**DSC324/424**
**Assignment #3 (Due Sunday, August 16th, 2020 at midnight)**

**1) (20 points, for data projects)** Choose a technique that we have covered so far in this course, and try applying that technique to your data. You may choose any of

a) Model building and Multiple Regression
b) PCA
c) CFA
d) CCA
e) CA (correspondence analysis)

If you are working as a group, each member of your group should try a different technique, or the same technique with different aspects of the data.

2) **Paper Review (10 points):** An academic paper from a conference or Journal will be posted to the Homework 3 content section of D2L. It contains a usage of Canonical Correlation. Review the paper and evaluate their usage of Canonical Correlation. In particular, address **(CCA of Regional Economic Development Based on the MVC Framework in Weka Platform)**

a) How suitable is their data for CC?
b) How are they applying CC? What two groups of variables are being correlated? Are they metric, ordinal, nominal?
c) What methods do they use to judge the quality of the correlation? Do they evaluate, and how do they evaluate the stability of the components?
d) How many correlates do they concentrate on in their analysis, and do they attempt to interpret the correlates in terms of the original variables?
e) What conclusions does CC allow them to draw?

3) **(20 points):** Perform the following Canonical Correlation Analysis on the Young People Survey from Lab 2: PCA/FA. Perform a canonical correlation analysis describing the relationships between the hobbies_interests and music variables using the data under the Lab 2: PCA/FA in the content folder).

    1. Answer the following questions regarding the canonical correlations.

        a. Test the null hypothesis that the canonical correlations are all equal to zero. Give your test statistic, d.f., and p-value.

        b. How many significant canonical variates are there?

        c. Present the first two canonical correlations (Cancor)?

        d. What can you conclude from the above analyses?

    2. Answer the following questions regarding the canonical variates.

        a. Give the formulae for the first canonical variate for the hobbies_interests and music variables.

        b. Give the correlations between the first canonical variate for hobbies_interests and the music variables.

        c. What can you conclude from the above analyses?

**EXTRA CREDIT (10 points)** Perform a correspondence analysis on the countries and sports liking data in Sports.csv. In this file you are provided with the table for the two sets of categories. In particular perform the following

a) Create a mosaic plot of the two categorical variables.
b) Plot the results of the correspondence analysis
c) With each country, create a profile for the sports likings. Which sports liking are most highly and least highly represented. For each country, draw the scale for that country and demonstrate that sports liking profile on the graph.