6·12. SPLIT PLOT DESIGN

Suppose we have a factor A which has p-levels, say, $a_1, a_2, \ldots, a_p$. These p-levels of A are allocated at random to the p plots of, each of say, r blocks. [This is same as an RBD with p treatments in r blocks.] It may sometimes be desirable to compare several levels of another factor B, on each of these plots. Suppose the factor B has q levels, say, $b_1, b_2, \ldots, b_q$. We now divide each plot within a block into q sub-plots and allocate the q levels of B at random in these sub-plots within each block. This scheme enables us to compare different levels of B, along with A, at a little extra cost. Obviously, this type of design is possible only if the levels of factor A require bigger plots whereas levels of B require smaller plots. The bigger plots are called *whole-plots* and the smaller plots are called *sub-plots*. The treatments under factor A allotted to whole-plots are called *whole-plot (W.P.) treatments* and the treatments under factor B allotted to the sub-plots are called *sub-plot (S.P.) treatments*.

Such a layout in which one set of treatments called the whole-plot treatments is assigned or allotted at random to large parts called the main plots (or whole-plots), and the second set of treatments called the 'sub-plot treatments' is allocated to the sub-divisions of the main plot called sub-plots is termed as 'Split Plot Design'.

We come across a large number of situations in practice, specifically in agricultural and industrial experiments, when the very nature of the level of one factor necessitates the use of bigger plots (units). We give below some situations, where Split-Plot Design can be suitably adopted :

(*i*) In research on milking machines, large amount of milk is required. Methods of cooling or pasturising require smaller amounts of milk and may be used as split-plot treatments.

(*ii*) In agricultural experiments, it is usually found that certain treatments like irrigation, dates of sowing, varieties of crops, etc., require large plots and may be regarded as whole-plot treatments. On the other hard, the treatments like the seeds, the methods of ploughing or planting, manurial applications, etc., require smaller plots and may be regarded as sub-plot treatments.

(*iii*) In alloy preparation or smelting, large quantity of material is required by the machine, while for moulding, we need relatively much smaller quantity.

(*iv*) In greenhouse temperature studies, the entire greenhouse which must be maintained at a constant temperature, is used as a main plot and several treatments that may be conducted in a green house are used as sub-plots.

(*v*) In management studies, the store (or farm) may be used as main-plot treatments and the methods of displaying (or producing) the commodity may be used as sub-plot treatments.

***Remarks* 1. Split-Plot-Design vs R.B.D. (R.B.D. with *rpq* Units).** In a split-plot design, the various levels of A (whole-plot treatments) are allocated at random within the whole-plots of a block and the levels of factor B (sub-plot treatments) are allocated at random among the sub-plots of the main-plot in each block. However, in R.B.D. all the $p \times q$ treatment combinations of two factors are allocated at random to the $p \times q$ plots in each block.

The average experimental error over all the treatment comparisons is same for the complete randomised block design as well as split-plot design. Therefore, there is no net gain in precision resulting from the use of split-plot design. However, in a split-plot design, there is an increased precision on :

(*i*) the main effect (B) of the sub-plot treatments, and

(*ii*) the interaction of W-P treatments and sub-plot treatments, *i.e.*, $(A \times B)$, than the main effects of the whole-plot treatments (A).

The increased precision on B and AB is obtained as the cost of the sacrifice of the precision on the main effect A, which is confounded with incomplete differences.

For the test of significance or the construction of confidence limits, the R.B.D. holds a slight advantage on the average because it provides more degrees of freedom for the estimate of single error variance.

2. Since the Split-Plot Design (SPD) is used in a number of fields other than agriculture, where the world 'plot' is not used *e.g.,* in industrial experiments, the SPD is also called a ***nested design.***

Statistical Analysis. Suppose there are r blocks (replications), each block divided into p whole-plots and each whole-plot is divided into q sub-plots. Within each block, the 'p' levels of the factor A are allocated at random to the p whole plots and within each whole plot, the q levels of the factor B are allocated at random to the q sub-plots. The mathematical model is :

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + e_{ij} + \delta_{jk} + e_{ijk} \; ; (i = 1, 2, ..., r \;\; ; j = 1, 2, ..., p; k = 1, 2, ..., q) \quad ...(6·246)$$

where y_{ijk} denotes the yield from the k-th sub-plot in the jth whole-plot of the ith block (replication).

μ is the general mean effect ;

α_i is the additional effect due to the ith replicate ;

β_j is the additional effect due to the jth whole-plot treatment;

γ_k is the additional effect due to the kth sub-plot treatment;

e_{ij} is the random error component associated with ith replication and jth whole-plot treatment ;

and δ_{jk} is the interaction effect between the jth whole plot treatment and kth sub-plot treatment.

Here, α_i, β_j, γ_k and δ_{jk} are fixed effects so that

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = \underset{(\text{for all } k)}{\sum_j \delta_{jk}} = \underset{(\text{for all } j)}{\sum_k \delta_{jk}} = 0 \qquad \ldots(6\cdot 246a)$$

and $$e_{ij} \overset{i.i.d}{\sim} N(0, \sigma_w^2) \text{ and } e_{ijk} \overset{i.i.d}{\sim} N(0, \sigma_s^2) \qquad \ldots(6\cdot 246b)$$

distributed independently of each other, the pooled error variance being $\sigma^2 = \sigma_w^2 + q\,\sigma_s^2$.

Remark. In the model (6·246), the sub-plots in the same main-plot are correlated *i.e.,*

$$E(e_{ijk}) = 0 \quad ; \quad E\left(e_{ijk}^2\right) = \sigma^2 \;\; ; \; E(e_{ijk}, e_{ijk'}) = \rho\,\sigma_s^2 \; ; \quad k \neq k'$$
$$= 0, \text{ otherwise}\} \qquad \ldots(6\cdot 246c)$$

i.e., the yields within the same whole-plot are correlated but within two different whole-plots are uncorrelated.

The least square estimate for various effects are obtained on minimising the error sum of squares.

$$E = \sum_i \sum_j \sum_k e_{ijk}^2 = \sum_i \sum_j \sum_k (y_{ij} - \mu - \alpha_i - \beta_j - e_{ij} - \gamma_k - \delta_{jk})^2 \text{ and are given by :}$$

$$\left.\begin{aligned} &\hat{\mu} = \bar{y}_{...} \quad ; \quad \hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...} \quad ; \quad \hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...} \\ &\hat{\gamma} = \bar{y}_{..k} - \bar{y}_{...} \; ; \; \hat{\delta}_{jk} = \bar{y}_{jk\cdot} - \bar{y}_{..k} - \bar{y}_{.j.} + \bar{y}_{...} \; ; \; \hat{e}_{ij} = \bar{y}_{ij\cdot} - \bar{y}_{i..} - \bar{y}_{j.} + \hat{e}_{ij} = \bar{y}_{...} \end{aligned}\right\} \ldots(6\cdot 247)$$

Substituting in (6·246), transposing, squaring both sides and summary over i, j & k, we get

$$\sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{...})^2 = pq \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 + rq \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 + rp \sum_k (\bar{y}_{..k} - \bar{y}_{...})^2$$
$$+ q \sum_i \sum_j (\bar{y}_{i..} - \bar{y}_{ij.} - \bar{y}_{.j.} + \bar{y}_{...})^2 + r \sum_j \sum_k (\bar{y}_{.jk} - \bar{y}_{..k} - \bar{y}_{.j.} + \bar{y}_{...})^2$$
$$+ \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.} - \bar{y}_{..k} + \bar{y}_{.jk})^2, \qquad \ldots(6\cdot 248)$$

(all the product terms vanish because algebraic sum of deviations from mean is zero.

$\Rightarrow$ T.S.S. = S.S.R. + S.S. (W × P) + S.S. (S × P) + S.S.E$_1$ + S.S. (A × B) + S.S.E$_2$ ...(6·249)

where T.S.S. $= \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{...})^2$, is the total sum of squares ;

S.S.R. $= pq \sum_{i=1}^{r} (\bar{y}_{i..} - \bar{y}_{...})^2$, is the S.S. due to replications (blocks) ,

$$\text{S.S. (WP)} = \text{SS(A)} = rq \sum_{j=1}^{p} (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2, \text{ is the S.S. due to whole plot treatments,}$$

$$\text{S.S. }(E_1) = q \sum_i \sum_j (\bar{y}_{ij\cdot} - \bar{y}_{i\cdots} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots})^2, \text{ is the error S.S. due to whole-plots ,}$$

$$\text{S.S. (SP)} = \text{S.S. (B)} = rp \sum_{k=1}^{q} (\bar{y}_{\cdot\cdot k} - \bar{y}_{\cdots})^2, \text{ is the S.S. due to sub-plots.}$$

$$\text{S.S. } (A \times B) = r \sum_j \sum_k (\bar{y}_{\cdot jk} - \bar{y}_{\cdot\cdot k} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots})^2, \text{ is the S.S. due to interaction effect } A \times B.$$

$$\text{S.S.}E_2 = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij\cdot} - \bar{y}_{\cdot jk} + \bar{y}_{\cdot j\cdot})^2, \text{ is the error S.S. due to sub-plots.}$$

Degrees of Freedom. The break-up of degrees of freedom for various sum of squares are as given in the following Table.

DEGREES OF FREEDOM FOR VARIOUS

S.S	T.S.S.	S.S.R.	S.S. (W × P) or S.S.(A)	S.S. E_1	S.S.(S × P) S.S.(B)	S.S.(W × P) × (S × P) S.S. (A × B)	S.S. E_2
d.f.	$pqr - 1$	$r - 1$	$p - 1$	$(p-1)(r-1)$	$q - 1$	$(p-1)(q-1)$	$p(r-1)(q-1)$

Since *d.f.* are additive, by Cochran's Theorem, the various S.S. are independent chi-square variates with respective degrees of freedom.

Expectation of Various S.S :

$$\text{S.S.R.} = \text{Sum of squares due to replicates} = pq \sum_i (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2$$

$$= pq \left[(\mu + \alpha_i + \bar{e}_{i\cdot} - \bar{e}_{i\cdot\cdot}) - (\mu + \bar{e}_{i\cdot\cdot} - \bar{e}_{\cdots})^2\right.$$

$$= pq \sum_i \left[\alpha_i + (\bar{e}_{i\cdot} - \bar{e}_{\cdot\cdot}) + (\bar{e}_{i\cdot\cdot} - \bar{e}_{\cdots})\right]^2$$

$$E(\text{SSR}) = pq \left[\sum_i \alpha_i^2 + E(\bar{e}_{i\cdot} - \bar{e}_{\cdot\cdot})^2 + E(\bar{e}_{i\cdot\cdot} - \bar{e}_{\cdots})^2\right]$$

[The expectations of all product terms are zero because of (6·246*b*)]

$$= pq \sum_i \alpha_i^2 + pq\, E\left[\sum_i \bar{e}_{i\cdot}^{\,2} - r\bar{e}^{\,2}\right] + pq\, E\left[\sum_i \bar{e}_{i\cdot\cdot}^{\,2} - r\bar{e}_{\cdot\cdot}^{\,2}\right]$$

$$= pq \sum_i \alpha_i^2 + pq\left\{r\left(\frac{\sigma_w^2}{p}\right) - r\left(\frac{\sigma_w^2}{rp}\right)\right\} + pq\left\{r\left(\frac{\sigma_s^2}{pq}\right) - r\left(\frac{\sigma_s^2}{pqr}\right)\right\}$$

$$= pq \sum_i \alpha_i^2 + q(r-1)\sigma_w^2 + (r-1)\sigma_s^2 \qquad \ldots(6\cdot250)$$

$$\Rightarrow \quad E(\text{MSR}) = E\left(\frac{\text{SSR}}{r-1}\right) = \frac{pq}{r-1}\sum_i \alpha_i^2 + q\,\sigma_w^2 + \sigma_s^2 \qquad \ldots(6\cdot250a)$$

$$\text{S.S. (A)} = \text{S.S. (WP)} = rq \sum_j (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2$$

$$= rq \sum_j \left[(\mu + \beta_j + \bar{e}_{\cdot j} + \bar{e}_{\cdot j\cdot}) - (\mu + \bar{e}_{\cdot\cdot} + \bar{e}_{\cdots})\right]^2$$

$$= rq \sum_j \left[\beta_j + (\bar{e}_{\cdot j} - \bar{e}_{\cdot\cdot}) + (\bar{e}_{\cdot j} - \bar{e}_{\cdots})\right]^2$$

$$\therefore \quad E\,(SSA) = rq\left[\sum_j \beta_j^2 + E\sum_j (\bar{e}_{.j} - \bar{e}_{..})^2 + E\sum_j (\bar{e}_{.j} - \bar{e}_{...})^2\right]$$

