

ESTADÍSTICA (Q)

Segundo Parcial

- 6 de julio de 2010 -

1	2	3	4	Nota

Apellido y Nombres:

Puntaje y criterio de aprobación: El puntaje figura al lado de cada ejercicio. Para aprobar es necesario obtener al menos 60 puntos.

JUSTIFIQUE CON CLARIDAD TODAS SUS AFIRMACIONES

Ejercicio 1: (20p) Una droga cura una rara enfermedad con probabilidad p . Se seleccionaron 90 personas que contrajeron esa enfermedad y se curaron 27.

- (6p) Hallar un intervalo de confianza para p de nivel asintótico 0.9.
- (4p) ¿Qué tamaño de muestra debería tomarse si se desea que el intervalo tenga una longitud menor que 0.1?
- (10p) Por estadísticas históricas se cree que la probabilidad que un paciente enfermo se cure de esa rara enfermedad es al menos 0,6. Se desea que la probabilidad de no confirmar esta creencia cuando en realidad es cierta sea aproximadamente 0.05. Plantee las hipótesis y un test asintótico adecuado para decidir entre ambas hipótesis ¿Qué decisión se toma en base a los datos mencionados en el enunciado?

Ejercicio 2: (20p) Una conocida empresa de cervezas quiere estudiar su nuevo producto. La composición de esta nueva cerveza fue modificada químicamente con el objetivo de disminuir el contenido de alcohol en casi la mitad de lo que es usual. La empresa sostiene que beber hasta 800 cm^3 de dicha bebida no afecta las condiciones neurológicas ni motrices del conductor. Para estudiar dicha afirmación se eligieron en forma aleatoria 10 jóvenes varones entre 18 y 25 años. Se les midió el tiempo de reacción en segundos a un determinado estímulo. Luego se les hizo ingerir 800 cm^3 del producto y a los 60 minutos se les midió nuevamente el tiempo de reacción al mismo estímulo.

ANTES	0,68	0,62	0,75	0,82	0,58	0,80	0,72	0,65	0,84	0,73
DESPUES	0,72	0,62	0,63	0,92	0,68	0,87	0,77	0,70	0,88	0,79

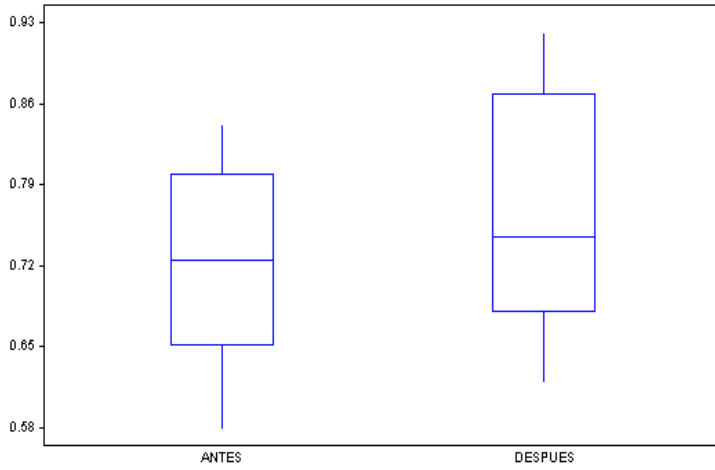
La empresa afirma que si el único error cometido es aleatorio no debería haber diferencia entre los tiempo medidos antes y después de ingerir los 800 cm^3 de dicha bebida. Para analizar esta afirmación queremos realizar un test con un nivel de significación del 5%.

A continuación se encuentran los box-plots correspondientes a las variables “antes”, “después” y “después – antes”. Además se muestran salidas del Statistix correspondientes a distintos tests para las muestras dadas.

