Ñ TEMOS

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
CADERNOS DE MATEMÁTICA E ESTATÍSTICA

PLANEJAMENTO E ANÁLISE DE EXPERIMENTOS
PARTE 1

JOÃO RIBOLDI

SÉRIE B, Nº 23
PORTO ALEGRE, FEVEREIRO DE 1994

PREFACIO

As presentes Notas destinam-se ao apoio didático das disciplinas de Planejamento de Experimentos do Curso de Bacharelado em Estatística. Surgiram da experiência acumulada ao longo dos anos e tem por objetivo servir como guia aos conteúdos abordados e não um limitante dos assuntos, não prescindindo, evidentemente, da consulta de bibliografia especializada para complementação .

Apesar de serem de objetivo específico, podem também servir como texto de apoio didático a outras disciplinas a nível de graduação e pós-graduação.

Agradecemos a todos que colaboraram na organização destas notas e em especial aos bolsistas Stela, Flávio, André e Silvana pelo trabalho de digitação.

Porto Alegre, 28 de Janeiro de 1994

Prof. João Riboldi

ÍNDICE

pag

1. ASPECTOS GERAIS DO PLANEJAMENTO DE EXPERIMENTOS 01
1.1 - Estatística Aplicada à Experimentação 01
1.2 - Método científico 01
1.3 - Experimentos 01
1.4 - Características de um bom Experimento 02
1.5 - Erro Experimental 03
1.6 - Etapas da Organização de um Experimento 06

2. DELINEAMENTO COMPLETAMENTE CASUALIZADOS OU
DELINEAMENTO INTEIRAMENTE CASUALIZADO 20
2.1 - Caracterização 20
2.2 - Análise de Variância 21
2.3 - Medidas de Precisão de Experimentos 23
2.4 - Exemplo 24
2.5 - Experimentos Inteiramente Casualizdos com Diferente
Número de Repetições por Tratamento 26
2.6 - Técnicas de Complementação na Análise de Variância...26
2.7 - Contrastes ou Comparação de Médias 27
2.8 - Testes de Comparações Múltiplas de Médias 28
2.9 - Alternativas para Comparações Múltiplas de Médias de
Tratamentos 36
2.10 - Teste Bayesiano ou Teste de Waller-Duncan ou Teste
da Razão Bayesiana k 42
2.11 - Método de Análise de Agrupamento ("Cluster
Analysis") ou Método de SCOTT e KNOTT para
Agrupamento de Médias 45
2.12 - Contrastes Ortogonais 52
2.13 - Análise de Regressão 66

2.14 - Considerações Teóricas sobre a Análise de Variância de Experimentos com um Fator Fixo em Delineamento Completamente Casualizàdo 78
2.15 - Análise de Variância de Experimentos com 1 Fator Aleatório em Delineamento Completamente Casualizado . 99
2.16 - Análise de Variância para Experimentos Inteiramente Casualizados Através do Modelo de Regressão 109
2.17 - Verificação da Adequabilidade do Modelo de Análise de Variância para o Delineamento Completamente Casualizado com um Fator 117
2.18 - Transformações de Dados 126
2.19 - Busca da Transformação 134

1. ASPECTOS GERAIS DO PLANEJAMENTO DE EXPERIMENTOS

1.1 - ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO

A estatística quando orientada para a área de investigação, dentro do chamado método científico, é definida como a ciência que se ocupa da experimentação no que diz respeito a sua:

1. Planificação (planejamento de experimentos)
2. Execução (instalação, condução e coleta de informações de experimentos)
3. Análise dos seus resultados

1.2 - MÉTODO CIENTÍFICO

Levantamento de um problema através da observação dos fatos ou resultado experimental

→ Proposta de uma explicação conjectural sobre os mesmos através da formulação de hipóteses (1)

↓

Realização de experimentos específicos para obtenção das observações (2)

↓

Verificação das hipóteses formuladas através da análise dos resultados, permitindo a elaboração de conclusões, previsões, generalizações (3)

Experimentação

← Desenvolvimento da teoria (4)

1.3 - EXPERIMENTOS

Pesquisa planejada para obter novos fatos, para confirmar ou não resultados obtidos, tendo por objetivo tomar decisões.

1. EXPERIMENTOS PRELIMINARES

Busca de informações para trabalhos futuros.

2. EXPERIMENTOS CRÍTICOS OU COMPARATIVOS

Compara-se os tratamentos utilizando-se observações suficientes para detectar diferenças existentes, com uma certa segurança.

3. EXPERIMENTOS DEMONSTRATIVOS

Divulgação de resultados. Comparação de novos tratamentos com padrão. Trabalhos de extensão.

1.4 - CARACTERÍSTICAS DE UM BOM EXPERIMENTO

1. Ausência de erro sistemático (vício) [CASUALIZAÇÃO].

2. Precisão: capacidade de detectar como significativa a diferença entre médias de tratamentos.

* Formas de expressar precisão de experimentos.

$$(i)\quad CV = \frac{\sqrt{QME}}{\bar{y}} \cdot 100$$

onde: QME = quadrado médio do erro experimental ou resíduo

$\bar{y}$ = média geral do experimento

(ii) Erro padrão da diferença entre duas médias de tratamentos

$$s_d = \sqrt{\frac{2QME}{r}}$$

(iii) Erro padrão da média de um tratamento

$$s_{\bar{y}} = \sqrt{\frac{QME}{r}}$$

* Erro experimental:

Quanto menor → Maior a precisão

- Causas do erro experimental:

(1) Variabilidade intrínseca das unidades experimentais (UE)

(2) Falhas de técnica experimental

- Controle do erro experimental:

(*i*) Uso de UE homogêneas (1)

(*ii*) Uso de técnica experimental cuidadosa (2)

(*iii*) Uso de delineamento experimental eficiente (1 e 2)

(*iv*) Uso de observações auxiliares ou concomitantes, através da análise de covariância (1 e 2)

3. Generalidade dos resultados.
4. Simplicidade.

1.5 - <u>ERRO EXPERIMENTAL</u>:

Variação das UE submetidas ao mesmo tratamento.

*<u>ESTIMATIVA</u>

(*i*) Repetindo-se todos os tratamentos

(*ii*) Repetindo-se parte dos tratamentos

(*iii*) Repetindo-se um tratamento

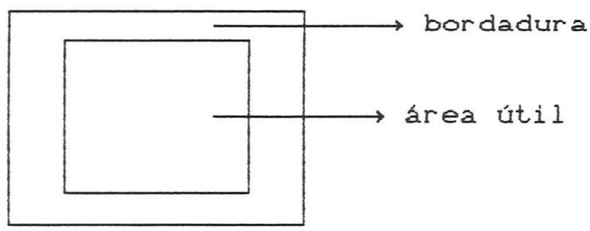
(*iv*) Usando-se estimativas anteriores

(*v*) Usando-se interações de ordem elevada (raramente atingem significância estatística) [REPETIÇÃO ÚNICA]

*<u>CONTROLE</u>

1. <u>TÉCNICA EXPERIMENTAL CUIDADOSA</u>:

Técnica que possibilite uniformidade de execução, tal como na aplicação dos tratamentos (evitando contaminação, erros de dosagem,...), evite competição inter-parcelas, utilizando-se bordadura nas parcelas.



E com isso eliminando efeito de bordo; uniformidade de condução; uniformidade de coleta de informações (de modo adequado e de tal forma a não favorecer nenhum tratamento); evite a influência de fatores estranhos.

2. OBSERVAÇÕES AUXILIARES OU CONCOMITANTES:

Em muitos experimentos a precisão pode ser aumentada pelo uso de observações auxiliares (concomitantes) através da análise de covariância. A análise de covariância é usada quando a variação entre as UE é em parte devida a algumas características não suficientemente controladas (condição própria das UE, ou devido a influência de fatores estranhos durante a execução do experimento). O uso de observações auxiliares elimina as competições intra-parcelas (efeito de falhas) decorrente de causas extrínsecas (não relacionadas aos tratamentos, tal como germinação desuniforme, ataque de moléstias, insetos e outros animais).

3. DELINEAMENTO EXPERIMENTAL EFICIENTE

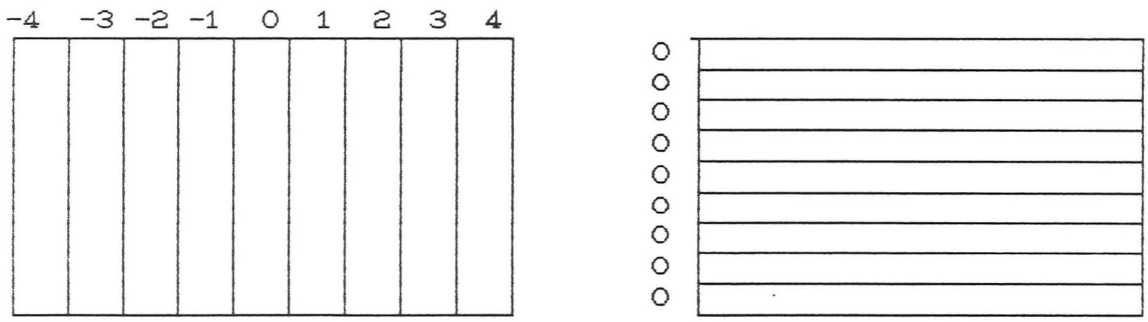
4. UNIDADES EXPERIMENTAIS HOMOGÊNEAS:

Forma e colocação das UE no campo, tamanho.

(a) FORMA E COLOCAÇÃO DAS PARCELAS DE CAMPO

(i) COLOCAÇÃO:

- Se existe um gradiente de fertilidade conhecido (GF) a dimensão mais longa deve estar seguindo o gradiente.



A
QM = 60/8

B
QM = 0

-3	0	3
-3	0	3
-3	0	3

C

GRADIENTE DE FERTILIDADE

$- \longrightarrow +$

QM = 54/8

IDEAL B

Se não houver condições de usar B usar C e não A.

- Todas as parcelas do bloco devem ter uma frente comum: B preferível.

(ii) FORMA:

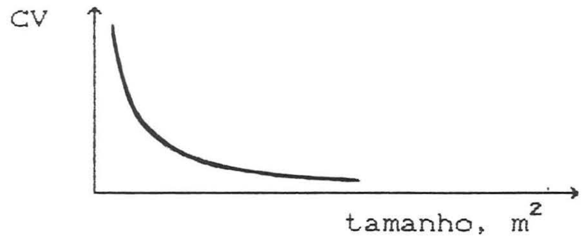
As parcelas devem ser de forma retangular estreitas e longas, uma vez que com esta forma as parcelas tendem a participar de todas as manchas de fertilidade existente no terreno, bem como, dessa maneira, o bloco tenderá a ser quadrado que é a forma mais compacta e com maior grau de homogeneidade.

5. TAMANHO DAS PARCELAS:

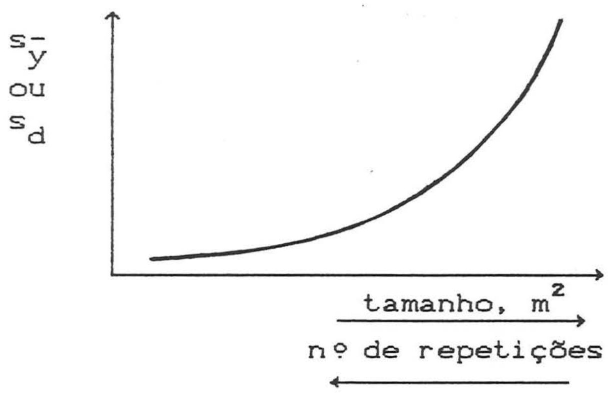
(i) ASPECTOS A CONSIDERAR:

- custo
- disponibilidade de área experimental, material (semente, fertilizantes,...), de tempo, mão-de-obra, maquinaria
- natureza dos fatores
- número de tratamentos

(ii) O QUE SE VERIFICA



Área Fixa: Quanto maior tamanho → Menor número de repetições



- Parcelas grandes variam menos do que parcelas pequenas.
- Acréscimo no tamanho da parcela → acréscimo no número de repetições (limitações de material experimental).
- Adequadas repetições de parcelas pequenas é mais fácil de se obter do que adequadas repetições de parcelas grandes.
- É preferível usar parcelas pequenas e maior número de repetições, ou seja, usar o menor tamanho possível, compatível com o material a ser experimentado, e compensar a perda de precisão com o aumento no número de repetições.

1.6 - ETAPAS DA ORGANIZAÇÃO DE UM EXPERIMENTO

1. ENUNCIADO DO PROBLEMA E FORMULAÇÃO DE HIPÓTESES.

2. ESCOLHA DOS FATORES QUE DEVEM SER INCLUIDOS NO EXPERIMENTO E DOS SEUS RESPECTIVOS NÍVEIS (= escolha dos tratamentos).

3. ESCOLHA DA UNIDADE EXPERIMENTAL E DA UNIDADE DE OBSERVAÇÃO.

4. ESCOLHA DAS VARIÁVEIS A SEREM MEDIDAS NA UNIDADE DE OBSERVAÇÃO.

5. DETERMINAÇÃO DAS REGRAS PARA ATRIBUIÇÃO DOS

TRATAMENTOS AS UNIDADES EXPERIMENTAIS (= escolha do delineamento experimental).

6. DETERMINAÇÃO DO NÚMERO DE REPETIÇÕES.

7. ESCOLHA DO PROCEDIMENTO DE ANALISE ESTATÍSTICA DOS RESULTADOS.

8. RELATÓRIO FINAL: CONCLUSÕES, PRECISÃO DAS ESTIMATIVAS, INTERPRETAÇÃO DOS RESULTADOS REFERINDO-SE A TRABALHOS SIMILARES, AVALIAÇÃO DA PESQUISA COM SUGESTÕES PARA PROSSEGUIR.

1.6.1. ENUNCIADO DO PROBLEMA E FORMULAÇÃO DE HIPÓTESES:

Uma pesquisa científica se inicia sempre com a formulação de hipóteses. As hipóteses são primeiramente formuladas em termos científicos dentro da área de estudo (hipóteses científicas) e em seguida devem ser expressas em termos estatísticos (hipóteses estatísticas).

Deve haver correspondência perfeita entre as hipóteses científica e estatística para evitar ambiguidade. Portanto, no enunciado do problema, a hipótese científica deve ser formulada de maneira precisa e objetiva.

No método científico, a formulação de um problema é fundamental. O cientista, o agente principal deste processo, é a pessoa que consegue visualizar problemas e suas soluções com uma habilidade maior que a da maioria das pessoas. A formulação de um problema é muitas vezes mais importante que sua solução a qual pode ser apenas uma questão de habilidade experimental.Propor problemas novos e encarar os velhos sob um novo ângulo, requer imaginação criadora e é o que promove o progresso da ciência.

Por hipótese,entende-se uma teoria que resume o conhecimento do pesquisador em relação ao problema em questão em determinado momento.

A determinação dos aspectos relevantes do fenômeno em estudo, e a formulação de hipóteses está sujeita a vários fatores:

- uma profunda familiarização com o problema sob investigação;
- uma sólida base de conhecimentos prévios (que permitirá associar e comparar o problema com situações análogas);
- intuição e a genialidade do pesquisador.

A formulação de hipóteses poderá conduzir a previsões sobre os resultados a serem obtidos, sendo uma etapa desenvolvida sob inteira responsabilidade do pesquisador.

<u>EXEMPLO</u>: Influência da aplicação de nitrogênio (t níveis) sobre o rendimento de cana-de-açúcar.

<u>HIPÓTESE 1</u>:

A cana-de-açúcar responde a adubação nitrogenada.

$\downarrow$ *definição operacional*

Ha: Pelo menos 2 médias de tratamentos diferem ($\tau_i \neq 0$)

Teste $\rightarrow$ Ho: $\mu_1 = \mu_2 = \ldots = \mu_t \Leftrightarrow (\tau_i = 0)$

<u>HIPÓTESE 2</u>:

O rendimento de cana-de-açúcar relaciona-se funcionalmente com doses (níveis) de N aplicadas ao solo.

$\downarrow$ definição operacional

y $\longrightarrow$ rendimento de cana-de-açúcar
x $\longrightarrow$ doses (níveis) de N

Ha: A relação y = f(x) explica as variações no rendimento de cana-de-açúcar $\Leftrightarrow$ (Ha: $R^2 \neq 0$)

Teste $\rightarrow$ Ho: A relação y = f(x) não explica as variações no rendimento de cana-de-açúcar $\Leftrightarrow$ (Ho: $R^2 = 0$)

Coeficiente de determinação:

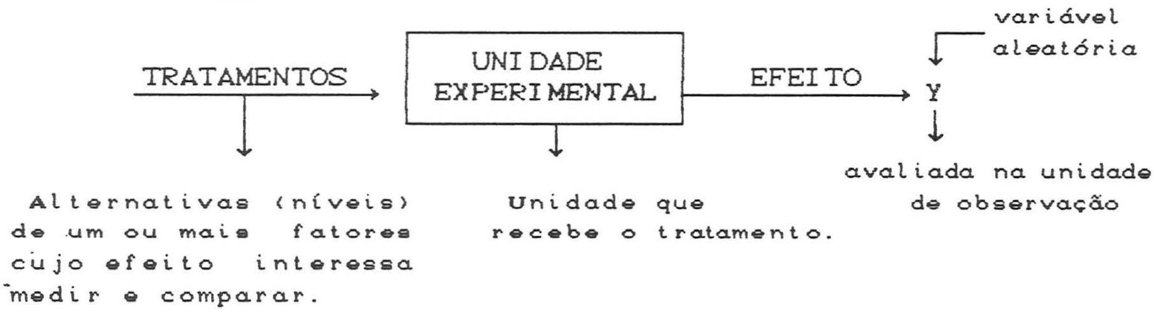
$$R^2 = \left.\frac{\text{SQREGRESSÃO}}{\text{SQTOTAL}}\right\}$$ Proporção da Variabilidade no rendimento de cana-de-açúcar explicada pela relação y = f(x).

1.6.2. ESCOLHA DOS FATORES QUE DEVEM SER INCLUIDOS NO EXPERIMENTO E DOS SEUS RESPECTIVOS NÍVEIS (=escolha dos tratamentos):

- Quais fatores e níveis a utilizar é função do pesquisador essencialmente.

- Como organizar os fatores e/ou níveis (escolha do delineamento de tratamentos) importa sob ponto de vista estatístico.

<u>EXPERIMENTO</u>:



Tratamentos:

- Experimentos unifatoriais: tratamentos são níves de um fator. Ex.: temperatura → T_1 , T_2 , T_3 , T_4 .
- Experimentos fatoriais: tratamentos são combinações de níveis de diferentes fatores. Ex.: temperatura → T_1 , T_2 , T_3 ; e pressão → P_1 e P_2 . Tratamentos: T_1P_1 T_1P_2

T_2P_1 T_2P_2

T_3P_1 T_3P_2

Fatores: Qualitativos (tipo de material, marcas) e Quantitativos (temperatura, pressão, ...)

Organização dos fatores e/ou níveis:

- considerar custo dos experimentos
- reduzir ao máximo o nº de tratamentos
- usar delineamento de tratamentos adequado visando a redução do nº de tratamentos e ainda assim estimar satisfatoriamente o efeito dos tratamentos.

Considerações sobre a escolha dos tratamentos:

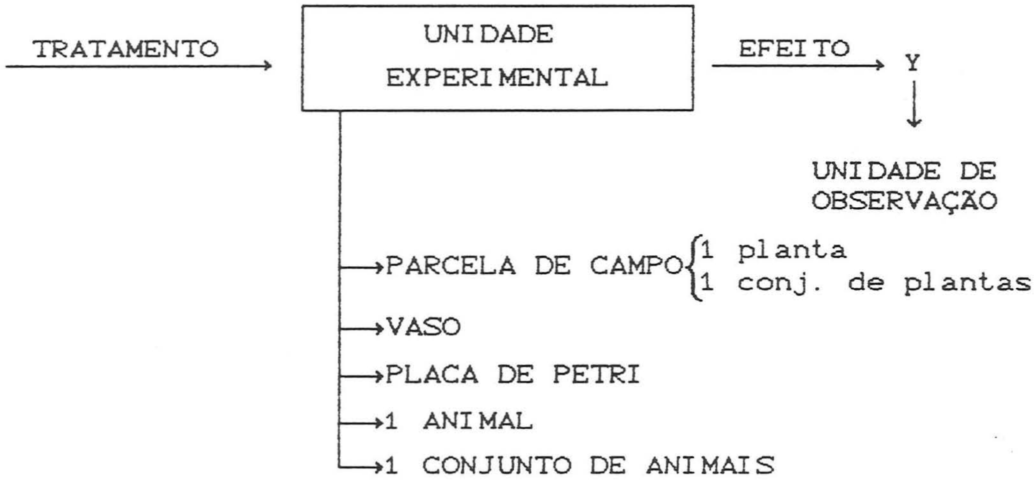
- quais fatores e níveis a utilizar é função do pesquisador essencialmente
- como organizar os fatores e/ou níveis (escolha do delineamento

de tratamentos) importa sob o ponto de vista estatístico

- a escolha dos tratamentos deve ser adequada, que permita a verificação das hipóteses formuladas.

1.6.3. ESCOLHA DA UNIDADE EXPERIMENTAL E DA UNIDADE DE OBSERVAÇÃO:

<u>EXPERIMENTO</u>:



<u>UNIDADE EXPERIMENTAL</u>

A escolha da unidade experimental deve ser no sentido de minimizar o erro experimental, ou seja, as unidades experimentais devem ser o mais homogêneas possíveis, controlando a sua heterogeneidade.

<u>PARCELAS DE CAMPO</u>

Principal fator de variabilidade das UE (parcelas) é a heterogeneidade do solo que é devida:

1) Diferença na constituição física e química do solo
2) Diferença de nivelação
3) Diferença de drenagem
4) Diferença de subsolo
5) Diferença de preparo do solo
6) Diferença de distribuição dos adubos

UNIDADE DE OBSERVAÇÃO

Dependendo da variável que está sendo avaliada pode-se ter como unidade de observação a unidade experimental como um todo (por exemplo,avaliação de rendimento) ou a unidade de observação é constituida por uma "amostra" de subunidades ou fração da unidade experimental (por exemplo, determinações tecnológicas).

1.6.4. ESCOLHA DAS VARIAVEIS A SEREM MEDIDAS NA UNIDADE DE OBSERVAÇÃO:

As medidas realizadas nas unidades experimentais após terem sido submetidas aos tratamentos constituem os valores da variável dependente. A variável dependente, em geral, é pré-determinada pelo pesquisador, isto é, ele estabelece "como critério" a variável a ser medida para verificação do efeito de tratamento.

O que constitui problema, às vezes, é a maneira como a variável é medida, pois disto depende a precisão das observações e a distribuição de probabilidade da variável a qual é essencial para a escolha do método de análise estatística.

Se os valores de uma variável são obtidas diretamente por meio de um instrumento de medida (régua, paquímetro, termômetro,...) a precisão das observações vai aumentar quando se utiliza, se possível, como observação a média de 3 ou mais medidas da mesma unidade experimental.

VARIAVEIS DE UM EXPERIMENTO

- Variáveis dependentes: medidas nas UE.
- Variáveis independentes: conjunto de fatores.
- Qualquer outra variável que possa influir nos resultados (valores) para a variável dependente deve ser mantida constante.

TIPO DE VARIAVEL (OBSERVAÇÕES) MEDIDAS NUM EXPERIMENTO

VARIÁVEIS DEPENDENTES

- Observações primárias: características que diretamente medem efeitos de tratamento.
  Ex.: rendimento; % PB; % açúcar.
- Observações primárias substitutivas: características que substituem as primárias pois são de mais fácil obtenção. Alta correlação com as primárias.
  Ex.: Digestibilidade "in vitro" quando objetivo digestibilidade "in vivo".

OUTROS FATORES

- Observações auxiliares ou suplementares ou concomitantes: medem efeito de outros fatores (Análise de covariância)

1.6.5. DETERMINAÇÃO DAS REGRAS PARA ATRIBUIÇÃO DOS TRATAMENTOS AS UNIDADES EXPERIMENTAIS (=escolha do delineamento experimental):

PRINCÍPIOS BASICOS DE EXPERIMENTAÇÃO:

REPETIÇÃO → -Uso de mais de 1 UE/tratamento
-Permite verificar a quem é devida a diferença observada

CASUALIZAÇÃO
-Distribuição casualda variabilidade
-Evita a introdução de vício no experimento

-Diminuição do erro (aumento de precisão)

→-Validade da estimativa do erro experimental

→-Validade da estimativa do efeito de tratamento

BLOQUEAMENTO
-Controle "a priori" da variabilidade das UE, agrupando as UE em blocos homogêneos

DELINEAMENTO EXPERIMENTAL:

Forma de atribuição dos tratamentos às unidades experimentais.

	DCC	DBC	DQL
REPETIÇÃO ↑	-nenhuma restrição	-1 restrição	-2 restrições
CASUALIZAÇÃO	na casualização VE	-1 bloqueamento	-2 bloqueamen
BLOQUEAMENTO ↓	-nenhum bloqueamento	completo	tos completos

DCC = Delineamento completamente casualizado ou inteiramente casualizado.

DBC = Delineamento blocos casualizados

DQL = Delineamento quadrado latino

1 BLOQUEAMENTO INCOMPLETO	PARCELA SUBDIVIDIDA CONFUNDIMENTO	FATORIAIS
	LATTICES BLOCOS INCOMPLETOS BALANCEADOS	1 FATOR

Eficiência Relativa

	Variância residual
Delineamento 1 ⟶	σ_1^2 (QME_1)
Delineamento 2 ⟶	σ_2^2 (QME_2)

A eficiência relativa do delineamento 1 em relação ao delineamento 2 é dada por:

$$ER\% = \frac{1}{\frac{\sigma_1^2}{\sigma_2^2}} \times 100 = \frac{\sigma_2^2}{\sigma_1^2} \times 100$$

A eficiência relativa é útil na escolha do delineamento experimental a ser utilizado.

1.6.6. DETERMINAÇÃO DO NÚMERO DE REPETIÇÕES:

REGRAS PRATICAS

- Nº de repetições por tratamento de tal forma que tenhamos no mínimo 20 parcelas.

GL erro ≥ 10

$\downarrow$ DEPENDE: do delineamento experimental
do nº de tratamentos

GL suficientes para uma estimativa representativa do erro experimental. Experimentos com poucos tratamentos necessitam de maior nº de repetições para GL suficientes para o erro.

NÚMERO DE REPETIÇÕES A USAR NUM EXPERIMENTO

Número adequado de repetições é importante no planejamento de um experimento:

- poucas repetições → pode-se não descobrir diferenças importantes
- muitas repetições → desperdício de tempo e material
- deve-se ter número suficiente de repetições para detectar como significativa a diferença no efeito de 2 tratamantos, se ela existir.

Para se determinar o número de repetições necessita-se:

- Estimativa de variabilidade: s^2 ou CV
- Tamanho da diferença entre médias a ser detectada como significativa: δ, expressa com % da média geral
- Nível de significância: α
- Segurança com que se deseja detectar a diferença: poder do teste, $P=1-\beta$
- Teste unilateral ou bilateral.

2 Grupos Independentes

População 1	População 2	Suposição
μ_1 , σ_1^2	μ_2 , σ_2^2	$\sigma_1^2 = \sigma_2^2 = \sigma^2$
$\downarrow$	$\downarrow$	
Amostra 1	Amostra 2	
n_1	n_2	
$\bar{y}_1$, s_1^2 , SQ_1	$\bar{y}_2$, s_2^2 , SQ_2	

$$Ho: \mu_1 = \mu_2 \Leftrightarrow Ho: \mu_1 - \mu_2 = 0$$
$$Ha: \mu_1 \neq \mu_2 \Leftrightarrow Ho: \mu_1 - \mu_2 \neq 0 \quad \text{[Bilateral]}$$
$$Ha: \mu_1 > \mu_2 \Leftrightarrow Ho: \mu_1 - \mu_2 > 0 \quad \text{[Unilateral]}$$

$$\frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{s_d} \cap t(n_1 + n_2 - 2)$$

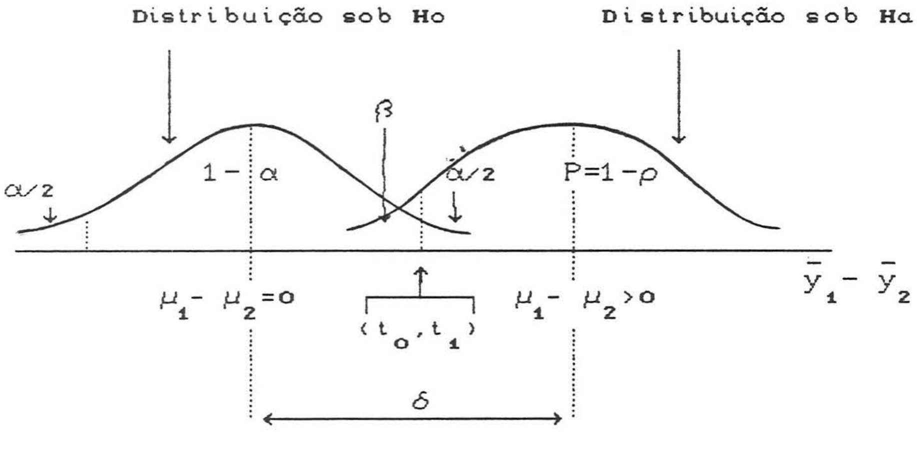
s_d ← Erro padrão da diferença entre 2 médias

$$s_d = \sqrt{s^2 \left[\frac{1}{n_1} + \frac{1}{n_2}\right]}$$

s^2 ← Variância ponderada

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{SQ_1 + SQ_2}{n_1 + n_2 - 2}$$

Se $n_1 = n_2 = n \Rightarrow s_d = \sqrt{\dfrac{2s^2}{n}}$ e $s^2 = \dfrac{s_1^2 + s_2^2}{2}$



Decisões

(i) Ho será rejeitada quando t calculado > t_0

[valor tabelado de t ignorando o sinal ⇒ t bilateral]

$$\text{Sob Ho} \quad t = \frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2 = 0)}{s_d} = \frac{\bar{y}_1 - \bar{y}_2}{s_d} = t_0$$

(ii) Ha será aceita quando t calculado > t_1

[Valor tabelado de t considerando o sinal $\Rightarrow$ t unilateral (do lado esquerdo da curva)]

$$\text{Sob Ha} \quad t = \frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{s_d} = t_1$$

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{s_d} \Rightarrow t_0\, s_d = \bar{y}_1 - \bar{y}_3 \qquad (1)$$

$$t_1 = \frac{(\bar{y}_1 - \bar{y}_2) - \overbrace{(\mu_1 - \mu_2)}^{\delta}}{s_d} \Rightarrow t_1\, s_d = \bar{y}_1 - \bar{y}_2 - \delta$$

$$\Rightarrow t_1\, s_d + \delta = \bar{y}_1 - \bar{y}_2 \qquad (2)$$

$$\therefore\ t_0\, s_d = t_1\, s_d + \delta$$

$$\delta = t_0\, s_d - t_1\, s_d \Rightarrow \delta = s_d\,(t_0 - t_1)$$

$$\text{Como} \quad s_d = \sqrt{\frac{2s^2}{n}} \quad \text{temos}$$

$$\delta = \sqrt{\frac{2s^2}{n}}\,(t_0 - t_1)$$

$$\delta^2 = \frac{2s^2(t_0 - t_1)^2}{n}$$

$$\therefore\ n = \frac{2s^2\,(t_0 - t_1)^2}{\delta^2}$$

Expressando a variabilidade em termos de CV e considerando que como t_1 será um valor negativo (lado esquerdo da distribuição) então $-t_1$ será um valor positivo, assim pode-se considerar o simétrico positivo, e então:

$$\boxed{n = \frac{2CV^2\,(t_0 + t_1)^2}{\delta^2}}$$

Exemplo: Duas rações devem ser comparadas com leitões, mantidos em baias individuais, sendo o aumento de peso o atributo a ser medido. Qual o nº de leitões a usar em cada grupo, isto é, o nº de repetições, se deseja-se detectar uma diferença no efeito das rações de pelo menos 10%, com uma segurança de 80%, e o coeficiente de variação previsto é de 8%?

