Agrobiologia

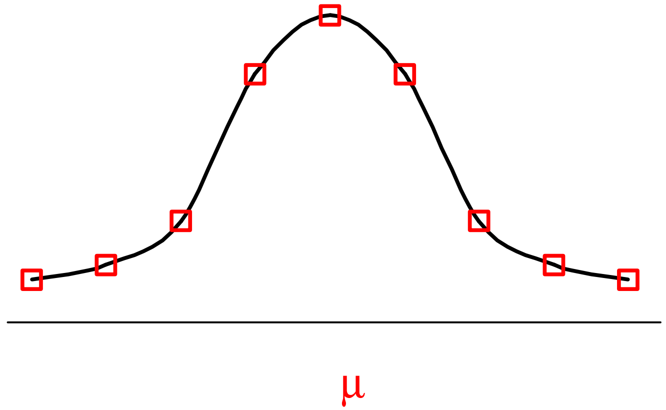
Ministério da Agricultura, Pecuária e Abastecimento

Governo Federal

Documentos

ISSN 1517-8498
Novembro/2003
163

Técnicas Experimentais aplicadas às Ciências Agrárias



Empresa Brasileira de Pesquisa Agropecuária
Centro Nacional de Pesquisa em Agrobiologia
Ministério da Agricultura, Pecuária e Abastecimento

ISSN 1517-8498
Novembro/2003

Documentos 163

Técnicas Experimentais aplicadas às Ciências Agrárias

Janaína Ribeiro Costa

Seropédica – RJ
2003

Exemplares desta publicação podem ser adquiridas na:

Embrapa Agrobiologia
BR465 – km 7
Caixa Postal 74505
23851-970 – Seropédica/RJ, Brasil
Telefone: (0xx21) 2682-1500
Fax: (0xx21) 2682-1230
Home page: www.cnpab.embrapa.br
e-mail: sac@cnpab.embrapa.br

Comitê Local de Publicações: Eduardo F. C. Campello (Presidente)
José Guilherme Marinho Guerra
Maria Cristina Prata Neves
Verônica Massena Reis
Robert Michael Boddey
Maria Elizabeth Fernandes Correia
Dorimar dos Santos Felix (Bibliotecária)

Expediente:
Revisor e/ou ad hoc: Guilherme Montandon Chaer
Normalização Bibliográfica: Dorimar dos Santos Félix
Editoração eletrônica: Marta Maria Gonçalves Bahia

1ª impressão (2003): 50 exemplares

COSTA, J. R. **Técnicas experimentais aplicadas às ciências agrárias.** Seropédica: Embrapa Agrobiologia, 2003. 102 p. (Embrapa Agrobiologia. Documentos, 163).

ISSN 1517-8498

1. Agricultura. 2. Ciência agrária. I. Embrapa. Centro Nacional de Pesquisa de Agrobiologia (Seropédica, RJ). II. Título. III. Série.

CDD 630

6. Referências Bibliográficas

BANZATTO, A. D.; KRONKA, S. do N. **Experimentação agrícola**. Jaboticabal: FUNEP, 1989. 249 p.

BEARZOTI, E.; OLIVEIRA, M. S. **Estatística básica**. Lavras: UFLA, 1997. 191 p.

FISHER, R. A. **The design of experiments**. Edinburgh: Oliver and Boyd, 1935.

HINKELMANN, K.; KEMPTHORNE, O. Design and analysis of experiments. New York: J. Wiley, 1994. 631 p.

MEAD, R.; CURNOW, R. N. **Statistical methods in agriculture and experimental biology.** New York: Chapman and Hall, 1983. 335 p.

NOGUEIRA, M. C. S. **Estatística experimental aplicada à experimentação agrícola.** Piracicaba: USP-ESALQ, 1997. 250 p.

PIMENTEL GOMES, F. **Curso de estatística experimental.** 13. ed. Piracicaba: Nobel/USP-ESALQ, 1990. 468 p.

RAMALHO, M. A.; FERREIRA, D. F.; OLIVEIRA, A. C. de. **A experimentação em genética e melhoramento de plantas.** Lavras: UFLA, 2000. 326 p.

STEEL, R. G. D.; TORRIE, J. H.; DICKEY, D. A. **Principles and procedures of statistics.** 3. ed. New York: McGraw-Hill, 1997. 666 p.

Autor

Janaína Ribeiro Costa

Pesquisadora da Embrapa Agrobiologia.

E-mail: janaina@cnpab.embrapa.br

Obtenção da produção máxima de milho (ton/ha)

Aqui cabe esclarecer que o sinal da estimativa do coeficiente $\hat{a}_2$ determina se a variável dependente y (no exemplo, produção), terá um valor máximo ou mínimo. Se $\hat{a}_2$ é negativo, y terá um máximo; caso contrário, se $\hat{a}_2$ for positivo, y terá um mínimo. No exemplo 5.7.3.1), para obtenção da produção máxima de milho é necessário antes maximizar a função de regressão polinomial quadrática, ou seja, derivar esta equação e igualar a zero:

$$\hat{y}_i = 8,8421 + 0,0950X_i - 0,00050X_i^2$$

$$\frac{d\hat{y}_i}{dX_i} = 0 + 0,0950 - 0,00100X_i$$

$$\frac{d\hat{y}_i}{dX_i} = 0 \quad \Rightarrow \quad 0 + 0,0950 - 0,00100X_i = 0$$

$X_i = \dfrac{0,0950}{0,00100} = 95$ kg/ha (Dose de adubo nitrogenado que levará a uma produção máxima).

Substituindo X_i = 95 na equação de $\hat{y}_i$ obtém-se a produção máxima de milho:

$$\hat{y}_i = 8,8421 + 0,0950.(95) - 0,00050.(95)^2$$

$$\hat{y}_i = 8,8421 + 0,0950.(95) - 0,00050.(95)^2$$

$\hat{y}_i = 13,3546$ ton/ha (produção máxima de milho para dose de adubo nitrogenado de 95 kg/ha).

$$\hat{b}_1 = \frac{\sum_{i=1}^{5} P_1(X_i).y_i}{3\sum_{i=1}^{5} P_1^2(X_i)} = \frac{(-2).27,5+(-1).32,0+(0).37,5+(1).42,6+(2).37,8}{3.[(-2)^2+(-1)^2+(0)^2+(1)^2+(2)^2]} = \frac{31,2}{3.(10)} = 1,0400$$

$$\hat{b}_2 = \frac{\sum_{i=1}^{5} P_2(X_i).y_i}{3\sum_{i=1}^{5} P_2^2(X_i)} = \frac{(2).27,5+(-1).32,0+(-2).37,5+(-1).42,6+(2).37,8}{3.[(2)^2+(-1)^2+(-2)^2+(-1)^2+(2)^2]} = \frac{-19,0}{3.(14)} = -0,4523$$

Lembrando que:

$$P_1(X_i) = x_i = \frac{X_i - 60}{30} = \frac{X_i}{30} - 2$$

$$P_2(X_i) = x_i^2 - \frac{n^2-1}{12} = \left(\frac{X_i}{30} - 2\right)^2 - 2$$

Portanto:

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 P_1(X_1) + \hat{b}_2 P_2(X_1)$$

$$\hat{y}_i = 11,8267 + 1,0400\left(\frac{X_i}{30} - 2\right) - 0,4523\left[\left(\frac{X_i}{30} - 2\right)^2 - 2\right]$$

Resolvendo a equação acima tem-se:

$\hat{y}_i = 8,8421 + 0,0950X_i - 0,00050X_i^2$ (Equação da Regressão Quadrática)

$\hat{y}_i = \hat{a}_o \quad + \hat{a}_1 X_i \quad - \quad \hat{a}_2 X_i^2$

Os valores observados (y_i) e estimados $(\hat{y}_i)$ para cada dose de adubo nitrogenado estão apresentados a seguir:

X_i	y_i	$\hat{y}_i$
0	27,5	8,8421
30	32,0	11,2421
60	37,5	12,7421
90	42,6	13,3421
120	37,8	13,0421

Apresentação

A preocupação crescente da sociedade com a preservação e a conservação ambiental tem resultado na busca pelo setor produtivo de tecnologias para a implantação de sistemas de produção agrícola com enfoque ecológicos, rentáveis e socialmente justos. O enfoque agroecológico do empreendimento agrícola se orienta para o uso responsável dos recursos naturais (solo, água, fauna, flora, energia e minerais).

Dentro desse cenário, a Embrapa Agrobiologia orienta sua programação de P&D para o avanço de conhecimento e desenvolvimento de soluções tecnológicas para uma agricultura sustentável.

A agricultura sustentável, produtiva e ambientalmente equilibrada apoia-se em práticas conservacionistas de preparo do solo, rotações de culturas e consórcios, no uso de adubação verde e de controle biológico de pragas, bem como no emprego eficiente dos recursos naturais. Infere-se daí que os processos biológicos que ocorrem no sistema solo/planta, efetivados por microrganismos e pequenos invertebrados, constituem a base sobre a qual a agricultura agroecológica se sustenta.

O documento 163/2003 atende uma demanda daqueles que atuam na pesquisa agropecuária, principalmente estudantes e profissionais recém ingressados na área, disponibilizando, de forma objetiva e prática, conceitos de estatística aplicados à experimentação em Ciências Agrárias. Na verdade, existem poucas publicações sobre o referido tema e este documento serve de roteiro para orientar aspectos básicos do planejamento da experimentação de campo e análise dos resultados obtidos.

S U M Á R I O

1) Noções básicas de experimentação agrícola 7
2) Distribuição de freqüências 10
2.1) Definição 10
2.2) Freqüência 10
2.3) Natureza da distribuição 19
3) Estatísticas descritivas **20**
3.1) Medidas de posição 20
3.2) Medidas de dispersão 23
3.3) Medidas de assimetria e curtose 27
4) Testes de comparações múltiplas **29**
4.1) Contrastes ortogonais de médias 29
4.2) Teste t de Student 33
4.3) Teste de Tukey 39
4.4) Teste de Duncan 41
4.5) Teste de SNK (Student Newman Keuls) 43
4.6) Teste de Scott-Knott 46
5) Análise de variância 54
5.1) Princípios básicos da experimentação 54
5.2) Pressuposições básicas da análise de variância 55
5.3) Delineamento Inteiramente Casualizado (DIC) 56
5.4) Delineamento em Blocos Casualizados (DBC) 61
5.5) Experimentos fatoriais 66
5.6) Experimentos em parcelas subdivididas 74
5.7) Análise de regressão 87
6) Referências Bibliográficas 102

FV	GL	SQ	QM	F	Prob>F
(Doses)	4	45,3160	11,3290	13,122	0,0005
Regressão Linear	1	32,4480	32,4480	37,586	0,0000
Regressão Quadrática	1	8,5952	8,5952	9,956	0,0100
Desvio	2	4,2728	2,1364	2,475	0,1340
Erro	10	8,6333	0,8633		
Total	14				
CV (%) =	7,86				
Média geral:	11,83		Número de observações:		15

Observa-se no quadro anterior que tanto a regressão linear quanto a quadrática foram significativas ao nível de significância estabelecido de 5% (Prob<0,05). O coeficiente de determinação (R^2) para a regressão linear e quadrática foram respectivamente:

$$R^2(\text{Linear}) = \frac{32,4480}{45,3160}.100 = 71,6\%\ ;$$

$$R^2(\text{Quadrática}) = \frac{8,5952}{45,3160}.100 = 19,0\%\ .$$

Apesar do R^2 da regressão quadrática ter sido baixo (19%), deve-se observar o valor de Prob>F do Desvio. Se este valor for maior que 0,05, indicando que o desvio foi não significativo, deve-se, portanto, considerar a equação de regressão significativa de maior grau, no caso, a quadrática:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 X^2 = \hat{b}_0 + \hat{b}_1 P_1(X_1) + \hat{b}_2 P_2(X_1)$$

em que:

$$\hat{b}_o = \bar{y} = \frac{\sum_{i=1}^{5} y_i}{3x5} = \frac{177,4}{15} = 11,8267$$

Para cada nível de X tem-se então:

Níveis	Dose de adubo	Totais y_i (das 3 repetições)	$P_1(X_i)$	$P_2(X_i)$
1	0	27,5	-2	+2
2	30	32,0	-1	-1
3	60	37,5	0	-2
4	90	42,6	+1	-1
5	120	37,8	+2	+2
Total		177,4		

As somas de quadrados (SQ's) da regressão linear e quadrática são dadas por:

$$\text{SQRegressão 1(Linear)} = \frac{\left(\sum_{i=1}^{5} P_1(X_i).y_i\right)^2}{3\sum_{i=1}^{5} P_1^2(X_i)}$$

$$= \frac{[(-2).27,5+(-1).32,0+(0).37,5+(1).42,6+(2).37,8]^2}{3.[(-2)^2+(-1)^2+(0)^2+(1)^2+(2)^2]} = \frac{973,44}{3.(10)} = 32,4480$$

$$\text{SQRegressão 2 (Quadrática)} = \frac{\left(\sum_{i=1}^{5} P_2(X_i).y_i\right)^2}{3\sum_{i=1}^{5} P_2^2(X_i)}$$

$$= \frac{[(2).27,5+(-1).32,0+(-2).37,5+(-1).42,6+(2).37,8]^2}{3.[(2)^2+(-1)^2+(-2)^2+(-1)^2+(2)^2]} = \frac{361}{3.(14)} = 8,5952$$

SQDesvio = SQ Doses – SQRegressão 1 – SQRegressão 2

= 45,3160 – 32.4480 – 8,5952 = 4,2728

O quadro de análise de variância com o desdobramento dos graus de liberdade da fonte de variação 'doses' em graus de liberdade devido a regressão polinomial está apresentado a seguir.

Técnicas Experimentais aplicadas às Ciências Agrárias

Janaína Ribeiro Costa

1. Noções básicas de experimentação agrícola

A Estatística Experimental é a ciência que tem como objetivo estudar experimentos (ensaios), englobando etapas como o planejamento, execução, coleta e análise dos dados experimentais e interpretação dos resultados obtidos. Ela foi proposta inicialmente na área de ciências biológicas por Ronald A. Fisher em 1919. Fisher propôs o uso da análise de variância (ANAVA) como ferramenta para análise e interpretação de dados. A ANAVA permite a decomposição do grau de liberdade e da soma de quadrados total em somas de quadrados correspondentes às fontes de variação previamente definidas no planejamento do experimento.

A fase de planejamento do experimento merece considerável atenção por parte do pesquisador pois dela dependerá o sucesso da análise e interpretação dos resultados sendo, portanto, recomendável uma consulta a um estatístico antes da instalação do experimento.

O planejamento envolve etapas como:

a) Formulação de hipóteses

A hipótese estatística formulada é denominada hipótese de nulidade e é simbolizada por H_o. Suponha que se deseja estudar qual estirpe de bactéria diazotrófica endofítica (considerando, por exemplo, três estirpes diferentes) proporcionará maior peso da parte área de cana-de-açúcar. No exemplo, H_o seria: não existem diferenças significativas entre os efeitos das estirpes (ou seja, qualquer diferença observada é devida a fatores não controlados). H_o poderá ser aceita ou rejeitada; caso seja rejeitada, aceitaremos uma hipótese denominada alternativa, simbolizada por H_1 que no exemplo seria: os efeitos das estirpes diferem significativamente entre si (ou as estirpes se comportam de modo diferente quanto ao peso da parte aérea).

b) Escolha dos fatores e seus respectivos níveis

Fatores (ou tratamentos) são aqueles que o pesquisador tem interesse em estudar o seu efeito sobre as variáveis respostas. As subdivisões de um fator são os níveis dos mesmos. Por exemplo, se o interesse for planejar um experimento para se estudar o efeito de 6 tipos diferentes de rotações de cultura, o fator em estudo é rotação e os níveis deste fator são os 6 tipos de rotação.

Em alguns casos, como por exemplo nos experimentos fatoriais ou em parcelas subdivididas, dois ou mais fatores são estudados. Suponha que se deseja estudar o efeito de 2 variedades de cana de açúcar e 3 doses de nitrogênio; neste caso se trata de um experimento em fatorial 2x3, em que se tem dois fatores (variedade e dose de nitrogênio); 2 níveis do fator variedade e 3 níveis do fator dose de nitrogênio.

Um fator pode ser classificado em:

b.1) Qualitativo: quando os níveis do fator são categorias, atributos. Por exemplo: nome de variedades de cana de açúcar (SP701143 e SP813250); métodos de extração de DNA (Cullen, Smalla, Sebach); origem de solos (MG, RJ, BA, SP); etc.

b.2) Quantitativo: quando os níveis do fator são mensurações de valores reais. Normalmente os níveis são valores numéricos acompanhados de uma unidade de medida. Por exemplo: dose de nitrogênio (0, 25 e 50 Kg/ha); concentrações de antibiótico (25, 50, 100, 200 μg/ml), etc.

c) Escolha da parcela (unidade experimental)

Parcela é a unidade experimental que receberá o tratamento. A parcela pode assumir diferentes formas e tamanhos. Por exemplo, uma parcela poderá ser constituída por uma ou várias plantas; um vaso contendo uma ou mais plantas; uma placa de Petri com

tem-se que $P_1(X_i) = x_i$; em que $x_i = \frac{X_i - \overline{X}}{q} = \frac{X_i - 60}{30}$ com i =1, 2, ..., 5. Portanto,

$$P_1(X_1 = 0) = \frac{0-60}{30} = -2$$

$$P_1(X_2 = 30) = \frac{30-60}{30} = -1$$

$$P_1(X_3 = 60) = \frac{60-60}{30} = 0$$

$$P_1(X_4 = 90) = \frac{90-60}{30} = +1$$

$$P_1(X_5 = 120) = \frac{120-60}{30} = +2$$

e tem-se que $P_2(X_i)$ = $x_i^2 - \frac{n^2-1}{12} = [P_1(X_i)]^2 - \frac{5^2-1}{12} = [P_1(X_i)]^2 - 2$ com i =1, 2, ..., 5. Portanto,

$$P_2(X_1 = 0) = (-2)^2 - 2 = +2$$

$$P_2(X_2 = 30) = (-1)^2 - 2 = -1$$

$$P_2(X_3 = 60) = (0)^2 - 2 = -2$$

$$P_2(X_4 = 90) = (+1)^2 - 2 = -1$$

$$P_2(X_5 = 120) = (+2)^2 - 2 = +2\,.$$

A análise de variância para os dados do exemplo 5.7.3.1) é:

FV	GL	SQ	QM	F	Prob>F
Doses	4	45,3160	11,3290	13,122	0,0005
Erro	10	8,6333	0,8633		
Total	14				

Rejeita-se H_o, concluindo-se pela existência do efeito de doses crescentes de adubo nitrogenado sobre a produção do milho (Prob < 0,05).

Considerando o modelo de regressão polinomial de $2^{\underline{o}}$ grau a seguir, foi realizada a análise de regressão:

$$y_i = \hat{a}_o + \hat{a}_1 X_i + \hat{a}_2 X_i^2 + \mathring{a}_i$$

reescrevendo este modelo pela expressão alternativa:

$$y_i = b_o + b_1 P_1(X_i) + b_2 P_2(X_i) + \varepsilon_i \quad ; \quad i = 1, 2, ..., 5.$$

As hipóteses testadas no modelo de regressão adotado são:

i) H_o: $b_1 = 0$ vs H_1: $b_1 \neq 0$.

ii) H_o: $b_2 = 0$ vs H_1: $b_2 \neq 0$.

Para obtenção das somas de quadrados das regressões linear e quadrática é necessário antes calcular os coeficientes dos polinômios $P_1(X_i)$ e $P_2(X_i)$.

Seja:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{5}(0 + 30 + 60 + 90 + 120) = 60\,;$$

q = 30 (correspondendo a 30-0 ou 60-30 ou 90-60 ou 120-90)

determinado meio de cultura; uma área com várias plantas; um animal; etc.

d) Escolha do delineamento experimental

Delineamento experimental é o plano de distribuição dos tratamentos na área experimental. Como exemplo de delineamentos tem-se o **d**elineamento **i**nteiramente **c**asualizado (DIC), o **d**elineamento em **b**locos **c**asualizados (DBC), o **d**elineamento em **q**uadrados **l**atinos (DQL), os delineamentos em blocos incompletos (por exemplo, os látices, blocos aumentados, etc.).

e) Escolha das variáveis a serem analisadas

Variáveis respostas ou variáveis dependentes ou simplesmente variáveis são características obtidas em cada parcela. Os dados (observações) são realizações de uma variável e serão analisados para verificar se há diferença entre os níveis dos fatores (tratamentos). Assim, exemplos de variáveis são: produção de grãos de feijão; altura de plantas de milho; pH, teor de Ca, Mg e P em amostras de solo; número de plantas de cana-de-açúcar atacadas por cercosporiose; etc.

