2°do Parcial - 1ra Fecha

1) Considere una muestra aleatoria de una distribución continua con densidad

$$f(x) = \begin{cases} \frac{2(\theta - x)}{\theta^2} & , \text{ si } 0 \leq x \leq \theta \\ 0 & , \text{ c.c} \end{cases}$$

a) Determinar el estimador, por el método de los momentos, del parámetro poblacional θ

b) ¿El estimador hallado en a) es insesgado? ¿es consistente?

Tenemos una muestra aleatoria de una distribución continua con densidad $f(x)$

a) Estimamos, a través del método de los momentos, el parámetro poblacional θ

Planteo de orden 1 $\quad \mu_1 = M_1 \rightarrow \int_0^\theta x f(x)\,dx = \frac{1}{m}\sum_{i=1}^{m} x_i$

Calculamos

$$\mu_1 = \int_0^\theta x f(x)\,dx = \int_0^\theta x \, \frac{2(\theta - x)}{\theta^2}\,dx = \frac{2}{\theta^2}\int_0^\theta (\theta x - x^2)\,dx =$$

$$= \frac{2}{\theta^2}\left(\frac{\theta x^2}{2} - \frac{x^3}{3}\right)\Big|_0^\theta = \left(\frac{x^2}{\theta} - \frac{2x^3}{3\theta^2}\right)\Big|_0^\theta = \frac{\theta^2}{\theta} - \frac{2\theta^3}{3\theta^2} = \theta(1 - 2/3) = \frac{\theta}{3}$$

Entonces $\quad \mu_1 = M_1$

$$\int_0^\theta x f(x)\,dx = \frac{1}{m}\sum_{i=1}^{m} x_i \rightarrow \frac{\theta}{3} = \frac{1}{m}\sum_{i=1}^{m} x_i$$

$$\rightarrow \theta = 3 \cdot \underbrace{\frac{1}{m}\sum_{i=1}^{m} X_i}_{=\overline{X}}$$

$$\rightarrow \hat{\theta}_{MM} = 3\overline{X}$$

$\wedge \rightarrow$ indica estimador

$MM \rightarrow$ obtenido por Método de momentos

b) Un estimador es insesgado si $E(\hat{\theta}) = \theta$

Entonces $E(\hat{\theta}_{MM}) = E(3\overline{X})$

$= 3E(\overline{X})$ $\quad$ * linealidad de la Esperanza.

$= 3E\left(\frac{1}{m}\sum_{i=1}^{m} X_i\right)$

$= \frac{3}{m}\sum_{i=1}^{m} \underbrace{E(X_i)}_{\theta/3}$ $\quad$ * linealidad de la Esperanza

$= \frac{3}{m}\sum_{i=1}^{m} \theta/3$

$= \frac{3}{m}\frac{\theta\, m}{3}$

$= \theta$ ✓

∴ Como $E(\hat{\theta}_{MM}) = \theta$, entonces $\hat{\theta}_{MM}$ es un estimador insesgado

• $\hat{\theta}_{MM}$ será consistente si cumple

$$\begin{cases} \lim\limits_{n\to\infty} E(\hat{\theta}_{MM}) = \theta \\ \lim\limits_{n\to\infty} V(\hat{\theta}_{MM}) = 0 \end{cases}$$

Calculamos $V(\hat{\theta}_{MM}) = V(3\bar{X}) = 3^2 V(\bar{X})$ * linealidad de la varianza

$$= 3^2 V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{3^2}{n^2}\sum_{i=1}^{n} V(X_i)$$ * linealidad de la varianza

$$= \frac{3^2 V(X_i)}{n}$$

$\rightarrow \lim\limits_{n\to\infty} E(\hat{\theta}_{MM}) = \lim\limits_{n\to\infty} \overset{cte}{\theta} = \theta$

$\rightarrow \lim\limits_{n\to\infty} V(\hat{\theta}_{MM}) = \lim\limits_{n\to\infty} \underbrace{\frac{\overset{cte}{3^2 V(X_i)}}{\underset{\to\infty}{n}}}_{\to 0} = 0$

∴ Como se cumplen ambas condiciones $\hat{\theta}_{MM}$ es un estimador consistente.

2) El tiempo de acceso al disco duro en un cierto modelo de computadoras es una variable aleatoria con media 15ms. Se ha propuesto una modificación técnica con objeto de disminuir este tiempo de acceso. Se prueba el nuevo sistema en 10 computadoras obteniéndose una media muestral $\bar{X} = 14$ms y una desviación estándar muestral $s = 2.286$. Suponga que los datos provienen de una población normal.

a) ¿Hay suficiente evidencia estadística, a favor de la hipótesis de que el nuevo modelo disminuye el tiempo medio de acceso? Decida con el p-valor

b) Construya un intervalo de confianza de nivel 95% para la verdadera media del tiempo de acceso al disco.

