

homework 01 for design of experiment

106354005 統碩一 余佑駿

Exercise 2.3

Suppose we are studying the effect of diet on height of children, and we have two diets to compare: **diet A (a well balanced diet with lots of broccoli)** and **diet B (a diet rich in potato chips and candy bars)**. We wish to find the diet that helps children grow (in height) fastest. We have decided to use **20 children** in the experiment, and we are contemplating the following methods for matching children with diets:

1. Let them choose.
2. Take the first 10 for A, the second 10 for B.
3. Alternate A, B, A, B.
4. Toss a coin for each child in the study: heads \rightarrow A, tails \rightarrow B.
5. Get 20 children; choose 10 at random for A, the rest for B.

Describe the benefits and risks of using these five methods.

Answer :

第二章中有講到很多種隨機化的方式，這一題主要就是考慮這五種方法隨機化的好壞。而我認為第四和第五種方法比起其他三種要來的好，以下是我分析的優缺點。

	優點	缺點
方法1	無	這個方法並不隨機，小朋友可能依照自己想要的洗好選擇diet，或者是不同年齡層的小孩會有不同的方案喜好，因此無法避免confounding的發生。
方法2	兩個diet方案的人數是一致的。	這個方法並不隨機，和上一題一樣無法避免confounding的發生，因為抵達時間可能會影響diet方案的分配。
方法3	兩個diet方案的人數是一致的。	因為依序分配的關係，所以研究人員可以預測到哪個孩子被分配到哪種diet方案，因此沒辦法保證實驗blinding，可能會產生biased。
方法4	在硬幣是公正的前提之下，這種方法是隨機的，因此可以避免confounding的發生，所以實驗的結果也可以做推論。	diet方案A、B的人數不一定一致。
方法5	這種方法是最好的，分配方案的方式隨機，且人數也相同。	無。

Exercise 2.5 設定 $\alpha = 0.05$

H_0 : the chlordane treatment has **NOT** affected the $\text{Na}^+ - \text{K}^+ \text{ATP}$ as activity. 即 $\mu = 0$

$H_1: \mu \neq 0$

題目是問說什麼樣的 experimental technique 必須一定要被使用，來驗證這是一個隨機試驗。

課文中有一段：

...in the randomization analogue to a paired t-test, the absolute values of the differences are taken to be fixed, and the signs of the differences are random, with each sign independent of the others and having equal probability of positive and negative.

因此，以下是要使用的 experimental technique

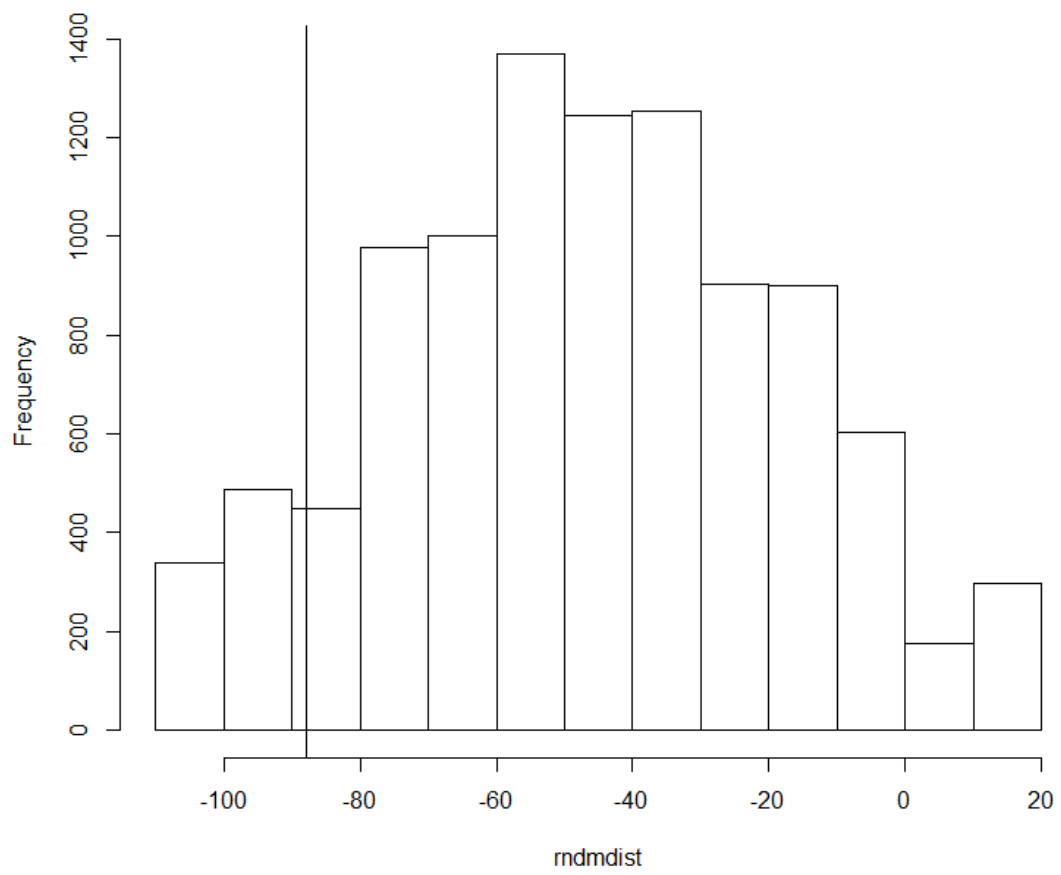
第一步：先用 sample 這個函數隨機生成1, -1, ...，將之和蟑螂的 percentage 相乘，每做出一組六個數字就相加，這樣的步驟重複 10000 次。

