# ANÁLISE DE VARIÂNCIA
# ANOVA

***Prof. Adriano Mendonça Souza, Dr.***
Departamento de Estatística
- PPGEMQ / PPGEP - UFSM

# UM EXEMPLO DE APLICAÇÃO

Digamos que temos 6 métodos de ensino aplicados a 30 crianças cada e gostaríamos de fazer uma comparação entre os métodos.

Fazendo-se a comparação 2 a 2 por meio do teste **Z** ou do teste **t** exigiria a execução de 15 testes, pois por meio de combinação temos $\binom{6}{2} = \frac{6!}{2!\,4!} = 15\ testes,$ ou então optar pela análise de variância, onde as hipóteses testadas seriam:

$H_0$: As médias são iguais ($\mu_1 = \mu_2 = \ldots = \mu_i$)

$H_1$: Existe pelo menos uma das médias diferentes.

Na tabela a seguir apresenta-se os métodos de ensino A, B, C, D, E e F. Apresenta-se também, a média, o desvio padrão, o n° de crianças em cada método e o respectivo grau de liberdade – $gl = \delta = N - 1$.

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| $\bar{X}$ | 75 | 72 | 76 | 79 | 82 | 72 |
| $S^2$ | 173,2 | 168,7 | 170,1 | 169,8 | 172 | 167,6 |
| N | 30 | 30 | 30 | 30 | 30 | 30 |
| gl | 29 | 29 | 29 | 29 | 29 | 29 |

Uma análise de variância permite que vários grupos sejam comparados a um só tempo, utilizando variáveis contínuas.

O teste é paramétrico (a variável de interesse deve ter distribuição normal) e os grupos devem ser independentes.

Considerando uma variável de interesse com média $\mu$ e variância $\sigma^2$ temos dois estimadores de $\sigma^2$:

$S_E^2$ = dispersão entre os grupos, que em inglês é representado por $S_B^2$ (between).

$S_D^2$ = dispersão dentro dos grupos, que em inglês é representado por $S_W^2$ (within).

O teste é aplicado utilizando a estatística calculada $F = \dfrac{S_B^2}{S_W^2}$

que é o teste que compara variâncias.

A variância das médias amostrais é calculada por

$$S^2_{\bar{x}} = \frac{\sum \bar{x}_i^2 - \frac{\left(\sum \bar{x}\right)^2}{k}}{k - 1}$$

onde:

k representa os grupos e $\delta = k - 1$ o grau de liberdade.

Como o N é igual para os 6 grupos, podemos proceder:

# Determinando $S_E^2$:

$$\sum \bar{x}_i = 75 + 72 + 76 + 79 + 82 + 72 = 456$$

$$\left(\sum \bar{x}_i\right)^2 = (456)^2$$

$$\sum \bar{x}_i^2 = 75^2 + 72^2 + 76^2 + 79^2 + 82^2 + 72^2 = 34734$$

Com k = 6

$$S_{\bar{x}}^2 = \frac{34.734 - \frac{(456)^2}{6}}{5} = 15,6$$

Pela distribuição amostral das médias temos :

$$\boxed{S_{\bar{x}}^2 = \frac{S^2}{N} \Rightarrow S^2 = S_{\bar{x}}^2 . N}$$

$S^2 = (15,6) \,.\, (30) = 468$

Mas $S^2 = S_E^2$, onde $S^2 = N \cdot S_x^2$ .

$$S_E^2 = N \left[ \frac{\sum \bar{x}_i^2 - \frac{\left(\sum \bar{x}_i\right)^2}{k}}{k-1} \right]$$

$S_E^2 = 468$

$k - 1 = 5$ gl

# Determinando $S_D^2$:

$$\boxed{S_D^2 = \frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2 + \dots + (N_k - 1)S_k^2}{(N_1 - 1) + (N_2 - 1) + (N_3 - 1) + \dots + (N_k - 1)}}$$

$$S_D^2 = \frac{(30-1)(173,2) + (30-1)(168,7) + \dots + (30-1)(167,6)}{180-6}$$

$$S_D^2 = 170,2 \text{ com } 174 \text{ graus de liberdade.}$$

Aplicando-se o teste, temos:

$$F = \frac{S_E^2}{S_D^2} = \frac{468}{170{,}2} = 2{,}75 \quad \begin{cases} \delta_1 = 5 \text{ graus de liberdade do numerador} \\ \delta_2 = 174 \text{ graus de liberdade do denominador} \end{cases}$$

$S_E^2$ sempre fica no numerador e $S_D^2$ no denominador.

