

Mais resoluções de problemas sob modelos lineares hierárquicos

Prof. Caio Azevedo

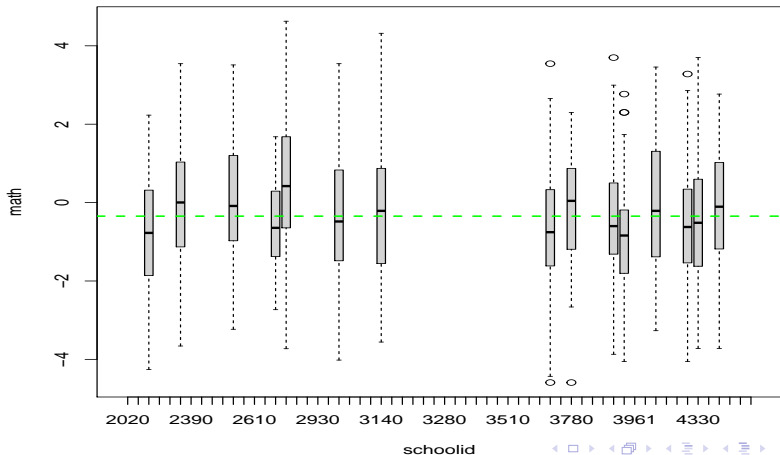
Dados “egsingle” disponíveis no pacote R `mlmRev`

- Informações sobre pontuações em matemática do “U.S. Sustaining Effects Study”. Um total de 60 escolas, 1720 alunos, acompanhados (há dados faltantes) em seis instantes com um total de 7230 observações.
- Portanto tem-se **dados longidutinais hierárquicos** (“clustered longitudinal data”).
- Temos três níveis: medidas repetidas agrupadas nos estudantes, os quais estão agrupados nas escolas.
- Utilizaremos parte do banco de dados (amostras de $J = 15$ escolas) e somente algumas variáveis, bem como desconsideraremos o nível “medida repetida”.

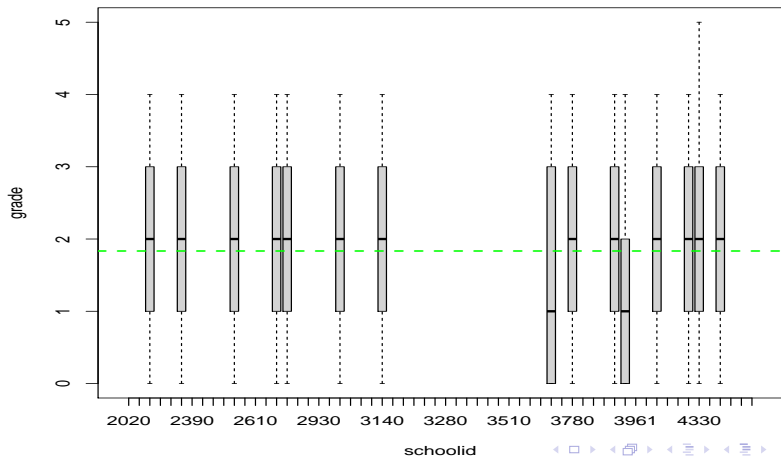
Dados “egsingle” disponíveis no pacote R `mlmRev`

- Objetivos são avaliar comportamento da variável “math” (nota via TRI com base em um teste Matemática, dos alunos) em função da variável “grade” (uma nota geral de cada aluno) ao longo das escolas (“schoolid”).
- Vamos considerar também a variável “lowinc” (porcentagem de alunos de baixa renda na escola) para compreender possíveis diferenças entre as escolas.

