Analysis of Variance

F Test

The difference between variance

Adjusted based on Cheryl Ann Willard's slides

- Analysis of Variance (ANOVA) is a statistical method used to test differences between two or more means.
- ANOVA is used to test general rather than specific differences among means.
- An ANOVA conducted on a design in which there is only one factor is called a one-way ANOVA. If an experiment has two factors, then the ANOVA is called a two-way ANOVA. For example, suppose an experiment on the effects of age and gender on reading speed were conducted using three age groups (8 years, 10 years, and 12 years) and the two genders (male and female). The factors would be age and gender. Age would have three levels and gender would have two levels.

One-Way Analysis of Variance

Introduction to Analysis of Variance

- The *t*-statistic covered earlier
  - examines **the difference between means** to see if their differences are significant.
  - is **limited** to comparisons involving **only two groups**.
  - can’t be used to conduct repeated *t*-tests because of risk of Type I error.
  - This risk gives rise to experiment-wise alpha level – the accumulated probability of a Type I error.

- If there are more than two groups, use analysis of variance instead.
  - We will use between-subjects design, and
  - One independent variable.
  - Analysis of variance is abbreviated as ANOVA.

Variance

In a one-way ANOVA, there are two kinds of variance:

- Within-Treatments Variance – the variability within a particular sample. It is made up of
  - Individual differences
  - Experimental error

Consider to be the result of Random error, or chance.

- Between-Treatments Variance – the variability between treatment groups. It is made up of
  - Individual differences
  - Experimental error
  - Treatment effects

To Summarize

Between-treatments Variance = individual differences + experimental error + treatment

Within-treatment Variance = individual differences + experimental error

- If there are no treatment effects, between treatments variance would approximately equal within-treatments variance.

- If treatment *is* effective, between-treatments variance would be greater than within-treatment variance.

The *F*-Statistic

- The statistic generated by ANOVA is the *F*-statistic.
- It is the ratio of between-treatments variance to within-treatment variance.

$$\boxed{F = \frac{\text{Between\_treatments Variance (individual differences, experimental eror, treatment effects)}}{\text{Within\_treatment Variance (individual differences, experimental error)}}}$$

- If there were no treatment effects, the value of F would be approximately 1.
- If there *were* treatment effects, the value of *F* would be larger than 1.
- Between-treatments variance is also called treatment variance. Thus,
  - Between-treatments variance = treatment variance
- Since individual differences and experimental error are considered to be the result of random error, then
  - Within-treatment variance = error variance
- Therefore, another formulation for the *F*-statistic is:

$$\boxed{F = \frac{\text{Treatment Variance}}{\text{Error Variance}}}$$

Hypothesis Testing with the F-Statistic

- The *null hypothesis* predicts that the value of *F* will be 1.00 because it asserts no difference between means.

  H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_k$

- The *alternative hypothesis* asserts that there will be significant differences between some of the means.

  H_1: Some μs are not equal

The F-Distribution Table

- If H_0 is true, $F \approx 1.00$.
- To reject H_0, $F > 1.00$.
- Will compare our obtained *F*-value to a theoretical distribution of *F*-values.
- The *F*-distribution is a family of curves with different *df*.

Characteristics of the F-distribution table:

- Two *df* - one associated with between-treatments variance (the numerator) and one associated with within-treatment variance (the denominator).
  - $df_{\text{bet}} = k - 1$
  - $df_{\text{wi}} = N - k$
- If *df* are not shown for your problem, use next lowest *df*.
- If your obtained *F*-value is greater than the critical *F*-value reject H_0.
- If your obtained *F*-value is less than the critical *F*-value, fail to reject H_0.

In addition,

- *F*-values will only be positive; thus no two-tailed tests.

Notations for ANOVA – (Some Old; Some New)

- Sum of Squares (SS) – sum of the squared deviations. We will be calculating three SS values - SS_{total}, SS_{wi}, and $SS_{bet.}$
- ΣX_{tot} – total of all the scores in the study.
- ΣX^2_{tot} – total of all of the squared scores in the study.
- N – total number of scores in all groups.
- n – the number of scores in each group; particular groups are designated by subscripts.
- t – as a subscript, refers to individual treatment groups.
- k – the number of groups in the study.

For Example,

Life Satisfaction in Late Adulthood

Research Problem. A researcher is interested in determining if interactive activities influence the subjective experience of life satisfaction in older people. She obtains a random sample of 15 residents from an assisted living residential center, all of whom report only moderate degrees of life satisfaction. The subjects are randomly assigned to one of three interactive conditions: 1) an on-line chat group, 2) caring for pets, 3) talking with volunteer students about their life experiences. After 3 months, the residents are given the Life Satisfaction Questionnaire. Higher scores designate greater life satisfaction. Do the scores indicate significant differences between the three groups? Use $\alpha = .05$.

