

Análise de Modelos de Regressão Binária com Eventos Raros

João Vieira¹

Orientadora: Prof^a Dr^a Esmeralda Ramalho²

Resumo

Nesta dissertação serão abordadas duas estratégias para lidar com eventos raros em variáveis dependentes binárias: a seleção amostral com base na variável dependente e a aplicação de modelos flexíveis. Enquanto que na primeira estratégia a proporção de eventos raros é inflacionada artificialmente na amostra usada para estimação, na segunda estratégia utilizam-se modelos cuja capacidade de descrição dos dados é potencialmente superior aos modelos binários standard (tipicamente o *logit* e o *probit*). Para estudar qual das opções será melhor empiricamente foi realizado um estudo de simulação avaliando vários aspetos destes modelos, como a capacidade preditiva e a dificuldade computacional na sua implementação. Neste estudo observou-se que a estimação dos modelos flexíveis era mais instável, apresentando mais dificuldade na sua implementação. Por outro lado, quando se realiza seleção amostral é necessário ter uma boa dimensão amostral para se notarem os efeitos desta seleção na probabilidade de uma observação pertencer ao conjunto dos uns. Tendo em conta os resultados do estudo de simulação, é recomendado utilizar a seleção amostral com correção, para ter em conta a alteração da probabilidade dos eventos. Recomenda-se, adicionalmente, que a amostra utilizada para a estimação contenha entre 20% a 35% de uns para não se perder informação que possa estar contida no zeros, caso a amostra tenha uma dimensão pequena/média, ou 50% caso se tenha uma amostra com uma grande dimensão.

Palavras-Chave: *Choise-Based Sampling*, Eventos Raros, Funções *Link*, *Links* Assimétricos, *Logit*, *Probit*, Seleção Amostral, Simulação.

¹ Contacto: joaovieira@aln.iseg.ulisboa.pt

² [ISEG Lisbon School of Economics & Management-Departamento de Matemática e REM](http://www.iseg.ulisboa.pt). Contacto: eramalho@iseg.ulisboa.pt