- (10p) Defina claramente las variables que intervienen. **Seleccione cuál test realizar** justificando su elección. Se espera que analice los supuestos y los valide con la información que dispone con un nivel de significación de 0.2. Escriba claramente las hipótesis a testear definiendo el o los parámetros involucrados en cada test.
- (7p) Escriba el estadístico del test seleccionado y su distribución bajo la hipótesis nula, si los conoce ¿Qué valor toma dicho estadístico para la muestra dada. Plantee la expresión para calcular el p-valor del test.
- (3p) ¿Respaldan estos datos la afirmación de la empresa? Responda de acuerdo al p-valor que corresponde al test seleccionado.

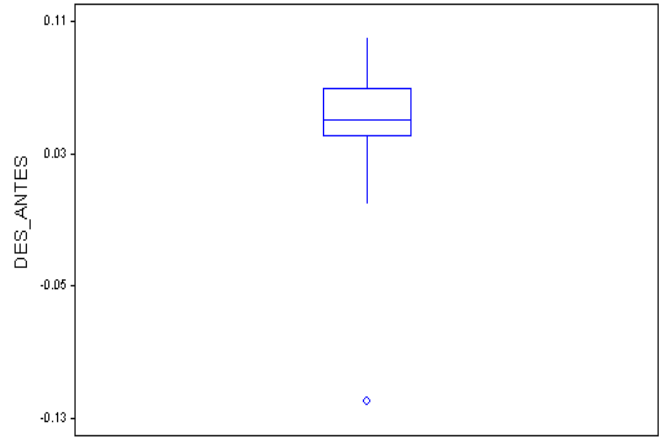
Salidas del Statistix:

Box and Whisker Plot



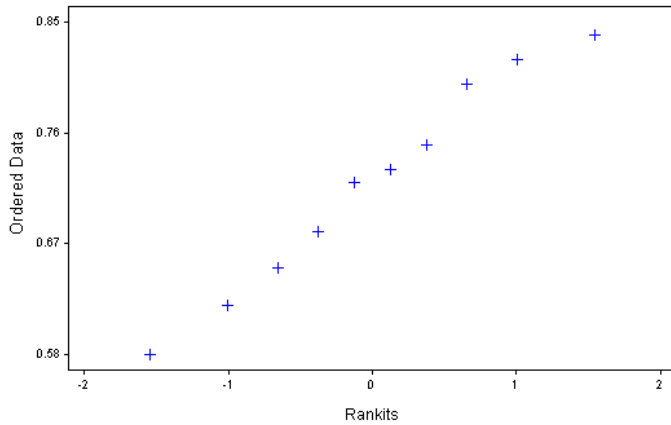
20 cases 2 missing cases

Box and Whisker Plot



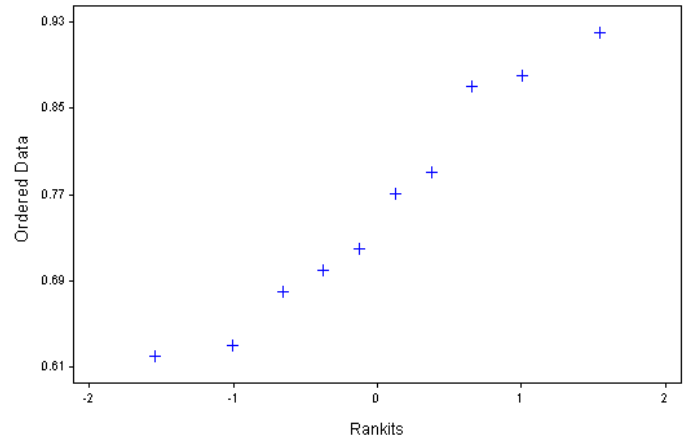
10 cases 1 missing cases

Normal Probability Plot of ANTES



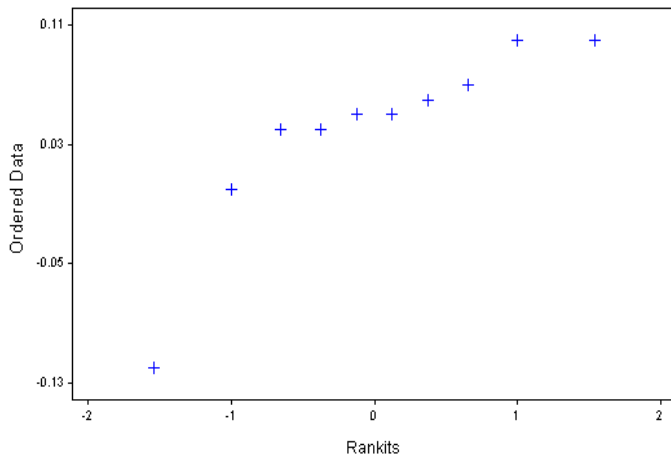
Shapiro-Wilk W 0.9661 P(W) 0.8530 10 cases

Normal Probability Plot of DESPUES



Shapiro-Wilk W 0.9386 P(W) 0.5376 10 cases

Normal Probability Plot of DES_ANTES



Shapiro-Wilk W 0.7864 P(W) 0.0099 10 cases

Shapiro-Wilk Normality Test

Variable	N	W	P
ANTES	10	0.9661	0.8530
DESPUES	10	0.9386	0.5376
DES_ANTES	10	0.7864	0.0099

Paired T Test for DESPUES - ANTES

Null Hypothesis: difference = 0
 Alternative Hyp: difference <> 0

Mean	0.0390
Std Error	0.0200
Mean - H0	0.0390
Lower 95% CI	-6.17E-03
Upper 95% CI	0.0842
T	1.95
DF	9
P	0.0826

Two-Sample T Tests for ANTES vs DESPUES

Variable	Mean	N	SD	SE
ANTES	0.7190	10	0.0869	0.0275
DESPUES	0.7580	10	0.1060	0.0335
Difference	-0.0390			

Null Hypothesis: difference = 0

Alternative Hyp: difference <> 0

Assumption	T	DF	P	95% CI for Difference	
				Lower	Upper
Equal Variances	-0.90	18	0.3801	-0.1301	0.0521
Unequal Variances	-0.90	17.3	0.3805	-0.1303	0.0523

Test for Equality of Variances	F	DF	P
	1.49	9,9	0.2810

Sign Test for DESPUES - ANTES

Number of Negative Differences	1
Number of Positive Differences	8
Number of Zero Differences (ignored)	1

Probability of a result as or more extreme than observed (one-tailed p-value) 0.0195

Wilcoxon Rank Sum Test for DESPUES VS ANTES

Variable	Rank Sum	N	U Stat	Mean Rank
DESPUES	114.50	10	59.500	11.5
ANTES	95.500	10	40.500	9.6
Total	210.00	20		

Exact Permutation Test Two-tailed P-value 0.4925

Normal Approximation with Corrections for Continuity and Ties 0.681

Two-tailed P-value for Normal Approximation 0.49

Wilcoxon Signed Rank Test for ANTES - DESPUES

Sum of Negative Ranks	-36.000
Sum of Positive Ranks	9.0000

Exact probability of a result as or more extreme than the observed ranks (one-tailed p-value) 0.0645

Normal Approximation with Continuity Correction 1.540

Two-tailed P-value for Normal Approximation 0.1235

Ejercicio 3: (35p) Un empresario de la construcción quiere importar una variedad de yeso que fragüe relativamente rápido y que sea económico. Es decir le interesa analizar el precio y la prestación del producto. Para ello podría comprar el que se comercializa en alguno de los siguientes países: Alemania, Brasil, Colombia o Dinamarca. A priori conoce los precios pero no las diferencias entre los tiempos que tarda cada variedad en fraguar. Le interesa conocer si hay diferencias significativas entre las medias del tiempo de fraguado de las cuatro variedades de yeso o si se comportan esencialmente iguales.

