Chapter 13 Multiple Testing

Exercise 1

a

We expect to make

$$\alpha \times m$$

type I errors.

b

The FWER is

$$1 - (1 - \alpha)^m$$

c

When m = 2, the FWER in b is

$$1 - (1 - \alpha)^2$$

which is in general relatively small.

If the 2 p-values are positive correlated, when one is small and smaller than

$$\alpha$$

then the other one tends to be small and also smaller than

$$\alpha$$

. The probability that we at least make one type I error raises (compared to the FWER in b) .

On the other hand, when one is large and larger

$$\alpha$$

then the other one tends to be large and larger than

$$\alpha$$

as well. The probability that we at least make one type I error declines (compared to the FWER in b) .

d

In general, if the 2 p-values are negatively correlated, when one is small and smaller than

$$\alpha$$

then the other one tends to be large and larger than

$$\alpha$$

. The probability that we at least make one type I error raises (compared to the FWER in b).

Exercise 2

a

The distribution of

$$A_j$$

is Bernoulli distribution.

b

The distribution of

$$\sum_{j=1}^{m} A_j$$

is Binomial distribution (sum of Bernoulli distribution). Each reject is equivalent to one success.

c

The standard deviation is

$$\sigma = \sqrt{m\alpha(1-\alpha)}$$

Exercise 3

We want to prove

$$1 - \prod_{j=1}^{m}(1-\alpha_j) \leq \sum_{j=1}^{m} \alpha_j \qquad (1)$$

Assuming (1) is true, we can easily get:

$$(1-\alpha_1)(1-\alpha_2)...(1-\alpha_m) \geq 1 - (\alpha_1 + \alpha_2 + ... + \alpha_m)$$
$$\Leftrightarrow (1-\alpha_1)(1-\alpha_2)...(1-\alpha_m)(1-\alpha_{m+1}) \geq (1-\alpha_{m+1})(1-(\alpha_1 + \alpha_2 + ... + \alpha_m))$$
$$\Leftrightarrow (1-\alpha_1)(1-\alpha_2)...(1-\alpha_m)(1-\alpha_{m+1}) \geq (1-\alpha_{m+1})(1-(\alpha_1 + \alpha_2 + ... + \alpha_m + \alpha_{m+1}) + \alpha_{m+1})$$

$$\Leftrightarrow \prod_{j=1}^{m+1}(1-\alpha_j) \geq (1-\alpha_{m+1})(1-\sum_{j=1}^{m+1}\alpha_j+\alpha_{m+1}) \qquad (2)$$

On the other hand, we have:

$$\sum_{j=1}^{m}\alpha_j \geq 0$$

$$\Leftrightarrow -1+\sum_{j=1}^{m+1}\alpha_j+1-\alpha_{m+1} \geq 0$$

$$\Leftrightarrow -(1-\sum_{j=1}^{m+1}\alpha_j)+1-\alpha_{m+1} \geq 0$$

$$\Leftrightarrow -(1-\sum_{j=1}^{m+1}\alpha_j)\alpha_{m+1}+(1-\alpha_{m+1})\alpha_{m+1} \geq 0$$

$$\Leftrightarrow (1-\sum_{j=1}^{m+1}\alpha_j)-(1-\sum_{j=1}^{m+1}\alpha_j)\alpha_{m+1}+(1-\alpha_{m+1})\alpha_{m+1} \geq (1-\sum_{j=1}^{m+1}\alpha_j)$$

$$\Leftrightarrow (1-\alpha_{m+1})(1-\sum_{j=1}^{m+1}\alpha_j+\alpha_{m+1}) \geq (1-\sum_{j=1}^{m+1}\alpha_j) \qquad (3)$$

From (2) and (3), we can write:

$$\prod_{j=1}^{m+1}(1-\alpha_j) \geq 1-\sum_{j=1}^{m+1}\alpha_j$$

$$\Leftrightarrow 1-\prod_{j=1}^{m+1}(1-\alpha_j) \leq \sum_{j=1}^{m+1}\alpha_j$$

By assuming (1) is true, we can prove that it's also true with

$$m+1$$

.

