浙江大学爱丁堡大学联合学院

ZJU-UoE Institute

Comparing multiple means

ADS 2, Lecture 13

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

Semester 1, 2019/20

So, you know how to do a t-test

So, you know how to do a t-test

But what if you want to compare more than two means?

Learning Objectives

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

- Design and interpret a simulation-based hypothesis test
- Use a simulation-based test to compare more than two means
- Discuss limitations of t-tests
- Discuss problems around multiple testing

What if I want to compare more than two groups?

Can you name situations where this problem would arise?

What if I want to compare more than two groups?

Can you name situations where this problem would arise?

More than 2 populations

- Effect of different drugs on recovery from injury

What if I want to compare more than two groups?

Can you name situations where this problem would arise?

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
- ...

What if I want to compare more than two groups?

Can you name situations where this problem would arise?

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

- Effect of diet *and* exercise on health

What if I want to compare more than two groups?

Can you name situations where this problem would arise?

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

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

OK, we can't just run a t-test here

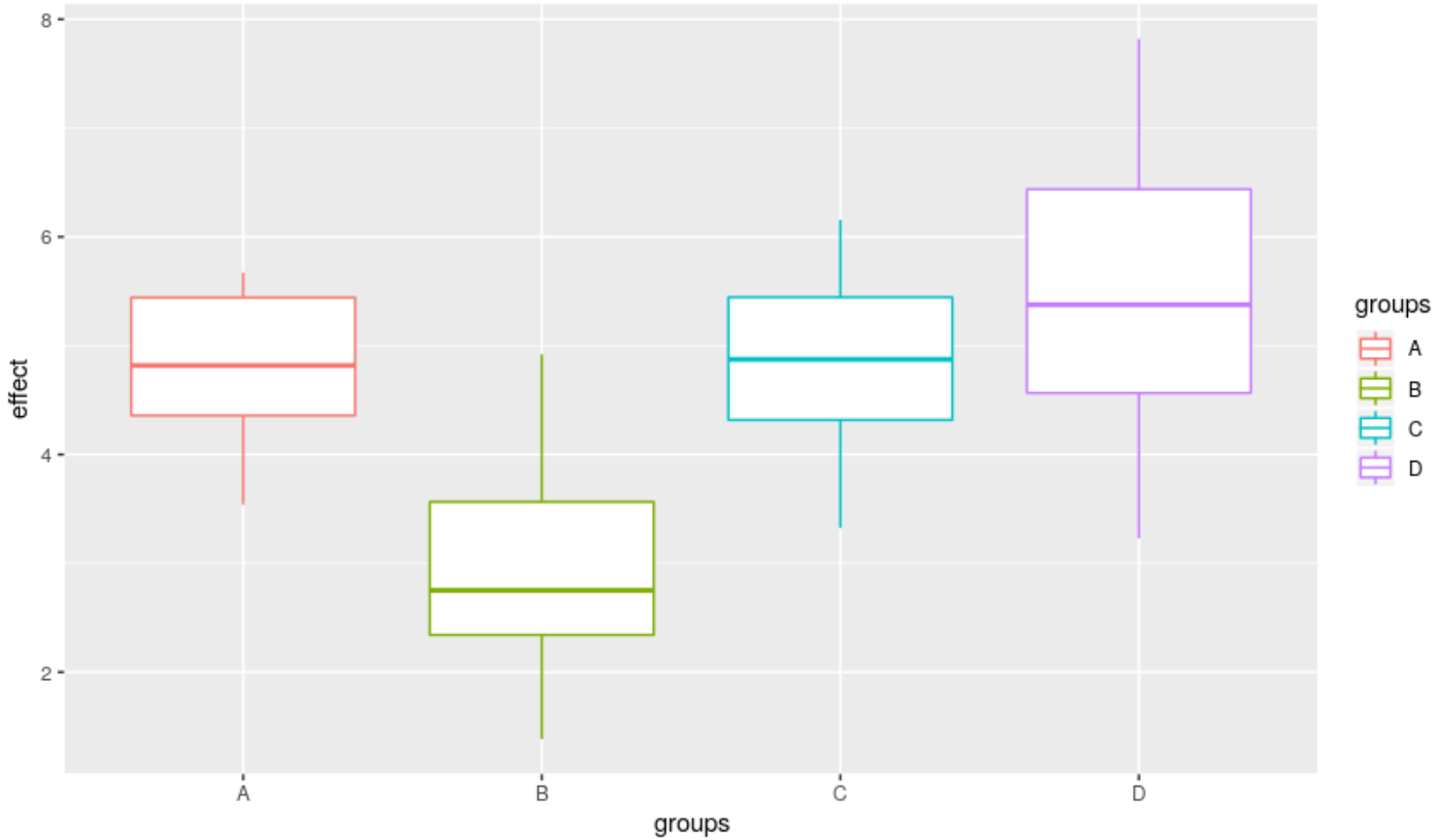
Why?

