



浙江大学爱丁堡大学联合学院 ZJU-UoE Institute

ANOVA

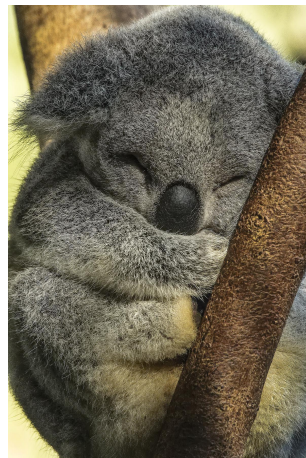
ADS 2, Lecture 14

Melanie Stefan - melanie.stefan@ed.ac.uk

Semester 1, 2019/20

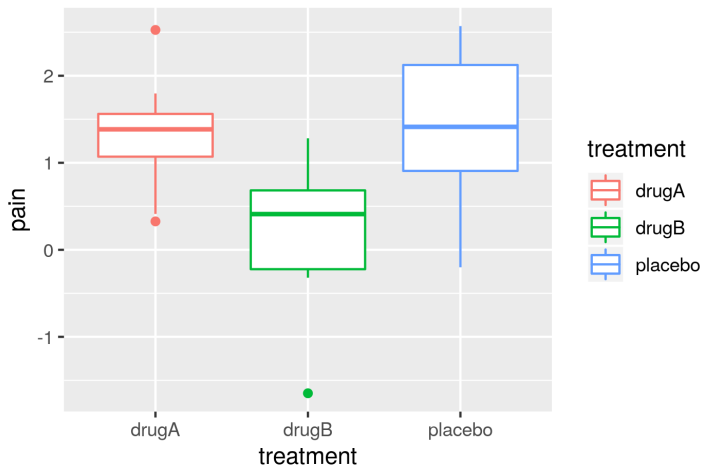
This lecture contains a lot of questions that I will ask you to think about in class. Providing the answers beforehand would defeat that purpose. Therefore, the version of the slides available to you before the lecture will not contain all of the information that is presented in the lecture.

A complete version will be uploaded to Learn after the lecture. In the meantime, here is a picture of an adorable baby koala.

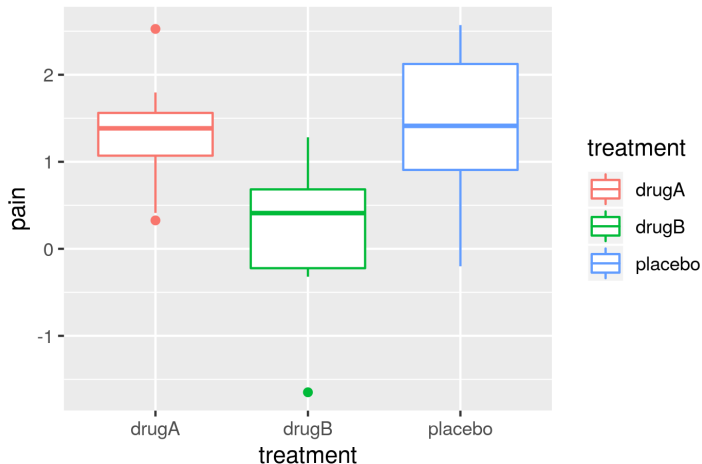


By Sheba_Also 43,000 photos [CC BY-SA 2.0 (<https://creativecommons.org/licenses/by-sa/2.0>)], via Wikimedia Commons

What did we do last week?



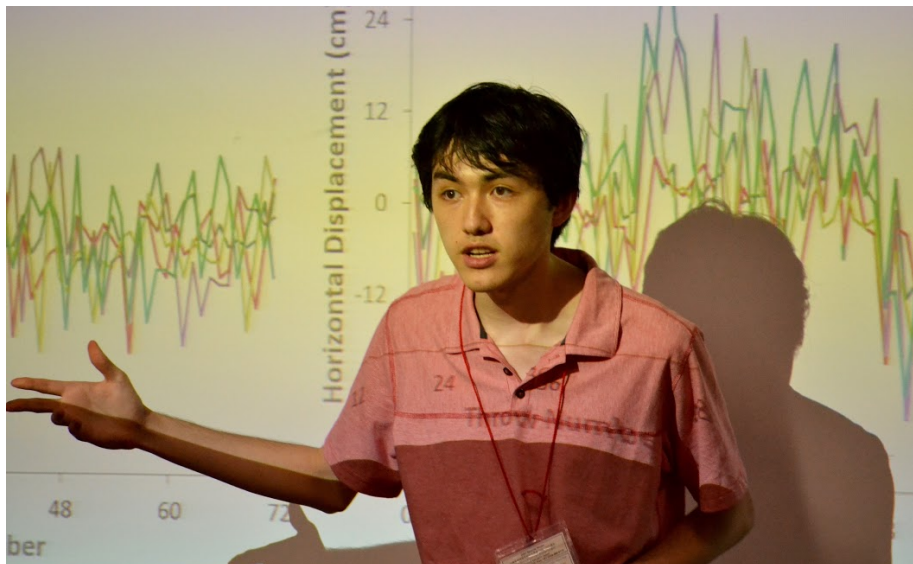
What did we do last week?



