

# Aplicación práctica: Base de datos del HATCO

**Datos:** 100 observaciones de 14 variables obtenidos desde encuestas a clientes de un distribuidor industrial.

## Variables de interés en la aplicación:

- Percepciones de HATCO:

- $X_1$  Velocidad de entrega
- $X_2$  Nivel de precios
- $X_3$  Flexibilidad de precios
- $X_4$  Imagen del fabricante
- $X_5$  Servicio conjunto
- $X_6$  Imagen de fuerza de ventas
- $X_7$  Calidad de producto

- Resultados de compra:

- $X_9$  Nivel de fidelidad

- Características del comprador:

- $X_8$  Tamaño de la empresa (Variable codificada 0-1)

## Primer paso: Objetivos del análisis

Predecir los niveles de fidelidad a los productos por parte de los clientes basándonos en las percepciones que estos tienen de la actividad de HATCO, así como identificar los factores que llevan al aumento de la utilización del producto para su aplicación en campañas de marketing diferenciadas.

## Segundo paso: Diseño de la investigación

### Análisis de Regresión múltiple (lineal)

Variable dependiente: Fidelidad al producto ( $X_9$ )

Variables independientes: 7 percepciones ( $X_1, X_2, X_3, X_4, X_5, X_6, X_7$ )

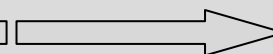
**Tamaño muestral: 100** (= n, N)

## Tercer paso: Supuestos del análisis de regresión múltiple

**1ª etapa:** Contrastación de las variables dependiente e independientes

- A. Relación de linealidad de cada v. independiente con la dependiente
- B. Constancia en la varianza (Homocedasticidad) Después del ajuste.
- C. Normalidad (tests asociados al análisis)

(Hacemos un primer sondeo antes de ajustar el modelo) □□



**2ª etapa:** Contrastación de la relación conjunta después de la estimación del modelo

(Se realizará después de haber estimado el modelo)

## A. Relación de linealidad de cada v. independiente con la dependiente

Hatco - Editor de datos SPSS

	x1	x2	x3
1	4,1	,6	6,9
2	1,8	3,0	6,3
3	3,4	5,2	5,7
4	2,7	1,0	7,1
5	6,0	,9	9,6
6	1,9	3,3	7,9
7	4,6	2,4	9,5
8	1,3	4,2	6,2
9	5,5	1,6	9,4

Gráficos > Interactivos > Diagramas de dispersión...

Diagrama de dispersión simple

Eje Y: y

Eje X: x1

Establecer marcas por:

Etiquetar los casos mediante:

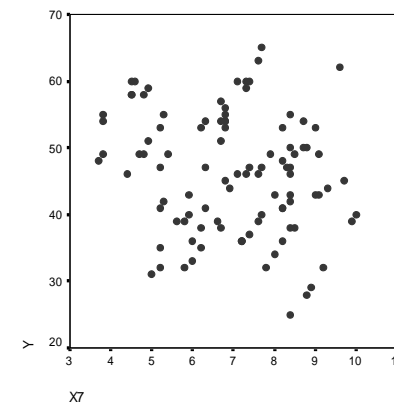
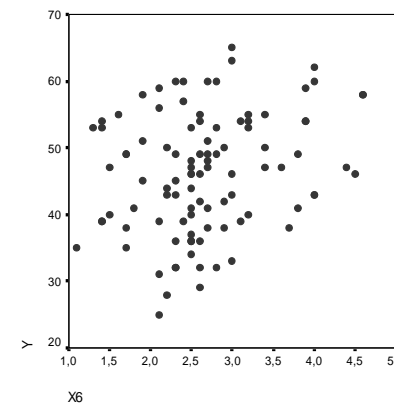
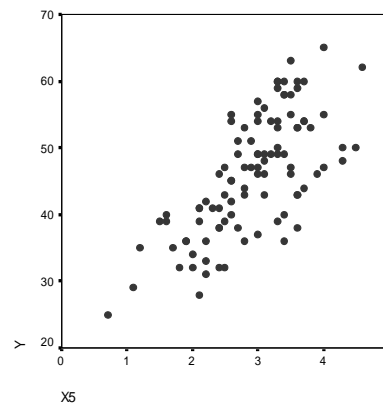
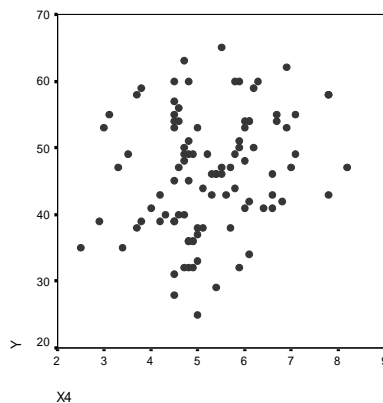
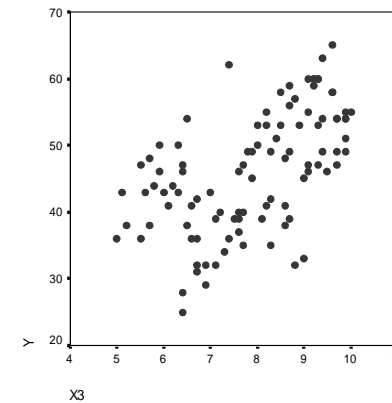
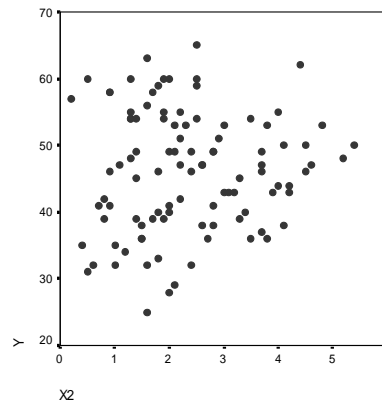
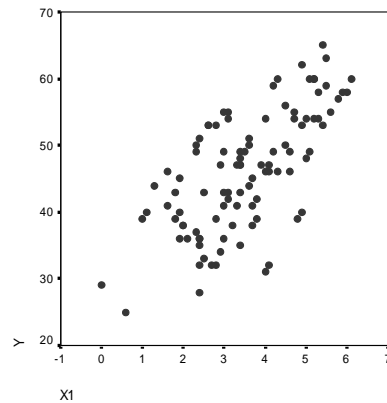
Plantilla

☐ Usar las especificaciones gráficas de:

