

Dos Casos Especiales de Regresión

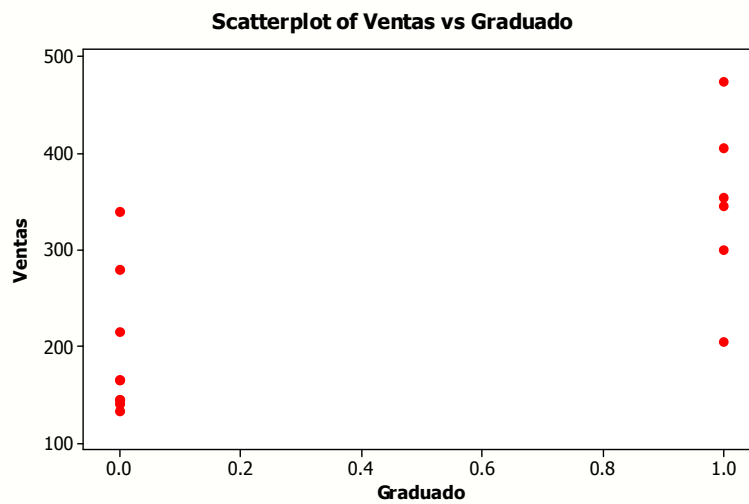
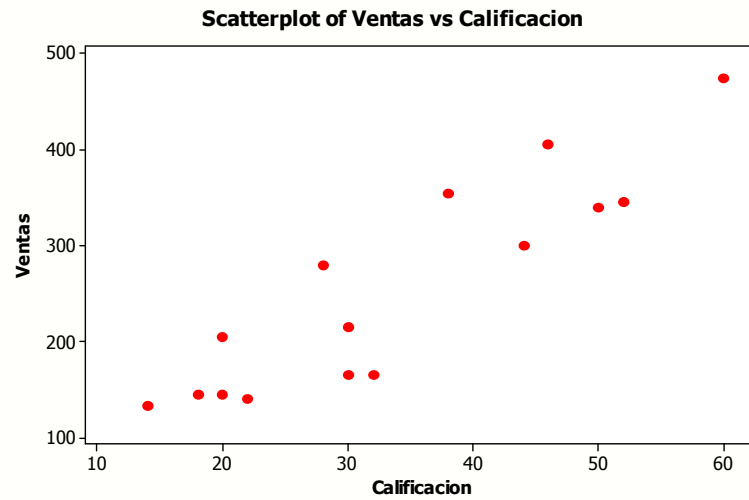
Ellie Burks

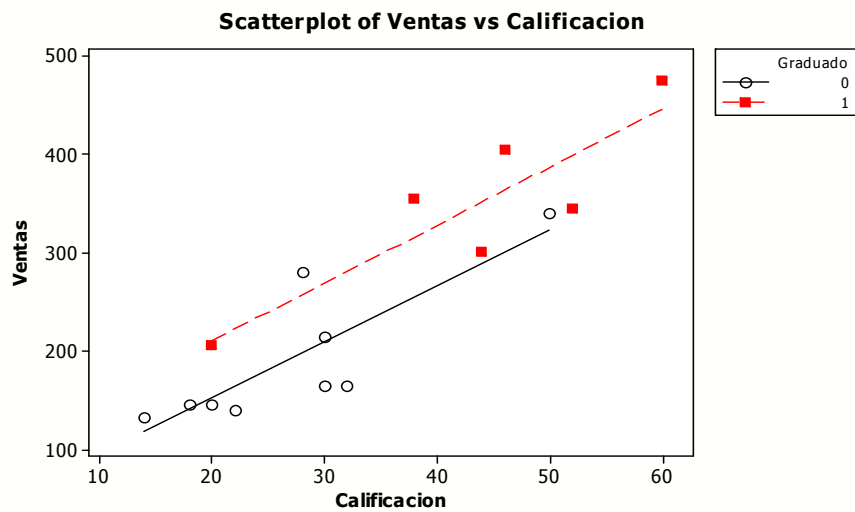
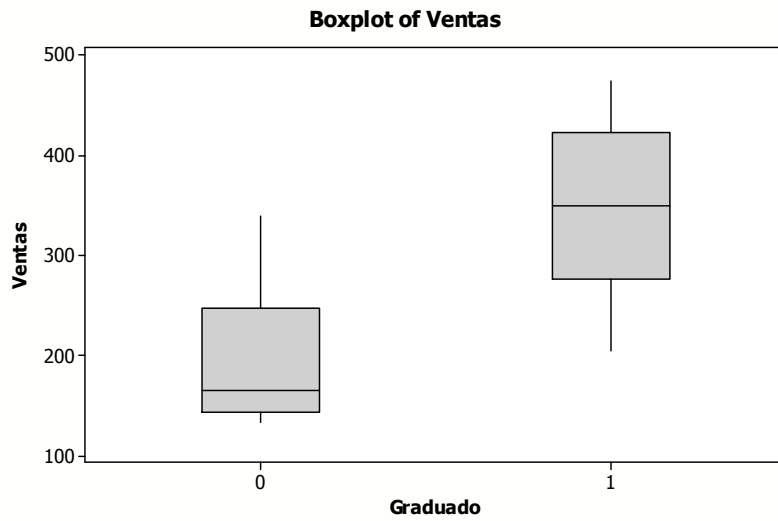
Ellie es la directora de personal para la corporación JVC. En este momento, JVC contrata a futuros vendedores y los capacita mediante un costoso programa. Ellie ha decidido aplicar una prueba de aptitudes para predecir qué solicitantes tendrán éxito en el trabajo. Ellie se pregunta además si los graduados universitarios llegan a ser mejores vendedores que los que no estudiaron una carrera. Ella decide probar con los siguientes 15 solicitantes. Determina si son graduados universitarios y codifica cada persona con 1 si lo es y con 0 si no. Los 15 futuros vendedores se capacitan en el programa y se les da el trabajo; después de un mes, se registran sus ventas totales y resultan los siguientes datos:

Solicitante	Ventas, Y	Calificación en la Prueba, X1	Graduado Universitario, X2
1	345	52	1
2	405	46	1
3	475	60	1
4	205	20	1
5	355	38	1
6	300	44	1
7	133	14	0
8	280	28	0
9	165	30	0
10	145	18	0
11	165	32	0
12	340	50	0
13	140	22	0
14	215	30	0
15	145	20	0

- Represente los datos en un diagrama de dispersión. Utilice 1 para los graduados universitarios y 0 para los que no lo son.
- ¿Es la calificación de la prueba una buena variable de predicción?
- ¿El hecho de que una persona esté titulada es una buena variable de predicción?

- d. Indique la ecuación de regresión que Ellie debe usar para predecir qué vendedores tendrán éxito.
- e. Interprete los coeficientes de regresión estimados.
- f. Estime las ventas del primer mes para un graduado universitario que obtuvo 50 en la prueba.
- g. Estime las ventas del primer mes para un solicitante no graduado que obtuvo 50 en la prueba.
- h. Escriba un memorándum a Ellie con el resumen de las conclusiones respecto a esta situación.





Regression Analysis: Ventas versus Calificacion

The regression equation is
 $Ventas = 15.7 + 7.10 \text{ Calificacion}$

Predictor	Coef	SE Coef	T	P
Constant	15.71	33.88	0.46	0.651
Calificacion	7.0979	0.9346	7.59	0.000

S = 49.2539 R-Sq = 81.6% R-Sq(adj) = 80.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	139937	139937	57.68	0.000
Residual Error	13	31537	2426		
Total	14	171474			

Regression Analysis: Ventas versus Calificacion, Graduado

The regression equation is
 $Ventas = 34.5 + 5.81 \text{ Calificacion} + 61.2 \text{ Graduado}$

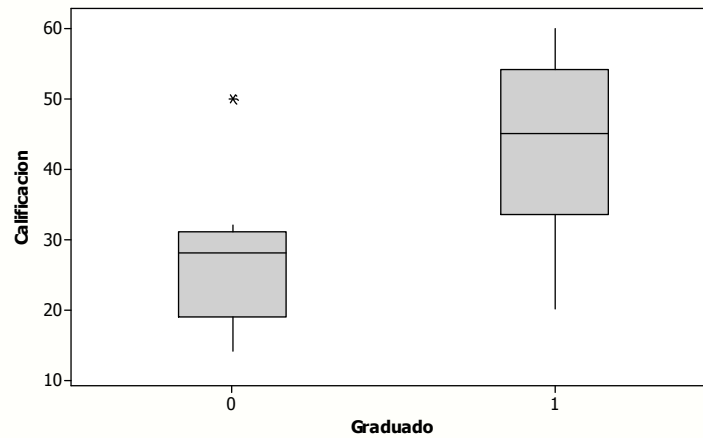
Predictor	Coef	SE Coef	T	P	VIF
Constant	34.48	31.10	1.11	0.289	
Calificacion	5.810	1.015	5.72	0.000	1.518
Graduado	61.25	28.20	2.17	0.051	1.518

