

# MAE 229 - Introdução à Probabilidade e Estatística II

## Resolução Lista 4

Professor: Pedro Morettin e Profa. Chang Chian

### Exercício 1

Antes de testar se a produtividade média dos operários do período diurno é igual a produtividade média do período noturno, precisamos verificar se a variância é igual para os dois grupos.

Em primeiro lugar note que

$$\begin{aligned} S^2 &= \frac{\sum_{i=1}^{15} (X_i - \bar{X})^2}{14} \\ &= \frac{\sum_{i=1}^{15} X_i^2 - 15 \frac{\sum_{i=1}^{15} X_i}{15}}{14} \\ &= \frac{\sum_{i=1}^{15} X_i^2 - \frac{1}{15} \left(\sum_{i=1}^{15} X_i\right)^2}{14} \end{aligned} \quad (1)$$

Seja  $S_1^2$  a variância amostral dos operários do diurno e  $S_2^2$  a variância amostral dos operários do noturno, então, pela equação (1), temos que

$$\begin{aligned} S_1^2 &= \frac{2660 - \frac{180^2}{15}}{14} \\ &= 35,7143 \\ S_2^2 &= \frac{2980 - \frac{150^2}{15}}{14} \\ &= 105,7143 \end{aligned}$$

Então, a estatística para testar  $H_0: \sigma_1^2 = \sigma_2^2$  versus  $H_1: \sigma_1^2 \neq \sigma_2^2$  é dada por

$$W = \frac{S_2^2}{S_1^2} \sim F(n_2 - 1, n_1 - 1)$$

sendo  $n_1$  o tamanho da amostra da população 1 e  $n_2$  o tamanho da amostra da população 2.

Sabendo que sob  $H_0$ , temos que  $W \sim F(14, 14)$ , logo a região crítica é tal que  $P(W \in RC) = P(W < f_1 \text{ ou } W > f_2)$  em que, considerando  $\alpha = 0,05$ ,  $P(W < f_1 | H_0) = 0,025$  e  $P(W < f_2 | H_0) = 0,975$ . Assim,  $f_1 = 0,3357$  e  $f_2 = 2,9785$  e  $RC = ]0,3357[ \cup ]2,9785; +\infty[$ . Como foi observado que  $W = 2,96 \notin RC$ , não rejeitamos a hipótese nula, ou seja, os operários do período diurno e noturno tem variâncias iguais.

Feito isso, agora, desejamos testar

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

em que  $\mu_1$  é a média dos operários do período diurno e  $\mu_2$  é a média dos operários do período noturno e as variâncias são iguais. Para essa situação, considere a seguinte estatística do teste

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{1/n_1 + 1/n_2}}$$

em que  $S_p^2 = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2}$  e, sob  $H_0$ ,  $T \sim t_{n_1+n_2-2}$ . Então, considerando um nível de significância  $\alpha = 0,05$ , a região crítica é  $]-\infty; f_1[ \cup ]f_2; +\infty[$  [ com  $P(T < f_1 | H_0) = 0,025$  e  $P(T > f_2 | H_0) = 0,025$ .

Usando os dados das amostras, temos que  $S_p^2 = 70,7143$ ,  $\bar{X} = 12$ ,  $\bar{Y} = 10$  e  $T = 0,6514$ . Ou seja, a região crítica é  $]-\infty; -2,0484[ \cup ]2,0484; +\infty[$  [ ao nível de significância  $\alpha = 0,05$ . Como  $T \notin RC$ , não temos evidência para rejeitar  $H_0$  e concluimos que não há diferença entre as produtividades média do período noturno e do período diurno.

## Exercício 2

Desejamos testar

$$H_0 : p_M - p_F = 0,1$$

$$H_1 : p_M - p_F \neq 0,1$$

em que  $p_M$  é a proporção de votantes masculinos do partido e  $p_F$  é a proporção de votantes femininos do partido.

Sob  $H_0$ , pode-se provar que

$$Z = \frac{(\hat{p}_M - \hat{p}_F) - (p_M - p_F)}{\sqrt{\frac{\hat{p}_M(1 - \hat{p}_M)}{n_M} + \frac{\hat{p}_F(1 - \hat{p}_F)}{n_F}}} \sim N(0,1)$$

com  $n_F = 625$ ,  $n_M = 400$ ,  $\hat{p}_F = 0,3104$  e  $\hat{p}_M = 0,425$ . A região de rejeição é da forma  $]-\infty; f_1[ \cup ]f_2; +\infty[$  [ com  $P(Z < f_1 | H_0) = \frac{\alpha}{2}$  e  $P(Z < f_2 | H_0) = 1 - \frac{\alpha}{2}$ .

