2º Teste Recuperação

1. Uma empresa envolvida no desenvolvimento de próteses de membros inferiores pretende personalizar soluções e contribuir significativamente para a melhoria da qualidade de vida de indivíduos amputados. Para tal pretende explorar fatores biomédicos e robóticos que possam influenciar o desempenho e a satisfação dos utilizadores de próteses de membros inferiores. Para atingir este objetivo recolheu uma amostra de alguns pacientes amputados que utilizam próteses de membros inferiores, cuja informação foi colocada no ficheiro "ProtesesMI_T2.txt"que se encontra no Moodle, e tem os seguintes campos:
   - paciente = código de identificação do paciente
   - genero = Género do paciente (Feminino, Masculino)
   - amputacao = paciente com amputação Transtibial (0 = Não, 1 = Sim)
   - tempo = Tempo de uso da prótese (em meses)
   - Afisica = Tempo que pratica atividade física (em horas/semana)

[2.0] (a) Para estimar o tempo médio que os pacientes praticam atividade física foi construido um intervalo de confiança cujo limite inferior é 5.2 horas/semana. Qual o grau de confiança atribuído a este intervalo?

[1.5] (b) Indique estimadores pontuais e calcule as respetivas estimativas para:
- a média e variância do tempo de uso das próteses pelos pacientes do género feminino;
- a percentagem de pacientes do género feminino com amputação transtibial.

[1.5] (c) Mostre que, 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 e o tempo de uso das próteses pelos pacientes do género masculino também segue uma distribuição Normal.

[2.0] (d) Com base num intervalo de confiança a 90% verifique se a variabilidade dos tempos de uso das próteses pode ser considerado igual quando os pacientes são separados por género.

[2.0] (e) Para um nível de significância de 10% e com base na região crítica, verifique se há evidência estatística que o tempo de uso médio das próteses é inferior no género masculino quando comparado com o género feminino.

[2.0] (f) Seja p a proporção de amputados transtibial. Interprete e teste, para um nível de significância de 7%, as seguintes hipóteses:

$$H_0 : p = 0.50 \qquad \text{contra} \qquad H_1 : p = 0.60$$

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 $\mathrm{p}_f^* \times 100\%$ = 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 $H_1 : X_m \nsim Normal$

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:

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

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

Tomada de decisão pela Região Critica:

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

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. Em determinada região sabe-se que o peso dos recém-nascidos segue uma distribuição normal de média 3500 gramas e variância 2500 gramas^2. Devido às condições socioeconómicas das pessoas que procuram o hospital da região, um pediatra suspeita que a variabilidade do peso das crianças recém-nascidas nesse hospital é superior ao valor indicado. A equipa de pediatras desse hospital pesou 6 recém-nascidos, escolhidos aleatoriamente.

[2.0] (a) Qual é a probabilidade da média amostral ser superior à média populacional em mais de 20 gramas?

[2.0] (b) Qual o número de recém-nascidos que a equipa de pediatras deveria ter pesado, caso pretendesse que o peso médio dos recém-nascidos pesados fosse no mínimo de 3490 gramas com probabilidade 0.9?

[2.0] (c) Sabendo que a amostra recolhida foi a seguinte:

3599.000, 3495.655, 3433.602, 3477.521, 3541.972, 3454.400

Diga a partir de que nível de significância a suspeita do pediatra está correta.

(d) Pretende-se saber se existem diferenças significativas entre os pesos dos bebés amamentados exclusivamente com leite materno e só amamentados parcialmente com leite materno. Para tal selecionaram-se de forma aleatória 5 recém-nascidos que durante os primeiros 3 meses foram alimentados exclusivamente de leite materno e selecionaram-se aleatoriamente outros 5 recém-nascidos que durante o mesmo período de tempo foram alimentados parcialmente com leite materno. No fim dos três meses a diferença (em gramas) entre o peso final dos bebés e o peso inicial, ou seja, o peso ganho (em gramas) pelos bebés foi registado na tabela seguinte:

peso ganho se amamentado exclusivamente com leite materno	2166.13	2235.77	2324.14	2168.85	1268.05
peso ganho se amamentado parcialmente com leite materno	2831.39	1838.87	1836.51	1800.39	1776.31

[1.0] i. Mostre que, em relação aos bebés amamentados exclusivamente com leite materno, o peso ganho pode ser modelado por uma distribuição exponencial de média 2000 gramas. Considere um nível de significância de 5%.

[2.0] ii. Teste, para um nível de significância de 5%, se há evidência estatística para considerar que os ganhos de peso são diferentes quando os bebés são exclusivamente amamentados com leite materno e quando são parcialmente amamentados com leite materno.

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

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

(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 $H_1 : Y_1 \nsim Exp(2000)$

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