X1	X2	X3
Online Chat	Pets	Students

X	X^2	X	X^2	X	X^2
3	9	8	64	3	9
3	9	12	144	7	49
4	16	9	81	10	100
6	36	7	49	6	36
4	16	9	81	9	81

Online Chat	Pets	Students
$\Sigma X_1 = 20$	$\Sigma X_2 = 45$	$\Sigma X_3 = 35$
$\Sigma X_1^2 = 86$	$\Sigma X_2^2 = 419$	$\Sigma X_3^2 = 275$
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$
$M_1 = 4$	$M_2 = 9$	$M_3 = 7$

$\Sigma X_{tot} = 100$
$\Sigma X^2_{tot} = 780$
$N = 15$
$k = 3$

X1		X2		X3	
Online Chat		Pets		Students	
X	X^2	X	X^2	X	X^2
3	9	8	64	3	9
3	9	12	144	7	49
4	16	9	81	10	100
6	36	7	49	6	36
4	16	9	81	9	81
$\Sigma X_1 = 20$		$\Sigma X_2 = 45$		$\Sigma X_3 = 35$	
$\Sigma X_1^2 = 86$		$\Sigma X_2^2 = 419$		$\Sigma X_3^2 = 275$	
$n_1 = 5$		$n_2 = 5$		$n_3 = 5$	
$M_1 = 4$		$M_2 = 9$		$M_3 = 7$	

$\Sigma X_{tot} = 100$
$\Sigma X^2_{tot} = 780$
$N = 15$
$k = 3$

Step 1: Formulate Hypotheses

H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: Some μs are not equal

Step 2: Indicate the Alpha Level and Determine Critical Values

$\alpha = .05$
$df_{\text{wi}} = 12$
$df_{\text{bet}} = 2$

Step 3: Calculate Relevant Statistics

The calculations needed include: sum of squares values, mean square values, and the *F*-statistic.

Sums of Squares (*SS*) . The first step in ANOVA is calculating three sum of squares values – SS_{total}, SS_{wi}, SS_{bet}.

- SS_{tot} – refers to the sum of squares for all of the *N* scores. We will change the notation in the formula that we are used to :

$$SS = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

to make it relevant to our current statistical procedure for ANOVA:

$$SS_{total} = \Sigma X_{tot}^2 - \frac{(\Sigma X_{tot})^2}{N}$$

$\boxed{\text{For our research problem:}}$

$$SS_{\text{tot}} = 780 - \frac{(100)^2}{15}$$
$$= 113.33$$

- SS_{wi} refers to the variability within each group as measured by:

$$SS_{\text{wi}} = \Sigma\left[\Sigma X_t^2 - \frac{(\Sigma X_t)^2}{n_t}\right]$$

This formula instructs you to calculate the sum of squares (*SS*) for each treatment group and then to sum each of the *SS* values. For our research problem,

$$\begin{aligned} SS_{\text{wi}} &= \left(86 - \frac{(20)^2}{5}\right) + \left(419 - \frac{(45)^2}{5}\right) + \left(275 - \frac{(35)^2}{5}\right) \\ &= 6 + 14 + 30 \\ &= 50 \end{aligned}$$

- $\underline{SS_{\text{bet}}}$ refers to the variability between each group as measured by:

$$SS_{\text{bet}} = \Sigma\left[\frac{(\Sigma X_t)^2}{n_t}\right] - \frac{(\Sigma X_{\text{tot}})^2}{N}$$

This formula instructs you to perform the operations in the brackets for each group before subtracting the expression at the end. For our research problem,

$$SS_{\text{bet}} = \frac{20^2}{5} + \frac{45^2}{5} + \frac{35^2}{5} - \frac{100^2}{15}$$
$$= 80 + 405 + 245 + 666.67$$
$$= 63.33$$

We can check our calculations for SS because our within and between variability should equal our total variability. Thus,

$$SS_{\text{wi}} + SS_{\text{bet}} = SS_{\text{total}}$$

$$50 + 63.33 = 113.33$$

Mean Square (*MS*). In ANOVA, *MS* refers to variance, which is calculated by dividing *SS* by its corresponding *df*. We will calculate the between and within mean squares as follows:

For our data:

$$MS_{\text{bet}} = \frac{SS_{\text{bet}}}{df_{\text{bet}}}$$

$$MS_{\text{bet}} = \frac{63.33}{2}$$
$$= 31.67$$

$$MS_{\text{wi}} = \frac{SS_{\text{wi}}}{df_{\text{wi}}}$$

$$MS_{\text{wi}} = \frac{50}{12}$$
$$= 4.17$$

F-Statistic. The final calculation in ANOVA is the *F*-statistic. You have been presented with a couple of different formulations for the *F*-statistic, the latest being:

$$F = \frac{\text{Treatment Variance}}{\text{Error Variance}}$$

Let's reformulate it one more time to make it usable for our ANOVA calculations. MS_{bet} is an expression for treatment variance, and MS_{wi} is an expression for error variance. Thus, the working formula we will use for the *F*-statistic is:

$$F = \frac{MS_{\text{bet}}}{MS_{\text{wi}}}$$

For our data:

$$F_{\text{obt}} = \frac{31.67}{4.17} = 7.59$$

- We now compare our F_{obt} to F_{crit} .
- Our F_{obt} = 7.59 with df_{wi} = 12 and df_{bet} = 2. Our α = .05.
- Consulting the *F*-distribution table, we find that F_{crit} = 3.88.
- Thus, we will reject the H_0 that interactive activities have no effect on life satisfaction.

/	df_1=1	2
df_2=1	161.4476	199.5000
2	18.5128	19.0000
3	10.1280	9.5521
4	7.7086	6.9443
5	6.6079	5.7861
6	5.9874	5.1433
7	5.5914	4.7374
8	5.3177	4.4590
9	5.1174	4.2565
10	4.9646	4.1028
11	4.8443	3.9823
12	4.7472	3.8853

http://socr.ucla.edu/Applets.dir/F_Table.html

Step 4: Make a Decision and Report the Results

Arrange obtained values into a summary table:

Source	*SS*	*df*	*MS*	*F*	*p*
Between-treatments	63.33	2	31.67	7.59	< .05
Within-treatment	50.00	12	4.17		
Total	113.33	14			

Statement of Conclusion: Reject H_0. Interactive activities have a significant effect on life satisfaction.