Ou seja, a região crítica é  $]-\infty; -1,96[ \cup ]1,96; +\infty[$  [ ao nível de significância  $\alpha = 0,05$ . Como foi observado que sob  $H_0$ ,  $Z = 0,4728 \notin RC$ , concluimos que a hipótese do partido está correta, baseado nestas evidências (não rejeitamos  $H_0$ ).

## Exercício 3

Aqui temos o caso de uma amostra pareada em que podemos definir a variável  $D = Y - X$ , sendo  $X$  a variável aleatória que corresponde ao número de peças produzidas na semana sem intervalo e  $Y$  a variável aleatória que corresponde ao número de peças produzidas na semana com intervalo. Assumindo normalidade para a variável  $D$ , temos que  $D \sim N(\mu_D, \frac{\sigma_D^2}{n})$ .

Queremos testar se a produtividade efetivamente sobe com o intervalo para o café, ou seja

$$H_0 : \mu_D = \mu_Y - \mu_X = 0$$

$$H_1 : \mu_D = \mu_Y - \mu_X > 0$$

em que  $\mu_X$  a média de peças produzidas na semana sem intervalo e  $\mu_Y$  a média de peças produzidas na semana com intervalo.

Sabemos que  $\bar{D} = \bar{Y} - \bar{X}$ . Considere  $S_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2$ . Então, a estatística

$$T = \frac{\bar{n}(\bar{D} - \mu_D)}{S_D} \sim t_{n-1}.$$

Portanto, a região crítica é tal que  $P(T > t_c) = 0,05$ , assumindo um nível de significância de 5%. Como para esses dados,  $T = 1,2753$ , a região crítica do teste é  $]2,015; +\infty[$ . Foi observado que, sob  $H_0$  ( $\mu_D = 0$ ),  $T = 1,2753$ , em que foram usados os dados da tabela abaixo, sendo que a penúltima linha corresponde às médias e a última às variâncias amostrais ( $\bar{D} = 1,5$  e  $S_D^2 = 8,3$ ). Como  $T < RC$ , não rejeitamos  $H_0$ , e concluímos que baseado neste teste a produtividade não aumenta devido ao intervalo para café.

Variável D

D	X	Y
5	23	28
3	35	38
0	29	29
4	33	37
-1	43	42
-2	32	30
<b>1,5</b>	<b>32,5</b>	<b>34,0</b>
<b>8,3</b>	<b>43,9</b>	<b>33,2</b>

#### Exercício 4

(a) Desejamos testar

$$H_0 : \frac{\sigma_A^2}{\sigma_B^2} = \frac{2}{3}$$

$$H_1 : \frac{\sigma_A^2}{\sigma_B^2} > \frac{2}{3}$$

em que  $\frac{\sigma_A^2}{\sigma_B^2}$  é a variância da fábrica A e  $\frac{\sigma_B^2}{\sigma_A^2}$  é a variância da fábrica B. Considere a seguinte estatística

$$W = \frac{S_B^2}{S_A^2}$$

em que  $S_A^2$  é a variância amostral da fábrica A e  $S_B^2$  é a variância amostral da fábrica B.

Sob  $H_0$ , sabemos que  $W \sim F(n_B - 1, n_A - 1) = F(74, 99)$ . Então, a região crítica é  $]0; f_1[ \cup ]f_2; +\infty[$  com  $P(W < f_1 | H_0) = \frac{\alpha}{2}$  e  $P(W < f_2) = 1 - \frac{\alpha}{2}$ . Usando os dados do problema, temos a seguinte região de rejeição:  $]0; 0,6468[ \cup ]1,5252; +\infty[$  ao nível de  $\alpha = 0,05$ .

Como observou-se que  $W = 1,78 \in RC$ , rejeitamos a hipótese nula, ou seja, as variâncias não são iguais ao nível de significância  $\alpha = 0,05$ .

