



# CHAPTER W2

# REPEATED-MEASURES ANALYSIS OF VARIANCE

## Chapter Outline

- Basic Logic
- Figuring a One-Way Repeated-Measures Analysis of Variance
- Assumptions and Related Issues
- More Complex Repeated-Measures Designs and Mixed Designs
- Key Terms

### **Are You Ready? What You Need to Have Mastered Before Starting This Chapter**

*Chapters 1-2, and 5-13 (it is especially important to have mastered the procedures in chapter 13 for working out a two-way analysis of variance)*

In all the analysis of variance situations we considered in Chapters 11 through 13, the different cells or groupings were based on scores from different individuals. Sometimes, however, a researcher measures the same individual in several different situations. (If there are only two such conditions, such as before and after some treatment, you can use a  $t$  test for dependent means, as described in Chapter 9.) Consider a study in which participants' task is to recognizing a syllable when embedded in three word types flashed quickly on the screen: familiar words, unfamiliar words, and nonword sounds. The result is that for each participant you have an average number of errors for each word type. Or suppose you do a study of psychotherapy effects testing patients on their depression before, immediately following, and again 3 months after therapy. In both examples, you have three groups of scores, but all three scores are from the same people. These studies each employ a **repeated-measures design**.

Repeated measures designs are analyzed with a **repeated-measures analysis of variance**. It has this name because the same participants are being measured repeatedly. This kind of design and analysis is also called a **within-subjects design** and **within-subjects analysis of variance** because the comparison is within, not between, the different participants or subjects. In this chapter we consider the logic and show you how to figure out a one-way repeated measures analysis of variance, consider the assumptions for such analyses and how researchers handle situations

where the assumptions might not be met, and then describe briefly how to expand the basic one-way repeated measures analysis of variance to factorial repeated measures analysis of variance and to analyses of variance that include both repeated measures factors and the usual between-subjects factors.

## BASIC LOGIC

Ironically, you carry out a *one-way* repeated-measures analysis of variance using the procedures for a standard *two-way* between-subjects analysis of variance. But this standard two-way analysis is modified in one crucial way. Just as with an ordinary two-way analysis of variance, you treat your groupings (conditions) as one factor, usually as columns. The modification is that the rows, instead of another factor, are participants. That is, there is one participant per row, and each participant has a score in each column. (See the example in the top of Table W2-1.) Thus, one factor is conditions and the other factor is participants; the cells have only one score each. Your main effect for conditions, as in any analysis of variance, uses the population variance estimate based on the deviation of each score's condition's mean from the grand mean. However, what changes is the population variance estimate that is the denominator of the *F* ratio.

**TABLE W2-1** Scores and Scores Minus Participant's Mean for Four Participants Each Exposed to Target Syllable in Familiar, Unfamiliar, and Nonword-Sound Word Types (Fictional Data)

### Actual Scores:

*Measure: Number of errors*

Participant	Word Type		
	<i>Familiar Word</i>	<i>Unfamiliar Word</i>	<i>Nonword Sound</i>
A	9	3	0
B	6	2	1
C	11	6	4
D	10	5	3

### Scores minus participant's mean:

*Measure: Number of errors*

Participant	Word Type		
	<i>Familiar Word</i>	<i>Unfamiliar Word</i>	<i>Nonword Sound</i>
A	5	-1	-4
B	3	-1	-2
C	4	-1	-3
D	4	-1	-3

In a repeated-measures analysis of variance you can not use the within-cell variation as your denominator for the  $F$  ratio. One problem is that there is no within-cell variation! There is only one score per cell, and there cannot be any variation within a single score. An alternative that might seem reasonable on first glance would be to use the average variation within each condition, as you would do in an ordinary one-way between-subjects analysis of variance. This, however, ignores the fact that the scores in each condition are from the same participants. Ignoring this is a problem for technical reasons (it violates the requirement that all scores should be independent). More important, however, is that doing so would not take advantage of the fact that you are interested in differences among conditions *over and above individual overall differences among participants*. So how do you take this repeated-measures nature of the study into account?

One solution that takes the repeated-measures nature of the study into account would be to use for each score, not the actual score, but its difference from that participant's mean across conditions. For example, consider the scores in the top part of Table W2-1 showing the number of errors for four participants who each were exposed to the syllable embedded in 30 presentations each of Familiar Words, Unfamiliar Words, and Nonword Sounds. Participant A's average is 4 errors (that is,  $9 + 3 + 0$  divided by 3 comes out to 4). Thus, A's errors in the Familiar Word condition are 5 above A's average, A's errors in the Unfamiliar Word condition are 1 below A's average, and A's errors in the Nonword Sound condition are 4 below A's average. The bottom part of Table W2-1 shows these scores (each score minus the participant's own mean) for all four participants.

