Welcome to ANOVA! This kind of statistics problem can be tricky, but can be understood by breaking down each step.

Some General Tips for ANOVA:

1. Use a real calculator. You can use your phone calculator, but if you find yourself struggling with typing in equations, having a real calculator can help you track what you're doing.

2. Print out reference tables. Once again, you don't have to do this as it is on the class website, but it may help you organize your brain. Most mistakes with ANOVA come not from the actual math, but with making sure you're doing the right thing at the right time.

3. Double check your answers. You don't want to lose points for just typing in a number wrong or accidentally hitting the wrong symbol. Make sure your math is right.

4. Reach out to Dr. Yang and Abby. We are here to help you and are happy to do so. If you have questions or you get stuck, make sure you ask questions in class. You can also reach out for office hours.

ANOVA TABLE

ANOVA TABLE

	Sum of Square	df	MSS (Mean Sum of Square)	F ratio/p value	Eta-Square
Between ($SS_{between}$)	$\sum N_G \left(X_G - X_T\right)^2$	K-1	$\frac{SS_{between}}{df_{between}}$	$\frac{MSS_{between}}{MSS_{within}}$	$\frac{SS_{between}}{SS_{total}}$
Within (SS_{within})	$SS_{Total} - SS_{Between}$	N-K	$\frac{SS_{within}}{df_{within}}$		
Total (SS_{total})	$\sum \left(X_i - X_T\right)^2$	N-1 *rarely needed			

This is an ANOVA table with the formulas associated with those cells. This can help you know what formula to use when and where. Refer back to you as you complete each step to understand the process.

EXAMPLE PROBLEM 1

Year in School	Spirit Score
Freshman	6
Freshman	3
Freshman	6
Freshman	5
Sophomore	5
Sophomore	5
Sophomore	2
Sophomore	4
Junior	4
Junior	3
Junior	5
Junior	4
Senior	7
Senior	4
Senior	7
Senior	6

Independent variable (what you are changing): year in school

Dependent variable (what you measure): spirit score

Step 1: Set up your ANOVA Table

You just have to memorize this, or have it written in your notes.

	Sum of Square	df	MSS (Mean Sum of Square)	F ratio/p value	Eta-Square
Between ($SS_{between}$)					
Within (SS_{within})					
Total (SS_{total})					

Step 2: ANOVA formula

This is the complete ANOVA formula. In order to fully calculate, we need to calculate each variable and add the pieces together. First, let's break up the pieces so it's clear what is what.

$$\sum(\mathrm{X_i} - \overline{\mathrm{X_T}})^2 = \sum \mathrm{N_G}\,(\overline{\mathrm{X_G}} - \overline{\mathrm{X_T}})^2 + \sum(\mathrm{X_i} - \overline{\mathrm{X_G}})^2$$

$$\sum \overline{(X_i - X_T)^2} : Total\ Sum\ of\ Squre\ or\ SS_{Total}$$
$$\sum N_G(X_G - X_T)^2 : Between\ Sum\ of\ Square\ or\ SS_{Between}$$
$$\sum (X_i - X_G)^2 : Within\ Sum\ of\ Square\ or\ SS_{Within}$$

So, our formula can be rewritten as: $\mathrm{SS_{Total}} = \mathrm{SS_{Between}} + \mathrm{SS_{Within}}$

Calculate Group Means and Total Means

It's actually easier to start calculating SS_{Total} before $SS_{Between}$ or SS_{Within} , so we start with that. To begin calculating SS_{Total}, we need to find the Group Means and Total Means. This is similar to another mean (aka average) calculation we've done before in class.

We have 4 groups in the independent variable (year: freshman, sophomore, junior, senior)

$$\bar{X}_{FR} = \frac{6+3+6+5}{4} = \mathbf{5.0}$$
$$\bar{X}_{SO} = \frac{5+5+2+4}{4} = \mathbf{4.0}$$
$$\bar{X}_{JR} = \frac{4+3+5+4}{4} = \mathbf{4.0}$$
$$\bar{X}_{SR} = \frac{7+4+7+6}{4} = \mathbf{6.0}$$

Great! Hold onto these values. These are our group means. Next, calculate Total Mean by adding the values of the dependent variable, then divide by the total number of cases (in our case, 16)

$$\overline{X_T} = \frac{6+3+6+5+5+5+2+4+4+4+3+5+4+7+4+7+6}{16} = 4.75$$

Step 2: Calculating $\sum(X_i - \overline{X_T})^2$ aka SS_{Total}

Now that we have all the pieces to calculate SS_{Total}, let's do so. You need to take each individual out from the dependent variable and subtract it from the total mean. For example, the first response in our table is freshman, and the total mean is 4.75. Therefore, compute 6-4.75. Then, take that value and square it (1.5625). Now, do that for each value (16 times in our case).

$$(6-4.75)^2 + (3-4.75)^2 + (6-4.75)^2 + (5-4.75)^2 + (5-4.75)^2 + (5-4.75)^2 + (2-4.75)^2 + (4-4.75)^2 + (4-4.75)^2 + (3-4.75)^2 + (5-4.75)^2 + (4-4.75)^2 + (7-4.75)^2 + (4-4.75)^2 + (7-4.75)^2 + (6-4.75)^2$$

$$SS_{Total} = 31$$

Yay! Now update your ANOVA table.

