

Ronaldlee Ejalu

Student ID: 2020637

DSC324/424

Assignment #3 (Due Sunday, February 21, 2021 at midnight)

1. **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 **(The Association of Work Satisfaction and Burnout-CCA)**

a) How suitable is their data for CC? Their data is suitable enough since they are using a sample size of 674 endoscopy staff members.

b) How are they applying CC?

They are applying canonical correlation analysis to examine the association between work satisfaction and burnout in endoscopy nursing staff.

c) What two groups of variables are being correlated?

The two groups are work satisfaction and burnout.

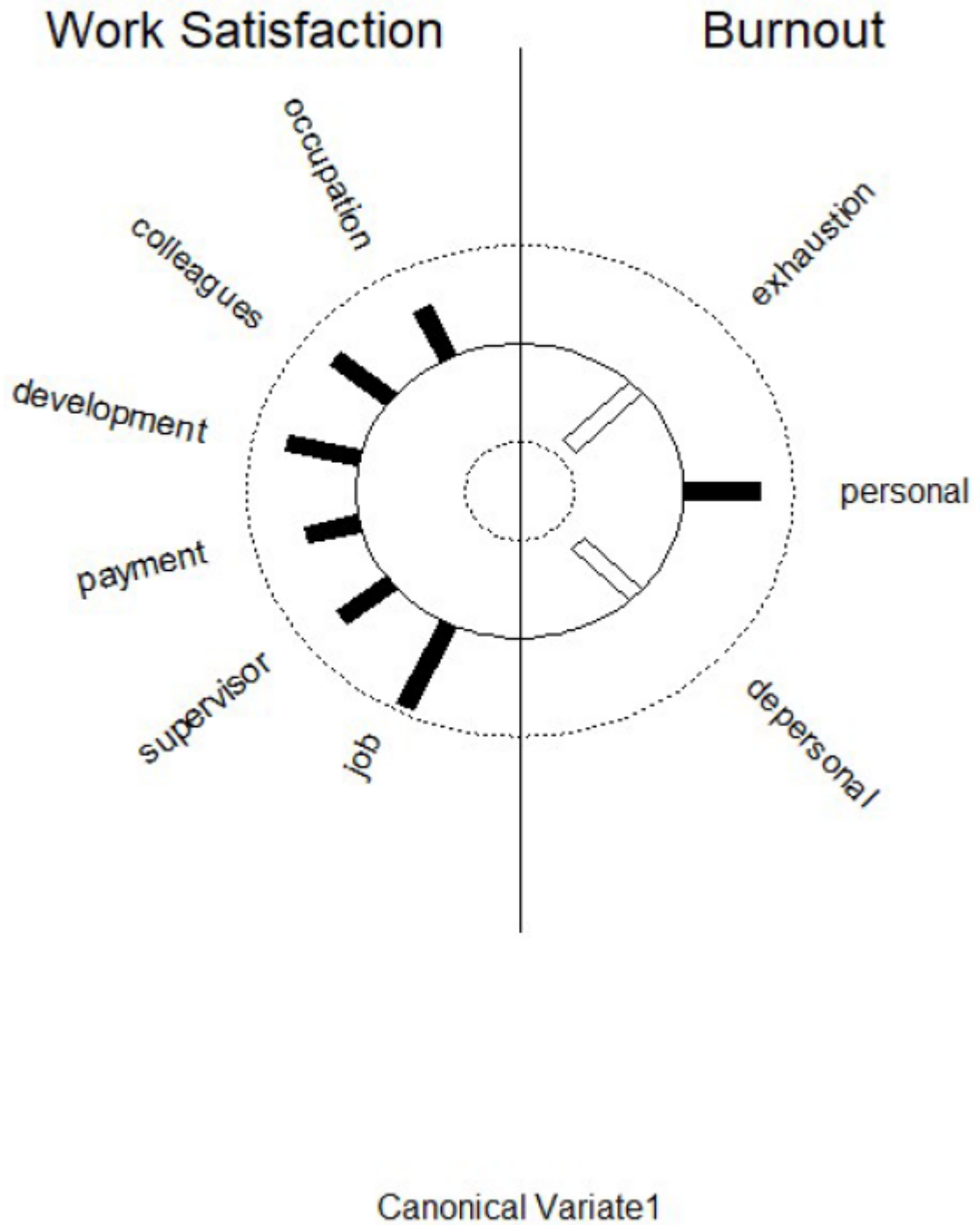
Are they metric, ordinal, nominal? They are ordinal

d) **What methods do they use to judge the quality of the correlation?**

First, the variables used in the study had a factor loading over the recommended cut-off values of 0.3, their ranges were between 0.37 and 0.90.