Is there an exact test to compare three or more groups?

This lecture is about ...

... ANalysis Of VAriance



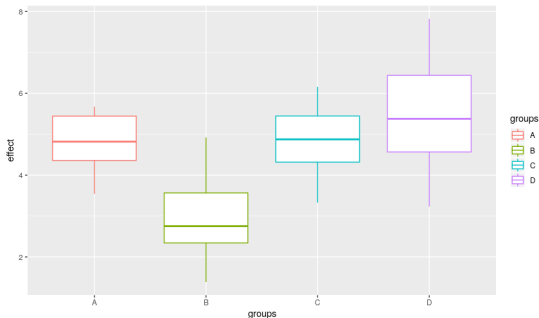
After this week, you will be able to . . .

- Explain the idea behind Analysis of Variance (ANOVA)
- State and test the assumptions of an ANOVA
- Use R to perform an ANOVA and appropriate post-hoc tests
- Interpret the results of an ANOVA

Outline

- 1 Quick refresher
- 2 Introduction to ANOVA
- 3 Types of ANOVA
- 4 Working through an example

Why can't we just do multiple t-tests?



What was the key idea instead?

What was the key idea instead?

Outline

- 1 Quick refresher
- 2 Introduction to ANOVA
- 3 Types of ANOVA
- 4 Working through an example

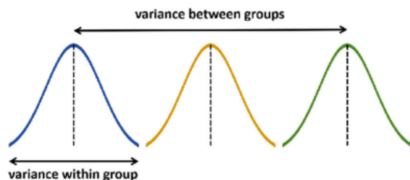
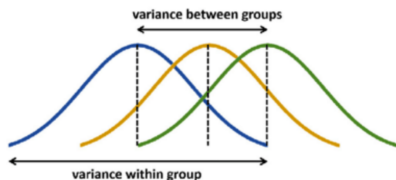
What is the Null Hypothesis of an ANOVA?

Principle behind ANOVA

(See also: Why is it called ANOVA?)

ANalysis Of VAriance

Within group variance vs Between group variance



Test Statistic

$$F = \frac{\text{between-group variability}}{\text{within-group variability}}$$

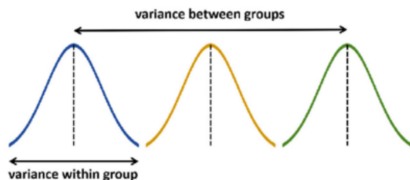
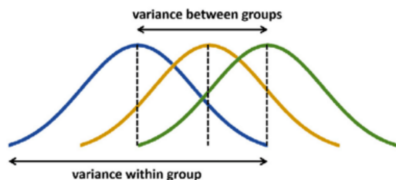
$$\text{between-group variability} = \frac{\sum_{i=1}^K n_i (\bar{Y}_i - \bar{Y})^2}{\text{degree of freedom } (K - 1)}$$

$$\text{within-group variability} = \frac{\sum_{i=1}^K \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{\text{every group, } n-1, \text{ sum up, } N-K \quad (N - K)}$$

K	...	number of groups
n_i	...	number of samples in group i
N	...	overall sample size
\bar{Y}_i	...	mean of group i
\bar{Y}	...	overall mean
Y_{ij}	...	j^{th} observation in group i

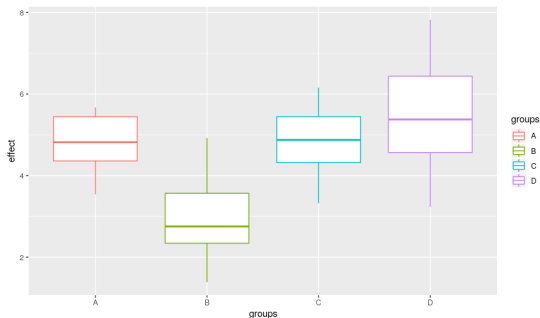
When might ANOVA fail?