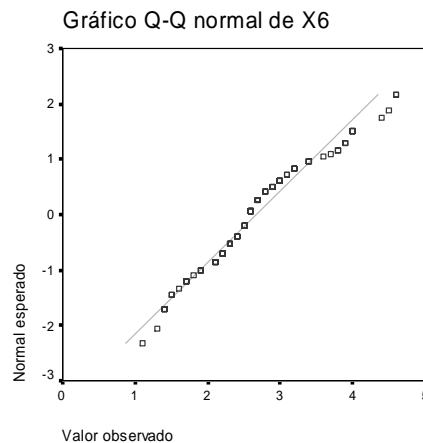
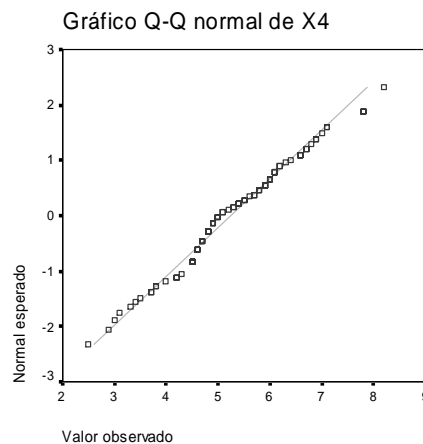
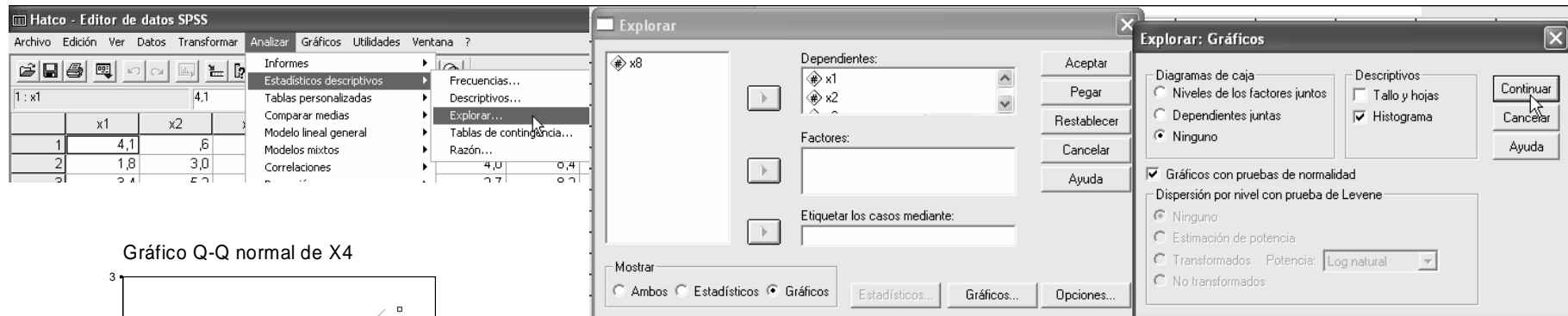
Archivo...

Títulos... Opciones...

Una a una poner cada v. independiente en el eje X



## C. Normalidad de todas las variables



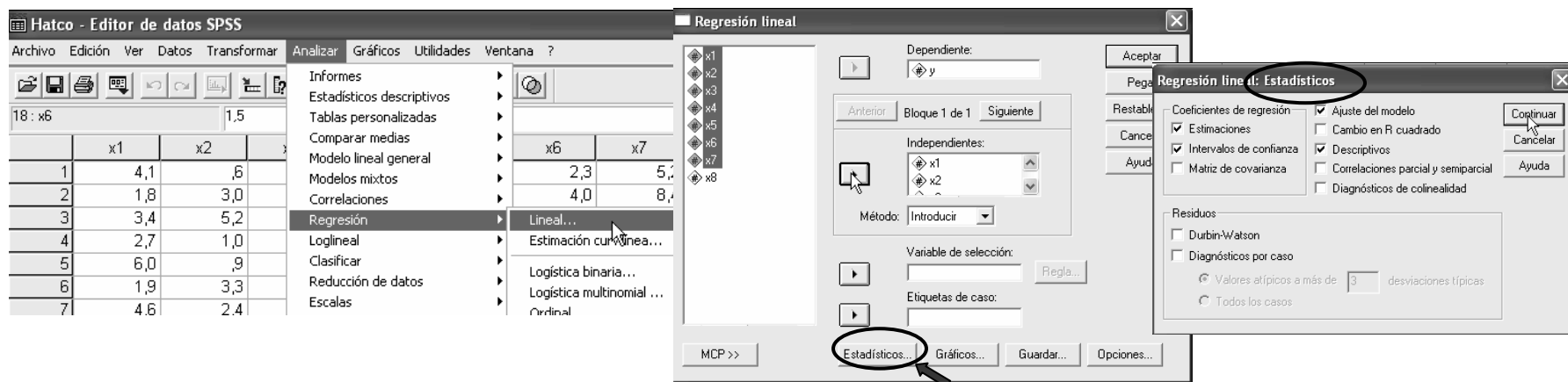
### Pruebas de normalidad

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
X1	,063	100	,200*	,985	100	,341
X2	,095	100	,028	,969	100	,017
X3	,095	100	,027	,950	100	,001
X4	,107	100	,007	,982	100	,183
X5	,085	100	,069	,986	100	,366
X6	,122	100	,001	,963	100	,007
X7	,091	100	,041	,971	100	,028
Y	,079	100	,131	,985	100	,320

\*. Este es un límite inferior de la significación verdadera.

