

5. ANÁLISE DE VARIÂNCIA

5.1. Caracterização

A análise de variância introduzida por R.A.Fisher é essencialmente um processo aritmético para decompor a variação total entre as unidades experimentais (variância total ou soma de quadrados de desvios totais ou soma dos quadrados total), em componentes associados a fontes ou causas previstas ou identificáveis de variação. Mais especificamente a variação total é decomposta em 3 grupos de causas ou fontes de variação:

- (a) Variação relacionada com os tratamentos.
- (b) Variação relacionada com causas controladas pelo delineamento experimental.
- (c) Variação relacionada com o erro experimental.

A análise de variância constitui-se na técnica estatística geralmente empregada na análise de dados experimentais quantitativos, nas diversas áreas de pesquisa, aplicando-se inclusive, nos casos em que se comparam dois tratamentos.

5.2. Análise de Variância para Classificações Simples (Classificações com um Único Fator) ou Análise de Variância para o Delineamento Completamente Casualizado

5.2.1. Caracterização

Quando as unidades experimentais são homogêneas, não existe um fundamento lógico para formação de blocos. No delineamento completamente casualizado cada unidade experimental recebe por sorteio um dos tratamentos que deverão ser comparados. Nesse caso nenhuma restrição é imposta quanto a casualização, ou seja a atribuição dos tratamentos às unidades experimentais é feita considerando o conjunto completo das unidades experimentais.

5.2.2. Vantagens do Delineamento Completamente Casualizado

O delineamento completamente casualizado é flexível quanto ao número de tratamentos e de repetições, limitando-se ao número de unidades experimentais disponíveis. O número de repetições pode variar de tratamento a tratamento; muito embora geralmente deseje-se um número igual de repetições por tratamento. O número desigual pode ocorrer por falta de material ou por perda de observação.

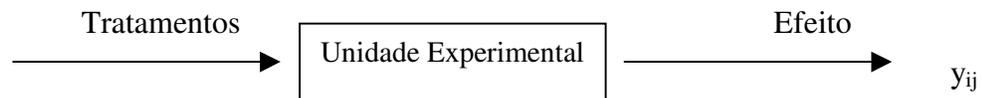
A análise estatística é simples, mesmo quando o número de repetições varia de tratamento a tratamento. Variâncias desiguais, referidas como heterogeneidade de variâncias podem ser acomodadas em procedimento específico para o teste de hipótese e construção de intervalos de confiança. A perda de informação com dados faltantes é pequena em relação a perda com outro delineamento. O número de graus de liberdade para estimativa do erro experimental é máximo, melhorando a precisão do experimento e permitindo sensibilidade aos testes estatísticos.

5.2.3. Desvantagens do Delineamento Completamente Casualizado

A principal objeção do delineamento completamente casualizado é que em algumas vezes é ineficiente. Desde que a casualização é irrestrita o erro experimental inclui todas as variações entre as unidades experimentais exceto aquela devida aos tratamentos. Em muitas situações é possível agrupar unidades experimentais de tal forma que a variação entre unidades dentro dos grupos seja menor do que entre unidades de grupos diferentes. Certos delineamentos, como o delineamento em blocos casualizados, usam a vantagem do agrupamento de unidades experimentais semelhantes, excluindo a variação entre grupos de unidades do erro experimental, e aumentando a precisão do experimento.

5.3. Análise de Variância com qualquer Número de Tratamentos (Grupos) e Repetições Iguais

Considerando esquematicamente um experimento, tem-se:



onde y_{ij} são as observações numéricas referente a uma variável resposta sobre as r_t unidades experimentais. As observações y_{ij} podem ser acomodadas numa estrutura conforme a que é apresentada na tabela 5.1.

Tabela 5.1 – Estrutura dos dados para o delineamento completamente casualizado com qualquer número de tratamentos e repetições iguais.

Tratamentos	Repetições				Totais de Tratamentos	Médias de Tratamentos
	1	2	Λ	r		
1	y_{11}	y_{12}	Λ	y_{1r}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	Λ	y_{2r}	$y_{2.}$	$\bar{y}_{2.}$
\mathbf{N}	\mathbf{M}	\mathbf{M}	y_{ij}	\mathbf{M}	\mathbf{M}	\mathbf{M}
t	y_{t1}	y_{t2}	Λ	y_{tr}	$y_{t.}$	$\bar{y}_{t.}$
					$y_{..}$	$\bar{y}_{..}$

Na tabela 5.1 y_{ij} denota a observação da j -ésima repetição do tratamento i , onde $i = 1, 2, \dots, t$ é o índice de tratamento; $j = 1, 2, \dots, r$ é o índice de repetição.

Os totais dos tratamentos são designados $y_{i.}$, em que o índice i . (i ponto) significa que as repetições j do tratamento foram somadas. Da mesma forma, $\bar{y}_{i.}$ representa a média ($y_{i.}/r$) do tratamento i . O total geral é:

$$y_{..} = \sum y_{ij}$$

e a média geral é

$$\bar{y}_{..} = \sum y_{ij} / rt = y_{..} / rt.$$

No delineamento completamente casualizado a variação total é decomposta em duas partes: a variação entre os tratamentos e a variação entre as unidades experimentais com o mesmo tratamento.

Comprova-se algebricamente que:

$$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = r \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$$

O termo $\sum (y_{ij} - \bar{y}_{..})^2$ é chamado **soma dos quadrados total** (SQ Total). Representa a variação de todas as observações em torno da média geral.

O termo $r \sum (\bar{y}_{i.} - \bar{y}_{..})^2$ é chamado **soma dos quadrados dos tratamentos** (SQ Tratamentos ou SQT). Representa a variação das médias dos tratamentos em torno da média geral, ou a variação entre os tratamentos ou devida a tratamentos.

Enfim, o termo $\sum (y_{ij} - \bar{y}_{i.})^2$ é chamado **soma dos quadrados do erro experimental** (SQ Erro ou SQE). Representa a variação dentro dos tratamentos, isto é, a variação entre as unidades experimentais com o mesmo tratamento, ou seja a variação devida ao erro experimental, que não é de responsabilidade dos tratamentos.

Os três termos têm, respectivamente, $(rt - 1)$, $(t - 1)$ e $t(r - 1)$ graus de liberdade, de forma que

$$(rt - 1) = (t - 1) + t(r - 1).$$

Ainda que essas somas dos quadrados possam ser obtidas pelas equações dadas, é preferível usar equações transformadas, mais adaptadas aos procedimentos computacionais:

$$\text{SQ Total} = \sum y_{ij}^2 - \frac{y_{..}^2}{rt} \quad (5.1)$$

$$\text{SQ Tratamentos} = \text{SQT} = \frac{\sum y_{i.}^2}{r} - \frac{y_{..}^2}{rt} \quad (5.2)$$

$$\text{SQ Erro Experimental} = \text{SQE} = \text{SQ Total} - \text{SQT} \quad (5.3)$$

A soma dos quadrados para erro experimental, mesmo que possa ser calculada diretamente, é determinada mais facilmente por **subtração**. Isto, como decorrência da equação geral de subdivisão da soma dos quadrados total. Por esta razão o erro experimental é também denominado **resíduo** ou **discrepância**.

O termo $y_{..}^2 / rt$, comum em (5.1) e (5.2), é o **fator de correção**, FC. A análise de variância é estruturada numa tabela especial denominada **tabela da análise de variância**. A tabela 5.2 é o modelo geral para a análise da variância de um experimento conduzido no delineamento completamente casualizado.

Tabela 5.2 – Análise de variância do delineamento completamente casualizado com qualquer número de tratamentos e repetições iguais.

Causas de variação	GL	SQ	QM	F
Tratamentos (Entre tratamentos)	t - 1	SQT	QMT	$\frac{QMT}{QME}$
Erro Experimental (Dentro dos tratamentos)	t(r - 1)	SQE	QME	
Total	rt - 1	SQ Total		

Após o cálculo das somas dos quadrados, calculam-se os quadrados médios QM_T , para tratamentos, e QM_E para o erro experimental, dividindo as somas dos quadrados pelos respectivos graus de liberdade.

A hipótese de nulidade (H_0) que se formula é de que não há diferença entre as médias dos tratamentos ($H_0: \mu_1 = \mu_2 = \dots = \mu_t$). Outras maneiras de formular a hipótese de nulidade são as seguintes: não há diferença entre os efeitos dos tratamentos ou os efeitos de tratamentos são nulos ($H_0: \tau_i = 0$), ou a variância dos efeitos dos tratamentos é igual a zero ($H_0: \sigma_\tau^2 = 0$).

O teste da hipótese de nulidade é dado por:

$$F = \frac{QM_{\text{Tratamentos}}}{QM_{\text{ErroExperimental}}}$$

O F calculado é comparado com o dado na tabela de distribuição F para (t - 1) e t(r - 1) graus de liberdade, respectivamente, de tratamentos e do erro experimental. Se for maior que o dado para o nível 5%, a diferença é dita significativa (P<0,05); será muito significativa quando F calculado for maior do que o dado para o nível 1% (P<0,01). No caso de F calculado ser menor do que o tabelado, não haverá diferença significativa entre os tratamentos.

