Emma Rand
Data Analysis in R

Week 6 Two sample tests

Last week

We will cover
- General intro to one- and two-sample tests
- one sample-tests
  - The one-sample t-test
  - The paired sample t-test
  - The one sample Wilcoxon

R practice
  - No workshop
  - Activities to consolidate previous skills: Workflow basics and projects

Summary of this week

- Independent and non-independent samples
- Two-sample-tests
  - The two-sample t-test
  - The two sample Wilcoxon, also known as the Mann-Whitney
- In RStudio
  - t-tests and their non-parametric equivalents
  - Summarising, plotting and reporting

Learning objectives

By actively following the material and carrying out the independent study the successful student will be able to:

- Recognise when two samples are not independent (MLO 2)

- Select, appropriately, *t*-tests and their non-parametric equivalents (MLO 2)

- Know what functions are used in R to run these tests and how to interpret them(MLO 3 and 4)

- Know how to state the results of these tests scientifically (MLO 3 and 4)

- Create figures for these tests which are suitable for including in a scientific report (MLO 3 and 4)

Introduction to one- and two-sample tests

Reminder

Types of *t*-test

1. One-sample
   - Compares the mean of sample to a particular value (compares the response to a reference)
     - Includes paired-sample test – compares the mean difference to zero (i.e., compares dependent means)

2. Two-sample
   - Compares two (independent) means to each other

t-tests in general
Assumptions

All *t*-tests assume the "residuals" are normally distributed and have homogeneity of variance

A residual is the difference between the predicted and observed value

Predicted value is the mean / group mean

Paired-sample *t*-tests

- Two samples but values are not independent (could not reorder)
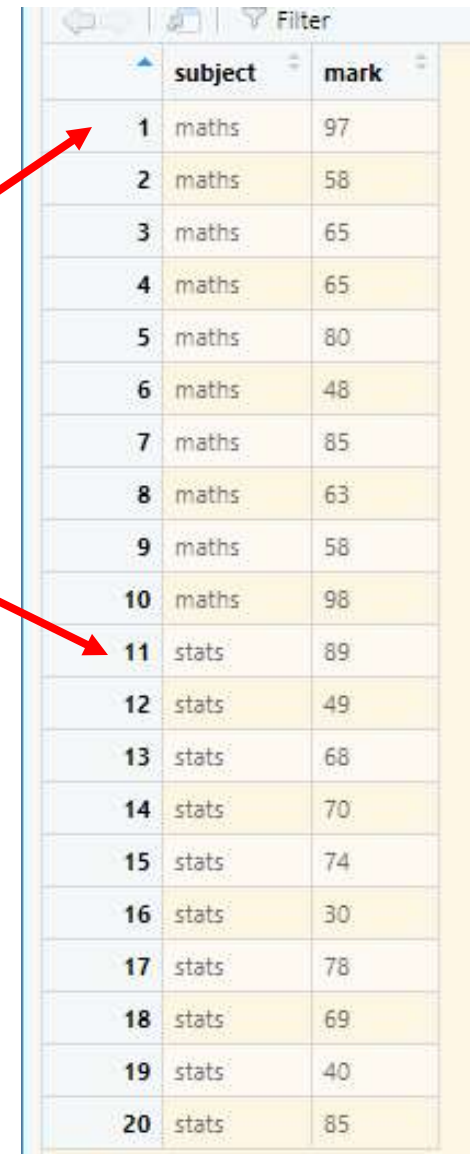
Actually a one-sample test

Paired-sample *t*-tests example

Is there a difference between the maths and stats marks of 10 students?

The one sample is the difference between the pairs of values

n.b. tidy data

Same student

	subject	mark
1	maths	97
2	maths	58
3	maths	65
4	maths	65
5	maths	80
6	maths	48
7	maths	85
8	maths	63
9	maths	58
10	maths	98
11	stats	89
12	stats	49
13	stats	68
14	stats	70
15	stats	74
16	stats	30
17	stats	78
18	stats	69
19	stats	40
20	stats	85

Paired-sample *t*-tests

Two samples but values are not independent.

They could be

- Same individual

- Same time or location

Key: do pairs of observations have something in common that make them more similar to each other than to other observations.

Two-sample *t*-tests

- Is there a difference between two independent means

    - Independent – values in one group not related to values in the other group

- Example: is there a difference between the masses of different subspecies of chaffinches?
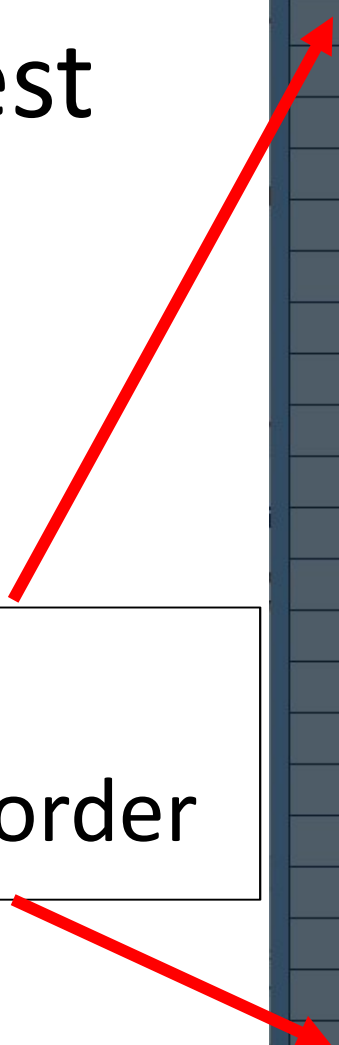
t-tests

two-sample *t*-test

Is there a difference between the masses of different subspecies of chaffinches?