a. Corrección de la significación de Lilliefors

## Cuarto paso: Estimación del modelo y valoración global del ajuste



Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,880 <sup>a</sup>	,775	,758	4,4237

a. Variables predictoras: (Constante), X7, X5, X6, X3, X2, X4, X1

Coeficientes<sup>a</sup>

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados		t	Sig.		
		B	Error típ.	Beta					
1	(Constante)	-10,187	4,977			-2,047	,04		
	X1	-5,76E-02	2,013	-,008		-,029	,97		
	X2	-,697	2,090	-,093		-,333	,74		
	X3	3,368	,411	,520		8,191	,000	2,551	4,185
	X4	-4,22E-02	,667	-,005		-,063	,950	-1,367	1,282
	X5	8,369	3,918	,699		2,136	,035	,587	16,151
	X6	1,281	,947	,110		1,352	,180	-,600	3,162
	X7	,567	,355	,100		1,595	,114	-,139	1,273

a. Variable dependiente: Y

### I. El modelo ajustado y la bondad del ajuste

$$Y = -10.187 - 0.057 X_1 - 0.697 X_2 + 3.368 X_3 - 0.042 X_4 + 8.369 X_5 + 1.281 X_6 + 0.567 X_7$$

El modelo ajustado explica aproximadamente el 75.8% de la variabilidad de Y

ANOVA<sup>b</sup>

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	6198,677	7	885,525	45,252	,000 <sup>a</sup>
	Residual	1800,323	92	19,569		
	Total	7999,000	99			

a. Variables predictoras: (Constante), X7, X5, X6, X3, X2, X4, X1

b. Variable dependiente: Y

## II. Tabla ANOVA

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_7 = 0 \quad (R = 0)$$

$$H_1 : \text{Algún } \beta_j \neq 0 \quad (R \neq 0)$$

Una significación de 0.000 nos dice que existe relación lineal entre las variables y tiene sentido el análisis de regresión

## III. Significación individual de las variables $X_j$ y la constante

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

- Al 5% las variables menos relevantes:  $X_1$ ,  $X_2$ ,  $X_4$ ,  $X_6$  y  $X_7$
- Las que más influencia parecen tener son:  $X_3$  y  $X_4$ , además de la constante que parece conveniente mantenerla en el modelo.
- Observar los coeficientes estandarizados (importancia relativa de cada variable  $X_j$  en la explicación de Y)

Coeficientes<sup>a</sup>

	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Intervalo de confianza para B al 95%	
	B	Error típ.	Beta			Límite inferior	Límite superior
(Constante)	-10,187	4,977		-2,047	,044	-20,071	-,303
X1	-5,76E-02	2,013	-,008	-,029	,977	-4,055	3,940
X2	-,697	2,090	-,093	-,333	,740	-4,848	3,454
X3	3,368	,411	,520	8,191	,000	2,551	4,185
X4	-4,22E-02	,667	-,005	-,063	,950	-1,367	1,282
X5	8,369	3,918	,699	2,136	,035	,587	16,151
X6	1,281	,947	,110	1,352	,180	-,600	3,162
X7	,567	,355	,100	1,595	,114	-,139	1,273

Variable dependiente: Y

Estas consideraciones no son concluyentes hay elegir las variables relevantes utilizando métodos secuenciales

#### IV. Examen de las correlaciones:

**Detectar posible multicolinealidad y qué variables guardan mayor correlación con la variable dependiente (de nuevo no podremos concluir nada).**

Correlaciones

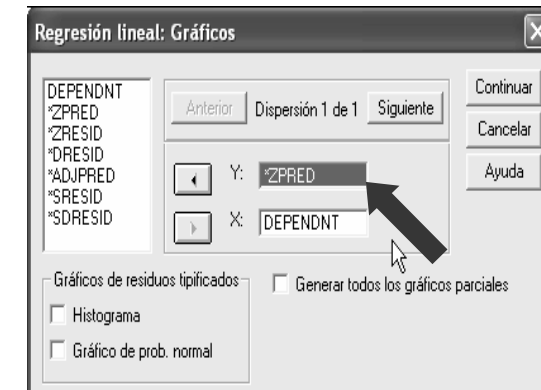
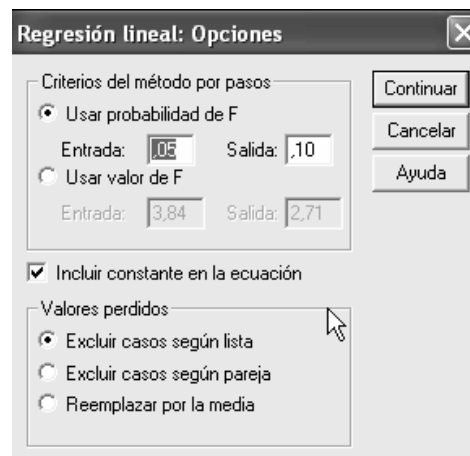
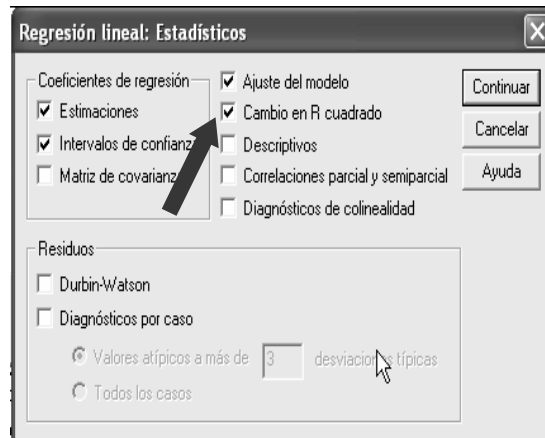
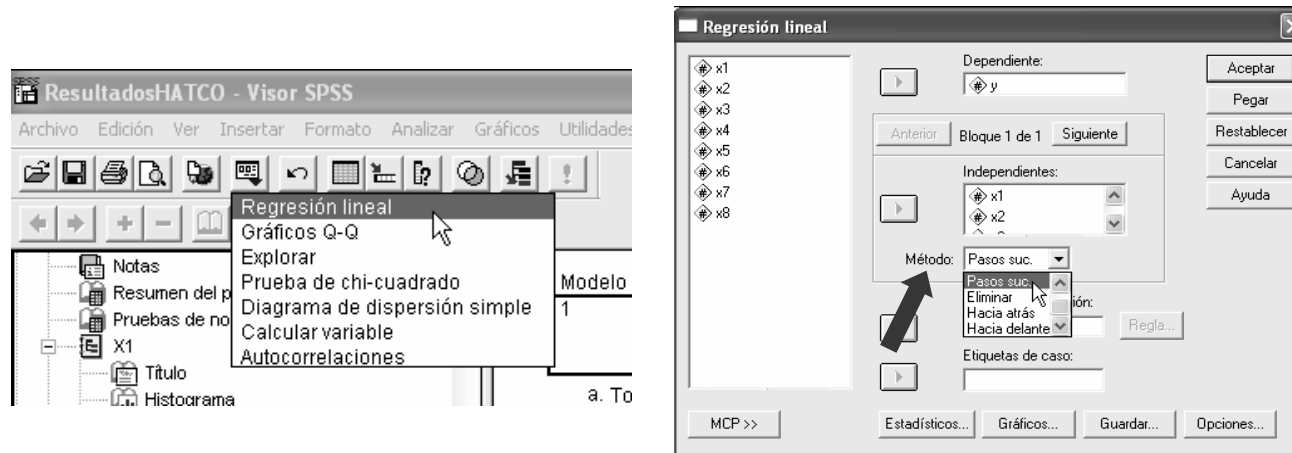
	Y	X1	X2	X3	X4	X5	X6	X7	
Correlación de Pearson	Y	1,000	,676	,082	,559	,224	,701	,256	-,192
	X1	,676	1,000	-,349	,509	,050	,612	,077	-,483
	X2	,082	-,349	1,000	-,487	,272	,513	,186	,470
	X3	,559	,509	-,487	1,000	-,116	,067	-,034	-,448
	X4	,224	,050	,272	-,116	1,000	,299	,788	,200
	X5	,701	,612	,513	,067	,299	1,000	,241	-,055
	X6	,256	,077	,186	-,034	,788	,241	1,000	,177
	X7	-,192	-,483	,470	-,448	,200	-,055	,177	1,000
Sig. (unilateral)	Y	,	,000	,209	,000	,012	,000	,005	,028
	X1	,000	,	,000	,000	,309	,000	,223	,000
	X2	,209	,000	,	,000	,003	,000	,032	,000
	X3	,000	,000	,000	,	,125	,255	,367	,000
	X4	,012	,309	,003	,125	,	,001	,000	,023
	X5	,000	,000	,000	,255	,001	,	,008	,293
	X6	,005	,223	,032	,367	,000	,008	,	,039
	X7	,028	,000	,000	,000	,023	,293	,039	,
N	Y	100	100	100	100	100	100	100	100
	X1	100	100	100	100	100	100	100	100
	X2	100	100	100	100	100	100	100	100
	X3	100	100	100	100	100	100	100	100
	X4	100	100	100	100	100	100	100	100
	X5	100	100	100	100	100	100	100	100
	X6	100	100	100	100	100	100	100	100
	X7	100	100	100	100	100	100	100	100

