PART 1: Poisson Regression

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

Examples of Poisson regression

- The number of awards earned by students at a secondary or 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.

Conventional 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.

Description of the data

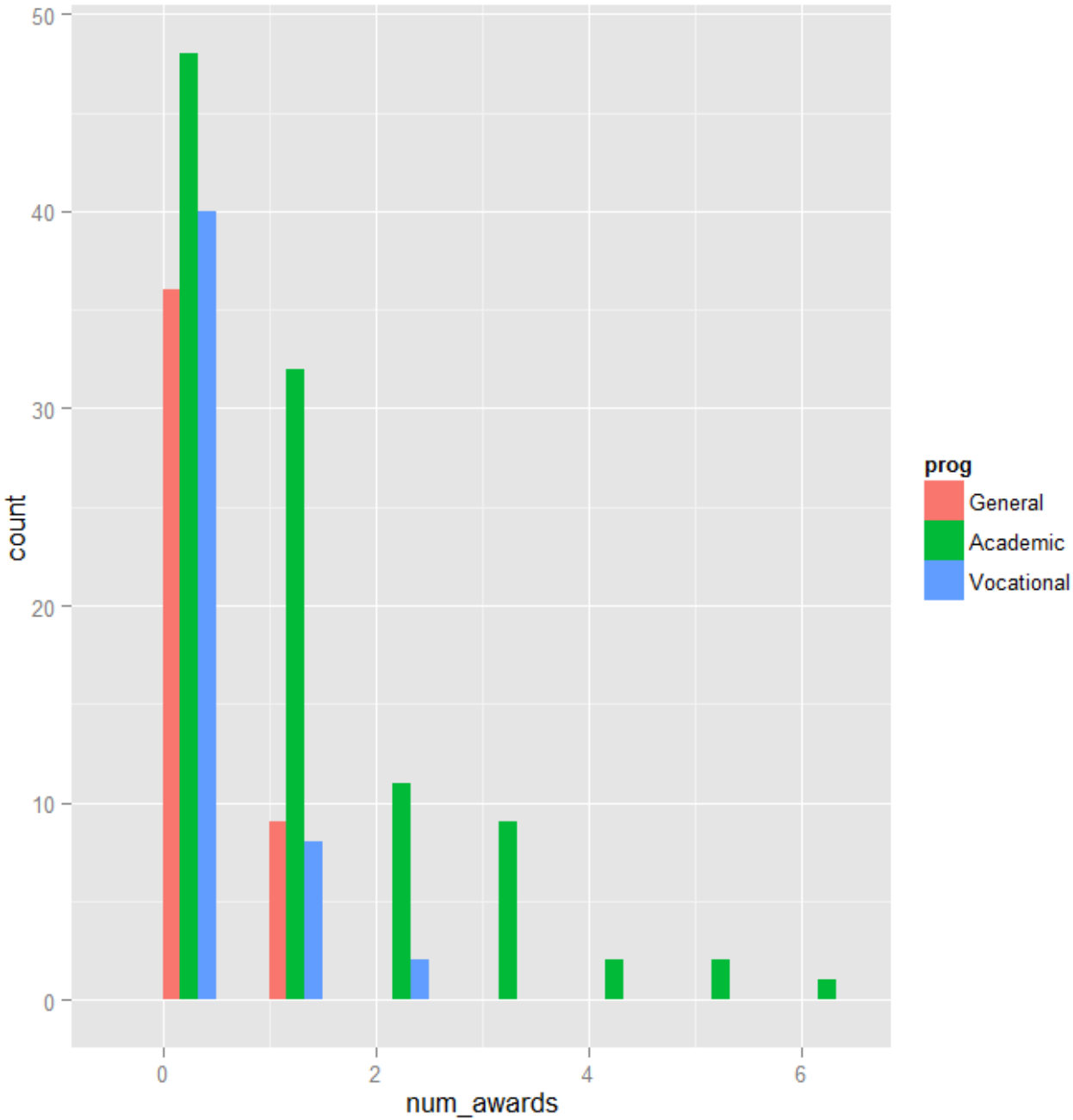
- For the purpose of illustration, we have simulated a data set for the last example.
- The data set is called *poisreg.csv*
- In this example, **num_awards** is the outcome variable and indicates the number of awards earned by students at a high school in a year.

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 
```


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

Poisson Regression with R

- Each variable has 200 valid observations and their distributions seem quite reasonable.
- The mean and variance of our outcome variable are more or less the same.
- Our model assumes that these values, conditioned on the predictor variables, will be equal (or at least roughly so).