Statistics with SpaRRows

Lecture 13

Julia Schroeder

Julia.schroeder@imperial.ac.uk

ANalysis Of VAriance – ANOVA

- ANOVA is really nothing scary
- It’s actually just a linear model
- You’ve run one already!

ANalysis **O**f **Va**riance – ANOVA

- ANOVA is really nothing scary
- It's actually just a linear model
- You've run one already!

- Tests for differences of variances between groups

ANOVA

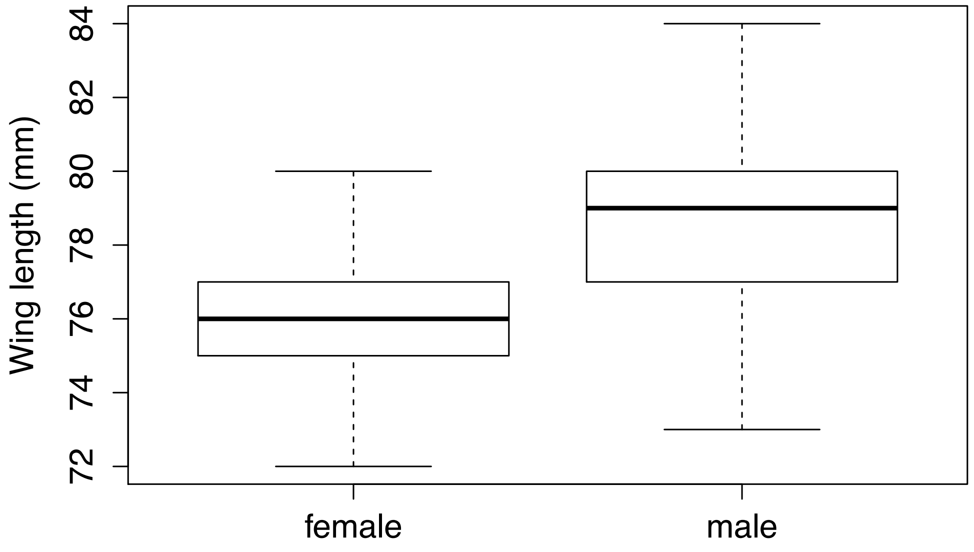
- Null-hypothesis: groups are equal

ANOVA

- Null-hypothesis: groups are equal
- Idea: variability between groups is larger than within groups

ANOVA

- Null-hypothesis: groups are equal
- Idea: variability between groups is larger than within groups

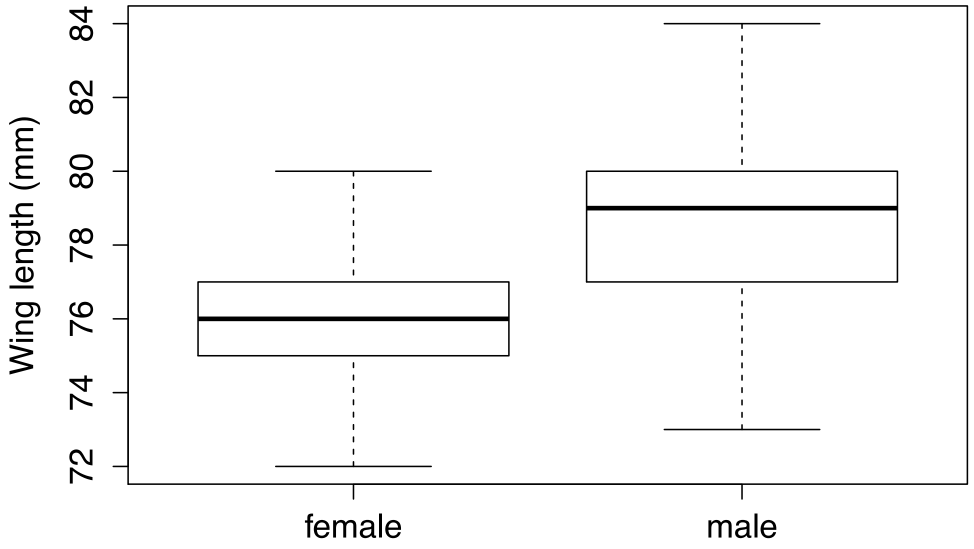


ANOVA

- Null-hypothesis: groups are equal

- Idea: variability between groups is larger than within groups

- Partition variance into parts between
- And within-groups:



ANOVA

- So, it’s slightly different from a t-test where we test for a difference in mean:

- Need three things
- Within-group variation
- Between-group variation
- Total variation (sum of two before)

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1

b_i =3,3,4,3,3,3

c_i =5,5,4,5,5,1

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1

b_i =3,3,4,3,3,3

c_i =5,5,4,5,5,1

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\mu = 1.17$

b_i =3,3,4,3,3,3 $\mu = 3.17$

c_i =5,5,4,5,5,1 $\mu = 4.17$

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\mu = 1.17$

b_i =3,3,4,3,3,3 $\mu = 3.17$

c_i =5,5,4,5,5,1 $\mu = 4.17$

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\mu = 1.17$ $SS = 0.83$

b_i =3,3,4,3,3,3 $\mu = 3.17$ $SS = 0.83$

c_i =5,5,4,5,5,1 $\mu = 4.17$ $SS = 12.83$

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\quad \mu = 1.17 \quad SS = 0.83$

b_i =3,3,4,3,3,3 $\quad \mu = 3.17 \quad SS = 0.83$

c_i =5,5,4,5,5,1 $\quad \mu = 4.17 \quad SS = 12.83$

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\mu = 1.17$ $SS = 0.83$ $\sigma^2 = 0.14$

b_i =3,3,4,3,3,3 $\mu = 3.17$ $SS = 0.83$ $\sigma^2 = 0.14$

c_i =5,5,4,5,5,1 $\mu = 4.17$ $SS = 12.83$ $\sigma^2 = 2.14$

Between-group stuff

1,1,2,1,1,1 5,5,4,5,5,1 3,3,4,3,3,3

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\mu = 1.17$ $SS = 0.83$

b_i =3,3,4,3,3,3 $\mu = 3.17$ $SS = 0.83$

c_i =5,5,4,5,5,1 $\mu = 4.17$ $SS = 12.83$

Between-group stuff

1,1,2,1,1,1 5,5,4,5,5,1 3,3,4,3,3,3

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\mu = 1.17$ $SS = 0.83$

b_i =3,3,4,3,3,3 $\mu = 3.17$ $SS = 0.83$

c_i =5,5,4,5,5,1 $\mu = 4.17$ $SS = 12.83$

$SS_E = 14.5$

Between-group stuff

1,1,2,1,1,1 5,5,4,5,5,1 3,3,4,3,3,3

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\mu = 1.17$ $\text{SS} = 0.83$

