Part2Report

Fernando Campaña

9/15/2019

Part 2: Basic Inferential Data Analysis

As part of the project, I will use the ToothGrowth dataset, executing the previous command gives us the help page of the dataset.

```
?ToothGrowth
```

The help page contain the description that reads: "The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC)."

In order to analyze the ToothGrowth dataset, it must be loaded, which can be accomplished with the following R chunk:

```
data(ToothGrowth)
```

Once succesfully loaded, we can see a sneak peek of the first values, so to get a sense of the data (types and columns names). What we get is a data frame with 60 observations on 3 variables

```
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

Alternatively, we can also get a summary as follows:

```
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

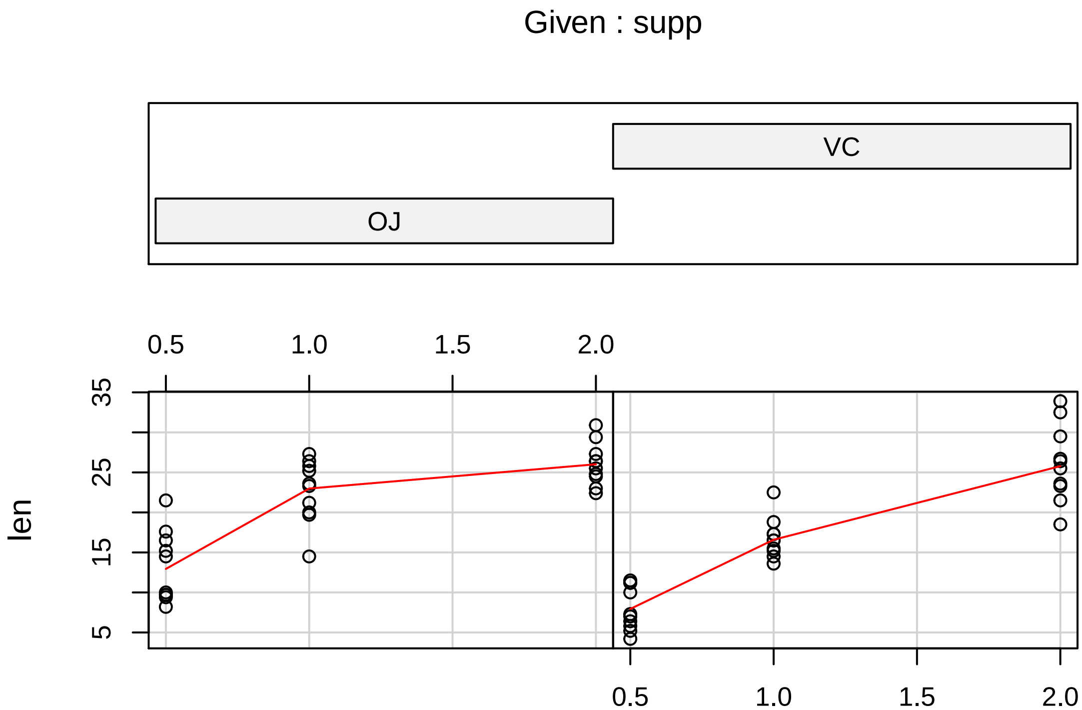
```
summary(ToothGrowth)
```

```
##       len        supp         dose      
##  Min.   : 4.20   OJ:30   Min.   :0.500  
##  1st Qu.:13.07   VC:30   1st Qu.:0.500  
##  Median :19.25           Median :1.000  
##  Mean   :18.81           Mean   :1.167  
##  3rd Qu.:25.27           3rd Qu.:2.000  
##  Max.   :33.90           Max.   :2.000  
```

So, we notice that the len and dose variables contain double values, the len range is from 4.20 to 33.90 and the dose range is from 0.5 to 2. While the supp variable contains a factor of 2 levels ("OJ" and "VC"), divided equitatively, 30 for each.

Let's graph this info, to get a visual insight.

```
require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```



TootGrowth data: length vs dose, given type of supplement As we can see, there is no 1.5 dose data, because they were only given 0.5, 1.0 and 2.0 mg/day of Vitamin C. It seems to be that the tooth length increases as the dose increases, in both the orange juice (OJ) and ascorbic acid (VC) groups.

To perform the hypothesis tests to compare tooth growth by supp and dose, we assume the variance is not equal between datasets. Also 0.05 is the assumed value for the tolerance of the error alpha.

Comparing supplements effect

- Ho: Mean is equal to 0 (No difference in tooth growth between OJ and VC supplement, equally effective)
- Ha: difference in means is not equal to 0 (One supplement is more effective than the other)

```
oj <- ToothGrowth %>% filter(supp == "OJ")
vc <- ToothGrowth %>% filter(supp == "VC")
t.test(oj$len, vc$len, paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  oj$len and vc$len
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
```

```
## mean of x mean of y
##  20.66333  16.96333
```

Since the p-value is higher than alpha (0.05), is not low enough to reject the null hypothesis. So we conclude, both supplements are equally effective at tooth growth.

Comparing supplement effect at different dossage

Dose: 0.5

- Ho: Mean is equal to 0 (No difference in tooth growth between OJ and VC supplement, equally effective, at 0.5 dose)
- Ha: difference in means is greater than 0 (OJ supplement is more effective than the other at a 0.5 dose)

```r
dose05VC <- ToothGrowth %>% filter(dose == 0.5 & supp == "VC")
dose05OJ <- ToothGrowth %>% filter(dose == 0.5 & supp == "OJ")
t.test(dose05OJ$len, dose05VC$len, alternative = "greater", paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  dose05OJ$len and dose05VC$len
## t = 3.1697, df = 18, p-value = 0.002652
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.377886      Inf
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

Due that the p-value is really small compared to the alpha, we can safely reject the null hypothesis and state that the OJ supplement is better than VC at 0.5 dose.

Dose: 1.0

- Ho: Mean is equal to 0 (No difference in tooth growth between OJ and VC supplement, equally effective, at 1.0 dose)
- Ha: difference in means is greater than 0 (OJ supplement is more effective than the other at a 1.0 dose)

```r
dose10VC <- ToothGrowth %>% filter(dose == 1.0 & supp == "VC")
dose10OJ <- ToothGrowth %>% filter(dose == 1.0 & supp == "OJ")
t.test(dose10OJ$len, dose10VC$len, alternative = "greater", paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  dose10OJ$len and dose10VC$len
## t = 4.0328, df = 18, p-value = 0.0003904
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.38014     Inf
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

The OJ supplement is better at this dose, because the p-value is lower than alpha and we reject the null hypothesis.

Dose: 2.0

- Ho: Mean is equal to 0 (No difference in tooth growth between OJ and VC supplement, equally effective, adjusting for dose 2.0)
- Ha: difference in means is not equal to 0 (One supplement is more effective than the other at 2.0 dose)

```
dose20VC <- ToothGrowth %>% filter(dose == 2.0 & supp == "VC")
dose20OJ <- ToothGrowth %>% filter(dose == 2.0 & supp == "OJ")
t.test(dose20VC$len, dose20OJ$len,  paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  dose20VC$len and dose20OJ$len
## t = 0.046136, df = 18, p-value = 0.9637
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.562999  3.722999
## sample estimates:
## mean of x mean of y
##     26.14     26.06
```

This time we can't reject the null hypothesis given the p-value is greater than alpha. So at this dossage (2.0) no supplement is more effective at tooth growth.