Lista 05 de Estatística II (aula 05)

1) Pesquisas apontam que o colesterol elevado é um fator de risco cardíaco. Um estudo comparou o nível de colesterol entre homens e mulheres entre 20 e 29 anos.

Homens		Mulheres
n1 = 24		n2 = 31
x1 = 167,16		x2 = 178,12
S1 = 30		S2 = 32

Usando $\alpha = 0.05$, determine se existe diferença significativa entre os níveis de colesterol entre homens e mulheres?

2) Pretende-se comparar a adesão a certificação de alimentos de acordo a Norma Global para a Segurança em Alimentos(de acordo com os requisitos do Codex Alimentarius, Programa Conjunto da Organização das Nações Unidas para a Agricultura e a Alimentação -FAO e da Organização Mundial da Saúde - OMS) entre as empresas de alimentos dos Estados Unidos e do Brasil. Selecionaram-se aleatoriamente 2500 empresas dos EUA e 1000 do Brasil, tendo-se observado que, das 2500 analisadas nos EUA, 1500 atendiam às normas exigidas, enquanto que das 1000 do Brasil, 500 às atendiam. A) Haverá razões para acreditar em diferentes níveis de adesão a estas normas nos dois países? B) Cite quais passos mudam caso se realizasse um teste monocaudal.

1) Pode existir ou não uma diferença significativa entre os níveis de colesterol entre homens e mulheres. Se os níveis de colesterol entre homens e mulheres forem iguais, não existe diferença entre eles (hipótese nula). Porém, se forem diferentes, existe diferença entre eles (hipótese alternativa).

As hipóteses são:

$$\begin{cases} H_0: \mu_H = \mu_M \;\; ou \;\; \mu_H - \mu_M = 0 \\ H_1: \mu_H \neq \mu_M \;\; ou \;\; \mu_H - \mu_M \neq 0 \end{cases}$$

Sob a hipótese nula, a estatística de teste é:

$$t = \frac{(\bar{X}_H - \bar{X}_M) - (\mu_H - \mu_M)}{\sqrt{\frac{S^2}{n_H} + \frac{S^2}{n_M}}} \sim t_{(n_H + n_M - 2)}$$

Onde:

$$S^2 = \frac{(n_H - 1) * S_H^2 + (n_M - 1) * S_M^2}{n_H + n_M - 2}$$

Dados os valores das amostras,

$$S^2 = \frac{(24-1)*(30)^2 \;+(31-1)*(32)^2}{24+31-2} = 970{,}19$$

$$t_{obs} = \frac{(167{,}16-178{,}12)-(0)}{\sqrt{\frac{970{,}19}{24}+\frac{970{,}19}{31}}} = -1{,}294 \quad \sim t_{(53)}$$

Como o teste é bilateral, o nível de significância é 0,05 e há 53 graus de liberdade, então da tabela de distribuição t de Student temos que a região crítica é:

$$RC = (-\infty; -2{,}006] \cup [2{,}006; +\infty)$$

Como o valor observado da estatística de teste, t_{obs}, não pertence à região crítica, então a evidência amostral indica que não se deve rejeitar a hipótese nula, ou seja, que os níveis de colesterol entre homens e mulheres são estatisticamente iguais (não existe diferença significativa entre eles).

2) a) Nível de adesão indica a proporção de firmas que atendem às normas exigidas em cada país. Assim, o que queremos saber é se os níveis de adesão nos EUA (p_{US}) e no Brasil (p_{BR}) são iguais, ou seja, se

$$\begin{cases} H_0: p_{US} = p_{BR} \;\; ou \;\; p_{US} - p_{BR} = 0 \\ H_1: p_{US} \neq p_{SP} \;\; ou \;\; p_{US} - p_{BR} \neq 0 \end{cases}$$

As amostras de firmas de São Paulo e Rio de Janeiro indicam as seguintes proporções amostrais de aprovação: $\hat{p}_{US} = \frac{1500}{2500} = 0{,}6 \;\; e \;\; \hat{p}_{BR} = \frac{500}{1000} = 0{,}5.$

No teste sobre a diferença de proporções, como a hipótese nula é que as proporções são iguais, devemos calcular a proporção comum: (Livro Morettin)

$$\hat{p} = \frac{n_{US} * \hat{p}_{US} + n_{BR} * \hat{p}_{BR}}{n_{US} + n_{BR}} = 0{,}5714$$

Sob a hipótese nula, a estatística de teste é:

$$Z = \frac{(\hat{p}_{US} - \hat{p}_{BR}) - (\hat{p}_{US} - \hat{p}_{BR})}{\sqrt{\hat{p} * (1-\hat{p}) * \left(\frac{1}{n_{US}} + \frac{1}{n_{BR}}\right)}} \sim N(0;1)$$

Como não dispomos dos valores das proporções populacionais, então esta estatística é calculada por:

$$Z = \frac{(p_{US} - p_{BR}) - (\pi_{US} - \pi_{BR})}{\sqrt{\hat{p} * (1 - \hat{p}) * \left(\frac{1}{n_{US}} + \frac{1}{n_{BR}}\right)}} \sim N(0;1)$$

Cujo valor observado é:

$$Z_{obs} = \frac{(0{,}6 - 0{,}5) - (0)}{\sqrt{0{,}5714 * (1 - 0{,}5714) * \left(\frac{1}{2500} + \frac{1}{1000}\right)}} = 5{,}40.$$

Ou (Livro Newbold et al.)

$$\boxed{z = \frac{(\hat{p}_x - \hat{p}_y)}{\sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n_x} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_y}}}}$$

E

$$\boxed{\hat{p}_0 = \frac{n_x\hat{p}_x + n_y\hat{p}_y}{n_x + n_y}}$$

Ao nível de significância (α) de 5%, a região crítica do teste seria: $(-\infty;-1{,}96]\cup[1{,}96;+\infty)$. Como o valor observado da estatística de teste, Z_{obs}, pertence à região crítica, então há evidência que indique que se deva rejeitar a hipótese (nula) de que os níveis de adesão nos EUA(p_{US}) e no Brasil (p_{BR}) são diferentes.

b) Caso fosse realizado um teste monocaudal, o primeiro aspecto que mudaria é a definição das hipóteses testadas.

Outro aspecto que mudaria é a forma de definição da região crítica, que passaria a ser unilateral. Especificamente, o valor limite da região crítica seria determinado ao se alocar todo nível de significância para uma das caudas da distribuição. Consequentemente, o valor-p do valor observado da estatística de teste seria definido para apenas uma das caudas da distribuição.