MAE 5776

ANÁLISE MULTIVARIADA

Júlia M Pavan Soler

pavan@ime.usp.br

1º Semestre/2022

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

MAE5776

Já vimos ☺

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

- Inferência sobre $\mu \in \Re^p$:

Caso de Uma Única População

O problema de Comparações Múltiplas

Caso de Duas Populações: Amostras pareadas e Amostras Independentes

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

✓ Delineamento Completamente Aleatorizado (DCA) com Um Fator em G níveis

$$\begin{cases} \text{SST} = \overset{\text{Entre}}{\text{SSB}} + \overset{\text{Dentro}}{\text{SSW}} \\ (T = H + E) \Rightarrow |H - \lambda E| = 0 \end{cases}$$

Teste de H0:$\mu_g = \mu$ sob Observações Independentes, Np e Homocedasticidade: Há diferentes estatísticas de teste (lambda de Wilks, Pillai, Roy).

- DCA Fatorial
- DCA Hierárquico
- Delineamento Aleatorizado em Blocos

MANOVA: Delineamento 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)

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

	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

Matrizes de Covariância

Trat=1	T6	T12	T18
T6	0.02	-0.02	0.00
T12	-0.02	0.09	0.00
T18	0.00	0.00	0.08

Trat=2	T6	T12	T18
T6	0.04	0.04	0.04
T12	0.04	0.07	0.06
T18	0.04	0.06	0.11

Trat=3	T6	T12	T18
T6	0.04	0.01	0.05
T12	0.01	0.11	0.01
T18	0.05	0.01	0.07

Trat=4	T6	T12	T18
T6	0.05	0.02	0.06
T12	0.02	0.06	0.05
T18	0.06	0.05	0.12

Sc	T6	T12	T18
T6	0.04	0.01	0.03
T12	0.01	0.08	0.03
T18	0.03	0.03	0.09

Box's M-test:
Chi-Sq=34.61, df = 18,
p-value = 0.01058

α=1% $\Rightarrow$ Não há evidência para a rejeição da hipótese de Homocedasticidade

	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

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

Atenção à decomposição imposta à Matriz de Covariância

SST = SSB + SSW

Tabela de MANOVA:

FV	no.gl	SQPC
Trat	4-1	**SSB** T6 T12 T18 T6 **0.98** 0.53 0.98 T12 0.53 **1.08** 1.63 T18 0.98 1.63 **2.51**
Resíduo	48-4	**SSW** T6 T12 T18 T6 **1.67** 0.55 1.53 T12 0.55 **3.66** 1.24 T18 1.53 1.24 **4.15**
Total	48-1	**SST** T6 T12 T18 T6 2.65 1.08 2.51 T12 1.08 4.74 2.87 T18 2.51 2.87 6.67

Fonte de Variação ENTRE grupos

Fonte de Variação DENTRO de grupos

	Estat.	approxF	numDf	denDf	Pr(>F)	
Pillai	0.7651	5.0206	9	132	8.062e-06	***
Wilks	0.3807	5.5401	9	102.37	3.354e-06	***
Hotel.-Lawley	1.2444	5.6229	9	122	1.742e-06	***
Roy	0.7004	10.273	3	44	3.013e-05	***

Concl.?

	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

Decomposição Espectral de $E^{-1}H$:

Autovalores

[,1]	[,2]	[,3]
0.7004284	0.5425676	0.001408624

Atovetores

	[,1]	[,2]	[,3]
T6	0.01632176	0.94481561	-0.08938875
T12	-0.20483386	-0.07142574	-0.50725864
T18	-0.40074216	-0.30089017	0.36425094

Contribuição das variáveis para a discriminação dos Trats!

Modelo estrutural e distribucional adotado:

$$Y_{ig\,3\times 1} = \mu_g + e_{ig}; \quad e_{ig} \sim N_3\left(\mu_g; \Sigma\right)$$

$$= \mu + \tau_g + e_{ig}; \quad \sum_{g=1}^{4} \tau_g = 0$$

Parametrização de desvios

$$= \begin{cases} \mu_1 + e_{i1} \\ \mu_1 + \tau_g + e_{ig}; \quad g = 2,3,4 \end{cases}$$

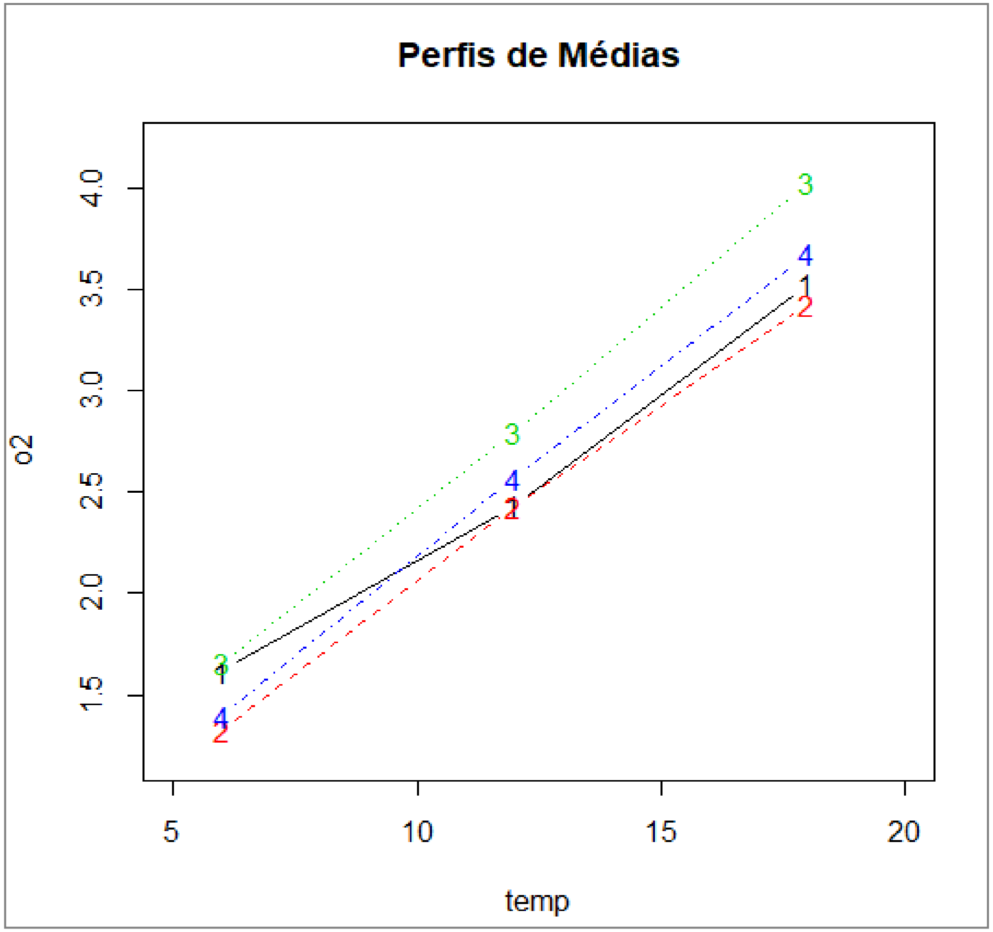
Parametrização casela de referência

Estimativas:	T6	T12	T18
(Intercept)	1.6183333	2.434166667	3.5266667
Trat=2	-0.2966667	-0.004166667	-0.1016667
Trat=3	0.0375000	0.365000000	0.5025000
Trat=4	-0.2241667	0.135833333	0.1500000