Uma variável também pode ser classificada, semelhantemente aos fatores (tratamentos), em:

e.1) Qualitativa

e.1.1) Nominal**:** quando são categorias, atributos, sem uma ordenação natural. Por exemplo: cor dos grãos do feijoeiro (marrom, preto, branco); textura do solo (arenoso, argiloso, silte); etc.

e.1.2) Ordinal**:** quando são atributos com uma ordenação natural. Por exemplo: suscetibilidade do cafeeiro à ferrugem (alta, média, baixa); nota para o ataque de cercosporiose em cana-de-açúcar (escala de 1, para ausência da doença, até 9, para o máximo de doença); etc.

e.2) Quantitativa

e.2.1) Discretas**:** quando são contagens de números inteiros positivos com uma ordenação natural. Por exemplo: número de chuvas em 2002 superior a 80 mm/h (ex. 20 chuvas); número de plantas atacadas com a broca do fruto do cafeeiro (ex. 200 plantas); número de minhocas encontradas em determinada amostra de solo (ex. 50 minhocas).

e.2.2) Contínuas**:** quando são mensurações de valores reais; normalmente existe uma unidade de medida acompanhando a variável. Por exemplo: produtividade (100,0 kg/ha); renda (R$2050,73/mês); altura (2,5 m); diâmetro (8,18 cm); peso (98,5 g); pH (5,5); teor de P, Ca, Mg, K, matéria orgânica, etc.

f) Análise dos dados obtidos com o experimento.

2) Distribuição de freqüências

2.1) Definição

Consiste em uma função que associa os valores que uma variável assume com suas freqüências de ocorrência, podendo ser elas absolutas, relativas ou porcentuais.

2.2) Freqüência

É uma medida que quantifica a ocorrência dos valores de uma variável.

2.2.1) Freqüência absoluta (fa) é o número de observações ocorridos em cada classe da variável estudada.

2.2.2) Freqüência relativa (fr) é dada pela divisão da fa pelo número total (n) de dados ou observações:

$$fr = \frac{fa}{n}.$$

2.2.3) Freqüência porcentual (fp) é dada pela multiplicação de fr por 100:

$$fp\,(\%) = fr.100\,.$$

$$SQRegressão\,k = \frac{\left(\sum_{i=1}^{n} P_k(X_i).y_i\right)^2}{r\sum_{i=1}^{n} P_k^2(X_i)}$$, associada a 1 grau de liberdade.

O coeficiente de determinação (R^2) em experimentos com repetição é dado por:

$$R^2(\%) = \frac{SQ\,Regressão\,k}{SQTratamento}.100 \quad , \quad 0 \le R^2 \le 100.$$

5.7.3.1) Exemplo de análise de regressão em dados com repetição: modelos de regressão polinomial

Um experimento foi instalado conforme o delineamento inteiramente casualizado, com três repetições para testar o efeito de 5 doses de adubo nitrogenado (0, 30, 60, 90 e 120 kg/ha). Os resultados obtidos em ton/ha de milho são:

Rep\Doses	0	30	60	90	120
1	8,6	10,5	12,5	12,6	13,7
2	9,5	10,0	12,8	15,1	12,8
3	9,4	11,5	12,2	14,9	11,3
Total	27,5	32,0	37,5	42,6	37,8

O modelo do exemplo anterior adotado foi:

$$y_{ij} = \mu + d_i + \varepsilon_{ij} \quad ; \quad i = 1, 2, ..., 5 \quad e \quad j = 1, 2, 3.$$

em que y_{ij} é o valor observado referente a i-ésima dose de adubo nitrogenado na j-ésima repetição; d_i é a i-ésima dose de adubo nitrogenado e ε_{ij} é o erro experimental associado a y_{ij} com $\varepsilon_i \cap N(0, \sigma^2)$ e independentes.

As hipóteses testadas na análise de variância são:

H_o: $d_1 = d_2 = ... = dn = 0$; $i=1, 2, ..., n$

H_1: pelo menos um d_i difere de 0.

$$P_2(X_i) = x_i^2 - \frac{n^2-1}{12};$$

$$P_3(X_i) = x_i^3 - \frac{3n^2-7}{20}.x_i;$$

$$P_4(X_i) = x_i^4 - \frac{3n^2-13}{14}.x_i^2 + \frac{3(n^2-1)(n^2-9)}{560};$$

$$P_5(X_i) = x_i^5 - \frac{5(n^2-7)}{18}.x_i^3 + \frac{15n^4-230n^2+407}{1008}x_i;$$

em que,

X_i são os níveis da variável independente;

$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ é a média dos níveis de X;

q é amplitude entre dois níveis consecutivos de X;

n é o número de níveis da variável independente (X).

O estimador de quadrados mínimos de b_k, é dado por:

$$\hat{b}_k = \frac{\sum_{i=1}^{n} P_k(X_i).y_i}{r\sum_{i=1}^{n} P_k^2(X_i)},$$

em que,

$P_k(X_i)$ são os coeficientes do polinômio ortogonal de grau k associado ao nível do fator;

y_i é o total do nível i da variável dependente (y);

r é o número de repetições.

A hipótese de nulidade é H_o: $b_k = 0$ e a hipótese alternativa é H_1: $b_k \neq 0$. A soma de quadrados da regressão de grau k, na análise de variância, é dada por:

Exemplo 1. No quadro a seguir está disposta a atividade agrícola predominante em cada uma das 20 propriedades rurais do município Vida Alegre.

Milho	Soja	Olericultura	Leite
Soja	Soja	Milho	Milho
Leite	Cana-de-açúcar	Trigo	Milho
Milho	Leite	Soja	Trigo
Milho	Laranja	Milho	Olericultura

A variável em estudo, atividade agrícola, é classificada como qualitativa nominal. Uma maneira mais informativa de descrever o conjunto de dados do Exemplo 1 é através da distribuição de freqüências das categorias desta variável, podendo ser feita por meio de representação tabular ou gráfica.

a) Representação tabular:

Tabela 1. Distribuição de freqüência das atividades agrícolas de 20 propriedades rurais do município de Vida Alegre

Atividade predominante	fa	fr	fp (%)
Milho	7	0,3500	35,0
Soja	4	0,2000	20,0
Leite	3	0,1500	15,0
Trigo	2	0,1000	10,0
Olericultura	2	0,1000	10,0
Cana-de-açúcar	1	0,0500	5,0
Laranja	1	0,0500	5,0
Total	20	1,0000	100,0

Fonte: Apostila de Estatística Básica (Bearzoti & Oliveira, 1997).

b) Representação gráfica:

Gráfico é uma figura para ilustração de fenômenos ou tendências onde existem escalas definidas.

Para a representação gráfica de variáveis qualitativas, como é o caso do Exemplo 1, os gráficos mais utilizados são:

- Gráfico de linhas: possui dois eixos, com fa ou fr ou fp disposta no eixo vertical e as classes (categorias) da variável dispostas no eixo horizontal.

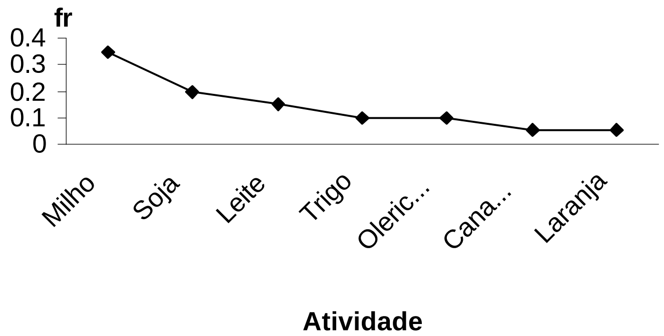


Figura 1. *Gráfico de linhas representando a distribuição de freqüência relativa referente à atividade agrícola predominante em propriedades do município de Vida Alegre.*

Análise de regressão em dados com repetição: modelos de regressão polinomial

O modelo de regressão polinomial de grau p, para uma única variável independente é representado por

$$y_i = \hat{a}_o + \hat{a}_1 X_i + \hat{a}_2 X_i^2 + \hat{a}_3 X_i^3 + \ldots + \hat{a}_p X_i^p + \mathring{a}_i \qquad \text{(i)}$$

com i = 1, 2, ..., n ; $\varepsilon_i \cap N(0, \sigma^2)$ independentes. β_0, β_1, ..., β_n são parâmetros da regressão a serem estimados.

Considerando n pares de dados (y_1, X_1), (y_2, X_2), ..., (y_n, X_n) em que n>p e que os níveis referentes a variável X são todos eqüidistantes, ou seja, $X_1=X_1$; $X_2=X_1+q$, $X_3=X_2+q$, ..., $X_n=X_{n-1}+q$, o modelo em (i) pode ser reescrito como:

$$Y_i= b_o + b_1P_1(X_i) + b_2P_2(X_i) + \ldots + b_pP_p(X_i) + \varepsilon_i$$

com i = 1, 2, ..., n ; $\varepsilon_i \cap N(0, \sigma^2)$ independentes. b_0, b_1, ..., b_n são parâmetros da regressão a serem estimados e $P_k(X_i)$ sendo um polinômio ortogonal de ordem k = 1, 2, ..., p que deve atender às seguintes restrições:

i) $P_0(X_i)=1$;

ii) $\sum_{i=1}^{n} P_k(X_i) = 0$;

iii) $\sum_{i=1}^{n} P_k(X_i).P_{K'}(X_i) = 0$ para $k \neq k'$;

iv) $\sum_{i=1}^{n} P_k^2(X_i) \neq 0$.

Os valores de $P_k(X_i)$ (k =1, 2, ...,p), quando os níveis da variável X são eqüidistantes, podem ser obtidos através das seguintes expressões:

$$P_1(X_i) = x_i \text{ ; em que } x_i = \frac{X_i - \overline{X}}{q} \text{ ;}$$

$$\hat{y}_i = \hat{b}_o + \hat{b}_1 X_i$$

em que

$$\hat{b}_1 = \frac{10.(1{,}388) + 15.(1{,}426) + ... + 55.(1{,}139) - \frac{(325).(12{,}577)}{10}}{10^2 + 15^2 + ... + 55^2 - \frac{(325)^2}{10}}$$

$$\hat{b}_1 = \frac{393{,}6650 - 408{,}7525}{2062{,}5000} = \frac{-15{,}0875}{2062{,}5000} = -0{,}0073$$ (estimativa de b_1),

$$\hat{b}_o = \frac{12{,}577}{10} - (-0{,}0073).\left(\frac{325}{10}\right) = 1{,}2577 + 0{,}2373 = 1{,}4950$$ (estimativa de b_o).

O modelo de regressão ajustado (estimado) é:

$$\hat{y}_i = 1{,}4950 - 0{,}0073X_i \,.$$

O R^2 foi de:

$$R^2 = \frac{0{,}1104}{0{,}1255}.100 = 90\%$$

indicando que 90% da variação na densidade do solo é explicada pelo modelo de regressão utilizado.

No quadro a seguir para cada valor de X_i tem-se o valor observado, o estimado e o desvio correspondente.

X_i	y_i (valores observados)	$\hat{y}_i$ (valores estimados)	$y_i - \hat{y}_i$
10	1,388	1,422	-0,034
15	1,426	1,386	0,040
20	1,393	1,349	0,044
25	1,341	1,313	0,029
30	1,26	1,276	-0,016
35	1,16	1,240	-0,080
40	1,177	1,203	-0,026
45	1,153	1,167	-0,014
50	1,14	1,130	0,010
55	1,139	1,094	0,045
Total	12,577	12,577	0
Média	1,2577	1,2577	0

- Gráfico de barras ou colunas: semelhantes aos gráficos de linhas, com a diferença que são usadas barras (colunas) ao invés de linhas.

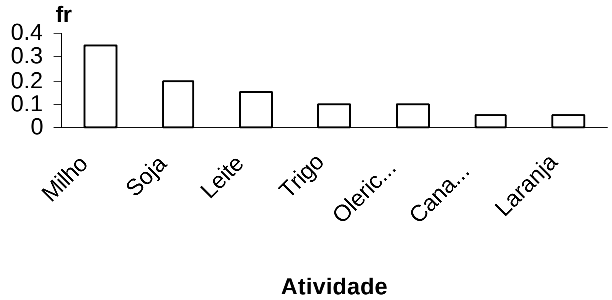


Figura 2. *Gráfico de barras verticais representando a distribuição de freqüência relativa referente à atividade agrícola predominante em propriedades do município de Vida Alegre.*

- Setograma (gráfico circular ou gráfico de setores): gráfico circular no qual os setores correspondem as categorias com áreas proporcionais as freqüências de cada classe. Para construção do setograma é necessário obter o ângulo referente ao setor de cada categoria, por meio de uma regra de três. Por exemplo, para a atividade milho do Exemplo 1, tem-se a regra de três para as freqüências porcentuais dada por:

$$100\% \;\text{———}\; 360^o$$
$$35\% \;\text{———}\; x$$
$$x = 126^o.$$

E assim por diante são calculados os outros ângulos correspondentes aos setores das outras categorias que serão traçados no gráfico.

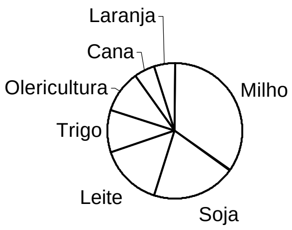


Figura 3. Setograma representando a distribuição de freqüência relativa referente à atividade agrícola predominante em propriedades do município de Vida Alegre.

Exemplo 2. O quadro seguinte apresenta o número de lagartas rosca encontradas em cada um dos 16 canteiros de um viveiro de mudas de eucalipto.

1	1	3	5
4	2	4	4
3	1	2	1
5	0	0	4

A variável número de lagartas rosca é classificada como quantitativa discreta. A distribuição de freqüências para variáveis quantitativas discretas são semelhantes à das variáveis qualitativas, como no caso do Exemplo 1, com os valores inteiros que a variável assume podendo ser considerados como “categorias” ou “classes naturais”.

a) Representação tabular:

Tabela 2. Distribuição de freqüência do número de lagartas rosca em canteiros de um viveiro de eucalipto

Nº de lagartas rosca	fa	fr	fp (%)
0	2	0,1250	12,5
1	4	0,2500	25,0
2	2	0,1250	12,5
3	2	0,1250	12,5
4	4	0,2500	25,0
5	2	0,1250	12,5
Total	16	1,0000	100,0

Fonte: Notas de aula.

H_1: $b_1 \neq 0$.

As somas de quadrados para o exemplo anterior foram:

$$\text{SQRegressão} = \frac{\left[10.(1{,}388) + 15.(1{,}426) + \ldots + 55.(1{,}139) - \frac{(325).(12{,}577)}{10}\right]^2}{10^2 + 15^2 + \ldots + 55^2 - \frac{(325)^2}{10}}$$

$$\text{SQRegressão} = \frac{[393{,}6650 - 408{,}7525]^2}{2062{,}5000} = \frac{(-15{,}0875)^2}{2062{,}5000} = 0{,}1104$$

$$\text{SQTotal} = 1{,}388^2 + 1{,}426^2 + \ldots + 1{,}139^2 - \frac{(12{,}577)^2}{10}$$

SQTotal = 15,9436 – 15,8181 = 0,1255

SQDesvios = 0,1255 – 0,1104 = 0,0151

O Quadro de análise de variância resultante é:

FV	GL	SQ	QM	F	Prob>F
Regressão	1	0,1104	0,1104	58,105	0,0001
Desvios	8	0,0151	0,0019		
Total	9	0,1255			

Da Tabela de F tem-se que $F_{(0{,}05;\ 1;\ 8)}$ é 5,32 e como 58,105 > 5,32, rejeita-se H_o ao nível de 5% de significância. Atualmente, os programas computacionais apresentam uma coluna a mais no quadro de análise de variância correspondente a Prob>F, não havendo a necessidade de procurar o valor de F em Tabela. Quando Prob>F for menor que 0,05, significa que o teste F foi significativo, ou seja, o pesquisador poderá rejeitar H_o e aceitar H_1. No exemplo, conclui-se então que as densidades (g/cm^3) em diferentes profundidades X (cm) podem ser explicadas por meio do seguinte modelo de regressão linear:

SQDesvios = $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = SQTotal - SQ\,Regessão$, associada a (n-2) graus de liberdade.

A decisão de rejeitar H_o ao nível α de significância se dará se

$$\frac{QM\,Regressão}{QMDesvios} = F \geq F_{(\alpha,\,1,\,n-2)}$$

em que $F_{(\alpha,\,1,\,n-2)}$ é o valor tabelado obtido através da Tabela de F-Snedecor para o nível α de significância, 1 e (n-2) graus de liberdade.

O coeficiente de determinação (R^2) é a estatística dada por:

$$R^2(\%) = \frac{SQ\,Regressão}{SQTotal}.100,\quad 0 \leq R^2 \leq 100.$$

O R^2 procura quantificar a proporção da variação da variável y que é explicada pelo modelo de regressão. Quanto mais próximo de 100 estiver R^2, melhor a qualidade de ajuste do modelo de regressão aos dados.

5.7.2.1) Exemplo de análise de regressão em dados sem repetição

Um estudo foi realizado sobre zonas de compactação em perfis de um solo, obtendo-se os seguintes dados de densidade (g/cm^3) em diferentes profundidades X (cm)

											Total
X (cm)	10	15	20	25	30	35	40	45	50	55	325
y (g/cm^3)	1,388	1,426	1,393	1,341	1,260	1,160	1,177	1,153	1,140	1,139	12,577

O modelo adotado foi:

$$y_i = b_0 + b_1X_i + \varepsilon_i\,,\quad i=1, 2, ..., 10 \quad e \quad \varepsilon_i \cap N(0, \sigma^2).$$

E as hipóteses testadas foram:

$$H_o: b_1 = 0$$

A representação gráfica também é semelhante à do Exemplo 1, com os valores inteiros no eixo horizontal, representando as classes da variável (número de lagartas).

Exemplo 3. Considere os valores a seguir referentes ao diâmetro à altura do peito (DAP), em cm, de 54 árvores de um talhão

10,7	17,2	21,2	22,9	24,2	25,9	28,8	32,8	36,1
12,4	17,6	21,6	23,3	24,4	26,1	29,5	33,6	37,5
13,8	18,8	21,8	23,5	24,4	26,1	30,2	34,2	38,1
14,6	19,2	22,2	23,8	24,6	26,8	30,9	34,5	39,0
16,1	20,5	22,3	23,9	24,8	27,5	31,3	34,7	39,7
16,8	20,9	22,8	24,2	25,5	28,1	32,0	35,5	41,2

A variável DAP é classificada como quantitativa contínua. A distribuição de freqüências para variáveis quantitativas contínuas são diferentes daquelas discretas e das variáveis qualitativas.

Primeiramente, para a realização de uma distribuição de freqüências de uma variável contínua, os dados devem ser ordenados em ordem crescente para uma melhor manipulação dos mesmos.

Depois segue-se a um algoritmo para a obtenção da distribuição de freqüências. Neste algoritmo, alguns passos são diferenciados se os dados são referentes a uma população ou a uma amostra.

i) Para população: escolher um número de classes (k) entre 5 e 20.

Para amostra:

Tamanho da amostra (n)	Número de classes (k)
Até 100	$\sqrt{n}$
> 100	5 $\log_{10}$ n

ii) Calcular a amplitude total (A) dos dados:

$$A = MVO - mvo$$

em que MVO é o maior valor observado e mvo é o menor valor observado;

iii) Calcular a amplitude de classe (c):

$$c = \frac{A + \Delta x}{k} \text{ (população)} \quad \text{ou} \quad c = \frac{A + \Delta x}{k - 1} \text{ (amostra)}$$

em que Δx é a precisão de medida (menor valor detectável pelo instrumento ou método de medição). O valor de c deverá ser arredondado para o mesmo número de casas decimais dos dados;

iv) Calcular o limite inferior da 1ª classe (LI_1):

$$LI_1 = mvo - \frac{\Delta x}{2} \text{ (população)} \quad \text{ou} \quad LI_1 = mvo - \frac{c}{2} \text{ (amostra);}$$

v) Calcular o limite superior da 1ª classe (LS_1):

$$LS_1 = LI_1 + c$$

LS_1 além de limite superior da 1ª classe, também é o limite inferior da 2ª classe:

$$LS_1 = LI_2$$

$$LS_2 = LI_2 + c$$

e assim sucessivamente até terminar as k classes;

vi) Calcular as freqüências absolutas (fa) e, opcionalmente, as relativas (fr) e porcentuais (fp) de cada classe:

Aplicando-se então o algoritmo nos dados do Exemplo 3, considerando que eles são referentes a uma população tem-se:

i) Escolhe-se, por exemplo, k = 10 classes;

ii) A = 41,2 – 10,7 = 30,5;

iii) $c = \frac{30{,}5 + 0{,}1}{10} = 3{,}06 = 3{,}1 (\text{arredondando})$;

$$b_1 = \frac{\sum_{i=1}^{n} X_i y_i - \frac{\left(\sum_{i=1}^{n} X_i \sum_{i=1}^{n} y_i\right)}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}.$$

A partir destes estimadores tem-se o modelo de regressão linear simples estimado (ajustado):

$$\hat{y}_i = \hat{b}_o + \hat{b}_1 X_i .$$

5.7.2) Análise de regressão em dados sem repetição

Seja a hipótese de nulidade em uma análise de regressão H_o: $b_1 = 0$ e a hipótese alternativa H_1: $b_1 \neq 0$, o esquema da análise de variância da regressão para se testar estas hipóteses é:

FV	GL	SQ	QM	F
Regressão	1	SQRegressão	SQRegressão/1	QMRegressão/QMDesvios
Desvios	n-2	SQDesvios	SQDesvios /(n-2)	
Total	n-1	SQTotal		

Em que as somas de quadrados (SQ's) são dadas pelas seguintes expressões:

SQTotal = $\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$, associada a (n-1) graus de liberdade.

