Outline for Ch. 11 – Analysis of Variance (ANOVA)

1. Overview of when to use ANOVA
2. What is ANOVA (*F*-ratios – aka, "omnibus tests"), conceptually?
   - What question do they help us answer?
   - **Null and alternative hypotheses**
   - How are F-ratios calculated?
   - What can we conclude from the *F* ratio, and its associated p-value?
3. Comparing subsets of means
   i. Post-hoc tests
   ii. Planned contrasts

What hypotheses does ANOVA (an *F* ratio) test?

- We're testing a hypothesis about the size of the difference between multiple groups.

How does this compare to the hypothesis we're testing with a **t-test?**

  - **null hypothesis:** the difference between group means is equal to zero.

  (There is no effect of our manipulation (IV) on the dependent variable (DV).)

    - $\mu_1 = \mu_2 = \mu_3 =$..... *etc.*

  - **alternative hypothesis:** the difference btwn group means is NOT equal to 0. (There IS an effect of our manipulation. Our DV systematically varies based on conditions [groups/levels] of the IV.)

Null and Alternative Hypotheses for Note-Taking Example

- DV/outcome = after-class level of understanding of material (quiz score)
- IV/predictor = method of note-taking
  - Condition 1: take notes by hand
  - Condition 2: take notes w/laptop
  - Condition 3: take no notes

μ_1 *means*
"the **mean** quiz score among those who **took notes by hand**"

- Hypotheses
  - H_0: $\mu_1 = \mu_2 = \mu_3$
    - Method of note-taking has **no effect** on quiz score.
  - H_1: *TBD* Symbolic version is a bit long and complex, but we'll get to it later!
    - Method of note-taking has **an effect** on quiz score.

Practice with Null and Alternative Hypotheses

Research Question: Does a student's political ideology relate to the number of nights per week that they drink alcohol?

Students were asked:

- Which of these best describes your political ideology?
  - 1=liberal, 2=moderate, 3=conservative
- On average, how many nights per week do you drink alcoholic beverages?
  (a quantitative variable measuring *nights* with values 1-7)

Write the null hypothesis & alternative hypothesis using words.
Use the word "difference" in your answers.

Practice with Null and Alternative Hypotheses

Research Question: Does a student's political ideology relate to the number of nights per week that they drink alcohol?

- Which of these best describes your political ideology?
  - 1=liberal, 2=moderate, 3=conservative
- On average, how many nights per week do you drink alcoholic beverages?

- **Null hypothesis:** There is no difference among students with different political ideologies in terms of how many nights per week they drink.
- **Alternative hypothesis:** There is a difference among students with different political ideologies in terms of how many nights per week they drink.
- **Null hypothesis:** There are no differences among liberals, moderates, and conservatives in terms of how many nights per week they drink.
- **Alternative hypothesis:** There are differences among liberals, moderates, and conservatives in terms of how many nights per week they drink.

Practice your understanding

1. Which is **true** of the similarities and differences between ANOVA and t-tests?

A. They both help us understand whether there are any real differences between groups.

B. ANOVA can be used when there are two or more predictors, but t-tests can only be used with one predictor.

C. ANOVA can be used when the study contains a predictor with more than two levels, but t-tests can only be used with binary predictors.

D. All of the above are true.

2. Which is an accurate description of the **null** hypothesis for ANOVA?

A. There are no relationships between the groups.

B. The are no differences among the group means.

C. The are no differences among the variables.

D. There are no relationships between three or more variables.

Slide 6

1) D
2) B

Practice with Null and Alternative Hypotheses (Step 1 of NHST)

Research Question: Does a student's political ideology relate to the number of nights per week that they drink alcohol?

- Which of these best describes your political ideology?
  - 1=liberal, 2=moderate, 3=conservative
- On average, how many nights per week do you drink alcoholic beverages?

