

Regressão para dados de contagem

Mirre Milanez

Patrícia Dias Soares

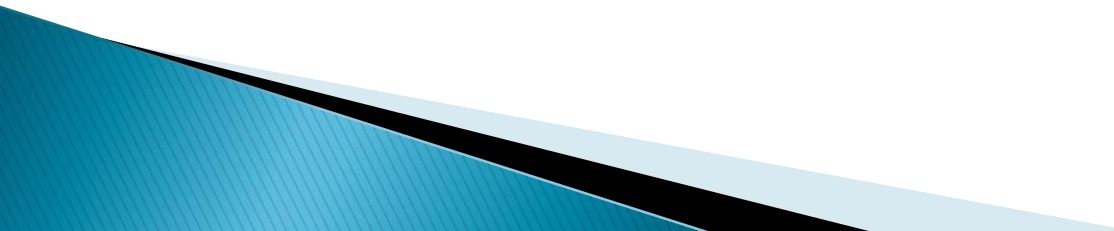
Disciplina de Estatística – Prof. Paulo Justiniano

- ▶ Em uma regressão de contagem as variáveis são dependentes.
- ▶ Os dados possuem significado por eles mesmos, ou seja, eles representam o número de vezes que um evento ocorreu.

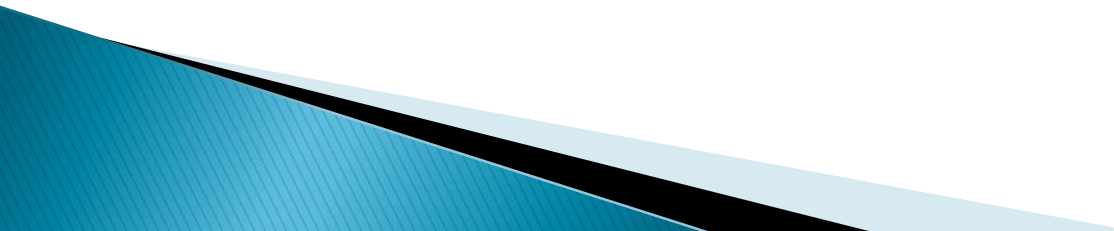
Exemplo:

- Quantas vezes o indivíduo foi ao médico;
- Número de vezes foi prescrita uma determinada droga para um indivíduo;
- Quantas vezes o indivíduo parou de fumar em um ano;
- Quantas vezes uma pessoa teve acidente de trabalho em um determinado período de tempo;
- Número de vezes uma pessoa bebeu em um dado período de tempo.
- Número de vezes uma pessoa come frutas ou vegetais.

Regressão de Poisson

- ▶ Regressão de Poisson – é usada para variáveis do modelo de contagem com o pressuposto de que os médias condicionais são iguais as variâncias condicionais.
 - ▶ Dados adequados para a regressão de Poisson não acontecem com muita frequência.
 - ▶ No entanto, a regressão de Poisson é frequentemente utilizado como um ponto de partida para a modelagem de dados de contagem.
- 

Devemos considerar

- ▶ Uma causa comum de mais de dispersão é zeros em excesso, o que, por sua vez, são gerados por um processo adicional de dados de geração. Nesta situação, modelo de zero-inflado deve ser considerado.
 - ▶ A variável de desfecho em uma regressão de Poisson não pode ter números negativos e a exposição não pode ter 0s.
 - ▶ Regressão de Poisson é estimada através de máxima verossimilhança. Ela geralmente requer uma amostra grande.
- 

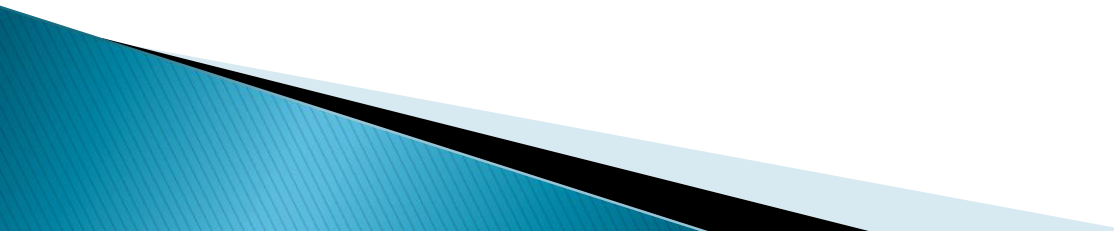
Regressão Binomial Negativa

- ▶ Regressão binomial negativa – regressão binomial negativa pode ser usada para dados mais dispersos de contagem, que é quando a variância condicional excede a média condicional.
- ▶ Pode ser considerado como uma generalização da regressão de Poisson, uma vez que tem a mesma estrutura como média de regressão de Poisson e tem um parâmetro adicional para modelar a dispersão excessiva.

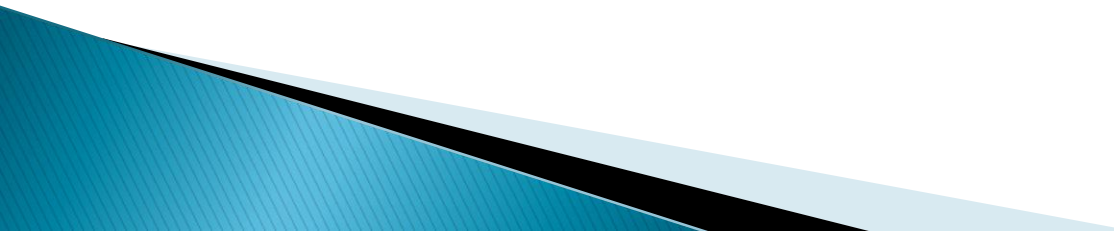
Devemos considerar

- ▶ Não é recomendável que nos modelos binomiais negativos sejam aplicados a amostras de pequenas dimensões.
- ▶ Uma causa comum de maior dispersão é o excesso de zeros por um processo adicional de geração de dados. Nesta situação, o modelo de zero-inflado deve ser considerado.
- ▶ Se o processo de geração de dados não permite qualquer zero (tais como o número de dias passados no hospital), então um modelo de zero truncada pode ser mais apropriada.
- ▶ Dados de contagem, muitas vezes têm uma variável de exposição, o que indica o número de vezes que o evento poderia ter acontecido.
- ▶ A variável de desfecho em uma regressão binomial negativa não pode ter números negativos.

Regressão de Zero-Inflado

- ▶ O modelo de Regressão Zero-inflado tenta explicar os zeros em excesso.
 - ▶ Em outras palavras, dois tipos de zeros que existam nos dados, "zeros verdadeiros" e "zeros em excesso".
 - ▶ O modelo de Zero-inflado estima duas equações simultaneamente, uma para o modelo de contagem e um para os zeros em excesso.
- 

Devemos considerar

- ▶ Uma vez que o zip tenha tanto um modelo de contagem e um modelo logístico, cada um dos dois modelos devem ter bons preditores. Os dois modelos não precisam necessariamente usar os mesmos preditores.
 - ▶ Problemas de predição perfeita, de separação ou separação parcial pode ocorrer na parte logística do modelo de zero-inflado.
 - ▶ Não é recomendável que os modelos de Poisson zero-inflado sejam aplicados a amostras de pequenas dimensões.
- 

Referencias

- ▶ <http://www.ats.ucla.edu/stat/dae/>