BIOMETRICS INFORMATION

(You're 95% likely to need this information)

PAMPHLET NO. # 48 DATE: January 20, 1995

SUBJECT: ANOVA: Why a fixed effect is tested by its interaction with a random effect

Many factorial ANOVA's contain both random and fixed factors in the model. Treatments are usually fixed factors while factors like blocks, seedlots, and families are often random and are included as a form of replication so that treatment effects can be generalized. The actual blocks, seedlots, or families used are considered to be a sample from a larger set (population) that could have been used in the experiment, and this sampling justifies the generalization of the results to these other blocks, seedlots, or families. One of the perplexing things about these factorial designs is that the treatment is tested by its interaction with the random factor and not by the usual experimental error. This pamphlet will discuss why this makes sense.

For example, suppose we are interested in a study with the factors seedlot and a treatment of three watering regimes. Further, suppose that the researcher has a total population of 18 seedlots to consider. For discussion purposes[1], we will assume that we know the actual response of each seedlot to each of three watering regimes and that these values are:

Watering Regime	Seedlot 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Mean
1	7	6	5	7	6	5	4	4	4	1	2	3	4	4	4	1	2	3	4
2	4	4	4	1	2	3	7	6	5	7	6	5	1	2	3	4	4	4	4
3	1	2	3	4	4	4	1	2	3	4	4	4	7	6	5	7	6	5	4
Mean	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4

Notice that the variability between seedlots and the treatment main effects are nil because all the marginal means, i.e., the treatment and seedlot means, have the same value. Nevertheless, the means inside the table are not all the same since there is an interaction between seedlot and watering regime. This occurs because seedlots respond differently to the treatments.

If it were possible to study all 18 seedlots then both seedlot and watering regime would be fixed factors and we would expect that any observed means from an experiment, for instance, would follow the pattern of the above means within experimental error. On the other hand, if including all seedlots is impractical, then a random set of, say, 3 seedlots out of the 18 could be chosen. Seedlot would now be a random factor.

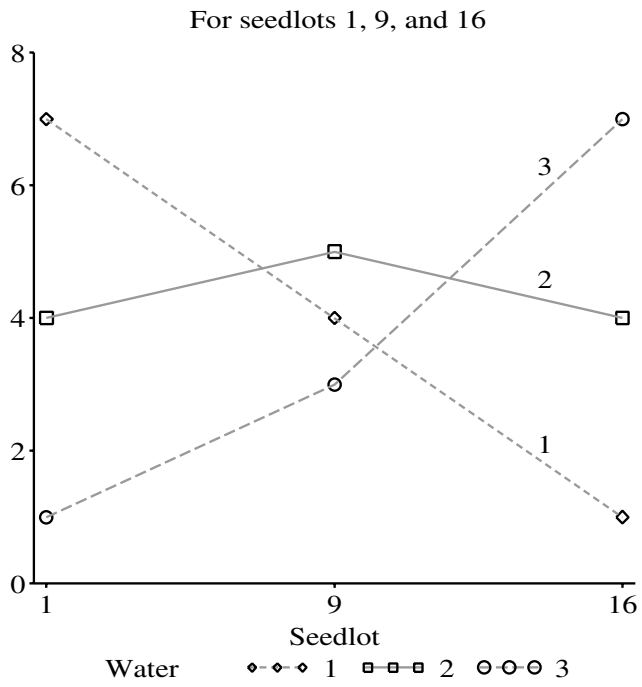
[1] The first part of this discussion relies on Maxwell and Delaney, pages 427 to 429, with the table of actual means taken directly from page 428.

Suppose that seedlots 1, 9, and 16 are chosen for study. The resulting sub-table is:

Watering Regime	Seedlot 1	9	16	Mean
1	7	4	1	4.00
2	4	5	4	4.33
3	1	3	7	3.67
Mean	4	4	4	4.00

Notice that the sampling has not changed the seedlot marginal means, nor the overall mean, but it has introduced variability into the watering regime means. They no longer have the same value of four and we would expect the observed means to reflect these changes. **Sampling seedlots has added variability to the treatment means.** The consequence is that the treatment differences should not be tested by the usual experimental error of a factorial ANOVA with only fixed factors. As we will see below, the correct error term for this example is the seedlot/treatment interaction. This error term takes into account the additional source of variability introduced by the sampling of seedlots.

