SDSE Homework 2

Due date: October 3rd, 2023 (by midnight)

Problem 1 (8+8+1 points)

Let $\{Y_i\}_N \overset{\text{iid}}{\sim} Y$ and consider these estimators of μ_Y:

$$\bar{Y}_A = Y_1$$
$$\bar{Y}_B = \frac{Y_1 + 2\,Y_N}{3}$$
$$\bar{Y}_C = \frac{1}{M}\sum_{i=1}^{M} Y_i \qquad 1 < M < N$$
$$\bar{Y}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

(a) Determine whether each of these estimators is biased or unbiased.

(b) Compute the variance of each estimator as a function of σ_Y^2.

(c) Order the estimators from least to highest variance.

Problem 2 (3+2+3+3 points)

We wish to compute a confidence interval for the yield strength of an aimuminum alloy. We have 100 independent measurements of the yield strength. The mean of this sample in 92 MPa and the sample standard deviation is 8 MPa.

(a) Find a 95% confidence interval for the yield strength of the alloy.

(b) Find a 99% confidence interval for the yield strength of the alloy.

(c) How many measurements must be used to specify the yield strength of the alloy to within 0.5 MPa with 95% confidence?

(d) How many measurements must be used to specify the yield strength of the alloy to within 0.5 MPa with 99% confidence?

Problem 3 (3+2 points)

Eight independent measurements of the boiling point of ethanol produced a sample mean of 78.37°C Celsius and a sample standard deviation of 0.14 degrees.

(a) Find a 98% confidence interval for the boiling point of ethanol.

(b) What assumption is tacit in the problem statement?

Problem 4 (3+2+3 points) A non-profit organization wishes to guage support for a proposed expansion of bike lanes along a busy corridor in Berkeley. They survey 400 residents asking whether they support or do not support the measure. The results of this survey are provided in `survey.txt`. Here a '1' means support and a '0' means does-not-support.

(a) Find the sample mean and the (biased) sample standard deviation.

(b) Find a 95% confidence interval for the proportion of residents that support the measure.

(c) Estimate the number of additional surveys that should be collected if we wish to reduce the radius of the confidence interval by half.

Problem 5 (2+3+3 points)

The CDC recommends that chlorine levels in swimming pools be kept at 1 ppm (parts per million). Lower levels can result in increased bacterial count, while higher levels can lead to skin and eye irritation. A pool maintenance company wishes to design an inspection protocol that will result in either a "pass" or a "fail" decision for a pool. Assume that the chlorine level measurements are normally distributed.

(a) State the null and alternative hypotheses for the test.

(b) An employee for the company takes 3 measurement: 0.92 ppm, 0.84 ppm, and 0.96 ppm. The equipment that they use is known to have a standard deviation of 0.2 ppm in chlorine level measurements. Find the p-value for this test.

(c) Suppose that we want to fail pools whose chlorine level is more than 0.1 ppm from the recommendation, and we want the rejection to have a significance level of 0.04. How many measurements should be taken?.

Problem 6 (2+2+3+2 points)

In Problem 3, does the data challenge the hypothesis that the boiling point of ethanol is 78.5°C Celsius to a level of significance of 0.01?

(a) State the null and alternative hypotheses for the test.

(b) What is the appropriate statistic to use for this test?

(c) Find the p-value.

(d) State the conclusion

Problem 7 (2+2+3+2 points)

A brand of trail mix advertises a 1-to-1 fruit-to-nuts ratio in their product. To test their manufacturing process, the company takes a single bag of the mix and counts 62 nuts and and 76 fruit pieces. Is the process performing correctly to a 0.05 level of significance?

(a) State the null and alternative hypotheses for the test.

(b) What is the appropriate statistic to use for this test?

(c) Find the p-value.

(d) State the conclusion