Experimental Design and Analysis Assignment 1

110024516 邱繼賢

Problem 1.

(a)

Fit statistical model :

$$y_{ij(k)} \ = \ \eta \ + \ \alpha_i \ + \ \beta_j \ + \ \gamma_{jk} \ + \ \epsilon_{ij(k)} \quad , \quad \epsilon_{ij(k)} \ \overset{iid}{\sim} \ N(0, \sigma^2)$$

where

$$\begin{cases} \alpha_i \text{ is the keyboard (treatment) effect,} & i \ = \ A, B \\ \beta_j \text{ is the manuscript (block) effect,} & j \ = \ 1, 2, ..., 6 \\ \gamma_{ij(k)} \text{ is the learning (block) effect,} & i \ = \ A, B \ , \ j \ = \ 1, 2, ..., 6 \ , \ k \ = \ 1 \text{ (first order)}, \ 2 \text{ (second order)} \end{cases}$$

我們假設 learning effect 並不會因為 manuscript 的差別而有所不同：

$\gamma_{ij(1)} \ = \ 0$ which has no learning effect, $\gamma_{ij(2)} \ = \ L$ which has L learning effect for all $j = 1, ..., 6$.

此題我們主要想要探討的是不同的 keyboard (treatment) 之間的差異性，即

$$\alpha_A \ - \ \alpha_B \ = \ (y_{Aj(k)} - y_{Bj(k)}) \ + \ (\gamma_{Aj(k)} - \gamma_{Bj(k)}) \ + \ (\epsilon_{Aj(k)} - \epsilon_{Bj(k)})$$

可以使用 $\frac{1}{6}\sum_{j=1}^{6}(y_{Aj(k)} - y_{Bj(k)})$ 統計量做為估計式來估計 $\alpha_A - \alpha_B$，以下分為不同情況來討論：

(1) **Without randomization :**

AB, AB, AB, AB, AB, AB all A's are in the first order (k=1)

$$\begin{cases} y_{Aj(1)} \ = \ \eta \ + \ \alpha_A \ + \ \beta_j \ + \ \epsilon_{Aj(1)} \\ y_{Bj(2)} \ = \ \eta \ + \ \alpha_B \ + \ \beta_j \ + \ L \ + \ \epsilon_{Bj(2)} \end{cases}$$

$$\Rightarrow \ E\left[\frac{1}{6}\sum_{j=1}^{6}(y_{Aj(1)} - y_{Bj(2)})\right] \ = \ \alpha_A \ - \ \alpha_B \ - \ L$$

which is a biased estimator with $bias \ = \ L$

(2) **With randomization :**

AB, BA, AB, BA, AB, AB

$$\begin{cases} y_{Aj(k)} \; = \; \eta \; + \; \alpha_A \; + \; \beta_j \; + \; L \times I(j=2,4) \; + \; \epsilon_{Aj(k)} \\ y_{Bj(k)} \; = \; \eta \; + \; \alpha_B \; + \; \beta_j \; + \; L \times I(j=1,3,5,6) \; + \; \epsilon_{Bj(k)} \end{cases}$$

$$\Rightarrow \; E\left[\frac{1}{6}\sum_{j=1}^{6}(y_{Aj(k)} - y_{Bj(k)})\right] \; = \; \alpha_A \; - \; \alpha_B \; - \; \frac{1}{3}L$$

which is a biased estimator with $bias \; = \; \frac{1}{3}L$ smaller than without randomization case above.

(3) **Balanced randomization**

AB, BA, BA, AB, BA, AB

$$\begin{cases} y_{Aj(k)} \; = \; \eta \; + \; \alpha_A \; + \; \beta_j \; + \; L \times I(j=2,3,5) \; + \; \epsilon_{Aj(k)} \\ y_{Bj(k)} \; = \; \eta \; + \; \alpha_B \; + \; \beta_j \; + \; L \times I(j=1,4,6) \; + \; \epsilon_{Bj(k)} \end{cases}$$

$$\Rightarrow \; E\left[\frac{1}{6}\sum_{j=1}^{6}(y_{Aj(k)} - y_{Bj(k)})\right] \; = \; \alpha_A \; - \; \alpha_B$$

which is an unbiased estimator.

(b) 我不會使用題目的序列來進行實驗，雖然它符合 balanced randomization 的特性，但是它有著一種特別的順序，如果 learning effect 不為一固定常數，而是隨著時間推進而遞增或遞減的函數，則使用此序列做出的估計值將不再 unbiased ，因此我會選擇重新做一次 balanced randomization 將 AB 和 BA 更加平均分散在時間線上，但若是 learning effect 已經可以確定為一固定常數，則使用此序列所估計出的估計式依舊不偏。

Problem 2.

(1) What are the differences between their observations?

此實驗共有 2 個 factor，其各別有 2 個 level，所以總共有 $2 \times 2 = 4$ 個 treatment，在 Group 1 中，每個 treatment 只被分配到一個 experimental unit，然後由兩位 students 重複進行觀察，這樣的操作方式稱為 repetition，其結果的變異來自於測量誤差和學生間的變異。

在 Group 2 中，每個 treatment 被分配到了兩個 experimental units，然後由兩個 students 各別進行觀察，這樣的操作方式稱為 replication，其結果的變異來自於測量誤差、學生間的變異和實驗單位間的變異。

(2) In what situation would one group obtain more accurate results than the other?

如果此實驗的主要變異來自於學生間的變異而不是實驗單位的變異，則 Group 1 的實驗方法可以較為精準，因為相同的 experimental unit 和 treatment 下，被不同的學生進行 repetition 可以降低學生所帶來的變異。

如果此實驗的主要變異來自於實驗單位的變異而不是學生間的變異，則 Group 2 的實驗方法可以較為精準，因為對不同的 units 進行 replication，增加 experimental unit 的數量可以降低其所帶來的變異，雖然是由不同的學生進行觀察，會把學生間的變異一起混淆進去，但在現在的假設下，學生間的變異並不是主要影響因素。

(3) Is there a block factor?

有。block factor 是指會對實驗結果帶來變異，但不是我們主要想探討的變數，在此實驗中就是不同學生觀測數據所帶來的變異。

Problem 3.

觀察 factor 設定為 (+,+) 的實驗，可以發現只有這組設定值的時候 setting 和 reading 都是由 A 來完成，這樣沒有對於 student 這個 block factor 做 balanced randomization，會造成如果 A 的 setting 或 reading 本身就有著較高或較低的偏差，他所做出的結果會 biased 掉實驗真正的結果，同理這種學生本身的偏差造成的實驗結果變異，也同樣會發生在實驗失敗後再由同一個學生重新進行實驗。另外讀取自己 setting 後的實驗結果，也可能會有較高的傾向主觀認為自己的實驗是成功的，進而對實驗結果產生偏差。

以下為我基於以上幾點原因對原本的實驗進行修改：

Run	Factor One	Factor Two	Setting	Reading
1	+	+	A	B
2	+	+	B	A
3	+	-	A	B
4	+	-	B	A
5	-	+	B	A
6	-	+	A	B
7	-	-	A	B
8	-	-	B	A

對於每個 treatment 做的兩次 replicate 都是不同的學生組合，同一次實驗中 setting 和 reading 的學生都是不同人，若這次實驗失敗了，則交換 setting 和 reading 的角色，然後重新進行一次實驗。