Planejamento de Experimentos
Experimentos Hierárquicos

Enrico A. Colosimo/UFMG

Depto. Estatística - ICEx - UFMG

Exemplo: Fornecedores - Matéria Prima (Montgomery, 1997)

- Uma empresa compra matéria prima de 3 diferentes fornecedores.
- A variação da pureza da matéria prima causa problema no produto final.
- Existem quatro lotes disponíveis de matéria prima de cada fornecedor e determinações de pureza (%) são obtidas para cada lote.

Exemplo: Fornecedores - Matéria Prima (Montgomery, 1997)

- Uma empresa compra matéria prima de 3 diferentes fornecedores.
- A variação da pureza da matéria prima causa problema no produto final.
- Existem quatro lotes disponíveis de matéria prima de cada fornecedor e determinações de pureza (%) são obtidas para cada lote.
- Foram feitos três determinações da pureza em cada lote.
- Os dados são mostrados na tabela a seguir, após subtrair 93.

Exemplo: Fornecedores - Matéria Prima (Montgomery, 1997)

	Fornecedor											
	1				2				3			
Lote	1	2	3	4	5	6	7	8	9	10	11	12
	1	-2	-2	1	1	0	-1	0	2	-2	1	3
	-1	-3	0	4	-2	4	0	3	4	0	-1	2
	0	-4	1	0	-3	2	-2	2	0	2	2	1
$Y_{ij.}$	0	-9	-1	5	-4	6	-3	5	6	0	2	6

Delineamentos hierárquicos

- Porque o delineamento não é fatorial?
- Em certos experimentos envolvendo dois ou mais fatores, os níveis de um determinado fator B são similares mas não idênticos aos níveis de outro fator, digamos A.
- Tais arranjos são chamados delineamentos aninhados ou hierárquicos.

Outro Exemplo: Eficácia de duas drogas

- Considere um experimento realizado para avaliar a eficácia relativa de duas drogas em relação a algum critério específico.
- Suponha que a droga 1 é administrada a pacientes provenientes de 3 hospitais, digamos H1, H2 e H3.
- Analogamente, a droga 2 é administrada a pacientes provenientes de **outros** 3 hospitais, digamos H4, H5 e H6.
- Suponha ainda que, em cada hospital, um total de n pacientes recebem a droga alocada aquele hospital.
- Tal experimento pode ser representado esquematicamente como segue:

droga 1			droga 2		
H1	H2	H3	H4	H5	H6
n	n	n	n	n	n

Outro Exemplo: Eficácia de duas drogas

- A diferença entre o efeito da droga 1 e o efeito da droga 2 pode ser em partes atribuída às diferenças entre os efeitos (específicos) associados aos hospitais 1,2, e 3, e os efeitos (específicos) associados aos hospitais 4, 5, e 6.
- Os efeitos específicos associados aos hospitais 1, 2 e 3 estão confinados (restritos) à droga 1 enquanto os efeitos específicos associados aos hospitais 4, 5 e 6 estão confinados (restritos) à droga 2.
- Dizemos que efeitos que são restritos a um único nível de um fator são aninhados dentro desse fator.
- No experimento em questão os efeitos específicos dos hospitais estão aninhados sob o fator de drogas.
- Uma vez que um determinado hospital só aparece em uma das duas drogas, não há nenhuma maneira de avaliar o efeito da interação entre os hospitais e as drogas!

Modelo para Delineamentos hierárquicos

- O modelo linear é da forma

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \epsilon_{ijk},$$

em que $\epsilon_{ijk} \sim N(0, \sigma^2)$.

- Temos que

$$\begin{aligned}
\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{...})^2 &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left[(\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..}) + (y_{ijk} - \bar{y}_{ij.})\right]^2 \\
&= bn\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2 + n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..})^2 \\
&+ \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij.})^2.
\end{aligned}$$

- Esquematicamente, temos

$$SQ_{Total} = SQ_A + SQ_{B(A)} + SQ_E.$$

- Se ambos os fatores, A e B, são fixos, então

$$\sum_{i=1}^{a} \tau_i = 0 \quad \text{e} \quad \sum_{j=1}^{b} \beta_{j(i)} = 0, \; \forall \; i = 1, \ldots, a.$$

- O fator A é fixo e B é aleatório, então

$$\sum_{i=1}^{a} \tau_i = 0 \quad \text{e} \qquad \beta_{j(i)} \sim N(0, \sigma_\beta^2)$$

- Se ambos os fatores, A e B, são aleatórios, então

$$\tau_i \sim N(0, \sigma_\tau^2) \quad \text{e} \quad \beta_{j(i)} \sim N(0, \sigma_\beta^2)$$

Tabela: ANOVA para uma Estrutura Hierárquica.