In other words, the family-wise error rate is no greater than

$$\sum_{j=1}^{m+1}\alpha_j$$

Exercise 4

a

We reject these six hypotheses

$$H_{01}, H_{02}, H_{03}, H_{08}, H_{09}, H_{10}$$

.

b

To control FWER at level 0.05 while testing m = 4 null hypotheses, we should reject null hypotheses whose p-values are less than

$$0.05/4 = 0.0125$$

. And they are hypotheses

$$H_{01}, H_{08}, H_{09}, H_{10}$$

.

c

```r
null_hypothesis = c('H01', 'H02', 'H03', 'H04', 'H05',
                    'H06', 'H07', 'H08', 'H09', 'H10')
p_values = c(0.0011, 0.31, 0.017, 0.32, 0.11,
             0.90, 0.07, 0.006, 0.004, 0.0009)
data = data.frame(null_hypothesis, p_values)
data
```

```
##    null_hypothesis p_values
## 1              H01   0.0011
## 2              H02   0.3100
## 3              H03   0.0170
## 4              H04   0.3200
## 5              H05   0.1100
## 6              H06   0.9000
## 7              H07   0.0700
## 8              H08   0.0060
## 9              H09   0.0040
## 10             H10   0.0009
```

```r
ordered_data = data[order(p_values), ]
q = 0.05
p_j = matrix()
m = dim(data)[1]
for (i in 1:10){
  ordered_data$p_j[i] = q * i / m
  ordered_data$reject[i] = ordered_data$p_values[i] < ordered_data$p_j[i]
}
```

As a result, we reject five null hypotheses

$$H_{01}, H_{03}, H_{08}, H_{09}, H_{10}$$

.

```r
ordered_data
```

```
##    null_hypothesis p_values   p_j reject
## 10             H10   0.0009 0.005   TRUE
## 1              H01   0.0011 0.010   TRUE
```

```
## 9               H09   0.0040 0.015   TRUE
## 8               H08   0.0060 0.020   TRUE
## 3               H03   0.0170 0.025   TRUE
## 7               H07   0.0700 0.030  FALSE
## 5               H05   0.1100 0.035  FALSE
## 2               H02   0.3100 0.040  FALSE
## 4               H04   0.3200 0.045  FALSE
## 6               H06   0.9000 0.050  FALSE
```

d

As a result, we reject seven null hypotheses

$$H_{01}, H_{03}, H_{05}, H_{07}, H_{08}, H_{09}, H_{10}$$

.

```
ordered_data = data[order(p_values), ]
q = 0.2
p_j = matrix()
m = dim(data)[1]
for (i in 1:10){
  ordered_data$p_j[i] = q * i / m
  ordered_data$reject[i] = ordered_data$p_values[i] < ordered_data$p_j[i]
}

ordered_data
```

```
##    null_hypothesis p_values  p_j reject
## 10             H10   0.0009 0.02   TRUE
## 1              H01   0.0011 0.04   TRUE
## 9              H09   0.0040 0.06   TRUE
## 8              H08   0.0060 0.08   TRUE
## 3              H03   0.0170 0.10   TRUE
## 7              H07   0.0700 0.12   TRUE
## 5              H05   0.1100 0.14   TRUE
## 2              H02   0.3100 0.16  FALSE
## 4              H04   0.3200 0.18  FALSE
## 6              H06   0.9000 0.20  FALSE
```

e

We have rejected seven null hypotheses from (d) and there are approximately

$$0.2 \times 7 \approx 1.4$$

(from 1 to 2) false positives.

