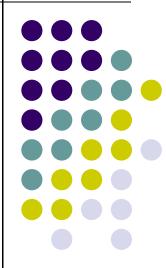
ANÁLISE DA VARIÂNCIA

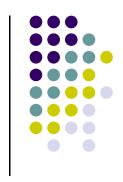




PLANEAMENTO EXPERIMENTAL

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

ANÁLISE DA VARIÂNCIA



O objectivo da Análise da Variância é isolar e avaliar as fontes de variação associadas com as variáveis experimentais, independentes e determinar como estas variáveis interactuam e afectam 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 reflectem os efeitos de uma ou mais variáveis independentes.

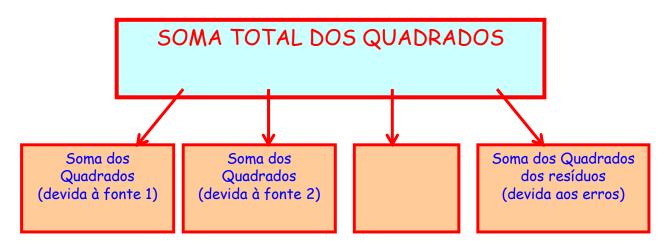
ANOVA – <u>An</u>alysis <u>of Va</u>riance

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



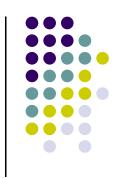
$$STQ = \sum_{i=1}^{n} (y_i - \overline{y})^2$$

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



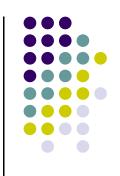
O número de fontes de variação e as fórmulas para as componentes estão relacionadas como tipo de *planeamento* escolhido e com o modelo estatístico mais apropriado para a análise.

PLANEAMENTO COMPLETAMENTE CASUAL



Tratamento		Observ				
Tratamento	1	2		n	Total	Média
1	y_{11}	\mathcal{Y}_{12}	•••	\mathcal{Y}_{1n}	$T_{1.}$	$\overline{\mathcal{Y}}_{1.}$
2	y_{21}	<i>y</i> ₂₂	• • •	\mathcal{Y}_{2n}	<i>T</i> _{2.}	\overline{y}_2
:	:	•	• • •	:	•	•
k	\mathcal{Y}_{k1}	y_{k2}	• • •	\mathcal{Y}_{kn}	$T_{k.}$	$\overline{\mathcal{Y}}_{k.}$

PARTIÇÃO DA SOMA DOS QUADRADOS



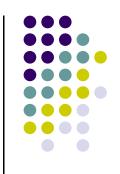
$$(y_{ij} - \overline{y}_{..}) = (\overline{y}_{i.} - \overline{y}_{..}) + (y_{ij} - \overline{y}_{i.})$$

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

$$STQ = SQT + SQR$$

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

PLANEAMENTO COMPLETAMENTE CASUAL

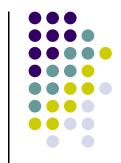


$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \begin{cases} i = 1, 2, ..., k \\ j = 1, 2, ..., n \end{cases}$$

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

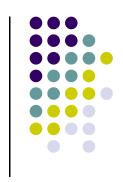
$$H_{01}: \alpha_i = 0 \quad i = 1, 2, ..., k$$

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

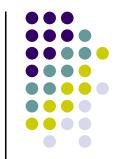


Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Média dos Quadrados	F
Tratamentos	SQT	k-1	MQT	
Resíduos	SQR	k(n-1)	MQR	F=MQT/MQR
Total	STQ	kn-1		





 Um estudo realizado para estudar o desenvolvimento de moscas consistiu na sua criação em três meio de cultura diferentes. A tabela apresenta o comprimento (mm×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.



