1. Define detalladamente cada uno de los siguientes parámetros e indica para qué sirven:
   a. Rango
   b. Coeficiente de asimetría
   c. Coeficiente de curtosis
   d. Desviación típica
   e. Cuartiles

Nota: El escribir la fórmula no puntuará.

Solución:

Rango: *En una muestra, es la diferencia entre el mayor y el menor valor de los datos observados.*

Coeficiente de asimetría: *Parámetro que sirve para detectar si los datos estan agrupados de forma simétrica alrededor de un valor central (media o mediana) o no.*

Coeficiente de curtosis: *En función del valor que se obtiene, la muestra puede ser leptocúrtica, mesocúrtica o planicúrtica. Además sirve para detectar posibles datos anómalos*

Desviación típica: *Parámetro de dispersión que sirve para medir la variabilidad de los datos. Si la distribución es normal, prácticamente toda la muestra se encuentra repartida en un intervalo centrado en la media y cuya amplitud es seis veces la desviación típica.*

Cuartiles: *Són parámetros de posición, e indican el valor de la variable que deja por debajo un porcentaje de los datos de la muestra. En concreto tenemos el cuartil C1, que deja por debajo el 25% de los datos, el cuartil C2(o mediana), que deja por debajo el 50% de los datos, y el cuartil C3, que deja por debajo el 75% de los datos.*

2. La Consultora **ALFAGRAND S.A.** ha publicado un estudio donde considera que el gasto medio mensual de una familia media española está en función del número de hijos menores a su cargo. Dicha consultora estima que una familia, independientemente del número de hijos, gasta mensualmente 650 euros. Además, este gasto se incrementa en 64 euros por hijo menor a cargo de la familia. También

sabemos que de los datos de la muestra estudiada se desprende una dispersión del gasto medio de 400 euros y una dispersión de 5 hijos correspondientes al número de hijos menores a cargo de la familia.

Nota: considere a efectos de cálculo que el número de hijos es una variable contínua.

a. ¿Qué modelo estadístico ha utilizado la consultora en su análisis? Justifica razonadamente tu respuesta.

Solución: *Ha utilizado el modelo de regresión lineal simple, porqué toda familia tiene un gasto fijo de 650€, independientemente del número de hijos que tenga, y además, a este gasto se le debe sumar una cantidad de 64€ por hijo:*

*Gasto mensual = 64*nºhijos menores + 650*

b. Si la familia Martínez está formada por cinco miembros, tres de ellos menores, ¿qué cantidad debería ingresar mensualmente como mínimo la familia para cubrir el gasto medio esperado?

Solución:

*Gasto mensual = 64*3+650 = 842 €* *deberá ingresar como mínimo la familia para cubrir el gasto medio esperado*

c. ¿Qué % de la variabilidad que la consultora detecta en el gasto medio mensual no es capaz de explicar el modelo planteado por la misma?. A la vista de este resultado, ¿se podria considerar que la covarianza entre las variables del modelo es prácticamente nula?

Solución:

El % de la variabilidad que la consultora detecta en el gasto medio mensual y es explicada por el modelo planteado por la misma es el coeficiente de determinación: $R = r^2$ *, donde* r *es el coeficiente de correlación lineal*

Y como sabemos que $r = b \cdot \dfrac{S_{hijos\_menores}}{S_{gasto\_medio}}$ *tenemos que*

$$r = 64 \cdot \frac{5}{400} = 0.8 \longrightarrow R = r^2 = 0.8^2 = 0.64$$

Es decir, el modelo planteado explica el 64% de la variabilidad. Por tanto, el 36% de la variabilidad no es explicada por el modelo.

Como el coeficiente de correlación lineal vale 0.8, podemos decir que existe relación entre las variables gasto esperado y número de hijos, y por tanto la covarianza va a ser distinta de cero.

d. ¿Entre qué valores se situaría en el 95% de los casos el gasto esperado mensual para una cantidad de hijos igual a 3?. ¿Qué pasaría con la amplitud del intervalo si en lugar del 95% nos planteásemos un % inferior?

