

Assignment 3

Due: 7/8/2018

Student name: Lavinia Wang #1473704

Problem 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?

Chinese family travelers were the population of interest in the study; thus, purposive sampling was used. The data collection consisted of two phases. In the first phase with participants at age 18 or older, a total of 253 valid questionnaires were returned, yielding a response rate of 83.8%. The second data collection phase was implemented by recruiting families with younger children, making the total number of sampled families 306. One set of variables has 19 attributes and the other has 32 attributes. Canonical correlation analysis is deemed as appropriate and especially useful when the researcher has limited a priori knowledge about whether the two sets of variables are related and how strong the relationship is.

b) How are they applying CC? What two groups of variables are being correlated? Are they metric, ordinal, nominal?

The study explored the interrelationships between the benefit-sought domains (including Communication and Togetherness, Shared Exploration, Escape and Relaxation, and Experiential Learning for Children) and destination activities. The results identified significant linkages between the two.

Exploratory factor analysis was then performed on the 19-item measurement of benefits to identify possible underlying dimensions. At the next stage, canonical correlation analysis was used to assess the nature and magnitude of the relationship between benefits sought and vacation activities. These are nominal data.

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?

Baseline statistical assumptions including sample size, linearity, and multicollinearity were checked to ensure the appropriateness of canonical correlation analysis for this study.

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?

Four separate analyses were performed between the benefit items of each of the four benefit factors and the 32 activities. Each benefit dimension was treated as one set, and activity items constituted the other set.

The first canonical variate pair shows a significant relationship between taking pictures and videos and four items under the factor of Communication and Togetherness. This variate indicates that the pursued benefits of enhanced communication and togetherness seem to be related to the activity of digitally capturing and recording family trip experiences.

The second canonical variate reveals a significant relationship between Shared Exploration and activity participation. This suggests that the sought benefit of sharing and exploring new things together is matched with a set of family activities that contribute to both intellectual exploration and utilitarian gratification.

The third significant canonical variate pair consists of two benefit items, escaping from the routine life (.75) and relaxing (.62), both from the factor of Escape and Relaxation, and two activity items,

Assignment 3

Due: 7/8/2018

Student name: Lavinia Wang #1473704

canoeing and kayaking and farm visits and agritourism sites (.42). This result suggests that families who consider escaping and relaxation as important vacation benefits are more likely to participate in physical activities offered in outdoor, nature-based environments.

In the fourth canonical variate, a significant relationship exists between the three items in the benefit dimension of Experiential Learning for Children and seven activity items. These results show that benefit items relating to children's intellectual development and holistic learning are associated with mostly culture- and nature-based activities, as well as some consumptive activities at the destinations.

e) What conclusions does CC allow them to draw?

The prominence of picture and video taking as a vacation activity for Chinese family travelers is evident. The benefit-seeking and vacation activity choices of Chinese families may be influenced to some degree by the child-centric and education-centric disposition of the culture.

Problem 3 (20 points): 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.

	wilksL	F	df1	df2	p
[1,]	0.80	5.20	20	1513.33	0.00
[2,]	0.93	2.71	12	1209.40	0.00
[3,]	0.99	0.54	6	916.00	0.78
[4,]	1.00	NaN	2	NaN	NaN

The first tells that there is at least one difference from zero. Degree of freedom equals 20 with p value 0.00, which is significantly less than 0.05.

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

The second suggests that there is some difference from zero among the last three canonical correlations. F value equals to 2.71. Degree of freedom is 12 and p-value is 0.00, which is less than 0.05

- c. Present the two canonical correlations

```
> cc$cor
[1] 0.37345882 0.24745350 0.07819567 0.03104307
```

- d. What can you conclude from the above analyses?

After the second canonical correlations, the p value is no longer statistically significant at 0.05 level. Therefore, there should be two canonical variates generated from the variables, one represents the attitudinal set and the other represents the health set.