$\hat{\mu}_1$ Referência: Trat=1

$\hat{\tau}_g; \quad g = 2,3,4$

Comparações Múltiplas



Intervalos de Confiança de Bonferroni (correção por variável)

		Li	Ls	Conclusão
T6	[2-1]	-0.52	-0.08	$\mu_{21} < \mu_{11}$
	[3-1]	-0.18	0.26	$\mu_{31} = \mu_{11}$
	[4-1]	-0.44	0.00	$\mu_{41} = \mu_{11}$
	[3-2]	0.11	0.55	$\mu_{31} > \mu_{21}$
	[4-2]	-0.15	0.29	$\mu_{41} = \mu_{21}$
	[4-3]	-0.48	-0.04	$\mu_{41} < \mu_{31}$

K=6 compa-rações

		Li	Ls	Conclusão
T12	[2-1]	-0.33	0.32	$\mu_{22} = \mu_{12}$
	[3-1]	0.04	0.69	$\mu_{32} > \mu_{12}$
	[4-1]	-0.19	0.46	$\mu_{42} = \mu_{12}$
	[3-2]	0.04	0.69	$\mu_{32} > \mu_{22}$
	[4-2]	-0.19	0.47	$\mu_{42} = \mu_{22}$
	[4-3]	-0.55	0.10	$\mu_{42} = \mu_{32}$

K=6

		Li	Ls	Conclusão
T18	[2-1]	-0.45	0.24	$\mu_{23} = \mu_{13}$
	[3-1]	0.16	0.85	$\mu_{33} > \mu_{13}$
	[4-1]	-0.20	0.50	$\mu_{43} = \mu_{13}$
	[3-2]	0.26	0.95	$\mu_{33} > \mu_{23}$
	[4-2]	-0.09	0.60	$\mu_{43} = \mu_{23}$
	[4-3]	-0.70	-0.01	$\mu_{43} < \mu_{33}$

K=6

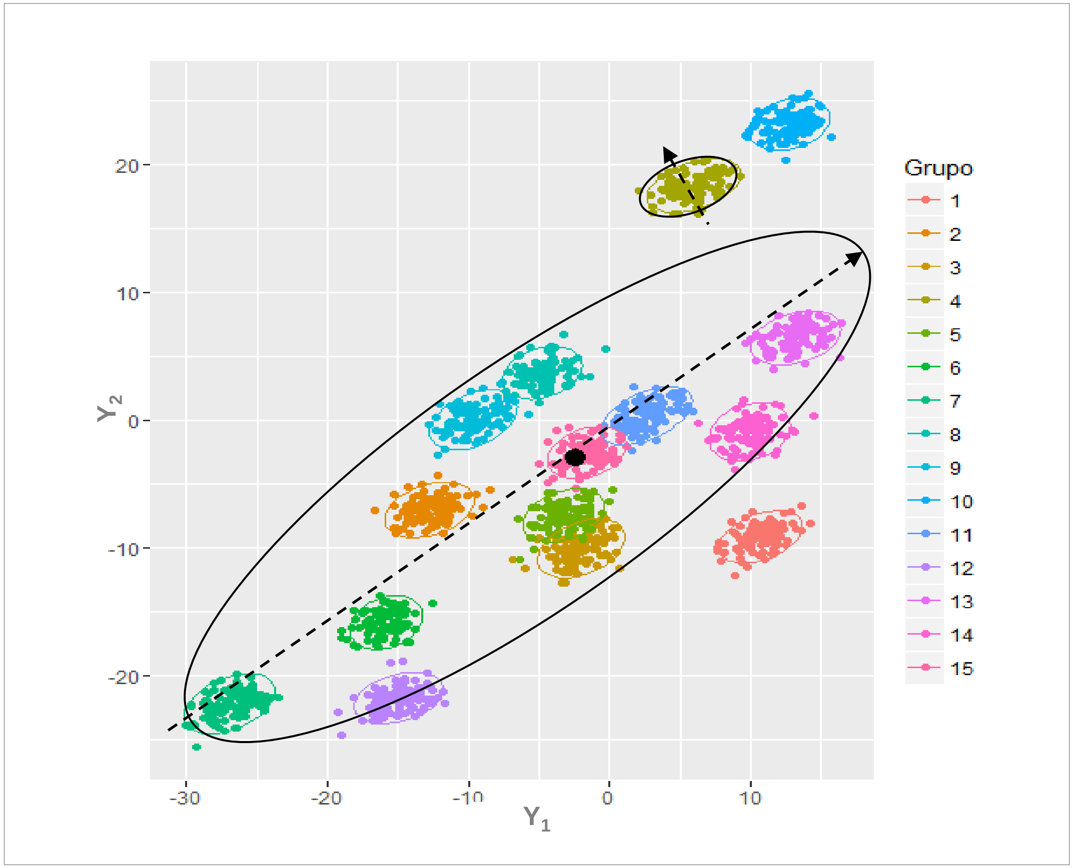
$$ICB\left(\mu_{gj}-\mu_{hj}\right)\ a\ (1-\alpha)100\% = \left(\bar{Y}_{g\,j}-\bar{Y}_{h\,j}\right) \pm t_{n-G}\left(\alpha/2K\right)\sqrt{V\left(\bar{Y}_{g\,j}-\bar{Y}_{h\,j}\right)}$$

$$\left(\frac{1}{n_g}+\frac{1}{n_h}\right)\frac{E_{jj}}{n-G}$$

MANOVA - Fontes de Variação

Dados simulados: Delineamento com População Estratificada em Muitos Grupos (G=15) e p=2

$$Y_{n\times p};\ Y_{ig} \sim N_p\left(\mu_g;\Sigma\right)$$



$\Rightarrow$ Delineamento com um Fator (Grupo, Tratamento) em 15 níveis

SST = SSB + SSW

SSB: Fonte de Variabilidade Entre grupos (elipse maior)