Se eligieron aleatoriamente 30 paquetes de yeso de cada una de las cuatro variedades de yeso y de cada bolsa se tomó una cantidad fija de yeso. Los datos del tiempo de fraguado (en minutos) de cada alícuota se identifican usando la letra inicial del país de origen del paquete (variedad), es decir A, B, C y D y se muestran a continuación.

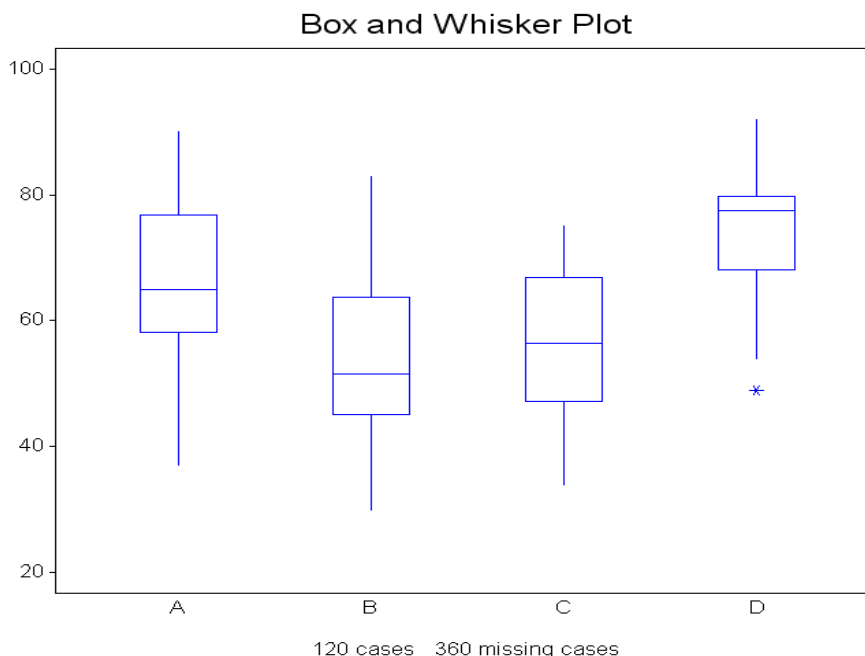
A	51	64	70	63	78	55	67	75	82	61	53	60	62	83	77	90	85	60	70	58	40	61	66	37	54	77	75	57	85	82
B	30	51	68	45	56	49	42	50	72	45	53	47	57	83	54	50	64	65	46	68	33	52	52	42	42	66	58	44	71	39
C	39	54	69	47	66	44	56	55	67	47	58	39	42	45	72	72	69	75	57	54	34	62	50	58	48	63	74	45	71	59
D	92	73	86	84	83	49	68	66	83	80	67	74	63	77	77	54	79	80	85	78	64	80	80	57	75	76	78	83	74	78

