Chapter 10

An Introduction to the Analysis of Variance

The analysis of variance (ANOVA) is one of the most powerful tools in the in the statistician's toolkit. *The purpose of ANOVA is to use available estimates of a population variance to determine if there are meaningful differences in sample means.* Mathematically quite simple, the elegant analytic structure of ANOVA allows us to think about and examine the organization of the world in new and exciting ways. As anthropology and archaeology are comparative disciplines, it is not surprising that we might be interested in determining if there are meaningful differences in the means of two or more groups.

For example, does the mean uselife of shell-tempered ceramics $\overline{Y}_1$ differ from the mean uselife of fiber-tempered ceramics, $\overline{Y}_2$, in an ethnoarchaeological study? How does sand tempered ceramics $\overline{Y}_3$ compare? We would like to test the null hypothesis about the ceramic uselife that $H_o: \mu_1 = \mu_2 = \mu_3$. The alternative hypothesis would be $H_a: \mu_1 \neq \mu_2 \neq \mu_3$.

ANOVA is useful for determining if differences do or do not indeed exist, thereby providing us with the rational to explain the observed patterns. In other words, whether or not there are significant differences, we might wish to explaining *why* that is the case. This is the core of the comparative method.

ANOVA can be used inductively to identify differences and similarities that may have otherwise gone unnoticed in our analysis. This could in turn cause us to view our data differently, or to identify important questions that we would have otherwise overlooked.

However, ANOVA is most powerful when we have theoretical, mathematical, or empirical reasons for expecting one or another outcome to be true *a priori*. In cases when we have reasons to believe that either $H_o: \mu_1 = \mu_2 = \mu_3$ or $H_a: \mu_1 \neq \mu_2 \neq \mu_3$ is in fact true, we can use ANOVA to test our underlying models and premises. If this context one or the other of our hypothesis, either H_0 or H_a, is actually the *hypothesis of interest*. The actual hypothesis test of H_0 or any H_a is then ultimately a test of whether or not our set of data meets our theoretical, mathematical, or empirical expectations – our hypothesis of interest. While ANOVA is designed to test the null hypothesis H_0, our hypothesis of interest may be either H_0 or H_a. Using an example, consider the following possible outcomes of our comparisons of two means with respect to our hypothesis of interest.

Let us say that we wish to evaluate a model of Paleoindian adaptations in North America that has theoretical expectations regarding Clovis and Folsom technologies. Let us say that the model generates hypotheses or implications that can be evaluated by examining aspects of these technologies. Furthermore, let us state that hypothesis have been created that can be tested by examining one attribute of these technologies, the comparative length of Clovis and Folsom points.

1. If the model predicts no difference between the means (we believe *a priori* that H_0 reflects the true outcome) and find one by rejecting H_0, we have rejected our hypothesis of interest H_0 and, by extension, our model.

2. If the model predicts no difference, and we fail to reject our hypothesis of interest H_0, the model is not rejected.

3. If the model predicts a difference (that is, we believe *a priori* that H_a reflects the true outcome), and we reject H_0, we have supported our hypothesis of interest H_a, and the model is not rejected.

4. If the model predicts a difference (we believe *a priori* that H_a reflects the true outcome) and found none, we have rejected our hypothesis of interest H_a in favor of H_0. Our model is rejected.

The above four possible outcomes are presented in Table 10.1 below.

Table 10.1. Possible outcomes of ANOVA analysis.

	Model predicts H_0 is true	Model predicts H_0 is false
Reject H_0	reject the model	do not reject the model
Fail to reject H_0	do not reject the model	reject the model

The consequences of these different outcomes have meaning only within the research question being addressed. If the model is rejected, one may wish to reformulate it, or consider an alternative model. If the model is not rejected, one needs to consider the full implications of the conclusion. What does it really mean, and what are the full consequences? Hopefully new questions will arise that need to be addressed.

