Statistical Programming Using the R Language

Lecture 5 – Introducing Multivariate Analysis

Preliminaries:

Before you begin, ensure you have done the following (instructions are contained within the slides):

- Loaded your script from yesterday, **problems.R** into RStudio.
- Loaded your session, **problems.RData** into RStudio.
- Set the working directory as the r_course folder

Your dataset should contain the entire dataframe for colon_cancer_data_set.txt, and two subsets, one containing gene expression for the tumour (affected) samples and other containing data for the normal (unaffected) samples.

Note, you can use the `ls()` function to list the variables contained within your R session. Alternatively (or simultaneously), you can look at the **environment window** (top right) in RStudio, which should also list the contents of your session.

Plots can saved by using the Export option in the plotting window.

Q1. Produce a cluster dendrogram of the samples in the colon cancer data set based on data from the first 20 genes. Use 1 - |r| for Spearman's rho as your distance metric. Do the tumour and normal samples form distinct clusters? (**Hint**: correlation works on the columns of a dataset so you will have to transpose your data). Try the option `hang=-1` in your `plot()` function to make a tidy dendrogram.

Q2. Produce a cluster dendrogram, again using 1-|r| for Spearman's rho, of the first 20 genes in your data set. (Hint: the initial data is in the format samples x genes so no transposition required).

Q3. Install and load the Heatplus package from bioconductor.

Q4. Produce a heatmap for all samples and the first 20 genes using 1 -|r| as in Q1 and Q2 as your distance metric. Are the dendrograms from the heatmap "roughly" similar to those from Q1 and Q2 for the samples? Differences in shape are to be expected but the groupings should be broadly similar.

Q4. Once completed:

4.1 Save your script.

4.2 Save your session.

These may useful for future reference!

I hope you found the workshop worthwhile and useful. Thanks!