Partitioning Sums of Squares in ANOVA

George H Olson, Ph. D.
Leadership and Educational Studies
Appalachian State University
(Fall 2010)

In this brief paper, I show how the total sums of squares (*SS*) for variable, Y_{ij} can be partitioned into two sources, sums of squares between groups (SS_{B}) and sums of squares within groups (SS_{W}).

We begin with deviation scores, $(Y_{ij} - \bar{Y}_{..})$, where Y_{ij} is the score for an individual, *i*, in group *j*. Next, square and sum the deviation scores over all individuals, over all groups:

$$\begin{aligned}\sum_j \sum_i (Y_{ij} - \bar{Y}_{..})^2 &= \sum_j \sum_i [(Y_{ij} - \bar{Y}_{.j}) + (\bar{Y}_{.j} - \bar{Y}_{..})]^2 \\ &= \sum_j \sum_i (Y_{ij} - \bar{Y}_{.j})^2 + \sum_j \sum_i (\bar{Y}_{.j} - \bar{Y}_{..})^2 \\ &\quad + 2\sum_j \sum_i (Y_{ij} - \bar{Y}_{.j})(\bar{Y}_{.j} - \bar{Y}_{..})\end{aligned}$$

Note, however, that

$$2\sum_j \sum_i (Y_{ij} - \bar{Y}_{.j})(\bar{Y}_{.j} - \bar{Y}_{..}) = 2\sum_j (\bar{Y}_{.j} - \bar{Y}_{..}) \sum_i (Y_{ij} - \bar{Y}_{.j})$$

since $(Y_{ij} - \bar{Y}_{.j})$ is the same for all individuals, *i*, in group *j*.

It might be easier to consider the expression,

$$2\sum_j \sum_i (y_{ij})(\bar{y}_{.j}) = 2\sum_j (\bar{y}_{.j}) \sum_i (y_{.j}),$$

where $y_{ij} = (Y_{ij} - \bar{Y}_{.j})$ is the deviation score for individuals in Group *j*, and $\bar{y}_{.j}$ is the deviation of $\bar{y}$ for Group *j* from the over all mean.

Furthermore,

$$\sum_j \sum_i (\bar{Y}_{.j} - \bar{Y}_{..})^2 = \sum_j n_j (\bar{Y}_{.j} - \bar{Y}_{..})^2$$

because, within each group, j, $(\bar{Y}_{.j} - \bar{Y}_{..})$ is constant (i.e., $\sum c = nc$.)

Hence,

$$\sum_j \sum_i (Y_{ij} - \overline{Y}_{\bullet\bullet})^2 = \sum_j \sum_i (Y_{ij} - \overline{Y}_{\bullet j})^2 + \sum_j n_j (\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})^2 .$$

The first term on the right, $\sum_j \sum_i (Y_{ij} - \overline{Y}_{\bullet j})^2$, is called the *within group* sums of squares (it is the sum of squared deviations of all individuals from the individual's respective group mean.) The second term on the right, $\sum_j n_j (\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})^2$, is called the *between groups* sums of squares (it is equal to the sum of squared deviations of each group mean from the overall mean.)