Lista 5 - Exercício 1a

Lucas e Nicolas

20/11/2019

TRV e IC para comparações de grupos

Exercício 1

Seja $\mathbf{X} = (X_1, ..., X_n)$ uma a.a. de $X \sim Normal(\mu_X, \sigma_X^2)$ e $\mathbf{Y} = (Y_1, ..., Y_m)$ uma a.a. de $Y \sim Normal(\mu_Y, \sigma_Y^2)$, tal que $\mathbf{X}$ e $\mathbf{Y}$ são independentes. Encontre o TRV para testar

(a)

$H_0 : \mu_X = \mu_Y$ contra $H_1 : \mu_X \neq \mu_Y$ assumindo que $\sigma_X^2 = \sigma_Y^2 = \sigma^2$

Solução:

Lembrando que

- Distribuição Normal com média μ e variância σ^2:

$$f(t) = (2\pi\sigma^2)^{-\frac{1}{2}} exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)$$

- Verossimilhança da $Normal(\mu, \sigma^2)$:

$$L(\theta) = \prod_{i=1}^{n}\left[(2\pi\sigma^2)^{-\frac{1}{2}} exp\left(-\frac{(t_i-\mu)^2}{2\sigma^2}\right)\right] = (2\pi\sigma^2)^{-\frac{n}{2}} exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(t_i-\mu)^2\right)$$

- TRV:

$$\lambda(\underset{\sim}{x}) = \frac{sup_{\theta\in\Theta_0} L(\theta)}{sup_{\theta\in\Theta} L(\theta)} = \frac{L(\hat{\theta_0})}{L(\hat{\theta})}$$

Com os dados do exercício, temos

- Identificando os espaços paramétricos (lembrando que $\Theta = \Theta_0 \cup \Theta_1$):

$$\Theta_0 = (\theta = (\mu_X, \mu_Y, \sigma) : \mu_X = \mu_Y = \mu_0 \in \mathbb{R},\ \sigma > 0)$$

$$\Theta = (\theta = (\mu_X, \mu_Y, \sigma) : \mu_X, \mu_Y \in \mathbb{R},\ \sigma > 0)$$

- Verossimilhança para as amostras independentes $\underset{\sim}{x}$ e $\underset{\sim}{y}$:

$$L(\theta|\underset{\sim}{x}, \underset{\sim}{y}) = (2\pi\sigma^2)^{-\frac{n+m}{2}} exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_X)^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{m}(y_j-\mu_Y)^2\right)$$

- TRV:

$$\lambda(\underset{\sim}{x}, \underset{\sim}{y}) = \frac{sup_{\theta \in \Theta_0} L(\theta)}{sup_{\theta \in \Theta} L(\theta)} = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$$

<u>Calculando $L(\hat{\theta}_0)$:</u>

$$\hat{\theta}_{\mathbf{0}} = (\hat{\mu}_0, \hat{\mu}_0, \hat{\sigma}_0^2)$$

$$log(L(\theta_0)) = -\frac{n+m}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu_0)^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{m}(y_j - \mu_0)^2$$

Para $\hat{\mu}_0$:

$$\frac{\partial L(\theta_0)}{\partial \mu_0} = 0 \Leftrightarrow \sum_{i=1}^{n}(x_i - \mu_0) + \sum_{j=1}^{m}(y_j - \mu_0) = 0 \Leftrightarrow \sum_{i=1}^{n} x_i + \sum_{i=1}^{m} y_j = (n+m)\mu_0$$

Logo,

$$\hat{\mu}_0 = \frac{\sum_{i=1}^{n} x_i + \sum_{i=1}^{m} y_j}{n+m}$$

Para $\hat{\sigma}_0^2$:

$$\frac{\partial L(\theta_0)}{\partial \sigma_0^2} = 0 \Leftrightarrow -\left(\frac{n+m}{2}\right)\frac{1}{\sigma_0^2} - \frac{1}{2(\sigma_0^2)^2}\left(\sum_{i=1}^{n}(x_i - \hat{\mu}_0)^2 + \sum_{j=1}^{m}(y_j - \hat{\mu}_0)^2\right) = 0$$

$$\Leftrightarrow \left(\sum_{i=1}^{n}(x_i - \hat{\mu}_0)^2 + \sum_{j=1}^{m}(y_j - \hat{\mu}_0)^2\right)\frac{1}{\sigma_0^2} = (n+m)$$

Logo,

$$\hat{\sigma}_0^2 = \frac{\left(\sum_{i=1}^{n}(x_i - \hat{\mu}_0)^2 + \sum_{j=1}^{m}(y_j - \hat{\mu}_0)^2\right)}{n+m}$$

Finalmente,

$$L(\hat{\theta}_{\mathbf{0}}|\underset{\sim}{x}, \underset{\sim}{y}) = (2\pi\hat{\sigma}_0^2)^{-\frac{n+m}{2}} exp\left(-\frac{1}{2\hat{\sigma}_0^2}\left(\sum_{i=1}^{n}(x_i - \hat{\mu}_0)^2 + \sum_{j=1}^{m}(y_j - \hat{\mu}_0)^2\right)\right) = (2\pi\hat{\sigma}_0^2)^{-\frac{n+m}{2}} exp(-\frac{n+m}{2})$$

