Lista 4 - Análise Estatística

Professor: Silvia Nagib Elian.
Monitor: Alberto Rodrigues Ferreira.

1. Os dados a seguir correspondem a comprimento craniano (x_1) e Largura (x_2) para uma amostra de 35 fêmeas e 14 machos de uma espécie:
   Fêmeas
   $\bar{x_1} = \begin{bmatrix} 22,860 \\ 24,397 \end{bmatrix} S_1 = \begin{bmatrix} 17,683 & 20,292 \\ 20,292 & 24,407 \end{bmatrix}$

   Machos
   $\bar{x_2} = \begin{bmatrix} 21,821 \\ 22,834 \end{bmatrix} S_2 = \begin{bmatrix} 18,479 & 19,095 \\ 19,095 & 20,755 \end{bmatrix}$

   Testa ao nível 0,05 a hipótese de igualdade dos vetores de média para as populações de machos e fêmeas.
   Solução:
   $H_0 : \mu_2 = \mu_1$

   $n_1 = 35$ e $n_2 = 14$
   $S_c = \dfrac{1}{n_1 + n_2 - 2}[(n_1 - 1)S_1 + (n_2 - 1)S_2] = \begin{bmatrix} 17.90317 & 19.96091 \\ 19.96091 & 23.39687 \end{bmatrix}$
   Estatística de teste

   $$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\bar{x_1} - \bar{x_2})^T S_c^{-1}(\bar{x_1} - \bar{x_2})$$

   $$= \frac{35 * 14}{35 + 14} \left(\begin{bmatrix} 22,860 \\ 24,397 \end{bmatrix} - \begin{bmatrix} 21,821 \\ 22,834 \end{bmatrix} \right)^T \begin{bmatrix} 17.90317 & 19.96091 \\ 19.96091 & 23.39687 \end{bmatrix}^{-1} \left(\begin{bmatrix} 22,860 \\ 24,397 \end{bmatrix} - \begin{bmatrix} 21,821 \\ 22,834 \end{bmatrix} \right) = 2.036672$$

   Quantil da distribuição T^2

   $$\frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{(p, n_1 + n_2 - p - 1)}(\alpha) = \frac{94}{46} F_{(2,46)}(0.05) = 6.538276$$

   Conclusão
   Como a estatística de teste é menor que $\dfrac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{(p, n_1 + n_2 - p - 1)}(\alpha)$ existem evidências estatísticas para não rejeitarmos a hipótese de igualdade de médias entre machos e fêmeas ao nível de 5%.

2. Considere o vetor aleatório $\begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix}^T \sim N(\mu, \Sigma)$, com $X_1 =$ Nota da informação, X_2= Nota de similaridades e X_3= Nota de aritmética. Para uma amostra de 49 pessoas obteve-se a matriz de

covariância amostral.

$$S = \begin{bmatrix} 121 & 81 & 49 \\ 81 & 169 & 49 \\ 49 & 49 & 121 \end{bmatrix}$$

Com base nessas informações, teste ao nível de significância de 10% a hipótese

$$\Sigma = \Sigma_0 = \begin{bmatrix} 100 & 60 & 60 \\ 60 & 100 & 60 \\ 60 & 60 & 100 \end{bmatrix}$$

Solução:
$H_0 : \Sigma = \Sigma_0$

$n = 49$
$det(\Sigma_0) = det\left(\begin{bmatrix} 100 & 60 & 60 \\ 60 & 100 & 60 \\ 60 & 60 & 100 \end{bmatrix}\right) = 352000,\ det(S) = det\left(\begin{bmatrix} 121 & 81 & 49 \\ 81 & 169 & 49 \\ 49 & 49 & 121 \end{bmatrix}\right) = 1373120$

$$\Sigma_0^{-1} = \begin{bmatrix} 0.018181818 & -0.006818182 & -0.006818182 \\ -0.006818182 & 0.018181818 & -0.006818182 \\ -0.006818182 & -0.006818182 & 0.018181818 \end{bmatrix}$$

$$S\Sigma_0^{-1} = \begin{bmatrix} 1.31363636 & 0.3136364 & -0.4863636 \\ -0.01363636 & 2.1863636 & -0.8136364 \\ -0.26818182 & -0.2681818 & 1.5318182 \end{bmatrix}$$

$tr(S\Sigma_0^{-1}) = 5.031818$

Estatística de teste

Para n grande

$$\begin{aligned} L &= (n-1)[ln(det(\Sigma_0)) - ln(det(S)) + tr(S\Sigma_0^{-1}) - p] \\ &= 48 * [ln(352000) - ln(1373120) + 5.031818 - 3] = 32.18921 \end{aligned}$$

Para n moderado

$$L^{'} = L\left[1 - \frac{1}{6(n-1)}\left(2p + 1 - \frac{2}{p+1}\right)\right] = 31.46272$$

Quantil da distribuição qui quadrado
$\chi^2(\alpha)_{(p(p+1)/2)} = \chi^2(0.1)_6 = 10.64464$

Conclusão
Como L e $L^{'}$ são maiores que $\chi^2(\alpha)_{(p/p+1)}$ existem evidências estatísticas para rejeitarmos a hipótese nula com nível de significação de 10%.

