ANALYSIS OF VARIANCE FOR THE COMPLETE TWO-WAY MODEL

The first thing we need to test for in two-way analysis of variance is whether there is interaction. Since "no interaction" means that all of the lines in the interaction plot have the same slopes, we can state the null hypothesis

H_0^{AB}: There is no interaction

as

$$H_0^{AB}: [(\alpha\beta)_{ij} - (\alpha\beta)_{iq}] - [(\alpha\beta)_{sj} - (\alpha\beta)_{sq}] = 0 \text{ for all } i \neq s,\ j \neq q$$

The alternate hypothesis is then

$$H_a^{AB}: [(\alpha\beta)_{ij} - (\alpha\beta)_{iq}] - [(\alpha\beta)_{sj} - (\alpha\beta)_{sq}] \neq 0 \text{ for at least one instance of } i \neq s,\ j \neq q$$

For *equal sample sizes* we can test the null hypothesis in a manner similar to the test for one-way analysis of variance: with an F-test testing the submodel (reduced model) determined by H_0^{AB} against the full model. We do this by comparing the sum of squares for error ssE under the full model with the sum of squares for error ssE_a^{AB} under the reduced model. This difference

$$ssAB = ssE_a^{AB} - ssE$$

is called the *sum of squares for the interaction AB.* We reject H_a^{AB} in favor of H_a^{AB} when ssAB is large relative to ssE.

Recall that the full model states:

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

Since this is equivalent to the cell-means model, which is a one-way model, we know that

$$ssE = \sum_i \sum_j \sum_t \hat{e}_{ijt}^{\,2} = \sum_i \sum_j \sum_t \left(y_{ijt} - \bar{y}_{i\cdot}\right)^2,$$

The usual types of algebraic manipulations show that ssE has the alternate formulas

$$ssE = \sum_i \sum_j \sum_t y_{ijt}^2 - \sum_i \sum_j r_{ij} \bar{y}_{ij\cdot}^{\,2}$$

$$= \sum_i \sum_j \sum_t y_{ijt}^2 - \sum_i \sum_j y_{ij\cdot}^2 / r_{ij}$$

If H_a^{AB} is true, then averaging over s and q gives the equations

$$[(\alpha\beta)_{ij} - (\overline{\alpha\beta})_{i\cdot}] - [(\overline{\alpha\beta})_{\cdot j} - (\overline{\alpha\beta})_{\cdot\cdot}] = 0 \qquad \text{for each } i, j$$

So under the reduced model,

$$(\alpha\beta)_{ij} = (\overline{\alpha\beta})_{i\cdot} + (\overline{\alpha\beta})_{\cdot j} - (\overline{\alpha\beta})_{\cdot\cdot}$$

so

$$\begin{aligned} Y_{ijt} &= \mu + \alpha_i + \beta_j + (\overline{\alpha\beta})_{i\cdot} + (\overline{\alpha\beta})_{\cdot j} - (\overline{\alpha\beta})_{\cdot\cdot} + \varepsilon_{ijt} \\ &= [\mu - (\overline{\alpha\beta})_{\cdot\cdot}] + [\alpha_i + (\overline{\alpha\beta})_{i\cdot}] + [\beta_j + (\overline{\alpha\beta})_{\cdot j}] + \varepsilon_{ijt} \\ &= \mu^* + \alpha_i^* + \beta_j^* + \varepsilon_{ijt} \end{aligned}$$

Thus the reduced model is of the form of the main effects model.

Estimates for the main effects model, assuming equal sample sizes:

Least squares may be used to find estimators of the parameters under the Main Effects Model assumption

$$Y_{ijt} = \mu + \alpha_i + \beta_j + \varepsilon_{ijt}\,.$$

(See p. 161 of the text for more details.)

For *equal* sample sizes (i.e., *balanced* anova), the resulting normal equations are readily solvable (with added constraints), yielding least squares estimator

$$(*) \qquad \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = \bar{y}_{i\cdot\cdot} + \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot}$$

for $E[Y_{ijt}] = \mu + \alpha_i + \beta_j$.

Note: 1. Recall that for the complete model, the least squares estimators were

$$\hat{\mu} = \bar{y}_{\cdot\cdot\cdot}$$
$$\hat{\alpha}_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot}$$
$$\hat{\beta}_j = \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot},$$

from which it follows that the least squares estimate for $\mu + \alpha_i + \beta_j$ is the same in both models. However, in the complete model, $E[Y_{ijt}] = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$, which is not the same as $E[Y_{ijt}]$ for the main effects model unless $(\alpha\beta)_{ij} = 0$.

