Distinguishing Between Random and Fixed: Variables, Effects, and Coefficients[1]

The terms “random” and “fixed” are used frequently in the multilevel modeling literature. The distinction is a difficult one to begin with and becomes more confusing because the terms are used to refer to different circumstances. Here are some summary comments that may help.

Random and Fixed Variables

A “fixed variable” is one that is assumed to be measured without error. It is also assumed that the values of a fixed variable in one study are the same as those same values of the fixed variable in another study. “Random variables” are assumed to be values that are drawn from a larger population of values and thus will represent them rather than be the exact same value invariantly. You can think of the values of random variables as representing a random sample of all possible values or instances of that variable. Thus, we expect to generalize the results obtained with a random variable to all other possible instances of that value (e.g., a job candidate with a strong résumé). Most of the time in ANOVA and regression analysis we assume the independent variables are fixed.

Random and Fixed Effects

The terms “random” and “fixed” are used in the context of ANOVA and regression models and refer to a certain type of statistical model. Almost always, researchers in psychology and social sciences use fixed effects regression or ANOVA and they are rarely faced with a situation involving random effects analyses. A fixed-effects ANOVA refers to assumptions about the independent variable and that error distribution for the variable. An experimental design is the easiest example for illustrating the principle. Usually, the researcher is interested in only generalizing the results to experimental values used in the study. For instance, a drug study might use 0 mg, 5 mg, and 10 mg of an experimental drug. This is a circumstance when a fixed effects ANOVA would be appropriate. In this example, the extrapolation is to other studies or treatments that might use the same values of the drug (i.e., 0 mg, 5 mg, and 10 mg). However, if the researcher wants to make inferences beyond the particular values of the independent variable used in the study, a random effects model is appropriate. A common example would be the use of some sort of stimulus, like a picture of a person or thing representing certain qualities. For example, public art works representing low, moderate, and high abstractness (e.g., statue of a war hero vs. a pivoting geometric design). In this case, the researcher would like to make inferences beyond just one art piece representing each category of abstractness, so the art pieces are conceptualized as particular pieces randomly drawn from a larger universe of possible pieces that are sampled from the domain for that level of abstractness. One could imagine using several instances of high abstract pieces that are randomly drawn from a larger population of high abstract pieces and are thought to be only a few of the possible particular instances of high abstract art. Thus, the statistical inferences are made to a larger universe of art works with variations of abstractness within each category. Such a generalization is more of an inferential leap, and, consequently, the random effects model is less powerful because we are taking into account some additional expected random variation on the independent variable. Random effects models are sometimes referred to as “Model II” or “variance component models.” Analyses using both fixed and random effects are called “mixed models” or "mixed effects models" which is one of the terms given to multilevel models.

Fixed and Random Coefficients in Multilevel Regression (MLR)

The random vs. fixed terminology is commonly used in multilevel modeling.[2] The ***intercepts and slopes*** in the level-1 equations may be allowed to vary randomly across groups and are therefore referred to as “random coefficients.” The variance of the intercepts, represented by $\tau^2_0 = var(U_{0j})$, is

---

[1] My distinctions between random variables, effects, and coefficients were inspired by and rely heavily on the one given by Kreft and de Leeuw (1998).
[2] In MLR, it is still the case that both level-1 and level-2 ***predictors*** are assumed to be fixed just like in the traditional ANOVA or (single-level) regression model.

nearly always estimated by default. The variance of the slope for the (first) predictor, represented by $\tau^2_1 = var(U_{1j})$, may or may not be estimated depending on the circumstances (don't worry, we will discuss the concept of varying slopes much more later).[3] The within-group variance, $var(R_{ij}) = \sigma^2$, also can be referred to as a "random" effect, although it is less likely to be mentioned because it is not usually the focus of hypothesis and significance testing.

In MLR, the term "fixed effect," on the other hand, refers to the average intercept (γ_{00} from the last handout) or average slope (e.g., γ_{01}), which has not really been discussed yet. These are usually the initial primary interest for researchers and are typically discussed first. Output from software packages will usually have sections labeled as "fixed effects" and "random effects." The values under the fixed effects section are essentially just the regression coefficients (intercept, slope) as we usually think about them in regression analysis. The random effects are the variances of the intercepts or slopes across groups and are often of key interest to researchers as well.

In the HLM program, variances for the intercepts and slopes are estimated by default (U_{0j} and U_{1j}, respectively). In SPSS Mixed and R (nlme or lme4), the user must specify whether and which intercept or slope variances should be estimated. If any variance, intercept or slope, is not specified their values are set to zero. By setting variances to zero, we are testing a model in which we assume β_{0j} and β_{1j} do not vary randomly across groups. Thus, the intercept or slope value is assumed to be constant or "nonvarying" across groups. For example, fixed, nonvarying intercepts would imply the group average for the dependent variable is assumed to be equal in each group—this implies no variance, so would be rarely assumed for intercepts. Note that although researchers sometimes refer to this constraint as "fixing the intercepts" or "fixing the slopes," the term is somewhat loosely applied, because we are really assuming they are fixed *and nonvarying*. We will discuss these concepts pertaining to varying slopes more in coming weeks.

