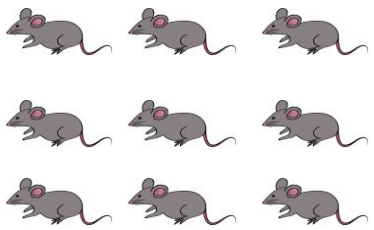
ANOVA and Multiple Comparisons

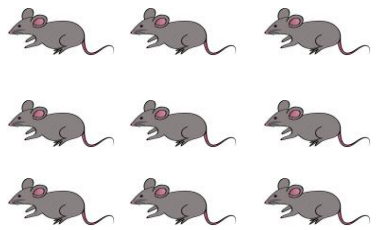
Keegan Flanagan, Jeffrey LeDue, Will Casazza, Katlyn Richardson, Megan Pawluk

More than 2 treatment groups.

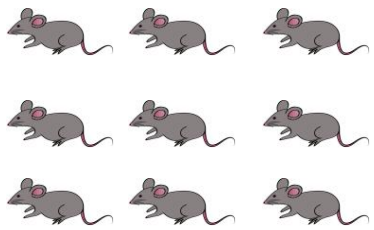
- Two sample t-tests are great, but they can only compare between two groups by design.
- What if we want to do 3 or more groups?
- Example: You have three groups of mice. One group receives a half dose of a drug, one group receives a full dose, and the last group is a control.



Control



Half Dose



Full Dose

Note: We cannot just do multiple t-tests, more on this later.

ANOVA.

- ANOVA = analysis of variance.
- In ANOVA, your null hypothesis is that all of the treatment groups have the same mean value for our variable of interest.

$$\mu_1=\mu_2=\mu_3$$

- Your alternative hypothesis is that at least one of the treatment groups has a significantly different mean

$$\mu_1=\mu_2\neq\mu_3 \text{ or } \mu_1\neq\mu_2=\mu_3 \text{ or } \mu_1\neq\mu_2\neq\mu_3$$

NOTE: ANOVA does not show us which groups are different from each other, it just tells us that at least one group is significantly different from the rest.

The Idea of Anova.

- If all of the groups are the same, then how much variation do we expect to see between the groups due to sample variation.
- If there are no differences between groups then taking a random sample from each group is the same as taking one big sample from one group.
- We treat all the datapoints like they were from one big sample and then analyze the variation within that combined sample. If there is much more variation than we would expect under the null hypothesis, then we conclude that at least one of the groups is different which is the source of the extra variation.
- How do we do this practically?

How to do ANOVA.

- There are at least 8 different values that need to be calculated to perform ANOVA.
- The calculations are not that difficult, but they are time consuming, and it is easy to get lost among the different steps.
- Practically speaking, ANOVA is always performed using coding software like R.
- Let's jump into R and give it a try.
- Go into the ANOVA_work subfolder and open up the t-ANOVA_script_working rmd file with Rstudio.

The ANOVA Table

	Df (Degrees of Freedom)	Sum.sq (Sum of Squares)	Mean.sq (Mean Square)	F value	Pr(>F) P-Value
Group	2 (DFg)	23.535 (SSg)	11.7677 (MSg)	4.3198 (F)	0.01791 (P)
Residuals (Error)	57 (DFe)	155.276 (SSe)	2.7241 (MSe)		
Total	59 (DFt)	178.811 (SSt)			

The Sum of Squares

- Sum of Squares total: All of the variation observed in the "Grand" sample.
- Sum of Squares group: All of the variation between groups.
- Sum of Squares error: All of the variation within groups.

