Factorial Design

Peter Yeh

12/19/2022

Introduction

A study design meant to study the effects of two or more factors. The effect of a factor is the change in response produced by a change in the level of the factor. When all possible combinations of the levels of the factors are investigated, the design is crossed.

Factor Effects

$A = \bar{y}_{A^+} - \bar{y}_{A^-}$

$B = \bar{y}_{B^+} - \bar{y}_{B^-}$

$AB =$ difference in diagonal averages

General Design

F_1, F_2, $\ldots$, F_r factors with number of levels l_1, l_2, $\ldots$, l_r. The number of all possible level combinations (treatments): l_1 x l_2 x $\ldots$ x l_r. We are interested in (main) effects, 2-factor interactions, 3-factor interactions, etc.

Two-factor Factorial Design

Statistical Model: Effects

$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$, $i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$; $k = 1, 2, \ldots, n$

$\mu =$ grand mean

$\tau_i = i$th level effect of factor A (ignores B)(main effects of A); $\sum_i \tau_i = 0$

$\beta_j = j$th level effect of factor B (ignores A)(main effects of B); $\sum_j \beta_j = 0$

$(\tau\beta)_{ij} =$ interaction effect of combination ij, (explains variation not explained by main effects), $\sum_i (\tau\beta)_{ij}$ $= \sum_j (\tau\beta)_{ij} = 0$

$\epsilon_{ijk} \sim N(0, \sigma^2)$

Estimates

$\hat{\mu} = \bar{y}_{...} = \frac{y_{...}}{abn}$

$\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...} = \frac{y_{i..}}{bn} - \frac{y_{...}}{abn}$

$\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...} = \frac{y_{.j.}}{an} - \frac{y_{...}}{abn}$

$(\hat{\tau\beta})_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} = \frac{y_{ij.}}{n} - \frac{y_{i..}}{bn} - \frac{y_{.j.}}{an} + \frac{y_{...}}{abn}$

Predicted value at level combination ij: $\hat{y}_{ijk} = \bar{y}_{ij.}$

Residuals: $\hat{\epsilon}_{ijk} = y_{ijk} - \bar{y}_{ij.}$

Sum of Squares

$SS_A = bn \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 = \frac{1}{bn} \sum_{i=1}^{a} y_{i..}^2 - \frac{y_{...}^2}{abn}$

$SS_B = an \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 = \frac{1}{an} \sum_{j=1}^{b} y_{.j.}^2 - \frac{y_{...}^2}{abn}$

$SS_{Subtotals} : \frac{1}{n} \sum_i^a \sum_j^b y_{ij.}^2 - \frac{y_{...}^2}{abn}$

$SS_{AB} = SS_{Subtotals} - SS_A - SS_B = n \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$

$SS_E = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2$

$SS_T = \sum (y_{ijk} - \bar{y}_{...})^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn}$

Expected Mean Squares

$E(MS_E) = \sigma^2$

$E(MS_A) = \sigma^2 + bn \sum \tau_i^2/(a-1)$

$E(MS_B) = \sigma^2 + an \sum \beta_j^2/(b-1)$

$E(MS_{AB}) = \sigma^2 + n \sum \sum (\tau\beta)_{ij}^2/(a-1)(b-1)$

Hypotheses Tests

1) Main effects of A: $H_0 : \tau_1 = ... = \tau_a = 0$ vs. H_1 : at least one $\tau_i \neq 0$
2) Main effects of B: $H_0 : \beta_1 = ... = \beta_b = 0$ vs. H_1 : at least one $\beta_j \neq 0$
3) Interaction effects of AB: $H_0 : (\tau\beta)_{ij} = 0$ for all i, j vs. H_1 : at least one $(\tau\beta)_{ij} \neq 0$

Note:

$df_E > 0$ only if the number of replicates (n) > 1. When n = 1, no SS_E is available and we cannot test the effects.

If we can assume that the interactions are negligible, MS_{AB} becomes a good estimate of σ^2 and can be used as MS_E. Need to check for this assumption.

■ **TABLE 5.3**
The Analysis of Variance Table for the Two-Factor Factorial, Fixed Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
A treatments	SS_A	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$F_0 = \frac{MS_A}{MS_E}$
B treatments	SS_B	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$F_0 = \frac{MS_B}{MS_E}$
Interaction	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
Error	SS_E	$ab(n-1)$	$MS_E = \frac{SS_E}{ab(n-1)}$	
Total	SS_T	$abn-1$		

Figure 1: Two-Factor Factorial Table

Checking Assumptions

1) Errors are normally distributed: QQ plot of residuals
2) Constant Variance: residuals vs. ($\hat{y}_{ij.}$, factor A, and factor B)
3) If n=1, no interaction: check with $pred^2/SS3$, add in the model to see if interaction can be present

Tukey's Test of Nonadditivity Assumes $(\tau\beta)_{ij} = \gamma\tau_i\beta_j$ and $H_0 : \gamma = 0$

Two-Factor Analysis with No Replicates

$SS_N = \frac{[\sum\sum y_{ij}y_{i.}y_{.j} - y_{..}(SS_A + SS_B + y_{..}^2/ab)]^2}{abSS_ASS_B}$

$SS_E = SS_{Residual} - SS_N$

$F_0 = \frac{SS_N/1}{(SS_E - SS_N)/((a-1)(b-1)-1)} \sim F_{1,(a-1)(b-a)-1}$

Extra Notes

Statistical Model: Means

$y_{ijk} = \mu_{ij} + \epsilon_{ijk}$, $i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$; $k = 1, 2, \ldots, n$

With $\mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij}$

■ **TABLE 5.9**

Analysis of Variance for a Two-Factor Model, One Observation per Cell

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Expected Mean Square
Rows (A)	$\sum_{i=1}^{a} \frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab}$	$a - 1$	MS_A	$\sigma^2 + \frac{b \sum \tau_i^2}{a - 1}$
Columns (B)	$\sum_{j=1}^{b} \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab}$	$b - 1$	MS_B	$\sigma^2 + \frac{a \sum \beta_j^2}{b - 1}$
Residual or AB	Subtraction	$(a - 1)(b - 1)$	MS_{Residual}	$\sigma^2 + \frac{\sum \sum (\tau\beta)_{ij}^2}{(a - 1)(b - 1)}$
Total	$\sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{ab}$	$ab - 1$		

Figure 2: Two-Factor with No Replicates Table