Teste Bilateral

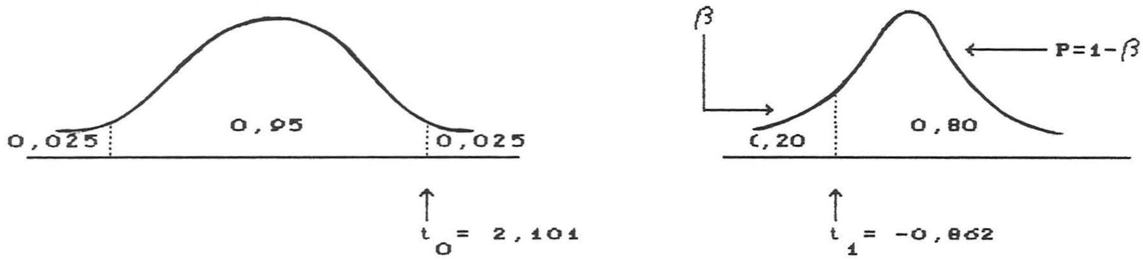
$\alpha = 0{,}05 \qquad CV = 8\% \qquad \delta = 10\% \qquad P = 0{,}80$

Processo iterativo

1º passo: $n = 10 \qquad GL = 2(n-1) = 2(10-1) = 2(9) = 18$

$t_0 = t_{.05}(18) = 2{,}101$ [$\alpha = 0{,}05$ GL = 18 ; ignorando o sinal]

$t_1 = t_{.20}(18) = 0{,}862$ [$1-P = 0.20$ GL=18 ; considerando o sinal]



$$n' = \frac{2(8)^2\,(2{,}101 + 0{,}862)^2}{(10)^2} = 11{,}2$$

2º passo: $n = 12 \qquad GL = 2(n-1) = 2(12-1) = 2(11) = 22$

$t_0 = t_{.05}(22) = 2{,}074$

$t_1 = t_{.20}(22) = 0{,}858$

$$n'' = \frac{2(8)^2\,(2{,}074 + 0{,}858)^2}{(10)^2} = 11$$

$\therefore \quad \boxed{n = 11}$

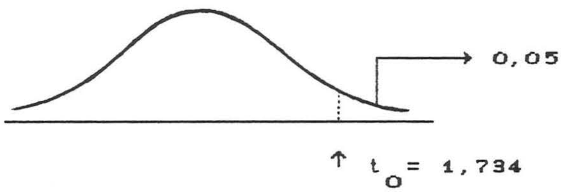
<u>Teste unilateral</u> $\alpha = 0{,}05$

<u>1º passo:</u> $n = 10 \rightarrow GL = 18$

$t_o = t.05(18) = 1{,}734$ [α = 0,05 ; GL = 18 ; considerando o sinal]

$t_1 = 0{,}862$

$$n' = \frac{2(8^2)(1{,}734 + 0{,}862)^2}{(10)^2} = 8{,}6$$



<u>2º passo:</u> $n = 9 \rightarrow GL = 2(n-1) = 2(9-1) = 2(8) = 16$

$t_o = t.05(16) = 1{,}746$

$t_1 = t.20(16) = 0{,}865$

$$n'' = \frac{2(8^2)(1{,}746 + 0{,}865)^2}{(10)^2} = 8{,}7$$

$\therefore \boxed{n = 9}$

<u>TABELA PARA DETERMINAÇÃO DO NÚMERO DE REPETIÇÕES</u>

UNILATERAL

δ	CV 8
10	9 → α = 5%; P = 80%
	12 → α = 5%; P = 90%
	22 → α = 1%; P = 95%

BILATERAL

δ	CV 8
10	11
	15
	24

1.6.7. ESCOLHA DO PROCEDIMENTO DE ANÁLISE ESTATÍSTICA DOS RESULTADOS:

ANÁLISE CLÁSSICA:

ANÁLISE DE VARIANCIA

VARIAÇÃO TOTAL
→1. Variação relacionada com os tratamentos
→2. Variação relacionada com causas controladas pelo delineamento experimental
→3. Variação relacionada com o erro experimental

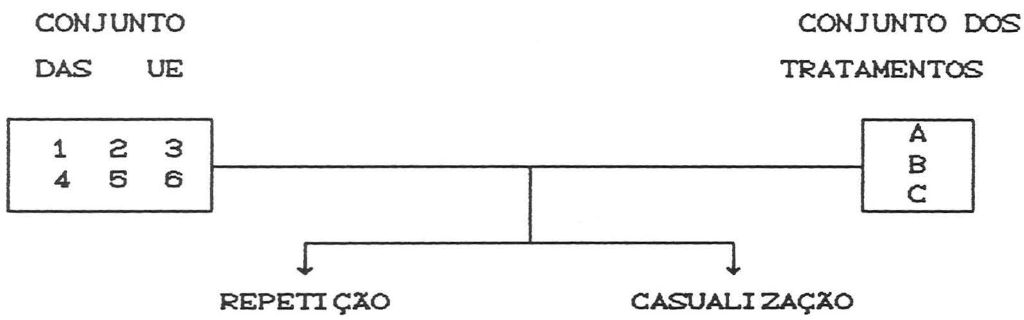
EXPERIMENTOS INTEIRAMENTE CASUALIZADOS: 1 e 3
EXPERIMENTOS EM BLOCOS CASUALIZADOS: 1, 2 e 3

*TÉCNICAS DE COMPLEMENTAÇÃO DA ANÁLISE DE VARIANCIA

1. Análise de regressão: Fatores quantitativos.
2. Contrastes ortogonais: Fatores quantitativos ou Fatores qualitativos que permitem estruturação.
3. Comparações múltiplas de médias: Fatores quantitativos ou Fatores qualitativos que permitem ou não permitem estruturação.

2. DELINEAMENTO COMPLETAMENTE CASUALIZADO OU DELINEAMENTO INTEIRAMENTE CASUALIZADO

2.1 CARACTERIZAÇÃO:



Atribuição por sorteio dos tratamentos a serem aplicados para as diferentes UE.

TRATAMENTOS

A	B	C
3	1	2
4	6	5

* Escolha casual das UE (nenhuma restrição quanto a casualização)

*Extensão de grupos independentes

USO:

-Uniformidade das UE, pois quanto menos uniformes → mais variável a informação → menos precisos os resultados.

- Execução uniforme sobre todas UE (instalação, condução e coleta de informações).

CARACTERÍSTICA:

-GL erro experimental (resíduo) maior possível → alta sensibilidade dos testes.

2.2 ANÁLISE DE VARIÂNCIA:

TRATAMENTOS → UNIDADE EXPERIMENTAL → EFEITO → y_{ij}

$i = 1, 2, \ldots, t$

└→ índice de tratamento

$j = 1, 2, \ldots, r$

└→ índice de repetição

TRATAMENTOS	REPETIÇÕES 1 2 ... r	TOTAIS DE TRATAMENTOS	MÉDIAS DE TRATAMENTOS
1	y_{11} y_{12} $\cdots$ y_{1r}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21} y_{22} $\cdots$ y_{2r}	$y_{2.}$	$\bar{y}_{2.}$
$\vdots$	$\vdots$ $\vdots$ y_{ij} $\vdots$	$\vdots$ $y_{i.}$	$\vdots$ $\bar{y}_{i.}$
t	y_{t1} y_{t2} $\cdots$ y_{tr}	$y_{t.}$	$\bar{y}_{t.}$
		$y_{..}$ TOTAL GERAL	$\bar{y}_{..}$ MÉDIA GERAL

$$\bar{y}_{..} = \frac{y_{..}}{r\,t} \quad ; \quad \bar{y}_{i.} = \frac{y_{i.}}{r}$$

VARIAÇÃO TOTAL (SQTotal)

→ Variação entre tratamentos ou devida a tratamentos (SQTratamentos ou SQT)

→ Variação dentro dos tratamentos ou não devida aos tratamentos ou devida ao erro experimental (SQErro Experimental ou SQE)

$$y_{ij} - \bar{y}_{..} = y_{ij} - \bar{y}_{i.} + \bar{y}_{i.} - \bar{y}_{..}$$

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

$$\underbrace{\sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij}-\bar{y}_{..})^2}_{\text{SQTotal}} = \underbrace{\sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij}-\bar{y}_{i.})^2}_{\text{SQE}} + \underbrace{2\sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij}-\bar{y}_{i.})(\bar{y}_{i.}-\bar{y}_{..})}_{\text{igual a zero}} + \underbrace{\sum_{i=1}^{t} r(\bar{y}_{i.}-\bar{y}_{..})^2}_{\text{SQT}}$$

TABELA DA ANALISE DE VARIANCIA

CAUSAS DE VARIAÇÃO	GL	SQ	QM	F
TRATAMENTOS (entre tratamentos)	$t - 1$	SQT	QMT	$\frac{QMT}{QME}$
ERRO EXPERIMENTAL (dentro dos tratamentos)	$t(r - 1)$	SQE	QME	
TOTAL	$r\,t - 1$	SQTotal		

GL:

$\bar{y}_{..} \longrightarrow$ GLTotal $= r\,t - 1$

$\bar{y}_{..} \longrightarrow$ GLT $= t - 1$

$\bar{y}_{i.} \longrightarrow (r - 1)$ GL / Tratamento => GLE $= t(r - 1)$

SQ:

$$\text{SQTotal} = \sum_{i,j}(y_{ij}-\bar{y}_{..})^2 = \underbrace{\sum_{i,j} y_{ij}^2 - \frac{y_{..}^2}{r\,t}}_{\text{FORMA USUAL}}$$

$$\frac{y_{..}^2}{r\,t} = \text{CORREÇÃO (C) OU FATOR DE CORREÇÃO (FC)}$$

$$\longrightarrow \frac{(r\,t\,\bar{y}_{..})^2}{r\,t} = r\,t\,\bar{y}_{..}^2$$

$$\text{SQT} = \sum_{i} r\,(\bar{y}_{i.}-\bar{y}_{..})^2 = \underbrace{\sum_{i} \frac{\bar{y}_{i.}}{r} - \text{FC}}_{\text{FORMA USUAL}}$$

$$\text{SQE} = \underbrace{\sum_{i,j}(y_{ij}-\bar{y}_{i.})^2}_{\text{FORMA DIRETA}} = \underbrace{\text{SQTotal} - \text{SQT}}_{\text{FORMA USUAL}}$$

QM:

$$QM = \frac{SQ}{GL} = \text{VARIANCIA}$$

$$QMT = \frac{SQT}{GLT} = \frac{SQT}{t-1}$$

$$QME = \frac{SQE}{GLE} = \frac{SQE}{t(r-1)}$$

F:

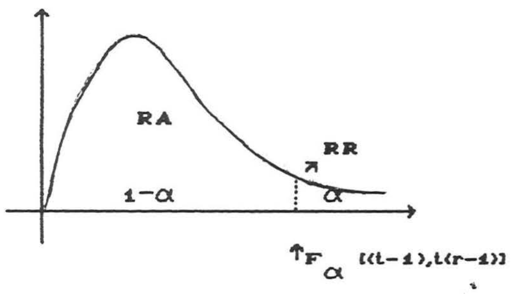
$$F = \frac{QMT}{QME}$$

TESTE DE HIPÓTESES:

Ho: $\mu_1 = \mu_2 = \ldots = \mu_t$

Ha: Pelo menos 2 médias de tratamentos diferem.

Decisão:



$$F\ \text{Calculado} = \frac{QMT}{QME}$$

Rejeita-se Ho se F calculado > $F\alpha\ [(t-1),\ t(r-1)]$

2.3 MEDIDAS DE PRECISÃO DE EXPERIMENTOS:

(1) Coeficiente de variação (CV):

$$CV = \frac{\sqrt{QME}}{\bar{y}_{..}} \times 100$$

(2) Erro padrão da média de 1 tratamento:

$$s_{\bar{y}} = \sqrt{\frac{QME}{r}}$$

(3) Erro padrão da diferença entre 2 médias:

$$s_d = \sqrt{\frac{2\ QME}{r}}$$

2.4 EXEMPLO:

Os dados abaixo referem-se a rendimento de cana em t/ha de um experimento inteiramente casualizado de competição de variedades de cana-de-açúcar.

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

$$\text{SQTotal} = 134192 - \underbrace{\frac{(1770)^2}{24}}_{\text{FC}} = 134192 - 130558 = 3654$$

$$\text{SQT} = \frac{432 + 510 + 456^2 + 372^2}{6} - \text{FC} = 132174 - 130559 = 1636$$

$$\text{SQE} = \text{SQTotal} - \text{SQT} = 3654 - 1636 = 2018$$

ou

$$\text{SQE} = \text{SQ/T}_A + \text{SQ/T}_B + \text{SQ/T}_C + \text{SQ/T}_D = 890+302+488+338 = 2018$$

ANÁLISE DE VARIÂNCIA:

CAUSAS DE VARIAÇÃO	GL	SQ	QM	F
VARIEDADES (entre variedades)	3	1636	545.3	5.40**
ERRO EXPERIMENTAL (dentro de variedades)	20	2018	100.9	
TOTAL	23	3654		

$CV = (\sqrt{QME} \,/\, \bar{y}..) \times 100 = (\sqrt{100{,}9} \,/\, 73.75) \times 100 = 13.6\%$

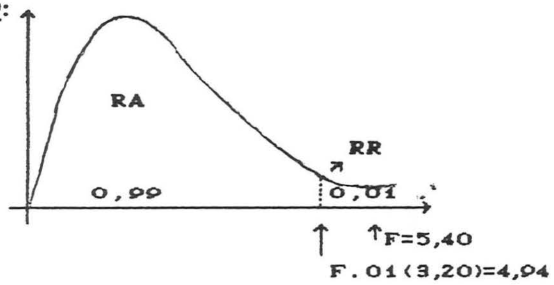
$F_{.01}\,(3.20) = 4.94$ ** SIGNIFICATIVO A 1%

HIPÓTESES:

H_o: $\mu_A = \mu_B = \mu_C = \mu_D$

H_a: Pelo menos 2 médias de variedades diferem

DECISÃO:



$F = 5.40 > F_{.01}\,(3.20) = 4.94$

A diferença entre médias de tratamentos é significativa ($P < 0.01$)

Rejeita-se Ho

CONCLUSÃO:

As variedades de cana-de-açúcar investigadas se diferenciam em termos de rendimento de cana.

2.5 EXPERIMENTOS INTEIRAMENTE CASUALIZADOS COM DIFERENTE NÚMERO DE REPETIÇÕES POR TRATAMENTO:

TRATAMENTOS	REPETIÇÕES
1	r_1
2	r_2
$\vdots$	$\vdots$
t	r_t
TOTAL	n

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

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

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

$$SQE = SQTotal - SQT$$

Erro Padrão da Diferença entre 2 Médias de Tratamentos:

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

Erro Padrão da Média de um Tratamento:

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

2.6 TÉCNICAS DE COMPLEMENTAÇÃO NA ANÁLISE DE VARIÂNCIA:

1- Ajustamento de funções de resposta através de técnicas de análise de regressão: fatores quantitativos.

2- Contrastes ortogonais: fatores quantitativos e qualitativos que permitem estruturação.

3- Comparações múltiplas de médias: fatores quantitativos, fatores qualitativos que permitem ou que não permitem estruturação.

2.7 <u>CONTRASTES OU COMPARAÇÕES DE MÉDIAS</u>:

Uma função linear de médias de tratamentos do tipo

$$C = \sum_{i=1}^{t} c_i \mu_i = c_1\mu_1 + c_2\mu_2 + \ldots + c_t\mu_t$$

é dita contraste ou comparação se $\sum_{i=1}^{t} c_i = 0$

<u>EXEMPLOS</u>:

$$\left.\begin{array}{ll} (1) & c_1 = \mu_1 - \mu_2 \\ (2) & c_2 = \mu_1 - \mu_3 \\ (3) & c_3 = \mu_1 + \mu_2 - \mu_3 \\ & \quad \equiv (\mu_1 + \mu_2)/2 - \mu_3 \end{array}\right\} \text{são contrastes}$$

$$(4) \quad c_4 = \mu_1 + \mu_2 - \mu_3 \quad \Big\} \text{ não é contraste}$$

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

são ortogonais se $\sum_{i=1}^{t} b_i c_i = 0$

C_1 e C_2 não são ortogonais: $(1)(1) + (-1)(0) + (0)(-1) = 1$
C_1 e C_3 são ortogonais: $(1)(1) + (-1)(1) + (0)(-2) = 0$
C_2 e C_3 não são ortogonais: $(1)(1) + (0)(1) + (-1)(-2) = 3$

Para diferente nº de repetições a condição de ortogonalidade é dada por:

$$\sum_{i=1}^{t} \frac{b_i c_i}{r_i} = 0$$

2.8 TESTES DE COMPARAÇÕES MÚLTIPLAS DE MÉDIAS

2.8.1 - DIFERENÇA MÍNIMA SIGNIFICATIVA (DMS)

$$DMS = \overbrace{t\ \alpha\ (GLE)}^{\text{confiabilidade}} \cdot \overbrace{s_d}^{\text{precisão}}$$

s_d → Erro Padrão da Diferença entre 2 médias de tratamentos

(GLE) → GL do Erro Experimental

α → Nível de Significância

DMS → Valor que deve atingir a diferença entre médias de tramentos para ser considerada significativa

$$s_d = \sqrt{\frac{2\ QME}{r}}$$

QME → QM Erro Experimental

r → nº de repetições

. Para ≠ nº de repetições:

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

-Qualquer diferença entre médias de tratamentos > DMS é dita significativa.

-Duas amostras independentes (2 grupos independentes)

$$t = \frac{\bar{y}_1 - \bar{y}_2}{sd} \quad ; \quad \underbrace{\bar{y}_1 - \bar{y}_2}_{\text{diferença}} = t \cdot s_d$$

<u>USO</u>

1- Não usar indiscriminadamente pois quando muitos forem os tratamentos em comparação (grande nº de comparações) algumas diferenças são consideradas como significativas por mero acaso.

2- Usar em comparações que envolvem 1 tratamento testemunha.

3- Usar em contrastes de médias planejados antes de se examinar os dados (ortogonais e de nº igual ao GL de tratamentos).

2.8.2 - TESTE DE TUKEY

$$\Delta = \overbrace{q\,\alpha\,(t, GLE)}^{\text{confiabilidade}} \cdot \overbrace{s\bar{y}}^{\text{precisão}}$$

$s\bar{y}$ → Erro Padrão da Média de um tratamento

$q\,\alpha\,(t, GLE)$ → Amplitude Total Studentizada

Δ → Diferença Mínima Significativa ou Diferença Significativa Honesta

t = nº de tratamentos

$$s_{\bar{y}} = \sqrt{\frac{QME}{r}}$$

Qualquer diferença entre 2 médias > Δ é dita significativa.

Para ≠ nº de repetições por tratamento:

$$\Delta = q\,\alpha\,(t, GLE)\sqrt{\frac{1}{2}\,\hat{V}\,(\hat{C})}$$

$$\text{onde } \hat{V}\,(\hat{C}) = QME\left[\frac{1}{r_i} + \frac{1}{r_j}\right]$$ Variância do contraste de 2 médias

$$C = \mu_i - \mu_j \; ; \quad \hat{C} = \bar{y}_{i.} - \bar{y}_{j.}$$

<u>DISTRIBUIÇÃO DA AMPLITUDE STUDENTIZADA</u>:

$$y_1, y_2, \ldots, y_n \overset{IID}{\cap} N\,(\mu, \sigma^2)$$

$$\mathcal{U} = MAX\,(y_j) - MIN\,(y_j)$$

$\mathcal{U}$ → Amplitude do conjunto de observações

s^2 com ν GL é estimador de σ^2

Então $q\,(r, \nu) = \dfrac{U}{s}$ é chamada de amplitude studentizada ou amplitude total studentizada.

$$\underbrace{\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_{t.}}_{\text{médias de tratamentos}} \overset{IID}{\cap} N\left(\mu, \frac{\sigma^2}{r}\right)$$

$$\mathcal{U} = \mathrm{MAX}\,(\bar{y}_{i.}) - \mathrm{MIN}\,(\bar{y}_{i.})$$

$$s^2 = s^2_{\bar{y}_{i.}} = \frac{QME}{r} \quad \left[\begin{array}{l}\text{VARIANCIA DA MÉDIA} \\ \text{DE UM TRATAMENTO}\end{array}\right]$$

$$s = s_{\bar{y}_{i.}} = \sqrt{\frac{QME}{r}} \quad \left[\begin{array}{l}\text{ERRO PADRÃO DA MÉDIA} \\ \text{DE UM TRATAMENTO}\end{array}\right]$$

Então:

$$\frac{\mathrm{MAX}\,(\bar{y}_{i.}) - \mathrm{MIN}\,(\bar{y}_{i.})}{\sqrt{\frac{QME}{r}}} \cap q\,(t,\ GLE)$$

CONSIDERAÇÕES SOBRE O TESTE DE TUKEY

(i) O teste de Tukey é válido para a totalidade dos contrastes entre 2 médias.

(ii) O teste de Tukey exige, em princípio, balanceamento, isto é igual nº de repetições por tratamento.

(iii) O teste de Tukey é exato para testar a maior ≠. Nos demais casos é conservador.

Relação entre t e q:

$$\underbrace{\mathrm{DMS}}_{\bar{y}_1 - \bar{y}_2} = t\, s_d \Rightarrow t = \frac{\bar{y}_1 - \bar{y}_2}{s_d} \Rightarrow t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{2\,QME}{r}}} \Rightarrow$$

$$\Rightarrow t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{2}\sqrt{\frac{QME}{r}}} \Rightarrow \sqrt{2}\; t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{QME}{r}}}$$

TUKEY:

$$\underbrace{\Delta}_{\bar{y}_1 - \bar{y}_2} = q\, s_{\bar{y}} \Rightarrow q = \frac{\bar{y}_1 - \bar{y}_2}{s_{\bar{y}}} \Rightarrow q = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{QME}{r}}} \quad \therefore q = \sqrt{2}\; t$$

2.8.3 - TESTE DE SCHEFFÉ

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

$$C = \sum_{i=1}^{t} c_i \mu_i \qquad \hat{C} = \sum_{i=1}^{t} c_i \bar{y}_{i.}$$

Para igual nº de repetições:

$$\hat{V}(\hat{C}) = \frac{QME}{r}(c_1^2 + c_2^2 + \ldots + c_t^2)$$

ou

Para diferente nº de repetições:

$$\hat{V}(\hat{C}) = QME \left[\frac{c_1^2}{r_1} + \frac{c_2^2}{r_2} + \ldots + \frac{c_t^2}{r_t} \right]$$

$\hat{V}(\hat{C})$ é a variância de um contraste de médias

Qualquer contraste > S é dito significativo.

CONSIDERAÇÕES DOBRE O TESTE DE SCHEFFÉ

Para a totalidade dos contrastes $C = \sum_{i=1}^{t} c_i \mu_i$

$$P\left[\hat{C} - Fo \sqrt{\hat{V}(\hat{C})} \leq C < \hat{C} + Fo \sqrt{\hat{V}(\hat{C})} \right] = 1 - \alpha$$

onde: $Fo = \sqrt{(t - 1)\, F\alpha\, (GLT, GLE)}$

- O Teste de Scheffé é válido para a totalidade dos contrastes.
- Para um contraste, ou para um nº pequeno deles, o Teste de Scheffé é bastante conservador.

2.8.4 - TESTE DE BONFERRONI:

DESIGUALDADE DE BONFERRONI: "Para um conjunto de k contrastes, se cada um é testado com um coeficiente de confiança $1 - \alpha$, o coeficiente de confiança conjunto é, pelo menos, $1 - k\alpha$"

Sejam dois intervalos de confiança, obtidos de uma mesma amostra, para os contrastes C_1 e C_2.

Seja o evento A_1 o evento correspondente ao complemento do intervalo de confiança para C_1 e A_2 ,analogamente, para C_2; com $P(A_1) = P(A_2) = \alpha$.

Sabe-se que:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

e

$$P(\overline{A_1 \cup A_2}) = P(\bar{A}_1 \cap \bar{A}_2) = 1 - P(A_1 \cup A_2)$$

$$\therefore \underbrace{P(\bar{A}_1 \cap \bar{A}_2)}_{} = 1 - P(A_1) - P(A_2) + P(A_1 \cap A_2)$$

→ Evento correspondente à região de confiança conjunta para C_1 e C_2

Como $P(A_1 \cap A_2) \geq 0$, serve a desigualdade:

$$P(\bar{A}_1 \cap \bar{A}_2) \geq 1 - P(A_1) - P(A_2) = 1 - 2\alpha$$

Para o caso geral, de k eventos, tem-se a desigualdade:

$$P(\bar{A}_1 \cap \bar{A}_2 \cap \ldots \cap \bar{A}_k) \geq 1 - k\alpha$$

<u>TESTE DE BONFERRONI</u>:

Consiste no teste DMS, fazendo-se correção do nível de significância em função do número de contrastes (k).

Considera-se $\alpha' = \frac{\alpha}{k}$ e obtem-se

$$DMS_B = \underbrace{t_{\alpha'}(GLE)}_{\text{confiabilidade}} \cdot \underbrace{\hat{s}_d}_{\text{precisão}}$$

- Qualquer diferença entre médias > DMS_B é dita significativa.
- Válido para um conjunto de k contrastes e é um tanto conservador. Útil quando k é pequeno.

2.8.5 - TESTE DE DUNNETT

É utilizado para comparar médias de tratamentos com média da testemunha (padrão ou controle)

$$d' = \underbrace{t^{*}\alpha\,(t',\ GLE)}_{\text{confiabilidade}} \cdot \underbrace{s\,d}_{\text{precisão}}$$

d' → diferença mínima significativa

- t^{*} de Dunnett, considerando probabilidade α, t' tratamentos (excluindo a testemunha) e GL do Erro Experimental.
- Qualquer diferença entre médias > d' é dita significativa.

CONSIDERAÇÕES DOBRE O TESTE DE DUNNET

(i) Pode ser usado também para comparações unilaterais.

(ii) Estendido para comparações com grupos selecionados de contrastes.

2.8.6 - TESTE DE STUDENT - NEWMAN - KEULS (Teste S-N-K) ou TESTE DE NEWMAN - KEULS (Teste N-K)

Teste de amplitude múltipla; isto é usa-se vários valores para as comparações, na dependência do nº de médias abrangidas pela comparação.

$$\Delta p = \overbrace{q\ \alpha\ (p, GLE)}^{\text{confiabilidade}} \cdot \overbrace{s\bar{y}}^{\text{precisão}}$$

$q\ \alpha\ (p, GLE)$ → amplitude total studentizada

Δp → diferença mínima significativa

p = nº de médias abrangidas pela comparação

→ $s\bar{y} = \sqrt{\dfrac{QME}{r}}$ para igual nº de repetições

→ $$s\bar{y} = \sqrt{\frac{QME}{2}\left(\frac{1}{r_i} + \frac{1}{r_j}\right)}$$

Qualquer diferença > Δp é dita significativa.

CONSIDERAÇÕES SOBRE O TESTE S-N-K OU TESTE N-K:

(i) Procedimento sequencial para a totalidade dos contrastes 2 a 2.

(ii) O Teste N-K, exige, em princípio, balanceamento.

(iii) O Teste N-K é um teste aproximado, pois as médias ordenadas não são independentes.