Note the following:

- Only the *SS* and *df* will have totals. *MS* is not additive.
- The *F* column will have only one value, and it refers to F_{obt}.
- The *p*-value is the alpha level being used.
- Following the summary table, write a verbal statement of your conclusions.

Effect Size for ANOVA

A popular measure of effect size for ANOVA is called *eta squared*, symbolized as η^2. The formula is:

$$\eta^2 = \frac{SS_{\text{bet}}}{SS_{\text{tot}}}$$

- SS_{total} is a measure of total variability.
- SS_{bet} is a measure of the variability between groups as a result of manipulating the IV.
- Thus, what eta squared tells us is the proportion of the total variability (SS_{total}) that is accounted for by treatment (SS_{bet}).

For our data,

$$\eta^2 = \frac{63.33}{113.33} = .56$$

In other words, 56% of all of the score differences in the dependent variable can be explained by treatment (i.e., changing the types of interactive activities).

Post Hoc Tests

- The results of an *F*-test don’t specify which means differ significantly.
- Have to undertake further analysis.
- Post hoc tests – tests which are conducted after a null hypothesis has been rejected and there are three or more treatment groups.
  - Used to distinguish which means differ significantly.

Tukey's Honestly Significant Difference Test (*HSD*)

- Post hoc test that makes multiple comparisons between means, two at a time.
- Has built-in safeguards that minimize experiment-wise alpha level.

$$HSD = q\left(\sqrt{\frac{MS_{\text{wi}}}{n}}\right)$$

- Sample size (n) must be the same for each group.
- Value for q can be found here https://www.real-statistics.com/statistics-tables/studentized-range-q-table/ (Studentized Range Statistic).
- Need a value for k (number of treatment groups) and a value for df_{wi}.
- Use the same alpha level as for the F-test.

Alpha = 0.05

	k -->		
df	2	3	4
1	17.969	26.976	32.819
2	6.085	8.331	9.798
3	4.501	5.910	6.825
4	3.926	5.040	5.757
5	3.635	4.602	5.218
6	3.460	4.339	4.896
7	3.344	4.165	4.681
8	3.261	4.041	4.529
9	3.199	3.948	4.415
10	3.151	3.877	4.327
11	3.113	3.820	4.256
12	3.081	3.773	4.199

Three steps for conducting Tukey's *HSD* test:

1. <u>Locate the critical *q* value</u>

   For our problem: $k = 3$
   $df_{\text{within}} = 12$
   $\alpha = .05$

   Thus, $q = 3.77$ (needed for formula below)

2. <u>Calculate the *HSD* value</u>

$$HSD = q\left(\sqrt{\frac{MS_{\text{wi}}}{n}}\right)$$

$$= 3.77\left(\sqrt{\frac{4.17}{5}}\right)$$

$$= 3.44$$

3. Compare Mean Differences and Write Verbal Conclusion.

- Mean differences for all sets of means are computed.
- Arrange the means so that only positive difference values are obtained.
- Only those differences that are greater than the HSD value calculated above are significant.
- Place an asterisk beside all mean differences that reach significance.
- Write a verbal conclusion.

$M_2 - M_1 = 9 - 4 = 5^*$ (pets and online chat groups)
$M_2 - M_3 = 9 - 7 = 2$ (pets and students groups)
$M_3 - M_1 = 7 - 4 = 3$ (students and online chat groups)

There was a significant difference in life satisfaction between the pets (M_2) and online chat groups (M_1), with those caring for pets showing a significant increase in life satisfaction. There were no significant differences between the means of any of the other groups.

Assumptions for One-Way ANOVA, Independent Measures Design

- Independent and random selection of subjects.
- The dependent variable can be measured on an interval or ratio scale.
- The dependent variable is normally distributed in the populations of interest.
- Homogeneity of variance of the dependent variable in the populations of interest.

Factorial Analysis of Variance

Factorial Design

- **Factorial analysis of variance** – ANOVA that involves examining the independent and combined effects for two or more independent variables.
- **Factor** – an independent variable that can be manipulated at several different levels.

  - For this class, we will focus on a two-factor, independent samples design (akin to the independent samples t test and one-way ANOVA).
  - We will also use equal n's for all of our sample groups.

Table 1. Gender x Age Design.

Group	Gender	Age
1	Female	8
2	Female	10
3	Female	12
4	Male	8
5	Male	10
6	Male	12

This design has two factors: age and gender. Age has three levels and gender has two levels. When all combinations of the levels are included (as they are here), the design is called a *factorial design*

How to Write Notation for a Factorial ANOVA

- The notation might look something like this: 2 x 3 x 3 and describes how many independent variables (or factors) are being used as well as the number of levels at which each factor is presented.
- How many numbers there are reflect how many independent variables there are.
- The value of each number reflects how many levels (or conditions) there are for each of those independent variables.

For example,

- A two-way ANOVA with two independent variables, each being presented at three levels, would be notated as a 3 x 3 ANOVA.
- A three-way ANOVA with three independent variables, the first with three levels, the second with two levels, and the third with three levels would be notated as 3 x 2 x 3.

Main Effects and Interactions

- **Main effect** – the effect of a single independent variable on a dependent variable.

  - In a one-way ANOVA, there is only one main effect since there is only one independent variable.
  - In a factorial ANOVA, there are as many main effects as there are factors.
  - Main effects are determined by examining the differences among the means for each separate factor.
  - They provide the same information that would be attained by conducting separate one-way ANOVAs for each independent variable.

