

PAC 1

Práctica Evaluable

García Prado, Sergio
Taboada Rodero, Ismael José

Universidad de Valladolid - Estadística Descriptiva

14/12/2016

“Con el fin de analizar el consumo energético de una empresa productores de acero se inspeccionaron durante cinco días cada una de las tres líneas de producción. En cada una de ellas se anotaron las variables más relevantes para distintas horas del turno, salvo en la última hora donde sólo se inspeccionó durante cuatro días.”

Ejercicio 1 (1/4)

$$n(\text{consumo}) = 117 \quad (1)$$

$$\overline{\text{consumo}} = \frac{\sum \text{consumo}_i}{n(\text{consumo})} = 139,456 \quad (2)$$

$$me(\text{consumo}) = 140,07 \quad (3)$$

$$\sigma_{\text{consumo}} = S_x = \sqrt{S_x^2} = \sqrt{\sum \frac{X_i^2}{n(X)} - \bar{X}^2} = 55,18353 \quad (4)$$

$$range(\text{consumo}) = |290,72 - 17,5| = 273,22 \quad (5)$$

Ejercicio 1 (2/4)

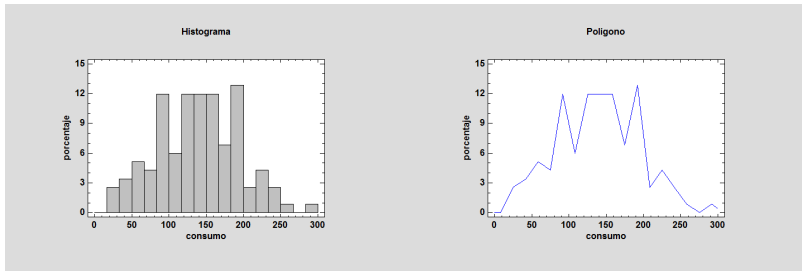


Figura 1: Histograma y Polígono de Frecuencias de *consumo*

Ejercicio 1 (3/4)

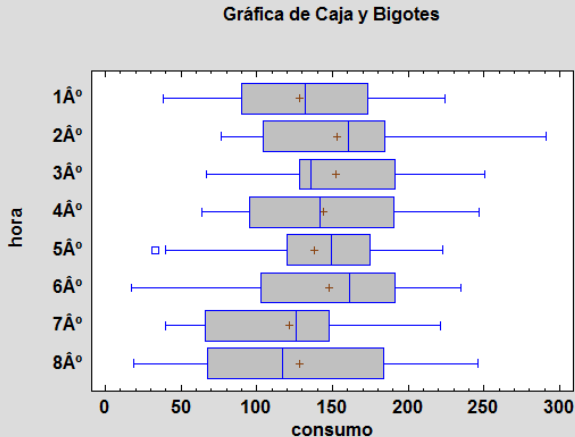


Figura 2: Diagrama de Cajas de *consumo* particionada sobre *hora*

Ejercicio 1 (4/4)

Parámetro	Línea A	Línea B	Línea C
Tamaño	39	39	39
Media	109,841	137,548	170,98
Mediana	99,37	143,82	173,1
Desviación Típica	39,6747	59,6624	47,31871
Máximo	226,38	245,74	290,72
Mínimo	33,18	17,5	68,3
Rango	193,2	228,24	222,42

Cuadro 1: Estadísticos de la variable *consumo* particionada sobre *línea*

Ejercicio 2 (1/2)

	Alta	Media	Baja
OFF	24 52,17 %	16 48,48 %	19 50,00 %
ON	22 47,83 %	17 51,52 %	19 50,00 %

Cuadro 2: Tabla de Frecuencias para temperatura por sistema

Ejercicio 2 (2/2)

