**Serviço Público Federal**

**Universidade Federal do Pará**

**Instituto de Ciências Exatas e Naturais**

**Faculdade de Estatística**

Dionisio Alves da Silva Neto

Matrícula: 202007840008

**Atividade 2 de Análise Multivariada II**:

Análise de Variância Multivariada (MANOVA)

Belém, PA

2022

**Questão 1: Pontuação dos juízes sobre peixes preparados por três métodos.**

**Item a):** Transformação da tabela *wide* para o formato *long*.

Para a criação do novo formato tabular, necessitou-se de alterações no padrão de ponto para vírgula nos dados disponibilizados. Posteriormente, utilizando funções e manipulação de objetos na linguagem R, foi possível construir o formato *long* para as pontuações dos juízes sobre os peixes preparados considerando os critérios: Aroma (y1), Sabor (y2), Textura (y3) e Umidade (y4). Também, foi elaborada uma nova coluna indicando os métodos empregados no preparo dos peixes (1, 2 e 3). A **Tabela 1** aborda a criação do novo banco de dados, o qual foi estruturado em 36 linhas e 4 colunas no formato *long*.

**Tabela 1**: Banco de dados da pontuação dos juízes sobre os peixes preparados, por método.

| y1 | y2 | y3 | y4 | método |
| --- | --- | --- | --- | --- |
| 5,40 | 6,00 | 6,30 | 6,70 | 1 |
| 5,20 | 6,20 | 6,00 | 5,80 | 1 |
| 6,10 | 5,90 | 6,00 | 7,00 | 1 |
| 4,80 | 5,00 | 4,90 | 5,00 | 1 |
| 5,00 | 5,70 | 5,00 | 6,50 | 1 |
| 5,70 | 6,10 | 6,00 | 6,60 | 1 |
| 6,00 | 6,00 | 5,80 | 6,00 | 1 |
| 4,00 | 5,00 | 4,00 | 5,00 | 1 |
| 5,70 | 5,40 | 4,90 | 5,00 | 1 |
| 5,60 | 5,20 | 5,40 | 5,80 | 1 |
| 5,80 | 6,10 | 5,20 | 6,40 | 1 |
| 5,30 | 5,90 | 5,80 | 6,00 | 1 |
| 5,00 | 5,30 | 5,30 | 6,50 | 2 |
| 4,80 | 4,90 | 4,20 | 5,60 | 2 |
| 3,90 | 4,00 | 4,40 | 5,00 | 2 |
| 4,00 | 5,10 | 4,80 | 5,80 | 2 |
| 5,60 | 5,40 | 5,10 | 6,20 | 2 |
| 6,00 | 5,50 | 5,70 | 6,00 | 2 |
| 5,20 | 4,80 | 5,40 | 6,00 | 2 |
| 5,30 | 5,10 | 5,80 | 6,40 | 2 |
| 5,90 | 6,10 | 5,70 | 6,00 | 2 |
| 6,10 | 6,00 | 6,10 | 6,20 | 2 |
| 6,20 | 5,70 | 5,90 | 6,00 | 2 |
| 5,10 | 4,90 | 5,30 | 4,80 | 2 |
| 4,80 | 5,00 | 6,50 | 7,00 | 3 |
| 5,40 | 5,00 | 6,00 | 6,40 | 3 |
| 4,90 | 5,10 | 5,90 | 6,50 | 3 |
| 5,70 | 5,20 | 6,40 | 6,40 | 3 |
| 4,20 | 4,60 | 5,30 | 6,30 | 3 |
| 6,00 | 5,30 | 5,80 | 6,40 | 3 |
| 5,10 | 5,20 | 6,20 | 6,50 | 3 |
| 4,80 | 4,60 | 5,70 | 5,70 | 3 |
| 5,30 | 5,40 | 6,80 | 6,60 | 3 |
| 4,60 | 4,40 | 5,70 | 5,60 | 3 |
| 4,50 | 4,00 | 5,00 | 5,90 | 3 |
| 4,40 | 4,20 | 5,60 | 5,50 | 3 |

Fonte: Construído pelo autor, 2022.

**Item b):** Análise Exploratória dos dados (AED).

Na fase inicial da Análise de variância multivariada (ANOVA) foi realizada a estatística descritiva dos dados no intuito se ter uma ideia inicial do banco de dados das notas atribuídas pelos juízes para cada variável (y1, y2, y3 e y4), tanto de uma forma geral, quanto em relação a cada método.

**1. b.1.1 Medidas descritivas do banco em geral**

**Tabela 2**: Valores descritivos de cada variável do banco de dados, por estatística.

| Estatística | y1 | y2 | y3 | y4 |
| --- | --- | --- | --- | --- |
| Mínimo | 3,900 | 4,000 | 4,000 | 4,800 |
| Q1 | 4,80, | 4,975 | 5,175 | 5,775 |
| Mediana | 5,250 | 5,200 | 5,700 | 6,000 |
| Média | 5,206 | 5,258 | 5,553 | 6,031 |
| Q3 | 5,700 | 5,750 | 6,000 | 6,425 |
| Máximo | 6,200 | 6,200 | 6,800 | 7,000 |

Fonte: Construído pelo autor, 2022.

**1. b.1.2 Vetor de médias geral**

|  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- |
|  | 5,206 | 5,258 | 5,553 | 6,031 |  |

**1. b.1.3 Matriz de variâncias e covariâncias geral**

|  |  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- | --- |
| y1 |  | 0,413 | 0,283 | 0,209 | 0,146 |  |
| y2 |  | 0,283 | 0,363 | 0,143 | 0,144 |  |
| y3 |  | 0,209 | 0,143 | 0,400 | 0,241 |  |
| y4 |  | 0,146 | 0,144 | 0,241 | 0,325 |  |

**1. b.1.4 Matriz de correlações geral**

|  |  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- | --- |
| y1 |  | 0,413 | 0,283 | 0,209 | 0,146 |  |
| y2 |  | 0,283 | 0,363 | 0,143 | 0,144 |  |
| y3 |  | 0,209 | 0,143 | 0,400 | 0,241 |  |
| y4 |  | 0,146 | 0,144 | 0,241 | 0,325 |  |

**1. b.2. Medidas descritivas por método**

**Tabela 3**: Valores descritivos de cada método (1, 2 e 3) para a variável aroma (y1), por estatística.

| Estatística | 1 | 2 | 3 |
| --- | --- | --- | --- |
| Mínimo | 4,0 | 3,9 | 4,2 |
| Q1 | 5,15 | 4,95 | 4,57 |
| Mediana | 5,5 | 5,25 | 4,85 |
| Média | 5,38 | 5,26 | 5,93 |
| Q3 | 5,72 | 5,93 | 5,32 |
| Máximo | 6,10 | 6,20 | 6,00 |

Fonte: Construído pelo autor, 2022.

**Tabela 4**: Valores descritivos de cada método (1, 2 e 3) para a variável sabor (y2), por estatística.

| Estatística | 1 | 2 | 3 |
| --- | --- | --- | --- |
| Mínimo | 5,00 | 4,00 | 4,00 |
| Q1 | 5,35 | 4,90 | 4,55 |
| Mediana | 5,90 | 5,20 | 5,00 |
| Média | 5,71 | 5,23 | 4,83 |
| Q3 | 6,02 | 5,55 | 5,2 |
| Máximo | 6,20 | 6,10 | 5,40 |

Fonte: Construído pelo autor, 2022.

**Tabela 5**: Valores descritivos de cada método (1, 2 e 3) para a variável textura (y3), por estatística.

| Estatística | 1 | 2 | 3 |
| --- | --- | --- | --- |
| Mínimo | 4,00 | 4,20 | 5,00 |
| Q1 | 4,97 | 5,02 | 5,68 |
| Mediana | 5,60 | 5,35 | 5,85 |
| Média | 5,44 | 5,31 | 5,91 |
| Q3 | 6,00 | 5,72 | 6,25 |
| Máximo | 6,30 | 6,10 | 6,80 |

Fonte: Construído pelo autor, 2022.

**Tabela 6**: Valores descritivos de cada método (1, 2 e 3) para a variável umidade (y4), por estatística.

| Estatística | 1 | 2 | 3 |
| --- | --- | --- | --- |
| Mínimo | 5,00 | 4,80 | 5,50 |
| Q1 | 5,60 | 5,75 | 5,85 |
| Mediana | 6,00 | 6,00 | 6,40 |
| Média | 5,98 | 5,88 | 6,23 |
| Q3 | 6,52 | 6,20 | 6,50 |
| Máximo | 7,00 | 6,50 | 7,00 |

Fonte: Construído pelo autor, 2022.

**1. b.2.1 Vetor de médias para o método 1**

|  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- |
|  | 5,38 | 5,71 | 5,44 | 5,98 |  |

**1. b.2.2 Vetor de médias para o método 2**

|  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- |
|  | 5,26 | 5,23 | 5,31 | 5,88 |  |

**1. b.2.3 Vetor de médias para o método 3**

|  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- |
|  | 5,93 | 4,83 | 5,91 | 6,23 |  |

**1. b.2.4 Matriz de variâncias e covariâncias para o método 1**

|  |  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- | --- |
| y1 |  | 0,3420 | 0,1540 | 0,2620 | 0,2310 |  |
| y2 |  | 0,1540 | 0,1950 | 0,2300 | 0,2280 |  |
| y3 |  | 0,2620 | 0,2300 | 0,4410 | 0,3270 |  |
| y4 |  | 0,2310 | 0,2280 | 0,3270 | 0,4850 |  |

**1. b.2.5 Matriz de variâncias e covariâncias para o método 2**

|  |  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- | --- |
| y1 |  | 0,5830 | 0,3590 | 0,3720 | 0,1920 |  |
| y2 |  | 0,3590 | 0,3300 | 0,2500 | 0,1780 |  |
| y3 |  | 0,3720 | 0,2500 | 0,3540 | 0,1610 |  |
| y4 |  | 0,1920 | 0,1780 | 0,1610 | 0,2680 |  |

**1. b.2.6 Matriz de variâncias e covariâncias para o método 3**

|  |  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- | --- |
| y1 |  | 0,2950 | 0,1950 | 0,1550 | 0,1110 |  |
| y2 |  | 0,1950 | 0,2120 | 0,1900 | 0,1630 |  |
| y3 |  | 0,1550 | 0,1900 | 0,2610 | 0,1520 |  |
| y4 |  | 0,1110 | 0,1630 | 0,1520 | 0,2080 |  |

**1. b.2.7 Matriz de correlações para o método 1**

|  |  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- | --- |
| y1 |  | 1,0000 | 0,5950 | 0,6740 | 0,5670 |  |
| y2 |  | 0,5950 | 1,0000 | 0,7820 | 0,7420 |  |
| y3 |  | 0,6740 | 0,7820 | 1,0000 | 0,7070 |  |
| y4 |  | 0,5670 | 0,7420 | 0,7070 | 1,0000 |  |

**1. b.2.8 Matriz de correlações para o método 2**

|  |  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- | --- |
| y1 |  | 1,0000 | 0,8190 | 0,8200 | 0,4850 |  |
| y2 |  | 0,8190 | 1,0000 | 0,7310 | 0,6000 |  |
| y3 |  | 0,820 | 0,7310 | 1,0000 | 0,5240 |  |
| y4 |  | 0,485 | 0,6000 | 0,5240 | 1,0000 |  |

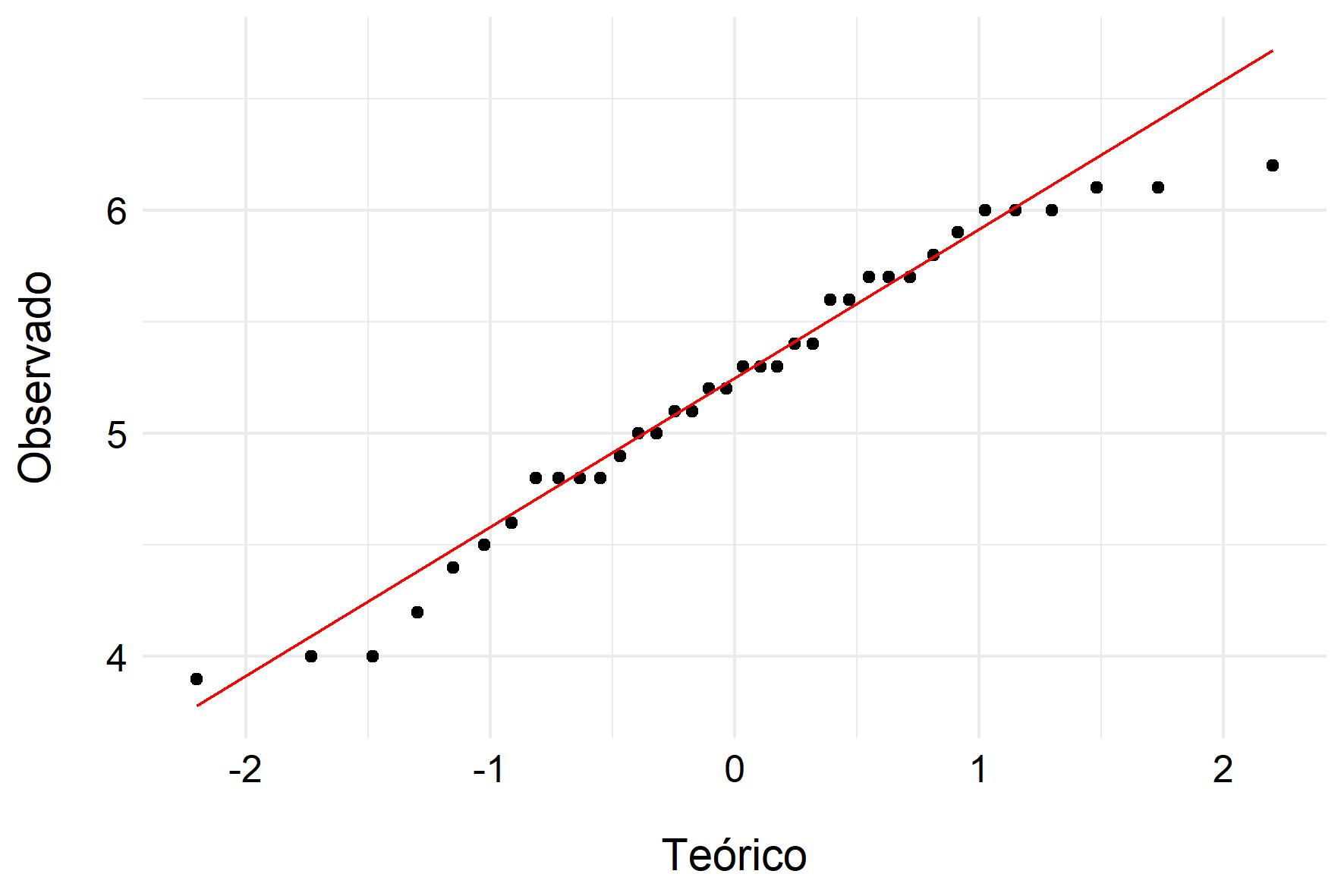
**1. b.2.9 Matriz de correlações para o método 3**

|  |  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- | --- |
| y1 |  | 1,0000 | 0,7790 | 0,5580 | 0,4480 |  |
| y2 |  | 0,7790 | 1,0000 | 0,8080 | 0,7790 |  |
| y3 |  | 0,5580 | 0,8080 | 1,0000 | 0,6510 |  |
| y4 |  | 0,4480 | 0,7790 | 0,6510 | 1,0000 |  |

**b.3 Gráficos Univariados**

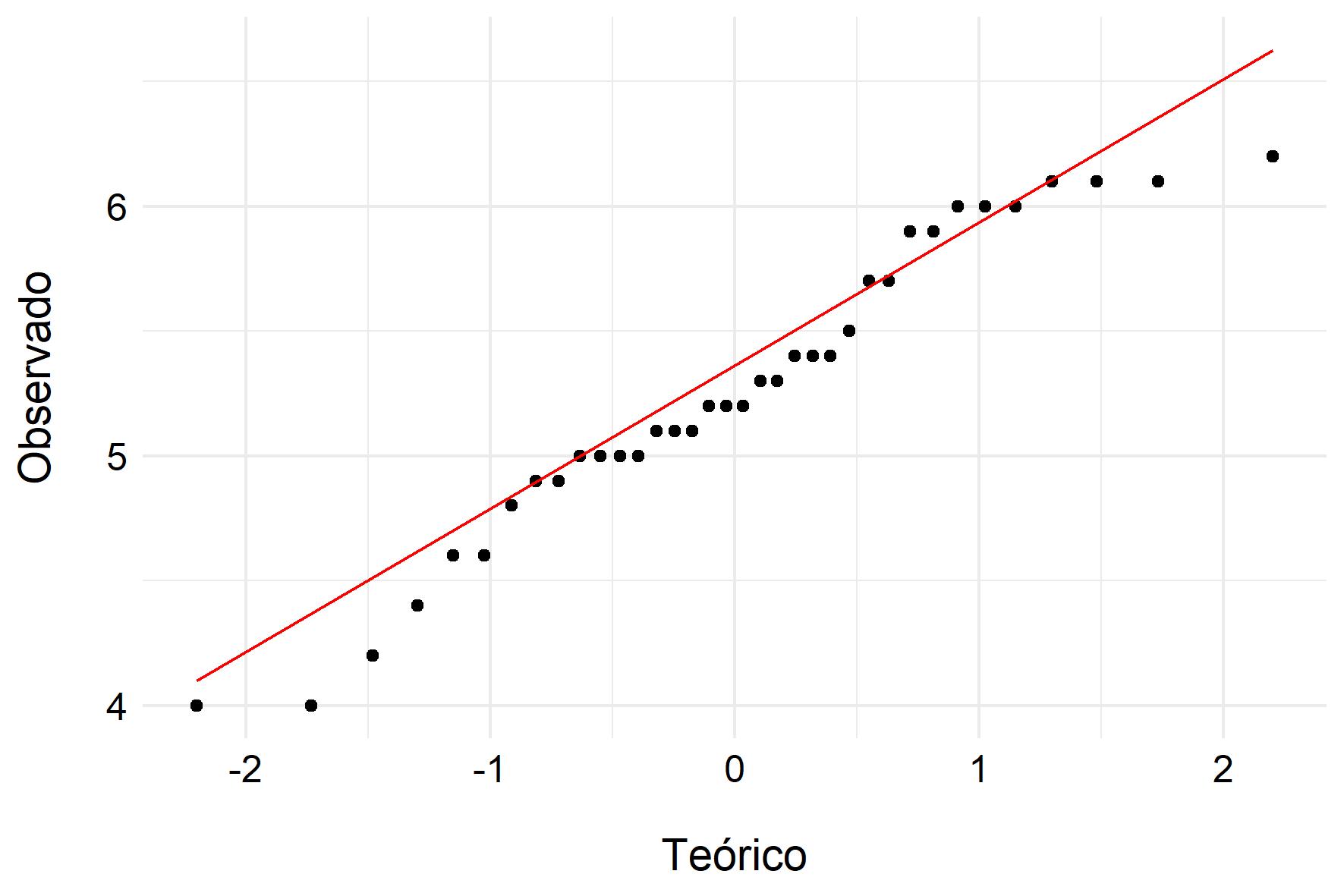
**1. b.3.1 Q-Q plot**

**Figura 1**: Figura Q-Q plot para a variável aroma (y1).



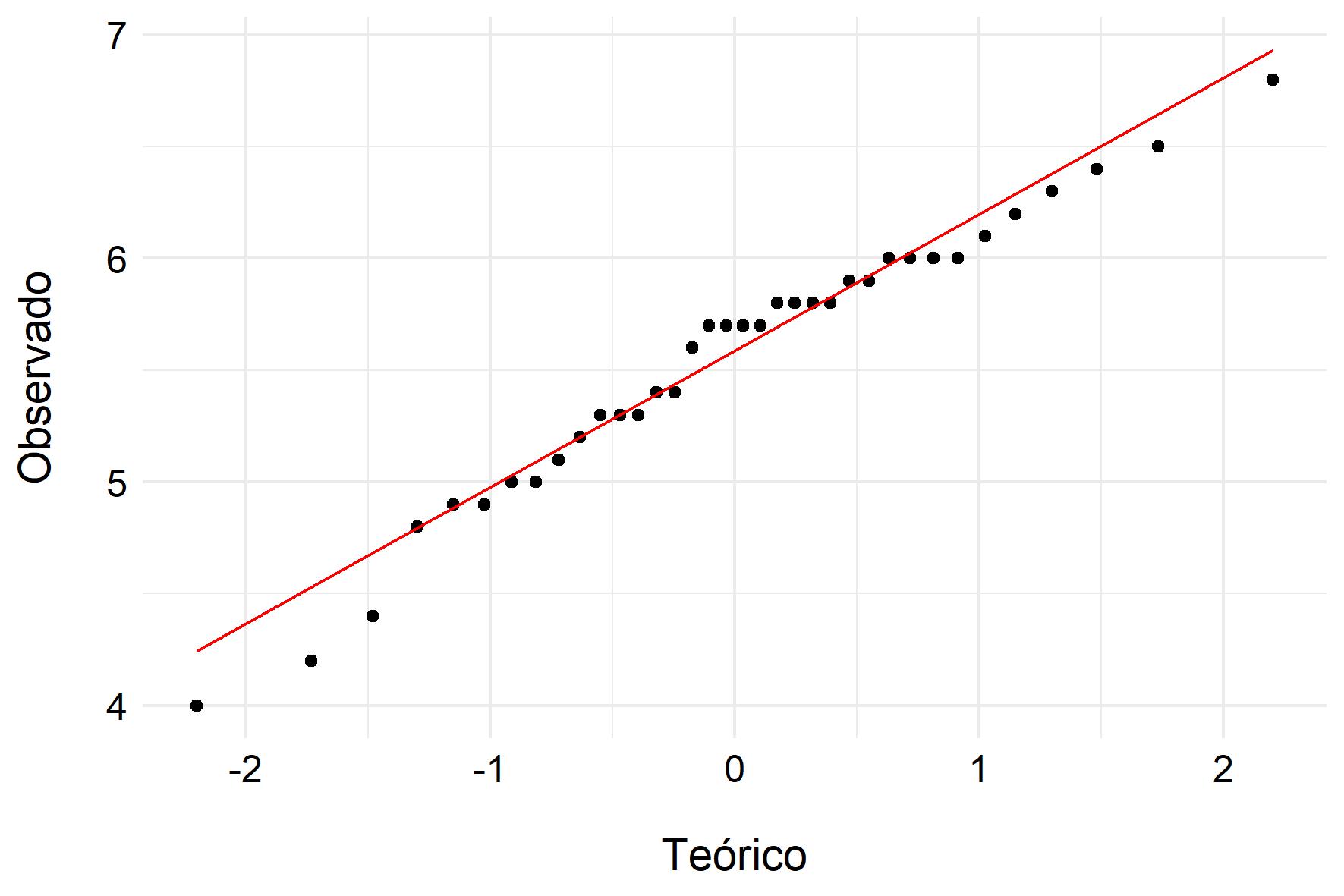
Fonte: Construído pelo autor, 2022.

**Figura 2**: Figura Q-Qplot para a variável sabor (y2).



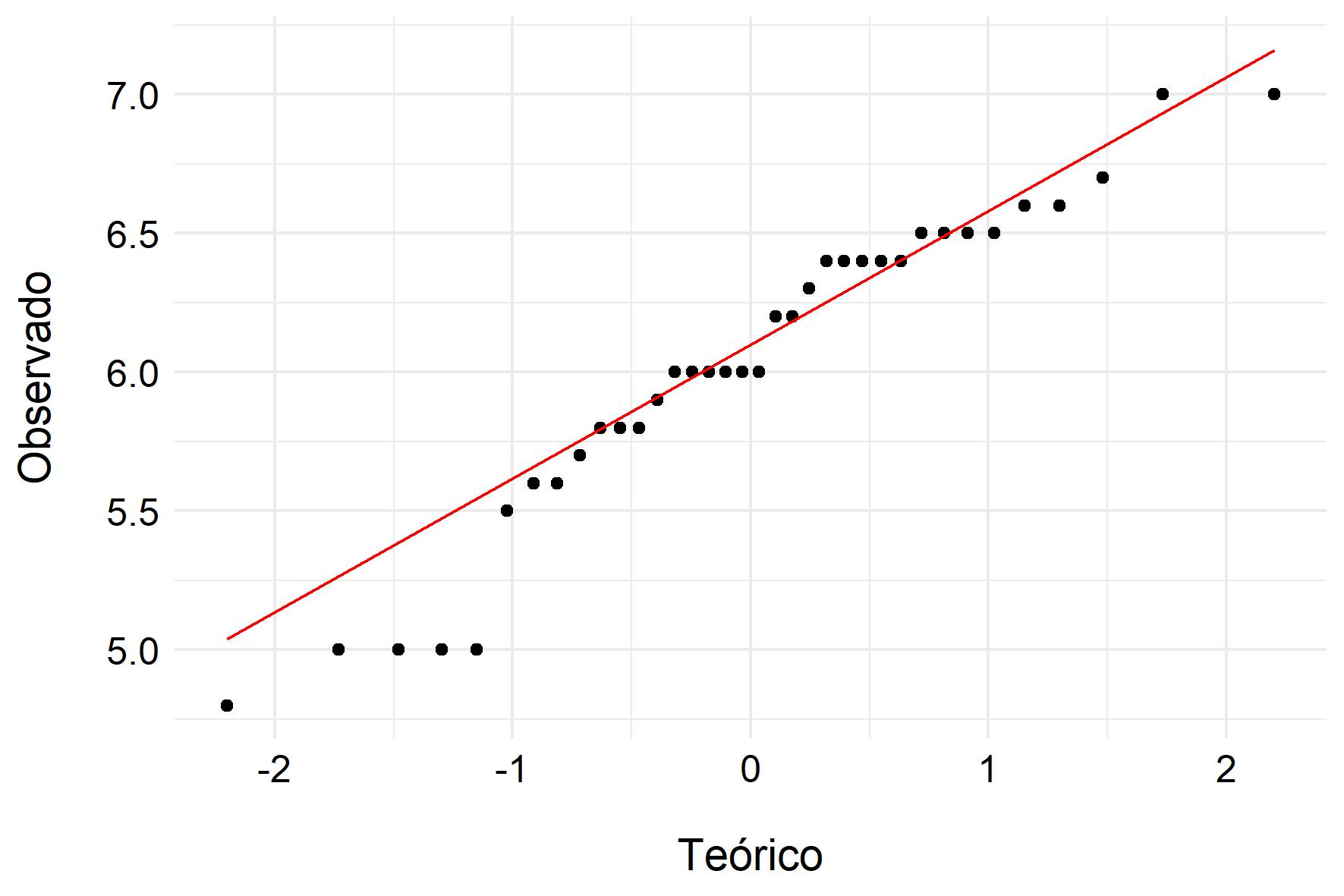
Fonte: Construído pelo autor, 2022.

**Figura 3**: Figura Q-Qplot para a variável textura (y3).



Fonte: Construído pelo autor, 2022.

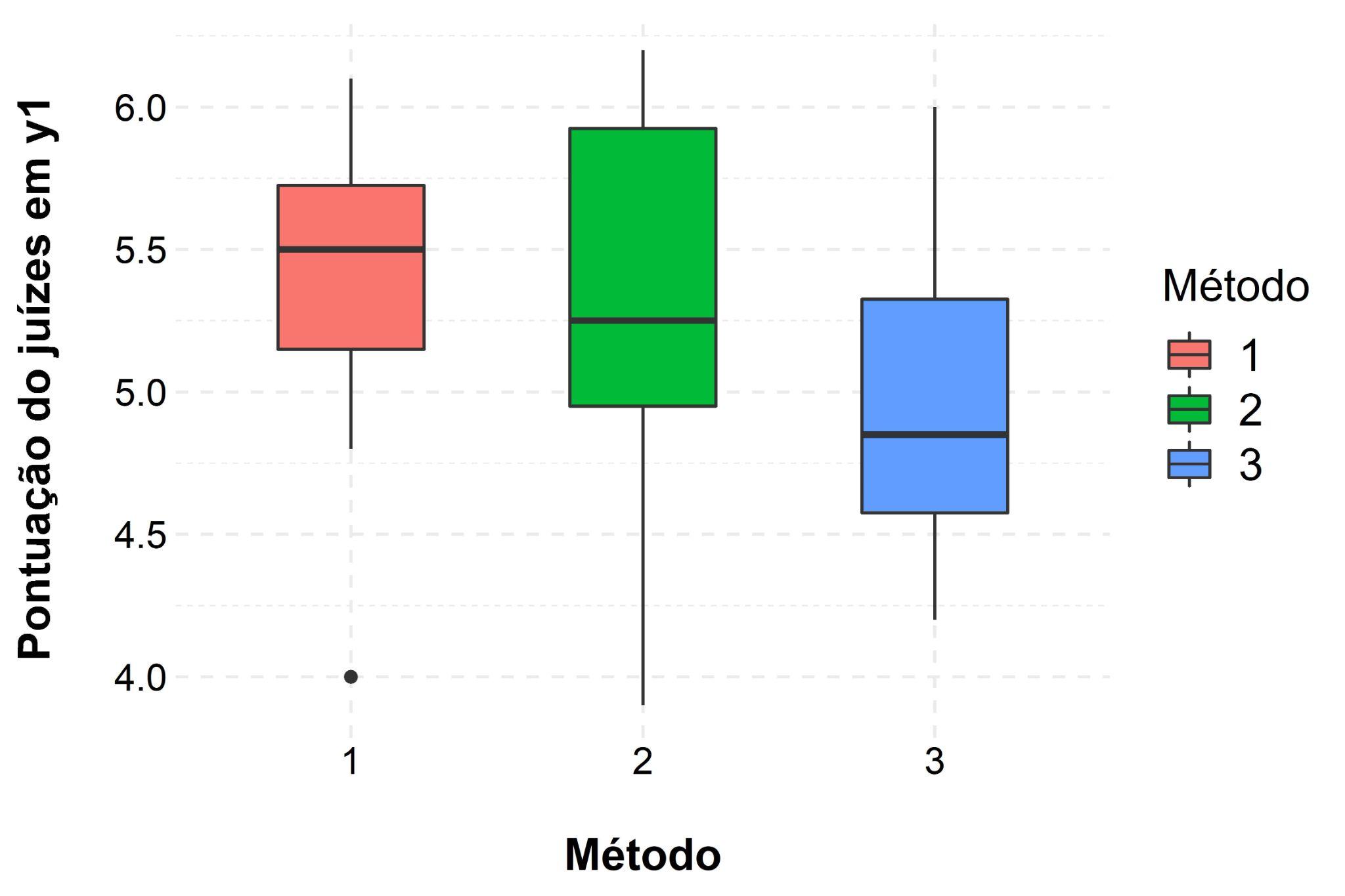
**Figura 4**: Figura Q-Qplot para a variável umidade (y4).



Fonte: Construído pelo autor, 2022.

