Técnicas Computacionais em Probabilidade e Estatística I

Aula VII

Chang Chiann
MAE 5704- IME/USP
1º Sem/2008

Análise Exploratória e Ajustes Robustos em ANOVA

Objetivo

Modelos ANOVA: Efeitos Fixos (e Aleatórios)

Ajustes Clássicos e Modelos Robustos

Análises de Diagnóstico

Estudos de Simulação/Testes de Aleatorização

Revisando

ANOVA
Modelos de Efeitos Fixos

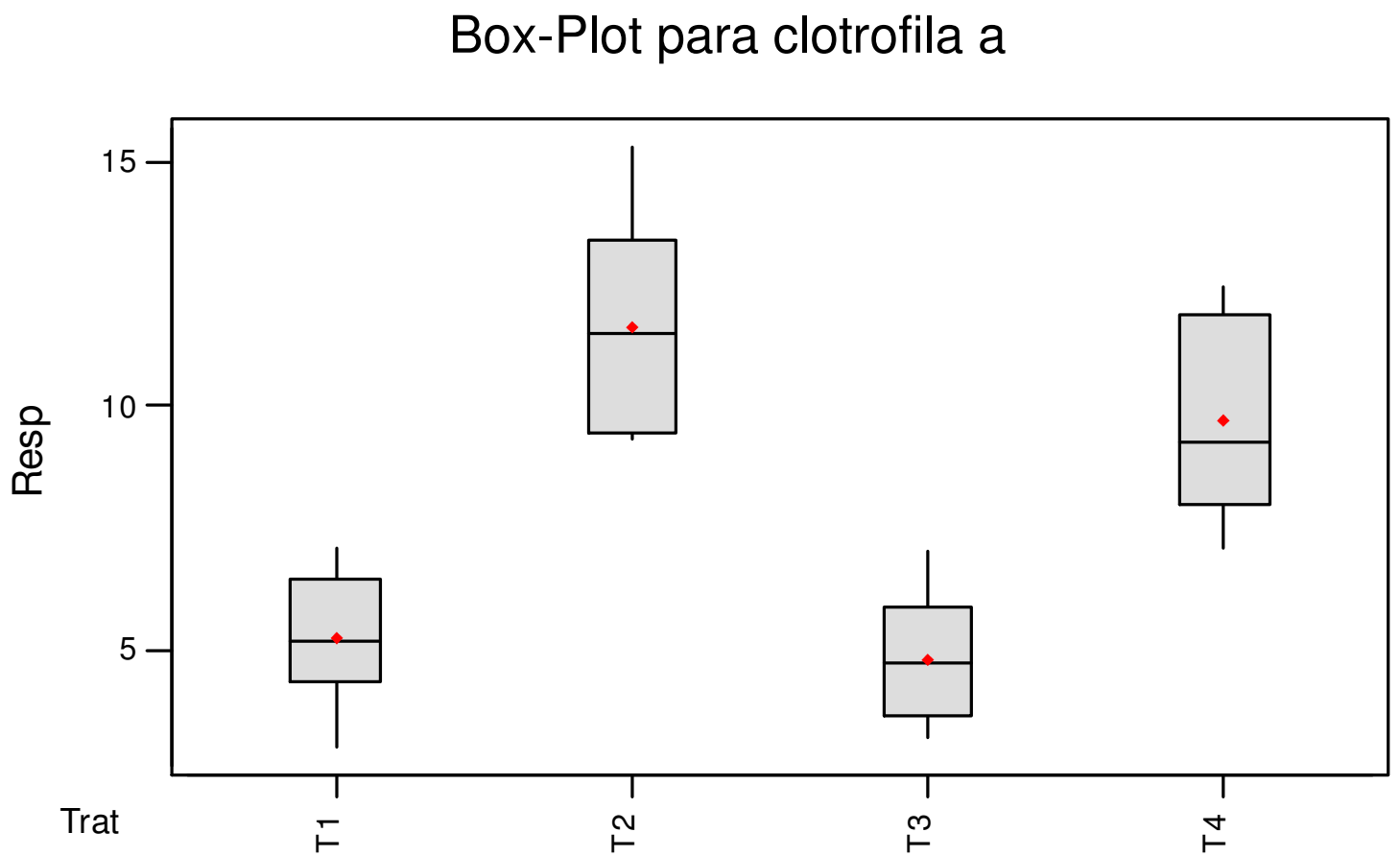
- Experimento Completamente Aleatorizado com 1 Fator
- Experimento Aleatprizado em Blocos Completos (Exp. Com Duas Entradas)
- Experimentos Fatoriais

Exemplo

Revisando

<u>Dados</u>: Medidas de clorofila a

T1	T2	T3	T4
6,2	12,7	7,0	8,3
4,8	11,3	4,4	7,1
3,0	9,3	3,8	11,7
5,6	9,5	5,0	10,0
7,1	11,7	5,5	8,5
4,8	15,3	3,2	12,4



Dados bem comportados!

Delineamento Completamente Aleatorizado - DCA

$N(\mu_1;\sigma^2)$	$N(\mu_2;\sigma^2)$	...	$N(\mu_k;\sigma^2)$	**População**
T1	**T2**	. . .	**Tk**	**Amostra**
Y11	**Y21**	. . .	**Yk1**	
. . .	. . .	**Yij**	. . .	
Y1n1	**Y2 n2**	. . .	**Yk nk**	
n1	**n2**	. . .	**nk**	
$\overline{y}_1$	$\overline{y}_2$	. . .	$\overline{y}_k$	
s_1	s_2	. . .	s_k	

- ✓ **Normalidade**
- ✓ **Variância constante**
- ✓ **Independência**

Modelo Estrutural e Distribucional

$$y_{ij} = \mu_j + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0;\sigma^2)$$

⇧ **componente fixo** ⇧ **componente aleatório**

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

⇧ **efeito do tratamento: componente da Média de Y**

$$E(y_{ij}) = \mu_j = \mu + \tau_j$$

(k+1) parâmetros definem o valor esperado de *y:* $\mu, \tau_1, \tau_2, \ldots, \tau_k$

$\Rightarrow$ *Restrições de Identificabilidade dos Parâmetros* $\sum_{j=1}^{k} \tau_j = 0$

Modelo Estrutural / Estimadores

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

⬇ *Identidade útil*

$$y_{ij} = \underbrace{\bar{y}}_{\hat{\mu}} + \underbrace{(\bar{y}_j - \bar{y})}_{\hat{\tau}_j} + \underbrace{(y_{ij} - \bar{y}_j)}_{\hat{e}_{ij}}$$

⬇

$$\underbrace{y_{ij} - \bar{y}}_{\Rightarrow SQ_{Total}} = \underbrace{(\bar{y}_j - \bar{y})}_{\Rightarrow SQ_{Modelo}} + \underbrace{(y_{ij} - \bar{y}_j)}_{\Rightarrow SQ_{\text{Re}síduo}}$$

Fontes de Variação

$N(\mu_1;\sigma^2)$ $N(\mu_k;\sigma^2)$

T_1	. . .	T_k
Y_{11}	. . .	Y_{k1}
. . .	. . .	. . .
Y_{1n1}	. . .	$Y_{k\,nk}$
n_1	. . .	n_k
$\bar{y}_1$	. . .	$\bar{y}_k$
s_1	. . .	s_k

Variância DENTRO

$$s_j^2 = \frac{\sum (y_{ij} - \bar{y}_j)^2}{n_j - 1} \qquad j = 1,\ldots,k$$

$$s_T^2 = \frac{\sum n_j (\bar{y}_j - \bar{y})^2}{k-1}$$

Var. ENTRE

Variação DENTRO

$N(\mu_1;\sigma^2)$ $N(\mu_k;\sigma^2)$

T_1	. . .	T_k
Y_{11}	. . .	Y_{k1}
. . .	. . .	. . .
Y_{1n1}	. . .	$Y_{k\,nk}$
n_1	. . .	n_k
$\bar{y}_1$	. . .	$\bar{y}_k$
s_1	. . .	s_k

$$s_j^2 = \frac{\sum (y_{ij} - \bar{y}_j)^2}{n_j - 1} \qquad j = 1,...k$$

$$s_R^2 = \frac{(n_1-1)s_1^2 + ... + (n_k-1)s_k^2}{n_1 + ... + n_k - k} = \frac{SQR}{n-k}$$

Quadrado Médio Residual (QMRes)

Estimativa da consistência interna dos dados

$$s_R^2 \quad \Rightarrow \quad \sigma^2$$

$$E(QM\,\mathrm{Re}\,s) = \sigma^2$$

Variação ENTRE

T_1	. . .	T_k
Y_{11}	. . .	Y_{k1}
. . .	. . .	. . .
Y_{1n1}	. . .	$Y_{k\,nk}$
n_1	. . .	n_k
$\bar{y}_1$	. . .	$\bar{y}_k$
s_1	. . .	s_k

$$s_T^2 = \frac{\sum n_j (\bar{y}_j - \bar{y})^2}{k-1}$$ **QMModelo**

Sob H e Balanceamento

⇩

k amostras aleatórias de tamanho n da Normal $N(\mu;\sigma^2)$

Lembra-se da distribuição amostral da média ?

