Poisson Regression

- Poisson regression is used to model count variables.
- Poisson regression has a number of extensions useful for count models.

Negative Binomial regression

- Negative binomial regression can be used for over-dispersed count data, that is when the conditional variance exceeds the conditional mean.
- It can be considered as a generalization of Poisson regression since it has the same mean structure as Poisson regression and it has an extra parameter to model the over-dispersion.
- If the conditional distribution of the outcome variable is over-dispersed, the confidence intervals for Negative binomial regression are likely to be narrower as compared to those from

Zero-inflated Regression models

- Zero-inflated models attempt to account for excess zeros.
- In other words, two kinds of zeros are thought to exist in the data, "true zeros" and "excess zeros".
- Zero-inflated models estimate two equations simultaneously, one for the count model and one for the excess zeros.

OLS regression

- Count outcome variables are sometimes log-transformed and analyzed using OLS regression.
- Many issues arise with this approach, including loss of data due to undefined values generated by taking the log of zero (which is undefined) and biased estimates.

Examples of Poisson regression

- 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.

Poisson Regression with R

Examples of Poisson regression

- 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.

Examples of Poisson regression

- 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.

Description of the data

- For the purpose of illustration, we have simulated a data set for the last example.
- The data set is called *poissonreg.csv*
- 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

Predictor Variables

- **math** is a continuous predictor variable and represents students' scores on their math final exam,
- **prog** is a categorical predictor variable with three levels indicating the type of program in which the students were enrolled.
- **prog** is coded as $1 =$ "*General*", $2 =$ "*Academic*" and $3 =$ "*Vocational*".

Poisson Regression with R

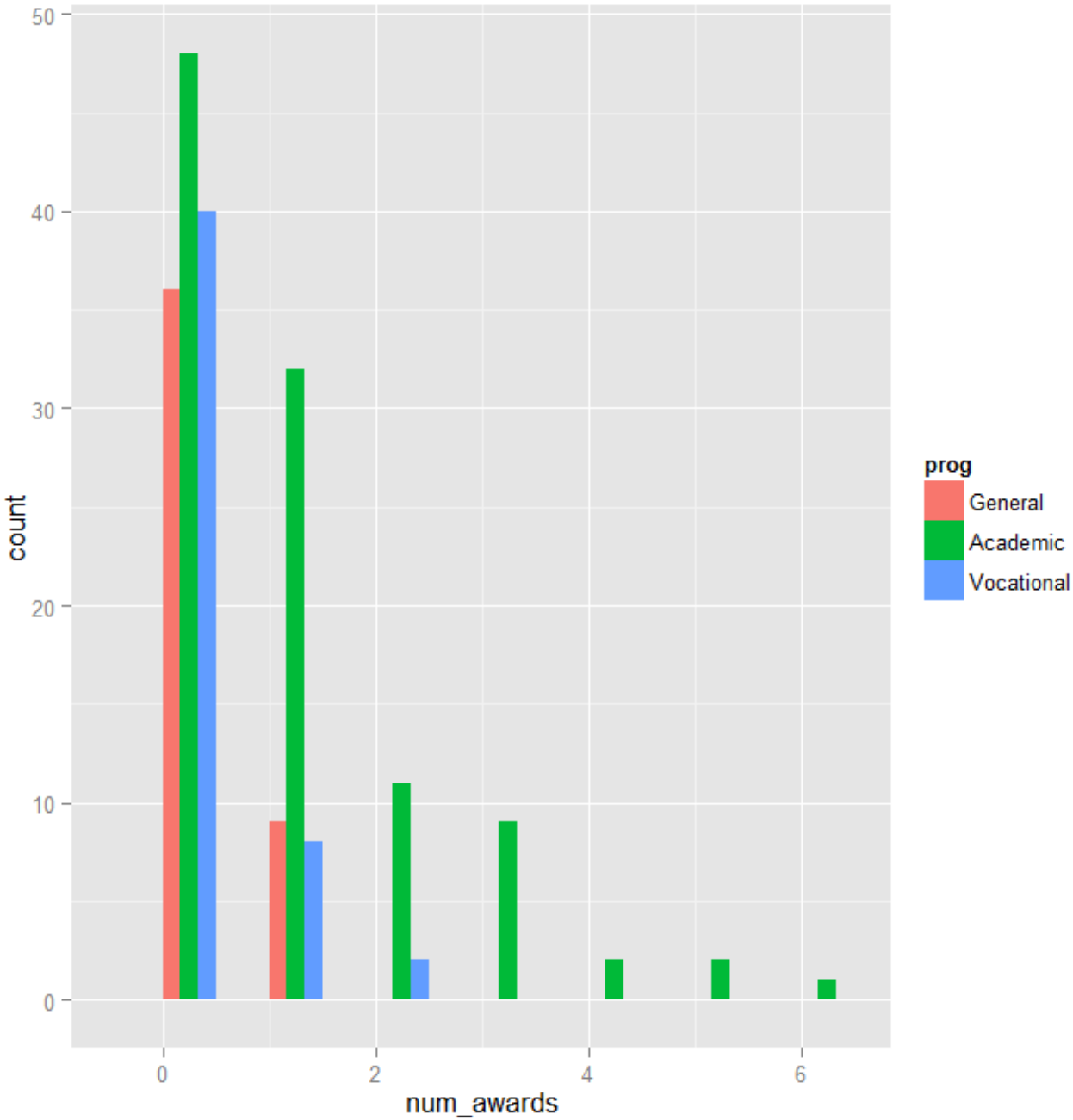
```
       id         num_awards          prog          math      
 1     :  1   Min.   :0.00   General   : 45   Min.   :33.0  
 2     :  1   1st Qu.:0.00   Academic  :105   1st Qu.:45.0  
 3     :  1   Median :0.00   Vocational: 50   Median :52.0  
 4     :  1   Mean   :0.63                    Mean   :52.6  
 5     :  1   3rd Qu.:1.00                    3rd Qu.:59.0  
 6     :  1   Max.   :6.00                    Max.   :75.0  
 (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

- 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.


50
40
30
20
10
0
count
0
2
4
6
num_awards
prog
General
Academic
Vocational