

Planejamento e Análise Estatística de Experimentos de parcela dividida (“split-plot”)

Prof. Caio Azevedo

Contexto

- Em alguns experimentos fatoriais não é possível alocar as unidades experimentais de modo aleatório aos tratamentos (combinação entre os níveis dos fatores), mesmo em blocos ou em quadrado latino.
- Neste caso, em geral, tem-se um experimento fatorial chamado de parcela dividida ou “split-plot”.

Exemplo 18: Manufaturadora de papel

- Uma manufaturadora de papel está interessada em estudar o efeito de 3 tipos de métodos de preparação de polpa e 4 temperaturas diferentes de cozimento dela (polpa) para a produção de papel.
- Resposta: resistência à tração.
- Neste caso, temos $3 \times 4 = 12$ tratamentos e, portanto, precisaríamos de no mínimo 12 observações para testar a significância de todos os efeitos (fatores principais e interações).
- O analista decide realizar 3 repetições por tratamento (ou seja, obter 36 observações).

Exemplo 18: Manufaturadora de papel (cont.)

- No entanto, durante um dia de experimento, é possível realizar somente 12 ensaios.
- Então, o analista decidiu realizar o experimento em 3 dias. Em um determinado dia, uma porção de polpa é preparada sob um dos 3 métodos sob estudo.
- Então, esta porção é dividida em 4 amostras e cada uma é cozida sob uma determinada temperatura, escolhida aleatoriamente.

Exemplo 18: Manufaturadora de papel (cont.)

- Posteriormente, uma segunda porção é preparada, sob outro método. Esta segunda porção também é dividida em 4 amostras e cada uma é cozida sob uma determinada temperatura, escolhida aleatoriamente.
- Repete-se o processo em relação ao terceiro método.

Exemplo 18: Manufaturadora de papel (cont.)

- Em princípio, seríamos levados a analisar tal experimento com um fatorial (com dois fatores: método e temperatura) em blocos aleatorizados (dias).
- Se fosse este o caso, a ordem de alocação das amostras a cada tratamento seria completamente aleatória, dentro de cada bloco. Como descrito anteriormente, isto não se verifica.

Dados do Exemplo 18

Método de preparo	Repetição (Bloco) 1			Repetição (Bloco) 2			Repetição (Bloco) 3		
	1	2	3	1	2	3	1	2	3
Temperatura °F									
200	30	34	29	28	31	31	31	35	32
225	35	41	26	32	36	30	37	40	34
250	37	38	33	40	42	32	41	39	39
275	36	42	36	41	40	40	40	44	45

Cada repetição (bloco) é dividida em 3 partes (chamadas de parcelas totais ou “whole plots”). Cada parcela total é dividida em 4 partes (chamadas de subparcelas ou “split-plot”).

Modelo (geral) (M1)

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

$$i = 1, 2, \dots, r; j = 1, 2, \dots, a; k = 1, 2, \dots, b$$

- Os parâmetros $\alpha_i, \beta_j, (\alpha\beta)_{ij}$ estão associados as parcelas totais e correspondem, respectivamente, à: repetição (bloco), método de preparo e ao erro da parcela total (interação entre método de preparo e bloco).

Modelo (geral) (M1) (cont.)

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

$$i = 1, 2, \dots, r; j = 1, 2, \dots, a; k = 1, 2, \dots, b$$

- Os parâmetros $\gamma_k, (\alpha\gamma)_{ik}, (\beta\gamma)_{jk}, (\alpha\beta\gamma)_{ijk}$ estão associados às subparcelas e correspondem, respectivamente, à: temperatura, interação entre repetição (bloco) e temperatura, interação entre os tratamentos e o erro da subparcela (interação entre repetição, método de preparo e temperatura).

Comentários

- As somas de quadrados são calculadas como num modelo de fatores (completo) sem repetição por tratamentos.
- Vamos considerar os fatores e a interação entre eles como efeitos fixos (tem-se interesse apenas nos níveis observados).
- Vamos considerar as repetição (blocos) e as interações que dependam dele como efeitos aleatórios (interesse na população de repetições/blocos e considerar possível dependência entre os dados).