Model I and Model II ANOVA

Two basic types of ANOVA exist; *Fixed Effects ANOVA (Model I)* and *Random Effects ANOVA (Model II)*. Computationally identical, the differences in the two models only become important when we construct our explanation of the source of the differences. In Model I, these effects are imposed upon the data by the investigator, and our hypothesis is testing whether or not the imposition of the *treatment* of the phenomena under investigation results in a significant difference in means. In Model II ANOVA, the investigator wants to explain differences that can be observed, but the source of the differences is beyond the investigator's control. Most archaeological applications will be Model II for the obvious reason that past behavior is outside of the archaeologist's control, although experimental studies such as those used to evaluate taphonomic processes can use Model I ANOVA.

To consider the differences in Model I and Model II ANOVA, let us consider the following anthropological example. Table 10.2 presents primate body temperatures in Centigrade for four groups of primates, A, B, C, and D (the symbol a is used to represent the number of groups. Here, a=4). Each group has data on 6 individuals (n=6).

Table 10.2. Primate body temperatures for four groups of primates.

	A	B	C	D
	37.90	37.70	37.70	37.90
	37.60	37.60	37.80	38.10
	38.10	37.90	38.00	38.40
	38.20	37.80	38.00	38.20
	37.80	37.50	37.80	37.60
	37.60	37.60	37.50	37.70
$\sum^{n} Y$	227.20	226.10	226.80	227.90
$\bar{Y}$	37.87	37.68	37.80	37.98
$\sum^{n} Y^2$	8603.62	8520.31	8573.22	8656.87
$\sum^{n} y^2$	0.31	0.11	0.18	0.47
$\bar{\bar{Y}} = 37.83$				

Table 10.2 also presents $\sum^{n} Y$, the sum of the individual variates calculated within each group; $\bar{Y}$, the mean for each group; $\sum^{n} Y^2$, the sum of all of the squared variates within each group; $\sum^{n} y^2$, the sum of the squared deviations of each variate within each group from its group mean; and $\bar{\bar{Y}}$, the grand mean for the table. These values will be used below in our calculation of the ANOVA.

This table represents a sample, and as a sample, we can use this data to estimate population parameters. As the name Analysis of Variance suggests, we are interested in various estimates we can make of the population variance σ^2.

The most obvious estimates of σ^2 is the s^2 for each groups. With four groups, A, B, C, and D, we can calculate four estimates s^2_A, s^2_B, s^2_C, and s^2_D where we use our usual formula for the sample variance:

$$s^2 = \frac{\sum y^2}{n-1}$$

our estimates are as follows:

$$s_A^2 = \frac{.31}{5} = .062$$

$$s_B^2 = \frac{.11}{5} = .022$$

$$s_C^2 = \frac{.18}{5} = .036$$

$$s_D^2 = \frac{.47}{5} = .094$$

Each of these is an estimate of σ^2, but which is the best estimate? We don't know, so instead we can calculate an average estimate as follows:

$$s^2 = \frac{\sum y_A^2 + \sum y_B^2 + \sum y_C^2 + \sum y_D^2}{(n_A - 1) + (n_B - 1) + (n_C - 1) + (n_D - 1)},$$

where the numerator is the sum of the sum of squares for each group, and the denominator is the sum of the degrees of freedom for each group. For our example:

$$s^2 = \frac{.31 + .11 + .18 + .47}{5 + 5 + 5 + 5} = \frac{1.07}{20} = .053$$

This value is called the average variance within groups, or more commonly, the *variance within groups*. It is an average of the sum of deviations of each variate from its group mean for the four groups. If the four groups are drawn from the same population, the average variance within groups should be a better estimate of the population variance than would any single group variance. This sum is called the *variance within groups* because the sum of squares is calculated *within* the groups.

Another estimate of the population variance σ^2 can be calculated by treating the four group means as if they were a sample of four observations, and calculate the sum of the squared deviations of each group mean $\overline{Y}$ from the grand mean $\overline{\overline{Y}}$, or $\sum^{a}(\overline{Y} - \overline{\overline{Y}})^2 = .047$.

