

ANÁLISE DA VARIÂNCIA



1

PLANEAMENTO EXPERIMENTAL



- Seleção dos fatores e identificação dos parâmetros que são objeto do estudo;
- Decisão sobre a magnitude dos erros padrão pretendidos;
- Escolha dos tratamentos (combinações de níveis de fatores) a serem incluídos na experiência, bem como o número de observações em cada tratamento;
- Atribuição dos tratamentos às unidades experimentais.



ANÁLISE DA VARIÂNCIA

O objetivo da Análise da Variância é isolar e avaliar as fontes de variação associadas com as variáveis experimentais, independentes e determinar como estas variáveis interatuam e afetam a variável resposta.

Nota histórica: Foi Sir Ronald Fisher quem desenvolveu esta técnica e a aplicou ao planeamento das experiências. Os seus livros “*Statistical Methods for Research Workers*”, editado em 1925 e “*The Design of Experiments*”, editado em 1935, são considerados clássicos na literatura.

Na Análise da Variância, a variação nas medidas observadas (resposta) é particionada em componentes que refletem os efeitos de uma ou mais variáveis independentes.

ANOVA – Analysis of Variance

Profª Ana Cristina Braga, DPS

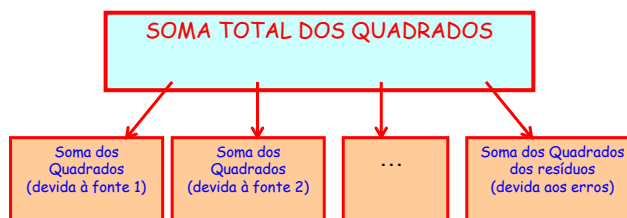
3



Se o conjunto de dados consiste em n resultados y_1, y_2, \dots, y_n e se a média é \bar{y} , a variação total das observações em relação à média, soma dos quadrados das variações, é:

$$STQ = \sum_{i=1}^n (y_i - \bar{y})^2$$

e designa-se por soma total dos quadrados (STQ) das variações.



O número de fontes de variação e as fórmulas para as componentes estão relacionadas como tipo de *planeamento* escolhido e com o modelo estatístico mais apropriado para a análise.

Profª Ana Cristina Braga, DPS

4



PLANEAMENTO COMPLETAMENTE CASUAL (PCA)

Tratamento	Observações					
	1	2	...	n	Total	Média
1	y_{11}	y_{12}	\cdots	y_{1n}	$T_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	\cdots	y_{2n}	$T_{2.}$	$\bar{y}_{2.}$
\vdots	\vdots	\vdots	\cdots	\vdots	\vdots	\vdots
k	y_{k1}	y_{k2}	\cdots	y_{kn}	$T_{k.}$	$\bar{y}_{k.}$

Profª Ana Cristina Braga, DPS

5



PARTIÇÃO DA SOMA DOS QUADRADOS

$$(y_{ij} - \bar{y}_{..}) = (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})$$
$$\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^k (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2$$
$$STQ = SQT + SQR$$

- STQ – Soma Total dos Quadrados
- SQT – Soma dos Quadrados dos Tratamentos
- SQR – Soma dos Quadrados dos Resíduos

Profª Ana Cristina Braga, DPS

6



PLANEAMENTO COMPLETAMENTE CASUAL (PCA)

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, k \\ j = 1, 2, \dots, n \end{cases}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$H_{01} : \alpha_i = 0 \quad i = 1, 2, \dots, k$$

$$H_{11} : \alpha_i \neq 0 \quad \text{para pelo menos um valor de } i$$

$$R.R : F > c$$

Profª Ana Cristina Braga, DPS

7



TABELA ANOVA

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Média dos Quadrados	F
Tratamentos	SQT	k-1	MQT	F=MQT/MQR
Resíduos	SQR	k(n-1)	MQR	
Total	STQ	kn-1		

Profª Ana Cristina Braga, DPS

8



EXEMPLO

Um estudo realizado para avaliar o desenvolvimento de moscas consistiu na sua criação em três meios de cultura diferentes. A tabela apresenta o comprimento ($\text{mm} \times 10^{-1}$) das asas de 5 moscas recolhidas aleatoriamente de cada meio. Verifique se existem diferenças entre os comprimentos das asas das moscas recolhidas de cada meio.

