

# Planejamento e Análise Estatística de Experimentos fatoriais: análise de dados de experimentos completamente aleatorizados - Parte 3

Prof. Caio Azevedo

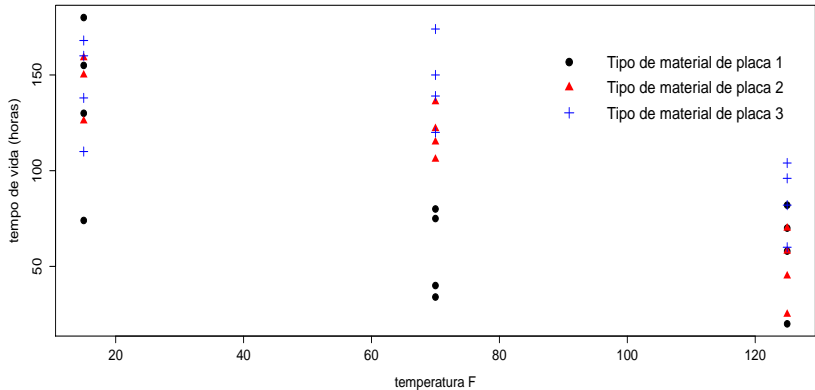
## Exemplo 4: Resistência de materiais (Continuação)

- Um engenheiro está desenvolvendo um tipo de bateria para ser usado em um dispositivo eletrônico sujeito à variações extremas de temperatura.
- Fatores de interesse:
  - Tipo de material da placa: 1, 2 e 3.
  - Temperatura: 15°F, 70°F e 125°F. Equivalente à -9,44°C, 21,11°C e 51,67 °C, respectivamente
- Para cada combinação (tipo de material da placa × temperatura) 4 baterias foram feitas.
- Variável resposta: tempo de vida em horas de cada bateria .

# Resultados oriundos do Modelo Fatorial com interação

- Identificamos um padrão de diferença entre as médias através da análise anterior.
- Podemos ajustar um modelo de regressão (linear ou quadrático) para quantificar o impacto no tempo de vida com o aumento dos níveis dos fatores.
- Naturalmente, podemos levar em consideração os resultados anterior para propor um modelo de regressão que leve em consideração o padrão de diferenças entre as médias.
- Entretanto, vamos começar com um modelo mais simples.

# Diagrama de Dispersão.

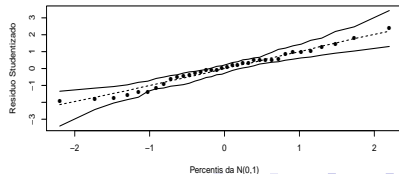
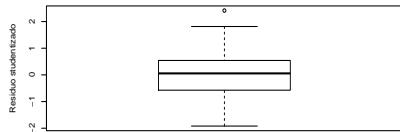
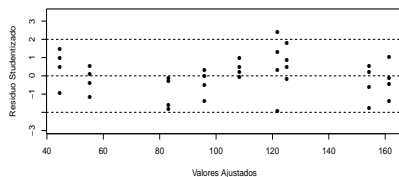
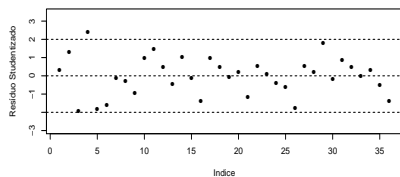


# Modelo linear 1: reta

$$Y_{ij} = \alpha_i + \beta_i x_{ij} + \xi_{ij}, i = 1, 2, 3, j = 1, \dots, 12$$

- $y_{ij}$  : tempo de vida (horas) do  $j$ -ésimo componente construído com o tipo de material  $i$  foi submetido.
- $x_{ij}$  : temperatura ao qual o  $j$ -ésimo componente construído com o tipo de material  $i$  foi submetido.
- $\alpha_i$  : valor esperado (média) do tempo de vida de componentes construídos com o tipo de material  $i$  submetidos à uma temperatura de  $0^\circ F$ .
- $\beta_i$  : incremento no valor esperado do tempo de vida componentes construídos com o tipo de material  $i$  quando a temperatura de submissão aumenta em  $1^\circ F$ .
- $\xi_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$ .

# Análise de resíduos



# Comentários

- Parece que as suposições do modelo não são válidas para o conjunto de dados em questão (embora o ajuste tenha melhorado em relação à situação anterior).
- Ausência de homocedasticidade e normalidade (leve).
- Uma alternativa: modelos de regressão com distribuição positiva e assimétrica para a variável resposta, que permita variâncias diferentes entre os grupos e com diferentes coeficientes de variação.
- Distribuições positivas: família gama (mãe não a tradicional), família normal inversa, família Weibull, família lognormal, família Birbaun-Saunders, normal assimétrica (apesar de ter suporte na reta)

# Comentários

- O modelo de regressão normal linear, aparentemente, não é adequado para analisar os dados em questão, apesar do ajuste ter melhorado em relação à situação anterior (considerando apenas dois fatores).
- Contudo, seguiremos com ele por questões pedagógicas.



## Estimativas dos parâmetros do modelo

Parâmetro	Estimativa	EP	IC(95%)	Estat. t	pvalor
$\alpha_1$	132,33	15,62	[ 101,71; 162,94]	8,47	<0,0001
$\alpha_2$	175,95	15,62	[145,33; 206,56]	11,27	<0,0001
$\alpha_3$	162,31	15,62	[131,70; 192,92]	10,39	<0,0001
$\beta_1$	-0,70	0,19	[-1,07 ; -0,33]	-3,74	0,0008
$\beta_2$	-0,97	0,19	[-1,34 ; -0,60]	-5,14	<0,0001
$\beta_3$	-0,53	0,19	[-0,90 ; -0,16]	-2,83	0,0082

Há uma aparente semelhança entre os interceptos (entre si) e os coeficientes angulares (entre si).