Suppose you were to figure an ordinary between-subjects one-way analysis of variance using these difference-from-the-participant's-own-mean scores? Your results would come out in a way that takes into account, and properly takes advantage of, the fact that the same participants are in each condition. You can see that there is much less variance within conditions in this lower table than in the original scores in the upper table. This is because we have eliminated the variation due to overall between person tendencies—everyone's scores only vary from their own mean so that any variation among different participants' means is eliminated. The result of there being less variance is that the within-group population variance is smaller, making the overall  $F$  larger. (This is the same principle we saw in Chapter 9 with the  $t$  test for dependent means that a repeated measures design in general has more power. Now you can see the reason for this: the repeated-measures design removes the variance due to individual overall differences among participants.)

If you were actually to carry out a standard between-subjects one-way analysis of variance with these scores in the bottom table (that is, taking the conditions as groups and figuring the denominator of the  $F$  ratio based on the variation within conditions) you would have to correct the degrees of freedom for the fact that you are subtracting out mean scores. You would then get the correct results. But it would also be a tedious process in a larger analysis because you would have to do all these subtractions of scores from each participants' means. This method also gets complicated in more complex designs.

Thus, in practice, researchers use a slightly different approach that accomplishes the same thing. When figuring the denominator of the  $F$  ratio in a repeated-measures analysis like this, they take advantage of the two-way set up and use the interaction of participants by conditions. This interaction, like any interaction, is the difference in pattern across one factor according to the level of the other factor. In this situation it means the average variation in pattern across conditions for different participants. Since it is looking at patterns, it is in effect subtracting out each participant's mean.

In fact, using this interaction term gives exactly the same result as using the scores based on differences from each participant's mean we described in the preceding paragraph (after adjusting the degrees of freedom appropriately).

Here is another way to think of this participant  $\times$  condition interaction. Like any interaction, its sum of squared deviations is the total squared deviations from the grand mean minus all the other deviations—the deviations within cells and the deviations for each of the factors. In this situation, however, the deviations within cells (the deviation of each score from its cell's mean) are all 0, because there is only one score per cell. This means that in addition to subtracting out the deviations for the main effect for conditions (each score's condition's mean minus the grand mean), you are *also* subtracting out the deviations of each score's participant's mean from the grand mean. That is, by subtracting out the differences among participant's means, you are subtracting out the overall differences among participants, which is as we have seen what gives such an advantage to repeated-measures designs.

In sum, you can think of a one-way repeated measures analysis of variance across conditions as a two-way analysis of variance in which one factor is experimental condition and the other factor is participant, and in which your between-conditions population variance estimate is as usual based on the deviations of each score's condition mean from the grand mean, but in which the within-groups population variance estimate is based only on the deviation of scores within conditions after adjusting for each participant's overall mean.

## FIGURING A ONE-WAY REPEATED MEASURES ANALYSIS OF VARIANCE

---

Following the logic we have just considered, you figure a one-way repeated measures analysis of variance by setting it up as a condition  $\times$  participant two-way design, figuring the two main effects and interaction in the usual way, and testing the condition main effect with an  $F$  ratio of the between-condition population variance estimate divided by the interaction population variance estimate.

Here are the steps:

- ❶ **Set up the scores with groupings across the top** (the conditions factor on which each participant has a score for each condition) **and a row for each participant.**
- ❷ **Figure the sums of squares in the usual way for total** (squared deviations of each score from the grand mean), **columns** (squared deviations of each score's column's mean from the grand mean), **and rows** (squared deviations of each score's row's mean from the grand mean). Notice that in figuring the sum of squares for rows, the row means are the same as each participant's mean score.
- ❸ **Figure the sum of squares for the interaction**, figuring the deviation for each score as its deviation from the grand mean minus the deviation of its column's mean from the grand mean and minus the deviation of its row's mean from the grand mean. (That is, you don't need also to subtract the scores deviation from its cell's mean because all cells only have one score so this deviation is always 0.)
- ❹ **Figure the degrees of freedom in the usual way for columns** (the number of columns minus 1), **rows** (the number of rows minus 1), **and interaction** (the number of cells minus the degrees of freedom for rows and columns, minus 1). Notice that the degrees of freedom for rows is the number of participants minus 1.