- **Null hypothesis:** There is no difference among students with different political ideologies in terms of how many nights per week they drink.
- **Alternative hypothesis:** There is a difference among students with different political ideologies in terms of how many nights per week they drink.

- Once we have set our alpha (step 2), and collected our data, what do we do in order to find out if the null hypothesis is credible?
  - *Next step, calculate the F-ratio (step 3) and interpret it (step 4)*

Outline for Ch. 11 – Analysis of Variance (ANOVA)

1. Overview of when to use ANOVA
2. What is ANOVA (*F*-ratios – aka, "omnibus tests"), conceptually?
   - What question does it help us answer?
   - Null and alternative hypotheses
   - **How are *F*-ratios calculated?**
   - What can we conclude from the *F* ratio, and its associated p-value?
3. Comparing subsets of means
   i. Post-hoc tests
   ii. Planned contrasts

Sums of Squares

$SS_T = SS_B + SS_R$

The "O.G." SS from Ch. 3!

SS_T - Total variability: overall amount of variability in the scores of your outcome variable

- **SS_B – Between Groups variability**: variability in the scores of the outcome variable that's *explained* by the predictor/manipulation (systematic variability)
  - AKA, the part of the total variability in the outcome that is produced by being in one group vs. another group.

- **SS_R - Residual (Error) variability**: variability in scores of the outcome variable *unexplained* by the predictor (unsystematic variability, "noise").

sometimes called SS_W
Within Groups variability

E.g., Natural differences across Ps; Lack of standardized procedures; Measurement error

F ratio (aka, *F* test; aka, *F* statistic), a type of "test statistic"

Remember from Ch. 7…

$$\text{a test statistic} = \frac{effect}{error} = \frac{systematic\ variation}{unsystematic\ variation}$$

$$F = \frac{MS_B}{MS_R}$$

Mean Squared Error (MS)

- SS's can be expressed as "averages" ("means")
  (Don't worry about how to calculate,
  *just know that **MS** is calculated from **SS**, and*
  MS is like an average.)

F-ratio (aka, *F*-test; *F*-statistic) in ANOVA – a new way of thinking about it

$$F = \frac{MS_B}{MS_R} = \frac{\text{between-group variance}}{\text{within-group (error) variance}}$$

To what extent (if at all) does the variation **between** the groups of your predictor exceed the variation **within** the groups?

E.g., variation in quiz scores **within** each of the 3 groups (within no notes, within by hand, within laptop)

E.g., variation in quiz scores **between** the no notes, by hand, and laptop groups

F-ratio (aka, *F*-test; *F*-statistic) in ANOVA – a new way of thinking about it

$$F = \frac{MS_B}{MS_R} = \frac{\text{between-group variance}}{\text{within-group (error) variance}}$$

- Like with all test statistics, **larger** *F*-ratios are more likely to be statistically significant (i.e., have smaller *p*-values)
  - and lead us to reject the null (i.e., conclude there IS an effect/ a difference).

Practice Your Understanding – indicate true or false, and correct the false statements

1) Holding all else constant, as the size of the differences among the group means increases, the *F*-ratio increases.
2) A researcher works to decrease "noise" in their data, in order to decrease the *F*-ratio.
3) As the total variation (SS_T) in your data increases, the *F*-ratio decreases.
4) As within-group differences increase, the overall F-ratio is more likely to be significant.
5) In the study on note-taking behavior and quiz scores, differences in students' quiz scores produced by differences in # of hours students slept the night before would be an example of unsystematic variation.

