MAE 5776

ANÁLISE MULTIVARIADA

Júlia M Pavan Soler

pavan@ime.usp.br

1º Semestre/2022

$$Y_{n\times p} = \left(Y_{ij}\right) \in \Re^{n\times p}$$

MAE5776

Já vimos ☺

- ✓ Estatísticas descritivas multivariadas: $\Re^{nxp}$, $\Re^{pxp}$, $\Re^{nxn}$

  Regiões (elipsóides) de Concentração: $R\left(Y_i\right) = \left(Y_i \in \Re^p;\ \left(Y_i - \bar{Y}\right)' S_u^{-1}\left(Y_i - \bar{Y}\right) \le \chi^2_p(\alpha) \right)$

- ✓ Inferência sobre $\mu \in \Re^p$:

  Caso de Uma Única População: $R\left(\mu \mid Y\right) = \left\{ T^2 = n\left(\bar{Y} - \mu\right)' S_u^{-1}\left(\bar{Y} - \mu\right) \le \frac{(n-1)p}{(n-p)} F_{p,(n-p)}(\alpha) \right\}$

  Caso de Duas Populações:

$$R\left(\mu_D \mid Y_1, Y_2\right) = \left\{ T^2 = n\left(\bar{D} - \mu_D\right)' S_D^{-1}\left(\bar{D} - \mu_D\right) \le \frac{(n-1)p}{(n-p)} F_{p,(n-p)}(\alpha) \right\}$$

$$R\left(\mu_D \mid Y_1, Y_2\right) = \left\{ T^2 = \left(\bar{D} - \mu_D\right)' \left(S_c \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right)^{-1} \left(\bar{D} - \mu_D\right) \le \frac{(n_1+n_2-2)p}{(n_1+n_2-p-1)} F_{(p;n_1+n_2-p-1)}(\alpha) \right\}$$

O problema de Comparações Múltiplas

$\Rightarrow$ Caso de Duas ou Mais Populações (**MANOVA**):

Delineamento Completamente Aleatorizado com Um Fator em G níveis

$$Y_{ig\ p\times 1} \sim N_p\left(\mu_{g\ p\times 1}; \Sigma_{g\ p\times p}\right); \quad H_0: \mu_g = \mu,\ \ g = 1, 2, \ldots, G$$

Delineamentos “Completamente Aleatorizado” com Um Único Fator

	T6	T12	T18	Trat
1	1.48	2.81	3.56	1
2	1.04	2.07	2.81	2
3	1.48	2.52	3.41	1
4	1.04	1.93	2.89	2
5	1.80	2.15	3.20	1
6	1.50	2.70	3.75	2
…				
23	1.80	2.15	3.90	1
24	1.20	2.25	3.30	2
25	1.78	2.96	4.00	3
26	1.48	2.81	3.85	4
27	1.33	2.52	3.84	3
28	1.03	2.07	2.96	4
29	1.65	3.00	3.98	3
30	1.50	2.85	3.75	4
…				
45	1.65	3.00	4.05	3
46	1.20	2.70	3.90	4
47	1.35	2.55	3.67	3
48	1.20	2.70	3.60	4

MANOVA.RM do R

Considere os seguintes dados de um Delineamento Completmente Aleatorizado com 1 Fator Tratamento em 4 níveis (Trat=1, Trat=2, Trat=3 e Trat=4)

p=3 variáveis (medidas repetidas de O2): T6, T12 e T18

$$Y_{48\times(3+1)}$$

3 variáveis resposta quantitativas e 1 categórica (identificando grupo)

$$H_0 : \mu_g = \mu_{3\times1}, \ \ g = 1, 2, 3, 4$$

DCA com Um Único Fator

	T6	T12	T18	Trat
1	1.48	2.81	3.56	1
2	1.04	2.07	2.81	2
3	1.48	2.52	3.41	1
4	1.04	1.93	2.89	2
5	1.80	2.15	3.20	1
6	1.50	2.70	3.75	2
…				
23	1.80	2.15	3.90	1
24	1.20	2.25	3.30	2
25	1.78	2.96	4.00	3
26	1.48	2.81	3.85	4
27	1.33	2.52	3.84	3
28	1.03	2.07	2.96	4
29	1.65	3.00	3.98	3
30	1.50	2.85	3.75	4
…				
45	1.65	3.00	4.05	3
46	1.20	2.70	3.90	4
47	1.35	2.55	3.67	3
48	1.20	2.70	3.60	4