2.8.7 - TESTE DE DUNCAN (Teste de Amplitude Múltipla de Duncan)

Usa-se vários valores para as comparações, na dependência do nº de médias abrangidas na comparação.

$$\underbrace{AMS}_{\text{Amplitude mínima significativa}} = \overbrace{\underbrace{q^{*} \alpha p \; (p, GLE)}_{\text{Amplitude Studentizada Significativa (ASS)}}}^{\text{confiabilidade}} \; . \; \overbrace{s_{\bar{y}}}^{\text{precisão}}$$

p = nº de médias abrangidas na comparação

$$AMS_p = ASS \; \alpha \; (p, GLE) \; . \; s_{\bar{y}}$$

$$s_{\bar{y}} = \sqrt{\frac{QME}{r}} \quad ; \text{ para igual nº de repetições}$$

$$s_{\bar{y}} = \sqrt{\frac{QME}{2}\left(\frac{1}{r_i} + \frac{1}{r_j}\right)} \quad ; \text{ para} \neq \text{nº de repetições}$$

- TAXA DE ERRO:

$$\alpha_p = 1 - (1 - \alpha)^{p-1}$$

$\alpha = 0.05$

$p = 2 \Rightarrow \alpha_2 = 1 - (1 - 0.05)^{2-1} = 1 - 0.95 = 0.05$

$p = 3 \Rightarrow \alpha_3 = 1 - (1 - 0.05)^{3-1} = 1 - (0.95)^2 = 0.10$

$p = 4 \Rightarrow \alpha_4 = 1 - (1 - 0.05)^{4-1} = 1 - (0.95)^3 = 0.14$

$\vdots$

Valores de $q^{*}\alpha_p(p, GLE)$ correspondem a valores de amplitude studentizada $[q(p, GLE)]$ considerando $\alpha = \alpha_p$

- Qualquer diferença entre médias > do que AMS é dita significativa.

CONSIDERAÇÕES SOBRE O TESTE DE DUNCAN:

(i) Procedimento sequencial para a totalidade dos contrastes 2 a 2.

(ii) O teste de Duncan, exige, em princípio, balanceamento.

(iii) O teste de Duncan é um teste aproximado, pois as médias ordenadas não são independentes.

CASOS PARTICULARES DO TESTE DE DUNCAN:

* DMS: $\alpha_p = \alpha$

$p = 2$

*TUKEY: $\alpha_p = \alpha$

$p = t$

* S-N-K: $\alpha_p = \alpha$

$p = p$

2.8.8 - TESTE DE AMPLITUDE MÚLTIPLA DE TUKEY:

$$\Delta^*_p = \frac{1}{2}\left[\underbrace{\Delta}_{\text{TUKEY}} + \underbrace{\Delta_p}_{\text{N-K}}\right]$$

2.9 ALTERNATIVAS PARA COMPARAÇÕES MÚLTIPLAS DE MÉDIAS DE TRATAMENTOS

<u>Análise de Médias de Tratamentos</u>:

(<u>Técnicas de Detalhamento na Análise de Variância</u>)

1 - Comparações de médias tomadas 2 a 2 (comparações múltiplas)

* para tratamentos qualitativos não estruturados.
* DMS, Duncan, Tukey.

2 - Contrastes planejados (ortogonais) entre médias ou grupos de médias.

* para tratamentos qualitativos que permitam estruturação.

3 - Ajustamento de funções de resposta através de técnicas de regressão.

* tratamentos são níveis quantitativos.
* Curva de resposta: 1 variável independente.
* Superfície de resposta: 2 ou mais variáveis independentes.

Hipótese geral sobre efeito de tratamentos na análise de variância.

Ho: $\mu_1 = \mu_2 = \ldots = \mu_t \Leftrightarrow$ Ho: $\tau_i = 0$

COMPARAÇÕES MÚLTIPLAS:

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

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

$$\frac{\text{nº de inferências erradas}}{\text{nº de inferências}}$$

Usada no teste DMS.

(*ii*) Taxa de erro por experimento ("experimentwise"): E

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

Usada no teste Tukey, Scheffé, Dunnett.

(*ii*.2) $$\frac{\text{nº de inferências erradas}}{\text{nº de experimentos}}$$

Usada no teste Bonferroni.

RELAÇÃO ENTRE α e E:

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

Nº de tratamentos no experimento	α = 0.05 E	E = 0.05 α
2	0.05	0.05
3	0.0975	0.0253
4	0.1426	0.0169
5	0.1835	0.0127
10	0.3698	0.0057
15	0.5124	0.0037
20	0.6227	0.0028

MÉTODOS DE COMPARAÇÕES MÚLTIPLAS:

Testes de amplitude simples:

(1) DMS:

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

(2) TUKEY:

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

(3) SCHEFFÉ:

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

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

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

(4) BONFERRONI:

$$DMS_B = t\alpha'\ (GLE)\ s_d$$

$$\alpha' = \frac{\alpha}{k}$$

(5) <u>DUNNETT</u>: Comparações tratamentos vs testemunha

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

Testes de amplitude múltipla:

(6) <u>S-N-K</u> <u>OU</u> <u>N-K</u>:

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

→ nº de médias incluidas na comparação

(7) <u>DUNCAN</u>:

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

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

$\alpha = 0.05$

$p = 2 \Rightarrow \alpha_2 = 1 - (1 - 0.05)^{2-1} = 1 - 0.95 = 0.05$

$p = 3 \Rightarrow \alpha_3 = 1 - (1 - 0.05)^{3-1} = 1 - (0.95)^2 = 0.10$

$\vdots \qquad \vdots$

Valores de $q^*\ \alpha p\ (p, \text{GLE})$ correspondem aos valores de amplitude Studentizada $[q\ (p, \text{GLE})]$, considerando $\alpha = \alpha p$.

Casos particulares do teste de Duncan:

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

(8) <u>AMPLITUDE</u> <u>MÚLTIPLA</u> <u>DE</u> <u>TUKEY</u>:

$$\frac{1}{2}\ [\Delta + \Delta p]$$

USO DOS TESTES

1) Todas as comparações de médias 2 a 2:

TUKEY: maior rigor (maior responsabilidade)

S-N-K

DUNCAN: menor rigor (menor responsabilidade)

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

SCHEFFÉ

BONFERRONI: poucos contrastes

3) Comparações tratamentos vs testemunha:

DUNNETT

BONFERRONI: poucas comparações

DMS

RESUMO DO EXEMPLO

VARIEDADES	MÉDIAS t/ha	DMS	TUKEY	SCHEFFÉ	BONFERRONI	S-N-K	DUNCAN	DUNNETT
B C A D	85 76 72 62	a ab bc c	a ab ab b	a ab ab b	a ab ab b	a ab ab b	a ab bc c	B ≠ D C = D A = D
VALOR P/ O TESTE	2 3 4	12.1	16.2	17.7	16.5	12.1 14.7 16.2	12.10 12.71 13.04	14.9
Nº DE ≠ SIGNIFICATIVAS	-	3	1	1	1	1	3	1

RESUMO DOS TESTES:

DMS:

TAXAS DE ERRO: com base na comparação.

EXATIDÃO: exato para igual e diferente nº de repetições.

UTILIZAÇÃO: para qualquer contraste de médias.

INCOVENIENTE: não controle no acréscimo da ocorrência de erro de conclusão do tipo I.

RECOMENDAÇÃO QUANTO AO USO: -Para testar grupos específicos de contrastes de médias escolhidos a priori.

-Para testar contrastes que envolvam tratamentos x testemunha.

TUKEY:

TAXA DE ERRO: com base no experimento.

EXATIDÃO: exato para igual nº de repetições.

UTILIZAÇÃO: para comparar médias 2 a 2.

INCONVENIENTE: exagerado rigor, controlando excessivamente erro tipo I; reduzindo o poder de discriminação.

RECOMENDAÇÃO QUANTO AO USO: -Para se comparar médias 2 a 2 nos casos onde a decorrência de erro de conclusão do tipo I é extremamente grave.

SCHEFFÉ:

TAXA DE ERRO: com base no experimento.

EXATIDÃO: exato para igual e diferente nº de repetições.

UTILIZAÇÃO: para qualquer contraste de médias.

INCONVENIENTE: exagerado rigor, reduzindo o poder de discriminação.

RECOMENDAÇÃO QUANTO AO USO: -Para testar contrastes que envolvam mais do que 2 médias (não-ortogonais).

- BONFERRONI:

TAXA DE ERRO: com base no experimento.

EXATIDÃO: exato para igual e diferente nº de repetições.

UTILIZAÇÃO: para qualquer contraste de médias.

INCONVENIENTE: exagerado rigor quando o nº de contrastes for grande ⇒ resultados pouco discriminativos.

RECOMENDAÇÃO QUANTO AO USO: -Para testar contrastes (médias 2 a 2 ou qualquer contraste) quando forem de nº reduzido.

S-N-K:

TAXA DE ERRO: a cada estágio do teste (amplitude de t médias, (t-1) médias,...) a probabilidade de se rejeitar a hipótese de igualdade de médias, se verdadeira, é α.

EXATIDÃO: exato para igual nº de repetições.

UTILIZAÇÃO: para comparar médias 2 a 2.

RECOMENDAÇÃO QUANTO AO USO: -Para se comparar médias 2 a 2 para qualquer situação, decorrência grave ou não para erro de conclusão do tipo I, substituindo razoavelmente com vantagem o teste de Tukey e o teste de Duncan, para os casos específicos.

AMPLITUDE MÚLTIPLA DE TUKEY:

EXATIDÃO: exato para igual nº de repetições.

UTILIZAÇÃO: para comparar médias 2 a 2.

INCOVENIENTE: o mesmo incoveniente, em parte, do teste de Tukey.

RECOMENDAÇÃO QUANTO AO USO: -Mesmas considerações do teste de Tukey; com decorrência menos grave na ocorrência de erro tipo I.

DUNCAN:

TAXA DE ERRO: α cresce a medida que cresce a distância entre as médias.

EXATIDÃO: exato para igual nº de repetições.

UTILIZAÇÃO: para comparar médias 2 a 2.

INCOVENIENTE: pouco controle com a ocorrência de erro tipo I, dado que a preocupação é com o erro tipo II, isto é, com o poder de discriminação dos resultados.

RECOMENDAÇÃO QUANTO AO USO: -Para se comparar médias 2 a 2 nos casos de investigação que se exige alto grau de discriminação e a decorrência de erro de conclusão do tipo I não é tão grave.

DUNNETT:

TAXA DE ERRO: com base no experimento.

EXATIDÃO: exato para igual nº de repetições para tratamentos.

UTILIZAÇÃO: para comparar médias 2 a 2 do tipo tratamento x testemunha.

INCOVENIENTE: rigor muitas vezes exagerado, o que pode conduzir a pouca discriminação.

RECOMENDAÇÃO QUANTO AO USO: -Comparações unilaterais ou bilaterais do tipo tratamento x testemunha, substituindo o DMS para os casos que cresce o nº de comparações.

INCONVENIENTES DOS PROCEDIMENTOS DE COMPARAÇÕES MÚLTIPLAS DE MÉDIAS

(1) taxa de erro

(2) a não consideração do valor da estatística F obtida na análise de variância: teste t bayesiano ou teste de Waller-Duncan ou teste da Razão bayesiana k.

(3) a não formação de grupos mutuamente exclusivos de médias de tratamentos: técnica multivariada de análise de agrupamento "cluster analysis".

2.10 - TESTE BAYESIANO OU TESTE DE WALLER-DUNCAN OU TESTE DA RAZÃO BAYESIANA K:

O teste encontra-se descrito em:
Duncan (1965); Waller e Duncan (1969 e 1972); Steel e Torrie (1980); Gomes (1985)e é um dos procedimentos de comparações múltiplas disponíveis no SAS

$$DMS_k = \overbrace{t\,(k,\ F,\ GLT,\ GLE)}^{\text{confiabilidade}} \cdot \overbrace{s_d}^{\text{precisão}}$$

F → valor calculado de F

k → razão bayesiana

t → valor tabelado

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

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

α = 0.01	0.05	0.01
k = 50	100	500

F alto: heterogeneidade entre tratamentos
valor de t é reduzido
alto poder de discriminação
taxa de erro "a nível de comparação" ou por comparação

F baixo: homogeneidade aproximada de tratamentos
valor de t é aumentado
rigoroso
taxa de erro "a nível de experimento" ou por experimento

O teste usa as vantagens dos procedimentos de taxa de erro por experimento e por comparação sem as desvantagens.

TABELA: INTERPOLAÇÃO

F	f	q ≤ 100	q > 100	q ≤ 20	q > 20
≤ 2.4	≤ 60	a	a	-	-
	> 60	a	b	-	-
> 2.4	≤ 20	-	-	a	b
	> 20	-	-	b	b

f = GLE ; q = GLT

$a = \sqrt{\frac{1}{F}}$; $b = \sqrt{F/(F-1)}$ e interpola-se nas tabelas.

Para t ≥ 15 , onde t é o nº de tratamentos e GLE ≥ 30

$t(100, F, \infty, \infty) = 1.72\ b$

$t(500, F, \infty, \infty) = 2.23\ b$

EXEMPLO: Cultivares de cana-de-açúcar.

C. Variação	GL	QM	F
VARIEDADES	3		5.4
ERRO	20	100.9	

t = 4

r = 6

F = 5.4 f = GLE = 20 q = GLT = 3

$a = \sqrt{1/F} = \sqrt{1/5.4} = 0.430$

$\alpha = 0.05 \Rightarrow k = 100$

f = 20 q = 2 F = 4 a = 0.5 t = 2.18

f = 20 q = 3 F = 4 a = 0.5 t = ? = 2.18 + 0.027 = 2.207

f = 20 q = 4 F = 4 a = 0.5 t = 2.22

INTERPOLAÇÃO HARMÔNICA:

$$\frac{1}{2} - \frac{1}{4} = \frac{1}{4} \longrightarrow 0.04$$
$$\frac{1}{2} - \frac{1}{3} = \frac{1}{6} \longrightarrow x \quad \Rightarrow x = \frac{(0.04)\ 1/6}{1/4} = 0.027$$

$f = 20 \quad q = 2 \quad F = 6 \quad a = 0.408 \quad t = 2.08$

$f = 20 \quad q = 3 \quad F = 6 \quad a = 0.408 \quad t = ? = 2.08 + 0.007 = 2.087$

$f = 20 \quad q = 4 \quad F = 6 \quad a = 0.408 \quad t = 2.08$

$$\frac{1}{2} - \frac{1}{4} = \frac{1}{4} \longrightarrow 0.01$$

$$\frac{1}{2} - \frac{1}{3} = \frac{1}{6} \longrightarrow x$$

$$\Rightarrow x = \frac{(0.01)\,(4)}{6} = 0.07$$

$f = 20 \quad q = 3 \quad F = 4 \quad a = 0.5 \quad t = 2.207$

$f = 20 \quad q = 3 \quad F = 5.4 \quad a = 0.430 \quad t = ? = 2.087 + 0.029 = 2.116$

$f = 20 \quad q = 3 \quad F = 6 \quad a = 0.408 \quad t = 2.087$

$$0.5 - 0.408 = 0.092 \longrightarrow 2.207 - 2.087 = 0.12$$

$$0.430 - 0.408 = 0.022 \longrightarrow x$$

$$x = \frac{(0.12)\,(0.022)}{0.092} = 0.029$$

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

$$= t(100;\ 5.4;\ 3;\ 20)\ s_d = (2.114)\,(5.2) = 12.3\ t/ha$$

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

$$DMS = (2.086)\,(5.8) = 12.1\ t/ha$$

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

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

2.11. MÉTODO DE ANÁLISE DE AGRUPAMENTO ("CLUSTER ANALISYS") OU MÉTODO DE SCOTT E KNOTT PARA AGRUPAMENTO DE MÉDIAS

SCOTT E KNOTT (1974) [BIOMETRICS]

1- ANÁLISE DE AGRUPAMENTO "CLUSTER ANALISYS"

UNIDADES AMOSTRAIS	VARIÁVEIS 1	2 . . .	p
1	X_{11}	X_{12}	X_{1p}
2	X_{21}	X_{22}	X_{2p}
⋮	⋮	⋮	⋮
n	X_{n1}	X_{n2}	X_{np}

* OBJETIVO: Reunir as unidades amostrais em um número de grupos de tal forma que exista homogeneidade dentro e heterogeneidade entre os grupos.
* MEDIDA DE PROXIMIDADE
* MÉTODO DE AGRUPAMENTO

2. MEDIDAS DE PROXIMIDADE

- SIMILARIDADE: semelhança; medida de correlação
- DISSIMILARIDADE: dessemelhança; medida de distância

2.1. MEDIDAS DE DISSIMILARIDADE

(1) DISTANCIA EUCLIDIANA

Entre os indivíduos (unidades amostrais) i e i' é dada por

$$d_{ii'} = \left[\sum_{j=1} \left\{ X_{ij} - X_{i'j} \right\}^2 \right]^{1/2}$$

(2) DISTÂNCIA EUCLIDIANA MÉDIA

$$\Delta_{ii'} = \frac{1}{\sqrt{p}} \, d_{ii'}$$

(3) DISTÂNCIA GENERALIZADA DE MAHALANOBIS

Entre as unidades amostrais i e i' é dada por

$$D^2_{ii'} = \left(\underset{\sim}{\bar{X}i} - \underset{\sim}{\bar{X}i'} \right)' S^{-1} \left(\underset{\sim}{\bar{X}i} - \underset{\sim}{\bar{X}i'} \right) ,$$

onde S é a matriz de dispersão amostral comum a todas as unidades amostrais, $\underset{\sim}{\bar{X}i}$, $\underset{\sim}{\bar{X}i'}$ são os vetores p dimensionais de médias da unidade i, i', com $i, i' = 1, 2, \ldots, n$ ($i \neq i'$).

2.2 MEDIDAS DE SIMILARIDADE

-Coeficiente de correlação Momento-Produto de Pearson

Para os indivíduos (unidades amostrais) i e i' é dado por

$$r_{ii'} = \frac{\sum_j X_{ij} X_{i'j} - \frac{1}{p}\left(\sum_j X_{ij}\right)\left(\sum_j X_{i'j}\right)}{\sqrt{\left[\sum_j X^2_{ij} - \frac{1}{p}\left(\sum_j X_{ij}\right)^2\right]\left[\sum_j X^2_{i'j} - \frac{1}{p}\left(\sum_j X_{i'j}\right)^2\right]}}$$

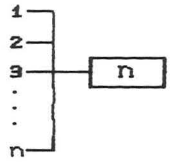
3) TÉCNICAS DE AGRUPAMENTO

(1) <u>MÉTODOS HIERARQUICOS</u>: Os indivíduos são reunidos em grupos e o processo repete-se em diferentes níveis até formar uma árvore.

(2) <u>MÉTODOS DE OTIMIZAÇÃO</u>: Os grupos são formados pela otimização de um critério de agrupamento. Os grupos são mutuamente exclusivos, formando uma partição do conjunto de entidades.

3.1 - TÉCNICAS HIERARQUICAS DE AGRUPAMENTO

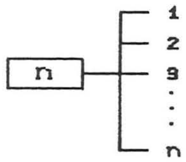
(1) <u>MÉTODOS AGLOMERATIVOS</u>: Produzem uma série de fusões sucessivas das <u>n</u> entidades, terminando o processo no estágio onde todas as entidades estão num único grupo.



(a) Método do vizinho mais próximo ou método do encadeamento simples;

(b) Método do vizinho mais distante ou método do encadeamento completo.

(2) <u>MÉTODOS DIVISIVOS</u>: Particionam o conjunto de <u>n</u> entidades sucessivamente até a partição final, formando grupos individuais. -Método de Edwards e Cavalli-Sforza (1965).



3.2-TÉCNICAS DE OTIMIZAÇÃO.

Critérios de agrupamento derivados da identidade:

$$T=H+R$$

(1) Minimização do traço (R);

(2) Minimização do $|R|$ <=> maximização $|T| \,/\, |R|$;

(3) Minimização do traço ($H\ R^{-1}$).

<u>Método de Edwards e Cavalli-Sforza</u>: Calculam-se os quadrados das distâncias euclidianas entre os indivíduos e comparam-se todas as possíveis partições dos indivíduos em dois grupos, de maneira a produzir uma partição que é caracterizada pela menor SQ dentro dos grupos $\Leftrightarrow$ maior SQ entre grupos.

Processo continua dentro de cada subgrupo, chegando no final a grupos individuais.

<u>EXEMPLO</u>: Consideremos os dados a seguir ordenados de forma crescente e a respectiva matriz de distâncias euclidianas ao quadrado.

Indivíduos	Variável(x)
5	1
1	2
2	7
4	9
3	12

$$d_{ii'}^{2} = (X_i - X_{i'})^{2}$$

$$D = \begin{array}{c} \\ 5 \\ 1 \\ 2 \\ 4 \\ 3 \end{array} \begin{bmatrix} 5 & 1 & 2 & 4 & 3 \\ 0 & 1 & 36 & 64 & 121 \\ & 0 & 25 & 49 & 100 \\ & & 0 & 4 & 25 \\ & & & 0 & 9 \\ (\text{sim.}) & & & & 0 \end{bmatrix}$$

Aplicando-se o método obteremos:

Nº de grupos	Partição	SQ entre	SQ dentro	
1	1 2 3 4 5	0,0 (0%)	86,8(100%)	= (1+36+...+9)/5
2	1 5/2 3 4	73,6 (85%)	13,2 (15%)	= 1/2+(4+25+9)/3
3	1 5/2 4/3	84,3 (97%)	2,5 (3%)	= 1/2+4/2
4	1 5/2/4/3	86,3 (99%)	0,5 (1%)	
5	1/2/3/4/5	86,8(100%)	0,0 (0%)	

k-1 caminhos ; com valores ordenados

k=5
k-1=4 caminhos

5/1 2 4 3   SQD=53
5 1/2 4 3   SQD=13,2
5 1 2/4 3   SQD=25,2
5 1 2 4/3   SQD=44,75

MÉTODO DE SCOTT E KNOTT (1974)

Utilizam o método divisivo de Edwards e Cavalli-Sforza (1965) como técnica de agrupamento das médias de tratamentos e para verificar se os grupos formados em determinada partição são significativamente diferentes definem a estatística

$$\lambda = \frac{\Pi}{2(\Pi' - 2)} \; \frac{\beta_0}{\hat{\sigma}_0^2}$$

onde β_0 = SQ entre grupos máximo = BSS máximo

$$\hat{\sigma}_0^2 = \left[\sum_1^k \left(\bar{y}_i - \bar{y} \right)^2 + \nu\, s^2_{\bar{y}} \right] / k + \nu$$

sendo k o nº de médias

$s^2_{\bar{y}}$ a variância comum às médias de tratamentos $\left[s^2_{\bar{y}} = QME/r \right]$

ν seu GL associado

λ tem distribuição aproximada de χ^2 com ν_0 GL

onde $\nu_0 = \dfrac{k}{\Pi - 2}$

Exemplo: Variedades de cana-de-açúcar

Variedades	Médias (t/ha)
B	85
C	76
A	72
D	62
média	73,75

C. Variação	GL	QM
var	3	
erro	20	100,9
t=4	r=6	

$$s^2_{\bar{y}} = \frac{QME}{r} = \frac{100,9}{6} = 16,82$$

matriz de distâncias Euclidianas ao quadrado

$$d_{ii'} = (X_i - X_{i'})^2$$

$$\begin{array}{c} B \\ C \\ A \\ D \end{array} \begin{bmatrix} 0 & 81 & 169 & 529 \\ & 0 & 16 & 196 \\ & & 0 & 100 \\ (sim.) & & & 0 \end{bmatrix}$$

Partição	SQDentro (SQD)	SQEntre (SQE)
B C A D	272,75	0
B C A/D	88,7	184,05
B/C A/D	8	264,75
B/C/A/D	0	272,75

$$(81 + 169 + \ldots + 100)/ 4 = 272,75$$

1ª Partição:

B/CAD $\quad SQD = (16 + 196 + 100)/ 3 = 104$

BC/AD $\quad SQD = 81/ 2 + 100/ 2 = 40,5 + 50 = 90,5$

BC/AD $\quad SQD = (81 + 169 + 16)/ 3 = 88,7 \quad$ <= mínimo

$\therefore$ SQE máximo = BSS máximo = 272,75 - 88,7 = 184,05

$$\lambda = \frac{\pi}{2(\pi-2)} \; \frac{BSS}{\hat{\sigma}^2} = \frac{3,1416}{2(3,1416-2)} \; \frac{184,05}{25,38} = 9,98$$

$$\hat{\sigma}^2 = \left[\sum_1^k \left(y_i - \bar{y} \right)^2 + \upsilon s^2_{\bar{y}} \right] / k + \upsilon \qquad \upsilon = 20 =$$

$$GL. = \upsilon_0 = \frac{k}{\pi - 2} = \frac{4}{3,1416-2} = 3,5 \simeq 3$$

$$\chi^2_{.05(3)} = 7,81$$

$$\hat{\sigma}^2 = \Big[\underbrace{(85-73,75)^2 + \ldots + (62-73,75)^2}_{272,75}\ (20) + (16,82)\Big] / 4+20 = 25,38$$

k= nº de médias na partição = 4

<u>2ª Partição:</u>

B/C A SQD = 16 / 2 = 8 ← mínimo → SQE máximo = BSS máximo = = 88,7 − 8 = 80,7

B C/A SQD = 81/ 2 = 40,5

$$\lambda = \frac{3,1416}{2(3,1416-2)}\ \frac{80,7}{18,48} = 6,01$$

$$GL = \nu_0 = \frac{k}{\pi - 2} = \frac{3}{3,1416-2} = 2,65 \simeq 2$$

$$\hat{\sigma}^2 = \Big[\ 88,7 + 20(16,82)\Big] / 3 + 20 = 18,48$$

$$\chi^2_{.05(2)} = 5,99$$

<u>3ª Partição:</u>

C/A SQD = 0 ← mínimo ⇒ SQE máximo = BSS máximo = 8

$$\lambda = \frac{3,1416}{2(3,1416-2)}\ \frac{8}{15,65} = 0,70$$

$$GL = \nu_0 = \frac{k}{\pi - 2} = \frac{2}{3,1416-2} = 1,75 \simeq 1$$

$$\hat{\sigma}^2 = \Big[\ 8 + 20(16,82)\Big] / 2 + 20 = 15,65$$

$$\chi^2_{.05(1)} = 3,84$$

B C A D

- B C A (SQD=88,7)
  - B (SQD=8)
    λ= 6,01* GL= 2
    BSS máximo = 80,7
    $\chi^2_{.05(2)}$ = 5,99
  - C A
    - C (SQD = 0)
      BSS máximo = 8
      λ=0,7 NS GL= 1
      $\chi^2_{.05(1)}$ = 3,84
    - A
- BSS máximo = 184,05
  λ= 9,98*
  GL= 3
  $\chi^2_{.05(3)}$ = 7,81
- D

$\boxed{B}$ $\boxed{C\ A}$ $\boxed{D}$

Var	TUKEY N-K	Duncan DMS DMSK
B	a	a
C	a b	a b
A	a b	b c
D	b	c

2.12 <u>CONTRASTES</u> <u>ORTOGONAIS</u>

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

Uma função linear de médias de tratamentos do tipo

$C = \sum_{i=1}^{t} c_i\mu_i = c_1\mu_1 + c_2\mu_2 + \ldots + c_t\mu_t$ é dita contraste ou comparação se $\sum_{i=1}^{t} c_i = 0$

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

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

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

Exemplos

$$\left.\begin{array}{l} (1)\ C_1 = \mu_1 - \mu_2 \\ (2)\ C_2 = \mu_1 - \mu_3 \\ (3)\ C_3 = \frac{\mu_1 + \mu_2}{2} - \mu_3 = \mu_1 + \mu_2 - 2\mu_3 \end{array}\right\} \text{são contrastes}$$

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

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

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

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

Causas de Variação		GL
tratamentos		$t-1$
contrastes ortogonais	C_1	1
	C_2	1
	:	:
	:	:
	C_{t-1}	1
Erro experimental		$t(r-1)$
Total		$rt-1$

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

Para cada contraste calcula-se a SQ dada por

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

Onde $y_{i.}$ = total do tratamento i

r = número de repetições

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

$j = 1 , 2 , \ldots , t-1$ (numero de cantrastes)

Testa-se cada contraste pela estatística F obtendo-se

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

Se os contrastes forem ortogonais :

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

Se os contrastes não forem ortogonais :

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

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

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

Exemplo 1 :

Variedades da cana-de-açúcar

	Variedades		médias , t/ha	totais , t/ha
origem genética	A	novas	72	432
origem genética2	B	novas	85	510
origem genética	C	novas	76	456
padrão ou testemunha ou controle	D		62	372

t=4 r=6

Análise de Variância :

Causas da variação	GL	SQ	QM
Variedades	3	1636	
erro experimental	20	2018	100,9

3 GL ⇒ 3 contrastes ortogonais

Contrastes :

C_1 : D versus (A+B+C) ⇔ [testemunha vs resto] ⇔

⇔ [novas variedades vs variedade padrão]

C_2 : B versus (A+C) ⇔ [origem genética2 vs origem genética1]

C_3 : A versus C ⇔ [entre variedades de origem genética1]

↓

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

$C_2 = \mu A + \mu C - 2\mu B$ $\qquad C_1 , C_2 , C_3$ ortogonais

$C_3 = \mu A - \mu C$

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

$$SQC_1 = \frac{(282)^2}{(6)(12)} = 1104 \qquad SQC_3 = \frac{(-24)^2}{(6)(2)} = 48$$

$$SQC_2 = \frac{(-132)^2}{(6)(6)} = 484$$

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

Análise de Variância

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

$F_{.01(1,20)} = 8,10$

$F_{.05(1,20)} = 4,35$

Conclusões :

· Novas variedades (em média) superiores à variedade padrão

· A variedade de origem genética 2 (variedade B) é superior as variedades de origem genética 1 (variedades A e C) .

· Não se evidenciam diferenças entre as variedades de origem genética 1 .

