

ESTADÍSTICA (Q)

Recuperatorio del Segundo Parcial

- 20 de julio de 2010 -

1	2	3	4	Nota

Apellido y Nombres: **LU:**

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: (17 puntos)

- a) (6 p) A partir de una muestra aleatoria de 9 observaciones de una cierta variable aleatoria X se obtiene una media muestral de 43.6 cm y un desvío estándar muestral de 11.3 cm. Hallar un intervalo de nivel de confianza exacto o aproximado 90% para $E(X)$. ¿Qué supuestos debe hacer?
- b) (6 p) Suponga que se toma una muestra aleatoria de 256 observaciones de la misma variable aleatoria y se obtiene una media muestral de 43.8 cm y un desvío estándar muestral de 11.6 cm. Hallar un intervalo de nivel de confianza exacto o aproximado 90% para $E(X)$. ¿Qué supuestos debe hacer?
- c) (2 p) Compare los dos intervalos en posición y longitud.
- d) (3 p) ¿Por qué no resultó ser la desviación estándar "s" de la segunda muestra mucho menor que la de la primera siendo que la segunda muestra tiene un número mucho mayor de observaciones?.

Ejercicio 2: (25 puntos) Se consideran dos muestras de gusanos de seda, una de machos (LONG_M) y otra de hembras (LONG_H), elegidos todos de diferentes localizaciones en el norte de Asia. Se le mide a cada uno la longitud en cm durante la cuarta fase de la metamorfosis. Los datos se encuentran en la siguiente tabla.

LONG_H	LONG_M
5.2	7.2
4.3	4.2
5.4	5.7
4.2	6.6
5	5
5.2	6.9
5.2	6.6
7	7
4.8	6.5
5.7	6.2

Se sabe que la longitud de los machos tiene distribución $N(\mu_M, \sigma_M^2 = 0.9)$ y la de hembras $N(\mu_H, \sigma_H^2 = 0.6)$. Se desea saber si existen diferencias significativas entre las longitudes medias de machos y hembras.

- a) (12 p) Defina las variables que intervienen. Escriba claramente las hipótesis a testear definiendo el o los parámetros involucrados. Proponga un test a nivel 0.1 para las hipótesis planteadas. Escriba el estadístico del test seleccionado y su distribución bajo la hipótesis nula. Escriba la conclusión del test.
- b) (6 p) Construir un intervalo de confianza de nivel 0.9 para la diferencia de medias ¿Hay razones para sospechar que la longitud de los machos es superior al de las hembras?
- c) (7 p) Calcular la probabilidad de cometer un error de tipo II cuando la diferencia de medias entre machos y hembras es de 1,5.

Variable	Descriptive Statistics			
	N	Sum	Mean	SD
LONG_H	10	52	5.2000	0.7846
LONG_M	10	61.900	6.1900	0.9562

