Basic Inferential Data Analysis

Allan Kavuma

23 May 2019

Overview

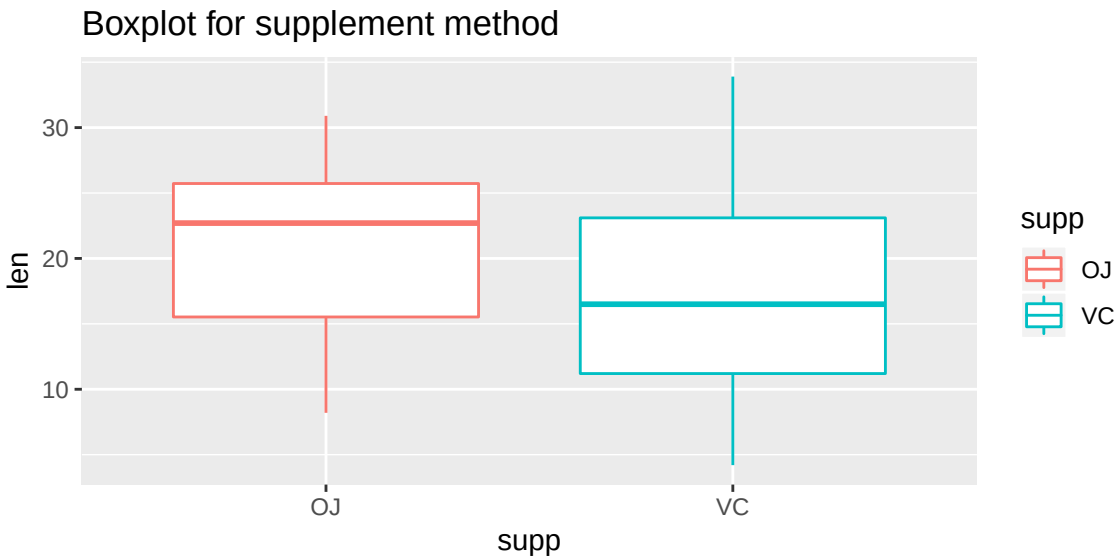
This project involves using Basic Inferential Data Analysis to analyze the ToothGrowth data in the R datasets package.
The steps below will be followed in this process:
1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence 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. Load ToothGrowth data and perform some basic exploratory data analyses

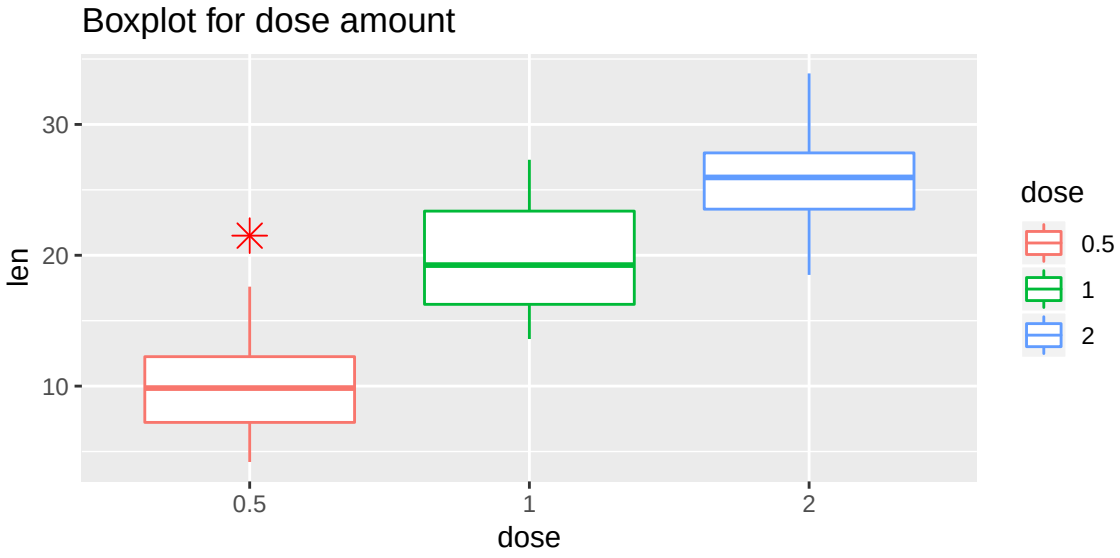
```
## Import libraries for the exploratory analysis
library(ggplot2)
## Load the ToothGrowth data and store
data("ToothGrowth")
## Show summary of the data
knitr::kable(summary(ToothGrowth))
```

