

## Problem set 7

S520

**Upload your answers through the Assignments tab on Canvas by 11:59 pm, Thursday 24th October.** Draw graphs in R and include code.

1. (5 points.) Trosset chapter 8.4 exercise 1.
2. (5 points.) Trosset chapter 8.4 exercise 4. Assume that twenty pairs of batteries is enough to use the Central Limit Theorem.
3. (5 points.) Trosset chapter 8.4 exercise 6.
4. (5 points.) Let  $X$  be a discrete random variable with probability mass function

$$P(X = x) = \begin{cases} 0.3 & x = -2 \\ 0.6 & x = -1 \\ 0.1 & x = 12 \\ 0 & \text{otherwise.} \end{cases}$$

Let  $X_1, \dots, X_n$  be an iid sequence of random variables with the same distribution as  $X$ . Let  $\bar{X}$  be the sample mean (of  $X_1, \dots, X_n$ ).

- (a) Find  $EX$ .
  - (b) Find  $\text{Var}(X)$ .
  - (c) What is the expected value of  $\bar{X}$ ?
  - (d) What is the variance of  $\bar{X}$ ? (Note: This will depend on  $n$ .)
  - (e) Suppose  $n = 100$ . Use the R function `pnorm()` to find the approximate probability that  $\bar{X}$  is greater than 0.5.
5. (10 points.) As part of the 2016 American National Election Studies (ANES) pilot study conducted in January 2016, a sample of 1200 respondents were asked to rate various public figures on a “feeling thermometer” from 0 to 100, where 0 indicates “very cold” and 100 indicates “very warm” feelings toward the figure. The file `ANES2016.txt` contains the feeling thermometer scores for three Presidential candidates: Donald Trump, Hillary Clinton, and Bernie Sanders. (The full data set for the study is in `anes_pilot_2016.csv`, but you do not have to use it here.) To simplify things, assume the respondents are a simple random sample from the population of adult U.S. citizens (this is not quite true but is close enough for our purposes.)
    - (a) Draw side-by-side boxplots (on the same plot) of the feeling thermometer scores for Trump, Clinton, and Sanders. What’s wrong with the data?

- (b) Get rid of the feeling thermometer observations that are greater than 100, e.g. by using the `subset()` function. Find the sample mean and sample standard deviation for the new variables for Trump, Clinton, and Sanders.
- (c) Find 99% confidence intervals for:
- i. Trump's mean feeling thermometer score if all adult U.S. citizens were asked in January 2016.
  - ii. Clinton's mean feeling thermometer score if all adult U.S. citizens were asked in January 2016.
  - iii. Sanders' mean feeling thermometer score if all adult U.S. citizens were asked in January 2016.
- (Note that since we've excluded scores over 100, the sample sizes are no longer all 1200.)