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

Prof. Caio Azevedo

# Contexto

- Vimos até agora experimentos envolvendo um único fator.
- Em muitas situações, o pesquisador tem interesse em como dois ou mais fatores afetam o comportamento da variável resposta.
- Nem todos os fatores são, necessariamente, de interesse. Contudo, em princípio, todos devem ser controlados de alguma forma.
- Comecemos com dois fatores dentro de uma estrutura balanceada.

# Descrição

- Fator A: possui $a$ níveis.
- Fator B: possui $b$ níveis.
- 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 $n$ observações (balanceado). Cada uma das $n$ observações são alocadas aleatoriamente à cada uma das combinações (fatores). Temos uma PCA (planejamento completamente casualizado).

# Descrição (Cont.)

- Note que tem-se um total de $n \times a \times b$ observações.
- Conceito importante: interação entre os fatores.
- Interação: a diferença entre as médias da resposta, entre dois níveis do Fator A, são iguais ao longo dos níveis do Fator B (vice-versa).

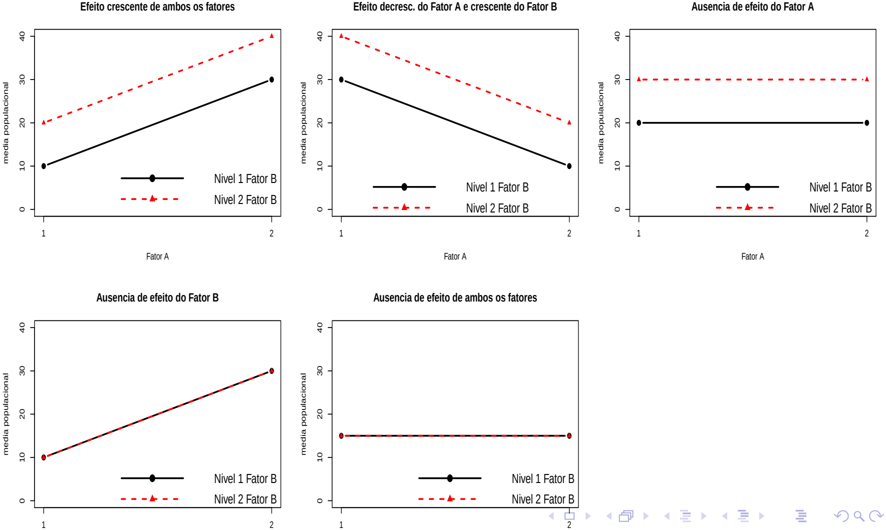
# Exemplo 4: Resistência de materiais

- 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$^o$F, 70$^o$F e 125$^o$F. Equivalente à -9,44$^o$C, 21,11$^o$C e 51,67 $^o$C, respectivamente
- Para cada combinação (tipo de material da placa $\times$ temperatura) 4 baterias foram feitas.
- Variável resposta: tempo de vida em horas de cada bateria .

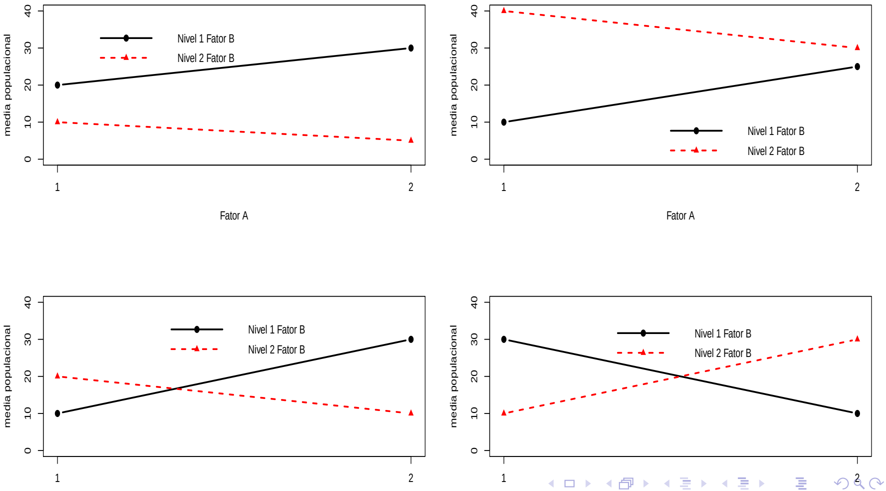
# Exemplo 4: continuação

- Experimento balanceado: 4 observações por tratamento.
- Um fator quantitativo (temperatura) e um fator qualitativo (tipo de material da placa).
- Como analisar o experimento?
- Qual seria um modelo apropriado?
- Como estimar os parâmetros e comparar as médias de interesse?
- Como verificar as suposições do modelo?

