Analysis of Variance and Design of Experiments-II

MODULE I

LECTURE - 8

INCOMPLETE BLOCK DESIGNS

Dr. Shalabh
Department of Mathematics & Statistics
Indian Institute of Technology Kanpur

Generally, we are not interested merely in the interblock analysis of variance but we want to utilize the information from interblock analysis along with the intrablock information to improve upon the statistical inferences.

After obtaining the interblock estimate of treatment effects, the next question that arises is how to use this information for an improved estimation of treatment effects and use it further for the testing of significance of treatment effects. Such an estimate will be based on the use of more information, so it is expected to provide better statistical inferences.

We now have two different estimates of the treatment effect as

- based on intrablock analysis $\hat{\tau} = C^{-}Q$ and
- based on interblock analysis $\tilde{\tau} = (N'N)^{-1}N'B - \frac{GE_{v1}}{bk}$.

Let us consider the estimation of linear contrast of treatment effects $L = l'\tau$. Since the intrablock and interblock estimates of τ are based on Gauss-Markov model and least squares principle, so the best estimate of L based on intrablock estimation is

$$\begin{aligned} L_1 &= l'\hat{\tau} \\ &= l'C^{-}Q \end{aligned}$$

and the best estimate of L based on interblock estimation is

$$\begin{aligned} L_2 &= l'\tilde{\tau} \\ &= l'\left[(N'N)^{-1}N'B - \frac{GE_{v1}}{bk}\right] \\ &= l'(N'N)^{-1}N'B \quad \text{(since } l'E_{v1} = 0 \text{ being contrast.)} \end{aligned}$$

The variances of L_1 and L_2 are

$$Var(L_1) = \sigma^2 l' C^- l$$

and

$$Var(L_2) = \sigma_f^2 l'(N'N)^{-1} l,$$

respectively. The covariance between Q (from intrablock) and B (from interblock) is

$$\begin{aligned} Cov(Q,B) &= Cov(V - N'K^{-1}B^*, B) \\ &= Cov(V,B) - Cov(N'K^{-1}B^*, B) \\ &= N'\sigma_f^2 - N'K^{-1}K\sigma_f^2 \\ &= 0. \end{aligned}$$

Note that B^* denotes the block total based on intrablock analysis and B denotes the block totals based on interblock analysis . We are using two notations B and B^* just to indicate that the two block totals are different. The reader should not misunderstand that it follows from the result of $Cov(Q,B) = 0$ in case of intrablock analysis.

Thus

$$Cov(L_1, L_2) = 0$$

irrespective of the values of l.

The question now arises that given the two estimators $\hat{\tau}$ and $\tilde{\tau}$ of τ , how to combine them and obtain a minimum variance unbiased estimator of τ. It is illustrated with following example:

Example

Let $\hat{\varphi}_1$ and $\hat{\varphi}_2$ be any two unbiased estimators of a parameter φ with $Var(\hat{\varphi}_1) = \sigma_1^2$ and $Var(\hat{\varphi}_2) = \sigma_2^2$. Consider a linear combination $\hat{\varphi} = \theta_1\hat{\varphi}_1 + \theta_2\hat{\varphi}_2$ with weights θ_1 and θ_2. In order that $\hat{\varphi}$ is an unbiased estimator of θ_1, we need

$$E(\hat{\varphi}) = \varphi$$

or $\quad \theta_1 E(\hat{\varphi}_1) + \theta_2 E(\hat{\varphi}_2) = \varphi$

or $\quad \theta_1\varphi + \theta_2\varphi = \varphi$

or $\quad \theta_1 + \theta_2 = 1.$

So modify $\hat{\varphi}$ as $\dfrac{\theta_1\hat{\varphi}_1 + \theta_2\hat{\varphi}_2}{\theta_1 + \theta_2}$ which is the weighted mean of $\hat{\varphi}_1$ and $\hat{\varphi}_2$.

Further, if $\hat{\varphi}_1$ and $\hat{\varphi}_2$ are independent , then $Var(\hat{\varphi}) = \theta_1^2\sigma_1^2 + \theta_2^2\sigma_2^2$.

Now we find θ_1 and θ_2 such that $Var(\hat{\varphi})$ is minimum such that $\theta_1 + \theta_2 = 1$.

$$\frac{\partial Var(\hat{\varphi})}{\partial \theta_1} = 0 \Rightarrow 2\theta_1\sigma_1^2 - 2(1-\theta_1)\sigma_2^2 = 0$$

or $\quad \theta_1\sigma_1^2 - \theta_2\sigma_2^2 = 0$

or $\quad \dfrac{\theta_1}{\theta_2} = \dfrac{\sigma_2^2}{\sigma_1^2}$

or $\quad$ weight $\propto \dfrac{1}{\text{variance}}$.

Alternatively, the Lagrangian function approach can be used to obtain such result as follows. The Lagrangian function with λ^* as Lagrangian multiplier is given by $\phi = Var(\hat{\varphi}) - \lambda^*(\theta_1 + \theta_2 - 1)$.

Solving $\dfrac{\partial \phi}{\partial \theta_1} = 0, \dfrac{\partial \phi}{\partial \theta_2}$ and $\dfrac{\partial \phi}{\partial \lambda^*} = 0$ also gives the same result that $\dfrac{\theta_1}{\theta_2} = \dfrac{\sigma_2^2}{\sigma_1^2}$.

