Lab Assignment #3

Dr. Wynne's Partial Solutions

Due February 22, 2023

Instructions

There are two purposes to this lab. First, you will get comfortable with the family-wise error rate and false discovery rate. Then, we will get more practice with coding a nonparametric bootstrap.

```
library(ISLR2)
library(ggplot2)
library(dplyr)
```

This lab assignment is worth a total of **15 points**.

Problem 1: Family-Wise Error Rate

Part b (Explanation: 1 pt)

What does the `p.adjust` function do exactly when `method = bonferroni`? What about when `method = holm`? Why does it make more sense for R to adjust the p-values rather than the significance level when controlling the FWER?

Explanation

Since we never told R what FWER we wanted to use, R doesn't know what significance level to compare each of the p-values to. But if we adjust the p-values, then we can compare the adjusted p-values to a *single* significance level.

When `method = bonferroni`, the `p.adjust` function multiplies all p-values by m, unless the adjusted p-value would be greater than 1, then it returns 1.

Rather than multiplying all the p-values by m, the `p.adjust` function multiplies the smallest p-value by m, the second-smallest p-value by m-1, etc., unless the adjusted p-value would be greater than 1, and then it cuts off at 1. (Note that if two p-values are adjusted and the larger p-value yields the smaller adjusted p-value, `p.adjust` replaces the smaller adjusted p-value with the larger one. This may not be apparent in part (b).)

Part c (Explanation: 1 pt)

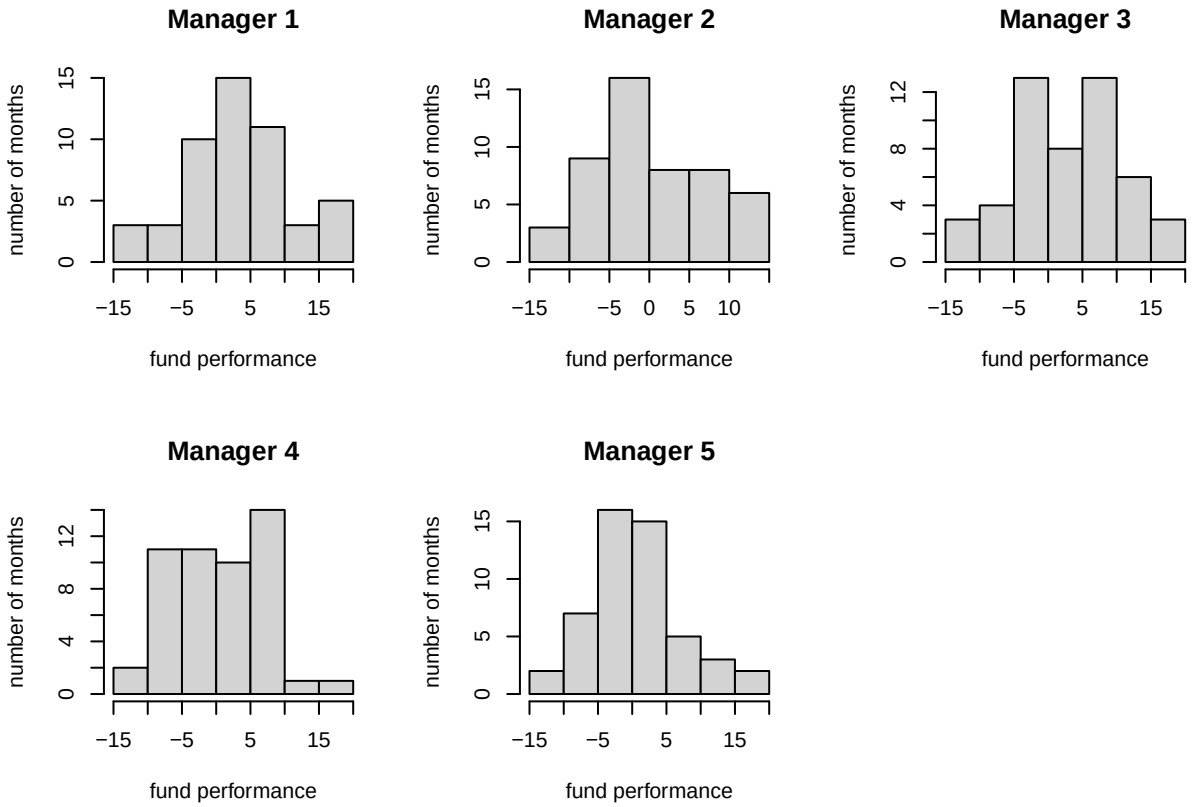
Consider the second-to-last chunk in ISLR Lab 13.6.2 (the one using `TukeyHSD`). Is it appropriate to do a one-way ANOVA with this data? Explain why or why not. (You may want to produce some graphs and/or numerical summaries to support your answer.)