→ Utilizar a variedade B (melhor variedade)

<u>OUTRA FORMA DE OBTENÇÃO DAS SOMAS DE QUADRADOS DOS CONTRASTES</u>

C_1 : $\underset{18}{\boxed{ABC}}$ vs $\underset{6}{\boxed{D}}$:

$$SQC_1 = \frac{\left(y_A+y_B+y_C\right)^2}{16} + \frac{y_O^2}{6} - \frac{\left(y_A+y_B+y_C+y_D\right)^2}{24}$$

$$= \frac{(432+510+456)^2}{18} + \frac{372^2}{6} - \frac{(1770)^2}{24} = 1107$$

C_2 : $\underset{12}{\boxed{AC}}$ vs $\underset{6}{\boxed{B}}$:

$$SQC_2 = \frac{\left(y_A+y_C\right)^2}{12} + \frac{y_B^2}{6} - \frac{\left(y_A+y_B+y_C\right)^2}{18}$$

$$= \frac{(432+456)^2}{12} + \frac{510^2}{6} - \frac{(432+510+456)^2}{18} = 484$$

C_3 : $\underset{6}{\boxed{A}}$ vs $\underset{6}{\boxed{C}}$:

$$SQC_3 = \frac{y_A^2 + y_B^2}{6} - \frac{\left(y_A+y_C\right)^2}{12}$$

$$= \frac{432^2+456^2}{6} - \frac{(432+456)^2}{12} = 48$$

TESTES DE CONTRASTES UTILIZANDO-SE A ESTATÍSTICA t

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

$$\frac{\hat{C}_j - C_j}{\sqrt{\hat{V}(\hat{C}_j)}} \cap t(GLE) \quad \xrightarrow[H_o]{sob} \quad t = \frac{\hat{C}_j}{\sqrt{\hat{V}(\hat{C}_j)}}$$

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

[para igual número de repetições por tratamento]

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

[para diferente número de repetições por tratamento]

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

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

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

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

$$= (2)(100,9) = 201,8$$

$t_{.01}(20)=2,845$

GLE=20

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

└ rejeita-se H_0

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

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

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

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

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

$$t = \frac{\hat{C}_2}{\sqrt{\hat{V}(\hat{C}_2)}} = \frac{-22}{\sqrt{100,9}} = -2,19\ * \qquad t_{.05(20)}=2,086$$

rejeita-se H_o

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

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

$C_3 = \mu_A - \mu_C$

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

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

$$t = \frac{\hat{C}_3}{\sqrt{\hat{V}(\hat{C}_3)}} = \frac{-4}{\sqrt{33,63}} = -0,69\ NS \qquad t_{.05(20)}=2,086$$

Aceita-se H_o

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

Exemplo 2:

Num experimento , realizado em DCC com quatro repetições por tratamento , para estudar o efeito de aplicações de enxofre (S) para reduzir a incidência da "Sarna Comum" em batatinha , obtiveram-se os seguintes resultados , expressos em termos de "índice de infecção de sarna" , nos cinco seguintes tratamentos :

Enxofre (S)	Sem (testemunha)	Com S_1		Com S_2	
Época de aplicações		outono	primavera	outono	primavera
Totais dos tratamentos	100	64	80	40	68

S_1 = dose simples de S

S_2 = dose dupla de S

SQE = 158,2

CONTRASTES :

↓

5 tratamentos

↓

4 GL

↓

4 contrastes ortogonais

C_1: Efeito de Enxofre [sem S vs com S]

C_2: Efeito de dose de S [S_1 vs S_2]

C_3: Efeito de época para S_1 [outono vs primavera para S_1]

C_4: Efeito de época para S_2 [outono vs primavera para S_2]

SOMA DOS QUADRADOS DOS CONTRASTES

Tratamentos	totais de tratamentos ($y_{i.}$)	coeficiente dos contrastes				$c_{1i}y_{i.}$	$c_{2i}y_{i.}$	$c_{3i}y_{i.}$	$c_{4i}y_{i.}$
		c_{1i}	c_{2i}	c_{3i}	c_{4i}				
Testemunha	100	4	0	0	0	400	0	0	0
S_1 no outono	64	-1	1	1	0	-64	64	64	0
S_1 na primavera	80	-1	1	-2	0	-80	80	-80	0
S_2 no outono	40	-1	-1	0	1	-40	-40	0	40
S_2 na primavera	68	-1	-1	0	-1	-68	-68	0	-68
Total	352	0	0	0	0	148	36	-16	-28

$$SQC_j = \frac{\left(\sum_{i=1}^{t} c_{ji} y_{i.}\right)^2}{r \sum_{i=1}^{t} c_{ji}^2} \qquad t=5 \quad r=4$$

$$SQC_1 = \frac{(148)^2}{(4)(20)} = 273,8$$

$$SQC_2 = \frac{(36)^2}{(4)(4)} = 81$$

$$SQC_3 = \frac{(-16)^2}{(4)(2)} = 32$$

$$SQC_4 = \frac{(-28)^2}{(4)(2)} = 98$$

$$\overline{}$$
$$484,8 = SQtratamentos$$

Análise de Variância

C. Variação	GL	SQ		QM	F
Tratamentos	4	484,8		121,2	11,49 *
C_1	1	273,8	484,8	273,8	25,95 *
C_2	1	81		81	7,68 *
C_3	1	32		32	3,03 NS
C_4	1	98		98	9,29 *
Erro experimental	15	158,92		10,55	
Total	19				

$F_{.05(1,15)} = 4,54$

- Efeito de S / com S $\Rightarrow$ menor incidência de sarna
- Efeito de dose / S_2 $\Rightarrow$ menor incidência de sarna
- Não se evidencia efeito de época / S_1
- Efeito de época / S_2 ; outono menor incidência que na primavera

$\rightarrow$ Aplicar dose dupla de S (S_2) no outono

Considerações

(1) Sempre que possível usar essa técnica em substituição aos testes de comparações múltiplas , não apresentando a técnica de contrastes ortogonais inconvenientes quanto a taxa de erro , pois consiste simplesmente na decomposição (partição) da SQtratamentos .

(2) Escolher sempre contrastes lógicos que devem ser estabelecidos ao planejar o experimento , ou no mínimo antes de se examinar os dados , evitando com isso a introdução de vicio no teste .

(3) Procurar sempre estabelecer contrastes que sejam ortogonais

(4) Em casos de contrastes não-ortogonais , usar preferencialmente , se pelo menos um contraste incluir mais do que duas médias , o teste de Scheffé .

Exemplo 3 : Experimento inteiramente casualizado com diferente número de repetições por tratamento .

Exemplo das variedades de cana-de-açúcar

variedades

A	B	C	D
54	78	75	55
x	x	x	x
x	x	x	x
77	x	x	x
	85	x	x
		76	68

r_i	4	5	6	6	21
$y_{i.}$	281	433	456	372	1542
$\bar{y}_{i.}$	70,25	86,60	76	62	73,43
variância dentro	30,92	56,3	97,6	67,6	67,3

$$QME = \frac{\overbrace{(4-1)30,92+(5-1)56,3+(6-1)(97,6+67,6)}^{SQE}}{\underbrace{3+4+5+5}_{17\ =\ GLE}}$$

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

n m s = P > F = 0,001376

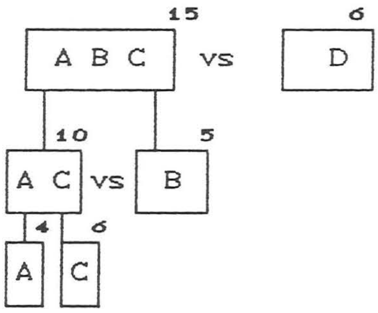
$$SQ_{total} = 64^2 + \ldots + 68^2 - (1542)^2 / 21$$

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

SQE = SQtotal − SQT

1144 = SQE

<u>CONTRASTE</u> :



: c_1: A + B + C vs D $\Leftrightarrow$ [novas vs padrão]

c_2: A + C vs B $\Leftrightarrow$ [origem genética1 vs origem genética2]

c_3: A vs C $\Leftrightarrow$ [dentro da origem genética1]

Forma simplificada para cálculo das SQ dos Contrastes

$$SQc_1 = \frac{\left(y_{A.} + y_{B.} + y_{C.}\right)^2}{15} + \frac{y_{D.}^2}{6} - \frac{\left(y_{A.} + y_{B.} + y_{C.} + y_{D.}\right)^2}{21}$$

$$= \frac{1170^2}{15} + \frac{372^2}{6} - \frac{1542^2}{21} = 1097$$

$$SQc_2 = \frac{\left(y_{A.} + y_{C.}\right)^2}{10} + \frac{y_{B.}^2}{5} - \frac{\left(y_{A.} + y_{B.} + y_{C.}\right)^2}{15}$$

$$= \frac{(737)^2}{10} + \frac{433^2}{5} - \frac{1170^2}{15} = 554{,}7$$

$$SQc_3 = \frac{y_{A.}^2}{4} + \frac{y_{C.}^2}{6} - \frac{\left(y_{A.} + y_{C.}\right)^2}{10} = \frac{281^2}{4} + \frac{456^2}{6} - \frac{737^2}{10} = 79{,}3$$

$$SQc_1 + SQc_2 + SQc_3 = 1097 + 554{,}7 + 79{,}3 = 1731 = SQT$$

Cálculo das SQ dos contrastes usando os coeficientes

$$c_1 = 6\,(4\,\mu A + 5\,\mu B + 6\,\mu C) - 15(6\,\mu D)$$

$$= 4\,\mu A + 5\,\mu B + 6\,\mu C - 15\,\mu D$$

$$c_2 = 5\,(4\,\mu A + 6\,\mu C) - 10(5\,\mu B)$$

$$= 4\,\mu A + 6\,\mu C - 10\,\mu B$$

$$c_3 = 6\,(4\,\mu A) - 4\,(6\,\mu C) = \mu A - \mu C$$

Verificação de Ortogonalidade : $\sum \frac{b_{ji}c_{ji}}{r_i} = 0$

Expressão para calculo da SQ :

$$SQc_j = \frac{\left(\sum_{i=1}^{t} c_{ji}\,\bar{y}_{i.}\right)^2}{\sum_{i=1}^{t} \frac{c_{ji}^2}{ri}}$$

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

$$SQc_1 = \frac{(240)^2}{\frac{4^2}{4} + \frac{5^2}{5} + \frac{6^2}{6} + \frac{(-15)^2}{6}} = 1097$$

$$SQc_2 = \frac{(-129)^2}{\frac{4^2}{4} + \frac{(-10)^2}{5} + \frac{6^2}{6}} = 554,7$$

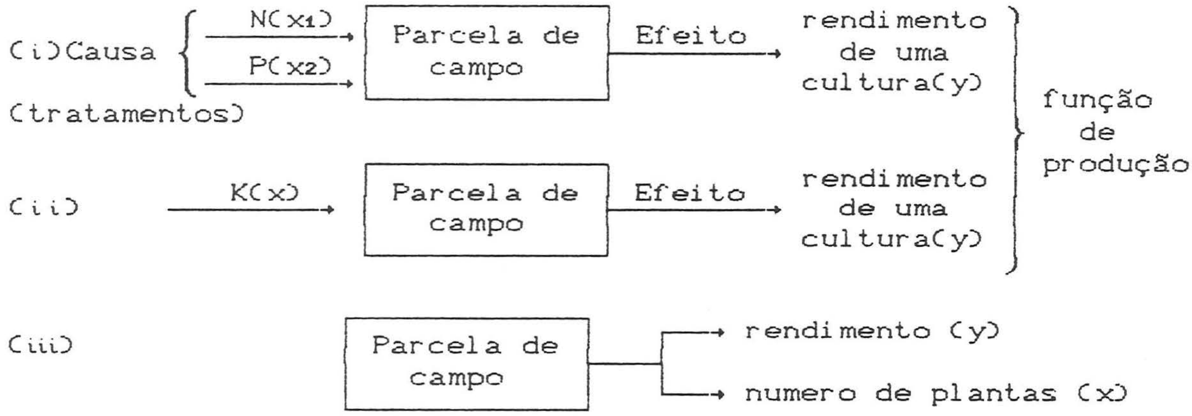
$$SQc_3 = \frac{(-5,75)^2}{\frac{1^2}{4} + \frac{(-1)^2}{6}} = 79,3$$

$$\left.\begin{matrix} SQc_1 \\ SQc_2 \\ SQc_3 \end{matrix}\right\} 1731 = SQT$$

erro padrão do contraste

Contraste	GL	SQ	QM	F	$\hat{c}_j$	$\sqrt{\hat{V}(\hat{c}_j)}$	t	decisão
c_1	1	1097	1097	16,23	240	59,44	4,038	*
c_2	1	554,7	554,7	8,24	-129	49,93	-2,871	*
c_3	1	79,3	79,3	1,18	-5,75	5,30	-1,086	NS

2.13. ANÁLISE DE REGRESSÃO



(i) $\underset{\text{variável dependente}}{y} = f\underbrace{(x_1, x_2)}_{\text{variáveis independentes}}$

(ii) e (iii) $y = f(x)$

· Interessa estabelecer o relacionamento funcional entre os dois tipos de variáveis .

REGRESSÃO LINEAR SIMPLES

Relação entre y e x é dada por uma reta

$$\hat{y} = a + bx$$

b → coeficiente de regressão linear (inclinação da reta)

a → intercepto da reta

$$b = \frac{\text{covariância } (x, y)}{\text{variância } (x)} = \frac{SP_{xy}}{SQ_x}$$

SP_{xy} = soma dos produtos dos desvios de x e y

$$SP_{xy} = \Sigma(x-\bar{x})(y-\bar{y}) = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}$$

$n \longrightarrow$ numero de pares de valores

SQ_x = soma dos quadrados dos desvios de x

$$SQ_x = \Sigma(x-\bar{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

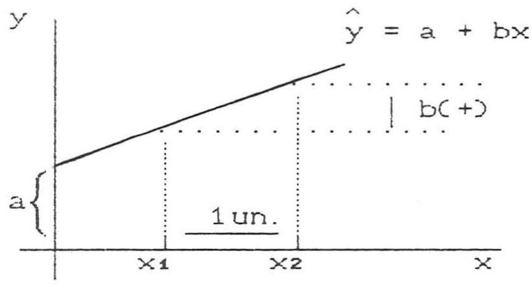
$$\text{Covariância } (x,y) = \frac{SP_{xy}}{\underbrace{n-1}_{GL}}$$

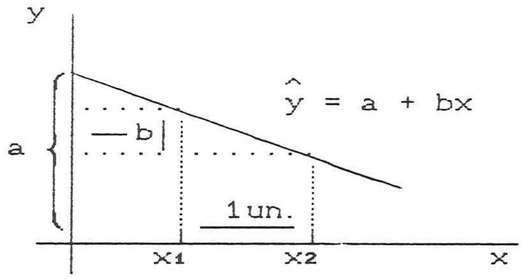
Medida de variação simultânea de x e y

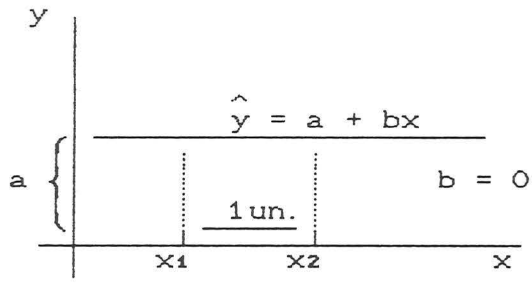
$$a = \bar{y} - b\bar{x}$$

<u>a</u> : Intercepto da reta . Ponto onde a reta corta o eixo dos y. É o valor estimado para y quando x = 0 .

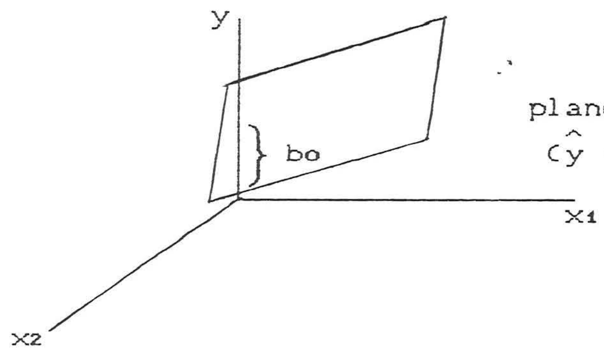
<u>b</u> : Inclinação da reta . <u>Coeficiente de regressão linear</u> . Representa quanto aumenta ou diminui y quando x cresce de uma unidade .







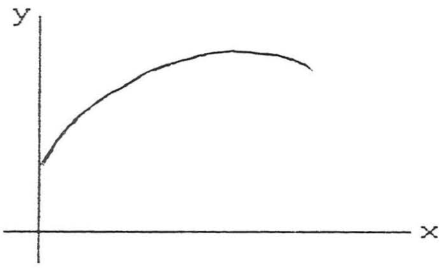
<u>REGRESSÃO MÚLTIPLA</u>



plano de regressão
$(\hat{y} = b_0 + b_1x_1 + b_2x_2)$
$\quad \rightarrow$ Intercepto
b_1 e b_2 coeficientes
de regressão parciais

b_1 : Quantidade que afeta y quando x_1 varia de uma unidade permanecendo constante x_2 .

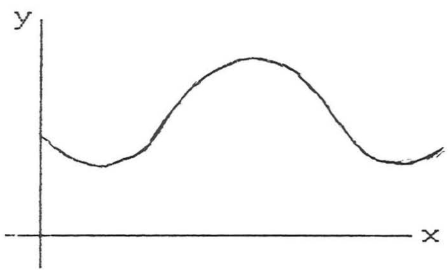
REGRESSÃO QUADRATICA



$\bar{y} = a + bx + cx^2$

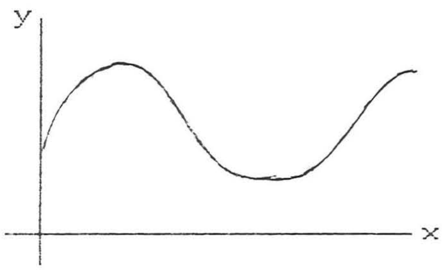
(curva de resposta)

REGRESSÃO CÚBICA OU DE 3º GRAU



$\hat{y} = a + bx + cx^2 + dx^3$

REGRESSÃO QUARTICA OU DE 4º GRAU



$\hat{y} = a + bx + cx^2 + dx^3 + ex^4$

SUPERFICIE DE RESPOSTA

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_{11}x_1^2 + b_{22}x_2^2 + b_{12}x_1x_2$$

ANÁLISE DE REGRESSÃO PARA COMPARAÇÃO DE MÉDIAS DE TRATAMENTOS

(Ajuste de regressões curvilineares pela técnica de polinômios ortogonais)

Sempre que os tratamentos representam níveis de um ou mais fatores quantitativos , a técnica mais recomendável (por ser mais informativa e mais eficiente para identificar diferenças entre médias de tratamentos) na decomposição da SQtratamentos é a análise de regressão .

O esquema básico de análise de variância é o do delineamento experimental adotado , somente com a subdivisão da SQtratamentos , em regressão linear , regressão quadrática , ... ;isto é em tantos componentes quantos forem os GL de tratamentos.

A equação de regressão a ser ajustada incluirá todos os componentes até o de mais alto grau significativo , mesmo que no intervalo exista algum componente não significativo .

O procedimento se simplifica quando os níveis de x forem equidistantes , através da utilização de polinômios ortogonais , cujos coeficientes são tabelados .

Exemplo:

Num experimento de adubação potássica de cana-de-açucar utilizou-se 4 níveis de K_2O (0 ; 60 ; 120 e 180 Kg/ha) em DCC com 8 repetições por tratamento .

A análise de variância é a seguinte :

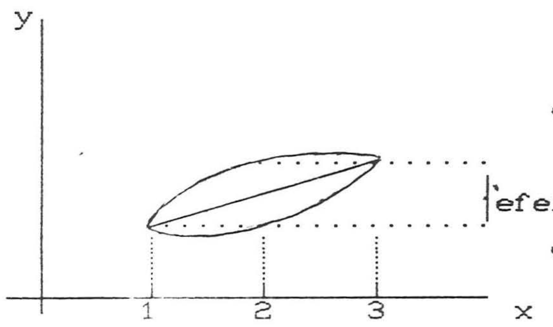
Causas de variação	GL	SQ	QM	F
tratamentos	3	638,50	212,83	3,20*
resíduos	28	1860,60	66,45	

CV = 21,2 %

$F_{.05}(3,28) = 2,96$

GLtratamentos = 3
- componente linear (K') ou reg. linear → 1 GL
- componente quadrático (K'') ou reg. quadrática → 1 GL
- componente cúbico (K''') ou reg. cúbica → 1 GL

Para três níveis a situação seria :



efeito quadrático positivo

efeito linear

efeito quadrático nulo

efeito quadrático negativo

2 GL
- efeito linear : linearidade → 1 GL
- efeito quadrático : curvatura → 1 GL

níveis de x	coeficiente dos componentes linear	quadrático
1	-1	1
2	0	-2
3	1	1

Para o exemplo tem-se

(x) trat.	totais(yi.) de tratametos	coeficientes (cji)			cjiyi.		
		1ºgrau	2ºgrau	3ºgrau	1ºgrau	2ºgrau	3ºgrau
0	302	-3	1	-1	-906	302	-302
60	370	-1	-1	+3	-370	-370	1110
120	400	1	-1	-1	400	-400	-1200
180	348	3	1	+1	1044	348	348
total	1420	0	0	0	168	-120	-44
	K	20	4	20			
	M	2	1	10/3			

K : soma dos quadrados dos coeficientes

M : constante para tornar os coeficientes inteiros

$$SQK' = SQ_{regressão}\ 1º\ grau = \frac{(168)^2}{8 \times 20} = 176,40$$

$$SQK'' = SQ_{regressão}\ 2º\ grau = \frac{(-120)^2}{8 \times 4} = 450,00$$

$$SQK''' = SQ_{regressão}\ 2º\ grau = \frac{(-44)^2}{8 \times 20} = 12,10$$

$$\therefore \quad SQcj = \frac{\left(\sum_i cjiyi.\right)^2}{r\ \underset{\sum_i c_{ji}^2}{K}}$$

C. VARIAÇÃO	GL	SQ	QM	F
K'(reg. linear)	1	176,40	176,40	2,65
K''(reg. quadrática)	1	450,00	450,00	6,77*
K'''(reg. cúbica)	1	12,10	12,10	0,18
Tratamentos	(3)	(638,50)		
Resíduo(erro)	28	1860,60	66,45	

$F.05(1,28) = 4,20$

·A resposta da cana-de-açúcar à adubação potássica é quadrática, indicando que para doses baixas de K_2O aplicadas ao solo , o rendimento da cana apresenta um acréscimo linear ; à medida que doses de K_2O crescem , o acréscimo vai se tornando menor tendendo a estabilizar nas doses mais altas .

COEFICIENTE DE DETERMINAÇÃO

$$R^2 = \frac{\text{SQR linear} + \text{SQR quadrática}}{\text{SQ tratamentos}} = \frac{SQK' + SQK''}{SQK}$$

$$= \frac{176,40 + 450}{638,50}$$

$$= 0,98$$

98% da variação no rendimento de cana-de-açúcar devida a aplicação de K_2O é explicada pela reg. quadrática do rendimento de cana para doses de K_2O aplicadas ao solo .

AJUSTE DA EQUAÇÃO DE REGRESSÃO

$$\hat{y} = \underbrace{\bar{y}}_{\text{média geral}} + B_1\, M_1\, P_1 + B_2\, M_2\, P_2 + \ldots$$

$$B_j = \frac{\sum_i c_{ji}\, y_{i.}}{r\, K}$$

B_j coeficiente de regressão de grau j

P_j polinômio ortogonal de grau j

No exemplo

$$\hat{y} = \bar{y} + B_1\, M_1\, P_1 + B_2\, M_2\, P_2$$

$$\bar{y} = \frac{1420}{32} = 44,375 \qquad M_1 = 2$$

$$M_2 = 1$$

$$B_1 = \frac{168}{8 \times 20} = 1,05$$

$$B_2 = \frac{-120}{8 \times 4} = -3,75$$

$$P_1 = x = \frac{x - \bar{x}}{q} \begin{array}{l} \leftarrow \text{médias dos níveis} \\ \leftarrow \text{diferença entre os níveis} \end{array}$$

$$P_2 = x^2 - \frac{n^2 - 1}{12} = x^2 - 5/4$$; n = nº de níveis , no exemplo $n = 4$

$$\hat{y} = 44,375 + 1,05(2)x - 3,75(1)(x^2 - 5/4)$$

substituindo $x = \dfrac{x - 90}{60} = \dfrac{x}{60} - \dfrac{3}{2}$ obteremos

$$\hat{y} = 37,475 + 0,2225x - 0,001042x^2$$

x	$\bar{y}$	$\hat{y}$	$\bar{y} - \hat{y}$
0	37,75	37,4750	0,2750
60	46,25	47,0738	-0,8238
120	50,00	49,1702	0,8298
180	43,50	43,7642	-0,2642
			0,0168

DOSE ÓTIMA

$$\hat{y} = a + bx + cx^2$$

Ponto de máximo : (i) $\dfrac{d\hat{y}}{dx} = 0 \rightarrow x^*$(ponto crítico)

(ii) $\dfrac{d\hat{y}^2}{dx^2} < 0 \Rightarrow$ ponto de máximo

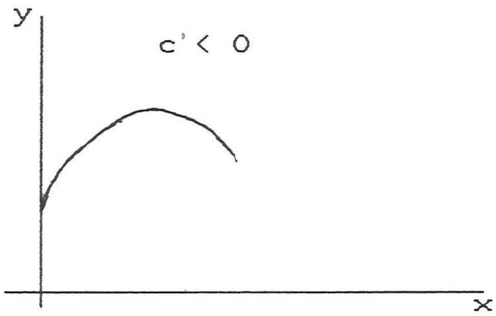
$> 0 \rightarrow$ ponto de mínimo

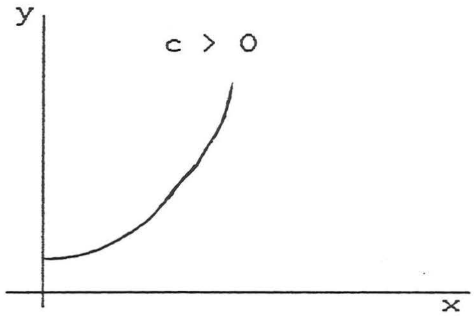
$$\frac{d\hat{y}}{dx} = b + 2cx = 0 \rightarrow x^* = \frac{-b}{2c}$$

$$\frac{d\hat{y}^2}{dx^2} = 2c$$

$\therefore$ $c < 0 \rightarrow$ ponto de máximo

$c > 0 \rightarrow$ ponto de mínimo





No exemplo

$c = -0{,}001042$

$b = 0{,}2225$

$$x^* = \frac{-0{,}2225}{2(-0{,}001042)} = 106{,}76 \cong 107$$

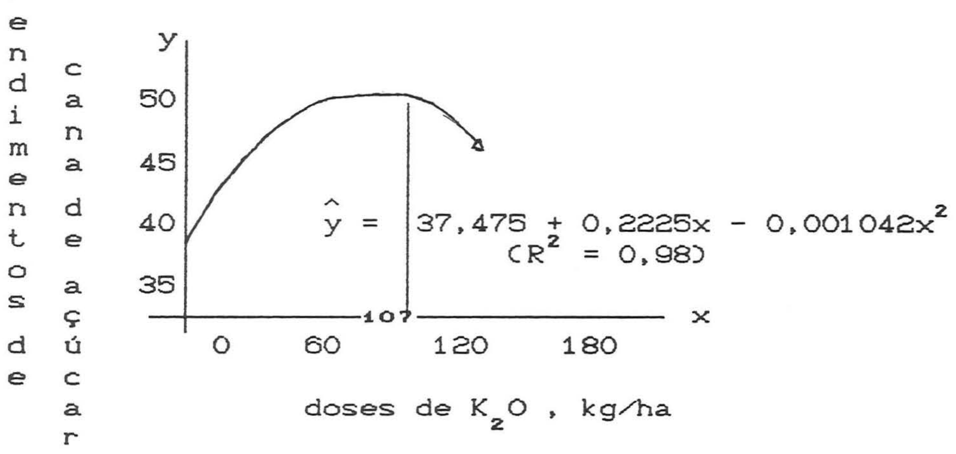
É ponto de máximo pois $c < 0$

$\therefore$ dose ótima 106,76(107) kg/ha de K_2O que nos leva a um rendimento máximo.

(melhor tratamento)

- O ponto determinado é o ponto de máxima eficiência técnica

– Poder-se-ia determinar o ponto de máxima eficiência econômica .



para $x = 107 \Rightarrow \hat{y} = 49,3526$

PONTO DE MAXIMA EFICIÊNCIA ECONÔMICA

Deseja-se tornar máxima a receita líquida

$$L = w\,\hat{y} - u\,x - m$$

- m — custos fixos
- x — dose do nutriente
- u — preço unitário do nutriente
- $\hat{y}$ — estimativa de produção dada pela curva de resposta
- w — preço do produto

Para determinação de máximo , deriva-se e iguala-se a zero :

$$\frac{dL}{dx} = w\,\hat{y}\,' - u = 0 \quad ; \text{ como} \quad \begin{array}{l} \hat{y} = a + bx + cx^2 \\ \hat{y}\,' = b + 2cx \end{array} ,$$

então temos

$$w(b + 2cx) - u = 0 \rightarrow b + 2cx - \frac{u}{w} = 0 \rightarrow 2cx = u/w - b$$

Logo $x^* = \dfrac{u/w - b}{2c}$, x^* depente da relação de preços

dose de máxima eficiência econômica

2.14 CONSIDERAÇÕES TEÓRICAS SOBRE A ANÁLISE DE VARIANCIA DE EXPERIMENTOS COM UM FATOR FIXO EM DELINEAMENTO COMPLETAMENTE CASUALIZADO .

