Stat 587, section 2 - Lab 4 self assessment

The data in Darwin.csv are classic data collected by Charles Darwin. He was interested in the effect of cross pollination in plants. He collected seeds from flowers on a single plant that were self-pollinated (pollen from that plant) or cross-pollinated (pollen from a distant plant). He then planted these seeds in pots, grew the plants and measured their height, in inches. Two seeds were planted in each pot. One was cross-pollinated seed; the other was self-pollinated. The data are in Darwin.csv.

1) Are these data two independent samples or paired? Explain your choice.

2) How many pots were used in this study?

3) Estimate the mean difference, as cross – self pollinated.

4) Calculate the standard error of that difference.

5) Use a t-test to test the null hypothesis of no difference in mean height between cross- and self-pollinated seeds. Report the p-value and a short conclusion.

6) Estimate a 95% confidence interval for the difference.

Our answers:

1) These data are paired because two seeds, one from each treatment, were planted in each pot. A cross-pollinated observation is “linked” to the self-pollinated observation in the same pot.

2) 15

Look at the number of rows of data

3) 2.62 inches.

If you wanted a sentence describing the mean difference, you could write something like: Cross-pollinated seeds produce plants that are 2.62 inches taller than the plants from self-pollinated seeds.

4) 1.22 inches.

Calculated as the sd / sqrt(n), if not reported by your software. t.test() in R doesn’t automatically report this, so you have to “hand” calculate it, although you use R to calculate the sd, which is the tedious bit.

5) p = 0.0498. Evidence of a difference in average height between cross- and self-pollinated seeds (p = 0.0498).

6) (0.0032, 5.23)