⑤ **Figure the mean squares** (the population variance estimates) **in the usual way for columns and the interaction.** For columns, this means the columns' sum of squares divided by the columns' degrees of freedom; for the interaction, this

**TABLE W2-2 One-Way Repeated Measures Analysis of Variance for Study of Errors in Recognition for Four Participants Each Exposed to Target Syllable in Familiar, Unfamiliar, and Nonword-Sound Word Types (Fictional Data)**

Measure: Number of errors

Participant	①	Word Type			Row	
		Familiar Word	Unfamiliar Word	Nonword Sound	$\Sigma$	$M$
A		9	3	0	12	4
B		6	2	1	9	3
C		11	6	4	21	7
D		10	5	3	18	6
$\Sigma$		36	16	8		
$M$		9	4	2		

$GM = 5$

Squared deviations from grand mean

Participant	Familiar Word				Unfamiliar Word				Nonword Sound			
	$X$	Column	Row	Int	$X$	Column	Row	Int	$X$	Column	Row	Int
A	16	16	1	1	4	1	1	0	25	9	1	1
B	1	16	4	1	9	1	4	0	16	9	4	1
C	36	16	4	0	1	1	4	0	1	9	4	0
D	25	16	1	0	0	1	1	0	4	9	1	0
$\Sigma$	78	64	10	2	14	4	10	0	46	36	10	2

$SS_{Total} = 78 + 14 + 46 = 138$

②  $SS_{Columns} = 64 + 4 + 36 = 104$

$SS_{Rows} = 10 + 10 + 10 = 30$

③  $SS_{Interaction} = 2 + 0 + 2 = 4$

Check:  $SS_{Total} = SS_{Columns} + SS_{Rows} + SS_{Interaction} = 104 + 30 + 4 = 138$

$df_{Total} = 12 - 1 = 11$

④  $df_{Columns} = 3 - 1 = 2$

$df_{Rows} = 4 - 1 = 3$

$df_{Interaction} = 11 - 2 - 3 = 6$

Check:  $df_{Total} = df_{Columns} + df_{Rows} + df_{Interaction} = 2 + 3 + 6 = 11$

$F$  needed for repeated measures (columns) effect ( $df 2, 6; p < .05$ ) : 5.14

**Analysis of Variance Table**

Source	SS	df	⑤ MS	F
Between Conditions (Columns)	104	2	52	77.6
Participants (Rows)	30	3		
Error (Interaction)	4	6	.67	
Total	138	11		

Conclusion: Reject the null hypothesis. ⑥

means the interaction's sum of squares divided by the interaction's degrees of freedom. Notice that you do not figure a population variance estimate based on rows. This would be the variance among participants, which is not usually of interest.

❶ **Figure the  $F$  ratio for the repeated-measures condition effect:** Divide the mean square for columns by the mean square for the interaction.

To test the significance of the repeated measures condition effect, you compare your  $F$  to a cutoff  $F$  based on the appropriate numerator (columns) and denominator (interaction) degrees of freedom.

Table W2-2 shows scores, figuring, and an analysis of variance table for the repeated-measures analysis of variance for the example we have been considering.

## ASSUMPTIONS AND RELATED ISSUES

The assumptions for a repeated-measures analysis of variance include the usual ones for an ordinary analysis of variance—normal distributions and equal variances of the populations for each condition or condition combination. There is also an additional assumption, called **sphericity**. Sphericity means that not only are the population variances for each condition the same, but the population correlations among the different conditions are the same. For example, if you have sphericity, the population correlation between errors in the Familiar Word condition and errors in the Unfamiliar Word condition will be the same as the correlation between errors in the Familiar Word condition and errors in the Nonword Sounds condition.

With the small sample size usually used in repeated-measures analysis of variance, it is hard to know whether you meet any of the assumptions. For example, regarding sphericity, correlations in small samples can vary quite considerably even if in the population they are equal and correlations in small samples could be quite similar even if in the population they are different. The difficulty of knowing whether you have met the sphericity assumption is especially a problem because the accuracy of the  $F$  test for repeated-measures analysis of variance can be strongly affected by violations of this assumption.

There are two widely used solutions for dealing with possible violations of the sphericity assumption. One solution is to estimate how much the assumption is violated and statistically correct for it by using a comparison  $F$  distribution with fewer degrees of freedom (so that you need a stronger result to be significant). Thus, when you do a repeated measures analysis of variance with most computer programs (including SPSS), in addition to the standard results, you get various results with corrected degrees of freedom and the resulting more conservative (less extreme) significance levels.

For example, Figure W2-1 shows a key section of the SPSS output for the example we have been considering. (The slight difference in the  $F$  figured in Table W2-2 is due to rounding error.) Notice that in addition to the standard output ("sphericity assumed"), you get three other results. These refer to different ways of correcting for violation of the sphericity assumption. The correction is to the degrees of freedom. The  $F$ s all come out the same, but since they are being compared to different  $F$  distributions, the  $p$  levels are not the same. In this present example, the uncorrected result (with  $df = 2, 6$ ) has a  $p$  that rounds off to .000. But the corrected results (with  $dfs = 1, 3$ ) are all about .003.