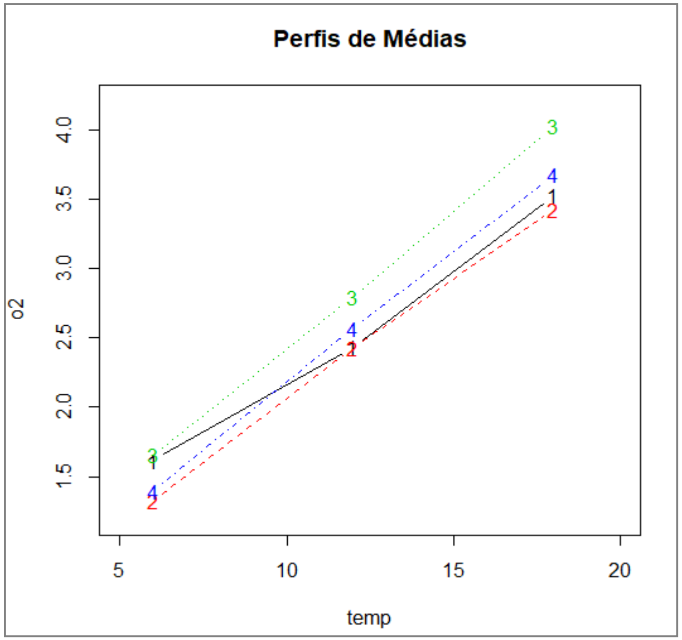
MANOVA.RM do R

$$H_0 : \mu_g = \mu_{3\times1}, \;\; g = 1,2,3,4$$

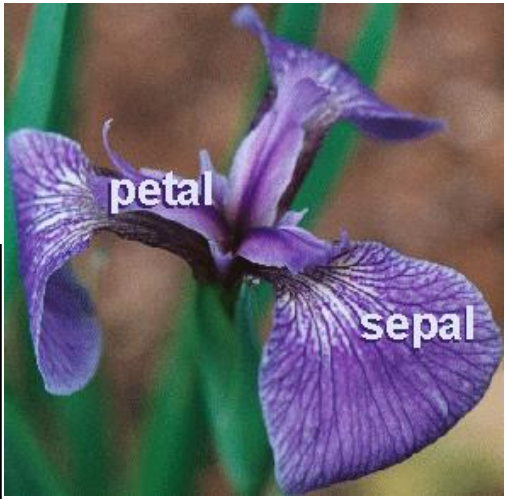
Centróides

	T6	T12	T18
Total(n=48)	1.497500	2.558333	3.664375

Trat	T6	T12	T18
1 (n=9)	1.618333	2.434167	3.526667
2 (n=9)	1.321667	2.430000	3.425000
3 (n=9)	1.655833	2.799167	4.029167
4 (n=9)	1.394167	2.570000	3.676667



Delineamento “Completamente Aleatorizado” com Um Único Fator



Dados “Iris” do R: Medidas do comprimento e largura da pétala e sépala de 50 flores de íris de três espécies (setosa, versicolor e virginica).

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
....					
50	5.0	3.3	1.4	0.2	setosa
51	7.0	3.2	4.7	1.4	versicolor
....					
100	5.7	2.8	4.1	1.3	versicolor
101	6.3	3.3	6.0	2.5	virginica
102	5.8	2.7	5.1	1.9	virginica
....					
150	5.9	3.0	5.1	1.8	virginica

$$Y_{150\times 4} = \begin{pmatrix} Y_{G=1\ 50\times 4} \\ Y_{G=2\ 50\times 4} \\ Y_{G=3\ 50\times 4} \end{pmatrix}$$

150=50+50+50

4=2+2

Comparar os Centróides das 3 espécies (considerando as 4 variáveis)?