- Las más correlacionadas con Y son  $X_1$ ,  $X_2$  y  $X_3$
- Entre si la mayor correlación (moderada) se observa entre  $X_1$  y  $X_5$ ,  $X_2$  y  $X_5$ ,  $X_4$  y  $X_6$



## V. Selección de variables independientes: Métodos secuenciales

### Utilizaremos el procedimiento de selección por pasos (Stepwise)



Variables introducidas/eliminadas<sup>a</sup>

Modelo	Variables introducidas	Variables eliminadas	Método
1	X5		Por pasos (criterio: Prob. de F para entrar <= ,050, Prob. de F para salir >= ,100).
2	X3		Por pasos (criterio: Prob. de F para entrar <= ,050, Prob. de F para salir >= ,100).
3	X6		Por pasos (criterio: Prob. de F para entrar <= ,050, Prob. de F para salir >= ,100).

a. Variable dependiente: Y

**El modelo ajustado final (después de tres pasos)**

$$Y = -6.520 + 3.376 X_3 + 7.621 X_5 + 1.406 X_6$$

**El modelo ajustado explica aproximadamente el 76.1% de la variabilidad de Y**

Variables excluidas<sup>d</sup>

					Estadísticos de colinealidad	
Modelo	Beta dentro	t	Sig.	Correlación parcial	Tolerancia	
1	X1	,396 <sup>a</sup>	4,812	,000	,439	,626
	X2	-,377 <sup>a</sup>	-5,007	,000	-,453	,737
	X3	,515 <sup>a</sup>	10,210	,000	,720	,996
	X4	,016 <sup>a</sup>	,216	,830	,022	,911
	X6	,093 <sup>a</sup>	1,252	,214	,126	,942
	X7	-,154 <sup>a</sup>	-2,178	,032	-,216	,997
2	X1	,016 <sup>b</sup>	,205	,838	,021	,405
	X2	-,020 <sup>b</sup>	-,267	,790	-,027	,464
	X4	,095 <sup>b</sup>	1,808	,074	,181	,892
	X6	,121 <sup>b</sup>	2,378	,019	,236	,939
	X7	,094 <sup>b</sup>	1,683	,096	,169	,799
3	X1	,030 <sup>c</sup>	,389	,698	,040	,403
	X2	-,029 <sup>c</sup>	-,405	,687	-,041	,462
	X4	-,002 <sup>c</sup>	-,021	,983	-,002	,357
	X7	,071 <sup>c</sup>	1,273	,206	,130	,768

a. Variables predictoras en el modelo: (Constante), X5

b. Variables predictoras en el modelo: (Constante), X5, X3

c. Variables predictoras en el modelo: (Constante), X5, X3, X6

d. Variable dependiente: Y

Coefficientes<sup>a</sup>

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Intervalo de confianza para B al 95%	
		B	Error típ.				Límite inferior	Límite superior
1	(Constante)	21,653	2,596		8,341	,000	16,502	26,804
	X5	8,384	,862	,701	9,722	,000	6,673	10,095
2	(Constante)	-3,489	3,057		-1,141	,257	-9,556	2,578
	X5	7,974	,603	,666	13,221	,000	6,777	9,171
	X3	3,336	,327	,515	10,210	,000	2,688	3,985
3	(Constante)	-6,520	3,247		-2,008	,047	-12,965	-,075
	X5	7,621	,607	,637	12,547	,000	6,416	8,827
	X3	3,376	,320	,521	10,562	,000	2,742	4,010
	X6	1,406	,591	,121	2,378	,019	-,232	2,579

Resumen del modelo<sup>d</sup>

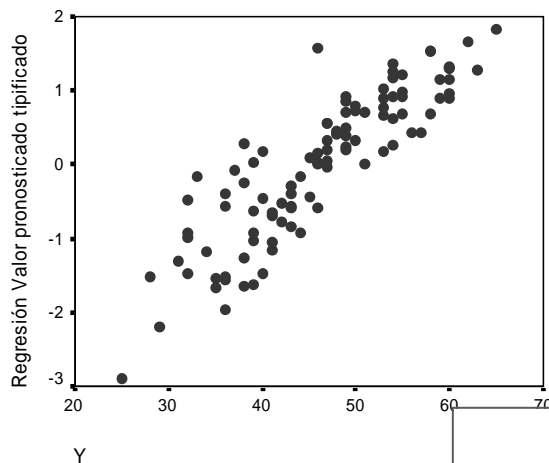
Modelo	R	R cuadrado	R cuadrado corregido	Error típ. de la estimación	Estadísticos de cambio				
					Cambio en R cuadrado	Cambio en F	gl1	gl2	Sig. del cambio en F
1	,701 <sup>a</sup>	,491	,486	6,4458	,491	94,525	1	98	,000
2	,869 <sup>b</sup>	,755	,750	4,4980	,264	104,252	1	97	,000
3	,877 <sup>c</sup>	,768	,761	4,3938	,014	5,656	1	96	,019