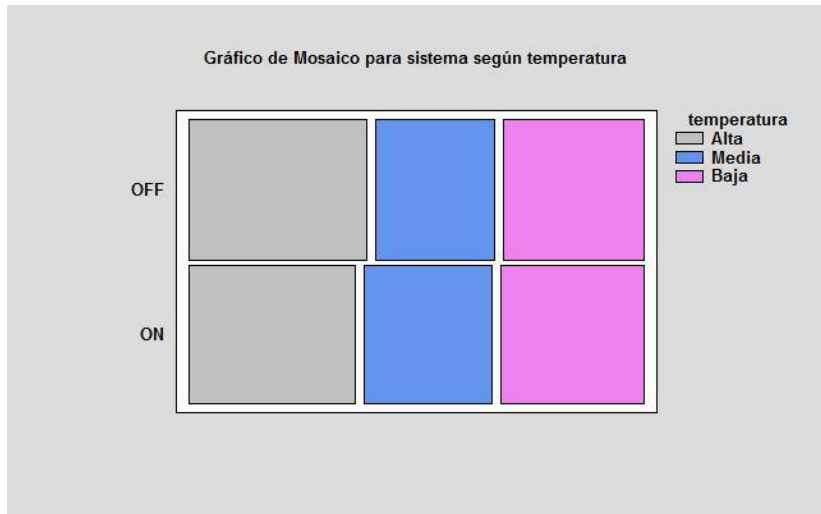


Figura 3: Diagrama de mosaico sobre las variables *sistema* y *temperatura*

Ejercicio 3 (1/4)

“El número de datos recogidos en momentos en que no hubo averías fue **89**, esto es **76.968 %** del total de datos disponibles. Para esos momentos, el *consumo* promedio fue **136.759 Megavatios/Hora**”

Ejercicio 3 (2/4)

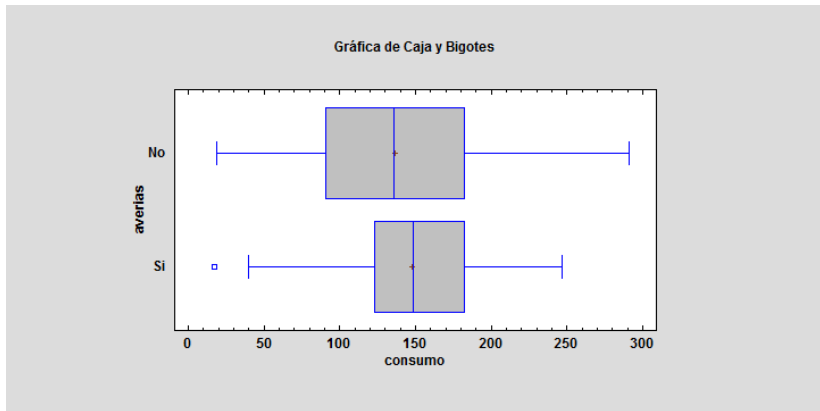


Figura 4: Diagrama de Cajas de *consumo* particionada sobre *averías*

Ejercicio 3 (3/4)

“El percentil 85 de la variable *consumo* es **194.92 Megavatios/Hora.**”

Ejercicio 3 (4/4)

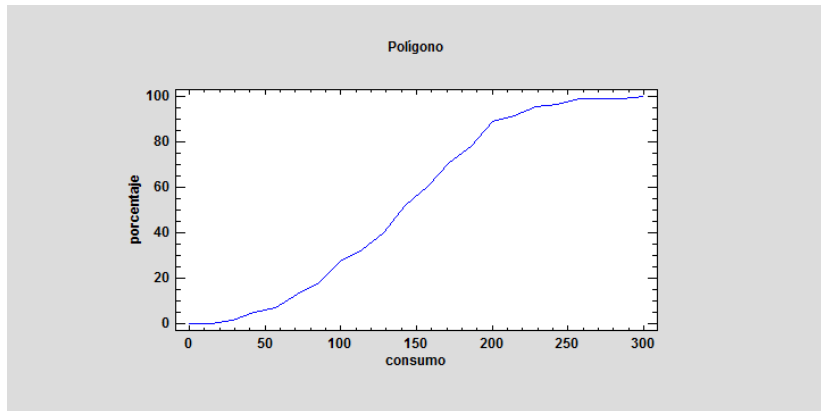


Figura 5: Polígono de frecuencias relativas acumuladas de *consumo*

Ejercicio 4 (1/3)