[All product terms are zero because of (6·246b)] ...(6·246b)

$$= rq\sum_j \beta_j^2 + rq\,E\left\{\sum_j \bar{e}_{.j}^{\,2} - p\bar{e}_{..}^{\,2}\right\} + rq\,E\left\{\sum_j \bar{e}_{..}^{\,2} - p\bar{e}_{...}^{\,2}\right\}$$

$$= rq\sum_j \beta_j^2 + rq\left\{p\left(\frac{\sigma_w^2}{r}\right) - p\,\frac{\sigma_w^2}{rp}\right\} + rq\left\{p\,\frac{\sigma_s^2}{rq} - p\,\frac{\sigma_s^2}{pqr}\right\}$$

$$= rq\sum_j \beta_j^2 + q(p-1)\,\sigma_w^2 + (p-1)\,\sigma_s^2$$

$$\Rightarrow \quad E(MSA) = E\left[\frac{(S.S.A.)}{p-1}\right] = \frac{rq}{p-1}\sum_j \beta_j^2 + q\sigma_w^2 + \sigma_s^2 \qquad ...(6·251)$$

$$S.S.B. = S.S.\,(S.P.) = rp\sum_k (\bar{y}_{..k} - \bar{y}_{...})^2$$

$$= rp\sum_k\left[(\mu + \bar{e}_{..} + \gamma_k + \bar{e}_{..k}) - (\mu + \bar{e}_{..} + \bar{e}_{...})\right]^2 = rp\sum_k\left[\gamma_k + (\bar{e}_{..k} - \bar{e}_{...})\right]^2$$

$$\therefore \quad E\,[SSB] = rp\left[\sum_k \gamma_k^2 + E\sum_k (\bar{e}_{..k} - \bar{e}_{...})^2\right]$$

[The product terms are zero because of (6·246b)]

$$= rp\sum_k \gamma_k^2 + rpE\left[\sum_k \bar{e}_{..k}^{\,2} - q\bar{e}_{...}^{\,2}\right]^{rp} = rp\sum_k \gamma_k^2 + rp\left[q\left(\frac{\sigma_s^2}{rp}\right) - q\left(\frac{\sigma_s^2}{rpq}\right)\right]$$

$$= rp\sum_k \gamma_k^2 + (q-1)\,\sigma_s^2$$

$$\Rightarrow \quad E\,[MSB] = E\left(\frac{SSB}{q-1}\right) = \frac{rp}{q-1}\sum_k \gamma_k^2 + \sigma_s^2 \qquad ...(6·252)$$

$$S.S.\,(A\times B) = r\sum_j\sum_k (\bar{y}_{.jk} - \bar{y}_{..k} - \bar{y}_{.j.} + \bar{y}_{...})^2$$

$$= r\sum_j\sum_k\Big[(\mu + \beta_j + \gamma_k + \bar{e}_{.j} + \delta_{jk} + \bar{e}_{.jk}) - (\mu + \bar{e}_{.j} + \gamma_k + \bar{e}_{..k})$$

$$- (\mu + \beta_j + \bar{e}_{.j} + \bar{e}_{.j.}) + (\mu + \bar{e}_{..} + \bar{e}_{...})\Big]^2$$

$$S.S.\,(A\times B) = r\sum_j\sum_k\left[\delta_{jk} + (\bar{e}_{.jk} - \bar{e}_{.j.} - \bar{e}_{..k} + \bar{e}_{...})\right]^2$$

$$E\,[S.S.\,(A\times B)] = r\sum_j\sum_k \delta_{jk}^2 + r\,E\left[\sum_j\sum_k\left\{(\bar{e}_{.jk} - \bar{e}_{.j.}) - (\bar{e}_{..k} - \bar{e}_{...})\right\}^2\right] \qquad ...(*)$$

[Product terms are zero because of (6·246(b).]

$$E\sum_j\sum_k = \left\{(\bar{e}_{.jk} - \bar{e}_{.j.}) - (\bar{e}_{..k} - \bar{e}_{...})\right\}^2 = E\left[\sum_k\left(\sum_j\left\{(\bar{e}_{.jk} - \bar{e}_{.j.}) - (\bar{e}_{..k} - \bar{e}_{...})^2\right\}\right)\right]$$

$$= E\left[\sum_k\left\{\sum_j (\bar{e}_{.jk} - \bar{e}_{.j.})^2 - p\,(\bar{e}_{..k} - \bar{e}_{...})^2\right\}\right]$$

$$= E\left[\sum_j\left\{\sum_k (\bar{e}_{.jk} - \bar{e}_{.j.})^2\right\} - p\sum_k (\bar{e}_{..k} - \bar{e}_{...})^2\right]$$

$$= E\left[\sum_j\left\{\sum_k \bar{e}^2_{jk} - q\,\bar{e}^2_{j.}\right\} - p\left\{\sum_k \bar{e}^2_{..k} - q\bar{e}^2_{...}\right\}\right]$$

$$= \sum_j\sum_k\left(\frac{\sigma_s^2}{r}\right) - q\sum_j\left(\frac{\sigma_s^2}{rq}\right) - p\sum_k\left(\frac{\sigma_s^2}{rq}\right) + pq\,\frac{\sigma_s^2}{pqr}$$

$$= pq\,\frac{\sigma_s^2}{r} - qp\,\frac{\sigma_s^2}{rq} - pq\,\frac{\sigma_s^2}{rp}$$

$$= \frac{1}{r}(pq - p - q + 1)\,\sigma_s^2 = \frac{1}{r}(p-1)(q-1)\,\sigma_s^2$$

Substituting in (*), we get

$$E\,[S.S.\,(A\times B) = r\sum_j\sum_k \delta^2_{jk} + (p-1)(q-1)\,\sigma_s^2$$

$$E\,[M.S.\,(A\times B) = E\left[\frac{S.S.\,(A\times B)}{(p-1)(q-1)}\right] = \frac{r}{(p-1)(q-1)}\sum_j\sum_k \delta^2_{jk} + \sigma_s^2 \qquad \ldots(6\cdot253)$$

$$S.S.E_1 = q\sum_i\sum_j\left(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}\right)^2$$

$$= q\sum_i\sum_j\Big[\left(\mu + \alpha_i + \beta_j + e_{ij} + \bar{e}_{ij.}\right) - \left(\mu + \alpha_i + \bar{e}_{i.} + \bar{e}_{i..}\right) - \left(\mu + \beta_j + \bar{e}_{.j} + \bar{e}_{.j.}\right) + \left(\mu + \bar{e}_{..} + \bar{e}_{...}\right)\Big]^2$$

$$= q\sum_i\sum_j\Big[\left(e_{ij} - \bar{e}_{i.} - \bar{e}_{.j} + \bar{e}_{..}\right) + \left(\bar{e}_{ij.} - \bar{e}_{i..} - \bar{e}_{.j} + \bar{e}_{...}\right)\Big]^2$$

$$E\,(S.S.E_1) = q\sum_i\sum_j\Bigg[\left\{\sigma_w^2 + \frac{\sigma_w^2}{p} + \frac{\sigma_w^2}{r} + \frac{\sigma_w^2}{pr} - 2\frac{\sigma_w^2}{p} - \frac{\sigma_w^2}{r} + \frac{2\sigma_w^2}{rp} + 2\frac{\sigma_w^2}{rp} - \frac{2\sigma_w^2}{rp} - \frac{2\sigma_w^2}{rp}\right\}$$
$$+ \left\{\frac{\sigma_s^2}{q} + \frac{\sigma_s^2}{pq} + \frac{\sigma_s^2}{rq} + \frac{\sigma_s^2}{pqr} - \frac{2\sigma_s^2}{pq} - \frac{2\sigma_s^2}{rq}\right\} + \frac{2\sigma_s^2}{rpq} + \frac{2\sigma_s^2}{rpq} - \frac{2\sigma_s^2}{rpq} - \frac{2\sigma_s^2}{rpq}\Bigg]$$

$$= q\sum_i\sum_j\left[\left\{\sigma_w^2 + \frac{\sigma_w^2}{pr} - \frac{\sigma_w^2}{r} - \frac{\sigma_w^2}{p}\right\} + \left\{\frac{\sigma_s^2}{q} - \frac{\sigma_s^2}{pq} - \frac{\sigma_s^2}{rq} + \frac{\sigma_s^2}{rpq}\right\}\right]$$

$$= q.\,pr.\,\sigma_w^2\left(\frac{pr + 1 - p - r}{pr}\right) + \sigma_s^2\,(pr - r - p + 1)$$

$$= q\,(p-1)(r-1)\,\sigma_w^2 + (p-1)(r-1)\,\sigma_s^2$$

$$\therefore \quad E(M.S.\,E_1) = E\left[\frac{S.S.\,E_1}{(p-1)(r-1)}\right] = q\sigma_w^2 + \sigma_s^2 \qquad \ldots(6\cdot254)$$

$$S.S.E_2 = \sum_i\sum_j\sum_k\left(y_{ijk} - \bar{y}_{ij.} - \bar{y}_{.jk} + \bar{y}_{.j.}\right)^2 = \sum_i\sum_j\sum_k\left(e_{ijk} - \bar{e}_{ij.} - \bar{e}_{.jk} + \bar{e}_{.jk}\right)^2$$

Expanding the R.H.S. and then taking expectation of both sides we get

$$E\,(S.S.E_2) = \sum_i\sum_j\sum_k\left[\sigma_s^2 + \frac{\sigma_s^2}{q} + \frac{\sigma_s^2}{r} + \frac{\sigma_s^2}{qr} - \frac{2\sigma_s^2}{q} - \frac{2\sigma_s^2}{r} + \frac{2\sigma_s^2}{rq} + \frac{2\sigma_s^2}{rq} - \frac{2\sigma_s^2}{rq} - \frac{2\sigma_s^2}{rq}\right]$$

$$= \sum_i \sum_j \sum_k \left(\sigma_s^2 - \frac{\sigma_s^2}{q} - \frac{\sigma_s^2}{r} + \frac{\sigma_s^2}{qr} \right) = pqr \left(1 - \frac{1}{q} - \frac{1}{r} + \frac{1}{qr} \right) \sigma_s^2$$

$$= p\,(qr - r - p + 1)\,\sigma_s^2 = p\,(q-1)\,(r-1)\,\sigma_s^2$$

$$\Rightarrow \quad E\,(MSE_2) = E\left[\frac{SSE_2}{p\,(q-1)\,(r-1)} \right] = \sigma_s^2 \qquad \text{...(6·255)}$$

Null Hypotheses

$$\left.\begin{aligned} H_{01} &: \beta_j = 0 \;\forall\; j = 1, 2, \ldots, p \\ H_{02} &: \gamma_k = 0 \;\forall\; k = 1, 2, \ldots, q \\ H_{03} &: \delta_{jk} = 0 \;\forall\; (j, k) \end{aligned}\right\} \qquad \text{...(6·256)}$$

Alternate Hypotheses

$$\left.\begin{aligned} H'_{01} &: \text{At least two of } \beta_j\text{'s are different.} \\ H'_{02} &: \text{At least two of } \gamma'_k\text{'s are different.} \\ H'_{03} &: \text{At least two of } \delta'_{jk} \text{ s are different.} \end{aligned}\right\} \qquad \text{...(6·256a)}$$

Test Statistics. Under $H_{01} : \beta_j = 0, \;\forall\; j = 1, 2, \ldots, p$, the test statistic for testing the homogeneity of the whole-plot treatments is :

$$F_1 = F_A = \frac{MS\,(W \times P)}{MS\,E_1} = \frac{MSA}{MSE_1} \sim F_{p-1,\,(p-1)\,(r-1)} \qquad \text{...(6·257)}$$

Under $H_{02} : \gamma_k = 0, \;\forall\; j = 1, 2, \ldots q$, the test statistic for testing the homogeneity of the sub-plot treatments is :

$$F_2 = F_B = \frac{MS\,(S \times P)}{MS\,E_2} = \frac{MSB}{MSE_2} \sim F_{q-1,\,p\,(q-1)\,(r-1)} \qquad \text{...(6·257a)}$$

Under $H_{03} : \delta_{jk} = 0, \;\forall\; (j, k)$, the test statistic for testing the independence of A and B is :

$$F_3 = F_{A \times B} = \frac{MS\,[(W \times P) \times (S \times P)]}{MSE_2} = \frac{MS\,(A \times B)}{MSE_2} \sim F_{(p-1)\,(q-1),\; p(q-1)\,(r-1)} \qquad \text{...(6·257b)}$$

The ANOVA Table for split plot design is basically split into two parts :
(*i*) The whole-plot analysis, and (*ii*) The sub-plot analysis, and is given in Table 6·74.