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrado Médio	F_0
A	SQ_A	a-1	QM_A	
B(A)	$SQ_{B(A)}$	a(b-1)	QM_B	
Erro	SQ_E	ab(n-1)	QM_E	
Total	SQ_{Total}	$abn - 1$		

Tabela: Valores esperados dos quadrados médios.

	A e B fixos	A fixo e B aleatório	A e B aleatórios
$E[QM_A]$	$\sigma^2 + \frac{bn\sum\tau_i^2}{a-1}$	$\sigma^2 + n\sigma_\beta^2 + \frac{bn\sum\tau_i^2}{a-1}$	$\sigma^2 + n\sigma_\beta^2 + bn\sigma_\tau^2$
$E[QM_{B(A)}]$	$\sigma^2 + \frac{n\sum\sum\beta_{j(i)}^2}{a(b-1)}$	$\sigma^2 + n\sigma_\beta^2$	$\sigma^2 + n\sigma_\beta^2$
$E[QM_E]$	σ^2	σ^2	σ^2

Testes na ANOVA

- Se ambos os fatores, A e B, são fixos, então

$$H_0 : \tau_1 = \ldots = \tau_a = 0 \quad F_0 = QM_A/QM_E,$$

$$H_0 : \beta_{1(1)} = \ldots = \beta_{b(a)} = 0 \quad F_0 = QM_{B(A)}/QM_E$$

- Se A é fixo e B é aleatório, então

$$H_0 : \tau_1 = \ldots = \tau_a = 0 \quad F_0 = QM_A/QM_{B(A)},$$

$$H_0 : \sigma^2_\beta = 0 \quad F_0 = QM_{B(A)}/QM_E,$$

- Se A e B são aleatórios, então

$$H_0 : \sigma^2_\tau = 0 \quad F_0 = QM_A/QM_{B(A)},$$

$$H_0 : \sigma^2_\beta = 0 \quad F_0 = QM_{B(A)}/QM_E,$$

Exemplo: Descritiva - Fornecedores

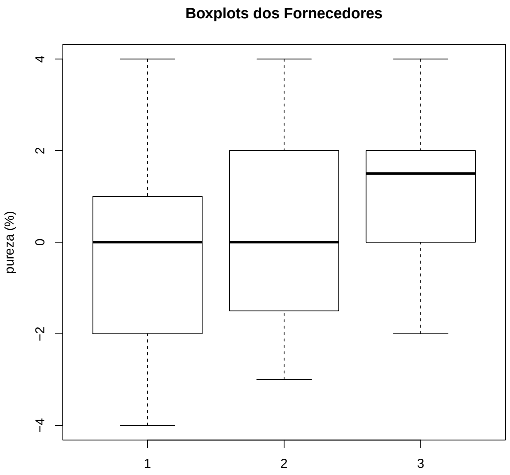
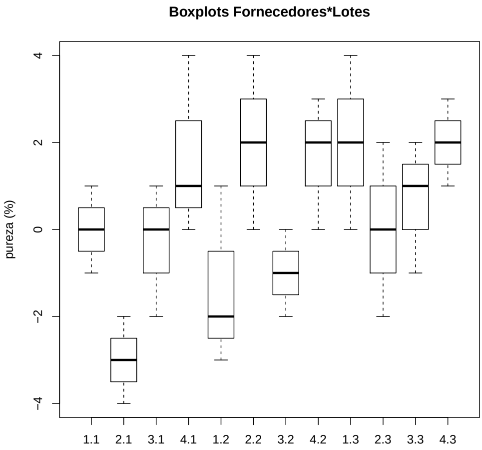


Figura: % Pureza.

Exemplo: Descritiva - Fornecedores e Lotes

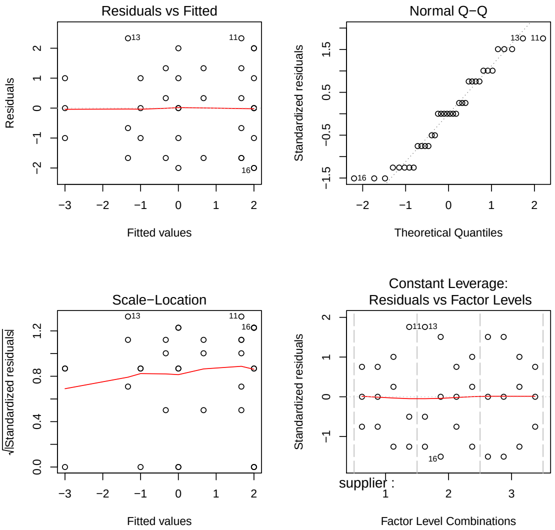


- Pode ser visualizado alguma heterogeneidade entre os lotes do mesmo fornecedor.

