Analysis of Variance and Design of Experiment-I

MODULE – VIII

LECTURE - 34

ANALYSIS OF VARIANCE IN RANDOM-EFFECTS MODEL AND MIXED-EFFECTS MODEL

Dr. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur
**Fixed-effects model**

All the factors in a fixed-effects model for an experiment have a predetermined set of levels, i.e., they are fixed. The statistical inferences are drawn only for those levels of the factors which are actually used in the experiment.

**Random-effects model**

The levels of factors used in experiment are randomly drawn from a population of possible levels in case of a random-effects model for an experiment. The statistical inferences are drawn from the data for all levels of the factors in the population from which the levels were selected and not only the levels used in the experiment.

For example, in case of quality control experiment about the daily production of five machines from an assembly line, we have the following set ups of fixed and random effect models:

i. Fixed-effects: The daily production of five particular machines from an assembly line.

ii. Random-effects: The daily production of five machines, chosen at random, that represent the machines as a class.
Many studies involve factors having a predetermined set of levels and factors in which the levels used in the study are randomly selected from a population of levels.

For example, the blocks in a randomized complete block design may represent a random sample of \( b \) plots of land taken from a population of plots in an agricultural research land. Then the effects due to the blocks are considered to be random effects.

Suppose the treatments are four new varieties of wheat that have been developed to be resistant to a specific bacteria. The levels of the treatment are fixed because there are only varieties of interest to the researchers, whereas the levels of the plots of land are random because the researchers are not interested in only those plots of land but are interested in the effects of these treatments on a wide range of plots of land.

When some of the factors to be used in the experiment have levels randomly selected from a population of possible levels and other factors have predetermined levels, the model used to relate the response variable to the levels of the factors is referred to as a **mixed-effects model**.
In a mixed-effects model for an experiment, the levels of some of the factors used in the experiment are randomly selected from a population of levels, whereas the levels of the other factors in the experiment are predetermined.

The inferences from the data in the experiment concerning factors with fixed levels are only for the levels of the factors used in the experiment, whereas the inferences concerning factors with randomly selected levels are for all levels of the factors in the population from which the levels were selected.
Analysis of variance in one way random-effects model

The model with random effects is of the same structure as the model with fixed effects given as

\[ y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (i = 1, 2, \ldots, s; \ h = 1, 2, \ldots, n_i). \]

Here \( \mu \) is the general mean effect as usual and \( \epsilon_{ij} \)'s are the usual random error component. The meaning of the parameter \( \alpha_i \) however has now changed. The \( \alpha_i \)'s are now the random effects of the \( i^{th} \) treatment (\( i^{th} \) machine). Hence, \( \alpha_i \)'s are the random variables whose distributions we have to specify. We assume

\[
E(\alpha_i) = 0, \\
Var(\alpha_i) = \sigma^2\alpha
\]

and

\[
E(\epsilon_{ij}\alpha_i) = 0, \\
E(\alpha_i\alpha_j) = 0 \ (i \neq j). 
\]

Then

\[ y_{ij} \sim (\mu, \sigma^2\alpha + \sigma^2) \]

holds.
In the model with fixed effects, the treatment effect $A$ was represented by the parameter estimates $\hat{\alpha}_i$, or $\hat{\mu} = \hat{\mu} + \hat{\alpha}_i$, respectively. In the model with random effects, a treatment effect can be expressed by the variance components. The variance $\sigma^2_a$ is estimated as a component of the entire variance. The absolute or relative size of this component then makes conclusions about the treatment effect possible.

