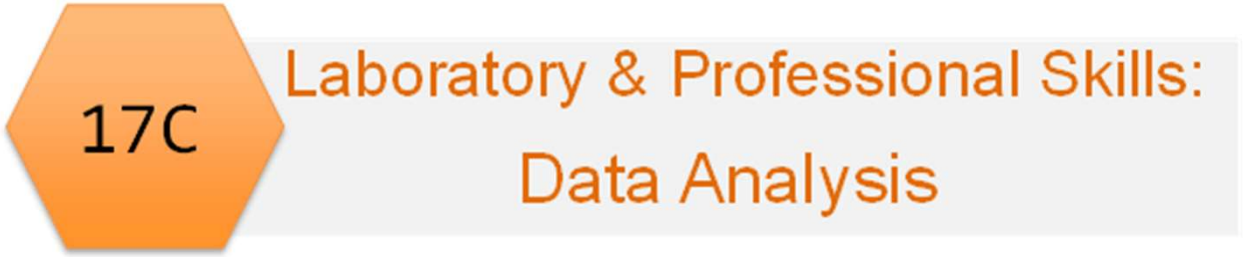


Laboratory & Professional skills for Bioscientists
Term 2: Data Analysis in R

More than one explanatory variable: Two-way ANOVA

Summary of this week

- Two-way ANOVA for more than one explanatory variable
  - Comparing to one-way
  - Rationale
  - The 3 null hypotheses
  - Running and interpreting the test
  - Understanding the interaction
  - Investigating the assumptions
  - Reporting the result

Learning objectives for the week

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

- Explain the rationale behind ANOVA and complete a partially filled ANOVA table (MLO 1 and 4)
- Read in data formatted for other statistical packages (MLO 3)
- Apply (appropriately), interpret and evaluate the legitimacy of, two-way ANOVA in R (MLO 2, 3 and 4)
- Explain the meaning of a significant interaction (MLO 4)
- Summarise and illustrate with appropriate figures test results scientifically (MLO 3 and 4)
- Use RStudio projects (MLO 4)

Revision (Lectures 6 and 7) Choosing tests

Steps - iterative

- Identify explanatory and response variables.
- The type of test depends on the type of type of data.
  - Categorical explanatory
    - Continuous response
      - One categorical explanatory variable: t-tests or one-way ANOVA
      - Two categorical explanatory variables: two-way ANOVA
  - Continuous explanatory
    - regression

Choosing between *t*-tests and one-way ANOVA

filecopy.R* × Untitled1* ×

Filter

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

Showing 1 to 23 of 40 entries

Two groups: two-sample *t*-test

Three groups: ANOVA

Without increasing Type I error

Untitled1* × Untitled2* ×

Filter

	values	population
1	10.31	A
2	13.07	A
3	10.33	A
4	10.52	A
5	11.67	A
6	7.27	A
7	10.31	B
8	13.07	B
9	10.33	B
10	10.52	B
11	11.67	B
12	7.27	B
13	10.31	C
14	13.07	C
15	10.33	C
16	10.52	C
17	11.67	C
18	7.27	C

Choosing between one-way and two-way ANOVA?

Response:
wing lengths

Explanatory:
species

Untitled1* × butter ×

Filter

	winglen	spp
1	23.6	F.flappa
2	23.3	F.flappa
3	18.2	F.flappa
4	22.6	F.flappa
5	29.3	F.flappa
6	22.2	F.flappa
7	24.5	F.flappa
8	26.3	F.flappa
9	20.6	F.flappa
10	23.9	F.flappa
11	26.5	F.flappa
12	24.7	F.flappa
13	28.3	F.flappa
14	22.3	F.flappa
15	21.8	F.flappa
16	30.0	F.flappa
17	21.5	F.flappa
18	20.1	F.flappa
19	24.3	F.flappa
20	27.2	F.flappa
21	28.6	F.concocti
22	17.2	F.concocti
23	20.4	F.concocti
24	21.9	F.concocti
25	26.3	F.concocti
26	27.8	F.concocti

Showing 1 to 26 of 40 entries

Choosing between one-way and two-way ANOVA?

What if we have two explanatory variables?

