Final exercise

Version 2019-03
Licence

This manual is © 2015-2019, Anne Segonds-Pichon.

This manual is distributed under the creative commons Attribution-Non-Commercial-Share Alike 2.0 licence. This means that you are free:

- to copy, distribute, display, and perform the work
- to make derivative works

Under the following conditions:

- Attribution. You must give the original author credit.
- Non-Commercial. You may not use this work for commercial purposes.
- Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under a licence identical to this one.

Please note that:

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.
- Nothing in this license impairs or restricts the author's moral rights.

Full details of this licence can be found at http://creativecommons.org/licenses/by-nc-sa/2.0/uk/legalcode
**Final exercise: Tomato Rooting Experiment**
(Data from: [www.stats tutor.ac.uk](http://www.stats tutor.ac.uk))

An experimenter is interested in the genetics of the rooting properties of tomato plants. He has been working with a population of near isogenic lines (i.e. almost genetically identical), 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 A and B here) 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 relative to M82, and the relationship between root and canopy development.

**tomato.data.csv**: In this experiment, he studies the ability of 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 growth. When the plants are a few weeks old, for each plant he counts the roots which have penetrated the membrane (**Number of penetrating roots NORP**), records the length by which the longest root has penetrated the membrane (**Length (mm)**), and then cuts the roots off immediately below the membrane and weighs them (**Penetrating root fresh weight (g)**). He also cuts off the above ground canopy of the plant at the top of the tube (soil level) and weighs that (**Canopy fresh weight (g)**). He oven dries both the excised roots and the canopy and later weighs the dried material (**Penetrating root dry weight (g)** and **Canopy dry weight (g)**).

**Questions:**

**tomato.power.csv**

1) The experimenter has some preliminary data on the number of penetrating roots for Line A and M82. Use the data to estimate the sample sizes.

The experimenter decides to be on the safe side and to go for n=15 in each group, including Line B.

**tomato.data.csv**

2) Root length
   a. What is the mean across the 3 lines?
   b. What is the mean for each line?
   c. What is the median for each line?
   d. Notice something weird?

3) What is the variability in the root length for each line?

4) Arrange graphs in a multi-panel way (2x3) so that you can compare at once:
   a. the distributions of number of penetrating root, canopy fresh weight and root length between the lines with boxplots
   b. the distributions of number of penetrating root, canopy fresh weight and root length between the lines with beanplots
   c. which one do you think is more informative?
5) Plot the mean root length for each line on a stripchart

6) Plot the mean for canopy dry weight for each line on a stripchart

7) Construct an interval within which the true (population) mean total root dry weight for M82 would fall (with 95% confidence)
   Hint: you will need to do some filtering and the plotrix package (or not!).

8) Is there a linear association between:
   a. size of the canopy dry weight and root dry weight?
   b. root fresh weight and root dry weight?
   c. root dry weight and number of penetrating roots?
   Hint: Maybe it is easier to look at all pairwise associations at once with cor()?

9) Are there differences in the mean canopy dry weight between lines A and B? Don’t forget to check the assumptions.

10) Are there differences between the 3 lines for:
    a. canopy dry weight?
    b. root dry weight?
    c. number of penetrating roots?
       • Don’t forget to check the assumptions!
    d. Construct a linear model for canopy dry weight and check graphically the assumptions.

11) Construct linear model to find out if:
    a. the canopy dry weight linearly related to the root dry weight?
    b. the canopy dry weight linearly related to the number of penetrating roots?
    c. the canopy dry weight linearly related to the canopy fresh weight?
    d. the root fresh weight linearly related to the root dry weight?
    e. Find $R^2$ for a. and d.

12) Plot all for 4 relationships on one panel and add a line of best-fit.

   If you feel up for some extra R, try below.
   • Assign a colour to each line with \texttt{col = tomato1$Genotype}
   • Add a legend \texttt{legend(x, levels(tomato1$Genotype), pch=, col=palette())}
   • This is a simplified version of the arguments. Check \texttt{?legend()} for all of them.
     o \texttt{x = position e.g. “topright”} (See Details in \texttt{?legend()})
     o \texttt{levels(tomato1$Genotype)}: a vector of names of the 3 lines that will be displayed in the legend.
     o \texttt{pch} how to display the colours
     o \texttt{col=palette()}: will take the first 3 default colours from the default R palette \texttt{palette()}. 
13) Is the relationship between the canopy dry weight and the root dry weight varying between lines?

   a. Analyse, check the assumptions and plot the data. Interpret the output, including $R^2$. Hint: remember that `abline()` needs a vector of 2 values to draw a line.
   
   b. One value looks dodgy. Run the analysis without it. Does it look better?
   
   c. Compare $R^2$ for the model with and without lines as a factor. Does it make a difference?