Note on Groups in MLR

The variance of the coefficients across groups (i.e., the random effects) in multilevel model implies something different about the groups than is what is expected in fixed effects ANOVA. The statistical theory behind multilevel models means that we think about β_{0j} and β_{1j} as akin to the *random variables* that I described above in the first section. This implies that, as with the notion of attempting to generalize beyond the particular values of a certain variable, we are attempting to generalize beyond the particular groups in the study. For instance, we may have 100 tech companies in our study, but we wish to generalize to a larger universe of tech companies. The groups or level-2 units used in multilevel models are thus of a different nature than the groups we use for a traditional fixed-effects ANOVA (or single-level regression). Using a small number of fixed groups, such as racial categories, gender, intervention groups, or organization type are not the right kind group to be use as a level-2 unit in MLR. We have to have a grouping variable that can be conceived of as a sampling unit where the subset used in our sample can be conceptualized as representative of a larger set of possible groups.

References

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage.
Raudenbush, S.W., & Bryk, A.S., (2002) *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
Snijders, T.A.B., & Bosker, R.J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling (2nd Edition)*. London: Sage.

[3] My notation follows that of Snijders & Bosker (2012), but notation differs among texts. In Raudenbush & Bryk (2002), for instance, the symbol for intercept variance is τ_{00}, and the subscript refers to the diagonal element corresponding to the row and column of the variance-covariance matrix for U_{0j}, and the subscript for the slope variance, τ_{11} (with no superscripted 2), refers to the diagonal element for the U_{1j} column and row.

Summary Table

Random vs. Fixed	Definition	Example	Use in Multilevel Regression
Variables	***Random variable***: (1) is assumed to be measured with measurement error. The scores are a function of a true score and random error; (2) the values come from and are intended to generalize to a much larger population of possible values with a certain probability distribution (e.g., normal distribution); (3) the number of values in the study is small relative to the values of the variable as it appears in the population it is drawn from. ***Fixed variable***: (1) assumed to be measured without measurement error; (2) desired generalization to population or other studies is to the same values; (3) the variable used in the study contains all or most of the variable's values in the population. It is important to distinguish between a variable that is *varying* and a variable that is *random*. A fixed variable can have different values, it is not necessarily invariant (equal) across groups.	***Random variable:*** photographs representing individuals with differing levels of attractiveness manipulated in an experiment, a subset of census tracks ***Fixed variable:*** gender, race, or intervention vs. control group.	Predictor variables in MLR generally assumed to be fixed
Effects	***Random effect***: (1) different statistical model of regression or ANOVA model which assumes that an independent variable is random; (2) generally used if the levels of the independent variable are thought to be a small subset of the possible values which one wishes to generalize to; (3) will probably produce larger standard errors (less powerful). ***Fixed effect***: (1) statistical model typically used in regression and ANOVA assuming independent variable is fixed; (2) generalization of the results apply to similar values of independent variable in the population or in other studies; (3) will probably produce smaller standard errors (more powerful).	**Random effect**: random effects ANOVA, random effects regression ***Fixed effect***: fixed effects ANOVA, fixed effects regression	Intercept only models in MLR are equivalent to random effects ANOVA and inclusion of one or more level-1 predictors makes the model equivalent to a random effects ANCOVA when slopes do not vary across groups.
Coefficients	***Random coefficient:*** term applies only to MLR analyses in which intercepts, slopes, and variances can be assumed to be random. MLR analyses most typically assume random coefficients. One can conceptualize the coefficients obtained from the level-1 regressions as a type of random variable which comes from and generalizes to a distribution of possible values. Groups are conceived of as a subset of the possible groups. ***Fixed coefficient:*** a coefficient can be fixed to be non-varying (invariant) across groups by setting the between-group variance to zero. Random coefficients must be variable across groups. Conceptually, fixed coefficients may be nonvarying *or* varying across groups.	***Random coefficient***: the level-2 predictor, average income, is used to predict school performance in each school. Intercept values for school performance are assumed to be a sample of the intercepts from a larger population of schools. ***Fixed coefficient***: slopes or intercepts constrained to be equal over different schools.	Both used in MLR. Slopes and intercept values can be considered to be fixed or random, depending on researchers' assumptions and how the model is specified. The average intercept or slope is referred to as a "fixed effect." Variances of the slopes and intercepts (if allowed to vary across groups) are called “random coefficients."