DEPARTAMENTO DE MATEMÁTICA
MÉTODOS ESTATÍSTICOS
2.º SEMESTRE - 2022/2023
2.º Teste (recuperação)

Data: 11 de julho de 2023 Duração: 2 horas

Resolução

O teste foi resolvido recorrendo ao software R: ver script_2TesteRecup_ME_22_23.

---

1. Populações:

   X_A- velocidade do vento da localização A

   X_B- velocidade do vento da localização B

   Amostras: $n_A = 11$ e $n_B = 13$

   (a) verificar a normalidade dos dados da Localização A

   Hipóteses: $H_0 : X_A \sim Normal$ contra $H_1 : X_A \nsim Normal$

   Escolha do Teste: Como se pretende saber se X_A tem comportamento normal (sem mais informação) numa amostra com número de elementos inferior a 50, deve-se recorrer ao teste de ajustamento Shapiro-Wilk.

   Tomada de Decisão: Como valor-p$= 0.6715 > 0.1$, então Não se rejeita H_0

   verificar a normalidade dos dados da Localização B

   Hipóteses: $H_0 : X_B \sim Normal$ contra $H_1 : X_B \nsim Normal$

   Escolha do Teste: À semelhança da localização A deve-se recorrer ao teste de ajustamento Shapiro-Wilk.

   Tomada de Decisão: Como valor-p$= 0.208 > 0.1$, então Não se rejeita H_0

   Conclusão: Com base nas amostras e para um nível de signicfiância de 1% há evidência estatística que os dados de ambas as localizações podem ser de populações com distribuição Normal.

   (b) Estimativas pontuais para μ e para σ^2

   Localização A: $\overline{x}_A = 42$ e $s^2_A = 108.2$

   Localização B: $\overline{x}_B = 39.92$ e $s^2_B = 209.7436$

   (c) Como a População pode ser considerada Normal (alínea (a)), então o Intervalo de confiança a $(1 - \alpha) \times 100\%$ para σ^2 é

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

   e sabe-se que o intervalo de confiança obtido para o desvio padrão é $]7.6878; 16.5710[$, ou seja, para a variância fica $]7.6878^2; 16.5710^2[$ então considerando, por exemplo, o limite superior tem-se

   $$\frac{(n-1)s^2}{\chi^2_{n-1;\frac{\alpha}{2}}} = 16.5710^2 \Leftrightarrow \frac{(11-1)\times 108.2}{\chi^2_{11-1;\frac{\alpha}{2}}} = 16.5710^2 \Leftrightarrow \chi^2_{10;\frac{\alpha}{2}} = \frac{10 \times 108.2}{16.5710^2} \Leftrightarrow$$

   $$\Leftrightarrow \chi^2_{10;\frac{\alpha}{2}} = 3.940305 \Leftrightarrow \frac{\alpha}{2} = F\,(3.940305) \underset{X^2 \sim \chi^2_{(10)}}{\Leftrightarrow} \frac{\alpha}{2} = 0.05 \Leftarrow \alpha = 0.10$$

   O grau de confiança utilizado foi de 90%.

(d) Interpretação das Hipóteses:

$$\begin{cases} H_0 : \mu_A = 40 \rightarrow & \text{a velocidade média do vento na localização A é de 40 nós} \\ \text{contra} & \\ H_1 : \mu_A \neq 40 \rightarrow & \text{a velocidade média do vento na localização A não é de 40 nós} \end{cases}$$

Teste das Hipóteses:

$$H_0 : \mu_A = 40 \quad \text{contra} \quad H_1 : \mu_A \neq 40$$

Estatística de Teste: População Normal e σ_A desconhecido

$$T = \frac{\overline{X}_A - \mu_A}{\frac{S_A}{\sqrt{n_A}}} \sim t_(n_A - 1)$$

Nível de significância $= \alpha = 0.01$

Tipo de teste: o teste é Bilateral

Estatística de Teste: $T_{obs} = 0.63769$

Região Crítica: $\left]-\infty, t_{\left(\frac{0.01}{2},10\right)}\right] \cup \left[t_{\left(1-\frac{0.01}{2},10\right)}, +\infty\right[=]-\infty, -3.1693] \cup [3.1693, +\infty[$

Decisão: Como T_{obs} não pertence à região crítica Não se Rejeita H_0

Conclusão: Com base na amostra e para um nível de significância de 1%, existe evidência estatística de que a média da velocidade do vento seja de 40 nós.