This is called the sum of squares of means. We can calculate the sum of squares of means by literally subtracting the grand mean from each sample mean, squaring these values, and then adding them together, i.e., $\sum^{a}(\overline{Y} - \overline{\overline{Y}})^2$, but it is slightly easier to calculate the sum of squares of means with the following formula:

$$\sum^{a} \overline{Y}^2 - \frac{(\sum^{a} \overline{Y})^2}{a}$$

Both formula are computationally equivalent and produce the same result. Using the date from Table 10.2, the sum of squares of means is:

$$5725.24 - \frac{151.33}{4} = .047$$

After the sum of squares of means is calculated, we can calculate the variance among means $s^2_{\overline{Y}}$ as follows:

$$s^2_{\overline{Y}} = \frac{\sum^{a} (\overline{Y} - \overline{\overline{Y}})^2}{a - 1}$$

for our example:

$$s^2_{\overline{Y}} = \frac{.047}{3} = .015$$

Our goal in calculating $s^2_{\overline{Y}}$ is to provide an estimate of σ^2. However, $\sigma^2 = \frac{\sum (Y - \mu)^2}{n}$ whereas $s^2_{\overline{Y}} = \frac{\sum (\overline{Y} - \overline{\overline{Y}})^2}{a - 1}$. Given that $\frac{\sum (Y - \mu)^2}{n} \neq \frac{\sum (\overline{Y} - \overline{\overline{Y}})^2}{a - 1}$ how can we use $s^2_{\overline{Y}}$ to estimate σ^2? We must first recall the relationship between the standard error and the standard deviation and use our knowledge of $s^2_{\overline{Y}}$ to solve for s^2, which can be use as an estimate of the population parameter σ^2. Consider the following relationships:

First, with knowledge of the standard deviation σ, the standard error can be calculated as follows:

$$\sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{n}}$$

Squaring both sides of the equation we get:

$$\sigma^2_{\overline{Y}} = \frac{\sigma^2}{n}$$

Multiplying both sides of the equation by n we obtain:

$$n\sigma^2_{\overline{Y}} = \sigma^2$$

The same relationship holds for s^2 such that $s^2 = n(s^2_{\overline{Y}})$. Given this relationship, once we have calculated the variance among means $s^2_{\overline{Y}}$, multiplying this value by n provides us with s^2, an additional estimate of our population parameter σ^2. For our example:

$$s^2 = n(s^2_{\overline{Y}}) = 6(.015) = .090$$

This estimate of the population variance σ^2 is called the *variance among groups*.

For our example, we now have used our data to create two independent estimates of the population variance σ^2:

We have the *variance within groups*: .053.

We have the *variance among groups*: .090.

Remember that the sum of squares of the variance *within groups* considered the squared differences of each variate from its group mean $(Y_{ij} - \overline{Y}_i)^2$, while the sum of squares of the variance *among groups* considered the squared differences of each group mean from the grand mean $(\overline{Y} - \overline{\overline{Y}})^2$. (Don't worry for now about Y_{ij}, which is the symbol for an individual variate. This symbolism will be explained below.) Conceptually, we can consider the former a measure of the average dispersion of variates around their group means, and the later a measure of the average dispersion of the group means around the grand mean.

Here is the core of ANOVA. If these two estimates of the population variance are close, we know that they are equally good estimates of the population parameter. This can only happen if the group means are sufficiently close enough to each other, and are therefore statistically close to the grand mean, i.e., $\overline{Y}_1 = \overline{Y}_2 = \overline{Y}_3 = ... = \overline{Y}_a = \overline{\overline{Y}}$. Yet, if the estimate of σ^2 are different—if the variance among groups exceeds the variance within groups in a meaningful way—there is a significant difference among the means. In other words, if the variation *between* groups exceed the variation *within* groups such that there is more difference between Groups A, B, C, and D than within the groups, the means of the groups are not close enough to be considered equal. Establishing whether or not significant difference among means exist is the purpose of ANOVA.

Let us consider these relationships by presenting the proper formulas. As a first step, we need to recognize that we need symbols to represent the location of columns, rows, and individual observations. Table 10.3 presents only the data from Table 10.2

Table 10.3. The body temperature of four samples of primates.

A	B	C	D
37.90	37.70	37.70	37.90
37.60	37.60	37.80	38.10
38.10	37.90	38.00	38.40
38.20	37.80	38.00	38.20
37.80	37.50	37.80	37.60
37.60	37.60	37.50	37.70

Table 10.4 shows that each observation can be specified by which group it belongs to (i), and which observation it is within that group (j).