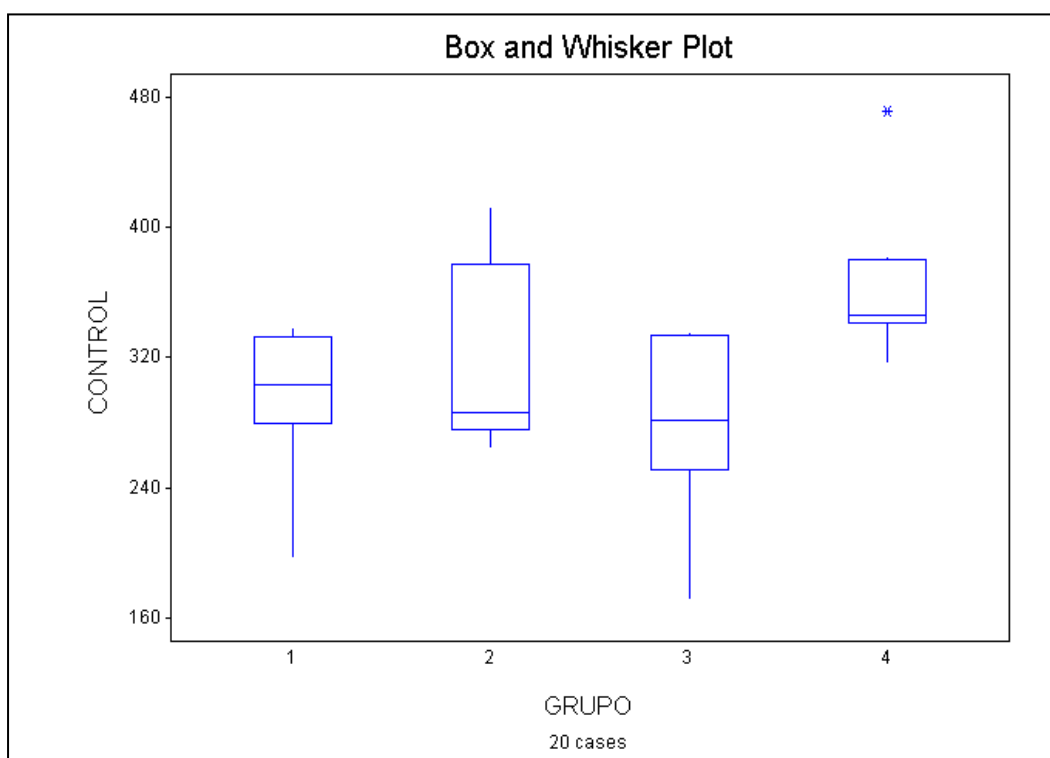
Ejercicio 3: (28 puntos) Los siguientes datos representan el número de bacterias aisladas del tracto gastrointestinal de cuatro grupos de ratas. Estos cuatro grupos fueron grupos testigo o control (es decir que no recibieron tratamientos) de otros estudios.

