Homework2

Swetha Jayapathy

10/13/2020

R Markdown

```
library(Sleuth3)
library(ggplot2)
```

```
HW2Dat <- read.csv("D:/D drive contents/Fall 2020/Stats 511/Class/HW/HW2/NavDat.csv")
```

```
head(HW2Dat)
```

```
##   Subject       DT        M
## 1       1 169.2650 152.9350
## 2       2 154.7725 102.3075
## 3       3 149.7275 136.2075
## 4       4 147.6100 141.2450
## 5       5 154.8400 137.3550
## 6       6 146.0300 111.2400
```

Question 1a :

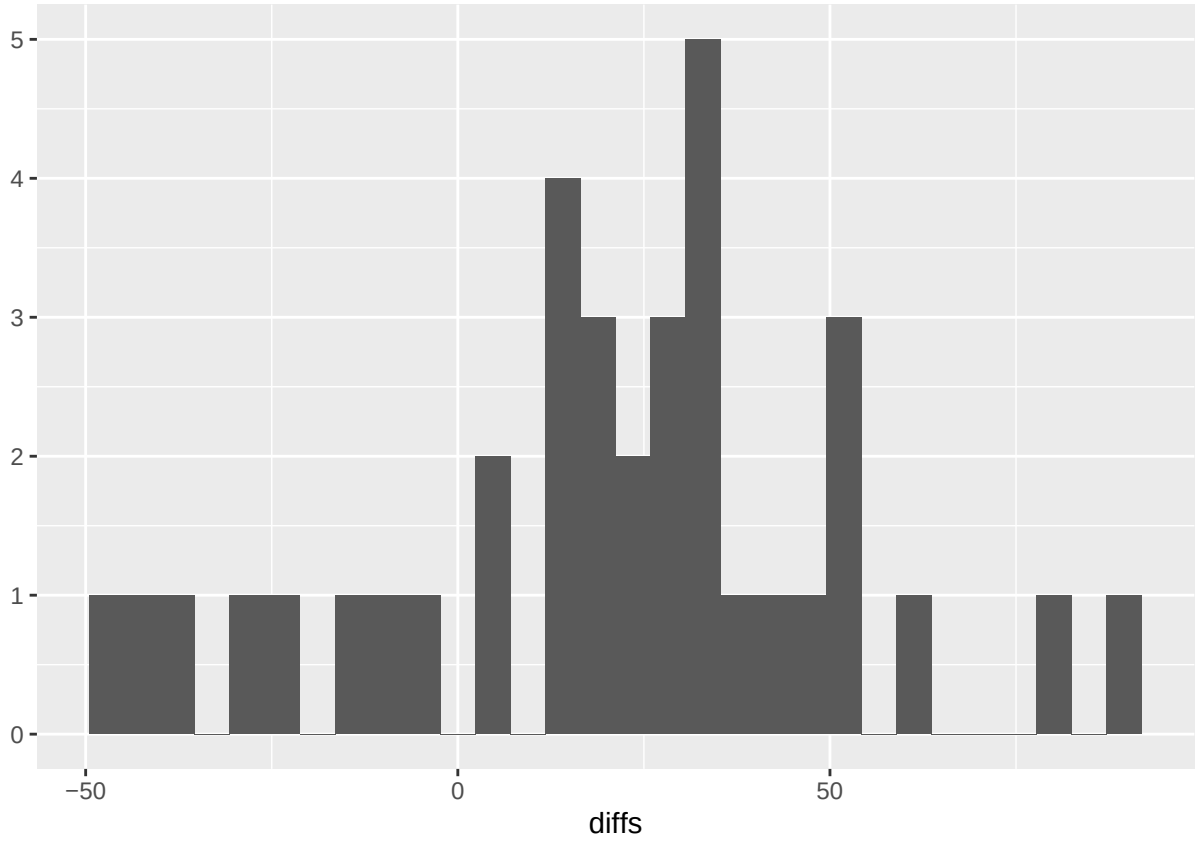
```
diffs <- with(HW2Dat, DT-M)
diffs
```

```
##  [1]  16.3300  52.4650  13.5200   6.3650  17.4850  34.7900 -38.0625  32.3100
##  [9]  17.4375 -14.8075  15.6125  91.8775  47.1175  39.3650 -21.4200  30.9425
## [17] -40.2625  29.4650  25.6900  -3.8600  78.6925  33.4375 -44.7850  41.5200
## [25]  13.6550   3.7500  22.5325  -7.6175  29.3800  61.3300  50.0825 -26.0500
## [33]  18.7250  27.3850  34.0300  54.1650
```

Question 1b : qplot(diffs, geom="histogram")

```
qplot(diffs, geom="histogram")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Question 1c

