Cuestiones resueltas

Primer examen

2024

Academia DEIMOS
Oposiciones: a) Secundaria.
b) Diplomados en
Estadística del Estado.
☎ 669 31 64 06
MADRID
www.academiadeimos.es
http://academiadeimos.blogspot.com.es
academia@academiadeimos.es
editorial@academiadeimos.es

OPOSICIONES AL CDEE. CONVOCATORIA 2023.
EJERCICIO 1.

ESTADÍSTICA TEÓRICA

1.- Una urna contiene dos bolas blancas y tres bolas negras. Cuatro jugadores A, B, C, D extraen por orden una bola de la urna sin devolverla. El primer jugador que extrae una bola blanca recibe 10 €. Calcular la esperanza de ingresos de cada jugador.

Definamos:

- B_i :="*la i-ésima bola extraída es blanca*" para $i = 1, 2, 3, 4$
- B_i :="*la i-ésima bola extraída es blanca*" para $i = 1, 2, 3, 4$
- X_i :="*ganancia del jugador i*"

Dado que

$$P(B_1) = \frac{2}{5} \quad , \quad P(N_1 \cap B_2) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10} \quad , \quad P(N_1 \cap N_2 \cap B_3) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} = \frac{1}{5} \quad , \quad P(N_1 \cap N_2 \cap N_3 \cap B_4) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} \cdot \frac{2}{2} = \frac{1}{10}$$

Entonces:

- La ganancia esperada del jugador A será:
$$E[X_1] = 10 \cdot P(B_1) = 10 \cdot \frac{2}{5} = 4\,€$$
- La ganancia esperada del jugador B será:
$$E[X_2] = 10 \cdot P(B_2) = 10 \cdot \frac{3}{10} = 3\,€$$
- La ganancia esperada del jugador C será:
$$E[X_3] = 10 \cdot P(B_3) = 10 \cdot \frac{1}{5} = 2\,€$$
- La ganancia esperada del jugador D será:
$$E[X_2] = 10 \cdot P(B_4) = 10 \cdot \frac{1}{10} = 1\,€$$

2.- Sean cuatro variables independientes X_1, X_2, X_3 y X_4, todas con distribución de Poisson de media $\frac{1}{2}$. Sea la variable

$$T = \frac{X_1 + X_2 + X_3 + X_4}{4}$$

Calcular la probabilidad de que $T \leq 1$

Por la Propiedad Reproductiva de la Poisson, la variable $S_4 = X_1 + X_2 + X_3 + X_4$ seguirá una distribución de Poisson de media $4 \cdot \frac{1}{2} = 2$, por lo que

$$P\left(R \leq 1\right) = P\left(S_4 \leq 4\right) = \sum_{i=0}^{4} P(S_4 = i) = \sum_{i=0}^{4} e^{-2}\ \frac{2^i}{i!} = 7e^{-2}$$

3.- Dada la función

$$\varphi(t) = \left(k + pe^{it}\right)^4$$

a) Calcule el valor de k para que $\varphi(t)$ sea función característica.

b) Calcule la esperanza de esa variable aleatoria.

a) Dado que $\varphi(0) = 1$, entonces

$$(k + pe^0)^4 = 1 \quad \Rightarrow \quad k + p = 1 \quad \Rightarrow \quad k = 1 - p = q$$

por lo que

$$\varphi(t) = \left(q + pe^{it}\right)^4$$

b) La derivada de esta función es

$$\varphi'(t) = 4pie^{it}\left(q + pe^{it}\right)^3$$

por lo que

$$E[iX] = \varphi'(0) = 4pi \quad \Rightarrow \quad E[X] = 4p$$

Observación: Esta función característica corresponde a la de una distribución Binomial de parámetros $n = 4$ y p.

4.- Un investigador está estudiando el efecto de un nuevo medicamento en la presión arterial. En un estudio con 100 participantes, seleccionados por muestreo aleatorio simple, la presión arterial media después del tratamiento fue de 120 mm Hg. El investigador quiere saber si el medicamento reduce la presión arterial. Si la media poblacional es de 125 mm Hg y la desviación típica poblacional de 15 mm Hg.

a) Plantee las hipótesis nula y alternativa para este estudio.

b) ¿Qué tipo de hipótesis ha usado (simple, unilateral o bilateral) para la hipótesis nula y por qué?

c) Proponga una región crítica para realizar el contraste de hipótesis.

d) Si el nivel de significación α es de 0,05 , ¿rechazaría la hipótesis nula?

