

Análise Multivariada II

Lista II

Breno Cauã Rodrigues da Silva

Questão 1. Análise A Posteriori

Pode-se estimar os vetores τ_k para os tratamentos A, B, C e D de forma que

$$\begin{aligned}\hat{\tau}_A &= (\bar{\mathbf{X}}_A - \bar{\mathbf{X}})^\top = (-274.95 \quad -0.59)^\top \\ \hat{\tau}_B &= (\bar{\mathbf{X}}_B - \bar{\mathbf{X}})^\top = (-202.75 \quad 0.37)^\top \\ \hat{\tau}_C &= (\bar{\mathbf{X}}_C - \bar{\mathbf{X}})^\top = (179.45 \quad 0.03)^\top \\ \hat{\tau}_D &= (\bar{\mathbf{X}}_D - \bar{\mathbf{X}})^\top = (298.25 \quad 0.20)^\top\end{aligned}$$

Usando a Soma de Quadrados do Resíduo de Within (\mathbf{W}) expressa por

$$\begin{aligned}\mathbf{W} &= \sum_{k=1}^g \sum_{i=1}^{n_k} (\mathbf{x}_{ki} - \bar{\mathbf{x}}_k)(\mathbf{x}_{ki} - \bar{\mathbf{x}}_k)^\top = n_1 \mathbf{S}_1 + n_2 \mathbf{S}_2 + \dots + n_g \mathbf{S}_g \\ \mathbf{W} &\approx \begin{bmatrix} 29058.55 & 10.26 \\ 10.26 & 0.32 \end{bmatrix},\end{aligned}$$

com $n - g = 20 - 4 = 16$ graus de liberdade. Onde $n = n_1 + n_2 + \dots + n_g$ é o tamanho total da amostra e $g = 4$ o número de tratamentos em comparação. Temos que a diferença entre a produtividade de dois tratamentos quaisquer pode ser obtida por

$$\hat{\tau}_{k_1} - \hat{\tau}_{j_1},$$

com $\widehat{Var}[\hat{\tau}_{k_1} - \hat{\tau}_{j_1}] = \left(\frac{1}{n_k} + \frac{1}{n_j}\right) \frac{\mathbf{W}_{11}}{n - g}$. De posse da estatística $\hat{\tau}_{k_1} - \hat{\tau}_{j_1}$ e de sua variância, podemos montar o intervalo de $100(1 - \alpha)\%$ de confiança, expresso por

$$(\hat{\tau}_{k_1} - \hat{\tau}_{j_1}) \pm t_{n-g; \alpha/pg(p-1)} \sqrt{\widehat{Var}[\hat{\tau}_{k_1} - \hat{\tau}_{j_1}]}.$$

Comparação entre a *Produtividade A e B*

Temos que

$$\hat{\tau}_{A_1} - \hat{\tau}_{B_1} = -274.95 - (-202.75) = -72.20,$$

e $\widehat{Var}[\hat{\tau}_{A_1} - \hat{\tau}_{B_1}] = \left(\frac{1}{5} + \frac{1}{5}\right) \frac{29058.55}{20 - 4} = 726.46$, logo, $\sqrt{726.46} \approx 26.95$. Fixando um $\alpha = 0.05$, temos que o quantil de acordo com a correção de Bonferroni $t_{20-4; 0.05/2 \cdot 4 \cdot (4-1)} = t_{16; 0.002} \approx 3.34$. Então, o intervalo de confiança fica

$$-72.20 \pm 3.34 \times 26.95 \approx [-162, 21; 17.81].$$

Como o intervalo contém o 0, diz-se que não há diferença entre a Produtividade do Tratamento A e do Tratamento B, ao nível de significância de 5%.

