STEVENS INSTITUTE of TECHNOLOGY
School of Business

Multivariate Data Analytics

Poisson Regression and Other Models

Prof. Feng Mai
School of Business

Poisson Regression Model

- Poisson Distribution: occurrence (count) of events occurring in an interval of time or space

$$P(Y = y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$$

  - λ is the average rate of occurrence
- We would like to predict a count response/outcome variable Y, let the rate depend on Xs: $\lambda = \exp\{\mathbf{X}\beta\}$.
- Generalized Linear Model (GLM)

$$g(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k = x_i^T \beta$$

  - *g(μ)* is the link function
  - **Logistic regression** is a GLM with a logit link function *g(μ)* = ln(p/1-p), which is the inverse of the logistic function
  - **Linear regression** is a GLM with an identity link *g(μ)* = *μ*
  - **Poisson regression** is a GLM with a log link function *g(μ)* = ln(*μ*), the log of average count
- Fit using Maximum Likelihood Estimation (MLE)

$$P(Y_i = y_i|\mathbf{X}_i, \beta) = \frac{e^{-\exp\{\mathbf{X}_i\beta\}}\exp\{\mathbf{X}_i\beta\}^{y_i}}{y_i!}.$$

One observation

$$L(\beta; \mathbf{y}, \mathbf{X}) = \prod_{i=1}^{n} \frac{e^{-\exp\{\mathbf{X}_i\beta\}}\exp\{\mathbf{X}_i\beta\}^{y_i}}{y_i!}.$$

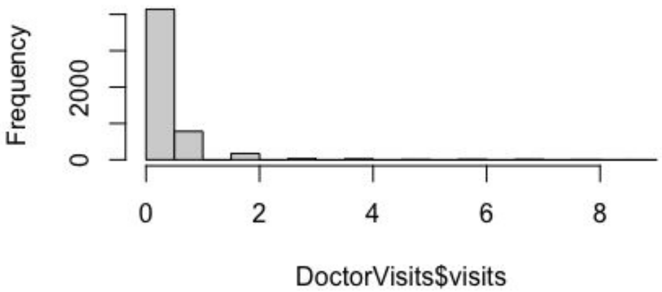
Entire dataset

Example of Poisson Regression

- Y: Number of doctor visits in past 2 weeks.
- X: gender, age, income, illness, number of days of reduced activity, general health score
- R code:

```
poi_mod <- glm(visits ~ gender + age + income + illness + reduced + health, family="poisson", data=DoctorVisits)
```

visits	gender	age	income	illness	reduced	health
9	male	62	0.25	5	14	10
8	female	19	0.25	2	7	0
8	male	52	0.25	5	14	7
8	male	57	0.01	1	9	4
8	female	62	1.50	1	14	1
8	male	67	0.25	2	14	5
7	male	22	0.00	1	14	1



Example of Poisson Regression (continued)

Interpretation using Incidence Rate Ratio

EXP(b) provides the incidence rate ratios

The ratio of

$$\frac{\text{average rate of Y after a z units increase in x}}{\text{the average rate of Y}}$$

$= EXP(z*b_1)$

Example:

$b_{age} = 0.005$, $p < 0.01$

As age increases by 10 years, the average number of doctors' visits increases by a factor of EXP(0.005*10) = 1.05

As age increases by 30 years, the average number of doctors' visits increases by a factor of EXP(0.005*30) = 1.16

	Dependent variable:
	visits
genderfemale	0.188***
	(0.055)
age	0.005***
	(0.001)
income	-0.126
	(0.080)
illness	0.198***
	(0.018)
reduced	0.128***
	(0.005)
health	0.031***
	(0.010)
Constant	-2.136***
	(0.096)
Observations	5,190
Log Likelihood	-3,364.467
Akaike Inf. Crit.	6,742.933
Note:	*p<0.1; **p<0.05; ***p<0.01

Other Regression Models

Negative Binomial Regression

- Similar to Poisson Regression, with over dispersed y (variance >> mean)

Ordered Logistic Regression

- The dependent variable y is a factor (categorical variable) with orders.
  - Survey responses (disagree, neutral, agree) or (likely, somewhat likely, unlikely, very unlikely)

Multinomial Logistic Regression

- The dependent variable y is a factor (categorical variable) with more than two categories.
  - Choice data: product colors, college majors
- Alternative: (One vs Rest)

Beta Regression

- The dependent variable y takes values between (0, 1)
  - Rates, proportions, and indices such as Gini

Thank you!

Prof. Feng Mai
School of Business