Robust Poisson Regression

Siyu Zhao, Zheming Yu, Christopher Liu, Runnan Liu, Shiming Zhang

10/13/2018

San Diego State University, Stats 610

The Poisson Distribution

A random variable Y is said to have a Poisson distribution with parameter λ if its probability is given by the probability mass function

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

for $\lambda > 0$ and y = 0, 1, 2, ...

The mean and variance of this distribution can be shown to be

$$E(Y) = Var(Y) = \lambda$$

What is Poisson Regression?

In Poisson regression

- Model used when the response variable, Y, is a count (eg. Number of vehicle accidents per year, number of visits to a website over a certain time span, etc)
- We can also have the response variable be Y/t, the rate of the event happening with t being an interval representing time, space, or some other grouping.

Robust Poisson Regression

[Placeholder Text]

Smoking and Lung Cancer Dataset

Smoking and Lung Cancer

We will be using data originally collected from a Canadian study of mortality by age and smoking status

The file containing cleaned data comes from the website of Professor German Rodriguez of Princeton University

link: http://data.princeton.edu/wws509/datasets/smoking.raw

Smoking and Lung Cancer Dataset

1	1	656	18
2	1	359	22
3	1	249	19
4	1	632	55
5	1	1067	117
6	1	897	170
7	1	668	179
8	1	361	120
9	1	274	120
1	2	145	2
2	2	104	4
3	2	98	3
4	2	372	38
5	2	846	113
6	2	949	173
7	2	824	212
8	2	667	243
9	2	537	253
1	3	4531	149
2	3	3030	169
3	3	2267	193
4	3	4682	576

- age at start of study: coded 1 to 9 for 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80+ respectively
- smoking status: 1 = never smoked, 2 = smoked cigars or pipe only, 3 = smoked cigarettes and cigar or pipe, 4 = smoked cigarettes only
- population: number of male pensioners followed
- deaths: number of deaths in a six-year period