Nota: $z_{0,05} = 1,645\,(\alpha = P\,(Z > z_\alpha)$ siendo Z la distribución normal estándar)

a) Las hipótesis nula y alternativa serán:

$$H_0:\ \mu = 15 \qquad , \qquad H_1: \mu < 15$$

b) La hipótesis nula es simple, dado que es el dato poblacional que tenemos, mientras que la hipótesis alternativa es compuesta y el contraste es unilateral. Dado que se intenta probar que el medicamento reduce la presión arterial, la hipótesis alternativa (la que se quiere probar) será "afirmar que la media poblacional es inferior a 15".

c) Como se trata de una población normal con varianza poblacional conocida, el estadístico contraste será

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}}$$

cuya distribución será una $\mathcal{N}(0,1)$. La región crítica para este contraste será

$$Z_0 < -z_\alpha$$

d) Como el valor del estadístico contraste, para la muestra obtenida, es:

$$z_0 = \frac{120 - 125}{15/\sqrt{100}} = -3,33$$

Dado que $z_0 < -z_\alpha$, la decisión será rechazar la hipótesis nula.

5.- Se precisa estimar la estatura media de los alumnos de una clase. Para ello se toma una muestra seleccionando a los 5 primeros alumnos que salgan de clase y se define la función

$$f\left(x_1, x_2, \ldots, x_5\right) = \frac{1}{5}\sum_{i=1}^{5} x_i$$

donde cada x_i $(i = 1, \ldots, 5)$ es la estatura del alumno i-ésimo de la muestra. ¿Es f un estimador insesgado de la estatura media? ¿Por qué?

Sea μ la estatura media de los alumnos dela clase. Dado que el estimador elegido es la media muestral de los 5 primeros alumnos que salen de clase, es decir

$$E[\bar{X}] = \frac{1}{5}\sum_{i=1}^{5} X_i = \frac{1}{5} 5\mu = \mu$$

Dado que la población a estudio es finita, esta cuestión también puede considerarse una cuestión de muestreo en poblaciones finitas:

Si asumimos que el orden de salida de clase de los alumnos no tiene ninguna relación su altura y que podemos por tanto tratar a esos 5 alumnos como una muestra aleatoria simple sin reemplazamiento, entonces f es un estimador insesgado de la estatura media, pues en el muestreo sin reemplazamiento de poblaciones finitas la media muestral es un estimador insesgado para la media poblacional.

Para justificarlo, definimos e_i, que es una variable aleatoria dicotómica que determina si el alumno i-ésimo es seleccionado o no en la muestra; y N como la cantidad total de alumnos en la clase.

De esta manera:

$$f(x_1, x_2, \ldots, x_5) = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{1}{5}\sum_{i=1}^{N} x_i e_i \Rightarrow E\left(f(x_1, x_2, \ldots, x_5)\right) = \frac{1}{5}\sum_{i=1}^{N} x_i E(e_i).$$

Dicha esperanza coincide con la media poblacional si los e_i tienen esperanza $\frac{5}{N}$, que es lo que ocurre en el muestreo aleatorio simple sin reposición.

26.- Sea $X_1, \ldots, X_n$ una muestra aleatoria simple de una $\mathcal{N}(\mu, \sigma)$. ¿Cuál es la distribución de $Y = \dfrac{\bar{X} - \mu}{S}\sqrt{n-1}$? ¿Por qué?

Nota: $S^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$

Dado que

$$\frac{\bar{X} - \mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0, 1)$$

Y que por el el Teorema de Fisher sabemos que $\bar{X}$ y S^2 son independientes y que, además:

$$n\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Entonces, basta recordar como se construye una variable con distribución $t-$Student para concluir que:

$$Y = \frac{\dfrac{\bar{X} - \mu}{\sigma}\sqrt{n}}{\dfrac{\sqrt{n\dfrac{S^2}{\sigma^2}}}{\sqrt{n-1}}} = \frac{\bar{X} - \mu}{S}\sqrt{n-1} \sim t_{n-1}$$

27.- Se sabe que la función generatriz de los momentos de una variable aleatoria X discreta viene dada por

$$g(t) = \frac{1-\alpha}{1-\alpha e^t} \quad \alpha \in (0,1)$$

Determinar cuál es la distribución de la variable aleatoria X.

Consideremos una v.a. con distribución Geométrica de parámetro $p \in (0,1)$. Su función generatriz de momentos es:

$$g_X(t) = E\left[e^{tX}\right] = \sum_{n=0}^{\infty} e^{nt} \cdot P(X=n) = \sum_{n=0}^{\infty} e^{nt} q^n p = p \sum_{n=0}^{\infty} (qe^t)^n = \frac{p}{1-qe^t} = \frac{1-q}{1-qe^t}$$

Haciendo $q = \alpha$ se obtiene la función generatriz de momentos que aparece en el enunciado, por lo que la distribución de la variable será una Geométrica de parámetro $p = 1-\alpha$.