$$H_0 : \mu_g = \mu_{4\times 1},\ \ g = 1, 2, 3$$

Inferência sobre Vetores de Médias de "Muitas" Populações

MANOVA

Comparações de Duas Populações $\Rightarrow$ Comparações de Muitas Populações (G≥2)

	População 1	**População 2**		**População G**
População estratificada	$N_p(\mu_1;\Sigma_1)$	$N_p(\mu_2;\Sigma_2)$	...	$N_p(\mu_G;\Sigma_G)$
Amostra	$\underbrace{Y_{11}, Y_{12}, \ldots, Y_{1n_1}}_{\bar{Y}_1 \quad S_1}$	$\underbrace{Y_{21}, Y_{22}, \ldots, Y_{2n_2}}_{\bar{Y}_2 \quad S_2}$		$\underbrace{Y_{G1}, Y_{G2}, \ldots, Y_{Gn_G}}_{\bar{Y}_G \quad S_G}$

Matriz de Dados: $Y_{n\times(p+1)} = \left(Grupo \ \boxed{Y_1 \ Y_2 \ \ldots Y_p} \right)$

Var. qualitativa

Var. quantitativas

Inferência sobre Vetores de Médias de "Muitas" Populações

$\mathbf{Y}_{ig\ p\times 1} = \left(Y_{ig1}, Y_{ig2}, \ldots, Y_{igp}\right)'$: vetor de observações da unidade i no grupo g

Modelo distribucional: $Y_{ig} \overset{iid}{\sim} N_p\left(\mu_g ; \Sigma_g\right)$

$$Y_{n\times p} \in \Re^{n\times p}; \quad Y_{n\times p} \overset{H_0}{\sim} N_{n\times p}\left(\mu_{n\times p}\ ;\ \Omega_{np\times np} = diag_{g=1}^{G}\left(I_{n_g} \otimes \Sigma_g\right)\right)$$

$\boxed{H_0 : \mu_1 = \mu_2 = \ldots = \mu_G; \quad \Sigma_g = \Sigma}$ Sob Homocedasticidade!

Sob H_0

$$Y_{n\times p} \in \Re^{n\times p}; \quad Y_{n\times p} \overset{H_0}{\sim} N_{n\times p}\left(\mu_{n\times p} = 1_n \mu' ; \Omega_{np\times np} = I_n \otimes \Sigma\right)$$

Independência das observações entre e dentro dos grupos

$$\Rightarrow \mu_{n\times p} = 1_n \mu_{p\times 1}' = \begin{pmatrix} \mu' \\ \mu' \\ \ldots \\ \mu' \end{pmatrix} \qquad \Rightarrow \Omega = I_n \otimes \Sigma = \begin{pmatrix} \Sigma & 0 & \cdots & 0 \\ 0 & \Sigma & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \Sigma \end{pmatrix}$$

Inferência sobre Vetores de Médias de Muitas Populações

$$Y_{g\,n_g \times p}; \quad Y_{gi} \overset{iid}{\sim} N_p\left(\mu_g; \Sigma_g\right); \quad g = 1, \dots, G$$

Hipótese condicional: sob homocedasticidade

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_G; \quad \Sigma_g = \Sigma$$

$$H_1 :\in \text{ pelo menos um centróide diferente}; \quad \Sigma_g = \Sigma$$

Estimador de máxima verossimilhança **EMVS sob H_0:**

$$\Rightarrow \begin{cases} \bar{Y} = 1/nY'1_n; \quad n = n_1 + \dots + n_G \\ \hat{\Sigma} = S_{p\times p} = \dfrac{1}{n}\displaystyle\sum_{g=1}^{G}\sum_{i=1}^{n_g}\left(Y_{ig} - \bar{Y}\right)\left(Y_{ig} - \bar{Y}\right)' \end{cases}$$

Divisor n

EMVS sob H_1:

$$\Rightarrow \begin{cases} \bar{Y}_g; \quad g = 1, \dots, G \\ \hat{\Sigma} = S_{c.MVS} = \dfrac{n_1 S_1 + \dots + n_G S_G}{n} \end{cases}$$