Exemplo: ANOVA - Fornecedores e Lotes

```
>  mod.gad1 <- lm(y˜supplier + supplier%in%batch)
>  gad(mod.gad1)
Analysis of Variance Table

Response: y
Df Sum Sq Mean Sq F value  Pr(>F)
supplier         2 15.056  7.5278  0.9690 0.41578
supplier:batch   9 69.917  7.7685  2.9439 0.01667 *
Residual        24 63.333  2.6389
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ?
```

Exemplo: Análise de Resíduos



- Indicação que as suposições são válidas.

Exemplo: Conclusões

- A maior fonte de heterogeneidade é entre lotes do mesmo fabricante.
- Não existe evidência de diferença entre fornecedores.
- O problema a ser solucionado é reduzir a variabilidade entre lotes junto aos fornecedores.

Componentes de Variância

- $$\widehat{\sigma}^2 = QM_E$$

  No exemplo: $\widehat{\sigma}^2 = 2,64$

- $$\widehat{\sigma}^2_\beta = \frac{QM_{B(A)} - QM_E}{n}$$

  No exemplo: $\widehat{\sigma}^2_\beta = \frac{7.77-2.64}{3} = 1,71$

- $$\widehat{\sigma}^2_\tau = \frac{QM_A - QM_{B(A)}}{bn}$$

Extensões do Modelo Hierárquico

- Três ou mais níveis.

$$y_{ijkl} = \mu + \tau_i + \beta_{j(i)} + \gamma_{k(ij)} + \epsilon_{ijkl},$$

- Modelo Fatorial Hierárquico.
    - Experimento Fatorial: Gênero (M/F) e Trabalha fora (s/n).
    - Unidade Experimental: aluno; Resposta: nota da primeira prova.
    - Incluir a nota da segunda prova na análise.
    - Qual delineamento é este?

- Experimento Split-plot ou parcelas sub-divididas.

Experimento Split-plot

- Experimento Split-plot envolve uma restrição operacional na execução da aleatorização nos níveis do bloco.
- Para tal, vamos dividir uma parcela principal em sub-paracelas para executar o experimento.
- Este delineamento é chamado de split-plot ou parcelas subdividas. Ou seja, vamos ter uma uma parcela principal, e a mesma será dividida em sub-parcelas para a realização do experimento.

Exemplo: Resistência do Papel (Montgomery, 1997)

- Um fabricante de papel quer comparar três métodos de preparação da celulose e quatro diferentes temperaturas de cozimento.
- Quer avaliar estes dois fatores na resistência final do papel (resposta).
- O experimento fatorial completo consiste de 12 rodadas e o fabricante decide fazer três replicas.

Exemplo: Resistência do Papel (Montgomery, 1997)

- Um fabricante de papel quer comparar três métodos de preparação da celulose e quatro diferentes temperaturas de cozimento.
- Quer avaliar estes dois fatores na resistência final do papel (resposta).
- O experimento fatorial completo consiste de 12 rodadas e o fabricante decide fazer três replicas.
- Na planta, somente é possível fazer um fatorial completo por dia (12 rodadas). Podemos, então, considerar o dia como bloco.

Exemplo: Resistência do Papel (Montgomery, 1997)

- Um fabricante de papel quer comparar três métodos de preparação da celulose e quatro diferentes temperaturas de cozimento.
- Quer avaliar estes dois fatores na resistência final do papel (resposta).
- O experimento fatorial completo consiste de 12 rodadas e o fabricante decide fazer três replicas.
- Na planta, somente é possível fazer um fatorial completo por dia (12 rodadas). Podemos, então, considerar o dia como bloco.
- No entanto, devido a restrições de produção, em cada dia não é possível repetir um mesmo método de produção.

Realização do Experimento: Resistência do Papel

- Em cada dia, um lote de celulose foi produzido por um dos três métodos, escolhido de forma aleatória.
- O lote é dividido em quatro partes, e cozinhado nas quatros temperaturas.
- O processo é repetido com outro método de produção, e mais uma vez repetido com o terceiro.
- A princípio, parece um fatorial completo (2 fatores) em bloco (dia). No entanto, o experimento não foi realizado desta forma.

