periment has 8 male autistics, 2 female autistics, 2 male controls, and 8 female controls (disproportionate cell frequencies), you can make a better-than chance guess as to whether a subject is male or female if you know whether they're autistic or not — the two factors are correlated. It is, of course, possible to have a middle ground — unequal but proportionate cell frequencies (see above, p. 70, for an example), which still involves orthogonal sums of squares.

6.4 Expected mean squares (EMS) and error terms

First we need to consider the **sampling fraction** for fixed and random factors (fixed and random factors are defined on p. 31). If we have factor A with a levels and it is a fixed factor, we have sampled all the levels. We can say that the maximum number of levels of A is $a_{max} = a$, and the sampling fraction $a / a_{max} = 1$. On the other hand, if our factor is a random factor, a_{max} is likely to be very large, so $a / a_{max} = 0$, approximately. Take the example of subjects: we presume that our s subjects are sampled from a very large population, $s_{max} \approx \infty$, so the sampling fraction $s / s_{max} = 0$.

It is possible to have sampling fractions between 0 and 1 (Howell, 1997, p. 423) — but you will have to work out some messy EMSs yourself. Software packages such as SPSS assume that the sampling fraction is 1 for fixed factors and 0 for random factors.

6.4.1 Rules for obtaining expected mean squares (EMS)

From Myers & Well (1995, p. 299). Let's list the rules with an illustrative example. Suppose we have one between-subjects factor A with 3 levels. There are 6 subjects *per level* of the between-subjects factor ($n = 6$). There are 4 levels of a within-subjects factor B.

1. Decide for each independent variable, including Subjects, whether it is **fixed or random.** Assign a letter to designate each variable. Assign another letter to represent the number of levels of each variable. (In our example, the variables are designated A, B, and S; the levels are a, b, and n respectively. A and B are fixed and S is random.)

2. Determine the **sources of variance** (SS) from the structural model. (We've already seen what this produces for our example design, when we discussed it earlier: SS_{total} is made up of $SS_A + SS_{S/A} + SS_B + SS_{AB} + SS_{SB/A}$. These are our sources of variance.)

3. List σ_e^2 as part of each EMS.

4. For each EMS, list the null hypothesis component — that is, the component corresponding directly to the source of variance under consideration. (Thus, we add $nb\sigma_A^2$ to the EMS for the A line, and $b\sigma_{S/A}^2$ to the EMS for the S/A line.) Note that a component consists of three parts:

   - A coefficient representing the number of scores at each level of the effect (for example, nb scores at each level of A, or b scores for each subject).
   - σ^2

     [Myers & Well (1995, pp. 299) use σ_A^2 if A is a random factor, and θ_A^2 if A is a fixed factor; Howell (1997, p. 423) doesn't, and I think it's clearer not to.]
   - As subscripts, those letters that designate the effect under consideration.

5. Now add to each EMS all components whose subscripts contain all the letters designating the source of variance in question. (For example, since the subscript SB/A contains the letters S and A, add $\sigma_{SB/A}^2$ to the EMS for the S/A line.)

6. Next, examine the components for each source of variance. If a slash (/) appears in the subscript, define only the letters to the left of the slash as 'essential'. If there are several slashes, only the letters preceding the leftmost slash are essential. If there is no slash, all letters are essential.

7. Among the essential letters, ignore any that are necessary to designate the source of variance. (If the source of variance is *A*, for example, then when considering $n\sigma^2_{AB}$, ignore the *A*. If the source is *S*/*A*, then when considering the $\sigma^2_{SB/A}$ component, *S* and *B* are essential subscripts and *S* is to be ignored.) If any of the remaining (non-ignored) essential letters designate fixed variables, delete the entire component from the EMS.

An example:

Term	EMS so far
Step 1: identify variables and numbers of levels.	
A, a (between-subjects factor)	
B, b (within-subjects factor)	
S, n (number of subjects per group)	
Step 2: identify sources of variance.	
A	
S / A	
B	
BA	
SB / A	
Step 3: List σ^2_e as part of each EMS.	
A	σ^2_e
S / A	σ^2_e
B	σ^2_e
BA	σ^2_e
SB / A	σ^2_e
Step 4: list the null hypothesis component.	
A	$\sigma^2_e + nb\sigma^2_A$
S / A	$\sigma^2_e + b\sigma^2_{S/A}$
B	$\sigma^2_e + an\sigma^2_B$
BA	$\sigma^2_e + n\sigma^2_{BA}$
SB / A	$\sigma^2_e + \sigma^2_{SB/A}$
Step 5: add all components whose subscripts contain all the letters designating the source of variance in question.	
A	$\sigma^2_e + nb\sigma^2_A + b\sigma^2_{S/A} + n\sigma^2_{BA} + \sigma^2_{SB/A}$
S / A	$\sigma^2_e + b\sigma^2_{S/A} + \sigma^2_{SB/A}$
B	$\sigma^2_e + an\sigma^2_B + n\sigma^2_{BA} + \sigma^2_{SB/A}$
BA	$\sigma^2_e + n\sigma^2_{BA} + \sigma^2_{SB/A}$
SB / A	$\sigma^2_e + \sigma^2_{SB/A}$

Steps 6 and 7: for each component, define 'essential' letters; ignore any that are part of the designation of the source of variance; if any remaining essential letters contain fixed factors, delete the component.

A	$\sigma_e^2 + nb\sigma_A^2 + b\sigma_{S/A}^2$
S / A	$\sigma_e^2 + b\sigma_{S/A}^2$
B	$\sigma_e^2 + an\sigma_B^2 + \sigma_{SB/A}^2$
BA	$\sigma_e^2 + n\sigma_{BA}^2 + \sigma_{SB/A}^2$
SB / A	$\sigma_e^2 + \sigma_{SB/A}^2$

6.4.2 Choosing an error term

A mean square qualifies as an error term for testing an effect if its *E*(MS) matches the $E(\text{MS}_{\text{effect}})$ in all respects except the null-hypothesis component (Keppel, 1991, p. 568). In our example above, therefore, we'd test MS_A against $\text{MS}_{\text{S/A}}$, and we'd test both MS_B and MS_{BA} against $\text{MS}_{\text{SB/A}}$.