No link
Could reorder

n.b. tidy data

	subspecies	mass
1	coelebs	18.3
2	coelebs	22.1
3	coelebs	22.4
4	coelebs	18.5
5	coelebs	22.2
6	coelebs	19.3
7	coelebs	17.8
8	coelebs	20.2
9	coelebs	22.1
10	coelebs	16.6
11	coelebs	20.7
12	coelebs	18.7
13	coelebs	22.6
14	coelebs	21.5
15	coelebs	21.7
16	coelebs	19.9
17	coelebs	23.1
18	coelebs	17.8
19	coelebs	19.5
20	coelebs	24.6
21	canariensis	22.7
22	canariensis	20.6
23	canariensis	25.4
24	canariensis	20.4

Showing 1 to 24 of 40 entries, 2 total columns

The two-sample *t*-test

A parametric test

two-sample *t*-test example

A number of [subspecies of the common chaffinch](#) have been described, based principally on the differences in the pattern and colour of the adult male plumage. Two of these subspecies are:

- "coelebs group" that occurs in Europe and Asia
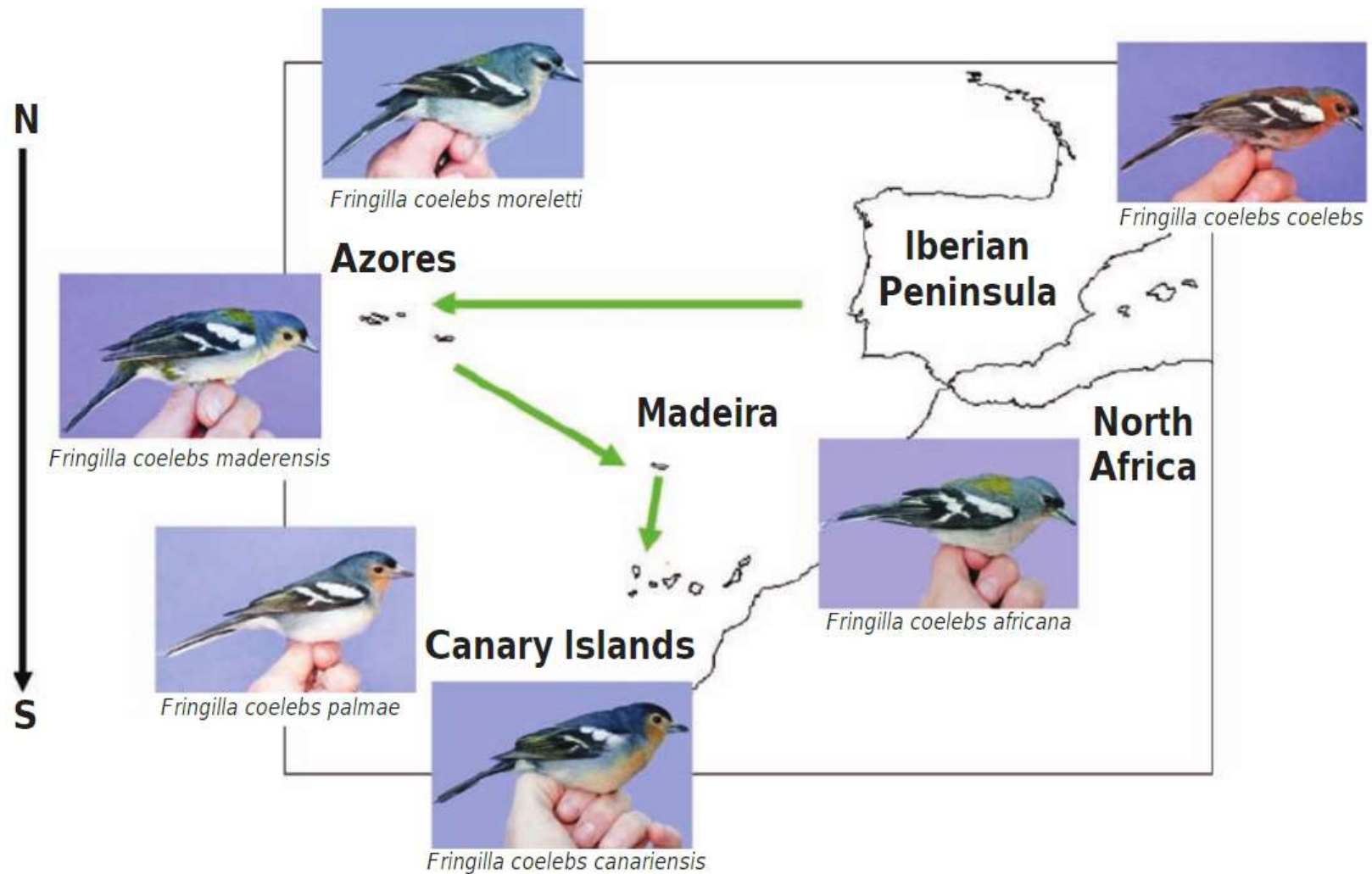- "canariensis group" on the Canary Islands