Assignment 3

Due: 7/8/2018

Student name: Lavinia Wang #1473704

2. Answer the following questions regarding the canonical variates.

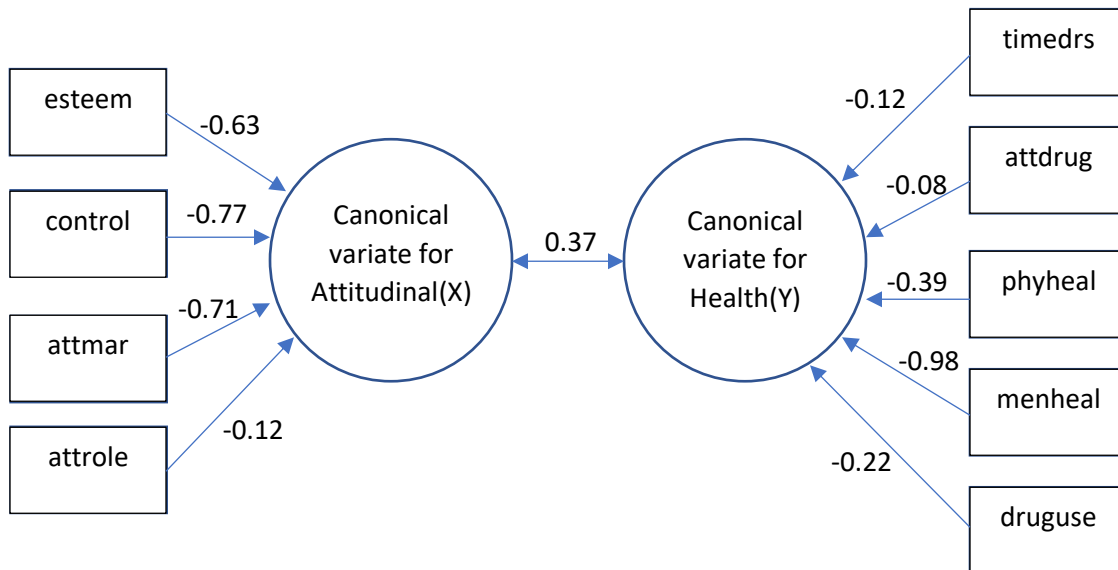
a. Give the formulae for the significant canonical variates for the attitudinal and health variables.

```
> round(cc$xcoef, 3)
      [,1] [,2] [,3] [,4]
esteem -0.079 -0.180  0.206 -0.042
control -0.446  0.035 -0.379  0.611
attmar  -0.058  0.058 -0.040 -0.086
attrole  0.019 -0.079 -0.121 -0.045
> round(cc$ycoef, 3)
      [,1] [,2] [,3] [,4]
timedrs  0.008  0.017  0.081  0.000
attdrug -0.057 -0.735 -0.096  0.400
phyheal  0.028 -0.103  0.145  0.224
menheal -0.259 -0.013 -0.022 -0.090
druguse  0.015  0.085 -0.061  0.052
```

Attitudinal = $-0.079 \cdot \text{esteem} - 0.446 \cdot \text{control} - 0.058 \cdot \text{attmar} + 0.019 \cdot \text{attrole}$

Health = $0.008 \cdot \text{timedrs} - 0.057 \cdot \text{attdrug} + 0.028 \cdot \text{phyheal} - 0.259 \cdot \text{menheal} + 0.015 \cdot \text{druguse}$

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.



Assignment 3

Due: 7/8/2018

Student name: Lavinia Wang #1473704

```
> loadings$corr.X.xscores
      [,1]      [,2]      [,3]      [,4]
esteem -0.6348309 -0.6439979  0.3886898 -0.1847383
control -0.7689471 -0.1040628 -0.2678367  0.5737731
attmar  -0.7129370  0.3336807 -0.1056965 -0.6051434
attrole  0.1191717 -0.7164603 -0.6339145 -0.2670676

> loadings$corr.Y.yscores
      [,1]      [,2]      [,3]      [,4]
timedrs -0.11823858  0.28151460  0.79322551  0.3757976
attdrug -0.08445857 -0.59978939 -0.17076486  0.6640656
phyheal -0.38882908  0.08083039  0.40933456  0.6348407
menheal -0.97959013  0.13655969  0.06755947  0.1310460
druguse -0.22906107  0.53887129 -0.22406244  0.7607851
```

c. What can you conclude from the above analyses?

For predictor variables(attitudinal), self-esteem, locus of control and attitudes toward current marital status load most strongly (typically look for 0.3 and above). For dependent variables(health), mental health symptoms variable has the highest correlations with its variate followed by mental health symptoms.

EXTRA CREDIT (10 points) Perform a correspondence analysis on the countries and sports liking data in Sports.csv. In particular perform the following:

a) Create a mosaic plot of the two categorical vari

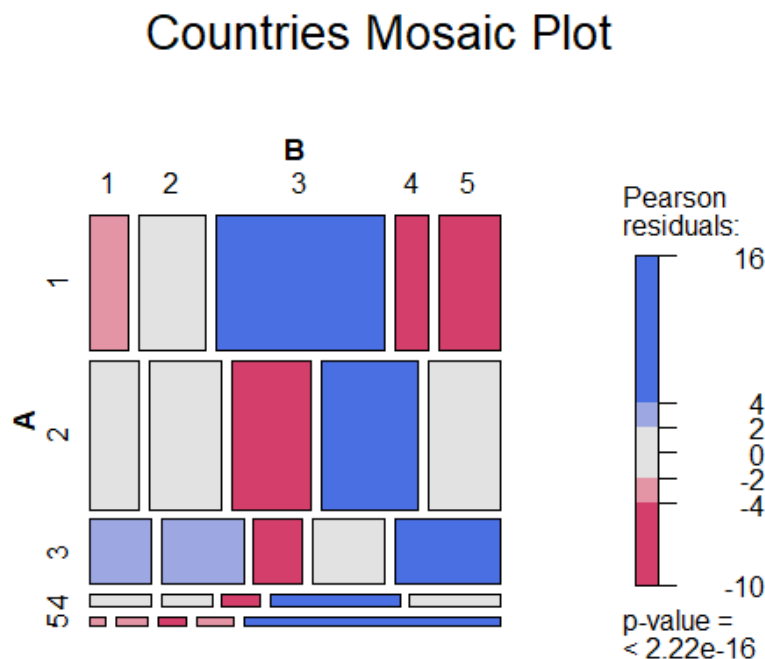


Figure 1 Mosaic plot of countries vs sports liking

b) Plot the results of the correspondence analysis

Assignment 3

Due: 7/8/2018

Student name: Lavinia Wang #1473704

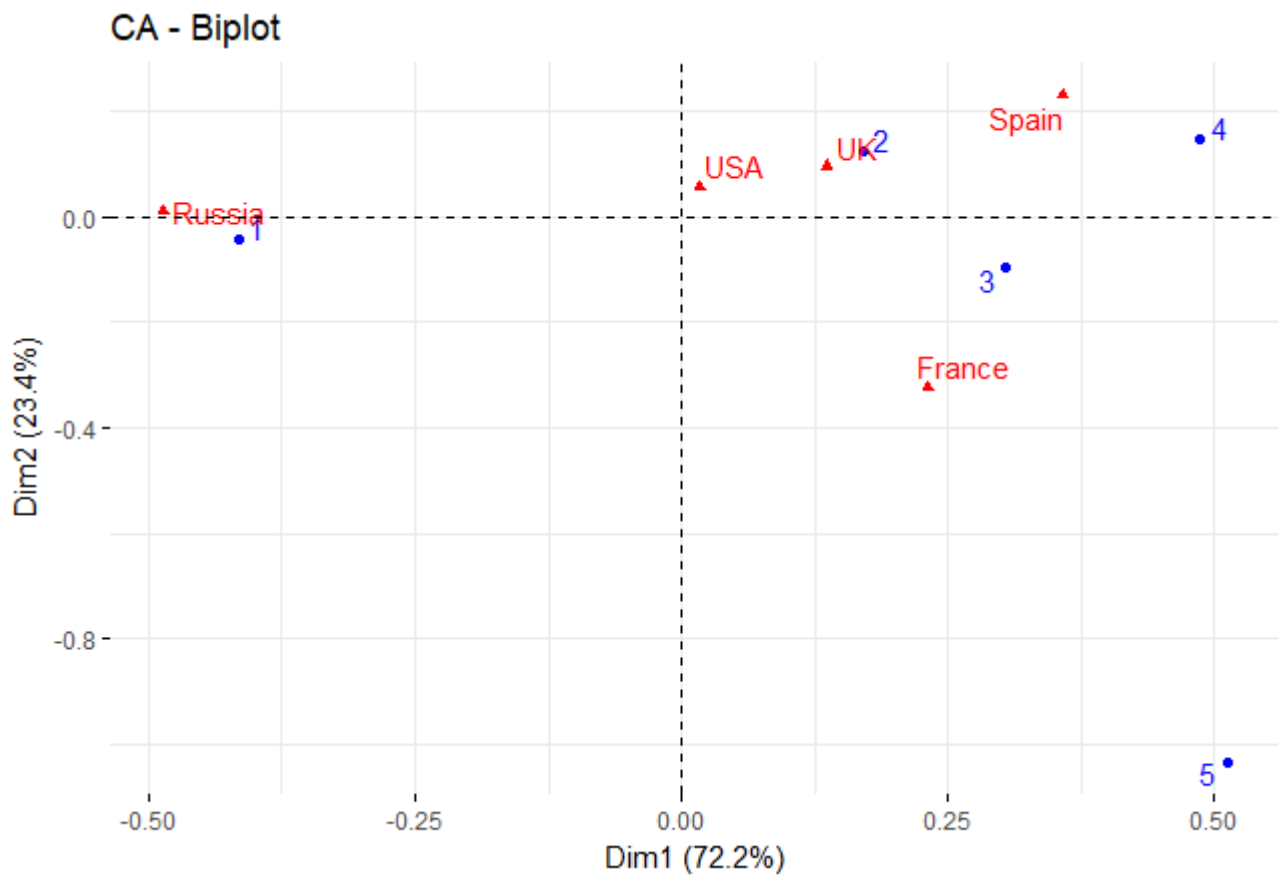


Figure 2 Correspondence analysis result