# Perfis médios: ausência de interação



# Perfis médios: presença de interação



# Voltando ao Exemplo 4

- Vamos considerar, inicialmente, somente os dois primeiros níveis de cada fator.
- Dados:

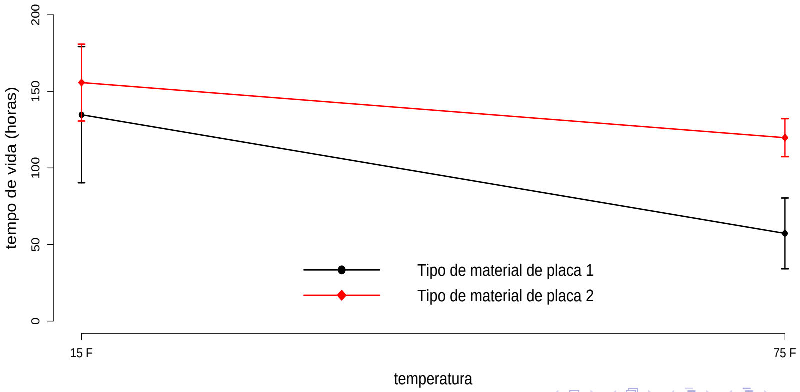
| Material | Temperatura ($^oF$) | | | |
|---|---|---|---|---|
| | **15** | | **70** | |
| 1 | 130 | 155 | 34 | 40 |
| | 74 | 180 | 80 | 75 |
| 2 | 150 | 188 | 136 | 122 |
| | 159 | 126 | 106 | 112 |

# Análise descritiva

Não há sentido em construir box-plots ou histogramas (poucas observações por grupo).

| Material | Temp. | Medida descritiva | | | | | |
|---|---|---|---|---|---|---|---|
| | | **Média** | **DP** | **Var.** | **CV%** | **Máximo** | **Mínimo** |
| 1 | 15 F | 134,75 | 45,35 | 2056,92 | 74,00 | 180,00 | 33,66 |
| | 70 F | 57,25 | 23,60 | 556,92 | 34,00 | 80,00 | 41,22 |
| 2 | 15 F | 155,75 | 25,63 | 656,25 | 126,00 | 188,00 | 16,45 |
| | 70 F | 119,75 | 12,66 | 160,25 | 106,00 | 136,00 | 10,57 |

# Gráfico de perfis médios



# Modelo (casela de referência)

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

(Fator A), $i = 1, 2$; (Fator B), $j = 1, 2$; (unidades experimentais), $k = 1, 2, 3, 4$

- Erros $\xi_{ijk} \overset{i.i.d}{\sim} N(0, \sigma^2)$, $\mu, \alpha_i, \beta_j, (\alpha\beta)_{ij}$ não aleatório.
- $\mathcal{E}_{\xi_{ijk}}(Y_{ijk}) = \mu_{ij}, \mathcal{V}_{\xi_{ijk}}(Y_{ijk}) = \sigma^2$.
- Restrições : $\alpha_1 = \beta_1 = (\alpha\beta)_{1j} = (\alpha\beta)_{i1} = 0, \forall i, j$.
- $Y_{ijk} \overset{ind.}{\sim} N(\alpha_i + \beta_j + (\alpha\beta)_{ij}, \sigma^2)$.

# Interpretações dos parâmetros

- Neste caso

$$
\begin{aligned}
\mu_{11} &= \mu \\
\mu_{21} &= \mu + \alpha_2 \\
\mu_{12} &= \mu + \beta_2 \\
\mu_{22} &= \mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}
\end{aligned}
$$