Table 10.4. Illustration of the determination of Y_{ij}.

		a groups (i)			
		1	2	3	4
	1	1,2	2,1	3,1	4,1
	2	1,2	2,2	3,2	4,2
n items (j)	3	1,3	2,3	3,3	4,3
	4	1,4	2,4	3,4	4,4
	5	1,5	2,5	3,5	4,5
	6	1,6	2,6	3,6	4,6

For example, $Y_{1,1}$=37.90 and $Y_{1,4}$=38.20 in Table 10.3.

To illustrate ANOVA, consider the following conceptual framework and how it integrates with Table 10.4. To illustrate how the variance of a sample is calculated, let us consider the symbolism of the calculation of the variance of Group 1:

$$\frac{1}{n-1}\sum_{j=1}^{j=n}(Y_{1,j}-\overline{Y}_1)^2$$

This symbolism tells one to subtract from each individual variate the group mean, square these values, sum them, and then divide the total by the degrees of freedom. This results in the variance for the first group.

The variance within groups is calculated by simply expanding the formula to include all of the groups. It is calculated in the following way:

$$\frac{1}{a(n-1)}\sum_{i=1}^{i=a}\sum_{j=1}^{j=n}(Y_{i,j}-\overline{Y}_i)^2$$

This set of instructions is informing one to subtract from each variate its own groups mean, square these values, sum all of them together, and then divide by the degrees of freedom for all of the groups. This results in the variance *within* groups, and is one of the two estimates of the population variance that we can create. The symbol for double summation $\sum_{i=1}^{i=a}\sum_{j=1}^{j=n}$ looks frightening, but is not. As a set of instructions it is telling one to sum, for the first group and continuing through all groups, the first through the last observation in each group.

The variance *among* groups can be illustrated by the following set of instructions:

$$\frac{n}{a-1}\sum_{i=1}^{i=a}(\overline{Y}_i - \overline{\overline{Y}})^2$$

This set of instructions tells us to subtract the grand mean from each group mean, square these values, sum them, multiple the summed value by the number of variates in each group, and then divide by the degrees of freedom. This results in the variance among groups, our second estimate of the population variance.

The next question to ask is whether or not the two statistics estimate the same population parameter. To put it another way, we can also ask if the two sample variances were drawn from the same population. The F distribution (Appendix C) provides the means of answering this question.

The F distribution allows us to determine if the ratio of two variances is larger than we would expect if they estimate the same parameter, σ^2. If the two sample variances s^2_{among} and s^2_{within} estimate the same parameter, the values of each should be approximately equal and the ratio $\frac{s^2_{among}}{s^2_{within}}$ should be equal to 1. The F distribution allows us to determine if the ratio deviates sufficiently from 1 such that we would have to conclude that these estimates are of different parameters (were drawn from different populations). By convention, s^2_{among} is placed in the numerator, and s^2_{within} in the denominator. While we have expressed the ratio as $F_s = \frac{s^2_{among}}{s^2_{within}}$, it is more common to use the symbolism of $F_s = \frac{s_1^2}{s_2^2}$ where $s_1^2 = s^2_{among}$ and $s_2^2 = s^2_{within}$.

The F distribution depends on the degrees of freedom (and therefore the sample size of the two values being compared), and therefore, there are an infinite number of F distributions. Appendix 3 provides a list of critical values for different degrees of freedom ν, for ν_1 and ν_2. For our example $\frac{s_1^2}{s_2^2} = \frac{.090}{.053} = 1.698$, where $\nu_1 = a - 1 = 4 - 1 = 3$ degrees of freedom, and $\nu_2 = a(n-1) = 4(6-1) = 20$ degrees of freedom. If $\alpha = .05$, we must accept the null hypothesis that there is no difference in the two estimates, as Appendix C illustrates that the value of 1.698 is significant only between alpha being equal to .25 and alpha equal to .10. Therefore, we conclude that variation among means does not exceed significantly the variation within. As a result, we must conclude that there is no significant differences in the means, and that they estimate the same population parameter, μ.

This is a Model II, or random effects ANOVA, as the influences on primate body temperature are beyond the control of the investigator in this case.