OK, we can't just run a t-test here

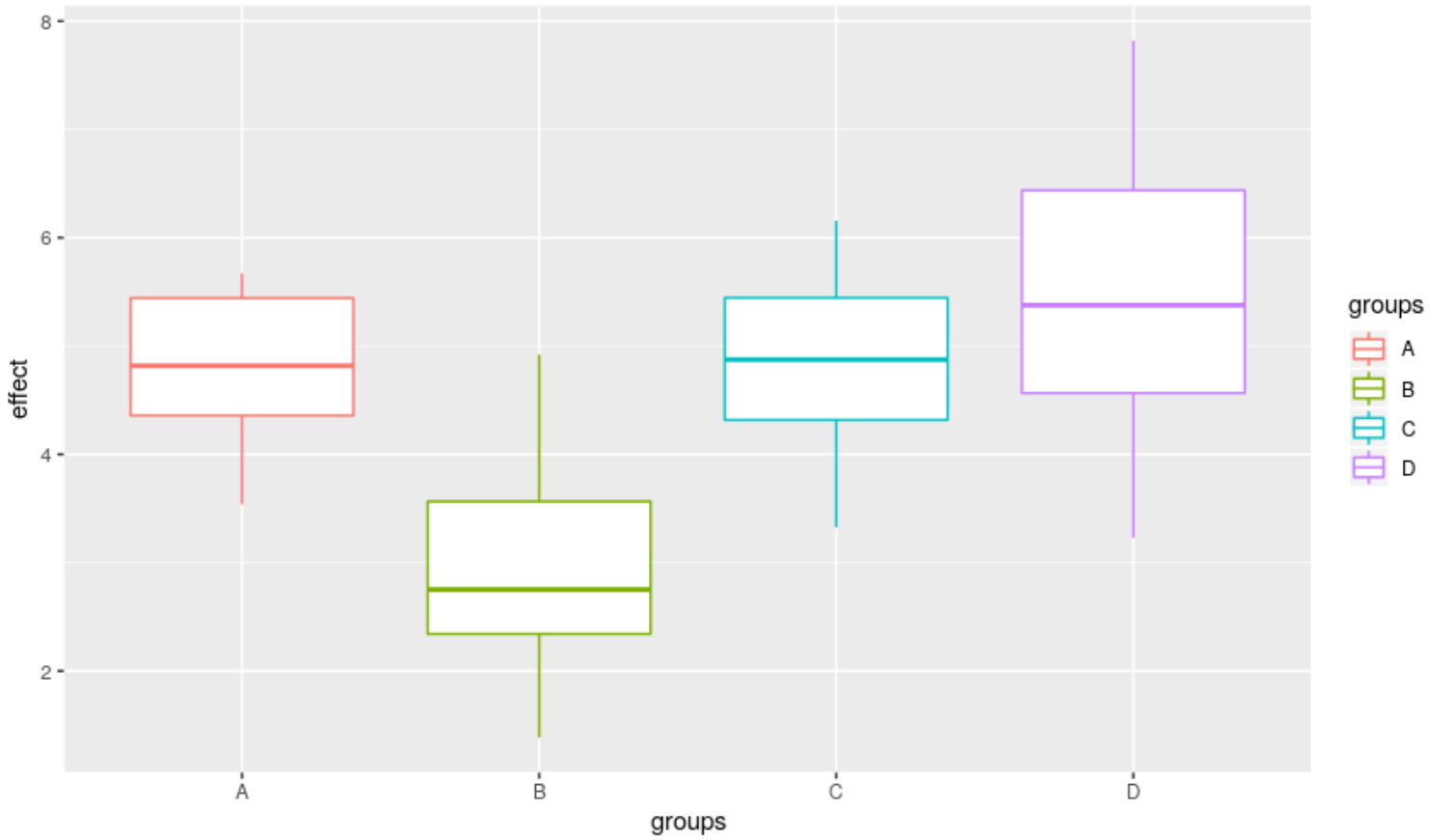
Why?

OK. But maybe we can run several t-tests?

Example: Comparing four groups



Example: Comparing four groups



How many t-tests would you need to run?

Example: Comparing four groups

Comparing:

A to B
A to C
A to D
B to C
B to D
C to D

t-test review

What is the probability of getting a false-positive result if there really is no difference?

