es la proporción de los cambios en las ventas que pueden ser explicados por el modelo.

De acuerdo al C_p -Mallow, el mejor modelo es aquel que tiene las siguientes 10 variables predictoras: variables indicadoras de departamento de carnes, pescados y mariscos, delicatessen, panadería y tortillería, así como quincenas (x_3), cantidad de producto vendido (x_4), cantidad de socios completos (x_6), horas disponibles de trabajo efectivo (x_8), *add loss* (x_{10}), y margen de ganancia (x_{17}) con un valor de C_p igual a 25.6.

Stepwise Regression: Y versus X21, X22, ...			
Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15			
Response is Y on 15 predictors, with N = 936			
Step	14	Step	14
Constant	-29906		
X17	0.248	X21	208094
T-Value	13.28	T-Value	31.45
P-Value	0.000	P-Value	0.000
X24	-246048	X4	4.32
T-Value	-7.49	T-Value	12.17
P-Value	0.000	P-Value	0.000
X10	0.290	X3	17623
T-Value	8.22	T-Value	5.54
P-Value	0.000	P-Value	0.000
X8	88	X6	9975
T-Value	3.51	T-Value	4.60
P-Value	0.000	P-Value	0.000
X18		X7	
T-Value		T-Value	
P-Value		P-Value	
X20	889668	X5	
T-Value	32.56	T-Value	
P-Value	0.000	P-Value	
X22	594276	S	48373
T-Value	39.50	R-Sq	99.14
P-Value	0.000	R-Sq(adj)	99.13
		Mallows C-p	25.6

Figura 4.6. Procedimiento Stepwise - B.

4.2.4.3.3. Modelo D

La ecuación de regresión es:

$$Y = - 31360 + 569546 X_{20} + 94271 X_{21} + 381572 X_{22} - 3717 X_{23} - 215637 X_{24} + 14721 X_3 + 2.84 X_4 + 4160 X_5 - 4902 X_6 - 4846 X_7 + 137 X_8 + 0.256 X_9 + 0.263 X_{10} + 0.135 X_{17} + 168 X_{18}$$

Las variables indicadoras de sucursal (x_{11} hasta x_{16}) son eliminadas de la ecuación debido a que los valores de las columnas son igual a cero. Por otro lado, la variable de departamento correspondiente a Tortillería se remueve del modelo debido a que se encuentra altamente correlacionada con otras variables independientes.

La tabla 4.12. muestra los coeficientes estimados del modelo D.

Predictor	Coef	SE Coef	T	P
Constant	-31360	9022	-3.48	0.001
X ₂₀	569546	31584	18.03	0.000
X ₂₁	94271	8461	11.14	0.000
X ₂₂	381572	20255	18.84	0.000
X ₂₃	-3717	7151	-0.52	0.603
X ₂₄	-215637	47972	-4.50	0.000
X ₃	14721	2512	5.86	0.000
X ₄	2.8407	0.3830	7.42	0.000
X ₅	4159.6	952.0	4.37	0.000
X ₆	-4902	4752	-1.03	0.303
X ₇	-4846	8252	-0.59	0.557
X ₈	136.82	39.52	3.46	0.001
X ₉	0.2564	0.5915	0.43	0.665
X ₁₀	0.26287	0.03630	7.24	0.000
X ₁₇	0.13467	0.01662	8.10	0.000
X ₁₈	168.30	26.58	6.33	0.000

Tabla 4.12. Coeficientes del modelo etapa 3 - D.

Los variables de departamento de Café, cantidad de socios completos (x_6), cantidad de bajas (x_7), y nómina por departamento (x_9), no son significativos a un nivel α de 0.05.

Los estimados del modelo de regresión para D se presentan en la figura 4.7.

$$S = 38080.1 \quad R\text{-Sq} = 98.8\% \quad R\text{-Sq}(\text{adj}) = 98.8\%$$

$$R\text{-Sq}(\text{pred}) = 98.77\%$$

Figura 4.7. Estimados del modelo de regresión etapa 3 - D.

La variación de las ventas con respecto a la media de la distribución de los datos es de 38,080.1 pesos. Existe un 98.8% de varianza justificado por las variables independientes estudiadas, mientras que el valor ajustado del coeficiente de determinación permanece siendo 98.8%. El valor de $R^2_{\text{predicción}}$ indica un 98.77% de utilidad potencial del modelo para identificar observaciones en las que es probable que el modelo produzca buenas predicciones futuras.

4.2.4.3.3.1. Método Stepwise

Para analizar la muestra de 936 datos de la sucursal D con una variable de respuesta y (ventas) sobre 15 variables predictivas, se aplica el procedimiento *stepwise* con el método de eliminación "hacia adelante y hacia atrás", con valores de entrada de $\alpha = 0.15$ y $F = 4$, y de salida de $\alpha = 0.15$ y $F = 4$.

La pantalla de resultados de Minitab arroja los resultados de todos los pasos del método *stepwise*, para fines prácticos se presenta solamente los valores de los pasos finales, siendo el 13 el mejor modelo, el procedimiento completo se presenta en el anexo 12.

De los resultados del método de eliminación paso a paso presentado en la figura 4.8., se concluye la necesidad de remover de la ecuación de regresión las variables: cantidad de socios completos (x_6), cantidad de bajas (x_7), y nómina por departamento (x_9).

Los estimados del modelo indican que la variación de las ventas con respecto a la media de la distribución de los datos es de \$38,080 pesos, este valor disminuye en comparación con el valor de la desviación estándar (S) resultante de la regresión. El R^2 indica que el 98.84% de la variabilidad de las ventas es explicada por su relación lineal con las variables independientes del modelo.

Así mismo, el R^2_{ajustado} sigue siendo similar a los resultados de la regresión, explicando que un 98.82% es la proporción de los cambios en las ventas que pueden ser explicados por el modelo.

El estadístico C_p -Mallow indica que el mejor modelo es el que se forma a partir de los regresores que se indican a continuación:

De acuerdo al C_p -Mallow, el mejor modelo es aquel que tiene las siguientes 10 variables predictoras: variables indicadoras de departamento Carnes, Pescados y Mariscos, y Panadería, quincenas (x_3), cantidad de producto vendido (x_4), cantidad de socios parciales (x_5), horas disponibles de trabajo efectivo (x_8), *add loss* (x_{10}), margen de ganancia (x_{17}), y venta por hora hombre (x_{18}) con un valor de C_p igual a 12, muy próximo a $p+1=11$.

Stepwise Regression: Y versus X21, X22, ...					
Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15					
Response is Y on 15 predictors, with N = 936					
Step	13	14	Step	13	14
Constant	-32813	-31407			
X10	0.270	0.262	X20	590937	577625
T-Value	7.51	7.25	T-Value	22.49	21.21
P-Value	0.000	0.000	P-Value	0.000	0.000
X17	0.130	0.133	X22	378602	386141
T-Value	7.92	8.08	T-Value	20.40	20.34
P-Value	0.000	0.000	P-Value	0.000	0.000
X24	-194988	-205837	X21	98495	96430
T-Value	-4.80	-5.02	T-Value	12.70	12.32
P-Value	0.000	0.000	P-Value	0.000	0.000
X9			X3	14542	14617
T-Value			T-Value	5.80	5.83
P-Value			P-Value	0.000	0.000
X18	172	167	X5	3381	4028
T-Value	6.55	6.32	T-Value	4.01	4.41
P-Value	0.000	0.000	P-Value	0.000	0.000
X8	136	143	X6		-6461
T-Value	3.72	3.89	T- Value		-1.83
P-Value	0.000	0.000	P-Value		0.000
X4	3.05	2.90	S	38080	38032
T-Value	8.44	7.86	R-Sq	98.84	98.84
P-Value	0.000	0.000	R-Sq(adj)	98.82	98.82
			Mallows C-p	12.0	10.7

Figura 4.8. Procedimiento Stepwise - D.

4.2.4.3.4. Modelo Combo

En el modelo combo se toma en cuenta las sucursales C, E y F. Debido a que se eliminan los valores iguales a cero correspondientes a la tienda E del departamento de Café, la muestra que se analiza es de 2648 datos para la variable de respuesta y con respecto a 20 variables explicativas.

La ecuación de regresión resultante es la siguiente:

$$Y = 28459 + 15645 X_{13} + 4716 X_{15} + 414791 X_{20} + 47785 X_{21} + 191230 X_{22} - 4181 X_{23} - 127750 X_{24} + 17633 X_3 + 1.64 X_4 - 3108 X_5 + 2304 X_6 + 5081 X_7 + 127 X_8 + 1.14 X_9 + 0.390 X_{10} + 0.302 X_{17} + 55.6 X_{18}$$

De acuerdo a los resultados del paquete de cómputo, las variables X_{16} y X_{25} se encuentran altamente correlacionadas con otra variable x ; por tanto, dicha variable ha sido eliminada de la ecuación. Por otro lado, las variables de la sucursal A (x_{11}), B (x_{12}) y D (x_{14}) fueron removidas del modelo debido a que las columnas se conforman de valores iguales a cero.

Los coeficientes de regresión del modelo se observan en la tabla 4.13.

Predictor	Coef	SE Coef	T	P
Constant	-28459	4240	-6.71	0.000
X_{13}	15645	2120	7.38	0.000
X_{15}	4716	2147	2.20	0.028
X_{20}	414791	11556	35.89	0.000
X_{21}	47785	3219	14.84	0.000
X_{22}	191230	6717	28.47	0.000
X_{23}	-4181	3464	-1.21	0.228
X_{24}	-127750	18273	-6.99	0.000
X_3	17633	1481	11.91	0.000
X_4	1.6445	0.1525	10.78	0.000
X_5	-3108.0	688.6	-4.51	0.000
X_6	2304	1262	1.83	0.068
X_7	5081	4886	1.04	0.298
X_8	126.92	21.59	5.88	0.000
X_9	1.1428	0.4030	2.84	0.005
X_{10}	0.38976	0.02514	15.50	0.000
X_{17}	0.30230	0.01484	20.37	0.000
X_{18}	55.643	5.743	9.69	0.000

Tabla 4.13. Coeficientes del modelo etapa 3 - Combo.

Los coeficientes estimados de la variable indicadora del departamento Café (x_{23}), así como la cantidad de bajas (x_7) con valores-p iguales a 0.228 y 0.298, respectivamente, indican que dichos regresores no están relacionados con las ventas a un nivel de significancia α de 0.05.

Los estimados del modelo de regresión del conjunto de sucursales combo son:

$S = 37842.1$ $R\text{-Sq} = 97.9\%$ $R\text{-Sq}(\text{adj}) = 97.9\%$
$R\text{-Sq}(\text{pred}) = 97.88\%$

Figura 4.9. Estimados del modelo de regresión etapa 3 - Combo.

La desviación estándar (S) indica que la variación esperada de las ventas con respecto de la media aritmética de los datos es de \$37,842 pesos.

El coeficiente de determinación (R^2) expresa que el 97.99% de la variación total de las ventas puede ser explicada por el modelo cuantitativo. El coeficiente de determinación ajustado siendo un valor más realista que R^2 , indica que el 97.99% es la proporción de cambio de las ventas que se aplica al modelo.

4.2.4.3.4.1. Método Stepwise

Al igual que en los grupos anteriores, se aplica el procedimiento *stepwise* con el método paso a paso, utilizando valores de entrada de $\alpha = 0.15$ y $F = 4$, así como valores de salida de $\alpha = 0.15$ y $F = 4$. La pantalla de resultados de Minitab arroja los resultados de todos los pasos del método *stepwise*, para fines prácticos se presenta solamente los valores de los pasos finales, el procedimiento completo se presenta en el anexo 13.

Las variables indicadoras de los niveles de sucursal (x_{11} hasta x_{15}) son omitidas en el procedimiento. De la figura 4.10. se determinan las variables independientes que deben ser eliminadas de la ecuación de regresión final. Para el caso del Combo se remueven los datos de las variables: sucursal E, departamento Café, cantidad de socios completos (x_6), cantidad de bajas (x_7), y nómina por departamento (x_9).

Al aplicar el método *stepwise* se calculan los estimados del modelo de regresión, se observa que la variación de las ventas con respecto a la media de la distribución de los datos es de \$37,922 pesos. El coeficiente

de determinación múltiple de 97.91% indica el porcentaje de los datos de la muestra que se ajustan al modelo. Así mismo, el $R^2_{ajustado}$ sigue siendo similar a los resultados de la regresión, explicando que un 97.90% es la proporción de los cambios en las ventas que pueden ser explicados por el modelo.

Los 12 predictores que se incluyen son: las variables indicadoras de la sucursal C, así como los departamentos Carnes, Pescados y Mariscos, Delicatessen, y Panadería, quincenas (x_3), cantidad de producto vendido (x_4), cantidad de socios parciales (x_5), horas disponibles de trabajo efectivo (x_8), *add loss* (x_{10}), margen de ganancia (x_{17}), y venta por hora hombre (x_{18}).

Stepwise Regression: Y versus X21, X22, ...			
Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15			
Response is Y on 15 predictors, with N = 2648			
Step	14	Step	14
Constant	-29051		
X17	0.313	X13	13020
T-Value	21.62	T-Value	7.85
P-Value	0.000	P-Value	0.000
X10	0.386	X18	57.8
T-Value	15.36	T-Value	10.71
P-Value	0.000	P-Value	0.000
X20	418070	X23	
T-Value	41.97	T-Value	
P-Value	0.000	P-Value	
X22	191656	X5	-2845
T-Value	29.28	T-Value	-4.44
P-Value	0.000	P-Value	0.000
X21	51179	X8	164
T-Value	18.57	T-Value	10.41
P-Value	0.000	P-Value	0.000
X3	17593	X24	-110733
T-Value	11.86	T-Value	-8.13
P-Value	0.000	P-Value	0.000
X4	1.66	S	37922
T-Value	12.05	R-Sq	97.91
P-Value	0.000	R-Sq(adj)	97.90
		Mallows C-p	24.1

Figura 4.10. Procedimiento Stepwise - Combo.

4.2.4.4. Etapa 4: Eliminación de variables no significativas

En la etapa anterior se identificaron los grupos de sucursales para la modelación de las ecuaciones de regresión correspondientes. En la etapa 4 se procede a la eliminación de las variables no significativas de cada conjunto de datos, tomando como base los resultados del método *stepwise*.

Los modelos finales para cada grupo de tiendas, se estudian a partir de pruebas estadísticas y análisis de residuales que permiten determinar normalidad, homocedasticidad, aleatoriedad e independencia. Los resultados se presentan en las siguientes páginas.

4.2.4.4.1. Modelo final A

El modelo para la sucursal A se determina a través de diversas pruebas de regresión que incluyen el análisis de las variables que resultaron ser significativas a un nivel de confianza del 95% en el procedimiento *stepwise*. Se considera necesario estudiar la influencia de las interacciones entre variables, así como los gráficos que demuestran un comportamiento cuadrático y/o cúbico de algunos regresores de interés. Además se lleva a cabo el procedimiento *stepwise* mostrado en el anexo 14 tomando en cuenta el paso 18 para definir las variables de entrada al modelo.

Después del extendido análisis de las variables seleccionadas se define que la ecuación final de la tienda A es:

$$Y = - 43462 + 8.26 X_4 + 70.1 X_8 + 0.186 X_{17} + 242426 X_{20} + 41300 X_{23} + 29.2 X_4 X_{20} + 50.3 X_4 X_{21} + 27.5 X_4 X_{22} - 0.000149 X_4^2 + 0.000000000943 X_4^3$$

El modelo muestra que existe una relación inversamente proporcional entre las ventas (Y) y las variable correspondiente a cantidad de producto vendido elevado a la segunda potencia (x_4^2).

Por otro lado, las ventas tienen una relación directamente proporcional con los departamentos de Carnes (x_{20}) y Café (x_{23}), con la cantidad de producto vendido (x_4), horas disponibles de trabajo efectivo (x_8), margen de ganancia (x_{17}), cantidad de producto vendido elevado al cubo (x_4^3), así como con las interacciones conformadas por las variables cantidad de producto vendido con Carnes ($x_4 x_{20}$), Pescados y Mariscos ($x_4 x_{21}$), y

Delicatessen (x_4x_{22}), debido a que el signo de los parámetros de regresión es positivo.

La interpretación de los coeficientes de regresión indica que el intercepto negativo de 43462 es el valor promedio de las ventas (Y) cuando las x_i son cero. Mientras que los estimados $\hat{\beta}_j$ muestran el cambio promedio de las ventas cuando x_i se incrementa en una unidad.

4.2.4.4.1.1. Estimación de los parámetros del modelo

Los coeficientes del modelo, así como los valores correspondientes al error estándar de los parámetros, se presentan en la tabla 4.14. Se concluye que los valores-p de todos los estimados son significativos a un nivel de significancia α de 0.05.

Los estimados del modelo de regresión de la figura 4.11., indican que la variación esperada de las ventas con respecto a la media de la distribución de los datos es de 36868.0 pesos. El valor de los errores comparado con las observaciones de las ventas es relativamente pequeño, por lo que se sugiere que no afecta significativamente al estudio.

El coeficiente de determinación (R^2) expresa que el 99.4% de la variación total de las ventas puede ser explicada por el modelo cuantitativo. El coeficiente de determinación ajustado (R^2_{ajustado}) de 99.4% es un valor alto, por lo que es posible deducir que los cambios que sufre la variable ventas (y) se pueden calcular a partir de las variables independientes.

El valor R^2 de predicción se calcula a partir del estadístico PRESS e indica que tan bueno es el modelo para pronosticar ventas para nuevas observaciones. El $R^2_{\text{predicción}}$ de 99.39% es cercano a los valores de R^2 y R^2_{ajustado} , se sugiere que el modelo es bueno para predecir nuevos datos.

Predictor	Coef	SE Coef	T	P
Constant	-43462	13855	-3.14	0.002
X ₄	8.260	1.503	5.50	0.000
X ₁₇	0.18553	0.01569	11.82	0.000
X ₈	70.11	14.80	4.74	0.000
X ₂₀	242426	30169	8.04	0.000
X ₂₃	41300	9609	4.30	0.000
X ₄ ²	-0.00015	3.39E-10	-4.41	0.000
X ₄ ³	9.43E-10	2.26E-10	4.17	0.000
X ₄ X ₂₀	29.231	1.046	27.94	0.000
X ₄ X ₂₁	50.303	1.336	37.66	0.000
X ₄ X ₂₂	27.5431	0.4828	57.05	0.000

Tabla 4.14. Coeficientes del modelo etapa 4 - A.

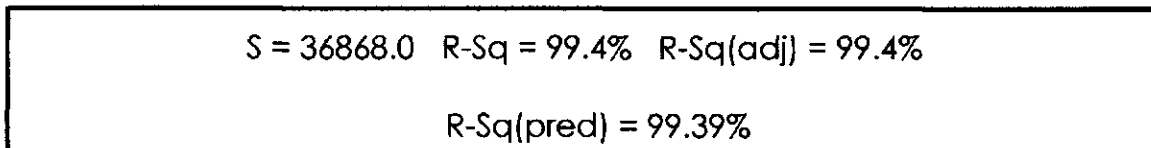


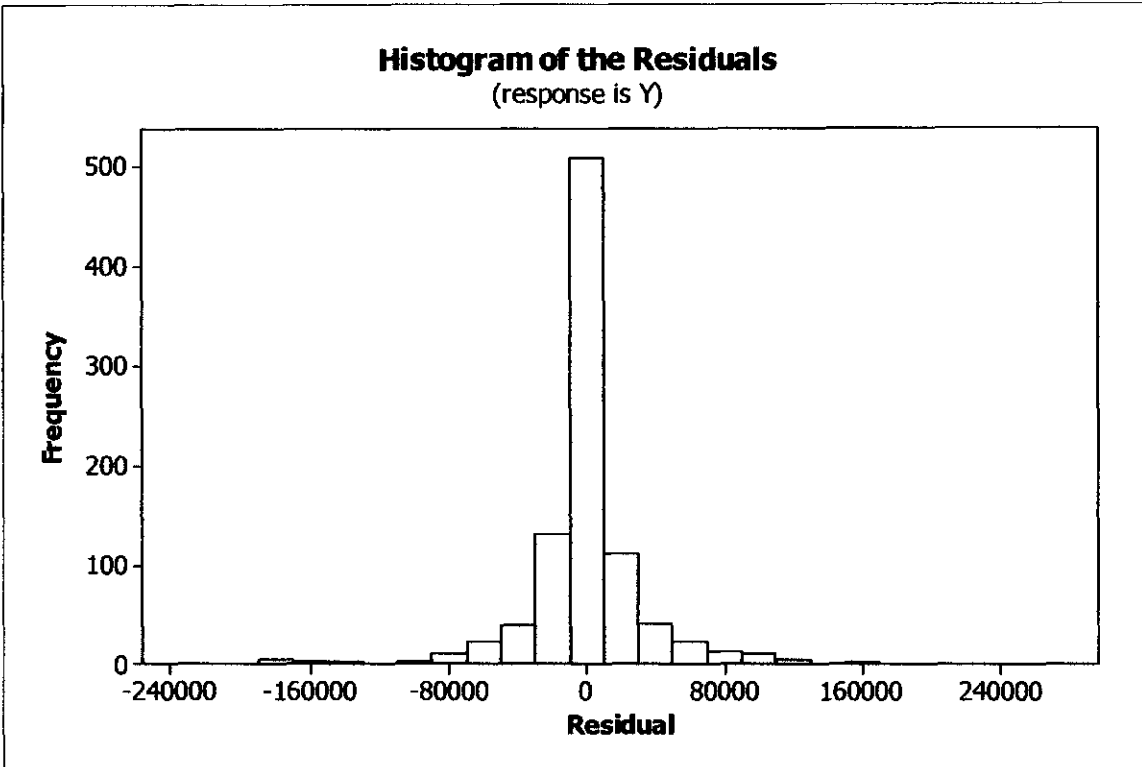
Figura 4.11. Estimados del modelo de regresión etapa 4 - A.

4.2.4.4.1.2. Supuestos del análisis de regresión

Los supuestos del análisis de regresión permiten comprobar si los errores son igual a $\varepsilon \sim N(0, \sigma^2)$, lo que implica que deben ser: aleatorios e independientes entres sí, se ajustan a una distribución normal, su media es cero y presentan varianza constante.

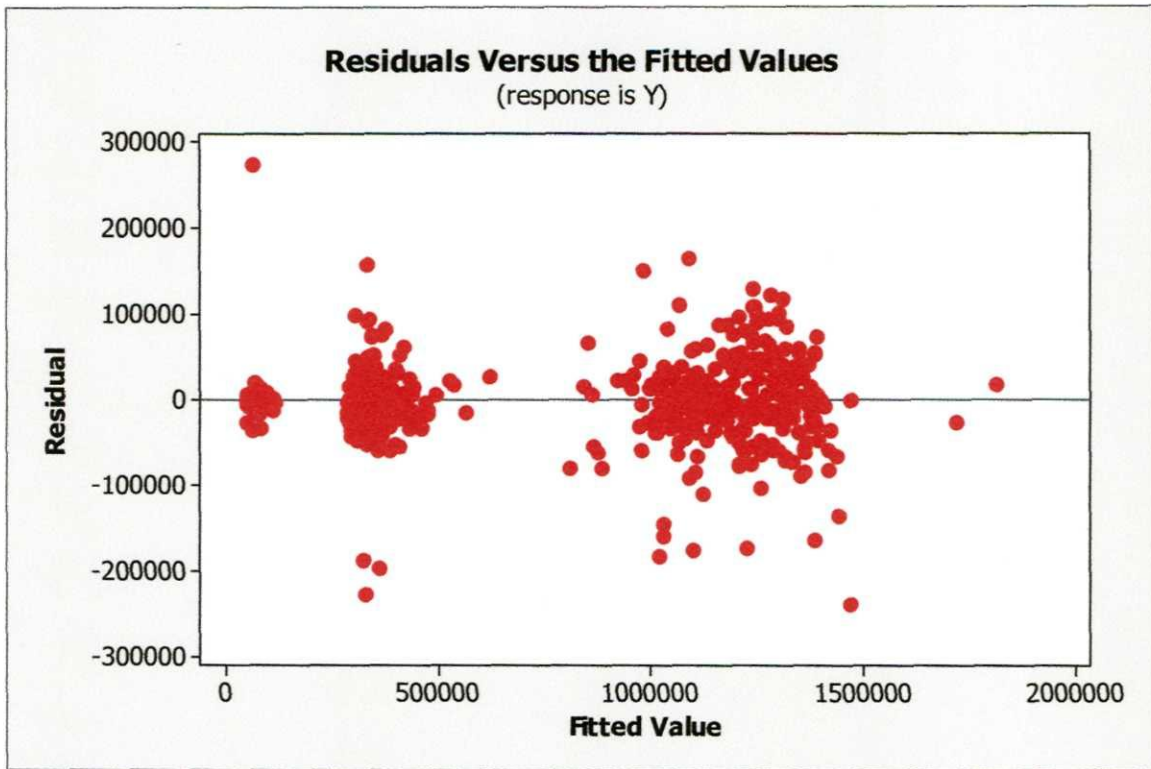
Análisis de residuales

Para análisis posteriores se utilizan las siguientes gráficas de residuos.



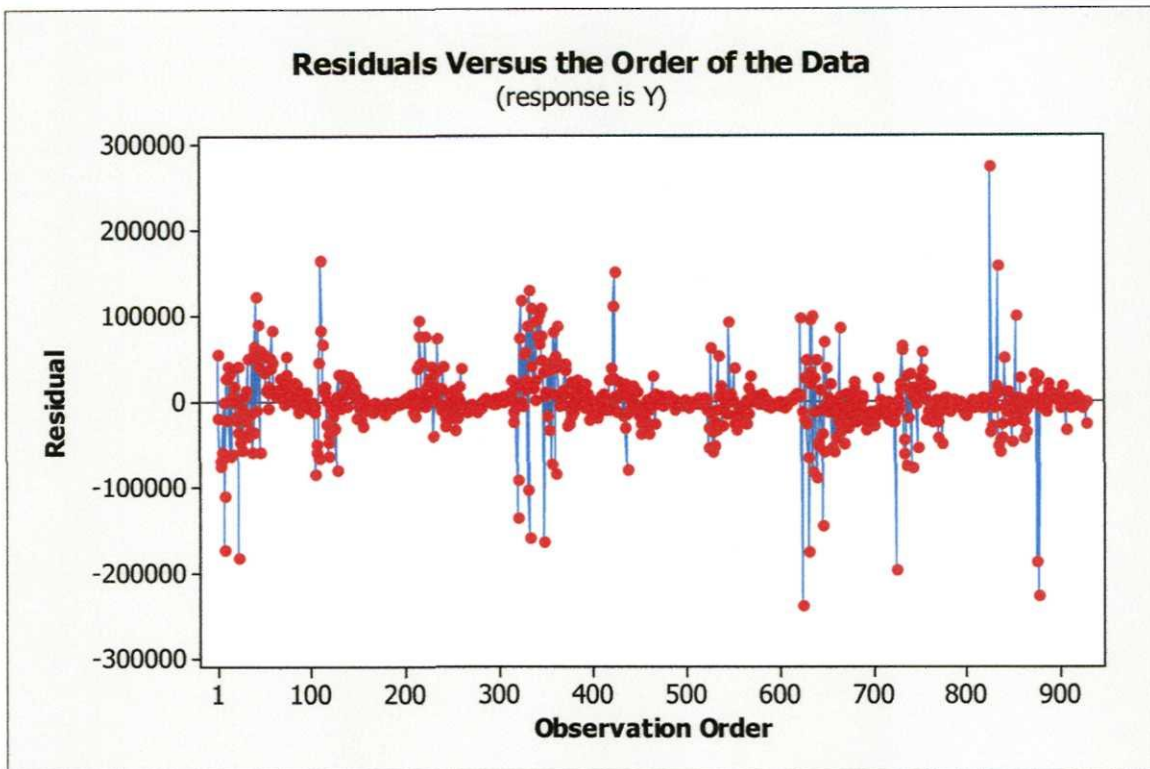
Gráfica 4.10. Histoarama de residuales etapa 4 - A.

El histograma de residuales arroja información sobre la existencia de datos atípicos debido a la presencia de barras aisladas en los extremos de la gráfica. Por este motivo se procede al tratamiento de las observaciones atípicas por medio de su eliminación de la base de datos.



Gráfica 4.11. Residuales versus los valores ajustados etapa 4 - A.

La gráfica de residuales versus los valores ajustados sugiere que la varianza de los errores es constante debido a que no se observa una tendencia en el comportamiento de los datos.



Gráfica 4.12. Residuales versus el orden de los datos etapa 4 - A.

La gráfica de residuales versus el orden de los datos muestra la presencia de diversas observaciones que se alejan de la línea central y que se encuentran separados significativamente al resto de los puntos.

A continuación se revisan los supuestos por medio de análisis más formales.

Normalidad

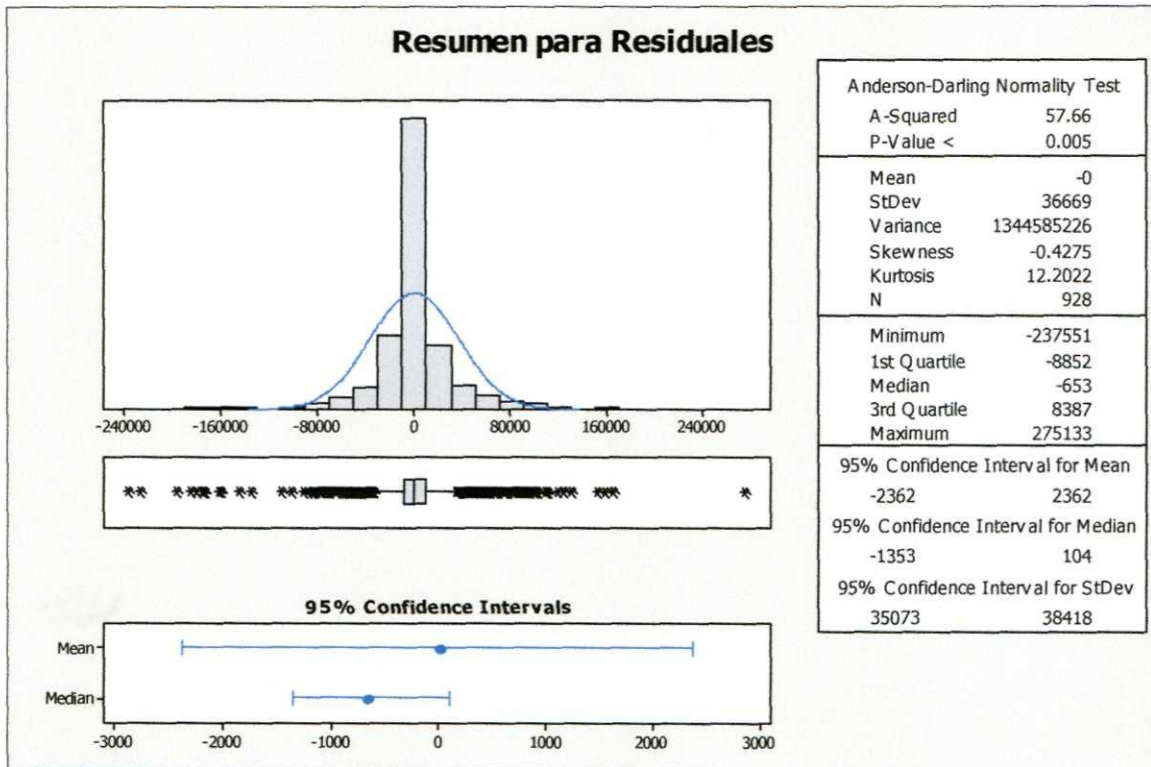
Para verificar la normalidad de los residuales, se presenta la gráfica 4.13. en Minitab.

La prueba de normalidad Anderson-Darling, revela que con un valor-p menor a 0.005, existe evidencia estadística suficiente para rechazar el supuesto de normalidad. La misma conclusión puede generarse a partir de la comprobación del supuesto por medio del análisis de los residuos estandarizados, los cuales se ajustan a una distribución normal cuando los valores se encuentran en el rango ± 2 . Al estudiar el anexo 15, se observa que los valores no se encuentran en dicho rango, debido a la presencia

de las observaciones que se identifican como puntos atípicos del conjunto de datos.

A pesar de que los residuos no son normales, lo más importante de este diagrama es representado por kurtosis, la cual es una medida estadística que describe el apuntamiento o ensanchamiento de una cierta distribución con respecto a una distribución normal. En este caso se presenta una kurtosis positiva excesiva, lo cual indica que hay una mayor probabilidad de que los residuos se encuentren más alejados de la media que en una distribución normal. Sin embargo, esto es una característica deseable del modelo dado a que la mayor parte de los residuos se distribuyen alrededor de la media igual a cero. *

Se observa además en el histograma que el valor del sesgo de los residuales es mínimo, por lo que se sugiere que los errores son simétricos y con varianza constante, lo cual sugiere el cumplimiento de homocedasticidad, lo cual se considera de varianza "homogénea" cuando en la gráfica de los residuos se encuentran distribuidos con cierta uniformidad.



Gráfica 4.13. Resumen para residuales etapa 4 - A.

* Fe de erratas

Aleatoriedad e independencia de los errores

Debido a que los datos no siguen una distribución normal, es necesario probar el supuesto de aleatoriedad e independencia de los errores por medio de la aplicación de una prueba no paramétrica, lo cual implica que no existe una asunción sobre una distribución en específico para la población, además los resultados de estas pruebas son más robustos contra la violación de los supuestos.

La prueba no paramétrica a utilizar es la "prueba de corridas", la cual permite observar si los datos de la población son aleatorios. Los resultados arrojados por Minitab son:

```
Runs Test: RESI4  
  
Runs test for RESI4  
  
Runs above and below K = -3.42956E-11  
  
The observed number of runs = 305  
The expected number of runs = 463.759  
440 observations above K, 488 below  
P-value = 0.000
```

Figura 4.12. Resultados de Run Test etapa 4 - A.

Debido a que no se especifica la opción de otro valor que no sea la media como criterio de comparación (K), la media igual a $-3.42956E-11$ es utilizada. Se realizaron 305 corridas. Para determinar un número inusual de corridas, se calcula el número de observaciones que se encuentran por arriba o debajo de K. El resultado del valor-p (0.000) es menor que el nivel de alpha de 0.05, por tanto no hay evidencia estadística concluyente que indique que los datos no son aleatorios.

Homogeneidad

Se utiliza la gráfica de residuales versus los valores ajustados para determinar que existe varianza homogénea debido a la uniformidad de la distribución de los residuos.

4.2.4.4.1.3. Prueba de hipótesis en regresión lineal múltiple

Prueba de la significancia de la regresión

De la tabla del análisis de varianza, se determina si existe una relación lineal entre la respuesta y cualquiera de las variables regresoras x_1, x_2, \dots, x_k .

Las pruebas de hipótesis que se evalúan son:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_a: \text{por lo menos una } \beta_i \neq 0 \ (i = 1, \dots, k)$$

El estadístico de prueba se calcula según los datos arrojados por Minitab como:

$$F_0 = \frac{MS_R}{MS_{Res}} = \frac{2.09906E + 13}{1359248096} = 15442.81$$

Para probar la hipótesis nula, se utiliza el criterio de rechazo siguiente:

$$F_0 > F_{\alpha, k, n-k-1}$$

Dado un valor de tablas $F_{\alpha, k, n-k-1}$ igual a 2.223812 a un nivel de significancia α igual a 0.05, se concluye que la hipótesis nula $H_0: \beta = 0$ se rechaza, indicando que al menos uno de los regresores x_k contribuye al modelo de forma significativa.

En una prueba equivalente, se puede rechazar H_0 cuando el valor-p de la prueba de F es menor que el nivel de alpha (α). En este caso, $0.00 < 0.05$, por tanto se rechaza la hipótesis nula.

Prueba sobre coeficientes individuales de regresión

Las pruebas de cualquier coeficiente de regresión son:

$$H_0: \beta_j = 0$$

$$H_a: \beta_j \neq 0$$

Se rechaza la hipótesis nula si $|t_0| > t_{\alpha/2, n-k-1}$.

Con n de 928 datos y k igual a 5, se obtiene que: $t_{\alpha/2, n-k-1} = 2.24507$

Predictor	T
Constant	-3.14
X ₄	5.50
X ₁₇	11.82
X ₈	4.74
X ₂₀	8.04
X ₂₃	4.30
X ₄ ²	-4.41
X ₄ ³	4.17
X ₄ X ₂₀	27.94
X ₄ X ₂₁	37.66
X ₄ X ₂₂	57.05

Tabla 4.15. Valores T - A.

De acuerdo a los resultados arrojados por Minitab en la tabla 4.15., encontramos que todos los valores T de los coeficientes de regresión caen en la región de rechazo a un nivel de significancia de 0.05, se concluye que existe una contribución importante de las variables independientes dado los demás regresores del modelo.

4.2.4.4.1.4. Análisis de varianza (ANOVA)

Los resultados del análisis de varianza se enuncian en la tabla 4.16.

Source	DF	SS	MS	F	P
Regression	10	2.09906E+14	2.09906E+13	15442.81	0.000
Residual Error	917	1.24643E+12	1359248096		
Lack of Fit	914	1.24640E+12	1363672353	120.42	0.001
Pure Error	3	33973882	11324627		
Total	927	2.11153E+14			

Tabla 4.16. Tabla ANOVA - A.

El valor-p igual a 0.000 indica que el modelo estimado por el procedimiento de regresión es significativo a un nivel α de 0.05, indicando que al menos un coeficiente es diferente de cero.

Las hipótesis para el ANOVA son:

$$H_0: MS_R \leq MS_{RES}$$

$$H_a: MS_R > MS_{RES}$$

El estadístico de prueba de la ecuación es $F_0 = MS_R/MS_{RES}$, y el criterio de rechazo para H_0 es $F_0 > F_{\alpha, 1, n-2}$.

Según los datos arrojados por Minitab, el $MS_R=2.09906E+13$ y $MS_{RES} = 1359248096$. El estadístico de prueba F_0 es igual a 15442.81 y el valor F de tablas a un nivel de significancia de 0.05 es igual a 3.85151.

Siendo $F_0 > F_{\alpha, 1, n-1}$, la hipótesis nula cae en la región de rechazo, lo que significa que la variación estimada por el modelo es mayor que la variación aleatoria, y por consiguiente se puede decir que el modelo es bueno.

La tabla 4.17 de los grados de libertad y la suma al cuadrado secuencial (Seq SS) de los estimados, indican que la variable cantidad de producto vendido al cubo, no explica una cantidad substancial de varianza única, lo cual sugiere la posibilidad de eliminar dicha variable del modelo.

Source	DF	Seq SS
X ₄	1	1.01418E+13
X ₁₇	1	1.87756E+14
X ₈	1	25694291711
X ₂₀	1	3.27651E+12
X ₂₃	1	1.04755E+12
X ₄ ²	1	3.15416E+11
X ₄ ³	1	14886834565
X ₄ X ₂₀	1	1.21717E+12
X ₄ X ₂₁	1	1.68598E+12
X ₄ X ₂₂	1	4.42463E+12

Tabla 4.17. Grados de libertad y suma al cuadrado secuencial - A.

4.2.4.4.1.5. Intervalos de confianza

A continuación se muestran los intervalos de confianza del modelo A a un nivel de significancia del 95%.

Intervalos de confianza de los coeficientes de regresión

En la tabla 4.18. se presentan los estimados de intervalo de confianza de los coeficientes de regresión β_j .

Coeficientes	95% Intervalo de Confianza para β	
	Límite Inferior	Límite Superior
Constante	-70653.7	-16269.9
X_4	5.311097	11.20912
X_{17}	0.154732	0.216326
X_8	41.06039	99.16894
X_{20}	183218	301634.8
X_{23}	22440.91	60159.24
$X_4^{\wedge}2$	-0.00022	-8.3E-05
$X_4^{\wedge}3$	4.99E-10	1.39E-09
X_4X_{20}	27.17749	31.28407
X_4X_{21}	47.68188	52.92474
X_4X_{22}	26.59563	28.49047

Tabla 4.18. Intervalo de confianza - A.

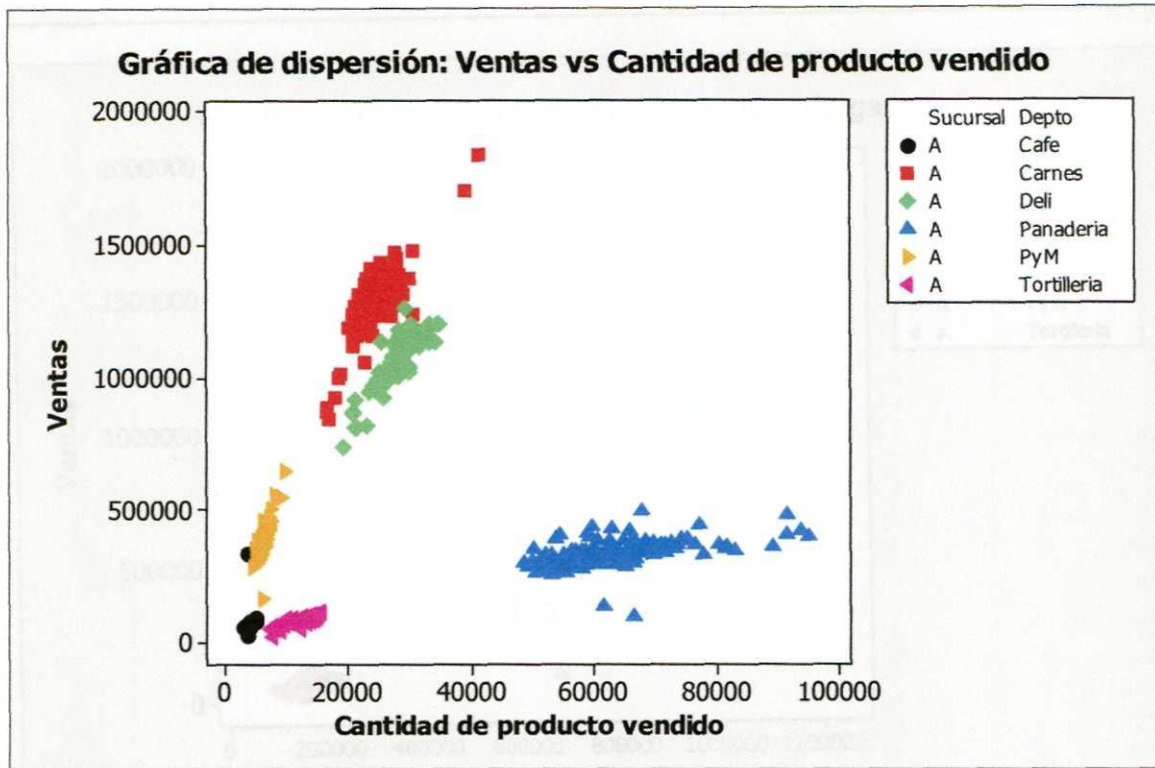
La construcción de estimaciones de intervalos de confianza es útil para establecer si el modelo es preciso. En este caso los intervalos de confianza de las variables no son muy amplios, lo que indica que su aportación al modelo es alta, lo que sugiere que la aportación de las variables al modelo es importante.

Estimación del intervalo de confianza de la respuesta media

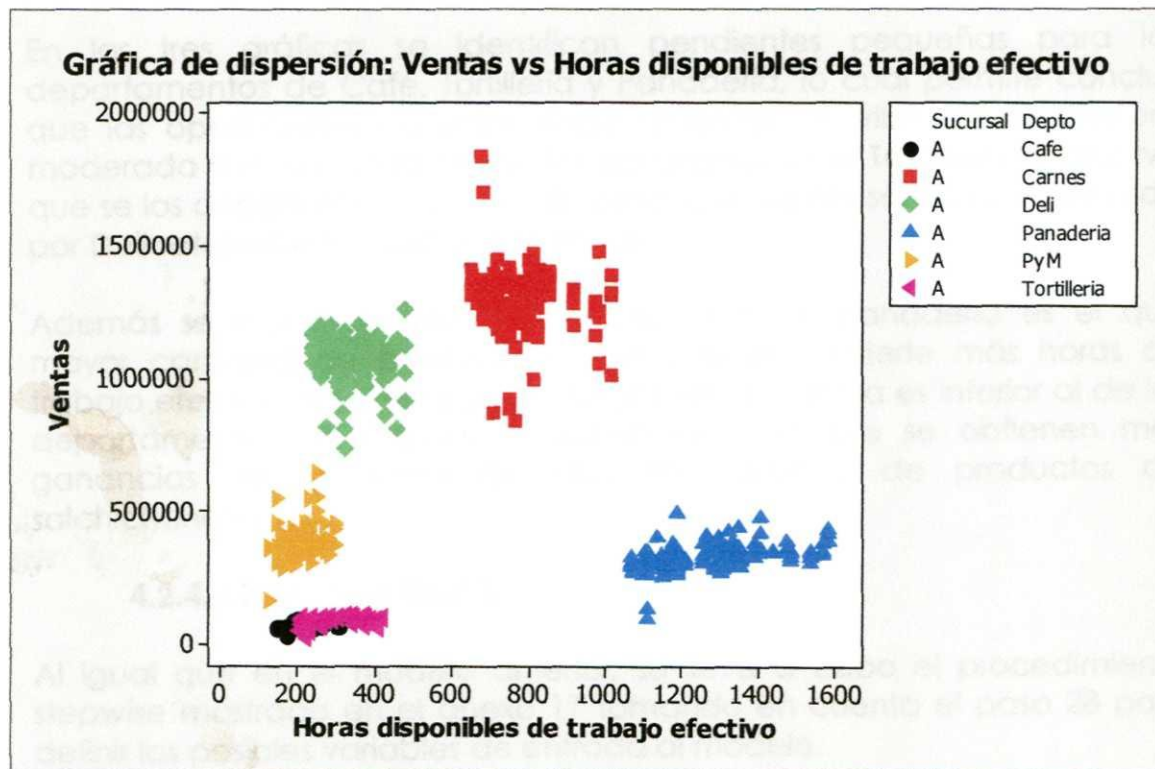
Los intervalos de confianza de la respuesta media, así como los intervalos de predicción del modelo a un nivel del 95%, se presentan en el anexo 16.