	consumo	pr.tbc	pr.cc	pr.galv1	pr.galv2
consumo		0,7433 0,0000	0,3853 0,0000	0,4013 0,0000	0,2407 0,0089
pr.tbc	0,7433 0,0000		0,1540 0,0974	0,0661 0,4786	0,1022 0,2727
pr.cc	0,3853 0,0000	0,1540 0,0974		0,3001 0,0010	0,0711 0,4463
pr.galv1	0,4013 0,0000	0,0661 0,4786	0,3001 0,0010		0,0496 0,5950
pr.galv2	0,2407 0,0089	0,1022 0,2727	0,0711 0,4463	0,0496 0,5950	

Cuadro 3: Matriz de correlaciones

Ejercicio 4 (2/3)

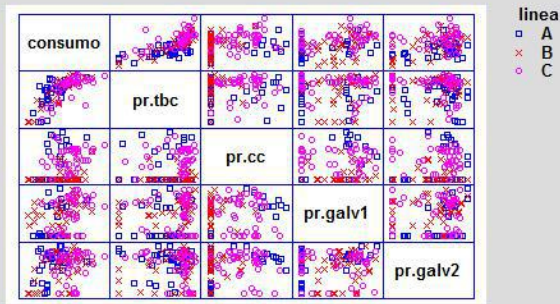


Figura 6: Matriz de gráficos planos

Ejercicio 4 (3/3)

	consumo	pr.tbc	pr.cc	pr.galv1	pr.galv2
consumo		0,7167 0,0000	0,4028 0,0001	0,4343 0,0000	0,2794 0,0080
pr.tbc	0,7167 0,0000		0,1617 0,1300	0,0973 0,3644	0,1208 0,2596
pr.cc	0,4028 0,0001	0,1617 0,1300		0,2855 0,0067	0,1262 0,2385
pr.galv1	0,4343 0,0000	0,0973 0,3644	0,2855 0,0067		0,0837 0,4354
pr.galv2	0,2794 0,0080	0,1208 0,2596	0,1262 0,2385	0,0837 0,4354	

Cuadro 4: Matriz de correlaciones para momentos sin averías

Ejercicio 5 (1/6)

La recta de regresión correspondiente a la variable **consumo** sobre la **producción de acero galvanizado de tipo 1** es:

$$Y = 112,825 + 0,661151X \quad (6)$$

Ejercicio 5 (2/6)

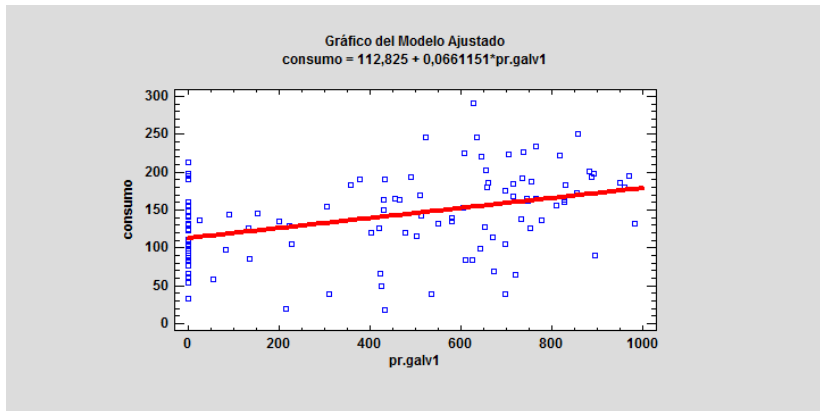


Figura 7: Diagrama de regresión lineal de *consumo* sobre *pr.galv1*

Ejercicio 5 (3/6)

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{7416,625}{55,1853 * 334,929} = \frac{7416,625}{18483,157} = 0,401264 \quad (7)$$

Ejercicio 5 (4/6)