Exercise 5

a

```r
null_hypothesis = c('H01', 'H02', 'H03', 'H04', 'H05')
p_values = c(0.0011, 0.1, 0.2, 0.025, 0.034)
data = data.frame(null_hypothesis, p_values)

ordered_data = data[order(p_values), ]
alpha = 0.1
p_j = matrix()
m = dim(data)[1]

for (i in 1:m){

  ordered_data$Bonferroni[i] = alpha / m
  ordered_data$Bonferroni_reject[i] = ordered_data$p_values[i] < ordered_data$Bonferroni[i]

  ordered_data$Holm[i] = alpha / (m + 1 - i)
  ordered_data$Holm_reject[i] = ordered_data$p_values[i] < ordered_data$Holm[i]

}

ordered_data
```

```
##   null_hypothesis p_values Bonferroni Bonferroni_reject       Holm Holm_reject
## 1             H01   0.0011       0.02              TRUE 0.02000000        TRUE
## 4             H04   0.0250       0.02             FALSE 0.02500000       FALSE
## 5             H05   0.0340       0.02             FALSE 0.03333333       FALSE
## 2             H02   0.1000       0.02             FALSE 0.05000000       FALSE
## 3             H03   0.2000       0.02             FALSE 0.10000000       FALSE
```

b

```r
null_hypothesis = c('H01', 'H02', 'H03', 'H04', 'H05')
p_values = c(0.0011, 0.1, 0.2, 0.021, 0.034)
data = data.frame(null_hypothesis, p_values)

ordered_data = data[order(p_values), ]
alpha = 0.1
p_j = matrix()
m = dim(data)[1]

for (i in 1:m){

  ordered_data$Bonferroni[i] = alpha / m
  ordered_data$Bonferroni_reject[i] = ordered_data$p_values[i] < ordered_data$Bonferroni[i]

  ordered_data$Holm[i] = alpha / (m + 1 - i)
  ordered_data$Holm_reject[i] = ordered_data$p_values[i] < ordered_data$Holm[i]
```

```
}

ordered_data
```

```
##   null_hypothesis p_values Bonferroni Bonferroni_reject       Holm Holm_reject
## 1             H01   0.0011       0.02              TRUE 0.02000000        TRUE
## 4             H04   0.0210       0.02             FALSE 0.02500000        TRUE
## 5             H05   0.0340       0.02             FALSE 0.03333333       FALSE
## 2             H02   0.1000       0.02             FALSE 0.05000000       FALSE
## 3             H03   0.2000       0.02             FALSE 0.10000000       FALSE
```

Exercise 6

a

Table 1: Bonferroni procedure

	Panel 1	Panel 2	Panel 3
False positives	0	0	0
False negatives	1	1	5
True positives	7	7	3
True negatives	2	2	2
Type I errors	0	0	0
Type II errors	1	1	5

b

Table 2: Holm procedure

	Panel 1	Panel 2	Panel 3
False positives	0	0	0
False negatives	1	0	0
True positives	7	8	8
True negatives	2	2	2
Type I errors	0	0	0
Type II errors	1	0	0

c

The false discovery rate associated with using the Bonferroni procedure to control the FWER at level

$$\alpha$$

= 0.05 is 0.

d

The false discovery rate associated with using the Holm procedure to control the FWER at level

$$\alpha$$

= 0.05 is 0.

e

Table 3: Bonferroni procedure

	Panel 1	Panel 2	Panel 3
False positives	0	0	0
False negatives	5	7	6
True positives	3	1	2
True negatives	2	2	2
Type I errors	0	0	0
Type II errors	5	7	6

There is no change in the FDR, it's still 0.

Exercise 7

```
library(ISLR2)
head(Carseats)
```

```
##   Sales CompPrice Income Advertising Population Price ShelveLoc Age Education
## 1  9.50       138     73          11        276   120       Bad  42        17
## 2 11.22       111     48          16        260    83      Good  65        10
## 3 10.06       113     35          10        269    80    Medium  59        12
## 4  7.40       117    100           4        466    97    Medium  55        14
## 5  4.15       141     64           3        340   128       Bad  38        13
## 6 10.81       124    113          13        501    72       Bad  78        16
##   Urban  US
## 1   Yes Yes
## 2   Yes Yes
## 3   Yes Yes
## 4   Yes Yes
## 5   Yes  No
## 6    No Yes
```

