

Problem Set 2.2: Understanding ANOVA

ADS2

Semester 2 2023/24

We expect this problem set to take around an hour to complete. But professors are sometimes wrong!^[citation missing] If this or future problem sets are too long, please let us know, so we can adjust and plan accordingly.

Comparing t-tests to ANOVAs

“A t-test is essentially just a one-way ANOVA with only two samples”

True or false? And why?

Testing assumptions for ANOVAs

You have seen both in lecture and in practical how to test the assumptions for ANOVAs, especially normality of residuals and equality of variances. But you may feel uncomfortable looking at diagnostic plots and making a decision: how normal is normal? How similar or different can/should variances be?

In order to increase your experience with diagnostic plots, make some simulated ANOVA datasets (for instance, comparing 3 groups) with specific properties, and look at different diagnostic plots for them. For instance:

- Make a dataset where residuals are normally distributed. What does the qq-plot look like? Is it different for different sample sizes?
- Make a dataset where residuals are *not* normally distributed
- Make a dataset where all variances are equal. Do this a few times. How much variation do you see in the diagnostic plot?
- Make a dataset where different groups have very different variances. What does the plot look like?
- **Bonus:** (If you have more time than an hour for this problem set and want to get a deeper understanding of diagnostic plots.) What is a normal-QQ plot? Read up on the definition and try to produce one yourself (without using the `aov` function). What is plotted on the x axis? What is plotted on the y axis? Why is this a good way to see if the residuals are normally distributed?
- **Bonus:** (If you have more time than an hour for this problem set and want to get a deeper understanding of diagnostic plots.) What is a residual-vs-fitted plot? Read up on the definition and try to produce one yourself (without using the `aov` function). It helps to think through what the “fitted value” is for each group in an ANOVA, and how the residuals are defined. Why is this a good way to see if there is equality of variances?

Interactions

A 2-way ANOVA with interactions will give three different p-values, answering three different questions:

1. Does the first factor influence the outcome?
2. Does the second factor influence the outcome?
3. Is there an interaction between both factors?

In the practical, you plotted data from a mouse experiment, where diet and genotype both had an effect on weight gain *and* there was an interaction between diet and genotype.

Use pencil and paper to sketch what this plot would look like ...

1. ... if diet affects weight gain, but genotype does not
2. ... if diet does not affect weight gain, but genotype does
3. ... if both diet and genotype affect weight gain, but there is no interaction between the two

Originally created by MI Stefan in 2019, CC-BY-SA 3.0

Last update by DJ MacGregor in 2024