

CHAP 7 - additional exercises

EXO 1

Draw three values from two distributions $N(6,2)$ and $N(8,2)$, perform a t-test and interpret the results.

```
x <- rnorm(3,6,2)
y <- rnorm(3,8,2)
t.test(x,y)

##
## Welch Two Sample t-test
##
## data: x and y
## t = -0.80742, df = 2.5873, p-value = 0.487
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.049075 1.903270
## sample estimates:
## mean of x mean of y
## 5.648934 6.221836
```

Repeat the above experiment multiple times. Do you get similar results? Why?
p-values vary widely, sometimes significant (<0.05) sometimes not

Repeat with values from distributions $N(5,1)$ and $N(8,1)$.

```
x <- rnorm(3,5,1)
y <- rnorm(3,8,1)
t.test(x,y)

##
## Welch Two Sample t-test
##
## data: x and y
## t = -2.2525, df = 2.9096, p-value = 0.1124
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -6.376496 1.145582
## sample estimates:
## mean of x mean of y
## 5.120602 7.736059
```

p-values generally lower, significant most of the time

EXO 2

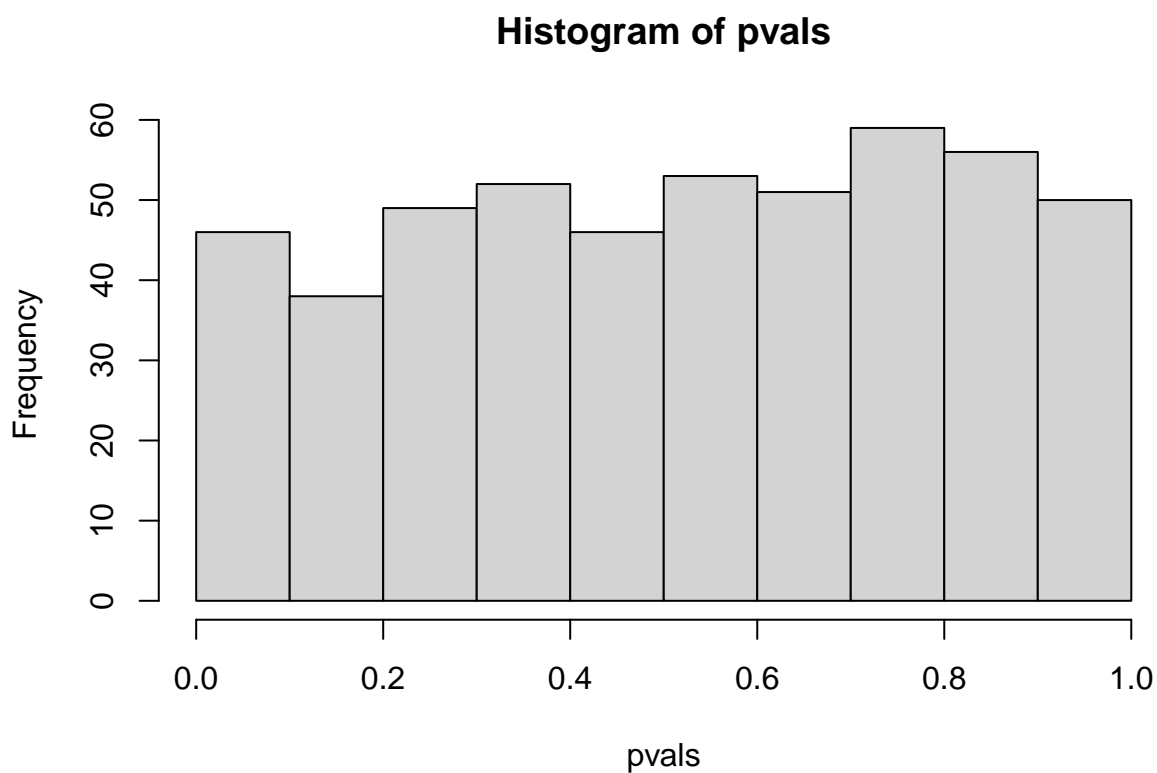
Generate random data using `rnorm` for 100 genes and 6 samples and test for differential expression, comparing for each gene the 3 first samples against the 3 last samples. Verify