Comparação entre a *Produtividade* A e C

Temos que

$$\hat{\tau}_{A_1} - \hat{\tau}_{C_1} = -274.95 - 179.45 = -454.4,$$

e $\widehat{Var}[\hat{\tau}_{A_1} - \hat{\tau}_{C_1}] = \left(\frac{1}{5} + \frac{1}{5}\right) \frac{29058.55}{20-4} = 726.46$, logo, $\sqrt{726.46} \approx 26.95$. Fixando um $\alpha = 0.05$, temos que o quantil de acordo com a correção de Bonferroni $t_{20-4; 0.05/2 \cdot 4 \cdot (4-1)} = t_{16; 0.002} \approx 3.34$. Então, o intervalo de confiança fica

$$-454.4 \pm 3.34 \times 26.95 \approx [-364.39; -544.41].$$

Como o intervalo não contém o 0, diz-se que há diferença entre a Produtividade do Tratamento A e do Tratamento C, ao nível de significância de 5%.

Comparação entre a *Produtividade* A e D

Temos que

$$\hat{\tau}_{A_1} - \hat{\tau}_{D_1} = -274.95 - 289.25 = -564.2,$$

e $\widehat{Var}[\hat{\tau}_{A_1} - \hat{\tau}_{D_1}] = \left(\frac{1}{5} + \frac{1}{5}\right) \frac{29058.55}{20-4} = 726.46$, logo, $\sqrt{726.46} \approx 26.95$. Fixando um $\alpha = 0.05$, temos que o quantil de acordo com a correção de Bonferroni $t_{20-4; 0.05/2 \cdot 4 \cdot (4-1)} = t_{16; 0.002} \approx 3.34$. Então, o intervalo de confiança fica

$$-564.2 \pm 3.34 \times 26.95 \approx [-654.21; -474.19].$$

Como o intervalo não contém o 0, diz-se que há diferença entre a Produtividade do Tratamento A e do Tratamento D, ao nível de significância de 5%.

Comparação entre a *Produtividade* B e C

Temos que

$$\hat{\tau}_{B_1} - \hat{\tau}_{C_1} = -202.25 - 179.45 = -381.7,$$

e $\widehat{Var}[\hat{\tau}_{B_1} - \hat{\tau}_{C_1}] = \left(\frac{1}{5} + \frac{1}{5}\right) \frac{29058.55}{20-4} = 726.46$, logo, $\sqrt{726.46} \approx 26.95$. Fixando um $\alpha = 0.05$, temos que o quantil de acordo com a correção de Bonferroni $t_{20-4; 0.05/2 \cdot 4 \cdot (4-1)} = t_{16; 0.002} \approx 3.34$. Então, o intervalo de confiança fica

$$-381.7 \pm 3.34 \times 26.95 \approx [-471, 71; -291, 69].$$

Como o intervalo não contém o 0, diz-se que há diferença entre a Produtividade do Tratamento A e do Tratamento D, ao nível de significância de 5%.

Comparação entre a *Produtividade* B e D

Temos que

$$\hat{\tau}_{B_1} - \hat{\tau}_{D_1} = -202.25 - 298.75 = -501,$$

e $\widehat{Var}[\hat{\tau}_{B_1} - \hat{\tau}_{D_1}] = \left(\frac{1}{5} + \frac{1}{5}\right) \frac{29058.55}{20-4} = 726.46$, logo, $\sqrt{726.46} \approx 26.95$. Fixando um $\alpha = 0.05$, temos que o quantil de acordo com a correção de Bonferroni $t_{20-4; 0.05/2 \cdot 4 \cdot (4-1)} = t_{16; 0.002} \approx 3.34$. Então, o intervalo de confiança fica

$$-501 \pm 3.34 \times 26.95 \approx [-591.01; -410.99].$$

Como o intervalo não contém o 0, diz-se que há diferença entre a Produtividade do Tratamento A e do Tratamento D, ao nível de significância de 5%.