SQRegressão = $\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \frac{\left[\sum_{i=1}^{n} X_i y_i - \frac{\left(\sum_{i=1}^{n} X_i \sum_{i=1}^{n} y_i\right)}{n}\right]^2}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}$, associada a 1 grau de liberdade.

A princípio, qualquer relação funcional entre um conjunto de variáveis regressoras e um conjunto de variáveis dependentes, representada por $y = f(X_1, X_2, ..., X_k)$, pode ser chamada de modelo de regressão, sendo tal modelo fixo para determinado conjunto de dados.

Normalmente esta relação funcional é desconhecida e uma função alternativa pode ser usada para aproximar *f* como, por exemplo, os modelos polinomiais que estão incluídos entre os tipos de modelos de regressão linear simples e são amplamente utilizados (Nogueira, 1997).

Um modelo de regressão linear é dito simples quando envolve somente uma variável regressora X. Os exemplos anteriores a) e c) se enquadram em casos de regressão linear simples. Já o exemplo b) é típico de regressão linear múltipla pois envolve mais de uma variável regressora (no caso, duas). O exemplo d) é um caso de regresão linear múltipla multivariada (múltipla pois apresenta 3 variáveis regressoras e multivariada pelas duas variáveis respostas, y_1 e y_2, utilizadas).

Sejam n pares de dados de duas variáveis (X_i, y_i) com i = 1, 2, ..., n. Admitindo que a relação entre y_i e X_i é uma reta, tem-se o modelo de regressão linear simples:

$$y_i = b_0 + b_1X_i + \varepsilon_i$$

em que ε_i é o erro experimental associado a observação y_i; b_0 e b_1 são parâmetros correspondentes ao coeficiente linear ou termo constante (intercepto da reta) e coeficiente angular ou de regressão, respectivamente. Os estimadores de quadrados mínimos de b_0 e b_1 são dados por:

$$b_o = \frac{\sum_{i=1}^{n} y_i}{n} - b_1\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \bar{y} - b_1\bar{X}$$

Observação: Note que, como os dados têm apenas uma casa decimal após a vírgula, o Δx é 0,1, se houvesse 2 casas após a vírgula, Δx seria 0,01 e assim por diante.

iv) $LI_1 = 10{,}7 - \frac{0{,}1}{2} = 10{,}65$

v) $LS_1 = 10{,}65 + 3{,}1 = 13{,}75$

$LI_2 = LS1 = 13{,}75$

$LS_2 = 13{,}75 + 3{,}1 = 16{,}85$ e assim por diante.

a) Representação tabular:

Tabela 3. Distribuição de freqüência do diâmetro à altura do peito (DAP), em cm, de 54 árvores de um talhão

Classes de DAP	Ponto médio	fa	fr	dfr	fp (%)
[10,65; 13,75)	12,2	2	0,0370	0,0119	3,70
[13,75; 16,85)	15,3	4	0,0741	0,0239	7,41
[16,85;19,95)	18,4	4	0,0741	0,0239	7,41
[19,95; 23,05)	21,5	9	0,1667	0,0538	16,67
[23,05; 26,15)	24,6	14	0,2592	0,0836	25,92
[26,15; 29,25)	27,7	4	0,0741	0,0239	7,41
[29,25; 32,35)	30,8	5	0,0926	0,0299	9,26
[32,35; 35,45)	33,9	5	0,0926	0,0299	9,26
[35,45; 38,55)	37,0	4	0,0741	0,0239	7,41
[38,55; 41,65)	40,1	3	0,0555	0,0179	5,55
Total	–	54	1,0000	–	100,00

Fonte: Notas de aula.

b) Representação gráfica:

Normalmente em gráficos de distribuição de freqüências de variáveis quantitativas contínuas usa-se no eixo vertical do gráfico a densidade de freqüência (df) de cada classe dada por:

$$\text{densidade de freqüência (df)} = \frac{\text{freqüência da classe}}{\text{amplitude da classe}}$$

Assim, pode-se usar a densidade de freqüência absoluta (dfa) ou a relativa (dfr) ou, ainda, a porcentual (dfp) obtidas, respectivamente, por:

$$dfa = \frac{fa}{c} \quad ; \quad dfr = \frac{fr}{c} \quad ; \quad dfp = \frac{fp}{c}.$$

Na Tabela 3 foram apresentadas as dfr´s (com c=3,1). O uso de df se torna importante nas situações onde as amplitudes de classes (c) são desiguais e, também, permite o cálculo de freqüências a partir de áreas do gráfico. Mas se c é igual para todas as classes pode-se utilizar, no eixo vertical do gráfico, tanto freqüências como densidades de freqüência.

Visto o conceito de df, os dois gráficos mais usais para distribuição de freqüências de variáveis contínuas são o histograma e o polígono de freqüência.

b.1) Histograma: é semelhante ao gráfico de barras, com barras dispostas lado a lado, e larguras iguais às amplitudes de classes.

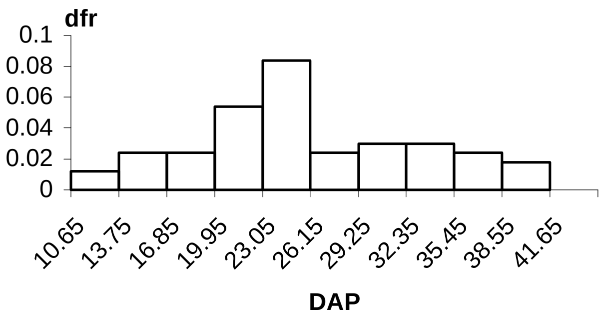


Figura 4. *Histograma de distribuição de freqüência relativa referente ao diâmetro à altura do peito (DAP), em cm, de 54 árvores de um talhão.*

Os resultados do teste de Tukey comparando as médias das Variedades para 1 e 2 linhas de irrigação está apresentado a seguir:

Variedades\Linhas	T_1'	T_2'
T_1	17,80 c	17,40 b
T_2	19,10 bc	19,10 ab
T_3	20,50 ab	19,80 a
T_4	21,18 a	17,40 b

Médias seguidas de mesma letra não diferem entre si pelo teste de Tukey (Prob > 0,05).

5.7) Análise de regressão

5.7.1) Características

Na pesquisa agropecuária é freqüente o interesse no estudo de relações funcionais entre variáveis quantitativas, como por exemplo:

a) Estudar a resposta na produção de grãos (y) em função de doses (X) de nitrogênio aplicadas ao solo, simbolizado por y = *f*(X);

b) Estimar o volume de madeira (y) em árvores de um povoamento florestal sem ter que derrubá-las, através da medida de seus diâmetros (X_1) e alturas (X_2), simbolizado por $y = f(X_1, X_2)$;

c) Expressar a curva de crescimento (y) de aves em função do tempo (X), simbolizado por y = *f*(X)

d) Determinar como o número de brotos (y_1) e seu peso seco (y_2) são afetados pelas doses de meio de cultura MS (X_1), de sacarose (X_2) e pH (X_3), simbolizado por $y_1, y_2 = f(X_1, X_2, X_3)$.

As variáveis y's dos exemplos anteriores que se deseja descrever são chamadas **variáveis dependentes ou respostas** e as variáveis X's são denominadas **independentes ou regressoras.**

Na natureza, certamente uma variável y que se deseja descrever, é determinada por um conjunto de outras variáveis, $X_1, X_2, ..., X_k$.

Variedade 2/ Linha 1 = $\overline{T_2T_1'} = \frac{76,4}{4} = 19,10$

Variedade 3/ Linha 1 = $\overline{T_3T_1'} = \frac{82,0}{4} = 20,50$

Variedade 4/ Linha 1 = $\overline{T_4T_1'} = \frac{84,7}{4} = 21,18$

- Comparando Médias de T para T_2':

Variedade 1/ Linha 2 = $\overline{T_1T_2'} = \frac{69,6}{4} = 17,40$

Variedade 2/ Linha 2 = $\overline{T_2T_2'} = \frac{76,4}{4} = 19,10$

Variedade 3/ Linha 2 = $\overline{T_3T_2'} = \frac{79,2}{4} = 19,80$

Variedade 4/ Linha 2 = $\overline{T_4T_2'} = \frac{69,6}{4} = 17,40$

Teste de Tukey:

$$DMS = q\sqrt{\frac{QMErro\,Médio}{r}}$$

sendo q para α=0,05; I = 4 tratamentos principais (Variedades) e GLErro Médio = $n' = 21 \Rightarrow q = 3,95$:

$$DMS = 3,95\sqrt{\frac{1,0207}{4}} = 2,00\,.$$

b.2) Polígono de freqüência: quando as amplitudes de classe (c) são iguais, o polígono é obtido pela união dos pontos médios das classes, nas alturas correspondentes às df's.

O polígono deve ser unido, no eixo horizontal, nos pontos:

$$LI_1 - \frac{c}{2} \qquad e \qquad LS_k + \frac{c}{2}$$

em que LS_k é o limite superior da última classe (k). No Exemplo 3 os pontos de união ao eixo horizontal são:

$$10,65 - \frac{3,1}{2} = 9,1 \quad e \quad 41,65 + \frac{3,1}{2} = 43,2\,.$$

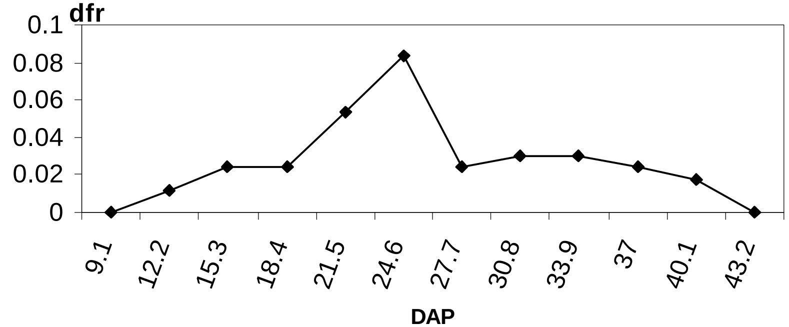


Figura 5. *Polígono de freqüência relativa referente ao diâmetro à altura do peito (DAP), em cm, de 54 árvores de um talhão.*

2.3) Natureza da distribuição

O objetivo da distribuição de freqüência é descrever o comportamento da variável. A natureza desse comportamento pode ser simétrica, assimétrica à direita ou à esquerda, como pode ser visualizado na Figura 6. Adiante será visto como se quantifica a assimetria.

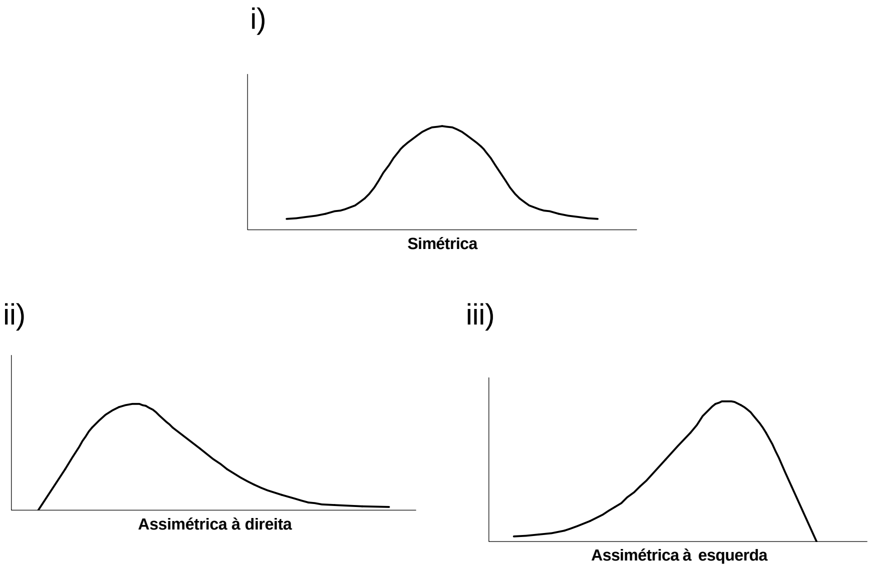


Figura 6. *Natureza da distribuição dos dados i) simétrica, ii) assimétrica à direita ou iii) assimétrica à esquerda.*

3) Estatísticas descritivas

3.1) Medidas de posição

Definição: é um número que descreve um conjunto de dados, pela indicação da posição que o conjunto ocupa na escala de valores possíveis que a variável em questão pode assumir.

3.1.1) Média $(\bar{X}$ ou $Me)$

$$Me = \frac{\sum_{i=1}^{N} X_i}{N}.$$

$$SQ\,Variedade/Linha\,2 = SQ\,T/T_2^{'} = \frac{1}{4}\left(69{,}6^2 + 76{,}4^2 + 79{,}2^2 + 69{,}6^2\right) - \frac{(294{,}8)^2}{16}$$

$SQ\,T/T_2^{'}$ = 5449,4800 – 5431,6900 = 17,7900.

Para certificar se o cálculo das somas de quadrados do desdobramento Variedades dentro de Linhas foi realizado corretamente basta verificar:

$$SQ\ T + SQ\ T \times T^{'} = SQ\ T/T_1^{'} + SQ\ T/T_2^{'}$$

$$26{,}9635 + 17{,}9184 = 27{,}0919 + 17{,}7900$$

$$44{,}8819 = 44{,}8819 \text{ ok!}$$

A análise de variância para o desdobramento T/T' é:

FV	GL	SQ	QM	F	Prob>F
$T/T_1^{'}$	(I-1) = 4-1 =3	27,0919	9,0306	8,848	0,0005
$T/T_2^{'}$	(I-1) = 4-1 =3	17,7900	5,9300	5,810	0,0047
Erro Médio	21	-	1,0207		

Da análise de variância anterior observa-se que houve diferença significativa entre efeitos de Variedades (T), no comprimento da banana, tanto para 1 linha de irrigação quanto para 2 linhas de irrigação (Prob < 0,05). Podemos então utilizar, por exemplo, o teste de Tukey para comparar as médias de T (Variedades) para $T_1^{'}$(1 linha de irrigação) e também para $T_2^{'}$ (2 linhas de irrigação).

Médias:

- Comparando Médias de T para $T_1^{'}$, do Quadro 3 pode-se obter:

Variedade 1/ Linha 1 = $\bar{T}\,T_1^{'} = \frac{71{,}2}{4} = 17{,}80$

Linha 1/ Variedade 4 = 21,18 a
Linha 2/ Variedade 4 = 17,40 b

d) Comparações entre médias de tratamentos principais dentro de cada nível de tratamento secundário (médias de Variedades dentro de cada Linha – T/T'):

Esta comparação envolve os dois erros por meio de um erro médio, sendo portanto um pouco mais complicada que as demais.

$$\text{QMErro Médio} = \frac{\text{QMErro(a)} + (K-1).\text{QMErro(b)}}{K}$$

$$\text{QMErro Médio} = \frac{1{,}0011 + (2-1).1{,}0403}{2} = 1{,}0207\,.$$

O número de graus de liberdade (n') associado a este Erro Médio é calculado de modo aproximado pela fórmula de Satterthwaite:

$$n' = \frac{[\text{QMErro(a)} + (K-1).\text{QMErro(b)}]^2}{\frac{[\text{QMErro(a)}]^2}{\text{GLErro(a)}} + \frac{[(K-1)\text{QMErro(b)}]^2}{\text{GLErro(b)}}}$$

$$n' = \frac{[1{,}0011 + (2-1).1{,}0403]^2}{\frac{[1{,}0011]^2}{9} + \frac{[(2-1).1{,}0403]^2}{12}} = 20{,}67 \approx 21 \text{ (arredondando).}$$

Observação: GLErro (a) ≤ n' ≤ [GLErro (a) + GLErro (b)].

Do Quadro 3 obtém-se:

$$\text{SQ Variedade /Linha 1} = \text{SQ T}/T_1' = \frac{1}{4}\left(71{,}2^2 + 76{,}4^2 + 82{,}0^2 + 84{,}7^2\right) - \frac{(314{,}3)^2}{16}$$

$\text{SQ T}/T_1'$ = 6201,1225 – 6174,0306 = 27,0919;

Para o Exemplo 3 a média é:

$$\text{Me} = \frac{10{,}7 + 12{,}4 + \cdots + 39{,}7 + 41{,}2}{54} = 25{,}9.$$

Propriedades da média:

i) Somando-se uma constante K a todos os dados, a média (Me) também é acrescida de K:

$$\text{Me}(x+K) = \text{Me}(x) + K\,;$$

ii) Multiplicando-se K a todos os dados, a média também é multiplicada por K:

$$\text{Me}(x.K) = K.\text{Me}\,;$$

iii) A soma dos desvios (d_i's) em relação a média é zero:

$$d_i = x_i - \text{Me}\,;$$

Exemplo 4. Para as N = 3 observações (x_i) a seguir, os desvios d_i em relação a média são:

	x_i	d_i
	3	3-5 = -2
	5	5-5 = 0
	7	7-5 = 2
Média (Me) =	5	$\sum_{i=1}^{N} d_i$ = 0

iv) A média minimiza a soma dos quadrados dos desvios (SQD), ou seja, o valor da SQD seria aumentada se colocássemos qualquer outro valor que não Me.

$$\text{SQD} = \sum_{i=1}^{N}\left[x_i - \text{Me}\right]^2.$$

Observações: A média é muito influenciada por valores discrepantes, extremos. Ela é a medida de posição mais utilizada.

3.1.2) Mediana (Md)

É o valor que é precedido e seguido pelo mesmo número de observações, em um conjunto de dados ordenados.

Exemplo 5. Para as N = 5 observações (x_i) a seguir, a mediana é:

x_1	x_2	x_3	x_4	x_5
8	9	10	15	40

Md = 10 (este valor é precedido e seguido por duas observações).

Observação: Se o número de observações (N) for par, toma-se a média dos dois valores centrais.

Exemplo 6. Para as N = 4 observações (x_i) a seguir, a mediana é:

x_1	x_2	x_3	x_4
9	10	14	20

$$\text{Md} = \frac{10+14}{2} = 12.$$

Propriedades da mediana:

i) Md (x+K) = Md (x) +K;

ii) Md (x.K) = K.Md (x);

iii) A mediana é o valor que minimiza a soma dos módulos dos desvios:

$$\sum \left| \, x_i - a \, \right| \text{ é mínima se a = Md(x).}$$

Observação: A Md é uma medida de posição para medidas assimétricas.

Da análise de variância anterior observa-se que houve diferença significativa entre efeitos de Linhas (T´), no comprimento da banana, somente para a Variedade 4 (Prob < 0,05). Para as demais variedades T_1,T_2 e T_3 não houve diferenças significativas (Prob > 0,05) entre 1 e 2 linhas de irrigação no comprimento do fruto central da terceira penca de banana. Podemos então utilizar, por exemplo, o teste de Tukey para comparar as médias de T´ (1 e 2 Linhas de irrigação) para T_4 (Variedade 4).