6.4.3 Pooling error terms

When we have random factors in a model, important variables are often tested against an interaction term. Since interaction terms have few *df* (and since power depends on *F* being large when the null hypothesis is false, and since *F* is the ratio of $\text{MS}_{\text{effect}}$ to MS_{error}, and since MS_{error} is $\text{SS}_{\text{error}} / df_{\text{error}}$), this means we may have poor power to detect such effects.

One possibility is to **test** interaction terms in a full model with a conservative criterion, like this (Howell, 1997, p. 425). If there is an interaction ($p < 0.05$), we declare that there's an interaction. If there isn't ($0.05 < p < 0.25$), we just look at the results for other terms. But if there is no interaction ($p > 0.25$), we **remove** the interaction term from the model. In the example above, if we found that the AB interaction was not significant ($p > 0.25$), we could remove any terms including it and its *df* would contribute to the within-subjects error term, which might increase power to detect effects of B (see p. 51).

6.5 Contrasts

See Howell (1997, pp. 354-369); Myers & Well (1995, chapter 6).

6.5.1. Linear contrasts

Linear contrasts are comparisons between linear combinations of different groups. Suppose we want to know whether students are more bored on Wednesdays than other weekdays, because Wednesday is statistics day, and whether they're more bored on weekdays than weekends. We could measure their boredom on all days of the week, and use DayOfWeek as a factor (with 7 levels) in an ANOVA. If this turned up significant, we would know that all days were not the same — but it wouldn't answer our original questions. We can do that with linear contrasts.

In general, a linear contrast is a linear combination of a set of treatment means. Each mean μ_j is *weighted* by a weight w_j:

$$L = w_1\mu_1 + w_2\mu_2 + \ldots + w_k\mu_k = \sum_j w_j\mu_j$$

$$\text{such that } \sum_j w_j = 0$$

In our example, suppose μ_1 is the Monday mean, μ_2 is the Tuesday mean, and so on. Our 'Wednesdays versus other weekdays' question can be written as a linear contrast:

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	$\sigma^2_e + \sigma^2_{SCD/AB}$	

EXPECTED MEAN SQUARES AND MIXED MODEL ANALYSES

Fixed vs. Random Effects

- The choice of labeling a factor as a fixed or random effect will affect how you will make the *F*-test.

- This will become more important later in the course when we discuss interactions.

Fixed Effect

- All treatments of interest are included in your experiment.

- You cannot make inferences to a larger experiment.

Example 1: An experiment is conducted at Fargo and Grand Forks, ND. If location is considered a fixed effect, you cannot make inferences toward a larger area (e.g. the central Red River Valley).

Example 2: An experiment is conducted using four rates (e.g. ½ X, X, 1.5 X, 2 X) of a herbicide to determine its efficacy to control weeds. If rate is considered a fixed effect, you cannot make inferences about what may have occurred at any rates not used in the experiment (e.g. ¼ x, 1.25 X, etc.).

Random Effect

- Treatments are a sample of the population to which you can make inferences.

- You can make inferences toward a larger population using the information from the analyses.

Example 1: An experiment is conducted at Fargo and Grand Forks, ND. If location is considered a random effect, you can make inferences toward a larger area (e.g. you could use the results to state what might be expected to occur in the central Red River Valley).

Example 2: An experiment is conducted using four rates (e.g. ½ X, X, 1.5 X, 2 X) of an herbicide to determine its efficacy to control weeds. If rate is considered a random effect, you can make inferences about what may have occurred at rates not used in the experiment (e.g. ¼ x, 1.25 X, etc.).

Assumptions Underlying ANOVA

- Experimental errors are random, independently, and normally distributed about a mean of zero and with a common variance (i.e. treatment variances are homogenous).

- The above assumption can be express as NID $(0, \sigma^2)$.

- Departure from this assumption can affect both the level of significance and the sensitivity of *F*- or *t*-tests to real departures from H_o:

- This results in the rejection of Ho when it is true (i.e. a Type I Error) more often than α calls for.

- The power of the test also is reduced if the assumption of NID$(0, \sigma^2)$ is violated.

- Violation of the assumption NID$(0, \sigma^2)$ with the fixed model is usually of little consequence because ANOVA is a very robust technique.

- Violation of the basic assumptions of ANOVA can be investigated by observing plots of the residuals.

- Residuals will be discussed in more detail when Transformations are discussed later in the semester.

- So far in class we have assumed that treatments are always a fixed effect.

- If some or all factors in an experiment are considered random effects, we need to be concerned about the denominator of the F-test because it may not be the Error MS.

- To determine the appropriate denominator of the F-test, we need to know how to write the Expected Mean Squares for all sources of variation.

Expected Mean Squares

All Random Model

- Each source of variation will consist of a linear combination of σ^2 plus variance components whose subscript matches at least one letter in the source of variation.

- The coefficients for the identified variance components will be the letters not found in the subscript of the variance components.

Example – RCBD with a 3x4 Factorial Arrangement

Sources of variation	σ^2	$r\sigma^2_{AB}$	$ra\sigma^2_B$	$rb\sigma^2_A$	$ab\sigma^2_R$
Rep	$\sigma^2 + ab\sigma^2_R$				
A	$\sigma^2 + r\sigma^2_{AB} + rb\sigma^2_A$				
B	$\sigma^2 + r\sigma^2_{AB} + ra\sigma^2_B$				
AxB	$\sigma^2 + r\sigma^2_{AB}$				
Error	σ^2				

Step 1. Write the list of variance components across the top of the table.

- There will be one variance component for each source of variation except Total.
- The subscript for each variance component will correspond to each source of variation.
- The variance component for error receives no subscript.