Voltando ao Modelo

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \xi_{ijk}$$
$$i = 1, 2, \dots, r; j = 1, 2, \dots, a; k = 1, 2, \dots, b$$

- $\beta_j, \gamma_k, (\beta\gamma)_{jk}$ não aleatórios.
- $\beta_1 = 0 = \gamma_1 = (\alpha\gamma)_{1k} = (\alpha\gamma)_{j1} = 0, \forall j, k.$
- $\xi_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2), \alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2), (\alpha\beta)_{ij} \stackrel{iid}{\sim} N(0, \sigma_{\alpha\beta}^2),$
 $(\alpha\gamma)_{ij} \stackrel{iid}{\sim} N(0, \sigma_{\alpha\gamma}^2), (\alpha\beta\gamma)_{ijk} \stackrel{iid}{\sim} N(0, \sigma_{\alpha\beta\gamma}^2),$ mutuamente independentes $\forall i, j, k.$
- Exercício: calcular as esperanças e variâncias marginais e condicionais, bem como a estrutura de correlação.

Esperanças das somas de quadrados (M1)

Parcela	Fator	Esperança do QM
Total	α_i (réplicas ou blocos, RB)	$\sigma^2 + ab\sigma_\alpha^2$
	β_j (Fator A, A)	$\sigma^2 + ab\sigma_{\alpha\beta}^2 + \frac{rb \sum_{j=1}^b \beta_j^2}{a-1}$
	$(\alpha\beta)_{ij}$ (RB \times Fator A, Erro da Parcela Total, RBA)	$\sigma^2 + b\sigma_{\alpha\beta}^2$
Subparcela	γ_k (Fator B, B)	$\sigma^2 + a\sigma_{\alpha\gamma}^2 + \frac{ra \sum_{k=1}^r \gamma_k^2}{b-1}$
	$(\alpha\gamma)_{ik}$ (RB \times Fator B, RBB)	$\sigma^2 + a\sigma_{\alpha\gamma}^2$
	$(\beta\gamma)_{jk}$ (Fator A \times Fator B, AB)	$\sigma^2 + \sigma_{\alpha\beta\gamma}^2 + \frac{r \sum_{j=1}^b \sum_{k=1}^r (\beta\gamma)_{jk}^2}{(a-1)(b-1)}$
	$(\alpha\beta\gamma)_{ijk}$ (RB \times Fator A \times Fator B, Erro da Subparcela, RBAB)	$\sigma^2 + \sigma_{\alpha\beta\gamma}^2$
	ξ_{jk}	σ^2 (não estimável)

Cont. (Estatísticas de teste) (M1)

Parcela	Fator	Estatística F
Total	α_i (réplicas ou blocos, RB)	-
	β_j (Fator A, A)	QM_A/QM_{RBA}
	$(\alpha\beta)_{ij}$ (RB \times Fator A, Erro da Parcela Total, RBA)	-
Subparcela	γ_k (Fator B, B)	QM_B/QM_{RBAB}
	$(\alpha\gamma)_{ik}$ (RB \times Fator B, RBB)	-
	$(\beta\gamma)_{jk}$ (Fator A \times Fator B, AB)	QM_{AB}/QM_{RBB}
	$(\alpha\beta\gamma)_{ijk}$ (RB \times Fator A \times Fator B, Erro da Subparcela, RBAB)	-
	ξ_{jk}	-

Comentários

- Essencialmente, a justificativa para as fórmulas das estatísticas F é o comportamento dos quadrados médios (Teorema de Cochran) e de seus respectivos valores esperados (este, em particular, sob H_0).
- Alguns autores propõem um outro modelo (slide a seguir) que, basicamente, consiste em desconsiderar os termos relativos às interações: Bloco \times B $(\alpha\gamma)_{ik}$ e Bloco \times A \times B $(\alpha\beta\gamma)_{ijk}$, transportando os respectivos “efeitos” para os erros.

