Poisson Regression

STA 312 Fall 2022

See last slide for copyright information

Regression: Outcomes are Counts

- Poisson process model roughly applies
- Examples: Relationship of explanatory variables to
  - Number of children
  - Number of typos in a short document
  - Number of workplace accidents in a short time period
  - Number of marriages
- For large λ, CLT says a normality assumption is okay, but not constant variance

Linear Model for log λ

- $\log \lambda = \beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1}$
- Implicitly for $i = 1, \ldots n$
- Everybody in the sample has a different $\lambda = \lambda_i$
- Take exponential function of both sides
- Substitute into Poisson likelihood
- Maximum likelihood as usual
- Likelihood ratio tests, etc.

$$\log \lambda = \beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1}$$

- Increase x_k with everything else held constant, and
  - Log λ increases by β_k
  - λ is multiplied by $e^{\beta k}$

Back to the job study: n=200 Students

- 106 employed in a job related to field of study
- 74 employed in a job unrelated to field of study
- 20 unemployed
- Could be independent Poisson processes

- Conditionally on the total number of students, multinomial with
  - $\pi_1 = \lambda_1/(\lambda_1+\lambda_2+\lambda_3)$
  - $\pi_2 = \lambda_2/(\lambda_1+\lambda_2+\lambda_3)$
  - $\pi_3 = \lambda_3/(\lambda_1+\lambda_2+\lambda_3)$

Poisson regression with dummy variables

Job Status	d_1	d_2	$\log \lambda = \beta_0 + \beta_1 d_1 + \beta_2 d_2$
Related	0	0	$\beta_0 \quad = \log \lambda_1$
Unrelated	1	0	$\beta_0 + \beta_1 \quad = \log \lambda_2$
Unemployed	0	1	$\beta_0 + \beta_2 \quad = \log \lambda_3$

On average, we expect $e^{\beta 2}$ times as many unemployed students as students with jobs related to their fields of study.

Copyright Information