

Proposta: Modelos de Poisson inflacionados em zero

ME714 | ANÁLISE DE DADOS DISCRETOS
Profa. Dra. Hildete Prisco Pinheiro

Caroline da Silva Mangile 195539
Gabriela Inocente Yogi 141812
Rodrigo Resende Soares Rocha 186819
Wesley R. da Silva Satelis 188650

29 de abril de 2021

No trabalho proposto aqui, serão estudados modelos de regressão em que a variável resposta é de contagem, mais especificamente de distribuição Poisson com excesso de zeros.

A distribuição de Poisson é uma distribuição discreta de contagem das ocorrências de um evento em um determinado intervalo de tempo. Uma variável aleatória X segue uma distribuição de Poisson se sua função densidade de probabilidade é dada por

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!},$$

em que $\lambda > 0$, representa a taxa de ocorrência por unidade média e é também a média e a variância da distribuição.

Um modelo de regressão que tem a quantidade de ocorrências de um evento como variável resposta, é chamado de regressão de Poisson. Uma das suposições deste modelo é que a média e variância são iguais, que é violada quando o conjunto de dados possui muitas contagens iguais a zero. Assim, é necessário ajustar um modelo que leve em consideração essa quantidade de zeros na resposta.

Em casos como este, uma solução é utilizar Modelos de Poisson Inflacionados em Zero (*Zero-inflated Poisson* - ZIP), veja Lambert (1992). Em um modelo ZIP, a variável resposta assume uma distribuição Poisson com uma componente degenerativa, com ponto de massa entre 0 e 1. O modelo assume que existem dois tipos de respostas zero, aquelas que estão em excesso e as que não estão. Os zeros em excesso são classificados como estruturais, que apresentam a falta de uma característica não presente na população. Os demais zeros são denominados amostrais, que indicam a ausência de alguma característica no período de coleta da amostra.

Seja Y uma variável resposta de um modelo ZIP,

$$P(Y = y) = (1 - p) \frac{e^{-\lambda} (\lambda)^y}{y!}, \quad y = 1, 2, \dots$$

$$P(Y = 0) = p + (1 - p)e^{-\lambda}$$

em que $0 < p < 1$. Assim, são incorporados mais zeros do que os permitidos pela distribuição de Poisson. Podemos ainda mostrar que

$$E(Y) = (1 - P)\lambda$$

$$e \text{ Var}(Y) = E(Y) + E(Y)[\lambda - E(Y)]$$

Mais informações em Gupta, Gupta, e Tripathi (1996).

O modelo ZIP ajusta dois modelos de regressão, um logístico que considera os zeros em excesso e um com as contagens de interesse. Considerando a importancia dos Modelos de Poisson Inflacionados em Zeros para estes casos específicos da regressão de Poisson e sua utilidade na modelagem de problemas na agricultura, economia, manufaturas, aplicações de patentes, segurança em rodovias, medicina e comportamento sexual, (Hall 2000) e (Ridout, Demétrio, e Hinde 1998), o trabalho proposto tem o objetivo de estudar profundamente os métodos de estimação de seus parâmetros e aplicação a um conjunto de dados.

Serão feitas análises descritivas, diagnósticos de modelo, interpretações a respeito do problema e predição. Todo o trabalho será conduzido com o uso da linguagem e ambiente de computação estatística R (R Core Team 2021).

Referências

- Gupta, Pushpa L, Ramesh C Gupta, e Ram C Tripathi. 1996. “Analysis of zero-adjusted count data”. *Computational Statistics & Data Analysis* 23 (2): 207–18.
- Hall, Daniel B. 2000. “Zero-inflated Poisson and binomial regression with random effects: a case study”. *Biometrics* 56 (4): 1030–39.
- Lambert, Diane. 1992. “Zero-inflated Poisson regression, with an application to defects in manufacturing”. *Technometrics* 34 (1): 1–14.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ridout, Martin, Clarice GB Demétrio, e John Hinde. 1998. “Models for count data with many zeros”. In *Proceedings of the XIXth international biometric conference*, 19:179–92. 19. International Biometric Society Invited Papers. Cape Town, South Africa.