

## 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.

#### Testing the Breaking Strength of Metal Rods

A manufacturer of metal rods used in construction must make reasonably certain that the rods satisfy some strength specifications. If the strength requirement 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 several rods from each hour's production and to determine their average breaking strength. When the production process is operating properly, the strength of rods is approximately normally distributed with the target mean value of 64 ksi. In practice, the mean breaking strength ( $\mu$ ) depends on various factors that may change from hour to hour. Random samples of size 30 will be taken every hour to determine any change in the value of the mean.

Suppose we would like to estimate the population mean using a confidence interval. What is the percentage of samples producing confidence intervals containing the true population mean? How likely is shutting down the production process erroneously based on the sample data? You will answer all the questions using simulation in the first part of the assignment.

In Part 1, you must use Excel in a Windows environment to obtain data (different data may be produced by *Random Number Generation* in Excel on MAC OS). Excel 2010 or higher version should be used in this part (earlier versions may produce different sequence of random numbers for the same seed).

#### Part 1: Simulation

Open the file *lab4.xls* and the *Data* worksheet. The data in the range E2:CZ31 consist of 100 samples of size 30 obtained from a normal distribution with a mean of 64 ksi and a standard deviation of 1 ksi. This corresponds to randomly selecting 100 samples, each consisting of 30 rods and obtaining strength measurements for each of them. The data were generated with the *Random Number Generation* feature (seed 1000) and are provided here for your convenience and consistency. Select the data, copy it, and paste into the yellow area in the *Simulation* worksheet.

The following labels are in the range A1-A4 in the *Simulation* worksheet: *Confidence Level*, *Sigma*, *Hypothesized Mean*, and *Level of Significance*. In the range B1-B4, you will enter the current numerical values of the parameters defined in the range A1-A4. The sample means, confidence intervals, and the  $p$ -values of the tests will be calculated automatically for each of the 100 samples given the parameters provided in the range B1-B4. Some cells in the worksheet are protected so that you will not be able to change mistakenly any formulas embedded into the worksheet while working.

1. First, you will use the template to verify some theoretical predictions with simulation and examine the relationship between confidence level and margin of error. Answer the following questions:
  - (a) Change the confidence level in B1 to the values of 0.90, 0.95, and 0.99, keeping the other parameters constant. Report the margin of error of the confidence intervals in each case. How does the margin of error change as the confidence interval increases? Explain briefly.

- (b) For each of the three confidence levels (0.90, 0.95, and 0.99), also obtain the number of confidence intervals that failed to cover the hypothesized population mean of 64. The COUNTIF() function may be useful in this task. Are the observed counts consistent with the values predicted by the theory? Explain briefly.
2. In this question, you will use the data generated in the previous question to test hypotheses about the mean. Consider the following null and alternative hypotheses:

$$H_0: \mu = 64 \quad \text{vs.} \quad H_A: \mu \neq 64.$$

In other words, we test the null hypothesis that the process is working properly against the alternative that the process should be shut down for adjustments. Suppose that in fact  $H_0: \mu = 64$  is true and you use sample data to test it. Given the hypothesized value of the mean in cell B3 and the population standard deviation in B2, the corresponding  $p$ -values and the decisions about the null hypothesis (R = Reject, DR = Don't Reject) are calculated automatically.

- (a) Change the level of significance in cell B4 to the values of 0.10, 0.05, and 0.01, keeping the other parameters constant. For each value, determine the number of samples that led to the wrong decision about the null hypothesis (rejecting the true value). How does the number of samples change as the level of significance increases? Explain briefly.
- (b) Compare the outcome of the test at the 5% level of significance with the 95% confidence intervals that failed to cover the mean of 64 for each sample. Repeat the exercise with the 1% level of significance and the 99% confidence intervals. What do you conclude about the relationship between confidence intervals and two-sided tests?

## Part 2: Comparison of Two Alloys

Suppose the manufacturer has just developed two new alloys (*ALLOY 1* and *ALLOY 2*) and would like to see which of them has better strength qualities. In order to compare the strength of the two alloys, 30 rods made of each alloy were randomly selected and subjected to the strength test. The breaking strength of each rod in the two groups was recorded. The data are recorded in the two columns *ALLOY 1* and *ALLOY 2* available in the *Data* worksheet. After the strength test, the 30 rods made of *ALLOY 2* were subjected to a combination of high pressure and temperature and their strength values were determined again.

