Experimental Design and Analysis Homework 6

110024516 邱繼賢

Problem 1. (5-4)

如果一個 design 的 resolution $= R$，代表在它的 defining contrast subgroup 中最短的 wordlength $= R$，那在它投影到一個 $R-1$ 的 factorial design 時，由於只有 $R-1$ 個 factors，所以它一定無法形成長度為 R 的 word，故一定為一個 full factorial design。

Problem 2. (5-13)

(a)

There are five 2-level factors :

(1) **temperature** : 160°F (-) or 180°F (+)

(2) **concentration** : 30% (-) or 40% (+)

(3) **catalyst** : A (-) or B (+)

(4) **stirring rate** : 60rpm (-) or 100rpm (+)

(5) **pH** : low (-) or high (+)

and we need to avoid the combinations of (+,+, ,+,) & (+, ,+, ,+), so we construct the 2^{5-2} design with 2 generators :

$$4 \ = -12 \ , \ 5 \ = -13$$

then the defining contrast subgroup :

$$I \ = -124 \ = -135 \ = \ 2345$$

The design matrix as below :

Run	temperature	concentration	catalyst	stirring rate	pH
1	+	+	+	-	-
2	+	+	-	-	+
3	+	-	+	+	-
4	+	-	-	+	+
5	-	+	+	+	+
6	-	+	-	+	-
7	-	-	+	-	+
8	-	-	-	-	-

此處僅為展示方便，實際做實驗時需要將實驗順序進行隨機排序。

(b)

Use defining contrast subgroup :

$$I \ = -124 \ = \ 345 \ = -1235$$

and there are seven alias sets :

$$1 = -24 = 1345 = -235$$

$$2 = -14 = 2345 = -135$$

$$3 = -1234 = 45 = -125$$

$$12 = -4 = 12345 = -35$$

$$13 = -234 = 145 = -25$$

$$23 = -134 = 245 = -15$$

$$123 = -34 = 1245 = -5$$

We can see that the catalyst-by-temperature (13) and catalyst-by-concentration (23) interaction effects are neither aliased with the main effects nor with each other. The design matrix is shown below :

Run	temperature	concentration	catalyst	stirring rate	pH
1	+	+	+	-	-
2	+	+	-	-	+
3	+	-	+	+	+
4	+	-	-	+	-
5	-	+	+	+	+
6	-	+	-	+	-

Run	temperature	concentration	catalyst	stirring rate	pH
7	-	-	+	-	-
8	-	-	-	-	+

我們為了在此實驗設計中避免某些 2-factor interactions 和主效應有所混淆，相對的就無法避免 (+, ,+, ,+) 這種高風險組合的出現 (look at the 3rd run)。

一樣此處僅為展示方便，實際進行實驗仍須將實驗順序隨機排列。

Problem 3. (5-15)

(a)

Let's look at the defining contrast subgroups of both two designs :

$$\text{(i) } I = 12345 = 1246 = 356 \Rightarrow \text{resolution} = 3$$

$$\text{(ii) } I = 1235 = 1246 = 3456 \Rightarrow \text{resolution} = 4$$

I prefer design **(ii)** because it has larger resolution.

(b)

Take a look at only the alias sets, which contain 2-factor interaction effects, of design **(ii)** :

$$12 = 35 = 46 = 123456$$

$$13 = 25 = 2346 = 1456$$

$$14 = 2345 = 26 = 1356$$

$$23 = 15 = 1346 = 2456$$

$$24 = 1345 = 16 = 2356$$

$$34 = 1245 = 1236 = 56$$

$$1234 = 45 = 36 = 1256$$

We can ignore all the 2-factor interactions involving factor 6 and 3(or higher)-factor interactions in above alias sets.

Therefore, 2-factor interactions : **14, 24, 34, 45** are estimable.

However, **12=35, 13=25, 23=15** these three pairs of 2-factor interactions are still aliased.

(c)

Arranging the 2^{6-2} design **(ii)** in 2^1 blocks need 1 block factor. My blocking scheme is $B = 134$ because the confounding set under this condition is

$$B = 134 = 245 = 236 = 156$$

only contains 3-factor interactions which are negligible under the usual assumption in **(b)**. In this way, we will not confound any other main or 2-factor interaction effects which we more concern about.

Problem 4. (5-16)

(a)

Let's look at the defining contrast subgroups of two 2^{6-2} designs :

$$\text{A} : I = 12345 = 1236 = 456 \Rightarrow \text{resolution} = 3$$

$$\text{B} : I = 1235 = 2346 = 1456 \Rightarrow \text{resolution} = 4$$

I will choose design B because it has larger resolution.

