Advanced Statistical Methods Hw9

Do Hyup Shin

2021-12-12

Problem 15.1

Show that Holm's procedure (15.10) is more generous than Bonferroni in declaring rejections.

Solution

Let $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(N)}$ be ordered p-values and $I_0 = \{i : H_{0(i)} \text{ is true}\}$. Define $i_0 = min\{i : p_{(i)} > \alpha/(N-i+1)\}$. Holm's procedure reject for all $i < i_0$, which means $p_{(i)} \leq \alpha/(N-i+1) \forall i < i_0$.

We know that $\frac{\alpha}{N-i+1} \geq \frac{\alpha}{N}$ $\forall i \in \{1, 2, \ldots, N\}$. Thus, the cut off value of Holm's procedure is larger than Bonferroni. So Holm;s procedure is more generous than Bonferroni in declaring rejections.

Problem 15.2

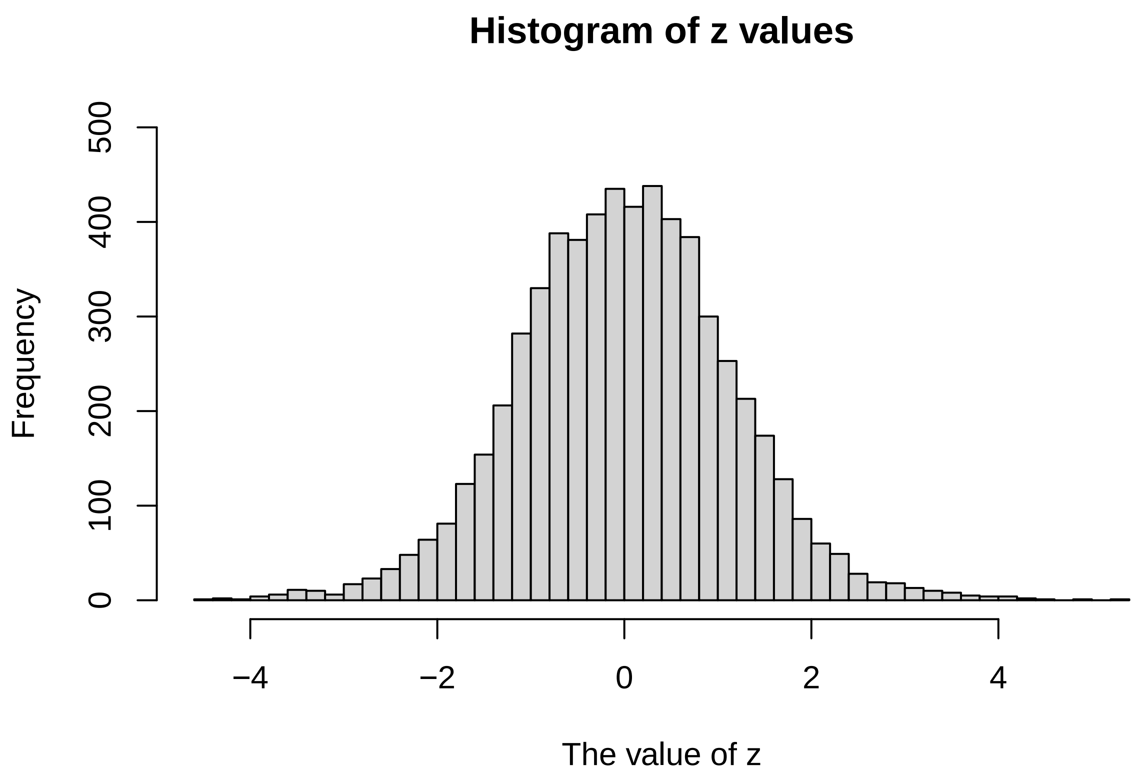
Redraw Figure 15.3 for q = 0.2.

