

## 11.2: Introduction to ANOVA's Sum of Squares

ANOVA is all about looking at the different sources of variance (i.e. the reasons that scores differ from one another) in a dataset. Fortunately, the way we calculate these sources of variance takes a very familiar form: the Sum of Squares.

### Sums of Squares

#### Between Groups Sum of Squares

One source of variability we can identify in § 11.1.3 of the previous example was differences or variability *between* the groups. That is, the groups seemed to have different average levels. The variability arising from these differences is known as the between groups variability, and it is quantified using Between Groups Sum of Squares.

Our calculations for sums of squares in ANOVA will take on the same form as it did for regular calculations of variance. Each observation, in this case the group means, is compared to the overall mean, in this case the grand mean, to calculate a deviation score. These deviation scores are squared so that they do not cancel each other out and sum to zero. The squared deviations are then added up, or summed. There is, however, one small difference. Because each group mean represents a group composed of multiple people, before we sum the deviation scores we must multiply them by the number of people within that group. Incorporating this, we find our equation for Between Groups Sum of Squares to be:

$$SS_B = \sum_{EachGroup} \left[ \left( \bar{X}_{group} - \bar{X}_T \right)^2 * (n_{group}) \right]$$

1. Subtract
2. Square
3. Multiply
4. Sum

I know, this looks a little extreme, but it really is what it says that it is, subtracting the mean of the total of all participants ( $\bar{X}_T$ ) from the mean of one of the groups ( $\bar{X}_{group}$ ), then squaring that subtraction. That gives you the difference score for that group. You then multiply that by the sample size for that group ( $n_{group}$ ). You do that for each group ( $\sum_{EachGroup}$ ). For example, if you had an IV with three levels ( $k = 3$ ) that were High, Medium, and Low, then you'd do the parts in the brackets for each group, then add all three groups together for one final Between Groups Sum of Squares.

The only difference between this equation and the familiar sum of squares for variance is that we are adding in the sample size. Everything else logically fits together in the same way.

#### Within Groups Sum of Squares (Error)

The formula for this within groups sum of squares is again going to take on the same form and logic. What we are looking for is the distance between each individual person and the mean of the group to which they belong. We calculate this deviation score, square it so that they can be added together, then sum all of them into one overall value:

$$SS_W = \sum_{EachGroup} \left[ \sum \left( (X - \bar{X}_{group})^2 \right) \right]$$

1. Subtract
2. Square
3. Sum

In this instance, because we are calculating this deviation score for each individual person, there is no need to multiply by how many people we have. It is important to remember that the deviation score for each person is only calculated relative to their group mean; this is what  $(X - \bar{X}_{group})$  is telling you to do: subtract the mean of a group from each score from that same group. You then square each of those subtractions. The sum ( $\sum$ ) is to sum all of the individual squared scores of all of the groups.

#### Total Sum of Squares

The calculation for this score is exactly the same as it would be if we were calculating the overall variance in the dataset (because that's what we are interested in explaining) without worrying about or even knowing about the groups into which our scores fall:

$$SS_T = \sum \left[ \left( X - \bar{X}_T \right)^2 \right]$$

1. Subtract
2. Square
3. Sum

We can see that our Total Sum of Squares is just each individual score minus the grand mean (the mean of the totality of scores,  $\bar{X}_T$ ). As with our Within Groups Sum of Squares, we are calculating a deviation score for each individual person, so we do not need to multiply anything by the sample size; that is only done for Between Groups Sum of Squares.

### Computation Check!

An important feature of these calculations in ANOVA is that they all fit together. We could work through the algebra to demonstrate that if we added together the formulas for  $SS_B$  and  $SS_W$ , we would end up with the formula for  $SS_T$ . That is:

$$SS_T = SS_B + SS_W$$

This will prove to be a very convenient way to check your work! If you calculate each  $SS$  by hand, you can make sure that they all fit together as shown above, and if not, you know that you made a math mistake somewhere.

### By Hand?

We can see from the above formulas that calculating an ANOVA by hand from raw data can take a very, very long time. For this reason, you will rarely be required to calculate the  $SS$  values by hand. Many professors will have you work out one problem on your own by hand, then either show you how to use statistical software or provide the Sums of Squares for you. However, you should still take the time to understand how they fit together and what each one represents to help understand the analysis itself; this will make it easier to interpret the results and make predictions (research hypotheses).

## Contributors and Attributions

- 

Dr. MO (Taft College)

---

This page titled [11.2: Introduction to ANOVA's Sum of Squares](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michelle Oja](#).