Second, they use a Helio plot, which shows a graphic display of structure correlations of the KAFA scales on the first predictor canonical variate and of the structure correlations of the MBI scales on the first criterion canonical variate. Black bars correspond to positive correlations and white bars correspond to negative correlations as shown below:

Helio Plot



Do they evaluate, and how do they evaluate the stability of the components?

They use classical test theory, such as Cronbach's alpha, discriminatory power or average intercorrelations.

- e) How many correlates do they concentrate on in their analysis, and do they attempt to interpret the correlates in terms of the original variables?

They concentrate on three canonical functions, with canonical correlations of 0.64, 0.32 and 0.17, but in actual sense they only concentrate on the first two because the third one was low with a correlation of 0.17, which they didn't even bother interpreting. Yes, they do attempt to interpret the correlates in terms of the original variables.

- f) What conclusions does CC allow them to draw?

First, Canonical Correlation revealed interesting results that had important consequences for clinical management. These results support the hypothesis of being burn out being a multidimensional construct, which must be investigated. Second, Canonical Correlation, also allows them to come up with various interventions aimed at reducing of burnout and to increase work satisfactions amongst endoscopy nursing staff. Also, CC allows to use them to use structural equation model to establish causation between locus of control and work-related stress and this enables them to conclude that both work-related stress and burn out contribute to poorer locus of control.

In summary, the results of Canonical Correlation could contribute to the design of various interventions that could be employed to address the issue of work satisfaction and burn out in endoscopy staff most effectively.

2. **(20 points):** Perform the following Canonical Correlation Analysis on the Young PeopleSurvey from Lab 2: PCA/FA. Perform a canonical correlation analysis describing the relationships between the hobbies_interests and spending 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.

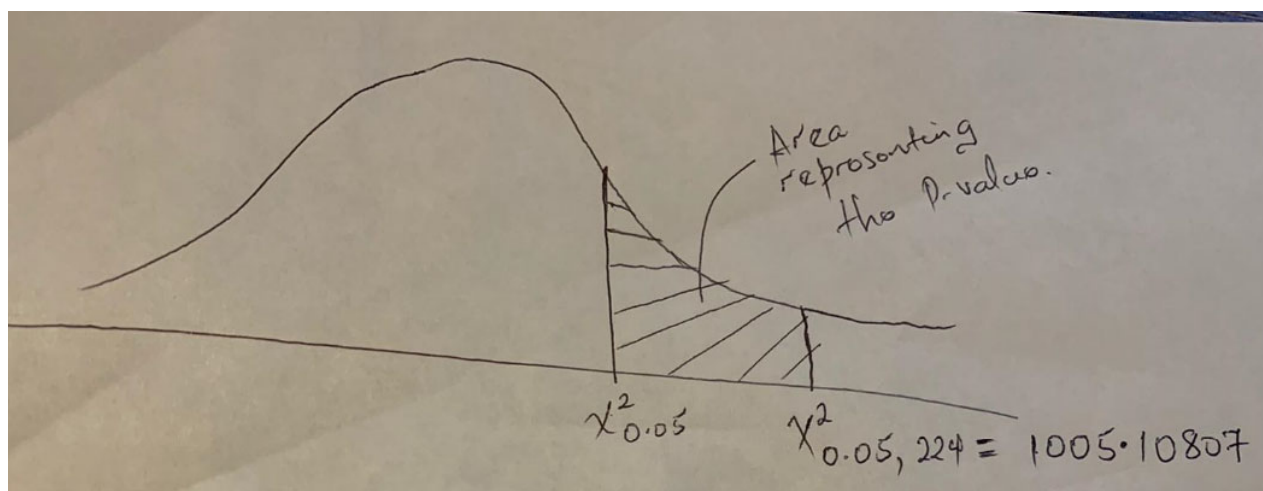
From the R computations below:

```
142 #calculate the chisquare test
143 ```{r}
144 round(pchisq(c2$chisq
145             , c2$df
146             , lower.tail=F
147             )
148             , 3
149             )
150 ...
```

CV 1	CV 2	CV 3	CV 4	CV 5	CV 6	CV 7
0.000	0.000	0.000	0.070	0.617	0.947	0.979

This shows that CV 1, CV 2, CV 3 are the only variates who p-values is less than 0.05

Looking the diagram below:



Shows whatever the critical value will be, the chi-square test lands on the upper tail as indicated above, in between will be the rejection region, which represents

the alpha of 5 percent of error and all the p-values for CV 1, CV 2 and CV 3 are less than 5 percent, which means that they fall within the rejection region implying that we should using the first three variates.