4.2.4.4.1.6. Gráficas de dispersión

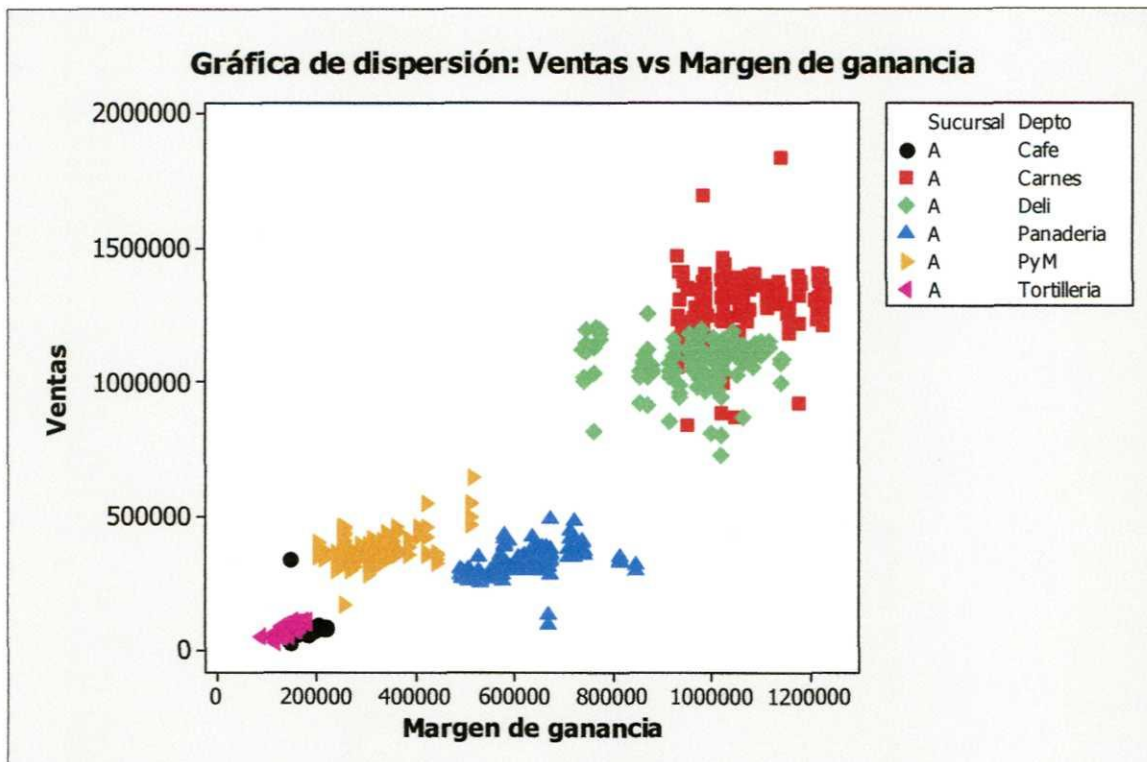
Las gráficas de dispersión 4.14., 4.15. y 4.16., se muestran como representación del comportamiento de los datos del modelo final de A. Es posible observar con claridad la presencia de puntos atípicos, los cuales puedan estar generando problemas en el cumplimiento de los supuestos de regresión lineal.



Gráfica 4.14. Ventas versus cantidad de producto vendido etapa 4 - A.



Gráfica 4.15. Ventas versus horas disponibles de trabajo efectivo etapa 4 - A.



Gráfica 4.16. Ventas versus margen de ganancia etapa 4 - A.

En las tres gráficas se identifican pendientes pequeñas para los departamentos de Café, Tortillería y Panadería, lo cual permite concluir que las operaciones de estos departamentos contribuyen de manera moderada a la cantidad de ventas generadas en A. También se observa que se los departamentos que más producen ventas son Carnes, seguido por Delicatessen y Pescados y Mariscos.

Además se identifica que el departamento de Panadería es el que mayor cantidad de producto vende y el que invierte más horas de trabajo efectivo, sin embargo el margen de ganancia es inferior al de los departamentos de Carnes y Delicatessen, ya que se obtienen más ganancias de la venta de kilos de carne y de productos de salchichonería.

4.2.4.4.2. Modelo final B

Al igual que en el modelo anterior, se lleva a cabo el procedimiento **stepwise** mostrado en el anexo 17 tomando en cuenta el paso 28 para definir las posibles variables de entrada al modelo.

Después de analizar cada regresor se concluye que la ecuación de regresión del modelo final para la tienda B es:

$$Y = 245640 - 271547 X_{21} - 244272 X_{23} - 277859 X_{25} + 6.85 X_4 + 1.41 X_9 - 0.0934 X_{10} + 0.170 X_{17} + 11661 X_3 X_{24} + 28.1 X_4 X_{20} + 40.0 X_8 X_{21} + 17.6 X_4 X_{22} - 10.0 X_4 X_{24} - 208 X_8 X_{20} + 142 X_8 X_{21} + 0.197 X_{10} X_{20} - 0.157 X_{17} X_{22} + 0.000000000228 X_4^3$$

Se observa que las variables indicadoras de sucursal (x_{11} hasta x_{15}) son eliminadas debido a que las columnas de datos se encuentran conformadas por valores iguales a cero.

El modelo muestra que existe una relación inversamente proporcional de las ventas (Y) debido al signo negativo de los coeficientes en las variables *add loss* (x_{10}), Pescados y Mariscos (x_{21}), Café (x_{23}), y Tortillería (x_{25}). También se observa este comportamiento con las interacciones conformadas por las variables cantidad de producto vendido y Panadería ($x_4 x_{24}$), horas disponibles de trabajo efectivo y Carnes ($x_8 x_{20}$), así como margen de ganancia y Delicatessen ($x_{17} x_{22}$).

Por otro lado, las ventas tienen una relación directamente proporcional con cantidad de producto vendido (x_4), nómina por departamento (x_9), y venta por hora hombre (x_{17}), ya que el signo de los parámetros de regresión es positivo. Lo mismo ocurre con las interacciones formadas por quincenas y Panadería ($x_3 x_{24}$), cantidad de producto vendido con: Carnes ($x_4 x_{20}$), Pescados y Mariscos ($x_4 x_{21}$), y Delicatessen ($x_4 x_{22}$); además de horas disponibles de trabajo efectivo con Pescados y Mariscos ($x_8 x_{21}$), *add loss* con Carnes ($x_{10} x_{20}$), y finalmente la cantidad de producto vendido elevado al cubo (x_4^3).

La interpretación de los coeficientes de regresión indica que el intercepto de 864165 es el valor promedio de las ventas (Y) cuando las x_i son cero. Mientras que los estimados $\hat{\beta}_j$ muestran el cambio promedio de las ventas cuando x_i se incrementa en una unidad.

4.2.4.4.2.1. Estimación de los parámetros del modelo

Los coeficientes del modelo, así como los valores correspondientes al error estándar de los parámetros, se presentan en la tabla 4.19. Se concluye que el valor-p del estimado cantidad de bajas (x_7) no es significativo a un nivel de significancia α de 0.05.

Los estimados del modelo de regresión de la figura 4.13., indican que la variación esperada de las ventas con respecto a la media de la distribución de los datos es de 18698.3 pesos. El valor de los errores comparado con las observaciones de las ventas es relativamente pequeño, por lo que se sugiere que no afecta significativamente al estudio.

El coeficiente de determinación R^2 y $R^2_{ajustado}$ muestran un ajuste igual o superior a 99.9%, un valor alto que permite deducir que las ventas pueden ser explicadas por el modelo de regresión.

Por otro lado, el valor R^2 de predicción de 99.86% de ajuste, indicando que el modelo es bueno para pronósticos futuros.

Predictor	Coef	SE Coef	T	P
Constant	245640	18900	13.00	0.000
X ₂₁	-271547	22452	-12.09	0.000
X ₂₃	-244272	18730	-13.04	0.000
X ₂₅	-277859	21669	-12.82	0.000
X ₄	6.8515	0.9671	7.08	0.000
X ₉	1.4072	0.1665	8.45	0.000
X ₁₀	-0.09341	0.02264	-4.13	0.000
X ₁₇	0.170424	0.009069	18.79	0.000
X ₃ X ₂₄	11661	3103	3.76	0.000
X ₄ X ₂₀	28.086	1.165	24.12	0.000
X ₄ X ₂₁	40.032	1.538	26.04	0.000
X ₄ X ₂₂	17.637	1.227	14.37	0.000
X ₄ X ₂₄	-10.027	1.027	-9.76	0.000
X ₈ X ₂₀	-207.64	20.88	-9.95	0.000
X ₈ X ₂₁	141.89	40.40	3.51	0.000
X ₁₀ X ₂₀	0.19676	0.03395	5.80	0.000
X ₁₇ X ₂₂	-0.15693	0.01966	-7.98	0.000
X ₄ ³	2.28E-10	2.33E-11	9.79	0.000

Tabla 4.19. Coeficientes del modelo etapa 4 - B.

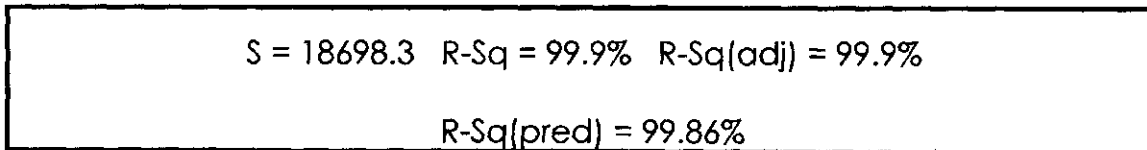


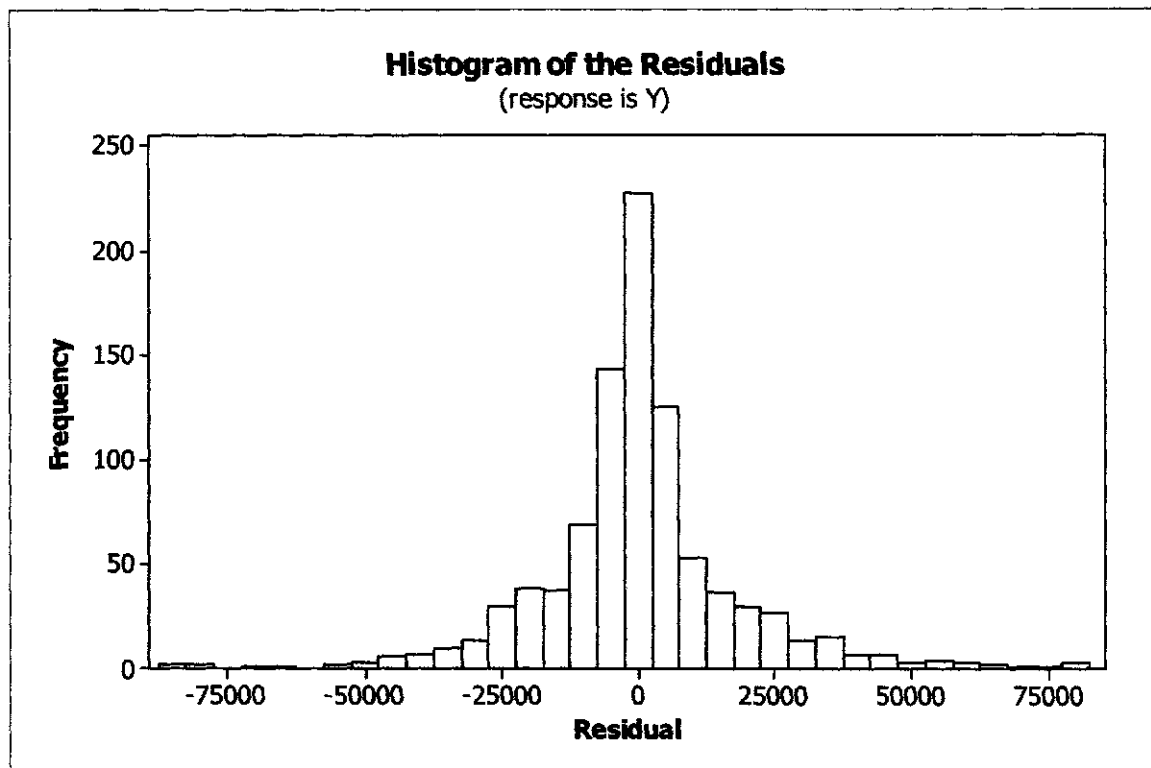
Figura 4.13. Estimados del modelo de regresión etapa 4 - B.

4.2.4.4.2. Supuestos del análisis de regresión

Al igual que en el modelo anterior, se analizan los supuestos del regresión de los datos de la muestra de B.

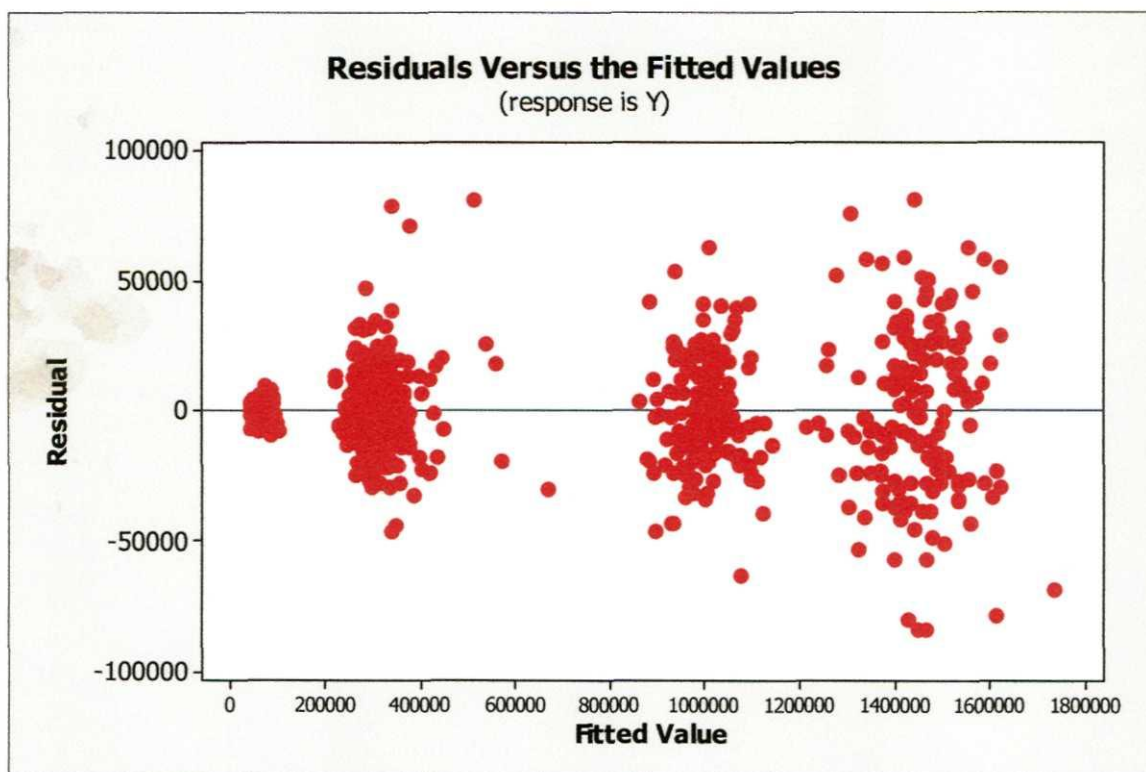
Análisis de residuales

Para análisis posteriores se utilizan las siguientes gráficas de residuos.



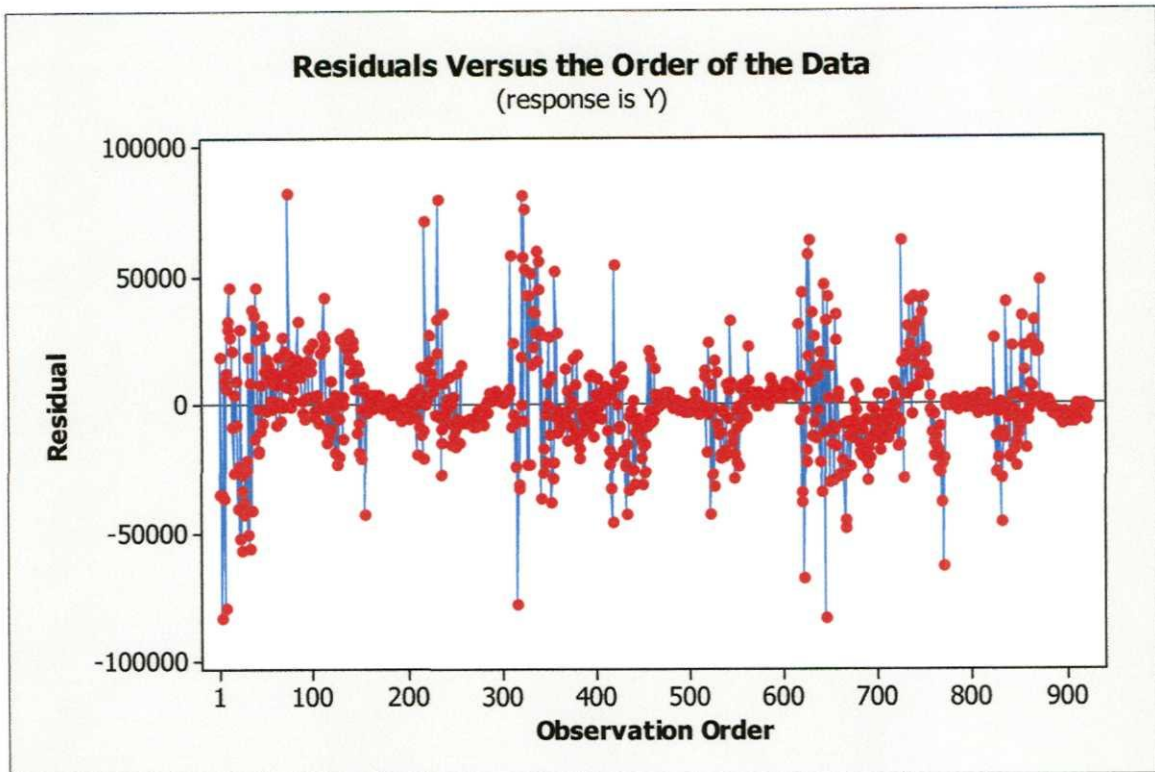
Gráfica 4.17. Gráfica de residuales etapa 4 - B.

El histograma de residuales arroja información sobre la existencia de datos atípicos en los extremos de la gráfica. Por este motivo se procede al tratamiento de las observaciones atípicas por medio de su eliminación de la base de datos.



Gráfica 4.18. Gráfica de residuales versus valores ajustados etapa 4 - B.

La gráfica de residuales versus los valores ajustados sugiere que la varianza de los errores es constante debido a que los puntos se distribuyen con cierta uniformidad.



Gráfica 4.19. Residuales versus el orden de los datos etapa 4 - B.

La gráfica de residuales versus el orden de los datos muestra la presencia de diversas observaciones que se alejan de la línea central y que pueden deberse a puntos atípicos.

A continuación se revisan los supuestos por medio de análisis más formales.

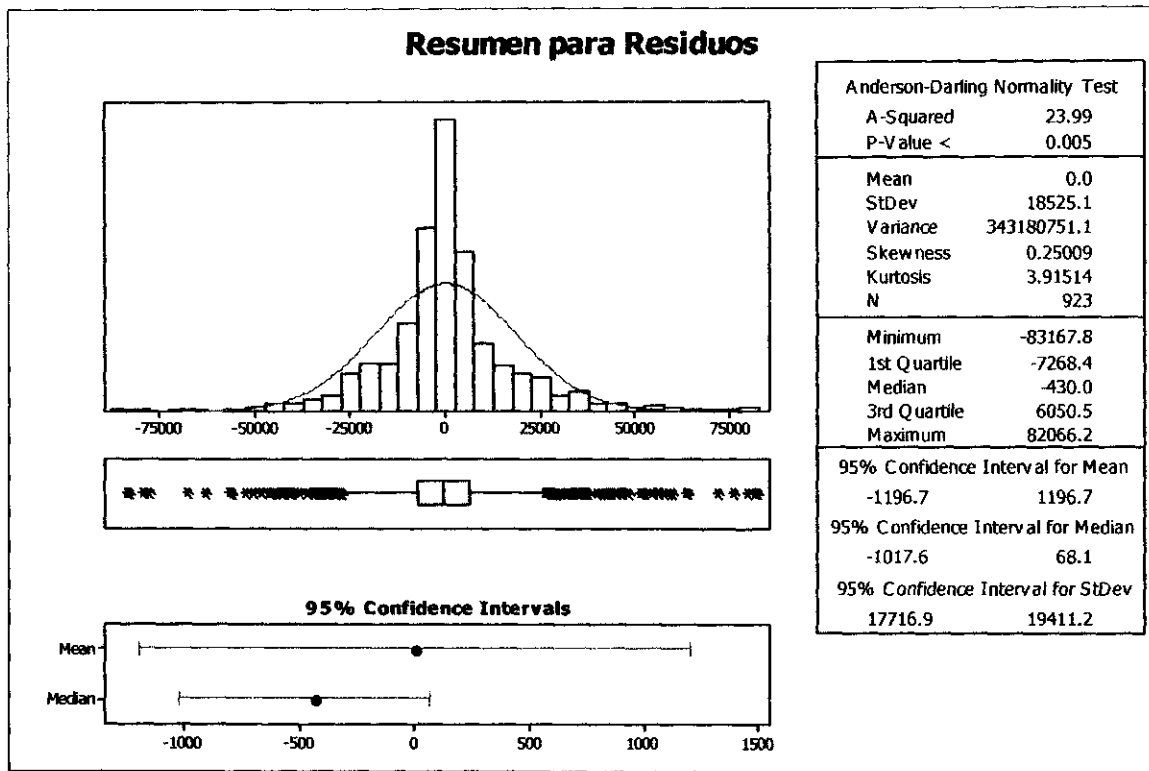
Normalidad

Para verificar la normalidad de los residuales, se presenta la gráfica 4.18., donde se incluye el estadístico de probabilidad normal Anderson-Darling. La prueba con un valor-p menor a 0.005 indica que existe evidencia estadística suficiente para rechazar el supuesto de normalidad.

Para comprobar dicho supuesto, se analizan los residuos estandarizados, los cuales se ajustan a una distribución normal cuando los valores se encuentran en el rango ± 2 . Al estudiar el anexo 19, se observa que los valores no se encuentran en dicho rango, debido a la presencia de las observaciones inusuales del conjunto de datos.

A pesar de que los residuos no son normales, lo más importante de este diagrama es representado por una alta kurtosis, que se refiere a la presencia de un pico afilado de la distribución. En este caso se considera este estadístico como una característica deseable en el modelo, dado a que la mayor parte de los residuos se distribuyen en la media igual a cero.

Se observa además en el histograma que la distribución los residuales alrededor del promedio es bastante simétrica e indica tener varianza constante, lo cual sugiere el cumplimiento de homocedasticidad, lo cual se considera de varianza "homogénea" dado a que los residuos se encuentran distribuidos con cierta uniformidad.



Gráfica 4.20. Gráfica para residuales etapa 4 - B.

Aleatoriedad e independencia de los errores

La aleatoriedad e independencia de los errores no puede verificarse por medio de la prueba de Durbin-Watson, debido a que los datos no siguen una distribución normal.

Por consiguiente, el supuesto de aleatoriedad e independencia de los errores por medio de la aplicación de la prueba no paramétrica de corridas. Los resultados arrojados por Minitab son:

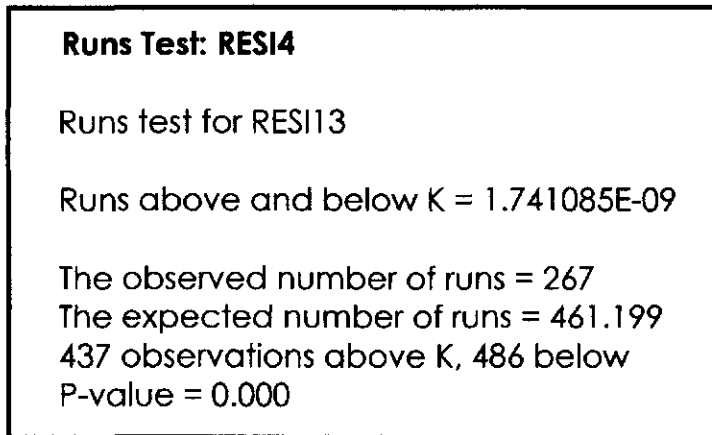


Figura 4.14. Resultados de Run Test etapa 4 - B.

Debido a que no se especifica la opción de otro valor que no sea la media como criterio de comparación (K), la media igual a 1.741085E-09 es utilizada. Se realizaron 267 corridas. Para determinar un número inusual de corridas, se calcula el número de observaciones que se encuentran por arriba o debajo de K. El resultado del valor-p (0.000) es menor que el nivel de alpha de 0.05, por tanto no hay evidencia estadística concluyente que indique que los datos no son aleatorios.

Homogeneidad

Se utiliza la gráfica de residuales versus los valores ajustados para determinar que existe varianza homogénea debido a la uniformidad de la distribución de los residuos.

4.2.4.4.2.3. Prueba de hipótesis en regresión lineal múltiple

Prueba de la significancia de la regresión

De la tabla del análisis de varianza, se determina si existe una relación lineal entre la respuesta y cualquiera de las variables regresoras x_1, x_2, \dots, x_k .

Las pruebas de hipótesis que se evalúan son:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_a: \text{por lo menos una } \beta_i \neq 0 \ (i = 1, \dots, k)$$

El estadístico de prueba se calcula según los datos arrojados por Minitab como:

$$F_0 = \frac{1.46389E + 13}{349627240} = 41870.04$$

Para probar la hipótesis nula, se utiliza el criterio de rechazo siguiente:

$$F_0 > F_{\alpha, k, n-k-1}$$

Dado un valor de tablas $F_{\alpha, k, n-k-1}$ igual a 2.01957 ($\alpha = 0.05$), se concluye que la hipótesis nula $H_0: \beta = 0$ se rechaza, indicando que al menos uno de los regresores x_k contribuye al modelo de forma significativa.

En una prueba equivalente, se puede rechazar H_0 cuando el valor-p de la prueba de F es menor que el nivel de alpha (α). En este caso, $0.00 < 0.05$, por tanto se rechaza la hipótesis nula.

Prueba sobre coeficientes individuales de regresión

Las pruebas de cualquier coeficiente de regresión son:

$$H_0: \beta_j = 0$$

$$H_a: \beta_j \neq 0$$

Se rechaza la hipótesis nula si $|t_0| > t_{\alpha/2, n-k-1}$.

Según tablas, $t_{\alpha/2, n-k-1} = 2.245098$

De acuerdo a los resultados arrojados por Minitab en la tabla 4.20., encontramos que el valor T para todos los coeficientes de regresión cae en la región de rechazo a un nivel de significancia de 0.05, se concluye que existe una contribución importante de las variables independientes dado los demás regresores del modelo.

Predictor	T
Constant	13.00
X ₂₁	-12.09
X ₂₃	-13.04
X ₂₅	-12.82
X ₄	7.08
X ₉	8.45
X ₁₀	-4.13
X ₁₇	18.79
X ₃ X ₂₄	3.76
X ₄ X ₂₀	24.12
X ₄ X ₂₁	26.04
X ₄ X ₂₂	14.37
X ₄ X ₂₄	-9.76
X ₈ X ₂₀	-9.95
X ₈ X ₂₁	3.51
X ₁₀ X ₂₀	5.80
X ₁₇ X ₂₂	-7.98
X ₄ ³	9.79

Tabla 4.20. Valores T - B.

4.2.4.4.2.4. Análisis de varianza (ANOVA)

El análisis de varianza de los datos de la sucursal B se presenta en la tabla 4.21.

Source	DF	SS	MS	F	P
Regression	17	2.48861E+14	1.46389E+13	41870.04	0.000
Residual Error	905	3.16413E+11	349627240		
Total	922	2.49178E+14			

Tabla 4.21. Tabla ANOVA - B.

El valor-p igual a 0.000 indica que el modelo estimado por el procedimiento de regresión es significativo a un nivel α de 0.05, indicando que al menos un coeficiente es diferente de cero.

Las hipótesis para el ANOVA son:

$$H_0: MS_R \leq MS_{RES}$$

$$H_a: MS_R > MS_{RES}$$

El estadístico de prueba de la ecuación es $F_0 = MS_R/MS_{RES}$, y el criterio de rechazo para H_0 es $F_0 > F_{\alpha, 1, n-2}$.

Según los datos arrojados por Minitab, el estadístico de prueba F_0 es igual a 41870.04 y el valor F de tablas a un nivel de significancia de 0.05 es igual a 3.851575.

Siendo $F_0 > F_{\alpha, 1, n-1}$, la hipótesis nula cae en la región de rechazo, lo que significa que la variación estimada por el modelo es mayor que la variación aleatoria, y por consiguiente se puede decir que el modelo es bueno.

La tabla 4.20. de los grados de libertad y la suma al cuadrado secuencial (Seq SS) de los estimados, indican que las variables cantidad de socios parciales (x_5) y cantidad de bajas (x_7), no explican una cantidad substancial de varianza única, lo cual sugiere la posibilidad de eliminar dichas variables del modelo.

Source	DF	Seq SS
X ₂₁	1	8.96481E+12
X ₂₃	1	5.40591E+13
X ₂₅	1	8.46049E+13
X ₄	1	7.13685E+13
X ₉	1	6.67059E+12
X ₁₀	1	9.28024E+12
X ₁₇	1	7.22462E+12
X ₃ X ₂₄	1	20457003823
X ₄ X ₂₀	1	4.35217E+12
X ₄ X ₂₁	1	3.73165E+11
X ₄ X ₂₂	1	1.84308E+12
X ₄ X ₂₄	1	15500745337
X ₈ X ₂₀	1	22509126444
X ₈ X ₂₁	1	3317445566
X ₁₀ X ₂₀	1	10224593113
X ₁₇ X ₂₂	1	14761616011
X ₄ ³	1	33478105940

Tabla 4.22. Grados de libertad y suma al cuadrado secuencial - B.

4.2.4.4.2.5. Intervalos de confianza

A continuación se muestran los intervalos de confianza del modelo B a un nivel de significancia del 95%.

Intervalos de confianza de los coeficientes de regresión

En la tabla 4.23. se presentan los estimados de intervalo de confianza de los coeficientes de regresión β_j .

Coeficientes	95% Intervalo de Confianza para β	
	Límite Inferior	Límite Superior
Constante	208545.7	282731.8
X ₂₁	-315609	-227481
X ₂₃	-281027	-207510
X ₂₅	-320375	-235322
X ₄	4.952651	8.748804
X ₉	1.08049	1.734042
X ₁₀	-0.13785	-0.04897
X ₁₇	0.152627	0.188224
X ₃ X ₂₄	5570.581	17752.12
X ₄ X ₂₀	25.80089	30.3722
X ₄ X ₂₁	37.01536	43.05081
X ₄ X ₂₂	15.22919	20.04593
X ₄ X ₂₄	-12.0423	-8.0102
X ₈ X ₂₀	-248.615	-166.671
X ₈ X ₂₁	62.59723	221.1754
X ₁₀ X ₂₀	0.13014	0.263389
X ₁₇ X ₂₂	-0.19552	-0.11835
X ₄ ³	1.82E-10	2.74E-10

Tabla 4.23. Intervalo de confianza - B.

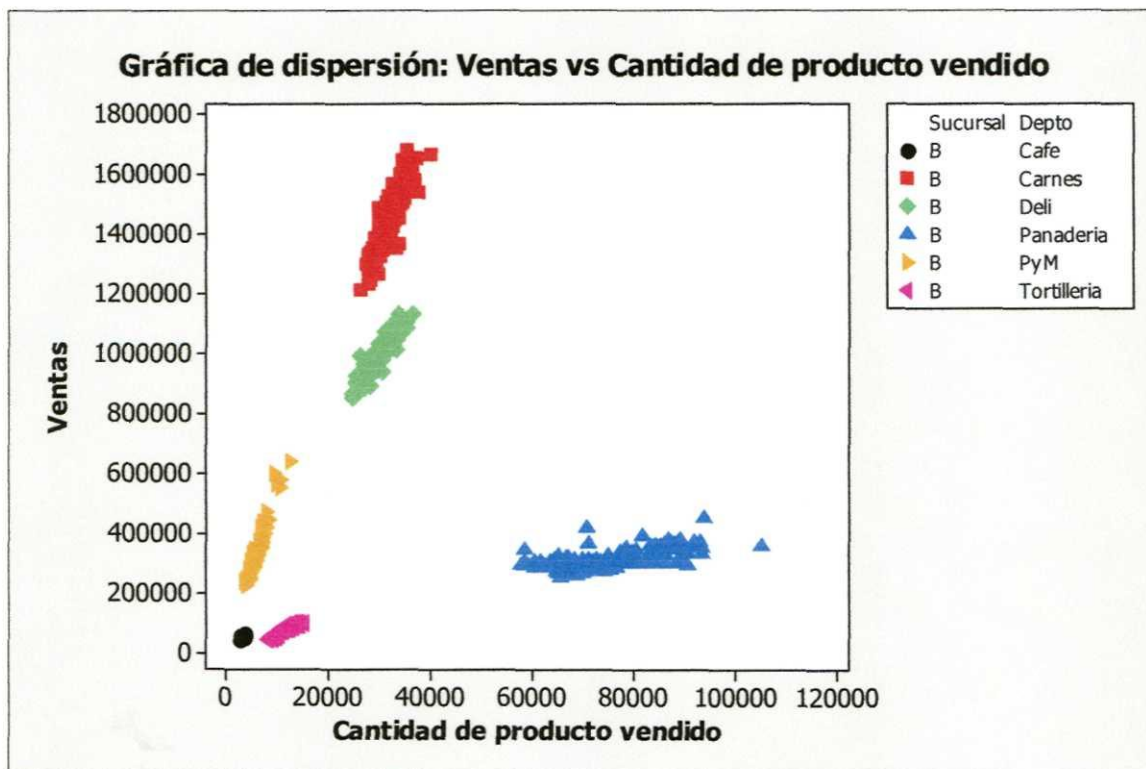
La construcción de estimaciones de intervalos de confianza es útil para establecer si el modelo es preciso. En este caso los intervalos de confianza de la mayor parte de las variables son moderadamente amplios. Los regresores que tienen un intervalo de confianza amplio son: Pescados y Mariscos, Café y Tortillería, así como algunas interacciones tales como quincenas y Panadería (X₃X₂₄), además de horas disponibles de trabajo efectivo con Carnes y Pescados respectivamente. Los intervalos grandes sugieren que la aportación de esas variables al modelo es baja.

Estimación del intervalo de confianza de la respuesta media

Los intervalos de confianza de la respuesta media, así como los intervalos de predicción del modelo a un nivel del 95%, se presentan en el anexo 19.

4.2.4.4.2.6. Gráficas de dispersión

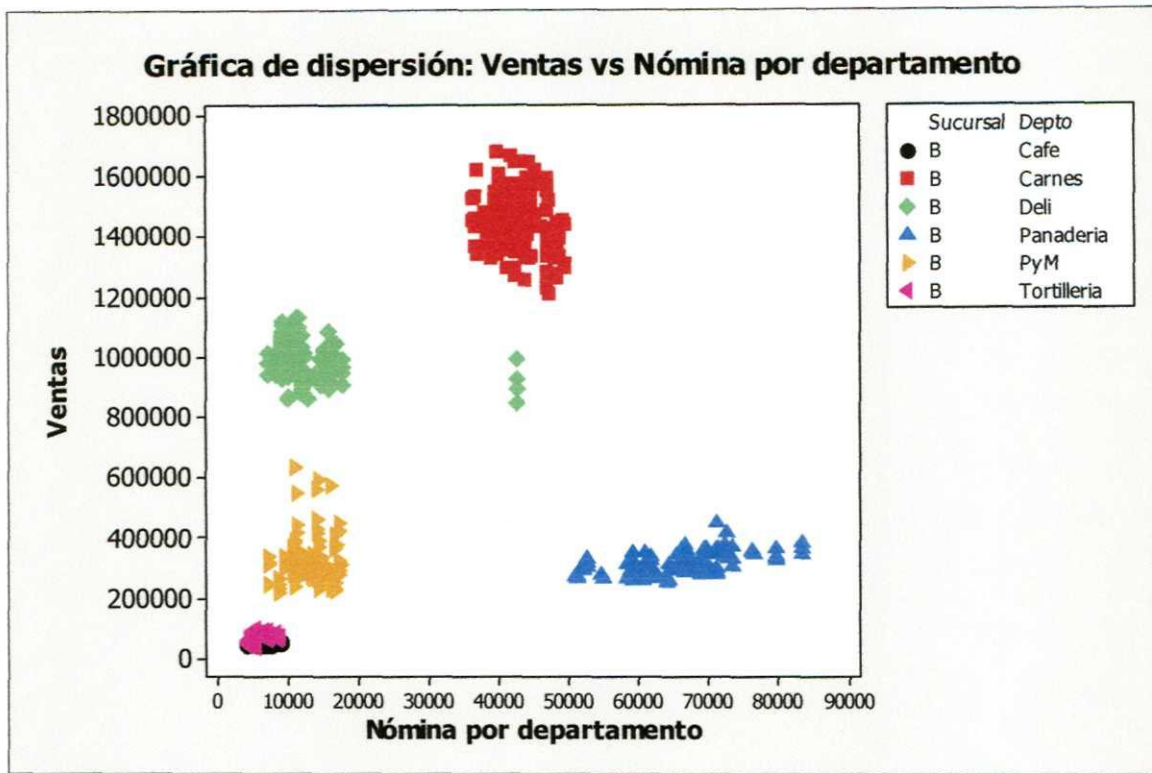
Para representar el comportamiento de los datos del modelo final de B, se muestran las gráficas de dispersión 4.21. a 4.24., en las cuales se observa un comportamiento lineal de las ventas con respecto a la variable independiente graficada. A pesar de la linealidad de los datos, se identifican puntos atípicos que pueden ser los causantes de los problemas de correlación e incumplimiento de los supuestos de regresión.



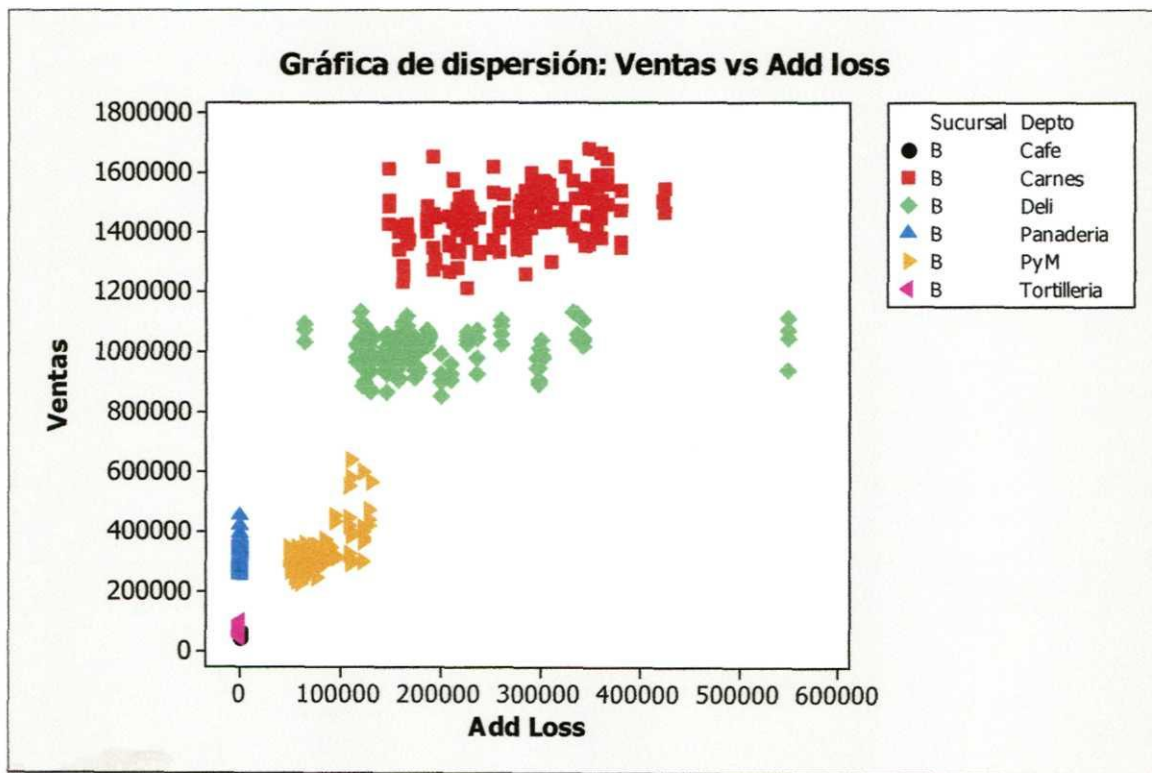
Gráfica 4.21. Ventas versus cantidad de producto vendido etapa 4 - B.

En la gráfica 4.21., es posible identificar que los departamentos de Carnes, Delicatessen y Pescados y Mariscos son las que mayores ventas generan para la sucursal de B, a pesar de que la cantidad de producto vendido sea pequeña. Por su lado, el área de Panadería representa el departamento con mayor cantidad de producto vendido, sin embargo la naturaleza del producto sugiere que las ventas generadas no son altas. Finalmente las pendientes para Café y Tortillería indican que la cantidad de producto vendido es pequeña y por tanto también lo son las ventas.

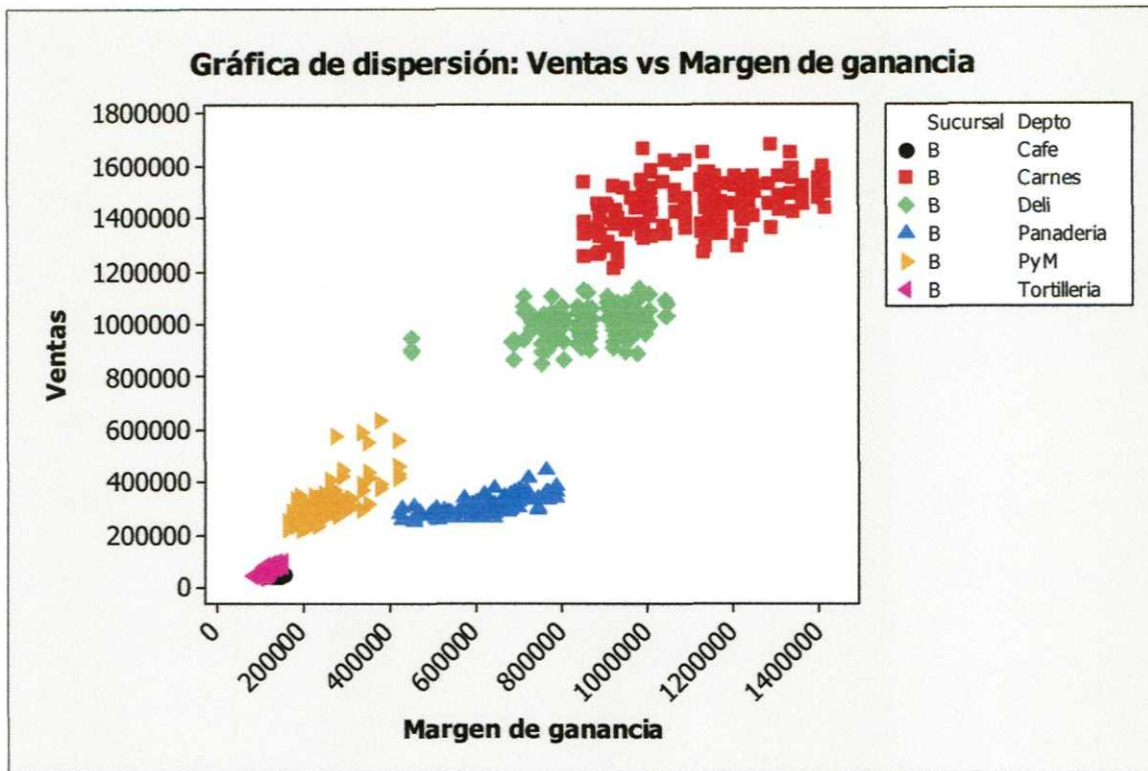
A continuación se presentan las gráficas de dispersión donde se comparan las ventas con respecto a otros regresores importantes.



Gráfica 4.22. Ventas versus nómina por departamento etapa 4 - B.



Gráfica 4.23. Ventas versus add loss etapa 4 - B.



Gráfica 4.24. Ventas versus margen de ganancia etapa 4 - B.

La gráfica 4.22., permite estudiar las ventas con respecto a la variable nómina por departamento. Se concluye que el departamento de Carnes genera mayor cantidad de ventas en la tienda de B. Posteriormente se identifica que Delicatessen y Pescados y Mariscos ocupan el segundo y tercer lugar en la cantidad de ventas generadas con una menor cantidad de nómina asignada en comparación con Carnes. En Panadería se asigna un mayor presupuesto para salarios; sin embargo, no se observa una tendencia a la alza de las ventas a medida que la nómina aumenta. Finalmente los departamentos de Café y Tortillería indican poco crecimiento en cuanto a nómina designada y ventas generadas.

De la gráfica 4.23., se concluye que entre mayores son las rebajas en los precios de los productos de los departamentos de Carnes y Delicatessen, mayor es el impacto en el aumento de las ventas. En el caso de los departamentos de Café, Panadería y Tortillería no se reportan rebajas.