- **Interaction effects** – situation in which one independent variable is affected by different levels of another independent variable.
  - For example, prescription medications may have side effects that, when taken alone, these side effects may not be serious.
  - However, when taken in combination with other medications, the effects could be quite harmful.

For Example,

Yameric Taste Test (No Interaction)

Research Problem. Food scientists from Ancient World Spice Company have been experimenting with a newly discovered spice, yameric, harvested from the root of a yameric tree. For this study, the same vegetable casserole dish is prepared using two different vegetables, cauliflower in one and eggplant in the other. The amount of the spice added was varied under three different conditions, ½ teaspoon, 1 teaspoon, and 1½ teaspoons.

- In the table below, means for the dependent variable (flavor ratings) are shown for each condition in the relevant cells.
- The mean of the means for each level of the factors is shown in the margins. These margin means are helpful in understanding main effects.
- The participants who ate the eggplant dish seasoned with yameric gave an overall mean rating of $M_{B_1} = 6$. The overall mean rating for those who ate the cauliflower dish was $M_{B_2} = 8$. The difference between these margin means indicates the main effect for Factor *B*.
- Similarly, the differences between the margin means for the varying amounts of the spices, Factor *A* (e.g., $M_{A_1} = 4$, $M_{A_2} = 7$,$M_{A_3} = 10$), indicate the main effect for Factor A.

Statistical tests are required to determine if the main effects are significant!

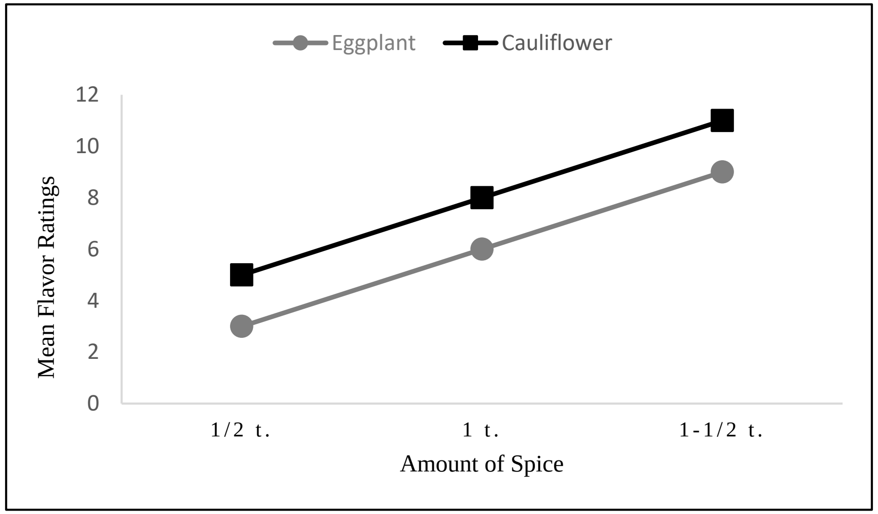
Vegetable (B)	Amount of Spice (A)			
	½ t. (A_1)	1 t. (A_2)	1½ t. (A_3)	
Eggplant (B_1)	3	6	9	$M_{B_1} = 6$
Cauliflower (B_2)	5	8	11	$M_{B_2} = 8$
	$M_{A_1} = 4$	$M_{A_2} = 7$	$M_{A_3} = 10$	

- In addition to checking for the significance of main effects, we also want to determine if there is any interaction between Factors *A* and *B*.
- That is, does the effect of one independent variable differ depending on the level of the other independent variable?
- For our example, does the effect of the amount of yameric that was added to the recipe change depending on which vegetable was used?

Vegetable (B)		Amount of Spice (A)			
		½ t. (A_1)	1 t. (A_2)	1½ t. (A_3)	
	Eggplant (B_1)	3	6	9	$M_{B_1} = 6$
	Cauliflower (B_2)	5	8	11	$M_{B_2} = 8$
		$M_{A_1} = 4$	$M_{A_2} = 7$	$M_{A_3} = 10$	

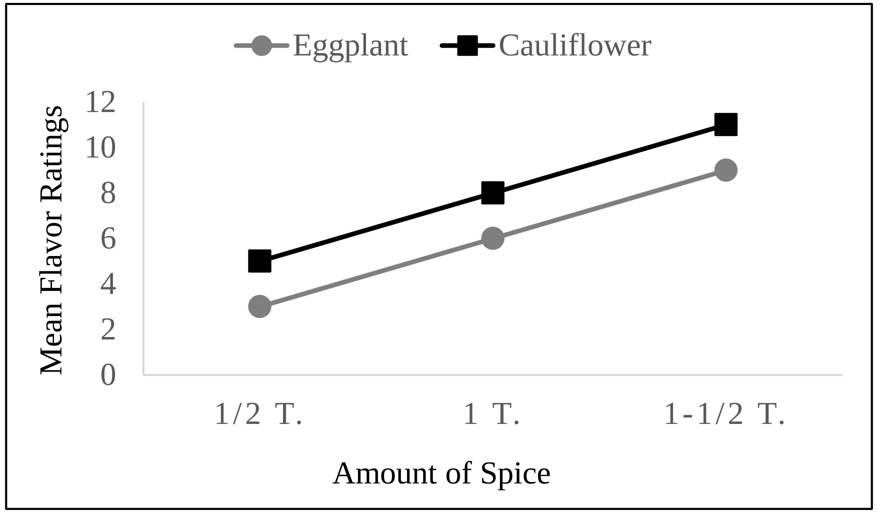
- One way to look for an interaction is to examine how the cell means change across different levels of one factor in comparison to the other factor.
- For Factor *B*, notice that subjects who ate the recipe prepared with cauliflower consistently gave 2-point higher ratings than those who ate the same dish prepared with eggplant. This was true regardless of the *amount* of the spice that was added (Factor *A*).
- Similarly, increasing the amount of the spice (Factor *A*) consistently increased the ratings by 3 points regardless of *which* vegetable was used (Factor *B*).
- Thus, the effect of one independent variable did not depend on the level of the other independent variable and there was no interaction.