Within group variance vs Between group variance



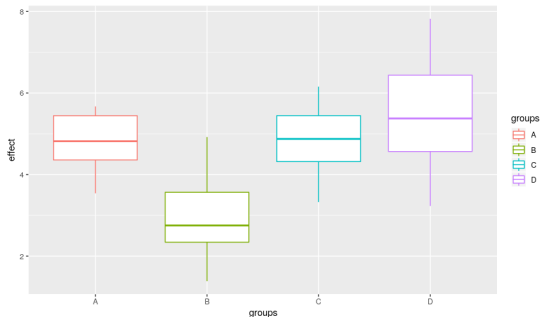
Assumptions for ANOVA

OK, so we run an ANOVA. And then what?



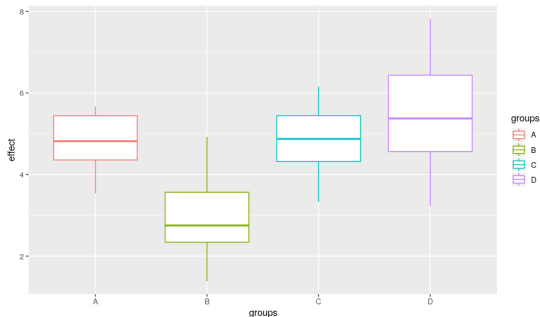
Let's say we do find that the groups are not the same. What do we do with this information?

Post-hoc tests



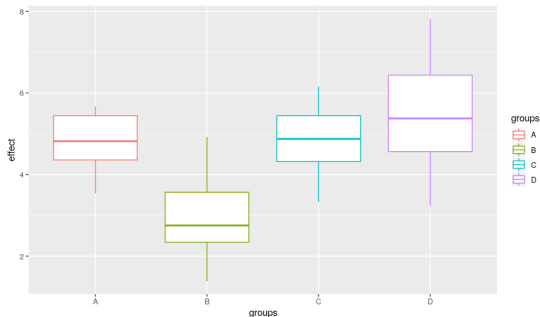
- We *do* want to know what groups exactly are different from each other.

Post-hoc tests



- We *do* want to know what groups exactly are different from each other.
- Idea: Do “something like” t-tests to compare pairs of groups.

Post-hoc tests



- We *do* want to know what groups exactly are different from each other.
- Idea: Do “something like” t-tests to compare pairs of groups.
- Why not exactly t-tests?

- Solution: Tukey's HSD test
- Honestly Significant Difference
- This runs multiple comparisons, with the appropriate corrections of p values to account for multiple testing

- Solution: Tukey's HSD test
- Honestly Significant Difference
- This runs multiple comparisons, with the appropriate corrections of p values to account for multiple testing
- Watch the spelling!



John Tukey

Outline

- 1 Quick refresher
- 2 Introduction to ANOVA
- 3 Types of ANOVA**
- 4 Working through an example

When do you use **ANOVA**?

Examples of situations in which you would use ANOVA

1. You want to compare the effect of three drugs on muscle growth.
Measured outcome (dependent variable) → muscle growth
1 factor (independent variable): drug type
3 groups (or levels) : drug A, B and C
2. You want to compare blood glucose levels in children, adult or elderly control or diabetic patients.
Measured outcome (dependent variable) → glycaemia
2 factors (independent variables):
 - (i) age – 3 levels: children, adult, old
 - (ii) diabetic status – 2 levels: yes/no

Types of ANOVA

What you will frequently use:

1-way ANOVA → 1 factor (e.g. effect of 3 doses of a drug on heart rate)

2-way ANOVA → 2 factors (e.g. effect of age and sex on salary)

3-way ANOVA → 3 factors (e.g. effect of age, sex and education on salary)
[less commonly used]

Also (not covered in this lecture)

Repeated measure ANOVA → when measuring the same subject multiple times, e.g. for a time-course