3. Medidas de comprimento, largura e altura das carapaças de 24 tartarugas macho e 24 tartarugas fêmeas foram avaliadas. As matrizes de covariância observadas foram

$$S_M = \begin{bmatrix} 138.77 & 79.15 & 37.38 \\ 79.15 & 50.04 & 21.65 \\ 37.38 & 21.65 & 11.26 \end{bmatrix} \text{ e } S_F = \begin{bmatrix} 451.39 & 271.17 & 168.70 \\ 271.17 & 171.73 & 103.29 \\ 168.70 & 103.29 & 66.65 \end{bmatrix}$$

Teste ao nível de 0.05 a hipótese de igualdade das matrizes de covariâncias populacionais para machos e fêmeas.
Solução:
$H_0 : \Sigma_M = \Sigma_F$
$K = 2, n_1 = 24, n_2 = 24, p = 3$
Estatística de teste

$$M = C\left[\sum_{l=1}^{2}(n_l - 1)ln[det(S_c)] - \sum_{l=1}^{2}(n_l - 1)ln[det(S_l)]\right]$$

$$\begin{aligned} C &= 1 - \frac{(2p^2 + 3p - 1)(K + 1)}{6(p + 1)Kn} \\ &= 1 - \frac{(2 * 9 + 3 * 3 - 1)(2 + 1)}{6 * (3 + 1)2 * 24} \\ &= 1 - \frac{104}{1632} = 0.9507576 \end{aligned}$$

$$\begin{aligned} S_c &= \frac{1}{\sum_{l=1}^{2}(n_l - 1)}\sum_{l=1}^{2}(n_l - 1)S_l \\ &= \frac{1}{46}\left(23\begin{bmatrix} 138.77 & 79.15 & 37.38 \\ 79.15 & 50.04 & 21.65 \\ 37.38 & 21.65 & 11.26 \end{bmatrix} + 23\begin{bmatrix} 451.39 & 271.17 & 168.70 \\ 271.17 & 171.73 & 103.29 \\ 168.70 & 103.29 & 11.26 \end{bmatrix}\right) \\ &= \begin{bmatrix} 295.08 & 175.160 & 103.040 \\ 175.16 & 110.885 & 62.470 \\ 103.04 & 62.470 & 38.955 \end{bmatrix} \end{aligned}$$

$$det(S_c) = det\left(\begin{bmatrix} 295.08 & 175.160 & 103.040 \\ 175.16 & 110.885 & 62.470 \\ 103.04 & 62.470 & 38.955 \end{bmatrix}\right) = 5561.227$$

$$det(S_M) = det(S_1) = det\left(\begin{bmatrix} 138.77 & 79.15 & 37.38 \\ 79.15 & 50.04 & 21.65 \\ 37.38 & 21.65 & 11.26 \end{bmatrix}\right) = 794.0539$$

$$det(S_F) = det(S_2) = det\left(\begin{bmatrix} 451.39 & 271.17 & 168.70 \\ 271.17 & 171.73 & 103.29 \\ 168.70 & 103.29 & 66.65 \end{bmatrix}\right) = 12639.89$$

$M = 23.5471$

$\chi^2(\alpha)_{((K-1)p(p+1)/2)} = \chi^2(0.05)_{(6)} = 12.59159$

Conclusão

Como M é maior que $\chi^2(\alpha)_{((K-1)p(p+1)/2)}$, existem evidências estatísticas para rejeitarmos a hipótese de que a matriz de covariâncias populacionais sejam iguais.