Utilizando-se a estatística tabelada F a 5% tem-se que $F_{(5,174);\,5\%} = 2{,}21$, onde $F_{calc} > F_{tab}$, onde a Hipótese $H_0$ é rejeitada, isto é,

existe pelo menos 1 média diferente das demais.

# EXPERIMENTOS INTEIRAMENTE AO ACASO

## (AMOSTRAS DE MESMO TAMANHO)

A idéia da ANOVA é comparar a variação devida aos tratamentos com a variação devida ao acaso ou resíduo.

As hipóteses básicas à aplicação da ANOVA são de que:

- as K populações tenham a mesma variância $\sigma^2$ - condição de homocedasticidade;

- a variável de interesse seja normalmente distribuída em todas as populações.

# Experimento inteiramente ao acaso

| | Tratamento | | | | | Total |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | ... | k | |
| | $y_{11}$ | $y_{21}$ | $y_{31}$ | | $y_{k1}$ | |
| | $y_{12}$ | $y_{22}$ | $y_{32}$ | | $y_{k2}$ | |
| | $y_{13}$ | $y_{23}$ | $y_{33}$ | | $y_{k3}$ | |
| | . | . | . | | . | |
| | . | . | . | | . | |
| | . | . | . | | . | |
| | $y_{1r}$ | $y_{2r}$ | $y_{3r}$ | ... | $y_{kr}$ | |
| **Total** | $T_1$ | $T_2$ | $T_3$ | ... | $T_k$ | $\Sigma T = \Sigma y$ |
| **Nº de repetições** | r | r | r | ... | r | $n = kr$ |
| **Média** | $\bar{y}_1$ | $\bar{y}_2$ | $\bar{y}_3$ | ... | $\bar{y}_k$ | |

# Cálculos para realizarmos uma ANOVA:

**Determinar os graus de liberdade**

tratamentos = k - 1; total = kr - 1; resíduo = k(r-1)

$$FC = C = \frac{\left(\sum y\right)^2}{n}$$

$$SQ_{Tot} = \sum y^2 - C$$

$$SQ_{Trat} = \frac{\sum T^2}{r} - C$$

$$SQ_{Res} = SQ_{Tot} + SQ_{Trat}$$

$$QM_{Res} = \frac{SQ_{Res}}{k(r-1)}$$

$$QM_{Trat} = \frac{SQ_{Trat}}{k-1}$$

$$F = \frac{QM_{Trat}}{QM_{Res}}$$

# Quadro da ANOVA de um experimento inteiramente ao acaso

| Causas de variação | SQ | GL | QM | F |
|---|---|---|---|---|
| **Tratamentos** | $SQ_{Trat}$ | $k - 1$ | $QM_{Trat.}$ | $F = \dfrac{QM_{Trat.}}{QM_{Res.}}$ |
| **Resíduo** | $SQ_{Res.}$ | $k(r - 1)$ | $QM_{Res.}$ | |
| **Total** | $SQ_{Tot.}$ | $kr - 1$ | | |

Note que os quadrados médios são obtidos dividindo as somas de quadrados pelos respectivos graus de liberdade.

Para testar as hipóteses é utilizada a estatística F de Snedecor, com (k – 1) graus de liberdade no numerador e k.(r – 1) graus de liberdade no denominador.

Se $\mathbf{F_c > F_{\alpha,\, \delta 1,\, \delta 2}}$ **rejeita-se** $\mathbf{H_0}$ e conclui-se que existe pelo menos uma média que difere de outra.

**Se $F_{calc} > F_{tab}$, rejeitar $H_0$.**

Neste caso dizemos que existem diferenças
Estatisticamente significativas entre as médias.

**Se $F_{calc} < F_{tab}$, não rejeitar $H_0$.**

Quando isso ocorre, dizemos que não existem evidências estatísticas de que as médias sejam diferentes.

# O p-valor

Um procedimento de teste equivalente usa a probabilidade de significância (p-valor), a qual é calculada pela maioria dos programas estatísticos.

O p-valor representa a probabilidade de ser obtida uma observação da distribuição F com k – 1 e k(r – 1) graus de liberdade maior ou igual ao valor observado pela $F_{calc}$.

**Note que se o p-valor for menor que α, rejeitamos $H_0$.**

# Se p-valor < $\alpha$, rejeita-se $H_0$.

Em outras palavras, o p-valor é a probabilidade, sob $H_0$, de ocorrência do valor particular observado para a estatística de teste ou de valores mais extremos.