6 La distribución de acciones de una empresa entre sus 15 accionistas es la siguiente:

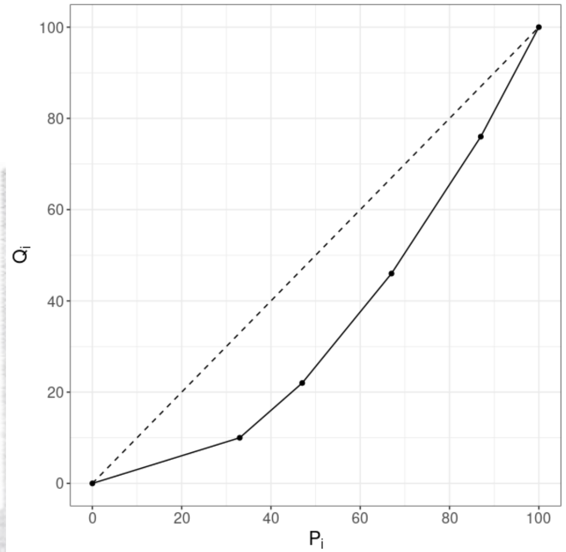
N acciones (x_i)	n_i	N_i	$x_i n_i$	$\sum x_i n_i$	$p_i = \frac{N_i}{N}$	$q_i = \frac{\sum_1^i x_j n_j}{\sum_k x_k n_k}$	$p_i - q_i$
2	5	5	10	10	33 %	10 %	23
6	2	7	12	22	47 %	22 %	25
8	3	10	24	46	67 %	46 %	21
10	3	13	30	76	87 %	76 %	11
12	2	15	24	100	100 %	100 %	0

Obtenga el índice de concentración de Gini y construya la curva de Lorenz correspondiente. Interprete los resultados.

El índice de concentración de Gini viene dado por:

$$IG = \frac{\sum_{i=1}^{n-1} p_i - q_i}{\sum_{i=1}^{n-1} p_i} = \frac{23 + 25 + 21 + 11}{33 + 47 + 67 + 87} = \frac{80}{234} = \frac{40}{117}.$$

Además, representando los puntos (p_i, q_i) obtenemos la curva de Lorenz:



Viendo el valor del índice de Gini que está cerca de un tercio y el gráfico la distribución de acciones no está muy concentrada. Hablaríamos de un caso de reparto o igualdad total si tuviéramos que el índice de Gini valiera 0, en cuyo caso la curva de Lorenz coincidiría con la diagonal. De hecho, el índice de Gini es una medida del área comprendido entre la curva de Lorenz y la diagonal. En el caso de máxima concentración o desigualdad de distribución de las acciones tendríamos un índice de Gini con valor 1 y una curva que iría prácticamente siguiendo a los ejes inferior y derecho del cuadrado.

7 Dada una variable estadística X:

a) Escriba la fórmula del momento ordinario (o respecto al origen) de orden 'p'.

b) Calcule los valores de los momentos ordinarios de orden 0, 1 y 2.

c) Escriba la fórmula del momento central (o respecto a la media) de orden 'p'.

d) Escriba, en su caso, la relación entre el momento central de orden 2 y los momentos ordinarios de orden 1 y 2.

e) Indique, en general, la fórmula de la relación entre los momentos centrales y los momentos ordinarios.

Vamos a resolver todo el ejercicio asumiendo que se tienen n datos, que pueden ser o no repetidos y sin agruparlos.

a) El momento respecto al origen de orden p viene dado por:

$$a_p = \frac{\sum_{i=1}^{n} x_i^p}{n}.$$

☎ 669 31 64 06 // academia@academiadeimos.com

b) El momento ordinario de orden 0 es $\sum_{i=1}^{n} 1/n = 1$ para cualquier variable X. El momento ordinario de orden 1 es:

$$a_1 = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x},$$

que es la media aritmética. Y el momento ordinario de orden 2 es:

$$a_2 = \frac{\sum_{i=1}^{n} x_i^2}{n}.$$

c) El momento central de orden p es:

$$m_p = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^p}{n}.$$

d) El momento central de orden 2 es la varianza y se relaciona con los momentos ordinarios de orden 1 y 2 a través de la expresión:

$$m_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = a_2 - a_1^2.$$

e) La relación entre los momentos centrales y los momentos ordinarios general es:

$$m_k = \sum_{i=0}^{k} (-1)^i \binom{k}{i} a_{k-i} a_1^i, \quad \forall i \geq 2,$$

siendo $\binom{k}{i}$ el número combinatorio k sobre i.