Variação ENTRE

$N(\mu;\sigma^2)$ $N(\mu;\sigma^2)$

T_1	. . .	T_k
Y_{11}	. . .	Y_{k1}
. . .	. . .	. . .
Y_{1n1}	. . .	$Y_{k\,nk}$
n_1	. . .	n_k
$\bar{y}_1$	. . .	$\bar{y}_k$
s_1	. . .	s_k

Sob H

$$\frac{\sum(\bar{y}_j-\bar{y})^2}{k-1} \Rightarrow \frac{\sigma^2}{n}$$

$$s_T^2=\frac{\sum n(\bar{y}_j-\bar{y})^2}{k-1} \Rightarrow \sigma^2$$

$$E(QME)=E(QMTr)=\sigma^2$$

$$s_T^2 \Rightarrow \sigma^2$$

$$s_R^2 \Rightarrow \sigma^2$$

Sob H

$$s_T^2 \cong s_R^2$$

ANOVA

$\Longrightarrow \quad Y_{ij} \sim N(\mu_j ; \sigma^2)$

H: $\mu_1 = \mu_2 = \ldots = \mu_k = \mu$

A: existe pelo menos uma diferença

Sob H $\Rightarrow$ duas estimativas de σ^2

- Quadrado Médio de Tratamento: S_T^2
- Quadrado Médio Residual: S_R^2

$$\mathbf{F} = \frac{S_T^2}{S_R^2}$$

Qual o comportamento de F ?

Sob H: Retirar amostras de tamanho n da mesma Normal

$\Rightarrow$ útil em estudos de simulação

Tabela de ANOVA

H: $\mu_1 = \mu_2 = \ldots = \mu_k = \mu$

F.V.	g l	SQ	QM	F	p
ENTRE	**K-1**	$\sum n_j (\bar{y}_j - \bar{y})^2$	**SQE / (K-1)**	**QME / QMR**	
DENTRO	**N-K**	$\sum_{ij} (y_{ij} - \bar{y}_j)^2$	**SQR / (N-K)**		
TOTAL	**N-1**	$\sum_{ij} (y_{ij} - \bar{y})^2$			

$$\mathbf{F} = \frac{S_T^2}{S_R^2} \sim \mathbf{F}\ (\textit{K-1}, \textit{N-K})$$

normalidade
homocedasticidade
independência

Tabela de ANOVA

$$H : \mu_1 = \mu_2 = ... = \mu_k = \mu \quad \Leftrightarrow \quad \tau_1 = \tau_2 = ... = \tau_k = 0$$

F.V.	g l	SQ	QM	F	p
ENTRE	**3**	**201.45**	**67.15**	**20.59**	**0.00**
DENTRO	**20**	**65.23**	**3.26**		
TOTAL	**23**	**266.68**			

Conclusão da análise? (Descritiva e Inferencial)

Diagnóstico

⇒ Checar as suposições do modelo

⇒ Identificar pontos “aberrantes”

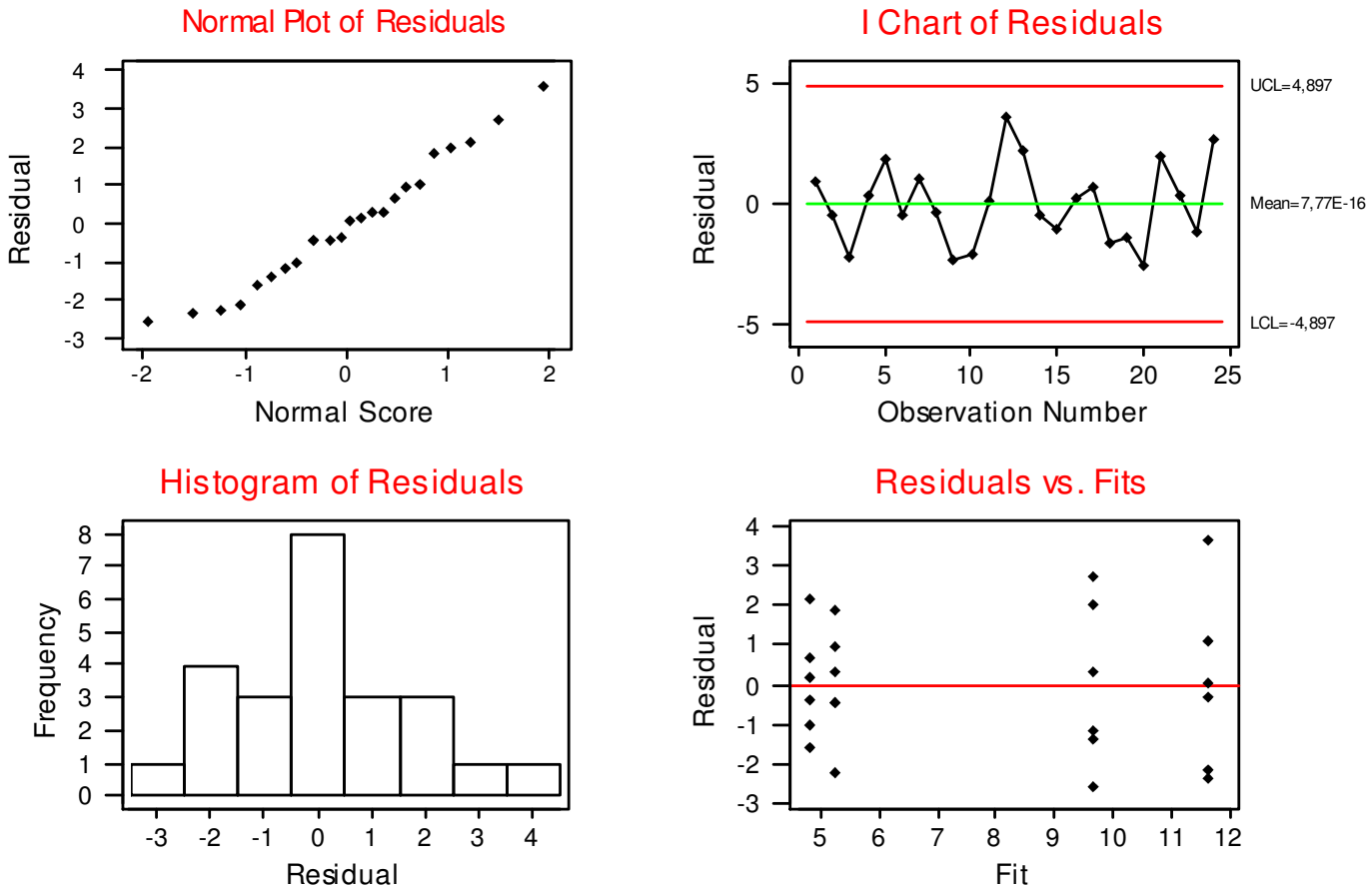
$$y_{ij} = \mu_j + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0;\sigma^2)$$

✓ Normalidade

✓ Variância constante (homocedasticidade)

✓ Independência

Residual Model Diagnostics



⇒ Outros procedimentos de diagnóstico:

Verificar a existência de pontos Aberrantes, de Alavanca e Influentes

Medidas de Diagnóstico

Modelo Clássico

$$\Rightarrow \qquad Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1} \qquad \Rightarrow \qquad \hat{Y} = X\,\hat{\beta} = X\left(X'X\right)^{-1}X'Y = H\,Y$$