Sources of variation	σ^2	σ^2_{AB}	σ^2_B	σ^2_A	σ^2_R
Rep					
A					
B					
AxB					
Error					

Step 2. Write in the coefficients for each variance component.

- Remember that the coefficient is the letter(s) missing in the subscript.
- The coefficient for Error is the number 1.

Sources of variation	σ^2	$r\sigma^2_{AB}$	$ra\sigma^2_B$	$rb\sigma^2_A$	$ab\sigma^2_R$
Rep					
A					
B					
AxB					
Error					

Step 3. All sources of variation will have σ^2 (i.e. the expected mean square for error as a variance component).

Sources of variation	σ^2	$r\sigma^2_{AB}$	$ra\sigma^2_B$	$rb\sigma^2_A$	$ab\sigma^2_R$
Rep	σ^2				
A	σ^2				
B	σ^2				
AxB	σ^2				
Error	σ^2				

Step 4. The remaining variance components will be those whose subscript matches at least one letter in the corresponding source of variation.

SOV	σ^2	$r\sigma^2_{AB}$	$ra\sigma^2_B$	$rb\sigma^2_A$	$ab\sigma^2_R$
Rep	$\sigma^2 + ab\sigma^2_R$		(Those variance components that have at least the letter R)		
A	$\sigma^2 + r\sigma^2_{AB} + rb\sigma^2_A$		(Those variance components that have at least the letter A)		
B	$\sigma^2 + r\sigma^2_{AB} + ra\sigma^2_B$		(Those variance components that have at least the letter B)		
AxB	$\sigma^2 + r\sigma^2_{AB}$		(Those variance components that have at least the letters A and B)		
Error	σ^2				

Example – RCBD with a 3x2 Factorial Arrangement Combined Across Environments

Sources of variation	σ^2	$r\sigma^2_{LAB}$	$rl\sigma^2_{AB}$	$ra\sigma^2_{LB}$	$rla\sigma^2_B$	$rb\sigma^2_{LA}$	$rlb\sigma^2_A$
Location	--						
Rep(Loc)	--						
A	$\sigma^2 + r\sigma^2_{LAB} + rl\sigma^2_{AB} + rb\sigma^2_{LA} + rlb\sigma^2_A$						
Loc x A	$\sigma^2 + r\sigma^2_{LAB} + rb\sigma^2_{LA}$						
B	$\sigma^2 + r\sigma^2_{LAB} + rl\sigma^2_{AB} + ra\sigma^2_{LB} + rla\sigma^2_B$						
Loc x B	$\sigma^2 + r\sigma^2_{LAB} + ra\sigma^2_{LB}$						
AxB	$\sigma^2 + r\sigma^2_{LAB} + rl\sigma^2_{AB}$						
Loc x A x B	$\sigma^2 + r\sigma^2_{LAB}$						
Error	σ^2						

Step 1. Write the list of variance components across the top of the table.

- There will be one variance component for each source of variation except Total.
- The subscript for each variance component will correspond to each source of variation.

- The variance component for error receives no subscript.

Sources of variation	σ^2	σ^2_{LAB}	σ^2_{AB}	σ^2_{LB}	σ^2_{B}	σ^2_{LA}	σ^2_{A}
Location	--						
Rep(Loc)	--						
A							
Loc x A							
B							
Loc x B							
A x B							
Loc x A x B							
Error							

Step 2. Write in the coefficients for each variance component.

- Remember that the coefficient is the letter(s) missing in the subscript.
- The coefficient for Error is the number 1.

Sources of variation	σ^2	$r\sigma^2_{LAB}$	$rl\sigma^2_{AB}$	$ra\sigma^2_{LB}$	$rla\sigma^2_{B}$	$rb\sigma^2_{LA}$	$rlb\sigma^2_{A}$
Location	--						
Rep(Loc)	--						
A							
Loc x A							
B							
Loc x B							
A x B							
Loc x A x B							
Error							

Step 3. All sources of variation will have σ^2 (i.e. the expected mean square for error as a variance component).

Sources of variation	σ^2	$r\sigma^2_{LAB}$	$rl\sigma^2_{AB}$	$ra\sigma^2_{LB}$	$rla\sigma^2_{B}$	$rb\sigma^2_{LA}$	$rlb\sigma^2_{A}$
Loc	--						
Rep(Loc)	--						
A	σ^2						
Loc x A	σ^2						
B	σ^2						
Loc x B	σ^2						
A x B	σ^2						
Loc x A x B	σ^2						
Error	σ^2						

Step 4. The remaining variance components will be those whose subscript matches at least one letter in the corresponding source of variation.

SOV	σ^2 $r\sigma^2_{LAB}$ $rl\sigma^2_{AB}$ $ra\sigma^2_{LB}$ $rla\sigma^2_{B}$ $rb\sigma^2_{LA}$ $rlb\sigma^2_{A}$	
Loc	--	
Rep(Loc)	--	
A	$\sigma^2 + r\sigma^2_{LAB} + rl\sigma^2_{AB} + rb\sigma^2_{LA} + rlb\sigma^2_{A}$	(Those variance components that have at least the letter A)
LocxA	$\sigma^2 + r\sigma^2_{LAB} + rb\sigma^2_{LA}$	(Those variance components that have at least the letter L and A)
B	$\sigma^2 + r\sigma^2_{LAB} + rl\sigma^2_{AB} + ra\sigma^2_{LB} + rla\sigma^2_{B}$	(Those variance components that have at least the letters B)
LocxB	$\sigma^2 + r\sigma^2_{LAB} + ra\sigma^2_{LB}$	(Those variance components that have at least the letters L and B)
AxB	$\sigma^2 + r\sigma^2_{LAB} + rl\sigma^2_{AB}$	(Those variance components that have at least the letters A and B)
LocxAxB	$\sigma^2 + r\sigma^2_{LAB}$	(Those variance components that have at least the letters L, A, and B)
Error	σ^2	

<u>All Fixed Effect Model</u>

Step 1. Begin by writing the expected mean squares for an all random model.

