

DISTRIBUIÇÃO POISSON-LINDLEY VERSUS MODELOS ZIP E ZINB PARA O AJUSTE DE DADOS DE CONTAGEM INFLACIONADOS DE ZEROS

Caroline Gomes Duarte, Silvia Maria de Freitas

Dados de contagem são frequentes em muitas áreas do conhecimento científico, e são denotados como dados discretos, pois são expressos em termos de contagem associados a uma característica de interesse. Por não ser possível associar uma distribuição normal a estes, a distribuição de Poisson associada a modelos de regressão, é a comumente utilizada como um procedimento inicial para a modelagem de dados de contagem. Entretanto, essa modelagem assume um modelo probabilístico bastante restrito, pois o parâmetro de média é igual à sua variância, e na prática, é comum que dados de contagem apresentem muitos valores nulos ou “zeros”, levando a uma variabilidade superior à variância nominal estipulada pela Poisson, caracterizada como superdispersão dos dados. Assim, surge a importância de usar alternativas da distribuição de Poisson de forma a incorporar a superdispersão e obter uma melhoria no ajuste através de um modelo mais adequado. Uma das formas é a utilização dos modelos para dados inflacionados de zeros, conhecidos como ZIP e ZINB, ou modelos em dois estágios, através dos modelos em dois estágios, com o uso das distribuições compostas. O objetivo principal deste trabalho é apresentar os modelos ZIP e ZINB já utilizados na literatura como alternativas para dados com excesso de zeros e introduzir a distribuição composta Poisson-Lindley Generalizada, visando estudar a performance de cada abordagem e o impacto dessas modelagens na qualidade dos ajustes, fazendo uso de conjuntos de dados reais. O trabalho foi desenvolvido com o apoio computacional do software R.

Palavras-chave: Superdispersão. Poisson-Lindley. ZIP. ZINB.