Problem Set 8 **Statistical Methods In Engineering And Science**

Due Date: 10:00 PM, March 6, 2023 Prof. Alexander Giessing

Last Update: March 3, 2023 Winter Quarter, 2023

Study Group: ______________ ______________ ______________

Please upload your solution in a single pdf file on Canvas. Include all calculations, R-code, and figures (if applicable). All data sets are available on Canvas https://canvas.uw.edu/courses/1614615.

Question 1. Consider a random sample $X_1, \dots X_n$ from a distribution with pdf

$$f(x) = \begin{cases} 0.5(1+\theta x) & \text{if } -1 \leq x \leq 1, \\ 0 & o/w. \end{cases}$$

(a) Show that $\hat{\theta} = 3n^{-1}\sum_{i=1}^{n} X_i$ is an unbiased estimator for θ.

(b) Show that the estimator $\hat{\theta}$ in part (a) is the method of moment estimator for θ.

(c) Obtain the standard error $SE(\hat{\theta})$. Compute an estimate $\widehat{SE}(\hat{\theta})$ of the standard error based on the data provided in `sample.txt`.

Question 2. A 95% CI for the unknown mean of some distribution contains the number zero.

(a) You construct the corresponding 98% CI, using the same data. Will it again contain the number zero?

(b) You collect new data, resulting in a data set of the same size. With this data, you construct a 95% CI for the unknown mean. Will the interval contain zero?

Question 3. Consider a shipment of 10 bags of cement, which are supposed to weigh 200lb each. The sample average weight of the 10 bags is 198.5lb with a sample standard deviation 1.65lb.

(a) Suppose that the 10 weighs can be viewed as a realization of a random sample from a normal distribution with unknown parameters. Construct a two-sided 95% CI for the expected weight of a bag.

(b) Based on these data, how many bags would you need to sample to make a two-sided 90% CI that is 0.45lb wide?

(c) Suppose that you actually do obtain the required number of bags calculated in part (b) and construct a new CI. Is it guaranteed to be at most 0.45lb wide?

Question 4. The data set `olympic1500m.txt` contains results of 23 races of the men's 1500m speed skating competition during the 2002 Winter Olympic Games in Salt Lake City. Variable `race` denotes the number of the race, variable `inner` denotes the time (in seconds) of the skater starting in the inner lane, variable `outer` denotes the time (in seconds) of the skater starting in the outer lane, and variable `diff` denotes the time difference "inner lane minus outer lane". We want to investigate whether speed skaters have an advantage in the 1500m race if they start in the outer lane.

(a) For this investigation, do you want to construct a one-sided or a two-sided confidence interval for the expected time difference? Explain.

(b) Do you think that the time differences are approximately normally distributed? Explain and justify. *(Hint: Create appropriate plots to support your argument.)*

(c) Using your answers from parts (a) and (b), construct the appropriate 95% and 98% CIs for the expected time difference.

(d) Based on your 95% CI, are you 95% confident that skaters starting in the outer line are indeed faster? What is your answer based on the 98% CI?

Question 5. [Optional. Recommended practice.] Let s_n be the sample standard deviation, σ^2 be the population variance, n be the sample size, and $\bar{X}_n$ be the sample mean. For the unknown population mean μ, state

- the two-sided 100(1-α) percent confidence interval
- the upper 100(1-α) percent confidence interval
- the lower 100(1-α) percent confidence interval
- whether the confidence intervals above are exact or approximate

under each of the following sets of assumptions:

(a) Population distribution is normal, σ is known

(b) Population distribution is unknown, σ is known, and n is large enough for CLT to apply.

(c) Population distribution is unknown, σ is unknown, and n is large enough for WLLN and CLT to apply.

(d) Population distribution is normal, σ is unknown, and n is not large enough for CLT to apply.