TABLE 6·74 : ANOVA TABLE FOR SPLIT PLOT DESIGN

Source of Variation	*d.f.*	*Sum of Squares*	*Mean Sum of Squares*	*Variance Ratio (F)*	*E(MSS)*
Replications (Blocks)	$r-1$	*SSR*	$s_R^2 = SSR/(r-1)$	$F_R = s_R^2/s_{E_1}^2 \sim F_{r-1,\,(r-1)(p-1)}$	$\frac{pq}{r-1} \sum_{i=1}^{r} \alpha_i^2 + q\sigma_w^2 + \sigma_s^2$
Whole-plot Treatments (A)	$p-1$	*SSA*	$s_A^2 = SSA/(p-1)$	$F_A = s_A^2/s_{E_1}^2 \sim F_{p-1,\,(r-1)(p-1)}$	$\frac{rq}{p-1} \sum_{j=1}^{p} \beta_j^2 + q\sigma_w^2 + \sigma_s^2$
W-P Error (Error 1)	$(r-1)\,(p-1)$	SSE_1	$s_{E_1}^2 = \frac{SSE_1}{[(r-1)(p-1)}$		$+q\,\sigma_w^2 + \sigma_a^2$
Total between whole-plots	$rp-1$	$S_w^2 = q \sum_i \sum_j (y_{ij.} - \overline{y}_{...})^2$			

Sub-plot Treatments (B)	$q-1$	SSB	$s_B^2 = SSB/(q-1)$	$F_B = s_B^2/s_{E_2}^2 \sim F_{q-1,\, p(r-1)(q-1)}$	$\frac{rp}{q-1}\sum_{k=1}^{q} \tau_k^2 + \sigma_s^2$
WPT × SPT = (A × B)	$(p-1)(q-1)$	$SS(A \times B)$	$s_{AB}^2 = \frac{SS(A\times B)}{[(p-1)(q-1)]}$	$F_{A\times B} = s_{AB}^2 / s_{E_2}^2 \sim F_{(p-1)(q-1),\, p(r-1)(q-1)}$	$\frac{r}{(p-1)(q-1)} \sum_j \sum_k \delta_{jk}^2 + \sigma_s^2$
S-P Error Error II	$p(r-1)(q-1)$	SSE_2	$s_{E_2}^2 = SSE_2/$ $[/(r-1)(q-1)]$		
Total between S–Ps within W-Ps	$rp(q-1)$				
Total	$rpq-1$	$S_T^2 = \sum_i \sum_j \sum_k (y_{ijk} - y)_{...})^2$			

***Remark.* Critical Difference Between Two Treatment Means.** If the nul hypotheses of homogeneity of treatment means are rejected, we would be interested to find which pairs of treatment means differ significantly. The standard errors of treatment means for various combinations are given in *Table 6·75.*

TABLE 6·75 : STANDARD ERRORS OF DIFFERENCE OF TWO TREATMENT MEANS

No.	*Type of Comparisons*	*Estimate of Standard Error*
1.	Difference between two whole-plot treatment means.	$\sqrt{2s_{E_1}^2/(rq)}$
2.	Difference between two sub-plot treatment means	$\sqrt{2\, s_{E_2}^2/(rp)}$
3.	Difference between two sub-plot treatment means at the same (fixed) level of whole-plot treatment.	$\sqrt{2\, s_{E_2}^2/r}$
4.	Difference between two whole-plot treatment means at the same or different level of the sub-plot treatment mean	$\sqrt{\left[2\,(q-1)\, s_{E_2}^2 + s_{E_1}^2\right]/rq}$

In the cases **1** to **3** in the *Table 6·75*,

CD between two treatment means = S.E. $(d) \times t$ (corresponding error *d.f.*) ...(6·258)

In case 4, the formula in (6·258) cannot be used because the ratio between the difference in treatment means to its S.E. does not follow the +-distribution. In this case we use an approximate method. The approximation tot he exact test involves the calculation of ∝% significant values of t say t_1 and t_2 corresponding to d.f. for SSE_1 and SSE_2 respectively *i.e.*,

$t_1 = t_{(r-1)(p-1)}\,(\alpha/2)$ and $t_2 = t_{p(r-1)(q-1)}\,(\alpha/2)$, and then calculate '$t$' by the formula :

$$t = \frac{(q-1)\, s_{E_2}^2 \cdot t_2 + s_{E_1}^2 \cdot t_1}{(q-1)\, s_{E_2}^2 + s_{E_1}^2} \qquad ...(6·258a)$$

2. Computation of Various S.S. For numerical computations, we proceed as follows :

Step 1. *Total Sum of Squares :*

Raw S.S. = R.S.S. = $\sum_{i,j,k} y_{ijk}^2$; $G = \sum_{i,j,k} y_{ijk}$; $N = r \times p \times q$

Correction Factor = C.F. = G^2/N.

Total S.S. = R.S.S. – CF. ...(6·259)

For numerical computation, we divide the analysis of Split Plot Design (S.P.D.) into two parts (*i*) Whole-plot analysis, and (*ii*) Sub-plot analysis.

Step 2 Whole-plot Analysis. We reduce the given data to into $p \times r$ two-way tale : "Whole plot treatment totals *Vs.* Blocks (Replicates) total", as given in following Table

Figures ($y_{ij.}$) in the Table represent the total of all the total of all the observations for the *j*th whole-plot in the *i*th replicate, total being taken over all the sub-plots in the given whole-plot.

$$(y_{ij.} = \sum_k y_{ijk})$$

Blocks (Replicates)	Whole-plot Treatments w_1	w_2	$w_i \ldots$	w_p	Totals
1	$y_{11.}$	$y_{12\cdot}$	$y_{1j.}$	$y_{1p.}$	T_1
2	$y_{21.}$	$y_{22.}$	$y_{2j.}$	$y_{2p.}$	T_2
⋮					⋮
j			$y_{ij.}$	…	$T_{i\cdot}$
⋮					⋮
r	$y_{r1.}$	$y_{r2.}$	$\ldots y_{rj.}$	$y_{rp.}$	T_r
Total	$T_{.1}$	$T_{.2} \ldots$	$T_{.j} \ldots$	$T_{\cdot p}$	G

Note that each of the y_{ij}. is the sum of q observations and the analysis of variance for this Table can be carried out in the usual way for two-way classification with q observations per cell.

$$\text{W.P. Treatment S.S.} = \frac{T_{.1}^2 + T_{.2}^2 + \ldots + T_{.p}^2}{r \times q} - \text{C.F.}$$

$$\text{W.P. Block S.S.} = \frac{T_{.1}^2 + T_{.2}^2 + \ldots + T_r^2}{p \times q} - \text{C.F.}$$

$$\text{W.P. Total S.S.} = \sum_i \sum_j T_{ij.}^2 - \text{C.F.}$$

$$\text{W.P. Error S.S.} = \text{SS } E_1 = \text{W.P. (TSS)} - \text{W.P. (SST)} - \text{W.P. (SSB)}$$

Step 3. Sub-plot Analysis

w_i : *i*th W.P.treatment ; s_j : *j*th sub-plot treatment

For sub-plot analysis *i.e.*, for computing S.P. treatment S.S. and SP × WP interaction S.S., we reduce the given data into the $p \times q$ table :, W.P. treatments *Vs.* S.P. treatments".

The figures (w_{ij}) in the Table corresponding to the cell $w_i\, s_j$ is the total of the combinations of w_i and s_j in all the replications of the given data. Note that each observation in this table is the sum of r observations.

S.P. Treatments	W.P. Treatments w_1	$w_2 \ldots$	$w_i \ldots$	w_p	Totals
1					T_1'
2					T_2'
⋮					⋮
j			(w_{ij})		T_j'
⋮					⋮
q					T_q'
Totals	$T_{.1}'$	$T_{.2}' \ldots$	$T_{.i}' \ldots$	$T_{.p}'$	$G' = G$

$$\text{S.P. (Treatment S.S.)} = \frac{T_1.'^2 + T_2.'^2 + \ldots + T_q.'^2}{pr} - \text{C.F.} = \text{S.P. (SST)}$$

(SP × WP) Sum of Squares (Interaction) $= \left\{ \sum_i \sum_j \left(\frac{w_{ij}^2}{r} \right) - \text{C.F.} \right\} - \text{S.P. (SST)} - \text{WP (SST) T.S.S.}$

Step 4. Error II Sum of Squares (SSE_2) :

$$SSE_2 = \text{Total S.S.} - \text{WP(BSS)} - \text{WP (SST)} - SSE_1 - \text{SP(SST)} - (\text{SP} \times \text{WP}) \text{ S.S.}$$
$$= \text{Total S.S.} - \text{Block S.S.} - \text{SSA} - SSE_1 - \text{SSB} - \text{SS}(A \times B)$$

6·12·1. Split-Plot Design *Vs.* Factorial Design

1. The layout in a split-plot design may be regarded as a factorial arrangement of two factors A and B, A at p levels and B at q levels. Regarding the sub-plot units as the experimental units and the whole plots as blocks, we find that the differences among the whole-plots are the same as the differences among the levels of the whole-point treatments and conquently the main effects of the whole-plot treatments are said to be confounded in this design. Hence, in SPD, the main effect A with $(p - 1)$ d.f. is completely confounded with the incomplete block or the whole-plot differences. Hence, *SPD is an incomplete block design.*

However, in a factorial experiments, our aim is to confound higher order interactions, which are supposed to be less important.

2. In two-factor factorial design, all the treatment combinations are allotted to the plots within a block at random while in split plot design, randomisation is done in two steps. Whole-plot treatments are allotted at random to the whole-plots within a block and then sub-plot treatments are randomly allotted to the sub-plots within each whole-plot.

3. Split-plot design is more useful than the factorial design in the following respects :

(*i*) In split-plot design sub-plot factor (B) and the interaction effect (AB) are estimated with greater precision than the whole-plot factor (A) but the average precision is same as in the case of factorial experiment. However, in a factorial experiment all the effects (main and interaction) are tested with same precision.

(*ii*) Split-plot design can be used even when all the factors are not of equal importance.

(*iii*) The size of the plots is according to the necessity of factors. Accordingly, the factor which require larger bulk of experimental material can be tested. However, in case of factorial designs, it is not possible to test the factors which require relatively large experimental material.

(*iv*) In case of split-plot design, we can include an extra factor at little extra cost without disturbing the original layout of the RBD. This is not possible in case of factorial experiment.

6·12·2. Advantages and Disadvantages of SPD

1. The split-plot design is advantageous when :

(*i*) The main effects of one of the two factors are large enough to be deducted even with a lower precision, so that the factor may be allotted to the main plots.

(*ii*) The main effects of the factor allotted to the main plots are not of much interest as compared to the effects of the factor allocated to the sub-plots and the interaction between the two factors is of primary interest.

2. Experimental units which are large by necessity or design may be used to compare subsidiary treatment at little extra cost. If it is possible to introduce another factor which requires small plots (units), in designing an SPD, of the two factors, one which is more important should be allocated to the split-plots, if possible (*i.e.*, provided it can be accommodated in smaller units).

3. As pointed out earlier, SPD is more useful if :

(*i*) Sub-plot treatment effects (*B*) and its interaction with the whole-plot treatment effects (*AB*) are of greater interest than the whole-plot treatment effects (A), *or*

(*ii*) the whole-plot effect cannot be estimated on small material.

4. In SPD, of the two errors, $SSE_2 < SSE_1$. This implies that, usually the main effect *B* and the interaction effect *AB* will be estimated and tested more precisely than the main effect *A*.

5. Overall precision SPD can, however, be increased by designing the whole-plot treatments in a Latin square or in an incomplete Latin square. (This is, however, not discussed in the book.)

Disadvantages

1. The whole-plot treatments are measured with less precision than they are in a randomised complete block design of *pq* treatments in each of *r* replications.

