Experimental Design
Note 7-1
2^K Factorial Design (II)

Keunbaik Lee

Sungkyunkwan University

General 2^K Design I

- k factors: A, B, $\cdots$, K each with 2 levels (+,-)
- consists of all possible level combinations (2^K treatments) each with n replicates
- Classify factorial effects:

Type of effect	Label	Number of effects
Main effects (of order 1)	$A, B, \cdots, K$	K
2-factor interactions (of order 2)	$AB, AC, \cdots, JK$	$\binom{K}{2}$
3-factor interactions (of order 3)	$ABC, ABD, \cdots, IJK$	$\binom{K}{3}$
$\cdots$	$\cdots$	$\cdots$
K-factor interaction (of order k)	$ABC\cdots K$	$\binom{K}{K} = 1$

- In total, how many effects?
- Each effect (main or interaction) has 1 degree of freedom full model (i.e., model consisting of all the effects) has $2^K - 1$ degrees of freedom.

General 2^K Design II

- Error component has $2^K(n-1)$ degrees of freedom (why?).
- One-to-one correspondence between effects and contrasts:
  - For main effect: convert the level column of a factor using - $\Rightarrow$ -1 and $+ \Rightarrow 1$.
  - For interactions: multiply the contrasts of the main effects of the involved factors, componentwisely.
- Estimates:

$$\text{grand mean: } \frac{\sum_i \bar{y}_{i\cdot}}{2^K}$$

For effect with contrast $C = (c_1, c_2, \cdots, c_{2^K})$, its estimate is

$$\text{effect} = \frac{\sum_i c_i \bar{y}_{i\cdot}}{2^{K-1}}$$

General 2^K Design III

- Variance:

$$var(\text{effect}) = \frac{\sigma^2}{n2^{K-2}} \sim MSE$$

  what is the standard error of the effect?

- t-test for H_0 : effect $= 0$. Using the confidence interval approach,

$$\text{effect} \pm t_{\alpha/2, 2^K(n-1)} SE(\text{effect})$$

Using ANOVA model:

- Sum of Squares due to an effect, using its contrast,

$$SS_{Effect} = \frac{(\sum_i c_i \bar{y}_{i.})^2}{2^K / n} = n2^{K-2}(\text{effect})^2$$

General 2^K Design IV

- SS_T and SSE can be calculated as before and a ANOVA table including SS due to the effects and SSE can be constructed and the effects can be tested by F-tests.

Using Regression:

- Introducing variables $x_1, \cdots, x_K$ for main effects, their products are used for interactions, the following regression model can be fitted

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_K + \cdots + \beta_{12\cdots k} x_1 x_2 \cdots x_K + \epsilon$$

The coefficients are estimated by half of effects they represent, that is,

$$\hat{\beta} = \frac{\text{effect}}{2}$$

Unreplicated 2^K Design I

- No degree of freedom left for error component if full model is fitted.
- Formulas used for estimates and contrast sum of squares are given in the previous slides 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
  - Often assume 3 or higher interactions do not occur.
  - Pool estimates together for error.
  - Warning: may pool significant interaction.
- Approach 2: Using the normal probability plot (QQ plot) to identify significant effects.

Unreplicated 2^K Design II

- Recall
$$var(\text{effect}) = \frac{\sigma^2}{2^{(K-2)}}.$$
If the effect is not significant $(= 0)$, then the effect estimate follows $N\left(0, \frac{\sigma^2}{2^{(K-2)}}\right)$.
- Assume all effects not significant, their estimates can be considered as a random sample from $N\left(0, \frac{\sigma^2}{2^{(K-2)}}\right)$.
- QQ plot of the estimates is expected to be a linear line.
- Deviation from a linear line indicates significant effects.

Unreplicated 2^K Design III

Filtration Rate Experiment

factor				
A	B	C	D	filtration
−	−	−	−	45
+	−	−	−	71
−	+	−	−	48
+	+	−	−	65
−	−	+	−	68
+	−	+	−	60
−	+	+	−	80
+	+	+	−	65
−	−	−	+	43
+	−	−	+	100
−	+	−	+	45
+	+	−	+	104
−	−	+	+	75
+	−	+	+	86
−	+	+	+	70
+	+	+	+	96

Unreplicated 2^K Design IV

See Design2level.SAS.

Fit a linear line based on small effects, identify the effects which are potentially significant, then use ANOVA or regression fit a sub-model with those effects.

- Potentially significant effects: A, AD, C, D, AC.
- Use main effect plot and interaction plot.
- ANOVA model involving only A, B, D and their interactions (projecting the original unreplicated 2^4 experiment onto a replicated 2^3 experiment).
- regression model only involving A, C, D, AC and AD.
- Diagnostics using residuals.

See Design2level-1.SAS.