EXEMPLO 3

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



Resolução

totais

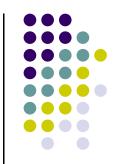
Ī						
	Meio 1	Meio 2	Meio 3			
	36	50	45			
	39	42	53			
	43	51	56			
	38	40	52			
	37	43	56			
Į	T1. =193	T2.= 226	T3.=262			
	T= 681					

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

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

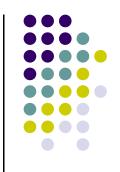
$$STQ = 31603 - 30917,4 = 685,6$$

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



Fonte de Variação	Soma dos Quadrado s	Graus de Liberdade	Média dos Quadrad os	F
Tratamentos	476,4	2	238,2	F=12
Resíduos	209,2	12	19,85	Γ-12
Total	685,6	14	$F_{2,12,0.05}$ =	= 3,89

Decisão: Como F> c, rejeita-se a H0 para um nível de significância de 5%, pelo que existem diferenças estatísticamente significativas entre os valores médios de crescimento nos 3 meios.



Exemplo 2 (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 α=0,05.

	Grupo	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				



Grupo de treino

	G1	G2	G3	G4		
	65	75	59	94		
	87	69	78	89		$\sum_{ij} y_{ij}^2 = 139511$
	73	83	67	80		$\overline{i,j}$
	79	81	62	88		
	81	72	83			
	69	79	76			
		90				
Totais	454	549	425	351	T = 1779	
nj	6	7	6	4		

H₀: Não existem diferenças significativas nas vendas devido aos diferentes programas de treino



$$\mu_1 = \mu_2 = \mu_3 = \mu_4$$

ou

$$\alpha j = 0 \text{ com } j = 1,2, 3, 4$$

H₁: Pelo menos 2 programas são diferentes

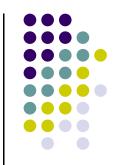
 $\alpha j \neq 0$ para pelo menos um valor de j.

R.R: F> c

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

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

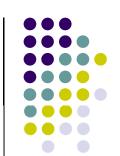
$$SQR = 1909, 2 - 712, 6 = 1196, 6$$



Fonte de Variação	Soma dos Quadrado s	Graus de Liberdade	Média dos Quadrad os	F
Tratamentos	712,6	3	237,5	□ -2 77
Resíduos	1196,6	19	62,97	F=3,77
Total	1909,2	22	$F_{3,19,0.05}$ =	= 3,13

Decisão: Como F> c, rejeita-se a H0 para um nível de significância de 5%, pelo que existem diferenças estatísticamente significativas entre os valores médios das vendas nos 4 grupos de treino.

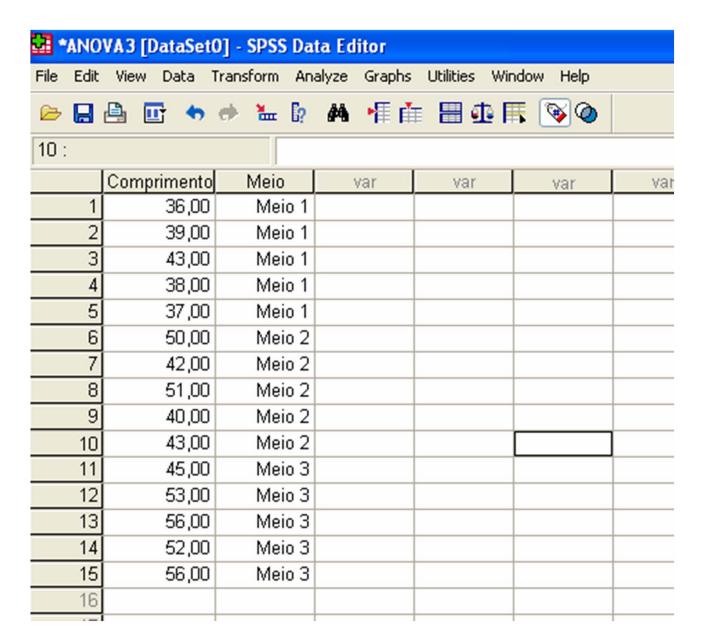
Intervalos de confiança para as médias



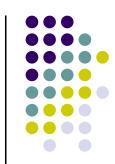
$$T = \frac{(\overline{y}_{.i} - \overline{y}_{.j}) - (\mu_i - \mu_j)}{\sqrt{\frac{SQR}{N-k}(\frac{1}{n_i} + \frac{1}{n_j})}} \sim t_{N-k}$$

