R Lessons Answers

Stefan Bekiranov
BIOC 8145

April 15, 2020

Contents

1 Introduction to R

```
> library(readr)
> library(dplyr)
> ydat <- read_csv(file="brauer2007_tidy.csv")
```

Exercise 1. What are the values after each statement in the following?

```
> mass <- 50              # mass?
> mass

[1] 50

> age  <- 30              # age?
> age

[1] 30

> mass <- mass * 2        # mass?
> mass

[1] 100

> age  <- age - 10        # age?
> age
```

```
[1] 20
```

```
> mass_index <- mass/age  # massIndex?
> mass_index
```

```
[1] 5
```

Exercise 2. Calculate the square root of the log-base-10 of the absolute value of `-4*(2550-50)`.

```
> sqrt(log10(abs(-4*(2550-50))))
```

```
[1] 2
```

Exercise 3. What is the standard deviation of expression? What's the range of rate in the data?

```
> sd(ydat$expression)
```

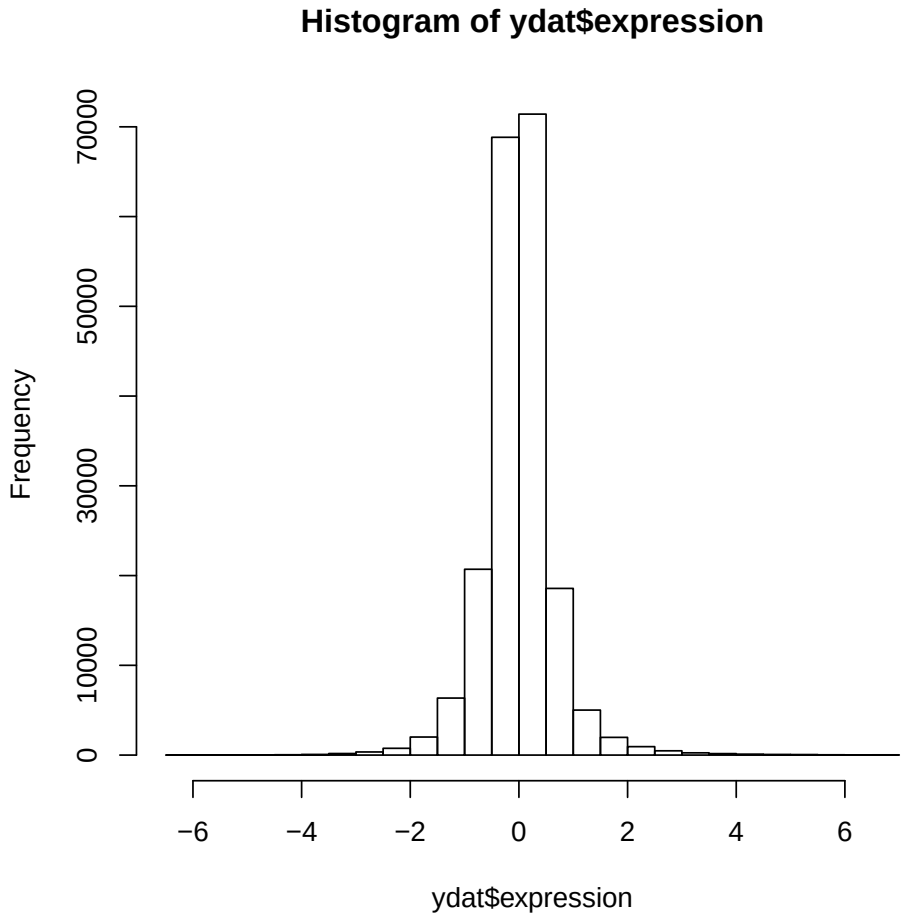
```
[1] 0.6675023
```

```
> range(ydat$rate)
```

```
[1] 0.05 0.30
```

6d. Generate a histogram of gene expression by executing the command `hist(ydat$expression)`.

```
> hist(ydat$expression)
```



2 R Data Manipulation

Exercise 1.1

```
> filter(ydat, bp=="leucine biosynthesis" & nutrient=="Leucine")

# A tibble: 24 x 7
   symbol systematic_name nutrient  rate expression bp          mf
   <chr>  <chr>           <chr>    <dbl>      <dbl> <chr>       <chr>
 1 LEU9   YOR108W         Leucine   0.05      0.44  leucine bi... 2-isopropylmala...
 2 LEU1   YGL009C         Leucine   0.05      3.84  leucine bi... 3-isopropylmala...
 3 LEU2   YCL018W         Leucine   0.05      1.54  leucine bi... 3-isopropylmala...
 4 LEU4   YNL104C         Leucine   0.05      1.94  leucine bi... 2-isopropylmala...
 5 LEU9   YOR108W         Leucine   0.1       0.570 leucine bi... 2-isopropylmala...
 6 LEU1   YGL009C         Leucine   0.1       3.36  leucine bi... 3-isopropylmala...
```

```
 7 LEU2   YCL018W         Leucine   0.1       1.23  leucine bi... 3-isopropylmala...
 8 LEU4   YNL104C         Leucine   0.1       1.71  leucine bi... 2-isopropylmala...
 9 LEU9   YOR108W         Leucine   0.15      0.46  leucine bi... 2-isopropylmala...
10 LEU1   YGL009C         Leucine   0.15      3.24  leucine bi... 3-isopropylmala...
# ... with 14 more rows
```

