HUDM 6026 - Computational Statistics
HW 07 – Cross Validation and Permutation Test

Instructions.

- You may use whatever software you like to write up your answers to the hw. For example, you might choose to use Rmarkdown for assignments that are primarily code-based; whereas, you might choose to use Word or LaTeX for more writing-heavy assignments. It's up to you.
- No matter which program you use, you should turn in a pdf or html version.
- This is an individual assignment (not a group assignment). Consult the syllabus for the rules about collaboration.

1. Do some research on the two-sample Kolmogorov-Smirnov (K-S) test for equality of distributions. Describe the null and alternative hypotheses and discuss how the test statistic is computed.
2. Install package **MatchIt** and load it. Then call `data(lalonde)`. Examine the help on `lalonde` and describe the meaning of the `treat` and `re78` variables.
3. Run the two-sample K-S test to test if participant income in 1978 is identically distributed across treatment group assignment or not. Use `ks.test()` to do this. Report the value of the test statistic, D, along with the p-value. Interpret the result in context.
4. Create a function that takes two arguments, x and y, each a vector of values, and outputs the value of the two-sample K-S statistic D.
5. Run an approximate permutation test with $B = 1000$ permutation replications to determine the estimated ASL for testing the null hypothesis that participant incomes are identically distributed across treatment arms. Use $\alpha = 0.05$.
6. Plot a histogram of the permutation distribution created by applying the K-S statistic to the $B = 1000$ permutation replications.