Analysis of Variance and Design of Experiments-I

MODULE – IV

LECTURE - 20

EXPERIMENTAL DESIGNS AND THEIR ANALYSIS

Dr. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur
**Analysis**

There is only one factor which is affecting the outcome – treatment effect. So the setup of one way analysis of variance is to be used.

\[ y_{ij} : \text{Individual measurement of } j^{th} \text{ experimental units for } i^{th} \text{ treatment } i = 1, 2, \ldots, v, j = 1, 2, \ldots, n_i. \]

\[ y_j : \text{Independently distributed following } N(\mu + \alpha, \sigma^2) \text{ with } \sum_{i=1}^{v} n_i \alpha = 0. \]

\[ \mu : \text{overall mean} \]

\[ \alpha : i^{th} \text{ treatment effect} \]

\[ H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_v = 0 \]

\[ H_1 : \text{All } \alpha_i \text{s are not equal.} \]

The data set is arranged as follows:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>...</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| \[ y_{11} \ | y_{21} \ | \ldots \ | y_{v1} \]
| \[ y_{12} \ | y_{22} \ | \ldots \ | y_{v2} \]
| \vdots  | \vdots | \ddots | \vdots |
| \[ y_{1n_1} \ | y_{2n_2} \ | \ldots \ | y_{vn_v} \]
| \[ T_1 \ | T_2 \ | \ldots \ | T_v \]

where \( T_i = \sum_{j=1}^{n_i} y_{ij} \) is the treatment total due to \( i^{th} \) effect, \( G = \sum_{i=1}^{v} T_i = \sum_{i=1}^{v} \sum_{j=1}^{n_i} y_{ij} \) is the grand total of all the observations.
In order to derive the test for $H_0$, we can use either the likelihood ratio test or the principle of least squares. Since the likelihood ratio test has already been derived earlier, so we choose to demonstrate the use of least squares principle.

The linear model under consideration is

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, ..., v, j = 1, 2, ..., n_i$$

where $\epsilon_{ij}$'s are identically and independently distributed random errors with mean 0 and variance $\sigma^2$. The normality assumption of $\epsilon$'s is not needed for the estimation of parameters but will be needed for deriving the distribution of various involved statistics and in deriving the test statistics.

Let

$$S = \sum_{i=1}^{v} \sum_{j=1}^{n_i} \epsilon_{ij}^2 = \sum_{i=1}^{v} \sum_{j=1}^{n_i} (y_{ij} - \mu - \alpha_i)^2.$$

Minimizing $S$ with respect to $\mu$ and $\alpha_i$, the normal equations are obtained as

$$\frac{\partial S}{\partial \mu} = 0 \Rightarrow n\mu + \sum_{i=1}^{v} n_i \alpha_i = 0$$

$$\frac{\partial S}{\partial \alpha_i} = 0 \Rightarrow n_i \mu + n_i \alpha_i = \sum_{j=1}^{n_i} y_{ij}, \quad i = 1, 2, ..., v.$$
Solving them using \( \sum_{i=1}^{v} n_i \alpha_i = 0 \), we get
\[
\hat{\mu} = \bar{y}_{oo},
\]
\[
\hat{\alpha}_i = \bar{y}_{io} - \bar{y}_{oo}
\]
where \( \bar{y}_{io} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \) is the mean of observation receiving the \( i \)th treatment and \( \bar{y}_{oo} = \frac{1}{n} \sum_{i=1}^{v} \sum_{j=1}^{n_i} y_{ij} \) is the mean of all the observations.

The fitted model is obtained after substituting the estimate \( \hat{\mu} \) and \( \hat{\alpha}_i \) in the linear model, we get
\[
y_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\epsilon}_{ij}
\]

or \( y_{ij} = \bar{y}_{oo} + (\bar{y}_{io} - \bar{y}_{oo}) + (y_{ij} - \bar{y}_{io}) \)

or \( (y_{ij} - \bar{y}_{oo}) = (\bar{y}_{io} - \bar{y}_{oo}) + (y_{ij} - \bar{y}) \).

Squaring both sides and summing over all the observation, we have
\[
\sum_{i=1}^{v} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{oo})^2 = \sum_{i=1}^{v} n_i (\bar{y}_{io} - \bar{y}_{oo})^2 + \sum_{i=1}^{v} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{oo})^2
\]

\[
\downarrow \quad \downarrow \quad \downarrow
\]

or \( \text{Total sum of squares} = \text{Sum of squares due to treatment effects} + \text{Sum of squares due to error} \)

or \( TSS = SSTr + SSE \)
- Since \( \sum_{i=1}^{v} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{io}) = 0 \), so TSS is based on the sum of \((n-1)\) squared quantities. Thus TSS carries only \((n-1)\) degrees of freedom.

- Since \( \sum_{i=1}^{v} n_i (\bar{y}_{io} - \bar{y}_{oo}) = 0 \), so SStr is based only on the sum of \((v-1)\) squared quantities. Thus SStr carries only \((v-1)\) degrees of freedom.

- Since \( \sum_{i=1}^{n} n_i (\bar{y}_i - \bar{y}_{io}) = 0 \) for all \( i = 1, 2, ..., v \), so SSE is based on the sum of squaring \( n \) quantities like \((y_{ij} - \bar{y}_{io})\) with \( v \) constraints \( \sum_{j=1}^{n} (y_{ij} - \bar{y}_{io}) = 0 \). So SSE carries \((n-v)\) degrees of freedom.

