15.8: Within-Subjects

Learning Objectives

- Explain why a within-subjects design can be expected to have more power than a between-subjects design
- Explain error in terms of interaction
- Be able to create the Source and df columns of an ANOVA summary table for a design with one between-subjects and one within-subjects variable
- Describe the consequences of violating the assumption of sphericity
- Discuss courses of action that can be taken if sphericity is violated

Within-subjects factors involve comparisons of the same subjects under different conditions. For example, in the "ADHD Treatment" study, each child's performance was measured four times, once after being on each of four drug doses for a week.Therefore, each subject's performance was measured at each of the four levels of the factor "Dose." Note the difference from between-subjects factors for which each subject's performance is measured only once and the comparisons are among different groups of subjects. A within-subjects factor is sometimes referred to as a repeated-measures factor since repeated measurements are taken on each subject. An experimental design in which the independent variable is a within-subjects factor is called a within-subjects design.

Advantage of Within-Subjects Designs

One-Factor Designs

Let's consider how to analyze the data from the "ADHD Treatment" case study. These data consist of the scores of 24 children with ADHD on a delay of gratification (DOG) task. Each child was tested under four dosage levels. For now, we will be concerned only with testing the difference between the mean in the placebo condition (the lowest dosage, $D0$) and the mean in the highest dosage condition ($D60$). The details of the computations are relatively unimportant since they are almost universally done by computers. Therefore we jump right to the ANOVA Summary table shown in Table $15.8.1$.

Table 15.8.1: ANOVA Summary Table

Source	df	SSQ	MS	F	p
Subjects	23	5781.98	251.39		
Dosage	1	295.02	295.02	10.38	0.004
Error	23	653.48	28.41		
Total	47	6730.48			

The first source of variation, "Subjects," refers to the differences among subjects. If all the subjects had exactly the same mean (across the two dosages), then the sum of squares for subjects would be zero; the more subjects differ from each other, the larger the sum of squares subjects.

Dosage refers to the differences between the two dosage levels. If the means for the two dosage levels were equal, the sum of squares would be zero. The larger the difference between means, the larger the sum of squares.

The error reflects the degree to which the effect of dosage is different for different subjects. If subjects all responded very similarly to the drug, then the error would be very low. For example, if all subjects performed moderately better with the high dose than they did with the placebo, then the error would be low. On the other hand, if some subjects did better with the placebo while others did better with the high dose, then the error would be high. It should make intuitive sense that the less consistent the effect of dosage, the larger the dosage effect would have to be in order to be significant. The degree to which the effect of dosage differs depending on the subject is the $Subjects \times Dosage$ interaction. Recall that an interaction occurs when the effect of one variable differs depending on the level of another variable. In this case, the size of the error term is the extent to which the effect of the variable "Dosage" differs depending on the level of the variable "Subjects." Note that each subject is a different level of the variable "Subjects."

Other portions of the summary table have the same meaning as in between-subjects ANOVA. The F for dosage is the mean square for dosage divided by the mean square error. For these data, the F is significant with $p = 0.004$. Notice that this F test is equivalent to the t test for correlated pairs, with $F = t^2$.

Table $15.8.2$ shows the ANOVA Summary Table when all four doses are included in the analysis. Since there are now four dosage levels rather than two, the df for dosage is three rather than one. Since the error is the $Subjects \times Dosage$ interaction, the df for error is the df for "Subjects" (23) times the df for Dosage (3) and is equal to 69.

Table 15.8.2: ANOVA Summary Table

Source	df	SSQ	MS	F	p
Subjects	23	9065.49	394.15		
Dosage	3	557.61	185.87	5.18	0.003
Error	69	2476.64	35.89		
Total	95	12099.74			

Carryover Effects

Often performing in one condition affects performance in a subsequent condition in such a way as to make a within-subjects design impractical. For example, consider an experiment with two conditions. In both conditions subjects are presented with pairs of words. In Condition A, subjects are asked to judge whether the words have similar meaning whereas in Condition B, subjects are asked to judge whether they sound similar. In both conditions, subjects are given a surprise memory test at the end of the presentation. If Condition were a within-subjects variable, then there would be no surprise after the second presentation and it is likely that the subjects would have been trying to memorize the words.

Not all carryover effects cause such serious problems. For example, if subjects get fatigued by performing a task, then they would be expected to do worse on the second condition they were in. However, as long as the order of presentation is counterbalanced so that half of the subjects are in Condition A first and Condition Bsecond, the fatigue effect itself would not invalidate the results, although it would add noise and reduce power. The carryover effect is symmetric in that having Condition A first affects performance in Condition B to the same degree that having Condition B first affects performance in Condition A.

Asymmetric carryover effects cause more serious problems. For example, suppose performance in ConditionB were much better if preceded by Condition A, whereas performance in Condition A was approximately the same regardless of whether it was preceded by Condition B. With this kind of carryover effect, it is probably better to use a between-subjects design.