La interpretación del modelo ajustado (efectos) la realizaremos en el paso quinto

Para visualizar gráficamente la bondad del ajuste:  
Dispersión de valores exactos frente a valores predichos

Gráfico de dispersión

Variable dependiente: Y



Opciones del diagrama de dispersión

Opciones de visualización

☐ Mostrar los subgrupos

Etiquetas de caso: Desactivado

Origen de las etiquetas:

☐ Variable de identificación

☒ Número de caso

Girasoles

☐ Mostrar girasoles

Opciones de los girasoles...

Ajustar línea

☒ Total ☐ Subgrupos

Opciones de ajuste...

Aceptar  
Cancelar  
Ayuda

Opciones de diagramas de dispersión: Ajustar línea

Método de ajuste

☒ Regresión lineal ☐ Minsce

☐ Regresión cuadrática

☐ Regresión cúbica

% de puntos a ajustar: 50

Nº de iteraciones: 3

Continuar  
Cancelar  
Ayuda

Líneas de predicción de la regresión

☒ Media ☒ Individual

Intervalo de confianza: 95 %

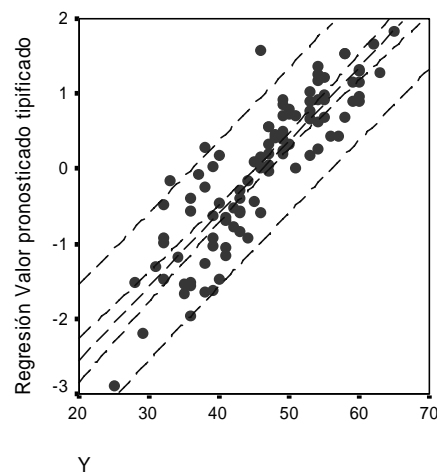
Opciones de la regresión

☒ Incluir constante en la ecuación

☒ Mostrar R-cuadrado en la leyenda

Gráfico de dispersión

Variable dependiente: Y



$R^2 = 0.7683$

Editando el gráfico obtenemos bandas de confianza y la línea de referencia

**Volvemos a comprobar la significación individual y conjunta de las variables:** SPSS ofrece los resultados en cada paso realizado

**ANOVA<sup>d</sup>**

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	3927,309	1	3927,309	94,525	,000 <sup>a</sup>
	Residual	4071,691	98	41,548		
	Total	7999,000	99			
2	Regresión	6036,513	2	3018,256	149,184	,000 <sup>b</sup>
	Residual	1962,487	97	20,232		
	Total	7999,000	99			
3	Regresión	6145,700	3	2048,567	106,115	,000 <sup>c</sup>
	Residual	1853,300	96	19,305		
	Total	7999,000	99			

Como ya sabíamos las variables guarda relación lineal con la v. dependiente en los tres modelos

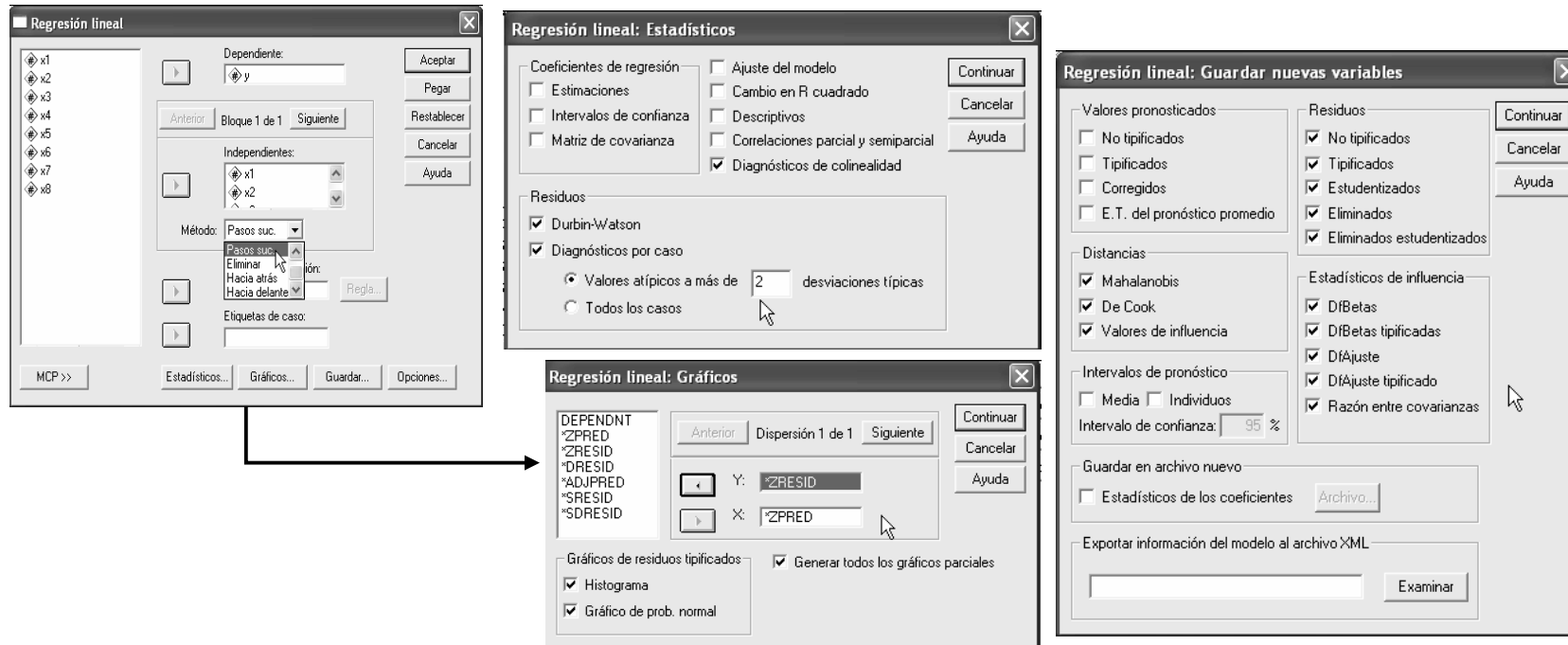
- a. Variables predictoras: (Constante), X5
- b. Variables predictoras: (Constante), X5, X3
- c. Variables predictoras: (Constante), X5, X3, X6
- d. Variable dependiente: Y

