

*ESTUDIO DE LAS VENTAS DE LAS
DELEGACIONES DEL GRUPO
DULCESOL EN ESPAÑA*

MEMORIA PRESENTADA POR:

Alba Alvado Fornás

GRADO EN ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

Convocatoria de defensa: Mayo de 2019

A mis padres, Gabriel Alvado y Amalia Fornás, gracias por ayudarme a crecer tanto personal como profesionalmente, brindándome la educación necesaria para ello y transmitiéndome los valores que han hecho, que hoy sea la persona que soy, y con ello, que haya podido llegar hasta aquí.

A Fran Javier Bravo, quien ha sido mi pareja y mi mejor amigo. Nunca podré agradecerte suficiente todo lo que has hecho por mí. Gracias por creer en mi cuando ni yo misma lo hacía. Por tus palabras de ánimo cada segundo que no tenía fuerzas para seguir. Por tu esfuerzo y dedicación, tu lucha constante y sobre todo por tu apoyo. Por hacer todo lo posible tan sólo para que yo pudiese cumplir mis sueños.

A vosotros, que cambiasteis el argumento de mi vida. Os quiero ∞.

ÍNDICE

1. INTRODUCCIÓN	7
2. METODOLOGÍA Y TÉCNICAS	8
2.1 METODOLOGÍA	8
2.2 PRIMERA ESTIMACIÓN DEL MODELO	11
2.3 TÉCNICAS UTILIZADAS EN EL DESARROLLO DEL MODELO	12
2.3.1 ANÁLISIS UNIVARIANTE.....	13
2.3.2 ANÁLISIS BIVARIANTE	16
2.3.3 ANÁLISIS MULTIVARIANTE	19
2.3.4 FORMACIÓN REQUERIDA.....	27
2.3.5 DATOS Y SOFTWARE	28
3. CÁLCULOS, RESULTADOS Y EXPLOTACIÓN	30
3.1 ANÁLISIS UNIVARIANTE	30
3.1.1 GRÁFICO DE CAJAS Y BIGOTES	30
3.1.2 CONCLUSIÓN ANÁLISIS UNIVARIANTE.....	44
3.1.3 INTERPRETACIÓN.....	45
3.2 ANÁLISIS BIVARIANTE	46
3.2.1 GRÁFICO X E Y, Y MODELO DE REGRESIÓN SIMPLE	46
3.2.2 MULTICOLINEALIDAD	60
3.2.3 MODELO FINAL.....	83
3.3 ANÁLISIS MULTIVARIADO	84
3.3.1 MODELO TEÓRICO PROPUESTO	84
3.3.2 INTERPRETACIÓN DE LOS PARÁMETROS	85
3.3.3 MODELO AJUSTADO.....	86
3.3.4 VALIDACIÓN DEL MODELO.....	87
3.3.5 NORMALIDAD DE RESIDUOS.....	90
3.4 PUNTOS ANÓMALOS Y PUNTOS INFLUYENTES	96
3.4.1 PUNTOS ANÓMALOS	96
3.4.2 PUNTOS INFLUYENTES	96
4. TABLA RESUMEN	98
5. SOLUCIÓN DEL MODELO MATEMÁTICO	100
6. PREDICCIONES	101
7. CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	108
7.1 CONCLUSIONES	108
7.2 FUTURAS LÍNEAS DE INVESTIGACIÓN	109
8. BIBLIOGRAFÍA	110

TABLAS

Tabla 1 - Datos obtenidos de datos de INE y de la base de datos de Dulcesol.....	11
Tabla 2 - Matriz de correlación	17
Tabla 3 - Matriz inversa de correlación.....	17
Tabla 4 - Tabla P-Valores de los diferentes parámetros.	19
Tabla 5 - Resumen estadístico para venta propia.....	30
Tabla 6 - Resumen estadístico para márgenes.	31
Tabla 7 - Resumen estadístico para número de comerciales.	32
Tabla 8 - Resumen estadístico para incentivos de contado.....	33
Tabla 9 - Resumen estadístico para incentivos de crédito.	34
Tabla 10 - Resumen estadístico para frecuencia visitas.	35
Tabla 11 - Resumen estadístico para ticket medio por visita.	36
Tabla 12 - Ticket medio por visita. Gráfico de caja y bigotes.....	36
Tabla 13 - Resumen estadístico para margen total medio diario.	37
Tabla 14 - Resumen estadístico para clientes.....	38
Tabla 15 - Resumen estadístico para regalos.....	39
Tabla 16 - Resumen estadístico para población.	40
Tabla 17 - Resumen estadístico para número de mujeres.	41
Tabla 18 - Resumen estadístico para IPC alimentos y bebidas no alcohólicas.	42
Tabla 19 - Resumen estadístico para tasa de riesgo de pobreza.	43
Tabla 20 – Puntos anómalos detectados en las variables.	44
Tabla 21 - Puntos anómalos encontrados en las delegaciones.	44
Tabla 22 - Relación entre las variables.....	59
Tabla 23 - Matriz de correlación.	60
Tabla 24 - Matriz inversa de correlación.....	61
Tabla 25 - Valores máximo y mínimo.....	62
Tabla 26 - Matriz de correlación eliminando la variable: Número de mujeres.	63
Tabla 27 - Matriz inversa de correlación eliminando la variable: Número de mujeres.....	64
Tabla 28 - Valores máximo y mínimo.....	65
Tabla 29 - Matriz de correlación eliminando las variables: Mujeres y ticket medio de visita.	67
Tabla 30 - Matriz inversa de correlación eliminando las variables: Mujeres y ticket medio de visita.	68
Tabla 31 - Valores máximos y mínimos.....	69
Tabla 32 - Matriz de correlación eliminando las variables: Mujeres, ticket medio por visita y márgenes.	70
Tabla 33 - Matriz inversa de correlación eliminando las variables: Mujeres, ticket medio por visita y márgenes.	71
Tabla 34 - Gráfica de sidementación.	72
Tabla 35 - Valores máximos y mínimos.....	72
Tabla 36 - Matriz de correlación eliminando las variables: Mujeres, ticket medio por visita, márgenes y número de comerciales.	73
Tabla 37 - Matriz inversa de correlación eliminando las variables: Mujeres, ticket medio por visita, márgenes y número de comerciales.	74
Tabla 38 - Valores máximos y mínimos.....	75
Tabla 39 - Matriz de correlación eliminando las variables: Mujeres, ticket medio por visita, márgenes, número de comerciales y clientes.	76

Tabla 40 - Matriz inversa de correlación eliminando las variables: Mujeres, ticket medio por visita, márgenes, número de comerciales y clientes.	77
Tabla 41 - Valores máximos y mínimos.....	78
Tabla 42 - Matriz de correlación eliminando las variables: Mujeres, ticket medio por visita, márgenes, número de comerciales, clientes y regalos.....	80
Tabla 43 - Matriz inversa de correlación eliminando las variables: Mujeres, ticket medio por visita, márgenes, número de comerciales, clientes y regalos.	81
Tabla 44 - Valores máximos y mínimos.....	82
Tabla 45 - Modelo final	83
Tabla 46 - Significatividad del parámetro.	87
Tabla 47 - Significatividad del modelo.	89
Tabla 48 - Test de normalidad.	91
Tabla 49 - Valor P parámetros.	91
Tabla 50 - Valor P modelo.....	91
Tabla 51 - Residuos atípicos.....	96
Tabla 52 - Puntos influyentes.....	96
Tabla 53 - Puntos influyentes a priori y a posteriori.....	96
Tabla 54 - Puntos influyentes a priori.....	97
Tabla 55 - Puntos influyentes a posteriori.....	97
Tabla 56 - Tabla resumen proceso.....	98
Tabla 57 - Residuo estudentizado.....	99
Tabla 58 - Valor-P parámetros.....	100
Tabla 59 - Valor-P modelo.....	100
Tabla 60 – Tabla explicativa proceso modificación de variables.....	101
Tabla 61 - Variables y delegaciones sin datos alterados.....	103
Tabla 62 - Variables modificadas según el resultado del proyecto econométrico para su posterior predicción.....	104
Tabla 63 - Sumatorio beneficio final inversión en incentivos crédito.....	104
Tabla 64 - Sumatorio beneficio final inversión en margen total medio.....	104
Tabla 65 - Sumatorio beneficio final inversión en regalos.....	105
Tabla 66 - Sumatorio beneficios finales.....	106
Tabla 67 - Reducción costes frecuencia de visitas.....	106
Tabla 68 - Predicciones puntuales y por intervalo.....	106
Tabla 69 - Predicción frecuencia de visitas.....	107
Tabla 70 - Predicción margen total medio diario.....	107
Tabla 71 - Predicción regalos.....	107

ILUSTRACIONES

Ilustración 1 - Técnicas utilizadas en el desarrollo del modelo.	12
Ilustración 2 - Cuadro resumen explicativo análisis univariante.	13
Ilustración 3 - Cuadro resumen explicativo análisis univariante.	14
Ilustración 4 - Gráfico de cajas y bigotes Fuente: Datavizcatalogue.....	15
Ilustración 5 - Contraste B_0	19
Ilustración 6 - Significatividad del modelo.	19
Ilustración 7 - Gráfico de probabilidad normal con 95% límites.	20
Ilustración 8- Distribuciones. Elaboración Daniela Ramello.	21
Ilustración 9 - Contraste de hipótesis.	21
Ilustración 10 - Contraste de hipótesis. Heterocedasticidad.....	22
Ilustración 11 - Autocorrelación de primer orden negativa	23
Ilustración 12 - Autocorrelación de primer orden positiva.....	23
Ilustración 13- FAS	24
Ilustración 14 - FAP	24
Ilustración 15 - Autocorrelación	25
Ilustración 16 - Contraste de hipótesis. Autocorrelación.	25
Ilustración 17 - Puntos influyentes	26
Ilustración 18 - Venta propia. Gráfico cajas y bigotes.	30
Ilustración 19 - Márgenes. Gráfico de cajas y bigotes.	31
Ilustración 20 - Número de comerciales. Gráfico de cajas y bigotes.....	32
Ilustración 21 - Incentivos de contado. Gráfico de cajas y bigotes.....	33
Ilustración 22 - Incentivos de crédito. Gráfico de caja y bigotes.	34
Ilustración 23 - Frecuencia visitas. Gráfico de caja y bigotes.....	35
Ilustración 24 - Margen total medio diario. Gráfico de caja y bigotes.	37
Ilustración 25 - Clientes. Gráfico de caja y bigotes.	38
Ilustración 26 - Regalos. Gráfico de caja y bigotes.	39
Ilustración 27 - Población. Gráfico de caja y bigotes.	40
Ilustración 28 - Número de mujeres. Gráfico de caja y bigotes.....	41
Ilustración 29 - IPC alimentos y bebidas no alcohólicas. Gráfico de caja y bigotes.....	42
Ilustración 30 - Tasa de riesgo de pobreza. Gráfico de caja y bigotes.....	43
Ilustración 31 - Márgenes vs Venta propia.	46
Ilustración 32 - Número de comerciales vs Venta propia.....	47
Ilustración 33 - Incentivos contado vs Venta propia.....	48
Ilustración 34 - Incentivos crédito vs Venta propia.	49
Ilustración 35 - Frecuencia visitas vs Venta propia.....	50
Ilustración 36 - Ticket medio por visita vs Venta propia.....	51
Ilustración 37 - Margen total medio diario vs Venta propia.....	52
Ilustración 38 - Clientes vs Venta propia.	53
Ilustración 39 - Regalos vs Venta propia.....	54
Ilustración 40 - Población vs Venta propia.	55
Ilustración 41 - Número de mujeres vs Venta propia.....	56
Ilustración 42 - IPC Alimentos y bebidas no alcohólicas vs Venta propia.....	57
Ilustración 43 - Tasa de riesgo de pobreza vs Venta propia.	58
Ilustración 44 – Gráfica de sedimentación.	62
Ilustración 45 - Gráfica de sedimentación.	65



Ilustración 46 - Gráfica de sedimentación.	69
Ilustración 47 - Gráfica de sedimentación.	75
Ilustración 48 - Gráfica de sedimentación.	78
Ilustración 49 - Gráfica de sedimentación.	82
Ilustración 50 – Regresión simple nuevo modelo.	86
Ilustración 51 - Gráfico probabilístico.	90
Ilustración 52 - FAS. Autocorrelación.	94
Ilustración 53 - FAP. Autocorrelación.	94
Ilustración 54 - Escala estadístico de Durbin - Watson.	95
Ilustración 55 - Observado vs predicho.	102
Ilustración 56 - Explicación aumento/reducción variables.	103

1. INTRODUCCIÓN

El **fundamento** de este proyecto final de grado, reside en demostrar estadísticamente porqué se factura una mayor cantidad económica en una delegación del grupo Dulcesol o en otra. Esto se comprobará realizando el estudio de un modelo econométrico que contendrá las variables externas e internas a la empresa asociadas al total de las ventas. Durante este proceso se podrán conocer las variables finales que provocan este aumento en la facturación de la compañía.

La **motivación** por la cual se realiza este proyecto es, porque el alumno ha estado realizando prácticas profesionales en la empresa, tanto curriculares como extracurriculares. El estudiante, propuso elaborar un proyecto econométrico, y la empresa quiso colaborar en dicho estudio, ofreciéndose como candidata para este. Para ello, el grupo, le ofreció la información necesaria para poder llevar a cabo este plan.

La cadena de suministros alimentarios y bebidas no alcohólicas, actualmente está valorando en adquirir nuevas delegaciones para continuar expandiéndose nacional e internacionalmente. Por lo que la **finalidad** de este proyecto econométrico, se encuentra en incrementar la facturación que registra la cadena Dulcesol actualmente además de obtener una reducción en costes, con el fin de destinar estos beneficios a futuras inversiones.

2. METODOLOGÍA Y TÉCNICAS

2.1 METODOLOGÍA

El primer paso para la ejecución de este proyecto será el de exponer el significado de cada uno de los componentes que se tratará en el modelo, con el fin de que se comprenda, mediante un razonamiento lógico y definido, porqué se han sugerido estas para explicar la variable regresada, dicho de otro modo, las ventas de la cadena Dulcesol. A posteriori, se perfeccionará una tabla, donde se podrán contemplar los valores de cada una de las delegaciones para cada uno de los factores seleccionados.

El segundo paso, tratará de indicar las técnicas que se emplearán posteriormente en el apartado de cálculos, resultados y explotación del modelo para poder desarrollar esta investigación. En este apartado, se podrán percibir los diferentes análisis que se deberán efectuar para obtener las predicciones y solución del modelo.

Como ya se ha indicado previamente, en este apartado se establecerán las variables con las cuales se definirá la variable Y, en otras palabras, las ventas de la cadena Dulcesol.

- **Márgenes:** Dicho regresor, mostrará cómo los márgenes, dependiendo de si son más elevados o más modestos, afectan directamente sobre la cantidad de ventas del grupo Dulcesol. A esta variable explicativa se le denominará **X1**.
- **Nº Comerciales:** El número de comerciales podría ser otra variable que repercutiese en que la cantidad de ventas aumentase, dado que dependiendo del número de comerciales que existan, la empresa tiene más posibilidades de abarcar a más municipios y con ello, a más clientes en menor tiempo. A este regresor se le nombrará **X2**.
- **Incentivos contado:** Los incentivos pueden ser claves para los comerciales a la hora de realizar las ventas, puesto que estarán más motivados a generar un mayor beneficio a la empresa a cambio de una comisión por dicho negocio.
Existen dos tipos de incentivos: los incentivos que se pagan por venta por contado y, los incentivos que se abonan por cada venta de crédito. Estos son diferentes. Los incentivos de crédito tendrán un valor más bajo puesto que el grupo, asume el riesgo de que los clientes no paguen a corto plazo (suelen darles un plazo de 90 días para presentar el pago), caso que no ocurre con los incentivos al contado, por eso son mayores los alicientes económicos de los segundos. A los incentivos se le otorgará la variable explicativa **X3**.
- **Incentivos crédito:** Esta variable ya se ha explicado anteriormente. Se le llamará **X4**.
- **Frecuencia visitas:** El parámetro indicado anteriormente, indica cuántas veces acude cada comercial al establecimiento donde se encuentra el cliente que va a realizar la compra. Este regresor, podría influir en la variable endógena, dado que se podría generar más venta si acuden un mayor número de ocasiones. A esta variable dependiente, se le denominará **X5**.
- **Ticket medio por visita:** Esta variable definirá cuánta es la repercusión de cada comercial a la hora de acudir a los establecimientos donde Dulcesol vende sus productos. Se podrá observar qué cantidad sobre el total de ventas consiguen en cada una de sus ventas. Al ticket medio por visita se le asignará **X6**.
- **Margen total medio diario:** Este elemento, indica cuál es el margen que hace Dulcesol de forma directa con sus proveedores en sus delegaciones diariamente. A esta variable se le otorgará el nombre de **X7**.

- **Nº de clientes:** Se pretende estimar con este regresor, si el que las delegaciones posean un gran número de clientes influye en mayor parte a que las ventas aumenten. El número de clientes, será la variable explicativa **X8**.
- **Regalos:** Mediante dicha variable independiente, se pretende comprobar, si una motivación extra, como son los regalos o promociones, puede repercutir a que los clientes obtengan más bienes con tal de beneficiarse de esta ventaja. A esta variable se le llamará **X9**.
- **Población:** En este caso, se podrá observar si se obtienen mayores ventas de la cadena Dulcesol cuanto mayor es la población. Podría suceder que este factor no fuese del todo relativo, sino que otros componentes contribuyeran más al aumento de las ventas, como por ejemplo la habilidad de los comerciales a la hora de comercializar con los productos del grupo. Pero esta es otra variable en la cual se indagará más adelante. La población recibirá **X10** como nombre.
- **Número de mujeres:** Según indican diversos estudios, como, por ejemplo, la investigación elaborada por la Universidad Nacional de Lanús (UNLa), las mujeres consumen más dulces que los hombres y lo incluyen en sus dietas. Por esta razón, se ha decidido estudiar la variable número de mujeres independientemente del regresor población. El número de mujeres recibirá el nombre de **X11**.
- **IPC de alimentos y bebidas no alcohólicas:** El IPC mide la evolución de los precios de la cesta de bienes y servicios de los consumidores en España. Es un indicador clave para conocer la inflación de la economía española y poder ajustar la subida de salarios a este incremento de precios de consumo. Este factor estudia, si los precios se incrementan o disminuyen por el paso del tiempo, de un mes a otro o, de un año a otro. El IPC en este caso, será una variable de gran importancia, la cual se deberá analizar junto a la variable explicar, las ventas.

Dentro del estudio del IPC general, se investigan componentes como la enseñanza, la medicina, transporte, vivienda... Estos factores no están relacionados con los productos que comercializa Dulcesol, así pues, se ha decidido focalizar en el IPC de alimentos y bebidas no alcohólicas de España. De esta forma se conseguirán unas estadísticas lo más veraces posibles y no se estudiará el conjunto del IPC sesgando de tal forma una información que podría resultar muy positiva para este estudio.

Si el IPC es alto, significa que existirá una pérdida importante de poder adquisitivo para adquirir los bienes y servicios del grupo Dulcesol, por lo que esto influiría negativamente a las ventas, estas se reducirían respectivamente. Sucedería justo al contrario si el IPC es menor, en este caso, las ventas serían mayores puesto que existiría un mayor poder adquisitivo sobre la renta. A los incentivos se le otorgará la variable explicativa **X12**.

- **Tasa de riesgo de pobreza:** Se ha decidido estudiar esta variable, puesto que, a mayor riesgo de pobreza, menor capacidad de compra tendrá la población española y, a menor riesgo de pobreza, mayor poder de adquisición poseerán. Esto influirá en la sensibilidad de la variable ventas. A esta variable dependiente, se le denominará **X13**.

Después de presentar las variables explicativas seleccionadas para la formulación del modelo econométrico, se expondrán los datos necesarios para el estudio de la variable endógena incluida esta, en otros términos, las ventas de Dulcesol. En este caso, **se estudiará el año 2017, desde el periodo enero a septiembre.**

Se ha apostado por este lapso, es decir, de enero a septiembre, teniendo en cuenta la información que otorgaba la empresa respecto a ciertas variables, puesto que **algunos factores para tener en cuenta solo son computados de forma trimestral o bimestral** y estos se consideraban de vital importancia para la investigación. Los componentes anteriormente comentados, afectados a este cómputo son: frecuencia de las visitas, ticket medio por visita y margen total diario.

Otra dificultad añadida a la realización de este modelo, y que por tanto ha limitado la selección de datos, ha sido la obtención de datos mediante los diferentes organismos estadísticos en cuanto a municipios se refiere. Las cadenas pertenecientes al grupo Dulcesol, efectúan sus ventas en distintas localidades y en algunos casos, diferentes delegaciones pueden estar vendiendo en una misma provincia. Por lo tanto, no se han podido seleccionar directamente datos de las diferentes comunidades autónomas, lo cual resultaría mucho más tratable y más contingente que encontrar la información por territorios.

Ambos agentes en conjunto han condicionado la recopilación de cifras para el modelo, puesto que se han excluido más variables de las que en realidad se deseaban.

Algunos de los componentes que se han querido tener en cuenta para la realización de este análisis y no se han podido estudiar dado que su periodo era anual, son los siguientes:

- **Dietas o régimen específico**
- **Número de empresas de alimentación**
- **PIB per cápita**
- **Número de hogares**
- **Tasa de paro**
- **Tasa de ocupación**

A continuación, como se ha dicho previamente, se mostrarán las variables escogidas en una tabla con las diferentes cifras correspondientes a cada una de ellas.

2.2 PRIMERA ESTIMACIÓN DEL MODELO

2018														
ENERO A SEPTIEMBRE														
DELEGACIÓN	NIVEL INTERNO										NIVEL EXTERNO			
	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
	VENTA PROPIA	MÁRGENES	Nº COMERCIALES	INCENTIVOS CONTADO	INCENTIVOS CRÉDITO	FRECUENCIA VISITAS	TICKET MEDIO POR VISITA	MARGEN TOTAL MEDIO DIARIO	CLIENTES	REGALOS	POBLACIÓN	NÚMERO DE MUJERES	IPC ALIM Y BEBI NO ALC	TASA DE RIESGO DE POBREZA
ARADUL	333351	95553,1	12	1,7	0,64	11,55	28,09	8,16	339	6573,72	925691	470906	2,5	11,3
ASTURSOL	810848	215019	14	1,7	1,06	9,24	26,13	5,99	596	1983,6	875889	458234	2,4	10,7
DOCES LUCENSES	1284000	399354	14	1,7	0,85	9,77	33,05	10,26	1148	3915,93	375333	192256	2,6	15,9
DULCADIS	2160360	602766	25	1,91	1,06	8,8	32,95	9,56	911	23017,4	998977	501428	2,7	23,9
DULCOSOL	5917520	1356950	32	3,83	1,7	9,29	53,24	15,6	2065	42753,2	1559626	791449	2,2	25
DULGALICIA	1074190	329614	17	1,06	1,06	11,78	32,31	9,41	1063	11446,8	902641	495536	2,6	15,9
DULGANDIA	2939750	756512	37	3,83	1,7	10,73	52,45	14,76	1385	36124,4	1737575	878807	2,6	21,8
DULMADRID	3997800	1201130	71	1,91	1,28	12,95	39,94	10,65	3491	63893,5	5531106	2882913	2,3	14,4
EUSKASOL	3005750	928854	42	1,7	0,85	11,63	42	13,34	1618	29696,3	1558090	797712	2,6	11
GADESDUL	2124480	560786	20	1,7	0,85	8,39	36,07	9,8	1276	26770	203304	104967	2,5	26,4
HERMANOS MUÑOZ	798290	226055	13	1,49	0,64	11,5	39,84	11,9	380	5519,83	672939	337270	2,4	12,8
HISPADUL	6584020	2024420	50	2,34	2,34	9,01	47,7	13,63	3182	40614,7	2080061	1059412	2,5	26,4
LEVANTINA	1262400	348255	15	1,7	0,85	11,35	45,36	10,48	791	9220,01	1679123	857839	2,6	21,8
TINERFEÑA	1250060	354989	19	1,7	0,85	9,93	38,59	9,89	889	12796,4	597518	310688	2	25,9
VEGODISA	1978430	606919	20	1,7	0,85	10,55	57,83	15,88	557	10815,2	935408	468999	2	25,9
VIGUESA	1163800	281090	14	1,7	0,85	8,94	30,1	8,56	671	2104,14	768060	397043	2,6	15,9

Tabla 1 - Datos obtenidos de datos de INE y de la base de datos de Dulcesol.

2.3 TÉCNICAS UTILIZADAS EN EL DESARROLLO DEL MODELO

A continuación, mediante una ilustración adjunta, se indicarán cuáles son las fases que se seguirán para averiguar la solución del modelo econométrico. Además, se indicará en qué consiste cada proceso de una forma más detallada. De tal forma, se podrá comprender con una mayor precisión cada parte de este proceso econométrico.

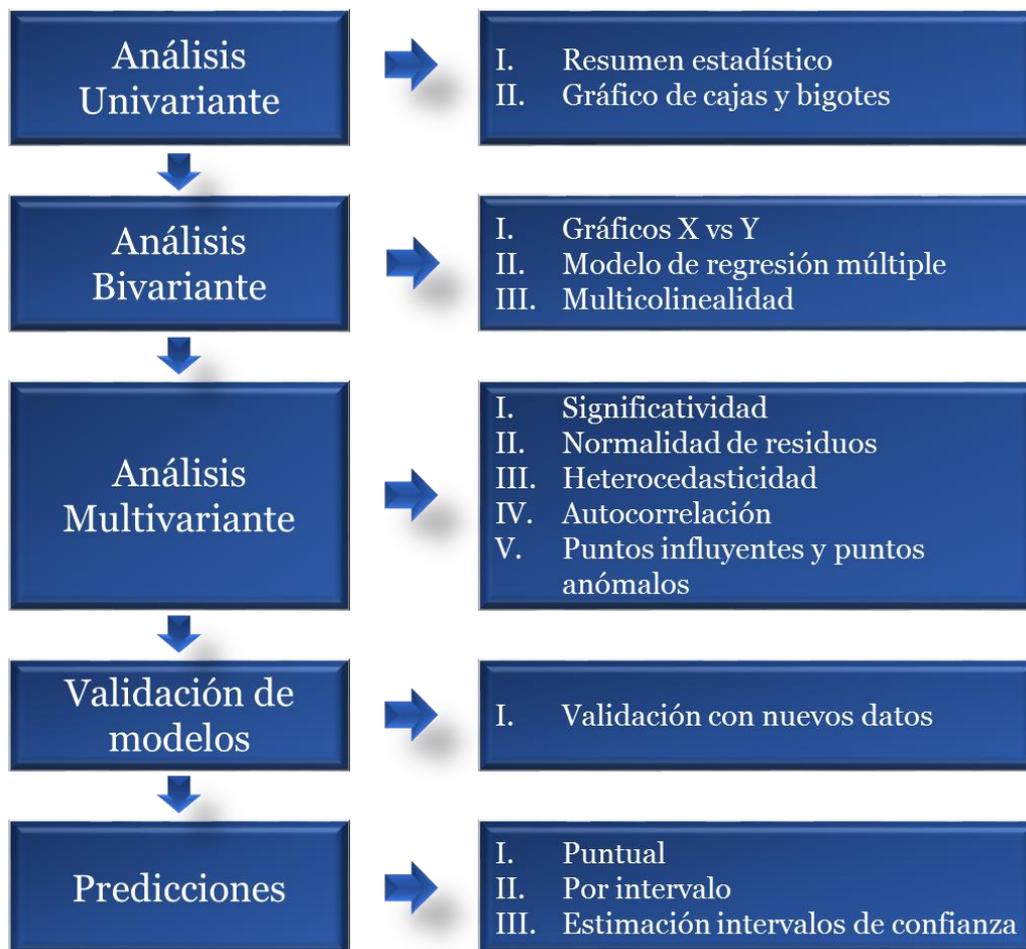


Ilustración 1 - Técnicas utilizadas en el desarrollo del modelo.

2.3.1 ANÁLISIS UNIVARIANTE

En la parte del análisis univariante, más en concreto en el resumen estadístico que posteriormente se analizará, se utilizarán algunos tecnicismos estadísticos como son los nombrados a continuación. En las dos tablas que se mostrarán ulteriormente, se nombrará cada medida, la expresión y definición, con el fin de aportar una mayor comprensión en la futura sección de este proyecto en la parte de cálculos, resultados y explotación.

RESUMEN ESTADÍSTICO	EXPRESIÓN	DEFINICIÓN
RECUENTO	N	Número total de delegaciones a estudiar.
PROMEDIO	$\bar{\mu} = \frac{x_1 + x_2 + x_3 \dots x_N}{N} = \frac{\sum_{i=1}^N x_i}{N}$	Suma de todos las cantidades observadas dividida entre el número total de datos.
MEDIANA	$\text{Posición de la mediana (par)} = \frac{N + 1}{2}$ $\text{Posición de la mediana (impar)} = \frac{\frac{N}{2} + \frac{N}{2}}{2}$	<p>Sistematización de los valores de menor a mayor, con el fin de elegir la cifra central del resultado de dicha clasificación.</p> <p>Seguir el procedimiento indicado en la posición de la mediana (par). En este caso, serán dos los valores que sean centrales, por lo que se sumarán ambos y se dividirán entre dos.</p>
DESVIACIÓN ESTÁNDAR	$s = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$	La desviación típica mostrará la dispersión de los puntos en comparación con la media.
MÍNIMO	-	Se selecciona el valor más pequeño de cada variable.
MÁXIMO	-	Se escoge el valor más grande de cada variable.
CUARTIL INFERIOR	$Q_1 = x(0,25*(n+1))$	El cuartil inferior o primer cuartil deja por debajo el 25% de los valores de la muestra.
CUARTIL SUPERIOR	$Q_3 = x(0,75*(n+1))$	El cuartil superior o tercer cuartil deja por debajo es 75% de los valores de la muestra.

Ilustración 2 - Cuadro resumen explicativo análisis univariante.