The information from a table can also be illustrated in the form of a graph with the dependent variable along the Y-axis. One of the dependent variables will be shown along the X-axis and the other one will be shown as lines on the graph.



For our example:

- Mean flavor ratings are shown along the Y-axis.
- The different levels of the spice are shown along the X-axis.
- The two different types of vegetables are shown as lines on the graph.



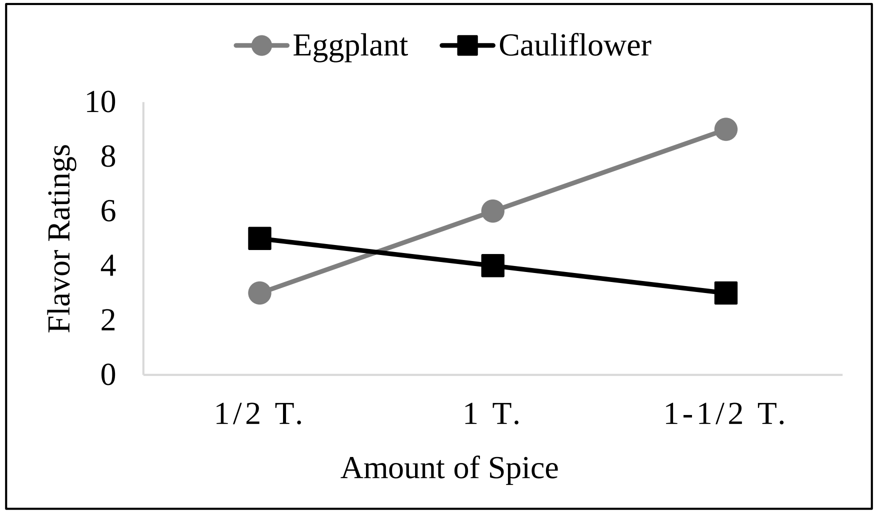
- Notice that the lines are parallel, reflecting the 2-point difference in flavor ratings between the different vegetables across all levels of the amount of spice that was added.
- Parallel lines where the data points for each line are approximately equidistant apart suggest that there is no interaction between the two independent variables.
- Lines that are not parallel, or that cross, suggest that there *may be* an interaction between Factors A and B.

Another Example,

Yameric Taste Test (This Time with an Interaction)

Vegetable (B)		Amount of Spice (A)			
		½ t. (A_1)	1 t. (A_2)	1½ t. (A_3)	
	Eggplant (B_1)	3	6	9	$M_{B_1} = 6$
	Cauliflower (B_2)	5	4	3	$M_{B_2} = 4$
		$M_{A_1} = 4$	$M_{A_2} = 5$	$M_{A_3} = 6$	

- Notice that the cell means for eggplant increase by 3 points as the amount of the spice increases.
- However, cell means for cauliflower decrease by 1 point as the spice increases.
- This infers an interaction because the effect that one independent variable has on the dependent variable depends on the level of the other independent variable.



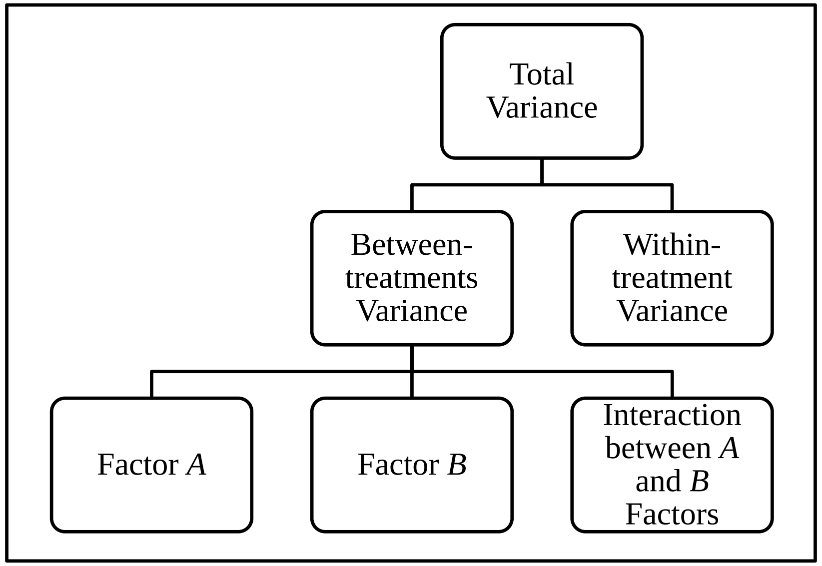
- Note that the lines cross and are not parallel. Thus, an interaction is suggested.
- But, as in the past, determining whether the interaction is significant is something that has to be determined by means of statistical tests.

Partitioning the Variance for the *F*-Statistic

total variance was broken down into two different kinds:

- Between-treatments variance – the variability between the treatment groups as a result of being treated differently (i.e., the effect of the independent variable).
- Within-treatment variance – the variability within a particular sample who are all treated the same (i.e., the result of chance, or error).