Comparação entre a *Produtividade C* e *D*

Temos que

$$\hat{\tau}_{C_1} - \hat{\tau}_{D_1} = 179.45 - 298.25 = -118.8,$$

e $\widehat{Var}[\hat{\tau}_{C_1} - \hat{\tau}_{D_1}] = \left(\frac{1}{5} + \frac{1}{5}\right) \frac{29058.55}{20 - 4} = 726.46$, logo, $\sqrt{726.46} \approx 26.95$. Fixando um $\alpha = 0.05$, temos que o quantil de acordo com a correção de Bonferroni $t_{20-4; 0.05/2 \cdot 4 \cdot (4-1)} = t_{16; 0.002} \approx 3.34$. Então, o intervalo de confiança fica

$$-118.8 \pm 3.34 \times 26.95 \approx [-208.81; -28.79].$$

Como o intervalo não contém o 0, diz-se que há diferença entre a Produtividade do Tratamento A e do Tratamento D, ao nível de significância de 5%.

Questão 2. MANOVA

(a) Converta o banco de dados para o formato longo

Os dados foram devidamente convertidos do formato *wide* para o formato *long*. O código utilizado para tal tarefa será anexado ao final deste documento. O conjunto de dados convertidos está disponível no link [DADOS_L2Q2_LONG.csv](#).

(b) Faça uma análise exploratória dos dados

Como análise preliminar, foram construídas a Table 1, Table 2, Table 3, Figure 1 e a @fig. Com o objetivo de identificar padrões que corroborem com os resultados inferenciais.

Table 1: Medidas de Resumo do Conjunto de Dados para o Método 1.

	Média	<i>S</i>	Mediana	Mín	Máx	1º Quartil	3º Quartil
AROMA	5,4	0,58	5,5	4	6,1	5,2	5,7
SABOR	5,7	0,44	5,9	5	6,2	5,3	6,0
TEXTURA	5,4	0,66	5,6	4	6,3	5,0	6,0
UMIDADE	6,0	0,70	6,0	5	7,0	5,6	6,5

Table 2: Medidas de Resumo do Conjunto de Dados para o Método 2.

	Média	<i>S</i>	Mediana	Mín	Máx	1º Quartil	3º Quartil
AROMA	5,3	0,76	5,2	3,9	6,2	5,0	5,9
SABOR	5,2	0,57	5,2	4,0	6,1	4,9	5,5
TEXTURA	5,3	0,59	5,3	4,2	6,1	5,0	5,7
UMIDADE	5,9	0,52	6,0	4,8	6,5	5,8	6,2

Table 3: Medidas de Resumo do Conjunto de Dados para o Método 3.

	Média	<i>S</i>	Mediana	Mín	Máx	1º Quartil	3º Quartil
AROMA	5,0	0,54	4,8	4,2	6,0	4,6	5,3
SABOR	4,8	0,46	5,0	4,0	5,4	4,5	5,2

Table 3: Medidas de Resumo do Conjunto de Dados para o Método 3.

	Média	S	Mediana	Mín	Máx	1º Quartil	3º Quartil
TEXTURA	5,9	0,51	5,8	5,0	6,8	5,7	6,2
UMIDADE	6,2	0,46	6,4	5,5	7,0	5,8	6,5

As Table 1, Table 2 e Table 3 apresentam as medidas descritivas para os métodos 1, 2 e 3, respectivamente. A seguir, destacam-se algumas observações importantes:

- O **Método 1** apresentou as maiores médias em praticamente todas as variáveis sensoriais, com destaque para **Sabor (5,7)** e **Umidade (6,0)**, sugerindo que esse método foi o mais bem avaliado pelos juízes.
- O **Método 2** obteve médias intermediárias em todas as variáveis, com valores próximos aos do Método 1, porém ligeiramente menores.
- O **Método 3** apresentou as **menores médias para Aroma (5,0)** e **Sabor (4,8)**, indicando avaliações sensoriais inferiores nesses atributos.
- Em relação à variabilidade, os **desvios padrão** foram relativamente baixos em todos os métodos, indicando **baixa dispersão entre as observações**. O maior desvio padrão ocorreu para **Aroma no Método 2 (0,76)**.
- As distribuições mostraram-se em geral **simétricas ou levemente assimétricas à esquerda**, e com curtose negativa na maioria dos casos, indicando distribuições levemente achatadas (platicúrticas).