# Hipóteses: igualdade entre os interceptos e os coeficiente angulares

- Temos que  $\beta = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3)$
- Interceptos:

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 \leftrightarrow \begin{cases} \alpha_1 = \alpha_2 \\ \alpha_1 = \alpha_3 \end{cases}$$

- Coeficientes angulares:

$$H_0 : \beta_1 = \beta_2 = \beta_3 \leftrightarrow \begin{cases} \beta_1 = \beta_2 \\ \beta_1 = \beta_3 \end{cases}$$

# Matrizes C

- Interceptos

$$C = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \end{bmatrix}$$

- Coeficiente angulares:

$$C = \begin{bmatrix} 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}$$

# Resultados dos testes para as as hipóteses anteriores

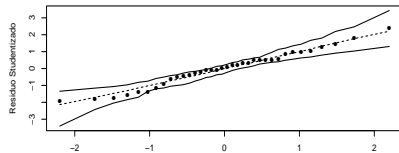
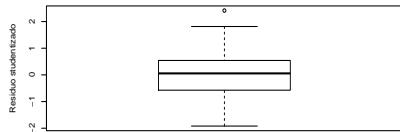
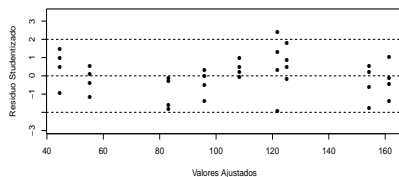
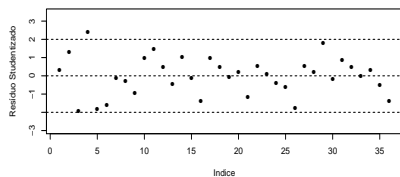
- Hipótese (1): 2,04 (pvalor = 0,1475).
- Hipótese (2): 1,36 (pvalor = 0,2730).

# Modelo reduzido

$$Y_{ij} = \alpha + \beta x_{ij} + \xi_{ij}, i = 1, 2, 3, j = 1, \dots, 12$$

- $y_{ij}$  : tempo de vida (horas) do  $j$ -ésimo componente construído com o tipo de material  $i$  foi submetido.
- $x_{ij}$  : temperatura ao qual o  $j$ -ésimo componente construído com o tipo de material  $i$  foi submetido.
- $\alpha$  : valor esperado (média) do tempo de vida de componentes construídos com qualquer tipo de material (entre os três considerados), submetidos à uma temperatura de  $0^\circ F$ .
- $\beta$  : incremento no valor esperado do tempo de vida componentes construídos com qualquer tipo de material (entre os três considerados), quando a temperatura de submissão aumenta em  $1^\circ F$ .

# Análise de resíduos: modelo reduzido



# Comentários

- Parece que as suposições do modelo não são válidas para o conjunto de dados em questão (embora o ajuste tenha melhorado em relação à situação anterior).
- Ausência de homocedasticidade e normalidade (leve).
- Uma alternativa: modelos de regressão com distribuição positiva e assimétrica para a variável resposta, que permita variâncias diferentes entre os grupos e com diferentes coeficientes de variação.
- Distribuições positivas: família gama (mãe não a tradicional), família normal inversa, família Weibull, família lognormal, família Birbaun-Saunders, normal assimétrica (apesar de ter suporte na reta)

# Comentários

- O modelo de regressão normal linear, aparentemente, não é adequado para analisar os dados em questão, apesar do ajuste ter melhorado em relação à situação anterior (considerando apenas dois fatores).
- Contudo, seguiremos com ele por questões pedagógicas.



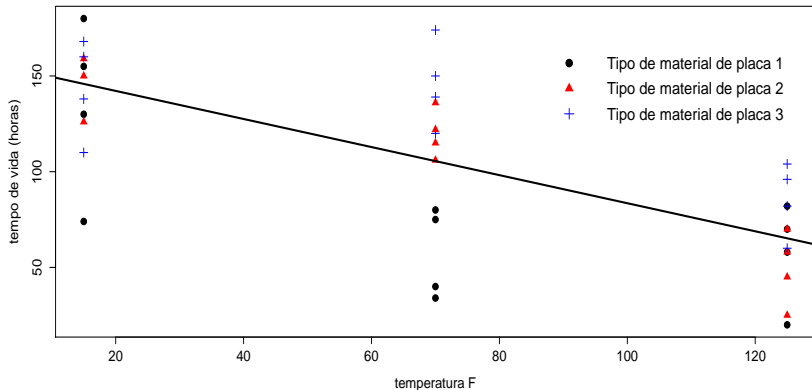
# Estimativas dos parâmetros do modelo

Parâmetro	Estimativa	EP	IC(95%)	Estat. t	pvalor
$\alpha$	156,86	10,40	[136,48;177,25]	15,08	< 0,0001
$\beta$	-0,73	0,13	[-0,98;-0,49]	-5,86	< 0,0001

Ambos os parâmetros são estatisticamente diferentes de zero.

Independentemente do tipo de material, o tempo esperado de vida sob  $0^{\circ}F$  bem como o decaimento com o aumento da temperatura, são os mesmos, independentemente do tipo de material utilizado. Exercício: ajustar um modelo quadrático com diferentes coeficientes.

# Modelo ajustado



# PCA desbalanceado com dois fatores

- Considere um experimento desbalanceado com dois fatores e um número geral de níveis.
- Os conceitos permanecem inalterados bem como as formas de análise.
- Mudanças nas fórmulas das somas de quadrados.
- Desbalanceamento desproporcional pode comprometer a ortogonalidade das somas de quadrados.

