

OPOSICIONES AL CDEE. CONVOCATORIA 2018. EJERCICIO 1

ESTADÍSTICA TEÓRICA

1. Se consideran dos sucesos, A y B , asociados a un experimento aleatorio $P(A)=0,7$, $P(B)=0,6$ y $P(\bar{A} \cup \bar{B}) = 0,58$ siendo \bar{A} y \bar{B} los sucesos complementarios de A y B respectivamente. Si el suceso M está incluido en A , ¿calcule el valor de $P(\bar{M}/\bar{A})$?

2. Dada la función

$$f(x) = \begin{cases} 2x + k & \text{si } 1 \leq x \leq 2 \\ 0 & \text{resto} \end{cases}$$

Determine el valor de k para que $f(x)$ sea función de densidad.

3. ¿Es la misma variable aleatoria la que resulta de sumar n variables aleatorias continuas independientes e idénticamente distribuidas y la que resulta de tomar una de ellas y multiplicarla por n ? Razone la respuesta.

4. Sea X una variable aleatoria con distribución de Poisson tal que $P(X=2)$ es la quinta parte de $P(X=1)$. Determine la función de probabilidad de X .

5. Un punto aleatorio X sigue una distribución uniforme en el intervalo $[0,1]$ y otro punto aleatorio Y tiene distribución uniforme en el intervalo $[2,3]$ independiente de X . Calcule la esperanza matemática y la varianza de la variable distancia entre X e Y definida por $d=Y - X$.

6. Sea (X,Y) una normal bivalente y se definen las nuevas variables aleatorias $Z=X+Y$ y $W=X-Y$. Determine la condición necesaria y suficiente para que Z y W sean variables aleatorias independientes.

7. Se consideran dos variables aleatorias independientes X e Y que siguen una distribución $N(0,1)$ y una distribución Chi-cuadrado de Pearson con 4 grados de libertad respectivamente. Se define una nueva variable aleatoria $Z = 2X/\sqrt{Y}$. ¿Cuál es la distribución de la variable aleatoria Z ? Razone la respuesta.

8. Sea (X_1, X_2, \dots, X_n) una muestra aleatoria simple de una variable aleatoria X con media μ y varianza σ^2 . Calcule el error cuadrático medio para los siguientes estimadores de la media μ :

$$\hat{\mu}_1 = X_1 \quad \text{y} \quad \hat{\mu}_2 = \frac{3X_1 - 2X_2 + X_3}{6}$$

9. Una impresora puede fallar 0, 1 ó 2 veces al día y se sabe que la probabilidad de fallar 1 vez es el doble que la probabilidad de fallar 2 veces. En 30 días la máquina ha fallado 1 vez dos días, 2 veces tres días y no ha fallado los 25 días restantes. Utilizando el método de los momentos, estime la probabilidad de que falle dos veces en un día.

10. Sea (X_1, X_2, \dots, X_n) una muestra aleatoria simple de una característica X que se distribuye según una normal de media μ y desviación típica σ . Razone si $P(X < \mu + 1)$ es mayor, menor o igual que $P(\bar{X} < \mu + 1)$.

11. Una muestra de una población normal ha dado los siguientes resultados:

$$n = 8, \quad \sum x_i = 16 \quad \text{y} \quad \sum x_i^2 = 46$$

Calcule de forma razonada un intervalo de confianza al 95% para la varianza de la población normal. (Nota: para un valor $\alpha = 0,025$ y $\alpha = 0,975$ el valor crítico necesario de la distribución del estadístico es 1,69 y 16,0 respectivamente).

12. La duración de un componente electrónico en horas se distribuye según una normal de media μ y desviación típica igual a 300. Se quiere contrastar la hipótesis de $\mu = 310$ con los resultados de una muestra de 100 componentes en la que se ha obtenido una duración media de 250 horas. Indique, razonando la respuesta, cuál es el valor crítico p del contraste.

13. En un sondeo previo a unas elecciones se preguntó a 500 personas cuál era su intención de voto. Los datos obtenidos se clasificaron en función de 3 grupos de edad y 4 partidos políticos. Se pide:

- ¿Qué contraste se debe plantear para estudiar la independencia entre la intención de voto y la edad de las personas?
- ¿Qué estadístico se utiliza para realizar el contraste y cuál es su distribución?
- Si el valor del estadístico del contraste es 70,86 y el valor crítico para el nivel de significación del 10% es 10,64, ¿qué se puede concluir?

14. Se prueba un método de selección de muestras en una población artificial, en la que se ha definido una característica cuyo total poblacional es igual a 10, y se obtienen todas las muestras posibles. Los resultados obtenidos son:

Muestra	Total muestral	Probabilidad muestra
1	10	3/8
2	7	1/8
3	14	4/8

A partir de los resultados obtenidos ¿es el total muestral un estimador insesgado del total poblacional? Justifique su respuesta.