Médias:

Linha 1/ Variedade 4 = $\overline{T_1^{'}T_4} = \dfrac{84{,}7}{4} = 21{,}18$

Linha 2/ Variedade 4 = $\overline{T_2^{'}T_4} = \dfrac{69{,}6}{4} = 17{,}40$

Teste de Tukey:

$$\text{DMS} = q\sqrt{\frac{\text{QMErro (b)}}{r}}$$

sendo q para α=0,05; K = 2 tratamentos secundários (Linhas) e GLErro (b) = 12 $\Rightarrow$ q = 3,08:

$$\text{DMS} = 3{,}08\sqrt{\frac{1{,}0403}{4}} = 1{,}57.$$

O contraste entre $T_1^{'}$ e $T_2^{'}$ para T_4 é:

$$\hat{y} = \overline{T}_1^{'} - \overline{T}_2^{'} = 21{,}18 - 17{,}40 = 3{,}78.$$

3,78 > 1,57 portanto $T_1^{'} \neq T_2^{'}$ para T_4. Ou seja, para Variedade 4 (T_4), 1 linha de irrigação ($T_1^{'}$) proporcionou significativamente maior comprimento (cm) do fruto central da terceira penca de banana do que 2 linhas de irrigação ($T_2^{'}$). Colocando as letras do teste:

$SQ\,T'/T_1$ = 2478,4000 – 2478,0800 = 0,3200;

$$SQ\,Linha\,/Variedade\,2 = SQ\,T'/T_2 = \frac{1}{4}\left(76,4^2 + 76,4^2\right) - \frac{(152,8)^2}{8}$$

$SQ\,T'/T_2$ = 2918,4800 – 2918,4800 = 0,0000;

$$SQLinha\,/\,Variedade\,3 = SQ\,T'\,/\,T_3 = \frac{1}{4}\left(82,0^2 + 79,2^2\right) - \frac{(161,2)^2}{8}$$

$SQ\,T'\,/\,T_3 =$ 3249,1600 – 3248,18 = 0,9800;

$$SQLinha\,/\,Variedade\,4 = SQ\,T'\,/\,T_4 = \frac{1}{4}\left(84,7^2 + 69,6^2\right) - \frac{(154,3)^2}{8}$$

$SQ\,T'\,/\,T_4$ = 3004,5625 – 2976,0613 = 28,5012.

Para certificar se o cálculo das somas de quadrados do desdobramento Linhas dentro de Variedades foi realizado corretamente basta verificar:

$$SQ\ T' + SQ\ T \times T' = SQ\ T'/T_1 + SQ\ T'/T_2 + SQ\ T'/T_3 + SQ\ T'/T_4$$

11,8828 + 17,9184 = 0,3200 + 0,0000 + 0,9800 + 28,5012.

29,8012 = 29,8012 ok!

A análise de variância para o desdobramento T'/T é:

FV	GL	SQ	QM	F	Prob>F
T'/T_1	(K-1) = 2-1 =1	0,3200	0,3200	0,308	0,6347
T'/T_2	(K-1) = 2-1 =1	0,0000	0,0000	0,000	0,9975
T'/T_3	(K-1) = 2-1 =1	0,9800	0,9800	0,942	0,4341
T'/T_4	(K-1) = 2-1 =1	28,5012	28,5012	27,397	0,0346
Erro (b)	12	12,4838	1,0403		

3.1.3) Moda (Mo)

É o valor mais freqüente no conjunto de dados.

Exemplo 7. Para as N = 5 observações (x_i) a seguir, a moda é:

x_1	x_2	x_3	x_4	x_5
8	9	9	12	18

Mo = 9 (valor mais freqüente; apareceu duas vezes no conjunto de dados).

Propriedades da moda:

i) Mo (x+K) = Mo (x) +K;

ii) Mo (x.K) = K.Mo (x).

Observações: A Mo também é uma medida de posição para medidas assimétricas. Ela é ainda menos afetada por valores extremos do que a mediana. Para variáveis contínuas, onde é difícil encontrar um mesmo valor repetido duas ou mais vezes, a moda é calculada de outra maneira, através do denominado método de Czuber; porém, tal método não será discutido neste material.

3.2) Medidas de dispersão

Definição: grandeza numérica que descreve a variabilidade em um conjunto de dados.

3.2.1) Amplitude (A)

A = MVO - mvo

Trata-se da diferença entre o maior valor observado (MVO) e o menor valor observado (mvo) como já foi visto anteriormente.

Exemplo 8. Considere dois conjuntos de dados (X e Y) medidos em metro (m):

						Totais
X	6	16	16	16	41	95
Y	6	11	21	31	41	110

A (X) = 41 – 6 = 35;

A (Y) = 41 – 6 = 35.

X e Y apresentam mesma amplitude (A), portanto o conjunto X apresenta claramente menor variabilidade (maior uniformidade) que o conjunto Y.

Observação: A amplitude é muito influenciada por valores extremos, uma vez que é calculada a partir deles. Assim, a medida que aumenta N, aumenta a chance de encontrar valores extremos, aumentando, portanto, a amplitude.

3.2.2) Variância (Var) e Desvio padrão (DP)

São medidas baseadas em todos os dados, a partir dos desvios em relação a média.

- Variância (Var ou σ^2): média dos quadrados dos desvios (também chamada de quadrado médio), cuja expressão é dada por:

$$\sigma^2 = \frac{\sum_{i=1}^{N}\left[x_i - Me\right]^2}{N} \text{ (população) ou } \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left[x_i - Me\right]^2}{n-1} \text{ (amostra).}$$

Ou ainda, pelas expressões alternativas:

$$\sigma^2 = \frac{\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}{N} \text{ (população) ou } \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} \text{ (amostra).}$$

$|-1,5| = 1,5 < 1,56$ portanto $\overline{T}_1 = \overline{T}_2$ e assim continua as comparações entre as outras médias de variedades duas a duas.

b) Comparações entre médias de tratamentos secundários (médias de Linhas - T'):

Comparando a média de T_1' com a de T_2' pelo teste de Tukey, do Quadro 3 pode-se obter:

$$\overline{T}_1' = \frac{T_{T_1'}}{rI} = \frac{314,3}{4x4} = 19,64 \quad \text{e} \quad \overline{T}_2' = \frac{T_{T_2'}}{rI} = \frac{294,8}{4x4} = 18,43$$

$$DMS = q\sqrt{\frac{QMErro(b)}{r.I}}$$

sendo q para α=0,05; K = 2 tratamentos secundários (Linhas) e GLErro (b) = 12 $\Rightarrow$ q = 3,08:

$$DMS = 3,08\sqrt{\frac{1,0403}{4.4}} = 0,78$$

O contraste entre T_1' e T_2' é:

$$\hat{y} = \overline{T}_1' - \overline{T}_2' = 19,64 - 18,43 = 1,21.$$

1,21 > 0,78 portanto $T_1' \neq T_2'$.

c) Comparações entre médias de tratamentos secundários dentro de cada nível de tratamento principal (médias de Linhas dentro de cada Variedade - T'/T):

Do Quadro 3 obtém-se:

$$SQ\,Linha/Variedade\,1 = SQ\,T'/T_1 = \frac{1}{4}\left(71,2^2 + 69,6^2\right) - \frac{(140,8)^2}{8}$$

T_{Ti} e $T_{Ti'}$: total do tratamento principal i e do tratamento secundário i', respectivamente.

$\overline{T}_i$ e $\overline{T}_i'$: média do tratamento principal i e do tratamento secundário i', respectivamente.

As comparações de médias que o pesquisador pode ter interesse em um experimento em parcelas subdivididas são as seguintes:

a) Comparações entre médias de tratamentos principais (médias de Variedades - T):

Comparando, por exemplo, a média de T_1 com a de T_2 pelo teste de Tukey, do Quadro 3 pode-se obter:

$$\overline{T}_1 = \frac{T_{T_1}}{rK} = \frac{140,8}{4x2} = 17,6 \quad \text{e} \quad \overline{T}_2 = \frac{T_{T_2}}{rK} = \frac{152,8}{4x2} = 19,1$$

$$DMS = q\sqrt{\frac{QMErro(a)}{r.K}}$$

sendo q para α=0,05; I = 4 tratamentos principais (Variedades) e GLErro (a) = 9 $\Rightarrow$ q = 4,41:

$$DMS = 4,41\sqrt{\frac{1,0011}{4.2}} = 1,56\,.$$

O contraste entre $\overline{T}_1$ e $\overline{T}_2$ é:

$$\hat{y} = \overline{T}_1 - \overline{T}_2 = 17,6 - 19,1 = -1,5\,.$$

Lembrando a interpretação do teste Tukey:

Se $|\hat{y}| \geq$ DMS $\Rightarrow$ as médias dos dois tratamentos em comparação podem ser consideradas estatisticamente diferentes.

- Desvio padrão (DP ou σ): é a raiz quadrada da variância, cuja expressão é dada por:

$$\sigma = \sqrt{\sigma^2} \text{ (população)} \qquad \text{ou} \qquad \hat{\sigma} = \sqrt{\hat{\sigma}^2} \text{ (amostra).}$$

Observações: Quanto maior σ^2 ou $\hat{\sigma}^2$, maior a variabilidade do conjunto de dados. O DP tem a vantagem, em relação a Var, de possuir a mesma unidade dos dados (por exemplo, se a unidade de medida dos dados é kg, a do DP também será kg enquanto que a da Var será kg^2), facilitando, assim, a visualização do quanto, em média, os dados se desviam da média.

Para o Exemplo 8 tem-se:

$$\text{Var(X)} = \frac{6^2 + 16^2 + 16^2 + 16^2 + 41^2 - \frac{(95)^2}{5}}{5} = \frac{2485 - 1805}{5} = 136,00 \text{ m}^2;$$

$$\text{Var(Y)} = \frac{6^2 + 11^2 + 21^2 + 31^2 + 41^2 - \frac{(110)^2}{5}}{5} = \frac{3240 - 2420}{5} = 164,00 \text{ m}^2;$$

$$\text{DP(X)} = \sqrt{136} = 11,66 \text{ m};$$

$$\text{DP(Y)} = \sqrt{164} = 12,81 \text{ m}.$$

Propriedades da variância e do desvio padrão:

i) Somado-se uma constante K a todos os dados, a Var e o DP não se alteram:

$$Var(x + K) = Var(x)\,;$$

$$DP(x + K) = DP(x)\,;$$

ii) Multiplicando-se K a todos os dados, a Var fica multiplicada por K^2 e o DP por K.

$$\text{Var(x.K)} = \text{K}^2[\text{Var(x)}];$$

$$\text{DP (x.K)} = \text{K [DP (x)]};$$

iii) O DP em relação a média é mínimo ao invés de qualquer outro valor devido ao fato da média ser o valor que torna mínima a soma de quadrados do desvio (SQD).

3.2.3) Coeficiente de Variação (CV)

$$CV(\%) = \left(\frac{DP}{Me}\right).100 .$$

O CV é uma medida relativa, porcentual, pois o desvio e a média possuem a mesma unidade.

Exemplo 9. Considere os pesos (Kg) de animais de dois rebanhos diferentes:

	Rebanho A	Rebanho B
	70	490
	90	510
	80	480
	100	500
Me	85	495
DP	11,18	11,18

É claro que pelos valores de pesos tratam-se de rebanhos de idades diferentes. Os rebanhos A e B possuem o mesmo DP, porém, é óbvio que diferenças de 5 kg, por exemplo, possuem um peso relativo muito maior no rebanho A do que no rebanho B. Assim, poderíamos afirmar que a variabilidade do rebanho A é maior do que a do rebanho B. Isto pode ser comprovado pelos valores de CV dos dois rebanhos:

CV (Rebanho A): $\frac{11,18}{85} \text{ x } 100 = 13,15\%$;

CV (Rebanho B): $\frac{11,18}{495} \text{ x } 100 = 2,26\%$.

FV	GL	SQ	QM	F	Prob>F
Bloco	3	15,7535	5,2512	5,245	0,0229
Variedades (T)	3	26,9635	8,9878	8,978	0,0045
Erro (a)	9	9,0102	1,0011		
Parcelas	15	51,7272			
Linhas (T')	1	11,8828	11,8828	11,422	0,0055
T x T'	3	17,9184	5,9728	5,741	0,0338
Erro (b)	12	12,4838	1,0403		
Total	31	94,0122			
CV (a) (%)	5,26				
CV (b) (%)	5,37				
Média geral $(\bar{y})$: 19,0		Número de observações:	32		

Nos experimentos em parcelas subdivididas tem-se dois coeficientes de variação (CV):

Para parcelas:

$$CV(a) = \frac{\sqrt{QMErro(a)}}{\bar{y}}.100 = \frac{\sqrt{1,0011}}{19,0}.100 = 5,26\% ;$$

Para subparcelas:

$$CV(b) = \frac{\sqrt{QMErro(b)}}{\bar{y}}.100 = \frac{\sqrt{1,0403}}{19,0}.100 = 5,37\% .$$

Considere:

I: número de tratamentos principais, I = 4 variedades;

K: número de tratamentos secundários, K = 2 linhas de irrigação;

r: número de blocos, r = 4 blocos;

$SQ\ Parcelas = 11645{,}5650 - 11593{,}8378 = 51{,}7272$;

SQ Erro (a) = SQ Parcelas – SQ Blocos – SQ Variedades

SQ Erro (a) = 51,7272 – 15,7535 – 26,9635 = 9,0102.

É necessário também fazer um outro quadro auxiliar com a combinação entre os níveis dos dois fatores (variedades e linhas de irrigação) para o cálculo da soma de quadrados do tratamento da subparcela (linhas de irrigação) e da interação variedades x linha (T x T´).

Quadro 3. Quadro auxiliar com os totais de todas as repetições para cada combinação entre os níveis dos fatores T e T´.

Linhas\Variedades	T_1	T_2	T_3	T_4	Totais
$T_1^{'}$	71,2 (4)	76,4	82,0	84,7	314,3(16)
$T_2^{'}$	69,6	76,4	79,2	69,6	294,8
Totais	140,8(8)	152,8	161,2	154,3	609,1

Do Quadro 3 é possível obter:

$$SQ\,Linhas = \frac{1}{16}(314{,}3^2 + 294{,}8^2) - 11593{,}8378$$

SQ Linhas = 11605,7206 – 11593,8378 = 11,8828;

$$SQ\,Variedades\,x\,Linhas\,(T\,x\,T^{'}) = \frac{1}{4}(71{,}2^2 + 69{,}6^2 + \ldots + 69{,}6^2) - C - SQ\,Variedades - SQ Linhas$$

SQ Variedades x Linhas (T x $T^{'}$) = 11650,6025 – 11593,8378 – 26,9635 – 11,8828 = 17,9184;

SQ Erro (b) = SQ Total – SQ Parcela – SQ Linhas – SQ Variedades x Linhas

SQ Erro (b) = 94,0122 – 51,7272 – 11,8828 – 17,9184 = 12,4838.

E o quadro de análise de variância para os dados do exemplo 5.6.5) conforme o esquema em parcela subdividida é:

Observação: O CV por ser adimensional é útil na comparação entre conjuntos de dados com mesma unidade mas permite, também, a comparação da variabilidade entre conjuntos de dados referentes a diferentes características.

3.3) Medidas de assimetria e curtose

Em estatística, freqüentemente é interessante saber se a população da qual a amostra foi coletada pode ser descrita por uma curva normal. Isso pode ser verificado por meio das seguintes medidas:

3.3.1) Coeficiente de assimetria (As): medida que quantifica o distanciamento de um conjunto de dados em relação à simetria. O coeficiente As é dado por:

$$As = \frac{m_3}{d^2\sqrt{d^2}} = \frac{m_3}{d^3}$$

sendo $m_3 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n}$ e $d^2 = \sigma^2$ (variância populacional) ou $\hat{\sigma}^2$ (variância amostral).

$$\text{Se As é} \begin{cases} \text{positivo}\,(As > 0)\text{: indica uma assimetria à direita,} \\ \text{negativo}\,(As < 0)\text{: indica uma assimetria à esquerda,} \\ \text{zero}\,(As = 0)\text{: indica uma simetria (amostra pode ser considerada vinda de uma} \\ \qquad\qquad \text{distribuição normal).} \end{cases}$$

Na Figura 6 pode ser visto a natureza do comportamento de uma variável, se simétrica, assimétrica à direita ou assimétrica à esquerda.

Observação: Na prática os valores de As dificilmente serão zero, podendo ser próximos de zero.

3.3.2) Coeficiente de curtose (K): medida que quantifica o grau de achatamento da distribuição de freqüência de um conjunto de dados, tendo a curva normal como referência. O coeficiente K é dado por:

$$K = \frac{m_4}{d^2.d^2} = \frac{m_4}{d^4}$$

sendo $m_4 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n}$ e $d^2 = \sigma^2$ (variância populacional) ou $\hat{\sigma}^2$ (variância amostral).

$$\text{Se K é} \begin{cases} > 3: \text{indica uma distribuição afiada chamada leptocúrtica,} \\ < 3: \text{indica uma distribuição achatada chamada platicúrtica,} \\ = 3: \text{indica uma distribuição semelhante a normal chamada mesocúrtica.} \end{cases}$$

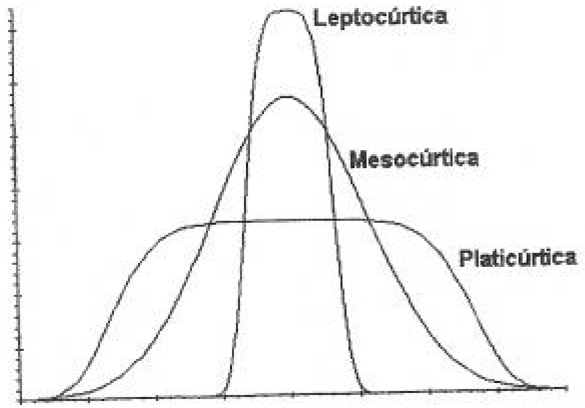


Figura 7. *Gráfico dos diferentes graus de achatamento relativos a uma distribuição de freqüência*

Exemplo 10. Seja as seguintes N = 4 observações, a média ($\bar{x}$) e a variância (d^2)destas observações dadas por

x_1	x_2	x_3	x_4	$\bar{x}$	d^2
2	15	16	17	12,5	37,25

Considerando que a unidade de cálculo é a subparcela, do Quadro de dados podemos tirar:

C = $\frac{(609,1)^2}{32}$ = 11593,8378;

SQ Blocos = $\frac{1}{8}(156,4^2 + 155,1^2 + 142,6^2 + 155,0^2) - 11593,8378$

SQ Blocos = 11609,5913 – 11593,8378 = 15,7535;

SQ Total = $19,0^2 + 17,1^2 + \ldots + 16,4^2 + 18,6^2$ – 11593,8378

SQ Total = 11687,8500– 11593,8378 = 94,0122.

Para o cálculo da soma de quadrados de parcelas, é necessário fazer um quadro auxiliar com os totais das parcelas.

Quadro 2. Quadro auxiliar com os totais das parcelas

Repetições	Tratamentos				Totais
	T_1	T_2	T_3	T_4	
1	37,9 (2)	39,0	41,5	38,0	156,4 (8)
2	34,7	37,8	42,2	40,4	155,1
3	32,4	36,8	36,0	37,4	142,6
4	35,8	39,2	41,5	38,5	155,0
Totais	140,8(8)	152,8	161,2	154,3	609,1

Do Quadro 2 calculamos:

$$\text{SQ Variedades} = \frac{1}{8}(140,8^2 + 152,8^2 + 161,2^2 + 154,3^2) - 11593,8378$$

SQ Variedades = 11620,8013 – 11593,8378 = 26,9635;

$$\text{SQ Parcelas} = \frac{1}{2}(37,9^2 + 34,7^2 + \ldots + 37,4^2 + 38,5^2) - 11593,8378$$

fruto central da terceira penca de banana estão dispostos na Tabela 8 a seguir.

