

# Estatística Multivariada sobre Dados Categóricos: Contribuições da Regressão Logística à Análise de Correspondência Múltipla

Aluno: Guilherme Guilhermino Neto

Orientador: Prof. Dr. Ronaldo Rocha Bastos

## RESUMO

Em diversas áreas do conhecimento, verifica-se a existência de bancos de dados compostos por variáveis que não são numéricas, mas categóricas. Suas escalas podem se referir a opiniões, comportamentos, sucesso / insucesso, entre outros.

Dentre as técnicas de análise estatística aplicáveis a dados categóricos, tem sido cada vez mais difundidas a análise de correspondência múltipla (ACM), que, análoga à análise de componentes principais, objetiva descrever geometricamente tabelas de contingência multivariadas, e a regressão logística, estratégia de modelagem destinada a criar classificadores baseados na relação entre covariáveis (categóricas ou não) e uma variável independente que denota um atributo qualquer descrito por variável categórica.

Este trabalho se vale de dados de clientes de uma seguradora de automóveis belga (GREENACRE et al., 2007) para abordar o uso destas duas técnicas em conjunto, a fim de discutir os benefícios desta complementariedade.

É realizada uma primeira ACM e ajustado um modelo de regressão logística. Em seguida, nova ACM é feita, desta vez incluindo na solução somente as variáveis mais significativas segundo o modelo logístico, cujas associações foram interpretadas na solução de ACM. Observa-se um aumento na variância explicada pela ACM, destacando o poder das análises exploratórias e confirmatória quando estas caminham lado a lado.

*Palavras-chave: Análise de correspondência. Regressão Logística. Data Mining.*