Solution

Let q = 0.2 and N = 6033. Then, there are N z-values and p-values corresponding to z-value. Define the ordered z-values such that $z_{(1)} \leq z_{(2)} \leq \cdots \leq z_{(N)}$ and the ordered p-values such that $p_{(1)} \geq p_{(2)\geq\cdots\geq p_{(N)}}$ where $p_{(i)}$ is corresponding $z_{(i)}$.

First, we draw the histogram of z-values.

```
pros_data <- read.csv("https://hastie.su.domains/CASI_files/DATA/prostz.txt", sep = " ")
hist(pros_data[,1], xlab = "The value of z", main = "Histogram of z values", breaks = 40, ylim = c(0,50
```



Next, we find the index j such that $j = max\{i : p_{(i)} \leq \frac{i}{N}q\}$ which is BH procedure. Also, we find the index k such that $k = min\{i : p_{(i)} > \frac{q}{N-i+1}\}$ which is Holm's procedure.

```
## Order the p-value
ordered_pv <- pnorm(sort(pros_data[,1], decreasing = T), lower.tail = F)

##Choose the 100 p-value
ordered_pv_100 = ordered_pv[1:100]

##FDR and Holm cut off values
q = 0.2
N = 6033
FDR = 1:100*q/N
Holm = q/(N - 1:100 + 1)

##Find the FDR index and Holm index
index_FDR = max(which(FDR > ordered_pv_100))
index_Holm = min(which(Holm < ordered_pv_100)) - 1
FDR_cor = c(index_FDR, FDR[index_FDR])
Holm_cor = c(index_Holm, Holm[index_Holm])
index_FDR
```

```
## [1] 61
```

```
index_Holm
```

```
## [1] 9
```

```
##z-value of rejection region
round(qnorm(FDR[61], lower.tail =F), 3)
```

```
## [1] 2.875
```

```r
round(qnorm(Holm[9], lower.tail =F), 3)
```

```
## [1] 3.989
```

Thus, $j = 61$ and $k = 10$. So, the FDR control boundary rejects H_{0i} for the 61 smallest values $p_{(i)}$ while Holm's FWER procedure rejects for only the 9 smallest values.

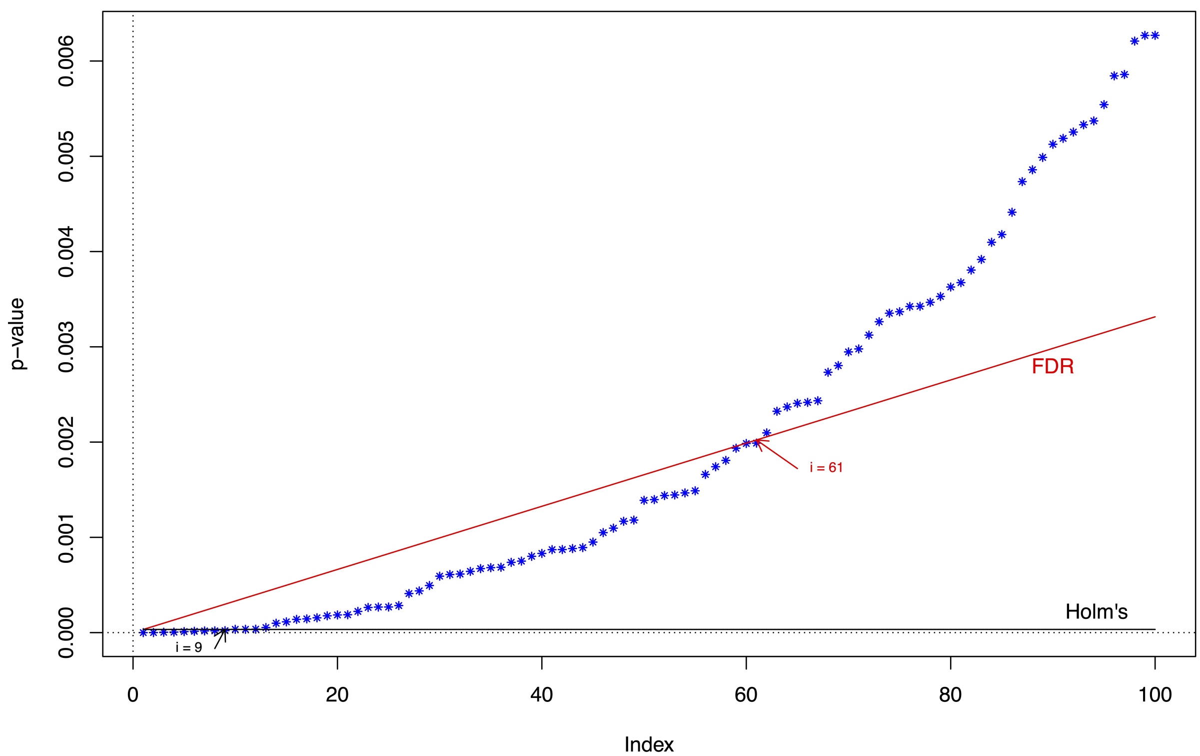
From z-values point of view, the FDR procedure rejects H_{0i} for the 61 largest z-values ($z_i \geq 2.875$),while FWER control rejects only the 9 most extreme z-values ($z_i \geq 3.989$).