Outro Modelo (M2)

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + (\beta\gamma)_{jk} + \xi_{ijk}$$
$$i = 1, 2, \dots, r; j = 1, 2, \dots, a; k = 1, 2, \dots, b$$

- $\beta_j, \gamma_k, (\beta\gamma)_{jk}$ não aleatórios.
- $\beta_1 = 0 = \gamma_1 = (\alpha\gamma)_{1k} = (\alpha\gamma)_{j1} = 0, \forall j, k.$
- $\xi_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2), \alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2), (\alpha\beta)_{ij} \stackrel{iid}{\sim} N(0, \sigma_{\alpha\beta}^2)$, mutuamente independentes $\forall i, j, k.$
- Exercício: calcular as esperanças e variâncias marginais e condicionais, bem como a estrutura de correlação.

Esperanças das somas de quadrados (M2)

Parcela	Fator	Esperança do QM
Total	α_i (réplicas ou blocos, RB)	$\sigma^2 + ab\sigma_\alpha^2$
	β_j (Fator A, A)	$\sigma^2 + ab\sigma_{\alpha\beta}^2 + \frac{rb \sum_{j=1}^b \beta_j^2}{a-1}$
	$(\alpha\beta)_{ij}$ (RB \times Fator A, Erro da Parcela Total, RBA)	$\sigma^2 + b\sigma_{\alpha\beta}^2$
Subparcela	γ_k (Fator B, B)	$\sigma^2 + a\sigma_{\alpha\gamma}^2 + \frac{ra \sum_{k=1}^r \gamma_k^2}{b-1}$
	$(\beta\gamma)_{jk}$ (Fator A \times Fator B, AB)	$\sigma^2 + \sigma_{\alpha\beta\gamma}^2 + \frac{r \sum_{j=1}^b \sum_{k=1}^r (\beta\gamma)_{jk}^2}{(a-1)(b-1)}$
	ξ_{jk} (Erro da subparcela, R)	σ^2

Cont. (Estatísticas de teste) (M2)

Parcela	Fator	Estatística F
Total	α_i (réplicas ou blocos, RB)	-
	β_j (Fator A, A)	QM_A/QM_{RBA}
	$(\alpha\beta)_{ij}$ (RB \times Fator A, Erro da Parcela Total, RBA)	-
Subparcela	γ_k (Fator B, B)	QM_B/QMR
	$(\beta\gamma)_{jk}$ (Fator A \times Fator B, AB)	QM_{AB}/QMR
	ξ_{jk} (Erro da subparcela, R)	-

Comentários

- Vamos apresentar as duas tabelas ANOVA.
- Contudo, apresentaremos somente a análise para o modelo M2 (considerando nula a interação entre bloco e fator A ($\sigma_{\alpha\beta}^2 = 0$)).
- O modelo M1 não é estimável, devido à quantidade de parâmetros/efeitos aleatórios presentes e a quantidade de observações disponíveis.
- Vamos abrir uma exceção e apresentar as tabelas Anova antes de apresentar o diagnóstico do modelo.
- Antes, vamos apresentar análises descritivas.

Número de observações

- Por cada nível de cada um dos três fatores: 12.
- Por cada combinação dos níveis dos fatores dois a dois: 3.
- Por cada combinação dos níveis dos três fatores: 1.
- Obs: é inviável, inclusive, estimar o modelo completo apenas com efeitos fixos.