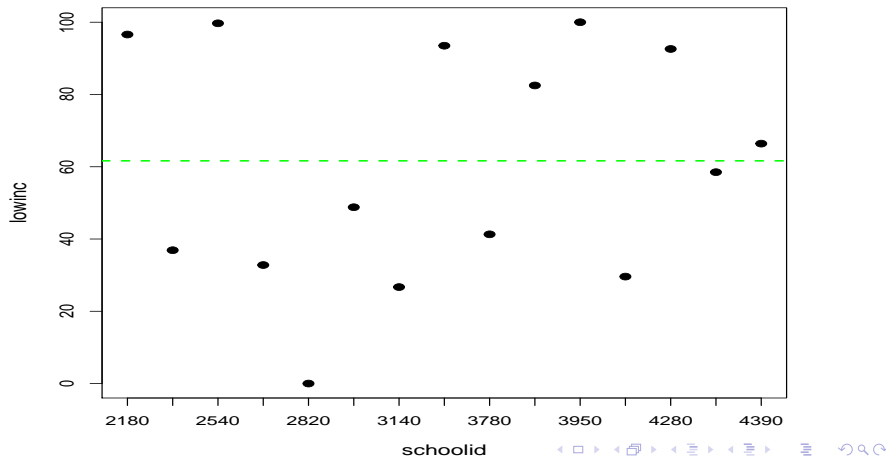
Boxplot: amth por escola



Boxplot: grade por escola



Lowinc por escola (linha verde: percentual médio)



Dispersão entre math e grade por escola

Dispersão entre math e grade

Modelo de dois níveis com covariáveis em dois níveis

■ Modelo 1

$$Y_{ji} = \beta_{0j} + \beta_{1j}(x_{ji} - 2) + \xi_{ji}, j = 1, 2, \dots, 15;$$

$$i = 1, 2, \dots, n_j \text{ (nível 1 : aluno)}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(z_j - 50) + u_{0j} \text{ (nível 2: escola)}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(z_j - 50) + u_{1j} \text{ (nível 2: escola)}$$

■ Erros e efeitos aleatórios: $\xi_{ji} \stackrel{iid}{\sim} N(0, \sigma^2)$,

$\mathbf{u}_j = (u_{0j}, u_{1j})' \stackrel{iid}{\sim} N_2(0, \Psi)$, $\xi_{ji} \perp \mathbf{u}_j, \forall i, j$ e $(\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11})'$ são

não aleatórios e $\Psi = \begin{bmatrix} \psi_{00} & \psi_{01} \\ \psi_{01} & \psi_{11} \end{bmatrix}$.

■ (momentos condicionais)

$$\begin{aligned}\mathcal{E}(Y_{ji}|\mathbf{u}_j) &= \gamma_{00} + \gamma_{01}(z_j - 50) + \gamma_{10}(x_{ji} - 2) + \gamma_{11}(z_j - 50)(x_{ji} - 2) \\ &\quad + u_{0j} + u_{1j}(x_{ji} - 2)\end{aligned}$$

$$\mathcal{V}(Y_{ji}|\mathbf{u}_j) = \sigma^2$$

■ (momentos marginais)

$$\mathcal{E}(Y_{ji}) = \gamma_{00} + \gamma_{01}(z_j - 50) + \gamma_{10}(x_{ji} - 2) + \gamma_{11}(z_j - 50)(x_{ji} - 2)$$

$$\mathcal{V}(Y_{ji}) = \psi_{00} + \psi_{11}(x_{ji} - 2)^2 + \psi_{01}(x_{ji} - 2) + \sigma^2$$

comentários

- Os pacotes “lme4” e “nlme” não apresentam uma sintaxe direta em termos dos níveis dos modelos.
- Ambos foram feitos, essencialmente, para ajuste de modelos mistos.
- Assim, na presença de três ou mais níveis e/ou covariáveis no nível 2 (ou acima) é necessário reescrever o modelo hierárquicos em termos do correspondente modelo misto.

Modelo misto correspondente

$$\begin{aligned} Y_{ji} &= \gamma_{00} + \gamma_{01}(z_j - 50) + u_{0j} + (\gamma_{10} + \gamma_{11}(z_j - 50) + u_{1j})(x_{ji} - 2) + \xi_{ji} \\ &= \gamma_{00} + \gamma_{01}(z_j - 50) + u_{0j} + \gamma_{10}(x_{ji} - 2) \\ &\quad + \gamma_{11}(z_j - 50)(x_{ji} - 2) + u_{1j}(x_{ji} - 2) + \xi_{ji} \\ &= \gamma_{00} + \gamma_{01}(z_j - 50) + \gamma_{10}(x_{ji} - 2) + \gamma_{11}(z_j - 50)(x_{ji} - 2) \\ &\quad + u_{0j} + u_{1j}(x_{ji} - 2) + \xi_{ji} \end{aligned}$$