SSW: Fonte de Variabilidade Dentro de grupos (elipses menores)
Em geral, sob homoceadticidade

Situação ideal:
- Efeito de Tratamento: SSB>SSW
- Poder da análise Multivariada: altas covariâncias e de sinais opostos nos componentes de variação SSB e SSW

MANOVA: Modelo Linear Multivariado

$Y_{n\times p}; \quad n = n_1 + \ldots + n_G \qquad Y_{ig\; p\times 1} = \mu_g + e_{ig}; \qquad Y_{n\times p} = X_{n\times G}\, \beta_{G\times p} + e_{n\times p}$

$$\mathbf{Y}_{n\times p} = \begin{pmatrix} Y_{11} & Y_{12} & & Y_{1p} \\ Y_{21} & Y_{22} & & Y_{2p} \\ \cdots & \cdots & \cdots & \\ Y_{n1} & Y_{n2} & & Y_{np} \end{pmatrix}; \qquad e_{n\times p} = \begin{pmatrix} e_{11} & e_{12} & & e_{1p} \\ e_{21} & e_{22} & & e_{2p} \\ \cdots & \cdots & \cdots & \\ e_{n1} & e_{n2} & & e_{np} \end{pmatrix}.$$

Parametrização de médias

$$X_{n\times G} = \begin{pmatrix} 1_{n_1} & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 1_{n_2} & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 1_{n_G} \end{pmatrix}; \qquad \boldsymbol{\beta}_{G\times p} = \begin{pmatrix} \mu_{11} & \mu_{12} & & \mu_{1p} \\ \mu_{21} & \mu_{22} & & \mu_{2p} \\ \cdots & \cdots & \cdots & \\ \mu_{G1} & \mu_{G2} & & \mu_{Gp} \end{pmatrix};$$

Parametrização de desvios

$$X_{n\times G} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & -1 & -1 & -1 \end{pmatrix}; \qquad \boldsymbol{\beta}_{G\times p} = \begin{pmatrix} \mu_{.1} & \mu_{.2} & & \mu_{.p} \\ \tau_{11} & \tau_{12} & & \tau_{1p} \\ \cdots & \cdots & \cdots & \\ \tau_{(G-1)1} & \tau_{(G-1)2} & & \tau_{(G-1)p} \end{pmatrix};$$

MANOVA: Modelo Linear Multivariado

$$Y_{n\times p} = X_{n\times G}\,\beta_{G\times p} + e_{n\times p}$$

Estimadores de Mínimos Quadrados e de MVS

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y \qquad \hat{Y} = X\hat{\beta} = X\left(X'X\right)^{-1} X'Y = PY$$

$$\hat{e} = Y - \hat{Y} = \left(I_n - X\left(X'X\right)^{-1} X'\right)Y = \left(I_n - P\right)Y \qquad \hat{e}'\hat{e}/n = \hat{\Sigma} = S$$

$$P = X\left(X'X\right)^{-1} X'$$

MANOVA: Modelo Linear Multivariado

$$Y_{n\times p} = X_{n\times G}\,\beta_{G\times p} + e_{n\times p}$$

Teste de Hipóteses Gerais

$$H_0 : C_{c\times G}\beta_{G\times p}U_{p\times u} = 0$$

$C_{c\times G}$: define contrastes entre as médias de grupos

$U_{p\times u}$: define contrastes entre as médias das variáveis

Estatísticas de Teste: Wilks, Pillai, Lawley-Hotelling, Roy

$$\text{Lambda de Wilks} : \lambda = \frac{|E|}{|H+E|}$$

Considerar os autovalores e autovetores de : $\left(H - \lambda E\right)l = 0$

$$H = \left(C\hat{\beta}U\right)'\left[C\left(X'X\right)^{-1}C'\right]^{-1}C\hat{\beta}U \qquad E = \left(YU\right)'\left[I - X\left(X'X\right)^{-1}X'\right]^{-1}YU$$

MANOVA: Modelo Linear Multivariado

$$Y_{n\times p} = X_{n\times G}\,\beta_{G\times p} + e_{n\times p}$$

Teste de Hipóteses Gerais $H_0 : C_{c\times G}\beta_{G\times p}U_{p\times u} = 0$

Exemplo: Considere um DCA balanceado e a parametrização de médias. Os seguintes ***Contrastes Ortogonais*** podem ser definidos (para comporem as linhas da matriz C): $C'_{l_1}C_{l_2} = 0$

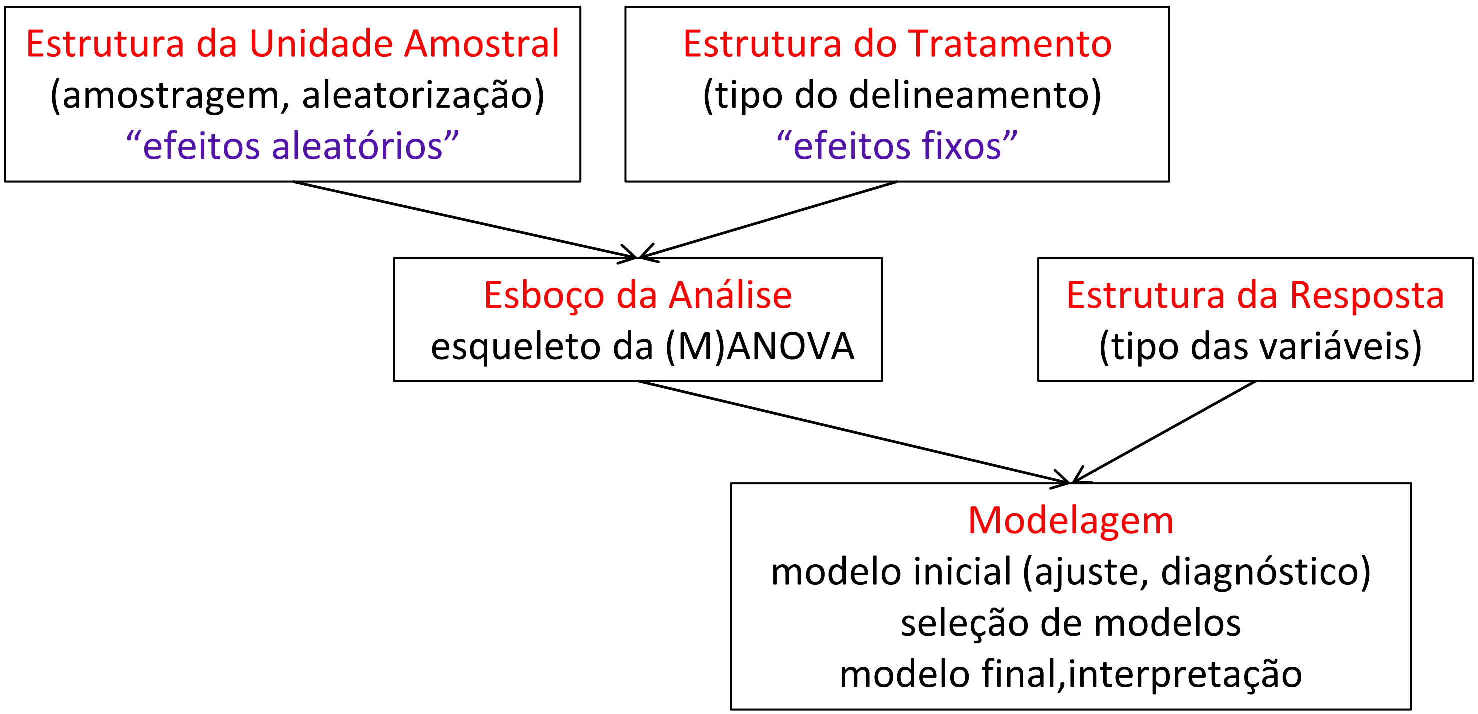
G=3 Tratamentos
$$\begin{cases} C1 = (1 \quad 0 \quad -1) \\ C2 = (\tfrac{1}{2} \quad -1 \quad \tfrac{1}{2}) \end{cases}$$