8 Multicolinealidad. Definición y consecuencias.

La multicolinealidad surge en el contexto de las regresión múltiple. Es un problema que ocurre cuando alguna de las variables independientes está relacionada linealmente de forma perfecta con otras variables explicativas. Es, decir, si el conjunto de variables explicativas está formado por $x_1, \ldots, x_k$ el problema surge cuando existen coeficientes α_i con algún coeficiente no nulo de manera que se cumple:

$$\alpha_0 + \sum_{i=1}^{k} \alpha_i x_i = 0.$$

Otra forma de verlo es que la correlación múltiple de alguno de los x_i con el resto de variables sea 1. Esto supone un problema porque entonces no se pueden resolver de forma unívoca las ecuaciones de la regresión que nos proporcionan los coeficientes, y tenemos un problema de identificabilidad.
La forma de lidiar con este problema consiste en localizar dichas variables independientes que están relacionadas linealmente con el resto y eliminarlas de la regresión.

9 Dos variables tienen las siguientes rectas de regresión mínimo cuadráticas

$$6x + 2y = 1, \quad 16x + 7y = 1.$$

Calcular el coeficiente de correlación.

Cuando tenemos las dos rectas de regresión simple el cuadrado del coeficiente de correlación viene dado por el producto de las pendientes. Es decir, si r es el coeficiente de correlación entonces:

$$r^2 = b \cdot b',$$

donde b y b' son las pendientes de la recta de regresión. Hemos decidir cual de las rectas proporcionadas es la de y en función de x. Para ello el criterio es que r^2 ha de estar entre 0 y 1 de manera que la elección tiene que ser:

$$x = \frac{1}{6} - \frac{1}{3}y, \quad y = \frac{1}{7} - \frac{16}{7}x.$$

Si hiciésemos la elección contraria tendríamos coeficientes 3 y $\frac{7}{16}$ que darían un r^2 superior a 1. El coeficiente de correlación resultante es:

$$r^2 = b \cdot b' = \frac{16}{3 \cdot 7} = \frac{16}{21} \Rightarrow r = \sqrt{\frac{16}{21}} = \frac{4}{\sqrt{21}}.$$

☎ 669 31 64 06 // academia@academiadeimos.com

10 ¿ Qué significa que una serie temporal sea aditiva o mutiplicativa? ¿Qué relación hay entre ambos modelos? Si una serie temporal es aditiva quiere decir que sus

componentes se combinan de manera aditiva, mediante sumas, y por tanto las variaciones debidas a la estacionalidad y la componente cíclica no dependen del nivel de la tendencia. Una manera habitual de escribir un modelo aditivo para una serie temporal es:

$$Y_{t,i} = T_{t,i} + e_{t,i} + c_{t,i} + r_{t,i}.$$

Si en cambio tenemos una serie temporal multiplicativa sus componentes se combinan mediante multiplicación. De esta manera tenemos un modelo de la forma:

$$Y_{t,i} = T_{t,i} \cdot e_{t,i} \cdot c_{t,i} \cdot r_{t,i}.$$

En este modelo, la amplitud de las variaciones debidas a la estacionalidad y el ciclo sí que dependen del nivel de la tendencia.

Se puede pasar fácilmente de un modelo a otro mediante el uso de logaritmos (o recíprocamente de la función exponencial). Por ejemplo, partiendo del modelo multiplicativo y aplicando logaritmo neperiano en ambos lados de la ecuación del modelo y teniendo en cuenta que el logaritmo del producto es la suma de logaritmos resulta:

$$\log Y_{t,i} = \log T_{t,i} + \log e_{t,i} + \log c_{t,i} + \log r_{t,i}.$$

Que cumple precisamente la ecuación de una serie aditiva.

11 Indique en qué fase y subfase del GSBPM se incluye el control del secreto estadístico. Defina el objetivo y contenido de dicha subfase.

El control del secreto estadístico es la subfase 6.4 de la fase análisis del GSBPM. Este subproceso tiene por objetivo asegurar que los datos (y metadatos) que se van a publicar no incumplan las reglas de confidencialidad y asegurar el cumplimiento del secreto estadístico a que están sujetos los productores de estadísticas. El grado y método de anonimización del control del secreto estadístico puede variar para distintos tipos de resultados, por ejemplo, el enfoque usado para conjuntos de microdatos con fines de investigación será diferente al que se usa para publicar tablas o mapas. Mientras que en microdatos por ejemplo es habitual eliminar o perturbar variables o crear categorías más agregadas, en las tablas es habitual no proporcionar información de ciertas celdas con tamaños pequeños.

