Multiple Testing Lab

Raven Ico

Use the following code to simulate a matric of categorical predictors $X \in \{0,1\}^{200 \times 100}$. Each column of X represents a predictor and each row of X indicates an observation. The code also simulates a response variable Y.

```
set.seed(2019)
# Simulate predictor matrix
X <- matrix(rbinom(10000, 1, .5), 200, 100) #100 predictors, 200 observations

# Simluate response vector
Y <- rnorm(200, mean = 2*X[,4]+5*X[,32]-.6*X[,90], sd = 1)
```

Problem 1

What is the "true" effect of each of the predictors on Y (we know this because we simulated the data)?

True effect of X[,4] is 2, X[,32] is 5, and X[,90] is 0.6 and the other predictor variables is 0.

Problem 2

Write a for loop that conducts a hypothesis test comparing whether the mean of the values of Y when the ith column of X is 0 are different than the mean of the values of Y when the ith column of X is 1 for each of the simulated predictors. Store the p-values. How many p-values are less than 0.05?

```
set.seed(2019)
pValues <- rep(NA, 100)
for(i in 1:100){
  pValues[i]<-t.test(Y~X[,i])$p.value
}
tibble("P-values<0.05 (No adjustment)"=sum(pValues<0.05))
```

```
## # A tibble: 1 x 1
##    `P-values<0.05 (No adjustment)`
##                              <int>
## 1                               14
```

Problem 3

Use an appropriate method to control the familywise error rate. How many corrected p-values are less than 0.05? Using 0.05 as the cutoff how many false positives are there? How many true positives?

```
Positives_Bonf = which(p.adjust(pValues,method="bonferroni")<0.05) #Correction method for familywise er
True_Positive_Bonf = sum(Positives_Bonf %in% c(4,32,90)) #True Positive
False_Positive_Bonf = length(Positives_Bonf)-True_Positive_Bonf

tibble("True Positive (Bonferroni)"=True_Positive_Bonf,"False Positive (Bonferroni)"=False_Positive_Bon
```

```
## # A tibble: 1 x 3
##    `True Positive (Bonfe~ `False Positive (Bonf~ `Total Positives (correcte~
##                    <int>                  <int>                       <int>
## 1                      2                      2                           4
```

Problem 4

Use an appropriate method to control the false discovery rate. How many corrected p-values are less than 0.05? Using 0.05 as the cutoff how many false positives are there? How many true positives?

```r
#Method of correction for False Discovery Rate = Benjamini & Hochberg
Positives_BH = which(p.adjust(pValues,method="BH")<0.05) #Correction method for familywise error rate
True_Positive_BH = sum(Positives_BH %in% c(4,32,90)) #True Positive
False_Positive_BH = length(Positives_BH)-True_Positive_BH

tibble("True Positive (BH)"=True_Positive_BH,"False Positive (BH)"=False_Positive_BH, "Total Positives
```

```
## # A tibble: 1 x 3
##   `True Positive (BH~ `False Positive (B~ `Total Positives (corrected p-va~
##               <int>               <int>                             <int>
## 1                  2                   2                                 4
```