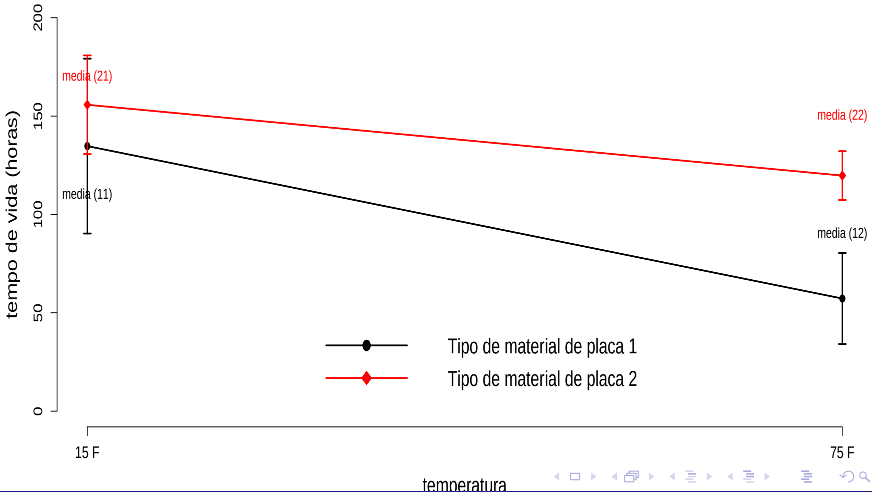
# Interpretações dos parâmetros

- Fator A (1: material 1, 2: material 2).
- Fator B (1: $15^o F$, 2: $75^o F$).
- Parâmetros $\boldsymbol{\beta} = (\mu, \alpha_2, \beta_2, (\alpha\beta)_{22})$ (modelo identificado).
- Se $(\alpha\beta)_{22} = 0$.
  - $\alpha_2$: incremento na vida média de baterias feitas com material 2 em relação àquelas feitas com material 1 submetidas à qualquer uma das duas temperaturas.
  - $\beta_2$: incremento na vida média de baterias submetidas à temperatura de $75^o F$ em relação submetidas à temperatura de $15^o F$, feitas com qualquer um dos dois tipos de material.

# Interpretações dos parâmetros (cont.)

- A não nulidade de $(\alpha\beta)_{22}$ faz com que os incrementos anteriores não dependam somente de $\alpha_2$ e $\beta_2$. Neste caso:
  - Dependendo da temperatura, a diferença entre a vida média de baterias feitas com os materiais 1 e 2 não é a mesma.
  - Dependendo do tipo de material, a diferença entre a vida média de baterias submetidas as temperaturas $15^oF$ e $75^oF$ não é a mesma.
- O parâmetro $(\alpha\beta)_{22}$ determina a existência ou não de interação.

# Visualização dos significados dos parâmetros



# Interpretações dos parâmetros (cont.)

- Não existe interação, neste naso, $\leftrightarrow H_0 : \mu_{21} - \mu_{11} = \mu_{22} - \mu_{12}$ for verdadeira.
- Por outro lado, a hipótese acima equivale à:

$$H_0 : \alpha_2 = \alpha_2 + \beta_2 + (\alpha\beta)_{22} - \beta_2 \leftrightarrow (\alpha\beta)_{22} = 0$$

- Portanto, inexiste interação $\leftrightarrow (\alpha\beta)_{22} = 0$.

# Interpretações dos parâmetros (cont.)

- Se existe interação, portanto se $(\alpha\beta)_{22} \neq 0$, temos que:
- $\alpha_2$: incremento na vida média de baterias feitas com material 2 em relação àquelas feitas com material 1 submetidas à temperatura de $15^o F$.
- $\beta_2$: incremento na vida média de baterias submetidas à temperatura de $75^o F$ em relação submetidas à temperatura de $15^o F$, feitas com material do tipo 1.
- $(\alpha\beta)_{22}$ : interação entre os fatores.

# Hipótese de interesse

- Comparar simultaneamente todas as médias deixa de ter sentido prático.
- Primeira hipótese (ausência de interação): $H_0 : (\alpha\beta_{22}) = 0$ vs $H_1 : (\alpha\beta_{22}) \neq 0$
- Se a hipótese acima ($H_0$) não for rejeitada, então:
  - Ausência de efeito principal de material: $H_0 : \alpha_2 = 0$ vs $H_1 : \alpha_2 \neq 0$.
  - Ausência de efeito principal de temperatura: $H_0 : \beta_2 = 0$ vs $H_1 : \beta_2 \neq 0$.
- Eventualmente, algum tipo de comparação entre as médias remanescentes.

# Hipótese de interesse (cont.)

- Se a hipótese acima de ausência de interação não for rejeitada, então não faz sentido estudar os efeitos principais isoladamente
- Portanto, deve-se efetuar algum tipo de comparação entre as médias.