A probabilidade de significância de um teste mede a força da evidência contra $H_0$ em uma escala numérica.

Um p-valor pequeno indica uma forte justificativa (evidência) para a rejeição de $H_0$.

# Exemplo 3.1

Suponhamos que um pesquisador conduziu um experimento inteiramente ao acaso em um conjunto de dados que se pressupõe que sejam normalmente distribuídos e que possuam homocedasticidade. O interesse do pesquisador é avaliar se existe uma diferença significativa entre os tratamentos $T_1$, $T_2$ e $T_3$. Como você ajudaria este pesquisador por meio da ANOVA utilizando um nível de significância de 5%?

| | $T_1$ | $T_2$ | $T_3$ | |
|---|---|---|---|---|
| | 3 | 11 | 16 | |
| | 5 | 10 | 21 | |
| | 4 | 12 | 17 | **Total** |
| **Soma** | 12 | 33 | 54 | 99 |
| **Médias** | 3 | 11 | 18 | 11 |

# Exemplo 3.2

Um fornecedor alimenta a linha de produção de uma determinada indústria com peças em que a sua espessura é medida em milímetros e produzidas pelas máquinas $M_A$, $M_B$ e $M_C$, verifique se existe diferença significativa na espessura média destes itens ao nível de 5%.

| | $M_A$ | $M_B$ | $M_C$ | |
|---|---|---|---|---|
| | 3,2 | 4,9 | 3,0 | |
| n = 5 | 4,1 | 4,5 | 2,9 | |
| | 3,5 | 4,5 | 3,7 | |
| | 3,0 | 4,0 | 3,5 | |
| | 3,1 | 4,2 | 4,2 | **Total** |
| **Soma** | 16,9 | 22,1 | 17,3 | 56,3 |
| **Médias** | 3,38 | 4,42 | 3,46 | 3,75 |

# Exemplo 3.3

A Hiperfértil desenvolveu 3 tipos de fertilizantes específicos para a cultura do milho. Para testá-los, aplicou-os às mesmas áreas em pequenos sítios do interior paulista, obtendo-se a produção mostrada na tabela a seguir.

Com esses dados podemos dizer que há significativas diferenças entre os fertilizantes utilizados? Teste essa hipótese ao nível de 5%.

| | Produção em sacas de 60kg<br>Fertilizantes | | | |
|---|---|---|---|---|
| **Região** | **1** | **2** | **3** | **Total** |
| Bragança | 30 | 32 | 26 | |
| Vargem | 35 | 31 | 29 | |
| Itu | 25 | 42 | 26 | |
| **Total** | **90** | **105** | **81** | **276** |
| **Médias** | **30** | **35** | **27** | **30,66** |

# Exercício 3.1

A tabela a seguir apresenta os dados de produção de milho, em toneladas por hectare, de quatro variedades. Faça a análise de variância para verificar se a produção média das variedades de milho é igual ao nível de 5%.

| Variedades | | |
|---|---|---|
| **A** | **B** | **C** |
| 4,00 | 4,00 | 5,52 |
| 4,48 | 4,72 | 4,72 |
| 4,16 | 5,28 | 5,44 |
| 4,40 | 4,72 | 5,76 |

# Exercício 3.2

Três grupos de ratos foram treinados para realizar exercício físico anaeróbio através de uma prancha inclinada com um trilho, sobre o qual corria um carrinho com pesos diferentes que o animal empurrava. Após vários meses de treinamento, um dos grupos foi submetido à exaustão de motores a gasolina pelo mesmo tempo, e um terceiro grupo foi mantido como controle, sem as atmosferas poluídas. A tabela abaixo mostra o resultado do desempenho físico dos três grupos.

Sabendo-se que esta variável (desempenho físico) distribui-se normalmente, determine se há diferença entre os grupos.

