

Biostatistics-Lecture6: Multiple Hypothesis testing

Ruibin Xi

3/16/2020

We perform some simulation to compare different multiple testing procedure. First generate a simulation data where all null hypotheses are true.

```
set.seed(1010093)
pValues <- rep(NA,1000)
for(i in 1:1000){
  y <- rnorm(20)
  x <- rnorm(20)
  pValues[i] <- summary(lm(y ~ x))$coeff[2,4]
}
```

Setting the significance level as $\alpha = 0.05$. If we do not do multiple testing adjustments, we will have many false positives.

```
sum(pValues < 0.05)
```

```
## [1] 51
```

If we control the FWER by Bonferroni, we have much less false positives.

校正

```
sum(p.adjust(pValues,method="bonferroni") < 0.05)
```

```
## [1] 0
```

选择合适的校正方法

control FWER by hochberg

```
sum(p.adjust(pValues,method="hochberg") < 0.05)
```

```
## [1] 0
```

Control FDR by BH

```
sum(p.adjust(pValues,method="BH") < 0.05)
```

```
## [1] 0
```

Control FDR by BY

```
sum(p.adjust(pValues,method="BY") < 0.05)
```

```
## [1] 0
```

Now we assume that 50% of NULL hypothesis are true

```
set.seed(1010093)
pValues <- rep(NA,1000)
for(i in 1:1000){
  x <- rnorm(20)
  # First 500 beta=0, last 500 beta=2
  if(i <= 500){y <- rnorm(20)}else{ y <- rnorm(20,mean=2*x)}
  pValues[i] <- summary(lm(y ~ x))$coeff[2,4]
}
trueStatus <- rep(c("Null","Alternative"),each=500)
```

No multiple testing adjustments.

```
tmp = table(pValues < 0.05, trueStatus)
rownames(tmp) = c("Retain","Reject")
tmp
```

```
##           trueStatus
##           Alternative Null
## Retain           0  476
## Reject          500   24
```

Control FWER,

```
tmp = table(p.adjust(pValues,method="bonferroni") < 0.05,trueStatus)
rownames(tmp) = c("Retain","Reject")
tmp
```

```
##           trueStatus
##           Alternative Null
## Retain           23  500
## Reject          477    0
```

```
tmp = table(p.adjust(pValues,method="hommel") < 0.05,trueStatus)
rownames(tmp) = c("Retain","Reject")
tmp
```

```
##           trueStatus
##           Alternative Null
## Retain           16  500
## Reject          484    0
```

Comparing bonferroni and hommel.

```
p1 = p.adjust(pValues,method="bonferroni")
p2 = p.adjust(pValues,method="hommel")
sum(p2<p1)

## [1] 1000
```

Contorl FDR

```
tmp = table(p.adjust(pValues,method="BH") < 0.05,trueStatus)
rownames(tmp) = c("Retain","Reject")
tmp

##           trueStatus
##           Alternative Null
##   Retain           0  487
##   Reject          500  13
```

