STA305/1004-Class 20

March 27, 2017

Today's Class

- Assessing significance in unreplicated factorial designs
  - Normal plots
  - half-Normal plots
  - Lenth's method
- Blocking factorial designs
  - Effect hierarchy principle
  - Generation of orthogonal blocks
  - Generators and deining relations

Exam review session

I will announce details by the end of the week.

Example - $2^{\cancel{3}\,4}$ design for studying a chemical reaction

A process development experiment studied four factors in a 2^4 factorial design.

- amount of catalyst charge **1**,
- temperature **2**,
- pressure **3**,
- concentration of one of the reactants **4**.
- The response y is the percent conversion at each of the 16 run conditions. The design is shown below.

Example - 2^4 design for studying a chemical reaction

x1	x2	x3	x4	conversion
-1	-1	-1	-1	70
1	-1	-1	-1	60
-1	1	-1	-1	89
1	1	-1	-1	81
-1	-1	1	-1	69
1	-1	1	-1	62
-1	1	1	-1	88
1	1	1	-1	81
-1	-1	-1	1	60
1	-1	-1	1	49
-1	1	-1	1	88
1	1	-1	1	82
-1	-1	1	1	60
1	-1	1	1	52
-1	1	1	1	86
1	1	1	1	79

The design is not replicated so it's not possible to estimate the standard errors of the factorial effects.

Example - 2^4 design for studying a chemical reaction

```r
fact1 <- lm(conversion~x1*x2*x3*x4,data=tab0510a)
round(2*fact1$coefficients,2)
```

```
(Intercept)          x1          x2          x3          x4       x1:x2
     144.50       -8.00       24.00       -0.25       -5.50        1.00
      x1:x3       x2:x3       x1:x4       x2:x4       x3:x4    x1:x2:x3
       0.75       -1.25        0.00        4.50       -0.25       -0.75
   x1:x2:x4    x1:x3:x4    x2:x3:x4 x1:x2:x3:x4
       0.50       -0.25       -0.75       -0.25
```

Half-Normal Plots

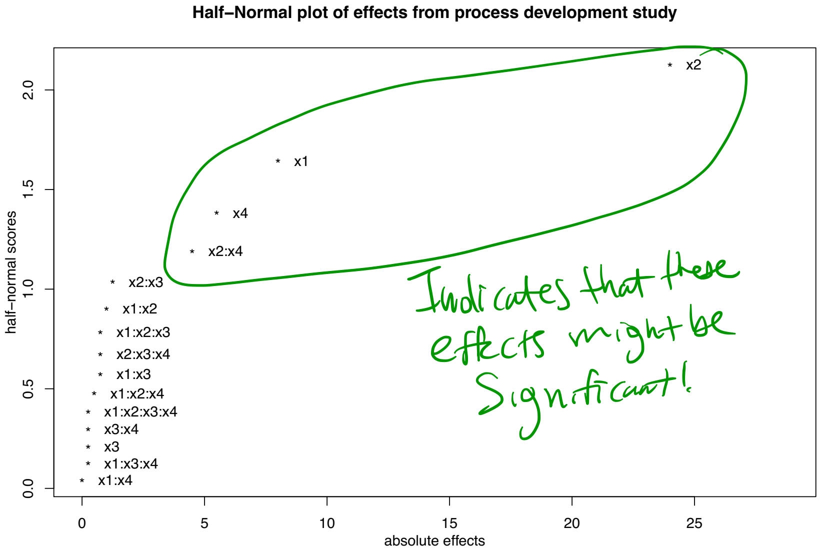
- A related graphical method is called the half-normal probability plot.
- Let
$$\left|\hat{\theta}\right|_{(1)} < \left|\hat{\theta}\right|_{(2)} < \cdots < \left|\hat{\theta}\right|_{(N)}.$$
denote the ordered values of the unsigned factorial effect estimates.
- Plot them against the coordinates based on the half-normal distribution - the absolute value of a normal random variable has a half-normal distribution.
- The half-normal probability plot consists of the points
$$\left|\hat{\theta}\right|_{(i)} \text{ vs. } \Phi^{-1}(0.5 + 0.5[i - 0.5]/N).\ i = 1, ..., N.$$

Half-Normal Plots

- An advantage of this plot is that all the large estimated effects appear in the upper right hand corner and fall above the line.
- The half-normal plot for the effects in the process development example can be obtained with `DanielPlot()` with the option `half=TRUE`.

