Design of experiment - Homework 08

統碩一 106354003 林健宏

※ Exercise 13.2

Grains or crystals adversely affect the sensory qualities of foods using dried fruit pulp. A factorial experiment was conducted to determine which factors affect graininess. The factors were drying temperature (three levels), acidity (pH) of pulp (two levels), and sugar content (two levels). The experiment has two replications, with each replication using a different batch of pulp. Response is a measure of graininess.

		Sugar low		Sugar high	
Temp.	Rep.	pH low	pH high	pH low	pH high
1	1	21	12	13	1
	2	21	18	14	8
2	1	23	14	13	1
	2	23	17	16	11
3	1	17	20	16	14
	2	23	17	17	5

Analyze these data to determine which factors effect combination of factors leads to the least graininess.

此題想要探討有哪些因素會影響顆粒(graininess)大小，這些因素分別為溫度(temperature)、酸鹼度(Ph)以及糖分(Sugar)，在這些因素交互組合下，每個組合皆實驗 2 次，得到表一。接著進行 RCBD 的實驗，去檢測實質影響顆粒大小的因素有哪些，並找出影響最大、最小的因素以及到至最小顆粒因素組合。

		Sugar low		Sugar high	
Temp.	Rep.	pH low	pH high	pH low	pH high
1	1	21	12	13	1
	2	21	18	14	8
2	1	23	14	13	1
	2	23	17	16	11
3	1	17	20	16	14
	2	23	17	17	5

(表一)

RCBD ANOVA table

H_0:該因子對於顆粒大小無顯著效果

H_1:該因子對於顆粒大小有顯著效果

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Ph	1	130.021	130.021	38.902	0.0004295
Sugar	1	196.021	196.021	58.6492	0.0001203
Temp	2	13.792	6.896	2.0632	0.1975129
Residuals	7	23.396	3.342		

在給定顯著水準$\alpha = 0.05$下，從 RCBD table 中可看見每個因素的 p-value，其中 p-value 小於 0.05 的因子有酸鹼度以及糖分兩個因素，意味著這兩個因素對於顆粒大小有顯著的影響，而溫度的 p-value 則比 0.05 來得大，顯示溫度對於顆粒沒有顯著效果。

接著比較各因素對於顆粒大小影響的程度

Table of effects

Ph	
Ph-High	Ph-Low
-3.292	3.292

Sugar	
Sugar-High	Sugar-Low
-4.042	4.042

Temperature		
Temp-1	Temp-2	Temp-3
-1.2917	-0.0417	1.3334

從各因子對於顆粒大小影響程度來看以糖分影響最大，而溫度影響程度最小。由於溫度在 RCBD 檢定中呈現不顯著的結果，因此在因素組合影響顆粒大小的探討中，僅考慮糖分與酸鹼度兩因素的組合。而從原始資料來看，在高糖分、高酸鹼度的數據很明顯比其他組的數據來得小，在從 Table of effects 中看，高糖分以及高酸鹼度對於顆粒大小是負影響，即該因素越高則顆粒越小。所以在高糖分以及高酸鹼度下，會讓顆粒較小。

※ Exercise 13.3

The data below are from a replicated Latin Square with four treatments; row blocks were reused, but column blocks were not. Test for treatment differences and use Tukey HSD with level .01 to analyze the pairwise treatment differences.

D 44	B 26	C 67	A 77	B 51	D 62	A 71	C 49
C 39	A 45	D 71	B 74	C 63	A 74	D 67	B 47
B 52	D 49	A 81	C 88	A 74	C 75	B 60	D 58
A 73	C 58	B 76	D 100	D 82	B 79	C 74	A 68

拉丁方格 ANOVA table

H_0：Treatment 不顯著

H_1：Treatment 顯著

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Row	3	1913.6	637.88	39.2538	$4.19*10^{-08}$
square	1	36.1	36.13	2.2231	0.1533
Treatment	3	635.4	211.79	13.0333	$9.17*10^{-05}$
square:Column	6	4833.3	805.54	49.5718	$3.18*10^{-10}$
Residuals	18	292.5	16.25		

在給定顯著水準$\alpha = 0.1$ 下，Treatment 的 P value 遠小於α，拒絕H_0的假設，有足夠的證據證明因子影響效果的，即 treatment 之間存在著差異。