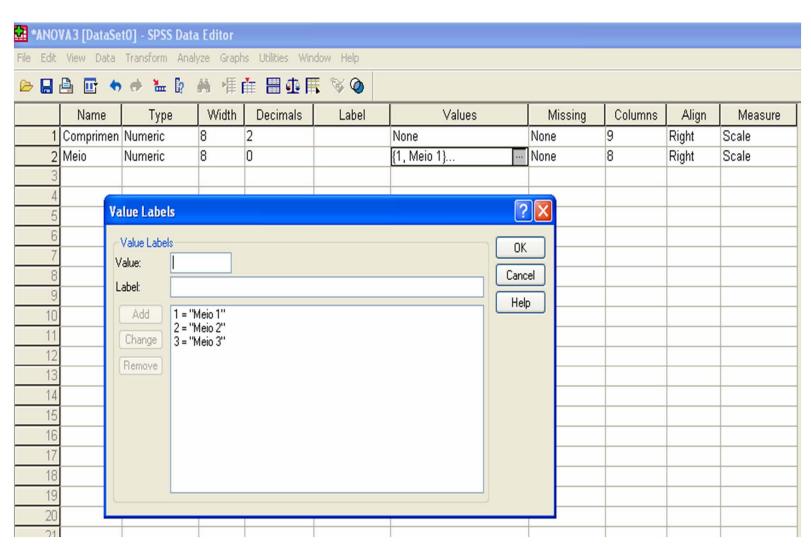
$$(\overline{y}_A - \overline{y}_B) - t_{(q_2),n-k}.s.\sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \le \mu_A - \mu_B \le (\overline{y}_A - \overline{y}_B) + t_{(q_2),n-k}.s.\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$$

$$1.36 \le \mu_4 - \mu_1 \le 22.81^*$$
 $-14.42 \le \mu_3 - \mu_1 \le 4.76$ $-1.09 \le \mu_4 - \mu_2 \le 19.73$ $-16.84 \le \mu_3 - \mu_2 \le 1.65$ $6.19 \le \mu_4 - \mu_3 \le 27.64^*$ $-6.48 \le \mu_2 - \mu_1 \le 12.00$