Solución:

El 95% de los valores del gasto esperado mensual para una cantidad de hijos igual a 3 seria

$I_{0.95}$*(gasto esperado mensual)=(gasto mensual para 3 hijos* $\pm$ *3* $\cdot S_{residuos}$*)*

Gasto mensual para 3 hijos = 64 ·3+650 = 842 €

$$S^2_{residuos} = S''_{gasto\_medio} \cdot (1 - r^2) = 400^2(1 - 0.64) = 57600 \longrightarrow S_{residuos} = \sqrt{57600} = 240$$

de donde,

$I_{0.95}$(gasto esperado mensual)=$\left(842 \pm 3 \cdot 240_s\right)$=$\left(362\ ,\ 1322\right)$

Si en lugar del 95%, nos planteásemos un % inferior, la amplitud del intervalo seria menor

3. En un cierto dispositivo electrónico utilizado en los rádares de control de tráfico el número de componentes que fallan a las 1000 horas se puede modelizar como una variable aleatoria discreta con función de probabilidad:

   $$P[X=k]= \alpha \cdot k \qquad k=[0,1,2,3,4]$$

   ¿Cuánto debe valer α?
   Solución:

   La suma de las probabilidades para todos los valores posibles de la variable debe ser igual a 1, de ahí que

   $$1=\sum_{k=o}^{4} P(X=k)=\alpha\cdot 0+\alpha\cdot 1+\alpha\cdot 2+\alpha\cdot 3+\alpha\cdot 4=\alpha+2\alpha+3\alpha+4\alpha=10\alpha \longrightarrow \alpha=0.1$$

4. Sabemos que el 15% de la población española tiene estudios superiores, estudios medios el 40%, estudios primarios el 35% y no tiene estudios el 10%. Los desempleados no se distribuyen proporcionalmente entre estas categorías, dado que de entre los de estudios superiores están sin trabajo el 10%, entre los de estudios medios el 35%, entre los estudios primarios el 18%, y entre los que no tienen estudios el 37%.
   a. ¿Qué probabilidad hay de que un individuo que está en paro tenga estudios superiores?
   Solución:

Definimos los sucesos:
S=Individuo con estudios superiores
M=Individuo con estudios medios
P=Individuo con estudios primarios
N=Individuo sin estudios
T=Individuo con trabajo

Sabemo que:

$P(S)=0.15$	$P(no\ T\mid S)=0.10$
$P(M)=0.40$	$P(no\ T\mid M)=0.35$
$P(P)=0.35$	$P(no\ T\mid P)=0.18$
$P(N)=0.10$	$P(no\ T\mid N)=0.37$

Con lo que por el teorema de Bayes

$$P(S\mid noT)=\frac{P(noT\mid S)\cdot P(S)}{P(noT)}=\frac{P(noT\mid S)\cdot P(S)}{P(noT\mid S)\cdot P(S)+P(noT\mid M)\cdot P(M)+P(noT\mid P)\cdot P(P)+P(noT\mid N)\cdot P(N)}=$$

$$=\frac{0.10\cdot 0.15}{0.10\cdot 0.15+0.35\cdot 0.40+0.18\cdot 0.35+0.37\cdot 0.10}=\frac{0.015}{0.255}=0.059$$

b. ¿Qué probabilidad hay de que si el individuo no tiene estudios, esté parado?
Solución:

$$P(\text{no } T|N)=0.37$$

c. ¿Qué probabilidad hay de que una persona tenga estudios primarios y esté trabajando?
Solución:

$$P(P \text{ y } T)=P(T|P)\cdot P(P)=0.82\cdot 0.35=0.287$$

d. ¿Qué probabilidad hay de que una persona que está trabajando tenga estudios superiores?
Solución:

$$P(S\,|\,T)=\frac{P(T\,|\,S)\cdot P(S)}{P(T)}=\frac{P(T\,|\,S)\cdot P(S)}{1-P(noT)}=\frac{0.90\cdot 0.15}{1-0.255}=0.181$$

Fuente: enunciados correspondientes a exámenes de diferentes años de la Universidad Politécnica de Valencia.