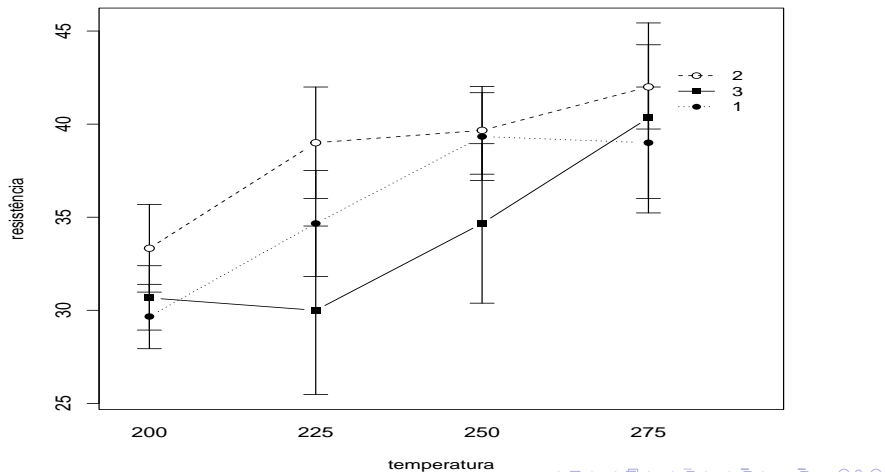
Medida resumo (Método \times Temperatura)

Método	Temp.	Média	DP	Var.	Min.	Max.	CV(%)
1	200	29,67	1,53	2,33	28,00	31,00	5,15
	225	34,67	2,52	6,33	32,00	37,00	7,26
	250	39,33	2,08	4,33	37,00	41,00	5,29
	275	39,00	2,65	7,00	36,00	41,00	6,78
2	200	33,33	2,08	4,33	31,00	35,00	6,24
	225	39,00	2,65	7,00	36,00	41,00	6,78
	250	39,67	2,08	4,33	38,00	42,00	5,25
	275	42,00	2,00	4,00	40,00	44,00	4,76
3	200	30,67	1,53	2,33	29,00	32,00	4,98
	225	30,00	4,00	16,00	26,00	34,00	13,33
	250	34,67	3,79	14,33	32,00	39,00	10,92
	275	40,33	4,51	20,33	36,00	45,00	11,18

Medida resumo (Bloco)

Bloco	Média	DP	Var.	Mín.	Max.	cv(%)
1	34,75	4,73	22,39	26,00	42,00	13,62
2	35,25	5,08	25,84	28,00	42,00	14,42
3	38,08	4,42	19,54	31,00	45,00	11,61

Gráfico de perfis médios



Comentários

- Parece haver interação entre os fatores de interesse e, conseqüentemente efeito de cada fator.
- Possível efeito crescente da temperatura (ao menos para algum(ns) método e eventual igualdade entre dois ou os três métodos, para algumas temperaturas.
- Parece não haver efeito (marginal) de bloco.

ANOVA (M1)

FV	SQ	GL	QM	Estat. F	pvalor
Repetição (bloco)	77,56	2	38,78		
Método	128,39	2	64,19	7,08	0,0485
Parcela total (método \times bloco)	36,28	4	9,07		
Temperatura	434,08	3	144,69	42,01	0,0002
Método \times temperatura	75,17	6	12,53	2,96	0,0520
Temperatura \times bloco	20,67	6	3,44		
(Subparcela) Método \times Temperatura \times Bloco	50,83	12	4,24		

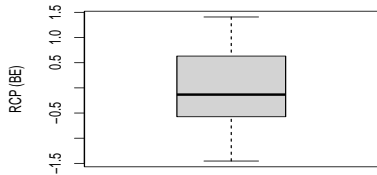
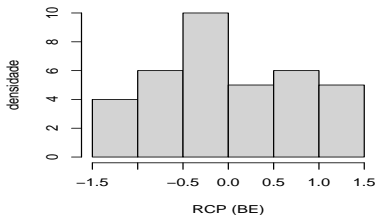
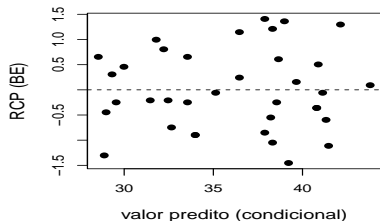
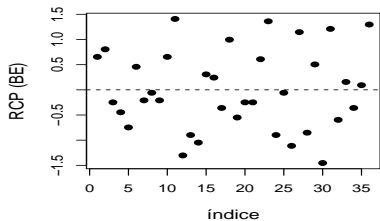
ANOVA (M2) com ($\sigma_{\alpha\beta}^2 = 0$)