Divisor n

$nS = T_{p\times p}$ Matriz de Soma de Quadrados e Produtos Cruzados TOTAL. SST

$nS_{c.MVS} = E_{p\times p}$ Matriz de Soma de Quadrados e Produtos Cruzados DENTRO de GRUPOS SSW

$H = T - E = \displaystyle\sum_{g=1}^{G} n_g\left(\bar{Y}_g - \bar{Y}\right)\left(\bar{Y}_g - \bar{Y}\right)'$ Matriz de Soma de Quadrados e Produtos Cruzados ENTRE GRUPOS SSB

Fontes de Variabilidade

Grupo	u.a.	Y_1	Y_2	…	Y_j	…	Y_p	
1	1	Y_{111}	Y_{112}	…	Y_{11j}	…	Y_{11p}	
		…	…					
1	n_1	$Y_{n_1 11}$	$Y_{n_1 12}$	…	$Y_{n_1 1j}$	…	$Y_{n_1 1p}$	
		$\bar{Y}_{.11}$	$\bar{Y}_{.12}$	…	$\bar{Y}_{.1j}$	…	$\bar{Y}_{.1p}$	$\bar{Y}_{.1.}$
2	1	Y_{121}	Y_{122}	…	Y_{12j}	…	Y_{12p}	
		…	…					
2	n_2	$Y_{n_2 21}$	$Y_{n_2 22}$	…	$Y_{n_2 2j}$	…	$Y_{n_1 2p}$	
		$\bar{Y}_{.21}$	$\bar{Y}_{.22}$	…	$\bar{Y}_{.2j}$	…	$\bar{Y}_{.2p}$	$\bar{Y}_{.2.}$
				…		…		
G	1	Y_{1G1}	Y_{1G2}	…	Y_{1Gj}	…	Y_{1Gp}	
		…	…					
G	n_G	$Y_{n_G G1}$	$Y_{n_G G2}$	…	$Y_{n_G Gj}$	…	$Y_{n_G Gp}$	
		$\bar{Y}_{.G1}$	$\bar{Y}_{.G2}$	…	$\bar{Y}_{.Gj}$	…	$\bar{Y}_{.Gp}$	$\bar{Y}_{.G.}$
								$\bar{Y}_{...}$

Variabilidade Dentro de Grupo (E)

Variabilidade Entre Grupos (H)

T = H + E

$$E_{p\times p} = \sum_{g=1}^{G}\sum_{i=1}^{n_g}\left(Y_{ig}-\bar{Y}_g\right)\left(Y_{ig}-\bar{Y}_g\right)'$$

$$H_{p\times p} = \sum_{g=1}^{G} n_g\left(\bar{Y}_g-\bar{Y}\right)\left(\bar{Y}_g-\bar{Y}\right)'$$

$$T_{p\times p} = nS = \sum_{g=1}^{G}\sum_{i=1}^{n_g}\left(Y_{ig}-\bar{Y}\right)\left(Y_{ig}-\bar{Y}\right)'$$

Identidade útil em $\Re^p$
(decomposição útil)

$$\boxed{Y_{ig} = \bar{Y} + \left(\bar{Y}_g-\bar{Y}\right) + \left(Y_{ig}-\bar{Y}_g\right)}$$

Lembrando "ANOVA": Delineamento Completamente Aleatorizado

Para p=1

$N(\mu_1;\sigma^2)$ $N(\mu_2;\sigma^2)$ ... $N(\mu_G;\sigma^2)$ **População**

T_1	T_2	. . .	T_G	**Amostra**
Y_{11}	Y_{21}	. . .	Y_{G1}	
. . .	. . .	Y_{ij}	. . .	
Y_{1n_1}	$Y_{2\,n_2}$	. . .	$Y_{G\,n_G}$	
n_1	n_2	. . .	n_G	
$\bar{Y}_1$	$\bar{Y}_2$	. . .	$\bar{Y}_G$	
S_1	S_2	. . .	S_G	

