Planejamento e Análise Estatística de Experimentos fatoriais hierárquicos

Prof. Caio Azevedo

Contexto

- Já vimos algumas possibilidades em termos de planejamento e análise de experimentos fatoriais não hierárquicos.
- Fatores não hierárquicos: os níveis do fator B, para cada nível do fator A, são os mesmos.
- Exemplo: Fator A: temperatura (^{o}C): (15,20,25), Fator B pressão (MPa): (100,120,140).
- Em alguns casos os níveis dos fatores são similares mas, não exatamente os mesmos. Neste caso existe uma hierarquia entre os fatores e a análise é diferente de quando os fatores não são hierárquicos.

Exemplo 16

- Considere que uma companhia compra a matéria prima necessária para desenvolver seus processos industriais de 3 diferentes fornecedores. A companhia deseja verificar se a pureza das matérias primas difere ou não entre os fornecedores. Dispõe-se de quatro amostras de matéria prima de cada forneceder e 3 medidas da pureza foram realizadas em cada amostra.

Representação dos dados do Exemplo 16

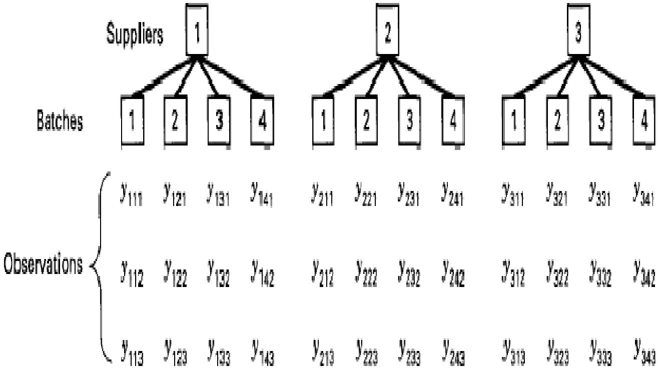


Figure 13-1 A two-stage nested design.

Cont. Exemplo 16

- Este é um exemplo de um experimento hierárquico de dois estágios (dois fatores).
- As amostras de matéria-prima estão hierarquizadas (encaixadas) dentro de cada fornecedor.
- Note que cada amostra é única, portanto, os níveis desse fator não são os mesmos ao longo dos fornecedores.
- Assim, trata-se de um experimento (fatorial) hierárquico.

Modelo (geral) efeitos fixos - casela de referência

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \xi_{(ij)k}$$

$i = 1, 2, ..., a$ (Fator A); $j = 1, 2, ..., b$ (Fator B); $k = 1, 2, ..., n$ (repetição)

- $\xi_{(ij)k} \overset{i.i.d}{\sim} N(0, \sigma^2)$.
- $\alpha_1 = \beta_{1(i)} = 0, \forall i$.
- No exemplo exemplo:

  $\boldsymbol{\beta} = (\mu, \alpha_2, \alpha_3, \beta_{2(1)}, \beta_{3(1)}, \beta_{4(1)}, \beta_{2(2)}, \beta_{3(2)}, \beta_{4(2)}, \beta_{2(3)}, \beta_{3(3)}, \beta_{4(3)})$.

Modelo (geral) efeitos aleatórios

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \xi_{(ij)k}$$

$i = 1, 2, ..., a$ (Fator A); $j = 1, 2, ..., b$ (Fator B); $k = 1, 2, ..., n$ (repetição)

- $\xi_{(ij)k} \overset{i.i.d}{\sim} N(0, \sigma^2)$.
- $\alpha_i \overset{i.i.d}{\sim} N(0, \sigma^2_\alpha), \beta_{j(i)} \overset{i.i.d}{\sim} N(0, \sigma^2_\beta)$
- $\xi_{(ij)k}, \alpha_i, \beta_{j(i)}$ mutuamente independentes $\forall i, j, k$.
- No exemplo exemplo $\boldsymbol{\beta} = (\mu)$.