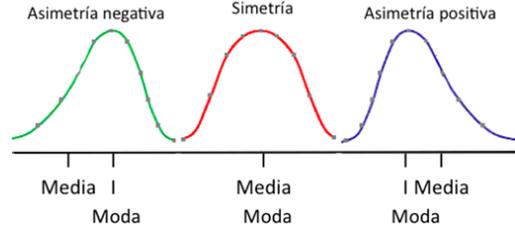
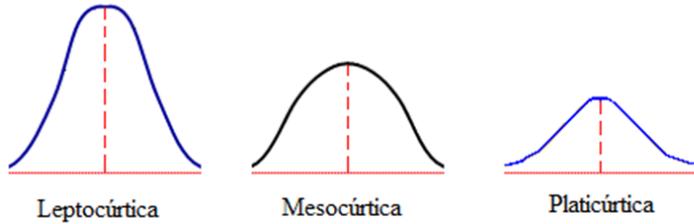
RESUMEN ESTADÍSTICO	EXPRESIÓN	DEFINICIÓN
SESGO	$CA_F = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{NS^3}$	<p>Es la diferencia entre la esperanza matemática del modelo y el valor numérico del parámetro que estima. Consiste en seleccionar de una forma errónea los datos. Existen tres tipos de asimetría:</p> 
CURTOSIS	$Curtosis = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{NS^4}$	<p>La curtosis muestra el número de datos que tienen cierta proximidad a la mediana. Cuantos más valores existan, más notable será la forma de la curva. Existen tres tipos de curtosis. La curtosis leptocúrtica, se identifica con la curtosis negativa, la mesocúrtica con la curtosis nula y la platicúrtica con la curtosis positiva.</p> 

Ilustración 3 - Cuadro resumen explicativo análisis univariante.

GRÁFICO CAJAS Y BIGOTES

Esta gráfica está representada por una caja rectangular y una línea horizontal. La caja rectangular ofrecerá los valores de los cuartiles primero, segundo y tercero. La recta (llamada bigotes) que está expuesta a los lados de la caja, presenta el recorrido intercuartílico, es decir, es la diferencia entre el tercer y el primer cuartil de la muestra. En la línea vertical está posicionada la mediana, esta está definida por el cuartil número dos.

Los datos que se sitúen fuera de este rectángulo, es decir, en los bigotes del gráfico, serán los valores anómalos o atípicos. Estos valores, son los que comportarán diferente a los demás. Este análisis los identificará individualmente y con detenimiento, dado que si existe un valor que se repite en un número elevado de ocasiones, se tendrá que eliminar para obtener un mejor modelo.

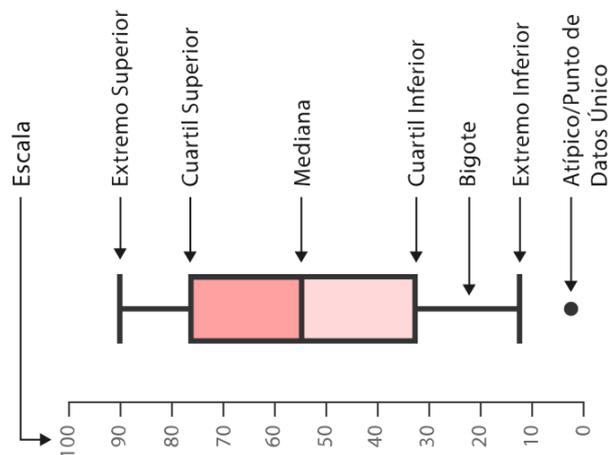


Ilustración 4 - Gráfico de cajas y bigotes
 Fuente: Datavizcatalogue [1].

DEFINICIÓN DE LOS CUARTILES

- **PRIMER CUARTIL:** Deja por **debajo el 25%** de los valores de la muestra.

$$Q_1 = x(0,25 \cdot (n+1))$$

- **SEGUNDO CUARTIL:** Mediana, puesto que deja **por debajo el 50%** de los valores de la muestra.

$$Q_2 = Me = x(0,5 \cdot (n+1))$$

- **TERCER CUARTIL:** Deja **por debajo el 75%** de los valores de la muestra.

$$Q_3 = x(0,75 \cdot (n+1))$$

2.3.2 ANÁLISIS BIVARIANTE

El análisis bivalente pretende mediante una **comparación** de la variable (**X**) **junto** a la variable (**Y**), **mostrar** un **comportamiento** conjunto en una muestra de **N** (en este caso de delegaciones de Dulcesol).

GRÁFICOS X VS Y

Este gráfico compara la variable explicativa elegida (x) con la variable explicada (y) con el fin de probar si existe una relación entre ambas.

En este caso se presentará un **gráfico de dispersión** donde los datos se representarán por puntos delimitados por los ejes X e Y. **Cuanto más juntos estén los puntos, más relacionados se encontrarán.**

MODELO DE REGRESIÓN SIMPLE

Este modelo, **relaciona** la variable predicha (**X**) **con** la variable predictora (**Y**). Esta regresión además de esto último, puede proporcionar información sobre cuáles pueden ser las **causas** de que la **variable predictora varíe** de una forma u otra.

La **función de regresión** que relaciona la variable endógena (Y) con las variables exógenas (X) es la siguiente:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + U$$

B₀: Es el término independiente. Es el valor que se espera cuando todas las variables independientes (x) valen 0.

B₁: Expresa el cambio en la variable dependiente (y) por cada cambio unitario en X1 cuando todas las demás variables independientes permanecen constantes.

B₂: Expresa el cambio en la variable dependiente (y) por cada cambio unitario en X2 cuando todas las demás variables independientes permanecen constantes.

B_nX_n: Expresa el cambio en la variable dependiente (y) por cada cambio unitario en Xn cuando todas las demás variables independientes permanecen constantes.

U = Representa las variables explicativas que no se han tenido en cuenta en el modelo.

Estudio de la **MULTICOLINEALIDAD**

En este caso cuando el valor se acerque más a uno, más relación tendrán las variables.

Los procedimientos que se pueden utilizar para comprender si existe multicolinealidad en el modelo que se va a estudiar, son tres y, son los siguientes:

- La primera de ellas expondrá a través del **ANÁLISIS DE LA MATRIZ DE CORRELACIÓN**, si existe una **relación entr e las variables**. Para ello, se formará una matriz donde cada una de las variables exógenas esté relacionada independientemente con cada variable explicativa a estudiar en el modelo. Su referencia será el coeficiente de correlación Rij.

	INCENTIVOS CONTADO	INCENTIVOS CRÉDITO	FRECUENCIA VISITAS	MARGEN TOTAL MEDIO DIARIO	REGALOS	IPC ALIM Y BEBI NO ALC	TASA DE RIESGO DE POBREZA
INCENTIVOS CONTADO	1	0,674525	-0,206089	0,599845	0,542071	-0,087815	0,378377
INCENTIVOS CRÉDITO	0,674525	1	-0,233171	0,471243	0,658557	0,0151081	0,424205
FRECUENCIA VISITAS	-0,206089	-0,233171	1	0,0950571	0,172496	-0,0473155	-0,514385
MARGEN TOTAL MEDIO DIARIO	0,599845	0,471243	0,0950571	1	0,479811	-0,31576	0,469954
REGALOS	0,542071	0,658557	0,172496	0,479811	1	-0,10083	0,271856
IPC ALIM Y BEBI NO ALC	-0,087815	0,0151081	-0,0473155	-0,31576	-0,10083	1	-0,32633
TASA DE RIESGO DE POBREZA	0,378377	0,424205	-0,514385	0,469954	0,271856	-0,32633	1

Tabla 2 - Matriz de correlación

Cuando **R = 1** se hablará de una **correlación perfecta entre las variables**. En este caso, se puede observar que la diagonal posee este valor, puesto que se relacionan las mismas variables. En este caso tan sólo se estudiarán los valores de debajo y encima de la diagonal puesto que se relacionan las variables exógenas.

- El segundo método para confirmar si existe multicolinealidad, se trata del **ANÁLISIS DE LA MATRIZ INVERSA DE CORRELACIÓN**. Este análisis mostrará si existe **una relación entre una variable exógena y el resto de ellas**. Esta técnica se realizará mediante Excel. Este estará definido por el coeficiente de correlación Rii-1.

INCENTIVOS CONTADO	2,527934115	-0,829373252	0,791318616	-1,17997582	-0,530571181	0,010530209	0,504564217
INCENTIVOS CRÉDITO	-0,829373252	2,82653579	0,649576319	-0,26693463	-1,38702847	-0,363498478	-0,167183988
FRECUENCIA VISITAS	0,791318616	0,649576319	2,467898592	-1,15745653	-1,145696106	0,225206524	1,623387062
MARGEN TOTAL MEDIO DIARIO	-1,179975817	-0,266934626	-1,15745653	2,5004249	0,159572631	0,270853884	-1,165745278
REGALOS	-0,530571181	-1,38702847	-1,14569611	0,15957263	2,471382963	0,04612658	-0,531988143
IPC ALIM Y BEBI NO ALC	0,010530209	-0,363498478	0,225206524	0,27085388	0,04612658	1,285314679	0,545664426
TASA DE RIESGO DE POBREZA	0,504564217	-0,167183988	1,623387062	-1,16574528	-0,531988143	0,545664426	2,58558824

Tabla 3 - Matriz inversa de correlación

En este caso, al estudiarse una variable exógena y el resto de ellas, tan sólo se estudiarán los valores de la diagonal. Puesto que recoge el valor de la variable exógena a analizar, el resto y las compara.

- Y, por último, el sistema denominado **ÍNDICE DE ACONDICIONAMIENTO**, calibra la **relación de todas las variables X al mismo tiempo**. Dicho índice, se fundamentará en los autovalores de autocorrelación. Se utilizará la fórmula siguiente:

$$I.C. = \sqrt{\frac{\text{Autovalor máximo}}{\text{Autovalor mínimo}}}$$

Estas son las **NORMAS** para saber si el modelo posee multicolinealidad en estos tres métodos:

- **ANÁLISIS DE LA MATRIZ DE CORRELACIÓN:** La matriz del modelo, presenta un valor R_{ij} **mayor a 0,7 o menor a -0,7**.
- **ANÁLISIS DE LA MATRIZ INVERSA DE CORRELACIÓN:** Se realiza una matriz cuadrada y la diagonal principal muestra coeficientes con **valores superiores a 10**.
- **ÍNDICE DE ACONDICIONAMIENTO:** El IC es **mayor o igual a 10**.

CONCLUSIÓN

Por lo tanto, se puede enunciar, que la multicolinealidad estudia las variables para conocer si existe una relación entre ellas. Existen tres formas de hacerlo. La primera estudia cada variable exógena comparándolas entre cada una de las variables explicativas. La segunda, relaciona una variable independiente con todas las demás a la vez. Y, por último, el índice de acondicionamiento las correlaciona todas entre sí.

DATOS IMPORTANTES

- Se debe resaltar, que todos los **métodos** anteriormente comentados son de **igual importancia**, y que se han de realizar todos y cada uno de los pasos nombrados anteriormente.
- Aunque exista multicolinealidad en uno de los métodos, no demostrará la existencia de multicolinealidad en los demás procesos.

SOLUCIONAR LA MULTICOLINEALIDAD

Si existe multicolinealidad durante los pasos previamente indicados, se deberá de eliminar una variable explicativa.

Para suprimir una variable se siguen dos criterios:

- Eliminar el mayor valor de la diagonal principal de la matriz inversa de correlación, pero si la diferencia del más grande y del segundo más grande es menor a uno, no se puede utilizar este criterio. En este caso se debería de seguir el segundo criterio.
- Hacer dos regresiones omitiendo una variable problemática en cada una, de manera que se eliminará la variable que genere un modelo con el R^2 más bajo.

2.3.3 ANÁLISIS MULTIVARIANTE

Este análisis **pretende explicar** la variable dependiente (**Y**) **a través de** las variables independientes (**X**). También es conocido como regresión lineal múltiple.

SIGNIFICATIVIDAD

Tras la supresión de las variables correlacionadas en la matriz de multicolinealidad, se elaborará la regresión múltiple, donde se conocerán los P-Valores de todos los parámetros, con el fin de **contrastar la hipótesis tanto del modelo como de los parámetros**. Este análisis ayudará a conocer si las variables son significativas o no.

	PARÁMETRO	VALOR-P
B₀	CONSTANTE	0,2258
B₁	INCENTIVOS CONTADO	0,1489
B₂	INCENTIVOS CRÉDITO	0,0078
B₃	FRECUENCIA VISITAS	0,0373
B₄	MARGEN TOTAL MEDIO DIARIO	0,0201
B₅	REGALOS	0,0034
B₆	IPC ALIM Y BEBI NO ALC	0,5218
B₇	TASA DE RIESGO DE POBREZA	0,1912

Tabla 4 - Tabla P-Valores de los diferentes parámetros.

El nivel de confianza será del **95%** y el error se estimará en un **5%**

$$\left. \begin{array}{l} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{array} \right\} \alpha = 0,05$$

Ilustración 5 - Contraste B_0

$$\left. \begin{array}{l} H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0 \\ H_1 : \text{Algún } \beta_i \neq 0 \end{array} \right\} \alpha = 0,05$$

Ilustración 6 - Significatividad del modelo.

P-Valor > α se aceptará H_0 y se concluirá que la variable analizada, **no es significativa**. Por lo que no debería pertenecer al modelo.

P-Valor < α se rechazará H_0 y se concluirá que la variable analizada, **sí es significativa**. Por lo que sí debería pertenecer al modelo.

NORMALIDAD DE RESIDUOS

En este apartado se examinará el **gráfico probabilístico** y el **histograma** de los **residuos**. Ambos **mostrarán** si los **residuos son normales o son anómalos**, pero no presentarán una evidencia clara de ello. Tan sólo serán mostrados de una **forma abstracta** para poder obtener a primera vista, una perspectiva general, sin desarrollar ningún proceso de cálculo. Por lo tanto, después de realizar estos análisis, se llevará a cabo un **contraste de hipótesis**, este último **confirmará con seguridad** si estos valores son anómalos o no a partir de unos cálculos matemáticos y una comparación que posteriormente se explicará.

- **GRÁFICO PROBABILÍSTICO:**

Estudia la normalidad de los residuos e identifica los valores anómalos.

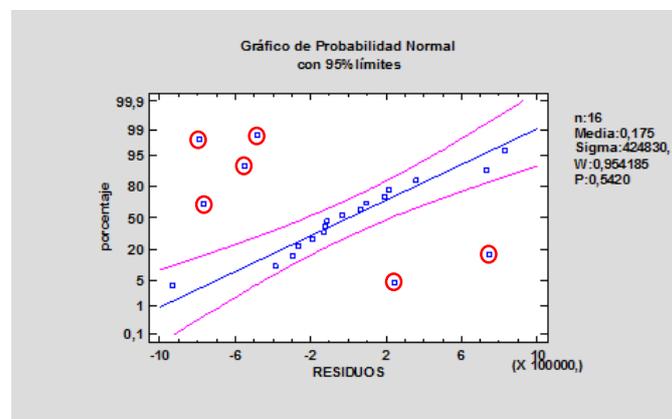


Ilustración 7 - Gráfico de probabilidad normal con 95% límites.

Las dos **líneas rosas** están representadas por los **intervalos de confianza** y la **línea de tendencia** por el **color azul**. Los **puntos cercanos a la línea de tendencia** mostrarán un **comportamiento normal** de los **residuos**. Mientras que, los puntos sobresalientes de los intervalos de confianza (redondeados en color rojo) podrían llegar a ser puntos anómalos o influyentes. Pero esto último, no se sabría con certeza hasta realizar el contraste de hipótesis.

- **HISTOGRAMA DE RESIDUOS:**

Estará compuesto por las **frecuencias** y los **intervalos de clase**. Los primeros, se situarán en el eje vertical y los segundos, en el eje horizontal, de donde se podrá observar a través de esta la normalidad de los residuos. El histograma se podrá mostrar de 4 formas diferentes:

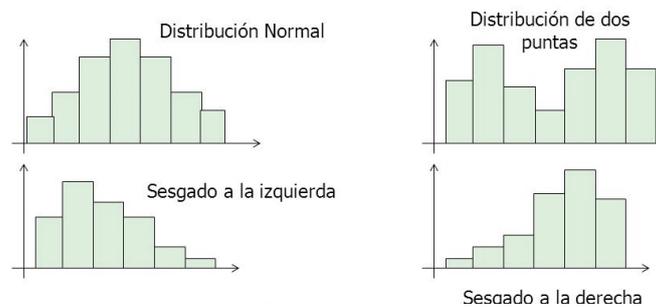


Ilustración 8- Distribuciones.

La primera imagen de esta ilustración, manifiesta una distribución normal. Esto significa que los **residuos** se encuentran **en la media**, por lo que se podría decir que los **residuos son normales**.

En los **demás casos no se sigue una distribución normal**, por lo que los residuos no se presentarían de forma normal.

- **CONTRASTE DE HIPÓTESIS:**

Se elabora un test de normalidad mediante Statgraphics **para averiguar** si los **residuos** se **distribuyen** de una forma **normal o anómala**. Dicho programa mostrará 4 P-Valores diferentes, los cuales están formados por: **Chi-cuadrado, Estadístico Shapiro-Wilk, Valor-Z para asimetría y Valor Z para curtosis**.

En este caso, no se utilizará Valor Z para curtosis, puesto que el número de delegaciones totales son insuficientes para poder obtener su valor.

Se escogerá la prueba con el Valor-P más restrictivo y se comparará con α .

$$\left. \begin{array}{l} H_0: \text{Los residuos son normales} \\ H_1: \text{Los residuos no son normales} \end{array} \right\} \alpha = 0,05$$

Ilustración 9 - Contraste de hipótesis.

Si **P-Valor > α (0,05)** se **aceptará H_0** y se concluirá con que los residuos se distribuyen de una forma **normal**.

Si **P-Valor < α (0,05)** se **rechazará H_0** y se concluirá con que los residuos **no** se distribuyen de una forma **normal**. Esto implicará que la variable dependiente asiste de forma significativa a la variable explicativa analizada. Por lo que esto implicará que no debe pertenecer al modelo.

HETEROCEDASTICIDAD

La heterocedasticidad se manifiesta cuando **se incumple una hipótesis del modelo de regresión lineal** ($Y = X\beta + u$). La varianza de la **perturbación no es constante**, es decir, cuando el error es distinto para cada una de las observaciones [2].

$$E(u^2) = \sigma^2 i, i = 1, 2, \dots, N.$$

El **término contrario** a este, se nombra **homocedasticidad**. En este caso todas las observaciones de la muestra la varianza se presentan constantes.

$$E(u^2) = \sigma^2$$

En la siguiente ilustración, se mostrarán algunos ejemplos de homocedasticidad y heterocedasticidad.

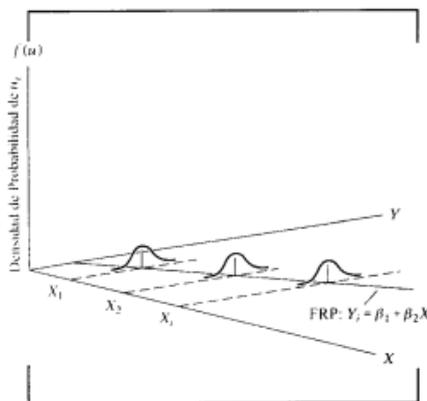


Figura 9.1

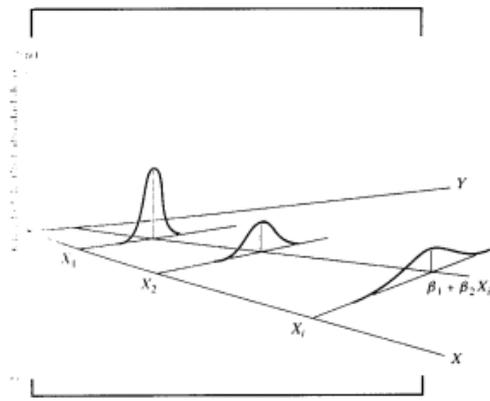


Figura 9.2

*Ilustración 13 – Homocedasticidad y heterocedasticidad.
 Fuente: Carlos Murillo Fort y Beatriz González López-Válcarcel [3]*

Para conocer si el modelo posee heterocedasticidad se deberá recurrir al contraste de hipótesis, y analizar los P-Valores.

$$\left. \begin{array}{l} H_0 : \sigma^2 = \text{Constante} \\ H_1 : \sigma^2 \neq \text{Constante} \end{array} \right\} \alpha = 0,05$$

Ilustración 10 - Contraste de hipótesis. Heterocedasticidad.

P-Valor > α se aceptará H_0 . La variable analizada **no** generará problemas de **heterocedasticidad**.

P-Valor < α se rechazará H_0 . La variable analizada generará problemas de **heterocedasticidad**.

AUTOCORRELACIÓN

La autocorrelación se puede definir como la **correlación entre miembros de series de observaciones ordenadas en el tiempo** (información de series de tiempo) **o en el espacio** (información de corte de transversal). El modelo de regresión lineal supone que no debe existir autocorrelación en los errores, es decir, el término de perturbación relacionado con una observación cualquiera no debería estar influenciado por el término de perturbación relacionado con cualquier otra observación [4].

Existen tres pruebas para poder **IDENTIFICAR LA AUTOCORRELACIÓN**, estas son:

- **REPRESENTACIÓN DE FAS Y DE FAP**
- **PRUEBA DE DURBIN-WATSON**
- **CONTRASTE DE HIPÓTESIS: GRÁFICO DE DURBIN-WATSON**

La **autocorrelación** de primer orden **positiva** y **negativa** se presentará de la siguiente manera:

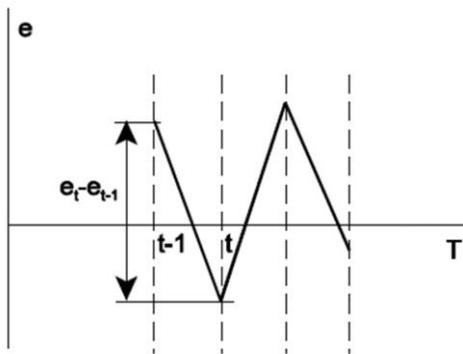


Ilustración 11 - Autocorrelación de primer orden negativa[4]

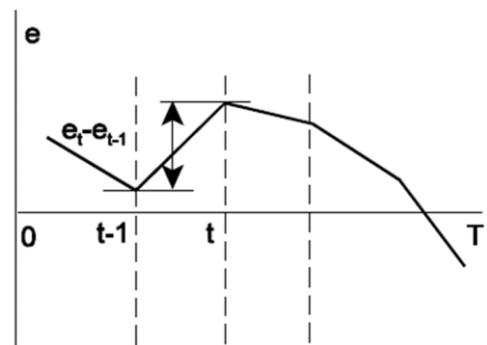


Ilustración 12 - Autocorrelación de primer orden positiva [4]

- REPRESENTACIÓN DE FAS Y DE FAP:

La representación gráfica de estos coeficientes de correlación para un conjunto **de retardos sucesivos** dará lugar a la función de autocorrelación simple o **FAS** [5].

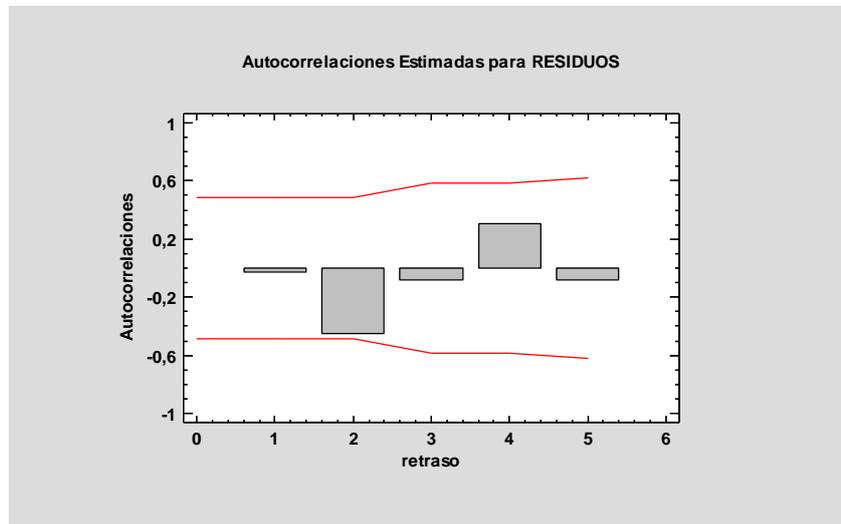


Ilustración 13- FAS

Junto con estos coeficientes de autocorrelación se definen los de autocorrelación parcial que **mide la correlación entre dos momentos de tiempo después de eliminar el efecto de los momentos intermedios**. La representación gráfica de estos coeficientes recibe el nombre de función de autocorrelación parcial [**FAP**]. [5]

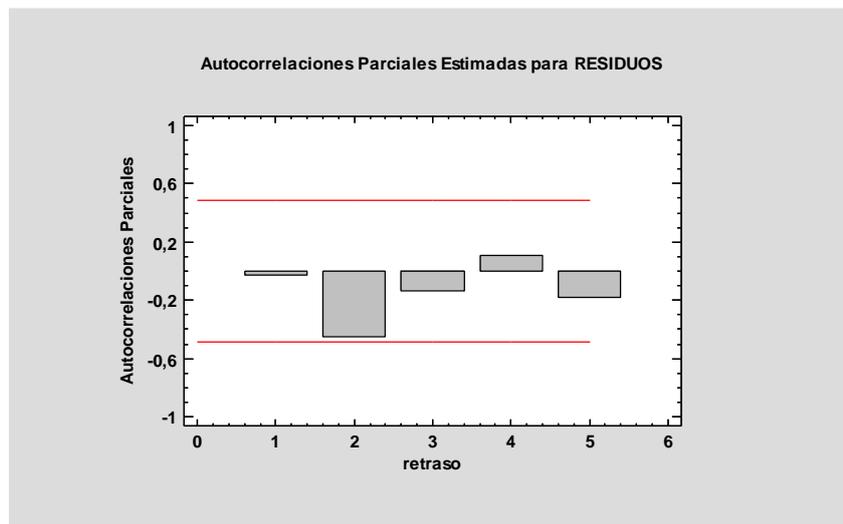


Ilustración 14 - FAP

- **PRUEBA DE DURBIN-WATSON:**

Se utilizará el test Durbin-Watson para detectar si existe **autocorrelación de primer orden** entre los residuales del modelo lineal de regresión simple (errores predictivos).

Los valores obtenidos se tendrán que situar en el gráfico que existe a continuación. Mediante este método se podrá conocer si existe correlación y en este caso, si es **positiva o negativa**.

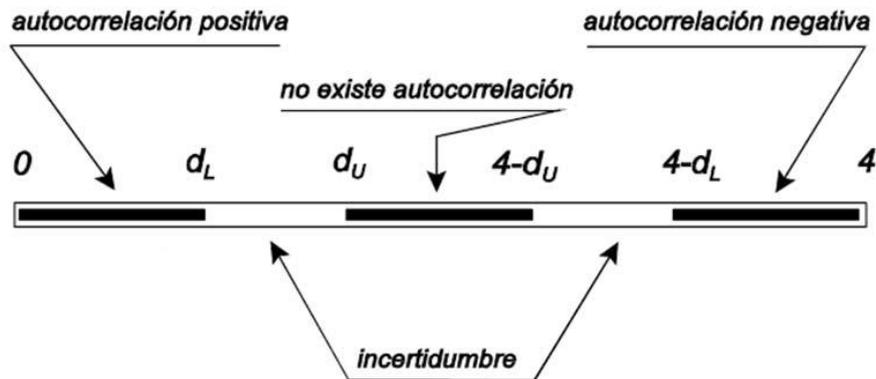


Ilustración 15 – Autocorrelación [6]

- **N** indicará el número de delegaciones que poseerá el modelo
- **K**: Número de variables independientes
- **A**: Error

- **CONTRASTE DE HIPÓTESIS:**

Otro método para conocer si existe autocorrelación de primer orden es el contraste de hipótesis.

$$\left. \begin{array}{l} H_0: \rho_1 = 0 \\ H_1: \rho_1 \neq 0 \end{array} \right\} \alpha = 0,05$$

Ilustración 16 - Contraste de hipótesis. Autocorrelación.

P-Valor > α (0,05) se aceptará H_0 y se concluirá que no existe autocorrelación de primer orden.

P-Valor < α (0,05) se aceptará H_1 y se concluirá que existe autocorrelación de primer orden.

PUNTOS ANÓMALOS Y PUNTOS INFLUYENTES

Los **puntos anómalos e influyentes**, como se ha indicado previamente, son aquellos que **se alejan de la línea de tendencia y no se encuentran entre los intervalos de confianza**. Todos los que estén presentes entre los intervalos de confianza o se encuentren sujetos a la línea de tendencia, mostrarán un comportamiento normal.

Para conocer si existen **puntos anómalos** candidatos a eliminar, se calculará el **residuo estudentizado** según las delegaciones pertenecientes al modelo. Este último, se analizará como **valor absoluto**. Si este muestra un valor **mayor a 3 o menor a -3 se eliminará la delegación del modelo**.

Posteriormente se analizarán los **puntos influyentes** tanto a priori como a posteriori analizando la influencia y el $|DFITS|$.

Cuando se habla de $|DFITS|$ se refiere a una medida escalada de la diferencia entre el valor estimado para la i -ésima observación utilizando toda la base de datos y el valor estimado para la i -ésima observación cuando ha sido eliminada de la base.

$$DFITS_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\sigma^2 \sqrt{p_{ii}}}$$

\tilde{n} será la influencia media de un solo punto

REGLAS PUNTOS INFLUYENTES:

A priori:

Si $n_{ii} \geq 2\tilde{n}$ se considerará de un punto influyente a priori

Si $n_{ii} < 2\tilde{n}$ No se considerará un punto influyente a priori

A posteriori:

Si $|DFIS| \geq 2\sqrt{\tilde{n}}$ Se considerará un punto influyente a posteriori

Si $|DFIS| \leq 2\sqrt{\tilde{n}}$ No se considerará un punto influyente a posteriori

A PRIORI	$n_{ii} \geq 2\tilde{n}$
A POSTERIORI	$ DFITS \geq 2\sqrt{\tilde{n}}$

Ilustración 17 - Puntos influyentes

2.3.4 FORMACIÓN REQUERIDA

Para la realización de este modelo econométrico, se ha requerido tener posesión de conocimientos sobre las distintas disciplinas académicas que se ofrecerán a continuación:

- **INTRODUCCIÓN A LA ESTADÍSTICA:**
Para poder ejecutar la estadística univariante y bivalente.
- **MODELOS MATEMÁTICOS (I) Y MODELOS MATEMÁTICOS (II):**
Se han aplicado estos conocimientos durante todo el transcurso de este proyecto.
- **MÉTODOS ESTADÍSTICOS PARA LA ECONOMÍA:**
Dicha asignatura ha contribuido a la posible interpretación de los test de hipótesis propuestos durante este modelo.
- **ECONOMETRÍA:**
La formación adquirida en esta área permitirá poseer la habilidad de analizar y de elaborar la técnica de regresión múltiple.

