

# Homework 5 - MATH 141

**Due Date:** Thursday 10/07/2021, 11:59 PM

**Instructions:**

- Please provide complete answers/solutions for each question/problem.
- If it involves mathematical computations, please provide your reasoning and/or detailed solutions.
- There are two ways you can write your answers, a: by handwriting (either physically or digitally), or b: by typing on a template document with file type options, Word or RMarkdown, which can be downloaded from the [course website](#).
- If you had handwritten your answers/solutions on a physical paper, make sure to label it properly and please scan your document using a scanner app for convenience. Suggestions: (1) “[Tiny Scanner](#)” for [Android](#) or (2) “[Scanner App](#)” for [iOS](#).
- If you have questions or concerns, please feel free to ask the instructor.
- **Please save your work as pdf file(s), don't put your name in any part of the document, and submit it to the Gradescope page for this course. Your document upload will correspond to your name automatically in Gradescope.**

## I. Hypothesis Testing

I.A. For the following statements, state the null and alternative hypothesis with its corresponding mathematical notations. Indicate whether the hypothesis test would be one-sided or two-sided.

*I.A.1.*

1. A teacher want to test if it takes fewer than 45 minutes to teach a lesson plan.
2. On statistics test, it is known that 60% of students pass it on the first try. We want to test if less than 60% pass on the first try.
3. You are testing that the mean speed of your cable Internet connection is less than three hundred Megabits per second.
4. We want to test if the average number of house spider appearances is more than 19 spiders per day.

*I.A.2.*

1. A study of emotions on Twitter determined that 60% of tweets sampled within the USA in 2020 are associated with negative emotions. A researcher wants to know if the actual percentage is actually not equal to 60%.
2. An observational study of sentiments on social media towards vaccine mandates is conducted in 2020. A sample of twitter users are labeled according to its political leanings (left or right leaning). For each group, the percentage of users with positive sentiments towards a vaccine mandate is computed. A researcher wants to determine if there is a significant difference in positive sentiments towards vaccine mandates between left and right leaning users.
3. Most adults need around 8 hours of sleep. You are testing that the mean sleeping time of young adults (20-30 year-old individuals in 2020) is less then 8 hours.
4. You are waiting in line to enter a ride at Disneyland Park but the wait time is announced to be no more than 30 minutes. You want to prove to the park managers that the mean wait time is actually more than 30 minutes.

I.B. In this exercise, we consider the data set [OpenIntro Data Sets: Malaria Vaccine Trial](#).

Volunteer patients were randomized into one of two experiment groups where they would receive an experimental vaccine or a placebo. They were subsequently exposed to a drug-sensitive strain of malaria and observed to see whether they came down with an infection.

In this study, volunteer patients were randomized into one of two experiment groups: 14 patients received an experimental vaccine or 6 patients received a placebo vaccine. Nineteen weeks later, all 20 patients were exposed to a drug-sensitive malaria virus strain; the motivation of using a drug-sensitive strain of virus here is for ethical considerations, allowing any infections to be treated effectively.

Consider the data set summarized as a contingency table shown in Table 1.

Table 1: Summary results of the Malaria Vaccine Trial

treatment	outcome		Total
	infection	no infection	
placebo	6	0	6
vaccine	5	9	14
Total	11	9	20

1. Suppose that “success” is when the individual gets no malaria infection. What is the point estimate? Use mathematical symbols such as  $\hat{p}_V$  for vaccine proportion and  $\hat{p}_P$  for placebo proportion.
2. State the null and alternative hypothesis. Use mathematical symbols such as  $H_0$  for the null and  $H_A$  for the alternative.
3. Under the null assumption, what is the expected number of people who do not get infected for each group? What is the assumption under the null hypothesis in terms of the variables?
4. Suppose that we use a randomization test. Explain the randomization simulation procedure in simple terms and make sure to include the assumptions, reasoning, and purpose of the simulation.
5. Using 1000 shuffles,  $\frac{16}{1000} = 0.016$  of the shuffles yielded a difference in proportion greater than or equal to the point estimate. What is this value called and interpret this value in the context of the vaccine trial data.
6. Notice that the observed difference in rates of infection is large but the sample size for the study is small. Do you think this is unclear if the observed difference did occur by chance? Explain why.
7. What are the two options of conclusions you can make based on the randomization test (explain your answer in the context of the vaccine trial)?
8. Based on a significance level of 0.05, what can you conclude?
9. Based on a significance level of 0.01, what can you conclude?
10. The significance level provides the cutoff for the p-value which will lead to a decision and it indicates a risk probability of concluding that a difference in proportion exists when there is no actual difference. In this particular study, you have made different conclusions in problems 8 and 9 based on two significance levels. What type of error corresponds to each decision?

## II. Confidence Intervals

Decide whether the following statements are true or false.

1. [True/False] Bootstrapping is best suited for modeling studies in which the data was obtained through random sampling from a population.
2. [True/False] A single plausible value for a parameter is provided by a point estimate. A point estimate, on the other hand, is rarely perfect; there is usually some error in the estimate. A confidence interval is a plausible range of values for a population parameter.
3. [True/False] “the 95% confidence interval for the population mean is (5, 10)”, is equivalent to “there is a 95% probability that the population mean is between 5 and 10”.
4. [True/False] Confidence intervals are always valid if the central limit theorem holds.
5. [True/False] If you take large random samples from the same population repeatedly and calculate 95 percent confidence intervals for the population average, approximately 95 percent of the intervals should contain the population average.

### III. Textbook Exercises

**Note:** To view the selected exercises below, please refer to the textbook, *OpenIntro: Introduction to Modern Statistics (2021)* by Mine Çetinkaya-Rundel and Johanna Hardin, First Edition.

**Section 11.5:** Choose any 3 exercise problems below to answer.

2. Identify the parameter, II.
4. True null hypothesis.
6. Identify hypotheses, II.
8. Heart transplants.

**Section 12.5:** Choose any 3 exercise problems below to answer.

2. Chronic illness.
4. Bootstrap distributions of  $\hat{p}$ , I.
6. Bootstrap distributions of  $\hat{p}$ , III.
8. Waiting at an ER.