One between- and one within-subjects factor

In the "Stroop Interference" case study, subjects performed three tasks: naming colors, reading color words, and naming the ink color of color words. Some of the subjects were males and some were females. Therefore, this design had two factors: gender and task. The ANOVA Summary Table for this design is shown in Table $15.8.3$.

Table $15.8.3$: ANOVA Summary Table for Stroop Experiment

Source	df	SSQ	MS	F	p
Gender	1	83.32	83.32	1.99	0.165
Error	45	1880.56	41.79		
Task	2	9525.97	4762.99	228.06	<0.001
Gender x Task	2	55.85	27.92	1.34	0.268
Error	90	1879.67	20.89		

The computations for the sums of squares will not be covered since computations are normally done by software. However, there are some important things to learn from the summary table. First, notice that there are two error terms: one for the between-subjects variable Gender and one for both the within-subjects variable Task and the interaction of the between-subjects variable and the within-subjects variable. Typically, the mean square error for the between-subjects variable will be higher than the other mean square error. In this example, the mean square error for Gender is about twice as large as the other mean square error.

The degrees of freedom for the between-subjects variable is equal to the number of levels of the between-subjects variable minus one. In this example, it is one since there are two levels of gender. Similarly, the degrees of freedom for the within-subjects variable is equal to the number of levels of the variable minus one. In this example, it is two since there are three tasks. The degrees of freedom for the interaction is the product of the degrees of freedom for the two variables. For the $Gender \times Task$ interaction, the degrees of freedom is the product of degrees of freedom Gender (which is 1) and the degrees of freedom Task (which is 2) and is equal to 2.

Assumption of Sphericity

Within-subjects ANOVA makes a restrictive assumption about the variances and the correlations among the dependent variables. Although the details of the assumption are beyond the scope of this book, it is approximately correct to say that it is assumed that all the correlations are equal and all the variances are equal. Table $15.8.4$ shows the correlations among the three dependent variables in the "Stroop Interference" case study.

Table 15.8.4: Correlations Among Dependent Variables

	word reading	color naming	interference
word reading	1	0.7013	0.1583
color naming	0.7013	1	0.2382
interference	0.1583	0.2382	1

Note that the correlation between the word reading and the color naming variables of 0.7013 is much higher than the correlation between either of these variables with the interference variable. Moreover, as shown in Table $15.8.5$, the variances among the variables differ greatly.

Table 15.8.5: Variances

Variable	Variance
word reading	15.77
color naming	13.92
interference	55.07

Naturally the assumption of sphericity, like all assumptions, refers to populations not samples. However, it is clear from these sample data that the assumption is not met in the population.

Consequences of Violating the Assumption of Sphericity

Although ANOVA is robust to most violations of its assumptions, the assumption of sphericity is an exception: Violating the assumption of sphericity leads to a substantial increase in the Type I error rate. Moreover, this assumption is rarely met in practice. Although violations of this assumption had at one time received little attention, the current consensus of data analysts is that it is no longer considered acceptable to ignore them.

Approaches to Dealing with Violations of Sphericity

If an effect is highly significant, there is a conservative test that can be used to protect against an inflated **Type I** error rate. This test consists of adjusting the degrees of freedom for all within-subjects variables as follows: The degrees of freedom numerator and denominator are divided by the number of scores per subject minus one. Consider the effect of Task shown in Table $15.8.3$. There are three scores per subject and therefore the degrees of freedom should be divided by two. The adjusted degrees of freedom are:

$(2)(1/2) = 1$ for the numerator and

$(90)(1/2) = 45$ for the denominator

The probability value is obtained using the F probability calculator with the new degrees of freedom parameters. The probability of an F of 228.06 or larger with 1 and 45 degrees of freedom is less than 0.001. Therefore, there is no need to worry about the assumption violation in this case.

Possible violation of sphericity does make a difference in the interpretation of the analysis shown in Table $15.8.2$. The probability value of an F of 5.18 with 1 and 23 degrees of freedom is 0.032, a value that would lead to a more cautious conclusion than the p value of 0.003 shown in Table $15.8.2$.

The correction described above is very conservative and should only be used when, as in Table $15.8.3$, the probability value is very low. A better correction, but one that is very complicated to calculate, is to multiply the degrees of freedom by a quantity called ε (the Greek letter epsilon). There are two methods of calculating ε. The correction called the Huynh-Feldt (or $H-F$) is slightly preferred to the one called the Greenhouse-Geisser (or $G-G$), although both work well. The $G-G$ correction is generally considered a little too conservative.

A final method for dealing with violations of sphericity is to use a multivariate approach to within-subjects variables. This method has much to recommend it, but it is beyond the scope of this text.