第二步：原本數據的總和當作當作臨界值，求出模擬的 10000 次總和中，比臨界值小的比例有多少個，作為 $p - value$

模擬出來的 $p - value = 0.0897$ ，故不拒絕 H_0 ，殺蟲劑中的化學藥劑沒有顯著效果。

```
1 roaches <- c(15.3, -31.8, -35.6, -14.5, 3.1, -24.5)
2
3 # 做出倒數的函數
4 reciprocal <- function(a){
5   a*(-1)
6 }
7
8 rndmdist = NULL
9 for (i in 1:10000) {
10   s1 = sample(c(15.3, reciprocal(0.950)),1,T)
11   s2 = sample(c(-31.8, reciprocal(0.978)),1,T)
12   s3 = sample(c(-35.6, reciprocal(0.762)),1,T)
13   s4 = sample(c(-14.5, reciprocal(0.733)),1,T)
14   s5 = sample(c(3.1, reciprocal(0.823)),1,T)
15   s6 = sample(c(-24.5, reciprocal(1.011)),1,T)
16
17   x = c(s1,s2,s3,s4,s5,s6)
18
19   rndmdist = c(rndmdist,sum(x))
20 }
21 # 6個數據，有倒數與否的隨機可能
22 # 模擬10000次，每次模擬就把六個數據做加總
23
24 t0 = sum(roaches)
25 t0
26
27 sum(rndmdist <= t0)/length(rndmdist)
28
29 hist(rndmdist)
30 abline(v=t0)
```

Histogram of rndmdist



Problem 2.1

和上一題的道理差不多，虛無假設為儀器 Visiplume 和標準儀器的測量是一致的。

所以 0.95、0.978、0.762、0.733、0.823、1.011，共有 2^6 種可能(倒數與否)，做 10000 次模擬，把原本數據加總作為臨界值，看 10000 次中，小於等於臨界值的比例有多少，即為 p-value 值，並劃出直方圖

因為 p-value = 0.0309

所以拒絕虛無假設，且儀器 Visiplume 的測量應該會比標準的測量較差。(題目說 higher than standard 感覺有誤)

```
1 x1 = c(0.950,0.978,0.762,0.733,0.823,1.011)
2
3 # 做出倒數的函數
4 reciprocal <- function(a){
5   1/a
6 }
7
8 rndmdist = NULL
9 for (i in 1:10000) {
10   s1 = sample(c(0.950, reciprocal(0.950)),1,T)
11   s2 = sample(c(0.978, reciprocal(0.978)),1,T)
12   s3 = sample(c(0.762, reciprocal(0.762)),1,T)
13   s4 = sample(c(0.733, reciprocal(0.733)),1,T)
14   s5 = sample(c(0.823, reciprocal(0.823)),1,T)
15   s6 = sample(c(1.011, reciprocal(1.011)),1,T)
16
17   x = c(s1,s2,s3,s4,s5,s6)
18
19   rndmdist = c(rndmdist,sum(x))
20 }
21 # 6個數據，有倒數與否的隨機可能
22 # 模擬10000次，每次模擬就把六個數據做加總
23
24 t0 = sum(x1)
25 t0
26
27 sum(rndmdist <= t0)/length(rndmdist)
28
29 hist(rndmdist)
30 abline(v=t0)
```