# Descrição

- Fator A: possui  $a$  níveis,  $i=1,\dots,a$ .
- Fator B: possui  $b$  níveis,  $j=1,\dots,b$ .
- Grupos: há um total de  $a \times b$  grupos (tratamentos), que são definidos pelas interseções dos níveis de cada grupo.
- Para cada grupos vamos considerar um total de  $n_{ij}$  observações. Cada uma das  $n_{ij}$  observações são alocadas aleatoriamente à cada uma das combinações (fatores). Temos uma PCA (planejamento completamente casualizado).
- Neste caso número total de observações  $n = \sum_{i=1}^a \sum_{j=1}^b n_{ij}$ .

# Modelo com interação (casela de referência)

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \xi_{ijk},$$

(Fator A),  $i = 1, \dots, a$ ; (Fator B),  $j =$

$1, \dots, b$ ; (unidades experimentais),  $k = 1, \dots, n_{ij}$

- Erros  $\xi_{ijk} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ ,  $\mu, \alpha_i, \beta_j, (\alpha\beta)_{ij}$  não aleatórios.
- Restrições :  $\alpha_1 = \beta_1 = (\alpha\beta)_{1j} = (\alpha\beta)_{i1} = 0, \forall i, j.$

# Modelo sem interação (casela de referência)

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \xi_{ijk},$$

(Fator A),  $i = 1, \dots, a$ ; (Fator B),  $j =$

$1, \dots, b$ ; (unidades experimentais),  $k = 1, \dots, n_{ij}$

- Erros  $\xi_{ijk} \stackrel{i.i.d}{\sim} N(0, \sigma^2)$ ,  $\mu, \alpha_i, \beta_j, (\alpha\beta)_{ij}$  não aleatórios.
- Restrições :  $\alpha_1 = \beta_1 = (\alpha\beta)_{1j} = (\alpha\beta)_{i1} = 0, \forall i, j.$

# Somas de quadrados

- Decomposição da soma de quadrados total:

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

# Tabela de análise de variância

- Para testar a igualdade simultânea das médias

FV	SQ	GL	QM	Estatística F	pvalor
Fator A	$SQF_A$	a-1	$QMF_A = \frac{SQF_A}{(a-1)}$	$F_A = \frac{QMF_A}{QMR}$	$\min(F(f_A H_0), S(f_A H_0))$
Fator B	$SQF_B$	b-1	$QMF_B = \frac{SQF_B}{(b-1)}$	$F_B = \frac{QMF_B}{QMR}$	$\min(F(f_B H_0), S(f_B H_0))$
Interação	SQInt	(a-1)(b-1)	$QMInt = \frac{SQInt}{[(a-1)(b-1)]}$	$F_{Int} = \frac{QMInt}{QMR}$	$\min(F(f_{Int} H_0), S(f_{Int} H_0))$
Resíduo	SQR	ab(n-1)	$QMR = \frac{SQR}{[ab(n-1)]}$		
Total	SQT	abn-1			

FV: fonte de variação, SQ: soma de quadrados, GL: graus de liberdade, QM: quadrado médio.  $F(x|H_0)$ ,  $S(x|H_0)$  fda e fds no ponto x sob  $H_0$ , respectivamente.



# Obtenção das somas de quadrados de forma matricial

- De forma semelhante ao caso de um único fator com vários níveis, as somas de quadrados associadas à cada fator (fatores principais e interação), é obtida através das diferenças das somas de quadrados de resíduos de um modelo sem o fator em questão e com o fator em questão (excluindo-se os fatores de “ordem” superior).
- Lembrando que, a forma geral da SQR é dada por  $SQR = Y' (I - H) Y$  e, em que  $H = X (X'X)^{-1} X'$

## Obtenção das SQ's sob a forma matricial (cont.)

- No caso de um experimento com dois fatores temos os seguintes modelos possíveis:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \xi_{ijk}, \text{ (Modelo 1)}$$

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \xi_{ijk}, \text{ (Modelo 2)}$$

$$Y_{ijk} = \mu + \beta_j + \xi_{ijk}, \text{ (Modelo 3)}$$

$$Y_{ijk} = \mu + \alpha_i + \xi_{ijk}, \text{ (Modelo 4)}$$

- Exercício: escrever a matriz  $X$  para cada modelo acima. Elas serão nomeadas  $X_1$ ,  $X_2$  e  $X_3$ , respectivamente

# Obtenção das somas de quadrados de forma matricial (cont.)

- Seja  $SQR_i$  associada ao modelo  $i$  e  $H_i = X_i (X_i' X_i)^{-1} X_i'$ .
- Assim, temos

$$SQF_A = SQR_3 - SQR_1 = Y'(H_2 - H_3)Y$$

$$SQF_B = SQR_4 - SQR_1 = Y'(H_2 - H_4)Y$$

$$SQInt = SQR_2 - SQR_1 = Y'(H_1 - H_2)Y$$

# Esperanças dos Quadrados Médios

- Expressões:

$$\mathcal{E}(QMF_A) = \sigma^2 + \frac{\sum_{j=1}^b n_{ij}(\bar{\mu}_{i.} - \bar{\mu}_{..})^2}{a-1}$$

$$\mathcal{E}(QMF_B) = \sigma^2 + \frac{\sum_{i=1}^a n_{ij}(\bar{\mu}_{.j} - \bar{\mu}_{..})^2}{a-1}$$

$$\mathcal{E}(QMF_{Int}) = \sigma^2 + \frac{\sum_{i=1}^a \sum_{j=1}^b n_{ij}[(\mu_{ij} + \bar{\mu}_{..}) - (\bar{\mu}_{i.} + \bar{\mu}_{.j})]^2}{a-1}$$

$$\mathcal{E}(QMR) = \sigma^2$$

## Exemplo 5: O problema de engarrafamento de refrigerantes

- Uma empresa está interessada que a quantidade de refrigerante colocada em cada garrafa seja mais uniforme entre os vasilhames.
- Fatores de interesse:
  - Percentual de carbonatação (CARB): 10%, 12% e 14%.
  - Pressão de operação no enchimento (PRE): 25 e 30 psi.
  - Velocidade na linha de produção (VELOC): 200 e 250 bpm.
- Para cada combinação dos fatores (temos um total de  $2 \times 2 \times 3 = 12$  tratamentos) foram medidas as diferenças entre o a quantidade de refrigerante inserida no vasilhame menos o valor padrão de dois refrigerantes escolhidos ao acaso.