(b)

There are $2^4 = 16$ runs in 2^{6-2} design, so we only have 15 degrees of freedom to estimate factorial effects. However, if resolution = 5, all main and 2-factor interaction effects are clear, namely all of them are fall in different alias sets. We will need at least $6 + \binom{6}{2} = 21$ degrees of freedom to estimate all of them. It is impossible.

Problem 5. (5-28)

(a)

All 2-factor interaction effects : **AB, AC, AD, AE, BC, BD, BE, CD, CE, DE** are clear.
All main effects : **A, B, C, D, E** are strongly clear.

(b)

計算每一組 level combinations 下的 $\bar{y}$ 和 s^2，若 $s^2 = 0$ 則以 0.001 代入

```
library(dplyr)
library(knitr)
welding = read.table("welding.txt", header = T)
colnames(welding) = c("A","B","C","D","E","y")
data = welding %>% group_by(A,B,C,D,E) %>%
    summarise(y_bar = mean(y), s_square = ifelse(var(y)==0,0.001,var(y))) %>%
    ungroup() %>%
    mutate(A=rep(c(-1,1),each=8),B=rep(c(-1,1),each=4,2),C=rep(c(-1,1),each=2,4),D=rep(c(-1,1),8),
           E=c(-1,1,1,-1,1,-1,-1,1,1,-1,-1,1,-1,1,1,-1))
kable(data, col.names = c("A","B","C","D","E","$\\bar{y}$","s^2"), digits = 3)
```

A	B	C	D	E	$\bar{y}$	s^2
-1	-1	-1	-1	-1	1275.000	9075.000
-1	-1	-1	1	1	1495.000	39325.000
-1	-1	1	-1	1	1220.000	63525.000
-1	-1	1	1	-1	1275.000	21175.000
-1	1	-1	-1	1	1495.000	9075.000
-1	1	-1	1	-1	1440.000	3025.000
-1	1	1	-1	-1	1275.000	0.001
-1	1	1	1	1	1256.667	1008.333
1	-1	-1	-1	1	1696.667	1008.333
1	-1	-1	1	-1	1898.333	1008.333
1	-1	1	-1	-1	1770.000	0.001
1	-1	1	1	1	1916.667	1008.333
1	1	-1	-1	-1	1916.667	1008.333
1	1	-1	1	1	1641.667	1008.333
1	1	1	-1	1	1990.000	0.001
1	1	1	1	-1	2118.333	1008.333

根據 **(a)** 所得到的結論，可以得知每一個 main 和 2-factor interaction effects 都是 clear effects，也就是說它們都只和 3(or 4)-factor interactions 落在同一個 alias sets，而那些高階的 interaction effects 都是可以忽略它們的效應的，故我們可以建構以下 location 和 dispersion models :

$$\bar{y} = X\beta + \epsilon \ , \ \ln s^2 = X\gamma + \delta$$

where X contains all main and 2-factor interaction effects of the five factors.