**Coefficientes<sup>a</sup>**

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Intervalo de confianza para B al 95%	
		B	Error típ.	Beta			Límite inferior	Límite superior
1	(Constante)	21,653	2,596		8,341	,000	16,502	26,804
	X5	8,384	,862	,701	9,722	,000	6,673	10,095
2	(Constante)	-3,489	3,057		-1,141	,257	-9,556	2,578
	X5	7,974	,603	,666	13,221	,000	6,777	9,171
	X3	3,336	,327	,515	10,210	,000	2,688	3,985
3	(Constante)	-6,520	3,247		-2,008	,047	-12,965	-,075
	X5	7,621	,607	,637	12,547	,000	6,416	8,827
	X3	3,376	,320	,521	10,562	,000	2,742	4,010
	X6	1,406	,591	,121	2,378	,019	,232	2,579

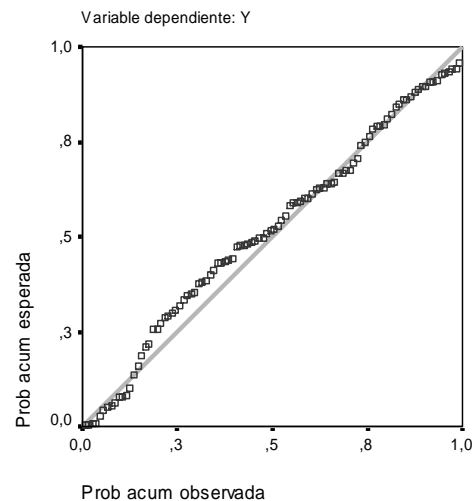
Todas las variables que se van introduciendo son significativas. Obsérvese que la constante en el modelo 2 resulta no significativa

## VI. ANÁLISIS DE LOS RESIDUOS: Evaluación de los supuestos del análisis

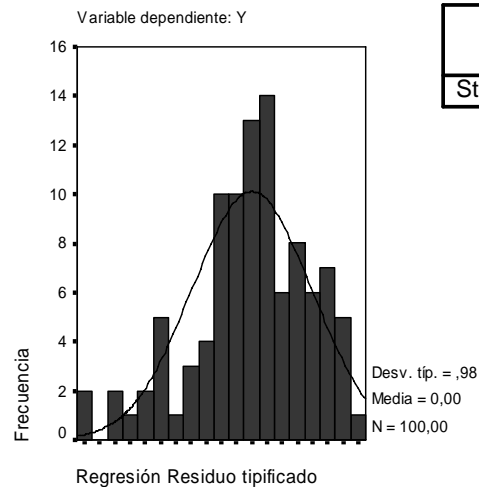


- Normalidad de los residuos: Histograma, gráfico de prob. normal**
  - No autocorrelación: Test de Durbin-Watson**
  - Homocedasticidad: Gráficos de dispersión (ZRESID frente a ZPRED)**
  - Falta de linealidad: Gráficos de regresión parcial**
  - No multicolinealidad: Diagnóstico de colinealidad**
- **Identificación de casos atípicos y datos influyentes: Diagnóstico por caso (atípicos), guardar residuos, estadísticos de influencia y distancias (Mahalanobis, Cook, valores de influencia)**

Gráfico P-P normal de los residuos tipificados



Histograma



Pruebas de normalidad

	Kolmogorov-Smirnov <sup>a</sup>		
	Estadístico	gl	Sig.
Standardized Residual	,079	100	,127

a. Corrección de la significación de Lilliefors

**a) Se puede admitir la normalidad de los residuos del modelo**

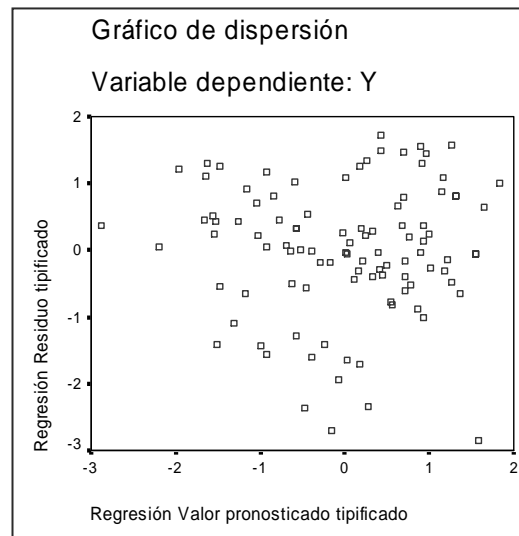
**b) Se puede admitir que no existe autocorrelación serial ( $d > 1.4$ )**

Resumen del modelo<sup>b</sup>

Modelo	Durbin-Watson
3	1,910 <sup>a</sup>

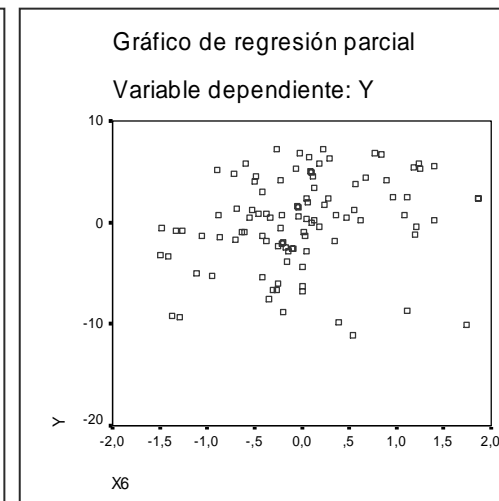
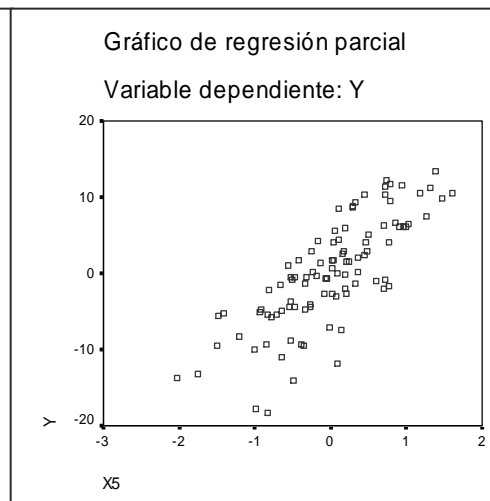
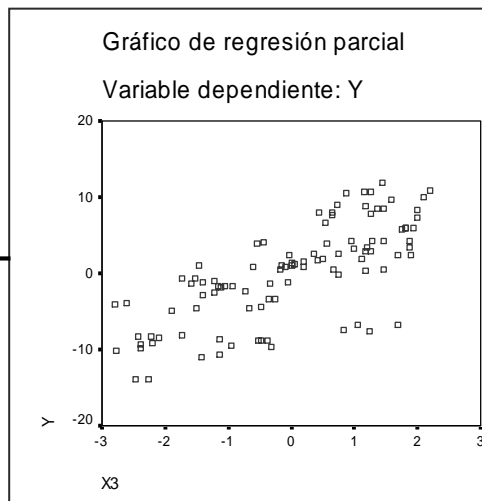
a. Variables predictoras: (Constante), X5, X3, X6