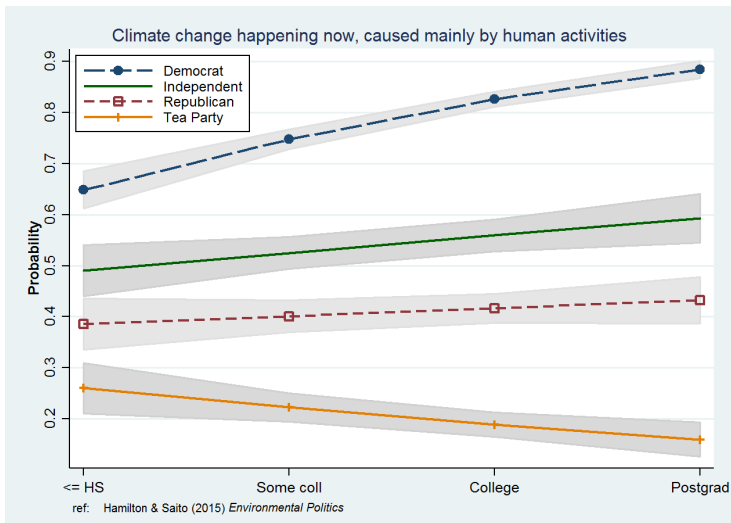
MANOVA (multivariate ANOVA) → when measuring more than one outcome, e.g. measure height and weight of patients treated with a drug vs control.

Interactions

If more than one factor is included, then the response to one factor may be affected by the other factor(s). This is called an **interaction**.

Interactions

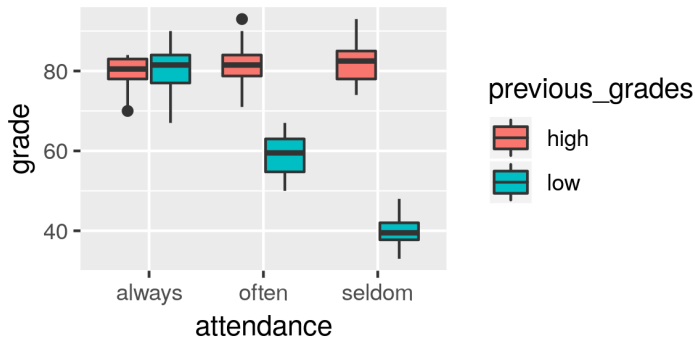
If more than one factor is included, then the response to one factor may be affected by the other factor(s). This is called an **interaction**.



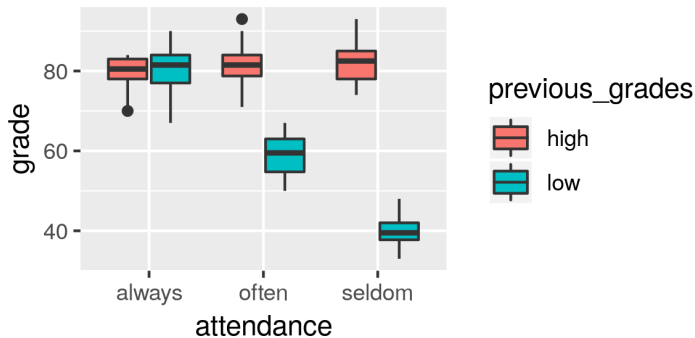
Outline

- 1 Quick refresher
- 2 Introduction to ANOVA
- 3 Types of ANOVA
- 4 Working through an example

Example: Effect of attendance and previous grades on course performance



Example: Effect of attendance and previous grades on course performance



What is H_0 ? What is H_A ?

Formulate H_0 , H_A

Select statistical test

Select statistical test

Independent random sampling

Independent random sampling

Assumptions of ANOVA

- Independent random sampling ✓
- Normality of residuals
- Equality of Variances

Normality of residuals

There are 3 ways to test for that!

First, make aov model:

```
model <- aov(grade ~ attendance * previous_grades ,  
             data=class)
```

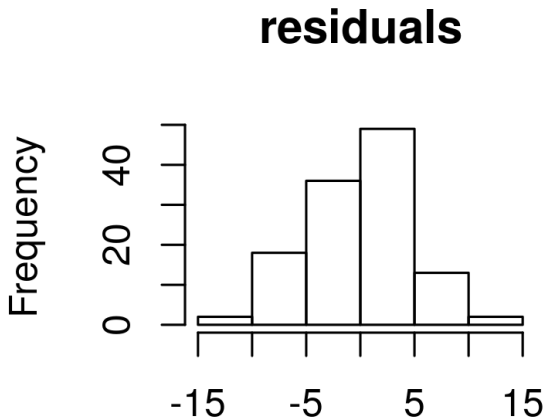
* is for model with interaction;

+ for model without interaction

Normality of residuals - Method 1:

Plot histogram of residuals and use visual inspection (i.e. “eyeball” it)