FV	SQ	GL	QM	Estat. F	pvalor
Repetição (bloco)	77,56	2	38,78		
Método	128,39	2	64,19	7,08	0,0485
Parcela total (método \times bloco)	36,28	4	9,07		
Temperatura	434,08	3	144,69	36,43	< 0,0001
Método \times temperatura	75,17	6	12,53	3,15	0,0271
Resíduos	18	71,50	3,97		

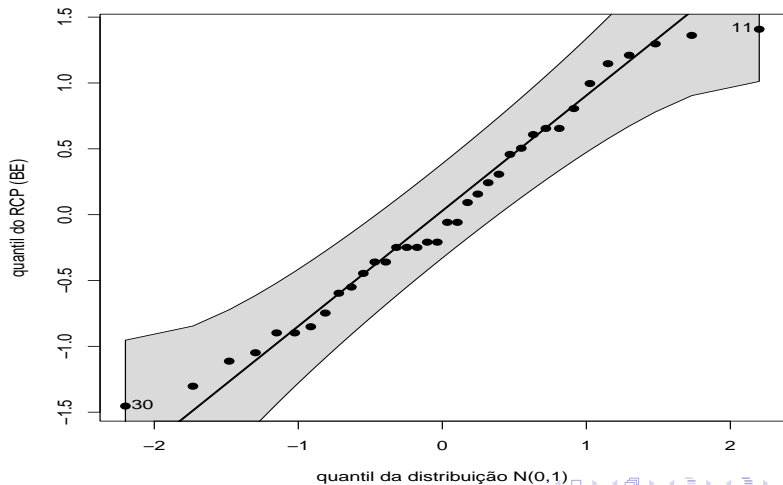
Comentários

- Sob ambos os modelos, há efeito marginal de interação.
- As hipóteses testadas pelas outras duas estatísticas F (Método e Temperatura) são ambas significativas, sendo que a primeira o é de forma marginal.
- Vamos continuar a análise via modelo M2 (com $(\sigma_{\alpha\beta}^2 = 0)$).
- Obs: O RMP (resíduo condicional padronizado (MQ)) não fora calculado pela função *hlm_diag* pois, provavelmente, há poucos blocos (níveis do efeito aleatório).

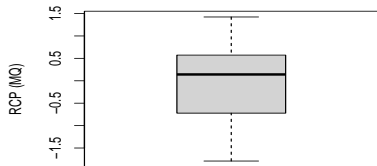
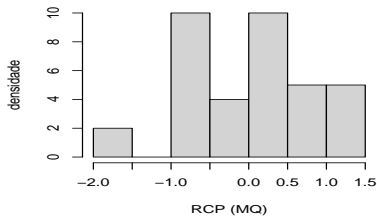
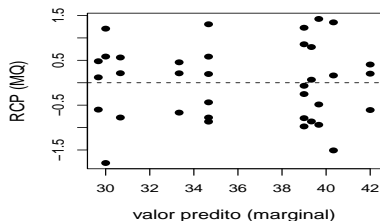
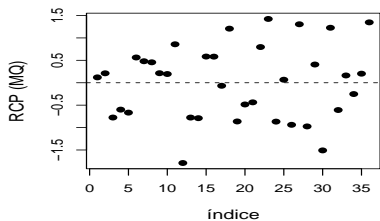
Análise de resíduos (RCP (BE)) (M2)



