

Planejamento e Análise Estatística de Experimentos fatoriais: análise de dados de experimentos completamente aleatorizados - Parte 2

Prof. Caio Azevedo

Contexto

- Veremos como analisar experimentos fatoriais com 2 fatores e mais de dois níveis.
- Em princípio, vamos considerar uma estrutura balanceada (mesmo número de observações por tratamento).

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

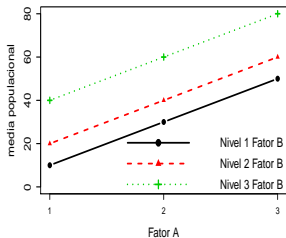
- Um engenheiro está desenvolvendo um tipo de bateria para ser usado em um dispositivo eletrônico sujeito à variações extremas de temperatura.
- Fatores de interesse:
 - Tipo de material da placa: 1, 2 e 3.
 - Temperatura: 15°F, 70°F e 125°F. Equivalente à -9,44°C, 21,11°C e 51,67 °C, respectivamente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

Voltando ao Exemplo 4: continuação

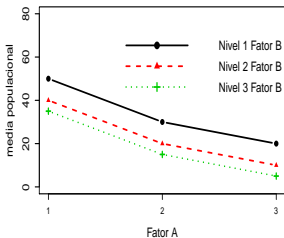
- Experimento balanceado: 4 observações por tratamento.
- Um fator quantitativo (temperatura) e um fator qualitativo (tipo de material da placa).
- Como analisar o experimento?
- Qual seria um modelo apropriado?
- Como estimar os parâmetros e comparar as médias de interesse?
- Como verificar as suposições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

Perfis médios: ausência de interação

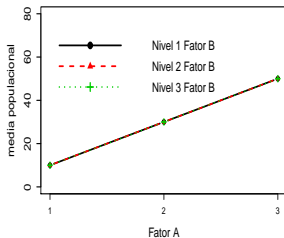
Efeito crescente de ambos os fatores



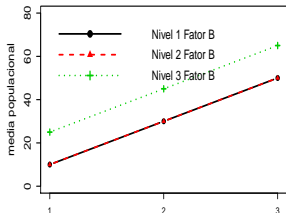
Efeito decresc. em A e em B



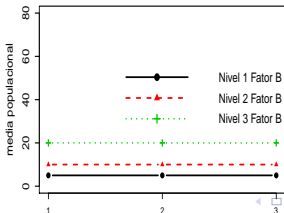
Efeito crescente em A e sem efeito em B



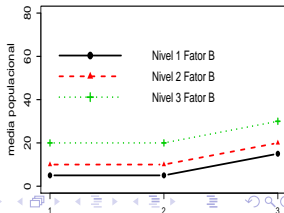
Efeito crescente em ambos os fatores (nao uniforme em B)



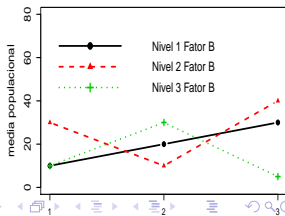
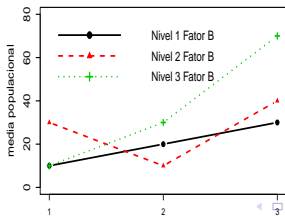
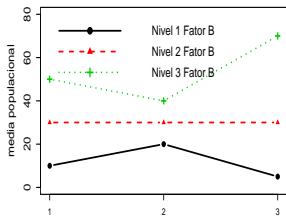
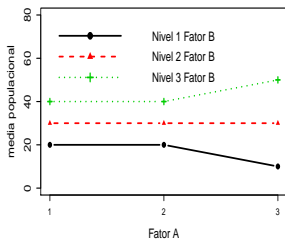
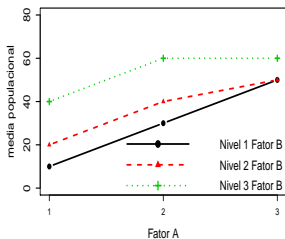
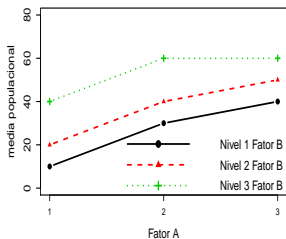
Efeito crescente em B e ausência de efeito em A