- Using the Fisher-Cochran theorem,

\[
TSS = SStr + SSE
\]

with degrees of freedom partitioned as

\[
(n - 1) = (v - 1) + (n - v).
\]
Moreover, the equality in $TSS = SSTr + SSE$ has to hold exactly. In order to ensure that the equality holds exactly, we find one of the sum of squares through subtraction. Generally, it is recommended to find $SSE$ by subtraction as

$$SSE = TSS - SSTr$$

where

$$TSS = \sum_{i=1}^{v} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{io})^2$$

$$= \sum_{i=1}^{v} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{G^2}{n}$$

$$G = \sum_{i=1}^{v} \sum_{j=1}^{n_i} y_{ij}$$

$$\frac{G^2}{n} : \text{correction factor}$$

$$SSTr = \sum_{j=1}^{n_0} n_j (\bar{y}_{io} - \bar{y}_{oo})^2$$

$$= \sum_{i=1}^{v} \left( \frac{T_i^2}{n_i} \right) - \frac{G^2}{n}$$

$$T_i = \sum_{j=1}^{n_i} y_{ij},$$
Now under $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_v = 0$, the model become

$$Y_{ij} = \mu + \varepsilon_{ij},$$

and minimizing $S = \sum_{i=1}^{v} \sum_{j=1}^{n_i} \varepsilon_{ij}^2$

with respect to $\mu$ gives

$$\frac{\partial S}{\partial \mu} = 0 \Rightarrow \hat{\mu} = \frac{G}{n} = \bar{y}_{oo}.$$

The SSE under $H_0$ becomes

$$\text{SSE} = \sum_{i=1}^{v} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{oo})^2$$

and thus $\text{TSS} = \text{SSE}$.

This TSS under $H_0$ contains the variation only due to the random error whereas the earlier $\text{TSS} = \text{SSTr} + \text{SSE}$ contains the variation due to treatments and errors both. The difference between the two will provides the effect of treatments in terms of sum of squares as

$$\text{SSTr} = \sum_{i=1}^{v} n_i (\bar{y}_i - \bar{y}_{oo})^2.$$
Expectations

\[ E(SSE) = \sum_{i=1}^{v} \sum_{j=1}^{n_i} E((y_{ij} - y_{io})^2) \]
\[ = \sum_{i=1}^{v} n_i \sum_{j=1}^{n_i} E(\varepsilon_{ij} - \overline{\varepsilon}_{io})^2 \]
\[ = \sum_{i=1}^{v} \sum_{j=1}^{n_i} E(\varepsilon_{ij}^2) - \sum_{i=1}^{v} n_i E(\overline{\varepsilon}_{io}^2) \]
\[ = n\sigma^2 - \sum_{i=1}^{v} n_i \frac{\sigma^2}{n_i} \]
\[ = (n - v)\sigma^2 \]

\[ E(MSE) = E\left(\frac{SSE}{n - v}\right) = \sigma^2 \]

\[ E(SSTr) = \sum_{i=1}^{v} n_i E(\overline{y}_{io} - \overline{y}_{oo})^2 \]
\[ = \sum_{i=1}^{v} n_i E(\alpha_i + \overline{\varepsilon}_{io} - \overline{\varepsilon}_{oo})^2 \]
\[ = \sum_{i=1}^{v} n_i \alpha_i^2 + \left[ \sum_{i=1}^{v} n_i \overline{\varepsilon}_{io}^2 - n \overline{\varepsilon}_{oo}^2 \right] \]
\[ = \sum_{i=1}^{v} n_i \alpha_i^2 + \left[ \sum_{i=1}^{v} n_i \frac{\sigma^2}{n_i} - n \frac{\sigma^2}{n} \right] \]
\[ = \sum_{i=1}^{v} n_i \alpha_i^2 + (v - 1)\sigma^2 \]

\[ E(MSTr) = E\left(\frac{SSTr}{v - 1}\right) \]
\[ = \frac{1}{v - 1} \sum_{i=1}^{v} n_i \alpha_i^2 + \sigma^2. \]

In general \( E(MSTr) \neq \sigma^2 \)

but under \( H_0 \), all \( \alpha_i = 0 \) and so
\[ E(MSTr) = \sigma^2 \]
Distributions and decision rules

Using the normal distribution property of $\epsilon_{ij}$'s, we find that $y_{ij}$'s are also normal as they are the linear combination of $\epsilon_{ij}$'s.

$$- \frac{SSTr}{\sigma^2} \sim \chi^2(v-1) \quad \text{under } H_0$$

$$- \frac{SSE}{\sigma^2} \sim \chi^2(n-v) \quad \text{under } H_0$$

$- SSTr$ and $SSE$ are independently distributed

$$- \frac{MStr}{MSE} \sim F(v-1, n-v) \quad \text{under } H_0.$$

So the decision rule is to

reject $H_0$ at $\alpha^*$ level of significance if $F > F_{\alpha^*, v-1, n-v}$.

[Note: We denote the level of significance here by $\alpha^*$ because has been used for denoting the factor]
The analysis of variance table in this case is given as following:

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>$F$ - value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between treatments</td>
<td>$v - 1$</td>
<td>$SSTr$</td>
<td>$MSTr$</td>
<td>$MSTr / MSE$</td>
</tr>
<tr>
<td>Error</td>
<td>$(n - v)$</td>
<td>$SSE$</td>
<td>$MSE$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$n - 1$</td>
<td>$TSS$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>