**DATA 557**

**Winter 2022**

**Homework Assignment 1**

**Instructions**

Submit your solutions **in pdf format** to the dropbox on the canvas page by **12:00PM, Friday January 21**. You may use any program to generate your pdf file. (RStudio is recommended but not required.)

For each question you will be given 1 point for complete credit, ½ point for partial credit, and 0 points for no credit. Assignment of credit will be based on the correctness of your answers as well as your reasoning (when requested as part of the question). You do not need to submit R code for this assignment.

You may work together to help each other solve problems, but you should create your own solutions and hand in your own work without copying others’ work.

**Question 1**

Suppose that you flip a coin 40 times and count the number of heads.

1.1. What is the distribution of the number of heads assuming the coin is fair?

1.2. The sample proportion of heads has an approximately normal distribution. What are the mean and standard deviation of this distribution assuming the coin is fair?

1.3. Define the Z-statistic for conducting a test of the null hypothesis that the coin is fair (i.e., has probability of a head equal to 0.5).

1.4. Suppose the experiment results in 15 heads and 25 tails. Conduct a test of the null hypothesis with type I error probability 0.05 using the normal approximation. State the Z statistic, the p-value, and the conclusion of the test (do you reject the null hypothesis or not).

1.5. If you had decided to use a type I error probability of 0.1 instead of 0.05 would your conclusion be different? Explain.

1.6. Calculate the p-value using the binomial distribution. Do you reach the same conclusion with the binomial distribution as with the normal approximation?

1.7. Calculate a 95% confidence interval for the probability of a head using the normal approximation. Does the confidence interval include the value 0.5?

1.8. Calculate a 90% confidence interval for the probability of a head using the normal approximation. How does it compare to the 95% confidence interval?

**Question 2**

A study is done to determine if enhanced seatbelt enforcement has an effect on the proportion of drivers wearing seatbelts. Prior to the intervention (enhanced enforcement) the proportion of drivers wearing their seatbelt was 0.7. The researcher wishes to test the null hypothesis that the proportion of drivers wearing their seatbelt after the intervention is equal to 0.7 (i.e., unchanged from before). The alternative hypothesis is that the proportion of drivers wearing their seatbelt is not equal to 0.7 (either < 0.7 or > 0.7). After the intervention, a random sample of 400 drivers was selected and the number of drivers wearing their seatbelt was found to be 305.

2.1. Calculate the estimated standard error of the proportion of drivers wearing seatbelts after the intervention.

2.2. Calculate a 95% confidence interval for the proportion of drivers wearing seatbelts after the intervention. What conclusion would you draw based on the confidence interval?

2.3. Conduct a test of the null hypothesis with type I error probability 0.05 using the normal approximation. Should the null hypothesis be rejected? How does your conclusion compare to the conclusion from the confidence interval?

2.4. Calculate the approximate p-value using the normal approximation and the exact p-value using the binomial distribution. Are the two p-values very different?

2.5. Calculate the power of the test to detect the alternative hypothesis that the proportion of drivers wearing their seatbelt after the intervention is equal to 0.8.

**Question 3**

Data set: ‘iq.csv’ (data set posted on canvas)

The data come from a study of lead exposure and IQ in children. IQ scores were measured on a sample of children living in a community near a source of lead. The IQ scores were age-standardized using established normal values for the US population. Such age-standardized scores have a mean of 100 and a standard deviation of 15 in the US population.

3.1. Create a histogram of the IQ variable. Is the distribution approximately normal?

3.2. Calculate the sample mean and sample SD of IQ. How do they compare numerically to the US population values?

3.3. Test the null hypothesis that the mean IQ score in the community is equal to 100 using the 2-sided 1-sample t-test with a significance level of 0.05. State the value of the test statistic and whether or not you reject the null hypothesis at significance level 0.05.

3.4. Give the p-value for the test in the previous question. State the interpretation of the p-value.

3.5. Compute a 95% confidence interval for the mean IQ. Do the confidence interval and hypothesis test give results that agree or conflict with each other? Explain.

3.6. Repeat the hypothesis test and confidence interval using a significance level of 0.01 and a 99% confidence interval.