4. Considere as variáveis X_1-Peso e X_2-Altura e as 39 medidas observadas no arquivo Peru do Minitab. Admitindo que (X_1, X_2) tem distribuição normal bivariada, teste a hipótese $\mu = (63.64kg, 161.5cm)$
Solução:
$H_0 : \mu = \mu_0$

$\bar{X} = [63.15897, 157.89231]^T$

$\mu_0 = [63.64, 161.5]^T$

$$S = \begin{bmatrix} 50.39406 & 16.84573 \\ 16.84573 & 27.76757 \end{bmatrix}$$

$$S^{-1} = \begin{bmatrix} 0.02489155 & -0.01510093 \\ -0.01510093 & 0.04517450 \end{bmatrix}$$

Estatística de teste

$T^2 = n(\bar{X} - \mu_0)^T S^{-1} (\bar{X} - \mu_0)$

$$= 39\left(\begin{bmatrix} 63.15897 \\ 157.89231 \end{bmatrix} - \begin{bmatrix} 63.64 \\ 161.5 \end{bmatrix}\right)^T \begin{bmatrix} 0.02489155 & -0.01510093 \\ -0.01510093 & 0.04517450 \end{bmatrix} \left(\begin{bmatrix} 63.15897 \\ 157.89231 \end{bmatrix} - \begin{bmatrix} 63.64 \\ 161.5 \end{bmatrix}\right) = 21.11123$$

$$\frac{(n-1)pF_{p,n-p}(\alpha)}{n-p} = \frac{38 * 2 * 3.251924}{37} = 6.679627$$

Conclusão
Como T^2 é maior que $\frac{(n-1)pF_{p,n-p}(\alpha)}{n-p}$, existem evidências estatísticas para acreditarmos que a

hipótese de que $\mu = \mu_0$ não seja verdadeira.

5. (a) Teste ao nível de 0,05 a hipótese de inexistência de efeito da condição para ambas as variáveis X_1 e X_2.
Solução:

$H_0 : \mu_{1i2} = \mu_{1i1}$ e $\mu_{2i2} = \mu_{2i1}$
Estatística de teste

$\bar{D} = [\bar{D}_1, \bar{D}_2]^T = [-22.90, -20.28]^T$

$S_D = \frac{1}{n-1}\sum_{i=1}^{10}(D_i - \bar{D})(D_i - \bar{D})^T = \begin{bmatrix} 496.52222 & 81.38889 \\ 81.38889 & 115.49067 \end{bmatrix}$
$S_D^{-1} = \begin{bmatrix} 0.002277045 & -0.001604685 \\ -0.001604685 & 0.009789566 \end{bmatrix}$

$T^2 = n\bar{D}^T S_D^{-1}\bar{D} = 10[22.90, 20.28]^T \begin{bmatrix} 0.002277045 & -0.001604685 \\ -0.001604685 & 0.009789566 \end{bmatrix} [22.90, 20.28] = 37.29872$
$\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha) = \frac{9*2}{8}F_{2,8}(0.05) = \frac{9}{4}4.45897 = 10.03268$
Conclusão
Como T^2 é maior que $\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$, existem evidências estatísticas para acreditarmos na inexistência de efeito da condição das variáveis X_1 e X_2.

6. **Solução:**
$H_0 : \mu_2 - \mu_1 = \mu_3 - \mu_2 = \mu_4 - \mu_3$

Estatística de teste
$T^2 = n(C\bar{X})^T(CSC^T)^{-1}(C\bar{X})$

$C = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$
$\bar{X} = [24.5, 27.5, 24.1, 30.5]^T$

$S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T = \begin{bmatrix} 23.3888889 & 1.50000 & -0.8333333 & -1.50000 \\ 1.5000000 & 22.50000 & 24.9444444 & 12.27778 \\ -0.8333333 & 24.94444 & 34.5444444 & 13.05556 \\ -1.5000000 & 12.27778 & 13.0555556 & 28.72222 \end{bmatrix}$

$C\bar{X} = [3, -3.4, 6.4]^T$

$$CSC^T = \begin{bmatrix} 42.888889 & 4.777778 & 12.000000 \\ 4.777778 & 7.155556 & -8.822222 \\ -12.000000 & -8.822222 & 37.155556 \end{bmatrix}$$

$$(CSC^T)^{-1} = \begin{bmatrix} 0.026124388 & -0.00995508 & 0.006073564 \\ -0.009955088 & 0.201390541 & 0.044603049 \\ 0.006073564 & 0.044603049 & 0.039465990 \end{bmatrix}$$

$$T^2 = 10[3, -3.4, 6.4] \begin{bmatrix} 0.026124388 & -0.00995508 & 0.006073564 \\ -0.009955088 & 0.201390541 & 0.044603049 \\ 0.006073564 & 0.044603049 & 0.039465990 \end{bmatrix} \begin{bmatrix} 3 \\ -3.4 \\ 6.4 \end{bmatrix} = 26.74905$$

$$\frac{(n-1)(p-1)F_{p-1,n-p+1}(\alpha)}{n-p+1} = \frac{27F_{3,7}(0.05)}{7} = 16.76635$$

Conclusão

Como T^2 é maior que $\frac{(n-1)(p-1)F_{p-1,n-p+1}(\alpha)}{n-p+1}$, podemos rejeitar a hipótese de igualdade dos ganhos médios de pesos nas quatro semanas com nível de significância de 5%.