G=4 Tratamentos
$$\begin{cases} C1 = (1 \quad -1/3 \quad -1/3 \quad -1/3) \\ C2 = (0 \quad -1/2 \quad -1/2 \quad 1) \\ C3 = (0 \quad 1 \quad -1 \quad 0) \end{cases}$$

Note que o número de graus de liberdade para estudar o efeito de Tratamento é (G-1): neste caso, a significância da estatística do teste coletivo das hipóteses é dada pela significância do correspondente efeito na tabela de MANOVA

Estrutura Geral de Análise de Dados

(Goos and Gilmour, 2012)



MANOVA – Diferentes Delineamentos

- Estrutura da Resposta: dados quantitativos multivariados ($\sim N_p$)

- Estrutura das Unidades Amostrais: Observações independentes

  Amostra aleatória simples de tamanho *n* de uma população sob estudo

  Como os Tratamentos foram aleatorizados às unidades amostrais?

  ✓ Delineamento Completamente Aleatorizado (DCA)
  Delineamento Aleatorizado em Blocos Completos (DABC)

- Estrutura dos Tratamentos:

  ✓ Delineamento com Um Único Fator (em G níveis)
  Delineamento Fatorial Cruzado (mais de um Fator)
  Delineamneto Fatorial Hierarquico

Delineamento Fatorial

Dados O2:

DCA com 1 Fator em 4 níveis → DCA Fatorial 2x2: Dois Fatores, cada um em dois níveis

	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

	Subj	Grup	Staph	T6	T12	T18
1	1	P	1	1.48	2.81	3.56
2	2	P	0	1.04	2.07	2.81
3	3	P	1	1.48	2.52	3.41
4	4	P	0	1.04	1.93	2.89
…						
21	21	P	1	1.50	2.85	3.12
22	22	P	0	1.65	2.70	3.40
23	23	P	1	1.80	2.15	3.90
24	24	P	0	1.20	2.25	3.30
25	25	V	1	1.78	2.96	4.00
26	26	V	0	1.48	2.81	3.85
27	27	V	1	1.33	2.52	3.84
28	28	V	0	1.03	2.07	2.96
…						
45	45	V	1	1.65	3.00	4.05
46	46	V	0	1.20	2.70	3.90
47	47	V	1	1.35	2.55	3.67
48	48	V	0	1.20	2.70	3.60

Estrutura de Dados

Dados: Medidas Repetidas
no formato "wide"

Alternativa de
análise: MANOVA

```
> library(reshape2)
> dat.wide <- dcast(dat.long, Subj +
Grup + Staph~ Time, value.var="O2")
> dat.wide
   Subj  Grup Staph    6    12    18
1     1     P     1 1.48 2.81 3.56
2     2     P     0 1.04 2.07 2.81
3     3     P     1 1.48 2.52 3.41
4     4     P     0 1.04 1.93 2.89
…
21   21     P     1 1.50 2.85 3.12
22   22     P     0 1.65 2.70 3.40
23   23     P     1 1.80 2.15 3.90
24   24     P     0 1.20 2.25 3.30
25   25     V     1 1.78 2.96 4.00
26   26     V     0 1.48 2.81 3.85
27   27     V     1 1.33 2.52 3.84
28   28     V     0 1.03 2.07 2.96
…
45   45     V     1 1.65 3.00 4.05
46   46     V     0 1.20 2.70 3.90
47   47     V     1 1.35 2.55 3.67
48   48     V     0 1.20 2.70 3.60
```

Dados: Medidas Repetidas
no formato "long"

Alternativa de
análise: Modelos
lineares mistos

```
> library(MANOVA.RM)
> dat.long<-o2cons
> dat.long
> dat.long
      O2 Staph Time  Grup Subj
1   1.48     1    6     P    1
2   2.81     1   12     P    1
3   3.56     1   18     P    1
4   1.04     0    6     P    2
5   2.07     0   12     P    2
6   2.81     0   18     P    2
…
67  1.80     1    6     P   23
68  2.15     1   12     P   23
69  3.90     1   18     P   23
70  1.20     0    6     P   24
71  2.25     0   12     P   24
72  3.30     0   18     P   24
73  1.78     1    6     V   25
74  2.96     1   12     V   25
75  4.00     1   18     V   25
76  1.48     0    6     V   26
77  2.81     0   12     V   26
78  3.85     0   18     V   26
…
139 1.35     1    6     V   47
140 2.55     1   12     V   47
141 3.67     1   18     V   47
142 1.20     0    6     V   48
143 2.70     0   12     V   48
144 3.60     0   18     V   48
```

