

Bios 312: Modern Regression Analysis
March 15, 2011
Lab 5: Poisson Regression

In this lab, we will use the Student Attendance data available from the course web site. School administrators are interested in studying the attendance behavior of high school juniors at two schools. Predictors of the number of days absent include: gender of the student and standardized test scores in math and language arts. Days of attendance was also available

1. Create a new variable for the total number of days at school. I will refer to this variable as 'total'. Use this variable as the offset in the following Poisson regression models. Remember to use a log transformation of total, if appropriate for the software you are using.
2. Use appropriate descriptive statistics to summarize the variables: daysabs, total, math, langarts, and male. Include the mean and variance in your summary statistics.
3. Create a histogram of daysabs. Describe the plot
4. Create a histogram of daysabs for females and males. Describe and compare the plots.
5. Fit a classical Poisson regression model using daysabs as the outcome and male as the predictor (without robust errors).
 - a. Provide full statistical inference for the output
 - b. One underlying assumption of the Poisson model is that mean equals the variance. Can the summary statistics created in question 2 be used to determine if the mean equals the variance in this model? Why or why not?
6. Fit a Poisson regression model with robust standard errors using daysabs as the outcome and male as the predictor. Interpret the output and compare to what was obtained using classical Poisson regression.
7. Fit a Poisson regression model with robust standard errors using daysabs as the outcome with covariates male and math.
 - a. Provide full statistical inference for the output.
 - b. Is there evidence that the association between daysabs and male is confounded by math? Explain
8. Fit a Poisson regression model with robust standard errors using daysabs as the outcome with covariates male and langarts.

- a. Provide full statistical inference for the output.
 - b. Is there evidence that the association between daysabs and male is confounded by langarts?
- 9. Fit a Poisson regression model with robust standard errors using daysabs as the outcome with covariates male, math, and langarts.
 - a. Provide full statistical inference for the output.
 - b. Is there evidence of confounding?
- 10. How does the scientific interpretation of male differ in models 6, 7, 8, and 9?
- 11. Finally, we will predict the number of days absent as a function of language arts score and gender using model 9. Note that model 9 also includes math score as a covariate, so we will need to fix the math score at some value.
 - a. Predict the number of days absent out of 100 days for males with a math score of 48.75 and a language arts score that ranges from 0 to 100
 - b. Predict the number of days absent out of 100 days for females with a math score of 48.75 and a language arts score that ranges from 0 to 100
 - c. Plot the predicted number of days absent out of 100 days versus the language arts score for male and female (holding math scores constant at 48.75)

Notes for R Users:

There are many methods for fitting Poisson regression with robust standard errors in R. One approach uses functions in the libraries 'lmtest' and 'sandwich'. The first time you use these packages, you will need to install (e.g. `install.packages("lmtest")`) the package; subsequently, you will need to load the package each time you begin R (e.g. `library("lmtest")`). After these libraries are loaded, the following command will fit the unadjusted model for male with robust standard errors.

```
m2 <- glm(daysabs ~ male, family="poisson", data=p)
coeftest(m2, vcov=sandwich)
```