(b) Neste contexto, desejamos testar

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A \neq \mu_B$$

em que  $\mu_A$  é a vida média da lâmpada da fábrica A e  $\mu_B$  é a vida média da fábrica B. Assumindo normalidade das populações, considere a seguinte estatística (utilizada quando as variâncias não são iguais e são desconhecidas)

$$T = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{S_A^2/n_A + S_B^2/n_B}}$$

em que  $\bar{X}_A$  é média amostral da fábrica A e  $\bar{X}_B$  é média amostral da fábrica B. Sob  $H_0$ , sabemos que  $T \sim t_v$ , com  $v = \frac{(A+B)^2}{A^2/(n_A-1) + B^2/(n_B-1)}$ , com  $A = s_A^2/n_A$  e  $B = s_B^2/n_B$ . Além disso, a região crítica é  $] -\infty; f_1[ \cup ]f_2; +\infty[$  com  $P(T < f_1 | H_0) = \frac{\alpha}{2}$  e  $P(T < f_2 | H_0) = 1 - \frac{\alpha}{2}$ .

Para os dados do problema, temos que  $s_A^2 = 8100$ ,  $s_B^2 = 14400$ ,  $n_A = 100$ ,  $n_B = 75$ ,  $\bar{X}_A = 1190$ ,  $\bar{X}_B = 1230$ ,  $A = 81$  e  $B = 192$ , de forma que  $v = 132$  e  $T \sim t_{132}$ . Ou seja, a região de rejeição é  $] -\infty; -1,978[ \cup ]1,978; +\infty[$  com nível de significância  $\alpha = 0,05$ . Como  $T = -2,421 \in RC$ , rejeitamos a hipótese nula, isto é, concluímos que a vida média das lâmpadas produzidas em fábricas distintas é diferente.

Além disso, o intervalo de confiança para diferença das médias com confiança  $\gamma = 0,95$  é  $IC(\mu_A - \mu_B; 0,95) = \left[ \bar{X}_A - \bar{X}_B - f_1 S_p \sqrt{\frac{1}{100} + \frac{1}{75}}; \bar{X}_A - \bar{X}_B + f_2 S_p \sqrt{\frac{1}{100} + \frac{1}{75}} \right] = (-71,394; -8,606)$ .

### Exercício 5

Supondo que as duas amostras são independentes, primeiramente vamos testar se a variância dos salários dos torneiros mecânicos é diferente da variância dos empregados da indústria mecânica em geral.

$$H_0 : \sigma_{TM}^2 = \sigma_{IM}^2$$

$$H_1 : \sigma_{TM}^2 \neq \sigma_{IM}^2$$

em que  $\sigma_{TM}^2$  é a variância dos torneiros mecânicos e  $\sigma_{IM}^2$  é a variância da indústria mecânica. Considere a seguinte estatística

$$W = \frac{S_{TM}^2}{S_{IM}^2}$$

em que  $S_{TM}^2$  é a variância amostral dos torneiros mecânicos e  $S_{IM}^2$  é a variância amostral da indústria mecânica.

Sob  $H_0$ , sabemos que  $W \sim F(n_{TM} - 1, n_{IM} - 1) = F(24, 35)$ . Então, a região crítica é  $]0; f_1[ \cup ]f_2; +\infty[$  com  $P(W < f_1 | H_0) = \frac{\alpha}{2}$  e  $P(W < f_2) = 1 - \frac{\alpha}{2}$ . Usando os dados do problema, temos a seguinte região de rejeição:  $]0; 0,4601[ \cup ]2,0617; +\infty[$  com nível  $\alpha = 0,05$ .

Como observou-se que  $W = 2,163 < RC$ , rejeitamos a hipótese nula, ou seja, as variâncias não são iguais ao nível de significância  $\alpha = 0,05$ .

Sendo assim, agora queremos testar

$$H_0 : \mu_{TM} = \mu_{IM}$$

$$H_1 : \mu_{TM} \neq \mu_{IM}$$

em que  $\mu_{TM}$  é o salário médio dos torneiros mecânicos e  $\mu_{IM}$  é o salário médio da indústria mecânica. Assumindo normalidade das populações, considere a seguinte estatística (utilizada quando as variâncias não são iguais e são desconhecidas)

$$T = \frac{\bar{X}_{TM} - \bar{X}_{IM}}{\sqrt{\frac{S_{TM}^2/n_{TM} + S_{IM}^2/n_{IM}}{2}}}$$

em que  $\bar{X}_{TM}$  é média amostral dos torneiros mecânicos e  $\bar{X}_{IM}$  é média amostral da indústria mecânica. Sob  $H_0$ , sabemos que  $T \sim t_v$ , com  $v = \frac{(n_{TM} + n_{IM})^2}{n_{TM}^2/(n_{TM} - 1) + n_{IM}^2/(n_{IM} - 1)}$ , com  $TM = s_{TM}^2/n_{TM}$  e  $IM = s_{IM}^2/n_{IM}$ . Além disso, a região crítica é  $] -\infty; f_1[ \cup ]f_2; +\infty[$  com  $P(T < f_1 | H_0) = \frac{\alpha}{2}$  e  $P(T < f_2 | H_0) = 1 - \frac{\alpha}{2}$ .