Dados: Medidas Repetidas no formato "wide"

```
> library(reshape2)
> dat.wide <- dcast(dat.long, Subj +
Grup + Staph~ Time, value.var="O2")
> dat.wide
   Subj  Grup Staph    6   12   18
1     1     P     1 1.48 2.81 3.56
2     2     P     0 1.04 2.07 2.81
3     3     P     1 1.48 2.52 3.41
4     4     P     0 1.04 1.93 2.89
…
21   21     P     1 1.50 2.85 3.12
22   22     P     0 1.65 2.70 3.40
23   23     P     1 1.80 2.15 3.90
24   24     P     0 1.20 2.25 3.30
25   25     V     1 1.78 2.96 4.00
26   26     V     0 1.48 2.81 3.85
27   27     V     1 1.33 2.52 3.84
28   28     V     0 1.03 2.07 2.96
…
45   45     V     1 1.65 3.00 4.05
46   46     V     0 1.20 2.70 3.90
47   47     V     1 1.35 2.55 3.67
48   48     V     0 1.20 2.70 3.60
```

Delineamento Fatorial 2x2,
ou
✓ Um Fator em 4 níveis

Entendendo a estrutura dos dados:

Delineamento Completamente Aleatorizado, Fatorial Cruzado:

1 Fator em 4 níveis

Fator Grupo (2 níveis): P e V

Fator Staphylococus (2 níveis): 1 e 0

Estrutura de aleatorização das unidades amostrais (experimentais) aos fatores: Delineamento Completamente Aleatorizado

Estrutura dos Fatores (Tratamentos): Fatorial Cruzado
(4 Tratamentos no total: 2 fatores, cada um em 2 níveis)

Delineamento balanceado: n=12 unidades experimentais em cada Tratamento (combinação dos fatores)

p=3 respostas avaliadas em cada sujeito (medidas repetidas de O2)

Dados O2:

DCA Fatorial 2x2

```
> library(reshape2)
> dat.wide <- dcast(dat.long, Subj +
Grup + Staph~ Time, value.var="O2")
> dat.wide
   Subj  Grup Staph    6    12    18
1      1     P     1 1.48 2.81 3.56
2      2     P     0 1.04 2.07 2.81
3      3     P     1 1.48 2.52 3.41
4      4     P     0 1.04 1.93 2.89
…
21    21     P     1 1.50 2.85 3.12
22    22     P     0 1.65 2.70 3.40
23    23     P     1 1.80 2.15 3.90
24    24     P     0 1.20 2.25 3.30
25    25     V     1 1.78 2.96 4.00
26    26     V     0 1.48 2.81 3.85
27    27     V     1 1.33 2.52 3.84
28    28     V     0 1.03 2.07 2.96
…
45    45     V     1 1.65 3.00 4.05
46    46     V     0 1.20 2.70 3.90
47    47     V     1 1.35 2.55 3.67
48    48     V     0 1.20 2.70 3.60
```

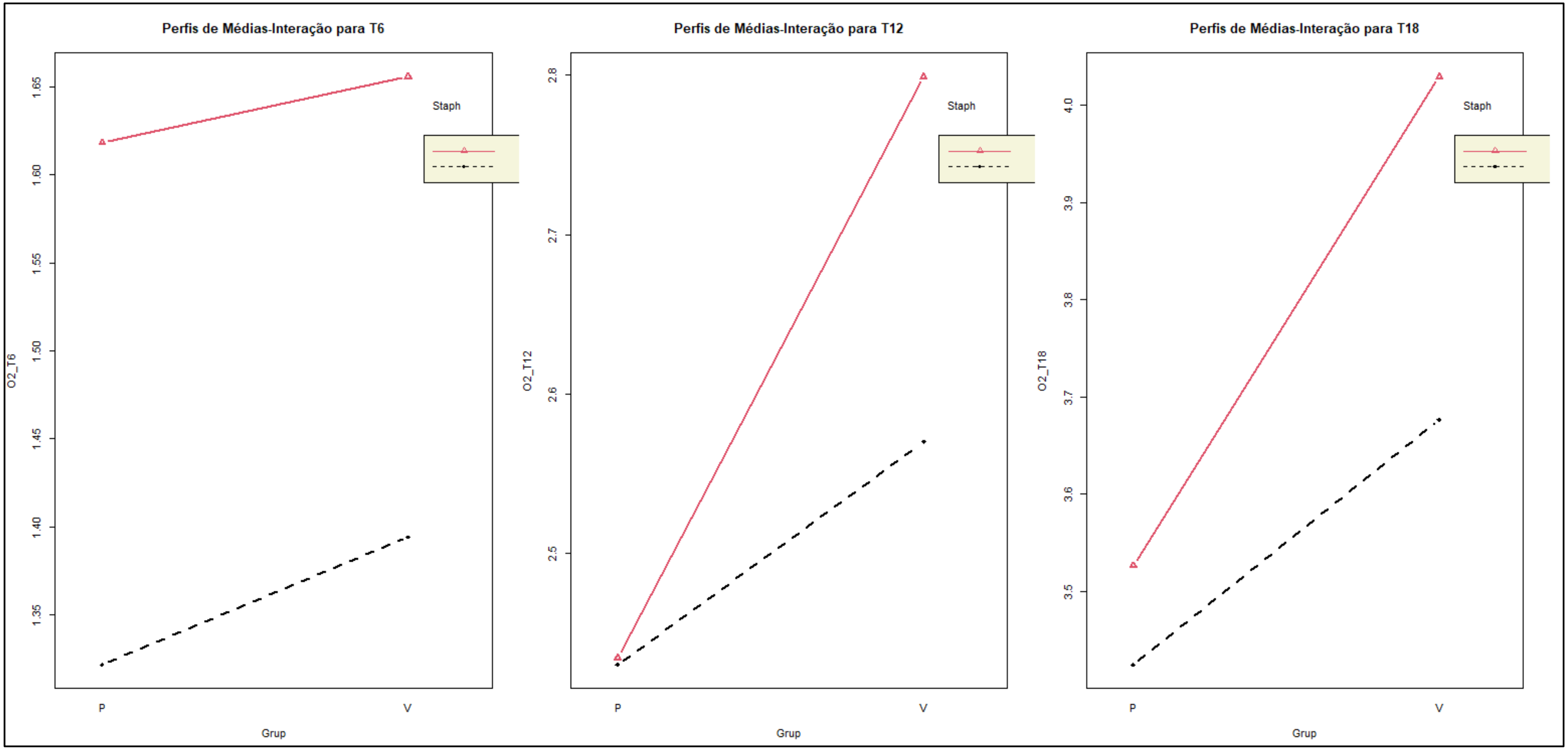
```
Centróides
      T6       T12      T18
Total 1.497500 2.558333 3.664375

Grup  T6       T12      T18
P     1.470    2.432083 3.475833
V     1.525    2.684583 3.852917

Staph T6       T12      T18
0     1.357917 2.500000 3.550833
1     1.637083 2.616667 3.777917

G*S   T6       T12      T18
1     1.618333 2.434167 3.526667
2     1.321667 2.430000 3.425000
3     1.655833 2.799167 4.029167
4     1.394167 2.570000 3.676667
```

- Ajuste de modelos MANOVA sob Delineamento Fatorial Cruzado!