Estimativas

Parâmetro	Estimativa	EP	IC(95%)	Estat. z	p-valor
γ_{00}	-0,1479	0,0702	[-0,2854 ; -0,0103]	-2,1067	0,0351
γ_{10}	0,8467	0,0262	[0,7952 ; 0,8981]	32,2653	<0,0001
γ_{01}	-0,0061	0,0021	[-0,0102 ; -0,0020]	-2,9011	0,0037
γ_{11}	-0,0018	0,0008	[-0,0033 ; -0,0003]	-2,3038	0,0212

ψ_{00}	ψ_{11}	ψ_{01}	ρ_{01}	σ^2
0,05606	0,00439	0,00880	0,56083	0,96902

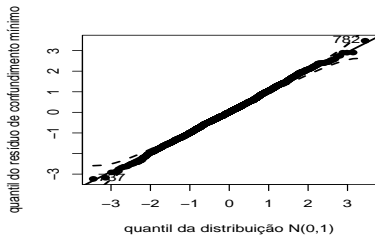
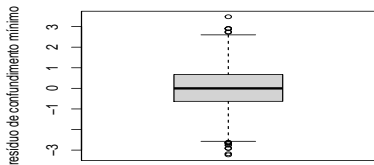
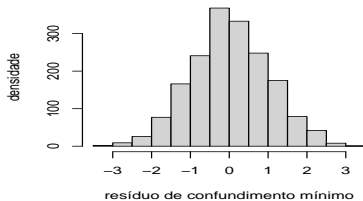
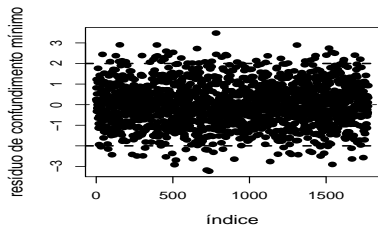
Comentários/observações

- Todos os parâmetros de regressão (γ) parecem ser significativos.
- A variabilidade, dentro de cada escola da variável math é constante entre os alunos, enquanto que ao longo das escolas ela depende da variável grade.
- As covariáveis grade (nível 1) e lowinc (nível 2), apresentaram efeitos (marginalmente) significativo.

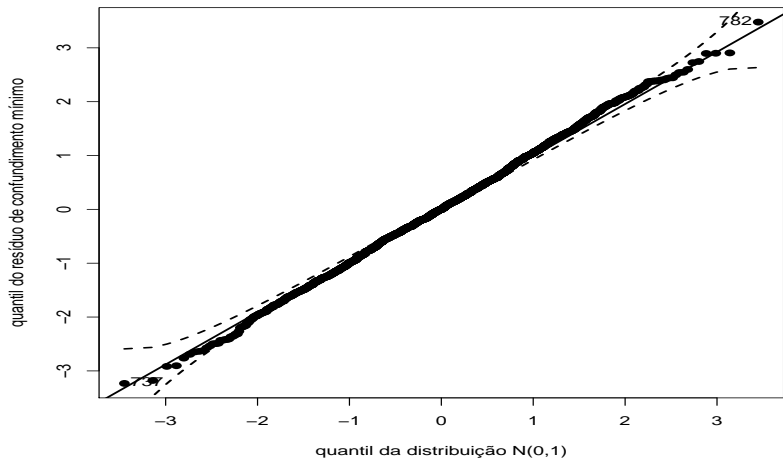
Comentários/observações

- Algumas interpretações:
 - Alunos (independentemente da escola em particular) de escolas com $\text{lowinc} = 0,50$, tem math esperado entre $[-0,2854;-0,0103]$.
 - Alunos (independentemente da escola em particular) de escolas com $\text{lowinc} = 0,50$, em um incremento esperado de $[0,7952;0,8981]$ no math para o aumento de uma unidade no grade.
 - O valor esperado (acima) parece variar (significativamente) entre as escolas com $\text{lowinc} = 0,5$ ($\tilde{\psi}_{00} = 0,05606$)
 - O incremento esperado (acima) parece variar (modicamente) entre as escolas com $\text{lowinc} = 0,5$ ($\tilde{\psi}_{11} = 0,00439$).

