LEAST SQUARES ESTIMATES FOR TWO-WAY MODELS

Cell-means Model

$Y_{ijt} = \mu + \tau_{ij} + \varepsilon_{ijt}$.
The ε_{ijt} are independent random variables.
Each $\varepsilon_{ijt} \sim N(0, \sigma^2)$

If we consider each combination of levels of A and levels of B as one treatment, the cell-means model is just a special case of the one-way ANOVA model, so the least squares method as developed there fits:

$\bar{Y}_{ij\bullet}$ is an unbiased estimator of $\mu + \tau_{ij}$,

with variance σ^2/r_{ij}

Notation: $\hat{\mu} + \hat{\tau}_{ij}$ is the estimate $\bar{y}_{ij\bullet}$ of $\mu + \tau_{ij}$

Two-way complete model.

$Y_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijt}$.
The ε_{ijt} are independent random variables.
Each $\varepsilon_{ijt} \sim N(0, \sigma^2)$

Using the method of least squares directly on this model gives 1 + a + b + ab normal equations, with a + b + 1 linear dependencies. Adding the constraints

$$\sum_{i=1}^{a} \hat{\alpha}_i = 0 \qquad \sum_{j=1}^{b} \hat{\beta}_j = 0$$

$$\sum_{i=1}^{a} (\alpha\beta)\hat{}_{ij} = 0, \ j = 1, \dots, b$$

$$\sum_{j=1}^{b} (\alpha\beta)\hat{}_{ij} = 0, \ i = 1, \dots, a,$$

gives a + b + 1 *independent* additional constraints.

The normal equations plus these constraints have solution

$$\hat{\mu} = \bar{y}_{\bullet\bullet\bullet}$$

$$\hat{\alpha}_i = \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}, \text{ i = 1, 2, … , a}$$

$$\hat{\beta}_j = \bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}, \text{ j = 1, 2, … , b}$$

$$(\alpha\beta)^\wedge_{ij} = \bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet},$$

i = 1, 2, … , a, j = 1, 2, … , b

Note: $\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + (\alpha\beta)^\wedge_{ij} = \ldots = \bar{y}_{ij\bullet}$, which is the same as $\hat{\mu} + \hat{\tau}_{ij}$ from the cell-means model.

Fits and residuals

Fits (or *fitted values*): the least squares estimates for the observations:

$$\hat{y}_{ijt} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + (\alpha\beta)^\wedge_{ij} = \hat{\mu} + \hat{\tau}_{ij} = \bar{y}_{ij\bullet}$$

Residuals:

$$\hat{e}_{ijt} = y_{ijt} - \hat{y}_{ijt}$$

Since the complete and cell-means models are equivalent, and the latter is a special case of the one-way ANOVA model, the sample variance of the residuals is

$$\sum_{i=1}^{t}\sum_{t=0}^{r_i}\sum_{j=1}^{r_j} \hat{e}_{ijt}^{\,2}/(n-1) = ssE/(n-1).$$

Define *standardized residuals* as before:

$$z_{ijt} = \frac{\hat{e}_{ijt}}{\sqrt{ssE/(n-1)}}$$

Estimable functions: As with the one-way model, a function of the parameters is called *estimable* if it has a unique least-squares estimate (without adding additional constraints). Examples of estimable functions include:

- $\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$, which has unique least squares estimate $\bar{y}_{ij\bullet}$

- Any function that is a linear combination of the left-hand sides of the normal equations.

- Most contrasts that are of interest. (More later)

Obtaining least squares estimates in Minitab

Cell-means model: Just use One-way ANOVA

Two-way complete model:

- The data need to be arranged so that there is a column for each factor. The command "Code Data Values" on the Manip menu is convenient for this.
- Then use Stat > ANOVA > Balanced ANOVA

Example: Battery experiment

Checking Model Assumptions

Do this before drawing any conclusions from the model.

Proceed as for one-way ANOVA, with some minor exceptions:

1. Check the fit of the model – plot (standardized) residuals against factors included in the model and, if possible, factors not included. A non-random pattern suggests lack of fit.

2. Check for outliers – using standardized residuals makes this easier.

3. Check for independence of error terms – plot residuals against order, and other time or spatial variables, or any other variables that might have an effect. A non-random pattern suggests lack of independence.

4. Check for equal variance
   - Plot residuals against fits and against each factor.
   - Rule of thumb: If the ratio s^2_{max}/ s^2_{min} of the largest treatment variance to the smallest does not exceed 3 (some say 4), then the inference procedures for the equal variance model are still valid. (Remember: even if the model assumptions are valid, a large ratio might occur by chance, especially if sample sizes are small. So consider any theoretical considerations available as well.)
   - p-values in tests may help make a decision in borderline cases.
   - Use the check applied to each factor separately if there are not enough observations in each cell to calculate the sample standard deviations for each cell.

5. Check for normality -- use a normal plot of residuals.

Example: Battery data – the only thing new is to plot against each factor separately.

Contrasts:

Treatment contrasts: Since the cell-means model is a special case of the one-way ANOVA model, we know that treatment contrasts such as the following are estimable:

- Pairwise contrasts

$$\tau_{ij} - \tau_{sh} = \alpha_i + \beta_j + (\alpha\beta)_{ij} - [\alpha_s + \beta_h + (\alpha\beta)_{sh}]$$

- *Simple contrasts* are of the form $\sum_{i=1}^{a} c_{ij}\tau_{ij}$ where $\sum_{i=1}^{a} c_{ij} = 0$, or $\sum_{j=1}^{b} c_{ij}\tau_{ij}$, where $\sum_{j=1}^{b} c_{ij} = 0$.

- *Simple pairwise differences* are of the form $\tau_{ij} - \tau_{sj}$ or $\tau_{ij} - \tau_{ih}$

- Differences of averages of the τ_{ij}'s.

The confidence interval methods of Chapter 4 are still applicable.

Interaction contrasts: Looking at an interaction plot, we can see that we can measure whether or not there is interaction by comparing slopes of the lines on the interaction plot: Non-parallel lines (i.e., different slopes) for two adjacent levels indicate interaction.

Exercise: In a model with two levels of each of the two factors, what contrast would measure the difference in the slopes of the lines between two adjacent levels of a factor? (Express the contrast both in terms of the cell-means model and in terms of the complete two-way model.)