a

```
model1 = lm(Sales ~ CompPrice, data = Carseats)
summary(model1)
```

```
##
## Call:
## lm(formula = Sales ~ CompPrice, data = Carseats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6618 -2.0422 -0.0108  1.7440  8.5846
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.021469   1.159945   5.191 3.34e-07 ***
## CompPrice   0.011801   0.009212   1.281    0.201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.822 on 398 degrees of freedom
## Multiple R-squared:  0.004106,   Adjusted R-squared:  0.001604
## F-statistic: 1.641 on 1 and 398 DF,  p-value: 0.2009
```

```
model2 = lm(Sales ~ Income, data = Carseats)
summary(model2)
```

```
##
## Call:
## lm(formula = Sales ~ Income, data = Carseats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.2629 -1.9447 -0.1772  1.7654  8.9064
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.44356    0.37061  17.386  < 2e-16 ***
## Income       0.01533    0.00500   3.067  0.00231 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.795 on 398 degrees of freedom
## Multiple R-squared:  0.02309,    Adjusted R-squared:  0.02063
## F-statistic: 9.407 on 1 and 398 DF,  p-value: 0.00231
```

```
model3 = lm(Sales ~ Advertising, data = Carseats)
summary(model3)
```

```
##
## Call:
## lm(formula = Sales ~ Advertising, data = Carseats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.3770 -1.9634 -0.1037  1.7222  8.3208
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.7370     0.1925  35.007  < 2e-16 ***
## Advertising   0.1144     0.0205   5.583 4.38e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.723 on 398 degrees of freedom
## Multiple R-squared:  0.07263,    Adjusted R-squared:  0.0703
## F-statistic: 31.17 on 1 and 398 DF,  p-value: 4.378e-08
```

```
model4 = lm(Sales ~ Population, data = Carseats)
summary(model4)
```

```
##
## Call:
## lm(formula = Sales ~ Population, data = Carseats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.5864 -2.0176 -0.0597  1.6892  8.7213
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.2401837  0.2906658  24.909   <2e-16 ***
## Population  0.0009672  0.0009593   1.008    0.314
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.824 on 398 degrees of freedom
## Multiple R-squared:  0.002547,   Adjusted R-squared:  4.116e-05
## F-statistic: 1.016 on 1 and 398 DF,  p-value: 0.314
```

```
model5 = lm(Sales ~ Price, data = Carseats)
summary(model5)
```

```
##
## Call:
## lm(formula = Sales ~ Price, data = Carseats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.5224 -1.8442 -0.1459  1.6503  7.5108
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.641915   0.632812  21.558   <2e-16 ***
## Price       -0.053073   0.005354  -9.912   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.532 on 398 degrees of freedom
## Multiple R-squared:  0.198,  Adjusted R-squared:  0.196
## F-statistic: 98.25 on 1 and 398 DF,  p-value: < 2.2e-16
```

```r
model6 = lm(Sales ~ Age, data = Carseats)
summary(model6)
```

```
##
## Call:
## lm(formula = Sales ~ Age, data = Carseats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.1900 -1.8648 -0.1261  1.7449  8.3969
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.65115    0.47365  20.376  < 2e-16 ***
## Age         -0.04041    0.00850  -4.754 2.79e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.751 on 398 degrees of freedom
## Multiple R-squared:  0.05374,    Adjusted R-squared:  0.05136
## F-statistic:  22.6 on 1 and 398 DF,  p-value: 2.789e-06
```

```r
model7 = lm(Sales ~ Education, data = Carseats)
summary(model7)
```

```
##
## Call:
## lm(formula = Sales ~ Education, data = Carseats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.4347 -2.0322 -0.0357  1.7788  8.6113
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.27461    0.76304  10.844   <2e-16 ***
## Education   -0.05599    0.05395  -1.038      0.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.824 on 398 degrees of freedom
## Multiple R-squared:  0.002699,   Adjusted R-squared:  0.0001936
## F-statistic: 1.077 on 1 and 398 DF,  p-value: 0.2999
```

```r
variables = c('CompPrice', 'Income', 'Advertising', 'Population',
              'Price', 'Age', 'Education')
p_values = c(0.201, 0.00231, 4.38e-08, 0.314, 2e-16, 2.79e-06, 0.3)
report = data.frame(variables, p_values)
report
```

```
##     variables p_values
```

```
## 1   CompPrice 2.01e-01
## 2      Income 2.31e-03
## 3 Advertising 4.38e-08
## 4  Population 3.14e-01
## 5       Price 2.00e-16
## 6         Age 2.79e-06
## 7   Education 3.00e-01
```