- Two one-way ANOVAs?? NO
- A Two-way ANOVA YES

- Note: tidy data format


lect 080 two-way ANOVA.R × butter ×

Filter

	winglen	spp	region
1	23.6	F.flappa	south
2	23.3	F.flappa	south
3	18.2	F.flappa	south
4	22.6	F.flappa	south
5	29.3	F.flappa	south
6	22.2		
7	24.5		
8	26.3		
9	20.6		
10	23.9		
11	26.5		
12	24.7	F.flappa	north
13	28.3	F.flappa	north
14	22.3	F.flappa	north
15	21.8	F.flappa	north
16	30.0	F.flappa	north
17	21.5	F.flappa	north
18	20.1	F.flappa	north
19	24.3	F.flappa	north
20	27.2	F.flappa	north
21	28.6	F.concocti	south
22	17.2	F.concocti	south
23	20.4	F.concocti	south
24	21.9	F.concocti	south

Showing 1 to 24 of 40 entries, 3 total columns

Explanatory:
species
region

Two-way ANOVA

Assumptions

Same as for one-way ANOVA

- Normality and ‘homoscedascity’ of residuals
- Common sense
- **Check after ANOVA** using the $residuals variable and diagnostic plots (as we did after one-way ANOVA)

Two-way ANOVA

Example

Response: wing lengths

Explanatory variables:

region: two levels

spp: two levels

lect 080 two-way ANOVA.R × | butter ×

Filter

	winglen	spp	region
1	23.6	F.flappa	south
2	23.3	F.flappa	south
3	18.2	F.flappa	south
4	22.6	F.flappa	south
5	29.3	F.flappa	south
6	22.2	F.flappa	south
7	24.5	F.flappa	south
8	26.3	F.flappa	south
9	20.6	F.flappa	south
10	23.9	F.flappa	south
11	26.5	F.flappa	north
12	24.7	F.flappa	north
13	28.3	F.flappa	north
14	22.3	F.flappa	north
15	21.8	F.flappa	north
16	30.0	F.flappa	north
17	21.5	F.flappa	north
18	20.1	F.flappa	north
19	24.3	F.flappa	north
20	27.2	F.flappa	north
21	28.6	F.concocti	south
22	17.2	F.concocti	south
23	20.4	F.concocti	south
24	21.9	F.concocti	south

Showing 1 to 24 of 40 entries, 3 total columns

Two-way ANOVA example

What does it test?

The null hypotheses here are:

1. mean of *F.flappa* (averaged over the regions) = mean of *F.concocti* (averaged over the regions),
2. mean of north (averaged over the spp) = mean of south (averaged over the spp) and
3. the effects of the two factors are independent.

Two-way ANOVA example

Reading in and examining the structure of the data

```
butter <- read.table("../data/butterf.txt", header=T)
glimpse(butter)
Observations: 40
Variables: 3
$ winglen <dbl> 23.6, 23.3, 18.2, 22.6, 29.3, 22.2, 24.5, 26.3, 20.6, 23.9...
$ spp     <fct> F.flappa, F.flappa, F.flappa, F.flappa, F.flappa, F.flappa...
$ region  <fct> south, south, south, south, south, south, south, south, so...
```

