

## 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

**Carlos Henrique de Oliveira**

Professor, existe uma lógica para entendermos quando devemos correlacionar duas variáveis? Por exemplo:  $\text{texp}:\text{mes}$ .

Resposta: Carlos, essa definição passa pela literatura, pela experiência gerencial, pela expertise do pesquisador. É mais uma decisão contextual do que matemática, nesse caso.

**Priscila Morgante Bohlsen**

não é a escola 1?

Resposta: Priscila, não entendemos sua pergunta. Caso queira, acione a monitoria no e-mail que iremos prontamente lhe auxiliar.

**Priscila Morgante Bohlsen**

a escola 8 tem  $-0.8$ , e a escola 1 tem  $-0.11$

Resposta: Priscila, não entendemos sua pergunta. Caso queira, acione a monitoria no e-mail que iremos prontamente lhe auxiliar.

**Thiago Colette Vegi**

Qual a vantagem de incluir a variável mês multiplicada por atividades e  $\text{texp}$  sendo que todas elas já estão no modelo? Essa inclusão não gera multicolinearidade?

Resposta: Thiago, a vantagem é a análise da interação dessas duas variáveis. Isoladamente elas estão no modelo original, o que estamos fazendo é incluir o efeito da interação destas variáveis. A multicolinearidade é sim uma preocupação neste caso, para termos certeza devemos sempre efetuar os testes segundo o que nos reza a boa prática em estatística.

**Benicio Antonio Cardoso**

Inferir desempenho entre universos diferentes sem trazer a avaliação ao mesmo nível para todos os universos, isso não pode gerar dados enviesados?

Resposta: Não consigo enxergar possibilidade de enviesamento dos dados, Benício. Estamos fazendo constar no modelo esta questão contextual presente no mundo real. Tudo isso está matematicamente adequado e já plenamente incluso no modelo proposto, visto que são traçadas "retas" para cada contexto. Estamos analisando universos distintos enquanto universos distintos, e estimando essas diferenças.

**Jean Carlos Zambrano Contreras**

Poderia recomendar um paper que tenha utilizado modelagem multinível?

Resposta: Jean, as leituras complementares trazem alguns exemplos. Caso ainda não seja suficiente, peço que acione a monitoria pelo e-mail que será prontamente atendido.

**Eduardo Garbin**

Tenho que colocar os betas na ordem que aparecem os níveis?

Resposta: Eduardo, se entendi bem sua pergunta, a resposta é não. As variáveis explicativas podem vir em qualquer ordem, o que não se pode confundir são as questões de nível.

**Thiago Colette Vegi**

Existe uma forma de aplicar um "modelo de inflação de 10?"

Resposta: Thiago, não entendemos sua pergunta. Caso queira, acione a monitoria no e-mail que iremos prontamente lhe auxiliar.

**Robert Fritz**

Mês não seria qualitativo?

Resposta: Robert, se entendi sua pergunta, mês é um indicativo temporal. Mostra quantos meses se passaram, portanto, uma variável quantitativa. Não é o mês em específico, mas sim uma quantidade de meses específica.

**Sandro Ramos**

É possível estimar a chance de um evento binário ou multinomial, feito por modelos logísticos, através de modelos multiníveis? Em caso positivo teria algum material que poderia adicionar?

Resposta: Sandro, é possível utilizar a regressão logística numa perspectiva multinível. Peço que volte na aula I de modelos multiníveis, em que o professor Fávero explica melhor esta questão. Há poucos trabalhos publicados nesta perspectiva, mas é possível sim.

**Miriam Miyashiro**

Qual seria uma configuração de computador capaz de rodar um modelo multinível com 1 milhão, 10 milhões, 30 milhões de observações, num tempo razoável? (poderia dar exemplos?)

Resposta: Miriam, impossível dizer. Depende de muitas questões, inclusive das características da base. O que consigo te dizer é que para alguns milhões de observações, precisaríamos, provavelmente, de um poder computacional gigantesco.

**Isadora Salvador Rocco**

Quando é indicado usar uma interação num modelo OLS ou logístico, exemplo, horas\* escola? No GMM ele já faz a interação automático?

Resposta: Isadora, essa definição passa pela literatura, pela experiência gerencial, pela expertise do pesquisador. É mais uma decisão contextual do que matemática, nesse caso.

**Ricardo José Pfitscher**

De acordo com a Expressão HLM2 completa, se o aluno estudar 0 horas e o professor tiver 0 de experiência, então o desempenho será negativo? Isso faz sentido?

Resposta: Ricardo, creio que esse exemplo hipotético está fora do raio em que o algoritmo foi treinado. Me parece uma extrapolação. É mais uma questão matemática que estatística.

**Rafael Marin Machado de Souza**

Professor, redes neurais MLP possuem muitas das características dos GLM apresentados, o Bias como Alpha e Pesos Sinápticos ( $w$ ) como Betas. Essa relação existe ou é uma comparação arbitrária?

Resposta: Rafael, não entendemos sua pergunta. Caso queira, acione a monitoria no e-mail que iremos prontamente lhe auxiliar.

**Juliano Sartorelli Dias**

Boa noite. No caso do gráfico da linha 174 do código, não seria interessante fazer um boxplot?

Resposta: Juliano, é possível. Seria útil para verificarmos as distribuições por escola (nas variáveis quanti), seria um bom complemento para a análise.

**Paulo Renato Leite**

Professor, posso ter perdido, mas em que momento foi estimado o desempenho?

Resposta: Paulo, o desempenho dos alunos está dado no exercício. É a nossa variável dependente, representa a nota de 0 a 100 pontos dos estudantes.

**Michelle da Costa Santana**

Professor, sobre monitoramento de modelos, quais seriam os principais indicadores que nos ajudariam a acompanhar a "saúde" do modelo?

Resposta: Michelle, o indicador para se comparar modelos e encontrar o melhor ajuste costuma ser o LL.

Hernandes Matias Junior 124.799.116-40