## Exemplo 5: continuação

- Variável resposta: diferença de preenchimento (DIP) .
- Planejamento completamente aleatorizado (PCA) balanceado, com 3 fatores, sendo dois com 2 níveis e 1 com três níveis.
- Fator A: CARB
- Fator B: PRE.
- Fator C: VELOC.
- Temos um total de 12 tratamentos.

## Modelo (casela de referência)

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \xi_{ijkl},$$

(Fator A),  $i = 1, 2, 3$ ; (Fator B),  $j = 1, 2$ ; (Fator C),  $k = 1, 2$ ;

(unidades experimentais),  $l = 1, 2$

- Erros  $\xi_{ijkl} \stackrel{i.i.d}{\sim} N(0, \sigma^2)$ ,  $\mu, \alpha_i, \beta_j, \gamma_k, (\alpha\beta)_{ij}, (\alpha\gamma)_{ik}, (\beta\gamma)_{jk}, (\alpha\beta\gamma)_{ijk}$  não aleatórios.
- Restrições :  $\alpha_1 = \beta_1 = \gamma_1 = (\alpha\beta)_{1j} = (\alpha\beta)_{i1} = (\alpha\gamma)_{i1} = (\alpha\gamma)_{1k} = (\beta\gamma)_{j1} = (\beta\gamma)_{1k} = (\alpha\beta\gamma)_{i11} = (\alpha\beta\gamma)_{1j1} = (\alpha\beta\gamma)_{11k} = 0, \forall i, j.$

# Interpretações dos parâmetros

- Agora tem-se: interações de segunda ordem (3 fatores simultaneamente), interação de primeira ordem (entre pares de fatores) e efeitos dos fatores principais.
- Se  $(\alpha\beta\gamma)_{ijk} = 0 \forall i, j, k$  não há interação de segunda ordem.
- Em não havendo interação de segunda ordem
  - Se  $(\alpha\beta)_{ij} = 0 \forall i, j$  não há interação de primeira ordem entre os Fatores A e B.
  - Se  $(\alpha\gamma)_{ik} = 0 \forall i, k$  não há interação de primeira ordem entre os Fatores A e C.
  - Se  $(\beta\gamma)_{jk} = 0 \forall j, k$  não há interação de primeira ordem entre os Fatores B e C.

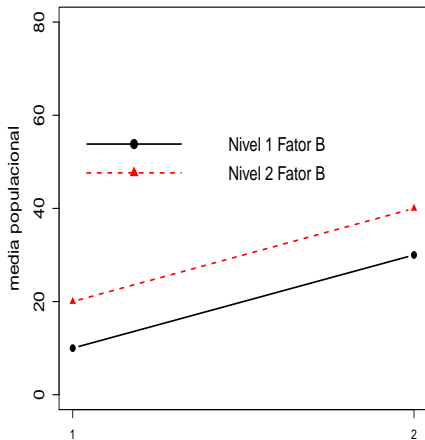


# Interpretações dos parâmetros (cont.)

- Em não havendo interação nenhuma interação de primeira ordem
  - Se  $(\alpha)_{ij} = 0 \forall i$  não há efeito do fator principal A.
  - Se  $(\beta)_j = 0 \forall j$  não há efeito do fator principal B.
  - Se  $(\gamma)_k = 0 \forall k$  não há efeito do fator principal C.

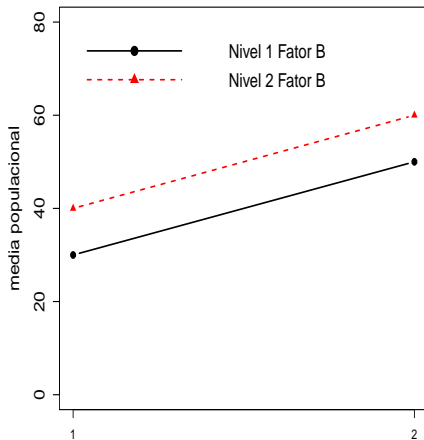
# Perfis médios: ausência de interação de segunda ordem

Nível 1 do Fator C



Fator A

Nível 2 do Fator C

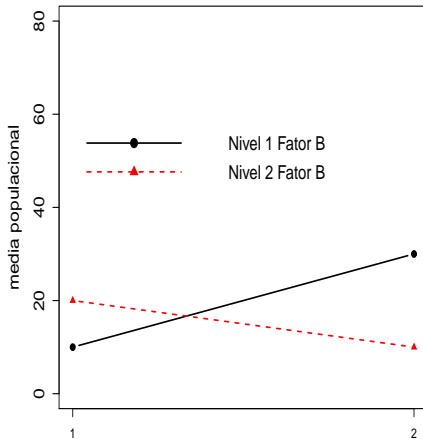


Fator A



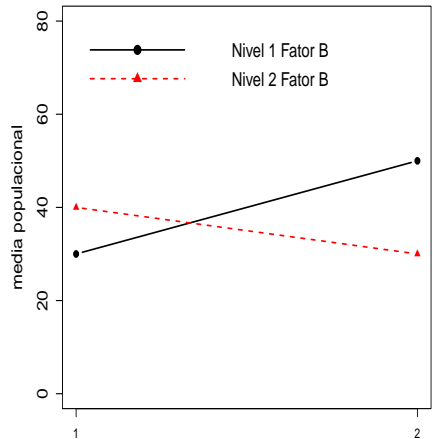
# Perfis médios: ausência de interação de segunda ordem