Assumptions

Common sense

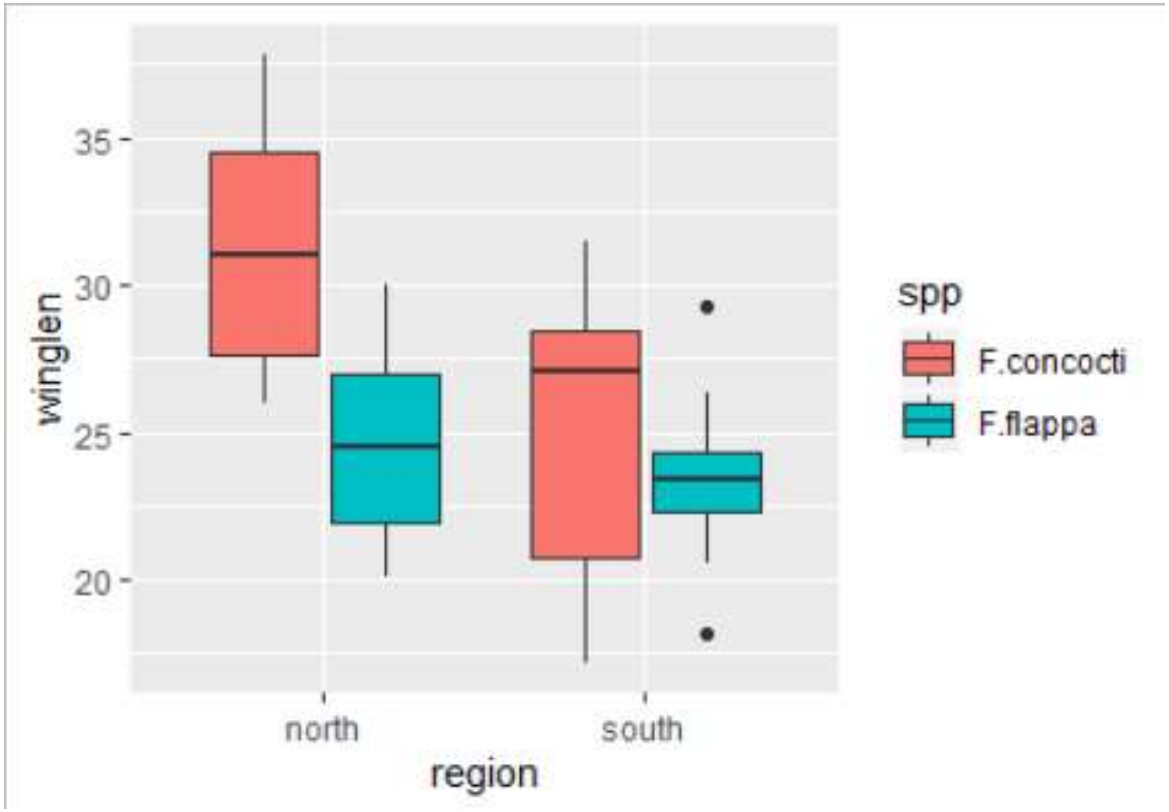
Can be checked after analysis

Two-way ANOVA example

Plot your data

Plot your data: roughly – perhaps..

```
ggplot(data = butter,
       aes(x = region, y = winglen, fill = spp)) +
  geom_boxplot()
```



Two-way ANOVA example

Plot your data

Sumarise

```
buttersum <- butter %>%
  group_by(region, spp) %>%
  summarise(mean = mean(winglen),
            median = median(winglen),
            sd = sd(winglen),
            n = length(winglen),
            se = sd/sqrt(n))
```

```
buttersum
# A tibble: 4 x 7
# Groups:   region [2]
  region spp          mean median    sd     n    se
  <fct>  <fct>       <dbl>  <dbl> <dbl> <int> <dbl>
1 north  F.concocti  31.4   31.0  4.28    10 1.35
2 north  F.flappa    24.7   24.5  3.27    10 1.03
3 south  F.concocti  25.0   27.0  4.96    10 1.57
4 south  F.flappa    23.4   23.5  3.01    10 0.953
```

Two-way ANOVA example

Plot your data

Run the anova

Name of the dataframe

```
mod <- aov(data = butter,
           winglen ~ region * spp)
```

The model: explain winglen by region, spp and the interaction between them

Assign result because we will be able to access residuals from this object later

Two-way ANOVA example

Understanding the test output

```
mod <- aov(data = butter, winglen ~ region * spp)
summary(mod)
             Df Sum Sq Mean Sq F value   Pr(>F)
region        1 145.16 145.161  9.2717 0.004334 **
spp           1 168.92 168.921 10.7893 0.002280 **
region:spp    1  67.08  67.081  4.2846 0.045692 *
Residuals    36 563.63  15.656
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

1. There is an effect of region (difference between regions)
2. There is an effect of species (difference between species)
3. There is an interaction between region and species…..

Two-way ANOVA example

Understanding the test output

```
mod <- aov(data = butter, winglen ~ region * spp)
summary(mod)
             Df Sum Sq Mean Sq F value   Pr(>F)
region        1 145.16 145.161  9.2717 0.004334 **
spp           1 168.92 168.921 10.7893 0.002280 **
region:spp    1  67.08  67.081  4.2846 0.045692 *
Residuals    36 563.63  15.656
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Total d.f. is no. of values – 1: 40 – 1 = 39
region d.f. is no. regions – 1: 2 – 1 = 1
spp d.f. is no. spp – 1: 2 – 1 = 1
Interaction d.f. is region d.f. * spp d.f. : 1* 1 = 1
Residual d.f. is total d.f. – all other d.f.: 39 – 1 – 1 – 1 = 36

