STAT 8007 Statistical Methods for Big Data

PCA Project Descriptor

**Please hand the project in to me on (or before) Tuesday 8th May.**

**This project is worth 50% of your final grade.**

**This analysis must be carried out using R.**

Polymer Manufacture

A company is in the process of developing a new polymer to be used in the manufacture of protective coatings. The polymer must be of high strength but must also be resistant to warping. The company is investigating the strength and warping properties of 60 different polymers, each composed of varying mixtures of 6 different materials. The materials are labelled as A, B, C, D E and F.

This project is split into three parts:

Part 1

Each of the 60 mixtures was tested for strength and resistance to warping and the mixtures were ranked according to a combined performance index in both qualities. The data set containing the ranking of the mixtures (1-60) and the corresponding levels of each of the 6 materials used in the mixture is in the file polymer1 available on Blackboard.

The data set polymer1 consists of 60 observations and 7 variables. The first variable is the ranking, (1:60) of the polymers where 1 corresponds to the polymer that performed best (a combination of strength and resistance to warping). The remaining 6 variables are the corresponding levels of each of the 6 materials used in the mixture.

The aim of the first part of this project is to determine:

1. whether PCA can reveal patterns in the ranking of the polymers
2. how the different materials relate to the ranking of the polymers

Part 2

The data set polymer2 consists of the same 60 observations and 8 variables. The first variable is the ranking, (1:60), of the polymers in relation to strength only (1 corresponds to the strongest polymer). The second variable is the ranking, (1:60), of the polymers in relation to its resistance to warping only (1 corresponds to the polymer that is most resistant to warping). The remaining 6 variables are the levels of each of the 6 materials used in the mixture.

The aim of the second part of this project is to:

1. relate the individual qualities of strength and resistance to warping to the principal components defined in part 1
2. determine how the different materials relate to the individual qualities of strength and resistance to warping

Part 3

Given a polymer with a known composition of materials that has not been tested for its strength and resistance to warping, is it possible to use PCA to assess its likely performance?

The data set containing the levels of the 6 materials for 10 polymers is in the file polymer3 available on Blackboard. The data set polymer3 consists of 10 observations and 7 variables. The first variable is the polymer label, (A:J), the remaining 6 variables are the levels of the composite materials. Note that the performance rankings of the samples are not given.

The aim of the third part of this project is to use PCA to comment on the composition of each sample and to relate this to the qualities of strength and resistance to warping. Which of these polymers do you expect to perform best in relation to strength and resistance to warping?

The project must have the following layout:

**Introduction**

**Methods**

**Results**

**Discussion**

The introduction should contain a brief description of the project aims and the data sets,

The materials and methods section should describe the methods used to analyse the data, the software this was implemented with and any additional packages or important functions that you used.

The results section will be the longest section. You may wish to split this section into further sections, for example:

1. Exploratory data analysis
2. PCA – this section should include summary tables, score plots, loading plots, goodness of fit and bar charts to illustrate how well each variable is explained by a PCA model with:
3. 1 principal component ()
4. 2 principal components )

The discussion gives you an opportunity to summarise your findings succinctly and discuss the implications of your findings.

The project should be between 15-20 pages (approximately). Note that marks are allocated to presentation.

Hints

The first column in the polymer1 data frame is a ranking– not a variable for use in the PCA. You will need the information contained in the first column to identify samples.

In part 3, use the polymer1 data set as your reference data set.