ESTIMATING PARAMETERS AND VARIANCE

Least Squares Estimates

Our model (in its various forms) involves various *parameters*: μ, σ, the μ_i's, and the τ_i's. Our purpose in doing an experiment is to estimate or compare certain of these parameters (and sometimes certain functions of these parameters) using our data.

Our data are the values y_{it} for the random variables Y_{it} that we observe in our experiment. In other words:

The data obtained from treatment 1 (or level 1 or population 1) are $y_{11}, y_{12}, \dots, y_{1r_1}$;
the data obtained from treatment 2 (or level 1 or population 2) are $y_{21}, y_{22}, \dots, y_{2r_2}$;
and so on.

To estimate certain parameters or functions of them, we use the *method of least squares*. We illustrate the idea for the means model:

Our model is $Y_{it} = \mu_i + \varepsilon_{it}$. We seek an estimate $\hat{\mu}_i$ for μ_i. We would like to find $\hat{\mu}_i$'s with the property that when we apply the estimated model to the data, the errors are as small as possible. In other words, if our estimates are $\hat{\mu}_i$, we would like the "error terms" (*residuals*) $e_{it} = y_{it} - \hat{\mu}_i$ to be as small as possible. But we want them to be small *collectively*. So we might try to minimize their sum. But positives and negative will cancel out, so this doesn't really seem like a very good idea. We might try to minimize the sum of the absolute values of the errors. This is reasonable, but technically not very easy. What does work pretty well is to minimize the sum of the squared errors: $\sum_{i,t} e_{it}^2 = \sum_{i=1}^{v}\sum_{t=1}^{r} e_{it}^2$. This amounts to minimizing the function

$$f(m_1, m_2, \dots, m_v) = \sum_{i,t} (y_{it} - m_i)^2,$$

which we can do by calculus.

Exercise: Do the calculus to find the least squares estimates $\hat{\mu}_i$ of the μ_i's.

Using least squares for the means model works out cleanly. However, if we try least squares with the effects model, we end up with the following v+1 equations ("normal equations") in the estimates $\hat{\mu}$ and $\hat{\tau}_i$ for μ and the τ_i's, respectively:

$$y_{..} - n\hat{\mu} - \sum_{i=1}^{v} r_i\hat{\tau}_i = 0$$

$$y_{i.} - r_i\hat{\mu} - r_i\hat{\tau}_i = 0, \qquad i = 1, 2, \ldots, v.$$

(The details of obtaining the equations will be a homework problem.)

If we add the last v equations, we get the first one. Thus we only have v independent equations in the v + 1 unknowns $\hat{\mu}$, $\hat{\tau}_1, \ldots, \hat{\tau}_v$-- so there are infinitely many solutions. To get around this problem, it is customary to impose the constraint

$$\sum_{i=1}^{v} \hat{\tau}_i = 0.$$

This gives us v + 1 equations in the v + 1 unknowns, and there is a unique solution to this set of n + 1 solutions. The constraint is not unreasonable, since we are thinking of the $\hat{\tau}_i$'s as measuring deviations around some common mean.

Comments:
1. Students who have had regression may wish to think about how the necessity of imposing an additional constraint here is connected to the need to have only v - 1 indicator variables for a categorical variable with v categories in regression.

2. Note that, even though the normal equations do not have a unique solution for $\hat{\mu}$, $\hat{\tau}_1, \ldots, \hat{\tau}_v$, the last v equations do give a unique solution for each $\hat{\mu} + \hat{\tau}_i$ -- the same one obtained for $\hat{\mu}_i$ by using least squares with the means model. Similarly, by subtracting pairs of the last v equations, we can obtain unique solutions for the differences $\hat{\tau}_i - \hat{\tau}_j$ -- that is, there are unique least squares estimators for the differences $\tau_i - \tau_j$ (the pair-wise differences in effects). The functions of the parameters $\mu, \tau_1, \ldots, \tau_v$ that do have unique least squares estimates are called *estimable functions*. You can read a little more about them in Sections 3.4.1 and 3.4.4.

3. Functions of the parameters that have the form $\sum_{i=1}^{v} c_i\tau_i$ where $\sum_{i=1}^{v} c_i = 0$ are called *contrasts*. For example, each difference of effects $\tau_i - \tau_j$ is a contrast. This is certainly a quantity which is often of interest in and experiment. Other contrasts, such as differences of averages, may be of interest as well in certain experiments.

