Tutorial Sheet 7: Statistical Tests and Regression
Course: CSEG 2036P | *School of Computing Sciences, UPES*
Faculty: Dr. Mrittunjoy Guha Majumdar

1. Let's consider a scenario where UPES wants to investigate whether a mindfulness meditation program has a significant impact on the stress levels of a group of individuals. The stress levels of each individual were measured before and after the mindfulness program, resulting in the following paired data:

   Before (stress levels): 65 72 68 74 60 80 77 68 75 62 70 78
   After (stress levels): 60 65 55 70 58 75 72 63 68 58 67 70

   Using the sign test for paired data ($\alpha = 0.05$), UPES aims to test whether there is a statistically significant decrease in stress levels after participating in the mindfulness program. What is your conclusion on this?

2. A professor has a hypothesis that the introduction of a new online learning curriculum has raised the median final exam score compared to the previous year's median score of 75. To test this hypothesis, a random sample of final exam scores from students who underwent the new curriculum was gathered. Test, using sign test, to see if the new curriculum is effective using = 0.05.

   Data: 78 100 75 64 87 80 72 91 89 70 82 76

3. A new training program has been introduced in Infosys and we are interested in seeing if it has had a significant impact on employees' job satisfaction. The job satisfaction of each employee was measured before and after the training program, resulting in the following paired data:

   Before Training: 78,85,72,90,68,76,82,88,79,75
   After Training: 82,88,74,92,70,78,84,90,85,80

   Now, the researcher wants to use the Wilcoxon Matched Pairs Signed Ranks Test to determine whether there is a significant difference in job satisfaction before and after the training program. *Given*: Critical Values for the one-tailed and two-tailed Wilcoxon test for $n = 10$ at significance level $\alpha = 0.05$ are 11 and 8 respectively.

4. Suppose we have a sample of 10 students who underwent a special tutoring program, and we want to determine if there is a significant improvement in their test scores after the program. The null hypothesis is that there is no difference in the test scores before and after the program. Let's consider the differences in test scores for each student (after - before): 2,-1,5,0,3,-2,4,1,0,-3

5. Let us consider the scores of two groups of students (Group A and Group B) using the Mann-Whitney U Test and referencing critical values, for a two-tailed test. *Data*: Group

A: 65, 72, 78, 85, 90 | Group B: 68, 75, 80, 82, 88. *Given*: For the two-tailed Mann-Whitney U Test with $n_1 = 5$ and $n_2 = 5$ for level of significance $\alpha = 0.05$, the critical value is 2.

6. Consider a study examining the time spent on smartphones by two different groups of participants: Group A, consisting of teenagers, and Group B, consisting of adults. The researchers want to determine if there is a significant difference in the time spent on smartphones between these two groups. The daily smartphone usage (in hours) for each group is as follows:

   Group A: 4, 5, 6, 3, 5, 4, 6, 5, 4, 3
   Group B: 2, 3, 4, 2, 3, 2, 4, 3, 2, 3

   Given: The critical value of the two-tailed Mann-Whitney U Test for $n_1 = 10$ and $n_2 = 10$ for $\alpha = 0.05$ is 23.

7. We have three independent groups (A, B, and C), and each group has different students with their test scores.

   Group A: 72, 78, 85, 90
   Group B: 68, 75, 80, 82
   Group C: 60, 65, 70, 76

   Determine if there is a significant difference in the test scores among three teaching methods.

   Given: The critical value for the Kruskal Wallis H test for 2 degrees of freedom and $n_1 = 4$, $n_2 = 4$ and $n_3 = 4$ at $\alpha = 0.05$ is 5.692.

8. A researcher is interested in understanding the relationship between the amount of time students spend studying for an exam and their exam scores. The researcher collects data on the study hours and exam scores for a group of students. The data is as follows:

   Study Hours: 10,15,8,12,18
   Exam Scores: 75,80,65,70,85

   Use Spearman's Rank Order Correlation to assess the strength and direction of the relationship between study hours and exam scores. Use a significance level of 0.05. For the Spearman's test, we take $n - 2$ degrees-of-freedom, which in this case is 3. *Given*: The critical value for Spearman's rank order correlation for 3 degrees-of-freedom at $\alpha = 0.05$ is 1.

9. In a study examining the relationship between academic achievement (A) and hours spent on extracurricular activities (B), researchers discovered that the students' sleep duration (C) might influence this relationship. Conduct a partial correlation analysis to assess the strength and direction of the relationship between academic achievement and extracurricular activity hours while controlling for sleep duration. The data is as follows:

   Academic Achievement (A): 85, 78, 92, 88, 75

Extracurricular Activity Hours (B): 10, 8, 12, 9, 11
Sleep Duration (C): 7, 8, 6, 7.5, 6.5

10. In a study investigating the relationship between study hours and final exam scores, a researcher collected data from a group of students. The goal is to predict final exam scores based on the number of hours spent studying. Perform a linear regression analysis using the principle of least squares and interpret the results. The data is as follows:

Study Hours (in hours): 8, 10, 12, 15, 9
Final Exam Scores: 75, 80, 85, 90, 78

11. *In a study examining the factors influencing product sales, a researcher collected data on the advertising expenditure (in dollars) and the number of retail locations for a sample of stores. The goal is to predict product sales (in units) based on both advertising expenditure and the number of retail locations. Develop a multiple regression model for this scenario and interpret the coefficients. The data relevant for the question is as follows:

Advertising Expenditure: $500, $700, $600, $800, $900
Number of Retail Locations: 2, 3, 2, 4, 3
Product Sales: 1000, 1200, 1100, 1400, 1300