2. The computation of two types of error sum of squares SSE_1 and SSE_2 makes the analysis more complex or difficult.

3. Sometimes, the whole-plot error (SSE_1) is much greater than the sub-plot error (SSE_2) so that the effects of the main plot treatments, though large and exciting are not significant, whereas those of the sub-plot treatments are too small to be of practical interest, are statistically significant.

4. The different treatment comparisons have different basic error variances (SEs) which make the analysis more complex as compared with the corresponding R.B.D.

6·12·3. Efficiency of Split plot Design. The ANOVA for the split-plot design is given in *Table 6·74*. The estimated information on the whole-plot treatments is proportional to $(1/s_{E_1}^2)$ and on the split-plot is $\left(1/s_{E_2}^2\right)$ Disregarding the difference in the number of d.f., the efficiency of the split-plot design relative to randomised complete block design on the *B* and the $A \times B$ comparison is :

$$E' = \frac{\{(\phi - 1) + (r-1)(p-1)\}\, s_{E_1}^2 + \{(q-1) + (q-1) = p(q-1)(r-1)\} s_{E_2}^2}{(pqr - r)\, s_{E_2}^2}$$

$$= \frac{r(p-1)\, s_{E_1}^2 + rp\,(q-1)\, s_{E_2}^2}{(pqr - r)\, s_{E_2}^2} = \frac{(p-1)\, s_{E_1}^2 + p(q-1)\, s_{E_2}^2}{(pq-1)\, s_{E_2}^2} \qquad \text{...(6·260)}$$

which is the weighted arithmetic means of $s_{E_1}^2$ and $s_{E_2}^2$, the corresponding weights being $(p - 1)$ and $p(q-1)$ respectively and hence is intermediate between $s_{E_1}^2$ and $s_{E_2}^2$.

On the other hand, the efficiency (E″) on the *A*-affects or whole-plot comparisons would be decreased and is given by :

$$E'' = \frac{(p-1)\, s_{E_1}^2 + p(q-1)\, s_{E_2}^2}{(pq-1)\, s_{E_1}^2} \qquad \text{...(6·260a)}$$

6·7. ANALYSIS OF COVARIANCE (ANOCOVA)

The basic objective of the designs considered so far is to make the treatment comparisons with the greatest precision by reducing the experimental error through the powerful tool of local control. *Analysis of Covariance* (ANOCOVA), like Randomised Block Design or Latin Square Design, is a technique of increasing the precision of the design by reducing the experimental error.

ANOCOVA is a technique in which it is possible to control certain sources of variation by taking additional observations on each of the experimental units. Let us suppose that in an experiment, y is the response variable and x is another variable which is linearly related to y. Moreover x cannot be controlled by the experimenter but can be observed alongwith the y's. The variable x is called the *covariate / concomitant / independent / ancillary variable*. In ANOCOVA we adjust for the variation in the response variable (y) for the linear regression (effect) of the independent variable (x). If this is not done, then the error mean square will be inflated due to the linear effect of x, thus making it difficult to detect the true differences in the response variable. ANOCOVA procedure is a combination of the Analysis of Variance (ANOVA) and the Regression Analysis. Whenever it is possible to take additional observations on one or more of the variables from each of the experimental units in the design alongwith the response variable under study, the ANOCOVA technique has proved to be useful in many fields of research.

We give below some illustrations for the use of ANOCOVA by identifying the response variable (y) and the concomitant variable (x).

***Illustration* 1.** Suppose we want to compare the effect of some rations (diets) on the weight of animals. We can analyse the data by performing the ANOCOVA by regarding :

y : the final weight of the animals taking the ration (diet), after a specified period (days / weeks / months) as the response variable.

x : the initial weight of the animals at the time of starting the experiment as the concomitant variable.

To ensure that the real differences in the final weights (y) are due to rations, we must adjust for the linear effect of the initial weight (x) on y.

Illustration 2. Suppose we want to compare the differences in the strength of the filament fibre (y) produced by different machines. Obviously y depends on the thickness (x) of the fibre-thicker the fibre, stronger it is. The effect of the thickness (x) on the strength (y) can be eliminated by performing ANOCOVA between the response variable (y) and the concomitant variable (x), for testing the differences in the strength of the fibre produced by different machines.

Illustration 3. In plant breeding experiments, suppose an equal number of seeds are sown per plot but at the time of harvest, the final number of plants in each plot will not be same due to certain reasons (like non-germination of certain seeds, early death of certain plants, attack by birds / cattle, etc.) and will vary from plot to plot. The yield (y) of a crop from different plots may depend on the number of plants (x) per plot. To study the real differences between the yields, we adjust for the linear effect of the number of plants per plot by performing ANOCOVA by regarding the yield per plot (y) as the response variable and the number of plants per plot (x) as the concomitant variable.

Illustration 4. Suppose we want to compare the different methods of teaching (classroom lectures, correspondence courses, on-line teaching, etc.) by observing the scores of the students, all of whom take the same final examination. The variation in their scores may not be simply due to the different methods of teaching but also due to the intelligence quotient (I.Q.) levels of the students. Hence the proper way to compare the different methods of teaching will be to adjust the observed data (scores) for the I.Q. of the students by carrying out the ANOCOVA by taking 'the score (y) of the student' as the response variable and 'the I.Q. (x) of the student' as the concomitant variable.

The above illustrations give us a glimpse of some diverse fields in which ANOCOVA can be used.

***Remark.* Choice of Concomitant Variable.** The concomitant variable need not necessarily be measurable. Even if it is a quality characteristic which cannot be measured quantitatively, *e.g.*, intelligence, poverty, indifference, good / bad, presence / absence, etc., but can be suitably converted into numerical scores, the use of ANOCOVA results in a considerable increase in precision.

6·7·1. ANOCOVA For One Way Classification With a Single Concomitant Variable in C.R.D. Layout. Let us suppose that we are comparing v treatments $t_1, t_2, \ldots, t_v$; ith treatment replicated r_i, $(i = 1, 2, \ldots, v)$ times so that $n \sum_{i=1}^{v} r_i$, is the total number of experimental units. Further suppose that the experiment is conducted with a Completely Randomised Design (CRD) layout.

Suppose that along with the response (dependent) variable y, we consider a single concomitant variable x. Then the linear ANOCOVA model will consist of the sum of two components-one is the same component as in ANOVA and the second component is due to the regression of y on the concomitant variable x.

Then assuming a linear relationship between the response variable (y) and the concomitant variable (x), the appropriate statistical model (for fixed effects) for ANOCOVA for Completely Randomised Design (CRD) with one concomitant variable is given by :

$$y_{ij} = \mu + \alpha_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij}\ ;\ (i = 1, 2, ..., r\ ;\ j = 1, 2, ..., r_i) \quad ...(6\cdot82)$$

where (in the usual notation of ANOVA)

(i) μ is the general mean effect,

(ii) α_i is the (fixed) additional effect due to the ith treatment ($i = 1, 2, ..., v$),

(iii) ε_{ij} is the random error effect,

(iv) β is the coefficient of regression of y on x, and

(v) x_{ij} is the value of concomitant variable corresponding to the response variable y_{ij}

so that (i) $\sum_{i=1}^{v} \alpha_i = 0$ and $\varepsilon_{ij} \overset{i.i.d}{\sim} N(0, \sigma_e^2)$ ...(6·83)

Estimation of Parameters in (6·82). We shall now estimate the parameters μ, α_i (i = 1, 2, ..., v) and β, using the principle of least squares by minimising the sum of squares of errors in (6·82), *viz.*,

$$SSE = \sum_i \sum_j \varepsilon_{ij}^2 = \sum_{i=1}^{v} \sum_{j=1}^{r_i} \left[y_{ij} - \mu - \alpha_i - \beta\,(x_{ij} - \bar{x}_{..}) \right]^2 \quad ...(6\cdot84)$$

The normal equations for estimating the parameters are :

$$\frac{\partial}{\partial \mu}(SSE) = 0 = -2 \sum_i \sum_j \left[y_{ij} - \mu - \alpha_i - \beta(x_{ij} - \bar{x}_{..}) \right] \quad ...(6\cdot85)$$

$$\frac{\partial}{\partial \alpha_i}(SSE) = 0 = -2 \sum_j \left[y_{ij} - \mu - \alpha_i - \beta\,(x_{ij} - \bar{x}_{..}) \right] \quad ...(6\cdot86)$$

$$\frac{\partial}{\partial \beta}(SSE) = 0 = -2 \sum_i \sum_j \left[\left\{ y_{ij} - \mu - \alpha_i - \beta(x_{ij} - \bar{x}_{..}) \right\} (x_{ij} - \bar{x}_{..}) \right] \quad ...(6\cdot87)$$

From (6·85), we get

$$\sum_i \sum_j y_{ij} - \mu \sum_i r_i - \sum_i \alpha_i - \beta \sum_i \sum_j (x_{ij} - \bar{x}_{..}) = 0 \;\Rightarrow\; \hat{\mu} = \frac{\sum_i \sum_j y_{ij}}{\sum_i r_i} = \bar{y}_{..} \quad ...(6\cdot88)$$

$$\left[\text{On using (6·83) and } \sum_i \sum_j (x_{ij} - \bar{x}_{..}) = 0\right]$$

From (6·86), we get

$$\sum_i y_{ij} - r_i \left(\hat{\mu} + \hat{\alpha}_i\right) - \hat{\beta} \sum_j (x_{ij} - \bar{x}_{..}) = 0$$

$$\Rightarrow \quad r_i\, \bar{y}_{i.} - r_i \left(\bar{y}_{..} + \hat{\alpha}_i\right) - \hat{\beta} \,.\, r_i \left(\bar{x}_{i.} - \bar{x}_{..}\right) = 0$$

$$\Rightarrow \quad \bar{y}_{i.} - \bar{y}_{..} - \hat{\alpha}_i - \hat{\beta}\,(\bar{x}_{i.} - \bar{x}_{..}) = 0$$

$$\Rightarrow \quad \hat{\alpha}_i = \left(\bar{y}_{i.} - \bar{y}_{..}\right) - \hat{\beta}\left(\bar{x}_{i.} - \bar{x}_{..}\right) \quad ...(6\cdot89)$$

From (6·87), we get

$$\sum_i \sum_j \left[\left\{ y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}\left(x_{ij} - \bar{x}_{..}\right) \right\} (x_{ij} - \bar{x}_{..}) \right] = 0$$

$$\Rightarrow \quad \sum_i \sum_j \left[\left\{ (y_{ij} - \bar{y}_{..}) - (\bar{y}_{i.} - \bar{y}_{..}) + \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}) - \hat{\beta}(x_{ij} - \bar{x}_{..}) \right\} \times \left\{ x_{ij} - \bar{x}_{..} \right\} \right] = 0$$

[Using (6·88) and (6·89)]

$$\Rightarrow \quad \sum_i \sum_j \left[\left\{ (y_{ij} - \bar{y}_{i.}) - \hat{\beta}(x_{ij} - \bar{x}_{i.}) \right\} (x_{ij} - \bar{x}_{..}) \right] = 0$$

$$\Rightarrow \quad \sum_i \sum_j \left[\left\{ (y_{ij} - \bar{y}_{i.}) - \hat{\beta}(x_{ij} - \bar{x}_{i.}) \right\} \left\{ x_{ij} - \bar{x}_{i.} + \bar{x}_{i.} - \bar{x}_{..} \right\} \right] = 0$$

$$\Rightarrow \quad \sum_i \sum_j (y_{ij} - \bar{y}_{i.})(x_{ij} - \bar{x}_{i.}) - \hat{\beta} \sum\sum (x_{ij} - \bar{x}_{i.})^2 = 0$$

$$\left[\because \sum_i \sum_j (y_{ij} - \bar{y}_{i.})(\bar{x}_{i.} - \bar{x}_{..}) = \sum_i \left\{ (\bar{x}_{i.} - \bar{x}_{..}) \sum_j (y_{ij} - \bar{y}_{i.}) \right\} \right] = 0$$

and $\sum_i \sum_j (x_{ij} - \bar{x}_{i.})(\bar{x}_{i.} - \bar{x}_{..}) = \sum_i \left[(\bar{x}_{i.} - \bar{x}_{..}) \sum_j (x_{ij} - \bar{x}_{i.}) \right] = 0$

$$\Rightarrow \quad \hat{\beta} = \frac{\sum_i \sum_j (y_{ij} - \bar{y}_{i.})(x_{ij} - \bar{x}_{i.})}{\sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2} \qquad \ldots (6·90)$$

Let us write

$$E_{xx} = \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2 \;\; ; \;\; E_{yy} = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 \qquad \ldots (6·91)$$

and $$E_{xy} = \sum_i \sum_j (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.})$$

Then,

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}} \qquad \ldots (6·92)$$

Substituting the estimated values of the parameter $\hat{\mu}$, $\hat{\alpha}_i$ and $\hat{\beta}$ in (6·84), we get

$$SSE = \sum_i \sum_j \left[y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}(x_{ij} - \bar{x}_{..}) \right]^2 = \sum_i \sum_j \left[(y_{ij} - \bar{y}_{i.}) - \hat{\beta}(x_{ij} - \bar{x}_{i.}) \right]^2$$

$$= \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + \hat{\beta}^2 \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2 - 2\hat{\beta} \sum_i \sum_j (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.})$$

$$= E_{yy} + \left(\frac{E_{xy}}{E_{xx}} \right)^2 . E_{xx} - 2 \frac{E_{xy}}{E_{xx}} . E_{xy} \qquad \text{[From (6·91) and (6·92)]}$$

$$\Rightarrow \quad SSE = E_{yy} - \frac{E_{xy}^2}{E_{xx}} \qquad \ldots (6·93)$$

E_{yy} is the Error Sum of Squares for *C.R.D.*

Since $(E_{xy}^2/E_{xx}) > 0$, there is a reduction in SSE if we apply ANOCOVA to CRD.

