ESTADISTICA-Examen de prácticas. Junio 2009

Nombre:...

Titulación:..

Grupo en el que estás matriculado:...

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

1. Una máquina produce por término medio un 8% de artículos defectuosos, apareciendo estos de forma fortuita pero a un ritmo medio estable. Se tiene un lote de 100 artículos. Calcula la probabilidad de que haya más de 7 artículos defectuosos

X=número de artículos defectuosos en el lote. $X \sim B(n=100; p=0.08)$

Por tanto $P(X>7)=0.55289$

2. Un servidor que alberga una página web recibe por término medio 10 accesos por minuto, pudiéndose considerar que los accesos son sucesos independientes. ¿Cuál es la probabilidad de que no acceda nadie durante 2 minutos?

X=número de accesos/minuto. $X \sim \text{Poi}(\lambda=10)$

Y=número de accesos/2 minutos. $Y \sim \text{Poi}(\lambda=20)$. $P(Y=0)=2.06 \in \times 10^{-9}$

El fichero IndiceMC.sf3 tiene el índice de masa corporal (kg/m²) de una muestra de estudiantes universitarios (Sexo=1: chicos Sexo=0: chicas). Se pide:

3. Ajusta una distribución normal a la variable 'Índice de Masa Corporal' utilizando una transformación de los datos si así se consiguiese mejorar el ajuste a la normal. ¿Qué resultados se obtienen? Justifica la respuesta

La variable es unimodal con asimetría positiva (Skewness=0.886), por lo que podemos mejorar su simetría, y por tanto su normalidad, con transformaciones del tipo X^c con c<1.

Sin transformaciones vemos que el p-valor de la chi-cuadrado es p=0.10 que podría ser suficiente. Sin embargo este ajuste se puede mejorar con una transformación. Por ejemplo con c=-1.6 tenemos un coeficiente de asimetría de sólo -0.01, y un p-valor de 0.5. El histograma de la variable transformada muestra además un comportamiento muy parecido a una normal.
Por tanto, usaremos el siguiente modelo

$$Y = X^{-1.6} \sim N(0.00725; 0.00117^2)$$

4. Utilizando dicho ajuste, haz una estimación de la proporción de estudiantes universitarios que tienen sobrepeso (índice de masa corporal mayor que 25) . Justifica la respuesta

$$P(X>25)=P(X^{-1.6}<25^{-1.6})$$
$$=P(Y<0.0058)=0.10$$

donde hemos usado que $X^{-1.6}=1/X^{1.6}$*, y que*

$$(X>25) \Rightarrow \left(\frac{1}{25}>\frac{1}{X}\right)$$

(Nota: en el caso de haber usado c>0, sale directamente que $P(X>25)=P(X^c>25^c)$ *)*

5. Demuestra, a partir de la muestra del fichero anterior, que la población de chicos universitarios tiene un índice de masa corporal medio superior al de las chicas. Justifica adecuadamente tu respuesta.

El interés en esta pregunta es hacer inferencia sobre la "población de chicos y chicas universitarios", y no en hacer una síntesis que sólo resuma las diferencias encontradas en los datos en las muestras. Es por tanto necesario realizar un análisis estadístico que permita formular conclusiones sobre la población con un determinado nivel de significación. Una herramienta estadística apropiada es el contraste de hipótesis, sobre la diferencia de medias poblacionales, aunque también se puede hacer un intervalo de confianza sobre la diferencia de medias poblacionales.

Llamemos μ_V *al índice de masa corporal (IMC) medio de los jóvenes varones, y* μ_M *al de las mujeres. Se quiere saber si* $\mu_M < \mu_V$ *por lo que el contraste a realizar es*

$$H_0 : \mu_M \geq \mu_V$$
$$H_1 : \mu_M < \mu_V$$

Se dispone de un total de 200 datos, 100 por cada sexo. En esta muestra se tiene que

$$\bar{x}_M = 20.89,$$
$$\bar{x}_V = 23.17.$$

Por tanto, en esta muestra los chicos tienen un IMC medio mayor que el de las chicas. Queremos saber si esa diferencia es o no significativa. Es decir, si es lo bastante grande, como para pensar que se debe a diferencias reales entre las poblaciones o se debe simplemente al azar del muestreo en poblaciones iguales. El número de datos es suficientemente grande para que no nos importe mucho la distribución del IMC en cada población. Con una muestra tan grande, podemos aplicar el teorema central del límite, y considerar que las medias muestrales se distribuyen como una normal, que es el fundamento del contrate de igualdad de medias que hemos visto en la asignatura.

Las desviaciones típicas del IMC para cada sexo son muy parecidas (1.9 y 2.3, respectivamente). Como la mayor no duplica a la menor, resulta más conveniente realizar el contraste de igualdad de medias asumiendo igualdad de varianzas poblacionales.

Este contraste proporciona un p-valor de 0 (<0.05) por lo que rechazamos de forma muy clara la hipótesis nula, incluso con niveles de significación inferiores al 5%. Los datos muestran entonces mucha evidencia a favor de que el IMC poblacional medio de los chicos es mayor que el de las chicas. Las diferencias encontradas en las muestras son, con mucha seguridad, debidas a una diferencia real y no al azar del muestreo.