

Exercise L3

2022-06-11

EXERCISE 1

1. Copy the “iris” dataset into a new data frame called “iris2”:

```
iris2 <- iris #or also 'iris2 <- data.frame(iris)  
head(iris2)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1         5.1         3.5         1.4         0.2  setosa  
## 2         4.9         3.0         1.4         0.2  setosa  
## 3         4.7         3.2         1.3         0.2  setosa  
## 4         4.6         3.1         1.5         0.2  setosa  
## 5         5.0         3.6         1.4         0.2  setosa  
## 6         5.4         3.9         1.7         0.4  setosa
```

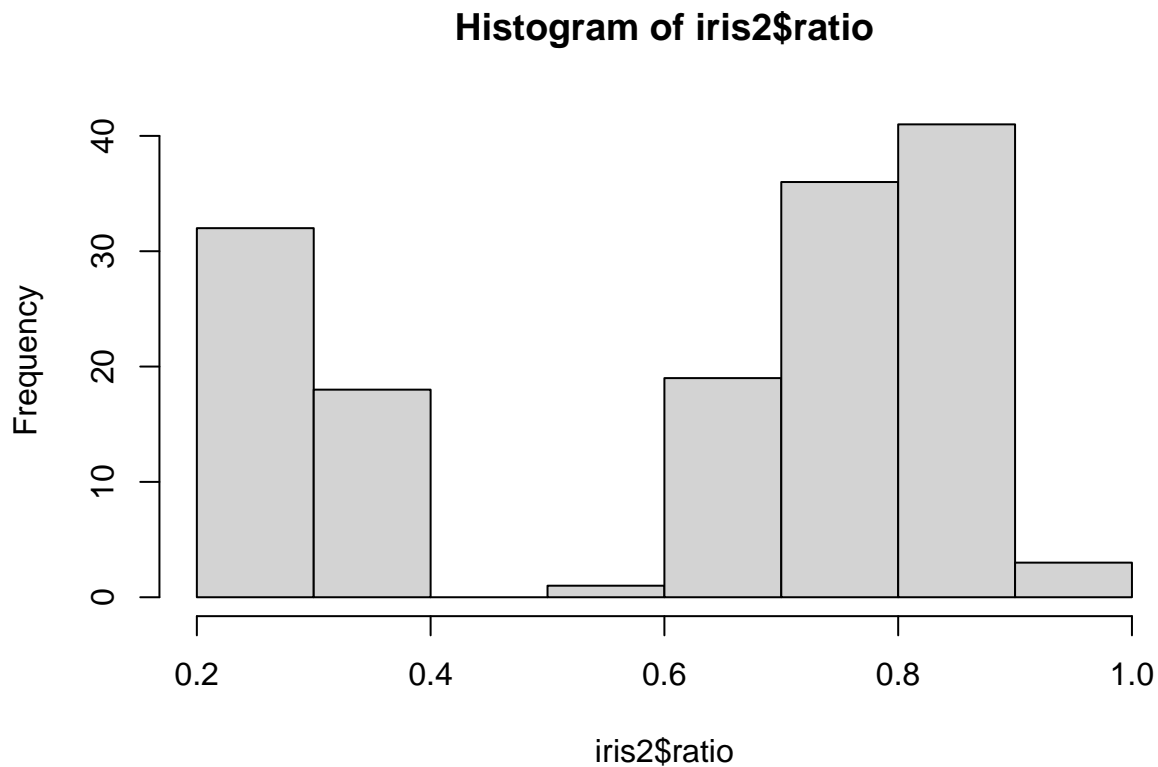
2. Add a column to “iris2”, named “ratio”, containing the ratio of Petal.Length to Sepal.Length for each flower:

```
iris2$ratio <- iris2$Petal.Length/iris2$Sepal.Length  
head(iris2)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species    ratio  
## 1         5.1         3.5         1.4         0.2  setosa 0.2745098  
## 2         4.9         3.0         1.4         0.2  setosa 0.2857143  
## 3         4.7         3.2         1.3         0.2  setosa 0.2765957  
## 4         4.6         3.1         1.5         0.2  setosa 0.3260870  
## 5         5.0         3.6         1.4         0.2  setosa 0.2800000  
## 6         5.4         3.9         1.7         0.4  setosa 0.3148148
```

3. Produce a histogram of the ratio:

```
hist(iris2$ratio)
```



EXERCISE 2

- Using the “diffexpr” data frame defined above, add a column named “FC” containing the true (non log) fold change:

```
load("C:/Users/seren/Desktop/Provero/image.RData")
```

```
FC <- 2^diffexpr[, "logFC"]
```

```
diffexpr$FC <- FC
```

```
head(diffexpr)
```

```
##           P      logFC      FC
## SERPINC1 1.085867e-05  0.3830255 1.30407377
## NETO2    4.284471e-05 -3.0850230 0.11784619
## ITCH     5.185839e-05 -1.1894569 0.43846788
## CLDN10   7.505848e-05 -3.5143170 0.08751554
## ERF      8.560227e-05 -0.9486367 0.51812184
## PAK2     1.124295e-04 -1.0847683 0.47146799
```

- Add a column of type “logical” and named “is_deg” which is TRUE if the P-value is < 0.01 and FALSE otherwise:

```
is_deg <- diffexpr$P < 0.01
```

```
class(is_deg)
```

```
## [1] "logical"
```

```
diffexpr$is_deg <- is_deg  
head(diffexpr)
```

```
##           P      logFC      FC is_deg  
## SERPINC1 1.085867e-05  0.3830255 1.30407377 TRUE  
## NETO2    4.284471e-05 -3.0850230 0.11784619 TRUE  
## ITCH     5.185839e-05 -1.1894569 0.43846788 TRUE  
## CLDN10   7.505848e-05 -3.5143170 0.08751554 TRUE  
## ERF      8.560227e-05 -0.9486367 0.51812184 TRUE  
## PAK2     1.124295e-04 -1.0847683 0.47146799 TRUE
```

```
tail(diffexpr)
```

```
##           P      logFC      FC is_deg  
## USP19    0.9982973 -3.526016e-04 0.9997556 FALSE  
## FGD6     0.9992218  4.395015e-04 1.0003047 FALSE  
## TNIP1    0.9993297 -4.234076e-04 0.9997066 FALSE  
## CHPF     0.9995846  2.072582e-04 1.0001437 FALSE  
## PCDHGB6  0.9996020  8.575406e-05 1.0000594 FALSE  
## KIAA1539 0.9996346 -1.121649e-04 0.9999223 FALSE
```

3. Order the dataframe alphabetically by gene name:

```
x <- rownames(diffexpr)  
diffexpr2 <- data.frame(diffexpr[order(x), ])  
head(diffexpr2)
```

```
##           P      logFC      FC is_deg  
## A2BP1    0.14651829 -0.51512168 0.6997339 FALSE  
## A2M      0.30715139  0.66331593 1.5837185 FALSE  
## A4GALT   0.72546554  0.05321985 1.0375780 FALSE  
## A4GNT    0.13041274  0.32306724 1.2509874 FALSE  
## AAAS     0.03670293 -0.28801857 0.8190262 FALSE  
## AACS     0.04747381 -0.68530650 0.6218737 FALSE
```