DEPARTAMENTO DE MATEMÁTICA
MÉTODOS ESTATÍSTICOS
2.º Semestre - 2020/2021
2.º Teste

Data: 26 de junho de 2021 Duração: 2 horas

Resolução

1. Sejam:

   X_1- o caudal mensal (em metros cúbicos por segundo) do rio Tejo, tal que $X_1 \sim N(451, 33)$

   X_2- o caudal mensal (em metros cúbicos por segundo) do rio Douro, tal que $X_2 \sim N(631, 42)$

   X_3- o caudal mensal (em metros cúbicos por segundo) do rio Sado, tal que $X_3 \sim N(\mu, \sigma)$

   (a) Pretende-se determinar $\sigma =$ desvio padrão do caudal mensal do rio Sado e sabe-se que:
   - $P(X_3 > 101) = 0.0228$
   - $\mu = 45\, m^3/s$

   Portanto tem-se

$$X_3 \sim N(\mu, \sigma) \Leftrightarrow X_3 \sim N(45, \sigma) \Leftrightarrow Z = \frac{X_3 - 45}{\sigma} \sim N(0,1)$$

   logo

$$P(X_3 > 101) = 0.0228 \Leftrightarrow P\left(Z > \frac{101-45}{\sigma}\right) = 0.0228 \Leftrightarrow P\left(Z > \frac{56}{\sigma}\right) = 0.0228 \Leftrightarrow$$
$$\Leftrightarrow 1 - P\left(Z \leq \frac{56}{\sigma}\right) = 0.0228 \Leftrightarrow 1 - \Phi\left(\frac{56}{\sigma}\right) = 0.0228 \Leftrightarrow \Phi\left(\frac{56}{\sigma}\right) = 0.9772 \Leftrightarrow$$
$$\Leftrightarrow \frac{56}{\sigma} = z_{0.9772} \Leftrightarrow \frac{56}{\sigma} = 2 \Leftrightarrow \sigma = 28\, m^3/s$$

   (b) Seja $S = X_1 + X_2$ a variável aleatória que representa a soma dos caudais mensais dos rios Tejo e Douro. Como X_1 e X_2 são variáveis aleatórias independentes e com distribuição Normal, então pela aditividade da distribuição Normal tem-se

$$S = X_1 + X_2 \sim N\left(1082, \sqrt{2853}\right) \Leftrightarrow Z = \frac{S - 1082}{\sqrt{2853}} \sim N(0,1)$$

   pois

$$\begin{aligned} \mu_S &= \mu_{X_1} + \mu_{X_2} = 451 + 631 = 1082 \\ \sigma_S &= \sqrt{\sigma^2_{X_1} + \sigma^2_{X_2}} = \sqrt{33^2 + 42^2} = \sqrt{2853} \end{aligned}$$

   Portanto pretende-se

$$\begin{aligned} P(S > 1000) &= P\left(Z > \tfrac{1000-1082}{\sqrt{2853}}\right) = P(Z > -1.54) = 1 - P(Z \leq -1.54) = 1 - \Phi(-1.54) = \\ &= 1 - (1 - \Phi(1.54)) = \Phi(1.54) = 0.9382 \end{aligned}$$

   (c) População: $Y-$ o caudal, em m^3/s, do rio Guadiana, tal que $Y \sim N(\mu_y, \sigma_y)$

   Amostra: $n = 19$

i. Estimativa pontual para μ_y = caudal médio do rio Guadiana:

$$\bar{y} = \frac{1328}{19} = 69.895\, m^3/s$$

Estimativa pontual para σ_y = desvio padrão do caudal do rio Guadiana:

$$s_y = \sqrt{\frac{94656 - 19 \times \left(\frac{1328}{19}\right)^2}{19-1}} = 10.099\, m^3/s$$

ii. Como a População é Normal, σ_y desconhecido e $n = 19 < 30$, então o Intervalo de confiança a $(1-\alpha) \times 100\%$ para μ_y é

$$\left]\bar{y} - t_{(n-1);1-\frac{\alpha}{2}} \frac{s_y}{\sqrt{n}}, \bar{y} + t_{(n-1);1-\frac{\alpha}{2}} \frac{s_y}{\sqrt{n}}\right[$$

