DEPARTAMENTO DE MATEMÁTICA
ESTATÍSTICA PARA A SAÚDE
1.º SEMESTRE - 2023/2024
2.º Teste (recuperação)

Data: 14 de fevereiro de 2024 Duração: 2 horas

- Parte do teste foi resolvido recorrendo ao software R: ver script_2teste_recup_ES_23_24.R.

Resolução

1. (a) Foi construído um intervalo de confiança para μ = tempo médio que os pacientes praticam atividade física, pretende-se determinar o grau de confiança $= 1-\alpha$ e sabe-se que o limite inferior desse intervalo é 5.2 horas/semana. A dimensão da amostra recolhida é $n = 47$.

Escolha do Intervalo de confiança: População qualquer, σ desconhecido e $n = 47 \geq 30$, então o Intervalo de confiança a $(1-\alpha)\times 100\%$ para μ é

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

Com base na amostra recolhida tem-se:

$$\left]5.7362 - z_{1-\frac{\alpha}{2}} \times \frac{2.3882}{\sqrt{47}}, 5.7362 + z_{1-\frac{\alpha}{2}} \times \frac{2.3882}{\sqrt{47}}\right[$$

Portanto tem-se

$$5.7362 - z_{1-\frac{\alpha}{2}} \times \frac{2.3882}{\sqrt{47}} = 5.2 \Leftrightarrow z_{1-\frac{\alpha}{2}} = 1.5392 \Leftrightarrow 1-\frac{\alpha}{2} = \Phi(1.5392) \Leftrightarrow$$

$$\Leftrightarrow 1-\frac{\alpha}{2} = 0.9381 \Leftrightarrow \alpha = 0.1238 \Leftrightarrow 1-\alpha = 0.8762$$

(b) Um estimador pontual para μ_f = tempo de uso das próteses pelos pacientes do género feminino é $\bar{X}_f$ = média amostral do tempo de uso das próteses pelos pacientes do género feminino e uma sua estimativa pontual é

$$\bar{x}_f = 23.0739 \text{ meses}$$

Um estimador pontual para σ_f^2 = variância do tempo de uso das próteses pelos pacientes do género feminino é S_f^2 = variância amostral do tempo de uso das próteses pelos pacientes do género feminino e uma sua estimativa pontual é

$$s_f^2 = 67.4229 \text{ meses}^2$$

Um estimador pontual para $p_f \times 100\%$ = percentagem de pacientes do género feminino com amputação transtibial é $\text{p}_f^* \times 100\%$ = percentagem amostral de pacientes do género feminino com amputação transtibial e uma sua estimativa pontual é

$$p_f^* \times 100\% = 26.087\% \text{ pacientes femininos}$$

(c) População: X_f = tempo de uso das próteses pelos pacientes do género feminino

Amostra aleatória de dimensão $n_f = 23$

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

Escolha do Teste: Como pretende-se testar se X_f comporta-se de acordo com uma Distribuição Normal (sem mais informação) e $n_x = 23 < 50$, vamos recorrer ao teste de ajustamento Shapiro-Wilk.

Tomada de Decisão: Como valor$-p = 0.8002 > \alpha = 0.10$, então Não se Rejeita H_0.

Conclusão: Com base na amostra e para um nível de significância de 10%, há evidência estatística que o tempo de uso das próteses pelos pacientes do género feminino segue uma distribuição Normal.

População: $X_m =$ tempo de uso das próteses pelos pacientes do género masculino

Amostra aleatória de dimensão $n_m = 24$

Hipóteses: $H_0 : X_m \sim Normal$ contra $H_1 : X_m \not\sim Normal$

Escolha do Teste: Como pretende-se testar se X_m comporta-se de acordo com uma Distribuição Normal (sem mais informação) e $n_m = 24 < 50$, vamos recorrer ao teste de ajustamento Shapiro-Wilk.

Tomada de Decisão: Como valor$-p = 0.9148 > \alpha = 0.10$, então Não se Rejeita H_0.

Conclusão: Com base na amostra e para um nível de significância de 10%, há evidência estatística que o tempo de uso das próteses pelos pacientes do género masculino segue uma distribuição Normal.

(d) Pela alínea anterior é possível considerar que as Populações são Normais, então o Intervalo de confiança a $(1-\alpha) \times 100\%$ para $\frac{\sigma_f^2}{\sigma_m^2}$ é

$$\left] \frac{1}{f_{1-\frac{\alpha}{2};n_f-1;n_m-1}} \times \frac{s_f^2}{s_m^2}, \frac{1}{f_{\frac{\alpha}{2};n_f-1;n_m-1}} \times \frac{s_f^2}{s_m^2} \right[$$

grau de confiança $= 1 - \alpha = 0.90$

nível de significância $= \alpha = 0.10$

Portanto, com base nas amostras recolhidas, o intervalo de confiança a 90% para o quociente de variâncias, $\frac{\sigma_f^2}{\sigma_m^2}$, é:

$$]0.4194, 1.7301[$$

Como $1 \in]0.4194, 1.7301[$, então, com 90% de confiança a variabilidade dos tempos de uso das próteses pode ser considerado igual quando os pacientes são separados por género.