Explanation

```
fund.mini <- Fund[,1:5]
```

According to the course notes, the assumptions of a one-way ANOVA are that $y_{ij} \sim N(\mu_j, \sigma)$ are mutually independent. Therefore, we need to check whether it is believable that the distribution of fund performance for each manager is roughly normal, that the managers have the same variability in fund performance, and that the data values can be assumed independent both *between* managers and *within* the same manager.

```
par(mfrow = c(2,3))
for(i in 1:5){
  hist(fund.mini[,i],
       main = paste("Manager", i),
       xlab = "fund performance",
       ylab = "number of months")
}
```



There are no clear outliers for any manager. For the first four managers, the sample distribution is roughly symmetric, suggesting that the population distribution *could* be normal. Although the sample distribution is skewed right for Manager 5, it is not so clearly skewed that it would be hard to believe a normal population distribution.

```
apply(fund.mini, 2, sd)
```

```
## Manager1 Manager2 Manager3 Manager4 Manager5
## 7.416198 6.855655 7.549834 6.708204 6.782330
```

The sample standard deviations are very close for all five managers, so it seems reasonable that they have the same variability.

However, as evidenced by the fact that the authors chose to do a *paired* t-test, there is a clear dependency structure in the data: if one manager has an above-average return in month i, it is very likely that the other managers will also have above-average returns in the same month. Furthermore, one could argue that if

a manager has an above-average return in month i, it is more likely that the manager will also have an above-average return in month $i+1$, so we may not even have independence *within* the same manager.

It is very likely that this *Fund* data was simulated without thinking about the inherent dependency structures in the data, so it is probably okay to run the initial one-way ANOVA on this simulated data, but I would be *very* hesitant to do it on real data.