The estimation of the variances $\sigma^2_a$ and $\sigma^2$ requires no assumptions about the distribution. For the test of hypothesis and the computation of confidence intervals, however, we assume the normal distribution, i.e.,

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

$$
\varepsilon_{ij}'s \text{ are assumed to be independent of each other,}
$$

$$
\alpha_i \sim N(0, \sigma^2_a),
$$

$$
\alpha_i's \text{ are assumed to be independent of each other.}
$$

and, hence,

$$
y_{ij} \sim N(\mu, \sigma^2_a + \sigma^2).
$$

Unlike, the model with effects, the response values $y_{ij}$ of a level $i$ of the treatment (i.e., of the $i^{th}$ sample) are no longer uncorrected

$$
E(y_{ij} - \mu)(y_{ij'} - \mu) = E(\alpha_i + \varepsilon_{ij})(\alpha_i + \varepsilon_{ij'}), \ j \neq j'
$$

$$
= E(\alpha_i^2) = \sigma^2_a.
$$
On the other hand, the response values of different samples are still uncorrelated \((i \neq i', \text{ for any } j, j')\):

\[
E(y_{ij} - \mu)(y_{i'j'} - \mu) = E(\alpha_i \alpha_{i'}) + E(\epsilon_{ij} \epsilon_{i'j'}) + E(\alpha_i \epsilon_{ij}) = 0.
\]

In the case of a normal distribution, uncorrelated can be replaced by independence.

### Test of the null hypothesis \(H_0 : \sigma^2 = 0\) against \(H_1 : \sigma^2 > 0\)

The hypothesis: “no treatment effect” for the two models is:

- fixed effects: \(H_0 : \alpha_i = 0\) for all \(i\).
- random effects: \(H_0 : \sigma^2 = 0\).

- If \(\sigma^2 = 0\), then the random \(\alpha_i\) effects are identically 0. In this case, each \(\hat{\alpha}_i\) estimate \((i = 1, 2, \ldots, s)\) should be close to 0 relative to the MSE.

- If \(\sigma^2 > 0\), then the random effects \(\alpha_i\) are not identically 0. In this case, the variability of the \(\hat{\alpha}_i\) estimate \((i = 1, 2, \ldots, s)\) should be large relative to the MSE.

- Testing hypotheses about the equality of means is meaningless in the random effects case. Therefore, we do not perform a multiple comparison procedure to compare the means.
The ANOVA table for a random factor is the same as the ANOVA table for a fixed factor with

\[
SS_{Total} = SS_{Treatment} + SS_{Error}.
\]

To see this we need to look at the expected mean squares for the random effects model. We can partly adopt some of the results of fixed effect model, we have for random effect model;

\[
E(MS_{Error}) = \sigma^2
\]

i.e., \( \hat{\sigma}^2 = MS_{Error} \) is an unbiased estimate of \( \sigma^2 \).

We compute \( E(MS_{Tr}) \) as follows:

\[
SS_{Tr} = \sum_{i=1}^{s} \sum_{j=1}^{n_i} (\bar{y}_{io} - \bar{y}_{oo})^2,
\]

\[
\bar{y}_{io} = \mu + \alpha_i + \epsilon_{io},
\]

\[
\bar{y}_{oo} = \mu + \alpha + \epsilon_{oo},
\]

\[
\alpha = \sum n_i \alpha_i / n,
\]

\[
(\bar{y}_{io} - \bar{y}_{oo}) = (\alpha_i - \alpha) + (\epsilon_{io} - \epsilon_{oo}).
\]
Then

\[ E(\bar{y}_{io} - \bar{y}_{oo})^2 = E(\alpha_i - \alpha)^2 + E(\bar{\varepsilon}_{io} - \bar{\varepsilon}_{oo})^2, \]

\[ E(\alpha_i - \alpha)^2 = E(\alpha_i^2) + E(\alpha^2) - 2E(\alpha_i \alpha) \]

\[ = \sigma^2_{\alpha} \left[ 1 + \sum \frac{n_i^2}{n^2} - 2 \frac{n_i}{n} \right], \]

\[ E(\bar{\varepsilon}_{io}^2 - \bar{\varepsilon}_{oo}^2)^2 = E(\varepsilon_{io}^2 + E(\varepsilon_{oo}^2) - 2E(\varepsilon_{io} \varepsilon_{oo}) \]

\[ = \frac{\sigma^2}{n_i} + \frac{\sigma^2}{n} - 2 \frac{\sigma^2}{n} \]

