

Homework 8 - MATH 141

Due Date: Wednesday 11/24/2021, 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 a problem asks you to show your R code, R outputs, or R plots, please provide them as additional pages into your current homework pdf while labeling them properly. This means that, **if you have handwritten your homework solutions and saved it as pdf, you would need to merge the separate pdf which contains your R code, R outputs, or R plots. Note that all of the problems that require R does not require you to show your R code - unless the problem specifically says so.**
- If you have questions or concerns, please feel free to ask the instructor.
- **Please save your work as one pdf file, 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. Inference for Two-Way Tables

The exercise problems shown below was taken and slightly modified from your textbook [OpenIntro: Introduction to Modern Statistics Section 18.4](#).

1. True or False.

Determine if the statements below are true or false. For each false statement, explain why it is false.

- As the degrees of freedom increases, the mean of the Chi-square distribution increases.
- If you found $\chi^2 = 10$ with $df = 5$ you would fail to reject H_0 at the 5% significance level.
- As the degrees of freedom increases, the variability of the Chi-square distribution decreases.

2. Sleep deprived transportation workers.

The National Sleep Foundation conducted a survey on the sleep habits of randomly sampled transportation workers and randomly sampled non-transportation workers that serve as a “control” for comparison. The code block below generates the necessary data. [N. S. Foundation 2012](#)

```
sleep_dep <- tibble(  
  hours = c(  
    rep("hours < 6", 139),  
    rep("6 <= hours <= 8", 692),  
    rep("hours > 8", 256)  
  ),  
  profession = c(  
    rep("non-transportation workers", 35), rep("transportation workers", 104),  
    rep("non-transportation workers", 193), rep("transportation workers", 499),  
    rep("non-transportation workers", 64), rep("transportation workers", 192)  
  )  
)
```

- What is the null and alternative hypothesis?
- Use the `infer` package in R to carry out the randomization procedure and theoretical method for the Chi-Squared test for independence. Plot the resulting distributions and provide a short description/analysis of the plot.
- Compute the p-value using the randomization and theory. Do they agree? What is the conclusion?

II. Inference for One Mean

The exercise problems shown below was taken and slightly modified from your textbook [OpenIntro: Introduction to Modern Statistics Section 19.4](#).

1. True or False.

Determine if the statements below are true or false. For each false statement, explain why it is false.

- A 99% confidence interval for a population mean μ based on a random sample of size 10 is 55.5 to 71.9. It follows that the p-value from the hypothesis test $H_0: \mu = 70$ versus $H_A: \mu \neq 70$ is significant at the $\alpha = 0.05$ level.
- The conditions for one mean inference is dependence and abnormality of the sampling distribution.
- As the degrees of freedom decreases, the t-distribution approaches the standard normal distribution.

2. Length of Gestation.

Every year, the United States Department of Health and Human Services releases to the public a large dataset containing information on births recorded in the country. This dataset has been of interest to medical researchers who are studying the relation between habits and practices of expectant mothers and the birth of their children.

In this exercise we work with a random sample of 1,000 cases from the dataset released in 2014. The length of pregnancy, measured in weeks, is commonly referred to as gestation. The `births14` data used in this exercise can be found in the `openintro` R package.

You might have heard that human gestation is typically 40 weeks. To assess the 40 week claim, the null and alternative hypotheses is as follows.

H_0 : The average length of pregnancy in the population is 40 weeks.

$$\mu = 40$$

H_A : The average length of pregnancy in the population is not 40 weeks.

$$\mu \neq 40$$

- Use R to compute, the summarized statistics of the data (e.g. mean, standard deviation, quartiles, etc). Provide a visualization (histogram and boxplot) of the number of weeks and make an assessment on whether the conditions for inference for one mean are satisfied.
- Use the `infer` package in R to carry out the simulation and theoretical method for the hypothesis test. Interpret the p-value and provide a conclusion in context of the data. Include the visualizations of the theoretical and simulated distributions with the test statistic and the shaded p-value. Compare the standard errors of the simulation and the theoretical results. Should it closely agree? Why or why not?
- Use the `infer` package in R to carry out the simulation and theoretical method for computing the 90% confidence interval. Interpret the 90% confidence interval and provide a conclusion in context of the data. Include the visualizations of the theoretical and simulated distributions with the 90% confidence interval. Compare the standard errors of the simulation and the theoretical results. Should it closely agree? Why or why not?