Distinguishing random and fixed factors. 28 May 2018

Eisenhart, C. 1947. The Assumptions Underlying the Analysis of Variance Biometrics, 3: 1-21.

Eisenhart distinguishes 2 roles for ANOVA
Class I: Detection and Estimation of Fixed (Constant) Relations Among the Means of Sub-Sets of the Universe of Objects Concerned.
Assumptions:
1 random variates,
2 additive effects,
3 equal variances and zero correlations,
4 normality
Class II: Detection and Estimation of Components of (Random) Variation Associated with a Composite Population.
Assumption A (Random Variables)
Assumption B (Additivity of Components)
Assumption C (Zero Correlations and Homogeneous Variances)
Assumption D: The deviations are normally distributed

"In summary, when the formulas and procedures of analysis of variance are used merely to summarize properties of the data in hand, no assumptions are needed to validate them. On the other hand, when analysis of- variance is used as a method of statistical inference, for inferring properties of the "population" from which the data in hand were drawn, then certain assumptions, about the "population" and the sampling procedure by means of which the data were obtained, must be fulfilled if the inferences are to be valid.

"Which Model--Model I or Model II?
In practical work a question that often arises is: which model is appropriate in the present instance-Model I or Model II?
Basically, of course, the answer is clear as soon as a decision is reached on whether the parameters of interest specify fixed relations, or components of random variation."

"The following parallel sets of questions serve to focus attention on the pertinent issues, and have been found helpful in answering the basic question of random versus fixed effects:
(1) Are the conclusions to be confined to the things actually studied (the animals, or the plots); to the immediate sources of these things (the herds, or the fields); or expanded to apply to more general populations (the species, or the farmland of the state) ?
(2) In complete repetitions of the experiment would the same things be studied again (the same animals, or the same plots); would new samples be drawn from the identical sources (new samples of animals from the same herds, or new experimental arrangements on the same fields); or would new samples be drawn from the more general populations (new samples of animals from new herds, or new experimental arrangements on new fields)?

Rasch, D., J. Pilz, R. Verdooren, A. Gebhardt. 2011 Optimal Experimental Design with R. Boca Raton, CRC Press.

Following Eisenhart (1947).
Fixed factors – There are exactly *a* levels, all included in the experiment. Model I or fixed effects model is all factors are fixed.
Random factors. Many levels, whose number in theory is infinite. Levels have to be selected randomly from the universe of levels. Model II analysis or random factor analysis if all factors are random.

Lawson, J. 2015. Design and Analysis of Experiments with R, Boca Raton, CRC Press

Experiment is an action where the experimenter changes at least one of the variables being studied and then observed the effect of his or her action(s). p 3.
When the purpose of experimentation is to study differences in the average response caused by differences in factor levels, the factors in the experiment are call *fixed factors.* P 142.
When the purpose of experimentation is to study the variance caused by changing levels of a factor, the factor is called a *random factor.* P142
Where the levels of fixed factors are specifically chosen by the experimenter, the levels of random factors are just samples of possible levels that could have been used. P142

Stroup, W.W. 2013. Generalized Linear Mixed Models. Boca Raton, CRC press.
P 38 Do each effect's levels represent a larger population of interest or are the levels observed in the study of explicit interest and hence the entire population?
Is there a probability distribution associated with the effect in question? (Typically normal and mutually independent for random factors)
P43 Fixed or Random: Tough Calls
The only all-encompassing right answer is " it depends." on the specifics of each study.
Broad inference defined pp 90-94. Narrow inference defined pp 95-99
P271 Broad inference uses estimable functions only….. This means that we do not restrict attention to a particular set of levels of the random effect, **b**, but intend inference to apply broadly across the entire population represented by the levels of **b.**
Narrow inference can serve one of two purposes.
1. Inference focuses on a particular level or set of levels of **b**. …. Examples include …. location specific effects.
2. Focus remains primarily on the fixed effects, but nonzero **M** is used to limit inference about **K'β** to particular subsets of the population represented bv **b**. [**K'β + M'b** is an estimable function)

Quinn G.P. and M.J. Keough 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press

p 276

There are two types of categorical predictor variables in linear models. The most common type is a fixed factor, where all the levels of the factor (i.e. all the groups or treatments) that are of interest are included in the analysis. We cannot extrapolate our statistical conclusions beyond these specific levels to other groups or treatments not in the study. If we repeated the study, we would usually use the same levels of the fixed factor again. Linear models based on fixed categorical predictor variables (fixed factors) are termed fixed effects models (or Model 1 ANOVAs). Fixed effect models are analogous to linear regression models where X is assumed to be fixed. The other type of factor is a random factor, where we are only using a random selection of all the possible levels (or groups) of the factor and we usually wish to make inferences about all the possible groups from our sample of groups. If we repeated the study, we would usually take another sample of groups from the population of possible groups. Linear models based on random categorical predictor variables (random factors) are termed random effects models (or Model 2 ANOVAs).