Figure 1. Distribution of the common chaffinch *Fringilla coelebs* in Macaronesia and nearby continental areas. Green lines depict path of colonisation of chaffinches in Macaronesia suggested by Marshall and Baker (1999).

t-tests

- Standard formula for all t-tests

$$t = \frac{statistic - hypothesised\ value}{s.e.\ of\ statistic}$$

- $t_{[d.f]} = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{s.e.\ of\ (\bar{x}_1 - \bar{x}_2)}$

To aid understanding, not for remembering

- d.f.= $n_1 + n_2 - 2$

	subspecies	mass
1	coelebs	18.3
2	coelebs	22.1
3	coelebs	22.4
4	coelebs	18.5
5	coelebs	22.2
6	coelebs	19.3
7	coelebs	17.8
8	coelebs	20.2
9	coelebs	22.1
10	coelebs	16.6
11	coelebs	20.7
12	coelebs	18.7
13	coelebs	22.6
14	coelebs	21.5
15	coelebs	21.7
16	coelebs	19.9
17	coelebs	23.1
18	coelebs	17.8
19	coelebs	19.5
20	coelebs	24.6
21	canariensis	22.7
22	canariensis	20.6
23	canariensis	25.4
24	canariensis	20.4

Showing 1 to 24 of 40 entries, 2 total columns

t-tests

Two-sample *t*-test example

```
chaff  <-  read_table("data/chaff.txt")
```

Note: these data are 'tidy'

All the responses in one column with other variables indicating the group

Organise your data this way

Tidy data

- Each variable should be in one column.
- Each different observation of that variable should be in a different row.
- There should be one table for each "kind" of data.
- If you have multiple tables, they should include a column in the table that allows them to be linked.

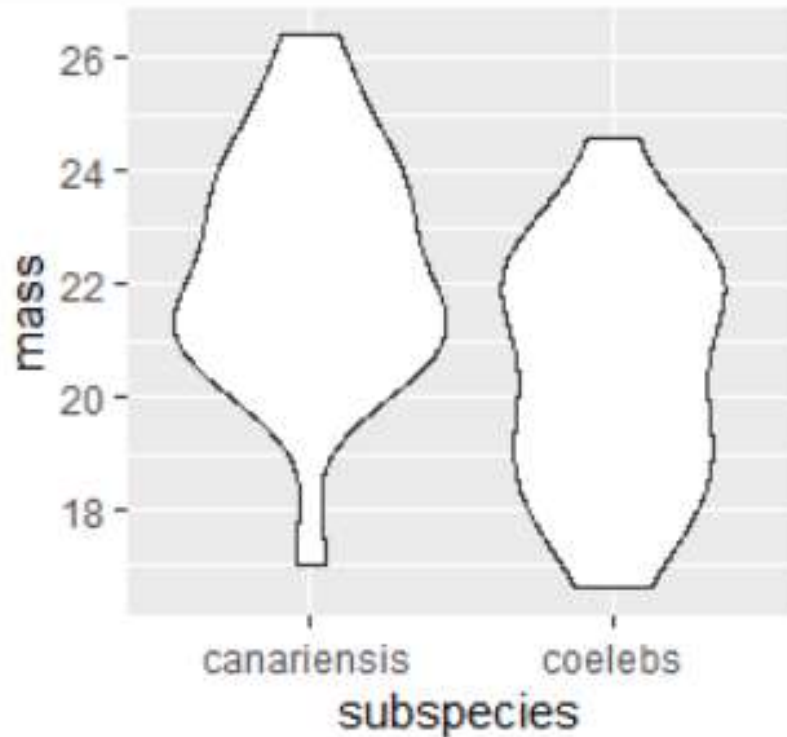Independent study: Wickham, H (2013). Tidy Data. Journal of Statistical Software.

