Analysis of Variance and Design of Experiments-I

MODULE – III

LECTURE - 18

EXPERIMENTAL DESIGN MODELS

Dr. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur
Note that if $\lambda = 0$, then

$$E\left[ \hat{\lambda} / \hat{\alpha}_i, \hat{\beta}_j \text{ for all } i, j \right] = E \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} \alpha_i \beta_j y_{ij} \right] = E \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} \alpha_i \beta_j (\mu + \alpha_i + \beta_j + 0 + \varepsilon_{ij}) \right]$$

$$= E \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} \alpha_i \beta_j \frac{(\sum_{i=1}^{I} \alpha_i^2)(\sum_{j=1}^{J} \beta_j^2)}{\sum_{i=1}^{I} \sum_{j=1}^{J} \alpha_i \beta_j} \right]$$

$$= 0 \frac{0}{\sum_{i=1}^{I} \sum_{j=1}^{J} \alpha_i^2 \beta_j^2}$$

$$= 0.$$ 

As $\alpha_i$ and $\beta_j$ remains valid irrespective of $\lambda = 0$, or not, in this sense $\hat{\lambda}$ is a function of $y_{ij}$ and hence normally distributed as

$$\hat{\lambda} \sim N\left(0, \frac{IJ \sigma^2}{S_A S_B}\right).$$
Thus the statistic

\[
\frac{(\hat{\lambda})^2}{\text{Var}(\hat{\lambda})} = \frac{IJ \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} \hat{\alpha}_i \hat{\beta}_j y_{ij} \right]^2}{\sigma^2 S_A S_B}
\]

\[
= \frac{IJ \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} (\bar{y}_{io} - \bar{y}_{oo})(\bar{y}_{oj} - \bar{y}_{oo})y_{ij} \right]^2}{\sigma^2 S_A S_B}
\]

\[
= \frac{IJ \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} (\bar{y}_{io} - \bar{y}_{oo})(\bar{y}_{oj} - \bar{y}_{oo})(y_{ij} - \bar{y}_{io} - \bar{y}_{oj} + \bar{y}_{oo}) \right]^2}{\sigma^2 S_A S_B}
\]

\[
= \frac{S_N}{\sigma^2}
\]

follows a $\chi^2$ - distribution with one degree of freedom where

\[
S_N = \frac{IJ \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} (\bar{y}_{io} - \bar{y}_{oo})(\bar{y}_{oj} - \bar{y}_{oo})(y_{ij} - \bar{y}_{io} - \bar{y}_{oj} + \bar{y}_{oo}) \right]^2}{S_A S_B}
\]

is the sum of squares due to non-additivity.
Note that
\[
\frac{S_{AB}}{\sigma^2} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2}{\sigma^2}
\]
follows \(\chi^2((I-1)(J-1))\),

so \(\left(\frac{S_N}{\sigma^2} - \frac{S_{AB}}{\sigma^2}\right)\) is nonnegative and follows \(\chi^2[(I-1)(J-1)-1]\).

The reason for this is as follows:

\[
y_{ij} = \mu + \alpha_i + \beta_j + \text{non-additivity} + \epsilon_{ij}
\]

and so

\[
TSS = SSA + SSB + S_N + SSE
\]

\[
\Rightarrow SSE = TSS - SSA - SSB - S_N
\]

has degrees of freedom

\[
= (IJ - 1) - (I - 1) - (J - 1) - 1
\]

\[
= (I - 1)(J - 1) - 1.
\]
We need to ensure that \( SSE > 0 \).

So using the result

“If \( Q, Q_1 \) and \( Q_2 \) are quadratic forms such that \( Q = Q_1 + Q_2 \) with \( Q \sim \chi^2(a) \), \( Q_2 \sim \chi^2(b) \) and \( Q_2 \) is non-negative, then \( Q_1 \sim \chi^2(a-b) \),”

ensures that the difference

\[
\frac{S_N}{\sigma^2} - \frac{S_{AB}}{\sigma^2}
\]

is nonnegative.

Moreover \( S_N \ (SS \ due \ to \ nonadditivity) \) and \( SSE \) are orthogonal.