(e) Como as populações podem ser consideradas Normais(alínea (a)) e as amostras são independentes, então o Intervalo de confiança a $(1-\alpha) \times 100\%$ para $\frac{\sigma_A^2}{\sigma_B^2}$ é

$$\left] \frac{1}{f_{n_A-1;n_B-1;1-\frac{\alpha}{2}}} \times \frac{s_A^2}{s_B^2}, \frac{1}{f_{n_A-1;n_B-1;\frac{\alpha}{2}}} \times \frac{s_A^2}{s_B^2} \right[$$

logo o Intervalo de confiança a 90% para $\frac{\sigma_A^2}{\sigma_B^2}$ é

$$]0.1873576; 1.5027114[$$

Como $\frac{\sigma_A^2}{\sigma_B^2} = 1$ pertence ao intervalo encontrado, para um grau de confiança de 90% e com base nas amostras dadas, pode-se considerar que a variabilidade da velocidade do vento nas duas localizações é idêntica.

(f) Pretende-se testar

$$\begin{cases} H_0 : \mu_A = \mu_B \\ contra \\ H_1 : \mu_A < \mu_B \end{cases} \Leftrightarrow \begin{cases} H_0 : \mu_A - \mu_B = 0 \\ contra \\ H_1 : \mu_A - \mu_B < 0 \end{cases}$$

Nível de significância: $\alpha = 0.10$

Tipo de teste: o teste é unilateral esquerdo

Estatística de Teste:

$$\left.\begin{array}{c} \text{Populações Normais (*)} \\ \sigma_A, \sigma_B \text{ desconhecidos} \\ \sigma_A = \sigma_B \text{ (**)} \\ \text{Amostras independentes} \end{array}\right\} T = \frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \sim t_{(n_A - n_B - 2)} \Leftrightarrow T \sim t_{(22)}$$

(*) As Populações podem ser consideradas Normais pois na alínea (*a*). vimos que para $\alpha = 0.10$ não se rejeita a hipótese de Normalidade.

(**) Os desvios padrão podem ser considerados iguais, pela alínea (e).

Tomada de Decisão pelo valor-p: Como valor-p$= 0.652 > 0.10 = \alpha$, então Não se Rejeita H_0.

Conclusão: Com 10% de significância e com base nas amostras, conclui-se que não existe evidência estatística que, em média, a velocidade do vento na localização A não seja inferior à velocidade média do vento na localidade B, pelo que a localização B não apresenta melhores condições para a instalação da central.

2. (a) Seja $X-$ quantidade de açúcar existente nos cereais de pequeno almoço

Hipóteses: $H_0 : X \sim Exp(8)$ contra $H_1 : X \not\sim Exp(8)$

Escolha do Teste: Como se pretende-se testar se X se comporta de acordo com uma Distribuição Exponencial de média 8 (uma distribuição contínua e completamente especificada) vamos recorrer ao teste de ajustamento de Kolmogorov-Smirnov.

Tomada de Decisão: Como valor$-p = 0.0452 \leq \alpha = 0.06$, então Rejeita-se H_0.

Conclusão: Com base na amostra e para um nível de significância de 6%, não há evidência estatística que a quantidade de açúcar existente nos cereais de pequeno almoço segue uma distribuição Exponencial de média 8 gramas.

(b) Populações:

$X_I =$ quantidade de açúcar nos cereais de pequeno almoço destinados aos consumidores infantis

$X_A =$ quantidade de açúcar nos cereais de pequeno almoço destinados aos consumidores adultos

Amostras: amostras aleatórias independentes

Escolha do teste: Como pretende-se um teste de hipóteses não paramétrico, então vamos recorrer ao testes de Mann-Whitney pois as amostras são independentes.

Hipóteses:

$$\begin{cases} H_0 : Mediana_{X_I} \leq Mediana_{X_A} & \text{os cereais de pequeno almoço destinados ao consumidor infantil} \\ & \text{não contêm mais açúcar do que os destinados a adultos} \\ \text{contra} & \\ H_1 : Mediana_{X_I} > Mediana_{X_A} & \text{os cereais de pequeno almoço destinados ao consumidor infantil} \\ & \text{contêm mais açúcar do que os destinados a adultos} \end{cases}$$

$$\Leftrightarrow \quad H_0 : Mediana_{X_I} - Mediana_{X_A} \leq 0 \quad \text{contra} \quad H_1 : Mediana_{X_I} - Mediana_{X_A} > 0$$

Tipo de teste: o teste é unilateral direito

Tomada de decisão: se valor-p$\leq \alpha$, então Rejeita-se H_0, ou seja, a suspeita pode ser considerada verdadeira.

Como valor-p$= 0.0151$, então a partir de $\alpha \geq 0.0151$ Rejeita-se H_0, ou seja, a partir de $\alpha \geq 0.0151$ há evidência estatística que a suspeita é verdadeira.

