Assignment 1

Team 7 - Aakash and Noufris

April 2023

### Clean the environment and import essentials

rm(list = ls())  
cat('\014')

library(psych)  
library(corrplot)

## corrplot 0.92 loaded

### Question 1

Import, scale, and summarize the data in R

df\_survey = read.csv("assg1.csv")  
describe(df\_survey)

## vars n mean sd median trimmed mad min max range skew kurtosis se  
## Resp 1 30 15.50 8.80 15.5 15.50 11.12 1 30 29 0.00 -1.32 1.61  
## X1 2 30 4.40 2.94 5.0 4.46 2.97 0 9 9 -0.18 -1.53 0.54  
## X2 3 30 5.13 3.06 6.0 5.29 2.97 0 9 9 -0.46 -1.18 0.56  
## X3 4 30 3.67 2.75 3.0 3.50 2.97 0 9 9 0.49 -0.94 0.50  
## X4 5 30 5.40 2.91 6.0 5.58 2.97 0 9 9 -0.48 -1.17 0.53  
## X5 6 30 4.53 2.85 4.0 4.50 2.97 0 9 9 0.19 -1.19 0.52

df\_survey.sc <- df\_survey  
df\_survey.sc[, 2:6] <- scale(df\_survey.sc[,2:6])  
describe(df\_survey.sc)

## vars n mean sd median trimmed mad min max range skew kurtosis  
## Resp 1 30 15.5 8.8 15.50 15.50 11.12 1.00 30.00 29.00 0.00 -1.32  
## X1 2 30 0.0 1.0 0.20 0.02 1.01 -1.50 1.56 3.06 -0.18 -1.53  
## X2 3 30 0.0 1.0 0.28 0.05 0.97 -1.68 1.26 2.94 -0.46 -1.18  
## X3 4 30 0.0 1.0 -0.24 -0.06 1.08 -1.34 1.94 3.28 0.49 -0.94  
## X4 5 30 0.0 1.0 0.21 0.06 1.02 -1.86 1.24 3.10 -0.48 -1.17  
## X5 6 30 0.0 1.0 -0.19 -0.01 1.04 -1.59 1.57 3.16 0.19 -1.19  
## se  
## Resp 1.61  
## X1 0.18  
## X2 0.18  
## X3 0.18  
## X4 0.18  
## X5 0.18

The survey responses were scaled (normalized) with mean 0 and std 1.

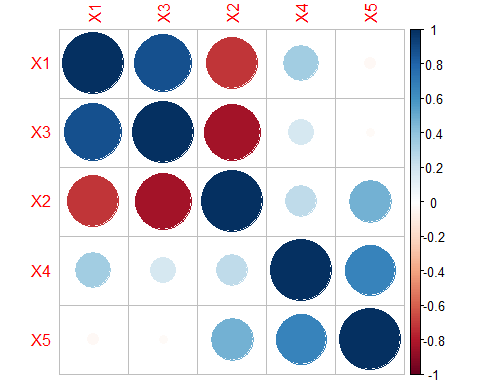
### Question 2

Present the matrix of correlations for the five variables (X1-X5) and plot the correlations. Explain (in words) your inferences based on this exploratory analysis

cor(df\_survey.sc[,2:6])

## X1 X2 X3 X4 X5  
## X1 1.00000000 -0.7107386 0.87895393 0.3432966 -0.03865034  
## X2 -0.71073859 1.0000000 -0.83182655 0.2573651 0.47015073  
## X3 0.87895393 -0.8318265 1.00000000 0.1727462 -0.02056604  
## X4 0.34329663 0.2573651 0.17274619 1.0000000 0.66421386  
## X5 -0.03865034 0.4701507 -0.02056604 0.6642139 1.00000000

corrplot(cor(df\_survey.sc[, 2:6]), order='hclust')



X1 and X3 are most highly positively correlated, X2 and X3 are most negatively correlated. This makes sense as X1 and X3 both rate the respect or attention the customer gets from the store, while X2 rates how much alone the salesperson leaves the user.

### Question 3

Perform PCA on the data and show the results (including the summary of the PCA and the matrix showing the relationship between the factors and original attributes)

df\_survey.pc = prcomp(df\_survey.sc[,2:6])  
summary(df\_survey.pc)

## Importance of components:  
## PC1 PC2 PC3 PC4 PC5  
## Standard deviation 1.6345 1.3610 0.58478 0.30601 0.20148  
## Proportion of Variance 0.5343 0.3704 0.06839 0.01873 0.00812  
## Cumulative Proportion 0.5343 0.9048 0.97315 0.99188 1.00000

df\_survey.pc$rotation

## PC1 PC2 PC3 PC4 PC5  
## X1 -0.55491712 0.2324895 -0.2795639 -0.7234233 0.19109786  
## X2 0.57063147 0.2117721 -0.1832902 -0.4621638 -0.61833777  
## X3 -0.57964069 0.1567520 0.3106045 0.1866707 -0.71282845  
## X4 -0.01698458 0.6859778 -0.5610983 0.4611287 0.04092556  
## X5 0.17370927 0.6371534 0.6906101 -0.1248104 0.26709683

### Question 4

The four equations are as follows:

Here,

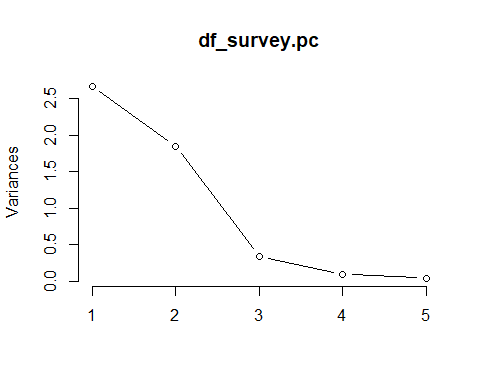
X1 = I never go back to a store that had treated me with disrespect. X2 = I find salespeople who fawn over me irritating, I just expect them to courteous. X3 = I have a lot of questions when I shop, I greatly appreciate a salesperson who is willing to spend time answering my questions. X4 = I care little for the fancy displays in departmental stores. X5 = Discount stores are much more of a bargain than departmental stores.

All responses are rated on the scale of 0 to 9, which are normalized by mean and standard deviation

### Question 5

How many factors or principal components should we retain? Explain your decision

plot(df\_survey.pc, type='l')

 Based on the PCA summary and the scree plot, we select to keep two principal components, PC1 and PC2. 1. PC1 captures 53.43% of the variance, while PC2 captures 37.04% of the variance. Both these components explain 90.48% of the variance in the dataset. 2. Looking at the scree-plot, we observe that PC1 and PC2 are behind the ‘elbow’ of the curve. From to PC3 and after, the curve flattens out. In addition, the variances of PC1 and PC2 are above 1, meaning that the amount of information that each of these two principal components provide is greater than 1 times the information of each dimension of the original data.

### Question 6

Can you think of intuitive names for the first and second factors? What do these factors represent? Explain how you arrived at them. Hint: Use the equations that quantify the relationship between the original attributes and the factors from the PCA (from answer to Question 4) and focus on strong correlations.

PC1 magnitude of variable coefficients rank: X3, X2, X1, X5, X4 PC2 magnitude of variable coefficients rank: X4, X5, X1, X2, X3

PC1 is highly correlated with X1(treating customers respectfully) and X3(salesperson willingness to assist customers). PC2 is highly correlated with X4(caring little about fancy displays in departmental stores) and X5(discount stores are much more of a bargain than departmental stores).

PC1 could be customer service PC2 could be value for money

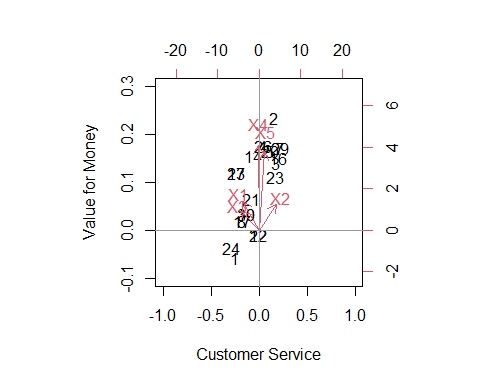
PC1 represents the assistance and support provided by a business to its customers during a purchase. High customer service includes aspects such as timely and effective communication, knowledgeable and friendly staff, quick resolution of problems and complaints.

PC2 represents the perception of a customer for a product based on cost and quality. If a product is of high quality, durable, and meets customer’s needs, and its price is reasonable or competitive, then a customer may perceive that they received good value for their money.

### Question 7

Now plot all the consumers on a two-dimensional map based on the first two principal components. Label the x and y-axis using the names from your answer to the previous question

biplot(df\_survey.pc,   
 xlim=c(-1,1),   
 ylim=c(-0.1,0.3),  
 xlab='Customer Service',  
 ylab='Value for Money')  
abline(h = 0, v = 0, col = 'gray60')



### Question 8

Based on the plot from the previous answer, how many segments (roughly) do you find in the market? Hint: A segment is a group of consumers who are similar to each other in their tastes and preferences.

Based on the above plot, we can identify 2 segments among the respondents of the survey. The first segment prioritize cost and quality when visiting discount stores and departmental stores. The second segment prioritize customer service when visiting discount stores and departmental stores.

### Question 9

Segment 1 (characteristics) 1. Value-conscious: Looking always for a good deal and making sure to get the best value. 2. Willing to pay more for quality: Willing to pay a premium price for a product or service that meets their expectations of quality. 3. Seek out reviews and recommendations: Doing research before making a purchase and seeking out reviews and recommendations from others to ensure they are getting a good deal. 4. Prefer long-term relationships: Valuing long-term relationships with businesses that consistently provide them with high-quality products and services at a reasonable price. 5. Pragmatic: Being practical in their purchasing decisions and not easily swayed by flashy marketing or branding.

Segment 2 (characteristics) 1. Loyalty: Tending to be more loyal to the companies which provide better customer service. 2. Communication: Being more communicative and open about their needs and expectations. 3. Relationship-focused: Being more interested in building a long-term relationship with the company rather than just making a one-time purchase. 4. Feedback-oriented: Willing to provide feedback on their experience with the company.

### Question 10

Which segment do you expect to be the most profitable and why? Assuming the data to be representative, what is the size of this segment (in percentage)

We expect segment 2 which values customer service more than anything else to be more profitable than segment 1. The reasons are: 1. customers in this segment are loyal. They tend to stick to the company that provides them good customer service. 2. Customers in this segment tend to pay more if they receive good customer service.

The size of this group could be around 30%.