2.3.5 DATOS Y SOFTWARE

Para obtener la información utilizada durante el modelo se ha recurrido a la **base de datos de Dulcesol** y a la página web del Instituto Nacional de Estadística (**INE**). Para recoger datos a nivel interno, se ha utilizado la primera opción, mientras que, para conocer los datos a nivel externo de la empresa, se ha servido del INE (Instituto Nacional de Estadística).

DATOS

- INE



- Base de datos de Dulcesol



SOFTWARE

- Microsoft Excel 2018



- Microsoft Word 2018



- Microsoft Power Point 2018



- Statgraphics



- Google Chrome 2018



- Paint



3. CÁLCULOS, RESULTADOS Y EXPLOTACIÓN

3.1 ANÁLISIS UNIVARIANTE

En este apartado, se conocerán **qué puntos anómalos** existen en cada uno de los regresores que explican la variable regresada. Esto se averiguará mediante el **gráfico de caja y bigotes**.

3.1.1 GRÁFICO DE CAJAS Y BIGOTES

VENTA PROPIA (EUROS): Esta variable indica la venta total que se hace en cada delegación durante los meses estudiados en este proyecto (desde enero a septiembre).

RESUMEN ESTADÍSTICO PARA VENTA PROPIA	
RECUENTO	16
PROMEDIO	2,29E+06
MEDIANA	1,63E+06
DESVIACIÓN ESTÁNDAR	1,82E+06
MÍNIMO	333351
MÁXIMO	6,58E+06
RANGO	6,25E+06
CUARTIL INFERIOR	1,12E+06
CUARTIL SUPERIOR	2,97E+06
SESGO	1,40835
CURTOSIS	1,2985

Tabla 5 - Resumen estadístico para venta propia

- La **desviación estándar** que se muestra en el siguiente gráfico es moderada (79,41).
- Se observa que los datos presentan una **asimetría** a la derecha de la caja

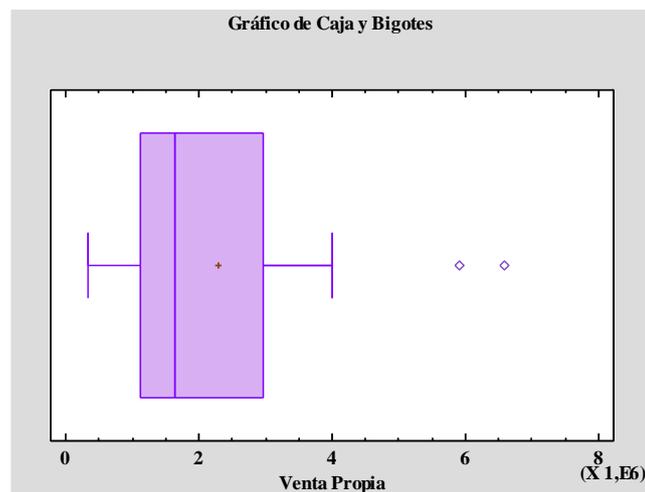


Ilustración 18 - Venta propia. Gráfico cajas y bigotes.

Los puntos anómalos que se pueden ver en el gráfico de caja y bigotes, se tratan de las delegaciones siguientes:

- Punto izquierdo: **Dulcosol**.
- Punto derecho: **Hispadul**.

MÁRGENES (EUROS):

RESUMEN ESTADÍSTICO PARA MÁRGENES	
RECuento	16
PROMEDIO	643017
MEDIANA	480070
DESVIACIÓN ESTÁNDAR	513090
MÍNIMO	95553,1
MÁXIMO	2,02E+06
RANGO	1,93E+06
CUARTIL INFERIOR	305352
CUARTIL SUPERIOR	842683
SESGO	1,54598
CURTOSIS	2,33423

Tabla 6 - Resumen estadístico para márgenes.

- En este caso, también se podría decir que la **desviación estándar** es moderada (79,794).
- Se puede apreciar que a la derecha se presenta una **asimetría** bastante significativa.

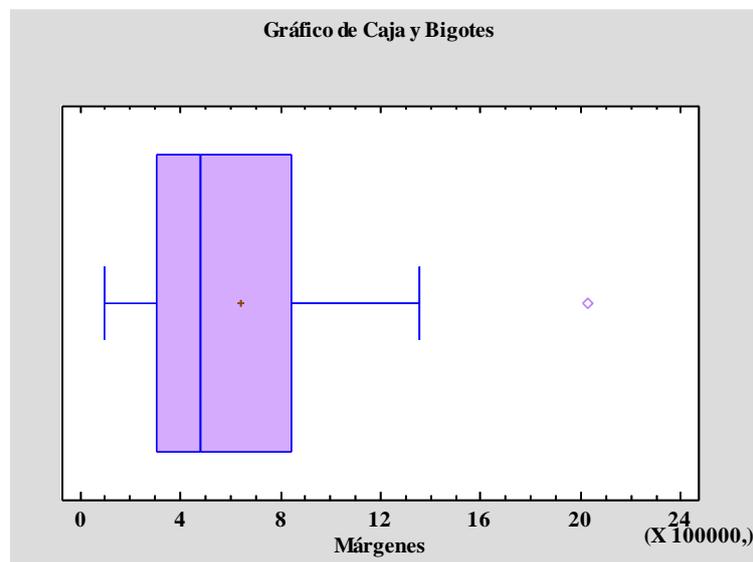


Ilustración 19 - Márgenes. Gráfico de cajas y bigotes.

El punto anómalo que se puede apreciar en la ilustración anterior se trata de **Hispadul**.

Nº DE COMERCIALES (PERSONAS):

RESUMEN ESTADÍSTICO PARA Nº COMERCIALES	
RECUENTO	16
PROMEDIO	25,9375
MEDIANA	19,5
DESVIACIÓN ESTÁNDAR	16,6031
MÍNIMO	12
MÁXIMO	71
RANGO	59
CUARTIL INFERIOR	14
CUARTIL SUPERIOR	34,5
SESGO	1,63816
CURTOSIS	2,44541

Tabla 7 - Resumen estadístico para número de comerciales.

- La **desviación estándar** que se presenta es baja (64,01195).
- Este gráfico de cajas y bigotes, presenta una **asimetría** a la derecha muy significativa.

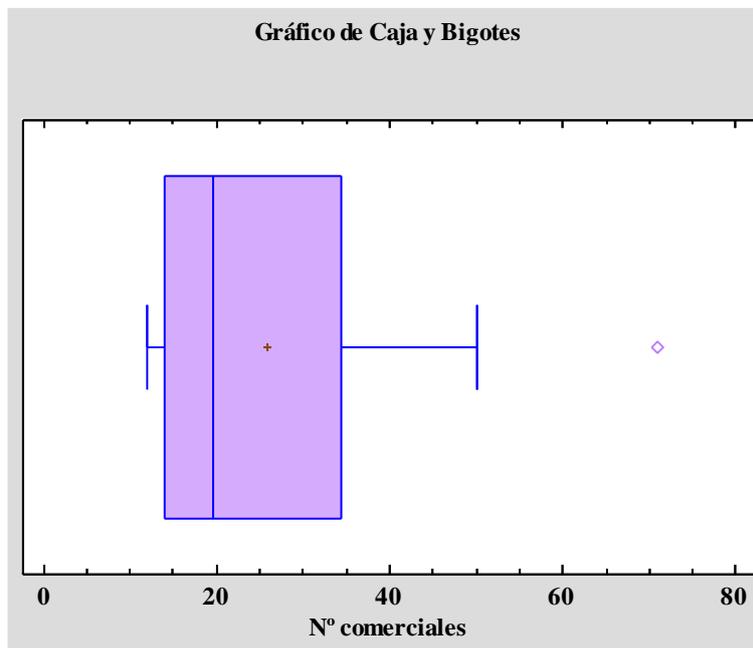


Ilustración 20 - Número de comerciales. Gráfico de cajas y bigotes.

El gráfico de cajas y bigotes relacionado con el número de comerciales muestra un punto anómalo en la delegación de **Dulmadrid**.

INCENTIVOS CONTADO (EUROS):

RESUMEN ESTADÍSTICO PARA INCENTIVOS CONTADO	
RECuento	16
PROMEDIO	1,97937
MEDIANA	1,7
DESVIACIÓN ESTÁNDAR	0,764918
MÍNIMO	1,06
MÁXIMO	3,83
RANGO	2,77
CUARTIL INFERIOR	1,7
CUARTIL SUPERIOR	1,91
SESGO	1,98017
CURTOSIS	3,36405

Tabla 8 - Resumen estadístico para incentivos de contado.

- La **desviación estándar** es muy pequeña (38,6445).
- Este gráfico de cajas y bigotes, presenta una **asimetría** muy significativa, donde muestra la no presencia del bigote derecho y un corto bigote izquierdo.

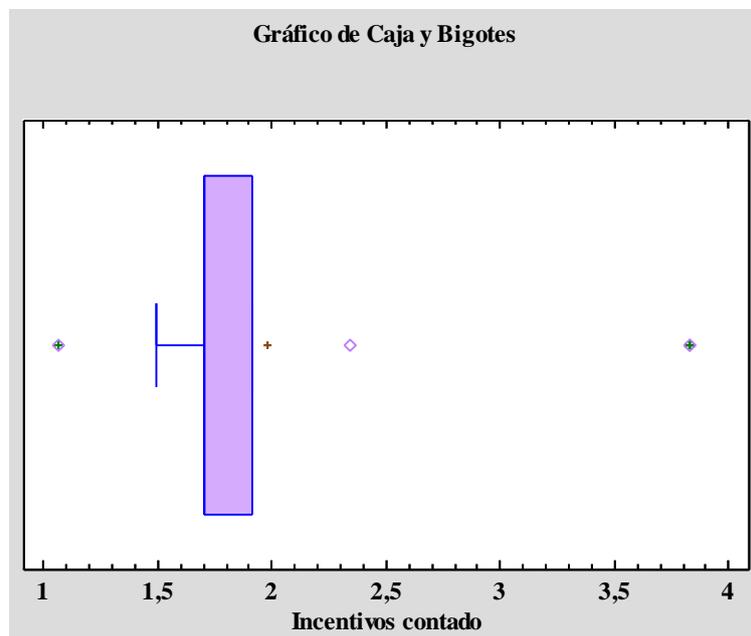


Ilustración 21 - Incentivos de contado. Gráfico de cajas y bigotes.

Este gráfico, muestra 3 puntos anómalos. Estos últimos son:

- Punto izquierdo: **Dulgalicia.**
- Punto central: **Hispadul.**
- Punto derecho: **Dulcosol.**

INCENTIVOS CRÉDITO (EUROS):

RESUMEN ESTADÍSTICO PARA INCENTIVOS CRÉDITO	
RECuento	16
PROMEDIO	1,08938
MEDIANA	0,85
DESVIACIÓN ESTÁNDAR	0,458526
MÍNIMO	0,64
MÁXIMO	2,34
RANGO	1,7
CUARTIL INFERIOR	0,85
CUARTIL SUPERIOR	1,17
SESGO	1,69191
CURTOSIS	2,67044

Tabla 9 - Resumen estadístico para incentivos de crédito.

- La **desviación estándar** que se muestra es pequeña (42,091).
- En cuanto a la **asimetría** de esta ilustración, se puede decir que el bigote izquierdo es más grande que el derecho por lo que muestra una asimetría significativa.

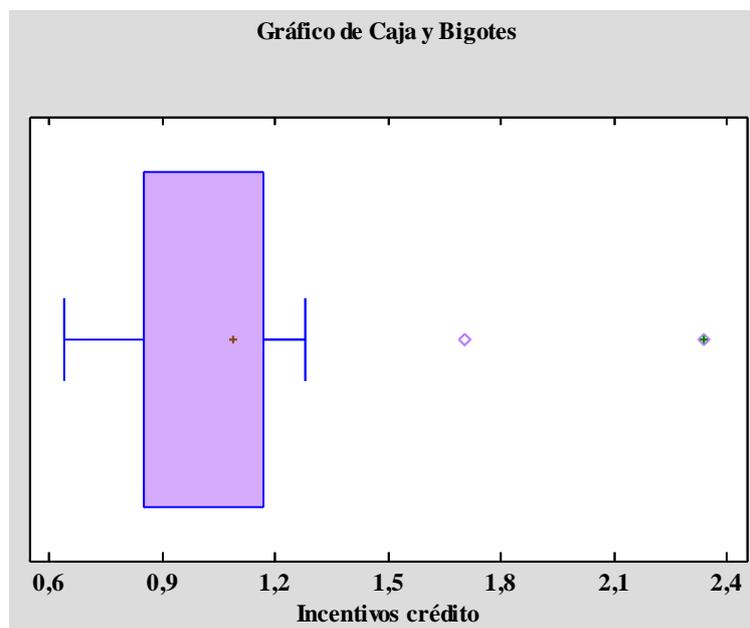


Ilustración 22 - Incentivos de crédito. Gráfico de caja y bigotes.

En este caso, las delegaciones que muestran anomalías en este gráfico son:

- Punto izquierdo: **Dulcosol**.
- Punto derecho: **Hispadul**.

FRECUENCIA VISITAS (UNIDADES):

RESUMEN ESTADÍSTICO PARA FRECUENCIA VISITAS	
RECuento	16
PROMEDIO	10,3381
MEDIANA	10,24
DESVIACIÓN ESTÁNDAR	1,34868
MÍNIMO	8,39
MÁXIMO	12,95
RANGO	4,56
CUARTIL INFERIOR	9,125
CUARTIL SUPERIOR	11,525
SESGO	0,266803
CURTOSIS	-1,04598

Tabla 10 - Resumen estadístico para frecuencia visitas.

- La **desviación estándar** presenta un valor muy bajo (13,0457).
- La **asimetría** que presenta es significativa, puesto que la parte derecha de la caja (bigote) es mucho más extensa que el de la izquierda.

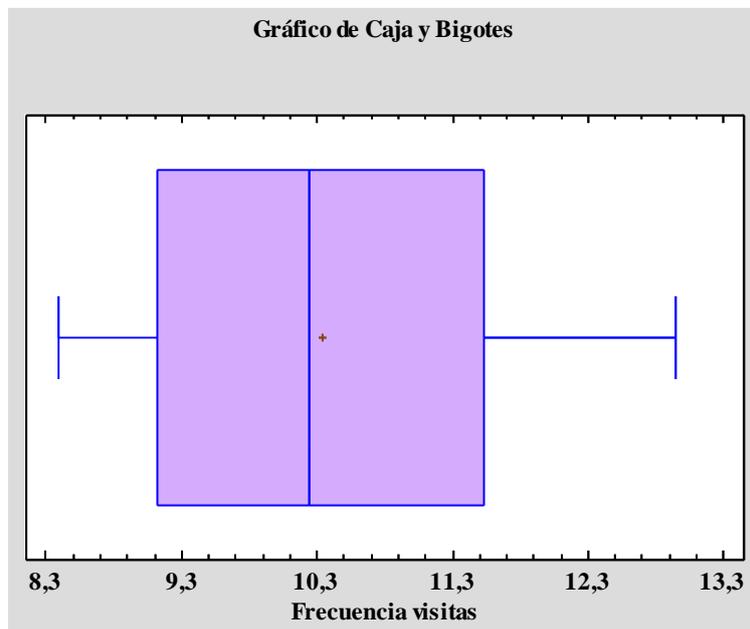


Ilustración 23 - Frecuencia visitas. Gráfico de caja y bigotes.

No se muestra **ningún punto anómalo** relacionado con la frecuencia de visitas.

TICKET MEDIO POR VISITA (UNIDADES):

RESUMEN ESTADÍSTICO PARA TICKET MEDIO POR VISITA	
RECuento	16
PROMEDIO	39,7281
MEDIANA	39,215
DESVIACIÓN ESTÁNDAR	9,47236
MÍNIMO	26,13
MÁXIMO	57,83
RANGO	31,7
CUARTIL INFERIOR	32,63
CUARTIL SUPERIOR	46,53
SESGO	0,457483
CURTOSIS	-0,733532

Tabla 11 - Resumen estadístico para ticket medio por visita.

- El valor de la **desviación estándar** es pequeño (64,8125).
- La **asimetría** de los datos es significativa, puesto el bigote de la derecha es mucho más largo que de la izquierda.

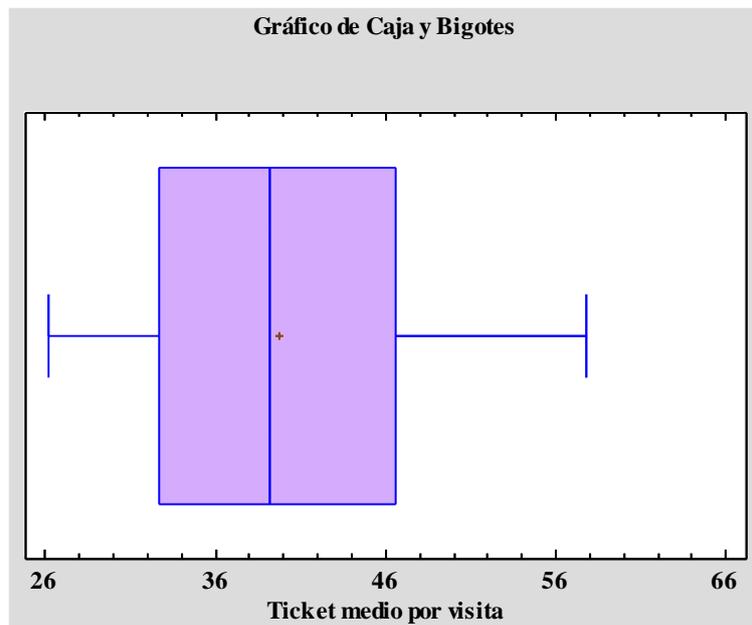


Tabla 12 - Ticket medio por visita. Gráfico de caja y bigotes.

No existen puntos anómalos en este gráfico de cajas y bigotes.

MARGEN TOTAL MEDIO DIARIO (EUROS):

RESUMEN ESTADÍSTICO PARA MARGEN TOTAL MEDIO DIARIO	
RECUENTO	16
PROMEDIO	11,1169
MEDIANA	10,37
DESVIACIÓN ESTÁNDAR	2,8217
MÍNIMO	5,99
MÁXIMO	15,88
RANGO	9,89
CUARTIL INFERIOR	9,485
CUARTIL SUPERIOR	13,485
SESGO	0,280347
CURTOSIS	-0,581184

Tabla 13 - Resumen estadístico para margen total medio diario.

- No existe una gran dispersión de los puntos sobre la media, puesto que el valor de la **desviación estándar** es mínimo (25,38).
- En esta ilustración se puede apreciar que la **asimetría** es significativa puesto que el bigote de la izquierda es en gran medida mayor que el de la derecha.

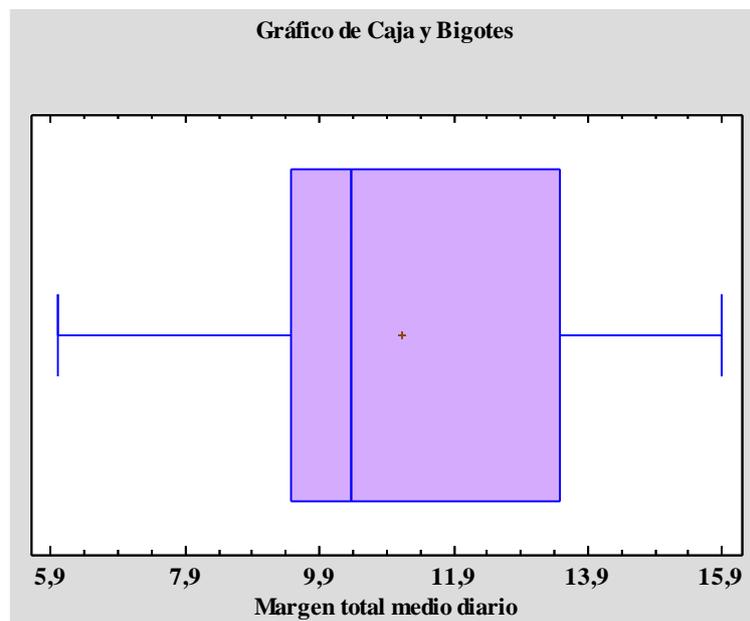


Ilustración 24 - Margen total medio diario. Gráfico de caja y bigotes.

No se aprecia **ningún punto anómalo** relacionado con el margen total medio diario.

NÚMERO DE CLIENTES (PERSONAS):

RESUMEN ESTADÍSTICO PARA CLIENTES	
RECUENTO	16
PROMEDIO	1272,63
MEDIANA	987
DESVIACIÓN ESTÁNDAR	927,232
MÍNIMO	339
MÁXIMO	3491
RANGO	3152
CUARTIL INFERIOR	633,5
CUARTIL SUPERIOR	1501,5
SESGO	1,50743
CURTOSIS	1,68412

Tabla 14 - Resumen estadístico para clientes

- El valor de la **desviación estándar** es bajo (72,86).
- La **asimetría** de los datos es significativa, puesto el bigote de la derecha es mucho más largo que de la izquierda.

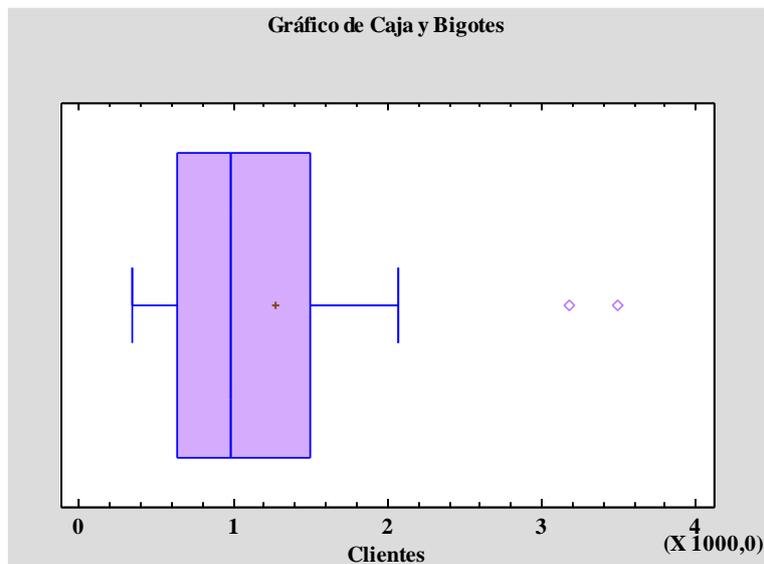


Ilustración 25 - Clientes. Gráfico de caja y bigotes.

Los puntos anómalos que se pueden observar, se tratan de las delegaciones de **Dulmadrid** e **Hispadul**.

Punto izquierdo: **Hispadul**

Punto derecho: **Dulmadrid**.

REGALOS (EUROS):

RESUMEN ESTADÍSTICO PARA REGALOS	
RECUENTO	16
PROMEDIO	20452,8
MEDIANA	12121,6
DESVIACIÓN ESTÁNDAR	18063
MÍNIMO	1983,6
MÁXIMO	63893,5
RANGO	61909,9
CUARTIL INFERIOR	6046,77
CUARTIL SUPERIOR	32910,3
SESGO	1,03003
CURTOSIS	0,47372

Tabla 15 - Resumen estadístico para regalos.

- La **desviación estándar** presenta una cifra alta (88,32).
- Se presenta una **asimetría** a la derecha en el gráfico cajas y bigotes.

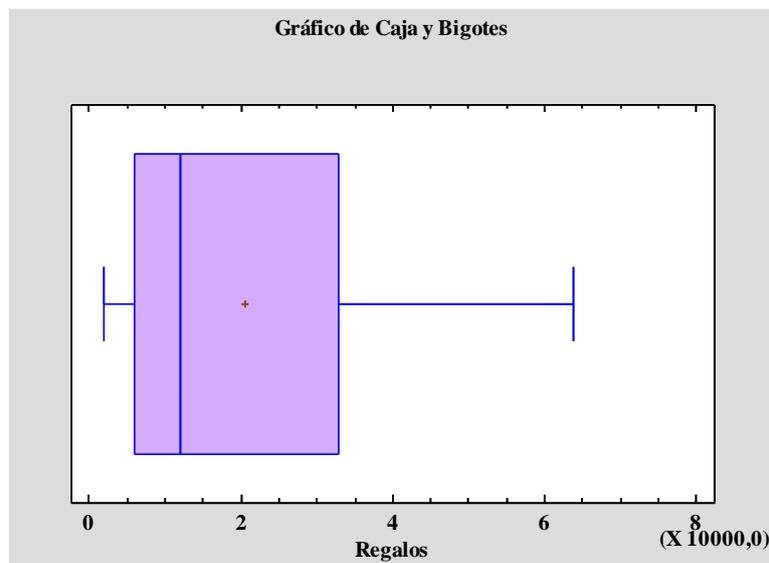


Ilustración 26 - Regalos. Gráfico de caja y bigotes.

En este caso, la ilustración **no muestra ningún punto anómalo existente.**

POBLACIÓ (PERSONAS):

RESUMEN ESTADÍSTICO PARA POBLACIÓN	
RECuento	16
PROMEDIO	1337580
MEDIANA	930550
DESVIACIÓN ESTÁNDAR	1235210
MÍNIMO	203304
MÁXIMO	5531110
RANGO	5327800
CUARTIL INFERIOR	720500
CUARTIL SUPERIOR	1619370
SESGO	2,85791
CURTOSIS	9,68175

Tabla 16 - Resumen estadístico para población.

- Existe una gran dispersión de los puntos respecto a la media, puesto que la **desviación estándar** es muy alta (92,35).
- El gráfico de caja y bigotes que se puede observar a continuación parece **simétrico**. No existe gran diferencia entre un bigote y otro.

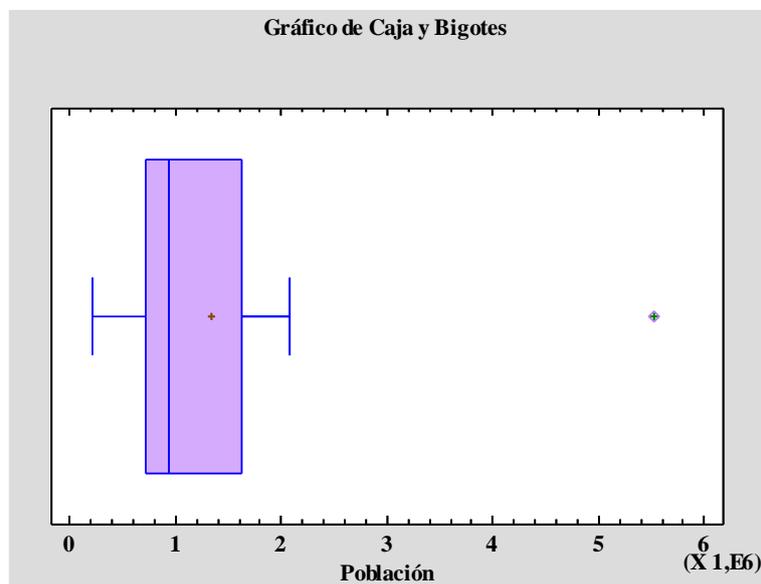


Ilustración 27 - Población. Gráfico de caja y bigotes.

En la ilustración anterior, se puede observar un punto anómalo. Este atiende a la delegación de **Dulmadrid**.

NÚMERO DE MUJERES (PERSONAS):

RESUMEN ESTADÍSTICO PARA NÚMERO DE MUJERES	
RECuento	16
PROMEDIO	687841
MEDIANA	483221
DESVIACIÓN ESTÁNDAR	642860
MÍNIMO	104967
MÁXIMO	2882910
RANGO	2777950
CUARTIL INFERIOR	367157
CUARTIL SUPERIOR	827776
SESGO	2,9122
CURTOSIS	9,98044

Tabla 17 - Resumen estadístico para número de mujeres.

- La dispersión de los puntos en el número de mujeres dista mucho de la media, la **desviación típica** es elevada (93,46).
- No parece que se presente una **asimetría** en ninguno de los dos bigotes.

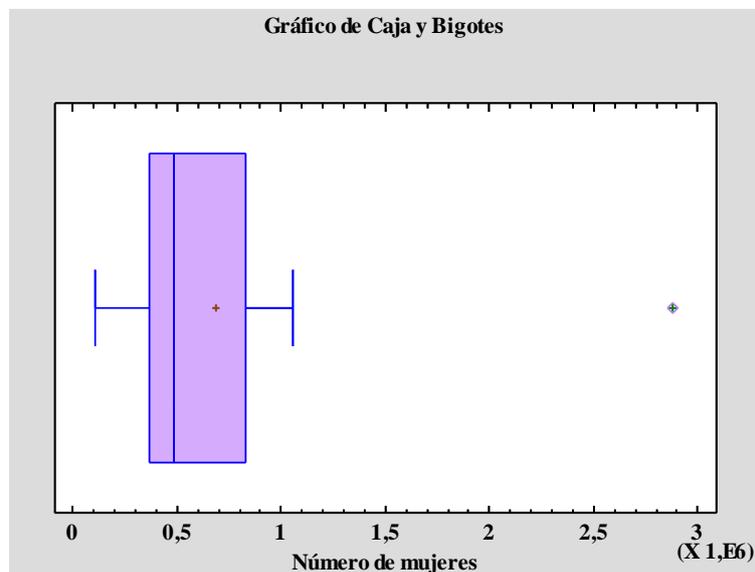


Ilustración 28 - Número de mujeres. Gráfico de caja y bigotes.

El punto anómalo presentado, lo origina la delegación de **Dulmadrid**.

IPC ALIMENTOS Y BEBIDAS NO ALCOHÓLICAS (EUROS):

RESUMEN ESTADÍSTICO PARA IPC ALIM Y BEBI NO ALC	
RECUENTO	16
PROMEDIO	2,44375
MEDIANA	2,5
DESVIACIÓN ESTÁNDAR	0,215928
MÍNIMO	2
MÁXIMO	2,7
RANGO	0,7
CUARTIL INFERIOR	2,35
CUARTIL SUPERIOR	2,6
SESGO	-1,15132
CURTOSIS	0,390088

Tabla 18 - Resumen estadístico para IPC alimentos y bebidas no alcohólicas.

- La **desviación típica** es muy baja, presenta un valor de 8,83. Por lo tanto, los puntos están muy cerca de la media.
- El gráfico de cajas y bigotes presenta una **asimetría** a la izquierda.

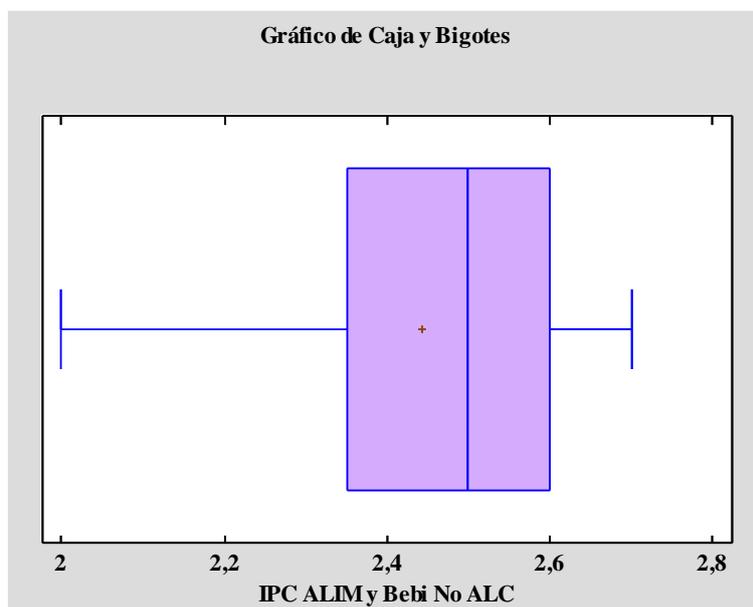


Ilustración 29 - IPC alimentos y bebidas no alcohólicas. Gráfico de caja y bigotes.

El gráfico de caja y bigotes relacionado con el IPC de alimentos y bebidas no alcohólicas, no muestra **ningún punto anómalo**.

TASA DE RIESGO DE POBREZA (PERSONAS):

RESUMEN ESTADÍSTICO PARA TASA DE RIESGO DE POBREZA	
RECuento	16
PROMEDIO	19,0625
MEDIANA	18,85
DESVIACIÓN ESTÁNDAR	6,11543
MÍNIMO	10,7
MÁXIMO	26,4
RANGO	15,7
CUARTIL INFERIOR	13,6
CUARTIL SUPERIOR	25,45
SESGO	-0,0751302
CURTOSIS	-1,77624

Tabla 19 - Resumen estadístico para tasa de riesgo de pobreza.

- No existe una gran dispersión de los puntos sobre la media, puesto que el valor de la **desviación estándar** es mínimo (32,08).
- Esta ilustración presenta una **asimetría** significativa a la izquierda.

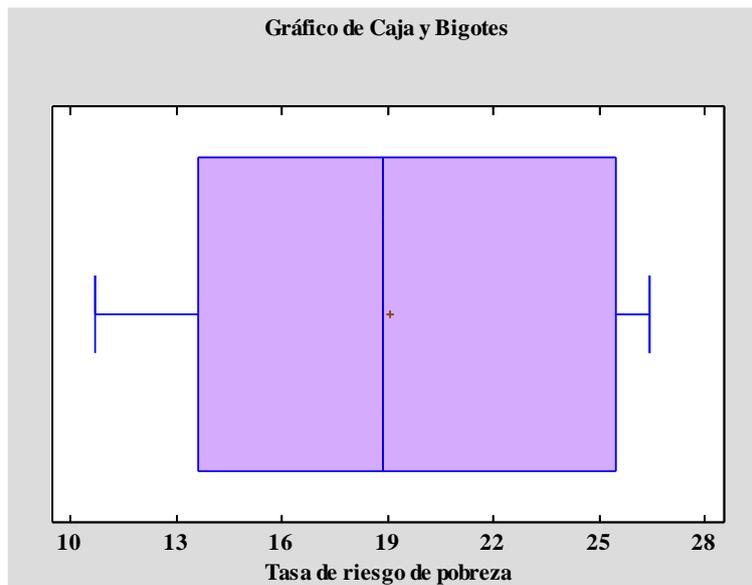


Ilustración 30 - Tasa de riesgo de pobreza. Gráfico de caja y bigotes.

El anterior gráfico **no presenta ningún punto anómalo** asociado a alguna delegación.

3.1.2 CONCLUSIÓN ANÁLISIS UNIVARIANTE

Después de realizar el análisis univariante, se ha decidido resumir mediante dos tablas los puntos anómalos detectados, con el fin de tener una visualización más global de los mismos.

- En la **primera ilustración**, se expondrán las **delegaciones** en las que **se han encontrado los puntos anómalos** según cada variable.
- En la **segunda ilustración**, se presentará **el número de veces** que cada **delegación** ha **mostrado puntos anómalos** en las variables exógenas.

VARIABLES	PUNTOS ANÓMALOS DETECTADOS
VENTA PROPIA	Dulcosol e Hispadul
MÁRGENES	Hispadul
NÚMERO DE COMERCIALES	Dulmadrid
INCENTIVOS CONTADO	Dulcosol, Dulgalicia e Hispadul
INCENTIVOS CRÉDITO	Dulcosol e Hispadul
FRECUENCIA VISITAS	Ningún punto anómalo
TICKET MEDIO POR VISITA	Ningún punto anómalo
MARGEN TOTAL O MEDIO DIARIO	Ningún punto anómalo
NÚMERO DE CLIENTES	Dulmadrid e Hispadul
REGALOS	Ningún punto anómalo
POBLACIÓN	Dulmadrid
NÚMERO DE MUJERES	Dulmadrid
IPC ALIMENTOS Y BEBIDAS NO ALCOHÓLICAS	Ningún punto anómalo
TASA DE RIESGO DE POBREZA	Ningún punto anómalo

Tabla 20 – Puntos anómalos detectados en las variables.

DELEGACIONES	PUNTOS ANÓMALOS ENCONTRADOS EN LAS DELEGACIONES
ARADUL	0
ASTURSOL	0
DOCES LUCENSES	0
DULCADIS	0
DULCOSOL	3
DULGALICIA	1
DULGANDIA	0
DULMADRID	4
EUSKASOL	0
GADESDUL	0
HERMANOS MUÑOZ	0
HISPADUL	5
LEVANTINA	0
TINERFEÑA	0
VEGODISA	0
VIGUESA	0

Tabla 21 - Puntos anómalos encontrados en las delegaciones.

3.1.3 INTERPRETACIÓN

Existen tres delegaciones que se comportan modestamente diferente al resto, pero tampoco de una forma desmedida y sin un motivo aparente. Sin contar con Dulgalicia que sólo se ha mostrado en una ocasión anómala en el anterior análisis y de la cual no cabe comentar, puesto que no se considera importante hacerlo. Estas son Dulcosol, Dulmadrid e Hispadul.

Si se analiza la tabla donde se resumen todos los datos, se puede apreciar que estas tres delegaciones comparten importantes variables, las cuales pueden ser el motivo de que conjuntamente resalten sobre el resto, puesto que tienen una relación entre ellas y, de esta forma crear la ilusión de que quizá no se hayan seleccionado datos correctos o haya sucedido otro factor que pueda influir a que estas se deban excluir del proyecto.

Las variables que dichas delegaciones presentan los mayores resultados sobre las demás en:

- **Márgenes**
- **Número de clientes**
- **Regalos**

Además, en la regresora **número de comerciales, Dulmadrid e Hispadul son las más altas**, puesto que el nivel de **población** allí es de los más altos que poseen las delegaciones.

Las demás delegaciones poseen ventas más bajas porque no detentan muchos de estos factores destacados y por esto, pueden dichas delegaciones destacar sobre el resto. Por lo tanto, **se decide continuar con todas las delegaciones** y *no se excluye ninguna* por tener un comportamiento desproporcionado al resto de forma no justificada.

3.2 ANÁLISIS BIVARIANTE

En el análisis bivalente se estudiará el gráfico X e Y, el modelo de regresión simple y el valor R.

3.2.1 GRÁFICO X E Y, Y MODELO DE REGRESIÓN SIMPLE

MÁRGENES VS VENTA PROPIA

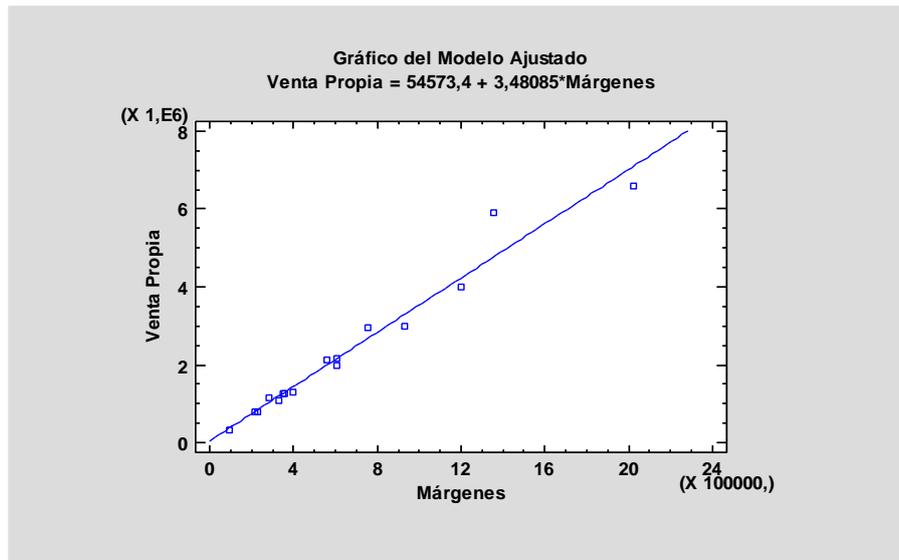


Ilustración 31 - Márgenes vs Venta propia.

MODELO DE REGRESIÓN SIMPLE

- $Venta Propia = 54573,4 + 3,48085 \cdot Márgenes$
- Coeficiente de Correlación = **0,980904**

CONCLUSIÓN

Después de comparar los márgenes con las ventas propias y, de analizar el coeficiente de correlación, se puede decir que las variables están *claramente relacionadas* entre sí de forma directa.

NÚMERO DE COMERCIALES VS VENTA PROPIA

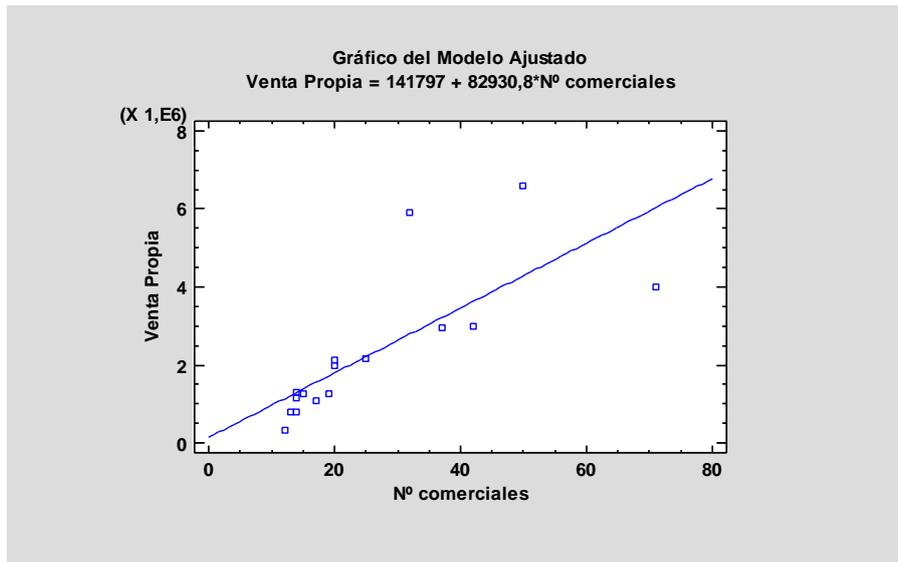


Ilustración 32 - Número de comerciales vs Venta propia.

MODELO DE REGRESIÓN SIMPLE

- Venta Propia = 141797 + 82930,8 · N° comerciales
- Coeficiente de Correlación = **0,756229**

CONCLUSIÓN

En este caso, se puede observar, que el número de comerciales y las ventas propias, presentan una *gran relación* entre ambas, puesto que el R2 es mayor a 0,7.

INCENTIVOS CONTADO VS VENTA PROPIA

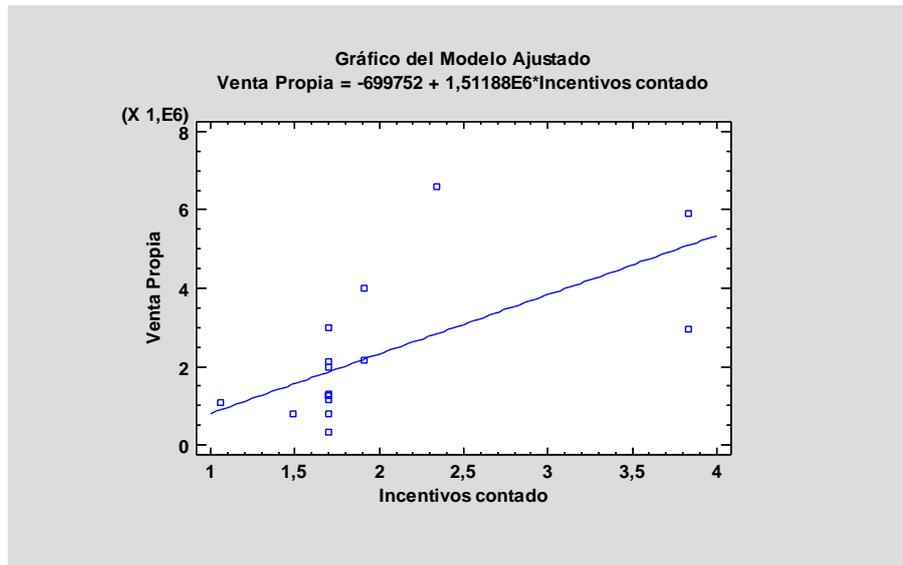


Ilustración 33 - Incentivos contado vs Venta propia.

MODELO DE REGRESIÓN LINEAL

- Venta Propia = -699752 + 1,51188E6 · Incentivos contado
- Coeficiente de Correlación = **0,635154**

CONCLUSIÓN

La variable regresora (incentivos contado) y la variable regresada (venta propia) evidencian una *severa asociación* entre ambas.

INCENTIVOS CRÉDITO VS VENTA PROPIA

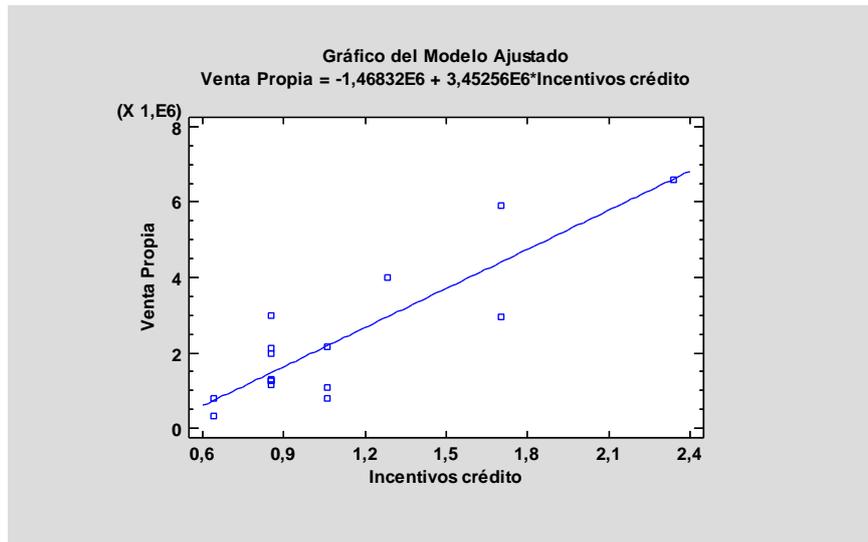


Ilustración 34 - Incentivos crédito vs Venta propia.

MODELO DE REGRESIÓN LINEAL

- Venta Propia = $-1,46832E6 + 3,45256E6 \cdot \text{Incentivos crédito}$
- Coeficiente de Correlación = **0,869469**

CONCLUSIÓN

Esta ilustración expone una *gran relación* entre la variable exógena (incentivos crédito) y la variable endógena (venta propia).

FRECUENCIA VISITAS VS VENTA PROPIA

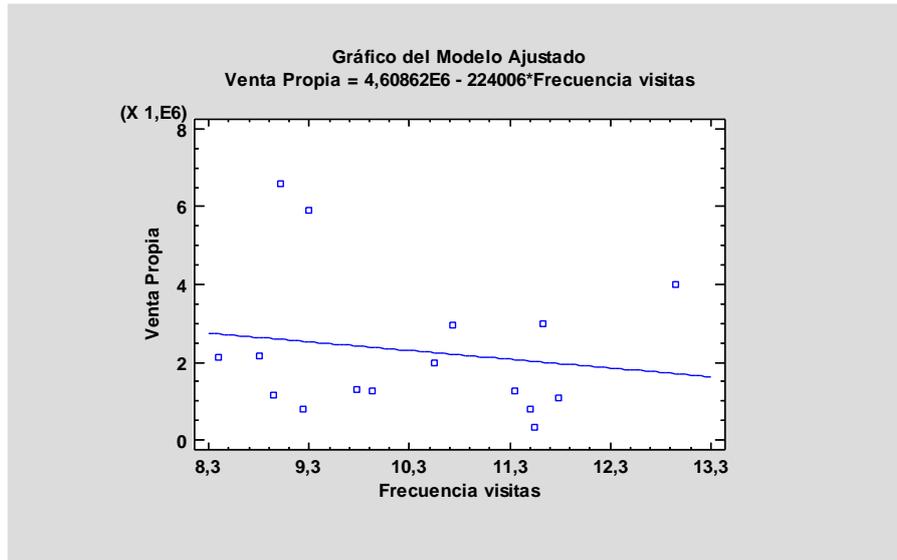


Ilustración 35 - Frecuencia visitas vs Venta propia.

MODELO DE REGRESIÓN LINEAL

- Venta Propia = 4,60862E6 - 224006 · Frecuencia visitas
- Coeficiente de Correlación = **-0,165927**

CONCLUSIÓN

Como conclusión al gráfico anterior, se puede añadir, que la frecuencia de visitas y la venta propia *no están relacionadas* entre sí.

TICKET MEDIO POR VISITA VS VENTA PROPIA

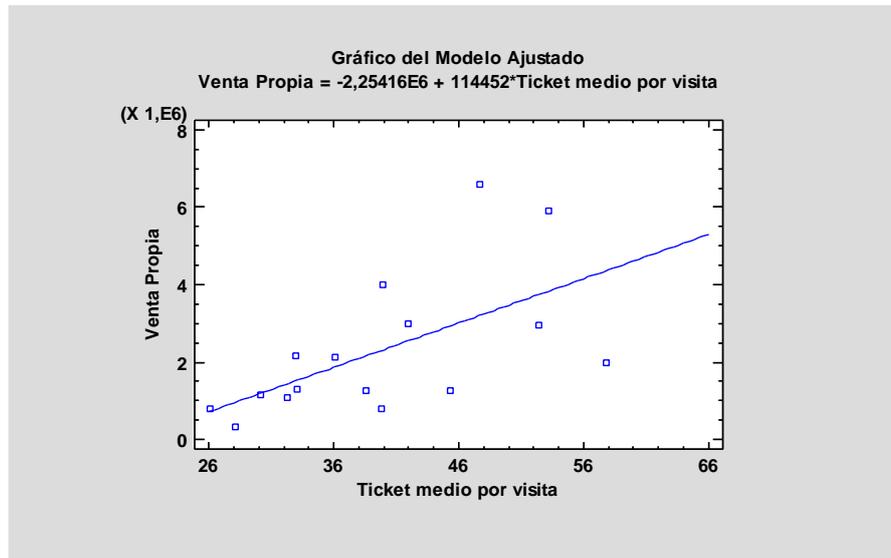


Ilustración 36 - Ticket medio por visita vs Venta propia.

MODELO DE REGRESIÓN LINEAL

- Venta Propia = -2,25416E6 + 114452 · Ticket medio por visita
- Coeficiente de Correlación = **0,59543**

CONCLUSIÓN

Se puede contemplar visualizando la ilustración expuesta anteriormente, que las variables ticket medio por visita y venta propia están *moderadamente* relacionadas entre ellas.

MARGEN TOTAL MEDIO DIARIO VS VENTA PROPIA

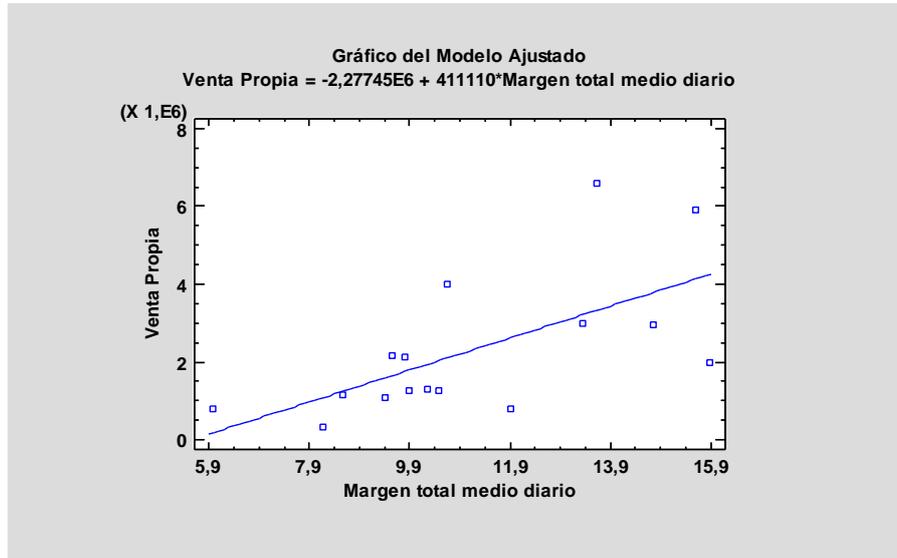


Ilustración 37 - Margen total medio diario vs Venta propia.

MODELO DE REGRESIÓN LINEAL

- Venta Propia = $-2,27745E6 + 411110 \cdot \text{Margen total medio diario}$
- Coeficiente de Correlación = **0,637114**

CONCLUSIÓN

Comparando la variable explicativa (margen total medio diario) y la variable explicada (venta propia), se puede presenciar que ambas están *modestamente* relacionadas.

CLIENTES VS VENTA PROPIA

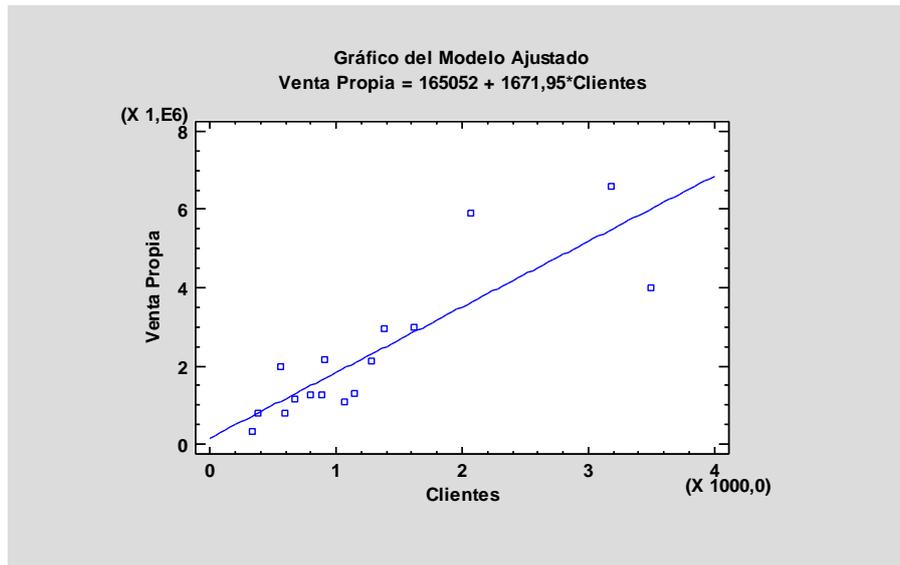


Ilustración 38 - Clientes vs Venta propia.

MODELO DE REGRESIÓN LINEAL

- Venta Propia = 165052 + 1671,95 · Clientes
- Coeficiente de Correlación = **0,851451**

CONCLUSIÓN

Seguidamente, después de tratar conjuntamente las variables clientes y venta propia, se puede considerar que cada una de ellas, están *muy relacionadas* entre sí.

REGALOS VS VENTA PROPIA

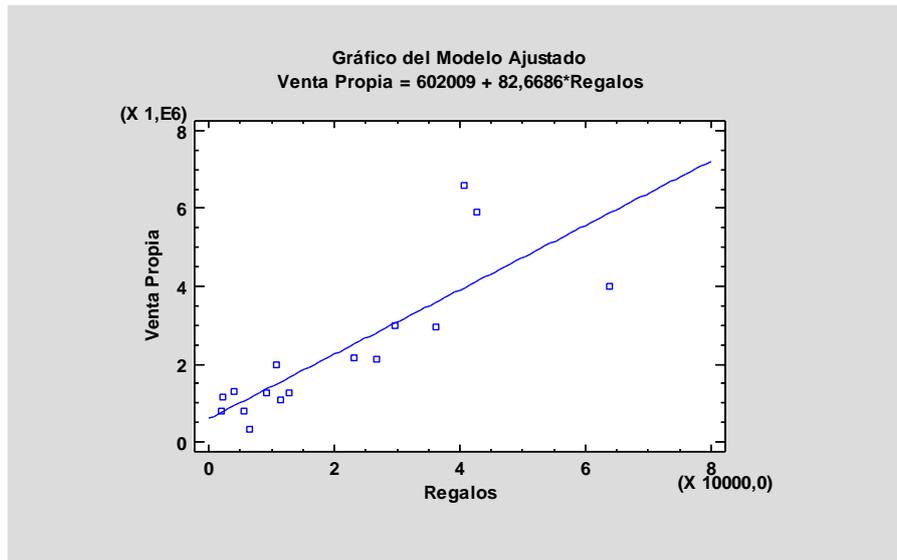


Ilustración 39 - Regalos vs Venta propia.

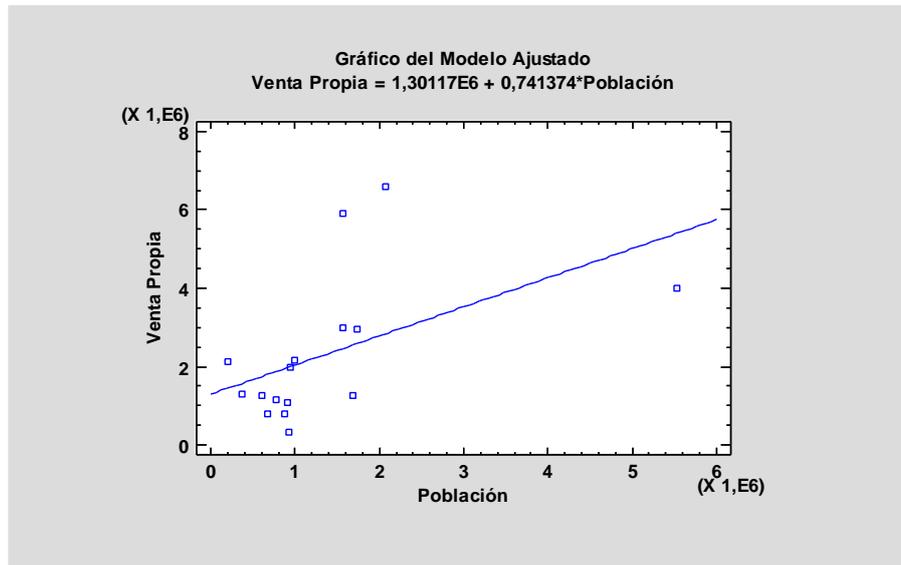
MODELO DE REGRESIÓN LINEAL

- Venta Propia = 602009 + 82,6686 · Regalos
- Coeficiente de Correlación = **0,820124**

CONCLUSIÓN

El siguiente gráfico manifiesta una *gran relación* entre las variables regalos y venta propia. Asimismo, el coeficiente de correlación lo indica, mostrando un valor superior al 0,7.

POBLACI3N VS VENTA PROPIA



Ilustraci3n 40 - Poblaci3n vs Venta propia.

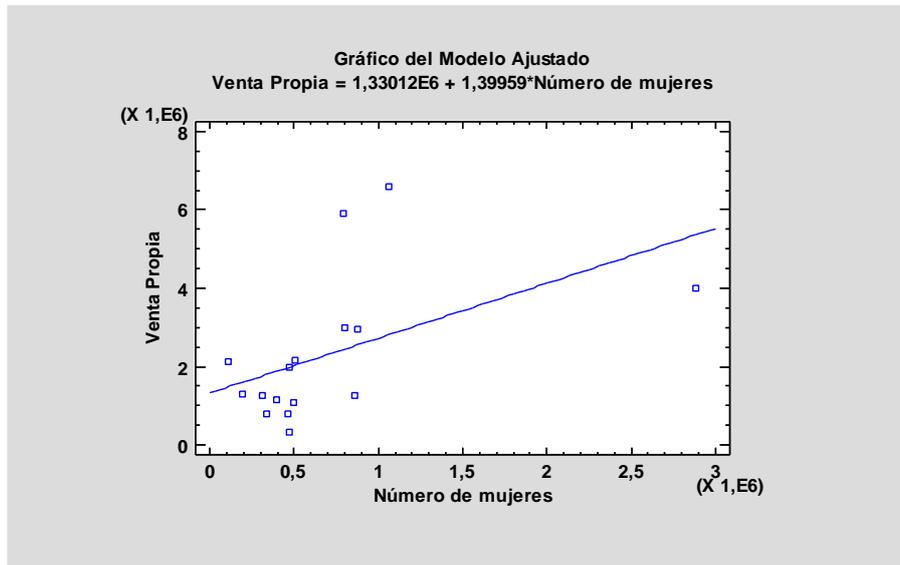
MODELO DE REGRESI3N LINEAL

- $Venta\ Propia = 1,30117E6 + 0,741374 \cdot Poblaci3n$
- Coeficiente de Correlaci3n = **0,502953**

CONCLUSI3N

La ilustraci3n anterior, presenta un coeficiente de correlaci3n del 0,502953 como se ha indicado anteriormente. Esto quiere decir, que tanto la variable poblaci3n como la variable venta propia, estn *humildemente relacionadas*.