然後可以藉由估計出來的迴歸係數來推得 location 和 dispersion effects：

$$\hat{\theta} = 2\hat{\beta} \ , \ \hat{\psi} = 2\hat{\gamma}$$

結果呈現如下表：

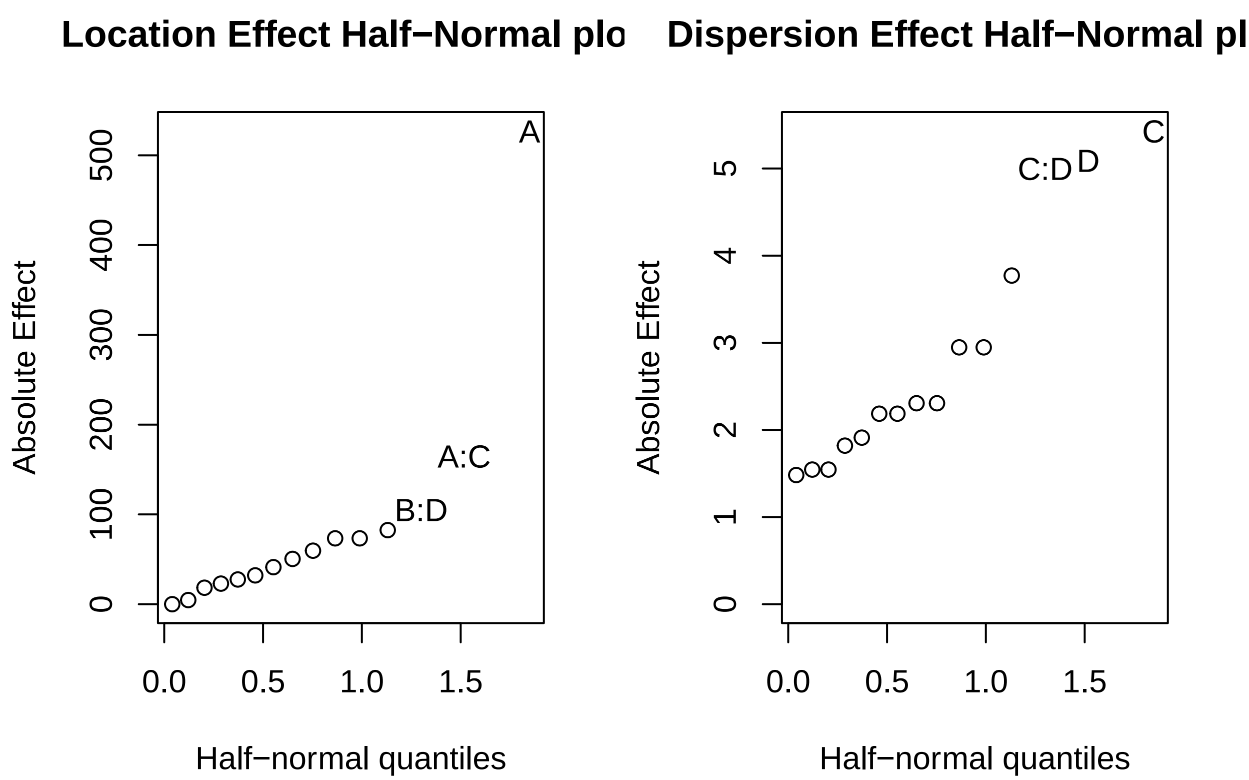
```
loc_mod = lm(y_bar ~ A+B+C+D+E+A:B+A:C+A:D+A:E+B:C+B:D+B:E+C:D+C:E+D:E, data)
loc_effect = 2*coef(loc_mod)[-1]
dis_mod = lm(log(s_square) ~ A+B+C+D+E+A:B+A:C+A:D+A:E+B:C+B:D+B:E+C:D+C:E+D:E, data)
dis_effect = 2*coef(dis_mod)[-1]
effect_tab = data.frame(y_bar = loc_effect, log_s = dis_effect)
kable(effect_tab, col.names = c("$\\bar{y}$","$ln s^2$"), digits = 3)
```

	$\bar{y}$	$ln s^2$
A	527.083	-3.771
B	73.333	-2.947
C	-4.583	-5.430
D	50.417	5.093
E	-32.083	2.186
A:B	22.917	2.947
A:C	165.000	-1.482
A:D	0.000	1.819
A:E	-82.500	-2.186
B:C	41.250	-2.306
B:D	-105.417	1.545
B:E	-59.583	-1.911
C:D	27.500	5.001
C:E	18.333	1.545
D:E	-73.333	-2.306

接下來以 Half-Normal plot 的方式來判斷哪些效應為顯著：

```
"halfnorm" <- function (x, nlab = 2, labs = as.character(1:length(x)), ylab = "Sorted Data") {
    x <- abs(x)
    labord <- order(x)
    x <- sort(x)
    i <- order(x)
```

```r
    n <- length(x)
    ui <- qnorm((n + 1:n)/(2 * n + 1))
    plot(ui, x[i], xlab = "Half-normal quantiles", ylab = ylab, ylim=c(0,max(x)),
         type = "n")
    if(nlab < n)
        points(ui[1:(n - nlab)], x[i][1:(n - nlab)])
    text(ui[(n - nlab + 1):n], x[i][(n - nlab + 1):n], labs[labord][(n - nlab + 1):n])
}
par(mfrow = c(1,2))
halfnorm(loc_effect,3,names(loc_effect), "Absolute Effect")
title("Location Effect Half-Normal plot")
halfnorm(dis_effect,3,names(dis_effect), "Absolute Effect")
title("Dispersion Effect Half-Normal plot")
```