Problem 2: False Discovery Rate

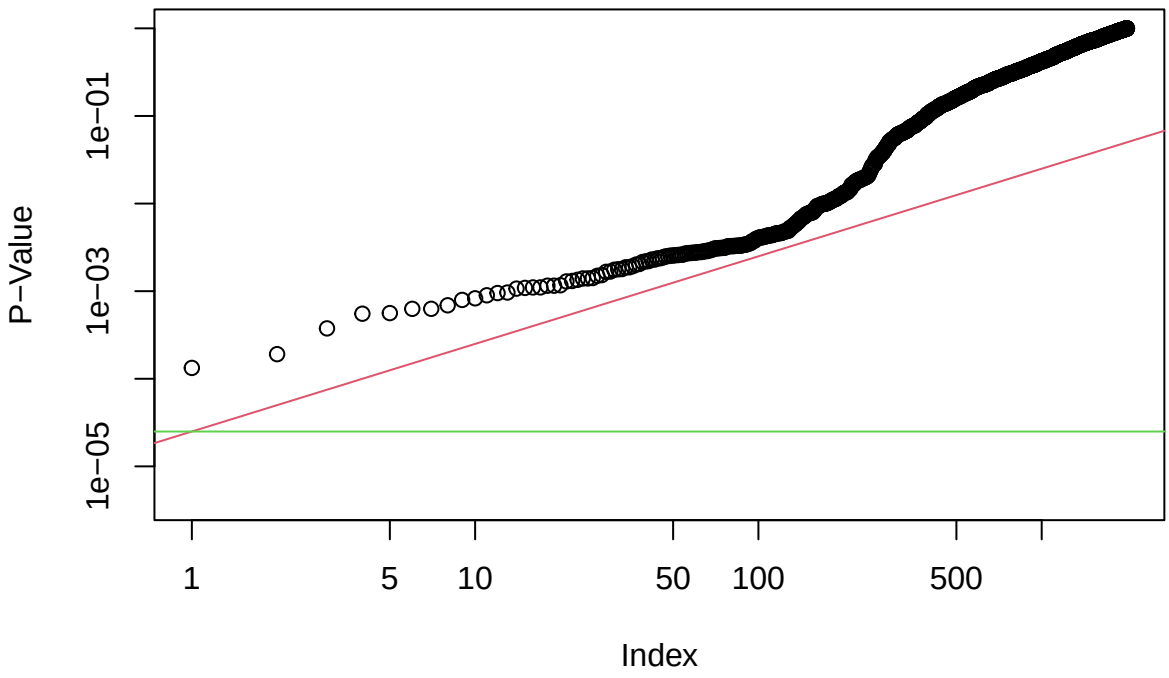
Part b (Code: 1 pt)

Finish writing the `FDR_plot` function in the code chunk below. This function should take two arguments: `p.values`, a vector of p-values, and `q`, the desired False Discovery Rate, and produce a graph like those in Figure 13.6. (Assume that on the graphs, q and α are set to the same value.)

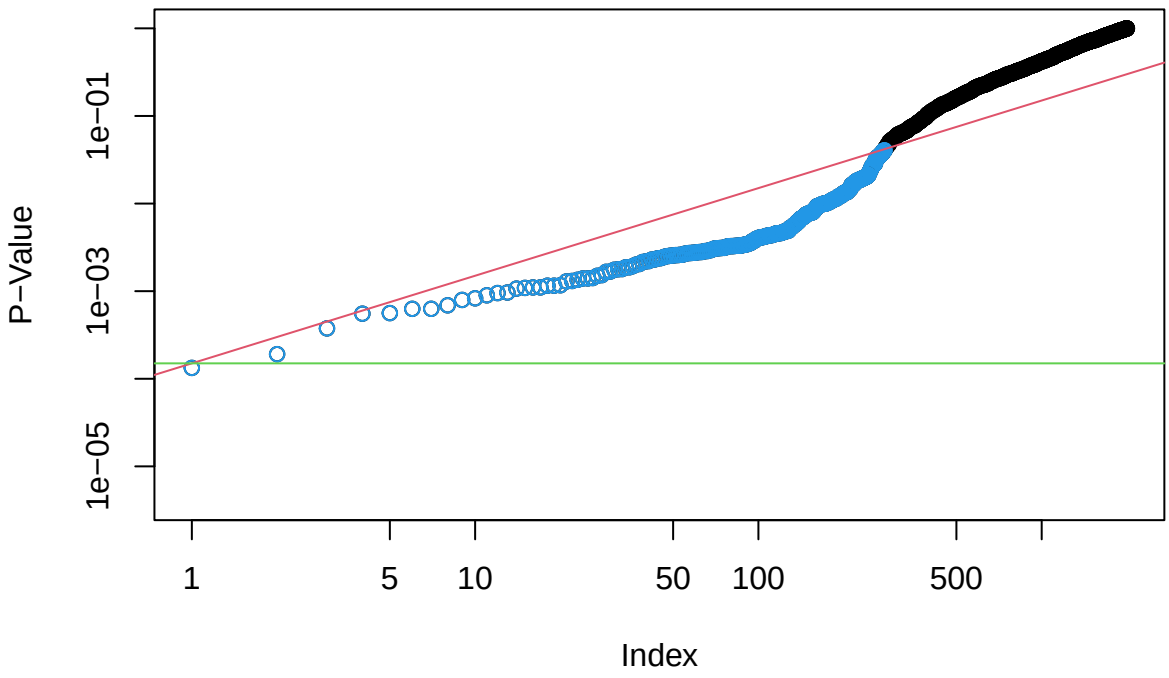
```
FDR_plot <- function(p.values, q){

  # Copy code from the last two chunks of ISLR Lab 13.6.3, but replace variable names and hard-coded va

  ps <- sort(p.values)
  m <- length(p.values)

  wh.ps <- which(ps < q * (1:m)/m)

  if (length(wh.ps) > 0){
    wh <- 1:max(wh.ps)
  } else {
    wh <- numeric(0)
  }


  plot(ps, log = "xy", ylim = c(4e-6, 1), ylab = "P-Value",
       xlab = "Index", main = "")
  points(wh, ps[wh], col = 4)
  abline(a = 0, b = q/m, col = 2, untf = TRUE)
  abline(h = q/m, col= 3)
  invisible(wh.ps) # invisibly return the significant p-values
}
```