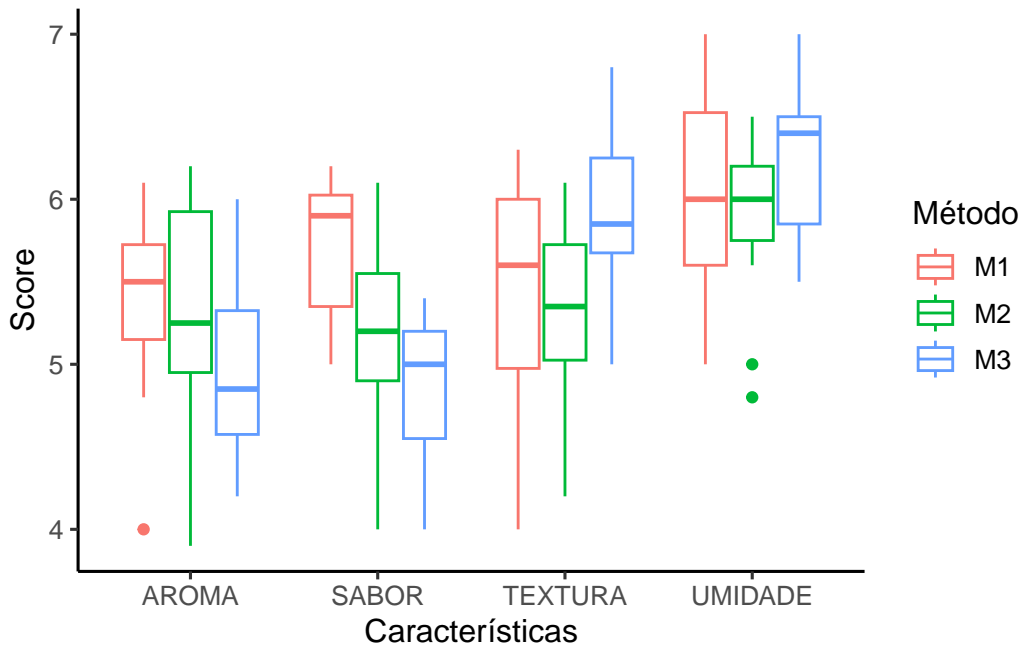


Figure 1: Boxplot das Variáveis do Conjunto de Dados Segmentadas pelo Método Aplicado.

A Figure 1 apresenta os boxplots das quatro variáveis segmentadas por método.

- Para **Aroma** e **Sabor**, observa-se que o **Método 1** possui as **maiores medianas** e menos dispersão, enquanto o **Método 3** apresenta as **menores medianas**.
- Na variável **Textura**, os três métodos apresentaram valores similares, com destaque para o **Método 3**, que apresentou uma leve vantagem em relação à mediana.

- Na variável **Umidade**, as medianas foram próximas entre os métodos, com **Método 3** apresentando uma mediana ligeiramente superior.

De modo geral, os boxplots indicam **vantagem do Método 1 nas variáveis Aroma e Sabor**, enquanto **Textura e Umidade** parecem menos discriminativas entre os métodos.