Según la gráfica 4.24., los departamentos de Carnes y Delicatessen son los que aportan mayores ventas a medida que el margen de ganancia aumenta. En Panadería se identifica una tendencia a tener un mayor

margen de ganancia mientras se mantiene un nivel de ventas menor a los \$500,000 pesos. En Pescados y Mariscos, el margen de ganancia con respecto a las ventas son moderadas en comparación con Carnes y Delicatessen. Finalmente, los departamentos de Café y Tortillería generan solamente un pequeño margen de ganancia, y por tanto también se reportan pocas ventas.

4.2.4.4.3. Modelo final D

El modelo para la sucursal D se determina a través de diversas pruebas de regresión que incluyen el análisis de las variables que resultaron ser significativas a un nivel de confianza del 95% en el procedimiento stepwise del anexo 20. Para definir las variables de entrada se utiliza el paso 23.

La ecuación de regresión para el modelo final es:

$$Y = 48827 + 208487 X_{22} + 185932 X_{24} + 0.0733 X_{17} + 20339 X_{3X20} + 32.2 X_{4X20} + 48.3 X_{4X21} + 22.1 X_{4X22} + 5128 X_{5X20} - 10238 X_{5X21} - 0.0580 X_{17X22} + 308 X_{18X20} - 0.000055 X_{18^3}$$

Se observa que las variables indicadoras de sucursal (x_{11} hasta x_{15}) son eliminadas debido a que las columnas de datos se encuentran conformadas por valores iguales a cero.

El modelo muestra que existe una relación inversamente proporcional de las ventas (Y) con las interacciones: departamento Pescados y Mariscos con cantidad de socios parciales (x_{5X21}), margen de ganancia y Delicatessen (x_{17X22}), y ventas por hora hombre al cubo (x_{18^3}).

El resto de los coeficientes de la ecuación tienen signo positivo, lo cual indica que las ventas tienen una relación directamente proporcional con los departamentos de Delicatessen (x_{22}) y Panadería (x_{24}), además de la variable margen de ganancia (x_{17}), y las interacciones de: quincenas y Carnes (x_{3X20}), cantidad de producto vendido con los departamentos Carnes (x_{4X20}), Pescados (x_{4X21}) y Delicatessen (x_{4X22}), cantidad de socios parciales y Carnes (x_{5X20}), y venta por hora hombre con Carnes (x_{18X20}).

La interpretación de los coeficientes de regresión indica que el intercepto de 48827 es el valor promedio de las ventas (Y) cuando las x_i son cero. Mientras que los estimados $\hat{\beta}_j$ muestran el cambio promedio de las ventas cuando x_i se incrementa en una unidad.

4.2.4.4.3.1. Estimación de los parámetros del modelo

Los coeficientes del modelo, así como los valores correspondientes al error estándar de los parámetros, se presentan en la tabla 4.24. Ninguna variable tiene un valor-p que no sea significativo a un nivel de significancia α de 0.05.

Los estimados del modelo de regresión de la figura 4.15., indican que la variación esperada de las ventas con respecto a la media de la distribución de los datos es de 16806.2 pesos. El valor de los errores comparado con las observaciones de las ventas es relativamente pequeño, por lo que se sugiere que no afecta significativamente al estudio.

El coeficiente de determinación R^2 y $R^2_{ajustado}$ muestran un ajuste igual o superior a 98.8%, un valor alto que permite deducir que las ventas pueden ser explicadas por el modelo de regresión.

Por otro lado, el valor R^2 de predicción de 99.76% de ajuste, indicando que el modelo es bueno para predecir futuras observaciones.

Predictor	Coef	SE Coef	T	P
Constant	48827	1339	36.48	0.000
X ₂₂	208487	20313	10.26	0.000
X ₂₄	185932	4176	44.53	0.000
X ₁₇	0.073341	0.008547	8.58	0.000
X ₃ X ₂₀	20339	2878	7.07	0.000
X ₄ X ₂₀	32.1761	0.7695	41.81	0.000
X ₄ X ₂₁	48.329	1.513	31.94	0.000
X ₄ X ₂₂	22.0751	0.8938	24.70	0.000
X ₅ X ₂₀	5127.6	602.2	8.51	0.000
X ₅ X ₂₁	-10238	2158	-4.74	0.000
X ₁₇ X ₂₂	-0.05800	0.01677	-3.46	0.001
X ₁₈ X ₂₀	307.51	56.19	5.47	0.000
X ₁₈ ³	-0.00005462	0.00000906	-6.03	0.000

Tabla 4.24. Coeficientes del modelo etapa 4 - D.

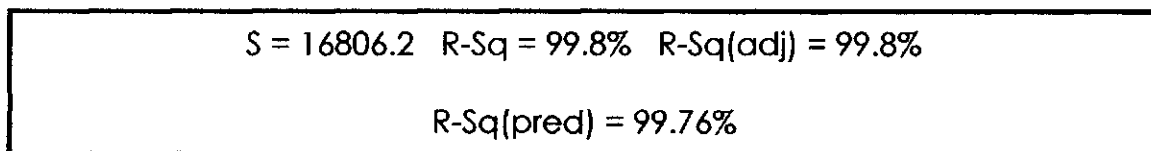


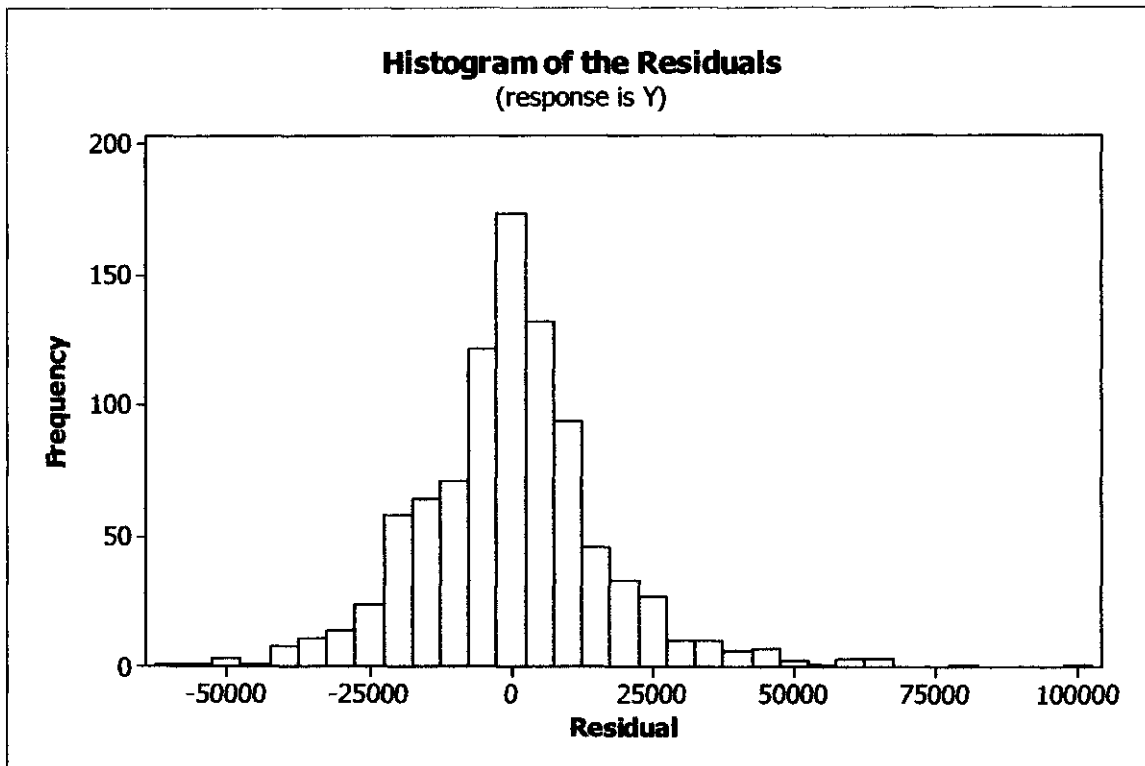
Figura 4.15. Estimados del modelo de regresión etapa 4 - D.

4.2.4.4.3.2. Supuestos del análisis de regresión

El análisis de los supuestos de regresión para el modelo de D se presenta a continuación.

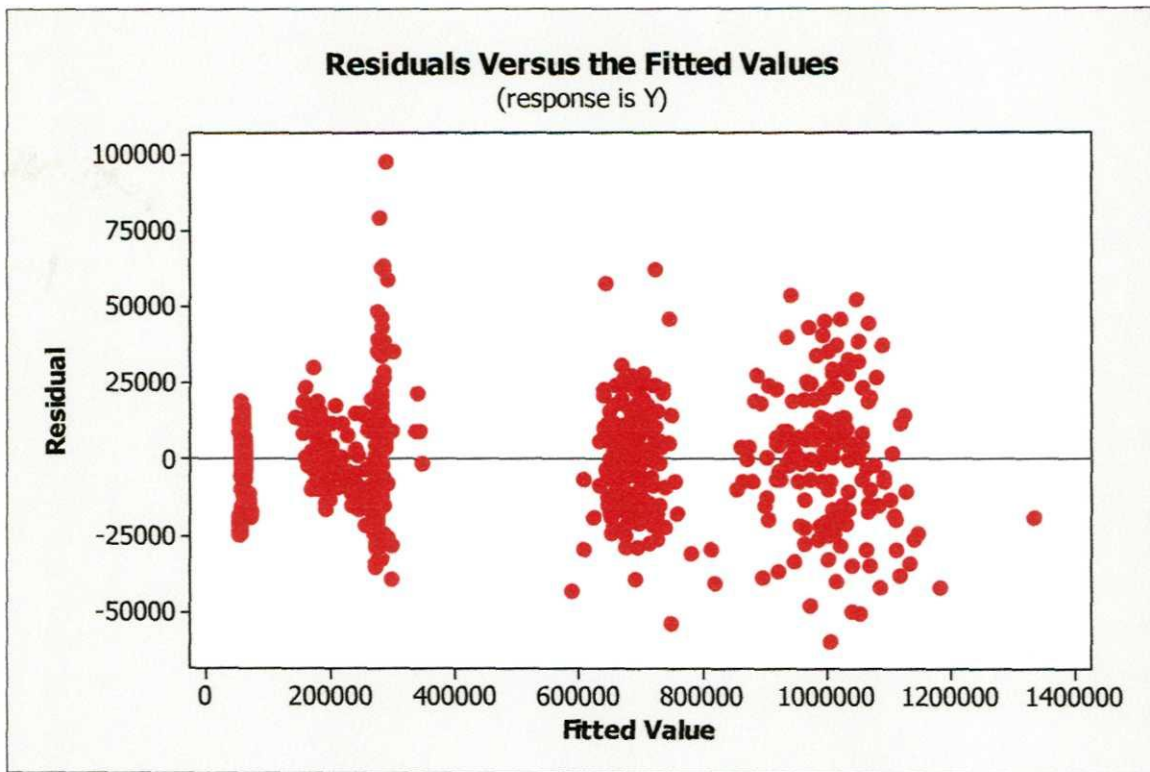
Análisis de residuales

Para análisis posteriores se utilizan las siguientes gráficas de residuos.



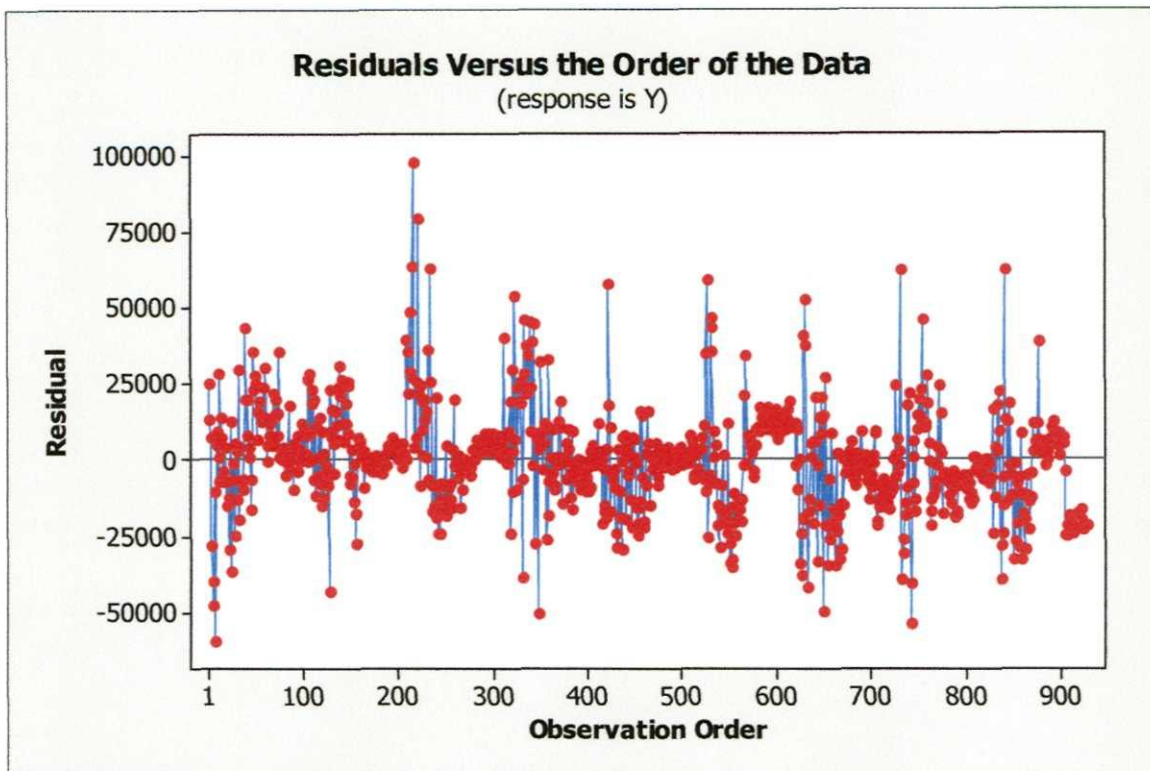
Gráfica 4.25. Histoarama de residuales para etapa 4 - D.

El histograma de residuales muestra un patrón asociado de asimetría, esto se debe a la ocurrencia de residuales grandes o por la presencia de observaciones atípicas.



Gráfica 4.26. Residuales versus los valores ajustados etapa 4 - D.

La gráfica de residuales versus los valores ajustados sugiere que la varianza de los errores es constante debido a que no se observa una tendencia en el comportamiento de los datos.



Gráfica 4.27. Residuales versus el orden de los datos etapa 4 - D.

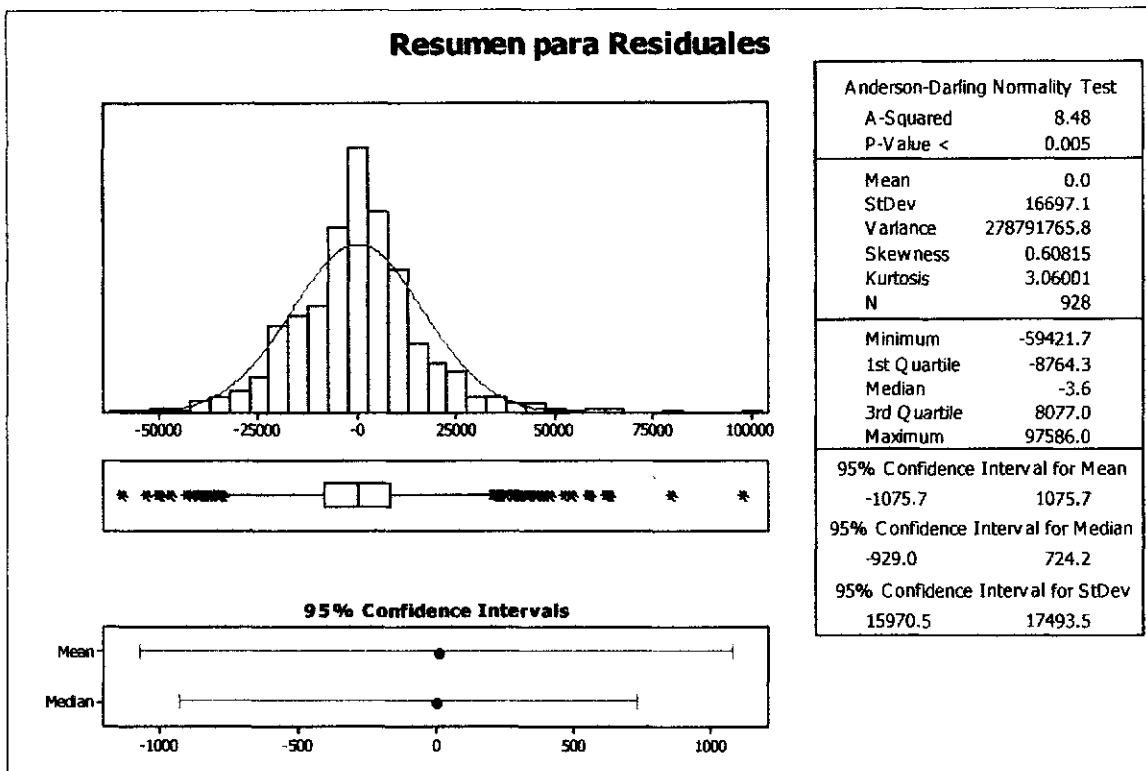
Finalmente, la gráfica de residuales versus el orden de los datos muestra la presencia de diversas observaciones que se encuentran lejos al resto de los puntos.

A continuación se revisan los supuestos por medio de análisis más formales.

Normalidad

Para verificar la normalidad de los residuales, se presenta con el estadístico de probabilidad normal Anderson-Darling. Con un valor-p menor a 0.005, existe evidencia estadística suficiente para rechazar el supuesto de normalidad.

Para comprobar dicho supuesto, se analizan los residuos estandarizados, los cuales se ajustan a una distribución normal cuando los valores se encuentran en el rango ± 2 . Al estudiar el anexo 21, se concluye que las observaciones de los valores que no se encuentran en dicho rango, debido a la presencia de los datos identificados como puntos atípicos.



Gráfica 4.28. Gráfica de normalidad de residuales etapa 4 - D.

A pesar de que los residuos no son normales, lo más importante de este diagrama es representado por la kurtosis mayor a un valor de 3 que indica una distribución relativamente apuntada, siendo esto una característica deseable en el modelo.

Aleatoriedad e independencia de los errores

La aleatoriedad e independencia de los errores se verifica por medio de la prueba no paramétrica de corridas cuyos resultados se observan en la figura 4.16.

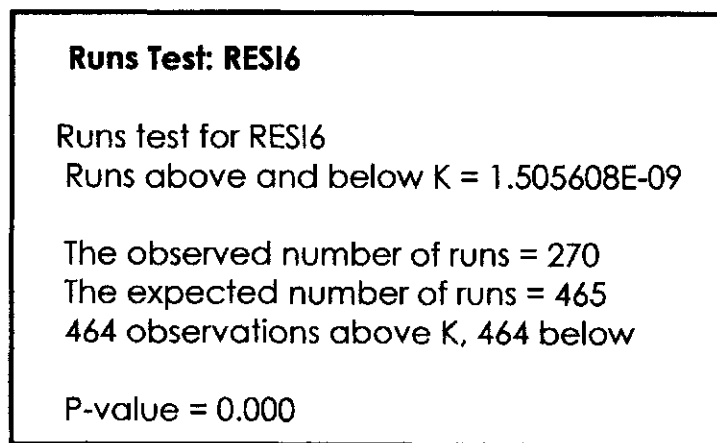


Figura 4.16. Resultados de Run Test etapa 4 - D.

Debido a que no se especifica la opción de otro valor que no sea la media como criterio de comparación (K), la media igual a 1.505608E-09 es utilizada. Se realizaron 270 corridas. Para determinar un número inusual de corridas, se calcula el número de observaciones que se encuentran por arriba o debajo de K. El resultado del valor-p (0.000) es menor que el nivel de alpha de 0.05, existe evidencia estadística suficiente para rechazar la falta de aleatoriedad de los datos.

Homogeneidad

Los resultados de la gráfica de residuales podrían sugerir que los datos tienen varianza constante, debido a que no se observa una tendencia particular en la distribución de los puntos. Por tanto se considera el supuesto de que existe varianza homogénea en los residuos.

4.2.4.4.3.3. Prueba de hipótesis en regresión lineal múltiple

Prueba de la significancia de la regresión

Esta prueba permite conocer si existe una relación lineal entre la respuesta y cualquiera de las variables regresoras x_1, x_2, \dots, x_k .

Las pruebas de hipótesis que se evalúan son:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_a: \text{por lo menos una } \beta_i \neq 0 \ (i = 1, \dots, k)$$

El estadístico de prueba (2.12.9) se calcula según los datos arrojados por Minitab como:

$$F_0 = \frac{9.54550E + 12}{282448051} = 33795.61$$

Para probar la hipótesis nula, se utiliza el criterio de rechazo siguiente:

$$F_0 > F_{\alpha, k, n-k-1}$$

Dado un valor de tablas $F_{\alpha, k, n-k-1}$ igual a 2.614537 con un nivel de significancia α igual a 0.05, se concluye que la hipótesis nula $H_0: \beta = 0$ se rechaza, indicando que al menos uno de los regresores x_k contribuye al modelo de forma significativa.

En una prueba equivalente, se puede rechazar H_0 cuando el valor-p de la prueba de F es menor que el nivel de alpha (α). En este caso, $0.00 < 0.05$, por tanto se rechaza la hipótesis nula.

Prueba sobre coeficientes individuales de regresión

Las pruebas de cualquier coeficiente de regresión son:

$H_0: \beta_j = 0$

$H_a: \beta_j \neq 0$

Se rechaza la hipótesis nula si $|t_0| > t_{\alpha/2, n-k-1}$.

Según tablas, $t_{\alpha/2, n-k-1} = 2.245058$

De acuerdo a los resultados arrojados por Minitab en la tabla 4.25., encontramos que el valor T para todos los coeficientes de regresión cae en la región de rechazo a un nivel de significancia de 0.05, se concluye que existe una relación causal entre las variables independientes del estudio.

Predictor	T
Constant	36.48
X ₂₂	10.26
X ₂₄	44.53
X ₁₇	8.58
X ₃ X ₂₀	7.07
X ₄ X ₂₀	41.81
X ₄ X ₂₁	31.94
X ₄ X ₂₂	24.70
X ₅ X ₂₀	8.51
X ₅ X ₂₁	-4.74
X ₁₇ X ₂₂	-3.46
X ₁₈ X ₂₀	5.47
X ₁₈ ³	-6.03

Tabla 4.25. Valores T - D.

4.2.4.4.3.4. Análisis de varianza (ANOVA)

La tabla 4.26., muestra los resultados del análisis ANOVA de la sucursal D.

Source	DF	SS	MS	F	P
Regression	12	1.14546E+14	9.54550E+12	33795.61	0.000
Residual Error	915	2.58440E+11	282448051		
Total	927	1.14804E+14			

Tabla 4.26. Tabla ANOVA - D.

El valor-p igual a 0.000 indica que el modelo estimado por el procedimiento de regresión es significativo a un nivel α de 0.05, indicando que al menos un coeficiente es diferente de cero.

Las hipótesis para el ANOVA son:

$$H_0: MS_R \leq MS_{RES}$$

$$H_a: MS_R > MS_{RES}$$

El estadístico de prueba de la ecuación es $F_0 = MS_R/MS_{RES}$, y el criterio de rechazo para H_0 es $F_0 > F_{\alpha, 1, n-2}$.

Según los datos arrojados por Minitab, el estadístico de prueba F_0 es igual a 33795.61 y el valor F de tablas a un nivel de significancia de 0.05 es igual a 3.85152.

Siendo $F_0 > F_{\alpha, 1, n-1}$, la hipótesis nula cae en la región de rechazo, lo que significa que la variación estimada por el modelo es mayor que la variación aleatoria, y por consiguiente se puede decir que el modelo es bueno.

La tabla 4.27, de los grados de libertad y la suma al cuadrado secuencial (Seq SS) de los estimados, indican que las interacciones formadas por las variables: cantidad de socios parciales con Pescados y Mariscos (x_5x_{21}), margen de ganancia con Delicatessen ($x_{17}x_{22}$), y venta por hora hombre con Carnes ($x_{18}x_{20}$) tienen los menores valores, por lo que se puede decir que no explican una cantidad substancial de varianza única, lo que sugiere la posibilidad de eliminar dichas variables del modelo.

Source	DF	Seq SS
X ₂₂	1	1.72090E+13
X ₂₄	1	3.47435E+11
X ₁₇	1	8.62086E+13
X ₃ X ₂₀	1	1.24650E+12
X ₄ X ₂₀	1	7.05677E+12
X ₄ X ₂₁	1	2.22435E+12
X ₄ X ₂₂	1	1.91466E+11
X ₅ X ₂₀	1	31151018939
X ₅ X ₂₁	1	5990708904
X ₁₇ X ₂₂	1	6660920847
X ₁₈ X ₂₀	1	7922985166
X ₁₈ ³	1	10267118545

Tabla 4.27. Grados de libertad y suma al cuadrado secuencial - D

4.2.4.4.3.5. Intervalos de confianza

A continuación se muestran los intervalos de confianza del modelo D a un nivel de significancia del 95%.

Intervalos de confianza de los coeficientes de regresión

En la tabla 4.28. se presentan los estimados de intervalo de confianza de los coeficientes de regresión β_j .

Coeficientes	95% Intervalo de Confianza para β	
	Límite Inferior	Límite Superior
Constante	46199.49	51453.64
X ₂₂	168622	248352
X ₂₄	177736.6	194126.5
X ₁₇	0.056567	0.090114
X ₃ X ₂₀	14689.28	25987.74
X ₄ X ₂₀	30.66588	33.68635
X ₄ X ₂₁	45.3589	51.29858
X ₄ X ₂₂	20.3209	23.82932
X ₅ X ₂₀	3945.673	6309.534
X ₅ X ₂₁	-14473.4	-6002.84
X ₁₇ X ₂₂	-0.09091	-0.02509
X ₁₈ X ₂₀	197.2421	417.7823
X ₁₈ ³	-7.2E-05	-3.7E-05

Tabla 4.28. Intervalo de confianza - D.

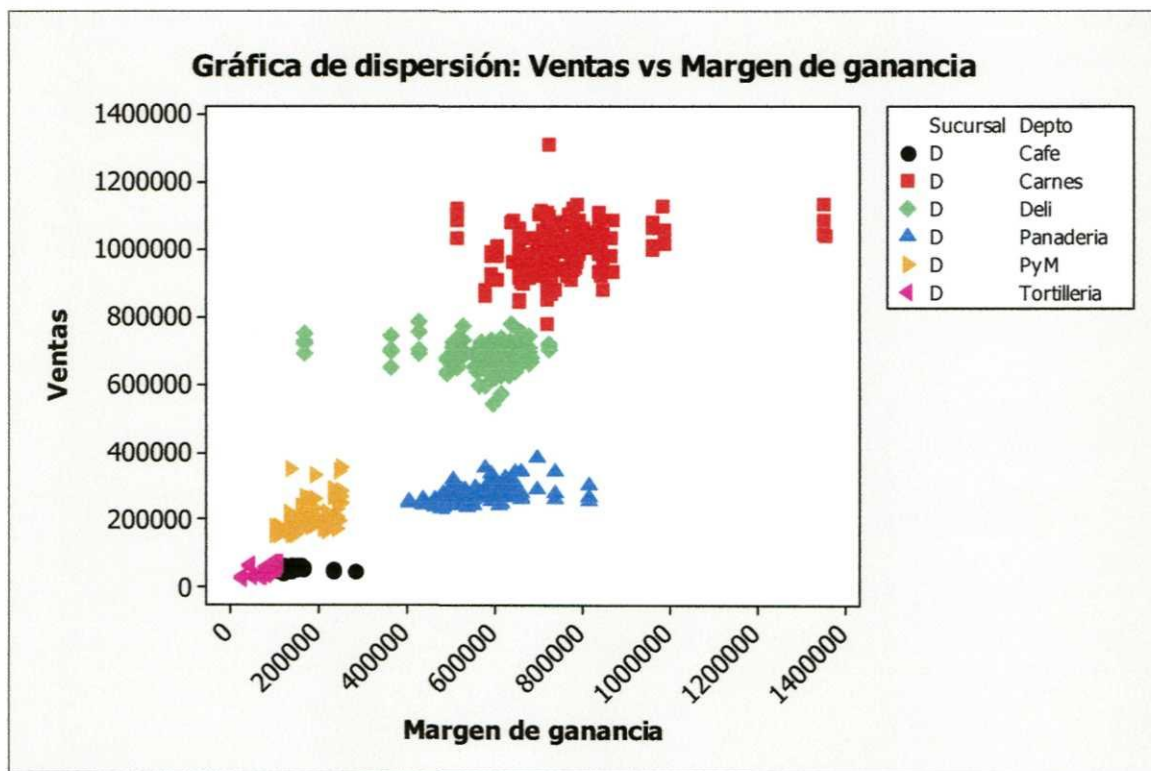
La construcción de estimaciones de intervalos de confianza es útil para establecer si el modelo es preciso. En este caso los intervalos de confianza de la mayor parte de las variables que son moderadamente amplios se forman por las interacciones: X5X20, X5X21, X17X22 y X18X20.

Estimación del intervalo de confianza de la respuesta media

Los intervalos de confianza de la respuesta media, así como los intervalos de predicción del modelo a un nivel del 95%, se presentan en el anexo 22.

4.2.4.4.3.6. Gráficas de dispersión

Para representar el comportamiento de los datos del modelo final de D, se muestra la gráfica de dispersión 4.29., en la cual se observa un comportamiento lineal de las ventas con respecto a la variable independiente graficada. Al igual que en los modelos anteriores se identifican puntos atípicos.



Gráfica 4.29. Ventas versus margen de ganancia etapa 4 - D.

Entre mayor es el margen de ganancia en los departamentos de Carnes y Delicatessen, las ventas incrementan. La pendiente del departamento de Panadería muestra que aunque el margen de ganancia incrementa,

las ventas de la sucursal no crecen a más de \$400,000 pesos. En Pescados y Mariscos las ventas oscilan en un intervalo de 2 a 4 mil pesos mientras que el margen de ganancia aumenta muy poco. Por otro lado, los departamentos de Café y Tortillería reportan poco margen de ganancia y ventas pequeñas.

4.2.4.4.4. Modelo final Combo

El modelo para las sucursales del Combo, al igual que los anteriores, se determina a través de diversas pruebas de regresión que incluyen el análisis de las variables que resultaron ser significativas a un nivel de confianza del 95% en el procedimiento *stepwise* mostrado en el anexo 23 y considerando el paso 22 para la selección de las variables de entrada.

La ecuación de regresión del modelo Combo compuesto por las tiendas C, E y F es:

$$Y = 7405 + 71861 X_{22} + 4.07 X_4 + 60.8 X_{18} + 33284 X_{3X20} + 12615 X_{3X22} + 18.1 X_{4X20} + 33.5 X_{4X21} + 11.4 X_{4X22} - 6.51 X_{4X24} - 872 X_{5X13} + 15408 X_{5X22} + 225 X_{8X20} + 142 X_{8X24} + 0.0512 X_{10X13} + 0.0606 X_{17X13} + 0.132 X_{17X22} + 16.8 X_{18X13} + 371 X_{18X20} - 59.7 X_{18X22} + 1499 X_{18X24} + 0.000041 X_4^2 - 0.000000000189 X_4^3$$

Existe una relación inversamente proporcional entre las ventas (Y) y las interacciones conformadas por: cantidad de producto vendido con Panadería (x_{4x24}), cantidad de socios parciales con sucursal C (x_{5x13}), venta por hora hombre y departamento de Delicatessen (x_{18x22}), y cantidad de producto vendido elevado a la tercera potencia (x_4^3).

Debido al signo positivo de los parámetros se puede decir que las ventas tienen una relación directamente proporcional con: Delicatessen (x_{22}), cantidad de producto vendido (x_4), ventas por hora hombre (x_{18}), y con las interacciones conformadas por: quincenas con Carnes (x_{3x20}) y Delicatessen (x_{3x22}); cantidad de producto vendido con Carnes (x_{4x20}), Pescados y Mariscos (x_{4x21}), Delicatessen (x_{4x22}), y Panadería (x_{4x24}); cantidad de socios parciales con Delicatessen (x_{5x22}), horas disponibles de trabajo efectivo con Carnes (x_{8x20}) y Panadería (x_{8x24}); *add loss* con C (x_{10x13}); margen de ganancia con Delicatessen (x_{17x22}) y C (x_{17x13}); venta por hora hombre con Carnes (x_{18x20}), Panadería (x_{18x24}) y C (x_{18x13}); y finalmente cantidad de producto vendido al cuadrado (x_4^2).

La interpretación de los coeficientes de regresión indica que el intercepto de 7405 es el valor promedio de las ventas (Y) cuando las x_i son cero. Mientras que los estimados $\hat{\beta}_j$ muestran el cambio promedio de las ventas cuando x_i se incrementa en una unidad.

4.2.4.4.1. Estimación de los parámetros del modelo

Los coeficientes del modelo, así como los valores correspondientes al error estándar de los parámetros, se presentan en la tabla 4.29. Se concluye que ninguno de los valores-p de los estimados tiene problemas de significancia a un nivel α de 0.05.

Predictor	Coef	SE Coef	T	P
Constant	7405	1097	6.75	0.000
X ₂₂	71861	9434	7.62	0.000
X ₄	4.0706	0.2093	19.45	0.000
X ₁₈	60.78	10.05	6.05	0.000
X ₃ X ₂₀	33284	1750	19.02	0.000
X ₃ X ₂₂	12615	1757	7.18	0.000
X ₄ X ₂₀	18.1313	0.6117	29.64	0.000
X ₄ X ₂₁	33.5478	0.6929	48.41	0.000
X ₄ X ₂₂	11.4096	0.6317	18.06	0.000
X ₄ X ₂₄	-6.5132	0.5731	-11.36	0.000
X ₅ X ₁₃	-872.1	230.6	-3.78	0.000
X ₅ X ₂₂	15408	2358	6.54	0.000
X ₈ X ₂₀	225.23	14.45	15.59	0.000
X ₈ X ₂₄	142.03	10.24	13.87	0.000
X ₁₀ X ₁₃	0.05124	0.01515	3.38	0.001
X ₁₇ X ₁₃	0.060632	0.007076	8.57	0.000
X ₁₇ X ₂₂	0.13187	0.01676	7.87	0.000
X ₁₈ X ₁₃	16.775	3.135	5.35	0.000
X ₁₈ X ₂₀	371.39	20.37	18.23	0.000
X ₁₈ X ₂₂	-59.67	10.38	-5.75	0.000
X ₁₈ X ₂₄	1499.0	100.7	14.88	0.000
X ₄ ²	4.124E-05	1.356E-05	3.04	0.002
X ₄ ³	-1.89E-10	8.68E-11	-2.19	0.029

Tabla 4.29. Coeficientes del modelo etapa 4 - Combo.

Los estimados del modelo de regresión de la figura 4.17., indican que la variación esperada de las ventas con respecto a la media de la distribución de los datos es de 17603.1 pesos. El valor de los errores

comparado con las observaciones de las ventas es relativamente pequeño, por lo que se sugiere que no afecta significativamente al estudio.

El coeficiente de determinación R^2 y $R^2_{ajustado}$ expresan que el 99.5% de la variación total de las ventas puede ser explicada por el modelo cuantitativo, por lo que es posible deducir que los cambios que sufre la variable ventas (y) se pueden calcular a partir de las variables independientes.

El valor R^2 de predicción se calcula a partir del estadístico PRESS e indica que tan bueno es el modelo para pronosticar ventas para nuevas observaciones. El $R^2_{predicción}$ de 99.54% sugiere que el modelo es bueno para predecir nuevos datos.

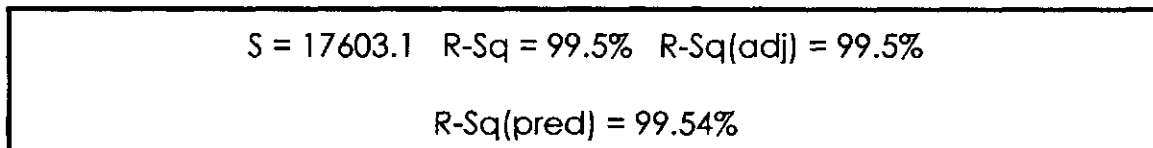


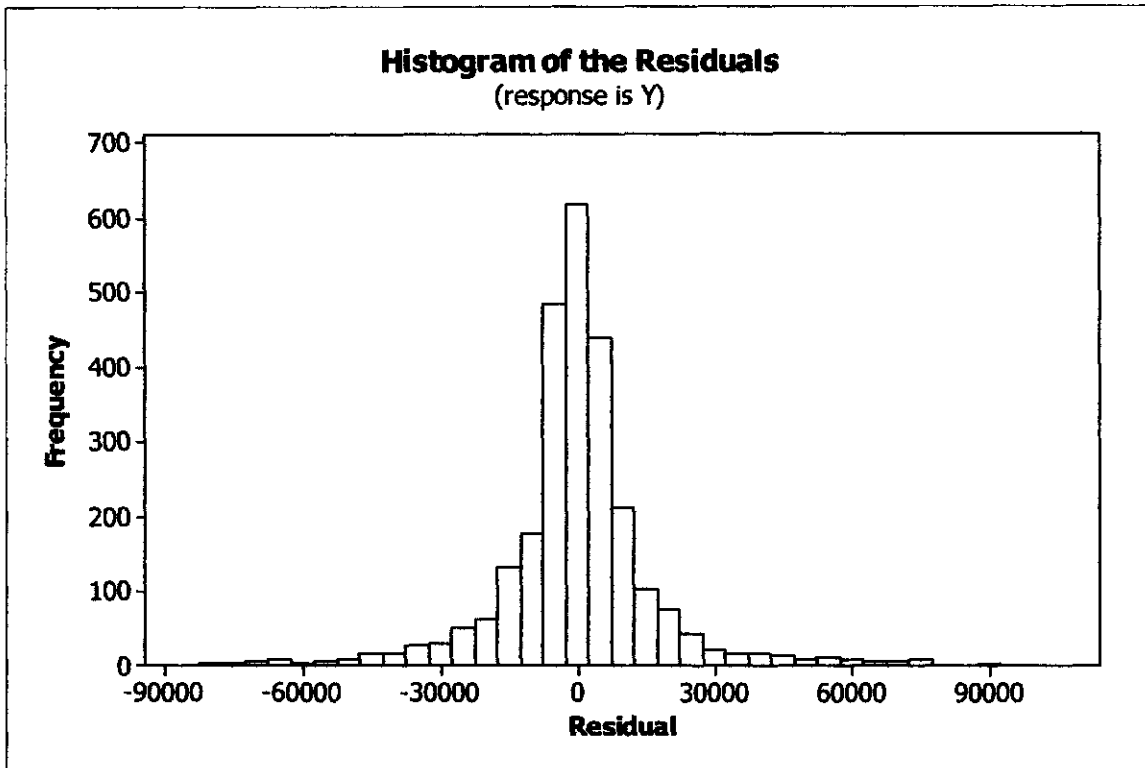
Figura 4.17. Estimados modelo de regresión etapa 4 - Combo.

4.2.4.4.4.2. Supuestos del análisis de regresión

El estudio de los supuestos de regresión se presenta con el objeto de identificar la naturaleza del comportamiento de los residuales del modelo.

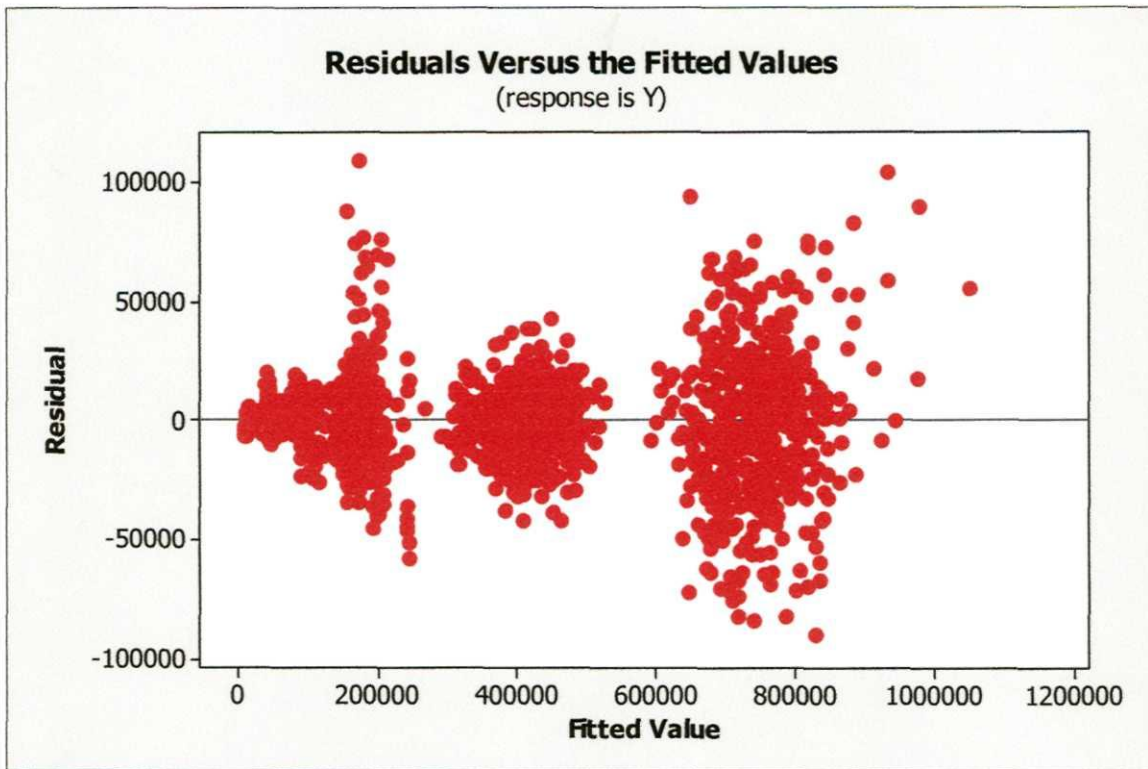
Análisis de residuales

A continuación se presentan las gráficas para el análisis de residuales.



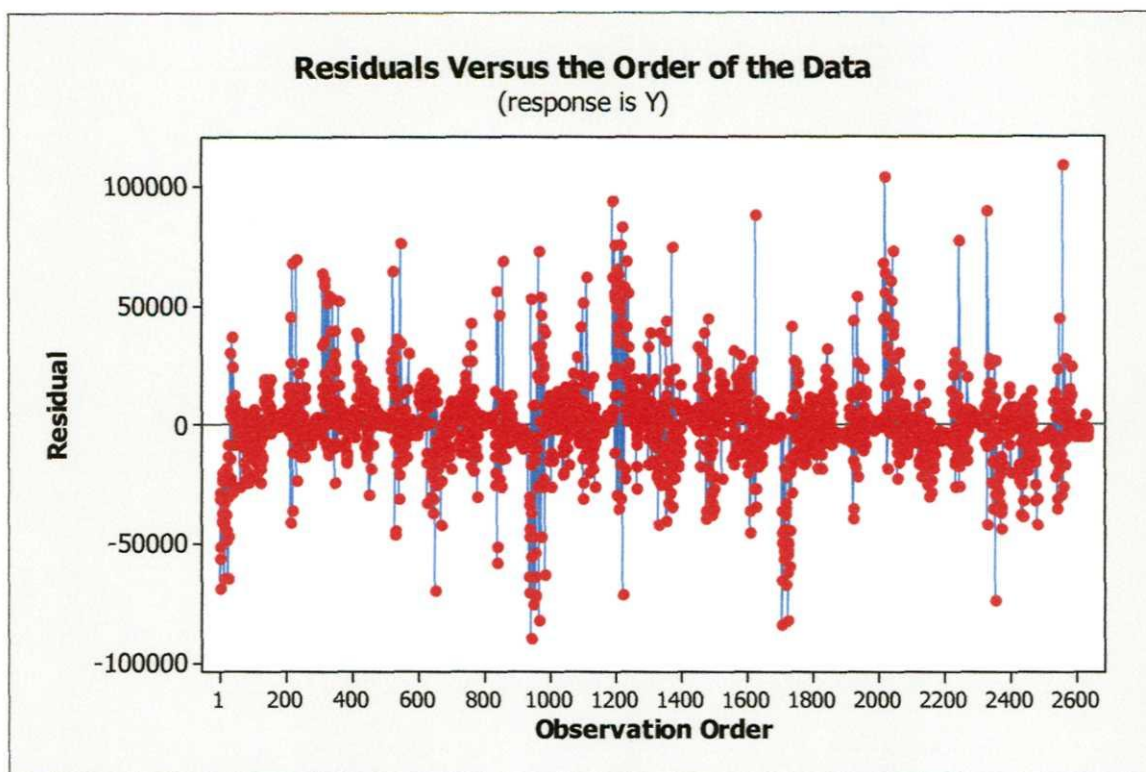
Gráfica 4.30. Histoarama de residuales etapa 4 - Combo.

El histograma de residuales arroja información sobre la existencia de datos atípicos debido a la presencia de barras pequeñas en los extremos de la gráfica. Por este motivo se procede al tratamiento de las observaciones atípicas por medio de su eliminación de la base de datos.



Gráfica 4.31. Residuales versus valores ajustados etapa 4 - Combo.

La gráfica de residuales versus los valores ajustados sugiere que la varianza de los errores es constante debido a que no se observa una tendencia en el comportamiento de los datos.



Gráfica 4.32. Residuales versus el orden de los datos etapa 4 - Combo.

