**Homework 6. Toyota Corolla Dataset**

Author: Rohit V. Akole

UConn ID: 3133252

University of Connecticut

OPIM5604 Predictive Modeling

Professor: Pankaj Prakash

Abstract

This homework is about the Toyota Corolla dataset PCA and a set of cereals for a public elementary school for the K-means cluster.

Table of Contents

[Question 1 4](#_Toc147665694)

[a. Principal Components Analysis 4](#_Toc147665695)

[b. Determine the correlation of 2 saved Principal Components 5](#_Toc147665696)

[Question 2 6](#_Toc147665697)

[a. Healthiest cluster of cereals 6](#_Toc147665698)

[b. Least healthy cluster of cereals 6](#_Toc147665699)

# Question 1

Q1. Using ToyotaCorolla.jmp dataset for the following:

## Principal Components Analysis

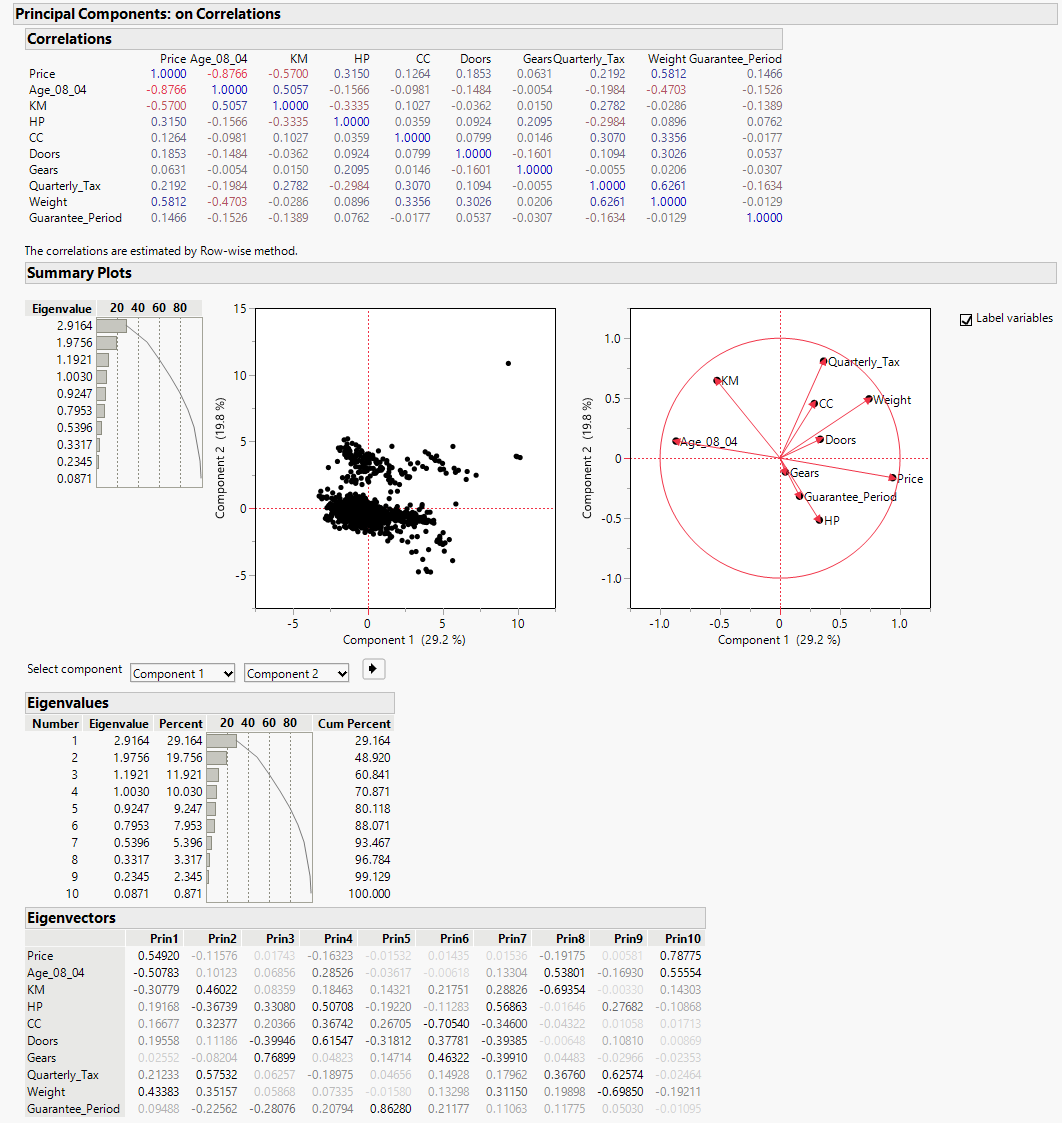


Figure 1. Principal Component Analysis

As we can see in PCA, we can maintain about 48.92% variability if I keep 2 principal components. In PC1, Price(54.92%) contribute the most information followed by Weight (43.48%) and least comes from Age\_08\_04 (-50.78%). For PC2, Quarterly Tax (57.53%), KM (46.02) provides the most information and HP (-36.74) provides the least. According to me, to get the ideal variability, we need to take at least 5 components to get the 80.12% variability.

## Determine the correlation of 2 saved Principal Components

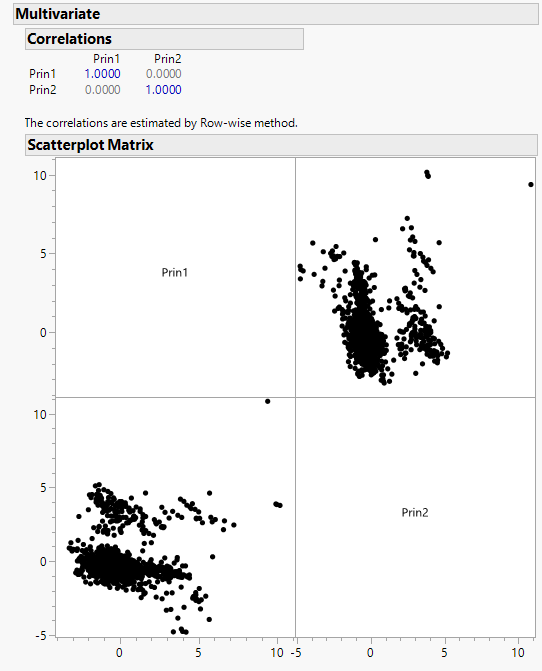


Figure 2. Multivariate Correlations