Two-way ANOVA example

Understanding the test output

```
mod <- aov(data = butter, winglen ~ region * spp)
summary(mod)
             Df Sum Sq Mean Sq F value   Pr(>F)
region        1 145.16 145.161  9.2717 0.004334 **
spp           1 168.92 168.921 10.7893 0.002280 **
region:spp    1  67.08  67.081  4.2846 0.045692 *
Residuals    36 563.63  15.656
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

‘Error term’ for all 3 tests

Two-way ANOVA example

Checking Assumptions

- Common sense
  - response should be continuous
  - No/few repeats
- Plot the residuals
- Using a test in R

Two-way ANOVA

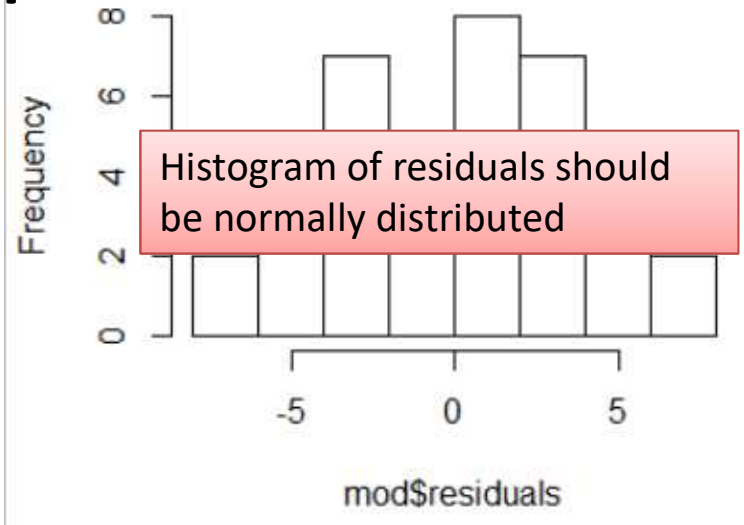
Checking Assumptions

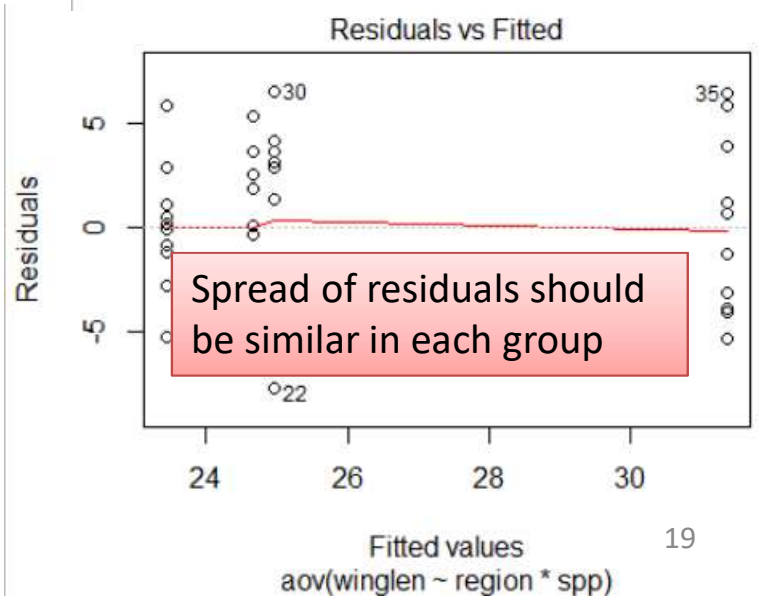
Residuals are calculated for you already!

```
hist(mod$residuals)
shapiro.test(mod$residuals)

   Shapiro-Wilk normality test

data:  mod$residuals
W = 0.97306, p-value = 0.4474
plot(mod, which=1)
```





Two-way ANOVA example

Reporting the result

Reporting the result: “significance, direction, magnitude”

There was a significant difference between the species (ANOVA: $F = 10.79$; $d.f. = 1,36$; $p = 0.002$) and between the regions ($F = 9.27$; $d.f. = 1,36$; $p = 0.004$). However, there was also a significant interaction between region and species ($F = 4.28$; $d.f. = 1,36$; $p = 0.046$)

What about direction and magnitude??

Two-way ANOVA example

Reporting the result: Post-hoc?

Post-hoc test e.g., Tukey

John Wilder Tukey

Wild Turkey

Wild Turkey

Two-way ANOVA example

Reporting the result

Which means differ? Post-hoc test needed e.g., Tukey

3 parts to the output. First two parts for region and spp

```
TukeyHSD(mod)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = winglen ~ region * spp, data = butter)

$region
              diff       lwr       upr    p adj
south-north -3.81 -6.347658 -1.272342 0.004334