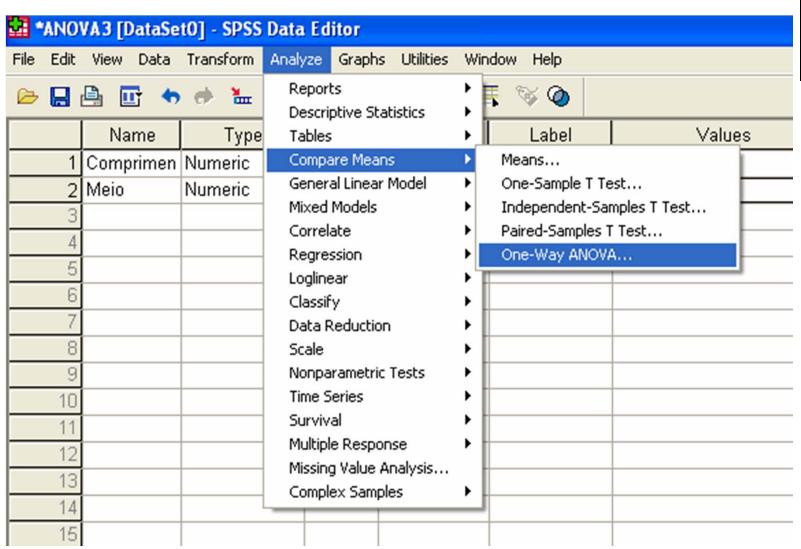








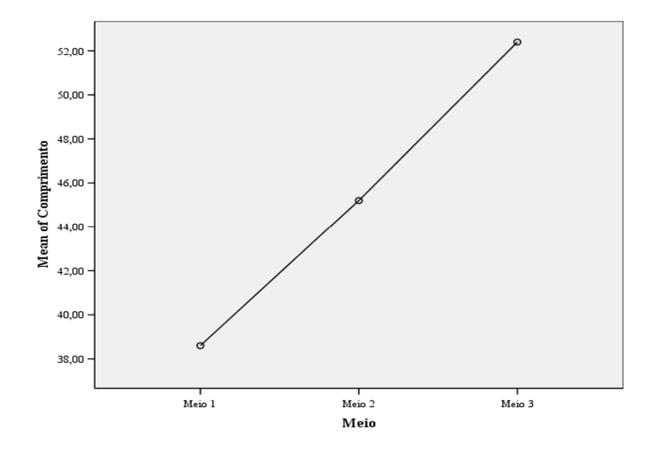




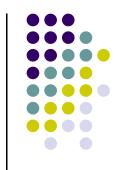
ANOVA

Comprimento

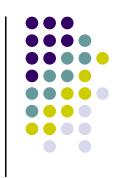
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	476,400	2	238,200	13,663	,001
Within Groups	209,200	12	17,433		
Total	685,600	14			





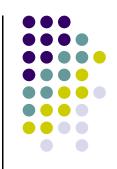


Permite comparar k tratamentos envolvendo n blocos, cada contendo k unidades experimentais relativamente homogéneas. Os k tratamentos são distribuídos aleatoriamente às unidades experimentais dentro de cada bloco, com uma unidade experimental por tratamento.



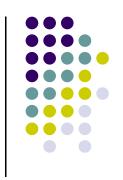
Tratamento						
Tratamonto	1	2		n	Total	Média
1	<i>y</i> ₁₁	\mathcal{Y}_{12}	•••	$\mathcal{Y}_{\mathrm{l}n}$	$T_{1.}$	$\overline{\mathcal{Y}}_{1.}$
2	y_{21}	<i>y</i> ₂₂	•••	\mathcal{Y}_{2n}	<i>T</i> _{2.}	\overline{y}_2
:	:	•	• • •	:	•	:
k	\mathcal{Y}_{k1}	\mathcal{Y}_{k2}	• • •	\mathcal{Y}_{kn}	$T_{k.}$	$\overline{\mathcal{Y}}_{k.}$

T.1 T.2 ... T.n



	Tratamento 1	Tratamento 2		Tratamento k
Bloco 1	y_{11}	\mathcal{Y}_{12}		${\cal Y}_{1k}$
Bloco 2	${\cal Y}_{21}$	y_{22}		y_{2k}
÷	÷	÷	÷	÷
Bloco n	\mathcal{Y}_{n1}	\mathcal{Y}_{n2}		${\cal Y}_{nk}$

 $R.R: F_2 > c_2$



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

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

$$H_{01} : \alpha_i = 0 \quad i = 1, 2, ..., a$$

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

$$R.R : F_1 > C_1$$

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

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



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

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

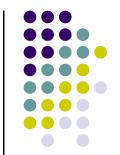
$$SQT = \frac{1}{b} \sum_{j=1}^{k} T_{.j}^{2} - \frac{1}{kb} T_{..}^{2}$$

$$SQB = \frac{1}{k} \sum_{i=1}^{b} T_{i.}^{2} - \frac{1}{kb} T_{..}^{2}$$

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

$$SQR = STQ - SQT - SQB$$

 T_i é o total dos valores obtidos para o bloco i ; $T_{.j}$ é o total dos valores obtidos para o tratamento 1



Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Média dos Quadrado s	F
Tratamentos	SQT	k-1	MQT	F ₁ =MQT/MQR
Blocos	SQB	b-1	MQB	F ₂ =MQB/MQ R
Resíduos	SQR	(k-1)(b-1)	MQR	K
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	T ₂ = 110	T ₃ = 11	3 T ₄ = 11	1 T ₅ =	123 T =556

Comparar os tempos de percurso para o emprego, considerando α = 0.05.