In most real data sets the three different corrected results (that is, their  $dfs$  and  $p$  values) also differ from each other. The "Lower-bound" correction, as the name suggests, is the most conservative, giving the highest  $p$  value. The "Greenhouse-

## Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
WORDTYPE	Sphericity Assumed	104.000	2	52.000	78.000	.000
	Greenhouse-Geisser	104.000	1.000	104.000	78.000	.003
	Huynh-Feldt	104.000	1.000	104.000	78.000	.003
	Lower-bound	104.000	1.000	104.000	78.000	.003
Error(WORDTYPE)	Sphericity Assumed	4.000	6	.667		
	Greenhouse-Geisser	4.000	3.000	1.333		
	Huynh-Feldt	4.000	3.000	1.333		
	Lower-bound	4.000	3.000	1.333		

**FIGURE W2-1** Key section of SPSS output for study of errors in recognition for four participants each exposed to target syllable in familiar, unfamiliar, and nonword-sound word types (fictional data)

Geisser” correction is the next most conservative. The **Huynh-Feldt correction** is the least conservative, but still widely (though not unanimously) agreed to be quite accurate. Thus, most psychologists focus on the Huynh-Feldt correction.

If the results using the Huynh-Feldt correction is not much different from the uncorrected (sphericity assumed) results, for simplicity, researchers usually just report the uncorrected results. However, if there is much difference between the corrected and uncorrected results, then the Huynh-Feldt corrected results are usually reported. In the present example, researchers would probably just report significance as  $p < .01$ . Another solution to the concern about violating the sphericity assumption is to use an entirely different approach, a special application of *multivariate analysis of variance* (see Chapter 17). Thus, most computer printouts for a repeated-measures analysis of variance also automatically include results of multivariate tests. The results using these tests are typically similar to those using the standard procedures. (In the word-type example we have been considering, the  $p$  for the multivariate tests is .003). Most psychologists use the standard procedures, although there is a strong minority opinion in favor of the multivariate approach. The advantage of the multivariate approach is that you don’t have to make the sphericity assumption. On the other hand, with small samples, the multivariate approach usually has lower power and you get around the sphericity assumption in any case by using the Huynh-Feldt correction.

## MORE COMPLEX REPEATED-MEASURES DESIGNS AND MIXED DESIGNS

So far we considered the one-way repeated measures analysis of variance, in which each of several participants are tested under different conditions. In this design, there is one repeated-measures variable, such as Word Type. Sometimes, however, a study includes two or more repeated-measures variables. For example, Word Type could be crossed with length of the words, so that each participant might have scores in six conditions, Familiar Long Words, Familiar Short Words, Unfamiliar Long Words, and so forth. This would be an example of a **two-way repeated-measures design**. The repeated-measures two-way analysis of variance would be set up like a three-way between-subjects analysis of variance, with one dimension for Word Type, one dimension for Word Length, and one dimension for Participants. The denominator of the  $F$  ratio for each of the two condition main effects (Word

Type and Word Length) and their interaction (Word Type  $\times$  Word Length) would in each case be the interaction of that term with the Participant factor. For example, you would test the Word Length main effect by figuring an  $F$  ratio in which the numerator is the population variance estimate based on the Word Length mean squares and the denominator is the population variance estimate based on the Word Length  $\times$  Participant interaction mean squares.

It is also possible to have in the same analysis both repeated-measures and ordinary between-subjects factors. For example, in the word recognition study each participant might have only the three Word Types, but some participants might be college age and other participants might be in their 70s. Thus, you would have one repeated-measures factor (Word Type) and one between-subjects factor (Age). This would be an example of a **mixed design**. In this example, the analysis of variance allows you to test the main effect for the repeated measures factor (Word Type), the main effect for the between-subject's factor (Age) and the interaction (Word Type  $\times$  Age).

The actual figuring of a mixed design is somewhat complicated, but two main points are of interest. First, the between-subjects part of the analysis would come out the same as if you were doing a one-way analysis for the between subjects factor using each participant's mean across the conditions. That is, in the example, it would be the same result as if you just had two levels of age and the scores within each were each participant's mean number of errors across the three words types. Second, the denominator of the  $F$  ratio for the repeated-measures main effect and the interaction is based on the interaction of participant with the repeated-measures factor, figured separately within each level of the between-subjects factor (for each age group) and averaged. That is, in the example, the main effect for Word Type and the Word Type  $\times$  Age interaction would both be figured using as the denominator the Word Type  $\times$  Participant interaction figured separately within each age group and then averaged. Thus, the computer output from a mixed analysis of variance, gives two sets of results: One for the between-subjects and one part for the repeated-measures aspect.

---

## KEY TERMS

Huynh-Feldt correction  
mixed design  
repeated-measures analysis of variance  
repeated-measures design  
sphericity assumption  
two-way repeated-measures design  
within-subjects analysis of variance  
within-subjects design