浙江大学爱丁堡大学联合学院

ZJU-UoE Institute

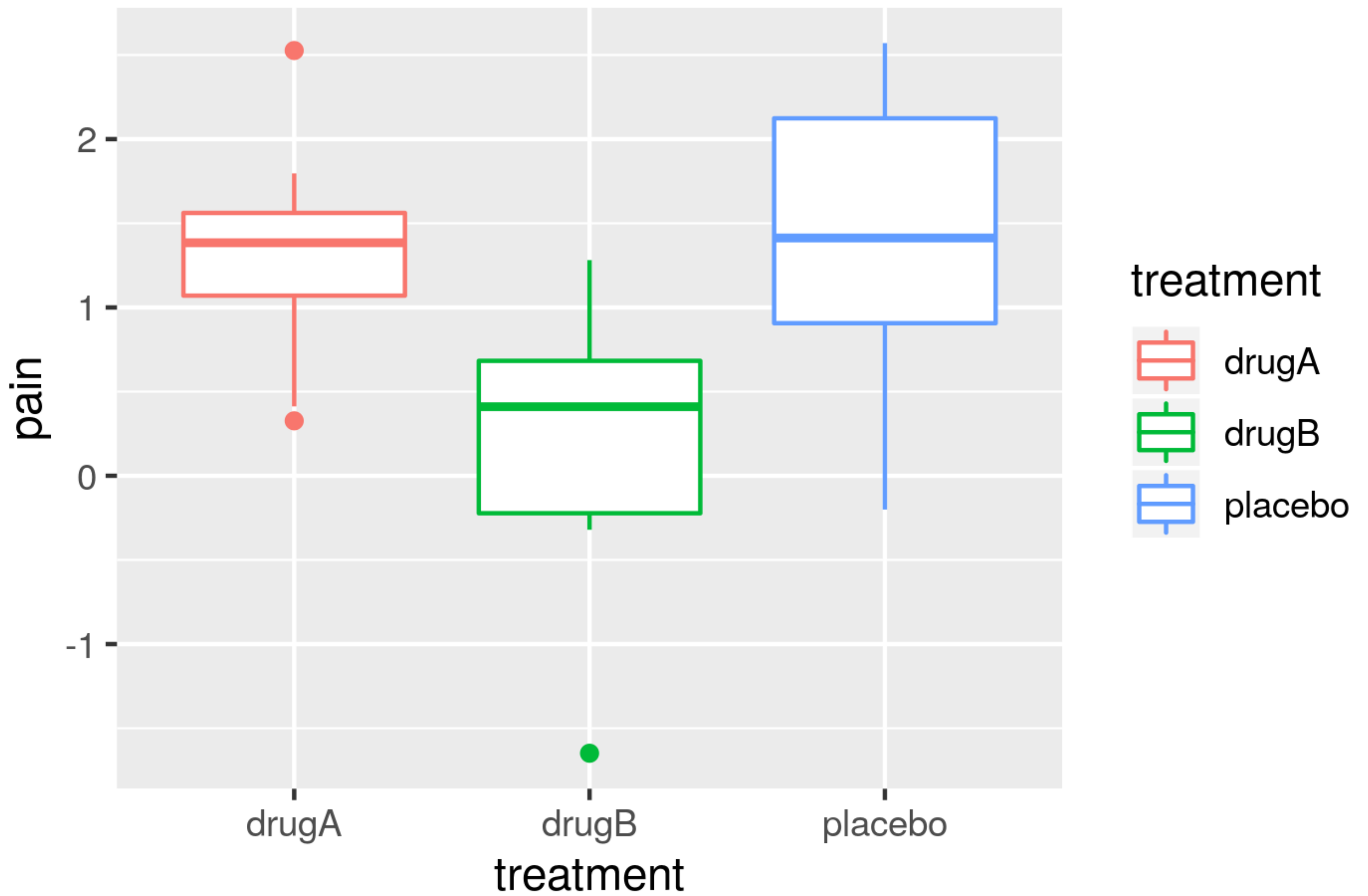
ADS2 Lecture 2.2
ANOVA

Dr Duncan MacGregor duncan.macgregor@ed.ac.uk

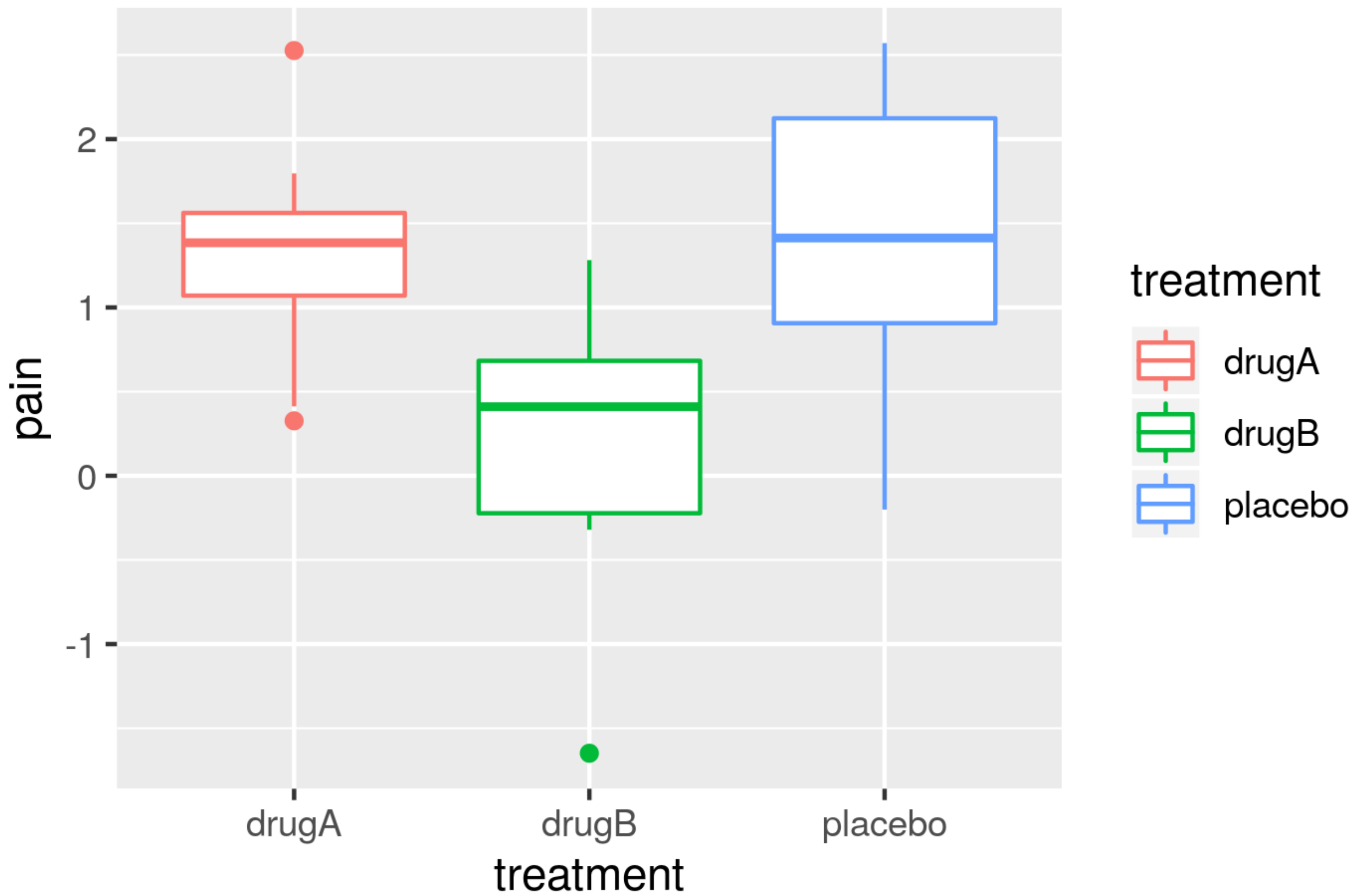
Semester 2, Week 2

2023-24

What did we do last week?



