BIOINFORMATICS MSc
PROBABILITY AND STATISTICS

SPLUS EXERCISE SHEET 5

ARRAY NORMALIZATION

Things to download:
SPLUS scripts and Data: On the web page

`http://stats.ma.ic.ac.uk/~das01/BioinformaticsMSc/material.html`

data and scripts relating to cDNA microarray analysis and processing can be found. The data relate to an anaysis of tissue samples from heterozygote wild-type animals on day15 and day 16 of the experiment.

There are four data files:

- `ReplicatesDataHE15.xls` (WildType Day 15)
- `ReplicatesDataHE16.xls` (WildType Day 16)
- `ReplicatesDataKO15.xls` (Knockout Day 15)
- `ReplicatesDataKO16.xls` (Knockout Day 16)

and scripts to normalize and analyze these data

- using no background correction
- using a background correction

The microarray experiments record measures of expression in the test (measure R) and control (measure G) samples. In these experiments, the control sample is the WildType homozygote on day 15. Key measures in the study of cDNA data are

$$A = \frac{1}{2}\log_2(R \times G) \qquad M = \log_2 \frac{R}{G} = \log_2 R - \log_2 G.$$

The measure M is the measure of differential expression in test over control samples. However, it has been observed that there is considerable variation in these data that is not present due to biological causes, but is rather due to systematic differences between different array experiments.

Download the data and scripts, and then use them to study effect the particular type of normalization used (which I am not claiming is the best method, but is one that appears to work ...). Your objective is to attempt to understand the code used.

Note there is also an SPLUS script to produce images of the microarrays data using the `image` function.