

CIÊNCIA DE DADOS (BIG DATA)

ANÁLISE ESTATÍSTICA

Professor curador: Mário Olímpio de Menezes



Mackenzie



TRILHA 6

PARTE B – MODELOS LINEARES GENERALIZADOS, REGRESSÃO LOGÍSTICA E DE POISSON

PARTE B – MODELOS LINEARES GENERALIZADOS, REGRESSÃO LOGÍSTICA E DE POISSON

REGRESSÃO DE POISSON

REGRESSÃO DE POISSON

- A Regressão de Poisson é útil quando queremos prever uma variável resposta que representa contagens.
- Uma variável aleatória Y é dita ter uma distribuição de Poisson com um parâmetro μ se ela assume valores inteiros $y = 0, 1, 2, \dots$, com probabilidade dada por:

$$\Pr\{Y = y\} = \frac{e^{-\mu} \mu^y}{y!}$$

considerando-se $\mu > 0$.

REGRESSÃO DE POISSON

A distribuição de Poisson tem **média** e **variância** iguais, assim a hipótese de *homocedasticidade* não é apropriada para dados que obedecem a distribuição de Poisson.

- Uma outra descrição da distribuição de Poisson é em termos de processos estocásticos.

REGRESSÃO DE POISSON

Suponha eventos que ocorram aleatoriamente no tempo:

- A probabilidade de, no mínimo, uma ocorrência do evento em um dado intervalo de tempo é proporcional ao comprimento do intervalo.
- A probabilidade de duas ou mais ocorrências do evento em um intervalo de tempo bem pequeno é negligível.

REGRESSÃO DE POISSON

- Os números de ocorrência do evento em intervalos de tempo disjuntos são mutuamente independente.

Então, a distribuição de probabilidade do número de ocorrências do evento em um intervalo de tempo fixo é Poisson com média $\mu = \lambda t$, em que λ é a taxa de ocorrência do evento por unidade de tempo e t é o comprimento do intervalo de tempo.

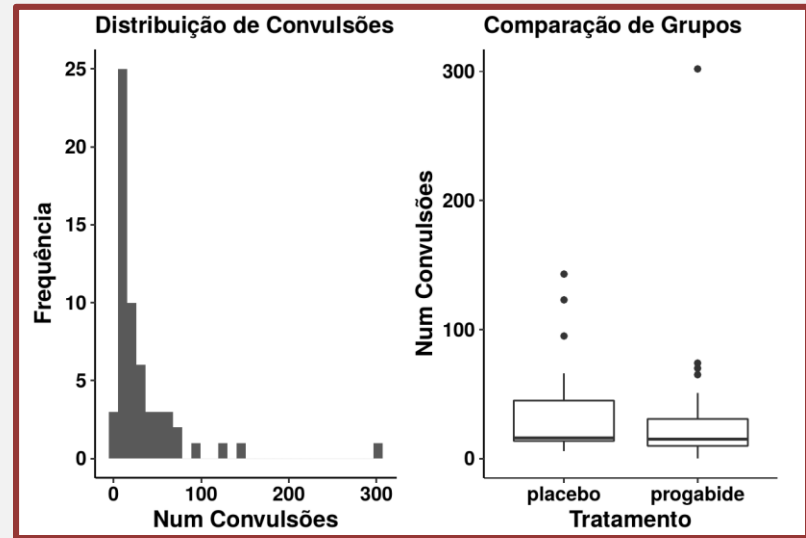
EXEMPLO – REGRESSÃO DE POISSON

Analisa se existe evidência ou não de que uma droga diminui o número de convulsões em pacientes.

- Variável Resposta é do tipo contagem (número de convulsões).
- Variáveis Explicativas são do tipo categóricas e numéricas.

EXEMPLO – VARIÁVEL RESPOSTA

- Distribuição distorcida.
- Número de convulsões com droga parece ter menor variância (média menor) – isto é esperado.



ANÁLISE DO MODELO DE POISSON

- Na regressão de Poisson, a variável dependente modelada é o log da média condicional.
- O modelo de regressão dos dados sobre convulsão apresenta os seguintes coeficientes:

```
> coef(fit.pois)
(Intercept)      Base      Age Trtprogabide
  1.94882593   0.02265174   0.02274013  -0.15270095
```

ANÁLISE DO MODELO DE POISSON

Normalmente é mais fácil interpretar os coeficientes de regressão na escala original da variável dependente. Então, exponenciamos os coeficientes:

```
> exp(coef(fit.pois))
```

(Intercept)	Base	Age	Trtprogabide
7.0204403	1.0229102	1.0230007	0.8583864

SUPER DISPERSÃO

- Em uma distribuição de Poisson, esperamos que a variância e a média sejam iguais.
- **Super dispersão** ocorre quando a variância observada da variável resposta é maior do que seria predito pela distribuição de Poisson.
- Tem impacto negativo na interpretação dos resultados!

SUPER DISPERSÃO

DUAS PRINCIPAIS RAZÕES PARA SUPER DISPERSÃO

- Omissão de uma variável preditora importante.
- Dependência de estado, ou seja, quando um evento depende da ocorrência/evento anterior.

No caso das convulsões, a probabilidade de uma convulsão não é independente do estado do paciente.

DETECTANDO SUPER DISPERSÃO

Quando a razão dos desvios residuais pelo número de graus de liberdade dos resíduos é muito maior do que 1 temos indícios fortes de super dispersão:

```
> deviance(fit.pois)/df.residual(fit.pois)
[1] 10.1717
```

LIDANDO COM SUPER DISPERSÃO

A função **glm** do **R** provê uma alternativa para lidar com a super dispersão:

- Substituímos `family="poisson"` por `family="quasipoisson"`
- Refazemos a regressão de Poisson.

LIDANDO COM A SUPER DISPERSÃO

Analizando os coeficientes do modelo, temos exatamente os mesmos coeficientes anteriores:

```
> coef(fit.od)
(Intercept)      Base      Age Trtprogabide
 1.94882593  0.02265174  0.02274013 -0.15270095
```

Entretanto, observamos pelo sumário do modelo que tanto **Age** como **Trtprogabide** não tem significância estatística.