- Identificação de pontos de alavanca (alto *leverage*):

$$\hat{y}_j = \sum_{j=1}^{n} x_j'\left(X'X\right)^{-1}x_j\,y_j \qquad \Rightarrow \qquad \hat{y}_j = h_{jj}\,y_j + \left(1-h_{jj}\right)X_j'\,\hat{\beta}_{(j)}$$

alavanca do valor ajustado $\left(h_{jj} > {}^{2p}/_{n}\right)$

- Identificação de pontos aberrantes:

$$t_j^* = \frac{\hat{\varepsilon}_j}{s_{(j)}\left(1-h_{jj}\right)^{1/2}} \sim t_{n-p-1}$$

resíduo studentizado (deletado)

- Identificação de pontos influentes (Cook):

$$D_j = \frac{\left(\hat{\beta}-\hat{\beta}_{(j)}\right)'X'X\left(\hat{\beta}-\hat{\beta}_{(j)}\right)}{p\,s^2} \quad > \quad F_{p,(n-p-1)}(1-\alpha)$$

Delineamento Aleatorizado em Blocos Completos

	Tratamentos			
Bloco	**T_1**	**T_2**	**. . .**	**T_k**
B_1	**Y_{11}**	**Y_{21}**	**. . .**	**Y_{k1}**
B_2	**Y_{12}**	**Y_{22}**	**. . .**	**Y_{k2}**
	. . .	**. . .**	**Y_{ij}**	**. . .**
B_n	**Y_{1n}**	**Y_{2n}**	**. . .**	**Y_{kn}**

“aleatorização restrita”

dentro dos blocos

k u.e. dentro de cada bloco são atribuídas aos tratamentos

n replicações em cada tratamento

⇒ Controlar FV externas

⇒ Ganhar precisão

Delineamento Completamente Aleatorizado
Delineamento Aleatorizado em Blocos Completos

DCA

T1	T2	…	Tk
.	.		.
.	.		.
.	.		.
.	.		.
.	.		.
n1	**n2**	**…**	**nk**

N

Aleatorização irrestrita das N unidades experimentais aos k Tratamentos

DAB

	T1	T2	…	Tk
B1	.	.		.
B2	.	.		.
…	.	.		.
Bn	.	.		.
	n	**n**	**…**	**n**

Aleatorização restrita das k unidades experimentais dentro de cada bloco

⇒ Mesmas suposições distribucionais

⇒ Diferentes esquemas de aleatorização

Modelo Estrutural / Estimadores

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

Identidade útil

$$y_{ij} = \underbrace{\bar{y}}_{\hat{\mu}} + \underbrace{(\bar{y}_j - \bar{y})}_{\hat{\tau}_i} + \underbrace{(\bar{y}_i - \bar{y})}_{\hat{\beta}_j} + \underbrace{(y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y})}_{\hat{e}_{ij}}$$

$\hat{e}_{ij}$: efeito de interação

$$\Rightarrow SQ_{Total} \quad \Rightarrow SQ_{Trat} \quad \Rightarrow SQ_{Bloco} \quad \Rightarrow SQ_{\text{Re}síduo}$$

Tabela de ANOVA

H: $\mu_1 = \mu_2 = \ldots = \mu_k = \mu$

Ganhar precisão?

F.V.	g l	SQ	QM	F	p
TRAT	**k-1**	$\sum n\ (\bar{y}_{.j} - \bar{y})^2$	**SQTR/(k-1)**	**QMTR/QMR**	
BLOCO	**n-1**	$\sum k\ (\bar{y}_{i.} - \bar{y})^2$			
RESÍDUO	**(k-1)(n-1)**	$\sum_{ij} (y_{ij} - \bar{y}_{.j} - \bar{y}_{i.} + \bar{y})^2$	**SQR/(k-1)(n-1)**		
TOTAL	**nk-1**	$\sum_{ij} (y_{ij} - \bar{y})^2$			

F ~ F [k-1 , (k-1)(n-1)]

Análise de Diagnóstico

"Resíduo" é o Efeito de Interação Bloco Trat*

Exemplo

Dados: Medidas de clorofila a

	Tratamento			
Bloco	**T1**	**T2**	**T3**	**T4**
B1	**6,2**	**12,7**	**7,0**	**8,3**
B2	**4,8**	**11,3**	**4,4**	**7,1**
B3	**3,0**	**9,3**	**3,8**	**11,7**
B4	**5,6**	**9,5**	**5,0**	**10,0**
B5	**7,1**	**11,7**	**5,5**	**8,5**
B6	**4,8**	**15,3**	**3,2**	**12,4**

hipotético

Tabela de ANOVA

H: $\mu_1 = \mu_2 = \ldots = \mu_k = \mu$

F.V.	g l	SQ	QM	F	p
TRAT	**3**	**201.448**	**67.149**	**19.79**	**0.000**
BLOCO	**5**	**14.343**	**2.869**	**0.85**	**0.538**
RESÍDUO	**15**	**50.346**	**3.392**		
TOTAL	**23**	**266.678**			

Com a inclusão de um suposto controle do efeito de uma variável bloco, houve ganho em precisão na identificação de efeito do tratamento ?

Delineamento Fatorial

A_1			A_2			...	A_a		
B_1	B_2	... B_b	B_1	B_2	... B_b	...	B_1	B_2	... B_b
-	-						-	-	
-	-		Y_{ijk}				-	-	
-	-						-	-	

- Estrutura de Tratamento $\Rightarrow$ 2 ou + Fatores Cruzados
- Delineamento com Replicações em cada combinação dos níveis dos fatores
- Compare este delineamento com o caso de Blocos

Exemplo

<u>Dados</u>: Medidas de clorofila a

T1	T2	T3	T4	
30%		100%		← Luminosidade
SN	N	SN	N	← Nutrientes
6,2	12,7	7,0	8,3	
4,8	11,3	4,4	7,1	
3,0	9,3	3,8	11,7	
5,6	9,5	5,0	10,0	
7,1	11,7	5,5	8,5	
4,8	15,3	3,2	12,4	

Tabela de ANOVA

F.V.	g l	SQ		
ENTRE	**K-1**	$\sum n_j (\bar{y}_j - \bar{y})^2$	**Luminosidade**	**a-1**
			Nutrientes	**b-1**
			Lumino*Nutrient	**(a-1)*(b-1)**
DENTRO	**N-K**	$\sum_{ij} (y_{ij} - \bar{y}_j)^2$		
TOTAL	**N-1**	$\sum_{ij} (y_{ij} - \bar{y})^2$		

$\Rightarrow$ *Análise de Diagnóstico*

Tabela de ANOVA

H: $\mu_1 = \mu_2 = \ldots = \mu_k = \mu$

F.V.	g l	SQ	QM	F	p
Luminosidade	**2-1**	**8.640**	**8.640**	**2.65**	**0.119**
Nutrientes	**2-1**	**189.282**	**189.282**	**58.04**	**0.000**
Lumino*Nutrient	**1**	**3.527**	**3.527**	**1.08**	**0.311**
DENTRO	**20**	**65.23**	**3.26**		
TOTAL	**23**	**266.68**			

Análise de Diagnóstico

Conclusão?

Exemplo

- Dados de Causas de Doenças

	Não fumante	1 -14	15-24	25+
Câncer				
Pulmão	0,07	0,47	0,86	1,66
Respiratório Superior	0,00	0,13	0,09	0,21
Estomago	0,41	0,36	0,10	0,31
Colon e Reto	0,44	0,54	0,37	0,74
Próstata	0,55	0,26	0,22	0,34
Outros tipos	0,64	0,72	0,76	1,02
Doenças Respiratórias				
Pulmonar	0,00	0,16	0,18	0,29
Bronquite	0,12	0,29	0,39	0,72
Outras	0,69	0,55	0,54	0,40
Trombose Coronária	4,22	4,64	4,60	5,99
Outras - cardiovascular	2,23	2,15	2,47	2,25
Hemorragia Cerebral	2,01	1,94	1,86	2,33
Ulcera Péptica	0,00	0,14	0,16	0,22
Violência	0,42	0,82	0,45	0,90
Outras Doenças	1,45	1,81	1,47	1,57

Resposta: Razão de mortes de homens por 1000 habitantes, de acordo com a causa de morte e o hábito de fumar (# de cigarros consumidos diariamente)

⇒ Qual é o tipo do Delineamento Experimental?