Step 2. All but the first and last components will drop out for each source of variation.

Step 3. Rewrite the last term for each source of variation to reflect the fact that the factor is a fixed effect.

<u>Example RCBD with 3x2 Factorial</u>

SOV	Before		After
Rep	$\sigma^2 + ab\sigma^2_R$	$\longrightarrow$	$\sigma^2 + ab\sigma^2_R$
A	$\sigma^2 + r\sigma^2_{AB} + rb\sigma^2_A$	$\longrightarrow$	$\sigma^2 + rb\dfrac{\sum \alpha_i^2}{(a-1)}$
B	$\sigma^2 + r\sigma^2_{AB} + ra\sigma^2_B$	$\longrightarrow$	$\sigma^2 + ra\dfrac{\sum \beta_j^2}{(b-1)}$
AxB	$\sigma^2 + r\sigma^2_{AB}$	$\longrightarrow$	$\sigma^2 + r\dfrac{\sum (\alpha\beta)^2_{ij}}{(a-1)(b-1)}$
Error	σ^2	$\longrightarrow$	σ^2

<u>Rules for Writing Fixed Effect Component</u>

Step 1. Coefficients don’t change.

Step 2. Replace σ^2 with $\sum$

Step 3. The subscript of the variance component becomes the numerator of the effect.

Step 4. The denominator of the effect is the degrees of freedom.

Example 2 CRD with a Factorial Arrangement

SOV	Before	After
Loc	--	
Rep(Loc)	--	
A	$\sigma^2 + r\sigma^2_{LAB} + rl\sigma^2_{AB} + rb\sigma^2_{LA} + rlb\sigma^2_A$	$\longrightarrow \sigma^2 + rlb\frac{\sum \alpha^2}{(a-1)}$
LxA	$\sigma^2 + r\sigma^2_{LAB} + rb\sigma^2_{LA}$	$\longrightarrow \sigma^2 + rb\frac{\sum \lambda\alpha^2}{(l-1)(a-1)}$
B	$\sigma^2 + r\sigma^2_{LAB} + rl\sigma^2_{AB} + ra\sigma^2_{LB} + rla\sigma^2_B$	$\longrightarrow \sigma^2 + rla\frac{\sum (\beta)^2}{(b-1)}$
LxB	$\sigma^2 + r\sigma^2_{LAB} + ra\sigma^2_{LB}$	$\longrightarrow \sigma^2 + ra\frac{\sum (\lambda\beta)^2}{(l-1)(b-1)}$
AxB	$\sigma^2 + r\sigma^2_{LAB} + rl\sigma^2_{AB}$	$\longrightarrow \sigma^2 + rl\frac{\sum (\alpha\beta)^2}{(a-1)(b-1)}$
LxAxB	$\sigma^2 + r\sigma^2_{LAB}$	$\longrightarrow \sigma^2 + r\frac{\sum (\lambda\alpha\beta)^2}{(l-1)(a-1)(b-1)}$
Error	σ^2	$\longrightarrow \sigma^2$

Mixed Models

- For the expected mean squares for all random models, all variance components remained.
- For fixed effect models, all components but the first and last are eliminated.
- For mixed effect models:
    1. The first and last components will remain.
    2. Of the remaining components, some will be eliminated based on the following rules:
        a. Always ignore the first and last variance components.

b. For the remaining variance components, any letter(s) in the subscript used in naming the effect is ignored.

c. If any remaining letter(s) in the subscript corresponds to a fixed effect, that variance component drops out.

Example 1 – RCBD with a Factorial Arrangement (A fixed and B random)

SOV	Before		After
Rep	$\sigma^2 + ab\sigma_R^2$	→	$\sigma^2 + ab\sigma_R^2$
A	$\sigma^2 + r\sigma_{AB}^2 + rb\sigma_A^2$	→	$\sigma^2 + r\sigma_{AB}^2 + rb\frac{\sum \alpha_i^2}{(a-1)}$
B	$\sigma^2 + r\sigma_{AB}^2 + ra\sigma_B^2$	→	$\sigma^2 + ra\sigma_B^2$
AxB	$\sigma^2 + r\sigma_{AB}^2$	→	$\sigma^2 + r\sigma_{AB}^2$
Error	σ^2	→	σ^2

Steps for each Source of Variation

Error - No change for Error.

AxB - No change for AxB since only the first and last variance components are present.

B - For the middle variance component, cover up the subscript for B, only A is present.
Since A is a fixed effect this variance component drops out.

A - For the middle variance component, cover up the subscript for A, only B is present.
Since B is a random effect this variance component remains.

Rep - Replicate is always a random effect, so this expected mean square remains the same.

Example 2 RCBD with a Factorial Arrangement (A and B fixed) combined across locations (random)

SOV	Before		After
Loc	--		--
Rep(Loc)	--		--
A	$\sigma^2 + r\sigma^2_{LAB} + rl\sigma^2_{AB} + rb\sigma^2_{LA} + rlb\sigma^2_{A}$	$\rightarrow$	$\sigma^2 + rb\sigma^2_{LA} + rlb\frac{\sum \alpha^2}{(a-1)}$
LxA	$\sigma^2 + r\sigma^2_{LAB} + rb\sigma^2_{LA}$	$\rightarrow$	$\sigma^2 + rb\sigma^2_{LA}$
B	$\sigma^2 + r\sigma^2_{LAB} + rl\sigma^2_{AB} + ra\sigma^2_{LB} + rla\sigma^2_{B}$	$\rightarrow$	$\sigma^2 + ra\sigma^2_{LB} + rla\frac{\sum \beta^2}{(b-1)}$
LxB	$\sigma^2 + r\sigma^2_{LAB} + ra\sigma^2_{LB}$	$\rightarrow$	$\sigma^2 + ra\sigma^2_{LB}$
AxB	$\sigma^2 + r\sigma^2_{LAB} + rl\sigma^2_{AB}$	$\rightarrow$	$\sigma^2 + r\sigma^2_{LAB} + rl\frac{\sum (\alpha\beta)^2}{(a-1)(b-1)}$
LxAxB	$\sigma^2 + r\sigma^2_{LAB}$	$\rightarrow$	$\sigma^2 + r\sigma^2_{LAB}$
Error	σ^2	$\rightarrow$	σ^2

Steps for Each Source of Variation

Error - Error remains the same.

