

Unequal Sample Sizes

Prerequisites

[ANOVA Designs](#), [Multi-Factor Designs](#)

The Problem of Confounding

Whether by design, accident, or necessity, the number of subjects in each of the various experimental conditions may not be equal. For example, the sample sizes for the [Obesity and Relationships](#) case study are shown in Table 1.

Although they were approximately equal, the most subjects were in the Acquaintance/Typical condition. Since n is used to refer to the sample size of an individual group, designs with unequal sample sizes are sometimes referred to as designs with unequal n .

Table 1. Sample Sizes for Obesity and Relationships

		Companion Weight	
		Obese	Typical
Relationship	Girl Friend	40	42
	Acquaintance	40	54

We consider an absurd design to illustrate the main problem caused by unequal n . Suppose an experimenter were interested in the effect of diet and exercise on cholesterol. The sample sizes are shown in Table 2.

Table 2. Sample Sizes for Diet and Exercise

		Exercise	
		Moderate	None
Diet	Low Fat	5	0
	High Fat	0	5

What makes this example absurd, is that there are no subjects in either the Low Fat/No exercise condition or the High Fat/Moderate exercise condition. The (hypothetical) data showing change in cholesterol are shown in Table 3.

Table 3. Data for Diet and Exercise

		Exercise		
		Moderate	None	Mean
Diet	Low Fat	-20 -25 -30 -35 -15	0	-25
	High Fat	0	-20 6 -10 -6 5	-5
	Mean	-25	-5	-15

The last column shows the mean change in cholesterol for the two Diet conditions whereas the last row shows the mean change for the two Exercise conditions. The value of -15 in the lower-right most cell in the table is the mean of all subjects.

We see from the last column that those on the low-fat diet lowered their cholesterol an average of 25 units whereas those on the high-fat diet lowered theirs by only 5 units. However, there is no way of knowing whether the difference is due to diet or to exercise since every subject in the low-fat condition was in the moderate-exercise condition and every subject in the high-fat condition was in the no-exercise condition. Therefore, Diet and Exercise are completely [confounded](#). The problem with unequal n is that it causes confounding.

Weighted and Unweighted Means

The difference between *weighted* and *unweighted means*, a difference critical for understanding how to deal with the confounding resulting from unequal n.

Weighted and unweighted means will be explained using the data shown in Table 4. Here, Diet and Exercise are confounded because 80% of the subjects in the low-fat condition exercised as compared to 20% of those in the high-fat condition. However, there is not complete confounding as there was with the data in Table 3.

The weighted mean is computed as the mean of the low-fat moderate-exercise condition and the low-fat no-exercise mean, weighted in accordance with sample size. To compute a weighted mean, you multiply each mean by its sample size and divide by N, the total number of observations. Since there are four subjects in the low-fat moderate-exercise condition and one subject in the low-fat no-exercise condition, the means are weighted by factors of 4 and 1 as

shown below where M_W is the weighted mean.

$$M_w = \frac{(4)(-27.5) + (1)(-20)}{5} = -26$$

The weighted mean for the low-fat condition is also the mean of all five scores in this condition. Thus if you ignore the factor "Exercise," you are implicitly computing weighted means.

The unweighted mean for the low-fat condition (M_U) is simply the mean of the two means.

$$M_u = \frac{-27.5 - 20}{2} = -23.75$$

Table 4. Data for Diet and Exercise with Partial Confounding

		Exercise			
		Moderate	None	Weighted Mean	Unweighted Mean
Diet	Low Fat	-20 -25 -30 -35 M=-27.5	-20 M=-20.0	-26	-23.75
	High Fat	-15 M=-15.0	6 -6 5 -10 M=-1.25	-4	-8.125
	Weighted Mean	-25	-5		
	Unweighted Mean	-21.25	-10.625		

Consider the the [main effect](#) of Diet by comparing the weighted mean for the low-fat diet (-26) with the weighted mean for the high-fat diet (-4). This difference of -22 is called "the effect of diet ignoring exercise" and is misleading since most of the low-fat subjects exercised and most of the high-fat subjects did not. However, the difference between the unweighted means of -15.5 (-23.75 minus -8.25) is not affected by this confounding and is therefore a better measure of the main effect. In short, the weighted means ignore the effects of other variables (exercise in this example) and result in confounding; unweighted means control for the effect of other variables and

therefore eliminate the confounding.