If you are not sure, think about what happens when we do a t-test. What does your p-value mean? How do you use it to decide?

t-test review

- When computing the p-value we ask:
  If H0 is true, what is the probability of seeing a result as or more extreme as the one we saw in our experiment?

t-test review

- When computing the p-value we ask:
  If H0 is true, what is the probability of seeing a result as or more extreme as the one we saw in our experiment?
- We compare the p-value to the "significance level" α to decide on whether or not to reject H0

t-test review

- When computing the p-value we ask:
  If H0 is true, what is the probability of seeing a result as or more extreme as the one we saw in our experiment?
- We compare the p-value to the "significance level" α to decide on whether or not to reject H0
- If we set α to 0.05, this means that we are happy with p-values lower than 0.05

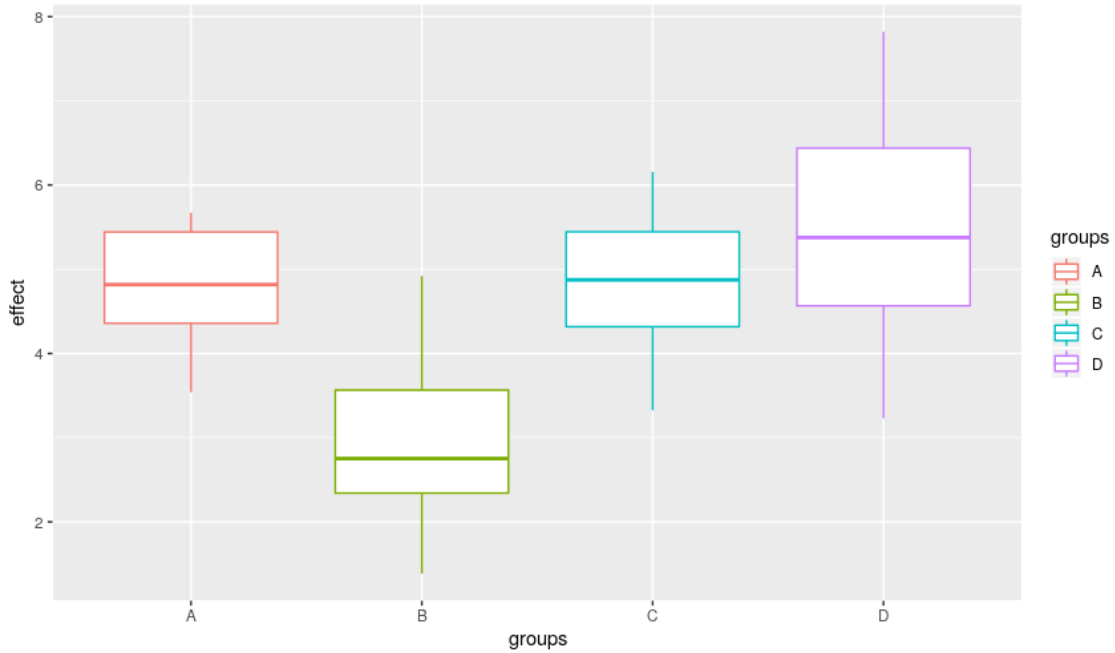
t-test review

- When computing the p-value we ask:
  If H0 is true, what is the probability of seeing a result as or more extreme as the one we saw in our experiment?
- We compare the p-value to the "significance level" α to decide on whether or not to reject H0
- If we set α to 0.05, this means that we are happy with p-values lower than 0.05
- But every so often, even if H0 is true, we will see a result as or more extreme as the one we saw. *How often exactly?*

t-test review

- When computing the p-value we ask:
  If H0 is true, what is the probability of seeing a result as or more extreme as the one we saw in our experiment?
- We compare the p-value to the "significance level" α to decide on whether or not to reject H0
- If we set α to 0.05, this means that we are happy with p-values lower than 0.05
- But every so often, even if H0 is true, we will see a result as or more extreme as the one we saw. *How often exactly?*
- By setting α, we accept a certain risk of seeing a false positive (given H0 is true). This is exactly what α is.

Assume there is really no difference

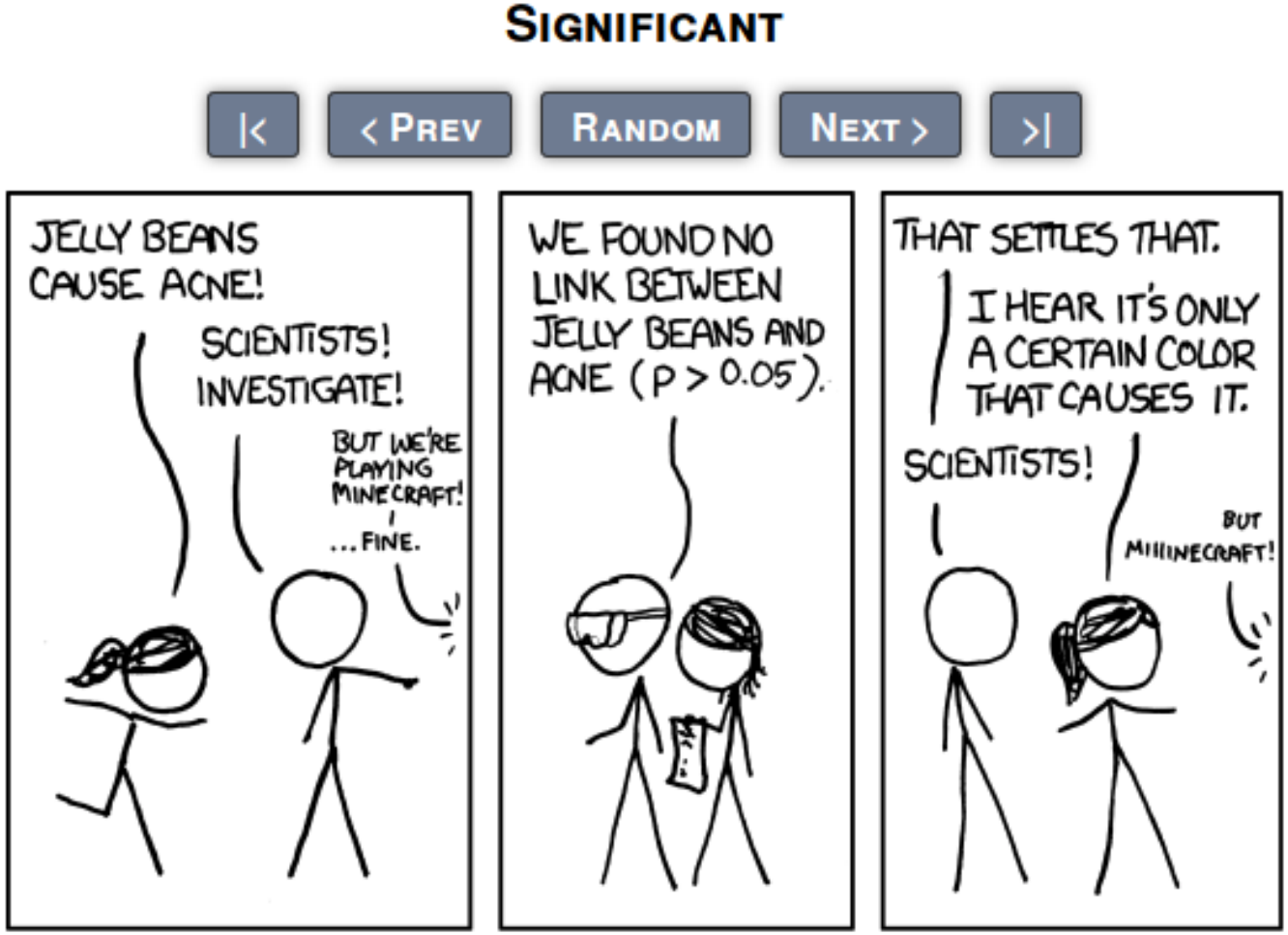


In 6 t-tests with $\alpha = 0.05$ for each, what is the probability of getting at least one false positive result?