LxAxB - The error mean square for LxAxB remains the same since there are only first and last terms.

AxB- Cover up the A and B in the subscript, L remains and corresponds to a random effect. Therefore the term remains.

LxB - Cover up the L and B in the subscript, A remains and corresponds to a fixed effect. Therefore the term drops out.

B – LAB term - Cover up the B in the subscript, L and A remain and A corresponds to a fixed effect; therefore, the term drops out.

AxB term - Cover up the B in the subscript, A remains and corresponds to a fixed effect; therefore, the term drops out.

LxB term - Cover up the B in the subscript, L remains and corresponds to a random effect; therefore, the term remains.

LxA - Cover up the L and A in the subscript, B remains and corresponds to a fixed effect. Therefore the term drops out.

A – LAB term - Cover up the A in the subscript, L and B remain and B corresponds to a fixed effect; therefore, the term drops out.

AxB term - Cover up the A in the subscript, B remains and corresponds to a fixed effect; therefore, the term drops out.

LxA term - Cover up the A in the subscript, L remains and corresponds to a random effect; therefore, the term remains.

Deciding What to Use as the Denominator of Your *F*-test

- For an all fixed model the Error MS is the denominator of all *F*-tests.

- For an all random or mix model,

  1. Ignore the last component of the expected mean square.

  2. Look for the expected mean square that now looks this expected mean square.

  3. The mean square associated with this expected mean square will be the denominator of the *F*-test.

  4. If you can't find an expected mean square that matches the one mentioned above, then you need to develop a Synthetic Error Term.

Example 1 – RCBD with a Factorial Arrangement (A fixed and B random)

SOV	Expected mean square	MS	*F*-test
Rep	$\sigma^2 + ab\sigma_R^2$	1	*F* = MS 1/MS 5
A	$\sigma^2 + r\sigma_{AB}^2 + rb\frac{\sum \alpha_i^2}{(a-1)}$	2	*F* = MS 2/MS 4
B	$\sigma^2 + ra\sigma_B^2$	3	*F* = MS 3/MS 5
AxB	$\sigma^2 + r\sigma_{AB}^2$	4	*F* = MS 4/MS 5
Error	σ^2	5	

Steps for *F*-tests

F_{AB} - Ignore $r\sigma_{AB}^2$. The expected mean square now looks like the expected mean square for Error. Therefore, the denominator of the *F*-test is the Error MS.

F_B - Ignore $ra\sigma_B^2$. The expected mean square now looks like the expected mean square for Error. Therefore, the denominator of the *F*-test is the Error MS.

F_A - Ignore $rb\frac{\sum \alpha_i^2}{(a-1)}$. The expected mean square now looks like the expected mean square for AxB. Therefore, the denominator of the *F*-test is the AxB MS.

Example 2 RCBD with a Factorial Arrangement (A and B fixed) combined across locations

SOV	Expected mean square	MS	*F*-test
Loc	--		non-valid
Rep(Loc)	--		non-valid
A	$\sigma^2 + rb\sigma^2_{LA} + rlb\frac{\sum \alpha^2}{(a-1)}$	1	*F* = MS 1/MS 2
LxA	$\sigma^2 + rb\sigma^2_{LA}$	2	*F* = MS 2/MS 7
B	$\sigma^2 + ra\sigma^2_{LB} + rla\frac{\sum \beta^2}{(b-1)}$	3	*F* = MS 3/MS 4
LxB	$\sigma^2 + ra\sigma^2_{LB}$	4	*F* = MS 4/MS 7
AxB	$\sigma^2 + r\sigma^2_{LAB} + rl\frac{\sum (\alpha\beta)^2}{(a-1)(b-1)}$	5	*F* = MS 5/MS 6
LxAxB	$\sigma^2 + r\sigma^2_{LAB}$	6	*F* = MS 6/MS 7
Error	σ^2	7	

Steps for *F*-tests

F_{LAB} - Ignore $r\sigma^2_{LAB}$. The expected mean square now looks like the expected mean square for Error. Therefore, the denominator of the *F*-test is the Error MS.

F_{AB} - Ignore $rl\sigma^2_{AB}$. The expected mean square looks like the expected mean square for LxAxB. Therefore, the denominator of the *F*-test is the LxAxB MS.

F_{LB} - Ignore $ra\sigma^2_{LB}$. The expected mean square now looks like the expected mean square for Error. Therefore, the denominator of the *F*-test is the Error MS.

F_B - Ignore $ra\sigma^2_B$. The expected mean square now looks like the expected mean square for LxB. Therefore, the denominator of the *F*-test is the LxB MS.

F_{LA} - Ignore $rb\sigma^2_{LA}$. The expected mean square now looks like the expected mean square for Error. Therefore, the denominator of the *F*-test is the Error MS.

F_A - Ignore $rb\frac{\sum \alpha_i^2}{(a-1)}$. The expected mean square now looks like the expected mean square for AxB. Therefore, the denominator of the *F*-test is the AxB MS.