Para os dados do problema, temos que  $s_{TM}^2 = 1,5625$ ,  $s_{IM}^2 = 0,7225$ ,  $n_{TM} = 25$ ,  $n_B = 36$ ,  $\bar{X}_{TM} = 4,22$ ,  $\bar{X}_{IM} = 3,64$ ,  $TM = 0,0625$  e  $IM = 0,0201$ , de forma que  $v = 39$  e  $T = 2,019 > t_{39}$ . Ou seja, a região de rejeição é  $] -\infty; -2,023[ \cup ]2,023; +\infty[$  com nível de significância  $\alpha = 0,05$ . Como  $T = 2,019 > RC$ , não rejeitamos a hipótese nula, isto é, concluímos que os salários médios não diferem, apesar da variância diferir.

## Exercício 6

Queremos testar se a nota média do grupo 1 é superior à nota média do grupo 2. Para isso, antes temos que avaliar se as variâncias são iguais, ou seja

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = \frac{2}{2}$$

$$H_1 : \frac{\sigma_1^2}{\sigma_2^2} = \frac{2}{2}$$

Considere a seguinte estatística

$$W = \frac{S_2^2}{S_1^2}$$

em que  $S_1^2$  é a variância amostral do grupo 1 e  $S_2^2$  é a variância amostral do grupo 2.

Sob  $H_0$ , sabemos que  $W \sim F(n_2 - 1, n_1 - 1) = F(31, 29)$ . Então, a região crítica é  $]0; f_1[ \cup ]f_2; +\infty[$  com  $P(W < f_1 | H_0) = \frac{1}{2}$  e  $P(W < f_2) = 1 - \frac{1}{2}$ . Usando os dados do problema, temos a seguinte região de rejeição:  $]0; 0,4841[ \cup ]2,0841; +\infty[$  ao nível de  $\alpha = 0,05$ .

Como observou-se que  $W = 1,78 < RC$ , não rejeitamos a hipótese nula, ou seja, as variâncias são iguais ao nível de significância  $\alpha = 0,05$ .

Feito isso, queremos testar se  $H_0 : \mu_1 = \mu_2$  contra  $H_1 : \mu_1 > \mu_2$ . Consciente que as variâncias são iguais e assumindo normalidade das populações temos para esse teste a estatística

$$T = \frac{\bar{X} - \bar{Y}}{S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

com  $S_P^2 = \frac{n_1(X_i - \bar{X})^2 + n_2(Y_j - \bar{Y})^2}{n_1 + n_2 - 2}$ , denotando por  $X$  a variável aleatória correspondente às notas do grupo 1 e por  $Y$  a variável aleatória das notas do grupo 2.

Sob  $H_0$ , sabemos que  $T \sim t_{n_1+n_2-2}$ . A região crítica neste caso é unilateral, da forma  $]f_2; +\infty[$  com  $P(T > f_2 | H_0) = \alpha$ .

Para os dados em questão, temos que  $n_1 = 30$  e  $n_2 = 32$ ,  $S_P^2 = 0,5047$ ,  $\bar{X} = 7,8$ ,  $\bar{Y} = 7,4$  e  $f_2 = 1,6707$ . Ou seja, a região crítica é  $]1,6707; +\infty[$  ao nível de significância  $\alpha = 0,05$ . Como foi observado que  $T = 2,216 > RC$ , rejeitamos  $H_0$ , ou seja, temos evidência para afirmar que as notas do grupo 1 são superiores.

### Exercício 7

- (a) Aqui temos o caso de uma amostra pareada em que podemos definir a variável  $D = X - Y$ , sendo  $X$  a variável aleatória que corresponde à pressão medida depois do remédio e  $Y$  a variável aleatória que corresponde à pressão medida antes do remédio. Assumindo normalidade para a variável  $D$ , temos que  $D \sim N(\mu_D, \frac{\sigma_D^2}{n})$ .