| Rato | Controle | | Rato | Álcool | | Rato | Gasolina |
|---|---|---|---|---|---|---|---|
| 1 | 2.4 | | $1_A$ | 2.3 | | $1_G$ | 2.5 |
| 2 | 3.1 | | $2_A$ | 2.5 | | $2_G$ | 3.1 |
| 3 | 1.9 | | $3_A$ | 1.8 | | $3_G$ | 1.9 |
| 4 | 3.0 | | $4_A$ | 2.4 | | $4_G$ | 3.0 |
| 5 | 3.0 | | $5_A$ | 2.6 | | $5_G$ | 3.0 |
| 6 | 2.2 | | $6_A$ | 2.9 | | $6_G$ | 2.2 |
| 7 | 2.2 | | $7_A$ | 2.0 | | $7_G$ | 2.2 |
| 8 | 2.3 | | $8_A$ | 2.7 | | $8_G$ | 2.2 |
| 9 | 2.5 | | $9_A$ | 2.8 | | $9_G$ | 2.5 |
| 10 | 2.5 | | $10_A$ | 2.2 | | $10_G$ | 2.5 |
| 11 | 1.9 | | $11_A$ | 2.1 | | $11_G$ | 2.0 |
| 12 | 2.8 | | $12_A$ | 2.5 | | $12_G$ | 2.9 |

# EXPERIMENTOS INTEIRAMENTE AO ACASO

## (COM NÚMEROS DIFERENTES DE REPETIÇÕES)

A análise estatística de um experimento inteiramente ao acaso com número diferentes de repetições não apresenta maior dificuldade. Todos os cálculos são feitos da maneira já apresentada antes, com exceção da soma de quadrados de tratamentos. A soma de quadrados de tratamentos é dada pela fórmula:

$$SQTr = \frac{T_1^2}{r_1} + \frac{T_2^2}{r_2} + \ldots + \frac{T_k^2}{r_k} - C$$

# Exemplo 3.4

Testes psicológicos para determinar o grau de satisfação profissional foram aplicados a 35 pacientes, agrupados por faixa etária. Os resultados são os seguintes:

| 18-20 | 21-24 | 25-34 | 35-44 | 45-54 | 55-70 |
|---|---|---|---|---|---|
| 8 | 20 | 20 | 18 | 13 | 12 |
| 11 | 12 | 18 | 20 | 13 | 16 |
| 9 | 21 | 21 | 21 | 18 | 13 |
| 13 | 12 | 29 | 17 | 9 | 13 |
| 10 | | 23 | 20 | 14 | |
| | | 15 | 16 | 8 | |
| | | 27 | 22 | | |
| | | 16 | | | |
| | | 20 | | | |

Sabendo-se que a distribuição desta variável é normal, determine se houve diferença significante entre as diversas faixas etárias, usando nível de significância 1%.

# Exemplo 3.5

Admitindo-se que as notas em Estatística, para cada turma, distribuem-se normalmente com mesma variância, quer-se saber se as médias obtidas nas provas de aproveitamento em cada uma das turmas são iguais, com $\alpha = 5\%$. Para tal, sortearam-se ao acaso alunos em cada uma das turmas e verificaram-se as suas notas, obtendo-se os seguintes resultados:

| | ~~ADM – Diu.~~ | ~~ADM – Not.~~ | ~~ECO – Not.~~ | ~~ECO – Diu.~~ |
|---|---|---|---|---|
| | 2,5 | 1,0 | 9,5 | 3,5 |
| | 6,5 | 0,5 | 2,0 | 5,0 |
| | 3,5 | 0,5 | 1,0 | 2,0 |
| | 4,0 | 8,0 | 5,0 | 7,0 |
| | 5,5 | 3,0 | 2,0 | 5,0 |
| | 5,5 | 0,5 | 4,5 | 4,0 |
| | 4,5 | 3,0 | 9,0 | 4,5 |
| | 4,0 | 7,0 | 5,5 | 2,0 |
| | 10,0 | 6,5 | 3,0 | 8,5 |
| | 5,5 | 5,5 | 4,5 | 4,0 |
| | 2,2 | 0,5 | 7,0 | 1,5 |
| | 4,0 | 6,5 | 3,5 | 2,5 |
| | 3,5 | 8,5 | 9,0 | 9,5 |
| | | | 8,0 | 3,0 |
| | | | 1,0 | 8,0 |
| | | | 5,5 | 6,5 |
| | | | 8,5 | 1,5 |
| **n** | 13 | 13 | | 17 |
| **Σx** | 61,5 | 51 | 17 | 78 |
| | 4,73 | 3,92 | 88,5 | 4,58 |
| **Σx²** | 338,25 | 320 | 5,20 | 460 |
| | | | 594,75 | |

# COMPARAÇÃO DE MÉDIAS

O objetivo principal da ANOVA é apontar se um grupo é estatisticamente diferente do outro ou não. Logo, se a hipótese nula é rejeitada a um determinado nível de significância, sabemos então que existe pelo menos uma das médias de um tratamento que é diferente das demais.