- ✓ **Normalidade**
- ✓ **Variância constante**
- ✓ **Independência**

Lembrando ANOVA - Modelo Linear Geral

Resposta da observação
i do grupo *g*

Modelo Estrutural: $\boxed{y_{ig}} = \mu_g + e_{ig}$ Parametrização de Médias

$$= \mu + \tau_g + e_{ig} \quad ; \quad \sum_{g=1}^{G} \tau_g = 0$$ Parametrização de Desvios

$$= \begin{cases} \mu_1 + e_{ig} \\ \mu_1 + \tau_g + e_{ig} ; \quad g = 2, \ldots, G \end{cases}$$ Parametrização Casela de Referência

Modelo Distribucional: $e_{ig} \overset{iid}{\sim} N_1\left(0; \sigma^2\right)$

Forma matricial:

$$Y_{n\times 1} = X_{n\times G} \quad \beta_{G\times 1} + e_{n\times 1}$$

vetor de observações	Matriz de Planejamento	vetor de parâmetros	vetor de erros

$$Y_{ig} = \mu + \tau_g + e_{ig}; \quad \sum_g \tau_g = 0$$ Usando a parametrização de desvios

G=3 grupos
com 5 observações
por grupo

$$\underset{n\times 1}{Y} = \underset{n\times G}{X}\ \underset{G\times 1}{\beta} + \underset{n\times 1}{e}$$

$$\begin{bmatrix} Y_{11} \\ Y_{21} \\ Y_{31} \\ Y_{41} \\ Y_{51} \\ Y_{12} \\ Y_{22} \\ Y_{32} \\ Y_{42} \\ Y_{52} \\ Y_{13} \\ Y_{23} \\ Y_{33} \\ Y_{43} \\ Y_{53} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{21} \\ e_{31} \\ e_{41} \\ e_{51} \\ e_{12} \\ e_{22} \\ e_{32} \\ e_{42} \\ e_{53} \\ e_{13} \\ e_{23} \\ e_{33} \\ e_{43} \\ e_{53} \end{bmatrix}$$

ANOVA - Fontes de Variação

Considere a seguinte identidade (decomposição útil para se obter as fontes de variação envolvidas no modelo):

$$y_{ig} = \bar{y} + (\bar{y}_g - \bar{y}) + (y_{ig} - \bar{y}_g)$$

$$\underbrace{y_{ig} - \bar{y}}_{\sum_{g,i}\left(y_{ig} - \bar{y}\right)^2} = \underbrace{(\bar{y}_g - \bar{y})}_{\sum_{g} n_g (\bar{y}_g - \bar{y})^2} + \underbrace{(y_{ig} - \bar{y}_g)}_{\sum_{g,i} (y_{ig} - \bar{y}_g)^2}$$

SQTotal

corrigida pela média

SQTratamento

“fonte de variação Entre grupos”

SQResidual

“fonte de variação Dentro do grupo”

Tabela de ANOVA

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_G = \mu \in \Re \quad \times \quad H_1: \exists \text{ pelo menos uma diferença}$$

F.V.	g l	SQ	QM	F	valor-p
ENTRE	**G-1**	$\sum_g n_g (\bar{y}_g - \bar{y})^2$	**SQEntre/(G-1)**	**QME/QMD**	
DENTRO	**n-G**	$\sum_{g,i} (y_{ig} - \bar{y}_g)^2$	**SQDentro/(n-G)**		
TOTAL	**n-1**	$\sum_{g,i} (y_{ig} - \bar{y})^2$			

$$\mathbf{F} = \frac{QMEntre}{QMDentro} \sim \mathbf{F}_{(G-1,\, n-G)}$$

Sob: $y_{ig} \overset{iid}{=} \mu_g + e_{ig}; \quad e_{ig} \overset{iid}{\sim} N_1\left(0; \sigma^2\right)$

normalidade
homocedasticidade
independência

Modelo MANOVA

vetor de observações da unidade i do grupo g

DCA: Delineamento Completamente Aleatorizado (1 fator em G níveis)

$$Y_{ig\,p\times 1} \overset{iid}{\sim} N_p\left(\mu_g ; \Sigma_g\right); \; i = 1, \ldots, n_g, \; g = 1, \ldots, G$$

Modelo estrutural:

$$Y_{ig\,p\times 1} = \underbrace{\boxed{\mu + \tau_g}}_{} + \underset{\text{Aleatório}}{e_{ig}} \quad ; \quad \sum_{g=1}^{G} \tau_g = 0$$

