Statistical Inference Project2 - Basic Inferential Data Analysis.

Sanat Das

6/23/2017

Overview

We need to analyze the ToothGrowth data in the R datasets package. The steps to be followed are as below:

1. Load the ToothGrowth data and perform some basic exploratory data analysis
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

1. Data Loading and basic exploratory data analysis

Loading required libraries:

```
library(ggplot2)
library(datasets)
```

Loading data:

```
data(ToothGrowth)
```

Basic exploratory data analyses:

```
nrow(ToothGrowth)
```

```
## [1] 60
```

```
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
```

```
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 ...
```

2. Basic Summary of the Data

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
summary(ToothGrowth)
```

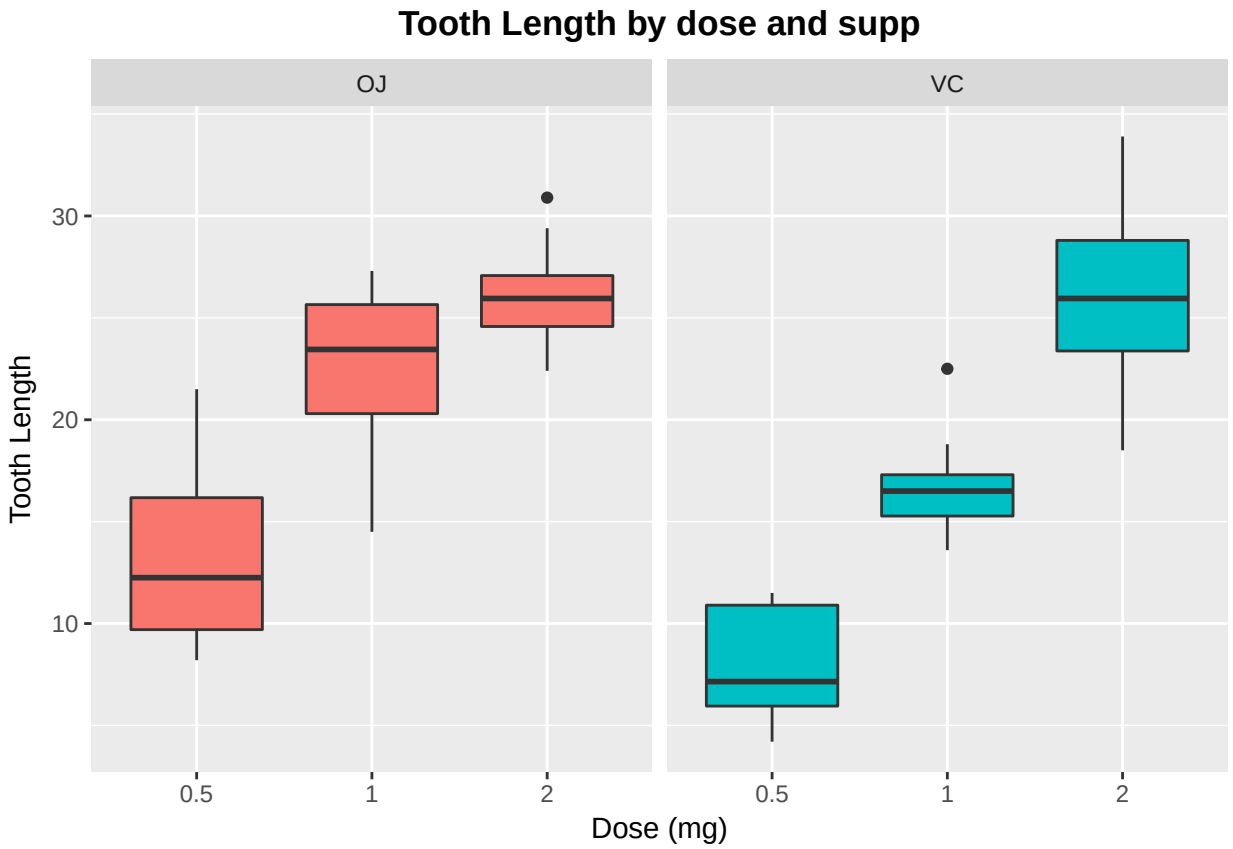
```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
table(ToothGrowth$supp,ToothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

Tooth growth by dose and supplement type:

```
ggplot(data=ToothGrowth,aes(x=factor(dose),y=len))+facet_grid(.~supp) + geom_boxplot(aes(fill=supp),sho
```



3. Use confidence intervals and/or hypothesis tests to compare Tooth growth by supp and dose

As seen above, the dataset has two levels of supp: OJ and VC and three levels of dose: 0.5, 1, 2.

Testing for Supp Types

Assuming the null hypothsis, H0: different supp types do not have any effect on tooth length.

```
t.test(len ~ supp, 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 confidence interval contains zero, and the p-value(0.06) is greater than the significance level 0.05. Hence, we can not reject the null hypothesis H0, which indicates that the supp types do not have any effect on teeth length.

Testing for Dose Level

Assuming the null hypothsis, H0: The dose level does not have any impact on the tooth length.

Create three sub-groups for dose levels:

```
ToothGrowth.dose_type1 <- subset(ToothGrowth, dose %in% c(0.5,1.0))
t.test(len ~ dose, data = ToothGrowth.dose_type1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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 in group 0.5   mean in group 1
##            10.605            19.735
```

```
ToothGrowth.dose_type2 <- subset(ToothGrowth, dose %in% c(0.5,2.0))
t.test(len ~ dose, data = ToothGrowth.dose_type2)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  len by dose
## 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 in group 0.5   mean in group 2
##            10.605            26.100
```

```
ToothGrowth.dose_type3 <- subset(ToothGrowth, dose %in% c(1.0,2.0))
t.test(len ~ dose, data = ToothGrowth.dose_type3)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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 in group 1 mean in group 2
##          19.735          26.100
```

In all the above three cases, the p-value is less than the significance level(0.05), and the confidence interval does not contain zero. Hence, the null hypothesis H0 can be rejected. As there is a positive relationship between dose level and teeth length, the increase in dose level seems to increase in tooth growth.

4. Assumptions and Conclusions

While analysing the data, the following assumptions were made:

1. The 60 guinea pigs taken are representative for the entire population of the guinea pigs.
2. The supplement and dosage were randomly assigned.
3. The distribution of the means is normal.

Based on the data analysis with above assumptions, we can conclude the following:

1. There is no strong evidence that Supp types have any effect on tooth growth.
2. There is strong evidence that increase in does level leads to increase in tooth growth.