Weed Science, 45:745. 1997

My view

D. Palmquist
USDA-ARS
Reno, NV 89512

When analyzing data using an analysis of variance (ANOVA), a researcher must be mindful of some simple considerations inherent in these statistical tests. These aspects apply to all types of ANOVA designs, from simple random complete blocks to completely randomized design factorials, and more complicated repeated measures and split-plot designs. I'm sure most readers have encountered this information from one or more statistics classes, but in the real world of research we sometimes become less rigorous than when first learning about these issues. I would like to stress their importance.

The first subject is the testing for homogeneity of variance between treatments **before** an ANOVA is conducted. Why is this important when an ANOVA tolerates mild departures from homogeneity of variance and normality? Well, the name of the statistical test itself should give the biggest clue. An analysis of variance is just what it says. It uses ratios of variances to construct *F*-tests that let an experimenter test for treatment mean differences. If there are mean treatment differences, one would expect that variation within a treatment would be much less than variation between treatments. Things become confusing if an experimenter does not test, and does indeed have heterogeneity of treatment variances. How can either a significant or nonsignificant *F*-test result be interpreted in this case? The error term in an ANOVA is a pooled estimate of the within-treatment variances that can either be biased upwards or downwards when the treatments have unequal variances. Is a significant *F*-test a result of true treatment differences or a downward biasing of the pooled error variance estimate due to heterogeneity of treatment variances? Similarly, is a nonsignificant *F*-test result indicative of no treatment differences or due to an upward bias of the pooled error variance component? In these times of sophisticated statistical software packages that print out P-values to the fifth decimal place, it takes very little effort to conduct a test to ensure homogeneity of variance before running an ANOVA. None of us like to work with transformed data for a variety of reasons: it's unrecognizable and somehow, in my bones, it still feels like cheating. But, I would much rather get believable, interpretable results from an ANOVA using transformed data than try and explain dubious, borderline effects that may be due to heterogeneity of variance problems.

The second issue concerns testing of main effects or lower order interactions in an ANOVA when a higher order interaction is significant. One should not do it! It does not make sense to interpret main effect mean differences for one factor when there is a dependence on the level of a second factor involved. Interactions provide information that cannot be provided by the sum of the respective main effects. According to B. J. Winer in his book *Statistical Principles in Experimental Design,* "From many points of view, interaction is a measure of the nonadditivity of the main effects."

One so-called solution I have seen used is to break a two-way ANOVA down into two one-way ANOVAs (one for each factor) when there is a significant interaction effect. The problem with this is that an experimenter does not escape the fact that there is still an interaction present between the two factors. One is still looking at the effects of the first factor averaged over the effects of the second factor, which is just as misleading as ignoring a significant two-way interaction in the first place. A more sophisticated approach is to break the analysis up into a series of one-ways, each at a specific level of the second factor. But, if a researcher goes to the trouble of finding an interaction, why would that valuable information then be discarded in this way? Another layer of complexity is added when models other than fixed are used. In the case of a mixed model for a simple two-way ANOVA, the *F*-statistic is derived by dividing the mean square of the fixed main effect by the mean square of the interaction effect. If there is a significant interaction and the analysis is subsequently broken down into a series of one-ways at each level of the second factor (as discussed above), the error term for testing main effect significance will be incorrect.

On the one hand, I know that it becomes increasingly difficult to interpret interactions when the number of factors being simultaneously tested is increased. On the other hand, one of the reasons for using an ANOVA design to analyze complex data is to find these interactions and try to make sense of them. Research is not simple—there is more to life than main effects!