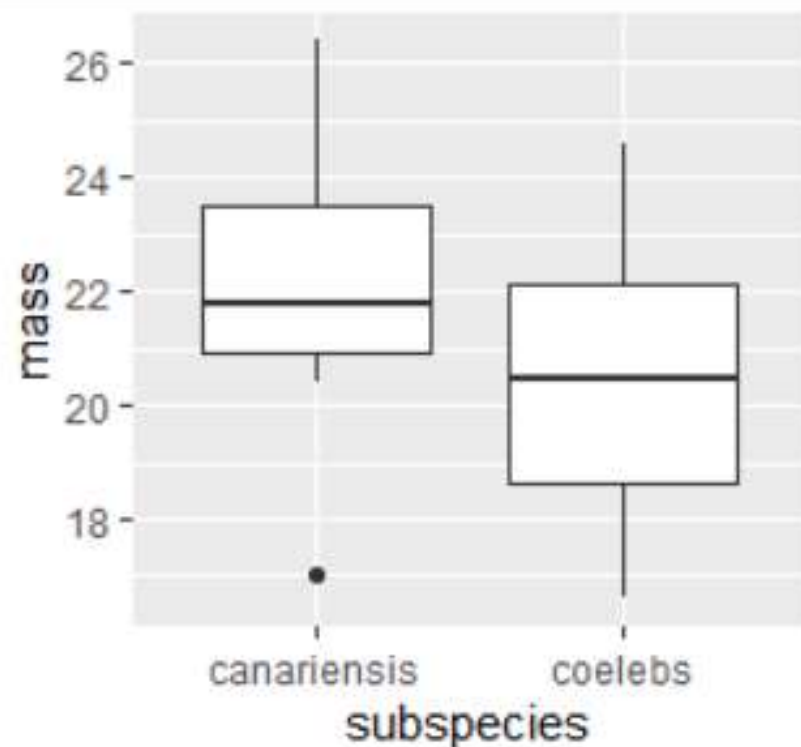https://www.jstatsoft.org/article/view/v059i10

Two-sample *t*-test example

Plot your data: roughly – perhaps one of these…

```
ggplot(data = chaff,
       aes(x = subspecies, y = mass)) +
  geom_violin()
```



```
ggplot(data = chaff,
       aes(x = subspecies, y = mass)) +
  geom_boxplot()
```

Two-sample *t*-test example

Summarise the data:

```
chaffsum <- chaff %>%
  group_by(subspecies) %>%
  summarise(mean = mean(mass),
            std = sd(mass),
            n = length(mass),
            se = std/sqrt(n))
```

```
chaffsum
# A tibble: 2 x 5
  subspecies    mean   std     n     se
  <chr>        <dbl> <dbl> <int>  <dbl>
1 canariensis  22.3  2.15    20  0.481
2 coelebs      20.5  2.14    20  0.478
```

Two-sample *t*-test example

Run the *t*-test

Name of the dataframe

```
t.test(data = chaff,
       mass ~ subspecies,
       var.equal = T)
```

The 'model' explain mass by subspecies

We are assuming homogeneity of variance

Paired = FALSE is the default.
No need to include

Two-sample *t*-test example

Run the *t*-test

```
t.test(data = chaff,
       mass ~ subspecies,
       var.equal = T)
```

Paired = FALSE is the default.
Two-sample test done by default

The model

```
        Two Sample t-test

data:  mass by subspecies
t = 2.6471, df = 38, p-value = 0.01175
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.422266 3.167734
sample estimates:
mean in group canariensis     mean in group coelebs
               22.275                       20.480
```
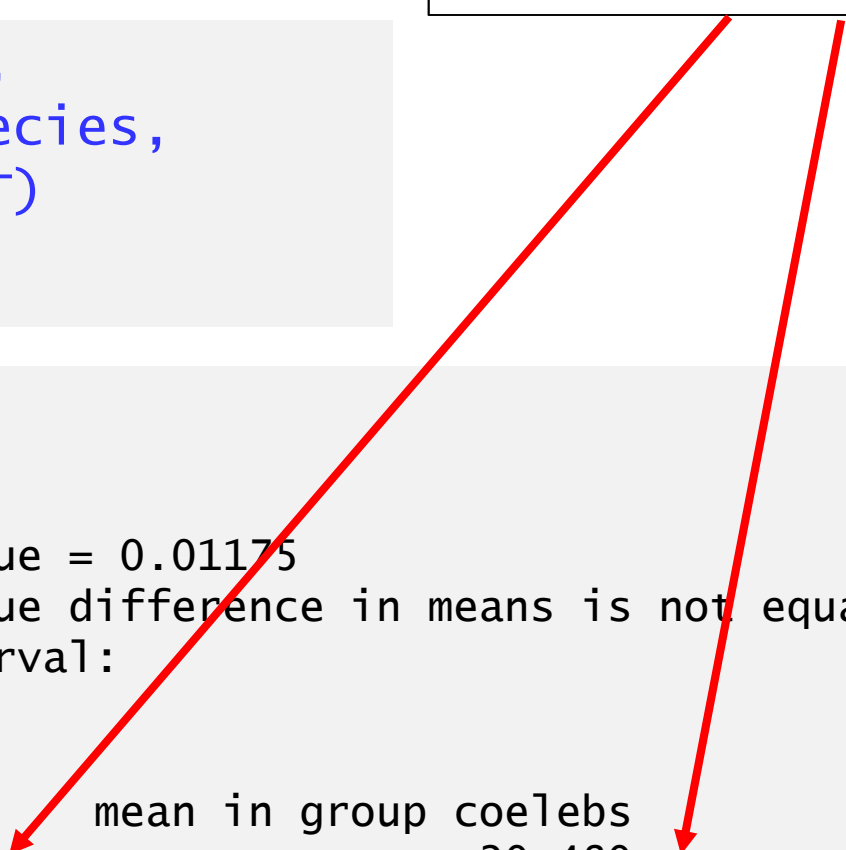
Two-sample *t*-test example

