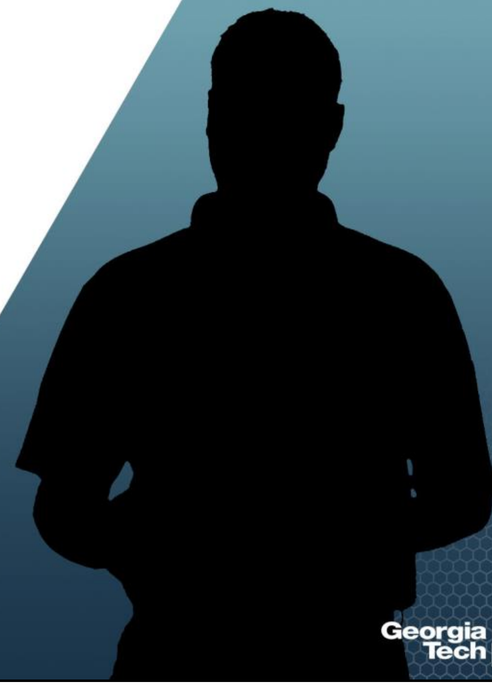
Regression Analysis

Poisson Regression

Nicoleta Serban, Ph.D.
Professor
Stewart School of Industrial and Systems Engineering



Model Description and Estimation

Georgia Tech

1

About This Lesson

Georgia Tech

2

Poisson Regression Model

Data: $\{(x_{11},.., x_{1p}), Y_1\},....,\{(x_{n1},.., x_{np}), Y_n\}$ where $Y_1, .., Y_n$ are event count data per observation unit with a Poisson distribution

Poisson Distribution: $Y \sim \text{Poisson}(\lambda)$: $P(Y=y) = \frac{e^{-\lambda}\lambda^y}{y!}$

$$E(Y) = V(Y) = \lambda$$

Model: Model the conditional expectation:

$Y_i | x_{i1,} \ldots, x_{ip} \sim \text{Poisson}(\lambda_i)$ with

$$\lambda_i = E(Y|x_1, \ldots, x_p) = e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}$$

OR

$$\log(\lambda_i) = \log(E(Y|x_1, \ldots, x_p)) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

3

Model Interpretation

The rate of event occurrence given predicting variable X = x:

$$\lambda = \lambda(x) = E(Y|x) = e^{\beta_0+\beta_1 x}$$

- The log function $\ln(\lambda(x)) = \beta_0 + \beta_1 x$ is the *log rate*.
- With an increase with one unit in x (if quantitative): $\frac{e^{\beta_0+\beta_1(x+1)}}{e^{\beta_0+\beta_1 x}} = e^{\beta_1}$
- If x categorical: $\frac{e^{\beta_0+\beta_1(x=1)}}{e^{\beta_0+\beta_1(x=0)}} = e^{\beta_1}$
- Interpretation of the regression coefficients in terms of log ratio of the rate.
- If other predicting variables are in the model, then we need to hold fixed all other predicting variables.

4

Model Estimation

Model the log rate given predictor(s):

$$\log(\lambda_i) = \log(\mathrm{E}(\mathrm{Y}|\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{p})) = \beta_0 + \beta_1 \mathrm{x}_1 + \ldots + \beta_p \mathrm{x}_\mathrm{p}$$

Parameters: $\beta_0, \beta_1, \ldots, \beta_p$

Approach: Maximum Likelihood Estimation:

$$\mathrm{L}(\beta_0, \beta_1, \ldots, \beta_p) = \prod_{i=1}^{n} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}$$

$$\max_{\beta_0, \beta_1, \ldots, \beta_p} \mathrm{l}(\beta_0, \beta_1, \ldots, \beta_p) = \log(\mathrm{L}(\beta_0, \beta_1, \ldots, \beta_p)) =$$

$$\sum_{i=1}^{n} \{y_i \log \lambda_i - \lambda_i\} = \sum_{i=1}^{n} \{y_i(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}) - e^{\beta_0 + \beta_1 \mathrm{x}_1 + \ldots + \beta_p \mathrm{x}_\mathrm{p}}\}$$

5

Model Estimation (cont'd)

Approach: Maximum Likelihood Estimation

$$\max_{\beta_0, \beta_1, \ldots, \beta_p} \mathrm{l}(\beta_0, \beta_1, \ldots, \beta_p) = \log(\mathrm{L}(\beta_0, \beta_1, \ldots, \beta_p)) =$$

$$\sum_{i=1}^{n} \{y_i \log \lambda_i - \lambda_i\} = \sum_{i=1}^{n} \{y_i(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}) - e^{\beta_0 + \beta_1 \mathrm{x}_1 + \ldots + \beta_p \mathrm{x}_\mathrm{p}}\}$$

- Maximizing the (log-)likelihood function with respect to $\beta_0, \beta_1, \ldots, \beta_p$ in close form expression is not possible because the (log-)likelihood function is a non-linear function in the model parameters
- Use numerical algorithm to estimate $\beta_0, \beta_1, \ldots, \beta_p \Rightarrow \hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$

Upshot: The estimated parameters and their standard errors are approximate estimates. Do not attempt to do it yourself! Use a statistical software to derive the estimated regression coefficients.

6

Summary

7