b. Variable dependiente: Y



**c) No muestra una pauta clara de aumento o disminución de los residuos. Admitimos la hipótesis de homocedasticidad**

**d) Las relaciones entre Y y  $X_3$  e Y y  $X_5$  son claramente lineales y muy significativas (más pendiente más efecto de la variable). La relación con  $X_6$  es menos clara y significativa, no obstante no muestra ninguna forma curvilínea. Admitimos la hipótesis de linealidad**



Variables excluidas<sup>d</sup>

Modelo	Beta dentro	Estadísticos de colinealidad	
		FIV	Tolerancia mínima
1	X1	,396 <sup>a</sup>	,626
	X2	-,377 <sup>a</sup>	,737
	X3	,515 <sup>a</sup>	,996
	X4	,016 <sup>a</sup>	,911
	X6	,093 <sup>a</sup>	,942
	X7	-,154 <sup>a</sup>	,997
2	X1	,016 <sup>b</sup>	,405
	X2	-,020 <sup>b</sup>	,464
	X4	,095 <sup>b</sup>	,892
	X6	,121 <sup>b</sup>	,936
	X7	,094 <sup>b</sup>	,797
3	X1	,030 <sup>c</sup>	,403
	X2	-,029 <sup>c</sup>	,462
	X4	-,002 <sup>c</sup>	,357
	X7	,071 <sup>c</sup>	,768

- a. Variables predictoras en el modelo: (Constante), X5  
b. Variables predictoras en el modelo: (Constante), X5, X3  
c. Variables predictoras en el modelo: (Constante), X5, X3, X6  
d. Variable dependiente: Y

Diagnósticos de colinealidad<sup>a</sup>

Modelo	Dimensión	Autovalor	Índice de condición	Proporciones de la varianza			
				(Constante)	X5	X3	X6
1	1	1,969	1,000	,02	,02		
	2	3,132E-02	7,928	,98	,98		
2	1	2,941	1,000	,00	,01	,00	
	2	4,595E-02	8,000	,03	,85	,19	
	3	1,347E-02	14,778	,97	,14	,80	
3	1	3,882	1,000	,00	,00	,00	,00
	2	5,997E-02	8,046	,01	,02	,11	,85
	3	4,541E-02	9,246	,02	,91	,14	,04
	4	1,237E-02	17,719	,97	,07	,75	,10

a. Variable dependiente: Y

**e) No se viola la hipótesis de no multicolinealidad: Los índices de condición en el modelo 3 no superan 30 en ningún caso.**

## Interpretación de los índices de condición

1. Identificar los índices que estén por encima del umbral: 30
2. Para los índices identificados, identificar las variables con proporciones de varianza por encima del 90%: Habrá multicolinealidad si ocurre con dos o más coeficientes.

## Factor de inflación de la varianza (FIV)

Localizar valores superiores a 10



## IDENTIFICACIÓN DE ATÍPICOS Y DATOS INFLUYENTES

➤ Detección de atípicos: Los residuos mayores corresponden a los casos 7, 11, 14 y 100.

Estadísticos sobre los residuos<sup>a</sup>

	Mínimo	Máximo	Media	Desviación típ.	N
Valor pronosticado	23,373	60,592	46,100	7,8789	100
Valor pronosticado tip.	-2,885	1,839	,000	1,000	100
Error típico del valor pronosticado	,4670	1,4292	,8468	,2359	100
Valor pronosticado corregido	23,180	60,388	46,104	7,9148	100
Residuo bruto	-42,552	7,574	-,000	4,3267	100
Residuo tip.	-2,857	1,724	,000	,985	100
Residuo estud.	-2,983	1,737	,000	1,004	100
Residuo eliminado	-13,687	7,694	-,004	4,4970	100
Residuo eliminado estud.	-3,115	1,756	-,004	1,017	100
Dist. de Mahalanobis	,129	9,485	2,970	2,185	100
Distancia de Cook	,000	,201	,010	,022	100
Valor de influencia centrado	,001	,096	,030	,022	100

a. Variable dependiente: Y

Diagnósticos por caso<sup>a</sup>

Número de caso	Residuo tip.	Y	Valor pronosticado	Residuo bruto
7	-2,857	46,0	58,552	-12,552
11	-2,360	32,0	42,367	-10,367
14	-2,342	38,0	48,292	-10,292
100	-2,696	33,0	44,848	-11,848

a. Variable dependiente: Y

Esta tabla es un resumen descriptivo, los valores para los 100 casos aparecen como variables añadidas en el editor de datos

➤ Detección de puntos influyentes

### Límites

Valor de influencia: 0.08

SDFBETA: (-0.2, 0.2)

SDFFIT: (-0.3244, 0.3244)

COVRATIO: (0.88, 1.12)