b_i =3,3,4,3,3,3 $\mu = 3.17$ $\text{SS} = 0.83$

c_i =5,5,4,5,5,1 $\mu = 4.17$ $\text{SS} = 12.83$

$\text{SS}_\text{E} = 14.5$

Between-group stuff

1,1,2,1,1,1 5,5,4,5,5,1 3,3,4,3,3,3

$\bar{a}, \bar{b}, \bar{c}$

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\mu = 1.17$ $SS = 0.83$

b_i =3,3,4,3,3,3 $\mu = 3.17$ $SS = 0.83$ $SS_E = 14.5$

c_i =5,5,4,5,5,1 $\mu = 4.17$ $SS = 12.83$

Between-group stuff

1,1,2,1,1,1 5,5,4,5,5,1 3,3,4,3,3,3

1.17, 3.17, 4.17

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\mu = 1.17$ $SS = 0.83$

b_i =3,3,4,3,3,3 $\mu = 3.17$ $SS = 0.83$ $SS_E = 14.5$

c_i =5,5,4,5,5,1 $\mu = 4.17$ $SS = 12.83$

Between-group stuff

1,1,2,1,1,1 5,5,4,5,5,1 3,3,4,3,3,3

$1.17, 3.17, 4.17$ $\mu = 2.84$

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\mu = 1.17$ $SS = 0.83$

b_i =3,3,4,3,3,3 $\mu = 3.17$ $SS = 0.83$ $SS_E = 14.5$

c_i =5,5,4,5,5,1 $\mu = 4.17$ $SS = 12.83$

Between-group stuff

1,1,2,1,1,1 5,5,4,5,5,1 3,3,4,3,3,3

$1.17, 3.17, 4.17$ $\mu = 2.84$

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\mu = 1.17$ $SS = 0.83$

b_i =3,3,4,3,3,3 $\mu = 3.17$ $SS = 0.83$ $SS_E = 14.5$

c_i =5,5,4,5,5,1 $\mu = 4.17$ $SS = 12.83$

Between-group stuff

1,1,2,1,1,1 5,5,4,5,5,1 3,3,4,3,3,3

$1.17, 3.17, 4.17$ $\mu = 2.84$ $SS = 28$ <- weigh each square with n of group!

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\quad \mu = 1.17 \quad SS = 0.83$

b_i =3,3,4,3,3,3 $\quad \mu = 3.17 \quad SS = 0.83$ $\quad SS_E = 14.5$

c_i =5,5,4,5,5,1 $\quad \mu = 4.17 \quad SS = 12.83$

Between-group stuff

1,1,2,1,1,1 $\quad$ 5,5,4,5,5,1 $\quad$ 3,3,4,3,3,3

$SS_B = \sum n_i (x_i - \bar{x})^2$

$1.17, 3.17, 4.17 \quad \mu = 2.84 \quad SS_B = 28$

<- weigh each square with n of group!

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\mu = 1.17$ $SS = 0.83$

b_i =3,3,4,3,3,3 $\mu = 3.17$ $SS = 0.83$ $SS_E = 14.5$

c_i =5,5,4,5,5,1 $\mu = 4.17$ $SS = 12.83$

Between-group stuff

1,1,2,1,1,1 5,5,4,5,5,1 3,3,4,3,3,3

$1.17, 3.17, 4.17$ $\mu = 2.84$ $SS_B = 28$

Total stuff

1,1,2,1,1,1,3,3,4,3,3,3,5,5,4,5,5,1

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\mu = 1.17$ $SS = 0.83$

b_i =3,3,4,3,3,3 $\mu = 3.17$ $SS = 0.83$ $SS_E = 14.5$

c_i =5,5,4,5,5,1 $\mu = 4.17$ $SS = 12.83$

Between-group stuff

1,1,2,1,1,1 5,5,4,5,5,1 3,3,4,3,3,3

$1.17, 3.17, 4.17$ $\mu = 2.84$ $SS_B = 28$

Total stuff

1,1,2,1,1,1,3,3,4,3,3,3,5,5,4,5,5,1

$\mu = 2.83$ $SS_T = 42.5$

ANOVA

- Within-group (SSE) + Between-group (SSG) = Total Sums of Squares (SST)

ANOVA

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Within-group stuff

a_i =1,1,2,1,1,1 $\mu = 1.17$ $SS = 0.83$

b_i =3,3,4,3,3,3 $\mu = 3.17$ $SS = 0.83$ $SS_E = 14.5$

c_i =5,5,4,5,5,1 $\mu = 4.17$ $SS = 12.83$

Between-group stuff

1,1,2,1,1,1 5,5,4,5,5,1 3,3,4,3,3,3

$1.17, 3.17, 4.17$ $\mu = 2.84$ $SS_B = 28$

Total stuff

1,1,2,1,1,1,3,3,4,3,3,3,5,5,4,5,5,1

$\mu = 2.83$ $SS_T = 42.5$

ANOVA

- Within-group (SSE) + Between-group (SSG) = Total Sums of Squares (SST)
- YES!

