

MAT02035 - Modelos para dados correlacionados

Modelos multiníveis: exemplo

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

Porto Alegre, 2022

TVSFP

- ▶ Embora a prevalência do tabagismo tenha diminuído entre os adultos nas últimas décadas, um número substancial de jovens começa a fumar e se torna viciado em tabaco.
- ▶ O *Television, School and Family Smoking Prevention and Cessation Project* (TVSFP) foi um estudo concebido para determinar a eficácia de um currículo escolar de prevenção do tabagismo em conjunto com um programa de prevenção baseado na televisão, em termos de prevenção do início do tabagismo e aumento da cessação do tabagismo.
- ▶ O estudo utilizou um desenho fatorial 2×2 , com quatro condições de intervenção determinadas pela classificação cruzada de um currículo escolar de resistência social (CC: codificado 1 = sim, 0 = não) com um programa de prevenção baseado em televisão (TV : codificado 1 = sim, 0 = não).
- ▶ A randomização para uma das quatro condições de intervenção ocorreu em nível escolar, enquanto grande parte da intervenção foi realizada em nível de sala de aula.

TVSFP

- ▶ O estudo original envolveu 6.695 alunos em 47 escolas no sul da Califórnia.
- ▶ Nossa análise se concentra em um subconjunto de 1.600 alunos da sétima série de 135 turmas em 28 escolas de Los Angeles.
- ▶ A variável resposta, uma escala de conhecimento sobre tabaco e saúde (THKS), foi administrada antes e após a randomização das escolas para uma das quatro condições de intervenção.
 - ▶ A escala avaliou o conhecimento de um estudante sobre tabaco e saúde.

TVSFP

- ▶ Consideramos um modelo linear para o escore THKS pós-intervenção, com o escore THKS basal ou pré-intervenção como covariável.
- ▶ Esse modelo para a mudança ajustada nos escores do THKS incluiu os principais efeitos de CC e TV e a interação $CC \times TV$.
- ▶ Os **efeitos da escola e da sala de aula foram modelados pela incorporação de efeitos aleatórios nos níveis 3 e 2**, respectivamente.

TVSFP

- ▶ Deixando Y_{ijk} denotar a pontuação THKS pós-intervenção do i -ésimo aluno dentro da j -ésima sala de aula dentro da k -ésima escola, nosso modelo é dado por

$$Y_{ijk} = \beta_1 + \beta_2 \text{Pre-THKS} + \beta_3 \text{CC} + \beta_4 \text{TV} + \beta_5 \text{CC} \times \text{TV} + b_k^{(3)} + b_{jk}^{(2)} + \epsilon_{ijk},$$

em que $\epsilon_{ijk} \sim N(0, \sigma_1^2)$, $b_{jk}^{(2)} \sim N(0, \sigma_2^2)$ e $b_k^{(3)} \sim N(0, \sigma_3^2)$.

TVSFP

```
mod1 <- lme(  
  fixed = postscore ~ prescore + curriculum*tvprevent,  
    random = ~1 | sid/cid,  
    data = thks,  
    na.action = na.omit)
```

TVSFP

- As estimativas REML dos efeitos fixos são exibidas na tabela a seguir.

	E	estimativa	EP	Z
(Intercept)	1.702	0.1254	13.57	
prescore	0.305	0.0259	11.79	
curriculum1	0.641	0.1609	3.98	
tvprevent1	0.182	0.1572	1.16	
curriculum1:tvprevent1	-0.331	0.2246	-1.47	

TVSFP

- ▶ As estimativas REML das três fontes de variabilidade indicam que há variabilidade tanto nos níveis de sala de aula quanto na escola, com quase duas vezes mais variabilidade entre as salas de aula dentro de uma escola do que entre as escolas.

```
VarCorr(mod1)
```

##	Variance	StdDev
## sid =	pdLogChol(1)	
## (Intercept)	0.0386	0.197
## cid =	pdLogChol(1)	
## (Intercept)	0.0647	0.254
## Residual	1.6023	1.266

TVSFP

- ▶ A correlação entre as pontuações do THKS para colegas (ou crianças da mesma sala de aula na mesma escola) é de aproximadamente **0.061** (ou $\frac{0.039+0.065}{0.039+0.065+1.602}$), enquanto a correlação entre as pontuações do THKS para crianças de diferentes salas de aula dentro da mesma escola é de aproximadamente **0.023** (ou $\frac{0.039}{0.039+0.065+1.602}$).