Estatisticamente para determinarmos qual ou quais tratamentos não são estatisticamente iguais, utilizamos uma **diferença mínima significativa** (**dms**) que é utilizada para comparar as médias dos tratamentos.

Nada impede que se a hipótese $H_0$ seja aceita, isto é, que as médias dos tratamentos sejam consideradas iguais que uma investigação seja conduzida.

Se $H_0$ for **aceita** (médias iguais) o método de comparação de médias é dito **não-protegido**;

Se $H_0$ for **rejeitada**, uma investigação será conduzida, então o método é dito **protegido**.

# -AMOSTRAS DE MESMO TAMANHO

Os testes utilizados para se encontrar a **dms** são os seguintes:

- o **teste t**,
- o teste de **Tukey**,
- o teste de **Dunnett**
- e o teste de **Duncan**.

# O teste t

Encontrar a dms $= t_{\delta,\alpha} \cdot \sqrt{\frac{2.QMR}{r}}$

onde:

$t_{\delta,\alpha}$ ⇨ valor de estatística t tabelada com

$\delta$ ⇨ graus de liberdade do resíduo

$\alpha$ ⇨ nível de significância

QMR ⇨ quadrado médio dos resíduos

r ⇨ número de repetições de cada tratamento

Sempre que o valor absoluto da diferença entre duas médias é igual ou maior do que o valor da **dms**, diz-se que as médias são estatisticamente diferentes.

# Exemplo 4.1

Consideremos o Exemplo 1 do item 3, onde se verificou por meio da ANOVA que existe pelo menos 1 dos tratamos $T_1$, $T_2$ e $T_3$ que apresentam uma média estatisticamente diferente ao nível de 5%. Logo nos perguntamos qual ou quais tratamentos são diferentes.

# O teste de Tukey

$$\text{Encontrar a dms} = q_{\alpha,(\delta,\text{k})}\sqrt{\frac{\text{QMR}}{\text{r}}}$$

Onde:

q ⇨ é o valor tabelado, levando-se em consideração os graus de liberdade do resíduo ($\delta$) e o número de tratamentos (k) e o nível de significância ($\alpha$).

QMR ⇨ quadrado médio dos resíduos

r ⇨ número de repetições de cada tratamento

# Exemplo 4.2

Considere-se o Exemplo 3.2 do item anterior onde a espessura de produção em milímetros das máquinas $M_A$, $M_B$ e $M_C$ são testados ao nível de 5%.

# O teste de Dunnett

Este teste deve ser aplicado toda vez que se pretende comparar as médias dos tratamentos apenas com a média controle.

$$\boxed{\text{Encontra-se a dms} = d_{\alpha\,(\delta,\,T)}\sqrt{\frac{2QMR}{r}}}$$

onde:

d ⇨ valor tabelado ao nível de significância estabelecido ($\alpha$); grau de liberdade do resíduo ($\delta$) e o número de grupos tratados (T).

QMR ⇨ quadrado médio dos resíduos

r ⇨ número de repetições de cada tratamento

# Exemplo 4.3

Suponhamos que no Exemplo 3.2 anterior a máquina $M_A$ seja considerada como controle e apliquemos então o teste de Dunnett.

# AMOSTRAS DE TAMANHOS DIFERENTES

O método para o cálculo da **diferença mínima significativa (dms)** é semelhante ao exposto anteriormente, apenas com o diferencial de que o número de repetições em cada tratamento deve ser levado em consideração e que a **dms** deve ser **calculada a cada diferença** que se queira investigar. Logo apresentamos uma tabela resumo de formulação.

| | |
|---|---|
| Teste t | $$dms = t_{\delta,\alpha} \cdot \sqrt{\left(\frac{1}{r_i} + \frac{1}{r_j}\right) QMR}$$ |
| Teste de Tukey | $$dms = q_{\alpha,(\delta,k)} \cdot \sqrt{\left(\frac{1}{r_i} + \frac{1}{r_j}\right) \frac{QMR}{2}}$$ |
| Teste de Dunnett | $$dms = d_{\alpha,(\delta,T)} \cdot \sqrt{\left(\frac{1}{r_t} + \frac{1}{r_c}\right) \frac{QMR}{2}}$$ |

Onde $r_i$ e $r_j$ são o número de repetições de cada tratamento. Aqui também o teste de **Dunnett é usado para comparar o grupo tratado com o grupo controle** e $r_t$ e $r_c$ representam o número de repetições de cada grupo respectivamente.

