

CSC334/424

Assignment #3 (Due Sunday, May 6th, 2018 at midnight)

Note: For each of the analysis problems, include a copy of the full analysis in your report along with your conclusions.

- 1) **(20 points, Group Submission, Due Friday, May 4th by midnight, and to be discussed with me during that class session)** For the first set of milestones on your final project, meet as a group and complete the following
 - a) Record minutes of the discussion
 - b) Choose a group liaison who will be responsible for interaction with me on group project issues
 - c) Perform an initial analysis on your data and summarize the data for its content. What are the variables involved and **what do they mean**? Don't just list the data dictionary from the download site. I want you to think and analyze these variables for how they might interact with others. Answer the following questions:
 - Is there an obvious parameter of interest, or are there several
 - What interesting metric variables does the data have
 - What interesting categorical variables does the data have
 - Are some of the variables ordinal (ordered but not metric like we discussed in class?)
 - What missing values are there and are there any patterns that you might exploit for filling them in?
 - What variables look like they might interact?
 - For the metric variables, what units are used and are different variables measured in related but different units? (e.g. cm and mm or feet and miles)?
 - For some of the categorical variables, what dependencies are there between any metric variables (parameters of interest?). Explore with boxplots, or if you need to investigate categorical/categorical interactions, look at correspondence analysis.
 - What obvious directions for investigation present themselves?
 - d) Each team-member should choose some initial individual direction to accomplish for the individual part (problem 2). These do not have to be incredibly deep at this point, but should show some progress towards determining how variables relate, what kinds of latent variables might be present or interaction between metric and non-metric variables.

Submit a group report on the Homework 3 – Part 1 dropbox where you detail your group's progress milestones a) through d) milestones. Each group member must submit an individual analysis of their aspect of the data with their homework.

- 2) **(20 points, Individual, to be turned in with the rest of the homework)** 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)

Each member of your group should try a different technique, or the same technique with different aspects of the data.

- 3) **Paper Review (20 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 **(Decent Work and Work Motivation in Lawyers An Empirical Research)**
- 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?
- 4) **Paper Review (20 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?

- 5) **(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.