Efeito crescente de ambos os fatores (nao uniforme em A)



Perfis médios: presença de interação



Voltando ao Exemplo 4

- Dados:

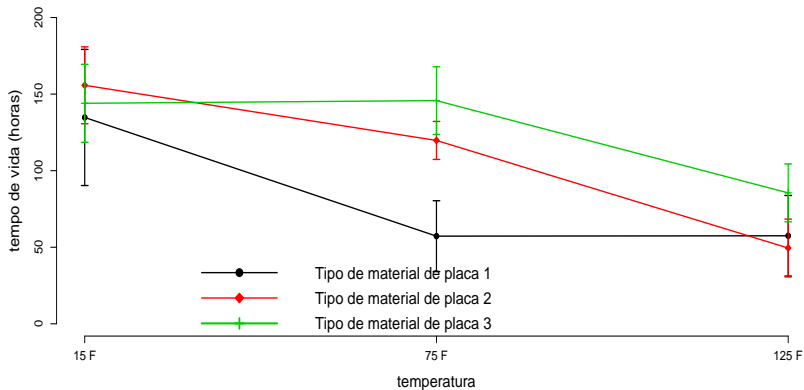
Material	Temperatura ($^{\circ}F$)					
	15		70		150	
1	130	155	34	40	20	70
	74	180	80	75	82	88
2	150	188	136	122	25	70
	159	126	106	112	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

Análise descritiva

Não há sentido em construir box-plots ou histogramas.

Material	Temp.	Medida descritiva					
		Média	DP	Var.	CV%	Mínimo	Máximo
1	15 F	134,75	45,35	2056,92	33,66	74,00	180,00
	70 F	57,25	23,60	556,92	41,22	34,00	80,00
	125 F	57,50	26,85	721,00	46,70	20,00	82,00
2	15 F	155,75	25,62	656,25	16,45	126,00	188,00
	70 F	119,75	12,66	160,25	10,57	106,00	136,00
	125 F	49,50	19,26	371,00	38,91	25,00	70,00
3	15 F	144,00	25,97	674,67	18,04	110,00	168,00
	70 F	145,75	22,54	508,25	15,47	120,00	174,00
	125 F	85,50	19,28	371,67	22,55	60,00	104,00

Gráfico de perfis médios



Modelo (casela de referência)

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

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

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

Interpretações dos parâmetros

- Neste caso

$$\mu_{11} = \mu$$

$$\mu_{21} = \mu + \alpha_2$$

$$\mu_{31} = \mu + \alpha_3$$

$$\mu_{12} = \mu + \beta_2$$

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

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

$$\mu_{13} = \mu + \beta_3$$

$$\mu_{23} = \mu + \alpha_2 + \beta_3 + (\alpha\beta)_{23}$$

$$\mu_{33} = \mu + \alpha_3 + \beta_3 + (\alpha\beta)_{33}$$

Casela de referência : matriz de planejamento (apenas uma observação por tratamento)

$$Y = \begin{bmatrix} Y_{111} \\ Y_{121} \\ Y_{131} \\ Y_{211} \\ Y_{221} \\ Y_{231} \\ Y_{311} \\ Y_{321} \\ Y_{331} \end{bmatrix} X = \begin{bmatrix} \mu & \alpha_2 & \alpha_3 & \beta_2 & \beta_3 & (\alpha\beta)_{22} & (\alpha\beta)_{23} & (\alpha\beta)_{32} & (\alpha\beta)_{33} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Interpretações dos parâmetros

- Fator A (1: material 1, 2: material 2, 3: material 3).
- Fator B (1: 15°F, 2: 75°F, 3: 124°F).
- Parâmetros $\beta = (\mu, \alpha_2, \alpha_3, \beta_2, \beta_3, (\alpha\beta)_{22}, (\alpha\beta)_{23}, (\alpha\beta)_{32}, (\alpha\beta)_{33})$
(modelo identificado).
- Exercício, escrever o modelo na forma matricial.
- Se $(\alpha\beta)_{22} = (\alpha\beta)_{23} = (\alpha\beta)_{32} = (\alpha\beta)_{33} = 0$. Os efeitos principais podem ser considerados separadamente.