# Somas de quadrados

- Decomposição da soma de quadrados total:

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

# Tabela de análise de variância

- Para testar a igualdade simultânea das médias

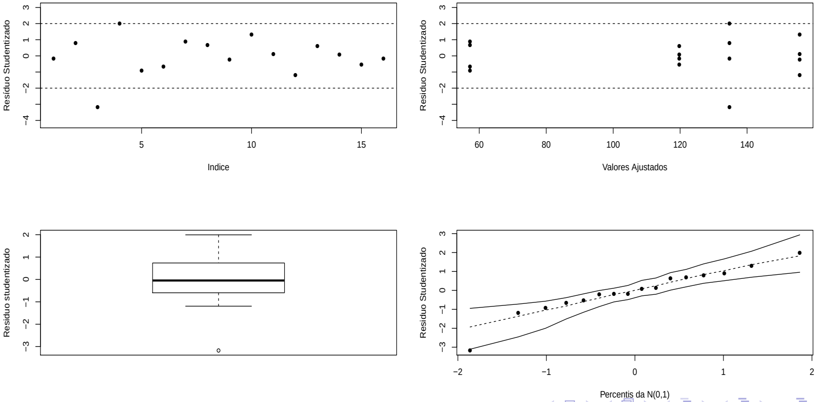
| FV | SQ | GL | QM | Estatística F | pvalor |
|---|---|---|---|---|---|
| Fator A | $SQF_A$ | a-1 | $QMF_A = \frac{SQF_A}{(a-1)}$ | $F_A = \frac{QMF_A}{QMR}$ | $S(f_A\|H_0)$ |
| Fator B | $SQF_B$ | b-1 | $QMF_B = \frac{SQF_B}{(b-1)}$ | $F_B = \frac{QMF_B}{QMR}$ | $S(f_B\|H_0)$ |
| Interação | SQInt | (a-1)(b-1) | $QMInt = \frac{SQInt}{[(a-1)(b-1)]}$ | $F_{Int} = \frac{QMInt}{QMR}$ | $S(f_{Int}\|H_0)$ |
| 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. $S(x|H_0)$ fds no ponto $x$ sob $H_0$.

# Testes para homocedasticidade

- Teste de Bartlett : 3,96 (0,2663).
- Teste de Levene : 1,51 (0,2633).
- Hipótese de homocedasticidade parace não ser desprezível (cautela).

# Análise de resíduos



# Tabela ANOVA

| FV | SQ | GL | QM | Estatística F | pvalor |
|---|---|---|---|---|---|
| Material | 6972,30 | 1 | 6972,30 | 8,13 | 0,0146 |
| Temperatura | 12882,20 | 1 | 12882,20 | 15,02 | 0,0022 |
| Interação | 1722,30 | 1 | 1722,30 | 2,01 | 0,1819 |
| Resíduo | 10291,00 | 12 | 857,60 | | |
| Total | 31867,75 | 15 | | | |

Ausência de interação e de efeito principal de Material (devido ao gráfico de perfis e ao resultado a seguir).

# Estimativas dos parâmetros do modelo

| Parâmetro | Estimativa | EP | IC(95%) | Estat. t | pvalor |
|---|---|---|---|---|---|
| $\mu$ | 134,75 | 14,64 | [102,85; 166,65] | 9,20 | $< 0{,}0001$ |
| $\alpha_2$ | 21,00 | 20,71 | [-24,12 ; 66,12] | 1,01 | 0,3305 |
| $\beta_2$ | -77,50 | 20,71 | [-122,62 ;-32,38] | -3,74 | 0,0028 |
| $(\alpha\beta)_{22}$ | 41,50 | 29,28 | [-22,31 ;105,31] | 1,42 | 0,1819 |

Apesar a inexistência de interação ser confirmada, há uma discordância em termos da inexistência do efeito de tipo de material. Provavelmente, pelo comprometimento no ajuste do modelo e pelo valor do erro-padrão associado à estimativa de $(\alpha_2)$. Optaremos por ajustar um modelo reduzido com os fatores principais, sem interação.

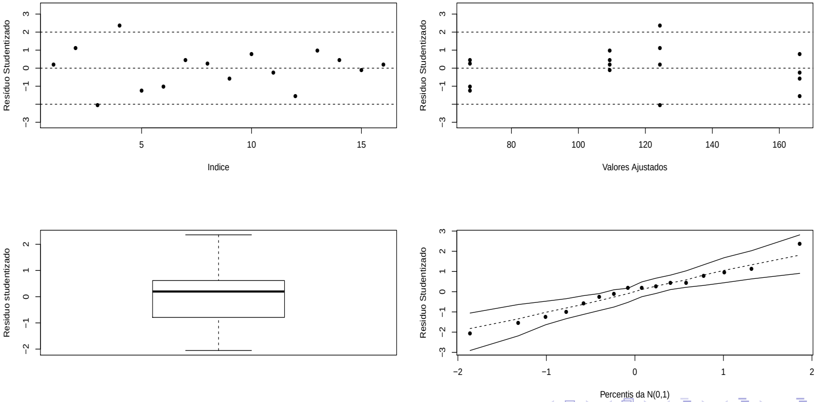
# Modelo reduzido (casela de referência)