We note that a pooled estimator of τ in the form of weighted arithmetic mean of uncorrelated L_1 and L_2 is the minimum variance unbiased estimator of τ when the weights θ_1 and θ_2 of L_1 and L_2, respectively are chosen such that

$$\frac{\theta_1}{\theta_2} = \frac{Var(L_2)}{Var(L_1)},$$

i.e., the chosen weights are reciprocal to the variance of respective estimators, irrespective of the values of l. So consider the weighted average of L_1 and L_2 with weights θ_1 and θ_2, respectively as

$$\begin{aligned}\tau^* &= \frac{\theta_1 L_1 + \theta_2 L_2}{\theta_1 + \theta_2} \\ &= \frac{l'(\theta_1 \hat{\tau} + \theta_2 \tilde{\tau})}{\theta_1 + \theta_2}\end{aligned}$$

with

$$\begin{aligned}\theta_1^{-1} &= l' C^{-1} l \sigma^2 \\ \theta_2^{-1} &= l'(N'N)^{-1} l \sigma_f^2.\end{aligned}$$

The linear contrast of τ^* is $L^* = l'\tau^*$

and its variance is

$$\begin{aligned}Var(L^*) &= \frac{\theta_1^2 Var(L_1) + \theta_2^2 Var(L_2)}{(\theta_1 + \theta_2)^2} l'l \quad (\text{since } Cov(L_1, L_2) = 0) \\ &= \frac{l'l}{(\theta_1 + \theta_2)}\end{aligned}$$

because the weights of estimators are chosen to be inversely proportional to the variance of the respective estimators.

We note that τ^* can be obtained provided θ_1 and θ_2 are known. But θ_1 and θ_2 are known only when σ^2 and σ_β^2 are known. So τ^* can be obtained if σ^2 and σ_β^2 are known. In case, if σ^2 and σ_β^2 are unknown, then their estimates can be used.

A question arises how to obtain such estimators?

One such approach to obtain the estimates of σ^2 and σ_β^2 is based on utilizing the results from intrablock and interblock analysis both and is as follows:

From intrablock analysis, we have

$$E(SS_{Error(t)}) = (n-b-v+1)\sigma^2,$$

so an unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{SS_{Error(t)}}{n-b-v+1}.$$

An unbiased estimator of σ_β^2 is obtained by using the following results based on the intrablock analysis:

$$SS_{Treat(unadj)} = \sum_{j=1}^{v} \frac{V_j^2}{\tau_j} - \frac{G^2}{n},$$

$$SS_{Block(unadj)} = \sum_{i=1}^{b} \frac{B_i^2}{k_i} - \frac{G^2}{n},$$

$$SS_{Treat(adj)} = \sum_{j=1}^{v} Q_j \hat{\tau}_j,$$

$$SS_{Total} = \sum_{i=1}^{b}\sum_{j=1}^{v} y_{ij}^2 - \frac{G^2}{n},$$

where

$$SS_{Total} = SS_{Treat(adj)} + SS_{Block(unadj)} + SS_{Error(t)}$$
$$= SS_{Treat(unadj)} + SS_{Block(adj)} + SS_{Error(t)}.$$

Hence

$$SS_{Block(adj)} = SS_{Treat(adj)} + SS_{Block(unadj)} - SS_{Treat(unadj)}.$$

Under the interblock analysis model

$$E[SS_{Block(adj)}] = E[SS_{Treat(adj)}] + E[SS_{Block(unadj)}] - E[SS_{Treat(unadj)}]$$

which is obtained as follows:

$$E[SS_{Block(adj)}] = (b-1)\sigma^2 + (n-v)\sigma_\beta^2$$

or

$$E\left[SS_{Block(adj)} - \frac{b-1}{n-b-v+1} SS_{Error(t)}\right] = (n-v)\sigma_\beta^2.$$

Thus an unbiased estimator of σ_β^2 is

$$\hat{\sigma}_\beta^2 = \frac{1}{n-v}\left[SS_{Block(adj)} - \frac{b-1}{n-b-v+1} SS_{Error(t)}\right].$$

Now the estimates of weights θ_1 and θ_2 can be obtained by replacing σ^2 and σ_β^2 by $\hat{\sigma}^2$ and $\hat{\sigma}_\beta^2$, respectively. Then the estimate of $\tau*$ can be obtained by replacing θ_1 and θ_2 by their estimates and can be used in place of $\tau*$. It may be noted that the exact distribution of associated sum of squares due to treatments is difficult to find when σ^2 and σ_β^2 are replaced by $\hat{\sigma}^2$ and $\hat{\sigma}_\beta^2$, respectively in $\tau*$. Some approximate results are possible which we will present while dealing with the balanced incomplete block design . An increase in the precision using interblock analysis as compared to intrablock analysis is measured by

$$\frac{1/\text{variance of pooled estimate}}{1/\text{variance of intrablock estimate}}-1.$$

In the interblock analysis, the block effects are treated as random variable which is appropriate if the blocks can be regarded as a random sample from a large population of blocks. The best estimate of the treatment effect from the intrablock analysis is further improved by utilizing the information on block totals. Since the treatments in different blocks are not all the same, so the difference between block totals is expected to provide some information about the differences between the treatments. So the interblock estimates are obtained and pooled with intrablock estimates to obtain the combined estimate of τ. The procedure of obtaining the interblock estimates and then obtaining the pooled estimates is called the **recovery of interblock information.**