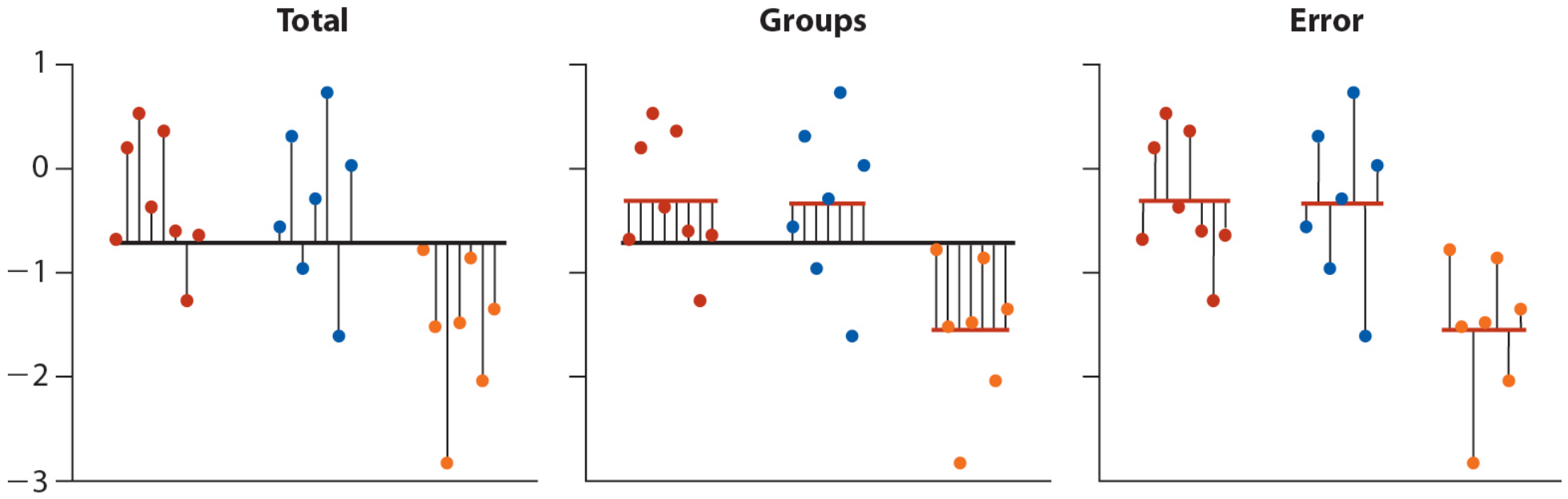


Figure 20.1: Whitock and Schluter, Fig 15.1.2 – Illustrating the partitioning of the sum of squares.

Degrees of Freedom and Mean Square.

- Degrees of Freedom.
  - Dependent on the sample size.
  - DF group is equal to the number of groups minus one (k-1).
  - DF error is equal to the total number of datapoints in all groups combined minus the number of groups (n-k)
- Mean Square.
  - Equal to the sum of squares divided by the degrees of freedom.
  - Mean Square Group is a measure of the mean variation between individuals belonging to different groups.
  - Mean Square Error is a measure of the mean variation between individuals belonging to the same group.
  - The Mean Square group and Mean Square error are comparable!

The F-statistic.

$$F = \frac{MSG}{MSE}$$

F = F-statistic. MSG= Mean Square Group MSE = Mean Square Error

- Equal to mean variation between groups over the mean variation within groups.
- If Ho is true, then the F-statistic should be equal to one.
- If Ho is false, then the F-statistic should be greater than one.
- How do we determine if F is far enough above one to conclude one of the means is significantly different?

Assumptions of ANOVA.

- Random Sampling (of course)
- The mean of the variable of interest is normally distributed in the population of every group.
  - Robust to this assumption if sample size is large (Central Limit Theorem)
- The variance of the variable of interest is the same in the population of every group.
  - Robust to this assumption if samples size is large and similar between groups.

The Shortcomings of ANOVA

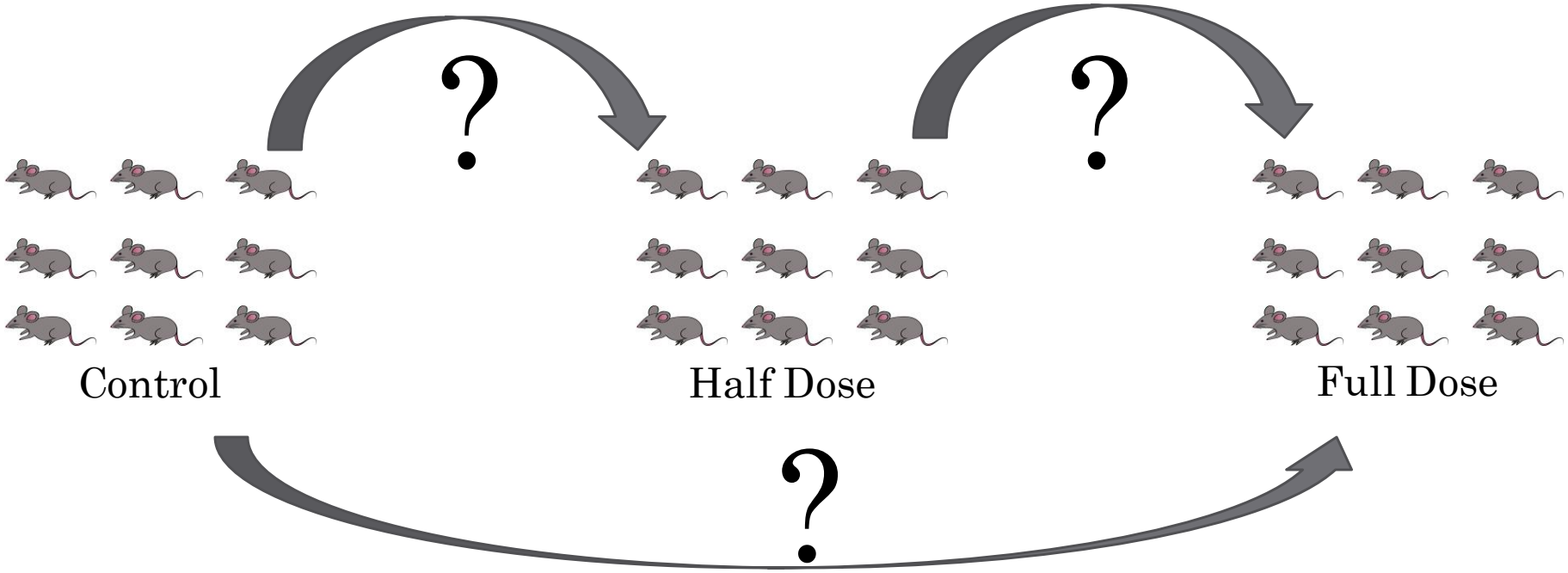
- ANOVA can only tell us if at least one group is different.
  - Often insufficient to answer the questions we wish to address.
- ANOVA allows us to quantify the sources of variation in our experiment.
  - A great way to analyze the magnitude of a treatment's effects.