Test your function. First, attempt to duplicate the left and right panels of Figure 13.6.

```
## from Lab 13.6.3, spruced up
fund.pvalues <- apply(Fund, 2, function(x) t.test(x, mu = 0)$p.value)

FDR_plot(fund.pvalues, q = 0.05)
```

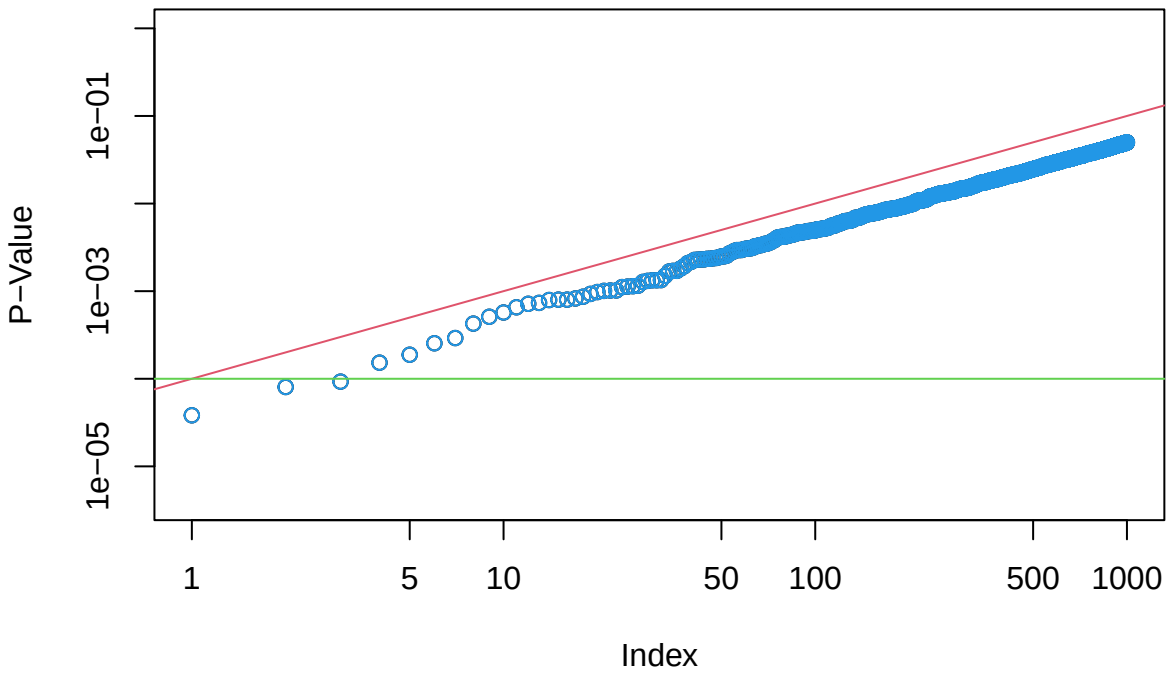



```
FDR_plot(fund.pvalues, q = 0.3)
```



