ANALYSIS OF BALANCED FACTORIAL DESIGNS

Estimates of model parameters and contrasts can be obtained by the method of Least Squares. Additional constraints must be added to estimate non-estimable parameters.

Example: The cell means are estimable. For three factors, the Least Squares estimates of the cell means are $\bar{y}_{ijk\bullet}$.

From the cell-means model and one-way analysis of variance, we have the following:

- The *error sum of squares* ssE is the sum of squared deviations from the fits $\bar{y}_{ijk\bullet}$.
  Example: For 3 factors, ssE = $\sum_i \sum_j \sum_k \sum_t (y_{ijkt} - \bar{y}_{ijk\bullet})^2$
- ssE has associated degrees of freedom n – v
  Example: For 3 factors, ssE has abcr – abc = abc(r-1) degrees of freedom.
- The *mean square error* msE is ssE divided by its degrees of freedom.
  Example: For 3 factors, msE = ssE/ abc(r-1) degrees of freedom.
- msE is an estimate for the variance σ^2.
- The standard error of the residuals $y_{ijkt} - \bar{y}_{ijk\bullet}$ is ssE/(n-1)

Before doing inference, we should use the fits and residuals to check the model assumptions of independent, normal errors with constant variance.

Example: Pollution filter data

Note: To get a column for Treatment easily, use 100*SIZE + 10*TYPE + SIDE

Hypothesis Tests and Analysis of Variance Table:

Submodels can be tested against larger models by F-tests, with F-statistic obtained as a ratio of mean squares.

The *error sum of squares* of a model is the sum of the squared deviations from the fits from that model.

Example: For three factors, the error sum of squares for a model is $\sum_i \sum_j \sum_k \sum_t (y_{ijkt} - \hat{y})^2$, where $\hat{y}$ is the fit for the model.

Special case I: Since the fit for the complete model is $\bar{y}_{ijk\bullet}$, the error sum of squares for the complete model is ssE mentioned above.

Special case II: The *total sum of squares* is the error sum of squares for the model $y_{ijkt} = \mu + \varepsilon_{ijkt}$

Since the fit for this model is the overall mean, the total sum of squares can be described as the sum of the squared deviations from the overall mean.

Example: For 3 factors, sstot = $\sum_i \sum_j \sum_k \sum_t (y_{ijkt} - \bar{y}_{....})^2$

The total sum of squares has n-1 associated degrees of freedom, where n = total number of observations (e.g., abcr for 3 factors).

Recall that for three factors, the null hypothesis for "no three-way interaction" is

H_0^{ABC}: $[(\alpha\beta\gamma)_{i+1,jk} - (\alpha\beta\gamma)_{ijk}] - [(\alpha\beta\gamma)_{i+1,qk} - (\alpha\beta\gamma)_{iqk}]$
- $[(\alpha\beta\gamma)_{i+1,jr} - (\alpha\beta\gamma)_{ijr}] - [(\alpha\beta\gamma)_{i+1,qr} - (\alpha\beta\gamma)_{iqr}] = 0$ for all i, j, k, q, r

If we let ssE_0^{ABC} denote the error sum of squares for the submodel where H_0^{ABC} is true, then we define *the sum of squares for three-way interaction* to be
$ssABC = ssE_0^{ABC} - ssE$
It has associated degrees of freedom (a-1)(b-1)(c-1). We define
msABC = ssABC/ (a-1)(b-1)(c-1).
Then MSABC/MSE has an F distribution with (a-1)(b-1)(c-1) and n-v degrees of freedom. This gives us an F-test for H_0^{ABC}.

Example: Pollution filter data.

To explain the rest of the ANOVA output, consider, for example
$H_0^{BC + ABC}$: There is no three-way interaction and no BC interaction.
This can be expressed in terms of the parameters as
$H_0^{BC + ABC}$: $[(\beta\gamma)_{ij} - (\beta\gamma)_{iq}] - [(\beta\gamma)_{sj} - (\beta\gamma)_{sq}] = 0$ for all $i \neq s$, $j \neq q$
And $[(\alpha\beta\gamma)_{i+1,jk} - (\alpha\beta\gamma)_{ijk}] - [(\alpha\beta\gamma)_{i+1,qk} - (\alpha\beta\gamma)_{iqk}]$
- $[(\alpha\beta\gamma)_{i+1,jr} - (\alpha\beta\gamma)_{ijr}] - [(\alpha\beta\gamma)_{i+1,qr} - (\alpha\beta\gamma)_{iqr}] = 0$ for all i, j, k, q, r

We define
ssE_0^{BC+ABC} = the sum of squares for the model assuming $H_0^{BC + ABC}$ is true,
and *the sum of squares for BC*
$ssBC = ssE_0^{BC+ABC} - ssE$.
Then ssBC has (b-1)(c-1) degrees of freedom. Defining msBC = ssBC/(b-1)(c-1), we get F-statistic msBC/msE, with degrees of freedom (b-1)(c-1) and n-v. This gives us an F-test for $H_0^{BC + ABC}$.

In general, the degrees of freedom for a sum of squares for a main effect or interaction is the product of one less than the number of levels for each factor involved in the effect or interaction.

Examples:

Sum of squares for	Degrees of freedom
A	a-1
AB	(a-1)(b-1)
BC	(b-1)(c-1)
ABC	(a-1)(b-1)(c-1)

The total sum of squares partitions into (is the sum of) the error sum of squares plus the sums of squares for all main effects and interactions in the model.

The degrees of freedom add correspondingly. For example, for the complete 3-way model, this is just the algebraic identity

$$(a-1)(b-1)(c-1) + (a-1)(b-1) + (a-1)(c-1) + (b-1)(c-1) + (a-1) + (b-1) + (c-1) + abcr - abc = abcr - 1$$

Note: If we are dealing with another model, the error sum of squares for that model will have a different number of degrees of freedom. The degrees of freedom for error are usually most easily found by subtracting the other degrees of freedom from the total degrees of freedom.

The *mean square* for a main effect or interaction is the sum of squares for that effect/interaction divided by its degrees of freedom.

Example: msBC = ssBC/(b-1)(c-1)

The F-statistic for a main effect or interaction is the mean square for the effect/interaction divided by the mean square error, with corresponding degrees of freedom in the numerator in denominator.

Contrasts

Contrasts are defined and estimated as in the cell means model. There are various labels that are given to certain types of contrasts; see p. 199 for details. Contrasts can also be defined relative to submodels, in which case parameters not present in the submodel are omitted from the contrast. The various methods for simultaneous confidence intervals still apply. There is a modification that can be used for finding simultaneous confidence intervals for contrasts in the levels of a single factor: Replace v by the number of levels of the factor in question, and replace r by the number of observations on each level of the factor of interest. (See p. 205 for an example.)