

## StatNews #61

# Specifying Fixed and Random Factors in Mixed or Multi-Level Models

## December 2003

### Updated 2012

Since SAS introduced Proc Mixed about ten years ago, S-Plus, Stata and SPSS have implemented procedures to analyze mixed models, greatly broadening the options available to researchers. These programs require correctly specifying the fixed and random factors of the model to obtain accurate analyses. The definitions in many texts often do not help with decisions to specify factors as fixed or random, since textbook examples are often artificial and hard to apply. Furthermore, the same factor can often be considered fixed or random, depending on the objective. This newsletter outlines a different way to think about fixed and random factors.

Consider an experiment that examines beetle damage on cucumbers. The experiment is replicated at five farms and on four fields at each farm. There are two varieties of cucumbers, and beetle damage is assessed on each of 50 plants at the end of the season. The researcher is interested in comparing differences in how much damage the two varieties sustain. The experiment then has the following factors: VARIETY, FARM, and FIELD.

Fixed factors can be thought of in terms of differences. The effect of a categorical fixed factor is defined by differences from the overall mean and the effect of a continuous fixed factor is defined by its slope—how the mean of the dependent variable differs with alternate values of the factor. The output for fixed factors provides estimates for mean-differences or slopes. Conclusions regarding fixed factors are particular to the values of these factors. For example, if one variety of cucumber is found to suffer significantly less damage than the other, this says nothing about cucumber varieties that were not tested.

Random factors, on the other hand, are defined by a distribution and not by differences. The values of a random factor are assumed to be chosen from a population with a normal distribution with a certain variance. The output for a random factor is an estimate of this variance and not a set of differences from a mean. Conclusions regarding random factors should be expressed in terms of variance. For example, we may find that the variability among fields makes up a certain percentage of the overall variability in beetle damage.

Situations that indicate fixed factors:

1. **The factor is the primary treatment that the researcher wants to compare.** In our example, VARIETY is definitely fixed as the researcher wants to compare the mean beetle damage on the two varieties.
2. **The factor is a secondary covariate that might be confounded with the treatment, and the researcher wants to control for differences in this covariate.** If these farms were specifically chosen for some feature they had, such as specific soil types or topographies that may affect beetle damage, and if the researcher would like to compare the farms as representatives of those soil types, then FARM should be fixed.
3. **The factor has only two values.** Even if everything else indicates that a factor should be random, if it has only two values, the variance cannot be calculated, and it should be fixed.

Situations that indicate random factors:

1. **The researcher is interested in quantifying how much of the overall variation to attribute to this factor.** If the researcher was interested in how much of the variation in beetle damage was attributable to the farm at which the damage took place, FARM would be random.
2. **The researcher is not interested in knowing which means differ, but wants to account for the variation in this factor.** If the farms were chosen at random, not for a specific feature, but because the researcher suspected that there is some variation in their soil types, which is representative of the variation across all farms, FARM should be random.
3. **The researcher would like to generalize the conclusions about this factor to the whole population.** There is nothing about comparing these specific fields that is of interest to the researcher. Rather, the researcher wants to generalize the results of this experiment to all fields, so FIELD is random.
4. **Any interaction with a random factor is also random.**

How the factors of a model are specified can have great influence on the results of the analysis and on the conclusions drawn. If you need assistance in specifying which factors are fixed and random, please contact any of the consultants in the Cornell Office of Statistical Consulting.

Author: Karen Grace-Martin

(This newsletter was distributed by the Cornell Statistical Consulting Unit. Please forward it to any interested colleagues, students, and research staff. Anyone not receiving this newsletter who would like to be added to the mailing list for future newsletters should contact us at [cscu@cornell.edu](mailto:cscu@cornell.edu). Information about the Cornell Statistical Consulting Unit and copies of previous newsletters can be obtained at <http://www.cscu.cornell.edu>).