MORE HYPOTHESIS TESTING FOR TWO-WAY ANOVA

What do we do after testing for interaction?

This depends on whether or not interaction is significant (statistically or otherwise) *and* on what the original questions were in designing the experiment *and* on whether or not the analyzer wishes to engage in data-snooping *and* on the context of the experiment. We will spend a while discussing this.

I. If we *reject* H_0^{AB} (i.e., assume there *is* interaction) then it is usually inappropriate to test for main effects (that is, the contributions of the two factors A and B separately), since the question of what a "main effect" is in the presence of interaction is unclear. (How can you "separate out" the effect of A from the interaction if there is interaction?) Instead, it is usually preferable to use the equivalent cell-means model to examine contrasts in the treatment combinations.

II. If we *do not reject* H_0^{AB} (i.e., decide there is *no* interaction), then we are usually interested in main effects. These can be tested within the complete model. *Staying with this model is advisable rather than switching to the inequivalent main-effects model.*

- Switching based on data essentially reduces power and increases type I error.
- Software is set up to continue the analysis conveniently within the complete model.

Testing the contribution of each factor in the complete model (equal sample sizes)

Note: We are still assuming equal sample sizes (i.e., balanced design).

We wish to test whether or not the factor A is needed in the model. Recall that the model states

$$Y_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijt}$$

A occurs in the model through the terms α_i and $(\alpha\beta)_{ij}$ So "A is not needed in the model" means that the contribution of these two terms is independent of the level of A. That is,

$$\alpha_i + (\alpha\beta)_{ij} = \alpha_s + (\alpha\beta)_{sj} \text{ for all i, s, and j.}$$

Thus, the null hypothesis "A is not needed in the model" can be stated as

$$H_0: \alpha_i + (\alpha\beta)_{ij} = \alpha_s + (\alpha\beta)_{sj} \text{ for all i, s, and j}$$

with alternate hypothesis

$$H_a: \alpha_i + (\alpha\beta)_{ij} \neq \alpha_s + (\alpha\beta)_{sj} \text{ for at least one combination i, s, and j}$$

The textbook does not explicitly mention this H_0. Instead, it lists two possible null hypotheses. The first is

$$H_0^A: \alpha_1^* = \alpha_2^* = \ldots \alpha_a^*$$

(with H_a^A: At least two of the α_i^*'s are different),

where $\alpha_i^* = \alpha_i + (\overline{\alpha\beta})_{i\bullet}$. That is, the test is whether or not the levels of A, averaged over the levels of B, have the same average effect on the response. (Recall that the α_i^*'s occurred previously in the notes Analysis of Variance for the Two-Way Complete Model.)

The second null hypothesis mentioned in the textbook is

H_0^{A+AB}: H_0^A and H_0^{AB} are both true.

What are the connections between these three possible null hypotheses?

i) Clearly, H_0^{A+AB} implies H_0^A.

ii) The following calculations show that H_0 implies H_0^{A+AB}:

If H_0 is true, then $\alpha_i + (\alpha\beta)_{ij} = \alpha_s + (\alpha\beta)_{sj}$ for all i, s, and j. Averaging over the subscript j gives

$$\alpha_i + (\overline{\alpha\beta})_{i\bullet} = \alpha_s + (\overline{\alpha\beta})_{s\bullet} \text{ for all i and s,}$$

which says H_0^A is true.

Subtracting this from the original equation,

$$(\alpha\beta)_{ij} - (\overline{\alpha\beta})_{i\bullet} = (\alpha\beta)_{sj} - (\overline{\alpha\beta})_{s\bullet} \text{ for all i, j, and s.}$$

Rearranging,

$$(\alpha\beta)_{ij} - (\alpha\beta)_{sj} = (\overline{\alpha\beta})_{i\bullet} - (\overline{\alpha\beta})_{s\bullet}$$

But the right side is independent of j, so we conclude

$$(\alpha\beta)_{ij} - (\alpha\beta)_{sj} = (\alpha\beta)_{iq} - (\alpha\beta)_{sq} \text{ for all i, s, j, and q,}$$

which says there is no interaction – i.e., H_0^{AB} is true

iii) The following shows that H_0^{A+AB} implies H_0:

If H_0^{A+AB} is true, then so is H_0^{AB}, so

$$(\alpha\beta)_{ij} - (\alpha\beta)_{sj} = (\alpha\beta)_{iq} - (\alpha\beta)_{sq} \text{ for all i, s, j, and q.}$$

Averaging over q and rearranging,

$$(\alpha\beta)_{ij} - (\overline{\alpha\beta})_{i\bullet} = (\alpha\beta)_{sj} - (\overline{\alpha\beta})_{s\bullet} \text{ for all i, j, and s.}$$

Add this to the equation for H^A to get

$$\alpha_i + (\alpha\beta)_{ij} = \alpha_s + (\alpha\beta)_{sj} \text{ for all i, j, and s,}$$

which says H_0 is true.

(Combining what we have so far: H_0 and H_0^{A+AB} are equivalent, and imply H_0^A.)

iv) Does H_0^A imply H_0^{A+AB} (equivalently, H_0)?

