Basic Differential Data Analysis

Sneh Bindesh Chitalia

11/06/2020

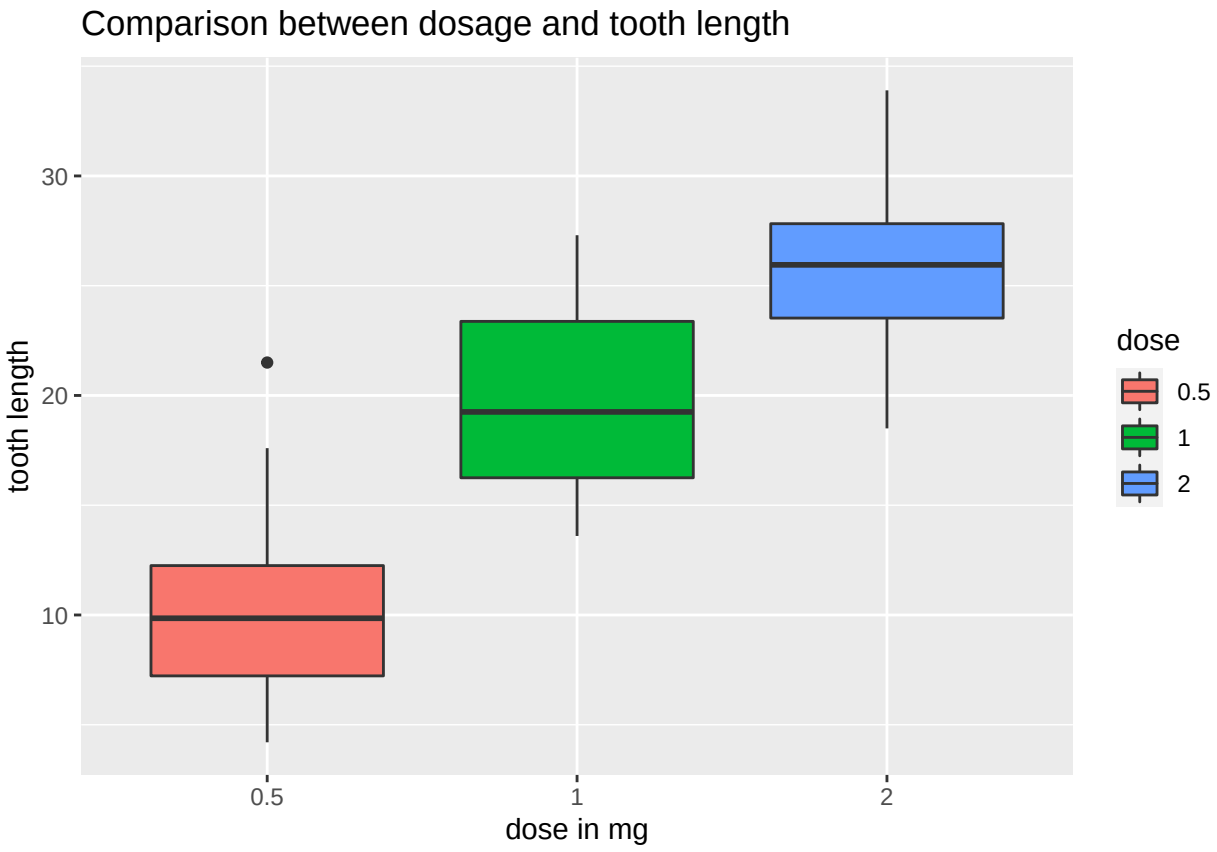
Load the dataset

```
library(datasets)
data("ToothGrowth")
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
```

Basic Exploratory Data Analysis

```
library(ggplot2)
#Plot 1 (dose vs len)
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
g<-ggplot(ToothGrowth,aes(x=dose,y=len,fill=dose))+geom_boxplot()+ggtitle("Comparison between dosage an
print(g)
```




```
#Plot 2 (supp vs len)
k<-ggplot(ToothGrowth,aes(x=supp,y=len,fill=supp))+geom_boxplot()+ggtitle("Comparison of delivery metho
print(k)
```

Summary

Check for any NA or Null or any type of missing values

```
sum(!complete.cases(ToothGrowth))
```

```
## [1] 0
```

Thus, there are no missing values in the data.
Further exploring the data

```
summary(ToothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

This gives us an overview of the data

```r
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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

This gives us information about data type of each variable.

Hypothesis Testing

We carry out hypothesis testing on whether delivery mode has any impact on teeth length.
The null hypothesis is:
H0: The delivery mode for Vitamin C does not have any impact on tooth length

```r
dosage<-ToothGrowth$dose
supp<-ToothGrowth$supp
len<-ToothGrowth$len
t.test(len[supp=="OJ"],len[supp=="VC"],paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len[supp == "OJ"] and len[supp == "VC"]
## 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 of x mean of y
##  20.66333  16.96333
```

Since the p-value > 0.05, we fail to reject the null hypothesis.
However, the confidence interval contains 0 which indicates that the test is not significant.

H0: The dosage in mg does not have any impact on tooth length

```r
# compare between dosage of 0.5 mg and 1 mg
t.test(len[dosage==0.5],len[dosage==1],paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len[dosage == 0.5] and len[dosage == 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
```

Since the p-value < 0.05, we can reject the null hypothesis.
Also, the confidence interval does not contain zero which indicates that the test is significant.

```
# compare between dosage of 1 mg and 2 mg
t.test(len[dosage==1],len[dosage==2],paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len[dosage == 1] and len[dosage == 2]
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

Since the p-value < 0.05, we can reject the null hypothesis.
Also, the confidence interval does not contain zero which indicates that the test is significant.
Thus we can succesfully rule out null hypothesis.

Conclusions

1. The delivery method of Vitamin C does not impact the teeth growth.
2. The amount of dosage given to the guinea pigs impact the teeth growth.

Assumptions

1. t-tests are considered.
2. These t-tests are considered to be unpaired.