BIOSTAT 651
Notes #13: Poisson Regression

- Lecture Topics:
  - Motivation: count, rate data
  - Inference procedures
  - Examples
- Text (Dobson & Barnett, 3rd Ed.): Chapter 9

Data Structure: Count Response

- Often in biomedical studies, the response variate is a *count*
  - response = event count

- Examples:
  - number of incident lung cancer cases in Michigan, 2000-2004
  - number of acute liver failure cases diagnosed between 01/01/1990 and 12/31/1999, among males age 40-64 working in either Michigan, Ohio or Illinois.
  - number of physician visits for influenza among University of Michigan students in 2005

Data Structure: Count, Rate

- If follow-up time (or, *exposure*) is constant across all subjects, we can model the mean count directly
  - more generally, we model rate, where:

$$\text{rate}_i = \frac{E(\text{count}_i)}{\text{exposure}_i}$$

- Rates are typically reported as events per unit time (or person-time)
  - e.g., cases per 100,000 patient-years

- Units of time dimension are arbitrary
  - must use same units for all observations.

Example: Equal Exposure

- Example: A group of clinicians wishes to study the association between air temperature and physician claims for asthma. The setting is Ann Arbor, with observed data given in the following table:

Month	Avg. Temp	Claims
Sep	51	16
Oct	44	14
Nov	40	9
Dec	29	6
Jan	23	11
Feb	21	8
Mar	32	15
Apr	41	17
May	62	7
Jun	78	5

Examples: Unequal Exposure

- Example: A study carried out at the University of Michigan Department of Internal Medicine examined hospital admission patterns among dialysis patients. Each patient was followed from the initiation of dialysis until 12/31/2009. Below is a subset of the records from the analysis file:

IDNUM	YEARS	ADMITS	DIAB	AGE
1	0.6	5	1	61
2	4.2	11	0	56
3	1.3	7	1	45
4	2.2	8	0	66

Poisson Distribution

- General data structure:

  Y_i = event count

  $\mathbf{x}_i$ = covariate vector

  T_i = exposure

  observed data: $(\mathbf{x}_i, Y_i, T_i)$ for $i = 1, \ldots, n$

- Assume that:
  - triplets $(\mathbf{x}_i, Y_i, T_i)$ are *iid*
  - Probability function:
  - Moments:
  - Rate function:

Poisson Regression Model

- Rate model:

$$\log \lambda_i = \mathbf{x}_i^T \boldsymbol{\beta}$$

  - written in terms of mean function:

$$\log E\left[\frac{Y_i}{T_i}\right] = \mathbf{x}_i^T \boldsymbol{\beta}$$

$$\log E[Y_i] = \log T_i + \mathbf{x}_i^T \boldsymbol{\beta}$$

  - the term $\log(T_i)$ is referred to as an offset

Poisson Regression as a GLM

- Poisson regression model is a special case of a GLM
  - link function:
  - mean function:
  - variance function:
- Standard implementation features the canonical link

Interpretation of Parameters

- Consider a Poisson model with a single covariate where (for now) X_i is continuous

$$\log \lambda_i = \beta_0 + \beta_1 (X_i - \overline{X})$$

- Interpretation of β_0:

- Interpretation of β_1:

Interpreting Parameters

- Suppose now that the covariate is binary:

  $X_i = I_i(\text{treated})$

$$\log \lambda_i \quad = \quad \beta_0 + \beta_1 X_i$$

- Interpretation of β_0:

- Interpretation of β_1:

Poisson Regression: Estimation

- Applying general GLM results to the case where $Y_i \sim$ follows the Poisson model:

$$\lambda_i = \lambda(\mathbf{x}_i) \quad = \quad e^{\mathbf{x}_i^T \boldsymbol{\beta}}$$

  - Score function:

$$\begin{aligned} U(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \mathbf{x}_i^T (y_i - \mu_i) = \sum_{i=1}^{n} \mathbf{x}_i^T (y_i - T_i \lambda_i) \\ &= \sum_{i=1}^{n} \mathbf{x}_i (y_i - T_i e^{\mathbf{x}_i^T \boldsymbol{\beta}}) \end{aligned}$$

  - Information matrix:

$$\begin{aligned} J(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T v(\mu_i) \\ &= \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T T_i e^{\mathbf{x}_i^T \boldsymbol{\beta}} \end{aligned}$$

Poisson Model: Maximum Likelihood

- From first principles,

$$
\begin{aligned}
L_i(\boldsymbol{\beta}) &\propto \frac{\exp\{-T_i e^{\mathbf{x}_i^T\boldsymbol{\beta}}\}\left(T_i e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)^{Y_i}}{Y_i!} \\
&= \exp\{-T_i e^{\mathbf{x}_i^T\boldsymbol{\beta}}\}\left(T_i e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)^{Y_i} \\
\ell_i(\boldsymbol{\beta}) &= -T_i e^{\mathbf{x}_i^T\boldsymbol{\beta}} + Y_i \log T_i + Y_i \mathbf{x}_i^T\boldsymbol{\beta} \\
U_i(\boldsymbol{\beta}) &= \mathbf{x}_i(Y_i - T_i e^{\mathbf{x}_i^T\boldsymbol{\beta}}) \\
J_i(\boldsymbol{\beta}) &= \mathbf{x}_i\mathbf{x}_i^T T_i e^{\mathbf{x}_i^T\boldsymbol{\beta}}
\end{aligned}
$$

Poisson Regression: SAS

- Like other GLMs, Poisson models can be fitted using PROC GENMOD
- e.g., Reconsider the data set from Slide #5, with input records of the form:

IDNUM	YEARS	ADMITS	DIAB	AGE	LOG_YRS
1	0.6	5	1	61	-0.5108
2	4.2	11	0	56	1.4351
3	1.3	7	1	45	0.2624
4	2.2	8	0	66	0.7885

```
PROC GENMOD DATA=dialysis;
 MODEL admits = diab age / DIST=Poisson LINK=log
                           OFFSET=log_yrs;
RUN;
```