

## 8.2: Partitioning the Sums of Squares

Time to introduce a new name for an idea you learned about last chapter, it's called **partitioning the sums of squares**. Sometimes an obscure new name can be helpful for your understanding of what is going on. ANOVAs are all about partitioning the sums of squares. We already did some partitioning in the last chapter. What do we mean by partitioning?

Imagine you had a big empty house with no rooms in it. What would happen if you partitioned the house? What would you be doing? One way to partition the house is to split it up into different rooms. You can do this by adding new walls and making little rooms everywhere. That's what partitioning means, to split up.

The act of partitioning, or splitting up, is the core idea of ANOVA. To use the house analogy. Our total sums of squares ( $SS_{\text{Total}}$ ) is our big empty house. We want to split it up into little rooms. Before we partitioned  $SS_{\text{Total}}$  using this formula:

$$SS_{\text{TOTAL}} = SS_{\text{Effect}} + SS_{\text{Error}}$$

Remember, the  $SS_{\text{Effect}}$  was the variance we could attribute to the means of the different groups, and  $SS_{\text{Error}}$  was the leftover variance that we couldn't explain.  $SS_{\text{Effect}}$  and  $SS_{\text{Error}}$  are the partitions of  $SS_{\text{TOTAL}}$ , they are the little rooms.

In the between-subjects case above, we got to split  $SS_{\text{TOTAL}}$  into two parts. What is most interesting about the repeated-measures design, is that we get to split  $SS_{\text{TOTAL}}$  into three parts, there's one more partition. Can you guess what the new partition is? Hint: whenever we have a new way to calculate means in our design, we can always create a partition for those new means. What are the new means in the repeated measures design?

Here is the new idea for partitioning  $SS_{\text{TOTAL}}$  in a repeated-measures design:

$$SS_{\text{TOTAL}} = SS_{\text{Effect}} + SS_{\text{Subjects}} + SS_{\text{Error}}$$

We've added  $SS_{\text{Subjects}}$  as the new idea in the formula. What's the idea here? Well, because each subject was measured in each condition, we have a new set of means. These are the means for each subject, collapsed across the conditions. For example, subject 1 has a mean (mean of their scores in conditions A, B, and C); subject 2 has a mean (mean of their scores in conditions A, B, and C); and subject 3 has a mean (mean of their scores in conditions A, B, and C). There are three subject means, one for each subject, collapsed across the conditions. And, we can now estimate the portion of the total variance that is explained by these subject means.

We just showed you a "formula" to split up  $SS_{\text{TOTAL}}$  into three parts, but we called the formula an idea. We did that because the way we wrote the formula is a little bit misleading, and we need to clear something up. Before we clear the thing up, we will confuse you just a little bit. Be prepared to be confused a little bit.

First, we need to introduce you to some more terms. It turns out that different authors use different words to describe parts of the ANOVA. This can be really confusing. For example, we described the SS formula for a between subjects design like this:

$$SS_{\text{TOTAL}} = SS_{\text{Effect}} + SS_{\text{Error}}$$

However, the very same formula is often written differently, using the words between and within in place of effect and error, it looks like this:

$$SS_{\text{TOTAL}} = SS_{\text{Between}} + SS_{\text{Within}}$$

Whoa, hold on a minute. Haven't we switched back to talking about a **between-subjects** ANOVA. YES! Then why are we using the word **within**, what does that mean? YES! We think this is very confusing for people. Here the word **within** has a special meaning. It **does not** refer to a within-subjects design. Let's explain. First,  $SS_{\text{Between}}$  (which we have been calling  $SS_{\text{Effect}}$ ) refers to variation **between** the group means, that's why it is called  $SS_{\text{Between}}$ . Second, and most important,  $SS_{\text{Within}}$  (which we have been calling  $SS_{\text{Error}}$ ), refers to the leftover variation within each group mean. Specifically, it is the variation between each group mean and each score in the group. "AAGGH, you've just used the word between to describe within group variation!". Yes! We feel your pain. Remember, for each group mean, every score is probably off a little bit from the mean. So, the scores within each group have some variation. This is the within group variation, and it is why the leftover error that we can't explain is often called  $SS_{\text{Within}}$ .