NÚMERO DE MUJERES VS VENTA PROPIA



IPC ALIMENTOS Y BEBIDAS NO ALCOHÓLICAS VS VENTA PROPIA

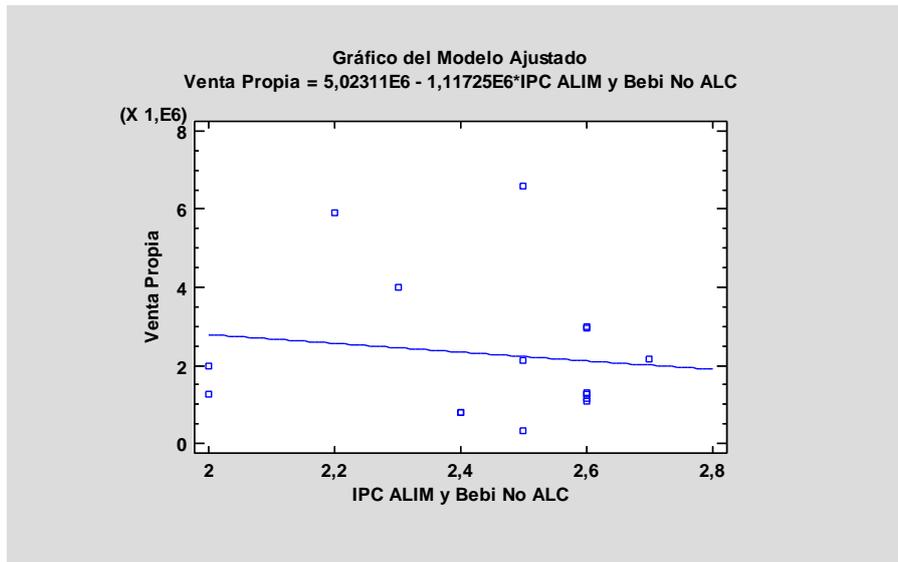


Ilustración 42 - IPC Alimentos y bebidas no alcohólicas vs Venta propia.

MODELO DE REGRESIÓN LINEAL

- Venta Propia = 5,02311E6 - 1,11725E6 · IPC ALIM y Bebi No ALC
- Coeficiente de Correlación = **-0,132498**

CONCLUSIÓN

El modelo de regresión lineal indica que existe una *mínima correlación* entre las variables IPC Alimentos y bebidas no alcohólicas y las ventas propias.

TASA DE RIESGO DE POBREZA VS VENTA PROPIA

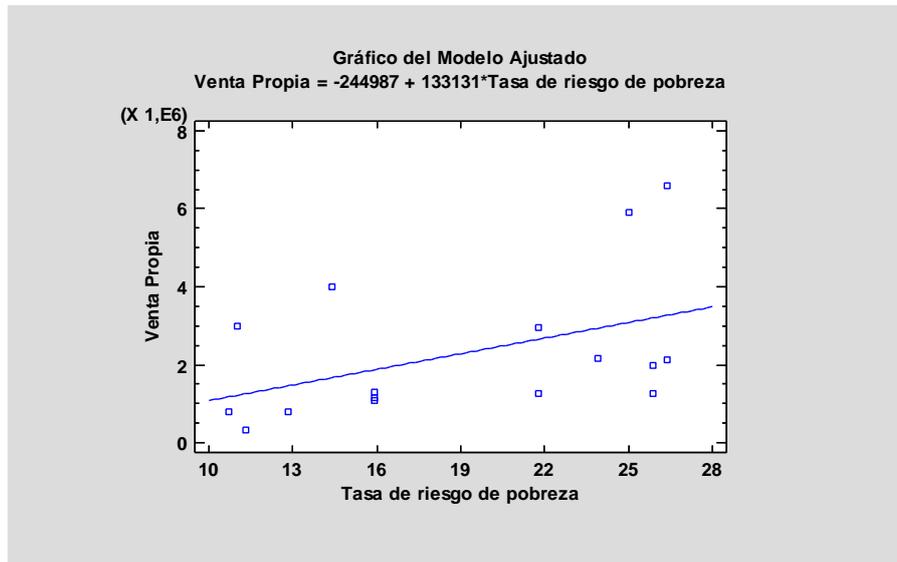


Ilustración 43 - Tasa de riesgo de pobreza vs Venta propia.

MODELO DE REGRESIÓN LINEAL

- $Venta Propia = -244987 + 133131 \cdot Tasa de riesgo de pobreza$
- Coeficiente de Correlación = **0,44715**

CONCLUSIÓN

Por último, si se examina conjuntamente la tasa de riesgo de pobreza y la venta propia, se puede observar que ambas están *modestamente relacionadas* entre sí.

3.2.2 CONCLUSIÓN ANÁLISIS BIVARIANTE GRÁFICO X VS Y:

	VARIABLES	VENTA PROPIA	RELACIÓN ENTRE AMBAS
B1	MÁRGENES	0,980904	Muy relacionadas
B2	NÚMERO DE COMERCIALES	0,756229	Muy relacionadas
B3	INCENTIVOS CONTADO	0,635154	Moderadamente relacionadas
B4	INCENTIVOS CRÉDITO	0,869469	Muy relacionadas
B5	FRECUENCIA VISITAS	-0,165927	No están relacionadas
B6	TICKET MEDIO POR VISITA	0,59543	Moderadamente relacionadas
B7	MARGEN TOTAL MEDIO DIARIO	0,637114	Moderadamente relacionadas
B8	CLIENTES	0,851451	Muy relacionadas
B9	REGALOS	0,820124	Muy relacionadas
B10	POBLACIÓN	0,502953	Moderadamente relacionadas
B11	NÚMERO DE MUJERES	0,494157	Moderadamente relacionadas
B12	IPC ALIMENTOS Y BEBIDAS NO ALCOHÓLICAS	-0,132498	No están relacionadas
B13	TASA DE RIESGO DE POBREZA	0,44715	Moderadamente relacionadas

Tabla 22 - Relación entre las variables.

Se ha esquematizado el resultado del análisis bivalente en una tabla, para que de esta forma se pueda visualizar todo conjuntamente de una manera más efectiva.

Como se ha podido apreciar y posteriormente analizar, existen 5 agentes que están **muy relacionados** entre sí, los valores de estos se hayan por *encima de 0,7* y se cree que se podrá modelizar las ventas propias de cada delegación. Estos, son los nombrados a continuación:

- Márgenes
- Número de comerciales
- Incentivos crédito
- Clientes
- Regalos.

Asimismo, se ha podido contemplar, que existen 6 componentes que están **moderadamente relacionados**. Este conjunto está comprendido por los valores *desde 0,4 y 0,69*. Dichos elementos están formados por los factores:

- Incentivos contado
- Ticket medio por visita
- Margen total medio diario
- Población
- Número de mujeres
- Tasa de riesgo de pobreza

Y, por último, también se han encontrado variables que **no estaban relacionadas**, es decir, que la variable explicativa no estaba relacionada con la variable explicada. Estas están compuestas por los valores *desde menos infinito hasta 0,4*. De este tipo se han encontrado dos variables:

- Frecuencia visitas
- IPC alimentos y bebidas no alcohólicas

3.2.2 MULTICOLINEALIDAD

MATRIZ DE CORRELACIÓN

	MÁRGENES	Nº COMERCIALES	INCENTIVOS CONTADO	INCENTIVOS CRÉDITO	FRECUENCIA VISITAS	TICKET MEDIO POR VISITA	MARGEN TOTAL MEDIO DIARIO	CLIENTES	REGALOS	POBLACIÓN	NÚMERO DE MUJERES	IPC ALIM Y BEBI NO ALC	TASA DE RIESGO DE POBREZA
MÁRGENES	1	0,80364	0,516965	0,861433	-0,112054	0,562942	0,605819	0,887321	0,812133	0,541109	0,532932	-0,101535	0,410145
Nº COMERCIALES	0,80364	1	0,365246	0,620429	0,297001	0,386361	0,408956	0,923173	0,932986	0,874481	0,871071	-0,0698498	0,0853318
INCENTIVOS CONTADO	0,516965	0,365246	1	0,674525	-0,206089	0,591856	0,599845	0,362882	0,542071	0,213624	0,20116	-0,087815	0,378377
INCENTIVOS CRÉDITO	0,861433	0,620429	0,674525	1	-0,233171	0,473548	0,471243	0,73531	0,658557	0,400782	0,39338	0,0151081	0,424205
FRECUENCIA VISITAS	-0,112054	0,297001	-0,206089	-0,233171	1	0,0980671	0,0950571	0,115648	0,172496	0,537048	0,541514	-0,0473155	-0,514385
TICKET MEDIO POR VISITA	0,562942	0,386361	0,591856	0,473548	0,0980671	1	0,944144	0,333727	0,462512	0,247632	0,235062	-0,420293	0,580496
MARGEN TOTAL MEDIO DIARIO	0,605819	0,408956	0,599845	0,471243	0,0950571	0,944144	1	0,355533	0,479811	0,186192	0,17361	-0,31576	0,469954
CLIENTES	0,887321	0,923173	0,362882	0,73531	0,115648	0,333727	0,355533	1	0,897977	0,779243	0,777655	-0,0476611	0,197408
REGALOS	0,812133	0,932986	0,542071	0,658557	0,172496	0,462512	0,479811	0,897977	1	0,784398	0,779898	-0,10083	0,271856
POBLACIÓN	0,541109	0,874481	0,213624	0,400782	0,537048	0,247632	0,186192	0,779243	0,784398	1	0,999792	-0,0974196	-0,113535
NÚMERO DE MUJERES	0,532932	0,871071	0,20116	0,39338	0,541514	0,235062	0,17361	0,777655	0,779898	0,999792	1	-0,0970856	-0,121243
IPC ALIM Y BEBI NO ALC	-0,101535	-0,0698498	-0,087815	0,0151081	-0,0473155	-0,420293	-0,31576	-0,0476611	-0,10083	-0,0974196	-0,0970856	1	-0,32633
TASA DE RIESGO DE POBREZA	0,410145	0,0853318	0,378377	0,424205	-0,514385	0,580496	0,469954	0,197408	0,271856	-0,113535	-0,121243	-0,32633	1

Tabla 23 - Matriz de correlación.

INTERPRETACIÓN

Si se investiga mediante la multicolinealidad, todas las variables independientes para descubrir cuáles de ellas están relacionadas, se presentará la anterior matriz. En este caso, las casillas destacadas en color rojo indican que las cifras son superiores a 0,7. Esto último, quiere decir, que las variables X en los casos anteriormente expuestos, presentan una correlación entre ellas, por lo que *existiría multicolinealidad*.

MATRIZ INVERSA DE CORRELACIÓN

MÁRGENES	85,91919987	23,08766155	33,229683	-38,29427286	14,56111949	1,805948423	-25,47066548	-50,0750859	-33,589703	-1029,0826	1035,00395	6,478360481	14,65451163
Nº COMERCIALES	23,08766155	39,16818982	18,051608	-15,85767234	7,10836239	2,982020297	-13,62889659	-15,42462006	-30,139184	-432,72034	423,556234	2,551689591	9,710827217
INCENTIVOS CONTADO	33,22968347	18,05160803	20,665593	-20,01528621	8,109360325	-0,745571224	-11,22807636	-14,13419208	-24,5964	-466,40005	465,918496	2,977770277	9,36789223
INCENTIVOS CRÉDITO	-38,29427286	-15,85767234	-20,01529	24,66978548	-6,907685528	-3,406313137	14,27061391	14,8154906	22,6129027	488,278561	-488,09843	-3,88682455	-8,197093905
FRECUENCIA VISITAS	14,56111949	7,10836239	8,1093603	-6,907685528	8,348573136	-1,960971555	-6,732445846	-4,783096373	-8,9829083	-104,24074	99,9302175	0,514350544	5,714790132
TICKET MEDIO POR VISITA	1,805948423	2,982020297	-0,745571	-3,406313137	-1,960971555	32,89047531	-27,16603954	3,434894431	6,47072944	-87,011994	74,3355976	2,075260848	-9,228972848
MARGEN TOTAL MEDIO DIARIO	-25,47066548	-13,62889659	-11,22808	14,27061391	-6,732445846	-27,16603954	35,06438771	8,245515224	6,5844799	264,886184	-250,45171	-2,49549001	1,10870574
CLIENTES	-50,0750859	-15,42462006	-14,13419	14,8154906	-4,783096373	3,434894431	8,245515224	48,22182443	11,4368224	695,049608	-705,13731	-3,65098127	-7,812276623
REGALOS	-33,58970261	-30,13918392	-24,5964	22,61290266	-8,982908345	6,470729443	6,5844799	11,43682242	42,5987356	512,462312	-514,17583	-3,55107677	-14,56370067
POBLACIÓN	-1029,082564	-432,7203381	-466,4001	488,278561	-104,2407421	-87,01199403	264,8861836	695,0496084	512,462312	19997,6613	-20109,436	-119,866778	-182,5856895
NÚMERO DE MUJERES	1035,003947	423,5562337	465,9185	-488,0984333	99,93021754	74,33559757	-250,4517056	-705,1373096	-514,17583	-20109,436	20239,4343	120,7462374	185,2178621
IPC ALIM Y BEBI NO ALC	6,478360481	2,551689591	2,9777703	-3,886824549	0,514350544	2,075260848	-2,495490013	-3,650981269	-3,5510768	-119,86678	120,746237	2,207909076	1,317125076
TASA DE RIESGO DE POBREZA	14,65451163	9,710827217	9,3678922	-8,197093905	5,714790132	-9,228972848	1,10870574	-7,812276623	-14,563701	-182,58569	185,217862	1,317125076	9,527235415

Tabla 24 - Matriz inversa de correlación.

INTERPRETACIÓN

Como se puede apreciar, después de hallar multicolinealidad mediante el primer método de matriz de correlación, se ha realizado este segundo método para conocer qué componente se debería eliminar del modelo. En esta ocasión, se han señalado en rojo todos los valores superiores a 10 en la diagonal, los cuales indicarían los *casos multicolineales*. Asimismo, se ha resaltado en color amarillo, la mayor cifra entre todas las que presentaban una correlación. Se puede contemplar, que el elemento explicativo con mayor valor en la diagonal es, el número de mujeres. Por lo que se eliminará la variable del modelo, con el fin de evitar posibles relaciones entre variables y asegurar un mejor modelo.

ÍNDICE DE ACONDICIONAMIENTO

COMPONENTE NÚMERO	EIGENVALOR	PORCENTAJE DE VARIANZA	PORCENTAJE ACUMULADO
1	6,55	50,40	50,40
2	2,80	21,52	71,92
3	1,50	11,54	83,46
4	0,86	6,65	90,11
5	0,50	3,81	93,92
6	0,35	2,72	96,64
7	0,22	1,71	98,35
8	0,12	0,92	99,27
9	0,04	0,34	99,61
10	0,02	0,17	99,78
11	0,02	0,12	99,90
12	0,01	0,10	100,00
13	2,47E-05	0,00	100,00

Tabla 25 - Valores máximo y mínimo.

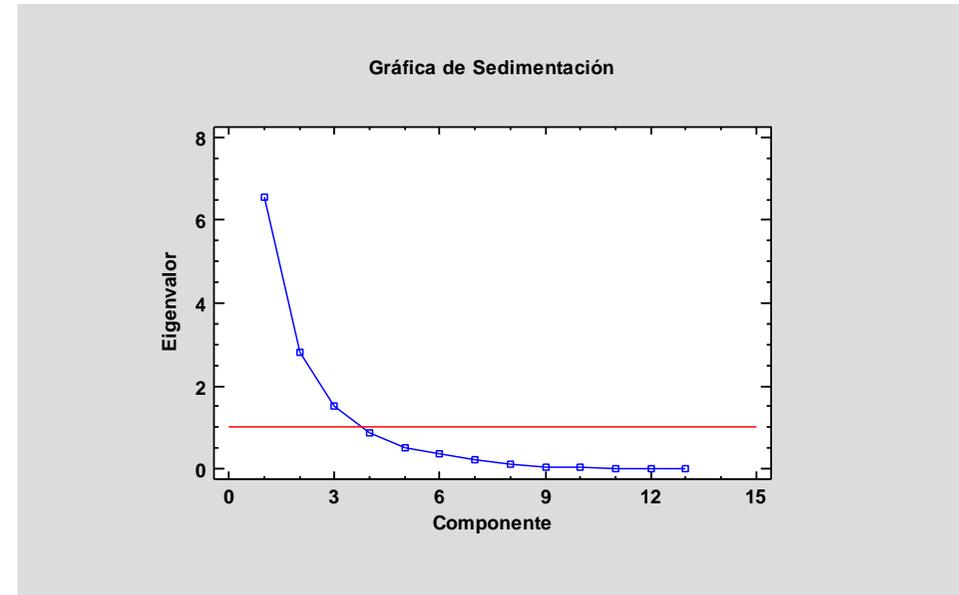


Ilustración 44 – Gráfica de sedimentación.

$$I.C. = \sqrt{\frac{\text{Autovalor máximo}}{\text{Autovalor mínimo}}} = \sqrt{\frac{6,55}{0,00002}} = 515,50$$

INTERPRETACIÓN

Existe una *multicolinealidad elevada* en el modelo, puesto que el Índice de acondicionamiento muestra unos valores superiores a 30.

MATRIZ DE CORRELACIÓN ELIMINANDO LA VARIABLE: NÚMERO DE MUJERES

	MÁRGENES	Nº COMERCIALES	INCENTIVOS CONTADO	INCENTIVOS CRÉDITO	FRECUENCIA VISITAS	TICKET MEDIO POR VISITA	MARGEN TOTAL MEDIO DIARIO	CLIENTES	REGALOS	POBLACIÓN	IPC ALIM Y BEBI NO ALC	TASA DE RIESGO DE POBREZA
MÁRGENES	1	0,80364	0,516965	0,861433	-0,112054	0,562942	0,605819	0,887321	0,812133	0,541109	-0,101535	0,410145
Nº COMERCIALES	0,80364	1	0,365246	0,620429	0,297001	0,386361	0,408956	0,923173	0,932986	0,874481	-0,0698498	0,0853318
INCENTIVOS CONTADO	0,516965	0,365246	1	0,674525	-0,206089	0,591856	0,599845	0,362882	0,542071	0,213624	-0,087815	0,378377
INCENTIVOS CRÉDITO	0,861433	0,620429	0,674525	1	-0,233171	0,473548	0,471243	0,73531	0,658557	0,400782	0,0151081	0,424205
FRECUENCIA VISITAS	-0,112054	0,297001	-0,206089	-0,233171	1	0,0980671	0,0950571	0,115648	0,172496	0,537048	-0,0473155	-0,514385
TICKET MEDIO POR VISITA	0,562942	0,386361	0,591856	0,473548	0,0980671	1	0,944144	0,333727	0,462512	0,247632	-0,420293	0,580496
MARGEN TOTAL MEDIO DIARIO	0,605819	0,408956	0,599845	0,471243	0,0950571	0,944144	1	0,355533	0,479811	0,186192	-0,31576	0,469954
CLIENTES	0,887321	0,923173	0,362882	0,73531	0,115648	0,333727	0,355533	1	0,897977	0,779243	-0,0476611	0,197408
REGALOS	0,812133	0,932986	0,542071	0,658557	0,172496	0,462512	0,479811	0,897977	1	0,784398	-0,10083	0,271856
POBLACIÓN	0,541109	0,874481	0,213624	0,400782	0,537048	0,247632	0,186192	0,779243	0,784398	1	-0,0974196	-0,113535
IPC ALIM Y BEBI NO ALC	-0,101535	-0,0698498	-0,087815	0,0151081	-0,0473155	-0,420293	-0,31576	-0,0476611	-0,10083	-0,0974196	1	-0,32633
TASA DE RIESGO DE POBREZA	0,410145	0,0853318	0,378377	0,424205	-0,514385	0,580496	0,469954	0,197408	0,271856	-0,113535	-0,32633	1

Tabla 26 - Matriz de correlación eliminando la variable: Número de mujeres.

INTERPRETACIÓN

Después de suprimir la variable número de mujeres, todavía se pueden apreciar diversos casos en los que *existe multicolinealidad*. Por lo que se deberá de rehacer la matriz inversa de correlación para conocer qué variable se debería suprimir del modelo.

MATRIZ INVERSA DE CORRELACIÓN ELIMINANDO LA VARIABLE: NÚMERO DE MUJERES

MÁRGENES	32,99118047	1,427847963	9,403549027	-13,3339	9,450889231	-1,995424472	-12,66306917	-14,01578265	-7,295784971	-0,7264606	0,30364085	5,18284282
Nº COMERCIALES	1,427847963	30,30431148	8,30120292	-5,643102	5,017095115	1,426378706	-8,387624542	-0,668016739	-19,37888416	-11,884611	0,02479972	5,83472184
INCENTIVOS CONTADO	9,403549027	8,30120292	9,93999443	-8,779098	5,808933544	-2,456801357	-5,462594955	2,098302867	-12,75990143	-3,4741551	0,19815181	5,10411556
INCENTIVOS CRÉDITO	-13,33390105	-5,643101556	-8,779098248	12,898702	-4,497747487	-1,613620308	8,230668162	-2,189748344	10,21293062	3,31519012	-0,974883	-3,7303412
FRECUENCIA VISITAS	9,450889231	5,017095115	5,808933544	-4,497747	7,855177509	-2,327996262	-5,495865166	-1,301550212	-6,444215698	-4,9523776	-0,0818221	4,80029514
TICKET MEDIO POR VISITA	-1,995424472	1,426378706	-2,456801357	-1,61362	-2,327996262	32,61745478	-26,246178	6,024729826	8,359199614	-13,153855	1,63178285	-9,9092429
MARGEN TOTAL MEDIO DIARIO	-12,66306917	-8,387624542	-5,462594955	8,2306682	-5,495865166	-26,246178	31,96518761	-0,480165512	0,221840861	16,0431326	-1,0013227	3,40067343
CLIENTES	-14,01578265	-0,668016739	2,098302867	-2,189748	-1,301550212	6,024729826	-0,480165512	23,65500016	-6,476947275	-5,5585953	0,55579031	-1,3593282
REGALOS	-7,295784971	-19,37888416	-12,75990143	10,212931	-6,444215698	8,359199614	0,221840861	-6,476947275	29,53627621	1,58903511	-0,4835603	-9,8583049
POBLACIÓN	-0,726460561	-11,88461149	-3,474155113	3,3151901	-4,952377608	-13,15385494	16,04313255	-5,558595295	1,589035113	17,3882915	0,10390438	1,44251597
IPC ALIM Y BEBI NO ALC	0,303640847	0,024799717	0,198151809	-0,974883	-0,081822136	1,631782854	-1,001322703	0,555790312	-0,483560336	0,10390438	1,48755031	0,2121357
TASA DE RIESGO DE POBREZA	5,182842821	5,834721837	5,104115558	-3,730341	4,800295142	-9,909242873	3,400673432	-1,359328228	-9,858304886	1,44251597	0,2121357	7,83224454

Tabla 27 - Matriz inversa de correlación eliminando la variable: Número de mujeres

INTERPRETACIÓN

A continuación, se ha elaborado la matriz inversa de correlación sin la variable número de mujeres, puesto que en la matriz de correlación *existía multicolinealidad*. Se puede constatar que la multicolinealidad sigue revelándose mediante valores superiores a 10 en la diagonal. Márgenes, es el elemento que mayor R_{ii}^{-1} tiene, pero el segundo más alto es ticket medio por visita. Ambos tienen una diferencia menor a uno, por lo que se realizarán dos modelos diferentes por separado y se averiguará cuál de las dos variables será más conveniente quitar.

ÍNDICE DE ACONDICIONAMIENTO ELIMINANDO LA VARIABLE: MUJERES

COMPONENTE NÚMERO	EIGENVALOR	PORCENTAJE DE VARIANZA	PORCENTAJE ACUMULADO
1	6,03	50,24	50,24
2	2,41	20,07	70,30
3	1,49	12,39	82,69
4	0,85	7,11	89,81
5	0,49	4,06	93,87
6	0,33	2,77	96,64
7	0,21	1,77	98,42
8	0,10	0,80	99,22
9	0,04	0,37	99,59
10	0,02	0,18	99,77
11	0,01	0,13	99,90
12	0,01	0,10	100,00

Tabla 28 - Valores máximo y mínimo

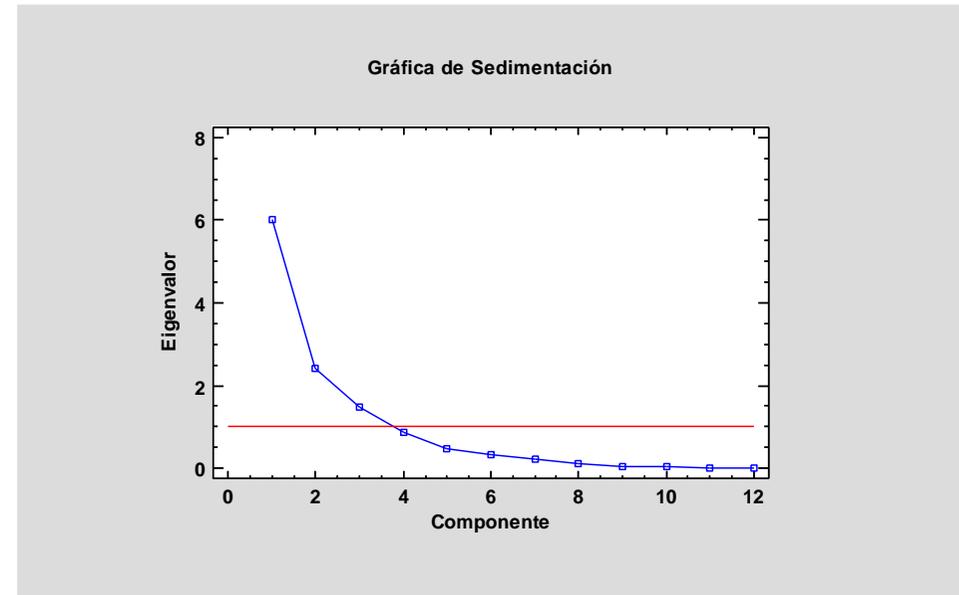


Ilustración 45 - Gráfica de sedimentación.

$$I. C. = \sqrt{\frac{\text{Autovalor máximo}}{\text{Autovalor mínimo}}} = \sqrt{\frac{6,03}{0,01}} = 22,12$$

INTERPRETACIÓN

Existe una *multicolinealidad elevada* en el modelo, puesto que el Índice de acondicionamiento muestra unos valores superiores a 20.

MODELOS A Y B:

A. MATRIZ DE CORRELACIÓN ELIMINANDO LAS VARIABLES: NÚMERO DE MUJERES Y MÁRGENES

$Ventas\ de\ Dulcesol = B_0 + B_1 \cdot \text{Número de comerciales} + B_2 \cdot \text{Incentivo contado} +$
 $B_3 \cdot \text{Incentivo crédito} + B_4 \cdot \text{Frecuencia visitas} + B_5 \cdot \text{Ticket medio por visita} + B_6 \cdot$
 $\text{Margen total medio diario} + B_7 \cdot \text{Número de clientes} + B_8 \cdot \text{Regalos} + B_9 \cdot$
 $\text{Población} + B_{10} \cdot \text{IPC Alimentos y bebidas no alcohólicas} + B_{11} \cdot$
 $\text{Tasa de riesgo de pobreza} + U \rightarrow R^2 = 96,662\%$

B. MATRIZ INVERSA DE CORRELACIÓN ELIMINANDO LA VARIABLES NÚMERO DE MUJERES Y TICKET MEDIO POR VISITA

$Ventas\ de\ Dulcesol = B_0 + B_1 * \text{Márgenes} + B_2 * \text{Número de comerciales} + B_3 *$
 $\text{Incentivo contado} + B_4 * \text{Incentivo crédito} + B_5 * \text{Frecuencia visitas} + B_6 *$
 $\text{Margen total medio diario} + B_7 * \text{Número de clientes} + B_8 * \text{Regalos} + B_9 \cdot$
 $\text{Población} + B_{10} * \text{IPC Alimentos y bebidas no alcohólicas} + B_{11} *$
 $\text{Tasa de riesgo de pobreza} + U \rightarrow R^2 = 99,6346\%$

Después de seguir este procedimiento, se puede observar que el R2 más alto es el de **ticket medio por visita**, por lo que será la variable a eliminar.

MATRIZ DE CORRELACIÓN ELIMINANDO LAS VARIABLES: MUJERES Y TICKET MEDIO DE VISITA

	MÁRGENES	Nº COMERCIALES	INCENTIVOS CONTADO	INCENTIVOS CRÉDITO	FRECUENCIA VISITAS	MARGEN TOTAL MEDIO DIARIO	CLIENTES	REGALOS	POBLACIÓN	IPC ALIM Y BEBI NO ALC	TASA DE RIESGO DE POBREZA
MÁRGENES	1	0,80364	0,516965	0,861433	-0,112054	0,605819	0,887321	0,812133	0,541109	-0,101535	0,410145
Nº COMERCIALES	0,80364	1	0,365246	0,620429	0,297001	0,408956	0,923173	0,932986	0,874481	-0,0698498	0,0853318
INCENTIVOS CONTADO	0,516965	0,365246	1	0,674525	-0,206089	0,599845	0,362882	0,542071	0,213624	-0,087815	0,378377
INCENTIVOS CRÉDITO	0,861433	0,620429	0,674525	1	-0,233171	0,471243	0,73531	0,658557	0,400782	0,0151081	0,424205
FRECUENCIA VISITAS	-0,112054	0,297001	-0,206089	-0,233171	1	0,0950571	0,115648	0,172496	0,537048	-0,0473155	-0,514385
MARGEN TOTAL MEDIO DIARIO	0,605819	0,408956	0,599845	0,471243	0,0950571	1	0,355533	0,479811	0,186192	-0,31576	0,469954
CLIENTES	0,887321	0,923173	0,362882	0,73531	0,115648	0,355533	1	0,897977	0,779243	-0,0476611	0,197408
REGALOS	0,812133	0,932986	0,542071	0,658557	0,172496	0,479811	0,897977	1	0,784398	-0,10083	0,271856
POBLACIÓN	0,541109	0,874481	0,213624	0,400782	0,537048	0,186192	0,779243	0,784398	1	-0,0974196	-0,113535
IPC ALIM Y BEBI NO ALC	-0,101535	-0,0698498	-0,087815	0,0151081	-0,0473155	-0,31576	-0,0476611	-0,10083	-0,0974196	1	-0,32633
TASA DE RIESGO DE POBREZA	0,410145	0,0853318	0,378377	0,424205	-0,514385	0,469954	0,197408	0,271856	-0,113535	-0,32633	1

Tabla 29 - Matriz de correlación eliminando las variables: Mujeres y ticket medio de visita.

