

Lab6

Jessica Temporal 7547611

September 29, 2016

Contents

1. Get the GEO Brain Aging study (agingStudy11FCortexAffy.txt, agingStudy11FCortexAffyAnn.txt). Also obtain the annotation file for this data frame. 2
2. Load into R, using `read.table()` function and the `header=T/row.names=1` arguments for each data file. 2
3. Prepare 2 separate vectors for comparison. The first is a comparison between male and female patients. The current data frame can be left alone for this, since the males and females are all grouped together. The second vector is comparison between patients ≥ 50 years of age and those < 50 years of age. To do this, you must use the annotation file and logical operators to isolate the correct arrays/samples. 2
4. Run the `t.test` function from the notes using the first gene vector below for the gender comparison. Then use the second gene vector below for the age comparison. Using these p-values, use either `p.adjust` in the base library or `mt.rawp2adjp` in the `multitest` library to adjust the values for multiple corrections with the Holm's method. 2
5. Sort the adjusted p-values and non-adjusted p-values and plot them vs. the x-axis of numbers (e.g. `1-length(p.adj)`) for each comparison data set. Make sure that the two lines are different colors. Hint: use `sort()` to sort the values. 3
6. Repeat #4 and #5 with the Bonferroni method. 5

1. Get the GEO Brain Aging study (agingStudy11FCortexAffy.txt, agingStudy11FCortexAffyAnn.txt). Also obtain the annotation file for this data frame.

```
study_file <- "agingStudy11FCortexAffy.txt"
anno_file <- "agingStudy11FCortexAffyAnn.txt"
```

2. Load into R, using read.table() function and the header=T/row.names=1 arguments for each data file.

```
study <- read.table(study_file, header = T, na.strings = "NA", blank.lines.skip = F, row.names = 1)
study_ann <- read.table(anno_file, header = T, na.strings = "NA", blank.lines.skip = F, row.names = 1)
```

3. Prepare 2 separate vectors for comparison. The first is a comparison between male and female patients. The current data frame can be left alone for this, since the males and females are all grouped together. The second vector is comparison between patients ≥ 50 years of age and those < 50 years of age. To do this, you must use the annotation file and logical operators to isolate the correct arrays/samples.

```
study_names <- data.frame(names = dimnames(study)[[2]])
study_names$sample <- NA
study_names$sex <- NA

for(i in 1:30) {
  study_names$sample[i] <- strsplit(as.character(study_names$names), '[.]')[[i]][1]
  study_names$sex[i] <- strsplit(as.character(study_names$names), '[.]')[[i]][3]
}

colnames(study) <- study_names$sample

study_m <- study[,subset(study_names, subset = sex=='M', select = 'sample')$sample]
study_f <- study[,subset(study_names, subset = sex=='F', select = 'sample')$sample]

study_lower <- study[,rownames(subset(study_ann, subset = Age < 50))]
study_higher <- study[,rownames(subset(study_ann, subset = Age >= 50))]
```