Half-Normal Plots

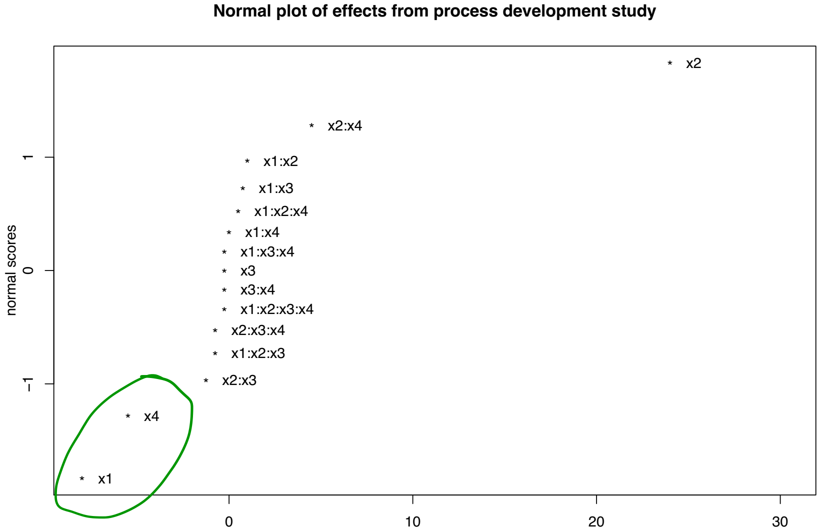
```r
library(FrF2)
DanielPlot(fact1,half=TRUE,autolab=F,
         main="Half-Normal plot of effects from process development study")
```



Half-Normal Plots

Compare with full Normal plot.

```
library(FrF2)
DanielPlot(fact1,half=F,autolab=F,
          main="Normal plot of effects from process development study")
```



Lenth's method: testing significance for experiments without variance estimates

- Half-normal and normal plots are informal graphical methods involving visual judgement.
- It's desirable to judge a deviation from a straight line quantitatively based on a formal test of significance.
- Lenth (1989) proposed a method that is simple to compute and performs well. (ref. pg. 205, Box, Hunter, and Hunter, 2005)

Lenth's method

- Let

$$\hat{\theta}_{(1)}, ..., \hat{\theta}_{(N)}$$

be estimated factorial effects of $\theta_1, \theta_2, ..., \theta_N$ in a 2^k design $N = 2^k - 1$.

- Assume that all the factorial effects have the same standard deviation.
- The pseudo standard error (PSE) is defined as

$$PSE = 1.5 \cdot \text{median}_{\left|\hat{\theta}_i\right| < 2.5 s_0} \left|\hat{\theta}_i\right|,$$

- The median is computed among the $\left|\hat{\theta}_i\right|$ with $\left|\hat{\theta}_i\right| < 2.5 s_0$ and

$$s_0 = 1.5 \cdot \text{median} \left|\hat{\theta}_i\right|.$$

Lenth's method

- $1.5 \cdot s_0$ is a consistent estimator of the standard deviation of $\hat{\theta}$ when $\theta_i = 0$ and the underlying distribution is normal.
- The $P\left(|Z| > 2.57\right) = 0.01, Z \sim N(0,1)$.
- $\left|\hat{\theta}_i\right| < 2.5 s_0$ trims approximately 1% of the $\hat{\theta}_i$ if $\theta_i = 0$.
- The trimming attempts to remove the $\hat{\theta}_i$ associated with non-zero (active) effects.
- By using the median in combination with the trimming means that *PSE* is not sensitive to the $\hat{\theta}_i$ associated with active effects.

Lenth's method

$N = 2^k - 1$

- To obtain a margin of error Lenth suggested multiplying the PSE by the $100 * (1 - \alpha)$ quantile of the t_d distribution, $t_{d,\alpha/2}$.
- The degrees of freedom is $d = N/3$. For example, the margin of error for a 95% confidence interval for θ_i is

$$ME = t_{d,.025} \times PSE.$$

- All estimates greater than the ME may be viewed as "significant", but with so many estimates being considered simultaneously, some will be falsely identified.
- A simultaneous margin of error that accounts for multiple testing can also be calculated,

Bonn Correction.

