Principal Components Analysis

Homework Wk 10

The assignment

The assignment

Perform a metric ordination on the `mtcars2` dataset based on the Euclidean distance matrix—*e.g.*, perform a Principal Components Analysis, or PCA.

Upload your product to `Blackboard` as a Word document created via `R markdown`. There is no template for this assignment; you are expected to create the `R markdown` file yourself using skills developed so far in the course and of course using the Internet and cheatsheets available from the folks at `R studio`.

Data preparation, identification

Load the `mtcars2` datset, available here: mtcars2 in Google Drive. Identify the structure of the data.

Your data should look like this:

```
## 'data.frame':    32 obs. of  16 variables:
##  $ make.model: Factor w/ 32 levels "AMC Javelin",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ make      : Factor w/ 20 levels "AMC","Cadillac",..: 1 2 3 4 5 6 16 7 8 8 ...
##  $ origin    : Factor w/ 2 levels "domestic","foreign": 1 1 1 1 2 1 1 2 2 2 ...
##  $ country   : Factor w/ 6 levels "Germany","Italy",..: 6 6 6 6 3 6 6 2 2 2 ...
##  $ continent : Factor w/ 3 levels "Asia","Europe",..: 3 3 3 3 1 3 3 2 2 2 ...
##  $ am        : Factor w/ 2 levels "Automatic","Manual": 1 1 1 1 2 1 1 2 2 2 ...
##  $ vs        : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 2 2 ...
##  $ gear      : Factor w/ 3 levels "3","4","5": 1 1 1 1 2 1 1 3 2 2 ...
##  $ carb      : Factor w/ 6 levels "1","2","3","4",..: 2 4 4 4 1 2 4 5 1 1 ...
##  $ cyl       : Factor w/ 3 levels "4","6","8": 3 3 3 3 1 3 3 2 1 1 ...
##  $ disp      : num  304 472 350 440 108 318 360 145 78.7 79 ...
##  $ hp        : int  150 205 245 230 93 150 245 175 66 66 ...
##  $ drat      : num  3.15 2.93 3.73 3.23 3.85 2.76 3.21 3.62 4.08 4.08 ...
##  $ wt        : num  3.44 5.25 3.84 5.34 2.32 ...
##  $ qsec      : num  17.3 18 15.4 17.4 18.6 ...
##  $ mpg       : num  15.2 10.4 13.3 14.7 22.8 15.5 14.3 19.7 32.4 27.3 ...
```

Text answers:

- Which variables look appropriate for inclusion in the site x species matrix (ordination data)?
- Which might be "environmental" variables to test?

Analysis

Fit and assess ordination

Fit the PCA

Provide code—but not results—for a PCA on the appropriate variables you identified above.

Assess the PCA

1. First, provide script for, and a graph of, a screeplot of the PCA. Interpret this graph.
2. Scond, use output from the PCA to answer the following questions.

Note: The results returned by `summary()` *calls on ordination objects with many variables and many observations can be unwieldly. It is therefore easier to assess the various parts of the ordination results individually. Code to extract two important sections follows:*

```
# Eigenvalues and proportion variance explained:
summary(round(eigenvals(cars.pca), 2) )
# Extract species scores
round(scores(cars.pca, choices=c(1:6))$species, 2)
```

Text answers:

- How much variation is explained by the first two axes (PCs) of the ordination? Would you consider this an acceptable amount?
- Which variable(s) contribute the most to PC1 and PC2? Along PC1, do you suspect they are negatively or positively correlated? How can you tell?

Plot the PCA

Produce code and a graph for a biplot of both site scores and species scores with informative text labels.

Text answers:

- True or False: The site scores are a mess.
- What might be done to improve this?

Handling environmental variables

1. Choose a categorical variable from the original dataset to treat as an environmental variable. Provide script and a graph for a new plot that shows site scores connected by the variable you selected.
2. Provide code and output for a statistical test of difference between these groups. Does the variable you chose explain any differences among the cars in the dataset? *Along which axis or axes does most of the variation between these groups occurr?*