O F-teste é essencialmente a comparação da variância das médias dos tratamentos com a variância do erro experimental. O erro experimental representa a variação aleatória entre as unidades experimentais com o mesmo tratamento, acrescida das variações de erro de técnica cometidos durante a condução do experimento. Se a variação entre as médias dos tratamentos for semelhante à variação do erro experimental, a relação QM_T/QM_E será aproximadamente igual à unidade. Neste caso a diferença entre as médias não será significativa e poderá ser atribuída à variação de amostragem. Para que a diferença entre as médias tenha significância estatística, o valor F calculado deverá ser bem maior do que a unidade. Quando isto sucede, a variação entre as médias dos tratamentos incluirá, além da variação do erro experimental, uma variação devida ao efeito intrínseco dos tratamentos.

Exemplo 1: Na tabela 5.3 encontram-se informações do conteúdo de nitrogênio em mg, de plantas de trevo vermelho inoculadas com culturas de estirpes *Rhizobium trifolii* mais um composto de cinco estirpes de *Rhizobium meliloti*.

Cada uma das cinco linhagens de trevo vermelho, *Rhizobium trifolii*, foi testada individualmente com um composto de cinco estirpes de alfafa, *Rhizobium meliloti*, e um composto de estirpes de trevo vermelho também foi testado com um composto de estirpes de alfafa, totalizando seis tratamentos. O experimento foi conduzido em casa de

vegetação, utilizando-se o delineamento completamente casualizado com cinco vasos (unidades experimentais) por tratamento.

Tabela 5.3 – Conteúdo de nitrogênio (mg) de plantas de trevo vermelho inoculadas em combinação de culturas de estirpes de *Rhizobium trifolii* e de *Rhizobium meliloti*.

	Tratamentos						Total
	3Dok1	3Dok5	3Dok4	3Dok7	3Dok1 3	Compos to	
	19,4	17,7	17,0	20,7	14,3	17,3	
	32,6	24,8	19,4	21,0	14,4	19,4	
	27,0	27,9	9,1	20,5	11,8	19,1	
	32,1	25,2	11,9	18,8	11,6	16,9	
	33,0	24,3	15,8	18,6	14,2	20,8	
1. $\sum_j y_{ij} = y_{i.}$	144,1	119,9	73,2	99,6	66,3	93,5	596,6 = y..
2. $\sum_j y_{ij}^2$	4287,53	2932,27	1139,42	1989,14	887,29	1758,71	12994,36
3. $(y_{i.})^2 / r$	4152,96	2875,20	1071,65	1984,03	879,14	1748,45	12711,43
4. $\sum_j (y_{ij} - \bar{y}_{i.})^2$	134,57	57,07	67,77	5,11	8,15	10,26	282,93
5. $\bar{y}_{i.}$	28,8	24,0	14,6	19,9	13,3	18,7	

Procedendo a análise de variância do exemplo da Tabela 5.3, obteve-se:

$$FC = \frac{y_{..}^2}{rt} = \frac{(\sum_{i,j} y_{ij})^2}{rt} = \frac{(596,6)^2}{(5)(6)} = 11864,38$$

$$SQ \text{ Total} = \sum_{i,j} y_{ij}^2 - FC = (19,4)^2 + \Lambda + (20,8)^2 - FC = 12994,36 - 11864,38 = 1129,98$$

$$SQ \text{ Tratamento} = \frac{Y_{1.}^2 + \Lambda + Y_{r.}^2}{r} - FC$$

$$= \frac{(144,1)^2 + \Lambda + (93,5)^2}{5} - 11.864,38 = 847,05$$

$$SQ \text{ Erro Experimental} = SQ \text{ Total} - SQ \text{ Tratamentos}$$

$$= 1.129,98 - 847,05 = 282,93$$

Os resultados numéricos da análise da variância encontram-se na tabela 5.4.

Tabela 5.4 – Análise da variância dos dados da Tabela 5.3

Causas da variação	GL	SQ	QM	F
Tratamentos	5	847,05	169,41	14,37**
Erro Experimental	24	282,93	11,79	
Total	29	1129,98		

**Muito significativo (P<0,01)

A hipótese estatística a ser testada no caso é $H_0: \mu_1 = \mu_2 = \dots = \mu_6$

Comparando F calculado com o teórico dado na tabela da distribuição F para 5 e 24 graus de liberdade, considerando-se um nível de significância de 1%, obtém-se :

$$F = 14,37 > F_{.01(5,24)} = 3,90$$

A diferença entre as médias dos tratamentos é muito significativa ($P < 0,01$). Rejeita-se H_0 . Os tratamentos pesquisados se diferenciam.

O F-teste, contudo, não indica quais as diferenças entre os tratamentos que devem ser considerados como estatisticamente significativas. Testes adicionais, em que se comparam as médias duas a duas são necessários para evidenciar quais os tratamentos que diferem sob o ponto de vista estatístico.

O erro padrão da média de um tratamento é dado por:

$$s_{\bar{y}} = \sqrt{\frac{s^2}{r}} = \sqrt{\frac{QM_E}{r}} = \sqrt{\frac{11,79}{5}} = 1,54 \text{ mg.}$$

O erro padrão da diferença de duas médias é calculado por:

$$s_d = \sqrt{\frac{2QM_E}{r}} = \sqrt{\frac{2(11,79)}{5}} = 2,17 \text{ mg}$$

Em ambos casos $s^2 = QME$, o quadrado médio do erro experimental. Em realidade s^2 é a variância ponderada das variâncias de unidades experimentais com o mesmo tratamento.

Estas estatísticas são úteis na comparação das médias de tratamentos.

O coeficiente de variação do experimento é dado por

$$CV = \frac{\sqrt{QM_E}}{\bar{Y}_{..}} 100 = \frac{\sqrt{11,79}}{14,89} 100 = 17,3\%$$

Pode-se proceder a análise utilizando o Excel ou o SAS, por exemplo. Os resultados obtidos utilizando-se o SAS encontram-se na tabela 5.5.

Tabela 5.5 – Resultados da Análise de Variância utilizando o SAS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Tratamentos	5	847.046667	169.409333	14.37	<.0001
Error	24	282.928000	11.788667		
Corrected Total	29	1129.974667			

R-Square	Coeff Var	Root MSE	nitrogenio Mean
0.749616	17.26515	3.433463	19.88667

tratamento	Nitrogenio LSMEAN	LSMEAN Number
3DOK1	28.8200000	1
3DOK13	13.2600000	2
3DOK4	14.6400000	3
3DOK5	23.9800000	4
3DOK7	19.9200000	5
Composto	18.7000000	6

Least Squares Means for Effect tratamento						
t for H0: LSMean(i)=LSMean(j) / Pr > t						
i/j	1	2	3	4	5	6
1		7.165512 <.0001	6.530011 <.0001	2.228861 0.0354	4.098526 0.0004	4.660346 <.0001
2	-7.16551 <.0001		-0.6355 0.5311	-4.93665 <.0001	-3.06699 0.0053	-2.50517 0.0194
3	-6.53001 <.0001	0.635502 0.5311		-4.30115 0.0002	-2.43148 0.0229	-1.86966 0.0738
4	-2.22886 0.0354	4.936651 <.0001	4.301149 0.0002		1.869665 0.0738	2.431485 0.0229
5	-4.09853 0.0004	3.066987 0.0053	2.431485 0.0229	-1.86966 0.0738		0.56182 0.5794
6	-4.66035 <.0001	2.505166 0.0194	1.869665 0.0738	-2.43148 0.0229	-0.56182 0.5794	

Os resultados obtidos utilizando-se o Excel encontram-se na tabela 5.6.

Tabela 5.6 – Resultados da Análise de Variância utilizando o Excel

Anova: fator único

RESUMO

<i>Grupo</i>	<i>Contagem</i>	<i>Soma</i>	<i>Média</i>	<i>Variância</i>
3DOK1	5	144,1	28,82	33,642
3DOK5	5	119,9	23,98	14,267
3DOK4	5	73,2	14,64	16,943
3DOK7	5	99,6	19,92	1,277
3DOK13	5	66,3	13,26	2,038
Composto	5	93,5	18,7	2,565

ANOVA

<i>Fonte da variação</i>	<i>SQ</i>	<i>gl</i>	<i>MQ</i>	<i>F</i>	<i>valor-P</i>	<i>F crítico</i>
Entre grupos	847,0467	5	169,4093	14,37053	1,48E-06	3,895082
Dentro dos grupos	282,928	24	11,78867			
Total	1129,975	29				

Exemplo 2: Os dados da tabela 5.7 referem-se a rendimento de cana em t/ha de um experimento no delineamento completamente casualizado de competição de variedades de cana-de-açúcar.

Tabela 5.7 – Rendimento de cana em t/ha de variedades de cana-de-açúcar.