Exercise 1.2

```
> quantile(ydat$expression, probs=0.99)

 99%
2.07

> filter(ydat, expression > quantile(ydat$expression, probs=0.99))

# A tibble: 1,971 x 7
   symbol systematic_name nutrient  rate expression bp          mf
   <chr>  <chr>           <chr>    <dbl>      <dbl> <chr>       <chr>
 1 ATO3   YDR384C         Glucose   0.05       2.27 transport*  transporter act...
 2 <NA>   YKL187C         Glucose   0.05       4.13 biological... molecular funct...
 3 <NA>   YGL117W         Glucose   0.05       2.3  biological... molecular funct...
 4 <NA>   YBR047W         Glucose   0.05       2.14 biological... molecular funct...
 5 SNZ1   YMR096W         Glucose   0.05       3.71 pyridoxine... protein binding
 6 SNO1   YMR095C         Glucose   0.05       2.4  pyridoxine... molecular funct...
 7 SSU1   YPL092W         Glucose   0.05       3.22 sulfite tr... sulfite transpo...
 8 BAP3   YDR046C         Glucose   0.05       2.19 amino acid... amino acid tran...
 9 PDR12  YPL058C         Glucose   0.05       2.38 transport*  xenobiotic-tran...
10 PHO89  YBR296C         Glucose   0.05       2.54 phosphate ... sodium:inorgani...
# ... with 1,961 more rows

View(filter(ydat, expression > quantile(ydat$expression, probs=0.99)))
```

They were certainly involved in various metabolic pathways/processes.

Exercise 2.1

```
> filter(ydat, bp=="leucine biosynthesis" & nutrient=="Leucine")

# A tibble: 24 x 7
   symbol systematic_name nutrient  rate expression bp          mf
   <chr>  <chr>           <chr>    <dbl>      <dbl> <chr>       <chr>
 1 LEU9   YOR108W         Leucine   0.05       0.44  leucine bi... 2-isopropylmala...
 2 LEU1   YGL009C         Leucine   0.05       3.84  leucine bi... 3-isopropylmala...
 3 LEU2   YCL018W         Leucine   0.05       1.54  leucine bi... 3-isopropylmala...
 4 LEU4   YNL104C         Leucine   0.05       1.94  leucine bi... 2-isopropylmala...
```

```
 5 LEU9   YOR108W         Leucine   0.1       0.570 leucine bi... 2-isopropylmala...
 6 LEU1   YGL009C         Leucine   0.1       3.36  leucine bi... 3-isopropylmala...
 7 LEU2   YCL018W         Leucine   0.1       1.23  leucine bi... 3-isopropylmala...
 8 LEU4   YNL104C         Leucine   0.1       1.71  leucine bi... 2-isopropylmala...
 9 LEU9   YOR108W         Leucine   0.15      0.46  leucine bi... 2-isopropylmala...
10 LEU1   YGL009C         Leucine   0.15      3.24  leucine bi... 3-isopropylmala...
# ... with 14 more rows
```

Exercise 2.2

```
> arrange(filter(ydat, bp=="leucine biosynthesis" & nutrient=="Leucine"), symbol)

# A tibble: 24 x 7
   symbol systematic_name nutrient  rate expression bp          mf
   <chr>  <chr>           <chr>    <dbl>      <dbl> <chr>       <chr>
 1 LEU1   YGL009C         Leucine   0.05       3.84 leucine bi... 3-isopropylmala...
 2 LEU1   YGL009C         Leucine   0.1        3.36 leucine bi... 3-isopropylmala...
 3 LEU1   YGL009C         Leucine   0.15       3.24 leucine bi... 3-isopropylmala...
 4 LEU1   YGL009C         Leucine   0.2        2.84 leucine bi... 3-isopropylmala...
 5 LEU1   YGL009C         Leucine   0.25       2.04 leucine bi... 3-isopropylmala...
 6 LEU1   YGL009C         Leucine   0.3        0.87 leucine bi... 3-isopropylmala...
 7 LEU2   YCL018W         Leucine   0.05       1.54 leucine bi... 3-isopropylmala...
 8 LEU2   YCL018W         Leucine   0.1        1.23 leucine bi... 3-isopropylmala...
 9 LEU2   YCL018W         Leucine   0.15       0.69 leucine bi... 3-isopropylmala...
10 LEU2   YCL018W         Leucine   0.2        0.39 leucine bi... 3-isopropylmala...
# ... with 14 more rows
```

Excercise 2.3

```
View(arrange(filter(ydat, bp=="leucine biosynthesis" & nutrient=="Leucine"), sym-
bol))
```