Sea X_i "tiempo de acceso al disco duro de la i-ésima computadora" $i = 1, 2, \dots, 10$

• $\mu_0 = 15$ ms • $n = 10$ • $\bar{X} = 14$ms • $s = 2{,}286$

Además $X_i \sim N(\mu, \sigma^2)$, σ^2 desconocida

a) Planteamos hipótesis $H_0: \mu = 15$ contra $H_1: \mu < 15$ → test unilateral

• El estadístico a utilizar será:

$T = \dfrac{\bar{X} - \mu}{s/\sqrt{n}}$. T tiene distribución student con parámetro $\underset{\text{grados de libertad}}{n-1}$ si H_0 es V.

• Criterio de aceptación / rechazo

$$\begin{cases} \text{Se acepta } H_0 \text{ si } t_0 \geq t_{\alpha, n-1} \\ \text{Se rechaza } H_0 \text{ si } t_0 < t_{\alpha, n-1} \end{cases}$$

$t_0 = \dfrac{14 - 15}{2{,}286/\sqrt{10}} = -1{,}3833$

El enunciado nos pide que utilicemos el p-valor para determinar si rechazamos o no la hipótesis nula

$$p\text{-valor} = P(T < t_0) = P(T < -1{,}3833) \underset{\uparrow \text{tabla}}{\simeq} 0{,}1$$

Criterio p-valor
$\begin{cases} p > \alpha & \times \\ p < \alpha & \checkmark \end{cases}$

Luego como p-valor $\simeq 0{,}1 > \underset{=\alpha}{0{,}05}$, no hay suficiente evidencia

$\therefore$ no hay suficiente evidencia estadística a favor de la hipótesis de que el nuevo modelo disminuye el tiempo medio de acceso al disco duro.

b) Queremos construir un intervalo de confianza del 95 % para la verdadera media de tiempo de acceso al disco.

$$1-\alpha = 0{,}95 \to \alpha = 0{,}05 \to \alpha/2 = 0{,}025$$

De acuerdo a los datos del ejercicio, la función pivote será $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$ y el intervalo de confianza lo calculamos con

$$IC_{(0{,}05)}(\mu) = \left[\bar{X} - t_{\frac{\alpha}{2}, n-1}\frac{S}{\sqrt{n}},\ \bar{X} + t_{\frac{\alpha}{2}, n-1}\frac{S}{\sqrt{n}}\right]$$

$$= \left[14 - t_{0{,}025, 9}\frac{2{,}286}{\sqrt{10}},\ 14 + t_{0{,}025, 9}\frac{2{,}286}{\sqrt{10}}\right]$$

$$= \left[14 - 2{,}262 \cdot \frac{2{,}286}{\sqrt{10}},\ 14 + 2{,}262 \cdot \frac{2{,}286}{\sqrt{10}}\right]$$

$$[12{,}3648,\ 15{,}6352]$$

3) Los tiempos de ejecución (en segundos) de 40 trabajos procesados por un centro de cálculo han resultado ser

10	19	90	40	15	11	32	17	4	152	23	13	36	101	2	14	2	23	34	15
27	1	57	17	3	30	50	4	62	48	9	11	20	13	38	54	46	12	5	26

a) Obtener el intervalo de confianza de 90 % para la media del tiempo de ejecución de un trabajo
b) ¿Hay suficiente evidencia estadística a favor de la hipótesis de que la verdadera media del tiempo de ejecución es mayor que 25 segundos? p-valor

Sea v.a. X_i = "tiempo de procesado del i-ésimo trabajo en segundos"

Sabemos que $n = 40$. Podemos estimar $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \dfrac{1186}{40} = 29{,}65$

σ^2 es desconocida y la distribución de X_i también, pero como $n = 40 > 30$ podemos hacer:

Escaneado con CamScanner

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{39}\sum_{i=1}^{40}(x_i - 29{,}65)^2} \underset{(*)_1}{=} \sqrt{\frac{1}{39}\cdot 36017{,}1} = 30{,}39$$

Calculemos la sumatoria

$$\sum_{i=1}^{40}(x_i - 29{,}65)^2 = (10-29{,}65)^2 + (19-29{,}65)^2 + (90-29{,}65)^2 + (40-29{,}65)^2$$
$$+ (15-29{,}65)^2 + (11-29{,}65)^2 + (32-29{,}65)^2 + (17-29{,}65)^2 + (4-29{,}65)^2$$
$$+ (152-29{,}15)^2 + (23-29{,}65)^2 + (13-29{,}65)^2 + (36-29{,}65)^2 + (101-29{,}65)^2$$
$$+ (2-29{,}65)^2 + (14-29{,}65)^2 + (26-29{,}65)^2 + (2-29{,}65)^2 + (23-29{,}65)^2$$
$$+ (34-29{,}65)^2 + (15-29{,}65)^2 + (27-29{,}65)^2 + (1-29{,}65)^2 + (57-29{,}65)^2$$
$$+ (17-29{,}65)^2 + (3-29{,}65)^2 + (30-29{,}65)^2 + (50-29{,}65)^2 + (4-29{,}65)^2$$
$$+ (62-29{,}65)^2 + (48-29{,}65)^2 + (9-29{,}65)^2 + (11-29{,}65)^2 + (20-29{,}65)^2$$
$$+ (13-29{,}65)^2 + (38-29{,}65)^2 + (54-29{,}65)^2 + (46-29{,}65)^2 + (12-29{,}65)^2 + (5-29{,}65)^2$$