2. For *unequal* sample sizes, the normal equations are much messier.

From (*), we see that for the main effects model,

$$ssE = \sum_i \sum_j \sum_t (y_{ijt} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2$$

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

which by the usual types of algebraic manipulations can be re-expressed as

$$\sum_i \sum_j \sum_t y_{ijt}^2 - br\sum_i \bar{y}_{i\cdot\cdot}^2 - ar\sum_j \bar{y}_{\cdot j\cdot}^2 + abr\,\bar{y}_{\cdots}^2$$

$$= \sum_i \sum_j \sum_t y_{ijt}^2 - \frac{1}{br}\sum_i y_{i\cdot\cdot}^2 - \frac{1}{ar}\sum_j y_{\cdot j\cdot}^2 + \frac{1}{abr} y_{\cdots}^2$$

Continuing with the test for interaction in the complete two-way model

Applying the above to the reduced model

$$Y_{ijt} = \mu^* + \alpha_i^* + \beta_j^* + \varepsilon_{ijt}$$

in the test for interaction in the complete two-way model, we get (assuming equal sample sizes)

$$ssE_a^{AB} = \sum_i \sum_j \sum_t (y_{ijt} - \hat{\mu}^* - \hat{\alpha}_i^* - \hat{\beta}_j^*)^2$$

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

which by the usual types of tricks can be re-expressed as

$$\sum_i \sum_j \sum_t (y_{ijt} - \bar{y}_{ij\cdot})^2 + \sum_i \sum_j \sum_t (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} + \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2.$$

Since the first term is just ssE for the full model, we have

$$ssAB = ssE_a^{AB} - ssE$$

$$= \sum_i \sum_j \sum_t (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} + \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2$$

$$= r\sum_i \sum_j (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} + \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2,$$

which can be re-expressed as

$$\frac{1}{r}\sum_i \sum_j y_{ij\cdot}^2 - \frac{1}{br}\sum_i y_{i\cdot\cdot}^2 - \frac{1}{ar}\sum_j y_{\cdot j\cdot}^2 + \frac{1}{abr} y_{\cdots}^2$$

Using the remaining two model assumptions, that the ε_{ijt} are independent random variables and each $\varepsilon_{ijt} \sim N(0, \sigma^2)$, it can be shown that for the corresponding random variables SSAB and SSE, when H_0^{AB} is true and sample sizes are equal,

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

ii) $SSE/\sigma^2 \sim \chi^2(n - ab)$

iii) SSAB and SSE are independent.

Thus, when sample sizes are equal and H_0^{AB} is true,

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

So we can use msAB/msE as a test statistic, rejecting for large values.

Further analysis after testing for interaction

I. If we reject H_0^{AB}, then it is usually inappropriate to test for main effects. 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}, then we are usually interested in main effects. These can be tested within the complete model (and staying with this model is advisable rather than switching to the inequivalent main-effects model.)

Testing main effects with the complete model (equal sample sizes)

We are now assuming that H_0^{AB} is true. So the model can be stated as

$$Y_{ijt} = \mu^* + \alpha_i^* + \beta_j^* + \varepsilon_{ijt}$$

where

$$\mu^* = \mu - (\overline{\alpha\beta})_{..}$$

$$\alpha_i^* = \alpha_i + (\overline{\alpha\beta})_{i.}$$

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

The hypothesis, "Factor A has no effect on the mean response," can be stated as

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

We will again use an F test comparing the full model with the reduced model where all α_i^* 's are equal. 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

$$ssA = ssE_0^A - ssE$$

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

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

which resembles the formula for ssT used to test equality of effects in one-way analysis of variance.

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^{AB} is true,

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

So 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

$$H_0^B: \beta_1 * = \beta_1 * = \ldots = \beta_b *$$

against

$$H_a^B: \beta_1 * = \beta_1 * = \ldots = \beta_b *$$

Analysis of Variance Table

The statistics for the three tests are typically summarized in an *Analysis of Variance Table* with one line each for A, B, AB, and "total sum of squares" sstot = ssA + ssB + ssAB + ssE

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

Example: Battery experiment