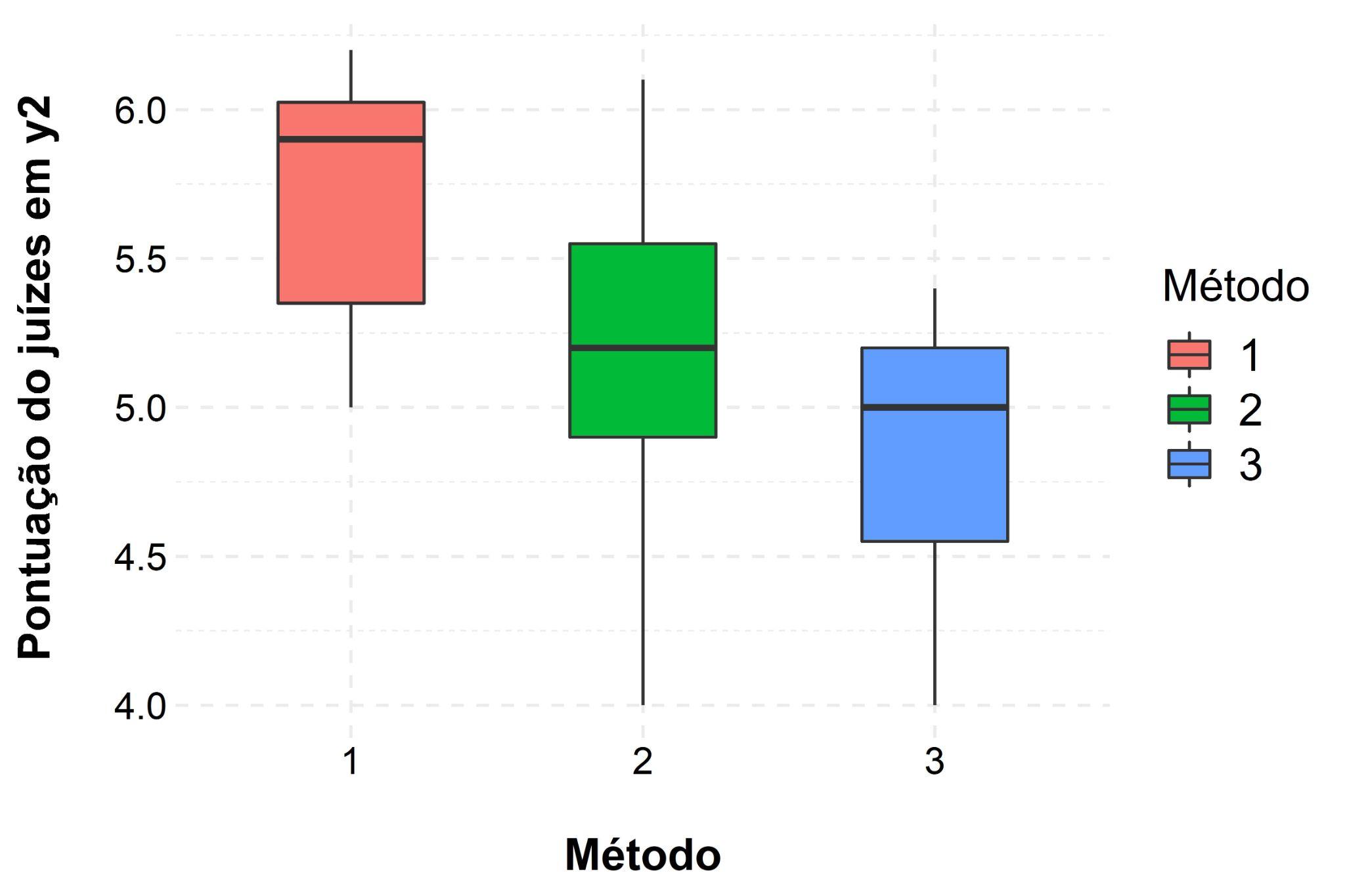
**1. b.3.2 Boxplots**

**Figura 5**: Figura Boxplot para a variável aroma (y1).



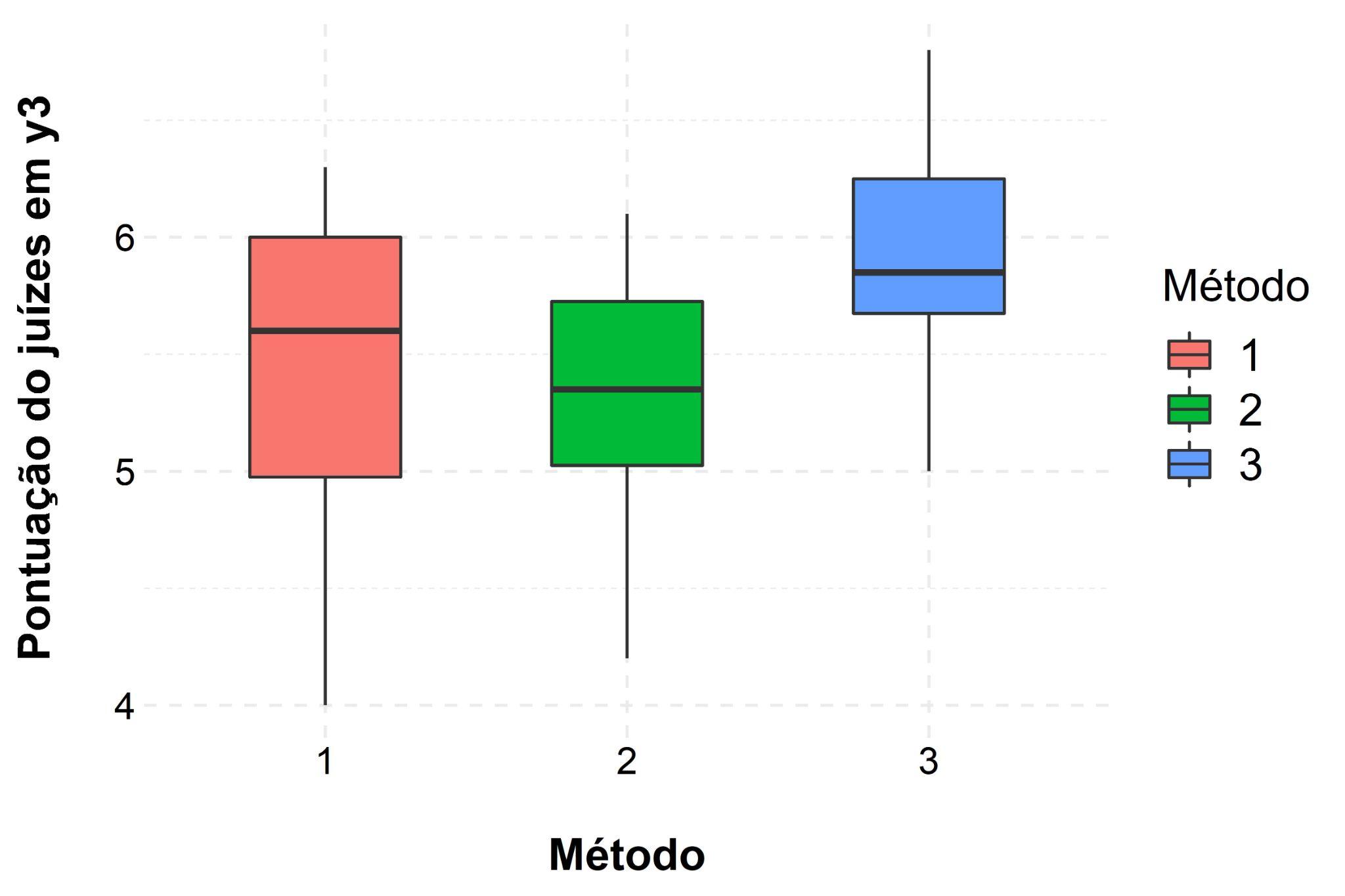
Fonte: Construído pelo autor, 2022.

**Figura 6**: Figura Boxplot para a variável sabor (y2).



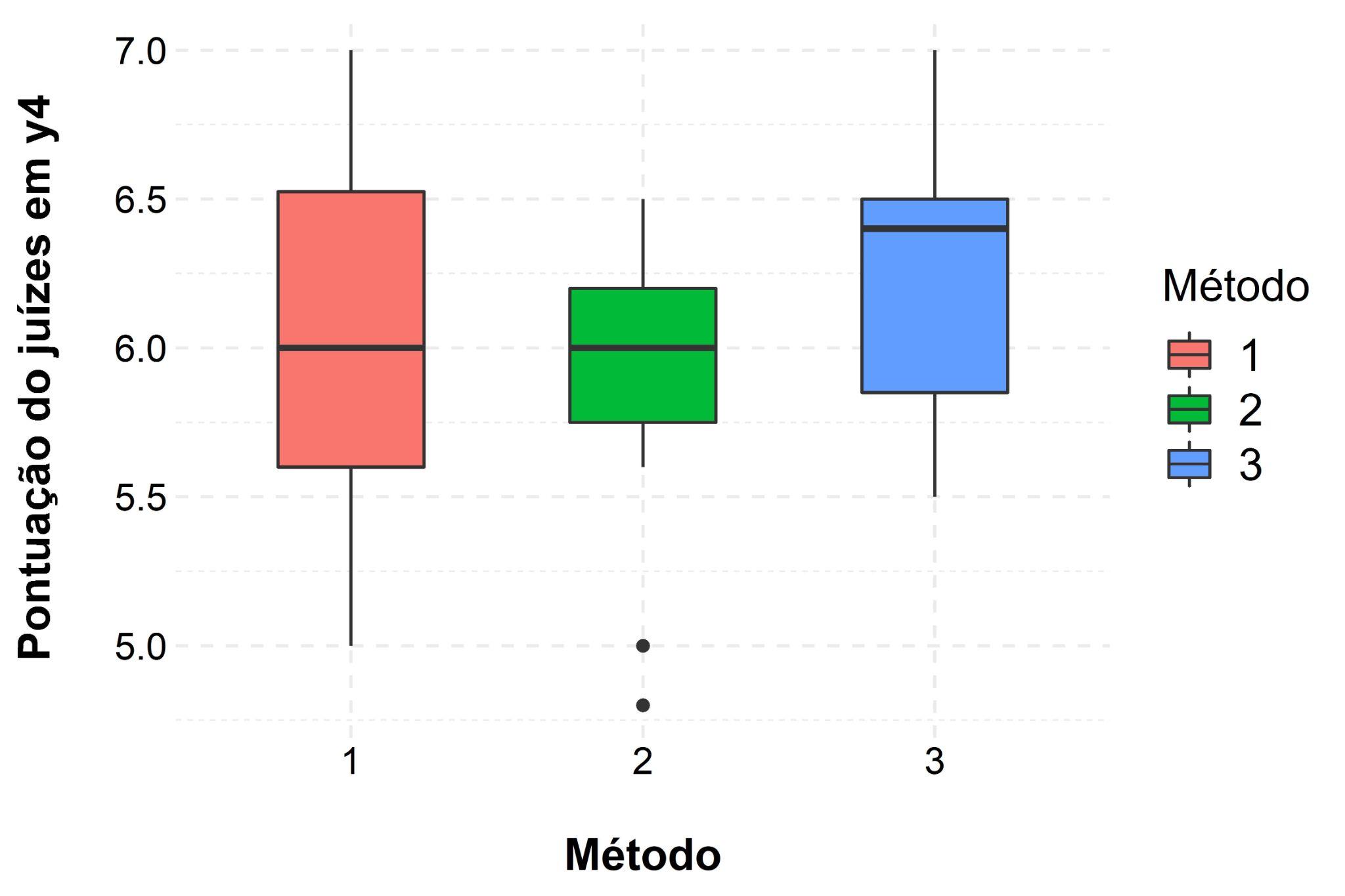
Fonte: Construído pelo autor, 2022.

**Figura 7**: Figura Boxplot para a variável textura (y3).



Fonte: Construído pelo autor, 2022.

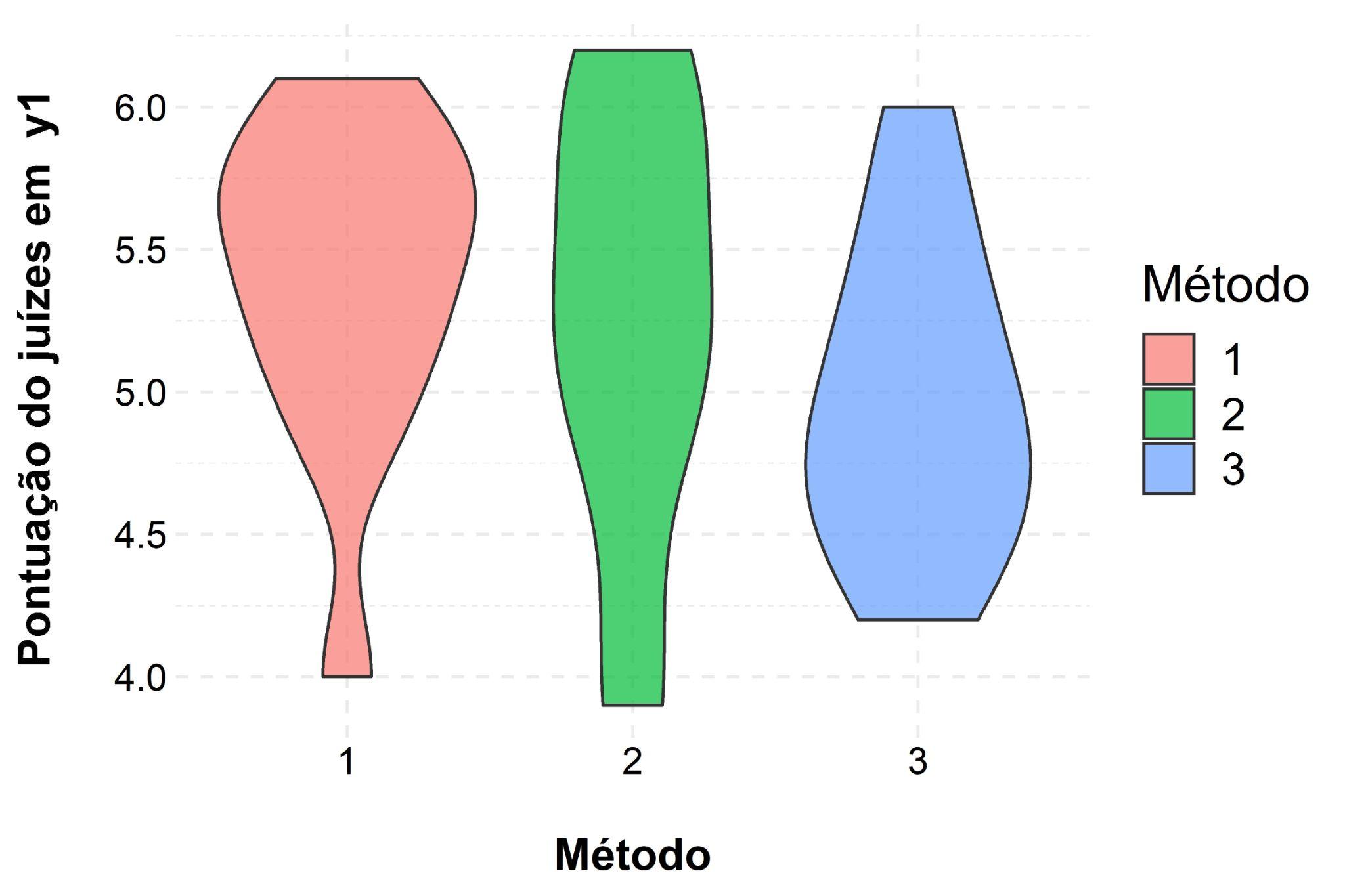
**Figura 8**: Figura Boxplot para a variável umidade (y4).



Fonte: Construído pelo autor, 2022.

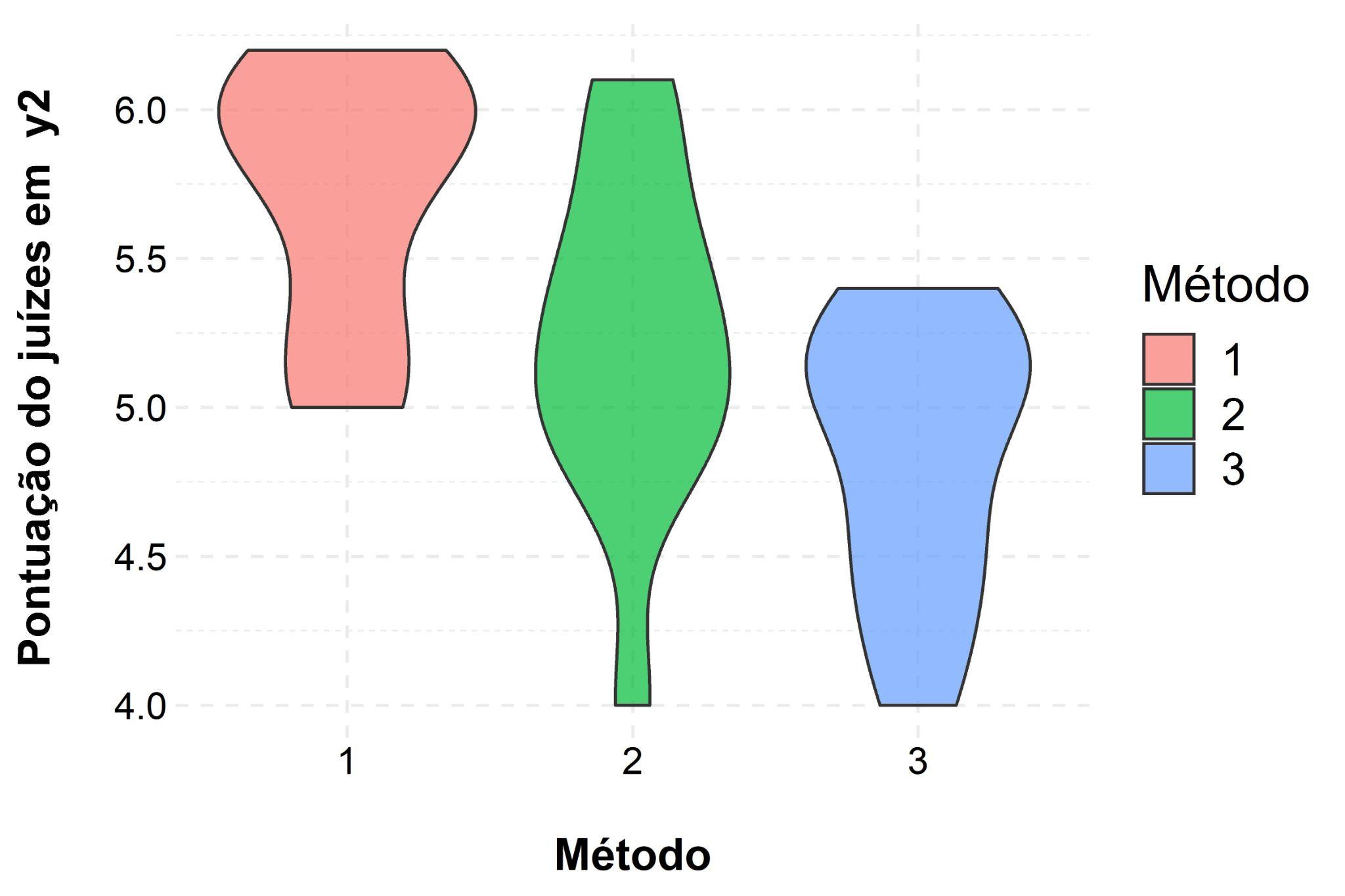
**1. b.3.3 Gráficos de Violino**

**Figura 9**: Figura de Violino para a variável aroma (y1).



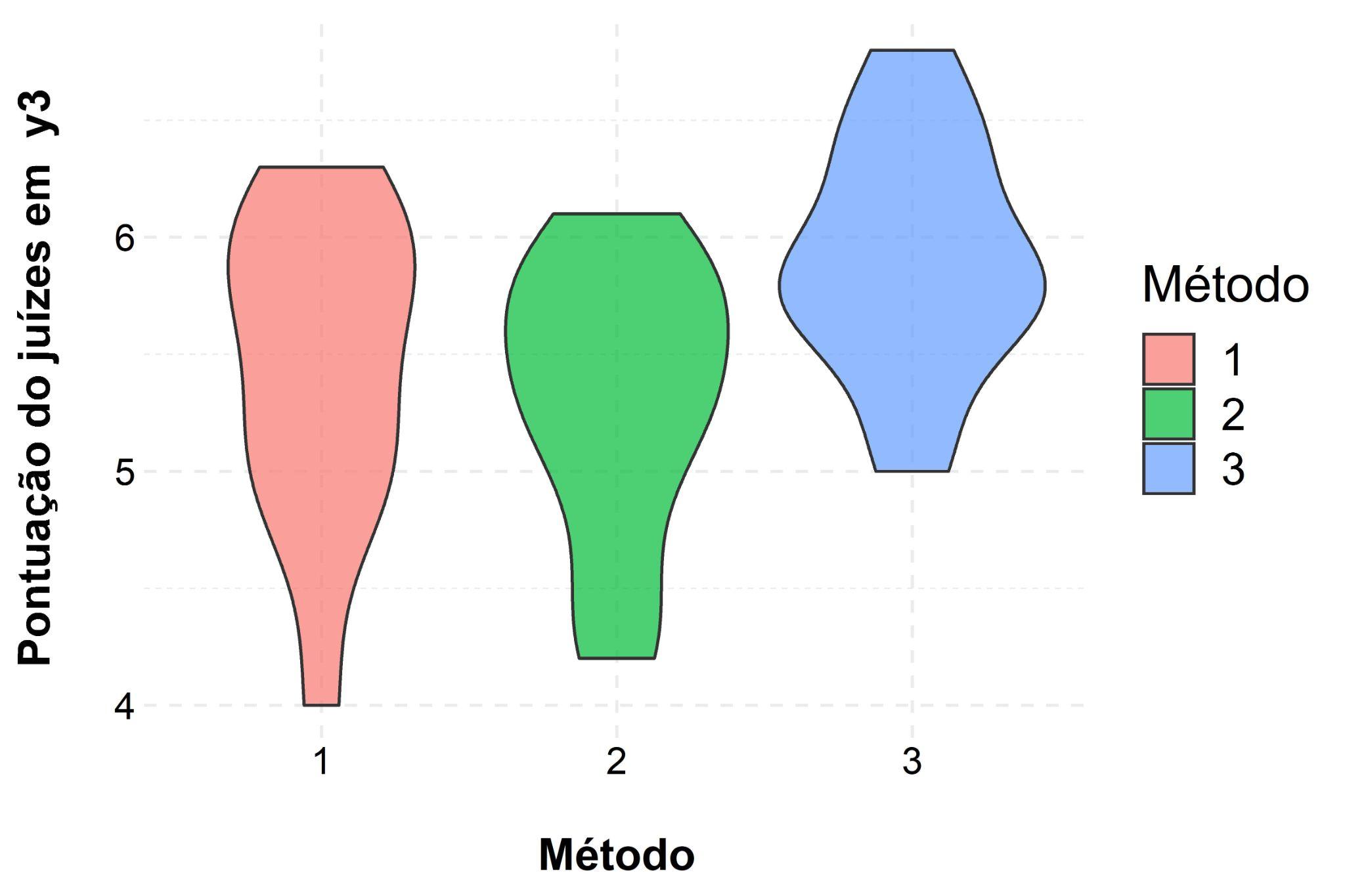
Fonte: Construído pelo autor, 2022.

**Figura 10**: Figura de Violino para a variável sabor (y2).



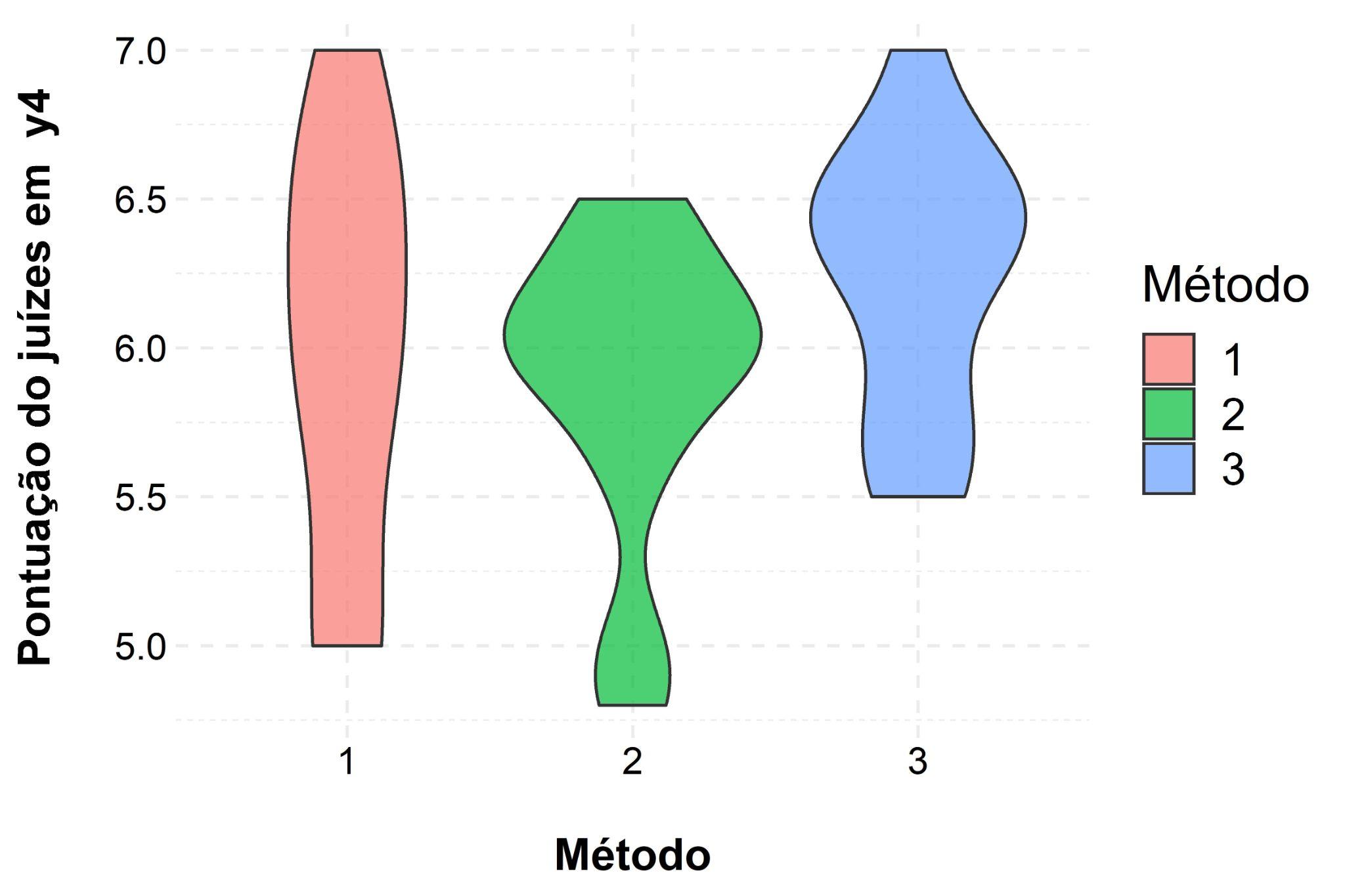
Fonte: Construído pelo autor, 2022.

**Figura 11**: Figura de Violino para a variável textura (y3).



Fonte: Construído pelo autor, 2022.

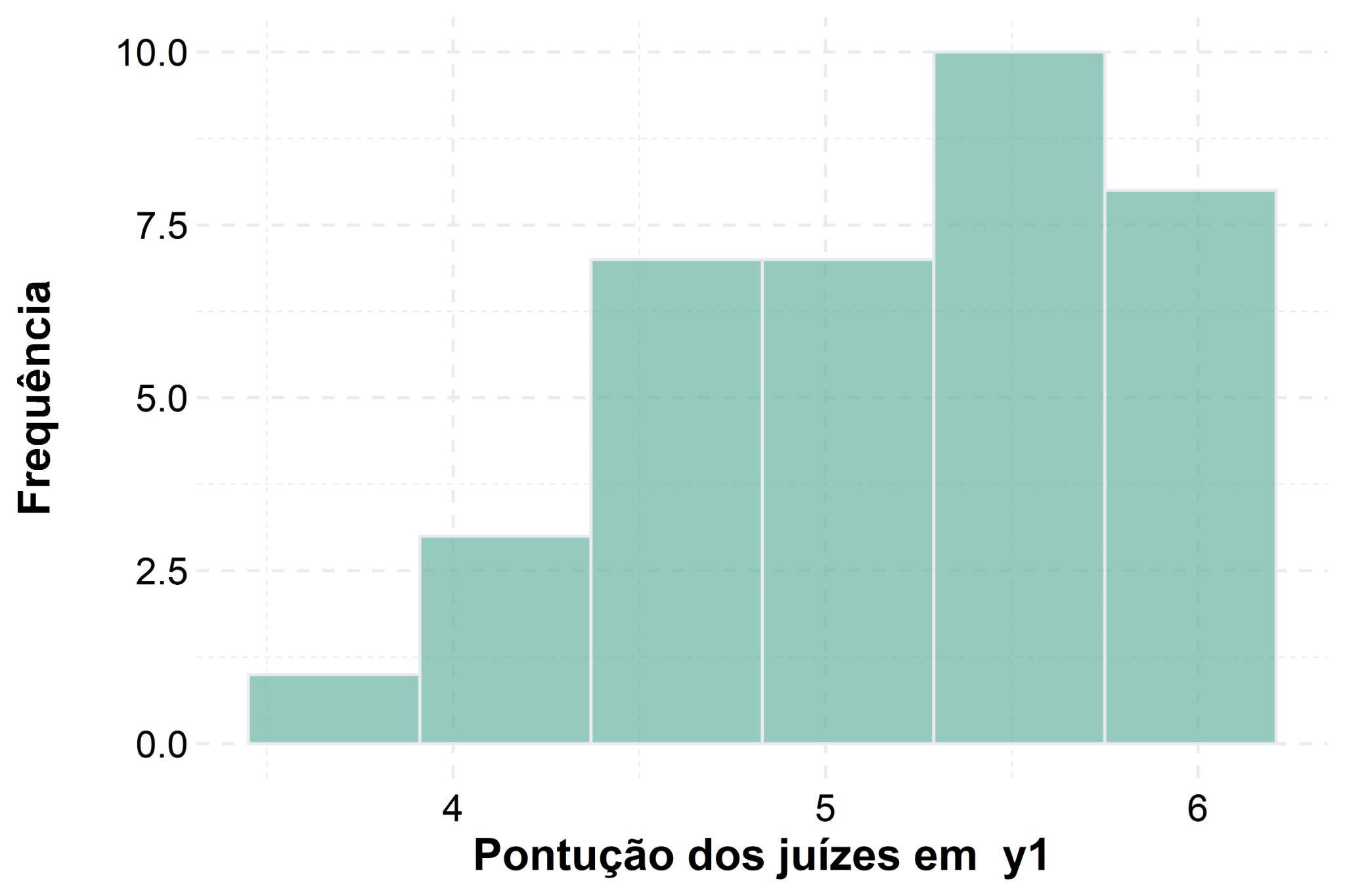
**Figura 12**: Figura de Violino para a variável umidade (y4).



Fonte: Construído pelo autor, 2022.

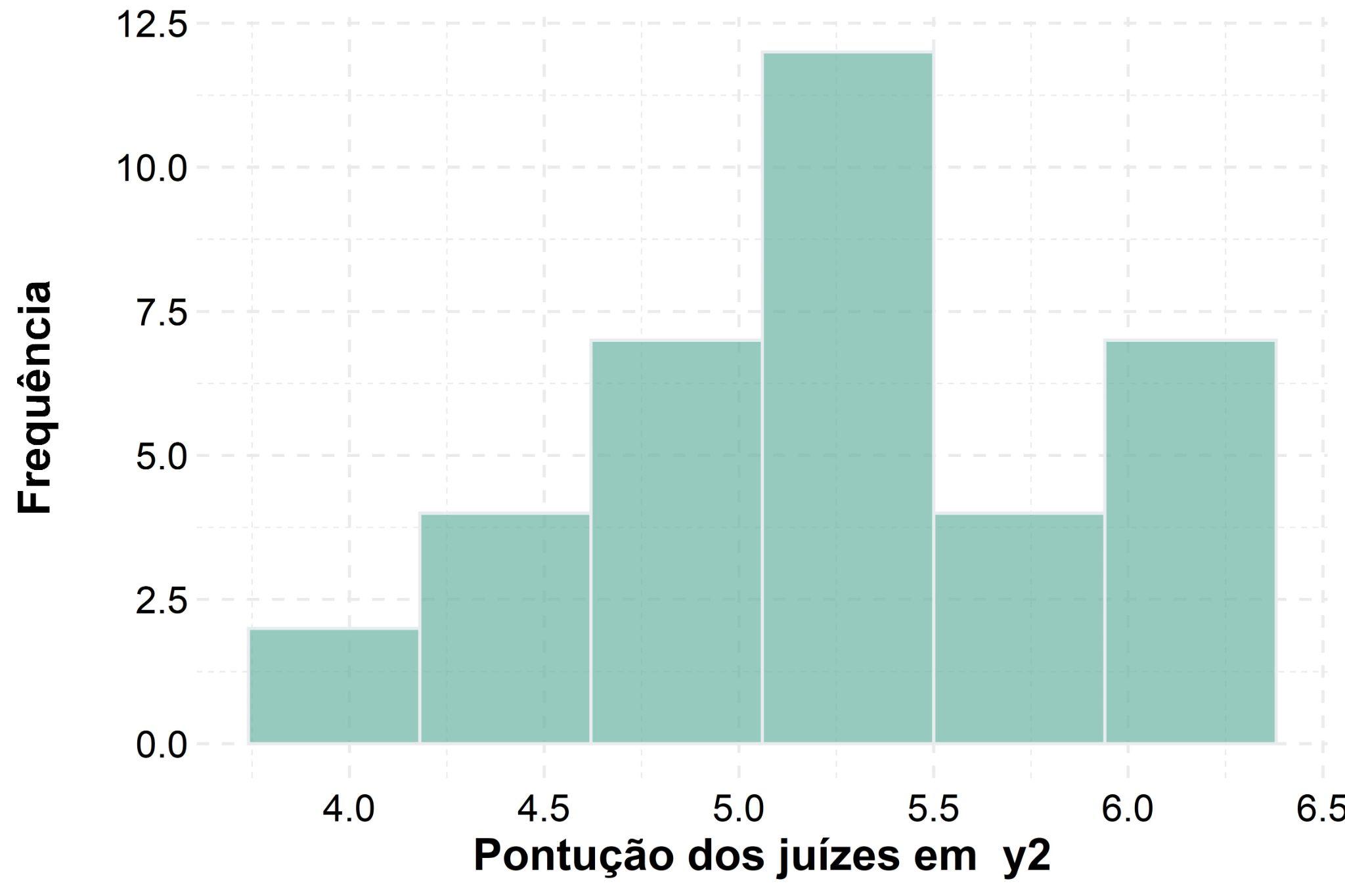
**1. b.3.4 Histograma**

**Figura 13**: Histograma para a variável aroma (y1).



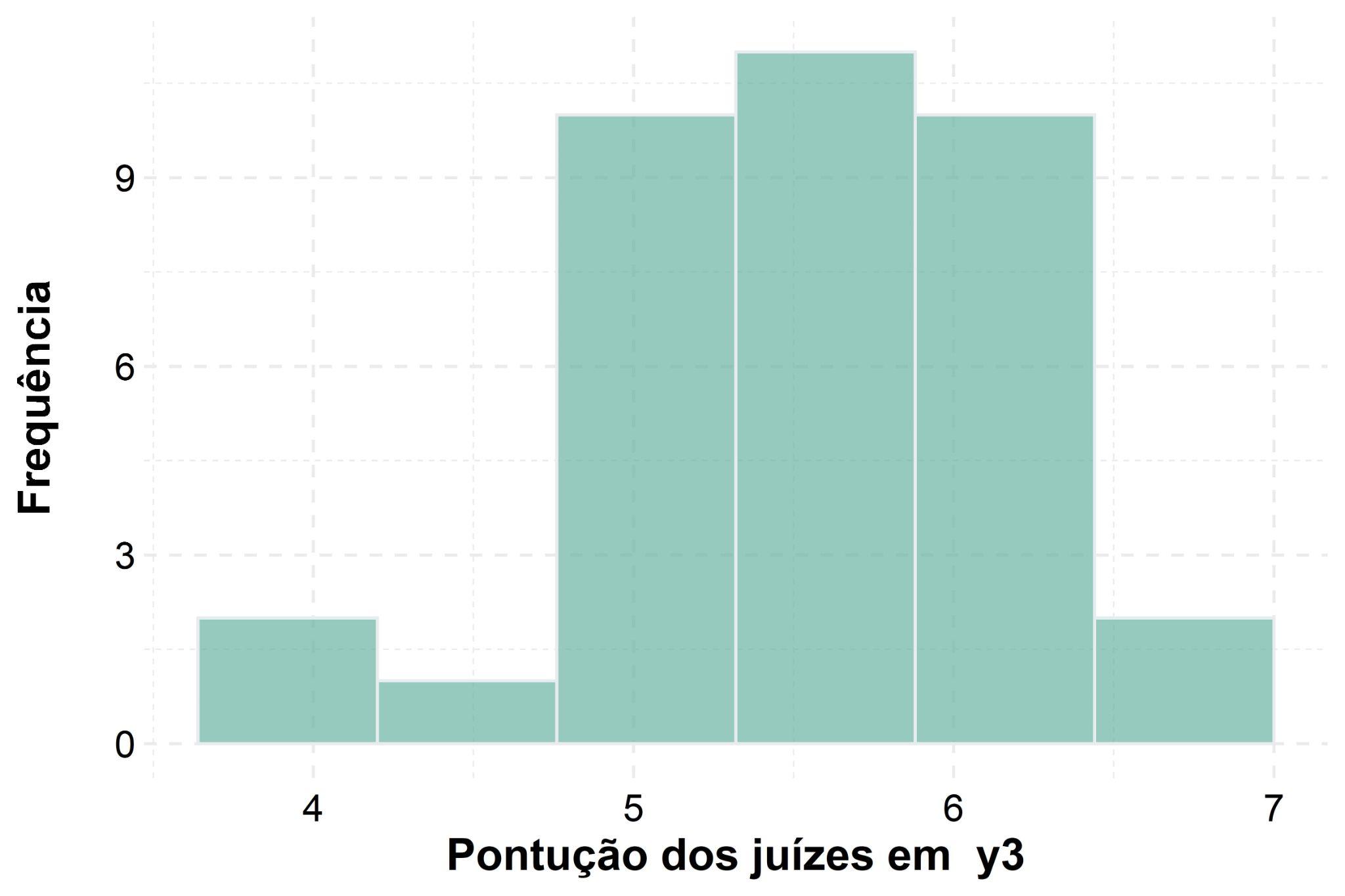
Fonte: Construído pelo autor, 2022.

