Homework policies: You should give a brief and concise explanation for each question. Just writing down an answer (e.g., “SD = 3.3”) with no explanation is not sufficient.

Homework is due at the beginning of class on the due date. Late homework will not be accepted, with one exception: you may hand in one assignment late (by the beginning of the next class) once during the semester. If you are planning to hand in a homework late, you must email me by the beginning of class on the original due date.

Collaboration: I encourage you to discuss homework problems with other students (and with me, of course). Your final answer should be written in your own words. If you work with other students, list their names on your paper.

(1) Review of Part III (p. 331), questions 9, 13 (medical reports). Don’t answer the questions in the book; instead, say whether each example is an observational study or an experiment, and briefly explain why. [9, 13 in the old edition, p. 339]

(2) Review of Part III (p. 331), question 4 (radon). Don’t answer the questions in the book; instead, identify the population, sample, parameter, and statistic. [4, p. 339]

(3) chapter 14, question 19 (car repairs) [13]

(4) chapter 14, question 33 (disjoint or independent) [27]

(5) chapter 14, question 41 (tires) [35]

(6) A general can plan a campaign to fight one major battle or three small battles. He believes that he has probability .56 of winning the large battle and probability .78 of winning each of the small battles. Assume that victories or defeats in the small battles are independent. The general must win either the large battle or all three small battles to win the campaign. Which strategy should he choose?

(7) Suppose a drug test is 97% accurate. All 150 employees at a company are made to take the drug test. If none of the 150 employees are in fact using drugs, what is the probability that at least one employee will be wrongly accused as a result of drug testing?

(8) chapter 15, question 46 (polygraphs) Hint: assume there are 10000 applicants and make a diagram like the one we made in class. Out of 10000 applicants, on average, how many will be identified as liars? Of these, how many are actually telling the truth? The correct answer is .81.) [42]

(8a) trivia question (optional; not for credit): The inventor of the modern polygraph also created which comic book superhero? How is the polygraph connected to one of this superhero’s special tools?

(9–14) The following questions deal with the computer simulation from the class of Wednesday, 2/16. Our goal was to estimate the number of people in a large class without having to count each person individually.

Notation: Let \( N \) represent the number of people in the class. Let \( n \) represent the number of people who were randomly sampled.

We compared three methods of estimating \( N \). Depending on which section you’re in, these may have included the following methods (the numbers may not match what we called them in class):

Method 1: Take the average of the \( n \) numbers in the sample and double it

Method 2: Take the average of the \( n \) numbers and add 2 SD’s

Method 3: Multiply the largest of the \( n \) numbers in the sample by \( \frac{n+1}{n} \)

(over)
We used the computer to take samples of size \( n = 5 \) from a hypothetical large class with size \( N = 100 \), and recorded the estimates from each method. The sampling distributions are shown below:

### Method 1
- Mean of sampling distribution: 99.8 people
- SD of sampling distribution: 25.4 people

### Method 2
- Mean of sampling distribution: 106.7 people
- SD of sampling distribution: 21.4 people

### Method 3
- Mean of sampling distribution: 99.8 people
- SD of sampling distribution: 17.0 people

(9) In this specific context of estimating \( N \), explain what a sampling distribution is.
Why is it important to look at the sampling distributions in comparing the three methods?

(10) I claimed in class that method 3 is the “best”. Based on the sampling distribution histograms, describe two criteria that the “best” method should satisfy.

(11) What should happen to the SD of the sampling distributions as \( n \) increases (assuming \( N \) does not change)? (Note: be sure you address the SD of the sampling distributions, not the SD of the values in the sample.) Explain briefly.

(12) Try applying Method 1 to the sample \{44, 92, 12, 21, 56\}. What is your prediction for \( N \)? Why is this prediction not sensible?

(13) The drawback in Method 1 pointed out in question (12) can be fixed as follows:

**Method 1a:** Take the average of the \( n \) values in the sample and double it. This is our estimate of \( N \). However, if this estimate is smaller than the largest value in the sample, then use that largest value as our estimate of \( N \) instead.

Using the sample given in (12) with Method 1a, how many people would you estimate are in the class?

(14) We saw that Method 1 is unbiased. Is Method 1a unbiased? Explain why or why not.