A Consideration of Treatment Effects

To illustrate Model I ANOVA, let us say that we are able to subject the primates under study to treatment effects that resulted in changes in their temperatures. These treatment effects are symbolized by α_i. For each group, we offer a treatment α_i, where the treatment results in the following increase or decrease in temperature:

$\alpha_1 = -.7$

$\alpha_2 = .2$

$\alpha_3 = -.3$

$\alpha_4 = .8$

With these treatments, our data now looks like the data in Table 10.5:

Table 10.5. Primate body temperature data from Table 10.3 modified by treatment effects.

	A ($\alpha_1 = -.7$)	B ($\alpha_2 = .2$)	C ($\alpha_3 = .3$)	D ($\alpha_4 = .8$)
	37.20	37.90	37.40	38.70
	36.90	37.80	37.50	38.90
	37.40	38.10	37.70	39.20
	37.50	38.00	37.70	39.00
	37.10	37.70	37.50	38.40
	36.90	37.80	37.20	38.50
$\sum^{n} Y + n\alpha_i$	223.00	227.30	225.00	232.70
$\overline{Y}_i + \alpha_i$	37.17	37.88	37.50	38.78
$\sum^{n} Y^2$	8288.48	8610.99	8437.68	9025.35
$\sum^{n} y^2$	0.31	0.11	0.18	0.47
$\overline{\overline{Y}} = 37.83$				

To evaluate if the variation among groups exceeds the variation within, we proceed with our ANOVA by calculating the sum of squares of means:

$$\sum^{a} \overline{Y} = 151.33$$

$$\overline{\overline{Y}} = 37.83$$

$$\sum^{a} \overline{Y}^2 = 5726.64$$

$$\sum^{a} (\overline{Y} - \overline{\overline{Y}})^2 = 1.449$$

The variance *within groups*:

$$s_w^2 = \frac{1}{a(n-1)}\sum_{i=1}^{i=a}\sum_{j=1}^{j=n}(Y_{ij} - \overline{Y}_i)^2$$

$$s_w^2 = \frac{1.07}{20}$$

$$s_w^2 = .053$$

The variance *among means*:

$$s_{\overline{Y}}^2 = \frac{\sum^{a}(\overline{Y} - \overline{\overline{Y}})^2}{a-1}$$

$$s_{\overline{Y}}^2 = \frac{1.449}{3}$$

$$s_{\overline{Y}}^2 = .483$$

The variance *among groups*:

$$s_a^2 = n(s_{\overline{Y}})^2$$

$$s_a^2 = 6(.483)$$

$$s_a^2 = 2.898$$

And finally, we can examine the relationship between the variance among groups and the variance within groups by using the F distribution:

$$F = \frac{s_a^2}{s_w^2}$$

$$F = \frac{2.898}{.053}$$

$$F = 54.679$$

This ratio shows that the variance among groups greatly exceeds the variation within groups (by 54.679 times). Using Appendix C where the degrees of freedom are: $\nu_a = a - 1 = 3$ and $\nu_w = a(n-1) = 20$, we see that 54.679 exceeds the critical value of 8.10 at α =.001. Notice that our variance within groups $s_w^2 = .053$ is identical in our original data, and only the variance among groups s_a^2 changes with the addition of treatment effects. This is because the addition of our treatment effects only impacts the difference among means, not the difference between individual observations and their group means. Given the results of our analysis, we must conclude that there is a difference in treatments, and that the two statistics s_w^2 and s_a^2 estimate different parameters, as a result of the added component due to treatment effects. The ratio actually estimates the following parameters where we have treatment effects:

$\frac{s_a^2}{s_w^2}$ estimates $\frac{\sigma^2 + \frac{n}{a-1}\sum^a \alpha^2}{\sigma^2}$ where $\frac{n}{a-1}\sum^a \alpha^2$ in the numerator represents the average treatment effects. If there are no treatment effects, $\frac{s_a^2}{s_w^2}$ estimates the ratio $\frac{\sigma^2}{\sigma^2}$.