**Figura 14**: Histograma para a variável sabor (y2).



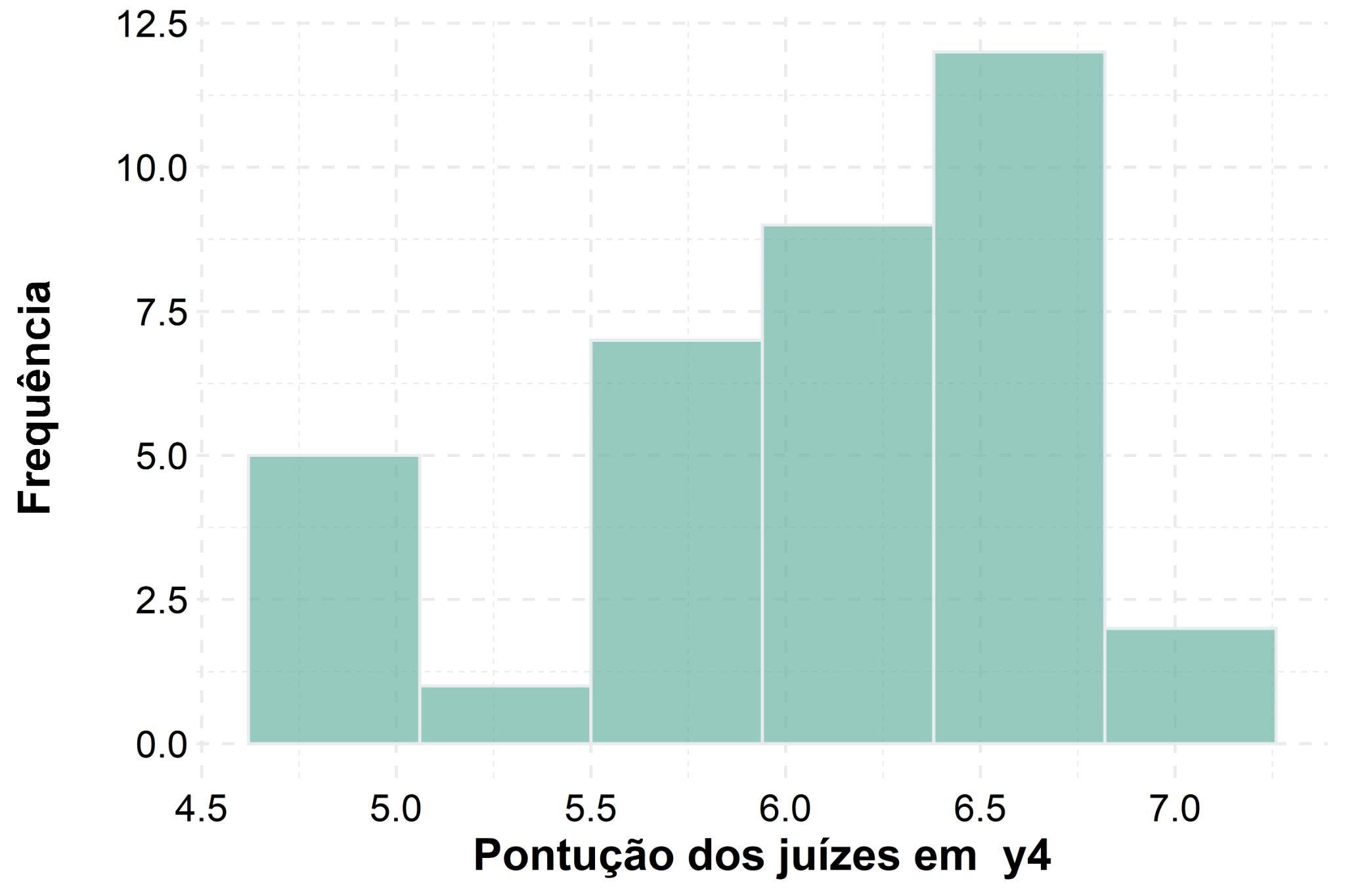
Fonte: Construído pelo autor, 2022.

**Figura 15**: Histograma para a variável textura (y3).



Fonte: Construído pelo autor, 2022.

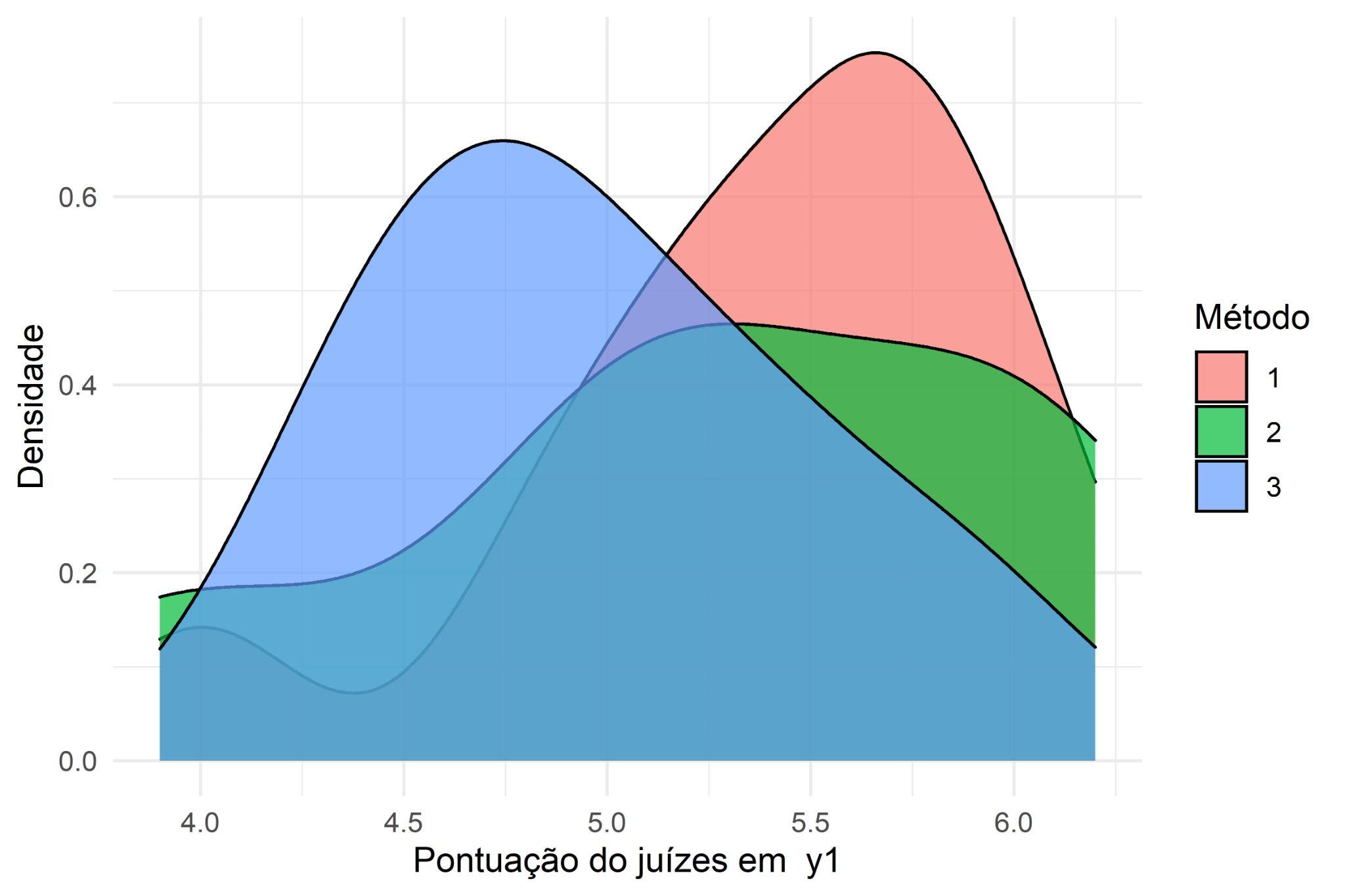
**Figura 16**: Histograma para a variável umidade (y4).



Fonte: Construído pelo autor, 2022.

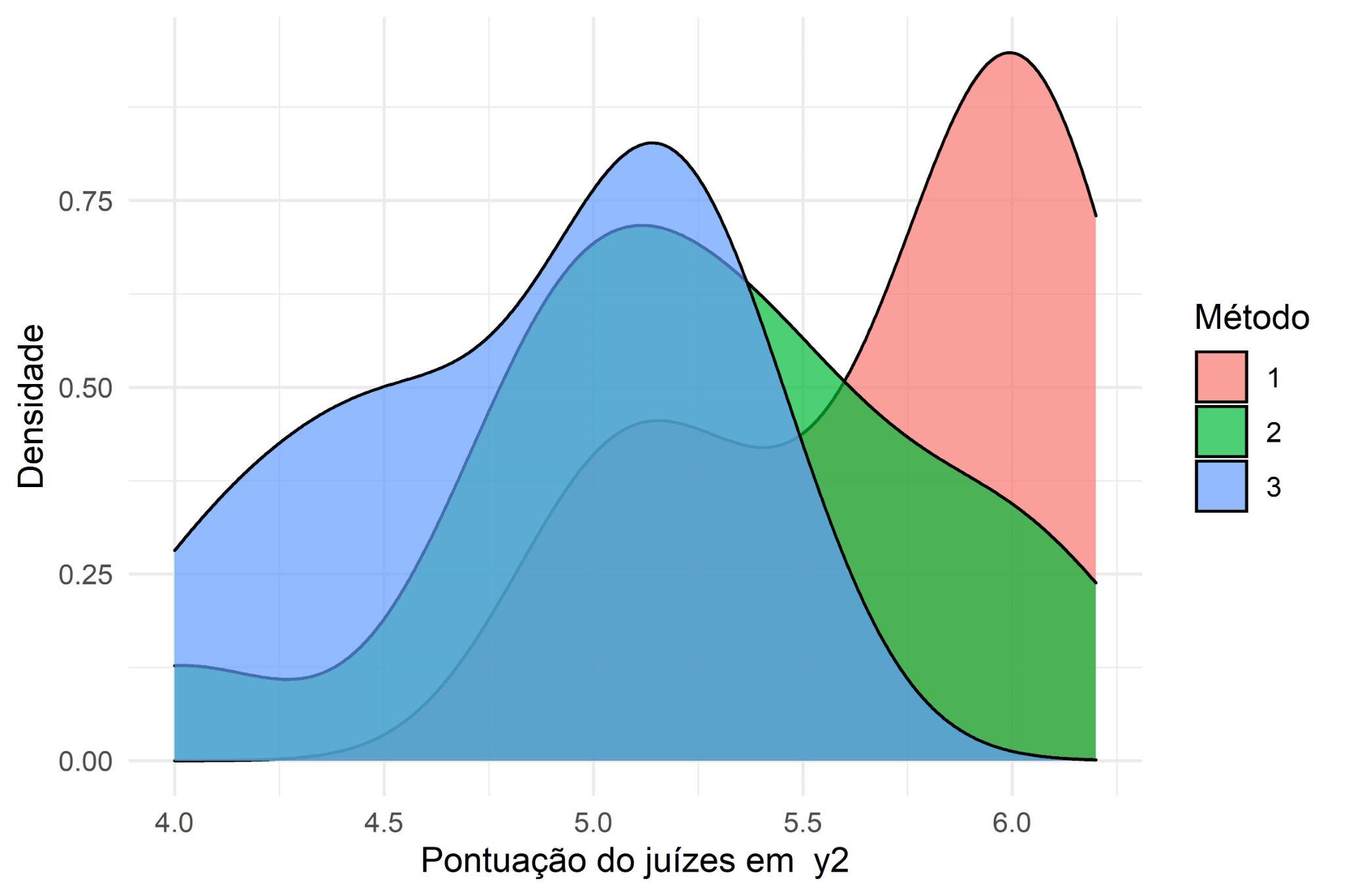
**1. b.3.5 Gráfico de Densidade**

**Figura 17**: Gráfico de densidade para a variável aroma (y1).



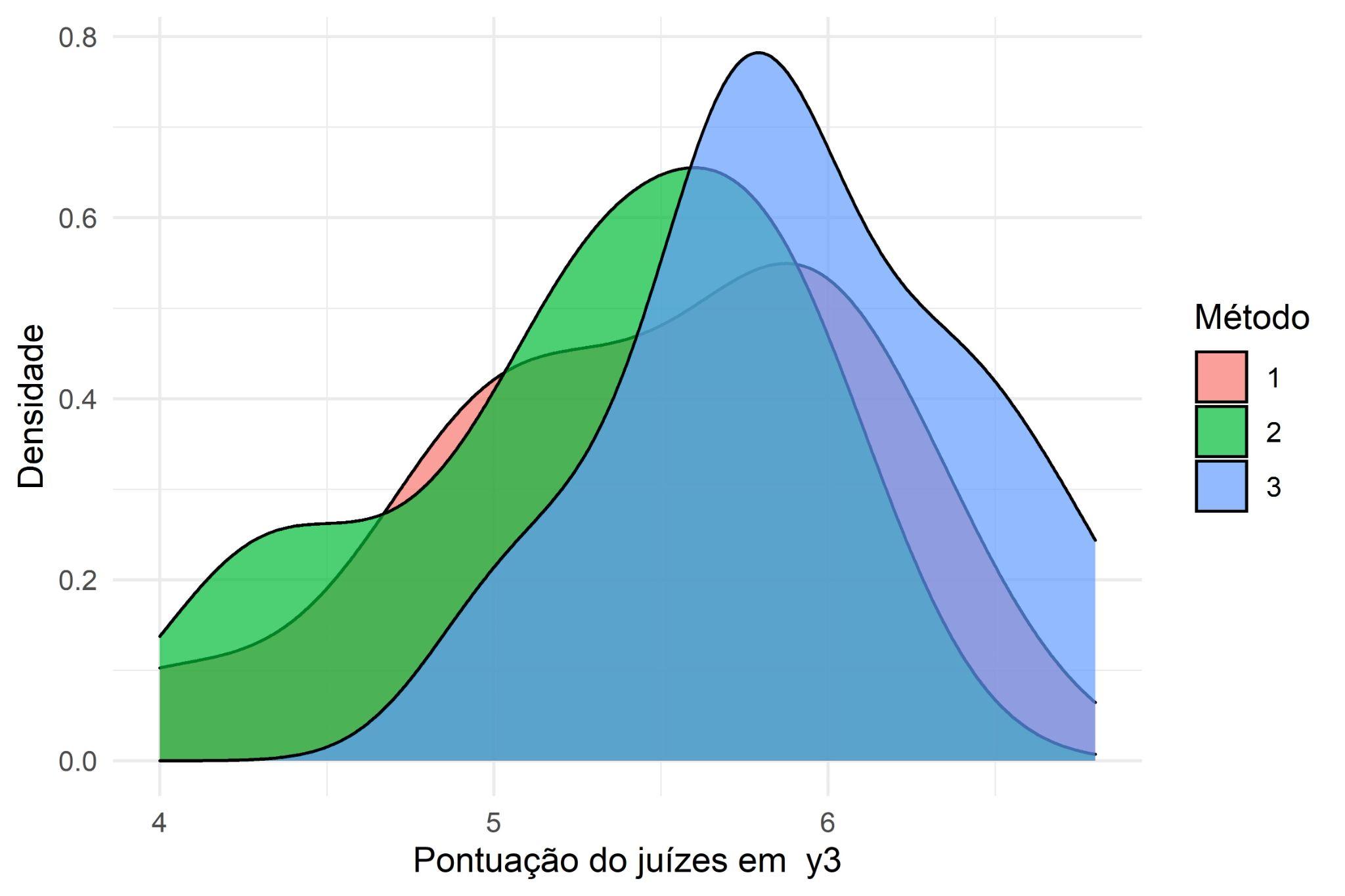
Fonte: Construído pelo autor, 2022.

**Figura 18**: Gráfico de densidade para a variável sabor (y2).



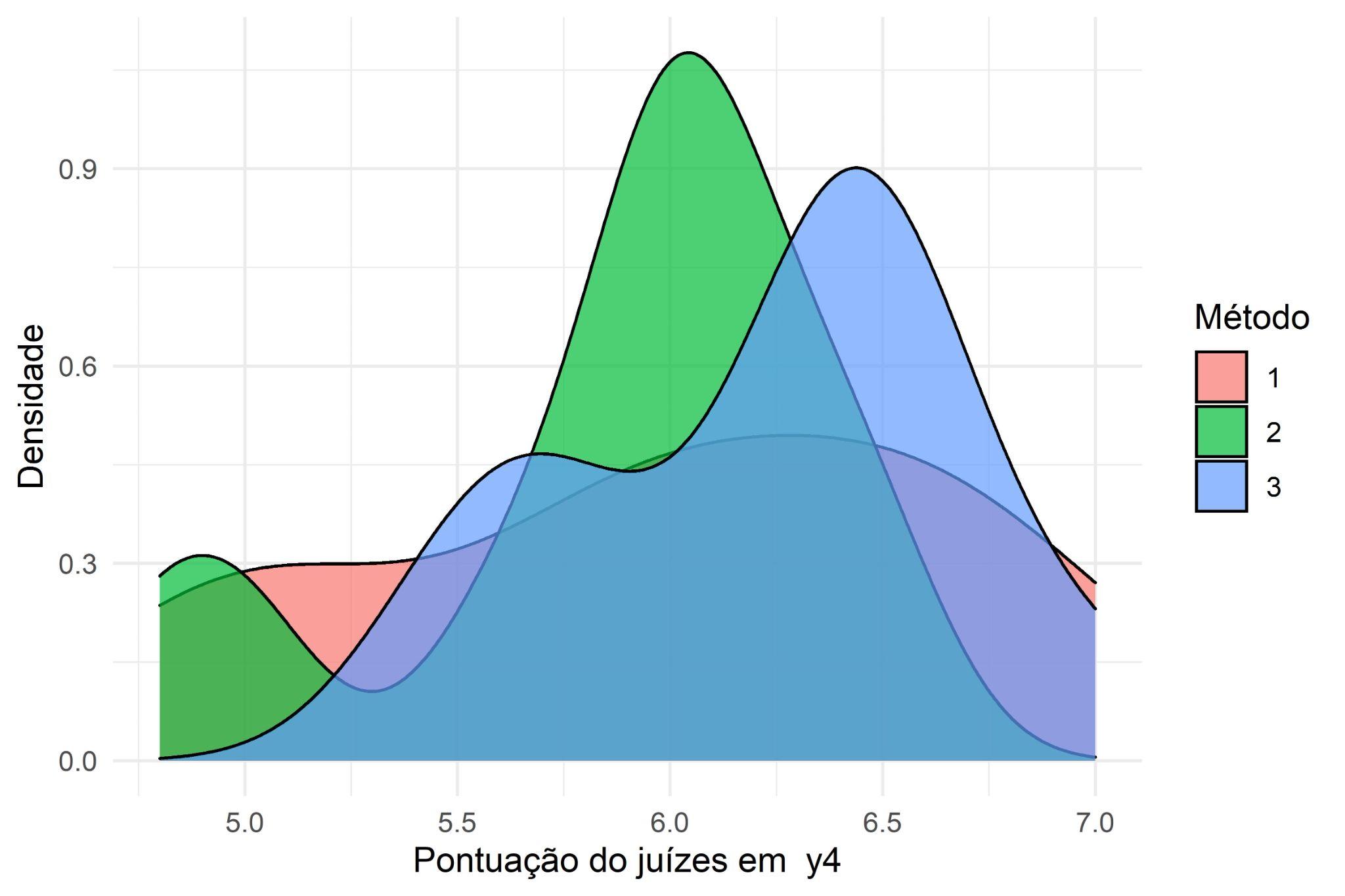
Fonte: Construído pelo autor, 2022.

**Figura 19**: Gráfico de densidade para a variável textura (y3).



Fonte: Construído pelo autor, 2022.

**Figura 20**: Gráfico de densidade para a variável umidade (y4).

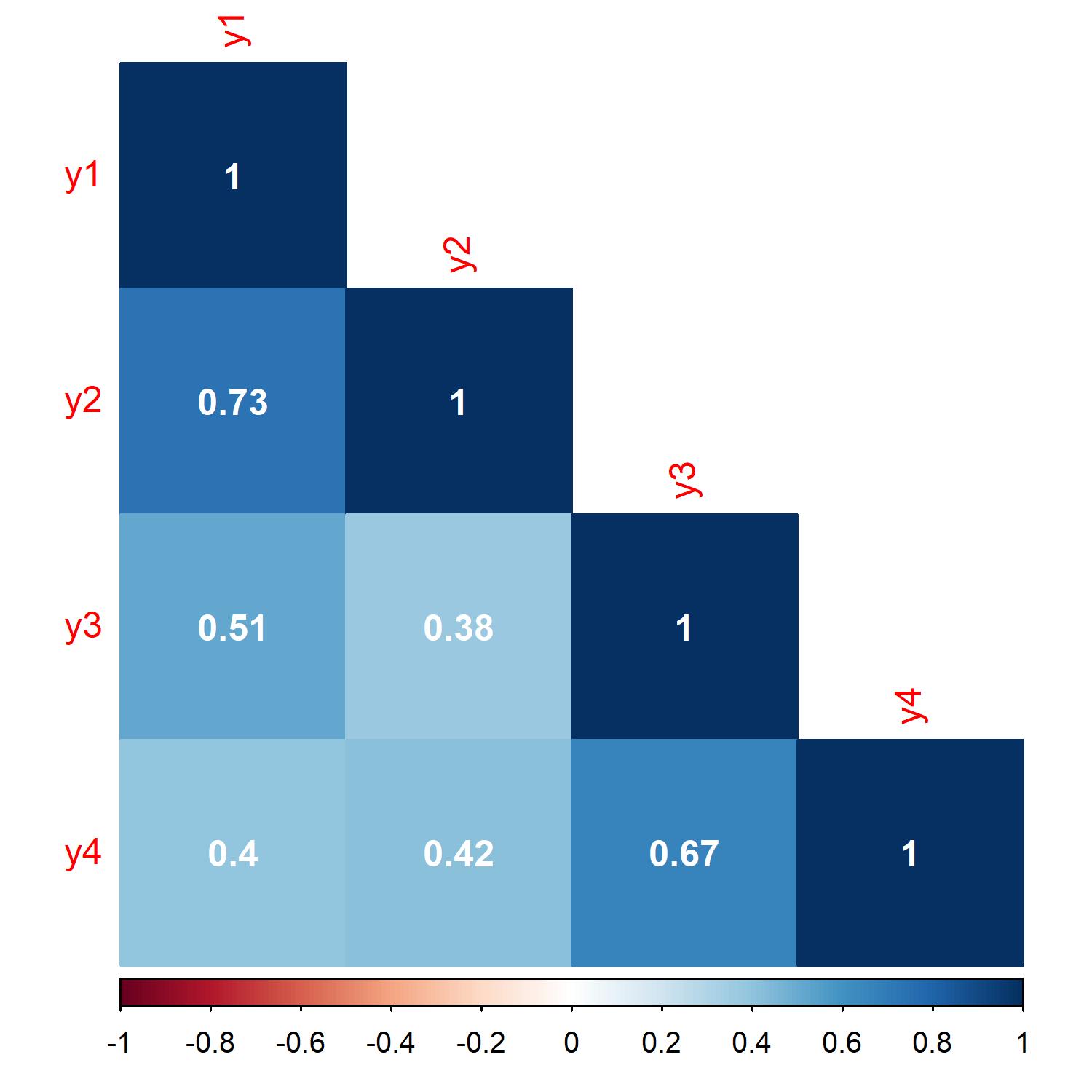


Fonte: Construído pelo autor, 2022.

**1. b.4 Gráficos Multivariados**

**1. b.4.1 Correlograma**

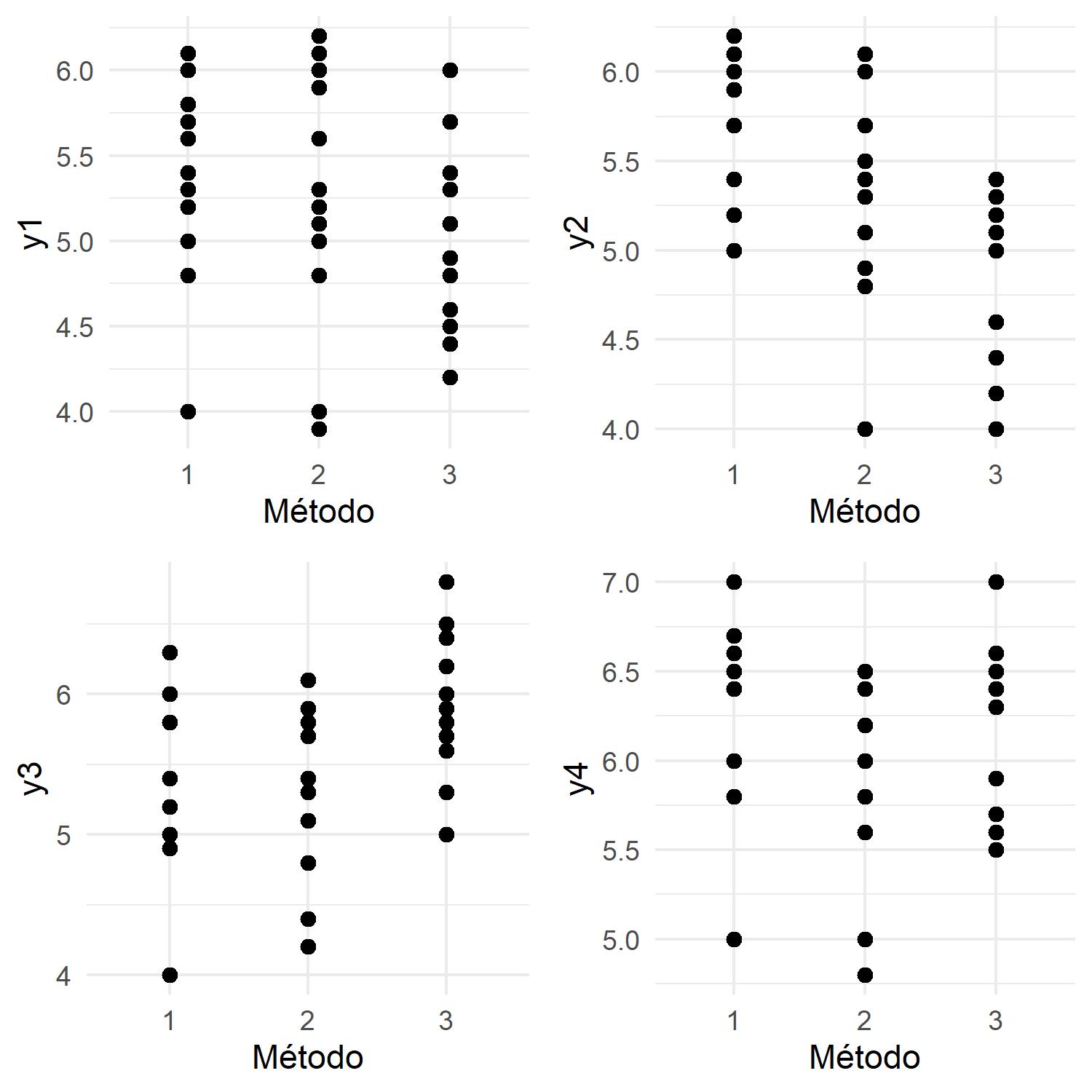
**Figura 21**: Correlograma entre as variáveis (y1, y2, y3 e y4) do estudo.



Fonte: Construído pelo autor, 2022.

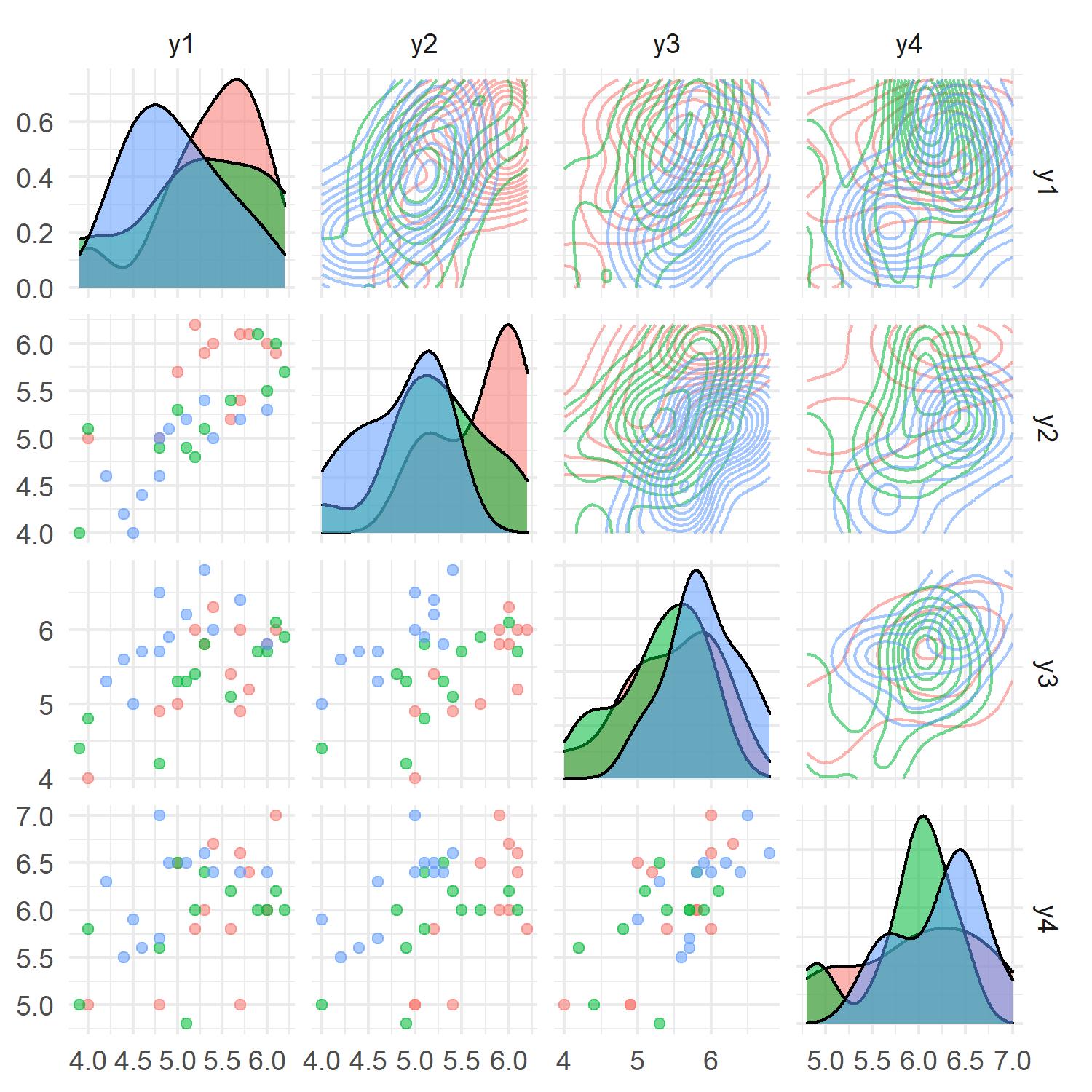
**1. b.4.2 Dispersão em relação aos métodos**

**Figura 22**: Dispersão individual de cada variável (y1, y2 e y3) em relação aos métodos (1, 2 e 3).



Fonte: Construído pelo autor, 2022.

**Figura 23**: Gráficos de dispersão, gráficos de densidade e curvas de nível entre os pares de variáveis do estudo, por método.



Fonte: Construído pelo autor, 2022.

**1. b.5 Análise de Variância Multivariada (MANOVA)**

**1. b.5.1 Verificação dos pressupostos**

**1. b.5.1.1 Normalidade multivariada dos dados**

Na fase inicial da análise de normalidade, realizou-se o teste de Shapiro-Wilk para verificar a normalidade univariada em cada variável quantitativa do banco de dados. Desse modo, a formalização do teste é dada por:

Hipóteses:

H0: A variável em questão segue uma distribuição normal.

H1: A variável em questão não segue uma distribuição normal.

De acordo com a **Tabela 7,** as variáveis y1, y2, y3 e y4, apresentaram um valor maior do que o nível de significância adotados, isto é, respectivamente, como 0,4374; 0,2340; 0,2747; 0,0416 são maiores do que 0,01, não rejeitamos H0 para o teste individual de cada variável. Logo, as variáveis y1, y2, y3 e y4 apresentam distintamente uma distribuição normal de probabilidade, com todas ao nível de 1% de significância.