	Tratamentos (Variedades)					
	A	B	C	D		
	64	78	75	55	t = 4 r = 6 rt = 24	
	72	91	93	66		
	68	97	78	49		
	77	82	71	64		
	56	85	63	70		
	95	77	76	68		
Total ($y_{i.}$)	432	510	456	372	1770	$y_{..}$
Média ($\bar{y}_{i.}$)	72	85	76	62	73.75	$\bar{y}_{..}$
$\sum_j y_{ij}^2$	31994	43652	35144	23402	134192	$\sum_{i,j} y_{ij}^2$
$y_{i.}^2/n$ (FC)	31104	43350	34656	23064	132174	$\sum_i y_{i.}^2/n$
SQ/T	890	302	488	338	2018	SQE

$$SQ_{Total} = 134192 - \frac{(1770)^2}{24} = 134192 - 130558 = 3654$$

$$SQT = \frac{432^2 + 510^2 + 456^2 + 372^2}{6} - FC = 132174$$

$$SQE = SQ_{Total} - SQT = 3654 - 1636 = 2018$$

$$SQE = SQ/T_A + SQ/T_B + SQ/T_C + SQ/T_D = 890 + 302 + 488 + 338 = 2018$$

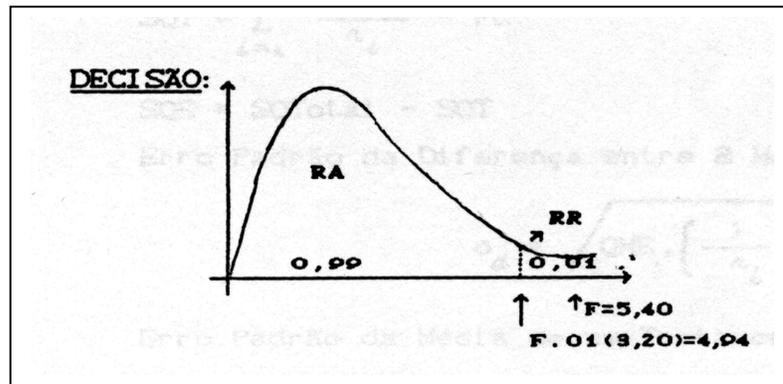
Os resultados numéricos da análise da variância encontram-se na tabela 5.8.

Tabela 5.8 – Análise da variância dos dados da Tabela 5.7

Causas de Variação	GL	SQ	QM	F
Variedades (entre variedades)	3	1636	545.3	5.40**
Erro Experimental (dentro de variedades)	20	2018	100.9	
Total	23	3654		

$$CV = (\sqrt{QME}/\bar{y}_{..}) \times 100 = (\sqrt{100,9}/73,75) \times 100 = 13,6\%$$

$$F_{.01}(3,20) = 4,94$$



$$F = 5.40 > F.01(3.20) = 4.94$$

A diferença entre médias de tratamentos é significativa ($P < 0.01$)
 Rejeita-se H_0

5.8 Os resultados da análise de variância utilizando-se o Excel aparecem na tabela

Tabela 5.8 – Resultados da Análise de Variância utilizando o Excel

Anova: fator único

RESUMO

<i>Grupo</i>	<i>Contagem</i>	<i>Soma</i>	<i>Média</i>	<i>Variância</i>
A	6	432	72	178
B	6	510	85	60,4
C	6	456	76	97,6
D	6	372	62	67,6

ANOVA

<i>Fonte da variação</i>	<i>SQ</i>	<i>gl</i>	<i>MQ</i>	<i>F</i>	<i>valor-P</i>	<i>F crítico</i>
Entre grupos	1636,5	3	545,5	5,406343	0,006876	4,938215
Dentro dos grupos	2018	20	100,9			
Total	3654,5	23				

5.2.4 Análise de Variância com qualquer Número de Tratamentos (Grupos) e Repetições Diferentes

Nesse caso tem-se:

TRATAMENTOS	REPETIÇÕES
1	r_1
2	r_2
.	.
.	.
.	.
t	r_t
TOTAL	n

A tabela de Análise de Variância apresenta a estrutura

CAUSAS DE VARIACÃO	GL	SQ	QM	F
TRATAMENTOS	t-1	SQT	QMT	$\frac{QMT}{QME}$
ERRO EXPERIMENTAL	n-t	SQE	QME	$\frac{QME}{QME}$
TOTAL	n-1	SQTotal		

As expressões para as somas de quadrados ficam:

$$SQTotal = \sum_{i,j} y_{ij}^2 - FC \quad ; \quad FC = \frac{y^2}{n}$$

$$SQT = \sum_{i=1}^t \frac{y_i^2}{r_i} - FC$$

$$SQE = SQTotal - SQT$$

As expressões para os erros padrões transformam-se para:

Erro padrão da Diferença entre duas médias de Tratamentos:

$$s_d = \sqrt{QME \left(\frac{1}{r_i} + \frac{1}{r_j} \right)}$$

Erro Padrão da Média de um Tratamento:

$$s_y = \sqrt{\frac{1}{2} QME \left(\frac{1}{r_i} + \frac{1}{r_j} \right)}$$

Exemplo 3: Considere-se que as duas últimas observações do Tratamento 3Dok1 e a última observação do tratamento 3Dok5 do exemplo 1 foram perdidas. Procedendo-se a análise de variância através do SAS e do Excel, obtém-se os resultados apresentados nas tabelas 5.9 e 5.10.

Tabela 5.9 - Análise de Variância pelo SAS considerando diferente número de repetições por tratamento.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Tratamentos	5	520.5954074	104.1190815	9.26	<.0001
Error	21	236.0186667	11.2389841		
Corrected Total	26	756.6140741			

R-Square	Coeff Var	Root MSE	nitrogenio Mean
0.688059	17.84629	3.352459	18.78519

tratamento	nitrogenio LSMEAN	LSMEAN Number
3DOK1	26.3333333	1
3DOK13	13.2600000	2
3DOK4	14.6400000	3
3DOK5	23.9000000	4
3DOK7	19.9200000	5
Composto	18.7000000	6

Least Squares Means for Effect tratamento						
t for H0: LSMean(i)=LSMean(j) / Pr > t						
i/j	1	2	3	4	5	6
1		5.339781 <.0001	4.776122 0.0001	0.950341 0.3527	2.619515 0.0160	3.117822 0.0052
2	-5.33978 <.0001		-0.65086 0.5222	-4.73121 0.0001	-3.14109 0.0049	-2.5657 0.0180
3	-4.77612 0.0001	0.650857 0.5222		-4.11757 0.0005	-2.49024 0.0212	-1.91484 0.0692
4	-0.95034 0.3527	4.731206 0.0001	4.117572 0.0005		1.769756 0.0913	2.312243 0.0310
5	-2.61952 0.0160	3.141092 0.0049	2.490235 0.0212	-1.76976 0.0913		0.575395 0.5711
6	-3.11782 0.0052	2.565697 0.0180	1.91484 0.0692	-2.31224 0.0310	-0.5754 0.5711	

Tabela 5.10 - Análise de Variância pelo Excel considerando diferente número de repetições por tratamento.

Anova: fator único

RESUMO

<i>Grupo</i>	<i>Contagem</i>	<i>Soma</i>	<i>Média</i>	<i>Variância</i>
3DOK1	3	79	26,33333	43,89333
3DOK5	4	95,6	23,9	18,98
3DOK4	5	73,2	14,64	16,943
3DOK7	5	99,6	19,92	1,277
3DOK13	5	66,3	13,26	2,038
composto	5	93,5	18,7	2,565

ANOVA

<i>Fonte da variação</i>	<i>SQ</i>	<i>gl</i>	<i>MQ</i>	<i>F</i>	<i>valor-P</i>	<i>F crítico</i>
Entre grupos	520,5954	5	104,1191	9,264101	8,93E-05	4,042136
Dentro dos grupos	236,0187	21	11,23898			
Total	756,6141	26				

Exemplo 4: Considere-se que as duas últimas observações da variedade A e a última observação da variedade B do exemplo 2 foram perdidas, então ter-se-ia a estrutura de dados apresentada na tabela 5.11

Tabela 5.11 – Exemplo 2 com número diferente de repetições

A	64	72	68	77		
B	78	91	97	82	85	
C	75	93	78	71	63	76
D	55	66	49	64	70	68

A análise de variância aparece na tabela 5.12

Tabela 5.12 – Análise de Variância para os dados da tabela 5.11

C. Variação	GL	SQ	QM	F
Variedades	3	1731	577	8,58
Erro experimental	17	1144	67,3	
Total	20	2875		

$$n \ m \ s = P > F = 0,001376$$

onde

$$SQ_{total} = 64^2 + 72^2 + 68^2 + 77^2 - 1542^2 / 21 = 2875$$

$$SQT = \frac{281^2}{4} + \frac{431^2}{5} + \frac{456^2 + 372^2}{6} - \frac{1542^2}{21} = 1731$$

$$SQE = SQ_{total} - SQT = 2875 - 1731 = 1144$$

Procedendo-se a análise de variância através do Excel, obtém-se os resultados apresentados nas tabela 5.13.

Tabela 5.13 - Análise de Variância pelo Excel considerando diferente número de repetições por tratamento para o Exemplo 4

Anova: fator único

RESUMO

Grupo	Contagem	Soma	Média	Variância
A	4	281	70,25	30,91667
B	5	433	86,6	56,3
C	6	456	76	97,6
D	6	372	62	67,6

ANOVA

Fonte da variação	SQ	gl	MQ	F	Valor-P	F crítico
Entre grupos	1731,193	3	577,0643	8,575631	0,001091	3,196774
Dentro dos grupos	1143,95	17	67,29118			
Total	2875,143	20				

5.4. Técnicas de Complementação na Análise de Variância

1 – Ajustamento de funções de resposta através de técnicas de análise de regressão para fatores quantitativos.