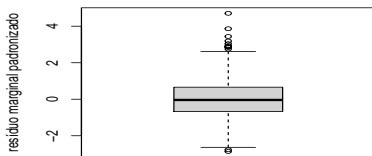
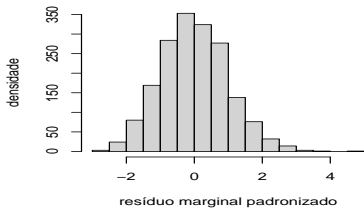
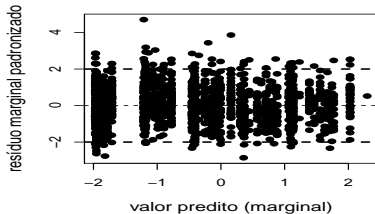
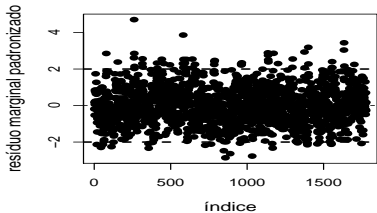
RCM: gráficos de diagnóstico



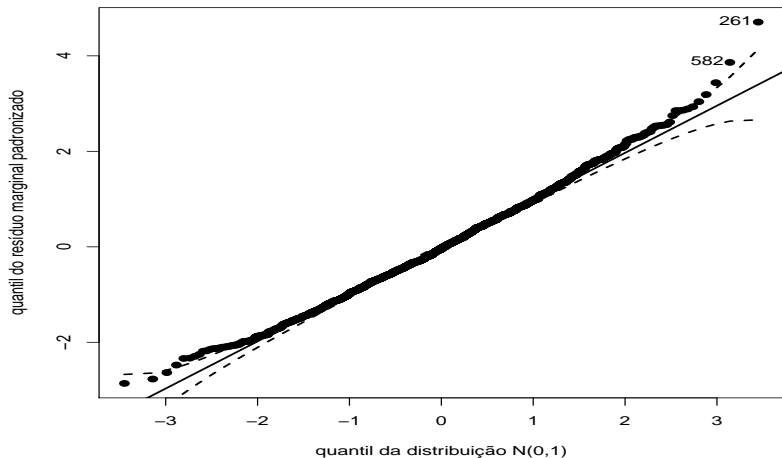
RCM: QQplot com envelope



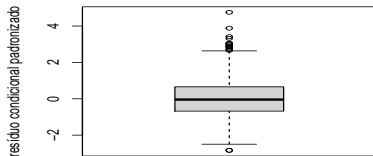
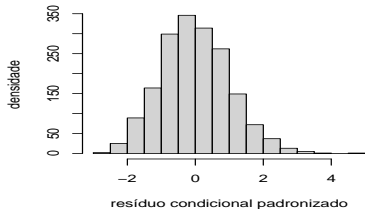
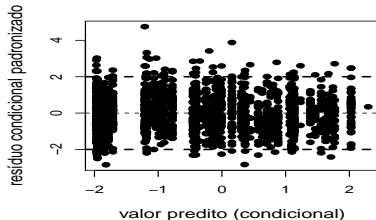
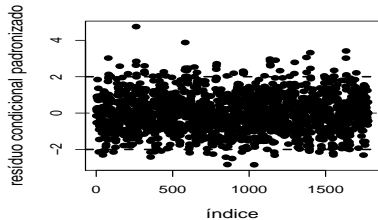
RMP: gráficos de diagnóstico



