ESTIMATING PARAMETERS AND VARIANCE FOR ONE-WAY ANOVA

**Least Squares Estimates**

Typical purpose of experiment: Use the data to estimate or compare certain of the model parameters (or certain functions of the parameters).

Data: The values \(y_{ij}\) for the random variables \(Y_{ij}\) obtained in our experiment:

From treatment 1: \(y_{11}, y_{12}, \ldots, y_{1n}\);
From treatment 2: \(y_{21}, y_{22}, \ldots, y_{2n}\);
etc.

**Method of least squares.**

Example: means model: \(Y_{ij} = \mu_i + \epsilon_{ij}\).

We seek an estimate \(\hat{\mu}_i\) for \(\mu_i\).

Idea: 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: Make the "estimated error terms" (residuals)
\[ e_{ir} = y_{ir} - \hat{\mu}_i \]
as small as possible -- collectively.

Picture:

How to do this?
Minimize the sum of the errors?
   Positives and negatives will cancel out

Minimize the sum of the absolute values of the errors?
   Technical problems.

Minimize the sum of the squared errors?
   Works reasonably well!

Details:
Least squares for the effects model:

(Some details might be homework.)

We get the following v + 1 equations ("normal equations") in the estimates \( \hat{\mu} \) and \( \hat{\tau}_i \) for \( \mu \) and the \( \tau_i \)'s, respectively:

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

\[
y_{i} - r_i \hat{\mu} - r_i \hat{\tau}_i = 0, \quad i = 1, 2, \ldots, v.
\]

Note: Adding the last \( v \) equations gives the first one.

Thus: only \( v \) independent equations in the \( v + 1 \) unknowns \( \hat{\mu}, \hat{\tau}_1, \ldots, \hat{\tau}_v \) -- hence infinitely many solutions.

Solution to dilemma: Add the constraint

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

Result: \( v + 1 \) independent equations in the \( v + 1 \) unknowns, so a unique solution.

Comments:

1. (For students with regression background) This is related to using only \( v - 1 \) indicator variables for a categorical variable with \( v \) categories in regression.

2. Solve for \( \hat{\mu} + \hat{\tau}_j \):

Solve for \( \hat{\tau}_i - \hat{\tau}_j \):

Estimable functions: The functions of the parameters \( \mu, \tau_1, \ldots, \tau_v \) that do have unique least squares estimates. Examples: \( \mu + \tau \) and \( \tau - \tau_i \).
(See 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
eexample, each difference of effects \( \tau_{i} - \tau_{j} \) is a contrast.
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
\[ 1/2 + 1/2 - 1/3 - 1/3 - 1/3 = 0. \]

(Similarly, we can show that every difference of
averages is a contrast.)

Exercise (Possible homework): 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 \( \sum_{i=1}^{v} c_{i}\bar{\tau}_{i} =
\sum_{i=1}^{v} c_{i}\bar{Y}_{i} \).

4. Since each \( Y_{ii} \) has the distribution of \( Y_{1} \),
and \( Y_{1} \sim N(\mu_{1}, \sigma^{2}) \), it follows from standard properties of
expected values that \( E(Y_{1}) = \mu_{1} \). Since the \( Y_{ii} \)'s are
independent, it follows from standard variance
calculations and properties of normal random
variables that \( \bar{Y}_{1} \sim N(\mu_{1}, \sigma^{2}/n) \).

Exercise: Go through the details of comment (4).
Also verify that the least squares estimator \( \sum_{i=1}^{v} c_{i}\bar{Y}_{i} \) 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 \( \sum_{i=1}^{v} c_{i}^{2}\sigma^{2} \).
[Hint: You need to establish and use the fact that the
\( \bar{Y}_{i} \)'s are independent.]
**Variance Estimate**

For the $i^{th}$ treatment group, the sample variance is

$$s_i^2 = \frac{\sum(y_i - \bar{y}_i)^2}{r_i - 1}.$$

The corresponding random variable

$$S_i^2 = \frac{\sum(y_i - \bar{y}_i)^2}{r_i - 1}$$

is an unbiased estimator for the population variance $\sigma^2$:

$$E(S_i^2) = \sigma^2.$$

(Ross, Chapter 4 or Wackerly, Chapter 8)

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 $\sigma^2$. To take into account different sample sizes we will take a weighted average:

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

Note: denominator equals $\sum r_i - \sum 1 = n - v$.

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

Using the definition of $S_i^2$, see that the numerator of $S^2$ is

$$\sum \sum (y_i - \bar{y}_i)^2$$

-- called SSE, the sum of squares for error or the error sum of squares.

So

$$S^2 = \text{SSE}/(n-v)$$

-- called MSE, the mean square for error or error mean square.
The above are random variables. Their values calculated from the data are:

\[
\text{ssE} = \sum_{i=1}^{n} \sum_{r=1}^{t} (y_{ir} - \bar{y}_r)^2
\]

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

\[
\text{msE} = \frac{\text{ssE}}{(n-v)}
\]

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

\[
\hat{s}^2 = \text{msE}
\]

-- the unbiased estimate of \(\sigma^2\) -- also denoted \(\hat{\sigma}^2\).

Note:

- \(y_{ir} - \bar{y}_r\) is sometimes called the \(i^{th}\) residual, denoted \(\hat{e}_{ir}\). So \(\text{ssE} = \sum_{i=1}^{n} \sum_{r=1}^{t} \hat{e}_{ir}^2\)

- Many people use SSE and MSE for ssE and msE.

- This unbiased estimate of \(\sigma^2\) is sometimes called the within groups (or within treatments) variation, since it calculates the sample variance within each group and then averages these estimates.

- Exercise (might be homework):

\[
\text{ssE} = \sum_{i=1}^{n} \sum_{r=1}^{t} y_{ir}^2 - \sum_{i=1}^{n} \bar{y}_r^2
\]