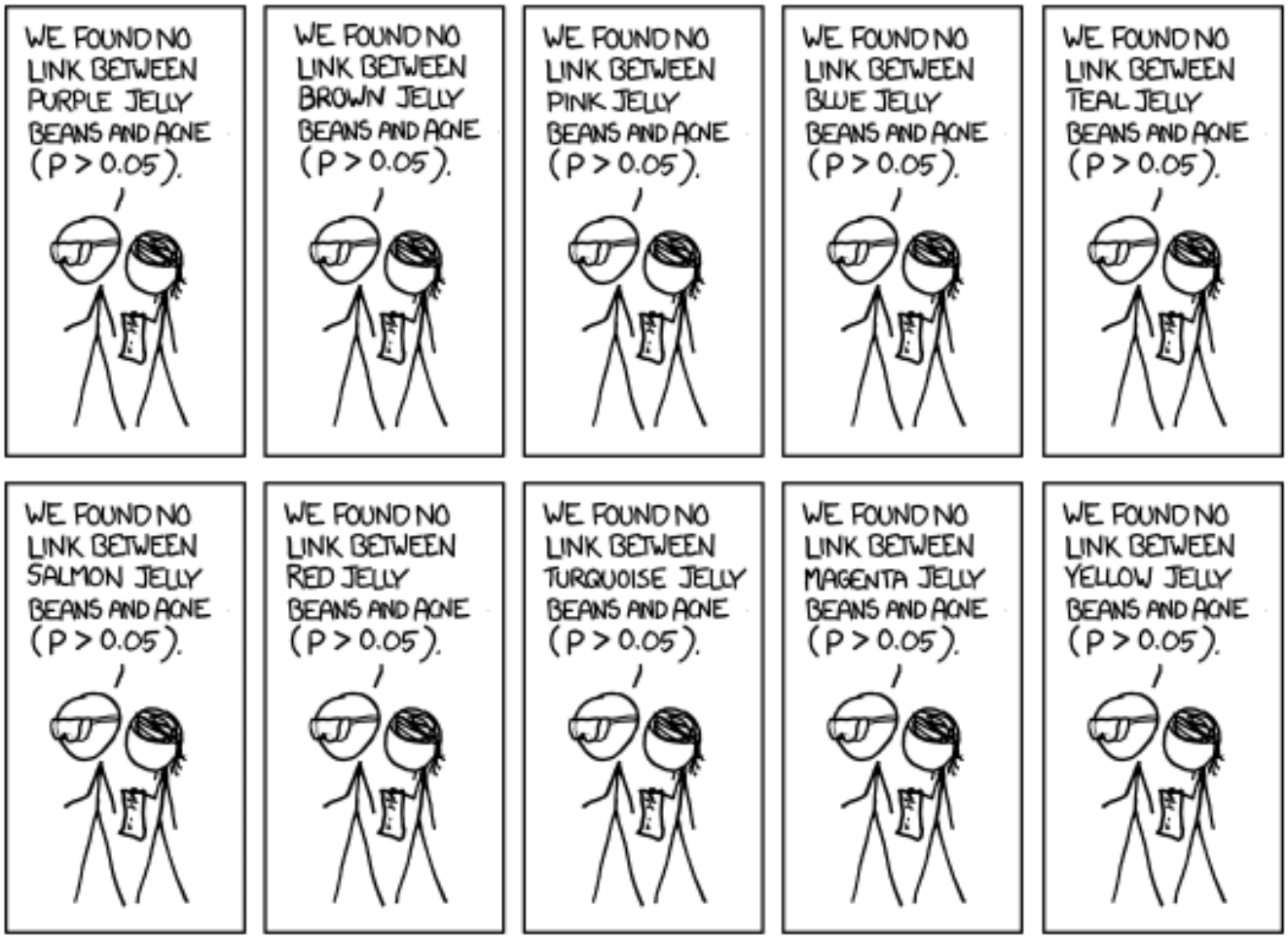
Let's do the math

$$
\begin{aligned}
& P(\text{at least one false positive}) = \\
= \quad & 1 - P(\text{no false positives}) = \\
= \quad & 1 - (0.95)^6 = \\
= \quad & 1 - 0.735 = \\
= \quad & 26.5\,\%
\end{aligned}
$$

