

# Lab 2 – Summary statistics, graphics, and the $t$ -test

August 20 & 21, 2018  
FANR 6750

Richard Chandler and Bob Cooper  
University of Georgia

## Last week we covered:

- Vectors
- Data frames
- Indexing
- Importing and exporting data
- Saving and loading workspaces

## TODAY'S TOPICS

### 1 INTRODUCTION

### 2 GRAPHICS

### 3 $t$ TESTS

- Two-sample  $t$  test
- Equality of variance test
- Paired  $t$ -test

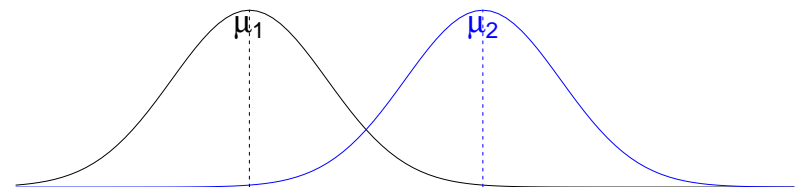
## SCENARIO

We have 2 samples of data and we want to know if they came from population.

The problem is that the true population means ( $\mu_1, \mu_2$ ) are unknown.

Under the assumption that the variances of the two populations are equal, the relevant hypotheses are:

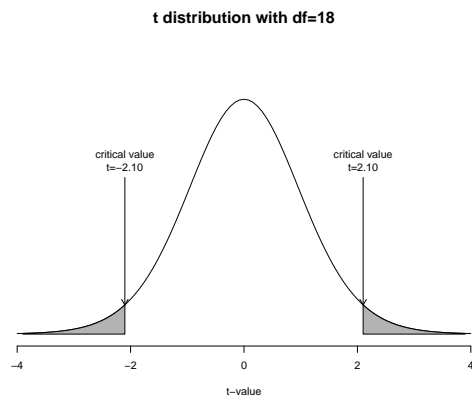
- $H_0 : \mu_1 = \mu_2$
- $H_A : \mu_1 \neq \mu_2$



## KEY POINTS

If the two sample means ( $\bar{y}_1, \bar{y}_2$ ) are very different and the standard error of the difference in means is small, the  $t$  statistic will be far from zero.

If the  $t$  statistic is more extreme than the critical values, you reject the null hypothesis ( $H_0$ ).



## EXERCISE I

- (1) Open **R** and set the working directory to a convenient location on your computer. Do this using `File > Change dir ...`, or the `setwd` function.
- (2) Put the file `treedata.csv` into your working directory.
- (3) Create a new **R** script and import `treedata.csv`. Name your object `treedata`.
- (4) Using the indexing methods we covered last time to create 2 objects: `yL` is the tree density data for the first 10 experimental units (low elevation), and `yH` is the tree density data for the last 10 units (high elevation).
- (5) Compute the mean, variance, and standard deviation of the 2 samples.

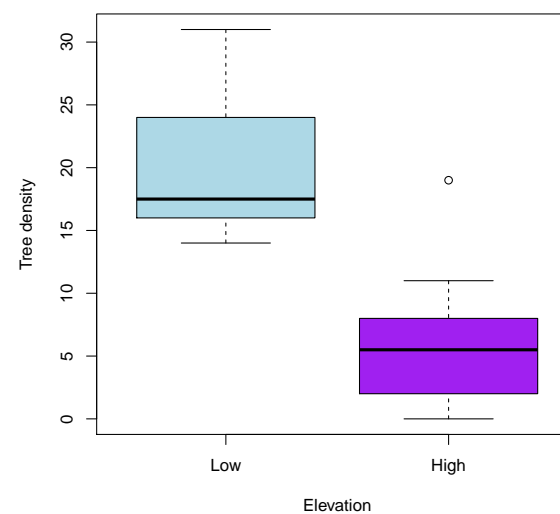
## TODAY'S TOPICS

1 INTRODUCTION

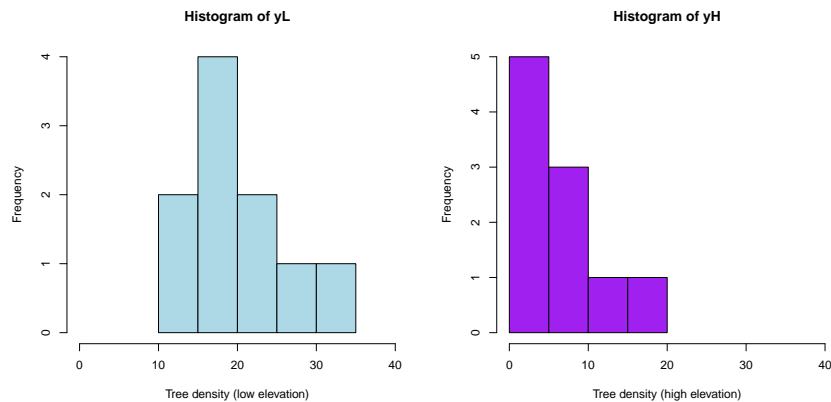
2 GRAPHICS

3  $t$  TESTS

## BOXPLOTS



```
hist(yL, xlab="Tree density (low elevation)", col="lightblue",
     xlim=c(0, 40))
hist(yH, xlab="Tree density (high elevation)", col="purple",
     xlim=c(0, 40))
```



Create the same boxplots and histograms as before, but change the colors of the boxplots and the number of break points in the histograms.