**Tabela 7**: Resumo do teste de normalidade, contendo a estatística de teste, P-valor e resultado, por variável.

| Variável | Estatística de Teste | P-valor | Resultado |
| --- | --- | --- | --- |
| y1 | 0,3568 | 0,4374 | Sim |
| y2 | 0,4691 | 0,2340 | Sim |
| y3 | 0,4409 | 0,2747 | Sim |
| y4 | 0,7681 | 0,0416 | Não |

Fonte: Construído pelo autor, 2022.

Ademais, foi realizado o teste conjunto para verificar se poderia haver a existência de uma distribuição normal conjunta de quarta dimensão para os dados. Desse modo, o teste de Doornik-Hansen comprovou a normalidade multivariada dos dados, com uma estatística de teste de **3,9425** e um valor associado de **0,8623**. Portanto, ao nível de 1% de significância, podemos afirmar que os dados apresentam uma distribuição normal de quarta dimensão. Os resultados na linguagem R do teste de Henze-Zirkler estão disponíveis no **Anexo III** deste trabalho.

**1. b.5.1.2 Identificação outliers multivariados**

A Distância de Mahalanobis é uma técnica matemática que realiza uma normalização dos dados em relação à variância. Desse modo, esta métrica é de suma importância na Análise Multivariada para a detecção de valores discrepantes (outliers) para as dimensões maiores do que o espaço bidimensional. Logo, a **Tabela 8** resume os resultados do cálculo da distância de Mahalanobis feita para cada observação do banco de dados e, por isso, podemos inferir que não existem outliers multivariados para as variáveis conjuntamente.

**Tabela 8**: Resumo do cálculo da distância de Mahalanobis na detecção de outliers multivariados, por observação.

| Observação | Método | y1 | y2 | y3 | y4 | Distância | Resultado |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 5,40 | 6,00 | 6,30 | 6,70 | 3,4730 | FALSO |
| 2 | 1 | 5,20 | 6,20 | 6,00 | 5,80 | 5,3380 | FALSO |
| 3 | 1 | 6,10 | 5,90 | 6,00 | 7,00 | 3,7620 | FALSO |
| 4 | 1 | 4,80 | 5,00 | 4,90 | 5,00 | 3,4920 | FALSO |
| 5 | 1 | 5,00 | 5,70 | 5,00 | 6,50 | 3,7410 | FALSO |
| 6 | 1 | 5,70 | 6,10 | 6,00 | 6,60 | 0,9360 | FALSO |
| 7 | 1 | 6,00 | 6,00 | 5,80 | 6,00 | 1,9590 | FALSO |
| 8 | 1 | 4,00 | 5,00 | 4,00 | 5,00 | 6,4000 | FALSO |
| 9 | 1 | 5,70 | 5,40 | 4,90 | 5,00 | 5,0710 | FALSO |
| 10 | 1 | 5,60 | 5,20 | 5,40 | 5,80 | 3,9730 | FALSO |
| 11 | 1 | 5,80 | 6,10 | 5,20 | 6,40 | 4,8400 | FALSO |
| 12 | 1 | 5,30 | 5,90 | 5,80 | 6,00 | 1,0140 | FALSO |
| 13 | 2 | 5,00 | 5,30 | 5,30 | 6,50 | 2,5860 | FALSO |
| 14 | 2 | 4,80 | 4,90 | 4,20 | 5,60 | 6,2490 | FALSO |
| 15 | 2 | 3,90 | 4,00 | 4,40 | 5,00 | 4,9120 | FALSO |
| 16 | 2 | 4,00 | 5,10 | 4,80 | 5,80 | 6,7950 | FALSO |
| 17 | 2 | 5,60 | 5,40 | 5,10 | 6,20 | 2,4810 | FALSO |
| 18 | 2 | 6,00 | 5,50 | 5,70 | 6,00 | 1,2950 | FALSO |
| 19 | 2 | 5,20 | 4,80 | 5,40 | 6,00 | 2,4300 | FALSO |
| 20 | 2 | 5,30 | 5,10 | 5,80 | 6,40 | 3,8150 | FALSO |
| 21 | 2 | 5,90 | 6,10 | 5,70 | 6,00 | 3,4790 | FALSO |
| 22 | 2 | 6,10 | 6,00 | 6,10 | 6,20 | 2,2860 | FALSO |
| 23 | 2 | 6,20 | 5,70 | 5,90 | 6,00 | 1,7360 | FALSO |
| 24 | 2 | 5,10 | 4,90 | 5,30 | 4,80 | 5,9360 | FALSO |
| 25 | 3 | 4,80 | 5,00 | 6,50 | 7,00 | 6,9390 | FALSO |
| 26 | 3 | 5,40 | 5,00 | 6,00 | 6,40 | 1,0170 | FALSO |
| 27 | 3 | 4,90 | 5,10 | 5,90 | 6,50 | 2,3080 | FALSO |
| 28 | 3 | 5,70 | 5,20 | 6,40 | 6,40 | 2,5380 | FALSO |
| 29 | 3 | 4,20 | 4,60 | 5,30 | 6,30 | 5,7990 | FALSO |
| 30 | 3 | 6,00 | 5,30 | 5,80 | 6,40 | 6,1670 | FALSO |
| 31 | 3 | 5,10 | 5,20 | 6,20 | 6,50 | 1,2040 | FALSO |
| 32 | 3 | 4,80 | 4,60 | 5,70 | 5,70 | 1,9580 | FALSO |
| 33 | 3 | 5,30 | 5,40 | 6,80 | 6,60 | 3,5040 | FALSO |
| 34 | 3 | 4,60 | 4,40 | 5,70 | 5,60 | 2,4880 | FALSO |
| 35 | 3 | 4,50 | 4,00 | 5,00 | 5,90 | 6,5810 | FALSO |
| 36 | 3 | 4,40 | 4,20 | 5,60 | 5,50 | 3,4960 | FALSO |

Fonte: Construído pelo autor, 2022.

Para a análise de outliers univariados, os boxplots desenvolvidos nas **Figuras 5, 6, 7** e **8**, que se referem às variáveis y1, y2, y3 e y4, nesta respectiva ordem, evidenciam que existe um valor aberrante para o método 1, na variável y1, tal como, dois outliers no método 2, para a variável y4. De um modo geral, não foram encontrados um quantitativo relevante que possa interferir na análise, por conta disto, é ainda seguro prosseguir com a MANOVA.

**1. b.5.1.3 Teste de homogeneidade da matriz de variâncias e covariâncias**

Uma parte importante na ANOVA é a realização do teste de homogeneidade de variâncias para as categorias da variável dependente. Porém, para a MANOVA, é preciso realizar o teste de homogeneidade na matriz de variâncias e covariâncias entre os grupos da variável dependente. Vale ressaltar que o teste Box, o qual é utilizado para verificar esta suposição, é bem sensível e, por este motivo, deve-se utilizar sempre o nível de significância a 1%. Segue abaixo a formalização do teste de Box:

Teste de Box para a homogeneidade da matriz de variâncias e covariâncias

Hipóteses:

H0: Existe homogeneidade na matriz de variâncias e covariâncias.

H1: Não existe homogeneidade na matriz de variâncias e covariâncias.

Estatística de teste = 13,50

P-valor = 0,8570

Conclusão:

Como o P-valor é maior que o nível de significância, isto é, 0,8570 > 0,01, não rejeitamos H0. Logo, existe homogeneidade na matriz de variâncias e covariâncias, ao nível de 1% de significância.

Como este pressuposto não foi rompido, espera-se que os testes de Hotelling e Pillai forneçam resultados confiáveis e, por isso, não existe ainda a necessidade de incrementar a robustez da MANOVA.

**1.b.5.1.4 Homogeneidade de Variâncias**

Para a análise posterior à MANOVA, geralmente é realizada a Análise de Variância (ANOVA) individual de cada variável dependente em relação às categorias da variável dependente do método *one way*. Dessa maneira, segue abaixo a formalização do Teste de Levene para investigar a homogeneidade de variâncias.

Teste de Levene para a homogeneidade de variâncias

Hipóteses:

H0: Existe homogeneidade de variâncias entre os métodos.

H1: Não existe homogeneidade de variâncias entre os métodos.

A **Tabela 9** mostra o resumo do Teste de Levene aplicado em cada variável do banco de dados, justificando a estatística F e o p-valor associado.

**Tabela 9**: Estatística de Teste e P-valor para o teste de Levene, por variável.

| Variável | Estatística F | P-valor |
| --- | --- | --- |
| y1 | 0,6840 | 0,5116 |
| y2 | 0,1577 | 0,8547 |
| y3 | 0,9190 | 0,5590 |
| y4 | 1,1674 | 0,3237 |

Fonte: Construído pelo autor, 2022.

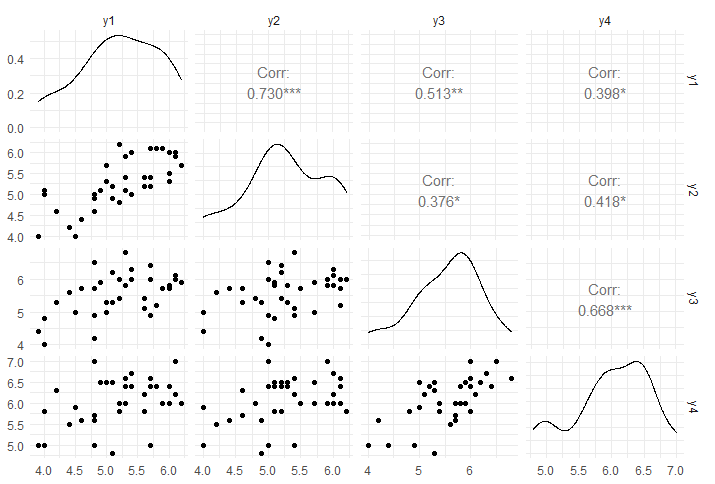
Conclusão:

Como em y1, y2, y3 e y4, o p-valor resultante é maior do que o nível de significância adotado, isto é, respectivamente, 0,5116; 0,8547; 0,5590; 0,3237 > 0,01, não se rejeita a hipótese nula para qualquer variável do estudo. Em suma, para y1, y2, y3 e y4, existe a homogeneidade de variâncias em relação aos métodos, ao nível de 1% de significância.

**1.b.5.1.5 Verificação da ausência de multicolinearidade**

A característica de multicolinearidade mostra que, duas ou mais variáveis quantitativas, acabam explicando a mesma informação sobre o fenômeno em estudo quando estamos interessados em inferir através de modelos. Nesse viés, é sempre um bom costume analisar em pares todas as variáveis que estamos trabalhando. Na **Figura 24**, é possível observar a dispersão e correlação linear em pares. Logo, podemos perceber que a única correlação forte está presente entre as variáveis y1 e y2, expressa pelo valor 0,7300, e este fator pode afetar a realização da MANOVA. Para o prosseguimento do projeto, visto que os demais pares não apresentaram valores maiores que 0,90, o que caracteriza uma correlação muito forte, iremos concluir que não existe multicolinearidade.

**Figura 24**: Gráficos de dispersão, gráficos de densidade e correlações entre os pares de variáveis do estudo.

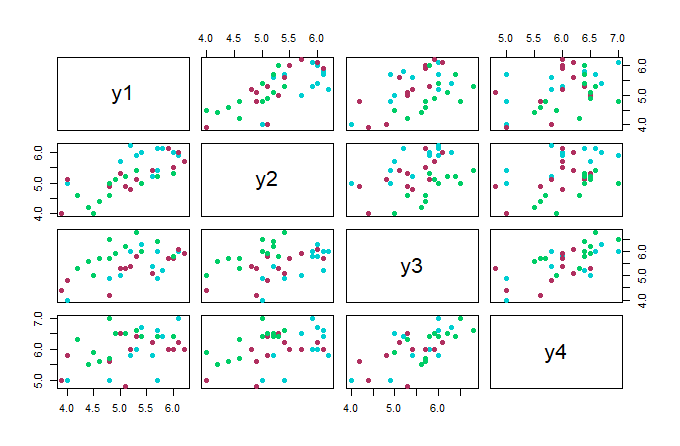


Fonte: Construído pelo autor, 2022.

**1.b.5.1.6 Verificação da relação linear das variáveis dependentes para cada grupo.**

Um outro aspecto importante para a realização da MANOVA, é a ausência da relação linear entre as variáveis dependentes para cada categoria da variável independente. Dessa forma, a **Figura 25** aborda a dispersão em pares entre as variáveis das notas dos juízes para a qualidade dos peixes, estratificadas pelos método 1, 2 e 3, destacados pelas cores azul, roxo e verde, respectivamente. A partir do gráfico, percebe-se a inconsistência em ajustar um modelo linear em qualquer par de variáveis.

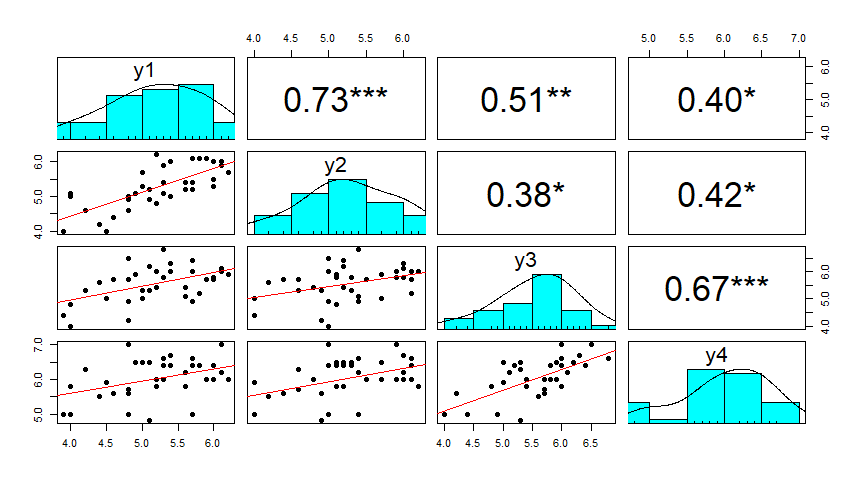
**Figura 25**: Gráficos de dispersão, gráficos de densidade e correlações entre os pares de variáveis do estudo.



Fonte: Construído pelo autor, 2022.

De forma complementar, a **Figura 26** aborda o ajuste de um modelo linear entre os pares de variáveis. Logo, reitera-se a inexistência de uma relação linear entre qualquer par de variáveis no estudo.

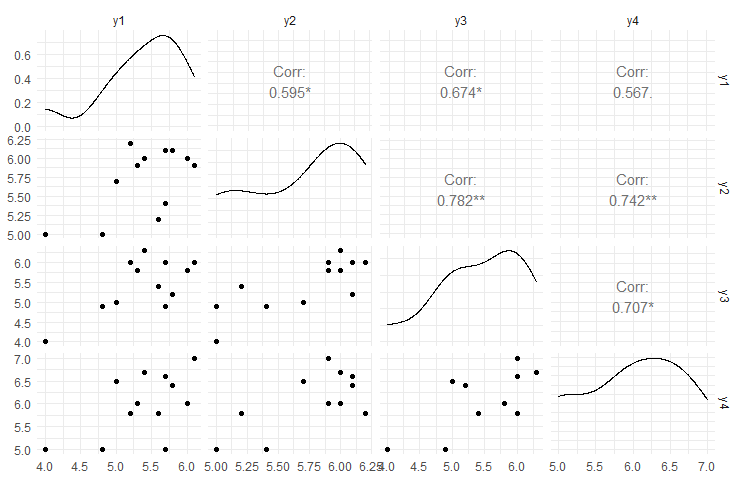
**Figura 26**: Histograma, ajustes lineares e correlações entre os pares de variáveis do estudo.



Fonte: Construído pelo autor, 2022.

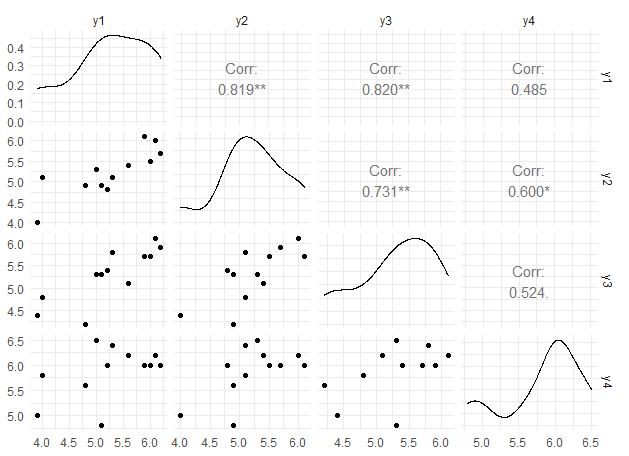
No estudo individual para cada método de preparo dos peixes (1, 2 e 3), os gráficos de densidade, dispersão e correlação linear, os quais estão destacados nas **Figuras 27, 28** e **29** para os métodos 1, 2 e 3, respectivamente, ratificam a existência de uma interação linear entre os pares de variáveis, mesmo existindo algumas correlações consideradas como forte.

**Figura 27**: Gráfico de densidade, gráfico de dispersão e correlações entre os pares de variáveis do estudo para o método 1.



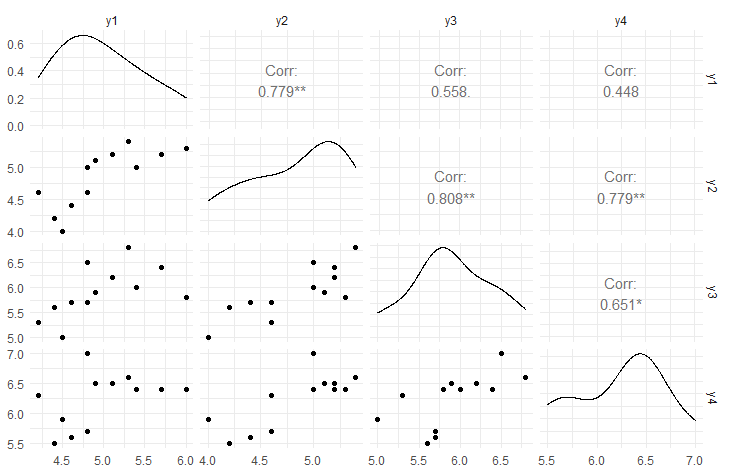
Fonte: Construído pelo autor, 2022.

**Figura 28**: Gráfico de densidade, gráfico de dispersão e correlações entre os pares de variáveis do estudo para o método 2.



Fonte: Construído pelo autor, 2022.

**Figura 29**: Gráfico de densidade, gráfico de dispersão e correlações entre os pares de variáveis do estudo para o método 3.



Fonte: Construído pelo autor, 2022.

**1.b.5.2** **Construção do modelo MANOVA**

Com apenas dois adendos na análise dos pressupostos, os quais foram: a presença de alguns outliers univariados entre os métodos e a existência de correlações fortes quando estratifica-se o banco de dados pelo métodos, é ainda seguro dizer que os resultados gerados pelo cálculo da MANOVA nos dados de avaliação dos juízes estão sujeitos a gerar resultados satisfatórios.

Na **Tabela 10**, são resumidos os resultados da MANOVA pelas estatísticas de teste: Lambda de Wilks, Traço de Pillai, Traço de Lawley-Hotelling e Raiz Máxima de Roy, estas quatro formas são propostas para o cálculo e interpretação da Análise de Variância Múltipla e, todas apresentam aproximações para distribuição F-Snedecor de probabilidade.

Para a MANOVA aplicado no problema de classificação dos juízes, estamos interessados em avaliar as seguintes hipóteses:

H0: O método de preparo dos peixes não afeta as variáveis aroma (y1), sabor (y2), textura (y3) e umidade (y4) conjuntamente.

H1: O método de preparo dos peixes afeta as variáveis aroma (y1), sabor (y2), textura (y3) e umidade (y4) conjuntamente.

Em conclusão a partir os valores para o p-valor na **Tabela 10** e, adotando 0,01 como nível de significância, podemos afirmar que, de acordo com os testes multivariados de Wilks, Pillai, Lawley-Hotelling e Roy, o método de preparo dos peixes afeta as variáveis aroma (y1), sabor (y2), textura (y3) e umidade (y4) conjuntamente, ou seja, rejeitamos H0 ao nível de 1% de significância nos 4 testes.

**Tabela 10**: Estatística de Teste, Valor aproximado a estatística F, Graus de liberdade e P-valor para a Análise de Variância Multivariada, por teste.

| Teste | Estatística de Teste | Valor F | GL1 | GL2 | P-valor |
| --- | --- | --- | --- | --- | --- |
| Lambda de Wilks | 0,2201 | 8,4883 | 8 | 60 |  |
| Traço de Pillai | 0,8642 | 5,8966 | 8 | 62 |  |
| Traço de Lawley-Hotelling | 3,1617 | 11,641 | 8 | 58 |  |
| Raiz máxima de Roy | 3,0356 | 23,526 | 4 | 31 |  |

Fonte: Construído pelo autor, 2022.

**1.b.5.3 ANOVA para cada variável**

Após a realização da Análise de Variância Múltipla, é considerado como uma boa prática a realização da Análise de Variância para cada variável utilizada, com o intuito de verificar se as categorias da variável dependente interferem no caso univariado, ou apenas conjuntamente que podemos perceber tal efeito.

Na **Tabela 11**, é exemplificado o resultado da ANOVA para a variável aroma (y1) levando em consideração os métodos de preparo 1, 2 e 3.

H0: Em média, a nota atribuída para o aroma para os três métodos de preparo (1, 2 e 3) é igual.

H1: Pelo menos um dos três métodos de preparo (1, 2 e 3) tem nota média para o aroma que se difere entre os demais.

Como o p-valor observado é maior do que o nível de significância adotado, isto é, 0,2880 > 0,01, não se rejeita H0. Logo, em média, a nota atribuída para o aroma para os três métodos de preparo (1, 2 e 3) é igual, ao nível de 1% de significância.

**Tabela 11**: Graus de Liberdade (GL), Soma de Quadrados (SQ), Quadrados Médios (QM), valor de F calculado e o P-valor de F, obtidos na ANOVA da variável aroma (y1).

| Causa de Variação | GL | SQ | QM | F calculado | P(> F calculado) |
| --- | --- | --- | --- | --- | --- |
| Método | 2 | 1,0506 | 0,52528 | 1,2928 | 0,2880 |
| Resíduos | 33 | 13,4083 | 0,40631 | - | - |

Fonte: Construído pelo autor, 2022.

Na **Tabela 12**, é possível visualizar os resultados obtidos para a ANOVA feita para a variável sabor (y2) em relação ao métodos de preparo dos peixes (1, 2 e 3).

H0: Em média, a nota atribuída para o sabor para os três métodos de preparo (1, 2 e 3) é igual.

H1: Pelo menos um dos três métodos de preparo (1, 2 e 3) tem nota média para o sabor que se difere entre os demais.

Como o p-valor observado é menor do que o nível de significância adotado, isto é, 0,0006 > 0,01, rejeita-se H0. Logo, em média, pelo menos um dos três métodos de preparo (1, 2 e 3) tem nota média para o sabor que se difere entre os demais, ao nível de 1% de significância.

**Tabela 12**: Graus de Liberdade (GL), Soma de Quadrados (SQ), Quadrados Médios (QM), valor de F calculado e o P-valor de F, obtidos na ANOVA da variável sabor (y2).

| Causa de Variação | GL | SQ | QM | F calculado | P(> F calculado) |
| --- | --- | --- | --- | --- | --- |
| Método | 2 | 4,6050 | 2,3025 | 9,3778 | 0,0006 |
| Resíduos | 33 | 8,1025 | 0,2455 | - | - |

Fonte: Construído pelo autor, 2022.

A **Tabela 13** tem por objetivo exibir os resultados da Análise de Variância feita para a variável textura (y3) em relação aos três métodos de preparo (1, 2 e 3).

H0: Em média, a nota atribuída para a textura para os três métodos de preparo (1, 2 e 3) é igual.

H1: Pelo menos um dos três métodos de preparo (1, 2 e 3) tem nota média para a textura que se difere entre os demais.

Como o p-valor observado é maior do que o nível de significância adotado, isto é, 0,0460 > 0,01, não se rejeita H0. Logo, em média, a nota atribuída para a textura para os três métodos de preparo (1, 2 e 3) é igual, ao nível de 1% de significância.

**Tabela 13**: Graus de Liberdade (GL), Soma de Quadrados (SQ), Quadrados Médios (QM), valor de F calculado e o P-valor de F, obtidos na ANOVA da variável textura (y3).

| Causa de Variação | GL | SQ | QM | F calculado | P(> F calculado) |
| --- | --- | --- | --- | --- | --- |
| Método | 2 | 2,3822 | 1,1911 | 3,3863 | 0,0460 |
| Resíduos | 33 | 11,6075 | 0,3517 | - | - |

Fonte: Construído pelo autor, 2022.

Por último, a **Tabela 14** mostra os resultados da ANOVA para a variável umidade (y4) em relação aos três métodos de preparo (1, 2 e 3).

H0: Em média, a nota atribuída para a umidade para os três métodos de preparo (1, 2 e 3) é igual.

H1: Pelo menos um dos três métodos de preparo (1, 2 e 3) tem nota média para a umidade que se difere entre os demais.

Como o p-valor observado é maior do que o nível de significância adotado, isto é, 0,2954 > 0,01, não se rejeita H0. Logo, em média, a nota atribuída para a umidade para os três métodos de preparo (1, 2 e 3) é igual, ao nível de 1% de significância.

**Tabela 14**: Graus de Liberdade (GL), Soma de Quadrados (SQ), Quadrados Médios (QM), valor de F calculado e o P-valor de F, obtidos na ANOVA da variável umidade (y4).

| Causa de Variação | GL | SQ | QM | F calculado | P(> F calculado) |
| --- | --- | --- | --- | --- | --- |
| Método | 2 | 0,8106 | 0,4053 | 1,2658 | 0,2954 |
| Resíduos | 33 | 10,5658 | 0,3202 | - | - |

Fonte: Construído pelo autor, 2022.

Dentre as variáveis utilizadas dentro do banco, apenas o sabor consegue mostrar com significância que existe diferença entre os métodos de preparo do peixe para os juízes, por outro lado, quando se olha conjuntamente para as 4 variáveis juntas, consegue-se captar o efeito de cada para também evidenciar com significância a diferença entre os modos de preparo do peixe.

**1.b.6** **Análise de da diferença entre grupos (Post-hoc)**

A Análise *Post-Hoc* é sempre posterior à realização da ANOVA, ela tem por objetivo identificar quais grupos diferem entre após verificar que pelo menos um deles se difere entre os demais. Para o presente trabalho, foram aplicados o teste de comparação de médias com a correção de Bonferroni e o teste de diferença entre médias de Tukey.

**1.b.6.1 Teste de Bonferroni para comparações**

Na **Tabela 15**, é possível verificar a estatística de teste, o p-valor e o p-valor ajustado de Bonferroni para o teste de comparação entre médias na variável aroma (y1). Desse modo, percebe-se que, ao nível de 1% de significância, não foram encontradas diferenças entre os grupos para a variável aroma (y1).

**Tabela 15**: Resultados do teste de Bonferroni de comparações na variável aroma (y1).

| Grupo 1 | Grupo 2 | Estatística de teste | P-valor | P-valor ajustado |
| --- | --- | --- | --- | --- |
| 1 | 2 | 0,4800 | 0,6340 | 1,000 |
| 1 | 3 | 1,5700 | 0,1260 | 0,3780 |
| 2 | 3 | 1,0900 | 0,2840 | 0,8520 |

Fonte: Construído pelo autor, 2022.

A **Tabela 16** mostra os resultados do teste de comparação de médias com a correção de Bonferroni para a variável sabor (y2). Nos três pares de variáveis do experimento, nota-se a diferença entre grupos. De fato, ao nível de 1%, a variável sabor (y2) mostra que o método que se difere do método 2, o método 1 se difere do método 3 e, por fim, o método 2 se difere do método 3, colaborando para a formação de três grupos distintos.

**Tabela 16**: Resultados do teste de Bonferroni de comparações na variável sabor (y2).

| Grupo 1 | Grupo 2 | Estatística de teste | P-valor | P-valor ajustado |
| --- | --- | --- | --- | --- |
| 1 | 2 | 2,3500 |  |  |
| 1 | 3 | 4,3300 |  |  |
| 2 | 3 | 1,9800 |  |  |

Fonte: Construído pelo autor, 2022.

Na **Tabela 17**, é possível verificar a estatística de teste, o p-valor e o p-valor ajustado de Bonferroni para o teste de comparação entre médias na variável textura (y3). Desse modo, percebe-se que, ao nível de 1% de significância, não foram encontradas diferenças entre os grupos para a variável textura (y3).

**Tabela 17**: Resultados do teste de Bonferroni de comparações na variável textura (y3).

| Grupo 1 | Grupo 2 | Estatística de teste | P-valor | P-valor ajustado |
| --- | --- | --- | --- | --- |
| 1 | 2 | 0,5510 | 0,5860 | 1,0000 |
| 1 | 3 | -1,9300 | 0,0626 | 0,1880 |
| 2 | 3 | -2,4800 | 0,0185 | 0,0555 |

Fonte: Construído pelo autor, 2022.

Na **Tabela 18**, é possível verificar a estatística de teste, o p-valor e o p-valor ajustado de Bonferroni para o teste de comparação entre médias na variável umidade (y4). Desse modo, percebe-se que, ao nível de 1% de significância, não foram encontradas diferenças entre os grupos para a variável umidade (y4).

**Tabela 18**: Resultados do teste de Bonferroni de comparações na variável umidade (y4).

| Grupo 1 | Grupo 2 | Estatística de teste | P-valor | P-valor ajustado |
| --- | --- | --- | --- | --- |
| 1 | 2 | 0,4690 | 0,6420 | 1,0000 |
| 1 | 3 | -1,0800 | 0,2870 | 0,8610 |
| 2 | 3 | -1,5500 | 0,1300 | 0,3910 |

Fonte: Construído pelo autor, 2022.

**1.b.6.2 Teste de Tukey para comparações**

A **Tabela 19** evidencia os resultados do teste de comparação de médias de Tukey para a formação de grupos na variável aroma (y1). Desse modo, como todos os pares formados entre os 3 métodos de preparo resultaram em p-valores ajustados maiores que 1%, isto é maior do que o nível de significância adotado (0,01), tem-se que todos os métodos apresentam notas iguais dos juízes.

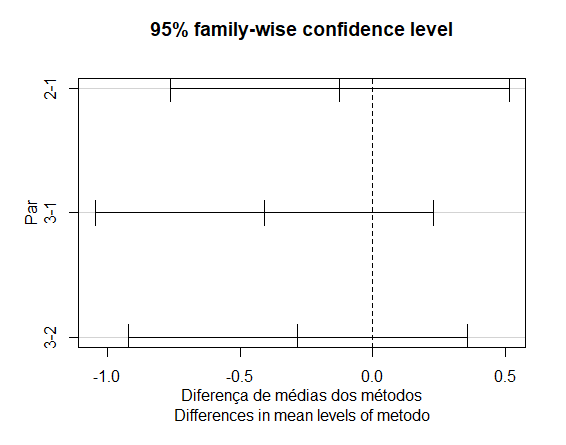
**Tabela 19**: Resultados do teste de Tukey para comparações na variável aroma (y1).

| Grupo 1 | Grupo 2 | Diferença | Limite inferior | Limite superior | P-valor ajustado |
| --- | --- | --- | --- | --- | --- |
| 2 | 1 | -0,1250 | -0,7635 | 0,5135 | 0,8810 |
| 3 | 1 | -0,4083 | -1,0469 | 0,5135 | 0,2730 |
| 3 | 2 | -0,2833 | -0,9219 | 0,3552 | 0,5276 |

Fonte: Construído pelo autor, 2022.

A **Figura 30** aborda o intervalo de confiança construído para cada par dos métodos de preparo formados a partir da variável aroma (y1). Diante disso, dado que todos os intervalos abrangem o valor de zero, temos que existem evidências estatísticas de que os métodos são iguais, ao nível de 1%.

**Figura 30**: Gráfico dos valores para o intervalo do teste de Tukey para a variável aroma (y1).



Fonte: Construído pelo autor, 2022.

A **Tabela 20** tem por objetivo elucidar os resultados do teste de Tukey para a diferença de médias entre grupos. Diferente do que acontece com os resultados do teste de Bonferroni, na **Tabela 16**, temos a formação de apenas 2 grupos ao invés de 3 para a variável sabor. Portanto, ao nível de 1% de significância, temos que os métodos 1 e 2 são iguais, os métodos 2 e 3 são iguais e os métodos 1 e 3 são diferentes.

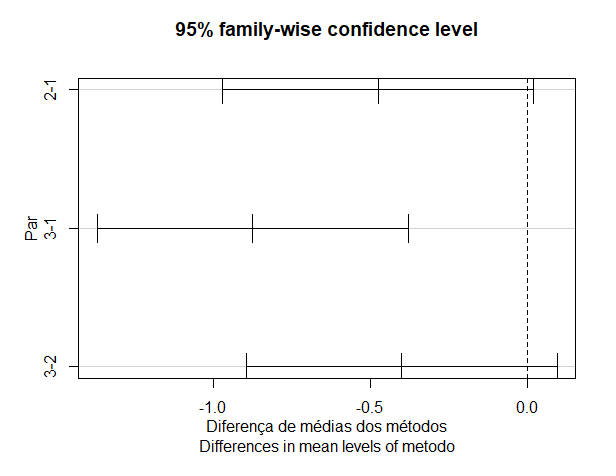
**Tabela 20**: Resultados do teste de Tukey para comparações na variável sabor (y2).

| Grupo 1 | Grupo 2 | Diferença | Limite inferior | Limite superior | P-valor ajustado |
| --- | --- | --- | --- | --- | --- |
| 2 | 1 | -0,4750 | -0,9714 | 0,0214 | 0,0630 |
| 3 | 1 | -0,8750 | -1,3714 | -0,3786 | 0,0004 |
| 3 | 2 | -0,4000 | -0,8964 | 0,0964 | 0,1337 |

Fonte: Construído pelo autor, 2022.