$spp
                      diff       lwr       upr     p adj
F.flappa-F.concocti -4.11 -6.647658 -1.572342 0.0022796
```

Two-way ANOVA example

Reporting the result

Which means differ? Post-hoc test needed e.g., Tukey

3 parts to the output. Third part for the interaction

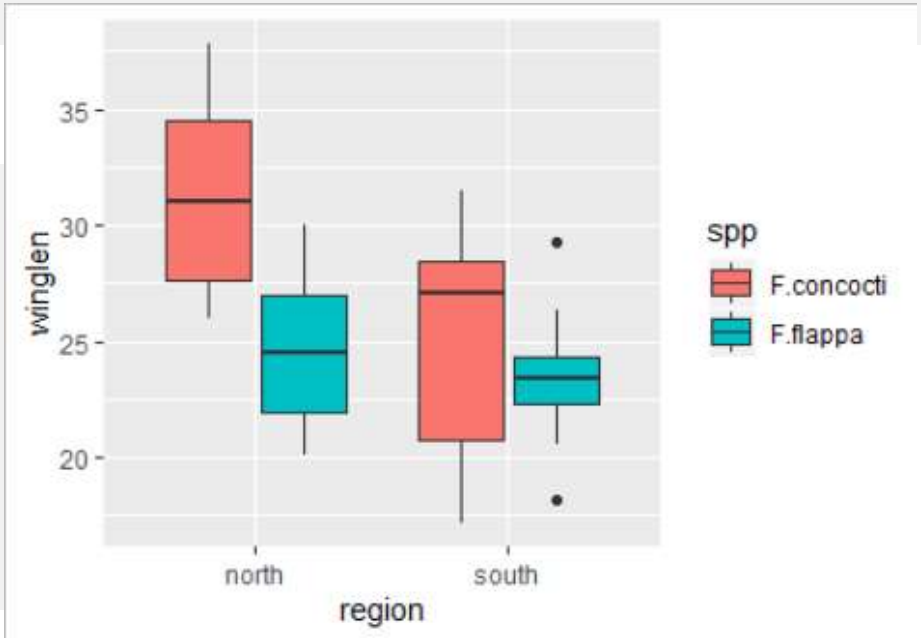
```
$`region:spp`
                                      diff        lwr       upr     p adj
south:F.concocti-north:F.concocti -6.40 -11.165769 -1.634231 0.0048102
north:F.flappa-north:F.concocti   -6.70 -11.465769 -1.934231 0.0030099
south:F.flappa-north:F.concocti   -7.92 -12.685769 -3.154231 0.0004123
north:F.flappa-south:F.concocti   -0.30  -5.065769  4.465769 0.9982343
south:F.flappa-south:F.concocti   -1.52  -6.285769  3.245769 0.8257284
south:F.flappa-north:F.flappa     -1.22  -5.985769  3.545769 0.9004525
```

Two-way ANOVA example

Reporting the result: direction and magnitude

```
$`region:spp`
                                        diff        lwr       upr     p adj
south:F.concocti-north:F.concocti -6.40 -11.165769 -1.634231 0.0048102
north:F.flappa-north:F.concocti   -6.70 -11.465769 -1.934231 0.0030099
south:F.flappa-north:F.concocti   -7.92 -12.685769 -3.154231 0.0004123
north:F.flappa-south:F.concocti   -0.30  -5.065769  4.465769 0.9982343
south:F.flappa-south:F.concocti   -1.52  -6.285769  3.245769 0.8257284
south:F.flappa-north:F.flappa     -1.22  -5.985769  3.545769 0.9004525
```