Example 3 CRD with a Factorial Arrangement (A fixed, B and C random)

SOV	Expected Mean Square	MS	*F*-test
A	$\sigma^2 + r\sigma^2_{ABC} + rb\sigma^2_{AC} + rc\sigma^2_{AB} + rbc\frac{\sum \alpha_i^2}{(a-1)}$	1	(MS 1 + MS 7)/(MS 4 + MS 5)
B	$\sigma^2 + ra\sigma^2_{BC} + rac\sigma^2_B$	2	MS 2/MS 6
C	$\sigma^2 + ra\sigma^2_{BC} + rab\sigma^2_C$	3	MS 3/MS 6
AxB	$\sigma^2 + r\sigma^2_{ABC} + rc\sigma^2_{AB}$	4	MS 4/MS 7
AxC	$\sigma^2 + r\sigma^2_{ABC} + rb\sigma^2_{AC}$	5	MS 5/MS 7
BxC	$\sigma^2 + ra\sigma^2_{BC}$	6	MS 6/MS 8
AxBxC	$\sigma^2 + r\sigma^2_{ABC}$	7	MS 7/MS 8
Error	σ^2	8	

Steps for *F*-tests

F_{ABC} - Ignore $r\sigma^2_{ABC}$. The expected mean square now looks like the expected mean square for Error. Therefore, the denominator of the F-test is the Error MS.

F_{BC} - Ignore $ra\sigma^2_{BC}$. The expected mean square now looks like the expected mean square for Error. Therefore, the denominator of the F-test is the Error MS.

F_{AC} - Ignore $rb\sigma^2_{AC}$. The expected mean square now looks like the expected mean square for AxBxC. Therefore, the denominator of the F-test is the AxBxC MS.

F_{AB} - Ignore $rcb\sigma^2_{AB}$. The expected mean square now looks like the expected mean square for AxBxC. Therefore, the denominator of the F-test is the AxBxC MS.

F_C - Ignore $rab\sigma^2_C$. The expected mean square now looks like the expected mean square for BxC. Therefore, the denominator of the F-test is the BxC MS.

F_B - Ignore $rac\sigma^2_B$. The expected mean square now looks like the expected mean square for BxC. Therefore, the denominator of the F-test is the BxC MS.

F_A - Ignore $rbc\frac{\sum \alpha_i^2}{(a-1)}$. The expected mean square now looks like none of the expected mean squares. Therefore, we need to develop a Synthetic Mean Square

- Need an Expected Mean Square that looks like: $\sigma^2 + r\sigma^2_{ABC} + rb\sigma^2_{AC} + rc\sigma^2_{AB}$.

AC = $\sigma^2 + r\sigma^2_{ABC} + rb\sigma^2_{AC}$ (missing $rc\sigma^2_{AB}$)

and

AB = $\sigma^2 + r\sigma^2_{ABC} + rc\sigma^2_{AB}$ (missing $rb\sigma^2_{AC}$)

- An expected mean square can be found that includes all needed variance components if you sum the expected mean squares of AC and AB.

$$\text{AC MS + AB MS} = 2\sigma^2 + 2r\sigma^2_{ABC} + rb\sigma^2_{AC} + rc\sigma^2_{AB}$$

- The problem with this sum is that it is too large by $\sigma^2 + r\sigma^2_{ABC}$.

- One method would be to get the needed expected mean square is by:

  AC MS + AB MS – ABC MS

  - Thus F_A could be: $\dfrac{\text{A MS}}{\text{AC MS} + \text{AB MS} - \text{ABC MS}}$

- However, this is not the preferred formula for this F-test.

- The most appropriate F-test is one in which the number of MS used in the numerator and denominator are similar.

- This allows for more accurate estimates of the degrees of freedom associate with the numerator and denominator.

- The formula above has one mean square in the numerator and three in the denominator.

  - The formula for F_A that is most appropriate is

$$\frac{\text{A MS} + \text{ABC MS}}{\text{AC MS} + \text{AB MS}}$$

- The numerator and the denominator then become: $2\sigma^2 + 2r\sigma^2_{ABC} + rb\sigma^2_{AC} + rc\sigma^2_{AB}$.

Calculation of Estimated Degrees of Freedom for the Synthetic Error Term

- Calculation of degrees of freedom for the numerator and denominator of the F-test cannot be calculated by adding together the degrees of freedom for the associated mean squares.

- For the F-test: $F_A = \frac{A\,MS + ABC\,MS}{AC\,MS + AB\,MS}$

- The numerator degrees of freedom = $\frac{(A\,MS + ABC\,MS)^2}{\left[\frac{(A\,MS)^2}{A\,df} + \frac{(ABC\,MS)^2}{ABC\,df}\right]}$

- The denominator degrees of freedom = $\frac{(AC\,MS + AB\,MS)^2}{\left[\frac{(AC\,MS)^2}{AC\,df} + \frac{(AB\,MS)^2}{AB\,df}\right]}$

Calculation of LSD Values – CRD with a Factorial Arrangement (A fixed, B and C Random)

LSD_{ABC} (0.05) = $t_{0.05/2;\,Error\,df}\sqrt{\frac{2Error\,MS}{r}}$

LSD_{BC} (0.05) = $t_{0.05/2;\,Error\,df}\sqrt{\frac{2Error\,MS}{ra}}$

LSD_{AC} (0.05) = $t_{0.05/2;\,ABC\,df}\sqrt{\frac{2(ABC\,MS)}{rb}}$

LSD_{AB} (0.05) = $t_{0.05/2;\,ABC\,df}\sqrt{\frac{2(ABC\,MS)}{rc}}$

LSD_{C} (0.05) = $t_{0.05/2;\,BC\,df}\sqrt{\frac{2(BC\,MS)}{rab}}$

LSD_{B} (0.05) = $t_{0.05/2;\,BC\,df}\sqrt{\frac{2(BC\,MS)}{rac}}$

LSD_{A} (0.05) = $t'_{0.05/2;\,Estimated\,df}\sqrt{\frac{2(AC\,MS + AB\,MS - ABC\,MS)}{rbc}}$

Where Estimated df for t' = $\frac{(ACMS + ABMS - ABC)^2}{\left[\frac{(ACMS)^2}{AC\,df} + \frac{(ABMS)^2}{AB\,df} + \frac{(ABCMS)^2}{ABC\,df}\right]}$

