

## LEAST SQUARES ESTIMATES FOR TWO-WAY MODELS

### *Cell-means Model*

$$Y_{ijt} = \mu + \tau_{ij} + \varepsilon_{ijt}.$$

The  $\varepsilon_{ijt}$  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:

$\bar{Y}_{ij}$  is an unbiased estimator of  $\mu + \tau_{ij}$ , with variance  $\sigma^2/r_{ij}$ .

We use  $\hat{\mu} + \hat{\tau}_{ij}$  to refer to the estimate  $\bar{y}_{ij}$ .

### *Two-way complete model.*

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

The  $\varepsilon_{ijt}$  are independent random variables.

Each  $\varepsilon_{ijt} \sim N(0, \sigma^2)$

If we use the method of least squares directly on this model, we obtain  $1 + a + b + ab$  normal equations, with  $a + b + 1$  linear dependencies. If we add the constraints

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

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

we obtain  $a + b + 1$  *independent* additional constraints, and the normal equations plus these constraints have the solution

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

$$\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y}_{\dots}, i = 1, 2, \dots, a$$

$$\hat{\beta}_j = \bar{y}_{\cdot j} - \bar{y}_{\dots}, j = 1, 2, \dots, b$$

$$(\alpha\beta)^{\wedge}_{ij} = \bar{y}_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\dots}, i = 1, 2, \dots, a, j = 1, 2, \dots, b$$

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

### *Fits and residuals*

The *fits* (or *fitted values*) are the least squares estimates for the observations:

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

The *residuals* are

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

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  $\sum_{i=1}^I \sum_{t=0}^{r_i} \sum_{j=1}^{r_j} \hat{e}_{ijt}^2 / (n-1) = ssE/(n-1)$ . We define the *standardized residuals* as before:

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

**Estimable functions:** As with the one-way model, we call a function of the parameters *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  $\bar{y}_{ij}$ .
- Any function that is a linear combination of the left-hand sides of the normal equations will be estimable.
- 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. If they are not already arranged this way, the command "Code Data Values" on the Manip menu is convenient to do so. Then use Stat > ANOVA > Balanced ANOVA

*Example:* Battery experiment

### **Checking Model Assumptions**

This should be done before drawing any conclusions from the model. We can 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 it easier to detect outliers.
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 (from simulation studies): If the ratio  $s_{\max}^2 / s_{\min}^2$  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. (Bear in mind that even if the model assumptions are valid, a large ratio might occur by chance, especially if sample sizes are small. So any theoretical considerations available might be helpful as well.)
  - p-values in tests may help make a decision in borderline cases.

- In two-way ANOVA, there might not be enough observations in each cell to calculate the sample standard deviations for each cell, in which case use the check applied to each factor separately.
5. Check for normality by using 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  $\tau_{ij}$ '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) between 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? (Express the contrast both in terms of the cell-means model and in terms of the complete two-way model.)