Interpretações dos parâmetros (cont.)

- A não nulidade de pelo menos um $(\alpha\beta)_{ij}$ faz com que os incrementos anteriores possam não depender somente de α_2, α_3 e β_2, β_3 . Neste caso:
 - Dependendo da temperatura, a diferença entre a vida média de baterias feitas com os materiais 1, 2 e 3 não é a mesma.
 - Dependendo do tipo de material, a diferença entre a vida média de baterias submetidas as temperaturas $15^\circ F$, $75^\circ F$ e $125^\circ F$ não é a mesma.
- O parâmetros $(\alpha\beta)_{ij}$ determinam a existência ou não de interação.

Interpretações dos parâmetros (cont.)

- Não existe interação, neste caso, por exemplo se (paralelismo simultâneo entre os perfis):

$$H_0 : \begin{cases} \mu_{21} - \mu_{11} = \mu_{22} - \mu_{12} = \mu_{23} - \mu_{13} \\ \mu_{31} - \mu_{21} = \mu_{32} - \mu_{22} = \mu_{33} - \mu_{23} \end{cases}$$

- for verdadeira. Por outro lado, a hipótese acima equivale à:

$$H_0 : (\alpha\beta)_{22} = (\alpha\beta)_{23} = (\alpha\beta)_{32} = (\alpha\beta)_{33} = 0$$

- for verdadeira.

Interpretações dos parâmetros (cont.)

- Se existe interação, cada um dos parâmetro tem interpretações mais específicas, em geral, fixando-se o nível de algum fator.

Hipótese de interesse

- Comparar simultaneamente todas as médias deixa de ter sentido prático.
- Primeira hipótese (ausência de interação):

$$\left\{ \begin{array}{l} H_0 : (\alpha\beta)_{22} = (\alpha\beta)_{23} = (\alpha\beta)_{32} = (\alpha\beta)_{33} = 0 \\ H_1 : \text{pelo menos um parâmetro diferente de } 0 \end{array} \right.$$

Hipótese de interesse

- Se a hipótese acima (H_0) não for rejeitada, então:

- Ausência de efeito principal de material:

$$\begin{cases} H_0 : \alpha_2 = \alpha_3 = 0 \\ H_1 : \text{pelo menos um parâmetro diferente de 0} \end{cases}$$

- Ausência de efeito principal de temperatura:

$$\begin{cases} H_0 : \beta_2 = \beta_3 = 0 \\ H_1 : \text{pelo menos um parâmetro diferente de 0} \end{cases}$$

- 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 comparações específicas entre as médias.

Somas de quadrados

- Decomposição da soma de quadrados total:

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

Tabela de análise de variância

- Para testar a igualdade simultânea das médias

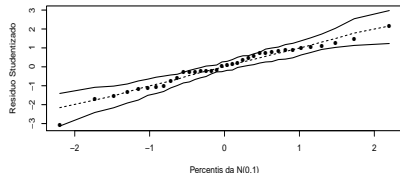
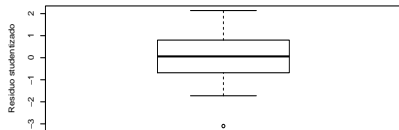
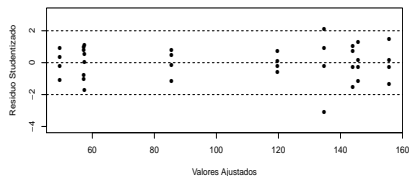
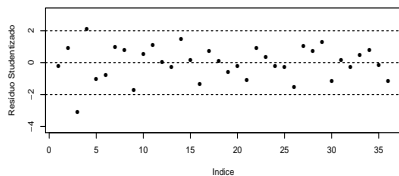
FV	SQ	GL	QM	Estatística F	pvalor
Fator A	SQF_A	a-1	$QMF_A = \frac{SQF_A}{(a-1)}$	$F_A = \frac{QMF_A}{QMR}$	$\min(F(f_A H_0), S(f_A H_0))$
Fator B	SQF_B	b-1	$QMF_B = \frac{SQF_B}{(b-1)}$	$F_B = \frac{QMF_B}{QMR}$	$\min(F(f_B H_0), S(f_B H_0))$
Interação	SQInt	(a-1)(b-1)	$QMInt = \frac{SQInt}{[(a-1)(b-1)]}$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

Testes para homocedasticidade

- Teste de Bartlett (pvalor) : 5,34 (0,7321).
- Teste de Levene (pvalor) : 0,80 (0,6081).
- Hipótese de homocedasticidade parece não ser desprezível (cautela).

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.

Tabela ANOVA

FV	SQ	GL	QM	Estatística F	pvalor
T. de Material	10684,00	2	5341,90	7,91	0,0020
Temperatura	39119,00	2	19559,40	28,97	<0,0001
Interação	9614,00	4	2403,40	3,56	0,0186
Resíduo	18231,00	27	675,20		
Total	77646,97	34			

Presença de interação. Como conduzir as comparações?

Estimativas dos parâmetros do modelo

Parâmetro	Estimativa	EP	IC(95%)	Estat. t	pvalor
μ	134,75	12,99	[109,28;160,22]	10,37	<0,0001
α_2	21,00	18,37	[-15,01;57,01]	1,14	0,2631
α_3	9,25	18,37	[-26,76;45,26]	0,50	0,6187
β_2	-77,50	18,37	[-113,51;-41,49]	-4,22	0,0002
β_3	-77,25	18,37	[-113,26;-41,24]	-4,20	0,0003
$(\alpha\beta)_{22}$	41,50	25,98	[-9,43;92,43]	1,60	0,1219
$(\alpha\beta)_{32}$	79,25	25,98	[28,32;130,18]	3,05	0,0051
$(\alpha\beta)_{23}$	-29,00	25,98	[-79,93;21,93]	-1,12	0,2742
$(\alpha\beta)_{33}$	18,75	25,98	[-32,18;69,68]	0,72	0,4768

Inferência

- Quais são as comparações de maior interesse?

Inferência

- Quais são as comparações de maior interesse?
- À que conclusões podemos chegar à partir dos resultados anteriores?

Inferência

- Quais são as comparações de maior interesse?
- À que conclusões podemos chegar à partir dos resultados anteriores?
- Como fazer as comparações de interesse?

Inferência

- Quais são as comparações de maior interesse?
- À que conclusões podemos chegar à partir dos resultados anteriores?
- Como fazer as comparações de interesse?
- Para cada temperatura, qual o material produz componentes que apresentam maior tempo de vida?

Hipóteses: comparação de tipos de material

- Temperatura 15°F:

$$H_0 : \begin{cases} \mu_{11} = \mu_{21} \\ e \\ \mu_{11} = \mu_{31} \end{cases} \leftrightarrow \begin{cases} \alpha_2 = 0 \\ e \\ \alpha_3 = 0 \end{cases}$$

- Temperatura 75°F:

$$H_0 : \begin{cases} \mu_{12} = \mu_{22} \\ e \\ \mu_{12} = \mu_{32} \end{cases} \leftrightarrow \begin{cases} \alpha_2 + (\alpha\beta)_{22} = 0 \\ e \\ \alpha_3 + (\alpha\beta)_{32} = 0 \end{cases}$$

Hipóteses: comparação de tipos de material (cont.)