```
(Ybar <- mean(diffs)) #Calculate the sample mean of the differences
```

```
## [1] 20.62757
```

```
(s <- sd(diffs))     #Calculate the sample standard deviation
```

```
## [1] 31.30602
```

```
(n <- length(diffs))         #Find the sample size
```

```
## [1] 36
```

Question 1d

```
(se_Ybar <- s/sqrt(n))    #Calculate the SE of the sample mean
```

```
## [1] 5.21767
```

```
Ybar/se_Ybar #t-statistics
```

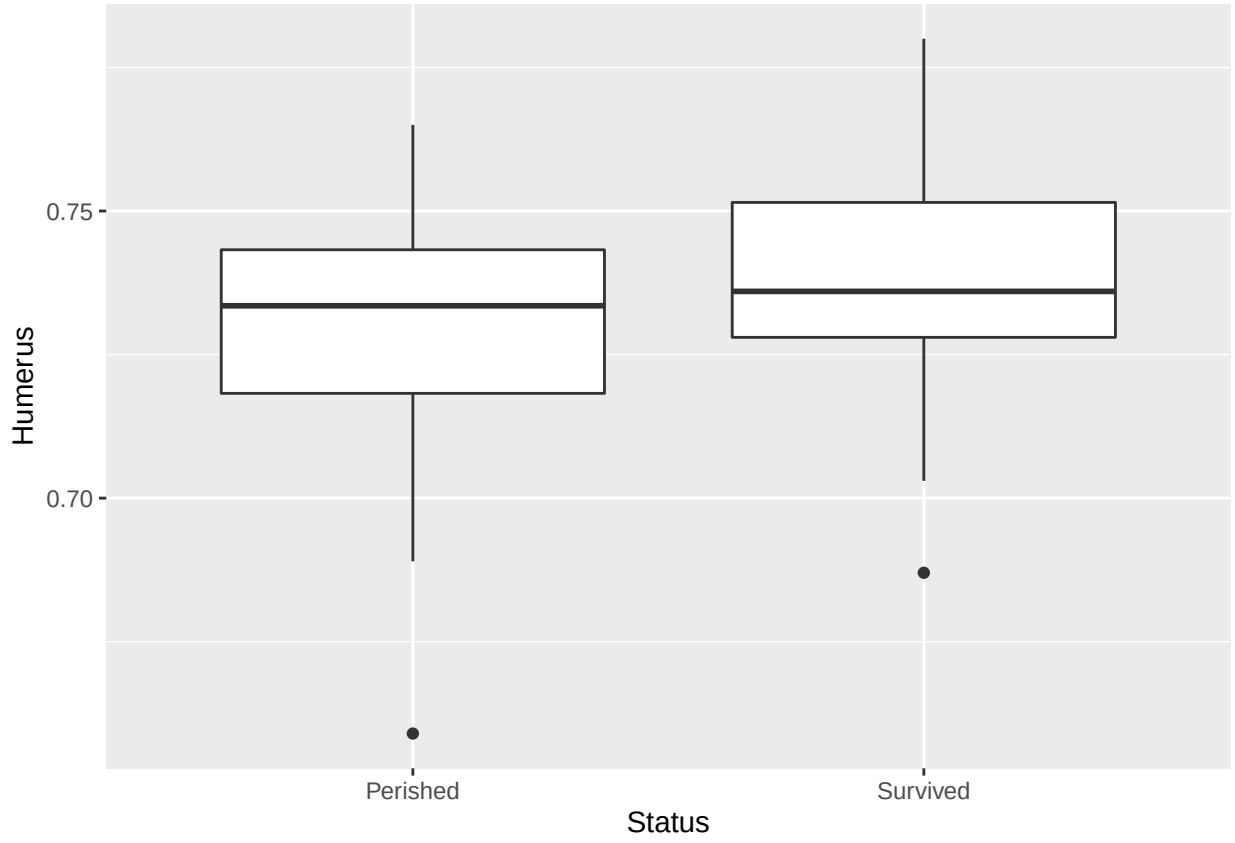
```
## [1] 3.953406
```

Question 1e

