INSY 7330 M2009 By S. Maghsoodloo

The Breakdown of Total Sum of Squares into Model and Error SS's for a 2-Factor Factorial Design with n_{ij} Responses in the ij^th Cell

Consider a 2-factor factorial design with factor A at *a* levels, factor B at *b* levels and n_{ij} observations at the ij^th cell of the *ab* FLCs (Factor Level Combinations) of the experiment. When n_{ij}'s are all equal, then the design is said to be balanced.

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(y_{ijk}-\bar{y}_{...})^2 \equiv \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}[(y_{ijk}-\bar{y}_{ij.})+(\bar{y}_{ij.}-\bar{y}_{...})]^2$$

$$\equiv \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(\bar{y}_{ij.}-\bar{y}_{...})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(y_{ijk}-\bar{y}_{ij.})^2$$

Because the cross-product term is identically equal to zero. As a result, we have

$$SS_T \equiv \sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}(\bar{y}_{ij.}-\bar{y}_{...})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(y_{ijk}-\bar{y}_{ij.})^2 = SS_{Model} + SS_{Error}$$

where $SS_{Model} = \sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}(\bar{y}_{ij.}-\bar{y}_{...})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}(\bar{y}_{ij.}^{2}-2\bar{y}_{ij.}\bar{y}_{...}+\bar{y}_{...}^{2})$

$$=\sum_{i=1}^{a}\sum_{j=1}^{b}[n_{ij}(y_{ij.}/n_{ij})^2-2y_{ij.}\bar{y}_{...}+n_{ij}\bar{y}_{...}^{2}]$$

$$=\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij.}^{2}/n_{ij}-2\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij.}\bar{y}_{...}+\sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}\bar{y}_{...}^{2}$$

$$=\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij.}^{2}/n_{ij}-2y_{...}\bar{y}_{...}+\bar{y}_{...}^{2}\sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}\text{, where }\sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij} = N.$$

Hence,

$$SS_{Model} = \sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij.}^{2}/n_{ij}-2y_{...}^{2}/N+N\bar{y}_{...}^{2} = \sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij.}^{2}/n_{ij}-2CF+CF$$

SS$_{Model}$ = $\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij.}^2 / n_{ij} - CF$, where CF = $y_{...}\bar{y}_{...}$ = $N\bar{y}_{...}^2$ = $y_{...}^2 / N$.

When the design is unbalanced, the SST always does break down into two additive components, namely SS$_{Model}$ and SS$_{Error}$ = $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij.})^2$.

However, in the unbalanced case, SS$_{Model}$ does not break down into 3 additive components SS$_A$, SS$_B$ and SS$_{A\times B}$ because

SS$_{Model}$ = $\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}(\bar{y}_{ij.} - \bar{y}_{...})^2$ =

$$= \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}[(\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})]^2$$

However, if the design is unbalanced, then the 3 cross-product terms do not vanish, i.e., the cross-product term such as

$$2\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}[(\bar{y}_{i..} - \bar{y}_{...})(\bar{y}_{.j.} - \bar{y}_{...})]$$

is not zero.

While when all n_{ij}'s = n, then all the 3 cross-product terms are zero such that in the balanced case

SS$_{Model}$ = SS$_A$ + SS$_B$ + SS$_{A\times B}$

where SS$_A$ = $n\sum_{i=1}^{a}\sum_{j=1}^{b} (\bar{y}_{i..} - \bar{y}_{...})^2$ = $nb\sum_{i=1}^{a} (\bar{y}_{i..} - \bar{y}_{...})^2$, SS$_B$= $na\sum_{j=1}^{b} (\bar{y}_{.j.} - \bar{y}_{...})^2$, and

SS$_{A\times B}$ = $n\sum_{i=1}^{a}\sum_{j=1}^{b} (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$.

When the design is an unbalanced factorial, the GLM (General Linear Models) approach has to be used to obtain an ANOVA table. Fortunately, Minitab provides GLM to analyze data from an unbalanced design.