Como grau de confiança $= 1 - \alpha = 0.99$, tem-se nível de significância $= \alpha = 0.01$, então:

$$t_{(n-1);1-\frac{\alpha}{2}} = t_{(19-1);1-\frac{0.01}{2}} = t_{(18);0.995} = 2.88$$

portanto o Intervalo de confiança a 99% para μ_y, o caudal médio do rio Guadiana, é:

$$\left]69.895 - 2.88 \times \frac{10.099}{\sqrt{19}}, 69.895 + 2.88 \times \frac{10.099}{\sqrt{19}}\right[=]63.222, 76.567[$$

Como $60 \notin]63.222, 76.567[$, com 99% de confiança pode afirmar-se que o caudal médio do rio Guadiana é significativamente diferente de 60 m^3/s, portanto a afirmação dos ambientalistas não parece ser válida.

iii. O responsável pelo estudo tem duas possibilidades para diminuir a margem de erro do Intervalo de Confiança:

- aumentar a dimensão da amostra, ou seja, efetuar mais do que 19 medições do caudal do rio Guadiana,

  ou
- diminuir o grau de confiança, o que significa que estará a aumentar o nível de significância.

iv. Sabe-se que o Intervalo de confiança a $(1-\alpha) \times 100\%$ para σ_y = desvio padrão do caudal do rio Guadiana é:

$$]7.970, 13.982[$$

Como tem-se: População Normal com μ_y desconhecido, então o Intervalo de confiança a $(1-\alpha) \times 100\%$ para σ_y é

$$\left]\sqrt{\frac{(n-1)\, s_y^2}{x^2_{n-1;1-\frac{\alpha}{2}}}}, \sqrt{\frac{(n-1)\, s_y^2}{x^2_{n-1;\frac{\alpha}{2}}}}\right[$$

considerando, por exemplo, o extremo superior tem-se

$$\sqrt{\frac{(n-1)\, s_y^2}{x^2_{n-1;\frac{\alpha}{2}}}} = 13.982 \Leftrightarrow \sqrt{\frac{(19-1) \times 10.099^2}{x^2_{19-1;\frac{\alpha}{2}}}} = 13.982 \Leftrightarrow \frac{(19-1) \times 10.099^2}{x^2_{18;\frac{\alpha}{2}}} = 13.982^2 \Leftrightarrow$$

$$\Leftrightarrow x^2_{18;\frac{\alpha}{2}} = \frac{(19-1) \times 10.099^2}{13.982^2} \Leftrightarrow x^2_{18;\frac{\alpha}{2}} = 9.39 \Leftrightarrow \frac{\alpha}{2} = 0.05 \Leftrightarrow \alpha = 0.10$$

logo o grau de confiança é

$$1 - \alpha = 1 - 0.10 = 0.90$$

2. Seja X- resistência do novo produto. O fabricante afirma que $\mu = E[X] = 4$ Ohms.

(a) Amostra de dimensão $n = 100$ componentes com 17 componentes defeituosas.

i. População: $Y \sim$ Binomial, onde p representa a proporção de componentes defeituosas produzidas.

Como tem-se: População Binomial e $n = 100 \geq 30$, o Intervalo de confiança a $(1-\alpha) \times 100\%$ para p é:

$$\left]p^* - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^* \times q^*}{n}}, p^* + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^* \times q^*}{n}}\right[$$

Como grau de confiança $= 1 - \alpha = 0.95$, tem-se nível de significância $= \alpha = 0.05$ e

$$z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.05}{2}} = z_{0.975} = 1.96$$

A proporção amostral é

$$p^* = \frac{17}{100} = 0.17 \qquad \text{e} \qquad q^* = 1 - p^* = 1 - 0.17 = 0.83$$

Então o Intervalo de confiança a 95% para p, a proporção de componentes defeituosas produzidas, é:

$$\left]0.17 - 1.96 \times \sqrt{\frac{0.17 \times 0.83}{100}}, 0.17 + 1.96 \times \sqrt{\frac{0.17 \times 0.83}{100}}\right[=]0.0964, 0.2436[$$

ii. Pretende-se determinar n tal que o erro de estimação seja menor que 2%, com um grau de confiança de 98%. O erro de estimação corresponde à margem de erro, portanto pretende-se:

$$\text{margem de erro } < 0.02 \Leftrightarrow \frac{\left(p^* + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^* \times q^*}{n}}\right) - \left(p^* - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^* \times q^*}{n}}\right)}{2} < 0.02 \Leftrightarrow$$

$$\Leftrightarrow z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^* \times q^*}{n}} < 0.02 \Leftrightarrow 2.326 \times \sqrt{\frac{p^* \times q^*}{n}} < 0.02$$

pois grau de confiança $= 1 - \alpha = 0.98$ logo nível de significância $= \alpha = 0.02$ e

$$z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.02}{2}} = z_{0.99} = 2.326$$