OK. So why did we introduce this new confusing way of talking about things? Why can't we just use  $SS_{\text{Error}}$  to talk about this instead of  $SS_{\text{Within}}$ , which you might (we do) find confusing. We're getting there, but perhaps a picture will help to clear things up.

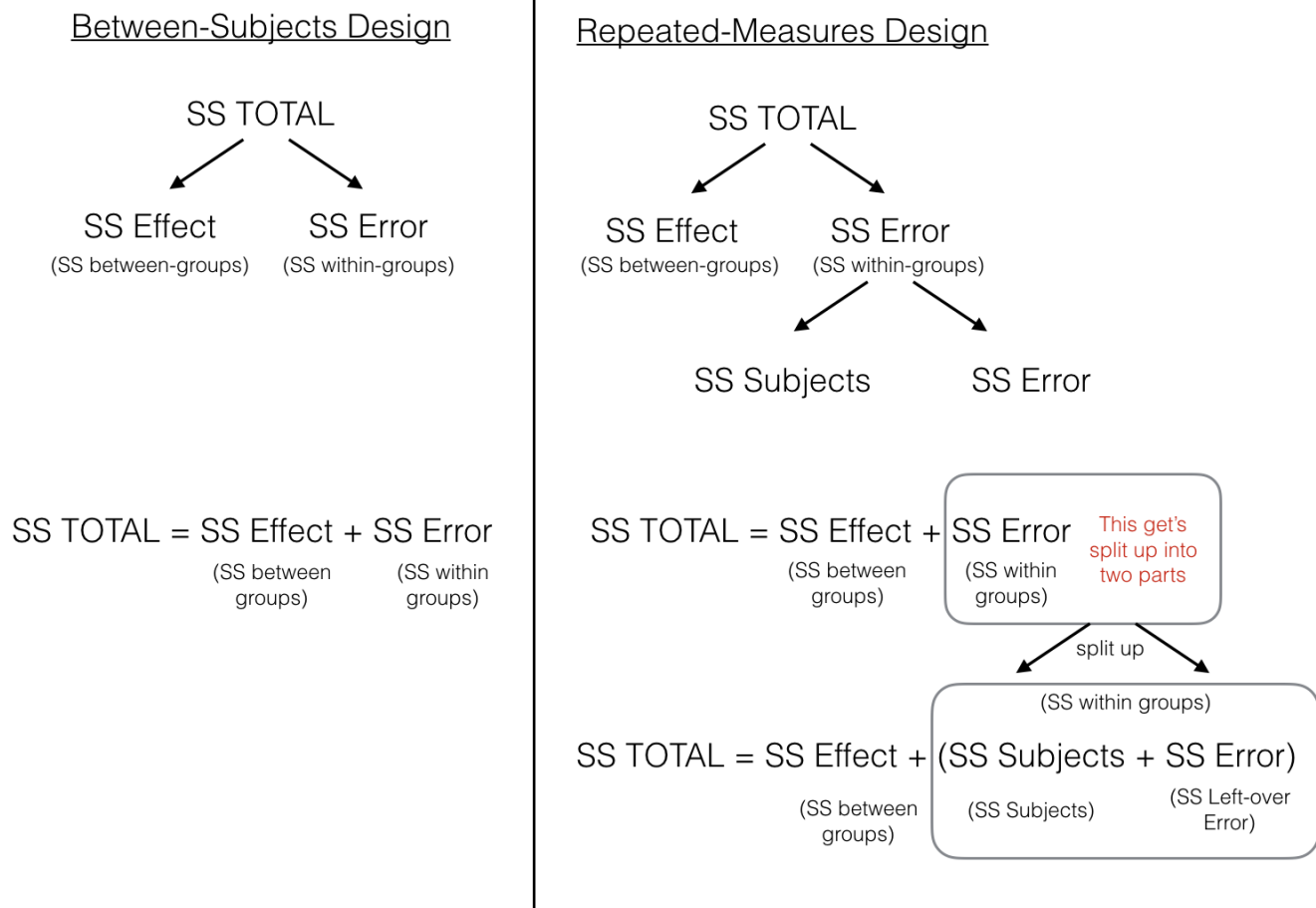


Figure 8.2.1: Illustration showing how the total sums of squares are partitioned differently for a between versus repeated-measures design.