INTERPRETACIÓN

Como se puede observar, sigue *hallándose multicolinealidad* después de eliminar las variables previamente mencionadas, por lo que se realizará la matriz inversa de correlación. El propósito de esto será, revelar qué componente se ha de eliminar del modelo en esta ocasión, con el fin de que deje de existir una relación entre las variables.

MATRIZ INVERSA DE CORRELACIÓN ELIMINANDO LAS VARIABLES MUJERES Y TICKET MEDIO DE VISITA

MÁRGENES	32,8691072	1,51510894	9,25325031	-13,4326168	9,30847037	-14,2687207	-13,6472103	-6,78439769	-1,53116848	0,40346775	4,57662919
Nº COMERCIALES	1,51510894	30,2419352	8,40864015	-5,57253708	5,11889963	-7,23986517	-0,93148139	-19,7444364	-11,3093864	-0,04655902	6,26805832
INCENTIVOS CONTADO	9,25325031	8,40864015	9,75494403	-8,90063883	5,63358496	-7,43950107	2,55209561	-12,1302725	-4,46492551	0,32106043	4,35773478
INCENTIVOS CRÉDITO	-13,4326168	-5,57253708	-8,90063883	12,818874	-4,61291595	6,93224172	-1,89169853	10,6264691	2,66445488	-0,89415698	-4,220562
FRECUENCIA VISITAS	9,30847037	5,11889963	5,63358496	-4,61291595	7,6890221	-7,36912611	-0,87154889	-5,84759694	-5,89120393	0,03464263	4,09304559
MARGEN TOTAL MEDIO DIARIO	-14,2687207	-7,23986517	-7,43950107	6,93224172	-7,36912611	10,8457635	4,36773364	6,94820999	5,45866419	0,31171854	-4,57296381
CLIENTES	-13,6472103	-0,93148139	2,55209561	-1,89169853	-0,87154889	4,36773364	22,5421797	-8,02096472	-3,12896298	0,25438571	0,4709958
REGALOS	-6,78439769	-19,7444364	-12,1302725	10,6264691	-5,84759694	6,94820999	-8,02096472	27,393981	4,96010437	-0,90175356	-7,31876464
POBLACIÓN	-1,53116848	-11,3093864	-4,46492551	2,66445488	-5,89120393	5,45866419	-3,12896298	4,96010437	12,0836502	0,76196415	-2,5536494
IPC ALIM Y BEBI NO ALC	0,40346775	-0,04655902	0,32106043	-0,89415698	0,03464263	0,31171854	0,25438571	-0,90175356	0,76196415	1,40591564	0,70787433
TASA DE RIESGO DE POBREZA	4,57662919	6,26805832	4,35773478	-4,220562	4,09304559	-4,57296381	0,4709958	-7,31876464	-2,5536494	0,70787433	4,82179829

Tabla 30 - Matriz inversa de correlación eliminando las variables: Mujeres y ticket medio de visita.

INTERPRETACIÓN

Después de haber encontrado multicolinealidad mediante el primer método de matriz de correlación, se ha realizado esta sistemática y se observa que la variable explicativa con mayor valor en la diagonal, y con un valor superior a 10, es márgenes. En consecuencia, se eliminará este elemento del modelo, con el fin de evitar posibles relaciones entre variables.

ÍNDICE DE ACONDICIONAMIENTO SIN VARIABLE: MUJERES Y TICKET MEDIO POR VISITA

COMPONENTE NÚMERO	EIGENVALOR	PORCENTAJE DE	PORCENTAJE
1	5,59	50,85	50,85
2	2,24	20,34	71,19
3	1,23	11,19	82,39
4	0,79	7,16	89,54
5	0,45	4,12	93,67
6	0,33	3,02	96,69
7	0,21	1,91	98,59
8	0,08	0,69	99,28
9	0,04	0,40	99,68
10	0,02	0,20	99,88
11	0,01	0,12	100,00

Tabla 31 - Valores máximos y mínimos.

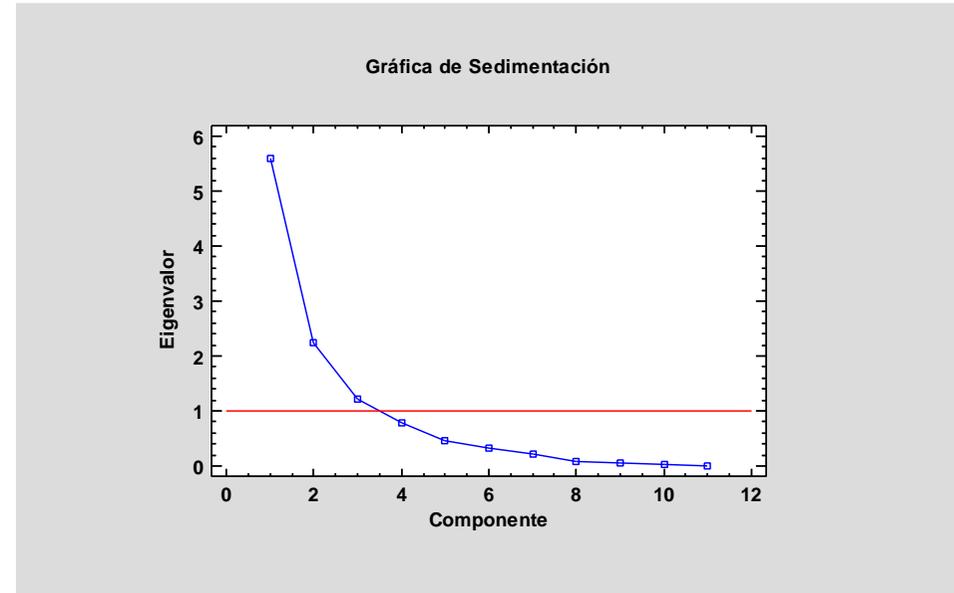


Ilustración 46 - Gráfica de sedimentación.

$$I.C. = \sqrt{\frac{\text{Autovalor máximo}}{\text{Autovalor mínimo}}} = \sqrt{\frac{5,59}{0,01}} = 20,95$$

INTERPRETACIÓN

Existe una *multicolinealidad elevada* en el modelo, puesto que el índice de acondicionamiento muestra un valor superior a 20.

MATRIZ DE CORRELACIÓN ELIMINANDO LAS VARIABLES: MUJERES, TICKET MEDIO POR VISITA Y MÁRGENES

	Nº COMERCIALES	INCENTIVOS CONTADO	INCENTIVOS CRÉDITO	FRECUENCIA VISITAS	MARGEN TOTAL MEDIO DIARIO	CLIENTES	REGALOS	POBLACIÓN	IPC ALIM Y BEBI NO ALC	TASA DE RIESGO DE POBREZA
Nº COMERCIALES	1	0,365246	0,620429	0,297001	0,408956	0,923173	0,932986	0,874481	-0,0698498	0,0853318
INCENTIVOS CONTADO	0,365246	1	0,674525	-0,206089	0,599845	0,362882	0,542071	0,213624	-0,087815	0,378377
INCENTIVOS CRÉDITO	0,620429	0,674525	1	-0,233171	0,471243	0,73531	0,658557	0,400782	0,0151081	0,424205
FRECUENCIA VISITAS	0,297001	-0,206089	-0,233171	1	0,0950571	0,115648	0,172496	0,537048	-0,0473155	-0,514385
MARGEN TOTAL MEDIO DIARIO	0,408956	0,599845	0,471243	0,0950571	1	0,355533	0,479811	0,186192	-0,31576	0,469954
CLIENTES	0,923173	0,362882	0,73531	0,115648	0,355533	1	0,897977	0,779243	-0,0476611	0,197408
REGALOS	0,932986	0,542071	0,658557	0,172496	0,479811	0,897977	1	0,784398	-0,10083	0,271856
POBLACIÓN	0,874481	0,213624	0,400782	0,537048	0,186192	0,779243	0,784398	1	-0,0974196	-0,113535
IPC ALIM Y BEBI NO ALC	-0,0698498	-0,087815	0,0151081	-0,0473155	-0,31576	-0,0476611	-0,10083	-0,0974196	1	-0,32633
TASA DE RIESGO DE POBREZA	0,0853318	0,378377	0,424205	-0,514385	0,469954	0,197408	0,271856	-0,113535	-0,32633	1

Tabla 32 - Matriz de correlación eliminando las variables: Mujeres, ticket medio por visita y márgenes.

INTERPRETACIÓN

Después de rehacer el modelo, reelaborar la tabla y posteriormente suprimir las variables mujeres, márgenes y ticket medio por visita, se puede verificar que diversas variables predictoras *siguen correlacionándose* entre sí. Por esto mismo, se examinará qué parámetro será el que se tenga que eliminar del modelo para evitar este conflicto entre ellos.

MATRIZ INVERSA DE CORRELACIÓN ELIMINANDO LAS VARIABLES: MUJERES, TICKET MEDIO POR VISITA Y MÁRGENES

Nº COMERCIALES	30,17209589	7,982109486	-4,95335756	4,689823585	-6,582145258	-0,302410142	-19,43170807	-11,23880684	-0,065157	6,05709756
INCENTIVOS CONTADO	7,982109486	7,14998611	-5,11911337	3,01308159	-3,422597192	6,394033021	-10,22034141	-4,033873792	0,20747693	3,06933059
INCENTIVOS CRÉDITO	-4,953357564	-5,11911337	7,32936696	-0,808823705	1,101044162	-7,468903438	7,853889594	2,038712311	-0,7292718	-2,3502311
FRECUENCIA VISITAS	4,689823585	3,01308159	-0,80882371	5,052880506	-3,328250805	2,993315808	-3,92626806	-5,45757985	-0,0796187	2,79695267
MARGEN TOTAL MEDIO DIARIO	-6,582145258	-3,42259719	1,10104416	-3,328250805	4,651607162	-1,556620496	4,003053159	4,793972718	0,48686685	-2,5862155
CLIENTES	-0,302410142	6,394033021	-7,46890344	2,993315808	-1,556620496	16,87587593	-10,83783775	-3,764702125	0,42190499	2,37120624
REGALOS	-19,43170807	-10,2203414	7,85388959	-3,92626806	4,003053159	-10,83783775	25,99363712	4,644061232	-0,8184752	-6,3741186
POBLACIÓN	-11,23880684	-4,03387379	2,03871231	-5,45757985	4,793972718	-3,764702125	4,644061232	12,01232248	0,78075922	-2,3404525
IPC ALIM Y BEBI NO ALC	-0,065156955	0,207476933	-0,72927183	-0,079618685	0,486866847	0,421904995	-0,818475191	0,780759222	1,40096308	0,65169629
TASA DE RIESGO DE POBREZA	6,057097558	3,069330588	-2,35023112	2,796952666	-2,586215491	2,37120624	-6,374118582	-2,340452544	0,65169629	4,18455754

Tabla 33 - Matriz inversa de correlación eliminando las variables: Mujeres, ticket medio por visita y márgenes.

INTERPRETACIÓN

Esta matriz muestra que *existen varias variables con valores superiores a 10* en la diagonal, pero la que posee el valor de R_{ii}^{-1} más alto, es el parámetro número de comerciales. Por lo que se prescindirá de esta última y se volverá a rehacer el modelo.

ÍNDICE DE ACONDICIONAMIENTO SIN VARIABLE: MUJERES, TICKET MEDIO POR VISITA Y MÁRGENES

COMPONENTE NÚMERO	EIGENVALOR	PORCENTAJE DE VARIANZA	PORCENTAJE ACUMULADO
1	4,75	47,55	47,55
2	2,21	22,12	69,67
3	1,22	12,15	81,82
4	0,78	7,79	89,61
5	0,44	4,35	93,96
6	0,26	2,64	96,60
7	0,21	2,06	98,66
8	0,08	0,76	99,42
9	0,04	0,43	99,85
10	0,02	0,16	100,00

Tabla 35 - Valores máximos y mínimos.

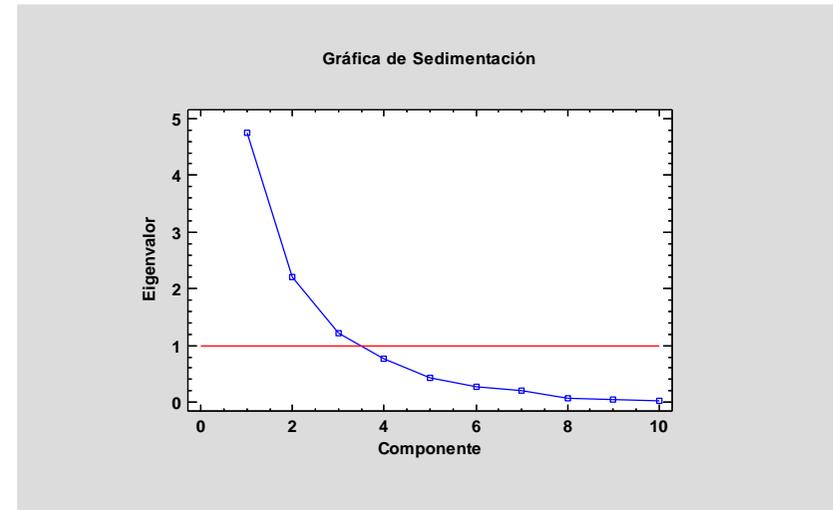


Tabla 34 - Gráfica de sidementación.

$$I. C. = \sqrt{\frac{\text{Autovalor máximo}}{\text{Autovalor mínimo}}} = \sqrt{\frac{4,75}{0,02}} = 17,51$$

INTERPRETACIÓN

Existe una *multicolinealidad moderada* en el modelo, puesto que el índice de acondicionamiento muestra un valor inferior a 20 pero superior a 10.

MATRIZ DE CORRELACIÓN ELIMINANDO LAS VARIABLES: MUJERES, TICKET MEDIO POR VISITA, MÁRGENES Y NÚMERO DE COMERCIALES

	INCENTIVOS CONTADO	INCENTIVOS CRÉDITO	FRECUENCIA VISITAS	MARGEN TOTAL MEDIO DIARIO	CLIENTES	REGALOS	POBLACIÓN	IPC ALIM Y BEBI NO ALC	TASA DE RIESGO DE POBREZA
INCENTIVOS CONTADO	1	0,674525	-0,206089	0,599845	0,362882	0,542071	0,213624	-0,087815	0,378377
INCENTIVOS CRÉDITO	0,674525	1	-0,233171	0,471243	0,73531	0,658557	0,400782	0,0151081	0,424205
FRECUENCIA VISITAS	-0,206089	-0,233171	1	0,0950571	0,115648	0,172496	0,537048	-0,0473155	-0,514385
MARGEN TOTAL MEDIO DIARIO	0,599845	0,471243	0,0950571	1	0,355533	0,479811	0,186192	-0,31576	0,469954
CLIENTES	0,362882	0,73531	0,115648	0,355533	1	0,897977	0,779243	-0,0476611	0,197408
REGALOS	0,542071	0,658557	0,172496	0,479811	0,897977	1	0,784398	-0,10083	0,271856
POBLACIÓN	0,213624	0,400782	0,537048	0,186192	0,779243	0,784398	1	-0,0974196	-0,113535
IPC ALIM Y BEBI NO ALC	-0,087815	0,0151081	-0,0473155	-0,31576	-0,0476611	-0,10083	-0,0974196	1	-0,32633
TASA DE RIESGO DE POBREZA	0,378377	0,424205	-0,514385	0,469954	0,197408	0,271856	-0,113535	-0,32633	1

Tabla 36 - Matriz de correlación eliminando las variables: Mujeres, ticket medio por visita, márgenes y número de comerciales.

INTERPRETACIÓN

El nuevo modelo *sigue mostrando multicolinealidad*, puesto que se pueden apreciar valores mayores a 0,7. Como consecuencia a esto último, se analizará mediante la matriz inversa de correlación la variable a eliminar.

MATRIZ INVERSA DE CORRELACIÓN ELIMINANDO LAS VARIABLES: MUJERES, TICKET MEDIO POR VISITA, MÁRGENES Y NÚMERO DE COMERCIALES

INCENTIVOS CONTADO	5,03829748	-3,808689246	1,77237609	-1,68127288	6,474036441	-5,079630538	-1,060617079	0,224714382	1,466909064
INCENTIVOS CRÉDITO	-3,808689246	6,516173501	-0,03889466	0,02045238	-7,518550157	4,663783129	0,193635686	-0,739968654	-1,355836493
FRECUENCIA VISITAS	1,77237609	-0,038894655	4,323914068	-2,30514985	3,040321168	-0,905885149	-3,710667022	-0,069490962	1,855462917
MARGEN TOTAL MEDIO DIARIO	-1,68127288	0,020452376	-2,30514985	3,2156898	-1,622592296	-0,236039996	2,342188811	0,472652636	-1,264839074
CLIENTES	6,474036441	-7,518550157	3,040321168	-1,6225923	16,87284492	-11,03259869	-3,877346907	0,421251937	2,43191557
REGALOS	-5,079630538	4,663783129	-0,90588515	-0,23604	-11,03259869	13,47905146	-2,594057536	-0,860438167	-2,473171431
POBLACIÓN	-1,060617079	0,193635686	-3,71066702	2,34218881	-3,877346907	-2,594057536	7,825978265	0,756488901	-0,084243705
IPC ALIM Y BEBI NO ALC	0,224714382	-0,739968654	-0,06949096	0,47265264	0,421251937	-0,860438167	0,756488901	1,400822372	0,664776652
TASA DE RIESGO DE POBREZA	1,466909064	-1,355836493	1,855462917	-1,26483907	2,43191557	-2,473171431	-0,084243705	0,664776652	2,968585302

Tabla 37 - Matriz inversa de correlación eliminando las variables: Mujeres, ticket medio por visita, márgenes y número de comerciales.

INTERPRETACIÓN

En esta ocasión tan sólo existen dos variables con valores mayores a 10 en la diagonal. Esto significa que cada vez el modelo está más cerca de no presentar una correlación entre sus variables y con ello, de no producirse malas estimaciones de los parámetros y/o una mala estimación de la desviación típica de la perturbación.

Las variables que *muestran valores superiores a 10* son clientes y regalos. Puesto que clientes es la variable más alta se eliminará del modelo.

ÍNDICE DE ACONDICIONAMIENTO ELIMINANDO LAS VARIABLES: MUJERES, TICKET MEDIO POR VISITA, MÁRGENES Y NÚMERO DE COMERCIALES

COMPONENTE NÚMERO	EIGENVALOR	PORCENTAJE DE VARIANZA	PORCENTAJE ACUMULADO
1	3,96	44,00	44,00
2	2,08	23,09	67,08
3	1,21	13,47	80,56
4	0,76	8,48	89,03
5	0,43	4,81	93,84
6	0,26	2,93	96,77
7	0,19	2,07	98,84
8	0,07	0,83	99,67
9	0,03	0,33	100,00

Tabla 38 - Valores máximos y mínimos.

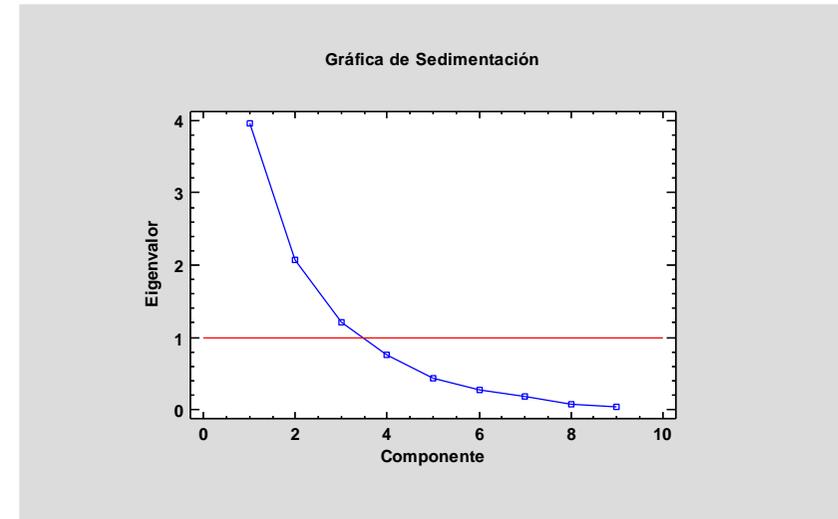


Ilustración 47 - Gráfica de sedimentación.

$$I.C. = \sqrt{\frac{\text{Autovalor máximo}}{\text{Autovalor mínimo}}} = \sqrt{\frac{3,96}{0,03}} = 11,58$$

INTERPRETACIÓN

Existe una *multicolinealidad moderada* en el modelo, puesto que el índice de acondicionamiento muestra un valor inferior a 20 pero superior a 10.

MATRIZ DE CORRELACIÓN ELIMINANDO LAS VARIABLES: MUJERES, TICKET MEDIO POR VISITA, MÁRGENES, NÚMERO DE COMERCIALES Y CLIENTES

	INCENTIVOS CONTADO	INCENTIVOS CRÉDITO	FRECUENCIA VISITAS	MARGEN TOTAL MEDIO DIARIO	REGALOS	POBLACIÓN	IPC ALIM Y BEBI NO ALC	TASA DE RIESGO DE POBREZA
INCENTIVOS CONTADO	1	0,674525	-0,206089	0,599845	0,542071	0,213624	-0,087815	0,378377
INCENTIVOS CRÉDITO	0,674525	1	-0,233171	0,471243	0,658557	0,400782	0,0151081	0,424205
FRECUENCIA VISITAS	-0,206089	-0,233171	1	0,0950571	0,172496	0,537048	-0,0473155	-0,514385
MARGEN TOTAL MEDIO DIARIO	0,599845	0,471243	0,0950571	1	0,479811	0,186192	-0,31576	0,469954
REGALOS	0,542071	0,658557	0,172496	0,479811	1	0,784398	-0,10083	0,271856
POBLACIÓN	0,213624	0,400782	0,537048	0,186192	0,784398	1	-0,0974196	-0,113535
IPC ALIM Y BEBI NO ALC	-0,087815	0,0151081	-0,0473155	-0,31576	-0,10083	-0,0974196	1	-0,32633
TASA DE RIESGO DE POBREZA	0,378377	0,424205	-0,514385	0,469954	0,271856	-0,113535	-0,32633	1

Tabla 39 - Matriz de correlación eliminando las variables: Mujeres, ticket medio por visita, márgenes, número de comerciales y clientes.

INTERPRETACIÓN

Población vs Regalos, es la única variable que presenta multicolinealidad en el modelo. Por lo tanto, se deberá calcular la matriz inversa de correlación para conocer qué variable se debe suprimir para que en el modelo final no existan relaciones entre variables.

MATRIZ INVERSA DE CORRELACIÓN ELIMINANDO LAS VARIABLES: MUJERES, TICKET MEDIO POR VISITA, MÁRGENES, NÚMERO DE COMERCIALES Y CLIENTES

INCENTIVOS CONTADO	2,55423815	-0,923854594	0,60581822	-1,05869135	-0,846470909	0,427103888	0,063081865	0,533793748
INCENTIVOS CRÉDITO	-0,923854594	3,16590289	1,315874342	-0,70257576	-0,252349687	-1,5341125	-0,55225853	-0,272173407
FRECUENCIA VISITAS	0,60581822	1,315874342	3,776078013	-2,01277463	1,082080929	-3,012007138	-0,14539643	1,417255583
MARGEN TOTAL MEDIO DIARIO	-1,058691347	-0,702575758	-2,01277463	3,05965176	-1,296999764	1,969320261	0,513162707	-1,030971728
REGALOS	-0,846470909	-0,252349687	1,082080929	-1,29699976	6,265197813	-5,129327231	-0,584995372	-0,883021776
POBLACIÓN	0,427103888	-1,5341125	-3,01200714	1,96932026	-5,129327231	6,934971496	0,853291776	0,474605757
IPC ALIM Y BEBI NO ALC	0,063081865	-0,55225853	-0,14539643	0,51316271	-0,584995372	0,853291776	1,390305284	0,604060801
TASA DE RIESGO DE POBREZA	0,533793748	-0,272173407	1,417255583	-1,03097173	-0,883021776	0,474605757	0,604060801	2,618068637

Tabla 40 - Matriz inversa de correlación eliminando las variables: Mujeres, ticket medio por visita, márgenes, número de comerciales y clientes.

INTERPRETACIÓN

En este caso se puede apreciar que *no existe multicolinealidad*. Dado que la matriz anterior mostraba una relación entre las variables, se deberá eliminar la que mayor R_{ii}^{-1} posea. En esta ocasión, vuelve a suceder lo anterior, tanto la variable población como regalos tienen entre ellos una diferencia menor a uno. Por lo tanto, se elaborarán dos regresiones y se eliminará la variable que genere un R^2 menor.

ÍNDICE DE ACONDICIONAMIENTO ELIMINANDO LAS VARIABLES: MUJERES, TICKET MEDIO POR VISITA, MÁRGENES, NÚMERO DE COMERCIALES Y CLIENTES

COMPONENTE NÚMERO	EIGENVALOR	PORCENTAJE DE VARIANZA	PORCENTAJE ACUMULADO
1	3,29	41,18	41,18
2	1,97	24,59	65,77
3	1,17	14,68	80,44
4	0,66	8,31	88,75
5	0,43	5,32	94,07
6	0,24	3,03	97,10
7	0,16	2,00	99,10
8	0,07	0,90	100,00

Tabla 41 - Valores máximos y mínimos.

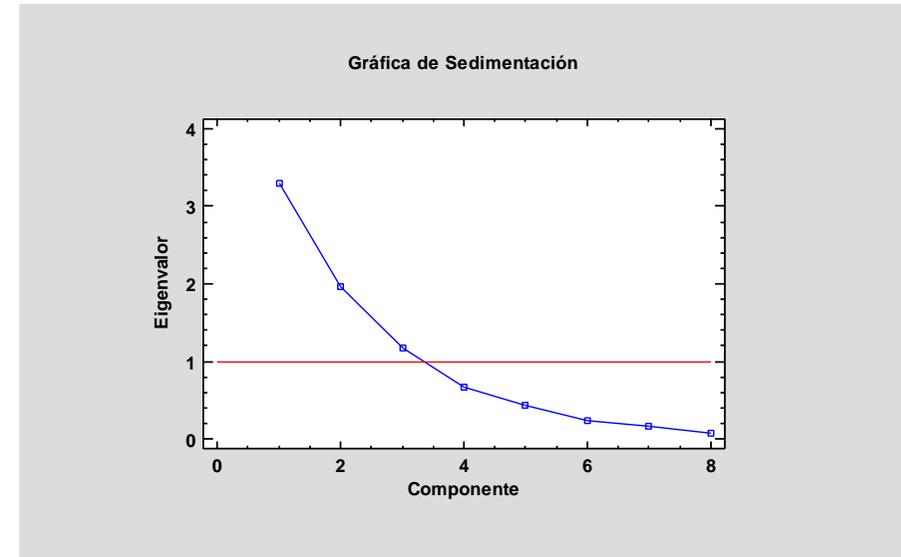


Ilustración 48 - Gráfica de sedimentación.

$$I.C. = \sqrt{\frac{\text{Autovalor máximo}}{\text{Autovalor mínimo}}} = \sqrt{\frac{3,29}{0,07}} = 6,76$$

INTERPRETACIÓN

No existe multicolinealidad en el modelo, puesto que el índice de acondicionamiento muestra un valor inferior a 10.

MODELOS A Y B:

A. MATRIZ DE CORRELACIÓN ELIMINANDO LAS VARIABLES: MUJERES, TICKET MEDIO POR VISITA, MÁRGENES, NÚMERO DE COMERCIALES, CLIENTES Y POBLACIÓN

$$\begin{aligned} \text{Ventas de Dulcesol} &= B_0 + B_1 \cdot \text{Incentivo contado} + B_2 \cdot \text{Incentivo crédito} + B_3 \cdot \\ &\text{Frecuencia visitas} + B_4 \cdot \text{Margen total medio diario} + B_5 \cdot \text{Regalos} + B_6 \cdot \\ &\text{IPC Alimentos y bebidas no alcohólicas} + B_7 \cdot \text{Tasa de riesgo de pobreza} + U = \\ R^2 &= 94,5559\% \end{aligned}$$

B. MATRIZ INVERSA DE CORRELACIÓN ELIMINANDO LA VARIABLES MUJERES, TICKET MEDIO POR VISITA, MÁRGENES, NÚMERO DE COMERCIALES, CLIENTES Y REGALOS

$$\begin{aligned} \text{Ventas de Dulcesol} &= B_0 + B_1 \cdot \text{Incentivo contado} + B_2 \cdot \text{Incentivo crédito} + B_3 \cdot \\ &\text{Frecuencia visitas} + B_4 \cdot \text{Margen total medio diario} + B_5 \cdot \text{Población} + B_6 \cdot \\ &\text{IPC Alimentos y bebidas no alcohólicas} + B_7 \cdot \text{Tasa de riesgo de pobreza} + U = \\ R^2 &= 90,9819\% \end{aligned}$$

Después de analizar ambos modelos, se puede contemplar que el R2 más alto es el modelo A, dónde se elimina la variable población.