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

(Fator A), $i = 1, 2$; (Fator B), $j = 1, 2$; (unidades experimentais), $k = 1, 2, 3, 4$

- Erros $\xi_{ijk} \overset{i.i.d}{\sim} N(0, \sigma^2)$, $\mu, \alpha_i, \beta_j, (\alpha\beta)_{ij}$ não aleatório.
- $\mathcal{E}_{\xi_{ijk}}(Y_{ijk}) = \mu_{ij} = \mu + \alpha_i + \beta_j, \mathcal{V}_{\xi_{ijk}}(Y_{ijk}) = \sigma^2$.
- Restrições : $\alpha_1 = \beta_1 = 0, \forall i, j$.
- $Y_{ijk} \overset{ind.}{\sim} N(\beta_j, \sigma^2)$.

# Análise de resíduos: modelo reduzido



# Tabela ANOVA

| FV | GL | SQ | QM | Estatística F | pvalor |
|---|---|---|---|---|---|
| Material | 1 | 6972,25 | 6972,25 | 7,54 | 0,0166 |
| Temperatura | 1 | 12882,25 | 12882,25 | 13,94 | 0,0025 |
| Resíduo | 13 | 12013,25 | 924,10 | | |
| Total | 15 | 31867,75 | | | |

# Estimativas dos parâmetros do modelo

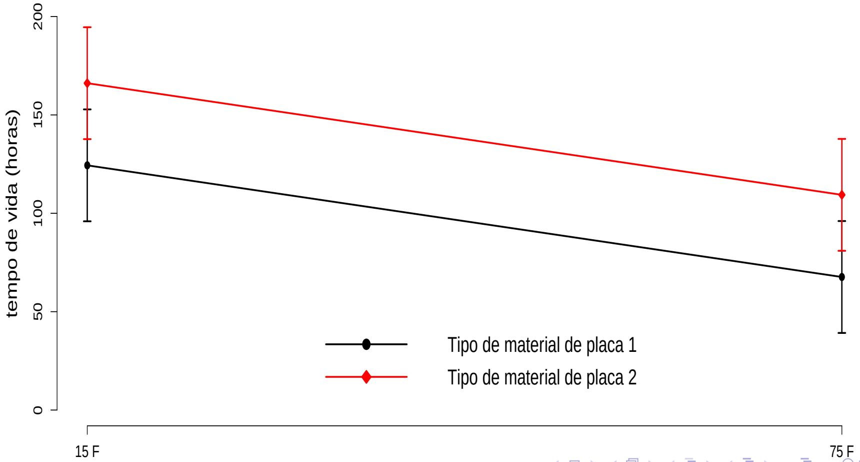
| Parâmetro | Estimativa | EP | IC(95%) | Estat. t | pvalor |
|---|---|---|---|---|---|
| $\mu$ | 124,38 | 13,16 | [ 95,94 ; 152,81 ] | 9,45 | $<$0,0001 |
| $\alpha_2$ | 41,75 | 15,20 | [ 8,91 ; 74,59 ] | 2,75 | 0,0166 |
| $\beta_2$ | -56,75 | 15,20 | [ -89,59 ; -23,91 ] | -3,73 | 0,0025 |

Para o modelo reduzido, os resultados da ANOVA e das estimativas dos parâmetros concordaram. No entanto, a magnitude dos erros-padrão continuam elevadas (isto é devido ao pequeno número de observações por grupo). Estimaremos as médias via modelo reduzido.

# Estimativas finais das médias

| Tratamento | Estimativa | EP | IC(95%) |
|---|---|---|---|
| Mater. do tipo 1 - temp. de 15$^oF$ | 124,38 | 13,16 | [95,94 ; 152,81 ] |
| Mater. do tipo 1 - temp. de 75$^oF$ | 67,62 | 13,16 | [39,19 ; 96,06] |
| Mater. do tipo 2 - temp. de 15$^oF$ | 166,12 | 13,16 | [137,69 ; 194,56] |
| Mater. do tipo 1 - temp. de 75$^oF$ | 109,38 | 13,16 | [80,94 ; 137,81] |

# Perfis médios ajustados pelo modelo reduzido



# Comentários

- Apesar dos intervalos de confianças para as médias se interceptarem, eles não incluem as estimativas pontuais do outro grupo.
- Os resultados (ANOVA, estimativas dos parâmetros e das médias) indicam que, apesar do fato anterior, as médias não são iguais.
- Os reduzidos tamanhos amostrais, por grupo, não permitiram a obtenção de IC's com comprimentos menores.