GA

Model testing approaches (types of Sums of Squares)

As I've mentioned during the lectures some statistics packages print several types of sums of squares for testing hypotheses in ANOVA, and these are normally referred to by their names given in SAS GLM. Basically, the strategies differ a bit in how the sums of squares are calculated. It is very important to note, already here, that **the three methods produce identical results if data is balanced** (equal number of observations per cell). If data is unbalanced, the three methods differ from one another and choosing type of SS becomes an issue. So, preferably have balanced data.... Below, I exemplify the three types of SS decomposition as simply as possible, using the two-way design to illustrate things.

Type I. Type I sums of squares are computed from the difference between the residual sums of squares of two different models. The particular models needed for the computation depend on the order of the variables are entered. For example, if the model is:

y = constant + a + b + a*b

then the sum of squares for A*B is produced from the difference between SSE (sum of squared error) in the two following models:

y = constant + a + b
y = constant + a + b + a*b

Similarly, the Type I sums of squares for B in this model are computed from the difference in SSE between the following models:

y = constant + a
y = constant + a + b

Finally, the Type I sums of squares for A is computed from the difference in SSE for the following:

y = constant
y = constant + a

In summary, to compute sums of squares, move from right to left and construct models which differ by the right-most term only. Note that for main effects, the order in which terms are enetered into the model matters (if data is unbalanced, that is). The SS attributable to a specific effect will depend on the order in which the effects are entered into the model if data is unbalanced!

Type II. Type II sums of squares are computed similarly to Type I except that main effects and interactions determine the ordering of differences instead of the statement order. For the above model, Type II sums of squares for the interaction are computed from the difference in residual sums of squares for the following models:

y = constant + a + b
y = constant + a + b + a*b

For the B effect, difference the following models:

y = constant + a + b
y = constant + a

For the A effect, difference the following (this is not the same as for Type I):

y = constant + a + b
y = constant + b

In summary, include interactions of the same order as well as all lower order interactions and main effects when differencing to get an interaction. When getting sums of squares for a main effect, difference against all other main effects only. Unlike for Type I SS, Type II sums of squares are invariant to the order in which effects are entered into the model in tests of main effects.

Type III. Type III sums of squares are the default for ANOVA in all programs but R, and are much simpler to understand. Simply difference from the full model, leaving out only the term in question. For example, the Type III sum of squares for A is taken from the following two models:

y = constant + b + a*b
y = constant + a + b + a*b

Σ Which of these three methods of preferable if data is unbalanced is a bit of a controversial and very complex topic. A first general advice is to have balanced data! If data is slightly unbalanced, **most scholars recommend using Type III SS**. However, if inbalance is more serious, you will need to read and think more about this in order to make maximal sense out of your analysis...

Note that in **ANODEV** (when you compare deviance between models “manually”) in generalized linear model fitting, normally follows a “**type II approach”**...