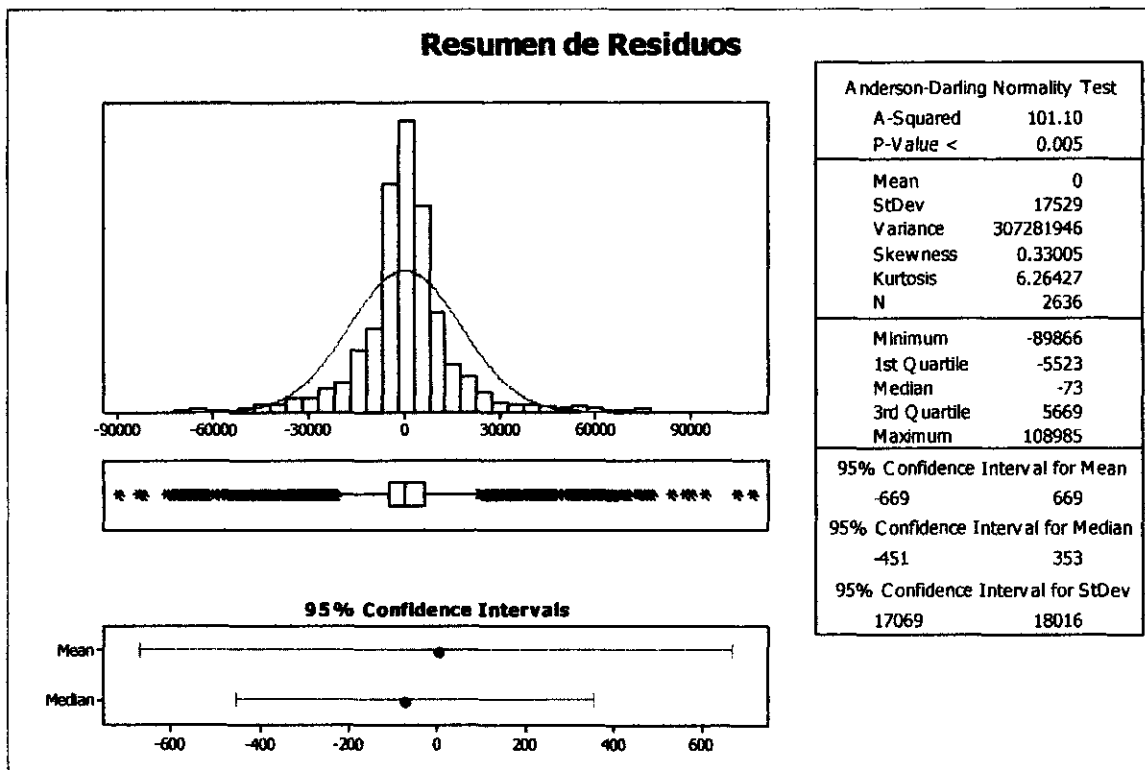
La gráfica de residuales versus el orden de los datos muestra la presencia de diversas observaciones que se alejan de la línea central y que se encuentran separados significativamente al resto de los puntos.

A continuación se revisan los supuestos por medio de análisis más formales.

Normalidad

Para verificar la normalidad de los residuales, se presenta la gráfica de la prueba de normalidad Anderson-Darling, la cual revela que con un valor p menor a 0.005, que existe evidencia estadística suficiente para rechazar el supuesto de normalidad.

El análisis de los residuos estandarizados que se ajustan a una distribución normal ocurre cuando los valores se encuentran en el rango ± 2 . Al estudiar el anexo 24, se observa que los valores no se encuentran en dicho rango, debido a la presencia de las observaciones que se identifican como puntos atípicos del conjunto de datos.



Gráfica 4.33. Resumen para residuales etapa 4 - Combo.

A pesar de que los residuos no son normales, lo más importante de este diagrama es representado por una kurtosis excesiva, que se refiere a la presencia de un pico afilado de la distribución. En este caso se considera este estadístico como una característica deseable en el modelo, debido a que la mayoría los residuos se encuentran alrededor de la media igual a cero.

Se observa además en el histograma que la distribución los residuales alrededor del promedio es bastante simétrica e indica tener varianza constante, lo cual sugiere el cumplimiento de homocedasticidad, lo cual se considera de varianza "homogénea" dado a que los residuos se encuentran distribuidos con cierta uniformidad.

Aleatoriedad e independencia de los errores

La prueba no paramétrica a utilizar es la "prueba de corridas", la cual arroja los siguientes resultados:

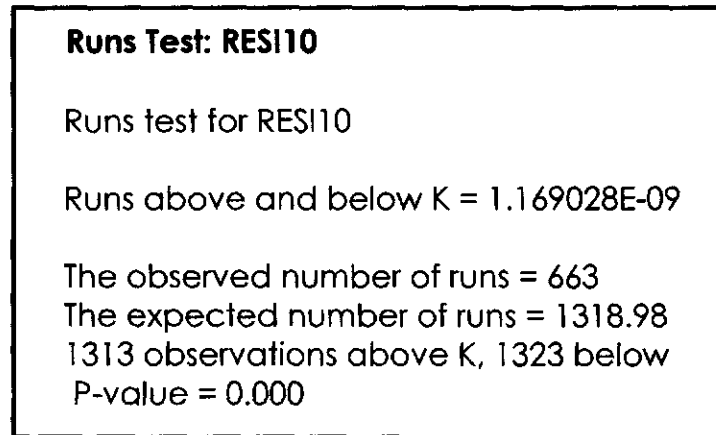


Figura 4.18. Resultados de Run Test etapa 4 - Combo.

Debido a que no se especifica la opción de otro valor que no sea la media como criterio de comparación (K), la media igual a 1.169028E-09 es utilizada. Se realizaron 663 corridas. Para determinar un número inusual de corridas, se calcula el número de observaciones que se encuentran por arriba o debajo de K. El resultado del valor-p (0.000) es menor que el nivel de alpha de 0.05, por tanto no hay evidencia estadística concluyente que indique que los datos no son aleatorios.

Homogeneidad

Se utiliza la gráfica de residuales versus los valores ajustados para determinar que existe varianza homogénea debido a la uniformidad de la distribución de los residuos.

4.2.4.4.3. Prueba de hipótesis en regresión lineal múltiple

Prueba de la significancia de la regresión

Esta prueba permite conocer si existe una relación lineal entre la respuesta y cualquiera de las variables regresoras x_1, x_2, \dots, x_k .

Las pruebas de hipótesis que se evalúan son:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_a: \text{por lo menos una } \beta_i \neq 0 \ (i = 1, \dots, k)$$

El estadístico de prueba se calcula según los datos arrojados por Minitab como:

$$F_0 = \frac{8.13437E + 12}{309869089} = 26251.00$$

Para probar la hipótesis nula, se utiliza el criterio de rechazo siguiente:

$$F_0 > F_{\alpha, k, n-k-1}$$

Dado un valor de tablas $F_{\alpha, k, n-k-1}$ igual a 2.608284 con un nivel de significancia α igual a 0.05, se concluye que la hipótesis nula $H_0: \beta = 0$ se rechaza, indicando que al menos uno de los regresores x_k contribuye al modelo de forma significativa.

En una prueba equivalente, se puede rechazar H_0 cuando el valor-p de la prueba de F es menor que el nivel de alpha (α). En este caso, $0.00 < 0.05$, por tanto se rechaza la hipótesis nula.

Prueba sobre coeficientes individuales de regresión

Las pruebas de cualquier coeficiente de regresión son:

$$H_0: \beta_j = 0$$

$$H_a: \beta_j \neq 0$$

Se rechaza la hipótesis nula si $|t_0| > t_{\alpha/2, n-k-1}$.

Según tablas, $t_{\alpha/2, n-k-1} = 2.242685$

De acuerdo a los resultados arrojados por Minitab en la tabla 4.30., encontramos que el valor T de la variable cantidad de producto vendido al cubo (x_4^3), es menor que el estadístico de prueba indicando que no guarda una relación causal con el resto de las variables independientes del estudio, mientras que el resto de los coeficientes de regresión si lo hace a un nivel de confianza del 95%.

Predictor	T
Constant	6.75
X ₂₂	7.62
X ₄	19.45
X ₁₈	6.05
X ₃ X ₂₀	19.02
X ₃ X ₂₂	7.18
X ₄ X ₂₀	29.64
X ₄ X ₂₁	48.41
X ₄ X ₂₂	18.06
X ₄ X ₂₄	-11.36
X ₅ X ₁₃	-3.78
X ₅ X ₂₂	6.54
X ₈ X ₂₀	15.59
X ₈ X ₂₄	13.87
X ₁₀ X ₁₃	3.38
X ₁₇ X ₁₃	8.57
X ₁₇ X ₂₂	7.87
X ₁₈ X ₁₃	5.35
X ₁₈ X ₂₀	18.23
X ₁₈ X ₂₂	-5.75
X ₁₈ X ₂₄	14.88
X ₄ ²	3.04
X ₄ ³	-2.19

Tabla 4.30. Valores T - Combo.

4.2.4.4.4. Análisis de varianza (ANOVA)

El análisis de varianza para el modelo Combo se muestra en la tabla siguiente:

Source	DF	SS	MS	F	P
Regression	22	1.78956E+14	8.13437E+12	26251.00	0.000
Residual Error	2613	8.09688E+11	309869089		
Total	2635	1.79766E+14			

Tabla 4.31. Tabla ANOVA - Combo.

El valor-p igual a 0.000 indica que el modelo estimado por el procedimiento de regresión es significativo a un nivel α de 0.05, indicando que al menos un coeficiente es diferente de cero.

Las hipótesis para el ANOVA son:

$$H_0: MS_R \leq MS_{RES}$$

$$H_a: MS_R > MS_{RES}$$

El estadístico de prueba de la ecuación es $F_0 = MS_R/MS_{RES}$, y el criterio de rechazo para H_0 es $F_0 > F_{\alpha, 1, n-2}$.

Según los datos arrojados por Minitab, el estadístico de prueba F_0 es igual a 26251.00 y el valor F de tablas a un nivel de significancia de 0.05 es igual a 3.844991.

Siendo $F_0 > F_{\alpha, 1, n-1}$, la hipótesis nula cae en la región de rechazo, lo que significa que la variación estimada por el modelo es mayor que la variación aleatoria, y por consiguiente se puede decir que el modelo es bueno.

La tabla 4.32. de los grados de libertad y la suma al cuadrado secuencial (Seq SS) de los estimados, indican que las variables formadas por las interacciones: cantidad de socios parciales con Delicatessen (x_5x_{22}), venta por hora hombre con la sucursal C ($x_{18}x_{13}$), y cantidad de producto vendido elevado al cubo (x_4^3), no explican una cantidad substancial de varianza única, lo cual sugiere la posibilidad de eliminar dichas variables del modelo.

Source	DF	Seq SS
X ₂₂	1	1.27456E+13
X ₄	1	5.46399E+12
X ₁₈	1	5.52769E+13
X ₃ X ₂₀	1	3.82382E+13
X ₃ X ₂₂	1	38041526270
X ₄ X ₂₀	1	6.33940E+13
X ₄ X ₂₁	1	1.99049E+12
X ₄ X ₂₂	1	3.16673E+11
X ₄ X ₂₄	1	3112527874
X ₅ X ₁₃	1	3.97702E+11
X ₅ X ₂₂	1	31373294
X ₈ X ₂₀	1	88865388691
X ₈ X ₂₄	1	3.53685E+11
X ₁₀ X ₁₃	1	2.61682E+11
X ₁₇ X ₁₃	1	1.20312E+11
X ₁₇ X ₂₂	1	18930568394
X ₁₈ X ₁₃	1	1070028305
X ₁₈ X ₂₀	1	1.56893E+11
X ₁₈ X ₂₂	1	16274476140
X ₁₈ X ₂₄	1	65941235040
X ₄ ²	1	6368518805
X ₄ ³	1	1483531534

Tabla 4.32. Grados de libertad y suma al cuadrado secuencial - Combo.

4.2.4.4.4.5. Intervalos de confianza

A continuación se muestran los intervalos de confianza del modelo Combo a un nivel de significancia del 95%.

Intervalos de confianza de los coeficientes de regresión

En la tabla 4.33. se presentan los estimados de intervalo de confianza de los coeficientes de regresión β_j .

Coeficientes	95% Intervalo de Confianza para β	
	Límite Inferior	Límite Superior
Constante	5253.092	9555.895
X ₂₂	53360.48	90359.61
X ₄	3.660246	4.480927
X ₁₈	41.06611	80.48806
X ₃ X ₂₀	29852.58	36715.47
X ₃ X ₂₂	9170.626	16059.76
X ₄ X ₂₀	16.93182	19.33089
X ₄ X ₂₁	32.18907	34.90657
X ₄ X ₂₂	10.17087	12.64839
X ₄ X ₂₄	-7.63689	-5.38929
X ₅ X ₁₃	-1324.31	-419.822
X ₅ X ₂₂	10785.56	20031.53
X ₈ X ₂₀	196.9077	253.5601
X ₈ X ₂₄	121.951	162.1105
X ₁₀ X ₁₃	0.021529	0.080944
X ₁₇ X ₁₃	0.046757	0.074507
X ₁₇ X ₂₂	0.099003	0.164741
X ₁₈ X ₁₃	10.62807	22.92131
X ₁₈ X ₂₀	331.4414	411.3249
X ₁₈ X ₂₂	-80.0126	-39.321
X ₁₈ X ₂₄	1301.493	1696.5
X ₄ ²	1.47E-05	6.78E-05
X ₄ ³	-3.6E-10	-2E-11

Tabla 4.33. Intervalo de confianza - Combo.

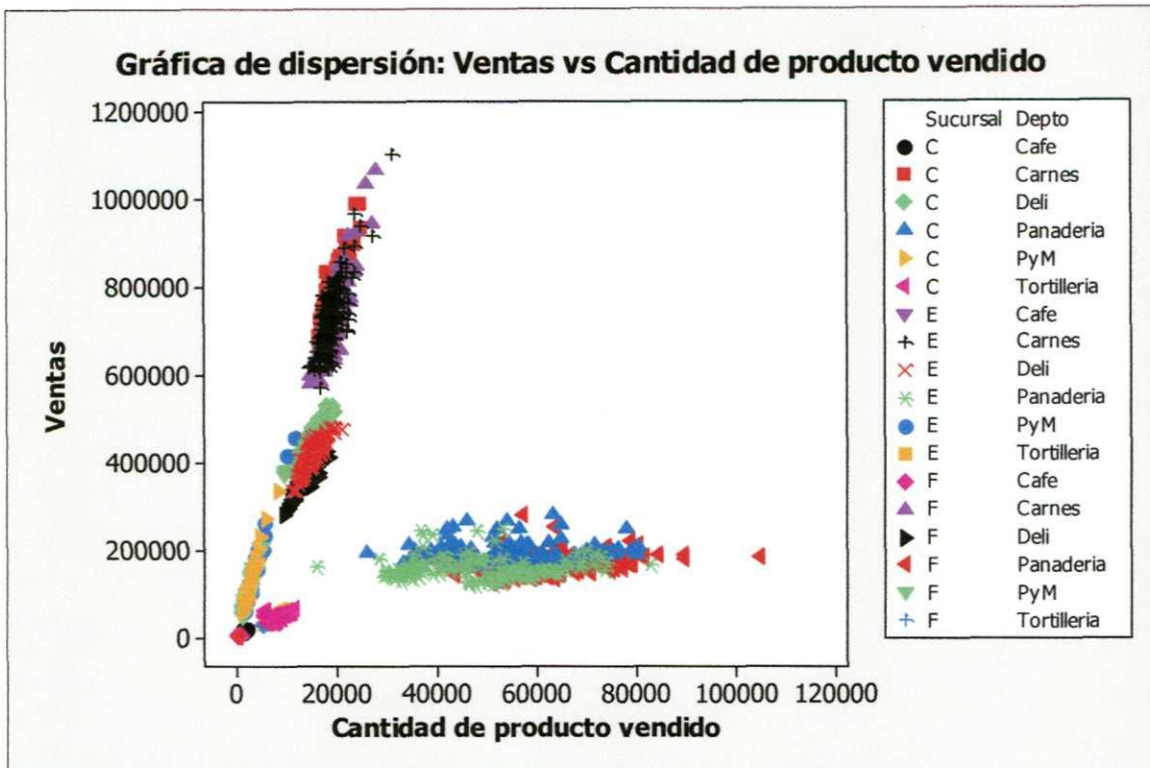
La construcción de estimaciones de intervalos de confianza es útil para establecer si el modelo es preciso. En este caso los intervalos de confianza de la mayor parte de las variables son amplios, lo que sugiere que la aportación de las variables al modelo es baja. Sin embargo, se observa que variables tales como cantidad de producto vendido y sus interacciones poseen intervalos de confianza reducidos.

Estimación del intervalo de confianza de la respuesta media

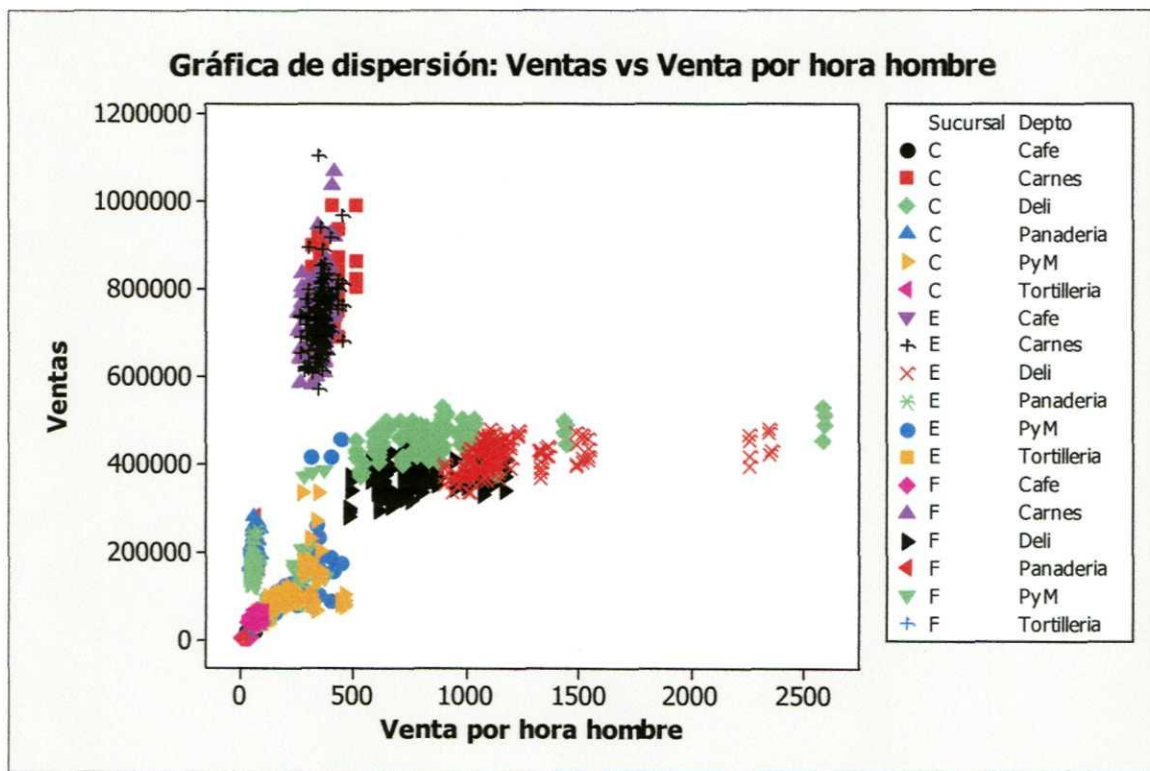
Los intervalos de confianza de la respuesta media, así como los intervalos de predicción del modelo a un nivel del 95%, se presentan en el anexo 25.

4.2.4.4.4.6. Gráficas de dispersión

En las gráficas de dispersión 4.34. y 4.35., se observa un comportamiento lineal de las ventas con respecto a la variable independiente graficada.



Gráfica 4.34. Ventas versus cantidad de producto vendido etapa 4 - Combo.



Gráfica 4.35. Ventas versus venta por hora hombre etapa 4 - Combo.

De la gráfica 4.34., se observa el comportamiento similar de las sucursales del Combo para cada uno de los departamentos analizados. La cantidad de producto vendido en Carnes, Delicatessen, y Pescados y Mariscos genera mayores ventas, aunque la cantidad de producto vendido sea pequeña. En Panadería se venden gran cantidad de productos, sin embargo las ventas no son tan grandes como en los departamentos antes mencionados. En Café y Tortillería la cantidad de producto vendido con respecto a las ventas generadas se presenta en una proporción pequeña.

Al igual que en todos los diagramas de dispersión mostrados, en la gráfica 4.35. se observa que el departamento de Carnes es el que más genera ventas, a pesar de que la venta por hora hombre es menor en comparación con el departamento de Delicatessen, en el cual se observan mayor cantidad de ventas por socio, pero menores ventas totales.

4.2.5. Análisis de sensibilidad

A través del análisis de sensibilidad se pretende conocer el cambio esperado en las ventas por cambio unitario en x_i cuando el resto de las variables independientes se mantienen constantes.

El análisis de sensibilidad se aplica en las variables de mayor importancia en la ecuación de regresión del modelo final de A, con el objeto de ejemplificar los cambios en el comportamiento de las ventas.

Se decide estudiar el comportamiento de la variable horas disponibles de trabajo efectivo con respecto al margen de ganancia cuando es bajo (-10%) y alto (+10%). El cambio en las ventas generadas se observa en base al incremento o decremento del 5, 10 y 15% del valor de la media de los datos de las horas de trabajo efectivo.

La media para la variable horas de trabajo se encuentra en un promedio de ventas de \$13,153,340.76 pesos. El incremento en las ventas a medida que el porcentaje de horas aumenta es de \$1,808.66 pesos.

En la tabla 4.34., se observan tres escenarios del comportamiento de las ventas totales de A cuando se analizan las horas con respecto a un margen de ganancia constante, bajo y alto.

De la gráfica 4.34., se observa el comportamiento similar de las sucursales del Combo para cada uno de los departamentos analizados. La cantidad de producto vendido en Carnes, Delicatessen, y Pescados y Mariscos genera mayores ventas, aunque la cantidad de producto vendido sea pequeña. En Panadería se venden gran cantidad de productos, sin embargo las ventas no son tan grandes como en los departamentos antes mencionados. En Café y Tortillería la cantidad de producto vendido con respecto a las ventas generadas se presenta en una proporción pequeña.

Al igual que en todos los diagramas de dispersión mostrados, en la gráfica 4.35. se observa que el departamento de Carnes es el que más genera ventas, a pesar de que la venta por hora hombre es menor en comparación con el departamento de Delicatessen, en el cual se observan mayor cantidad de ventas por socio, pero menores ventas totales.

4.2.5. Análisis de sensibilidad

A través del análisis de sensibilidad se pretende conocer el cambio esperado en las ventas por cambio unitario en x_i cuando el resto de las variables independientes se mantienen constantes.

El análisis de sensibilidad se aplica en las variables de mayor importancia en la ecuación de regresión del modelo final de A, con el objeto de ejemplificar los cambios en el comportamiento de las ventas.

Se decide estudiar el comportamiento de la variable horas disponibles de trabajo efectivo con respecto al margen de ganancia cuando es bajo (-10%) y alto (+10%). El cambio en las ventas generadas se observa en base al incremento o decremento del 5, 10 y 15% del valor de la media de los datos de las horas de trabajo efectivo.

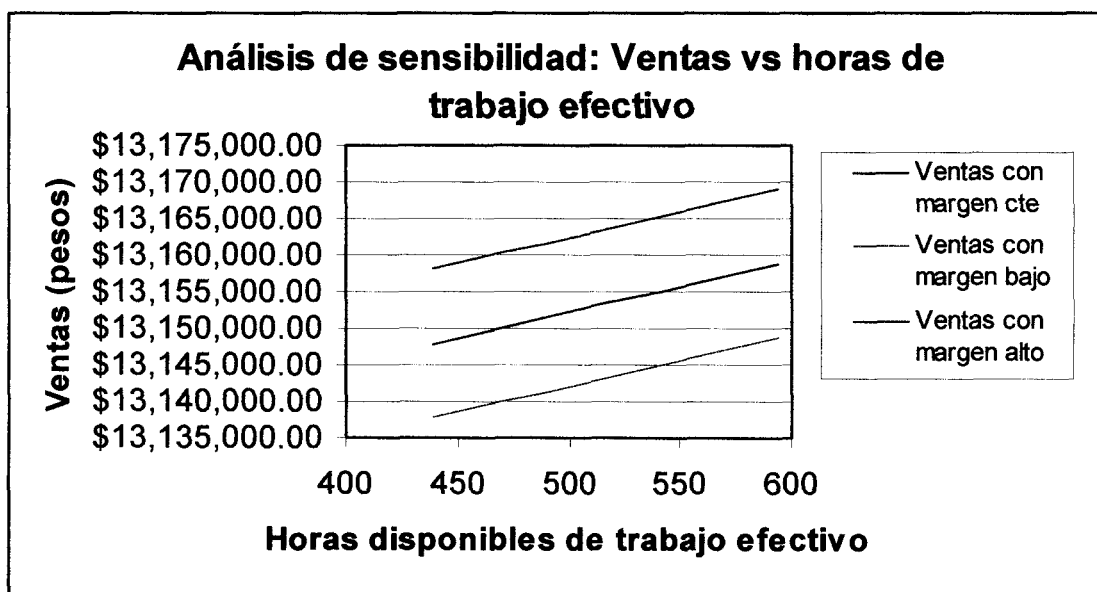
La media para la variable horas de trabajo se encuentra en un promedio de ventas de \$13,153,340.76 pesos. El incremento en las ventas a medida que el porcentaje de horas aumenta es de \$1,808.66 pesos.

En la tabla 4.34., se observan tres escenarios del comportamiento de las ventas totales de A cuando se analizan las horas con respecto a un margen de ganancia constante, bajo y alto.

	Horas de trabajo	Ventas con margen cte	Ventas con margen bajo	Ventas con margen alto
$\mu-15\%$	438.56	\$13,147,914.80	\$13,137,751.59	\$13,158,078.01
$\mu-10\%$	464.35	\$13,149,723.45	\$13,139,560.24	\$13,159,886.66
$\mu-5\%$	490.15	\$13,151,532.11	\$13,141,368.90	\$13,161,695.32
μ	515.95	\$13,153,340.76	\$13,143,177.55	\$13,163,503.98
$\mu+5\%$	541.75	\$13,155,149.42	\$13,144,986.21	\$13,165,312.63
$\mu+10\%$	567.54	\$13,156,958.08	\$13,146,794.87	\$13,167,121.29
$\mu+15\%$	593.34	\$13,158,766.73	\$13,148,603.52	\$13,168,929.94

Tabla 4.34. Análisis de sensibilidad horas de trabajo efectivo.

Las rectas de la gráfica 4.36., indican que las ventas incrementan a medida que aumentan las horas disponibles de trabajo efectivo. Las ventas son mayores cuando el margen de ganancia se encuentra en un nivel alto. En este escenario se requiere del incremento en las horas de trabajo efectivo, por lo que es importante tomar en consideración los costos a los que se puede incurrir al poner en práctica esta sugerencia, como lo son el incremento en nómina y los gastos de operación al mantener la tienda más tiempo al servicio de los consumidores.



Gráfica 4.36. Análisis de sensibilidad ventas vs. horas de trabajo efectivo.

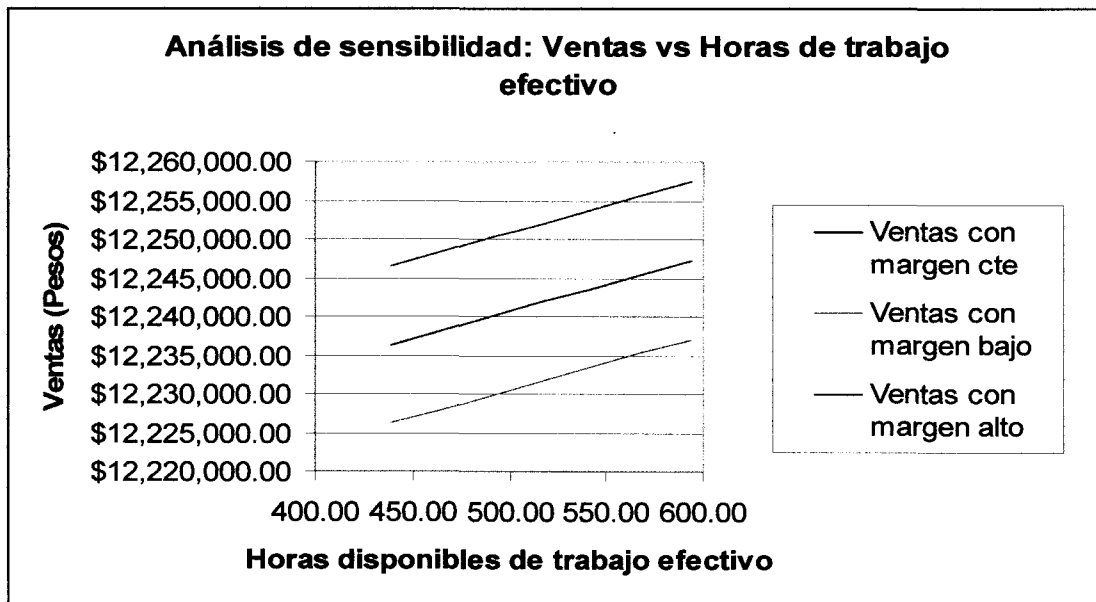
Por otro lado, se decide observar el cambio en las ventas de A cuando el departamento de Carnes se encuentra o no presente en la ecuación de regresión.

En la tabla 4.35., se observan los cambios en las ventas cuando en el modelo no se incluye la variable correspondiente al departamento de Carnes.

	Horas de trabajo	Ventas con margen cte	Ventas con margen bajo	Ventas con margen alto
$\mu-15\%$	438.56	\$12,236,354.59	\$12,226,191.38	\$12,246,517.80
$\mu-10\%$	464.35	\$12,238,163.25	\$12,228,000.04	\$12,248,326.46
$\mu-5\%$	490.15	\$12,239,971.90	\$12,229,808.69	\$12,250,135.12
μ	515.95	\$12,241,780.56	\$12,231,617.35	\$12,251,943.77
$\mu+5\%$	541.75	\$12,243,589.22	\$12,233,426.01	\$12,253,752.43
$\mu+10\%$	567.54	\$12,245,397.87	\$12,235,234.66	\$12,255,561.08
$\mu+15\%$	593.34	\$12,247,206.53	\$12,237,043.32	\$12,257,369.74

Tabla 4.35. Análisis de sensibilidad horas de trabajo efectivo sin Depto. Carnes.

Las pendientes de las rectas indican que las ventas de A son mayores cuando las horas de trabajo efectivo incrementan y el margen de ganancia se encuentra en un nivel alto.

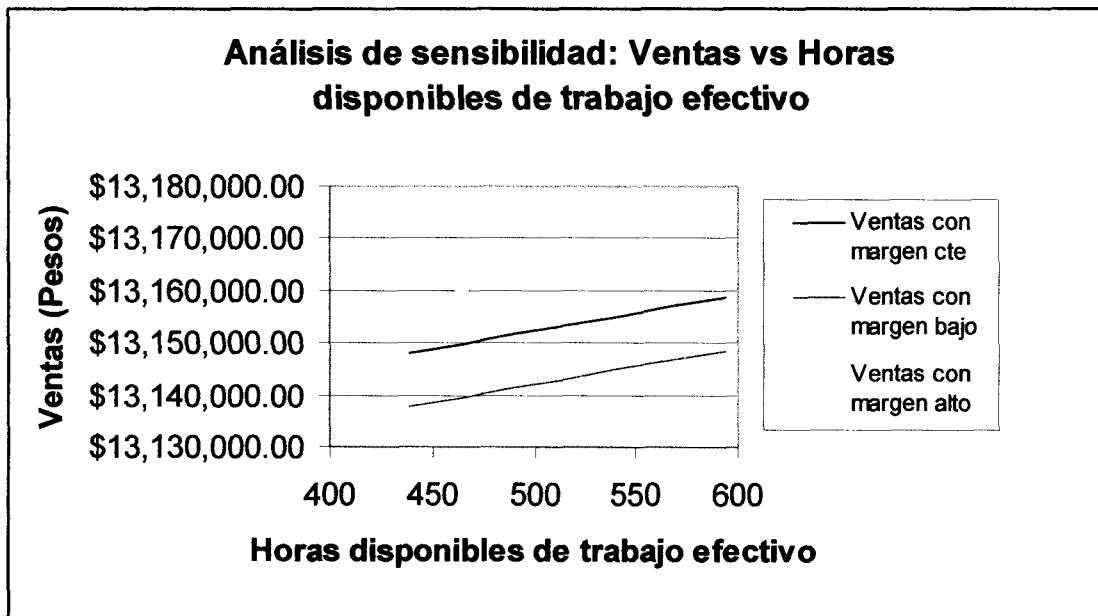


Gráfica 4.37. Análisis de sensibilidad ventas vs. horas de trabajo efectivo sin Carnes.

En la tabla 4.36., se observan los cambios en las ventas cuando se incluye en el modelo la variable departamento de Carnes.

	Horas de trabajo	Ventas con margen cte	Ventas con margen bajo	Ventas con margen alto
$\mu-15\%$	438.56	\$13,147,914.80	\$13,137,751.59	\$13,158,078.01
$\mu-10\%$	464.35	\$13,149,723.45	\$13,139,560.24	\$13,159,886.66
$\mu-5\%$	490.15	\$13,151,532.11	\$13,141,368.90	\$13,161,695.32
μ	515.95	\$13,153,340.76	\$13,143,177.55	\$13,163,503.98
$\mu+5\%$	541.75	\$13,155,149.42	\$13,144,986.21	\$13,165,312.63
$\mu+10\%$	567.54	\$13,156,958.08	\$13,146,794.87	\$13,167,121.29
$\mu+15\%$	593.34	\$13,158,766.73	\$13,148,603.52	\$13,168,929.94

Tabla 4.36. Análisis de sensibilidad horas de trabajo efectivo con Depto. Carnes.



Gráfica 4.38. Análisis de sensibilidad ventas vs. horas de trabajo efectivo con Carnes.

Los resultados de la tabla 4.35. y la gráfica 4.38., arrojan que las ventas son superiores cuando opera el departamento de Carnes. Las ventas sin el departamento de Carnes son de \$12,241,780.56 y con la presencia del departamento son de \$13,153,340.76, esto cuando las horas disponibles de trabajo efectivo se encuentran en el valor de la media y el margen de ganancia es constante.

Como se había supuesto, las ventas de A son superiores cuando existe el departamento de Carnes. Se recomienda aumentar la cantidad de horas disponibles de trabajo efectivo con el objeto de incrementar las ventas totales.

4.3. Conclusiones

Los resultados estadísticos que fueron descritos en este capítulo permiten llegar a dos conclusiones principales: (1) los cuatro modelos finales tienen un alto nivel de confiabilidad para el pronóstico de observaciones futuras, y (2) las variables de productividad presentes en cada ecuación de regresión, impactan de manera significativa en el incremento o decremento de las ventas de un supermercado.

A través del análisis de sensibilidad fue posible conocer el comportamiento de las ventas según el impacto de los valores de las variables independientes y los parámetros β_j . Encontramos que las ventas fluctúan de forma importante en el modelo al modificar el valor de la variable estudiada; sin embargo, favorablemente en la mayoría de los casos el incremento en el valor del regresor generó un aumento considerable en las ventas totales.

Capítulo 5. Conclusiones

5.1. Conclusiones

Este proyecto de tesis representa la oportunidad de aplicar en un contexto real uno de los temas de mayor interés en el área de estudio de la Ingeniería Industrial, el análisis de regresión lineal. El conocimiento sobre el uso de esta herramienta resulta ser de gran utilidad para el desarrollo de modelos matemáticos para la predicción de ventas futuras que se presentaron en la sección de resultados del Capítulo 4.

5.1.1. Del proceso

El proceso de elaboración del presente trabajo se considera un reto enriquecedor para el analista ya que le permite aprender sobre la historia de las empresas detallistas de autoservicio, así como de los estudios dirigidos por investigadores que identificaron las variables de productividad de interés para conducir un análisis de regresión bajo diversos criterios.

La revisión bibliográfica permite adquirir conocimiento sobre la herramienta estadística que se utiliza para la evaluación de la información y la generación de resultados que fueron validados por el experto de la empresa.

De la metodología se aprende sobre la importancia de seguir una serie de pasos estructurados que permitan lograr un objetivo; por esto, es necesario conocer el área de aplicación del estudio, definir las variables que se desean evaluar y proseguir con la recolección de datos para su análisis estadístico por medio de la aplicación de la herramienta.

Cabe mencionar que la oportunidad de utilizar regresión lineal para estudiar el trabajo que realiza una de las cadenas de supermercados más reconocidas en el norte del país, permite desarrollar habilidades analíticas sobre los factores que intervienen en el proceso productivo de una empresa del sector de servicios de venta de alimentos al detalle.

5.1.2. De la investigación

El desarrollo de la investigación se relaciona principalmente con el estudio de la información relacionada con las operaciones de la industria detallista de autoservicio, la cual tiene como objetivo principal alcanzar altos niveles de rentabilidad y productividad. Por esto es que los

indicadores de productividad para medir el desempeño de una empresa pueden variar según las necesidades de cada una.

Las variables de productividad identificadas en la revisión de la literatura son las horas de trabajo efectivo y los empleados clasificados como parciales o completos. Así como los niveles de producto que pueden medirse en términos de transacciones realizadas, unidades físicas y ventas.

De la investigación se aprende sobre el Índice Nacional de Precios al Consumidor (INPC), el cual permite determinar los precios de los alimentos en México, y su importancia radica en su uso para eliminar el efecto de la inflación al transformar los datos financieros en valores monetarios actuales.

Por otro lado, el proceso de investigación permite concluir que muchos detallistas han solicitado el uso de sofisticados métodos estadísticos para la predicción de ventas, entre ellos encontramos el análisis de regresión múltiple, el análisis de discriminante múltiple, y los modelos de gravedad.

Sin embargo, del estudio de las diversas opciones se decide utilizar el análisis de regresión lineal múltiple, la cual es una técnica estadística para el modelado de la relación entre dos o más variables. Esta técnica puede ser utilizada para construir un modelo matemático que permite estimar el comportamiento de la variable dependiente, a partir de valores específicos de las variables independientes.

5.1.3. Del uso de la herramienta

Además de conocer el impacto financiero sobre las utilidades que puede generar la empresa al conocer el comportamiento de las ventas en algunas de las operaciones de sus departamentos clave, un análisis de regresión provee una técnica estadística que se adapta al propósito específico de las necesidades e intereses particulares de la organización, ya que provee información para la evaluación de los factores de rendimiento existentes, siendo con ello una herramienta de gran utilidad para la toma de decisiones importantes.

Un análisis de regresión es tan fiable como los datos sobre los que se basa. Mediante el uso de datos históricos se corre el riesgo de sufrir de errores de transcripción y otros problemas con la calidad de los mismos. Estos errores hacen que los datos históricos sean propensos a tener observaciones atípicas poco deseables.

Por consiguiente, se recomienda llevar a cabo un procedimiento formal para la recolección exacta y fiable de información que sea de utilidad para vigilar el proceso y que tenga influencia significativa sobre la variable de respuesta.

El uso de la técnica de regresión lineal no siempre es útil para resolver problemas inherentes de los datos del modelo. En ocasiones las pruebas paramétricas no funcionan adecuadamente para estudiar el cumplimiento de los supuestos de normalidad, homocedasticidad, independencia y aleatoriedad de los residuales. De la misma forma ocurre con las herramientas de regresión Ridge, las cuales no aseguran la disminución de los efectos de la multicolinealidad en el modelo.

La aplicación de la herramienta estadística no debe limitarse solamente al uso de Minitab como paquete de cómputo. Existe diversos software capaces de procesar comandos para variados propósitos de regresión como los son SAS, SPSS, Mathlab, entre otros.

5.1.4. De las pruebas de hipótesis

Uno de los propósitos de esta tesis consiste en concretar las pruebas de hipótesis que se determinaron como guías de investigación. En tabla 5.1., se hace un reconocimiento de las hipótesis que se rechazan de acuerdo a cada uno de los modelos finales generados; la descripción de las mismas se encuentra en el Capítulo 1, sección 1.4.

Número	A	B	D	Combo
1	Se rechaza	Se rechaza	Se rechaza	Se rechaza
2	No se rechaza	No se rechaza	No se rechaza	No se rechaza
3	Se rechaza	Se rechaza	No se rechaza	Se rechaza
4	No se rechaza	No se rechaza	No se rechaza	No se rechaza
5	No se rechaza	No se rechaza	No se rechaza	No se rechaza
6	No se rechaza	No se rechaza	No se rechaza	No se rechaza
7	Se rechaza	No se rechaza	No se rechaza	No se rechaza
8	No se rechaza	Se rechaza	No se rechaza	No se rechaza
9	No se rechaza	Se rechaza	No se rechaza	No se rechaza
10	Se rechaza	Se rechaza	Se rechaza	No se rechaza
11	No se rechaza	No se rechaza	No se rechaza	Se rechaza

Tabla 5.1. Resumen de los resultados de pruebas de hipótesis.

En resumen, las pruebas de hipótesis para cada uno de los modelos finales presentados en la tabla 5.1., indican que existe relación significativa entre las variables independientes del estudio y las ventas de las sucursales A, B, D y Combo.

Las ventas de la sucursal A son afectadas significativamente por la cantidad de producto vendido, las horas disponibles de trabajo efectivo, y el margen de ganancia.

En el modelo para B se observa que las variables cantidad de producto vendido, nómina por departamento, *add loss*, y margen de ganancia, impactan significativamente a las ventas de la tienda. Así mismo, encontramos que existe relación entre la variable margen de ganancia y las ventas de la sucursal D.

Finalmente en el modelo de Combo se observa que las ventas se relacionan significativamente con la cantidad de producto vendido y con la venta por hora hombre.

5.1.5. De los resultados

Los objetivos planteados inicialmente en esta tesis, se cumplieron a través de los resultados estadísticos obtenidos en el Capítulo 4, los cuales permiten concluir que los cuatro modelos finales son potencialmente buenos para la predicción de observaciones futuras, referentes a las ventas de las sucursales estudiadas.

Por medio del análisis de regresión se logra identificar la influencia significativa de las variables de productividad definidas en un principio del estudio, de modo tal que fue posible eliminar aquellas que no tenían una relación de importancia en la descripción del comportamiento de las ventas en los departamentos de manufactura de alimentos de: Carnes, Pescados y Mariscos, Delicatessen, Café, Panadería y Tortillería.

Por consiguiente, cada una de las ecuaciones de regresión generadas para los modelos de A, B, D, y Combo, se conforman por las variables de productividad que resultaron ser significantes a un nivel de confianza del 95%.

Se concluye de los resultados que la tienda A genera mayores ventas de la cantidad de producto que se comercializa en los departamentos de Carnes, Pescados y Mariscos y Delicatessen. Además se observa que a medida que las horas de trabajo efectivo incrementan, las ventas de la sucursal mejoran y el margen de ganancia es mayor. Por tanto se sugiere la mejora continua en las operaciones de los departamentos antes

mencionados, por medio de mejores prácticas de manufactura que permitan eliminar desperdicios y reducir tiempos muertos.

En la sucursal B se identifica que las ventas totales se ven influenciadas por la cantidad de producto vendido, por la nómina asignada para salarios en cada uno de los departamentos estudiados, así como por el margen de ganancia. Es recomendable aplicar estrategias de mercadotecnia para promocionar los productos del departamento de Panadería, ya que las ventas semanales se observan en mayor medida cuando se trata de una quincena, además de que las ganancias de la cantidad de producto vendido no se reflejan de forma importante en las ventas totales de la tienda. Así mismo, se sugiere estudiar las rebajas que se aplican a los productos del departamento de Carnes, ya que es uno de los que genera mayor cantidad de ventas.

De la tienda D es posible concluir que el departamento de Carnes tiene mayor influencia sobre las ventas, debido a que se vende mayor cantidad de kilos de carne los días que corresponden a semana de quincena de pago salarial, además de que la flexibilidad de la fuerza laboral de los socios de medio tiempo permite que las ventas por hora hombre sean mayores y que los costos de operación disminuyan. Por otro lado, se recomienda emplear trabajadores de tiempo completo que tengan experiencia y capacitación sobre los productos que ofrecen en el departamento de Pescados y Mariscos. Este último junto con el departamento de Delicatessen, contribuyen de forma importante en la cantidad de ventas totales.

El modelo final que corresponde al Combo formado por las tiendas C, E y F, permite concluir que las ventas se generan principalmente de la cantidad de producto vendido en los departamentos de Carnes, Pescados y Mariscos, y Delicatessen. Se refleja que en los días de quincena se venden más kilos de carne y de productos de salchichonería. Para incrementar las ventas, se recomienda aumentar las horas de trabajo efectivo en el departamento de Carnes y Panadería, lo cual lleva a la generación de mayores ventas por hora hombre. Se observa además que la sucursal C contribuye de forma importante sobre las ventas debido a las rebajas en los precios de los productos de gran demanda, así como en el incremento del margen de ganancia. Se sugiere mejorar las tareas en las tiendas E y F, es necesario optimizar las actividades para que los departamentos sean productivos y utilicen eficientemente los recursos disponibles.

5.2. Recomendaciones para futuras investigaciones

A pesar de contar con modelos de regresión múltiple capaces de predecir ventas a un 90% de confiabilidad, los modelos tienen la limitación concerniente al conjunto de predictores seleccionados para su análisis. En consecuencia, futuras investigaciones deben ser dirigidas de modo que se incluyan diversas características de ventas que puedan tener influencia significativa sobre la variable de respuesta, como ejemplo factores sobre el rendimiento de la tienda.

Desde la perspectiva cuantitativa, la medición de la importancia relativa de los regresores podría ser de ayuda para comprender la selección de los mismos.

Dado que los presentes modelos son estadísticos, su estabilidad debe cuestionarse. Es necesario realizar ajustes periódicos a los modelos propuestos, así como el análisis de sensibilidad relacionado con la inclusión de nuevas observaciones de las variables bajo estudio.

