

Lab Worksheet 1 Solution

This worksheet uses the `iris` data set available in R. This data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are *Iris setosa*, *versicolor*, and *virginica*:

```
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
```

Problem 1: Is there a difference in sepal length between species *setosa* and species *virginica*? Perform a t test and discuss your results (1-2 sentences).

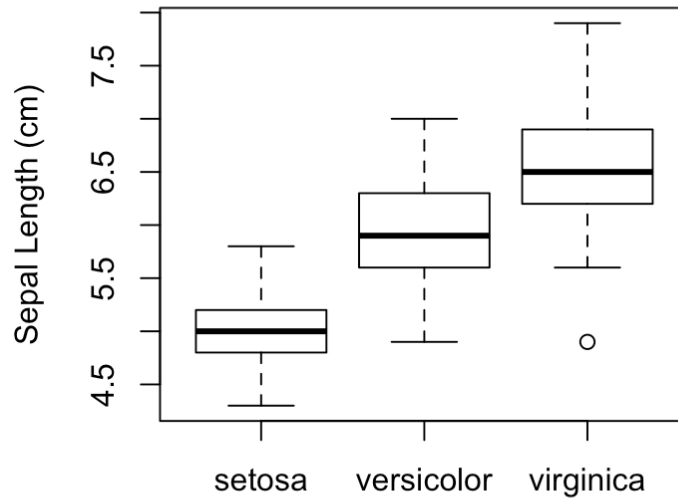
```
t.test(iris$Sepal.Length[iris$Species=='setosa'], iris$Sepal.Length[iris$Species=='virginica'])
```

```
##
## Welch Two Sample t-test
##
## data:  iris$Sepal.Length[iris$Species == "setosa"] and iris$Sepal.Length[iris$Species == "virginica"]
## t = -15.386, df = 76.516, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.78676 -1.37724
## sample estimates:
## mean of x mean of y
##      5.006      6.588
```

There is a significant difference in sepal length. Sepals of species *virginica* are on average 1.6 cm longer than sepals of species *setosa*.

Problem 2: Make side-by-side box plots of sepal length for the three species. Discuss what patterns you observe (1-2 sentences).

```
boxplot(iris$Sepal.Length ~ iris$Species, ylab="Sepal Length (cm)")
```



Sepal length seems to increase from *setosa* to *versicolor* to *virginica*.

Problem 3: Make a scatter plot of sepal length vs. petal length for the three species. Make a single plot that shows the data for all three species at once, in different colors. Hint: To see all data in one plot, you will have to manually set the plot limits, using the `xlim` and `ylim` parameters of the `plot` function. Discuss your results (1-2 sentences).

```
setosa <- iris[iris$Species=='setosa',]
versicolor <- iris[iris$Species=='versicolor',]
virginica <- iris[iris$Species=='virginica',]
plot(setosa$Sepal.Length, setosa$Petal.Length, pch=19, col='blue', xlim=c(3,8),
ylim=c(1,8), xlab = "Sepal Length (cm)", ylab = "Petal Length (cm)")
points(versicolor$Sepal.Length, versicolor$Petal.Length, pch=19, col='red')
points(virginica$Sepal.Length, virginica$Petal.Length, pch=19, col='green')
```

