Over-Dispersion

- Overdispersion is the presence of greater variability in a data set than would be expected based on a given simple statistical model.
- Poisson Distribution:

$$\mathrm{Var}(X) > \mathrm{E}(X)$$

Zero-Inflation

- One common cause of over-dispersion is excess zeros, which in turn are generated by an additional data generating process.
- In this situation, zero-inflated model should be considered.
- If the data generating process does not allow for any 0s (such as the number of days spent in the hospital), then a zero-truncated model may be more appropriate.

Over-Dispersion

- When there seems to be an issue of dispersion, we should first check if our model is appropriately specified, such as omitted variables and functional forms.
- For example, if we omitted the predictor variable prog in the example above, our model would seem to have a problem with over-dispersion.
- In other words, a misspecified model could present a symptom like an over-dispersion problem.

Poisson Regression with R

- Assuming that the model is correctly specified, the assumption that the conditional variance is equal to the conditional mean should be checked.
- There are several tests including the likelihood ratio test of over-dispersion parameter alpha by running the same model using negative binomial distribution.
- The R package **pscl** (Political Science Computational Laboratory, Stanford University) provides many functions for binomial and count data including `odTest` for testing over-dispersion.

The odTest() command (pscl package)

- The function odTest is a **likelihood ratio test** for over-dispersion in count data
- odTest() Compares the log-likelihoods of a negative binomial regression model and a Poisson regression model.

```
odTest(glmobj, alpha=.05,
  digits = max(3, getOption("digits") - 3))
```

Exposure Variables - SKIP

- Count data often have an exposure variable, which indicates the number of times the event could have happened.
- This variable should be incorporated into a Poisson model with the use of the `offset` option.
- The outcome variable in a Poisson regression cannot have negative numbers, and the exposure cannot have 0s.

Some Remarks

- Many different measures of pseudo-R-squared exist. They all attempt to provide information similar to that provided by R-squared in OLS regression, even though none of them can be interpreted exactly as R-squared in OLS regression is interpreted.
- Poisson regression is estimated via maximum likelihood estimation. It usually requires a large sample size.