$d.f.$ for SSE = Total $d.f.$ – $d.f.$ due to treatments – 1 $d.f.$ due to β.

$= (n - 1) - (v - 1) - 1 = n - v - 1 \qquad \ldots (6·93a)$

Under the *null hypotheses* :

H_0 : All the treatments are equally effective, i.e., $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_v = 0$ ... (6·94)

the model (6·82) reduces to :

$$y_{ij} = \mu + \beta'(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij}^* \quad \text{... (6·94a)}$$

Error Sum of squares under H_0 is given by :

$$(SSE)^* = \sum_i \sum_j \varepsilon_{ij}^{*2} = \sum_i \sum_j \left[y_{ij} - \mu - \beta'(x_{ij} - \bar{x}_{..})\right]^2 \quad \text{... (6·95)}$$

Normal equations for estimating μ and $\hat{\beta}'$ are :

$$\frac{\partial}{\partial \mu}(SSE)^* = 0 = -2 \sum_i \sum_j \left[y_{ij} - \mu - \beta'(x_{ij} - \bar{x}_{..})\right] \quad \text{... (6·96)}$$

$$\frac{\partial}{\partial \hat{\beta}'}(SSE)^* = 0 = -2 \sum_i \sum_j \left[\left\{y_{ij} - \mu - \beta'(x_{ij} - \bar{x}_{..})\right\}\left\{x_{ij} - \bar{x}_{..}\right\}\right] \quad \text{... (6·97)}$$

$$\Rightarrow \quad \hat{\mu} = \frac{\sum_i \sum_j y_{ij}}{\sum_i r_i} = \bar{y}_{..} \quad \text{... (6·98)}$$

$$\Rightarrow \quad \sum_i \sum_j \left[\left\{y_{ij} - \hat{\mu} - \hat{\beta}'(x_{ij} - \bar{x}_{..})\right\}(x_{ij} - \bar{x}_{..})\right] = 0$$

$$\Rightarrow \quad \sum_i \sum_j \left[\left\{(y_{ij} - \bar{y}_{..}) - \hat{\beta}'(x_{ij} - \bar{x}_{..})\right\}(x_{ij} - \bar{x}_{..})\right] = 0$$

$$\Rightarrow \quad \hat{\beta}' = \frac{\sum_i \sum_j (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..})}{\sum_i \sum_j (x_{ij} - \bar{x}_{..})^2} \quad \text{... (6·99)}$$

Let us write :

$$SS_{xx} = \sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 \; ; \; SS_{yy} = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 \quad \text{... (6·100)}$$

$$SS_{xy} = \sum_i \sum_j (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..})$$

$$\therefore \quad \hat{\beta}' = \frac{SS_{xy}}{SS_{xx}} \quad \text{... (6·101)}$$

Substituting from (6·98) and (6·101) in (6·95), we get the restricted error sum of squares under H_0 as :

$$(SSE)^* = \sum_i \sum_j \left[y_{ij} - \hat{\mu} - \hat{\beta}'(x_{ij} - \bar{x}_{..})\right]^2 = \sum_i \sum_j \left[(y_{ij} - \bar{y}_{..}) - \hat{\beta}'(x_{ij} - \bar{x}_{..})\right]^2$$

$$= \sum_i \sum_j \left[(y_{ij} - \bar{y}_{..})^2 + \hat{\beta}'^2 \sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 - 2\hat{\beta}' \sum_i \sum_j (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..})\right.$$

$$= SS_{yy} + \left(\frac{SS_{xy}}{SS_{xx}}\right)^2 . SS_{xx} - 2\frac{SS_{xy}}{SS_{xx}} . SS_{xy} \quad \text{[Using (6·100) and (6·101)]}$$

$$= SS_{yy} - \frac{SS_{xy}^2}{SS_{xx}} \quad \text{... (6·102)}$$

$d.f.$ for $(SSE)^*$ = Total $d.f.$ − $d.f.$ for $\beta = (n-1) - 1 = n - 2$ ... (6·102a)

Hence, the sum of squares due to treatments is given by :

$$SST\,(S_t^2) = (SSE)^* - (SSE) \qquad \ldots(6\cdot103)$$

$$d.f. \text{ for } SST = d.f.\,(SSE)^* - d.f.\,(SSE) = n - 2 - (n - v - 1) = v - 1$$

$$\therefore \quad MST = s_t^2 = \frac{SST}{d.f.} = \frac{(SSE)^* - (SSE)}{v - 1} \qquad \ldots(6\cdot104)$$

$$MSE = s_E^2 = \frac{SSE}{d.f} = \frac{SSE}{n - v - 1} \qquad \ldots(6\cdot105)$$

Hence, the test statistic for testing, $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$ is given by :

$$F = \frac{MST}{MSE} = \frac{s_t^2}{s_E^2} = \left[\frac{(SSE)^* - (SSE)}{v - 1}\right] \times \left(\frac{n - v - 1}{SSE}\right) \sim F_{(v-1,\, n-v-1)} \qquad \ldots(6\cdot106)$$

If $F > F_{v-1,\, n-v-1}(\alpha)$, then we reject H_0 at 'α' level of significance otherwise we fail to reject H_0.

ANOCOVA Table for one Way Classification (CRD Layout). Let us define :

$$T_{xx} = \sum_{i=1}^{v} r_i\,(x_{ij} - \bar{x}_{i\cdot})^2 \;;\quad T_{yy} = \sum_i r_i\,(y_{ij} - \bar{y}_{i\cdot})^2 \qquad \ldots(6\cdot107)$$

$$T_{xy} = \sum_i r_i\,(x_{ij} - \bar{x}_{i\cdot})\,(y_{ij} - \bar{y}_{i\cdot})$$

Then, we have

$$\sum_i \sum_j (x_{ij} - \bar{x}_{\cdot\cdot})^2 = \sum_{i=1}^{u} \sum_{j=1}^{r_i} \left[(x_{ij} - \bar{x}_{i\cdot}) + (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})\right]^2$$

$$= \sum_i \sum_j (x_{ij} - \bar{x}_{i\cdot})^2 + \sum_i r_i\,(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2,$$

The product term will be zero, since the algebraic sum of deviations from mean is zero.

$$\Rightarrow \quad SS_{xx} = \sum_i \sum_j (x_{ij} - \bar{x}_{\cdot\cdot})^2 = E_{xx} + T_{xx} \qquad \ldots(6\cdot108)$$

Similarly, we can prove that

$$SS_{yy} = E_{yy} + T_{yy} \qquad \ldots(6\cdot109)$$

Also

$$SS_{xy} = \sum_i \sum_j (x_{ij} - \bar{x}_{\cdot\cdot})\,(y_{ij} - \bar{y}_{\cdot\cdot})$$

$$= \sum_i \sum_j \left[(x_{ij} - \bar{x}_{i\cdot} + \bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})\,(y_{ij} - \bar{y}_{i\cdot} + \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})\right]$$

$$= \sum_i \sum_j (x_{ij} - \bar{x}_{i\cdot})\,(y_{ij} - \bar{y}_{i\cdot}) + \sum_i \left[(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) \sum_j (x_{ij} - \bar{x}_{i\cdot})\right]$$

$$+ \sum_i \left[(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot}) \sum_j (y_{ij} - \bar{y}_{i\cdot})\right] + \sum_i \sum_j (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})\,(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})$$

$$= E_{xy} + 0 + 0 + T_{xy}$$

$$\Rightarrow \quad SS_{xy} = E_{xy} + T_{xy} \qquad \ldots(6\cdot110)$$

Results in (6·108), (6·109), and (6·110) give the partitioning of the total sum of squares due to x (SS_{xx}), the total sum of squares due to y (SS_{yy}) and total sum of products (SP) of x and y (SS_{xy}) respectively.

Using these, notations the above statistical analysis can be expressed elegantly as given in the ANOCOVA *Table 6·25*.

TABLE 6·25 : ANOCOVA TABLE (CRD)

Source of Variation	Sum of Squares and Products				Estimate of β	Adjusted SS_{yy}	Adjusted d.f.
	d.f.	SS_{xx}	SS_{xy}	SS_{yy}			
Classes (Treatments)	$v-1$	T_{xx}	T_{xy}	T_{yy}			
Error	$n-v$	E_{xx}	E_{xy}	E_{yy}	E_{xy}/E_{xx}	SSE	$n-v-1$
Total	$n-1$	E'_{xx}	E'_{xy}	E'_{yy}	E'_{xy}/E'_{xx}	(SSE^*)	$n-2$
Difference (Total – Error)						$(SSE^*)-(SSE)$	$v-1$

Note.
$$E'_{xx} = \sum_i \sum_j (x_{ij}-\bar{x}_{..})^2 = SS_{xx} = T_{xx}+E_{xx} \quad \text{... (6·111)}$$
$$E'_{yy} = \sum_i \sum_j (y_{ij}-\bar{y}_{..})^2 = SS_{yy} = T_{yy}+E_{yy} \quad \text{... (6·111a)}$$
$$E'_{xy} = \sum_i \sum_j (x_{ij}-\bar{x}_{..})(y_{ij}-\bar{y}_{..}) = SS_{xy} = T_{xy}+E_{xy} \quad \text{... (6·111b)}$$

8·7·2. Analysis of Covariance for Two-way Classification (Random Block Design) with One Concomitant Variable. Suppose we want to compare v treatments, each treatment replicated r times so that total number of experimental units is $n = vr$. Suppose that the experiment is conducted with a Randomised Block Design (RBD) layout.