4. Run the t.test function from the notes using the first gene vector below for the gender comparison. Then use the second gene vector below for the age comparison. Using these p-values, use either p.adjust in the base library or mt.rawp2adjp in the multitest library to adjust the values for multiple corrections with the Holm's method.

```
# gender comparison gene vector
g.g <- c(1394, 1474, 1917, 2099, 2367, 2428, 2625, 3168, 3181, 3641, 3832, 4526,
4731, 4863, 6062, 6356, 6684, 6787, 6900, 7223, 7244, 7299, 8086, 8652,
8959, 9073, 9145, 9389, 10219, 11238, 11669, 11674, 11793)

# age comparison gene vector
g.a <- c(25, 302, 1847, 2324, 246, 2757, 3222, 3675, 4429, 4430, 4912, 5640, 5835, 5856, 6803,
12404, 12442, 67, 88, 100)
```

```

# t.test function from the notes
t.test.all.genes <- function(x,s1,s2) {
  x1 <- x[s1]
  x2 <- x[s2]
  x1 <- as.numeric(x1)
  x2 <- as.numeric(x2)
  t.out <- t.test(x1,x2, alternative="two.sided",var.equal=T)
  out <- as.numeric(t.out$p.value)
  return(out)
}

gender_test <- apply(study[g.g,], 1, t.test.all.genes, s1 = as.matrix(study_m), s2 = as.matrix(study_f))
age_test <- apply(study[g.a,], 1, t.test.all.genes, s1 = as.matrix(study_higher), s2 = as.matrix(study_lower))

gender_test_adjusted <- p.adjust(gender_test, method = "holm")
age_test_adjusted <- p.adjust(age_test, method = "holm")

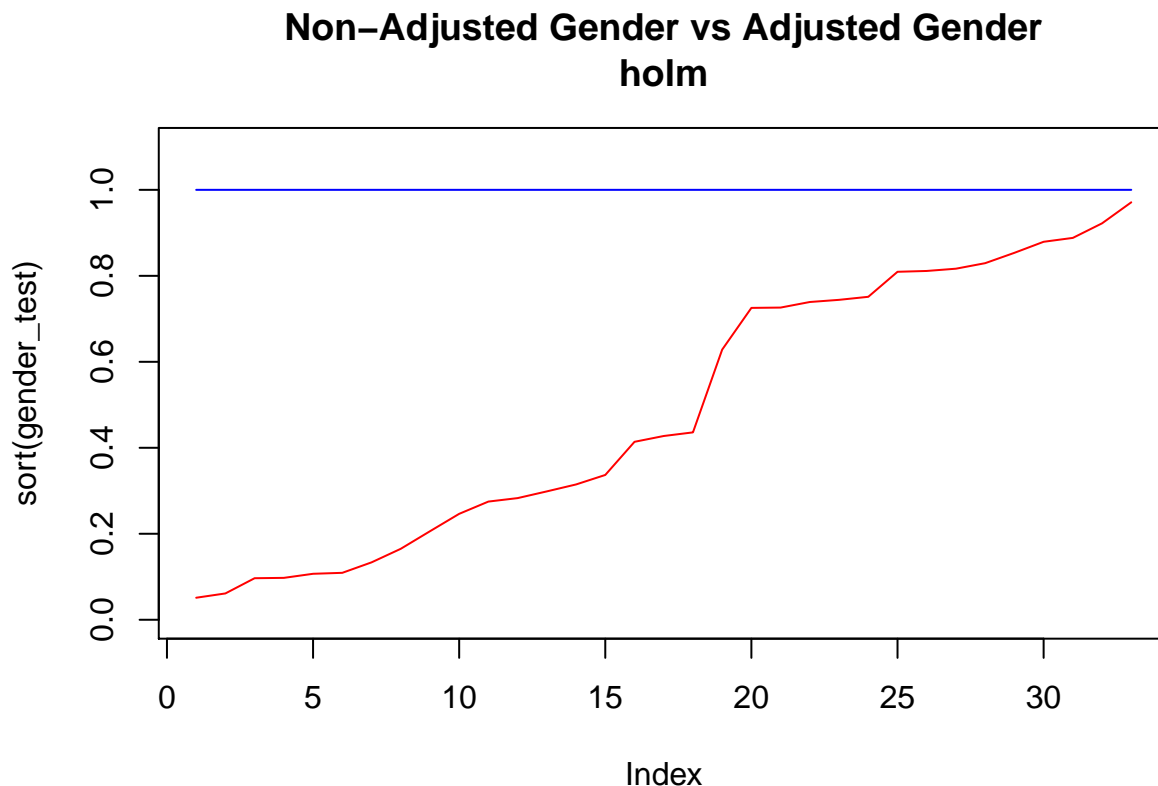
```

5. Sort the adjusted p-values and non-adjusted p-values and plot them vs. the x-axis of numbers (e.g. `1-length(p.adj)`) for each comparison data set. Make sure that the two lines are different colors. Hint: use `sort()` to sort the values.

```

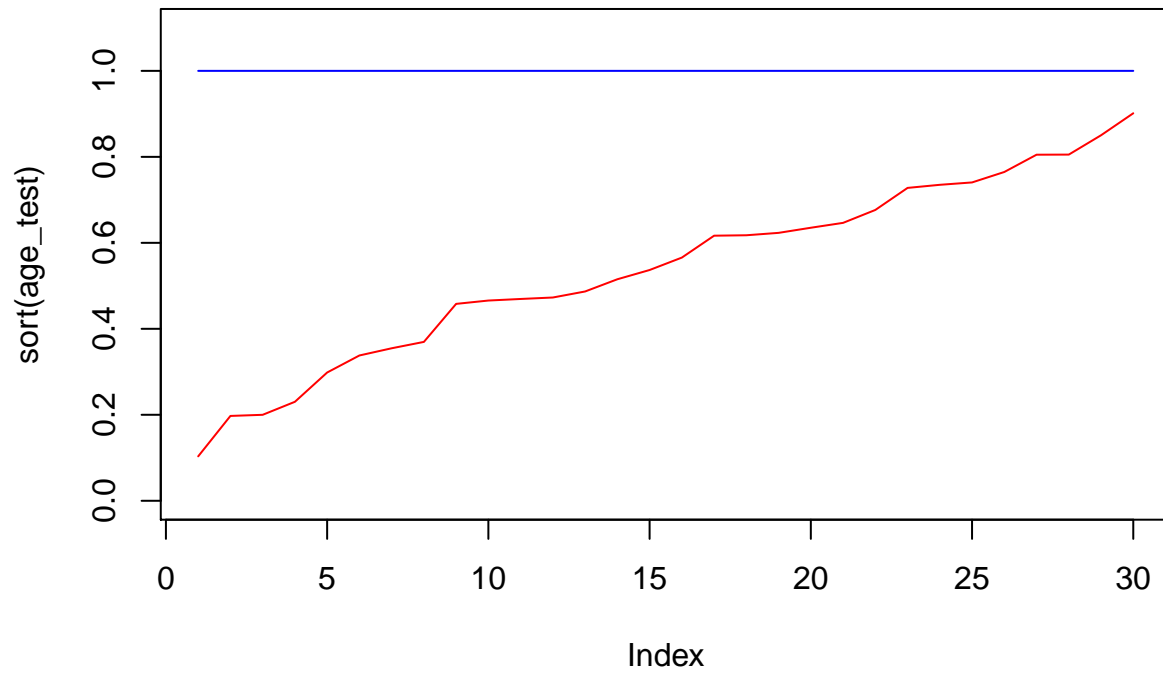
plot(sort(gender_test), col = "red", type = "l",
      main = "Non-Adjusted Gender vs Adjusted Gender\nholm", ylim = c(0,1.1))
points(sort(gender_test_adjusted), col = "blue", type = "l")

