## ISyE 6404 Enrichment Project #1 (EP-1): Rank-based Methods

This is a team project. Please ask a member of your team to submit a PDF file of your report to Canvas> Assignment> EP-1 before the due date. Old reports from a regression class provide examples of writing style.

For the problem below please provide citation of the problem and data collected from the Internet, textbook or other sources. State the goal of a study, problem/data background, descriptions of response- (y) and explanatory-variable (X). When a procedure is used to analyze a data set, please state the intention/*aim* of the procedure and the *assumption(s)* needed to apply the procedure.

*Please include your R- and Matlab-codes in the appendix; specify which part of the codes is for what purpose (e.g., data input, data preparation, WSiRT test, MW test).*

Please include a table-of-contents and **a workload distribution statement** specifying who is responsible for what task(s) in the project. In the end of the semester every member in a team will submit a **self-evaluation report** to provide a feedback about tasks performed on projects. If there is a consistent low-evaluation from several members in a team about a certain individual, his/her grade will be adjusted correspondingly.

1. **Two-sample Studies (40%):**

   Locate *a data set* in the field of your interest, e.g., eCommerce, medical study, drug development, supply-chain/logistics operations, for applying the following procedures for two-sample studies.

   i) Calculate Pearson and Spearman coefficient of correlation and Kendall's Tau. Use a Bootstrap resampling procedure with $B$ = #bootstrap-samples = 1000 to assess the standard deviation (sd) of *three* estimates. Comment on your findings.

   ii) Apply Wilcoxon Signed Rank Test, Wilcoxon Sum Rank Test, Mann-Whitney $U$ Test to compare two samples. For each test please state clearly what distribution is used to calculate the p-value.

   iii) Use Conover test for equal variances in these two samples. Explain how to calculate its p-value.

   iv) Use the parametric $F$-test for equal variances to the data; comment on the difference of the assumptions and results compared to them in (iii).

v) Depending on the outcomes from the F-test in (iv), apply an appropriate parametric two-sample *t*-test to the data; comment on the difference of the assumptions and results compared to them in (ii).

vi) Apply Kolmogorov-Smirnov, Anderson-Darling, Cramer-Von Mises, Shapiro-Wilk Tests for normality to the two samples separately; comment on the findings by *comparing results obtained from these four tests*. Make a statement about the situation that a particular procedure might be more appropriate. Moreover, based on the results learned here, comment on whether the parametric methods used in (iv) and (v) are appropriate.


2. **Multiple-Sample (ANOVA) Studies (60%):**

Locate *one data set each for the two problems below* in the field of your interest, e.g., eCommerce, medical study, drug development, supply-chain/logistics operations, for applying the following procedures for ANOVA studies.

i) Apply Kriskal-Wallis Test for an one-way ANOVA study. If it is suitable, perform a K-W pairwise comparisons. Make conclusions about your findings.

ii) Use Friedman test and also the F-Test discussed in the textbook page 148 for the study of one-way ANOVA with one blocking variable. Comment on your findings. If it is suitable, perform a K-W pairwise comparisons. Make conclusions about your findings.

iii) Conduct a variance test based on the procedure (Conover test) given in Section 8.3 textbook. Comment on your findings.

iv) Repeat the same studies in (i), (ii) and (iii) using parametric approaches (also include the possible pairwise comparisons). State the assumptions needed for the parametric approaches. Compare the results here against those in (i), (ii) and (iii), respectively. Note that if there are certain assumptions (e.g., normality and equal-variance) required in the parametric studies, please apply appropriate procedures to "test" the assumption.