Referencias

- Banco de México. (2002). "Metodología para el Cambio de Base del INPC". Disponible en: <http://www.banxico.org.mx/inpc/> 12 de Marzo de 2007.
- Belongia, M. (1979). "Domestic Food Programs and Their Related Impacts on Retail Food Prices". *American Journal of Agricultural Economics*. Vol. 61, No. 2, pp. 358-362.
- Berry, L.L.; Zeithaml, V.A.; Parasuraman, A. (1985). "Quality counts in services too". *Business Horizon*. Vol. 28, No. 3, pp. 44-52.
- Browell, S. and Ivers, R. (1998). "Part-time employees in food retailing". *Managing Service Quality*. Vol. 8, No. 4, pp. 281-293.
- Bucklin, L. (1980). "Technological Change and Store Operations: The Supermarket Case". *Journal of Retailing*. Vol. 3, No. 56, pp. 3-15.
- Budd, J. W.; McCall B. P. (2001). "The Grocery Stores Wage Distribution: A Semi-Parametric Analysis of the Role of Retailing and Labor Market Institutions". *Industrial and Labor Relations Review*. Vol. 54, No. 2, pp. 484-501.
- Carey, J.; P. Otto. (1977). "Output per Unit of Labor Input in the Retail Food Store Industry". *Monthly Labor Review*. Vol. 1, No. 100, pp. 42-47.
- Cooke J. (2003). "How labor standards lift productivity". *Logistics Management*. Vol. 42, No. 3, pp. 59-62.
- Devore, J. L. (1998). "Probabilidad y estadística para ingeniería y ciencias". México: Matemáticas Internacional Thomson.
- Draper, N.; Smith H. (1981). *Applied Regression Analysis*. New York: Wiley Interscience. Second Edition.
- Dubinsky, A.J.; Nataraajan, R.; Wen-Yeh H. (2004). "The Influence of Moral Philosophy on Retail Salespeople's Ethical Perceptions". *The Journal of Consumer Affairs*. Vol. 38, No. 2, pp. 297.
- Duff, M. (2001). "Activists target seafood market, seek compliance from retailers". *DSN Retailing Today*. Vol. 40, No. 9, pp. 19-21.

- Geber, G. (1994). "Making the most of a flexible work force". *Recruitment, Retention and Selection*. Vol. 3, No. 3, pp. 19-20.
- Gujarati, D. N. (2004). "Econometría". México: McGraw-Hill.
- Guajardo, C. G. (2002). "Contabilidad Financiera". México: McGraw-Hill.
- Hernández S.; Fernández; Baptista. (2003). "Metodología de la Investigación". México: McGraw-Hill.
- Hines, W.; Montgomery, D. (1993). "Probabilidad y Estadística para Ingeniería y Administración". México: CECSA.
- Hoch, S.J.; Kim, B.; Montgomery A.L.; Rossi, P.E. (1995). "Determinants of Store-Level Price Elasticity". *Journal of Marketing Research*. Vol. 32, No. 2, pp. 17-29.
- Ingene, C. (1982). "Labor Productivity in Retailing". *Journal of Marketing*. Vol. 46, No. 4, pp. 75-90.
- Koenes, A. (1996). *Técnicas de la venta minorista: Un enfoque estratégico*. España: Ediciones Díaz de Santos.
- Kutner M. H.; Nachtsheim, C.J.; Neter, J.; Li, W. (2005). *Applied linear statistical models*. United States of America: McGraw Hill.
- McClelland, W.G. (1958). "Sales per Person and Size in Retailing: Some Fallacies" *The Journal of Industrial Economics*. Vol. 6, No. 3, pp. 221-229.
- McTaggart J.; Heller W. (2005). "Forces of change". *Progressive Grocer*. Vol. 84, No. 6, pp. 48
- Minichiello, R. (1967). "The Real Challenge of Food Discounters". *Journal of Marketing*. Vol.31, No. 2, pp. 37-42.
- Montgomery, D.; Peck, E.; Vining, G. (2004). *Introducción al Análisis de Regresión Lineal*. México: CECSA.
- Moore, L.V.; Diez Roux, A.V. (2006). "Associations of Neighborhood Characteristics With the Location and Type of Food Stores". *American Journal of Public Health*. Vol. 92, No. 2, pp. 325.

- Moschis, G.; Curasi, C.; Bellenger, D. (2004). "Patronage motives of mature consumers in the selection of food and grocery stores". *Journal of Consumer Marketing*. Vol. 21, No. 2, pp. 123-133.
- Mullin, T. (2005). "Reinventing Supermarkets". *Chain Store Age*. Vol. 81, No. 7, pp. 28
- Parasuraman, A.; Zeithalm, V.A.; Berry, L.L. (1988). "SERVQUAL: A multiple-item scale for measuring consumer perception of service quality". *Journal of Retailing*. Vol. 67, No. 4, pp. 420-450.
- Parks, L. (2002). "Despite gains in pharmacy, supermarket sales decline again". *Drug Store News*. Vol. 24, No. 8, pp. 8.
- Reinartz, W.J.; Kumar, V. (1999) "Store, Market, and Consumer Characteristics: The Drivers of Store Performance". *Marketing Letters*. Vol. 10, No. 1, pp. 5.
- Rogers, D. (1992). "A Review of Sales Forecasting Models Most Commonly Applied in Retail Site Evaluation". *International Journal of Retail & Distribution Management*. Vol. 20, No. 4, pp. 3, 9
- Savin, N.E.; White, K.J. (1977). "The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors". *Econometría*. Vol. 45, No. 8, pp. 1989-1996.
- Silva, A.L; Cardoso M. GMS. (2005). "Predicting supermarket sales: The use of regression trees". *Journal of Targeting, Measurement and Analysis for Marketing*. ABI/INFORM Global. Vol. 13, No. 3, pp. 239
- Thuesen, G.; Fabrycky W.J. (2001). *Engineering Economy*. Estados Unidos: Prentice Hall.
- Valenzuela, J. (2002). "Regresión Lineal Múltiple". ITESM Campus Sinaloa.
- Vida, I.; Reardon, J.; Fairhurst A. (2000). "Determinants of International Retail Involvement: The Case of Large U.S. Retail Chains". *Journal of International Marketing*. Vol. 8, No. 4, pp. 37-60.
- Walters, R.G.; MacKenzie S.B. (1988). "A Structural Equations Analysis of the Impact of Price Promotions on Store Performance". *Journal of Marketing Research*. Vol. 25, No.1, pp. 51-63.

Wong, A.; Sohal, A. (2003). "Assessing customer-salesperson interactions in a retail chain: Differences between city and country retail districts". *Marketing Intelligence & Planning*. Vo. 21, No. 21, pp. 292.

Xin-lu M.; Lin-du Z. (2005). "Credit System of the Food Safety in Retail Trade". *Institute of System Engineering*. Vol. 9, No. 6, pp. 1242

Yan, H., Adebajo, D.; Mahoney, F. (2004). "Competitive Advantage in Food Retail Strategy and supply Chain – An Analysis of the UK and Chinese Markets". *IEEE*. Vol. 5, No. 4, pp. 90-93.

Anexos**Anexo 1. Índice Nacional de Precios al Consumidor 2003**

Actualización: 28/septiembre/2005, información vigente.

Para determinar el pago de las contribuciones y sus accesorios, se aplicará el índice nacional de precios al consumidor (INPC) que es calculado por el Banco de México.

Tabla I.- Índice General de Inflación
Base: 2ª quincena de Junio de 2002.

MES	AÑO 2003
ENERO	103.320
FEBRERO	103.607
MARZO	104.261
ABRIL	104.439
MAYO	104.102
JUNIO	104.188
JULIO	104.339
AGOSTO	104.652
SEPTIEMBRE	105.275
OCTUBRE	105.661
NOVIEMBRE	106.538
DICIEMBRE	106.996

Anexo 2. Índice Nacional de Precios al Consumidor 2004

Actualización: 28/septiembre/2005, información vigente.

Para determinar el pago de las contribuciones y sus accesorios, se aplicará el índice nacional de precios al consumidor (INPC) que es calculado por el Banco de México.

Tabla 2.- Índice General de Inflación
Base: 2ª quincena de Junio de 2002.

MES	AÑO 2004
ENERO	107.661
FEBRERO	108.305
MARZO	108.672
ABRIL	108.836
MAYO	108.563
JUNIO	108.737
JULIO	109.022
AGOSTO	109.695
SEPTIEMBRE	110.602
OCTUBRE	111.368
NOVIEMBRE	112.318
DICIEMBRE	112.550

Anexo 3. Índice Nacional de Precios al Consumidor 2005

Actualización: 15/febrero/2006, información vigente.

Para determinar el pago de las contribuciones y sus accesorios, se aplicará el índice nacional de precios al consumidor (INPC) que es calculado por el Banco de México.

Tabla 3.- Índice General de Inflación
Base: 2ª quincena de Junio de 2002.

MES	AÑO 2005
ENERO	112.554
FEBRERO	112.929
MARZO	113.438
ABRIL	113.842
MAYO	113.556
JUNIO	113.447
JULIO	113.891
AGOSTO	114.027
SEPTIEMBRE	114.484
OCTUBRE	114.765
NOVIEMBRE	115.591
DICIEMBRE	116.301

Anexo 4. Índice Nacional de Precios al Consumidor 2006

Actualización: 9/febrero/2007, información vigente.

Para determinar el pago de las contribuciones y sus accesorios, se aplicará el índice nacional de precios al consumidor (INPC) que es calculado por el Banco de México.

Tabla 4.- Índice General de Inflación
Base: 2ª quincena de Junio de 2002.

MES	AÑO 2006
ENERO	116.983
FEBRERO	117.162
MARZO	117.309
ABRIL	117.481
MAYO	116.958
JUNIO	117.059
JULIO	117.380
AGOSTO	117.979
SEPTIEMBRE	119.170
OCTUBRE	119.691
NOVIEMBRE	120.319
DICIEMBRE	121.015

Anexo 5. Índice Nacional de Precios al Consumidor 2007

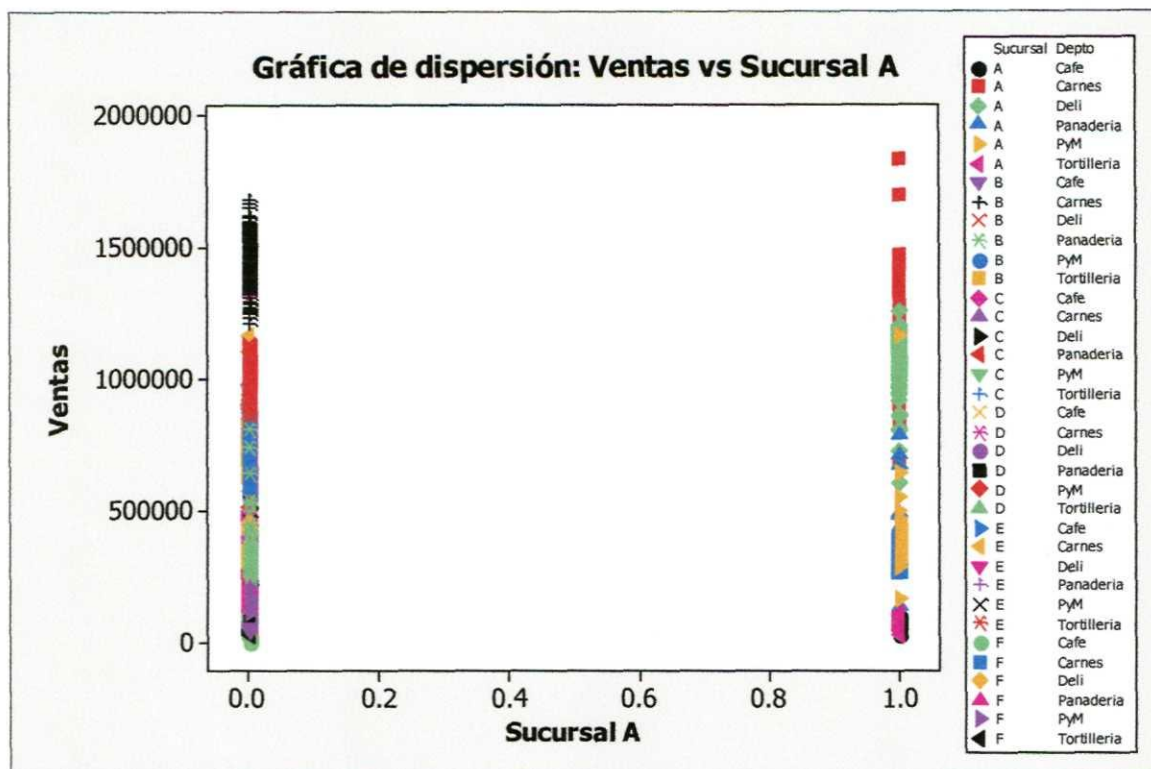
Actualización: 10/mayo/2007, información vigente.

Para determinar el pago de las contribuciones y sus accesorios, se aplicará el índice nacional de precios al consumidor (INPC) que es calculado por el Banco de México.

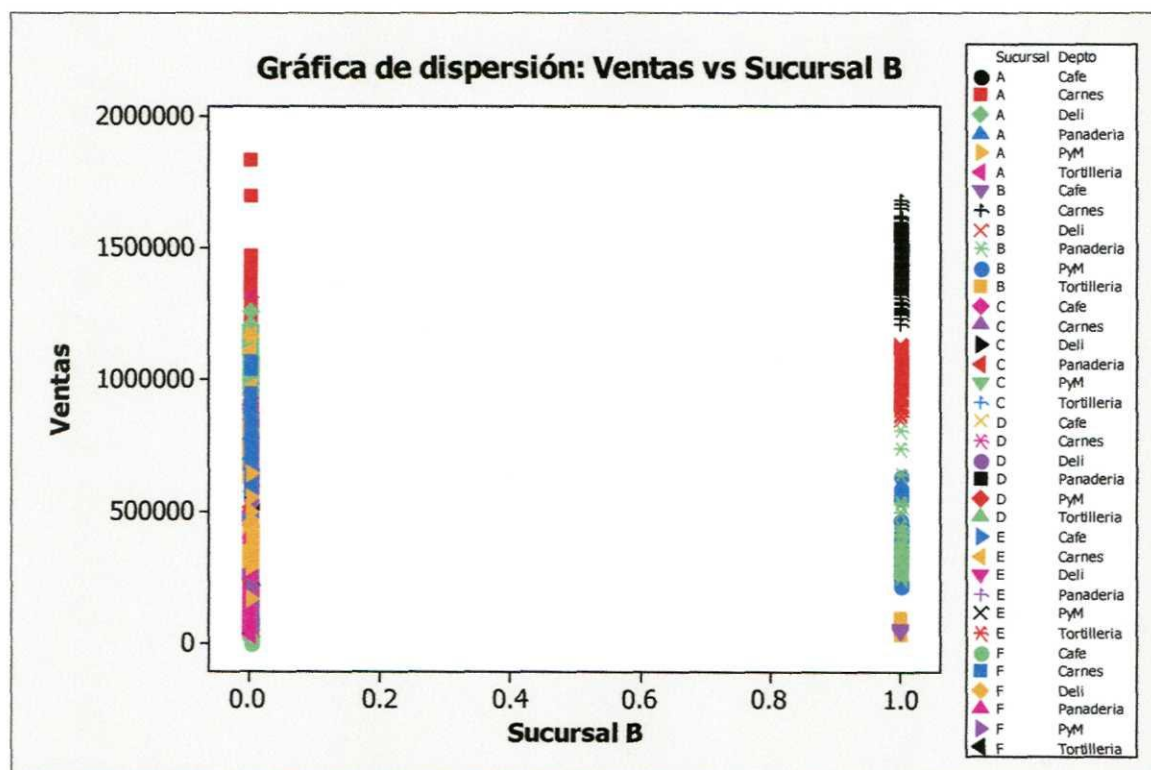
Tabla 5.- Índice General de Inflación
Base: 2ª quincena de Junio de 2002.

MES	AÑO 2007
ENERO	121.640
FEBRERO	121.980
MARZO	122.244
ABRIL	122.171

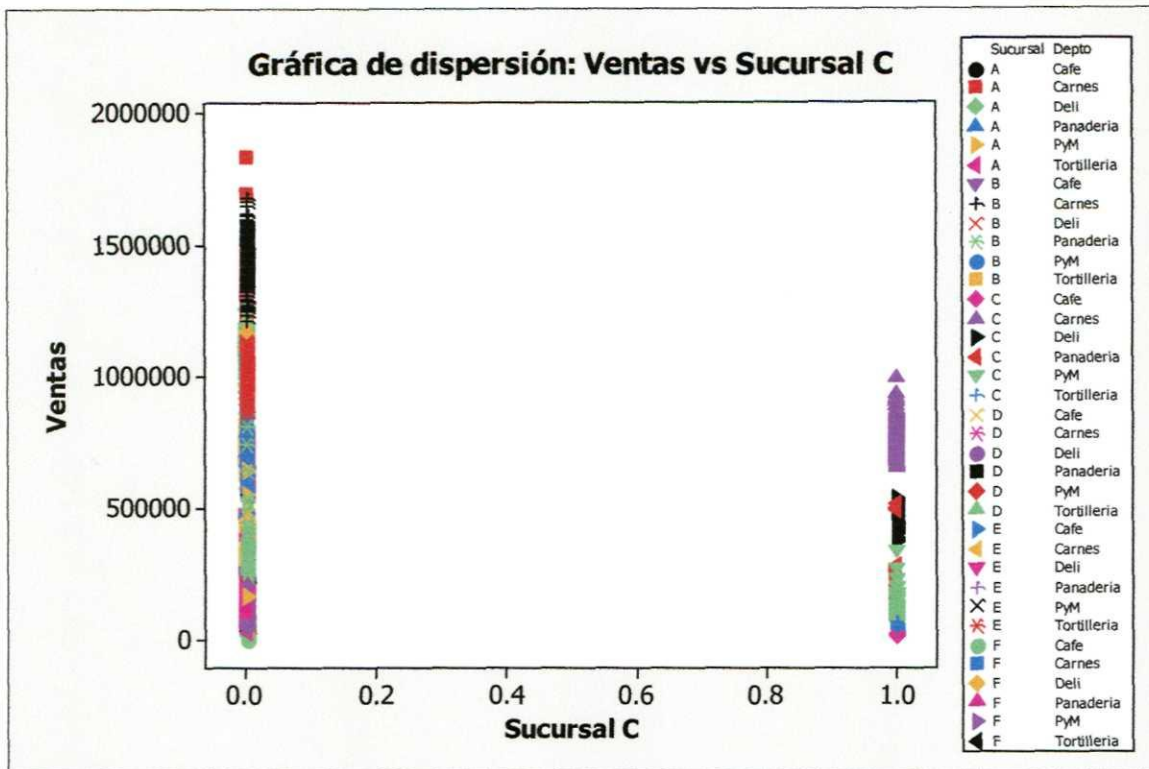
Anexo 6. Gráficas de dispersión de la etapa 1: modelo general



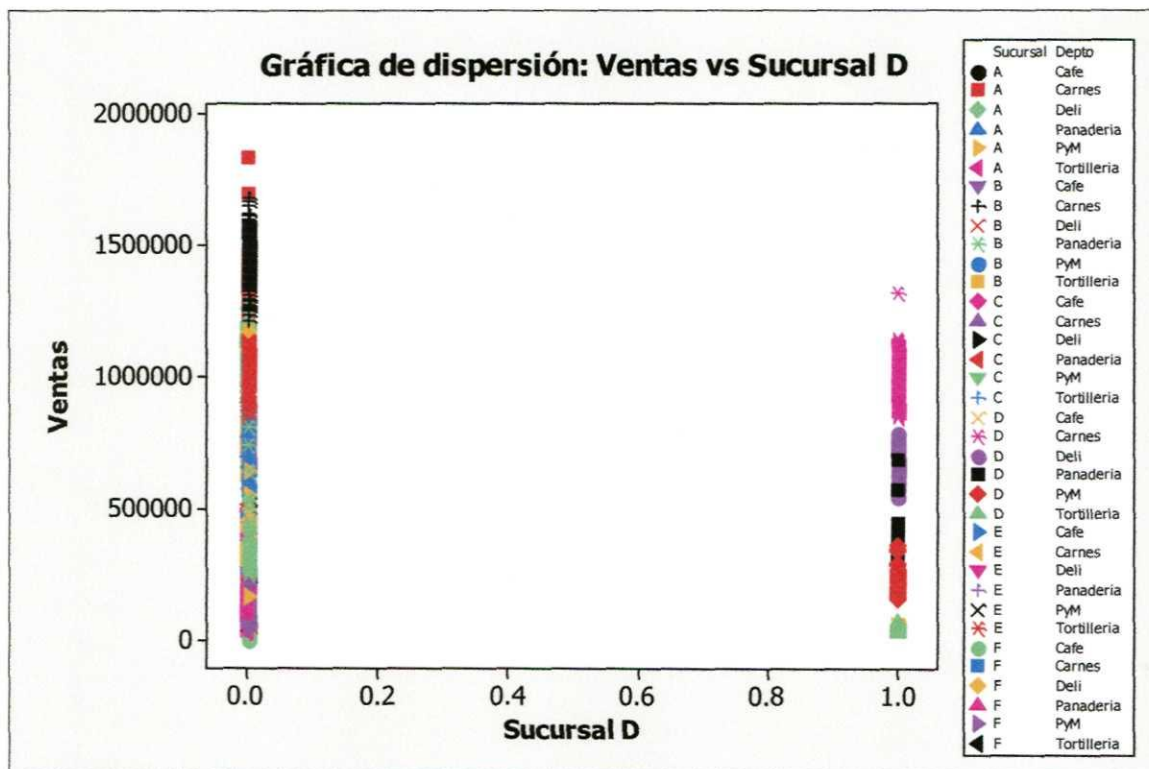
Gráfica 6.1. Ventas versus variable indicadora sucursal A (x_{11}).



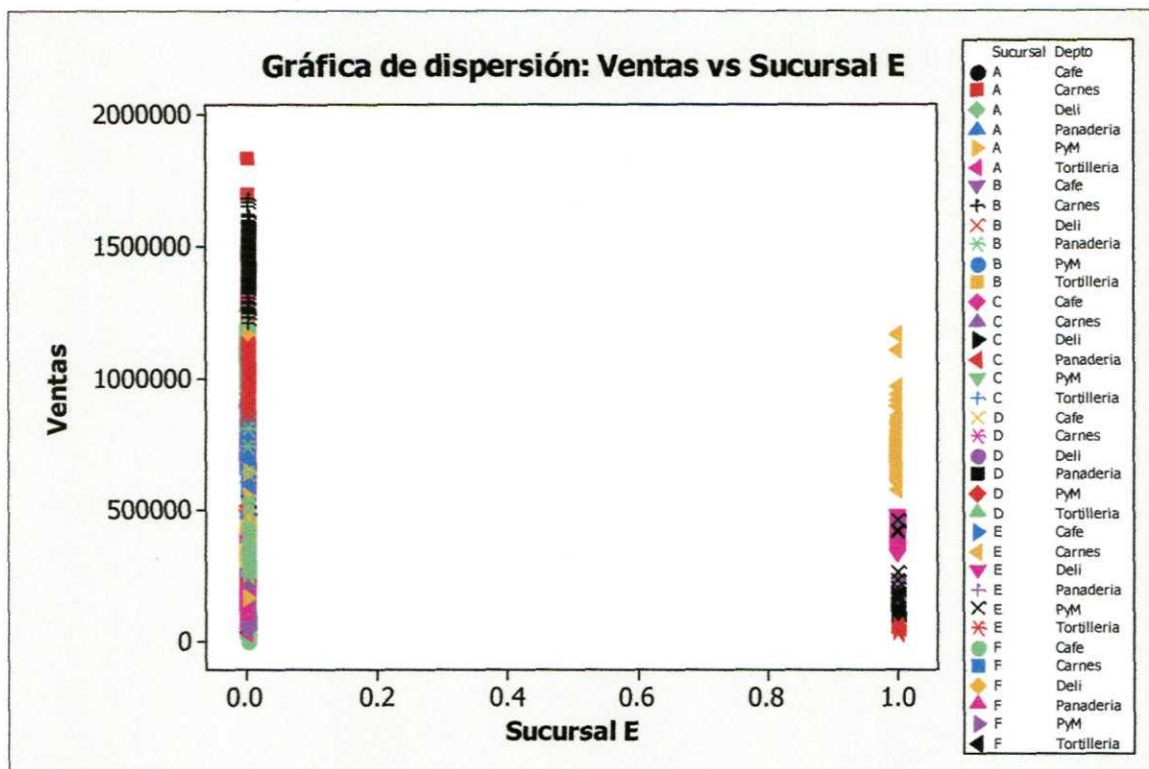
Gráfica 6.2. Ventas versus variable indicadora sucursal B (x_{12}).



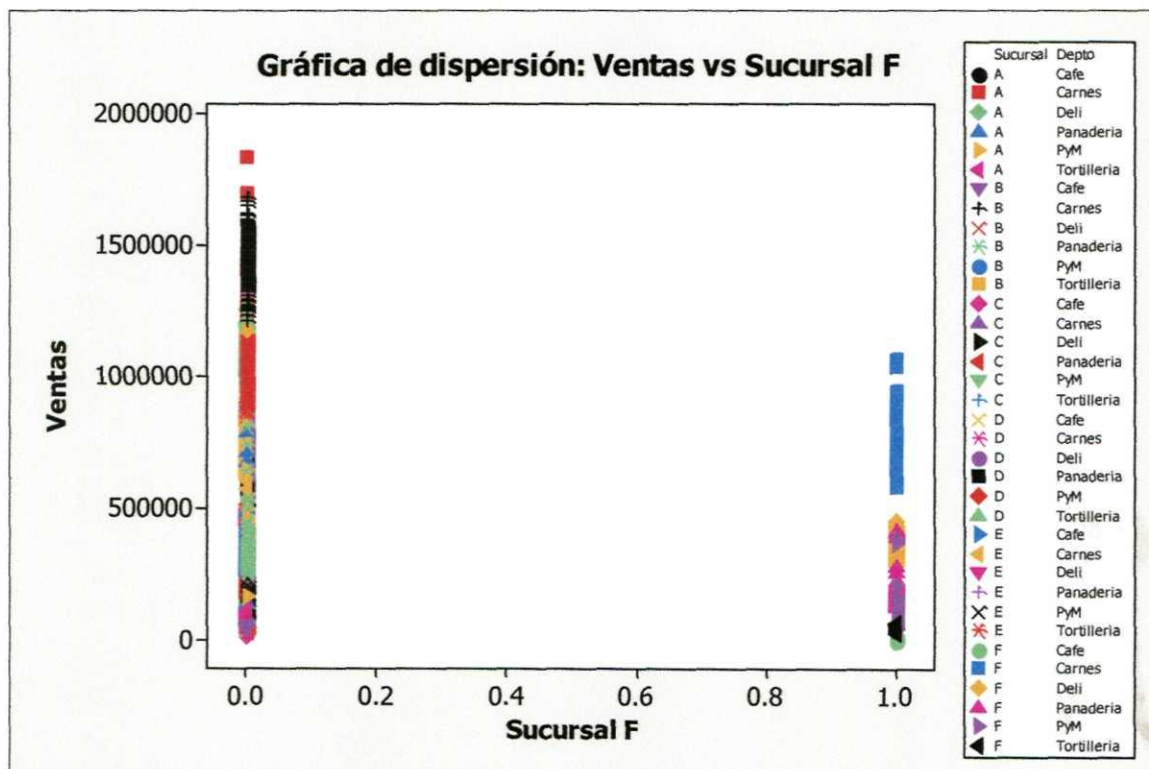
Gráfica 6.3. Ventas versus variable indicadora sucursal C (x_{13}).



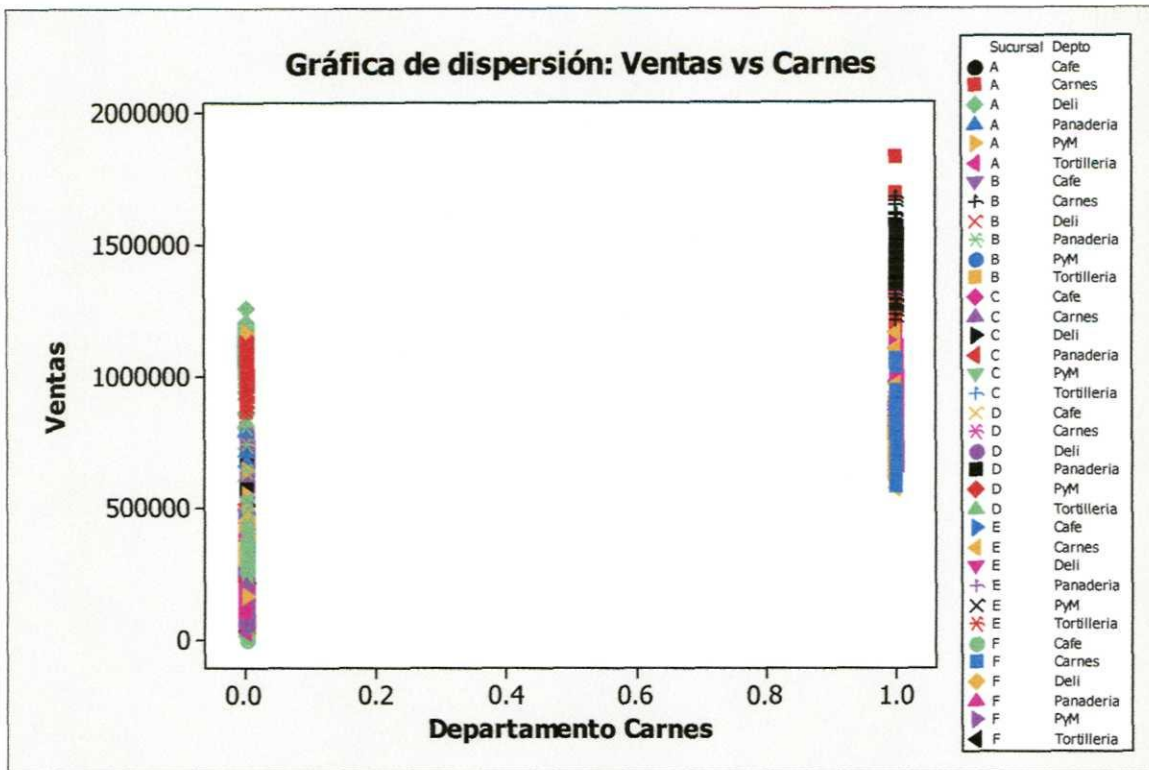
Gráfica 6.4. Ventas versus variable indicadora sucursal D (x_{14}).



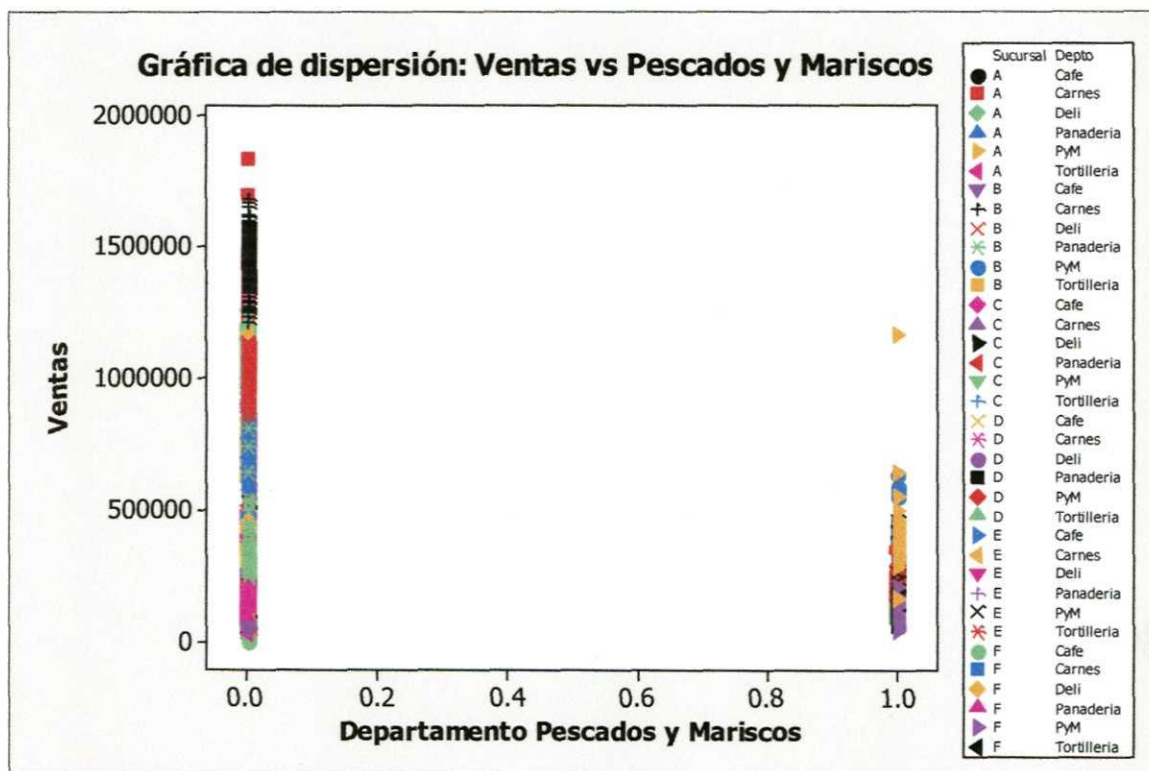
Gráfica 6.5. Ventas versus variable indicadora sucursal E (x_{15}).



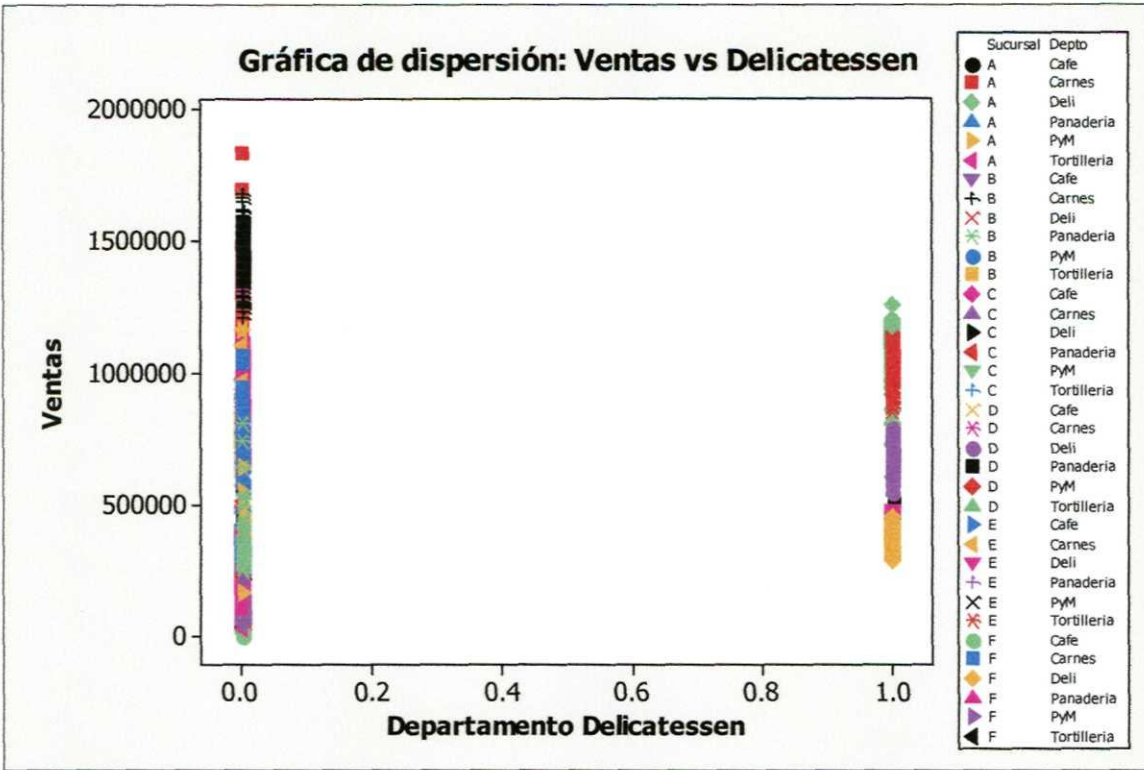
Gráfica 6.6. Ventas versus variable indicadora sucursal F (x_{16}).



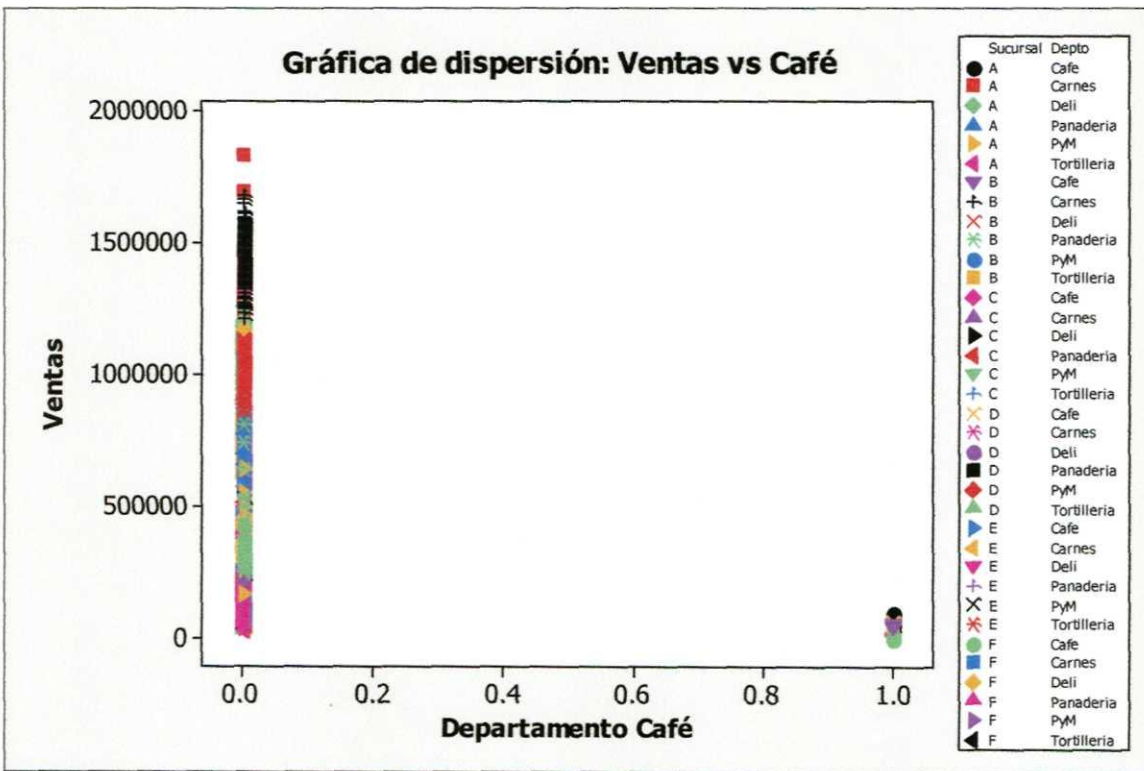
Gráfica 6.7. Ventas versus variable indicadora departamento Carnes (x_{20}).



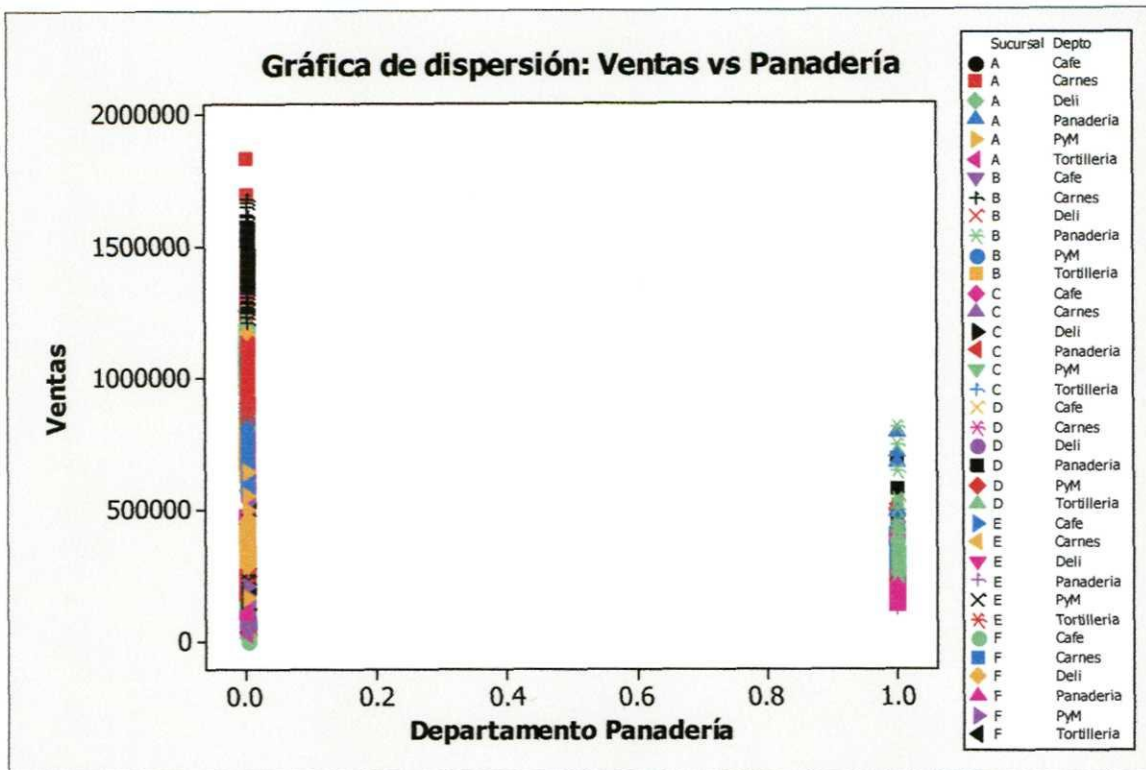
Gráfica 6.8. Ventas versus variable indicadora departamento Pescados (x_{21}).



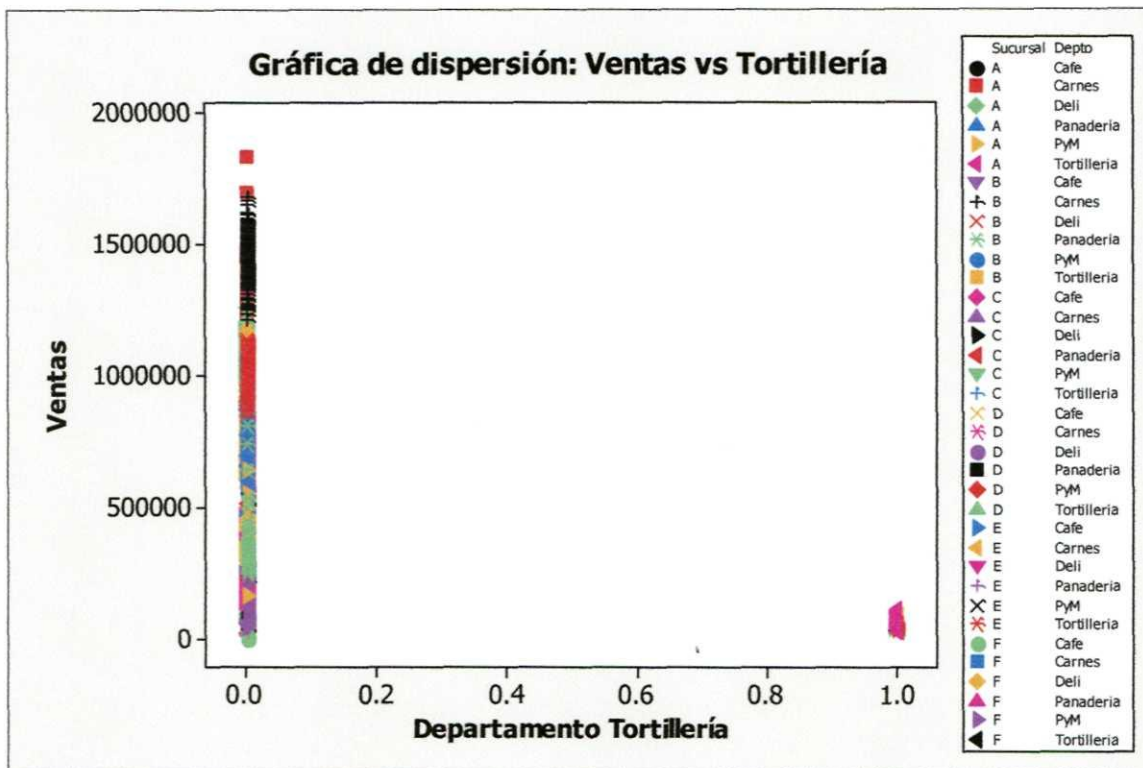
Gráfica 6.9. Ventas versus variable indicadora departamento Delicatessen (x22).