2 – Contrastes ortogonais para fatores quantitativos e qualitativos que permitem estruturação.

3 – Comparações múltiplas de médias para fatores quantitativos e qualitativos que permitem ou que não permitem estruturação.

5.5 Métodos de Comparações Múltiplas

5.5.1 Testes de amplitude simples

(1) **DMS** (Diferença Mínima Significativa)

$$\text{DMS} = t\alpha (GLE) s_d ; \quad s_d = \sqrt{\frac{2 \text{QME}}{r}}$$

(2) **TUKEY**:

$$\Delta = q\alpha (t, GLE) s_{\bar{y}} ; \quad s_{\bar{y}} = \sqrt{\frac{\text{QME}}{r}}$$

(3) **SCHEFFÉ**:

$$S = \sqrt{(t-1) F\alpha [GLT, GLE] \hat{V}(\hat{C})}$$

$$C = \sum_{i=1}^t c_i \mu_i ; \quad \hat{C} = \sum_{i=1}^t c_i \bar{y}_i$$

$$\hat{V}(\hat{C}) = \frac{\text{QME}}{r} \sum_{i=1}^t c_i^2$$

(4) **BONFERRONI**:

$$\text{DMS}_B = t\alpha' (GLE) s_d ; \quad \alpha' = \alpha / k$$

(5) **DUNNETT**:

Comparações de tratamentos versus testemunha

$$d' = t\alpha (t', GLE) s_d$$

5.5.2 Testes de amplitude múltipla:

(6) **S-N-K OU N-K**: (Student-Newman-Keuls ou Newman-Keuls)

$$\Delta p = q\alpha(p, GLE)s_{\bar{y}}$$

↳ n° de médias incluídas na comparação

(7) **DUNCAN**:

$$AMS_p = q * \alpha p(p, GLE)s_{\bar{y}}$$

Taxa de erro: $\alpha p = 1 - (1 - \alpha)^{p-1}$

$\alpha = 0,05$

$p = 2 \Rightarrow \alpha_2 = 1 - (1 - 0,05)^{2-1} = 1 - 0,95 = 0,05$

$p = 3 \Rightarrow \alpha_3 = 1 - (1 - 0,05)^{3-1} = 1 - 0,95^2 = 0,10$

M

M

Valores de $q * \alpha p(p, GLE)$ também conhecidos como ASS (Amplitude Studentizada Significativa) correspondem aos valores de amplitude Studentizada $[q(p, GLE)]$, considerando $\alpha = \alpha p$.

Casos particulares do teste de Duncan:

- DMS: $\alpha p = \alpha$ $p = 2$
- TUKEY: $\alpha p = \alpha$ $p = t$
- S-N-K: $\alpha p = \alpha$ $p = p$

Exemplo 5: Testes de comparações múltiplas para o exemplo 2

DMS $DMS_{5\%} = t_{.05(GLE)}s_d = t_{.05(20)}s_d = (2,086)(5,8) = 12,1$

$$s_d = \sqrt{\frac{2QME}{r}} = \sqrt{\frac{2(100,9)}{6}} = 5,8$$

TUKEY $\Delta_{5\%} = q_{.05(t,GLE)}s_{\bar{y}} = q_{.05(4,20)}s_{\bar{y}} = (3,96)(4,1) = 16,2$

$$s_{\bar{y}} = \sqrt{\frac{QME}{r}} = \sqrt{\frac{100,9}{6}} = 4,1$$

SCHEFFÉ $S_{5\%} = \sqrt{(t-1)F_{.05(GLT,GLE)}\hat{V}(\hat{C})} = \sqrt{(4-1)(3,10)(33,6)} = 17,7$

$$F_{.05(GLT,GLE)} = F_{.05(3,20)} = 3,10$$

$$\hat{V}(\hat{C}) = \hat{V}(\bar{y}_i - \bar{y}_j) = s_d^2 = \frac{2QME}{r} = \frac{2(100,9)}{6} = 33,6$$

$$V(\hat{C}) = \frac{QME}{r}((1)^2 + (-1)^2) = \frac{2QME}{r}$$

BONFERRONI $DMS_{B_{3\%}} = t_{\alpha(GLE)}s_d = (2,845)(5,8) = 16,5$

$$\alpha' = \frac{\alpha}{6} = \frac{0,05}{6} = 0,0083 \cong 0,01$$

$$t_{,01(20)} = 2.845$$

DUNNETT

$$d'_{5\%} = t_{\alpha'(t',GLE)}^D s_d = (2,54)(5,8) = 14,7$$

$$t_{,05(3,20)}^D = 2,54$$

D → Padrão B e D: $85 - 62 = 23 > 14,7 \Rightarrow B \neq D$ (B>D)
 C e D: $76 - 62 = 14 < 14,7 \Rightarrow C \tilde{=} deD$
 A e D: $72 - 62 = 10 < 14,7 \Rightarrow A \tilde{=} deD$

SNK

$$\Delta_p = q_{\alpha(p,GLE)} s_{\bar{y}} = q_{,05(p,20)} (4,1)$$

p	2	3	4	
$q_{,05(p,20)}$	2,95	3,58	3,96	
Δ_p	12,1	14,7	16,2	→ Δ de Tukey
	↳ DMS			

DUNCAN

$$AMS = ASS_{\alpha(p,GLE)} s_{\bar{y}} = ASS_{,05(p,20)} (4,1)$$

p	2	3	4
$q_{,05(p,20)}$	2,95	3,10	3,18
AMS_p	12,1	12,7	13,0
	↳ DMS		

Resumo do exemplo2:

Variedades	Médias t/hα	DMS	Tukey	Scheffé	Bonferroni	S-N-K	Duncan	Dunnnett
B	85	a	a	a	a	a	a	B ≠ D C = D A = D
C	76	ab	ab	ab	ab	ab	ab	
A	72	bc	ab	ab	ab	ab	bc	
D	62	c	b	b	b	b	c	
Valor p/ o teste	2					12,1	12,1	14,9
	3	12,1	16,2	17,7	16,5	14,7	12,7	
	4					16,2	13,0	
Nº de ≠ significativas	-	3	1	1	1	1	3	1

Exemplo 6: Testes de comparações múltiplas para o exemplo 1

Para os dados do *Rhizobium*, do exemplo 1 tem-se:

$$DMS_{0,05} = t_{0,025} s_{\bar{Y}_i - \bar{Y}_j} = 2,064 \sqrt{\frac{2(11,79)}{5}} = 4,5 \text{ mg}$$

Tratamentos						
	3DOK1 3	3DOK4	Compos to	3DOK7	3Dok5	3DOK1
\bar{y}_i	13,3(T1)	14,6(T2)	18,7(T3)	19,9(T4)	24,0(T5)	28,8(T6)

Então: (T6) – (T1) = 15,5 > 4,5 ⇒ **significante**

(T6) - (T2) = 14,2 > 4,5 ⇒ **significante**

...

(T6) – (T5) = 4,8 > 4,5 ⇒ **significante**

(T5) – (T1) = 10,7 > 4,5 ⇒ **significante**

(T5) – (T2) = 9,4 > 4,5 ⇒ **significante**

...

(T2) – (T1) = 1,3 < 4,5 ⇒ **não significante**

Diferenças entre as médias de nitrogênio no experimento de Rhizobium					
	(T6)	(T5)	(T4)	(T3)	(T2)
(T1)	15,5*	10,7*	6,6*	5,4*	1,3
(T2)	14,2*	9,4*	5,3*	4,1	
(T3)	10,1*	5,3*	1,2		
(T4)	8,9*	4,1			
(T5)	4,8*				

13,3 14,6 18,7 19,9 24,0 28,8

O resultado do DMS pelo SAS com o sistema de letras é:

T tests (LSD) for variable: MG

NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.

Alpha= 0.05 df= 24 MSE= 11.78867

Critical Value of T= 2.06

Least Significant Difference= 4.4818

Means with the same letter are not significantly different.

T Grouping	Mean	N	TRAT
A	28.820	5	T3DOK1
B	23.980	5	T3DOK5
C B	19.920	5	T3DOK7
C D	18.700	5	COMPO
E D	14.640	5	T3DOK4
E	13.260	5	T3DOK13

Resumo do exemplo1:

Tratamento	Média	DMS	Bonferroni	Scheffé	Tukey	Ducan	S-N-K	Waller
T3DOK1	28,82	a	a	a a	a a	a	a	a
T3DOK5	23,98	b b	ab	ab b	ab b	b b	b b	b b
T3DOK7	19,92	bc c	bc	bc bc	bc bc	bc c	bc bc	bc c
COMPO	18,7	cd d	bc	bc c	bc c	cd d	bc c	cd d
T3DOK4	14,64	de e	c	c c	c c	de e	cd d	de e
T3DOK13	13,26	e	c	c	c	e	d	e

Tratamento	Dunett
T3DOK1	T3DOK1 ≠ COMPO
T3DOK5	T3DOK5 = COMPO
T3DOK7	T3DOK7 = COMPO
COMPO	
T3DOK4	T3DOK4 = COMPO
T3DOK13	T3DOK13 = COMPO

5.5.3 Considerações sobre os Métodos de Comparações Múltiplas

Taxas de erro tipo I: Há diferentes formas de avaliar o erro tipo I, criando dificuldades para avaliar o mérito relativo dos procedimentos de comparações múltiplas.