Agora existem duas possibilidades:

- Supondo que se mantém a proporção amostral obtida com a amostra de dimensão 100, $p^* = 0.17$ e $q^* = 0.83$, então tem-se

$$2.326 \times \sqrt{\frac{0.17 \times 0.83}{n}} < 0.02 \Leftrightarrow n > 1908.5 \Rightarrow n \geq 1909$$

ou seja, devem ser recolhidas no mínimo mais $1909 - 100 = 1809$ componentes.

- Como vai ser recolhida uma amostra e nada garante que se mantenha o que foi observado na amostra de dimensão 100, então vamos supor a pior situação (a que dá origem à maior dimensão da amostra) $p^* = q^* = 0.5$ e tem-se

$$2.326 \times \sqrt{\frac{0.5 \times 0.5}{n}} < 0.02 \Leftrightarrow n > 3381.4 \Rightarrow n \geq 3382$$

ou seja, devem ser recolhidas no mínimo mais $3382 - 100 = 3282$ componentes.

(b) População: X- resistência do novo produto. O fabricante afirma que $\mu = E[X] = 4$ Ohms.

Amostra: $n = 50$, $\overline{x} = 3.7$ Ohms e $s^2 = 2.96$ Ohms^2.

Hipóteses:

$$\begin{cases} H_0 : \mu = 4 \rightarrow & \text{o fabricante deve manter a resistência anunciada} \\ vs & \\ H_1 : \mu < 4 \rightarrow & \text{o fabricante deve diminuir a resistência anunciada} \end{cases}$$

Estatística de Teste: População desconhecida mas $n = 50 \geq 30$ (permite recorrer ao Teorema do limite central) e σ é desconhecido

$$Z = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim N(0,1)$$

Valor observado da Estatística de Teste, considerando a amostra recolhida e supondo a hipótese H_0 verdadeira:

$$Z_{\text{obs}} = \frac{3.7 - 4}{\frac{\sqrt{2.96}}{\sqrt{50}}} = -1.23$$

Valor$-p$: o teste é Unilateral Esquerdo

$$\text{valor} - p = P(Z \leq Z_{\text{obs}}) = P(Z \leq -1.23) = \Phi(-1.23) = 1 - \Phi(1.23) = 1 - 0.8907 = 0.1093$$

Como o valor $- p > \alpha = 0.05$, então não se rejeita a H_0. Com base na amostra e para um nível de significância de 5%, não há evidência estatística que o fabricante deva diminuir a resistência anunciada, ou seja, o fabricante deve manter a resistência de 4 Ohms.

3. Populações:

   População 1: X = quilómetros por litro percorridos com a gasolina A

   População 2: Y = quilómetros por litro percorridos com a gasolina B

   Amostras Aleatórias Independentes:

   Amostra Aleatória da População 1: amostra 1

   Amostra Aleatória da População 2: amostra 2

   (a) Teste de Mann-Whitney $\rightarrow$ as amostras são independentes

   Hipóteses:

$$\begin{cases} H_0 : M_X = M_Y \rightarrow & \text{não existem diferenças nos km/l percorridos com as duas gasolinas} \\ vs & \\ H_1 : M_X < M_Y \rightarrow & \text{com a gasolina A são percorridos menos km/l do que com a gasolina B} \end{cases}$$

   Estatística de Teste:

$$U = S_1 - \frac{n \times (n+1)}{2}$$

   Valor observado da Estatística de Teste, considerando as amostras recolhidas e supondo a hipótese H_0 verdadeira:

$$U_{\text{obs}} = W = 44.5$$

   Região Crítica: $\alpha = 0.05$, $n = m = 11$ e o teste é Unilateral Esquerdo

$$RC =]-\infty, U_{n,m;\alpha}] =]-\infty, U_{11,11;0.05}] =]-\infty, 35]$$

   Como $U_{\text{obs}} = 44.5 \notin RC$, então Não se Rejeita H_0. Com base nas amostras e para um nível de significância de 5%, não há evidência estatística que são percorridos menos quilómetros por litro com a gasolina A do que com a gasolina B.