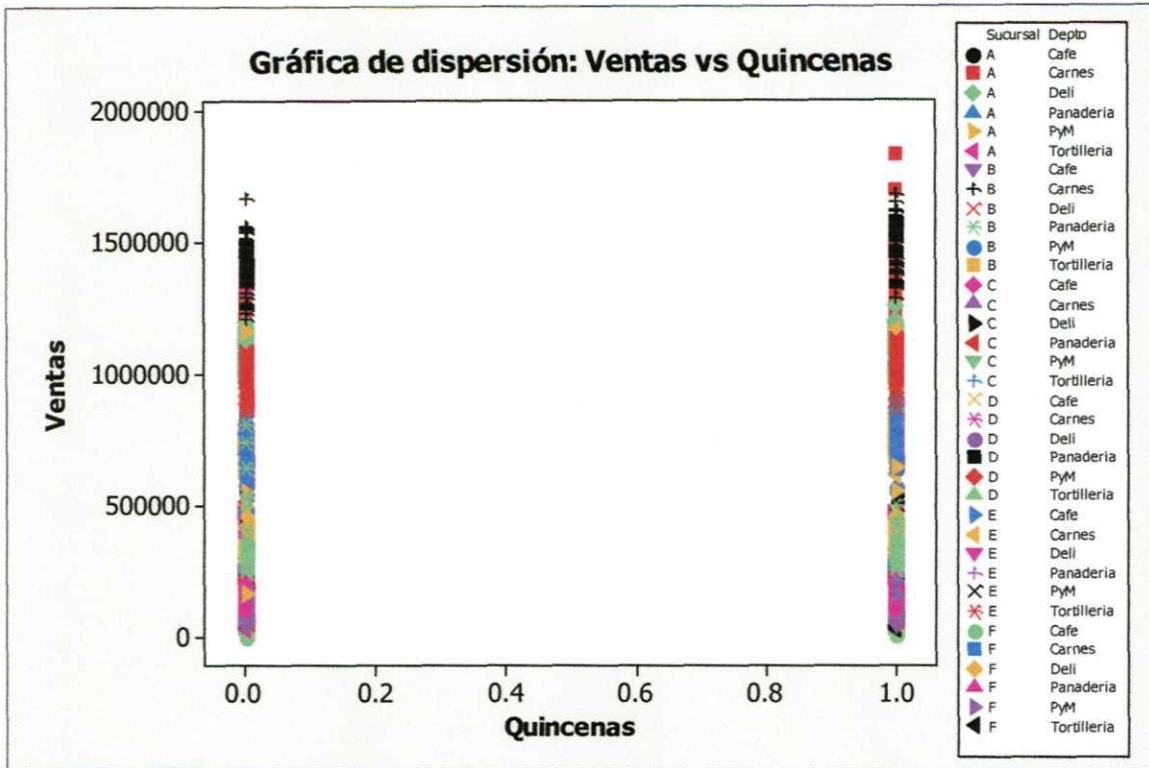
Gráfica 6.10. Ventas versus variable indicadora departamento Café (x23).



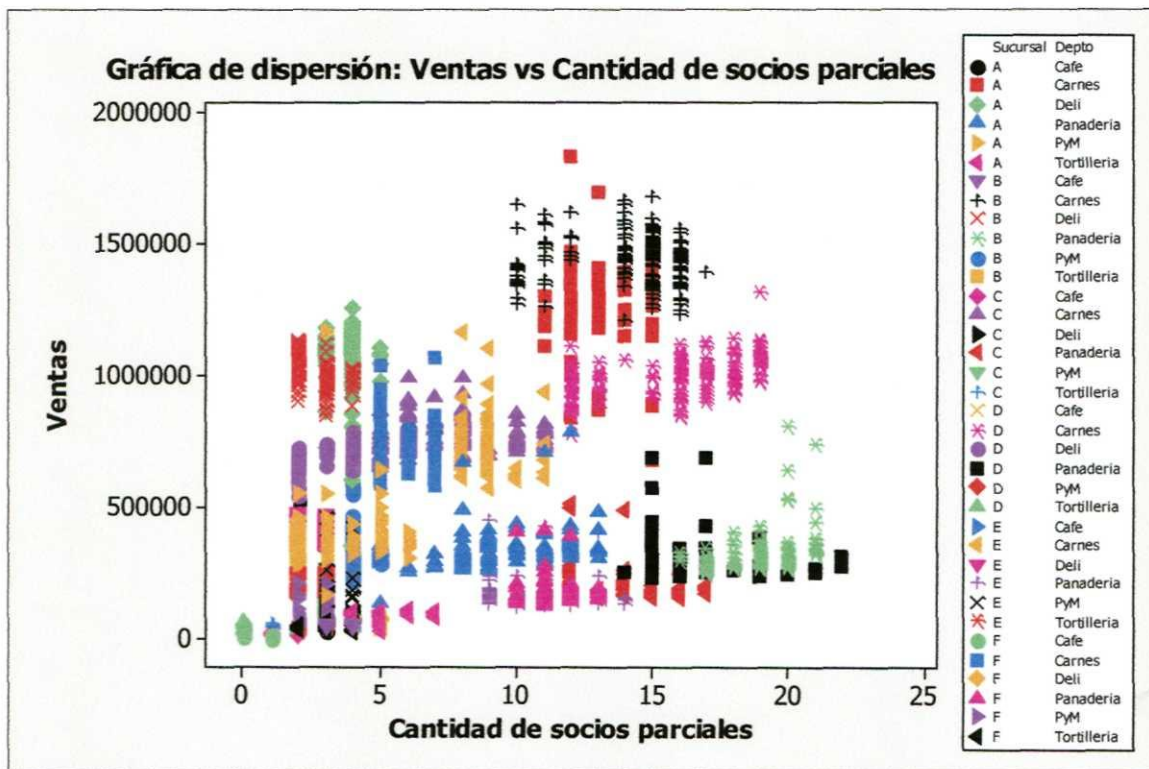
Gráfica 6.11. Ventas versus variable indicadora departamento Panadería (x_{24}).



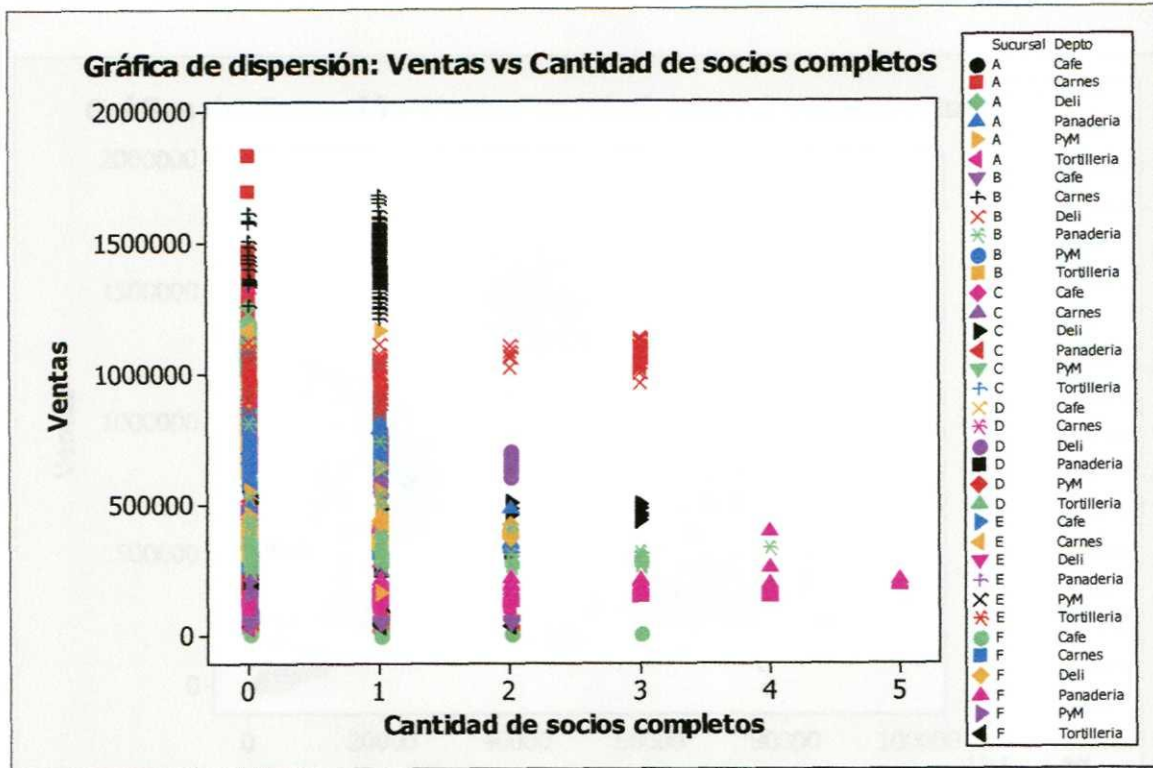
Gráfica 6.12. Ventas versus variable indicadora departamento Tortillería (x_{25}).



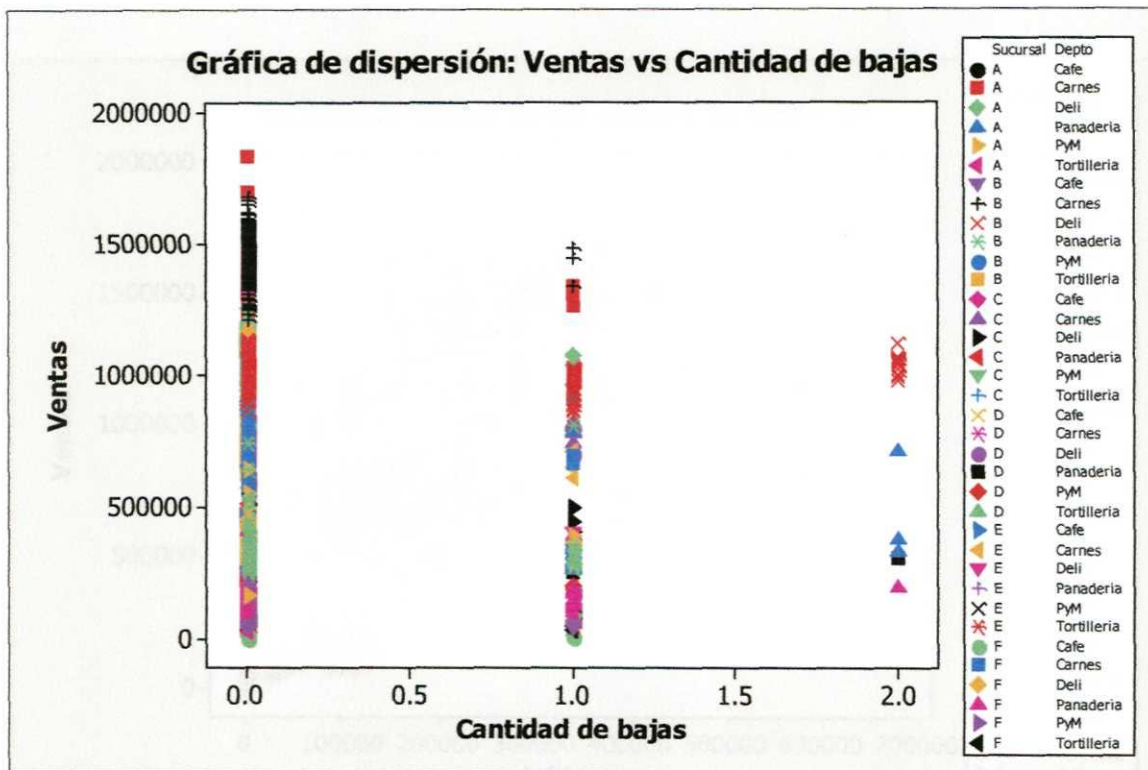
Gráfica 6.13. Ventas versus variable indicadora quincenas (x_3).



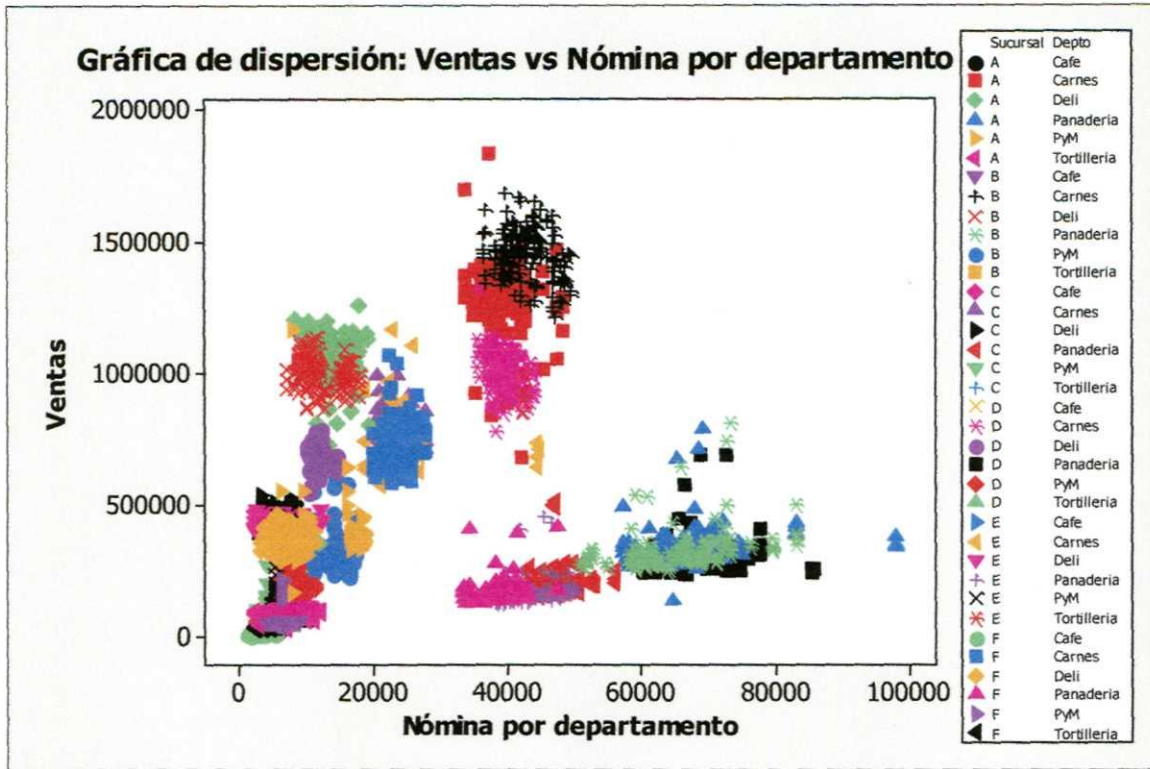
Gráfica 6.14. Ventas versus cantidad de socios parciales.



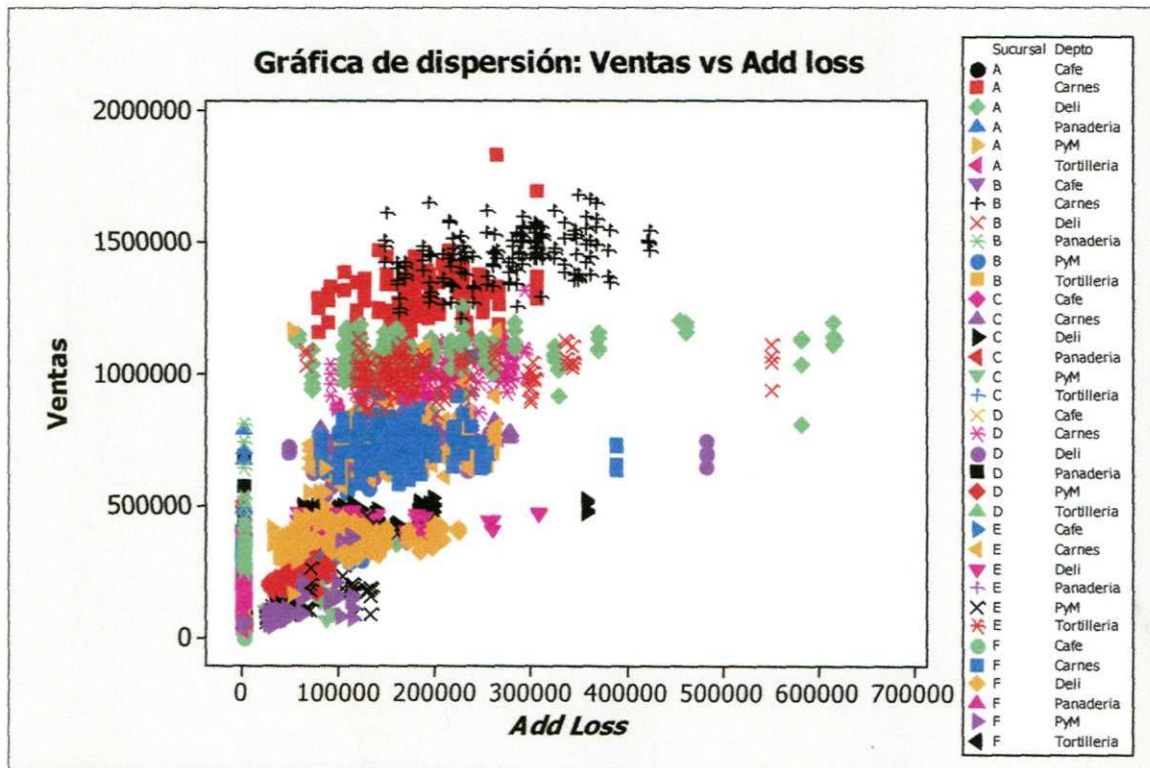
Gráfica 6.15. Ventas versus cantidad de socios completos.



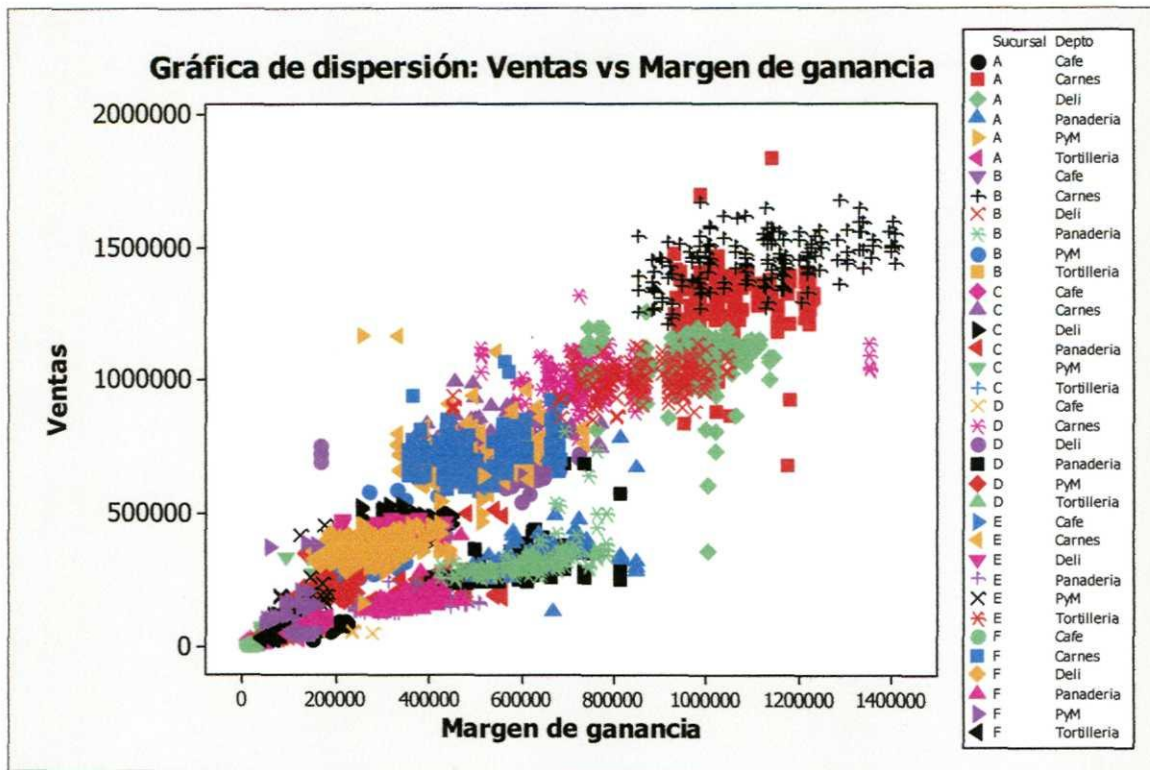
Gráfica 6.16. Ventas versus cantidad de bajas.



Gráfica 6.17. Ventas versus nómina por departamento.



Gráfica 6.18. Ventas versus add loss.



Gráfica 6.19. Ventas versus margen de ganancia.

Anexo 7. Matriz de correlaciones modelo general

Correlations	Y	X11	X12	X13	X14	X15	X20	X21	X22	X23	X24	X3	X4	X5	X6	X7	X8	X9	X10	X17	X18
Y	1	0.185873	0.183521	-0.129363	0.004512	-0.086964	0.714528	-0.202186	0.332924	-0.349823	-0.151972	0.021949	0.230382	0.390968	-0.037649	0.08524	0.352967	0.326873	0.845837	0.909712	0.544017
X11	0.185873	1	-0.20708	-0.20708	-0.20708	-0.185301	-0.0059	-0.0059	-0.005023	0.030817	-0.0059	3.01E-05	0.049798	0.026087	-0.140702	0.022476	0.139867	0.110765	0.031423	0.271798	0.032016
X12	0.183521	-0.20708	1	-0.20708	-0.20708	-0.185301	-0.0059	-0.0059	-0.005023	0.030817	-0.0059	3.01E-05	0.116623	0.161698	0.158611	0.122217	0.118875	0.105391	0.109521	0.218185	0.068773
X13	-0.129363	-0.20708	-0.20708	1	-0.20708	-0.185301	-0.0059	-0.0059	-0.005023	0.030817	-0.0059	3.01E-05	-0.083858	-0.10539	-0.142036	-0.043263	-0.103336	-0.09303	-0.06589	-0.171041	-0.043238
X14	0.004512	-0.20708	-0.20708	-0.20708	1	-0.185301	-0.0059	-0.0059	-0.005023	0.030817	-0.0059	3.01E-05	0.008659	0.11989	0.015954	-0.038729	0.074676	0.080698	-0.018928	0.02504	-0.067325
X15	-0.086964	-0.185301	-0.185301	-0.185301	-0.185301	1	0.031839	0.031839	0.027109	-0.16631	0.031839	-0.000162	-0.036018	-0.037984	-0.109122	-0.03566	-0.069036	-0.067289	-0.002869	-0.129841	0.091734
X20	0.714528	-0.0059	-0.0059	-0.0059	-0.0059	0.031839	1	-0.20708	-0.206545	-0.185857	-0.20708	3.01E-05	0.039948	0.427152	-0.116037	-0.031929	0.342702	0.283732	0.63917	0.559405	0.089274
X21	-0.202186	-0.0059	-0.0059	-0.0059	-0.0059	0.031839	-0.20708	1	-0.206545	-0.185857	-0.20708	3.01E-05	-0.358197	-0.256474	-0.205365	-0.059131	-0.308912	-0.259806	-0.078435	-0.295744	-0.013337
X22	0.332924	-0.005023	-0.005023	-0.005023	-0.005023	0.027109	-0.206545	-0.206545	1	-0.185377	-0.206545	-0.000146	-0.004526	-0.303493	0.042398	0.086463	-0.278257	-0.235575	0.406367	0.286974	0.835258
X23	-0.349823	0.030817	0.030817	0.030817	0.030817	-0.16631	-0.185857	-0.185857	-0.185377	1	-0.185857	2.7E-05	0.343181	0.267195	0.273384	-0.045746	-0.276984	-0.291025	-0.30613	-0.349644	-0.299813
X24	-0.151972	-0.0059	-0.0059	-0.0059	-0.0059	0.031839	-0.20708	-0.20708	-0.206545	-0.185857	-0.20708	3.01E-05	0.883065	0.654931	0.161278	0.085947	0.769255	0.808461	-0.341087	0.165076	-0.325284
X3	0.021949	3.01E-05	3.01E-05	3.01E-05	3.01E-05	-0.000162	3.01E-05	3.01E-05	-0.000146	2.7E-05	3.01E-05	1	0.013265	-0.000781	-0.004261	0.008527	-0.000745	-0.001002	0.004115	0.000205	0.004672
X4	0.230382	0.049798	0.116623	-0.083858	0.008659	-0.036018	0.039948	-0.358197	-0.004526	-0.343181	0.883065	0.013265	1	0.772937	0.163182	0.13061	0.85565	0.882786	-0.007295	0.513399	-0.065343
X5	0.390968	0.026087	0.161698	-0.10539	0.11989	-0.037984	0.427152	-0.256474	-0.303493	-0.267195	0.654931	-0.000781	0.772937	1	0.039955	0.041645	0.898899	0.889551	0.161827	0.57636	-0.209384
X6	-0.037649	-0.140702	0.158611	-0.142036	0.015954	-0.109122	-0.116037	-0.205365	0.042398	0.273384	0.161278	-0.004261	0.163182	0.039955	1	0.037981	0.030581	0.034809	-0.07553	0.019	-0.065942
X7	0.08524	0.022476	0.122217	-0.043263	-0.038729	-0.03566	-0.031929	-0.059131	0.086463	-0.045746	0.085947	0.008527	0.13061	0.041645	0.037981	1	0.067427	0.089319	0.026571	0.141026	0.134741
X8	0.352967	0.139867	0.118875	-0.103336	0.074676	-0.069036	0.342702	-0.308912	-0.278257	-0.276984	0.769255	-0.000745	0.85565	0.898899	0.030581	0.067427	1	0.982171	0.077095	0.590538	-0.246837
X9	0.326873	0.110765	0.105391	-0.09303	0.080698	-0.067289	0.283732	-0.259806	-0.235575	-0.291025	0.808461	-0.001002	0.882786	0.889551	0.034809	0.089319	0.982171	1	0.059075	0.576437	-0.202486
X10	0.845837	0.031423	0.109521	-0.06589	-0.018928	-0.002869	0.63917	-0.078435	0.406367	-0.30613	-0.341087	0.004115	-0.007295	0.161827	-0.07553	0.026571	0.077095	0.059075	1	0.656949	0.610513
X17	0.909712	0.271798	0.218185	-0.171041	0.02504	-0.129841	0.559405	-0.295744	0.286974	-0.349644	0.165076	0.000205	0.513399	0.57636	0.019	0.141026	0.590538	0.576437	0.656949	1	0.42627
X18	0.544017	0.032016	0.068773	-0.043238	-0.067325	0.091734	0.089274	-0.013337	0.835258	-0.299813	-0.325284	0.004672	-0.065343	-0.209384	-0.065942	0.134741	-0.246837	-0.202486	0.610513	0.42627	1

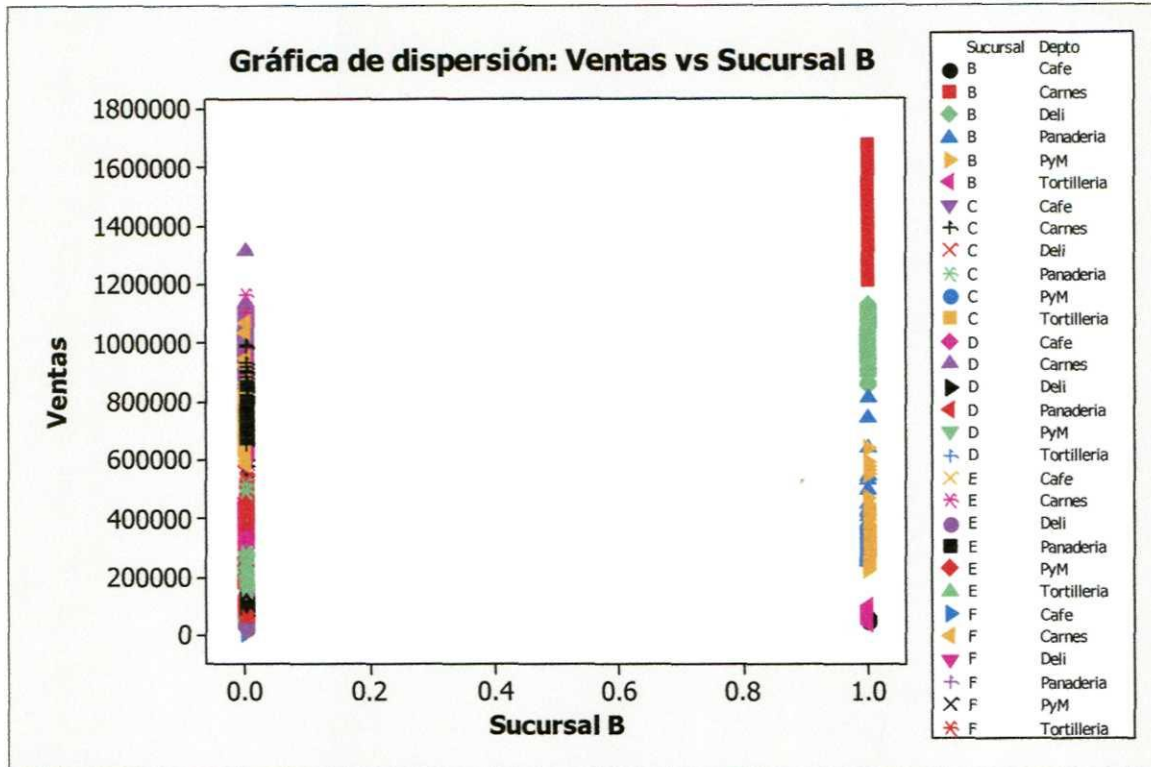
Paquete estadístico de cómputo SPSS

Anexo 8. Matriz de covarianza modelo general

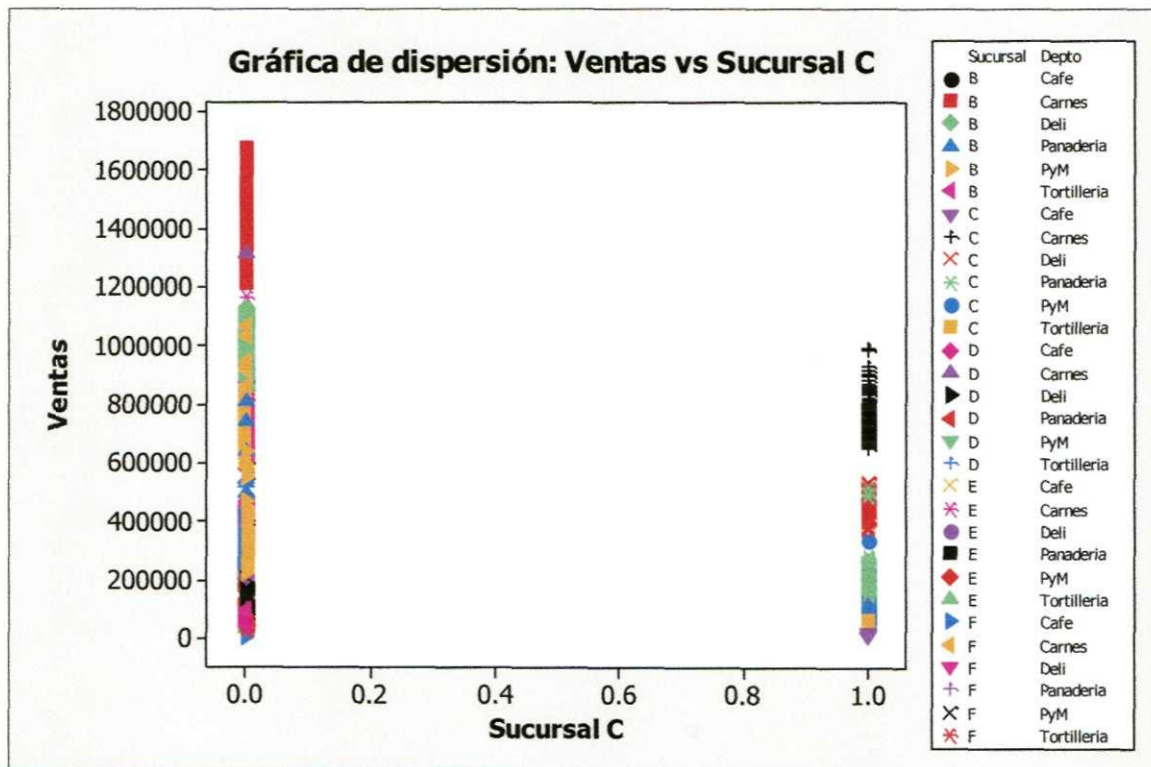
Covariances	X11	X12	X13	X14	X15	X20	X21	X22	X23	X24	X3	X4	X5	X6	X7	X8	X9	X10	X17	X18	
X11	0																				
X12	0	0																			
X13	0	0	0																		
X14	0	0	0	0																	
X15	0	0	0	0	0																
X20	0	0	0	0	0	0															
X21	0	0	0	0	0	0	0														
X22	0	0	0	0	0	0	0	0													
X23	0	0	0	0	0	0	0	0	0												
X24	0	0	0	0	0	0	0	0	0	0											
X3	0	0	0	0	0	0	0	0	0	0	0										
X4	407	954	-686	71	-273	327	-2930	-37	-2607	7224	143	470730521									
X5	0	0	0	0	0	0	-1	0	0	1	0	84839	26								
X6	0	0	0	0	0	0	0	0	0	0	0	2582	0	1							
X7	0	0	0	0	0	0	0	0	0	0	0	608	0	0	0						
X8	18	16	-14	10	-8	45	-41	-37	-34	102	0	6500959	1592	8	5	122628					
X9	828	788	-695	603	-466	2121	-1942	-1758	-2020	6043	-10	379731398	89222	503	380	6818946	3.93E+08				
X10	1069	3727	-2242	-644	-90	21751	-2669	13805	-9673	-11607	185	-14285130	73895	-4972	514	2436794	1.05E+08	8.15E+09			
X17	31709	25455	-19955	2921	-14035	65263	-344503	33423	-37874	19259	32	3.447E+09	902257	4288	9360	63990074	3.54E+09	1.83E+10	9.58E+10		
X18	4	9	-5	-8	11	11	-2	105	-35	-41	1	-473689	-354	-16	10	-28881	-1341340	18411947	44072090	111640	

Paquete estadístico de cómputo Minitab

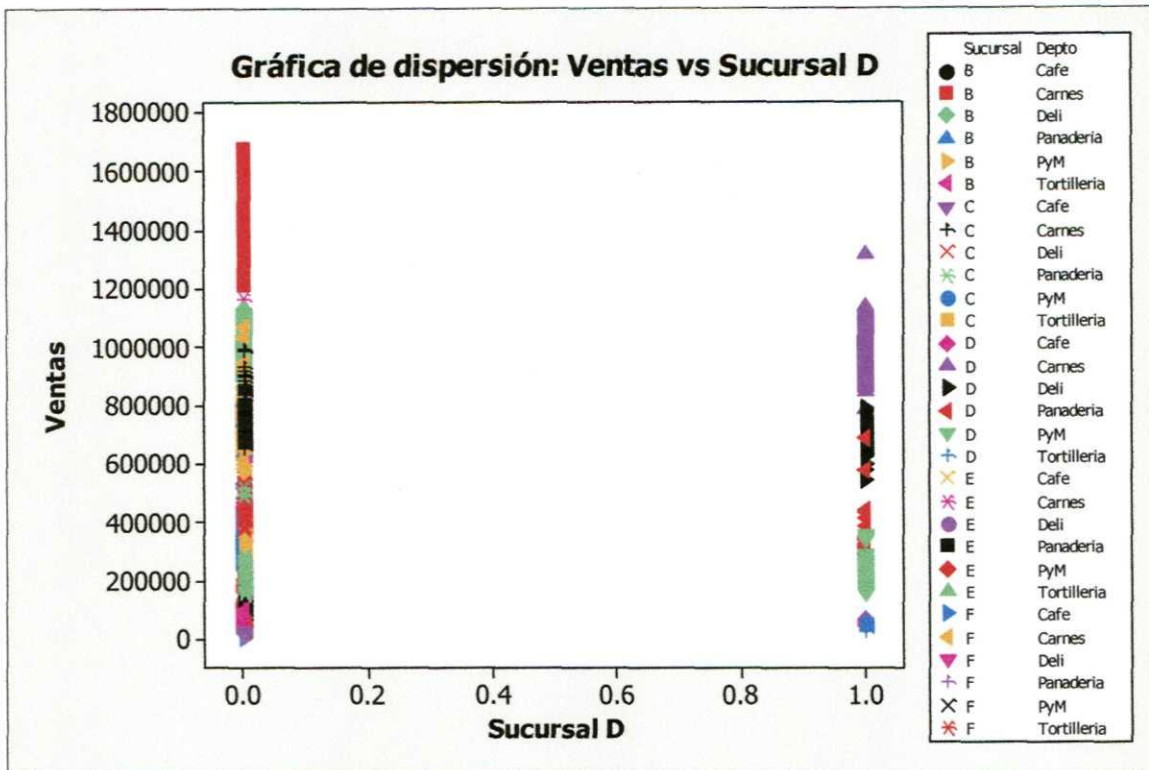
Anexo 9. Gráficas de dispersión de la etapa 2: segmentación



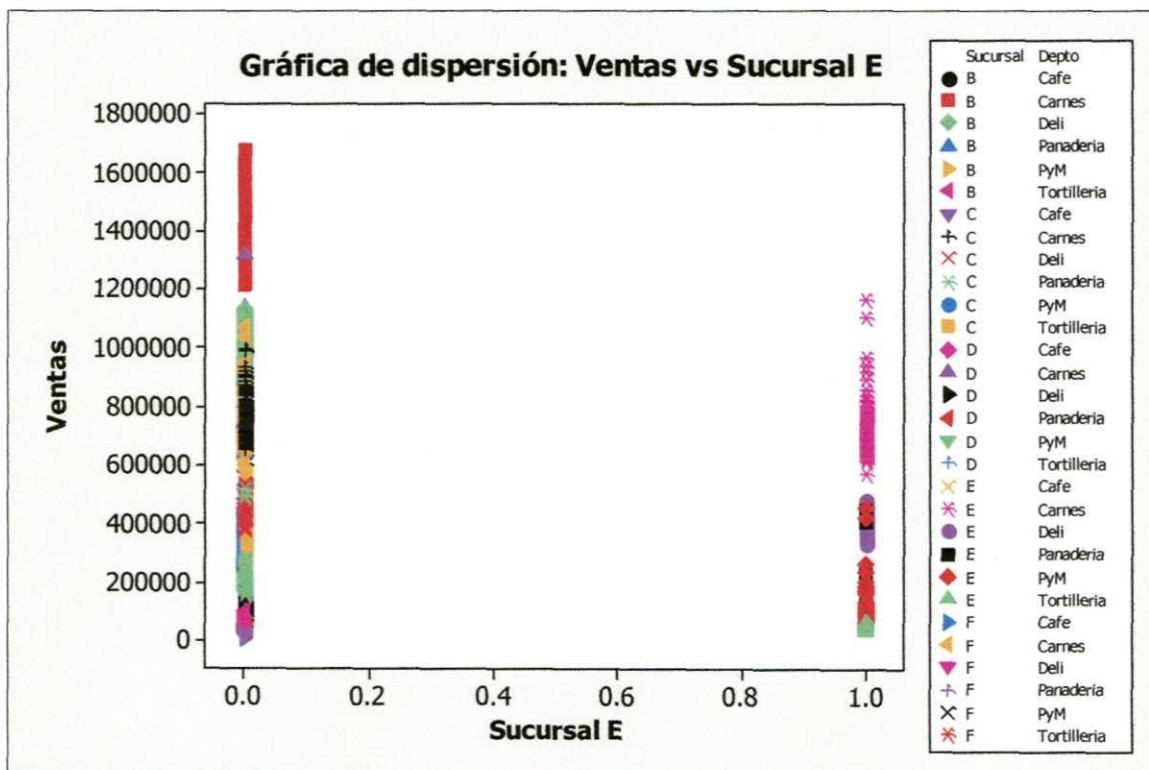
Gráfica 9.1. Ventas versus variable indicadora sucursal B (x_{12}).



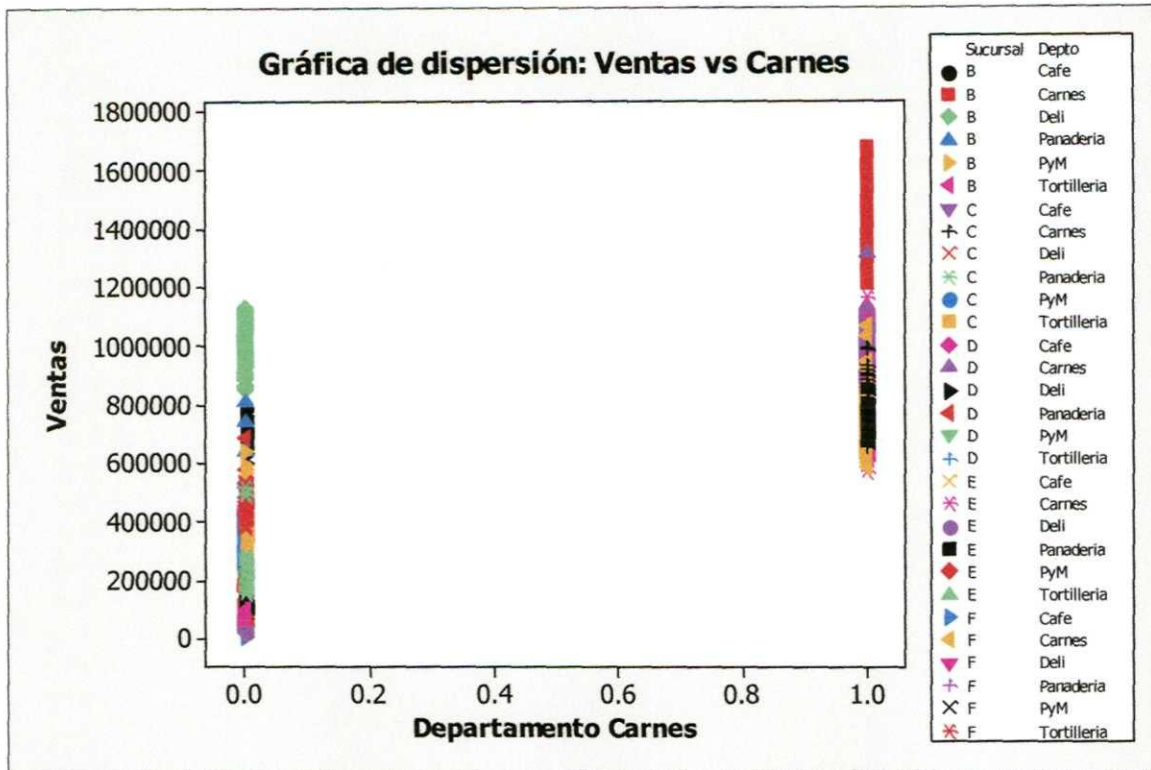
Gráfica 9.2. Ventas versus variable indicadora sucursal C (x_{13}).



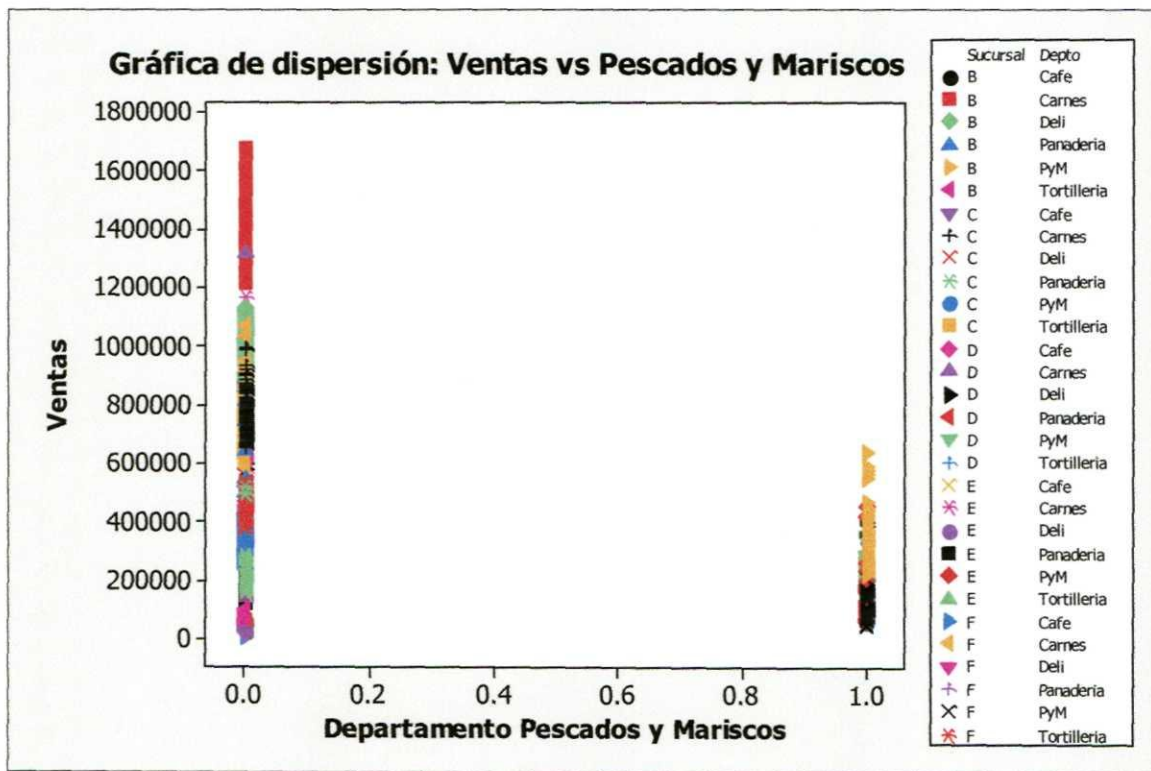
Gráfica 9.3. Ventas versus variable indicadora sucursal D (x_{14}).



