Basic Inferential Data Analysis Using R
Statistical Inference Course Project

Mouhammed Samer Houssien Hasn

July 21, 2019

Abstract. This project shows a Basic Inferential Data Analysis Using basic analysis tools in R. The analysis based on 60 observations of tooth growth of Guinea pigs, which is ToothGrowth data set, aims to provide simple descriptive statistics of data, and test the hypothesis of the effectiveness of two delivery methods ("OJ" (Orange Juice) and VC (Ascorbi Acid, a form of Vitamin C)), each of them contains three different value of doses (0.5 ,1 , and 2 mg/day).

1. Descriptive analysis

ToothGrowth contains 3 columns for tooth length(len), supplement method (supp), and value of doses. As the following simple view:

```
data("ToothGrowth") #Load the data
head(ToothGrowth) #short view of data
```

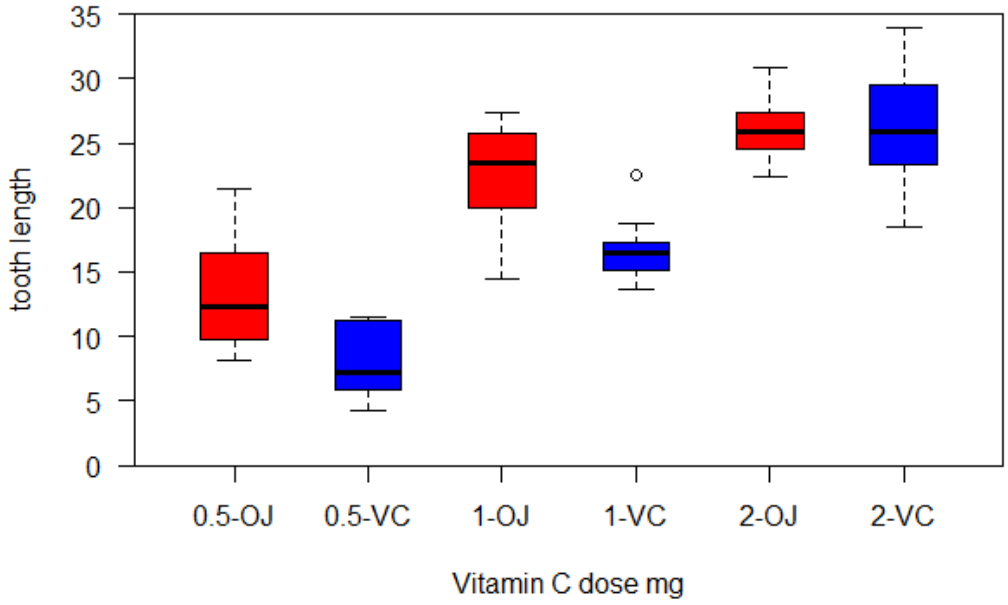
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

Then we make the simple descriptive analysis with a boxplot:

```
summary(ToothGrowth) #descrition of data
```

```
##       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
```

Fig.1
Tooth Growth boxplot



2. Hypothesis test

We want to test which supplement method is more effective, and how much the dose should be, or is there a significant difference between supplement methods, or the quantity of doses. We make three t-tests for difference in means, with a 95% confidence level.

2.1 Difference test for supplement methods

Null hypothesis (H_0): there is no significant difference in the effect of each supplement method. Then we run the test as follow:

```
#hypothesis test 1
attach(ToothGrowth)
t1<-t.test(len[supp=="VC"],len[supp=="OJ"],paired = FALSE)
t1
##  Welch Two Sample t-test
## data:  len[supp == "VC"] and len[supp == "OJ"]
## 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:
##  -7.5710156  0.1710156
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

We can not reject the null hypothesis as the p-value above the significant level of 5% (or the test value is between the Three-sigma limits (-1.9153), then we conclude that the supplement method makes no significant difference on tooth length.

2.2 Difference test for doses of 0.5 and 1 mg/day

Null hypothesis (H_0): there is no significant difference in the effect of 0.5 and 1 mg/day doses. Then we run the test as follow:

```
t2<-t.test(len[dose==0.5],len[dose==1],paired = FALSE)

t2
##  Welch Two Sample t-test
##
## data:  len[dose == 0.5] and len[dose == 1]
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

We reject the null hypothesis as the p-value under the significant level of 5% (or the test value is outside the Three-sigma limits (-6.4766), then we conclude that the effect of doses of 0.5 mg/day differs significantly from 1 mg/day doses effect.

2.3 Difference test for doses of 0.5 and 2 mg/day

Null hypothesis (H_0): there is no significant difference in the effect of 0.5 and 2 mg/day doses. Then we run the test as follow:

```
t3<-t.test(len[dose==0.5],len[dose==2],paired = FALSE)
t3
##  Welch Two Sample t-test
##
## data:  len[dose == 0.5] and len[dose == 2]
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

We reject the null hypothesis as the p-value under the significant level of 5% (or the test value is outside the Three-sigma limits (-11.799), then we conclude that the effect of doses of 0.5 mg/day differs significantly from 2 mg/day doses effect.