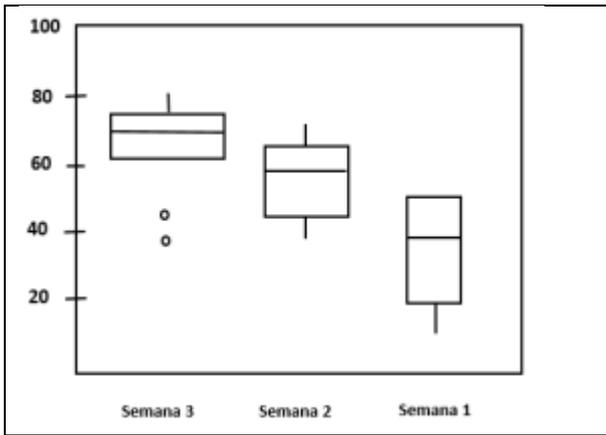
15. En una población dividida en 4 estratos de 100 viviendas cada uno, se desea estimar la proporción P de viviendas con más de un coche. Para ello se selecciona, sin reemplazamiento y con probabilidades iguales, una muestra independiente de 10 viviendas en cada estrato. Se obtienen las proporciones muestrales 0,1 ; 0,4 ; 0,4 ; 0,1 de viviendas con más de un coche en los estratos. Se pide un estimador insesgado de la proporción P y una estimación de su varianza.

ESTADÍSTICA DESCRIPTIVA Y PROCESO ESTADÍSTICO

1. Efectuados dos ensayos de regresión lineal de Y sobre X , se obtienen los siguientes resultados:

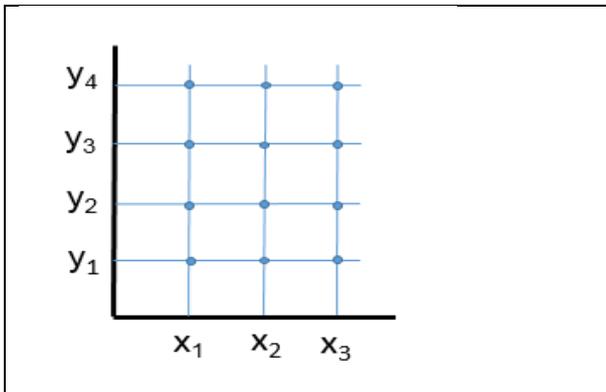
- La varianza residual del primer ensayo es el doble de la obtenida en el segundo.
- La varianza de Y en el primer ensayo es 2,5 veces la obtenida en el segundo.

¿Qué recta de regresión muestra un mejor grado de ajuste?



2. Un alumno aprueba el examen del carnet de conducir si obtiene al menos 60 puntos en un test. Se prepara durante tres semanas (1, 2 y 3), realizando 300 test en cada una de ellas. Los resultados obtenidos se presentan en el gráfico adjunto. Valore la evolución y sus posibilidades de aprobar.

3. La facturación media de las empresas agrícolas es de 13.500 euros y de 18.000 euros para las empresas no agrícolas. De entre estas últimas, el 40% son industriales y el 60% de servicios. Calcule la distribución porcentual de las empresas entre los tres sectores (agrícolas, industriales y de servicios) sabiendo que la facturación media del conjunto es de 16.500 euros. Si la facturación media de las empresas industriales es de 17.000 euros, calcule la facturación media de las empresas de servicios.



4. Se toman datos relativos a las variables X e Y de un conjunto de 12 individuos, resultando la nube de puntos que se representa en el gráfico adjunto.

Calcule formalmente la $Cov(X,Y)$.

- Se pretende analizar si una fuente administrativa es adecuada para producir información estadística. Para ello, se estudia una muestra representativa, obteniéndose diversos indicadores de calidad (cobertura, duplicados, etc.) que permitan tomar una decisión sobre el uso de la citada fuente. Justifique razonadamente en qué fase y subproceso del estándar GSBPM se clasifican todas estas tareas.
- Describa los fundamentos de la depuración selectiva y sus beneficios frente a otras técnicas de depuración estadística.
- El coeficiente de correlación lineal entre X e Y es 0,5. ¿Qué porcentaje de la variación total no es explicado por la recta de regresión?
- Bajo el supuesto de que se disponga de cifras de población y de renta por países a nivel mundial, comente qué medidas estadísticas se han utilizado y cómo se ha llegado a la siguiente afirmación: “El 0,7% de la población mundial (X millones) posee el 45,2% de la riqueza global”. A partir de esta información, ¿puede decir algo acerca del valor del índice de Gini asociado a esta distribución?

9. Explique la relación que existe entre la media aritmética, la geométrica y la armónica. Demuéstrelo para una muestra de tamaño 2.

10. Se presupone relación lineal sobre un conjunto de observaciones bidimensionales, por lo que se calcula el par de rectas de regresión asociadas:

$$Y/X: Y=108,9 - 0,125 X$$

$$X/Y: X=320,6 - 7,78 Y$$

A partir de estos resultados calcule el tipo y el grado de correlación lineal que respalde o refute la suposición inicial.

11. Comente cómo puede afectar la festividad de Semana Santa al estudio de la evolución temporal de una serie económica. ¿Qué componente recoge este comportamiento? ¿Cómo se incorpora este evento en la modelización?

12. Especifique cómo se utiliza el Índice de Laspeyres encadenado en el cálculo del Índice de Precios del Consumo (IPC) del INE.

13. En una distribución de frecuencias agrupadas de una variable estadística cuantitativa unidimensional, defina el intervalo modal y el intervalo mediano, explicando claramente cada uno de los símbolos empleados en las definiciones empleadas.

14. Explique la diferencia entre el coeficiente de correlación simple y el coeficiente de correlación parcial en una distribución n-dimensional.

15. ¿Qué tipo de comportamiento de una serie de datos observada a lo largo del tiempo recoge la componente residual? ¿Cómo se calcula?