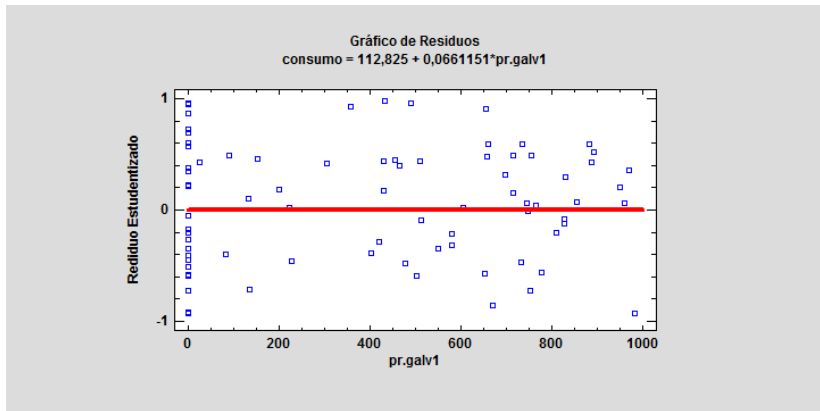


Figura 8: Diagrama de residuos studentizados correspondientes a la regresión lineal de *consumo* sobre *pr.galv1*

Ejercicio 5 (5/6)

Fila	X	Y	Residuos Studentizados.
55	216	19.07	-2.17
56	536	38.39	-2.21
60	698	39.72	-2.42
61	432	17.50	-2.51
88	627	290.72	2.08
117	0	213.23	2.03

Cuadro 5: Residuos studentizados correspondientes a la regresión de la variable *consumo* sobre *pr.galv1*

Ejercicio 5 (6/6)

Parámetro	Valor
Itercepto	112,825
Pendiente	0,661151
Coeficiente de Correlación	0,401264
Residual del caso 1	-0,311627

Cuadro 6: Tabla requerida en el enunciado del ejercicio

Ejercicio 6 (1/5)

	Media	DE	CV	Máximo	Rango
pr.tbc	7567,82	3002,7	39,6773 %	10979,0	10979,0
pr.cc	295,53	358,619	121,348 %	1204,0	1204,0
pr.ca	124,479	161,991	130,135 %	677,0	677,0
pr.galv1	402,803	334,929	83,1495 %	982,0	982,0
pr.galv2	1159,72	577,792	49,8217 %	1963,0	1963,0
pr.pint	188,556	289,446	153,507 %	898,0	898,0
Total	1623,15	2966,56	182,766 %	10979,0	10979,0

Cuadro 7: Tabla de comparación de estadísticos sobre variables de producción

Ejercicio 6 (2/5)

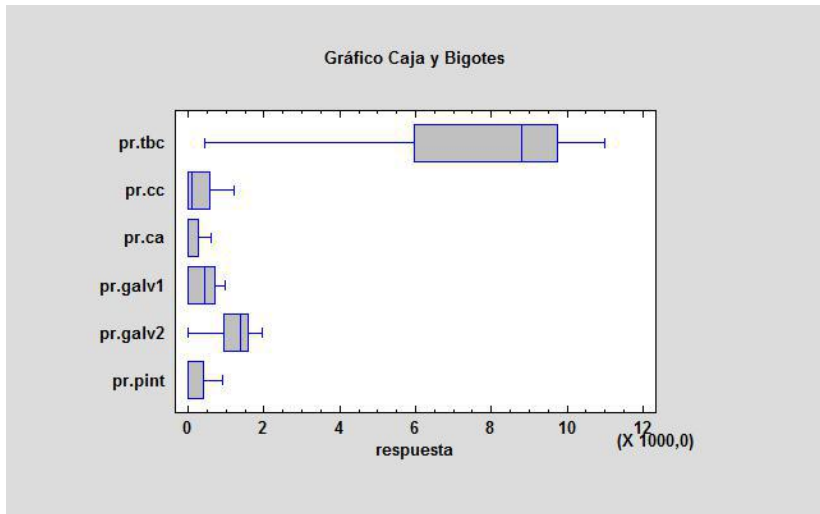


Figura 9: Diagramas de cajas sobre variables de producción

Ejercicio 6 (3/5)