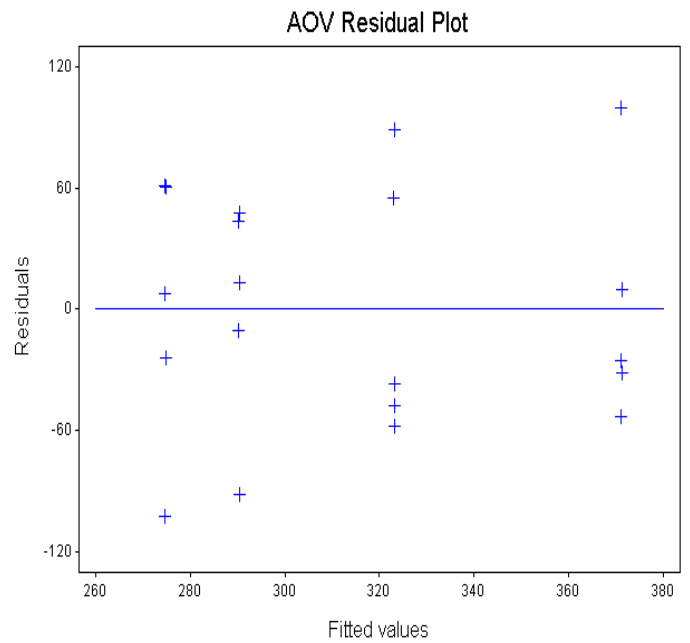
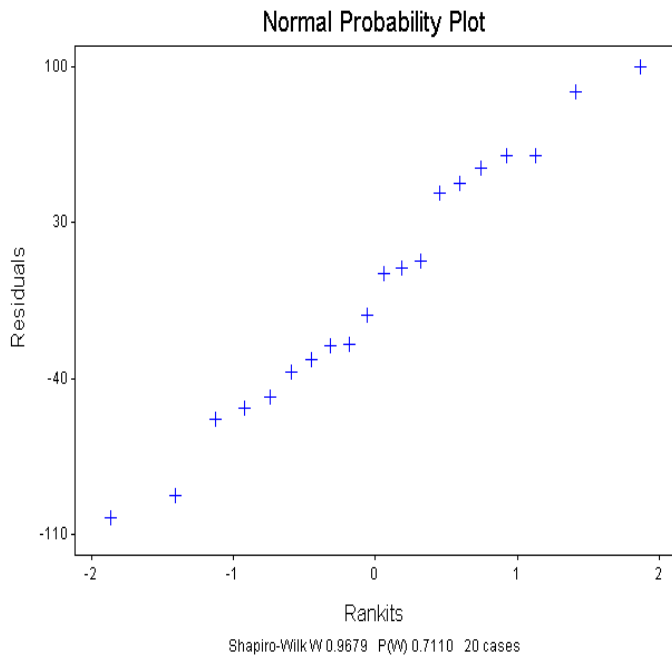
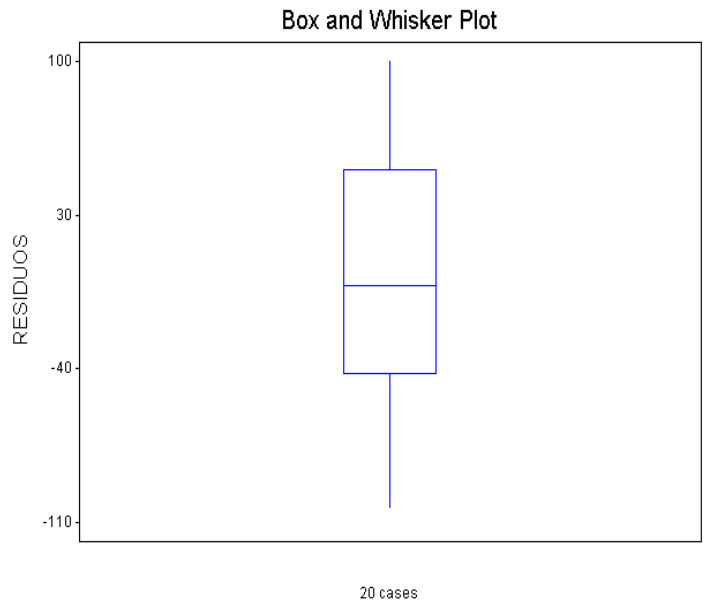
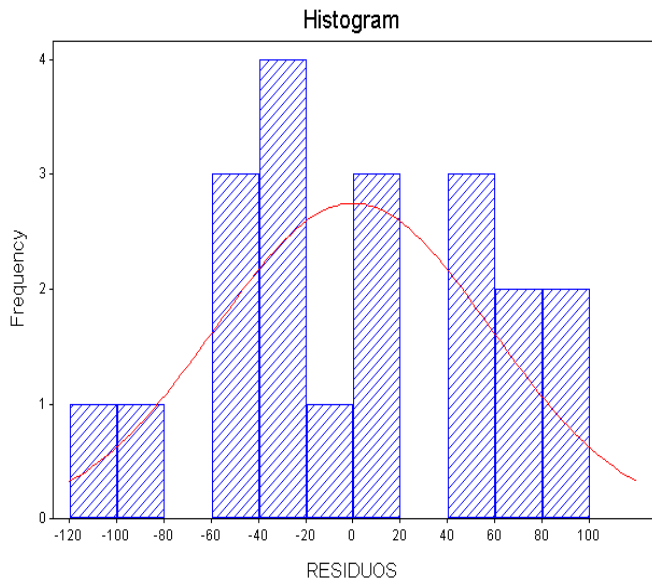
CONTROL1	CONTROL2	CONTROL3	CONTROL4
279	378	172	381
338	275	335	346
334	412	335	340
198	265	282	471
303	286	250	318

Nos interesa testear si hay diferencias significativas entre los números de bacterias en los cuatro grupos de ratas o si las diferencias observadas pueden ser atribuidas al azar.

- (6 p) Plante un modelo de análisis de la varianza (ANOVA) 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.
- (6 p) 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.
- (8 p) 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. Escriba la conclusión del test F a nivel 5%.
- (4 p) Supongamos que la presencia del dato atípico que se observa en el boxplot, influencia negativamente en la validación del supuesto de normalidad. Se quiere testear diferencias significativas entre los números de bacterias en los cuatro grupos de ratas, pero usando un test no paramétrico. En dicho test se asigna un rango a cada una de las observaciones. Se pide calcular los rangos de todas las observaciones y hacer un promedio de rangos por grupos.
- (4 p) Escriba las hipótesis que se desean testear en el test no paramétrico. Escribir la conclusión del test no paramétrico a nivel 5%.

Salidas del Statistix: alguna o algunas de las siguientes salidas pueden tener datos faltantes.