Assuming a linear relationship between the response variable (y) and concomitant variable (x), the appropriate statistical model for ANOCOVA for RBD (with one concomitant variable) is :

$$y_{ij} = \mu + \alpha_i + \theta_j + \beta(x_{ij}-\bar{x}_{..}) + \varepsilon_{ij} \quad \text{... (6·112)}$$

where

(*i*) μ is the general mean effect,

(*ii*) α_i is the (fixed) additional effect due to the *i*th treatment, ($i = 1, 2, ..., v$) ;

(*iii*) θ_j is the (*fixed*) additional effect due to the *j*th block, ($j = 1, 2, ..., r$) ;

(*iv*) β is the coefficient of regression of y on x ,

(*v*) x_{ij} is the value of the concomitant variable corresponding to the response variable y_{ij}; and

(*vi*) ε_{ij} is the random error effect so that :

$$\sum_{i=1}^{v} \alpha_i = 0 \ , \ \sum_{j=1}^{r} \theta_j = 0 \ , \ \text{and} \ \varepsilon_{ij} \overset{i.i.d}{\sim} N(0, \sigma_e^2) \quad \text{(6·113)}$$

Estimation of Parameters in (6·112). We shall estimate the parameters μ, α_i ($i = 1, 2, ..., v$), θ_j ($j = 1, 2, ..., r$) and β, using the principle of least squares by minimising the error sum of squares in (6·112), *viz.*,

$$SSE = \sum_i \sum_j \varepsilon_{ij}^2 = \sum_{i=1}^{v} \sum_{j=1}^{r} \left[y_{ij} - \mu - \alpha_i - v_j - \beta(x_{ij}-\bar{x}_{..}) \right]^2 \quad \text{... (6·114)}$$

Normal equations for estimating the parameters are :

$$\frac{\partial}{\partial\mu}(SSE) = 0 = -2\sum_i\sum_j\left[y_{ij}-\mu-\alpha_i-\theta_j-\beta(x_{ij}-\bar{x}_{..})\right] \quad \text{... (6·115)}$$

$$\frac{\partial}{\partial\alpha_i}(SSE) = 0 = -2\sum_j\left[y_{ij}-\mu-\alpha_i-\theta_j-\beta(x_{ij}-\bar{x}_{..})\right] \quad \text{... (6·116)}$$

$$\frac{\partial}{\partial\theta_j}(SSE) = -2\sum_i\left[y_{ij}-\mu-\alpha_i-\theta_j-\beta(x_{ij}-\bar{x}_{..})\right] \quad \text{... (6·117)}$$

$$\frac{\partial}{\partial\beta}(SSE) = -2\sum_i\sum_j\left[\left\{y_{ij}-\mu-\alpha_i-\theta_j-\beta(x_{ij}-\bar{x}_{..})\right\}(x_{ij}-\bar{x}_{..})\right] \quad \text{... (6·118)}$$

(6·115) on using (6·113) gives :

$$\hat{\mu} = \frac{\sum_i\sum_j y_{ij}}{rv} = \bar{y}_{..} \quad \text{... (6·119)}$$

Similarly, as in § 6·8·1, we shall get from (6·116) and (6·117) respectively :

$$\hat{\alpha}_i = (\bar{y}_{i.}-\bar{y}_{..})-\hat{\beta}(\bar{x}_{i.}-\bar{x}_{..})$$

and $$\hat{\theta}_j = (\bar{y}_{.j}-\bar{y}_{..})-\hat{\beta}(\bar{x}_{.j}-\bar{x}_{..}) \quad \text{... (6·120)}$$

Substituting these estimated values in (6·118), we get

$$0 = \sum_i\sum_j\left[\left\{y_{ij}-\bar{y}_{..}-\left\{\bar{y}_{i.}-\bar{y}_{..}-\hat{\beta}(\bar{x}_{i.}-\bar{x}_{..})\right\}-\left\{\bar{y}_{.j}-\bar{y}_{..}-\hat{\beta}(\bar{x}_{.j}-\bar{x}_{..})\right\}-\hat{\beta}(x_{ij}-\bar{x})\right\}(x_{ij}-\bar{x}_{..})\right]$$

$$\Rightarrow \quad 0 = \sum_i\sum_j\left[\left\{(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})-\hat{\beta}(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})\right\}(x_{ij}-\bar{x}_{..})\right]$$

$$\Rightarrow \quad \sum_i\sum_j\left[\left\{(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})-\hat{\beta}(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})\right\}\right.$$
$$\left.\times\left\{(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})+(\bar{x}_{i.}-\bar{x}_{..})+(\bar{x}_{.j}-\bar{x}_{..})\right\}\right] = 0$$

$$\Rightarrow \quad \sum_i\sum_j(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})-\hat{\beta}\sum_i\sum_j(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})^2 = 0,$$

(the product terms will be zero since algebraic sum of deviations from mean is zero).

$$\Rightarrow \quad \hat{\beta} = \frac{\sum_i\sum_j(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})}{\sum_i\sum_j(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})^2}$$

Let us write :

$$E_{xx} = \sum_i\sum_j(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})^2\ ;\ E_{yy} = \sum_i\sum_j(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2$$

and $$E_{xy} = \sum_i\sum_j(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})$$

Then, $$\hat{\beta} = \frac{E_{xy}}{E_{xx}} \quad \text{... (6·121)}$$

Substituting the values of $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\theta}_j$ and $\hat{\beta}$ in (6·114), the unrestricted error sum of squares for model (6·112) becomes :

$$SSE = \text{Minimum value of error } S.S.$$

$$= \sum_i \sum_j \left[y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\theta}_j - \hat{\beta}(x_{ij} - \bar{x}_{..})\right]^2 \qquad \text{[From (6·114)]}$$

$$= \sum_i \sum_j \left[(y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{..}) - \hat{\beta}(x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}_{..})\right]^2 \text{ (On simplification)}$$

$$= \sum_i \sum_j (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{..} + \bar{y}_{..})^2 + \hat{\beta}^2 \sum_i \sum_j (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}_{..})^2$$

$$- 2\hat{\beta} \sum_i \sum_j (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}_{..})(y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{..})$$

$$= E_{yy} + \left(\frac{E_{xy}}{E_{xx}}\right)^2 . E_{xx} - 2 . \frac{E_{xy}}{E_{xx}} . E_{xy}$$

$$= E_{yy} - \frac{E_{xy}^2}{E_{xx}} = E_{yy} - \hat{\beta} E_{xy} \qquad \ldots (6·122)$$

$$= \text{Error } S.S. \text{ for } y \text{ in } RBD - \frac{E_{xy}^2}{E_{xx}} \qquad \ldots (6·122a)$$

Since $E_{xy}^2/E_{xx} > 0$, there is reduction in SSE if we apply ANOCOVA to RBD, *the reduction in SSE due to the regression of y on x being* $(E_{xy}^2/E_{xx}) = \hat{\beta} E_{xy}$.

d.f. for SSE = Total *d.f.* – (*d.f.* due to treatments) – (*d.f.* due to blocks) – (*d.f.* due to β)

$$= (rv - 1) - (v - 1) - (r - 1) - 1 = (r - 1)(v - 1) - 1 \qquad \ldots (6·123)$$

Under the null hypothesis :

H_0 : All treatment effects are equal, *i.e.*, $H_0 : \alpha_i = \alpha_2 = \ldots = \alpha_v = 0$, ... (6·124)

the model (6·112) reduces to

$$y_{ij} = \mu + \theta_j + \beta'(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \qquad \ldots (6·124a)$$

Restricted error sum of squares under H_0 is given by :

$$(SSE)^* = \sum_i \sum_j \varepsilon_{ij}'^2 = \sum_{i=1}^{v} \sum_{j=1}^{r} \left[y_{ij} - \mu - \theta_j - \beta'(x_{ij} - \bar{x}_{..})\right]^2 \qquad \ldots (6·125)$$

The normal equations for estimating μ, θ_j and β' are given by :

$$\frac{\partial}{\partial \mu}(SSE)^* = 0 = -2 \sum_i \sum_j \left[y_{ij} - \mu - \theta_j - \beta'(x_{ij} - \bar{x}_{..})\right] \qquad \ldots (6·126)$$

$$\frac{\partial}{\partial \theta_j}(SSE)^* = 0 = -2 \sum_i \left[y_{ij} - \mu - \theta_j - \beta'(x_{ij} - \bar{x}_{..})\right] \qquad \ldots (6·127)$$

$$\frac{\partial}{\partial \beta'}(SSE)^* = 0 = -2 \sum_i \sum_j \left[\left\{y_{ij} - \mu - \theta_j - \beta'(x_{ij} - \bar{x}_{..})\right\}(x_{ij} - \bar{x}_{..})\right] \qquad \ldots (6·128)$$

Equation (6·126) $\Rightarrow$ $\hat{\mu} = \dfrac{\sum_i \sum_j y_{ij}}{vr} = \bar{y}_{..}$ ... (6·129)

Equation (6·127) $\Rightarrow$ $\sum_i \left[y_{ij} - \hat{\mu} - \hat{\theta}_j - \hat{\beta}'(x_{ij} - \bar{x}_{..})\right] = 0$

$$\Rightarrow \quad v\,\bar{y}_{.j} - v\,\bar{y}_{..} - v\hat{\theta}_j - \hat{\beta}'\,v\,(\bar{x}_{.j} - \bar{x}_{..}) = 0$$

$$\Rightarrow \quad \hat{\theta}_j = (\bar{y}_{.j} - \bar{y}_{..}) - \hat{\beta}'\,(\bar{x}_{.j} - \bar{x}_{..}) \qquad \text{... (6·130)}$$

From (6·128), we get

$$\sum_i \sum_j \Big[(x_{ij} - \bar{x}_{..})\,\big\{y_{ij} - \bar{y}_{..} - (\bar{y}_{.j} - \bar{y}_{..}) + \hat{\beta}'\,(\bar{x}_{.j} - \bar{x}_{..}) - \hat{\beta}'\,(x_{ij} - \bar{x}_{..})\big\}\Big] = 0$$

$$\Rightarrow \quad \sum_i \sum_j \Big[(x_{ij} - \bar{x}_{..})\,\big\{(y_{ij} - \bar{y}_{.j}) - \hat{\beta}'\,(x_{ij} - \bar{x}_{.j})\big\}\Big] = 0$$

$$\Rightarrow \quad \sum_i \sum_j \Big[(x_{ij} - \bar{x}_{.j} + \bar{x}_{.j} - \bar{x}_{..})\,\big\{(y_{ij} - \bar{y}_{.j}) - \hat{\beta}'\,(x_{ij} - \bar{x}_{.j})\big\}\Big] = 0$$

$$\Rightarrow \quad \sum_i \sum_j (x_{ij} - \bar{x}_{.j})(y_{ij} - \bar{y}_{.j}) - \hat{\beta}' \sum\sum (x_{ij} - \bar{x}_{.j})^2 = 0,$$ the other product terms are zero.

$$\Rightarrow \quad \hat{\beta}' = \frac{\sum_i \sum_j (x_{ij} - \bar{x}_{.j})(y_{ij} - \bar{y}_{.j})}{\sum_i \sum_j (x_{ij} - \bar{x}_{.j})^2} \qquad \text{... (6·131)}$$

Let us define :

$$E_{xx}' = \sum_i \sum_j (x_{ij} - \bar{x}_{.j})^2\,;\; E_{yy}' = \sum_i \sum_j (y_{ij} - \bar{y}_{.j})^2;\; E_{xy}' = \sum_i \sum_j (x_{ij} - \bar{x}_{.j})(y_{ij} - \bar{y}_{.j}) \qquad \text{... (6·132)}$$

$$\therefore \quad \hat{\beta}' = \frac{E_{xy}'}{E_{xx}'} \qquad \text{... (6·133)}$$

Hence, under H_0, the restricted error sum of squares is given by :

$(SSE)^*$ = Minimum value of error $S.S.$ in (6·125)

$$= \sum_i \sum_j \big[y_{ij} - \hat{\mu} - \hat{\theta}_j - \hat{\beta}'\,(x_{ij} - \bar{x}_{..})\big]^2 \qquad \text{[From (6·125)]}$$

$$= \sum_i \sum_j \big[(y_{ij} - \bar{y}_{.j}) - \hat{\beta}'\,(x_{ij} - \bar{x}_{.j})\big]^2$$

$$= \sum_i \sum_j (y_{ij} - \bar{y}_{.j})^2 + \hat{\beta}'^2 \sum_i \sum_j (x_{ij} - \bar{x}_{.j})^2 - 2\,\hat{\beta}' \sum_i \sum_j (x_{ij} - \bar{x}_{.j})(y_{ij} - \bar{y}_{..})$$

$$= E_{yy}' + \left(\frac{E_{xy}'}{E_{xx}'}\right)^2 . E_{xx}' - 2\left(\frac{E_{xy}'}{E_{xx}'}\right)(E_{xy}') \qquad \text{[Using (6·132) and (6·133)]}$$

$$= E_{yy}' - \frac{E_{xy}'^2}{E_{xx}'} \qquad \text{... (6·134)}$$

$d.f.$ for $(SSE)^*$ = Total $d.f.$ – $d.f.$ for Blocks – $d.f.$ for β

$$= (vr - 1) - (r - 1) - 1 = vr - r - 1 = r\,(v - 1) - 1 \qquad \text{... (6·135)}$$

∴ Adjusted sum of squares for treatments ($SST = S_t^2$) is given by :

$$SST\,(S_t^2) = (SSE)^* - (SSE) \qquad \text{... (6·136)}$$

where (SSE) and $(SSE)^*$ are given in (6·122) and (6·134) respectively.

