# Stat 20 Section Worksheet 2: 8/30/2018

## Problems from FPP, Chapter 2, and one not from there:

1. (A.13) (Simpson's paradox) A hypothetical university has 2 departments, A and B. There are 2,000 male applicants, of which half apply to each department. There are 1,100 female applicants of which 100 apply to department A and the rest to department B. Department A admits 60% of the men and 60% of the women. Department B admits 30% of the men and 30% of the women. "For each department, the percentage of men admitted equals the percentage of women admitted; this must be so for both departments together." True or false? Explain.
2. Since 2000, the median US wage has risen about 1%, adjusted for inflation, but over the same time period, the median wage for the different educational groups - high school dropouts, high school graduates with no college education, people with some college education, people with college degrees and higher - have fallen. That is, in each of these educational groups, the median wage has decline. How can this be? Explain.
3. (R.8) Ads for ADT Security Systems claim "when you go on vacation, burglars go to work… According to FBI statistics, over 25% of home burglaries occur between Memorial Day and Labor Day." Do the statistics prove that burglars go to work when other people go on vacation? Answer yes or no and please explain your answer.
4. (R.10) A study of young children found that those with more body fat tended to have more "controlling" mothers; the *San Francisco Chronicle* concluded that "Parents of Fat Kids Should Lighten Up."
   A. Was this an observational study or a controlled experiment?
   B. Did the study find an association between a mother's behavior and her child's level of body fat?
   C. If controlling behavior by the mother causes the child to eat more, would this explain the association?
   D. Suppose there was a gene that causes obesity? Would this explain the association?
   E. Can you think of another way to explain the association?
   F. Do the data support the *Chronicle*'s advice on child-rearing?

## Getting Going with R: vectors

Today you are going to play with some of the commands you saw in class.

1. Generate a vector that contains the integers from 1 to 10 and call this vector x
2. Generate another vector that goes from -4 to 0 and call this one y
3. Add x and y: What do you get?
4. Now input the vector (68, 73, 71, 74, 73, 73, 73, 74, 72) into R and call it berk.temp (these are the predicted high temperatures (in Fahrenheit) for Berkeley for the next ___ days. How many days (do not count the numbers, use R!)
5. What will be the average temperature over these days?

6. Now, there are many students in the class who can't quite figure out Fahrenheit since they are more used to the metric system. Convert these temperatures to Celsius, recalling that C=(F-32)/1.8.
7. Let's look at berk.temp again. Sample 4 values from it. What did you get? Did you sample with or without replacement (that is, if these numbers were written on tickets and then put in a box, did you put them back before taking out the next one)?
8. Now let's simulate rolling a fair six-sided die. How would you simulate **one** die roll? Hint: first create a vector. What are its elements? Now how would you use this vector to simulate a die roll? What function would you use? What are this function's possible inputs? What did you put in?
9. Let's go further. Simulate a roll of two dice (for example, if you are playing monopoly, you would roll a pair of dice). How can we do that? Now how would you add the values of the rolls? That is, you want a **compound** function.
10. Suppose now we want to simulate drawing **one** ticket from a box with **five** tickets marked 1, 3, 5, 15, 27. How would you do this using R since we don't actually have a box?
11. What about if you wanted to simulate drawing 2 tickets WITH replacement? That is, you put the first ticket back before drawing the next?
12. Write down the result of drawing 10 tickets "at random" from this box. "At random" means that all the tickets are equally likely.
13. Back to coin tosses. We can simulate a coin toss by creating a vector using c("H", "T"). (Remember if you are using objects that are not numbers, you have to put them in quotes.) This is nice, but not as useful, since we usually want to count the number of heads or the number of tails. So simulate a coin toss, marking 1 for heads and 0 for tails.
14. Now simulate the result of 10 flips of a fair coin, and count the number of heads. How many heads do you expect to see? How many do you actually see? Do it again.
15. Repeat this exercise, but flip the coin 100 times and count the number of heads in these 100 tosses. (You can look up the lecture notes if you get stuck.)
16. Download the dataset family.csv from bcourses -> Files -> Data. Save it to a folder on your computer. Load the file into RStudio (Chromebook users will first have to upload the file using the bottom right panel tab): Environment tab -> Import Dataset -> From Text (readr). A new window should pop up where you can click on Browse and you can select the file from where you have saved the data file.
17. What is the average age of this family, and their average weight? How many family members are there?
18. Convert the height to meters and their weight to kgs and check that the numbers in the vmi column are correct.