Modelo (geral) efeitos mistos - casela de referência

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \xi_{(ij)k}$$

$i = 1, 2, ..., a$ (Fator A); $j = 1, 2, ..., b$ (Fator B); $k = 1, 2, ..., n$ (repetição)

- $\xi_{(ij)k} \overset{i.i.d}{\sim} N(0, \sigma^2)$.
- $\alpha_1 = 0, \beta_{j(i)} \overset{i.i.d}{\sim} N(0, \sigma^2_{\beta})$. Em nosso exmplo: $\boldsymbol{\beta} = (\mu, \alpha_2, \alpha_3)$
- Ou $\alpha_i \overset{i.i.d}{\sim} N(0, \sigma^2_{\alpha}), \beta_{1(i)} = 0, \forall i$. Em nosso exemplo:

  $\boldsymbol{\beta} = (\beta_{2(1)}, \beta_{3(1)}, \beta_{4(1)}, \beta_{2(2)}, \beta_{3(2)}, \beta_{4(2)}, \beta_{2(3)}, \beta_{3(3)}, \beta_{4(3)})$.
- $\xi_{(ij)k}, \alpha_i$ mutuamente independentes $\forall i, j, k$ ou $\xi_{(ij)k}, \beta_{j(i)}$ mutuamente independentes $\forall i, j, k$.

Comentários

- **Exercício: obter as estruturas de covariância para cada modelo**.
- A decomposição das somas de quadrados é a mesma, independentemente do modelo utilizado.
- Contudo, as hipóteses de interesse e as estatísticas dos teste mudam conforme o modelo.
- Mais detalhes veja a página 558 do livro do Montgomery.

Decomposição das somas de quadrados

$$
\begin{aligned}
\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\overline{Y}_{...}\right)^2 &= bn\sum_{i=1}^{a}\left(\overline{Y}_{i..}-\overline{Y}_{...}\right)^2 + n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{Y}_{ij.}-\overline{Y}_{i..}\right)^2 \\
&+ \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\overline{Y}_{ij.}\right)^2 \\
SQT &= SQF_A + SQF_{B(A)} + SQR
\end{aligned}
$$

Esperança das somas de quadrados

Esperança	A & B Fixos	A Fixo e B aleatório	A e B aleatórios
$\mathcal{E}(QMF_A)$	$\sigma^2 + \dfrac{bn\sum_{i=1}^{a}\alpha_i^2}{a-1}$	$\sigma^2 + n\sigma_\beta^2 + \dfrac{bn\sum_{i=1}^{a}\alpha_i^2}{a-1}$	$\sigma^2 + n\sigma_\beta^2 + bn\sigma_\tau^2$
$\mathcal{E}(QMF_{B(A)})$	$\sigma^2 + \dfrac{n\sum_{i=1}^{a}\sum_{j=1}^{b}\alpha_i^2}{a(b-1)}$	$\sigma^2 + n\sigma_\beta^2$	$\sigma^2 + n\sigma_\beta^2$
$\mathcal{E}(QMR)$	σ^2	σ^2	σ^2

Tabela de análise de variância

FV	SQ	GL	QM	Estatística F	pvalor
Fator A	SQF_A	a-1	$QMF_A = \frac{SQF_A}{(a-1)}$	F_A	$\min(F(f_A\|H_0), S(f_A\|H_0))$
B dentro de A	$SQF_{B(A)}$	a(b-1)	$QMF_B = \frac{SQF_{B(A)}}{(a(b-1))}$	F_B	$\min(F(f_B\|H_0), S(f_B\|H_0))$
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.

Hipóteses de interesse/Estatística do Teste

- Ambos os fatores fixos:

$$H_0 : \alpha_i = 0, \forall i \text{ vs } H_1 : \text{ pelo menos um parâmetro diferente de 0}$$

$$H_0 : \beta_{j(i)} = 0, \forall i, j \text{ vs } H_1 : \text{ pelo menos um parâmetro diferente de 0}$$

Estatísticas: Respectivamente, $F_A = \frac{QMF_A}{QMR}; F_B = \frac{QMF_{B(A)}}{QMR}$

Hipóteses de interesse/Estatística do Teste (cont.)