(c) Seja $p =$ a proporção de cereais que estão na prateleira mais alta, uma estimativa pontual é

$$p^* = \frac{\text{número de cereais na prateleira mais alta na amostra}}{\text{número de cereais na amostra}} = \frac{36}{77} = 0.4675$$

(d) Populações:

$p_1 =$ proporção de cereais que estão na prateleira mais alta num hipermercado

$p_2 =$ proporção de cereais que estão na prateleira mais alta noutro hipermercado

Amostras Aleatórias:

amostras de dimensões $n_1 = 77$ e $n_2 = 96$

amostras aleatórias independentes

Escolha do Intervalo de confiança:

Como temos populações binomiais e $n_1 = 77 \geq 30$ e $n_2 = 96 \geq 30$ o intervalo de confiança a $(1-\alpha) \times 100\%$ para $p_1 - p_2$ é dado por:

$$\left]\left(p_1^* - p_2^*\right) - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}{n_2}}; \left(p_1^* - p_2^*\right) + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}{n_2}}\right[$$

logo, o intervalo de confiança a 97% para $p_1 - p_2$ é:

$$]-0.0493, 0.2760[$$

Como $0 \in]-0.0493, 0.2760[$, então, com 97% de confiança há a possibilidade da percentagem de embalagens de cereais na prateleira mais alta ser idêntica em ambos os hipermercados.

(e) População:

p_{3I} = proporção de cereais na prateleira mais alta destinados aos consumidores infantis

Amostra aleatória:

dimensão da amostras $n_3 = 36$

Hipóteses:

$$\begin{cases} H_0 : p_{3I} \geq 0.10 \longrightarrow & \text{na prateleira mais alta, a percentagem de embalagens destinadas} \\ & \text{ao consumidor infantil não é inferior a 10\%} \\ \text{contra} & \\ H_1 : p_{3I} < 0.10 \longrightarrow & \text{na prateleira mais alta, a percentagem de embalagens destinadas} \\ & \text{ao consumidor infantil é inferior a 10\%} \end{cases}$$

Nível de significância $= \alpha = 0.04$

Tipo de teste: o teste é Unilateral Esquerdo

Estatística de Teste: Como temos uma população binomial e $n_3 = 36 \geq 30$, a estatística de teste é:

$$Z = \frac{p^* - p}{\sqrt{\frac{pq}{n}}} \dot{\sim} N(0,1)$$

Tomada de Decisão pelo valor-p: Como valor-p $= 0.7817 > 0.04 = \alpha$, então Não se Rejeita H_0

OU

Tomada de Decisão pela Região crítica:

$$RC =]-\infty, z_\alpha] =]-\infty, z_{0.04}] =]-\infty, -1.7507]$$

Como $Z_{obs} = 0.7778 \notin RC$, então Não se Rejeita H_0.

Conclusão: Com base na amostra e para um nível de significância de 4%, não há evidência estatística que este hipermercado esteja a cumprir a orientação dos responsáveis.

(f) Variáveis:

shelf - variável qualitativa ordinal

client - variável qualitativa nominal

Escolha do Teste:

Como as variáveis são qualitativas e pretende-se verificar se as variáveis estão associadas, vamos recorrer ao teste de independência do Qui-Quadrado.

Hipóteses a testar:

H_0 : Não existe associação entre a prateleira onde se encontra a embalagem de cereais e o consumidor a quem se destina, ou seja, as variáveis shelf e client são independentes

contra

H_1 : Existe associação entre a prateleira onde se encontra a embalagem de cereais e o consumidor a quem se destina, ou seja, as variáveis shelf e client não são independentes

nível de significância $= \alpha = 0.06$

Tabela de contingência entre shelf e client:

		Consumidor		
		Adulto	Infantil	Total
prateleira	mais baixa	15	5	20
	do meio	6	15	21
	mais alta	31	5	36
	Total	52	25	77

a tabela de contingência tem $r = 3$ linhas e $c = 2$ colunas

Estatística de Teste:

$$Q \sim \chi^2_{(r-1)\times(c-1)} \Leftrightarrow Q \sim \chi^2_{(2)}$$

Tomada de Decisão pelo valor-p: Como valor-p $= 0.00003 \leq 0.06 = \alpha$, então Rejeita-se H_0.

<u>OU</u>

Tomada de Decisão pela Região crítica:

$$RC = [x^2_{(r-1)\times(c-1);1-\alpha}, +\infty[= [x^2_{(2);1-0.06}, +\infty[= [5.6268, +\infty[$$

Como $Q_{obs} = 20.7143 \in RC$, então Rejeita-se H_0.

Conclusão: Com base nas amostras e para um nível de significância de 6%, há evidência estatística que existe associação entre a prateleira onde se encontra a embalagem de cereais e o consumidor a quem se destina.

Como existe associação entre as variáveis, então vamos medir essa associação. Como uma das variáveis é qualitativa nominal, então só é possível calcular o coeficiente de contingência e o coeficiente V de Crámer:

- coeficiente de contingência $= 0.4604 \in [0.30, 0.50[\quad \mapsto$ associação moderada
- coeficiente V de Crámer $= 0.5186 \geq 0.50 \quad (k = 2) \quad \mapsto$ associação elevada

A associação existente pode ser considerada entre moderada e elevada.