Run the *t*-test

```
t.test(data = chaff,
       mass ~ subspecies,
       var.equal = T)
```

The two group means

```
        Two Sample t-test

data:  mass by subspecies
t = 2.6471, df = 38, p-value = 0.01175
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.422266 3.167734
sample estimates:
mean in group canariensis     mean in group coelebs
              22.275                     20.480
```

t-tests
Two-sample *t*-test example
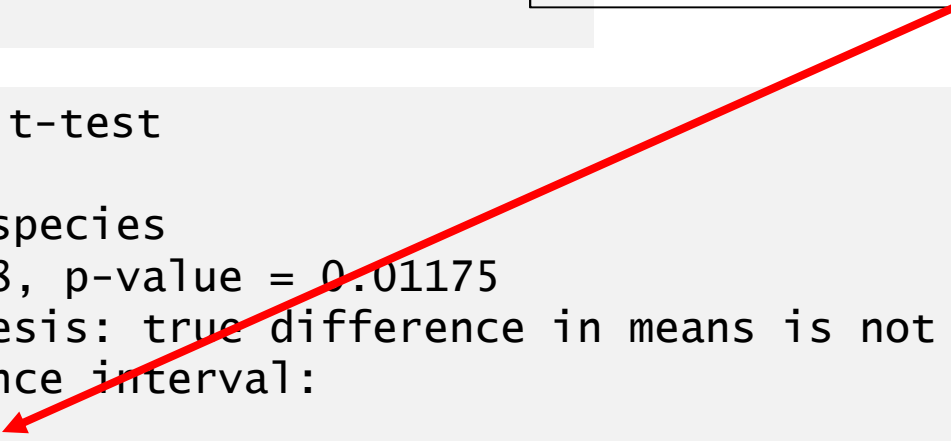
Run the *t*-test

```
t.test(data = chaff,
       mass ~ subspecies,
       var.equal = T)
```

95% CI on the difference between the two means

22.275 - 20.480 = 1.795

```
        Two Sample t-test

data:  mass by subspecies
t = 2.6471, df = 38, p-value = 0.01175
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.422266 3.167734
sample estimates:
mean in group canariensis     mean in group coelebs
              22.275                       20.480
```
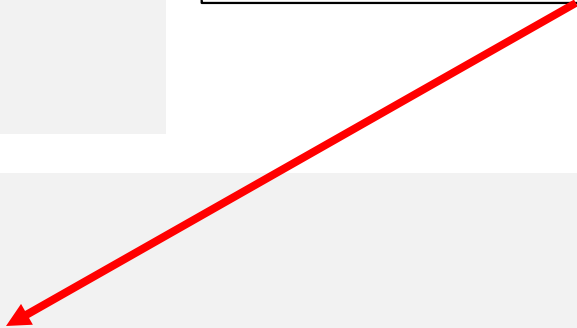
t-tests

Two-sample *t*-test example

Run the *t*-test

```
t.test(data = chaff,
       mass ~ subspecies,
       var.equal = T)
```

The bit that says whether it is significant *p* value

```
        Two Sample t-test

data:  mass by subspecies
t = 2.6471, df = 38, p-value = 0.01175
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.422266 3.167734
sample estimates:
mean in group canariensis      mean in group coelebs
                22.275                      20.480
```

Two-sample *t*-test example

Run the *t*-test

```
t.test(data = chaff,
       mass ~ subspecies,
       var.equal = T)
```

$p < 0.05$
Conclusion: there is a significant difference between the subspecies in mass

```
        Two Sample t-test

data:  mass by subspecies
t = 2.6471, df = 38, p-value = 0.01175
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.422266 3.167734
sample estimates:
mean in group canariensis     mean in group coelebs
              22.275                        20.480
```

Two-sample *t*-test example

Run the *t*-test

```
t.test(data = chaff,
       mass ~ subspecies,
       var.equal = T)
```

The values we will quote in results

```
        Two Sample t-test

data:  mass by subspecies
t = 2.6471, df = 38, p-value = 0.01175
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.422266 3.167734
sample estimates:
mean in group canariensis     mean in group coelebs
               22.275                    20.480
```

Two-sample *t*-test example

Check the assumptions

All *t*-tests assume the "residuals" are normally distributed and have homogeneity of variance

First use common sense: mass is a continuous and we would expect it to be normally distributed thus we would expect the residuals to be normally distributed