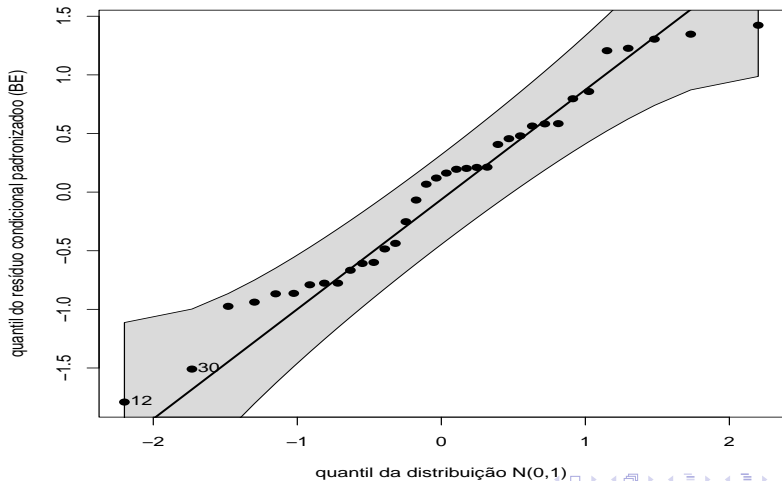
QQplot (RCP (BE)) (M2)



Análise de resíduos (RMP) (M2)

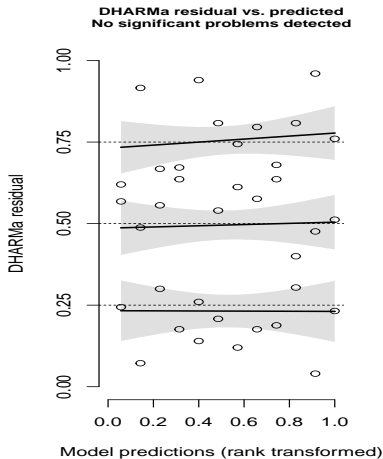
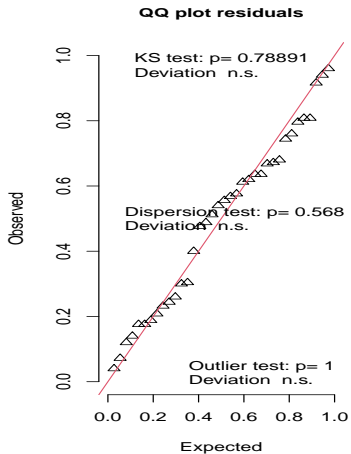


QQplot (RMP) (M2)

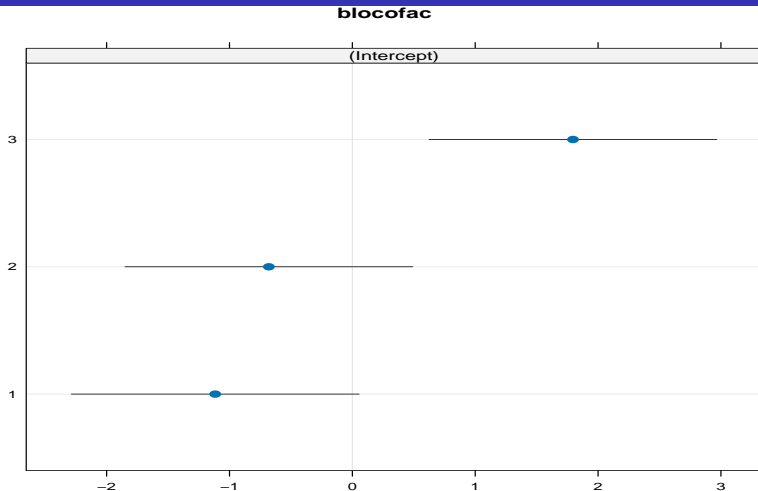


Análise de resíduos (RQ)

DHARMA residual



Estimativas dos efeitos aleatórios



Comentários

- O modelo parece bem ajustado.
- Há efeito de bloco e, aparentemente, os três parecem afetar as médias condicionais.
- Vamos fazer alguns testes de interesse.