The Chi-square test statistic are recorded as:

For CV 1 - $\chi^2_{0.05,224} = 1005 \cdot 10$ and the p-value is 0.000

CV 2 $\chi^2_{0.05,186} = 462 \cdot 48$ and the p-value is 0.000

CV 3 $\chi^2_{0.05,150} = 246 \cdot 66$ and the p-value is 0.000

b. How many significant canonical variates are there?

They are three significant variates, which are CV 1, CV 2 and CV 3

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

0.75 for CV 1 and 0.53 for CV 2

d. What can you conclude from the above analyses?

If you think of this as Pseudo R squared: $100 - 75 = 25\%$

The 25 percent represents error in the model meaning that we might not have everything that explains spending for variate CV 1.

Also, $100 - 53 = 47\%$ represents the error in the model implying that we might now have everything that explains spending for CV1

2. Answer the following questions regarding the canonical variates.

a. Give the formulae for the first canonical variate for the hobbies_interestsand spending variables.

Hobbies_interests

$$\begin{aligned}
CVI = & 0.145X_{\text{History}} - 0.009X_{\text{Psychology}} + 0.083 \\
& X_{\text{Politics}} + 0.248X_{\text{Mathematics}} + 0.257X_{\text{Physics}} \\
& + (-0.028)X_{\text{Internet}} + 0.247X_{\text{PC}} - 0.187X_{\text{Economy Mgt}} \\
& - 0.138X_{\text{Biology}} - 0.029X_{\text{Chemistry}} - 0.031X_{\text{Reading}} \\
& + 0.058X_{\text{Geography}} - 0.174X_{\text{Foreign Languages}} - 0.055X_{\text{Medicine}} \\
& - 0.142X_{\text{Law}} + 0.045X_{\text{Cars}} - 0.094X_{\text{Art Exhibitions}} \\
& + 0.073X_{\text{Religion}} + 0.059X_{\text{Countryside.. outdoors}} \\
& - 0.301X_{\text{Dancing}} + 0.026X_{\text{Musical Instruments}} \\
& - 0.062X_{\text{Writing}} - 0.038X_{\text{Passive Sport}} - 0.072X_{\text{Active Sport}} \\
& - 0.079X_{\text{Gardening}} - 0.546X_{\text{Celebrities}} - 0.967X_{\text{Shopping}} \\
& + 0.153X_{\text{Science and Technology}} \\
& - 0.133X_{\text{Theatre}} - 0.148X_{\text{Fun with friends}} \\
& + (-0.063)X_{\text{Adrenaline sports}} + (-0.205)X_{\text{Pets}}
\end{aligned}$$

Spending

$$\begin{aligned}
 CV1 = & 0.096X_{\text{Finances}} - 0.903X_{\text{shopping-centres}} \\
 & - 0.304X_{\text{Branded-clothing}} - 0.066X_{\text{Entertainment-spending}} \\
 & - 0.732X_{\text{spending-on-looks}} - 0.033X_{\text{spending-on-gadgets}} \\
 & - 0.088X_{\text{spending-on-healthy-eating}}
 \end{aligned}$$

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

Hobbies_interests

Structural Correlations (Loadings):

X Vars:	CV 1
History	0.144616124
Psychology	-0.008500419
Politics	0.082611865
Mathematics	0.247601799
Physics	0.256781864
Internet	-0.027648295
PC	0.246784835
Economy.Management	-0.187121670
Biology	-0.137794217
Chemistry	-0.028907505
Reading	-0.030896539
Geography	0.057778631
Foreign.languages	-0.173546229
Medicine	-0.055182491
Law	-0.141783551
Cars	0.045068648
Art.exhibitions	-0.093648453
Religion	0.072691528
Countryside..outdoors	0.058770496
Dancing	-0.300649613
Musical.instruments	0.025619023
Writing	-0.061687937
Passive.sport	-0.037614249
Active.sport	-0.071569875
Gardening	-0.078674653
Celebrities	-0.546270137
Shopping	-0.966803839
Science.and.technology	0.153422903
Theatre	-0.133480440
Fun.with.friends	-0.148223795
Adrenaline.sports	-0.062827647
Pets	-0.204902197