No gráfico do intervalo de confiança de 1% de significância da variável sabor, exemplificado na **Figura 31**, percebe-se queapenas o par entre os métodos 1 e 3 é o que não contém o valor zero, valor o qual mostra indícios de igualdade entre os grupos.

**Figura 31**: Gráfico dos valores para o intervalo do teste de Tukey para a variável sabor (y2).



A **Tabela 21** evidencia os resultados do teste de comparação de médias de Tukey para a formação de grupos na variável textura (y3). Desse modo, como todos os pares formados entre os 3 métodos de preparo resultaram em p-valores ajustados maiores que 1%, isto é maior do que o nível de significância adotado (0,01), tem-se que todos os métodos apresentam notas iguais dos juízes.

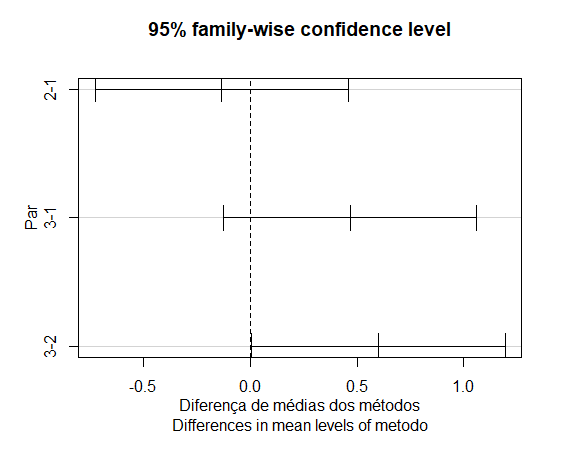
**Tabela 21**: Resultados do teste de Tukey para comparações na variável textura (y3).

| Grupo 1 | Grupo 2 | Diferença | Limite inferior | Limite superior | P-valor ajustado |
| --- | --- | --- | --- | --- | --- |
| 2 | 1 | -0,1333 | -0,7275 | 0,4608 | 0,8468 |
| 3 | 1 | 0,4667 | -0,1275 | 1,0608 | 0,1469 |
| 3 | 2 | 0,6000 | 0,0059 | 1,1941 | 0,0474 |

Fonte: Construído pelo autor, 2022.

A **Figura 32** aborda o intervalo de confiança construído para cada par dos métodos de preparo formados a partir da variável textura (y3). Diante disso, dado que todos os intervalos abrangem o valor de zero, temos que existem evidências estatísticas de que os métodos são iguais, ao nível de 1%.

**Figura 32**: Gráfico dos valores para o intervalo do teste de Tukey para a variável textura (y3).



Fonte: Construído pelo autor, 2022.

A **Tabela 22** evidencia os resultados do teste de comparação de médias de Tukey para a formação de grupos na variável umidade (y4). Desse modo, como todos os pares formados entre os 3 métodos de preparo resultaram em p-valores ajustados maiores que 1%, isto é maior do que o nível de significância adotado (0,01), tem-se que todos os métodos apresentam notas iguais dos juízes.

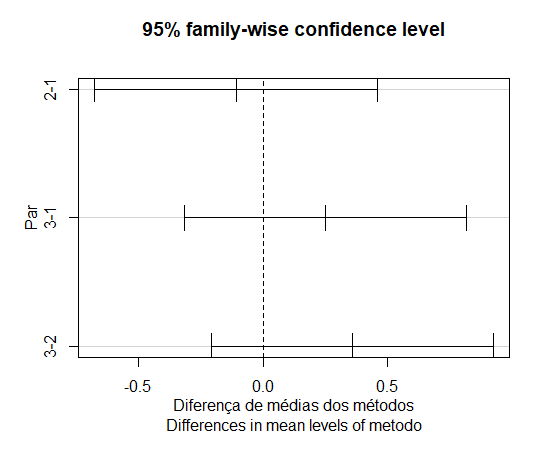
**Tabela 22**: Resultados do teste de Tukey para comparações na variável umidade (y4).

| Grupo 1 | Grupo 2 | Diferença | Limite inferior | Limite superior | P-valor ajustado |
| --- | --- | --- | --- | --- | --- |
| 2 | 1 | -0,1083 | -0,6752 | 0,4585 | 0,8863 |
| 3 | 1 | 0,2500 | -0,3168 | 0,8168 | 0,5316 |
| 3 | 2 | 0,3583 | -0,2085 | 0,9252 | 0,2808 |

Fonte: Construído pelo autor, 2022.

A **Figura 33** aborda o intervalo de confiança construído para cada par dos métodos de preparo formados a partir da variável umidade (y4). Diante disso, dado que todos os intervalos abrangem o valor de zero, temos que existem evidências estatísticas de que os métodos são iguais, ao nível de 1%.

**Figura 33**: Gráfico dos valores para o intervalo do teste de Tukey para a variável umidade (y4).



**Questão 2:** Dados sobre Feijão e Vagem**.**

**Item a):** Análise Exploratória de Dados (AED).

**2. a.1.1 Medidas descritivas do banco em geral**

**Tabela 23**: Valores descritivos de cada variável do banco de dados, por estatística.

| Estatística | y1 | y2 | y3 | y4 |
| --- | --- | --- | --- | --- |
| Mínimo | 58,90 | 1,30 | 36,60 | 233,00 |
| Q1 | 60,23 | 3,50 | 37,95 | 272,00 |
| Mediana | 63,45 | 4,30 | 40,10 | 285,50 |
| Média | 63,78 | 4,45 | 40,18 | 286,50 |
| Q3 | 68,33 | 5,38 | 41,83 | 301,20 |
| Máximo | 69,80 | 1,80 | 48,40 | 345,00 |

Fonte: Construído pelo autor, 2022.

**2. a.1.2 Vetor de médias geral**

|  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- |
|  | 63,78 | 4,45 | 40,18 | 286,50 |  |

**2. a.1.2 Matriz de variâncias e covariâncias geral**

|  |  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- | --- |
| y1 |  | 15,4257 | -4,6069 | 9,4333 | -45,8121 |  |
| y2 |  | -4,6069 | 2,1447 | -3,6790 | 20,7909 |  |
| y3 |  | 9,4333 | -3,6790 | 8,0306 | -44,1727 |  |
| y4 |  | -45,8121 | 20,790 | -44,1727 | 658,3788 |  |

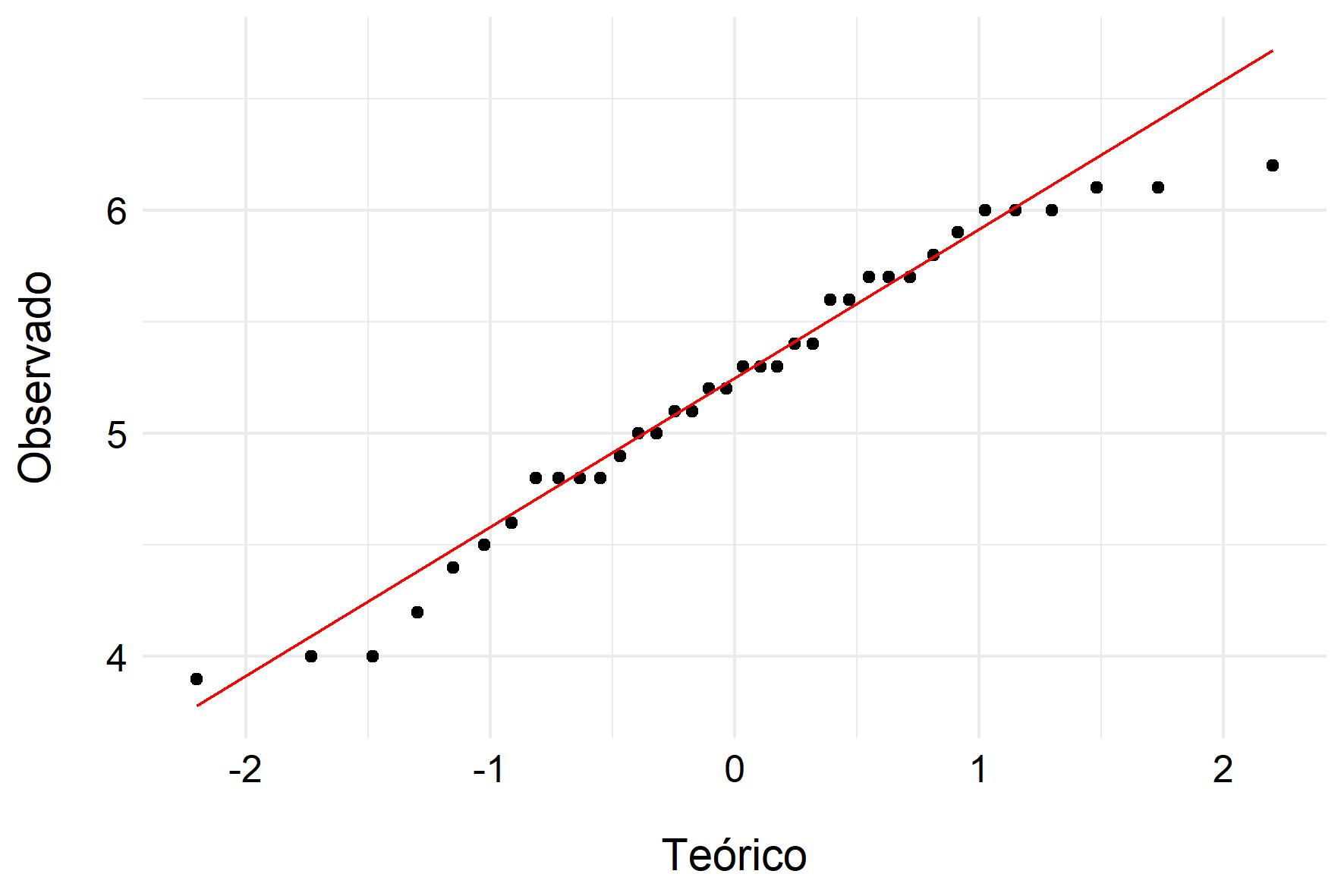
**2. a.1.2 Matriz de correlações geral**

|  |  | y1 | y2 | y3 | y4 |  |
| --- | --- | --- | --- | --- | --- | --- |
| y1 |  | 1,0000 | -0,8004 | 0,8476 | -0,4546 |  |
| y2 |  | -0,8004 | 1,0000 | -0,8859 | 0,5529 |  |
| y3 |  | 0,8476 | -0,8859 | 1,0000 | -0,6075 |  |
| y4 |  | -0,4546 | 0,5529 | -0,6075 | 1,0000 |  |

**2. a.2 Gráficos univariados**

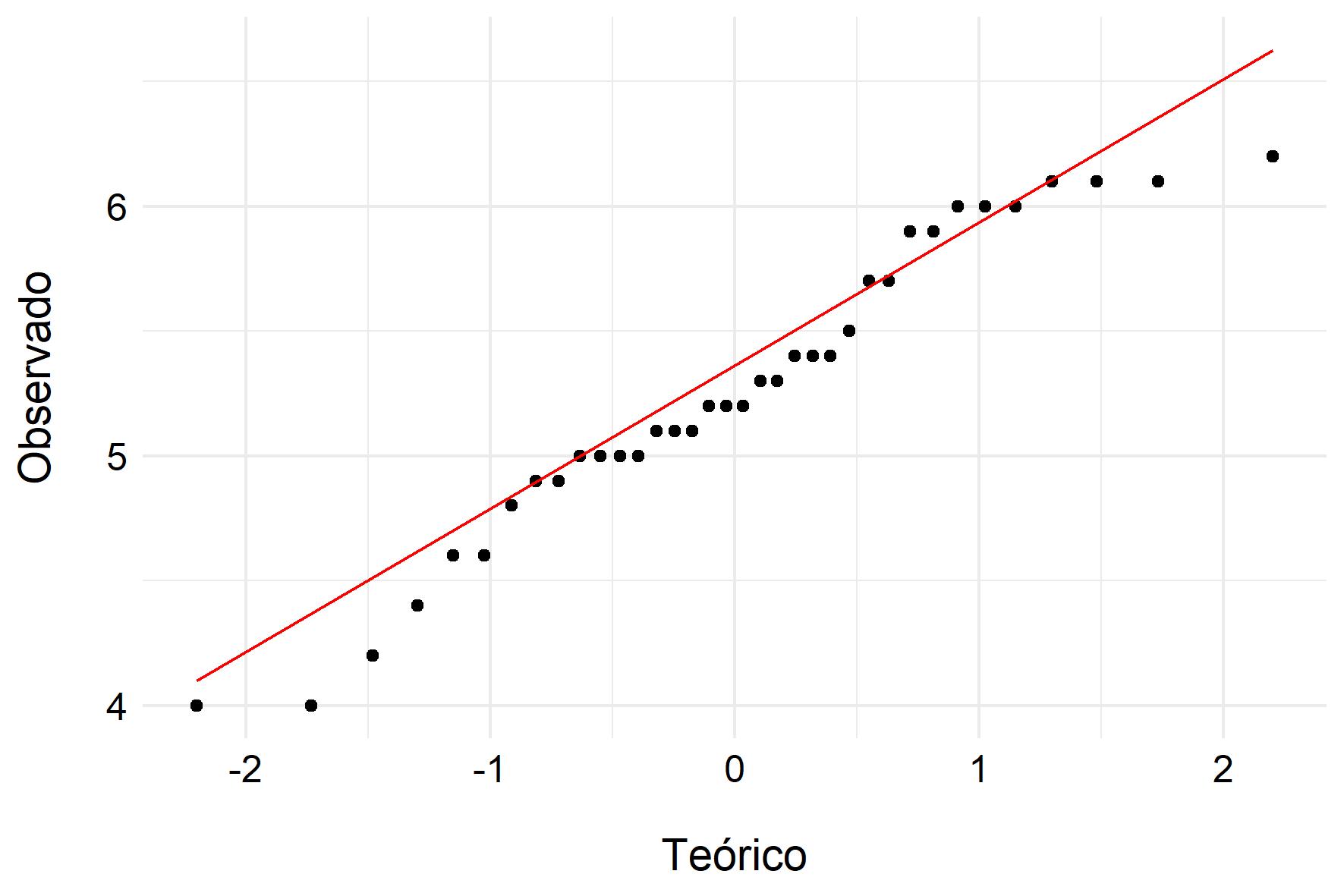
**2. a.2.1 Q-Q plot**

**Figura 34**: Q-Qplot para a variável precocidade do rendimento (y1).



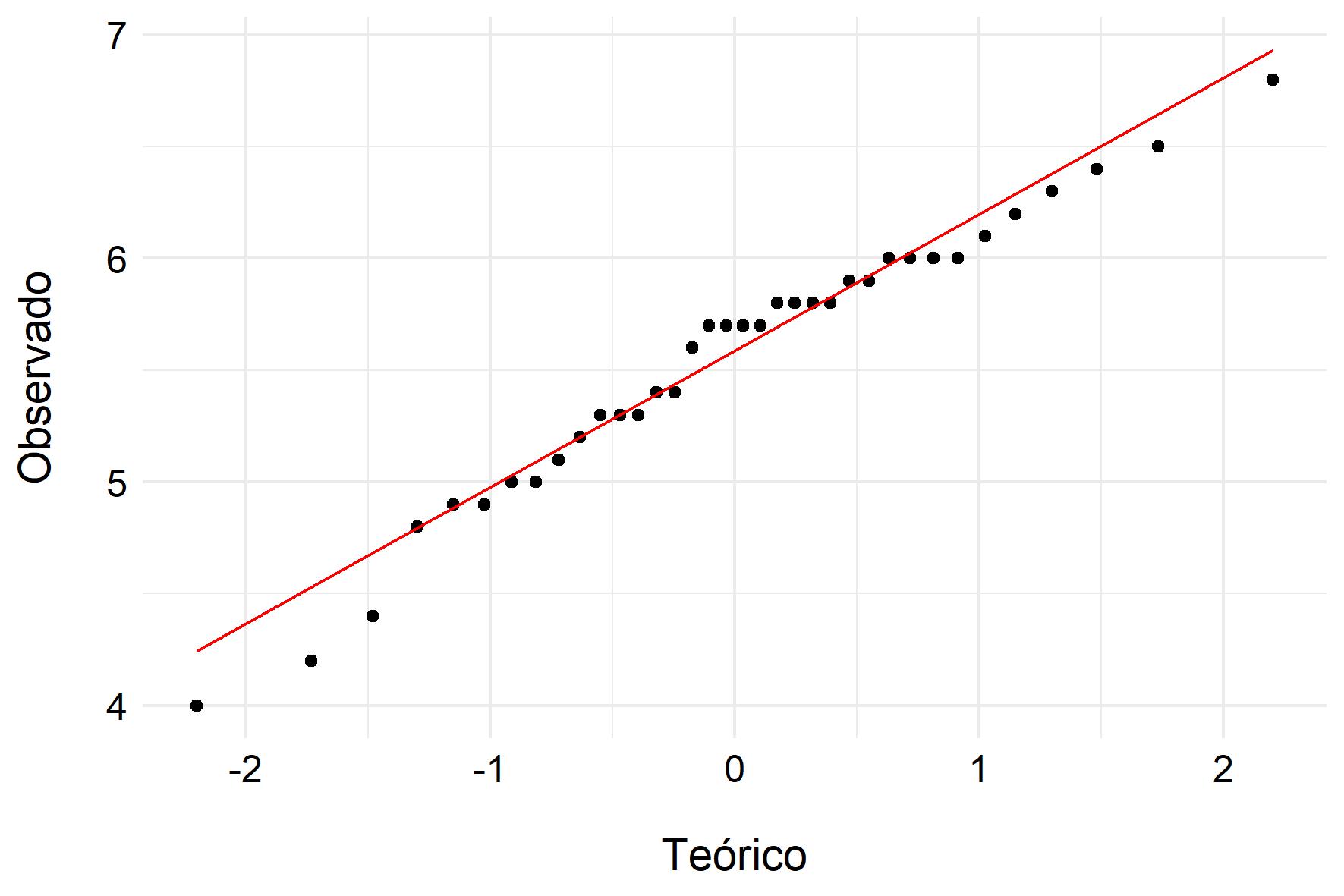
Fonte: Construído pelo autor, 2022.

**Figura 35**: Q-Qplot para a variável precocidade da área foliar específica (y2).



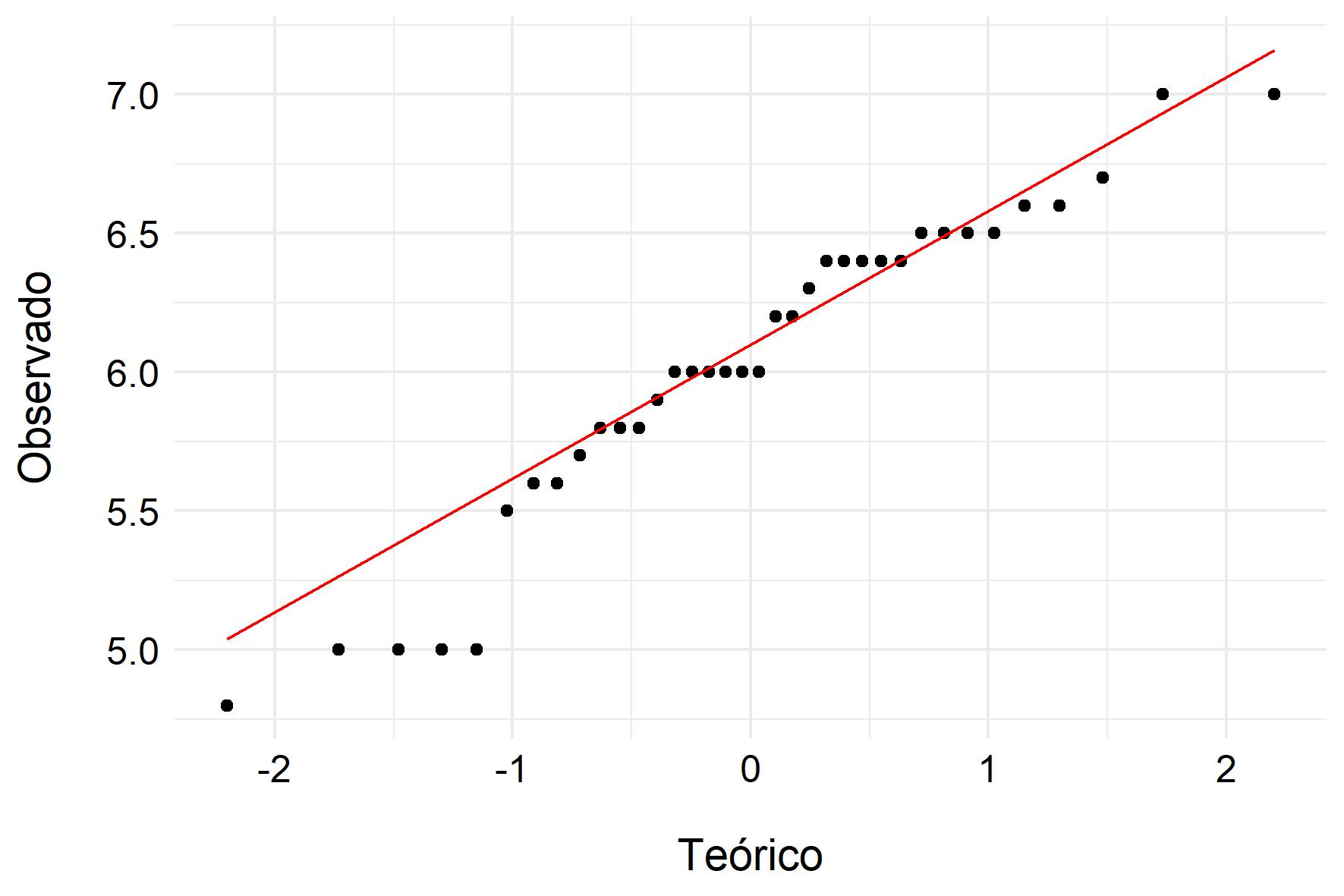
Fonte: Construído pelo autor, 2022.

**Figura 36**: Q-Qplot para a variável rendimento total (y3).



Fonte: Construído pelo autor, 2022.

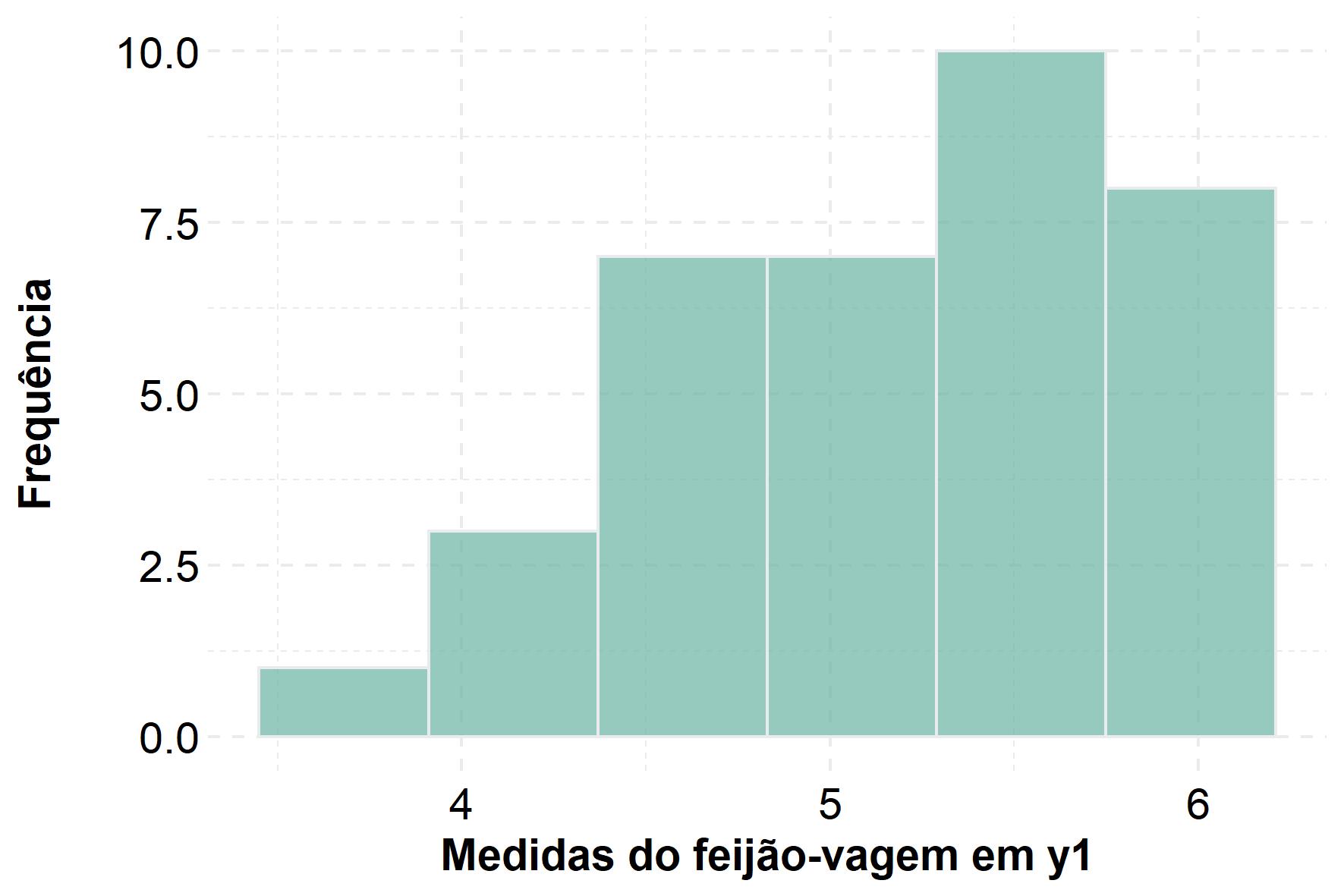
**Figura 37**: Q-Qplot para a variável SLA médio (y4).



Fonte: Construído pelo autor, 2022.

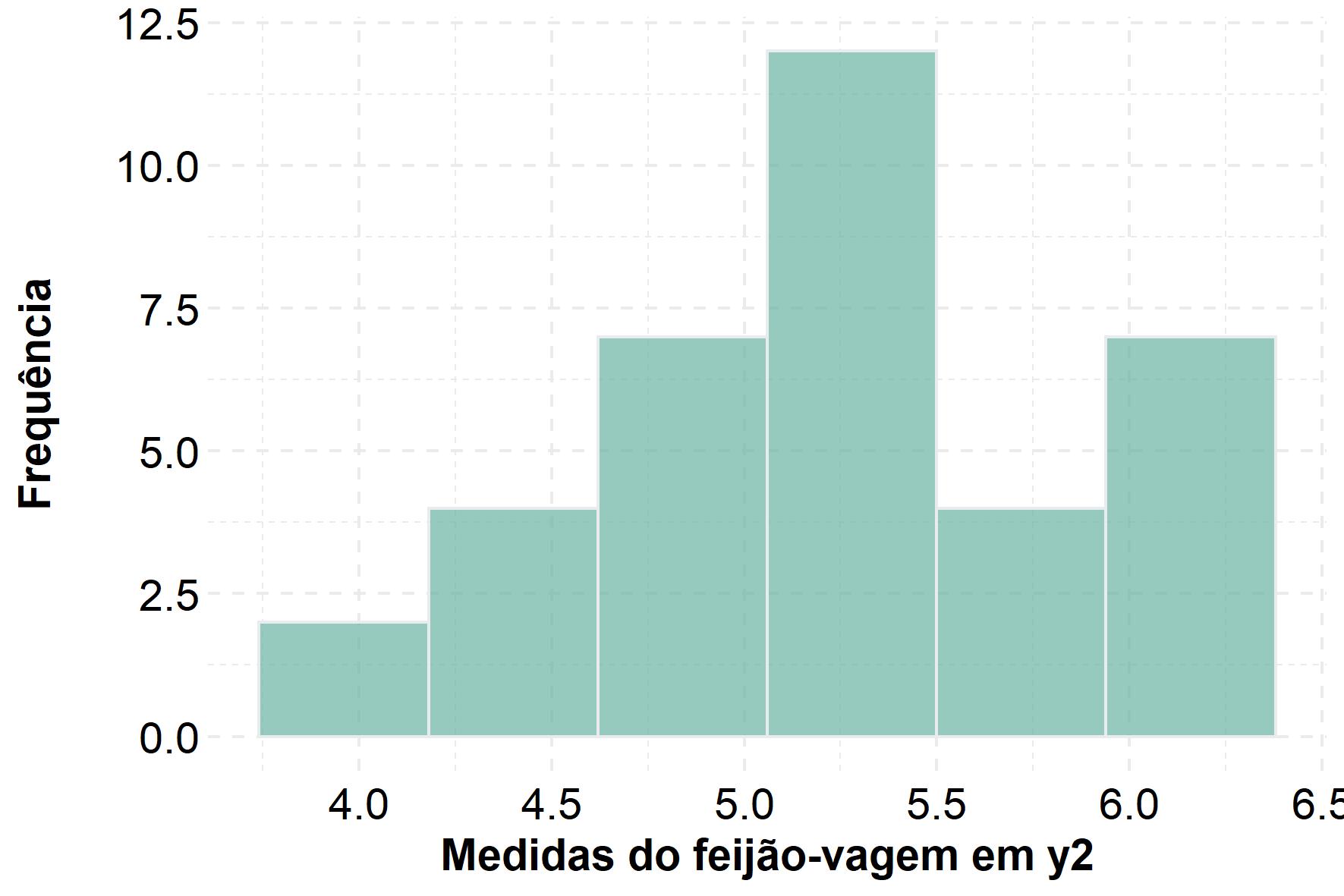
**2. a.2.2 Histograma**

**Figura 36**: Histograma para a variável precocidade do rendimento (y1).



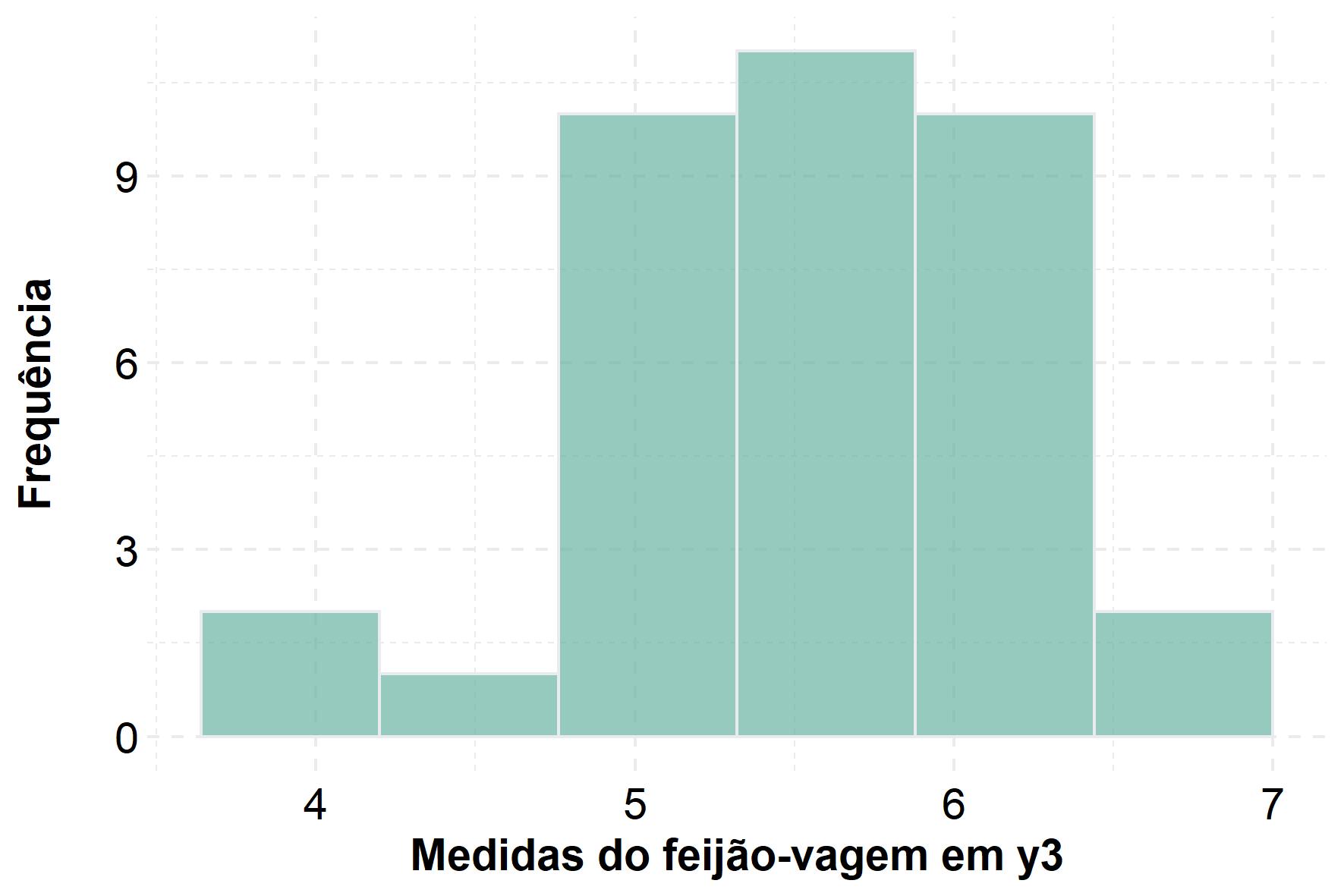
Fonte: Construído pelo autor, 2022.