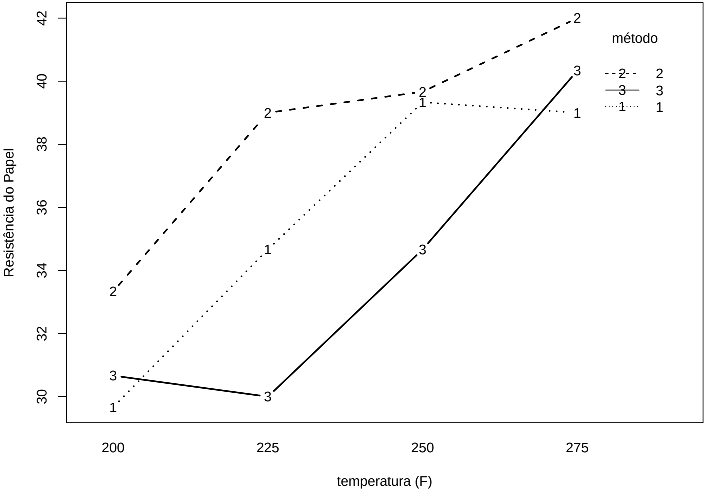
Realização do Experimento: Resistência do Papel

- Em cada dia, um lote de celulose foi produzido por um dos três métodos, escolhido de forma aleatória.
- O lote é dividido em quatro partes, e cozinhado nas quatros temperaturas.
- O processo é repetido com outro método de produção, e mais uma vez repetido com o terceiro.
- A princípio, parece um fatorial completo (2 fatores) em bloco (dia). No entanto, o experimento não foi realizado desta forma.
- Cada bloco (dia) é dividido em três partes (parcelas) que são os métodos de preparação. Cada parcela é dividida em sub-parcelas, e uma temperatura é aplicada em cada uma delas.

Exemplo: Resistência do Papel

	Blocos (dias)								
	1			2			3		
Método	1	2	3	1	2	3	1	2	3
Temperatura									
200	30	-34	29	28	31	31	31	35	32
225	35	41	26	32	36	30	37	40	34
250	37	38	33	40	42	32	41	39	39
275	36	42	36	41	40	40	40	44	45

Exemplo:Gráfico de Interação



- Indicação de possível interação entre método e temperatura.

Modelo Split-plot

- O modelo linear é da forma

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \gamma_k + (\beta\gamma)_{jk} + \epsilon_{ijk},$$

  em que $\epsilon_{ijk} \sim N(0, \sigma^2)$.
- Temos que

- τ: efeito de bloco (fator A); $i = 1, \ldots, a$;
- β: efeito de parcela (fator B); $j = 1, \ldots, b$;
- γ: efeito de sub-parcela (fator C): $k = 1, \ldots, c$;

Modelo Split-plot

- O modelo linear é da forma

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \gamma_k + (\beta\gamma)_{jk} + \epsilon_{ijk},$$

em que $\epsilon_{ijk} \sim N(0, \sigma^2)$.

- Temos que

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{...})^2 \quad = \quad SQ\text{ (parcela)} + SQ\text{ (subparcela)}$$

- Esquematicamente, temos

$$SQ_{Total} = SQ_A + SQ_B + SQ_{AB} + SQ_C + SQ_{BC} + SQ_E.$$

Tabela: Valores esperados dos quadrados médios: B e C fixos e A aleatório.

Parcela	$E[QM_A]$	$\sigma^2 + bc\sigma^2_\tau$
	$E[QM_B]$	$\sigma^2 + c\sigma^2_{\tau\beta} + \frac{ac\sum \beta_j^2}{b-1}$
	$E[QM_{AB}]$	$\sigma^2 + c\sigma^2_{\tau\beta}$
Sub-Parcela	$E[QM_C]$	$\sigma^2 + \frac{ab\sum \gamma_k^2}{c-1}$
	$E[QM_{BC}]$	$\sigma^2 + \frac{a\sum\sum(\beta\gamma)^2_{jk}}{(b-1)(c-1)}$
	$E[QM_E]$	σ^2