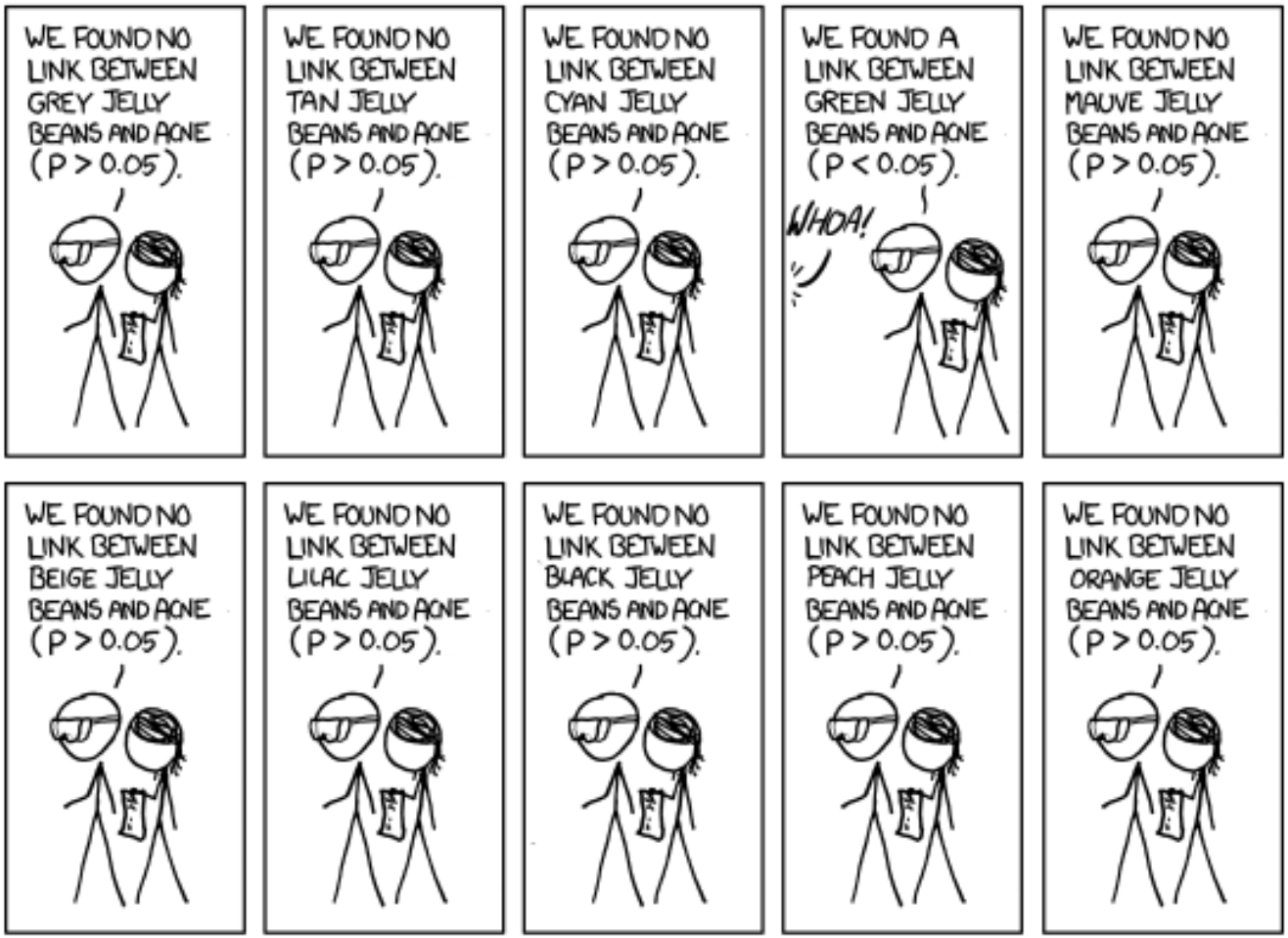
Very relevant xkcd comic



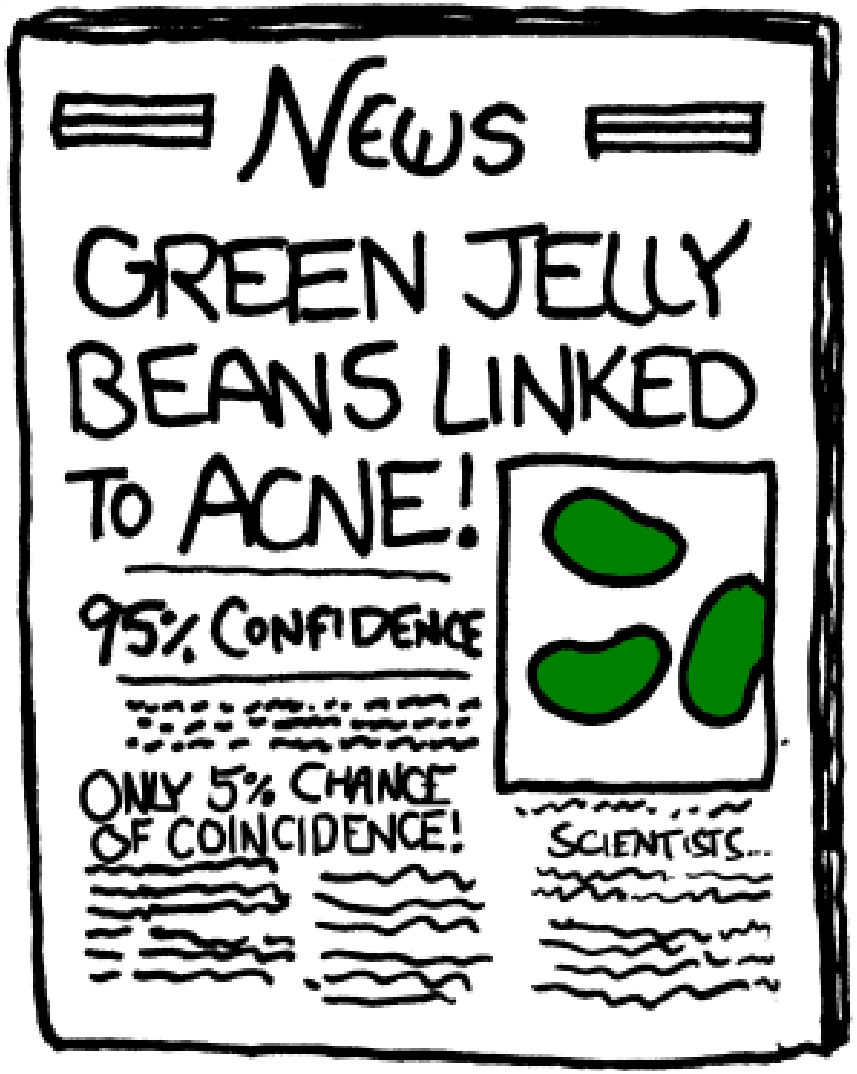
Very relevant xkcd comic



Very relevant xkcd comic



Very relevant xkcd comic



Yes, OK, but . . .

If we can't do a bunch of t-tests, what other option do we have?

Key idea

Looking not at group means, but at variation between individuals.

Key idea

Looking not at group means, but at variation between individuals.