# Exemplo 4.4

Para ilustração do procedimento, utilizamos o experimento que conta o número de ovos por poedeira, 35 dias após o início do experimento, conforme a tabela a seguir:

# N° de ovos por poedeira, 35 dias após o início do experimento.

| Tratamento | | | |
|---|---|---|---|
| A | B | C | D |
| 25 | 24 | 25 | 20 |
| 21 | 31 | 18 | 17 |
| 29 | 32 | 19 | 23 |
| | | 22 | 16 |

# ANOVA

| Causas de Variação | SQ | gl | QM | F |
|---|---|---|---|---|
| Tratamento | 200 | 3 | 66,67 | 5,13 |
| Resíduo | 130 | 10 | 13,00 | |
| Total | 330 | 13 | | |

Ao nível de 5% os tratamentos **não são iguais**.

Digamos que estamos interessados em calcular a dms entre as médias usando o teste de Tukey entre A e C, para exemplificar.

$$\mathrm{dms} = q_{\alpha,(\delta,k)} \cdot \sqrt{\left(\frac{1}{r_i} + \frac{1}{r_j}\right)\frac{QMR}{2}}$$

$$= q_{5\%(10,4)} \cdot \sqrt{\left(\frac{1}{r_A} + \frac{1}{r_C}\right)\frac{13}{2}}$$

$$= 4{,}33 \cdot \sqrt{\left(\frac{1}{3} + \frac{1}{4}\right) 6{,}5}$$

$$= 8{,}43$$

# Resumidamente teremos:

| Comparação | N° de repetições | dms Turkey | Valor absoluto da diferença |
|---|---|---|---|
| A – B | 3; 3 | 9,01 | \|25 – 29\| = 4 |
| A – C | 3; 4 | 8,43 | \|25 – 21\| = 4 |
| A – D | 3; 4 | 8,43 | \|25 – 19\| = 6 |
| B – C | 3; 4 | 8,43 | \|29 – 21\| = 8 |
| B – D | 3; 4 | 8,43 | \|29 – 19\| = 10* |
| C – D | 4; 4 | 7,81 | \|21 – 19\| = 2 |

Logo pode-se concluir que a média de B é significativamente maior que a D, logo este tratamento apresenta um resultado superior.

# OBSERVAÇÕES

⇨ Recomenda-se que quando se pretende comparar grupos tratados com o grupo controle, deve-se designar mais unidades ao grupo controle, de modo que a seguinte expressão seja satisfeita:

$$\boxed{\frac{r_c}{r_t} = \sqrt{k - 1} \quad \Rightarrow \quad r_c = r_t \, . \sqrt{k - 1}}$$

onde:

$r_c$ ⇨ repetições do grupo controle

$r_t$ ⇨ repetições do grupo tratado

$k$ ⇨ número de tratamentos

⇨ Embora o número de repetições não traga grande dificuldade para a análise de experimentos, convém lembrar, que o número igual de repetições tem alguma vantagem: a análise de variância é mais fácil e os testes de comparação de médias são exatos.

# EXPERIMENTOS EM BLOCOS AO ACASO

O experimento da análise de variância pode se tornar mais sensível se houver a possibilidade de identificar e isolar as causas que influenciam o experimento. Essas causas estranhas, quando não identificadas, contribuem para aumentar o valor de $S^2_A$ e mascarar a conclusão final

Assim, sempre que possível essas causas de variação devem ser isoladas através de um planejamento, onde as observações de cada amostra são divididas em sub-amostras e denominadas blocos. Isso equivale a fazer, em lugar de uma classificação simples, uma classificação dupla e cruzada das observações, segundo os tratamentos e segundo os blocos.

A **soma total dos quadrados** (STQ) é, agora, dividida em três componentes: **tratamento**, **bloco** e **erro**, testando-se simultaneamente dois valores de *F*, um correspondente aos tratamentos e outro aos blocos.

Se este último valor **F resultar significativo**, será indicativo de que **estivemos acertados em isolar a causa de variação**; em caso contrário, a conclusão seria a mesma se tivéssemos aplicado o modelo de classificação simples.

# OPERACIONALIZANDO EXPERIMENTOS EM BLOCOS AO ACASO

Para entender como se faz a análise de variância de um experimento em blocos ao acaso, primeiro observe a tabela a seguir. Nessa tabela estão indicados os dados de um experimento em blocos ao acaso com $k$ tratamentos e $r$ blocos. O total de cada tratamento é dado pela soma das $r$ unidades submetidas a esse tratamento; o total de bloco é dado pela soma das $k$ unidades do bloco.