$= 36017.1 \; (*)_1$

a) De acuerdo a los datos que tenemos, usamos

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}, \quad Z \simeq N(0,1) \text{ por T.C.L}$$

$\bar{x} = 29{,}65$ $\quad n = 40$

$s = 30{,}39$ $\quad \frac{\alpha}{2} = 0.05$

Buscamos el intervalo de confianza con

$z_{\frac{\alpha}{2}} = 1{,}645$ (tabla)

$$\left[\bar{x} - z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \; \bar{x} + z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right] = \left[29{,}65 \mp 1{,}645 \cdot \frac{30{,}39}{\sqrt{40}}\right]$$

$$= [21.7456, \; 37.5543]$$

b) Quiero saber si hay suficiente evidencia estadística a favor de la hipótesis de que la verdadera media del tiempo de ejecución es mayor a 25 seg

Planteamos hipótesis $H_0: \mu = 25$ contra $H_1: \mu > 25 \rightarrow$ test unilateral

El estadístico de prueba a utilizar es $Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$, $Z \simeq N(0,1)$ (por TCL), si H_0 es V

Criterio de rechazo / aceptación:

$$\begin{cases} \text{acepto } H_0 \text{ si } z_0 \leq z_\alpha \\ \text{rechazo } H_0 \text{ si } z_0 > z_\alpha \end{cases}$$

$$z_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{29{,}65 - 25}{30{,}39/\sqrt{40}} = 0{,}9677$$

Usamos el p-valor

criterio p-valor
$$\begin{cases} p > \alpha \; ✗ \\ p < \alpha \; ✓ \end{cases}, \quad \alpha = 0.05$$

$$P(Z > z_0) = 1 - P(Z \leq z_0)$$
$$= 1 - P(Z \leq 0{,}9677)$$
$$\approx 1 - \Phi(0{,}9677) = 1 - 0{,}834 = 0{,}166$$

Escaneado con CamScanner

Como p-valor = 0,166 > 0,05, no hay suficiente evidencia estadística a favor de la hipótesis.

(4) Se probaron algunas muestras de carbón de cada una de dos minas, y se midió la capacidad calorífica (en kilocalorías por libra) para cada muestra. Los resultados fueron

MINA 1 =	4167	4268	4159	4285	4229	4386	4103
MINA 2 =	3924	3988	4096	4026	4235	4178	

a) Determine un intervalo de confianza de 90% para la diferencia en la capacidad calorífica entre el carbón de la mina 1 y el de la mina 2. Suponga que los datos provienen de poblaciones normales pendientes

b) Pueden concluir que las capacidades caloríficas del carbón de las dos minas son diferentes? Utilice el intervalo de confianza encontrado en el inciso anterior ¿Cuál es el nivel de significancia del test?

Sea $\overline{X}_{i\ 1,2}$ = "capacidad calorífica (en kilocalorías por libra) para la i-ésima muestra de la mina"

Sabemos que $X_{1_i} \sim N(\mu_1, \sigma_1^2)$ y $X_{2_i} \sim N(\mu_2, \sigma_2^2)$ independientes

Calculamos

- $\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1_i} = \frac{29597}{7} = 4228{,}14$
- $\overline{X}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2_i} = \frac{24447}{6} = 4074{,}5$
- $n_1 = 7$
- $n_2 = 6$
- $S_1 = \sqrt{\frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_{1_i}-\overline{X}_1)^2}$
- $S_2 = \sqrt{\frac{1}{n_2-1}\sum_{i=1}^{n_2}(X_{2_i}-\overline{X}_2)^2}$

$S_1 = \sqrt{\frac{53920{,}8972}{6}} = 94{,}7988$

$S_2 = \sqrt{\frac{69419{,}5}{5}} = 117{,}83$

$\sum_{i=1}^{7}(X_{1_i}-4228{,}14)^2 = (4167-4228{,}14)^2 + (4268-4228{,}14)^2 + (4159-4228{,}14)^2 + (4285-4228{,}14)^2 + (4229-4228{,}14)^2 + (4386-4228{,}14)^2 + (4103-4228{,}14)^2 \Leftrightarrow$

$\Leftrightarrow 53920{,}8972$

$\sum_{i=1}^{6}(X_{2_i}-4074{,}5)^2 = (3924-4074{,}5)^2 + (3988-4074{,}5)^2 + (4096-4074{,}5)^2 + (4026-4074{,}5)^2 + (4235-4074{,}5)^2 + (4178-4074{,}5)^2 = 69419{,}5$