

Keppel's (1991) ANOVA notational system

Capital letters represent the sum of the scores in one level of the group represented by that letter, e.g., "A" means the sum of the scores in one level of factor A. There are as many of these sums as there are levels of A. If you have to identify them in particular, "A₁" is the sum of the scores in level 1 of factor A. The mean of group A₁ is notated "Y_{A1}," with a bar over the Y.

Lower case letters represent the number of levels. The number of subjects in a cell, i.e., the number of subjects who get the same treatment combination, is designated "n". For 3 levels of A, 2 levels of B, and 6 subjects per AB cell: a = 3, b = 2, n = 6. (Note that A becomes "a", B becomes "b", but S becomes "n" rather than "s"!)

SV	df	SS
A	a-1=2	[A]-[T]
B	b-1=1	[B]-[T]
AB	(a-1)(b-1)=2	[AB]-[A]-[B]+[T]
S/AB	ab(n-1)=30	[Y]-[AB]

1) From the list of sources of variance you can find df:

Multiply the (lower case) letters to the right of the slash by the (lower case) letters to the left of the slash **less one**: df for S/AB = ab(n-1) = 2*3*(6-1) = 30. If there is no slash, it's as if it's all the way to the right -- that is, as if everything in that term is to the left of the slash: for example, even though there is no slash in AB, pretend that those letters are to the left of the slash and subtract 1 from each before multiplying, giving (a-1)(b-1). (Once again, note that the lower case letter is the same as the SV it represents, except for S, which becomes "n"!)

2) and sums of squares in brackets:

Expand the df, substitute bracketed quantities for lower case letters, and use T for 1, so SS for AB is (a-1)(b-1) = ab-a-b+1 --> [AB]-[A]-[B]+[T]. Note: when the df contains an "n", use S in its place in writing the brackets; also note that when the bracket contains all the letters, it represents a single data point which should be notated as Y. Thus, for example, SS for S/AB is ab(n-1) = abn-ab --> [ABS]-[AB] -- but [ABS] contains all the letters, and since that means it represents a single data point, write it as [Y], giving you [Y]-[AB].

3) and sums of squares in equations:

For each bracketed quantity, (a) identify each of the sums represented by the capital letter(s) inside the brackets: to get [A], what are "A₁", "A₂", & "A₃"? (b) Square each of those sums. (c) Add them together. (d) Divide that new sum by the number of observations going into each original sum -- NOT by the number of groups. Example: each A is the sum of 12 subjects' data points, so divide the sum of the three A² by 12 -- NOT by 3!

This applies when there is more than one letter inside the brackets as well: AB represents the sum of the six scores in an AB cell (in AB₃₂, for instance). There are six of those cells whose sums need to be squared and then added together; then you divide that new sum by six -- NOT because there are six cells (that's coincidence!), but because each cell has six scores in it!

A single data point (notated as Y) is the combination of all the letters: Y (or "ABS") is one score. To calculate [Y] then (or "[ABS]", if you like), only one score "goes into" that "total", so you end up squaring each individual score (i.e., each Y "total") in the whole data set, adding them all up, and dividing by 1 -- or in other words, just squaring each score and adding them all up.

T is the total sum of the scores in the whole data set. You add all the scores together to get T, square that one number, and (since there can be no other T's to add to it) divide it by the total number of scores in the data set (also familiar from other notational systems as "big N").

Compute each bracket quantity and add and subtract them as specified by the df, and those resulting numbers are the sums of squares for each source of variance. Don't make the mistake of thinking that [T] is the total SS, or that [A] is the SS for the A treatment -- all those bracket quantities must be combined correctly before they tell you anything.