Nível 1 do Fator C



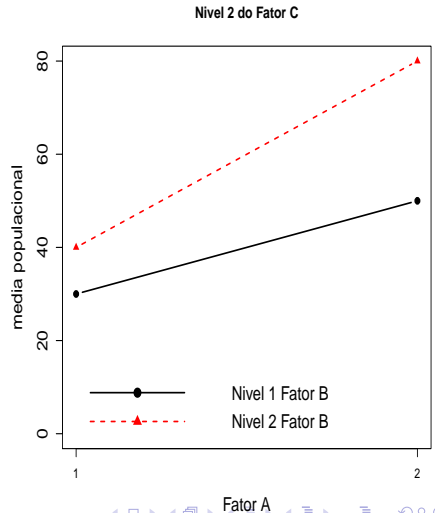
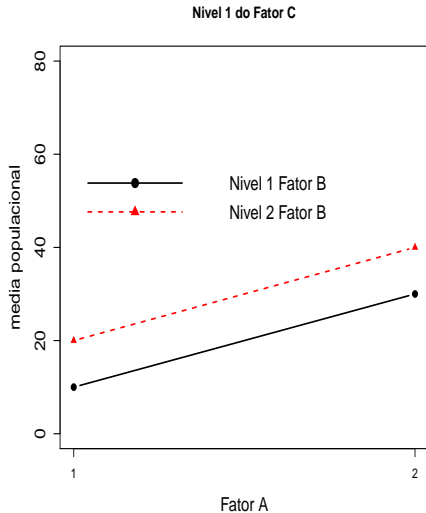
Fator A

Nível 2 do Fator C



Fator A

# Perfis médios: presença de interação de primeira ordem



# Médias

## ■ Neste caso

$$\mu_{111} = \mu, \mu_{211} = \mu + \alpha_2, \mu_{311} = \mu + \alpha_3$$

$$\mu_{121} = \mu + \beta_2, \mu_{221} = \mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22},$$

$$\mu_{321} = \mu + \alpha_3 + \beta_2 + (\alpha\beta)_{32}$$

$$\mu_{112} = \mu + \gamma_2, \mu_{212} = \mu + \alpha_2 + \gamma_2 + (\alpha\gamma)_{22},$$

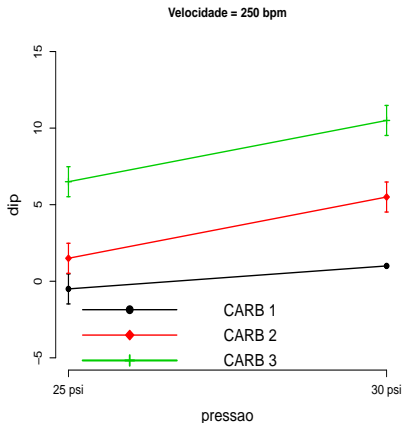
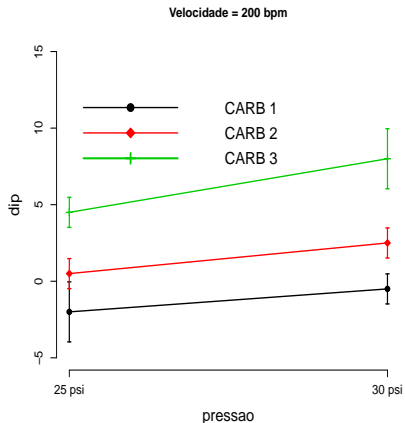
$$\mu_{312} = \mu + \alpha_3 + \gamma_2 + (\alpha\gamma)_{32}$$

$$\mu_{122} = \mu + \beta_2 + \gamma_2 + (\beta\gamma)_{22},$$

$$\mu_{222} = \mu + \alpha_2 + \beta_2 + \gamma_2 + (\alpha\beta)_{22} + (\beta\gamma)_{22} + (\alpha\gamma)_{22},$$

$$\mu_{322} = \mu + \alpha_3 + \beta_2 + \gamma_2 + (\alpha\beta)_{32} + (\beta\gamma)_{22} + (\alpha\gamma)_{32} + (\alpha\beta\gamma)_{322}$$

# Voltando ao exemplo: Gráfico de perfis médios

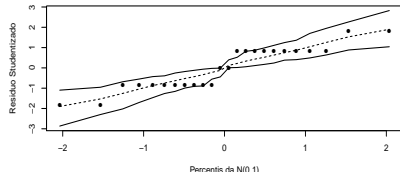
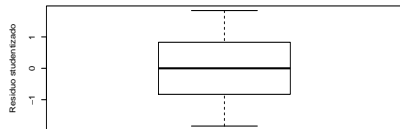
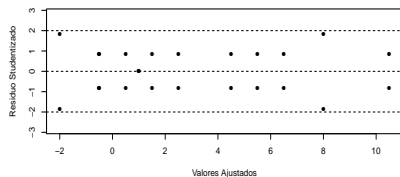
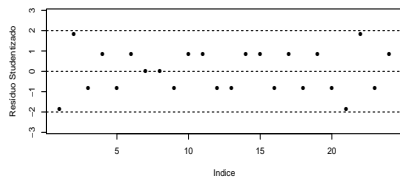


# Análise descritiva

Não há sentido em construir box-plots ou histogramas.

