Evaluación Continua
Segundo examen parcial

Duración: 90 minutos

- **No está permitido el uso de documentación, salvo el formulario que haya recibido**
- **Use 4 dígitos decimales en todos los cálculos y resultados**

1. **(2.5 Puntos)** Para estimar la media μ de una población con varianza $\sigma^2 = 2$ a partir de muestras de tamaño 5, se han propuesto los siguientes estimadores:

$$\hat{\mu_1} = \frac{X_1 + X_2 - 2X_3 + X_4 + 2X_5}{3}$$
$$\hat{\mu_2} = \frac{X_1 + 2X_2 - X_3 + X_4 + X_5}{3}$$

   a) **(0.75 Puntos)** Compruebe si ambos estimadores son insesgados.

   Solution

   Tomamos esperanzas en ambos estimadores:

   $E(\hat{\mu_1}) = \frac{E(X_1)+E(X_2)-2E(X_3)+E(X_4)+2E(X_5)}{3} = \frac{\mu+\mu-2\mu+\mu+2\mu}{3} = \frac{3\mu}{3} = \mu$ Insesgado.

   $E(\hat{\mu_2}) = \frac{E(X_1)+2E(X_2)-E(X_3)+E(X_4)+E(X_5)}{3} = \frac{\mu+2\mu-\mu+\mu+\mu}{3} = \frac{4}{3}\mu$ Sesgado.

   b) **(1 Punto)** Calcule el error cuadrático medio de ambos estimadores.

   Solution

   Necesitamos calcular las correspondientes varianzas.
   $var(\hat{\mu_1}) = \frac{1}{9}[var(X_1) + var(X_2) + 4var(X_3) + var(X_4) + 4var(X_5)] = \frac{11}{9}\sigma^2 = \frac{22}{9}$

   $var(\hat{\mu_2}) = \frac{1}{9}[var(X_1) + 4var(X_2) + var(X_3) + var(X_4) + var(X_5)] = \frac{8}{9}\sigma^2 = \frac{16}{9}$

   El sesgo del primer estimador es 0, luego el ECM coincide con su varianza.
   $ECM(\hat{\mu_1}) = \frac{22}{9}$

   El sesgo del segundo estimador es: $\frac{4}{3}\mu - \mu = \frac{1}{3}\mu$
   $ECM(\hat{\mu_2}) = (\frac{1}{3}\mu)^2 + \frac{16}{9}$

   c) **(0.75 Puntos)** ¿Puede ser μ_2 mejor que μ_1 en algún caso?

   Solution

   μ_2 será mejor que μ_1 si ECM(μ_2)<ECM(μ_1), es decir, si:
   $(\frac{1}{3}\mu)^2 + \frac{16}{9} < \frac{22}{9}$
   $\frac{1}{9}\mu^2 < \frac{6}{9}$; $\mu = \sqrt{6} = 2.45$

2. **(2.5 Puntos)** Se sabe que el número de horas que duermen los habitantes de una ciudad se puede aproximar mediante una distribución normal con desviación típica de 0.64 horas. Se tomó una muestra aleatoria simple de tamaño 16 habitantes a los que se les preguntó por el número de horas de sueño, obtienen los siguientes datos:

   6.5 8.5 6.5 8.5 7.5 7.0 5.5 7.5

   7.5 6.5 7.0 7.0 6.5 8.0 7.5 6.0

   a) **(1 Punto)** Halle el intervalo de confianza para la media de la población con una confianza del 95 %.

   Solution

   Se calcula $\overline{x} = 7.09$, se tiene $n = 16$, $\sigma = 0.64$, $z_{\alpha/2} = 1.96$, se sustituye en la fórmula y se obtiene:

$$IC_{95\,\%}(\mu) = (6.78\,;\,7.4).$$

   b) **(1 Punto)** Si se desea estimar la media de dicha población con una precisión de 0.15 (amplitud del intervalo de 0.30) y al mismo nivel de confianza, ¿cuál es el tamaño de muestra necesario?

   Solution

   El error es $E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < 0.15$. Despejando, $n > z^2_{\alpha/2}\frac{\sigma^2}{0.15^2} = 69.93$, por lo tanto el tamaño muestral minimo es de 70 personas.

c) **(0.5 Puntos)** Considere el siguiente intervalo de confianza para la media (6.61 ; 7.57). ¿Qué nivel de confianza se está considerando?

Solution

La longitud del intervalo (6.61 ; 7.57) es 0.96 que es igual a $2 \cdot z_{\alpha/2} \frac{0.64}{4}$. Despejando se obtiene $z_{\alpha/2} = 3$. Es sabido que $P(Z < z_{\alpha/2}) = 1 - \alpha/2$. Luego $P(Z < 3) = 0.9987 = 1 - \alpha/2$ y por tanto $\alpha \simeq 0.0026$.