Two-sample *t*-test example

Second by plotting residuals:

calculate the residuals – the difference between predicted and observed (i.e., group mean and value)

```
# add the group means to the data
chaff <- merge(chaff, chaffsum[,1:2], by = "subspecies")

# add the residuals
chaff <- chaff %>%
  mutate(residual = mass - mean)
```
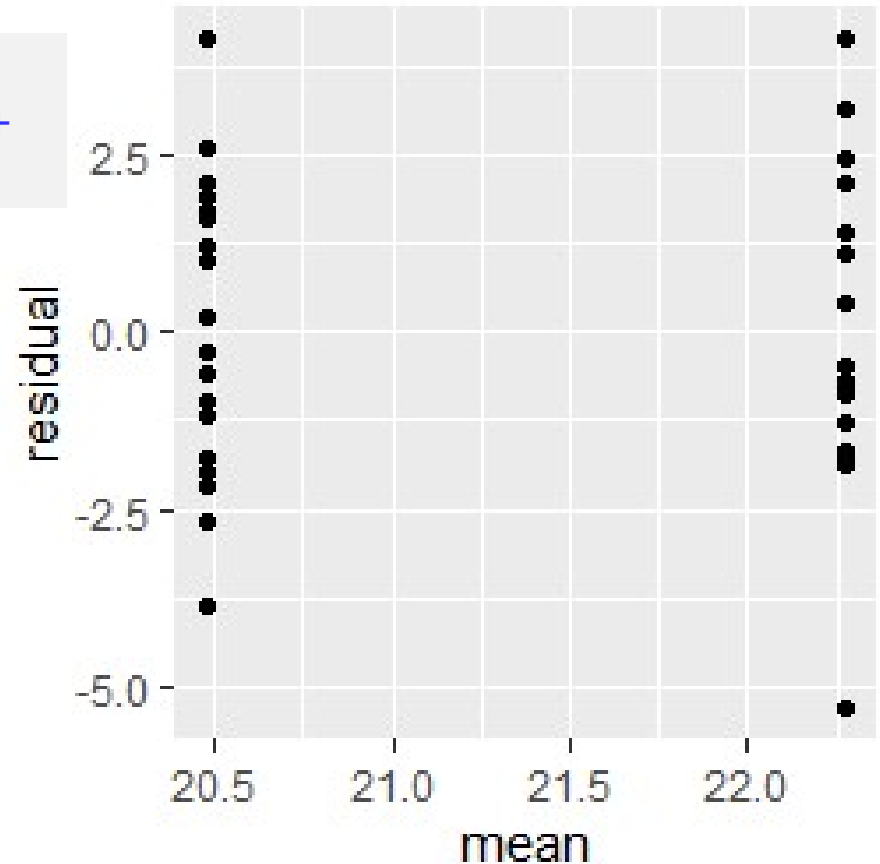
Two-sample *t*-test example

Checking the assumptions: normally and <u>homogenously</u> distributed residuals

```
ggplot(data = chaff,
       aes(x = mean, y = residual)) +
  geom_point()
```
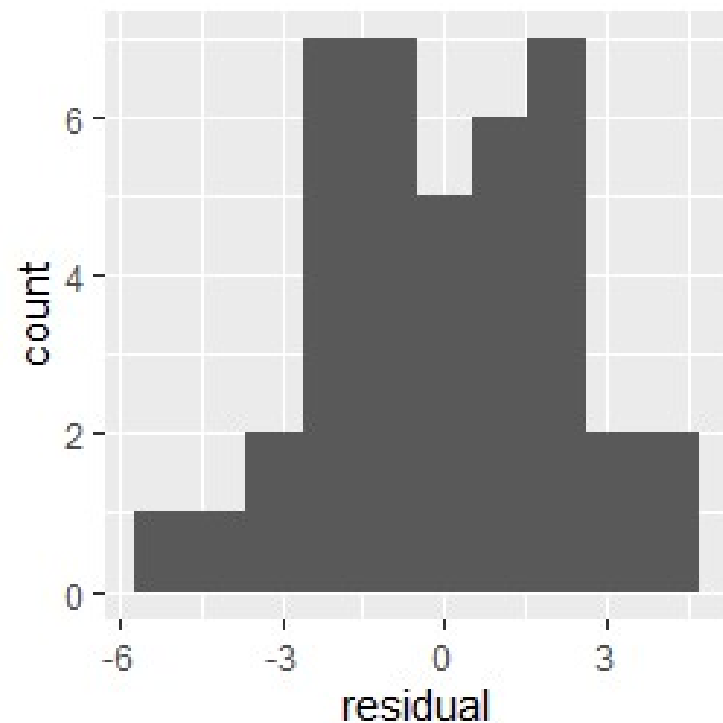
Variance is about the same for all values of x

Two-sample *t*-test example

Checking the assumptions: <u>normally</u> and homogenously distributed residuals

```
ggplot(data = chaff,
       aes(x = residual)) +
  geom_histogram(bins = 10)
```


Looks roughly normal symmetrical

Two-sample *t*-test example

Checking the assumptions: <u>normally</u> and homogenously distributed residuals

```
shapiro.test(chaff$residual)

        Shapiro-Wilk normality test

data:  chaff$residual
W = 0.98046, p-value = 0.7067
```

Not significantly different from a normal distribution

Two-sample *t*-test example

Reporting the result: "significance of effect, direction of effect, magnitude of effect"

Canariensis chaffinches ($\bar{x} \pm s.e$: 22.48 ± 0.48) are significantly heavier than coelebs (20.28 ± 0.48) ($t = 2.65$; $d.f. = 38$; $p = 0.012$). See figure 1.