(1) CARACTERIZAÇÃO

Modelo matemático :

observação da j-ésima repetição do tratamento i

constante inerente a todos os dados (média geral)

efeito do tratamento i

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

erro associado a UE ij (não-observável)

aleatório

fixo → modelo fixo ou modelo I

Aleatório→modelo aleatório ou modelo II

fixo

$i = 1,2,\ldots,t$ (índice de tratamento)

$j = 1,2,\ldots,ri$ (índice de repetição)

$$n = \sum_{i=1}^{t} ri = \text{n° total de UE} = \text{n° total de observações.}$$

FATOR FIXO : Escolha deliberada dos níveis do fator , isto é dos tratamentos , a serem incluídos no experimento . Conclusões são válidas para os tratamentos utilizados no experimento .

$$Ho : \mu_1 = \mu_2 = \ldots = \mu_t$$

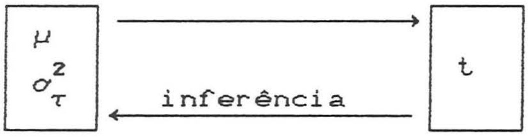
$$\Updownarrow$$

$$Ho : \tau_i = 0 \ , \ \forall\, i$$

FATOR ALEATÓRIO : Os níveis do fator (tratamentos) são escolhidos de forma aleatória de uma população de níveis , constituindo os tratamentos uma amostra aleatória da população de tratamentos . As conclusões são válidas para a população de tratamentos .

$$Ho : \sigma^2_\tau = 0$$

população de tratamentos   amostra de t tratamentos



para a população de tratamentos

σ^2_τ é a variação da população de tratamentos da qual os t tratamentos usados no experimento são amostra .

No caso presente tem-se o modelo linear

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad \text{e considera-se } \tau_i \text{ fixo .}$$

Na forma matricial o modelo linear é expresso por

$$\underset{\sim}{y} = x\underset{\sim}{\theta} + \underset{\sim}{\varepsilon}$$

onde

${}_{n}\underset{\sim}{y}_{1}$ é um vetor de realizações de variáveis aleatórios (vetor de observações)

${}_{n}x_{t+1}$ é uma matriz conhecida , é a matriz dos coeficientes dos parâmetros do modelo (matriz do delineamento)

${}_{t+1}\underset{\sim}{\theta}_{1}$ é o vetor dos parâmetros do modelo

${}_{n}\underset{\sim}{\varepsilon}_{1}$ é um vetor não observável , de erros aleatórios .

(2) PRESSUPOSIÇÕES ASSOCIADAS AO MODELO

$\varepsilon_{ij} \overset{IID}{\cap} N(0, \sigma^2)$ [IID = identicamente e independentemente distribuidos]

$\underset{\sim}{\varepsilon} \cap N(\emptyset, I\sigma^2)$ [modelo linear de Gauss-Markov normal (G.M.N) ou G.M.N. ordinário
Se $\underset{\sim}{\varepsilon} \cap N(\underset{\sim}{\emptyset}, D\sigma^2)$, onde D matriz diagonal $\neq$ I $\Rightarrow$ G.M.N. ponderado
SE $\underset{\sim}{\varepsilon} \cap N(\underset{\sim}{\emptyset}, V\sigma^2)$, V matriz qualquer $\Rightarrow$
$\Rightarrow$ G.M.N. generalizado]

Como y_{ij} são funções lineares de ε_{ij}, das pressuposições sobre os componentes do modelo decorre que :

(i) $\varepsilon(y_{ij}) = \mu + \tau_i = \mu_i$ [média do tratamento i]

$\Leftrightarrow \varepsilon(\underset{\sim}{y}) = x\underset{\sim}{\theta}$

(ii) $V(y_{ij}) = \sigma^2 \Leftrightarrow V(\underset{\sim}{y}) = I\sigma^2$

(iii) $y_{ij} \cap N(\mu_i, \sigma^2)$ e independentes $\Longleftrightarrow \underset{\sim}{y} \cap N(x\underset{\sim}{\theta}, I\sigma^2)$

(3) EXEMPLO

	TRATAMENTOS			
	T_1	T_2	T_3	
i=1,2,3	12	6	5	
	11	7	4	
	10	8	3	
			4	
$y_{i.}$	33	21	16	70=$y_{..}$
$\bar{y}_{i.}$	11	7	4	7=$\bar{y}_{..}$
s_i^2	1	1	2/3	
r_i	3	3	4	10=n

$$
\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \end{bmatrix}
=
\underbrace{\begin{bmatrix} 12 \\ 11 \\ 10 \\ 6 \\ 7 \\ 8 \\ 5 \\ 4 \\ 3 \\ 4 \end{bmatrix}}_{\underset{\sim}{y}}
=
\underbrace{
\begin{array}{c}
\begin{array}{cccc} \mu & \tau_1 & \tau_2 & \tau_3 \end{array} \\
\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}
\end{array}}_{x}
\underbrace{\begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}}_{\underset{\sim}{\theta}}
+
\underbrace{\begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \end{bmatrix}}_{\underset{\sim}{\varepsilon}}
$$

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

$$y_{11} = \mu + \tau_1 + \varepsilon_{11}$$

$$y_{12} = \mu + \tau_1 + \varepsilon_{12}$$

$$y_{13} = \mu + \tau_1 + \varepsilon_{13}$$

$$y_{21} = \mu + \tau_2 + \varepsilon_{21}$$

$$\vdots$$

$$y_{23} = \mu + \tau_2 + \varepsilon_{23}$$

$$y_{31} = \mu + \tau_3 + \varepsilon_{31}$$

$$\vdots$$

$$y_{34} = \mu + \tau_3 + \varepsilon_{34}$$

(4) SISTEMA DE EQUAÇÕES NORMAIS (SEN)

Usa-se o método dos quadrados mínimos

$$\underset{\sim}{y} = \underset{\sim}{x}\,\underset{\sim}{\theta} + \underset{\sim}{\varepsilon} \Rightarrow \underset{\sim}{\varepsilon} = \underset{\sim}{y} - \underset{\sim}{x}\,\underset{\sim}{\theta}$$

Obtém-se a função $z(\theta)$, onde

$$\begin{aligned} z(\theta) = \sum_{ij} \varepsilon_{ij}^2 = \| \varepsilon \|^2 &= \varepsilon'\varepsilon \\ &= (y - x\theta)'(y - x\theta) \\ &= (y' - \theta'x')(y - x\theta) \\ &= y'y - \underbrace{y'x\theta}_{(1xn)(nxt+1)(t+nx1)} - \underbrace{\theta'x'y}_{(1xt+1)(t+1xn)(n+1)} + \theta'x'x\theta \end{aligned}$$

são escalares, sendo 1 o transposto do outro

$$\therefore\ z(\theta) = y'y - 2\theta'x'y + \theta'x'x\theta$$

Buscando um mínimo para $z(\theta)$ temos :

$$\frac{\partial z}{\partial \theta} = \emptyset - 2\partial\theta'x'y + \underbrace{\partial\theta'x'x\theta + \theta'x'x\partial\theta}_{+\,2\partial\theta'x'x\theta}$$

$$= 2\partial\theta'(x'x\theta - x'y)$$

Fazendo-se $\frac{\partial z}{\partial \theta} = \emptyset$, obtem-se o SEN

$$x'x\hat{\theta} = x'y$$

No exemplo

$$\begin{bmatrix} 10 & 3 & 3 & 4 \\ 3 & 3 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 4 & 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\tau}_3 \end{bmatrix} = \begin{bmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ y_{3.} \end{bmatrix}$$

Genericamente o SEN é dado por

$$\begin{bmatrix} n & r_1 & r_2 & \dots & r_t \\ r_1 & r_1 & 0 & & 0 \\ r_2 & 0 & r_2 & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ r_t & 0 & 0 & \dots & r_t \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\tau}_3 \\ \vdots \\ \hat{\tau}_t \end{bmatrix} = \begin{bmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ y_{3.} \\ \vdots \\ y_{t.} \end{bmatrix}$$

(5) SOLUÇÃO DO SISTEMA DE EQUAÇÕES NORMAIS

Sistema de Equações Lineares

- <u>Consistente</u>: tem solução
  - Ex.: $\underset{\sim}{y} = x\underset{\sim}{\theta} + \underset{\sim}{\varepsilon}$

    $\downarrow$

    $x'x\theta = x'y$ é consistente
  - Determinado : Tem solução única. Para $x'x\hat{\theta}=x'y$ se $x'x$ é não singular, isto é, se $\exists (x'x)^{-1}$, então $\hat{\theta} = (x'x)^{-1} x'y$

    ex.: modelos de regressão $x'x$ tem posto completo
  - Indeterminado : A solução não é única. Para $x'x\hat{\theta} = x'y$ se $x'x$ é singular, isto é $\nexists (x'x)^{-1}$, a solução não é única. Ex.: Modelos de planejamento de experimentos; $x'x$ tem posto incompleto
- <u>Inconsistente</u>

  Não tem solução

  Ex.: $\underset{\sim}{y} = x\underset{\sim}{\theta}$

No presente caso $x'x$ tem posto incompleto, é singular, portanto o SEN $x'x\hat{\theta} = x'y$ é indeterminado.

A solução do SEN pode ser feita através de vários procedimentos :

(a) Usando inversas generalizadas
  - → Condicional
  - → Mínimos quadrados
  - → Moore-Penrose

(b) Reparametrizando o modelo

(c) Usando restrições nas soluções

(d) Usando restrições nos parâmetros

Restrições nas soluções

A restrição deve ser uma função paramétrica não-estimável. Geralmente se utiliza a restrição

$$\sum_{i=1}^{t} r_i\hat{\tau}_i = r_1\hat{\tau}_1 + r_2\hat{\tau}_2 + \ldots + r_t\hat{\tau}_t = 0$$

Se $r_1=r_2=\ldots=r_t=r \Rightarrow$ A restrião é $\sum_{i=1}^{t} \hat{\tau}_i = \hat{\tau}_1 + \hat{\tau}_2 + \ldots + \hat{\tau}_t = 0$

Estimabilidade

(i) Definição : Uma função linear paramétrica $\lambda'\theta$ é estimável se e só se $\exists$ ao menos uma combinação linear das observações $c'y$ tal que $E\left[\, c'y \,\right] = \lambda'\theta$.

· No modelo $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$

· $E(y_{ij}) = \mu + \tau_i$ são as funções básicas estimáveis

(1) $\mu + \tau_1$ é estimável

(2) μ não é estimável

(3) τ_1 não é estimável

(4) O contraste $\tau_1 - \tau_2$ é estimável pois

$$(\mu + \tau_1) - (\mu + \tau_2) = \tau_1 - \tau_2$$

(5) O contraste $\tau_1 + \tau_2 - 2\tau_3$ é estimável , pois

$$(\mu + \tau_1) + (\mu + \tau_2) - 2(\mu + \tau_3) = \tau_1 + \tau_2 - 2\tau_3$$

(6) Qualquer contraste de efeitos de tratamentos é estimável

(7) A função paramétrica $3\tau_1 + 3\tau_2 + \tau_3 = r_1\tau_1 + r_2\tau_2 + r_3\tau_3$

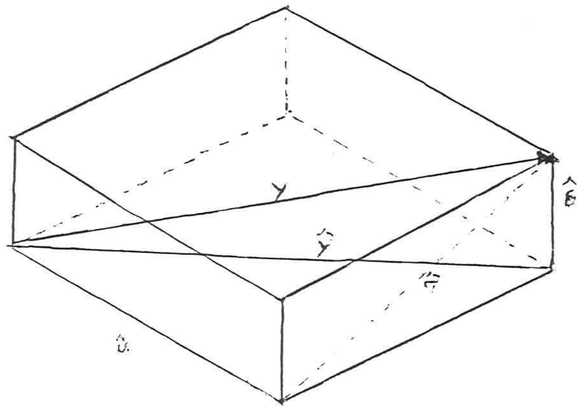
para o exemplo , não é estimável

(ii) <u>Regra prática para verificar estimabilidade</u> : A função linear paramétrica $\lambda'\theta$ é estimável se $H'\lambda = \lambda$ ou $\lambda' H = \lambda'$ onde $H = (x'x)^{-}(x'x)$.

(iii) <u>Teorema de Gauss-Markov</u> : Se $\lambda'\theta$ é estimável então seu BLUE ("Best Linear Umbiased Estimator") ou MELI ("Melhor Estimador Linear Imparcial") é dado por $\lambda'\hat{\theta}$ onde $\hat{\theta}$ é qualquer solução do SEN .

(6) ANÁLISE DE VARIANCIA E TESTE SOBRE TRATAMENTOS

Representação Geométrica



$$\underset{\sim}{y} = \underset{\sim}{x\theta} + \underset{\sim}{\varepsilon}$$

$$\underset{\sim}{\hat{y}} = \underset{\sim}{x\hat{\theta}}$$

$$\underset{\sim}{\hat{\varepsilon}} = \underset{\sim}{y} - \underset{\sim}{\hat{y}}$$

$$\underset{\sim}{\hat{\theta}} = \begin{bmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \vdots \\ \hat{\tau}_t \end{bmatrix}$$

Norma ou comprimento de um vetor

$$\|u\| = \left\{ u'u \right\}^{1/2} = \left\{ \sum_j u_j^2 \right\}^{1/2} = \sqrt{u_1^2 + \ldots + u_n^2}$$

$$\|u\|^2 = u'u = \sum_j u_j^2 = u_1^2 + u_2^2 + \ldots + u_n^2$$

Pitágoras

$$\|y\|^2 = \|\hat{y}\|^2 + \|\hat{\varepsilon}\|^2$$

$$y'y = \hat{y}'\hat{y} + \hat{\varepsilon}'\hat{\varepsilon} \quad ; \text{ Mas } \quad \hat{y} = x\hat{\theta} \quad \text{e} \quad \hat{\varepsilon} = y - \hat{y} = y - x\hat{\theta}$$

$$y'y = \hat{\theta}'x'x\hat{\theta} + (y - x\hat{\theta})'(y - x\hat{\theta}) = \hat{\theta}'x'y + (y - x\hat{\theta})'(y - x\hat{\theta})$$

SEN : $x'x\hat{\theta} = x'y$

$$(y - x\hat{\theta})'(y - x\hat{\theta}) = (y' - \hat{\theta}'x')(y - x\hat{\theta}) = y'y \underbrace{- \hat{\theta}'x'y - y'x\hat{\theta}} + \underbrace{\hat{\theta}'x'x\hat{\theta}}_{x'y}$$

$$= y'y - 2\hat{\theta}'x'y + \hat{\theta}'x'y$$

$$= y'y - \hat{\theta}'x'y$$

$$\therefore\ y'y = \hat{\theta}'x'y + (y'y - \hat{\theta}'x'xy)$$

$$y'y = \sum_{ij} y_{ij}^2 \Big\}\ \text{soma dos quadrados das observações}$$

$$y'y - \hat{\theta}'x'y \Big\}\ \text{soma de quadrados do erro}$$

[SQ devida ao erro de ajustamento pelo modelo escolhido]

$$\hat{\theta}'x'y = \begin{bmatrix} \hat{\mu} & \hat{\tau}_1 & \dots & \hat{\tau}_t \end{bmatrix} \begin{bmatrix} y_{..} \\ y_{1.} \\ \vdots \\ y_{t.} \end{bmatrix} = \hat{\mu}y_{..} + \hat{\tau}_1 y_{1.} + \dots + \hat{\tau}_t y_{t.} \Big\}\ \text{soma de quadrados de parâmetros}$$

[SQ devida aos parâmetros considerados no modelo]

$$\hat{\theta} = (x'x)^- x'y$$

$$y'y = \hat{\theta}'x'y + (y'y - \hat{\theta}'x'y)$$

$$= \underbrace{y'x(x'x)^-x'}_{P}y + (y'y - y'x(x'x)^- x'y)$$

$$= y'Py + y'\left[I - x(x'x)^- x' \right] y$$

$$= y'Py + y'\left[I - P \right] y$$

$P \Big\}$ projetor ortogonal de y no espaço dos parâmetros

$\Big[$ espaço coluna de x ; c(x) $\Big]$

$$\hat{y} = x\hat{\theta} = x(x'x)^{-} x'y = Py$$

$I - P \Big\}$ projetor ortogonal de y no espaço ortogonal ao espaço dos parâmetros , isto é no espaço do erro

$\left[C^{\perp}(x) \right]$

$$\hat{\varepsilon} = y - \hat{y} = y - Py = (I - P)\,y$$

$$\text{SQparâmetros} = SQ(\mu, \underset{\sim}{\tau}) = \underbrace{SQ(\mu)}_{y'P_1 y} + \underbrace{SQ(\underset{\sim}{\tau} / \mu)}_{\text{SQtratamentos}}$$

$SQ(\mu) = ?$

$$x = \left[x_1 \ \vdots \ x_2 \right] = \begin{array}{ccc} \mu & & \underset{\sim}{\tau} \\ \end{array} \begin{bmatrix} 1 & . & \\ 1 & . & x_2 \\ \vdots & . & \\ \vdots & . & \\ 1 & . & \end{bmatrix}$$

$$SQ(\mu) = y'\,P_1\,y = y'\,x_1 \underbrace{(x_1'\,x_1)^{-}}_{1/n} x_1'\,y$$

$$= y'\,\frac{1}{n}\,x_1\,x_1'\,y$$

$$= \frac{1}{n}\,(y_{..})^2 = \text{F C (Fator de Correção)}$$

$$\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ . & . & \dots & . \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

$$= E\,(n)$$

$$y'\,E\,(n) = \begin{bmatrix} y_{..} & y_{..} & \dots & y_{..} \end{bmatrix}$$

$$SQparâmetros = \hat{\theta}'x'y = \begin{bmatrix} 0 & \bar{y}_{1.} & \bar{y}_{2.} & \dots & \bar{y}_{t.} \end{bmatrix} \begin{bmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ \vdots \\ y_{t.} \end{bmatrix}$$

$$= \sum_{i=1}^{t} \bar{y}_{i.} \; y_{i.}$$

$$= \sum_{i=1}^{t} \frac{y_{1.}}{r_i} \; y_{1.} \; = \; \sum_{i=1}^{t} \frac{(y_{i.})^2}{r_i}$$

SQ

$SQTratamentos = SQParâmetros - SQ\mu$

$$= \sum_{i=1}^{t} \frac{(y_{ij})^2}{r_i} \; - \; FC$$

$SQTotal = y'y - SQ(\mu) = \sum_{ij} y_{ij}^2 - FC$

$SQTotal = y'(I-P_1)y = y'y - y'P_1y$

$SQErro = SQTotal - SQTratamentos$

$SQErro = y'(I-P)y$

$SQParâmetros = y'Py$

$SQTratamentos = y'Py - y'P_1y$

$= y'(P-P_1)y$

onde $P = x(x'x)^-x'$

$P_1 = x_1(x_1'x_1)^-x_1' = \frac{1}{n}x_1x_1' = \frac{1}{n}\ E(n)$

GL

O número de GL de uma soma de quadrados é o posto (Característica ou Rank) da matriz núcleo da forma quadrática correspondente

$$GL\left[SQP \right] = r\left[\overbrace{x(x'x)^{-}x'}^{P} \right] = t$$

$$GL\left[SQErro \right] = r\left[I - P \right] = r(I) - r(P) = n - t$$

$$GL\left[SQTratamentos \right] = r\left[P - P_1 \right] = r(P) - r(P_1) = t - 1$$

$$GL\left[SQTotal \right] = r\left[I - P_1 \right] = r(I) - r(P_1) = n - 1$$

FORMAS QUADRATICAS

$$Q(\tilde{x}) = \tilde{x}'\, \underbrace{A}_{\text{matriz núcleo da forma quadrática}}\, \tilde{x}$$

Exemplos : (i) $SQ(\mu) = y'P_1y = \dfrac{y^2_{..}}{n}$

(ii) $SQTratamentos = SQ(\tilde{\tau} / \mu) = y'(P - P_1)y$

$$= \underbrace{y'Py}_{} - \underbrace{y'P_1y}_{}$$

$$P = x(x'x)^{-}x' \qquad = \sum_{i=1}^{t} \frac{y^2_{i.}}{r_i} - \frac{y^2_{..}}{n}$$

T1	T2	T3
x	x	x
x	x	x
x	x	x
		x

$$x = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} ;$$

$$(X'X)^- = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/4 \end{bmatrix} \quad ; \quad X'X = \begin{bmatrix} 10 & 3 & 3 & 4 \\ 3 & 3 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 4 & 0 & 0 & 4 \end{bmatrix}$$

$$P = X(X'X)X' = \left[\begin{array}{ccc:ccc:cccc} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 \end{array}\right]$$

$r(P) = 3$

GENERICAMENTE

$$P = \begin{bmatrix} 1/r_1 & 1/r_1 & \dots & 1/r_1 & & & & \\ 1/r_1 & 1/r_1 & \dots & 1/r_1 & & & \emptyset & \\ \cdot & \cdot & \cdot & \cdot & & & & \\ 1/r_1 & 1/r_1 & \dots & 1/r_1 & & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & & 1/r_t & 1/r_t & \dots & 1/r_t \\ & \emptyset & & & 1/r_t & 1/r_t & \dots & 1/r_t \\ & & & & \cdot & \cdot & \cdot & \cdot \\ & & & & 1/r_t & 1/r_t & \dots & 1/r_t \end{bmatrix}$$

$r(P) = t$

$$y'\,P\,y = \begin{bmatrix} y_{11} \cdots y_{trt} \end{bmatrix} P \begin{bmatrix} y_{11} \\ \vdots \\ y_{trt} \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{y_{1.}}{r_1} & \frac{y_{1.}}{r_1} & \cdots & \frac{y_{1.}}{r_1} & \cdots & \frac{y_{t.}}{r_t} & \frac{y_{t.}}{r_t} & \cdots & \frac{y_{t.}}{r_t} \end{bmatrix} \begin{bmatrix} y_{11} \\ \vdots \\ y_{trt} \end{bmatrix}$$

$$= \frac{y_{1.}}{r_1} \underbrace{(y_{11} + \ldots + y_{1r1})}_{y_{1.}} + \ldots + \frac{y_{t.}}{r_t} \underbrace{(y_{t.} + \ldots + y_{trt})}_{y_{t.}}$$

$$= \frac{y_{1.}^2}{r_1} + \ldots + \frac{y_{t.}^2}{r_t} = \sum_{i=1}^{t} \frac{y_{i.}^2}{r_t}$$

ou

$$SQP = y'\,P\,y = y'x(x'x)^{-}x'y$$

$$= \begin{bmatrix} y_{..} & y_{1.} & \cdots & y_{t.} \end{bmatrix} \begin{bmatrix} 0 & 0 & \ldots & 0 \\ 0 & 1/r_1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 1/r_t \end{bmatrix} \begin{bmatrix} y_{..} \\ y_{1.} \\ \vdots \\ y_{t.} \end{bmatrix} =$$

$$= \begin{bmatrix} 0 & \frac{y_{1.}}{r_1} & \cdots & \frac{y_{t.}}{r_t} \end{bmatrix} \begin{bmatrix} y_{..} \\ y_{1.} \\ \vdots \\ y_{t.} \end{bmatrix} = \frac{y_{1.}^2}{r_1} + \ldots + \frac{y_{t.}^2}{r_t}$$

ESPERANÇA DE FORMAS QUADRATICAS

Método longo : usando somatórios

$$E(SQT) = E\left[\sum_{i=1}^{t} \frac{y_{i.}^2}{ri} - \frac{y_{..}^2}{n}\right] = E\left[\sum_{i=1}^{t} \frac{y_{i.}^2}{r_i}\right] - E\left[\frac{y_{..}^2}{n}\right]$$

Método curto :

$$E(SQT) = E\left[\, y'(P - P_1)y \,\right] = E\left[\, y'Py \,\right] - E\left[\, y'P_1y \,\right]$$

TEOREMA

Seja $\underset{\sim}{y} \cap N\left[\, x\theta \,;\, I\sigma^2 \,\right]$ [vale também sob não-normalidade]

então $E\left[\underset{\sim}{y}' A \underset{\sim}{y}\right] = \sigma^2 Tr(A) + \underset{\sim}{\theta}' x' A\, x \underset{\sim}{\theta}$ onde A é a matriz núcleo da forma quadrática.

$$E\left[\, SQT \,\right] = E\left[\, y'(P - P_1)y \right] = \sigma^2\, Tr(P - P_1) + \theta' x'(P - P_1)x\theta$$

$$Tr(P - P_1) = Tr(P) - Tr(P_1) = t - 1$$

$$\theta' x'(P - P_1)x\, \theta = \sum_{i=1}^{t} r_i\, (\mu_i - \mu)^2 = \sum_{i=1}^{t} r_i \tau_i^2$$

$$y_{ij} = \underbrace{\mu + \tau_i}_{\mu_i} + \varepsilon_{ij} \quad ; \quad \mu_i = \mu + \tau_i \quad ; \quad \tau_i = \mu_i - \mu$$

$$\therefore \quad E\left[\, SQT \,\right] = (t - 1)\, \sigma^2 + \sum_{i=1}^{t} r_i \tau_i^2$$

$$E\left[SQE\right] = E\left[y'(I - P)y\right] = Tr(I - P)\sigma^2 + \theta' x'(I - P)x\theta$$

$$Tr(I - P) = Tr(I) - Tr(P) = n - t$$

$$\theta' x'(I - P)\, x\, \theta = \theta' x' x\, \theta - \theta' x' P\, x\, \theta$$

$$\theta' x'\; P\, x\, \theta = \theta' x' \underbrace{x(x'x)^{-}\, x'}_{P} x\, \theta$$

Mas da definição de inversa condicional ,$(x'x)^-$ é invers condicional de $(x'x)$ se $(x'x)^-(x'x)\,(x'x) = x'x$; ainda s $x'x$ é simétrica além da condição da definição tem-se qu $(x'x)^-\; x'x\; (x'x)^- = (x'x)^-$.

$$\therefore\; \theta' x' P\, x\, \theta = \theta' \underbrace{x' x\, (x'x)^-\; x' x}\, \theta = \theta' x' x\, \theta$$

$$e\; \theta' x'(I - P) x\, \theta = \theta' x' x\, \theta - \theta' x' x\, \theta = \emptyset$$

$$\therefore\; E\left[SQE\right] = (n - t)\,\sigma^2$$

$$GLT = r\left[P - P_1\right] = t - 1 \qquad QMT = \frac{SQT}{GLT} = \frac{SQT}{t-1}$$

$$GLE = r\left[I - P\right] = n - t \qquad QME = \frac{SQE}{CLE} = \frac{SQE}{n-t}$$

$$E\left[QMT\right] = \frac{1}{t - 1}\quad E\left[SQT\right] = \frac{(t-1)\sigma^2}{(t-1)} + \frac{1}{t-1}\sum_{i=1}^{t} r_i \tau_i^2 = \sigma^2 + \frac{1}{t-1}\sum_{i=1}^{t} r_i \tau_i^2$$

$$E\left[QME\right] = \frac{1}{n-t}\quad E\left[SQE\right] = \frac{1}{(n-t)}\,(n-t)\,\sigma^2 = \sigma^2$$

DISTRIBUIÇÃO DAS FORMAS QUADRÁTICAS

TEOREMA DE FISHER-COCHRAN

Seja $\underset{\sim}{y} \cap N(\underset{\sim}{\mu}\ ,\ I\sigma^2)$, uma condição necessária e suficiente para

$$\frac{y' A_i y}{\sigma^2} \cap \chi^2 \left\{ r \left[A_i \right] ; \delta_i \right\}$$ onde δ_i é o parâmetro

de não-centralidade e $\delta_i = \dfrac{1}{2\sigma^2}\ \mu' A_i \mu$ e para que $y' A_i y$ e $y' A_j y$;

$i \neq j$; $i,j = 1,2,\ldots,p$; sejam independentes é que

$$r \left[\sum_{i=1}^{p} A_i \right] = \sum_{i=1}^{p} r \left[A_i \right]$$

$$\underset{\sim}{y} = x\underset{\sim}{\theta} + \underset{\sim}{\varepsilon}$$

$$SQT = y'(P - P_1)y$$

$$SQE = y'(I - P)y$$

$$\underset{\sim}{y} \cap N(x\underset{\sim}{\theta}\ ;\ I\sigma^2)$$

$$(P - P_1) + (I - P) = I - P_1$$

$$r \left[I - P_1 \right] = r \left[I \right] - r \left[P_1 \right] = n - 1$$

$$\therefore$$

$$r \left[P - P_1 \right] = t - 1 \quad ; \quad r \left[I - P \right] = n - t$$

$$r \left[P - P_1 \right] + r \left[I - P \right] = t - 1 + n - t = n - 1$$

$$\therefore \quad r \left[I - P_1 \right] = r \left[P - P_1 \right] + r \left[I - P \right] = n - 1$$

Logo $$\frac{y'(P - P_1)y}{\sigma^2} = \frac{SQT}{\sigma^2} \cap \chi^2 \left[(t-1) ; \delta_T \right]$$

onde $$\delta_T = \frac{1}{2\sigma^2}\ \theta' x'(P - P_1)x\ \theta = \frac{1}{2\sigma^2} \sum_{i=1}^{t} r_i \tau_i^2$$

$$\frac{y'(I - P)y}{\sigma^2} = \frac{SQE}{\sigma^2} \cap \chi^2(n - t)$$

e independentes

pois $\delta_E = \frac{1}{2\sigma^2} \theta' x'(I - P)x\, \theta = \emptyset$

DEFINIÇÃO

Se $u \cap \chi^2[p, \delta]$ e $v \cap \chi^2[s]$ e independentes então

$$\frac{u/p}{v/s} \cap F\left[p, s, \delta\right]$$

$$\frac{SQT}{\sigma^2} \cap x^2\left[(t-1), \delta\right] ; \frac{SQE}{\sigma^2} \cap x^2 (n-t) \quad \text{e independentes}$$

$$\therefore \quad \frac{\frac{SQT/\sigma^2}{t-1}}{\frac{SQE/\sigma^2}{n-t}} = \frac{\frac{SQT}{t-1}}{\frac{SQE}{n-t}} = \frac{QMT}{QME} \cap F\left[(t-1), (n-t), \delta\right]$$

onde $\delta = \frac{1}{2\sigma^2} \sum_{i=1}^{t} r_i \tau_i^2$

Sob Ho: $\tau_i = 0 \Leftrightarrow$ Ho: $\mu_1 = \mu_2 = \ldots = \mu_t$

$$\frac{QMT}{QME} \cap F\left[(t-1), (n-t)\right]$$

ANÁLISE DE CARIANCIA

causas de variação	GL	SQ	QM	E(QM)	F
tratamentos	$t-1$	SQT	QMT	$\sigma^2 + \frac{1}{t-1} \sum_{i=1}^{t} r_i \tau_i^2$	$\frac{QMT}{QME}$
erro	$n-t$	SQE	QME	σ^2	
Total	$n-1$	SQtotal			

(7) ESTIMAÇÃO POR PONTO E POR INTERVALO

(a) Estimação por ponto

Teorema de Gauss-Markov : Se $\lambda'\theta$ é estimável então seu Blue o Meli é dado de modo único por $\widehat{\lambda'\theta} = \lambda'\hat{\theta}$ onde $\hat{\theta}$ é qualque solução do sistema de equações normais .