<u>Name of Variable</u>	<u>Description of Variable</u>
ALLOY 1	The strength of rods made of ALLOY 1,
ALLOY 2	The strength of rods made of ALLOY 2,
ALLOY 2 + TREATMENT	The strength of ALLOY 2 rods subjected to high pressure & temperature

3. First you will obtain the summary statistics for each sample and then obtain a confidence interval for the mean strength of each alloy.
- (a) Use the *Descriptive Statistics* feature to obtain the summary statistics and a 95% confidence interval for the mean strength of each alloy. Paste the summary statistics into your report and report the 95% confidence interval for each alloy. Use the summaries to compare the two alloys. Which of the two alloys has better strength qualities? Explain briefly.
- (b) According to the specifications, the mean strength of each alloy is required to exceed 64 ksi. Is there any indication that the mean strength of either alloy is below the required threshold value of 64 ksi? Refer to the 95% confidence interval for each alloy to answer the question. Explain briefly.

4. Do the data provide evidence that the mean strength of each alloy exceeds the threshold value of 64 ksi? Now you will 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 alloy. 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? Notice that as there is no appropriate feature in *Data Analysis* to carry out the test directly, so you will have to calculate the value of the test statistic and the corresponding  $p$ -value by entering appropriate formulas into Excel worksheet. *Lab 3 Instructions* may be useful in this part.
  - (b) What are the assumptions about the distribution of strength 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.
5. In this part, you will compare the mean strength of ALLOY 1 and ALLOY 2 rods. Do the data provide any evidence of a difference in the mean strengths of ALLOY 1 and ALLOY 2 rods?
  - (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 3 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 distribution of strength 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.
6. The thirty ALLOY 2 rods were subjected to a combination of high pressure and temperature. In this question, you will estimate the effect of the treatment on the mean strength of the rods.
  - (a) Do the data provide evidence that the treatment increased the mean strength of the ALLOY 2 rods? 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 95% two-sided confidence interval for the mean change in strength of ALLOY 2 rods after the treatment. First create a new variable, EFFECT, defined as the difference in strength between ALLOY 2 + TREATMENT rods and ALLOY 2 rods. Is the interval consistent with the test outcome in part (a)? Explain briefly.
  - (c) What are the assumptions necessary to make the test in part (a) and confidence interval in part (b) valid? Do the assumptions hold? Explain briefly.
  - (d) Is the effect of the treatment independent of the initial strength of the rods? In order to answer the question, obtain the plot of the variable EFFECT versus ALLOY 2 measurements. What do you conclude?

## MARKING SCHEMA TO THE LAB 4 ASSIGNMENT

Proper Header and Appearance: 10 points

### Question 1 (13)

- (a) Margin of error for each confidence level: 1 point each (3 points total)  
Relationship between margin of error and confidence level: 2 points
- (b) Observed number of confidence intervals that failed to cover the population mean: 2 points for each level (6 points total)  
Consistency with the theory: 2 points

### Question 2 (14)

- (a) Number of samples that led to the wrong decision about the null hypothesis for each of the three levels of significance: 2 points each (6 points total)  
Relationship between number of samples and significance level: 2 points
- (b) Comparison of the test at  $\alpha = 0.05$  with 95% confidence intervals: 2 points  
Comparison of the test at  $\alpha = 0.01$  with 99% confidence intervals: 2 points  
Conclusion: 2 points

### Question 3 (14)

- (a) Summary statistics: 3 points  
95% confidence interval for each alloy: 2 points each (4 points total)  
Comparison and indicating the better alloy: 3 points
- (b) Relating to threshold value of 64 ksi: 2 points each (4 points total)

### Question 4 (16)

- (a) Test for ALLOY 1: 6 points  
(hypotheses: 1, distribution: 1, test statistic: 2,  $p$ -value: 1, conclusion: 1)  
Test for ALLOY 2: 6 points
- (b) Assumptions: 2 points  
Checking assumptions: 2 points

### Question 5 (12)

- (a) Choosing appropriate test: 2 points  
Test to compare ALLOY 1 and ALLOY 2 strengths: 6 points  
(hypotheses: 1, distribution: 1, test statistic: 2,  $p$ -value: 1, conclusion: 1)
- (b) Assumptions: 2 points  
Checking assumptions: 2 points

**Question 6 (25)**

- (a) Choosing appropriate test: 2 points  
Test for the treatment effect: 6 points  
(hypotheses: 1, distribution: 1, test statistic: 2,  $p$ -value: 1, conclusion: 1)
- (b) 95% confidence interval for the treatment effect: 3 points  
Consistency of the interval with the outcome of the test in part (a): 2 points
- (c) Assumptions: 2 points  
Checking assumptions: 2 points
- (d) Plot of EFFECT versus ALLOY 2 strengths: 6 points  
Comments: 2 points

**TOTAL = 104**