Then, test your function on the simulated dataset in the chunk below, using $q = 0.1$.

```
set.seed(12)
sim_pvalues <- runif(1000, min = 1e-5, max = 0.05001)
# Now put a line of code running your function on this set of "significant" p-values
FDR_plot(sim_pvalues, q = 0.1)
```



Notice that the Bonferroni method would only pick out 3 of these p-values as significant at the $\alpha = 0.1$ significance level, while using an FDR of 0.1, all of these p-values are significant.

Problem 3: Simulation Study of FWER and FDR

This problem is adapted from ISLR Chapter 13, Exercise 8.

Part a (Code: 1 pt)

Using the code in Exercise 13.7.8, create a 20 x 100 matrix where each column represents 20 random numbers from $N(0, 1)$.

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

Then, run a t-test on each column of the matrix testing $H_0 : \mu = 0$ against $H_a : \mu \neq 0$. We are going to use the `apply` function to do this rather than adapt the `for` loop from the ISLR Labs. The `apply` function applies a single function to each row (`MARGIN = 1`) or column (`MARGIN = 2`) of a matrix.

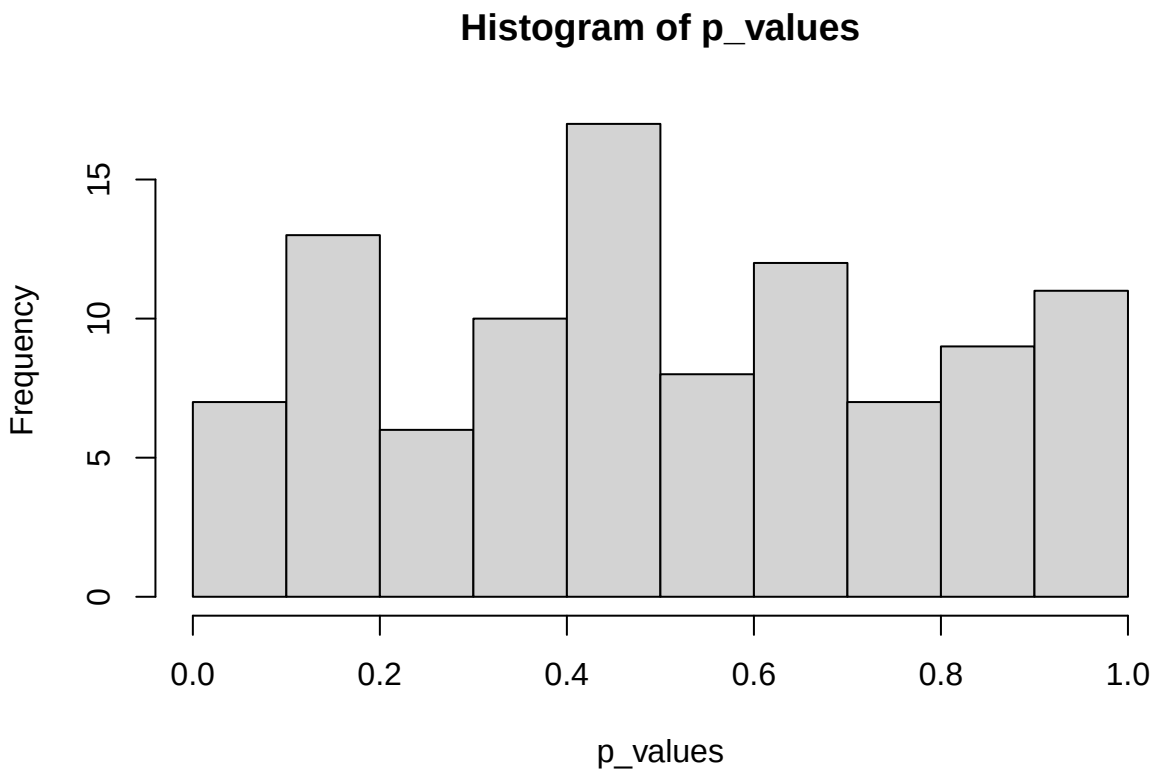
Don't forget to delete the `eval = FALSE` after you've fixed this code chunk to run properly!

```
t_test_0 <- function(x){
  # x: a vector of data

  p_value <- t.test(x, mu = 0)$p.value
```

```
  return(p_value)

}

p_values <- apply(X, 2, t_test_0)
```