(i) **Taxa de erro por comparação (“comparisonwise”):** α

$$\frac{\text{n}^\circ \text{ de inferências erradas}}{\text{n}^\circ \text{ de inferências}}$$

Usada no teste DMS.

(ii) **Taxa de erro por experimento (“experimentwise”):**

(ii.1) $\frac{\text{n}^\circ \text{ de experimentos com no mínimo uma inferência errada}}{\text{n}^\circ \text{ de experimentos}}$

Usada nos testes de Tukey, Scheffé, Dunnett.

(ii.2) $\frac{\text{n}^\circ \text{ de inferências erradas}}{\text{n}^\circ \text{ de experimentos}}$

Usada no teste de Bonferroni.

Relação entre α e E:

$$E = 1 - (1 - \alpha)^{t-1}; \quad \alpha = 1 - (1 - E)^{1/t-1}$$

Nº de tratamentos No experimento	$\alpha = 0,05$ E	E = 0,05 α
2	0,05	0,05
3	0,0975	0,0253
4	0,1426	0,0169
5	0,1835	0,0127
10	0,3698	0,0057
15	0,5124	0,0037
20	0,6227	0,0028

Uso dos testes:

1) Todas as comparações de médias 2 a 2:

TUKEY: maior rigor (maior responsabilidade)

S-N-K

DUNCAN: menor rigor (menor responsabilidade)

2) Contrastes não-ortogonais que envolvam mais do que duas médias (pelo menos um):

SCHEFFÉ

BONFERRONI: poucos contrastes

3) Comparações de tratamentos versus testemunha:

DUNNETT

BONFERRONI: poucas comparações

DMS

5.5.4 Intervalos de Confiança em Comparações Múltiplas

t:
$$\frac{(\bar{y}_i - \bar{y}_j) - (\mu_i - \mu_j)}{s_d} \sim t_{(GLE)}$$

IC 100(1- α)% p/ $\mu_i - \mu_j = \bar{y}_i - \bar{y}_j \pm t_{\alpha/2, (GLE)} s_d = \bar{y}_i - \bar{y}_j \pm \underset{\text{DMS}}{14,243} = \bar{y}_i - \bar{y}_j \pm \text{DMS}$

Tukey:
$$\frac{(\bar{y}_i - \bar{y}_j) - (\mu_i - \mu_j)}{s_{\bar{y}}} \sim q_{(t, GLE)}$$

IC 100(1- α)% p/ $\mu_i - \mu_j = \bar{y}_i - \bar{y}_j \pm q_{\alpha/2, (GLE)} s_{\bar{y}} = \bar{y}_i - \bar{y}_j \pm \underset{\Delta}{14,243} = \bar{y}_i - \bar{y}_j \pm \Delta$

Exemplo 7: Intervalos de confiança para o exemplo 2 das variedades de cana-de-açúcar:

t: IC 95% para $\mu_i - \mu_j = \bar{y}_i - \bar{y}_j \pm t_{0,05(20)} sd = \bar{y}_i - \bar{y}_j \pm (2,086)(5,8) = \bar{y}_i - \bar{y}_j \pm 12,1$

Tukey: IC 95% para $\mu_i - \mu_j = \bar{y}_i - \bar{y}_j \pm q_{0,05(4,20)} s_{\bar{y}} = \bar{y}_i - \bar{y}_j \pm (3,96)(4,1) = \bar{y}_i - \bar{y}_j \pm 16,2$

Contraste	Estimativa do Contraste	IC 95% p/ t	Decisão	IC 95% p/ Tukey	Decisão
$\mu_B - \mu_C$	9	[-3,1;21,1]	NS	[-7,2;25,2]	NS
$\mu_B - \mu_A$	13	[0,9;25,1]	*	[-3,2;29,2]	NS
$\mu_B - \mu_D$	23	[10,9;35,1]	*	[6,8;39,2]	*
$\mu_C - \mu_A$	4	[-8,1;16,1]	NS	[-12,2;20,2]	NS
$\mu_C - \mu_B$	14	[1,9;26,1]	*	[-2,2;30,2]	NS
$\mu_A - \mu_D$	10	[-2,1;22,1]	NS	[-6,2;26,2]	NS

Variedades	Médias, t/ha
B	85
C	76
A	72
D	62

5.5.5. Teste de Waller-Duncan ou Teste da Razão Bayesiana k:

O teste de Waller-Duncan descrito na literatura especializada é um dos procedimentos de comparações múltiplas disponíveis em aplicativos estatísticos tipo SAS e outros.

No teste calcula-se

$$DMS_k = t(k, F, GLT, GLE) \cdot sd$$

Confiabilidade
precisão

Valor calculado de F
 Razão bayesiana
 Valor tabelado

Onde $sd = \sqrt{\frac{2QME}{r}}$

Razão k: importância relativa do erro tipo I em relação ao erro tipo II.

$$k = \frac{\text{custo do erro tipo I}}{\text{custo do erro tipo II}}$$

$\alpha = 0.01$	0.05	0.01
k = 50	100	500

F alto: heterogeneidade entre tratamentos

Valor de t é reduzido

Alto poder de discriminação

Taxa de erro “a nível de comparação” ou por comparação

F baixo: homogeneidade aproximada de tratamentos
 Valor de t é aumentado
 Rigoroso
 Taxa de erro “a nível de experimento” ou por experimento

O teste usa as vantagens dos procedimentos de taxa de erro por experimento e por comparação sem as desvantagens. O teste é trabalhado pois na obtenção do valor para comparação necessita-se fazer interpolações sucessivas em tabelas especiais.

Exemplo 8: Consideramos os dados das cultivares de cana-de-açúcar do exemplo 2.

	C. Variação	GL	QM	F
t = 4	VARIEDADES	3		5.4
r = 6	ERRO	20	100.9	

F = 5.4 f = GLE = 20 q = GLT = 3
 Fazendo interpolações sucessivas em tabelas especiais obtém-se o valor de t tabelado.
 Então

$$DMS_k = t(k, F, GLT, GLE)sd = t(100; 5.4; 3; 20)sd = (2.116)(5.8) = 12.3 \text{ t/ha}$$

$$sd = \sqrt{\frac{2QME}{r}} = \sqrt{\frac{2(100.9)}{6}} = 5.8 \text{ t/ha}$$

VARIEDADES	MEDIAS, t/ha	
B	85	a
C	76	a b
A	72	b c
D	62	c

Médias seguidas de mesma letra não diferem significativamente pelo teste de Waller-Duncan a 5%.

O resultado do SAS aparece a seguir.

Waller-Duncan K-ratio t Test for RENDIMENTO

```

Kratio                                100
Error Degrees of Freedom              20
Error Mean Square                      100.9
F Value                                5.41
Critical Value of t                    2.11620
Minimum Significant Difference         12.273

```

Means with the same letter are not significantly different.

Waller Grouping	Mean	N	VARIEDADES
A	85.000	6	B
B	76.000	6	C
B	72.000	6	A
C	62.000	6	D

5.5.6 Comparações Múltiplas com dados desbalanceados(diferente número de repetições por tratamento)

i. Utilizando-se o número de repetições de cada tratamento

$$s_{\bar{y}} = \sqrt{\frac{QME}{2} \left(\frac{1}{r_i} + \frac{1}{r_j} \right)} \quad e \quad s_d = \sqrt{QME \left(\frac{1}{r_i} + \frac{1}{r_j} \right)}$$

ii. Utilizando-se a média harmônica do número de repetições

$$r_h = \bar{r} = \frac{t}{\frac{1}{r_1} + \frac{1}{r_2} + \dots + \frac{1}{r_t}}$$

$$s_{\bar{y}} = \sqrt{\frac{QME}{\bar{r}}} \quad e \quad s_{\bar{y}} = \sqrt{\frac{2 QME}{\bar{r}}} \quad (\text{menos trabalhoso; indicado quando o n}^\circ \text{ de repetições é muito diferente de tratamento a tratamento})$$

Exemplo 9: Comparação múltipla no caso de diferente número de repetições por tratamento do exemplo 4 das variedades de cana-de-açúcar

Variedades	Médias, t / ha	r _i
A	70,25	4
B	86,60	5
C	76	6
D	62	6

i. Teste de Tukey utilizando o número de repetições de cada tratamento:

$$\Delta_{5\%} = q_{0.05(4, GLE)} s_{\bar{y}}$$

$$= q_{0.05(4,17)} s_{\bar{y}}$$

$$= (4,02) s_{\bar{y}} \quad \text{onde: } s_{\bar{y}} = \sqrt{\frac{QME}{2} \left(\frac{1}{r_i} + \frac{1}{r_j} \right)}$$