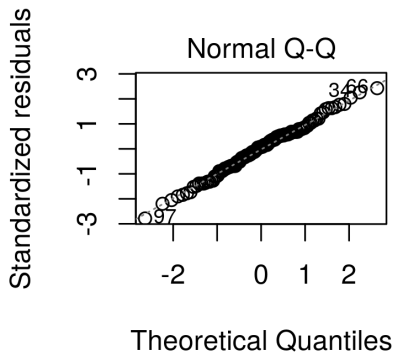
```
hist(resid(model), main="residuals")
```



Normality of residuals - Method 2:

Use one of the analytic plots provided by R when using aov and eyeball it. The plot is the second one shown and is called Normal Q-Q. Dots should be aligned along the diagonal.

```
plot(model, 2)
```



Normality of residuals - Method 3:

Use a formal test for normality, e.g. the Shapiro-Wilk test

```
> shapiro.test(resid(model))
```

Shapiro-Wilk normality test

data: resid(model)

W = 0.99355, p-value = 0.8574

What is H_0 for that test?

Normality of residuals - Now what?

Normality of residuals - Now what?

Q: Which of the three methods should I use?

Normality of residuals - Now what?

Q: Which of the three methods should I use?

A: It does not matter, as long as you state and explain your choice

Normality of residuals - Now what?

Q: Which of the three methods should I use?

A: It does not matter, as long as you state and explain your choice

Q: Can I use all three?

No, it increase false-positive.

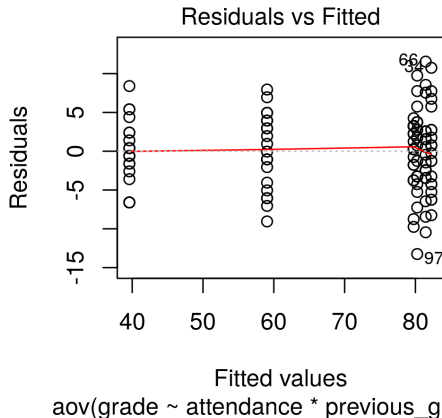
Assumptions of ANOVA

- Independent random sampling ✓
- Normality of residuals ✓
- Equality of Variances

Equality of Variances

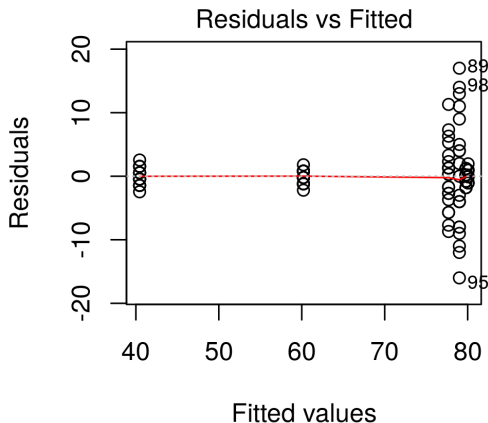
Use “Residuals vs Fitted” plot. Looking for similar heights of “columns”

`plot(model, 1)`



Equality of Variances

Here is a counterexample of what it would look like with different Variances:



Assumptions of ANOVA

- Independent random sampling ✓
- Normality of residuals ✓
- Equality of Variances ✓

OK, now finally . . .

We can do the actual ANOVA.

OK, now finally ...

We can do the actual ANOVA.

```
summary(model)
```

```
> summary(model)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
attendance	2	7278	3639	152.9	<2e-16	***
previous_grades	1	13889	13889	583.4	<2e-16	***
attendance:previous_grades	2	9321	4661	195.8	<2e-16	***
Residuals	114	2714	24			

```
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Are we done?

Now, you should be (more) able to ...

- Explain the idea behind Analysis of Variance (ANOVA)
- State and test the assumptions of an ANOVA
- Use R to perform an ANOVA and appropriate post-hoc tests
- Interpret the results of an ANOVA

What questions do you have?

Acknowledgments and Image Credits

This lecture uses materials from ABMS2 lectures from previous years by Paula Brunton and Nicola Romanò. Where not otherwise indicated, images are also from those lectures.

Image credits

- Belief in Climate change as a function of education and political party preference in the United States. By IChiloe - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=48240215>
- John Wilder Tukey. By Source, Fair use, <https://en.wikipedia.org/w/index.php?curid=17099473>
- R plots and screenshots for ANOVA example. My own work, 2019. CC-BY-SA 3.0.
- Student presenting data. By Yuuki Guzman and Agoston Tyll (Okinawa Institute of Science and Technology), 2015.