

Name:

uID:

Homework 6: Estimators and Confidence Intervals

Instructions: Write your answers directly on this pdf (via an editor, iPad, or pen/pencil). The answers should be in the specified place. Students will be responsible for loading their assignments to GradeScope, and identifying what page contains each answer.

The assignment should be uploaded by 11:50pm on the date it is due. There is some slack built into this deadline on GradeScope. Assignments will be marked late if GradeScope marks them late.

If the answers are too hard to read you will lose points (entire questions may be given 0).

Please make sure your name appears at the top of the page.

You may discuss the concepts with your classmates, but write up the answers entirely on your own. **Be sure to show all the work involved in deriving your answers! If you just give a final answer without explanation, you may not receive credit for that question.**

1. The following temperatures were recorded at the Salt Lake City airport at 12:00 PM each day from March 20 through March 26 :

3/20	3/21	3/22	3/23	3/24	3/25	3/26
48	50	45	45	37	37	34

- (a) Calculate the sample mean of the data.

- (b) Calculate the sample variance of the data.

- (c) Calculate the median of the data.

2. Let X_1, \dots, X_{50} be independent $U(2, 6)$ distributed random variables. Suppose that \bar{X} is

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

(a) Calculate $E[\bar{X}]$.

(b) Calculate $Var[\bar{X}]$.

(c) Use the Central Limit Theorem to approximate $\Pr(\bar{X} \geq 4.5)$.

3. Say a Utah pollster conducted $m=15$ polls among people who would vote in the 2020 presidential elections, and reports that 60% of the respondents would vote for Chris Stewart. But the pollster did *not* report how many people n were interviewed in each poll, although each poll they conduct always includes the same number of people n . However, the pollster did report how many people voted for Spencer Cox in the 2020 gubernatorial election. Also, suppose we know that exactly $3/5$ of all voters voted for Cox in 2020.
- (a) Each poll conducted can be represented as a random variable X_i for $i = 1, 2, \dots, 15$, representing the number of people who said they would vote for Cox. What is the distribution of each X_i ? (Remember to include the parameters of the distribution in your description.)
- (b) Given the random sample X_1, \dots, X_{15} , suppose we want to estimate the number of people n included in each poll. Give an unbiased statistic that you would use to estimate the parameter n . Please include a verification that the bias is zero. (Hint: Start with a mean statistic).

(c) What is the variance of this statistic?

4. In this problem you are going to analyze the built-in R data set `iris`. First, extract the petal width of the Virginica species and save it to a vector `x` using this command:

```
x = iris$Petal.Width[iris$Species == "virginica"]
```

Answer the following: **Write down the R code you use to compute parts (a), (b), (c), (d) - it should be very short.**

- (a) Produce the following plots for `x`: (1) A histogram, (2) a box-plot, and (3) an empirical cdf. (Remember to place your plots in the space provided below. Also remember that items that won't fit in the space provided should be included at the END of the assignment.)

(b) Using a Normal approximation, what is the 95% confidence interval for the mean of \mathbf{x} ? (Assume you know the true standard deviation of x is $\sigma = 0.2747$).

(c) Using a Student t distribution, what is the 95% confidence interval for the mean of \mathbf{x} ?

(d) Now assume that you only have the first 10 measurements. That is, create the vector $\mathbf{y} = \mathbf{x}[1:10]$. Repeat parts (a) and (b) for the vector \mathbf{y} .

- (e) How did the decrease in sample size affect the results? Were the two different confidence intervals affected differently?