Causas de Variação	GL	QM
Tratamentos	3	
Erro Experimental	17	67,3

$$B \text{ e } D: \Delta_1 = q_{0.05(4,17)} s_{\bar{y}} = (4,02) \sqrt{\frac{67,3}{2} \left(\frac{1}{5} + \frac{1}{6} \right)} = (4,02)(3,51) = 14,1$$

$$B \text{ e } A: \Delta_2 = q_{0.05(4,17)} s_{\bar{y}} = (4,02) \sqrt{\frac{67,3}{2} \left(\frac{1}{5} + \frac{1}{4} \right)} = (4,02)(3,89) = 15,6$$

$$B \text{ e } C: \Delta_1 = 14,1$$

$$C \text{ e } D: \Delta_3 = q_{0.05(4,17)} s_{\bar{y}} = (4,02) \sqrt{\frac{67,3}{2} \left(\frac{1}{6} + \frac{1}{6} \right)} = (4,02)(3,35) = 13,5$$

$$C \text{ e } A: \Delta_4 = q_{0.05(4,17)} s_{\bar{y}} = (4,02) \sqrt{\frac{67,3}{2} \left(\frac{1}{6} + \frac{1}{4} \right)} = (4,02)(3,74) = 15,1$$

$$A \text{ e } D: \Delta_4 = 15,1$$

$$\begin{aligned} \bar{y}_B - \bar{y}_D &= 86,6 - 62 = 24,6 > \Delta_1 = 14,1 \Rightarrow B \neq D && (B>D) \\ \bar{y}_B - \bar{y}_A &= 86,6 - 70,25 = 16,35 > \Delta_2 = 15,6 \Rightarrow B \neq A && (B>A) \\ \bar{y}_B - \bar{y}_C &= 86,6 - 76 = 10,6 < \Delta_1 = 14,1 \Rightarrow B \text{ não } \neq C \\ \bar{y}_C - \bar{y}_D &= 76 - 62 = 14 > \Delta_3 = 13,5 \Rightarrow C \neq D && (C>D) \\ \bar{y}_C - \bar{y}_A &= 76 - 70,25 = 5,75 < \Delta_4 = 15,1 \Rightarrow C \text{ não } \neq A \\ \bar{y}_A - \bar{y}_D &= 70,25 - 62 = 8,25 < \Delta_4 = 15,1 \Rightarrow A \text{ não } \neq D \end{aligned}$$

Variedades	Médias, t / ha
A	86,6 a
B	76 a b
C	70,25 b c
D	62 c

IC para Tukey a 95%

Contraste	Estimativa do Contraste	L_I	L_S	Decisão
$\mu_B - \mu_C$	10,6	-3,5	24,7	NS
$\mu_B - \mu_A$	16,35	0,75	31,95	*
$\mu_B - \mu_D$	14,6	10,5	38,7	*
$\mu_C - \mu_A$	5,75	-9,35	20,85	NS
$\mu_C - \mu_D$	14	0,5	27,5	*
$\mu_A - \mu_D$	8,25	-6,85	23,35	NS

ii. Teste de Tukey utilizando a Média Harmônica do número de repetições

$$\begin{aligned} \bar{r} &= \frac{4}{\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{6}} = 5,1 && \Delta_{5\%} = q_{0,05(4,17)} s_{\bar{y}} = (4,02)(3,63) = 14,6 \\ s_{\bar{y}} &= \sqrt{\frac{QME}{\bar{r}}} = \sqrt{\frac{67,3}{5,1}} = 3,63 && q_{0,05(4,17)} = 4,02 \end{aligned}$$

Variedades	Médias t / ha
B	86,60 a
C	76,00 a b
A	70,25 b
D	62,00 b

5.7. Contrates Ortogonais

(Comparações de médias de tratamentos por G.L. individuais na análise de variância)

Uma função linear de médias de tratamentos do tipo $C = \sum_{i=1}^t c_i \mu_i = c_1 \mu_1 + c_2 \mu_2 + \dots + c_t \mu_t$ é dita contraste ou comparação se $\sum_{i=1}^t c_i = 0$.

Os contrastes $C = \sum_{i=1}^t c_i \mu_i$ e $B = \sum_{i=1}^t b_i \mu_i$

São ortogonais se $\sum_{i=1}^t b_i c_i = 0$ [para igual número de repetições por tratamento]

se $\sum_{i=1}^t \frac{b_i c_i}{r_i} = 0$ [para diferente número de repetições por tratamento].

Exemplos

(1) $C_1 = \mu_1 - \mu_2$

(2) $C_2 = \mu_1 - \mu_3$

(3) $C_3 = \frac{\mu_1 - \mu_2}{2} - \mu_3 = \mu_1 + \mu_2 - 2\mu_3$

são contrastes

(4) $C_4 = \mu_1 + \mu_2 - \mu_3$ } não é contraste

C_1 e C_2 não são ortogonais: $(1)(1) + (-1)(0) + (0)(-1) = 1 \neq 0$

C_1 e C_3 são ortogonais: $(1)(1) + (0)(1) + (-1)(-2) = 0$

C_2 e C_3 não são ortogonais : $(1)(1) + (0)(1) + (-1)(-2) = 3 \neq 0$

ANÁLISE DE VARIÂNCIA, DCC

Causas da Variação		GL
contrast ortogonais	tratamentos	t - 1
	C_1	1
	C_2	1
	\vdots	\vdots
	C_{t-1}	1
Erro experimental		t (r - 1)
Total		rt - 1

A técnica consiste em decompor a SQ de tratamentos em tantas partes (contrast) quantos forem os GL de tratamentos, sendo estes contrastes ortogonais entre si e tendo 1 GL.

Para cada contraste calcula-se a SQ dada por

$$SQC_j = \frac{\left(\sum_{i=1}^t c_{ji} y_i \right)^2}{r \sum_{i=1}^t c_{ji}^2}$$

Onde y_i = total do tratamento i
 r = número de repetições

$$\sum_{i=1}^t c_{ji}^2 = \text{soma dos quadrados dos coeficientes dos contrastes}$$

$j = 1, 2, \dots, t-1$ (número de contrastes)

Testa – se cada contraste pela estatística F obtendo-se

$$F = \frac{QMC_j}{QME} \text{ onde } QMC_j = SQC_j$$

Se os contrastes forem ortogonais :

$$SQC_1 + SQC_2 + \dots + SQC_{t-1} = SQT$$

Se os contrastes não forem ortogonais :

$$SQC_1 + SQC_2 + \dots + SQC_{t-1} \neq SQT$$

Como se trabalha com um GL no numerador, está se testando um grupo contra outro levando-nos a uma conclusão específica.

Não apresenta os inconvenientes dos testes de comparações múltiplas de médias.

Exemplo 10: Consideremos o exemplo 2 das variedades de cana-de-açúcar.

		Variedades		Médias, t/ha	Totais, t/ha
origem genética 2	origem genética	A	novas	72	432
		B		85	510
		C		76	456
		D		62	372
	padrão ou testemunha ou controle			t = 4	r = 6

Análise de Variância

Causas da variação	GL	SQ	QM
Variedades	3	1636	
Erro experimental	20	2018	100.9

3 GL \Rightarrow 3 contrastes ortogonais

Contrastes:

$$C_1 : D \text{ versus } (A+B+C) \Leftrightarrow [\text{testemunha vs resto}] \Leftrightarrow$$

$$\Leftrightarrow \left[\begin{array}{cc} \text{novas} & \text{variedade} \\ \text{variedades} & \text{padrão} \end{array} \right] \text{ vs } \left[\begin{array}{c} \text{vs} \\ \text{vs} \end{array} \right]$$

$$C_2 : B \text{ versus } (A+C) \Leftrightarrow [\text{origem genética2 vs origem genética1}]$$

$$C_3 : A \text{ versus } C \Leftrightarrow [\text{entre variedades de origem genética1}]$$

$$C_1 = \mu A + \mu B + \mu C - 3\mu D$$

$$C_2 = \mu A + \mu C - 2\mu B$$

$$C_3 = \mu A - \mu C$$

C_1, C_2, C_3 ortogonais

Variedades	Totais (Y_i)	Coeficientes dos contrastes			$C_{1i}Y_i$	$C_{2i}Y_i$	$C_{3i}Y_i$
		C_{1i}	C_{2i}	C_{3i}			
A	432	1	1	1	432	432	432
B	510	1	-2	0	510	-1020	0
C	456	1	1	-1	456	456	-456
D	372	-3	0	0	-1116	0	0
Total	1770	0	0	0	282	-132	-24

$$SQC_1 = \frac{282^2}{6.12} = 1104$$

$$SQC_2 = \frac{(-132)^2}{6.6} = 484$$

$$SQC_3 = \frac{(-24)^2}{6.2} = 48$$

$$SQC_1 + SQC_2 + SQC_3 = 1104 + 484 + 48 = 1636 = SQT$$

Análise de Variância:

Causas de Variação	GL	SQ	QM	F	
Variedades	3	1636			
C1	1	1404	1104	10,94**	$H_0 : C_1 = 0$
C2	1	484	484	4,80*	$H_0 : C_2 = 0$
C3	1	48	48	0,48	$H_0 : C_3 = 0$
Erro experimental	20	2018	100,9		

$$F.01(1,20)=8,10$$

$$F.05(1,20)=4,35$$

Conclusões:

- Novas variedades (em média) superiores à variedade padrão.
- A variedade de origem genética 2 (variedade B) é superior em média às variedades de origem genética 1 (variedades A e C).
- Não se evidenciam diferenças entre as variedades de origem genética 1.