Variations:

- Random effect ANOVA
  - Determine the amount of variation that is being added to your experiment from random factors (e.g. Sampling error).
- Two-Way ANOVA
  - Determine how much variation can be contributed to different variables and the interaction between different variables.
- Take Home Message: It is not just about hypothesis testing.

Multiple Comparisons

- What if we want to know which groups are different?
  - Is the half dose of our drug enough to have an effect?
  - Is the full dose of our drug more effective than a half dose?
  - Is only the full dose enough to have an effect?



Why not Just Do Multiple Tests?

- Our type 1 error rate is decided by the alpha value we choose. Alpha = 0.05 means a type 1 error occurs 5% of the time.
- Reminder: type 1 error is when the null hypothesis is true but still ends up getting rejected due to sample variation.
- Multiple comparisons mean multiple opportunities to make type 1 errors.
- The more tests we do, the more likely we are to make at some type 1 errors. This is called the multiple comparisons problem.
- Many different statistical methods have been created that address this problem. These methods are collectively referred to as multiple comparison methods.
- Multiple comparison methods attempt solve this problem by attempting to control certain error rates…

Family Wise Error Rate

- The probability of at least one type 1 error occurring.
- Example: Let's say you run 100 two-sample t-tests.
  - All 100 tests are comparing samples from truly identical populations
  - Alpha = 0.05. We would expect about 5 instances of type 1 error.
  - What are the chances of 0 type 1 errors?
  - There is a 95% chance you will not get a type 1 error for each test, so $0.95^{100} = 0.004$ is the chance of 0 type 1 errors.
  - Family wise error rate for our example = $1-0.95^{100} = 0.996$ (BAD!)
  - FWER controlling methods attempt to reduce this family wise error rate to more reasonable levels (Often 0.05).

False Discovery Rate (FDR)

- The expected proportion of incorrectly rejected null hypotheses (type 1 error occurrences) among all of the rejections.
- Example: Let's say you run 100 two-sample t-tests.
  - 90 tests are comparing samples from truly identical populations
  - 10 tests are comparing samples from truly different populations.
  - The null is rejected 15 times.
  - 5 of your rejections must be type 1 errors.
  - Your false discovery rate is 5/15 = 0.333 (BAD!)
  - FDR controlling methods attempt to reduce this false discovery rate to a more reasonable level. (often 0.05).

Controlling FDR: Benjamini-Hockberg (BH)

P-value	Rank	Critical Value (I/M * Q)
0.001	1	0.005
0.008	2	0.01
0.013	3	0.015
0.022	4	0.02
0.04	5	0.025
0.06	6	0.03
0.15	7	0.035
0.38	8	0.04
0.45	9	0.045
0.59	10	0.05

Choosing a Method

- Choosing which method you want to use should be based on the goals of your experiment.
- If the conclusions of your paper will have immediate, important, and practical consequences (eg medical studies) use FWER controlling methods.
- If the goal of your research paper is to explore potential future research projects (eg genomics) use FDR controlling methods.
- There are a LOT of alternative methods.
  - If you cannot decide, perhaps come to databinge or visit TOG.