MATRIZ DE CORRELACIÓN ELIMINANDO LAS VARIABLES: MUJERES, TICKET MEDIO POR VISITA, MÁRGENES, NÚMERO DE COMERCIALES, CLIENTES Y REGALOS

	INCENTIVOS CONTADO	INCENTIVOS CRÉDITO	FRECUENCIA VISITAS	MARGEN TOTAL MEDIO DIARIO	REGALOS	IPC ALIM Y BEBI NO ALC	TASA DE RIESGO DE POBREZA
INCENTIVOS CONTADO	1	0,674525	-0,206089	0,599845	0,542071	-0,087815	0,378377
INCENTIVOS CRÉDITO	0,674525	1	-0,233171	0,471243	0,658557	0,0151081	0,424205
FRECUENCIA VISITAS	-0,206089	-0,233171	1	0,0950571	0,172496	-0,0473155	-0,514385
MARGEN TOTAL MEDIO DIARIO	0,599845	0,471243	0,0950571	1	0,479811	-0,31576	0,469954
REGALOS	0,542071	0,658557	0,172496	0,479811	1	-0,10083	0,271856
IPC ALIM Y BEBI NO ALC	-0,087815	0,0151081	-0,0473155	-0,31576	-0,10083	1	-0,32633
TASA DE RIESGO DE POBREZA	0,378377	0,424205	-0,514385	0,469954	0,271856	-0,32633	1

Tabla 42 - Matriz de correlación eliminando las variables: Mujeres, ticket medio por visita, márgenes, número de comerciales, clientes y regalos.

INTERPRETACIÓN

No se percibe existencia de multicolinealidad, puesto que los valores son menores 0,7 o mayores a -0,7.

MATRIZ INVERSA DE CORRELACIÓN ELIMINANDO LAS VARIABLES: MUJERES, TICKET MEDIO POR VISITA, MÁRGENES, NÚMERO DE COMERCIALES, CLIENTES Y POBLACIÓN

INCENTIVOS CONTADO	2,527934115	-0,829373252	0,791318616	-1,17997582	-0,530571181	0,010530209	0,504564217
INCENTIVOS CRÉDITO	-0,829373252	2,82653579	0,649576319	-0,26693463	-1,38702847	-0,363498478	-0,167183988
FRECUENCIA VISITAS	0,791318616	0,649576319	2,467898592	-1,15745653	-1,145696106	0,225206524	1,623387062
MARGEN TOTAL MEDIO DIARIO	-1,179975817	-0,266934626	-1,15745653	2,5004249	0,159572631	0,270853884	-1,165745278
REGALOS	-0,530571181	-1,38702847	-1,14569611	0,15957263	2,471382963	0,04612658	-0,531988143
IPC ALIM Y BEBI NO ALC	0,010530209	-0,363498478	0,225206524	0,27085388	0,04612658	1,285314679	0,545664426
TASA DE RIESGO DE POBREZA	0,504564217	-0,167183988	1,623387062	-1,16574528	-0,531988143	0,545664426	2,58558824

Tabla 43 - Matriz inversa de correlación eliminando las variables: Mujeres, ticket medio por visita, márgenes, número de comerciales, clientes y regalos.

INTERPRETACIÓN

Se percata a posteriori de reelaborar esta tabla, que *no existe ninguna variable con multicolinealidad*. Puesto que los valores de la diagonal son menores a 10.

ÍNDICE DE ACONDICIONAMIENTO ELIMINANDO LAS VARIABLES: MUJERES, TICKET MEDIO POR VISITA, MÁRGENES, NÚMERO DE COMERCIALES, CLIENTES Y POBLACIÓN

COMPONENTE NÚMERO	EIGENVALOR	PORCENTAJE DE VARIANZA	PORCENTAJE ACUMULADO
1	3,11	44,39	44,39
2	1,39	19,84	64,23
3	1,17	16,73	80,96
4	0,51	7,27	88,23
5	0,43	6,08	94,30
6	0,24	3,42	97,72
7	0,16	2,28	100,00

Tabla 44 - Valores máximos y mínimos.

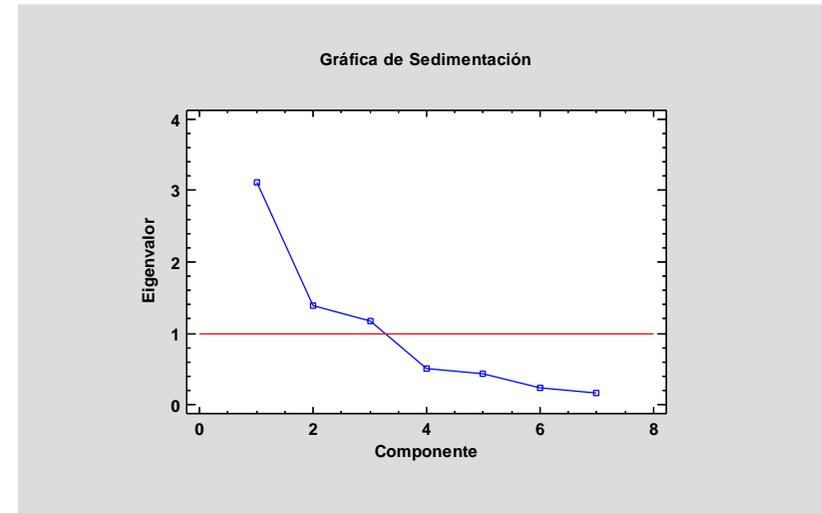


Ilustración 49 - Gráfica de sedimentación.

$$I.C. = \sqrt{\frac{\text{Autovalor máximo}}{\text{Autovalor mínimo}}} = \sqrt{\frac{3,29}{0,07}} = 6,76$$

INTERPRETACIÓN

No existe multicolinealidad en el modelo, puesto que el índice de acondicionamiento muestra un valor inferior a 10.

3.2.3 MODELO FINAL

2018								
ENERO A SEPTIEMBRE								
	NIVEL INTERNO						NIVEL EXTERNO	
	Y	X1	X2	X3	X4	X5	X6	X7
DELEGACIÓN	VENTA PROPIA	INCENTIVOS CONTADO	INCENTIVOS CRÉDITO	FRECUENCIA VISITAS	MARGEN TOTAL MEDIO DIARIO	REGALOS	IPC ALIM Y BEBI NO ALC	TASA DE RIESGO DE POBREZA
ARADUL	333351	1,7	0,64	11,55	8,16	6573,72	2,5	11,3
ASTURSOL	810848	1,7	1,06	9,24	5,99	1983,6	2,4	10,7
DOCES LUCENSES	1284000	1,7	0,85	9,77	10,26	3915,93	2,6	15,9
DULCADIS	2160360	1,91	1,06	8,8	9,56	23017,4	2,7	23,9
DULCOSOL	5917520	3,83	1,7	9,29	15,6	42753,2	2,2	25
DULGALICIA	1074190	1,06	1,06	11,78	9,41	11446,8	2,6	15,9
DULGANDIA	2939750	3,83	1,7	10,73	14,76	36124,4	2,6	21,8
DULMADRID	3997800	1,91	1,28	12,95	10,65	63893,5	2,3	14,4
EUSKASOL	3005750	1,7	0,85	11,63	13,34	29696,3	2,6	11
GADESDUL	2124480	1,7	0,85	8,39	9,8	26770	2,5	26,4
HERMANOS MUÑOZ	798290	1,49	0,64	11,5	11,9	5519,83	2,4	12,8
HISPADUL	6584020	2,34	2,34	9,01	13,63	40614,7	2,5	26,4
LEVANTINA	1262400	1,7	0,85	11,35	10,48	9220,01	2,6	21,8
TINERFEÑA	1250060	1,7	0,85	9,93	9,89	12796,4	2	25,9
VEGODISA	1978430	1,7	0,85	10,55	15,88	10815,2	2	25,9
VIGUESA	1163800	1,7	0,85	8,94	8,56	2104,14	2,6	15,9

Tabla 45 - Modelo final

3.3 ANÁLISIS MULTIVARIADO

3.3.1 MODELO TEÓRICO PROPUESTO

Los parámetros considerados para el modelo, en función de cada variable explicativa son los que se indicarán a continuación. Se clasificarán en niveles externos e internos de la empresa. Los niveles internos, serán aquellos que varíen en función de la actividad de la empresa, y los externos, los que afectarán a la empresa de forma indirecta sin que esta pueda variar su resultado. Para los datos internos se ha obtenido la información de la base de datos de la empresa. En cambio, los datos externos han sido escogidos de la página web del Instituto Nacional de Estadística (INE).

$$\begin{aligned} \text{Ventas de Dulcesol} = & B_0 + B_1 \cdot \text{Incentivo contado} + B_2 \cdot \text{Incentivo crédito} + B_3 \cdot \\ & \text{Frecuencia visitas} + B_4 \cdot \text{Margen total medio diario} + B_5 \cdot \text{Regalos} + B_6 \cdot \\ & \text{IPC Alimentos y bebidas no alcohólicas} + B_7 \cdot \text{Tasa de riesgo de pobreza} + U \end{aligned}$$

Seguidamente, se especificarán las unidades de cada variable a fin de construir una función con los mismos parámetros para poder llegar así a la solución correspondiente:

NIVELES INTERNOS

- **INCENTIVOS CONTADO:** Euros
- **INCENTIVOS CRÉDITO:** Euros
- **NÚMERO DE VISITAS/ FRECUENCIA VISITA:** Unidades de visita
- **MARGEN TOTAL MEDIO DIARIO:** Euros
- **REGALOS:** Euros

NIVELES EXTERNOS

- **IPC ALIMENTOS Y BEBIDAS NO ALCOHÓLICAS:** Euros
- **TASA DE RIESGO DE POBREZA:** Personas

Las **UNIDADES FÍSICAS** se incorporarán a la función para que las unidades sean las mismas:

$$\begin{aligned} \text{Ventas de Dulcesol} = & B_0 + B_1 \cdot \text{Incentivo contado (euros)} + B_2 \cdot \\ & \text{Incentivo crédito (euros)} + B_3 \cdot \text{Frecuencia visitas (unidades)} + B_4 \cdot \\ & \text{Margen total medio diario (euros)} + B_5 \cdot \text{Regalos (euros)} + B_6 \cdot \\ & \text{IPC Alimentos y bebidas no alcohólicas (euros)} + B_7 \cdot \\ & \text{Tasa de riesgo de pobreza (población)} + U \end{aligned}$$

3.3.2 INTERPRETACIÓN DE LOS PARÁMETROS

Los parámetros del modelo serán los siguientes:

B₀ = Valor medio de las ventas del grupo Dulcesol cuando las variables X valen 0 (euros)

B₁ = Es el incremento medio de las ventas del grupo Dulcesol cuando los **incentivos de contado** aumentan en un euro y el resto de las variables explicativas permanecen constantes $\left(\frac{\text{euros}}{\text{euros}}\right)$.

B₂ = Es el incremento medio de las ventas del grupo Dulcesol cuando los **incentivos de crédito** aumentan en un euro y el resto de las variables explicativas permanecen constantes $\left(\frac{\text{euros}}{\text{euros}}\right)$.

B₃ = Es el incremento medio de las ventas del grupo Dulcesol cuando los comerciales realizan **una visita más** a un cliente y el resto de las variables explicativas permanecen constantes $\left(\frac{\text{euros}}{\text{unidades}}\right)$.

B₄ = Es el incremento medio de las ventas del grupo Dulcesol cuando el **margen total diario** aumenta en un euro y el resto de las variables explicativas permanecen constantes $\left(\frac{\text{euros}}{\text{euros}}\right)$.

B₅ = Es el incremento medio de las ventas del grupo Dulcesol cuando la variable **regalo** aumenta en un euro y el resto de las variables explicativas permanecen constantes $\left(\frac{\text{euros}}{\text{euros}}\right)$.

B₆ = Es el incremento medio de las ventas del grupo Dulcesol cuando el **IPC de alimentos y bebidas no alcohólicas** aumenta en un euro y el resto de las variables explicativas permanecen constantes $\left(\frac{\text{euros}}{\text{euros}}\right)$.

B₇ = Es el incremento medio de las ventas del grupo Dulcesol cuando **el riesgo de pobreza** aumenta en una persona y el resto de las variables explicativas permanecen constantes $\left(\frac{\text{euros}}{\text{población}}\right)$.

U = Representa las variables explicativas que **no hemos tenido en cuenta en el modelo**

3.3.3 MODELO AJUSTADO

$$\begin{aligned} \text{Venta Propia} = & 4,23789E6 - 498630 \cdot (\text{Incentivos contado}) + 1,94108E6 \cdot \\ & (\text{Incentivos crédito}) - 436167 \cdot (\text{Frecuencia visitas}) + 243440 \cdot \\ & (\text{Margen total medio diario}) + 53,5899 \cdot (\text{Regalos}) - 528230 \cdot \\ & (\text{IPC ALIM y Bebi No ALC}) - 56393,6 \cdot (\text{Tasa de riesgo de pobreza}) \end{aligned}$$

El R^2 de este modelo es **94,56%**

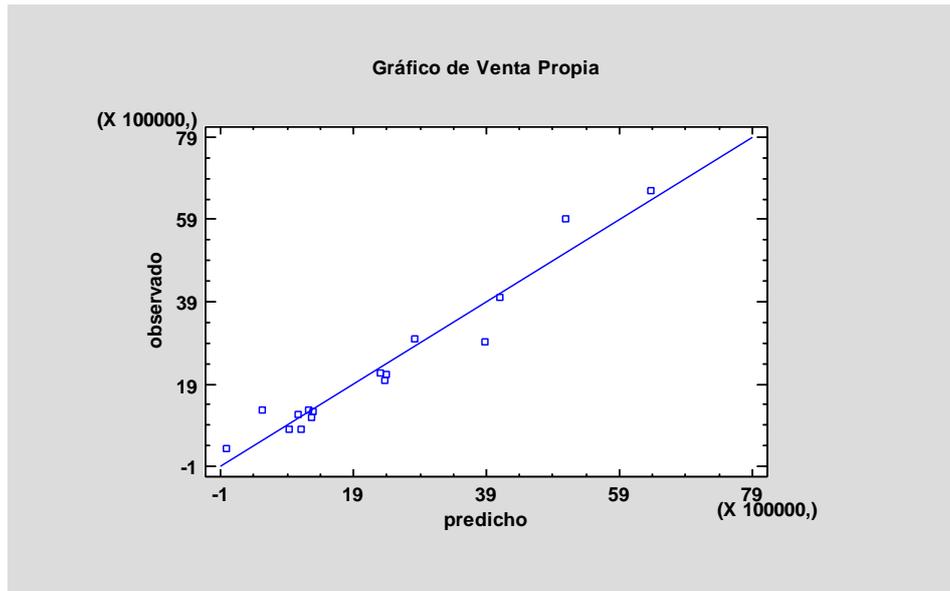


Ilustración 50 – Regresión simple nuevo modelo.

3.3.4 VALIDACIÓN DEL MODELO

SIGNIFICATIVIDAD DEL PARÁMETRO

	PARÁMETRO	VALOR-P
B ₀	CONSTANTE	0,2258
B ₁	INCENTIVOS CONTADO	0,1489
B ₂	INCENTIVOS CRÉDITO	0,0078
B ₃	FRECUENCIA VISITAS	0,0373
B ₄	MARGEN TOTAL MEDIO DIARIO	0,0201
B ₅	REGALOS	0,0034
B ₆	IPC ALIM Y BEBI NO ALC	0,5218
B ₇	TASA DE RIESGO DE POBREZA	0,1912

Tabla 46 - Significatividad del parámetro.

CONTRASTE β_0 :

$$\left. \begin{array}{l} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{array} \right\} \alpha = 0,05$$

P-Valor = 0,2258 > α por lo tanto, **se acepta H₀** y se concluye que la constante, *no es significativa*.

CONTRASTE β_1 :

$$\left. \begin{array}{l} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{array} \right\} \alpha = 0,05$$

P-Valor = 0,1489 > α por lo tanto, **se acepta H₀** y se concluye que la variable incentivos contado, *no es significativa*.

CONTRASTE β_2 :

$$\left. \begin{array}{l} H_0 : \beta_2 = 0 \\ H_1 : \beta_2 \neq 0 \end{array} \right\} \alpha = 0,05$$

P-Valor = 0,0078 < α por lo tanto, **se rechaza H₀** y se concluye que la variable incentivos crédito, *sí es significativa*.

CONTRASTE β_3 :

$$\left. \begin{array}{l} H_0 : \beta_3 = 0 \\ H_1 : \beta_3 \neq 0 \end{array} \right\} \alpha = 0,05$$

P-Valor = 0,0373 < α por lo tanto, **se rechaza H_0** y se concluye que la variable frecuencia visitas, *sí es significativa*.

CONTRASTE β_4 :

$$\left. \begin{array}{l} H_0 : \beta_4 = 0 \\ H_1 : \beta_4 \neq 0 \end{array} \right\} \alpha = 0,05$$

P-Valor = 0,0201 < α por lo tanto, **se rechaza H_0** y se concluye que la variable margen total medio diario, *sí es significativa*.

CONTRASTE β_5 :

$$\left. \begin{array}{l} H_0 : \beta_5 = 0 \\ H_1 : \beta_5 \neq 0 \end{array} \right\} \alpha = 0,05$$

P-Valor = 0,0034 < α por lo tanto, **se rechaza H_0** y se concluye que la variable regalos, *sí es significativa*.

CONTRASTE β_6 :

$$\left. \begin{array}{l} H_0 : \beta_6 = 0 \\ H_1 : \beta_6 \neq 0 \end{array} \right\} \alpha = 0,05$$

P-Valor = 0,5218 > α por lo tanto, **se acepta H_0** y se concluye que la variable IPC alimentos y bebidas no alcohólicas, *no es significativa*.

CONTRASTE β_7 :

$$\left. \begin{array}{l} H_0 : \beta_7 = 0 \\ H_1 : \beta_7 \neq 0 \end{array} \right\} \alpha = 0,05$$

P-Valor = 0,1912 > α por lo tanto, **se acepta H_0** y se concluye que la variable tasa de riesgo de pobreza, *no es significativa*.

SIGNIFICATIVIDAD DEL MODELO

FUENTE	VALOR-P
MODELO	0,0002

Tabla 47 - Significatividad del modelo.

$$\left. \begin{array}{l} H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0 \\ H_1 : \text{Algún } \beta_i \neq 0 \end{array} \right\} \alpha = 0,05$$

P-Valor = 0,0002 < α por lo tanto, **se rechaza la hipótesis nula** y se concluye con que el modelo, *sí es significativo*.

3.3.5 NORMALIDAD DE RESIDUOS

GRÁFICO PROBABILÍSTICO

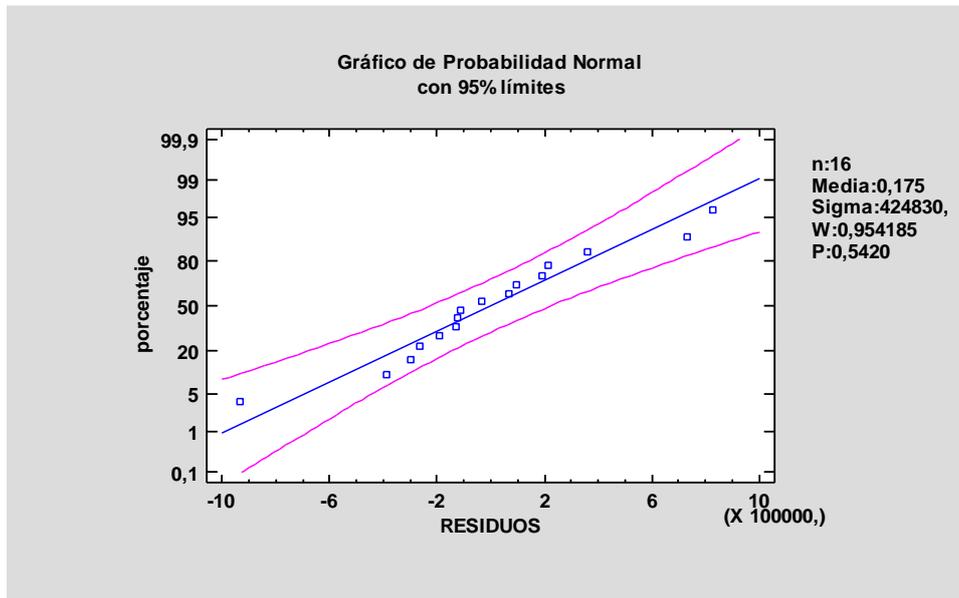


Ilustración 51 - Gráfico probabilístico.

INTERPRETACIÓN

En la ilustración anterior, se puede apreciar que **no existen valores anómalos**, puesto que todos los puntos están próximos a la línea de normalidad. Los puntos que se separan de la línea se dan puesto que el R^2 del modelo no es del 100%. Asimismo, cabe recalcar que ningún valor se sitúa fuera de los intervalos de confianza.

TEST DE NORMALIDAD

PRUEBA	VALOR-P
CHI-CUADRADO	0,689019
ESTADÍSTICO W DE SHAPIRO-WILK	0,542018
VALOR-Z PARA ASIMETRÍA	0,914747

Tabla 48 - Test de normalidad.

$$\left. \begin{array}{l} H_0: \text{Los residuos son normales} \\ H_1: \text{Los residuos no son normales} \end{array} \right\} \alpha = 0,05$$

P-Valor = 0,542018 > α (0,05), por lo que **se acepta H_0** . Los residuos se distribuyen *con normalidad*.

HETEROCEDASTICIDAD

$$\begin{aligned} RESIDUOS^2 = & 3,72035 \cdot 10^{-11} \cdot \text{Incentivos contado} - 8,00401 \cdot 10^{-10} \cdot \\ & \text{Incentivos crédito} + 2,08452 \cdot 10^{-10} \cdot \text{Frecuencia visitas} - 9,71672 \cdot 10^{-9} \cdot \\ & \text{Margen total medio diario} - 3,38529 \cdot 10^{-6} \cdot \text{Regalos} - 2,12165 \cdot 10^{-11} \cdot \\ & \text{IPC ALIM y Bebi No ALC} + 2,81866 \cdot 10^{-8} \cdot \text{Tasa de riesgo de pobreza} \end{aligned}$$

PARÁMETRO	VALOR-P
INCENTIVOS CONTADO	0,0045
INCENTIVOS CRÉDITO	0,6642
FRECUENCIA VISITAS	0,6105
MARGEN TOTAL MEDIO DIARIO	0,7292
REGALOS	0,4198
IPC ALIM Y BEBI NO ALC	0,2424
TASA DE RIESGO DE POBREZA	0,9776

Tabla 49 - Valor P parámetros.

FUENTE	VALOR-P
MODELO	0,0146

Tabla 50 - Valor P modelo.

- **INCENTIVOS CONTADO:**

P-Valor 0,0045 < α se rechaza H_0 , por lo que la variable incentivos contado *genera problemas de heterocedasticidad*.

$$\left. \begin{array}{l} H_0 : \sigma^2 = \text{Constante} \\ H_1 : \sigma^2 \neq \text{Constante} \end{array} \right\} \alpha = 0,05$$

- **INCENTIVOS CRÉDITO:**

P-Valor 0,6642 > α se acepta H_0 , el componente incentivos crédito *no genera problemas de heterocedasticidad*.

$$\left. \begin{array}{l} H_0 : \sigma^2 = \text{Constante} \\ H_1 : \sigma^2 \neq \text{Constante} \end{array} \right\} \alpha = 0,05$$

- **FRECUENCIA VISITAS:**

P-Valor 0,6105 > α se acepta H_0 , la frecuencia de las visitas *no genera problemas de heterocedasticidad*.

$$\left. \begin{array}{l} H_0 : \sigma^2 = \text{Constante} \\ H_1 : \sigma^2 \neq \text{Constante} \end{array} \right\} \alpha = 0,05$$

- **MARGEN TOTAL MEDIO DIARIO:**

P-Valor 0,7292 > α se acepta H_0 , el margen total medio diario *no genera problemas de heterocedasticidad*.

$$\left. \begin{array}{l} H_0 : \sigma^2 = \text{Constante} \\ H_1 : \sigma^2 \neq \text{Constante} \end{array} \right\} \alpha = 0,05$$

- **REGALOS:**

P-Valor 0,4198 > α se acepta H_0 , el regresor anteriormente nombrado, es decir regalos, *no genera problemas de heterocedasticidad*.

$$\left. \begin{array}{l} H_0 : \sigma^2 = \text{Constante} \\ H_1 : \sigma^2 \neq \text{Constante} \end{array} \right\} \alpha = 0,05$$

- **IPC ALIMENTOS Y BEBIDAS NO ALCOHÓLICAS:**

P-Valor 0,2424 > α se acepta H_0 , la variable IPC Alimentos y bebidas no alcohólicas, *no genera problemas de heterocedasticidad*.

$$\left. \begin{array}{l} H_0 : \sigma^2 = \text{Constante} \\ H_1 : \sigma^2 \neq \text{Constante} \end{array} \right\} \alpha = 0,05$$

- **TASA DE RIESGO DE POBREZA:**

P-Valor 0,9776 > α se acepta H_0 , la tasa de riesgo de pobreza *no genera problemas de heterocedasticidad*.

$$\left. \begin{array}{l} H_0 : \sigma^2 = \text{Constante} \\ H_1 : \sigma^2 \neq \text{Constante} \end{array} \right\} \alpha = 0,05$$

AUTOCORRELACIÓN

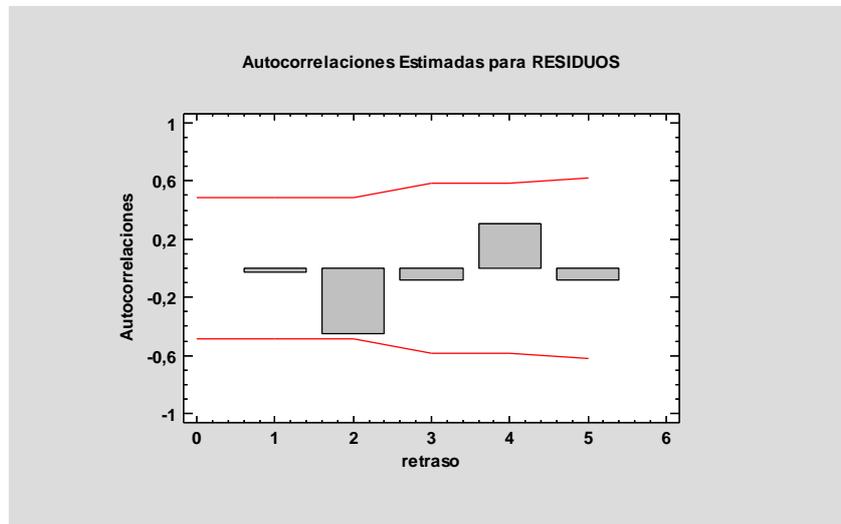


Ilustración 52 - FAS. Autocorrelación.

En la ilustración anterior, se puede contemplar que coeficientes de autocorrelación no superan los límites. Por lo tanto, se puede decir que no existe relación lineal entre el residuo a e_T y e_{T-S} (**FAS**). Es decir, **no existe autocorrelación**.

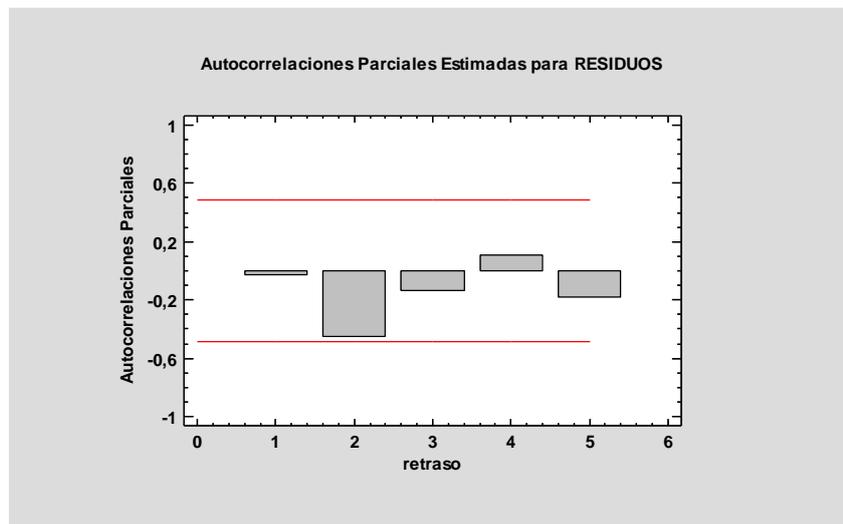


Ilustración 53 - FAP. Autocorrelación.

El **FAP** muestra también que no se superan los límites de autocorrelación. Por lo que **no existe autocorrelación de ningún orden**.

TEST DE AUTOCORRELACIÓN

En la siguiente ilustración, se mostrará dónde se situarán los valores correspondientes al modelo en el estadístico de Durbin-Watson.

- N (número de variables) = 16
- K (número de variables independientes) = 7
- D_U : 2,624
- D_L : 0,398

R-cuadrada = 94,5559 por ciento
 R-cuadrado (ajustado para g.l.) = 89,7923 por ciento
 Error estándar del est. = 581722,
 Error absoluto medio = 310645,
Estadístico Durbin-Watson = 2,00623 (P=0,5552)
 Autocorrelación de residuos en retraso 1 = -0,0284548

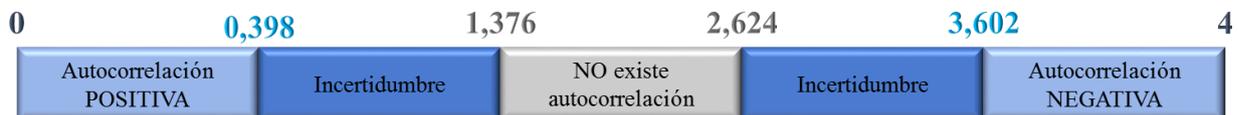


Ilustración 54 - Escala estadístico de Durbin - Watson.

INTERPRETACIÓN:

Se han buscado los valores D_U y D_L en la tabla del estadístico, conociendo los valores N y K. Estos últimos, respectivamente son el número de delegaciones como de la cantidad de variables que existen tras haber eliminado algunos regresores en la parte de multicolinealidad. Posteriormente de obtenerlos, se han situado en el gráfico para comprobar la existencia o no existencia de la autocorrelación.

Después de realizar la regresión múltiple, Statgraphics muestra un **valor de 2,00623** para el estadístico Durbin-Watson. Situando ese valor en la tabla, se puede decir que **no existe autocorrelación**.

A continuación, se estudiará a través del contraste de hipótesis la presencia de autocorrelación.

CONTRASTE DE HIPÓTESIS:

$$\left. \begin{array}{l} H_0: \rho_1 = 0 \\ H_1: \rho_1 \neq 0 \end{array} \right\} \alpha = 0,05$$

P-Valor 0,5552 > α (0,05) por lo tanto, se acepta H_0 y se concluye que no existe autocorrelación de primer orden.