Profª Ana Cristina Braga, DPS

9



EXEMPLO

Meio 1	36	39	43	38	37
Meio 2	50	42	51	40	43
Meio 3	45	53	56	52	56

Profª Ana Cristina Braga, DPS

10



Resolução:

	Meio 1	Meio 2	Meio 3
	36	50	45
	39	42	53
	43	51	56
	38	40	52
	37	43	56
totais	T1. =193	T2. = 226	T3. =262
	T..= 681		

$$\sum_{i,j} y_{ij}^2 = 31603$$

$$SQT = \frac{1}{5} (193^2 + 226^2 + 262^2) - \frac{1}{15} 681^2 = 476,4$$

$$STQ = 31603 - 30917,4 = 685,6$$

$$SQR = 685,6 - 476,4 = 209,2$$

Profª Ana Cristina Braga, DPS

11



TABELA ANOVA

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Média dos Quadrados	F
Tratamentos	476,4	2	238,2	F=12
Resíduos	209,2	12	19,85	
Total	685,6	14		

$$F_{2,12,0.05} = 3,89$$

Decisão: Como $F > c$, rejeita-se a H_0 para um nível de significância de 5%, pelo que existem diferenças estatisticamente significativas entre os valores médios de crescimento nos 3 meios.

Profª Ana Cristina Braga, DPS

12



Exemplo (Amostras desequilibradas)

Quatro grupos de vendedores foram sujeitos a diferentes programas de treino. Durante o programa de treino houve algumas desistências. No fim dos programas, a cada vendedor foi atribuída uma área de venda. A tabela regista as vendas ao fim de uma semana. Considere $\alpha=0,05$.

Grupo de treino			
G1	G2	G3	G4
65	75	59	94
87	69	78	89
73	83	67	80
79	81	62	88
81	72	83	
69	79	76	
	90		

Profª Ana Cristina Braga, DPS

13



Grupo de treino				
G1	G2	G3	G4	
65	75	59	94	
87	69	78	89	
73	83	67	80	
79	81	62	88	
81	72	83		
69	79	76		
	90			
Totais	454	549	425	351
				T.. = 1779
nj	6	7	6	4

$$\sum_{i,j} y_{ij}^2 = 139511$$

Profª Ana Cristina Braga, DPS

14



H_0 : Não existem diferenças significativas nas vendas devido aos diferentes programas de treino

$$\mu_1 = \mu_2 = \mu_3 = \mu_4 \quad \text{ou} \quad \alpha_j = 0 \text{ com } j = 1, 2, 3, 4$$

H_1 : Pelo menos 2 programas são diferentes

$\alpha_j \neq 0$ para pelo menos um valor de j .

R.R: $F > c$

$$SQT = \left(\frac{454^2}{6} + \frac{549^2}{7} + \frac{425^2}{6} + \frac{351^2}{4} \right) - \frac{1}{23} 1779^2 = 712,6$$

$$STQ = 139511 - \frac{1}{23} 1779^2 = 1909,2$$

$$SQR = 1909,2 - 712,6 = 1196,6$$

Profª Ana Cristina Braga, DPS

15



TABELA ANOVA

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Média dos Quadrados	F
Tratamentos	712,6	3	237,5	F=3,77
Resíduos	1196,6	19	62,97	
Total	1909,2	22		

$$F_{3,19,0.05} = 3,13$$

Decisão: Como $F > c$, rejeita-se a H_0 para um nível de significância de 5%, pelo que existem diferenças estatisticamente significativas entre os valores médios das vendas nos 4 grupos de treino.