- Fator A fixo e fator B aleatório:

$$H_0 : \alpha_i = 0, \forall i \text{ vs } H_1 : \text{ pelo menos um parâmetro diferente de } 0$$

$$H_0 : \sigma^2_\beta = 0; H_1 : \sigma^2_\beta > 0$$

Estatísticas: Respectivamente, $F_A = \frac{QMF_A}{QMF_{B(A)}}; F_B = \frac{QMF_{B(A)}}{QMR}$

Hipóteses de interesse/Estatística do Teste (cont.)

- Ambos os fatores aleatórios:

$$H_0 : \sigma_{\alpha}^2 = 0; H_1 : \sigma_{\alpha}^2 > 0$$

$$H_0 : \sigma_{\beta}^2 = 0; H_1 : \sigma_{\beta}^2 > 0$$

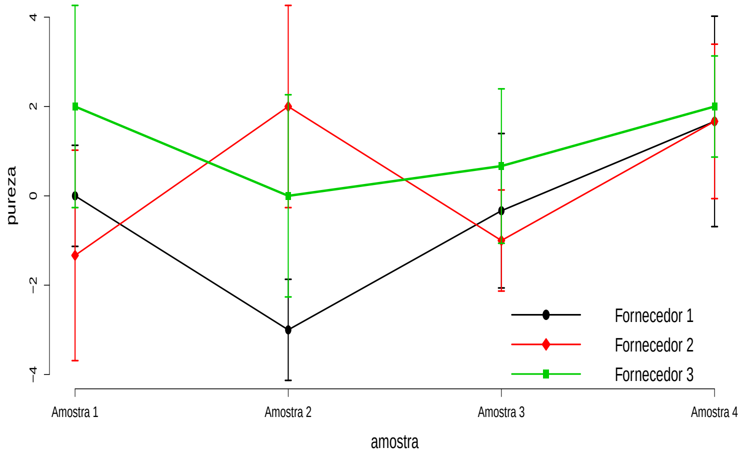
Estatísticas: Respectivamente, $F_A = \frac{QMF_A}{QMF_{B(A)}}; F_B = \frac{QMF_{B(A)}}{QMR}$

Dados do. Exemplo 16

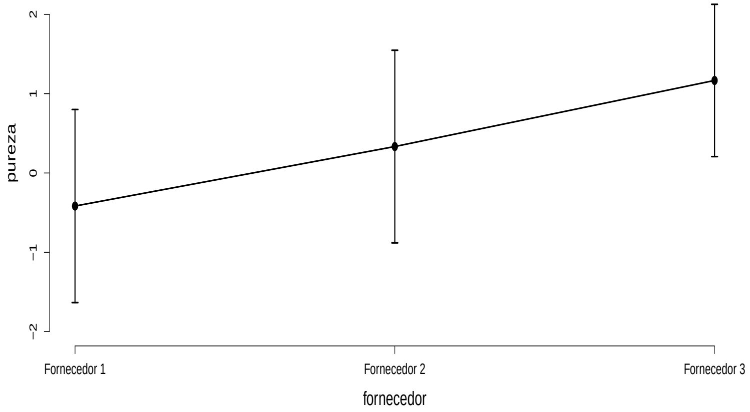
	Fornecedor 1				**Fornecedor 2**				**Fornecedor 3**			
Porção de matéria prima	1	2	3	4	1	2	3	4	1	2	3	4
	1	-2	-2	1	1	0	-1	0	2	-2	1	3
	-1	-3	0	4	-2	4	0	3	4	0	-1	2
	0	-4	1	0	-3	2	-2	2	0	2	2	1

Valores relativos à um índice de pureza após subtrair o valor 93.

