

Petal.Width	5.6450	4.8084	6.2718
	Petal.Width		
Sepal.Length	5.6450		
Sepal.Width	4.8084		
Petal.Length	6.2718		
Petal.Width	6.1566		

Multivariate Test:

	Df	test	stat	approx	F	num	Df	den	Df	Pr(>F)
Pillai	1	0.7452457	105.3127			4	144	<	2.22e-16	***

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05  
 '.' 0.1 ' ' 1

Concluimos então que tanto  $C_1$  quanto  $C_2$  são significativos ( $p < 0.001$ ). Logo, o vetor de médias da espécie 'setosa' difere do vetor de médias do grupo 'versicolor, virginica'; e os vetores de médias destas duas últimas também diferem entre si.

## Desafio

Com base no último modelo ajustado (`fit2`), mostre que:

$$\mathbf{B}_{Species} = \mathbf{B}_{C1} + \mathbf{B}_{C2}$$

Para isso, crie um objeto que receberá um valor lógico. *Dica:* use a função `all.equal()` ou `all()`.

## 8.5. Variável discriminante canônica

Dada uma MANOVA para  $p$  variáveis resposta,  $Y_1, Y_2, \dots, Y_p$ , uma função discriminante canônica ou simplesmente variável canônica (VC) pode ser determinada por combinação linear das variáveis originais, na forma:

$$VC = a_1 Y_1 + a_2 Y_2 + \dots + a_p Y_p$$

Sendo VC aplicada a cada uma das observações multivariadas do experimento, ela nos fornece informações sobre todas as variáveis resposta simultaneamente, de forma univariada. Para tal, VC deve ser obtida de forma que o teste  $F$  para tratamentos seja o maior possível.

A determinação dos coeficientes 'a' é, novamente, um problema de autovalores, semelhante àquele visto na determinação de componentes principais, com a seguinte diferença: autovalores e autovetores não são obtidos a partir de uma matriz de correlação ou covariância, mas sim a partir da matriz  $\mathbf{W}^{-1}\mathbf{B}$ , obtida da MANOVA. Então, tem-se que a solução de

$$(\mathbf{W}^{-1}\mathbf{B} - \lambda\mathbf{I})\mathbf{a} = \mathbf{0}$$

produz  $VC_1, VC_2, \dots, VC_k$ , em que  $k = \min(p, I - 1)$ . Entretanto, é sabido que  $VC_1$  é a que possui o maior poder de discriminação dos tratamentos<sup>2</sup>. Assim, se essa variável canônica reter uma parte considerável da variação de tratamentos, uma ANOVA pode ser realizada com os seus escores e, conseqüentemente, testes de comparações múltiplas, tais como o teste de Tukey, podem também ser utilizados para classificação multivariada dos tratamentos.

No R, usando o modelo ajustado anteriormente (`fit`), podemos fazer:

```
R> B <- summary(fit)$SS$Species
R> W <- summary(fit)$SS$Residuals
R> eig <- eigen(solve(W, B))
```

<sup>2</sup>A proporção da variação contida numa variável canônica pode ser calculada de forma idêntica àquela vista em componentes principais, isto é:  $\lambda_j / \sum_j \lambda_j$ .

```
# Calculando VC escalonadas
R> vc <- scale(as.matrix(iris[, -5]) %*%
+             eig$vectors[, 1:2],
+             center = TRUE, scale = FALSE)

# Anova com VC1
R> can1.aov <- aov(vc[, 1] ~ Species, data = iris)
R> anova(can1.aov)
Analysis of Variance Table

Response: vc[, 1]
          Df Sum Sq Mean Sq F value    Pr(>F)
Species    2  299.763  149.882   2366.1 < 2.2e-16 ***
Residuals 147    9.312    0.063
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
                 '.' 0.1 ' ' 1

# Teste Tukey com VC1
R> TukeyHSD(can1.aov, "Species")
Tukey multiple comparisons of means
95% family-wise confidence level
```

```
Fit: aov(formula = vc[, 1] ~ Species, data = iris)
```

```
$Species
```

	diff	lwr	upr	p adj
versicolor-setosa	2.374055	2.254872	2.49323	0
virginica-setosa	3.370098	3.250916	3.48928	0
virginica-versicolor	0.996043	0.876860	1.11522	0