CARB	PRE	VELOC	Medida descritiva					
			Média	DP	Var.	CV%	Mínimo	Máximo
10	25 psi	200 bpm	-2,00	1,41	2,00	-70,71	-3,00	-1,00
10	25 psi	250 bpm	-0,50	0,71	0,50	-141,42	-1,00	0,00
10	30 psi	200 bpm	-0,50	0,71	0,50	-141,42	-1,00	0,00
10	30 psi	250 bpm	1,00	0,00*	0,00*	0,00*	1,00	1,00
12	25 psi	200 bpm	0,50	0,71	0,50	141,42	0,00	1,00
12	25 psi	250 bpm	1,50	0,71	0,50	47,14	1,00	2,00
12	30 psi	200 bpm	2,50	0,71	0,50	28,28	2,00	3,00
12	30 psi	250 bpm	5,50	0,71	0,50	12,86	5,00	6,00
14	25 psi	200 bpm	4,50	0,71	0,50	15,71	4,00	5,00
14	25 psi	250 bpm	6,50	0,71	0,50	10,88	6,00	7,00
14	30 psi	200 bpm	8,00	1,41	2,00	17,68	7,00	9,00
14	30 psi	250 bpm	10,50	0,71	0,50	6,73	10,00	11,00

# Análise de resíduos





# Comentários

- Parece que as suposições do modelo não são válidas para o conjunto de dados em questão (embora o ajuste tenha melhorado em relação à situação anterior).
- Ausência de homocedasticidade e normalidade.
- Uma alternativa: modelos de regressão com distribuição assimétrica/caudas pesadas para a variável resposta, que permita variâncias diferentes: normal e t assimétricas.

# Comentários

- O modelo de regressão normal linear, aparentemente, não é adequado para analisar os dados em questão, apesar do ajuste ter melhorado em relação à situação anterior (considerando apenas dois fatores).
- Contudo, seguiremos com ele por questões pedagógicas.

# Tabela ANOVA

FV	SQ	GL	QM	Estatística F	pvalor
CARB	2	252,750	126,375	178,4118	<0,0001
PRE	1	45,375	45,375	64,0588	<0,0001
VELOC	1	22,042	22,042	31,1176	0,0001
CARB x PRE	2	5,250	2,625	3,7059	0,0558
CARB x VELOC	2	0,583	0,292	0,4118	0,6714
PRE x VELOC	1	1,042	1,042	1,4706	0,2486
CARB x PRE x VELOC	2	1,083	0,542	0,7647	0,4869
Resíduo	12	8,500	0,708		
Total					

Ausência de interação de segunda ordem e de interação de primeira entre CARB e VELOC e PRE e VELOC.

**OBS: Pesquisar as formas escalares das somas de quadrados**

# Estimativas dos parâmetros do modelo

Parâmetro	Estimativa	EP	IC(95%)	Estat. t	pvalor
$\mu$	-2,00	0,60	[-3,17 ; -0,83]	-3,36	0,01
$\alpha_2$	2,50	0,84	[0,85; 4,15]	2,97	0,01
$\alpha_3$	6,50	0,84	[4,85; 8,15 ]	7,72	0,00
$\beta_2$	1,50	0,84	[-0,15; 3,15]	1,78	0,10
$\gamma_2$	1,50	0,84	[-0,15 ; 3,15]	1,78	0,10
$(\alpha\beta)_{22}$	0,50	1,19	[-1,83; 2,83]	0,42	0,68
$(\alpha\beta)_{32}$	2,00	1,19	[-0,33 ; 4,33]	1,68	0,12
$(\alpha\gamma)_{22}$	-0,50	1,19	[-2,83; 1,83]	-0,42	0,68
$(\alpha\gamma)_{32}$	0,50	1,19	[-1,83 ; 2,83 ]	0,42	0,68
$(\beta\gamma)_{22}$	-0,00	1,19	[-2,33 ; 2,33]	-0,00	1,00
$(\alpha\beta\gamma)_{222}$	2,00	1,68	[ -1,30 ; 5,30]	1,19	0,26
$(\alpha\beta\gamma)_{322}$	0,50	1,68	[ -2,80 ; 3,80]	0,30	0,77

## Modelo reduzido (casela de referência)

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + \xi_{ijkl},$$

(Fator A),  $i = 1, 2, 3$ ; (Fator B),  $j = 1, 2$ ; (Fator C),  $k = 1, 2$ ;

(unidades experimentais),  $l = 1, 2$

- Erros  $\xi_{ijkl} \stackrel{i.i.d}{\sim} N(0, \sigma^2)$ ,  $\mu, \alpha_i, \beta_j, \gamma_k, (\alpha\beta)_{ij}$  não aleatórios.
- Restrições :  $\alpha_1 = \beta_1 = \gamma_1 = (\alpha\beta)_{1j} = (\alpha\beta)_{i1} = 0, \forall i, j$ .

# Médias: modelo reduzido

## ■ Neste caso

$$\mu_{111} = \mu, \mu_{211} = \mu + \alpha_2, \mu_{311} = \mu + \alpha_3$$

$$\mu_{121} = \mu + \beta_2, \mu_{221} = \mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22},$$

$$\mu_{321} = \mu + \alpha_3 + \beta_2 + (\alpha\beta)_{32}$$

$$\mu_{112} = \mu + \gamma_2, \mu_{212} = \mu + \alpha_2 + \gamma_2,$$

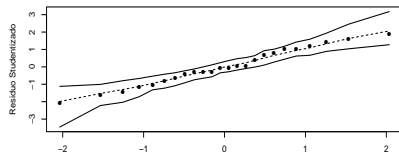
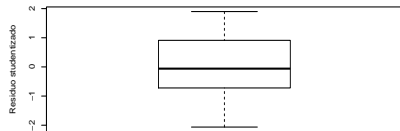
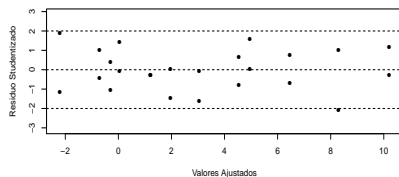
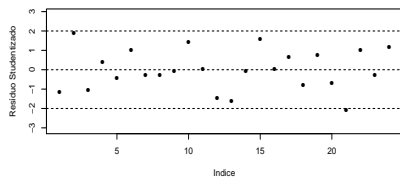
$$\mu_{312} = \mu + \alpha_3 + \gamma_2$$