<u>Calculando $L(\hat{\theta})$:</u>

$$\hat{\theta} = (\hat{\mu}_X, \hat{\mu}_Y, \hat{\sigma}_0^2)$$

$$log(L(\theta)) = -\frac{n+m}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu_X)^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{m}(y_j - \mu_Y)^2$$

Para $\hat{\mu}_X$:

$$\frac{\partial L(\theta)}{\partial \mu_X} = 0 \Leftrightarrow \sum_{i=1}^{n}(x_i - \mu_0) = 0 \Leftrightarrow \sum_{i=1}^{n} x_i = n\mu_X$$

Logo,

$$\hat{\mu}_X = \frac{\sum_{i=1}^{n} x_i}{n}$$

Para $\hat{\mu}_Y$:

$$\frac{\partial L(\theta)}{\partial \mu_Y} = 0 \Leftrightarrow \sum_{j=1}^{m}(y_j - \mu_Y) = 0 \Leftrightarrow \sum_{j=1}^{m} y_j = m\mu_Y$$

Logo,

$$\hat{\mu}_Y = \frac{\sum_{i=1}^{m} y_j}{m}$$

Para $\hat{\sigma}^2$:

$$\frac{\partial L(\theta)}{\partial \sigma^2} = 0 \Leftrightarrow -\left(\frac{n+m}{2}\right)\frac{1}{\sigma^2} - \frac{1}{2(\sigma^2)^2}\left(\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2 + \sum_{j=1}^{m}(y_j - \hat{\mu}_Y)^2\right) = 0$$

$$\Leftrightarrow \left(\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2 + \sum_{j=1}^{m}(y_j - \hat{\mu}_Y)^2\right)\frac{1}{\sigma^2} = (n+m) \Leftrightarrow \hat{\sigma}^2 = \frac{\left(\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2 + \sum_{j=1}^{m}(y_j - \hat{\mu}_Y)^2\right)}{n+m}$$

Logo,

$$\hat{\sigma}^2 = \frac{\left(\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2 + \sum_{j=1}^{m}(y_j - \hat{\mu}_Y)^2\right)}{n+m}$$

Finalmente,

$$L(\hat{\theta}|\underset{\sim}{x}, \underset{\sim}{y}) = (2\pi\hat{\sigma}^2)^{-\frac{n+m}{2}} exp\left(-\frac{1}{2\hat{\sigma}^2}\left(\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2 + \sum_{j=1}^{m}(y_j - \hat{\mu}_Y)^2\right)\right) = (2\pi\hat{\sigma}^2)^{-\frac{n+m}{2}} exp(-\frac{n+m}{2})$$

Portanto,

$$\lambda(\underset{\sim}{x}, \underset{\sim}{y}) = \frac{sup_{\theta\in\Theta_0} L(\theta)}{sup_{\theta\in\Theta} L(\theta)} = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} = \frac{(2\pi\hat{\sigma}_0^2)^{-\frac{n+m}{2}} exp(-\frac{n+m}{2})}{(2\pi\hat{\sigma}^2)^{-\frac{n+m}{2}} exp(-\frac{n+m}{2})} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{\frac{n+m}{2}}$$

E o TRV tem região crítica dada por

$$A_1 = \{(\underset{\sim}{x}, \underset{\sim}{y});\ \lambda(\underset{\sim}{x}, \underset{\sim}{y}) \leq c\},\ c \in [0, 1]$$

Ou seja, substituindo $\hat{\sigma}^2$ e $\hat{\sigma_0}^2$ com um pouco de álgebra, temos que o TRV rejeita H_0 quando

$$\left(\frac{1}{1+\frac{n(\bar{x}-\hat{\mu_0})^2+m(\bar{y}-\hat{\mu_0})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{j=1}^{m}(y_j-\bar{y})^2}}\right)^{\frac{(n+m)}{2}} \leq c$$

que é equivalente a rejeitar H_0 quando

$$\frac{n(\bar{x}-\hat{\mu_0})^2+m(\bar{y}-\hat{\mu_0})^2}{s_p^2}$$

em que $s_p^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{j=1}^{m}(y_j-\bar{y})^2}{n+m-2}$. Só que

$$\bar{x}-\hat{\mu_0} = \frac{m}{n+m}(\bar{x}-\bar{y})$$

$$\bar{y}-\hat{\mu_0} = \frac{n}{n+m}(\bar{y}-\bar{x})$$

O que implica que o TRV pode ser escrito como

$$\lambda^*(\underset{\sim}{x};\underset{\sim}{y}) = \frac{\bar{x}-\bar{y}}{s_p\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}}$$

Portanto, a região crítica do TRV é dada por

$$A_1^* = \left\{(\underset{\sim}{x},\underset{\sim}{y});\ \lambda^*(\underset{\sim}{x};\underset{\sim}{y}) \leq\ c_1 \text{ ou } \lambda^*(\underset{\sim}{x};\underset{\sim}{y}) \geq\ c_2\right\}$$

Sob H_0, $\lambda^*(\underset{\sim}{X};\underset{\sim}{Y}) \sim t_{n+m-2}$, lembrando que a densidade da distribuição t é

$$f(t|k) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)}\left(1+\frac{t^2}{k}\right)^{-\frac{(k+1)}{2}}$$

sendo k os graus de liberdade.

Vale destacar que os valores de c_1 e c_2 são obtidos utilizando a tabela da distribuição t com $n+m-2$ graus de liberdade.