Two Level Factorial Design

Peter Yeh

12/22/2022

Introduction

Also written as 2^k factorial design. It involves k factors which has two levels each (+ and -).

Factor Effect Estimation

$A = \bar{y}_{A^+} - \bar{y}_{A^-} = \frac{1}{2n}[ab + a - b - 1]$

$B = \bar{y}_{B^+} - \bar{y}_{B^-} = \frac{1}{2n}[ab + b - a - 1]$

$AB =$ difference in diagonal averages $= \frac{1}{2n}[ab + 1 - a - b]$

The portions in the brackets is called the total effect of the factor, or the factors' contrast.

There is a one-to-one correspondence between effects and contrasts:

effect $= \frac{1}{2}\sum c_i\bar{y}_i$

Sum of Squares

$SS_{Contrast} = \frac{(\sum c_i\bar{y}_i)^2}{\sum c_i^2/n}$, where the number of c is 2^k

Or,

$SS_A = \frac{[ab+a-b-1]^2}{4n}$ with $df = 1$

$SS_B = \frac{[ab+b-a-1]^2}{4n}$ with $df = 1$

$SS_AB = \frac{[ab+1-a-b]^2}{4n}$ with $df = 1$

$SS_E = SS_T - SS_A - SS_B - SS_{AB}$ with $df = N - 4 = abn - 4$

$SS_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{4n}$ with $df = N - 1 = abn - 1$

Regression Model

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon = y = \bar{y}_{..} + \frac{A}{2}x_1 + \frac{B}{2}x_2 + \frac{AB}{2}x_1 x_2$, which means that the estimated coefficients are half of the effects.

2^3 Factorial Design

Introduction

Can be extended to 2^3, … 2^k design. Every column has an equal number of + and - signs. The sum of the product of signs in any two columns is zero. Multiplying any column by I leaves that column unchanged. The product of any two columns in the table yields another column. Orthogonal design as in AB X B = A

Estimates

grand mean: $\frac{\sum \bar{y}_{i.}}{2^3}$, effect: $\frac{\sum c_i \bar{y}_{i.}}{2^{3-1}}$, Var(effect): $\frac{\sigma^2}{n 2^{3-2}}$

Contrast Sum of Squares: $SS_{effect} = \frac{(\sum c_i \bar{y}_{i.})^2}{2^3/n} = 2n(effect)^2$

Confidence Interval

t-test for effects

effect $\pm t_{\alpha/2, 2^k(n-1)}$S.E.(effect)

Model Summary Statistics

$\hat{\beta} - t_{\alpha/2, df_E} se(\hat{\beta}) \leq \beta \leq \hat{\beta} + t_{\alpha/2, df_E} se(\hat{\beta})$, where Standard Error: $se(\hat{\beta}) = \sqrt{\frac{MSE}{n2^k}} = \sqrt{\frac{\sigma^2}{n2^k}}$

$R^2 = \frac{SS_{Model}}{SS_T}$, $R^2_{Adj} = 1 - \frac{SS_E/df_E}{SS_T/df_T}$

Notes:

■ **TABLE 6.3**

Algebraic Signs for Calculating Effects in the 2^3 Design

Treatment Combination	Factorial Effect							
	I	*A*	*B*	*AB*	*C*	*AC*	*BC*	*ABC*
(1)	+	−	−	+	−	+	+	−
a	+	+	−	−	−	−	+	+
b	+	−	+	−	−	+	−	+
ab	+	+	+	+	−	−	−	−
c	+	−	−	+	+	−	−	+
ac	+	+	−	−	+	+	−	−
bc	+	−	+	−	+	−	+	−
abc	+	+	+	+	+	+	+	+

Figure 1: 2^3 Factorial Effect Table

General 2^k Design

Has k factors each with 2 levels. Consists of all possible level combinations 2^k treatments each with n replicates. The main effects are first order with k effects. 2-factor or second order interactions have $\binom{k}{2}$. Third... Has 2^{k-1} total effects. Each effect has 1 degree of freedom and error has $2^k - 1$ degrees of freedom that add up to 2^{k-1} degrees of freedom.

Has one-to-one correspondence between effects and contrasts. For main effect, column level only needs to be changed. For interactions, multiplying the contrasts of the main effects of the involved factors.

Estimates

grand mean: $\frac{\sum \bar{y}_{i.}}{2^k}$

For effect with contrast $C = (c_1, c_2, ..., c_{2^k})$, $effect = \frac{\sum c_i \bar{y}_i}{2^{(k-1)}}$

Variance: Var(effect) $= \frac{\sigma^2}{n2^{k-2}} = \frac{MSE}{n2^{k-2}}$

t-test: effect $\pm t_{\alpha/,2^k(n-1)} S.E.(effect)$

$SS_{effect} = \frac{(\sum c_i \bar{y}_{i.})^2}{2^k/n} = n2^{k-2}(effect)^2$

Coefficents are estimated by half of effects they represent in regression model.

Unreplicated 2^k Design

No degree of freedom left for error if full model is fitted. Estimates and contrast sum of squares is the same as the general model with $n = 1$. No error sum of squares available, cannot estimate σ^2 and test effects in both the ANOVA and Regression approaches.

Approach 1: Pooling high-order interactions

Usually 3 or higher interactions do not occur and are pooled. May pool siginificant interactions.

Approach 2: Using the normal probability plot to identify significant effects

Estimates of effects are assumed to follow $N(0, \frac{\sigma^2}{2^{k-2}})$

The QQ plot of the estimates is expected to be a linear line. Deviations from a linear line indicates significant effects.