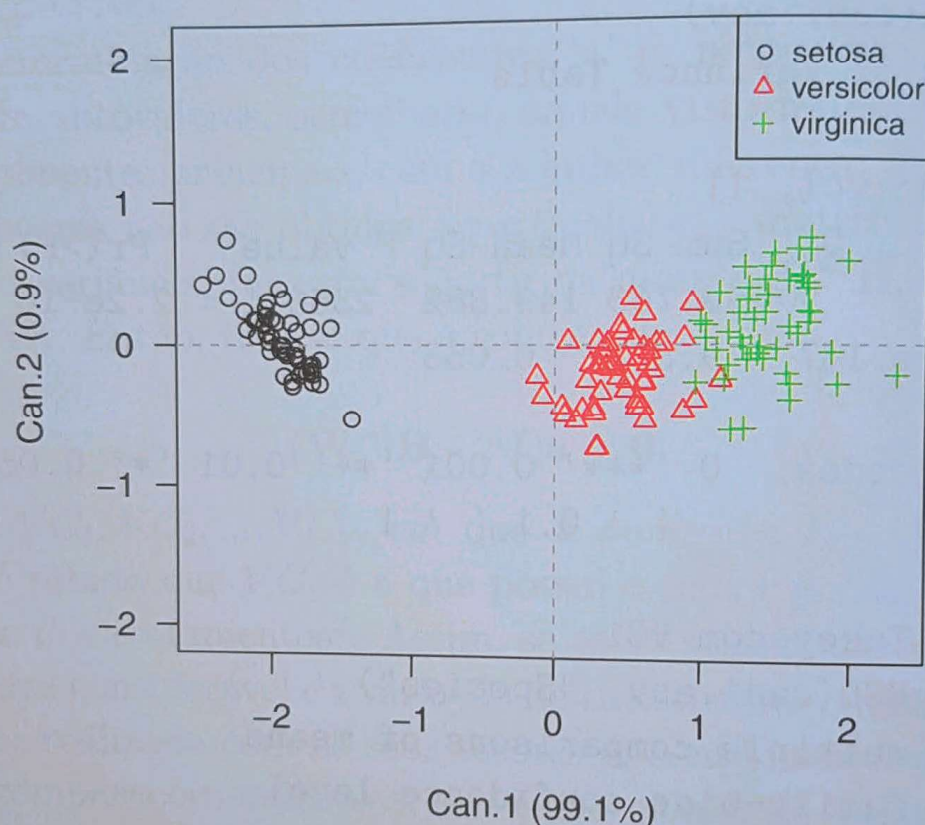
Observa-se que o valor- $p$  obtido com a ANOVA é bastante semelhante àquele obtido via MANOVA. Isso porque a  $VC_1$  foi responsável por quase toda a variabilidade contida nos tratamen-



tos.

Pelo teste de Tukey, observa-se que todos os vetores de médias diferem entre si ( $p < 0,001$ ).

A partir dos escores das variáveis canônicas (objeto `vc`), o seguinte gráfico de dispersão (Figura 8.1) pode ser útil para verificação das diferenças entre espécies.



**Figura 8.1.** Dispersão dos escores das duas primeiras variáveis discriminantes canônicas obtidas via MANOVA dos dados 'iris'.

Observe que, como escalonamos as variáveis canônicas para terem média zero, então podemos descrever as características de cada observação de acordo com a sua localização no plano bidimensional. Para isso, precisamos analisar as cargas (*loadings*) das variáveis originais em cada variável canônica. As cargas correspondem às correlações das observações de cada variável com os escores das variáveis canônicas.

```
R> cargas <- cor(cbind(vc, iris[, -5]))[1:2, -c(1:2)]
```

```
R> cargas
  Sepal.Length Sepal.Width Petal.Length Petal.Width
1    0.7918878   -0.5307590    0.98495127   0.9728120
2    0.2175931    0.7579893    0.04603709   0.2229024
```

Com isso, percebemos que  $VC_1$  reflete um contraste entre as variáveis *Sepal.Length*, *Petal.Length*, *Petal.Width* e *Sepal.Width*. Logo, como as observações da espécie setosa estão à esquerda do eixo Can.1, dizemos que elas apresentam os maiores valores para *Sepal.Width* e os menores para as demais variáveis. A espécie virginica tem comportamento oposto, assim como a versicolor, embora esta apresente valores intermediários.