Efeito Fixo

Aleatório

Modelo distribucional:

$$e_{ig} \overset{iid}{\sim} N_p\left(0; \Sigma\right) \Rightarrow Y_{ig} \overset{iid}{\sim} N_p\left(\mu_g ; \Sigma\right)$$

Suposições: observações independentes (Entre grupos e Dentro de grupo), Distribuição Normal p-variada, Matriz de Covariâncias homogênea

Modelo MANOVA

$$Y_{ig\,p\times 1} = \left(Y_{ig1}, Y_{ig2}, \ldots, Y_{igp}\right)' \qquad Y_{ig\,p\times 1} = \mu + \tau_g + e_{ig} \quad ; \quad \sum_{g=1}^{G} \tau_g = 0 \qquad e_{ig} \overset{iid}{\sim} N_p(0;\Sigma)$$

Identidade útil para descrever as fontes de variação:

$$\boxed{Y_{ig\,p\times 1} = \bar{Y} + \left(\bar{Y}_g - \bar{Y}\right) + \left(Y_{ig} - \bar{Y}_g\right)}$$

$$H = \sum_{g=1}^{G} n_g \left(\bar{y}_g - \bar{y}\right)\left(\bar{y}_g - \bar{y}\right)'$$

matriz de SQPC devido ao efeito do tratamento (Entre Grupos) - Notação: H=SSB (*Between)*

$$E = \sum_{g=1}^{G}\sum_{i=1}^{n_g} \left(y_{ig} - \bar{y}_g\right)\left(y_{ig} - \bar{y}_g\right)' = \left(n_1 - 1\right)S_{u1} + \ldots + \left(n_G - 1\right)S_{uG}$$

: matriz de SQPC devido ao erro (Dentro de Grupos)

Notação: E=SSW (*Within)*

$$H + E = \sum_{g=1}^{G}\sum_{i=1}^{n_g} \left(y_{ig} - \bar{y}\right)\left(y_{ig} - \bar{y}\right)'$$

: matriz de SQPC total corrigida pela média

Notação: H+E=T

Tabela de MANOVA

$$H_0 : \ \mu_1 = \mu_2 = \ldots = \mu_G = \mu$$

F.V.	g.l.	Matriz de SQPC
Trat	G-1	$H_{p\times p} = \sum_{g=1}^{G} n_g \left(\bar{Y}_g - \bar{Y}\right)\left(\bar{Y}_g - \bar{Y}\right)'$
Resíduo	n-G	$E_{p\times p} = \sum_{g=1}^{G}\sum_{i=1}^{n_g} \left(Y_{ig} - \bar{Y}_g\right)\left(Y_{ig} - \bar{Y}_g\right)'$
TOTAL	n-1	$T = H + E = \sum_{g=1}^{G}\sum_{i=1}^{n_g} \left(Y_{ig} - \bar{Y}\right)\left(Y_{ig} - \bar{Y}\right)'$

$$\Lambda^* = \frac{|\ E\ |}{|\ H + E\ |} = \left|T^{-1}E\right| = \left|I + E^{-1}H\right|^{-1}$$

Estatística lambda de Wilks (critério baseado na estatística Razão de Verossimilhanças sob normalidade multivariada, independência e homocedasticidade).

Quanto menor maior é a evidência para efeito de grupo!

Distribuição da Estatística Λ*

# Var.	# Grupos	Distribuição Amostral (sob $Y_{ig} \overset{iid}{\sim} N_p(\mu_g;\Sigma)$)
p = 1	g ≥ 2	$\left(\dfrac{N-g}{g-1}\right)\left(\dfrac{1-\Lambda^*}{\Lambda^*}\right) \sim F_{g-1,N-g}$
p = 2	g ≥ 2	$\left(\dfrac{N-g-1}{g-1}\right)\left(\dfrac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2(g-1),2(N-g-1)}$
p ≥ 1	g = 2	$\left(\dfrac{N-p-1}{p}\right)\left(\dfrac{1-\Lambda^*}{\Lambda^*}\right) \sim F_{p,N-p-1}$
p ≥ 1	g =3	$\left(\dfrac{N-p-2}{p}\right)\left(\dfrac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2p,2(N-p-2)}$

