**RESEARCH AND SURVEY STATISTICS – STA3022F**

**COMPUTER PRACTICAL 4**

**CORRESPONDENCE ANALYSIS (CA and MCA)**

Note on R packages

This practical uses the R package ca and FactoMineR.

Install the packages first (Select line 1 and run).

Once the package is installed, you need to load it every time you use R with the command library(ca)

Library(FactoMineR). (Select line 2 and 3 and Run)

1. Open the script file **prac4CAMCA.R**.
2. Example CA

We will use the tea dataset in FactoMineR package.

Select line 5 and 6 and Run

To see what the data looks like:

Select line 7 and Run

To construct the two-way table between Tea and where variables:

Select line 9 and 10 and Run

In order to view the table with row and column totals, we use the addmargins() command. Run line 11.

1. First we perform the -test of association for : no association between Tea and where vs : there is significant association between Tea and where.

Select line 14-15 and Run for the simple ca dataset.

Select line 17 and 18 and Run for the test. The p-value of 0.014 indicates that there is some statistically significant association between tea and where.

Perform the CA, run lines 20 to 22. When we construct a 2D CA map, we see that 98.8 + 1.2 = 100% of the inertia is accounted for.

Interpret the CA map

1. Example MCA

We will use the same tea dataset.

Import the data set mailorder.csv. To perform Joint Correspondence Analysis, the most efficient form of MCA, run lines 28 and 29.

To obtain the singular values

Select line 30 and run

To obtain the eigenvalues (inertias or variances – squares of d’s)

Select line 31 and run

To obtain the variances explained

Select line 32 and run

To plot the two dimensions:

Select line 34 and run

To see what kind of outputs you can get using mjca() function:

Select lines 36 and 37 and run

1. Donor data analysis for quiz:

Upload the donor data first

1. We will use the variable DOMAIN for the row categories of the frequency table we are going to create.

The variable DOMAIN is coded as follows:

1st character = Urbanicity level of the donor's neighbourhood

U=Urban

C=City

S=Suburban

T=Town

R=Rural

2nd character (number) = Socio-Economic status of the neighbourhood

1 = Highest SES

2 = Average SES

3 = Lowest SES

except for Urban communities, where 1 = Highest SES

2 = Above average SES,

3 = Below average SES

4 = Lowest SES

We are going to remove the donors with missing values, with the code in lines 41 to 45.

The column categories will be the interests reported by the donors. As you can see, a single donor can have interests in both Collections and Veterans, and so contribute to two columns. The different interests are given by the variables (columns):

COLLECT1 Collectables

VETERANS Veterans

BIBLE Bible reading

CATELOGUE Shop by catelogue

HOMEE Work from home

PETS Household pets

CDPLAY CD player owners (Remember this data was collected in 1997, today’s equivalent question might be do you own an i-phone?)

STEREO Stereo / Records / Tapes / CD

PCOWNERS Home user of PC

PHOTO Photography

CRAFTS Crafts

FISHER Fishing

GARDENIN Gardening

BOATS Power boating

WALKER Walk for health

KIDSTUFF Buys children’s products

CARDS Stationary / Cards buyer

PLATES Plate collector

Run the code in lines 47 to 66 to create the table interests. Line 66 shows the content of the table.

Now perform a -test and create a CA map of the table interests with code similar to lines 16 to 22.

1. Next, we will return to the detailed information on major donors only (DERIVED DATASET). The variable (column) MDMAUD (Major donor matrix) contains codes describing the frequency and amount of giving for donors who have given a 100+ donation at any time in their giving history.

1st character = Recency of giving

C=Current donor

L=Lapsed donor

I=Inactive donor

D=Dormant donor

2nd character = Frequency of giving

1=One gift in the period of recency

2=Two to four gifts in the period of recency

5=Five+ gifts in the period of recency

3rd character = Amount of giving

L=Less than $100 (low)

C=$100-$499 (core)

M=$500-$999 (major)

T=$1000+ (top)

Use the code in lines 72 and 73 to extract this information into a data matrix. The code in line 74 shows the first few rows of the data set.

Use code similar to lines 28 to 37 to perform a Joint Correspondence Analysis on the data.

Complete the quiz: Prac 3 – Quiz.