# Um experimento em blocos ao acaso

| Bloco | Tratamento | | | | | Total |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **...** | **K** | |
| 1 | $y_{11}$ | $y_{21}$ | $y_{31}$ | | $y_{k1}$ | $B_1$ |
| 2 | $y_{12}$ | $y_{22}$ | $y_{32}$ | | $y_{k2}$ | $B_2$ |
| 3 | $y_{13}$ | $y_{23}$ | $y_{33}$ | | $y_{k3}$ | $B_3$ |
| . | . | . | . | | . | . |
| . | . | . | . | | . | . |
| $r$ | $y_{1r}$ | $y_{2r}$ | $y_{3r}$ | | $y_{kr}$ | $B_r$ |
| **Total** | $T_1$ | $T_2$ | $T_3$ | ... | $T_k$ | $\Sigma T = \Sigma B = \Sigma y$ |
| **Nº de repetições** | $r$ | $r$ | $r$ | ... | $r$ | $n = kr$ |
| **Média** | $\bar{y}_1$ | $\bar{y}_2$ | $\bar{y}_3$ | ... | $\bar{y}_k$ | |

# Cálculos para ANOVA de um experimento em blocos ao acaso:

**os graus de liberdade:**

tratamentos: $k - 1$

blocos: $r - 1$

total: $kr - 1$

resíduo: $(kr - 1) - (k - 1) - (r - 1) = (k - 1)\,(r - 1)$

$$\boxed{FC = C = \frac{\left(\sum y\right)^2}{kr}}$$

# Cálculos intermediários para a ANOVA

$$SQ_{Tot} = \Sigma y2 - C$$

$$SQ_{Trat} = \frac{\sum T^2}{r} - C$$

$$SQ_{Bloco} = \frac{\sum B^2}{k} - C$$

$$SQ_{Res} = SQ_{Tot} - SQ_{Trat} - SQ_{Bloco}$$

# Quadro de análise de variância de um experimento em blocos ao acaso

| Causas de Variação | SQ | gl | QM | F |
|---|---|---|---|---|
| **Tratamentos** | $SQ_{Trat}$ | $k - 1$ | $QM_{Trat}$ | $F = \dfrac{QM_{Trat}}{QM_{Res}}$ |
| **Blocos** | $SQ_{Bloco}$ | $r - 1$ | $QM_{Bloco}$ | $F = \dfrac{QM_{Bloco}}{QM_{Res}}$ |
| **Resíduos** | $SQ_{Res}$ | $(k - 1)\ (r - 1)$ | $QM_{Res}$ | |
| **Total** | $SQ_{Tot}$ | $kr - 1$ | | |

# Exemplo 5.1

Vamos considerar a tabela a seguir, que condensa os tempos, em minutos, que quatro tipos de barcos com cascos diferentes levaram para percorrer determinado circuito, em três dias diferentes: dia calmo, com ondas moderadas e um dia com ondas fortes e muito vento. Verifique se existe uma relação entre o tipo de casco com a característica das ondas.

| | Casco | Blocos | | | | |
|---|---|---|---|---|---|---|
| | | Dia 1 | Dia 2 | Dia 3 | | |
| TRATAMENTOS | 1 | 45 | 46 | 51 | 142 | 20164 |
| | 2 | 42 | 44 | 50 | 136 | 18496 |
| | 3 | 36 | 41 | 48 | 125 | 15625 |
| | 4 | 49 | 47 | 54 | 450 | 22500 |
| | $T_j$ | 172 | 178 | 203 | 553 | 76785 |
| | $T_j^2$ | 29584 | 31684 | 41209 | 102477 | |

# Exercício 5.1

| Bloco | Tratamento | | |
|---|---|---|---|
| | A | B | C |
| I | 45 | 48 | 60 |
| II | 90 | 69 | 78 |
| III | 79 | 57 | 77 |
| IV | 82 | 70 | 97 |

Considerando um experimento em blocos ao acaso, faça a análise de variância destes dados tabelados:

# Exercício 5.2

São dados os pesos de três ratos aos 30, 34, 38, 42 e 46 dias de idade. Faça a análise de variância e interprete o resultado. Considere que cada animal é um bloco e que as idades são os tratamentos.