**Figura 37**: Histograma para a variável precocidade da área foliar específica (y2).



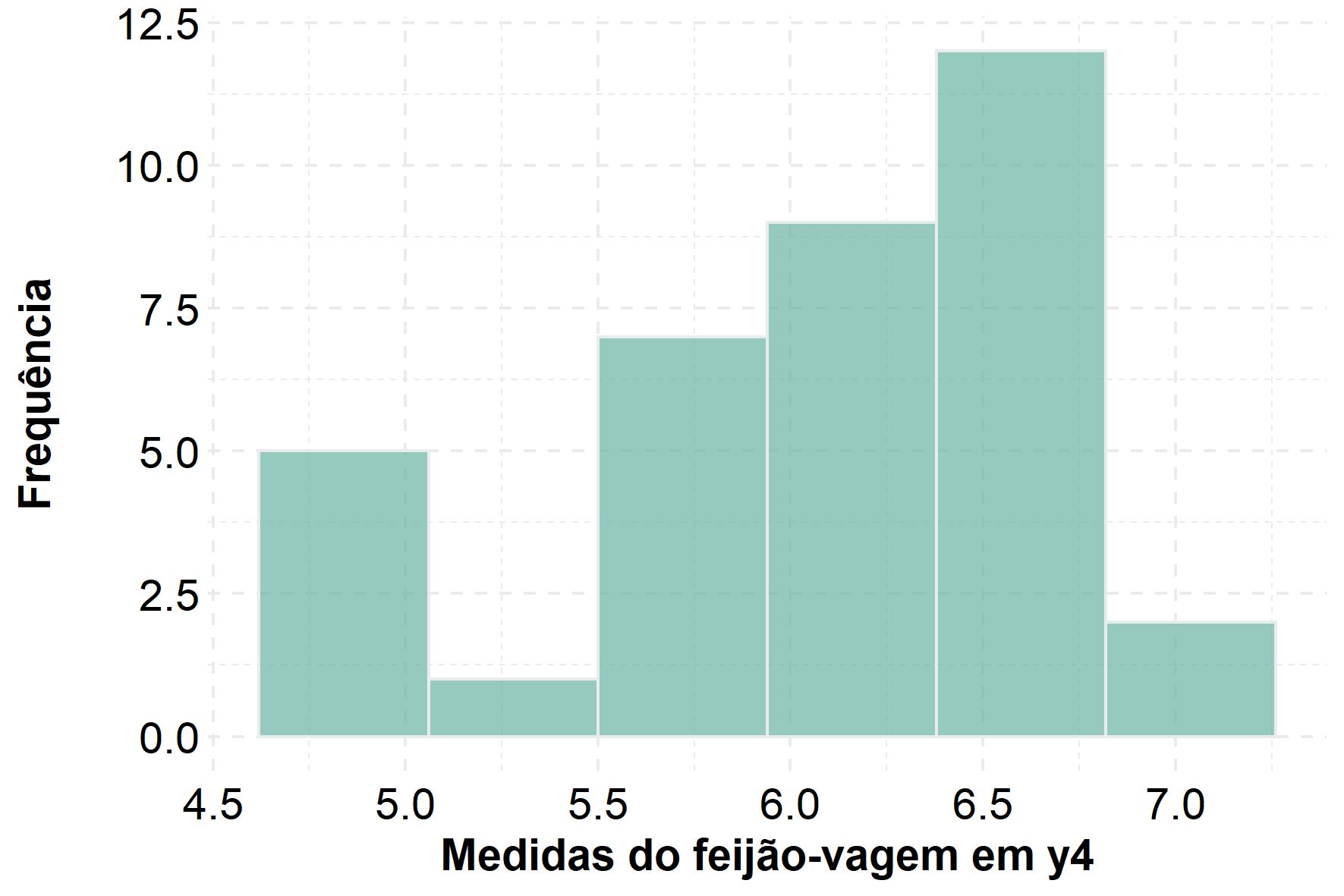
Fonte: Construído pelo autor, 2022.

**Figura 38**: Histograma para a variável rendimento total (y3).



Fonte: Construído pelo autor, 2022.

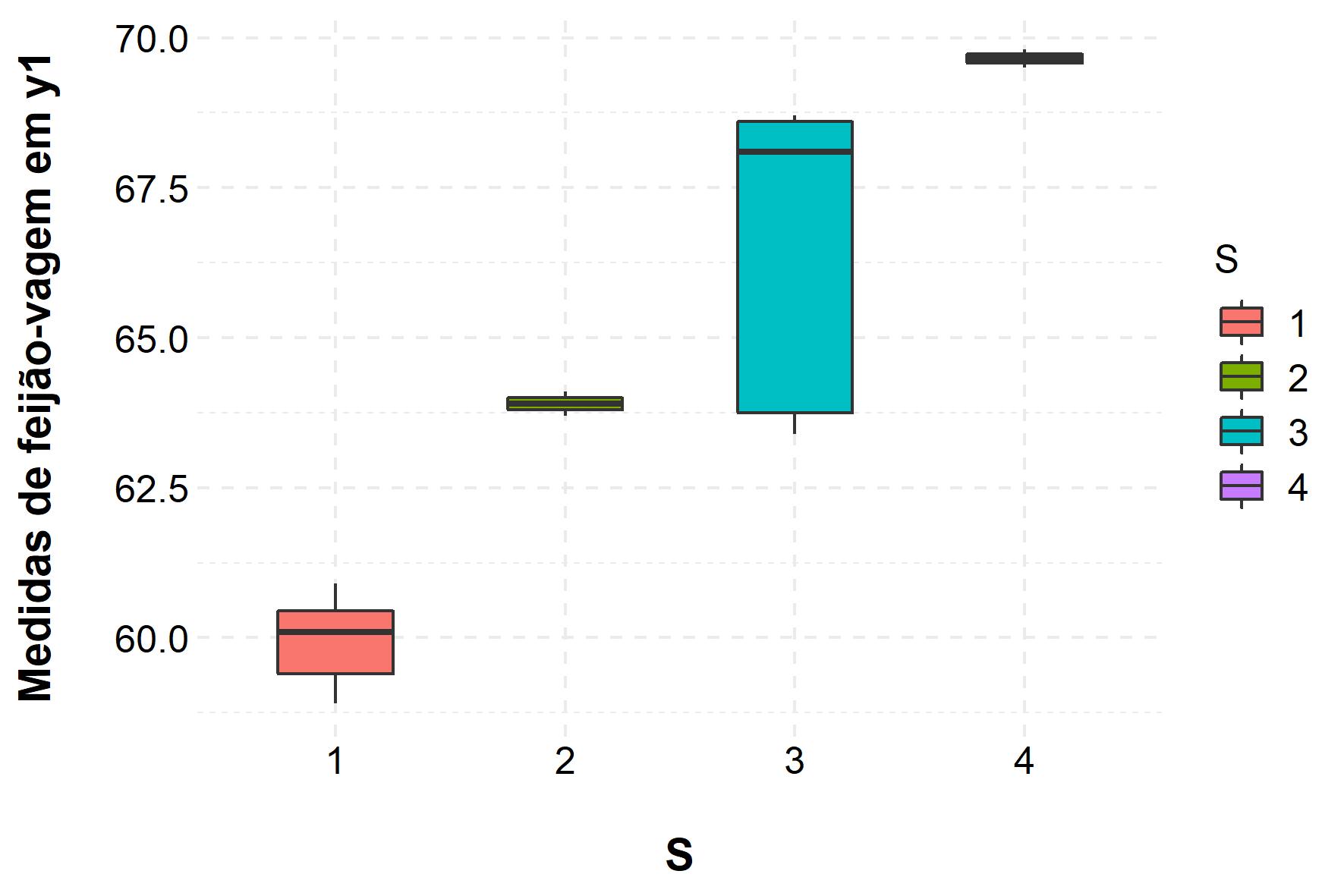
**Figura 39**: Histograma para a variável SLA médio (y4).



Fonte: Construído pelo autor, 2022.

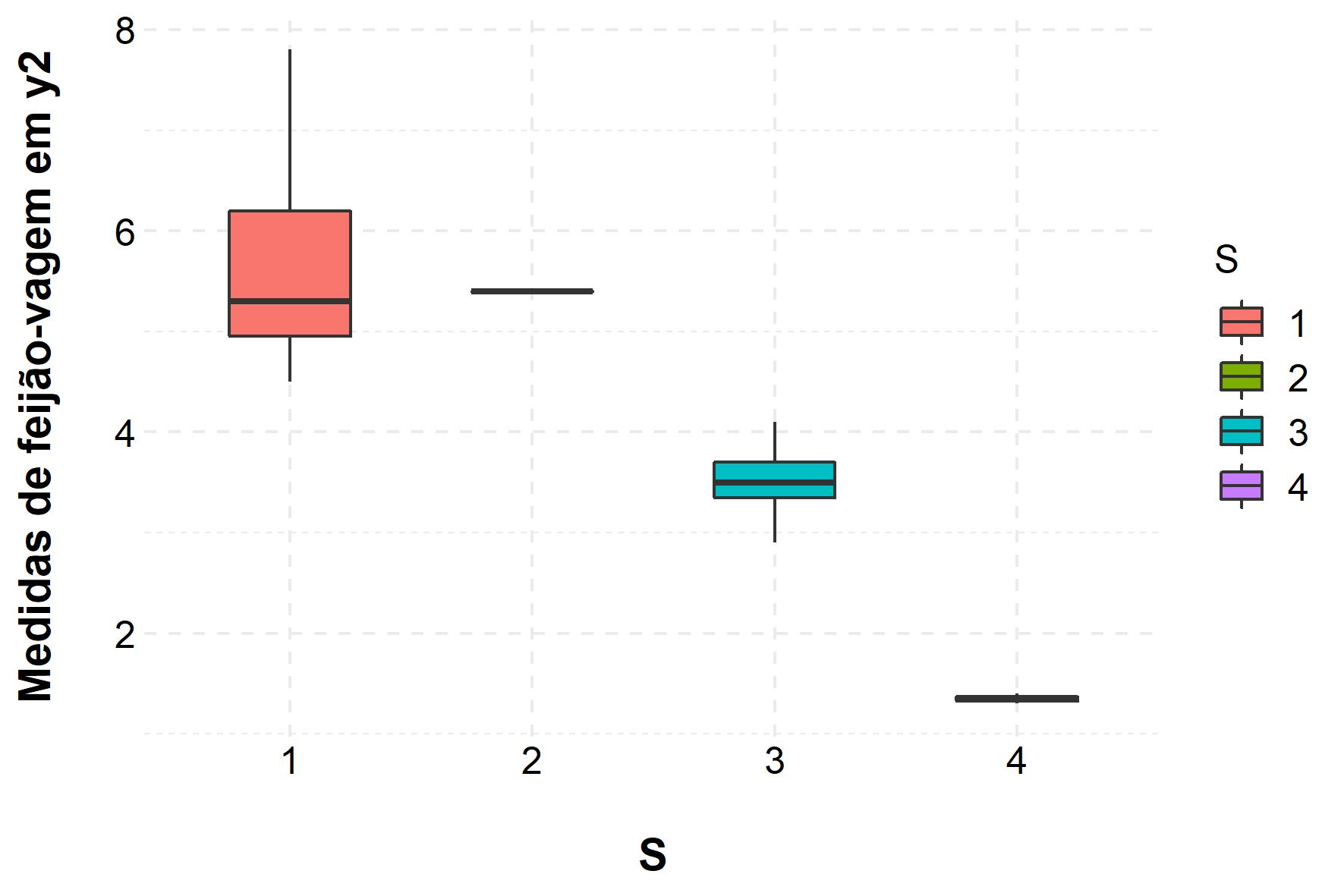
**2. a.2.3 Boxplot**

**Figura 40**: Boxplot para a variável precocidade do rendimento (y1), pelo tipo de semeadura (S).



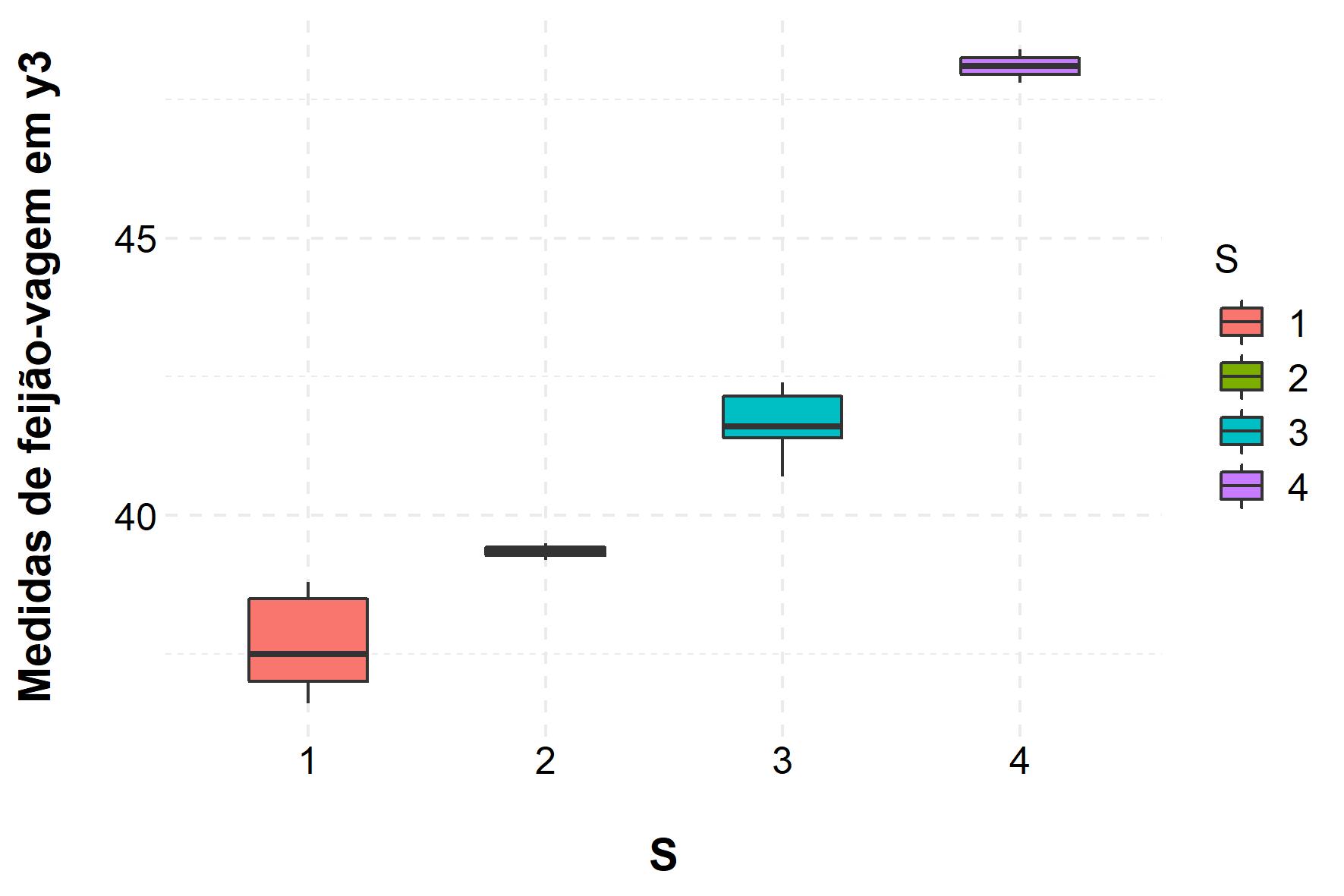
Fonte: Construído pelo autor, 2022.

**Figura 41**: Boxplot para a variável precocidade da área foliar específica (y2), pelo tipo de semeadura (S).



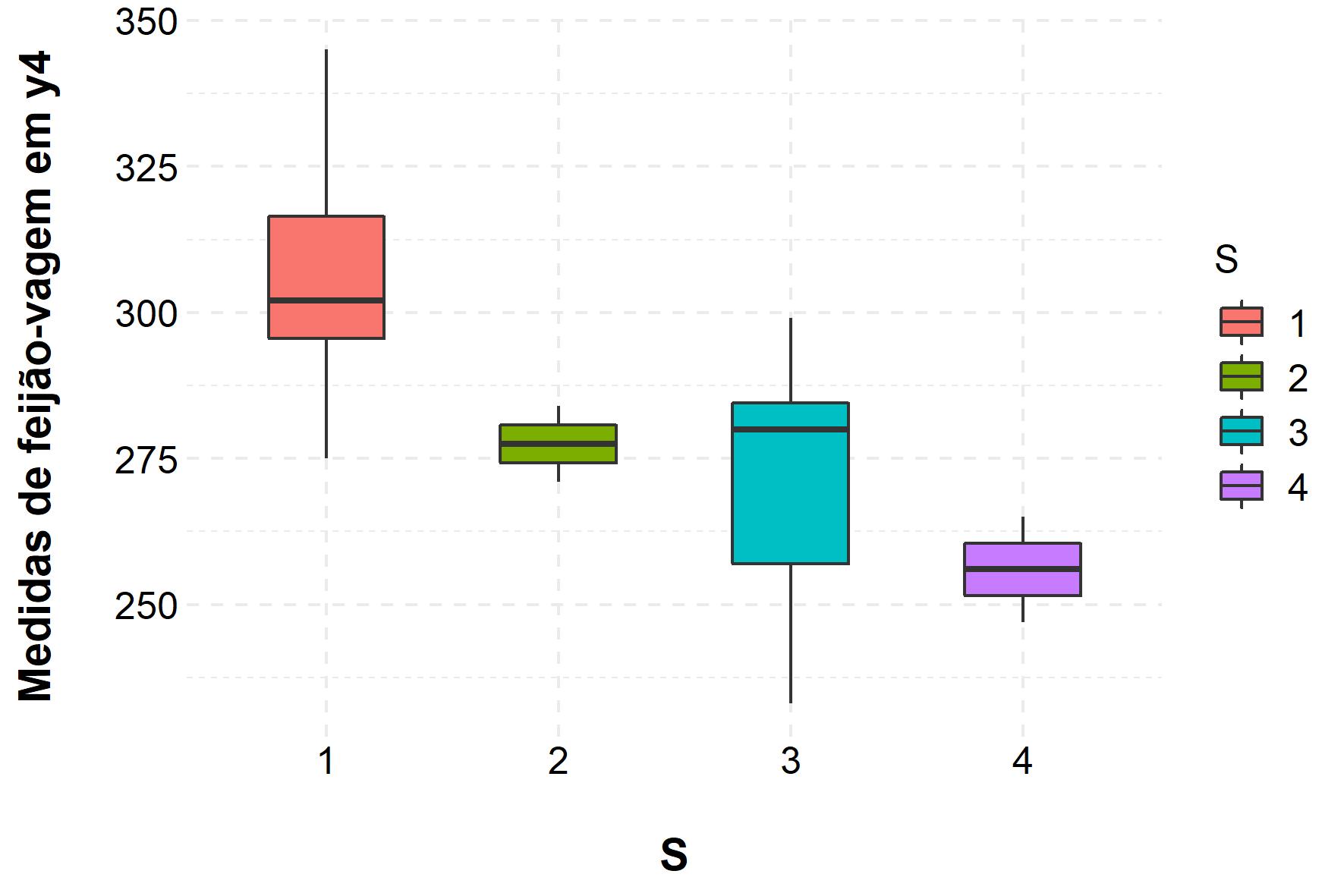
Fonte: Construído pelo autor, 2022.

**Figura 42**: Boxplot para a variável rendimento total (y3), pelo tipo de semeadura (S).



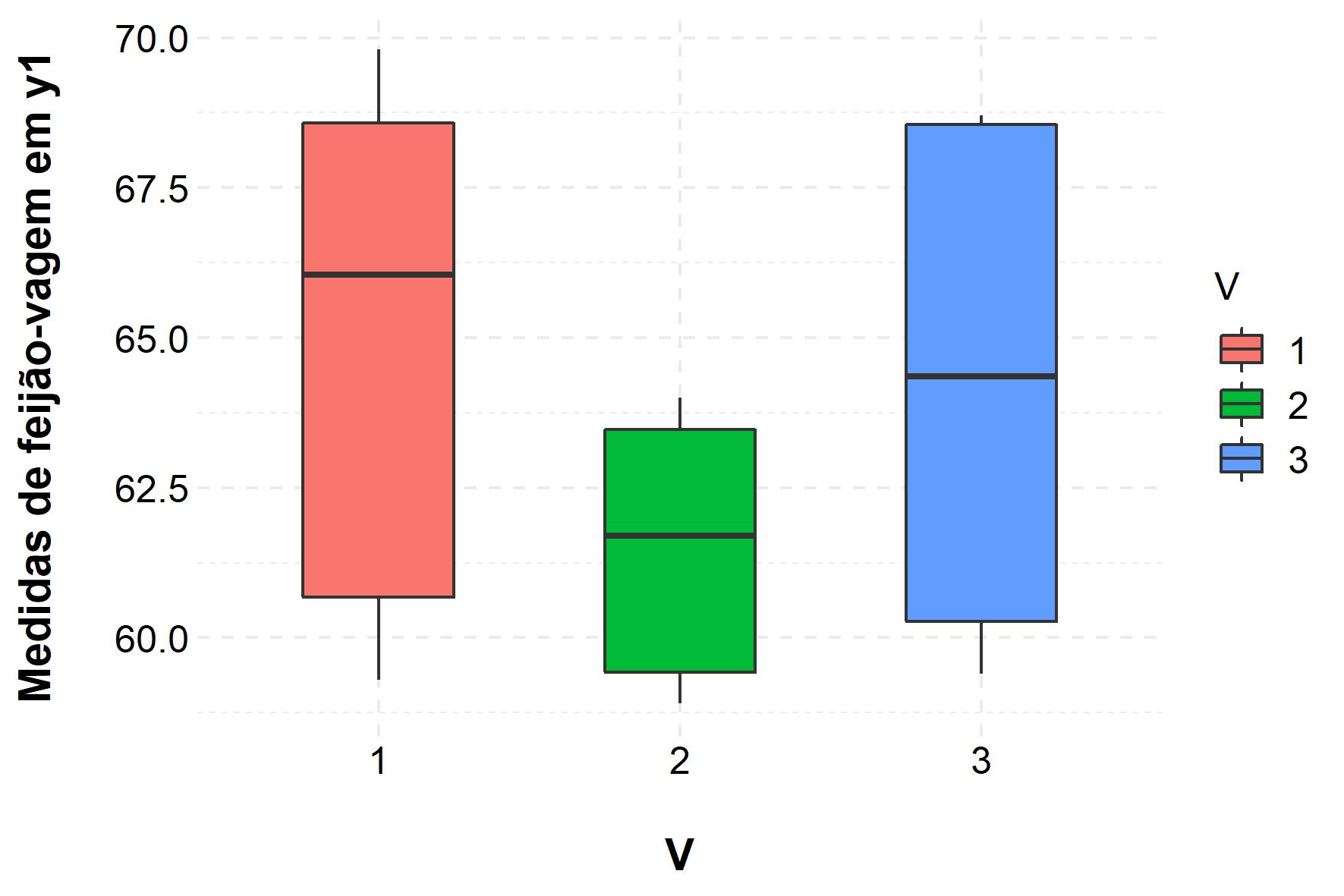
Fonte: Construído pelo autor, 2022.

**Figura 43**: Boxplot para a variável rendimento total (y4), pelo tipo de semeadura (S).



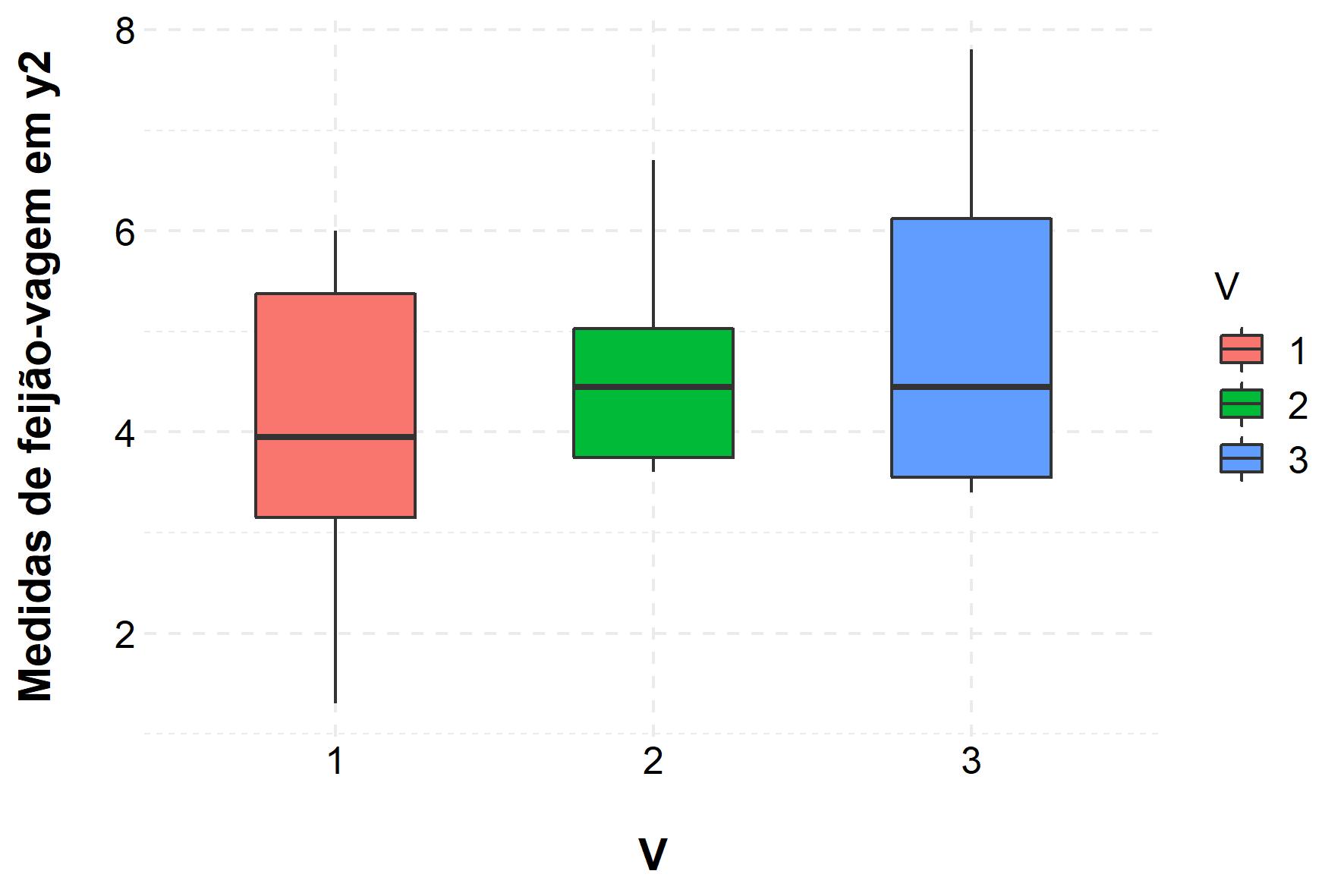
Fonte: Construído pelo autor, 2022.

**Figura 44**: Boxplot para a variável precocidade do rendimento (y1), pelo tipo de variedade (V).



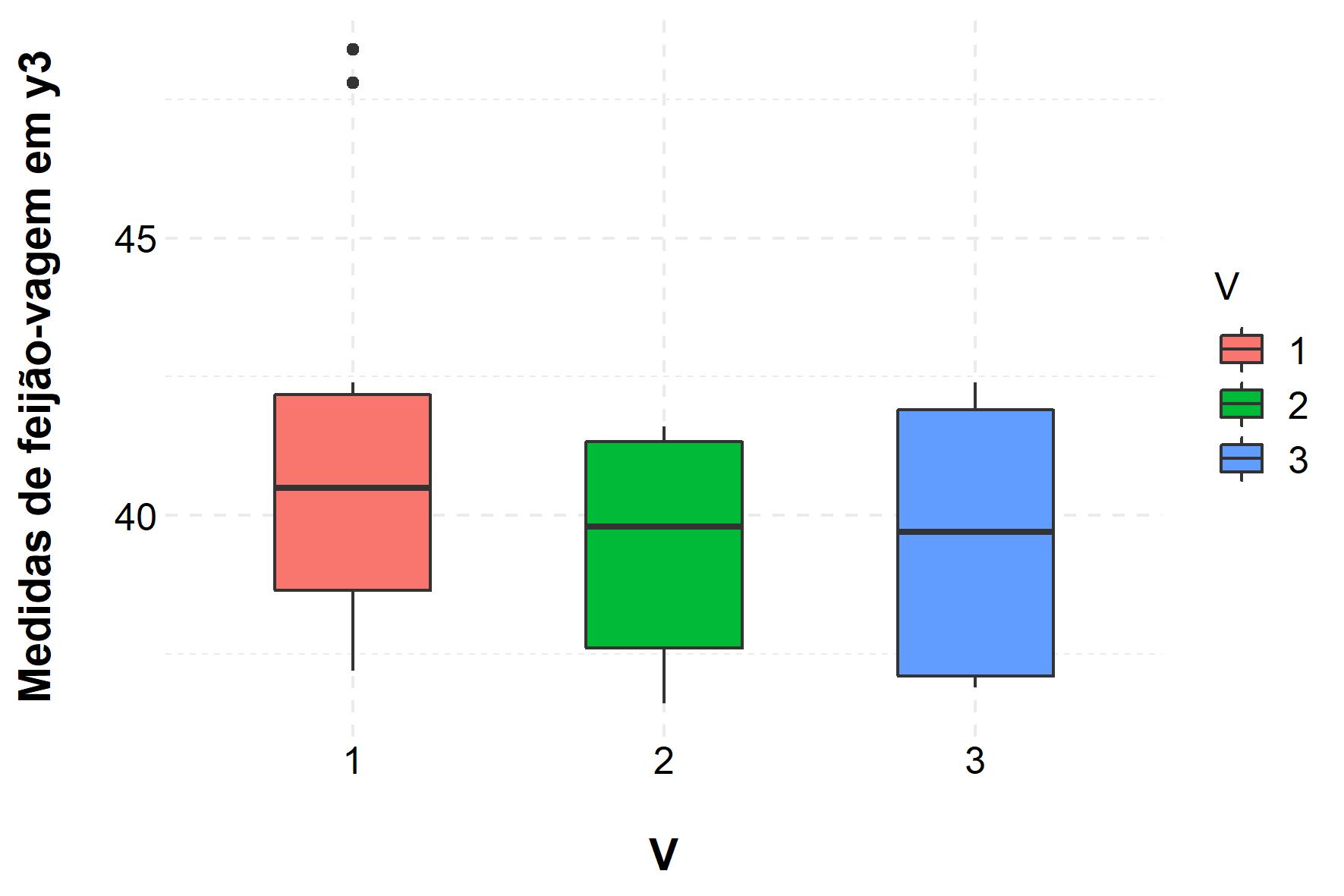
Fonte: Construído pelo autor, 2022.

**Figura 45**: Boxplot para a variável precocidade do rendimento (y2), pelo tipo de variedade (V).



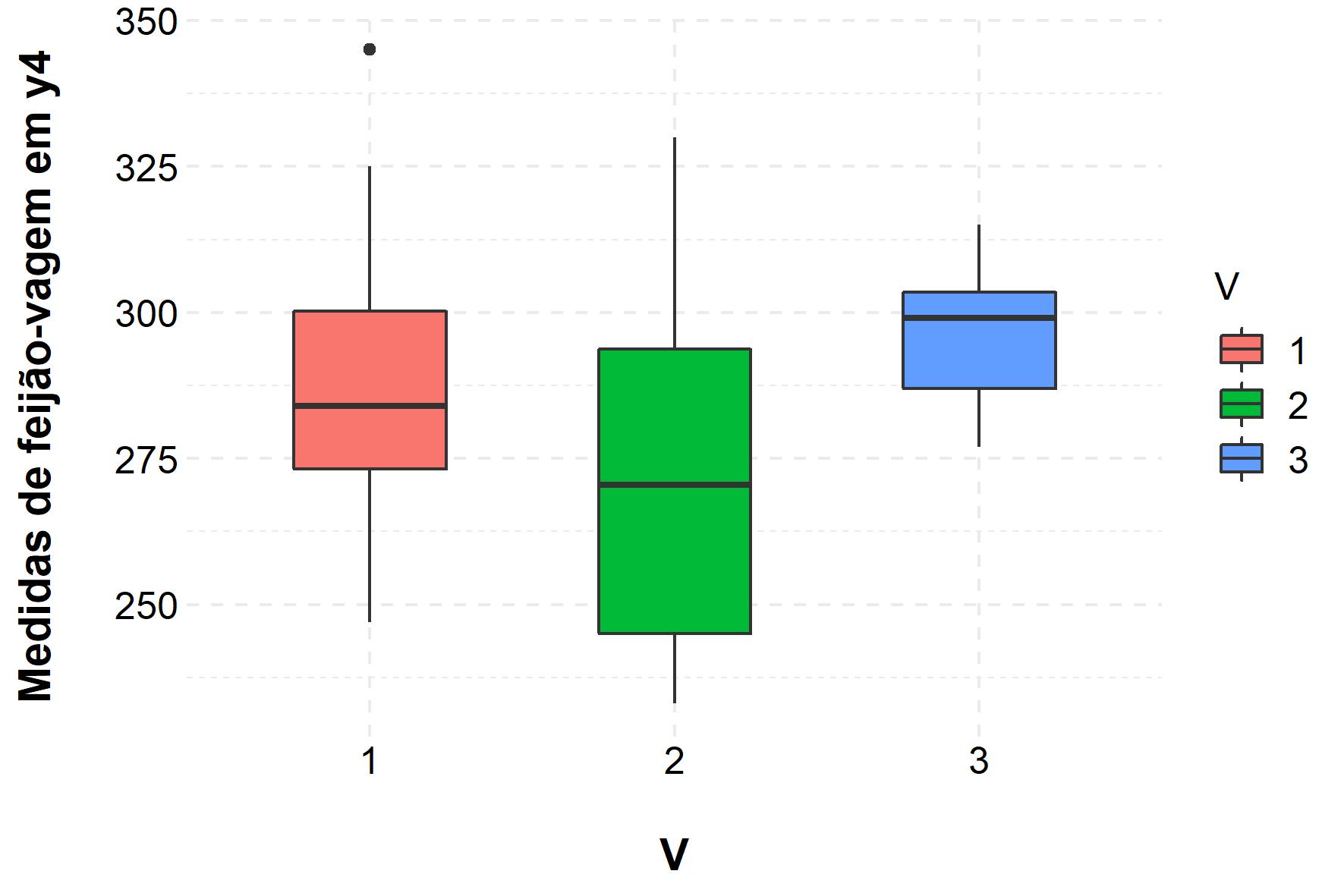
Fonte: Construído pelo autor, 2022.