Caso assintótico: $-\left(N-1-\dfrac{p+g}{2}\right)\ln\left(\dfrac{|E|}{|H+E|}\right) \underset{\sim}{\overset{\substack{n\to\infty \\ (n-p)\to\infty}}{}} \chi^2_{p(g-1)}(\alpha)$

Dados Iris (G=3,p=4)

Tabela de MANOVA

Qual é o modelo estrutural da MANOVA?

$$H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{\mu} \quad \Leftrightarrow \quad H_0: \boldsymbol{\tau}_1 = \tau_2 = \boldsymbol{\tau}_3 = 0$$

F.V.	No. g.l.	Matriz de Soma de Quadr e Prod Cruzados	
Grupo	3-1	SSB=H $\begin{matrix} 63.21 & -19.95 & 165.25 & 71.28 \\ & 11.34 & -57.24 & -22.93 \\ & & 437.10 & 186.77 \\ & & & 80.41 \end{matrix}$	Matriz de SQPC devido ao efeito de Grupo
Resíduo	150-3	SSW=E $\begin{matrix} 38.96 & 13.63 & 24.62 & 5.65 \\ & 16.96 & 8.12 & 4.81 \\ & & 27.22 & 6.27 \\ & & & 6.16 \end{matrix}$	Matriz de SQPC devido ao efeito do Erro
TOTAL	150-1	SST=H+E $\begin{matrix} 102.17 & -6.32 & 189.87 & 76.92 \\ & 28.31 & -49.12 & -18.12 \\ & & 464.33 & 193.05 \\ & & & 86.57 \end{matrix}$	Matriz de SQPC Total

$$\Lambda^* = \frac{|E|}{|H+E|} = 1.486462\text{e-}31 \quad \Rightarrow \underbrace{\left(\frac{N-p-2}{p}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right)}_{9.337386\text{e+}16} \sim F_{2p,2(N-p-2)}(\alpha = 0{,}01)$$

2.573429

Concl.?

Dados Iris (G=3,p=4)
Tabela de MANOVA $\Rightarrow$ Tabelas de ANOVA

$$H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{\mu} \quad \Rightarrow \quad H_{0j}: \mu_{1j} = \mu_{2j} = \mu_{3j} = \mu_j \quad j=1,2,\ldots,p$$

Tabela ANOVA para Variável 1:

F.V.	No. g.l.	SQ	QM	F
Grupo	3-1	63.21	31.61	31.61/0.265=119.27
Resíduo	150-3	38.96	0.265	
TOTAL	150-1	102.17		

qf(0.99;2;147) = 4.7525

Concl.?

Resultados da ANOVA para cada Variável:

Fontes de Var.	SQGrupo	SQResíduo	F	
resp 1	63.2121	38.9562	119.27	
resp 2	11.3449	16.9620	49.16	
resp 3	437.1028	27.2226	1180.16	
resp 4	80.4133	6.1566	960.01	
Número de g.l.	2	147		
Erro padrão residual:	0.5147894	0.3396877	0.4303345	0.20465

=√0.265

Obtenha as Tabelas ANOVA para cada variável!
Da análise individual das variáveis a de maior dignificância é a Varável 3 seguida da 4!

Estatísticas da MANOVA

Critério	Estatística	Aproximação F
Wilks	$\Lambda = \dfrac{\lvert E \rvert}{\lvert H+E \rvert} = \prod_i \dfrac{1}{1+\lambda_i}$	$\left(\dfrac{rt-2f}{pq}\right)\left(\dfrac{1-\Lambda^{1/t}}{\Lambda^{1/t}}\right) \sim F_{pq,(rt-2f)}$
Traço de Pillai	$V = tr\left(\dfrac{H}{H+E}\right) = \sum_i \dfrac{\lambda_i}{1+\lambda_i}$	$\left(\dfrac{2n+s+1}{2m+s+1}\right)\left(\dfrac{V}{s-V}\right) \sim F_{s(2m+s+1),s(2n+s+1)}$
Traço de Hotelling Lawley	$U = tr\left(\dfrac{H}{E}\right) = \sum_i \lambda_i$	$\dfrac{2(sn+1)U}{s^2(2m+s+1)} \sim F_{s(2m+s+1),2(sn+1)}$
Raiz Máxima de Roy	$\theta = \lambda_1$	$\dfrac{(\nu-d+q)\theta}{d} \sim F_{d,(\nu-d+q)}$