```



```
plot(sort(age_test), col = "red", type = "l",  
     main = "Non-Adjusted Age vs Adjusted Age\nholm", ylim = c(0,1.1))  
points(sort(age_test_adjusted), col = "blue", type = "l")
```

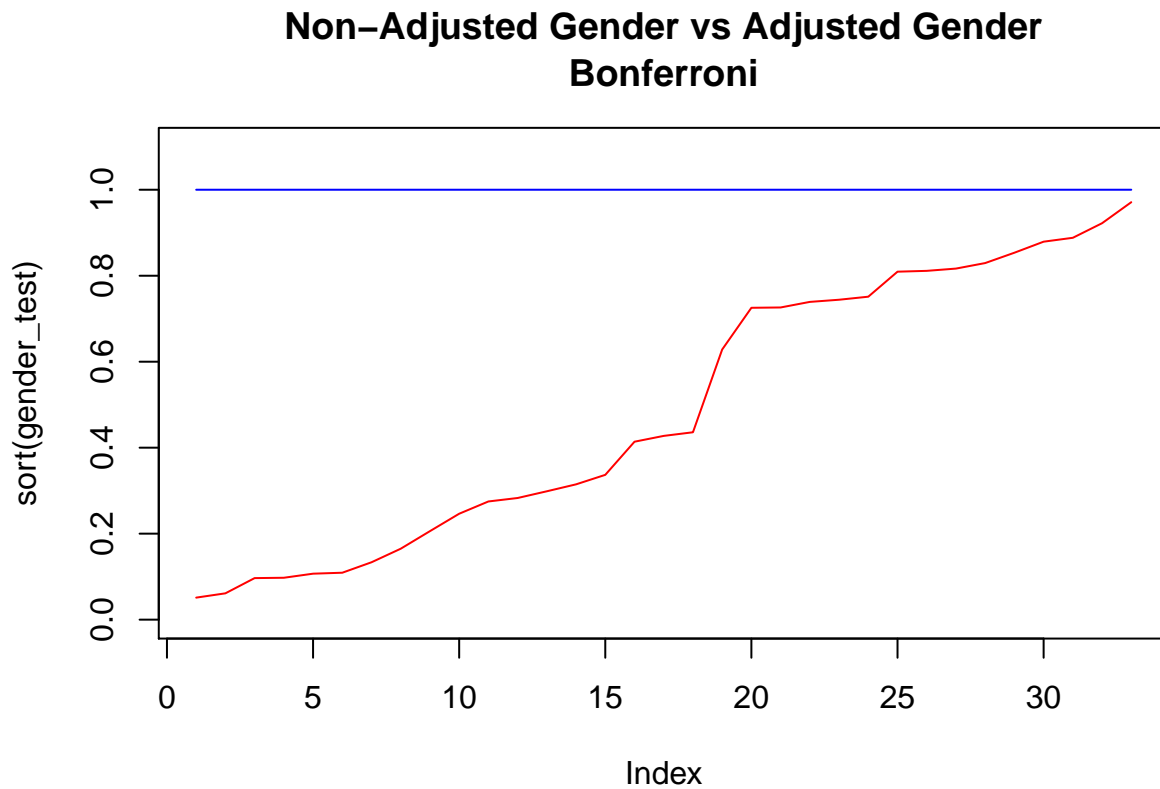
Non-Adjusted Age vs Adjusted Age holm



6. Repeat #4 and #5 with the Bonferroni method.

```
bon_gender_test_adjusted <- p.adjust(gender_test, method = "bonferroni")
bon_age_test_adjusted <- p.adjust(age_test, method = "bonferroni")

plot(sort(gender_test), col = "red", type = "l",
      main = "Non-Adjusted Gender vs Adjusted Gender\nBonferroni", ylim = c(0,1.1))
points(sort(bon_gender_test_adjusted), col = "blue", type = "l")
```



```
plot(sort(age_test), col = "red", type = "l",  
     main = "Non-Adjusted Age vs Adjusted Age\nBonferroni", ylim = c(0,1.1))  
points(sort(bon_age_test_adjusted), col = "blue", type = "l")
```

