IPS Instituto Politécnico de Setúbal
Escola Superior de Tecnologia de Setúbal

DEPARTAMENTO DE MATEMÁTICA
MÉTODOS ESTATÍSTICOS
2.º SEMESTRE - 2020/2021
2.º TESTE

Teste modelo | Duração: 2h00

Resolução

1. Seja X a variável aleatória contínua que representa o coeficiente de inteligência (QI), tal que $X \sim N(100, 16)$ pois $\mu = E[X] = 100$ e $\sigma = \sqrt{V[X]} = 16$.

   (a) $X \sim N(100, 16) \Leftrightarrow Z = \frac{X-100}{16} \sim N(0,1)$. Pretende-se

   $$P(X > c) = 0.95 \Leftrightarrow P\left(Z > \tfrac{c-100}{16}\right) = 0.95 \Leftrightarrow 1 - P\left(Z \leq \tfrac{c-100}{16}\right) = 0.95 \Leftrightarrow$$

   $$\Leftrightarrow 1 - \Phi\left(\tfrac{c-100}{16}\right) = 0.95 \Leftrightarrow \Phi\left(\tfrac{c-100}{16}\right) = 0.05 \Leftrightarrow \tfrac{c-100}{16} = z_{0.05} \Leftrightarrow$$

   $$\Leftrightarrow \tfrac{c-100}{16} = -z_{1-0.05} \Leftrightarrow \tfrac{c-100}{16} = -z_{0.95} \Leftrightarrow \tfrac{c-100}{16} = -1.645 \Leftrightarrow c = 73.68$$

   (b) Sejam:
   $QI = X$ o coeficiente de inteligência, tal que $X \sim N(100, 16)$
   $W-$ o coeficiente da inteligência emocional, tal que $W \sim N(200, \sqrt{50})$ pois $\mu_W = E[W] = 200$ e $\sigma_W^2 = V[W] = 50$.
   $Y = X + W$ a variável aleatória que representa o índice I. Como X e W são variáveis aleatórias independentes e com distribuição Normal, então pela aditividade da distribuição Normal tem-se

   $$Y = X + W \sim N\left(300, \sqrt{306}\right) \Leftrightarrow Z = \frac{Y - 300}{\sqrt{306}} \sim N(0,1)$$

   pois

   $$\begin{aligned} \mu_Y &= \mu_X + \mu_W = 100 + 200 = 300 \\ \sigma_Y &= \sqrt{\sigma_X^2 + \sigma_W^2} = \sqrt{16^2 + 50} = \sqrt{306} \end{aligned}$$

   Portanto tem-se

   $$\begin{aligned} P(Y < 290) &= P\left(Z < \tfrac{290-300}{\sqrt{306}}\right) = P(Z < -0.57) \underset{\text{v.a. contínua}}{=} \Phi(-0.57) = \\ &= 1 - \Phi(0.57) = 1 - 0.7157 = 0.2843 \end{aligned}$$

   (c) Amostra $n = 12$. População Normal com $\sigma = 16$ conhecido, então a distribuição amostral é

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

   Pretende-se

   $$\begin{aligned} P\left(100 < \bar{X} < 110\right) &= P\left(\frac{100-100}{\frac{16}{\sqrt{12}}} < Z < \frac{110-100}{\frac{16}{\sqrt{12}}}\right) = P(0 < Z < 2.17) \underset{\text{v.a. contínua}}{=} \\ &= \Phi(2.17) - \Phi(0) = 0.9850 - 0.5 = 0.4850 \end{aligned}$$

2. População: X variável aleatória contínua que representa quantidade de gordura nos hambúrgueres, tal que $X \sim N(\mu, \sigma)$. Amostra: $n = 101$, $\bar{x} = 30.2$ e $s = 3.8$.

   (a) População Normal com σ desconhecido e $n = 101 \geq 30$, então o Intervalo de confiança a $(1-\alpha) \times 100\%$ para μ é

   $$\left]\bar{x} - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right[$$

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

   $$z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.01}{2}} = z_{0.995} = 2.576$$

   portanto o Intervalo de confiança a 99% para a quantidade média de gordura dos hambúrgueres servidos na cantina é:

   $$\left]30.2 - 2.576 \times \frac{3.8}{\sqrt{101}}, 30.2 + 2.576 \times \frac{3.8}{\sqrt{101}}\right[=]29.226, 31.174[$$

   Como $35 \notin]29.226, 31.174[$, com 99% de confiança pode afirmar-se que a quantidade média de gordura dos hambúrgueres servidos na cantina é significativamente diferente de 35g.

