Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Math 227 / Fall 2019 / Prof. David Soto

**Module 4. Activity 1.**

##### The data set FatMice18 contains the body mass gain in mice after 4 weeks for two of the treatment conditions: a normal light/dark cycle (LD) or a bright light on at night (LL).

* Light Light treatment: LD= normal light/dark cycle or LL=bright light at night
* WgtGain4 Weight gain (grams over a four week period)

##### Source: Fonken, L., et. al., "Light at night increases body mass by shifting time of food intake," Proceedings of the National Academy of Sciences, October 26, 2010; 107(43): 18664-18669.

**Question 1.**

a) Generate histograms for the weight gain under the two different light conditions. What is the R code you use to generate this?

b) Generate boxplots for the weight gain under the two different light conditions. What is the R code you use to generate this?

c) Based on your visualization from parts (a) and (b), would you say that light exposure at night explains the weight gain of the mice? Explain.

d) Run the instruction **gf\_point(WgtGain4 ~ Light, data = FatMice18)**

Run the instruction **gf\_jitter(WgtGain4 ~ Light, data = FatMice18, width = .1)**

What is the difference between the two plots?

e) Write the statement “variation in weight gain is explained by variation in light exposure at nightplus variation in other stuff” using a word equation as described in the book:

**Question 2. Data frame: NutritionStudy**

The instruction **NutritionStudy$Fiber3group <-ntile(NutritionStudy$Fiber, 3)**  classifies all cases into three different levels of Fiber consumption. One third of all cases, those with the lowest consumption of fiber, will be assigned a “1”. The second third, those with a moderate consumption of fiber, will be assigned a “2”. The last third of the cases, those with a high consumption of fiber, will be assigned a “3”. These new values will be saved in a new variable **Fiber3group.**

a) What are the values of the variable **Fiber3group** for the first six cases of **NutritionStudy**?

b) Generate histograms of Cholesterol under the three different levels of fiber consumption. What is the R code you use to generate this?

c) Do you think that the variation in fiber consumption explains the variation in cholesterol for this sample of individuals? Explain.

d) Generate histograms of Cholesterol under two different levels of fat consumption. What is the R code you use to generate this?

e) Do you think that the variation in fat consumption explains the variation in cholesterol for this sample of individuals? Explain.

**Question 3. Data frame: CocaineTreatment**

##### Source: Gawin, F., et.al., "Desipramine Facilitation of Initial Cocaine Abstinence", Archives of General Psychiatry, 1989; 46(2): 117 - 121.

a) Generate an appropriate visual representation to evaluate if the variation in the variable **Drug** explains the variation in the variable **Relapse**. What is your choice for the type of graph? Why?

b) Would you say that, based on the sample data, one drug is more effective than the others to treat cocaine withdraw? Explain.

**Question 4**.

Note: Type I error is when we conclude that some variable we manipulated had an effect when in fact the observed difference was simply due to random sampling variation.

The following instruction creates a vector with two values: “1” and “2”

**RandomGroupValues <- c(1,2)**

The following instruction randomly selects “1” or “2” and it creates a list of 315 values. This list is assigned to a new variable **RandomLabel** in the data frame **NutritionStudy**. (That is, each case is randomly “assigned” to be a member of a category “1” or “2”).

**NutritionStudy$RandomLabel <- resample(RandomGroupValues,315)**

a) What does the following instruction accomplish?

**gf\_histogram(~Cholesterol, data = NutritionStudy) %>% gf\_facet\_grid(RandomLabel~.)**

b) Would you say that variation in **RandomLabel** explains the variation in **Cholesterol**?

c) Keep clicking “run” to generate plot, after plot of **Cholesterol** and **RandomLabel**. Try to look up for a plot that could indicate that **RandomLabel** explains **Cholesterol**. If you were to draw such conclusion, what type of error are you committing? Explain.