

10.3: The One-Way ANOVA Formula

There are two main processes in testing a hypothesis stating that the means of three or more groups are not all equal:

1. An ANOVA omnibus test and
2. A post-hoc test.

Omnibus means containing or representing many things at once. The first process, the omnibus test, is where the ANOVA formula is used. The omnibus test performed using the ANOVA formula is a blunt test that can tell us whether, either, all group means are approximately equal or at least one group mean is significantly different than at least one other group mean. In this part of the process, the ANOVA formula considers data from all independent groups together. When the result of the ANOVA omnibus test is significant, it indicates that not all group means are equal to each other. This test alone is sufficient to testing whether a non-directional hypothesis is or is not supported.

When an omnibus test is significant and a hypothesis is directional, additional analyses are needed to reveal which groups are significantly different from which other groups. This second round of analysis is known as post-hoc testing. In post-hoc testing, two group means are compared at a time to determine which group(s) are significantly different from which other group(s). Comparing two groups at a time is referred to as a *pairwise comparison*. However, when the result of the omnibus test is not significant, it indicates that group means are approximately equal to one another and that no further testing is warranted. Thus, post hoc tests are performed after the result of an ANOVA omnibus test is determined to be statistically significant and when a hypothesis was directional. We will tackle these in their appropriate order; therefore, we must start with the ANOVA omnibus formula and how to use it before moving on to post-hoc testing.

Components of the ANOVA Formula

The omnibus test is performed using the formula for a one-way ANOVA. Before we can use the formula, it is important to understand what it can tell us and how it gets there. The one-way ANOVA formula has two main parts. The numerator focuses on difference between groups and the denominator focuses on differences within groups. The obtained value is indicated with the symbol F . F tells us the ratio of difference between groups relative to difference within groups. Another way to say this is that it tells us how different the groups are from one another after taking into account how different members of each group are from other members of their own group. The goal is to see whether groups are more distinct from one another than individuals are from members within groups.

The variation between groups can be thought of as system differences and/or differences caused by the independent variable in an experiment plus some error. The error is the random, non-systemic and/or non-treatment related differences that occur; this random error can be estimated by calculating the variation among members within groups. Thus, the variation within groups can be thought of as noise we want to account for so we can see the how much variation is systemic and non-random. Note that the systematic variation is the estimate of how much a dependent variable was affected by an independent variable when a true experiment is used. What we want to know is how much of the variability between groups is systemic. Therefore, we can understand the formula's main construction and outcomes as follows:

$$F = \frac{\text{variation between groups}}{\text{variation within groups}} = \text{ratio of systemic differences between groups}$$

The denominator of the formula can be referred to as the error term. In ANOVA, an **error term** is a calculation of the amount of variation that is random/non-systemic (or which is not estimated to be caused by the independent variable in a true experiment).

Unfortunately, the formula looks simpler than it is so we will look at it first then expand it out to understand its two main parts and how to calculate them. The one-way ANOVA formula is as follows:

$$F = \frac{MSS_b}{MSS_w}$$

The numerator asks for the mean sum of squares between groups (MSS_b). The denominator asks for the mean sum of squares within groups (MSS_w). Calculating each of these requires that we first calculate an SS and a df . This is because $MSS = SS \div df$ so the formula can be rewritten as follows:

$$F = \frac{SS_b \div df_b}{SS_w \div df_w}$$

The denominator of the ANOVA formula

The parts of the denominator are more familiar so we will start with those. The two components needed are SS_w and df_w .

Sum of Squares Within (SS_w)

SS_w is the sum of squared deviations within the group (also known simply as Sum of Squares Within); this has its own formula which we must use to calculate the sum of squares within each independent group. The formula for this is the same one we used in Chapter 4 in route to finding standard deviations and variances. The formula is as follows:

$$SS_w = \Sigma(x - \bar{x})^2$$

The steps to findings the SS_w for each independent group are:

1. Find the mean.
2. Subtract the mean from each raw score to find each deviation.
3. Square each deviation.
4. Sum the squared deviations

Each group has its own SS_w and we need to put them all together for ANOVA. Therefore, we must find SS_w for each group and add them together to use in the formula. If we had three groups, this could be summarized as follows:

$$SS_w = SS_{w1} + SS_{w2} + SS_{w3}$$

Degrees of Freedom Within (df_w)

df_w is the degrees of freedom within the group; this has its own formula which we must use. The formula is as follows:

$$df_w = N - k$$

The steps to findings the df_w are:

1. Find the total sample size (N) by adding the sample sizes (n) of all groups together.
2. Find k which refers to the number of independent groups or factors for the qualitative, grouping variable.
3. Subtract the number of groups (k) from the sum of sample sizes (N).

The numerator of the ANOVA formula

The two components needed for the numerator are SS_b and df_b . Each of these has its own formula which we must use. The parts of the numerator are less familiar because we have not seen them in any prior chapter of this book. Therefore, be sure to take extra care in reviewing these new formulas.