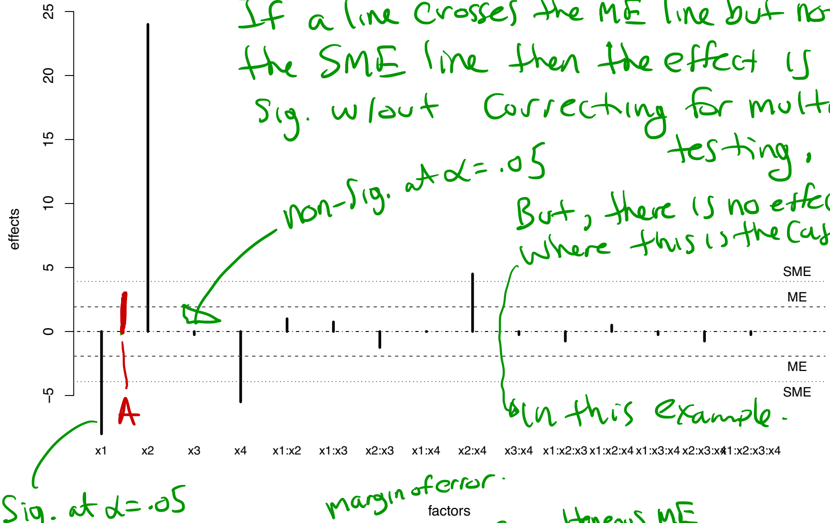
$$SME = t_{d,\gamma} \times PSE,$$

where $\gamma = \left(1 + (1 - \alpha)^{1/N}\right)/2$.

- The details of how to calculate MSE and PSE are given in the class notes.

Lenth's method - Lenth Plot for process development example

```r
LenthPlot(fact1,cex.fac = 0.8)
```



If a line crosses the ME line but not the SME line then the effect is sig. w/out correcting for multiple testing.

But, there is no effect where this is the case in this example.

non-sig. at $\alpha = .05$

Sig. at $\alpha = .05$

margin of error.

Simultaneous ME

```
##    alpha       PSE        ME       SME
## 0.050000 0.750000 1.927936 3.913988
```

Blocking factorial designs

- In a trial conducted using a 2^3 design it might be desirable to use the same batch of raw material to make all 8 runs.
- Suppose that batches of raw material were only large enough to make 4 runs. Then the concept of blocking could be used.

Blocking factorial designs

Consider the 2^3 design.

Run	1	2	3	12	13	23	123
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

Runs	Block
1, 4, 6, 7	I
2, 3, 5, 8	II

$123 = -$

$123 = +$

How are the runs assigned to the blocks?

Blocking factorial designs

Run	1	2	3	12	13	23	123
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

Runs	Block	sign of 123
1, 4, 6, 7	I	$-$
2, 3, 5, 8	II	$+$

Blocking factorial designs

- Any systematic differences between the two blocks of four runs will be eliminated from all the main effects and two factor interactions.
- What you gain is the elimination of systematic differences between blocks.
- But now the three factor interaction is confounded with any batch (block) difference.
- The ability to estimate the three factor interaction separately from the block effect is lost.

Effect hierarchy principle

1. Lower-order effects are more likely to be important than higher-order effects.
2. Effects of the same order are equally likely to be important.

- One reason that many accept this principle is that higher order interactions are more difficult to interpret or justify physically.
- Investigators are less interested in estimating the magnitudes of these effects even when they are statistically significant.

Generation of Orthogonal Blocks

In the 2^3 example suppose that the block variable is given the identifying number 4.

Run	1	2	3	4=123
1	-1	-1	-1	-1
2	1	-1	-1	1
3	-1	1	-1	1
4	1	1	-1	-1
5	-1	-1	1	1
6	1	-1	1	-1
7	-1	1	1	-1
8	1	1	1	1

- Think of your experiment as containing four factors.
- The fourth factor will have the special property that it does not interact with other factors.
- If this new factor is introduced by having its levels coincide exactly with the plus and minus signs attributed to 123 then the blocking is said to be **generated** by the relationship 4=123.
- This idea can be used to derive more sophisticated blocking arrangements.

An example of how not to block

$4 \cdot 5 = 123 \cdot 23$
$= 132 \cdot 23 = 13 \cdot 3 = 1$

Suppose we would like to arrange the 2^3 design into four blocks.

