

Assignment 1: Consumer Segmentation using PCA

Ali Goli

Overview

A pilot survey was conducted of 30 individuals to understand their different attitudes towards discount stores (e.g., K-Mart), which usually come with low levels of service versus departmental stores (e.g., Macys) which come with higher levels of service. Respondents were asked their opinion on a 0-9 agree-disagree scale (where 0 indicates strongly disagree and 9 indicates strongly agree) on the following questions.

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

The data for this study are provided in `assg1.csv`. Using this data, answer the following questions.

Questions

1. Import, scale, and summarize the data in R. (1 point)
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. (1 point)
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). (1 point)
4. Write down the five equations that formally define the relationship between the original attributes and the factors from the PCA. (1 point)
5. How many factors or principal components should we retain? Explain your decision. (1.5 points)
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. (2 points)
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. (1.5 points)
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. (2 points)
9. Describe the characteristics of each segment. Discuss the relationship between the segment characteristics and the original attributes? (2 points)

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)? (2 points)

Instructions

1. Late assignments will automatically receive zero.
2. The points for each question are shown next to it. The total is 15 points.
3. You are expected to write up the assignment using R Markdown (in the style of the handouts you were given for the in-class work). Please answer each question one by one. For each question, show the R commands you used and then briefly present any discussion/text that you want below it. Please submit the .Rmd file and the 'knitted' word document. You do not need to submit any data files.
4. If you have any questions or need any clarifications on the assignment, please post on your question on the discussion board for this assignment. I or the TA will try to answer it as soon as possible. If you see a question from another student that you know the answer to, please feel free to chip in.