從圖形判斷：

(1) Location： A, AC 兩效應顯著

(2) Dispersion : C, D, CD 三效應顯著

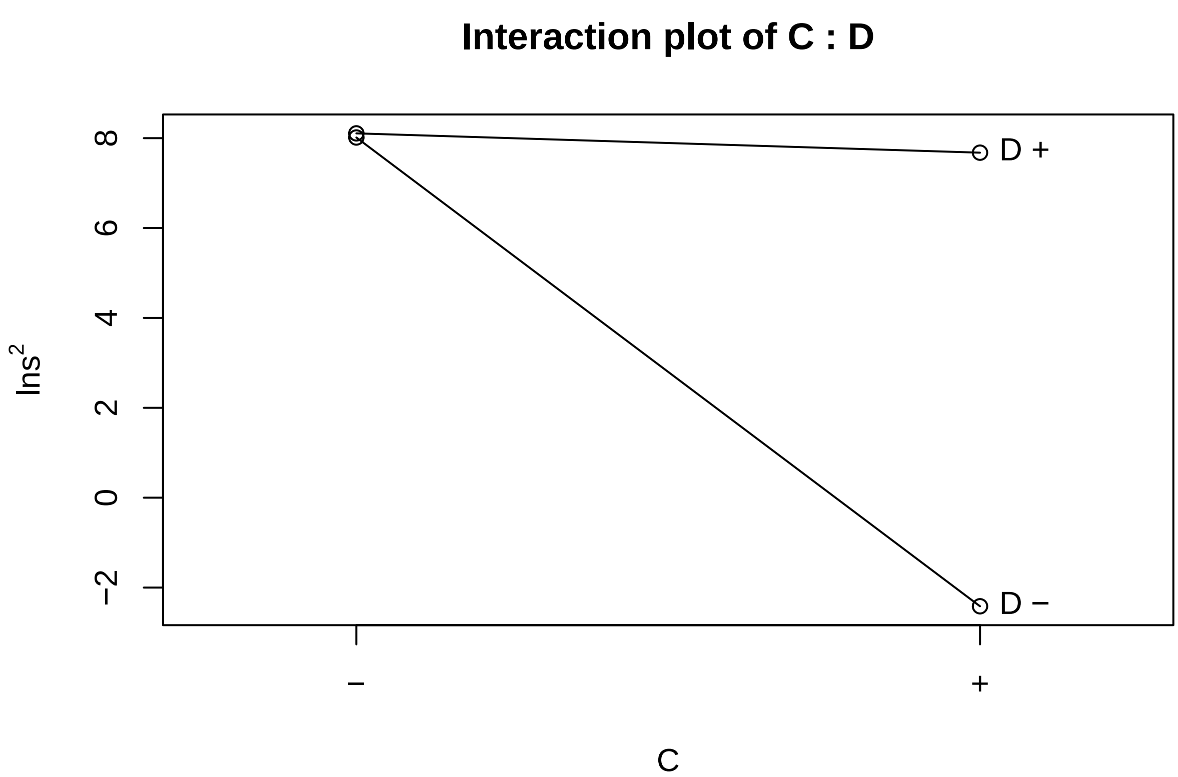
故能配飾 location 和 dispersion models :

$$\hat{\bar{y}} = \hat{\beta}_0 + \hat{\beta}_A x_A + \hat{\beta}_{AC} x_{AC} = 1605 + 263.5\, x_A + 82.5\, x_{AC}$$

$$\ln \hat{s}^2 = \hat{\gamma}_0 + \hat{\gamma}_C x_C + \hat{\gamma}_D x_D + \hat{\gamma}_{CD} x_{CD} = 5.3456 - 2.715\, x_C + 2.5463\, x_D + 2.5003\, x_{CD}$$

(c)

```
library(latex2exp)
data2 = data %>% group_by(C,D) %>% summarise(m = mean(log(s_square)))
plot(data2$C, data2$m, xaxt = "n",xlab = "C", ylab = TeX("$ln s^2$"),
     main = "Interaction plot of C : D", xlim = c(-1.5,1.5), pch = 1)
axis(1,c(-1,1),labels = c("-","+"))
segments(-1,data2$m[1],1,data2$m[3]) ; segments(-1,data2$m[2],1,data2$m[4])
text(1,data2$m[3],"D -",pos = 4) ; text(1,data2$m[4],"D +",pos = 4)
```



藉由 factor C, D 的 interaction plot 可以看出在 (C, D) 不同的設定值下，$\ln s^2$ 的大小也會有所不同，而從圖中看出 $(C, D) = (+, -)$ 會有最小的變異，從 dispersion model 觀察也能得到相同結論，故 $(C, D) = (+, -)$ 就是 optimal factor settings for minimizing variance

(d)

選擇 factor settings $(A, C) = (+, +)$ 可以得到 location model 的最大值

$$\hat{\hat{y}} = 1605 + 263.5 + 82.5 = 1951$$

故 $(A, C) = (+, +)$ 就是 optimal factor settings for maximizing the tensile strength

(e)

這是一個 Larger-the-better problem，adjustment factor 為那些對 dispersion model 有顯著效應但是對 location model 無顯著貢獻的 factor，故 factor D 為 adjustment factor
Two-step procedure :

(1) Choose $(A, C) = (+, +)$ to maximize the tensile strength

$$\hat{\hat{y}} = 1605 + 263.5 + 82.5 = 1951$$

(2) Choose $(C, D) = (+, -)$ to minimize variance

$$\hat{s}^2 = \exp[5.3456 - 2.715 - 2.5463 - 2.5003] = 0.08927802$$

PS : Larger-the-best 和 Nominal-the-best problem 在決定 adjustment factor 以及 two-step procedure 順序剛好相反。