**Homework 1 – Descriptive Statistics  
Due: February 7 as hard copy**

In this assignment, you will be analyzing data related to nutrition and health of 317 individuals. The data is available in the accompanying Excel file NutritionStudy.xlsx. A description of the dataset is available in the accompanying file NutritionStudyInformation.docx

You will need to import the Excel file into RStudio. To import Excel data, first save the Excel file to your local drive. Then, in RStudio, go **to File->Import Dataset->from Excel**. RStudio will tell you that you need to install a library. Accept this suggestion, and then wait until the next prompt. Enter the name of your file, and click **Import*.*** You will see this data in the top left window of RStudio, just like we viewed the built-in datasets. (If this doesn’t work, you may have to try again, possibly closing RStudio and reopening it.)

1. Create a pie chart of patients’ smoking status. Give the categories in your pie chart descriptive names, and display the percentages of each type of smoker next to the category names. Make sure to give your chart an appropriate title.

Submit:  
- The commands you used to create your pie chart, in order  
- The pie chart

1. Create a scatterplot of grams of fat consumed per day (on the y-axis) vs number of calories consumed per day. Give your scatterplot an appropriate title and axis labels, including units.

Submit:  
- The commands you used to create your scatterplot, in order  
- The scatterplot  
- Answers to the following questions:

1. Are there any outliers in the dataset? Give the number of calories consumed by the outlier(s).
2. Suppose a person consumes 2000 calories per day. How many grams of fat do you expect this person to consume per day? Give your answer in two forms: i) as a single number (this is called a *point estimate*), and ii) as a range along the lines of “it seems pretty certain that a person who consumes 2000 calories per day consumes between \_\_\_ and \_\_\_ milligrams of fat per day” (this is called an *interval estimate*). You don’t have to use any fancy commands to get a precise range for part ii) (we will learn how to do this later in the course) – just get your numbers from the graph and explain briefly how you got them.
3. Create side-by-side boxplots of weekly alcohol consumption grouped by sex, **only for the people in the study who consume alcohol.** If there are any values that look **very** unusual (hint: there are), then redo your boxplots without the value(s). Give your graph a descriptive title and descriptive axis labels, and include units.

Submit:

* The commands you used to create your boxplots, in order
* The boxplots
* An answer to the following question: How do the weekly alcohol consumptions of male and female drinkers compare? Your answer should contain at least one percentage and proportion, as well as a reference to the numbers in your boxplot.

1. Use only the results of the **favstats** command to answer the following questions:
   1. How do the calorie consumptions of men and women compare? Be sure to reference both a measure of centre and a measure of variation in your answer.
   2. How do the fiber consumptions of men and women compare? Be sure to reference both a measure of centre and a measure of variation in your answer.

Submit:

* Your **favstats** commands
* Outputs of your commands
* Answers to a. and b.

1. Consider now the BMI data. We saw two ways to define outliers in class. In this question, you will find the number of outliers in this dataset using both methods.

For **each** method, submit:

* The commands you used to count the number of outliers
* The output of your commands, which should provide the number of outliers

1. Now we will compare the distribution of fat consumption of men and women.

Submit:

* The commands and output you used to compare the two distributions
* An answer to the following questions: Based on your results, is the fat consumption for men symmetric, skewed left, or skewed right? How about for women?

**To submit: answers to all questions, including graphs, as a hard copy. Be sure to include your full name, your A0 number, and your set letter on the front page of your assignment.**