```
with(HW2Dat, t.test(DT, M, paired=TRUE))
```

```
##
##  Paired t-test
##
## data:  DT and M
## t = 3.9534, df = 35, p-value = 0.0003572
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  10.03514 31.22000
## sample estimates:
## mean of the differences
##                20.62757
```

Question 2a



Statistical Conclusion:

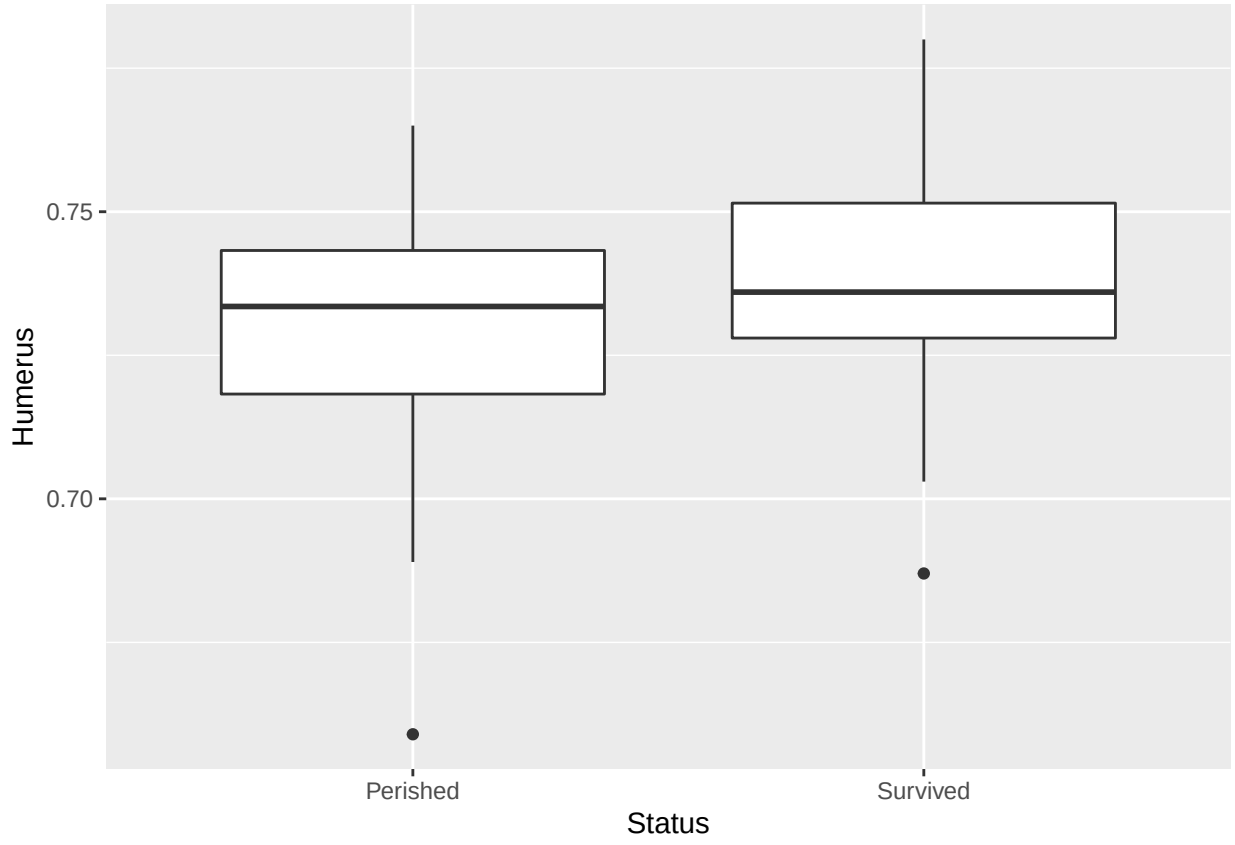
According to the 2 sample t-test, we can say that there is no significant evidence against the null hypothesis ; that is the mean difference of the length of humerus between the perished and the survived is non-zero (p-value = 0.0809). The mean differences between that of perished and survived from the t-test is 0.01 with a 95% confidence interval in the range -0.021446053 to 0.001279386.

Question 2b

```
# To get 2 sided Confidence Interval
t.test(Humerus~Status, data=ex0221, var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  Humerus by Status
## t = -1.777, df = 57, p-value = 0.0809
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.021446053  0.001279386
## sample estimates:
## mean in group Perished mean in group Survived
##              0.7279167              0.7380000
```

```
qplot(Status, Humerus, data=ex0221, geom="boxplot")
```



Question 3 (a) Describe what is meant by the sampling distribution of the sample standard deviation.

Sampling distribution of the sample standard deviation is the distribution under repeated sampling of the standard deviation which is computed by taking particular sample size of n samples from a population. This is the distribution which we get when we take an arbitrary number of samples and calculate the standard deviation of all the samples to plot in graph.

(b) Describe what is meant by the sampling distribution of the sample maximum.

Sampling distribution of the sample maximum is the distribution under repeated sampling of the maximum value which is computed by taking particular sample size of n samples from a population. This requires us to compute the sample maximum for each of the n samples. The distribution of these n sample maximum is the sampling distribution of sample standard deviations.