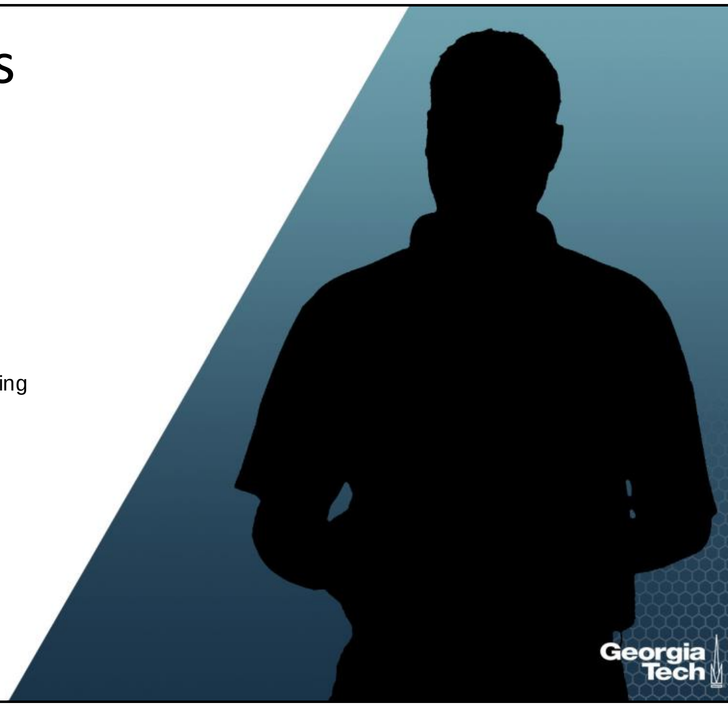
Regression Analysis

Poisson Regression

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

Statistical Inference



1

About This Lesson

2

Model Estimation

Model the log rate given predictor(s):

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

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

Approach: Maximum Likelihood Estimation:

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

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

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

Georgia Tech

3

Statistical Inference

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

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 large sample size ⇒ Statistical inference is not reliable for small sample data

1-α Approximate Confidence interval

$$\hat{\beta}_j \pm z_{\frac{\alpha}{2}} \sqrt{\mathrm{V}(\hat{\beta}_j)}$$

Georgia Tech

4

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
    $\mathrm{H_0}: \beta_j = 0$ vs. $\mathrm{H_a}: \beta_j \neq 0$
  - $\text{z}-\text{value} = \frac{\hat{\beta}_j - 0}{\text{se}(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)}$
  - Reject $\mathrm{H_0}$ if $|\text{z}-\text{value}|$ is too large
    - Implies that β_j is statistically significant

5

Statistical Inference (cont'd)

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

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?

$\mathrm{H_0}: \beta_j \leq 0$ versus $\mathrm{H_A}: \beta_j > 0$?

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

What if we want to test for negative relationship?

$\mathrm{H_0}: \beta_j \geq 0$ versus $\mathrm{H_A}: \beta_j < 0$?

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

6

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?

$H_0: \beta_j \le 0$ versus $H_A: \beta_j > 0$?

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

What if we want to test for negative relationship?

$H_0: \beta_j \ge 0$ 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

7

Testing for Subsets of Coefficients

Full model:

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

Reduced model:

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

The hypothesis test:

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

vs.

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

- 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
  - $\text{Deviance} = \log\left(\mathcal{L}(\bar{\beta}_0, \bar{\beta}_1, \ldots, \bar{\beta}_p)\right) - \log\left(\mathcal{L}(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p, \hat{\alpha}_1, \ldots, \hat{\alpha}_q)\right) \approx \chi_q^2$
  - $\text{P-value} = \Pr(\chi_q^2 > \text{Deviance})$

Georgia Tech

8

Testing for Subsets of Coefficients

Full model:

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

Reduced model:

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

The hypothesis test:

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

vs.

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

- 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

9

Testing for Overall Regression

Full model:

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

Reduced model:

$\text{Logit}\left(\text{p}\left(X_1,\cdots,X_p\right)\right)=\beta_0$

The hypothesis test:

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

vs.

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

- 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
  - $\text{Deviance}=\log\left(\mathcal{L}(\bar{\beta}_0)\right)-\log\left(\mathcal{L}(\hat{\beta}_0,\hat{\beta}_1,\ldots,\hat{\beta}_p)\right)\approx\chi_p^2$
  - $\text{P-value}=\Pr(\chi_p^2>\text{Deviance})$

10

Summary

11