

ESTADISTICA-Examen de prácticas. Junio 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 **COitv.sf3** tiene, en su variable CO, la concentración de monóxido de carbono (CO) de los gases emitidos por 100 vehículos en una estación ITV. Se pide:

Coefficiente de asimetría= 1.456

Rango intercuartílico= 0.82

Percentil 30= 0.2

Q1= 0.18

Q2= 0.3

Q3= 1.0

2. El fichero AlumnosIndustriales.sf3 tiene información de una muestra de 95 estudiantes. La variable Dinero indica la cantidad de dinero (en pesetas) que llevan cada uno en el bolsillo. Se pide:

- a. Ajusta una distribución normal a dicha variable, 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 $c=0.35$ 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). El ajuste es por tanto $X^{0.35}$

- b. Utilizando dicho ajuste, calcula la probabilidad de que un alumno lleve más de 500 pesetas en el bolsillo.

$$P(x > 500) = P(x^{0.35} > 500^{0.35}) = P(x^{0.35} > 8.803) = 1 - 0.43 = 0.57; \text{ donde } x^{0.35} = N(9.62, 4.65).$$

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. El fichero **cardata.asf** tiene varias variables de un conjunto de 155 automóviles. Queremos construir una recta de regresión que explique el consumo de un vehículo (variable **mpg**=millas por galón de combustible).
- a. ¿Qué variable es más apropiada para hacer esa regresión, el peso del vehículo (**weight**=kilos) o su potencia (**horsepower**= caballos)? Justifica brevemente tu respuesta.

Los gráficos XY de weight frente a mpg, y horsepower frente a mpg muestrna relaciones lineales aparentemente iguales de fuertes. Sin embargo, el coeficiente de correlación con weight es, en valor absoluto, mayor (-0.83). Por tato, la mejor variable para la regresión es weight

- b. Escribe la recta de regresión seleccionada

$$\text{mpg}=55.897-0.01 \text{ weight}$$

5. 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 en el proceso productivo, 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.

Nombre:.....

6. Un pediatra afirma que no hay diferencias entre niños y niñas en su peso al nacer. Para comprobar esta afirmación se dispone del fichero **reciennacidos.sf3**, que tiene el peso (gramos) de una muestra de recién nacidos en el hospital de San Pedro de Alcántara de Cáceres en la Semana Santa del año 2002. A la vista de los datos ¿Tiene razón el pediatra? Justifica y razona adecuadamente (de forma breve) tanto la metodología que vas a utilizar como las conclusiones.

Tanto el peso de los niños como de las niñas siguen distribuciones que pueden asumirse normales. Los histogramas son bastante normales y el pvalor de los test de chi-cuadrado son muy elevados. Asumimos normalidad de las poblaciones. Podemos así realizar comparación de medias y varianzas sin problemas.

Las distribuciones estimadas son: Peso niños: $X_1 \sim N(3162.07; 431.05^2)$. Peso niñas $X_2 \sim N(3163.09; 440.46^2)$. Para comparar las medias hacemos el contraste $H_0: \mu_1 = \mu_2; H_1: \mu_1 \neq \mu_2$. Para realizar ese contraste, hay que comparar primero las varianzas. Las cuasidesviaciones típicas son 431.05 y 440.46. La diferencia es pequeña (no es la mayor el doble de la menor) por lo que podemos hacer la comparación de medias asumiendo igualdad de varianzas. En cualquier caso, como tenemos normalidad, podemos hacer el contraste de igualdad de varianzas $H_0: \sigma_1^2 = \sigma_2^2; H_1: \sigma_1^2 \neq \sigma_2^2$ (NOTA: este contraste sólo es parte del temario de Ing. Informática, no de ITIG), que da un pvalor de $0.87 > 0.05$. Por tanto podemos asumir igualdad de varianzas poblacionales.

El contraste de igualdad de medias, asumiendo igualdad de varianzas, da un pvalor de $0.98 > 0.05$, por lo que podemos considerar que las medias son también iguales.

Por tanto las poblaciones del peso de niños y niñas son iguales con un nivel de significación del 5%