The same two sources of variance also apply to a two-way ANOVA. However, since we now have more than one independent variable, between-treatments variance is further partitioned into the variance associated with Factor *A*, the variance associated with Factor *B*, and the variance associated with the interaction between Factors *A* and *B*.



Remember the formulations of the *F*-statistic that we used in one way avova

$$t_{obt} = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{s_{M_1 - M_2}}$$

$$F = \frac{\text{Treatment Variance}}{\text{Error Variance}} = \frac{MS_{bet}}{MS_{wi}}$$

Two-way ANOVAs require calculating three different *F*-statistics, one for each component of between-treatments variance.

Comparison Table Between T-test and F-test

Parameter of Comparison	**T-test**	**F-test**
Implication	The T-test is used to test the hypothesis whether the given mean is significantly different from the sample mean or not	F-test is used to compare the two standard deviations of two samples and check the variability. An F-test is a ratio of two Chi-squares.
Null Hypothesis	H0: the sample mean is equal to u.	H0: the two samples have the same variance.
Test statistic	T = (mean – comparison value)/ Standard Error ~t(n-1) $t_{obt} = \frac{M - \mu}{s_M}$ $t_{obt} = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{s_{M_1 - M_2}}$	F = s^2_1 / s^2_2 ~ F(n1-1,n2-1)
Degree of freedom	The degree of freedom is (n-1) where n is the number of sample values	The degree of freedom is (n1-1,n2-1) where n1 and n2 are the numbers of observations in samples 1 and 2.

Hypothesis Testing for a Factorial ANOVA

The number of hypotheses to be tested in a factorial ANOVA depends on the number of factors involved.

- A separate hypothesis test for each factor is needed to determine if any of their main effects are significant.
- In addition, tests for determining any and all significant interactions between the factors are also necessary.
- The more factors involved, the more significance tests required.

Since we are limiting our application in this class to two factors, we will be conducting three significance tests:

- One for each of the main effect of Factor *A*.
- One for each of the main effect of Factor *B*.
- One for the interaction between Factors *A* and *B*.

The null and alternative hypotheses for a factorial ANOVA follow the same line of reasoning as other significance tests and will be written as indicated below.

- Main effect for Factor *A*:

  H_0: $\mu_{A_1} = \mu_{A_2} = \mu_{A_k}$

  H_1: Some μ_As are not equal

- Main effect for Factor *B*:

  H_0: $\mu_{B_1} = \mu_{B_2} = \mu_{B_k}$

  H_1: Some μ_Bs are not equal

- Interaction between Factors *A* and *B*

  H_0: Factors *A* and *B* do not have a significant interaction.

  H_1: Factors *A* and *B* have a significant interaction.

Sample Research Question without an Interaction

A sleep researcher is studying the effects of sleep deprivation on various age groups. Using two groups, teenagers and adults, the number of errors made on a math test are measured after either 3, 6, or 9 hours of sleep.

- This is a 3 x 2 ANOVA.
- Factor A is hours of sleep (and is manipulated at three levels).
- Factor B is age group (and is manipulated at two levels).

There are many steps in conducting a factorial ANOVA. They are the same calculations as required for a one-way ANOVA except that there are more of them.

- Sum of squares values, degrees of freedom, mean square values, and F-statistics are all required.
- We will not attempt to duplicate the examples in your book here, but a few areas will be highlighted.

Mean Values for Sleep Deprivation Study

		Hours of Sleep (Factor A)			
		A_1 3 Hours	A_2 6 Hours	A_3 9 Hours	
Age Group (Factor B)	Teenagers (B_1)	$M_{A_1B_1} = 7$	$M_{A_2B_1} = 5$	$M_{A_3B_1} = 2$	$M_{B_1} = 4.67$
	Adults (B_2)	$M_{A_1B_2} = 3$	$M_{A_2B_2} = 2$	$M_{A_3B_2} = 1$	$M_{B_2} = 2$
		$M_{A_1} = 5$	$M_{A_2} = 3.5$	$M_{A_3} = 1.5$	$M_{\text{tot}} = 3.33$

- Each cell is identified by its own subscript. For example, the subscript A_1B_1 refers to the treatment group involving teenagers (B_1) who received 3 hours of sleep (A_1).
- When performing the calculations for a factorial ANOVA where the subscript "$_{\text{cell}}$" is used, the calculations needed for each of the individual cells, or treatment groups, will be performed.
- The values in the row and column margins which are identified by subscripts with a single letter and number represent the values across the various levels of each of the factors. For example, M_{A_1} refers to the mean of the means for both teenagers and adults who received 3 hours of sleep (level one of Factor A).
- The subscript "$_{\text{tot}}$," represents values across all treatment groups. M_{tot}, for example, gives you the mean of means for all levels of A (i.e., A_1, A_2, and A_3). The mean of means for all levels of B will give you the same value.