No! Consider the example where $\mu = 0$, $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0$, $(\alpha\beta)_{11} = (\alpha\beta)_{22} = 0$, $(\alpha\beta)_{12} = (\alpha\beta)_{21} = 1$. Thus

$$Y_{11} = \varepsilon_{11},\ Y_{12} = 1 + \varepsilon_{12},\ Y_{21} = 1 + \varepsilon_{21},\ Y_{22} = \varepsilon_{22}$$

Then $\alpha_1^* = \alpha_1 + (\overline{\alpha\beta})_{1\bullet} = 0 + (0 + 1)/2 = ½$ and $\alpha_2^* = \alpha_2 + (\overline{\alpha\beta})_{2\bullet} = 0 + (1+0)/2 = ½$, so H_0^A *is* true. But H_0^{AB} is *not* true. (Draw an interaction plot!)

The test for H_0^A is the one that is automatic in most software. *We will take the perspective that it does not make sense to test for a main effect of A unless there is no interaction*, so using this test will not cause problems. (But if you ever see a paper that tests for "main effects" when there is interaction, be cautious in the interpretation. Do *not* interpret the null hypothesis as saying "A has no effect;" it just means that "the levels of A, *averaged over the levels of B*, have the same average effect on the response.")

To test H_0^A, we will again use an F test comparing the full model with a reduced model: the one where H_0^A is true. If sample sizes are equal, it can be shown that the least squares estimate of $E[Y_{ijt}]$ under this new reduced model (i.e, under H_0^A) is

$$\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} + \bar{y}_{\cdots},$$

giving sum of squares for the reduced model

$$ssE_0^A = \sum_i \sum_j \sum_t (y_{ijt} - \bar{y}_{ij\cdot} + \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2,$$

which by appropriate algebraic manipulations becomes

$$ssE_0^A = \sum_i \sum_j \sum_t (y_{ijt} - \bar{y}_{ij\cdot})^2 + br\sum_{i=1}^{a} (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2$$

$$= ssE + br\sum_{i=1}^{a} (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2,$$

so the *sum of squares for treatment factor A* is

$$ssA = ssE_0^A - ssE$$

$$= br\sum_{i=1}^{a} (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2$$

$$= (1/br)\sum_{i=1}^{a} (y_{i\cdot\cdot})^2 - (y_{\cdots})^2/abr,$$

which resembles the formula for ssT used to test equality of effects in one-way analysis of variance. The reasoning behind the test used is: If H_0^A is true, then ssA should be small compared to ssE, so we will have evidence lending doubt to H_0^A if ssA/ssE is unusually large.

If SSA is the random variable corresponding to ssA, it can be shown that when H_0^A is true and sample sizes are equal,

i) $SSA/\sigma^2 \sim \chi^2(a-1)$

ii) SSA and SSE are independent.

Thus, when sample sizes are equal and H_0^A is true,

$$\frac{SSA/(a-1)\sigma^2}{SSE/(n-ab)\sigma^2} = \frac{MSA}{MSE} \sim F(a-1,n-ab)$$

Since msA/msE is just a scalar multiple of the ratio ssA/ssE, we can use msA/msE as a test statistic, rejecting for large values.

Similarly, we can form the *sum of squares for treatment factor B* and obtain an F-test based on

$$\frac{SSB/(b-1)\sigma^2}{SSE/(n-ab)\sigma^2} = \frac{MSB}{MSE} \sim F(b-1,n-ab)$$

for

H_0^B: $\beta_1^* = \beta_2^* = \dots \beta_b^*$

where $\beta_j^* = \beta_j + (\overline{\alpha\beta})_{\cdot j}$

That is, the test is whether or not the levels of B, averaged over the levels of A, have the same average effect on the response.

The alternate hypothesis is

H_a^B: At least two of the β_j*'s are different.

Analysis of Variance Table

For each of the three tests (for interaction, effect of A and effect of B), we have a corresponding sum of squares, ssAB, ssA, and ssB. We also have the error sum of squares, ssE. If we add up the formulas for these three sums of squares and do appropriate algebraic manipulations, we will get

$$\text{ssA} + \text{ssB} + \text{ssAB} + \text{ssE} = \sum_i \sum_j \sum_t (y_{ijt} - \bar{y}_{...})^2.$$

This last sum of squares is called the *total sum of squares*, denoted ssT or sstot. It can be seen as a measure of the total variability of the data without taking into account either A or B. Similarly, ssE is a measure of the variability taking into account A, B and their interaction; ssA is a measure of the variability taking B into account but not A, and ssB is a measure of the variability taking A into account but not B.

The sums of squares and the additional information used in the tests for A, B and AB are traditionally summarized in an *Analysis of Variance Table* with one line each for A, B, AB, error, and "total sum of squares"

sstot (or ssT) = ssA + ssB + ssAB + ssE

Still assuming equal sample sizes, sstot = $\sum_i \sum_j \sum_t (y_{ijt} - \bar{y}_{...})^2$. It can be seen as a measure of the total variability of the data without taking into account either A or B. Similarly, ssE is a measure of the variability taking into account A, B and their interaction; ssA is a measure of the variability taking B into account but not A, and ssB is a measure of the variability taking A into account but not B.

Interpreting ANOVA tests

Interpretation requires thought -- we need to take into account the purpose of the study, the context, multiple comparisons, and whether or not we are willing to do data snooping. Interpretation can sometimes be frustrating -- for example, what if the test for interaction is significant, but the test for one of the factors is not?

Examples: Battery and reaction time.

Note: When sample sizes are unequal, the formulae for the sums of squares are more complicated, and the corresponding random variables are not independent. More on this later.