# Peso em gramas de três ratos segundo a idade em dias

| Rato | Idade | | | | |
|---|---|---|---|---|---|
| | **30** | **34** | **38** | **42** | **46** |
| 1 | 83 | 86 | 103 | 116 | 132 |
| 2 | 63 | 69 | 79 | 81 | 98 |
| 3 | 55 | 61 | 79 | 79 | 91 |

# EXPERIMENTOS EM BLOCOS AO ACASO COM REPETIÇÃO

A metodologia na aplicação deste modelo é a mesma da anterior.

# Experimento em blocos ao acaso com repetições

| **Bloco** | **Tratamento** | | | | **Total** |
|---|---|---|---|---|---|
| | **1** | **2** | **...** | ***k*** | |
| | $y_{111}$ | $y_{211}$ | | $y_{k11}$ | |
| 1 | $y_{112}$ | $y_{212}$ | | $y_{k12}$ | $B_1$ |
| | . | . | | . | |
| | . | . | | . | |
| | $y_{11m}$ | $y_{21m}$ | | $y_{k1m}$ | |
| | $y_{121}$ | $y_{221}$ | | $y_{k21}$ | |
| 2 | $y_{122}$ | $y_{222}$ | | $y_{k22}$ | $B_2$ |
| | . | . | | . | |
| | . | . | | . | |
| | $y_{12m}$ | $y_{22m}$ | | $y_{k2m}$ | |
| . | . | . | | . | . |
| . | . | . | | . | . |
| | $y_{1r1}$ | $y_{2r1}$ | | $y_{kr1}$ | |
| $r$ | $y_{1r2}$ | $y_{2r2}$ | | $y_{kr2}$ | $B_r$ |
| | . | . | | . | |
| | . | . | | . | |
| | $y_{1rm}$ | $y_{2rm}$ | | $y_{krm}$ | |
| **Total** | $T_1$ | $T_2$ | ... | $T_k$ | $\Sigma T = \Sigma B = \Sigma y$ |
| **Núm. de repetições** | $rm$ | $rm$ | ... | $rm$ | $n = krm$ |
| **Média** | $\bar{y}_1$ | $\bar{y}_2$ | ... | $\bar{y}_k$ | |

# Para fazer a análise de variância, é preciso calcular:

os graus de liberdade

total: $n - 1$

tratamentos: $k - 1$

blocos: $r - 1$

resíduo: $(n - 1) - (k - 1) - (r - 1) = n - k - r + 1$

$$C = \frac{(\sum y)^2}{n}$$

$$SQ_{Bloco} = \frac{\sum B^2}{km} - C$$

$$SQ_{Tot} = \Sigma y^2 - C$$

$$QM_{Trat} = \frac{SQ_{Trat}}{k-1}$$

$$SQ_{Trat} = \frac{\sum T^2}{rm} - C$$

$$QM_{Bloco} = \frac{SQ_{Bloco}}{r-1}$$

$$SQ_{Res} = SQ_{Tot} - SQ_{Trat} - SQ_{Bloco}$$

$$QM_{Res} = \frac{SQ_{Res}}{n - k - r + 1}$$

o valor de *F* para tratamentos

$$F = \frac{QM_{Trat}}{QM_{Res}}$$

o valor de *F* para blocos

$$F = \frac{QM_{Bloco}}{QM_{Res}}$$

# Quadro de ANOVA

| Causas de Variação | SQ | gl | QM | F |
|---|---|---|---|---|
| Tratamento | $SQ_{Trat.}$ | $k - 1$ | $\frac{SQ_{Trat.}}{k-1}$ | $\frac{QM_{Trat.}}{QM_{Res.}}$ |
| Blocos | $SQ_{Bloco}$ | $r - 1$ | $\frac{SQ_{Bloco}}{r-1}$ | $\frac{QM_{Bloco}}{QM_{Res.}}$ |
| Resíduo | $SQ_{Res.}$ | $n - k - r + 1$ | $\frac{SQ_{Res.}}{n-k-r+1}$ | |
| Total | | $n - 1$ | | |

# Exemplo 6.1

Notas dos alunos do teste segundo o tratamento (fonte de informação) e o bloco (faixa de idade)

| Bloco | Tratamento | | | | Total |
|---|---|---|---|---|---|
| | A | B | C | D | |
| | 65 | 56 | 58 | 38 | |
| I | 69 | 49 | 65 | 30 | 648 |
| | 73 | 54 | 57 | 34 | |
| | | | | | |
| | 72 | 73 | 76 | 71 | |
| II | 79 | 77 | 69 | 65 | 864 |
| | 80 | 69 | 71 | 62 | |
| Total | 438 | 378 | 396 | 300 | 1512 |