CSC334/424
Assignment #3 (Due Sunday, July 8th, 2018 at midnight)

1A) (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.

1B) (20 points, for literature review projects) Choose a technique that you will be covering in your literature review, research it, and write two paragraphs discussing when it is used, how does the technique work, and how do you interpret its results.

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 **(Vacation Benefits and Activities Understanding Chinese Family Travelers)**

   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. Individual samples were collected (n = 465) in a survey relating to attitudes and health variables. The following *Attitudinal* variables were measured:

control = Locus of control
attmar = Attitudes toward current marital status
esteem = Self esteem
attrole = Attitudes toward role of women

In addition, the following *Health* variables were measured:

menheal = Mental health symptoms
phyheal = Physical health symptoms
timedrs = Visits to health professionals
attdrug = Attitude toward use of drugs
druguse = Use of psychotropic drugs

Perform a canonical correlation analysis, describing the relationships between the attitudinal and health variables using the data under the course documents for HW3).

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

    c. Present the two canonical correlations

    d. What can you conclude from the above analyses?

2. Answer the following questions regarding the canonical variates.

    a. Give the formulae for the significant canonical variates for the attitudinal and health variables.

    b. Give the correlations between the significant canonical variates for attitudinal and the attitudinal variables, and the correlations between the significant canonical variates for health and the health 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.