<u>SAS Example for ana Mixed Model RCBD Combined Across Locations (Factor A fixed and Locations random)</u>

```
options pageno=1;
data threefct;
input Loc    Rep    A      Yield;
datalines;
0       1       1       25.7
0       1       2       31.8
0       1       3       34.6
0       1       4       27.7
0       1       5       38
0       1       6       42.1
0       2       1       25.4
0       2       2       29.5
0       2       3       37.2
0       2       4       30.3
0       2       5       40.6
0       2       6       43.6
0       3       1       23.8
0       3       2       28.7
0       3       3       29.1
0       3       4       30.2
0       3       5       34.6
0       3       6       44.6
0       4       1       22
0       4       2       26.4
0       4       3       23.7
0       4       4       33.2
0       4       5       31
0       4       6       42.7
1       1       1       48.9
1       1       2       67.5
1       1       3       58.4
1       1       4       35.8
1       1       5       66.9
1       1       6       44.2
1       2       1       64.7
1       2       2       71.5
1       2       3       42.5
1       2       4       31
1       2       5       81.9
1       2       6       61.6
1       3       1       27.8
1       3       2       31
1       3       3       31.2
1       3       4       29.5
1       3       5       31.5
1       3       6       38.9
```

```
1	4	1	23.4
1	4	2	27.8
1	4	3	29.8
1	4	4	30.7
1	4	5	35.9
1	4	6	37.6
2	1	1	23.4
2	1	2	25.3
2	1	3	29.8
2	1	4	20.8
2	1	5	29
2	1	6	36.6
2	2	1	24.2
2	2	2	27.7
2	2	3	29.9
2	2	4	23
2	2	5	32
2	2	6	37.8
2	3	1	21.2
2	3	2	23.7
2	3	3	24.3
2	3	4	25.2
2	3	5	26.5
2	3	6	34.8
2	4	1	20.9
2	4	2	24.3
2	4	3	23.8
2	4	4	23.1
2	4	5	31.2
2	4	6	40.2
;;
ods rtf file='cmbloc.rtf';
proc glm;
class loc rep a;
model yield=loc rep(loc) a loc*a/ss3;
test h=a e=loc*a;
means a/lsd e=loc*a;
means loc*a;
run;
ods rtf close;
```

Analysis of Mixed Models Using Proc GLM

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
Loc	3	0 1 2
Rep	4	1 2 3 4
A	6	1 2 3 4 5 6

Number of Observations Read	72
Number of Observations Used	72

Analysis of Mixed Models Using Proc GLM

The GLM Procedure

Dependent Variable: Yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	26	9971.27750	383.51067	9.65	<.0001
Error	45	1789.24250	39.76094		
Corrected Total	71	11760.52000			

R-Square	Coeff Var	Root MSE	Yield Mean
0.847860	18.26836	6.305628	34.51667

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Loc	2	3358.260833	1679.130417	42.23	<.0001
Rep(Loc)	9	3930.862500	436.762500	10.98	<.0001
A	5	1847.900000	369.580000	9.30	<.0001
Loc*A	10	834.254167	83.425417	2.10	0.0446

Tests of Hypotheses Using the Type III MS for Loc*A as an Error Term					
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
A	5	1847.900000	369.580000	4.43	0.0218

Analysis of Mixed Models Using Proc GLM

The GLM Procedure

t Tests (LSD) for Yield

Note: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	10
Error Mean Square	83.42542
Critical Value of t	2.22814
Least Significant Difference	8.3084

Means with the same letter are not significantly different.					
t Grouping			**Mean**	**N**	**A**
	A		42.058	12	6
	A				
B	A		39.925	12	5
B	A				
B	A	C	34.600	12	2
B		C			
B		C	32.858	12	3
		C			
		C	29.283	12	1
		C			
		C	28.375	12	4

Analysis of Mixed Models Using Proc GLM

The ANOVA Procedure

Level of Loc	Level of A	N	Yield	
			Mean	Std Dev
0	**1**	4	24.2250000	1.7017148
0	**2**	4	29.1000000	2.2286020
0	**3**	4	31.1500000	6.0058305
0	**4**	4	30.3500000	2.2487033
0	**5**	4	36.0500000	4.1677332
0	**6**	4	43.2500000	1.0908712
1	**1**	4	41.2000000	19.2175267
1	**2**	4	49.4500000	23.2460032
1	**3**	4	40.4750000	13.2336377
1	**4**	4	31.7500000	2.7766887
1	**5**	4	54.0500000	24.3493326
1	**6**	4	45.5750000	11.0581418
2	**1**	4	22.4250000	1.6255768
2	**2**	4	25.2500000	1.7616280
2	**3**	4	26.9500000	3.3550956
2	**4**	4	23.0250000	1.7969882
2	**5**	4	29.6750000	2.4676237
2	**6**	4	37.3500000	2.2649503

PROC MIXED for Analysis of Mixed Models

- The SAS procedure PROC MIXED offers an alternative for the analysis of mixed models.

- This method of analysis is becoming the preferred or even required method of analysis in some journals (e.g. *Agronomy Journal* and *Crop Science*).

- The PROC MIXED procedure uses a method of analysis called "Restricted or residual maximum likelihood" (REML) as its default method of analysis.

- The SAS manual (http://www.technion.ac.il/docs/sas/stat/chap41/sect1.htm) states for PROC MIXED that "A *mixed linear model* is a generalization of the standard linear model used in the GLM procedure, the generalization being that the data are permitted to exhibit correlation and nonconstant variability. The mixed linear model, therefore, provides you with the flexibility of modeling not only the means of your data (as in the standard linear model) but their variances and covariances as well."

- For many of us, the need for covariance parameters will be of minimal importance.

- Research areas where covariance parameters can be important include:

    - The experimental units from which you collect data can be grouped into clusters and the data from a common cluster are correlated.