⇒ Quais são os fatores sob estudo?

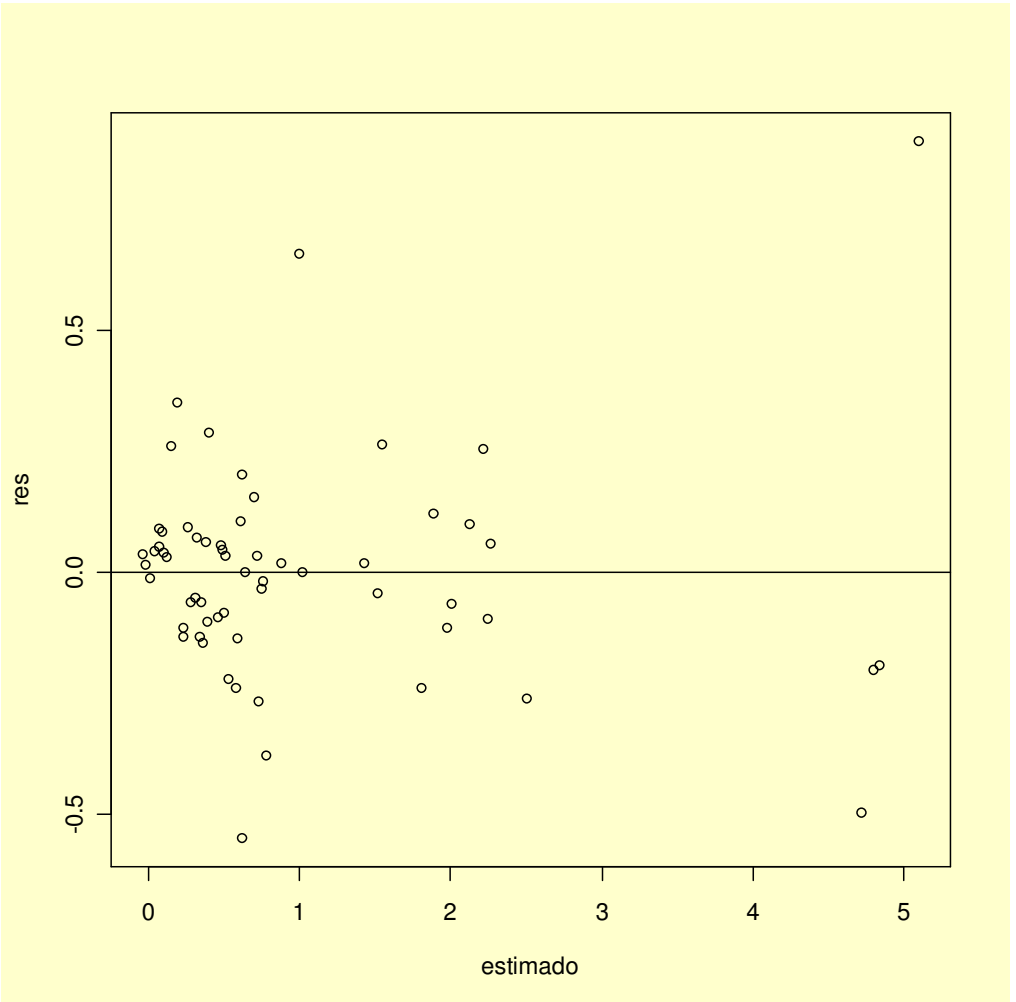
⇒ Há réplicas?

ANOVA Clássica

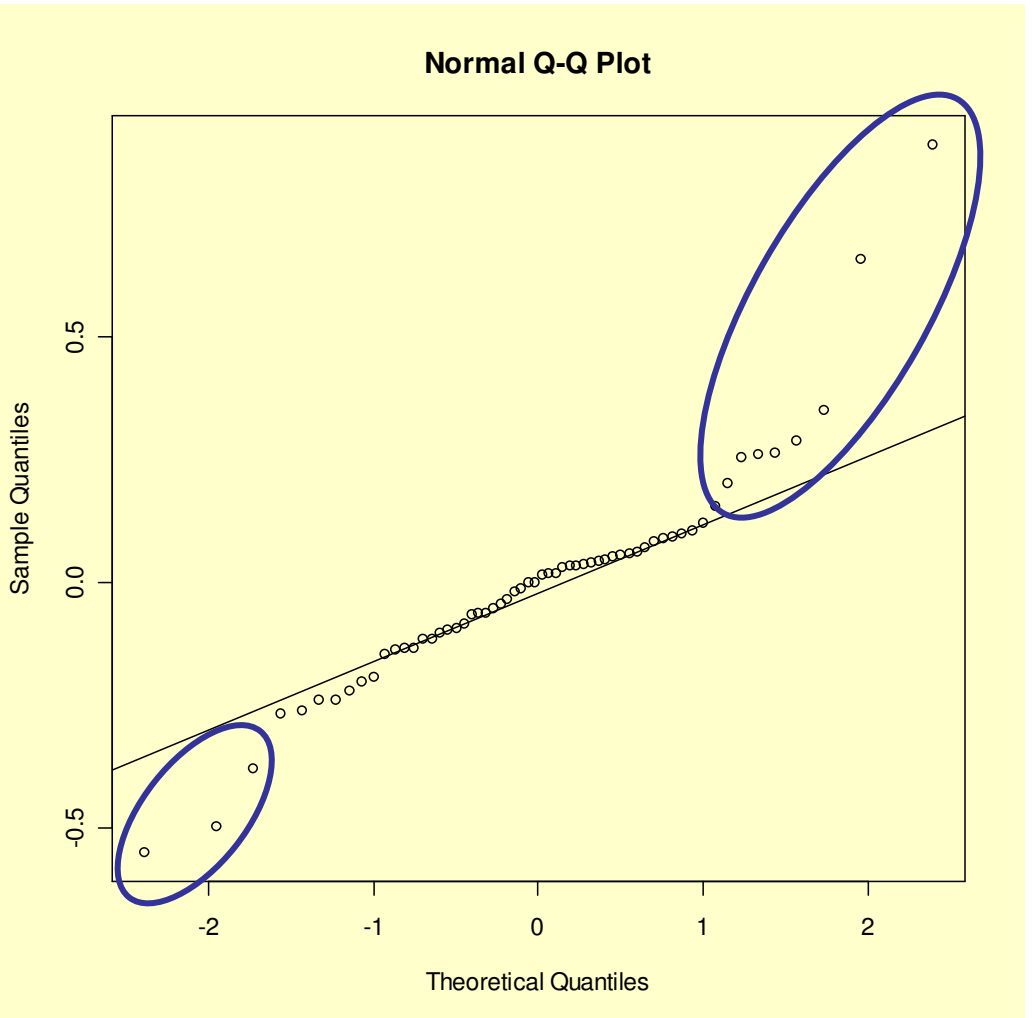
- Tabela de ANOVA

	DF	SQ	SQM	F	P
LINHA	14	88,69	6,33	87,94	0,000
COLUNA	3	1,21	0,40	5,61	0,003
ERRO	42	3,03	0,07		
TOTAL	59	92,92			

Análise de Resíduos



Análise de Resíduos



Análise de Tabelas de Duas Entradas por Medianas

- "Median Polish" é uma técnica da análise de dados de experimentos fatoriais mais robusta do que a ANOVA

- É utilizada em modelos aditivos de tabelas 2X2 $\Rightarrow$ pode ser generalizada para incluir efeitos de interação

- Este procedimento é similar à ANOVA, no entanto usa-se **valores das medianas em vez das médias**, assim adiciona-se robustez para o controle dos efeitos de *outliers*