b

```r
m = dim(report)[1]
alpha = 0.05
for (i in 1:m){
  report$reject[i] = report$p_values[i] < alpha
}
report
```

```
##     variables p_values reject
## 1   CompPrice 2.01e-01  FALSE
## 2      Income 2.31e-03   TRUE
## 3 Advertising 4.38e-08   TRUE
## 4  Population 3.14e-01  FALSE
## 5       Price 2.00e-16   TRUE
## 6         Age 2.79e-06   TRUE
## 7   Education 3.00e-01  FALSE
```

c

```r
FWER = 0.05
for (i in 1:m){
  report$FWER_reject[i] = report$p_values[i] < FWER/m
}
report
```

```
##     variables p_values reject FWER_reject
## 1   CompPrice 2.01e-01  FALSE       FALSE
## 2      Income 2.31e-03   TRUE        TRUE
## 3 Advertising 4.38e-08   TRUE        TRUE
## 4  Population 3.14e-01  FALSE       FALSE
## 5       Price 2.00e-16   TRUE        TRUE
## 6         Age 2.79e-06   TRUE        TRUE
## 7   Education 3.00e-01  FALSE       FALSE
```

d

```r
ordered_report = report[order(p_values), ]
q = 0.2
```

```r
for (i in 1:m){
  p_j = q / m * i
  ordered_report$FDR_reject[i] = ordered_report$p_values[i] < p_j
}
ordered_report
```

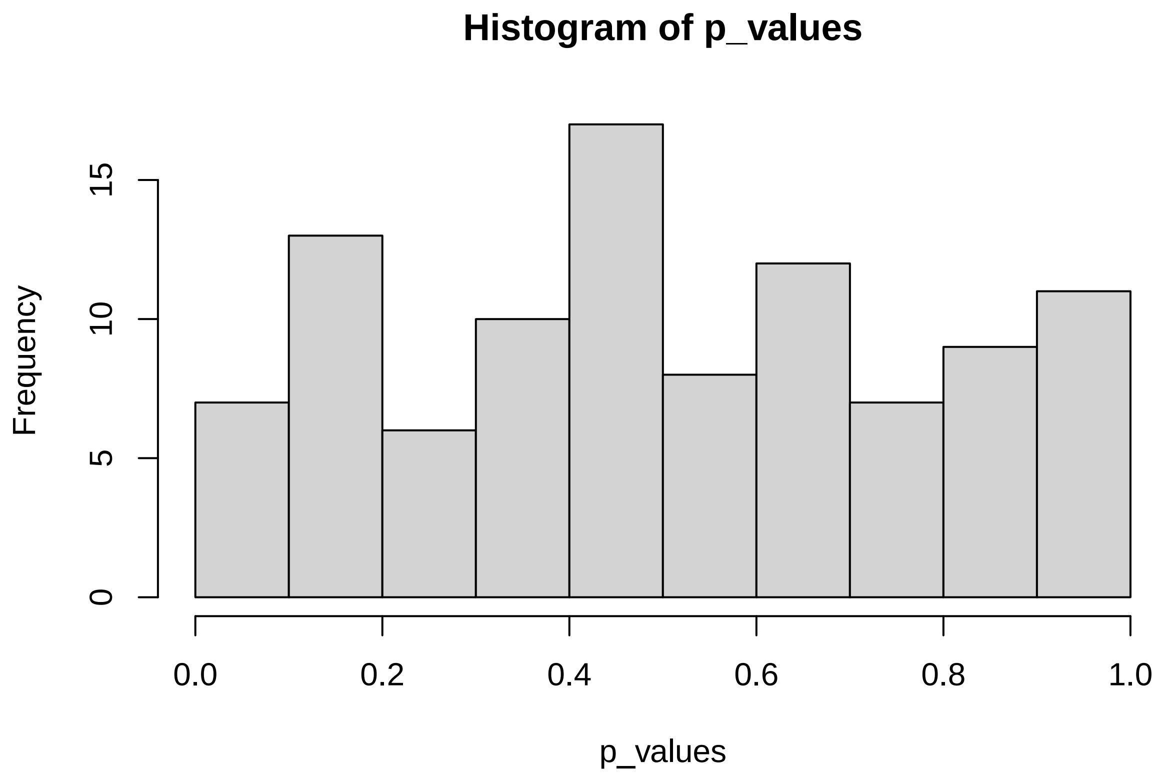
```
##     variables p_values reject FWER_reject FDR_reject
## 5       Price 2.00e-16   TRUE        TRUE       TRUE
## 3 Advertising 4.38e-08   TRUE        TRUE       TRUE
## 6         Age 2.79e-06   TRUE        TRUE       TRUE
## 2      Income 2.31e-03   TRUE        TRUE       TRUE
## 1   CompPrice 2.01e-01  FALSE       FALSE      FALSE
## 7   Education 3.00e-01  FALSE       FALSE      FALSE
## 4  Population 3.14e-01  FALSE       FALSE      FALSE
```