Spending:

Y Vars :	CV 1
Finances	0.09605158
Shopping.centres	-0.90272895
Branded.clothing	-0.30369881
Entertainment.spending	-0.06648889
Spending.on.looks	-0.73242054
Spending.on.gadgets	-0.03338746
Spending.on.healthy.eating	-0.08826853

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

When you look at the correlations for CV 1 for hobbies_interest, shopping is most important variable in this variate and there is a strong negative relationship between shopping and the CV 1 meaning that as shopping increases, the variate increases but in the opposite direction.

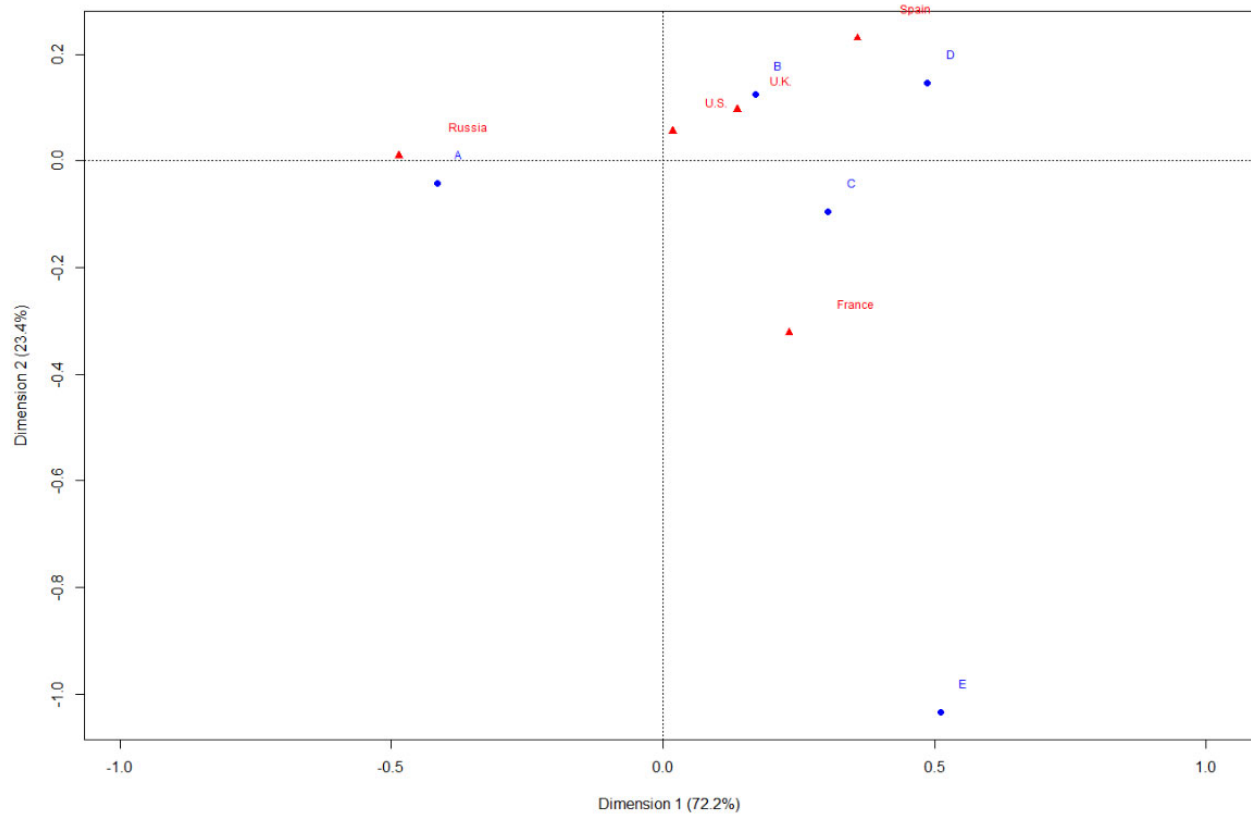
Also, on the spending side, Finance variable is the most important in this variate and there is a strong positive relationship between this variable and the variate.

Spending on gadgets yields little information about the data.

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

Create a mosaic plot of the two categorical variables.

The mosaic plot for countries:



Plot the results of the correspondence analysis

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.