TVSFP

- ▶ As estimativas dos efeitos fixos para as condições de intervenção, quando comparadas com seus erros padrão, indicam que:
 - ▶ **nem a intervenção na mídia de massa (TV) nem sua interação com o currículo de sala de aula de resistência social (CC) têm impacto nas mudanças ajustadas pelas pontuações THKS da linha de base.**
- ▶ Há um efeito significativo do currículo de sala de aula de resistência social, com crianças designadas ao currículo de resistência social mostrando maior conhecimento sobre tabaco e saúde.

Breve conclusão

Políticas educacionais abrangentes são mais eficazes para promover o conhecimento sobre o assunto do que campanhas de mídia de massa, que dependem fortemente da motivação individual.

TVSFP

- ▶ A estimativa do efeito principal do CC , no modelo que exclui a interação $CC \times TV$, é de 0.47 ($EP = 0.113$, $p < 0.001$).

```
mod2 <- lme(
  fixed = postscore ~ prescore + curriculum + tvprevent,
  random = ~1 | sid/cid,
  data = thks,
  na.action = na.omit)
```

	Estimativa	EP	Z
(Intercept)	1.7849	0.1129	15.80
prescore	0.3052	0.0259	11.79
curriculum1	0.4715	0.1133	4.16
tvprevent1	0.0196	0.1133	0.17

TVSFP

- ▶ As correlações intra-cluster nos níveis de escola e sala de aula são relativamente pequenas.
 - ▶ Podemos ficar tentados a considerar isso como uma indicação de que o agrupamento desses dados é irrelevante.
 - ▶ **No entanto, tal conclusão seria errônea.**
- ▶ Embora as correlações intra-cluster sejam relativamente pequenas, elas têm um impacto substancial nas inferências sobre os efeitos das condições de intervenção.

TVSFP

- Para ilustrar isso, considere o seguinte modelo para as mudanças ajustadas nas pontuações do THKS:

$$Y_{ijk} = \beta_1 + \beta_2 \text{Pre-THKS} + \beta_3 \text{CC} + \beta_4 \text{TV} + \beta_5 \text{CC} \times \text{TV} + \epsilon_{ijk},$$

em que $\epsilon_{ijk} \sim N(0, \sigma^2)$.

- Este modelo ignora o agrupamento dos dados nos níveis de sala de aula e escola; é um modelo de **regressão linear padrão** e **assume observações independentes** e **variância homogênea**.

TVSFP

```
mod3 <- lm(
  formula = postscore ~ prescore + curriculum*tvprevent,
  data = thks)
```

	Estimativa	EP	Z
(Intercept)	1.661	0.0844	19.69
prescore	0.325	0.0259	12.58
curriculum1	0.641	0.0921	6.95
tvprevent1	0.199	0.0900	2.21
curriculum1:tvprevent1	-0.322	0.1303	-2.47

TVSFP

- ▶ As estimativas dos efeitos fixos são semelhantes às relatadas no Modelo 1.
- ▶ No entanto, os erros padrão baseados no modelo (supondo que não haja agrupamento) são **enganosamente pequenos** para os efeitos da intervenção randomizada e **levam a conclusões substancialmente diferentes** sobre os efeitos das condições de intervenção.

TVSFP

- ▶ **Isso destaca uma lição importante:** o impacto do agrupamento depende tanto da magnitude da correlação intra-cluster quanto do tamanho do cluster.
 - ▶ Para os dados do TVSFP, os tamanhos dos clusters variam de 1 a 13 salas de aula dentro de uma escola e de 2 a 28 alunos dentro de uma sala de aula.
 - ▶ Com tamanhos de cluster relativamente grandes, mesmo uma correlação intra-cluster muito modesta pode ter um impacto discernível nas inferências.

Avisos

- ▶ **Fim das aulas síncronas!**
- ▶ **Para casa:** ler o Capítulo 22 do livro “**Applied Longitudinal Analysis**”.
 - ▶ Caso ainda não tenha lido, leia também os Caps. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 e 13.

Bons estudos!

