Reading: Section 1.3, 1.4, 2.1, 2.2

Problems for Self-Study: (Do Not Turn In)

- Exercise 1.9, 1.11, 1.13, 1.21, 1.27, on p.58-62 and Exercise 2.1, 2.3, 2.5, 2.7, 2.13, 2.15, 2.17 on p.116-120. Answers can be found at the end of the book (p.405-409).

Problems to Turn In: due 3 pm on Friday, Oct. 20, on Canvas.

1. Exercise 1.12 on p.58

2. In each of the following situations, identify the sampling method as one of the following: simple random sampling, stratified sampling, multistage sampling, or voluntary response sampling.
   (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) An online poll asks people who visit this site to choose their favorite television show.
   (c) 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.

3. 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 low 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?

4. Suppose you are on the staff of a member of Congress who is considering a bill that would provide government-sponsored insurance for nursing-home care. You report that 1128 letters have been received on the issue, of which 871 oppose the legislation. “I’m surprised that most of my constituents oppose the bill. I thought it would be quite popular,” says the congresswoman. Are you convinced that a majority of the voters oppose the bill? How would you explain the statistical issue to the congresswoman?

5. During the Second World War, the U.S. military collected data on bullet holes found in B-24 bombers that returned from flight missions. The data showed that most bullet holes were found in the wings and tail of the aircraft. Therefore, the military reasoned that more armor should be added to these regions, as they are more likely to be shot.
   (a) Identify the population of interest and the sample in this study.
   (b) Identify the population we can generalize to given the data. Hint: Were there some kind of planes that were left out by the data collection method?
   (c) Abraham Wald, a famous statistician of the era, is reported to have argued against this reasoning. In fact, he argued that based on these data more armor should be added to the center of the plane, and NOT the wings and tail. What was Wald’s argument?

6. Exercise 2.10 on p. 117

7. Exercise 2.16 on p. 119
8. Exercise 2.18 on p. 120

9. This problem is just Lab 4: http://www.stat.uchicago.edu/~yibi/s220/labs/lab04.html

Suppose a fair coin is tossed a huge number of times. Let $H(n)$ be the number of heads that occur in $n$ tosses. By the law of large numbers, we expect heads to occur in about half of the tosses, i.e., $H(n) \approx \frac{n}{2}$.

In this problem, we are going to simulate coin tossing in R, and see if this is true.

(a) Simulate tossing a fair coin 10000 times. Compute the value of $H(n)$ for $n = 1, 2, \ldots, 10000$ for these 10000 tosses. Make two plots, one with $H(n) - \frac{n}{2}$ vs. $n$, and the other with $\frac{H(n)}{n} - 0.5$ vs. $n$.

(b) Repeat the simulation in part (a) three times, and include those additional six plots (two plots per simulation). Since the outcomes of the tosses are different, the $H(n)$ and the two plots for the three simulations will be different.

(c) From the 8 plots made in part (a) and (b), as $n$ gets large, does $H(n) - \frac{n}{2}$ seem to approach 0, and does $\frac{H(n)}{n} - 0.5$ seem to approach 0? What’s the right statement of the law of large numbers for coin tossing? If you are not sure, you can make more simulations and/or increase the number of tosses in each simulation.

(d) Now let's check the law of large numbers for tossing an UNFAIR coin. Simulate tossing an unfair coin with only 0.2 probability to land heads 10000 times, and compute the values of $H(n)$. This time we expect $H(n) \approx 0.2n$, so please plot $H(n) - 0.2n$ vs. $n$, and plot $\frac{H(n)}{n} - 0.2$ vs. $n$. Repeat this simulation 4 times (so 8 plots in total). As $n$ gets large, does $H(n) - 0.2n$ seem to approach 0, and does $\frac{H(n)}{n} - 0.2$ seem to approach 0? What’s the right statement of the law of large numbers when tossing such an unfair coin?

10. You and your arch nemesis play games by tossing a fair coin. For each of the three games below, you can choose either to toss the coin 10 times or 100 times. Identify how many coin tosses (10 tosses or 100 tosses) you should choose for each game that is more beneficial for you. Explain your reasoning.

(a) If the number of heads $H$ is within 2 of half the number of tosses (i.e. $N/2 - 2 \leq H \leq N/2 + 2$, where $N$ is the number of tosses), your arch nemesis pays you $1. Otherwise you pay him $1.

(b) If heads come up between 40% and 60% of the time, your arch nemesis pays you $1. Otherwise you pay him $1.

(c) If heads come up more than 60% of the time, your arch nemesis pays you $1. Otherwise you pay him $1.

Hint: Think about what you observed in problem #9.