Statistical Inference Project - Part 2: Basic Inferential Data Analysis

Saeed Nusri

July 30, 2016

2. Data Analysis

In this section the ToothGrowth data is used for exploratory data analysis.

Loading data

The following codes reveal more information about the dataset.

```
#Loading the dataset
data(ToothGrowth)
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
```

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

```
unique(ToothGrowth$supp)
```

```
## [1] VC OJ
## Levels: OJ VC
```

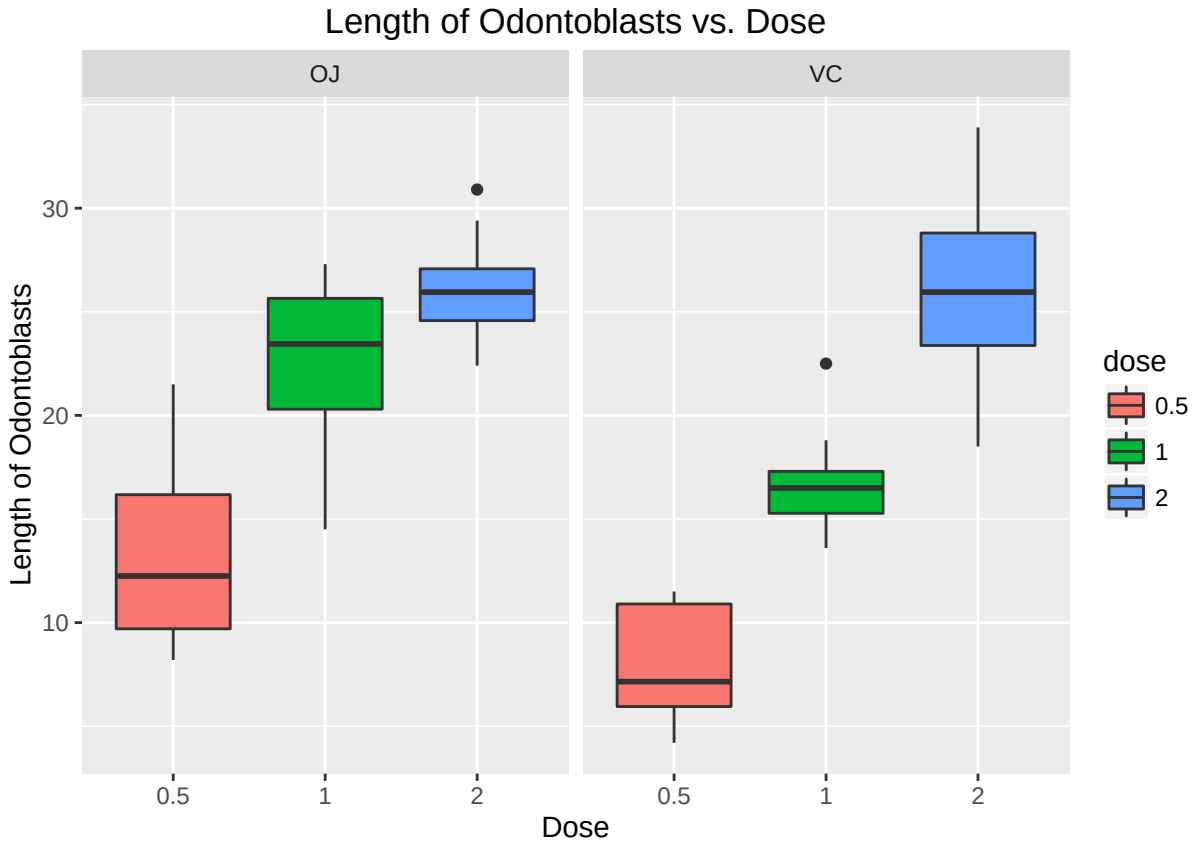
The data compreosis of 60 observations, length of odontoblasts in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1 and 2 mg) with each of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC).

Data Analysis

```
#Changing the dose variable into a factor class
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

#Loading ggplot2 package for plotting box plots of the datasets to see trnds
library(ggplot2)
g <- ggplot(data=ToothGrowth, aes(x=dose, y=len)) +
        geom_boxplot(aes(fill=dose)) +facet_grid(~supp) +
        xlab("Dose") + ylab("Length of Odontoblasts") +
        ggtitle("Length of Odontoblasts vs. Dose")

print(g)
```



It can be concluded that there is a clear positive correlation between the tooth length and the dose levels of Vitamin C, for both delivery methods.

One interesting question that can also be addressed is whether the supplement type (i.e. orange juice or ascorbic acid) has any effect on the tooth length.

```
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
```

```
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

The p-value is 0.06 and the confidence interval contains zero. Hence, we can not reject the null hypothesis, which states that the different supplement types have no effect on tooth length.

```
#Loading dplyr for using the 'filter' function

library(dplyr)
```

```
#Segragating the ToothGrowth data with the different dosage of OJ and VC
dose1 <- filter(ToothGrowth, dose == 0.5)
dose2 <- filter(ToothGrowth, dose == 1.0)
dose3 <- filter(ToothGrowth, dose == 2.0)


#Carrying out t tes
t.test(len ~ supp, paired = F, var.equal = F, data = dose1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len ~ supp, paired = F, var.equal = F, data = dose2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```r
t.test(len ~ supp, paired = F, var.equal = F, data = dose3)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

Since the 95% confidence interval (CI) of the first two tests, of doses 1.0 mg and 1.5 mg, do not include 0, it tells us there is a statistically significant difference between the supplementing orange juice and ascorbic acid. Moreover, a low p-value allows the rejection of the null hypothesis. However, the impact diminishes when the dosage is above certain level, in our case, 2 mg.

Conclusion

Based on the analysis under the assumption that the supplements, orange juice and ascorbic acid, were independently and identically distributed among the subjects, we can conclude that the orange juice, when adminstered with moderate dosage would have a significant impact on the ondontoblasts on Guinea pigs.