Now that we understand how ANOVA works, a more rapid way to proceed with the calculation is as follows. Calculate the following quantities:

Quantity 1. The Grand Total.

$$\sum^a \sum^n Y = 223.0 + 227.3 + 225.0 + 232.7$$

$$\sum^a \sum^n Y = 908.00$$

Quantity 2. The sum of the squared individual observations.

$$\sum^a \sum^n Y^2 = 8288.48 + 8610.99 + 8437.68 + 9025.35$$

$$\sum^a \sum^n Y^2 = 34362.50$$

Quantity 3. The sum of the squared group totals, each divided by its sample size.

$$\sum \frac{(\sum^n Y)^2}{n_i} = \frac{(223.0)^2}{6} + \frac{(227.3)^2}{6} + \frac{(225.0)^2}{6} + \frac{(232.7)^2}{6}$$

$$\sum \frac{(\sum^n Y)^2}{n_i} = 34361.43$$

Quantity 4. The grand total squared and divided by the total sample size. This is also called the *correction term*, or CT.

$$CT = \frac{(\sum^{a}\sum^{n} Y)^2}{\sum^{a} n_i}$$

$$CT = \frac{(908)^2}{24}$$

$$CT = 34352.67$$

Quantity 5. The total sum of squares.

$$SS_{total} = \sum^{a}\sum^{n} Y^2 - CT$$

$$SS_{total} = Quantity2(SS_{observations}) - Quantity4(CT)$$

$$SS_{total} = 34362.50 - 34352.67$$

$$SS_{total} = 9.83$$

Quantity 6. The group sum of squares.

$$SS_{groups} = \sum^{a} \frac{(\sum^{n} Y)^2}{n_i} - CT$$

$$SS_{groups} = Quantity\ 3 - Quantity\ 4(CT)$$

$$SS_{groups} = 34361.43 - 34352.67$$

$$SS_{groups} = 8.76$$

Quantity 7. The sum of squares within.

$$SS_{within} = SS_{total} - SS_{groups}$$

$$SS_{within} = Quantity\ 5 - Quantity\ 6$$

$$SS_{within} = 9.83 - 8.76$$

$SS_{within} = 1.07$

Although less intuitive than the original formula we presented, these quantities produce exactly the same results as calculating the variance among and within groups. The results of these calculations are then customarily presented in ANOVA table of the following general form, presenting the sources of variation, and the associated degrees of freedom, sum of squares, mean squares, and F_s. This is represented in Table 10.6.

Table 10.6. Generalized ANOVA table.

Source of Variation	df	SS	MS	F_s
$\bar{Y} - \bar{\bar{Y}}$ among group	$a - 1$	Quantity 6	$\frac{\text{Quantity 6}}{a - 1}$	$\frac{MS_{groups}}{MS_{within}}$
$Y - \bar{Y}$ within group	$\sum^{a} n_i - a$	Quantity 7	$\frac{\text{Quantity 7}}{\sum^{a} n_i - 1}$	
$Y - \bar{\bar{Y}}$ total	$\sum^{a} n_i - 1$	Quantity 5		

In Table 10.6 you will notice a new term, the total sum of squares. This is a new source of variation, one that demonstrates the sum of the squared deviations of each variate from the grand mean. Notice that the among group sum of squares and the within group sum of squares are additive to the total sum of squares.

The results of our analysis are presented in an ANOVA table in Table 10.7.

Table 10.7. ANOVA of treatment effects on primate temperatures.

Source of Variation	df	SS	MS	F_s
$\bar{Y} - \bar{\bar{Y}}$ among group	3	8.76	2.92	55.094
$Y - \bar{Y}$ within group	20	1.07	.053	
$Y - \bar{\bar{Y}}$ total	23	9.83		

The ANOVA table contains all of the information necessary to evaluate whether or not there is a difference in group means by examining the ratio between the amount of variation *among* groups to the amount *within* groups. Remember, if the ratio of among groups variation exceeds the within group variation beyond the critical value determined by alpha and the appropriate degrees of freedom for the F distribution, we conclude that there is a significant differences among means.

Let us examine the ANOVA table in some detail. First, we have the following sources of variation:

$\overline{Y} - \overline{\overline{Y}}$ *among groups*; that attributed to the differences from the group means to the grand means.

$Y - \overline{Y}$ *within groups*; that attributed to the differences of each Y from its mean.

$Y - \overline{\overline{Y}}$ *total*; that attributed to the differences of each Y from the grand mean.

Each of these sources of variation has an associated *degrees of freedom* (df). The degrees of freedom are necessary to calculate a variance, or a mean square. The among groups degrees of freedom and within groups degrees of freedom are additive to the total degrees of freedom.