**Figura 46**: Boxplot para a variável rendimento total (y3), pelo tipo de variedade (V).



Fonte: Construído pelo autor, 2022.

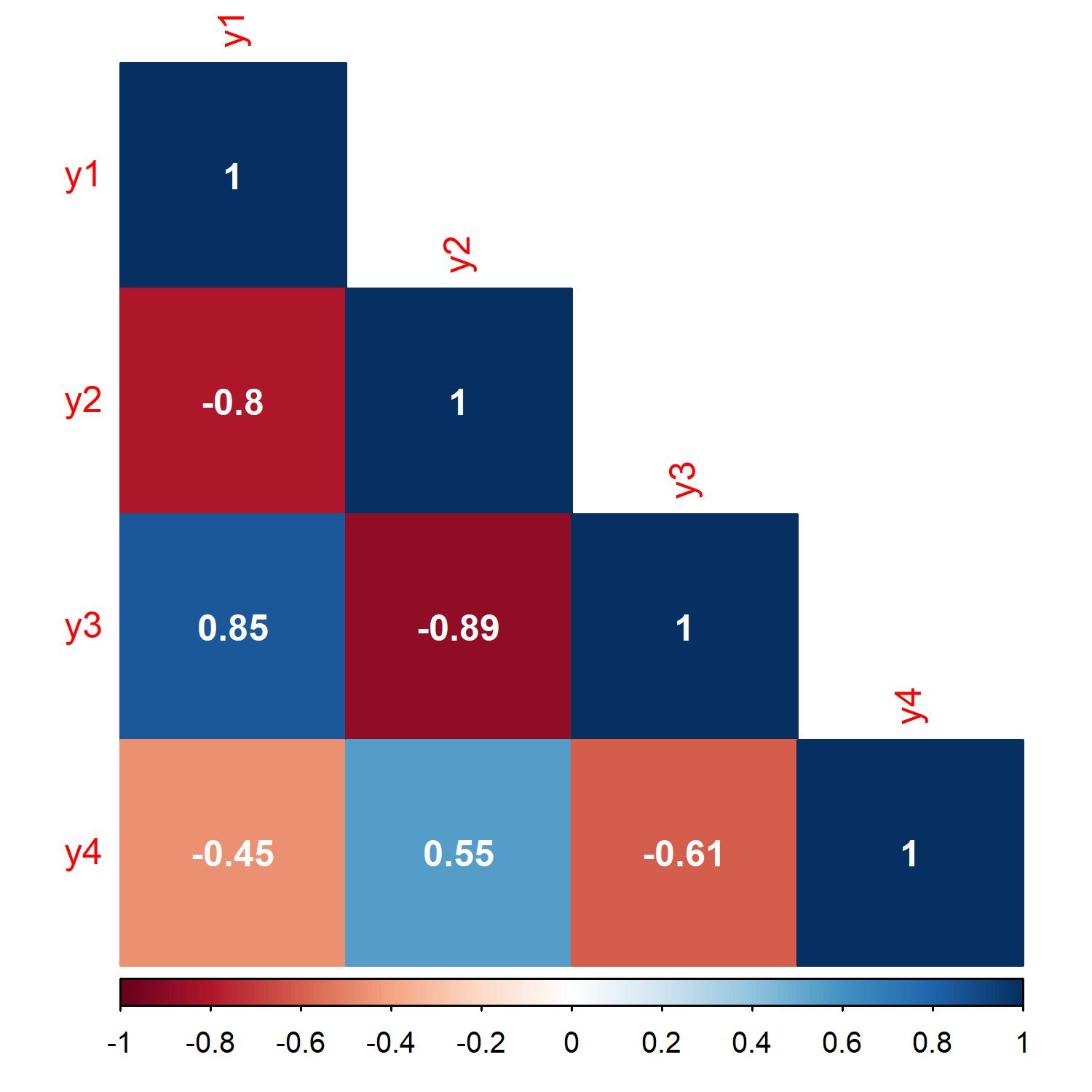
**Figura 47**: Boxplot para a variável SLA médio (y4), pelo tipo de variedade (V).



Fonte: Construído pelo autor, 2022.

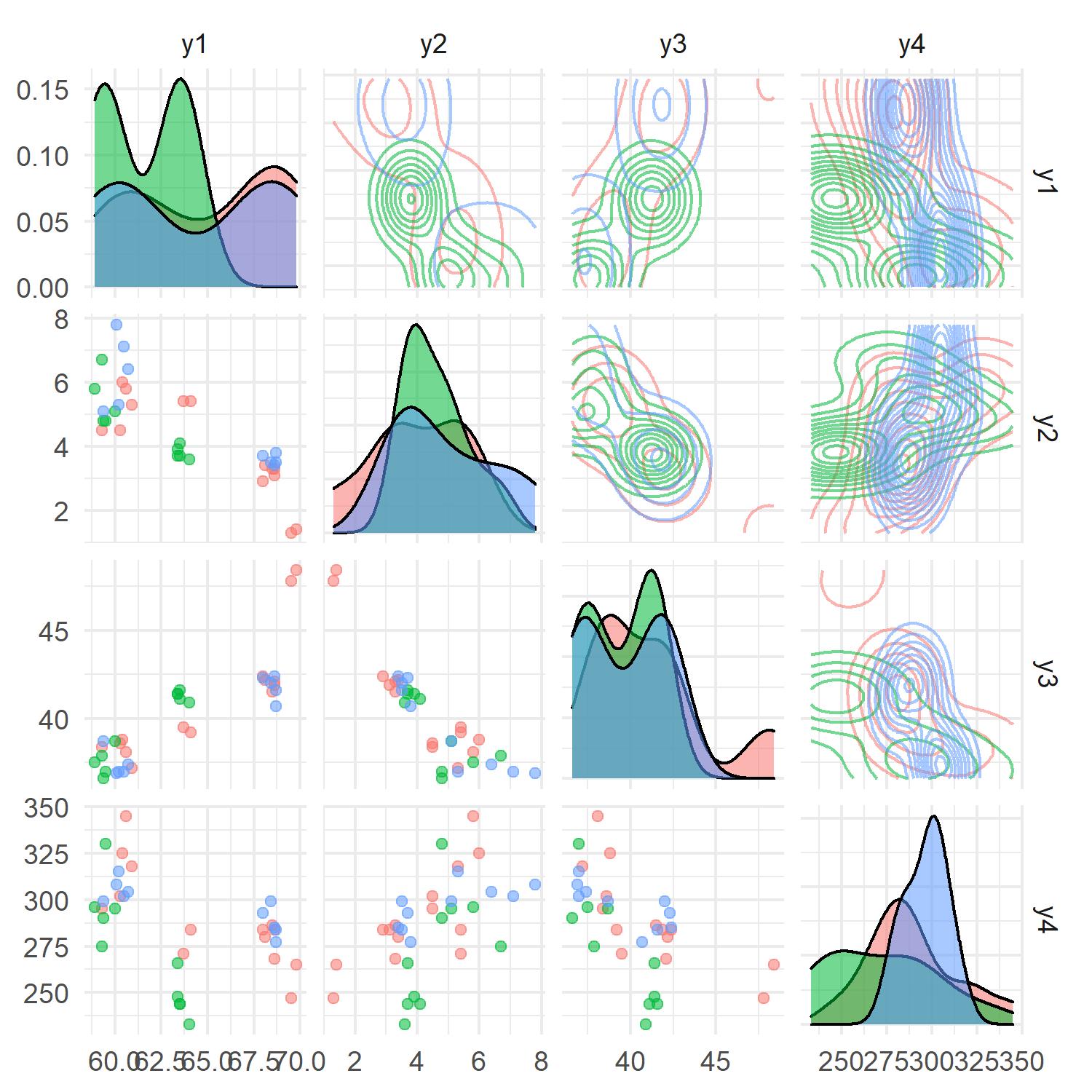
**2. a.3 Gráficos Multivariados**

**Figura 48**: Correlograma para as variáveis do estudo de de feijão e vagem.



Fonte: Construído pelo autor, 2022.

**Figura 49**: Gráficos de densidade, curvas de nível e dispersão dos dados de feijão e vagem, por variedade.



Fonte: Construído pelo autor, 2022.

**Item b):** Realização da MANOVA a dois fatores.

**2. b.1 Verificação dos pressupostos**

**2. b.1.1** **Normalidade multivariada dos dados**

Na realização do teste de normalidade univariada, foi utilizado a metodologia de Shapiro-wilk como forma de avaliar a distribuição normal de cada variável do banco de dados. Como pode ser visto na **Tabela 24**, apenas as variáveis y2 e y4 tem um comportamento probabilístico normal e, y1 e y3 não conseguiram atender tal requisito. Este fato pode quebrar uma das suposições do cálculo da MANOVA.

**Tabela 24**: Resumo do teste de normalidade, contendo a estatística de teste, P-valor e resultado, por variável.

| Variável | Estatística de Teste | P-valor | Resultado |
| --- | --- | --- | --- |
| y1 | 2,1217 | < 0,001 | Não |
| y2 | 0,4437 | 0,2694 | Sim |
| y3 | 1,1291 | 0,0051 | Não |
| y4 | 0,2944 | 0,5784 | Sim |

Fonte: Construído pelo autor, 2022.

A falta da distribuição normal para todas as variáveis do banco de dados já revela que, conjuntamente, y1, y2, y3 e y4 não apresenta uma distribuição normal de quarta dimensão. Entretanto, para comprovar esta hipótese, foi aplicado o teste multivariado de normalidade de Royston, o qual resultou em uma estatística de teste de **11,1145,** associada a um p-valor de **0,0029**. Em conclusão, temos que o vetor de variáveis formado y1, y2, y3 e y4 não segue uma distribuição normal de quarta dimensão, ao nível de 1% de significância.

**2. b.1.2** **Identificação de outliers multivariados**

Na realização do teste da distância de Mahalanobis para a identificação de outliers multivariados, apenas a estratificação pela variedade obteve dados suficientes para aplicação do método. Com isso, temos a ausência de outliers multivariados dentro do banco de dados, pois todas as observações apresentaram distâncias dentro do esperado.

Em uma análise adicional, os boxplots desenvolvidos nas **Figuras 39** a **45** mostram que, quando investigamos a estratificação por variedade, existe a presença de outliers univariados nas variáveis rendimento total (y3) e SLA médio (y4). Por isso, temos outro pressuposto quebrado para a realização da MANOVA.

**2. b.1.3** **Teste de homogeneidade da matriz de variâncias e covariâncias**

O Teste de Box para a homogeneidade da matriz de variâncias e covariâncias foi também aplicado para verificar se a matriz de variâncias e covariâncias são iguais entre as variedades e semeaduras. Todavia, o teste para os tipos de semeadura apresentou erros por haver uma categoria com poucos elementos para as operações matemáticas.

Para a aplicação nas variedades dos dados de feijão e vagem temos a formalização do teste como:

Hipóteses:

H0: Existe homogeneidade na matriz de variâncias e covariâncias das variedades.

H1: Não existe homogeneidade na matriz de variâncias e covariâncias das variedades.

Estatística de teste = 44,60

P-valor = 0,0013

Conclusão:

Como o P-valor é menor do que o nível de significância, isto é, 0,0013 < 0,01, não rejeitamos H0. Logo, existe homogeneidade na matriz de variâncias e covariâncias das variedades, ao nível de 1% de significância. Com isso, temos que o pressuposto de igualdade da matriz de variâncias e covariâncias foi atendido para as variedades.

**2. b.1.4** **Homogeneidade de variâncias**

Outro pressuposto para a realização da MANOVA é a verificação da igualdade de variâncias entre as variáveis independentes do estudo. Nesta visão, aplicamos o Teste de Levene para a homogeneidade de Variâncias para investigar tal característica.

Hipóteses:

H0: Existe homogeneidade de variâncias entre os tipos de semeadura.

H1: Não existe homogeneidade de variâncias entre os tipos de semeadura.

Na **Tabela 25**, pode-se verificar os resultados do Teste de Levene para cada variável, considerando os tipos de semeadura. Logo, temos que apenas a variável precocidade do rendimento (y1) apresentou um p-valor abaixo do nível de significância adotado (0,01). Desse modo, rejeita-se, com 1% de significância, a hipótese nula para a variável y1 e, assume-se para as demais a homogeneidade de variâncias.

**Tabela 25**: Estatística de Teste e P-valor para o teste de Levene para os tipos de semeadura, por variável.

| Variável | Estatística F | P-Valor |
| --- | --- | --- |
| y1 | 22,8580 |  |
| y2 | 6,8113 | 0,0012 |
| y3 | 3,6615 | 0,0232 |
| y4 | 0,9526 | 0,4277 |

Fonte: Construído pelo autor, 2022.

Em adição, aplicou-se também o teste de Levene para a homogeneidade de Variâncias considerando o tipo de variedade como a variável de grupos. Com isso, temos:

Hipóteses:

H0: Existe homogeneidade de variâncias entre os tipos de variedade.

H1: Não existe homogeneidade de variâncias entre os tipos de variedade.

Por meio da Tabela 26, temos os resultados do teste de Levene, considerando as variedades, para todas as variáveis do estudo (y1, y2, y3 e y4). Logo, novamente, percebe-se a rejeição da hipótese nula apenas para a variável preciosidade do rendimento (y1) e, a não rejeição de H0 para as demais variáveis. Em suma, ao nível de 1% de significância, a variável y1 não apresenta a homogeneidade de variâncias entre as categorias de variedades, em contraste, as variedades y2, y3 e y4 apresentam tal característica.

**Tabela 26**: Estatística de Teste e P-valor para o teste de Levene para os tipos de variedade, por variável.

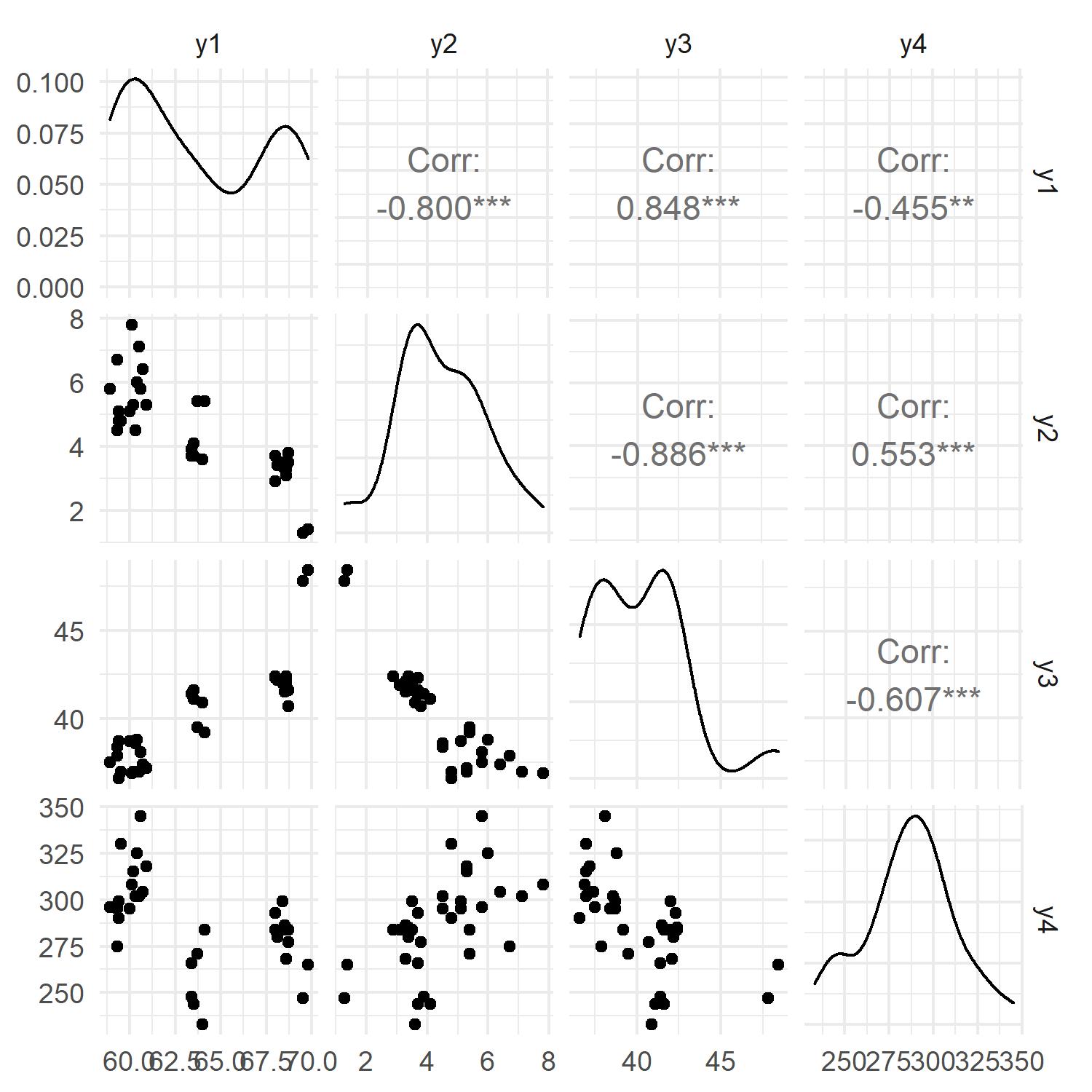
| Variável | Estatística F | P-Valor |
| --- | --- | --- |
| y1 | 15,1740 |  |
| y2 | 1,9171 | 0,1641 |
| y3 | 0,7846 | 0,4652 |
| y4 | 3,3037 | 0,4652 |

Fonte: Construído pelo autor, 2022.

**2. b.1.5** **Verificação da ausência de multicolinearidade**

A **Figura 50** aborda as relações de dispersão em pares e correlações lineares entre as variáveis do estudo. A partir disto, temos que as variáveis preciosidade do rendimento (y1) e preciosidade da área folicular específica (y2) apresentam uma correlação forte negativa de -0,8000. As variáveis preciosidade do rendimento (y1) e rendimento total (y3) apresentam uma correlação forte positiva de 0,8480. As variáveis preciosidade da área folicular específica (y2) e rendimento total (y3) apresentam uma correlação linear forte e negativa de -0,8860. Tais valores próximos de 0,90 podem ser um indício de multicolinearidade, porém, como nenhum ultrapassou este limite, iremos supor que os dados atendem à ausência de multicolinearidade.

**Figura 50**: Gráficos de densidade, correlações lineares e dispersão dos dados de feijão e vagem.

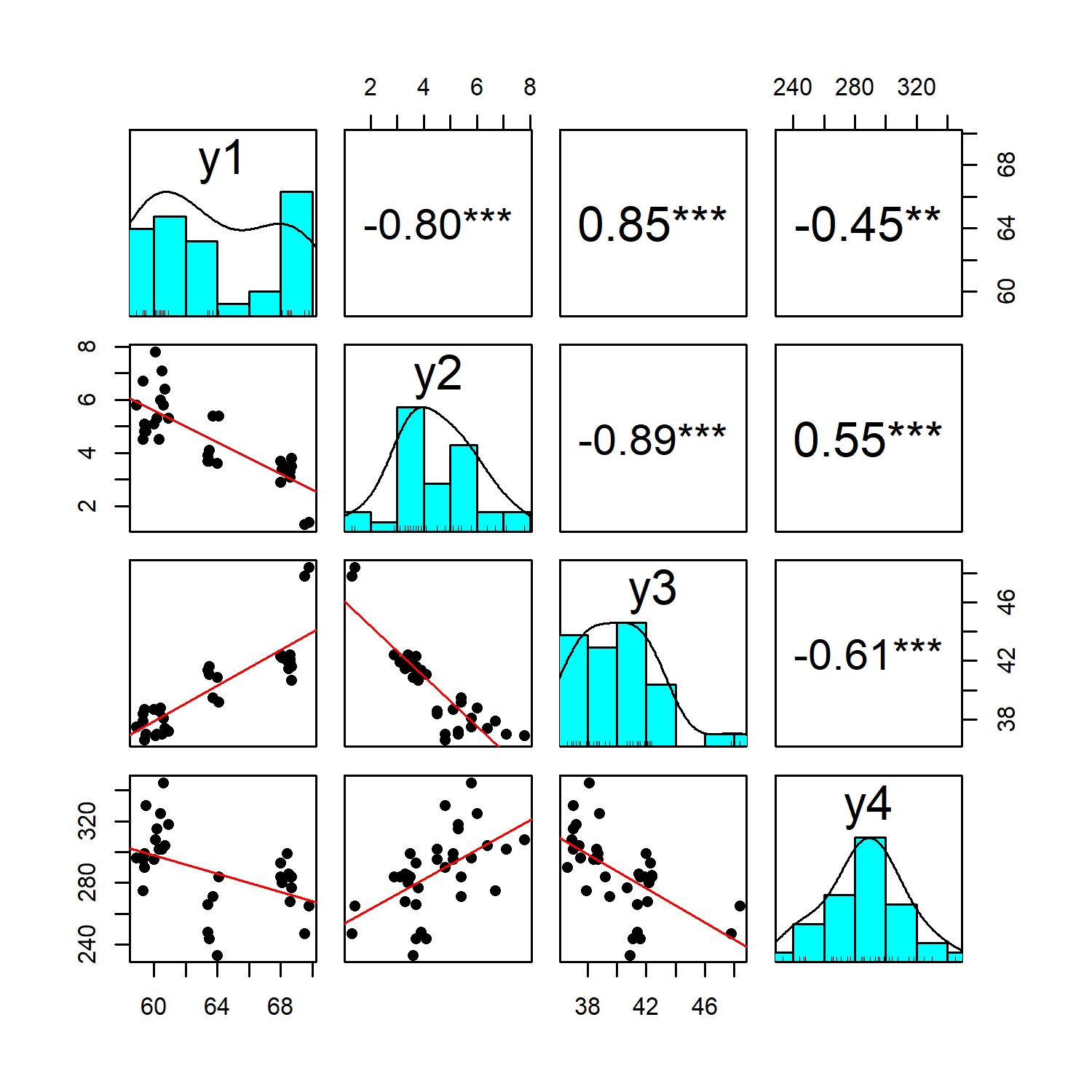


Fonte: Construído pelo autor, 2022.

**2. b.1.5** **Verificação da ausência da relação linear das variáveis dependentes para cada grupo**

A partir da análise na **Figura 51**, podemos ter uma noção se os pares de variáveis apresentam um ajuste linear. A partir disto, podemos afirmar que a reta vermelha criada para representar o modelo de regressão linear em qualquer par de variáveis não conseguia explicar uma relação de primeira ordem. Com isso, temos que não existe a relação linear entre as variáveis do estudo.

**Figura 51**: Histogramas, correlações e ajustes lineares para os dados de feijão e vagem.



Fonte: Construído pelo autor, 2022.

**2. b.2** **Realização da MANOVA a dois fatores (two way)**

Em revisão, para a realização da MANOVA a dois fatores, temos que a normalidade multivariada dos dados não foi atendida, a homogeneidade de variâncias não foi atendida para a variável y, quando leva-se em conta os tipos de variedade e, para os tipos de semeadura não foi possível aplicar o teste de homogeneidade da matriz de variâncias e covariâncias e o teste de Levene para homogeneidade de variâncias, por haver uma categoria com observações insuficientes.

Em contraste, mesmo com alguns pressupostos violados, iremos prosseguir com a construção da MANOVA a dois fatores. Dessa maneira, a **Tabela 27** aborda os resultados da MANOVA pelo teste de Lambda de Wilks; a **Tabela 28** aborda os resultados da MANOVA pelo teste do Traço de Pillai; a **Tabela 29** aborda os resultados da MANOVA pelo teste do Traço de Hotelling-Lawley e; a **Tabela 30** aborda os resultados da MANOVA pelo teste da Raiz máxima de Roy.

Em todos todos o testes iremos considerar as seguintes hipóteses:

**Para a Semeadura:**

H0: O tipo de semeadura não interfere nas variáveis preciosidade do rendimento (y1), preciosidade da área folicular específica (y2), rendimento total (y3) e SLA Médio (y4) conjuntamente.

H1: O tipo de semeadura interfere nas variáveis preciosidade do rendimento (y1), preciosidade da área folicular específica (y2), rendimento total (y3) e SLA Médio (y4) conjuntamente.

De acordo com os p-valores gerados nos 4 testes presentes na primeira linha das **Tabelas 27, 28, 29** e **30**, verifica-se a rejeição da hipótese nula ao nível de 1% de significância. Portanto, para as metodologias Lambda de Wilks, Traço de Pillai, Traço de Hotelling-Lawley e Raiz máxima de Roy, o tipo de semeadura interfere nas variáveis preciosidade do rendimento (y1), preciosidade da área folicular específica (y2), rendimento total (y3) e SLA Médio (y4) conjuntamente.

**Para a Variedade:**

H0: A variedade não interfere nas variáveis preciosidade do rendimento (y1), preciosidade da área folicular específica (y2), rendimento total (y3) e SLA Médio (y4) conjuntamente.

H1: A variedade interfere nas variáveis preciosidade do rendimento (y1), preciosidade da área folicular específica (y2), rendimento total (y3) e SLA Médio (y4) conjuntamente.

Em conformidade com os p-valores gerados nos 4 testes presentes na segunda linha das **Tabelas 27, 28, 29** e **30**, verifica-se a rejeição da hipótese nula ao nível de 1% de significância. Portanto, para as metodologias Lambda de Wilks, Traço de Pillai, Traço de Hotelling-Lawley e Raiz máxima de Roy, a variedade interfere nas variáveis preciosidade do rendimento (y1), preciosidade da área folicular específica (y2), rendimento total (y3) e SLA Médio (y4) conjuntamente.

**Para a interação entre Semeadura e Variedade:**

H0: A interação entre o tipo de semeadura e variedade não interfere nas variáveis preciosidade do rendimento (y1), preciosidade da área folicular específica (y2), rendimento total (y3) e SLA Médio (y4) conjuntamente.

H1: A interação entre o tipo de semeadura e variedade interfere nas variáveis preciosidade do rendimento (y1), preciosidade da área folicular específica (y2), rendimento total (y3) e SLA Médio (y4) conjuntamente.

Como os p-valores gerados nos 4 testes presentes na terceira linha das **Tabelas 27, 28, 29** e **30**, rejeitamos a hipótese nula ao nível de 1% de significância. Portanto, para as metodologias Lambda de Wilks, Traço de Pillai, Traço de Hotelling-Lawley e Raiz máxima de Roy, a interação entre o tipo de semeadura e variedade interfere nas variáveis preciosidade do rendimento (y1), preciosidade da área folicular específica (y2), rendimento total (y3) e SLA Médio (y4) conjuntamente.

**Tabela 27**: Estatística de Teste, Valor aproximado a estatística F, Graus de liberdade e P-valor para a Análise de Variância Multivariada, para o teste Lambda de Wilks.

| Causa de Variação | Estatística de Teste | Valor F | GL1 | GL2 | P-Valor |
| --- | --- | --- | --- | --- | --- |
| S | 0,0011 | 60,8510 | 12 | 61 |  |
| V | 0,0421 | 22,2850 | 8 | 46 |  |
| (S x V) | 0,0901 | 13,4020 | 8 | 46 |  |

Fonte: Construído pelo autor, 2022.

**Tabela 28**: Estatística de Teste, Valor aproximado a estatística F, Graus de liberdade e P-valor para a Análise de Variância Multivariada, para o teste Traço de Pillai.

| Causa de Variação | Estatística de Teste | Valor F | GL1 | GL2 | P-Valor |
| --- | --- | --- | --- | --- | --- |
| S | 1,8424 | 9,9475 | 12 | 75 |  |
| V | 1,1985 | 8,9716 | 8 | 48 |  |
| (S x V) | 1,0340 | 6,4218 | 8 | 48 |  |

Fonte: Construído pelo autor, 2022.

**Tabela 29**: Estatística de Teste, Valor aproximado a estatística F, Graus de liberdade e P-valor para a Análise de Variância Multivariada, para o teste Traço de Hotelling-Lawley.

| Causa de Variação | Estatística de Teste | Valor F | GL1 | GL2 | P-Valor |
| --- | --- | --- | --- | --- | --- |
| S | 212,385 | 383,4700 | 12 | 65 |  |
| V | 17,0540 | 46,9000 | 8 | 44 |  |
| (S x V) | 8,7170 | 23,9700 | 8 | 4 |  |

Fonte: Construído pelo autor, 2022.

**Tabela 30**: Estatística de Teste, Valor aproximado a estatística F, Graus de liberdade e P-valor para a Análise de Variância Multivariada, para o teste Raiz máxima de Roy.

| Causa de Variação | Estatística de Teste | Valor F | GL1 | GL2 | P-Valor |
| --- | --- | --- | --- | --- | --- |
| S | 209,5940 | 1309,9600 | 4 | 25 |  |
| V | 16,7120 | 100,2700 | 4 | 24 |  |
| (S x V) | 8,5560 | 51,3400 | 4 | 24 |  |

Fonte: Construído pelo autor, 2022.

**Anexo I**: Script da primeira Questão

## ------

## Universidade Federal do Para

## Analise Multivariada II

## Analise de Variancia Multipla (MANOVA)

## Discente: Dionisio Alves da Silva Neto

## ------

## ---

## Pacotes para analise

## ---

if(!require(pacman)) install.packages("pacman"); library(pacman)

p\_load(dplyr, car, rstatix, ggplot2, MVN, GGally,

emmeans, psych, plotly, MASS, openxlsx, corrplot,gridExtra)

## ---

## Importancao do banco de dados I

## Primeira Questao

## ---

setwd("E:/")

dados = read.xlsx('bd\_lista2\_multivariada2.xlsx',

fillMergedCells = F, colNames = T,

sheet = 1, detectDates = F)

## Preprocessamento dos dados

## Organizando do formato wide para long

head(dados)

metodo1 = dados[,1:4]

metodo2 = dados[,5:8]

metodo3 = dados[,9:12]

colnames(metodo1)= metodo1[1,]

colnames(metodo2)= metodo2[1,]

colnames(metodo3)= metodo3[1,]

# retirando a linha 1 com o nome de variaveis

metodo1 = metodo1[-1,]

metodo2 = metodo2[-1,]

metodo3 = metodo3[-1,]

# criando a coluna de identificacao do metodo

metodo1['metodo'] = rep(1, dim(metodo1)[1])

metodo2['metodo'] = rep(2, dim(metodo2)[1])

metodo3['metodo'] = rep(3, dim(metodo3)[1])

dados = rbind(metodo1,metodo2,metodo3)

rownames(dados) = 1:dim(dados)[1]

#View(dados)

summary(dados)

## transformando as variaveis

dados$y1 = as.numeric(dados$y1)

dados$y2 = as.numeric(dados$y2)

dados$y3 = as.numeric(dados$y3)

dados$y4 = as.numeric(dados$y4)

dados$metodo = as.factor(dados$metodo)

##--- Banco de dados organizado ---##

## ---

## Analise Exploratoria dos dados

## letra (b)

## ---

summary(dados)

describe(dados)

## vetor de medias

colMeans(dados[,1:4])

## matriz de covariancias

cov(dados[,1:4])

## matriz de correlacoes

cor(dados[,1:4])

## Descritivas de cada grupo

## y1:

dados %>% group\_by(metodo) %>%

summarise(n = n(),

Min = min(eval(as.name('y1'))),

Q1 = quantile(eval(as.name('y1')))[[2]],

Median = median(eval(as.name('y1'))),

Mean = mean(eval(as.name('y1'))),

Q3 = quantile(eval(as.name('y1')))[[4]],

Max = max(eval(as.name('y1'))))

## laco para a descritiva de todas as variaveis

for(variavel in 1:4){

c = paste('y',variavel, sep = '')

print("------------------------------")

print(paste("---Estatistica descritiva de ", c,"---", sep = ""))

print("------------------------------")

tabela\_disc = dados %>% group\_by(metodo) %>%

summarise(n = n(),

Min = min(eval(as.name(c))),

Q1 = quantile(eval(as.name(c)))[[2]],

Median = median(eval(as.name(c))),

Mean = mean(eval(as.name(c))),

Q3 = quantile(eval(as.name(c)))[[4]],

Max = max(eval(as.name(c))))

print(tabela\_disc)

}

# vetor de medias, por metodo.

dados %>% group\_by(metodo) %>%

summarise(n = n(), mean\_y1 = mean(y1),

mean\_y2 = mean(y2), mean\_y3 = mean(y3),

mean\_y4 = mean(y4))

# matriz de variancia e covariancia

dados %>% group\_by(metodo) %>%

do(data.frame(cov = t(cov(.[,1:4]))))

## Matrizes de correlacao por ingestao

dados %>% group\_by((metodo)) %>%

do(data.frame(cov = t(cor(.[,1:4]))))

## ---

## Construcao de graficos univariados

## ---

## qqplot

setwd("E:/graficos\_lista2")

qq1 = ggplot(data = dados, aes(sample = dados[,'y1'])) +

stat\_qq() +

stat\_qq\_line(color = 'red') + theme\_minimal()

qq1 + scale\_x\_continuous("Te?rico") +

scale\_y\_continuous("Observado") +

theme(axis.text.x = element\_text(color = 'black', size = 12),

axis.text.y = element\_text(color = 'black', size = 12),

axis.title.x = element\_text(size = 14,

margin = margin(t = 15, l = 0, r = 0, b = 0)),

axis.title.y = element\_text(size = 14,

margin = margin(t = 0, l = 0, r = 15, b = 0)),

)

for(variavel in 1:4){

var = paste('y',variavel, sep = '')

qq = ggplot(data = dados, aes(sample = dados[,var])) +

stat\_qq() +

stat\_qq\_line(color = 'red') + theme\_minimal()

qqq = qq + scale\_x\_continuous("Te?rico") +

scale\_y\_continuous("Observado") +

theme(axis.text.x = element\_text(color = 'black', size = 12),

axis.text.y = element\_text(color = 'black', size = 12),

axis.title.x = element\_text(size = 14,

margin = margin(t = 15, l = 0, r = 0, b = 0)),

axis.title.y = element\_text(size = 14,

margin = margin(t = 0, l = 0, r = 15, b = 0)),

)

ggsave(paste('qqplot\_', var, '.jpeg',sep = ''),

width = 15, height = 10,units = "cm",dpi = 300)

