

Statistical Inference Course Project - Part 2

Orumwese Esosa Cyriaque

8/2/2021

Exploratory Analysis and Statistical Inference on the ToothGrowth Data in R Datasets Package

Synopsis

This report aims to analyze the ToothGrowth data in the R datasets package. The aim is compare tooth growth by supplement delivery method and by dose.

Loading the Data

ToothGrowth data is a dataset which measures the effect of Vitamin C on tooth growth in Guinea pigs. It is a data frame with 60 observations on 3 variables;

- len: This is the response and it is the length of the cells responsible for tooth growth in the guinea pigs.
- supp: The two delivery methods with which the doses were administered. By orange juice (OJ) or by ascorbic acid (VC), a form of vitamin C.
- dose: Each animal received one of 3 dose levels of vitamin C (0.5, 1 and 2 mg/day).

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

Checking for NA values

```
sum(is.na(ToothGrowth))
```

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

Basic Summary Tests

Performing basic summary tests

```
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
```

```
# First 6 rows of the data
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
```

Unique Values of Each Variable

```
# Getting the unique values of each variable
unique(ToothGrowth$len)
```

```
## [1]  4.2 11.5  7.3  5.8  6.4 10.0 11.2  5.2  7.0 16.5 15.2 17.3 22.5 13.6 14.5
## [16] 18.8 15.5 23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5 23.3 29.5 17.6  9.7  8.2
## [31]  9.4 19.7 20.0 25.2 25.8 21.2 27.3 22.4 24.5 24.8 30.9 29.4 23.0
```

