

14.2: Sources of variation

Introduction

Sources of variation, or components of the two-way ANOVA, include two factors, each with two or more levels (groups); collectively, factors are often referred to as the **main effects** in these types of ANOVA. The other source of variation in a two-way ANOVA is the **interaction** between the two factors. Below, I have listed the important components, although I have not included how the sum of squares are calculated. You are expected to know the sources of variation for this most basic two-way ANOVA table (Table 14.2.1). You should also be able to solve any missing elements in one of these tables by utilizing any included information.

Table 14.2.1. ANOVA table for two-way, balanced, replicated design.

Source of variation	Degrees of Freedom	Mean Squares	F -statistic
Factor A (Diet)	$a - 1$	$\frac{\text{factor A } SS}{\text{factor A } DF}$	$\frac{\text{factor A } MS}{\text{error } MS}$
Factor B (Population)	$b - 1$	$\frac{\text{factor B } SS}{\text{factor B } DF}$	$\frac{\text{factor B } MS}{\text{error } MS}$
$A \times B$ interaction	$(a - 1)(b - 1)$	$\frac{A \times B \text{ } SS}{A \times B \text{ } DF}$	$\frac{A \times B \text{ } MS}{\text{error } MS}$
Error (within-cells)	$ab(n - 1)$	$\frac{\text{error } SS}{\text{error } DF}$	
Total	$N - 1$		

Taking each row from Table 14.2.1 one at a time, we have:

Source of variation	Degrees of Freedom	Mean Squares	F -statistic
First Factor	$a - 1$	$\frac{\text{factor A } SS}{\text{factor A } DF}$	$\frac{\text{factor A } MS}{\text{error } MS}$

where Source refers to the source of variation, DF refers to Degrees of Freedom, a is the number of levels (groups) of the first factor, SS refers to Sum of Squares, and MS refers to the Mean Squares.

Next is the second factor

Source of variation	Degrees of Freedom	Mean Squares	F -statistic
Second Factor	$b - 1$	$\frac{\text{factor B } SS}{\text{factor B } DF}$	$\frac{\text{factor B } MS}{\text{error } MS}$

where b is the number of levels (groups) of the second factor. Next is the interaction between the first and second factors.

Source of variation	Degrees of Freedom	Mean Squares	F -statistic
Interaction	$(a - 1)(b - 1)$	$\frac{A \times B \text{ } SS}{A \times B \text{ } DF}$	$\frac{A \times B \text{ } MS}{\text{error } MS}$

and lastly the Within-cell Error or residual source of variation

Source of variation	Degrees of Freedom	Mean Squares	F -statistic
Error (within-cells)	$ab(n - 1)$	$\frac{\text{error } SS}{\text{error } DF}$	N/A

where n is the number of experimental units for each group. Note that if the sample size differs for one or more groups (levels), then the design would be unbalanced and this formula does not work to determine the degrees of freedom. The total degrees of freedom for the two-way ANOVA is simply $N - 1$, where N is the sample size for the entire problem; a little algebra shows that N may be calculated as $N = abn$

Unbalanced designs

An **unbalanced design** implies that observations are missing value for one or more groups. What to do if data are missing? The decision depends on how the data are missing (see [Chapter 5](#)). For example, if data are missing at random with respect to treatment, then this should not affect inference. If data are missing not at random, then inference, logically, must be impacted. Calculating the ANOVA, moreover, becomes a different matter. In the one-way ANOVA, no real problem arises although setting up contrasts among the levels requires a weighting term to be factored into the calculations. For higher-level ANOVA involving two or more factors, the sums of squares for treatment effects are no longer simple partitioning into the different sources of variation. The sources overlap and the order by which the Factors enter into the statistical model now affects the calculations. Thus, while setting up the calculations for the balanced design is straightforward, perhaps surprisingly, if group sizes differ, this simple relationship for calculating the degrees of freedom, sums of squares, and Mean squares becomes an unsolvable problem. This problem is largely solved by the **general linear model**.

Questions

1. In two-way ANOVA, what should you always test first?
 - A. The significance of Factor 1.
 - B. The significance of Factor 2.
 - C. The significance of the interaction between Factor 1 and Factor 2.
 - D. Doesn't matter which is tested first because you have three null hypotheses in the 2-way ANOVA.
 2. Why is the cell empty for F statistic in the Within-cell Error or residual source of variation?
 3. Based on the results of a two-way ANOVA, the error sums of squares (SSE) was computed to be 160. If we ignore one of the factors and perform a one-way ANOVA using the same data, will the SSE be the same as in the two-way ANOVA, or will it increase? Decrease? Explain your choice.
 4. While conducting a two-way ANOVA, you conclude that a statistically significant interaction exists between factor 1 and factor 2. What should be your next step? Do you drop the interaction term from the model and redo the analysis or do you report the results of factor effects including the non-significant interaction?
-

This page titled [14.2: Sources of variation](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michael R Dohm](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.