

PERGUNTAS REALIZADAS DURANTE A AULA E RESPONDIDAS POR E-MAIL

Curso: MBA em Data Science e Analytics

Disciplina: Supervised Machine Learning: Modelagem Multinível II

Data: 05/10/2021

Ricardo José Pfitscher

Professor, se eu quiser prever o desempenho de um estudante não específico? No exemplo estamos apontando o aluno 1, se fosse um aluno qualquer?

Resposta: Boa pergunta. É só alterarmos os dados das variáveis preditoras deste estudante qualquer na função predict, e o resultado do desempenho previsto para este estudante será obtido.

Danilo Steckelberg

Por que os modelos HLM são considerados lineares se eles consideram interações entre variáveis preditoras (tempo_experiência * horas_estudo)? Eles não deveria entrar na categoria de mod. não lineares?

Resposta: Porque a especificação, para estes modelos específicos, apresenta forma funcional linear, e não quadrática, cúbica, etc. É importante lembrar que os modelos HLM fazem parte do grupo de modelos denominados de GLMM, que inclui os modelos não lineares, analogamente aos modelos GLM.

Ellen Nogueira

Por que são chamados efeitos aleatórios? Poderiam ser em virtude da correlação/multicolinearidade entre as variáveis? Poderiam ser em virtude de variáveis explicativas não consideradas?

Resposta: Ótima pergunta. Exatamente isso. Porque capturam as aleatoriedades decorrentes de comportamentos diferentes nos interceptos e nas inclinações entre os grupos distintos.

Victor Hugo Camargo do Nascimento

Em caso de amplo banco de dados, o correto seria ir criando o máximo de ninhos possível e formar um só modelo? E a partir desse modelo gerar outros modelos?

Resposta: Ótima pergunta. Sim, é um procedimento bastante recomendável. A maior restrição, neste caso, será a capacidade computacional.

Eduardo Garbin

No HLM3 com medidas repetidas, posso ter a inflação de algum dos valores? Posso tratar como no zero inflated?

Resposta: Ótima pergunta. Sem dúvidas. Os modelos multinível zero-inflated são extremamente recentes na literatura. Deixei um código adicional com estimação de modelos inflacionados de zeros com perspectiva multinível.

Alex De Lima Bassi

Professor, para seguir o modelo no R não seria interessante começar como nível zero para o tempo

Resposta: Seguindo a estratégia step-up, iniciamos com o modelo nulo, em que a variável referente ao tempo não entra como preditora.

Wagner Seizo Hokama

como foi calculado o V_{0j} e V_{1j} do exercício das 20h50?

Resposta: Ótima pergunta. Se fossem estimados os seguintes modelos por OLS do tipo $\beta_{0j} = \gamma_{00} + \gamma_{01} \cdot \text{temp}_j + v_{0j}$, os valores de v_{0j} seriam os termos de erro desta regressão para cada grupo, ou contexto. O mesmo vale para a equação das inclinações aleatórias para a obtenção de v_{1j} . É exatamente isso

que a modelagem multinível faz, por meio da estimação por máxima verossimilhança no conceito restrito (REML).

Eduardo Garbin

Se um efeito aleatorio não for significativo, ele fica fora?

Resposta: Sim, este é o procedimento recomendável quando da aplicação da estratégia step-up.

Vitor Bruno da Silveira Guimarães

Se adicionasse nesse exemplo uma variável contendo a cidade, a equação ainda sim continua a mesma? Depois do Pipe colocaria | Escola + Cidade?

Resposta: Boa pergunta. Neste caso, que seria um HLM3, o comando para o termo aleatório seria algo como: `random = list(cidade = ~1, escola = ~1)`

Matheus Garcia

Professor, por que no modelo final, no R, o efeito aleatório não está, também, em função do `texp`?

Resposta: Boa pergunta. Pela própria especificação algébrica do modelo (slide 14 da apostila em ppt). Em outras palavras, no componente de efeitos aleatórios entram variáveis preditoras de nível inferior multiplicadas pelos efeitos aleatórios de inclinação de nível superior. E a variável `texp` é uma preditora de nível superior no nosso exemplo.

Bruno Lopes Da Silva

Quando utilizo os termos de erro para compor a equação da regressão utilizo a média dos erros?

Resposta: A média dos efeitos aleatórios é utilizada em um modelo GLM. NO GLMM utilizam-se os próprios efeitos aleatórios de cada contexto, individualmente.

Alex De Lima Bassi

Professor teria como vc apresentar uma modelagem multinível com outputs de cluster

Resposta: Sugerimos que fique como exercício adicional, já que comentamos em aula, porém não mostramos por meio de código.

Leonardo Destro Bronzato

Quais vantagens obtemos ao rodar o GLMM ao inves de rodarmos um OLS para cada "grupo"/nível?

Resposta: Boa pergunta. A parcimônia é a maior das vantagens, além da consideração de interações no componente de efeitos aleatórios.

Leonardo Destro Bronzato

Oque seria no exemplo da escola este efeito aleatorio (V_{0j}) ?

Resposta: Os interceptos individuais de cada escola.

Leonardo Destro Bronzato

Eu poderia considerar que " $B_{0j} = Y_{00} + Y_{01} \cdot \text{texp}_j + v_{0j}$ " seria uma regressão linear (OLS) para o intercepto do nível 1 ? se sim, O V_{0j} (efeito aleatorio) seria o erro dessa regressão ?

Resposta: Excelente pergunta. É exatamente isso. Inclusive se você rodar esta regressão, serão obtidos exatamente os valores de v_{0j} . O mesmo vale para a equação das inclinações aleatórias para a obtenção de v_{1j} .