Profª Ana Cristina Braga, DPS

16

Intervalos de confiança para as comparações múltiplas



$$T = \frac{(\bar{y}_i - \bar{y}_j) - (\mu_i - \mu_j)}{\sqrt{MQR \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \sim t_{N-k}$$

$$(\bar{y}_i - \bar{y}_j) - t_{(\alpha/2), N-k} * \sqrt{MQR \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} < \mu_i - \mu_j < (\bar{y}_i - \bar{y}_j) + t_{(\alpha/2), N-k} * \sqrt{MQR \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

$$1.36 \leq \mu_4 - \mu_1 \leq 22.81^* \quad -14.42 \leq \mu_3 - \mu_1 \leq 4.76 \quad -1.09 \leq \mu_4 - \mu_2 \leq 19.73$$

$$-16.84 \leq \mu_3 - \mu_2 \leq 1.65 \quad 6.19 \leq \mu_4 - \mu_3 \leq 27.64^* \quad -6.48 \leq \mu_2 - \mu_1 \leq 12.00$$

Profª Ana Cristina Braga, DPS

17

***ANOVA3 [DataSet0] - SPSS Data Editor**

File Edit View Data Transform Analyze Graphs Utilities Window Help

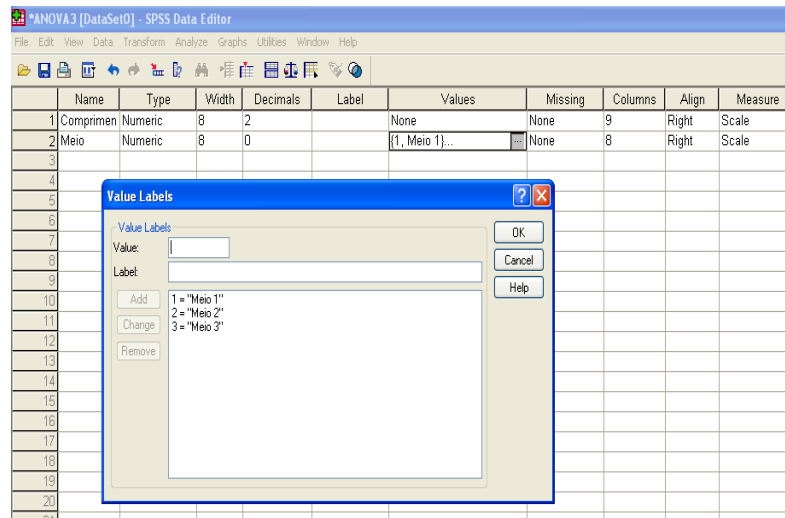
10 :

	Comprimento	Meio	var	var	var	var
1	36,00	Meio 1				
2	39,00	Meio 1				
3	43,00	Meio 1				
4	38,00	Meio 1				
5	37,00	Meio 1				
6	50,00	Meio 2				
7	42,00	Meio 2				
8	51,00	Meio 2				
9	40,00	Meio 2				
10	43,00	Meio 2				
11	45,00	Meio 3				
12	53,00	Meio 3				
13	56,00	Meio 3				
14	52,00	Meio 3				
15	56,00	Meio 3				
16						



