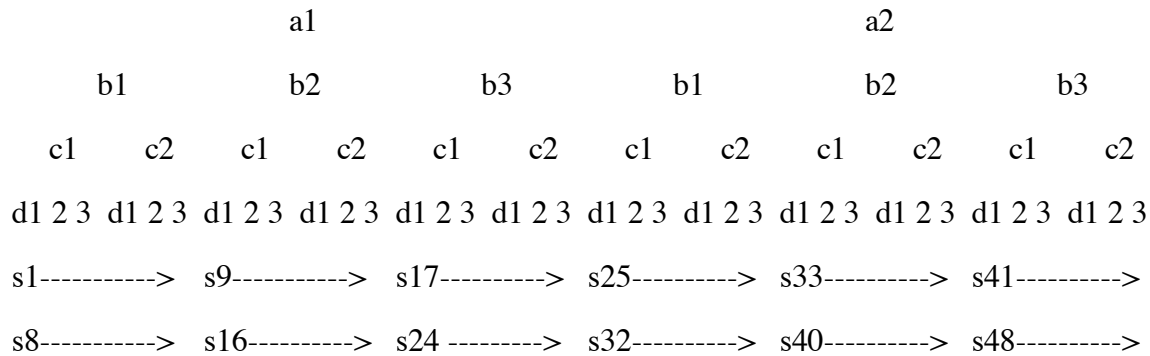


FINDING SOURCES OF VARIANCE WHEN YOU DON'T ALREADY HAVE THEM MEMORIZED

Why do an easy example when we could do a long drawn-out mixed design? Let's look at a 2 between, 2 within: $A \times B \times (C \times D \times S)$; $a=2$, $b=3$, $c=2$, $d=3$, $n=8$

If this is given to you verbally, write down all the factors and identify them as between or within. Then the first thing to do, no matter how smart you think you are, is to draw the whole tree, using arrows to indicate where subjects contribute more than one data point. (This also helps you determine what "n" is: it's the number of subjects who are all treated the same.) Put the between factors on the top and draw all the branches:



The most important question to get straight is, what are subjects nested in? To find out, look at your top variable and ask, are there completely different subjects at each level of this variable? Here 1-24 are in a1 and 25-48 are in a2, so the answer is yes. Then go to the cell at the next level down -- that is, divide the a cells into ab cells -- and ask the same question. Again the answer here is yes: each ab cell has completely different subjects in it (1-8 in a1b1, 9-16 in a1b2,... 41-48 in a2b3). Every time you answer yes, check the next level down. Next are the abc cells; do all the abc cells have different subjects? Here the answer is no: a1b1c1 has subjects 1-8 in it, but a1b1c2 also has subjects 1-8 in it. Stop once you've answered no. Go back up to the last time you said yes: those are the cells in which subjects are nested. It will be the lowest level in the tree at which different cells contain completely different subjects. That means that in this design, the between-factors subject term is S/AB (as opposed to S/A or S/ABC), since the ab cells were the lowest to hold different groups of subjects.

[Another way to do this is to look at the notation for labeling designs; subjects are nested in the cell representing the combination of everything outside the parentheses. For this design A and B are outside the parentheses, so subjects are nested in ab cells. For $A \times B \times C \times (D \times S)$, subjects would be nested in abc cells -- the subject term is S/ABC. To make this consistent, write parentheses around any completely within design: $A \times B \times S$ should be written $(A \times B \times S)$, and subjects aren't nested in anything -- rather, they cross A and B. But pay more attention to the design than to the notation. Generally, subjects are nested in the combination of all the between factors.]

Now you list the sources of variance. If you do it this way, each group of effect terms will be followed in your list by the error term used to test them in an F ratio -- how convenient! Write down all the between terms first, followed by their interactions, and then the (nested) subject term. In this case you have A, B, AB, and S/AB. (As another example, if the design were $A \times (B \times C \times D \times S)$, the between terms would simply be A and S/A. This design is pronounced "A, by...B-by-C-by-D-by-S".) Next, write down the first within term, in this case C. Then combine that term (C) with each and every one of the terms that preceded it (A, B, AB, and S/AB). The terms following C, then, are AC, BC, ABC, and SC/AB.

Note three things at this point. First, write the letters to the left of the slash in alphabetical order, except that S comes first. That means that instead of writing CA you write AC, and instead of CS/AB it's SC/AB. The only reason for this is that you'll be less likely to miss any terms in a long list, or write them twice, if they always have a uniform appearance. Second, when you combine nested terms, whatever's to the left of the slash stays to the left, and whatever's to the right stays on the right. That's why when you combine C with S/AB you get SC/AB and not S/ABC. If there's no slash pretend everything's to the left. Third, a letter can only appear once in any SV. (If you should accidentally violate this rule somehow you can probably correct it as follows: if both occurrences are on the same side of the slash, erase one; if they're on opposite sides, erase the one on the left of the slash. But this will hopefully never come up.)