## Nota

Alguns pacotes computacionais aceitam apenas matrizes simétricas nos algoritmos de decomposição espectral. Contudo, pode ser mostrado que os autovalores da matriz não simétrica  $\mathbf{W}^{-1}\mathbf{B}$  são os mesmos da matriz simétrica  $(\mathbf{U}^{-1})^T\mathbf{B}\mathbf{U}^{-1}$ , sendo  $\mathbf{W} = \mathbf{U}^T\mathbf{U}$  a decomposição de Choleski.<sup>3</sup> Entretanto, os autovetores não são os mesmos, precisando ser ajustados da seguinte forma: seja  $\mathbf{b}$  um autovetor de  $(\mathbf{U}^{-1})^T\mathbf{B}\mathbf{U}^{-1}$ , então  $\mathbf{a} = \mathbf{U}^{-1}\mathbf{b}$  é um autovetor de  $\mathbf{W}^{-1}\mathbf{B}$ . Usaremos esta abordagem no próximo capítulo, para calcular a matriz de covariâncias das variáveis canônicas.

## Desafio

A partir da Figura 8.1, construa um *biplot* adicionando ao gráfico as cargas das variáveis em cada eixo canônico. Faça ajustes de escala, se necessário. *Dica*: use a função `arrows()`.

<sup>3</sup>A decomposição recebe o nome do seu criador, André-Louis Choleski. A matriz triangular superior  $\mathbf{U}$  é conhecida como o triângulo de Choleski. Uma variante dessa decomposição é a forma  $\mathbf{W} = \mathbf{L}\mathbf{L}^T$ , em que  $\mathbf{L}$  é triangular inferior. Para mais detalhes, consulte Golub & Loan (1996).



## 8.6. Exercício

Carregue os dados com o nome 'Wolves', disponíveis no pacote **candisc**. Trata-se de dados de onze variáveis morfométricas de crânios de lobos machos e fêmeas de dois locais. Considere o fator **group** como sendo a combinação de sexo e local, e então:

1. realize a MANOVA;
2. determine variáveis discriminantes canônicas para **group** e calcule a proporção da variância retida em cada uma;
3. caso o autovalor da VC1 represente mais de 60% do total, aplique o teste de Tukey ( $\alpha = 0.05$ ) para comparar, indiretamente, vetores de médias de **group** a partir dos escores de VC1;
4. realize outra MANOVA, agora considerando não o fator **group**, mas os fatores **location** e **sex**, bem como a interação entre eles; peça o teste do Lambda de Wilks.

## 8.7. Outras implementações

O pacote **candisc** (Friendly & Fox, 2013) apresenta uma implementação para determinação de variáveis discriminantes canônicas via MANOVA.

## 8.8. Bibliografia

Ferreira, D. F. (2011) **Estatística multivariada**. 2. ed. Lavras: Editora UFLA.

Fox, J.; Weisberg, S. (2011) **An {R} Companion to applied regression**. 2. ed. Thousand Oaks CA: Sage. URL: [http:](http://)

`//socserv.socsci.mcmaster.ca/jfox/Books/Companion`

Friendly, M.; Fox, J. (2013) **candisc**: Visualizing generalized canonical discriminant and canonical correlation analysis. R package version 0.6-5. Disponível em:  
<http://CRAN.R-project.org/package=candisc>

Johnson, R. A.; Wichern, D. W. (2007) **Applied multivariate statistical analysis**. 6. ed. New Jersey: Pearson.

Krzanowski, W. J. (2000) **Principles of multivariate analysis: a user's perspective**. Oxford: Oxford University Press.

Manly, B. F. J. (2008) **Métodos estatísticos multivariados: uma introdução** (Tradução Sara I. Carmona). 3. ed. Porto Alegre: Bookman.

Rencher, A. (2002) **Methods of multivariate analysis**. 2. ed. New York: John Wiley & Sons.