Assignment 3 due by October 31, 2023
第七組
410650229 010 林可翰
410650377 015 張哲瑋
410650880 033 鄭暐瀚

1. (20 pt.) Do Problem 4.1(a). p126不同種類的石墨塗佈機對燈箱讀數的影響，挑三天連續測試，各種類的石墨塗佈機在不同天測試下，燈箱讀數會是隨機的

前提假設:

Y_{ij}=μ+β_i+τ_j+ε_{ij} for i=1,2,3 and j=1,2,3,4

Y_{ij}:第j個類型的石墨塗布機在第i天的燈箱讀數

μ:四種塗布機對燈箱讀數影響的共同效果。

β_i:天數 ,β_i ~ NID(0,σ^2_β)，σ^2_β第i天區集效果的共同變異數。

τ_j:第j個種類的石墨塗布機的處理效果, $\sum_{j=1}^{4} \tau_j = 0$

ε_{ij}:第j種石磨塗布機的第i天燈箱讀數的誤差 ε_{ij} ~ NID(0,σ^2)，σ^2是共同母體變異數。

假設β_i與ε_{ij}為獨立關係，天數與石墨塗布機間無交互作用。

檢定假設:

設τ_j第j個種類的石墨塗布機的處理效果。

H_0:τ_1=τ_2=τ_3=τ_4=0

H_1:至少有一個τ_j≠0

變異來源:D表示天數，B表示不同種類的石墨塗布機。

來源	DF	Anova SS	均方	F 值	Pr > F
D	2	0.20666667	0.10333333	1.15	0.3783
B	3	1.53000000	0.51000000	5.67	0.0348

p-value=0.0348則至少要α>0.0348才能拒絕H_0。

結論:

至少要顯著水準α>0.0348，我們才有充分證據認為不同種類的石墨塗佈機對燈箱讀數有顯著的影響。

2. (20 pt.) Do Problem 4.2.p126進行正交比對分析

欲檢定:

H_{01}:$\mu_M - \mu_A$=0

H_{11}:$\mu_M - \mu_A \neq 0$

H_{02}:$\mu_A - \mu_K$=0

H_{12}:$\mu_A - \mu_K \neq 0$

H_{03}:$\mu_K - \mu_L$=0

H_{13}:$\mu_K - \mu_L \neq 0$，μ_j表示不同種類的石墨塗布機燈箱讀數的平均值。

訂立正交比對:

機種	M	A	K	L	
	$T_{.1}$	$T_{.2}$	$T_{.3}$	$T_{.4}$	和
C_1	1	-1	0	0	0
C_2	0	1	-1	0	0
C_3	0	0	1	-1	0
乘積	0	0	0	0	

對比	DF	對比 SS	均方	F 值	Pr > F
12	1	0.24000000	0.24000000	2.57	0.1475
23	1	0.42666667	0.42666667	4.57	0.0650
34	1	0.32666667	0.32666667	3.50	0.0983

結論:

