# Durham University
# MATH1541 Statistics
# Exercise Sheet 16

Kamil Hepak
Tutorial Group 4

Mar 2019

## 1 Q1

### 1.1 With Main Outliers

Let differences be denoted $d$: $\bar{d} = 2.55$, $s_d = 4.7117$, $n_d = 15$.
$H_0 : \mu_d = 0$, $H_a : \mu_d \neq 0$, $\alpha = 5\%$ (two-tailed)
The critical values for the $t_{n-1}$ test in this case are thus the 0.025 values from the tails of $t_{14}$ distribution. These values are $\pm 2.1448$.
We calculate the test statistic, $t$, as follows:

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n_d}}}$$

$$t = \frac{2.55 - 0}{\frac{4.7117}{\sqrt{15}}}$$

$$t = 2.0961$$

We can go about checking the significance of the test statistic in one of many ways, but the most straightforward is to compare it to the critical value. Since $t < 2.1448$, we fail to reject the null hypothesis at the 5% significance level - the evidence does not suggest $\mu_d \neq 0$.

### 1.2 Assumptions

To make this a valid test, we must assume:

- Each observation is independent
- Each observation is selected with a simple random sample
- The original distribution of the plant heights is reasonably Normal (incl. no outliers and little to no skew)

## 1.3 Without Main Outliers

There are three outliers in the data set - two are contained in the cross-fertilised portion, while one is in the self-fertilised. The two cross-fertilised outliers are further from the lower outlier limit (17.65625, calculated as $Q_2 - 1.5 \cdot \text{IQR}$) for the cross-fertilised data than the other outlier is from the self-fertilised limit, as well as being absolutely lower than the other outlier. Therefore, they and their corresponding self-fertilised data points have been omitted. Additionally, when converted to differences, the only outliers are the ones stemming from the cross-fertilised data, further reinforcing the decision to omit those two data points.

Let differences be denoted $d$: $\bar{d} = 4.0481$, $s_d = 2.7260$, $n_d = 13$.
$H_0 : \mu_d = 0$, $H_a : \mu_d \neq 0$, $\alpha = 5\%$ (two-tailed)
The critical values for the $t_{n-1}$ test in this case are thus the 0.025 values from the tails of $t_{12}$ distribution. These values are $\pm 2.1788$.
We calculate the test statistic, $t$, as follows:

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n_d}}}$$

$$t = \frac{4.0481 - 0}{\frac{2.7260}{\sqrt{13}}}$$

$$t = 5.3542$$

We can go about checking the significance of the test statistic in one of many ways, but the most straightforward is to compare it to the critical value. Since $t > 2.1788$, we reject the null hypothesis at the 5% significance level - the evidence does not suggest $\mu_d = 0$.