   (b) Populações:

   População 1: X = quilómetros por litro percorridos com a gasolina A, $X \sim N(\mu_x, \sigma_x)$

   População 2: Y = quilómetros por litro percorridos com a gasolina B, $Y \sim N(\mu_y, \sigma_y)$

   Amostras Aleatórias Independentes:

Amostra Aleatória da População 1: $\bar{x} = 18.082$ e $s_x = 1.834$
Amostra Aleatória da População 2: $\bar{y} = 19.082$ e $s_y = 2.354$
Teste de hipóteses paramétrico:

$$\begin{array}{lcl} H_0: \sigma_x^2 = \sigma_y^2 & & H_0: \frac{\sigma_x^2}{\sigma_y^2} = 1 \rightarrow \text{ as variâncias dos km/l percorridos com os dois tipos de gasolina são iguais} \\ \text{contra} & \Leftrightarrow & \text{contra} \\ H_1: \sigma_x^2 \neq \sigma_y^2 & & H_1: \frac{\sigma_x^2}{\sigma_y^2} \neq 1 \rightarrow \text{ as variâncias dos km/l percorridos com os dois tipos de gasolina são diferentes} \end{array}$$

nível de significância $= \alpha = 0.01$
Estatística de Teste: Populações Normais e amostras independentes

$$\mathrm{F} = \frac{S_x^2}{S_y^2} \times \frac{\sigma_y^2}{\sigma_x^2} \sim F_{(n_x-1, n_y-1)}$$

Valor observado da Estatística de Teste, considerando as amostras recolhidas e supondo a hipótese H_0 verdadeira:

$$\mathrm{F}_{\text{obs}} = \frac{1.834^2}{2.354^2} \times \frac{1}{\frac{\sigma_x^2}{\sigma_y^2}} = \frac{1.834^2}{2.354^2} \times \frac{1}{1} = 0.607$$

Região Crítica: o teste é bilateral, $\mathrm{F} \sim F_{(11-1,11-1)} \Leftrightarrow \mathrm{F} \sim F_{(10,10)}$ e $\alpha = 0.01$

$$\begin{aligned} RC &= \left[0, f_{(n_x-1,n_y-1);\frac{\alpha}{2}}\right] \cup \left[f_{(n_x-1,n_y-1);1-\frac{\alpha}{2}}, +\infty\right[= \\ &= \left[0, f_{(11-1,11-1);\frac{0.01}{2}}\right] \cup \left[f_{(11-1,11-1);1-\frac{0.01}{2}}, +\infty\right[= \left[0, f_{(10,10);0.005}\right] \cup \left[f_{(10,10);0.995}, +\infty\right[= \\ &= \left[0, \frac{1}{f_{(10,10);1-0.005}}\right] \cup [5.85, +\infty[= \left[0, \frac{1}{f_{(10,10);0.995}}\right] \cup [5.85, +\infty[\\ &= \left[0, \frac{1}{5.85}\right] \cup [5.85, +\infty[= [0, 0.17] \cup [5.85, +\infty[\end{aligned}$$

Como $\mathrm{F}_{\text{obs}} = 0.607 \notin RC$, então Não se Rejeita H_0. Com base nas amostras e para um nível de significância de 1%, as variâncias dos quilómetros por litro percorridos com as duas gasolinas podem ser consideradas iguais.

(c) População 1: $X =$ quilómetros por litro percorridos com a gasolina A
Hipóteses:

$$\begin{cases} H_0: X \sim N(19, 1) \\ vs \\ H_1: X \nsim N(19, 1) \end{cases}$$

Decisão:
Como valor-p$= 0.03703 \leq \alpha = 0.10$, então rejeita-se H_0. Com base na amostra 1 e para um nível de significância de 10%, não há evidência estatística que essa amostra venha de uma população com distribuição Normal com média 19 e desvio padrão 1.

População 2: $Y =$ quilómetros por litro percorridos com a gasolina B
Hipóteses:

$$\begin{cases} H_0: Y \sim N(19, 1) \\ vs \\ H_1: Y \nsim N(19, 1) \end{cases}$$

Decisão:
Como valor-p$= 0.2849 > \alpha = 0.10$, então não se rejeita H_0. Com base na amostra 2 e para um nível de significância de 10%, há evidência estatística que essa amostra poderá vir de uma população com distribuição Normal com média 19 e desvio padrão 1.

Fim do teste