```
buttersum
# A tibble: 4 x 7
# Groups:   region [2]
  region spp          mean
  <fct>  <fct>       <dbl>
1 north  F.concocti   31.4
2 north  F.flappa     24.7
3 south  F.concocti   25.0
4 south  F.flappa     23.4
```



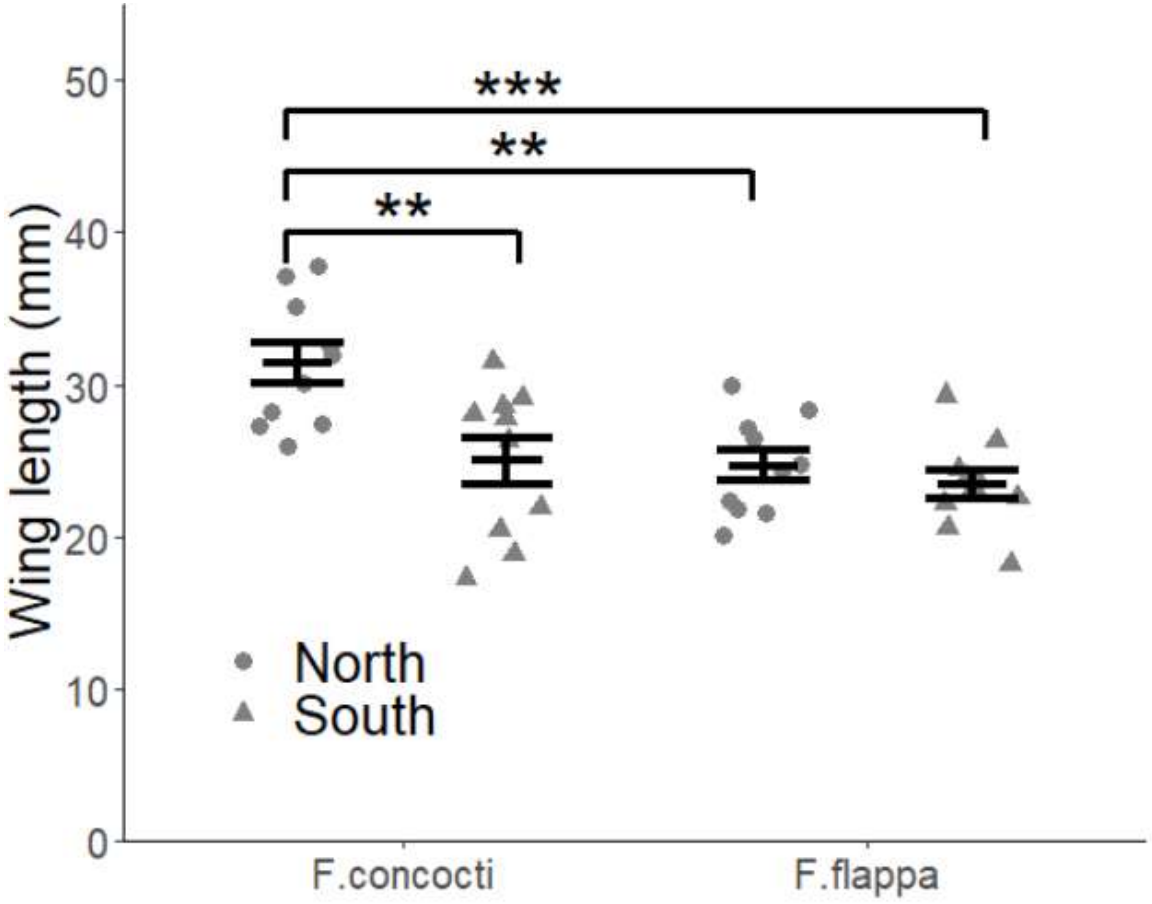
Two-way ANOVA example

Reporting the result: direction and magnitude

F.concocti had significantly longer wings than *F.flappa* (ANOVA: F = 10.79; *d.f*. = 1,36; p = 0.002) and individuals were significantly bigger in the North than the South (F = 9.27; *d.f.* = 1,36; p = 0.004). However, there was also a significant interaction between region and species (F = 4.28; *d.f.* = 1,36; p = 0.046) with a significant difference between regions for *F.concocti* (Tukey Honest Significant difference: p = 0.005) but not for *F.flappa.* (Figure 1).