3. **(2.5 Puntos)** Se piensa que el salario bruto medio anual en el País Vasco es superior al salario medio bruto anual en la Comunidad de Madrid. Para juzgar dicha afirmación se tomaron muestras aleatorias simples de 1000 individuos en las dos comunidades. Se obtuvieron los siguientes resultados (en euros):

	Madrid	País Vasco
Número de observaciones	1000	1000
Media	26179.6	26730.5
Cuasivarianza	5126.69^2	4973.83^2

a) **(1.25 Puntos)** ¿Corroboran los datos obtenidos la afirmación sobre los salarios en el País Vasco y en la Comunidad de Madrid para un nivel de significación del 5 %? Indique concretamente las hipótesis nula y alternativa, la expresión del estadístico del contraste, la región crítica o área de rechazo del contraste y la conclusión a la que llega.

Solution

El contraste de hipótesis que nos piden es:

$$H_0 : \mu_X - \mu_Y = 0$$
$$H_1 : \mu_X - \mu_Y > 0$$

donde X es el salario bruto anual de los trabajadores del País Vasco e Y representa el salario bruto anual de los trabajadores de la Comunidad de Madrid, siendo μ_X y μ_Y sus respectivas medias.

Se trata de muestras grandes en ambos casos, $n_X = n_Y = 1000$, por tanto el estadístico de contraste es:

$$T = \frac{\bar{X} - \bar{Y} - 0}{\sqrt{\frac{s_Y^2}{n_X} + \frac{s_Y^2}{n_Y}}} \sim_{H_0} N(0, 1),$$

que toma el valor

$$t = \frac{(26730.5 - 26179.6) - 0}{\sqrt{\frac{4973.83^2}{1000} + \frac{5126.69^2}{1000}}} = 2.4389.$$

La región de rechazo es:

$$RR_{0.05} = \{T > Z_{0.05}\} = \{T > 1.645\}$$

Como $2.4389 > 1.64$, tenemos suficiente evidencia estadística como para rechazar H_0 y, por tanto, la muestra obtenida proporciona evidencia de que el salario medio bruto en el País Vasco excede al salario medio en Madrid.

b) **(0.75 Puntos)** Calcule el p-valor asociado al contraste de hipótesis anterior.

Solution

$p - valor = P(Z > 2.44) = 1 - 0.9927 = 0.0073.$

c) **(0.5 Puntos)** ¿Qué hipótesis sobre la población y las muestras aseguran la validez del procedimiento empleado en los apartados anteriores?

Solution

Al tratarse de muestras grandes no es necesario suponer nada sobre la distribución de X e Y. Sobre las muestras sí: deben ser cada una de ellas aleatorias y simples y deben ser independientes entre sí.

4. **(2.5 Puntos)** Se quiere explicar el contenido en NO_2 en el aire de una zona urbana a partir de cuatro indicadores ambientales, X_1, X_2, X_3 y X_4. Se proponen los siguientes modelos:

Modelo 1: $y = -2.661 + 0.7333X_1 - 0.0397X_2 - 0.0031X_3 - 0.0032X_4$
$R^2 = 74.4235$ $R^2_{ajustado} = 73.8866$

Variable	p-valor
X_1	0
X_2	0.0304
X_3	0.3638
X_4	0.9432

Modelo 2: $y = -2.6667 - 0.0733X_1 - 0.0395X_2 - 0.0031X_3$
$R^2 = 74.42216$ $R^2_{ajustado} = 73.6228$

Variable	t-statistic
X_1	16.6
X_2	-2.23
X_3	-0.91

Modelo 3: $y = -2.7469 + 0.735X_1 - 0.0393X_2$
$R^2 = 74.1983$ $R^2_{ajustado} = 73.6363$

Variable	p-valor
X_1	0
X_2	0.0280

a) **(0.5 Puntos)** ¿Es el modelo 1 el mejor por tener el $R^2_{ajustado}$ más alto? Justifique la respuesta.

Solution

No, porque el modelo 1 tiene 2 variables no significativas.

b) **(1 Punto)** ¿Cuál es el mejor modelo? Justifique la respuesta. ¿Qué distribución de probabilidad sigue Y bajo dicho modelo?

Solution

El mejor modelo es el 3, porque tiene todas sus variables significativas y el $R^2_{ajustado}$ más alto.
La variable Y seguirá una distribución Normal.

c) **(1 Punto)** Si la varianza residual de ese modelo es 0.045^2, calcule $P(Y < 2)$ cuando todas las variables del modelo toman el valor 7.

Solution

La media de Y viene dada por la recta de regresión. Para el modelo 3:
$Y = -2.7469 + 0.735 \times 7 - 0.0393 \times 7 = 2.123$
$P(Y < 2) = P(Z < \dfrac{2 - 2.123}{0.045}) = P(Z < -2.73) = 0.0032.$