   (b) O Intervalo de confiança a $(1-\alpha) \times 100\%$ para μ é

   $$\left]\bar{x} - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right[$$

   logo

   $$\text{amplitude } = \left(\bar{x} + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right) - \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right) = 2 z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

   portanto pretende-se determinar o n tal que

   $$\text{amplitude } = 0.5 \Leftrightarrow 2 z_{1-\frac{\alpha}{2}} \tfrac{s}{\sqrt{n}} = 0.5 \Leftrightarrow 2 \times 2.576 \times \tfrac{3.8}{\sqrt{n}} = 0.5 \Leftrightarrow n = \left(\tfrac{2 \times 2.576 \times 3.8}{0.5}\right)^2 \Leftrightarrow$$

   $$\Leftrightarrow n = 1533.1 \Rightarrow n = 1534 \text{ hambúrgeres}$$

   (c) Intervalo de confiança a $(1-\alpha) \times 100\%$ para a variância da quantidade de gordura dos hambúrgueres servidos na cantina (σ^2) é:

   $$]11.617, 18.537[$$

   Sabe-se que:
   População Normal com μ desconhecido, então o Intervalo de confiança a $(1-\alpha) \times 100\%$ para σ^2 é

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

   considerando, por exemplo, o extremo superior tem-se

   $$\tfrac{(n-1)s^2}{x^2_{n-1;\frac{\alpha}{2}}} = 18.537 \Leftrightarrow \tfrac{(101-1)\times 3.8^2}{x^2_{101-1;\frac{\alpha}{2}}} = 18.537 \Leftrightarrow \tfrac{100 \times 3.8^2}{x^2_{100;\frac{\alpha}{2}}} = 18.537 \Leftrightarrow x^2_{100;\frac{\alpha}{2}} = \tfrac{100 \times 3.8^2}{18.537} \Leftrightarrow$$

   $$\Leftrightarrow x^2_{100;\frac{\alpha}{2}} = 77.9 \Leftrightarrow \tfrac{\alpha}{2} = 0.05 \Leftrightarrow \alpha = 0.10$$

   logo o grau de confiança é

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

3. População: X = despesa familiar mensal em Portugal, em euros, com $X \sim N(\mu_x, \sigma_x)$

   Amostra Aleatória: $n_x = 31$ e $\bar{x} = 1200$ euros e $s_x^2 = 12544$ euros^2

   (a) Teste de hipóteses paramétrico:

   $H_0 : \mu_x = 1250 \rightarrow$ a despesa média familiar mensal em Portugal é de 1250 euros
   contra
   $H_1 : \mu_x < 1250 \rightarrow$ a despesa média familiar mensal em Portugal é inferior a 1250 euros

   nível de significância = $\alpha = 0.05$

   Estatística de Teste: População Normal, σ_x desconhecido e $n_x = 31 \geq 30$

$$Z = \frac{\bar{X} - \mu_x}{\frac{S_x}{\sqrt{n_x}}} \sim N(0,1)$$

   Valor observado da Estatística de Teste sob a Hipótese H_0 e com base na amostra recolhida:

$$Z_{\text{obs}} = \frac{1200 - 1250}{\frac{\sqrt{12544}}{\sqrt{31}}} = -2.49$$

   Valor-p: o teste é unilateral esquerdo, $Z \sim N(0,1)$ e $\alpha = 0.05$

$$\text{valor-p} = P(Z \leq Z_{\text{obs}}) = P(Z \leq -2.49) = \Phi(-2.49) = 1 - \Phi(2.49) =$$
$$= 1 - 0.9936 = 0.0064$$

   Como valor-p = $0.0064 \leq 0.05 = \alpha$, então Rejeita-se H_0. Com base na amostra e para um nível de significância de 5%, a despesa média familiar mensal em Portugal é inferior a 1250 euros.

   (b) Populações:

   População 1: X = despesa familiar mensal em Portugal, em euros, $X \sim N(\mu_x, \sigma_x)$

   População 2: Y = despesa familiar mensal em Espanha, em euros, $Y \sim N(\mu_y, \sigma_y)$

   Amostras Aleatórias:

   Amostra Aleatória da População 1: $n_x = 31$ e $\bar{x} = 1200$ euros e $s_x^2 = 12544$ euros^2

   Amostra Aleatória da População 2: $n_y = 31$ e $\bar{y} = 1800$ euros e $s_y = 102$ euros

   Teste de hipóteses paramétrico:

   $H_0 : \sigma_x^2 = \sigma_y^2$ contra $H_1 : \sigma_x^2 \neq \sigma_y^2$ $\Leftrightarrow$ $H_0 : \frac{\sigma_x^2}{\sigma_y^2} = 1 \rightarrow$ as variâncias das despesas dos dois países são iguais
   contra
   $H_1 : \frac{\sigma_x^2}{\sigma_y^2} \neq 1 \rightarrow$ as variâncias das despesas dos dois países são diferentes

   nível de significância = $\alpha = 0.02$

   Estatística de Teste: Populações Normais e amostras independentes

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

Valor observado da Estatística de Teste sob a Hipótese H_0 e com base nas amostras recolhidas:

$$\mathrm{F}_{\mathrm{obs}} = \frac{12544}{102^2} \times \frac{1}{\frac{\sigma_x^2}{\sigma_y^2}} = \frac{12544}{102^2} \times \frac{1}{1} = 1.2057$$

Região Crítica: o teste é bilateral, $\mathrm{F} \sim F_{(31-1,31-1)} \Leftrightarrow \mathrm{F} \sim F_{(30,30)}$ e $\alpha = 0.02$

$$\begin{aligned} RC &= \left[0, f_{(n_x-1,n_y-1);\frac{\alpha}{2}}\right] \cup \left[f_{(n_x-1,n_y-1);1-\frac{\alpha}{2}}, +\infty\right[= \\ &= \left[0, f_{(31-1,31-1);\frac{0.02}{2}}\right] \cup \left[f_{(31-1,31-1);1-\frac{0.02}{2}}, +\infty\right[= \left[0, f_{(30,30);0.01}\right] \cup \left[f_{(30,30);0.99}, +\infty\right[= \\ &= \left[0, \frac{1}{f_{(30,30);1-0.01}}\right] \cup [2.39, +\infty[= \left[0, \frac{1}{f_{(30,30);0.99}}\right] \cup [2.39, +\infty[\\ &= \left[0, \frac{1}{2.39}\right] \cup [2.39, +\infty[= [0, 0.4184] \cup [2.39, +\infty[\end{aligned}$$

Como $\mathrm{F}_{\mathrm{obs}} = 1.2057 \notin RC$, então Não se Rejeita H_0. Com base nas amostras e para um nível de significância de 2%, as variâncias das despesas dos dois países podem ser consideradas iguais.

4. Teste de Ajustamento do Qui-Quadrado

Seja X a variável aleatória discreta que representa o último digito do peso de pacientes, com domínio $D_X = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$:

$$\begin{cases} H_0 : X \sim U_{(10)} \rightarrow & \text{os pacientes foram pesados} \\ vs & \\ H_1 : X \nsim U_{(10)} \rightarrow & \text{os pacientes indicaram o seu peso} \end{cases}$$

Como valor-p $= 0.001691 \leq 0.05 = \alpha$, então Rejeita-se H_0. Com base na amostra e para um nível de significância de 5%, não existe evidência estatística que os pacientes tenham sido pesados, ou seja, parece que foram os pacientes que indicaram o seu peso.

5. Teste de Wilcoxon $\rightarrow$ as amostras são emparelhadas

$$\begin{cases} H_0 : M_D = 0 \rightarrow & \text{não existem diferenças entre o tamanho do macho e da fêmea nos casais} \\ vs & \\ H_1 : M_D \neq 0 \rightarrow & \text{existem diferenças entre o tamanho do macho e da fêmea nos casais} \end{cases}$$

onde D = Fêmea - Macho

Valor observado da Estatística de Teste sob a Hipótese H_0:

$$T_{\mathrm{obs}} = min\{T_{\mathrm{obs}}^-, T_{\mathrm{obs}}^+\} = min\{29.5, 6.5\} = 6.5$$

pois

- a soma das $n = 8$ posições (não há zeros) é $\frac{8\times(8+1)}{2} = 36$
- soma das posições com o sinal "+" $= T_{\mathrm{obs}}^+ = V = 6.5$
- soma das posições com o sinal "-" $= T_{\mathrm{obs}}^- = 36 - 6.5 = 29.5$

Região Crítica: $\alpha = 0.05$, $n = 8$ e o teste é Bilateral

$$RC = [0, T_{n;\alpha}] = [0, T_{8;0.05}] = [0, 3]$$

Como $T_{\text{obs}} = 6.5 \notin RC$, então Não se Rejeita H_0. Com base nas amostras e para um nível de significância de 5%, não há evidência estatística que existam diferenças significativas entre o tamanho do macho e da fêmea nos casais destas aves.