# Analysis of Covariance (ANOCOVA)

## Definition:

The statistical technique which reduces the experimental error by eliminating the effect of variation in the ancillary or concomitant variate, resulting in increase in the precision of the estimates of treatment means in known as Analysis of Co-variance.

## Ancillary observations

Ancillary observations are values of the characteristics of those extraneous sources of variation which influence the characteristic that is compared for each treatment. These sources of variation cannot be controlled in the experiment and vary from one experimental unit to another experimental unit, but can be measured numerically in each experimental unit. Such type of variable or characteristic is called ancillary variable or concomitant variate or Co-variate.

**Example:** In the experiment conducted to compare different levels of nitrogen on the yield of wheat. The ancillary variable in this experiment may be plant population, number of tillers, age of wheat crop or straw yield per plot etc., which are not practically controlled and vary randomly from plot to plot. Each of these ancillary observations influence directly the yield of wheat in each plot.

## Statistical analysis of ANOCOVA for CRD

Let the number of treatmentsbe t (say $v_1, v_2, v_3, \ldots v_t$) and the number of replications for treatment $v_i$ be $r_i$. Also let y denote the main or response variate (say yield ) and x denote the corresponding value (for the same experimental unit) of the ancillary variate (say plant population).

The data from the original layout can be arranged in the following tabular form :

| Replication | Treatments | | | | | | |
|---|---|---|---|---|---|---|---|
| | $V_1$ | | $V_2$ | | ... | $V_t$ | |
| | y | x | y | x | | y | x |
| 1 | $y_{11}$ | $x_{11}$ | $y_{21}$ | $x_{21}$ | ... | $y_{t1}$ | $x_{t1}$ |
| 2 | $y_{12}$ | $x_{12}$ | $y_{22}$ | $x_{22}$ | ... | $y_{t2}$ | $x_{t2}$ |
| . | | | | | | . | . |
| . | | | | | | . | . |
| . | | | | | | . | . |
| $r_i$ | $y_{1r_1}$ | $x_{1r_1}$ | $y_{2r_2}$ | $x_{2r_2}$ | ... | $y_{tr_t}$ | $x_{tr_t}$ |
| Total | $T_{1(y)}$ | $T_{1(x)}$ | $T_{2(y)}$ | $T_{2(x)}$ | ... | $T_{t(y)}$ | $T_{t(x)}$ |

Where $G_y = T_{1(y)} + T_{2(y)} + \ldots + T_{t(y)} = \sum_{i=1}^{t} T_{i(y)}$

And $G_x = T_{1(x)} + T_{2(x)} + \ldots + T_{t(x)} = \sum_{i=1}^{t} T_{i(x)}$

Where $y_{ij}$ represents the value of y corresponding to $i^{th}$ treatment and $j^{th}$ replication. $i = 1, 2, \ldots t$

and $j = 1, 2, \ldots, r_i$

**a) Computation of sum of squares for the main variate y:**

i) $C.F = \frac{G^2_Y}{N}$

ii) $T.S.S._Y = \Sigma\Sigma y_{ij}^2 - C.F$

iii) Treatment Sum of Square $= \sum \frac{T^2_{i(y)}}{r_i} - C.F = A_1$ (say)

iv) Error Sum of Square

$E.S.S._Y = T.S.S._Y - T_r.S.S._Y = A$ (say)

**b) Computation of Sum of Squares for the ancillary Variate x :**

i) $C.F = \frac{G^2_x}{N}$

ii) $T.S.S._x = \sum\sum x^2_{ij} - C.F$

iii) Treatment Sum of Square $= \sum \frac{T^2_{i(x)}}{r_i} - C.F = B_1$ (say)

iv) Error Sum of Square

$E.S.S._x = T.S.S._x - T_r.S.S_x = B$ (say)

**c) Computation of Sum of products of x and y:**

i) $C.F = \frac{G_x G_y}{N}$

ii) Total Sum of Product of xy

$$T.S.P_{xy} = \sum\sum[(x_{ij}y_{ij}]) - C.F_{xy}$$

iii) Treatment sum of Product of xy

$$Tr.S.P_{xy} = \sum\left(\frac{T_{i(x)}.T_{i(y)}}{r_i}\right) - C.F_{xy} = C_1(say)$$

iv) Error sum of product of xy

$E.S.P_{xy} = T.S.P_{xy} - Tr.S.P_{xy} = C$ (say)

Now we summaries the above data in the following table

| Sources of variation | Degrees of freedom | Sum of square for x | Sum of square for y | Sum of product of xy | Regression co-efficient of Y on X ($b_{yx}$) |
|---|---|---|---|---|---|
| Treatments | $t-1=n_1$(say) | $B_1$ | $A_1$ | $C_1$ | $F=\left[\frac{C^2/B}{A-C^2/B}\right](n-1)$ |
| Error | $N-t=n$ | B | B | C | |
| Total | N-1 | $B_1+B$ | $A_1+A$ | $C_1+C$ | |

To compare the value of F with $F_{(1,n-1)}$ d.f ($\alpha$ )

i) If the value of F is not significant then we do not give any consideration to the ancillary variable and ignore it and compare the treatment directly by making use of the ANOVA technique.

ii) If the value of F is significant then it is concluded that the ancillary variable x influences the main or response variable y. In order to eliminate this effect of x on y we make use of the technique of analysis of covariance as explained below.

We calculated the following adjusted sum of squares:

i) Adjusted T.S.S. for y $= (A_1 + A) - \frac{(C_1 + C)^2}{(B_1 + B)} = S_2$ (say)

ii) Adjusted E.S.S for y $= A - \frac{C^2}{B} = S$(say)

iii) Adjusted Tr.S.S for y = adjusted $T.S.S_Y$ – adjusted E.S.S

$= S_2 - S = S_1$, (say)

ANOVA (for adjusted $S.S._Y$)

| S.V | d.f | Adjusted $S.S._Y$ | M.S.S | F- ratio | F - table |
|---|---|---|---|---|---|
| Treatment | $t-1 = n_1$ | $S_1$ | $S_1/n_1=T$ | $F = \frac{T}{E}$ | $F_{(n_1, n-1)d.f}(\alpha)$ |
| Error | $n-1$ | $S$ | $S/n-1 = E$ | | |
| Total | $N-2$ | $S_2$ | | | |