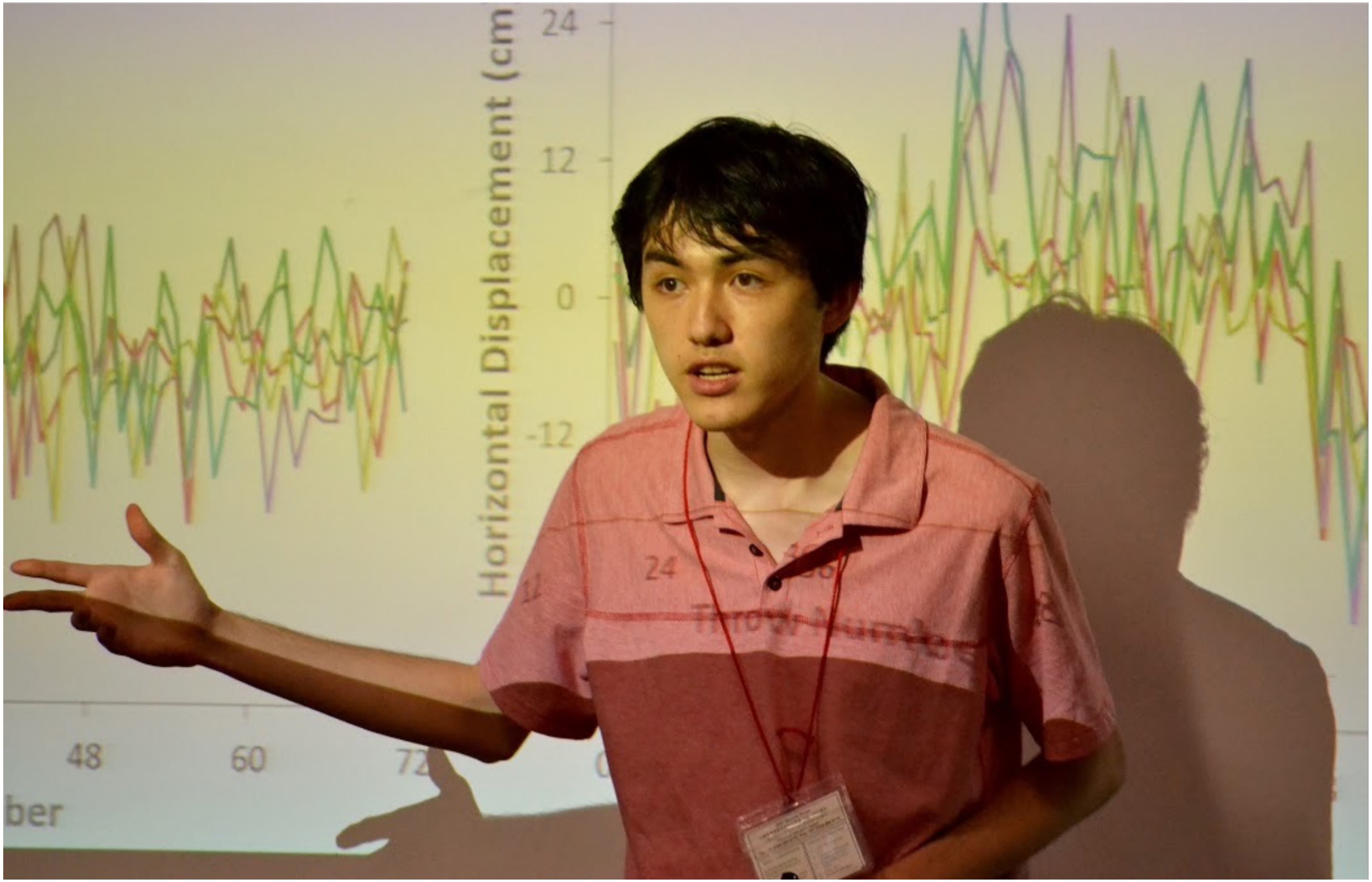
What did we do last week?



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

This lecture is about . . .

. . . ANalysis Of VAriance



Learning Objectives

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

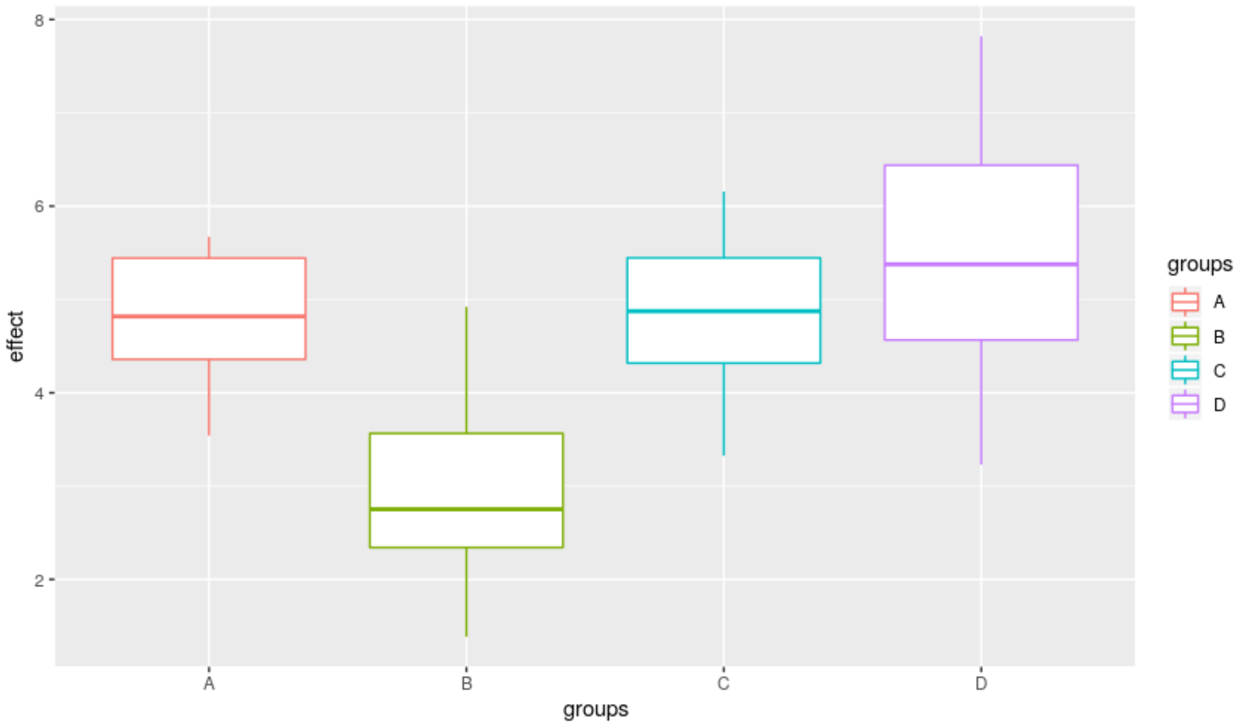
1. Quick Refresher

2. Introduction to ANOVA

3. Types of ANOVA

4. Working through an example

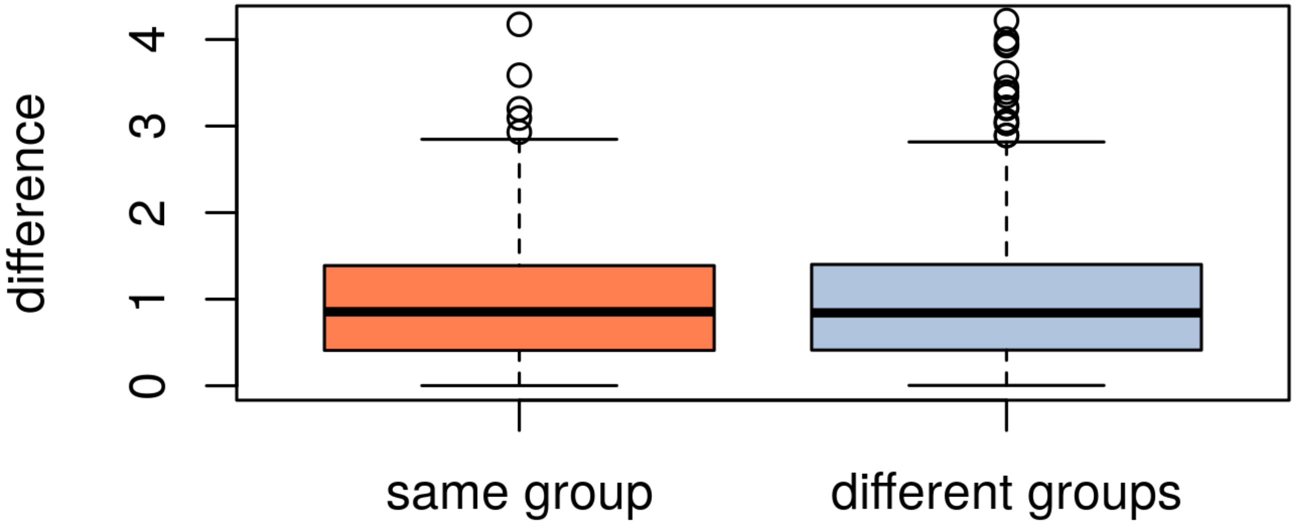
Why can’t we just do multiple t-tests?