Tabela 8. Comprimento (cm) do fruto central da terceira penca de banana para um experimento em blocos casualizados (DBC), com 4 repetições, em esquema de parcela subdividida com 4 variedades de banana (T_1, T_2, T_3 e T_4) nas parcelas e 2 linhas de irrigação ($T_1^{'}$ = 1 linha e $T_2^{'}$ = 2 linhas) nas subparcelas

Repetições	Tratamentos								Totais
	T_1		T_2		T_3		T_4		
	$T_1^{'}$	$T_2^{'}$	$T_1^{'}$	$T_2^{'}$	$T_1^{'}$	$T_2^{'}$	$T_1^{'}$	$T_2^{'}$	
1	19,0	18,9	19,2	19,8	20,8	20,7	21,1	16,9	156,4
2	17,1	17,6	19,5	18,3	20,9	21,3	22,7	17,7	155,1
3	17,5	14,9	17,5	19,3	18,6	17,4	21,0	16,4	142,6
4	17,6	18,2	20,2	19,0	21,7	19,8	19,9	18,6	155,0
Totais	71,2	69,6	76,4	76,4	82,0	79,2	84,7	69,6	609,1

5.6.6) Croqui de campo

	T_2		T_4		T_1		T_3	
BL I	$T_2^{'}$	$T_1^{'}$	$T_2^{'}$	$T_1^{'}$	$T_1^{'}$	$T_2^{'}$	$T_1^{'}$	$T_2^{'}$
	T_3		T_1		T_2		T_4	
BL II	$T_1^{'}$	$T_2^{'}$	$T_2^{'}$	$T_1^{'}$	$T_1^{'}$	$T_2^{'}$	$T_2^{'}$	$T_1^{'}$
	T_4		T_3		T_1		T_2	
BL III	$T_1^{'}$	$T_2^{'}$	$T_1^{'}$	$T_2^{'}$	$T_1^{'}$	$T_2^{'}$	$T_2^{'}$	$T_1^{'}$
	T_1		T_2		T_3		T_4	
BL IV	$T_2^{'}$	$T_1^{'}$	$T_1^{'}$	$T_2^{'}$	$T_2^{'}$	$T_1^{'}$	$T_1^{'}$	$T_2^{'}$

$$m_3 = \frac{(2-12,5)^3 + (15-12,5)^3 + (16-12,5)^3 + (17-12,5)^3}{4} = \frac{-1008}{4} = -252\ ;$$

$$m_4 = \frac{(2-12,5)^4 + (15-12,5)^4 + (16-12,5)^4 + (17-12,5)^4}{4} = \frac{12754,25}{4} = 3188,563\ ;$$

$$As = \frac{-252}{37,25\sqrt{37,25}} = \frac{m_3}{d^3} = -1,108$$ (As < 0 → Assimetria a esquerda);

$$K = \frac{3188,563}{(37,25).(37,25)} = 2,30$$ (K < 3 → Distribuição platicúrtica).

4) Testes de comparações múltiplas

4.1) Contrastes ortogonais de médias

Definição: São combinações lineares dadas por:

$$Y_1 = a_1m_1 + a_2m_2 + ... + a_nm_n$$

$$Y_2 = b_1m_1 + b_2m_2 + ... + b_nm_n$$

$$\vdots$$

$$Y_{I-1} = c_1m_1 + c_2m_2 + ... + c_nm_n$$

sendo a soma dos coeficientes de cada contraste igual a zero:

$$\sum_{i=1}^{n} a_i, \sum_{i=1}^{n} b_i \cdots \sum_{i=1}^{n} c_i = 0,$$

em que:

a_1, b_1, c_1, ..., a_n, b_n, c_n são os coeficientes dos contrastes;

m_1, m_2, ..., m_n são médias dos tratamentos 1, 2, ...,n.

Dois contrastes são ditos ortogonais quando há uma independência entre suas comparações, ou melhor, quando a variação de um contraste é independente da variação do outro. A exigência para que dois contrates sejam ortogonais é que a covariância (Cov) entre eles seja nula:

$$\text{Cov}(Y_i,\ Y_{i'}) = 0.$$

Seja s_i^2 a variância do tratamento i e r_i o número de repetições do tratamento i, a covariância entre dois contrastes é dada por uma das seguintes expressões:

- Se $s_1^2 \neq s_2^2 \neq ... \neq s_n^2$ e $r_1 \neq r_2 \neq ... \neq r_n$:

$$\text{Cov}(Y_1,\ Y_2) = \sum_{i=1}^{n} \frac{a_i b_i}{r_i} s_i^2 = \frac{a_1 b_1}{r_1} s_1^2 + \frac{a_2 b_2}{r_2} s_2^2 + \cdots + \frac{a_n b_n}{r_n} s_n^2 .$$

- Se $s_1^2 = s_2^2 = ... = s_n^2$ e $r_1 \neq r_2 \neq ... \neq r_n$:

$$\text{Cov}(Y_1,\ Y_2) = \sum_{i=1}^{n} \frac{a_i b_i}{r_i} = \frac{a_1 b_1}{r_1} + \frac{a_2 b_2}{r_2} + \cdots + \frac{a_n b_n}{r_n} .$$

- Se $s_1^2 = s_2^2 = ... = s_n^2$ e $r_1 = r_2 = ... = r_n$:

$$\text{Cov}(Y_1,\ Y_2) = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n .$$

A variância (Var) de um contraste Y é:

$$\text{Var}(Y) = s^2 \sum_{i=1}^{n} \frac{c_i^2}{r_i} \quad (\text{se } s_1^2 = s_2^2 = ... = s_n^2 = s^2)$$

ou

$$\text{Var}(Y) = \sum_{i=1}^{n} \frac{c_i^2}{r_i} s_i^2 \quad (\text{se } s_1^2 \neq s_2^2 \neq ... \neq s_n^2).$$

O erro padrão do contraste Y é:

$$s(Y) = \sqrt{\text{Var}(Y)} .$$

5.6.3) Desvantagem

Há uma redução do número de graus de liberdade do erro, comparativamente ao esquema fatorial, redução esta decorrente da existência de dois erros, o erro (a) referente às parcelas e o erro (b), correspondente às subparcelas dentro das parcelas.

5.6.4) Modelo estatístico do experimento em parcela subdividida

O modelo a seguir corresponde a um modelo de um DBC em esquema de parcela subdividida:

$$y_{ijk} = \mu + \beta_j + \alpha_i + \delta_{ij} + \gamma_k + (\alpha\gamma)_{ik} + e_{ijk}$$

em que y_{ijk} é o valor observado referente a parcela que recebeu o i-ésimo nível do tratamento principal α e o k-ésimo nível do tratamento secundário γ no j-ésimo bloco; μ representa uma constante geral associada a esta variável aleatória; β_j é o efeito do j-ésimo bloco; α_i é o efeito do i-ésimo nível do tratamento principal; $\delta_{ij} = (\alpha\beta)_{ij}$ é o efeito residual das parcelas, caracterizado como componente do erro (a); γ é o efeito do k-ésimo nível do tratamento secundário; $(\alpha\gamma)_{ij}$ é o efeito da interação do i-ésimo nível do tratamento principal α com o k-ésimo nível do tratamento secundário γ e e_{ijk} representa o efeito residual das subparcelas, caracterizado como componente do erro (b).

Sobre as distribuições de δ_{ij} e e_{ijk} pode-se considerar as seguintes pressuposições: i) $\delta_{ij} \cap N(0,\ \sigma_\delta^2)$; ii) $e_{ijk} \cap N(0,\ \sigma^2)$; iii) δ_{ij} e e_{ijk} são não correlacionados.

5.6.5) Exemplo de parcela subdividida

Foi realizado um experimento em blocos casualizados com 4 repetições, no esquema de parcelas subdivididas. Os tratamentos das parcelas foram 4 variedades de banana (T_1, T_2, T_3 e T_4) e os tratamentos das subparcelas foram uma e duas linhas de irrigação (T_1' = 1 linha e T_2' = 2 linhas). Os dados do comprimento (cm) do

5.6) Experimentos em parcelas subdivididas

5.6.1) Características

O esquema experimental em parcelas subdivididas se caracteriza como sendo uma variação do experimento fatorial com dois fatores (Steel et al., 1997). A principal característica destes experimentos é que as parcelas são divididas em subparcelas. Os tratamentos das parcelas são chamados de primários ou principais e são dispostos segundo um tipo qualquer de delineamento, sendo os mais usados os delineamentos em blocos casualizados, com o objetivo de procurar controlar a variabilidade que possa haver no material experimental. Os tratamentos das subparcelas são chamados secundários e são dispostos aleatoriamente dentro de cada parcela. Assim, cada parcela funciona como um bloco para os tratamentos secundários. Primeiro casualizam-se os níveis do fator primário nas parcelas de cada bloco; em seguida, casualizam-se os níveis do fator secundário nas subparcelas de cada parcela. Pimentel Gomes (1990) e Hinkelmann & Kempthorne (1994), dentre outros autores, são unânimes em afirmar a maior precisão existente no teste de tratamentos secundários.

5.6.2) Vantagens

Os experimentos em parcelas subdivididas apresentam uma grande utilidade na pesquisa agropecuária, além de outras diversas áreas.

Tais experimentos são úteis em situações como: a) quando os níveis de um dos fatores exigem grandes quantidades de material experimental (por exemplo, níveis de irrigação), devendo ser casualizados nas parcelas; b) quando informações prévias asseguram que as diferenças entre os níveis de um dos fatores são maiores que as do outro fator; c) quando se deseja maior precisão para comparações entre níveis de um dos fatores; d) quando existe um fator de maior importância (que deverá ser casualizado na subparcela) e outro de importância secundária, sendo este incluído para aumentar a extensão dos resultados e e) nas situações práticas, onde é difícil a instalação do experimento no esquema fatorial.

Observações: Em um experimento com I tratamentos, o número máximo de contrastes ortogonais possíveis é dado por I-1 comparações. Os contraste são formulados de acordo com o interesse do pesquisador.

Exemplo 11. Considere as médias de produtividade de grãos (t/ha) de 4 cultivares de milho:

$$\hat{m}_1 = 5{,}2$$

$$\hat{m}_2 = 3{,}3$$

$$\hat{m}_3 = 4{,}0$$

$$\hat{m}_4 = 9{,}0$$

$r_1 = r_2 = r_3 = r_4 = 5$ e $s_1^2 = s_2^2 = s_3^2 = s_4^2 = s^2 = 0{,}19$.

i) Escolher os (I-1) = 4-1 = 3 contrastes:

$Y_1 = m_1 + m_2 - m_3 - m_4$ em que $a_1 = 1$, $a_2 = 1$, $a_3 = -1$, $a_4 = -1$

$Y_2 = m_1 - m_2$ em que $a_1 = 1$, $a_2 = -1$, $a_3 = 0$, $a_4 = 0$

$Y_3 = m_3 - m_4$ em que $a_1 = 0$, $a_2 = 0$, $a_3 = 1$, $a_4 = -1$.

ii) Verificar se o somatório dos coeficientes de cada contraste é igual a zero:

$$Y_1 \Rightarrow \sum_{i=1}^{4} a_i = 1 + 1 - 1 - 1 = 0$$

$$Y_2 \Rightarrow \sum_{i=1}^{4} b_i = 1 - 1 + 0 + 0 = 0$$

$$Y_3 \Rightarrow \sum_{i=1}^{4} c_i = 0 + 0 + 1 - 1 = 0.$$

iii) Verificar se a covariância entre dois contrastes é igual a zero:

$$\hat{Cov}(\hat{Y}_1, \hat{Y}_2) = \sum_{i=1}^{4} a_i b_i = 1.1 + 1.(-1) + (-1).0 + (-1).0 = 0$$

$$\hat{Cov}(\hat{Y}_1, \hat{Y}_3) = \sum_{i=1}^{4} a_i c_i = 1.0 + 1.0 + (-1).1 + (-1).(-1) = 0$$

$$\hat{Cov}(\hat{Y}_2, \hat{Y}_3) = \sum_{i=1}^{4} b_i c_i = 1.0 + (-1).0 + 0.1 + 0.(-1) = 0.$$

iv) Calcular a variância de cada contraste:

$$\hat{Var}(\hat{Y}_1) = s^2 \sum_{i=1}^{n} \frac{a_i^2}{r_i} = 0{,}19\left(\frac{1^2 + 1^2 + (-1)^2 + (-1)^2}{5}\right) = 0{,}1520$$

$$\hat{Var}(\hat{Y}_2) = s^2 \sum_{i=1}^{n} \frac{b_i^2}{r_i} = 0{,}19\left(\frac{1^2 + (-1)^2 + 0^2 + 0^2}{5}\right) = 0{,}0760$$

$$\hat{Var}(\hat{Y}_3) = s^2 \sum_{i=1}^{n} \frac{c_i^2}{r_i} = 0{,}19\left(\frac{0^2 + 0^2 + 1^2 + (-1)^2}{5}\right) = 0{,}0760.$$

v) Calcular o erro padrão de cada contraste:

$$s(\hat{Y}_1) = \sqrt{\hat{Var}(\hat{Y}_1)} = \sqrt{0{,}1520} = 0{,}3899 \text{ t/ha}$$

$$s(\hat{Y}_2) = \sqrt{\hat{Var}(\hat{Y}_2)} = \sqrt{0{,}0760} = 0{,}2757 \text{ t/ha}$$

$$s(\hat{Y}_3) = \sqrt{\hat{Var}(\hat{Y}_3)} = \sqrt{0{,}0760} = 0{,}2757 \text{ t/ha.}$$

vi) Calcular as estimativas destes contrastes:

$$\hat{Y}_1 = 5{,}2 + 3{,}3 - 4{,}0 - 9{,}0 = -4{,}5 \text{ t/ha}$$

$$\hat{Y}_2 = 5{,}2 - 3{,}3 = 1{,}9 \text{ t/ha}$$

$$\hat{Y}_3 = 4{,}0 - 9{,}0 = -5{,}0 \text{ t/ha.}$$

Aplicando o teste de Scott-Knott para variedades dentro de cada nível de inoculante tem-se:

- Variedade dentro do inoculante 1:

Variedade	Médias	Resultado do teste
1	231,4	b
2	385,3	a

A variedade 2 apresentou peso do colmo estatisticamente superior ao da variedade 1 quando foi utilizado o inoculante 1 (Prob<0,05).

- Variedade dentro do inoculante 2:

Variedade	Médias	Resultado do teste
1	209,0	b
2	374,8	a

A variedade 2 apresentou peso do colmo estatisticamente superior ao da variedade 1 quando foi utilizado o inoculante 2 (Prob<0,05).

- Variedade dentro do inoculante 3:

Variedade	Médias	Resultado do teste
1	244,3	b
2	379,3	a

A variedade 2 também apresentou peso do colmo estatisticamente superior ao da variedade 1 quando foi utilizado o inoculante 3 (Prob<0,05).

- Inoculante dentro da variedade 2:

Inoculantes	Médias	Resultado do teste
1	385,3	a
2	374,8	a
3	379,3	a

Também não houve diferenças significativas (Prob>0,05) com relação ao peso do colmo entre os 3 inoculantes utilizados para a variedade 2.

b) Estudar o comportamento das variedades para cada inoculante

Do Quadro 1 tem-se:

SQ Variedade / I_1 = $\frac{1}{4}(925{,}4^2 + 1541{,}0^2) - \frac{(2466{,}4)^2}{8}$ = 47370,4200

SQ Variedade / I_2 = $\frac{1}{4}(835{,}8^2 + 1499{,}0^2) - \frac{(2334{,}8)^2}{8}$ = 54979,2800

SQ Variedade / I_3 = $\frac{1}{4}(977{,}1^2 + 1517{,}1^2) - \frac{(2494{,}2)^2}{8}$ = 36450,0000.

FV	GL	SQ	QM	F	Prob>F
Variedade / I_1	1	47370,4200	47370,4200	25,700	0,0001
Variedade / I_2	1	54979,2800	54979,2800	29,828	0,0001
Variedade / I_3	1	36450,0000	36450,0000	19,775	0,0005
Erro	15	27648,1067	1843,2071		

Neste segundo desdobramento da interação (variedade dentro de inoculante) conclui-se que as duas variedades apresentaram pesos de colmos diferentes (Prob<0,05) para cada inoculante utilizado (I_1 ou I_2 ou I_3).

vii) Conclusões dos contrates:

a) $\hat{Y}_1' = \frac{\hat{m}_1 + \hat{m}_2}{2} - \frac{\hat{m}_3 + \hat{m}_4}{2} =$ -2,25 t/ha

O contraste Y_1 nos indica que o grupo das cultivares 1 e 2 produz em média 2,25 t/ha a menos que o grupo das cultivares 3 e 4.

b) $\hat{Y}_2' = \frac{\hat{m}_1}{1} - \frac{\hat{m}_2}{1} = 1{,}9$ t/ha

O contraste Y_2 nos indica que a cultivar 1 superou em média a produção da cultivar 2 em 1,9 t/ha.

c) $\hat{Y}_3' = \frac{\hat{m}_3}{1} - \frac{\hat{m}_4}{1} = -5{,}0$ t/ha.

O contraste Y_3 nos indica que a cultivar 3 produziu em média 5,0 t/ha a menos que a cultivar 4.

4.2) Teste t de Student

4.2.1) Teste t para contrastes ortogonais

Considerações:

- O teste t pode ser usado para contrastes ortogonais, sugeridos pela estrutura dos tratamentos.

- De acordo com Banzatto & Kronka (1989), deve-se escolher os contrastes antes de avaliar os dados ou, se possível, na fase de planejamento do experimento para evitar que sejam escolhidos contrastes correspondentes as maiores diferenças observadas entre médias, o que aumentaria, assim, a probabilidade de erro tipo I (α). O α consiste no erro que se comete ao rejeitar Ho, sendo que ela é verdadeira.

Dada uma hipótese de nulidade (H_o) e sua hipótese alternativa (H_1) dada por:

H_o: Y= 0, ou seja, as médias ou grupos de médias comparadas no contraste não diferem entre si.

H_1: Y≠ 0, ou seja, pelo menos uma média difere das demais ou um grupo de médias difere de outro grupo.

A estatística t é calculada por:

$$t = \frac{\hat{Y} - 0}{\sqrt{\hat{V}(\hat{Y})}} = \frac{\hat{Y} - 0}{s(\hat{Y})}$$

sendo $\hat{Y}$ a estimativa do contraste de interesse e $s(\hat{Y})$ a estimativa do erro padrão do contraste.

A estatística t é comparada (em valor absoluto) com um valor tabelado (t_t), procurando-se na Tabela de t (encontrada em livros de estatística) o número de graus de liberdade (GL) associado a variância e o nível de significância α. Se $|t| < t_t$, aceita-se a hipótese H_o e conclui-se que as médias ou os grupos de médias em comparação são iguais; caso contrário, se $|t| \geq t_t$, rejeita-se a hipótese H_o e conclui-se que as médias ou o grupo de médias em comparação são diferentes.

Exemplo 12. Aplicar o teste t nos contrates Y_1, Y_2 e Y_3 do Exemplo 11, considerando que o GLErro da análise de variância é 16.

$Y_1 = m_1 + m_2 - m_3 - m_4$ $\quad \hat{Y}_1$ = -4,5 t/ha $\quad s(\hat{Y}_1)$ = 0,3899 t/ha

$Y_2 = m_1 - m_2$ $\quad \hat{Y}_2$ = 1,9 t/ha $\quad s(\hat{Y}_2)$ = 0,2757 t/ha

$Y_3 = m_3 - m_4$ $\quad \hat{Y}_3$ = -5 t/ha $\quad s(\hat{Y}_3)$ = 0,2757 t/ha

- Teste t para Y_1:

$$t_{c(Y_1)} = \frac{-4,5 - 0}{0,3899} = -11,541$$

$t_{t(Y_1)}$ para α=0,05 e GL Erro=16 $\Rightarrow$ $t_{t(Y_1)}$ = 2,12

variedades e inoculantes, recomenda-se proceder o desdobramento da interação V x I para certificar tal informação.

O desdobramento, no caso deste exemplo com dois fatores, pode ser realizado das seguintes maneiras:

a) Estudar o comportamento dos inoculantes para cada variedade

Do Quadro 1 tem-se:

SQ Inoculante / V_1 = $\frac{1}{4}(925,4^2 + 835,8^2 + 977,1^2) - \frac{(2738,3)^2}{12}$ = 2555,5617;

SQ Inoculante / V_2 = $\frac{1}{4}(1541,0^2 + 1499,0^2 + 1517,1^2) - \frac{(4557,1)^2}{12}$ = 221,9017.

FV	GL	SQ	QM	F	Prob>F
Inoculante / V_1	2	2555,5617	1277,7808	0,693	0,5110
Inoculante / V_2	2	221,9017	110,9508	0,060	0,9427
Erro	15	27648,1067	1843,2071		

Neste primeiro desdobramento da interação (inoculante dentro de variedade) conclui-se que tanto para variedade 1 quanto para a variedade 2, não há diferença significativa (Prob>0,05) no peso do colmo entre os três inoculantes aplicados.

Aplicando o teste de Scott-Knott para inoculantes dentro de cada nível de variedade tem-se:

- Inoculante dentro da variedade 1:

Inoculantes	Médias	Resultado do teste
1	231,4	a
2	209,0	a
3	244,3	a

Realmente, não houve diferenças significativas (Prob>0,05) com relação ao peso do colmo entre os 3 inoculantes utilizados para a variedade 1.

SQ Variedades x Inoculantes = 140612,1900 - 137834,7267– 1812,4900 = 964,9733.