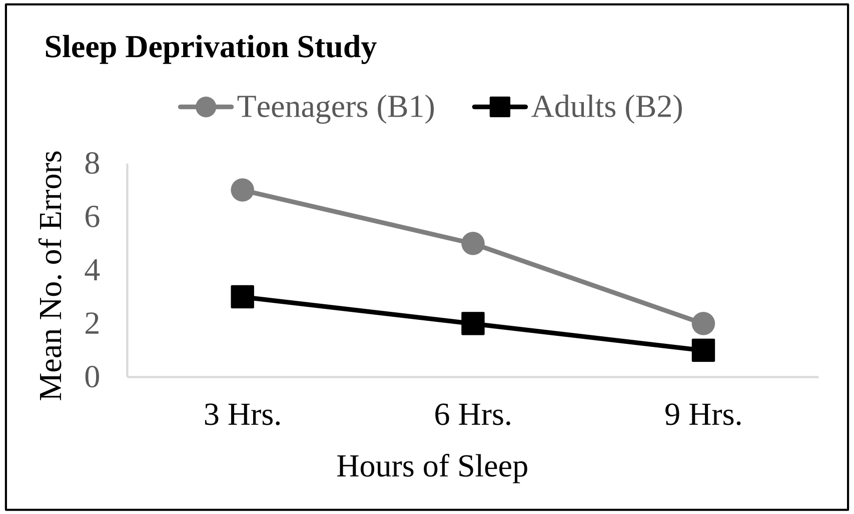
F-Statistics

Using the appropriate *df*, each of the three obtained *F*-values will need to be compared with F_{crit} values to determine significance. An alpha level of $\propto = .05$ was used. The three obtained *F*-values and the associated critical values of *F* for this problem were as follows:

- $F_A = 18.5$ for Factor A ($F_{crit} = 3.88$)
- $F_B = 32$ for Factor B ($F_{crit} = 4.75$)
- $F_{AB} = 3.5$ for the interaction between Factors A and B ($F_{crit} = 3.88$)

Graphing the Cell Means

Creating a graph of the cell means will help to ascertain if there was an interaction between Factors A and B. This should be done before proceeding with any interpretation because this result will determine how to proceed with interpreting the main effects.



Notice that the lines tend towards being parallel, suggesting no significant interaction.

Interpreting Main Effects and Interactions

- If your calculations show a non-significant interaction and significant main effects, the main effects can be interpreted in the same way as with a one-way ANOVA.
  - Each factor can be evaluated independently by means of post hoc tests.
- However, if there is a significant interaction then the main effects cannot be interpreted directly, even if their obtained F-value exceeds F_{crit}.
  - This is because the interaction makes it necessary to qualify the main effects. Remember that if a significant interaction exists, this means that the effect that one independent variable has on the dependent variable depends on the level of the other independent variable. Thus, the factors cannot be interpreted independently because they are creating a combined effect.

As we did before with a one-way ANOVA, we will arrange our obtained values into a summary table as shown. The table includes values for *SS*, *df*, *MS*, F_{obt}, and *p*-values for each source of between-treatments variance (i.e., Factors *A*, *B*, and the interaction between *A* and *B*).

Summary Table for Sleep Deprivation Study					
Source	SS	df	MS	F	p
Factor A	37	2	18.5	18.5	<.05
Factor B	32	1	32	32	<.05
Interaction	7	2	3.5	3.5	>.05
Within	12	12	1		
Total	88	17			

- Look first for any significant interaction. Our obtained F_{AB}-value was 3.5 and F_{crit} was 3.88. Thus, we can conclude that there was no significant interaction between Factors *A* and B.
    - Since that is the case, Factors *A* and *B* can each be interpreted separately.
- For Factor *A*, hours of sleep, our obtained *F*-value was 18.5 and F_{crit} was 3.88. We can therefore reject the null hypothesis for Factor *A* and conclude that increasing the number of hours slept resulted in significantly fewer math errors.
- For Factor *B*, age group, our obtained *F*-value was 32 and F_{crit} was 4.75. Thus, the null hypothesis for Factor *B* can also be rejected and we can conclude that teenagers made significantly more math errors than adults.
- For any of the individual factors that have more than two levels and that show significance, researchers will usually follow up with a post hoc test similar to Tukey's *HSD* that you learned previously.

Effect Size

Eta squared (η^2)

- In a factorial ANOVA, an effect size will be calculated for each of the significant results.
- The same statistic that was used for a one-way ANOVA, eta squared (η^2), can be used here as well.
- In this case, however, three sources of between-treatments variance are possible and so our eta squared formulas will be as follows:

$$\eta_A^2 = \frac{SS_A}{SS_{tot}} \qquad \eta_B^2 = \frac{SS_B}{SS_{tot}} \qquad \eta_{AB}^2 = \frac{SS_{AB}}{SS_{tot}}$$

For this problem, since only Factors, *A* and *B* were significant, only the effect sizes for these variables need to be considered.

For Factor *A*: $\eta_A^2 = \frac{37}{88} = .42$

For Factor *B*: $\eta_B^2 = \frac{32}{88} = .36$

- The effect size obtained by eta squared for Factor *A* suggests that 42% of the total variance in the number of errors made can be explained by the number of hours slept.
- Similarly, 36% of the total variance can be explained by age group (Factor *B*). In each case, the effect size is substantial.

- A drawback to using eta squared in a factorial ANOVA is that, as more variables are added, the proportion of variance accounted for by a single factor automatically decreases.
    - This makes it difficult to interpret the contribution of one factor to the total amount of variability.
- Another effect size value, omega squared (ω^2), is often recommended as a supplement to η^2, especially for small samples.

Omega squared (ω^2)

The formulas for omega squared are:

$$\omega_A^2 = \frac{SS_A - (df_A)(MS_{\text{wi}})}{SS_{\text{tot}} + MS_{\text{wi}}} \qquad \omega_B^2 = \frac{SS_B - (df_B)(MS_{\text{wi}})}{SS_{\text{tot}} + MS_{\text{wi}}} \qquad \omega_{AB}^2 = \frac{SS_{AB} - (df_{AB})(MS_{\text{wi}})}{SS_{\text{tot}} + MS_{\text{wi}}}$$

For our problem, again we need only to consider the omega squared values for Factors *A* and *B* separately:

For Factor *A*: $\omega_A^2 = \frac{37-(2)(1)}{88+1} = .39$

For Factor *B*: $\omega_A^2 = \frac{32-(1)(1)}{88+1} = .35$

- The effect sizes obtained by omega squared are somewhat smaller but still sizable.
- There is a tendency for eta squared to overestimate the size of an effect while omega squared tends to be more conservative.