ANOVA

- Within-group (SSE) + Between-group (SSG) = Total Sums of Squares (SST)

Source	SS	df	MS	F
Group	SSG	k-1	MSG = SSG /(k-1)	MSG/MSE
Residual	SSE	n-k	MSE = SSE /(n-k)	
Total	SST	n-1		

- K = n of groups
- N = sample size

ANOVA

- Within-group (SSE) + Between-group (SSG) = Total Sums of Squares (SST)

Source		SS	df		MS	F
Group	28	SSG	k-1	2	MSG = SSG /(k-1)	MSG/MSE
Residual	14.5	SSE	n-k	15	MSE = SSE /(n-k)	
Total	42.5	SST	n-1	17		

- K = n of groups 3
- N = sample size 3 *6 = 18

ANOVA

- Within-group (SSE) + Between-group (SSG) = Total Sums of Squares (SST)

Source		SS	df		MS		F
Group	28	SSG	k-1	2	MSG = SSG /(k-1)	14	MSG/MSE
Residual	14.5	SSE	n-k	15	MSE = SSE /(n-k)	0.97	
Total	42.5	SST	n-1	17			

- K = n of groups 3
- N = sample size 3 *6 = 18

ANOVA

- Within-group (SSE) + Between-group (SSG) = Total Sums of Squares (SST)

Source		SS	df		MS		F	
Group	28	SSG	k-1	2	MSG = SSG /(k-1)	14	MSG/MSE	
Residual	14.5	SSE	n-k	15	MSE = SSE /(n-k)	0.97		14.43
Total	42.5	SST	n-1	17				

- K = n of groups 3
- N = sample size 3 *6 = 18

ANOVA

```
> anova(lm(y~as.factor(x)))
Analysis of Variance Table

Response: y
             Df Sum Sq Mean Sq F value    Pr(>F)    
as.factor(x)  2   28.0 14.0000  14.483 0.0003143 ***
Residuals    15   14.5  0.9667                      
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Source		SS	df		MS		F
Group	28	SSG	k-1	2	MSG = SSG /(k-1)	14	MSG/MSE
Residual	14.5	SSE	n-k	15	MSE = SSE /(n-k)	0.97	14.43
Total	42.5	SST	n-1	17			

ANOVA

```
Analysis of Variance Table

Response: Wing
            Df Sum Sq Mean Sq F value    Pr(>F)
Sex.1        1 2722.0 2721.98  643.15 < 2.2e-16 ***
Residuals 1693 7165.3    4.23
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