What was the key idea instead?

What was the key idea instead?

Compare differences between and within groups



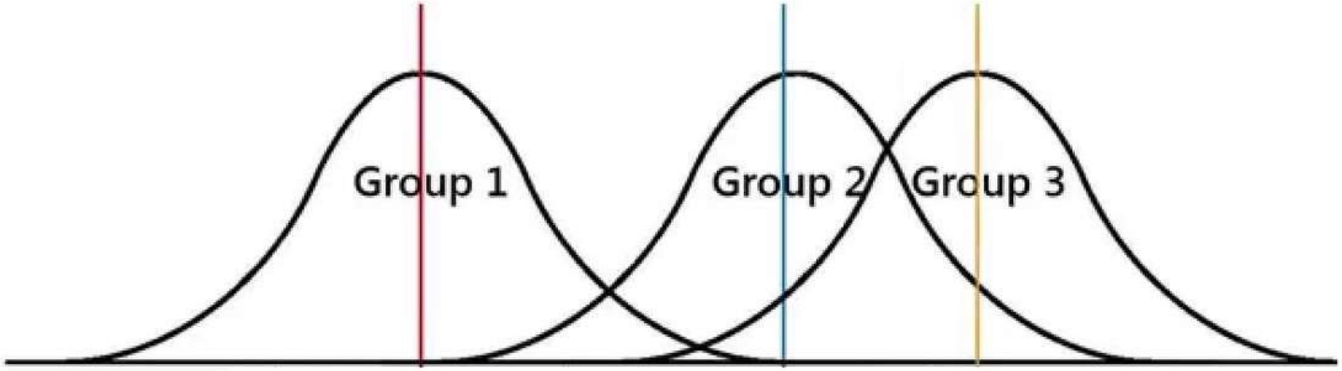
Outline

1. Quick Refresher
2. Introduction to ANOVA
3. Types of ANOVA
4. Working through an example

What is the Null Hypothesis of an ANOVA?

What is the Null Hypothesis of an ANOVA?

ANOVA



ANOVA tests the **null hypothesis** i.e.

"There is no difference between any of the groups"

or

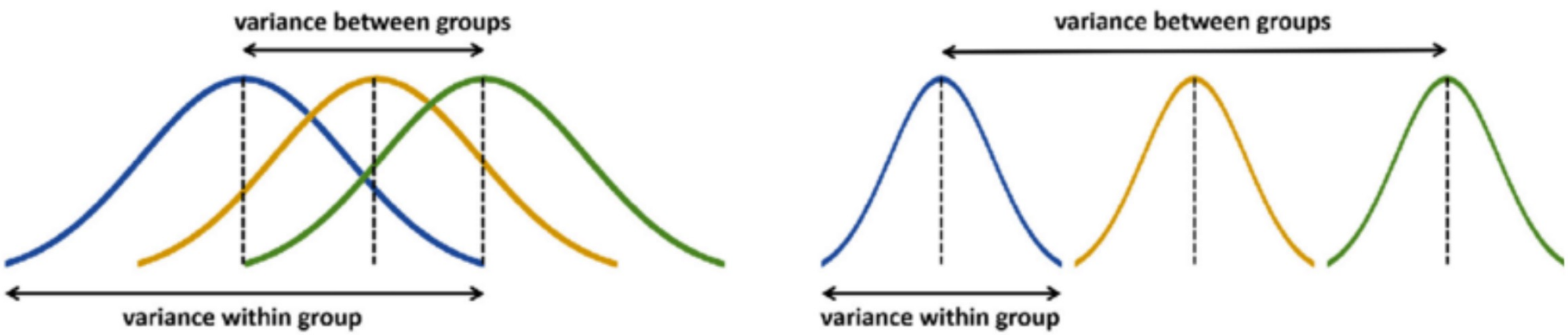
"The group does not influence the response"

Principle behind ANOVA

(See also: Why is it called ANOVA?)

ANalysis **O**f **VA**riance

Within group variance vs Between group variance



Test Statistic

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

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

$$\text{within–group variability} = \frac{\sum_{i=1}^{K} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_i\right)^2}{(N-K)}$$

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

Test Statistic

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

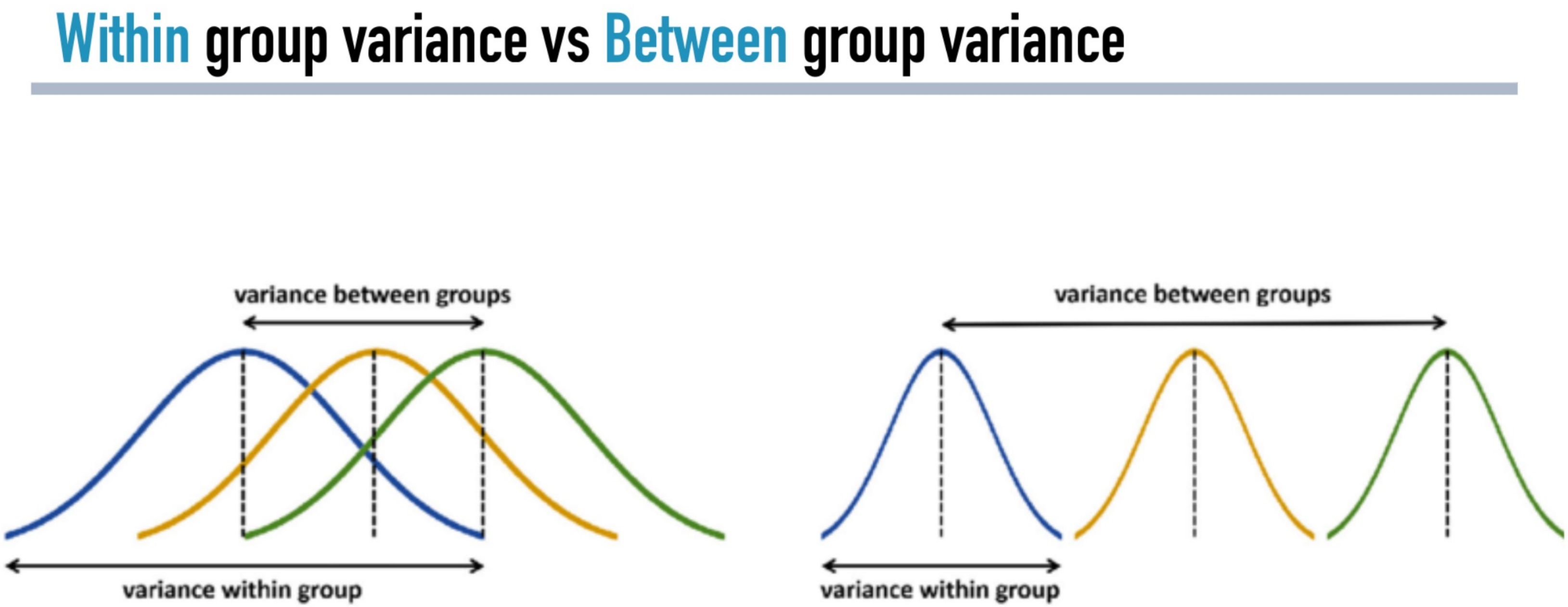
$$\text{between–group variability} = \frac{\sum_{i=1}^{K} n_i (\bar{Y}_i - \bar{Y})^2}{(K-1)}$$

$$\text{within–group variability} = \frac{\sum_{i=1}^{K} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_i\right)^2}{(N-K)}$$

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

Explain these equations in your own words!