Pacotes no R para testes $H_0 : C\beta = \mathbf{0}$.

- Funções *emmeans* e *contrast* do pacote **emmeans** (usaremos este).
- Função *linearHypothesis* do pacote **car**.
- Função *lsmeans* do pacote **lsmeans**.
- Obs: pelo que observei o pacote *emmeans* admite apenas hipóteses com **C** sendo um vetor (não uma matriz).
- Testamos a igualdade entre as médias da resistências, dentro de cada temperatura, entre os diferentes tipos de material.

Testes

tempfac = 200:

contrast	estimate	SE	df	t.ratio	p.value
PS	-3.667	1.81	22	-2.029	0.0547
PT	-1.000	1.81	22	-0.553	0.5856
ST	2.667	1.81	22	1.476	0.1542

tempfac = 225:

contrast	estimate	SE	df	t.ratio	p.value
PS	-4.333	1.81	22	-2.398	0.0254
PT	4.667	1.81	22	2.582	0.0170
ST	9.000	1.81	22	4.980	0.0001

Testes

tempfac = 250:

contrast	estimate	SE	df	t.ratio	p.value
PS	-0.333	1.81	22	-0.184	0.8554
PT	4.667	1.81	22	2.582	0.0170
ST	5.000	1.81	22	2.767	0.0113

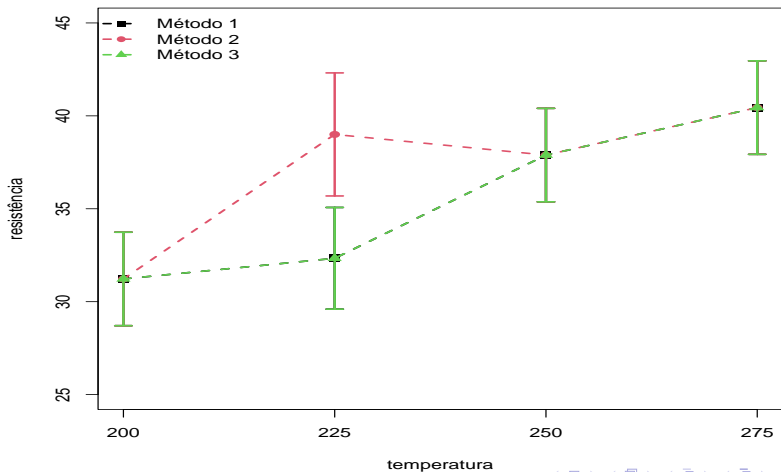
tempfac = 275:

contrast	estimate	SE	df	t.ratio	p.value
PS	-3.000	1.81	22	-1.660	0.1111
PT	-1.333	1.81	22	-0.738	0.4684
ST	1.667	1.81	22	0.922	0.3664

Testes

- Com o cuidado de não serem testes simultâneos dentro de cada temperatura (nível de significância global), conclui-se que:
 - Para as temperaturas 200, 350 e 275, não há nenhuma diferença.
 - Para a temperatura 225, a média para o tipo de material 2 é diferente das outras duas.
- Obs: há detalhes sobre os graus de liberdade (df) que devem ser melhor investigados.

Gráfico de perfis médios ajustados



Comentários

- Com base nos resultados anteriores e no gráfico de perfis médios, conjecturamos a possibilidade de que as seguintes médias podem ser iguais entre si (Grupo 1):
 - Média do método 2 sob a temperatura de 225 °C.
 - Média das médias dos método 1, 2 e 3 sob a temperatura de 250 °C.
 - Média das médias dos método 1, 2 e 3 sob a temperatura de 275 °C.
- Outra igualdade (Grupo 2), média das médias dos três métodos sob a temperatura 200 e a média das médias dos métodos 1 e 3 sob a temperatura 225.
- Em termos dos parâmetros do modelo temos (próximo slide)

Hipótese adicional (cont.)