$$\mu_{122} = \mu + \beta_2 + \gamma_2,$$

$$\mu_{222} = \mu + \alpha_2 + \beta_2 + \gamma_2 + (\alpha\beta)_{22},$$

$$\mu_{322} = \mu + \alpha_3 + \beta_2 + \gamma_2 + (\alpha\beta)_{32}$$

# Análise de resíduos



# Comentários

- Parece que as suposições do modelo não são válidas para o conjunto de dados em questão (embora o ajuste tenha melhorado em relação à situação anterior).
- Ausência de homocedasticidade e normalidade.
- Uma alternativa: modelos de regressão com distribuição assimétrica/caudas pesadas para a variável resposta, que permita variâncias diferentes: normal e t assimétricas.



# Comentários

- O modelo de regressão normal linear, aparentemente, não é adequado para analisar os dados em questão, apesar do ajuste ter melhorado em relação à situação anterior (considerando apenas dois fatores).
- Contudo, seguiremos com ele por questões pedagógicas.

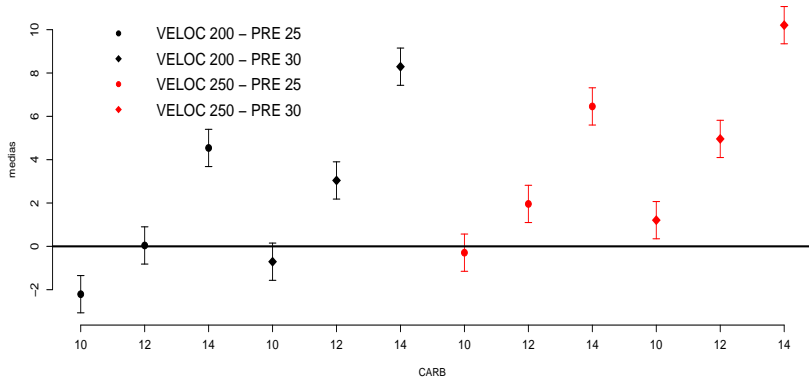
# Tabela ANOVA: modelo reduzido

FV	SQ	GL	QM	Estatística F	pvalor
CARB	2	252,75	126,38	191,68	<0,0001
PRE	1	45,38	45,38	68,82	<0,0001
VELOC	1	22,04	22,04	33,43	<0,0001
CARB x PRE	2	5,25	2,63	3,98	0,03818
Resíduos	17	11,21	0,66		
Total	22	336,63			

# Estimativas dos parâmetros do modelo

Parâmetro	Estimativa	EP	IC(95%)	Estat. t	pvalor
$\mu$	-2,21	0,44	[-3,07; -1,35]	-5,04	0,00
$\alpha_2$	2,25	0,57	[1,12 ; 3,38]	3,92	0,00
$\alpha_3$	6,75	0,57	[5,62 ; 7,88]	11,76	0,00
$\beta_2$	1,50	0,57	[0,37 ; 2,63]	2,61	0,02
$\gamma_2$	1,92	0,33	[1,27 ; 2,57]	5,78	0,00
$(\alpha\beta)_{22}$	1,50	0,81	[-0,09 ; 3,09]	1,85	0,08
$(\alpha\beta)_{32}$	2,25	0,81	[0,66 ; 3,84 ]	2,77	0,01

# Médias estimadas: modelo reduzido



# Comparações entre as médias

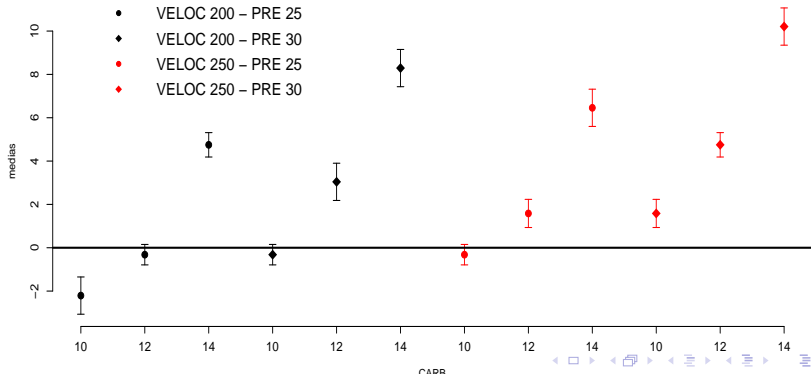
- Perguntas de interesse : quais tratamento fornecem DIP's médios:
  - Em torno do 0?
  - Negativos?
  - Positivos?
- Quais são os padrões de igualdade entre as médias.
- Após realizadas algumas comparações (em termos de  $C\beta = \mathbf{0}$ ) discutidas em classe (veja também o arquivo com os programas relacionados), chegou-se ao padrão a seguir.

# Padrões

- Considere que os tratamentos (de acordo com o gráfico anterior) estão numeradas de 1 a 12 (conforme o padrão apresentado).
- Identificou-se 8 grupos, a saber:
  - Grupo 1: tratamento 1.
  - Grupo 2: tratamentos 2, 4 e 7.
  - Grupo 3: tratamentos 3 e 11.
  - Grupo 4: tratamento 5.
  - Grupo 5: tratamento 6.
  - Grupo 6: tratamentos 8 e 10.
  - Grupo 7: tratamento 9.
  - Grupo 8: tratamento 12.