When might ANOVA fail?

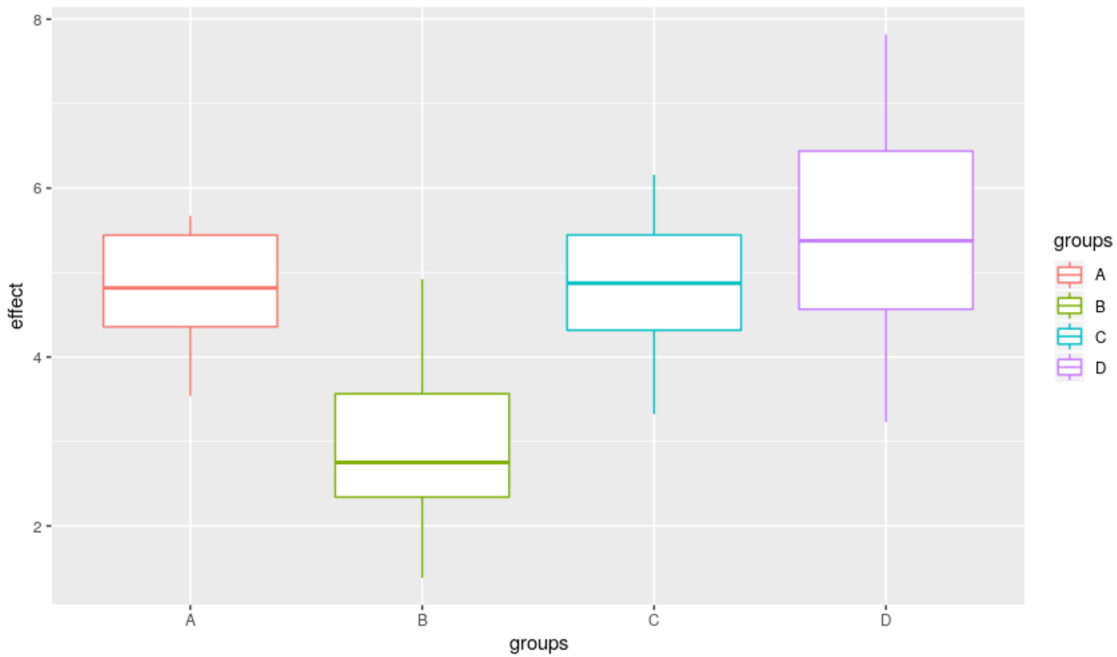


Assumptions for ANOVA

Assumptions for ANOVA

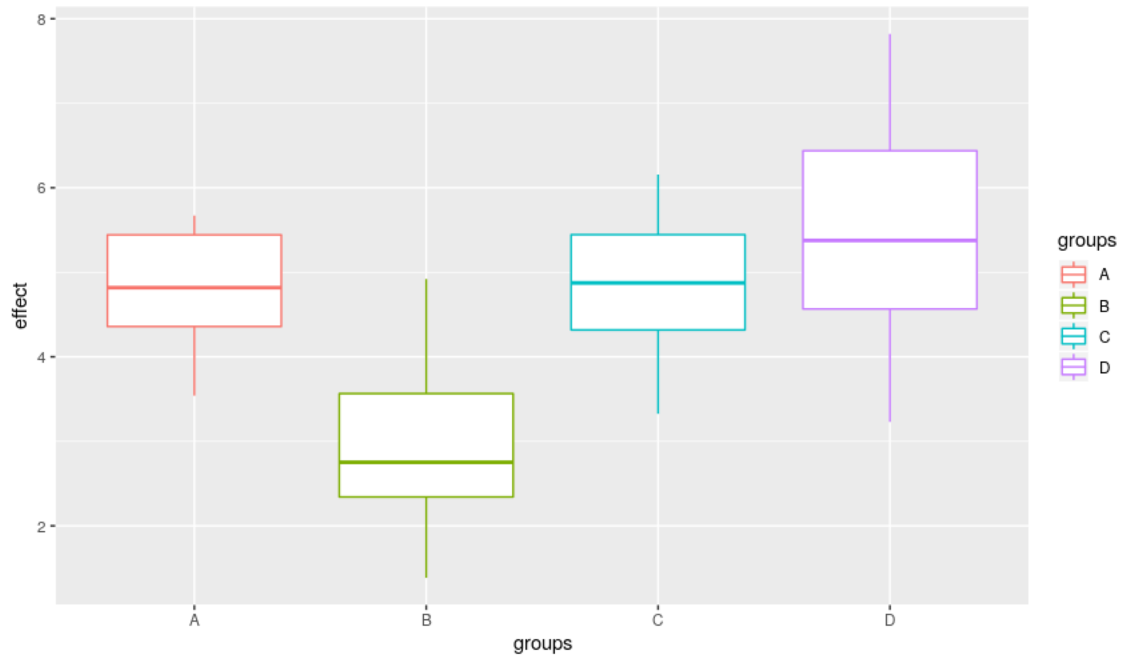
- Independent random sampling
- Normality of residuals (distances from group mean)
- Equality of Variances

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



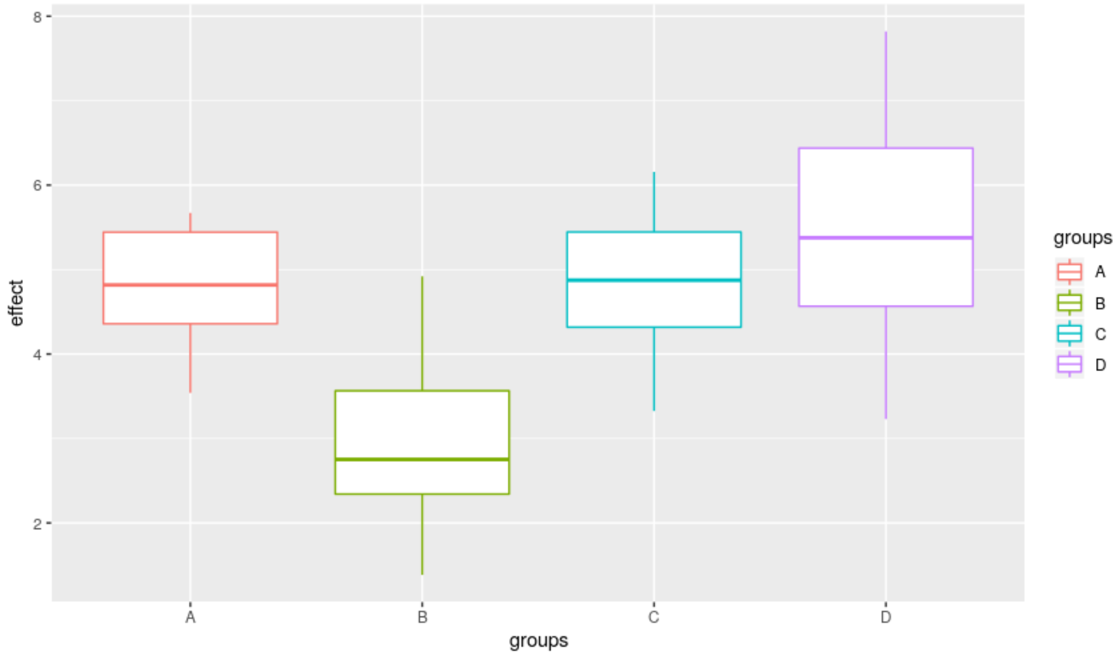
Let's say we do find that the groups are not the same. What do we do with this information?