p= # de var.; q=g.l. trat (ou do contraste); ν=g.l. erro; s=min(p,q); r=(p+q+1)/2; f=(pq-2)/4

d= max(p,q); m=(|p-q|-1)/2; n=(ν-p-1)/2; **λ: autovalor de |H-λE|=0;**

$$t = \sqrt{(p^2q^2-4)/(p^2+q^2-5)} \text{ se } (p^2+q^2-5) > 0, \text{ ou } 1 \text{ c.c.}$$

Contribuição das Variáveis na Discriminação dos Grupos

Considere a seguinte decomposição espectral:

$$|H-\lambda E|=0;\ (H-\lambda E)l=0 \Rightarrow \frac{l_k' H l_k}{l_k' E l_k}=\lambda_k$$

$$l_k=(l_{k1}\ \ l_{k2}\ \dots l_{kp})'$$

Notação: H=SSB; E=SSW
A informação sobre discriminação entre os grupos está na decomposição espectral de $E^{-1}H$

A avaliação dos coeficientes dos autovetores, l, associados aos maiores autovalores, λ, define a importância de cada variável no efeito de tratamento. ***Este resultado decorre também da Análise Discriminante de Fisher.***

Dados Iris:

➢ L **Matriz de Autovetores**

	l_1	l_2	l_3	l_4
[Var1,]	-0.06840592	-0.001987912	0.23777665	0.1120017
[Var2,]	-0.12656121	-0.178526702	-0.09841836	-0.1981247
[Var3,]	**0.18155288**	0.076863566	-0.09653934	-0.2289487
[Var4,]	**0.23180286**	-0.234172267	-0.01460600	0.3759916

```
> eigen(M)$values   Autovalores
[1]  3.219193e+01  2.853910e-01  1.286765e-15 -1.554312e-15
```

λ_1 λ_2 λ_3 λ_4

A variável Y4, seguida de Y3, mais contribuem para a discriminação entre os grupos (respeitando a estrutura multivariada).

Comparações Múltiplas

Comparações múltiplas entre tratamentos (dois a dois) para cada variável (com correção de Bonferroni):

Dados Iris: Realize comparações múltiplas com correção para os múltiplos testes!

Comparacão entre Tratamentos g e *h*: $\underbrace{\mu_g - \mu_h}_{\tau_g - \tau_h} \Rightarrow \bar{Y}_g - \bar{Y}_h$

Avaliar os componentes do vetor de diferenças (p variáveis)

$$\hat{\tau}_{g_{p\times 1}} = \bar{Y}_g - \bar{Y} \Rightarrow \underbrace{\hat{\tau}_{g\,j} = \bar{Y}_{g\,j} - \bar{Y}_j}_{\text{Trat g Variável j}} \qquad \hat{\tau}_h = \bar{Y}_h - \bar{Y} \Rightarrow \underbrace{\hat{\tau}_{h\,j} = \bar{Y}_{h\,j} - \bar{Y}_j}_{\text{Trat } h \text{ Variável j}}$$

$$V\left(\bar{Y}_{g\,j} - \bar{Y}_{h\,j}\right) = \left(\frac{1}{n_g} + \frac{1}{n_h}\right)\boxed{\frac{E_{jj}}{n-G}} \Rightarrow \boxed{\left(\bar{Y}_{g\,j} - \bar{Y}_{h\,j}\right) \pm t_{n-G}(\alpha/2K)\sqrt{V\left(\bar{Y}_{g\,j} - \bar{Y}_{h\,j}\right)}}$$

QMRes
Diag(S_{uc})

Intervalo de confiança 100(1-α)% com Correção de Bonferroni para um total de K comparações.

(Ex., K=p + G(G-1)/2)