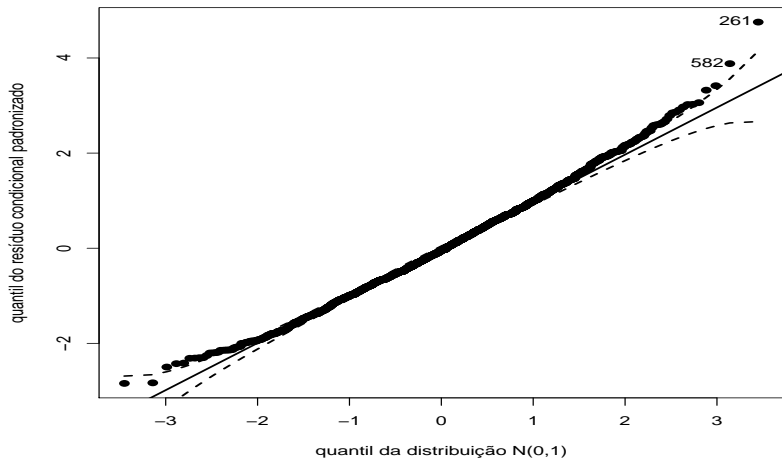
RMP: QQplot com envelope



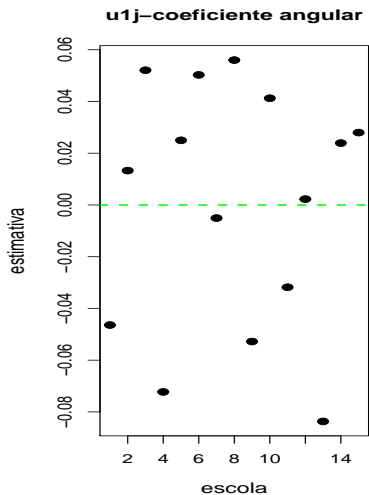
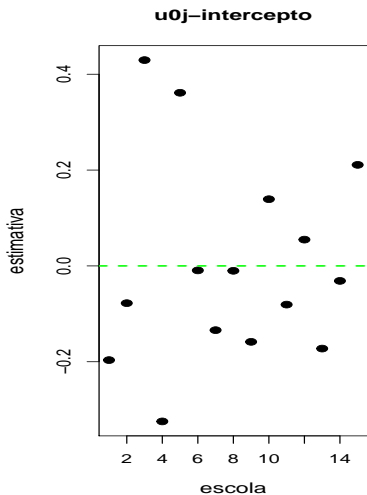
RCP: gráficos de diagnóstico



