# STAT22000, Autumn 2013 Homework 3

All page, section, and exercise numbers below are for the course text (Moore, McCabe and Craig, Introduction to the Practice of Statistics, 7th edition).

**Reading**: Section 2.3, 2.4, 3.0 (Introduction), 3.1, 3.2
**Problems for Self-Study**: (Do Not Turn In. Solutions are at the end of the textbook.)

1. Exercise 2.101 on p.134
2. Exercise 3.27 on p.185
3. Exercise 3.53 on p.198
4. Exercise 3.73 on p.201

**Problems to Turn In**: due Wednesday, Oct. 23, in class
**Grading: 50 total points**

1. [Revision of Exercise 2.84 on p.121] The mean height of American women in their early twenties is about 64.5 inches and the standard deviation is about 2.5 inches. The mean height of American men at the same age is about 68.5 inches with standard deviation about 2.7 inches. Suppose the correlation between heights of husbands and wives is about $r = 0.5$.

   (a) What is the equation of the regression line to predict the husband's height from the wife's height in young couples?

   (b) Predict the height of the husband of a women who is 67 inches tall.

   (c) What is the equation of the regression line to predict the wife's height from the husband's height in young couples?

   (d) Manually draw a graph of both regression lines in part (a) and (c). Give the coordinate of the point that the two lines intersect. For each of the two lines, give the coordinate of another point on it.

2. **Exercise 2.110 on p. 135**
   Drilling down beneath a lake in Alaska yields chemical evidence of past changes in climate. Biological silicon, left by the skeletons of single-celled creatures called *diatoms* measures the abundance of life in the lake. A rather complex variable based on the ratio of certain isotopes relative to ocean water gives an indirect measure of moisture, mostly from snow. As we drill down, we look farther into the past. Here are data from 2300 to 12,000 years ago:

   | Isotope (%) | Silicon (mg/g) | Isotope (%) | Silicon (mg/g) | Isotope (%) | Silicon (mg/g) |
   |---|---|---|---|---|---|
   | $-19.90$ | 97 | $-20.71$ | 154 | $-21.63$ | 224 |
   | $-19.84$ | 106 | $-20.80$ | 265 | $-21.63$ | 237 |
   | $-19.46$ | 118 | $-20.86$ | 267 | $-21.19$ | 188 |
   | $-20.20$ | 141 | $-21.28$ | 296 | $-19.37$ | 337 |

   (a) Make a scatter plot of silicon (response) against isotope (explanatory). Ignoring the outlier, describe the direction, form, and strength of the relationship. The researchers say that this and the relationships among other variables they measured are evidence for cyclic changes in climate that are linked to changes in the sun's activity.

(b) The researchers single out one point: "The open circle in the plot is an outlier that was excluded in the correlation analysis." Circle this outlier on your graph. What is the correlation with and without this point? The point strongly influences the correlation.

(c) Is the outlier also strongly influential for the regression line? Calculate and draw on your graph two regression lines, and discuss what you see.

3. **Exercise 3.18 on p.184**
Explain what is wrong with each of the following randomization procedures and describe how you would do the randomization correctly.

(a) Twenty students are to be used to evaluate a new treatment. Ten men are assigned to receive the treatment and 10 women are assigned to be the controls.

(b) Ten subjects are to be assigned to two treatments, 5 to each. For each subject, a coin is tossed. If the coin comes up heads, the subject is assigned to the first treatment; if the coin comes up tails, the subject is assigned to the second treatment.

(c) An experiment will assign 40 rats to four different treatment conditions. The rats arrive from the supplier in batches of 10 and the treatment lasts two weeks. The first batch of 10 rats is randomly assigned to one of the four treatments, and data for these rats are collected. After a one-week break, another batch of 10 rats arrives and is assigned to one of the three remaining treatments. The process continues until the last batch of rats is given the treatment that has not been assigned to the three previous batches.

4. **Exercise 3.52 on p.198**
In each of the following situations, describe the sample as an SRS, a stratified random sample, a multistage sample, or a voluntary response sample. Explain your answers

(a) There are seven sections of an introductory statistics course. A random sample of three sections is chosen and then random samples of 8 students from each of these sections are chosen.

(b) A student organization has 55 members. A table of random numbers is used to select a sample of 5.

(c) An online poll asks people who visit this site to choose their favorite television show.

(d) Separate random samples of male and female first-year college students in an introductory psychology are selected to receive a one-week alternate instructional method.

5. A survey is carried out by the finance department to determine the distribution of household size in a certain city. They draw a simple random sample of 1,000 households. After several visits, the interviewers find people at home in only 653 of the sample households. Rather than face such a high non-response rate, the department draws a second batch of households, and uses the first 347 completed interviews in the second batch to bring the sample up to its planned strength of 1,000 households. The department counts 3,087 people in these 1,000 households, and estimates the average household size in the city to be about 3.1 persons. Is this estimate likely to be too low, too high, or about right? Why?