Shapiro-Wilk Normality Test

Variable	N	W	P
RESIDUOS	20	0.9679	0.7110

One-Way AOV for CONTROL by GRUPO

Source	DF	SS	MS	F	P
GRUPO	3	27234.2			0.1195
Error					
Total	19	91187.8			

	Chi-Sq	DF	P
Bartlett's Test of Equal Variances	0.16	3	0.9844
Cochran's Q			0.2890
Largest Var / Smallest Var			1.4225

GRUPO	Mean
1	290.40
2	323.20
3	274.80
4	371.20
Total	314.90
Observations per Mean	5

Kruskal-Wallis One-Way Nonparametric AOV for CONTROL by GRUPO					
GRUPO	Mean Rank	Sample Size			
1	X	5			
2	X	5			
3	X	5			
4	X	5			
Total	10.5	20			
Kruskal-Wallis Statistic			6.2047		
P-Value, Using Chi-Squared Approximation			0.1021		
Parametric AOV Applied to Ranks					
Source	DF	SS	MS	F	P
Between	3	217.000	72.333	2.59	0.0892
Within	16	447.500	27.9688		
Total	19	664.500			

Ejercicio 4: (30 puntos) Los siguientes datos representan el peso al nacer (PESOS) medidos en onzas de 30 bebés y su incremento porcentual (TASA) entre los 90 y 100 días después del nacimiento.
1 oz = 28,35 gramos

PESO	72	112	111	107	119	92	126	80	81	84	115	118	128	128	116
TASA	68	63	66	72	52	75	76	118	120	114	29	42	48	50	59
PESO	125	126	122	126	127	86	142	132	87	123	133	106	103	118	94
TASA	27	60	71	88	63	88	53	50	111	59	76	72	90	68	91

Supondremos que los valores de la variable PESO no tienen error. Luego del enunciado se presentan algunas salidas del Statistix. Indique en qué gráfico o salida se basan sus justificaciones o conclusiones.

- (8 p) Se sospecha que el incremento porcentual de peso de los bebés se relaciona linealmente con el peso al nacer. Escriba un modelo adecuado indicando los parámetros y definiendo claramente las variables involucradas. Gráficamente ¿le parece 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.
- (8 p) 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 0,2.
- (4 p) ¿Le parece que el peso al nacimiento influye o está asociada con su incremento porcentual? Responda la pregunta realizando un test a nivel 5%. Plantear las hipótesis, el estadístico, su distribución, y el p valor.
- (4 p) Calcular el valor predicho y un intervalo de predicción del 95% para la variable TASA cuando el peso al nacer (PESO) es 100.
- (6 p) Estimar el valor de la variable PESO sabiendo que la variable TASA = 80. Calcular un intervalo de confianza aproximado del 95% para el valor verdadero de la variable PESO cuando la variable TASA = 80.

Salidas del Statistix: alguna o algunas de las siguientes salidas pueden tener datos faltantes.

Descriptive Statistics					Aclaración:
Variable	N	Mean	SD	Sum	
PESO	30	112.97	17.248		$(PESO^2)_i = PESO_i \times PESO_i$ $(PESO \times TASA)_i = PESO_i \times TASA_i$
TASA	30	72.800	22.970		
PESO ²	30			391471	
PESO*TASA	30			237801	

Predicted/Fitted Values of TASA			
Lower Predicted Bound	55.229	Lower Fitted Bound	79.266
Predicted value		Fitted Value	
Upper Predicted Bound	117.18	Upper Fitted Bound	93.143
SE (Predicted Value)	15.122	SE (Fitted Value)	3.3872
Percent Coverage	95.0		
Corresponding T	2.05		