3.4 PUNTOS ANÓMALOS Y PUNTOS INFLUYENTES

3.4.1 PUNTOS ANÓMALOS

RESIDUOS ATÍPICOS		
FILA	DELEGACIONES	RESIDUO ESTUDENTIZADO
5	Dulcosol	2,96
7	Dulgandía	-7,37

Tabla 51 - Residuos atípicos.

El residuo estudentizado de la delegación **Dulcosol** (fila 5) muestra un valor del 2,96. Se trata de un punto anómalo. Pero dado que no es mayor a 3 **no se eliminará del modelo**.

En cuanto a **Dulgandía** se puede observar que el residuo estudentizado presenta un valor muy anómalo. Por lo que **se eliminará del modelo**.

3.4.2 PUNTOS INFLUYENTES

PUNTOS INFLUYENTES			
FILA	DELEGACIONES	INFLUENCIA	DFITS
5	Dulcosol	0,548608	3,26763
7	Dulgandía	0,635927	-9,74455
12	Hispadul	0,839822	2,09712
13	Levantina	0,455263	1,82014
15	Vegodisa	0,647371	-1,55348
Influencia media de un solo punto = 0,5			

Tabla 52 - Puntos influyentes.

A PRIORI	$n_{ij} \geq 2\bar{n}$	1
A POSTERIORI	$ DFITS \geq 2\sqrt{\bar{n}}$	1,4142136

Tabla 53 - Puntos influyentes a priori y a posteriori.

PUNTOS INFLUYENTES A PRIORI

PUNTOS INFLUYENTES A PRIORI			
FILA	DELEGACIONES	INFLUENCIA	¿PUNTO INFLUYENTE A PRIORI?
5	Dulcosol	$0,548608 \leq 2\tilde{n}$	No es un punto influyente a priori
7	Dulgandía	$0,635927 \leq 2\tilde{n}$	No es un punto influyente a priori
12	Hispadul	$0,839822 \leq 2\tilde{n}$	No es un punto influyente a priori
13	Levantina	$0,455263 \leq 2\tilde{n}$	No es un punto influyente a priori
15	Vegodisa	$0,647371 \leq 2\tilde{n}$	No es un punto influyente a priori

Tabla 54 - Puntos influyentes a priori.

PUNTOS INFLUYENTES A POSTERIORI

PUNTOS INFLUYENTES A POSTERIORI			
FILA	DELEGACIONES	DFITS	¿PUNTO INFLUYENTE A POSTERIORI?
5	Dulcosol	$3,26763 \geq 2\sqrt{\tilde{n}}$	Sí es un punto influyente a posteriori
7	Dulgandía	$9,74455 \geq 2\sqrt{\tilde{n}}$	Sí es un punto influyente a posteriori
12	Hispadul	$2,09712 \geq 2\sqrt{\tilde{n}}$	Sí es un punto influyente a posteriori
13	Levantina	$1,82014 \geq 2\sqrt{\tilde{n}}$	Sí es un punto influyente a posteriori
15	Vegodisa	$1,55348 \geq 2\sqrt{\tilde{n}}$	Sí es un punto influyente a posteriori

Tabla 55 - Puntos influyentes a posteriori.

4. TABLA RESUMEN

VARIABLES	ANÁLISIS UNIVARIANTE	ANÁLISIS BIVARIANTE		NUEVO MODELO	ANÁLISIS MULTIVARIANTE	HETEROCEDASTICIDAD
	PUNTOS ANÓMALOS DETECTADOS	RELACIÓN ENTRE VARIABLES			SIGNIFICATIVIDAD	
		GRÁFICO X VS Y	MULTICOLINEALIDAD			
VENTA PROPIA	DULCOSOL E HISPADUL			VENTA PROPIA (MODELO)	SÍ ES SIGNIFICATIVO	
MÁRGENES	HISPADUL	MUY RELACIONADAS	sí	MÁRGENES		
NÚMERO DE COMERCIALES	DULMADRID	MUY RELACIONADAS	sí	NÚMERO DE COMERCIALES		
INCENTIVOS CONTADO	DULCOSOL, DULGALICIA E HISPADUL	MODERADAMENTE RELACIONADAS	NO	INCENTIVOS CONTADO	NO ES SIGNIFICATIVA	sí
INCENTIVOS CRÉDITO	DULCOSOL E HISPADUL	MUY RELACIONADAS	NO	INCENTIVOS CRÉDITO	SÍ ES SIGNIFICATIVA	NO
FRECUENCIA VISITAS	NINGÚN PUNTO ANÓMALO	NO ESTÁN RELACIONADAS	NO	FRECUENCIA VISITAS	SÍ ES SIGNIFICATIVA	NO
TICKET MEDIO POR VISITA	NINGÚN PUNTO ANÓMALO	MODERADAMENTE RELACIONADAS	sí	TICKET MEDIO POR VISITA		
MARGEN TOTAL O MEDIO DIARIO	NINGÚN PUNTO ANÓMALO	MODERADAMENTE RELACIONADAS	NO	MARGEN TOTAL O MEDIO DIARIO	SÍ ES SIGNIFICATIVA	NO
NÚMERO DE CLIENTES	DULMADRID E HISPADUL	MUY RELACIONADAS	sí	NÚMERO DE CLIENTES		
REGALOS	NINGÚN PUNTO ANÓMALO	MUY RELACIONADAS	NO	REGALOS	SÍ ES SIGNIFICATIVA	NO
POBLACIÓN	DULMADRID	MODERADAMENTE RELACIONADAS	sí	POBLACIÓN		
NÚMERO DE MUJERES	DULMADRID	MODERADAMENTE RELACIONADAS	sí	NÚMERO DE MUJERES		
IPC ALIMENTOS Y BEBIDAS NO ALCOHÓLICAS	NINGÚN PUNTO ANÓMALO	NO ESTÁN RELACIONADAS	NO	IPC ALIMENTOS Y BEBIDAS NO ALCOHÓLICAS	NO ES SIGNIFICATIVA	NO
TASA DE RIESGO DE POBREZA	NINGÚN PUNTO ANÓMALO	MODERADAMENTE RELACIONADAS	NO	TASA DE RIESGO DE POBREZA	NO ES SIGNIFICATIVA	NO
				CONSTANTE	NO ES SIGNIFICATIVA	

Tabla 56 - Tabla resumen proceso.

Como se ha dicho previamente, la **variable de Dulgandía** también presenta un residuo estudentizado demasiado bajo, por lo **que se suprimirá** también del modelo.

RESIDUOS ATÍPICOS		
FILA	DELEGACIONES	RESIDUO ESTUDENTIZADO
5	Dulcosol	2,96
7	Dulgandía	-7,37

Tabla 57 - Residuo estudentizado.

5. SOLUCIÓN DEL MODELO MATEMÁTICO

Este sería el resultado de excluir la delegación de Dulgandía, realizar una selección paso a paso hacia detrás y eliminar la constante dado que no salía significativa en este transcurso:

$$\begin{aligned} \text{Venta Propia} - \text{AVG}(\text{Venta Propia}) = & B_0 + B_1 \cdot \text{Incentivos crédito} - \\ & - \text{AVG}(\text{Incentivos crédito}) - B_2 \cdot \text{Frecuencia visitas} - \text{AVG}(\text{Frecuencia visitas}) + B_3 \cdot \\ & \text{Margen total medio diario} - \text{AVG}(\text{Margen total medio diario}) + B_4 \cdot \text{Regalos} - \\ & \text{AVG}(\text{Regalos}) - B_5 \cdot \text{Tasa de riesgo de pobreza} - \text{AVG}(\text{Tasa de riesgo de pobreza}) \end{aligned}$$

PARÁMETRO	VALOR-P
INCENTIVOS CRÉDITO-AVG (INCENTIVOS CRÉDITO)	0
FRECUENCIA VISITAS-AVG(FRECUENCIA VISITAS)	0,0015
MARGEN TOTAL MEDIO DIARIO-AVG(MARGEN TOTAL MEDIO DIARIO)	0
REGALOS-AVG(REGALOS)	0
TASA DE RIESGO DE POBREZA-AVG(TASA DE RIESGO DE POBREZA)	0,0187

Tabla 58 - Valor-P parámetros.

ANÁLISIS DE VARIANZA	
FUENTE	VALOR-P
MODELO	0
RESIDUO	
TOTAL	

Tabla 59 - Valor-P modelo.

El R^2 del modelo, muestra un valor de **99,017%**. La ecuación del modelo ajustado sería la siguiente:

$$\begin{aligned} \text{Venta Propia} - \text{AVG}(\text{Venta Propia}) = & 2,0598E6 \cdot \text{Incentivos crédito} - \\ & - \text{AVG}(\text{Incentivos crédito}) - 271308 \cdot \text{Frecuencia visitas} - \text{AVG}(\text{Frecuencia visitas}) + \\ & + 231726 \cdot \text{Margen total medio diario} - \text{AVG}(\text{Margen total medio diario}) + 44,5875 \cdot \\ & \text{Regalos} - \text{AVG}(\text{Regalos}) - 39111,2 \cdot \text{Tasa de riesgo de pobreza} - \\ & - \text{AVG}(\text{Tasa de riesgo de pobreza}) \end{aligned}$$

6. PREDICCIONES

DELEGACIÓN	MEDIA	VALOR	AHORRO/INVERSIÓN	VARIABLES				
				INCENTIVOS CRÉDITO (+)	FRECUENCIA VISITAS (-)	MARGEN DIARIO (+)	REGALOS (+)	TASA DE RIESGO DE POBREZA (-)
				SE PUEDEN MODIFICAR				NO MODIFICABLE
ARADUL	Encima de la línea	(-)	Reducción de costes		↓			↓
ASTURSOL	Debajo de la línea	(+)	Inversión	↑		↑	↑	
DOCES LUCENSES	Encima de la línea	(-)	Reducción de costes		↓			↓
DULCADIS	Debajo de la línea	(+)	Invertir	↑		↑	↑	
DULCOSOL	Encima de la línea	(-)	Reducción de costes		↓			↓
DULGALICIA	Debajo de la línea	(+)	Invertir	↑		↑	↑	
DULMADRID	Debajo de la línea	(+)	Invertir	↑		↑	↑	
EUSKASOL	Encima de la línea	(-)	Reducción de costes		↓			↓
GADESDUL	Debajo de la línea	(+)	Invertir	↑		↑	↑	
HERMANOS MUÑOZ	Debajo de la línea	(+)	Invertir	↑		↑	↑	
HISPADUL	Encima de la línea	(-)	Reducción de costes		↓			↓
LEVANTINA	Encima de la línea	(-)	Reducción de costes		↓			↓
TINERFEÑA	Encima de la línea	(-)	Reducción de costes		↓			↓
VEGODISA	Debajo de la línea	(+)	Invertir	↑		↑	↑	
VIGUESA	Encima de la línea	(-)	Reducción de costes		↓			↓

Tabla 60 – Tabla explicativa proceso modificación de variables.

ANÁLISIS

Para analizar las predicciones se ha decidido elaborar una tabla para que el proceso explicativo resulte más visual y sencillo. La **reducción de costes** se mostrará con una **flecha en rojo** en las variables correspondientes. Mientras que las **inversiones** se presentarán mediante una **flecha en verde** en las variables oportunas.

En el caso de la **tasa de riesgo de pobreza**, al ser una **variable externa inalterable** a efectos de la empresa, tan sólo se podrá observar su comportamiento. Se tendrá que tener en cuenta que un aminoramiento de está, incrementará las ventas de la empresa.

Para concluir de tal forma el modelo, se ha planteado de la siguiente manera:

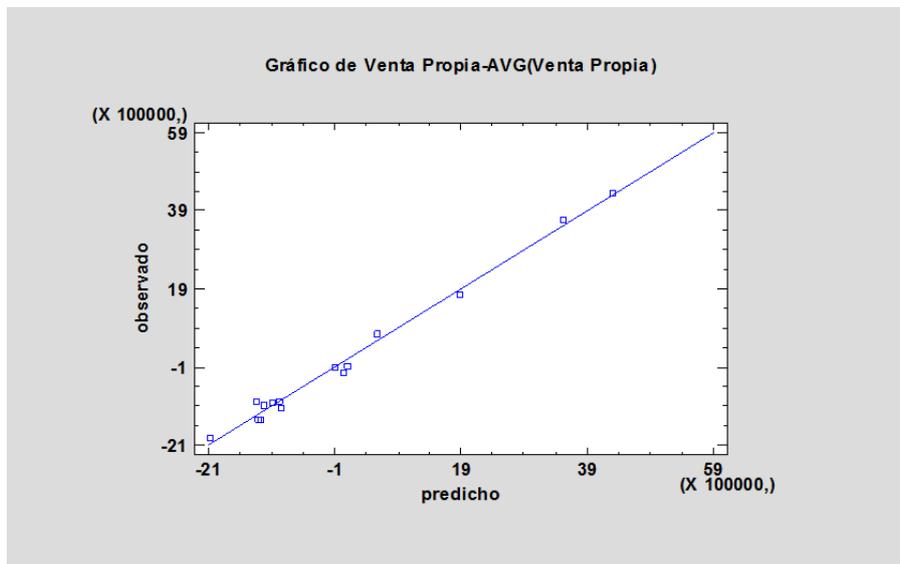


Ilustración 55 - Observado vs predicho.

Primeramente, se ha analizado el observado vs predicho del gráfico de venta propia, y se ha estudiado cada delegación en el gráfico para poder conocer la posición que presentaba en el modelo frente a la media.

Posteriormente, se ha examinado con detenimiento el valor de beta de cada variable y se han clasificado en variables con **betas positivos y negativos**.

- En el caso de los **betas negativos** (los cuales **provocan una disminución en la variable de ventas**) interesará que sean bajos. Los que estén por encima de la media indicarán que las delegaciones están utilizando **más recursos económicos de los que actualmente necesitan**. A estos últimos se les provocará una **reducción del 15%**. En las conclusiones finales del proyecto, tan sólo se focalizará en las delegaciones que estén por **encima de la media**. En este caso se ha decidido mermar el regresor **frecuencia de visitas** para aminorar los costos de la empresa. Como ya se ha explicado previamente la **tasa de riesgo de pobreza** no se podrá disminuir dado que es un factor externo.
- Los **betas positivos** (los cuales afectan al **incremento de ventas**) interesará que sean altos. Los situados por debajo de la media apuntarán a que las delegaciones están **desaprovechando recursos que tienen a su alcance**. Las delegaciones que se muestren por debajo de la media,

se les provocará un **aumento del 15%** con el fin de producir un acrecentamiento de la variable dependiente.

BETA	MEDIA	SOLUCIÓN
-	Encima	Reducir costes
+	Debajo	Invertir

Ilustración 56 - Explicación aumento/reducción variables.

En los casos de **beta positivo por encima de la media** y **beta negativo por debajo de la media**, no se modificarán los valores en este proyecto, puesto que esto significa que la delegación está **obteniendo más ventas que la media** y el otro caso, que está **ahorrando más costes que la media**. Este estudio tratará de aumentar las ventas de las delegaciones que estén desaprovechando sus recursos y de ahorrar costes de las delegaciones que estén sobrepasando los límites y abusando de la economía de la empresa. Por lo tanto, el caso anteriormente comentado, se dejará para futuras líneas de investigación. Se podrá estudiar la obtención de un mayor rendimiento tanto a nivel de reducción de costes como de inversión, de las delegaciones que actualmente son eficaces y eficientes con su economía.

Tabla principal con todos los valores sin modificar:

	INCENTIVO CRÉDITO (%)	FRECUENCIA VISITAS	MARGEN TOTAL MEDIO DIARIO	REGALOS	TASA DE RIESGO DE POBREZA (%) DATOS DEL 2017
ARADUL	0,64	11,55	8,16	6573,72	11,3
ASTURSOL	1,06	9,24	5,99	1983,6	10,7
DOCES LUCENSES	0,85	9,77	10,26	3915,93	15,9
DULCADIS	1,06	8,8	9,56	23017,39	23,9
DULCOSOL	1,7	9,29	15,6	42753,23	25
DULGALICIA	1,06	11,78	9,41	11446,84	15,9
DULMADRID	1,28	12,95	10,65	63893,54	14,4
EUSKASOL	0,85	11,63	13,34	29696,28	11
GADES DUL	0,85	8,39	9,8	26770,04	26,4
HERMANOS MUÑOZ	0,64	11,5	11,9	5519,83	12,8
HISPADUL	2,34	9,01	13,63	40614,67	26,4
LEVANTINA	0,85	11,35	10,48	9220,01	21,8
TINERFEÑA	0,85	9,93	9,89	12796,35	25,9
VEGODISA	0,85	10,55	15,88	10815,15	25,9
VIGUESA	0,85	8,94	8,56	2104,14	15,9

Tabla 61 - Variables y delegaciones sin datos alterados.

Tras **modificar los datos** de cada variable según el análisis del modelo:

	INCENTIVO CRÉDITO (%)	FRECUENCIA VISITAS	MARGEN TOTAL MEDIO DIARIO	REGALOS	TASA DE RIESGO DE POBREZA (%) DATOS DEL 2017
ARADUL	0,64	9,82	8,16	6573,72	11,30
ASTURSOL	1,22	9,24	6,89	2281,14	10,70
DOCES LUCENSES	0,85	8,30	10,26	3915,93	15,90
DULCADIS	1,22	8,80	10,99	26470,00	23,90
DULCOSOL	1,70	7,90	15,60	42753,23	25,00
DULGALICIA	1,22	11,78	10,82	13163,87	15,90
DULMADRID	1,47	12,95	12,25	73477,57	14,40
EUSKASOL	0,85	9,89	13,34	29696,28	11,00
GAESDUL	0,98	8,39	11,27	30785,55	26,40
HERMANOS MUÑOZ	0,74	11,50	13,69	6347,80	12,80
HISPADUL	2,34	7,66	13,63	40614,67	26,40
LEVANTINA	0,85	9,65	10,48	9220,01	21,80
TINERFEÑA	0,85	8,44	9,89	12796,35	25,90
VEGODISA	0,98	10,55	18,26	12437,42	25,90
VIGUESA	0,85	7,60	8,56	2104,14	15,90

Tabla 62 - Variables modificadas según el resultado del proyecto econométrico para su posterior predicción.

El **beneficio final** se quedaría de la siguiente forma para cada una de las delegaciones que se muestran a continuación:

	INCENTIVO CRÉDITO (%)	AJUSTADO	MEDIA - AJUSTADO	BENEFICIO FINAL
ASTURSOL	1,22	-1028050,00	1221636,25	410.787,85€
DULCADIS	1,22	353154,00	2602840,25	442.484,43€
DULGALICIA	1,22	-709612,00	1540074,25	465.881,32€
DULMADRID	1,47	2191830,00	4441516,25	443.717,58€
GAESDUL	0,98	102572,00	2352258,25	227.782,54€
HERMANOS MUÑOZ	0,74	-1179420,00	1070266,25	271.976,63€
VEGODISA	0,98	222711,00	2472397,25	493.962,93€
TOTAL				2.756.593,30€

Tabla 63 - Sumatorio beneficio final inversión en incentivos crédito.

	MARGEN TOTAL MEDIO	AJUSTADO	MEDIA - AJUSTADO	BENEFICIO FINAL
ASTURSOL	6,8885	-1142110	1107576,25	296.727,85€
DULCADIS	10,994	343234	2592920,25	432.564,43€
DULGALICIA	10,8215	-704141	1545545,25	471.352,32€
DULMADRID	12,2475	2160040	4409726,25	411.927,58€
GAESDUL	11,27	153508	2403194,25	278.718,54€
HERMANOS MUÑOZ	13,685	-1006360	1243326,25	445.036,63€
VEGODISA	18,262	455367	2705053,25	726.618,93€
TOTAL				3.062.946,3€

Tabla 64 - Sumatorio beneficio final inversión en margen total medio.

	REGALOS	AJUSTADO	MEDIA - AJUSTADO	BENEFICIO FINAL
ASTURSOL	2281,14	-1292310	957376,253	146.527,85€
DULCADIS	26469,9985	213538	2463224,25	302.868,43€
DULGALICIA	13163,866	-905765	1343921,25	269.728,32€
DULMADRID	73477,571	2257150	4506836,25	509.037,58€
GADESDUL	30785,546	40653,6	2290339,85	165.864,14€
HERMANOS MUÑOZ	6347,8045	-1301720	947966,253	149.676,63€
VEGODISA	12437,4225	70484	2320170,25	341.735,93€
TOTAL				1.885.438,9€

Tabla 65 - Sumatorio beneficio final inversión en regalos.

El **beneficio final total** para cada una de las variables, si reciben estas un incremento del 15% para las delegaciones mencionadas, se quedaría de tal manera:

VARIABLES	BENEFICIOS
INCENTIVOS CRÉDITO	2.756.593,3€
MARGEN TOTAL MEDIO	3.062.946,3€
REGALOS	1.885.438,9€
Total	7.704.978,5€

Tabla 66 - Sumatorio beneficios finales.

En cuanto a la **reducción de costes** se quedaría de la siguiente manera:

	FRECUENCIA VISITAS	AJUSTADO	MEDIA - AJUSTADO	REDUCCIÓN DE COSTES FINAL
ARADUL	9,8175	-1758600	491086,2527	157.734,91
DOCES LUCENSES	8,3045	-766558	1483128,253	199.127,79
DULCOSOL	7,8965	3697670	5947356,253	29.832,66
EUSKASOL	9,8855	836661	3086347,253	80.600,78
HISPADUL	7,6585	4531200	6780886,253	196.866,40
LEVANTINA	9,6475	-986451	1263235,253	835,94
TINERFEÑA	8,4405	-827257	1422429,253	172.372,49
VIGUESA	7,599	-1065990	1183696,253	19.895,71
TOTAL				-857.266,70 €

Tabla 67 - Reducción costes frecuencia de visitas.

A continuación, se mostrarán las **predicciones puntuales y por intervalo**:

PREDICCIÓN INCENTIVOS CRÉDITO		
	PREDICCIÓN PUNTUAL	PREDICCIÓN POR INTERVALO
ASTURSOL	-1028050	$[-1667530 \leq \text{Incentivos contado} \leq -388563]_{95\%}$
DULCADIS	353154	$[-209319 \leq \text{Incentivos contado} \leq 915627]_{95\%}$
DULGALICIA	-709612	$[-1319570 \leq \text{Incentivos contado} \leq -99650]_{95\%}$
DULMADRID	2191830	$[1503550 \leq \text{Incentivos contado} \leq 2880110]_{95\%}$
GADESDUL	102572	$[-514756 \leq \text{Incentivos contado} \leq 719903]_{95\%}$
HERMANOS MUÑOZ	-1179420	$[-1754680 \leq \text{Incentivos contado} \leq -604156]_{95\%}$
VEGODISA	222711	$[-405312 \leq \text{Incentivos contado} \leq 850730]_{95\%}$

Tabla 68 - Predicciones puntuales y por intervalo.

PREDICCIÓN FRECUENCIA DE VISITAS		
	PREDICCIÓN PUNTUAL	PREDICCIÓN POR INTERVALO
ARADUL	-1758600	$[-2367200 \leq \text{Frecuencia visitas} \leq -1150000]_{95\%}$
DOCES LUCENSES	-766558	$[-1394920 \leq \text{Frecuencia visitas} \leq -138196]_{95\%}$
DULCOSOL	3697670	$[3008480 \leq \text{Frecuencia visitas} \leq 4386870]_{95\%}$
EUSKASOL	836661	$[112877 \leq \text{Frecuencia visitas} \leq 1560450]_{95\%}$
HISPADUL	4531200	$[3801120 \leq \text{Frecuencia visitas} \leq 5261290]_{95\%}$
TINERFEÑA	-827257	$[-1446210 \leq \text{Frecuencia visitas} \leq -208305]_{95\%}$
VIGUESA	-1065990	$[-1721140 \leq \text{Frecuencia visitas} \leq -410843]_{95\%}$

Tabla 69 - Predicción frecuencia de visitas.

PREDICCIÓN MARGEN TOTAL MEDIO DIARIO		
	PREDICCIÓN PUNTUAL	PREDICCIÓN POR INTERVALO
ASTURSOL	-1142110,00	$[-1779370 \leq \text{Margen total medio diario} \leq -504839]_{95\%}$
DULCADIS	343234,00	$[-234873 \leq \text{Margen total medio diario} \leq 921341]_{95\%}$
DULGALICIA	-704139,00	$[-1299360 \leq \text{Margen total medio diario} \leq -108922]_{95\%}$
DULMADRID	2160040,00	$[1463500 \leq \text{Margen total medio diario} \leq 2856590]_{95\%}$
GADESUL	153510,00	$[-493805 \leq \text{Margen total medio diario} \leq 800824]_{95\%}$
HERMANOS MUÑOZ	-1006360,00	$[-1631030 \leq \text{Margen total medio diario} \leq -381682]_{95\%}$
VEGODISA	455365,00	$[-253938 \leq \text{Margen total medio diario} \leq 1164670]_{95\%}$

Tabla 70 - Predicción margen total medio diario.

PREDICCIÓN REGALOS		
	PREDICCIÓN PUNTUAL	PREDICCIÓN POR INTERVALO
ASTURSOL	-1292310,00	$[-1885440 \leq \text{Regalos} \leq -699183]_{95\%}$
DULCADIS	213539,00	$[-332345 \leq \text{Regalos} \leq 759423]_{95\%}$
DULGALICIA	-905765,00	$[-1461480 \leq \text{Regalos} \leq -350052]_{95\%}$
DULMADRID	2257150,00	$[1555840 \leq \text{Regalos} \leq 2958470]_{95\%}$
GADESUL	40654,30	$[-574026 \leq \text{Regalos} \leq 655334]_{95\%}$
HERMANOS MUÑOZ	-1301720,00	$[-1847690 \leq \text{Regalos} \leq -755743]_{95\%}$
VEGODISA	70482,60	$[-525813 \leq \text{Regalos} \leq 666778]_{95\%}$

Tabla 71 - Predicción regalos.

Después de conocer las predicciones, se llevarán a cabo una serie de conclusiones y futuras líneas de investigación de este proyecto.

7. CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

7.1 CONCLUSIONES

- Para explicar el aumento de las ventas del grupo Dulcesol en España, se ha utilizado un modelo matemático econométrico. Se han presentado 13 variables independientes y una variable dependiente. El modelo presentaba un R^2 99,9032%.
- Después de analizar el modelo mediante el **análisis univariante** se ha podido observar que 8 delegaciones presentaban puntos anómalos en las variables analizadas en el gráfico de cajas y bigotes. Los más destacados han sido: Hispadul, con 5 puntos anómalos. Dulmadrid con 4 puntos atípicos. Dulcosol con 3 puntos. Dulgalicia presentaba tan sólo un punto atípico. El punto que más se ha llegado a plantear eliminar, ha sido Hispadul, pero esta empresa posee mucho nivel de venta debido a su gran población, además sus variables incrementan debido a lo mismo. Por lo tanto, se ha decidido no eliminar ninguna delegación del modelo.
- Posteriormente se ha estudiado el **análisis bivariante** mediante el gráfico x vs y y la multicolinealidad para medir la relación entre variables. Se han encontrado variables muy relacionadas y que mostraban la presencia de multicolinealidad, por lo que después de realizar los cálculos oportunos, se han suprimido 6 variables en el modelo. El nuevo R^2 era 94,56%
- A continuación, se ha indagado en el **análisis multivariante**. En este se ha podido observar la significatividad, test de normalidad, heterocedasticidad, autocorrelación, puntos anómalos y puntos influyentes. De lo nombrado anteriormente cabe destacar, que se han presentado 4 variables no significativas incluyendo las ventas y la constante del modelo. En cuanto a la heterocedasticidad del modelo, se ha concluido que la variable incentivos crédito generaba problemas, por lo que se ha suprimido del modelo. Y por último, analizando los residuos atípicos, se ha podido contemplar que Dulgandía presentaba un residuo estudentizado anómalo, por lo que se ha eliminado del modelo.
- **El R^2 del modelo final es 99,017%**
- Se ha propuesto incrementar el valor de las variables con beta positivo, de las delegaciones inferiores a la media en un 15%, con el fin de aumentar las ventas en las delegaciones Astursol, Dulcadis, Dulgalicia, Dulmadrid, Gadesdul, Hermanos Muñoz y Vegodisa. El resultado final han sido unos **beneficios de 7.704.978,5€**.
- Otra medida que se expuesto ha sido disminuir en un 15% la variable frecuencia de visitas, de las delegaciones que estaban por encima de la media. Estas eran: Aradul, Doces lucenses, Dulcosol, Euskasol, Hispadul, Levantina, Tinerfeña y Viguessa. Se predice reducir los costes siguiendo esta propuesta en **857.266,70 €**
- Se ha expresado el interés por **controlar la variable tasa de riesgo de pobreza**, puesto que, a mayor disminución, mayor será el aumento de las ventas de la empresa.

7.2 FUTURAS LÍNEAS DE INVESTIGACIÓN

- Será interesante que se **estudiase financieramente el coste** que supondría **implementar** este **proyecto** en el grupo, puesto que se prevén altos beneficios y un gran ahorro en costes.
- **Disminución** de los **betas negativos** por debajo de la media y **aumento** de los **betas positivos** por encima de la media con el fin de maximizar el rendimiento de la empresa aportándoles mayores ventas
- Desempeñar un estudio econométrico del mismo, pero recogiendo los **datos anuales** e incluir las variables que no se han podido escoger.
- Realizar un **estudio** econométrico de la **tasa de riesgo de pobreza** y qué variables influyen a que esta ocurra.
- **Estudiar** la delegación de **Dulgandía**, conocer porqué está fue eliminada y conocer qué valores debería de mostrar para que ésta tuviese un comportamiento normal.
- **Efectuar** este proyecto mediante la **estructura semilogarítmica, logarítmica, inversa y logarítmica inversa**.

8. BIBLIOGRAFÍA

- [1] https://datavizcatalogue.com/ES/metodos/diagrama_cajas_y_bigotes.html [10 de Noviembre]
- [2] https://www.academia.edu/16716695/Manual_de_Econometria_9 [08 de Enero]
- [3] http://www2.ulpgc.es/hege/almacen/download/6/6090/Manual_de_Econometria_9.pdf [20 de Enero]
- [4] Jordán Núñez, Jorge (2019) Apuntes econometría (UPV).
- [5] <http://ciberconta.unizar.es/Leccion/autocorrelacion/analisis%20de%20autocorrelacion.PDF> [05 de Febrero]
- [6] Jordán Núñez, Jorge (2019) Apuntes econometría (UPV).