Assignment 9

Prepared by:

- Josh Levine (jl2108)
- Harsh Patel (hkp49)
- Jaini Patel (jp1891)
- Yifan Liao (yl1463)
- Aayush Shah (avs93)

False Discovery Rate (FDR)

The FDR is known as the expected proportions of type I errors. A type I error is where we incorrectly reject the null hypothesis. FDR is closely related to FWER (family-wise error rate), the probability of making any type I error at all.

FDR Formula:

FDR = E (V/R | R>0) P(R>0)

where:

```
*   V = Number of Type I errors
*   R = Number of rejected hypotheses
```

Below is the function for FDR implementation:

```r
false_discovery_rate <- function(p_values, Q){

  sorted_p_values = sort(p_values) #Sorting the p-values -> Step1

  len = length(sorted_p_values)      #Number of p-values

  #Hypothesis 1

  hypothesis_1 = Q*c(1:len)/len

  #Hypothesis 2 - If not independent

  hep = len * sum(1/c(1:len))

  hypothesis_2 = (Q * c(1:len))/hep

  #the p-vlaues less than the hypothesis line are considered to be interesting

  is_interesting_true = (sorted_p_values < hypothesis_1)

  #compute the index below which the sorted p-values are less than the line

  threshold_index = max(which(is_interesting_true == "TRUE"))
```

```r
    print(paste0("The Threshold Index: ", threshold_index)) #The no. of TRUE values

    p_star = sorted_p_values[threshold_index]


    hypothesis = c(1:len)

    plot(hypothesis, sorted_p_values, col="black")

    lines(c(1:threshold_index), sorted_p_values[c(1:threshold_index)],
          col = "red", type="o")

    lines(hypothesis, hypothesis_1, type="l", col="black")


    #list index of hypothesis which are interesting in the original unsorted list of p values

    temp = (p_values < p_star)

    fd = which(temp == "TRUE")

    print("Index of intersting p values in the unsorted list:")

    print(fd)



    #false rejection rate

    frr = -log(p_star)/length(fd)

    print(paste0("False Rejection Rate: ", frr))



    ###############################################################################

    # NOTE: FRR depends on the Q value given as an input to the function

    # Here, the Q value was 0.05 hence the frr is approximately 0.05

    ###############################################################################
}
```

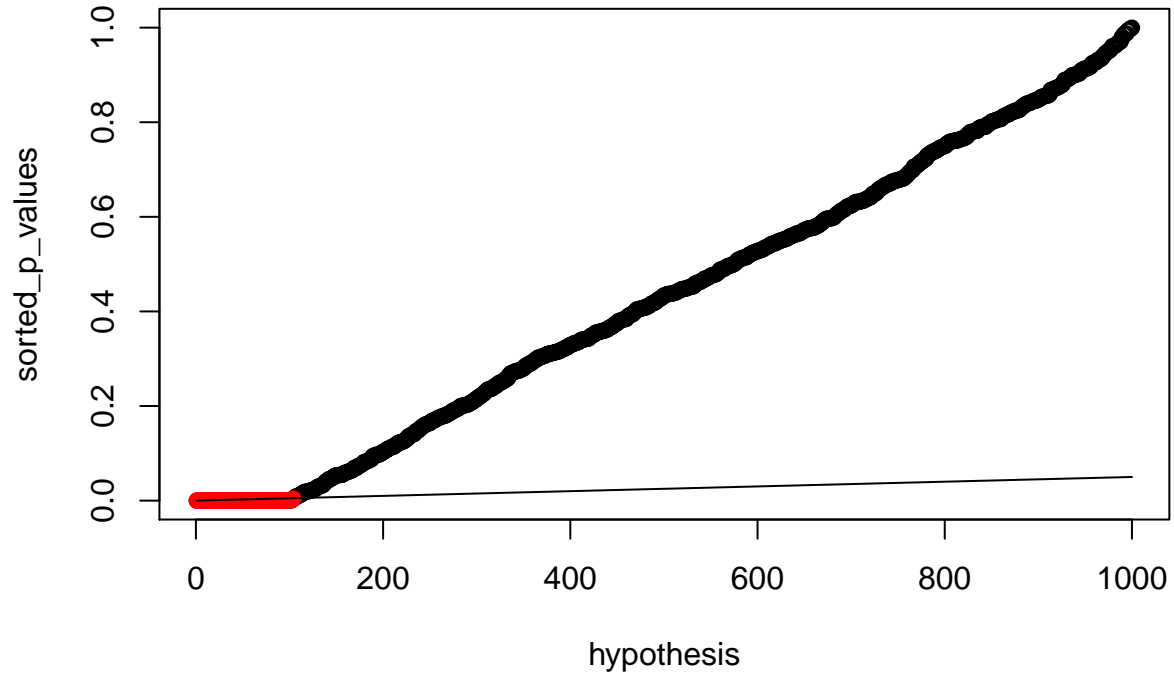Below is the code to call the fdr function:

```r
# Test Vector of p-values

vec <- c(1e-5*runif(100),runif(900))

Q = 0.05

false_discovery_rate(vec, Q)
```

[1] "The Threshold Index: 104"


```
## [1] "Index of intersting p values in the unsorted list:"
##   [1]   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
##  [19]  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
##  [37]  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54
##  [55]  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72
##  [73]  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90
##  [91]  91  92  93  94  95  96  97  98  99 100 434 554 665
## [1] "False Rejection Rate: 0.0530982765886552"
```