E o quadro de análise de variância para os dados do exemplo 5.5.5) conforme o esquema fatorial 3x2 é:

FV	GL	SQ	QM	F	Prob>F
Bloco	3	3806,8083	1268,9361	0,688	0,5730
(Tratamentos)	(5)	(140612,1900)	28122,4380	15,257	0,0000
Variedades (V)	1	137834,7267	137834,7267	74,780	0,0000
Inoculantes (I)	2	1812,4900	906,2450	0,492	0,6211
V x I	2	964,9733	482,4867	0,262	0,7731
Erro	15	27648,1067	1843,2071		
Total	23	172067,1050			
CV (%) =	14,12				
Média geral:	303,98	Número de observações:	24		

Aplicando o teste de Scott-Knott para variedades (pois esta fonte de variação foi significativa: Prob<0,05) tem-se:

Variedades	Médias	Resultado do teste
1	228,2	b
2	379,8	a

Aplicando o teste de Scott-Knott para inoculantes, apesar de seu efeito ter sido não significativo (Prob>0,05), tem-se:

Inoculantes	Médias	Resultado do teste
1	308,3	a
2	291,8	a
3	311,8	a

Embora a interação V x I não seja significativa (Prob > 0,05), indicando não haver uma dependência entre os efeitos dos fatores

Como $\left| t_{c(Y_1)} \right| > t_{t(Y_1)} \Rightarrow \left| -11{,}541 \right| > 2{,}12 \Rightarrow$ rejeita-se H_o: $Y_1 = 0$ e portanto $m_1 + m_2 \neq m_3 + m_4$ (os dois grupos de médias de cultivares diferem entre si ao nível de 5% de significância)

- Teste t para Y_2:

$$t_{c(Y_2)} = \frac{1{,}9 - 0}{0{,}2757} = 6{,}892$$

$t_{t(Y_2)}$ para α=0,05 e GL Erro=16 $\Rightarrow$ $t_{t(Y_2)}$ = 2,12

Como $\left| t_{c(Y_2)} \right| > t_{t(Y_2)} \Rightarrow 6{,}892 > 2{,}12 \Rightarrow$ rejeita-se H_o: $Y_2 = 0$ e portanto $m_1 \neq m_2$ (a média da cultivar 1 difere da cultivar 2 ao nível de 5% de significância)

- Teste t para Y_3:

$$t_{c(Y_3)} = \frac{-5{,}0 - 0}{0{,}2757} = 18{,}136$$

$t_{t(Y3)}$ para α=0,05 e GL Erro=16 $\Rightarrow$ $t_{t(Y_2)}$ = 2,12

Como $\left| t_{c(Y_3)} = \right| > t_{t(Y3)} \Rightarrow 18{,}136 > 2{,}12 \Rightarrow$ rejeita-se H_o: $Y_3 = 0$ e portanto $m_3 \neq m_4$ (a média da cultivar 3 difere da cultivar 4 ao nível de 5% de significância).

4.2.2) Teste t para comparação de duas médias

Passos para realização do teste:

i) Definir a hipótese de nulidade: H_o: $\bar{y}_1 = \bar{y}_2$;

ii) Estabelecer o nível de significância (α);

iii) Calcular a média de cada grupo ($\bar{y}_i$);

iv) Calcular a variância de cada grupo (s_i^2);

v) Calcular a variância ponderada s_p^2, por meio da expressão:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \; ;$$

vi) Calcular a estatística t, por meio da expressão:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)s_p^2}} \; ;$$

vii) Encontrar o valor Tabelado de t (t_t) procurando na Tabela de t o valor correspondente a combinação entre o nível de significância estabelecido, α, e o grau de liberdade (GL) dado por $n_1 + n_2 - 2$;

viii) Concluir o resultado do teste:

Se $|t| < t_t$, aceita-se a hipótese H_o e conclui-se que as duas médias em comparação são iguais;

Se $|t| \geq t_t$, rejeita-se a hipótese H_o e conclui-se que as duas médias em comparação são diferentes.

Exemplo 13. Foi avaliado o volume estimado (em m^3/ha) de madeira produzida por 2 espécies de eucalipto. Verifique se há diferença entre as médias das 2 espécies por meio do teste t.

Espécie	Volume (m^3/ha)
1	24
1	26
1	29
1	32
1	38
2	60
2	63
2	71

SQ Bloco = $\frac{1}{6}(1819{,}5^2 + 1941{,}7^2 + 1802{,}4^2 + 1731{,}8^2) - 2217619{,}2150$

SQ Bloco = 2221426,0233 – 2217619,2150 = 3806,8083.

$\text{SQ Tratamento} = \frac{1}{4}(925{,}4^2 + 835{,}8^2 + 977{,}1^2 + 1541{,}0^2 + 1499{,}0^2 + 1517{,}1) - 2217619{,}2150$

SQ Tratamento = 2358231,4050 – 2217619,2150 = 140612,1900.

SQ Total = $238{,}1^2 + 223{,}6^2 + \ldots + 298{,}4^2 + 363{,}8^2 - 2217619{,}2150$

SQ Total = 2389686,3200 – 2217619,2150 = 172067,1050.

SQ Erro = 172067,1050 – 3806,8083 – 140612,1900 = 27648,1067.

Deve-se montar um quadro auxiliar com os totais de todas as repetições para cada combinação entre os níveis dos fatores.

Quadro 1. Quadro auxiliar com os totais de todas as repetições para cada combinação entre os níveis dos fatores.

	I_1	I_2	I_3	Totais
V_1	925,4 (4)	835,8	977,1	2738,3 (12)
V_2	1541,0	1499,0	1517,1	4557,1
Totais	2466,4 (8)	2334,8	2494,2	7295,4

() os valores dentro de parênteses correspondem ao número de parcelas que deu origem a cada total.

Do Quadro 1 obtém-se:

SQ Variedades = $\frac{1}{12}(2738{,}3^2 + 4557{,}1^2) - 2217619{,}2150$

SQ Variedades = 2355453,9417 – 2217619,2150 = 137834,7267.

SQ Inoculantes = $\frac{1}{8}(2466{,}4^2 + 2334{,}8^2 + 2494{,}2^2) - 2217619{,}2150$

SQ Inoculantes = 2219431,7050– 2217619,2150 = 1812,4900.

SQ Variedades x Inoculantes = [SQ V, I – C] – SQ Variedades – SQ Inoculantes

$\text{SQ Variedades x Inoculantes} = \left[\frac{1}{4}(925{,}4^2 + \cdots + 1517{,}1^2) - 2217619{,}2150\right] - 137834{,}7267 - 1812{,}4900$

peso do colmo (ton/ha). Os dados estão apresentados na Tabela 7 a seguir.

Tabela 7. Peso do colmo (ton/ha) para os 6 tratamentos de um experimento em blocos casualizados (DBC), com 4 repetições, em esquema fatorial 2x3

Tratamentos	Repetições				Totais
	1	2	3	4	
1 – V_1I_1	238,1	256,0	267,7	163,6	925,4
2 – V_1I_2	223,6	217,0	184,7	210,5	835,8
3 – V_1I_3	286,8	205,8	231,6	252,9	977,1
4 – V_2I_1	347,5	403,9	347,0	442,6	1541,0
5 – V_2I_2	351,2	452,5	396,9	298,4	1499,0
6 – V_2I_3	372,3	406,5	374,5	363,8	1517,1
Totais	1819,5	1941,7	1802,4	1731,8	7295,4

5.5.6) Croqui de campo

BL I	2	4	1	3	6	5
BL II	5	2	6	1	4	3
BL III	3	4	5	2	1	6
BL IV	6	1	3	4	5	2

Assim, os valores das somas de quadrados para o exemplo 5.5.5) são:

C = $\frac{(7295{,}4)^2}{24}$ = 2217619,2150.

i) Hipótese de nulidade: Ho: $\bar{y}_1 = \bar{y}_2$;

ii) α=0,05;

iii) $\bar{y}_1 = 29{,}80$ e $\bar{y}_2 = 64{,}67$;

iv) $s_1^2 = 30{,}20$ e $s_2^2 = 32{,}33$;

v) $s^2_{p_{(1,2)}} = \frac{(5-1).30{,}20 + (3-1).32{,}33}{5+3-2}$ = 30,91;

vi) $t_{(1,2)} = \frac{29{,}80 - 64{,}67}{\sqrt{\left(\frac{1}{5}+\frac{1}{3}\right)30{,}91}}$ = -8,588;

vii) $t_{t(1,2)}$ para α=0,05 e GL = 5 + 3 –2 = 6 $\Rightarrow$ $t_{t(1,2)}$ = 2,447;

viii) Comparando a média da espécie 1 com a média da espécie 2 de eucalipto:

$|t_{(1,2)}| > t_{t(1,2)}$

$|-8{,}588|$ = 8,588 > 2,447 $\Rightarrow$ Rejeita-se H_o $\Rightarrow$ Portanto $\bar{y}_1 \neq \bar{y}_2$ (a média da espécie 1 de eucalipto difere da média da espécie 2 de eucalipto ao nível de 5% de probabilidade).

4.2.3) Teste t para comparação de duas médias em uma análise de variância

A diferença mínima significante (DMS ou LSD-Least Significant Difference) entre duas médias pelo teste t de é dada por:

$$DMS = t_t \sqrt{\frac{2.QME}{r}}$$

em que t_t é o valor de t tabelado, o qual corresponde o valor obtido da combinação entre o nível de significância estabelecido (α) e o grau de liberdade do erro (GLE) da análise de variância, na Tabela unilateral de t. O QME é o quadrado médio do erro da análise de variância e r é o número de repetições de cada tratamento.

Quando o valor absoluto da diferença entre duas médias for igual ou maior que a DMS, as médias podem ser consideradas estatisticamente diferentes.

Exemplo 14. Foi realizada a análise de variância para os dados de porcentagem de absorção de água de 5 linhagens de feijão, com 3 repetições por linhagem. O valor do grau de liberdade do erro (GLE) foi 10 e o quadrado médio do erro (QME) foi 4,08. Compare as médias dos tratamentos a seguir pelo teste t:

$$\bar{y}_1 = 95{,}5$$

$$\bar{y}_2 = 87{,}8$$

$$\bar{y}_3 = 86{,}9$$

$$\bar{y}_4 = 26{,}3$$

$$\bar{y}_5 = 108{,}2$$

i) t_t para α=0,05 e GLE = 10 $\Rightarrow$ t_t = 2,228;

ii) $\text{DMS} = 2{,}228\sqrt{\dfrac{2.(4{,}08)}{3}} = 3{,}67$;

iii) Coloque as médias em ordem decrescente e faça a diferença entre elas duas a duas, começando da diferença entre a maior e a menor média e assim por diante:

$$\bar{y}_5 = 108{,}2$$

$$\bar{y}_1 = 95{,}5$$

$$\bar{y}_2 = 87{,}8$$

$$\bar{y}_3 = 86{,}9$$

$$\bar{y}_4 = 26{,}3$$

5.5.3) Desvantagens

como os tratamentos correspondem a todas as combinações possíveis entre os níveis dos fatores, o número de tratamentos a ser avaliado pode aumentar muito, não podendo ser distribuídos em blocos completos casualizados devido à exigência de homogeneidade das parcelas dentro de cada bloco. Isto pode levar a complicações na análise, sendo preciso lançar mão de algumas técnicas alternativas (como por exemplo, o uso de blocos incompletos).

A análise estatística e a interpretação dos resultados pode tornar-se um pouco mais complicada que nos experimentos simples.

5.5.4) Modelo estatístico do fatorial

O modelo a seguir corresponde a um modelo de um delineamento em blocos casualizados (DBC) em esquema fatorial com 2 fatores (α e γ), mas pode ser estendido para os casos em que há mais fatores, incluindo os fatores isolados e as interações duplas, triplas e outras entre os fatores.

$$y_{ijk} = \mu + \beta_j + \alpha_i + \gamma_k + (\alpha\gamma)_{ik} + e_{ijk}$$

em que, y_{ijk} é o valor observado referente a parcela que recebeu o i-ésimo nível do fator α e o k-ésimo nível do fator γ no j-ésimo bloco; μ representa uma constante geral; β_j representa o efeito do j-ésimo bloco; α_i representa o efeito do i-ésimo nível do fator α; γ representa o efeito do k-ésimo nível do fator γ; $(\alpha\gamma)_{ik}$ representa a interação entre o efeito do i-ésimo nível do fator α e o efeito do do k-ésimo nível do fator γ e e_{ijk} representa o erro experimental associado à observação y_{ijk}, suposto ter distribuição normal com média zero e variância comum.

5.5.5) Exemplo de fatorial

Em um experimento em blocos casualizados com 4 repetições, no esquema fatorial 2x3 foi avaliado o efeito de 2 variedades de cana-de-açúcar (V_1 e V_2) e 3 tipos de inoculantes (I_1, I_2 e I_3) quanto ao

5.5) Experimentos fatoriais

5.5.1) Características

Em alguns experimentos, o pesquisador avalia dois ou mais tipos de tratamentos e deseja verificar se há interação entre estes tipos. Tais experimentos são denominados experimentos fatoriais e os tipos de tratamentos são denominados fatores. As categorias (subdivisões) de cada fator são ditas níveis do fator. Como exemplo, considere um experimento em que se comparou o efeito de 3 estirpes de rizóbio (BR 9001, BR 9004 e BR 4812) e o efeito de um determinado fungo (presença e ausência do fungo) na variável número de nódulos produzido pelo feijão. Neste caso, existem dois fatores: estirpe de rizóbio e a ocorrência do fungo. Os níveis do fator estirpe são 3 (BR 9001, BR 9004 e BR 48122) e do fungo são 2 (presença e ausência).

Costuma-se representar o fatorial pela multiplicação dos níveis. No exemplo anterior o fatorial é 3x2 (fatorial 3 por 2), assim fica claro que existem dois fatores, o primeiro fator com 3 níveis de estirpe e o segundo com 2 níveis de fungo. O número total de tratamentos avaliados também é dado pela multiplicação dos níveis, ou seja, no exemplo são avaliados 3x2 = 6 tratamentos avaliados (1: BR 9001 na presença do fungo; 2: BR 9004 na presença do fungo; 3: BR 4812 na presença do fungo; 4: BR 9001 na ausência do fungo; 5: BR 9004 na ausência do fungo; 6: BR 4812 na ausência do fungo. Se fossem, por exemplo, 3 fatores com 5, 2 e 3 níveis para cada fator respectivamente, a representação seria: fatorial 5x2x3, sendo avaliado um total de 30 tratamentos e assim por diante.

Vale lembrar que os experimentos fatoriais não são delineamentos e sim um esquema de desdobramento de graus de liberdade de tratamentos, e podem ser instalado em qualquer dos delineamentos experimentais, DIC, DBC, etc. (Banzatto & Kronka, 1989).

5.5.2) Vantagens

- Permite estudar os efeitos principais dos fatores e os efeitos das interações entre eles.

$\bar{y}_5 - \bar{y}_4 = 108{,}2-26{,}3 = 81{,}9 \Rightarrow 81{,}9 > 3{,}67 \therefore \bar{y}_5 \neq \bar{y}_4;$

$\bar{y}_5 - \bar{y}_3 = 108{,}2-86{,}9 = 21{,}3 \Rightarrow 21{,}3 > 3{,}67 \therefore \bar{y}_5 \neq \bar{y}_3;$

$\bar{y}_5 - \bar{y}_2 = 108{,}2-87{,}8 = 20{,}4 \Rightarrow 20{,}4 > 3{,}67 \therefore \bar{y}_5 \neq \bar{y}_2;$

$\bar{y}_5 - \bar{y}_1 = 108{,}2-95{,}5 = 12{,}7 \Rightarrow 12{,}7 > 3{,}67 \therefore \bar{y}_5 \neq \bar{y}_1;$

$\bar{y}_1 - \bar{y}_4 = 95{,}5-26{,}3 = 69{,}2 \Rightarrow 69{,}2 > 3{,}67 \therefore \bar{y}_1 \neq \bar{y}_4;$

$\bar{y}_1 - \bar{y}_3 = 95{,}5-86{,}9 = 8{,}6 \Rightarrow 8{,}6 > 3{,}67 \therefore \bar{y}_1 \neq \bar{y}_3;$

$\bar{y}_1 - \bar{y}_2 = 95{,}5-87{,}8 = 7{,}7 \Rightarrow 7{,}7 > 3{,}67 \therefore \bar{y}_1 \neq \bar{y}_2;$

$\bar{y}_2 - \bar{y}_4 = 87{,}8-26{,}3 = 61{,}5 \Rightarrow 61{,}5 > 3{,}67 \therefore \bar{y}_2 \neq \bar{y}_4;$

$\bar{y}_2 - \bar{y}_3 = 87{,}8-86{,}9 = 0{,}9 \Rightarrow 0{,}9 < 3{,}67 \therefore \bar{y}_2 = \bar{y}_3;$

$\bar{y}_3 - \bar{y}_4 = 86{,}9-26{,}3 = 60{,}6 \Rightarrow 60{,}6 > 3{,}67 \therefore \bar{y}_3 \neq \bar{y}_4;$

iv) Coloque letras iguais para médias semelhantes e letras distintas para médias que diferem entre si e interprete o teste.

$\bar{y}_5 = 108{,}2$ a

$\bar{y}_1 = 95{,}5$ b

$\bar{y}_2 = 87{,}8$ c

$\bar{y}_3 = 86{,}9$ c

$\bar{y}_4 = 26{,}3$ d

A linhagem 5 foi a que apresentou maior porcentagem de absorção de água diferindo das demais linhagens (Prob < 0,05).

4.3) Teste de Tukey

A diferença mínima significante (D.M.S.) entre duas médias pelo teste de Tukey é dada por:

$$DMS = q\sqrt{\frac{QME}{r}}$$

em que q é um valor tabelado, o qual corresponde o valor obtido da combinação entre o número de tratamentos (I) e o grau de liberdade do erro (GLE) da análise de variância, para um nível de significância estabelecido (α). O QME e r já foram descritos no teste t.

A interpretação é a mesma do teste t, ou seja, quando o valor absoluto da diferença entre duas médias for igual ou maior que a DMS, as médias podem ser consideradas estatisticamente diferentes.

Exemplo 15: Compare as médias dos tratamentos do Exemplo14 pelo teste de Tukey.

i) q para α=0,05; I = 5 tratamentos e GLE = 10 $\Rightarrow$ q = 4,65;

ii) $DMS = 4{,}65\sqrt{\frac{4{,}08}{3}}$ = 5,42;

iii) $\bar{y}_5 - \bar{y}_4$ = 81,9 > 5,42 $\therefore$ $\bar{y}_5 \neq \bar{y}_4$;

$\bar{y}_5 - \bar{y}_3$ = 21,3 > 5,42 $\therefore$ $\bar{y}_5 \neq \bar{y}_3$;

$\bar{y}_5 - \bar{y}_2$ = 20,4 > 5,42 $\therefore$ $\bar{y}_5 \neq \bar{y}_2$;

$\bar{y}_5 - \bar{y}_1$ = 12,7 > 5,42 $\therefore$ $\bar{y}_5 \neq \bar{y}_1$;

$\bar{y}_1 - \bar{y}_4$ = 69,2 > 5,42 $\therefore$ $\bar{y}_1 \neq \bar{y}_4$;

$\bar{y}_1 - \bar{y}_3$ = 8,6 > 5,42 $\therefore$ $\bar{y}_1 \neq \bar{y}_3$;

$\bar{y}_1 - \bar{y}_2$ = 7,7 > 5,42 $\therefore$ $\bar{y}_1 \neq \bar{y}_2$;

$\bar{y}_2 - \bar{y}_4$ = 61,5 > 5,42 $\therefore$ $\bar{y}_2 \neq \bar{y}_4$;

$\bar{y}_2 - \bar{y}_3$ = 0,9 < 5,42 $\therefore$ $\bar{y}_2 = \bar{y}_3$;

$\bar{y}_3 - \bar{y}_4$ = 60,6 > 5,42 $\therefore$ $\bar{y}_3 \neq \bar{y}_4$;

SQ Bloco = $\frac{1}{4}(253{,}6^2 + 250{,}7^2 + 249{,}0^2 + 218{,}8^2 + 215{,}0^2) - 70460{,}3205$

SQ Bloco = 70815,7225 – 70460,3205 = 355,4020.

SQ Tratamento = $\frac{1}{5}(291{,}1^2 + 343{,}8^2 + 291{,}6^2 + 259{,}6^2) - 70460{,}3205$

SQ Tratamento = 71188,7140 – 70460,3205 = 728,3935.

SQ Total = $72{,}8^2 + 58{,}3^2 + \ldots + 27{,}4^2 + 39{,}0^2$ – 70460,3205

SQ Total = 73209,0700 – 70460,3205 = 2748,7495.

SQ Erro = 2748,7495 – 355,4020 - 728,3935 = 1664,9540.