7. (a) Admitindo normalidade para as 4 variáveis, teste ao nível de 5% a hipótese de igualdade de médias das porcentagens de erro nos 4 altímetros.

Solução:

Estatística de teste

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

$T^2 = n(C\bar{X})^T(CSC^T)^{-1}(C\bar{X})$

$$C = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

$\bar{X} = [2.750, 6.250, 3.500, 8.625]^T$

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T \begin{bmatrix} 2.2142857 & 1.3571429 & 1.1428571 & -0.8214286 \\ 1.3571429 & 1.0714286 & 0.7142857 & -0.6071429 \\ 1.1428571 & 0.7142857 & 0.8571429 & -0.5000000 \\ -0.8214286 & -0.6071429 & -0.5000000 & 1.1250000 \end{bmatrix}$$

$C\bar{X} = [3.5, -2.7, 5.125]^T$

$$CSC^T = \begin{bmatrix} 0.5714286 & -0.14285714 & 0.64285714 \\ 0.1428571 & 0.50000000 & -0.03571429 \\ 0.6428571 & -0.03571429 & 2.98214286 \end{bmatrix}$$

$$(CSC^T)^{-1} = \begin{bmatrix} 2.5234568 & 0.6827160 & -0.5358025 \\ 0.6827160 & 2.1864198 & -0.1209877 \\ -0.5358025 & -0.1209877 & 0.4493827 \end{bmatrix}$$

$$T^2 = 8[3.5, -2.7, 5.125]\begin{bmatrix} 2.5234568 & 0.6827160 & -0.5358025 \\ 0.6827160 & 2.1864198 & -0.1209877 \\ -0.5358025 & -0.1209877 & 0.4493827 \end{bmatrix}\begin{bmatrix} 3.5 \\ -2.7 \\ 5.125 \end{bmatrix} = 242.3728$$

$$\frac{(n-1)(p-1)F_{p-1,n-p+1}(\alpha)}{n-p+1} = \frac{21F_{3,5}(0.05)}{5} = 22.7197$$

Conclusão

Como T^2 é maior que $\frac{(n-1)(p-1)F_{p-1,n-p+1}(\alpha)}{n-p+1}$, existem evidências estatísticas para rejeitarmos a igualdade de médias das porcentagens de erro nos 4 altímetros.

8. Amostras de tamanho $n_1 = 45$ e $n_2 = 55$ foram tomadas de proprietários de residências com e sem ar condicionado, respectivamente. Foram medidos consumos de energia no horário normal $(X1)$ e no horário de pico $(X2)$ durante determinado mês. Os resultados foram:

Com ar condicionado: $\bar{x}_1 = \begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix}$, $S_1 = \begin{bmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{bmatrix}$

Sem ar condicionado: $\bar{x}_2 = \begin{bmatrix} 130 \\ 355 \end{bmatrix}$, $S_2 = \begin{bmatrix} 8632 & 19616.7 \\ 19616.7 & 55964.5 \end{bmatrix}$

b) **Solução:**

$H_0 : \mu_2 = \mu_1$

$$S_c = \frac{1}{n_1 + n_2 - 2}[(n_1 - 1)S_1 + (n_2 - 1)S_2] = \begin{bmatrix} 10963.69 & 21505.42 \\ 21505.42 & 63661.31 \end{bmatrix}$$

$$S_c^{-1} = \begin{bmatrix} 2.703477 * 10^{-4} & -9.132612 * 10^{-5} \\ -9.132612 * 10^{-5} & 4.655900 * 10^{-5} \end{bmatrix}$$

Estatística de teste

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\bar{x_1} - \bar{x_2})^T S_c^{-1}(\bar{x_1} - \bar{x_2})$$
$$= \frac{45 * 55}{45 + 55}\left(\begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix} - \begin{bmatrix} 130 \\ 355 \end{bmatrix}\right)^T \begin{bmatrix} 2.703477 * 10^{-4} & -9.132612 * 10^{-5} \\ -9.132612 * 10^{-5} & 4.655900 * 10^{-5} \end{bmatrix}\left(\begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix} - \begin{bmatrix} 130 \\ 355 \end{bmatrix}\right)$$
$$= 16.06622$$

Quantil da distribuição T^2

$$\frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1}F_{(p,n_1+n_2-p-1)}(\alpha) = \frac{(35 + 17 - 2)2}{35 + 17 - 2 - 1}F_{(2,35+17-2-1)}(0.05) = 6.538276$$

Conclusão

Como T^2 é maior que $\frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1}F_{(p,n_1+n_2-p-1)}(\alpha)$, existem evidências estatísticas para rejeitarmos a hipótese de igualdade dos vetores de médias.