Análise de Tabelas de Duas Entradas por Medianas

Dados de um experimento com dois fatores (sem replicação)

i	j		
	1	…	J
1	$y_{1,1}$	…	$y_{1,J}$
⋮	⋮	⋱	⋮
I	$y_{I,1}$	…	$y_{I,J}$

Modelos Aditivos $$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

Notação Alternativa (Hoaglin et al., 1983):

$$y_{ij} = m + a_i + b_j + e_{ij}$$

Análise de Tabelas de Duas Entradas por Medianas

Dados de um experimento com dois fatores (sem replicação)

i	j		
	1	…	J
1	$y_{1,1}$	…	$y_{1,J}$
⋮	⋮	⋱	⋮
I	$y_{I,1}$	…	$y_{I,J}$

Modelo Aditivo: $y_{ij} = m + a_i + b_j + e_{ij}$

Ajuste por processo Iterativo:

$$y_{ij} = m^{(0)} + a_i^{(0)} + b_j^{(0)} + e_{ij}^{(0)}$$

$$y_{ij} = m^{(n)} + a_i^{(n)} + b_j^{(n)} + e_{ij}^{(n)}$$

Análise de Tabelas de Duas Entradas por Medianas

Modelo ANOVA com Interação:

$$y_{ij} = m + a_i + b_j + \gamma_{ij} + e_{ij}$$

$$y_{ij} = m + a_i + b_j + \frac{a_i \times b_j}{m} + e_{ij}$$

$$y_{ij} = m \times \left(1 + \frac{a_i}{m}\right) \times \left(1 + \frac{b_j}{m}\right) + e_{ij}$$

Análise de Tabelas de Duas Entradas por Medianas

Dados de um experimento com dois fatores (sem replicação)

i	j		
	1	…	J
1	$y_{1,1}$	…	$y_{1,J}$
⋮	⋮	⋱	⋮
I	$y_{I,1}$	…	$y_{I,J}$

Primeiro Passo: $$y_{ij} = m^{(0)} + a_i^{(0)} + b_j^{(0)} + e_{ij}^{(0)}$$

$$m^{(0)} = 0 \qquad a_i^{(0)} = 0 \quad i = 1,\ldots,I$$

$$b_j^{(0)} = 0 \quad j = 1,\ldots,J$$

Análise de Tabelas de Duas Entradas por Medianas Processo Iterativo

Linhas:

$$\Delta a_i^{(n)} = med\{e_{ij}^{(n-1)} \,|\, j = 1,...,J\} \;\; ; \qquad \text{i=1,...,I}$$

$$\Delta m_b^{(n)} = med\{b_j^{(n-1)} \,|\, j = 1,...,J\} \;\; ;$$

$$d_{ij}^{(n)} = e_{ij}^{(n-1)} - \Delta a_i^{(n)} \;\; ; \qquad \text{j=1,...,J} \qquad \text{i=1,...,I}$$

Colunas:

$$\Delta b_j^{(n)} = med\{d_{ij}^{(n)} \,|\, i = 1,...,I\} \;\; ; \qquad \text{j=1,...,j}$$

$$\Delta m_a^{(n)} = med\{a_i^{(n-1)} + \Delta a_i^{(n)} \,|\, i = 1,...,I\} \;\; ;$$

$$e_{ij}^{(n)} = d_{ij}^{(n)} - \Delta b_j^{(n)} \;\; ; \qquad \text{i=1,...,I} \qquad \text{j=1,...,J}$$

Análise de Tabelas de Duas Entradas por Medianas Processo Iterativo

Iterações na linha:

$$\Delta a_i^{(n)} = med\{e_{ij}^{(n-1)} \,|\, j = 1,\ldots,J\}$$

$$\Delta m_b^{(n)} = med\{b_j^{(n-1)} \,|\, j = 1,\ldots,J\}$$

$$d_{ij}^{(n)} = e_{ij}^{(n-1)} - \Delta a_i^{(n)}$$

Iterações na coluna:

$$\Delta b_j^{(n)} = med\{d_{ij}^{(n)} \,|\, i = 1,\ldots,I\}$$

$$\Delta m_a^{(n)} = med\{a_i^{(n-1)} + \Delta a_i^{(n)} \,|\, i = 1,\ldots,I\}$$

$$e_{ij}^{(n)} = d_{ij}^{(n)} - \Delta b_j^{(n)}$$

Valores Comuns e Efeitos:

$$m^{(n)} = m^{(n-1)} + \Delta m_a^{(n)} + \Delta m_b^{(n)}$$

$$a_i^{(n)} = a_i^{(n-1)} + \Delta a_i^{(n)} - \Delta m_a^{(n)} \qquad i = 1,\ldots,I$$

$$b_j^{(n)} = b_j^{(n-1)} - \Delta m_b^{(n)} + \Delta b_j^{(n)} \qquad j = 1,\ldots,J$$

Processo Iterativo

Ajuste do Modelo de ANOVA por Medianas

Linha na iteração *n:*

i	j: 1	…	J	Nova Mediana	Prev
1	$e_{1,1}^{(n-1)}$	…	$e_{1,J}^{(n-1)}$	$[\Delta a_1^{(n)}]$	$a_1^{(n-1)}$
⋮	⋮	⋱	⋮	⋮	⋮
I	$e_{I,1}^{(n-1)}$	…	$e_{I,J}^{(n-1)}$	$[\Delta a_I^{(n)}]$	$a_I^{(n-1)}$
Prev	$b_1^{(n-1)}$	…	$b_j^{(n-1)}$	$[\Delta m_b^{(n)}]$	$m^{(n-1)}$

Coluna na iteração *n:*

i	j: 1	…	J	Prev
1	$d_{1,1}^{(n)}$	…	$d_{1,J}^{(n)}$	$a_1^{(n-1)} + \Delta a_1^{(n)}$
⋮	⋮	⋱	⋮	⋮
I	$d_{I,1}^{(n)}$	…	$d_{I,J}^{(n)}$	$a_I^{(n-1)} + \Delta a_I^{(n)}$
Nova Mediana	$\Delta b_1^{(n)}$	…	$\Delta b_j^{(n)}$	$[\Delta m_a^{(n)}]$
Prev	$b_1^{(n-1)} - \Delta m_b^{(n)}$	…	$b_j^{(n-1)} - \Delta m_b^{(n)}$	$m^{(n-1)} + \Delta m_b^{(n)}$

Ajuste do Modelo ANOVA por Medianas "Median Polish"

- Dados de Causas de Doenças

	não fuma	1_14	15-25	25+	mediana linha
Câncer					
Pulmão	0,070	0,470	0,860	1,660	**0,665**
Respiratório Superior	0,000	0,130	0,090	0,210	**0,110**
Estomago	0,410	0,360	0,100	0,310	**0,335**
Colon e Reto	0,440	0,540	0,370	0,740	**0,490**
Próstata	0,550	0,260	0,220	0,340	**0,300**
Outros tipos	0,640	0,720	0,760	1,020	**0,740**
Doenças Respiratórias					
Pulmonar	0,000	0,160	0,180	0,290	**0,170**
Bronquite	0,120	0,290	0,390	0,720	**0,340**
Outras	0,690	0,550	0,540	0,400	**0,545**
Trombose Coronária	4,220	4,640	4,600	5,990	**4,620**
Outras - cardiovascular	2,230	2,150	2,470	2,250	**2,240**
Hemorragia Cerebral	2,010	1,940	1,860	2,330	**1,975**
Ulcera Péptica	0,000	0,140	0,160	0,220	**0,150**
Violência	0,420	0,820	0,450	0,900	**0,635**
Outras Doenças	1,450	1,810	1,470	1,570	**1,520**

	não fuma	1_14	15-25	25+	mediana linha
Câncer					
Pulmão	0,070	0,470	0,860	1,660	**0,665**
Respiratório Superior	0,000	0,130	0,090	0,210	**0,110**
Estomago	0,410	0,360	0,100	0,310	**0,335**
Colon e Reto	0,440	0,540	0,370	0,740	**0,490**
Próstata	0,550	0,260	0,220	0,340	**0,300**
Outros tipos	0,640	0,720	0,760	1,020	**0,740**
Doenças Respiratórias					
Pulmonar	0,000	0,160	0,180	0,290	**0,170**
Bronquite	0,120	0,290	0,390	0,720	**0,340**
Outras	0,690	0,550	0,540	0,400	**0,545**
Trombose Coronária	4,220	4,640	4,600	5,990	**4,620**
Outras - cardiovascular	2,230	2,150	2,470	2,250	**2,240**
Hemorragia Cerebral	2,010	1,940	1,860	2,330	**1,975**
Ulcera Péptica	0,000	0,140	0,160	0,220	**0,150**
Violência	0,420	0,820	0,450	0,900	**0,635**
Outras Doenças	1,450	1,810	1,470	1,570	**1,520**