Key question

If I select two individuals from different groups, are they going to be more different than if I select two individuals from the same group?

Can you think of another way of phrasing this question?

Key idea

Looking not at group means, but at variation between individuals.

Key question

If I select two individuals from different groups, are they going to be more different than if I select two individuals from the same group?

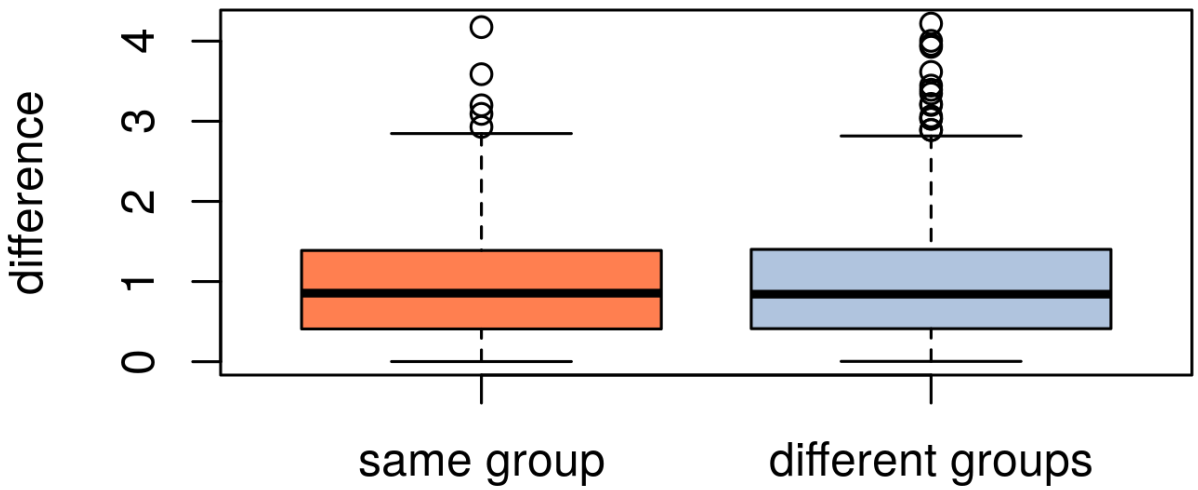
Can you think of another way of phrasing this question?

Alternative formulation

How much of the variation between individuals is explained by differences **between** groups (as opposed to differences **within** the same group)?

How does this help?

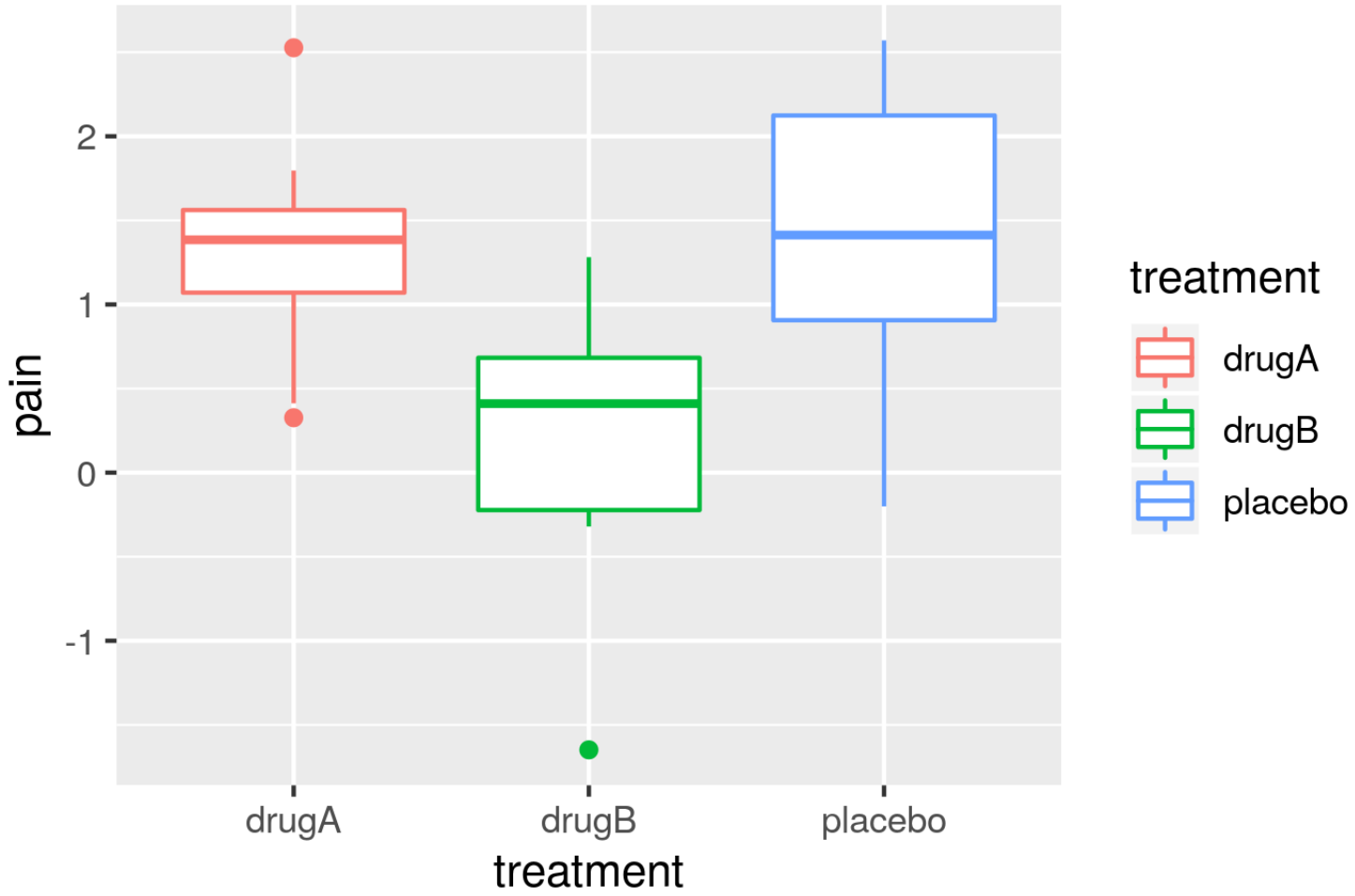
How does this help?