Delineamento Fatorial
Efeito de Interação entre Fatores

Dados O2:

DCA Fatorial 2x2



Delineamento Fatorial

Um estudo tem como objetivo avaliar as condições de fabricação de um filme plástico. Três variáveis resposta (Y1, Y2 e Y3) foram observadas sob dois níveis (baixo e alto) dos fatores F1 e F2.

DCA Fatorial Cruzado (2x2) Balanceado (n=20)

Dados do Arquivo EXH

		Maq2 F2					
		Baixo			Alto		
Maq1	F1	Y1	Y2	Y3	Y1	Y2	Y3
		6,5	9,5	4,4	6,9	9,1	5,7
		6,2	9,9	6,4	7,2	10	2
Baixo		5,8	9,6	3	6,9	9,9	3,9
		6,5	9,6	4,1	6,1	9,5	1,9
		6,5	9,2	0,8	6,3	9,4	5,7
		6,7	9,1	2,8	7,1	9,2	8,4
		6,6	9,3	4,1	7	8,8	5,2
Alto		7,2	8,3	3,8	7,2	9,7	6,9
		7,1	8,4	1,6	7,5	10,1	2,7
		6,8	8,5	3,4	7,6	9,2	1,9

$Y_{20\times(3+1)}$

p=3 variáveis e a categoria de grupo.

$\Rightarrow$ Realizar uma análise de Variância Multivariada destes dados.

Delineamento Fatorial - ANOVA

Ef. principal de F1 Ef. principal de F2 Ef. de interação Resíduo

$$y_{ijk} = \mu_{jk} + e_{ijk} = \mu + \tau_j + \beta_k + \gamma_{jk} + e_{ijk}$$

Resposta da observação i avaliada no nível j do fator 1 e nível k do fator 2

Restrições de identificabilidade

$$\sum_{j=1}^{a} \tau_j = 0, \sum_{k=1}^{b} \beta_k = 0, \sum_{j=1}^{a} \gamma_{jk} = 0, \sum_{k=1}^{b} \gamma_{jk} = 0$$

"Identidade útil" para obtenção das Somas de Quadrados e dos estimadores dos efeitos de interesse:

Ef. principal de F1 Ef. principal de F2 Ef. de interação Resíduo

$$y_{ijk} = \bar{y} + (\bar{y}_{j.} - \bar{y}) + (\bar{y}_{.k} - \bar{y}) + (\bar{y}_{jk} - \bar{y}_{j.} - \bar{y}_{.k} + \bar{y}) + (y_{ijk} - \bar{y}_{jk})$$

SQ_F1 SQ_F2 SQ_F1*F2 SQ_Residual

Caso Multivariado $\Rightarrow$ formulação para o vetor de resposta p-dimensional.

Tabela de MANOVA

Delineamento Completamente Aleatorizado com estrutura Fatorial de Grupos (G=axb) e r réplicas (balanceado)

F.V.	g.l.	Matriz de SSCP
Fator 1	a-1	$HF1=\sum_{j=1}^{a} br\left(\bar{Y}_{j.}-\bar{Y}\right)\left(\bar{Y}_{j.}-\bar{Y}\right)'$
Fator 2	b-1	$HF2=\sum_{k=1}^{b} ar\left(\bar{Y}_{.k}-\bar{Y}\right)\left(\bar{Y}_{.k}-\bar{Y}\right)'$
Interação	(a-1)(b-1)	$HInt=\sum_{j=1}^{a}\sum_{k=1}^{b} r\left(\bar{Y}_{jk}-\bar{Y}_{j.}-\bar{Y}_{.k}+\bar{Y}\right)\left(\bar{Y}_{jk}-\bar{Y}_{j.}-\bar{Y}_{.k}+\bar{Y}\right)'$
Resíduo	ab(r-1)	$E=\sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{r}\left(Y_{ijk}-\bar{Y}_{jk}\right)\left(Y_{ijk}-\bar{Y}_{jk}\right)'$
TOTAL	rab-1	$HF1+HF2+HInt+E=\sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{r}\left(Y_{ijk}-\bar{Y}\right)\left(Y_{ijk}-\bar{Y}\right)'$

Tabela de MANOVA

F.V.	Estatística Multivariada	Distribuição (Bartlet)
Interação	$\Lambda^*_{Int} = \dfrac{\lvert E \rvert}{\lvert HInt + E \rvert}$	$-\left(ab(r-1) - \dfrac{p+1-(a-1)(b-1)}{2} \right) \ln \Lambda^*_{Int} \sim \chi^2_{(a-1)(b-1)p}$
Fator 1	$\Lambda^*_{F1} = \dfrac{\lvert E \rvert}{\lvert HF1 + E \rvert}$	$-\left(ab(r-1) - \dfrac{p+1-(a-1)}{2} \right) \ln \Lambda^*_{F1} \sim \chi^2_{(a-1)p}$
Fator 2	$\Lambda^*_{F2} = \dfrac{\lvert E \rvert}{\lvert HF2 + E \rvert}$	$-\left(ab(r-1) - \dfrac{p+1-(b-1)}{2} \right) \ln \Lambda^*_{F2} \sim \chi^2_{(b-1)p}$

Testar a interação com os efeitos principais no modelo!

Testar os efeitos principais somente sob *inexistência de interação* (modelo aditivo)!

Delineamento Fatorial

Dados O2:

DCA Fatorial 2x2

```
               Df  Pillai approx F num Df den Df    Pr(>F)
Grup            1 0.36121   7.9165      3     42 0.0002686 ***
Staph           1 0.36152   7.9270      3     42 0.0002660 ***
Grup:Staph      1 0.10131   1.5782      3     42 0.2088765
Residuals      44

Conclusão: Não há evidência para efeito de interação significante
Modelo Reduzido (sem a interação)

              Df  Pillai approx F num Df den Df    Pr(>F)
Grup           1 0.33988   7.3798      3     43 0.0004280 ***
Staph          1 0.36086   8.0927      3     43 0.0002194 ***
Residuals     45
```

```
  Grup    X6       X12      X18
1    P 1.470 2.432083 3.475833
2    V 1.525 2.684583 3.852917
```

```
  Staph       X6      X12      X18
1     0 1.357917 2.500000 3.550833
2     1 1.637083 2.616667 3.777917
```

Realize comparações múltiplas para estudar os efeitos principais dos fatores

Delineamento Aleatorizado em Blocos Completos

Considere os dados de fabricação de um filme plástico: três variáveis (Y1, Y2 e Y3) foram observadas sob dois níveis (baixo e alto) de regulação das Máquinas Maq1 e Maq2. Os materiais de filme plástico estão blocados de acordo com o fornecedor.