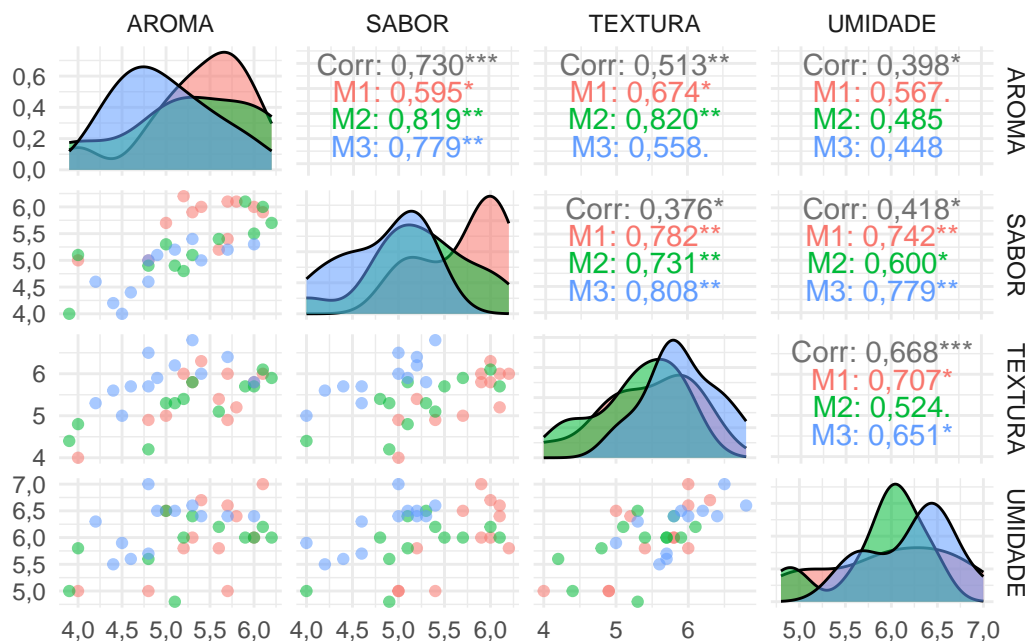


Figure 2: Pairplot do Conjunto de Dados Segmentadas pelo Método Aplicado.

A Figure 2 apresenta o pairplot das variáveis sensoriais, com gráficos de dispersão e os respectivos coeficientes de correlação de Pearson.

- De forma geral, observou-se **associação positiva** entre as variáveis, especialmente entre:
 - **Aroma e Sabor** ($r = 0,730$)
 - **Textura e Sabor** ($r = 0,668$)
 - **Umidade e Textura** ($r = 0,418$)
- Ao segmentar por método, notam-se correlações moderadas e significativas nos três grupos, com destaque para:
 - **Método 1:** Aroma e Sabor ($r = 0,595$), Textura e Umidade ($r = 0,742$)
 - **Método 2:** Aroma e Sabor ($r = 0,820$)
 - **Método 3:** Aroma e Sabor ($r = 0,779$), Textura e Sabor ($r = 0,651$)

Esses resultados indicam que os julgadores **avaliaram os atributos de forma coerente**, ou seja, peixes bem avaliados em um aspecto tendem a receber boas avaliações nos demais.

(c) Faça a verificação de pressupostos

As suposições da Manova estão relacionadas à normalidade dos erros e igualdade das matrizes de covariância nas populações.

Table 4: Teste de Mardia para Normalidade Multivariada.

Teste	Estatística de Teste	P-Valor	Normalidade
Mardia Skewness	11,0661582905076	0,944488761492132	YES
Mardia Kurtosis	-1,591217415925	0,111560654735673	YES
MVN	NA	NA	YES

Table 5: Teste de Henze-Zirkler para Normalidade Multivariada.

Teste	Estatística de Teste	P-Valor	Normalidade
Henze-Zirkler	0,9	0,07	YES

Table 6: Teste de Royston para Normalidade Multivariada.

Teste	Estatística de Teste	P-Valor	Normalidade
Royston	8,3	0,08	YES

De acordo com os dados das Table 4, Table 5 e Table 6 os dados seguem uma distribuição normal multivariada. Agora, basta verificar se a Matriz de Covariância dos grupos são iguais (homogeneidade).

Table 7: Teste M de Box para Homogeneidade das Matrizes de Covariâncias.

Estatística de Teste	Graus de Liberdade	P-Valor
13	20	0,86

Como o $p - valor > 0,05$, não se pode rejeitar a hipótese nula de homogeneidade das matrizes de covariâncias.