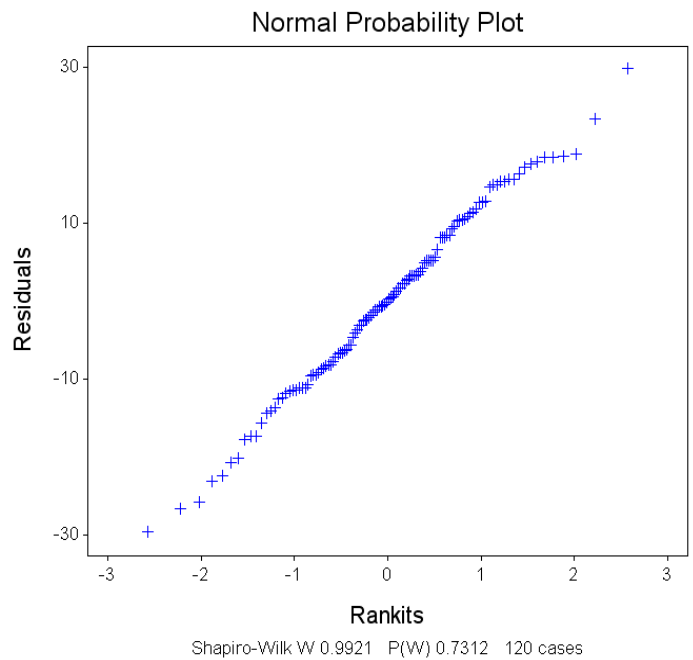
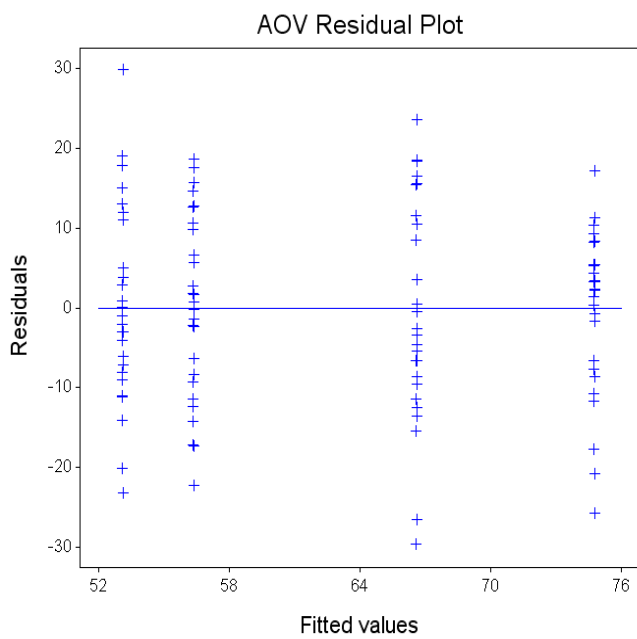
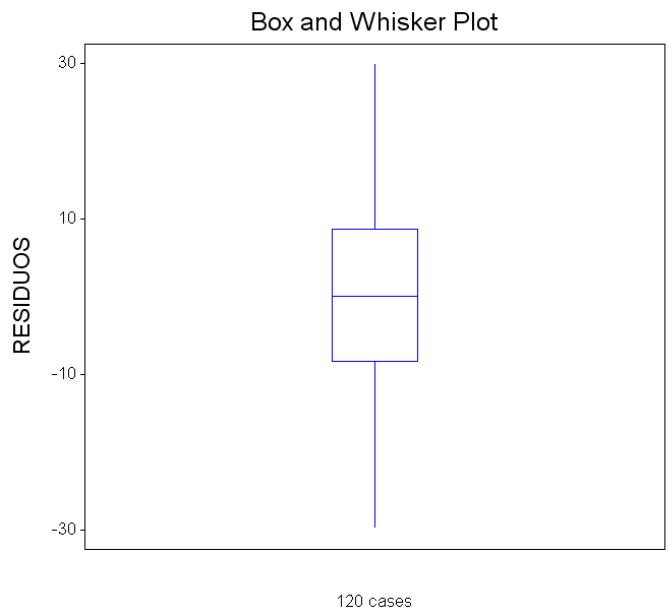
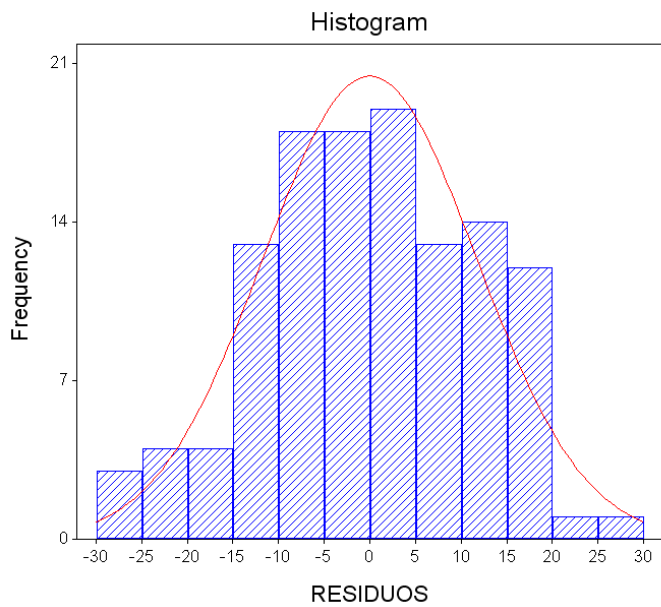
Nos interesa testear si hay diferencias significativas entre los tiempos que tardan los materiales provenientes de los distintos países en fraguar o si las diferencias observadas se atribuyen al azar.

