

## ESTADÍSTICA- PRÁCTICAS Septiembre 2008

Nombre:.....

Titulación:.....

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

- No está permitido utilizar (ni acceder a) ningún tipo de documentación
- Los datos están en la web de la asignatura. Descarga los datos en el Escritorio del ordenador
- Cuando realices un contraste de hipótesis o de normalidad, escribe el p-valor que obtienes
- Cuando hagas un contraste, escribe la hipótesis nula y la alternativa que estás utilizando
- En cualquier inferencia, escribe qué suposiciones estás haciendo (normalidad, muestras grandes, igualdad de varianzas, etc) y en qué te basas para hacerlas.

1. El fichero **radiacestación.sf3** contiene datos de radiación solar de dos localizaciones alternativas en un conjunto de días. A mayor valor de las variables, mayor es la radiación solar. Compara de forma descriptiva la distribución de los datos de ambas localizaciones, señalando los rasgos más importantes que les distinguen.

*Utilizando las opciones de Compare/Two samples /two sample comparison podemos comparar los histogramas y las medidas características.*

*Puede verse que ambas distribuciones tienen histogramas bastante similares. Ambas tienen rangos parecidos (entre 100 y 3000), varianzas similares y asimetrías negativas. Los valores medios son también similares y en torno a 2000. La principal diferencia está en que Loc1 es bimodal, con modas en torno a 100 y 1500; mientras que Loc2 es unimodal en torno a 2500. Esta diferencia se ve reflejada en la kurtosis, que en la distribución bimodal es mucho menor.*

2. Las variables Radiac\_Loc1 y Radiac\_Loc2 del fichero **radiacestación.sf3** corresponden a mediciones tomadas los mismos días. Son por tanto datos emparejados. ¿Se podrían utilizar los datos de una de las localizaciones para predecir la radiación de la otra localización? Justifica adecuadamente tu respuesta

*Se podría hacer una regresión simple del tipo  $Radiac\_Loc1 = a + b * Radiac\_Loc2$ . Para que esta regresión tuviese sentido, los datos deberían mostrar fuerte relación lineal. Sin embargo, el gráfico XY muestra una nube muy dispersa. Esta falta de relación se confirma con el coeficiente de correlación, que es de sólo 0.07.*

*Por tanto, no se puede predecir la radiación de una zona a partir de la otra.*

3. Un sistema productivo produce artículos defectuosos en una proporción de 3 de cada 1000. Se tiene un lote de 200 artículos de dicho proceso productivo. ¿Cuál es la probabilidad de que tenga más de 1 artículo defectuoso? Justifica brevemente la respuesta.

$$X = \text{número de artículos defectuosos en el lote de 200. } B(200, 0.003)$$
$$P(X > 1) = 0.1217$$

4. Un servidor gestiona los accesos a una red. A partir de 20 accesos en un mismo minuto el servidor se satura ¿Qué proporción del tiempo estará ese servidor saturado si los usuarios acceden a esa red de forma independiente y con un ritmo medio de 15 accesos por minuto.

$$X = \text{número de accesos por minuto} = \text{Poi}(15)$$
$$P(X > 20) = 0.08$$

5. Continuamos con el fichero **radiacestación.sf3**. Ajusta una normal a la variable Radiac\_Loc2 utilizando alguna transformación con el fin de que el ajuste sea lo mejor posible. A partir de ese modelo transformado calcula la probabilidad de que la radiación de un día supere las 1000 unidades.

*Como la distribución es asimétrica negativa, hay que hacer transformaciones del tipo  $X^c$  con  $c > 1$  para buscar un mejor ajuste a la normal. El test de la chi-cuadrado proporciona los p-valores más altos con transformaciones en torno a  $c = 1.5 - 2$ . Por ejemplo con  $c = 1.75$  se tiene un histograma bastante simétrico y un pvalor = 0.726.*

$$\text{La probabilidad de que en un día haya más de 1000 unidades de radiación es}$$
$$P(X > 1000) = P(X^{1.75} > 17783) = 0.9666$$