Statistical analysis programs use different terms for means that are computed controlling for other effects. SPSS calls them "*estimated marginal means*" whereas SAS and SAS JMP call them *least squares means*.

Types of Sums of Squares

The section on [Multi-factor ANOVA](#) stated that the sum of squares total was not equal to the sum of the sum of squares for all the other sources of variation when there is unequal n. This is because the confounded sums of squares are not apportioned to any source of variation. For the data in Table 4, the sum of squares for Diet is 390.625, the sum of squares for Exercise is 180.625, and the sum of squares confounded between these two factors is 819.375 (the calculation of this value is beyond the scope of this introductory text). In the ANOVA Summary Table shown in Table 5, this large portion of the sum of squares is not apportioned to any source of variation and represents the "missing" sums of squares. That is, if you add up the sums of squares for Diet, Exercise, D x E, and Error, you get 904.625. If you add the confounded sum of squares of 819.375 to this value you get the total sum of squares of 1722.00. When confounded sums of squares are not apportioned to any source of variation, the sums of squares are called *Type III sums of squares*. Type III sums of squares are, by far, the most common and if sums of squares are not otherwise labeled, it can safely be assumed that they are Type III.

Table 5. ANOVA Summary Table for Type III SSQ

Source	df	SSQ	MS	F	p
Diet	1	390.625	390.625	7.43	0.034
Exercise	1	180.625	180.625	3.423	0.113
D x E	1	15.625	15.625	0.2969	0.605
Error	6	315.750	52.625		
Total	9	1722.000			

When confounded sums of squares are apportioned to sources of variation, the sums of squares are called *Type I sums of squares*. The order in which the confounded sums of squares is apportioned is determined by the order in which the effects are listed. The first effect gets any sums of squares confounded between it and any of the other effects. The second gets the sums of squares confounded between it and subsequent effects, but not confounded with the first effect, etc. The Type I sums of squares are shown in Table 6. As you can

see, with Type I sums of squares, the sum of all sums of squares is the total sum of squares.

Table 6. ANOVA Summary Table for Type III SSQ

Source	df	SSQ	MS	F	p
Diet	1	1210.000	390.625	7.43	0.034
Exercise	1	180.625	180.625	3.423	0.113
D x E	1	15.625	15.625	0.2969	0.605
Error	6	315.750	52.625		
Total	9	1722.000			

In *Type II sums of squares*, sums of squares confounded between main effects is not apportioned to any source of variation whereas sums of squares confounded between main effects and interactions is apportioned to the main effects. In our example, there is no confounding between D x E interaction and either of the main effects. Therefore, the Type II sums of squares are equal to the Type III sums of squares.

Unweighted Mean Analysis

Type III sums of squares are tests of difference in unweighted means. However, there is an alternative method to testing the same hypotheses tested using Type III sums of squares. This method, *unweighted means analysis*, is computationally simpler than the standard method but is an approximate test rather than an exact test. It is, however, a very good approximation in all but extreme cases. Moreover, it is exactly the same as the traditional test for effects with one degree of freedom. The Analysis Lab uses unweighted means analysis and therefore may not match the results of other computer programs exactly when there is unequal n and the df are greater than one.

Causes of Unequal Samples

None of the methods for dealing with unequal sample sizes are valid if the experimental treatment is the source of the unequal sample sizes. Imagine an experiment seeking to determine whether publicly performing an embarrassing act would affect one's attitude toward public disclosure. In this imaginary experiment, the experimental group is asked to reveal to a group of people the most embarrassing thing they have ever done. The control group is asked to describe what they had at their last meal. Twenty subjects are recruited for the experiment and randomly divided into two equal groups of 10, one for the experimental treatment and one for the control. Following the description,

subjects are given an attitude survey concerning public disclosure. So far, this seems like a valid experimental design. However, of the 10 subjects in the experimental group, four withdrew from the experiment because they did not wish to publicly describe an embarrassing situation. None of the subjects in the control group withdrew. Even if the data analysis shows a significant effect, it would not be valid to conclude that the treatment had an effect because a likely alternative explanation cannot be ruled out. Namely, subjects who were willing to describe an embarrassing situation differed from those who were not even before the experiment began. Thus, the differential drop-out rate destroyed the [random assignment](#) of subjects to conditions, a critical feature of the experimental design. No amount of statistical adjustment can compensate for this flaw.