To more clearly illustrate the difference between fixed and random effects, consider a study intended to test **only** seedlots 1, 9, and 16[2]. Seedlot would now be a fixed factor. In this case, an interaction between seedlot and treatment can be usefully interpreted. We could say, for instance, that seedlots 1, 9, and 16 responded similarly to watering regime 2 (see graph) although seedlot 1 responded better to watering regime 1 while seedlot 16 did the best with watering regime 3. These conclusions could be used during future work on these particular seedlots. But if the seedlots used in the study are only a sample of the seedlots we could have used, and if we want to generalize these results to other seedlots, then only large patterns in the seedlot/treatment interaction can be interpreted in a meaningful way. The above

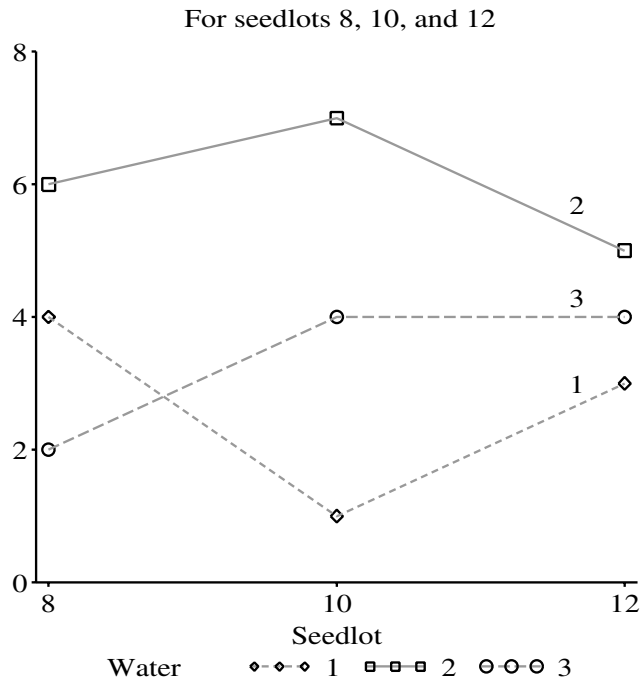


[2] I believe that this rarely occurs in practice. People will claim that they are not interested in generalizing the results to other seedlots (thus treating them as fixed), but I have yet to meet someone who does a study of this type without wanting to generalize, at least subconsciously, to other 'similiar' situations. And even if they can convince me that the only interest is in a limited subset of seedlots (blocks, families, etc.), others interested in their results are likely to generalize the results beyond those included in the experiment.

description of the interaction between seedlots and treatments is of little help in predicting the responses of other seedlots to the treatments. Would a new seedlot behave as did seedlot 1, seedlot 9, or seedlot 16? How would we know which pattern to pick? For that matter, might not a new seedlot exhibit a different pattern of response to the treatments?

For results to be generalized to seedlots **not** included in the study, treatment effects must be consistent from one seedlot to another. Another way to say this is that treatment effects must be large relative to the random interactions. A treatment by seedlot interaction occurs when the differences between treatment means change from one seedlot to the next. Nevertheless, if the relative rankings of the treatment responses remain similar from one seedlot to the next, even in the presence of an interaction, then the differences between treatment means will be substantially larger and consistent than the differences due to the interaction.

Looking at the example, we see that seedlots 1, 9, and 16 do not show this sort of consistent pattern. But suppose that we had used seedlots 8, 10 and 12. In this case, treatment 2 shows a consistently higher response than treatments 1 and 3 for the seedlots even though the difference between the means changes with seedlot (see graph on left). Treatments 1 and 3 do not show a consistent pattern. An experiment on these three seedlots would be expected to show the same pattern in the observed means. If the observed pattern is the same and is supported by the appropriate F-test of the experiment then we might argue that this pattern would also be true for other seedlots[3]. The test for treatment using the seedlot/treatment interaction is a test for large consistent differences between treatments regardless of seedlot variations in response.



Notice that for this example, we are reaching an incorrect conclusion (Type I error). It just happens that this small sample of seedlots shows a consistently high response to treatment 2. This type of error is less likely to occur when larger samples of seedlots are used in experiments.

The next step is to examine the Expected Mean Squares for the ANOVA that we will use to test for treatments differences. The table on the first page presented actual population means. To

[3] If the difference between any two treatments varies with seedlot, then we can only make weak statements about the specific differences between treatments for other seedlots.

study these means suppose that 5 pots each containing 1 seedling are randomly assigned both a seedlot and a watering regime treatment. There is an experimental error due to the variability between pots/seedlings. The study includes a total of 45 pots in a completely randomized design. The ANOVA table is:

Source of Variation		df	Expected Mean Square[4]	Error
Seedlot	S	2	$\sigma_e^2 +$ $+ 15\sigma_s^2$	pots
Treatment	T	2	$\sigma_e^2 + 5\sigma_{st}^2 + 15\sigma_t^2$	S x T
S x T		4	$\sigma_e^2 + 5\sigma_{st}^2$	pots
Pots	P(ST)	36	σ_e^2	--

The error term to be used in the denominator of the F-test for any particular source is found by noting which other source in the table has the same components for its Expected Mean Square (EMS) **except** for the variance component of interest[5]. For example, the test for treatment should be $\sigma_e^2 + 5\sigma_{st}^2$. The term with this EMS is S x T. Similarly, both seedlot[6] and S x T would be tested by pots (P(ST)), and there is no term available to test Pots because there is only one seedling per pot. If seedlot had been a fixed factor then the term $5\sigma_{st}^2$ would not have appeared in the EMS for treatment and thus treatment would also have been tested by the P(ST) (the usual experimental error). **The inclusion of $5\sigma_{st}^2$ in the EMS accounts for the extra variability between treatment means injected by the sampling of seedlots from a population of seedlots.**

The determination of a factor as fixed or random can make substantial changes to the appropriate ANOVA and to the conclusions of a study. Accordingly, their effect should be carefully considered when designing a study. For the example study, 45 pots may have seemed sufficient replication. But the main source of replication for the treatments was the seedlots, of which there were only three. Thus the degrees of freedom for the treatment test is 2 and 4 instead of 2 and 36. Thus a better design would contain more seedlots and perhaps fewer pots per seedlot/treatment combination.

