Chapter 15: PCA and Factor Analysis

**Principal Components Analysis (PCA)**

The data in the mtcars.csv file was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles.

The variables are as follows:

mpg Miles per(US) gallon

cyl Number of cylinders

disp Displacement (cu.in.)

hp Gross horsepower

drat Rear axle ratio

wt Weight (1000 lbs)

qsec 1/4 mile time

vs Engine (0 = V-shaped, 1 = straight)

am Transmission (0 = automatic, 1 = manual)

gear Number of forward gears

carb Number of carburetors

Conduct a Principal Components Analysis (PCA) to reduce the number of variables that describe the characteristics of the automobiles.

Describe the assumptions test results and the PCA results.

Data set-up

library(psych)

data = read.csv("mtcars.csv")

data$X = NULL # Deletes the text column with car names

 # For data handling tips, see Chapter 21

Use KMO to examine the sampling adequacy of the data set, as shown in the book.



The variables all meet the recommended threshold, so no deletions are required before continuing.

Run the Bartlett test, as shown in the book.



The Bartlett’s test result allows us to reject the null hypothesis that all the correlations tested simultaneously are ’not statistically different from zero’. This result is necessary to carry on with the interpretation of the analysis’ results.

Follow the instructions in the book to see a neat correlation matrix, which may help if you want to analyze the data more closely.



Continue with the instructions to see the eigenvalues, which indicate 2 usable components.



and the summary of the PCA:

 

The read-out (examine 'Comp.1' and 'Comp.1') indicates that the largest component accounts for 60% of the variance, with the second accounting for 24%. The cumulative proportion figures demonstrate that together they account for 84% of the variance. Taking the guidance in the book into consideration, a 2-factor solution seems optimal from this analysis.

Then follow the instructions in the book to create a scree plot:



The scree plot supports the result of having two components, but there is some ambiguity about a possible third component before the scree (the rocks at the bottom of the hill). We should at least take into account that the eigenvalue of the third component is quite low.



Loadings: This is a snapshot of the first three components. Consider loadings greater than 0.3 or lower than -0.3.



The plot rather puts things in perspective.

And this visualisation of the observations against the variables is also rather useful:



We can also use rotation. Below are loadings based on 2 and then 3 components, using varimax rotation:





**EFA**

Use the Job Burnout.csv file.

This is a survey of 99 employees using the Maslach-Burnout Inventory General Survey (MBI-GS). The survey questions are as follows:

Question 1 I feel emotionally drained from my work.

Question 2 I feel used up at the end of the workday.

Question 3 I feel tired when I get up in the morning and have to face another day on the job.

Question 4 Working all day is really a strain for me.

Question 5 I can effectively solve the problems that arise in my work.

Question 6 I feel burned out from my work.

Question 7 I feel I am making an effective contribution to what this organization does.

Question 8 I have become less interested in my work since I started this job.

Question 9 I have become less enthusiastic about my work.

Question 10 In my opinion, I am good at my job.

Question 11 I feel exhilarated when I accomplish something at work.

Question 12 I have accomplished many worthwhile things in this job.

Question 13 I just want to do my job and not be bothered.

Question 14 I have become more cynical about whether my work contributes anything.

Question 15 I doubt the significance of my work.

Question 16 At my work, I feel confident that I am effective at getting things done.

The MBI-GS suggests that job burnout comprises of three major subscales namely exhaustion, cynicism, and professional efficacy. Based on this scale, questions 1,2,3,4 and 6 correspond to exhaustion, questions 8,9,13, 14, and 15 correspond to cynicism, and questions 5,7,10,11, and 12 correspond to professional efficacy

Conduct an exploratory factor analysis to see if the data will show the same three categories of underlying factors.

If you used KMO, you would find that there is no problem with the sampling adequacy, so we can continue the analysis without removing any variables. While following the code used in the book, we will just focus on the 3-factor solution.



The oblimin rotation method has been used.

Although there is some discrepancy from the original grouping of questions, there is still a resemblance to the subscales as suggested by the MBI-GS.

And below is a heat plot visualisation:



And one with pies!



And below is a triangular chart that is probably clearer to interpret:



And a combination chart:

