

# Introduction to Poisson Regression with Robust Standard Errors - Part 4

---

Christopher Liu, Siyu Zhao, Zheming Yu, Runnan Liu, Shiming Zhang  
11/29/2018

San Diego State University, Stats 610

# Objectives of the Video

- Discuss the motivation for the Poisson Regression model with Robust Standard Errors
- Understand the difference between this model and the "regular" Poisson Regression model
- Understand the interpretation of the model's parameters

Our focus, however, is to discuss **Poisson Regression with Robust Standard Errors**

- Modified Poisson Regression that can work with response variables with binary outcomes
- Addresses problems with overdispersion

# Poisson Regression with Robust Standard Errors

Main problem is with the Poisson assumption of

$$E(Y_i) = \text{Var}(Y_i)$$

- With binomial data, Poisson regression usually underestimates variance of data
- We need a way to address this problem when making inferences

# Poisson Regression with Robust Standard Errors

The only adjustment to make:

- The standard errors of the estimated predictors ( $\hat{\beta}$ ) are replaced with "robust" standard errors from the sandwich estimator

$$\text{Var}(\hat{\beta}) = (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1}$$

where  $\Sigma$  is the covariance matrix of the residuals

# Poisson Regression with Robust Standard Errors

- Simple to implement through software such as R with the "sandwich" library
- Everything else is kept the same, such as the values of  $\hat{\beta}$  (the maximum likelihood estimates of  $\beta$ ) as well as their interpretations

# Regression for Data with Binary Outcomes

Multiple methods for regression of data with binary outcomes

- Logistic Regression is a popular alternative

Many real-world scenarios can be modeled with binary regression

- Health-related problems such as association of disease and certain factors
- Widely used methods in Epidemiology and Public Health

## Pima Indians Diabetes dataset

- Contains health measurements for subjects and their status of being diabetic or not  
→ Diabetes status is the outcome variable
- Originally from the University of California, Irvine Machine Learning Repository. Still available through Kaggle