Deconstructing Type III

L. R. LaMotte[1]

[1]*LSU Health - New Orleans U.S.A.*

Abstract

R. A. Fisher expounded analysis of variance (ANOVA) for settings in which responses are observed under experimental conditions described by combinations of levels of one or more factors. For two factors, ANOVA partitions differences among cell means into three groups, "variation between classes of type A and between classes of type B" and "interaction of causes" (Fisher 1938, p. 240), commonly named A and B main effects and AB interaction effects. This scheme extends readily, resolving differences for f factors into $2^f - 1$ effects. For three factors, for example, the effects are the three main effects, three two-factor interaction effects, and the three-factor interaction effect.

ANOVA quickly became the main statistical methodology in diverse disciplines. It provides concepts and terminology that have become a common language of applied statistics.

In balanced models, with equal numbers of observed responses over all factor-level combinations (FLCs or cells), sums of squares (SSs) for statistics to test the effects have simple formulations, and computing them is straightforward. Their distributional properties (assuming normally-distributed responses) are apparent.

The situation is unsettled for unbalanced models and settings in which there are no observations in some cells. ANOVA SSs do not provide appropriate test statistics. While there is no theoretical impediment to mimicking the balanced-model partition, the results are less (often un-) informative because some or all subspaces of effects are not estimable. No statistical computing package, as far as I know, takes this approach.

Most packages use Type III SSs to produce an ANOVA-like partition of effects. Type III was introduced by SAS in the 1970s: see Goodnight (1976) and SAS (1978). It is defined by an algorithm; the process begins and ends with a set of rules for formulating the Type III estimable functions of an effect. No hypothesis is formulated. The SS for an effect is the squared norm of the orthogonal projection of the vector of observed responses onto the space of Type III estimable functions.

The lack of a compact mathematical definition of Type III estimable functions and SSs gives them a mystical, black-box aura. A system of beliefs has evolved. For example, it is widely asserted that, in balanced models, Type III SSs are the same as ANOVA SSs and that they test ANOVA-effect hypotheses in unbalanced models if there are no empty cells. Such assertions typically are illustrated in terms of two-factor settings with models that do not involve covariates. General formulations of models that include factor effects are hard to find (see Hocking 2013, for example), and no formulation of Type III in general settings has appeared. Proofs of these assertions do not exist. Apparently what is thought to be known about Type III is based only on experience and observation, and even then only in uncomplicated settings.

My objective in this lecture is to provide an explicit formulation of Type III estimable functions and SSs and to establish their properties. It is shown that what is believed to be true, is true, when all of an effect is estimable, but not otherwise.

Keywords ANOVA, Factor Effects

References

[1] Fisher, R. A. (1938). *Statistical Methods for Research Workers, 7th Edition.* Oliver and Boyd, London.

[2] Goodnight, J. H. (1976). The general linear models procedure. Proceedings of the First International SAS User's Group. SAS Institute Inc., Cary, NC.

[3] Hocking, R. R. (2013). Methods and Applications of Linear Models, Third Edition. John Wiley & Sons, Inc., Hoboken, New Jersey.

[4] SAS Institute Inc. (1978). SAS Technical Report R-101, Tests of hypotheses in fixed-effects linear models. SAS Institute Inc., Cary, NC.