Example: An experimenter is trying to determine which type of non-rechargeable battery is most economical. He tests five types and measures the lifetime per unit cost for a sample of each. He also is interested in whether alkaline or heavy duty batteries are most economical as a group. He has selected two types of heavy duty (groups 1 and 2) and three types of alkaline batteries (groups 3, 4, and 5). So to study his second question, he tests the difference in averages, $(\tau_1 + \tau_2)/2 - (\tau_3 + \tau_4 + \tau_5)/3$. Note that this *is* a contrast, since the coefficient sum is 1/2 + 1/2 - 1/3 - 1/3 - 1/3 = 0.

Exercise: Every contrast $\sum_{i=1}^{v} c_i \tau_i$ is a linear combination of the effect differences $\tau_i - \tau_j$ and is estimable, with least squares estimate $\sum_{i=1}^{v} c_i \hat{\tau}_i = \sum_{i=1}^{v} c_i \overline{Y}_{i\cdot}$

3. Since each Y_{it} has the distribution of Y_i and $Y_i \sim N(\mu_i, \sigma^2)$, it follows from standard properties of expected values that $E(\overline{Y}_{i\cdot}) = \mu_i$. Since the Y_{it}'s are independent, it follows from standard variance calculations and properties of normal random variables that $\overline{Y}_{i\cdot} \sim N(\mu_i, \sigma^2/r_i)$.

Exercise: Go through the details of comment (3). Also verify that the least squares estimator $\sum_{i=1}^{v} c_i \overline{Y}_{i\cdot}$ of the contrast $\sum_{i=1}^{v} c_i \tau_i$ (where $\sum_{i=1}^{v} c_i = 0$) has normal distribution with mean $\sum_{i=1}^{v} c_i \tau_i$ and variance $\sum_{i=1}^{v} \frac{c_i^2}{r_i} \sigma^2$. [Hint: You need to establish and use the fact that the $\overline{Y}_{i\cdot}$'s are independent.]

Variance Estimate

If we just consider a single treatment group, the data for that group give sample variance

$$s_i^2 = \frac{\sum_{t=1}^{r_i} (y_{it} - \bar{y}_{i\cdot})^2}{r_i - 1}.$$

The corresponding random variable $S_i^2 = \frac{\sum_{t=1}^{r_i} (Y_{it} - \overline{Y}_{i\cdot})^2}{r_i - 1}$ is an unbiased estimator for the population variance σ^2: $E(S_i^2) = \sigma^2$. (See Ross, Chapter 4 or Wackerly, Chapter 8 if you are not familiar with this.)

As in our discussion of the two-sample t-test, the average of the S_i^2's will then also be an unbiased estimator of σ^2. To take into account different sample sizes we will take a weighted average:

$$S^2 \text{ (or } \hat{\sigma}^2) = \frac{\sum_i (r_i - 1) S_i^2}{\sum_i (r_i - 1)}$$

Note that the denominator equals $\sum_i r_i - \sum_i 1$ = n - v.

Exercise: Check that S^2 is an unbiased estimator of σ^2 -- that is, check that $E(S^2) = \sigma^2$.

Note (using the definition of S_i^2) that the numerator of S^2 is $\sum_{i=1}^{t} \sum_{t=0}^{r_i} \left(Y_{it} - \bar{Y}_i\right)^2$. This expression is called SSE -- the *sum of squares for error* or the *error sum of squares.* So the estimator for variance is often written as

S^2 = SSE/(n-v).

This expression is called MSE -- the *mean square for error* or *error mean square*.

The above are random variables. Their values calculated from the data are:

ssE = $\sum_{i=1}^{t} \sum_{t=0}^{r_i} (y_{it} - \bar{y}_i)^2$

-- also called the *sum of squares for error* or the *error sum of squares*

msE = ssE/(n-v)

-- also called the *mean square for error* or *error mean square*

s^2 = msE -- the unbiased estimate of σ^2 -- also denoted $\hat{\sigma}^2$.

Note:

- $y_{it} - \bar{y}_i$ is sometimes called the it^{th} *residual*, denoted $\hat{e}_{it}$. So ssE = $\sum_{i=1}^{t} \sum_{t=0}^{r_i} \hat{e}_{it}^{\,2}$
- Many people use SSE and MSE for ssE and msE.
- This unbiased estimate of σ^2 is sometimes called the *within groups* (or *within treatments*) *variation*, since it calculates the sample variance within each group and then averages these estimates.