Thus the \( F \)–test for nonadditivity is

\[
F = \frac{\left( \frac{S_N}{\sigma^2} - \frac{1}{1} \right)}{\frac{SSE / \sigma^2}{(I-1)(J-1)-1}}
\]

\[
= \frac{[(I-1)(J-1)-1]}{SSN \ SSE}
\]

\[
\sim F[1,(I-1)(J-1)-1] \quad \text{under } H_0.
\]

So the decision rule is

Reject \( H_0 : \lambda = 0 \) whenever

\[
F > F_{1-\alpha}[1,(I-1)(J-1)-1].
\]
The analysis of variance table for the model including a term for nonadditivity is as follows:

<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>Factor A</td>
<td>(( I - 1))</td>
<td>(S_A)</td>
<td>(MS_A = \frac{S_A}{I - 1})</td>
<td></td>
</tr>
<tr>
<td>Factor B</td>
<td>((J - 1))</td>
<td>(S_B)</td>
<td>(MS_B = \frac{S_B}{J - 1})</td>
<td></td>
</tr>
<tr>
<td>Non-additivity</td>
<td>1</td>
<td>(S_N)</td>
<td>(MS_N = S_N)</td>
<td>(\frac{MS_N}{MSE})</td>
</tr>
<tr>
<td>Error</td>
<td>((I - 1)(J - 1) - 1)</td>
<td>(SSE) (By subtraction)</td>
<td>(MSE = \frac{SSE}{(I - 1)(J - 1) - 1})</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>((IJ - 1))</td>
<td>TSS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Comparison of variances

One of the basic assumptions in the analysis of variance is that the samples are drawn from different normal populations with different means but same variances. So before going for the analysis of variance, the test of hypothesis about the equality of variance is needed to be done.

We discuss the test of equality of two variances and more than two variances.

Case 1: Equality of two variances

\( H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2. \)

Suppose there are two independent random samples

\[ A : x_1, x_2, \ldots, x_{n_1} \sim N(\mu_A, \sigma_A^2) \]
\[ B : y_1, y_2, \ldots, y_{n_2} \sim N(\mu_B, \sigma_B^2). \]

The sample variance corresponding to the two samples are

\[ s^2_x = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2 \]
\[ s^2_y = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2. \]

Under \( H_0 : \sigma_A^2 = \sigma_B^2 = \sigma^2, \)

\[ \frac{(n_1 - 1)s^2_x}{\sigma^2} \sim \chi^2(n_1 - 1) \]
\[ \frac{(n_2 - 1)s^2_y}{\sigma^2} \sim \chi^2(n_2 - 1). \]
Moreover, the sample variances $s_x^2$ and $s_y^2$ are independent. So

$$
\left( \begin{array}{c}
\frac{(n_1 - 1)s_x^2}{\sigma^2} \\
n_1 - 1
\end{array} \right)
\sim F_{n_1 - 1, n_2 - 1},
\quad
\left( \begin{array}{c}
\frac{(n_2 - 1)s_y^2}{\sigma^2} \\
n_2 - 1
\end{array} \right)
$$

So for testing $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 \neq \sigma_2^2$, the null hypothesis $H_0$ is rejected if

$$F > F_{1 - \frac{\alpha}{2}, n_1 - 1, n_2 - 1}$$

or

$$F < F_{\frac{\alpha}{2}, n_1 - 1, n_2 - 1}$$

where

$$F_{\frac{\alpha}{2}, n_1 - 1, n_2 - 1} = \frac{1}{F_{\frac{\alpha}{2}, n_2 - 1, n_1 - 1}}.$$

If the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ is rejected, then the problem is termed as Fisher-Behran’s problem. The solution are available for this problem.
**Case 2: Equality of more than two variances: Bartlett’s test**

\[ H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 \quad \text{and} \quad H_1 : \sigma_i^2 \neq \sigma_j^2 \quad \text{for atleast one} \quad i \neq j = 1, 2, \ldots, k. \]

Let there be \( k \) independent normal population \( N(\mu_i, \sigma_i^2) \) each of size \( n_i, i = 1, 2, \ldots, k. \) Let \( s_1^2, s_2^2, \ldots, s_k^2 \) be \( k \) independent unbiased estimators of population variances \( \sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2 \) respectively with \( \nu_1, \nu_2, \ldots, \nu_k \) degrees of freedom. Under \( H_0, \) all the variances are same as \( \sigma^2, \) say and an unbiased estimate of \( \sigma^2 \) is

\[
s^2 = \sum_{i=1}^{k} \frac{V_i s_i^2}{\nu} \quad \text{where} \quad \nu_i = n_i - 1, \nu = \sum_{i=1}^{k} \nu_i.
\]

Bartlett has shown that under \( H_0, \)

\[
\sum_{i=1}^{k} \left( \nu_i \ln \frac{s^2}{s_i^2} \right) \left[ 1 + \frac{1}{3(k-1)} \left( \sum_{i=1}^{k} \left( \frac{1}{\nu_i} - \frac{1}{\nu} \right) \right) \right]
\]

is distributed as \( \chi^2(k-1) \) based on which \( H_0 \) can be tested.