(d) Faça uma análise de variância multivariada usando todos os quatro testes da MANOVA a um fator para comparar os três métodos relativamente às variáveis estudadas

```
# Testando a Hipótese Nula
```

```
summary(manova.model, test = "Wilks")
```

```

          Df Wilks approx F num Df den Df  Pr(>F)
Method      2  0,22      8,49      8    60 1,2e-07 ***
Residuals 33
---
Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
```

```
summary(manova.model, test = "Pillai")
```

```

          Df Pillai approx F num Df den Df  Pr(>F)
Method      2  0,864      5,9      8    62 1,3e-05 ***
Residuals 33
---
Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

```

```
summary(manova.model, test = "Hotelling-Lawley")
```

```

          Df Hotelling-Lawley approx F num Df den Df  Pr(>F)
Method      2           3,16    11,5      8    58 1,4e-09 ***
Residuals 33
---
Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

```

```
summary(manova.model, test = "Roy")
```

```

          Df  Roy approx F num Df den Df  Pr(>F)
Method      2 3,04    23,5      4    31 5,1e-09 ***
Residuals 33
---
Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

```

Todos os testes foram altamente significativos (***), sugerindo que existem diferenças significativas entre os métodos para pelo menos uma das variáveis.

0.1 e) Faça análise a posteriori se os testes multivariados rejeitarem a hipótese nula. Faça a conclusão da análise.

```
# AROMA
TukeyHSD(x = aov(AROMA ~ Method, data = df.WIDE), "Method", conf.level = 0.95)
```

```

Tukey multiple comparisons of means
 95% family-wise confidence level

```

```
Fit: aov(formula = AROMA ~ Method, data = df.WIDE)
```

```

$Method
      diff    lwr    upr p adj
M2-M1 -0,12 -0,76 0,51  0,88
M3-M1 -0,41 -1,05 0,23  0,27
M3-M2 -0,28 -0,92 0,36  0,53

```

```
# SABOR
TukeyHSD(x = aov(SABOR ~ Method, data = df.WIDE), "Method", conf.level = 0.95)
```

```

Tukey multiple comparisons of means
 95% family-wise confidence level

```

```
Fit: aov(formula = SABOR ~ Method, data = df.WIDE)
```

```
$Method
```

	diff	lwr	upr	p	adj
M2-M1	-0,48	-0,97	0,021	0,06	
M3-M1	-0,88	-1,37	-0,379	0,00	
M3-M2	-0,40	-0,90	0,096	0,13	

```
# TEXTURA
```

```
TukeyHSD(x = aov(TEXTURA ~ Method, data = df.WIDE), "Method", conf.level = 0.95)
```

Tukey multiple comparisons of means
95% family-wise confidence level

```
Fit: aov(formula = TEXTURA ~ Method, data = df.WIDE)
```

```
$Method
```

	diff	lwr	upr	p	adj
M2-M1	-0,13	-0,7275	0,46	0,85	
M3-M1	0,47	-0,1275	1,06	0,15	
M3-M2	0,60	0,0059	1,19	0,05	

```
# UMIDADE
```

```
TukeyHSD(x = aov(UMIDADE ~ Method, data = df.WIDE), "Method", conf.level = 0.95)
```

Tukey multiple comparisons of means
95% family-wise confidence level

```
Fit: aov(formula = UMIDADE ~ Method, data = df.WIDE)
```

```
$Method
```

	diff	lwr	upr	p	adj
M2-M1	-0,11	-0,68	0,46	0,89	
M3-M1	0,25	-0,32	0,82	0,53	
M3-M2	0,36	-0,21	0,93	0,28	

Após a análise a posteriori, verificou-se que os métodos aplicados só apresentaram diferenças no “SABOR” e “TEXTURA”.

1 Códigos Utilizados

Para ter acesso aos códigos utilizados, acesse o link [CODE UTILIZADO](#).