ANOVA is useful when you want to see if the difference between the means of three or more groups is statistically significant. You need two variables: one interval (for which you have those means) and one that has three or more groups. For example, years of education and racial/ethnic groups (white, black, Latino, Asian). We also assume random sampling and a large sample size or a normal distribution.

The procedure is the same as for other hypothesis tests: calculate your test statistic (F), compare it to the critical value for that test statistic, make a decision about the Null Hypothesis (accept or reject), and interpret what you've learned about the two variables and their relationship to each other.

$$F = \frac{MS_{betweenGroups} = \frac{SS_{betweenGroups}}{df_{betweenGroups}}}{MS_{withinGroups} = \frac{SS_{withinGroups}}{df_{withinGroups}}}$$

$$SS_{withinGroups} = \sum X^2_{total} - \sum \left(N_{group} \bar{X}^2_{group} \right)$$

$$df_{withinGroups} = N_{total} - k$$

$$SS_{betweenGroups} = \sum \left(N_{group} \bar{X}^2_{group} \right) - \left(N_{total} \bar{X}^2_{total} \right)$$

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

These are all the formulas you need to calculate ANOVA's F statistic.

To calculate this, it is extremely useful to make a summary table for the data. For example, (Research Hypothesis H_1) Does the residence of college students have an effect on their happiness? Consider the data below (left) from a study on college students, their residence type (dorm, apartment, or home) and their happiness (1=not happy, 10=happy). (Chapter 8, question #6)

	Dorm X	Dorm X^2	**Apartment** X	Apt X^2	**Home** X	Home X^2	Totals
	8	64	**2**	4	**5**	25	
	9	81	**1**	1	**4**	16	
	7	49	**3**	9	**3**	9	
	8	64	**3**	9	**4**	16	
Σ	**32**	258	**9**	23	**16**	66	
N	**4**		**4**		**4**		Σ = **12**
$\bar{X}$	32/4 = **8**		9/4 = **2.25**		16/4 = **4**		(32+9+16)/12 = **4.75**
$\bar{X}^2$	8^2 = **64**		2.25^2 = 5.06		4^2 = 16		4.75^2 = **22.56**
$\sum X^2$	258		23		66		Σ = **347**
$N * \bar{X}^2$	4*64 = **256**		4*5.06 = **20.24**		4*16 = **64**		**270.72**

Note that you could also add a row to summarize variance and/or standard deviation if you need it.

So – to calculate ANOVA, start with k and the dfs, then do the SSs, that will lead you to the MSs and voila! You've got what you need to get the F statistic!

So – what is k? k is the number of groups – which would be 3 for this data!

$df_{betweenGroups} = k - 1$ = 3 - 1 = **2** $df_{withinGroups} = N_{total} - k$ = 12 – 3 = **9**

$$SS_{betweenGroups} = \sum\left(N_{group}\bar{X}^2_{group}\right) - \left(N_{total}\bar{X}^2_{total}\right)$$

= (256+20.24+64) – 270.72 = 340.24 – 270.72 = **69.52**

$$SS_{withinGroups} = \sum X^2_{total} - \sum\left(N_{group}\bar{X}^2_{group}\right)$$ = 347 – 340.24 = **6.76**

$$MS_{between} = \frac{SS_{between}}{df_{between}} = \frac{69.52}{2} = \boxed{34.76}$$

$$MS_{within} = \frac{SS_{within}}{df_{within}} = \frac{6.76}{9} = \boxed{0.75111}$$

$$F = \frac{MS_{between}}{MS_{within}} = \frac{34.76}{.75} = \boxed{46.35}$$

So – the calculated test statistic F is 46.35!

To find the critical value of our tests statistic F, refer to the appropriate table with $df_{numerator} = df_{between}$ & $df_{denominator} = df_{within}$

From that table:

Critical Value of F (Table D) • .05 ($df_{numerator}$ $df_{between}$ = 2 & $df_{denominator}$ df_{within} =9) = 4.26

Critical Value of F (Table D) • .01 ($df_{numerator}$ $df_{between}$ = 2 & $df_{denominator}$ df_{within} =9) = 8.02

[Comparison of calculated and critical values of the test statistic and Null Hypothesis decision]:

The value of the calculated F exceeds the critical value of F, thus we REJECT THE NULL HYPOTHESIS!

[Statement of the relationship]:

There is a relationship between happiness and where students live.

[Statement of the nature or pattern of the relationship]:

Students who live in the dorms are happier (X =8) than students who live in an apartment (X =2.25).