

Assignment 9 STAT 581

Tianzhi Cao tc324
Abhishek Bhatt ab2016
Siva Harshini Dev Bonam sdb202

November 3, 2020

Problem

Write a program to identify potentially interesting discoveries from a collection of P values

Input Vector of P values, indicator T or F for independence

Output, on following page

Numbers corresponding to experiments which are potentially interesting

Graph as in following page (lecture 14)

Answer

step 1.Sort P values

step 2.Count tests

step 3.Set Q

step 4.Plot sorted P-values(smallest to largest) vs line $Q * c(1 : m)/m$ (If not independent, $Q * c(1 : m)/(m * (\sum(1/i) i = 1, \dots, m))$)

step 5.Find P^* =largest P value \downarrow line

step 6.Every $P_i = P^*$ is “interesting”

```
1 FDR<-function(vec,q=0.05,ind=1){
2   n=length(vec)
3   qline=c(1:n)*q/n
4   argsort=order(vec)
5   vec=sort(vec)
6   rtn=vector()
7   bkpoint=0
8   if (ind==0){
9     qline=c(1:n)*q/(n*sum(1/c(1:n)))
10  }
11  for (i in n:1){
12    if (qline[i]>vec[i]){
13      bkpoint=i
14      break
15    }
16  }
17  if (bkpoint!=0){
```

```

18     for (i in 1:bkpoint){
19         rtn=c(rtn, argsort[i])
20     }
21 }
22 cat(rtn, "\n")
23 cat("FDR=", length(rtn)/n)
24 plot(c(1:n), vec, pch=16, cex=0.5, xlab="data_rank", ylab="p_value")
25 points(c(1:n), qline, pch=16, cex=0.5, col="red")
26 legend("topright", lty=c(0,0), pch=16, col=c("black", "red"), legend=c
27       ("sorted_p", "qline"))

```

Independent Result:

```

1 > FDR(c(0.0001*runif(100), runif(900)), 0.05)
2 45 96 73 36 99 68 28 84 77 3 69 100 59 54 62 94 31 81 9 42 95 26 5
   10 4 30 71 33 72 46 70 44 8 88 82 12 66 74 80 49 52 89 63 83 29
   90 35 50 86 98 27 78 43 61 39 24 25 58 19 93 64 2 97 7 67 41
   20 32 87 1 53 92 17 75 6 16 55 51 57 14 37 34 40 56 13 60 15 76
   23 79 85 47 18 22 65 38 48 11 21 91 383 375 467 769 305 788
3 FDR= 0.106

```

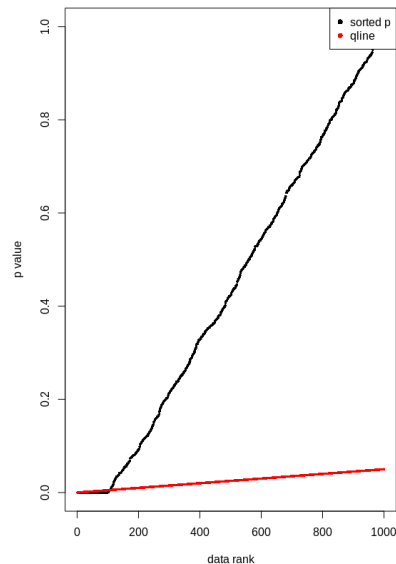


Figure 1: Independent P-value Result.

Dependent Result:

```
1 > FDR(c(0.000001*runif(100),runif(900)),0.05,0)
2 18 60 79 89 91 11 67 84 43 39 14 80 20 74 62 19 66 44 30 56 23 78
   40 34 13 24 6 38 75 12 58 69 63 1 64 96 72 8 9 55 7 77 87 73 81
   15 47 93 95 83 37 35 28 48 4 54 52 99 5 36 88 65 41 3 94 50 82
   98 22 27 71 2 32 33 25 42 16 21 17 45 57 31 92 70 97 86 90 76
   29 100 61 68 85 26 59 46 53 49 10 51 291
3 FDR= 0.101
```

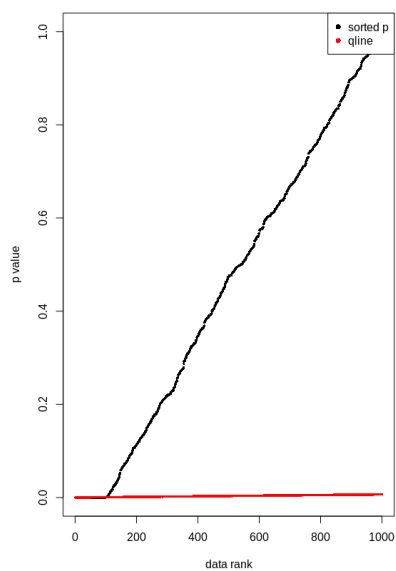


Figure 2: Dependent P-value Result.