It once more gives us two things to compare, reducing the problem to a single test.

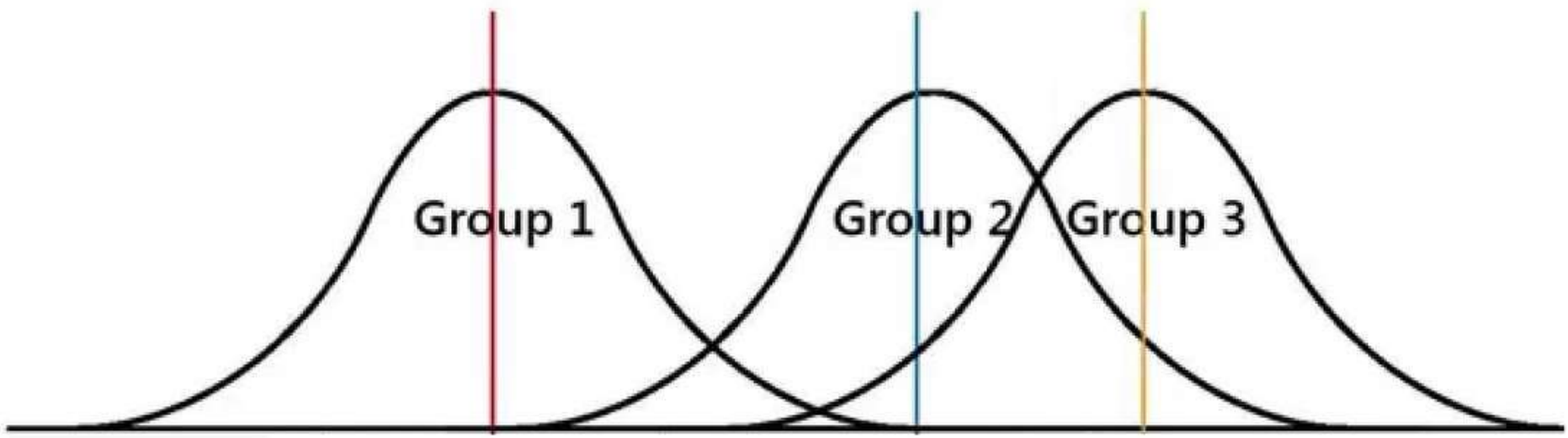
Preview: This afternoon's practical

Using a simulation-based approach to determine whether there are differences between 3 groups.



Preview: Next week's lecture

A more formal look at ANalysis Of VAriance



What questions do you have?

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

- Design and interpret a simulation-based hypothesis test
- Use a simulation-based test to compare more than two means
- Discuss limitations of t-tests
- Discuss problems around multiple testing

Image credits

- ANOVA. From a lecture by Nicola Romano for ABMS2, 2018/10.
- Four glasses of beer. By FASTILY - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=78103101
- R plots of various data sets. My own work, 2019. CC BY-SA 3.0.
- Significant. By Randall Munroe, xkcd comics https://xkcd.com/882/. This work is licensed under a Creative Commons Attribution-NonCommercial 2.5 License.
- Students working together. By Yuuki Guzman and Agoston Tyll, Okinawa Institute of Science and Technology, 2015.
- Two pints of Guinness. By Harald Bischoff - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=23344636