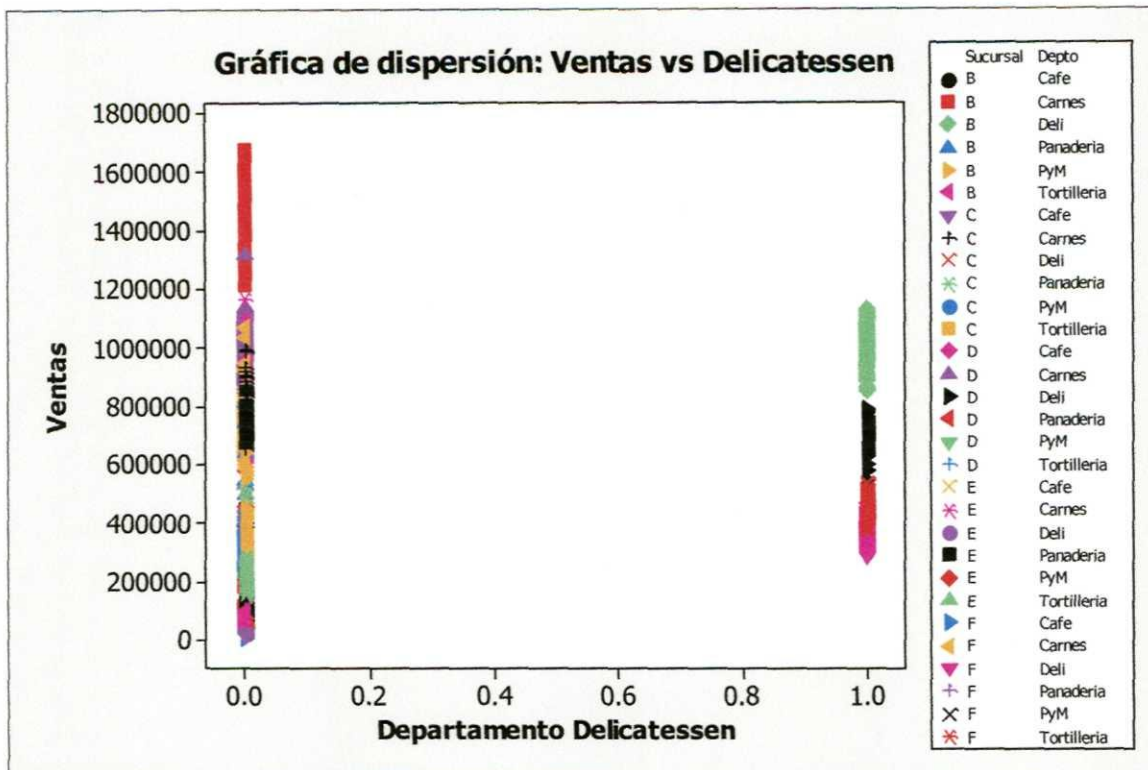
Gráfica 9.4. Ventas versus variable indicadora sucursal E (x_{15}).



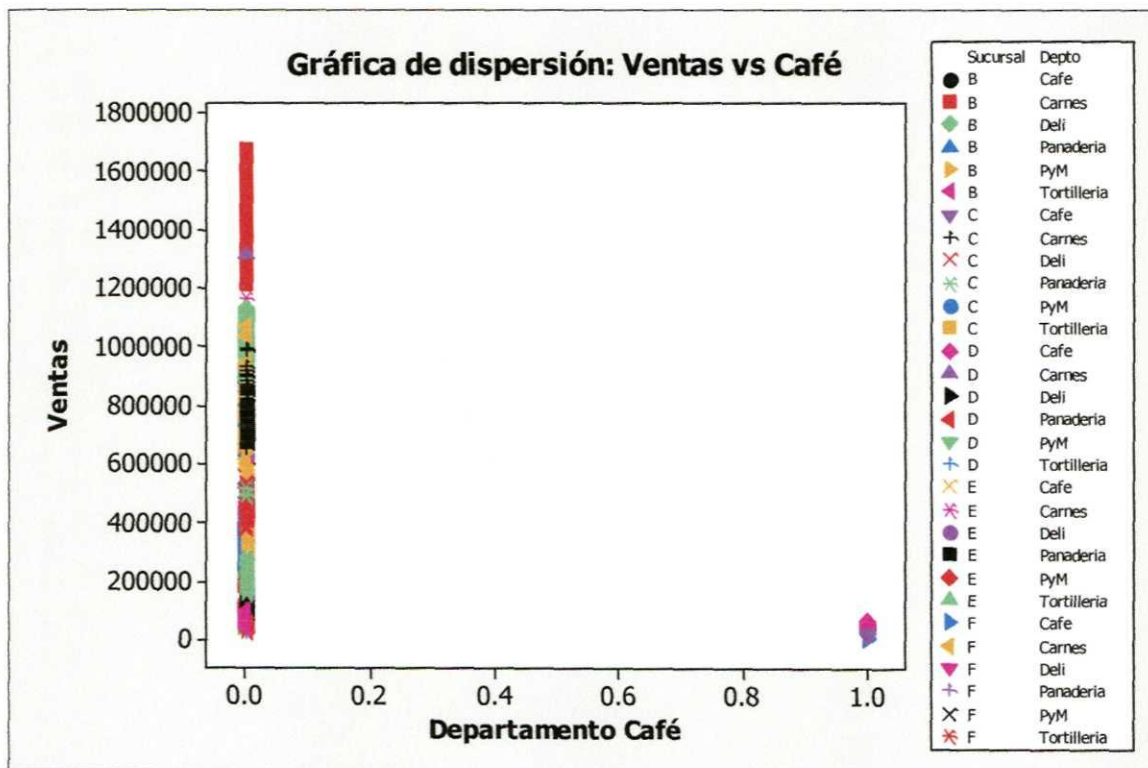
Gráfica 9.5. Ventas versus variable indicadora departamento Carnes (x_{20}).



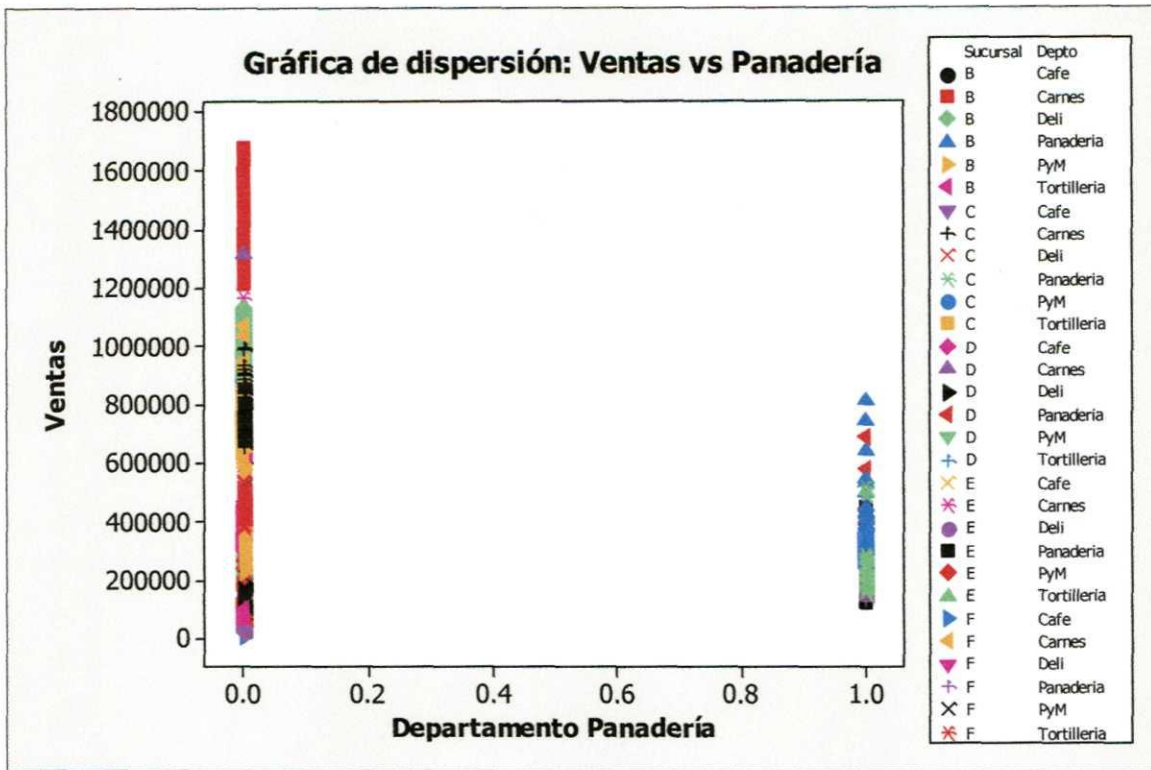
Gráfica 9.6. Ventas versus variable indicadora departamento Pescados (x_{21}).



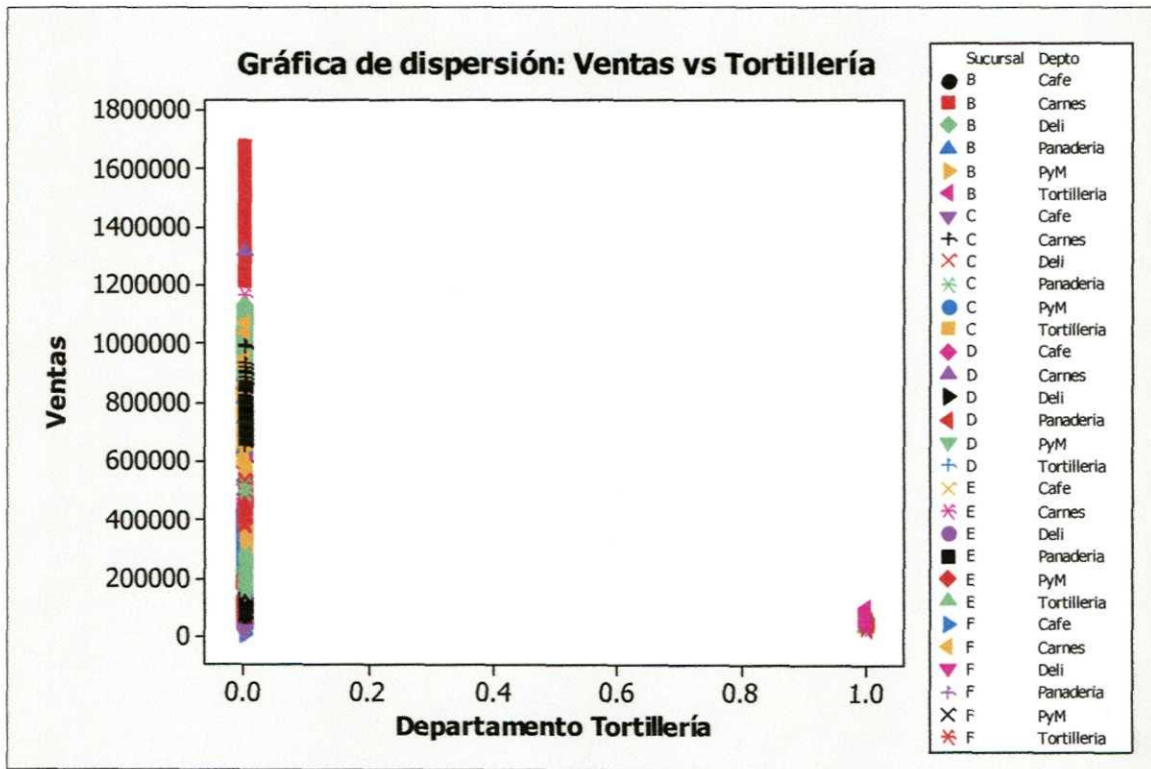
Gráfica 9.7. Ventas versus variable indicadora departamento Delicatessen (x_{22}).



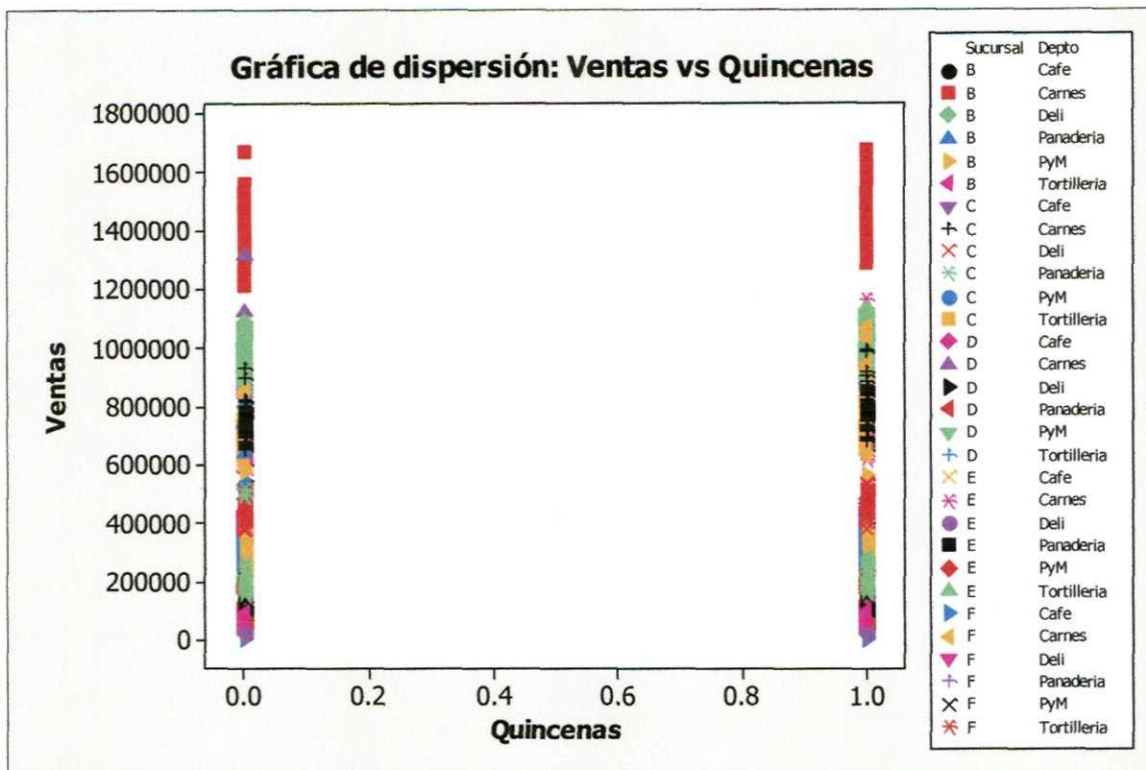
Gráfica 9.8. Ventas versus variable indicadora departamento Café (x_{23}).



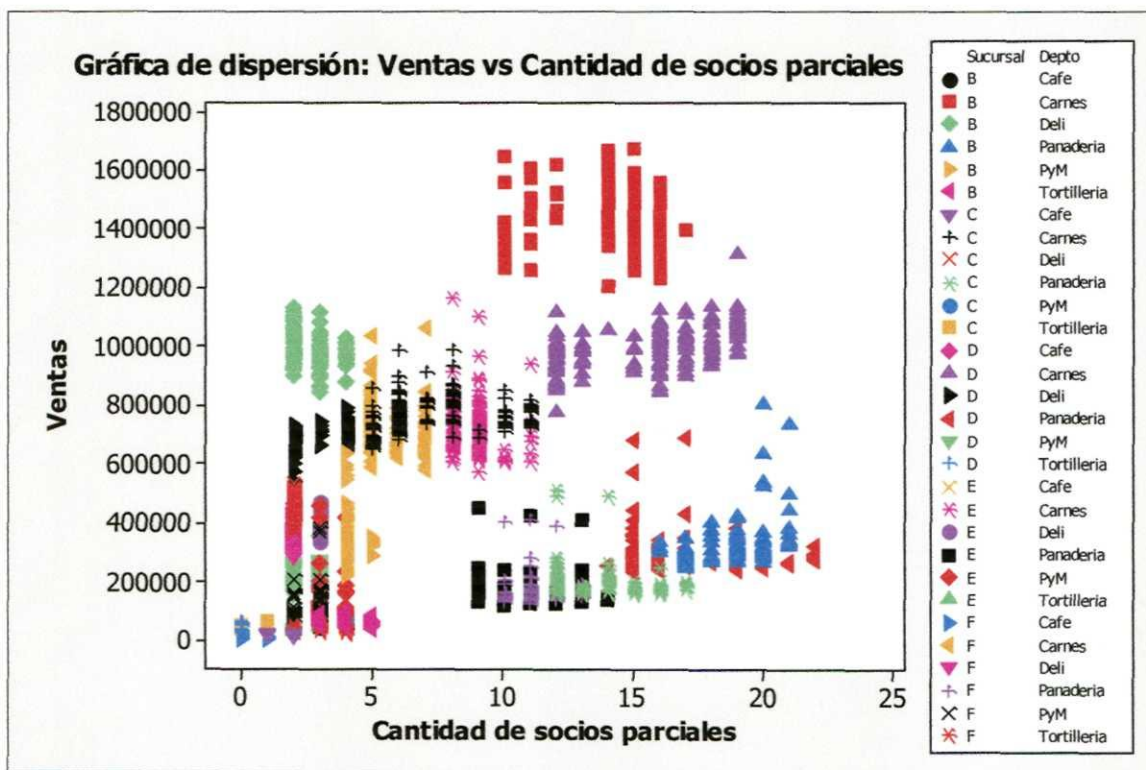
Gráfica 9.9. Ventas versus variable indicadora departamento Panadería (x_{24}).



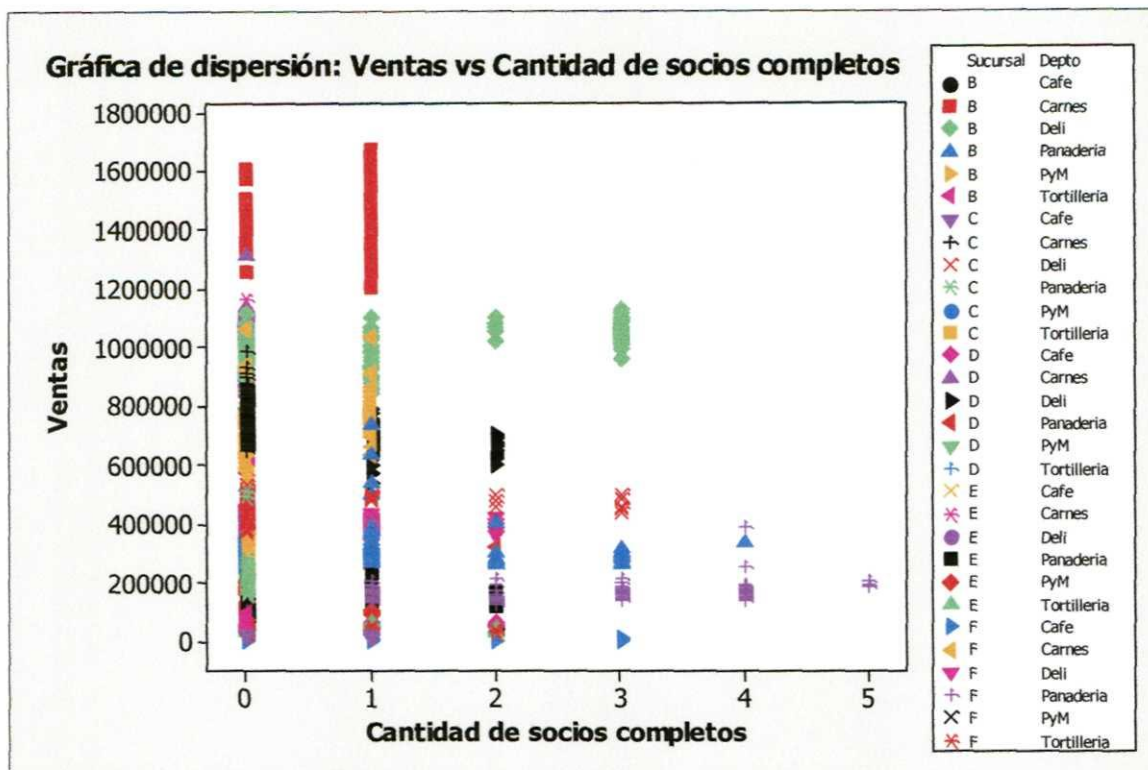
Gráfica 9.10. Ventas versus variable indicadora departamento Tortillería (x_{25}).



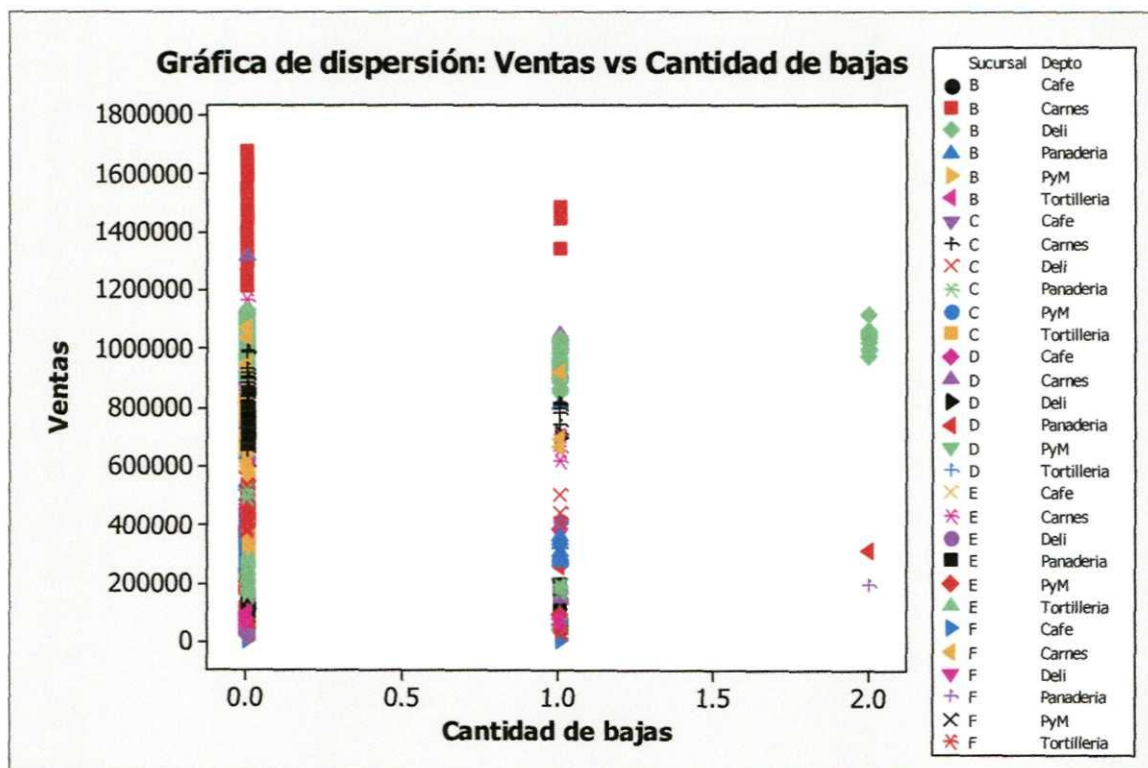
Gráfica 9.11. Ventas versus quincenas.



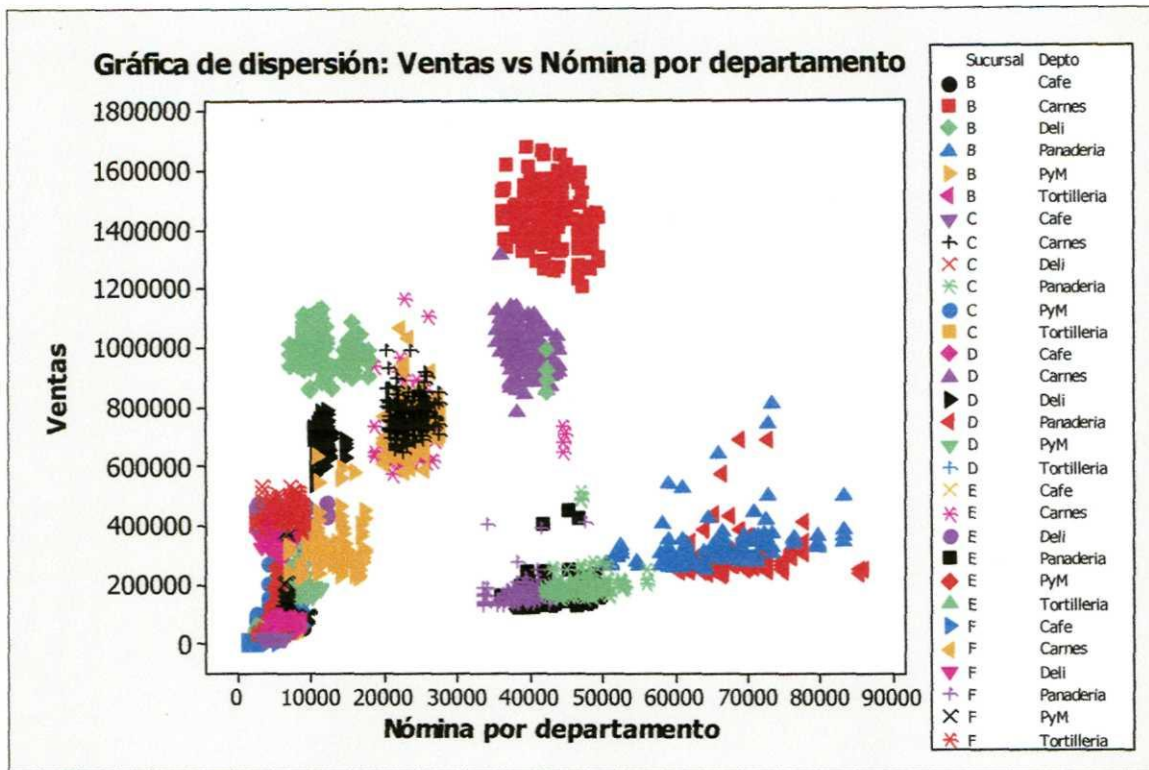
Gráfica 9.12. Ventas versus cantidad de socios parciales.



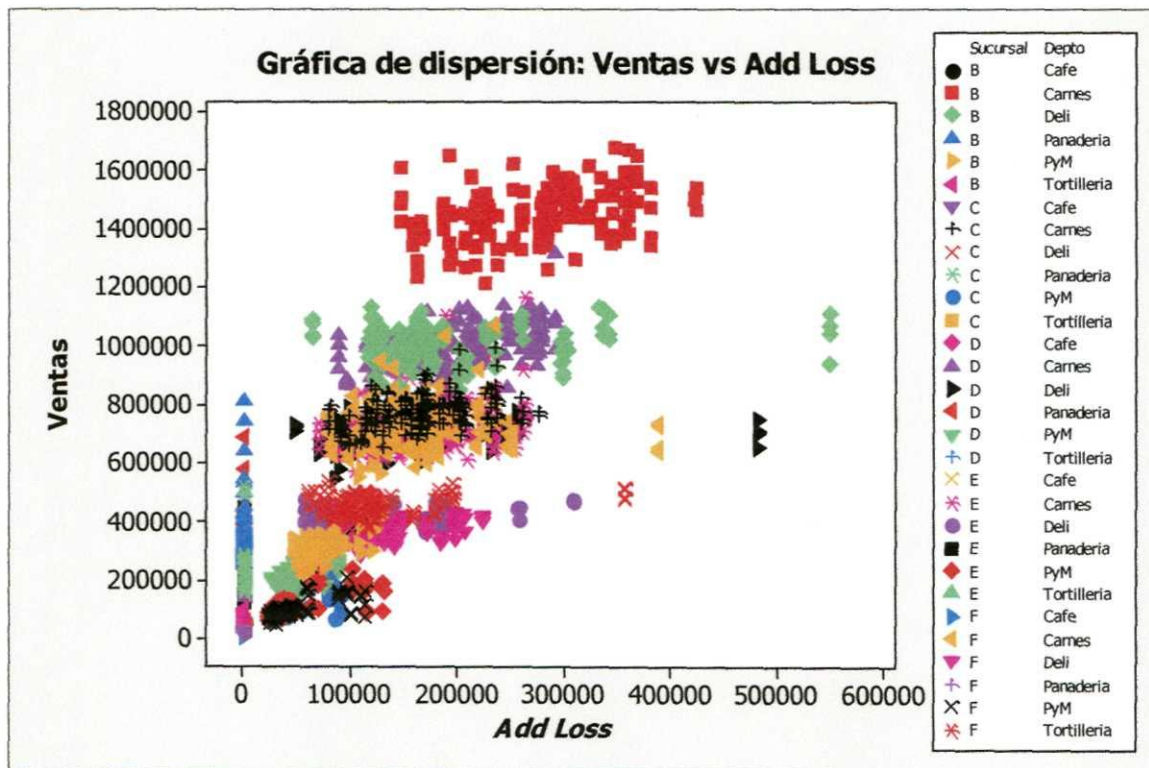
Gráfica 9.13. Ventas versus cantidad de socios completos.



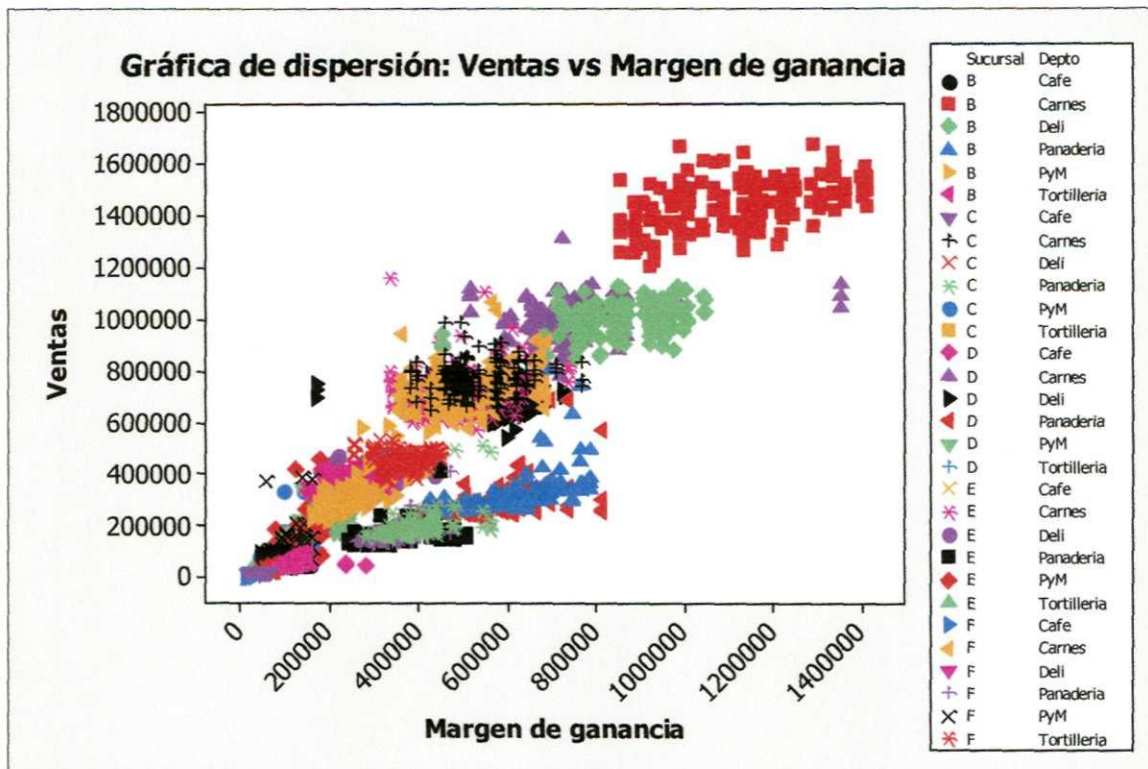
Gráfica 9.14. Ventas versus cantidad de bajas.



Gráfica 9.15. Ventas versus nómina por departamento.



Gráfica 9.16. Ventas versus add loss.



Gráfica 9.17. Ventas versus margen de ganancia.

Anexo 10. Método Stepwise etapa 3: A

Stepwise Regression: Y versus X20, X21, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is Y on 15 predictors, with N = 936

Step	1	2	3	4	5	6
Constant	-115891	-75127	-58555	-76892	-119434	-81812
X17	1.190	1.224	1.051	0.923	0.758	0.710
T-Value	76.23	116.91	72.93	56.20	35.52	33.62
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X24		-355534	-263531	-515460	-631830	-315782
T-Value		-34.11	-24.04	-23.80	-27.68	-7.70
P-Value		0.000	0.000	0.000	0.000	0.000
X10			0.855	0.896	0.718	0.669
T-Value			15.73	17.91	14.48	14.00
P-Value			0.000	0.000	0.000	0.000
X9				5.19	8.85	3.98
T-Value				13.14	17.93	5.57
P-Value				0.000	0.000	0.000
X18					236	314
T-Value					11.23	14.35
P-Value					0.000	0.000
X20						203312
T-Value						9.11
P-Value						0.000
S	177265	118323	105234	96706	90796	87042
R-Sq	86.15	93.84	95.13	95.89	96.38	96.68
R-Sq(adj)	86.14	93.82	95.11	95.87	96.36	96.66
Mallows C-p	4405.7	1445.6	949.1	657.5	470.4	358.1

Step	7	8	9	10	11	12
Constant	-38123	16024	11445	18838	21236	5426
X17	0.414	0.277	0.284	0.181	0.190	0.155
T-Value	10.24	7.04	7.56	4.75	5.07	3.95
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X24	-232495	28990				
T-Value	-5.71	0.66				
P-Value	0.000	0.508				
X10	0.470	0.344	0.349	0.254	0.261	0.234
T-Value	9.08	6.97	7.15	5.28	5.45	4.82
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X9	4.80	1.67	2.08	-0.81	-1.03	-1.38
T-Value	6.91	2.39	6.62	-1.81	-2.44	-3.18
P-Value	0.000	0.017	0.000	0.071	0.015	0.002

X18	367	91	96	41		
T-Value	16.69	2.93	3.24	1.41		
P-Value	0.000	0.003	0.001	0.159		
X20	457068	807955	784961	940980	949327	997754
T-Value	12.41	17.88	27.20	28.53	29.24	27.77
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X22	241350	614673	600674	680091	700140	732531
T-Value	8.49	15.00	17.12	19.47	21.93	21.88
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X21		193669	188304	252670	268546	295065
T-Value		11.94	13.41	16.43	25.63	21.82
P-Value		0.000	0.000	0.000	0.000	0.000
X4				4.14	4.28	5.28
T-Value				8.75	9.27	9.39
P-Value				0.000	0.000	0.000
X23						31671
T-Value						3.08
P-Value						0.002
S	83894	78145	78121	75121	75161	74820
R-Sq	96.92	97.33	97.33	97.53	97.53	97.55
R-Sq(adj)	96.90	97.31	97.31	97.51	97.51	97.53
Mallows C-p	267.9	111.5	110.0	33.4	33.5	25.8
Step	13	14	15	16		
Constant	-18697	-17910	-18381	-21354		
X17	0.176	0.176	0.172	0.173		
T-Value	4.46	4.48	4.38	4.39		
P-Value	0.000	0.000	0.000	0.000		
X24	-153382	-146032	-147890	-146857		
T-Value	-3.38	-4.65	-4.71	-4.68		
P-Value	0.001	0.000	0.000	0.000		
X10	0.242	0.241	0.241	0.241		
T-Value	5.02	5.02	5.02	5.02		
P-Value	0.000	0.000	0.000	0.000		
X9	0.14					
T-Value	0.22					
P-Value	0.822					
X18						
T-Value						
P-Value						
X20	914571	918723	922135	922004		
T-Value	21.08	23.42	23.50	23.51		
P-Value	0.000	0.000	0.000	0.000		
X22	687015	687501	690576	690574		
T-Value	19.13	19.19	19.27	19.28		
P-Value	0.000	0.000	0.000	0.000		
X21	289675	290270	290983	290756		
T-Value	21.39	21.87	21.93	21.93		

P-Value	0.000	0.000	0.000	0.000
X4	6.21	6.23	6.26	6.24
T-Value	9.96	10.08	10.13	10.09
P-Value	0.000	0.000	0.000	0.000
X23	36410	36605	37399	37164
T-Value	3.53	3.56	3.64	3.62
P-Value	0.000	0.000	0.000	0.000
X7			17977	17363
T-Value			1.69	1.63
P-Value			0.091	0.103
X3				7095
T-Value				1.45
P-Value				0.146
S	74402	74363	74289	74245
R-Sq	97.58	97.58	97.59	97.59
R-Sq(adj)	97.56	97.56	97.57	97.57
Mallows C-p	16.3	14.3	13.4	13.3

Anexo 11. Método Stepwise etapa 3: B

Stepwise Regression: Y versus X20, X21, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is Y on 15 predictors, with N = 936

Step	1	2	3	4	5	6
Constant	-98654	-53228	-37694	-71681	-154174	-78973
X17	1.2372	1.2795	0.9964	0.9178	0.6328	0.4520
T-Value	78.75	133.87	60.54	59.31	29.18	19.20
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X24		-402812	-245987	-451688	-670027	-303277
T-Value		-40.26	-21.25	-27.10	-34.31	-9.60
P-Value		0.000	0.000	0.000	0.000	0.000
X10			1.158	1.024	0.772	0.583
T-Value			19.71	19.31	15.83	12.57
P-Value			0.000	0.000	0.000	0.000
X8				241	622	336
T-Value				15.70	23.62	10.68
P-Value				0.000	0.000	0.000
X18					275	423
T-Value					16.83	23.22
P-Value					0.000	0.000
X20						391824
T-Value						14.04
P-Value						0.000
S	187486	113375	95302	84787	74272	67497
R-Sq	86.91	95.22	96.63	97.33	97.95	98.31
R-Sq(adj)	86.90	95.21	96.61	97.32	97.94	98.30
Mallows C-p	13320.8	4276.3	2746.8	1979.5	1303.1	915.4
Step	7	8	9	10	11	12
Constant	-55789.3	-3516.2	-917.9	20013.2	22040.6	-8165.8
X17	0.357	0.301	0.298	0.238	0.242	0.255
T-Value	14.73	14.89	15.30	11.96	12.56	13.42
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X24	-298188	-12690				-196468
T-Value	-9.94	-0.45				-6.28
P-Value	0.000	0.653				0.000
X10	0.596	0.336	0.334	0.293	0.296	0.305
T-Value	13.54	8.75	8.75	7.93	8.05	8.45
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X8	367	128	116	-40	-46	58
T-Value	12.24	4.68	10.86	-1.97	-2.42	2.33
P-Value	0.000	0.000	0.000	0.049	0.016	0.020

X18	296	55	54	21		
T-Value	13.85	2.63	2.59	1.03		
P-Value	0.000	0.009	0.010	0.302		
X20	492253	887563	898824	1024100	1028257	909810
T-Value	17.38	29.55	54.37	48.60	49.71	32.86
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X22	209296	588397	592437	621604	638824	597592
T-Value	10.07	23.60	25.49	27.59	42.15	36.80
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X21		168348	170092	206301	210740	201556
T-Value		20.94	24.16	26.27	32.06	30.52
P-Value		0.000	0.000	0.000	0.000	0.000
X4				2.74	2.79	4.00
T-Value				9.03	9.37	11.44
P-Value				0.000	0.000	0.000
S	64123	52865	52842	50689	50690	49672
R-Sq	98.48	98.97	98.97	99.05	99.05	99.09
R-Sq(adj)	98.47	98.96	98.96	99.04	99.04	99.08
Mallows C-p	736.5	206.7	204.9	116.0	115.2	74.9
Step	13	14	15	16	17	
Constant	-15864	-29906	-37845	-37293	-26452	
X17	0.258	0.248	0.235	0.239	0.243	
T-Value	13.75	13.28	12.20	12.43	12.61	
P-Value	0.000	0.000	0.000	0.000	0.000	
X24	-189797	-246048	-260322	-248617	-205605	
T-Value	-6.16	-7.49	-7.84	-7.44	-5.41	
P-Value	0.000	0.000	0.000	0.000	0.000	
X10	0.304	0.290	0.279	0.271	0.287	
T-Value	8.56	8.22	7.91	7.66	7.99	
P-Value	0.000	0.000	0.000	0.000	0.000	
X8	59	88	112	107	117	
T-Value	2.41	3.51	4.21	4.04	4.36	
P-Value	0.016	0.000	0.000	0.000	0.000	
X18			52	75	79	
T-Value			2.66	3.48	3.64	
P-Value			0.008	0.001	0.000	
X20	910009	889668	872353	869103	890947	
T-Value	33.39	32.56	31.15	31.09	30.31	
P-Value	0.000	0.000	0.000	0.000	0.000	
X22	598974	594276	548901	531498	517102	
T-Value	37.47	37.50	23.59	21.97	20.77	
P-Value	0.000	0.000	0.000	0.000	0.000	
X21	200972	208094	197511	191499	189472	
T-Value	30.91	31.45	25.63	23.83	23.50	
P-Value	0.000	0.000	0.000	0.000	0.000	
X4	3.87	4.32	4.28	4.16	4.18	
T-Value	11.21	12.17	12.08	11.68	11.75	

P-Value	0.000	0.000	0.000	0.000	0.000
X3	17814	17623	17525	17426	17406
T-Value	5.54	5.54	5.52	5.51	5.51
P-Value	0.000	0.000	0.000	0.000	0.000
X6		9975	11241	10215	9919
T-Value		4.60	5.08	4.55	4.42
P-Value		0.000	0.000	0.000	0.000
X7				-14973	-14627
T-Value				-2.55	-2.49
P-Value				0.011	0.013
X5					-3652
T-Value					-2.35
P-Value					0.019
S	48896	48373	48215	48073	47956
R-Sq	99.12	99.14	99.14	99.15	99.15
R-Sq(adj)	99.11	99.13	99.13	99.14	99.14
Mallows C-p	45.1	25.6	20.5	16.0	12.5

Anexo 12. Método Stepwise etapa 3: D

Stepwise Regression: Y versus X20, X21, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is Y on 15 predictors, with N = 936

Step	1	2	3	4	5	6
Constant	154415	-16637	-23656	-36918	-99968	-177846
X10	3.554	2.231	1.522	1.263	0.752	0.525
T-Value	54.97	48.01	27.12	24.18	16.09	12.15
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X17		0.670	0.890	0.760	0.478	0.310
T-Value		45.31	50.42	43.19	25.95	16.29
P-Value		0.000	0.000	0.000	0.000	0.000
X24			-191000	-478494	-710622	-717682
T-Value			-18.03	-23.71	-37.79	-43.49
P-Value			0.000	0.000	0.000	0.000
X9				5.48	11.79	3.17
T-Value				16.09	30.94	5.16
P-Value				0.000	0.000	0.000
X18					382	585
T-Value					23.48	31.22
P-Value					0.000	0.000
X8						657
T-Value						16.72
P-Value						0.000
S	170538	95375	82162	72719	57650	50573
R-Sq	76.39	92.62	94.53	95.72	97.31	97.93
R-Sq(adj)	76.36	92.61	94.51	95.70	97.30	97.92
Mallows C-p	17800.4	4922.7	3410.8	2469.0	1207.5	716.6
Step	7	8	9	10	11	12
Constant	-176353	-138034	-40948	-22918	-23558	-29594
X10	0.481	0.447	0.402	0.319	0.322	0.322
T-Value	11.74	11.09	11.08	9.21	9.34	9.50
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X17	0.261	0.260	0.155	0.140	0.141	0.142
T-Value	14.11	14.41	8.80	8.45	8.55	8.76
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X24	-933463	-686053	-232417	-162615	-178117	-168055
T-Value	-36.97	-15.69	-4.69	-3.48	-4.34	-4.16
P-Value	0.000	0.000	0.000	0.001	0.000	0.000
X9	3.69	1.33	1.45	-0.40		
T-Value	6.35	2.01	2.43	-0.69		
P-Value	0.000	0.045	0.015	0.490		

X18	541	563	418	189	189	190
T-Value	29.89	31.34	22.23	7.11	7.14	7.31
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X8	590	456	186	176	167	168
T-Value	15.70	10.98	4.49	4.54	4.57	4.67
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X4	4.10	4.50	1.92	2.96	2.94	2.77
T-Value	10.85	12.05	5.10	8.15	8.12	7.76
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X20		158460	464865	603330	594531	596163
T-Value		6.85	15.94	20.26	22.09	22.54
P-Value		0.000	0.000	0.000	0.000	0.000
X22			198828	356904	354409	355715
T-Value			14.96	19.32	19.57	19.98
P-Value			0.000	0.000	0.000	0.000
X21				90875	89379	88724
T-Value				11.56	11.84	11.95
P-Value				0.000	0.000	0.000
X3						14669
T-Value						5.80
P-Value						0.000
S	47669	46531	41778	39071	39060	38389
R-Sq	98.17	98.26	98.60	98.77	98.77	98.82
R-Sq(adj)	98.15	98.24	98.58	98.76	98.76	98.80
Mallows C-p	534.2	466.1	198.6	59.8	58.3	26.1
Step	13	14				
Constant	-32813	-31407				
X10	0.270	0.262				
T-Value	7.51	7.25				
P-Value	0.000	0.000				
X17	0.130	0.133				
T-Value	7.92	8.08				
P-Value	0.000	0.000				
X24	-194988	-205837				
T-Value	-4.80	-5.02				
P-Value	0.000	0.000				
X9						
T-Value						
P-Value						
X18	172	167				
T-Value	6.55	6.32				
P-Value	0.000	0.000				
X8	136	143				
T-Value	3.72	3.89				
P-Value	0.000	0.000				
X4	3.05	2.90				
T-Value	8.44	7.86				

P-Value	0.000	0.000
X20	590937	577625
T-Value	22.49	21.21
P-Value	0.000	0.000
X22	378602	386141
T-Value	20.40	20.34
P-Value	0.000	0.000
X21	98495	96430
T-Value	12.70	12.32
P-Value	0.000	0.000
X3	14542	14617
T-Value	5.80	5.83
P-Value	0.000	0.000
X5	3381	4028
T-Value	4.01	4.41
P-Value	0.000	0.000
X6		-6461
T-Value		-1.83
P-Value		0.068
S	38080	38032
R-Sq	98.84	98.84
R-Sq(adj)	98.82	98.82
Mallows C-p	12.0	10.7

Anexo 13. Método Stepwise etapa 3: Combo

Stepwise Regression: Y versus X13, X15, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is Y on 17 predictors, with N = 2648