So we draw the graphs of each lines.

```r
##Plot the FDR and Holm line
plot(ordered_pv_100, ylab = "p-value", col = "blue", pch = 8)
lines(FDR, type = "l", col= 'red')
text(90, FDR[90],"FDR", col = "red",  pos = 1)
lines(Holm, type = "l", col= 'black')
text(90,0.0002,"Holm\'s", col = "black",  pos = 4)
abline(h = 0, lty = 3)
abline(v = 0, lty = 3)


arrows(FDR_cor[1]+4,  FDR_cor[2]-0.0003, FDR_cor[1], FDR_cor[2], col = 'red', angle = 30, length = 0.1)
text(FDR_cor[1]+4, FDR_cor[2]-0.0003,"i = 61", col = "red",  pos = 4, cex = 0.7)


arrows(Holm_cor[1]-1, Holm_cor[2]-0.0002, Holm_cor[1], Holm_cor[2], col = 'black', angle = 30, length =
text(Holm_cor[1]-1, Holm_cor[2]-0.0002,"i = 9", col = "black",  pos = 2, cex = 0.7)
```


p−value
Index
FDR
Holm's
i = 61
i = 9
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0
20
40
60
80
100

Problem 15.4

For an observed data set of z-values $z_1, z_2, \ldots, z_N$, a case z_i of particular interest just barely made it into the Benjamini–Hochberg $\mathcal{D}_q$ rejection region. Later you find out that 25 of the very negative other z-values were actually positive, and exceed z_i. Is H_{0i} still rejected?

Solution

Let $z_{(1)} \leq z_{(2)} \leq \cdots \leq z_{(N)}$ be ordered z-values of $z_1, z_2, \ldots, z_N$ and $p_{(N)} \leq p_{(N-1)} \leq \cdots \leq p_{(1)}$ be ordered p-value corresponing $z_{(i)}$.

Let j be ordered index of z_i. This means $z_{(j)} = z_i$ and $p_{(j)} = p_i = P(Z \geq z_i)$. By assuption, z_i just barely made it into the Benjamini–Hochberg $\mathcal{D}_q$ rejection region. So j is the largest index for which $p_{(i)} \leq \frac{i}{N}q$. Thus, $p_{(j)} = p_i \leq \frac{j}{N}q$.

Suppose that 25 of the very negative other z-values were actually positive, and exceed z_i. This means that the p-values of the 25 changed values becomes smaller than $p_{(j)} = p_i$. If we reorder the p-value, the index of $p_{(j)} = p_i$ is changed to $p_{(j+25)} = p_i$.

Since we know that $\frac{j}{N}q \leq \frac{j+25}{N}q$, $p_{(j+25)} = p_i \leq \frac{j}{N}q \leq \frac{j+25}{N}q$.

Therefore, H_{0i} is still rejected.