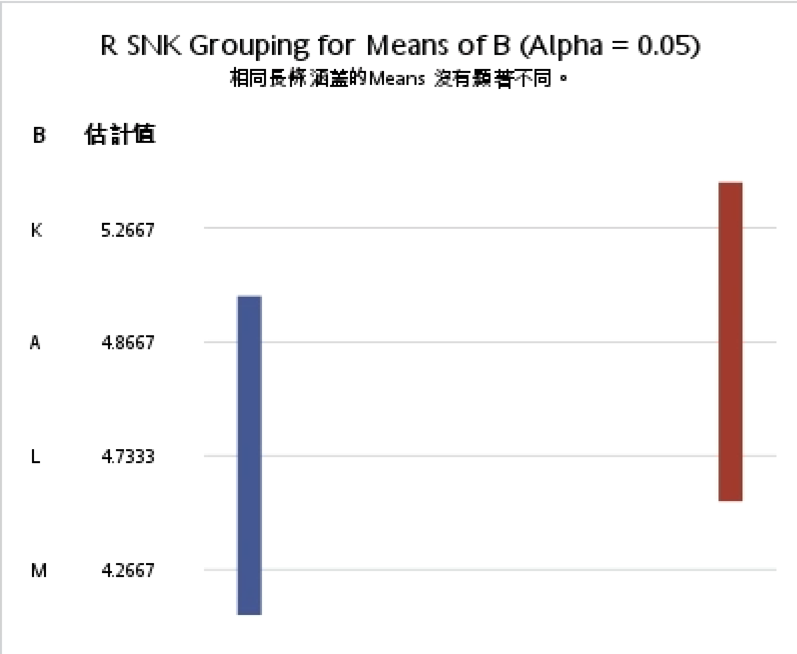
1.p-value>0.1，機種M與A的燈箱讀數之間沒有顯著差異。

2.至少要顯著水準α>0.0650，機種A與K的燈箱讀數才有顯著差異。

3.至少要顯著水準α>0.0983，機種K與L的燈箱讀數才有顯著差異。

3. (20 pt.) Do Problem 4.3.p126用Newman-Keuls 多重比較法，比較四種塗佈機對燈箱讀數的平均。

因為我們有足夠證據來支持四種不同種類的石墨塗佈機對燈箱讀數有顯著的影響。想知道個別處理間是否存在差異。



結論:

在顯著水準α=0.05下，機種K、A、L的燈箱讀數之間沒有顯著差異;機種A、L、M的燈箱讀數之間沒有顯著差異。

4. (20 pt.) Do Problem 4.4p126在K類的第二天資料值遺失下進行分析，需補遺失值。

遺失值的處理:

$SS_{error} = SS_{total} - SS_{treament} - SS_{block}$

$$= \sum_i \sum_j Y_{ij}^2 - \sum_j \frac{T_{.j}^2}{n} - \sum_j \frac{T_{i.}^2}{k} + \frac{T_{..}^2}{nk}$$

SS_{error}=(249.8+y^2)

$$-\frac{163.84+213.16+(y+10.6)^2+201.64}{3} - \frac{338.56+(y+14.4)^2+376.36}{4} + \frac{(y+52.2)^2}{12}$$

$\rightarrow \frac{d(SS_{error})}{dy}$=2y-$\frac{2(y+10.6)}{3}$-$\frac{2(y+14.4)}{4}$+$\frac{2(y+52.2)}{12}$=0

→y≈ 5.5667

(前提與4.1一致，將種類K的石墨塗佈機第二天所缺失的燈箱讀數設為5.5667。)

檢定假設:

設τ_j第j個種類的石墨塗布機的處理效果。

H_0:τ_1=τ_2=τ_3=τ_4=0

H_1:至少有一個τ_j≠0

來源	DF	平方和	均方	F 值	Pr > F
模型	5	2.23273537	0.44654707	5.67	0.0284
誤差	6	0.47277778	0.07879630		
已校正的總計	11	2.70551315			

來源	DF	Anova SS	均方	F 值	Pr > F
D	2	0.31464148	0.15732074	2.00	0.2164
B	3	1.91809389	0.63936463	8.11	0.0156

由於SAS無法自動辨識遺失值的調整(誤差需自由度多減一)，需要手動修正MSE、F值、P值:

$MS_{error}=\frac{SS_{error}}{5}$=0.09455555

F值=$\frac{MS_{treament}}{MS_{error}}\approx 6.76$

P值:利用google 試算表 =F.DIST.RT(6.76, 3, 5) 得p-value≈ 0.03283

結論:

經過遺失值修正後，至少要在顯著水準$\alpha>0.03283$下，不同種類的石墨塗佈機間的燈箱讀數才有顯著差異。

5. (20 pt.) Do Problem 4.25.p130 特定三個檢驗員皆一同採用特定三個磅秤，測量特定三個供應商的材料重量。檢定檢驗員、磅秤和供應商三者帶來重量上的不同。

前提假設:

model:

$Y_{ijk}= \mu+ \beta_i +\tau_j + \gamma_k +\varepsilon_{ijk}$

Y_{ijk}為第i個檢驗員用第k個磅秤所測得第j個供應商的材料重量。

μ為材料重量的共同效果。

β_i為檢驗員的區集效果。

τ_j為供應商處理效果。

γ_k為磅秤的區集效果。

ε_{ijk}為第i個檢驗員用第k個磅秤所測得第j個供應商材料重量的誤差和

因為檢測員、磅秤與供應商皆為特定，並非從母體抽出的結果。則判斷為三個固定效果:

$$\sum_{i=1}^{3}\beta_i = 0$$

$$\sum_{j=1}^{3}\tau_j = 0$$

$$\sum_{k=1}^{3}\gamma_k = 0$$

$$\varepsilon_{ijk}\sim NID(0,\sigma^2)$$

檢定假設:
依題檢定三個效果個別帶來的差異:
a.檢驗員

H_0:$\beta_1=\beta_2=\beta_3=0$

H_1:至少有一個$\beta_i\neq0$

p-value=0.5>α不拒絕H_0。

b.供應商

H_0:$\tau_1=\tau_2=\tau_3=0$

H_1:至少有一個$\tau_j\neq0$

p-value=0.02，則至少要α>0.02才能拒絕H_0

c.磅秤

H_0:$\gamma_1=\gamma_2=\gamma_3=0$

H_1:至少有一個$\gamma_k\neq0$

p-value=0.0067<0.01拒絕H_0。

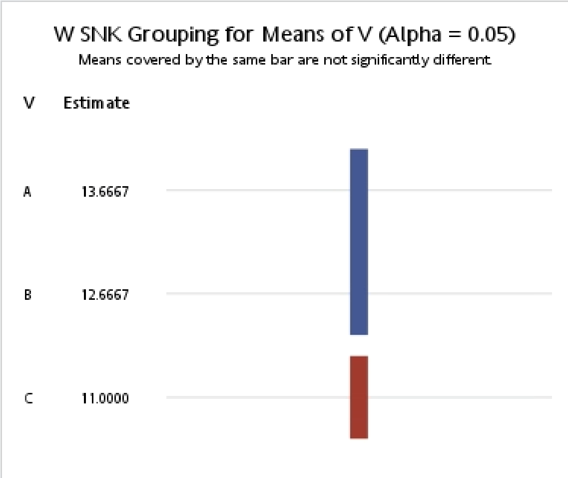
變異來源:I表示檢測員;S表示磅秤;V表示供應商。

Source	DF	Type III SS	Mean Square	F Value	Pr > F
I	2	0.22222222	0.11111111	1.00	0.5000
S	2	32.88888889	16.44444444	148.00	0.0067
V	2	10.88888889	5.44444444	49.00	0.0200

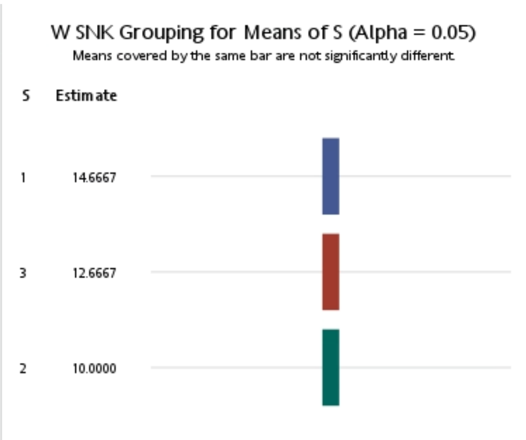
因為我們有足夠證據來支持磅秤、供應商對材料重量有顯著差異。
若要更進一步了解不同供應商和磅秤的差異:

用SAS進行Student-Newman-Keuls 多重比較法。

供應商:



磅秤:



結論:

a.我們沒有充分證據顯示這三個檢驗員所觀測的材料重量有顯著差異。

b.顯著水準α至少要>0.02，才我們有充分證據顯示這三個供應商所提供的材料重量有顯著差異。其中供應商A、B所提供的材料重量與供應商C所提供的在顯著水準α=0.05下有顯著差異。

c.我們有充分證據顯示這三個磅秤所測量的材料重量有顯著差異。其中，在顯著水準α=0.05下，磅秤1號、2號、3號之間所測量的材料重量都有顯著差異。

SAS:

```
data d1;
input B $ D R@@;
cards;
M 1 4.0 M 2 4.8 M 3 4.0
A 1 4.8 A 2 5.0 A 3 4.8
K 1 5.0 K 2 5.2 K 3 5.6
L 1 4.6 L 2 4.6 L 3 5.0
;
run;
proc anova;
class D B;
model R=D B;
means B/SNK;
run;

proc GLM;
class B;
model R=B;
contrast '12' B 1 -1 0 0;x
contrast '23' B 0 1 -1 0;
contrast '34' B 0 0 1 -1;
run;

data d44;
input B $ D R@@;
cards;
M 1 4.0 M 2 4.8 M 3 4.0
A 1 4.8 A 2 5.0 A 3 4.8
K 1 5.0 K 2 5.5667 K 3 5.6
L 1 4.6 L 2 4.6 L 3 5.0
;
run;
proc anova;
class D B;
model R=D B;
run;

data d425;
input I$ V$ S$ W@@;
CARDS;
1 A 1 16 1 B 2 10 1 C 3 11
2 B 1 15 2 C 2 9 2 A 3 14
3 C 1 13 3 A 2 11 3 B 3 13
;
RUN;
PROC glm;
class I V S;
MODEL W= I S V;
MEANS S V/SNK;
RUN;
```