LAB 4 ASSIGNMENT

CONFIDENCE INTERVALS AND HYPOTHESIS TESTING

This lab assignment will give you the opportunity to explore the concept of confidence intervals and hypothesis testing in the context of a real-world problem. In particular, you will study the relationship between the sample size, confidence level, and the margin of error of confidence intervals. Moreover, you will investigate the idea of the *p*-value in a corresponding hypothesis-testing problem. One-sample and two-sample inferences will be considered for the corresponding data.

Metal Conductivity

Electric conductivity is the movement of electrically charged particles. All metals conduct electricity to a certain extent, but certain metals are more highly conductive. The most highly conductive metals are silver, copper and gold. Copper, for example, is highly conductive and commonly used in metal wiring. Brass, on the other hand, contains copper but other materials in its makeup reduce conductivity. Electrical conductivity is commonly measured by siemens per meter ($\mathrm{S/m}$). As the values can be quite large, this lab will scale the values by a factor of 10^6.

A new steel manufacturer would like to generally satisfy a particular conductivity. If the particular conductivity is not met, the production process is discontinued while adjustments are made. One recognized way of monitoring production is to take a random sample of steel pieces. When the production process is operating properly, the conductivity is approximately normally distributed with the target mean value of 6.200×10^6 S/m (or 6.200 mega-Siemens per meter). In practice, the mean conductivity (μ) depends on various factors that may change from piece to piece of steel. Random samples of size 30 will be taken periodically through a week to determine any change in the value of the mean.

The manufacturer constructs two types of steel (a base design and a newer design to improve conductivity) and would like to see which of them has better conductivity. In order to compare the conductivity of the two steel types, 30 steel pieces of each type were randomly selected and their conductivity measured. The data are recorded in two columns (*Steel A* and *Steel B*) in the *Data* worksheet. For the purpose of further improving conductivity, the 30 selected pieces of Steel B were combined with tunqstoid to potentially obtain an even higher conductivity before their conductivities were measured again. **Unless otherwise stated, numerical answers should be to four decimal places.**

Name of Variable	Description of Variable
Steel A	Conductivity of randomly selected base design pieces of steel,
Steel B	Conductivity of randomly selected newer design pieces of steel,
Tunqstoid	Conductivity of randomly selected pieces combined with tunqstoid.

1. Before carrying out statistical analysis for the data, examine the study design. Is this an observational study or an experiment? Are population inferences applicable? Causal inferences? Explain briefly.

2. First obtain the summary statistics for the Steel A and Steel B samples.

    (a) Obtain and paste the descriptive statistics for each group into your report. (Consider reading ahead to obtain all needed output.) Compare the means and standard deviations of the two distributions. Which of the two steel types has better conductivity? Explain briefly.

    (b) Obtain and paste the side-by-side boxplots into your report. (Consider reformatting the data to obtain the same boxplot format seen in the *Lab 1 Instructions*.) Comment about the shape (symmetric, skewed) of each distribution. Are there are any outliers? Compare the medians and the interquartile ranges of the two distributions.

(c) Obtain and paste the histograms for each group into your report using the following bins: 6.10, 6.12, 6.14, ..., 6.28. The format of your histogram should be the same as the format from previous labs and the *Lab 1 Instructions*. Describe the shape of the histogram (modality, skewness, outliers). Is there any evidence that the assumption of normality necessary to apply the t-tools is satisfied in each group? Explain briefly.

3. Use the summary statistics from Question 2 to answer the following.

   (a) Construct and interpret a 98% two-sided confidence interval for the mean conductivity of each steel type.

   (b) Since the target mean value is 6.200×10^6 S/m, is there any evidence that the mean conductivity of either type not meeting this target value? Refer to the 98% two-sided confidence interval for each type obtained in part (a) to answer the question. Explain briefly.

4. Do the data provide evidence that the mean conductivity of each type does not meet the target value of 6.200×10^6 S/m? Answer the question by carrying out the appropriate statistical tests.

   (a) Carry out an appropriate test to check the above claim using the data for each type. Use an appropriate significance level corresponding to the confidence level in Question 3. In particular, state the null and alternative hypotheses in terms of the population parameters, obtain the value of the test statistic, specify the distribution of the test statistic under the null hypothesis, and obtain the *p*-value of the test. What do you conclude? (Note that you will have to calculate the value of the test statistic and the corresponding *p*-value by entering appropriate formulas into Excel worksheet. *Lab 4 Instructions* may be useful in this part.)

