

Distinguishing Between Random and Fixed: Variables, Effects, and Coefficients

The terms “random” and “fixed” are used frequently in the hierarchical linear 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 the 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. You can think of the values of random variables as representing a random sample of all possible values of that variable. Thus, we expect to generalize the results obtained with a random variable to all other possible values of that random variable. 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 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 the 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 using 0 mg, 5 mg, or 10 mg of an experimental drug. This is when a fixed effects ANOVA would be appropriate. In this case, 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 used. A common example would be the use of public art works representing low, moderate, and high abstractness (e.g., statue of a war hero vs. a pivoting geometric design). The researcher would like to make inferences beyond the three pieces used, so the art pieces are conceptualized as pieces randomly drawn from a larger universe of possible pieces and the inferences are made to a larger universe of art work and range of abstractness values. Such a generalization is more of an inferential leap, and, consequently, the random effects model is less powerful. 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.”

Fixed and Random Coefficients in Multilevel Regression

The random vs. fixed distinction for variables and effects is important in multilevel regression. In multilevel regression models, both level-1 and level-2 *predictors* are assumed to be fixed. However, level-1 *intercepts and slopes* are typically assumed to vary randomly across groups. Because of the assumptions about their error distributions, we call their variances, $\tau_{00}=\text{var}(U_{0j})$ and $\tau_{11}=\text{var}(U_{1j})$, “random coefficients.”

This means that we can think about β_{0j} and β_{1j} as akin to the random *variables* I described above. Instead of attempting to generalize beyond the particular values of the independent variable, we are attempting to generalize beyond the particular groups in the study. For instance, we may have 50 companies, but we wish to generalize to a larger universe of companies when we examine the means (intercepts) or the x-y relationship (slopes).

In the HLM program, variances for the intercepts and slopes are estimated by default (U_{0j} and U_{1j} , respectively). In SPSS Mixed, the user must specify which intercepts or slopes 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. Although we typically 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*.

Intercept Model and Random Effects ANOVA

In the simplest HLM model, there are no predictors. We have one level-1 equation and one level-2 equation:

$$Y_{ij} = \beta_{0j} + R_{ij}$$

where Y is the dependent variable, β_{0j} is the intercept, and R_{ij} is the residual or error. β_{0j} can be thought of as the mean of each group. The level-2 equation also has no predictors in its simplest form:

$$\beta_{0j} = \gamma_{00} + U_{0j}$$

where β_{0j} is the dependent variable, γ_{00} is the level-2 intercept, and U_{0j} is the level-2 error. In this equation, γ_{00} represents the grand mean or the mean of the intercepts. U_{0j} represents the deviation of each mean from the grand mean. When the average deviation is large, there are large group differences.

It can be shown that this is really the ANOVA model. ANOVA examines the deviations of group means from the grand mean. Because we assume that the group means, represented by β_{0j} and, thus, their deviations are randomly varying, this model is equivalent to the *random effects* ANOVA model. We can rewrite our two HLM equations as a single equation, if we plug in the level-2 terms into level-1 equation.

$$Y_{ij} = \gamma_{00} + U_{0j} + R_{ij}$$

In this equation, γ_{00} represents the grand mean, the U_{0j} term represents the deviations from the mean (or mean differences or treatment effect, if you would like), and R_{ij} represents the within-group error or variation. In more mathematically oriented ANOVA textbooks, you will see an ANOVA model written something like:

$$Y_{ij} = \mu + \alpha_j + e_{ij}$$

If the model is a random effects ANOVA, α_j is assumed to have a random distribution. In multilevel regression, because we assume U_{0j} to vary randomly, the simple HLM model with no level-1 or level-2 predictors is equivalent to the random effects ANOVA model.

When we add predictors to the level-1 equation, they are covariates and the model becomes a random effects ANCOVA in which the means are adjusted for the covariate.

Summary Table

Random vs. Fixed	Definition	Example	Use in Multilevel Regression
Variables	<p>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.</p> <p>It is important to distinguish between a variable that is <i>varying</i> and a variable that is <i>random</i>. A fixed variable can have different values, it is not necessarily invariant (equal) across groups.</p>	<p>Random variable: photographs representing individuals with differing levels of attractiveness manipulated in an experiment, census tracks</p> <p>Fixed variable: gender, race, or intervention vs. control group.</p>	<p>Predictor variables in MLR generally assumed to be fixed</p>
Effects	<p>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).</p>	<p>Random effect: random effects ANOVA, random effects regression</p> <p>Fixed effect: fixed effects ANOVA, fixed effects regression</p>	<p>Intercept only models in MLR are equivalent to random effects ANOVA or ANCOVA.</p>
Coefficients	<p>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.</p> <p>Fixed coefficient: a coefficient can be fixed to be non-varying (invariant) across groups by setting its between group variance to zero.</p> <p>Random coefficients must be variable across groups. Conceptually, fixed coefficients may be invariant or varying across groups.</p>	<p>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.</p> <p>Fixed coefficient: slopes or intercepts constrained to be equal over different schools.</p>	<p>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 variance of the slopes or intercepts are considered random coefficients.</p>