Propriedades

(1) $E(\lambda'\hat{\theta}) = \lambda' E(\theta) = \lambda' E\left[(x'x)^{-}x'y\right] = \lambda'(x'x)^{-}x'E(y)$

$\underset{\sim}{y} \cap N(x\underset{\sim}{\theta}, I\sigma^2) \qquad \hat{\theta} = (x'x)^{-}x'y$

$v(\underset{\sim}{y}) = I\sigma^2$

$$= \lambda' \underbrace{(x'x)^{-}x'x}_{H}\theta$$

$$= \underbrace{\lambda' H}_{\lambda'}\theta = \lambda'\theta$$

(2) $v(\widehat{\lambda'\theta}) = ?$

Regra : $v(Ay) = A\,v(y)A'$

$$v(\widehat{\lambda'\theta}) = v(\lambda'\hat{\theta}) = v\left[\lambda'(x'x)^{-}x'y\right] = \lambda'(x'x)^{-}x'\underbrace{v(y)}_{I\sigma^2}x(x'x)^{-}\lambda$$

$$= \lambda'\underbrace{(x'x)^{-}x'x(x'x)^{-}}_{(x'x)^{-}}\lambda\,\sigma^2 = \lambda'(x'x)^{-}\lambda\,\sigma^2$$

$$\therefore \quad \hat{v}(\widehat{\lambda'\theta}) = \hat{v}(\lambda'\hat{\theta}) = \lambda'(x'x)^{-}\lambda \text{ QME}$$

(3) $\lambda'\hat{\theta} \cap N\left[\lambda'\theta \,;\, \lambda'(x'x)^{-}\lambda\,\sigma^2\right]$

De modo geral se $\lambda'\theta = \sum_{i=1}^{t} c_i \tau_i$, isto é se $\lambda'\theta$ é um contraste ,

então $\widehat{\lambda'\theta} = \lambda'\hat{\theta} = \sum_{i=1}^{t} c_i\hat{\tau}_i = \sum_{i=1}^{t} c_i\bar{y}_{i.}$ e

$$\hat{v}(\widehat{\lambda'\theta}) = \hat{v}(\lambda'\hat{\theta}) = \sum_{i=1}^{t}\frac{c_i^2}{r_i}\text{ QME} = \left(\frac{c_1^2}{r_1} + \frac{c_2^2}{r_2} + \ldots + \frac{c_t^2}{r_t}\right)\text{QME}$$

(b) Estimação por intervalo

$$\frac{SQE}{\sigma^2} \cap \chi^2(n-t) \Rightarrow \frac{(n-t)\ QME}{\sigma^2} \cap \chi^2(n-t)$$

$$\lambda'\hat{\theta} \cap N(\lambda'\theta\ ;\ \lambda'(x'x)^{-}\lambda\ \sigma^2)$$

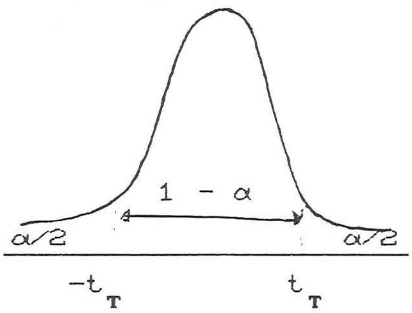
$$z = \frac{\lambda'\hat{\theta} - \lambda'\theta}{\sqrt{\lambda'(x'x)^{-}\lambda\ \sigma^2}} = \frac{\lambda'\hat{\theta} - \lambda'\theta}{\sqrt{\lambda'(x'x)^{-}\lambda\ \sigma^2}} \cap N(0,1)$$

$$\left.\begin{array}{l} u \cap N(0,1) \\ v \cap \chi^2(n) \end{array}\right\} \longrightarrow \frac{u}{\sqrt{v/n}} \cap t(n)$$

$$\frac{\dfrac{\lambda'\hat{\theta} - \lambda'\theta}{\sqrt{\lambda'(x'x)^{-}\lambda\ \sigma^2}}}{\sqrt{\dfrac{\dfrac{(n-t)QME}{\sigma^2}}{(n-t)}}} = \frac{\lambda'\hat{\theta} - \lambda'\theta}{\sqrt{\left[\lambda'(x'x)^{-}\lambda\right] QME}} =$$

$$\therefore$$

$$\frac{\lambda'\hat{\theta} - \lambda'\theta}{\sqrt{\hat{v}(\lambda'\hat{\theta})}} \cap t(n-t)$$



$$t_T = t_{\alpha}\ (n-t)$$

$$P\left\{-t_T \le t \le t_T\right\} = 1 - \alpha$$

$$p\left\{-t_T \le \frac{\lambda'\hat{\theta} - \lambda'\theta}{\sqrt{\hat{v}(\lambda'\hat{\theta})}} \le t_T\right\} = 1 - \alpha$$

$$\therefore\ IC\left[\lambda'\theta\right]100\ (1 - \alpha)\% = \lambda'\hat{\theta} \pm t_{\alpha}\ (n-t)\sqrt{\hat{v}(\lambda'\hat{\theta})}$$

CRITÉRIO DE TUKEY : Só para contrastes que envolvem duas médias

$$IC \left[\tau_i - \tau_i' \right] 100\,(1 - \alpha)\% = \hat{c} \pm \Delta$$

$$\text{onde } \Delta = q\,(\alpha, t, n-t) \sqrt{\frac{1}{2}\ \hat{v}(\hat{c})}$$

$$\hat{c} = \hat{\tau}_i - \hat{\tau}_i'$$

$$\hat{v}(\hat{c}) = \left(\frac{1}{r_i} + \frac{1}{r_i'} \right) QME$$

(8) Soma de quadrados devida à hipótese $H_0 : B'\theta = \emptyset$

$$\text{A forma quadrática } (B'\hat{\theta})' \left[B'(x'x)^{-} B \right]^{-1} (B'\hat{\theta})$$

é definida como soma de quadrados devida à hipótese $H_0 : B'\theta = \emptyset$, onde B' é uma matriz de posto linha completo e que define a hipótese H_0.

(9) Regiões de confiança

Define-se região de confiança para um conjunto de p funções estimáveis, linearmente independentes, $B'\emptyset$, ao nível de confiança $1 - \alpha$, como a região delimitada pelo elipsóide de centro $B'\hat{\emptyset}$ e

$$(B'\hat{\theta} - B'\theta)' \left[B'(x'x)^{-} B \right]^{-1} (B'\hat{\theta} - B'\theta) \leq p\, \hat{\sigma}^2\, F_{\alpha}\left[p, n-t\right]$$

$$\text{onde } \hat{\sigma}^2 = QME$$

Para $p = 2$ tem-se uma elipse.

$$\begin{array}{c} \lambda_1' \\ \lambda_2' \end{array} \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & -2 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\tau}_3 \end{bmatrix}$$

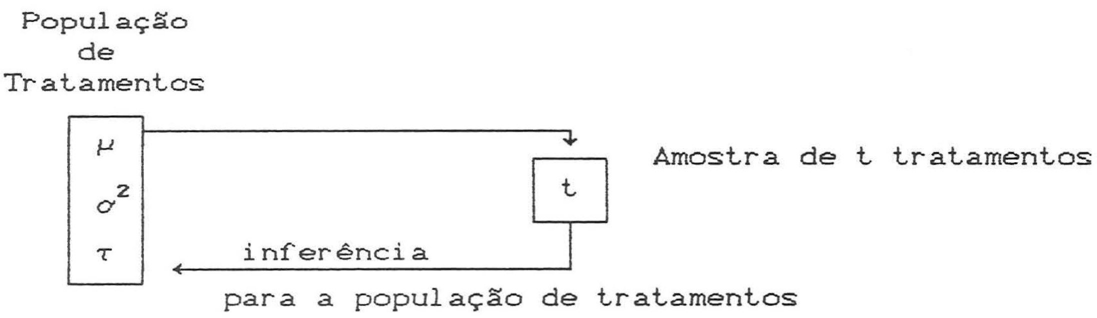
2.15. ANÁLISE DE VARIÂNCIA DE EXPERIMENTOS COM 1 FATOR ALEATÓRIO EM DELINEAMENTO COMPLETAMENTE CASUALIZADO

(a) Caracterização:

$$y_{ij} = \underbrace{\mu}_{\text{fixo}} + \tau_i + \underbrace{\varepsilon_{ij}}_{\text{aleatória}}$$

Aleatório ⇒ modelo aleatório
modelo II

Fator aleatório: Os níveis de fator, isto é os tratamentos, são escolhidos de forma aleatória de uma população de níveis, constituindo os tratamentos uma amostra aleatória da população de tratamentos. As conclusões são válidas para a população de tratamentos.



Ho: $\sigma^2_\tau = 0$

↑ Variância da população de tratamentos da qual os t tratamentos usados no experimento são uma amostra aleatória.

(b) Modelo matemático:

$$\boxed{y_{ij} = \mu + \tau_i + \varepsilon_{ij}}$$

$i = 1, 2, \ldots, t$ (índice de tratamento)

$j = 1, 2, \ldots, r_i$ (índice de repetição)

$r_1 + r_2 + \ldots + r_t = n =$ nº total de observações

(c) Suposições do modelo:

(i) $\varepsilon_{ij} \overset{IID}{\cap} N(0,\sigma^2)$

(ii) $\tau_i \overset{IID}{\cap} N(0,\sigma^2)$

(iii) ε_{ij} e τ_i são independentes

(d) Decorrências das suposições do modelo:

Modelo Aleatório	Modelo Fixo
(i) $E(y_{ij}) = \mu$	$E(y_{ij}) = \mu + \tau_i = \mu_i$
(ii) $V(y_{ij}) = \sigma^2_\tau + \sigma^2$ } Variância Total	$V(y_{ij}) = \sigma^2$

Variância dentro dos tratamentos (variabilidade interna dos dos tratamentos). É constante para todos os tratamentos.

Variância entre os tratamentos (variabilidade de μ_i ao redor de μ)

(iii) $\text{cov}(y_{ij}\,,\ y_{ij'}) = E\,[(y_{ij} - \mu)(y_{ij'} - \mu)]$

$j \neq j'$

$$= E\,[(\tau_i + \varepsilon_{ij})(\tau_i + \varepsilon_{ij'})]$$

$$= E\,[\tau^2 + \tau_i\ \varepsilon_{ij} + \tau_i\ \varepsilon_{ij'} + \varepsilon_{ij}\ \varepsilon_{ij'}]$$

$$= \sigma^2_\tau$$

$\text{cov}(y_{ij}\,,\ y_{i'j'}) = 0$

[Pelo fato de sortear os tratamentos e as amostras dentro dos tratamentos tem-se covariância entre observações de um mesmo tratamento]

(iv) $\text{cov}(y_{ij}\ ,\ y_{i'j}) = E\,[(y_{ij} - \mu)(y_{i'j} - \mu)]$

$i \neq i'$ $\qquad = E\,[(\tau_i + \varepsilon_{ij})(\tau_{i'} + \varepsilon_{i'j})]$

$\qquad = E\,[\tau_i\tau_{i'} + \tau_i\ \varepsilon_{i'j} + \tau_{i'}\ \varepsilon_{ij} + \varepsilon_{ij}\ \varepsilon_{i'j}]$

$\qquad = 0$

$\text{cov}(y_{ij}\ ,\ y_{i'j}) = 0$

(v) $y_{ij} \ \cap \ N(\mu, \sigma^2_\tau + \sigma^2)$ $\qquad\qquad$ $y_{ij} \ \cap \ N(\mu + \tau_i\ ,\ \sigma^2)$

(e) Hipóteses:

Ho: $\sigma^2_\tau = 0$ $\qquad$ [As médias de todos os possíveis níveis do fator em estudo, isto é, de todos os níveis da população de níveis, são iguais a μ]

Ha: $\sigma^2_\tau \neq 0$

(f) Análise de Variância:

Causas de Variação	GL	SQ	QM	F	E(QM)
Tratamentos	t-1	SQT	QMT	$\frac{QMT}{QME}$	$\sigma^2 + \varphi\ \sigma^2_\tau$
Erro	n-t	SQE	QME		σ^2
Total	n-1				

$$\varphi = \frac{1}{t-1}\left[n - \frac{\sum_{i=1}^{t} r_i^{\,2}}{n} \right]$$

Se $r_1 = r_2 = \ldots = r_t = r \ \Rightarrow \ n = rt$, então

$$\varphi = \frac{1}{t-1}\left[r - \frac{r^2 + r^2 + \ldots + r^2}{r} \right] = \frac{1}{t-1}\left[rt - \frac{r^2 t}{rt} \right]$$

$$= \frac{1}{t-1}\left[r(t-1) \right] = r$$

$\therefore \quad E(QMT) = \sigma^2 + r\ \sigma^2_\tau$

$$\frac{QMT}{QME} \overset{\text{Sob Ho}}{\cap} F\,[(t-1),(n-t)]$$

$$\text{Rejeita-se Ho se} \quad F = \frac{QMT}{QME} > F_{\alpha}\,[(t-1),(n-t)]$$

(g) Exemplo:

Um pesquisador está interessado em avaliar se a temperatura média do corpo de animais de uma mesma espécie é constante. Como é impossível a realização de um experimento com todos os animais da espécie, ele sorteou 5 animais e fez 4 medidas de suas temperaturas de forma completamente casualizada. Os dados obtidos foram os seguintes:

	Animais					
	1	2	3	4	5	
	26	23	25	28	30	
	28	20	28	27	32	
	25	24	24	29	28	
	29	22	27	31	31	
Total	108	89	104	115	121	537
Média	27	22,25	26	28,75	30,25	26,85
r	4	4	4	4	4	20

Causas de Variação	GL	SQ	QM	F
Entre animais	4	148,30	37,08	12,03
Erro Experimental	15	46,25	3,08	
Total	19	194,55		

Ho: $\sigma^2_\tau = 0$ vs Ha: $\sigma^2_\tau \neq 0$

$F = 12,03 > F_{.01(4,15)} = 4,89 \Rightarrow$ Rejeita-se Ho

Conclusão: A temperatura média corporal não é constante para todos os animais da espécie estudada, isto é a temperatura média corporal varia de animal a animal da espécie estudada.

(h) Estimação no modelo aleatório:

Consideremos $r_1 = r_2 = \ldots = r_t = r$

(1) Estimação de μ

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

$$\hat{\mu} = \bar{y}_{..} = \frac{y_{..}}{rt}$$

$$y_{..} = \sum_{i=1}^{t} \sum_{j=1}^{r} y_{ij} = rt\mu + r \sum_{i=1}^{t} \tau_i + \sum_{i=1}^{t} \sum_{j=1}^{r} \varepsilon_{ij}$$

$$\bar{y}_{..} = \frac{y_{..}}{rt} = \mu + \frac{1}{t} \sum_{i} \tau_i + \frac{1}{rt} \sum_{i} \sum_{j} \varepsilon_{ij}$$

$$= \mu + \frac{1}{t} \underbrace{[\tau_1 + \tau_2 + \ldots + \tau_t]}_{t \text{ termos independentes}} + \frac{1}{rt} \underbrace{[\varepsilon_{11} + \ldots + \varepsilon_{tr}]}_{rt \text{ termos independentes}}$$

$E(\bar{y}_{..}) = \mu \quad \Rightarrow \quad \bar{y}_{..}$ é estimador imparcial

$$\text{Var}(\bar{y}_{..}) = \frac{1}{t^2} \left[\sigma^2_\tau + \ldots + \sigma^2_\tau\right] + \frac{1}{r^2 t^2} \left[\sigma^2 + \ldots + \sigma^2\right]$$

$$= \frac{1}{t^2} t\sigma^2_\tau + \frac{1}{r^2 t^2} r + \sigma^2 = \frac{1}{t} \sigma^2_\tau + \frac{1}{rt} \sigma^2$$

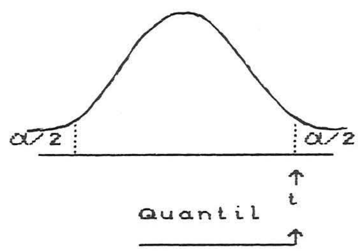
$$\therefore \text{Var}(\bar{y}_{..}) = \frac{1}{rt} (\sigma^2 + r\sigma^2_\tau)$$

$$E(QMT) = \sigma^2 + r\sigma^2_\tau \quad \Rightarrow \quad \hat{\sigma}^2 + r\hat{\sigma}^2_\tau = QMT$$

$$\therefore \hat{\text{Var}}(\bar{y}_{..}) = \frac{1}{rt} QMT \;; \text{ pois } \hat{\text{Var}}(\bar{y}_{..}) = \frac{1}{rt} (\hat{\sigma}^2 + r\hat{\sigma}^2_\tau)$$

$$\frac{\bar{y}_{..} - \mu}{\sqrt{\hat{V}(\bar{y}_{..})}} = \frac{\bar{y}_{..} - \mu}{\sqrt{\dfrac{QMT}{rt}}} \cap t_{(t-1)}$$

$$\text{IC } 100(1-\alpha)\% \text{ para } \mu = y_{..} \pm \underset{\substack{\uparrow \\ t_{\alpha/2}}}{t_{\alpha\,(t-1)}} \sqrt{\frac{QMT}{rt}}$$



No exemplo

$$\bar{y}_{..} = 26{,}85 \qquad \sqrt{\hat{V}(\bar{y}_{..})} = \sqrt{\frac{QMT}{rt}} = \sqrt{\frac{37{,}08}{20}} = 1{,}36$$

$$\begin{aligned} \text{IC } 95\% \text{ p/ } \mu &= 26{,}85 \pm t_{.05(4)}(1{,}36) \\ &= 26{,}95 \pm 2{,}776(1{,}36) \\ &= 26{,}95 \pm 3{,}78 \begin{cases} \rightarrow 30{,}63 \\ \rightarrow 23{,}07 \end{cases} \end{aligned}$$

Estima-se com confiabilidade de 95% que a temperatura média corporal da espécie em estudo esteja no intervalo [23,07;30,63]

(2) Estimação de σ^2: $E(QME) = \sigma^2$

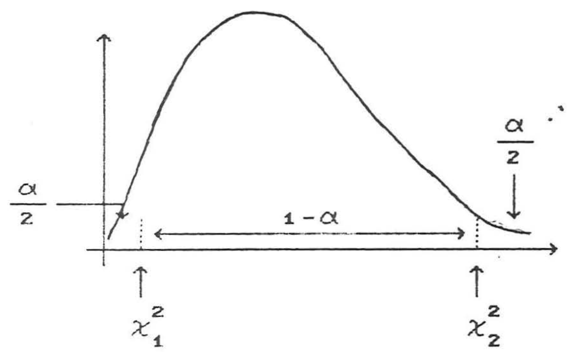
$\hat{\sigma}^2 = QME$ $\left[\begin{array}{c}\text{Estimação} \\ \text{por ponto}\end{array}\right]$ $\left.\begin{array}{c}\sigma^2 \\ \text{e} \\ \sigma^2_\tau\end{array}\right\}$ Componentes de Variância

↑

(Método dos Momentos ou da Análise de Variância)

→ Estimação por intervalo

$$\frac{(n-t)QME}{\sigma^2} \cap \chi^2(n-t)$$



$$P\left[\chi^2_1 < \chi^2 < \chi^2_2\right] = 1-\alpha$$

$$P\left[\chi^2_1 < \frac{(n-t)QME}{\sigma^2} < \chi^2_2\right] = 1-\alpha$$

$$P\left[\frac{1}{\chi^2_1} > \frac{\sigma^2}{(n-t)QME} > \frac{1}{\chi^2_2}\right] = 1-\alpha$$

$$P\left[\frac{(n-t)QME}{\chi^2_1} > \sigma^2 > \frac{(n-t)QME}{\chi^2_2}\right] = 1-\alpha$$

$$\therefore \quad P\left[\frac{(n-t)QME}{\chi^2_2} < \sigma^2 < \frac{(n-t)QME}{\chi^2_1}\right] = 1-\alpha$$

$$\therefore \quad \text{IC } 100(1-\alpha)\% \text{ p/ } \sigma^2 = \left[L_I \,;\, L_S\right]$$

Para o exemplo:

$\hat{\sigma}^2 = QME = 3,08$

IC 95% p/ $\sigma^2 = \left[L_I \; , \; L_S\right]$

$$L_I = \frac{(n-t)QME}{\chi_2^2} = \frac{SQE}{\chi_2^2} \quad ; \quad L_S = \frac{(n-t)QME}{\chi_1^2} = \frac{SQE}{\chi_1^2}$$

$$\chi_2^2 = \chi^2_{.025(n-t)} = \chi^2_{.025(15)} = 27,5$$

$$\chi_1^2 = \chi^2_{.975(n-t)} = \chi^2_{.975(15)} = 6,26$$

$$L_I = \frac{46,25}{27,5} = 1,68 \quad ; \quad L_S = \frac{46,25}{6,26} = 7,38$$

IC 95% p/ $\sigma^2 = [1,68 \; ; \; 7,38]$

(3) Estimação de σ_τ^2

→ Estimação por ponto: pelo Método da Análise de Variância ou dos Momentos.

$$E(QMT) = \sigma^2 + r\,\sigma_\tau^2$$

$$\hat{\sigma}^2 + r\,\hat{\sigma}_\tau^2 = QMT \quad \Rightarrow \quad QME + r\,\hat{\sigma}_\tau^2 = QMT$$

$$\Rightarrow \quad \therefore\; r\,\hat{\sigma}_\tau^2 = QMT - QME$$

$$\Rightarrow \quad \hat{\sigma}_\tau^2 = \frac{QMT - QME}{r}$$

Este método pode conduzir a se ter $\hat{\sigma}_\tau^2$ negativa, o que não tem sentido.

→ Estimação por intervalo: IC aproximado

IC $100(1-\alpha)\%$ p/ $\sigma_\tau^2 = \left[L_I \; ; \; L_S\right]$, onde

$$L_I = \frac{n'\,\hat{\sigma}_\tau^2}{\chi_2^2} \quad e \quad L_S = \frac{n'\,\hat{\sigma}_\tau^2}{\chi_1^2} \quad ; \quad n' = \frac{(QMT-QME)^2}{\frac{(QMT)^2}{t-1} + \frac{(QME)^2}{n-t}}$$

Para o exemplo:

$$\hat{\sigma}^2_\tau = \frac{37,08 - 3,08}{4} = \frac{34}{4} = 8,5$$

IC 95% p/ $\sigma^2_\tau = ?$

$$n' = \frac{(34)^2}{\frac{(37,08)^2}{4} + \frac{(3,08)^2}{15}} \cong 4$$

IC 95% p/ $\sigma^2_\tau = \left[L_I \; , \; L_S\right]$

$$\chi^2_2 = \chi^2_{.025(4)} = 11,1$$

$$\chi^2_1 = \chi^2_{.975(4)} = 0,484$$

$$L_I = \frac{(4)(8,5)}{11,1} = 3,06 \qquad L_S = \frac{(4)(8,5)}{0,484} = 70,25$$

IC 95% p/ $\sigma^2_\tau = [3,1 \; ; \; 70,2]$

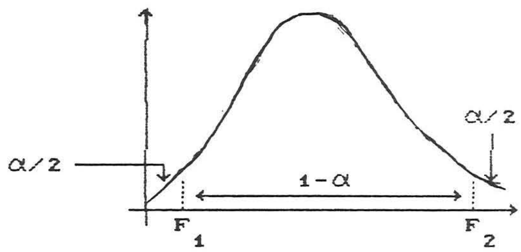
(4) Estimação do Quociente $\sigma^2_\tau \,/\, \sigma^2$

Medida de relação da variabilidade entre tratamentos e a variação dentro dos tratamentos.

O IC 100(1-α)% para $\frac{\sigma^2_\tau}{\sigma^2} = \left[L_I \; , \; L_S\right]$

onde $L_I = \frac{1}{r}\left(\frac{QMT}{QME} \cdot \frac{1}{F_2} - 1\right)$

$$L_S = \frac{1}{r}\left(\frac{QMT}{QME} \cdot \frac{1}{F_1} - 1\right)$$



Para o exemplo

$$\frac{\hat{\sigma}^2_\tau}{\hat{\sigma}^2} = \frac{8,50}{3,08} = 2,76$$

IC 95% para $\sigma^2_\tau / \sigma^2 = \left[L_I \;,\; L_S\right]$

$$L_I = \frac{1}{r}\left(\frac{QMT}{QME}\cdot\frac{1}{F_2} - 1\right) = \frac{1}{4}\left(\frac{37,08}{3,08}\cdot\frac{1}{3,80} - 1\right) = 0,54$$

$$F_2 = F_{.025(4,15)} = 3,80$$

$$L_S = \frac{1}{r}\left(\frac{QMT}{QME}\cdot\frac{1}{F_1} - 1\right) = \frac{1}{4}\left(\frac{37,08}{3,08}\cdot\frac{1}{0,1155} - 1\right) = 25,81$$

$$F_1 = F_{.975(4,15)} = \frac{1}{F_{.025(15,4)}} = \frac{1}{8,66} = 0,1155$$

IC 95% para $\dfrac{\sigma^2_\tau}{\sigma^2} = [0,54 \;;\; 25,81]$

(5) Estimação do Quociente $\dfrac{\sigma^2_\tau}{\sigma^2_\tau + \sigma^2}$

$$\rho = \frac{\sigma^2_\tau}{\sigma^2_\tau + \sigma^2}$$

Representa a proporção da variância total atribuída aos tratamentos.

↑ Coeficiente de Correlação Intraclasse

Ic a 100(1-α)% para $\rho = \left[L_I \;,\; L_S\right]$

onde $L_I = \left[\dfrac{L_i}{1 + L_i}\right]$; $L_S = \left[\dfrac{L_s}{1 + L_s}\right]$

sendo

$$L_i = \frac{1}{r}\left(\frac{QMT}{QME}\cdot\frac{1}{F_2} - 1\right)$$

$$L_s = \frac{1}{r}\left(\frac{QMT}{QME}\cdot\frac{1}{F_1} - 1\right)$$

pois

$$\rho = \frac{\sigma^2_\tau}{\sigma^2_\tau + \sigma^2} = \frac{\sigma^2_\tau / \sigma^2}{\dfrac{\sigma^2_\tau}{\sigma^2} + 1}$$

Para o exemplo

$$\hat{\rho} = \frac{\hat{\sigma}^2_\tau}{\hat{\sigma}^2_\tau + \hat{\sigma}^2} = \frac{8,5}{8,5 + 3,08} = 0,73$$

73% da variabilidade total da temperatura é explicada pela variação entre animais.

IC 95% para $\rho = \left[L_I \ , L_S\right]$

$$L_I = \frac{L_i}{1 + L_i} = \frac{0,54}{1 + 0,54} = 0,35$$

$$L_S = \frac{L_s}{1 + L_i} = \frac{25,81}{1 + 25,81} = 0,96$$

IC 95% para ρ = [0,35 ; 0,96]

2.16. ANÁLISE DE VARIÂNCIA PARA EXPERIMENTOS INTEIRAMENTE CASUALIZADOS ATRAVÉS DO MODELO DE REGRESSÃO.

1. Considerações gerais:

Considerando o modelo linear

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad \begin{array}{l} i = 1, 2, \ldots, t \\ j = 1, 2, \ldots, r \end{array}$$

O sistema de equações normais é dado por

$$\begin{array}{l} rt\hat{\mu} + r\hat{\tau}_1 + r\hat{\tau}_2 + \ldots + r\hat{\tau}_t = y_{..} \\ r\hat{\mu} + r\hat{\tau}_1 = y_{1.} \\ r\hat{\mu} + r\hat{\tau}_2 = y_{2.} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ r\hat{\mu} + r\hat{\tau}_t = y_{t.} \end{array}$$

Considerando a restrição nas soluções

$$\boxed{\sum_{i=1}^{t} \hat{\tau}_i = 0}$$

tem-se

$$\hat{\mu} = \bar{y}_{..}$$ média Geral

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} \qquad , \quad i = 1, 2, \ldots, t$$

$$\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \bar{y}_{i.}$$

A redução na soma de quadrados devido ao ajuste do modelo linear considerado é dada por

$$R(\mu,\tau) = \hat{\mu}\, y_{..} + \sum_{i=1}^{t} \hat{\tau}_i\, y_{i.}$$

$$= \left(\bar{y}_{..}\right)\left(y_{..}\right) + \sum_{i=1}^{t}\left(\bar{y}_{i.} - \bar{y}_{..}\right) y_{i.}$$

$$= \frac{y_{..}^2}{rt} + \sum_{i=1}^{t} \bar{y}_{i.}\, y_{i.} - \bar{y}_{..} \underbrace{\sum_{i=1}^{t} y_{i.}}_{y_{..}}$$

$$= \sum_{i=1}^{t} \frac{y_{i.}^2}{r}$$

que tem t GL uma vez que são t equações normais linearmente independentes.