Predictor Values: PESO = 100.00

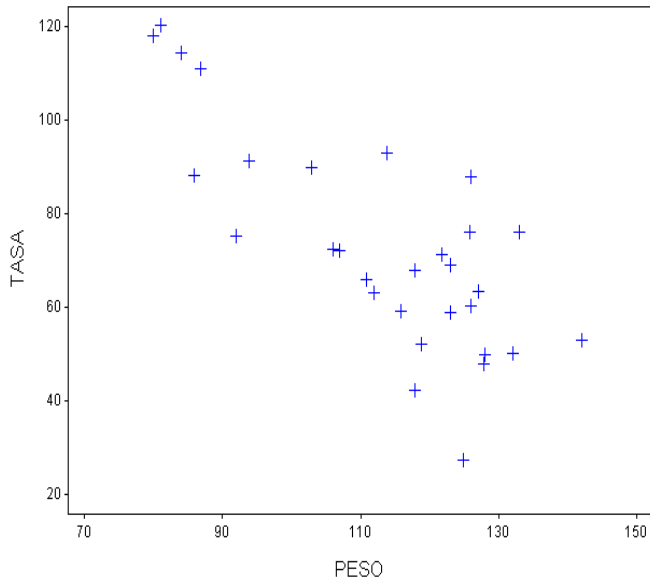
Unweighted Least Squares Linear Regression of TASA

Predictor Variables	Coefficient	Std Error	T	P
Constant		18.1254		0.0000
PESO			-6.52	0.0000

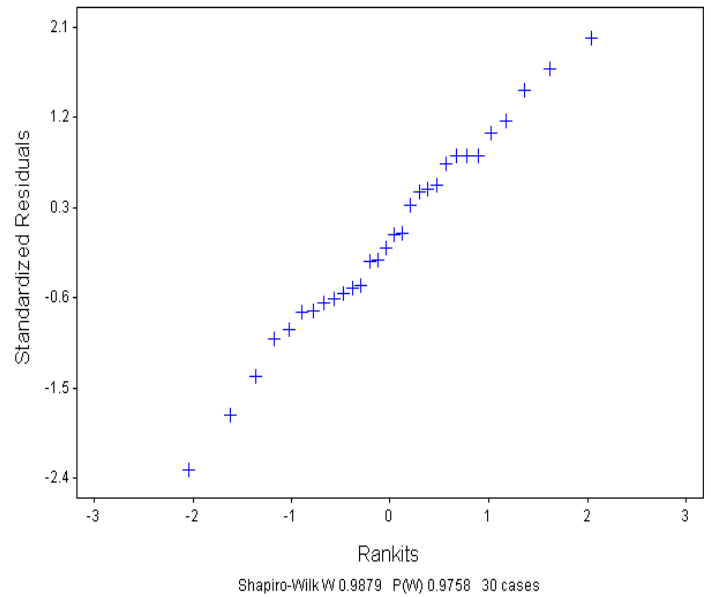
R-Squared 0.6025 Resid. Mean Square (MSE) 217.198
Standard Deviation 14.7376

Source	DF	SS	MS	F	P
Regression	1	9219.3	9219.26	42.45	0.0000
Residual	28	6081.5	217.20		
Total	29	15300.8			

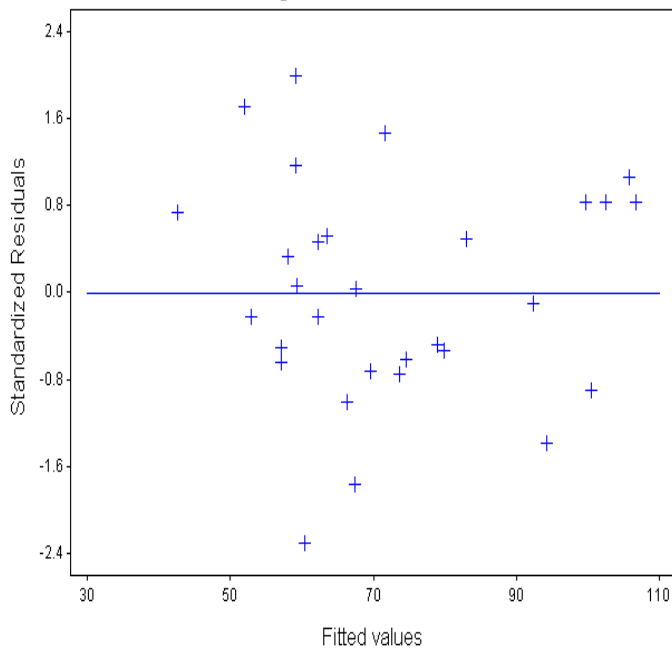
Scatter Plot of TASA vs PESO



Normal Probability Plot



Regression Residual Plot



Box and Whisker Plot