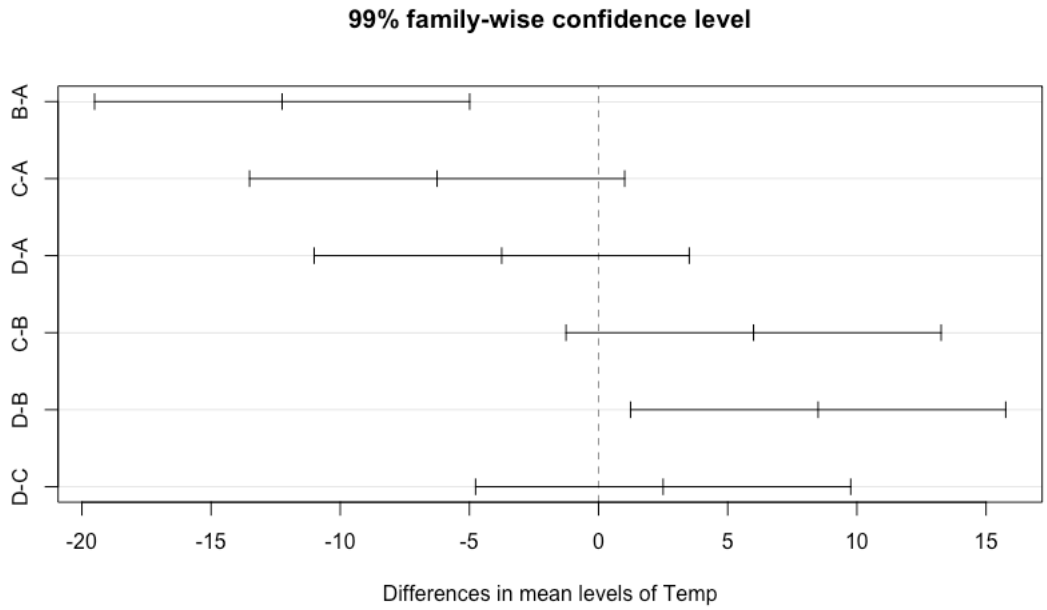
接著用 Tukey HSD 去對 treatment 做倆倆比較，看 treatment 之間的差異性。以下表格為兩倆比較之 99%的信賴區間以及圖。

$$H_0: \text{treatment}_i \text{與 treatment}_j \text{無顯著差異}; u_i - u_j = 0$$

$$H_1: \text{treatment}_i \text{與 treatment}_j \text{存在差異}; u_i - u_j \neq 0$$

	diff	ㄎ	upr
B-A	-12.25	-19.510396	-4.989604
C-A	-6.25	-13.510396	1.010396
D-A	-3.75	-11.010396	3.510396
C-B	6	-1.260396	13.260396
D-B	8.5	1.239604	15.760396
D-C	2.5	-4.760396	9.760396

（倆倆 treatment 平均比較之 99%信賴區間）



在這 6 個倆倆比較中，僅有 B-A, B-D 這兩組組合並無包含 0 在內，意味著拒絕H_0的假設，其餘都沒有足夠的證據去拒絕H_0的假設。

※ Appendix

```
# Exerciese 13.2 ----------------------------------------------------------

library(dplyr)
Ph = rep(c("PhLow","PhHigh"), time = 6)
Sugar = rep(c("SugarLow","SugarHigh"),each = 2,time = 3)
Temp <- rep(c("1","2","3"),each = 4)%>% as.factor()
Y <- c(21, 15, 13.5, 4.5, 23, 15.5, 14.5, 6, 20, 18.5, 16.5, 9.5)
data <- cbind.data.frame(Ph, Sugar, Temp,Y)
str(data)
la <- aov(Y~Ph+Sugar+Temp,data)
anova(la)

model.tables(la)

# Exercise 13.3 ----------------------------------------------------------

# L1
coll <- rep(c("C1","C2","C3","C4"),times = 4)
roww <- rep(c("R1","R2","R3","R4"),each = 4)
temp1 <-c("D","B","C","A","C","A","D","B","B","D","A","C","A","C","B","D")
Y1 <- c(44,26,67,77,39,45,71,74,52,49,81,88,73,58,76,100)
square1 <- rep(1,16) %>% as.factor()
data1 <- cbind.data.frame(coll,roww,temp1,Y1,square1)
colnames(data1) <- c("Column","Row","Temp","Y","square")
#L2
coll <- rep(c("C5","C6","C7","C8"),times = 4)
roww <- rep(c("R1","R2","R3","R4"),each= 4)
temp2 <- c("B","D","A","C","C","A","D","B","A","C","B","D","D","B","C","A")
Y2 <- c(51,62,71,49,63,74,67,47,74,75,60,58,82,79,74,68)
square2 <- rep(2,16) %>% as.factor()
data2 <- cbind.data.frame(coll,roww,temp2,Y2,square2)
colnames(data2) <- c("Column","Row","Temp","Y","square")
data <- rbind.data.frame(data1,data2)

temp3 <- lm(Y~Row+square+Temp+square:Column,data)
```

```
anova(temp3)
aov.model <- aov(Y~Row+Temp+square:Column+square,data)

sss <- TukeyHSD(aov.model,conf.level = 0.99)

plot(sss)
sss$Temp
```