$d.f.\ (SST) = d.f.\ (SSE)^* - d.f.\ (SSE)$

$$= [r(v - 1) - 1] - [(r - 1)(v - 1) - 1] = v - 1 \qquad \text{[Using (6·135) and (6·123)]} \qquad \text{... (6·137)}$$

$$\therefore \quad MST(s_t^2) = \frac{SST}{d.f.} = \frac{S_t^2}{d.f.} = \frac{(SSE)^* - (SSE)}{v-1} \qquad \ldots (6\cdot138)$$

Hence the Test statistic for testing, $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_v$, is given by :

$$F = \frac{MST}{MSE} = \left[\frac{(SSE)^* - SSE}{(v-1)}\right] \times \frac{(r-1)(v-1)-1}{SSE} \sim F_{v-1,\,(r-1)(v-1)-1} \qquad \ldots (6\cdot139)$$

If $F > F_{v-1,\,(r-1)(v-1)-1}(\alpha)$ then H_0 is rejected at 'α' level of significance, otherwise we fail to reject H_0.

ANOCOVA Table for RBD. Let us write :

$$SS_{yy} = \text{Sum of squares due to } y = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$$

$$= \sum_i \sum_j \left[(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..})\right]^2$$

$$= \sum_i \sum_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 + r \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 + v \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2$$

$$= E_{yy} + T_{yy} + B_{yy} \qquad \ldots (6\cdot140)$$

where $T_{yy} = r \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2$ is the treatment $S.S.$ for y for RBD

$B_{yy} = v \sum_i (\bar{y}_{.j} - \bar{y}_{..})^2$ is the block S.S. for y for RBD

and $E_{yy} = \sum_i \sum_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$ is the error $S.S.$ for y for RBD.

Similarly, we have

$$SS_{xx} = \sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 = E_{xx} + T_{xx} + B_{xx} \qquad \ldots (6\cdot141)$$

$$SP_{xy} = \sum_i \sum_j (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..})$$

$$= \sum_i \sum_j \Big[\left\{(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}) + (\bar{x}_{i.} - \bar{x}_{..}) + (\bar{x}_{.j} - \bar{x}_{..})\right\}$$
$$\times \left\{(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..})\right\}\Big]$$

$$= \sum_i \sum_j (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})$$
$$+ \sum_i r(\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..}) + \sum_j v(\bar{x}_{.j} - \bar{x}_{..})(\bar{y}_{.j} - \bar{y}_{..}),$$

all other product terms will be zero.

$$\therefore \quad SP_{xy} = E_{xy} + T_{xy} + B_{xy} \qquad \ldots (6\cdot142)$$

where $T_{xy} = \sum_i r(\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..})$ and $B_{xy} = \sum_j v(\bar{x}_{.j} - \bar{x}_{..})(\bar{y}_{.j} - \bar{y}_{..}) \ldots (6\cdot142a)$

Results in (6·140), (6·141) and (6·142) give the partitioning of the total sum of squares due to y (SS_{yy}), the total sum of squares due to x (SS_{xx}) and the total sum of products (SP) of x and y (SP_{xy}).

Using the above notations, the above statistical analysis can be elegantly expressed in the ANOCOVA Table 6·26.)

TABLE 6·26 : ANOCOVA TABLE (RBD)

Sources of Variation	*d.f.*	*Sum of Squares and Products* SS_{xx}	SS_{yy}	SP_{xy}	*Estimate of β*	*Adjusted* SS_{yy}	*Adjusted d.f.*
Blocks	$r-1$	B_{xx}	B_{yy}	B_{xy}			
Treatments	$v-1$	T_{xx}	T_{yy}	T_{xy}			
Error	$(r-1)(v-1)-1$	E_{xx}	E_{yy}	E_{xy}	$\hat{\beta}=\dfrac{E_{xy}}{E_{xx}}$	*SSE*	$(r-1)(v-1)-1$
Treatment + Error	$r(v-1)$	E_{xx}'	E_{yy}'	E_{xy}'	$\hat{\beta}'=\dfrac{E_{xy}'}{E_{xx}'}$	SSE*	$r(v-1)-1$
Difference						(SSE)* − (SSE)	$v-1$

Note that $E_{xx}' = T_{xx} + E_{xx}$; $E_{yy}' = T_{yy} + E_{yy}$ and $E_{xy}' = T_{xy} + E_{xy}$ ... (6·143)

Example 6·8. *In an experiment on cotton with 5 manurial treatments, it was observed that the number of plants per plot is varying from plot to plot. The yields of cotton (Kapas) along with the number of plants per plot are given in Table 6·27. Analyse the yield data removing the effect of the variation in plant population on the yield by analysis of covariance technique and draw your conclusions. The design adopted was a R.B.D. with 4 replications.*

Treatments : 5 *Levels of Nitrogen :*

$$N_0 = 0,\ N_1 = 20,\ N_2 = 40,\ N_3 = 60 \text{ and } N_4 = 80\ kg/ha.$$

TABLE 6·27 : YIELD OF COTTON (NUMBER OF PLANTS) PER PLOT

Replicate I	N_1 12·0 (24)	N_0 10·5 (30)	N_4 27·0 (30)	N_2 16·5 (28)	N_3 25·0 (35)
Replicate II	N_3 26·0 (40)	N_2 20·0 (25)	N_0 12·0 (25)	N_4 26·0 (22)	N_1 15·5 (28)
Replicate III	N_2 22·0 (32)	N_4 30·0 (35)	N_3 20·0 (24)	N_1 20·0 (35)	N_0 14·5 (30)
Replicate IV	19·0(26)	18·5 (16)	8·5 (24)	29·0 (30)	25·0 (35)

Also obtain :

(*i*) *Average variance for the comparisons of treatment means, and*

(*ii*) *the gain in precision obtained on using ANOCOVA over RBD.*

Solution. We set up the following hypotheses :

Null Hypotheses.

$H_{0t} : \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5$, *i.e.*, the treatments are homogeneous.

$H_{0R} : b_1 = b_2 = b_3 = b_4$, *i.e.*, the blocks or replicates are homogeneous.

Alternative Hypotheses :

H_{1t} : At least two τ_i's are different.

H_{1b} : At least two b_j's are different.

We shall use the ANOCOVA technique to test these hypotheses.

y : Yield of cotton per plot;

x (Concomitant variable) : Number of plants per plot

TABLE 6·28 : CALCULATIONS FOR VARIOUS SUM OF SQUARES

Treatments	*Yield of Cotton (in kg.) along with the number of plants per plot*									
	Replication I		*Replication II*		*Replication III*		*Replication IV*		*Total*	
	x	y	x	y	x	y	x	y	x	y
N_0	30·0	10·5	25·0	12·0	30·0	14·5	24·0	8·5	109·0	45·5
N_1	24·0	12·0	28·0	15·5	35·0	20·0	26·0	19·0	113·0	66·5
N_2	28·0	16·5	25·0	20·0	32·0	22·0	35·0	25·0	120·0	83·5
N_3	35·0	25·0	40·0	26·0	24·0	20·0	16·0	18·5	115·0	89·5
N_4	30·0	27·0	22·0	26·0	35·0	30·0	30·0	29·0	117·0	112·0
Total	147·0	91·0	140·0	99·5	156·0	106·5	131·0	100·0	574·0 $= G(x)$	397·0 $= G(y)$

In the usual notations, we have

$$n = vr = 5 \times 4 = 20;\ G = \text{Grand total of all the observations.}$$

For x. $G = G(x) = 574{\cdot}0\ ;\ n = 20$

$$CF \text{ (Correction Factor)} = \frac{[G(x)]^2}{n} = \frac{(574)^2}{20} = \frac{3{,}29{,}476}{20} = 16{,}473{\cdot}80$$

$$RSS_{xx} = \sum_i \sum_j x_{ij}^2 = 30^2 + 25^2 + \ldots + 35^2 + 30^2$$

$$= 3{,}001 + 3{,}261 + 3{,}658 + 3{,}657 + 3{,}509 \text{ (Row-wise } S.S.)$$

$$= 17{,}086{\cdot}00$$

$$\text{Total } S.S.\ (SS_{xx}) = RSS_{xx} - CF(x) = 17{,}086{\cdot}00 - 16{,}473{\cdot}80 = 612{\cdot}20$$

$$R_{xx} = SS \text{ (Replications)} = \frac{147^2 + 140^2 + 156^2 + 131^2}{5} = CF(x) = \frac{82{,}706}{5} - CF(x)$$

$$= 16{,}541{\cdot}20 - 16{,}473{\cdot}80 = 67{\cdot}40$$

$$T_{xx} = SS \text{ (Treatments)} = \frac{109^2 + 113^2 + 120^2 + 115^2 + 117^2}{4} - CF(x)$$

$$= \frac{65{,}964}{4} - C.F. = 16{,}491{\cdot}00 - 16{,}473{\cdot}80 = 17{\cdot}20$$

$$E_{xx} = SS(\text{Error}) = T_{xx} - R_{xx} - T_{xx} = 612{\cdot}20 - 67{\cdot}40 - 17{\cdot}20 = 527{\cdot}60$$

For y.

$$CF(y) = \frac{[G(y)]^2}{n} = \frac{(397{\cdot}0)^2}{20} = \frac{1{,}57{,}609}{20} = 7{,}880{\cdot}45$$

$$RSS_{yy} = \sum_i \sum_j y_{ij}^2 = (10{\cdot}5)^2 + (12{\cdot}0)^2 + \ldots + (30)^2 + (29)^2$$

$$= 536{\cdot}75 + 1{,}145{\cdot}25 + 1{,}781{\cdot}25 + 2{,}043{\cdot}25 + 3{,}146{\cdot}00 \text{ (Row-wise } S.S.)$$

$$= 8{,}652{\cdot}50$$

$$\text{Total } SS\ (SS_{yy}) = RSS_{yy} - CF(y) = 8{,}652{\cdot}50 - 7{,}880{\cdot}45 = 772{\cdot}05$$

$$R_{yy} = SS(\text{Replicates}) = \frac{(91{\cdot}0)^2 + (99{\cdot}5)^2 + (106{\cdot}5)^2 + (100{\cdot}0)^2}{5} - CF(y)$$

$$= \frac{39{,}523}{5} - CF(y) = 7{,}904{\cdot}70 - 7{,}880{\cdot}45 = 24{\cdot}25$$

$$T_{yy} = SS\ (\text{Treatments}) = \frac{(45{\cdot}5)^2 + (66{\cdot}5)^2 + (83{\cdot}5)^2 + (89{\cdot}5)^2 + (112{\cdot}0)^2}{4} - CF(y)$$

$$= \frac{34{,}019}{4} - CF(y) = 8{,}504{\cdot}75 - 7{,}880{\cdot}45 = 624{\cdot}30$$

$$E_{yy} = SS(\text{Error}) = SS_{yy} - R_{yy} - T_{yy} = 772{\cdot}05 - 24{\cdot}25 - 624{\cdot}30 = 123{\cdot}50$$

For Product *xy*. $S.P. \equiv$ Sum of Products

$$CF(xy) = \frac{G(x)\ G(y)}{n} = \frac{(574{\cdot}0) \times (397{\cdot}0)}{20} = \frac{2{,}27{,}878}{20} = 11{,}393{\cdot}90$$

$$RSS\ (\text{Products}) = RSS_{xy} = \sum\sum xy = (30{\cdot}0 \times 10{\cdot}5) + (25{\cdot}0) \times (12{\cdot}0)$$

$$+ \ldots + (35{\cdot}0) \times (30{\cdot}0) + (30{\cdot}0) \times 29{,}0)$$

$$= 1{,}254 + 1{,}916 + 2{,}541 + 2{,}691 + 3{,}302$$ [Row-wise sum of products *xy*]

$$= 11{,}704{\cdot}00$$

$$\text{Total } SP(xy) = RSS(\text{Products}) - CF\ (xy) = 11{,}704{\cdot}00 - 11{,}393{\cdot}90 = 310{\cdot}10$$

$$R_{xy} = SP\ (xy) \text{ for Replicates}$$

$$= \frac{1}{5}\left[(147{\cdot}0 \times 91{\cdot}0) + (140{\cdot}0 \times 99{\cdot}5) + (156{\cdot}0 \times 106{\cdot}5) + (131{\cdot}0 \times 100{\cdot}0)\right] - CF\ (xy)$$

$$= \frac{57{,}021{\cdot}0}{5} - CF\ (xy) = 11{,}404{\cdot}20 - 11{,}393{\cdot}90 = 10{\cdot}30$$

$$T_{xy} = SP(xy) \text{ for Treatments}$$

$$= \frac{1}{4}\left[(109{\cdot}0 \times 45{\cdot}5) + (113{\cdot}0 \times 66{\cdot}5) + (120{\cdot}0 \times 83{\cdot}5)\right.$$

$$\left. + (115{\cdot}0 \times 89{\cdot}5) + (117{\cdot}0 \times 112{\cdot}0)\right] - CF\ (xy)$$

$$= \frac{45{,}890{\cdot}5}{4} - CF(xy) = 11{,}472{\cdot}63 - 11{,}393{\cdot}90 = 78{\cdot}73$$

$$E_{xy} = SP(xy) \text{ for Error} = \text{Total } SP(xy) - R_{xy} - T_{xy}$$

$$= 310{\cdot}10 - 10{\cdot}30 - 78{\cdot}73 = 221{\cdot}07$$

TABLE 6·29 : SUM OF SQUARES AND PRODUCTS

Source of Variation	*d.f.*	*SS(x²)*	*SS(xy)*	*SS y²)*	*MS(yy)*	*F(yy)*
(1)	(2)	(3)	(4)	(5)	(6) = (5) ÷ (2)	
Replications	4 − 1 = 3	$R_{xx} = 67{\cdot}40$	$R_{xy} = 10{\cdot}30$	$R_{yy} = 24{\cdot}25$	8·08	
Treatments	5 − 1 = 4	$T_{xx} = 17{\cdot}20$	$T_{xy} = 78{\cdot}73$	$T_{yy} = 624{\cdot}30$	156·08	$\frac{156{\cdot}08}{10{\cdot}29} = 15{\cdot}17$
Error	3 × 4 = 12	$E_{xx} = 527{\cdot}60$	$E_{xy} = 221{\cdot}07$	$E_{yy} = 123{\cdot}50$	10·29	
Total	20 − 1 = 19	$SS_{xx} = 612{\cdot}20$	$S_{xy} = 310{\cdot}10$	$SS_{yy} = 772{\cdot}05$		