	Sum of Square	df	MSS (Mean Sum of Square)	F ratio/p value	Eta-Square
Between (*SSbetween*)					
Within (*SSwithin*)					
Total (*SStotal*)	31				

Step 3: Calculating $SS_{Between}$ aka $\sum N_G (\overline{X_G} - \overline{X_T})^2$

Now that we have SS_{Total} we can move on to other SS. Start with $SS_{Between}$. Take each group mean (5, 4, 4, and 6 in our case) and subtract the total mean (4.75). Square it. Then add up all our new values.

$$SS_{Between} = 4(5 - 4.75)^2 + 4(4 - 4.75)^2 + 4(4 - 4.75)^2 + 4(6 - 4.75)\text{^}2$$

$$SS_{Between} = 0.25 + 4.5 + 6.25$$

$$SS_{Between} = 11$$

Update ANOVA:

	Sum of Square	df	MSS (Mean Sum of Square)	F ratio/p value	Eta-Square
Between ($SS_{Between}$)	11				
Within (SS_{Within})					
Total (SS_{Total})	31				

Step 4: Calculating SS_{Within} aka $\sum(X_i - \overline{X_G})^2$

We can rearrange this formula to get SS_{Within}:

$$SS_{Total} = SS_{Between} + SS_{Within}$$

$$SS_{Within} = SS_{Total} - SS_{Between}$$

Therefore:

$$SS_{Within} = 31 - 11 = 20$$

Update ANOVA:

	Sum of Square	df	MSS (Mean Sum of Square)	F ratio/p value	Eta-Square
Between ($SS_{Between}$)	11				
Within (SS_{Within})	20				

Total (SS_{Total})	31				

Step 5: Calculating df for between ($df_{between}$) and df for within (df_{within})

$$df_{between} = k - 1$$

k: total number of groups in the independent variable

N: total number of cases

In this case, independent variable is year in school (FR, SO, JR, SR), and N is 16 (because we have 16 responses).

So therefore, we have 4 groups:

$$df_{between} = k - 1 = 4 - 1 = 3$$

$$df_{within} = N - K$$

$$df_{within} = 16 - 4 = 12$$

Update:

	Sum of Square	df	MSS (Mean Sum of Square)	F ratio/p value	Eta-Square
Between ($SS_{Between}$)	11	3			
Within (SS_{Within})	20	12	1.67		
Total (SS_{Total})	31				

Step 6: Calculating Mean Sum of Square for between ($MSS_{between}$) and Mean Sum of Square for within (MSS_{within})

$$MSS_{between} = \frac{SS_{between}}{df_{between}} = \frac{11}{3} = 3.7$$

$$MSS_{within} = \frac{SS_{within}}{df_{within}} = \frac{20}{12} = 1.67$$

Step 7: Calculating f ratio

$$F_{df_{between}, df_{within}} = F_{3,12} = \frac{MSS_{between}}{MSS_{within}} = \frac{3.7}{1.67} = 2.19$$

Hold onto this value for now.

Step 8: determine the p value based on f ration table on page 473-475

Use Appendix E - F distribution (see reference tables on class website)

Row is $\mathrm{df}_{\mathrm{within}}(12)$

Column is $\mathrm{df}_{\mathrm{between}}(3)$

Use that number

If that number is lower than f ratio, then continue to p>0.1. if not, STOP

In this case 3.49, so our f ratio is higher than that (2.19), so we STOP

F ratio < crit value, then STOP

F ratio > crit value, CONTINUE to lower level (until stop)

P>0.05

Update:

	Sum of Square	df	MSS (Mean Sum of Square)	F ratio/p value	Eta-Square
Between ($SS_{Between}$)	11	3	3.7	2.21, p>0.05	
Within (SS_{Within})	20	12	1.67		
Total (SS_{Total})	31				

Step 9: decision regarding the null hypothesis, type of error committed

Cannot reject the null hypothesis, committing type II error

Rejects the null hypothesis (P<0.05)	Type I error	OK
Not to reject the null hypothesis (P>=0.05)	OK	Type II Error

While this statement is not in the ANOVA table, it is crucial to understanding ANOVA, and you will be asked for in assessments.

Step 10: calculating eta-square (E^2)

$$E^2 = \frac{SS_{between}}{SS_{total}} = \frac{11}{31} = 0.35 = 35\%$$

Update ANOVA:

	Sum of Square	df	MSS (Mean Sum of Square)	F ratio/p value	Eta-Square
Between ($SS_{Between}$)	11	3	3.7	2.21, p>0.05	35%
Within (SS_{Within})	20	12	1.67		
Total (SS_{Total})	31				

Step 11: interpreting eta-square

$$0 \leq E^2 \leq 1 \text{ or } 100\%$$

It is a PRE: knowing independent variable reduces errors in estimating the value of the dependent variable by X%

So,

Knowing year in school reduces errors in estimating spirit score by 35%.

Congrats! You just solved an ANOVA problem. Although the values may change in the data set, the math and steps involved are the same every time.