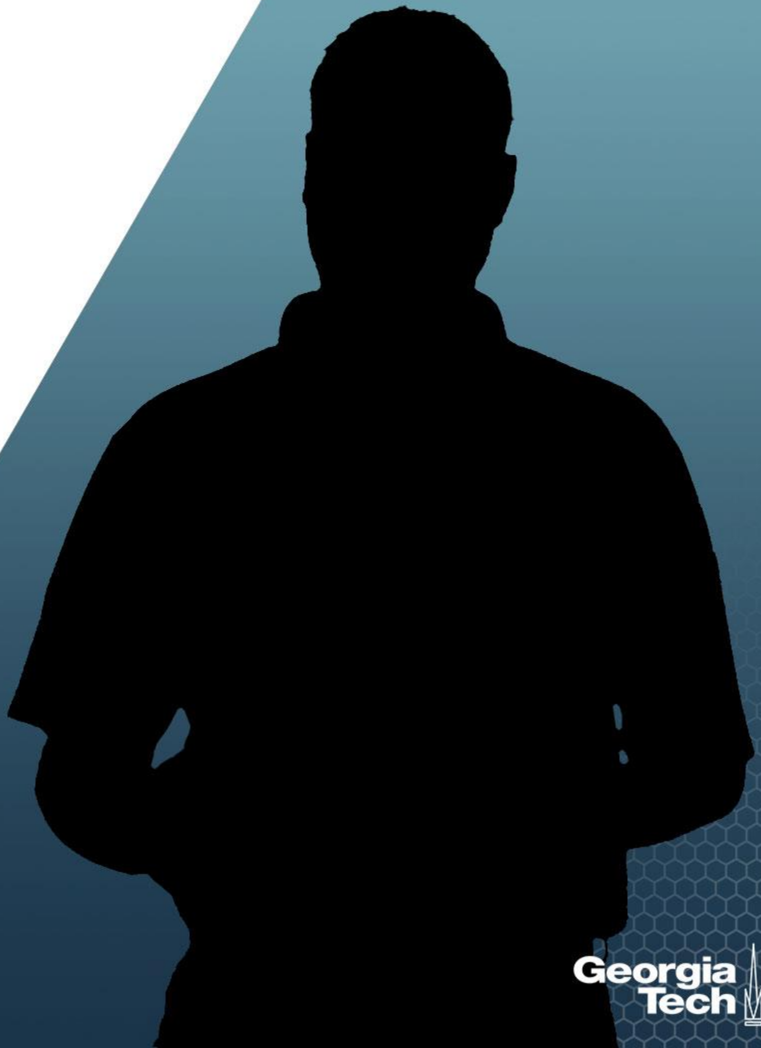
Regression Analysis

Poisson Regression

Nicoleta Serban, Ph.D.

Professor

Stewart School of Industrial and Systems Engineering



Statistical Inference

About This Lesson

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:

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

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

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

Georgia Tech

Statistical Inference

Maximum Likelihood Estimators (MLEs): $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)$

Statistical Properties of MLEs:

- Approximate Sampling Distribution: $\hat{\boldsymbol{\beta}} \approx \boldsymbol{N}(\boldsymbol{\beta}, \boldsymbol{V})$
- The normal approximation relies on the assumption of <u>large sample size</u> ⇒ Statistical inference is not reliable for small sample data

1-α Approximate Confidence interval $\left\{ \hat{\beta}_j \pm z_{\frac{\alpha}{2}} \sqrt{V(\hat{\beta}_j)} \right.$

Statistical Inference (cont'd)

- Hypothesis testing and Confidence Intervals rely on the approximately normal distribution of large sample sizes

- Use the z-test (Wald test)
  - Test is for the statistical significance of $\hat{\beta}_j$ *given all other predicting variables in the model*
  - Null hypothesis is that β_j is not significant
    $\mathrm{H_0}: \beta_j = 0 \text{ vs. } \mathrm{H_a}: \beta_j \neq 0$
  - $\mathrm{z{-}value} = \frac{\widehat{\beta}_j - 0}{\mathrm{se}(\widehat{\beta}_j)} = \frac{\widehat{\beta}_j}{\mathrm{se}(\widehat{\beta}_j)}$
  - Reject $\mathrm{H_0}$ if $|\mathrm{z{-}value}|$ is too large
    - Implies that β_j is statistically significant

Statistical Inference (cont'd)

$\text{z-value} = \frac{\widehat{\beta_j} - b}{se(\widehat{\beta_j})}$ how large to reject $H_0: \beta_j = b$?

For significance level α , Reject if $\text{z-value} > z_{\frac{\alpha}{2}}$

Alternatively, compute $\text{P-value} = 2P(Z > |\text{z-value}|)$

What if we want to test for positive relationship?

$H_0: \beta_j \leq 0$ **versus** $H_A: \beta_j > 0$?

$\text{P-value} = P(Z > \text{z-value})$

What if we want to test for negative relationship?

$H_0: \beta_j \geq 0$ **versus** $H_A: \beta_j < 0$?

$\text{P-value} = P(Z < \text{z -value})$

Georgia Tech

Statistical Inference (cont'd)

$\text{z-value} = \frac{\widehat{\beta_j} - b}{se(\widehat{\beta_j})}$ how large to reject $H_0: \beta_j = b$?

For significance level α , Reject if $\text{z-value} > z_{\frac{\alpha}{2}}$

Alternatively, compute $\text{P-value} = 2P(Z > |\text{z-value}|)$

What if we want to test for positive relationship?