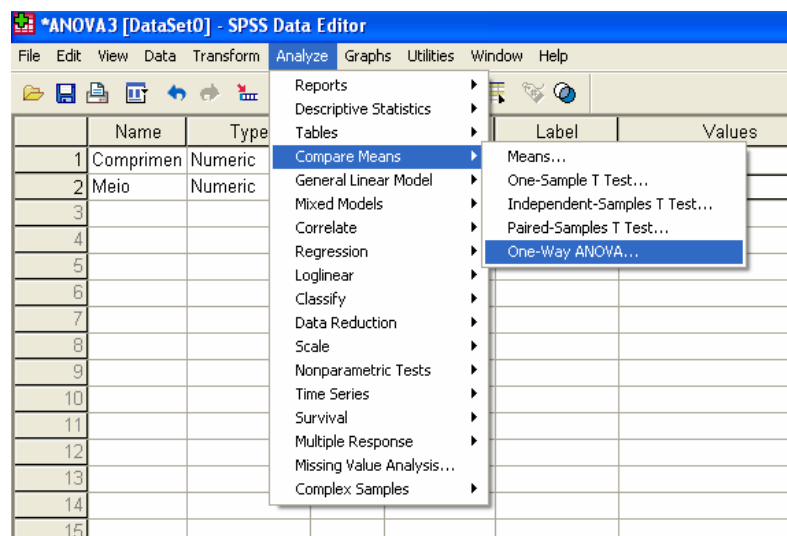
Profª Ana Cristina Braga, DPS

18



Profª Ana Cristina Braga, DPS

19



Profª Ana Cristina Braga, DPS

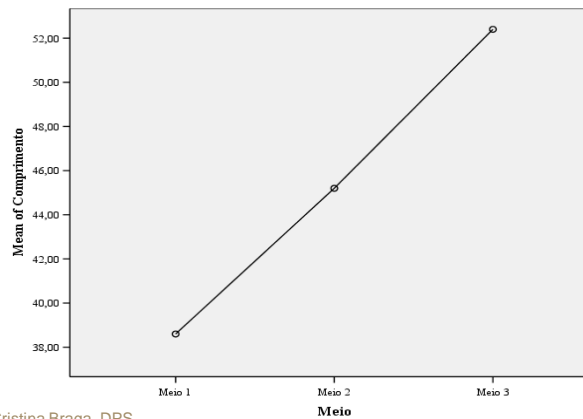
20

TABELA ANOVA



ANOVA

Comprimido	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	476,400	2	238,200	13,663	,001
Within Groups	209,200	12	17,433		
Total	685,600	14			



Profª Ana Cristina Braga, DPS

21

PLANEAMENTO COM BLOCOS ALEATÓRIOS (PBA)



Permite comparar k tratamentos envolvendo n blocos, cada contendo k unidades experimentais relativamente homogêneas. Os k tratamentos são distribuídos aleatoriamente às unidades experimentais dentro de cada bloco, com uma unidade experimental por tratamento.

Profª Ana Cristina Braga, DPS

22



PLANEAMENTO COM BLOCOS ALEATÓRIOS (PBA)

Tratamento	Blocos					
	1	2	...	n	Total	Média
1	y_{11}	y_{12}	...	y_{1n}	$T_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	...	y_{2n}	$T_{2.}$	$\bar{y}_{2.}$
...
k	y_{k1}	y_{k2}	...	y_{kn}	$T_{k.}$	$\bar{y}_{k.}$
	T.1	T.2	...	T.n		

Profª Ana Cristina Braga, DPS

23



PLANEAMENTO COM BLOCOS ALEATÓRIOS (PBA)

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, k \\ j = 1, 2, \dots, b \end{cases} \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

$$H_{01} : \alpha_i = 0 \quad i = 1, 2, \dots, k$$

$$H_{11} : \alpha_i \neq 0 \quad \text{para pelo menos um valor de } i$$

$$R.R : F_1 > c_1$$

$$H_{02} : \beta_j = 0 \quad j = 1, 2, \dots, b$$

$$H_{12} : \beta_j \neq 0 \quad \text{para pelo menos um valor de } j$$

$$R.R : F_2 > c_2$$

Profª Ana Cristina Braga, DPS

24

PLANEAMENTO COM BLOCOS ALEATÓRIOS (PBA)