$$\begin{cases} \frac{\mu_{11} + \mu_{21} + \mu_{31}}{3} = \frac{\mu_{12} + \mu_{32}}{2} \\ \mu_{22} = \frac{\mu_{13} + \mu_{23} + \mu_{33}}{3} \\ \mu_{22} = \frac{\mu_{14} + \mu_{24} + \mu_{34}}{3} \end{cases}$$

$$\Leftrightarrow \begin{cases} \frac{3\mu + \beta_2 + \beta_3}{3} = \frac{2\mu + \beta_3 + 2\gamma_2 + (\beta\gamma)_{32}}{2} \\ \mu + \beta_2 + \gamma_2 + (\beta\gamma)_{22} = \frac{3\mu + \beta_2 + \beta_3 + 3\gamma_3 + (\beta\gamma)_{23} + (\beta\gamma)_{33}}{3} \\ \mu + \beta_2 + \gamma_2 + (\beta\gamma)_{22} = \frac{3\mu + \beta_2 + \beta_3 + 3\gamma_4 + (\beta\gamma)_{24} + (\beta\gamma)_{34}}{3} \end{cases}$$

Hipótese adicional (cont.)

$$\Leftrightarrow \begin{cases} -2\beta_2 + \beta_3 + 6\gamma_2 + 3(\beta\gamma)_{32} = 0 \\ -2\beta_2 + \beta_3 - 3\gamma_2 + 3\gamma_3 - 3(\beta\gamma)_{22} + (\beta\gamma)_{23} + (\beta\gamma)_{33} = 0 \\ -2\beta_2 + \beta_3 - 3\gamma_2 + 3\gamma_4 - 3(\beta\gamma)_{22} + (\beta\gamma)_{24} + (\beta\gamma)_{34} = 0 \end{cases}$$

Hipótese adicional (cont.)

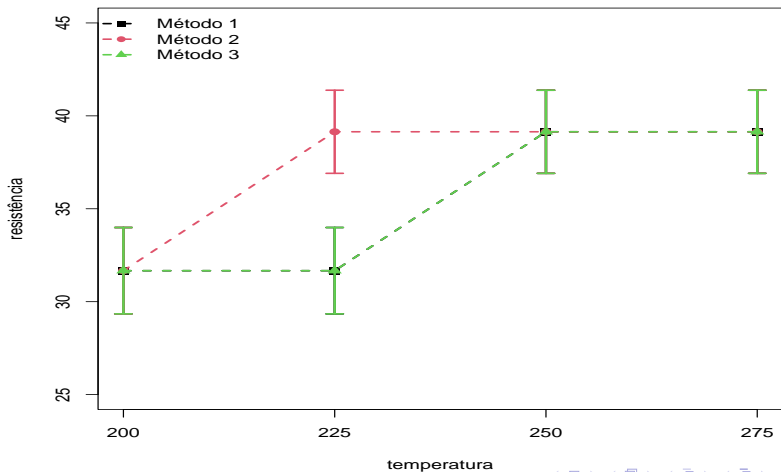
- Resultado: Estatística do teste(p-valor) - 6,92 (0,07475) (usando a correção de Bonferroni, não rejeitamos H_0).
- No próximo slide temos a configuração final dos perfis médios, adicionado a última igualdade.
- Média das médias (Grupo 1)

$$\mu^* = \frac{\mu_{13} + \mu_{14} + \mu_{22} + \mu_{23} + \mu_{24} + \mu_{33} + \mu_{34}}{7}.$$

(Grupo 2)

$$\mu^{**} = \frac{\mu_{11} + \mu_{12} + \mu_{21} + \mu_{31} + \mu_{32}}{5}.$$

Gráfico de perfis médios ajustados finais



Comentários

- Melhores desempenhos (equivalentes entre si): métodos 1, 2, 3 sob as temperaturas 250 e 275 e via método 2 sob temperatura 225
- Piores desempenhos (equivalentes entre si): métodos 1, 2, 3 sob a temperatura 200 e métodos 1 e 3, sob a temperatura 225.