Resolução:

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

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

Hoj: Não existem diferenças significativas nos tempos devido aos diferentes caminhos

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

Hoz: Não existem diferenças significativas nos tempos devido aos diferentes dias da semana

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

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



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

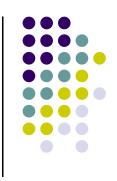
$$\Sigma \Sigma y_{ij}^2 = 15610$$
 n = 5 k = 4 STQ = 153.2 SQT = 52.8 SQB = 73.2 SQR = 27.2

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		

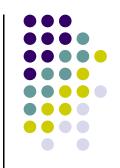
<u>Decisão</u>: Como $F_1 = 7.75 > F_{3, 12, (0.05)} = 3.49$ e $F_2 = 8.06 > F_{4, 12, (0.05)} = 3.26$, rejeitam-se ambas hipóteses nulas para um nível de significância 0.05, pelo que existem diferenças significativas nos tempos de percurso, quer devido aos diferentes caminhos quer devido aos diferentes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



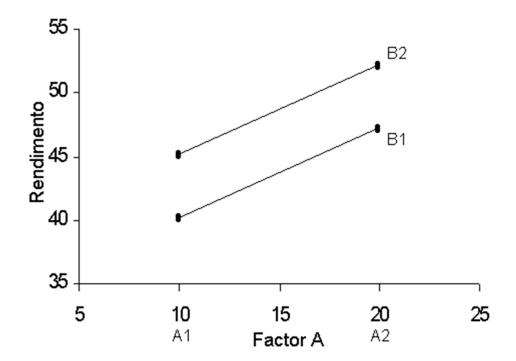


- O Planeamento Factorial é um método de seleccionar os tratamentos (combinações factores-níveis) a serem incluídos numa experiência.
- Uma Experiência Completamente Factorial é aquela em que os tratamentos consistem em todas as combinações de níveis de factores.
- Uma Experiência Factorial envolvendo o FACTOR A com a níveis e o FACTOR B com b níveis é uma experiência factorial axb com t = axb tratamentos.

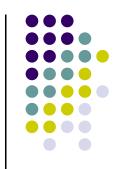




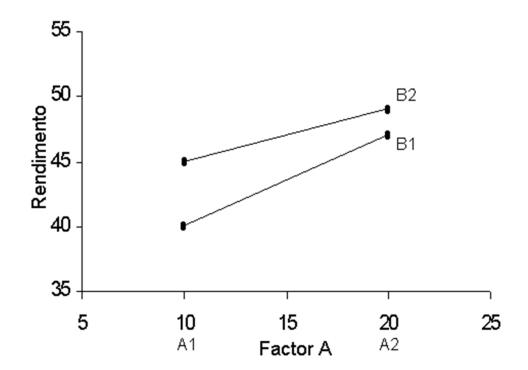
	Factor B Temperatura						
	B1 -150°C B2 -200°C						
Factor A	A1 -10g/l	40	45				
Concentração	A2 -20g/l	47	52				



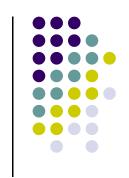




	Factor B Temperatura						
	B1 -150°C B2 -200°C						
Factor A	A1 -10g/l	40	45				
Concentração	A2 -20g/l	47	48				



PLANEAMENTO COM DOIS FACTORES



$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \begin{cases} i = 1, 2, ..., p \\ j = 1, 2, ..., q \end{cases} \qquad \varepsilon_{ijk} \sim N(0, \sigma^2)$$

$$k = 1, 2, ..., r$$

$$H_{01}: \alpha_i = 0 \quad i = 1, 2, ..., p$$

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

$$H_{02}: \beta_j = 0 \quad j = 1, 2, ..., q$$

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

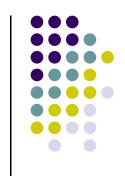
$$H_{02}: \gamma_{ii} = 0 \quad i = 1, 2, ..., p; j = 1, 2, ..., q$$