# Médias estimadas: modelo reduzido (após os testes

$$C\beta = \mathbf{0}$$



## Estimativas finais das médias

Grupo	Estimativa	EP	IC(95%)
Grupo 1	-2,21	0,44	[-3,07 ; -1,35]
Grupo 2	-0,32	0,24	[-0,79 ; 0,15]
Grupo 3	4,75	0,29	[4,19 ; 5,31]
Grupo 4	3,04	0,44	[2,18; 3,90]
Grupo 5	8,29	0,44	[7,43 ; 9,15]
Grupo 6	1,58	0,33	[0,93 ; 2,23]
Grupo 7	6,46	0,44	[5,60 ; 7,32]
Grupo 8	10,21	0,44	[9,35 ; 11,07]



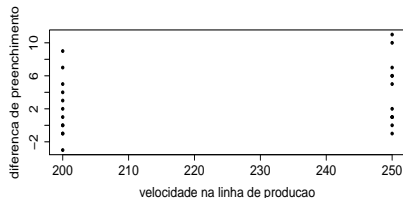
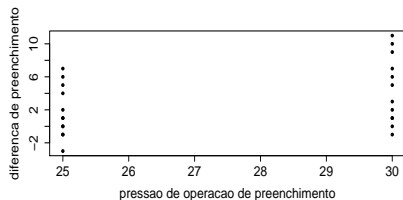
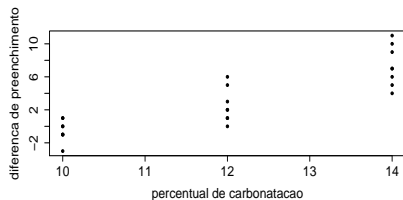
# Comentários

- Em geral, é melhor estimar as médias com o modelo reduzido (todas as médias são estimadas com maior precisão).
- Entretanto, também é válido estimar as médias, com o modelo “atual” .
- Conclusões finais:
  - Grupo 1: DIP médio negativo.
  - Grupo 2: DIP médio em torno do zero.
  - Grupos 3 a 8: DIP médio positivo.

# Análise via modelo de regressão

- Identificamos um padrão de diferença entre as médias através da análise anterior.
- Podemos ajustar um modelo de regressão (linear ou quadrático) para quantificar o impacto que aumento de uma ou mais de uma unidade nos fatores causam no DIP.
- Naturalmente, podemos levar em consideração os resultados anterior para propor um modelo de regressão que leve em consideração o padrão de diferenças entre as médias.
- Entretanto, vamos considerar um modelo de regressão linear múltipla.

# Diagramas de Dispersão.

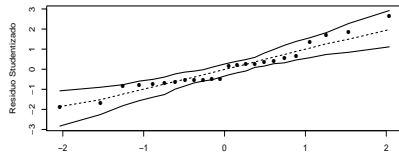
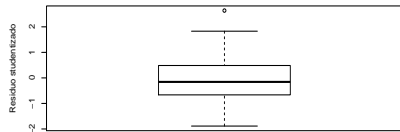
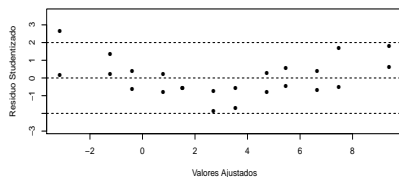
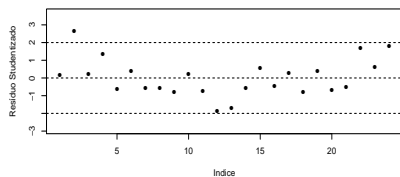


# Modelo linear 1: reta

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \xi_{ij}, i = 1, \dots, 24$$

- $Y_i$  : DIP associado ao  $i$ -ésimo vasilhame.
- $x_{1i}$  : CARB ao qual foi submetido o  $i$ -ésimo vasilhame.
- $x_{2i}$  : PRE ao qual foi submetido o  $i$ -ésimo vasilhame.
- $x_{3i}$  : VELOC ao qual foi submetido o  $i$ -ésimo vasilhame.
- $\beta_0$  : valor esperado (média) do DIP para vasilhames submetidos ao valor 0 (simultaneamente) de todas as covariáveis (CAR,PRE,VELOC).
- $\beta_1$  : incremento no valor esperado do DIP com o aumento em uma unidade de CARB mantendo-se as outras covariáveis fixas.
- $\beta_2$  : incremento no valor esperado do DIP com o aumento em uma unidade de PRE mantendo-se as outras covariáveis fixas.
- $\beta_3$  : incremento no valor esperado do DIP com o aumento em uma unidade de VELOC mantendo-se as outras covariáveis fixas.

# Análise de resíduos



# Comentários

- Parece que as suposições do modelo não são válidas para o conjunto de dados em questão (embora o ajuste tenha melhorado em relação à situação anterior).
- Ausência de homocedasticidade e normalidade.
- Uma alternativa: modelos de regressão com distribuição assimétrica/caudas pesadas para a variável resposta, que permita variâncias diferentes: normal e t assimétricas.

# Comentários

- O modelo de regressão normal linear, aparentemente, não é adequado para analisar os dados em questão, apesar do ajuste ter melhorado em relação à situação anterior (considerando apenas dois fatores).
- Contudo, seguiremos com ele por questões pedagógicas.

# Estimativas dos parâmetros do modelo

Parâmetro	Estimativa	EP	IC(95%)	Estat. t	pvalor
$\beta_0$	-44,25	3,36	[-50,84 ; -37,66]	-13,15	<0,0001
$\beta_1$	1,97	0,13	[ 1,72 ; 2,22 ]	15,32	<0,0001
$\beta_2$	0,55	0,08	[0,39 ; 0,71]	6,55	<0,0001
$\beta_3$	0,04	0,01	[ 0,02 ; 0,05]	4,57	0,0002

Todas as covariáveis apresentam impacto positivo e significativo na média da variável resposta.