S = 43.4354 R-Sq = 86.8% R-Sq(adj) = 84.6%

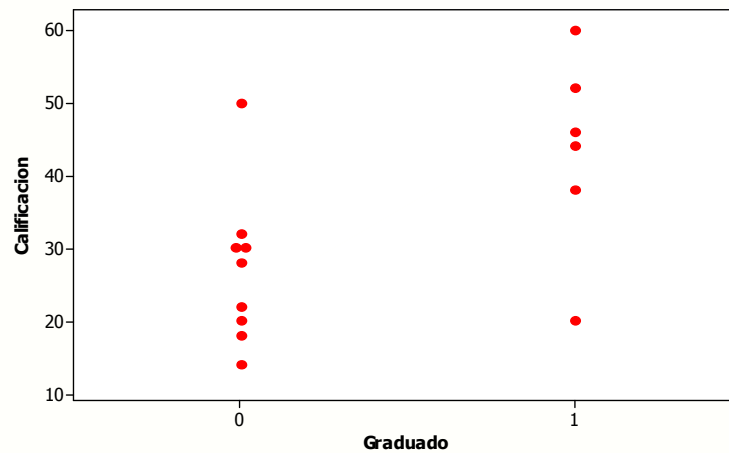
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	148835	74417	39.44	0.000
Residual Error	12	22640	1887		
Total	14	171474			

Boxplot of Calificacion



Individual Value Plot of Calificacion



Henry Hancock

Henry Hancock, gerente de ventas de la compañía NAPE (un distribuidor mayoritario de refacciones para automóviles), quiere desarrollar un modelo para predecir, en el mes de abril, las ventas anuales totales de la región. Si las ventas regionales se pueden predecir, entonces se podrán estimar las ventas totales de la compañía. El número de distribuidoras de la región que mantiene en inventario las refacciones de la compañía y el número de automóviles registrados para cada región, desde el primero de abril, son las dos variables de predicción que Henry quiere investigar. Para esto obtiene los siguientes datos:

Región	Ventas (millones)	Número de Distribuidoras	Número de Autos Registrados
1	52.3	2011	24.6
2	26.0	2850	22.1
3	20.2	650	7.9
4	16.0	480	12.5
5	30.0	1694	9.0
6	46.2	2302	11.5
7	35.0	2214	20.5
8	3.5	125	4.1
9	33.1	1840	8.9
10	25.2	1233	6.1
11	38.2	1699	9.5

Usando los resultados estadísticos presentados, discuta las siguientes preguntas

- Analice la matriz de correlación.
- ¿Son válidos los coeficientes de regresión?
- ¿Cómo puede mejorarse esta ecuación de regresión?

Correlations: Ventas, Distribuidoras, Autos, Ingreso

	Ventas	Distribuidoras	Autos
Distribuidoras	0.739		
Autos	0.548	0.670	
Ingreso	0.936	0.556	0.281

Regression Analysis: Ventas versus Distribuidoras, Autos

The regression equation is
 $Ventas = 10.1 + 0.0110 \text{ Distribuidoras} + 0.195 \text{ Autos}$

Predictor	Coef	SE Coef	T	P	VIF
Constant	10.109	7.220	1.40	0.199	
Distribuidoras	0.010989	0.005200	2.11	0.068	1.813
Autos	0.1947	0.6398	0.30	0.769	1.813

S = 10.3051 R-Sq = 55.1% R-Sq(adj) = 43.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	1043.7	521.8	4.91	0.041
Residual Error	8	849.6	106.2		
Total	10	1893.2			

CONTINUACIÓN

Henry decide investigar una nueva variable de predicción: el ingreso personal en la región. Los datos para esta nueva variable son:

Región	Ingreso Personal (miles de Millones)
1	98.5
2	31.1
3	34.8
4	32.7
5	68.8
6	94.7
7	67.6
8	19.7
9	67.9
10	61.4
11	85.6

Regression Analysis: Ventas versus Distribuidoras, Autos, Ingreso

The regression equation is

Ventas = - 3.92 + 0.00238 Distribuidoras + 0.457 Autos + 0.401 Ingreso

Predictor	Coef	SE Coef	T	P	VIF
Constant	-3.918	2.290	-1.71	0.131	
Distribuidoras	0.002384	0.001572	1.52	0.173	2.473
Autos	0.4574	0.1675	2.73	0.029	1.854
Ingreso	0.40058	0.03779	10.60	0.000	1.481

S = 2.66798 R-Sq = 97.4% R-Sq(adj) = 96.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	1843.40	614.47	86.32	0.000
Residual Error	7	49.83	7.12		
Total	10	1893.23			

Regression Analysis: Ventas versus Autos, Ingreso

The regression equation is

Ventas = - 4.03 + 0.621 Autos + 0.430 Ingreso

Predictor	Coef	SE Coef	T	P	VIF
Constant	-4.027	2.468	-1.63	0.141	
Autos	0.6209	0.1382	4.49	0.002	1.086
Ingreso	0.43017	0.03489	12.33	0.000	1.086

S = 2.87655 R-Sq = 96.5% R-Sq(adj) = 95.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	1827.03	913.52	110.40	0.000
Residual Error	8	66.20	8.27		
Total	10	1893.23			