}

## ---

## Boxplots

## ---

b2 = ggplot() + aes(dados,

x = dados$metodo,

y = dados[,'y2'],

fill = dados$metodo) +

geom\_boxplot(width = 0.5) +

scale\_fill\_discrete(labels = c("1", "2", "3"),

name = "M?todo") +

scale\_x\_discrete("M?todo", labels = c("1", "2", "3")) +

scale\_y\_continuous("Pontua??o do ju?zes") +

theme\_minimal()

b2 + theme(axis.text.x = element\_text(color="black",

size=22,

margin = margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=22,

margin = margin(t = 0, r = 0, b = 0, l = 15)),

axis.title=element\_text(size=26,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"),

legend.title=element\_text(size=22),

legend.text=element\_text(size=20))

ggsave("boxplot\_teste.jpeg", width = 20, height = 10, dpi = 800)

for (variavel in colnames(dados)[-5]){

b = ggplot() + aes(dados,

x = dados$metodo,

y = dados[,variavel],

fill = dados$metodo) +

geom\_boxplot(width = 0.5) +

scale\_fill\_discrete(labels = c("1", "2", "3"),

name = "M?todo") +

scale\_x\_discrete("M?todo", labels = c("1", "2", "3")) +

scale\_y\_continuous(paste("Pontua??o do ju?zes em ", variavel, sep ="")) +

theme\_minimal()

b = b + theme(axis.text.x = element\_text(color="black",

size=12,

margin = margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=12,

margin = margin(t = 0, r = 0, b = 0, l = 15)),

axis.title=element\_text(size=14,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"),

legend.title=element\_text(size=14),

legend.text=element\_text(size=14))

ggsave(plot = b, paste('boxplot\_', variavel, '.jpeg',sep = ''),

width = 15, height = 10,units = "cm",dpi = 800)

}

## ---

## Grafico de violino

## ---

v2 = ggplot() + aes(y = dados[,'y1'],

x = dados$metodo,

fill = dados$metodo) +

geom\_violin(alpha = 0.7) +

scale\_fill\_discrete(labels = c("1", "2", "3"),

name = "M?todo") +

scale\_y\_continuous("Pontua??o do ju?zes") +

scale\_x\_discrete("M?todo", labels = c("1", "2", "3")) +

theme\_minimal()

v2 + theme(axis.text.x = element\_text(color="black",

size=22,

margin = margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=22,

margin = margin(t = 0, r = , b = 0, l = 15)),

axis.title=element\_text(size=26,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"),

legend.title=element\_text(size=22),

legend.text=element\_text(size=20))

ggsave("violino\_teste.jpeg", width = 20, height = 10, dpi = 800)

for(variavel in colnames(dados)[-5]){

v = ggplot() + aes(y = dados[,variavel],

x = dados$metodo,

fill = dados$metodo) +

geom\_violin(alpha = 0.7) +

scale\_fill\_discrete(labels = c("1", "2", "3"),

name = "M?todo") +

scale\_y\_continuous(paste("Pontua??o do ju?zes em ", variavel)) +

scale\_x\_discrete("M?todo", labels = c("1", "2", "3")) +

theme\_minimal()

v = v + theme(axis.text.x = element\_text(color="black",

size=12,

margin = margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=12,

margin = margin(t = 0, r = , b = 0, l = 15)),

axis.title=element\_text(size=14,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"),

legend.title=element\_text(size=16),

legend.text=element\_text(size=14))

ggsave(plot = v, paste('violino\_', variavel, '.jpeg',sep = ''),

width = 15, height = 10,units = "cm",dpi = 800)

}

## ---

## Histograma

## ---

for(variavel in colnames(dados)[-5]){

h = dados %>%

ggplot(aes(x = dados[,variavel])) +

geom\_histogram(position = 'identity', bins = 6,

alpha = 0.7, fill="#69b3a2", color="#e9ecef") +

theme\_minimal() +

scale\_x\_continuous(paste("Pontu??o dos ju?zes em ", variavel)) +

scale\_y\_continuous("Frequ?ncia")

h = h + theme(axis.text.x = element\_text(color="black",

size=12,

margin(t = 0, r = 15, b = 0, l = 0)),

axis.text.y = element\_text(color="black",

size=12,

margin = margin(t = 0, r = , b = 0, l = 15)),

axis.title=element\_text(size=14,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"))

ggsave(plot = h, paste('histograma\_', variavel, '.jpeg',sep = ''),

width = 15, height = 10,units = "cm",dpi = 800)

}

## ---

## Grafico de densidade

## ---

d1 = dados %>% ggplot() + aes(x = dados[,'y1'], fill = metodo) +

geom\_density(alpha =0.7) +

scale\_x\_continuous("Pontua??o do ju?zes") +

scale\_fill\_discrete(labels = c("1", "2", "3"),

name = "M?todo") +

scale\_y\_continuous("Densidade") +

theme\_minimal()

d1 + theme(axis.text.x = element\_text(color="black",

size=22,

margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=22,

margin = margin(t = 0, r = , b = 0, l = 15)),

axis.title=element\_text(size=26,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"),

legend.title=element\_text(size=22),

legend.text=element\_text(size=20)

)

for(variavel in colnames(dados)[-5]){

d = dados %>% ggplot() + aes(x = dados[,variavel], fill = metodo) +

geom\_density(alpha =0.7) +

scale\_x\_continuous(paste("Pontua??o do ju?zes em ", variavel)) +

scale\_fill\_discrete(labels = c("1", "2", "3"),

name = "M?todo") +

scale\_y\_continuous("Densidade") +

theme\_minimal()

d + theme(axis.text.x = element\_text(color="black",

size=12,

margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=12,

margin = margin(t = 0, r = , b = 0, l = 15)),

axis.title=element\_text(size=14,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"),

legend.title=element\_text(size=14),

legend.text=element\_text(size=12)

)

print(d)

ggsave(plot = d, paste('densidade\_', variavel, '.jpeg',sep = ''),

width = 15, height = 10,units = "cm",dpi = 800)

}

## ---

## Construcao de graficos multivariados

## ---

## Dispersao de pares por metodo

## grafico de correlacao

jpeg("correlacao\_y.jpeg", units="in", width=5, height=5, res=300)

corrplot(cor(dados[,1:4]), method = 'color',type = 'lower',

cl.pos = 'b', addCoef.col = 'white', xlab ='hhh')

dev.off()

## Grafico de pares

jpeg("pares\_y.jpeg", units="in", width=5, height=5, res=300)

ggpairs(dados[,1:4],

upper = list(continuous = "density", combo = "box\_no\_facet"),

ggplot2::aes(color = factor(dados$metodo), alpha = 0.7)) +

theme\_minimal()

dev.off()

## dispersao em grupos

d1 = ggplot(dados, aes(y = y1, x = metodo))+

geom\_point(size = 2)+ scale\_x\_discrete("M?todo") +

theme\_minimal()

d2 = ggplot(dados, aes(y = y2, x = metodo))+

geom\_point(size = 2)+ scale\_x\_discrete("M?todo")+

theme\_minimal()

d3 = ggplot(dados, aes(y = y3, x = metodo))+

geom\_point(size = 2)+ scale\_x\_discrete("M?todo")+

theme\_minimal()

d4 = ggplot(dados, aes(y = y4, x = metodo))+

geom\_point(size = 2)+scale\_x\_discrete("M?todo") +

theme\_minimal()

jpeg("espalhamento\_y.jpeg", units="in", width=5, height=5, res=300)

grid.arrange(d1, d2, d3, d4)

dev.off()

## grafico 3d de densidade conjunta

den3d = kde2d(dados$y3, dados$y4)

fig = plot\_ly(x = den3d$x, y = den3d$y, z = den3d$z,

type = 'surface')

fig %>% add\_surface() %>%

layout(scene = list(xaxis = list(title = "y3"),

yaxis = list(title = "y4"),

zaxis = list(title= "Densidade")))

## ---

## Testes de normalidade multivariada entre os pares das variaveis

## Sera utilizadoo teste de Dornik-Hansen

## ---

## y1 & y2

mvn(cbind(dados$y1, dados$y2), mvnTest = 'dh')

## y1 & y3

mvn(cbind(dados$y1, dados$y3), mvnTest = 'dh')

## y1 & y4

mvn(cbind(dados$y1, dados$y4), mvnTest = 'dh')

## y2 & y3

mvn(cbind(dados$y2, dados$y3), mvnTest = 'dh')

## y2 & y4

mvn(cbind(dados$y2, dados$y4), mvnTest = 'dh')

## y3 & y4

mvn(cbind(dados$y3, dados$y4), mvnTest = 'dh')

mvn(dados[,-5], mvnTest = 'dh')

## ---

## Realizacao da MANOVA

## ---

## ---

## Passo 1: Verificacao dos pressupostos da MANOVA

## ---

## ---

## 1.1 Normalidade dos dados (Univariada e Multivariada)

## ---

# O output retorna tambem os testes de Anderson-Darling para cada variavel

mvn(dados[,-5], mvnTest = 'dh')

# Foi comprovada a distribuicao multivariada dos dados

# Apenas a variavel y4 nao atendeu os a distribuico normal univariada

## ---

## 1.2 Identificar outliers

## Sera utilizada a distancia de Mahalanobis

## ---

outliers\_multi = dados %>% dplyr::select(1:4) %>% group\_by(dados$metodo) %>%

doo(~mahalanobis\_distance(.))

## Verificar se tem outliers

outliers\_multi %>% filter(is.outlier == TRUE)

## sem outliers multivariados segunda distancia de Mahablnobis

## Na descritiva, temos um outlier em y1 e dois em y4

## Identificacao de outliers por grupo, fazer o boxplot.

## ---

## 1.3 Teste de homogeneidade da matriz de variancias e covariancias

## ---

# Se nao rompido e n iguais por grupo: Pillai e Hotelling sao confiavaeis

# Caso os n sejam diferentes, uma opcao e usar uma manova robusta

# Alfa: 0,01

# H0: Existe homogeneidade da matriz de variancias e covariancias

# H1: Nao Existe homogeneidade da matriz de variancias e covariancias

box\_m(dados[,1:4], dados$metodo)

## ------

## 1.4 Homogenenidade de Variancias (matriz de variancias e covariancias)

## ------

# Se nao rompido e n iguais por grupo: Pillai e Hotelling sao confiavaeis

# Caso os n sejam diferentes, uma opcao e usar uma manova robusta

# Alfa: 0,01

# H0: Existe homogeneidade das variancias

# H1: Nao Existe homogeneidade de variancias

box\_m(dados[, c("Memoria", "Latencia")], dados$Alcool)

# Como o p-valor resultou em 0,8570, nao rejeitamos h0. Logo,

# existe homogeneidade na matriz de variancias e covariancias

# Veficacao da homogeneidade de variancias - teste de Levene no pacote car

# ANOVA

# alpha = 0.01

## H0: As variancias sao homogeneas

## H1: As varincias nao sao homogeneas

leveneTest(y1 ~ metodo, dados, center = mean) # sao homogeneas

leveneTest(y2 ~ metodo, dados, center = mean) # sao homogeneas

leveneTest(y3 ~ metodo, dados, center = mean) # sao homogeneas

leveneTest(y4 ~ metodo, dados, center = mean) # sao homogeneas

## ------

## 1.4 Verificacao da ausencia de multicolinearidade

## ------

# baixa correlacao indica ausencia de multicolinearidade

ggpairs(dados[,1:4]) + theme\_minimal()+

theme(axis.title.x.top = element\_text(size = 22),

axis.title.y.right = element\_text(size = 22))

ggsave("correlacao.jpeg", width = 20, height = 20, dpi = 800)

cor(dados[,1:4])

# tem correlacao moderad em y1 e y2, e em y3 e y4

## ------

## 1.5 Verificacao da relacao linear das variaveis dependentes para cada grupo

## ------

pairs(dados[,1:4], pch = 19,

col = c('darkturquoise', "maroon", "springgreen3")[factor(dados$metodo)])

pairs.panels(dados[,1:4], ellipses = F,

lm = T, stars = T, cor = T, pch = 16)

## graficos de dipersao separados por grupos

graph\_ln = dados%>% group\_by(metodo) %>%

doo(~ggpairs(.))

graph\_ln$.results.[[1]] + theme\_minimal()

graph\_ln$.results.[[2]] + theme\_minimal()

graph\_ln$.results.[[3]] + theme\_minimal()

## ------

## MANOVA

## ------

# construcao do modelo

modelo = manova(cbind(y1,y2,y3,y4) ~ metodo, # variaveis dependentes ~ variaveis independnetes

data = dados)

# analise dos resultados

summary(modelo)

summary(modelo, test = "Wilks")

summary(modelo, test = "Pillai") # mais robusto

summary(modelo, test = "Roy")

summary(modelo, test = "Hotelling-Lawley")

# ANOVA Univariada para cada vari?vel

summary.aov(modelo)

## ------

## Analisando a diferenca entre grupos pos-hoc

## ------

# Pela medias marginais estimadas

dados %>% emmeans\_test(y1 ~ metodo, p.adjust.method = "bonferroni")

dados %>% emmeans\_test(y2 ~ metodo, p.adjust.method = "bonferroni")

dados %>% emmeans\_test(y3 ~ metodo, p.adjust.method = "bonferroni")

dados %>% emmeans\_test(y4 ~ metodo, p.adjust.method = "bonferroni")

# Nao existe diferenca entre os grupos duas caneas e uma caneca

# Existe diferenca para os grupos 2 canecas e 4 canecas, assim como para os

# grupos 4 anecas e nenhum

# Nao existe diferenca entre os grupos 2 canecas e nenhuma caneca

# Existe diferenca entre os grupos 2 e 4 canecas, tal como os grupos

# 4 canecas e nenhum

## Outra op??o de Post-Hoc

TukeyHSD(x = aov(y1 ~ metodo, data = dados), "metodo", conf.level = 0.95)

TukeyHSD(x = aov(y2 ~ metodo, data = dados), "metodo", conf.level = 0.95)

TukeyHSD(x = aov(y3 ~ metodo, data = dados), "metodo", conf.level = 0.95)

TukeyHSD(x = aov(y4 ~ metodo, data = dados), "metodo", conf.level = 0.95)

plot(TukeyHSD(x = aov(y1 ~ metodo, data = dados)))

title(xlab = 'DiferenÃ§a de mÃ©dias dos mÃ©todos', ylab = "Par", line = 2)

plot(TukeyHSD(x = aov(y2 ~ metodo, data = dados)))

title(xlab = 'DiferenÃ§a de mÃ©dias dos mÃ©todos', ylab = "Par", line = 2)

plot(TukeyHSD(x = aov(y3 ~ metodo, data = dados)))

title(xlab = 'DiferenÃ§a de mÃ©dias dos mÃ©todos', ylab = "Par", line = 2)

plot(TukeyHSD(x = aov(y4 ~ metodo, data = dados)))

title(xlab = 'DiferenÃ§a de mÃ©dias dos mÃ©todos', ylab = "Par", line = 2)

**Anexo II**: Script da segunda Questão

## ------

## Universidade Federal do Para

## Analise Multivariada II

## Analise de Variancia Multipla (MANOVA) parte 2

## Discente: Dionisio Alves da Silva Neto

## ------

## ---

## Pacotes para analise

## ---

if(!require(pacman)) install.packages("pacman"); library(pacman)

p\_load(dplyr, car, rstatix, ggplot2, MVN, GGally,

emmeans, psych, plotly, MASS, openxlsx, corrplot)

## ---

## Importancao do banco de dados I

## Primeira Questao

## ---

setwd("D:/")

dados2 = read.csv('BD2\_multi2\_part2.csv', header = T, dec = ',', sep = ';')

## Preprocessamento dos dados

## Organizando do formato wide para long

head(dados2)

parte1 = dados2[,1:7]

parte2 = dados2[,8:dim(dados2)[2]]

colnames(parte2) = colnames(parte1)

## Juncao para formar o banco de dados final e limpo

dados2 = rbind(parte1, parte2)

dados2$S = as.factor(dados2$S)

dados2$V = as.factor(dados2$V)

dados2$y1 = as.numeric(unlist(dados2$y1))

dados2$y2 = as.numeric(unlist(dados2$y2))

dados2$y3 = as.numeric(unlist(dados2$y3))

dados2$y4 = as.numeric(unlist(dados2$y4))

## ---

## Analise Exploratoria dos dados

## letra (b)

## ---

summary(dados2)

describe(dados2)

quantitativas = dados2[,4:7]

## vetor de medias

colMeans(quantitativas)

## matriz de covariancias

cov(quantitativas)

## matriz de correlacoes

cor(quantitativas)

## Descritivas de cada grupo

## y1:

## laco para a descritiva de todas as variaveis

setwd("E:/graficos\_lista2/questao2")

for (fator in c('S', 'V')){

print("------------------------------")

print(paste("---Analisando ", fator,"---", sep = ""))

print("------------------------------")

for(variavel in 1:4){

c = paste('y',variavel, sep = '')

print("------------------------------")

print(paste("---Estatistica descritiva de ", c,"---", sep = ""))

print("------------------------------")

tabela\_disc = dados2 %>% group\_by(dados2[,fator]) %>%

summarise(n = n(),

Min = min(eval(as.name(c))),

Q1 = quantile(eval(as.name(c)))[[2]],

Median = median(eval(as.name(c))),

Mean = mean(eval(as.name(c))),

Q3 = quantile(eval(as.name(c)))[[4]],

Max = max(eval(as.name(c))))

print(tabela\_disc)

}

}

#-----------------------------

# vetor de medias, por S.

dados2 %>% group\_by(S) %>%

summarise(n = n(), mean\_y1 = mean(y1),

mean\_y2 = mean(y2), mean\_y3 = mean(y3),

mean\_y4 = mean(y4))

# vetor de medias, por V.

dados2 %>% group\_by(V) %>%

summarise(n = n(), mean\_y1 = mean(y1),

mean\_y2 = mean(y2), mean\_y3 = mean(y3),

mean\_y4 = mean(y4))

#-----------------------------

#-----------------------------

# matriz de variancia e covariancia

dados2 %>% group\_by(S) %>%

do(data.frame(cov = t(cov(.[4:7]))))

dados2 %>% group\_by(V) %>%

do(data.frame(cov = t(cov(.[4:7]))))

#-----------------------------

## Matrizes de correlacao por ingestao

dados2 %>% group\_by((S)) %>%

do(data.frame(cov = t(cor(.[4:7]))))

dados2 %>% group\_by((V)) %>%

do(data.frame(cov = t(cor(.[4:7]))))

#-----------------------------

## ---

## Construcao de graficos univariados

## ---

## qqplot

qq1 = ggplot(data = dados2, aes(sample = dados2[,'y1'])) +

stat\_qq() +

stat\_qq\_line(color = 'red') + theme\_minimal()

qq1 + scale\_x\_continuous("Teórico") +

scale\_y\_continuous("Observado") +

theme(axis.text.x = element\_text(color = 'black', size = 12),

axis.text.y = element\_text(color = 'black', size = 12),

axis.title.x = element\_text(size = 14,

margin = margin(t = 15, l = 0, r = 0, b = 0)),

axis.title.y = element\_text(size = 14,

margin = margin(t = 0, l = 0, r = 15, b = 0)),

)

qq2 = ggplot(data = dados2, aes(sample = dados2[,'y2'])) +

stat\_qq() +

stat\_qq\_line(color = 'red') + theme\_minimal()

qq2 + scale\_x\_continuous("Teórico") +

scale\_y\_continuous("Observado") +

theme(axis.text.x = element\_text(color = 'black', size = 12),

axis.text.y = element\_text(color = 'black', size = 12),

axis.title.x = element\_text(size = 14,

margin = margin(t = 15, l = 0, r = 0, b = 0)),

axis.title.y = element\_text(size = 14,

margin = margin(t = 0, l = 0, r = 15, b = 0)),

)

qq3 = ggplot(data = dados2, aes(sample = dados2[,'y3'])) +

stat\_qq() +

stat\_qq\_line(color = 'red') + theme\_minimal()

qq3 + scale\_x\_continuous("Teórico") +

scale\_y\_continuous("Observado") +

theme(axis.text.x = element\_text(color = 'black', size = 12),

axis.text.y = element\_text(color = 'black', size = 12),

axis.title.x = element\_text(size = 14,

margin = margin(t = 15, l = 0, r = 0, b = 0)),

axis.title.y = element\_text(size = 14,

margin = margin(t = 0, l = 0, r = 15, b = 0)),

)

qq4 = ggplot(data = dados2, aes(sample = dados2[,'y4'])) +

stat\_qq() +

stat\_qq\_line(color = 'red') + theme\_minimal()

qq4 + scale\_x\_continuous("Teórico") +

scale\_y\_continuous("Observado") +

theme(axis.text.x = element\_text(color = 'black', size = 12),

axis.text.y = element\_text(color = 'black', size = 12),

axis.title.x = element\_text(size = 14,

margin = margin(t = 15, l = 0, r = 0, b = 0)),

axis.title.y = element\_text(size = 14,

margin = margin(t = 0, l = 0, r = 15, b = 0)),

)

#ggsave(paste('qqplot\_', variavel, '.jpeg',sep = ''),

#width = 15, height = 10,units = "cm",dpi = 300)

## ---

## Boxplots

## ---

b2 = ggplot() + aes(dados2,

x = dados2$S,

y = dados2[,'y2'],

fill = dados2$S) +

geom\_boxplot(width = 0.5) +

scale\_fill\_discrete(labels = unique(dados2$S),

name = "S") +

scale\_x\_discrete("S", labels = unique(dados2$S)) +

scale\_y\_continuous("y2") +

theme\_minimal()

b2 + theme(axis.text.x = element\_text(color="black",

size=22,

margin = margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=22,

margin = margin(t = 0, r = 0, b = 0, l = 15)),

axis.title=element\_text(size=26,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"),

legend.title=element\_text(size=22),

legend.text=element\_text(size=20))

ggsave("boxplot\_teste.jpeg", width = 20, height = 10, dpi = 800)

b2 = ggplot() + aes(dados2,

x = dados2$V,

y = dados2[,'y2'],

fill = dados2$V) +

geom\_boxplot(width = 0.5) +

scale\_fill\_discrete(labels = unique(dados2$V),

name = "Método") +

scale\_x\_discrete("Método", labels = unique(dados2$V)) +

scale\_y\_continuous("y2") +

theme\_minimal()

b2 + theme(axis.text.x = element\_text(color="black",

size=22,

margin = margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=22,

margin = margin(t = 0, r = 0, b = 0, l = 15)),

axis.title=element\_text(size=26,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"),

legend.title=element\_text(size=22),

legend.text=element\_text(size=20))

for (variavel in colnames(dados2)[4:7]){

b = ggplot() + aes(dados2,

x = dados2$S,

y = dados2[,variavel],

fill = dados2$S) +

geom\_boxplot(width = 0.5) +

scale\_fill\_discrete(labels = unique(dados2$S),

name = "S") +

scale\_x\_discrete("S", labels = c("1", "2", "3", "4")) +

scale\_y\_continuous(paste("Medidas de feijão-vagem em ", variavel, sep ="")) +

theme\_minimal()

b = b + theme(axis.text.x = element\_text(color="black",

size=12,

margin = margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=12,

margin = margin(t = 0, r = 0, b = 0, l = 15)),

axis.title=element\_text(size=14,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"),

legend.title=element\_text(size=12),

legend.text=element\_text(size=12))

print(b)

ggsave(paste('boxplot\_S\_', variavel, '.jpeg',sep = ''),

width = 15, height = 10,units = "cm",dpi = 300)

}

for (variavel in colnames(dados2)[4:7]){

b = ggplot() + aes(dados2,

x = dados2$V,

y = dados2[,variavel],

fill = dados2$V) +

geom\_boxplot(width = 0.5) +

scale\_fill\_discrete(labels = unique(dados2$V),

name = "V") +

scale\_x\_discrete("V", labels = unique(dados2$V)) +

scale\_y\_continuous(paste("Medidas de feijão-vagem em ", variavel, sep ="")) +

theme\_minimal()

b = b + theme(axis.text.x = element\_text(color="black",

size=12,

margin = margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=12,

margin = margin(t = 0, r = 0, b = 0, l = 15)),

axis.title=element\_text(size=14,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"),

legend.title=element\_text(size=12),

legend.text=element\_text(size=12))

print(b)

ggsave(paste('boxplot\_V\_', variavel, '.jpeg',sep = ''),

width = 15, height = 10,units = "cm",dpi = 300)

}

## ---

## Grafico de violino

## ---

v2 = ggplot() + aes(y = dados[,'y1'],

x = dados$metodo,

fill = dados$metodo) +

geom\_violin(alpha = 0.7) +

scale\_fill\_discrete(labels = c("1", "2", "3"),

name = "Método") +

scale\_y\_continuous("Medidas de feijão-vagem em ") +

scale\_x\_discrete("S", labels = c("1", "2", "3")) +

theme\_minimal()

v2 + theme(axis.text.x = element\_text(color="black",

size=22,

margin = margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=22,

margin = margin(t = 0, r = , b = 0, l = 15)),

axis.title=element\_text(size=26,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"),

legend.title=element\_text(size=22),

legend.text=element\_text(size=20))

ggsave("violino\_teste.jpeg", width = 20, height = 10, dpi = 800)

for(variavel in colnames(dados2)[4:7]){

v = ggplot() + aes(y = dados2[,variavel],

x = dados2$S,

fill = dados$S) +

geom\_violin(alpha = 0.7) +

scale\_fill\_discrete(labels = unique(dados$S),

name = "Método") +

scale\_y\_continuous(paste("Medidas de feijão-vagem em ", variavel, sep ="")) +

scale\_x\_discrete("Método", labels = c("1", "2", "3")) +

theme\_minimal()

v = v + theme(axis.text.x = element\_text(color="black",

size=5,

margin = margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=5,

margin = margin(t = 0, r = , b = 0, l = 15)),

axis.title=element\_text(size=16,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"),

legend.title=element\_text(size=14),

legend.text=element\_text(size=14))

ggsave(paste('violino\_', variavel, '.jpeg',sep = ''),

width = 15, height = 10,units = "cm",dpi = 300)

}

## ---

## Histograma

## ---

for(variavel in colnames(dados2)[4:7]){

h = dados %>%

ggplot(aes(x = dados[,variavel])) +

geom\_histogram(position = 'identity', bins = 6,

alpha = 0.7, fill="#69b3a2", color="#e9ecef") +

theme\_minimal() +

scale\_x\_continuous(paste("Medidas do feijão-vagem em", variavel)) +

scale\_y\_continuous("Frequência")

h = h + theme(axis.text.x = element\_text(color="black",

size=14,

margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=14,

margin = margin(t = 0, r = , b = 0, l = 15)),

axis.title=element\_text(size=14,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"))

ggsave(paste('histograma\_', variavel, '.jpeg',sep = ''),

width = 15, height = 10,units = "cm",dpi = 300)

}

## ---

## Grafico de densidade

## ---

d1 = dados %>% ggplot() + aes(x = dados[,'y1'], fill = metodo) +

geom\_density(alpha =0.7) +

scale\_x\_continuous("Pontuação do juízes") +

scale\_fill\_discrete(labels = c("1", "2", "3"),

name = "Método") +

scale\_y\_continuous("Densidade") +

theme\_minimal()

d1 + theme(axis.text.x = element\_text(color="black",

size=22,

margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=22,

margin = margin(t = 0, r = , b = 0, l = 15)),

axis.title=element\_text(size=26,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"),

legend.title=element\_text(size=22),

legend.text=element\_text(size=20)

)

for(variavel in colnames(dados)[-5]){

d = dados %>% ggplot() + aes(x = dados[,variavel], fill = metodo) +

geom\_density(alpha =0.7) +

scale\_x\_continuous(paste("Pontuação do juízes em ", variavel)) +

scale\_fill\_discrete(labels = c("1", "2", "3"),

name = "Método") +

scale\_y\_continuous("Densidade") +

theme\_minimal()

d + theme(axis.text.x = element\_text(color="black",

size=22,

margin(t = 0, r = 0, b = 15, l = 0)),

axis.text.y = element\_text(color="black",

size=22,

margin = margin(t = 0, r = , b = 0, l = 15)),

axis.title=element\_text(size=26,face="bold"),

panel.grid.major = element\_line(linetype = "dashed"),

panel.grid.minor = element\_line(linetype = "dashed"),

legend.title=element\_text(size=22),

legend.text=element\_text(size=20)

)

ggsave(paste('densidade\_', variavel, '.jpeg',sep = ''),

width = 15, height = 10,units = "cm",dpi = 300)

}

## grafico de correlacao

jpeg("correlacao\_y.jpeg", units="in", width=5, height=5, res=300)

corrplot(cor(dados2[,4:7]), method = 'color',type = 'lower',

cl.pos = 'b', addCoef.col = 'white', xlab ='hhh')

dev.off()

## Grafico de pares

jpeg("pares\_y.jpeg", units="in", width=5, height=5, res=300)

ggpairs(dados2[,4:7],

upper = list(continuous = "density", combo = "box\_no\_facet"),

ggplot2::aes(color = factor(dados2$V), alpha = 0.7)) +

theme\_minimal()

dev.off()

jpeg("pares\_y\_S.jpeg", units="in", width=5, height=5, res=300)

ggpairs(dados2[,4:7],

upper = list(continuous = "density", combo = "box\_no\_facet"),

ggplot2::aes(color = factor(dados2$S), alpha = 0.7)) +

theme\_minimal()

dev.off()

## ------

## Verificacao dos pressupostos

## ------

## ---

## 1.1 Normalidade multivariada dos dados

## ---

mvn(quantitativas, mvnTest = 'royston')

## ---

## 1.2 Identificacao de outliers multivariados

## Sera utilizada a distancia de Mahalanobis

## ---

outliers\_multi = dados2 %>% dplyr::select(4:7) %>% group\_by(dados$S) %>%

doo(~mahalanobis\_distance(.))

outliers\_multi

outliers\_multi2 = dados2 %>% dplyr::select(4:7) %>% group\_by(dados$V) %>%

doo(~mahalanobis\_distance(.))

outliers\_multi

outliers\_multi2

## Verificar se tem outliers

outliers\_multi %>% filter(is.outlier == TRUE)

outliers\_multi2 %>% filter(is.outlier == TRUE)

## ---

## 1.3 Teste de homogeneidade da matriz de variâncias e covariâncias

## ---

# H0: Existe homogeneidade da matriz de variancias e covariancias

# H1: Nao Existe homogeneidade da matriz de variancias e covariancias

box\_m(dados2[,4:7], dados2$S)

box\_m(dados2[,4:7], dados2$V)

## ---

## 1.4 Homogeneidade de variancias (ANOVAS)

## ---

# Veficacao da homogeneidade de variancias - teste de Levene no pacote car

# ANOVA

## H0: As variancias sao homogeneas

## H1: As varincias nao sao homogeneas

leveneTest(y1 ~ S, dados2, center = mean) # nao sao homogeneas

leveneTest(y2 ~ S, dados2, center = mean) # nao sao homogeneas

leveneTest(y3 ~ S, dados2, center = mean) # nao sao homogeneas

leveneTest(y4 ~ S, dados2, center = mean) # sao homogeneas

leveneTest(y1 ~ V, dados2, center = mean) # nao sao homogeneas

leveneTest(y2 ~ V, dados2, center = mean) # sao homogeneas

leveneTest(y3 ~ V, dados2, center = mean) # sao homogeneas

leveneTest(y4 ~ V, dados2, center = mean) # sao homogeneas

## ---

## 1.5 Verificacao da ausencia de multicolineariadade

## ---

#baixa correlacao indica ausencia de multicolinearidade

jpeg("multicolinearidade.jpeg", units="in", width=5, height=5, res=300)

ggpairs(dados2[,4:7]) + theme\_minimal()+

theme(axis.title.x.top = element\_text(size = 22),

axis.title.y.right = element\_text(size = 22))

dev.off()

cor(dados2[,4:7])

## ---

## 1.6 Verificacao da ausencia da relacao linear das variaveis dependentes

## para cada grupo

## ---

pairs(dados2[,4:7], pch = 19,

col = c('darkturquoise', "maroon", "springgreen3")[factor(dados2$V)])

jpeg("relacao\_linear.jpeg", units="in", width=5, height=5, res=300)

pairs.panels(dados2[,4:7], ellipses = F,

lm = T, stars = T, cor = T, pch = 16)

dev.off()

## graficos de dipersao separados por grupos

graph\_ln = dados2[,4:7] %>% group\_by(dados$V) %>%

doo(~ggpairs(.))

graph\_ln$.results.[[1]] + theme\_minimal()

graph\_ln$.results.[[2]] + theme\_minimal()

graph\_ln$.results.[[3]] + theme\_minimal()

## ------

## Realizacao da MANOVA a dois fatores

## ------

modelo <- manova(cbind(Latencia, Memoria) ~ Alcool, data=dados)

md1 = manova(cbind(y1, y2, y3, y4)~(S+V+(S\*V)), data = dados2)

summary(md1, test="Wilks")

summary(md1, test="Pillai")

summary(md1, test="Hotelling-Lawley")

summary(md1, test="Roy")

**Anexo III**: Resultados do teste de Doornik-Hansen da primeira Questão

$multivariateNormality

Test E df p value MVN

1 Doornik-Hansen 3.942508 8 0.862273 YES

$univariateNormality

Test Variable Statistic p value Normality

1 Anderson-Darling y1 0.3568 0.4374 YES

2 Anderson-Darling y2 0.4691 0.2340 YES

3 Anderson-Darling y3 0.4409 0.2747 YES

4 Anderson-Darling y4 0.7681 0.0416 NO

$Descriptives

n Mean Std.Dev Median Min Max 25th 75th Skew Kurtosis

y1 36 5.205556 0.6427372 5.25 3.9 6.2 4.800 5.700 -0.3338277 -0.8648950

y2 36 5.258333 0.6025541 5.20 4.0 6.2 4.975 5.750 -0.2630816 -0.7023753

y3 36 5.552778 0.6322233 5.70 4.0 6.8 5.175 6.000 -0.4945770 -0.1497976

y4 36 6.030556 0.5701225 6.00 4.8 7.0 5.775 6.425 -0.5154752 -0.5165972