that you identify about 5 p-values smaller than 0.05. Visualise and interpret the histogram of these p-values.

```
exprs <- matrix(rnorm(3000),ncol=6)

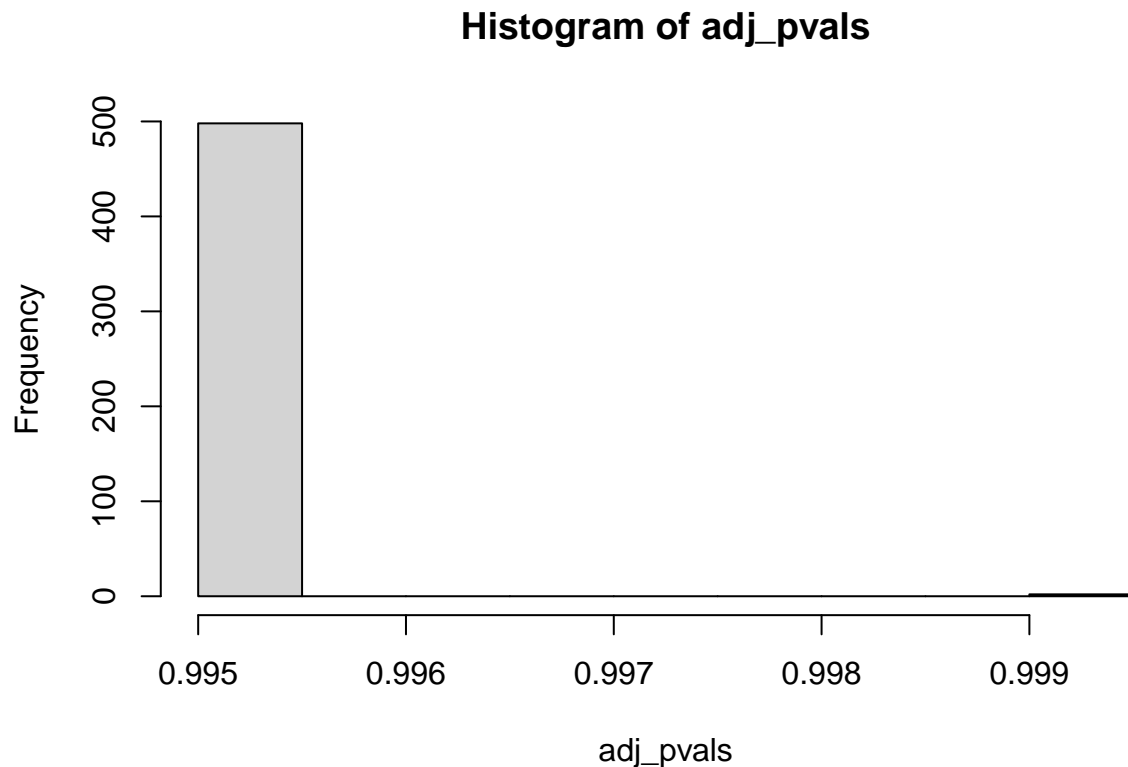
my_t_test <- function(x) {
  t.test(x[1:3],x[4:6])$p.value }
pvals <- apply(exprs, 1, my_t_test)

hist(pvals)
```



Adjust the 100 p-values for multiple testing and compare the initial and adjusted p-values.

```
adj_pvals <- p.adjust(pvals, method = "BH")
hist(adj_pvals)
```



After adjustment, big p-values

EXO 3

We have used two-sided two-sample t-tests above. Also familiarise yourselves with one-sided, one-sample and paired tests. In particular, verify how to implement these with the `t.test` function.

arguments of `t.test()` function : one-sample t-test : `mu = x` one-sided t-test : `alternative = "less"/"greater"`
 paired t-test : `paired = TRUE`

EXP 4

Stimulate a dataset of log2 fold-changes measured in triplicate for 1000 genes.

What function would you use to generate these data?

What test would you use to test for differential expression? Apply it to calculate 1000 p-values.

Visualise and interpret the histogram of p-values.