-0,595 = 0,070-(0,665)

	não fuma	1_14	15-25	25+	prev
Câncer					
Pulmão	**-0,595**	-0,195	0,195	0,995	0,665
Respiratório Superior	-0,110	0,020	-0,020	0,100	0,110
Estomago	0,075	0,025	-0,235	-0,025	0,335
Colon e Reto	-0,050	0,050	-0,120	0,250	0,490
Próstata	0,250	-0,040	-0,080	0,040	0,300
Outros tipos	-0,100	-0,020	0,020	0,280	0,740
Doenças Respiratórias					
Pulmonar	-0,170	-0,010	0,010	0,120	0,170
Bronquite	-0,220	-0,050	0,050	0,380	0,340
Outras	0,145	0,005	-0,005	-0,145	0,545
Trombose Coronária	-0,400	0,020	-0,020	1,370	4,620
Outras - cardiovascular	-0,010	-0,090	0,230	0,010	2,240
Hemorragia Cerebral	0,035	-0,035	-0,115	0,355	1,975
Ulcera Péptica	-0,150	-0,010	0,010	0,070	0,150
Violência	-0,215	0,185	-0,185	0,265	0,635
Outras Doenças	-0,070	0,290	-0,050	0,050	1,520
mediana col	**-0,100**	**-0,010**	**-0,020**	**0,120**	**0,545**

prev: dados do ajuste prévio

-0,030 = -0,040-(-0,010)

0,120 = 0,665-(0,545)

	não fuma	1_14	15-25	25+	prev
Câncer					
Pulmão	-0,595	-0,195	0,195	0,995	0,665
Respiratório Superior	-0,110	0,020	-0,020	0,100	0,110
Estomago	0,075	0,025	-0,235	-0,025	0,335
Colon e Reto	-0,050	0,050	-0,120	0,250	0,490
Próstata	0,250	-0,040	-0,080	0,040	0,300
Outros tipos	-0,100	-0,020	0,020	0,280	0,740
Doenças Respiratórias					
Pulmonar	-0,170	-0,010	0,010	0,120	0,170
Bronquite	-0,220	-0,050	0,050	0,380	0,340
Outras	0,145	0,005	-0,005	-0,145	0,545
Trombose Coronária	-0,400	0,020	-0,020	1,370	4,620
Outras - cardiovascular	-0,010	-0,090	0,230	0,010	2,240
Hemorragia Cerebral	0,035	-0,035	-0,115	0,355	1,975
Ulcera Péptica	-0,150	-0,010	0,010	0,070	0,150
Violência	-0,215	0,185	-0,185	0,265	0,635
Outras Doenças	-0,070	0,290	-0,050	0,050	1,520
mediana col	**-0,100**	**-0,010**	**-0,020**	**0,120**	**0,545**

	não fuma	1_14	15-25	25+	mediana lin	prev
Câncer						
Pulmão	-0,495	-0,185	0,215	0,875	**0,015**	**0,120**
Respiratório Superior	-0,010	0,030	0,000	-0,020	**-0,005**	-0,435
Estomago	0,175	0,035	-0,215	-0,145	**-0,055**	-0,210
Colon e Reto	0,050	0,060	-0,100	0,130	**0,055**	-0,055
Próstata	0,350	**-0,030**	-0,060	-0,080	**-0,045**	-0,245
Outros tipos	0,000	-0,010	0,040	0,160	**0,020**	0,195
Doenças Respiratórias						
Pulmonar	-0,070	0,000	0,030	0,000	**0,000**	-0,375
Bronquite	-0,120	-0,040	0,070	0,260	**0,015**	-0,205
Outras	0,245	0,015	0,015	-0,265	**0,015**	0,000
Trombose Coronária	-0,300	0,030	0,000	1,250	**0,015**	4,075
Outras - cardiovascular	0,090	-0,080	0,250	-0,110	**0,005**	1,695
Hemorragia Cerebral	0,135	-0,025	-0,095	0,235	**0,055**	1,430
Ulcera Péptica	-0,050	0,000	0,030	-0,050	**-0,025**	-0,395
Violência	-0,115	0,195	-0,165	0,145	**0,015**	0,090
Outras Doenças	0,030	0,300	-0,030	-0,070	**0,000**	0,975
prev	-0,100	-0,010	-0,020	0,120	**-0,015**	0,545

-0,005 = -0,010-(-0,005)

0,005 = -0,010-(-0,015)

-0,265 = 0,055+(-0,210)

	não fuma	1_14	15-25	25+	mediana	prev
Câncer						
Pulmão	-0,495	-0,185	0,215	0,875	**0,015**	0,120
Respiratório Superior	-0,010	0,030	0,000	-0,020	**-0,005**	-0,435
Estomago	0,175	0,035	-0,215	-0,145	**-0,055**	-0,210
Colon e Reto	0,050	0,060	-0,100	0,130	**0,055**	-0,055
Próstata	0,350	-0,030	-0,060	-0,080	**-0,045**	-0,245
Outros tipos	0,000	-0,010	0,040	0,160	**0,020**	0,195
Doenças Respiratórias						
Pulmonar	-0,070	0,000	0,030	0,000	**0,000**	-0,375
Bronquite	-0,120	-0,040	0,070	0,260	**0,015**	-0,205
Outras	0,245	0,015	0,015	-0,265	**0,015**	0,000
Trombose Coronária	-0,300	0,030	0,000	1,250	**0,015**	4,075
Outras - cardiovascular	0,090	-0,080	0,250	-0,110	**0,005**	1,695
Hemorragia Cerebral	0,135	-0,025	-0,095	0,235	**0,055**	1,430
Ulcera Péptica	-0,050	0,000	0,030	-0,050	**-0,025**	-0,395
Violência	-0,115	0,195	-0,165	0,145	**0,015**	0,090
Outras Doenças	0,030	0,300	-0,030	-0,070	**0,000**	0,975
prev	-0,100	-0,010	-0,020	0,120	**-0,015**	0,545

	não fuma	1_14	15-25	25+	prev
Câncer					
Pulmão	-0,510	-0,200	0,200	0,860	0,135
Respiratório Superior	**-0,005**	0,035	0,005	-0,015	-0,440
Estomago	0,230	0,090	-0,160	-0,090	**-0,265**
Colon e Reto	-0,005	0,005	-0,155	0,075	0,000
Próstata	0,395	0,015	-0,015	-0,035	-0,290
Outros tipos	-0,020	-0,030	0,020	0,140	0,215
Doenças Respiratórias					
Pulmonar	-0,070	0,000	0,030	0,000	-0,375
Bronquite	-0,135	-0,055	0,055	0,245	-0,190
Outras	0,230	0,000	0,000	-0,280	0,015
Trombose Coronária	-0,315	0,015	-0,015	1,235	4,090
Outras - cardiovascular	0,085	-0,085	0,245	-0,115	1,700
Hemorragia Cerebral	0,080	-0,080	-0,150	0,180	1,485
Ulcera Péptica	-0,025	0,025	0,055	-0,025	-0,420
Violência	-0,130	0,180	-0,180	0,130	0,105
Outras Doenças	0,030	0,300	-0,030	-0,070	0,975
mediana coluna	**-0,005**	**0,005**	**0,000**	**0,000**	**0,015**
prev	-0,085	0,005	-0,005	0,135	0,530

0,175 = 0,180-(0,005)

0,120 = 0,135-(0,015)

