Poisson regression is used to model count variables.
This page uses the following packages. Make sure that you can load them before trying to run the examples on this page. If you do not have a package installed, run: install.packages("packagename"), or if you see the version is out of date, run: update.packages().

Poisson Regression with R

Please note: The purpose of this page is to show how to use various data analysis commands. It does not cover all aspects of the research process which researchers are expected to do. In particular, it does not cover data cleaning and checking, verification of assumptions, model diagnostics or potential follow-up analyses.

Examples of Poisson regression

Example 1. The number of persons killed by mule or horse kicks in the Prussian army per year. Ladislaus Bortkiewicz collected data from 20 volumes of Preussischen Statistik. These data were collected on 10 corps of the Prussian army in the late 1800s over the course of 20 years.

Example 2. The number of people in line in front of you at the grocery store. Predictors may include the number of items currently offered at a special discounted price and whether a special event (e.g., a holiday, a big sporting event) is three or fewer days away.

Example 3. The number of awards earned by students at one high school. Predictors of the number of awards earned include the type of program in which the student was enrolled (e.g., vocational, general or academic) and the score on their final exam in math.

Poisson Regression with R

Description of the data
For the purpose of illustration, we have simulated a data set for Example 3 above. In this example, num_awards is the outcome variable and indicates the number of awards earned by students at a high school in a year, math is a continuous predictor variable and represents students' scores on their math final exam, and prog is a categorical predictor variable with three levels indicating the type of program in which the students were enrolled. It is coded as 1 = "General", 2 = "Academic" and 3 = "Vocational". Let's start with loading the data and looking at some descriptive statistics.

Poisson Regression with R

```
p <- read.csv("http://www.ats.ucla.edu/stat/da
p <- within(p, {
  prog <- factor(prog, levels=1:3, labels=c("G
  id <- factor(id)
})
summary(p)
```

Poisson Regression with R

```
       id          num_awards          prog
1       :  1   Min.   :0.00   General    : 45
2       :  1   1st Qu.:0.00   Academic   :105
3       :  1   Median :0.00   Vocational: 50
4       :  1   Mean   :0.63
5       :  1   3rd Qu.:1.00
6       :  1   Max.   :6.00
(Other):194
```

Poisson Regression with R

Each variable has 200 valid observations and their distributions seem quite reasonable. The unconditional mean and variance of our outcome variable are not extremely different. Our model assumes that these values, conditioned on the predictor variables, will be equal (or at least roughly so).

Poisson Regression with R

We can use the tapply function to display the summary statistics by program type. The table below shows the average numbers of awards by program type and seems to suggest that program type is a good candidate for predicting the number of awards, our outcome variable, because the mean value of the outcome appears to vary by prog. Additionally, the means and variances within each level of prog–the conditional means and variances–are similar. A conditional histogram separated out by program type is plotted to show the distribution.