We shall now adjust for variation in yield (y) from plot to plot for the linear (regression) effect of the number of plants (x) per plot. An estimate of the coefficient of regression (β) of y on x is given by :

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}} = \frac{221\cdot07}{527\cdot60} = 0\cdot4190 \simeq 0\cdot42$$

The adjusted (corrected) error sum of squares for y, adjusted for this linear effect is given by :

$$\text{Adjusted Error } S.S. \text{ for } y = \text{Adjusted } (E_{yy}) = E_{yy} - \hat{\beta} E_{xy} = E_{yy} - \frac{(E_{xy})^2}{E_{xx}}$$

$$= 123\cdot50 - \frac{(221\cdot07)^2}{527\cdot60} = 123\cdot50 - 92\cdot63 = 30\cdot87$$

$$(OR : \hat{\beta} E_{xy} = 0\cdot4190 \times 221\cdot07 = 92\cdot63)$$

The estimation of $\hat{\beta}$ results in loss of 1 *d.f.* for error sum of squares, which now becomes $12 - 1 = 11$.

Next, the variation in treatments is also to be adjusted for variation in x. For this, we prepare the following *Table 6·30,* for (Treatments + Error) sum of squares.

TABLE 6·30 : *S.S AND S.P.* FOR (TREATMENTS + ERROR)

Source of Variation	$SS(x^2)$	$SP(xy)$	$SS(y^2)$
Treatments	$T_{xx} = 17\cdot20$	$T_{xy} = 78\cdot73$	$T_{yy} = 624\cdot30$
Error	$E_{xx} = 527\cdot60$	$E_{xy} = 221\cdot07$	$E_{yy} = 123\cdot50$
Treatments + Error	$E_{xx}' = 544\cdot80$	$E_{xy}' = 299\cdot80$	$E_{yy}' = 747\cdot80$

$$E_{xx}' = T_{xx} + E_{xx} \,;\; E_{xy}' = T_{xy} + E_{xy} \;;\; E_{yy}' = T_{yy} + E_{yy}$$

The *S.S.* for (Treatments + Error) for y is adjusted for linear (regression) effect of x on y exactly similarly as the error *S.S.* and is given by :

$$SSE^* = \text{Adjusted } S.S. \text{ for (Treatments + Error) for } y = E_{yy}' - \frac{(E_{xy}')^2}{E_{xx}'}$$

$$= 747\cdot80 - \frac{(299\cdot80)^2}{544\cdot80} = 947\cdot80 - 164\cdot98 = 582\cdot82$$

Finally, the 'Treatment *S.S.*' for y adjusted for the linear effect of x on y is given by :

Adjusted (Treatment *S.S.*) for $y = SSE^* -$ Adjusted (E_{yy})

= Adjusted (Treatment + Error) *S.S.* for y – Adjusted Error *S.S.* for y

$= 582\cdot82 - 30\cdot87 = 551\cdot95$

TABLE 6·31 : ADJUSTED ANALYSIS OF VARIANCE TABLE

Source of Variation	*d.f.*	*S.S.*	*M.S.S.*	*Variance Ratio*
(1)	*(2)*	*(3)*	*(4) = (3) ÷ (2)*	*(F)*
Treatments	4	551·95	137·99	$F_T = \frac{137\cdot99}{2\cdot81} = 49\cdot11^{**}$
Error	$12 - 1 = 11$	30·87	2·81	
Treatment + Error	$16 - 1 = 15$	582·82		

** Highly significant

Conclusion. Tabulated $F_{4, 11}$ (0·05) = 3·36.

Since the calculated value F_T = 49·11 is much grater than the tabulated (critical) value, it is highly significant. Hence, we reject the null hypothesis of equality of treatment means and conclude that the treatments differ significantly as regards their effect on increase of yield of cotton. Moreover, from *Table 6·28*, we conclude that the treatment N_4 is the most effective, followed by N_3, N_2, N_1 and N_0 respectively.

(*i*) **Comparing two Treatment Means.** Since the yields from different treatments differ significantly, we proceed to study which pairs of treatment means differ significantly. For this, we shall first adjust the treatment mean yields for the regression of yields (y) on the number of plants (x) per plot. The adjusted mean yield ($\overline{Y}_i$) for the ith treatment is given by :

$$\overline{Y}_i = \bar{y}_i - \hat{\beta}(\bar{x}_i - \bar{x}), \qquad \dots (6\cdot144)$$

where $\bar{x}_i$ and $\bar{y}_i$ are the observed mean values of the number of plants and the plot yield respectively for the ith treatment, so that

$$\bar{x}_i = \frac{1}{r}\sum_{j=1}^{r} x_{ij} = \frac{x_{i\cdot}}{4} \quad ; \quad \bar{y}_i = \frac{1}{r}\sum_{j=1}^{r} y_{ij} = \frac{y_{i\cdot}}{4}$$

and

$$\bar{x} = \frac{1}{n}\sum_i \sum_j x_{ij} = \frac{G(x)}{n} = \frac{574}{20} = 28\cdot7$$

Substituting in (6·144), we get

$$\overline{Y}_i = \bar{y}_i - 0\cdot42\,(\bar{x}_i - 28\cdot7) \qquad \dots (*)$$

TABLE 6·32 : ADJUSTED MEAN YIELDS

Treatment	*Average Plant No.* $\bar{x}_i$	$\bar{x}_i - \bar{x}$ $= \bar{x}_i - 28\cdot7$	$\hat{\beta}(\bar{x}_i - \bar{x})$ $= 0\cdot42\,(\bar{x}_i - 28\cdot7)$	*Average yield (Un-adjusted)* ($\bar{y}_i$)	*Adjusted mean* $\overline{Y}_i$ $= \bar{y}_i - \hat{\beta}(\bar{x}_i - 28\cdot7)$
$1 : N_0$	$\frac{109}{4} = 27\cdot25$	− 1·45	− 0·61	$\frac{45\cdot5}{4} = 11\cdot38$	$\overline{Y}_1 = 11\cdot99$
$2 : N_1$	$\frac{113}{4} = 28\cdot25$	− 0·45	− 0·19	$\frac{66\cdot5}{4} = 16\cdot63$	$\overline{Y}_2 = 16\cdot82$
$3 : N_2$	$\frac{120}{4} = 30\cdot00$	1·30	0·55	$\frac{83\cdot5}{4} = 20\cdot88$	$\overline{Y}_3 = 20\cdot33$
$4 : N_3$	$\frac{115}{4} = 28\cdot75$	0·05	0·02	$\frac{89\cdot5}{4} = 22\cdot38$	$\overline{Y}_4 = 22\cdot36$
$5 : N_4$	$\frac{117}{4} = 29\cdot25$	0·55	0·23	$\frac{112}{4} = 28\cdot00$	$\overline{Y}_5 = 27\cdot77$

Next we have to obtain the standard error (*S.E.*) of the difference between pair of adjusted treatment means. ***Finney*** proved that the *average S.E.* of the difference between a pair of two treatment means is given by :

$$S.E.\,(\overline{Y}_i - \overline{Y}_j) = \left[\frac{2s_E'^2}{r}\left(1 + \frac{M.S.S. \text{ due to treatments for } x}{E_{xx}}\right)\right]^{1/2} = \left[\frac{2s_{E'}^2}{r}\left(1 + \frac{T_{xx}}{(v-1)E_{xx}}\right)\right]^{1/2} \qquad \dots (6\cdot145)$$

where $s_{E'}^2$ = Adjusted *M.S.S.* due to error (*Table 6·31*)

and r = Number of replications ; v = Number of treatments.

Substituting the values from *Tables 6·31 and 6·29*, we get

$$S.E.\ (\overline{Y}_i - \overline{Y}_j) = \left[\frac{2 \times 2{\cdot}81}{4}\left(1 + \frac{17{\cdot}20}{4 \times 527{\cdot}60}\right)\right]^{1/2} = \left[1{\cdot}405\left(1 + \frac{4{\cdot}30}{527{\cdot}60}\right)\right]^{1/2}$$

$$= \sqrt{1{\cdot}405 \times 1{\cdot}0081} = \sqrt{1{\cdot}4164} = 1{\cdot}19 \qquad \ldots (**)$$

∴ Average variance for comparing two treatment means = 1·4164.

Critical difference (*C.D.*) between two adjusted treatment means is given by :

$$C.D.\ (\overline{Y}_i - \overline{Y}_j) = S.E.\ (\overline{Y}_i - \overline{Y}_j) \times t_{0 \cdot 025} \text{ (for adjusted error } d.f.)$$

$$= 1{\cdot}19 \times t_{11}\ (0{\cdot}025) = 1{\cdot}19 \times 2{\cdot}201 = 2{\cdot}62$$

Two treatment means will differ significantly if :

$$\overline{Y}_i - \overline{Y}_j > C.D.\ (\overline{Y}_i - \overline{Y}_j) = 2{\cdot}62.$$

From *Table 6·32*, we conclude as follows :

1. Treatment 1 (N_0) differs significantly from each of the treatments 2, 3, 4 and 5.
2. Treatment 2 (N_1) differs significantly from all other treatments.
3. Treatment 3 (N_2) differs significantly from treatments 1, 2 and 5.
4. Treatment 4 (N_3) differs significantly from treatments 1, 2 and 5.
5. Treatment 5 (N_4) differs significantly from the treatments 1, 2, 3 and 4.
6. Treatments 3 and 4 do not differ significantly.
7. Since the average yield $\overline{Y}_5$ is maximum, treatment 5(N_4) is the most effective in increasing the yield of crop.

(*ii*) ***Efficiency of ANOCOVA over Simple R.B.D.***

The variance of the difference between two treatment means without adjusting for the linear effect of the concomitant variable x, *i.e.*, for simple *R.B.D.* is given by :

$$V_1 = V(\overline{y}_i - \overline{y}_j) = \frac{2\, s_E^2 \text{ (for } y)}{r} = \frac{2\, E_{yy}}{r\,(r-1)\,(v-1)} = \frac{2}{4} \times 10{\cdot}29 = 5{\cdot}145$$

[From Table 6·29]

(*i*) The average variance (V_2) of difference between two adjusted treatment means, adjusted for the linear effect of x on y is given by :

$$V_2 = V(\overline{Y}_i - \overline{Y}_j) = 1{\cdot}4164 \qquad \text{[From (*)]}$$

(*ii*) Efficiency of ANOCOVA in *RBD* over simple *RBD* is given by :

$$\frac{V_1}{V_2} = \frac{5{\cdot}1450}{1{\cdot}4164} = 3{\cdot}6324$$

Percentage gain in efficiency = (3·6324 – 1) × 100% = 263·24% ≃ 263%

The gain in precision is 2·63 times.

Source :

1. S.C. Gupta and V.K. Kapoor : Fundamental of Applied Statistics – Sultan Chand & Sons, Fourth Edition, 2015.
2. Panneer Selvam: Design And Analysis of Experiments, Prentice Hall.