Maq1	Baixo						Alto					
Maq2	Baixo			Alto			Baixo			Alto		
Bloco	Y1	Y2	Y3	Y1	Y2	Y3	Y1	Y2	Y3	Y1	Y2	Y3
1	6,5	9,5	4,4	6,9	9,1	5,7	6,7	9,1	2,8	7,1	9,2	8,4
2	6,2	9,9	6,4	7,2	10	2	6,6	9,3	4,1	7	8,8	5,2
3	5,8	9,6	3	6,9	9,9	3,9	7,2	8,3	3,8	7,2	9,7	6,9
4	6,5	9,6	4,1	6,1	9,5	1,9	7,1	8,4	1,6	7,5	10,1	2,7
5	6,5	9,2	0,8	6,3	9,4	5,7	6,8	8,5	3,4	7,6	9,2	1,9

Estrutura dos tratamentos: Fatorial 2x2

Estrutura de aleatorização das unidades amostrais aos tratamentos é restrita a **Blocos**.

⇒ Considere que as unidades amostrais (total de 20) estão

Blocadas, de tal forma que há 5 blocos de 4 observações (homogêneas).

Dentro de cada bloco os 4 tratamentos foram aleatorizados às observações.

Note que NÃO há replicas dentro dos níveis do fator Bloco.

Delineamento Aleatorizado em Blocos Completos - ANOVA

Ef. principal de Trat

Ef. de Bloco

$$y_{gk} = \mu + \tau_g + \beta_k + e_{gk}$$

Resposta da observação avaliada no nível g do Fator de interesse e no nível k do Fator Bloco (não há réplica)

Restrições de identificabilidade dos parâmetros

$$\sum_{g=1}^{G} \tau_g = 0, \sum_{k=1}^{b} \beta_k = 0$$

"Identidade útil" para obtenção das Somas de Quadrados e dos estimadores dos efeitos de interesse:

Ef. principal do Fator de interesse

Efeito do fator Bloco

Resíduo: é o ef. de interação entre o fator de interesse e Bloco

$$y_{gk} = \bar{y} + (\bar{y}_{g.} - \bar{y}) + (\bar{y}_{.k} - \bar{y}) + (y_{gk} - \bar{y}_{g.} - \bar{y}_{.k} + \bar{y})$$

SQ_F1 SQ_F2 SQ_Residual

Caso Multivariado $\Rightarrow$ formilação para o vetor de resposta p-dimensional.

Tabela de MANOVA
Delineamento Aleatorizado em Blocos Completos

F.V.	g.l.	Matriz de SSCP
Fator	**G-1**	$HF1 = \sum_{g=1}^{G} b\left(\overline{Y}_{g.} - \overline{Y}\right)\left(\overline{Y}_{g.} - \overline{Y}\right)'$
Bloco	**b-1**	$HF2 = \sum_{k=1}^{b} G\left(\overline{Y}_{.k} - \overline{Y}\right)\left(\overline{Y}_{.k} - \overline{Y}\right)'$
Resíduo	**(G-1)(b-1)**	$E = \sum_{g=1}^{G}\sum_{k=1}^{b}\left(Y_{gk} - \overline{Y}_{g.} - \overline{Y}_{.k} + \overline{Y}\right)\left(Y_{gk} - \overline{Y}_{g.} - \overline{Y}_{.k} + \overline{Y}\right)'$
TOTAL	**Gb-1**	$HF1 + HF2 + E = \sum_{g=1}^{G}\sum_{k=1}^{b}\left(Y_{gk} - \overline{Y}\right)\left(Y_{gk} - \overline{Y}\right)'$

Delineamento Hierárquico (*Nested*)

Considere o seguinte experimento em que as notas dos alunos foram avaliadas segundo Escola e Método de Ensino (A, B, C e D)

Escola 1				Escola 2			
Método A		Método B		Método C		Método D	
Nota1	Nota2	Nota1	Nota2	Nota1	Nota2	Nota1	Nota2
7.6	8.2	9.2	9.2	5.6	10.0	6.2	8.8
5.9	7.7	4.3	4.3	7.7	6.9	4.9	4.8
…	…	…	…	…	…	…	…
4.8	6.5	9.0	9.0	5.8	7.8	8.8	7.3

Estrutura hierárquica dos fatores

p=2

DCA: atribuição aleatória dos estudantes às salas de aula

Estrutura de Tratamentos: há dois fatores hierárquicos Método(Escola).

O fator Método de Ensino está definido DENTRO do fator Escola.

Delineamento Hierárquico (*Nested*) - ANOVA

Efeitos Fixos dos Fatores

Ef. principal de F1

Ef. de F2(F1)

$$y_{ijk} = \mu + \tau_j + \beta_{k(j)} + e_{ijk}$$

Resposta da observação i avaliada no nível **k do Fator 2 dentro do nível j do Fator 1**

Restrições de identificabilidade dos parâmetros

$$\sum_{j=1}^{a} \tau_j = 0, \sum_{k=1}^{b} \beta_{k(j)} = 0$$

"Identidade útil" para obtenção das Somas de Quadrados e dos estimadores dos efeitos de interesse:

Ef. principal de F1

Ef. de F2 dentro de F1: é a soma do ef. de F2 e da interação

Resíduo

$$y_{ijk} = \bar{y} + (\bar{y}_{j.} - \bar{y}) + (\bar{y}_{jk} - \bar{y}_{j.}) + (y_{ijk} - \bar{y}_{jk})$$

SQ_F1 SQ_F2(F1) SQ_Residual

Caso Multivariado $\Rightarrow$ descrever os resultados para o vetor de resposta p-dimensional.