 $H_{12}: \gamma_{ii} \neq 0$ para pelo menos um par ij



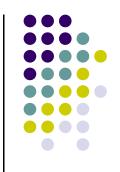
Fonte de Variação			Média dos Quadrados	F
Factor A	SQF _A	p-1	MQF _A	F ₁ =MQF _A /MQR
Factor B	SQF _B	q-1	MQF _B	F ₂ = MQF _B /MQR
Interacção AxB	SQI _{AxB}	(p-1)(q-1)	MQI _{AxB}	F ₃ = MQI _{AxB} /MQR
Resíduos	SQR	pq(r-1)	MQR	
Total	STQ	pqr-1		





 Os dados representam as concentrações de cálcio no plasma (mg/100ml) de pássaros de ambos os sexos, com metade dos pássaros de cada sexo tratado com uma hormona e a outra metade sem hormona. O que pode concluir acerca dos efeitos dos dois factores?





		Factor	B (sexo)			
		Macho Fêmea				
Factor A	Sem	14.5;11.0;10.8;	16.5;18.4;12.7;			
(tratamento		14.3;10.0	14.0;12.8			
hormonal)	Com	32.0;23.8;28.8;	39.1;26.2;21.3;			
		25.0;29.3	35.8;40.2			



	Macho				Fêmea				Ti.		
Sem	14,5	11	10,8	14,3	10	16,5	18,4	12,7	14	12,8	135
Sem	T ₁₁ =60,6			T ₁₂ =74,4							
Com	32	24	28,8	25	29,3	39,1	26,2	21,3	3 35,8	40,2	301,5
Com			T ₂₁ =13	38,9		T ₂₂ =162,6					
T.j	199,5					23	7		436,5		

$$\sum_{i,j,k} y_{ijk}^2 = 11354,31$$

$$SQF_{A} = \frac{\sum_{i=1}^{p} T_{i}^{2}}{rq} - \frac{T^{2}}{pqr}$$

$$SQF_{B} = \frac{\sum_{j=1}^{q} T_{j}^{2}}{rp} - \frac{T^{2}}{pqr}$$

$$SQF_{B} = \frac{\sum_{j=1}^{q} T_{j}^{2}}{rp} - \frac{T^{2}}{pqr}$$

$$SQF_{B} = \frac{1}{5*2} (135^{2} + 301,5^{2}) - \frac{1}{2*2*5} 436,5^{2} = 1386,113$$

$$SQF_{B} = \frac{1}{5*2} (199,5^{2} + 237^{2}) - \frac{1}{2*2*5} 436,5^{2} = 70,313$$

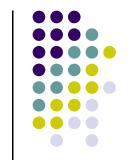
$$SQR = \sum_{ijk} y_{ijk}^{2} - \frac{\sum_{ij} T_{ij}^{2}}{r}$$

$$SQR = 11354,31 - \frac{1}{5} (60,6^{2} + 74,4^{2} + 138,9^{2} + 162,6^{2}) = 366,372$$

