

Intro to R for Epidemiologists

Lab 1 (1/15/15)

Many of these questions go beyond the information provided in the lecture. Therefore, you may need to use R help files and the internet to search for answers. Feel free to ask questions of the instructor, the TAs, or your classmates, but try to work through as much as you can independently.

For the lab, you are expected to create an R script (.R file in the R editor) with your code corresponding to each question. Begin each question with a commented line of code indicating the question. As an example:

```
# Jenna Krall  
  
# Question 1.  
head(iris)
```

Part 1. Characteristics of irises

1. The dataset `iris` in R comes with the preloaded `datasets` R package. Read about the data in the corresponding R help file. What are the variables included in this dataset? Take a look at the data using `head(iris)`.

```
# Read help page  
`?`(iris)  
# Find variables  
colnames(iris)
```

```
## [1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width"  
## [5] "Species"
```

```
# Look at the data  
head(iris)
```

```
##   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. What is the class of the `iris` dataset?

```
# Class of dataset  
class(iris)
```

```
## [1] "data.frame"
```

3. How many rows and columns does the `iris` dataset have? What R functions directly give the number of rows and columns?

```
# Number of rows  
nrow(iris)
```

```
## [1] 150
```

```
# Number of columns  
ncol(iris)
```

```
## [1] 5
```

4. Read the help page for `$` (use `?"$"`). How can you use `$` to subset the `iris` dataset? Use `$` to compute the mean sepal length in the `iris` dataset.

```
# Find mean sepal length  
mean(iris$Sepal.Length)
```

```
## [1] 5.843333
```

5. What is the R function for standard deviation? Read the help page for this function. Compute the standard deviation of sepal width in the `iris` dataset.

```
# Standard deviation of sepal width  
sd(iris$Sepal.Width)
```

```
## [1] 0.4358663
```

6. What is the R function for obtaining quantiles of a vector? Compute the 20th and 80th percentiles of petal length.

```
# Quantiles of petal length  
quantile(iris$Petal.Length, probs = c(0.2, 0.8))
```

```
## 20% 80%  
## 1.50 5.32
```