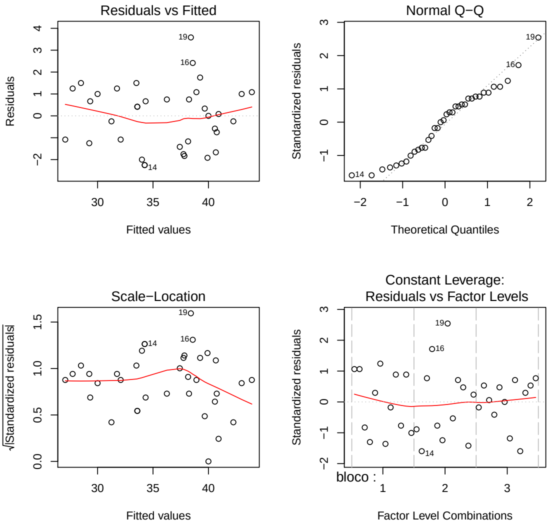
Exemplo: ANOVA - Resistência do Papel

```
> modelo1 <- aov(resist ~ bloco + metodo*temperatura + Erro
>  summary(modelo1)
    Error: bloco
Df Sum Sq Mean Sq
bloco  2  77.56   38.78

Error: bloco:metodo
Df Sum Sq Mean Sq F value Pr(>F)
metodo      2 128.39   64.19   7.078 0.0485 *
Residuals   4  36.28    9.07
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ?

Error: Within
Df Sum Sq Mean Sq F value   Pr(>F)
temperatura          3  434.1  144.69  36.427 7.45e-08 ***
metodo:temperatura   6   75.2   12.53   3.154   0.0271 *
Residuals           18   71.5    3.97
```

Exemplo: Resistência do Papel (Montgomery, 1997)

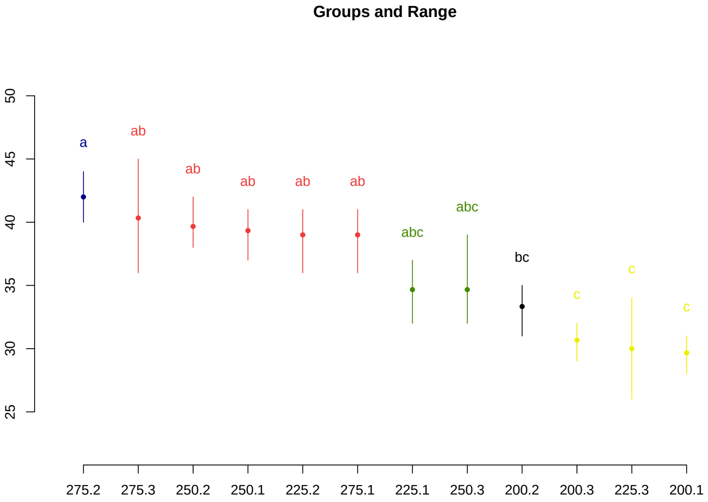
- Devemos verificar a adequação do modelo antes de acreditarmos nos resultados da ANOVA.
- É necessário um termo de erro para encontrarmos os resíduos.
- Uma forma de fazer isso é ajustar o modelo fatorial em blocos. Neste caso os testes F não são válidos mas podemos utilizar estes resultados para validar o modelo do split-plot.
- Procedimento similar será utilizado nas comparações múltiplas.

Exemplo:Gráfico de Resíduos



- Indicação de validação do modelo.

Exemplo:Comparações Múltiplas



- 12 combinações método e temperatura.

Conclusôes: Resistência do Papel

- Temperatura
  - Temperatura 200F: não existe diferença entre os métodos.
  - Temperatura 225F: existe diferença entre os métodos 2 e 3.
  - Temperatura 250F: não existe diferença entre os métodos.
  - Temperatura 275F: não existe diferença entre os métodos.
- Métodos
  - Método 1: as temperaturas de 250 e 275 se diferem da de 200F.
  - Método 2: a temperatura de 275 se diferem da de 200F.
  - Método 3: as temperaturas de 200 e 225 se diferem da de 275F
- Se objetivo é escolher a(s) combinação (ões) com a maior resistência, qual seria a sua escolha?

Conclusôes: Resistência do Papel

- Temperatura
  - Temperatura 200F: não existe diferença entre os métodos.
  - Temperatura 225F: existe diferença entre os métodos 2 e 3.
  - Temperatura 250F: não existe diferença entre os métodos.
  - Temperatura 275F: não existe diferença entre os métodos.
- Métodos
  - Método 1: as temperaturas de 250 e 275 se diferem da de 200F.
  - Método 2: a temperatura de 275 se diferem da de 200F.
  - Método 3: as temperaturas de 200 e 225 se diferem da de 275F
- Se objetivo é escolher a(s) combinação (ões) com a maior resistência, qual seria a sua escolha? Não escolher a temperatura de 200F.

  Escolher as temperaturas de 275 ou 250, com qualquer método (ou o mais econômico).