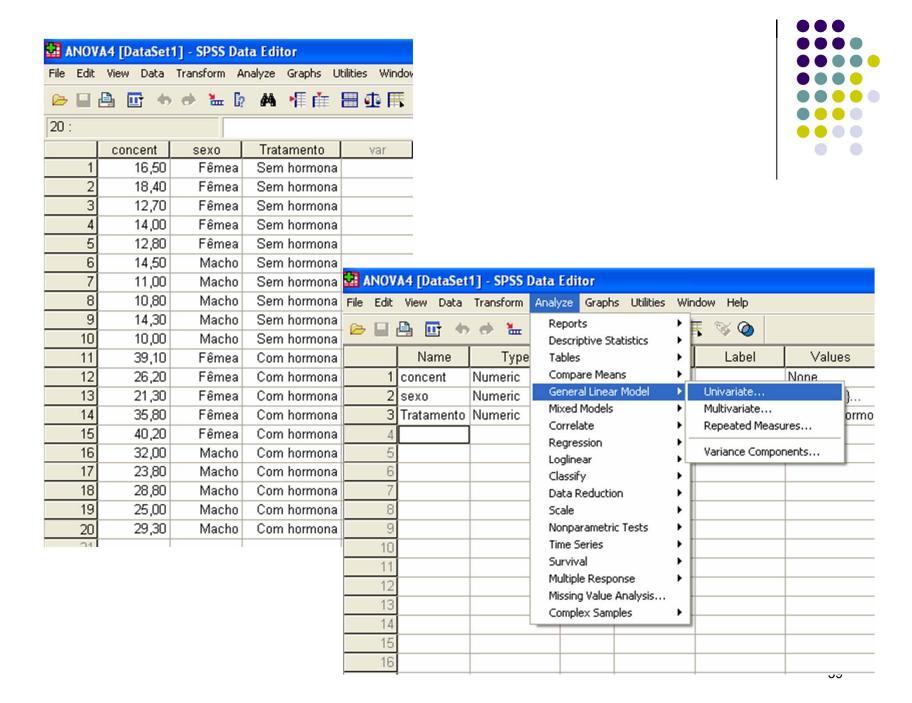
$$STQ = \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{r} y_{ijk}^{2} - \frac{T^{2}}{pqr}$$

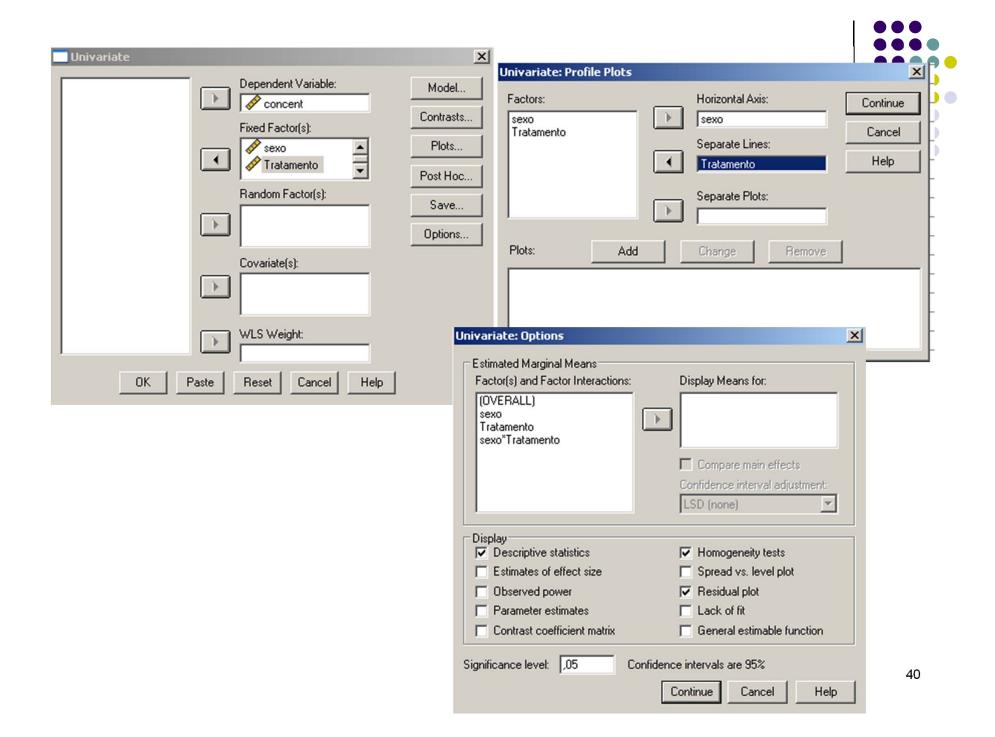
$$SQI_{AB} = STQ - SQF_{A} - SQF_{B} - SQR$$