Suponga que se verifican las hipótesis del modelo de análisis de la varianza (ANOVA). Al final del ejercicio se presenta la salida del Statistix para el ANOVA. Responda en base a dicha información justificando claramente cada elección.

- (5p) Plante un modelo para este problema definiendo claramente las variables y parámetros involucrados en el problema. Indique cuáles son los supuestos necesarios para aplicar el análisis de la varianza. Escriba las hipótesis que se desean testear.
- (6p) Analice la validez cada uno de esos supuestos. Indique claramente qué resultados está usando, es decir en qué test y/o gráfico basa su conclusión. Cualquier test que utilice para evaluar supuestos hágalo con nivel de significación 0.20.
- (4p) Escriba el estadístico del test del ANOVA y halle su valor para los datos observados ¿Cuál es la distribución del estadístico bajo la hipótesis nula? ¿Cuál es la regla de decisión del test? Plantee el p-valor.
- (3p) Escriba la conclusión del test F a nivel 5%
- (8p) Utilice el método que le parece más adecuado para construir todos los intervalos de confianza para las diferencias de medias, con nivel de confianza simultáneo de al menos 95% ¿En qué se basa su elección? El nivel global, ¿es exactamente 95% o al menos 95%? Encontrará una ayuda junto con las salidas.
- (4p) Detecte mediante tests con nivel simultáneo, qué pares de medias difieren al 5%. Plantee las hipótesis correspondientes.
- (5p) Elegir la salida de Tukey o la de Bonferroni (abajo) para marcar cuáles son los grupos homogéneos. ¿Le parece razonable lo que se observa en los boxplots de los cuatro tipos de yeso? Si tuviera que elegir dos de los cuatro yesos en base al tiempo de fraguado, qué le recomendaría al constructor? Explícite en qué basa su elección.

Salidas del Statistix:





Shapiro-Wilk Normality Test

Variable	N	W	P
RESIDUOS	120	0.9921	0.7312

One-Way AOV for: A B C D

Source	DF	SS	MS	F	P
Between	3	8773.4	2924.46	20.8	0.0000
Within	116	16317.0	140.66		
Total	119	25090.4			

Grand Mean 62.717 CV 18.91

	Chi-Sq	DF	P
Bartlett's Test of Equal Variances	2.56	3	0.4646
Cochran's Q			0.3151
Largest Var / Smallest Var			1.8107

Variable	Mean
A	66.600
B	53.133
C	56.367
D	74.767

Bonferroni All-Pairwise Comparisons Test

Variable	Mean	Homogeneous Groups
D	74.767	-----
A	66.600	-----
C	56.367	-----
B	53.133	-----

D	74.767	-----
A	66.600	-----
C	56.367	-----
B	53.133	-----

Completar si utilizó este método

Alpha 0.05 Standard Error for Comparison 3.0623

Critical T Value 2.684 Critical Value for Comparison 8.2200

Tukey HSD All-Pairwise Comparisons Test

Variable	Mean	Homogeneous Groups
D	74.767	-----
A	66.600	-----
C	56.367	-----
B	53.133	-----