A soma de quadrados do erro é dada por

$$SQE = \sum_{i=1}^{t} \sum_{j=1}^{r} y_{ij}^2 - \sum_{i=1}^{t} \frac{y_{i.}^2}{r}$$

que tem t(r-1) GL.

Para obter a SQ de tratamentos considera-se o modelo restritivo sob a hipótese de nulidade Ho: $\tau_i=0 \quad \forall\ i$, ou seja considera-se o modelo reduzido

$$\boxed{y_{ij} = \mu + \varepsilon_{ij}}$$

Nesse caso tem-se somente uma equação normal $rt\ \hat{\mu}=y_{..}$, da qual obtemos $\hat{\mu}=\bar{y}_{..}$. A soma de quadrados devido ao ajuste do modelo somente com μ é dada por

$$R(\mu) = \left(\bar{y}_{..}\right)\left(y_{..}\right) = \frac{y_{..}^2}{rt} \quad \text{com 1 GL.}$$

A soma de quadrados devida a tratamentos é dada por

$$SQ\ Tratamentos = R(\tau/\mu) = R(\mu,\tau) - R(\mu)$$
$$= \sum_{i=1}^{t} \frac{y_{i.}^2}{r} - \frac{y_{..}^2}{rt}$$

com t-1 GL.

Para testar a hipótese Ho: $\tau_i=0$, $\forall i$, utiliza-se a Estatística

$$\boxed{F = \frac{R(\tau/\mu)\ /\ (t-1)}{(SQE)\ /\ t(r-1)}}$$

que sob Ho tem distribuição F com (t-1) e t(r-1) GL.

2. Conexão entre o modelo de regressão e o modelo de análise de variância:

Consideremos a existência de três tratamentos, o modelo de análise de variância é dado por

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad i = 1, 2, 3$$
$$j = 1, 2, \ldots, r$$

o equivalente modelo de regressão é

$$y_{ij} = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \varepsilon_{ij} \qquad i = 1, 2, 3$$

$$j = 1, 2, \ldots, r$$

onde X_{1j} e X_{2j} são variáveis indicadoras (Variáveis "Dummy"), definidas por

$$X_{1j} = \begin{cases} 1 & \text{se a observação } j \text{ é do tratamento 1} \\ 0 & \text{em caso contrário} \end{cases}$$

$$X_{2j} = \begin{cases} 1 & \text{se a observação } j \text{ é do tratamento 2} \\ 0 & \text{em caso contrário} \end{cases}$$

A relação entre os parâmetros $\beta_0, \beta_1 e \beta_2$ no modelo de regressão e os parâmetros μ e τ_i ($i = 1, 2, 3$) no modelo de análise de variância é facilmente determinada.

Se as observações são do tratamento 1 então $X_{1j} = 1$ e $X_{2j} = 0$.

Para o modelo de regressão tem-se

$$y_{1j} = \beta_0 + \beta_1(1) + \beta_2(0) + \varepsilon_{1j} = \beta_0 + \beta_1 + \varepsilon_{1j}$$

No modelo de análise de variância uma observação do tratamento 1 é dada por

$$y_{1j} = \mu + \tau_1 + \varepsilon_{1j} = \mu_1 + \varepsilon_{1j}$$

Então a correspondência entre os dois modelos é dada por

$$\beta_0 + \beta_1 = \mu_1 = \mu + \tau_1$$

Similarmente se as observações são do tratamento 2, então $X_{1j} = 0$ e $X_{2j} = 1$ e para o modelo de regressão tem-se

$$y_{2j} = \beta_0 + \beta_1(0) + \beta_2(1) + \varepsilon_{2j} = \beta_0 + \beta_2 + \varepsilon_{2j}$$

Considerando o correspondentes modelo de análise de variância tem-se

$$y_{2j} = \mu + \tau_2 + \varepsilon_{2j} = \mu_2 + \varepsilon_2$$

Logo

$$\beta_0 + \beta_2 = \mu + \tau_2 = \mu_2$$

Considerando as observações do tratamento 3, para o qual

$X_{1j} = X_{2j} = 0$ o modelo de regressão torna-se

$$y_{3j} = \beta_0 + \beta_1(0) + \beta_2(0) + \varepsilon_{3j} = \beta_0 + \varepsilon_{3j}$$

Considerando o correspondente modelo de análise de variância, tem-se

$$y_{3j} = \mu + \tau_3 + \varepsilon_{3j} = \mu_3 + \varepsilon_{3j}$$

Logo

$$\beta_0 = \mu_3 = \mu + \tau_3$$

Desta forma a correspondência entre os parâmetros do modelo de regressão e do modelo de análise de variância, no presente caso, é dada por:

$$\beta_0 = \mu_3$$

$$\beta_0 + \beta_1 = \mu_1 \Rightarrow \beta_1 = \mu_1 - \mu_3$$

$$\beta_0 + \beta_2 = \mu_2 \Rightarrow \beta_2 = \mu_2 - \mu_3$$

Considerando-se o caso geral com $\underline{t}$ tratamentos, então o modelo de regressão terá t-1 variáveis indicadoras e é dado por

$$y_{ij} = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_{t-1} X_{t-1,j} + \varepsilon_{ij}$$

$$i = 1, 2, \ldots, t$$

$$j = 1, 2, \ldots, r$$

onde $X_{ij} = \begin{cases} 1 & \text{se o observação } j \text{ é do tratamento } i \\ 0 & \text{em caso contrário} \end{cases}$

$$i = 1, 2, \ldots, t-1$$

A relação entre os parâmetros do modelo de regressão e do modelo de análise de variância é dada por

$$\beta_0 = \mu_t$$

$$\beta_1 = \mu_i - \mu_t \quad , \; i = 1, 2, \ldots, t-1$$

Teste de Hipótese:

No modelo de análise de variância deseja-se testar $Ho: \mu_1 = \mu_2 = \mu_3 = \mu \Leftrightarrow Ho: \tau_1 = \tau_2 = \tau_3 = 0$.

Se Ho é verdadeira, então os parâmetros do modelo de regressão tornam-se

$$\beta_0 = \mu$$
$$\beta_1 = 0$$
$$\beta_2 = 0$$

Desta forma testar a hipótese

Ho: $\beta_1 = \beta_2 = 0$ no modelo de regressão é equivalente ao teste de Ho: $\mu_1 = \mu_2 = \mu_3$ ou seja o teste F para a regressão considerando o modelo de regressão com variáveis indicadoras é equivalente ao teste F para tratamentos no modelo de análise de variância.

Exemplo

	Tratamentos t_1	t_2	t_3	
	12	6	5	
	11	7	4	
	10	8	3	
			4	
$y_{i.}$	33	21	16	$70 = y_{..}$
$\bar{y}_{i.}$	11	7	4	$7 = \bar{y}_{..}$
r	3	3	4	$10 = n$

Modelo de análise de variância

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

$$\hat{\mu} = \bar{y}_{..} = 7 \qquad \hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$$

$$\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \bar{y}_{..} + \bar{y}_{i.} - \bar{y}_{..} = \bar{y}_{i.}$$

Modelo de regressão correspondente

$$y_{ij} = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \varepsilon_{ij}$$

onde

$X_{1j} \rightarrow$ 1 p/ obs. do tratamento 1; $\rightarrow$ 0 em caso contrário

$X_{2j} \rightarrow$ 1 p/ obs. do tratamento 2; $\rightarrow$ 0 em caso contrário

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \end{bmatrix} = \underbrace{\begin{bmatrix} 12 \\ 11 \\ 10 \\ 6 \\ 7 \\ 8 \\ 5 \\ 4 \\ 3 \\ 4 \end{bmatrix}}_{\underset{\sim}{y}} = \underbrace{\begin{array}{ccc} \beta_0 & \beta_1 & \beta_2 \\ \hline 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{array}}_{X} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{\underset{\sim}{\beta}} + \underbrace{\begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \end{bmatrix}}_{\underset{\sim}{\varepsilon}}$$

O sistema de Equações Normais (sen) é dado por:

$$X'X\,\hat{\beta} = X'y \qquad \begin{bmatrix} 10 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 70 \\ 33 \\ 21 \end{bmatrix} \begin{array}{l} \to y_{..} \\ \to y_{1.} \\ \to y_{2.} \end{array}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (X'X)^{-1} X'y = \begin{bmatrix} 1/4 & -1/4 & -1/4 \\ -1/4 & 7/12 & 1/4 \\ -1/4 & 1/4 & 7/12 \end{bmatrix} \begin{bmatrix} 70 \\ 33 \\ 21 \end{bmatrix} = \begin{bmatrix} 4 \\ 7 \\ 3 \end{bmatrix} \begin{array}{l} \leftarrow \hat{\mu}_3 = \hat{\mu} + \hat{\tau}_3 \\ \leftarrow \hat{\mu}_1 - \hat{\mu}_3 = \hat{\tau}_1 - \hat{\tau}_3 \\ \leftarrow \hat{\mu}_2 - \hat{\mu}_3 = \hat{\tau}_2 - \hat{\tau}_3 \end{array}$$

$\therefore$ se $\hat{\mu}_1 - \hat{\mu}_3 = 7$ e $\hat{\mu}_3 = 4 \Rightarrow \hat{\mu}_1 = 11$

se $\hat{\mu}_2 - \hat{\mu}_3 = 3$ e $\hat{\mu}_3 = 4 \Rightarrow \hat{\mu}_2 = 7$

Teste de Hipótese

$$Ho: \beta_1 = \beta_2 = 0 \quad \Leftrightarrow \quad Ho: \mu_1 = \mu_2 = \mu_3$$

Ha: $\beta_i \neq 0$ para pelo menos 1 i (i = 1, 2)

A redução da soma de quadrados considerando o modelo de regressão completo $y_{ij} = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \varepsilon_{ij}$ é dada por

$$R(\beta_0, \beta_1, \beta_2) = \text{SQ Parâmetros} = \hat{\beta}'X'y = [4 \quad 7 \quad 3] \begin{bmatrix} 70 \\ 33 \\ 21 \end{bmatrix} = 574$$

com 3 GL

Para se obter a SQ Regressão considera-se o modelo de regressão restrito sob Ho: $\beta_1 = \beta_1 = 0$, ou seja considera-se o modelo $y_{ij} = \beta_0 + \varepsilon_{ij}$

Nesse caso obtém-se a equação normal

$$10\,\hat{\beta}_0 = 70 \rightarrow \hat{\beta}_0 = \frac{70}{10} = 7 = \bar{y}_{..}$$

A redução da soma de quadrados nesse caso é dada por

$$R(\beta_0) = \hat{\beta}'X'y = 7(70) = 490 \quad \text{com 1 GL}$$

$\therefore$ SQ Regressão é dada por

$$SQ\ Regressão = R(\beta_1, \beta_2 / \beta_0) = R(\beta_0, \beta_1, \beta_2) - R(\beta_0)$$

$$= 574 - 490 = 84 = SQT \quad c/\ 2\ GL$$

$$SQ\ Total = 12^2 + \ldots + 4^2 - (70)^2/10 = 90$$

$$SQ\ Erro = SQ\ Total - SQ\ Regressão = 90 - 84 = 6$$

Causas de Variação	GL	SQ	QM	F
Regressão	2	84	42	49**
Erro	7	6	6/7	
Total	9	90		

$F = 49 \quad > F_{.01}(2,7) = 9,55 \Rightarrow$ rejeita-se Ho

$\Leftrightarrow$ Teste de tratamentos

<u>Estimação por ponto e por intervalo</u>

$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{V}(\hat{\beta}_i)}} \cap t\ (GLE)$$

$$\hat{\tau}_1 - \hat{\tau}_3 = \hat{\beta}_1 = 7$$

$$IC\left[\tau_1 - \tau_3\right] 95\% = IC\left(\beta_1\right) 95\% = \hat{\beta}_1 \pm t_{.05}(GLE)\sqrt{\hat{V}(\hat{\beta}_1)}$$

$$\hat{V}\left(\hat{\underset{\sim}{\beta}}\right) = (X'X)^{-1}\hat{\sigma}^2 = \begin{bmatrix} \hat{V}(\hat{\beta}_0) & \hat{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \hat{Cov}(\hat{\beta}_0, \hat{\beta}_2) \\ & \hat{V}(\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_1, \hat{\beta}_2) \\ & sim. & \hat{V}(\hat{\beta}_2) \end{bmatrix} = (X'X)^{-1} QME$$

$$= \begin{bmatrix} 1/4 & -1/4 & -1/4 \\ -1/4 & 7/12 & 1/4 \\ -1/4 & 1/4 & 7/12 \end{bmatrix} 6/7$$

$$\therefore \quad \hat{V}(\hat{\beta}_1) = \left(\frac{7}{12}\right)\left(\frac{6}{7}\right) = 1/2 = 0{,}5$$

$$\therefore \quad IC\ [\tau_1 - \tau_3]\ 95\% = IC\ (\beta_1)\ 95\% = \hat{\beta}_1 \pm t_{.05}(7)\ \sqrt{0{,}5}$$

$$= 7 \pm (2{,}365)(0{,}707)$$

$$= 7 \pm 1{,}7 \begin{cases} \rightarrow 8{,}7 \\ \rightarrow 5{,}3 \end{cases}$$

$$IC\ [\tau_1 + \tau_2 - 2\tau_3]\ 95\% = IC\ [\mu_1 + \mu_2 - 2\mu_3]\ 95\% = IC\ [\beta_1 + \beta_2]\ 95\%$$

$$= \hat{\beta}_1 + \hat{\beta}_2 \pm t_{.05}(7)\ \sqrt{\hat{V}(\hat{\beta}_1 + \hat{\beta}_2)} = (7+3) \pm 2{,}315\ \sqrt{10/7}$$

$$= 10 \pm 2{,}8 \begin{cases} \rightarrow 12{,}8 \\ \rightarrow 7{,}2 \end{cases}$$

$$\hat{V}(\hat{\beta}_1 + \hat{\beta}_2) = \hat{V}(\hat{\beta}_1) + \hat{V}(\hat{\beta}_2) + 2\ \hat{Cov}(\hat{\beta}_1, \hat{\beta}_2)$$

$$= [7/12 + 7/12 + 2(1/4)]\ \frac{6}{7} = 10/7$$

$$IC\ [\tau_1 - \tau_2]\ 95\% = IC\ [\beta_1 - \beta_2]\ 95\% = \hat{\beta}_1 - \hat{\beta}_2 \pm t_{.05}(7)\ \sqrt{\hat{V}(\hat{\beta}_1 - \hat{\beta}_2)}$$

$$= 4 \pm 2{,}365\ \sqrt{4/7}$$

$$= 4 \pm 1{,}8 \begin{cases} \rightarrow 5{,}8 \\ \rightarrow 2{,}2 \end{cases}$$

$$\hat{V}(\hat{\beta}_1 - \hat{\beta}_2) = \hat{V}(\hat{\beta}_1) + \hat{V}(\hat{\beta}_2) - 2\ \hat{Cov}(\hat{\beta}_1, \hat{\beta}_2)$$

$$= \left[\frac{7}{12} + \frac{7}{12} - 2\left(\frac{1}{4}\right)\right] = \frac{4}{7}$$

2.17. VERIFICAÇÃO DA ADEQUABILIDADE DO MODELO DE ANÁLISE DE VARIÂNCIA PARA O DELINEAMENTO COMPLETAMENTE CASUALIZADO COM UM FATOR.

1. SUPOSIÇÃO DA ANÁLISE DE VARIÂNCIA.

1.1. Aditividade:

O modelo é linear e aditivo

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

Erro, não observável, associado à observação ij

Efeito do tratamento $\underline{i}$

Média Geral

Observação da repetição $\underline{j}$ do tratamento $\underline{i}$

$i = 1, 2, \ldots, t \qquad j = 1, 2, \ldots, r$ (ou r_i)

1.2. Normalidade, homogeneidade e independência:

$$\varepsilon_{ij} \cap N(0, \sigma^2) \quad \text{e} \quad \text{independentes}$$

Para o modelo de efeito aleatório acrescenta-se as suposições que $\tau_i \cap N(0, \sigma_\tau^2)$ e independentes e que τ_i e ε_{ij} são independentes entre si.

2. VERIFICAÇÃO DAS SUPOSIÇÕES DA ANÁLISE DE VARIÂNCIA.

2.1. Análise de resíduos:

Para o modelo $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, tem-se que

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

Os desvios ou resíduos e_{ij} são dados por:

$$\boxed{e_{ij} = \hat{\varepsilon}_{ij} = y_{ij} - \bar{y}_{i.}} \quad \text{pois} \quad \boxed{e_{ij} = \hat{\varepsilon}_{ij} = y_{ij} - \hat{y}_{ij}}$$

Através da análise dos resíduos e_{ij} pode-se verificar se as suposições do modelo estão satisfeitas, pois pode-se verificar normalidade, homogeneidade, independência, bem como permite detectar a presença de valores aberrantes ("outliers"), que é uma das causas, a mais comum, de não-aditividade.

2.2. Verificação estatística de aditividade:

Teste de aditividade de Tukey.

2.3. Verificação estatística de normalidade:

Testes de normalidade (testes de aderência)

- Teste χ^2
- Teste de Lilliefors (Kolmogorov-Smirnov)

2.4. Verificação estatística de independência:

Teste de independência

- Teste do sinal (teste Cox-Stuart)

2.5. Verificação estatística de homocedasticidade:

Testes de homogeneidade de variâncias

- Teste de Bartlett
- Teste de Cochran
- Teste de Hartley ou F máximo

2.1. Análise de resíduos

Exemplo: DCC y = ganho de peso = peso final - peso inicial

4 rações; 5 repetições/ração suínos

Ração	Repetições 1	2	3	4	5	Médias	Variâncias
A	35	19	31	15	30	26	73
B	40	35	46	41	33	39	26,5
C	39	27	20	29	45	32	99
D	27	12	13	28	30	22	76,5
Médias			———			29,75	68,75

Resíduos: $e_{ij} = y_{ij} - \bar{y}_{i.}$

Ração	Repetições 1	2	3	4	5
A	9	-7	5	-11	4
B	1	-4	7	2	-6
C	7	-5	-12	-12	13
D	5	-10	-9	-9	8

Resíduos Padronizados:

$$e_{ij} = e_{ij} \,/\, \sqrt{QME} = e_{ij} \,/\, \sqrt{68,75} = e_{ij} \,/\, 8,2916$$

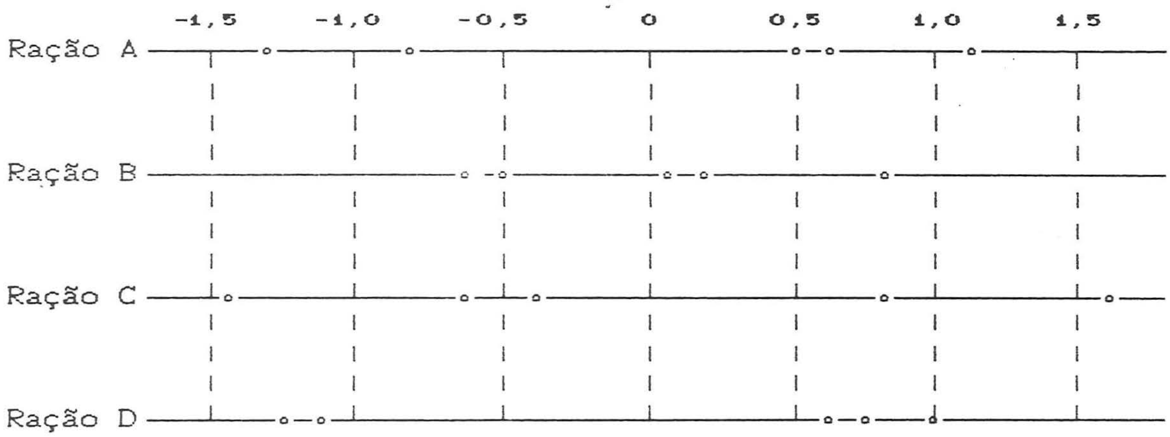
Ração	Respetições				
	1	2	3	4	5
A	1,1	-0,8	0,6	-1,3	0,5
B	0,1	-0,5	0,8	0,2	-0,7
C	0,8	-0,6	-1,4	-0,4	1,6
D	0,6	-1,2	-1,1	0,7	1,0

ANOVA

C. Variação	GL	SQ	QM	F
Rações	3	823,75	274,57	3,99
Erro	16	1100	68,75	
Total	19	1923,75		

$$\varepsilon_{ij} \cap N(0,\sigma^2) \Rightarrow \frac{\varepsilon_{ij}}{\sqrt{\sigma^2}} \cap N(0,1)$$

VERIFICAÇÃO DE HOMOCEDASTICIDADE E PRESENÇA DE VALORES ABERRANTES.



(i) A variabilidade dos resíduos é aproximadamente igual nos 4 tratamentos (rações) o que comprova a suposição de igualdade das variâncias dentro de cada tratamento (homocedasticidade dos erros).

(ii) Nenhum resíduo é muito maior ou menor que os demais, garantindo a não-existência de valores aberrantes ("outliers").

Como todos $|e_{ij}^{*}| < 1,96$, onde $e_{ij}^{*} = \dfrac{e_{ij}}{\sqrt{QME}}$,

considera-se a não-existência de "outliers" com confiabilidade de 95% .

VERIFICAÇÃO DE INDEPENDÊNCIA.

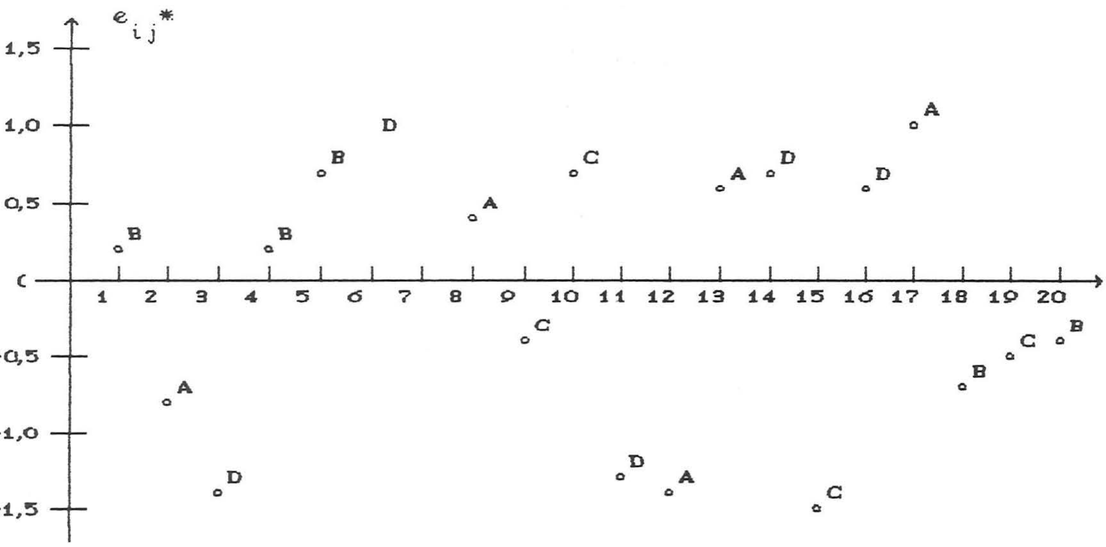
Conjunto das UE

1 2 3 4
5 6 7 8 9 10
11 12 13 14 15
16 17 18 19
20

Conjunto dos Tratamentos

A
B
C
D

Ração	Repetição 1	2	3	4	5
A	17	2	13	12	8
B	1	20	5	4	18
C	10	19	15	9	7
D	16	3	11	14	6



Os resíduos se distribuem erraticamente (sem nenhuma ordem) em relação ao eixo das abcissas, o que indica a validade da suposição de independência.

2.2. Verificação Estatística de Aditividade

2.3. Verificação Estatística de Normalidade

2.4. Verificação Estatística de Independência

2.5. Verificação estatística de homocedasticidade

Tratamentos	1	2	...	t
Variâncias	s_1^2	s_2^2	...	s_t^2
Nº de repetições	r_1	r_2	...	r_t
GL	$r_1 - 1$	$r_2 - 1$	...	$r_t - 1$

Ho: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_t^2$ [HOMOGENEIDADE DAS VARIÂNCIAS DE TRATAMENTOS]

Ha: Pelo menos 2 variâncias diferem [HETEROGENEIDADE DAS VARIÂNCIAS DE TRATAMENTOS]

2.5.1. Relação Empírica:

Se $\frac{\text{maior variância}}{\text{menor variância}} \leq 4$ (6 ou 7) $\Rightarrow$ homogeneidade de variâncias

2.5.2. Testes de homogeneidade de variâncias:

(1) Teste de Bartlett

Para o teste de Bartlett calcula-se a estatística $\chi^2 = \frac{M}{C}$ onde

$$M = \underbrace{2,3026}\left[\sum_{i=1}^{t}(r_i-1)\log\bar{s}^2 - \sum_{i=1}^{t}(r_i-1)\log s_i^2\right]$$

Fator que converte logarítmo na base decimal para logarítmo na base e.

$$\bar{s}^2 = \frac{\sum_{i=1}^{t}(r_i-1)\, s_i^2}{\sum_{i=1}^{t}(r_i-1)} = \text{QME}$$

$$C = 1 + \frac{1}{3(t-1)}\left[\sum_{i=1}^{t}\frac{1}{(r_i-1)} - \frac{1}{\sum_{i=1}^{t}(r_i-1)}\right]$$

Se $r_1 = r_2 = \ldots = r_t = r$ então

$$M = 2,3026\,(r-1)\left[t\log\bar{s}^2 - \sum_{i=1}^{t}\log s_i^2\right]$$

$$\bar{s}^2 = \frac{\sum_{i=1}^{t} s_i^2}{t} = \text{QME}$$

$$C = 1 + \frac{1}{3(t-1)}\left[\frac{t}{r-1} - \frac{1}{t(r-1)}\right] = 1 + \frac{1}{3(t-1)(r-1)}\underbrace{\left[t - \frac{1}{t}\right]}_{\frac{t^2-1}{t}}$$

$$= 1 + \frac{1}{3(t-1)(r-1)}\left[\frac{(t+1)(t-1)}{t}\right] = 1 + \frac{t+1}{3t(r-1)}$$

Sob Ho $\frac{M}{C} \cap \chi^2(t-1)$

Logo rejeita-se Ho ao nível α de significância se

χ^2 calculado $> \chi^2\,\alpha\,(t-1)$

(2) Teste de Cochran

Calcula-se a estatística C, onde

$$C = \frac{\text{maior variância}}{\sum_{i=1}^{t} s_i^2}$$

Rejeita-se Ho ao nível α de significância se

$$C_{calculado} > \underbrace{C_{\alpha}\,[t,(r-1)]}_{\text{Valor tabelado}}$$

(3) Teste de Hartley ou Teste F Máximo

Calcula-se a estatística H, onde $H = \frac{\text{maior variância}}{\text{menor variância}}$

Rejeita-se Ho ao nível α de significância se

$$H_{calculado} > \underbrace{H_{\alpha}\,(t,r)}_{\text{Valor tabelado}}$$

Exemplo:

$t = 4$

$r = 5$

Ração	s_i^2	$\log s_i^2$
A	73	1,8633
B	26,5	1,4232
C	99	1,9956
D	76,5	1,8837
Total	275	7,1658

Relação Empírica:

$$\frac{\text{maior var}}{\text{menor var}} = \frac{99}{26,5} = 3,7424$$

HOMOCEDASTICIDADE

Ho: $\sigma_A^2 = \sigma_B^2 = \sigma_C^2 = \sigma_D^2$

Ha: Pelo menos 2 variâncias diferem.

(1) Teste de Bartlett

$$\bar{s}^2 = QME = 68,75$$

log (68,75)

$$M = 2,3026\ (4)\ [(4)(1,8373) - 7,1658]$$

$$= 9,2104\ [7,3491 - 7,1658]$$

$$= 9,2104\ [0,1833] = 1,6883$$

$$C = 1 + \frac{4 + 1}{3(4)(4)} = 1 + \frac{5}{48} = 1,104$$

$$\chi^2 = \frac{M}{C} = \frac{1,6883}{1,104} = 1,53$$

$$\chi^2 = 1,53 < \chi^2_{.05}(3) = 7,81$$

Aceita-se Ho. As evidências amostrais não são suficientes para comprovar que as variâncias sejam diferentes. Logo é razoável admitir a hipótese de homogeneidade de variâncias (homocedasticidade).

(2) <u>Teste de Cochran</u>

$$C = \frac{98}{275} = 0,36$$

$\alpha = 0,05 \Rightarrow C_{.05}(4,4) = 0,6287$

$C = 0,36 < C_{.05}(4,4) = 0,6287 \Rightarrow$ aceita-se Ho.

(3) <u>Teste de Hartley ou Teste F Máximo</u>

$$H = \frac{99}{26,5} = 3,74$$

$\alpha = 0,05 \Rightarrow H_{.05}(4,5) = 20,6$

$H = 3,74 < H_{.05}(4,5) = 20,6 \Rightarrow$ aceita-se Ho.

3. CONSEQÜÊNCIAS DE FALHAS NAS SUPOSIÇÕES DA ANALISE DE VARIÂNCIA.

3.1. Não-Normalidade:

(i) <u>Modelo de efeito fixo</u>

- nível de significância levemente superior do especificado
- poder levemente menor
- estimativas por ponto continuam imparciais

(ii) <u>Modelo de efeito aleatório</u>

- estimativas pontuais imparciais dos componentes de variância
- influência acentuada nas estimativas por intervalo para componentes de variância

3.2. Heterogeneidade de variâncias (= heterocedasticidade):

(i) Teste F para dados balanceados (= nº de repetições por tratamento) é robusto, no caso de modelo de efito fixo, à heterogeneidade de variâncias.