Two-sample *t*-test: figures

Supports your claim:

— Show the data (all if possible)

— Show the 'model' (the predicted values i.e., means and error bars)

— Say what kind of error bars

— Full but concise figure legends
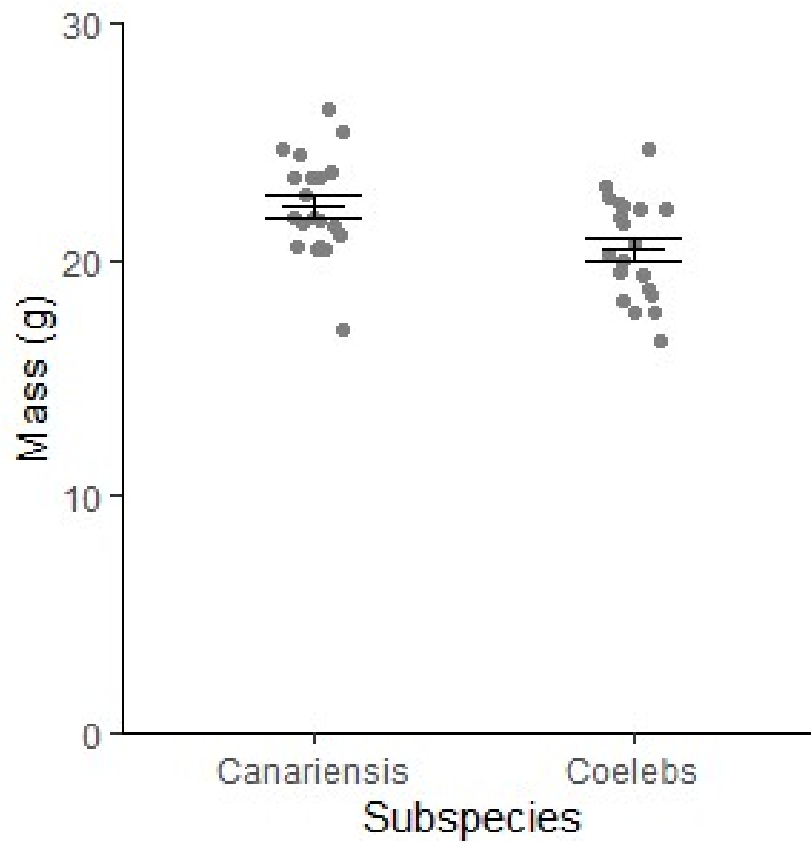
t-tests
Two-sample *t*-test: figures

Figure 1. Mean two subspecies of chaffinch. Error
bars are means +/- one standard error.

35

Two-sample *t*-tests - summary

- Parametric

- To test for a difference between two independent means

- *t* is size of difference relative to the s.e.

- Function in R:
  ```
  t.test(data = df, response ~ explanatory,
         var.equal = TRUE)
  ```

- If $p < 0.05$ the test is significant

- assumptions: normally and homogenously distributed residuals

- Significance, direction, magnitude

- Figure: data and 'model'

The two sample Wilcoxon
also known as the Mann-Whitney

Non-parametric equivalent of the
two-sample *t*-test

When the *t*-test assumptions are not met: non- parametric tests

- Non-parametric tests make fewer assumptions

- Based on the ranks rather than the actual data

- Null hypotheses are about the *mean rank* (not the mean)

t-test equivalents

i,.e., the type of question is the same but the response variable is not normally distributed or it is impossible to tell (small samples)

- one – sample *t*-test and paired-sample *t*-test: the one-sample Wilcoxon

- Two-sample *t*-test (next week): two-sample Wilcoxon aka Mann-Whitney

Non-parametric tests
two-sample Wilcoxon (Mann-Whitney)

Example: comparing the number of leaves on 8 mutant and wild type plants (small samples, counts)

Non-parametric tests
two-sample Wilcoxon (Mann-Whitney)