也比原来假阳性的数量少了，也算控制住
了type 1 error

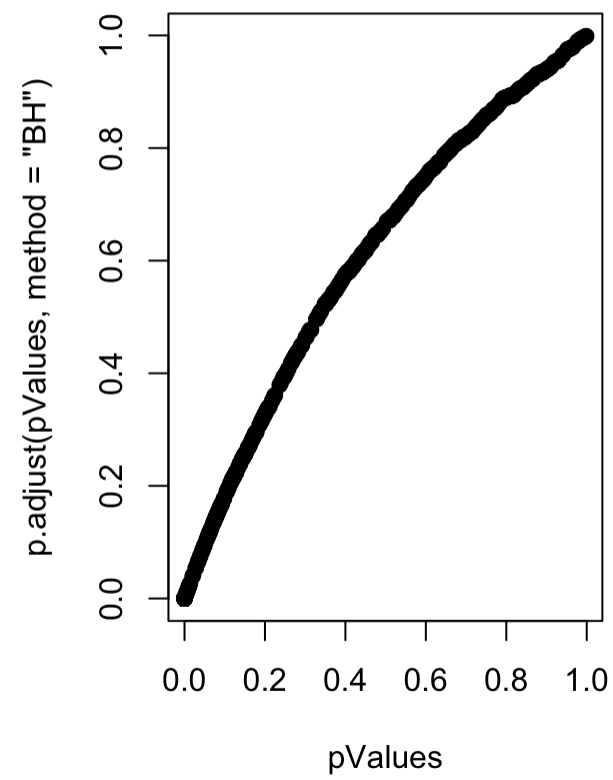
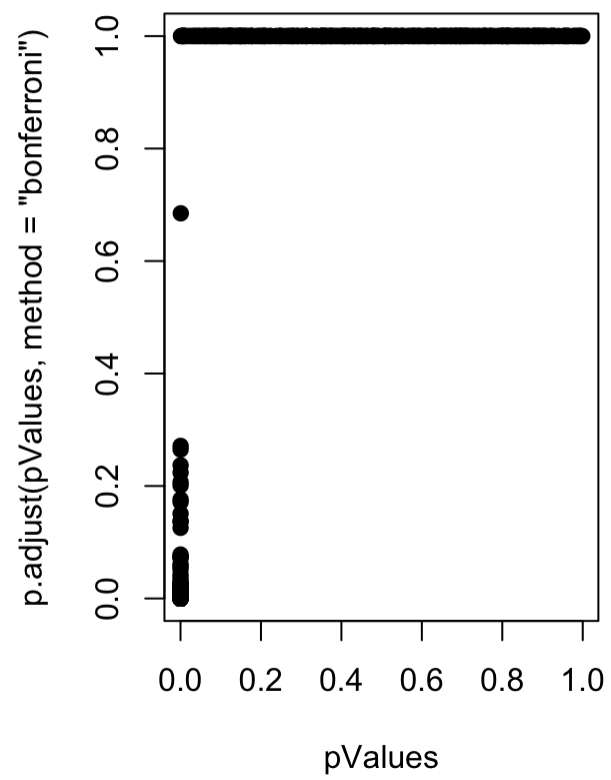
```
tmp = table(p.adjust(pValues,method="BY") < 0.05,trueStatus)
rownames(tmp) = c("retain","reject")
tmp

##           trueStatus
##           Alternative Null
##   retain           2  498
##   reject          498   2
```

不知道是否独立的情况下用BY

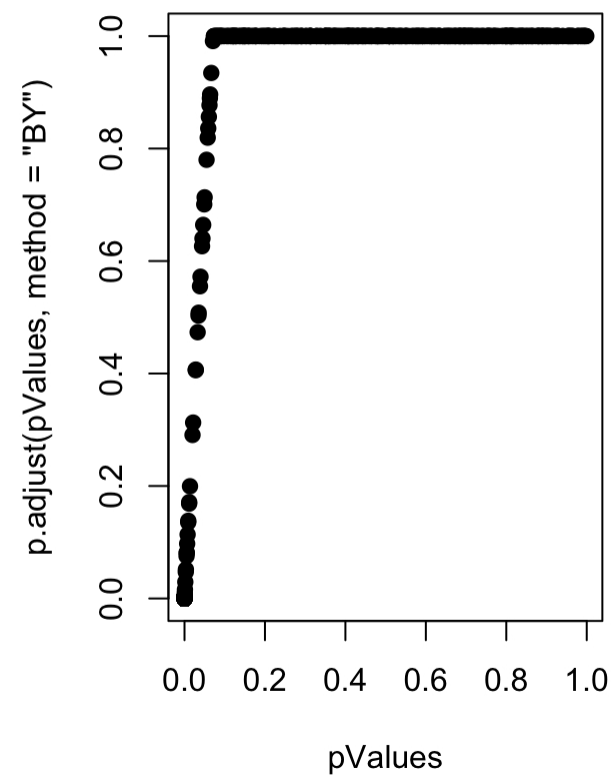
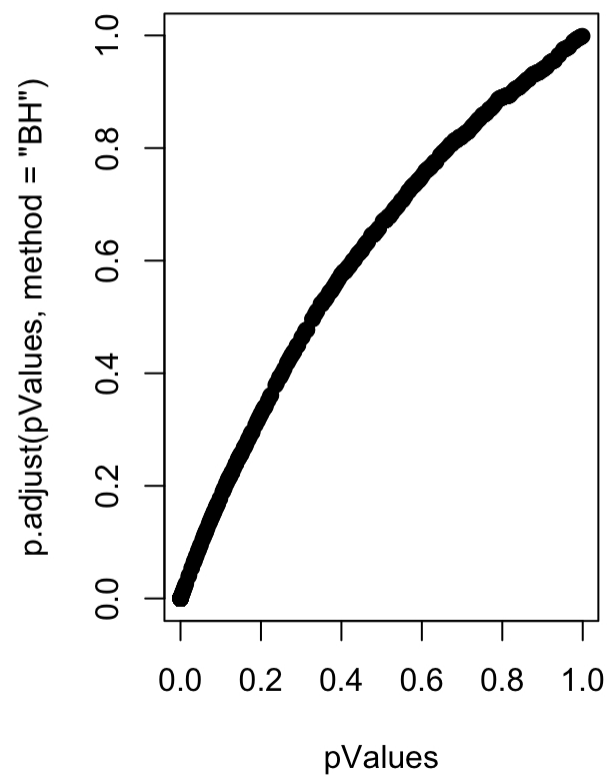
plot the p-values

```
par(mfrow=c(1,2))
plot(pValues,p.adjust(pValues,method="bonferroni"),pch=19)
plot(pValues,p.adjust(pValues,method="BH"),pch=19)
```