D	74.767	-----
A	66.600	-----
C	56.367	-----
B	53.133	-----

Completar si utilizó este método

Alpha 0.05 Standard Error for Comparison 3.0623

Critical Q Value 3.687 Critical Value for Comparison 7.9829

Diferencia de promedios

A-B	13.467
A-C	10.233
A-D	-8.167
B-C	-3.234
B-D	-21.634
C-D	-18.4

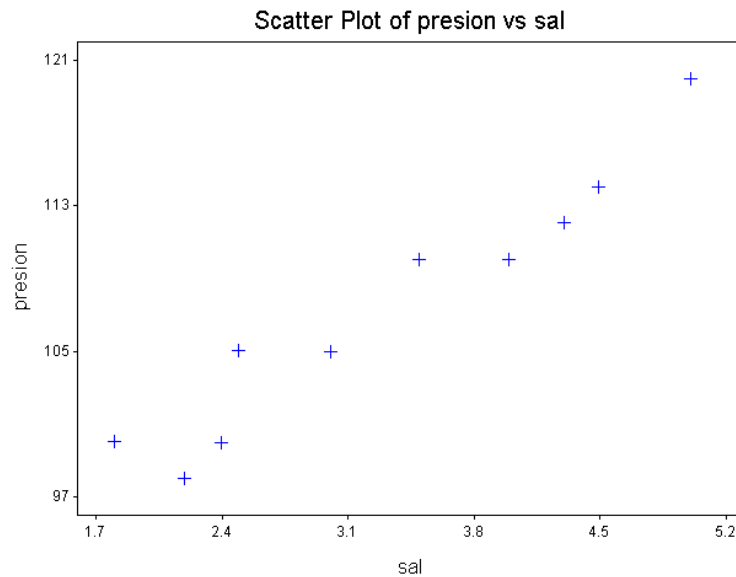
Ejercicio 4: (25p) Se quiere estudiar la asociación entre consumo de sal y presión arterial. A un conjunto de diez voluntarios se les administra distintas dosis de sal (en gramos) en su dieta durante un mes y se mide su presión arterial máxima (mm de Hg) al finalizar el mes. Los resultados fueron los siguientes:

sal	1,8	2,2	2,4	2,5	3	3,5	4	4,3	4,5	5
presion	100	98	100	105	105	110	110	112	114	120

Luego del enunciado se presentan algunas salidas del Statistix. Para cada una de las justificaciones indique en qué gráfico o salida se basa para darlas.

- (9p) El investigador propone que la presión arterial máxima crece linealmente con el contenido de sal en la dieta. Escriba un modelo adecuado indicando los parámetros y definiendo claramente las variables involucradas ¿Es razonable suponer este modelo? Escribir la recta estimada planteando las expresiones de los estimadores que intervienen. Interprete los parámetros de la recta estimada en términos de este problema.
- (5p) Escriba los supuestos necesarios para que las conclusiones respecto de los tests y los intervalos de confianza sean válidas. Analice si se cumplen en este caso indicando claramente en qué se basa para obtener sus conclusiones. Cualquier test que utilice para evaluar supuestos hágalo con nivel de significación 0,1.
- (5p) ¿Le parece que la cantidad de sal influye o está asociada con la presión arterial? Responda la pregunta realizando un test a nivel 5%. Recuerde plantear las hipótesis, el estadístico y su distribución, y el p valor.
- (6p) Plantee una fórmula para el coeficiente de determinación y llegue a su valor utilizando la salida del Statistix ¿Cómo interpreta esta medida para estos datos?