Post-hoc tests



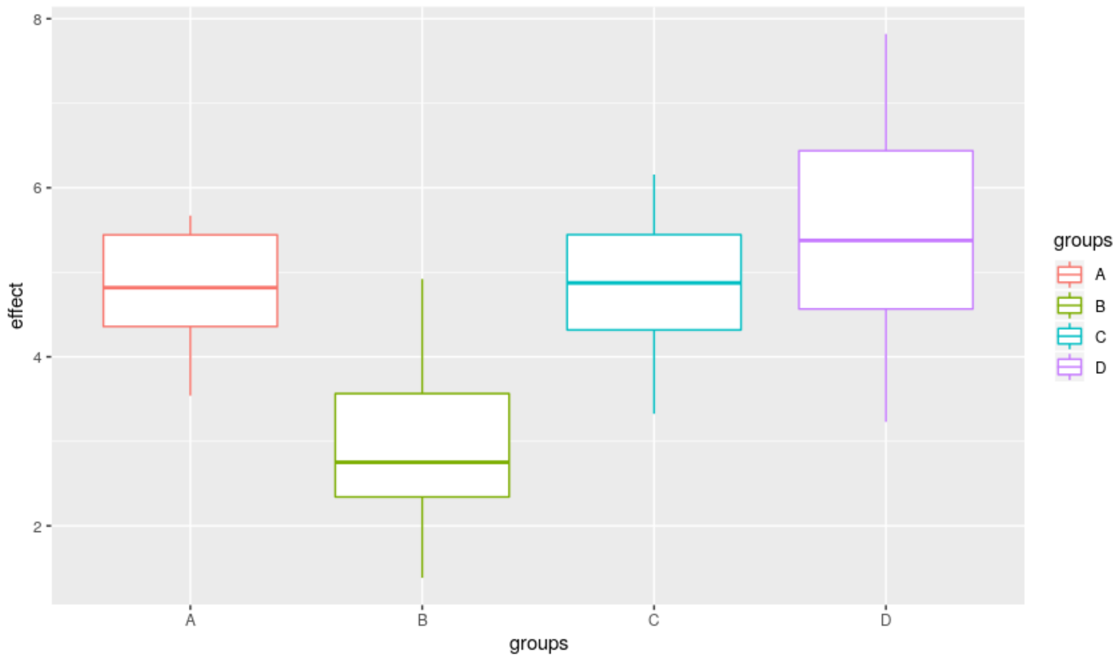
- We *do* want to know what groups exactly are different from each other.

Post-hoc tests



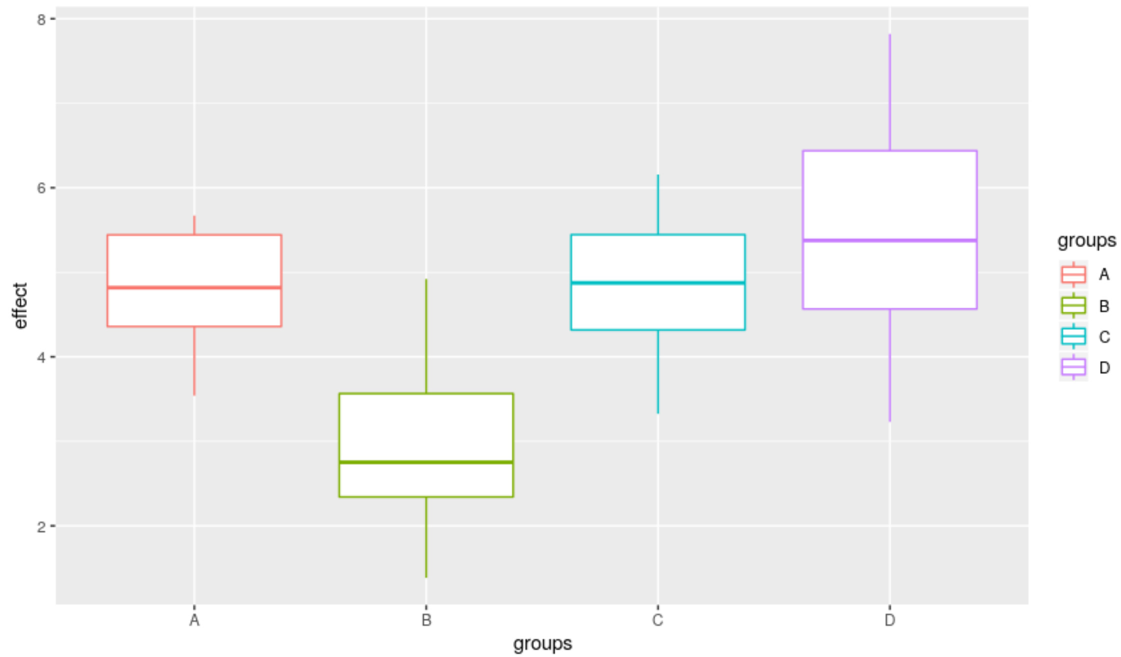
- We *do* want to know what groups exactly are different from each other.
- Idea: Do “something like" t-tests to compare pairs of groups.

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?

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?
- Because we need to correct for multiple testing to reduce the risk of false positives

Post-hoc 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

Post-hoc 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
- Watch the spelling!

John Tukey

Outline

1. Quick Refresher
2. Introduction to ANOVA
3. Types of ANOVA
4. Working through an example

When to use an ANOVA

More than 2 populations

- Effect of different drugs on recovery from injury

When to use an ANOVA

More than 2 populations

- Effect of different drugs on recovery from injury
- Feeding behaviour of different bird species
- Comparing healthcare in China, the UK and Austria
- . . .

When to use an ANOVA

More than 2 populations

- Effect of different drugs on recovery from injury
- Feeding behaviour of different bird species
- Comparing healthcare in China, the UK and Austria
- . . .

More than 1 predictive variable (factor)

- Effect of diet *and* exercise on health
- Effect of genetic background and drugs on stress levels
- Differences in height by gender *and* birth province
- . . .

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

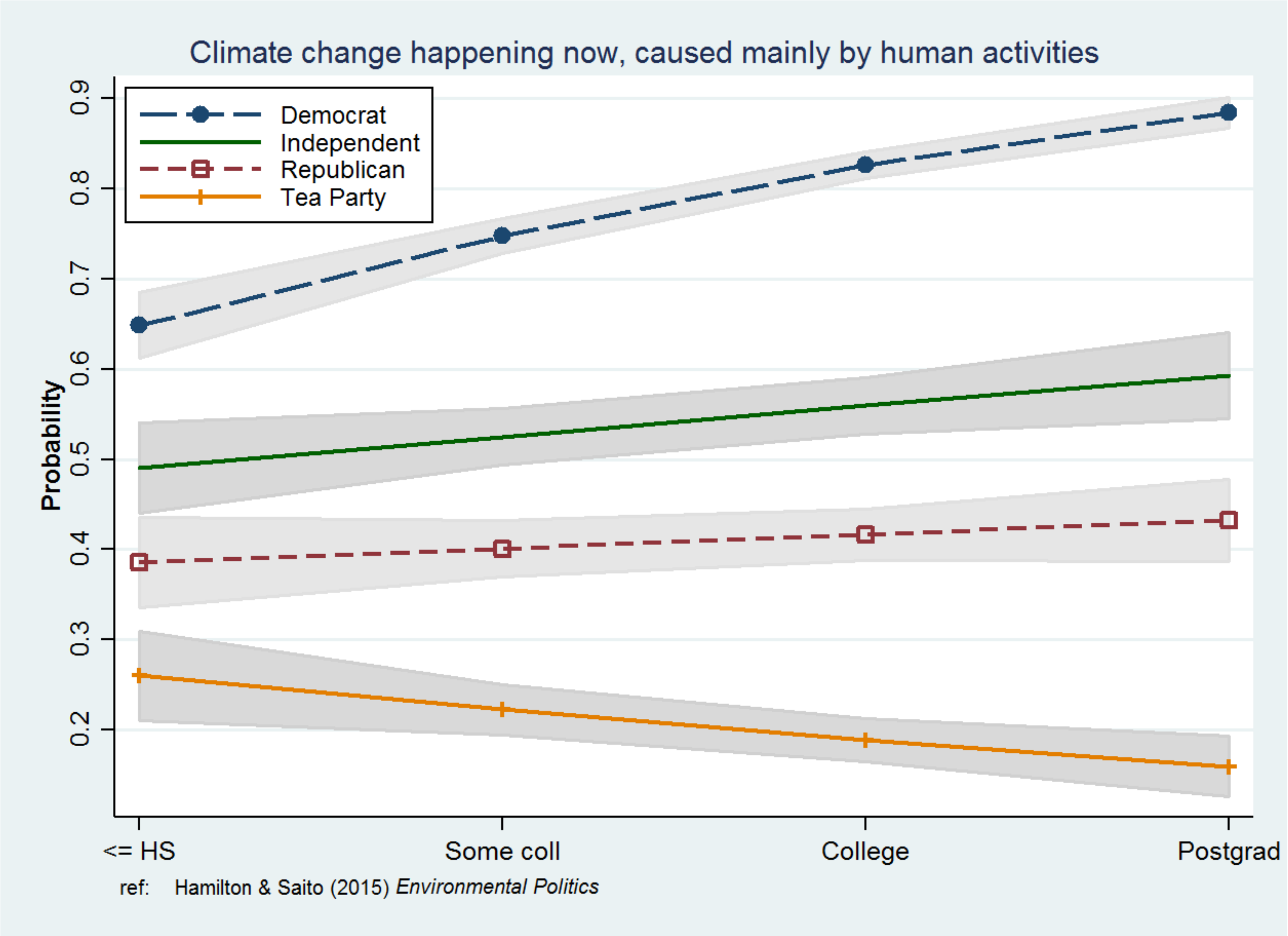
MANOVA (multivariate ANOVA) → when measuring more than one outcome, e.g. measure height and weight of patients treated with a drug vs control.