Plot a histogram of the p-values obtained.

```
hist(p_values)
```



Part b (Code: 1 pt; Explanation: 0.5 pts)

Without any adjustment for multiple hypothesis tests, how many null hypotheses would be rejected at $\alpha = 0.05$? We can take advantage of the fact that R implicitly converts *logical* (TRUE/FALSE) variables to *numerical* variables.

```
alpha <- 0.05
sum(p_values <= alpha)
```

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

Obtain the adjusted p-values using the Holm step-down procedure. How many null hypotheses would be rejected if we control the FWER at 0.05?

```
p_adjust_holm <- p.adjust(p_values, method = "holm")
sum(p_adjust_holm <= alpha)
```

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

Obtain the adjusted p-values using the Benjamini-Hochberg procedure. How many null hypotheses would be rejected if we control the FDR at 0.05?

```
p_adjust_bh <- p.adjust(p_values, method = "BH")
sum(p_adjust_bh <= alpha)
```

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

Explanation

Controlling the Type I Error rate for each individual test, we would incorrectly reject 4 simulated H_0, but controlling the FWER and FDR, we would not incorrectly reject any simulated H_0.

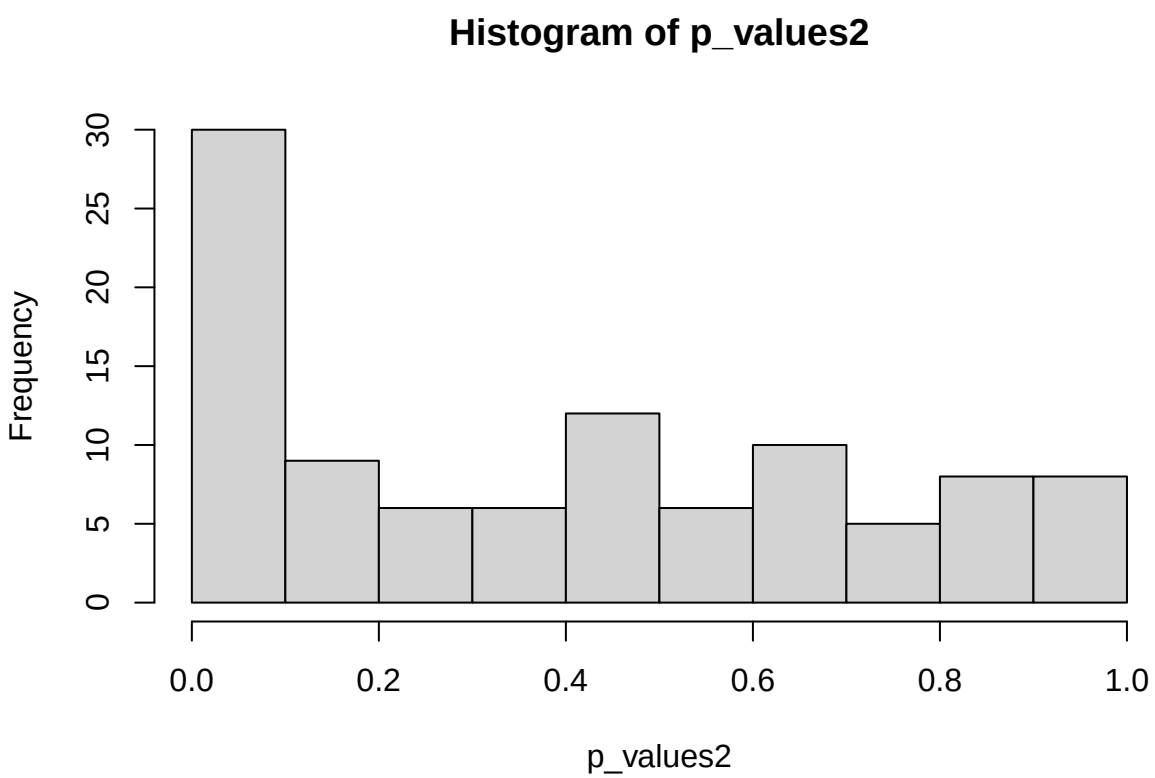
Part c (Code: 1 pt)