- Temperatura 125°F:

$$H_0 : \begin{cases} \mu_{13} = \mu_{23} \\ e \\ \mu_{13} = \mu_{33} \end{cases} \leftrightarrow \begin{cases} \alpha_2 + (\alpha\beta)_{23} = 0 \\ e \\ \alpha_3 + (\alpha\beta)_{33} = 0 \end{cases}$$

Matrizes C

- Temperatura $15^{\circ}F$:

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

- Temperatura $75^{\circ}F$:

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

Matrizes C (cont.)

- Temperatura 215°F:

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

Resultados dos testes para as as hipóteses anteriores

- Hipótese (1): 0,66 (pvalor = 0,5269).
- Hipótese (2): 12,26 (pvalor = <0,0001).
- Hipótese (3): 2,12 (pvalor = 0,1400).

Hipóteses: comparação de tipos de material para a temperatura 75 °F

- Materias 2 e 3

$$H_0 : \left\{ \mu_{22} = \mu_{32} \right\} \leftrightarrow \left\{ \alpha_2 - \alpha_3 + (\alpha\beta)_{22} - (\alpha\beta)_{32} = 0 \right\}$$

- Materiais 1 e 2

$$H_0 : \left\{ \mu_{12} = \mu_{22} \right\} \leftrightarrow \left\{ \alpha_2 + (\alpha\beta)_{22} = 0 \right\}$$

Matrizes C

- Materiais 2 e 3:

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

- Materiais 1 e 2:

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

Resultados dos testes para as as hipóteses anteriores

- Hipótese (4): 2,00 ($p_{\text{valor}} = 0,1685$).
- Hipótese (5): 11,57 ($p_{\text{valor}} = <0,0021$).

Grupos

- Parece haver 4 grupos em relação ao tempo de vida:
 - Grupo 1: Componentes submetidos à temperatura de $15^{\circ}F$ independentemente do material utilizado.
 - Grupo 2: Componentes submetidos à temperatura de $75^{\circ}F$ feitos com material do tipo 1 .
 - Grupo 3: Componentes submetidos à temperatura de $750^{\circ}F$ feitos com material do tipo 2 ou do tipo 3 .
 - Grupo 4: Componentes submetidos à temperatura de $125^{\circ}F$ independentemente do material utilizado.

Modelo reduzido 1

$$Y_{ijk} = \mu_{ij} + \xi_{ijk},$$

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

- Erros $\xi_{ijk} \stackrel{i.i.d}{\sim} N(0, \sigma^2)$, μ_{ij} não aleatórios.
- Restrições (a seguir):

Modelo reduzido 1

- Neste caso

$$\mu_{11} = \mu$$

$$\mu_{21} = \mu$$

$$\mu_{31} = \mu$$

$$\mu_{12} = \mu + \delta_1$$

$$\mu_{22} = \mu + \delta_2$$

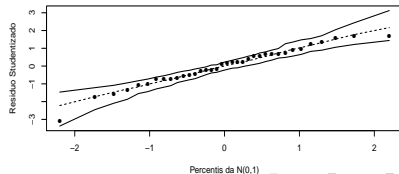
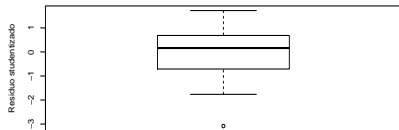
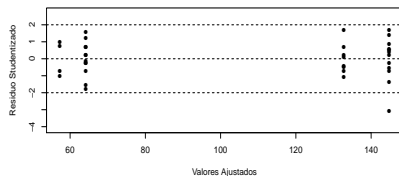
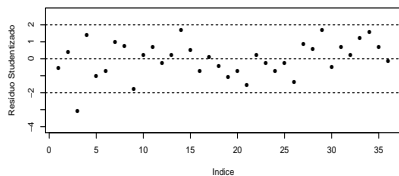
$$\mu_{32} = \mu + \delta_2$$