    - Collecting data over time from the same experimental unit and the repeated measurements are correlated.

- The output from the PROC MIXED analysis is quite different from that produced using PROC GLM.

- Major differences will be the *F*-tests **are done only** on the sources of variation have all fixed parameters and mean separation is done only on fixed effects. *F*-tests and mean separation **are not done** on random or mixed effects.

- If the data are balanced, the results for the *F*-tests on the fixed effect sources of variation in the PROC MIXED analysis will be the same as those obtained from the PROC GLM analysis.

- If the data are unbalanced, the results for the *F*-tests from the two procedures will be different.

Example of SAS PROC MIXED to Analyze a Single Factor Experiment Over Multiple Location (A is a fixed a effect and location is a random effect).

```
options pageno=1;
data threefct;
input Loc   Rep   A     Yield;
datalines;
0      1      1      25.7
0      1      2      31.8
0      1      3      34.6
0      1      4      27.7
0      1      5      38
0      1      6      42.1
0      2      1      25.4
0      2      2      29.5
0      2      3      37.2
0      2      4      30.3
0      2      5      40.6
0      2      6      43.6
0      3      1      23.8
0      3      2      28.7
0      3      3      29.1
0      3      4      30.2
0      3      5      34.6
0      3      6      44.6
0      4      1      22
0      4      2      26.4
0      4      3      23.7
0      4      4      33.2
0      4      5      31
0      4      6      42.7
1      1      1      48.9
1      1      2      67.5
1      1      3      58.4
1      1      4      35.8
1      1      5      66.9
1      1      6      44.2
1      2      1      64.7
1      2      2      71.5
1      2      3      42.5
1      2      4      31
1      2      5      81.9
1      2      6      61.6
1      3      1      27.8
1      3      2      31
1      3      3      31.2
1      3      4      29.5
1      3      5      31.5
1      3      6      38.9
1      4      1      23.4
1      4      2      27.8
1      4      3      29.8
1      4      4      30.7
1      4      5      35.9
1      4      6      37.6
2      1      1      23.4
```

```
2       1       2       25.3
2       1       3       29.8
2       1       4       20.8
2       1       5       29
2       1       6       36.6
2       2       1       24.2
2       2       2       27.7
2       2       3       29.9
2       2       4       23
2       2       5       32
2       2       6       37.8
2       3       1       21.2
2       3       2       23.7
2       3       3       24.3
2       3       4       25.2
2       3       5       26.5
2       3       6       34.8
2       4       1       20.9
2       4       2       24.3
2       4       3       23.8
2       4       4       23.1
2       4       5       31.2
2       4       6       40.2
;;
ods rtf file='cmbloc.rtf';
proc mixed;
class loc rep a;
model yield= a;
*comments In Proc Mixed only effects that are solely fixed are included
in the model statement;
random loc rep(loc) loc*a;
*Comment In Proc Mixed the Random statement includes sources of
variation that include solely random effects or mixed effects;
lsmeans a/diff;
title 'ANOVA done using Proc Mixed';
run;
ods rtf close;
```

ANOVA Done Using Proc Mixed

The Mixed Procedure

Model Information	
Data Set	WORK.THREEFCT
Dependent Variable	Yield
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information		
Class	**Levels**	**Values**
Loc	3	0 1 2
Rep	4	1 2 3 4
A	6	1 2 3 4 5 6

Dimensions	
Covariance Parameters	4
Columns in X	7
Columns in Z	33
Subjects	1
Max Obs Per Subject	72

Number of Observations	
Number of Observations Read	72
Number of Observations Used	72
Number of Observations Not Used	0

ANOVA Done Using Proc Mixed

The Mixed Procedure

Iteration History			
Iteration	**Evaluations**	**-2 Res Log Like**	**Criterion**
0	1	532.99533545	
1	1	481.74525633	0.00000000

Convergence criteria met.

Covariance Parameter Estimates	
Cov Parm	**Estimate**
Loc	49.9460
Rep(Loc)	66.1669
Loc*A	10.9161
Residual	39.7609

Fit Statistics	
-2 Res Log Likelihood	481.7
AIC (smaller is better)	489.7
AICC (smaller is better)	490.4
BIC (smaller is better)	486.1

Type 3 Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
A	5	10	4.43	0.0218

Least Squares Means						
Effect	**A**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
A	1	29.2833	5.3958	10	5.43	0.0003
A	2	34.6000	5.3958	10	6.41	<.0001
A	3	32.8583	5.3958	10	6.09	0.0001

ANOVA Done Using Proc Mixed

The Mixed Procedure

Least Squares Means						
Effect	**A**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
A	4	28.3750	5.3958	10	5.26	0.0004
A	5	39.9250	5.3958	10	7.40	<.0001
A	6	42.0583	5.3958	10	7.79	<.0001

Differences of Least Squares Means							
Effect	**A**	**_A**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
A	1	2	-5.3167	3.7288	10	-1.43	0.1844
A	1	3	-3.5750	3.7288	10	-0.96	0.3603
A	1	4	0.9083	3.7288	10	0.24	0.8125
A	1	5	-10.6417	3.7288	10	-2.85	0.0171
A	1	6	-12.7750	3.7288	10	-3.43	0.0065
A	2	3	1.7417	3.7288	10	0.47	0.6505
A	2	4	6.2250	3.7288	10	1.67	0.1260
A	2	5	-5.3250	3.7288	10	-1.43	0.1838
A	2	6	-7.4583	3.7288	10	-2.00	0.0734
A	3	4	4.4833	3.7288	10	1.20	0.2569
A	3	5	-7.0667	3.7288	10	-1.90	0.0873
A	3	6	-9.2000	3.7288	10	-2.47	0.0333
A	4	5	-11.5500	3.7288	10	-3.10	0.0113
A	4	6	-13.6833	3.7288	10	-3.67	0.0043
A	5	6	-2.1333	3.7288	10	-0.57	0.5799