EXPECTED MEAN SQUARES: not **Greek to me!**

Expected Mean Squares are theoretical descriptions of group differences broken into their logical components of variability. The Mean Squares in the ANOVA table are numbers obtained to represent group variances; the EMS are abstract representations of the MS.

This discussion isn't replacing the stated rules, or conflicting with them in any way. It's just elucidating them a bit. Let's say you have all the sources of variance listed. First determine whether each effect is fixed or random. (This has nothing to do with whether you have a "completely randomized design", which just means "between-subjects": no subject appears in more than one cell. In any design, each effect is considered "fixed" or "random".)

How do you know if it's Fixed or Random? You might be told outright, for one thing. But if you have to figure it out, start by assuming it's probably fixed. *Fixed* effects are what you're used to and what are most common by way far. *Random* effects are those for which you are really not interested in the particular levels you've chosen for your design; instead you want to generalize from the levels you've randomly selected to other possible levels. Good examples of this are: Subjects -- would you really want to discover how well Ivan, Penelope, Herb, and Wilhelmina do on your task? No, you'd want to generalize from their performance to how people like them generally would perform. Groups -- would you care about the particular combinations of subjects with certain characteristics that you happened to use? No, you'd want to generalize to the possible other classrooms, or other therapy groups, or other sets of roommates, or other groups of whatever type you used.

[There are also cases you most likely won't encounter, in which items selected to be representative are random effects (if they're even analyzed in the design): in a reading test you're not interested in how well children read the particular fifty words on your test, but rather how those words are indicative of their ability to read the rest of the words in the language, so Words would be random. Or if you're sampling food from restaurant chains -- do you want to make statements about particular McDonald's franchises in Willimantic, in New Haven, and in Vernon? No, you want your conclusions to apply to the food at any McDonald's, not just the ones you randomly selected, so Locations is random. What's a bad example of a random variable? Well, Gender sure is -- does it make any sense at all to say you used the levels "male" and "female" in order to generalize to all the other possible sexes? Those are the only levels of interest. And you don't have to exhaust all the possible levels: Drug Dosage is not a random effect -- you choose something like 10, 50, and 100 mg because you want to know what those amounts will do; you don't choose them randomly so that you can also make statements about what happens when you give 200mg or 500 mg or 20g. As long as the levels you use are the only ones you want to make statements about, it's a fixed effect.]

So you have to say for each single factor whether it's fixed or random. For interactions of factors, if any one of the combined factors is random itself, then the whole combined term counts as random.

The Rules. Once you know for each term whether it's fixed or random, it's trivial to write down a variance symbol: θ^2 for anything fixed, and σ^2 for anything random. Then, equally trivial, you put the SV itself as a subscript: for B you have θ^2_B (it's θ^2 assuming B is fixed) and for SC/AB you have $\sigma^2_{SC/AB}$ (it's σ^2 even if C is fixed because it only takes one random term -- S -- to make the whole thing random). Finally, use all the remaining letters as coefficients, in lower case since they're representing numbers. If you have variables up through D, then your terms become $acdn\theta^2_B$ and $d\sigma^2_{SC/AB}$. Now you have a complete variance term, the hypothesis term, for each SV.

[You use all the other letters in the design as coefficients because you want to multiply the variance by the number of times it enters into your pattern of differences. If you drew a design with the correct terms -- draw AxBx(CxDxS) -- you'd see that S/AB interacted with C in d different cases, for instance SC/AB at D1, at D2, and at D3; thus, $d\sigma^2_{SC/AB} = 3\sigma^2_{SC/AB}$. Likewise you can see an effect of B at every combination of A, C, and D, so there are acd=2*2*3=12 places that the variance associated with B enters into the group differences.]

Once you have a complete variance term for each effect, the only question is what other variances are components of the differences in those effects -- or, concretely, which other variance terms should be added to the hypothesis term as components of each effect. Well, everything gets the random population variance, σ^2_e. Then, according to the rules, consider the hypothesis term of another SV if that SV itself (or, same thing, the subscript of its hypothesis term) contains all the letters of the effect you're working on: then looking only to the left of any slashes, if all the letters aside from the effect you're working on are random, then you add that variance term to the effect you're working on.

What you're doing by these rules is just mechanically adding other terms that represent the effects of random variables on your term. Notice that this means you'll only ever add σ^2 terms as components, since a θ^2 would represent some kind of fixed variable; if you find yourself adding a θ^2 term, you've either made a mistake in the rules or chosen the wrong Greek letter. A more useful thing to notice is that you can do all your combining of terms before you write a single Greek letter. All the information you use to make those decisions is right in the list of your sources of variance: which terms are fixed and random, which terms contain which letters, and where the slashes are. You might find it easier, then, to work out your combinations first and then worry about the simple stuff, i.e., the Greek letters, subscripts, and coefficients.