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

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

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

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

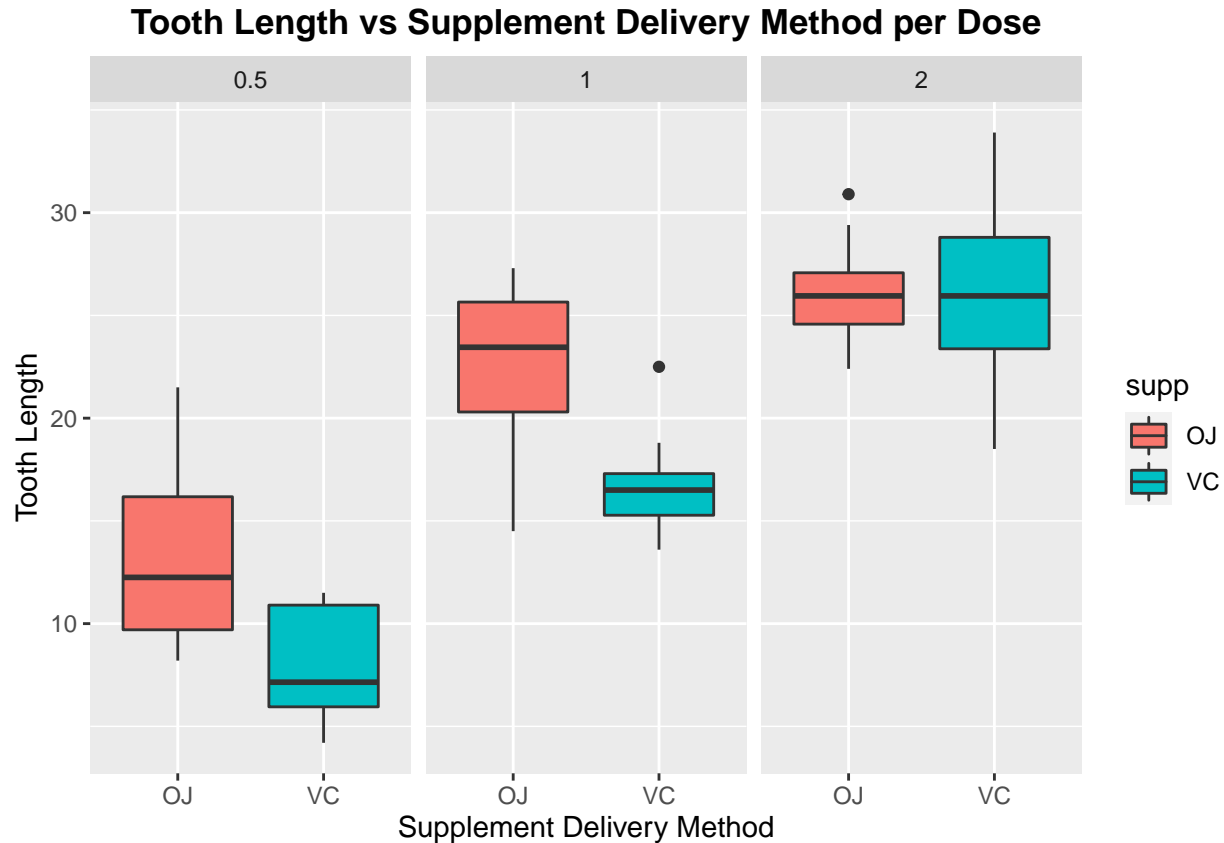
Convert dose column to factor class

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

Exploratory Data Analysis

The relationship between the supplement (supp variable) and the tooth length (len) for each dose. This was visualized using the boxplot.

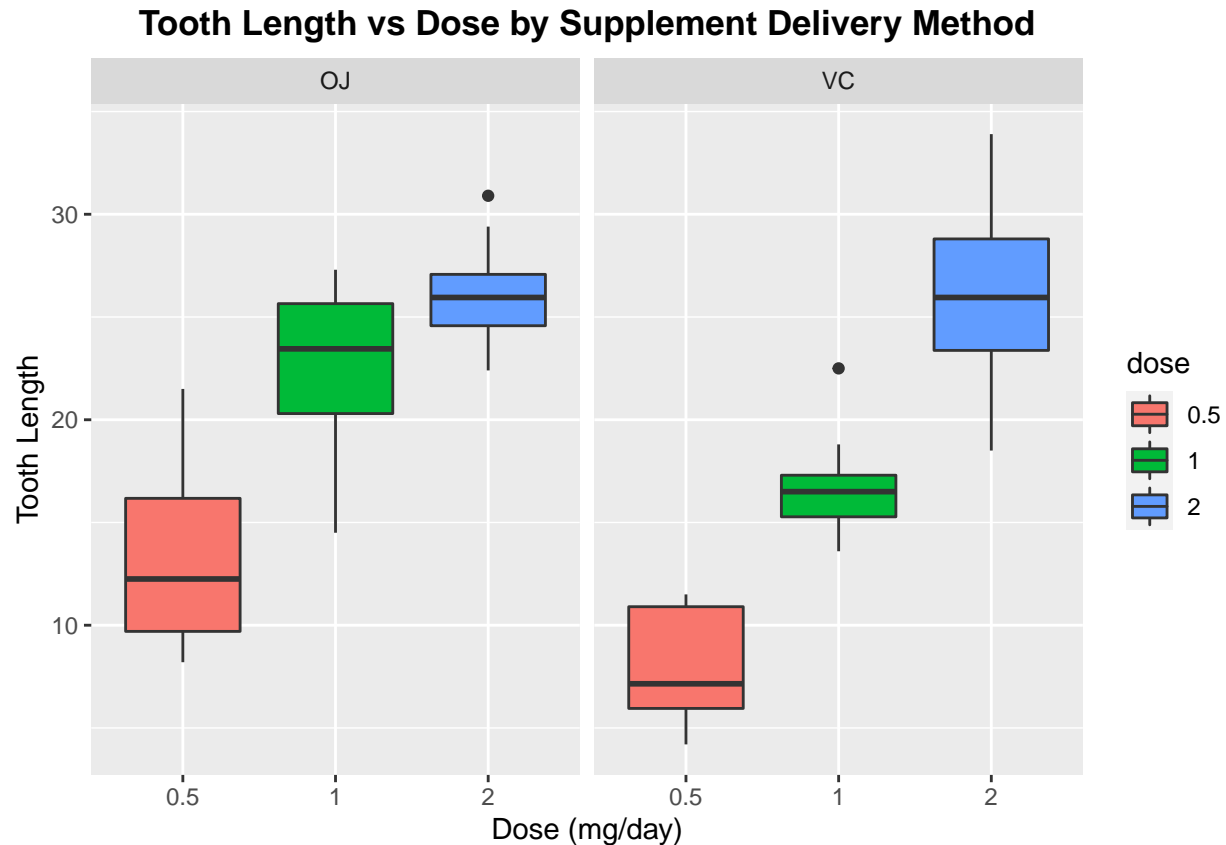
```
ggplot(ToothGrowth, aes(x=supp, y=len)) +
  geom_boxplot(aes(fill=supp)) + facet_grid(~dose) +
  labs(x="Supplement Delivery Method", y="Tooth Length", title="Tooth Length vs Supplement Delivery Method", title="Tooth Length vs Supplement Delivery Method", title="Tooth Length vs Supplement Delivery Method")
  theme(plot.title=element_text(hjust=0.5, face="bold"))
```