Create a new matrix, `X2`, that is exactly the same as X, except that the first 25 fund managers do actually have a slight long-term return of +1%. Conduct your 100 t-tests and plot a histogram of the new p-values.

```
set.seed(1)
X2 <- X
X2[,1:25] <- rnorm(25*n, mean = 1, sd = 1)
```

```
# don't rewrite the function, just apply it to the new dataset!
p_values2 <- apply(X2, 2, t_test_0)
```

```
hist(p_values2)
```



Part d (Code: 1 pts, Explanation: 1 pt)

Without any adjustment for multiple hypothesis tests, how many of the 75 true null hypotheses would be rejected at $\alpha = 0.05$? How many of the 25 false null hypotheses would be rejected?

```
reject.falseH0 <- sum(p_values2[1:25] <= alpha)
reject.trueH0 <- sum(p_values2[26:100] <= alpha)
c(true = reject.trueH0, false = reject.falseH0)
```

```
##  true false
##     3    24
```

Obtain the adjusted p-values using the Holm step-down procedure. How many null hypotheses of each type would be rejected if we control the FWER at 0.05?

```
p_adjust_holm2 <- p.adjust(p_values2, method = "holm")
reject.falseH0 <- sum(p_adjust_holm2[1:25] <= alpha)
reject.trueH0 <- sum(p_adjust_holm2[26:100] <= alpha)
c(true = reject.trueH0, false = reject.falseH0)
```

```
##  true false
##     0    18
```

Obtain the adjusted p-values using the Benjamini-Hochberg procedure. How many null hypotheses of each type would be rejected if we control the FDR at 0.05?

```
p_adjust_bh2 <- p.adjust(p_values2, method = "BH")
reject.falseH0 <- sum(p_adjust_bh2[1:25] <= alpha)
reject.trueH0 <- sum(p_adjust_bh2[26:100] <= alpha)
c(true = reject.trueH0, false = reject.falseH0)
```

```
##  true false
##     1    22
```

Explanation

When controlling the Type I Error rate of each individual test at the 0.05 significance level, we rejected 24 of the 25 false H_0 but also 3 of the 75 true H_0. When we controlled the FWER rate instead, we rejected none of the true H_0, but only rejected 18 of the 25 false H_0. When we controlled the FDR, we were in the middle: we still rejected 1 of the 75 true H_0, but 22 of the 25 false H_0.

It is important to consider the situation where *some* null hypothesis are true (instead of *all*) because we should care about the power of our procedure in addition to the Type I Error rate. In addition, this is the more realistic situation - if we suspected that *all* of the null hypotheses were true, why are we doing the test?