(e) Hipóteses:

$$\begin{cases} H_0 : \mu_m \geq \mu_f \rightarrow & \text{o tempo de uso médio das próteses não é inferior no género masculino quando comparado com o género feminino} \\ \text{contra} & \\ H_1 : \mu_m < \mu_f \rightarrow & \text{o tempo de uso médio das próteses é inferior no género masculino quando comparado com o género feminino} \end{cases}$$

$$\begin{cases} H_0 : \mu_m \geq \mu_f \\ vs \\ H_1 : \mu_m < \mu_f \end{cases} \Leftrightarrow \begin{cases} H_0 : \mu_m - \mu_f \geq 0 \\ vs \\ H_1 : \mu_m - \mu_f < 0 \end{cases}$$

nível de significância $= \alpha = 0.10$

Tipo de teste: Teste Unilateral Esquerdo

Estatística de Teste:

$$\left.\begin{array}{c} \text{Populações Normais (alínea (c))} \\ \sigma_m \text{ e } \sigma_f \text{ desconhecidos} \\ \sigma_m = \sigma_f \text{ (alínea (d))} \\ \text{amostras aleatórias independentes} \end{array}\right\} T = \frac{\left(\overline{X}_m - \overline{X}_f\right) - \left(\mu_m - \mu_f\right)}{\sqrt{\left(\frac{1}{n_m} + \frac{1}{n_f}\right) \times \frac{(n_m-1)S_m^2 + \left(n_f-1\right)S_f^2}{n_m+n_f-2}}} \sim t_{(n_m+n_f-2)}$$

Como $n_m + n_f - 2 = 24 + 23 - 2 = 45$, tem-se $T \sim t_{(45)}$

Tomada de decisão pela Região Critica:

$$RC = \left]-\infty, t_{\alpha, n_m+n_f-2}\right] = \left]-\infty, t_{0.10,45}\right] = \left]-\infty, -1.3006\right]$$

Como $T_{obs} = 0.5649 \notin RC$, então não se rejeita H_0.

Conclusão: Com base nas amostras e para um nível de significância de 10%, há evidência estatística que o tempo de uso médio das próteses não é inferior no género masculino quando comparado com o género feminino.

(f) Interpretação:

$$\begin{cases} H_0 : p = 0.50 \rightarrow & \text{a proporção populacional de amputados transtibial é 0.50} \\ \text{contra} & \\ H_1 : p = 0.60 \rightarrow & \text{a proporção populacional de amputados transtibial é 0.60} \end{cases}$$

Hipóteses:

$$\begin{cases} H_0 : p = 0.50 \\ vs \\ H_1 : p = 0.60 \end{cases} \Leftrightarrow \begin{cases} H_0 : p = 0.50 \\ vs \\ H_1 : p > 0.50 \end{cases}$$

nível de significância $= \alpha = 0.07$

Tipo de teste: Teste Unilateral Direito

Estatística de Teste:

$$\left.\begin{array}{c} \text{População Binomial} \\ n = 47 \geq 30 \end{array}\right\} Z = \frac{p^* - p}{\sqrt{\frac{pq}{n}}} \dot{\sim} N(0,1)$$

Tomada de decisão pelo valor$-p$:

Como valor$-p = 0.8464 > \alpha = 0.07$, então não se rejeita H_0.

OU

Tomada de decisão pela Região crítica:

$$RC = [z_{1-\alpha}, +\infty[= [z_{1-0.07}, +\infty[= [1.4758, +\infty[$$

Como $Z_{obs} = -1.0211 \notin RC$, então não se rejeita H_0.

Conclusão: Com base na amostra e para um nível de significância de 7%, há evidência estatística que a proporção populacional de amputados transtibial é 0.50.

2. População: X = peso, em gramas, dos recém-nascidos, $X \sim N(3500, 50)$ pois $\mu = E[X] = 3500$ gramas e $\sigma = \sqrt{V[X]} = \sqrt{2500} = 50$ gramas.

   Um pediatra suspeita que a variabilidade do peso dos recém-nascidos é superior ao valor indicado: $\sigma^2 > 2500$.

   Amostra aleatória: dimensão $n = 6$ recém-nascidos.