TESTE DE CONTRASTES UTILIZANDO-SE A ESTATÍSTICA t

$$H_0 : C_j = 0 \quad \text{vs} \quad H_a : C_j \neq 0$$

$$\frac{\hat{C}_j - C_j}{\sqrt{\hat{V}(\hat{C}_j)}} \cap t(GLE) \xrightarrow{\text{sob } H_0} t = \frac{\hat{C}_j}{\sqrt{\hat{V}(\hat{C}_j)}}$$

$$\hat{V}(\hat{C}_j) = \frac{QME}{r} (c_{j1}^2 + K + c_{jt}^2) = \frac{QME}{r} \left(\sum_{i=1}^t c_{ji}^2 \right)$$

[Para igual número de
repetições por tratamento]

$$\hat{V}(\hat{C}_j) = QME \left(\frac{c_{j1}^2}{r_1} + K + \frac{c_{jt}^2}{r_t} \right) = QME \left(\sum_{i=1}^t \frac{c_{ji}^2}{r_i} \right)$$

(Para diferente número de
repetições por tratamento)

(1) $H_0 : C_1 = 0$ vs $H_a : C_1 \neq 0$

$$C_1 = \mu_A + \mu_B + \mu_C - 3\mu_D \quad r=6$$

$$\hat{C}_1 = \bar{y}_A + \bar{y}_B + \bar{y}_C - 3\bar{y}_D = 72 + 83 + 76 - 3(68) = 47$$

$$\hat{V}(\hat{C}_1) = \frac{QME}{6} [1^2 + 1^2 + 1^2 + (-3)^2] = \frac{QME}{6} (12) = 2QME = 2(100,9) = 201,8$$

t.01(20)=2,845 GLE=20

$$t = \frac{\hat{C}_1}{\sqrt{\hat{V}(\hat{C}_1)}} = \frac{47}{\sqrt{201,8}} = 3,31^{**}$$

└─ rejeita-se H_0

$$t^2 = (3,31)^2 = 10,94 = F$$

(2) $H_0 : C_2 = 0$ vs $H_a : C_2 \neq 0$

$$C_2 = \mu_A + \mu_C - 2\mu_B$$

$$\hat{C}_2 = \bar{y}_A + \bar{y}_C - 2\bar{y}_B = 72 + 76 - 2(85) = -22$$

$$\hat{V}(\hat{C}_2) = \frac{QME}{6} [1^2 + 1^2 + (-2)^2] = \frac{QME}{6} (6) = QME = 100,9$$

t.05(20)=2.086

$$t = \frac{\hat{C}_2}{\sqrt{\hat{V}(\hat{C}_2)}} = \frac{-22}{\sqrt{100,9}} = -2,19^*$$

└─ rejeita-se H_0

$$t^2 = (-2,19)^2 = 4,80 = F$$

(3) $H_0 : C_3 = 0$ vs $H_a : C_3 \neq 0$

$$C_3 = \mu_A - \mu_C$$

$$\hat{C}_3 = \bar{y}_A - \bar{y}_C = 72 - 76 = -4$$

$$\hat{V}(\hat{C}_3) = \frac{QME}{6} [1^2 + (-1)^2] = \frac{2.QME}{6} = \frac{2(100,9)}{6} = 33,63$$

t.05(20)=2,086

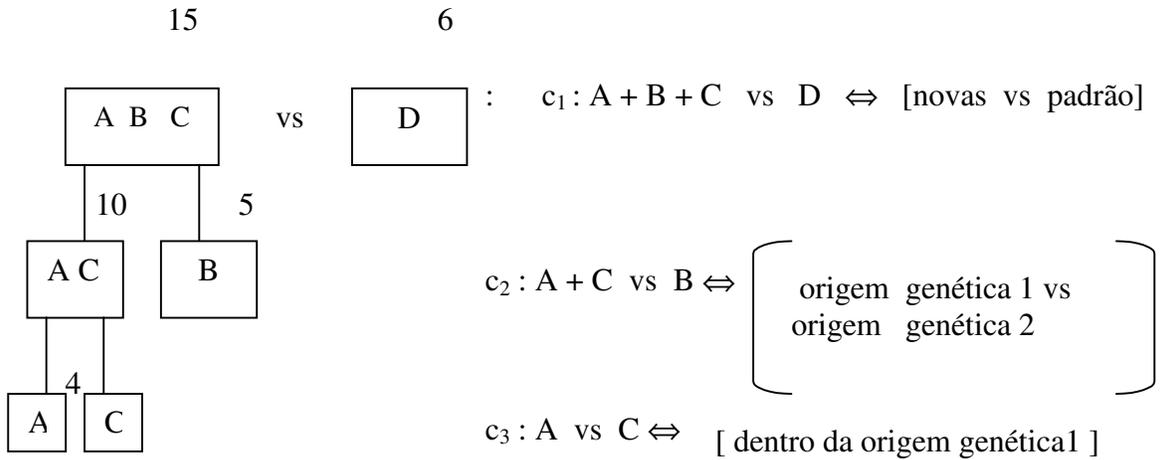
$$t = \frac{\hat{C}_3}{\sqrt{\hat{V}(\hat{C}_3)}} = \frac{-4}{\sqrt{33,63}} = -0,69^{NS}$$

└─ aceita-se H_0

$$t^2 = (-0,69)^2 = 0,48 = F$$

CONTRASTES PARA DIFERENTES NÚMEROS DE REPETIÇÕES POR TRATAMENTO

Exemplo 11: Consideremos o exemplo 4 das variedades de cana-de-açúcar com diferente número de repetições.



Cálculo das SQ dos contrastes usando os coeficientes

$$C_1 = 6 (4 \mu_A + 5 \mu_B + 6 \mu_C) - 15 (6 \mu_D)$$

$$= 4 \mu_A + 5 \mu_B + 6 \mu_C - 15 \mu_D$$

$$C_2 = 5 (4 \mu_A + 6 \mu_C) - 10 (5 \mu_B)$$

$$= 4 \mu_A + 6 \mu_C - 10 \mu_B = 2 \mu_A + 3 \mu_C - 5 \mu_B$$

$$C_3 = 6 (4 \mu_A) - 4 (6 \mu_C) = \mu_A - \mu_C$$

Verificação de Ortogonalidade:

$$\sum \frac{b_{ji} c_{ji}}{r_i} = 0$$

Expressão para cálculo da SQ :

$$SQ_{c_j} = \frac{\left(\sum_{i=1}^t c_{ji} \bar{y}_i \right)^2}{\sum_{i=1}^t \frac{c_{ji}^2}{r_i}}$$

Variedades	r _i	\bar{y}_i	c _{1i}	c _{2i}	c _{3i}	c _{1i} \bar{y}_i	c _{2i} \bar{y}_i	c _{3i} \bar{y}_i
A	4	70,25	4	4	1	281	281	70,25
B	5	86,60	5	-10	0	433	-866	0
C	6	76	6	6	-1	456	456	-76
D	6	62	-15	0	0	-930	0	0
Total	21	-	0	0	0	240	-129	-5,75

$$SQ_{c_1} = \frac{(240)^2}{\frac{4^2}{4} + \frac{5^2}{5} + \frac{6^2}{6} + \frac{(-15)^2}{6}} = 1097$$

$$SQ_{c_2} = \frac{(-129)^2}{\frac{4^2}{4} + \frac{(-10)^2}{5} + \frac{6^2}{6}} = 554,7$$

$$SQ_{c_3} = \frac{(-5,75)^2}{\frac{1^2}{4} + \frac{(-1)^2}{6}} = 79,3$$

1731 = SQT

Contraste	GL	SQ	QM	F	\hat{c}_j	$\sqrt{\hat{v}(\hat{c}_j)}$	t	decisão
c1	1	1097	1097	16,23	240	59,44	4,038	*
c2	1	554,7	554,7	8,24	-129	49,93	-2,871	*
c3	1	79,3	79,3	1,18	-5,75	5,30	-1,086	NS
				F.01(1,17)=8,40			t.01(17)=2,898	
				F.05(1,17)=4,45			t.05(17)=2,110	

5.7. Modelo Linear Aditivo Para A Análise Da Variância Do Delineamento Completamente Casualizado.

Para a análise de um delineamento completamente casualizado pressupõe-se, para cada observação Y_{ij} , o seguinte modelo linear aditivo:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (5.7.1)$$

onde Y_{ij} é o valor observado na j -ésima repetição ($j=1,2,\dots,r$) do tratamento \underline{i} ($i=1,2,\dots,t$). μ é a média geral sobre todas populações, τ_i é o efeito do tratamento i e representa o desvio da população com média μ_i em relação à média geral μ , mais um desvio aleatório ε_{ij} em relação à média da população \underline{i} .

O desvio ε_{ij} é também denominado efeito da unidade experimental j submetida ao tratamento \underline{i} . Pressupõe-se que estes desvios aleatórios têm uma distribuição normal e independente, com média zero e variância σ^2 .

Os desvios τ_i conhecidos por efeitos dos tratamentos i, representam a diferença $\mu_i - \mu$ entre a média da população i e a média geral, uma vez que o modelo também poderia ser escrito por

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad (5.7.2)$$

onde $\mu_i = \mu + \tau_i$.