Exercise 8

```r
set.seed(1)
n = 20
m = 100
X = matrix(rnorm(n * m), ncol = m)
```

a

```r
p_values = matrix()
for (i in 1:m){
  p_values[i] = t.test(X[, i], mu = 0)$p.value
}
hist(p_values)
```



b

We would reject 4 null hypothesis. This is roughly equal to 100 * 0.05 = 5

```
sum(p_values < 0.05)
```

```
## [1] 4
```

c

If we control FWER at level 0.05, we would reject 0 null hypotheses.

```
FWER = 0.05
m = 100
sum(p_values < FWER/m)
```

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

d

If we control FDR at level 0.05, we would reject 0 null hypotheses.

```r
ordered_pvalues = data.frame(p_values[order(p_values)])
colnames(ordered_pvalues) = 'p_values'
q = 0.05
for (i in 1:m){
  p_j = q / m * i
  ordered_pvalues$FDR_reject[i] = ordered_pvalues$p_values[i] < p_j
}
sum(ordered_pvalues$FDR_reject)
```

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

e

If we control the FWER for just these 10 cherry-picked managers, we would reject all 1 hypothesis.

```r
ave_returns = apply(X, 2, mean)
cherry_picked_returns = sort(ave_returns, decreasing = T)[1:10]
cherry_picked_indices = match(cherry_picked_returns, ave_returns)
```

```r
m = 10

p_values = matrix()
for (i in 1:m){
  index = cherry_picked_indices[i]
  p_values[i] = t.test(X[, index], mu = 0)$p.value
}

FWER = 0.05
sum(p_values < FWER/m)
```

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

If we control the FDR for just these 10 cherry-picked managers, we would reject 1 hypothesis.

```r
ordered_pvalues = data.frame(p_values[order(p_values)])
colnames(ordered_pvalues) = 'p_values'
q = 0.05
for (i in 1:m){
  p_j = q / m * i
  ordered_pvalues$FDR_reject[i] = ordered_pvalues$p_values[i] < p_j
}
sum(ordered_pvalues$FDR_reject)
```

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

f

By cherry-picking the best managers, we have accidentally choose managers whose p-values are the smallest and this violates the assumption that all tested null hypotheses are adjusted for multiplicity.