

Homework#1 (Total score = 100)
COMPSCI- 5590-0012 Econometrics of Data Science
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF MISSOURI-KANSAS CITY

Due on Friday, September 30, 2022 at 11:59 PM (in Canvas)

“It must be an individual submission. Any kind of copying or corroboration will be severely penalized”.

Suppose our goal is to quantify the impact of the repeal of the national maximum speed limit on average speeds on U.S. roads. Representative speed data are collected for particular time periods before and after the repeal. The spreadsheet “hw1_speed_data.csv” contains vehicular speed data before and after the repeal of a speed limit law. Use **R** to do the solve the following questions:

- *1. Generate summary statistics for vehicular speeds data and compare the results. (6)
- *2. Generate and interpret box plots for vehicular speeds data (make them look nice). (10)
- *3. Generate and interpret histograms for vehicular speeds data (make them look nice). (10)
4. Find out the mean and median values of the after-speed data for those particular vehicles whose speeds before the repeal were greater than 60 mph. (5)
5. Find out the frequency distribution of vehicular after-speed data and interpret results. (5)
6. Generate 99% confidence intervals for mean vehicular after-speed data assuming the population variance is unknown. Explain each step and interpret the results. (8)
7. Generate 95% confidence intervals for the variance of before-speed data. Explain each step and interpret the results. (8)
8. Test whether the mean speed is 55 mph before and 60 mph after at the $\alpha=5\%$ significance level. Explain each step and interpret the results. (10)
9. Test whether the variance of after-speed data is less than 19 mph^2 at the $\alpha=5\%$ significance level. Explain each step and interpret the results. (10)
10. Test that the mean vehicular speeds before and after are equal at the $\alpha=10\%$ significance level. Explain each step and interpret the results. (10)
11. Test that the vehicular speed variances before and after are equal at the $\alpha=5\%$ significance level. Explain each step and interpret the results. (10)
12. Use a Mann-Whitney-Wilcoxon test to assess whether the distributions of speeds before and after are equal. Also draw density plots using before and after speeds data. Interpret the results based on the test and drawing. (8)

In each case, clearly explain what relationships you used, give each step of your reasoning as appropriate, and state your conclusions.

*Notes: for questions 1, 2, 3 provide solutions for both of the vehicular before and after speeds data. Consider before and after speeds data as *independent* samples.

Submission guidelines:

- In the document, first, write down a question and then write your answer after that.
- In the answer script, write your name and student ID on the right-hand side of the header.
- Please type your response. No handwritten submission will be accepted.
- Upload a pdf file containing the results, graphs, reasoning, and interpretations.
- Upload the R file.
- Please write the file name (both pdf and R) in this format: COMPSCI-5590-hw1-your last name