Modelos de Regressão 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

2 de julho de 2021

1 Introdução

ainda tá bem ruim

Neste trabalho são expostos conceitos teóricos a respeito dos modelos de Poisson Inflacionados em Zero (ZIP) e uma aplicação utilizando um conjunto de dados real. O conjunto é formado por viagens de acampamento feitas por 250 grupos de pessoas à um parque nos Estados Unidos.

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

2 Métodos

2.1 Distribuição de Poisson

2.2 Modelos de Poisson Inflacionados em Zero (ZIP)

Uma propriedade importante da distribuição de Poisson é que a média e variância são iguais, $Var(y_i|x_i) = E(y_i|x_i) = \lambda_i$, esta propriedade é referida como equidispersão. Na prática a suposição de equidispersão é violada quando a variância das contagens observadas é maior que a média por conta de heterogeneidade não observada ou quando a frequência de zeros é maior que o número de zeros esperado em uma distribuição de Poisson.

Assumindo que a variável resposta tem distribuição de Poissson e que o logaritmo de seu valor esperado pode ser modelado por uma combinação linear de parâmentros desconhecidos. Seja y_i, $i = 1, ..., n$ a variável resposta de um modelo de regressão, assumimos que y_i tem distribuição de Poisson com média λ_i, definida em função das covariáveis x_i. Assim, um modelo de regressão de Poisson é dado por

$$P(y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$$

em que a esperança condicional é dado por $\lambda_i = E(y_i|x_i) = \exp(x'\beta)$. O vetor $\mathrm{x}_i' = (x_{i,1}, x_{i,2}, ..., x_{i,p})$ contem as covariáveis e $\beta' = (\beta_1, ..., \beta_p)$ é o vetor de parâmetros do modelo de regressão.

2.3 Teste de superdispersão

O teste de superdispersão proposto por Cameron e Trivedi (1990) é baseado em uma regressão linear sem o intercepto com as hipóteses

$$H_0 : Var\left(y_i\right) = \lambda_i \; vs \; H_1 : Var\left(y_i\right) = \lambda_i + \alpha g\left(\lambda_i\right)$$

em que α é um parâmetro desconhecido e $g(.)$ é uma função definida, comumente $g(\lambda_i) = \lambda_i^2$ ou $g(\lambda_i) = \lambda_i$. Este teste é condusido estimando-se o modelo de Poisson, construindo $\hat{\lambda}_i = \exp\left(\mathrm{x}_i'\hat{\beta}\right)$ e ajustando um modelo por mínimos quadrados ordinários sem o intercepto

$$\frac{(y_i - \hat{\lambda}_i)^2 - y_i}{\hat{\lambda}_i} = \alpha \frac{g(\hat{\lambda}_i)}{\hat{\lambda}_i} + e_i$$

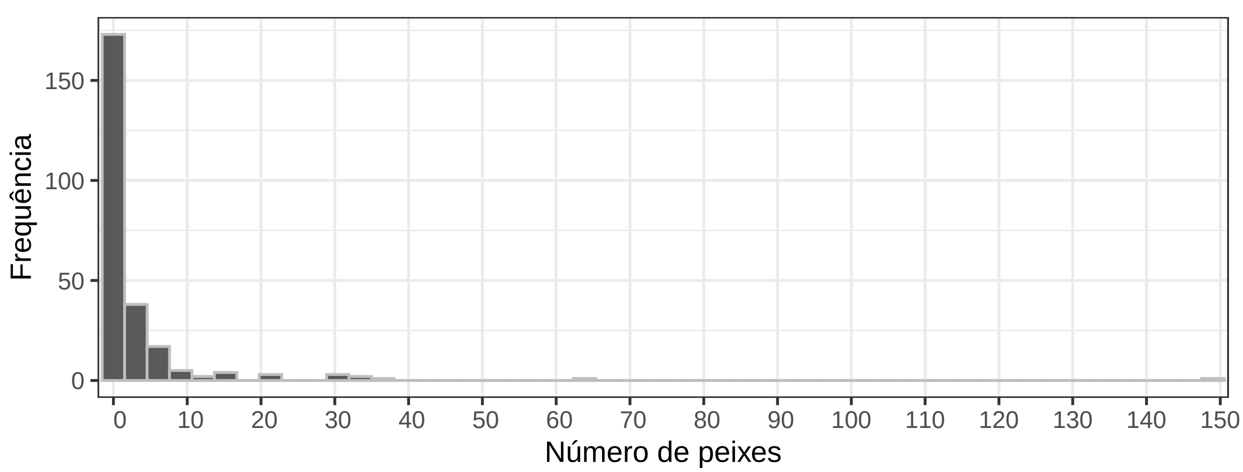
em que e_i é o erro. A significância do coeficiente α implica na existencia de superdispersão nos dados.

3 Aplicação

Os dados são provenientes de 250 acampamentos familiares em um parque nos Estados Unidos. Cada grupo foi questionado sobre o número de peixes capturados, quantas pessoas o grupo tinha e quantas crianças o grupo tinha, e se eles foram acompanhados por um guia.

- **LIVE_BAIT:** Variável binária. Indica se foram usadas iscas vivas ou não;
- **CAMPER:** Variável binária. Indica se o o grupo foi acompanhado por um guia ou não;
- **PERSONS:** Variável numérica. Número de pessoas no grupo;
- **CHILDREN:** Variável numérica. Número de crianças no grupo;
- **FISH_COUNT:** Variável numérica. Número de peixes pegos pelo grupo;

```
##  LIVE_BAIT CAMPER     PERSONS         CHILDREN       FISH_COUNT
##  0: 34     0:103   Min.   :1.000   Min.   :0.000   Min.   :  0.000
##  1:216     1:147   1st Qu.:2.000   1st Qu.:0.000   1st Qu.:  0.000
##                    Median :2.000   Median :0.000   Median :  0.000
##                    Mean   :2.528   Mean   :0.684   Mean   :  3.296
##                    3rd Qu.:4.000   3rd Qu.:1.000   3rd Qu.:  2.000
##                    Max.   :4.000   Max.   :3.000   Max.   :149.000
```



4 Conclusões

Referências

Cameron, A.Colin, e Pravin K. Trivedi. 1990. "Regression-based tests for overdispersion in the Poisson model". *Journal of Econometrics* 46 (3): 347–64. https://doi.org/https://doi.org/10.1016/0304-4076(90)90014-K.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.