Anova – Type I/II/III SS explained

ANOVA (and R)
The ANOVA Controversy

ANOVA is a statistical process for analysing the amount of variance that is contributed to a sample by different factors. It was initially derived by R. A. Fisher in 1925, for the case of balanced data (equal numbers of observations for each level of a factor).

When data is unbalanced, there are different ways to calculate the sums of squares for ANOVA. There are at least 3 approaches, commonly called Type I, II and III sums of squares (this notation seems to have been introduced into the statistics world from the SAS package but is now widespread). Which type to use has led to an ongoing controversy in the field of statistics (for an overview, see Heer [2]). However, it essentially comes down to testing different hypotheses about the data.
Type I, II and III Sums of Squares

Consider a model that includes two factors A and B; there are therefore two main effects, and an interaction, AB. The full model is represented by SS(A, B, AB).

Other models are represented similarly: SS(A, B) indicates the model with no interaction, SS(B, AB) indicates the model that does not account for effects from factor A, and so on.

The influence of particular factors (including interactions) can be tested by examining the differences between models. For example, to determine the presence of an interaction effect, an F-test of the models SS(A, B, AB) and the no-interaction model SS(A, B) would be carried out.

It is convenient to define incremental sums of squares to represent these differences. Let

SS(AB | A, B) = SS(A, B, AB) − SS(A, B)
SS(A | B, AB) = SS(A, B, AB) − SS(B, AB)
SS(B | A, AB) = SS(A, B, AB) − SS(A, AB)
SS(A | B) = SS(A, B) − SS(B)
SS(B | A) = SS(A, B) − SS(A)

The notation shows the incremental differences in sums of squares, for example SS(AB | A, B) represents "the sum of squares for interaction after the main effects", and SS(A | B) is "the sum of squares for the A main effect after the B main effect and ignoring interactions" [1].

The different types of sums of squares then arise depending on the stage of model reduction at which they are carried out. In particular:

Type I, also called "sequential" sum of squares:

> SS(A) for factor A.

> SS(B | A) for factor B.

> SS(AB | B, A) for interaction AB.

> This tests the main effect of factor A, followed by the main effect of factor B after the main effect of A, followed by the interaction effect AB after the main effects.

> Because of the sequential nature and the fact that the two main factors are tested in a particular order, this type of sums of squares will give different results for unbalanced data depending on which main effect is considered first.

> For unbalanced data, this approach tests for a difference in the weighted marginal means. In practical terms, this means that the results are dependent on the realized sample sizes, namely the proportions in the particular data set. In other words, it is testing the first factor without controlling for the other factor (for further discussion and a worked example, see Zahn [4]).

> Note that this is often not the hypothesis that is of interest when dealing with unbalanced data.

Type II:

> SS(A | B) for factor A.

> SS(B | A) for factor B.

> This type tests for each main effect after the other main effect.

Note that no significant interaction is assumed (in other words, you should test for interaction first (SS(AB | A, B)) and only if AB is not significant, continue with the analysis for main effects).

If there is indeed no interaction, then type II is statistically more powerful than type III (see Langsrud [3] for further details).

Computationally, this is equivalent to running a type I analysis with different orders of the factors, and taking the appropriate output (the second, where one main effect is run after the other, in the example above).

Type III:

SS(A | B, AB) for factor A.

SS(B | A, AB) for factor B.

This type tests for the presence of a main effect after the other main effect and interaction. This approach is therefore valid in the presence of significant interactions.

However, it is often not interesting to interpret a main effect if interactions are present (generally speaking, if a significant interaction is present, the main effects should not be further analysed).

If the interactions are not significant, type II gives a more powerful test.

NOTE: when data is balanced, the factors are orthogonal, and types I, II and III all give the same results.
Summary: Usually the hypothesis of interest is about the significance of one factor while controlling for the level of the other factors. This equates to using type II or III SS. In general, if there is no significant interaction effect, then type II is more powerful, and follows the principle of marginality. If interaction is present, then type II is inappropriate while type III can still be used, but results need to be interpreted with caution (in the presence of interactions, main effects are rarely interpretable).

The anova and aov Functions in R

The anova and aov functions in R implement a sequential sum of squares (type I). As indicated above, for unbalanced data, this rarely tests a hypothesis of interest, since essentially the effect of one factor is calculated based on the varying levels of the other factor. In a practical sense, this means that the results are interpretable only in relation to the particular levels of observations that occur in the (unbalanced) data set. Fortunately, based on the above discussion, it should be clear that it is relatively straightforward to obtain type II SS in R.
Type II SS in R

Since type II SS tests each main effect after the other main effects, and assumes no interactions, the correct SS can be obtained using anova() and varying the order of the factors.

For example, consider a data frame (search) for which the response variable is the time that it takes users to find a relevant answer with an information retreival system (time). The user is assigned to one of two experimental search systems on which they run the test (sys). They are also assigned a number of different search queries (topic).

To obtain type I SS:

anova(lm(time ~ sys * topic, data=search))

If the data is unbalanced, you will obtain slightly different results if you instead use:

anova(lm(time ~ topic * sys, data=search))

The type II SS is obtained by using the second line of output from each of the above commands (since in type I SS, the second component will be the second factor, after the first factor). That is, you obtain the type II SS results for topic from the first command, and the results for sys from the second.

Type III SS in R

This is slightly more involved than the type II results.
First, it is necessary to set the contrasts option in R. Because the multi-way ANOVA model is over-parameterised, it is necessary to choose a contrasts setting that sums to zero, otherwise the ANOVA analysis will give incorrect results with respect to the expected hypothesis. (The default contrasts type does not satisfy this requirement.)

options(contrasts = c("contr.sum","contr.poly"))

Next, store the model:

model <- lm(time ~ topic * sys, data=search)

Finally, call the drop1 function on each model component:

drop1(model, .~., test="F")

The results give the type III SS, including the p-values from an F-test.

Type II and III SS Using the car Package

A somewhat easier way to obtain type II and III SS is through the car package. This defines a new function, Anova(), which can calculate type II and III SS directly.

Type II, using the same data set defined above:

Anova(lm(time ~ topic * sys, data=search, type=2))

Type III:

Anova(lm(time ~ topic * sys, data=search, contrasts=list(topic=contr.sum, sys=contr.sum)), type=3))

NOTE: *Again, due to the way in which the SS are calculated when incorporating the interaction effect, for type III you must specify the contrasts option to obtain sensible results.*

The Sum of Squares for the Main Effects in a Type II ANOVA do not take the respective interaction terms into account while a Type III does. Thus, the estimates of the main effects in a Type III ANOVA are mathematically/statistically valid even if the interaction term is significant, while they are not for a Type II ANOVA. Some statisticians have argued that this is a moot point because you cannot interpret your main effects if an interaction term exists anyway- that's precisely what a significant interaction term is telling you, the outcome depends on the level of both factors in a non-additive way. Thus, some people argue that Type II ANOVAs are superior because they have more power, and anything with a significant interaction term should be re-examined in a different way.