E o quadro de análise de variância para os dados do Exemplo 5.4.5) é:

FV	GL	SQ	QM	F	Prob>F
Bloco	4	355,4020	88,8505	0,640	0,6441
Cobertura morta	3	728,3935	242,7978	1,750	0,2100
Erro	12	1664,9540	138,7462		
Total	19	2748,7495			
CV (%) =	19,83				
Média geral:	59,4	Número de observações: 20			

Como Prob > 0,05 para cobertura morta, conclui-se que as quatro coberturas mortas tiveram influência semelhante no peso seco do brócolis. Neste caso, não há necessidade de aplicação de um teste de comparação múltipla.

Observação: Se o valor de F para tratamento for significativo a determinado nível α de significância, o pesquisador pode usar um teste de comparação múltipla para comparar as médias dos tratamentos (caso este seja qualitativo), diz-se então que o teste usado é protegido; caso contrário, se F for não significativo, o pesquisador poderá optar ou não pelo uso do teste e, então, diz-se que o teste é não protegido.

5.4.7) Esquema de análise de variância do DBC com fontes de variação e graus de liberdade

Considerando a mesma representação da Tabela 5, porém aqui, as repetições representam os blocos, o quadro de análise de variância para os dados de um delineamento em blocos casualizados (DBC) é expresso de uma maneira geral por:

FV	GL	SQ	QM	F
Bloco	J-1	$\frac{1}{I}\sum_{j=1}^{J} B_j^2 - C$	SQBloco/GLBloco	QMBloco/QMErro
Tratamento	I-1	$\frac{1}{J}\sum_{i=1}^{I} T_i^2 - C$	SQTrat./GLTrat	QMTrat./QMErro
Erro	(I-1)(J-1)	SQTotal -SQBloco-SQTrat.	SQErro/GLErro	
Total	IJ-1	$\sum_{i,j=1}^{I,J} y_{ij}^2 - C$		
CV(%)= $\frac{\sqrt{QMErro}}{\bar{y}}$ .100				
$\bar{y} = \sum_{i,j=1}^{I,J} y_{ij} / I.J$				

No exemplo 5.4.5) tem-se:

- Delineamento: DBC;

- Tratamentos: I = 4 tipos de cobertura morta (sorgo, crotalária, milheto e vegetação espontânea);

- Repetições: J = 5;

- Variável a analisar: peso seco (g/parcela).

Assim, os valores das somas de quadrados para o exemplo 5.3.5) são:

iv)

$\bar{y}_5$ =108,2 a

$\bar{y}_1$ = 95,5 b

$\bar{y}_2$ = 87,8 c

$\bar{y}_3$ = 86,9 c

$\bar{y}_4$ = 26,3 d

4.4) Teste de Duncan

A diferença mínima significante (DMS) entre duas médias pelo teste de Duncan é dada por:

$$DMS = z_n \sqrt{\frac{QME}{r}} .$$

No teste de Duncan, se estão envolvidos I tratamentos no estudo, é necessário calcular I-1 DMS's. O que diferencia uma DMS da outra é o valor de z_n que é um valor tabelado, o qual corresponde ao valor obtido da combinação entre o número de médias ordenadas abrangidas na comparação (n), e o grau de liberdade do erro (GLE) da análise de variância, ao nível de significância estabelecido (α). O QME e r já foram descritos nos testes anteriores.

Para realização deste teste deve-se também ordenar as médias em ordem decrescente e ir fazendo a diferença sempre entre a maior e menor média, observando assim, o número (n) de médias ordenadas abrangidas na comparação.

A interpretação é a mesma dos testes anteriores, ou seja, quando o valor absoluto da diferença entre duas médias for igual ou maior que a D.M.S, as médias podem ser consideradas estatisticamente diferentes. A única diferença é que na comparação entre duas médias deve-se considerar o valor de DMS correspondente ao n em questão. Com o exemplo seguinte ficará mais fácil o entendimento.

Exemplo 16: Compare as médias dos tratamentos do Exemplo 14 pelo teste de Duncan.

i) Como no exemplo tem-se I=5 médias de tratamentos é necessário calcular z_5, z_4, z_3, e z_2 , ou seja, é necessário o cálculo de $z_I = z_5$ até z_2:

z_5 para α=0,05; n = 5 e GLE = 10 $\Rightarrow z_5 = 3{,}430$;

z_4 para α=0,05; n = 4 e GLE = 10 $\Rightarrow z_4 = 3{,}376$;

z_3 para α=0,05; n = 3 e GLE = 10 $\Rightarrow z_3 = 3{,}293$;

z_2 para α=0,05; n = 2 e GLE = 10 $\Rightarrow z_2 = 3{,}151$;

ii) Calcula-se então as I-1= 4 DMS's:

$$\mathrm{DMS}_5 = 3{,}430\sqrt{\frac{4{,}08}{3}} = 4{,}00\,;$$

$$\mathrm{DMS}_4 = 3{,}376\sqrt{\frac{4{,}08}{3}} = 3{,}94\,;$$

$$\mathrm{DMS}_3 = 3{,}293\sqrt{\frac{4{,}08}{3}} = 3{,}84\,;$$

$$\mathrm{DMS}_2 = 3{,}151\sqrt{\frac{4{,}08}{3}} = 3{,}67.$$

iii) Lembrando que as médias colocadas em ordem decrescente são:

$\bar{y}_5$ =108,2

$\bar{y}_1$ = 95,5

$\bar{y}_2$ = 87,8

$\bar{y}_3$ = 86,9

$\bar{y}_4$ = 26,3

Tabela 6. Peso seco (kg/parcela) de brócolis em um experimento em blocos casualizados (DBC) com 5 repetições em que foi avaliada a influência de 4 tipos de cobertura morta (1: sorgo, 2: crotalária; 3: milheto e 4: vegetação espontânea)

Rep. \ Trat.	1	2	3	4	Total
1	72,8	69,0	45,3	66,5	253,6
2	58,3	64,1	60,9	67,4	250,7
3	50,4	72,1	67,2	59,3	249,0
4	51,6	73,6	66,2	27,4	218,8
5	59,0	65,0	52,0	39,0	215,0
Total	292,1	343,8	291,6	259,6	1187,1
Média	58,4	68,8	58,3	51,9	59,4
Correção (C)					C = $(1187{,}1)^2/20 = 70460{,}3205$
n					n = 4.5 = 20

5.4.6) Croqui de campo

BL I	2	3	1	4
BL II	4	1	2	3
BL III	2	1	4	3
BL IV	3	2	1	4
BL V	1	4	3	2

A disposição dos tratamentos é realizada de forma aleatória dentro dos blocos.

5.4.2) Vantagens

- Controla diferenças nas condições ambientais de um bloco para outro.

- Leva a uma estimativa mais exata da variância residual ($\hat{\sigma}^2$), uma vez que a variação ambiental entre blocos é isolada.

5.4.3) Desvantagens

- Há uma redução no número de graus de liberdade do erro pois o DBC utiliza o princípio do controle local.

- O número de tratamentos a ser utilizado é limitado pela exigência de homogeneidade dentro dos blocos, não podendo ser muito elevado.

5.4.4)Modelo estatístico do DBC

$$y_{ij} = \mu + b_j + t_i + e_{ij}$$

em que, y_{ij} representa a observação do i-ésimo tratamento no j-ésimo bloco; μ representa uma constante geral associada a esta variável aleatória; bj representa o efeito do j-ésimo bloco; ti representa o efeito do i-ésimo tratamento; e eij representa o erro experimental associado a observação yij, suposto ter distribuição normal com média zero e variância comum.

5.4.5) Exemplo de DBC

Estudou-se a influência de 4 tipos de cobertura morta (sorgo, crotalária, milheto e vegetação espontânea) no peso seco de brócolis. O experimento foi instalado em DBC com 5 repetições. Os dados de peso seco estão dispostos na Tabela 6 a seguir.

E as diferenças entre elas duas a duas:

$\bar{y}_5 - \bar{y}_4 = 81{,}9 \Rightarrow n = 5 \therefore$ compara-se 81,9 com a $DMS_5 \Rightarrow 81{,}9 > 4{,}00 \therefore \bar{y}_5 \neq \bar{y}_4$;

$\bar{y}_5 - \bar{y}_3 = 21{,}3 \Rightarrow n = 4 \therefore$ compara-se 21,3 com a $DMS_4 \Rightarrow 21{,}3 > 3{,}94 \therefore \bar{y}_5 \neq \bar{y}_3$;

$\bar{y}_5 - \bar{y}_2 = 20{,}4 \Rightarrow n = 3 \therefore$ compara-se 20,4 com a $DMS_3 \Rightarrow 20{,}4 > 3{,}84 \therefore \bar{y}_5 \neq \bar{y}_2$;

$\bar{y}_5 - \bar{y}_1 = 12{,}7 \Rightarrow n = 2 \therefore$ compara-se 12,7 com a $DMS_2 \Rightarrow 12{,}7 > 3{,}67 \therefore \bar{y}_5 \neq \bar{y}_1$;

$\bar{y}_1 - \bar{y}_4 = 69{,}2 \Rightarrow n = 4 \therefore$ compara-se 69,2 com a $DMS_4 \Rightarrow 69{,}2 > 3{,}94 \therefore \bar{y}_1 \neq \bar{y}_4$;

$\bar{y}_1 - \bar{y}_3 = 8{,}6 \Rightarrow n = 3 \therefore$ compara-se 8,6 com a $DMS_3 \Rightarrow 8{,}6 > 3{,}84 \therefore \bar{y}_1 \neq \bar{y}_3$;

$\bar{y}_1 - \bar{y}_2 = 7{,}7 \Rightarrow n = 2 \therefore$ compara-se 7,7 com a $DMS_2 \Rightarrow 7{,}7 > 3{,}67 \therefore \bar{y}_1 \neq \bar{y}_2$;

$\bar{y}_2 - \bar{y}_4 = 61{,}5 \Rightarrow n = 3 \therefore$ compara-se 61,5 com a $DMS_3 \Rightarrow 61{,}5 > 3{,}84 \therefore \bar{y}_2 \neq \bar{y}_4$;

$\bar{y}_2 - \bar{y}_3 = 0{,}9 \Rightarrow n = 2 \therefore$ compara-se 0,9 com a $DMS_2 \Rightarrow 0{,}9 < 3{,}67 \therefore \bar{y}_2 = \bar{y}_3$;

$\bar{y}_3 - \bar{y}_4 = 60{,}6 \Rightarrow n = 2 \therefore$ compara-se 60,6 com a $DMS_2 \Rightarrow 60{,}6 > 3{,}67 \therefore \bar{y}_3 \neq \bar{y}_4$.

iv) Coloque letras iguais para médias semelhantes e letras distintas para médias que diferem entre si e interprete o teste.

$\bar{y}_5 = 108{,}2$ a

$\bar{y}_1 = 95{,}5$ b

$\bar{y}_2 = 87{,}8$ c

$\bar{y}_3 = 86{,}9$ c

$\bar{y}_4 = 26{,}3$ d

4.5) Teste de SNK (Student Newman Keuls)

A diferença mínima significante (DMS) entre duas médias pelo teste de SNK é dada por:

$$DMS_n = q_n \sqrt{\frac{QME}{r}}.$$

O procedimento para a realização deste teste é bastante semelhante ao do teste de Duncan. A diferença é que nas DMS's do SNK são usados os valores tabelados de q_n ao invés de z_n, ou seja, deve-se procurar o valor tabelado na Tabela de q ao nível de significância estebelecido (α), correspondente a combinação entre o número de médias abrangidas na comparação (n) e o grau de liberdade do erro (GLE) da análise de variância.

Exemplo 17. Compare as médias dos tratamentos do Exemplo 14 pelo teste de SNK.

i) Como no exemplo tem-se I=5 médias de tratamentos, é necessário calcular q_5, q_4, q_3, e q_2 , ou seja, é necessário o cálculo de $q_I = q_5$ até q_2:

q_5 para α=0,05; n = 5 e GLE = 10 $\Rightarrow q_5 = 4,65$;

q_4 para α=0,05; n = 4 e GLE = 10 $\Rightarrow q_4 = 4,33$;

q_3 para α=0,05; n = 3 e GLE = 10 $\Rightarrow q_3 = 3,88$;

q_2 para α=0,05; n = 2 e GLE = 10 $\Rightarrow q_2 = 3,15$;

ii) Calcula-se então as I-1= 4 DMS's:

$$DMS_5 = 4,65\sqrt{\frac{4,08}{3}} = 5,42;$$

$$DMS_4 = 4,33\sqrt{\frac{4,08}{3}} = 5,05;$$

$$DMS_3 = 3,88\sqrt{\frac{4,08}{3}} = 4,52;$$

$$DMS_2 = 3,15\sqrt{\frac{4,08}{3}} = 3,67.$$

E o quadro de análise de variância para os dados do Exemplo 5.3.5) é:

FV	GL	SQ	QM	F	Prob>F
Variedades	3	13019,0330	4339,6776	5,668	0,0056
Erro	20	15314,0178	765,7009		
Total	23	28333,0508			
CV (%):	18,41				
Média: $\bar{y}$:	150,31	Número de observações:	24		

Como Prob < 0,05 (valor fornecido por alguns programas computacionais de análise de variância), conclui-se que há diferença estatística significativa entre as médias de peso seco da parte aérea das quatro variedades de cana-de açúcar. Deve-se então aplicar algum dos testes de comparação múltipla nestas médias.

5.4) Delineamento em Blocos Casualizados (DBC)

5.4.1) Características

Os tratamentos são distribuídos aleatoriamente em blocos (princípio do controle local) de modo que haja maior uniformidade possível dentro de cada bloco.

O número de parcelas por bloco é igual ao número de tratamentos, ou seja, cada bloco deverá conter todos os tratamentos.

O DBC possui os três princípios básicos da experimentação: casualização, repetição e controle local e, portanto, as repetições são organizadas em blocos.

Normalmente, é o delineamento mais utilizado em condições de campo. A eficiência do DBC depende da uniformidade dentro de cada bloco, podendo haver heterogeneidade entre blocos. Os blocos podem ser instalados na forma quadrada, retangular ou irregular, desde que seja respeitada a uniformidade dentro do bloco.

O quadro de análise de variância para os dados da Tabela 5 é:

FV	GL	SQ	QM	F
Tratamento	I-1	$\frac{1}{J}\sum_{i=1}^{I} T_i^2 - C$	SQTrat./GLTrat.	QMTrat./QMErro
Erro	I(J-1)	SQTotal –SQTrat.	SQErro/GLErro	
Total	IJ-1	$\sum_{i,j=1}^{I,J} y_{ij}^2 - C$		
CV(%)= $\frac{\sqrt{QMErro}}{\bar{y}}$.100				
$\bar{y} = \sum_{i,j=1}^{I,J} y_{ij} / I.J$				

No exemplo 5.3.5) tem-se:

- Delineamento: DIC;
- Tratamentos: I = 4 variedades de cana-de-açúcar (A, B, C, D);
- Repetições: J = 6;
- Variável a analisar: peso seco da parte aérea (g/parcela).

Assim, os valores das somas de quadrados para o exemplo 5.3.5) são:

SQ Tratamento = $\frac{1}{6}(667{,}59^2 + 1005{,}30^2 + 1011{,}87^2 + 922{,}59^2) - 542207{,}25$09

SQ Tratamento = 555226,28389 – 542207,2509 = 13019,0330.

SQ Total = $113{,}83^2 + 133{,}89^2 + \ldots + 922{,}59^2 + 153{,}77^2$ – 542207,25

SQ Total = 570540,3017 - 542207,2509 = 28333,0508.

SQ Erro = 28333,0508 – 13019,0330 = 15314,0178.

iii) Lembrando que as médias colocadas em ordem decrescente são:

$\bar{y}_5$ =108,2

$\bar{y}_1$ = 95,5

$\bar{y}_2$ = 87,8

$\bar{y}_3$ = 86,9

$\bar{y}_4$ = 26,3

E as diferenças entre elas duas a duas:

$\bar{y}_5 - \bar{y}_4$ = 81,9 $\Rightarrow$ n = 5 $\therefore$ compara-se 81,9 com a $DMS_5 \Rightarrow$ 81,9 > 4,65 $\therefore$ $\bar{y}_5 \neq \bar{y}_4$;

$\bar{y}_5 - \bar{y}_3$ = 21,3 $\Rightarrow$ n = 4 $\therefore$ compara-se 21,3 com a $DMS_4 \Rightarrow$ 21,3 > 4,33 $\therefore$ $\bar{y}_5 \neq \bar{y}_3$;

$\bar{y}_5 - \bar{y}_2$ = 20,4 $\Rightarrow$ n = 3 $\therefore$ compara-se 20,4 com a $DMS_3 \Rightarrow$ 20,4 > 3,88 $\therefore$ $\bar{y}_5 \neq \bar{y}_2$;

$\bar{y}_5 - \bar{y}_1$ = 12,7 $\Rightarrow$ n = 2 $\therefore$ compara-se 12,7 com a $DMS_2 \Rightarrow$ 12,7> 3,15 $\therefore$ $\bar{y}_5 \neq \bar{y}_1$;

$\bar{y}_1 - \bar{y}_4$ = 69,2 $\Rightarrow$ n = 4 $\therefore$ compara-se 69,2 com a $DMS_4 \Rightarrow$ 69,2 > 4,33 $\therefore$ $\bar{y}_1 \neq \bar{y}_4$;

$\bar{y}_1 - \bar{y}_3$ = 8,6 $\Rightarrow$ n = 3 $\therefore$ compara-se 8,6 com a $DMS_3 \Rightarrow$ 8,6 > 3,88 $\therefore$ $\bar{y}_1 \neq \bar{y}_3$;

$\bar{y}_1 - \bar{y}_2$ = 7,7 $\Rightarrow$ n = 2 $\therefore$ compara-se 7,7 com a $DMS_2 \Rightarrow$ 7,7 > 3,15 $\therefore$ $\bar{y}_1 \neq \bar{y}_2$;

$\bar{y}_2 - \bar{y}_4$ = 61,5 $\Rightarrow$ n = 3 $\therefore$ compara-se 61,5 com a $DMS_3 \Rightarrow$ 61,5 > 3,88 $\therefore$ $\bar{y}_2 \neq \bar{y}_4$;

$\bar{y}_2 - \bar{y}_3$ = 0,9 $\Rightarrow$ n = 2 $\therefore$ compara-se 0,9 com a $DMS_2 \Rightarrow$ 0,9 < 3,15 $\therefore$ $\bar{y}_2 = \bar{y}_3$;

$\bar{y}_3 - \bar{y}_4$ = 60,6 $\Rightarrow$ n = 2 $\therefore$ compara-se 60,6 com a $DMS_2 \Rightarrow$ 60,6 > 3,15 $\therefore$ $\bar{y}_3 \neq \bar{y}_4$.

v) Coloque letras iguais para médias semelhantes e letras distintas para médias que diferem entre si e interprete o teste.

$\bar{y}_5$ =108,2 a

$\bar{y}_1$ = 95,5 b

$\bar{y}_2$ = 87,8 c

$\bar{y}_3$ = 86,9 c

$\bar{y}_4$ = 26,3 d

4.6) Teste de Scott-Knott

O procedimento de Scott e Knott (1974) utiliza a razão de verossimilhança para atestar a significância de que os n tratamentos podem ser divididos em dois grupos que maximizem a soma de quadrados entre grupos (Ramalho et al., 2000).

Seja por exemplo 3 tratamentos, A, B e C. O processo consiste em determinar uma partição, em dois grupos, que maximize a soma de quadrados. Veja que nesse caso são possíveis 2^{n-1} grupos, isto é, A vs B e C, B vs A e C e C vs A e B. Com um número pequeno de tratamentos como o do exemplo, é fácil obter todos os grupos. Contudo, quando o número (n) de tratamentos é grande, o número de grupos cresce exponencialmente, dificultando a aplicação do teste. Para atenuar esse problema, basta ordenar as médias dos tratamentos. Nessa situação, o número de partições possíveis passa a ser obtido por n-1. Uma vez ordenada as médias, procede-se do seguinte modo, fazendo inicialmente o número de tratamentos envolvidos no grupo de médias considerado(g) igual ao o número total de tratamentos (n).

i) Determinar a partição entre dois grupos que maximiza a soma de quadrados (SQ) entre grupos. Seja T_1 e T_2 os totais

A disposição das repetições de cada tratamento é realizada de forma totalmente aleatória às parcelas.

5.3.7) Esquema de análise de variância do DIC com fontes de variação e graus de liberdade

Imagine um experimento com I tratamentos e cada tratamento com J repetições representado na Tabela a seguir.