Run	1	2	3	4=123	5=23	45=1
1	-1	-1	-1	-1	1	-1
2	1	-1	-1	1	1	1
3	-1	1	-1	1	-1	-1
4	1	1	-1	-1	-1	1
5	-1	-1	1	1	-1	-1
6	1	-1	1	-1	-1	1
7	-1	1	1	-1	1	-1
8	1	1	1	1	1	1

- Runs are placed in different blocks depending on the signs of the block variables in columns 4 and 5.
- Consider two block factors called 4 and 5.
- 4 is associated with ? 123
- 5 is associated ? 23

An example of how not to block

Run	1	2	3	4=123	5=23	45=1
1	-1	-1	-1	-1	1	-1
2	1	-1	-1	1	1	1
3	-1	1	-1	1	-1	-1
4	1	1	-1	-1	-1	1
5	-1	-1	1	1	-1	-1
6	1	-1	1	-1	-1	1
7	-1	1	1	-1	1	-1
8	1	1	1	1	1	1

Block	Run
I	4,6
II	3,5
III	1,7
IV	2,8

4	5
−	−
+	−
−	+
+	+

An example of how not to block

- 45 is confounded with the main effect of 1.
- Therefore, if we use 4 and 5 as blocking variables we will not be able to separately estimate the main effect 1.
- Main effects should not be confounded with block effects.

An example of how not to block

- Any blocking scheme that confounds main effects with blocks should not be used.
- This is based on the assumption:

  The block-by-treatment interactions are negligible.

- This assumption states that treatment effects do not vary from block to block.
- Without this assumption estimability of the factorial effects will be very complicated.

An example of how not to block

- For example, if $B_1 = 12$ then this implies two other relations:

$$1B_1 = 112 = 2 \text{ and } B_12 = 122 = 122 = 1.$$

- If there is a significant interaction between the block effect B_1 and the main effect 1 then the main effect 2 is confounded with $1B_1$.
- If there is a significant interaction between the block effect B_1 and the main effect 2 then the main effect 1 is confounded with B_12.

How to do it

Run 4 5

Block I 2,7 − −

II 3,6 − +

III 4,5 + −

IV 1,8 + +

$4 = 12$

$4 \cdot 4 = 124$

$I = 124$

$I =$ Column of $+1$

Run	1	2	3	4=12	5=13
1	-1	-1	-1	1	1
2	1	-1	-1	-1	-1
3	-1	1	-1	-1	1
4	1	1	-1	1	-1
5	-1	-1	1	1	-1
6	1	-1	1	-1	1
7	-1	1	1	-1	-1
8	1	1	1	1	1

$124 \cdot 135 = 2345$

- Set 4=12, 5=13.
- Then $I = 124 = 135 = 2345$.
- Estimated block effects 4, 5, 45 are assoicated with the estimated two-factor interaction effects 12, 13, 23 and not any main effects.
- Which runs are assigned to which blocks?

∴ a better blocking scheme compared to previous.

Generators and Defining Relations

- A simple calculus is available to show the consequences of any proposed blocking arrangement.
- If any column in a 2^k design are multiplied by themselves a column of plus signs is obtained. This is denoted by the symbol I.

$$I = 11 = 22 = 33 = 44 = 55,$$

where, for example, 22 means the product of the elements of column 2 with itself.

- Any column multiplied by I leaves the elements unchanged. So, $I3 = 3$.

Generators and Defining Relations

- A general approach for arranging a 2^k design in 2^q blocks of size 2^{k-q} is as follows.
- Let $B_1, B_2, ..., B_q$ be the block variables and the factorial effect v_i is confounded with B_i,

$$B_1 = v_1, B_2 = v_2, ..., B_q = v_q.$$

- The block effects are obtained by multiplying the B_i's:

$$B_1B_2 = v_1v_2, B_1B_3 = v_1v_3, ..., B_1B_2 \cdots B_q = v_1v_2 \cdots v_q$$

- There are $2^q - 1$ possible products of the B_i's and the I (whose components are $+$).

Generators and Defining Relations

$7 = 2^{3} - 1$

Example: A 2^5 design can be arranged in 8 blocks of size $2^{5-3} = 4$.

Consider two blocking schemes.

1. Define the blocks as

$B_1B_2 = 12$, $B_1B_3 = 245$, $B_2B_3 = 145$, $B_1B_2B_3 = 34$

$$B_1 = 135, B_2 = 235, B_3 = 1234.$$

2 - two factor interactions

The remaining blocks are confounded with the following interactions:

2. Define the blocks as:

$B_1B_2 = 23$, $B_1B_3 = 1245$ 4 - two factor interactions

$$B_1 = 12, B_2 = 13, B_3 = 45.$$

$B_2B_3 = 1345$

$B_1B_2B_3 = 2345$

Which is a better blocking scheme?

$\therefore$ first scheme is better! $\because$ confounds less lower order interactions.