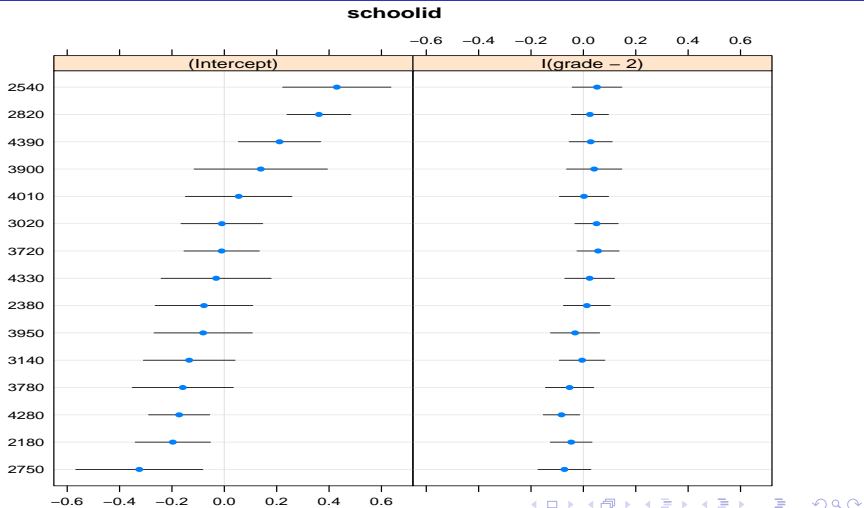
RMP: QQplot com envelopes



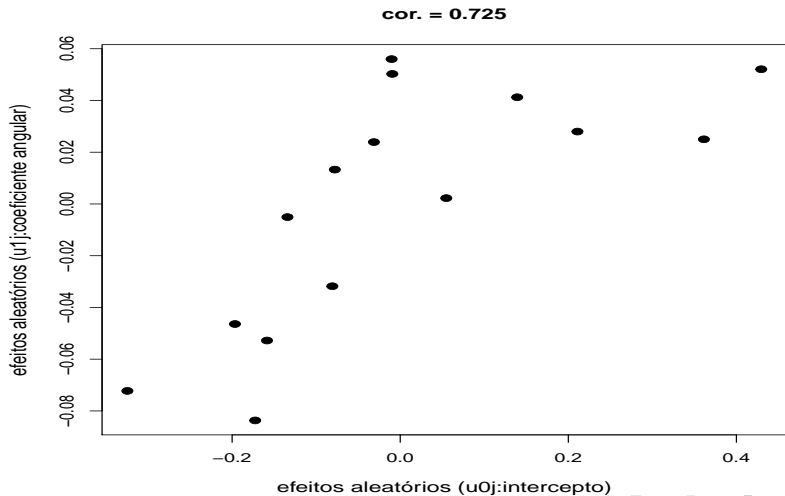
Estimativas pontuais dos efeitos aleatórios



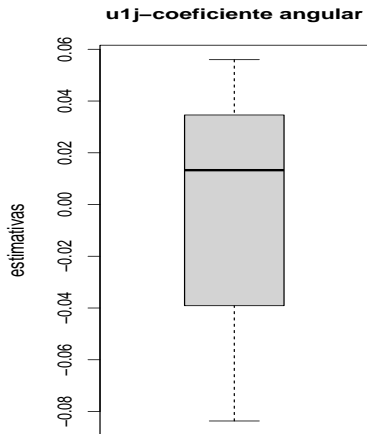
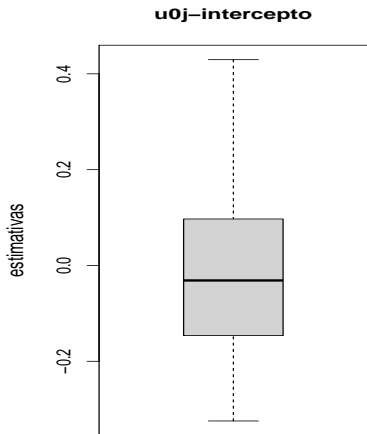
Estimativas pontuais e intervalares dos efeitos aleatórios



Dispersão entre as estimativas dos efeitos aleatórios

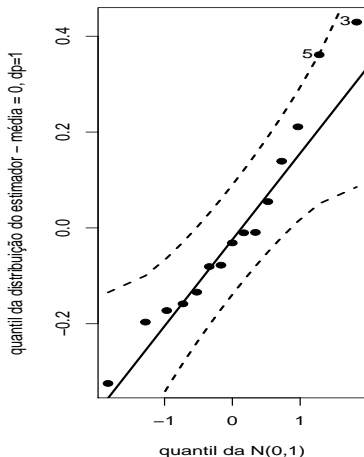


Box-plots das estimativas dos efeitos aleatórios

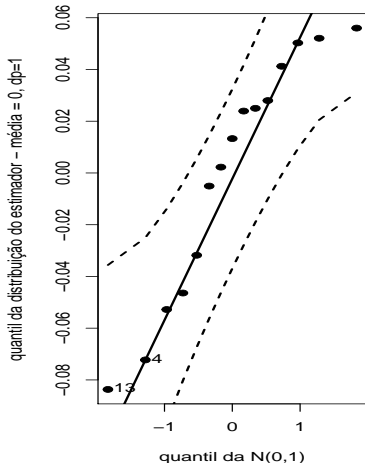


QQplot com envelopes das estimativas dos EA's

u0j-intercepto, KS : 0.6215



u1j-coef. angular, KS : 0.7648



Comentários

- O RCM não apresenta maiores problemas (eventualmente considerar uma distribuição de caudas pesadas, com a t de Student, normal potência etc).
- O RCP e o RMP apresentam assimetria positiva e heterocedasticidade.
- Pelos gráficos dos valores preditos dos efeitos aleatórios, e pelos resultados acima, aparentemente, há ausência de normalidade nos efeitos aleatórios.

Comentários

- Contudo, aparentemente, os resultados indicam que os interceptos e os coeficientes angulares variam ao longo das escolas.
- Sugestões: utilizar distribuição de caudas pesadas para os erros condicionais (ξ) ou a distribuição dos efeitos aleatórios (\mathbf{u}) (alguma distribuição assimétrica).