**NOTE: After installing packages, run library(ca) and library(FactoMineR). If you get the following error:**

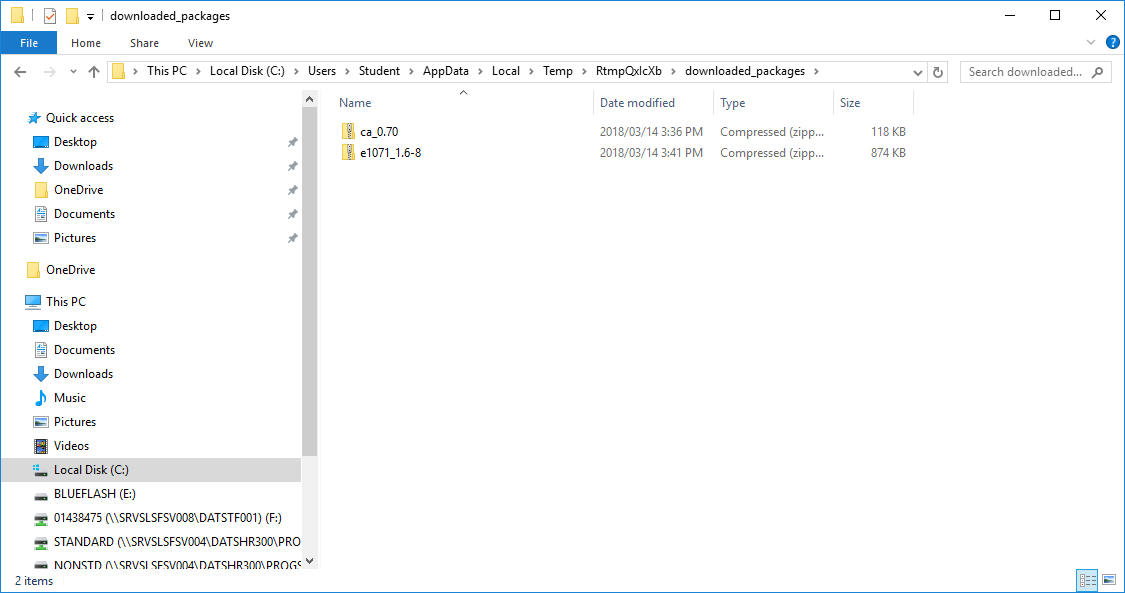
**“unable to move temporary installation C:\Users\STUDENTNUMBER\Documents\R\win-library\3.4\file…\ca to C:\Users\STUDENTNUMBER\Documents\R\win-library\3.4\ca”**

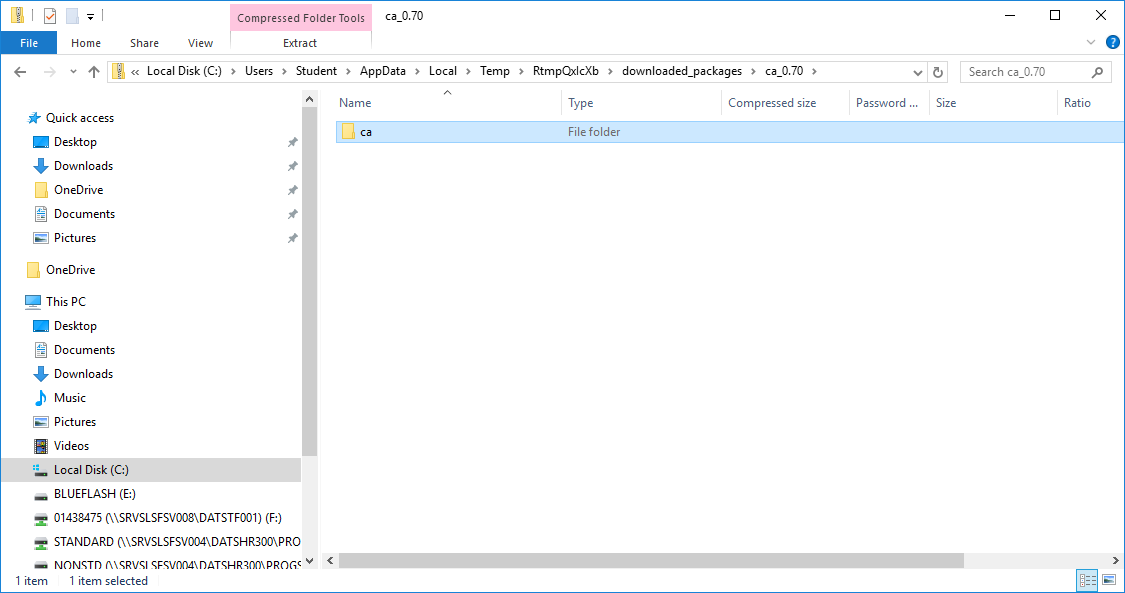
**PLEASE FOLLOW THESE STEPS:**

1. NAVIGATE TO THE FOLDER:

C:\Users\STUDENTNUMBER\AppData\Local\Temp\Rtmp..\downloaded\_packages\ca\_0.70

As shown below:





COPY THE CA FOLDER from the directory to the following directory:

**C:\Program Files\R\R-3.4.2\library**