Sample Research Question with an Interaction

Let's review an example that involves an interaction.

Research Question

A political scientist is studying whether interest in politics differs between males and females and if that interest changes as a function of education. She recruits 24 individuals, half male and half female. Four participants from each gender were randomly selected for each group from populations of high school graduates, individuals with some college education, and university graduates. The Political Interest Survey (PIS) survey was administered to all participants. Higher scores reflect greater political interest. Use $\propto = .05$.

This is a 3 x 2 ANOVA with two independent variables. One independent variable, level of education (Factor *A*), has three levels and the other independent variable, gender (Factor *B*), has two levels. The dependent variable is the degree of interest in politics.

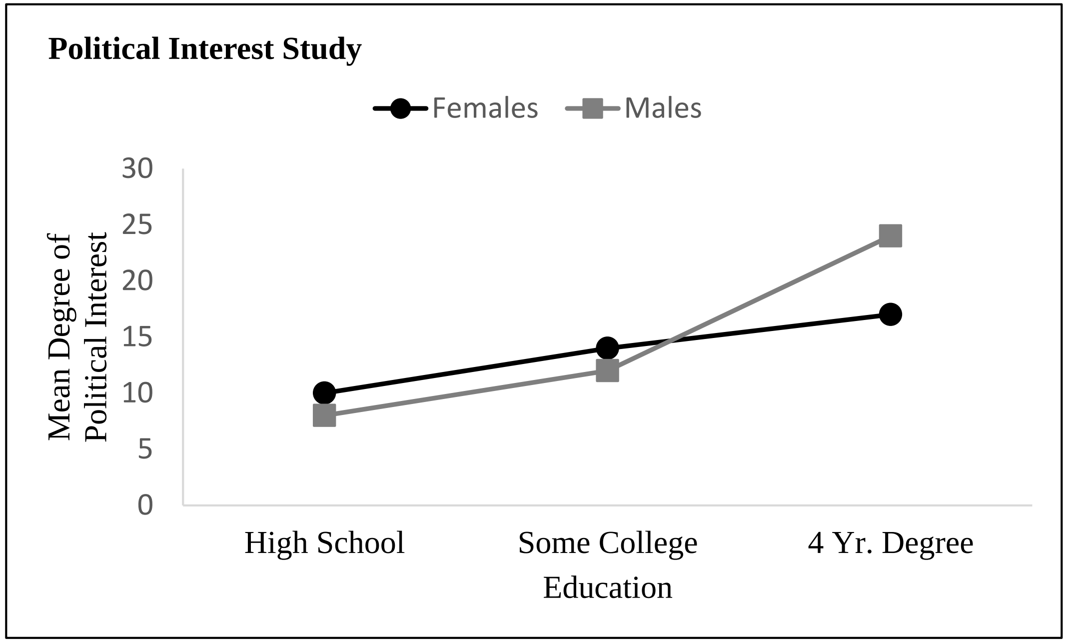
Mean Values for Political Interest Study

		Level of Education (Factor A)			
		A_1 High School	A_2 Some College	A_3 University	
Gender (Factor B)	Females (B_1)	$M_{A_1B_1} = 10$	$M_{A_2B_1} = 14$	$M_{A_3B_1} = 17$	$M_{B_1} = 13.67$
	Males (B_2)	$M_{A_1B_2} = 8$	$M_{A_2B_2} = 12$	$M_{A_3B_2} = 24$	$M_{B_2} = 14.67$
		$M_{A_1} = 9$	$M_{A_2} = 13$	$M_{A_3} = 20.5$	$M_{\text{tot}} = 14.17$

Summary Table for Political Interest Study

Source	SS	df	MS	F	p
Factor A	545.33	2	272.67	36.07	<.05
Factor B	5.99	1	5.99	.79	>.05
Interaction	108.01	2	54	7.14	<.05
Within	136	18	7.56		
Total	795.33	23			

- Remember to look first to see if there is a significant interaction.
    - If there is, it is the interaction that tells the most complete story and should be the focus of the results.
    - This is because even significant main effects can be misleading if interpreted directly since they do not consider the effect of the other independent variable.
- This is where a graph of the cell means is helpful. Notice that the lines representing gender intersect dependent on educational attainment, suggesting an interaction.



- Even though our obtained results show Factor *A* (level of education) to be significant, it is not that straightforward because the effect that educational attainment has on political interest depends on gender.
- It is best to describe interactions by looking at the pattern of cell means.
    - We can see from the graph that interest in politics for both genders increases slightly along closely parallel lines but then it increases sharply for males with a university education but does not increase at the same rate for females with a university education.
- In addition to a statement about how the variables interact, effect sizes for each of the main effects and any interactions would also be computed.

Assumptions and Requirements

- The assumptions that should be met for a factorial ANOVA are the same as those for a one-factor ANOVA, **independent samples design**. These include:
    - Normal distribution of the dependent variable in the population.
    - Independent and random selection of subjects for each group.
    - Interval or ratio scores for the dependent variable.
    - Homogeneity of variance of the scores for each treatment group.
- In addition, the independent samples model that we have examined in this class requires that the sample size be the same for each treatment group.