Salidas del Statistix:



Descriptive Statistics

Variable	N	Mean	SD
sal	10	3.3200	1.0983
presion	10	107.40	7.0427

Unweighted Least Squares Linear Regression of presion

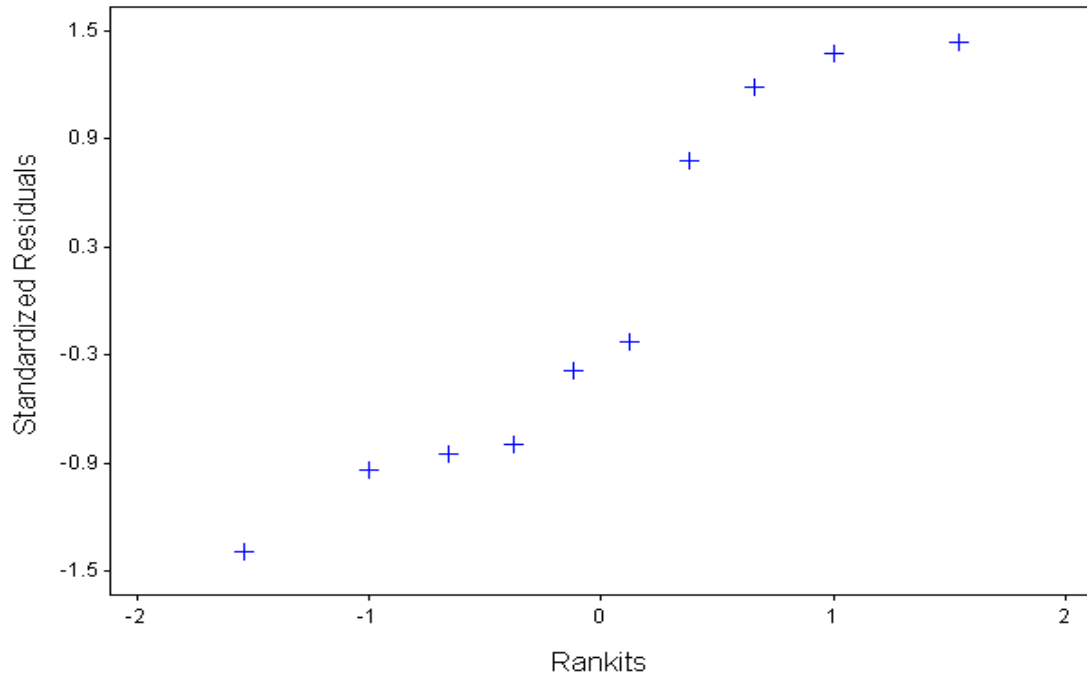
Predictor Variables	Coefficient	Std Error	T	P
Constant	86.9038	2.13349	40.73	0.0000
sal	6.17354			

Atención!!
Esta salida tiene datos faltantes

Source	DF	SS	MS	F	P
Regression	1	413.751		101.38	0.0000
Residual					
Total	9	446.400			

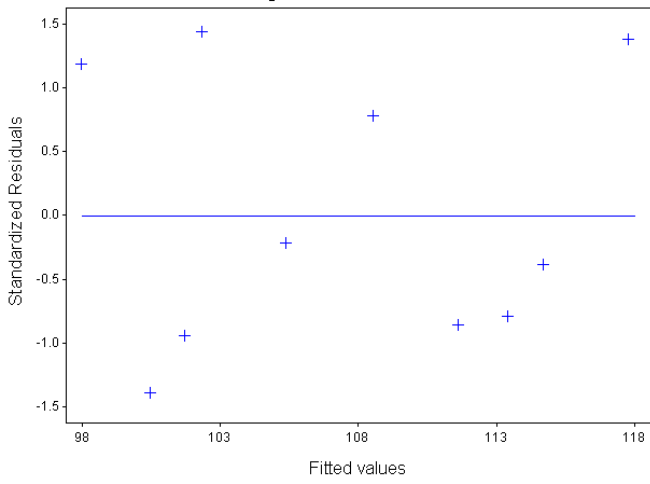
R-Squared Resid. Mean Square (MSE)
Standard Deviation

Normal Probability Plot



Shapiro-Wilk W 0.8823 P(W) 0.1387 10 cases

Regression Residual Plot



Box and Whisker Plot

