DSSC 221 – Probability and Statistics
Lab 10

The data files "tt_ug.csv" and "tt_g.csv" are in LMS. The first file contains the results of the travel time survey for *undergraduate*, and the second contains results for *graduate* students.

Assume that these observations represent random samples drawn from the populations of KU students.

For all questions, decide what parameter you're testing for, what you know/don't know, and therefore which equation to use in calculating the test statistic and doing the hypothesis test.

Part 1: Data Preparation

a) Import "tt_ug.csv" and "tt_g.csv" into Python.
b) Find how many observations are in each sample.
c) Find the mean of each sample.
d) Find the variance of each sample.
e) **Challenge Question (Extra pts.) -** Find the pooled variance.

Part 2: Testing equality of means, undergraduate versus graduate students

Assume that the variances of the undergraduate and graduate students' travel times are equal. Test the hypothesis that undergraduate and graduate students have the same average travel time to the library (H_o: $\mu_{ug} = \mu_g$) against

a) The hypothesis that the average travel times of undergraduate and graduate students are different (H_1: $\mu_{ug} \neq \mu_g$), at the 5% significance level. What is the p-value?

b) **Challenge Question (Extra pts.) -** The hypothesis that the average travel times of graduate students are longer (H_1: $\mu_{ug} < \mu_g$), at the 5% significance level.

Part 3: Testing equality of variances

Challenge Question (Extra pts.) - Test the hypothesis that the variances of undergraduate and graduate students' travel times to the library are equal (H_o: $\sigma_{ug}^2 = \sigma_g^2$) against the hypothesis that the variances of undergraduate and graduate students are different (H_o: $\sigma_{ug}^2 \neq \sigma_g^2$), at the 5% significance level.

Hint: Use F-statistic which defines $F = \frac{S_{ug}^2}{S_g^2}$. This follows F-distribution with $(n_{ug} - 1, n_g - 1)$ degree of freedom.