0,135= 0,000+(0,135)

	não fuma	1_14	15-25	25+	prev
Câncer					
Pulmão	-0,510	-0,200	0,200	0,860	0,135
Respiratório Superior	-0,005	0,035	0,005	-0,015	-0,440
Estomago	0,230	0,090	-0,160	-0,090	-0,265
Colon e Reto	-0,005	0,005	-0,155	0,075	0,000
Próstata	0,395	0,015	-0,015	-0,035	-0,290
Outros tipos	-0,020	-0,030	0,020	0,140	0,215
Doenças Respiratórias					
Pulmonar	-0,070	0,000	0,030	0,000	-0,375
Bronquite	-0,135	-0,055	0,055	0,245	-0,190
Outras	0,230	0,000	0,000	-0,280	0,015
Trombose Coronária	-0,315	0,015	-0,015	1,235	4,090
Outras - cardiovascular	0,085	-0,085	0,245	-0,115	1,700
Hemorragia Cerebral	0,080	-0,080	-0,150	0,180	1,485
Ulcera Péptica	-0,025	0,025	0,055	-0,025	-0,420
Violência	-0,130	0,180	-0,180	0,130	0,105
Outras Doenças	0,030	0,300	-0,030	-0,070	0,975
mediana	**-0,005**	**0,005**	**0,000**	**0,000**	**0,015**
prev	-0,085	0,005	-0,005	0,135	0,530

	não fuma	1_14	15-25	25+	efeito linha
Câncer					
Pulmão	-0,505	-0,205	0,200	0,860	**0,120**
Respiratório Superior	0,000	0,030	0,005	-0,015	**-0,455**
Estomago	0,235	0,085	-0,160	-0,090	**-0,280**
Colon e Reto	0,000	0,000	-0,155	0,075	**-0,015**
Próstata	0,400	0,010	-0,015	-0,035	**-0,305**
Outros tipos	-0,015	-0,035	0,020	0,140	**0,200**
Doenças Respiratórias					
Pulmonar	-0,065	-0,005	0,030	0,000	**-0,390**
Bronquite	-0,130	-0,060	0,055	0,245	**-0,205**
Outras	0,235	-0,005	0,000	-0,280	**0,000**
Trombose Coronária	-0,310	0,010	-0,015	1,235	**4,075**
Outras - cardiovascular	0,090	-0,090	0,245	-0,115	**1,685**
Hemorragia Cerebral	0,085	-0,085	-0,150	0,180	**1,470**
Ulcera Péptica	-0,020	0,020	0,055	-0,025	**-0,435**
Violência	-0,125	**0,175**	-0,180	0,130	**0,090**
Outras Doenças	0,035	0,295	-0,030	-0,070	**0,960**
efeito coluna	**-0,090**	**0,010**	**-0,005**	**0,135**	**0,545**

Resultado Final do Ajuste por Medianas do Modelo

	não fuma	1_14	15-25	25+	efeito linha
Câncer					
Pulmão	-0,505	-0,205	0,200	0,860	**0,120**
Respiratório Superior	0,000	0,030	0,005	-0,015	**-0,455**
Estomago	0,235	0,085	-0,160	-0,090	**-0,280**
Colon e Reto	0,000	0,000	-0,155	0,075	**-0,015**
Próstata	0,400	0,010	-0,015	-0,035	**-0,305**
Outros tipos	-0,015	-0,035	0,020	0,140	**0,200**
Doenças Respiratórias					
Pulmonar	-0,065	-0,005	0,030	0,000	**-0,390**
Bronquite	-0,130	-0,060	0,055	0,245	**-0,205**
Outras	0,235	-0,005	0,000	-0,280	**0,000**
Trombose Coronária	-0,310	0,010	-0,015	1,235	**4,075**
Outras - cardiovascular	0,090	-0,090	0,245	-0,115	**1,685**
Hemorragia Cerebral	0,085	-0,085	-0,150	0,180	**1,470**
Ulcera Péptica	-0,020	0,020	0,055	-0,025	**-0,435**
Violência	-0,125	0,175	-0,180	0,130	**0,090**
Outras Doenças	0,035	0,295	-0,030	-0,070	**0,960**
efeito coluna	**-0,090**	**0,010**	**-0,005**	**0,135**	**0,545**

- Efeito comum: 0.545 a cada 1000 homens

- –0.090 indica que há uma menor proporção de homens não fumantes que morrem por estas doenças.

- 0.135 nos mostra que os homens que fumam mais, morrem mais devido a estas doenças.

Resultado Final do Ajuste por Medianas do Modelo de ANOVA

- Modelo Aditivo (sem o termo de interação):

$$y_{ij} = m^{(n)} + a_i{}^{(n)} + b_j{}^{(n)} + e_{ij}{}^{(n)}$$

$$0{,}36_{32} = 0{,}545^{(2)} + (-0{,}280)^{(2)} + 0{,}010^{(2)} + 0{,}085^{(2)}$$

- Modelo com o termo de interação

$$y_{ij} = m^{(n)} + a_i{}^{(n)} + b_j{}^{(n)} + \frac{a_i \times b_j}{m}{}^{(n)} + e_{ij}{}^{(n)}$$

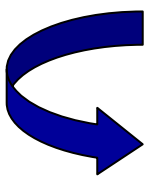
$$0{,}36_{32} = 0{,}545^{(2)} + (-0{,}280)^{(2)} + 0{,}010^{(2)} + \frac{(-0{,}280)\times 0{,}010}{0{,}545}{}^{(2)} + \left(0{,}085 - \frac{(-0{,}280)\times 0{,}010}{0{,}545}\right)^{(2)}$$

Ajuste do Modelo de ANOVA Aditivo usando “Median Polish” no R

- O algoritmo trabalha removendo a mediana da linha e da coluna, e continua até que a soma absoluta dos resíduos seja menor que '*eps*' ou até o número de iterações desejada, pré-estabelecida pelo usuário.
  - library(eda)
  - mediana<-medpolish(matriz,maxiter=***n***,eps=*0.01*)
- No R o método começa pela linha.
  - Para começar pela coluna é só modificar a matriz de entrada.
  - matriz<-rbind(matriz[,1] , matriz[,2], ... , matriz[,2])

ANOVA Robusta
Median Polish

- A principal vantagem deste método é a resistência aos valores *outliers*
- Produz bons resultados em tabelas com "*missing*"
- Converge de maneira rápida e aproximada para a soma dos resíduos absolutos
- Não possui as mesmas propriedades da análise com médias, porém na prática pode e deve ser usada como uma análise exploratória preliminar dos dados (Tukey, 1977).



Outras metodologias robustas:

- Análise (Clássica) dos dados transformados em Postos: cuidados na atribuição dos postos
- Obtenção de M-estimadores

Procedimentos de Inferência
ANOVA Clássica

$$Y_{N\times 1} = X_{N\times k}\ \beta_{k\times 1} + \varepsilon_{N\times 1}$$

$$\hat{\beta};\quad \min \hat{\varepsilon}'\,\hat{\varepsilon} = \min \left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)$$

$$\hat{\beta} = \left(X'\,X\right)^{-1} X'\,Y$$

$$\hat{Y} = \underbrace{X\,\left(X'\,X\right)^{-1} X'}\ Y$$

Método de Mínimos Quadrados e Verossimilhança (Distr. Normal) conduzem aos mesmos estimadores

H: Matriz de Projeção $\Rightarrow$ Solução Não Robusta

“Regressão” Robusta
Predição e Mínima Distância

- **Mínimos Quadrados (soluções na norma L_2)**

$$\hat{Y} = X\,\hat{\beta}; \quad \left\| Y - \hat{Y}_L \right\|^2 = \min_{\beta=\hat{\beta}} \sum_{ij} \varepsilon_{ij}^2$$

Soluções não robustas/resistentes para (um único) *outlier* em ambas direções, Y e X

- **Mínimos Quadrados “Aparados” (*trimmed*)**

$$\hat{Y} = X\,\hat{\beta}; \quad \min_{\beta=\hat{\beta}} \sum_{k=1}^{h} \left(\hat{\varepsilon}^2\right)_{(1:n)}; \qquad h < n$$

$\Rightarrow$h$\cong$ n/2 : Soluções com as “melhores” propriedades de robustez

“Regressão” Robusta
Predição e Mínima Distância

- **Mínimos Valores Absolutos (soluções na norma L_1)**

$$\hat{Y} = X\,\hat{\beta}; \quad \left\| Y - \hat{Y}_L \right\|^1 = \min_{\beta=\hat{\beta}} \sum_{ij} \left| \varepsilon_{ij} \right|$$

$\Rightarrow$Soluções robustas/resistentes para *outliers* na direção Y

$\Rightarrow$ Soluções “não”robustas/resistentes para (um único) *outlier* na direção X

$\Rightarrow$ O método “Median Polish” é robusto a valores aberrantes gerias e não precisa minimizar a soma dos resíduos absolutos. No entanto, em alguns casos, converge para tal resultado.