As pressuposições sobre os efeitos dos tratamentos τ_i produzem dois modelos básicos para a análise de variância:

- 1) Modelo fixo ou Modelo I: Os efeitos τ_i são fixos e $\sum \tau_i = 0$.
- 2) Modelo aleatório ou Modelo II: Os efeitos τ_i são variáveis aleatórias de uma população de τ_s , com distribuição normal e independente, com média zero e variância σ_τ^2 .

A distinção está em que no modelo fixo, uma repetição do experimento introduziria os mesmos τ_i no novo experimento; toda atenção está concentrada nos τ_i . É o caso mais comum na experimentação. O experimentador escolhe deliberadamente os tratamentos, tais como, adubos, espaçamentos de plantio, rações, cultivares, medicamentos, etc. No modelo aleatório, a repetição do estudo introduziria um novo conjunto de τ_s , mas da mesma população de τ_s . Neste caso interessa a variabilidade dos τ_s , pois os tratamentos variam de um experimento a outro. Quando se estuda, por exemplo, adaptação generalizada de uma variedade de trigo, os locais em que é testada constituem uma amostra de todos os locais possíveis da região.

Um experimento conduzido no modelo fixo, estima os parâmetros μ , τ_i e σ^2 . No modelo aleatório estimam-se os parâmetros μ , σ^2 e σ_τ^2 a variância da população de tratamentos.

Considerando-se o modelo 5.7.1 o valor estimado (predito) para a repetição j do tratamento i é

$$\hat{Y}_{ij} = \hat{\mu} + \hat{\tau}_i \quad (5.7.3)$$

onde considerando-se que $\sum \hat{\tau}_i = 0$ tem-se $\hat{\mu} = \bar{Y}_{..}$ e $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$

Logo

$$\hat{Y}_{ij} = \hat{\mu} + \hat{\tau}_i = \bar{Y}_{..} + (\bar{Y}_{i.} - \bar{Y}_{..}) = \bar{Y}_{i.}$$

e os resíduos ou desvios

$$\hat{\varepsilon}_{ij} = e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_{i.}$$

Também

$$\sum_{ij} e_{ij}^2 = \sum_{ij} (Y_{ij} - \hat{Y}_{ij})^2 = \sum_{ij} (Y_{ij} - \bar{Y}_{i.})^2 = \text{SQE}$$

Para se estruturar o teste F na análise de variância necessita-se dos quadrados médios esperados que são estimados pelos quadrados médios calculados na análise da variância. Assim, o quadrado médio do erro experimental QM_E é uma estimativa de σ^2 , a variância dos componentes aleatórios das observações Y . O quadrado médio para tratamentos QM_T é uma estimativa de

$\sigma^2 + r \sigma_\tau^2$, no Modelo Aleatório e de $\sigma^2 + r \sum_{t-1} \frac{\tau_i^2}{t-1}$ no modelo fixo.

As variâncias dos efeitos aleatórios do modelo σ^2 e σ_τ^2 são chamados **componentes de variância**.

$\sum \tau_i^2$ não é um **componente de variância**, isto porque os τ_i não são aleatórios, mas sim, constantes fixas no modelo fixo.

Os quadrados médios esperados para os modelos fixos e aleatórios encontram-se na Tabela 5.14.

Tabela 5.14 Quadrados médios esperados nos Modelos fixo e aleatório no delineamento completamente casualizado com número igual de repetições.

Causas de Variação	GL	QM	Quadrados Médios Esperados	
			Modelo Fixo	Modelo Aleatório
Tratamentos	t-1	QM_T	$\sigma^2 + r \sum_{t-1} \frac{\tau_i^2}{t-1}$	$\sigma^2 + r \sigma_\tau^2$
Erro Experimental	t(r-1)	QM_E	σ^2	σ^2
Total	rt-1			

Nota-se que na classificação “tratamentos” dois elementos contribuem para a magnitude do quadrado médio esperado. Um é a variação entre os τ_i , isto é, entre os efeitos dos tratamentos. O outro é a variação σ^2 entre as unidades experimentais sobre as quais foram aplicados os tratamentos. Ambos são inseparáveis. Quando se pressupõe o Modelo fixo, a soma dos quadrados $\sum \tau_i^2$ representa a variação entre os tratamentos específicos do experimento. Quando se pressupõe o Modelo Aleatório σ_τ^2 representa a variância verdadeira entre todos os possíveis tratamentos, dos quais os t tratamentos usados no experimento constituem uma amostra.

Pela análise dos quadrados médios esperados fica evidente o mecanismo do F-teste. Para o modelo fixo

$$F = \frac{\sigma^2 + r \sum_{t-1} \frac{\tau_i^2}{t-1}}{\sigma^2}$$

A hipótese de nulidade que se formula no modelo fixo é $H_0: \tau_i = 0$, que é nulo o efeito dos tratamentos em comparação. Se a hipótese é verdadeira, $F = \frac{\sigma^2}{\sigma^2} = 1$, e H_0 não será rejeitada.

Para o modelo aleatório formula-se a hipótese $H_0 : \sigma_\tau^2 = 0$, de que não há variância entre os tratamentos. Se for verdadeira a hipótese,

$$F = \frac{(\sigma^2 + \sigma_\tau^2)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1$$

e H_0 não será rejeitada.

Em ambos os casos a hipótese será rejeitada quando a contribuição ao numerador da variação dos tratamentos é suficientemente grande para resultar num F significativo.

Na análise da variância em discussão, as hipóteses para os Modelos fixos e aleatório são testadas pelo mesmo valor F. Em análises mais complexas, o teste poderá tomar forma diferente, de acordo com o modelo adotado.

Em muitas situações pode-se estar interessado em estimar os componentes de variância. No caso, utilizando o método da Análise de Variância, tem-se

$$E(QME) = \sigma^2 \Rightarrow \hat{\sigma}^2 = QME$$

$$E(QMT) = \sigma^2 + r\sigma_\tau^2 \Rightarrow \hat{\sigma}^2 + r\hat{\sigma}_\tau^2 = QMT \Rightarrow \hat{\sigma}_\tau^2 = \frac{QMT - QME}{r}$$

Os quadrados médios esperados no caso de diferentes números de repetições por tratamento aparecem na Tabela 5.15.

Tabela 5.15 Quadrados Médios esperados nos modelos fixo e aleatório no delineamento completamente casualizado com diferente número de repetições.

Causas de Variação	GL	Quadrados Médios Esperados	
		Modelo Fixo	Modelo Aleatório
Tratamentos	t-1	$\sigma^2 + \frac{\sum r_i \tau_i^2}{t-1}$	$\sigma^2 + \varphi \sigma_\tau^2$
Erro experimental	n-t	σ^2	σ^2
Total	n-1		

Onde $\varphi = \frac{\sum r_i - \frac{\sum r_i^2}{\sum r_i}}{t-1}$.

5.8. Exercícios

1) Num experimento de alimentação de suínos, em que se usaram quatro rações (A, B, C, D) cada uma fornecida a cinco animais ao acaso e mantidos em baias individuais, os aumentos de peso observados, em quilogramas, foram os seguintes:

RAÇÕES			
A	B	C	D
27	35	39	40
12	19	27	35
13	31	20	46
28	15	29	41
30	30	45	33
TOTAL 110	130	160	195

1.1 Construa a tabela de Análise de Variância.

1.2 Formule as hipóteses, proceda ao teste para Tratamentos (Rações) a 5% e conclua.

1.3 Se pertinente, complemente a Análise de Variância a 5%, com o procedimento de

1.3.1 Tukey

1.3.2 Duncan

Apresente os resultados pelo sistema de letras e conclua.

1.4 Admita que a ração D é uma ração padrão e que as demais são novas formulações, tais que a fonte protéica de B é diferente de A e C. Complemente a Análise de Variância pela técnica de contrastes ortogonais, construindo o conjunto lógico de contrastes, estabelecendo os coeficientes, procedendo ao teste de significância dos contrastes a 5% e concluindo.

1.5 Suponha que o primeiro animal da ração A morreu durante a execução do experimento.

1.5.1 Construa a tabela de Análise de Variância.

1.5.2 Formule as hipóteses, proceda ao teste para Tratamentos (Rações) a 5% e conclua.

1.5.3 Se pertinente, complemente a Análise de Variância a 5%, com o procedimento de Tukey. Apresente os resultados pelo sistema de letras e conclua.

1.5.4 Repita o item (1.4) nesta nova realidade.

2) Use o teste S-N-K, o teste de Duncan, o teste de Tukey e os testes DMS, Scheffé e Bonferroni, todos ao nível de significância de 5%, para comparar as seguintes médias de tratamentos:

A	B	C	D	E	F	G
173	132	103	151	129	107	124

Foi empregado o Delineamento Completamente Casualizado, com cinco repetições por tratamento, e o quadrado médio do erro experimental foi de 245.

a) Apresente os resultados dos testes, assinalando com mesma letra as médias de tratamentos que não apresentam diferenças significativas, e conclua em cada caso.

b) Há diferenças nas conclusões pelos testes com relação ao número de diferenças significativas?