Tabela de MANOVA
Delineamento Hierárquico (“Nested”)

Fonte de Variação	Número de g.l.	Matriz de SQPC
F1	**a-1**	$H_{F1_{p\times p}} = \sum_{j=1}^{a} a\left(\bar{Y}_{j} - \bar{Y}\right)\left(\bar{Y}_{j} - \bar{Y}\right)'$
F2(F1)	**a(b-1)**	$H_{F2(F1)_{p\times p}} = \sum_{j=1}^{4}\sum_{k=1}^{4} r\left(\bar{Y}_{jk} - \bar{Y}_{j}\right)\left(\bar{Y}_{jk} - \bar{Y}_{j}\right)'$
Resíduo	**ab(r-1)**	$E_{p\times p} = \sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{r}\left(Y_{ijk} - \bar{Y}_{jk}\right)\left(Y_{ijk} - \bar{Y}_{jk}\right)'$
Total	**abr-1**	$\sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{r}\left(Y_{ijk} - \bar{Y}\right)\left(Y_{ijk} - \bar{Y}\right)'$

Modelos MANOVA

Pense nas possíveis decomposições da matriz de observações Y_{nxp}

Decomposições (Identidades) úteis para a construção das SQPC

Modelo de um único fator: $y_{ig} = \bar{y} + \boxed{\left(\bar{y}_g - \bar{y}\right)} + \left(y_{ig} - \bar{y}_g\right)$ **T=H+E**

DCA

Fatorial Cruzado: $y_{ijk} = \bar{y} + \left(\bar{y}_{j.} - \bar{y}\right) + \boxed{\left(\bar{y}_{.k} - \bar{y}\right) + \left(\bar{y}_{jk} - \bar{y}_{j.} - \bar{y}_{.k} + \bar{y}\right)} + \left(y_{ijk} - \bar{y}_{jk}\right)$

T=H1+H2+H1*H2+E

O efeito de F2 dentro de F1 é a soma do efeito principal de F1 e do efeito de interação

Fatorial Hierárquico: $y_{ijk} = \bar{y} + \left(\bar{y}_{j.} - \bar{y}\right) + \boxed{\left(\bar{y}_{jk} - \bar{y}_{j.}\right)} + \left(y_{ijk} - \bar{y}_{jk}\right)$ **T=H1+H2(H1)+E**

O ef. de interação entre Bloco (F2) e F1 é o resíduo (**modelo aditivo**)

DABC

Modelo com fator Bloco: $y_{jk} = \bar{y} + \boxed{(\bar{y}_{j.} - \bar{y})} + (\bar{y}_{.k} - \bar{y}) + \boxed{(y_{jk} - \bar{y}_{j.} - \bar{y}_{.k} + \bar{y})}$

T=H1+Bloco+E

Modelos MANOVA
Decomposição da Matriz Y_{nxp}

ASCA: ANOVA-Simultaneous Component Analysis
(Smilde et al., 2005)

Modelo de um único fator: (p=1) $y_{ig} = \bar{y} + \left(\bar{y}_g - \bar{y}\right) + \left(y_{ig} - \bar{y}_g\right)$

⬇

Para p>1: $Y_{ig\ p\times 1} = \bar{Y}_{p\times 1} + \left(\bar{Y}_g - \bar{Y}\right)_{p\times 1} + \left(Y_{ig} - \bar{Y}_g\right)_{p\times 1}$

⬇

$$Y_{n\times p};\quad n = \sum_{g=1}^{G} n_g$$

Decomposição de Y devido à estrutura de grupos.

$$Y_{n\times p} = M_{n\times p} + T_{n\times p} + E_{n\times p}$$

Matriz de Médias	Componente da variabilidade ENTRE grupos	Componente da variabilidade DENTRO de grupos

Exemplo

Duas variáveis avaliadas em unidades amostrais submetidas a 3 tratamentos

T1		T2		T3	
Y11	**Y12**	**Y21**	**Y22**	**Y31**	**Y32**
9	3	0	4	3	8
6	2	2	0	1	9
9	7			2	7
8	4	1	2	2	8

Média geral = (4 , 5)

$$Y_{8\times 2} = M_{8\times 2} + T_{8\times 2} + E_{8\times 2}$$

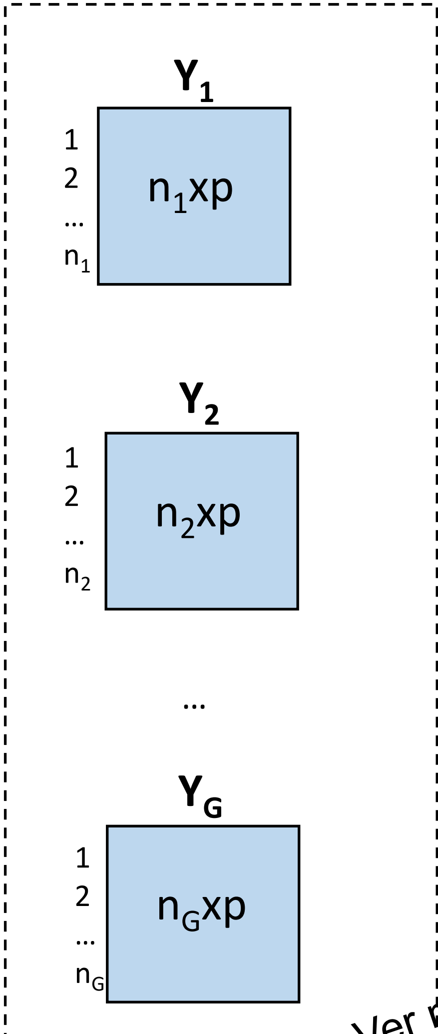
$$\begin{pmatrix} 9 & 3 \\ 6 & 2 \\ 9 & 7 \\ 0 & 4 \\ 2 & 0 \\ 3 & 8 \\ 1 & 9 \\ 2 & 7 \end{pmatrix} = \begin{pmatrix} 4 & 5 \\ 4 & 5 \\ 4 & 5 \\ 4 & 5 \\ 4 & 5 \\ 4 & 5 \\ 4 & 5 \\ 4 & 5 \end{pmatrix} + \begin{pmatrix} 4 & 3 \\ 4 & 2 \\ 4 & 7 \\ -3 & 4 \\ -3 & 0 \\ -2 & 3 \\ -2 & 3 \\ -2 & 3 \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ -2 & -2 \\ 1 & 3 \\ -1 & 2 \\ 1 & -2 \\ 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix}$$

Dependendo do problema, pode haver interesse na análise multivariada, não de Y mas do componente T de Y, ou do componente E de Y (E: resíduo ou resposta Y normalizada)

Pense, por ex., no componente T definindo ***Multicentros*** do estudo. Pode haver interesse no efeito de Centros nos dados (usar T) ou mesmo em eliminar este efeito dos dados (usar E)!

P-Integração de Bancos de Dados

Dados de G Centros



$$Y_{n \times p}; \quad n = \sum_{g=1}^{G} n_g$$

$$Y_{n \times p} = M_{n \times p} + \boxed{T_{n \times p}} + \boxed{E_{n \times p}}$$

Realizar análises (redução de dimensionalidade) nos componentes da decomposição de Y

Ex.: Obter "componentes principais" de T (ou mesmo de E)

Ver mix-Omics_R