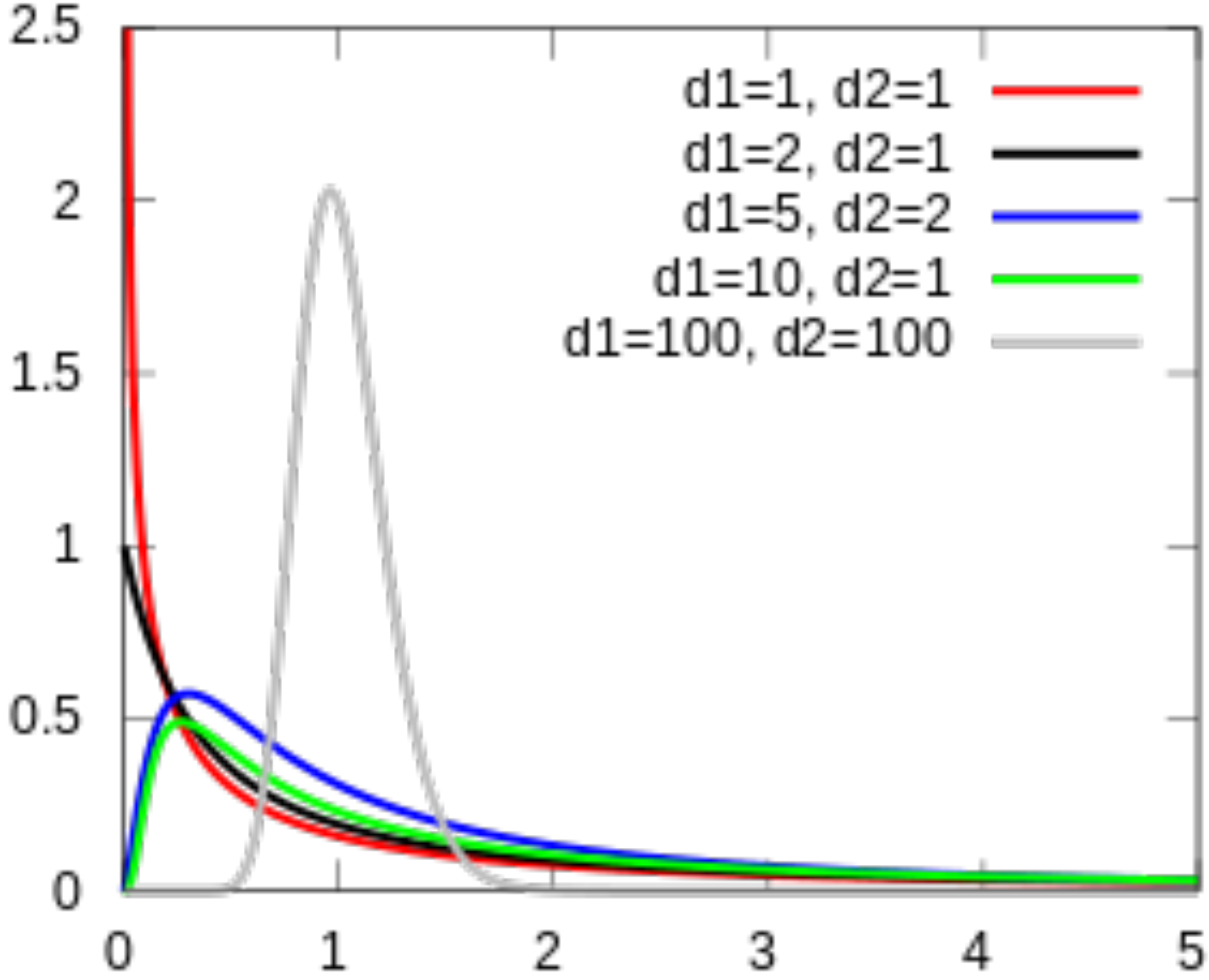
Source	SS	df	MS	F
Group	SSG	k-1	MSG = SSG /(k-1)	MSG/MSE
Residual	SSE	n-k	MSE = SSE /(n-k)	
Total	SST	n-1		

And what do you think do we do with the F value we calculated?

Fisher's F-distribution

Characterised by d_1 and d_2

Both are df's – k-1 and n-1!



Let’s examine the ingredients of an ANOVA:

- A continuous response variable (wing)
- A factorial explanatory variable (sex) with two or more levels

Let’s examine the ingredients of a LM:

- A continuous response variable (mass)
- A continuous explanatory variable (tarsus)

Let’s examine the ingredients of an ANOVA:

- A continuous response variable (wing)
- A factorial explanatory variable (sex) with two or more levels

Let’s examine the ingredients of a LM:

- A continuous response variable (mass)
- A continuous explanatory variable (tarsus)

- But, but, but we also ran a lm with sex, didn’t we? That was a factor!

Learning aims:

- ANOVA: to test for differences in many groups
- LM can use not only continuous, but also factors as explanatory variable!

LM vs ANOVA: do it now!

- We can also run a linear model with a factorial response.

- Work through HO 13.

DO it NOW!

- What is a *post-hoc* test?
- What is the problem with *post-hoc* tests? When are they useful?
- Make a table of how many repeated measures (1, 2, 3,…) there are of wing length within individuals (hint: tapply, dplyr or something like that is good for this! Google!)
- Make a table of how make repeated measures (1, 2, 3,…) there are of body mass within individuals.
- Run a linear model on a subset of data, explaining mass with year as factor this time excluding the year 2000. Examine both, the anova and summary output, and come to a conclusion about have a good think about biological relevance.
- What is the problem running a linear model with BirdID as factor?