   (b) What are the assumptions about the distribution of conductivity required to make the tests in part (a) valid? Do the assumptions hold? Explain briefly. It is not required to verify the assumptions with Excel. Are these the same assumptions as the intervals in Question 3? Why?

5. In this part, compare the mean conductivity of A and B steel types. Do the data provide any evidence of that the mean conductivity of steel B type is higher than the steel A type?

   (a) Answer the above question by carrying out the appropriate test in the *Data Analysis* menu. Before you choose an appropriate test, you might refer to the output in Question 2 to decide what test would be appropriate. In particular, state the null and alternative hypotheses in terms of the population parameters, obtain the value of the test statistic, specify the distribution of the test statistic under the null hypothesis, and obtain the *p*-value of the test. What do you conclude?

   (b) What are the assumptions about the distributions of conductivity required to make the tests in part (a) valid? Do the assumptions hold? Explain briefly. It is not required to verify the assumptions with Excel.

   (c) From the result in part (a), what is the most appropriate 95% confidence interval/bound? What could be stated about the claimed value in part (a) in relation to this interval/bound?

6. For the purpose of further improving conductivity, the 30 selected pieces of Steel B were combined with tunqstoid to potentially obtain an even higher conductivity before their conductivities were measured again. In this question, estimate the effect of the combination on the mean conductivity.

   (a) Do the data provide evidence that the combination increased the mean conductivity of Steel B pieces? Answer the question by carrying out an appropriate test in Excel. In particular, state the null and alternative hypotheses in terms of the population parameters, obtain the value of the test statistic, specify the distribution of the test statistic under the null hypothesis, and obtain the *p*-value of the test. What do you conclude?

(b) Use the *Descriptive Statistics* feature in the *Data Analysis* menu to obtain a corresponding 95% one-sided confidence interval/bound for the mean change in conductivity of Steel B pieces after combination. First create a new variable, *Change*, defined as the difference in conductivity between *Tunqstoid* and *Steel B* measurements. Is the interval consistent with the test outcome in part (a)? Explain briefly.

(c) What are the assumptions required to make the tests in part (a) valid? Do the assumptions hold? Explain briefly. It is not required to verify the assumptions with Excel.

MARKING SCHEMA TO THE LAB 4 ASSIGNMENT

Question 1 (6)

Type of study: 2 points
Population inferences: 2 points
Causal inferences: 2 points

Question 2 (37)

(a) Summary statistics: 3 points
Comparison of means and standard deviations: 2 points each (4 points total)
Better type: 2 points
(b) Correctly formatted side-by-side boxplots: 4 points
Analysis of the shape of each boxplot: 2 points each (4 points total)
Comparing the centers and spreads: 2 points each (4 points total)
(c) Correctly formatted histogram: 4 points each (8 points total)
Analysis of the shape of each histogram: 3 points each (6 points total)
Comments about normality: 2 points

Question 3 (12)

(a) 98% two-sided confidence interval for each type: 2 points each (4 points total)
Interpretation of each confidence interval: 2 points each (4 points total)
(b) Relating to target mean value: 2 points each (4 points total)

Question 4 (23)

(a) Appropriate significance level: 1 point
Test for Steel A: 8 points
(hypotheses: 2, distribution: 1, test statistic: 2, *p*-value: 1, conclusion: 2)
Test for Steel B: 8 points
(b) Assumptions: 2 points
Checking assumptions: 2 points
Relating assumptions to Question 3: 2 points

Question 5 (21)

(a) Choosing appropriate test: 2 points
Output: 3 points
Test to compare Steel A and B types: 8 points
(hypotheses: 2, distribution: 1, test statistic: 2, *p*-value: 1, conclusion: 2)
(b) Assumptions: 2 points
Checking assumptions: 2 points
(c) Appropriate confidence interval/bound: 2 points
Conclusion about claimed value in part (a) in relation to interval/bound: 2 points

Question 6 (27)

(a) Choosing appropriate test: 2 points
Output: 3 points
Test for the combination change: 8 points
(hypotheses: 2, distribution: 1, test statistic: 2, *p*-value: 1, conclusion: 2)

(b) Output: 3 points
Confidence interval for the combination change: 3 points
Interpretation of confidence interval: 2 points
Consistency of the interval with the outcome of the test in part (a): 2 points

(c) Assumptions: 2 points
Checking assumptions: 2 points

TOTAL = 126