PS5841

Data Science in Finance & Insurance

Poisson Regression

Yubo Wang

Spring 2022

Contingency Tables

- No constraints

$$E(Y_i) = \mu_i = n_i\theta_i$$

  - μ_i Exposure dependent
    - $\frac{\mu_i}{n_i}$ occurence rate
  - Constant exposure, $n_i = 1$
- Model with Poisson Distribution

$$f(\boldsymbol{y}; \boldsymbol{\mu}) = \prod_{i=1}^{N} \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, E(Y_i) = \mu_i$$

Poisson Distribution

- $Y_i \sim Po(\mu_i), i = 1, \ldots N$

$$f(y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y!}, y = 0,1,2, \ldots$$

$$l(\boldsymbol{\mu}; \boldsymbol{y}) = \sum_{i=1}^{N} [y_i \log(\mu_i) - \mu_i - \log y_i!]$$

Poisson Regression

$$E(Y_i) = \mu_i = n_i\theta_i$$
$$\log(\mu_i) = \log(n_i) + \boldsymbol{x}_i^T\boldsymbol{\beta}$$

offset

Goodness of Fit (1)

- Pearson chi-squared statistic

$$X^2 = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i} \sim \chi^2(N - p)$$

- Pearson residuals

$$r_i = \frac{o_i - e_i}{\sqrt{e_i}}$$

$$\sum_{i=1}^{N} r_i^2 = X^2$$

Goodness of Fit (2)

- Pseudo R^2

$$pseudo\ R^2 = \frac{l(\widehat{\boldsymbol{\beta}}_{min}) - l(\widehat{\boldsymbol{\beta}})}{l(\widehat{\boldsymbol{\beta}}_{min})}$$

That was