Gráficos de perfis médios



Gráficos de perfis médios: por fornecedor



Medidas descritivas

Fornecedor	Amostra	Média	DP	Var	Mínimo	Máximo	$\lvert CV(\%)\rvert$
1	1	0,00	1,00	1,00	-1,00	1,00	-
	2	-3,00	1,00	1,00	-4,00	-2,00	33,33
	3	-0,33	1,53	2,33	-2,00	1,00	458,26
	4	1,67	2,08	4,33	0,00	4,00	124,90
2	5	-1,33	2,08	4,33	-3,00	1,00	156,12
	6	2,00	2,00	4,00	0,00	4,00	100,00
	7	-1,00	1,00	1,00	-2,00	0,00	100,00
	8	1,67	1,53	2,33	0,00	3,00	91,65
3	9	2,00	2,00	4,00	0,00	4,00	100,00
	10	0,00	2,00	4,00	-2,00	2,00	-
	11	0,67	1,53	2,33	-1,00	2,00	229,13
	12	2,00	1,00	1,00	1,00	3,00	50,00

Medidas descritivas: por fornecedor

Fornecedor	Média	DP	Var	Mínimo	Máximo	$\|CV(\%)\|$
1	-0,42	2,15	4,63	-4,00	4,00	516,35
2	0,33	2,15	4,61	-3,00	4,00	643,85
3	1,17	1,70	2,88	-2,00	4,00	145,43

Comentários

- A rigor, neste problema, faz sentido:
  - Tratar o fator **fornecedor** como fixo, pois a companhia só trabalha com esses 3 fornecedores.
  - Tratar o fator **amostra** como aleatório pois, tem-se interesse na produção de cada fornecedor como um todo.
- Vejamos como fica a análise em se tratando ambos os fatores como fixos e como descrito acima.

Situação 1: Ambos os fatores fixos

- Média para cada amostra dentro de cada fornecedor

$$\mathcal{E}(Y_{ijk}) = \mu_{ij} = \mu + \alpha_i + \beta_{j(i)}$$

- Média para cada fornecedor:

$$\mu_{i.} = \frac{1}{b}\sum_{j=1}^{b} \mu_{ij} = \mu + \alpha_i + \overline{\beta}_{.(i)}, \overline{\beta}_{.(i)} = \frac{1}{b}\sum_{j=1}^{b} \beta_{j(i)}$$

Situação 2: Fornecedor (fixo) e amostra (aleatório)

- Média para cada amostra dentro de cada fornecedor

  $\mathcal{E}(Y_{ijk}) = \mu_{ij} = \mu + \alpha_i$, pois $\mathcal{E}(\boldsymbol{\beta}_{j(i)}) = 0$

- Média para cada fornecedor:

  $\mu_{i.} = \mu + \alpha_i$

- Pode-se pensar no valor predito para cada amostra dentro de cada fornecedor, ou seja:

  $\widehat{Y}_{ijk} = \widehat{\mu} + \widehat{\alpha}_i + \widehat{\beta}_{(j)i}$

- Distribuição condicional dado a amostra:

  $Y_{ijk}|\beta_{j(i)} \sim N(\mu + \alpha_i + \beta_{j(i)}, \sigma^2)$

Aspectos Inferenciais

- Para os modelos de feitos fixos, os estimadores para $\boldsymbol{\beta}$ são obtidos através do método de mínimos quadrados, σ^2 é estimado pelo QMR e a ANOVA pode ser obtida a partir da decomposição das somas de quadrados.
- Para os modelos de feitos aleatórios (mistos), os estimadores para $\boldsymbol{\beta}$ são obtidos, em geral, através do método de máxima verossimilhança restrita assim como as componentes de variância ($\sigma^2_{(.)}$) (estes podem ou não coincidir com os estimadores pelo método dos momentos). Em muitos casos, também, a ANOVA pode ser obtida a partor da decomposição das somas de quadrados.