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

```
## Subset the Orange Juice data
ojData <- ToothGrowth[ToothGrowth$supp == "OJ",]
## Subset the ascobic acid data
vcData <- ToothGrowth[ToothGrowth$supp == "VC",]
##Convert dose to factor for proper plotting
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
##Convert supp to factor for proper plotting
ToothGrowth$supp <- as.factor(ToothGrowth$supp)
## Boxplot for tooth length as per supplement type
ggplot(data = ToothGrowth, aes(x = supp, y = len)) +
        geom_boxplot(aes(colour = supp)) +
        labs(title = "Boxplot for supplement method")
```




```
## Boxplot for tooth length as per dose quantity
ggplot(data = ToothGrowth, aes(x = dose, y = len)) +
        geom_boxplot(aes(colour = dose), outlier.colour="red",outlier.shape=8,
                outlier.size=4) +
        labs(title = "Boxplot for dose amount")
```



2. Basic summary of the data.

- The more Vitamin C dose levels the Guinea pigs received, the longer their teeth grew (graph 2).
- Guinea pigs that received Vitamin C thorough orange juice method had longer teeth than those that received ascorbic acid (graph 1).
- There are more extremes/variability (i.e variance) in teeth length for guinea pigs that received ascorbic acid than those that received orange juice (graph 1).

3. Confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

```
## Create tooth length vectors for orange juice and ascorbic acid
ojlen <- ToothGrowth[ToothGrowth$supp == "OJ",]$len
vclen <- ToothGrowth[ToothGrowth$supp == "VC",]$len
## Run the t test to compare tooth growth by supplement method
t.test(len ~ supp, paired = FALSE, var.equal = TRUE, data = ToothGrowth)
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##         20.66333         16.96333
```

```
## Run t test to compare tooth growth by dose amount
halfDose <- ToothGrowth[ToothGrowth$dose == 0.5,]$len ## half dose tooth len
oneDose <- ToothGrowth[ToothGrowth$dose == 1,]$len ## one dose tooth length
twoDose <- ToothGrowth[ToothGrowth$dose == 2,]$len ## two dose tooth length
## comparing tooth growth of half dose and one dose
t.test(oneDose, halfDose, paired = FALSE, var.equal = TRUE)$p.value
```

```
## [1] 1.266297e-07
```

```
t.test(oneDose, halfDose, paired = FALSE, var.equal = TRUE)$conf
```

```
## [1]  6.276252 11.983748
## attr(,"conf.level")
## [1] 0.95
```

```
## Comparing tooth growth of half dose and two dose
t.test(twoDose, halfDose, paired = FALSE, var.equal = TRUE)$p.value
```

```
## [1] 2.837553e-14
```

```
t.test(twoDose, halfDose, paired = FALSE, var.equal = TRUE)$conf
```

```
## [1] 12.83648 18.15352
## attr(,"conf.level")
## [1] 0.95
```

```
## Comparing tooth growth of one dose and two dose
t.test(twoDose, oneDose, paired = FALSE, var.equal = TRUE)$p.value
```

```
## [1] 1.810829e-05
```

```
t.test(twoDose, oneDose, paired = FALSE, var.equal = TRUE)$conf
```

```
## [1] 3.735613 8.994387
## attr(,"conf.level")
## [1] 0.95
```

4. Conclusions and the assumptions

Assumptions

- Samples are randomly selected.
- Population distribution is approximatelty normally distributed.
- Null Hypothesis 1: There is no difference in tooth growth between the two supplement delivery methods, orange juice and ascorbic acid.
- Null Hypothesis 2: There is no difference in tooth growth between the 3 dose amounts, 0.5, 1 and 2.

Conclusions

- Comparing the supplement method of orange juice and ascorbic acid, the p-value for the t-test is 0.0603934, which is greater than 0.05 and the confidence interval includes a zero; hence we fail to reject the null hypothesis that the means of the two methods are equal(no difference between means). With a 95% confidence level, there is no difference in tooth growth by delivery method.
- Also, comparing the doses; the p values are very small and confidence intervals per comparison all positive; hence with a 95% confidence level, there is statistical evidence that one dose results in longer tooth length than half a dose, and two doses result in longer tooth length than one dose.