ESTADÍSTICA

Solución Examen Final Julio 2012

Lee esto antes de empezar

- La duración del examen es de 2h.
- No se puede salir del aula y volver a entrar, debiendo permanecer en ella durante la primera media hora.
- No se puede utilizar ni lápiz ni bolígrafo rojo.

P1

Una empresa del sector de la construcción está muy preocupada por la calidad de las vigas de hormigón que adquiere. El motivo de dicha preocupación es que el 20% de las vigas de sus proveedores tiene el hormigón con alguna enfermedad. Con el objetivo de seleccionar las vigas que van a emplear en sus edificaciones, han decidido adquirir un test para comprobar el estado del hormigón. Saben que este test da positivo (es decir, nos alerta sobre una posible enfermedad del hormigón) en el 95% de las vigas en mal estado y en el 10% de las vigas en buen estado.
Si el hormigón de una viga da positivo en el test ¿con qué probabilidad se encuentra en mal estado? (1.5 puntos)

Solución:

Denotamos por

M≡ viga con hormigón en mal estado
B≡ viga con hormigón en buen estado
P≡ el test da positivo
N≡ el test da negativo

Sabemos que:

$$P(M) = 0.2 \qquad P(B) = 0.8$$

$$P(P/M) = 0.95 \qquad P(P/B) = 0.1$$

$$P(\text{viga en mal estado}/\text{test da positivo}) = P(M/P) = \frac{P(M \cap P)}{P(P)}$$

$$P(P) = P\big((P \cap M) \cup (P \cap B)\big) = P(P \cap M) + P(P \cap B) = P(P/M)P(M) + P(P/B)P(B) =$$

$$= 0.95 \cdot 0.2 + 0.1 \cdot 0.8$$

$$P(M/P) = \frac{0.95 \cdot 0.2}{0.95 \cdot 0.2 + 0.1 \cdot 0.8} = \frac{19}{27} = 0.703$$

P2

El tiempo, en horas, de inspección de una solicitud en Hacienda es una v.a. con la siguiente función de densidad

$$f(x) = \begin{cases} kx & 0 < x < 3 \\ 0 & resto \end{cases}$$

Se pide:

a) Halle el valor de k para que sea densidad.
b) Calcule la media y la varianza de la v.a. anterior.
c) Determine la probabilidad de que el tiempo de inspección de una solicitud supere las dos horas.
d) Si se inspeccionan 50 solicitudes de forma consecutiva y se supone independencia y que no se pierde tiempo entre las inspecciones, ¿cuál es la probabilidad de que el tiempo total de las 50 solicitudes supere las 95 horas?

(3.5 puntos)

Solución:

a) $1=\int_{\mathbb{R}} f=\int_0^3 kxdx=k\left.\frac{x^2}{2}\right|_0^3=\frac{9}{2}k \quad \Rightarrow \quad k=\frac{2}{9}$

b)

$$f(x)=\begin{cases}\frac{2}{9}x & 0<x<3\\ 0 & resto\end{cases}$$

$\mathrm{E}[X]=\int_{\mathbb{R}} xf(x)dx=\int_0^3 x\frac{2}{9}xdx=\left.\frac{2}{9}\frac{x^3}{3}\right|_0^3=2$ $\qquad$ $\mathrm{E}\left[X^2\right]=\int_{\mathbb{R}} x^2 f(x)dx=\int_0^3 x^2\frac{2}{9}xdx=\left.\frac{2}{9}\frac{x^4}{4}\right|_0^3=\frac{9}{2}$

$\mathrm{V}[X]=\mathrm{E}\left[X^2\right]-\left(\mathrm{E}[X]\right)^2=\frac{9}{2}-2^2=\frac{1}{2}$

c) $P(X>2)=\int_2^{+\infty} f(x)dx=\int_2^3\frac{2}{9}xdx=\left.\frac{x^2}{9}\right|_2^3=\frac{5}{9}=0.555$

d)

$X_i\equiv$ v.a. tiempo de la solicitud i-ésima $\qquad i=1,....,50$

$T\equiv$ v.a. tiempo total $\qquad T=X_1+\cdots+X_{50}$

¿P(tiempo total supere las 95 horas)=$P(T>95)$?

$X_1,\cdots,X_{50}$ v.a.i.i.d. con $\mathrm{E}[X_i]=2 \qquad \mathrm{V}[X_i]=\frac{1}{2} \qquad \forall i=1,....,50$

Por el TCL, T se distribuye aproximadamente como una Normal: $T\approx N\left(50\cdot 2,\sqrt{50\cdot\frac{1}{2}}\right)$

$T\approx N(100,5)$

$P(T>95)\approx P(N(100,5)>95)=P\left(N(0,1)>\frac{95-100}{5}\right)=P(N(0,1)>-1)=P(N(0,1)<1)=0.8413$

P3

Cada sucursal del *Banco Paraiso* recibe cheques de manera independiente y con una tasa media constante de 10 cheques a la hora. Se sabe que la probabilidad de que un cheque no tenga fondos es del 0.01.

a) Para una sucursal determinada ¿cuál es la probabilidad de que reciba exactamente tres cheques en su primera media hora de funcionamiento?

b) La entidad bancaria dispone de 12 sucursales en la ciudad. Si en una jornada laboral cada una de esas 12 sucursales ha recibido 20 cheques, ¿cuál es la probabilidad de que al menos 2 de las sucursales hayan recibido algún cheque sin fondos?