Do the same for the next within term: write down D; then D combined successively with A, B, AB, S/AB, C, AC, BC, ABC, and SC/AB, gives AD, BD, ABD, SD/AB, CD, ACD, BCD, ABCD, and SCD/AB. Note that you combine the newest term with all of the preceding terms (in order), not just the between terms. If there were an E within factor as well, it would be combined, in order, with all the SV's we've listed so far. The resulting list of Sources of Variance is as follows:

A	
B	
AB	
S/AB	error term for A, B, and AB
C	
AC	
BC	
ABC	
SC/AB	error term for C, AC, BC, and ABC
D	
AD	
BD	
ABD	
SD/AB	error term for D, AD, BD, and ABD
CD	
ACD	
BCD	
ABCD	
SCD/AB	error term for CD, ACD, BCD, and ABCD

From this list you can find df (use the lower case letters to the right of the slash, times the lower case letters to the left of the slash less one, so df for SD/AB = $ab(n-1)(d-1) = 2*3*(8-1)*(3-1) = 84$); sums of squares from brackets (expand the df and use T for 1, so SS for SD/AB is $ab(dn-n-d+1) = abdn-abn-abd+ab \rightarrow [Y]-[ABS]-[ABD]+[AB]$; SS for C is $c-1 \rightarrow [C] - [T]$); and EMS -- so once you can find sources of variance the whole ANOVA practically does itself for you.

Note that the error terms in this list are exactly those predicted by the EMS for this design when all the effects (aside from subjects) are fixed. When there are random effects besides subjects, this system produces all the sources of variance, but does not match them up with their error terms! Usually all the factors are fixed, though.

[An alternative way to be sure you've found all the sources of variance -- use this only if the first method overwhelms you or if you need to check yourself: first, simply list all the factors (including subjects) singly as they appear in your tree, and make all possible combinations of them. Then for every term containing the letter S, put a slash to the right of it followed by the combination of terms in which subjects are nested (there's no getting around figuring that out). Then erase any term that has the same letter on both sides of the slash. Here's how this alternative method works...

[For the present example, start with A,B,C,D, and S; list all the two way interactions: AB, AC, AD, AS, BC, BD, BS, CD, CS, DS; all the three ways: ABC, ABD, ABS, ACD, ACS, ADS, BCD, BCS, BDS, CDS; the four ways: ABCD, ABCS, ABDS, ACDS, BCDS; and the five way: ABCDS. (I hope that's all of them.) Next step is to add in the slash to any subject terms, in this case with AB to the right since that's what subjects are nested in; in effect we're substituting S/AB for every plain old S in our list. That gives us the single terms A, B, C, D, and S/AB; the two way interactions: AB, AC, AD, AS/AB, BC, BD, BS/AB, CD, CS/AB, DS/AB; the three ways: ABC, ABD, ABS/AB, ACD, ACS/AB, ADS/AB, BCD, BCS/AB, BDS/AB, CDS/AB; the four ways: ABCD, ABCS/AB, ABDS/AB, ACDS/AB, BCDS/AB; and the five way: ABCDS/AB. Finally, we erase any terms with the same letter on both sides of the slash. This is because it would be impossible for one factor to interact with another factor that it's nested in -- how could a subject behave differently depending on the level of the other factor, if a subject only gets one level of the other factor? That is, if subjects only get one level of A or B, how can A or B interact with subjects? It can't. Erasing the impossible terms AS/AB, BS/AB, ABS/AB, ACS/AB, ADS/AB, BCS/AB, BDS/AB, ABCS/AB, ABDS/AB, ACDS/AB, BCDS/AB, and ABCDS/AB, we're left with A, B, C, D, S/AB, AB, AC, AD, BC, BD, CD, CS/AB, DS/AB, ABC, ABD, ACD, BCD, CDS/AB, and ABCD.

This is exactly the same list we arrived at using the other method, but this rote listing method doesn't tell us the error terms. The way to get the error term for any SV, assuming all fixed effects, is this, I think. Combine the term with the design's subject term (in this case S/AB), and if any letter appears on both sides of the slash, erase it from the left side. For example, for B we form the combination BS/AB, then since B appears on both sides, we erase B from the left side, leaving S/AB. Similarly, for ABD we form ABDS/AB, and erase A and B from the left side, leaving DS/AB, or more familiarly SD/AB.

[Can we do this all in one step? For an $A \times (B \times S)$ design, with all factors except subjects fixed, we know that the subject term is S/A since A is outside the parentheses; we know that AB will be one of the terms since it is a combination that does not have any letter appearing on both sides of a slash; and combining it with S/A gives ABS/A, so we erase the A from the left side to get BS/A for an error term. Another example: for an $A \times B \times C \times (D \times E \times S)$ design, with all factors except subjects fixed, we know that the subject term is S/ABC since A, B, and C are outside the parentheses; we know that BCD will be one of the terms since it is a combination that does not have the same letter appearing on both sides of a slash; and combining it with S/ABC gives BCDS/ABC, so we erase the B and C from the left side to get DS/ABC for an error term. Final example: in $(A \times B \times S)$, the subject term is just S, since nothing is outside the parentheses; for the term AB, the error term will be its combination with S, or simply the ABS interaction. What could be simpler, right?]