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

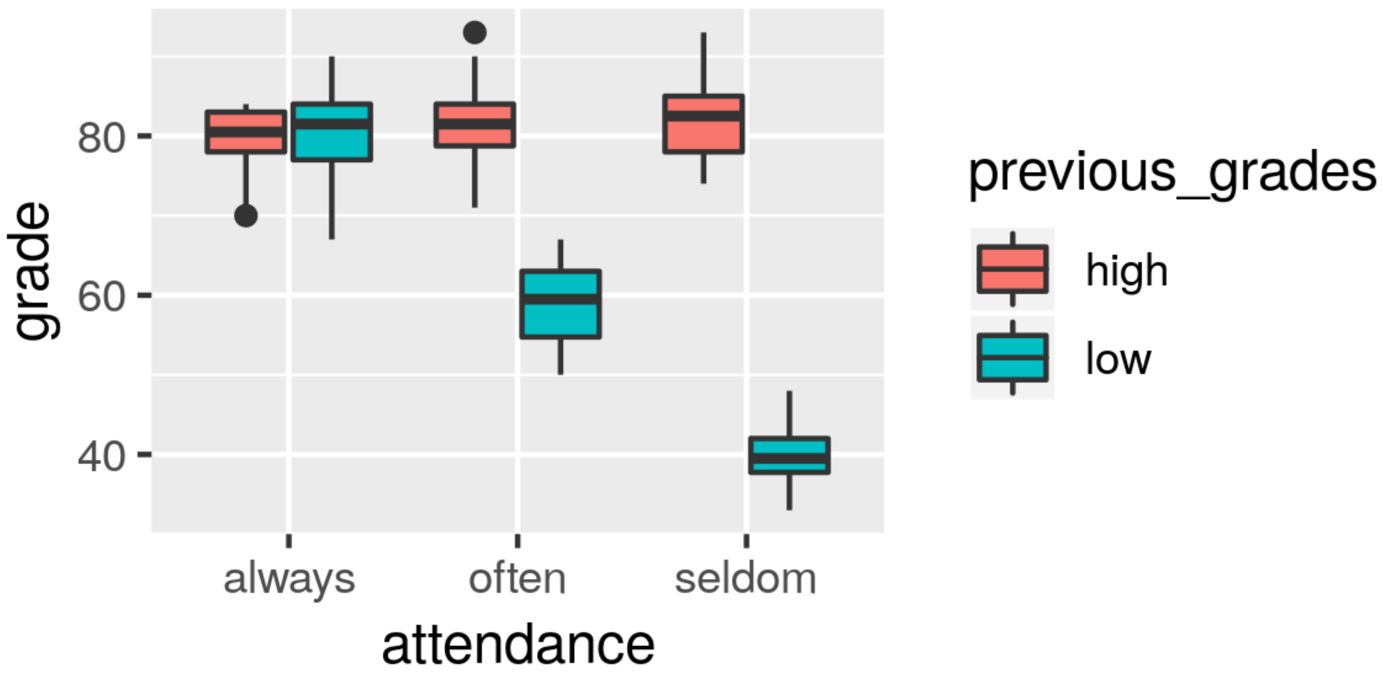
If more than one factor is included, then the response to one factor may be affected by the other factor(s). This is called an **interaction**.



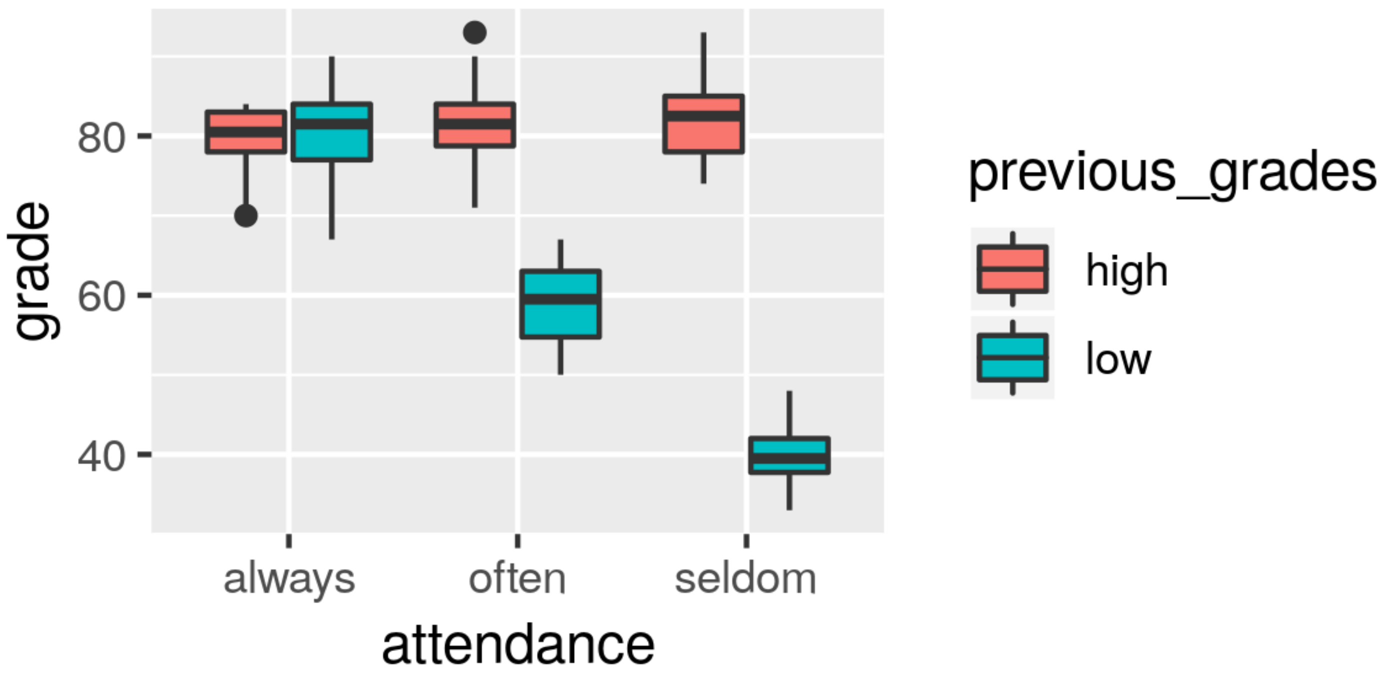
Outline

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

Hypotheses:

- H_0: There is no effect of class attendance or previous grades on course performance
- H_A: At least one of those factors (class attendance or previous grades) influences course performance.