(2.5 puntos)

Solución:

a)

$X\equiv$ v.a. número de cheques recibidos durante media hora

Como los cheques se reciben independientemente y con una tasa media constante de 10 cheques a la hora, el numero medio de cheques recibidos en $1/2$ hora será 5 y la v.a. X seguirá una distribucion de Poisson de parámetro $\lambda=5$: $X\equiv P(5)$

$P(X=3)=\frac{e^{-5}5^3}{3!}=P(X\le 3)-P(X\le 2)\underset{tablas}{=}0.265-0.125=0.14$

b) Denotamos por: E≡ cheque sin fondos F≡ cheque con fondos

Sabemos $P(E)=0.01$ $P(F)=0.99$

$$P\left(\text{algún cheque sin fondos entre los 20}\right)=1-P(\text{todos los cheques con fondos}) \underset{independientes}{=} 1-\left(P(F)\right)^{20}=$$

$$=1-(0.99)^{20}=0.18$$

Sea $Y \equiv$ v.a. número de sucursales, entre las 12, que reciben algún cheque sin fondos

$Y \equiv \mathrm{B}(12, 0.18)$

$$P\left(\text{al menos 2 de las sucursales hayan recibido algún cheque sin fondos}\right)=P(Y\geq 2)=1-P(Y<2)=$$

$$=1-P(Y\leq 1)=1-\left(P(Y=0)+P(Y=1)\right)=1-\left((0.82)^{12}+12\cdot(0.18)\cdot(0.82)^{11}\right)=1-0.335=0.665$$

P4

Se ha registrado el caudal horario (en m^3/seg) de dos crecidas del río Duero a su paso por la localidad de Soria. Se han observado $n_1=43$ datos para la primera crecida y $n_2=48$ datos para la segunda crecida, con medias muestrales de $\bar{x}_1=83.2$ y $\bar{x}_2=71.1$ y con desviaciones típicas muestrales de $s_1=4.4$ y $s_2=2.9$ .Se supone que los datos se han tomado de manera independiente y que provienen de poblaciones normales. A la vista de esta información:

a) Obtenga un intervalo de confianza del 90% para el cociente de varianzas de los caudales horarios de las crecidas.

DATOS: $P\left\{F_{42,47}>0.604\right\}=0.95$ $P\left\{F_{47,42}>0.609\right\}=0.95$

b) En virtud del resultado del apartado anterior ¿podemos admitir la hipótesis de varianzas iguales? ¿por qué?

c) Obtenga un intervalo de confianza para la diferencia de caudales horarios medios de ambas crecidas con nivel de confianza del 95%.

d) En virtud del resultado del apartado anterior ¿podemos afirmar que el caudal medio de la primera crecida del río supera al valor medio de la segunda crecida? ¿por qué?

(2.5 puntos)

Solución:

$X \equiv$ v.a. caudal horario en la primera crecida $X \equiv \mathrm{N}(\mu_1, \sigma_1)$

m.a.s. $n=43$: $\bar{x}=83.2$ $s_1=4.4$

$Y \equiv$ v.a. caudal horario en la segunda crecida $Y \equiv \mathrm{N}(\mu_2, \sigma_2)$

m.a.s. $m=48$: $\bar{y}=71.1$ $s_2=2.9$

a) Intervalo de confianza del 90% para el cociente de varianzas poblacionales:

$$\mathrm{I.C.}_{0.9}=\left(\frac{S_1^2}{S_2^2}\cdot\frac{1}{f_{0.05;42,47}}, \frac{S_1^2}{S_2^2}\cdot f_{0.05;47,42}\right)$$

Sabemos que $P\left\{F_{42,47}>0.604\right\}=0.95 \Leftrightarrow f_{0.95;42,47}=0.604$

$$f_{0.05;47,42}=\frac{1}{f_{0.95;42,47}} \Leftrightarrow f_{0.05;47,42}=\frac{1}{0.604}=1.655$$

Sabemos que $P\left\{F_{47,42}>0.609\right\}=0.95 \Leftrightarrow f_{0.95;47,42}=0.609$

$$f_{0.05;42,47}=\frac{1}{f_{0.95;47,42}} \Leftrightarrow \frac{1}{f_{0.05;42,47}}=f_{0.95;47,42}=0.609$$

La realización muestral del intervalo de confianza es

$$\text{I.C.}_{0.9} = \left(\left(\frac{4.4}{2.9}\right)^2 \cdot 0.609, \left(\frac{4.4}{2.9}\right)^2 \cdot 1.655 \right) = (1.401, 3.809)$$

b) Con una seguridad del 90% sabemos que $\frac{\sigma_1^2}{\sigma_2^2}$ está entre 1.401 y 3.809. Como el uno no pertenece al intervalo, con una seguridad del 90%, podemos concluir que el cociente de varianzas es distinto de uno y por tanto las varianzas no son iguales.

c) Por el apartado anterior sabemos que las <u>varianzas poblacionales son distintas</u>, por lo que el intervalo de confianza del 95% para la diferencia de medias poblacionales viene dado por:

$$\text{I.C.}_{0.95} = \left((\bar{X} - \bar{Y}) - z_{0.025}\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}, (\bar{X} - \bar{Y}) + z_{0.025}\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}} \right)$$

$$P(\text{N}(0,1) > 1.96) = 0.025$$

La realización muetral del intervalo de confianza es

$$\text{I.C.}_{0.95} = \left((83.2 - 71.1) \mp 1.96\sqrt{\frac{(4.4)^2}{43} + \frac{(2.9)^2}{48}} \right) = (12.1 \mp 1.55) = (10.54, 13.65)$$

d) Con una seguridad del 95% podemos afirmar que $\mu_1 - \mu_2$ está entre 10.54 y 13.65. Como los extremos del intervalo son ambos positivos, podemos concluir que $\mu_1 - \mu_2 > 0 \Leftrightarrow \mu_1 > \mu_2$ con una confianza del 95% y por tanto afirmar que el caudal medio en la primera crecida supera al caudal medio de la segunda.