Using the multivariate to see the correlation between Principal Component 1 (PC1) and Principal Component 2 (PC2), using the correlation matrix we can see, P1 to P1 and P2 to P2 both are perfectly correlated to itself and hence the correlation is 1.00 for both. Similarly, P1 to P2 and P2 to P1 are perfectly uncorrelated to each other. With this outcome, we can say that both principal components are orthogonal to each other. Both components are uncorrelated and capture different aspects.

# Question 2

Q2. A public elementary school would like to choose a set of cereals to include in their daily cafeteria.

Every day a different cereal is offered, but all cereals should support a healthy diet. For this you are

requested to find a cluster of "healthy cereals." Note: No data preprocessing is required for this

assignment.

## Healthiest cluster of cereals

Based on the given cluster summary, cluster 2 is the healthiest cluster of cereals. It has the lowest calories, highest protein, no fat, lowest sodium, highest fiber, and no sugar. The overall rating is 93.7049 as well, which is the highest rating of all. Hence, Cluster 2 includes the healthiest cereals.

## Least healthy cluster of cereals

According to me, cluster 4 is the least healthy cluster of cereals. It has the lowest rating of all at 38.0984, the highest calories (116.667), the highest sodium (208.333), the highest carbs (18.167), and sugar (6.333) is also high.