There's one important exception to the way the terms are written. When subjects give only a single data point (i.e., the design is completely randomized), the subject term's EMS is just σ^2_e regardless of what subjects are nested in. That is, S/A for a one-factor CR design, S/AB for a two-factor, S/ABC for three, etc., would each have just the population variance term σ^2_e. Do not write "$\sigma^2_e + \sigma^2_{S/AB}$", for instance, and do not try to add "$\sigma^2_{S/AB}$" as a component of the EMS for A as the rules would suggest; the term is just σ^2_e, and that's already part of all the other terms, including A. This only holds for completely randomized designs, where you always find that the bottom SV is S to the left of a slash with all the other SV's to the right. In mixed designs you also may find terms like S/AB, but in those cases you do write "$\sigma^2_e + \sigma^2_{S/AB}$".

Using EMS. Yes, there are reasons to bother with this stuff. First, it tells you what MS terms are error terms for what other MS terms, i.e., how to make F ratios. An error MS is the denominator of an F ratio, and it has all the same EMS components as the numerator except for one -- the numerator's hypothesis term. For instance, in a one-way repeated measures design (AxS), EMS_A is "$\sigma^2_e + \sigma^2_{AS} + n\theta^2_A$", so its error term is "$\sigma^2_e + \sigma^2_{AS}$", or EMS_{AS}. When H_O is true, $\theta^2_A = 0$, thus the F ratio has the same numerator and denominator, which should make it equal 1. The bigger the effect, the more is added to the numerator and the bigger the F ratio gets.

[The idea of a "true" F ratio refers to the fact that mathematically, F should have the same numerator and denominator; what we like, experimentally, is when our F is not a true F ratio, i.e., when the numerator has something extra in it, namely a hypothesis term bigger than zero. Then the p-value tells us the chances that we really do have a true F ratio. If that p is really small, it tells us that we probably don't have a real F, and we conclude that the culprit is the hypothesis term we threw into the numerator. When we test a strong effect it is very unlikely that the F we compute is a true F, since we see things like $p = .0001$, and that makes us happy.]

You also find out that some effects don't even have error terms. In the simplest (AxS) case, that's true of S, since its EMS is "$\sigma^2_e + a\sigma^2_S$" and there is no other term with just "σ^2_e" as its EMS. But we might try to make an F ratio anyway, using "$\sigma^2_e + \sigma^2_{AS}$", or EMS_{AS}, as the denominator. Write out that fraction and you'll see that the denominator is bigger than it should be according to the above definition of an error term-- so your F will be smaller than it should be. This is known as a conservative F ratio: given that it's biased toward being small by its puffed up denominator, if it turns out to be significant anyway you'll know that it's really significant (or in technical terms, "way significant"). If the conservative F is not significant, the S effect might still be significant but you have no way to find out. Luckily, you rarely care about the effects without error terms anyhow. (For the gullible I should mention that "way significant" isn't really a technical term.)

Pooling error terms is another neat thing to do based on EMS. Keep in mind that two things can make an F more significant: a small error term (so the F ratio is larger), or lots of df (even a fairly small F can be significant on lots of df). Under certain circumstances you can combine error terms in your design to make it more likely that your F will reach significance. Read this very slowly: say you have a hypothesis MS term for your numerator and an error MS term for your denominator. Looking through your EMS for the design, you may see that there is a third MS term that has nearly the same EMS as your error term. This third EMS should differ from your error EMS by just one component. Now, if that component represents a hypothesis term that was found to be really small, then you can pretend that extra component is not even there, and add the third term into the error term. Before doing the mechanics of that, look at a concrete example (and isn't it amazing what can count as a "concrete example"?):

In the Ax(BxS) mixed design, the EMS for the B hypothesis term is "$\sigma^2_e + \sigma^2_{SB/A} + an\theta^2_B$", and for its error term SB/A the EMS is "$\sigma^2_e + \sigma^2_{SB/A}$". Fine -- you could compute your F ratio as is, and that would be the standard thing to do. But if your output says that F isn't significant, you do the advanced thing. You look at the EMS for the interaction term AB, which is "$\sigma^2_e + \sigma^2_{SB/A} + n\theta^2_{AB}$". It's the same as the error term except for the "$n\theta^2_{AB}$" part. Look at your output again; what's the p-value for the AB interaction? If it's bigger than .25, that tells you that the "$n\theta^2_{AB}$" part is small enough to be negligible. It's as if the interaction term has the same EMS as the error term. So use them both as the error term! Important: it's not good enough for the third term to just be non-significant, $p > .05$; it has to be ridiculously non-significant in order for you to disregard the hypothesis component in its EMS. The rule of thumb is that $p > .25$ counts as ridiculously non-significant.

The mechanics of combining the terms is really simple. Any MS is just the SS divided by the corresponding df -- nothing new -- so to combine MS terms, add up all the SS and divide by the added-up df. In the case above, the pooled MS is just ($SS_{SB/A} + SS_{AB}$) divided by ($df_{SB/A} + df_{AB}$). The reason this helps is that even if the size of the new MS error term is exactly the same, resulting in the exact same F ratio, the df for the denominator has gone from a(b-1)(n-1) to a(b-1)(n-1) + (a-1)(b-1), which means you need to reach a smaller critical F value to get significance.

If all this hasn't convinced you that EMS are both useful and fun, go read about "quasi-F ratios".

