MOAC – Module CH923
2008/09

Assignment 1 – Data Analysis

For all questions you should present the results of your statistical analyses, an interpretation of these results, and, where you have used R (or some other statistical package) to generate the statistical analyses, the code that you have used. For questions 2 to 6, if you conclude that the assumptions behind an analysis you intended to use are not met, please describe the problem as well as the analysis you eventually perform

All data files are provide in two formats – as sheets within an Excel file ("Assignment 1 – Data Analysis.xls"), and as tab-delimited files for loading into R

This assignment must be submitted by 17:00 on Monday 17th November to the MOAC administrator (moac2). We would prefer electronic submission but hard-copy is also acceptable.

1. An experimenter is interested in the genetics of root properties in tomato plants. He has been working with a population of near isogenic lines, which have been derived from a cross between an inbred commercial cultivar 'M82' and a wild relative. The lines are genetically identical to M82 except for one region (different for each line) where the lines have the genes of the wild relative. He has identified two lines (called here A and B), with introgressions on different linkage groups, which appear to differ from M82 in some rooting properties. He performs two experiments to examine in more detail the rooting properties of these two lines and M82.

    - In the first experiment he studies the ability of the roots to penetrate a barrier. He grows tomato plants in individual plastic tubes of compost 30cm tall, with a membrane across the bottom of the tube which resists root penetration. When the plants are a few weeks old, for each plant he counts the roots which have penetrated the membrane, records the length by which the longest root has penetrated, and then cuts the roots off at the membrane and weighs them. He also cuts off the above-ground canopy of the plant at ground level and weighs that. He then oven-dries the excised roots and canopy at 100°C for 48 hours, and later weighs the dried material.

    - In the second experiment he grows the plants hydroponically in individual, tall (2m) tubes. When the plants come into flower he lifts them from their tubes and

measures the length of the longest root. He then cuts the roots into four sections, each having one quarter the length of the longest root, so that section 1 is the quarter of roots closest to the surface, and section four is the deepest quarter. These sections are dried and weighed. The above-ground canopy is also weighed both before and after drying.

Explore these data, using appropriate summary statistics and graphical methods but no formal tests, assessing possible relationships between variables, and potential differences between the lines.

Excel sheets: *Question 1, experiment A*
Question 1, experiment B
R data files: *Root penetration.txt*
Root length.txt

2. An experimenter has obtained twenty samples of soil from around the country, all supposedly infected with a soil borne disease. She is interested in testing whether a seed treatment will give protection against the disease, which if uncontrolled results in poor growth. She divides each of the twenty soil samples into two parts, and sows treated seed into one part, and untreated seed into the other. After the seedlings are two weeks old she cuts the seedlings off at the soil surface, dries and weighs them. Has the seed treatment provided protection?

Excel sheet: *Question 2*
R data file: *Seed treatment.txt*

3. A set of 95 Arabidopsis accessions have been assessed for a gene which gives resistance to a specific fungal disease pathovar. They have also been sequenced at many different loci, and classified according to their haplotypes (arbitrarily labelled) at these loci. If the resistance gene is tightly linked to one of the loci at which the accessions have been sequenced, then an association between the resistance phenotype and the haplotypes at the loci might be expected (this association is known as linkage disequilibrium). Data for the haplotypes for a locus and for the resistance phenotype are provided. Is there evidence of such an association?

Excel sheet: *Question 3*
R data file: *Association.txt*

4. An PhD student has attempted to transform a cultivar of tomato with a construct containing a gene which gives resistance to a herbicide, Basta. He has obtained three lines each deriving from a separate transformation event. To confirm Basta resistance, he performs a small screening experiment, growing five plants from each line and from an untransformed control, spraying them with Basta, and then weighing the above-ground part of the plant two weeks later. Are there any differences in the performance of the three transformed lines?

Excel sheet: *Question 4*

R data file: *Herbicide resistance.txt*

5. At Warwick HRI we have a weather station, which automatically measures a wide range of different weather variables daily. A small subset of the data for 2006 is provided, giving mean (air) temperature in degrees centigrade and total solar radiation in MJm^{-2}. Is there a linear relationship between these variables?

   Excel sheet: *Question 5*
   R data file: *Weather data.txt*

6. A scientist is interested in the processes involved in leaf senescence. She generated a dataset by growing 36 arabidopsis plants, and over a period of twelve days during which the seventh leaf goes from being mature to being heavily senescent, harvesting the leaf from three plants a day. The level of gene expression of a gene previously linked to senescence is measured in each leaf by qRT-PCR. Gene expression levels relative to a housekeeping gene (which is believed not to vary in expression levels over time) are provided. Does the expression level of the gene vary over the course of the experiment? If so, is the change linear with time?

   Excel sheet: *Question 6*
   R data file: *Senescence.txt*