

Regression Analysis

Poisson Regression

Nicoleta Serban, Ph.D.

Professor

Stewart School of Industrial and Systems Engineering

Introduction



1

About This Lesson



2

1

Other Distributions of the Response

- ➔ The response variable (e.g. rate) has a **Poisson distribution**
 - What drives the rate of phone calls per day in a calling service center?
 - What predicts the density per mile of trees in a forest?

- ➔ The response variable (e.g. wait time) has an **exponential distribution**
 - What explains the wait time for a wellness visit at your physician offices?

- ➔ The response variable can have other distributions from the **exponential family of distributions**

Generalized Linear Model



3

Standard Linear Regression

Model: $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i, i=1, \dots, n$

Assumptions:

- *Linearity/Mean Zero Assumption:* $E(\varepsilon_i) = 0$
- *Constant Variance Assumption:* $\text{Var}(\varepsilon_i) = \sigma^2$
- *Independence Assumption:* $\{\varepsilon_1, \dots, \varepsilon_n\}$ are independent random variables
- **Normality Assumption:** $\varepsilon_i \sim \text{Normal}$



4

Generalized Linear Model

Data: $\{(x_{11}, \dots, x_{1p}), Y_1\}, \dots, \{(x_{n1}, \dots, x_{np}), Y_n\}$ where Y_1, \dots, Y_n response variable with a **distribution from the exponential family**

Model: Model the conditional expectation:

$$g(E(Y|x_1, \dots, x_p)) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

OR

$$E(Y|x_1, \dots, x_p) = g^{-1}(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)$$

where $g(\cdot)$ is a *link function* and $g^{-1}(\cdot)$ the *inverse link function* depending on the distribution of Y .



5

Generalized Linear Model

Y ~ *distribution in the exponential family* if its density function can be written as:

$$f(y; \theta) = h(y) e^{g(\theta)T(y) - B(\theta)}$$

where θ is the parameter of the distribution and $g(\theta)$ is the link function.

Distribution	Link	Regression Function
Normal	$g(m) = m$	$m = x^T \beta$
Poisson	$g(m) = \log(m)$	$m = e^{x^T \beta}$
Bernoulli	$g(m) = \log(m/1-m)$	$m = \frac{e^{x^T \beta}}{1 + e^{x^T \beta}}$
Gamma	$g(m) = 1/m$	$m = \frac{1}{x^T \beta}$



6

Poisson Regression

Data: $\{(x_{11}, \dots, x_{1p}), Y_1\}, \dots, \{(x_{n1}, \dots, x_{np}), Y_n\}$ where Y_1, \dots, Y_n response variable with a **Poisson distribution**

Model: Model the conditional expectation:

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

OR

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



7

Linear Regression versus Poisson Regression

Standard Linear Regression with log-transformation:

- $E(\log(Y)|x_1, \dots, x_p) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$
- $V(\log(Y)|x_1, \dots, x_p)$ constant

Poisson Regression:

- $\log(E(Y|x_1, \dots, x_p)) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$
- $V(Y|x_1, \dots, x_p) = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}$

OR

$$\log(V(Y|x_1, \dots, x_p)) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$



8

Linear Regression versus Poisson Regression

Standard Linear Regression with log-transformation:

- $E(\log(Y)|x_1, \dots, x_p) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$
- $V(\log(Y)|x_1, \dots, x_p)$ constant

Poisson Regression:

- $\log(E(Y|x_1, \dots, x_p)) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$
- $V(Y|x_1, \dots, x_p) = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}$

OR

$$\log(V(Y|x_1, \dots, x_p)) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

- Using Standard Linear Regression with log-transformation instead of Poisson Regression will result in violations of the assumption of constant variance.
- Alternatively, Standard Linear Regression could be used if the number of counts are large and with the variance stabilizing transformation $\sqrt{\mu + 3/8}$.

9



Summary



10