Specifying Fixed and Random Factors in Mixed or Multilevel Models

Cornell Statistical Consulting Unit

1 Introduction

In many experimental designs, measurements can be grouped according to experimental unit on which those measurements were taken, the experimental conditions applied to the experimental unit, as well as other sources of variation such as plots within a field or batches within a given experimental run. Most statistical analyses will incorporate these sources of variation as explanatory variables or "factors" in a statistical model. This can improve the precision of parameter estimates and is often necessary to account for correlations among measurements taken on the same experimental unit.

There are two common approaches to incorporating potential sources of variation as factors in a statistical model. In one approach, an explanatory factor is considered a "fixed effect" and the statistical model estimates the average difference in the dependent variable between each group defined by that factor *for the population of all measurements that could be taken within those same groups*. Alternatively, the factor can be treated as a "random effect", and the statistical model estimates the *variance* of the dependent variable, in the population of *groups* defined by that factor. This newsletter provides guidelines for incorporating explanatory factors as fixed or random effects in a statistical model.

2 Example

The concepts of fixed and random effects can be better understood through an example.

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. We might expect variation in the amount of damage due to the variety, the farm, the field. We can identify VARIETY, FARM, and FIELD as sources of variation, or factors, that we can incorporate into a statistical model for cucumber damage.

3 Fixed and random factors

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 (the change in the dependent variable associated with a one-unit increase in the factor). Including a categorical variable as a fixed factor yields estimates of the average difference in the dependent variable between each category, for the population of measurements that conceivably exist within those fixed categories. For example, if the variety of cucumber is treated as a fixed effect, then we are estimating the average difference in beetle damage for the population of cucumbers within each fixed variety.

In contrast, including a categorical factor as a random effect leads to an estimate of the variance of the dependent variable across the population of categories or groups defined by that factor. In the cucumber example, including FIELD as a random effect would lead to an estimate of the between-field variance in cucumber damage, and we could quantify the contribution of field-to-field variation to the total variance in cucumber damage. In this case, we are less interested in comparing the average damage between specific fields, but we still want to control for this source of variation and to understand how much of the variation in cucumber damage can be attributed to variation among field characteristics.

4 Guidelines

Here are some guidelines for choosing whether a factor should be included as a fixed effect or a random effect in a statistical analysis.

Situations that indicate fixed factors:

1. The factor is the primary treatment that the researcher wants to compare. In our example, VARIETY is a fixed effect since 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 (categorical) factor has very few unique categories or groups; in this case, it may not be feasible to estimate group-to-group variance.

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.

Created December 2003. Last updated April 2022.