D. Cook: 0.042

res_1	dre_1	zre_1	sre_1	sdr_1	mah_1	coo_1	lev_1	cov_1	dff_1	sdf_1
-6,29856	-6,43100	-1,43352	-1,44851	-1,45696	1,04876	,01103	,01059	,97463	-,13244	-,21127
3,57540	3,80703	,81374	,83969	,83839	5,03337	,01142	,05084	1,07807	,23163	,21339
-1,29032	-1,39691	-,29367	-,30556	-,30411	6,56411	,00193	,06630	1,12453	-,10659	-,08741
-2,40091	-2,48741	-,54643	-,55619	-,55418	2,45292	,00279	,02478	1,06649	-,08650	-,10519
-,26802	-,29492	-,06100	-,06399	-,06366	8,04023	,00010	,08121	1,14734	,02689	,02847
2,36338	2,41237	,53789	,54344	,54144	1,02013	,00153	,01030	1,05101	,02689	,02847
-12,5520	-13,6870	-2,85677	-2,98313	-3,11547	7,21926	,20116	,07292	,77000		

Puntos influyentes y atípicos:

7, 11, 14 y 100

(posible eliminación)

## Quinto paso: Interpretación del resultado teórico

**El modelo ajustado (Bondad del ajuste del 76.1% )**

$$Y = - 6.520 + 3.376 X_3 + 7.621 X_5 + 1.406 X_6$$

### Interpretación de los efectos de cada variable explicativa

- **Efectos tipificados (Importancia relativa de cada variable en la predicción del nivel de fidelidad)**

El mayor efecto en la predicción del nivel de fidelidad al producto es el asociado a las variables servicio global (63.7%) y flexibilidad de los precios (52.1%).

Aunque con menos efecto, la variable imagen de los vendedores (12.1%) tiene influencia en el nivel de fidelidad.

- **Efectos sin tipificar: (Incremento medio en el nivel de fidelidad cuando la variable aumenta en una unidad y el resto de variables valen 0)**

Servicio global: 7.621

Flexibilidad de precios: 3.376

Imagen de los vendedores: 1.406

Término constante: -6.520      (Incremento medio en el nivel de fidelidad cuando el resto de variables valen 0)

### Predicciones

Nivel de fidelidad (cliente que responde 4.0 en las tres v. independientes) = 43.9

-En SPSS basta añadir un caso con datos en todas las variables excepto en el nivel de fidelidad-

## Sexto paso: Validación de los resultados

- ✓ Comparación entre  $R^2$  y  $R^2$ -ajustado (grandes diferencias puede indicar un sobreajuste del modelo y que no se mantiene una razón adecuada de observaciones con respecto al número de variables)

En nuestro caso: 0.768, 0.761

- ✓ Obtener una segunda muestra y corroborar los resultados.
- ✓ Métodos de exploración intensiva de la muestra (bootstrap, jackknife, PRESS)

## Evaluación de modelos de regresión alternativos

- Incorporación de nuevas variables.
- Inclusión de términos de interacción entre las presentes en el modelo
- Inclusión de términos cuadráticos o cúbicos en alguna variable independiente (si se alejaba de la linealidad)



**Inclusión de una variable no métrica: tamaño de la empresa ( $X_8$ )**

## Inclusión de una variable no métrica: Tamaño de la empresa ( $X_8$ ) 0 / 1 (empresa pequeña / grande)

### OBSERVACIONES

- Al estar codificada con los valores 0 y 1 se puede introducir directamente en el modelo.  
Si está sin codificar o los códigos no se corresponden: Variables ficticias (en SPSS el módulo de Escalamiento óptimo)
- La interpretación del efecto asociado a la variable no métrica es distinto: Interpretamos el coeficiente como el valor de las grandes empresas comparado con el de las pequeñas. Un coeficiente positivo indica que las grandes empresas cuentan con un nivel de fidelidad mayor que el de las pequeñas, y al contrario si es negativo.

Coeficientes<sup>a</sup>

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	21,653	2,596		8,341	,000
	X5	8,384	,862	,701	9,722	,000
2	(Constante)	-3,489	3,057		-1,141	,257
	X5	7,974	,603	,666	13,221	,000
	X3	3,336	,327	,515	10,210	,000
3	(Constante)	-12,956	4,170		-3,107	,002
	X5	8,411	,593	,703	14,190	,000
	X3	4,184	,410	,645	10,193	,000
	X8	3,762	1,182	,206	3,182	,002
4	(Constante)	-16,335	4,254		-3,840	,000
	X5	8,055	,592	,673	13,613	,000
	X3	4,245	,399	,655	10,630	,000
	X8	3,852	1,149	,211	3,353	,001
5	(Constante)	-15,963	4,186		-3,813	,000
	X5	7,008	,772	,586	9,080	,000
	X3	3,991	,411	,616	9,700	,000
	X8	4,946	1,248	,271	3,964	,000
	X6	1,574	,555	,135	2,834	,006
	X1	1,123	,544	,165	2,064	,042

a. Variable dependiente: Y

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,701 <sup>a</sup>	,491	,486	6,4458
2	,869 <sup>b</sup>	,755	,750	4,4980
3	,882 <sup>c</sup>	,778	,771	4,3002
4	,890 <sup>d</sup>	,793	,784	4,1766
5	,895 <sup>e</sup>	,802	,791	4,1067

a. Variables predictoras: (Constante), X5

b. Variables predictoras: (Constante), X5, X3

c. Variables predictoras: (Constante), X5, X3, X8

d. Variables predictoras: (Constante), X5, X3, X8, X6

e. Variables predictoras: (Constante), X5, X3, X8, X6, X1

**Las empresas grandes cuentan con un nivel de fidelidad un 4.946% mayor que las pequeñas**

## Ejercicios propuestos

1. Con los datos contenidos en el fichero “coches.sav” ajustar mediante el procedimiento de regresión no lineal un modelo del tipo:

$$Y = a X_1 + b X_2 + c \sqrt{X_3} + X_4^{**} d + e$$

Donde:

Y: Consumo

$X_1$ : Motor

$X_2$ : Potencia (cv)

$X_3$ : Peso

$X_4$ : Acel

Parámetros: a,b,c,d,e (Valor inicial 1)

2. Con los datos del fichero “coches.sav” ajustar mediante el procedimiento de regresión curvilínea el mejor modelo que permita explicar la variable Consumo en función de la variable Potencia.
3. Ajustar mediante el procedimiento de regresión lineal un modelo que permita predecir el consumo en función de: motor, potencia y la variable codificada origen (utilizar variables ficticias).
4. Con los datos del fichero “encuesta.sav” ajustar mediante el procedimiento de regresión categórica un modelo que permita predecir el grado de felicidad (feliz) en términos del sexo, la raza y el tipo de vida (vida) de los encuestados. (Nota: en este caso los coeficientes resultantes, tipificados, indican en qué medida influye los valores de cada variable independiente en la variable dependiente.)