Exeprimentos com fatores (efeitos) aleatórios

Prof. Caio Azevedo

Contexto

- Até agora vimos experimentos com fatores fixos.
- Fator fixo: tem-se interesse apenas em se fazer inferência para os níveis observados.
- Ou, para níveis intermediários (fatores quantitativos) via modelos de regressão.
- Em muitos casos, tem-se interesse em se fazer inferência para um número muito maior (as vezes infinito) de níveis dos fatores.
- Como devemos planejar e analisar experimentos sob esse intuito?

Exemplo 15

- Uma empresa têxtil fabrica tecidos com um grande número de teares. É de interesse da fábrica que os teraes tenham um comportamento homogêneo entre si de tal forma que os tecidos fabricados tenham resistência uniforme. Um engenheiro de produção suspeita que, além de existir a variablidade usual na resistência dentro das amostras de tecido do mesmo tear, pode existir variações significativas na resistência entre teares. Para investigar este aspecto, foram selcionados, aleatoriamente, quatro teares e quatro medidas de resistência foram feitas no mesmo tecido de cada terar (4 tecidos).

Resultados do experimento

Tear	Observação			
	1	2	3	4
1	98	97	99	96
2	91	90	93	92
3	96	95	97	95
4	95	96	99	98

Modelagem

$$Y_{ij} = \mu + \tau_i + \xi_{ij}$$

$i = 1, 2, ..., a$ (níveis do fator) $j = 1, 2, ..., n$ (repetição), pode-se considerar desbalanceamento: n_i

- μ não aleatório, $\tau_i \overset{i.i.d.}{\sim} N(0, \sigma_\tau^2)$, $\xi_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$.
- $\tau_i \perp \xi_{ij}, \forall i, j$ (mutuamente independentes).

Propriedades do modelo (exercício)

- $\mathcal{E}(Y_{ij}) = \mu$, $\mathcal{V}(Y_{ij}) = \sigma_\tau^2 + \sigma^2$.
- $Cov(Y_{ij}, Y_{i'j'}) = \begin{cases} 0, \text{se } i \neq i' \\ \sigma_\tau^2 + \sigma^2, \text{se } i = i' \text{ e } j = j' \\ \sigma_\tau^2, \text{se } i = i' \end{cases}$
- As observações não são mais mutuamente independentes.
- A decomposição usual da soma de quadrados permanece válida $SQT = SQF + SQR$ contudo, o teste usual deixa de ter sentido, dado que queremos comparar o desempenho de todos os níveis (de todos os teares) inclusive considerando aqueles que não foram observados.

Ausência de efeito de fator

- Os teares são equivalente entre si (ou seja, não existe feito de fator) se $\tau_i \equiv 0$. Em outras palavras, nossas hipóteses de interesse são:

$$H_0 : \sigma^2_\tau = 0$$

$$H_1 : \sigma^2_\tau > 0$$

- Assim como antes, temos que:
  $SQF/\sigma^2 \sim \chi^2_{(a-1)}$, sob H_0 e $SQR/\sigma^2 \sim \chi^2_{(N-a)}$
  $N = n \times a$.

Inferência

- Assim, sob H_0

  $F_0 = \frac{SQF/(a-1)}{SQR/(N-a)} = \frac{QMF}{QMR} \sim F_{(a-1,N-a)}$

- Logo, rejeita-se H_0 se $f_0 > f_c$, em que $P(F > f_c), F \sim F_{(a-1,N-a)}$ e f_0 é valor observado de F_0 (ou de modo equivalente, pelo cáculo do pvalor).

- Pode-se provar que (exercício): $\mathcal{E}(QMF) = \sigma^2 + n\sigma_{\tau}^2, \mathcal{E}(QMR) = \sigma^2$

Inferência (cont.)

- Estimadores usuais (método dos momentos) para σ^2 e σ^2_τ são dados por:

  $\widehat{\sigma}^2 = QMR; \widehat{\sigma}^2_\tau = \frac{QMF - QMR}{n}$.

- Se o experimento for besbalanceado, devemos considerar:

  $n_0 = \frac{1}{a-1}\left[\sum_{i=1}^{a} n_i - \frac{\sum_{i=1}^{a} n_i^2}{\sum_{i=1}^{a} n_i}\right]$

- Tal modificação preserva as boas propriedades dos estimadores (ENVUM, em nosso modelo). Para modelos mais complexos tais propriedades podem não mais valer.

- Em alguns casos, pode-se obter estimativas negativas para σ^2_τ usando-se $\widehat{\sigma}^2_\tau$. Uma alternativa é a utilização e métodos Bayesianos.

Tabela de análise de variância

- Para testar a igualdade simultânea das médias

FV	SQ	GL	QM	Estatística F	pvalor
Fatores	SQF	k-1	QMF=SQF/(k-1)	$F_0 = QMF/QMR$	$\min(P(F_0 > f\|H_0), P(F_0 < f\|H_0))$
Resíduo	SQR	n-k	QMR=SQR/(n-k)		
Total	SQT	n-1			

FV: fonte de variação, SQ: soma de quadrados, Gl: graus de liberdade, QM: quadrado médio.

Estimativa dos parâmetros (quantidades)

- A construção da tabela anova não requer à obtenção de estimativas para μ nem de valores preditos para τ_j.
- A estimação dos efeitos fixos (μ) pode ser feita usando o método da máxima verossimilhança restrita.
- Com as estimativas de μ e de (σ, σ_τ^2) pode-se obter valores preditos para τ_j usando-se métodos Bayesianos empíricos, por exemplo.

Estimativa dos parâmetros (quantidades) cont.

- Veja as seguintes referências (e as referências neles contidas) (disponíveis no site do curso):
  - Boeck, Paul ; Azevedo, C. L. N. ; TAVARES, H. R. (2011). Linear and Nonlinear Generalized Mixed Models: inference and applications. São Paulo: Associação Brasileira de Estatística.
  - Singer, J.M., Nobre, J.S. e Rocha, F.M.M. (2010). Análise de dados longitudinais. (Versão preliminar parcial)

Estimativa dos parâmetros (quantidades) cont.

- Os pacores do R *nlme* e *lme4* permitem realizar inferências em modelos mistos (estimativas dos parâmetros e verificação da qualidade do ajuste do modelo).
- A análise de resíduos deve ser feita utilizando-se os **resíduos de confundimento mínimo**.
- A tabela ANOVA pode ser obtida utilizando-se a função *lm* da forma usual.

Tabela ANOVA

FV	df	SQ	QM	Estatística F	pvalor
Tear	3	89,19	29,73	15,68	0,0002
Resíduo	12	22,75	1,90		
Total	15	111,94			

Estimativas dos parâmetros

Parâmetro	Estimativa	EP	IC(95%)
μ	95,44	1,36	[92,77;98,10]
σ^2	6,96	-	[3,57;18,96]
σ_τ^2	1,90	-	-
$\rho = \frac{\sigma_\tau^2}{\sigma^2+\sigma_\tau^2}$	0,78	-	[0,38;0,98]

Pesquisar as fórmulas para os IC's no livro do Montgomery Capítulo 12.

Apesar do comprimento considerável do IC associado ao parâmetro ρ nota-se que a variabilidade entre os teares não é desprezível.