Check the following EMS examples; I hope I did them right, but if I didn't, sue me.

AxBx(CxDxS), all fixed effects:

A	$\sigma^2_e + cd\sigma^2_{S/AB} + bcdn\theta^2_A$	
B	$\sigma^2_e + cd\sigma^2_{S/AB} + acdn\theta^2_B$	
AB	$\sigma^2_e + cd\sigma^2_{S/AB} + cdn\theta^2_{AB}$	
S/AB	$\sigma^2_e + cd\sigma^2_{S/AB}$	error term for A, B, and AB
C	$\sigma^2_e + d\sigma^2_{SC/AB} + abdn\theta^2_C$	
AC	$\sigma^2_e + d\sigma^2_{SC/AB} + bdn\theta^2_{AC}$	
BC	$\sigma^2_e + d\sigma^2_{SC/AB} + adn\theta^2_{BC}$	
ABC	$\sigma^2_e + d\sigma^2_{SC/AB} + dn\theta^2_{ABC}$	
SC/AB	$\sigma^2_e + d\sigma^2_{SC/AB}$	error term for C, AC, BC, and ABC
D	$\sigma^2_e + c\sigma^2_{SD/AB} + abcn\theta^2_D$	
AD	$\sigma^2_e + c\sigma^2_{SD/AB} + bcn\theta^2_{AD}$	
BD	$\sigma^2_e + c\sigma^2_{SD/AB} + acn\theta^2_{BD}$	
ABD	$\sigma^2_e + c\sigma^2_{SD/AB} + cn\theta^2_{ABD}$	
SD/AB	$\sigma^2_e + c\sigma^2_{SD/AB}$	error term for D, AD, BD, and ABD
CD	$\sigma^2_e + \sigma^2_{SCD/AB} + abn\theta^2_{CD}$	
ACD	$\sigma^2_e + \sigma^2_{SCD/AB} + bn\theta^2_{ACD}$	
BCD	$\sigma^2_e + \sigma^2_{SCD/AB} + an\theta^2_{BCD}$	
ABCD	$\sigma^2_e + \sigma^2_{SCD/AB} + n\theta^2_{ABCD}$	
SCD/AB	$\sigma^2_e + \sigma^2_{SCD/AB}$	error term for CD, ACD, BCD, and ABCD

AxBx(CxDxS), A,B,C fixed effects, D random -- note the paucity of error terms due to D being a random effect:

A	$\sigma^2_e + c\sigma^2_{SD/AB} + cd\sigma^2_{S/AB} + bcn\sigma^2_{AD} + bcdn\theta^2_A$	no error term!
B	$\sigma^2_e + c\sigma^2_{SD/AB} + cd\sigma^2_{S/AB} + acn\sigma^2_{BD} + acdn\theta^2_B$	no error term!
AB	$\sigma^2_e + c\sigma^2_{SD/AB} + cd\sigma^2_{S/AB} + cn\sigma^2_{ABD} + cdn\theta^2_{AB}$	no error term!
S/AB	$\sigma^2_e + c\sigma^2_{SD/AB} + cd\sigma^2_{S/AB}$	
C	$\sigma^2_e + \sigma^2_{SCD/AB} + d\sigma^2_{SC/AB} + abn\sigma^2_{CD} + abdn\theta^2_C$	no error term!
AC	$\sigma^2_e + \sigma^2_{SCD/AB} + d\sigma^2_{SC/AB} + bn\sigma^2_{ACD} + bdn\theta^2_{AC}$	no error term!
BC	$\sigma^2_e + \sigma^2_{SCD/AB} + d\sigma^2_{SC/AB} + an\sigma^2_{BCD} + adn\theta^2_{BC}$	no error term!
ABC	$\sigma^2_e + \sigma^2_{SCD/AB} + d\sigma^2_{SC/AB} + n\sigma^2_{ABCD} + dn\theta^2_{ABC}$	no error term!
SC/AB	$\sigma^2_e + \sigma^2_{SCD/AB} + d\sigma^2_{SC/AB}$	
D	$\sigma^2_e + c\sigma^2_{SD/AB} + abcn\sigma^2_D$	
AD	$\sigma^2_e + c\sigma^2_{SD/AB} + bcn\sigma^2_{AD}$	
BD	$\sigma^2_e + c\sigma^2_{SD/AB} + acn\sigma^2_{BD}$	
ABD	$\sigma^2_e + c\sigma^2_{SD/AB} + cn\sigma^2_{ABD}$	
SD/AB	$\sigma^2_e + c\sigma^2_{SD/AB}$	
CD	$\sigma^2_e + \sigma^2_{SCD/AB} + abn\sigma^2_{CD}$	
ACD	$\sigma^2_e + \sigma^2_{SCD/AB} + bn\sigma^2_{ACD}$	
BCD	$\sigma^2_e + \sigma^2_{SCD/AB} + an\sigma^2_{BCD}$	
ABCD	$\sigma^2_e + \sigma^2_{SCD/AB} + n\sigma^2_{ABCD}$	
SCD/AB	$\sigma^2_e + \sigma^2_{SCD/AB}$	