$H_0: \beta_j \leq 0$ **versus** $H_A: \beta_j > 0$?

$\text{P-value} = P(Z > \text{z-value})$

What if we want to test for negative relationship?

$H_0: \beta_j \geq 0$ **versus** $H_A: \beta_j < 0$?

$\text{P-value} = P(Z < \text{z-value})$

- Because the approximation of the normal distribution relies on large sample size, so do the hypothesis testing procedures.
- What if *n* is small?
  - The hypothesis testing procedure will have a probability of type I error larger than the significance level.
  - In other words, there will likely be more type I errors than expected.

Georgia Tech

Testing for Subsets of Coefficients

Full model:

$$\mathrm{Log}\left(\mathrm{p}\left(X_1,\cdots,X_p,Z_1,\cdots,Z_q\right)\right)=$$
$$\beta_0+\beta_1X_1+\cdots+\beta_pX_p+\alpha_1Z_1+\cdots+\alpha_qZ_q$$

Reduced model:

$$\mathrm{Log}\left(\mathrm{p}\left(X_1,\cdots,X_p\ \right)\right)=\beta_0+\beta_1X_1+\cdots+\beta_pX_p$$

The hypothesis test:

$$\mathbf{H_0}\text{: }\boldsymbol{\alpha_1=\alpha_2=\cdots=\alpha_q=0}$$

vs.

$$\mathbf{H_a}\text{: }\boldsymbol{\alpha_i\neq 0\ \textbf{for at least one}\ \alpha_i,\mathrm{i}=1,\cdots,q}$$

- Maximize the likelihood function under reduced model: $\mathcal{L}\,(\bar{\beta}_0,\bar{\beta}_1,\ldots,\bar{\beta}_p)$
- Maximize the likelihood function under full model: $\mathcal{L}(\hat{\beta}_0,\hat{\beta}_1,\ldots,\hat{\beta}_p,\hat{\alpha}_1,\ldots,\hat{\alpha}_q)$
- Test Statistics
  - $\mathrm{Deviance}=\log\left(\mathcal{L}(\bar{\beta}_0,\bar{\beta}_1,\ldots,\bar{\beta}_p)\right)-\log\left(\mathcal{L}(\hat{\beta}_0,\hat{\beta}_1,\ldots,\hat{\beta}_p,\hat{\alpha}_1,\ldots,\hat{\alpha}_q)\right)\approx\chi_q^2$
  - $\text{P-value}=\Pr(\chi_q^2>\mathrm{Deviance})$

Georgia Tech

Testing for Subsets of Coefficients

Full model:

$$\text{Log}\left(\text{p}\left(X_1, \cdots, X_p, Z_1, \cdots, Z_q\right)\right) =$$
$$\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \alpha_1 Z_1 + \cdots + \alpha_q Z_q$$

Reduced model:

$$\text{Log}\left(\text{p}\left(X_1, \cdots, X_p\right)\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

The hypothesis test:

$$\mathbf{H_0}\text{: } \boldsymbol{\alpha_1 = \alpha_2 = \cdots = \alpha_q = 0}$$

vs.

$$\mathbf{H_a}\text{: } \boldsymbol{\alpha_i \neq 0 \text{ for at least one } \alpha_i, \mathrm{i} = 1, \cdots, q}$$

- Maximize the likelihood function under reduced model: $\mathcal{L}\,(\bar{\beta}_0, \bar{\beta}_1, \ldots, \bar{\beta}_p)$
- Maximize the likelihood function under full model: $\mathcal{L}(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p, \hat{\alpha}_1, \ldots, \hat{\alpha}_q)$
- Test Statistics
  - $\text{Deviance} = \log\left(\mathcal{L}(\bar{\beta}_0, \bar{\beta}_1, \ldots, \bar{\beta}_p)\right) - \log\left(\mathcal{L}(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p, \hat{\alpha}_1, \ldots, \hat{\alpha}_q)\right) \approx \chi_q^2$
  - $\text{P-value} = \Pr(\chi_q^2 > \text{Deviance})$

- The hypothesis test for subsets of coefficients is approximate
- This is not a test for goodness of fit!
  - It only compares two models

Georgia Tech

Testing for Overall Regression

Full model:

$$\mathrm{Log}\left(\mathrm{p}\left(X_1, \cdots, X_p\right)\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

Reduced model:

$$\mathrm{Log}\left(\mathrm{p}\left(X_1, \cdots, X_p\right)\right) = \beta_0$$

The hypothesis test:

$$\mathbf{H_0}: \boldsymbol{\beta_1} = \boldsymbol{\beta_2} = \cdots = \boldsymbol{\beta_q} = \mathbf{0}$$

vs.

$$\mathbf{H_a}: \boldsymbol{\beta_i} \neq \mathbf{0} \text{ for at least one } \boldsymbol{\beta_i}, \mathrm{i} = \mathbf{1}, \cdots, \boldsymbol{p}$$

- Maximize the likelihood function under reduced model: $\mathcal{L}\,(\bar{\beta}_0)$
- Maximize the likelihood function under full model: $\mathcal{L}(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)$
- Test Statistics
  - $\mathrm{Deviance} = \log\left(\mathcal{L}(\bar{\beta}_0)\right) - \log\left(\mathcal{L}(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)\right) \approx \chi_p^2$
  - $\text{P-value} = \Pr(\chi_p^2 > \mathrm{Deviance})$

Summary