$$\mu_{13} = \mu + \delta_3$$

$$\mu_{23} = \mu + \delta_3$$

$$\mu_{33} = \mu + \delta_3$$

Análise de resíduos



Comentários

- Os testes sobre homocedasticidade continuam não rejeitando tal hipótese.
- O ajuste ficou um pior em relação ao modelo completo.
- 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
μ (Grupo 1)	144,833	7,79	[129,56; 160,10989]	18,58	< 0,0001
δ_1 (Grupo 2)	-87,58	15,59	[-118,13; -57,03]	-5,619	< 0,0001
δ_2 (Grupo 3)	-12,08	12,32	[-36,24; 12,07]	-0,980	0,3340
δ_3 (Grupo 4)	-80,67	11,02	[-102,27; -59,06]	-7,318	< 0,0001

Os grupos 1 e 3 são equivalentes (δ_2 não é significativamente diferente de 0). Os grupos 2 e 4 parecem ser equivalentes (em termos das médias). O teste $C\beta = \mathbf{0}$ forneceu os seguintes resultados: $q = 0,20$ (pvalor = 0,6602).

Grupos finais: segundo o modelo reduzido

- Parece haver somente 2 grupos em relação ao tempo de vida:
 - Grupo 1: Componentes submetidos à temperatura de $15^{\circ}F$ independentemente do material utilizado e componentes submetidos à temperatura de $750^{\circ}F$ feitos com materia do tipo 2 ou do tipo 3 .
 - Grupo 2: Componentes submetidos à temperatura de $750^{\circ}F$ feitos com material do tipo 1 e componentes submetidos à temperatura de $125^{\circ}F$ independentemente do material utilizado.

Modelo reduzido 2

$$Y_{ijk} = \mu_{ij} + \xi_{ijk},$$

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

- Erros $\xi_{ijk} \stackrel{i.i.d}{\sim} N(0, \sigma^2)$, μ_{ij} não aleatórios.
- Restrições (a seguir):

Modelo reduzido 1

- Neste caso

$$\mu_{11} = \mu$$

$$\mu_{21} = \mu$$

$$\mu_{31} = \mu$$

$$\mu_{12} = \mu + \delta_1$$

$$\mu_{22} = \mu$$

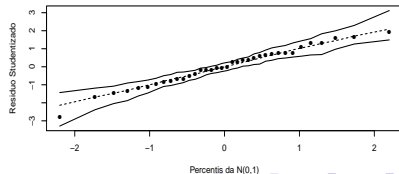
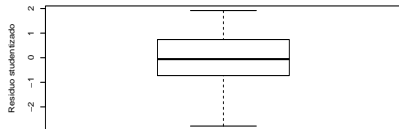
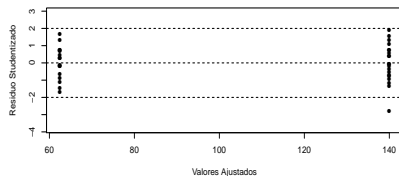
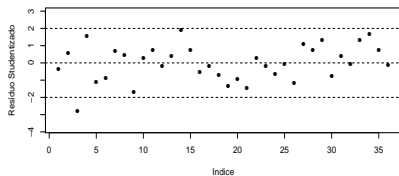
$$\mu_{32} = \mu$$

$$\mu_{13} = \mu + \delta_1$$

$$\mu_{23} = \mu + \delta_1$$

$$\mu_{33} = \mu + \delta_1$$

Análise de resíduos



Comentários

- Os testes sobre homocedasticidade continuam não rejeitando tal hipótese.
- O ajuste não mudou em relação ao modelo reduzido 1.
- 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
μ	140,00	5,96	[128,31; 151,69]	23,48	< 0,0001
δ_1	-77,56	8,94	[-95,09;-60,03]	-8,67	< 0,0001

Estimativas finais das médias

Grupo	Estimativa	EP	IC(95%)
Grupo 1	140,00	5,96	[128,31;151,69]
Grupo 2	62,44	6,67	[49,37; 75,50]

Gráfico de perfis médios ajustados via modelo reduzido 2