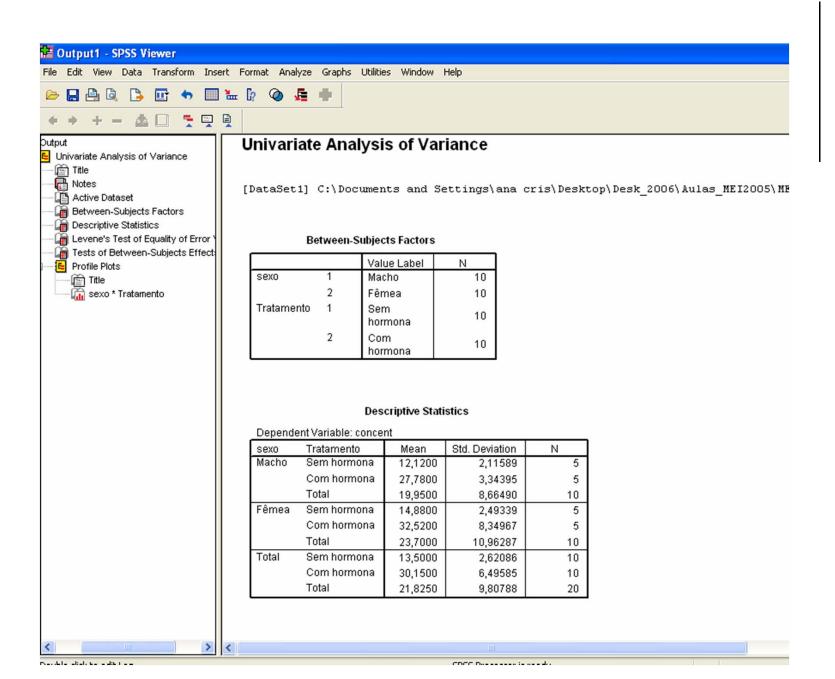
$$SQI_{AxB} = 4,901$$



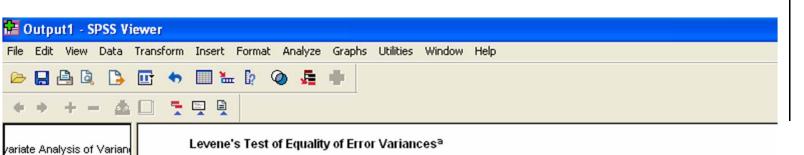
Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Média dos Quadrados	F
Factor A	1386,113	1	1386,113	F ₁ =60,53
Factor B	70,313	1	70,313	F ₂ = 3,07
Interacção AxB	4,901	1	4,901	E - 0.21
Resíduos	366,372	16	22,898	F ₃ = 0,21
Total	1827,698	19	$F_{1,16,0.05}$	= 4,49











variate Analysis of Varia Notes

Levene's Test of Equality Tests of Between-Subjec Profile Plots Dependent Variable: concent

F	df1	df2	Sig.	
10,365	3	16	,000	

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+sexo+Tratamento+sexo * Tratamento

Tests of Between-Subjects Effects

Dependent Variable: concent

	Type III Sum				
Source	of Squares	df	Mean Square	F	Sig.
Corrected Model	1461,326ª	3	487,109	21,273	,000
Intercept	9526,613	1	9526,613	416,041	,000
sexo	70,313	1	70,313	3,071	,099
Tratamento	1386,113	1	1386,113	60,534	,000
sexo * Tratamento	4,901	1	4,901	,214	,650
Error	366,372	16	22,898		
Total	11354,310	20			
Corrected Total	1827,698	19			

a. R Squared = ,800 (Adjusted R Squared = ,762)



Dependent Variable: concent

Observed		6	80 80 90	° ° 6	88 80 80 80 80
Predicted	00 00 00 00 00 00			00	0 00 000 000 000 000
Std. Residual	& & ° ° °	•8 96	9 9 9 0		
	Observed	Predi	icted	Std. Re	sidual

Model: Intercept + sexo + Tratamento + sexo * Tratamento