## TODAY'S TOPICS

1 INTRODUCTION

2 GRAPHICS

3 t TESTS

TWO-SAMPLE *t*-TEST WITH EQUAL VARIANCES

**Step 1:** Compute the *t* statistic<sup>1</sup>:

$$t = \frac{(\bar{y}_L - \bar{y}_H) - (\mu_L - \mu_H)}{\sqrt{s_p^2/n_L + s_p^2/n_H}}$$

where  $s_p^2$  is the pooled variance:

$$s_p^2 = \frac{(n_L - 1)s_L^2 + (n_H - 1)s_H^2}{n_L + n_H - 2}$$

**Step 2:** Compare *t* statistic to critical values

Critical value for 1-tailed test  $t_{\alpha=0.05,18} = -1.73$  or  $1.73$

Critical values for 2-tailed test  $t_{\alpha=0.05,18} = -2.10$  and  $2.10$

<sup>1</sup>Remember,  $H_0$  states that  $\mu_L - \mu_H = 0$ .

Step 1: Compute the  $t$  statistic:

```
mean.L <- mean(yL)
mean.H <- mean(yH)
s2.L <- var(yL)
s2.H <- var(yH)
n.L <- length(yL) # length returns the number of elements in a vector
n.H <- length(yH)
s2.p <- ((n.L-1)*s2.L + (n.H-1)*s2.H)/(n.L+n.H-2)
SE <- sqrt(s2.p/n.L + s2.p/n.H)
t.stat <- (mean.L - mean.H) / SE
t.stat

## [1] 5.404896
```

Step 2: Compare  $t$  statistic to critical values (two-tailed)

```
alpha <- 0.05
## NOTE: qt returns critical values. No need to use "t tables"
critical.vals <- qt(c(alpha/2, 1-alpha/2), df=n.L+n.H-2)
critical.vals

## [1] -2.100922  2.100922
```

Conclusion: Reject  $H_0$  because 5.4 is more extreme than the critical values.

```
t.test(yH, yL, var.equal=TRUE,
       paired=FALSE, alternative="two.sided")

##
## Two Sample t-test
##
## data:  yH and yL
## t = -5.4049, df = 18, p-value = 3.898e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -19.580772 -8.619228
## sample estimates:
## mean of x mean of y
##      6.1      20.2
```

Make sure you set `var.equal=TRUE`. Otherwise, R will assume that the variances of the two populations are unequal.

```
t.test(treeDensity ~ Elevation, data=treedata, var.equal=TRUE,
       paired=FALSE, alternative="two.sided")

##
## Two Sample t-test
##
## data:  treeDensity by Elevation
## t = -5.4049, df = 18, p-value = 3.898e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -19.580772 -8.619228
## sample estimates:
## mean in group High mean in group Low
##      6.1      20.2
```

This second option returns identical results, but it is preferred because the notation is much more similar to the notation used to fit ANOVA models.

The standard 2 sample  $t$ -test assumes that the variances are equal. Here's how you can test this assumption:

```
var.test(yL, yH)

##
## F test to compare two variances
##
## data:  yL and yH
## F = 1.1499, num df = 9, denom df = 9, p-value = 0.8386
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2856132 4.6293987
## sample estimates:
## ratio of variances
##      1.149877
```

The Caterpillar Data from class

```
location <- 1:12
untreated <- c(23,18,29,22,33,20,17,25,27,30,25,27)
treated <- c(19,18,24,23,31,22,16,23,24,26,24,28)
```

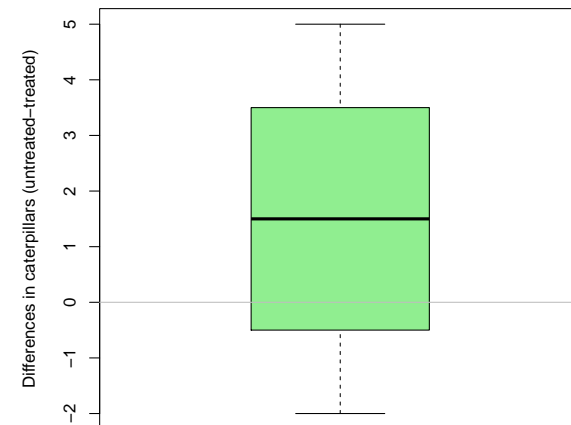
For paired  $t$ -tests, we want to know if the **mean of the differences** differs from zero

```
diff <- untreated-treated
diff

## [1] 4 0 5 -1 2 -2 1 2 3 4 1 -1

mean(diff) ## Estimate of the mean of the differences

## [1] 1.5
```



## PAIRED $t$ -TEST

**Recall:** Paired  $t$ -test is the same as a one-sample  $t$ -test on the differences. The hypothesis *in the Caterpillar example* is one-tailed:

- $H_0 : \mu_d \leq 0$
- $H_A : \mu_d > 0$

**Step 1.** Calculate the standard deviation of the differences.

$$s_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}$$

**Step 2.** Calculate the test statistic.

$$t = \frac{\bar{y} - 0}{s_d / \sqrt{n}}$$

**Step 3.** Compare to critical value.

## ASSIGNMENT

**Create a script to do the following:**

- (1) Do a paired  $t$  test on the caterpillar data without using the `t.test` function. Use only the functions `mean`, `sd`, and possibly `length`.
- (2) Do the paired  $t$  test again, but this time using the `t.test` function.
  - ▶ You will need to use the “paired” argument when using the `t.test` function
  - ▶ Assume variances are equal
- (3) Do a standard (*unpaired*) two-sample  $t$  test using the `t.test` function.
- (4) Add a comment to the end of your script explain your results. Also include the null and alternative hypotheses for each test.

**Upload your script<sup>2</sup> to ELC before next week’s lab.**

- The script must be self-contained.
- In other words, you should be able to copy and paste the entire thing into the **R** console, and it should return the correct answers to the questions.

**Read pp. 127–131 in “Introductory Statistics with R”**

<sup>2</sup>Or upload an Rmarkdown (.Rmd) file