Sum of Squares Between (SS_b)

SS_b is the sum of squared deviations between groups (also known simply as Sum of Squares Between). This must be calculated for each group before they can be added together. The formula for calculating SS_b for each independent group is as follows:

$$SS_{b_group} = n_{group} \left[(\bar{x}_{group} - \bar{x}_{grand})^2 \right]$$

We have some new pieces here so let's review all of them. The parts of the formula and their translations are as follows:

n_{group} : the sample size for a group

\bar{x}_{group} : the mean for a group

\bar{x}_{grand} : the mean when all data for all groups are treated as one grand group

The group sample sizes and means are the same ones we have used in prior chapters, however, the grand mean has not appeared in a prior chapter. The grand mean (\bar{x}_{grand}) is the mean for all scores together, regardless of their individual group memberships. To find the grand mean, we must sum all the raw scores for all groups to get something known as the grand total or G . Then, we divide G by the total sample size which is N . The formula for the grand mean, therefore, is as follows:

$$\bar{x}_{grand} = \frac{G}{N}$$

The steps to findings the SS_b for each independent group are:

1. Find the grand mean.
2. Subtract the grand mean from the group mean to find the deviation between these two means.
3. Square the deviation between the means.
4. Multiply the squared deviation by the size of the group.

Each group has its own SS_b and we need to put them all together for ANOVA. Therefore, we must find SS_b for each group and add them together to use in the formula. If we had three groups, this could be summarized as follows:

$$SS_b = SS_{b1} + SS_{b2} + SS_{b3}$$

The formula for sum of squared deviation between can also be written showing the steps per group and the step to sum those group values together in one formula as follows:

$$SS_b = \sum n_i \left[(\bar{x}_i - \bar{x}_{\text{grand}})^2 \right]$$

The subscript i stands in for the names of all groups being tested such that the computations should each be computed for Group 1, then Group 2, then Group 3 and so on until computations for all groups have been completed. Thus, what this is saying is that, if you want to know the overall SS_b , you need to first find SS_b for each group and then sum those to get the overall SS_b .

You may have noticed some differences between the sum of squares within and the sum of squares between. Let's take a moment to consider their similarities and differences. They both find deviations and square those deviations. However, the within calculations focus on individual scores versus their group mean whereas between calculations focus on the group versus the grand group. Further, each individual in a group has their deviation overtly calculated in a within calculation (using $x - \bar{x}$); however, deviation is calculated at the group level in the between calculations (using $\bar{x}_{\text{group}} - \bar{x}_{\text{grand}}$). The group mean is representing all members of the group so the sample size needs to be taken into account to address this. This is why the squared group deviations in the SS_b formula are multiplied by their sample sizes. This adds weight to the squared deviation for each group proportional to the number of scores being represented by the group means.

Degrees of Freedom Between (df_b)

df_b is the degrees of freedom between the groups; this has its own formula which we must use. This is a new version of df . The formula is as follows:

$$df_b = k - 1$$

The steps to findings the df_b are:

1. Find k which refers to the number of independent groups or factors for the qualitative, grouping variable.
2. Subtract 1 from k .

df_b is always the number of groups minus 1. For example, if three independent groups were being compared, the df_b would be 2 but if four independent groups were being compared the df_b would be 3, and so on.

Putting the Formula Together

Once the four components are calculated, their results are put into the ANOVA formula and used to solve for F .

$$F = \frac{SS_b \div df_b}{SS_w \div df_w} = \frac{MSS_b}{MSS_w}$$

Interpreting Obtained F -Values

Obtained F -values are always positive so only their magnitude, and not their direction, is interpreted. The magnitude represents the ratio of variation that is systematic and non-random relative to the amount of variation that is non-systematic and random. Remember that the difference within groups represents the error term and is calculated in the denominator. When the F -value is 1.00 it means there is as much difference between groups as there is within groups; this indicates that all the difference observed is just random and does not represent actual differences between groups. When this occurs, the null hypothesis is retained. Thus, the closer F is to 1.00, the less the difference observed is attributed to group by group (between) variation.

Consistent with this, when the difference between the groups is greater that the error term (i.e. differences within groups), the F -value will be greater than 1.00. The larger the F -value, the more the difference observed is attributed to group by group (between) variation. Another way to say this is that the larger the F -value, the greater the non-random differences are between groups and,

thus, the more evidence there is in support of the alternative hypothesis and against the null hypothesis. When the F -value is large enough to surpass the critical value, it means that the differences observed between groups (after accounting for difference within groups) is unlikely to be due to chance (i.e. it is unlikely to be random). When this occurs, the result can be declared statistically significant.

Reading Review 10.2

1. What is being calculated and represented by the numerator of the one-way ANOVA formula?
2. What is being calculated and represented by the denominator of the one-way ANOVA formula?
3. What is a grand mean and how is it calculated?

This page titled [10.3: The One-Way ANOVA Formula](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by .