Plot your data: roughly – perhaps one of these…
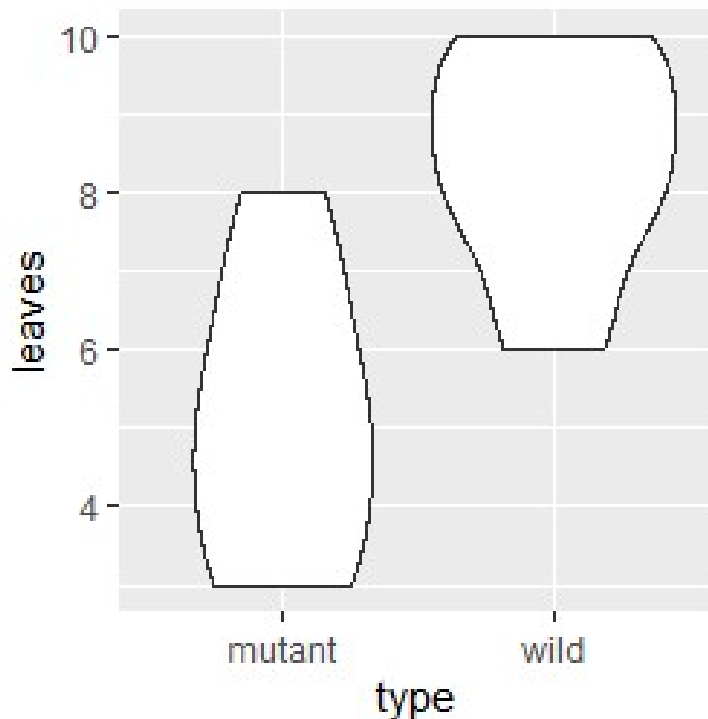
two-sample Wilcoxon (Mann-Whitney)

Summarise the data:

```
arabidopsissum <- arabidopsis %>%
   group_by(type) %>%
   summarise(median = median(leaves),
             n = length(leaves))
```

```
arabidopsissum
# A tibble: 2 x 3
  type    median      n
  <chr>    <dbl> <int>
1 mutant       5      8
2 wild       8.5      8
```

two-sample Wilcoxon (M-W): example

Carrying out the test two-sample Wilcoxon

```
wilcox.test(data = arabidopsis, leaves  ~ type)

        Wilcoxon rank sum test with continuity correction

data:  leaves by type
W = 5, p-value = 0.005051
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(x = c(3, 5, 6, 7, 3, 4, 5, 8), y = c(8, 9,  :
  cannot compute exact p-value with ties
```

No need to worry!

Non-parametric tests
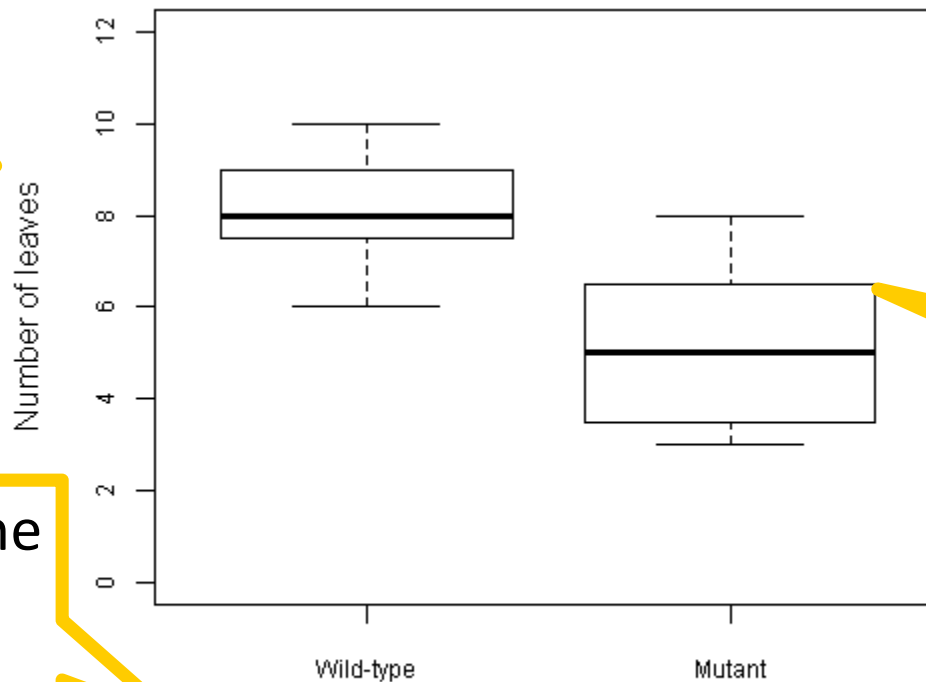# two-sample Wilcoxon (M-W): example

Reporting the result: "significance of effect, direction of effect, magnitude of effect"

There are significantly more leaves on wild-type (median = 8.5) than mutant (median = 5) plants (Mann-Whitney: W=5, n_1=8, n_2=8, p = 0.005051)

Non-parametric tests
two-sample Wilcoxon (M-W): example

Figure 2. Median (heavy lines) number of leaves on Wild-type and mutant plants. Boxes indicate the inter-quartile range.

Label with units

Non-parametric tests: use median+IQR

Measure of dispersion - IQR

What is the figure?

Always refer to figure in the text

Wilcoxon- summary

- Non-parametric

- when assumptions for *t*-test not met

- To test whether the mean rank in one group differs from another

- Function in R:
  `wilcox.test(data = ` *df* `, ` *response ~ explanatory* `)`

- If $p < 0.05$ the test is significant

- Few assumptions

- Figure: boxplot