$$SQT = b \sum_{j=1}^k (\bar{y}_{.j} - \bar{Y})^2$$
$$SQB = k \sum_{i=1}^b (\bar{y}_{i.} - \bar{Y})^2$$
$$STQ = \sum_{i=1}^b \sum_{j=1}^k (y_{ij} - \bar{Y})^2$$
$$SQR = \sum_{i=1}^b \sum_{j=1}^k (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{Y})^2$$

$$SQT = \frac{1}{b} \sum_{j=1}^k T_{.j}^2 - \frac{1}{kb} T_{..}^2$$
$$SQB = \frac{1}{k} \sum_{i=1}^b T_{i.}^2 - \frac{1}{kb} T_{..}^2$$
$$STQ = \sum_{i=1}^b \sum_{j=1}^k y_{ij}^2 - \frac{1}{kb} T_{..}^2$$
$$SQR = STQ - SQT - SQB$$

$T_{i.}$ é o total dos valores obtidos para o bloco i ; $T_{.j}$ é o total dos valores obtidos para o tratamento j

TABELA ANOVA



Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Média dos Quadrados	F
Tratamentos	SQT	k-1	MQT	F ₁ =MQT/MQR F ₂ =MQB/MQR
Blocos	SQB	b-1	MQB	
Resíduos	SQR	(k-1)(b-1)	MQR	
Total	STQ	kb-1		



Exemplo: Considere o tempo (em minutos) que levou uma certa pessoa a conduzir de casa até ao emprego, de segunda a sexta, por 4 caminhos diferentes.

dias	Seg.	Ter.	Qua.	Qui.	Sex.	
Caminho 1	22	26	25	25	31	$T_1 = 129$
Caminho 2	25	27	28	26	29	$T_2 = 135$
Caminho 3	26	29	33	30	33	$T_3 = 151$
Caminho 4	26	28	27	30	30	$T_4 = 141$

$$T_1 = 99 \quad T_2 = 110 \quad T_3 = 113 \quad T_4 = 111 \quad T_5 = 123 \quad T = 556$$

Comparar os tempos de percurso para o emprego, considerando $\alpha = 0.05$.

Resolução:

Trata-se de um planeamento com blocos aleatórios (dias da semana), cujo modelo é:

$$y_{ij} = \mu_{ij} + e_{ij} = \mu + \alpha_j + \beta_i + e_{ij}$$

H_{01} : Não existem diferenças significativas nos tempos devido aos diferentes caminhos

$$\mu_1 = \mu_2 = \mu_3 = \mu_4 \quad \text{ou} \quad \alpha_j = 0 \quad \text{com } j = 1, 2, 3, 4$$

H_{02} : Não existem diferenças significativas nos tempos devido aos diferentes dias da semana

$$\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 \quad \text{ou} \quad \beta_i = 0 \quad \text{com } i = 1, 2, 3, 4, 5$$

H_{11} : $\alpha_j \neq 0$ para pelo menos um valor de j .

Profª Ana Cristina Braga, DPS

27



H_{12} : $\beta_i \neq 0$ para pelo menos um valor de i .

$$\sum y_{ij}^2 = 15610 \quad n = 5 \quad k = 4 \quad STQ = 153.2 \quad SQT = 52.8 \quad SQB = 73.2 \quad SQR = 27.2$$

Tabela ANOVA

Fonte de variação	Soma dos Quadrados	Graus de liberdade	Média dos Quadrados	Estatística de teste, F
Tratamentos	52.8	3	17.6	$F_1 = 7.75$
Blocos	73.2	4	18.3	
Resíduos	27.2	12	2.27	$F_2 = 8.06$
Total	153.2	19		

Decisão: Como $F_1 = 7.75 > F_{3, 12, (0.05)} = 3.49$ e $F_2 = 8.06 > F_{4, 12, (0.05)} = 3.26$, rejeitam-se ambas hipóteses nulas para um nível de significância 0.05, pelo que existem diferenças significativas nos tempos de percurso, quer devido aos diferentes caminhos quer devido aos diferentes dias da semana.

Profª Ana Cristina Braga, DPS

28