

## Introduction to Clustered Data Analysis with Generalized Mixed-Effect Models: theory, application and a case study

Thiago Cortez Costa<sup>1</sup>

[thiago.cortez@uol.com.br](mailto:thiago.cortez@uol.com.br)

### Modelos Hierárquicos Generalizados: teoria e aplicação para análise de determinantes dos resultados eleitorais

Contexto eleitoral ou atributos dos candidatos? O uso de Modelos Hierárquicos Generalizados para compreender as eleições

**Área temática:** técnicas estatísticas e análise de resultados eleitorais

**Resumo:** Modelos de regressão convencionais têm entre seus pressupostos que a variável resposta é numérica e normalmente distribuída; e que as unidades pesquisadas são independentes entre si. Contudo, para grande parte dos problemas de pesquisa em investigações sociais esse não é o caso. Os modelos generalizados de efeito misto, também chamados de modelos hierárquicos generalizados, constituem uma solução para essas limitações. técnica mais apropriada para o estudo de casos que se acredita que os atributos de uma determinada unidade observada são influenciados pelo conglomerado ou agrupamento a que ela pertence. O presente trabalho pretende difundir o uso da modelagem hierárquica como ferramenta de análise de dados organizados em níveis ou agrupamentos quando a variável resposta não é variável numérica contínua (isto é, não possui uma distribuição normal). Para tanto, apresenta a teoria estatística subjacente à técnica, acompanhada de um estudo de caso em que a técnica é aplicada para explicação dos resultados eleitorais de 2006.

#### Contexto geral e problema

Uma grande parte dos fenômenos estudados por cientistas sociais, naturais ou da saúde apresentam dados estruturados em agrupamentos. Em certos casos, a estrutura de agrupamento é simples, e cada unidade de observação pertence a um único grupo, havendo apenas um fator de agrupamento. Outros casos mais complexos possuem uma estrutura de níveis hierarquizados (aninhados e sobrepostos uns aos outros).

Por exemplo, um estudo longitudinal observa uma determinada característica em um conjunto de indivíduos, fazendo medições repetidas ao longo do tempo em ocasiões não-idênticas para todos os indivíduos. Essas observações estão agrupadas por indivíduo – um caso simples. Para complicar, no entanto, cada indivíduo pode estar agrupado dentro de outra unidade (ou outro nível), como um domicílio, uma família, uma escola, ou um local de trabalho. Tais unidades, por sua vez, poderão estar agrupadas dentro de localidades geográficas, como bairros, cidades, estados e países. Para tomar um exemplo nas ciências naturais, imaginemos um estudo que acompanhe o crescimento de uma dada espécie de planta, com espécimes agrupados por tipo de solo em que se encontram, e os tipos de solo agrupados por região climática. E em um exemplo de administração de empresas, as vendas de uma empresa podem estar agrupadas por vendedores, e os vendedores por lojas em que trabalham.

O ponto comum a todos esses casos é que as observações feitas não são independentes entre si. Há uma correlação entre as observações feitas dentro do mesmo grupo. Ocorrências de uma e de

<sup>1</sup> Doutorando em Ciência Política pela Universidade de Brasília

Mestre em Estudos Populacionais e Pesquisas Sociais pela Escola Nacional de Ciências Estatísticas (ENCE/IBGE)  
Analista Legislativo e Pesquisador do Instituto DataSenado, do Senado Federal