

# ESTADISTICA-Examen de prácticas. Septiembre 2007

Nombre:.....

Titulación:.....

Grupo en el que estás matriculado:.....

- No está permitido utilizar (ni acceder a) ningún tipo de documentación, ni usar calculadora.
- Los datos están en la web de la asignatura. Descarga los datos en el Escritorio del ordenador

1. El fichero **parqueeolico1.sf3** tiene la velocidad del viento (variable velocidad) en metros/segundo que se ha registrado en un parque eólico. Cada observación corresponde a valores medidos en una hora. Se pide:

Coefficiente de asimetría= 0.48

Rango intercuartílico= 6.25

Varianza= 17.69

Percentil 30= 4.86

Q1= 4.16

Q2= 7.22

Q3= 10.4

2. El fichero **VelViento730.sf3** tiene información de la velocidad del viento (en metros por segundo) en un conjunto de horas en dos parques eólicos (Parque1 y Parque2)

- a. Ajusta una distribución normal a la velocidad del viento del parque 1 (Parque1), buscando una transformación si fuese necesario. ¿Qué resultados se obtienen? Justifica la respuesta

*La distribución no es normal: el histograma es asimétrico, y el test de la chi-cuadrado tiene un p-valor muy pequeño. Como la no normalidad es por su asimetría positiva intentamos una transformación del tipo  $X^c$ , con  $c < 1$ .*

*Con  $c = [0.40; 0.50]$  resulta un histograma parecido a la normal, y el test de la chi-cuadrado tiene  $p\text{-valor} < 0.05$  (nota: son válidas transformaciones en el rango 0.26-0.59). Elegimos la transformación  $X^{0.45}$*

- b. Utilizando dicho ajuste, calcula la probabilidad de que en una hora el viento supere los 5 m/s

$$P(x > 5) = P(x^{0.45} > 5^{0.45}) = P(x^{0.45} > 2.063) = 1 - 0.46 = 0.54; \text{ donde } x^{0.45} = N(9.62, 0.57^2).$$

3. A un servidor web acceden clientes de forma independiente a un ritmo medio de 60 accesos/minuto. El servidor tiene una capacidad máxima de gestión de 80 accesos/minuto. A partir de ese número de accesos se dice que el servidor está saturado. Cuando el servidor se encuentra saturado, los usuarios que intentan acceder son rechazados. Calcula, utilizando exclusivamente el Statgraphics:

- a. ¿Qué proporción del tiempo estará el servidor saturado?

$$X = \text{accesos en un minuto} = \text{Poi}(\lambda = 60).$$

$$P(X > 80) = 0.0056$$

- b. ¿Qué capacidad mínima debería tener el servidor para que la probabilidad de saturarse fuese menor de 0.001?

*Buscamos en icdf el valor C tal que  $P(X > C) = 0.001$ , y se obtiene que  $C = 85$*

- c. ¿Cuál es la probabilidad de que haya más de 5 segundos sin accesos?

$$T = \text{tiempo, en segundos, entre accesos} = \text{Exp}(\lambda^* = 1 \text{ acc / seg})$$

$$P(T > 5) = 0.0067$$

Nombre:.....

4. Un proceso de fabricación produce diodos de valor de Tensión nominal de 650 mv (milivoltios). Tras analizar el proceso, el ingeniero de calidad ha establecido que cuando el proceso funciona adecuadamente, la tensión de los diodos tiene un valor medio de 650 y una varianza de 40. Si dichos parámetros cambiasen sería indicio de que algo anómalo podría estar sucediendo, y habría que investigar las causas. El fichero **caracteristicas\_diodo.sf3** tiene valores de una muestra de 100 diodos de dicho proceso. A la vista de esta muestra ¿Puede decirse que el proceso funciona según lo previsto? Justifica y razona adecuadamente (de forma breve) tanto la metodología que vas a utilizar como las conclusiones.

*Si el proceso funciona según lo previsto debe mantener su media y su varianza.*

*Para analizar si se mantiene la media hacemos el contraste  $H_0 : \mu = 650; H_1 \neq 650$ . Como el tamaño muestra es grande, no necesitamos preocuparnos por la distribución de la variable Tensión. El pvalor es inferior a 0.05, por lo que tenemos que rechazar que la media se mantenga. También podemos hacer un intervalo de confianza para la media. Un intervalo del 95% es*

*95,0% confidence interval for mean: 688,02 +/- 1,20216 [686,818;689,222]*

*que no contiene al 650. Rechazamos con un nivel de significación del 5% que la media se mantenga*

*Para ver si se mantiene la varianza hacemos un intervalo de confianza para la desviación típica. Este intervalo sólo es válido si la distribución es normal. El histograma de la variable tensión parece normal, y el test de la chi-cuadrado muestra un pvalor > 0.05. Podemos por tanto asumir normalidad de la población. El intervalo para la desviación típica es*

*95,0% confidence interval for standard deviation: [5,31949;7,03813]*

*donde se comprueba que sí contiene a 6.32. Por tanto no podemos rechazar con un nivel de significación del 5% que la varianza no haya cambiado.*

*Como conclusión podemos decir que el proceso parece haber cambiado su media, y ha dejado entonces de funcionar según lo previsto.*

5. El fichero **cardata.sf** tiene varias variables de un conjunto de 155 automóviles. Queremos construir una recta de regresión que explique la aceleración del coche (variable **accel**=segundos que tarda en alcanzar una determinada velocidad).
- ¿Qué variable es más apropiada para hacer esa regresión, el peso del vehículo (**weight**=kilos) o su potencia (**horsepower**= caballos)? Justifica brevemente tu respuesta.

*El gráfico XY de weight frente a accel muestran una ausencia de relación (el coeficiente de correlación es - 0.04). El gráfico de accel con horsepower muestra una mayor relación (a valores bajos de horsepower la relación parece no lineal)*

- Escribe la recta de regresión seleccionada  
*accel = 20,7949 - 0,0509812\*horsepower*