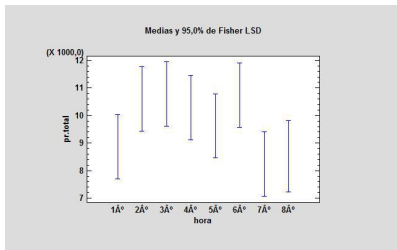


Figura 10: Diagramas de comparación de medias sobre 95 % de confianza

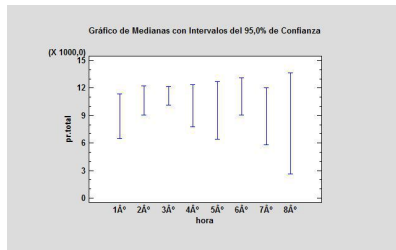


Figura 11: Diagramas de comparación de medianas sobre 95 % de confianza

Ejercicio 6 (4/5)

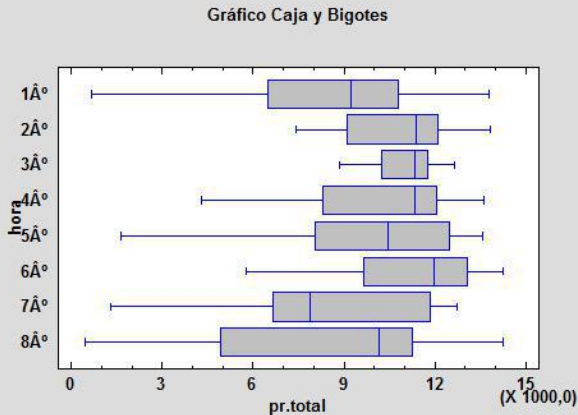


Figura 12: Diagramas de cajas sobre variables de producción agrupadas por la variable *hora*

Ejercicio 6 (5/5)

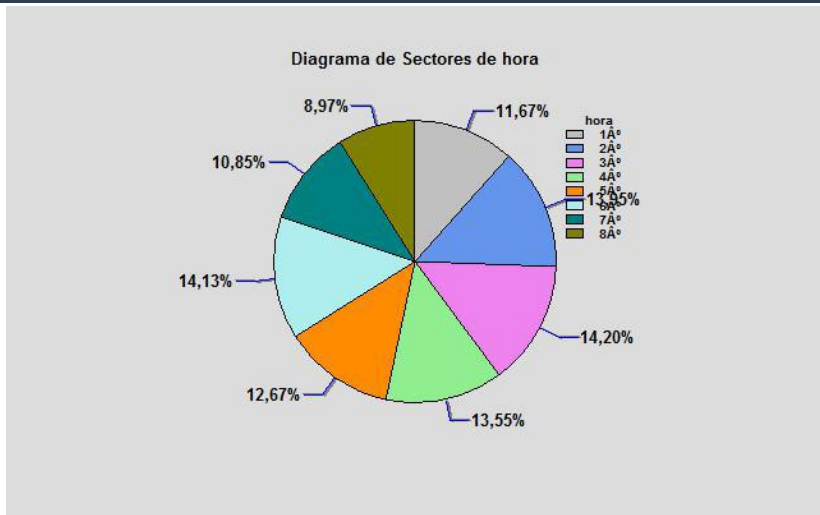


Figura 13: Diagrama de sector sobre porcentaje de producción agrupadas por la variable *hora*

- ▶ Relación entre *consumo* con *líneas*, *averías* o *producción*.Ejercicios 1, 3, 4 y 5
- ▶ Independencia entre *sistema* de control de sobrecalentamiento y *temperatura*.Ejercicio 2
- ▶ Taxonomía de la producción de la empresa.Ejercicio 6(*libre*)
 - ▶ Gran importancia de **bandas calientes de tren**.
 - ▶ Alta producción en horas medias de la jornada.
 - ▶ Baja producción en las últimas horas.
- ▶ Gran importancia de la **estadística descriptiva** en el análisis de variables y conjuntos de datos.