Universidad de Oviedo

ESTADÍSTICA

Ejemplo de examen de prácticas - Abril 2013

APELLIDOS, NOMBRE	NOTA

1. **(1 punto)** Si la duración en años de un determinado modelo de pieza sigue una distribución Weibull con parámetro de forma $k = 2$ y parámetro de escala $\lambda = 3$, ¿cuánto vale la probabilidad de que una pieza elegida al azar dure más de 4 años? $0'1690$
¿y la probabilidad de que, elegidas ocho de estas piezas de forma independiente, al menos 3 de ellas duren más de 4 años? $0'1393$

Todos los ejercicios que se proponen a continuación se refieren a la base de datos utilizada habitualmente en clase, cuyo nombre es **acero.rda**.

2. **(1'5 puntos)** En la muestra obtenida:
¿Qué % de las horas hay entre 1 y 3 averías (ambos inclusive)? $17'09\,\%$
¿Qué % de las horas hay más de 2 averías? $14'53\,\%$
¿Qué % de las horas con averías ha estado encendido el sistema de detección de sobrecalentamiento? $42'9\,\%$

3. **(2'5 puntos)** ¿Puede admitirse como significativamente mayor, en promedio, la producción total en las horas en las que ha habido averías de la producción total en las horas en la que no ha habido averías? No
¿Por qué? (Justifica detalladamente todos los pasos dados)

   - Puesto que dada la naturaleza de los datos estamos trabajando con **muestras independientes**, debemos estudiar la normalidad de la variable producción total en cada uno de los dos grupos que se comparan (cuando hay y cuando no hay averías).
   - Si realizamos el **test de normalidad** para la producción cuando hay averías, se obtiene que el p-valor asociado a esta muestra para dicho test es $0'2406$ y cuando no hay averías se obtiene que el correspondiente p-valor es $0'9579$. En ambos casos no se rechaza la hipótesis nula H_0, con lo que se puede **admitir la normalidad** de la variable en ambos grupos. Esto hace que el test adecuado para comparar en promedio la producción con y sin averías sea el **test t para muestras independientes**.
   - Para poder rellenar la ventana asociada a dicho test, tenemos que determinar si suponemos que las varianzas son iguales o no. Puesto que se ha admitido la normalidad, el test adecuado para realizar este contraste es el **test F para dos varianzas** y las hipótesis contrastadas son:

H_0: la varianza de la producción total es igual cuando hay avería y cuando no ($\sigma^2_{No} = \sigma^2_{Si}$)

H_1: la varianza de la producción total es distinta cuando hay avería y cuando no ($\sigma^2_{No} \neq \sigma^2_{Si}$)

Como el p-valor asociado a dicho test es $0'9636$ no existen evidencias para rechazar H_0, con lo que asumiremos la suposición de que las varianzas pueden considerarse iguales.

- Si realizamos el **test t para muestras independientes**, es un test para las medias y, por tanto, las hipótesis a contrastar, según el enunciado, serían:

  H_0: la producción media total no es mayor cuando hay avería que cuando no ($\mu_{No} \geq \mu_{Si}$)

  H_1: la producción media total es mayor cuando hay avería que cuando no ($\mu_{No} < \mu_{Si}$)

  Realizado este test, se obtiene que el p-valor es $0'2392$, con lo que no se rechaza H_0, es decir, **no existen evidencias significativas de que la producción total sea mayor cuando hay averías que cuando no**.

4. **(1 punto)** Selecciona los datos muestrales correspondientes a las horas en las que la temperatura fue baja y hay menos de 3 averías, y responde a las siguientes cuestiones:
   ¿Cuál es el nuevo tamaño muestral? 32 horas
   ¿Cuál es el consumo medio muestral? $135'85$ megavatios-hora
   ¿Y la desviación típica muestral? $41'17$ megavatios-hora

5. **(2 puntos)** Analicemos ahora, **para todas las horas**, las variables "temperatura" y "linea":
   ¿Están relacionadas? Sí ¿Por qué? Al tratarse de dos variables cualitativas, analizamos su relación mediante el test chi-cuadrado de independencia. Aplicado dicho test a los datos de nuestra muestra, se obtiene el p-valor $1'505 \cdot 10^{-5}$, que supone rechazar H_0, es decir, que existen evidencias significativas de relación entre ambas variables (no hay la misma temperatura cuando se utiliza una línea que cuando se utiliza otra)
   ¿Qué % de las horas hay temperatura alta y se usa la línea A? $20'5\,\%$
   ¿Qué % de las horas que hay temperatura alta se usa la línea A? $52'2\,\%$

6. **(0'5 puntos)** Crea una nueva variable que mida la producción de galvanizado tipo 1 más la producción de galvanizado tipo 2. ¿Cuál es la media de dicha variable para las horas en las que se ha producido en la línea A? $1502'667$ toneladas

7. **(1'5 puntos)** Si queremos predecir la emisión de CO en función de la emisión de COV, la recta de regresión sería:

$$\text{emisión de CO} = -0'6904 + 8'1609 \cdot \text{ emisión COV}$$

   ¿Cuánto se estima que vale la variabilidad del porcentaje de emisión de CO explicado por la emisión de COV? $R^2 = 99'01\,\%$

El resultado del examen de prácticas supone $2'5$ puntos de la valoración final de la asignatura. Así, por ejemplo, una persona que saque en este examen un 8 sobre 10 puntos, tiene 2 puntos en la nota final de la asignatura. Los otros $7'5$ puntos restantes se podrán obtener en el examen final que se realizará el lunes 20 de Mayo a las 16:00.