Formulate H_0, H_A

Hypotheses:

- H_0: There is no effect of class attendance or previous grades on course performance
- H_A: At least one of those factors (class attendance or previous grades) influences course performance.

Additional Hypotheses if we test for interactions:

- H_0: There is no interaction between class attendance and previous grades
- H_A: There is an interaction between class attendance and previous grades

Select statistical test

Select statistical test

2-way ANOVA seems reasonable since we are looking at two factors (class attendance and previous grades)

Select statistical test

2-way ANOVA seems reasonable since we are looking at two factors (class attendance and previous grades) (and this is a lecture about ANOVA).

Select statistical test

2-way ANOVA seems reasonable since we are looking at two factors (class attendance and previous grades) (and this is a lecture about ANOVA).

BUT: We need to test assumptions first!

Assumptions of ANOVA

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

Independent random sampling

Independent random sampling

Cannot be assessed from data set itself. We have to believe that this is true given the description of the experiment itself.

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)
```

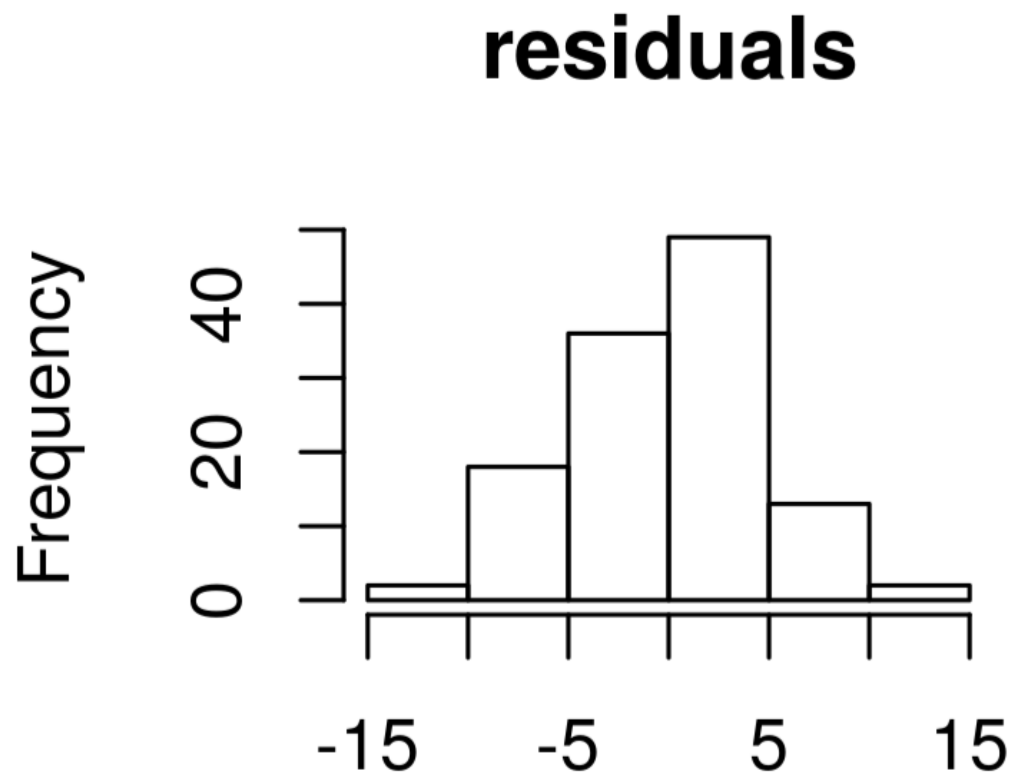
* is for model with interaction;

+ is for model without interaction

Normality of residuals – Method 1:

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

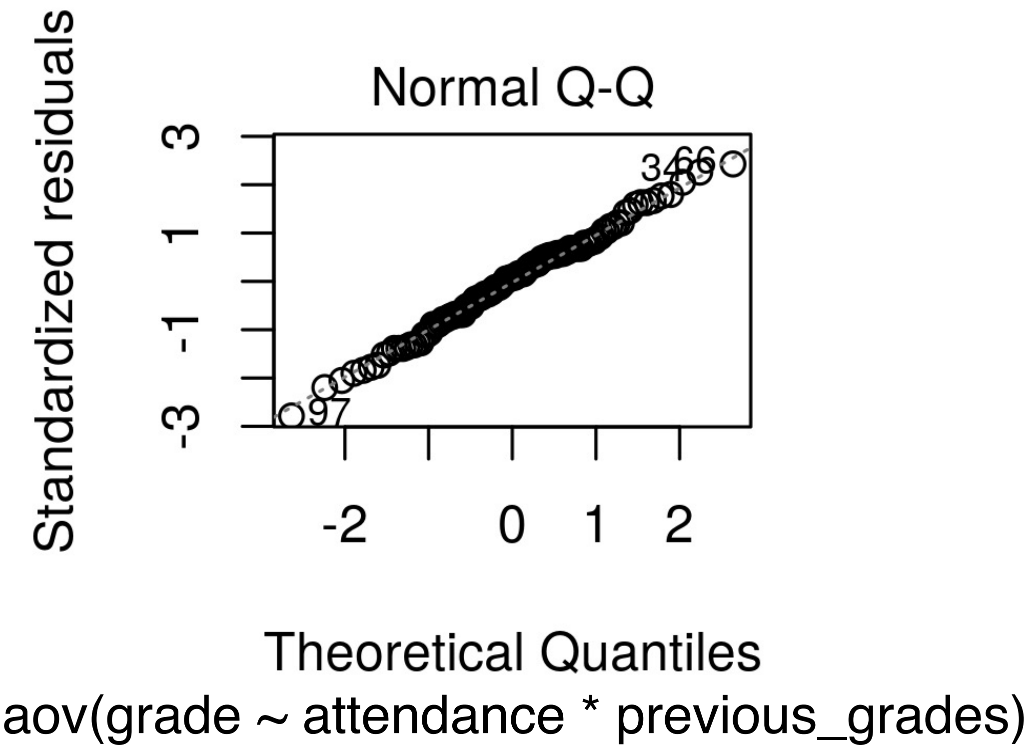
```
hist(resid(model), main = ”residuals”)
```



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?

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?

A: **No, no you may not!**

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?

A: **No, no you may not!** (How will you decide?)

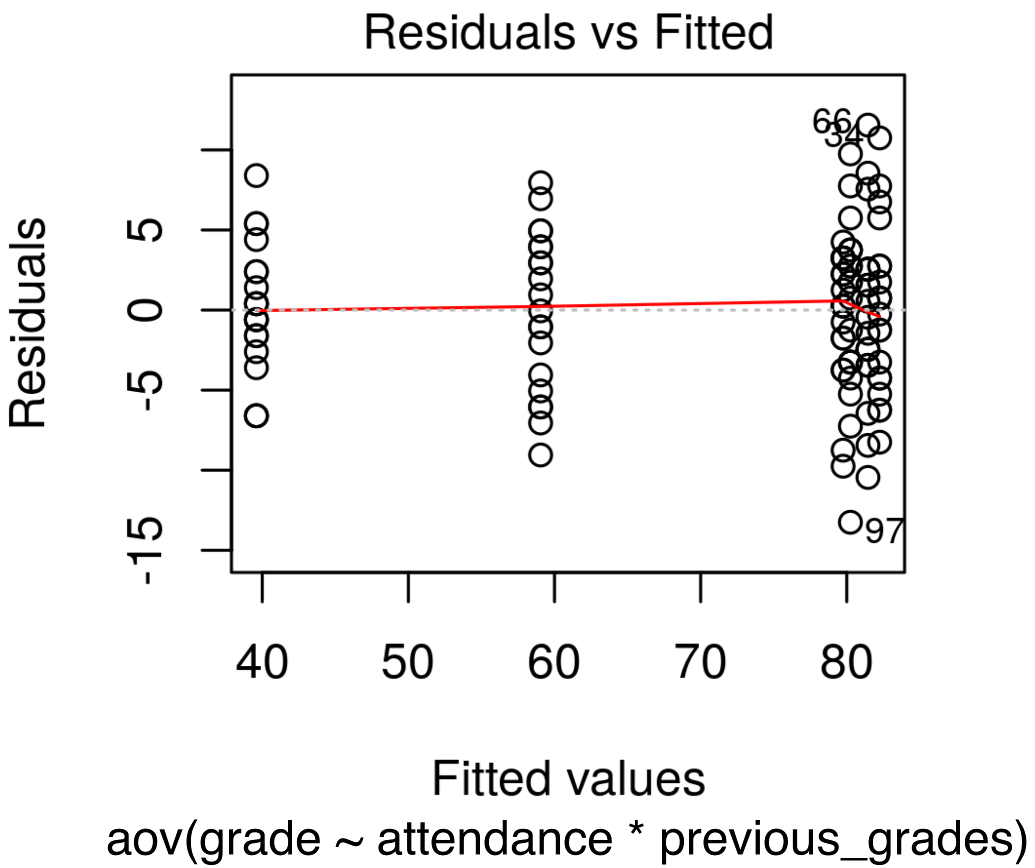
Assumptions of ANOVA

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

Equality of Variances

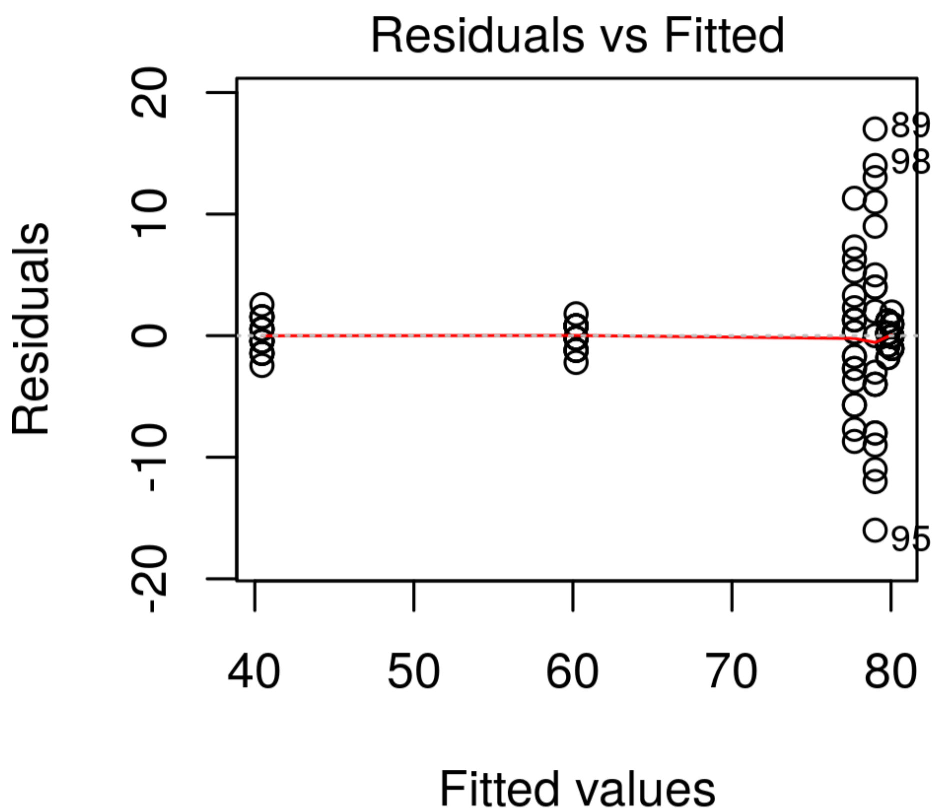
Use "Residuals vs Fitted" plot. Looking for similar heights of "columns"

