ABSTAT17

IGC, April 10–13, 2017

EXERCISES: Multiple Testing

In this exercise we will work with the `multtest` R package for multiple testing, although there are several packages which have in their procedures multiple testing approaches.

Sandrine Dudoit and colleagues implemented the `R multtest` package, which is part of the Bioconductor software, to perform multiple testing analyses. See the `multtest` documentation here.

1. Load the `multtest` package and explore the `mt.teststat` function. Load the data `golub` from the leukemia microarray study of Golub et al. (1999).

   ```
   source("http://bioconductor.org/biocLite.R")
   library(Biobase)
   biocLite("multtest")
   library(multtest)
   ?mt.teststat

   data(golub)
   ?golub
   head(golub)
   golub.cl
   ncol(golub)
   nrow(golub)
   ```

2. Compute the t-statistic and Wilcoxon statistic and compare their normal QQ plots. Are the Wilcoxon results similar to those for the t-test?

The mt.teststat function calculates a test statistic (for example, the t-statistic or the Wilcoxon) for each row in a data frame.

```
?mt.teststat
welch_t<-mt.teststat(golub,golub.cl,test="t")
head(welch_t)
qqnorm(welch_t)
qqline(welch_t)

wilk<-mt.teststat(golub,golub.cl,test="wilcoxon")
qqnorm(wilk)
qqline(wilk)
```

3. Calculate the unadjusted p-value corresponding to each t-statistic in the list determined earlier.

```
raw_p_t<-2*(1-pnorm(abs(welch_t)))
hist(raw_p_t)
plot(sort(raw_p_t))

length(raw_p_t)*0.01
length(raw_p_t[raw_p_t<0.01])
```

4. Calculate the adjusted p-values using the Bonferroni, Holm, and Benjamini-Hochberg methods

The mt.rawp2adjp function computes adjusted p-values for each raw-pvalue.

```
?mt.rawp2adjp
procs = c("Bonferroni", "Holm", "BH")
res = mt.rawp2adjp(raw_p_t, procs)
names(res)
adjp = res$adjp[order(res$index),]
round(adjp,3)
```

5. Look at the number of rejected null hypotheses at successive p-values from alpha = 0.05 to alpha=1, using the Bonferroni, Holm, and Benjamini-Hochberg procedures and compare the results.

The mt.reject function returns the number of rejected hypotheses corresponding to each adjusted p-value for the multiple testing correction procedures you specify.

```
?mt.reject
mt.reject(adjp, c(0.01,0.05))$r
mt.reject(adjp, seq(0,1, 0.05))$r
```

6. Compare de adjusted p-values obtained via packages `multtest`, `qvalue`, `p.adjust` and `fdrtool`. If possible, print also the FDR.

```
biocLite("fdrtool")
library(fdrtool)
fdr<-fdrtool(raw_p_t,statistic="pvalue")
names(fdr)
head(fdr$pval)
head(fdr$qval)
head(fdr$lfdr)

# or

biocLite("qvalue")
library(qvalue)
fdr.q <- qvalue(raw_p_t)
names(fdr.q)
head(fdr.q$pvalues)
head(fdr.q$qvalues)
head(fdr.q$lfdr)

# or

fdr.p <- p.adjust(raw_p_t,"BH")
head(fdr.p)
```