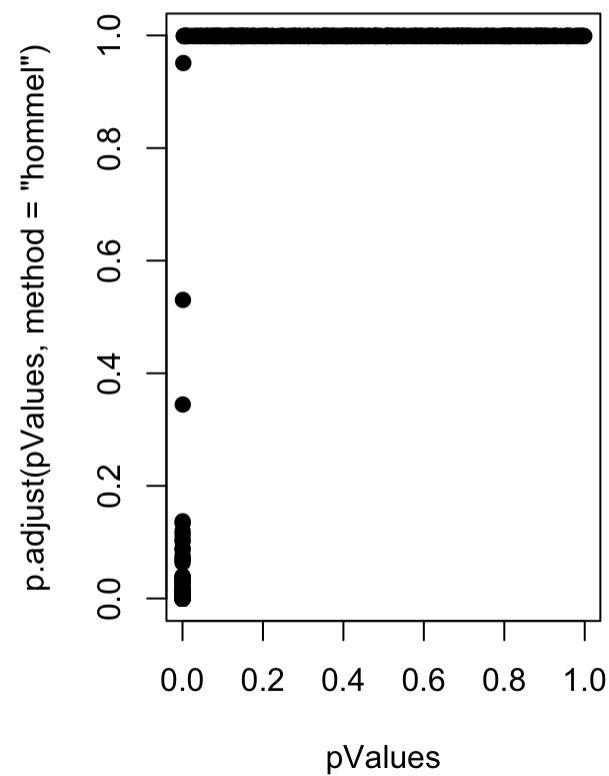
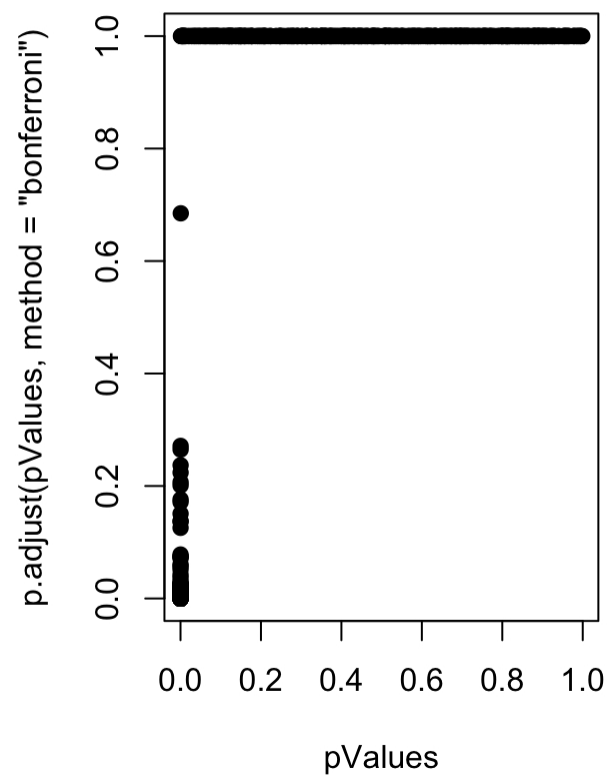