```
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.

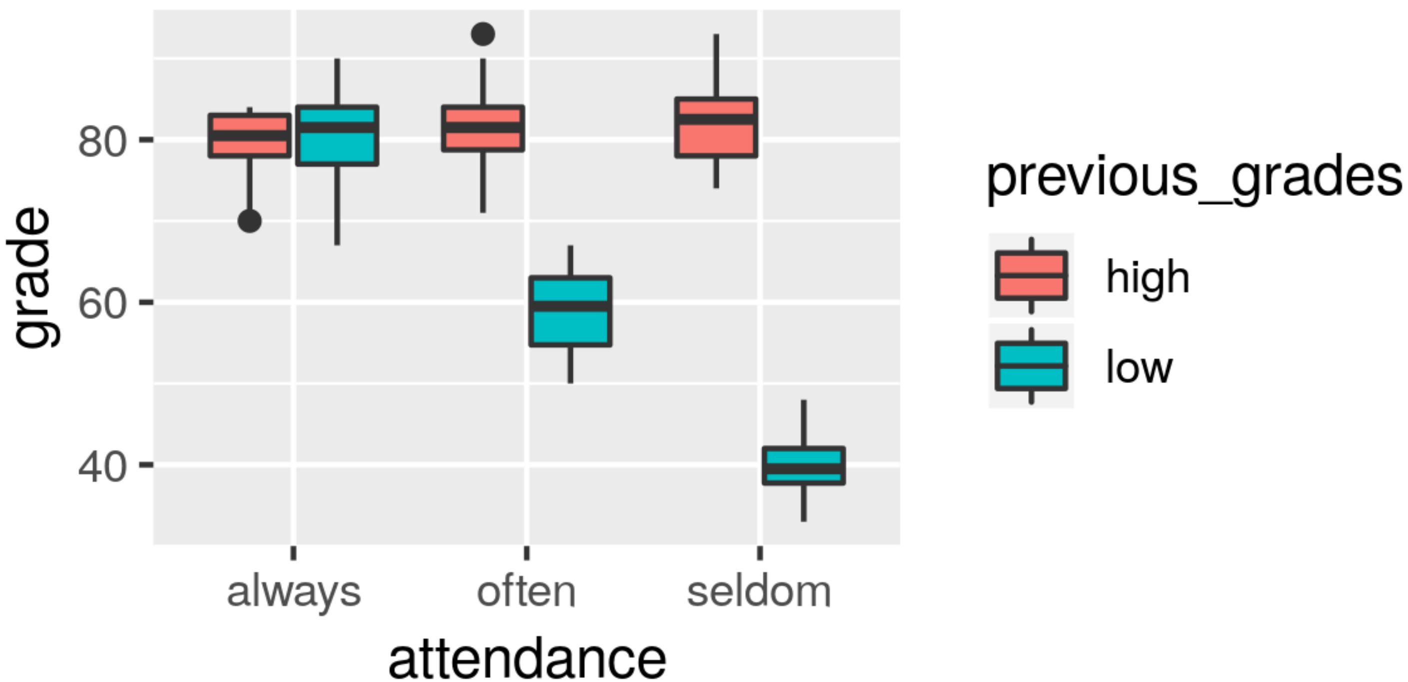
```
summary(model)
```

```
> 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?

Are we done?

NO! We need to do a post-hoc test to learn more details! What groups exactly are different?



Post-hoc test

`TukeyHSD(`**`model`**`)`

	diff	lwr	upr	p adj
often:high-always:high	1.70	-2.772633	6.172633	0.8797129
seldom:high-always:high	2.50	-1.972633	6.972633	0.5869676
always:low-always:high	0.50	-3.972633	4.972633	0.9995135
often:low-always:high	-20.70	-25.172633	-16.227367	0.0000000
seldom:low-always:high	-40.15	-44.622633	-35.677367	0.0000000
seldom:high-often:high	0.80	-3.672633	5.272633	0.9953540
always:low-often:high	-1.20	-5.672633	3.272633	0.9707621
often:low-often:high	-22.40	-26.872633	-17.927367	0.0000000

What do you conclude?

Review

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?

Acknowledgements and Image Credits

This lecture uses materials from ABMS2 lectures from previous years by Paula Brunton and Nicola Romano, and ADS2 lectures by Melanie Stefan. 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.