   (a) Pretende-se calcular uma probabilidade sobre a média amostral, então é necessário determinar a sua distribuição amostral: População Normal e $\sigma = 50$ conhecido, logo $Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$.

$$P\left(\bar{X} > \mu + 20\right) = P\left(Z > \frac{\mu + 20 - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(Z > \frac{20}{\frac{50}{\sqrt{6}}}\right) = P(Z > 0.9798) =$$
$$= 1 - P(Z \leq 0.9798) = 1 - \Phi(0.9798) = 0.1636$$

(b) Já sabemos $Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ e agora pretende-se determinar n tal que:

$$P\left(\bar{X} \geq 3490\right) = 0.9 \Leftrightarrow P\left(Z \geq \frac{3490-3500}{\frac{50}{\sqrt{n}}}\right) = 0.9 \Leftrightarrow P\left(Z \geq -\frac{\sqrt{n}}{5}\right) = 0.9 \Leftrightarrow$$

$$\Leftrightarrow 1 - P\left(Z < -\frac{\sqrt{n}}{5}\right) = 0.9 \underset{\text{v.a. contínua}}{\Leftrightarrow} 1 - \Phi\left(-\frac{\sqrt{n}}{5}\right) = 0.9 \Leftrightarrow \Phi\left(-\frac{\sqrt{n}}{5}\right) = 0.1 \Leftrightarrow$$

$$\Leftrightarrow -\frac{\sqrt{n}}{5} = z_{0.1} \Leftrightarrow -\frac{\sqrt{n}}{5} = -1.2816 \Leftrightarrow n = 41.0594 \Rightarrow n = 42 \text{ recém-nascidos}$$

(c) Hipóteses:

$$\begin{cases} H_0 : \sigma^2 = 2500 \rightarrow & \text{o pediatra não tem razão} \\ \text{contra} & \\ H_1 : \sigma^2 > 2500 \rightarrow & \text{o pediatra tem razão} \end{cases}$$

Tipo de teste: Teste Unilateral Direito

Estatística de Teste:

$$\text{População Normal:} \quad X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)} \Leftrightarrow X^2 \sim \chi^2_{(5)}$$

Tomada de decisão: valor$-p = 0.1906$

Pretende-se determinar o nível de significância a partir do qual a suspeita do pediatra está correta, ou seja, pretende-se determinar o nível de significância a partir do qual a hipótese H_0 é rejeitada. Como rejeita-se H_0 se valor$-p \leq \alpha$, então para $\alpha \geq 0.1906$ rejeita-se a hipótese nula. Portanto, para $\alpha \geq 0.1906$ a suspeita do pediatra está correta.

(d) População 1: Y_1 = peso ganho (em gramas) pelos bebés amamentados exclusivamente com leite materno.

População 2: Y_2 = peso ganho (em gramas) pelos bebés amamentados parcialmente com leite materno.

Amostras aleatórias independentes, com $n_1 = 5$ e $n_2 = 5$

i. Hipóteses: $H_0 : Y_1 \sim Exp(2000)$ contra $H_1 : Y_1 \not\sim Exp(2000)$

Escolha do Teste: Como pretende-se testar se Y_1 comporta-se de acordo com uma Distribuição Exponencial de média 2000, que é uma distribuição de probabilidade contínua e está completamente especificada, vamos recorrer ao teste de ajustamento Kolmogorov-Smirnov.

Tomada de Decisão: Como valor$-p = 0.1581 > \alpha = 0.05$, então Não se Rejeita H_0.

Conclusão: Com base na amostra e para um nível de significância de 5%, há evidência estatística que o peso ganho, em relação aos bebés amamentados exclusivamente com leite materno, pode ser modelado por uma distribuição exponencial de média 2000 gramas.

ii. Como as amostras têm dimensão inferior a 30 e na alínea (d)i. vimos que há dados que podem ser modelados por uma distribuição exponencial (logo não é possível considerar que seguem uma distribuição Normal), então não é possível utilizar testes de hipóteses paramétricos. Para ultrapassar estes problemas vamos recorrer aos testes de hipóteses não paramétricos, neste caso ao teste de Mann-Whitney pois as amostras são independentes. Seja $Mediana_1$ a mediana referente à população Y_1 e seja $Mediana_2$ a mediana referente à população Y_2.

Hipóteses a testar:

$$\begin{cases} H_0 : Mediana_1 = Mediana_2 \rightarrow & \text{ganhos de peso iguais com os dois tipos de amamentação} \\ vs & \\ H_1 : Mediana_1 \neq Mediana_2 \rightarrow & \text{ganhos de peso diferentes com os dois tipos de amamentação} \end{cases}$$

$$\begin{cases} H_0 : Mediana_1 = Mediana_2 \\ vs \\ H_1 : Mediana_1 \neq Mediana_2 \end{cases} \Leftrightarrow \begin{cases} H_0 : Mediana_1 - Mediana_2 = 0 \\ vs \\ H_1 : Mediana_1 - Mediana_2 \neq 0 \end{cases}$$

Tipo de teste: Teste Bilateral

nível de significância $= \alpha = 0.05$

Tomada de decisão: Como valor$-p = 0.5476 > \alpha = 0.05$, Não se rejeita H_0.

Com 5% de significância e com base nas amostras, não há evidência estatística que os ganhos de peso sejam diferentes quando os bebés são exclusivamente amamentados com leite materno e quando são parcialmente amamentados com leite materno.

Fim do Teste