\[ = \sigma^2 \left( \frac{1}{n_i} - \frac{1}{n} \right). \]
Hence
\[
\sum_{j=1}^{n_i} E(\bar{y}_{io} - \bar{y}_{oo})^2 = n_i E(\bar{y}_{io} - \bar{y}_{oo})^2
\]
\[
= \sigma^2_\alpha \left[ n_i + \frac{n_i \sum n_i^2}{n^2} - 2 \frac{n_i}{n} \right] + \sigma^2 \left( 1 - \frac{n_i}{n} \right)
\]

and
\[
\sum_{i=1}^{s} n_i E(\bar{y}_{io} - \bar{y}_{oo})^2 = \sigma^2_\alpha \left[ n - \frac{\sum n_i^2}{n} \right] + \sigma^2 (s - 1).
\]

We have now
i. In the unbalanced case, i.e., all sample sizes \( n_i \)'s are not the same, we have
\[
E(MS_{T}) = \frac{1}{s-1} E(SS_A) = \sigma^2 + k\sigma^2_\alpha
\]
with
\[
k = \frac{1}{s-1} \left( n - \frac{1}{n} \sum n_i^2 \right);
\]

ii. In the balanced case, we have \((n_i = r \text{ for all } i, n = rs)\)
\[
k = \frac{1}{s-1} \left( rs - \frac{1}{rs} rs^2 \right) = r,
\]
\[
E(MS_{T}) = \sigma^2 + r\sigma^2_\alpha.
\]
This yields the unbiased estimate \( \hat{\sigma}_2^2 \) of \( \sigma_2^2 \) as follows:

(i) In the unbalanced case

\[
\hat{\sigma}_2^2 = \frac{MS_{Tr} - MS_{\text{Error}}}{k}
\]

(ii) in the balanced case

\[
\hat{\sigma}_2^2 = \frac{MS_{Tr} - MS_{\text{Error}}}{r}.
\]

In the case of an assumed normal distribution we have

\[
MS_{\text{Error}} \sim \sigma^2 \chi^2_{n-s}
\]

and

\[
MS_{Tr} \sim (\alpha^2 + k\sigma_2^2) \chi^2_{s-1}.
\]

The two distributions are independent, hence the ratio

\[
\frac{MS_{Tr}}{MS_{\text{Error}}} \cdot \frac{\sigma^2}{\sigma^2 + k\sigma_2^2}
\]

has a central \( F \)-distribution under the assumption of equal variances, i.e., under \( H_0 : \sigma_2^2 = 0 \).
Under $H_0 : \sigma^2 = 0$ we thus have

$$\frac{MS_{Tr}}{MS_{Error}} \sim F_{s-1,n-s}.$$

Hence $H_0 : \sigma^2 = 0$ is tested with the same test statistic as $H_0 : \alpha_i = 0$ (all $i$) in the model with fixed effects. The table of the analysis of variance remains unchanged. It is given as follows:

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of squares</th>
<th>Degrees of freedom</th>
<th>$E(\text{MS})$ Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fixed</td>
</tr>
<tr>
<td>Treatments</td>
<td>$SS_{Tr}$</td>
<td>$s - 1$</td>
<td>$\sigma^2 + \frac{\sum n_i \alpha^2_i}{s-1}$</td>
</tr>
<tr>
<td>Error</td>
<td>$SS_{Error}$</td>
<td>$n - s$</td>
<td>$\sigma^2 + k\sigma^2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>Random</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>
Note: The estimate of $\sigma_\alpha^2$ can be negative also ($\hat{\sigma}_\alpha^2 < 0$). But we know that a variance component cannot be negative.

The following are 3 possible ways to handle this situation:

1. Assume $\sigma_\alpha^2 = 0$ and the negative estimate occurred due to random sampling. The problem is that using zero instead of a negative number can affect the other estimates.

2. Estimate $\sigma_\alpha^2$ using the restricted maximum likelihood method because it always yields a nonnegative estimate. This method will adjust other variance component estimates.

3. Assume the model is incorrect, and examine the problem in another way. For example, add or remove an effect from the model, and then analyze the new model.