---

[4] Note that the terms, σ_e^2 , σ_{st}^2 , σ_t^2 , and σ_s^2 have the usual meanings, that is, that they are the components of variance for the experimental error, the seedlot/treatment interaction, the treatment effect and the seedlot effect respectively.

[5] Determining the correct error term from an EMS table has also been discussed in BI #19 and #40.

[6] There is some controversy about the test for seedlot. It is argued by some that the S x T variance component should also be included in its EMS so that seedlot should be tested by S x T. In most cases, the test for seedlot is of little interest so that a definitive answer is not so important. See Schwarz (1993) for a recent discussion.

To further examine the influence of a crossed random factor, let us consider what happens to the example experiment if only one seedlot had been used instead of three. The previous ANOVA table becomes:

Source of Variation		df	Expected Mean Square	Error
Seedlot	S	0	$\sigma_e^2 +$ $+ 45\sigma_s^2$	pots
Treatment	T	2	$\sigma_e^2 + 1\sigma_{st}^2 + 15\sigma_t^2$	S x T
S x T		0	$\sigma_e^2 + 1\sigma_{st}^2$	pots
Pots	P(ST)	42	σ_e^2	--

While S x T is still the correct error term for the treatment, it is no longer possible to estimate its mean square for use in a test because it has zero degrees of freedom. Thus the only test available for the treatment is P(ST). There are two ways to justify the use of this error term. The first is to restrict the study results and inferences to the seedlot used in the experiment. The other is to assume that there is no seedlot by treatment interaction (that is $\sigma_{st}^2 = 0$); however, while this assumption makes an analysis possible, it may be difficult to justify on biological grounds. If it is not reasonable then the probability of finding treatment differences can be much greater than it should be.

If we had not thought to consider seedlot as a factor in the experiment then we would have used the following ANOVA table.

Source of Variation		df	Expected Mean Square	Error
Treatment	T	2	$\sigma_e^2 + 15\sigma_t^2$	P(T)
Pots	P(T)	42	σ_e^2	--

As this appears to be a simple Completely Randomized design, we would use P(T) to test treatment. We might not have recognized that we were assuming that treatments had no interaction with seedlots and/or that the results were limited to the seedlot used.

This example points out the need to clearly understand which factors/sources of variation have not been included in a study and how this limits the ability to generalize the results. Obviously, it is not practical to include ALL potential sources of variation in a study but it is instructive to consider how a crossed random factor can increase the variability between the treatment means. If this random factor is not taken into account when testing for treatment differences it is likely that differences will be found more often than they should be. That is, the probability level for the Type I error will not be 0.05, for instance, but will be substantially greater.

This was a simple example with only one random crossed factor. The situation is more complicated with two or more crossed random factors, usually leading to approximate or pseudo F-tests (briefly described in BI # 18).

References:

Keppel, G., 1973, Design and Analysis: A Researcher's Handbook, Prentice-Hall, Inc., Englewood Cliffs, N.J., see Chapter 16, pp. 333-338.

Keppel, G., 1982, Design and Analysis: A Researcher's Handbook, 2nd ed., Prentice-Hall, Inc., Englewood Cliffs, N.J., see Chapter 21.

Maxwell, S.E. and H.D. Delaney, 1990, Designing Experiments and Analyzing Data: A model comparison perspective, Wadsworth Publ. Co. Belmont, Calif. see Chap. 10.

Milliken, G.A. and D.E. Johnson, 1984, Analysis of Messy Data, Volume 1: Designed Experiments, Lifetime Learning Publications, Belmont, Calif. see pg 51, Chaps 18 to 23

Schwarz, C. J. (1993) "The Mixed-Model ANOVA: The Truth, the Computer Packages, the Books (Part I: Balanced Data)", The American Statistician, 47, pp 48-59.

Searle, S.R., 1971, Linear Models. John Wiley & Sons, New York, N.Y. (see section 9.3, p. 382)

Winer, B.J., 1971, Statistical Principles in Experimental Design, 2nd ed., McGraw-Hill Book Co., New York, N.Y.

Contact: Wendy Bergerud
387-5676

PROBLEM FROM BI #47

Use `do plot = 1, 3, 4, 5, 6, 8;` or use `do plot = 1, 3 to 6, 8;`. The `do` statement can include lists of values (separated by commas) or ranges like `3 to 12 by 3`, or combinations of both.