ANOVA Robusta
Predição e Mínima Distância

- **M-estimadores: minimizam uma função dos resíduos padronizados**

Huber, (1973)

$$\hat{Y} = X\,\hat{\beta}; \quad \min_{\beta=\hat{\beta}} \sum_{ij} \rho\left(\frac{\varepsilon_{ij}}{\hat{\sigma}}\right);$$

$$\rho(u) = \rho(-u) \text{ com um único mínimo em } 0 \qquad \rho(u) = \begin{cases} 1/2u^2 & |u| \le 1 \\ |u| - 1/2 & |u| > 1 \end{cases}$$

$$\Rightarrow \quad \sum_{ij} \psi\left(\frac{\hat{\varepsilon}_{ij}}{\hat{\sigma}}\right) X_{ij} = 0; \qquad \psi(u) = \rho'(u)$$

está associada uma certa projeção

$\Rightarrow$Mais eficientes assintoticamente que as soluções sob norma L_1

$\Rightarrow$ Soluções “não” robustas/resistentes na direção X (ponto de corte = 1/n)

ANOVA Robusta
Predição e Mínima Distância

- **M-estimadores: versão generalizada**

$$\Rightarrow \quad \sum_{ij} w(x_{ij})\,\psi\left(\frac{\hat{\varepsilon}_{ij}}{\hat{\sigma}}\right) x_{ij} = 0; \qquad \psi(u) = \rho'(u)$$

$$\Rightarrow \quad \sum_{ij} w(x_{ij})\,\psi\left(\frac{\hat{\varepsilon}_{ij}}{w(x_{ij})\,\hat{\sigma}}\right) x_{ij} = 0; \qquad \psi(u) = \rho'(u)$$

$\Rightarrow$Critério: garantir robustez/resistência e encontrar soluções inferenciais

$\Rightarrow$ Análise da Função de Influência e Análise das propriedades assintóticas dos M-estimadores

ANOVA Robusta
Predição e Mínima Distância

- **M-estimadores: versão generalizada**

$$\Rightarrow \quad \sum_{ij} w(x_{ij})\,\psi\left(\frac{\hat{\varepsilon}_{ij}}{\hat{\sigma}}\right) x_{ij} = 0; \qquad \psi(u) = \rho'(u)$$

$$\Rightarrow \quad \sum_{ij} w(x_{ij})\,\psi\left(\frac{\hat{\varepsilon}_{ij}}{w(x_{ij})\,\hat{\sigma}}\right) x_{ij} = 0; \qquad \psi(u) = \rho'(u)$$

$\Rightarrow$Critério: garantir robustez/resistência

$\Rightarrow$ Análise das propriedades assintóticas dos M-estimadores (sob condições de regularidade), realização de inferências, análises de resíduos

ANOVA Robusta

$$H_0 : C\beta = 0 \quad \times \quad H_1 : C\beta \neq 0$$

$\Rightarrow$Sob a solução de Mínimos Quadrados (ou premissas clássicas):

$$\Lambda = \left(L_0 / L_1\right) = e^{-D_0} \Big/ e^{-D_1}\,; \qquad D = \left\|\hat{\varepsilon}\right\|^2 = \hat{\varepsilon}'\,\hat{\varepsilon}$$

$$-2\ln\Lambda = 2\ \left(D_0 - D_1\right) \quad \sim \quad \chi^2_q$$

$\Rightarrow$Sob a solução de M-estimadores (e condições de regularidade):

$$-2\ln\Lambda = 2\,\tau^{-1}\left(D_0 - D_1\right) \quad \sim \quad \chi^2_q; \qquad \tau = E\left[\psi^2\left(\hat{\varepsilon}/\hat{\sigma}\right)\right] \Big/ E\left[\psi'\left(\hat{\varepsilon}/\hat{\sigma}\right)\right]$$

$$F_M = 2\left(D_0 - D_1\right)/\hat{\tau} \quad \sim \quad \chi^2_q$$

ANOVA Robusta

Delineamento Fatorial 3^4 (McKean and Schrader, 1982; John, 1978*)

Tabela ANOVA: valores da estatística F

	Mínimos Quadrados		Análise Robusta	
F.V.	Y	Y(j)*	Y	
Linear A	5.23[a]	13.80[b]	9.75[b]	
Quadrático A	0.79	9.82[b]	4.71[a]	
Linear B	0.25	5.18	1.52	$\Rightarrow F_M$
Quadrático B	0.42	3.63	2.14	
Linear C	0.97	2.56	1.69	
Quadrático C	0.14	5.27[a]	1.57	
Linear D	0.19	0.54	0.01	
Quadrático D	0.11	0.08	0.00	
Linear A * Linear B	0.06	0.17	0.00	
Linear A * Linear C	0.32	0.86	1.07	
Linear B * Linear C	2.68	7.09[a]	6.50[a]	

a: $p \leq 0.05$ *b: $p \leq 0.01$*

Y: dados originais Y(j): dados com uma estimativa da obs *outlier*

$\Rightarrow$ Análise de Diagnóstico sob ajustes robustos (!!)

Medidas de Diagnóstico

Modelo Clássico

$$\Rightarrow \qquad Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1} \qquad \Rightarrow \qquad \hat{Y} = X\,\hat{\beta} = X\left(X'X\right)^{-1}X'Y = H\,Y$$

- Identificação de pontos de alavanca (alto *leverage*):

$$\hat{y}_j = \sum_{j=1}^{n} x_j'\left(X'X\right)^{-1}x_j\,y_j \qquad \Rightarrow \qquad \hat{y}_j = h_{jj}\,y_j + \left(1-h_{jj}\right)X_j'\,\hat{\beta}_{(j)}$$

alavanca do valor ajustado $\left(h_{jj} > {}^{2p}/_{n}\right)$

- Identificação de pontos aberrantes:

$$t_j^* = \frac{\hat{\varepsilon}_j}{s_{(j)}\left(1-h_{jj}\right)^{1/2}} \sim t_{n-p-1}$$

resíduo studentizado (deletado)

- Identificação de pontos influentes (Cook):

$$D_j = \frac{\left(\hat{\beta}-\hat{\beta}_{(j)}\right)'X'X\left(\hat{\beta}-\hat{\beta}_{(j)}\right)}{p\,s^2} \quad > \quad F_{p,(n-p-1)}(1-\alpha)$$

Qual a distribuição destas medidas sob estimadores robustos ?

ANOVA Robusta

$\Rightarrow$ Métodos Robustos de análise: ajuste por medianas, transformação por postos, soluções aparadas, operadores de projeção mais gerais (M-estimadores)

$\Rightarrow$ Especificação de Modelos Robustos: adotar modelos distribucionais mais gerais para as observações (por exemplo a classe das distribuições elipticas)

Testes de Aleatorização

Considere conjuntos de dados amostrais gerados sob diferentes delineamentos experimentais:

$\Rightarrow$ Como os dados efetivamente observados podem ser usados para construir uma distribuição de referência empírica ?

$\Rightarrow$ E se os dados amostrais apresentarem observações aberrantes ?

$\Rightarrow$ Como atribuir postos às observações segundo diferentes delineamentos ?

Referência Bibliográfica

- Beckman, RJ; Natchtsheim, CJ and Cook, RD. (1987). Diagnostics for Mixed-Model Analysis of Variance. Technometrics 29(4):413:426.
- Box, G.E.; Hunter, W.G and Hunter, J.S. (1978). *Statistics for Experimenters. An Introduction to Designs, Data Analysis and Model Building*. John Wiley & Sons.
- Hoaglin, DC; Mosteller, F and Tukey, JW. (1983). Understanding robust and exploratory data analysis. Wiley.
- Launer,R.L. and Siegel, A.F. (1982). Modern Data Analysis. Academic Press.
- Lesaffre, E. and Verbeke, G.(1998). Local Influence in Linear Mixed Models. Biometrics 54:570-582.
- Neter, J. et al. (1996). *Applied Linear Statistical Models*. Irwin.
- Rousseeuw, P.J. and Leroy, A. (1987). Robust Regression and Outlier Detection. Wiley & Sons.
- Tukey, JW. (1977). ABC"s of EDA. Wiley.