The figure lines up the partitioning of the Sums of Squares for both between-subjects and repeated-measures designs. In both designs,  $SS_{Total}$  is first split up into two pieces  $SS_{Effect}$  (between-groups) and  $SS_{Error}$  (within-groups). At this point, both ANOVAs are the same. In the repeated measures case we split the  $SS_{Error}$  (within-groups) into two more littler parts, which we call  $SS_{Subjects}$  (error variation about the subject mean) and  $SS_{Error}$  (left-over variation we can't explain)

So, when we earlier wrote the formula to split up SS in the repeated-measures design, we were kind of careless in defining what we actually meant by  $SS_{Error}$ , this was a little too vague:

$$SS_{TOTAL} = SS_{Effect} + SS_{Subjects} + SS_{Error}$$

The critical feature of the repeated-measures ANOVA, is that the  $SS_{Error}$  that we will later use to compute the MSE in the denominator for the  $F$ -value, is smaller in a repeated-measures design, compared to a between subjects design. This is because the  $SS_{Error}$  (within-groups) is split into two parts,  $SS_{Subjects}$  (error variation about the subject mean) and  $SS_{Error}$  (left-over variation we can't explain)

To make this more clear, we made another figure:

## Repeated-Measures Design

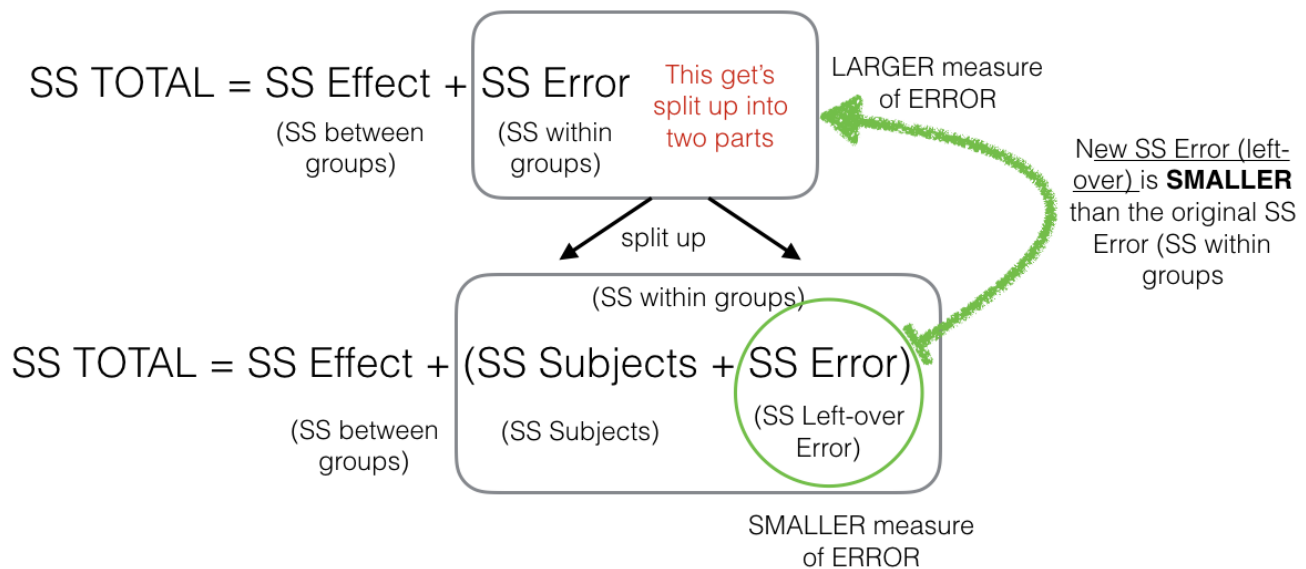


Figure 8.2.2: Close-up showing that the Error term is split into two parts in the repeated measures design.

As we point out, the  $SS_{\text{Error (left-over)}}$  in the green circle will be a smaller number than the  $SS_{\text{Error (within-group)}}$ . That's because we are able to subtract out the  $SS_{\text{Subjects}}$  part of the  $SS_{\text{Error (within-group)}}$ . As we will see shortly, this can have the effect of producing larger F-values when using a repeated-measures design compared to a between-subjects design.

This page titled [8.2: Partitioning the Sums of Squares](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Matthew J. C. Crump](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.