Step	1	2	3	4	5	6
Constant	-33003.9	-32157.2	100.2	18896.1	6569.6	-1893.3
X17	1.2317	0.7620	0.5913	0.4495	0.4732	0.4728
T-Value	87.70	76.45	67.48	61.18	65.82	67.48
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X10		1.900	1.480	0.624	0.433	0.429
T-Value		76.39	67.89	24.66	15.98	16.26
P-Value		0.000	0.000	0.000	0.000	0.000
X20			190471	388672	420925	421673
T-Value			41.66	69.02	73.00	75.03
P-Value			0.000	0.000	0.000	0.000
X22				176377	204626	205135
T-Value				44.57	48.81	50.21
P-Value				0.000	0.000	0.000
X21					40974	41101
T-Value					15.69	16.14
P-Value					0.000	0.000
X3						18743
T-Value						11.87
P-Value						0.000
S	132340	73924	57446	43414	41531	40474
R-Sq	74.40	92.02	95.18	97.25	97.48	97.61
R-Sq(adj)	74.39	92.01	95.17	97.24	97.48	97.60
Mallows C-p	29717.0	7451.5	3453.0	840.6	546.2	387.2

Step	7	8	9	10	11	12
Constant	-2325.8	-7400.3	-9271.7	-3302.4	-768.7	-5184.1
X17	0.373	0.343	0.329	0.321	0.331	0.298
T-Value	27.97	25.43	24.56	23.81	24.10	20.39
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X10	0.401	0.402	0.380	0.380	0.387	0.391
T-Value	15.31	15.57	14.85	14.89	15.18	15.42
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X20	464585	476323	471451	471015	471009	473890
T-Value	62.96	64.64	64.63	64.79	64.94	65.69
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X22	229840	236388	201819	200082	194246	199532
T-Value	46.82	48.43	31.98	31.75	29.94	30.72
P-Value	0.000	0.000	0.000	0.000	0.000	0.000

X21	49762	51826	46194	40972	42613	46042
T-Value	18.46	19.47	17.05	13.90	14.32	15.33
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X3	18363	18267	18019	18076	17850	17968
T-Value	11.79	11.92	11.91	11.99	11.86	12.02
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X4	0.732	0.914	0.976	0.910	1.198	1.032
T-Value	8.80	10.87	11.71	10.78	10.38	8.78
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X13		15410	16493	17322	17834	17433
T-Value		9.36	10.12	10.59	10.89	10.71
P-Value		0.000	0.000	0.000	0.000	0.000
X18			44.7	43.7	42.8	52.0
T-Value			8.47	8.30	8.13	9.59
P-Value			0.000	0.000	0.000	0.000
X23				-13134	-14228	-13612
T-Value				-4.41	-4.76	-4.59
P-Value				0.000	0.000	0.000
X5					-1977	-4090
T-Value					-3.65	-6.45
P-Value					0.000	0.000
X8						65
T-Value						6.30
P-Value						0.000
S	39900	39262	38746	38611	38521	38241
R-Sq	97.68	97.75	97.81	97.83	97.84	97.87
R-Sq(adj)	97.67	97.75	97.80	97.82	97.83	97.86
Mallows C-p	303.0	210.7	137.5	119.2	107.5	68.9
Step	13	14	15	16	17	
Constant	-26990	-29051	-28764	-28402	-31334	
X17	0.312	0.313	0.308	0.305	0.305	
T-Value	21.28	21.62	21.14	20.82	20.82	
P-Value	0.000	0.000	0.000	0.000	0.000	
X10	0.387	0.386	0.386	0.388	0.388	
T-Value	15.37	15.36	15.35	15.43	15.45	
P-Value	0.000	0.000	0.000	0.000	0.000	
X20	420826	418070	409909	415727	411485	
T-Value	39.50	41.97	39.21	38.40	36.98	
P-Value	0.000	0.000	0.000	0.000	0.000	
X22	191710	191656	189148	191620	191307	
T-Value	29.29	29.28	28.61	28.53	28.48	
P-Value	0.000	0.000	0.000	0.000	0.000	
X21	50237	51179	48712	49034	49405	
T-Value	16.50	18.57	16.69	16.78	16.87	
P-Value	0.000	0.000	0.000	0.000	0.000	
X3	17615	17593	17575	17552	17596	
T-Value	11.88	11.86	11.86	11.86	11.89	

P-Value	0.000	0.000	0.000	0.000	0.000
X4	1.63	1.66	1.70	1.75	1.71
T-Value	11.12	12.05	12.28	12.45	12.01
P-Value	0.000	0.000	0.000	0.000	0.000
X13	13363	13020	12263	14639	15055
T-Value	7.75	7.85	7.29	7.17	7.32
P-Value	0.000	0.000	0.000	0.000	0.000
X18	57.3	57.8	58.3	54.7	56.2
T-Value	10.55	10.71	10.80	9.67	9.81
P-Value	0.000	0.000	0.000	0.000	0.000
X23	-2466				
T-Value	-0.73				
P-Value	0.466				
X5	-2933	-2845	-2885	-3259	-3002
T-Value	-4.50	-4.44	-4.51	-4.90	-4.40
P-Value	0.000	0.000	0.000	0.000	0.000
X8	159	164	137	126	134
T-Value	9.17	10.41	7.19	6.35	6.56
P-Value	0.000	0.000	0.000	0.000	0.000
X24	-105131	-110733	-132993	-126723	-134840
T-Value	-6.72	-8.13	-8.22	-7.70	-7.85
P-Value	0.000	0.000	0.000	0.000	0.000
X9			1.02	1.09	1.12
T-Value			2.54	2.71	2.78
P-Value			0.011	0.007	0.006
X15				4389	4712
T-Value				2.05	2.19
P-Value				0.040	0.028
X6					2020
T-Value					1.64
P-Value					0.101
S	37925	37922	37882	37859	37847
R-Sq	97.91	97.91	97.91	97.92	97.92
R-Sq(adj)	97.90	97.90	97.90	97.90	97.91
Mallows C-p	25.5	24.1	19.6	17.4	16.7

Anexo 14. Método Stepwise etapa 4: A

Stepwise Regression: Y versus X20, X21, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is Y on 32 predictors, with N = 928

Step	1	2	3	4
Constant	111486	103201	60349	72616
X17^2	0.00000	0.00000	0.00000	0.00000
T-Value	83.91	62.76	71.25	61.73
P-Value	0.000	0.000	0.000	0.000
X10		1.490	1.331	1.293
T-Value		28.42	27.84	29.94
P-Value		0.000	0.000	0.000
X4X21			26.10	25.90
T-Value			15.60	17.16
P-Value			0.000	0.000
X4X20				6.71
T-Value				14.60
P-Value				0.000
X4X22				
T-Value				
P-Value				
X8				
T-Value				
P-Value				
S	162800	119023	105953	95545
R-Sq	88.38	93.79	95.09	96.01
R-Sq(adj)	88.36	93.78	95.07	95.99
PRESS	24652565367168	13290853760841	10511460571124	8578066327081
R-Sq(pred)	88.32	93.71	95.02	95.94
Step	5	6		
Constant	101354	32048		
X17^2	0.00000	0.00000		
T-Value	35.66	15.37		
P-Value	0.000	0.000		
X10	0.274	0.053		
T-Value	7.00	2.01		
P-Value	0.000	0.045		
X4X21	36.16	48.00		
T-Value	36.00	65.15		
P-Value	0.000	0.000		
X4X20	27.75	37.85		
T-Value	42.97	74.25		
P-Value	0.000	0.000		

X4X22	18.90	29.01		
T-Value	36.58	65.52		
P-Value	0.000	0.000		
X8		182.8		
T-Value		35.24		
P-Value		0.000		
S	61057	39863		
R-Sq	98.37	99.31		
R-Sq(adj)	98.36	99.30		
PRESS	3493966335355	1494361384541		
R-Sq(pred)	98.35	99.29		
Step	7	8	9	10
Constant	2042	7647	9720	9948
X17^2	-0.00000			
T-Value	-1.44			
P-Value	0.150			
X10	0.041	0.048	0.017	
T-Value	1.61	1.90	0.66	
P-Value	0.109	0.058	0.511	
X4X21	42.09	43.03	44.50	44.76
T-Value	39.70	51.39	51.88	59.08
P-Value	0.000	0.000	0.000	0.000
X4X20	36.79	36.76	31.63	31.71
T-Value	71.56	71.52	31.76	32.12
P-Value	0.000	0.000	0.000	0.000
X4X22	26.75	26.97	28.25	28.47
T-Value	51.13	53.67	52.54	68.05
P-Value	0.000	0.000	0.000	0.000
X8	113.9	124.4	140.5	142.1
T-Value	10.93	16.72	18.05	19.26
P-Value	0.000	0.000	0.000	0.000
X17	0.317	0.261	0.222	0.218
T-Value	7.55	17.48	13.85	14.61
P-Value	0.000	0.000	0.000	0.000
X20			159713	163334
T-Value			5.98	6.25
P-Value			0.000	0.000
X4				
T-Value				
P-Value				
X21				
T-Value				
P-Value				
S	38702	38725	38014	38002
R-Sq	99.35	99.35	99.37	99.37
R-Sq(adj)	99.34	99.34	99.37	99.37
PRESS	1412224898011	1409928493655	1371949256915	1368129620131
R-Sq(pred)	99.33	99.33	99.35	99.35

Step	11	12		
Constant	18370	18830		
X17^2				
T-Value				
P-Value				
X10				
T-Value				
P-Value				
X4X21	45.52	54.34		
T-Value	60.54	13.17		
P-Value	0.000	0.000		
X4X20	30.12	30.14		
T-Value	30.14	30.21		
P-Value	0.000	0.000		
X4X22	28.39	28.38		
T-Value	69.24	69.36		
P-Value	0.000	0.000		
X8	75.3	75.7		
T-Value	5.87	5.92		
P-Value	0.000	0.000		
X17	0.190	0.190		
T-Value	12.44	12.46		
P-Value	0.000	0.000		
X20	238333	237479		
T-Value	8.45	8.43		
P-Value	0.000	0.000		
X4	1.53	1.51		
T-Value	6.31	6.24		
P-Value	0.000	0.000		
X21		-53823		
T-Value		-2.17		
P-Value		0.030		
S	37227	37151		
R-Sq	99.40	99.40		
R-Sq(adj)	99.39	99.39		
PRESS	1316319546704	1310464399559		
R-Sq(pred)	99.38	99.38		
Step	13	14	15	16
Constant	11371	2803	-15365	-18491
X17^2				
T-Value				
P-Value				
X10				
T-Value				
P-Value				
X4X21	53.5	53.1	51.4	48.0

T-Value	12.88	12.79	12.23	44.43
P-Value	0.000	0.000	0.000	0.000
X4X20	29.8	29.4	28.3	28.1
T-Value	29.27	28.24	24.95	25.23
P-Value	0.000	0.000	0.000	0.000
X4X22	27.72	27.63	26.88	26.77
T-Value	49.79	49.50	42.02	42.87
P-Value	0.000	0.000	0.000	0.000
X8	86	88	75	74
T-Value	6.13	6.30	5.00	4.96
P-Value	0.000	0.000	0.000	0.000
X17	0.203	0.202	0.205	0.205
T-Value	11.98	11.89	12.06	12.12
P-Value	0.000	0.000	0.000	0.000
X20	223308	231094	246205	248116
T-Value	7.63	7.82	8.16	8.25
P-Value	0.000	0.000	0.000	0.000
X4	1.87	2.28	4.17	4.42
T-Value	5.88	5.84	4.69	5.32
P-Value	0.000	0.000	0.000	0.000
X21	-49398	-41103	-21875	
T-Value	-1.99	-1.63	-0.83	
P-Value	0.047	0.104	0.409	
X17X24	-0.059	-0.090	-0.126	-0.135
T-Value	-1.75	-2.38	-3.11	-3.45
P-Value	0.080	0.017	0.002	0.001
X4X23		2.4	5.7	6.3
T-Value		1.80	2.95	3.45
P-Value		0.073	0.003	0.001
X4^2			-0.00002	-0.00002
T-Value			-2.37	-2.75
P-Value			0.018	0.006
X4^3				
T-Value				
P-Value				
X23				
T-Value				
P-Value				
S	37110	37065	36972	36966
R-Sq	99.40	99.40	99.41	99.41
R-Sq(adj)	99.40	99.40	99.40	99.40
PRESS	1308842684727	1305341440641	1299560213196	1299615085236
R-Sq(pred)	99.38	99.38	99.38	99.38
Step	17	18		
Constant	-33340	-49251		
X17^2				
T-Value				
P-Value				

X10		
T-Value		
P-Value		
X4X21		
	49.2	50.4
T-Value	40.63	37.36
P-Value	0.000	0.000
X4X20		
	28.7	28.3
T-Value	25.07	24.56
P-Value	0.000	0.000
X4X22		
	27.05	26.73
T-Value	42.40	40.64
P-Value	0.000	0.000
X8		
	76	74
T-Value	5.06	4.97
P-Value	0.000	0.000
X17		
	0.197	0.199
T-Value	11.36	11.47
P-Value	0.000	0.000
X20		
	240744	241436
T-Value	7.96	8.00
P-Value	0.000	0.000
X4		
	6.54	8.13
T-Value	4.97	5.28
P-Value	0.000	0.000
X21		
T-Value		
P-Value		
X17X24		
	-0.077	-0.087
T-Value	-1.62	-1.82
P-Value	0.105	0.070
X4X23		
	8.4	-1.5
T-Value	4.02	-0.27
P-Value	0.000	0.785
X4^2		
	-0.00009	-0.00012
T-Value	-2.53	-3.13
P-Value	0.012	0.002
X4^3		
	0.00000	0.00000
T-Value	2.07	2.65
P-Value	0.039	0.008
X23		
		49664
T-Value		1.98
P-Value		0.048
S		
	36900	36841
R-Sq	99.41	99.41
R-Sq(adj)	99.40	99.40
PRESS	1296076840578	1291831545849
R-Sq(pred)	99.39	99.39

Step	19	20	21
Constant	-48324	-55794	-54607
X17^2			
T-Value			
P-Value			
X10			
T-Value			
P-Value			
X4X21	50.3	50.8	46.5
T-Value	37.71	37.49	14.76
P-Value	0.000	0.000	0.000
X4X20	28.4	28.2	28.2
T-Value	24.62	24.51	24.52
P-Value	0.000	0.000	0.000
X4X22	26.76	26.89	26.86
T-Value	41.20	41.22	41.18
P-Value	0.000	0.000	0.000
X8	74	60	59
T-Value	4.97	3.59	3.50
P-Value	0.000	0.000	0.000
X17	0.199	0.195	0.197
T-Value	11.50	11.27	11.34
P-Value	0.000	0.000	0.000
X20	241821	236521	236155
T-Value	8.02	7.83	7.82
P-Value	0.000	0.000	0.000
X4	8.0	8.2	8.1
T-Value	5.34	5.45	5.38
P-Value	0.000	0.000	0.000
X21			
T-Value			
P-Value			
X17X24	-0.086	-0.079	-0.079
T-Value	-1.80	-1.64	-1.64
P-Value	0.072	0.101	0.101
X4X23			
T-Value			
P-Value			
X4^2	-0.00012	-0.00012	-0.00012
T-Value	-3.13	-3.17	-3.09
P-Value	0.002	0.002	0.002
X4^3	0.00000	0.00000	0.00000
T-Value	2.64	2.62	2.53
P-Value	0.008	0.009	0.011
X23	43321	46791	46020
T-Value	4.48	4.76	4.68
P-Value	0.000	0.000	0.000

X5		2096	2115
T-Value		1.88	1.90
P-Value		0.061	0.058
X10X21			0.44
T-Value			1.48
P-Value			0.139
S	36823	36772	36748
R-Sq	99.41	99.41	99.42
R-Sq(adj)	99.40	99.41	99.41
PRESS	1290832712135	1289822218735	1287751787234
R-Sq(pred)	99.39	99.39	99.39

Anexo 15. Método Stepwise etapa 4: B**Stepwise Regression: Y versus X20, X21, ...**

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is Y on 43 predictors, with N = 923

Step	1	2	3	4	5	6
Constant	-97635	-55114	-81183	-200918	-176328	-153414
X17	1.2397	1.2851	1.1121	1.8076	2.1141	1.8861
T-Value	78.91	140.95	96.27	60.91	78.20	68.51
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X17X24		-0.689	-1.201	-1.505	-0.851	-0.700
T-Value		-43.03	-41.46	-58.62	-24.45	-21.88
P-Value		0.000	0.000	0.000	0.000	0.000
X9			6.85	8.50	0.79	0.57
T-Value			19.93	30.89	1.99	1.64
P-Value			0.000	0.000	0.047	0.102
X17^2				-0.00000	-0.00000	-0.00000
T-Value				-24.59	-38.57	-37.29
P-Value				0.000	0.000	0.000
X4X20					11.22	10.19
T-Value					23.18	23.67
P-Value					0.000	0.000
X10						0.539
T-Value						16.42
P-Value						0.000
S	186707	107627	89975	69902	55539	48844
R-Sq	87.12	95.72	97.01	98.20	98.86	99.12
R-Sq(adj)	87.10	95.71	97.00	98.19	98.86	99.12
Step	7	8	9	10	11	12
Constant	-159959	-119700	26034	27294	30793	16840
X17	1.8738	1.5459	0.2216	0.2119	0.1791	0.1841
T-Value	75.04	40.03	6.01	6.36	19.01	20.91
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X17X24	-0.793	-0.703	0.066	0.073	0.084	-0.127
T-Value	-26.69	-24.02	2.76	3.49	4.59	-5.13
P-Value	0.000	0.000	0.006	0.000	0.000	0.000
X9	1.59	2.46	1.79	1.77	1.82	1.68
T-Value	4.89	7.75	10.00	10.05	10.80	10.64
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X17^2	-0.00000	-0.00000	-0.00000	-0.00000		
T-Value	-40.36	-29.86	-1.18	-1.02		
P-Value	0.000	0.000	0.238	0.306		
X4X20	32.92	33.66	35.41	34.96	35.26	33.74

T-Value	19.90	21.55	40.27	70.76	89.00	86.05
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X10	0.520	0.495	0.099	0.097	0.093	0.094
T-Value	17.47	17.55	5.44	5.43	5.33	5.79
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X20	-758217	-646874	-20022			
T-Value	-14.14	-12.53	-0.62			
P-Value	0.000	0.000	0.535			
X4X22		4.94	25.18	25.33	25.71	24.12
T-Value		10.71	48.06	54.77	93.87	83.20
P-Value		0.000	0.000	0.000	0.000	0.000
X4X21			39.18	39.42	40.00	40.59
T-Value			44.46	49.75	71.74	77.63
P-Value			0.000	0.000	0.000	0.000
X4						1.96
T-Value						11.67
P-Value						0.000
S	44274	41755	23481	23473	23474	21911
R-Sq	99.28	99.36	99.80	99.80	99.80	99.82
R-Sq(adj)	99.27	99.35	99.80	99.80	99.80	99.82
Step	13	14	15	16	17	18
Constant	15422	17557	12839	15684	6977	-2326
X17	0.1928	0.1849	0.2250	0.2028	0.2003	0.2029
T-Value	22.45	22.00	22.20	20.11	20.25	20.77
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X17X24	-0.141	-0.112	-0.142	-0.123	-0.048	-0.075
T-Value	-5.84	-4.75	-6.04	-5.42	-1.91	-2.94
P-Value	0.000	0.000	0.000	0.000	0.057	0.003
X9	1.77	1.44	1.38	1.39	1.61	1.25
T-Value	11.52	9.26	9.05	9.50	10.90	7.72
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X17^2						
T-Value						
P-Value						
X4X20	32.91	33.40	32.72	32.07	31.00	30.80
T-Value	83.64	86.18	83.64	83.52	74.76	74.96
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X10	0.101	0.111	0.047	-0.090	-0.088	-0.091
T-Value	6.43	7.23	2.65	-3.87	-3.86	-4.01
P-Value	0.000	0.000	0.008	0.000	0.000	0.000
X20						
T-Value						
P-Value						
X4X22	23.86	18.62	20.69	23.10	22.17	22.29
T-Value	84.48	24.87	26.10	28.44	27.36	27.87
P-Value	0.000	0.000	0.000	0.000	0.000	0.000

X4X21	40.18	40.67	40.70	43.08	43.11	43.29
T-Value	79.03	81.67	83.69	79.34	80.99	82.25
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X4	1.95	2.04	2.07	2.05	3.15	3.09
T-Value	11.95	12.85	13.35	13.74	13.64	13.55
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X3X20	28339	27494	28746	29318	29238	29437
T-Value	7.99	7.98	8.53	9.04	9.20	9.39
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X22		162537	206325	214482	212408	221736
T-Value		7.51	9.33	10.07	10.18	10.73
P-Value		0.000	0.000	0.000	0.000	0.000
X17X22			-0.144	-0.187	-0.182	-0.190
T-Value			-6.75	-8.88	-8.80	-9.29
P-Value			0.000	0.000	0.000	0.000
X10X20				0.293	0.295	0.246
T-Value				8.63	8.87	7.18
P-Value				0.000	0.000	0.000
X4X24					-1.76	-1.85
T-Value					-6.17	-6.54
P-Value					0.000	0.000
X5						3119
T-Value						5.10
P-Value						0.000
S	21195	20580	20095	19331	18949	18693
R-Sq	99.84	99.84	99.85	99.86	99.87	99.87
R-Sq (adj)	99.83	99.84	99.85	99.86	99.87	99.87
Step	19	20	21	22	23	24
Constant	-2966	-2529	-10473	-14293	-11785	-8884
X17	0.2032	0.2034	0.2011	0.1996	0.1995	0.1976
T-Value	21.10	21.26	21.10	21.05	21.10	20.94
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X17X24	-0.073	-0.071	-0.071	-0.067	-0.062	-0.059
T-Value	-2.91	-2.82	-2.87	-2.68	-2.49	-2.40
P-Value	0.004	0.005	0.004	0.007	0.013	0.017
X9	1.09	1.09	1.13	1.03	1.33	1.40
T-Value	6.71	6.74	7.01	6.37	6.66	6.98
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X17^2						
T-Value						
P-Value						
X4X20	30.90	30.93	28.20	27.41	27.76	28.37
T-Value	76.23	76.79	32.04	30.32	30.44	30.43
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X10	-0.078	-0.078	-0.077	-0.083	-0.082	-0.084
T-Value	-3.49	-3.51	-3.47	-3.75	-3.73	-3.82
P-Value	0.001	0.000	0.001	0.000	0.000	0.000

X20						
T-Value						
P-Value						
X4X22	21.41	21.38	18.52	17.56	17.88	18.52
T-Value	26.58	26.73	16.19	15.00	15.23	15.57
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X4X21	43.25	43.26	41.58	37.80	37.57	39.58
T-Value	83.37	83.95	58.96	28.99	28.83	26.86
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X4	3.06	3.07	5.97	6.96	6.98	6.45
T-Value	13.62	13.74	6.92	7.70	7.73	7.03
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X3X20	29310	29282	29003	28692	28121	27979
T-Value	9.48	9.54	9.50	9.45	9.26	9.25
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X22	188041	188605	192998	199716	182267	176106
T-Value	8.81	8.89	9.14	9.47	8.24	7.96
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X17X22	-0.176	-0.176	-0.173	-0.175	-0.172	-0.171
T-Value	-8.62	-8.68	-8.56	-8.72	-8.62	-8.58
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X10X20	0.224	0.226	0.224	0.227	0.208	0.207
T-Value	6.61	6.71	6.68	6.82	6.12	6.10
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X4X24	-1.81	-1.88	-4.75	-5.71	-5.68	-5.16
T-Value	-6.51	-6.79	-5.47	-6.29	-6.27	-5.61
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X5	3615	3483	3743	4007	4387	4480
T-Value	5.93	5.74	6.16	6.58	7.02	7.18
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X8X22	216	216	216	220	241	246
T-Value	5.29	5.31	5.35	5.48	5.90	6.04
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X3X24		11029	11020	11055	11409	11462
T-Value		3.63	3.65	3.68	3.81	3.84
P-Value		0.000	0.000	0.000	0.000	0.000
X25			-27654	-35618	-35074	-30776
T-Value			-3.48	-4.33	-4.28	-3.71
P-Value			0.001	0.000	0.000	0.000
X8X21				73	79	177
T-Value				3.44	3.71	4.42
P-Value				0.001	0.000	0.000
X8					-30	-37
T-Value					-2.54	-3.11
P-Value					0.011	0.002
X21						-37371
T-Value						-2.89

P-Value						0.004
S	18421	18299	18187	18080	18026	17953
R-Sq	99.88	99.88	99.88	99.88	99.88	99.88
R-Sq(adj)	99.87	99.88	99.88	99.88	99.88	99.88
Step	25	26	27	28		
Constant	51739	125794	135244	278086		
X17	0.1942	0.1861	0.1825	0.1772		
T-Value	20.60	19.11	20.74	20.38		
P-Value	0.000	0.000	0.000	0.000		
X17X24	-0.048	-0.023				
T-Value	-1.94	-0.88				
P-Value	0.053	0.380				
X9	1.32	1.31	1.29	1.47		
T-Value	6.59	6.58	6.52	7.47		
P-Value	0.000	0.000	0.000	0.000		
X17^2						
T-Value						
P-Value						
X4X20	26.9	25.0	24.9	25.4		
T-Value	26.64	21.08	21.08	21.82		
P-Value	0.000	0.000	0.000	0.000		
X10	-0.083	-0.083	-0.083	-0.082		
T-Value	-3.83	-3.84	-3.83	-3.87		
P-Value	0.000	0.000	0.000	0.000		
X20						
T-Value						
P-Value						
X4X22	17.8	17.5	17.4	17.1		
T-Value	14.82	14.58	14.56	14.59		
P-Value	0.000	0.000	0.000	0.000		
X4X21	39.1	39.3	39.4	39.7		
T-Value	26.62	26.84	26.93	27.60		
P-Value	0.000	0.000	0.000	0.000		
X4	7.01	7.02	7.04	6.92		
T-Value	7.59	7.63	7.65	7.66		
P-Value	0.000	0.000	0.000	0.000		
X3X20	29268	31508	31551	30377		
T-Value	9.68	10.17	10.19	9.96		
P-Value	0.000	0.000	0.000	0.000		
X22	125415	59213	52179	-80620		
T-Value	4.82	1.75	1.59	-2.04		
P-Value	0.000	0.080	0.112	0.042		
X17X22	-0.168	-0.159	-0.156	-0.148		
T-Value	-8.46	-8.00	-8.00	-7.74		
P-Value	0.000	0.000	0.000	0.000		
X10X20	0.213	0.218	0.222	0.173		

T-Value	6.31	6.47	6.65	5.13
P-Value	0.000	0.000	0.000	0.000
X4X24	-6.04	-7.53	-7.79	-10.42
T-Value	-6.39	-7.11	-7.67	-9.50
P-Value	0.000	0.000	0.000	0.000
X5	3468	3045	2893	1447
T-Value	5.11	4.41	4.33	2.06
P-Value	0.000	0.000	0.000	0.040
X8X22	257	259	260	232
T-Value	6.34	6.40	6.45	5.81
P-Value	0.000	0.000	0.000	0.000
X3X24	11278	11724	11841	11862
T-Value	3.81	3.97	4.01	4.09
P-Value	0.000	0.000	0.000	0.000
X25	-90004	-160829	-168985	-308466
T-Value	-4.92	-5.46	-6.04	-8.43
P-Value	0.000	0.000	0.000	0.000
X8X21	192	198	201	183
T-Value	4.80	4.96	5.06	4.67
P-Value	0.000	0.000	0.000	0.000
X8	-51	-55	-56	-43
T-Value	-4.10	-4.39	-4.56	-3.49
P-Value	0.000	0.000	0.000	0.001
X21	-92810	-164695	-173198	-310833
T-Value	-4.65	-5.35	-5.93	-8.33
P-Value	0.000	0.000	0.000	0.000
X23	-55529	-126356	-134411	-274896
T-Value	-3.63	-4.56	-5.14	-7.77
P-Value	0.000	0.000	0.000	0.000
X4^3		0.00000	0.00000	0.00000
T-Value		3.06	3.52	6.58
P-Value		0.002	0.000	0.000
X8X20				-154
T-Value				-5.78
P-Value				0.000
S	17833	17751	17749	17438
R-Sq	99.89	99.89	99.89	99.89
R-Sq(adj)	99.88	99.88	99.88	99.89

Anexo 16. Método Stepwise etapa 4: D

Stepwise Regression: Y versus X20, X21, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is Y on 50 predictors, with N = 936

Step	1	2	3	4	5	6
Constant	154415	-16637	-26225	-38765	-101144	-82031
X10	3.554	2.231	1.463	1.268	1.119	0.988
T-Value	54.97	48.01	26.16	24.20	23.65	22.15
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X17		0.670	0.912	0.810	1.315	1.493
T-Value		45.31	51.51	46.19	37.34	42.56
P-Value		0.000	0.000	0.000	0.000	0.000
X24X17			-0.359	-0.749	-0.868	-0.376
T-Value			-19.32	-23.50	-29.71	-8.15
P-Value			0.000	0.000	0.000	0.000
X9				4.33	4.62	-0.32
T-Value				14.40	17.29	-0.71
P-Value				0.000	0.000	0.479
X17^2					-0.00000	-0.00000
T-Value					-15.98	-21.44
P-Value					0.000	0.000
X20X4						9.16
T-Value						13.14
P-Value						0.000
S	170538	95375	80638	72964	64660	59410
R-Sq	76.39	92.62	94.73	95.69	96.62	97.15
R-Sq(adj)	76.36	92.61	94.72	95.67	96.60	97.13
Step	7	8	9	10	11	12
Constant	-83262	8883	11243	-10424	42524	43483
X10	0.990	0.596	0.595	0.205	0.032	
T-Value	22.21	15.23	15.23	6.47	1.20	
P-Value	0.000	0.000	0.000	0.000	0.229	
X17	1.483	0.719	0.698	0.691	0.133	0.123
T-Value	45.79	17.35	31.78	43.05	4.90	4.75
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X24X17	-0.407	-0.012				
T-Value	-27.48	-0.60				
P-Value	0.000	0.550				
X9						
T-Value						
P-Value						
X17^2	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000

T-Value	-22.56	-12.38	-14.67	-18.05	-1.67	-1.45
P-Value	0.000	0.000	0.000	0.000	0.096	0.148
X20X4	8.75	24.65	24.99	28.82	39.27	39.70
T-Value	22.94	33.25	53.41	78.46	73.64	99.87
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X22X4		15.46	15.79	19.14	28.92	29.27
T-Value		23.49	41.89	63.93	60.15	76.62
P-Value		0.000	0.000	0.000	0.000	0.000
X21X4				27.97	39.54	40.09
T-Value				28.51	42.87	50.24
P-Value				0.000	0.000	0.000
X24X18					2891	2933
T-Value					23.39	24.74
P-Value					0.000	0.000
S	59394	47069	47052	34381	27283	27289
R-Sq	97.15	98.21	98.21	99.05	99.40	99.40
R-Sq(adj)	97.13	98.20	98.20	99.04	99.40	99.39
Step	13	14	15	16	17	18
Constant	51354	49440	50376	50031	49737	51019
X10						
T-Value						
P-Value						
X17	0.043	0.062	0.053	0.055	0.059	0.046
T-Value	1.58	5.39	4.64	4.92	5.23	3.95
P-Value	0.114	0.000	0.000	0.000	0.000	0.000
X24X17						
T-Value						
P-Value						
X9						
T-Value						
P-Value						
X17^2	0.00000					
T-Value	0.75					
P-Value	0.455					
X20X4	40.98	40.81	37.01	37.00	36.20	32.20
T-Value	97.02	116.10	54.15	54.56	51.24	27.77
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X22X4	20.3	20.4	20.3	22.5	22.6	22.5
T-Value	16.21	16.37	16.62	16.81	16.96	17.02
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X21X4	41.35	41.12	41.30	41.94	41.87	42.15
T-Value	52.14	56.24	57.66	57.49	57.78	58.48
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X24X18	3316	3253	3322	3302	3277	3376
T-Value	26.33	35.15	36.41	36.42	36.30	36.58
P-Value	0.000	0.000	0.000	0.000	0.000	0.000

X22	202583	196750	203434	174616	172175	181284
T-Value	7.51	7.62	8.04	6.68	6.63	7.02
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X20X5			5701	5707	6223	5209
T-Value			6.42	6.48	7.03	5.74
P-Value			0.000	0.000	0.000	0.000
X18^3				-0.00005	-0.00005	-0.00006
T-Value				-3.93	-3.96	-4.06
P-Value				0.000	0.000	0.000
X20X3					16170	20147
T-Value					3.78	4.64
P-Value					0.000	0.000
X20X18						363
T-Value						4.33
P-Value						0.000
S	26510	26504	25947	25748	25566	25324
R-Sq	99.43	99.43	99.46	99.47	99.47	99.48
R-Sq(adj)	99.43	99.43	99.45	99.46	99.47	99.48
Step	19	20	21	22	23	24
Constant	51597	51997	52570	53721	51714	51670
X10						
T-Value						
P-Value						
X17	0.044	0.035	0.034	0.024	0.042	0.042
T-Value	3.86	3.13	3.06	2.03	3.19	3.21
P-Value	0.000	0.002	0.002	0.042	0.001	0.001
X24X17						
T-Value						
P-Value						
X9						
T-Value						
P-Value						
X17^2						
T-Value						
P-Value						
X20X4	32.2	30.5	30.5	30.5	30.5	30.5
T-Value	27.96	26.41	26.54	26.62	26.77	26.69
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X22X4	22.4	22.3	22.3	23.8	23.8	23.6
T-Value	17.13	17.34	17.47	17.39	17.45	17.27
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X21X4	42.05	42.26	49.22	49.58	49.19	50.48
T-Value	58.73	60.19	22.36	22.59	22.47	21.78
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X24X18	4397	4458	4466	4549	4404	4404
T-Value	15.42	15.96	16.07	16.35	15.67	15.68
P-Value	0.000	0.000	0.000	0.000	0.000	0.000

X22	181796	188399	188046	178248	227006	164443
T-Value	7.09	7.50	7.53	7.10	7.61	3.44
P-Value	0.000	0.000	0.000	0.000	0.000	0.001
X20X5	5213	4777	4780	4803	4763	4764
T-Value	5.79	5.40	5.43	5.48	5.46	5.46
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X18^3	-0.00006	-0.00005	-0.00005	-0.00006	-0.00006	-0.00014
T-Value	-4.09	-3.95	-4.07	-4.58	-4.65	-2.96
P-Value	0.000	0.000	0.000	0.000	0.000	0.003
X20X3	20147	21899	21903	21877	21938	22022
T-Value	4.68	5.18	5.21	5.23	5.26	5.29
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X20X18	365	440	442	464	427	441
T-Value	4.38	5.34	5.39	5.66	5.17	5.31
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X24	-64696	-137201	-137675	-139391	-136383	-136182
T-Value	-3.78	-6.77	-6.83	-6.94	-6.81	-6.80
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X8^3		0.00005	0.00005	0.00005	0.00005	0.00005
T-Value		6.34	6.38	6.51	6.33	6.31
P-Value		0.000	0.000	0.000	0.000	0.000
X21X5			-10485	-10626	-10482	-11012
T-Value			-3.33	-3.39	-3.36	-3.52
P-Value			0.001	0.001	0.001	0.000
X22X10				-0.088	-0.141	-0.134
T-Value				-2.93	-4.05	-3.83
P-Value				0.003	0.000	0.000
X22X17					-0.085	-0.080
T-Value					-2.99	-2.78
P-Value					0.003	0.006
X22X18						128
T-Value						1.68
P-Value						0.094
S	25144	24627	24493	24393	24288	24264
R-Sq	99.49	99.51	99.52	99.52	99.53	99.53
R-Sq(adj)	99.49	99.51	99.51	99.52	99.52	99.52

Anexo 17. Método Stepwise etapa 4: Combo

Stepwise Regression: Y versus X22, X24, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is Y on 24 predictors, with N = 2636

Step	1	2	3	4	5	6
Constant	159380	90681	61572	33761	31577	31230
X4X20	30.676	34.250	35.764	37.211	36.463	30.027
T-Value	85.70	216.51	299.88	439.29	469.07	70.77
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X4X22		21.758	23.695	25.545	24.911	25.005
T-Value		107.15	154.81	235.06	257.63	269.39
P-Value		0.000	0.000	0.000	0.000	0.000
X18X24			1977	2466	2282	2308
T-Value			48.40	85.10	88.12	92.84
P-Value			0.000	0.000	0.000	0.000
X4X21				34.47	34.30	34.50
T-Value				55.89	64.11	67.31
P-Value				0.000	0.000	0.000
X17X13					0.0935	0.0849
T-Value					29.45	27.46
P-Value					0.000	0.000
X18X20						350
T-Value						15.41
P-Value						0.000
S	134221	57983	42184	28529	24744	23702
R-Sq	73.60	95.08	97.39	98.81	99.10	99.18
R-Sq(adj)	73.59	95.07	97.39	98.81	99.10	99.18
Mallows C-p	150567.9	25949.3	12493.2	4287.2	2574.9	2146.2

Step	7	8	9	10	11	12
Constant	30116	30235	30144	29779	26204	11258
X4X20	30.02	28.45	28.28	23.53	22.81	19.10
T-Value	73.35	68.65	70.46	40.85	40.21	34.97
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X4X22	25.099	25.098	20.590	20.575	20.042	17.125
T-Value	279.54	288.46	59.06	60.41	59.51	51.04
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X18X24	1262	1262	1270	1273	610	1214
T-Value	16.13	16.64	17.30	17.75	6.56	13.56
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X4X21	34.91	34.88	35.01	35.15	35.57	37.22
T-Value	70.48	72.66	75.33	77.40	79.75	89.52
P-Value	0.000	0.000	0.000	0.000	0.000	0.000

X17X13	0.0825	0.0817	0.0714	0.0700	0.0730	0.0676
T-Value	27.63	28.21	24.60	24.63	26.13	26.26
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X18X20	354	402	416	403	403	427
T-Value	16.18	18.66	19.93	19.75	20.19	23.29
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X8X24	77.1	77.1	78.4	78.8	73.7	91.6
T-Value	14.05	14.50	15.22	15.67	14.91	19.90
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X3X20		28166	28326	33328	33259	33393
T-Value		13.10	13.61	16.01	16.33	17.88
P-Value		0.000	0.000	0.000	0.000	0.000
X17X22			0.218	0.221	0.219	0.237
T-Value			13.33	13.79	13.96	16.45
P-Value			0.000	0.000	0.000	0.000
X8X20				182	185	204
T-Value				11.25	11.71	14.03
P-Value				0.000	0.000	0.000
X4					0.790	4.368
T-Value					10.87	25.12
P-Value					0.000	0.000
X4X24						-4.16
T-Value						-22.28
P-Value						0.000
S	22864	22156	21447	20951	20499	18801
R-Sq	99.24	99.28	99.33	99.36	99.39	99.48
R-Sq(adj)	99.23	99.28	99.33	99.36	99.38	99.48
Mallows C-p	1815.3	1545.4	1283.6	1106.1	947.8	383.4
Step	13	14	15	16	17	18
Constant	9984	9601	9443	9446	9578	9956
X4X20	19.02	18.86	18.92	18.91	18.96	18.55
T-Value	35.48	35.55	36.05	36.36	36.56	35.38
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X4X22	13.16	12.80	13.32	12.15	11.91	11.62
T-Value	25.66	25.15	26.21	22.93	22.39	21.78
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X18X24	1206	1261	1261	1263	1299	1478
T-Value	13.73	14.48	14.64	14.79	15.18	15.74
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X4X21	37.46	37.21	37.12	37.12	36.97	36.91
T-Value	91.66	91.83	92.55	93.41	92.93	93.07
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X17X13	0.0710	0.0536	0.0526	0.0522	0.0434	0.0394
T-Value	27.86	16.07	15.93	15.97	11.11	9.89
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X18X20	424	427	422	423	415	422
T-Value	23.55	23.99	23.97	24.21	23.71	24.10

P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X8X24	92.1	94.6	94.8	94.8	96.2	123.0
T-Value	20.37	21.13	21.39	21.60	21.91	16.76
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X3X20	33313	33412	33342	33350	33305	33377
T-Value	18.17	18.44	18.60	18.78	18.81	18.92
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X17X22	0.183	0.141	0.112	0.124	0.137	0.139
T-Value	12.11	8.93	6.95	7.70	8.37	8.58
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X8X20	205	208	208	209	209	215
T-Value	14.33	14.76	14.92	15.07	15.12	15.55
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X4	4.53	4.59	4.59	4.59	4.63	4.49
T-Value	26.42	27.05	27.38	27.64	27.89	26.74
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X4X24	-4.30	-4.41	-4.41	-4.42	-4.48	-5.63
T-Value	-23.42	-24.22	-24.52	-24.75	-25.10	-18.22
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X22	75008	89265	51924	61759	62140	64612
T-Value	10.07	11.78	5.79	6.87	6.93	7.22
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
X18X13		20.2	26.2	26.2	20.9	21.3
T-Value		7.97	9.97	10.06	7.21	7.36
P-Value		0.000	0.000	0.000	0.000	0.000
X5X22			16864	16032	15455	15357
T-Value			7.60	7.28	7.02	7.01
P-Value			0.000	0.000	0.000	0.000
X3X22				12435	12669	12729
T-Value				7.05	7.20	7.26
P-Value				0.000	0.000	0.000
X10X13					0.058	0.066
T-Value					4.11	4.67
P-Value					0.000	0.000
X4^2						0.00001
T-Value						4.55
P-Value						0.000
S	18451	18235	18041	17876	17822	17755
R-Sq	99.50	99.52	99.53	99.53	99.54	99.54
R-Sq(adj)	99.50	99.51	99.52	99.53	99.53	99.54
Mallows C-p	274.0	207.8	149.2	99.9	84.7	65.6
Step	19	20	21	22	23	
Constant	10285	10726	10477	7405	7604	
X4X20	18.68	18.01	18.05	18.13	17.34	
T-Value	35.55	29.27	29.34	29.64	22.29	
P-Value	0.000	0.000	0.000	0.000	0.000	

X4X22	11.82	11.35	10.90	11.41	10.90
T-Value	22.00	19.52	17.32	18.06	15.50
P-Value	0.000	0.000	0.000	0.000	0.000
X18X24	1476	1555	1548	1499	1435
T-Value	15.73	15.40	15.33	14.88	13.31
P-Value	0.000	0.000	0.000	0.000	0.000
X4X21	37.00	36.99	36.68	33.55	33.52
T-Value	93.14	93.19	85.34	48.41	48.37
P-Value	0.000	0.000	0.000	0.000	0.000
X17X13	0.0561	0.0552	0.0572	0.0606	0.0616
T-Value	8.01	7.87	8.07	8.57	8.68
P-Value	0.000	0.000	0.000	0.000	0.000
X18X20	419	429	422	371	382
T-Value	23.88	23.61	22.78	18.23	17.93
P-Value	0.000	0.000	0.000	0.000	0.000
X8X24	124.7	139.8	140.0	142.0	131.6
T-Value	16.96	13.58	13.59	13.87	10.94
P-Value	0.000	0.000	0.000	0.000	0.000
X3X20	33251	33309	33280	33284	33335
T-Value	18.87	18.91	18.90	19.02	19.05
P-Value	0.000	0.000	0.000	0.000	0.000
X17X22	0.127	0.128	0.131	0.132	0.132
T-Value	7.56	7.62	7.77	7.87	7.89
P-Value	0.000	0.000	0.000	0.000	0.000
X8X20	213	222	222	225	237
T-Value	15.38	15.28	15.27	15.59	14.79
P-Value	0.000	0.000	0.000	0.000	0.000
X4	4.45	4.19	4.18	4.07	3.78
T-Value	26.45	19.97	19.91	19.45	13.89
P-Value	0.000	0.000	0.000	0.000	0.000
X4X24	-5.62	-6.65	-6.62	-6.51	-8.15
T-Value	-18.22	-11.53	-11.50	-11.36	-7.12
P-Value	0.000	0.000	0.000	0.000	0.000
X22	63493	67893	72424	71861	77052
T-Value	7.10	7.40	7.63	7.62	7.75
P-Value	0.000	0.000	0.000	0.000	0.000
X18X13	19.6	19.7	17.6	16.8	16.8
T-Value	6.65	6.69	5.59	5.35	5.36
P-Value	0.000	0.000	0.000	0.000	0.000
X5X22	16193	16202	14600	15408	15416
T-Value	7.33	7.34	6.17	6.54	6.54
P-Value	0.000	0.000	0.000	0.000	0.000
X3X22	12545	12565	13000	12615	12589
T-Value	7.16	7.17	7.36	7.18	7.17
P-Value	0.000	0.000	0.000	0.000	0.000
X10X13	0.052	0.055	0.059	0.051	0.047
T-Value	3.45	3.64	3.90	3.38	3.03
P-Value	0.001	0.000	0.000	0.001	0.002

X4^2	0.00001	0.00004	0.00004	0.00004	0.00007
T-Value	4.67	2.98	2.98	3.04	3.10
P-Value	0.000	0.003	0.003	0.002	0.002
X5X13	-658	-714	-775	-872	-840
T-Value	-2.89	-3.11	-3.35	-3.78	-3.63
P-Value	0.004	0.002	0.001	0.000	0.000
X4^3		-0.00000	-0.00000	-0.00000	-0.00000
T-Value		-2.10	-2.10	-2.19	-2.68
P-Value		0.036	0.036	0.029	0.007
X18			5.5	60.8	62.2
T-Value			1.86	6.05	6.16
P-Value			0.063	0.000	0.000
X18X22				-60	-61
T-Value				-5.75	-5.88
P-Value				0.000	0.000
X24					48484
T-Value					1.65
P-Value					0.100
S	17731	17719	17711	17603	17597
R-Sq	99.54	99.54	99.54	99.55	99.55
R-Sq(adj)	99.54	99.54	99.54	99.55	99.55
Mallows C-p	59.1	56.7	55.2	24.1	23.4

Fe de erratas: Corrección en página 76.

A pesar de que los residuos no son normales, lo más importante de este diagrama es representado por kurtosis, la cual es una medida estadística que describe el apuntamiento o ensanchamiento de una cierta distribución con respecto a una distribución normal. En este caso se presenta una kurtosis positiva excesiva, siendo esto es una característica deseable del modelo dado a que la mayor parte de los residuos se distribuyen alrededor de la media igual a cero.