1. T
2. F – decreasing noise will decrease the denominator of the F-ratio, which will *increase* the F ratio
3. F – if total variation increases, we do not know how the F-ratio will change. It could go up or down, depending on how the SS_B and SS_R change (do they both go up? Does only one go up?)
4. F – if SS_R (i.e., within-group diffs) increases, that means the denominator of the F-ratio is increasing, which means the overall F ratio decreases, which means the p-value increases (implying that the F-ratio is *less* likely to be significant)
5. T

Outline for Ch. 11 – Analysis of Variance (ANOVA)

1. Overview of when to use ANOVA
2. **What is ANOVA (*F*-ratios – aka, "omnibus tests"), conceptually?**
   - What question do they help us answer?
   - Null and alternative hypotheses
   - How are F-ratios calculated?
   - **What can we conclude from the *F* ratio and its associated *p*-value?**
3. Comparing subsets of means
   i. Post-hoc tests
   ii. Planned contrasts

F-ratio (aka, *F*-test; *F*-statistic) in ANOVA – a new way of thinking about it

$$F = \frac{MS_B}{MS_R} = \frac{\text{between-group variance}}{\text{within-group variance}}$$

- Larger *F*-ratios (smaller *p*-values) make it more likely you will reject the null and conclude there IS an effect (or, there ARE differences).

H_0: $\mu_1 = \mu_2 = \mu_3$
Method of note-taking has **no effect** on quiz scores.

H_1: *TBD*
Method of note-taking has **an effect** on quiz scores.
(Mean quiz scores differ across groups.)

From last time…

ANOVA (the *F*-test) is an "omnibus test"

- It tests for an overall difference between group means.
- ***However,*** *it doesn't tell us exactly which groups' means differ.*

- For a 3-level predictor, if your *F*-ratio (& p-value, & CI) reveal that there *are* differences in the means, there are multiple possible ways the means might differ . . .
  - $\mu_1 \neq \mu_2 \neq \mu_3$
  - $\mu_1 = \mu_2 \neq \mu_3$
  - $\mu_1 \neq \mu_2 = \mu_3$
  - $\mu_2 \neq \mu_1 = \mu_3$

Example research question: Does taking **notes by hand**, **w/a laptop**, or **not at all** cause better understanding of the material?

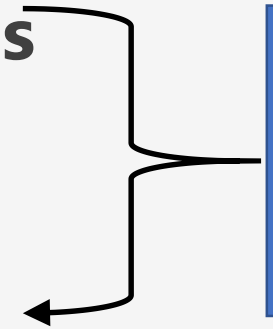
- Suppose we collected data and ran an ANOVA to answer this question. Results showed $F(2, 82) = 6.09$, $p = .012$. Assume $\alpha = .05$
- **Q**: What can we conclude from this *F*-test & significant *p*-value?

  A: *that there is a difference in quiz score across the 3 groups.*

- How do we know which groups differ? What's the next step?

Outline for Ch. 11 – Analysis of Variance (ANOVA)

1. Overview of when to use ANOVA
2. What is ANOVA (*F*-ratios – aka, "omnibus tests"), conceptually?
   - What question do they help us answer?
   - Null and alternative hypotheses
   - How are they calculated?
   - What can we conclude from the *F* ratio, and its associated p-value?
3. **Comparing subsets of means**
   i. **Post-hoc tests**
   ii. **Planned contrasts**



Note that we do not have to do these follow-up analyses for t-tests, because t-tests involve only two groups. If the t-test is statistically significant, we know the two groups' means are significantly different.

Comparing subsets of means

- The researcher must run follow-up analyses **if** the overall *F*-test is statistically significant, to figure out which groups' means differ.

- Two common types of follow-up analyses
  i. Post-hoc tests
  ii. Planned contrasts

When do you use each type?

i. Post-hoc tests

- Not testing specific hypotheses about which group means differ
- A post-hoc test compares *all pairs* of means

ii. Planned contrasts

- The comparisons you make are *hypothesis driven*, and
- the comparisons are *planned a priori*
- Planned contrasts compare certain subsets of means

When do you use each type of follow-up test?
Example: *Note-taking behavior and quiz score*

i. Post-hoc tests

- Use if there is *no or little* relevant existing research on how note-taking behavior relates to understanding of course material.

I think note-taking behavior will matter, but I'm not sure which note-taking method will lead to better vs. worse understanding.

ii. Planned contrasts

- Use if there *is* existing research suggesting that different note-taking behaviors should have certain effects on understanding of material.

