**PSY 653 Module 10: Analyses Involving Categorical Dependent Variables**

**Gemma Wallace & Neil Yetz**

**April 15, 2020**

*Description of the datasets for the demo activity:*

This is a simulated dataset with 164 observations of 5 variables.

* **Y:** A binary variable (Coded as 0 or 1).
* **X1**: A binary variable (Coded as 0 or 1)
* **X2:** A continuous variable ranging from 0 to 10
* **X3:** A continuous variable ranging from 0 to 5
* **X4:** A continuous variable ranging from 0 to 4

**Demo Activity**

1. Create a new R notebook and load the following libraries: tidyverse, psych, olsrr
2. Read in the datafile “Logistic2.csv”.
3. Use ordinary least squares multiple regression to regress Y on all X variables
   1. Describe what the values of each of the regression coefficients tells you
   2. How well can you predict Y?
4. Use logistic regression to regress Y on all X variables
   1. Describe what the values of each of the regression coefficients tells you
   2. How do you know whether or not the model fits?
   3. How well can you predict Y?

**Try it Yourself Activity**

Description of the dataset (courtesy of Dr. Kim Henry):

*A research team sought to examine factors associated with 21st birthday drinking among female students at a large University. Female students who were nearing age 21 and self-classified as regular drinkers were eligible for the study. In total, 200 students were recruited and agreed to take part in the study. Students were instructed to report to the lab two weeks prior to their 21st birthday. During this lab session, students completed a brief survey that measured alcohol use during the past month (using the Timeline Follow Back Method) and their weight was recorded. One week prior to their 21st birthday, participants were sent a link for an online survey to measure positive alcohol expectancies for drinking on their 21st birthday. Within three days prior to their 21st birthday, students reported to the lab and were given a diary-based data collection form to record several items on their 21st birthday. Students were instructed to record the food that they consumed during the day, the de- gree to which they were in a partying mood just prior to the celebration, and the quantity and type of drinks that they con- sumed during the first two hours of the celebration. The students were also given a small breathalyzer machine to measure BAC 2 hours after consumption of their first drink.*

The dataset called bac\_module10.csv contains the following variables:

* **weight:** weight in kilograms
* **weight\_low:** participants with body weight lower than 1 standard deviation below the mean were coded as , and participants with body weight above 1 standard deviation below the mean were coded as 0
* **alcexp:** positive alcohol expectancy for drinking on the impending 21st birthday, a multi-item scale that ranges from 1-7, where a higher score indicates more positive expectations about the role alcohol will play
* **typ\_drks:** the number of standard alcohol drinks consumed in the past 30 days
* pmood: a rating on a scale from 1-9 on the respondent’s mood to party on the 21st birthday, where 1 means never been less in the mood to party, and 9 means never been more in the mood to party
* **absorb:** a score calculated from the food diaries to determine how full the participant was when they began drinking, the score ranges from 1 to 8, where 1 means a completely full stomach, and 8 means a completely empty stomach
* **alc\_gm:** a score calculated from the drinking diary to estimate the grams of alcohol consumed on the 21st birthday
* **bac:** the participant’s blood alcohol content, measured as grams of alcohol per deciliter of blood on the 21st birthday
* **bac\_over:** participants were coded as 1 if their measured blood alcohol content was > 0.08 and coded as 0 if their measured BAC was < 0.08.

1. Read in the datafile “bac\_module10.csv”.
2. Use ordinary least squares multiple regression to regress bac\_over on alcexp, pmood, weight\_low, and typ\_drks
   1. Describe what the values of each of the regression coefficients tells you
   2. How well can you predict bac\_over?
3. Use logistic regression to regress to regress bac\_over on alcexp, pmood, weight\_low, and typ\_drks
   1. Describe what the values of each of the regression coefficients tells you
   2. How do you know whether or not the model fits?
   3. How well can you predict bac\_over?