Tabela 5. Representação esquemática dos dados de um delineamento inteiramente casualizado

Rep. \ Trat.	1	2	3	$\cdots$	I	
1	y_{11}	y_{21}	y_{31}	$\cdots$	y_{I1}	
2	y_{12}	y_{22}	y_{32}	$\cdots$	y_{I2}	
3	y_{13}	y_{23}	y_{33}	$\cdots$	y_{I3}	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
J	y_{1J}	y_{2j}	y_{3j}	$\cdots$	y_{Ij}	
Total	T_1	T_2	T_3	$\cdots$	T_I	$\sum_{i=1}^{I} T_i = \sum_{i=1}^{I} y_{ij}$
Média	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\cdots$	$\bar{y}_I$	$\bar{y} = \sum_{i,j=1}^{I,J} y_{ij} / I.J$
Correção (C)						$c = \left(\sum_{i,j=1}^{I,J} y_{ij}\right)^2 / I.J$
n						n=IJ

5.3.5) Exemplo de DIC

Suponha que foi avaliado o peso seco da parte aérea (g/parcela) de 4 variedades de cana-de-açúcar. O experimento foi instalado em casa de vegetação. O delineamento foi o inteiramente casualizado com 6 repetições. Cada parcela era constituída de 1 vaso com 3 plantas. Os dados de peso estão dispostos na Tabela a seguir:

Tabela 4. Peso seco da parte aérea (g/parcela) de 4 variedades de cana-de-açúcar (A, B, C e D) em um delineamento inteiramente casualizado com 6 repetições

Rep. \ Trat.	A	B	C	D	
1	113,83	174,94	213,39	166,76	
2	133,89	168,76	86,69	131,17	
3	96,15	156,35	157,65	177,88	
4	101,22	144,89	174,44	121,23	
5	95,16	181,57	187,00	180,94	
6	127,34	178,79	192,70	144,61	
Total	667,59	1005,30	1011,87	922,59	3607,35 (total geral)
Média	111,27	167,55	168,65	153,77	150,31 (média geral)
Correção (C)					C = $(3607,35)^2/24 = 542207,2509$
n					n = 4.6 = 24

5.3.6) Croqui de campo

C	A	B	B
D	D	C	A
C	A	D	B
B	C	B	A
C	A	D	B
A	C	D	D

dos dois grupos com k_1 e k_2 tratamentos em cada um, a soma de quadrados B_o é estimada por:

$$B_o = \frac{T_1^2}{k_1} + \frac{T_2^2}{k_2} - \frac{(T_1 + T_2)^2}{k_1 + k_2}$$

$$T_1 = \sum_{i=1}^{k_1} \bar{y}_{(i)} \quad \text{e} \quad T_2 = \sum_{i=k_1+1}^{g} \bar{y}_{(i)}$$

em que $\bar{y}_{(i)}$ é a média do tratamento da posição ordenada i.

Os dois grupos deverão ser identificados por meio da inspeção das somas de quadrados das g-1 partições possíveis, sendo g o número de tratamentos envolvidos no grupo de médias considerado.

ii) Determinar o valor da estatística λ:

$$\lambda = \frac{\pi}{2(\pi - 2)} \cdot \frac{B_o}{\hat{\sigma}_o^2}$$

em que $\hat{\sigma}_o^2$ é o estimador de máxima verossimilhança de $\sigma_{\bar{y}}^2$ dado por:

$$\hat{\sigma}_o^2 = \frac{1}{g+v}\left[\sum_{i=1}^{g} (\bar{y}_{(i)} - \bar{y})^2 + v.s_{\bar{y}}^2\right]$$

em que v é o grau de liberdade do erro associado a este estimador, $\bar{y}$ é a média das médias dos g tratamentos e $s_{\bar{y}}^2 = \frac{QME}{r}$ é o estimador não viesado de $\sigma_{\bar{y}}^2$, sendo QME o quadrado médio do erro e r o número de repetições.

iii) Se $\lambda \geq \chi^2_{(\alpha; g/(\pi-2))}$ rejeita-se a hipótese de que os dois grupos são idênticos em favor da hipótese alternativa de que os dois grupos diferem. $\chi^2_{(\alpha; g/(\pi-2))}$ é um valor tabelado obtido na Tabela de

Qui-quadrado (encontrada em alguns livros de estatística), correspondente a combinação entre o nível de significância estebelecido (α) e o valor dado por $g/(\pi-2)$.

iv) No caso de rejeitar esta hipótese, os dois subgrupos formados serão independentemente submetidos aos passos i) a iii), fazendo respectivamente $g=k_1$ e $g=k_2$. O processo em cada subgrupo se encerra ao se aceitar H_o no passo iii) ou se cada subgrupo contiver apenas uma média.

Exemplo 18. Agora vamos aplicar o algoritmo do teste de Scott e Knott nas médias do Exemplo 14 em que o quadrado médio do erro foi de 4,08 com 10 graus de liberdade, e as médias das 5 linhagens de feijão estimadas a partir de 3 repetições foram:

$\bar{y}_4 = \bar{y}_{(1)} = 26{,}3$

$\bar{y}_3 = \bar{y}_{(2)} = 86{,}9$

$\bar{y}_2 = \bar{y}_{(3)} = 87{,}8$

$\bar{y}_1 = \bar{y}_{(4)} = 95{,}5$

$\bar{y}_5 = \bar{y}_{(5)} = 108{,}2$

lembrando que $\bar{y}_{(i)}$ é a média do tratamento da posição ordenada i, com $i = 1,..., 5$.

i) SQ da partição (1) vs (2), (3), (4) e (5)

$$B_o = \frac{26{,}3^2}{1} + \frac{(86{,}9+87{,}8+95{,}5+108{,}2)^2}{4} - \frac{(26{,}3+86{,}9+87{,}8+95{,}5+108{,}2)^2}{5}$$

$B_o = 691{,}6900 + 35796{,}6400 - 32756{,}4180 = 3731{,}9120$;

alternância das parcelas evita-se uma possível vantagem de algum tratamento. A instalação do DIC no campo experimental exige uma certa homogeneidade das condições ambientais (como por exemplo quanto a fertilidade do solo, distribuição uniforme de água, etc.).

5.3.2) Vantagens

- Possui grande flexibilidade quanto ao número de tratamentos e repetições, sendo dependente, entretanto, da quantidade de material e área experimental disponíveis.

- Pode-se ter DIC não balanceado, ou seja, com números de repetições diferentes entre tratamentos, o que não leva a grandes alterações n
- a análise de variância; mas os testes de comparações múltiplas passam a ser aproximados e não mais exatos. O ideal é que os tratamentos sejam igualmente repetidos.

- Considerando o mesmo número de parcelas e tratamentos avaliados, é o delineamento que possibilita o maior grau de liberdade do erro.

5.3.3) Desvantagens

- Exige homogeneidade das condições experimentais. Se as condições não forem uniformes, como se esperava antes da instalação do experimento, toda variação (exceto à devida a tratamentos) irá para o erro, aumentando sua estimativa e reduzindo, portanto, a precisão do experimento.

5.3.4) Modelo estatístico do DIC

$$y_{ij} = \mu + t_i + e_{ij}$$

em que, y_{ij} representa a observação do i-ésimo tratamento na j-ésima repetição; μ representa uma constante geral associada a esta variável aleatória; t_i representa o efeito do i-ésimo tratamento; e e_{ij} representa o erro experimental associado a observação y_{ij}, suposto ter distribuição normal com média zero e variância comum.

5.2.4) Homogeneidade: os erros devem apresentar variâncias comuns (homogeneidade = homocedasticidade de variâncias).

Estas pressuposições visam facilitar a interpretação dos resultados e testar a significância nos testes de hipóteses. Na prática, o que pode ocorrer é a validade aproximada e não exata de alguma (s) destas pressuposições; neste caso, o pesquisador não perderia tanto com a aproximação visto que os testes aplicados na análise de variância são robustos quanto a isto. A homogeneidade de variância é que, na maioria das vezes, é necessária pois, caso não seja verificada, o teste F e de comparações múltiplas poderão ser alterados.

Quando alguma (s) das pressuposições da análise não se verifica(m), existem alternativas que podem ser usadas, entre elas a transformação de dados com a posterior análise de variância destes dados transformados; ou a utilização dos recursos da estatística não paramétrica.

Feitas as considerações iniciais necessárias para o entendimento dos próximos assuntos, iniciaremos agora os conceitos e exemplos dos delineamentos mais usuais.

5.3) Delineamento Inteiramente Casualizado (DIC)

5.3.1) Características

- Os tratamentos são distribuídos nas parcelas de forma inteiramente casual (aleatória).

- O DIC possui apenas os princípios da casualização e da repetição, não possuindo controle local e, portanto, as repetições não são organizadas em blocos.

- Normalmente é mais utilizado em experimentos de laboratório; experimentos em vasos ou bandejas em casa de vegetação, onde há possibilidade de controle das condições ambientais. Nos experimentos em casa de vegetação recomenda-se constantemente mudar as parcelas de posição para evitar diferenças ambientais devido a posição da parcela na casa de vegetação. Com esta

SQ da partição (1) e (2) vs (3), (4) e (5)

$$B_o = \frac{(26{,}3+86{,}9)^2}{2} + \frac{(87{,}8+95{,}5+108{,}2)^2}{3} - \frac{(26{,}3+86{,}9+87{,}8+95{,}5+108{,}2)^2}{5}$$

B_o = 6407,1200 + 28324,0833 - 32756,4180 = 1974,7853;

SQ da partição (1), (2) e (3) vs (4) e (5)

$$B_o = \frac{(26{,}3+86{,}9+87{,}8)^2}{3} + \frac{(95{,}5+108{,}2)^2}{2} - \frac{(26{,}3+86{,}9+87{,}8+95{,}5+108{,}2)^2}{5}$$

B_o = 13467,0000 + 20746,8450 - 32756,4180 = 1457,4270;

SQ da partição (1), (2), (3) e (4) vs (5)

$$B_o = \frac{(26{,}3+86{,}9+87{,}8+95{,}5)^2}{4} + \frac{108{,}2^2}{1} - \frac{(26{,}3+86{,}9+87{,}8+95{,}5+108{,}2)^2}{5}$$

B_o = 21978,0625 + 11707,2400 - 32756,4180 = 928,8845.

A partição (1) vs (2), (3), (4) e (5) foi a que maximizou a soma de quadrados entre grupos (B_o = 3731,9120).

ii) Considerando g=5, v=10 e $\bar{y} = \frac{26{,}3+86{,}9+87{,}8+95{,}5+108{,}2}{5} = 80{,}94$ tem-se:

$$\hat{\sigma}_o^2 = \frac{1}{5+10}\left[(26{,}3-80{,}94)^2 + \ldots + (108{,}2-80{,}94)^2 + 10.\frac{4{,}08}{3}\right]$$

$$\hat{\sigma}_o^2 = \frac{1}{15}[4023{,}2120 + 13{,}6000] = 269{,}1208$$

$$\lambda = \frac{\pi}{2(\pi-2)}.\frac{3731{,}9120}{269{,}1208} = 19{,}0806.$$

O valor de $\chi^2_{(0,05;\ 5/(\pi-2))} = \chi^2_{(0,05;\ 4,380)}$ é 10,089. Como λ > 10,089 rejeita-se H_o, ou seja dois grupos são formados ao nível de 5%, o grupo 1 com apenas o tratamento (linhagem) 4=(1) e o grupo 2 com os tratamentos 3=(2), 2=(3), 1=(4) e 5=(5).

Deve-se então repetir o algoritmo apenas para os subgrupos que contém mais de um tratamento, no caso apenas para o grupo 2.

i) SQ da partição (2) vs (3), (4) e (5)

$$B_o = \frac{86,9^2}{1} + \frac{(87,8+95,5+108,2)^2}{3} - \frac{(86,9+87,8+95,5+108,2)^2}{4}$$

B_o = 7551,6100 + 28324,0833 - 35796,6400 = 79,0533;

SQ da partição (2) e (3) vs (4) e (5)

$$B_o = \frac{(86,9+87,8)^2}{2} + \frac{(95,5+108,2)^2}{2} - \frac{(86,9+87,8+95,5+108,2)^2}{4}$$

B_o = 15260,0450 + 20746,8450 - 35796,6400 = 210,2500;

SQ da partição (2), (3) e (4) vs (5)

$$B_o = \frac{(86,9+87,8+95,5)^2}{3} + \frac{108,2^2}{1} - \frac{(86,9+87,8+95,5+108,2)^2}{4}$$

B_o = 24336,0133 + 11707,2400 - 35796,6400 = 246,6133.

A partição (2), (3) e (4) vs (5) foi a que maximizou a soma de quadrados entre grupos (B_o = 246,6133).

seria possível realizar testes de hipóteses. O uso de um número adequado de repetições, possibilita uma boa estimativa do erro experimental, melhorando as estimativas de interesse. No entanto, o número de repetições pode ser limitado, por exemplo, pelo número de tratamentos que serão comparados, pela disponibilidade de material e de área experimental, entre outros fatores.

5.1.2) Casualização: refere-se à distribuição aleatória dos tratamentos às parcelas de modo que todas as parcelas tenham a mesma chance de receber qualquer um dos tratamentos. Com isso, a casualização evita que determinado tratamento seja favorecido e garante que os erros sejam independentes (Mead & Curnow, 1983). Alguns programas computacionais elaboram planilhas de campo já com os tratamentos aleatorizados, como por exemplo o MSTAT, SISVAR e outros.

5.1.3) Controle local: a idéia básica do controle local é a partição do conjunto total de parcelas em subconjuntos (blocos) que sejam os mais homogêneos possíveis. Para Hinkelmann & Kempthorne (1994), o princípio do controle local é o reconhecimento de padrões supostamente associados às parcelas. Este princípio é utilizado para atenuar problemas de heterogeneidade ambiental (por exemplo de solo, de distribuição de água no caso de experimentos irrigados, etc).

5.2) Pressuposições básicas da análise de variância

Para realização de uma análise de variância deve-se aceitar algumas pressuposições básicas:

5.2.1) Aditividade: os efeitos de tratamentos e erro devem ser aditivos;

5.2.2) Independência: os erros devem ser independentes, ou seja, a probabilidade de que o erro de uma observação qualquer tenha um determinado valor não deve depender dos valores dos outros erros;

5.2.3) Normalidade: os erros devem ser normalmente distribuídos;

ii) Teste de Tukey:

$$\text{DMS} = q\sqrt{\left(\frac{1}{r_i}+\frac{1}{r_{i'}}\right)\frac{\text{QME}}{2}}$$

iii) Teste de Duncan:

$$\text{DMS} = z_n\sqrt{\left(\frac{1}{r_i}+\frac{1}{r_{i'}}\right)\frac{\text{QME}}{2}}$$

iv) Teste de SNK:

$$\text{SNK} = q_n\sqrt{\left(\frac{1}{r_i}+\frac{1}{r_{i'}}\right)\frac{\text{QME}}{2}}$$

com r_i e $r_{i'}$ sendo o número de repetições do tratamento i e i', respectivamente.

5) Análise de variância

A análise de variância (ANAVA) é um dos métodos para análise dos dados que visa decompor a variação total entre parcelas em fontes (causas) de variação devidas a efeitos principais dos fatores, efeitos de interações entre fatores, efeitos de aninhamento e resíduo (erro).

Para facilitar o entendimento, antes de partirmos para exemplos de análises de variância, é necessário fazer alguns comentários sobre os princípios básicos da experimentação e também sobre as pressuposições da análise de variância.

5.1) Princípios básicos da experimentação

Os delineamentos experimentais clássicos são baseados nos três conceitos a seguir, estabelecidos por Fisher (1935).

5.1.1) Repetição: refere-se ao número de parcelas que receberão um mesmo tratamento. Os tratamentos devem ser repetidos, possibilitando, assim, estimar o erro experimental sem o qual não

ii) Considerando g=4, v=10 e $\bar{y} = \frac{86,9+87,8+95,5+108,2}{4} = 94,60$ tem-se:

$$\hat{\sigma}_o^2 = \frac{1}{4+10}\left[(86,9-94,60)^2+\ldots+(108,2-94,60)^2+10.\frac{4,08}{3}\right]$$

$$\hat{\sigma}_o^2 = \frac{1}{14}[291,3000+13,6000] = 21,7786$$

$$\lambda = \frac{\pi}{2(\pi-2)}.\frac{246,6133}{21,7786} = 15,5810.$$

O valor de $\chi^2_{(0,05;\ 4/(\pi-2))} = \chi^2_{(0,05;\ 3,504)}$ é 10,253. Como λ > 10,253 rejeita-se H_o, ou seja dois grupos são formados ao nível de 5%, o grupo 1 com os tratamentos (linhagens) 3=(2), 2=(3) e 1=(4) e o grupo 2 com apenas o tratamento 5=(5).

Deve-se então repetir o algoritmo apenas para o grupo 1.

i) SQ da partição (2) vs (3) e (4)

$$B_o = \frac{86,9^2}{1}+\frac{(87,8+95,5)^2}{2}-\frac{(86,9+87,8+95,5)^2}{3}$$

B_o = 7551,6100 + 16799,4500 – 24336,0133 = 15,0417;

SQ da partição (2) e (3) vs (4)

$$B_o = \frac{(86,9+87,8)^2}{2}+\frac{95,5^2}{1}-\frac{(86,9+87,8+95,5)^2}{3}$$

B_o = 15260,0450 + 9120,2500 – 24336,0133 = 44,2817;

A partição (2) e (3) vs (4) foi a que maximizou a soma de quadrados entre grupos (B_o = 44,2817).

ii) Considerando g=3, v=10 e $\bar{y} = \dfrac{86,9 + 87,8 + 95,5}{3} = 90,07$ tem-se:

$$\hat{\sigma}_o^2 = \frac{1}{3+10}\left[(86,9-90,07)^2 + (87,8-90,07)^2 + (95,5-90,07)^2 + 10.\frac{4,08}{3}\right]$$

$$\hat{\sigma}_o^2 = \frac{1}{13}[44,6867 + 13,6000] = 4,4836$$

$$\lambda = \frac{\pi}{2(\pi-2)}.\frac{44,2817}{4,4836} = 13,5896.$$

O valor de $\chi^2_{(0,05;\ 3/(\pi-2))} = \chi^2_{(0,05;\ 2,628)}$ é 7,136. Como λ > 7,136 rejeita-se H_o, ou seja dois grupos são formados ao nível de 5%, o grupo 1 com os tratamentos (linhagens) 3=(2), 2=(3) e o grupo 2 com apenas o tratamento e 1=(4).

Deve-se então repetir novamente o algoritmo para o grupo 1.

i) SQ da partição (2) vs (3)

$$B_o = \frac{86,9^2}{1} + \frac{87,8^2}{1} - \frac{(86,9+87,8)^2}{2}$$

B_o = 7551,6100 + 7708,8400 – 15260,0450 = 0,4050;

Neste caso, a partição (2) vs (3) por ser única foi a que maximizou a soma de quadrados entre grupos (B_o = 0,4050).

ii) Considerando g=2, v=10 e $\bar{y} = \dfrac{86,9 + 87,8}{2} = 87,35$ tem-se:

$$\hat{\sigma}_o^2 = \frac{1}{2+10}\left[(86,9-87,35)^2 + (87,8-87,35)^2 + 10.\frac{4,08}{3}\right]$$

$$\hat{\sigma}_o^2 = \frac{1}{12}[0,4050 + 13,6000] = 1,1671$$

$$\lambda = \frac{\pi}{2(\pi-2)}.\frac{0,4050}{1,1671} = 0,4775.$$

O valor de $\chi^2_{(0,05;\ 2/(\pi-2))} = \chi^2_{(0,05;\ 1,752)}$ é 5,458. Como λ < 5,458 aceita-se H_o, ou seja, os dois grupos são idênticos ao nível de 5%, formando um único grupo com os tratamentos (linhagens) 3=(2) e 2=(3), finalizando assim o algoritmo.

Colocando letras diferentes para médias distintas e letras iguais para médias semelhantes por meio do teste Scott e Knott tem-se então:

$\bar{y}_4 = \bar{y}_{(1)}$ = 26,3 d

$\bar{y}_3 = \bar{y}_{(2)}$ = 86,9 c

$\bar{y}_2 = \bar{y}_{(3)}$ = 87,8 c

$\bar{y}_1 = \bar{y}_{(4)}$ = 95,5 b

$\bar{y}_5 = \bar{y}_{(5)}$ = 108,2 a

Observações: Nestes exemplos os resultados de todos os testes realizados foram iguais mas poderiam ter diferenciado entre um ou outro teste. Quando o número de repetições é diferente entre os tratamentos as DMS's podem ser calculadas por:

i) Teste t:

$$DMS = t_t\sqrt{\left(\frac{1}{r_i} + \frac{1}{r_{i'}}\right)QME}$$