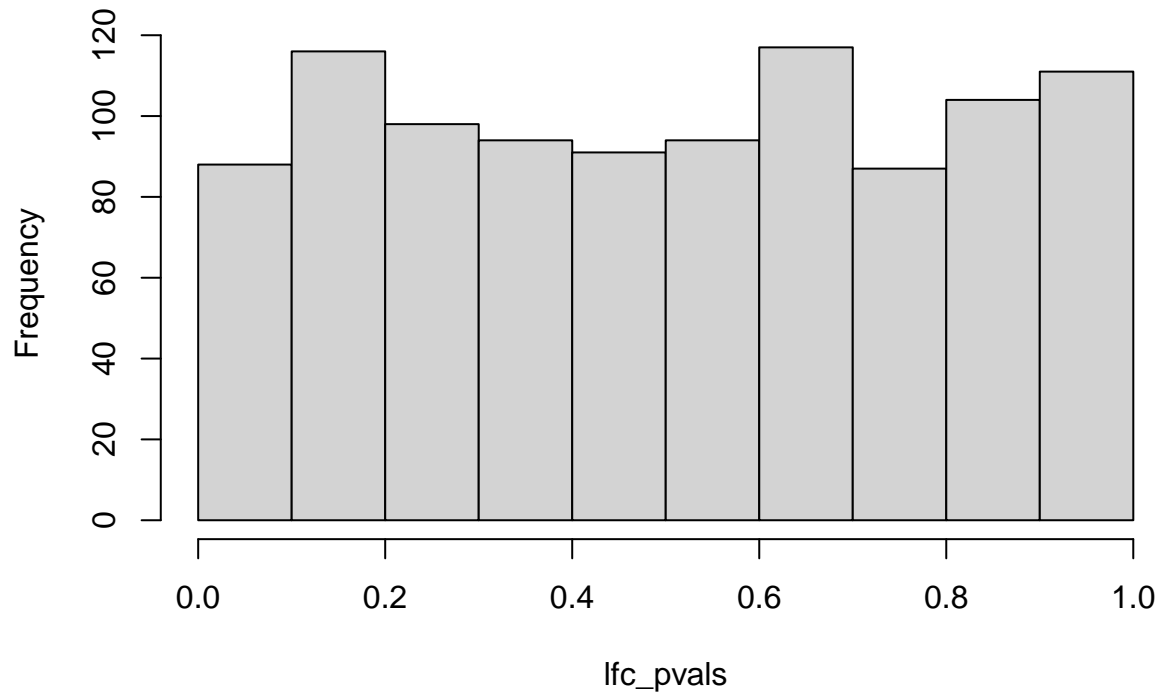
FDR-adjust the p-values.

```
lfc <- matrix(rnorm(3000),ncol=3)

lfc_test <- function(x) {
  t.test(x[1:3], mu = 0)$p.value }
lfc_pvals <- apply(lfc,1,lfc_test)

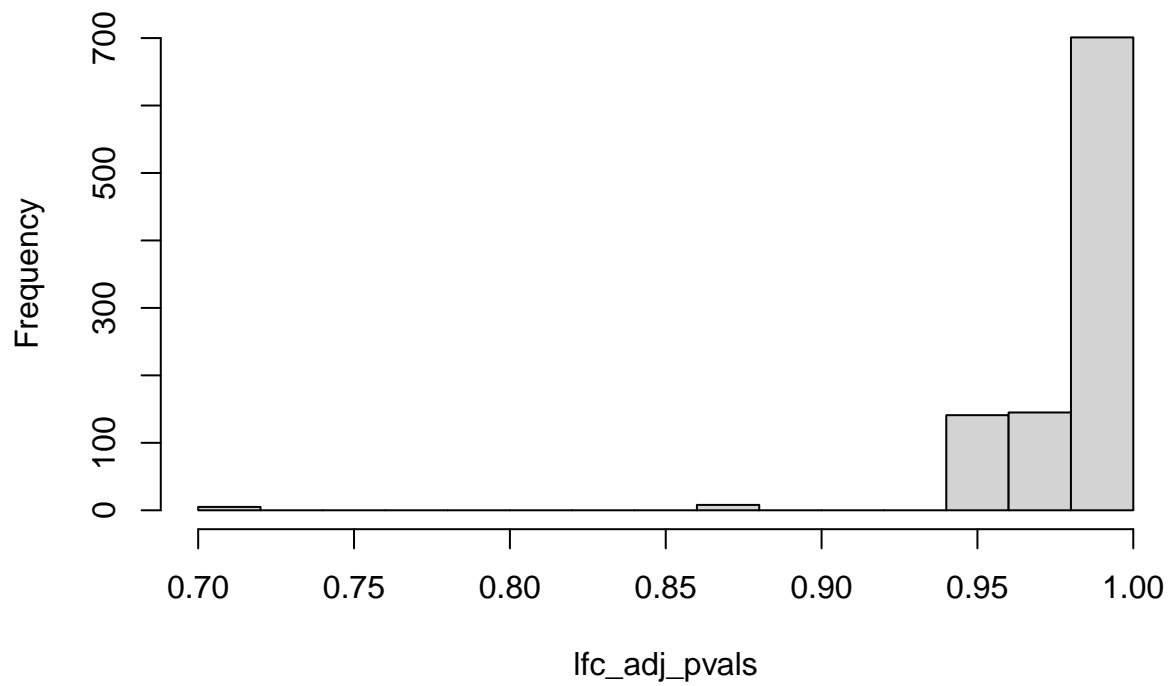
hist(lfc_pvals)
```

Histogram of lfc_pvals



```
lfc_adj_pvals <- p.adjust(lfc_pvals, method = "BH")  
hist(lfc_adj_pvals)
```

Histogram of lfc_adj_pvals



EXO 5

Load the `tdata2` dataset from the `rWSBIM1322` package. These data represent the measurement of a inflammation biomarker in the blood of 15 patients, before and after treatment for a acute liver inflammation. Visualise the data and run a test to verify if the treatment has had an effect or not.