(ii) Teste de contrastes ficam sensivelmente afetados pela heterogeneidade de variâncias.

(iii) Para modelo de feito aleatório, heterogeneidade de variâncias tem acentuado efeito na inferência de componentes de variância, no caso de dados balanceados ou não.

3.3. Não-independência dos erros:

- Não-independência tem sérios efeitos na análise de variância, tanto no caso de modelo de efeito fixo como aleatório.
- Influência sobre nível de significância e poder.
- Casualização adequada é a principal recomendação para evitar ocorrência de não-independência.
- Mudança do modelo com inclusão de termos não considerados, podem eliminar dependência dos erros.

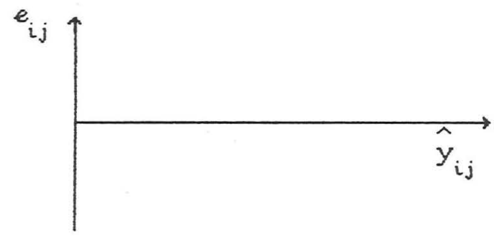
3.4. Não- aditividade:

- Influência sobre as inferências da análise de variância (estimação e teste de hipótese).
- Perda de informação.
- Menor precisão.

4. TRANSFORMAÇÃO DE DADOS.

4.1. Gráfico dos resíduos versus valores ajustados $\hat{y}_{ij}$:

Se o modelo é correto e se as suposições são satisfeitas, os resíduos não deveriam apresentar estrutura nenhuma, em particular eles deveriam ser não-relacionados com qualquer outra variável incluindo a resposta y_{ij} . Uma verificação simples é o gráfico de resíduos com os valores ajustados $\hat{y}_{ij}$. Este tipo de gráfico detecta especialmente problemas de variância não-constante (heterocedasticidade), como por exemplo quando os resíduos crescem ou decrescem com a magnitude dos dados, apresentando o gráfico uma estrutura tipo megafone ou funil.



Variância não-constante decorre de não-normalidade, (heterogeneidade regular) principalmente por problemas de assimetria, onde para distribuições assimétricas a variância tende a ser função da média. Resposta errática de tratamentos também podem provocar variância não-constante (heterogeneidade do tipo irregular).

O procedimento usual para resolver problemas de variância não-constante é a aplicação de uma transformação para estabilizar a variância e então realizar a análise de variância com os dados transformados.

<u>Exemplo</u>

Realizou-se um experimento com o objetivo de verificar o efeito de cinco inseticidas no controle do curculio da ameixa. Observou-se o número de larvas de curculio da ameixa que surgem em gaiolas sobre solo tratado, obtendo-se os seguintes resultados:

Tratamentos (inseticida)

	A	B	C	D	E	Testemunha
	14	7	6	95	37	212
	6	1	1	133	31	172
	8	0	1	86	13	202
	36	15	4	115	69	217
$\bar{y}_{i.}$	16	6	3	107	38	201
A	30	15	5	47	56	45
s_i^2	190	48	6	442	545	406
s_i	14	7	2	21	23	20

Hartley

$$H = \frac{\text{maior var}}{\text{menor var}} = \frac{545}{6} = 90,8 \quad > \quad H_{.05}(6,4) = 62,3$$

↓ Variâncias heterogêneas

ANOVA p/ dados originais

C. Variação	GL	SQ	QM	F
Trat	5	122639	24528	89,97
Erro	18	4907	273	
Total	23	127546		

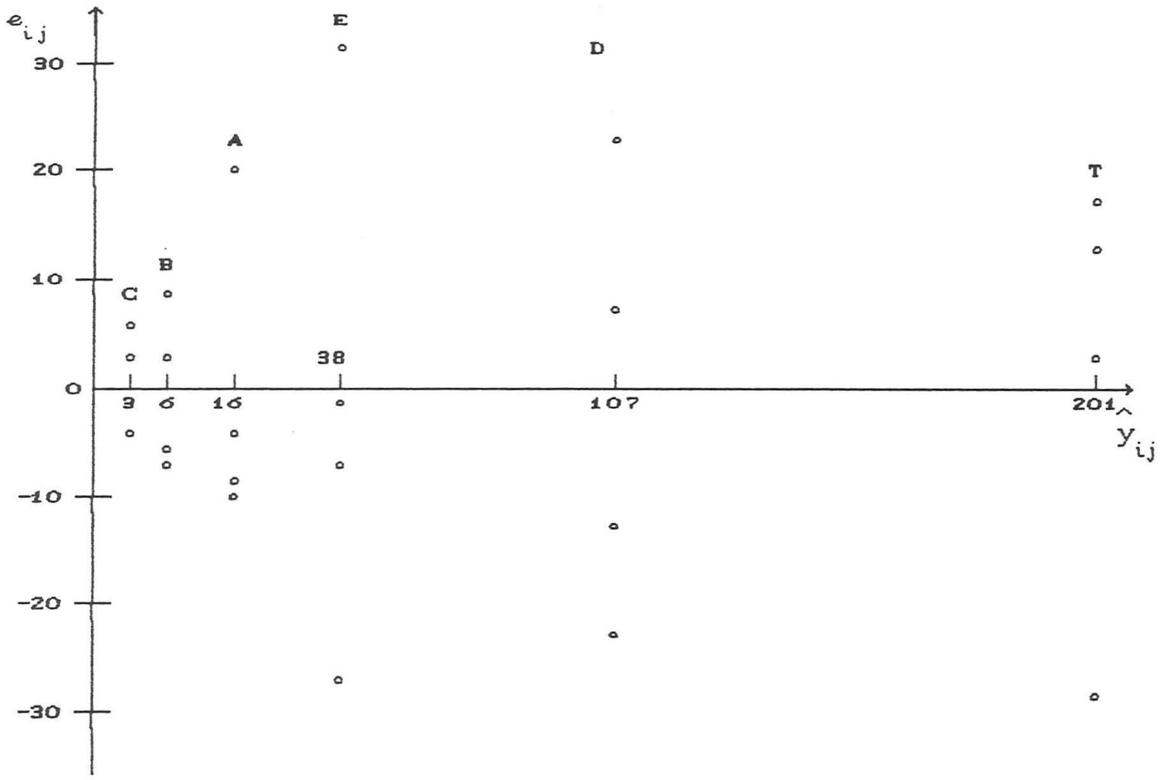
Não Válido

Resíduos: $e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i.}$

Tratamentos (inseticidas)

	A	B	C	D	E	Test.
	-2	1	3	-12	-1	11
	-10	-5	-2	23	-7	-29
	-8	-6	-2	-21	-25	1
	20	9	1	8	31	16
$\bar{y}_{i.}$	16	6	3	107	38	201

Gráfico e_{ij} versus $\hat{y}_{ij}$



4.2. Seleção de uma transformação para estabilizar a variância:

Se o pesquisador conhece a distribuição teórica das observações, ele pode utilizar esta informação na escolha da transformação. Por exemplo se as observações seguem a distribuição de Poisson, então a transformação raíz quadrada $y_{ij}*=\sqrt{y_{ij}}$ ou $y_{ij}* = \sqrt{y_{ij}+1}$ seria usada. Se os dados seguem uma distribuição log-normal, então uma transformação logarítmo é adequada $(y_{ij}* = \log y_{ij})$. Para dados com distribuição binomial expressos em proporção a transformação arco seno é apropriada, $y_{ij}* = \text{arco seno } y_{ij}$.

Quando o pesquisador não conhece a distribuição das observações (não conhece a relação entre variância e média) pode-se empiricamente estimar a forma da transformação requerida pelos dados.

4.3. Principais transformações:

(1) Transformação raíz quadrada ($y_{ij}* = \sqrt{y_{ij}}$)

Adequada principalmente para dados de contagem, quando os valores são pequenos relacionados com o universo de medida, tal como contagens por unidade de área; variáveis que tem distribuição de Poisson (nº de insetos/m^2, nº de plantas/m^2, nº de grãos/espiga, ...) na qual existe relação e proporcionalidade entre variâncias e médias, isto é, $\hat{\sigma}_i^2/\bar{y}_{i.}$ tende a ser constante.

A transformação raíz quadrada também é usada para dados de contagem, expressos em % , quando os valores estiverem no intervalo 0 - 20% ou 80 - 100% .

Quando existir valores inferiores a 10 e/ou existir valores zeros é comum utilizar-se

$$y_{ij}* = \sqrt{y_{ij} + 1/2} \quad \text{ou} \quad y_{ij}* = \sqrt{y_{ij} + 1} \quad \text{ou}$$

$$y_{ij}* = \sqrt{y_{ij}} + \sqrt{y_{ij} + 1}$$

que são mais eficientes para estabilizar a variâncias dentro dos tratamentos.

(2) Transformação logarítmo [$y_{ij}* = \log (y_{ij})$]

Transformação adequada para valores que cobrem grande amplitude de variação.

Quando existir valores menores do que 10 e/ou ocorrer valores nulos utiliza-se

$$\log (y + 1/2) \quad \text{ou} \quad \log (y + 1) \quad \text{ou} \quad \log (y + 10)$$

Transformação utilizada para dados de contagem ou de medição, quando o desvio padrão (amplitude) é proporcional a média, isto é, quando $\hat{\sigma}_i \, / \, \bar{y}_{i.}$ tende a ser constante.

(3) Tranformação angular ou arco-seno

($y_{ij}* = \text{arco-seno} \sqrt{y_{ij}}$, onde y_{ij} expresso em %)

Transformação para dados de contagem, expressos em % , que representam a ocorrência de certa característica, produzindo uma classificação em 2 categorias ou classes, ou seja, dados que se distribuem conforme a distribuição binomial.

Exemplo: % de sementes germinadas, % de frutas atacadas, % de plantas infectadas, % de animais doentes, % de peças defeituosas, etc...

Transformação:

$$\% \longrightarrow \text{ângulos (graus)}$$
$$100 \longrightarrow 90^o$$

Quando existir valor zero é preferível substituí-lo por $\frac{1}{4n}$ % e o valor 100% por $(100 - \frac{1}{4n})\%$

Para se proceder a transformação pode-se utilizar a tabela de Bliss, que é uma tabela especial para este fim.

(4) Transformação recíproca

Quando o desvio padrão é proporcional ao quadrado da média, isto é, quando $\hat{\sigma}_i/(\bar{y}_{i.})^2$ tende a ser constante, utiliza-se a transformação recíproca, dada por

$$\boxed{y_{ij}^* = \frac{1}{y_{ij}}}$$

Resumo:

Distúrbio		Solução
(1) Não-Normalidade (2) Heterocedasticidade	→	(i) Transformação de dados (ii) ANOVA não-paramétrica (iii) Modelos lineares generalizados (MLG)
(3) Não-independência	→	(i) Casualização adequada (ii) ANOVA não-paramétrica (iii) MLG
(4) Não-aditividade: "outliers"	→	(i) Eliminar os "outliers" (ii) Transformação de dados (iii) ANOVA não-paramétrica

Exemplo: Larvas de curcúlio da ameixa.

	Tratamentos						
	A	B	C	D	E	Testemunha	
$A / \bar{y}$	1,9	2,5	1,7	0,4	1,5	0,2	
$s / \bar{y}$	0,9	1,2	0,7	0,2	0,6	0,1	} certa proporcionalidade
$s^2 / \bar{y}$	11,9	8	2	4,1	14,3	2,0	

Transformação: Logarítmo

Valor zero $\Rightarrow$ $\log(y_{ij} + 10) = y_{ij}^*$

	Tratamentos						
	A	B	C	D	E	Testemunha	
$y_{i.}$	1,38	1,23	1,20	2,02	1,67	2,35	
	1,20	1,04	1,04	2,16	1,61	2,26	
	1,26	1,00	1,04	1,98	1,36	2,33	
	1,66	1,40	1,15	2,10	1,90	2,36	
$y_{i.}$	5,50	4,67	4,43	8,26	6,54	9,30	38,7
$\bar{y}_{i.}$	1,375	1,1675	1,1075	2,065	1,635	2,325	1,6125
s_i^2	0,0417	0,0341	0,0065	0,0065	0,0492	0,0020	

Teste de Homogeneidade de Variâncias: Teste de Hartley

Ho: $\sigma^2_A = \sigma^2_B = \sigma^2_C = \sigma^2_D = \sigma^2_E = \sigma^2_{Test}$

Ha: Pelo menos duas variâncias diferem.

$$H = \frac{\text{maior variância}}{\text{menor variância}} = \frac{0,0492}{0,0020} = 24,6$$

$\alpha = 0,05$ $t = 6$ $r = 4$ $H_{.05}(6,4) = 62,0$

$H = 24,6 < H_{.05}(6,4) = 62,0$

As evidências amostrais não são suficientes para comprovar que as variâncias são heterogêneas $\rightarrow$ homocedasticidade.

Análise de variância para os dados transformados por log (y+10)

Causas de variação	GL	SQ	QM	F
Tratamentos	5	4,8895	0,9779	41,97 **
Erro Experimental	18	0,4202	0,0233	Válida
Total	23	5,3097		
CV = 9,47%		$F_{.01}(5,18) = 4,25$		

Os tratamentos se diferenciam quanto a sua eficiência no controle de larvas de curcúlio de ameixa.

Teste de Tukey

$$\Delta = q_{.05}(6,18)\ s_{\bar{y}} = (4,49)(0,0763) = 0,3427$$

$$s_{\bar{y}} = \sqrt{\frac{QME}{r}} = \sqrt{\frac{0,0233}{4}} = 0,0763$$

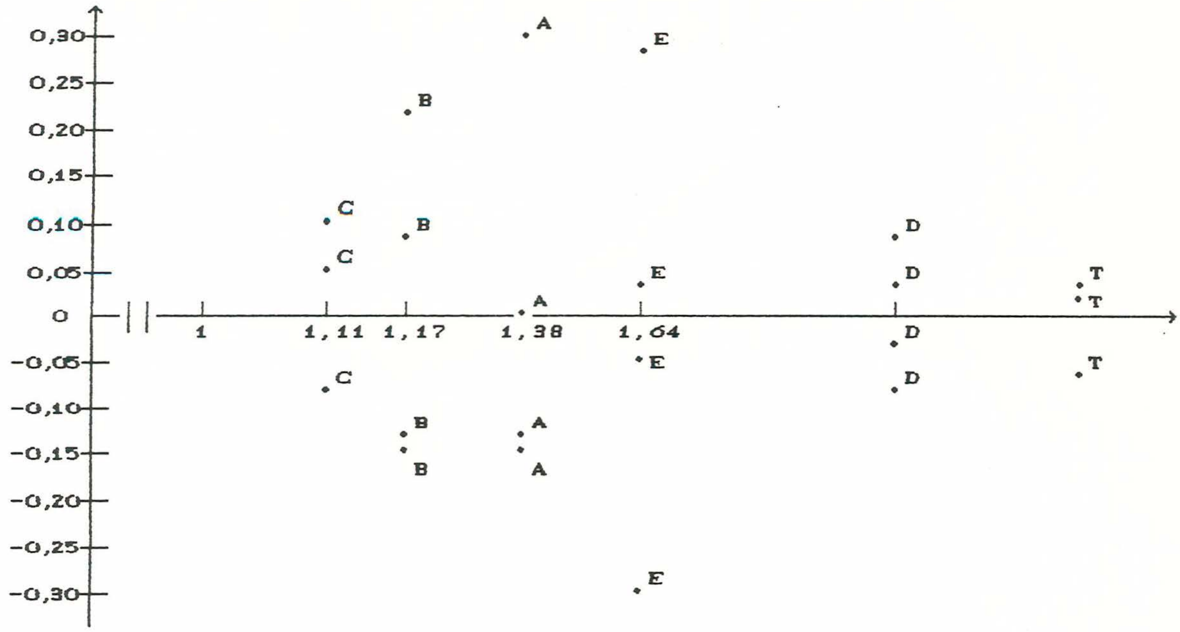
Tratamentos	Médias Transformadas	Médias Originais	
Testemunha	2,375	201	a
D	2,065	107	a
E	1,635	38	b
A	1,375	16	bc
B	1,1675	6	c
C	1,1075	3	c

Médias seguidas de mesma letra não diferem significativamente pelo teste de Tukey a 5% ∴ B e C mais eficientes.

Resíduos: $e_{ij}* = y_{ij}* - \hat{y}_{ij}* = y_{ij}* - \bar{y}_{i.}*$

Tratamentos

	A	B	C	D	E	Test
	0	0,06	0,04	-0,04	0,03	0,03
	-0,18	-0,13	-0,07	0,10	-0,03	-0,06
	-0,12	-0,17	-0,07	-0,08	-0,28	0,01
	0,28	0,23	0,04	0,04	0,26	0,04
$\bar{y}_{i.}$	1,38	1,17	1,11	2,06	1,64	2,32


0,30
0,25
0,20
0,15
0,10
0,05
0
-0,05
-0,10
-0,15
-0,20
-0,25
-0,30
1
1,11
1,17
1,38
1,64
A
B
C
D
E
T

2.19. BUSCA DA TRANSFORMAÇÃO

Seja $E_{(y)} = \mu$ a média de y e supõe-se que o desvio padrão de y é proporcional a uma potência da média de y tal que

$$\boxed{\sigma_y \propto \mu^{\alpha}}$$

Desejamos encontrar uma transformação de y que produz uma variância constante. Supõe-se que a transformação é uma potência dos dados originais, por exemplo $y^{*} = y^{\lambda}$.

Então pode-se mostrar que

$$\boxed{\sigma^{*}_{y} \propto \mu^{\lambda+\alpha-1}}$$

Logo se tomarmos $\lambda=1-\alpha$, então a variância dos dados transformados y^{*} é constante.

As transformações mais comuns são sintetizadas a seguir:

Relação entre σ_y e μ	α	$\lambda=1-\alpha$	Transformações
$\sigma_y \propto$ constante	0	1	Nenhuma
$\sigma_y \propto \mu^{1/2}$	1/2	1/2	Raíz Quadrada ← dados ∩ Poisson
$\sigma_y \propto \mu$	1	0	Logarítmo
$\sigma_y \propto \mu^{3/2}$	3/2	-1/2	Raíz quadrada recíproca
$\sigma_y \propto \mu^{2}$	2	-1	Recíproca

SELEÇÃO EMPÍRICA DE α

Em muitas situações experimentais, onde existem repetições, podemos empiricamente estimar α dos dados. Desde que no i-ésimo tratamento

$$\boxed{\sigma_{yi} \propto \mu_i^{\alpha} = \theta\ \mu_i^{\alpha}}$$

onde θ é uma constante de proporcionalidade ($\sigma_{yi} = \theta\ \mu_i^{\alpha}$).

Usando logarítmo obtemos

$$\boxed{\log \sigma_{yi} = \log \theta + \alpha \log \mu_i}$$

Assim um gráfico de $\log \sigma_{yi}$ e $\log \mu_i$, produziria uma reta com inclinação α .

Desde que não conhecemos σ_{yi} e $\log \mu_i$, utiliza-se a média $\bar{y}_{i.}$ e o desvio padrão s_i do tratamento i como estimadores de μ_i e σ_{yi} e a inclinação da linha reta ajustada como estimativa de α, ou seja

$$\boxed{\underbrace{\log s_i}_{y} = \log \theta + \hat{\alpha} \underbrace{\log \bar{y}_{i.}}_{x}}$$

Exemplo:

Um engenheiro está interessado em determinar se quatro diferentes métodos de estimação da freqüência de ocorrência de enchentes produzem equivalentes estimativas do pico de vazão quando aplicados à mesma bacia de um rio. Cada procedimento é usado 6 vezes na bacia do rio e os resultados de vazão foram os seguintes:

Método de estimação	Observações						$\bar{y}_{i.}$	s_i
1	0,34	0,12	1,23	0,70	1,75	0,12	0,71	0,66
2	0,91	2,94	2,14	2,36	2,86	4,55	2,63	1,09
3	6,31	8,37	9,75	6,09	9,82	7,24	7,93	1,66
4	17,15	11,52	10,95	17,20	14,35	16,82	14,72	2,77

ANOVA para dados originais

C. Variação	GL	SQ	QM	F
Métodos	3	708,3471	236,1157	76,07 * *
Erro	20	62,0811	3,1041	
Total	23	770,4282		

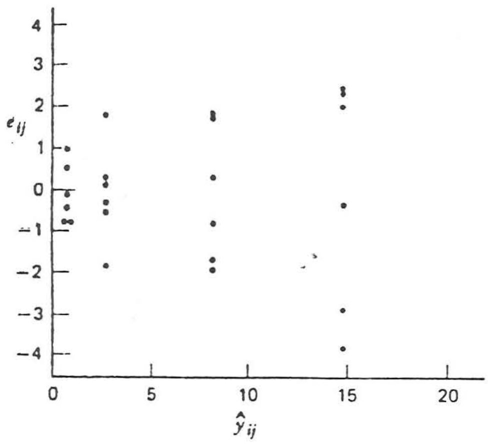
ANOVA para dados transformados $y_{ij}^{*} = \sqrt{y_{ij}}$

C. Variação	GL	SQ	QM	F
Métodos	3	32,6842	10,8947	81,05 **
Erro	20	2,6884	0,1344	
Total	23	35,3726		

. Resíduos $e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i.}$

Método	Observações						$\hat{y}_{ij} = \bar{y}_{i.}$
1	-0,37	-0,59	0,52	-0,01	1,04	-0,59	0,71
2	-1,72	0,31	-0,49	-0,27	0,23	1,92	2,63
3	-1,62	0,44	1,82	-1,84	1,89	-0,69	7,93
4	2,43	-2,90	-3,77	2,48	-0,37	2,10	14,72

. Gráfico resíduos e_{ij} versus valores ajustados $\hat{y}_{ij}$



Estrutura de funil, os resíduos crescem com a medida que crescem $\hat{y}_{ij}$.

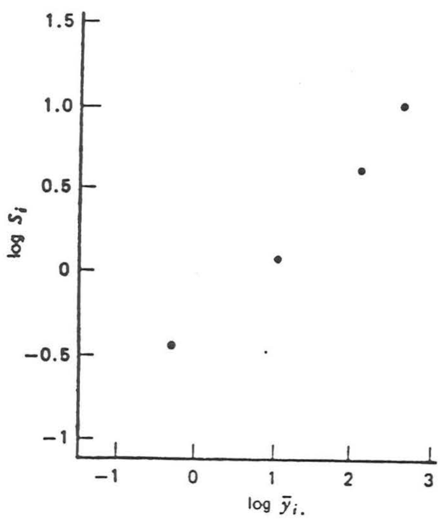
↓

Variância não constante, ou seja, de Heterocedasticidade

. $\log \bar{y}_{i.}$ e $\log s_i$

$\bar{y}_{i.}$	s_i	x: $\log \bar{y}_{i.}$	y: $\log s_i$
0,71	0,66	-0,1487	-0,1805
2,63	1,09	0,4200	0,0374
7,93	1,66	0,8993	0,2201
14,72	2,77	1,1679	0,4425

Gráfico $\log s_i$ versus $\log \bar{y}_{i.}$



. Equações de Regressão: $\hat{y} = -0,1342 + 0,4517x$

$\hat{y} = \log \theta + \hat{\alpha}\, x$ $\quad\quad \hat{\alpha} = 0,4517 \cong 1/2$

$\uparrow$ $\log s_i$ $\quad\quad$ $\uparrow$ $\log \bar{y}_{i.}$ $\quad\quad$ $\downarrow$ Transformação raíz quadrada

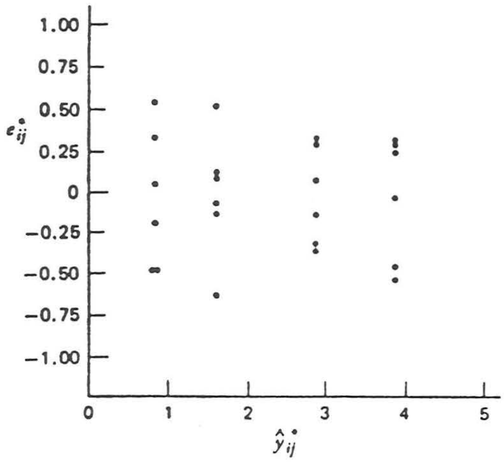
. Transforma-se dos dados para $y_{ij}{*} = \sqrt{y_{ij}}$

Método	Observações						$\bar{y}_{i.}{*}$
1	0,5831	0,3464	1,1091	0,8367	1,3229	0,3464	0,7574
2	0,9539	1,7146	1,4629	1,5362	1,6912	2,1331	1,5820
3	2,5120	2,8931	3,1225	2,4678	3,1337	2,6907	2,8033
4	4,1413	3,4380	3,3091	4,1473	3,7881	4,1012	3,8208

. Resíduos $e_{ij}* = y_{ij}* - \hat{y}_{ij}* = y_{ij}* - \bar{y}_{i.}*$

Método	Observações						$\hat{y}_{ij}* = y_{i.}*$
1	-0,1743	-0,4110	0,3517	0,0793	0,5655	-0,4110	0,7574
2	-0,6281	0,1326	-0,1191	-0,0458	0,1092	0,5511	1,5820
3	-0,2913	0,0898	0,3192	-0,3355	0,3304	-0,1126	2,8033
4	0,3205	-0,3828	-0,5117	0,3265	-0,0327	0,2804	3,8208

. Gráfico resíduos $e_{ij}*$ versus valores ajustados $\hat{y}_{ij}*$



Não se verifica estrutura de funil, não se evidenciando relação entre $e_{ij}*$ e $\hat{y}_{ij}*$.

. Análise de variância com os valores transformados $y_{ij}^{*} = \sqrt{y_{ij}}$

. ANOVA, técnicas de complementação com dados transformados. Apresentação de resultados com dados originais.

A transformação utilizada simplesmente para atender as suposições do modelo.

Publicações do Instituto de Matemática da UFRGS

Cadernos de Matemática e Estatística

Série B: Trabalho de Apoio Didático

1. Elsa Mundstock - Curso Básico Sobre Wordstar 3.45 - MAR/89

2. Jaime B. Ripoll - Introdução ao Cálculo Diferencial Via Funções de Uma Variável Real - OUT/89

3. Edmund R. Puczylowski - Dimension of Modular Lattices - JUN/90

4. Marcos Sebastiani - Geometrias Não Euclidianas - JUL/90

5. Sandra R. C. Pizzatto - Cálculo Numérico - AGO/91

6. Vera Clotilde G. Carneiro - Elementos de Cálculo para Biologia - AGO/91

7. Elsa Mundstock - Iniciação ao SPSS/PC - SET/91

8. Elisa Hagg, Loiva C. de Zeni, Maria Alice Gravina e Vera Carneiro - Notas da 1ª Oficina de Matemática da UFRGS - JAN/92

9. Paulo Werlang de Oliveira, Elisabete Rambo, Suzana Lima dos Santos, Coordenação: Profª Maria Alice Gravina - A Tartaruga no Espaço Tridimensional - FEV/92

10. Silvio Possoli - Análise Multivariada - JUL/92

11. Dinara Westphalen Fernandez - Números Índices - OUT/92

12. Maria Teresinha Albanese - Coeficiente de Fidedignidade de um Instrumento de Medida - OUT/92

13. Vera Clotilde Carneiro e Sérgio Cláudio Ramos - Gráficos na Escola - DEZ/92

14. João Riboldi - Elementos Básicos de Estatística - JAN/93

15. Paulo W. de Oliveira e M. Alice Gravina - Logo: Manual do Usuário - MAR/93

16. Ruben Markus, Elsa C. de Mundstock, Dinara W. X. Fernandez e João Riboldi - Exercícios de Métodos Estatísticos - AGO/93

17. Loiva C. de Zeni e M. Alice Gravina - Sugestões de Atividades no Ambiente Logo para a Exploração de Conteúdos Matemáticos dos Currículos Escolares de 1º e 2º Grau - SET/93

18. João Riboldi - Delineamentos Experimentais de Campo, Parte 1 - SET/93

19. Marlusa Benedetti, Patrícia P. Gil, Shirley I. Techera, Angela Andreotti, Milene Milan, Marlise Moraes, Luciana Santos, Augustinho Zimmermann, Coordenação: Profª Maria Alice Gravina - Atividades em Geometria Usando Recortes - OUT/93

20. João Riboldi - Delineamentos Experimentais de Campo, Parte 2 - OUT/93

21. Anne C. Rutsatz, Édina R. de C. Alexandre, Gorete Losada, M. Alice Gravina, Rosemary P. Disconzi, Shirley Techera e Vera C. G. Carneiro - O Pensamento e a Linguagem da Álgebra: Tabelas, Gráficos e Equações - DEZ/93

22. Dinara W. X. Fernandez - Estatística Descritiva I - JAN/94

23. João Riboldi - Planejamento e Análise de Experimentos, Parte 1- FEV/94

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE MATEMÁTICA

NÚCLEO DE ATIVIDADES EXTRA CURRICULARES

Os Cadernos de Matemática e Estatística publicam as seguintes séries:

Série A: Trabalho de Pesquisa

Série B: Trabalho de Apoio Didático

Série C: Colóquio de Matemática SBM/UFRGS

Série D: Trabalho de Graduação

Série F: Trabalho de Divulgação

Série G: Textos para Discussão

Toda correspondência com solicitação de números publicados e demais informações deverá ser enviada para:

NAEC - NÚCLEO DE ATIVIDADES EXTRA CURRICULARES

INSTITUTO DE MATEMATICA - UFRGS

AV. BENTO GONÇALVES, 9500 - PRÉDIO 43111

CEP 91509 - 900 AGRONOMIA - POA/RS

FONE: 336 92 22 OU 339 13 55 OU 228 16 33

RAMAL 6197

FAX: 336 15 12