相对平缓一点的变化

```
par(mfrow=c(1,2))
plot(pValues,p.adjust(pValues,method="BH"),pch=19)
plot(pValues,p.adjust(pValues,method="BY"),pch=19)
```



```
par(mfrow=c(1,2))
plot(pValues,p.adjust(pValues,method="bonferroni"),pch=19)
plot(pValues,p.adjust(pValues,method="hommel"),pch=19)
```



非常保守，大部分p-value直接变成1

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

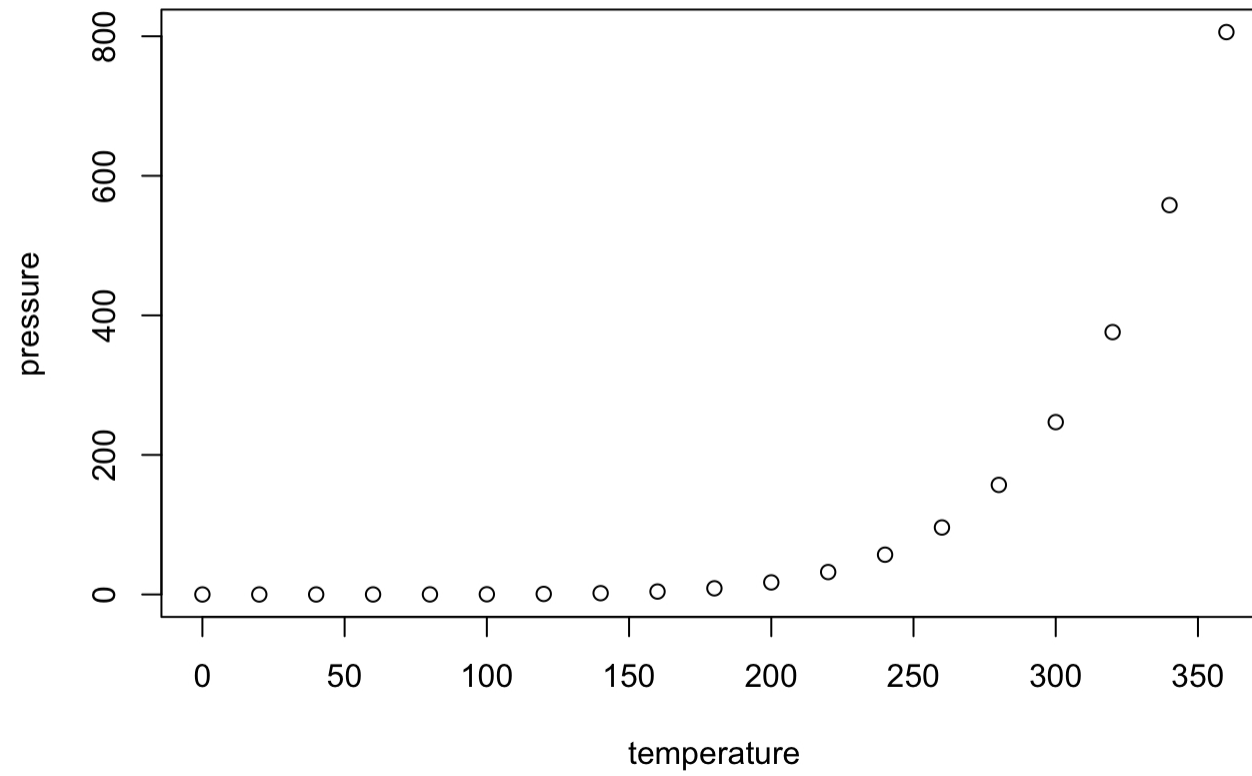
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   : 2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.