We can see that the tooth length increased generally with increase in dose. The Oranje Juice (OJ) supplement performed better than the Ascorbic Acid (VC) generally but for the 2mg dose, the VC supplement showed a wider range of lengths and also similar median with the OJ supplement.

The relationship between tooth length and dosage per supplement delivery. This was visualized using boxplot.

```
ggplot(ToothGrowth) +
  geom_boxplot(aes(x=dose, y=len, fill=dose)) + facet_grid(~supp) +
  labs(x="Dose (mg/day)", y="Tooth Length", title="Tooth Length vs Dose by Supplement Delivery Method", title="Tooth Length vs Dose by Supplement Delivery Method", title="Tooth Length vs Dose by Supplement Delivery Method")
  theme(plot.title=element_text(hjust=0.5, face="bold"))
```



An increase in tooth length can be seen with increase in dose. The OJ supplement performed better than the VC supplement but for the 2mg dose where the VC supp showed a wider range of tooth growth.

Tooth Growth Comparisons

Hypothesis tests and confidence intervals are used to compare tooth growth by supplement and dose.

By Supplement

Hypothesis:

Suppliment delivery methods have no effect on tooth growth.

- H_0 : Both group have the same mean.
- H_A : Means are different.

T-test

```
supp.test <- t.test(data=ToothGrowth, len~supp)
supp.test
```

```
##
## 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 test returns a p-value of 0.06. Therefore since the p-value is greater than 0.05 and the confidence interval of the test contains zero, we can say that supplement delivery method seems to have no impact on Tooth growth based on this test. We fail to reject the null hypothesis.

By Dose

Hypothesis:

Higher doses of Vitamin C cause less tooth growth.

- H_0 : Mean of larger dose value is smaller than or equal to smaller dose value.
- H_A : Mean of larger dose value is greater than smaller dose value.

T-test Running t-tests on different pairs of dose values

```
# Using dose values 0.5 and 1.0
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,1.0))
dose.test <- t.test(data=ToothGrowth_sub, len~dose)
dose.test
```

```
##
## 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
```

```
# Using dose values 0.5 and 2.0
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,2.0))
dose.test <- t.test(data=ToothGrowth_sub, len~dose)
dose.test
```

```
##
## 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

# Using dose values 1.0 and 2.0
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,2.0))
dose.test <- t.test(data=ToothGrowth_sub, len~dose)
dose.test

##
## 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
```

For all tests, the p-value is zero and the confidence interval does not contain zero. Therefore since the p-value is less than the significance level of 0.5 and the confidence interval does not contain 0, we reject the null hypothesis. This means that higher doses of Vitamin C result in greater tooth growth

Assumption and Conclusion

Assuming that:

1. The sample is representative of the population;
2. The distribution of the sample means follows the Central Limit Theorem.

In reviewing our t-test analysis from above we can conclude that:

1. Supplement delivery method has no effect on tooth growth and,
2. Increase in dosage of Vitamin C results in increased tooth growth.