Aspectos Computacionais

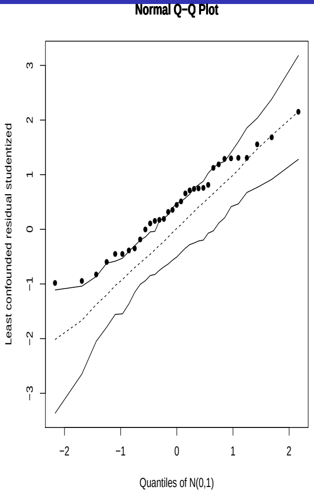
- Em nosso contexto:
  - Modelos de efeitos fixos: pode-se usar a função **lm** tanto para se obter a ANOVA quanto as estimativas dos parâmetros.
  - Modelos de efeitos mistos (aleatórios): pode-se usar a função **lm** ou **aov**, em geral, para se obter a ANOVA. As estimativas dos parâmetros, em geral, não podem ser obtidas utilizando-se a função **lm**. Tem-se que usar a função **lmer**, por exemplo.

ANOVA considerando a Situação 1

FV	SQ	GL	QM	Estatística F	pvalor
Fornecedor	15,06	2	7,53	2,85	0,0774
Amostra dentro de fornec.	69,92	9	7,77	2,94	0,0167
Resíduo	63,33	24	2,64		

Há efeito de ambos os fatores. Faz sentido comparar as amostras dentro de cada fornecedor (cautela, depende do interesse da companhia) e os fornecedores entre si.

Análise residual: Sit. 2 (resíduo de confundimento mínimo)

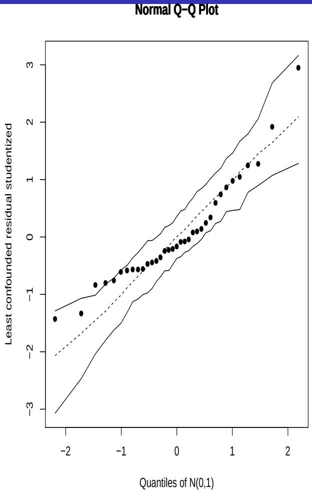


ANOVA considerando a Situação 2

FV	SQ	GL	QM	Estatística F	pvalor
Fornecedor	15,06	2	7,53	0,97	0,4158
Amostra dentro de fornec.	69,92	9	7,77	2,94	0,0167
Resíduo	63,33	24	2,64		

Há efeito apenas de amostra. Ajustar um modelo reduzido considerando equivalência entre os fornecedores. Note que, como estamos assumindo uma única variância para os efeitos aleatórios de todas as amostras, tal variabilidade é devida a amostras de pelo menos um fornecedor (não necessariamente todos).

Análise residual: Sit. 2 - modelo reduzido (RCM)



Comentários

- Para ambos os modelos, o ajuste aos dados não foi adequado.
- Outras distribuições para a variável resposta e para os efeitos aleatórios, devem ser usadas.
- Há variabilidade significativa (nos termos expostos anteriormente) entre as amostras.
- Os níveis de pureza médios das amostras, entre os fornecedores, mostraram-se equivalentes à (estimativa, erro-padrão e IC 95%): 0,36 (0,46)[-0,54;1,27].

Componentes de variância

- Pode-se obter estimativas para os componentes de variância $(\sigma^2, \sigma^2_\beta, \sigma^2_\alpha)$ de modo semelhante ao modelo de efeitos aleatórios (ver página 565 do livro do Montgomery).
- Computacionalmente, tais estimativas podem ser obtidas através da função **lmer**.
- Em nosso exemplo, temos que $\widetilde{\sigma}^2 = 2,64$ e $\widetilde{\sigma}^2_\beta = 1,71$ o que implica que $\widetilde{\rho} = \frac{\widetilde{\sigma}^2_\beta}{\widetilde{\sigma}^2_\beta + \widetilde{\sigma}^2} = 0,40$.

Comentários

- **Caso tivesse sido detectada diferença dos índices médios de pureza entre os fornecedores, poderíamos usar a estatística dos testes $C\boldsymbol{\beta}$.**
- Contudo, as estimativas do vetor $\boldsymbol{\beta} = (\mu, \alpha_2, \alpha_3)$ devem ser obtidas usando-se a função **lmer** e não a função **lm**.
- O programa disponibilizado para se realizar o teste $C\boldsymbol{\beta}$ tem de ser adaptado para se usar a saída da função **lmer** (fazer as adaptações!).