

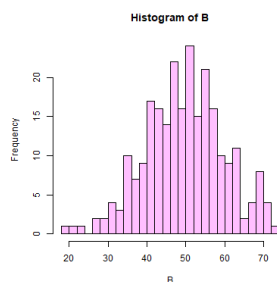
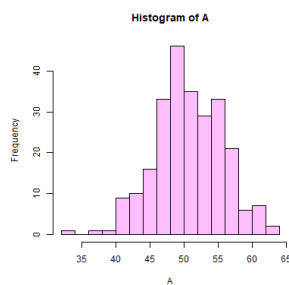
統計分析法 第6週レポート

202212022 田島瑞起

2023/11/21

1 設問 1

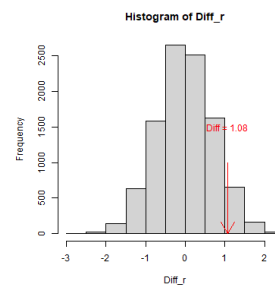
A,B についてのヒストグラムを下記の図で示す。



2 設問 2

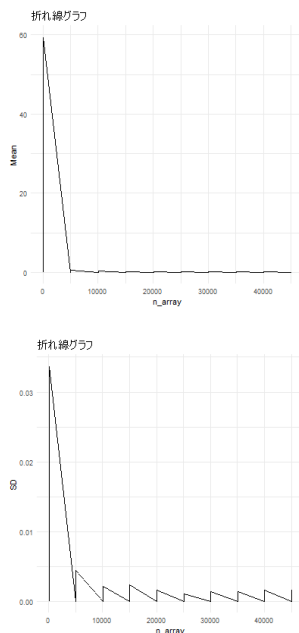
A 群のデータ平均値が B 群のデータ平均値よりも大きいことを示すために、A と B を統合したデータから 10,000 回のサンプリングを実行し、サンプリングしたデータを二分割し、平均値の差をとり、平均値の分布

上で実測値がどれほどの確率で起こっているか確かめ、その後有意水準を満たしているか確認する。この方法で p を算出した結果、 $p=0.0656$ となり、A の平均値は B の平均値よりも大きいという帰無仮説が支持されることが分かった。また、理想分布及び実測値は下の図で示す。



3 設問 3

設問 2 で使用したコードを編集して、 $n=50$ における p の平均値は 13.086 標準偏差は 0.0365 と算出された。次に n の値を 50 50000 まで変化させたとき p の平均値及び標準偏差がどのように推移するか確認する。 n の値の間隔は 5000 とし実行した結果が下記の図になる。今回の課題では計算量を考慮した結果 n のサンプリング間隔が大きくなってしまったため、正確な値は不明であるが $n=5000$ あたりから結果が安定して



いるため、その付近の値でサンプリングすれば良いと考えられる。

4 設問 4

perm を利用して検定したところ、 $p=0.06709$ と算出され、これは優位水準を満たす値なので、A のデータのほうが B のデータよりも大きいことが分かる。t-test では、 p は有意水準を満たしており、2つのデータに差は無いと結論できる。今課題で個人的に考えたのは、ランダムイゼーションは、2群を混ぜてからサンプリングすることによって、偏りを減らすことにより、2群の差異を t-test よりもシビアに判定することが出来るのではないかと考えた。

5 ソースコード

図 1 s2212022-1.c

```
1 #課題1
2 ReadData <- read.table('data_r.txt')
3 A <- ReadData$V1
```

```
4 B <- ReadData$V2
5 A <- A[2:length(A)]
6 B <- B[2:length(B)]
7 A <- as.numeric(A)
8 B <- as.numeric(B)
9 png("6-1.png", width = 400, height =
  400)
10 hist(A,breaks=20,col='#ff00ff40')
11 dev.off()
12 png("6-2.png", width = 400, height =
  400)
13 hist(B,breaks=20,col='#ff00ff40')
14 dev.off()
15 #課題2
16 Diff <- mean(A) - mean(B)
17 n <- 9999
18 Diff_r <- numeric(n+1)
19 Data <- c(A,B)
20 for(i in 1:n){
21   data_r <- sample(Data,replace=F)
22   A_r <- data_r[1:length(A)]
23   B_r <- data_r[(length(A)+1):length(
     data_r)]
24   Diff_r[i] <- (mean(A_r) - mean(B_r))
25 }
26 Diff_r[n+1] <- Diff
27 p <- sum(Diff_r >= Diff)(n+1)
28
29 png("6-3.png",width=400,height=400)
30 hist(Diff_r)
31 arrows(x0 = Diff, y0 = 0, x1 = Diff, y1
  = 1000, col = "red", code = 1, angle
  = 30)
32 text(Diff,1500,labels = paste("Diff_=",
  round(Diff, 2)), col = "red")
33 dev.off()
34
35 #課題3
36 n <- 49
37 p_array <- numeric(10)
38 for(j in 1:10){
39   for(i in 1:n){
40     data_r <- sample(Data,replace=F)
41     A_r <- data_r[1:length(A)]
42     B_r <- data_r[(length(A)+1):
       length(data_r)]
43     Diff_r[i] <- (mean(A_r) - mean(
       B_r))
44   }
45   Diff_r[n+1] <- Diff
46   p <- sum(Diff_r >= Diff)/(n+1)
47   p
48   p_array[j] <- p
49 }
```

```

50 mean(p_array)
51 sd(p_array)
52
53 n_array <- seq(50,50000,5000)
54 mean_array <- numeric(length(n_array))
55 sd_array <- numeric(length(n_array))
56 for(k in n_array){
57   n <- k
58   p_array <- numeric(10)
59   for(j in 1:10){
60     for(i in 1:n){
61       data_r <- sample(Data,
62         replace=F)
63       A_r <- data_r[1:length(A)]
64       B_r <- data_r[(length(A)+1):
65         length(data_r)]
66       Diff_r[i] <- (mean(A_r) -
67         mean(B_r))
68     }
69     Diff_r[n+1] <- Diff
70     p <- sum(Diff_r >= Diff)/(n+1)
71     p_array[j] <- p
72   }
73   mean_array <- append(mean_array,mean
74     (p_array))
75   sd_array <- append(sd_array,sd(
76     p_array))
77 }
78
79 install.packages("ggplot2")
80 # ライブラリの読み込み
81 library(ggplot2)
82 # グラフの作成
83 df <- data.frame(n = n_array, mean =
84   mean_array, sd = sd_array)
85 # 折れ線グラフの描画
86 png("6-4.png",width=400,height=400)
87 ggplot(df, aes(x = n, y = mean)) +
88   geom_line() +
89   labs(title = "折れ線グラフ",
90     x = "n_array",
91     y = "Mean") +
92   theme_minimal()
93
94 dev.off()
95
96 png("6-5.png",width=400,height=400)
97 ggplot(df, aes(x = n, y = sd)) +
98   geom_line() +
99   labs(title = "折れ線グラフ",
100     x = "n_array",
101     y = "SD") +
102   theme_minimal()
103
104 dev.off()

```

```

98
99 #課題4
100
101 # permパッケージをインストールして読み込
102   む
103 install.packages("perm")
104 library(perm)
105 # permTS()関数を使用して順位和検定を実行
106 perm_test_result <- permTS(A, B,
107   alternative = "greater")
108
109 # 結果の表示
110 print(perm_test_result)
111
112 # t検定を実行
113 t_test_result <- t.test(A, B,
114   alternative = "greater")
115
116 # 結果の表示
117 print(t_test_result)

```