

## **PART 2: Negative Binomial Regression**

Negative binomial regression is for modeling over-dispersed count outcome variables.

## Negative Binomial regression with R

- ▶ 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 ( $r$ ) to model the over-dispersion.

$$\Pr(X = k) = \binom{k + r - 1}{k} p^k (1 - p)^r \quad \text{for } k = 0, 1, 2, \dots$$

# Negative Binomial Regression with R

## Examples of negative binomial regression

- ▶ **Example 1** School administrators study the attendance behavior of high school juniors at two schools. Predictors of the number of days of absence include the type of program in which the student is enrolled and a standardized test in math.
- ▶ **Example 2** A health-related researcher is studying the number of hospital visits in past 12 months by senior citizens in a community based on the characteristics of the individuals and the types of health plans under which each one is covered.

## Negative Binomial Regression with R

**Description of the data** Let's pursue Example 1 from above.

- ▶ We have attendance data on 314 high school juniors from two urban high schools in the file **negbin.csv** .
- ▶ The response variable of interest is days absent, **daysabs**.
- ▶ The variable **math** gives the standardized math score for each student.
- ▶ The variable **prog** is a three-level nominal variable indicating the type of instructional program in which the student is enrolled.

# Negative Binomial Regression with R

## Exploratory Data Analysis

```
summary(dat)
```

	id	gender	math	daysabs
1001	: 1	female:160	Min. : 1.0	Min. : 0.00
1002	: 1	male :154	1st Qu.:28.0	1st Qu.: 1.00
1003	: 1		Median :48.0	Median : 4.00
1004	: 1		Mean :48.3	Mean : 5.96
1005	: 1		3rd Qu.:70.0	3rd Qu.: 8.00
1006	: 1		Max. :99.0	Max. :35.00

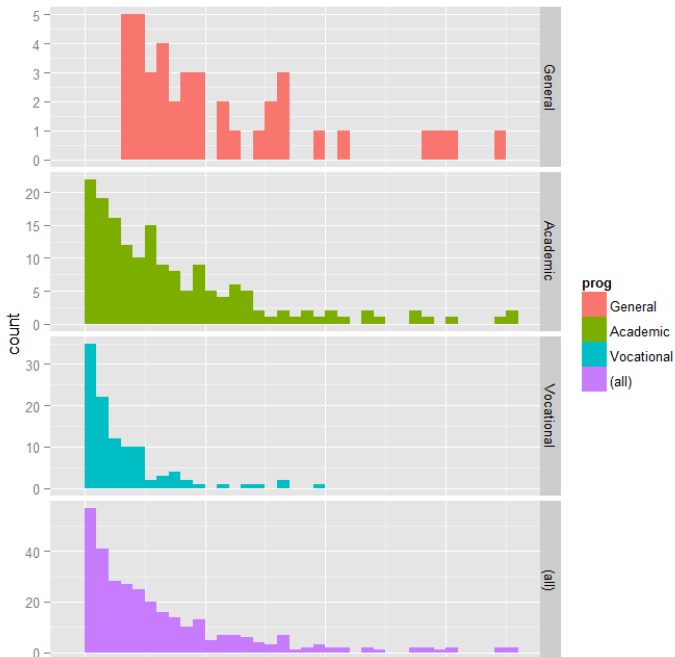
```
(Other):308
```

```
prog
```

```
General : 40
```

```
Academic :167
```

```
Vocational:107
```



## Negative Binomial Regression with R

- ▶ Each variable has 314 valid observations and their distributions seem quite reasonable.
- ▶ The unconditional mean of our outcome variable is much lower than its variance.