Queremos testar se a pressão efetivamente sobe depois de tomar o remédio, ou seja

$$H_0 : \mu_D = \mu_X - \mu_Y = 0$$

$$H_1 : \mu_D = \mu_X - \mu_Y > 0.$$

Sabemos que  $\bar{D} = \bar{X} - \bar{Y} \sim N(\mu_D, \frac{\sigma_D^2}{n})$ . Considere  $S_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2$ .

Então, a estatística

$$T = \frac{\bar{n}(\bar{D} - \mu_D)}{S_D} \sim t_{n-1}.$$

Portanto, a região crítica é tal que  $P(T > t_c) = 0,05$ , assumindo um nível de significância de 5%. Como para esses dados,  $T = t_6$ , a região crítica do teste é  $]1,9432; +\infty[$ . Foi observado que, sob  $H_0$  ( $\mu_D = 0$ ),  $T = 2,6737$ , em que foram usados os dados da tabela abaixo. Como  $T > RC$ , rejeitamos  $H_0$ , e concluímos que o remédio realmente tem o efeito colateral de aumentar a pressão diastólica.

Variável D

	Dados							Média	Variância
X	125	126	138	117	143	128	136	130,43	80,29
Y	120	124	130	118	140	128	130	127,14	54,48
D	5	2	8	-1	3	0	6	3,29	10,57

- (b) Um intervalo de confiança de  $\alpha = 0,9$  para o efeito médio da droga  $\mu_D$  é:  $IC(\mu_D; 0,9) = ]\bar{D} - t_{90\%} \frac{S_D}{\bar{n}}; \bar{D} + t_{90\%} \frac{S_D}{\bar{n}}[$ , em que  $t_{90\%}$  é tal que  $P(-t_{90\%} < T < t_{90\%}) = 0,9$ . Assim,  $t_{90\%} = 1,9432$  e  $IC = ]0,90; 5,67[$ .

### Exercício 8

Vamos testar primeiramente se as variâncias das populações são iguais:

$$H_0 : \sigma_A^2 = \sigma_B^2$$

$$H_1 : \sigma_A^2 \neq \sigma_B^2$$

em que  $\sigma_A^2$  é a variância da população A e  $\sigma_B^2$  é a variância da população B. Considere a seguinte estatística

$$W = \frac{S_B^2}{S_A^2}$$

em que  $S_A^2$  é a variância amostral da população A e  $S_B^2$  é a variância amostral da população B.

Sob  $H_0$ , sabemos que  $W \sim F(n_B - 1, n_A - 1) = F(7, 7)$ . Então, a região crítica é  $]0; f_1[ \cup ]f_2; +\infty[$  com  $P(W < f_1 | H_0) = \frac{\alpha}{2}$  e  $P(W < f_2) = 1 - \frac{\alpha}{2}$ . Usando os dados do problema, temos a seguinte região de rejeição:  $]0; 0,1840[ \cup ]5,4355; +\infty[$  ao nível de  $\alpha = 0,02$ .

Como observou-se que  $W = 1,625 > RC$ , não rejeitamos a hipótese nula, ou seja, as variâncias são iguais ao nível de significância  $\alpha = 0,02$ .

Neste contexto, desejamos testar

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A \neq \mu_B$$

em que  $\mu_A$  é a média da população A e  $\mu_B$  é a média da população B. Assumindo normalidade das populações, considere a seguinte estatística

$$T = \frac{\bar{X}_A - \bar{X}_B}{S_p \sqrt{1/n_A + 1/n_B}}$$

em que  $S_p^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}$ ,  $\bar{X}_A$  é a média amostral de A e  $\bar{X}_B$  é a média amostral de B.

Sob  $H_0$ , sabemos que  $T \sim t_{n_A + n_B - 2}$  e a região crítica é  $] -f_1[ \quad ] f_2[$ ; [ com  $P(T < f_1 | H_0) = \frac{\alpha}{2}$  e  $P(T < f_2 | H_0) = 1 - \frac{\alpha}{2}$ .

Para os dados do problema, temos que  $s_A^2 = 2,286$ ,  $s_B^2 = 3,714$ ,  $s_p^2 = 3$ ,  $n_A = 8$ ,  $n_B = 8$ ,  $\bar{X}_A = 14$ ,  $\bar{X}_B = 15,5$  e  $T \sim t_{14}$ . Ou seja, a região de rejeição é  $] -2,625[ \quad ] 2,625[$ ; - [ com nível de significância  $\alpha = 0,02$ .

Como foi observado que  $T = -1,732 / RC$ , não rejeitamos a hipótese nula, isto é, concluímos que as médias são iguais ao nível de significância de 2%.