Based on past research, I predict that taking no notes will lead to worse understanding than either form of taking notes, and I predict that notes by hand will cause better understanding than laptop notes.

When do you use each type?

i. Post-hoc tests

- Not testing specific hypotheses
- Compare *all* pairs of means
- More "exploratory" in nature

I can't wait to explore my data & run some **post-hoc tests**!!!

ii. Planned contrasts

- Hypothesis driven
- Planned a priori
- Compares certain subsets of means
- More "confirmatory" in nature

i. Post-hoc tests (not testing specific hypotheses; exploratory)

- They compare each group mean against every other group mean, one by one.
- Reveal which pairs are significantly different.

- Example: Running a post-hoc test in our note-taking study would lead to three comparisons . . .

by hand vs. w/laptop

by hand vs. no notes

no notes vs. w/laptop

ii. Planned contrasts (planned a priori; hypothesis driven; confirmatory)

- Based on hypotheses you had *prior to* data collection, you compare specific subsets of means.
- Planned contrasts reveal which of these subsets are significantly different.
- Note-taking behavior example:
  - No notes group VS. group taking notes by hand and group taking notes w/laptop
    - *Here, we compare the mean of one group with the mean of the other two groups combined.*
  - AND then, notes by hand vs. laptop notes

	Group 1	Group 2	Group 3
	By hand	W/laptop	No notes
	7	3	5
	4	2	2
	5	1	4
	3	1	2
	6	4	3
Mean	5.00	2.20	**3.20**

***M* = 3.6**

Practice Your Understanding

1. The output from an ANOVA (F-ratio) reveals:
   a) whether your grand mean is identical to any of your individual group means.
   b) whether your data are exploratory or confirmatory.
   c) which conditions of your IV are different from which other conditions of your IV.
   d) whether there are overall differences between your group means.
   e) All of these

2. If you are running planned contrasts for a study that involved five conditions, which of the following are possible comparisons you could make?
   a) Compare group 1 with group 5.
   b) Compare group 1 with groups 2-5 combined.
   c) Compare group 1 with groups 2, 4, and 5 combined.
   d) All of these
   e) None of these

3. What is the primary purpose of conducting follow-up tests after obtaining a significant *F*-ratio?
   a) To ensure that you replicate the effects you observed with the *F*-ratio itself
   b) To reduce the likelihood of committing a type I error when interpreting your *F*-ratio
   c) To learn exactly where the differences are among the groups of your IV
   d) All of these
   e) None of these

4. True or False?
The alternative hypothesis for ANOVA for a study with 4 levels of an independent variable would be that all group means would differ from one another.

Practice Your Understanding

1. The output from an ANOVA (F-ratio) reveals:
   a) whether your grand mean is identical to any of your individual group means.
   b) whether your data are exploratory or confirmatory.
   c) which conditions of your IV are different from which other conditions of your IV.
   d) whether there are overall differences between your group means.
   e) All of these

2. If you are running planned contrasts for a study that involved five conditions, which of the following are possible comparisons you could make?
   a) Compare group 1 with group 5.
   b) Compare group 1 with groups 2-5 combined.
   c) Compare group 1 with groups 2, 4, and 5 combined.
   d) All of these
   e) None of these

3. What is the primary purpose of conducting follow-up tests after obtaining a significant *F*-ratio?
   a) To ensure that you replicate the effects you observed with the *F*-ratio itself
   b) To reduce the likelihood of committing a type I error when interpreting your *F*-ratio
   c) To learn exactly where the differences are among the groups of your IV
   d) All of these
   e) None of these

4. True or **False**?
The alternative hypothesis for ANOVA for a study with 4 levels of an independent variable would be that all group means would differ from one another.

Q4 -There are many aspects to the alternative hypothesis. This statement states only one aspect of it. See slide 16.