19. En una población estratificada en tres estratos, se consideran los siguientes datos de una característica X:

$$\sum_{h=1}^{3} W_h S_h^2 = 600, \quad N = 1000.$$

Siendo:
W_h los pesos relativos del estrato h
S_h^2 cuasivarianza poblacional del estrato h
N Tamaño de la población
En un muestreo aleatorio simple sin reposición, determine el tamaño de muestra que, con afijación proporcional, proporciona una varianza del estimador de la media igual a 5,4.

En el muestreo aleatorio simple estratificado la varianza del estimador de la media viene dada por:

$$V(\bar{X}) = \sum_{h=1}^{3} W_h^2(1-f_h)\frac{S_h^2}{n_h} = 5{,}4$$

Siendo f_h la fracción de muestreo en el estrato h. Como la afijación es proporcional tenemos que:

$$n_h = W_h \cdot n \Rightarrow f_h = \frac{n_h}{N_h} = \frac{W_h \cdot n}{W_h \cdot N} = \frac{n}{N}.$$

Por tanto, queda:

$$\left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{3} W_h S_h^2 = 5{,}4 \Rightarrow \frac{1}{n} = \frac{5{,}4}{600} + \frac{1}{1000} = \frac{54+6}{6000} = \frac{60}{600} = \frac{1}{100} \Rightarrow n = 100.$$

20. Sean dos variables aleatorias X, Y que toman los siguientes pares de valores en una población de tamaño $N = 3$:

U_i	X_i	Y_i
U_1	2	1
U_2	6	1
U_3	10	3

Obtenga todas las muestras posibles de tamaño $n = 2$ para (X, Y) con muestreo sin reposición y probabilidades iguales, considerando $(U_i, U_j) = (U_j, U_i)$. Calcular la Esperanza del estimador de la razón poblacional $R = X/Y$.

En el muestreo sin reposición y con probabilidades iguales y considerando $(U_i, U_j) = (U_j, U_i)$ las posibles muestran S_i son $(U_1, U_2), (U_1, U_3)$ y (U_2, U_3). Todas ellas son equiprobables, con probabilidad $1/3$ y el estimador de la razón poblacional viene dado por la razón muestral $\hat{R} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} Y_i}$.

Con todo ello en cuenta la distribución en el muestreo del estimador de la razón es:

S_i	$P(S_i)$	$\hat{R}$
(U_1, U_2)	$1/3$	$8/2=4$
(U_1, U_3)	$1/3$	$12/4=3$
(U_2, U_3)	$1/3$	$16/4=4$

Así, la esperanza del estimador de la razón es:

$$E(\hat{R}) = \frac{1}{3} \cdot (4+3+4) = \frac{11}{3}.$$

El valor de la razón poblacional es $18/5$, que vemos que no coincide con la esperanza del estimador, lo cual es de esperar porque el estimador de razón no es insesgado en general.

21. ¿ Cuál es el primer software que se ejecuta al encender un ordenador? ¿Dónde se encuentra?

El primer software que se ejecuta al encender un ordenador es la BIOS. Se encuentra en la placa base en un chip de memoria ROM, que es una memoria de sólo lectura para que la BIOS no pueda ser modificada.

22. Disponemos de 1 byte de memoria para almacenar números enteros no negativos,

a) ¿ cuál es el mayor entero que se podría almacenar?

b) Si el código ASCII para la letra minúscula 'a' es 97. ¿ Cómo se representa en bits la letra 'm' en código ASCII?

a) Un byte está formado por 8 bits. Por tanto, el mayor número entero no negativo que se puede almacenar es $2^8 - 1 = 255$ que se representaría como una secuencia de ocho 1s consecutivos.

b) En el código ASCII las letras siguen el orden habitual del abecedario. Como la letra 'mócupa la posición de la letra á'más 12 siguiendo el abecedario, el código ASCII para la letra 'm'minúscula es 97+12=109. Ahora, para saber su representación en bits nos basta con obtener la representación del entero no negativo 109:

$$109 = 64 + 32 + 8 + 4 + 1 = 2^6 + 2^5 + 2^3 + 2^2 + 2^0 \Rightarrow 01101101$$

es la representación en bits de la letra 'm'minúscula en código ASCII.