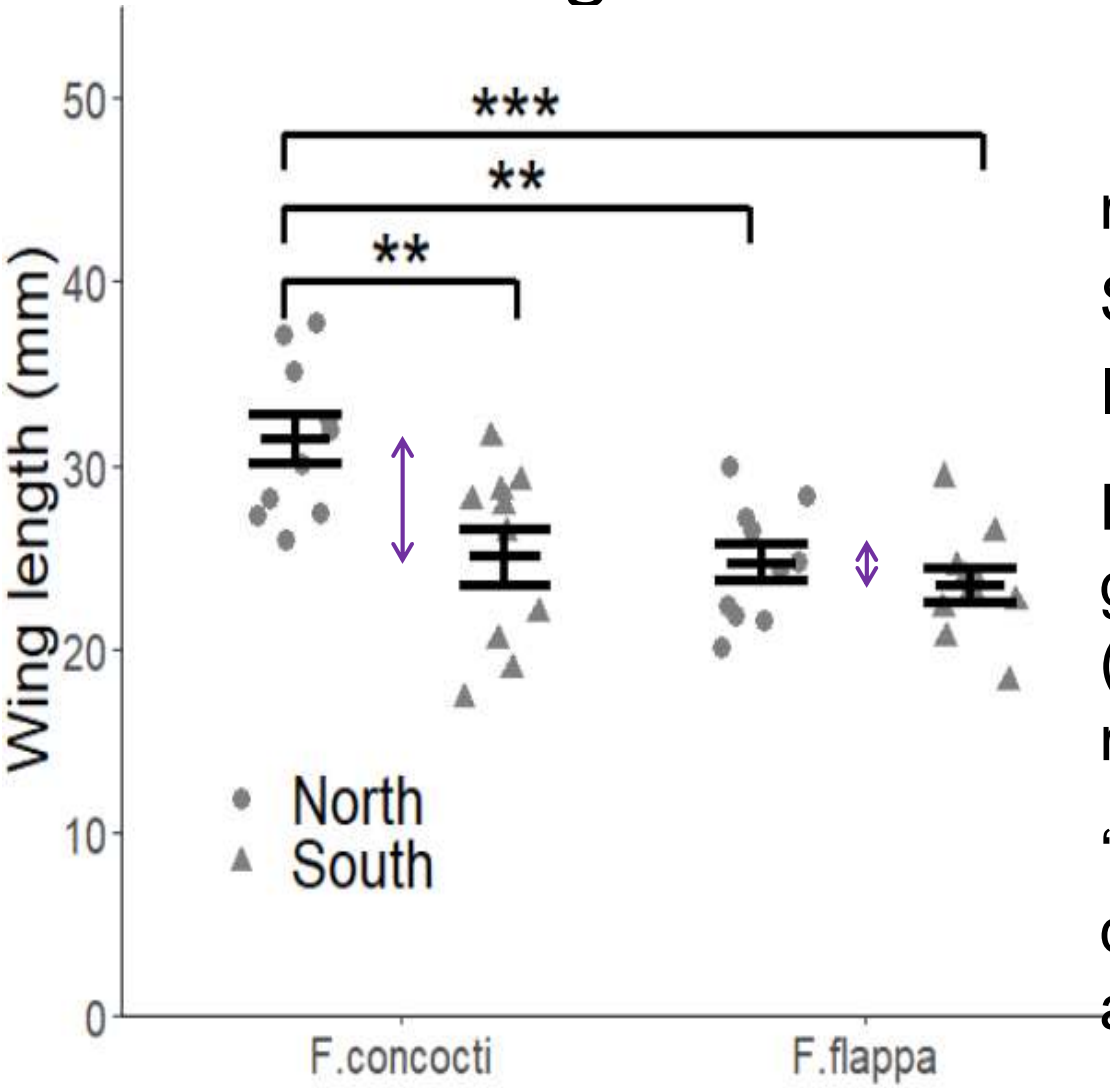
Two-way ANOVA example

Reporting the result: figure



Two-way ANOVA example

Understanding the interaction from the figure



region – Sig

Spp – Sig

Int – Sig

Effect of region is greater in *F.concocti* (i.e., the gap between regions is bigger)

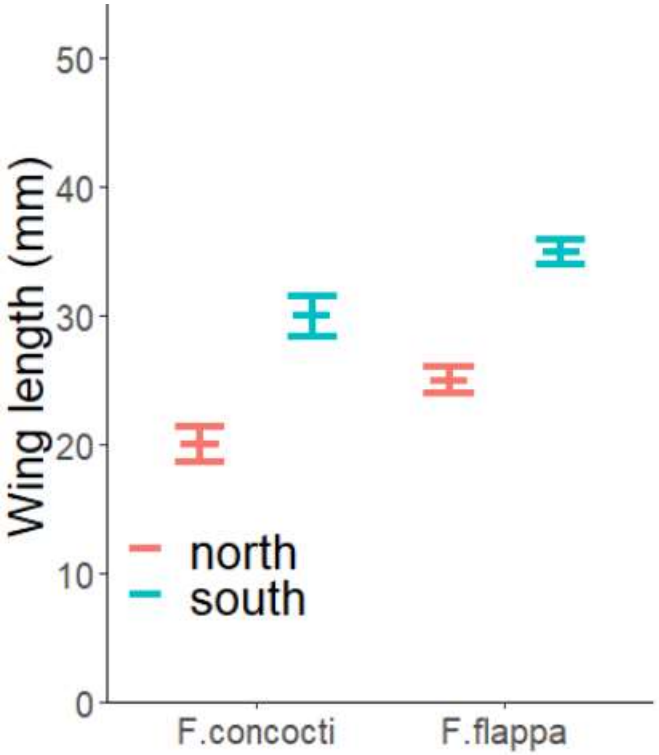
'effect of one factor depends on the level of another'

Two-way ANOVA example

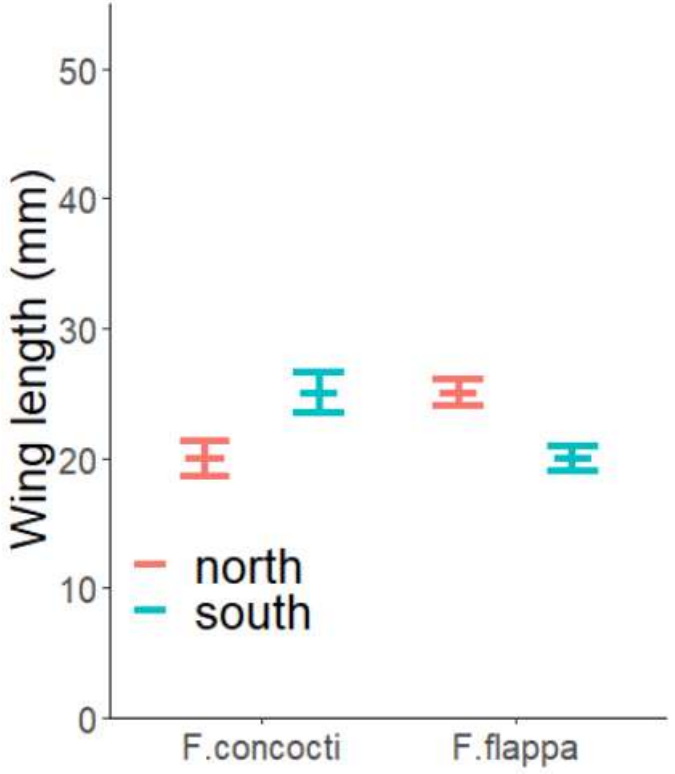
Understanding the interaction from the figure

Some other possible results

No interaction:Gap the same

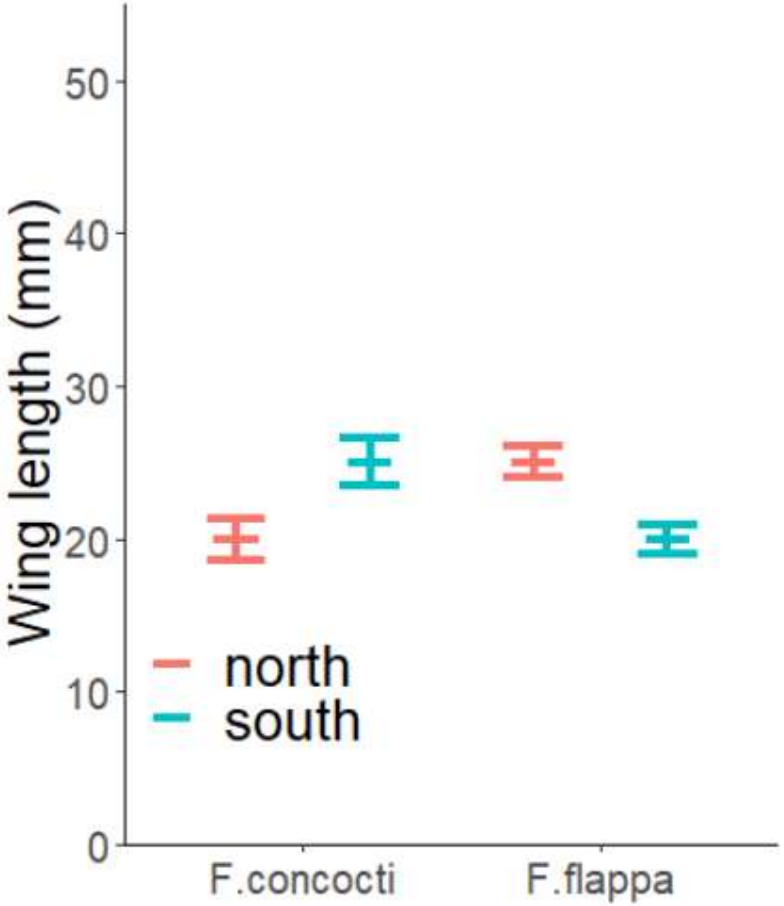


Interaction: Gap the reversed



Two-way ANOVA example

Understanding the interaction from the figure



Region – NS

Spp – NS

Int – Sig

But region does have an effect!

It is just reversed!

If you have a significant interaction, interpret main effects with care. Look at the Post-hoc test

Learning objectives for the week

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

- Explain the rationale behind ANOVA and complete a partially filled ANOVA table (MLO 1 and 4)
- Read in data formatted for other statistical packages (MLO 3)
- Apply (appropriately), interpret and evaluate the legitimacy of, two-way ANOVA in R (MLO 2, 3 and 4)
- Explain the meaning of a significant interaction (MLO 4)
- Summarise and illustrate with appropriate figures test results scientifically (MLO 3 and 4)
- Use RStudio projects (MLO 4)