Each source of variation also has a *sum of squares* (SS), the sum of the deviations squared. The sum of squares are also necessary to calculate the variance, or mean square. As noted above, the among groups sum of squares and the within group sum of squares are additive to the total sum of squares.

The *mean square* (abbreviated MS) for each source of variation is the sum of squares for each source of variation divided by the appropriate degrees of freedom. It may be conceptualized as a kind of "average" deviation, although it is not a true mean, as the sum of squares is divided by the degrees of freedom, rather than the sample size.

The among groups mean square describes the dispersion of the group means around the grand mean. If there are no random or fixed effects, the among groups mean square estimates σ^2. If there are random or fixed effects, the among groups mean square estimates σ^2 plus the sum of the random or fixed effects.

The within groups mean square describes the average dispersion of the observations in each group around the group means. The within groups mean square estimates σ^2 if the groups are random samples from the same distribution.

The total mean squares is calculated by dividing the total sum of squares by the total degrees of freedom, and describes the dispersion of the variates around the grand mean. This describes the variation due to all causes. It estimates σ^2 if there are no fixed or random effects.

To illustrate the additive effects of deviations, consider that for an individual variate, the difference between its group mean and the grand mean, plus the deviation of that variate from its group mean, is equal to the difference between that variate and the grand mean. Mathematically, this statement is symbolized as:

$$(\overline{Y} - \overline{\overline{Y}}) + (Y - \overline{Y}) = (Y - \overline{\overline{Y}})$$

For the individual variate $Y_{2,3} = 37.9$ in Table 10.3 where $\overline{Y} = 37.68$ and $\overline{\overline{Y}} = 37.83$:

$$(37.68 - 37.83) + (37.90 - 37.68) = (37.90 - 37.83)$$

$$-.15 + .22 = .07$$

$$.07 = .07$$

Now let us think about our data in a slightly different way.

What would be a good predictor of any individual variate Y_{ij}? Assuming there are no treatment or random effects between the groups, one of the best predictors of any individual variate would be the grand mean, $\overline{\overline{Y}}$. If we had no knowledge of the value of each variate, and used $\overline{\overline{Y}}$ to predict all individual variates, our predictions would look like Table 10.8.

Table 10.8. The data resulting from the use of the grand mean to estimate each Y_{ij}.

	A	B	C	D
	37.83	37.83	37.83	37.83
	37.83	37.83	37.83	37.83
	37.83	37.83	37.83	37.83
	37.83	37.83	37.83	37.83
	37.83	37.83	37.83	37.83
	37.83	37.83	37.83	37.83
$\sum^{n} Y$	226.98	226.98	226.98	226.98
$\overline{Y}$	37.83	37.83	37.83	37.83
$\overline{\overline{Y}} = 37.83$				

Of course, if we used the grand mean for such a prediction, there would be no variation in our table. All values would be the same—yet, we know that other factors influence the value attained by a variate. To make our most accurate prediction possible we need to consider what sources of variation influence that value.

Let us present of general model considering these influences. In both Model I and Model II ANOVA, the use of μ (because we are building a general model here, we use Greek letters to illustrate population parameters) to predict Y_{ij} has clear limitations. We depict that prediction as $Y_{ij} = \mu$. In Model I or fixed effects ANOVA, $Y_{ij} = \mu$ is insufficient because each variate is potentially subject to treatment effects. Therefore $Y_{ij} = \mu + a_i$ is a better model.

Yet, $Y_{ij} = \mu + a_i$ is insufficent as well, as there is most often some deviation, however small, of the actual variate from the resultant prediction. This error is symbolized as e_{ij}. The source of this error is unknown. It may be caused by unknown factors that influence the value of the variate, or it may be a product of measurement error. Importantly, errors are expected, as our models are almost always at variance from the real world in some way. This error term allows us to see exactly how much our real world is at odds with the model.

Our best model then is $Y_{ij} = \mu + a_i + e_{ij}$. To illustrate this prediction consider $Y_{3,1} = \mu + a_i + e_{ij}$, where $Y_{3,1} = 37.7$. Continuing, $Y_{3,1} = \mu + a_i + e_{ij}$ may be expressed as $37.7 = 37.83 + (-.3) + e_{3,1}$. Solving for $e_{3,1}$, we find that $e_{3,1}$=.17. Notice that this deviation is calculated *within groups*—which is the reason the within groups MS is often called the *error* MS.

