

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 apresenta uma aplicação prática utilizando um conjunto de dados real. O conjunto de dados 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

Os dados que temos, são dados de contagem, portanto, primeiro podemos tentar utilizar o modelo de poisson que é o modelo generalizado linear mais simples e mais utilizado para modelar esse tipo de dados. A distribuição de poisson é uma distribuição que possui apenas uma moda e apenas 1 parâmetro μ que é igual a sua média e a sua variância.

2.2 Modelos de Poisson Inflacionados em Zero (ZIP)

Ter dados com zeros inflacionados significa que, apesar de claramente termos um modelo de contagem, é possível perceber que tem-se mais zeros do que o esperado para dados que seguem uma distribuição de poisson. Apesar da quantidade de zeros variar de acordo com a média da poisson, através do gráfico 1 fica evidente que no banco de peixe há um número de zeros maior do que o esperado. Ainda, é possível perceber que os dados estão sobrediversados, logo a variância não é igual à média e as suposições do modelo de poisson é violada. Ignorar os zeros inflacionados dos dados de contagem pode ter alguns problemas como por exemplo os parâmetros estimados e os erros serem viesados. O modelo ZIP ajusta dois modelos de regressão, um logístico que considera os zeros em excesso e um com as contagens de interesse. Este tipo de modelo foi proposto por Lambert (1992)

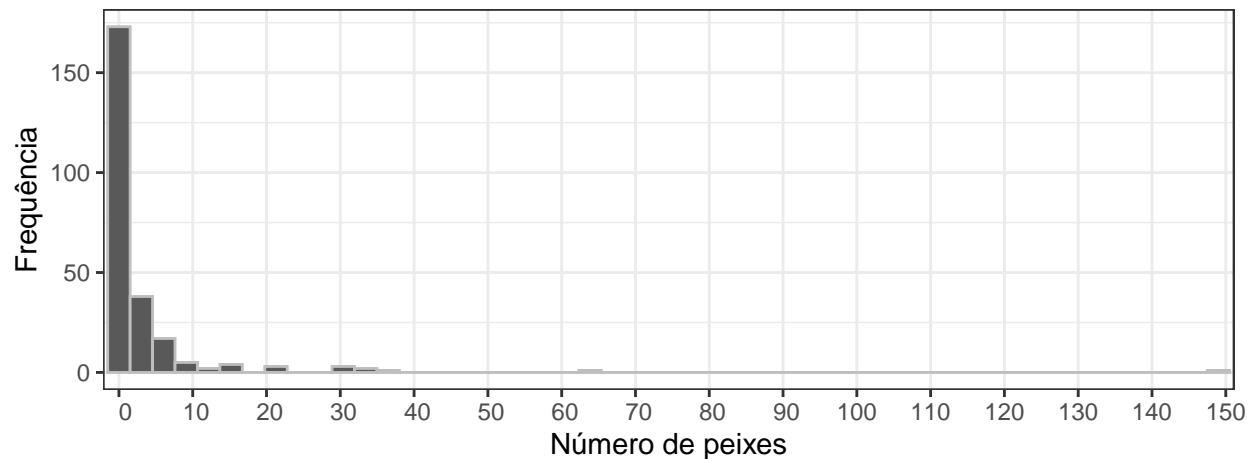
3 Aplicação

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

- **LIVE_BAIT:** Variável binária. Indica se foram usadas iscas vivas ou não;
- **CAMPER:** Variável binária. Indica se 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

Através das estatísticas sumárias, pode-se ter uma ideia do banco de dados. Por exemplo, pode-se ver que há pesca com iscas vivas foram maiores do que com iscas não vivas. Além disso, é possível ver que apesar da média da peixes pescados ser 3.286, há alguns dados que puxam essa média para esse valor, uma vez que a mediana é de 0.

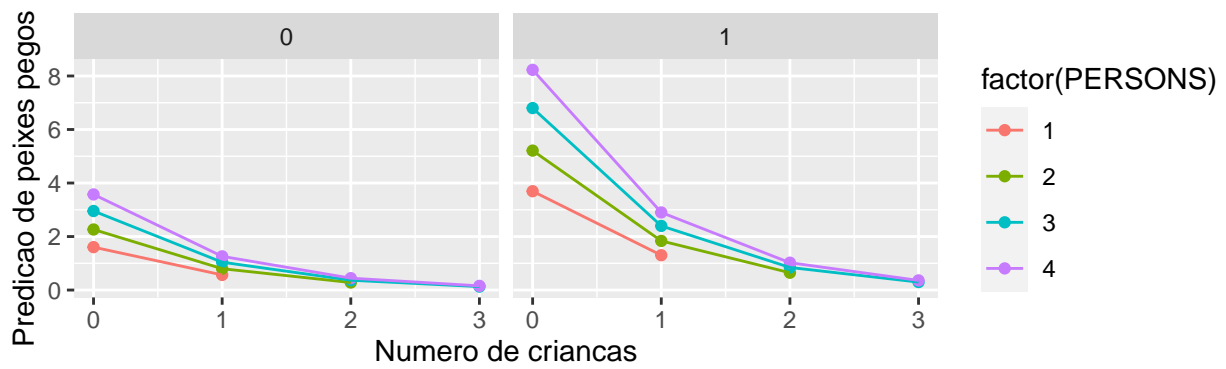


Através do gráfico 1 é possível perceber que tem-se mais zeros do que o esperado para dados de contagem que seguem uma distribuição de poisson. Ainda, é possível perceber que os dados estão sobredispersados, logo a variância não é igual à média e as suposições do modelo de poisson é violada.

O gráfico 2, foi realizado para saber aplicar uma transformação logaritmica na contagem de peixes seria uma opção para aplicação do modelo de poisson. Entranto, claramente os dados ainda não ficaram da forma ideal para a utilização deste tipo de modelo. Diante disto, há indícios de que uma das opções possíveis para realizar a modelagem é utilizar algum modelo do tipo ZIP, que lida melhor com essas violações. Portanto, foi aplicado um modelo de regressão de poisson para zeros inflacionados

Pode-se perceber que é retornado dois modelos diferentes. O primeiro modelo ajustado foi uma regressão de poisson, enquanto que o segundo modelo corresponde ao modelo inflacionado. Na saída dos modelos pode-se observar os coeficientes estimados, os erros padronizados e os p-valores. Neste caso, todos os preditores dos modelos demonstraram ser significantes.

Por fim, é possível realizar predições com o objetivo de prever o número de peixes pegos de acordo com as variáveis regressoras.



Fica claro que ter crianças no grupo, faz com que o número esperado de peixe pegos seja menor do que não ter crianças no grupo. Ainda, o número esperado de peixes pegos nos grupos que foram acampar é o dobro dos grupos que não foram acampar.

4 Conclusões

Referências

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