When plotting our e_{ij}, they too would have a distribution, and that distribution should approximate normality because the variation should be *random*. If the distribution is not normal, the variation is likely nonrandom and we likely have an additional treatment effect that we need to identify and control through a redesign of the experiment or through an ANOVA that considers multiple treatment effects.

If we reject our null hypothesis that the $\overline{Y}$ s are the same, we are actually saying that the treatments have different effects. To put it another way, we have been testing for differences in treatment effects from the beginning. Therefore, we could have actually phrased our null hypothesis in the following ways: $H_o : a_1 = a_2 ... = a_n$, or $H_o : (a_1 - a_2) = 0$.

Recall here that our general model for Model I ANOVA is $Y_{ij} = \mu + a_i + e_{ij}$. We have a similar general model for Model II ANOVA, where $Y_{ij} = \mu + A_i + e_{ij}$. The only change in the Model II is A_i, the random effects component, in substitution of the treatment effects component a_i. This distinction is critical, and leads us why we must interpret Model I and II ANOVAS in clearly different manners.

Interpretation of Model I and Model II ANOVA

With Model I ANOVA, if we have significant treatment effects (our F value is high), we can conclude that the different means were *caused* by our treatments—if our experiment were designed correctly. Others may repeat our experiment, verify our conclusions, and we have obtained what may be called an *explanation* of the differences in means, in terms of our treatments.

With a Model II ANOVA, we have no treatment effects, only random effects—those beyond the control of the investigator. If we have significant random effects (sometimes called added variance components), we must try to explain them, which often leads to the need for additional research rather than secure conclusions. For example, consider the following data presented in Table 10.9 for sherd thickness of the ceramic types listed below.

Table 10.9. Summary information for the sherd thickness of three Southwestern pottery types.

Pottery Type	n	mean	standard deviation
Wingate Polychrome	25	5.6	.661
Tularosa B/W	25	5	.924
St. John's Polychrome	25	5.8	.707

These three ceramic types are quite prominent in archaeological sites in western New Mexico and Eastern Arizona in the 1100 and 1200's. We know that they are different in one way, two are polychromes, while the third is a bichrome. Intrigued that they co-occur so commonly, we may ask the question, "Do these ceramics serve similar or different functions?" If similar, we would expect there to be no significant differences in mean ceramic width. If dissimilar, we expect there to be differences. Our null hypothesis is $H_o : \mu_{wingate} = \mu_{tularosa} = \mu_{st. john's}$, our alternative hypothesis is $H_a : \mu_{wingate} \neq \mu_{tularosa} \neq \mu_{st. john's}$.

Table 10.10 presents the results of our Model II ANOVA.

Table 10.10. ANOVA analysis comparing the sherd thickness of various pottery types.

Source of Variation	Df	SS	MS	F_s
$\bar{Y} - \bar{\bar{Y}}$ among group	2	8.67	4.33	7.26

$Y - \bar{Y}$ within group	72	43.00	.597	
$Y - \bar{\bar{Y}}$ total	74	51.67		

Inspection of Appendix C tells us that our F ratio is significant as $p < .001$, and there is a significant difference among the means. We therefore reject $H_o : \mu_{wingate} = \mu_{tularosa} = \mu_{st. john's}$, and conclude $H_a : \mu_{wingate} \neq \mu_{tularosa} \neq \mu_{st. john's}$.

Inspecting the means and standard deviations of the data, we conclude that most of the among group variation is a result of the small mean value for the Tularosa ceramics. The formal presentation of this conclusion should ideally present a construction of confidence intervals around these means, with the appropriate alpha level as presented in Table 9.3, and an assessment of the power of the test in order to support the ANOVA.

With respect to interpretation, we now must seek to explain the random effects. Why are Tularosa sherds thinner? Addressing this question will likely lead to the construction of a number of other hypotheses that remain to be tested; one of the consequences of sources of variation being beyond the control of the investigator.

We now move to a related procedure, regression.