

Statistical Inference Course Project: Inferential Data Analysis

anckur khaitan

Inferential Data Analysis ToothGrowth

This project analyses the ToothGrowth dataset available in R and tries to answer questions about whether there is a relationship between tooth length growth in guinea pigs and the dose and supplement given to them.

Data Summary

```
library(datasets)
library(ggplot2)

# Load the dataset
data = ToothGrowth

# Look at data summary
summary(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
```

```
# Check for any missing values
sum(complete.cases(data)) == nrow(data)
```

```
## [1] TRUE
```

```
# How are doses and supplement divided
table(data$dose, data$supp)
```

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

```
# Check number of unique dose values
unique(data$dose)
```

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

```
# Convert the dose to factor since only three different levels exist
data$dose = as.factor(data$dose)
```

The data frame has 60 observations, each with 3 variables. The variables are,

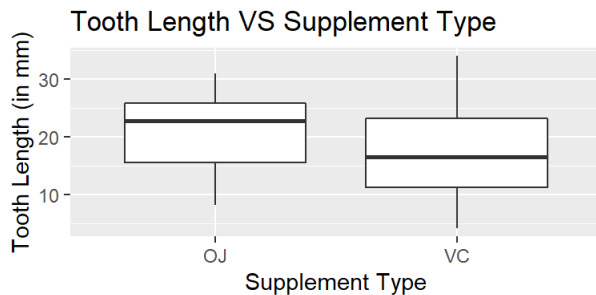
- len: numeric Tooth length
- supp factor: Supplement type (VC or OJ)
- dose: numeric Dose in milligrams/day

No missing values are present in the data frame. Also, there are 10 samples for each supplement and dose combination. The dose amount has only 3 unique values 0.5, 1.0 and 2.0 mg/day and so that variable has been converted to a factor variable.

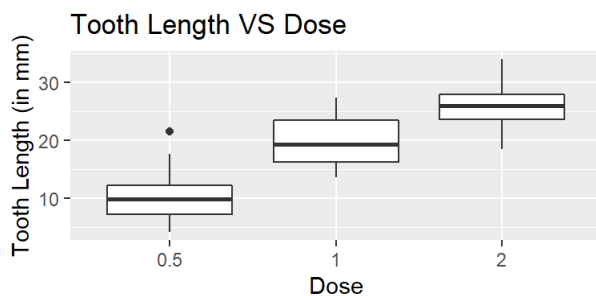
Exploring The Data

Let us see visually using boxplots how the tooth length varies with supplement and dose types.

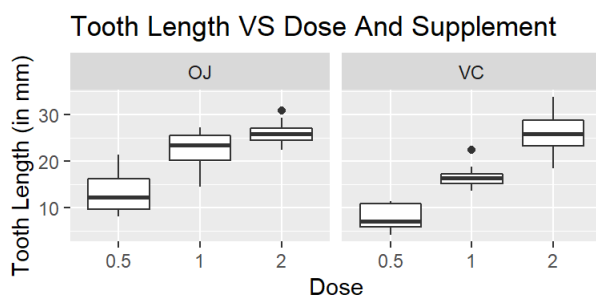
```
# Tooth Length Variation with supplement
p = ggplot(data, aes(supp, len))
p = p + geom_boxplot() + xlab('Supplement Type') + ylab('Tooth Length (in mm)') +
  labs(title='Tooth Length VS Supplement Type')
print(p)
```



```
# Tooth Length Variation with doses
p = ggplot(data, aes(dose, len))
p = p + geom_boxplot() + xlab('Dose') + ylab('Tooth Length (in mm)') +
  labs(title='Tooth Length VS Dose')
print(p)
```



```
# Tooth Length Variation with doses
p = ggplot(data, aes(dose, len))
p = p + geom_boxplot() + xlab('Dose') + ylab('Tooth Length (in mm)') +
  labs(title='Tooth Length VS Dose And Supplement') + facet_wrap(~supp)
print(p)
```



Tooth length seems to be higher overall for Orange Juice as compared to Vitamin C, but there is not a major difference. Furthermore, for both types of supplements, it seems that tooth length increases with the dose amount. Also, for a dose of 0.5 and 1.0 mg/day, tooth length seems to be less for Vitamin C than Orange Juice, but for a higher dose of 2.0 mg/day there doesn't seem to be much difference. We will test these claims using t-tests.

Hypothesis Testing

Hypothesis 1

Does Tooth Length vary with dose amount for Orange Juice?

```
# Does Tooth Length vary with Dose for Orange Juice?
df = data[data$supp=="OJ" & data$dose %in% c(0.5,1.0),]
t.test(len ~ dose, data = df, var.equal=FALSE, paired = FALSE, alternative="greater")
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -5.0486, df = 17.698, p-value = 1
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -12.72568 Inf
## sample estimates:
## mean in group 0.5 mean in group 1
## 13.23 22.70
```

```
df = data[data$supp=="OJ" & data$dose %in% c(1.0,2.0),]
t.test(len ~ dose, data = df, var.equal=FALSE, paired = FALSE, alternative="greater")
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -2.2478, df = 15.842, p-value = 0.9804
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -5.971376 Inf
## sample estimates:
## mean in group 1 mean in group 2
## 22.70 26.06
```

We see that the p-value is greater than 0.05 in both cases. The null hypothesis that for Orange Juice, Tooth Length is less for 0.5 mg/day dose than that for 1.0 mg/day and growth for 1.0 mg/day is less than that of 2.0 mg/day, cannot be rejected.

Hypothesis 2

Does Tooth Length vary with dose amount for Vitamin C?

```
# Does Tooth Length vary with Dose for Vitamin C?
df = data[data$supp=="VC" & data$dose %in% c(0.5,1.0),]
t.test(len ~ dose, data = df, var.equal=FALSE, paired = FALSE, alternative="greater")
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -7.4634, df = 17.862, p-value = 1
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -10.83313 Inf
## sample estimates:
## mean in group 0.5 mean in group 1
## 7.98 16.77
```

```
df = data[data$supp=="VC" & data$dose %in% c(1.0,2.0),]
t.test(len ~ dose, data = df, var.equal=FALSE, paired = FALSE, alternative="greater")
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -5.4698, df = 13.6, p-value = 1
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -12.39348 Inf
## sample estimates:
## mean in group 1 mean in group 2
## 16.77 26.14
```

p-value is greater than 0.05 in both cases. The null hypothesis that for Vitamin-C, Tooth Length is less for 0.5 mg/day dose than that for 1.0 mg/day and growth for 1.0 mg/day is less than that of 2.0 mg/day, cannot be rejected.

Hypothesis 3

Does Tooth Length vary with just Supplement Type?

```
# Does Tooth Length vary with just supplement type?
df = data
t.test(len ~ supp, data = df, var.equal=FALSE, paired = FALSE)
```

```
##
##  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
```

p-value is greater than 0.05 and so the null hypothesis that the mean tooth length is the same cannot be rejected.

Hypothesis 4

Does Tooth Length vary with supplement type for particular dose amounts?

```
# Does Tooth Length vary with supplement type for Dose of 0.5
df = data[data$dose %in% c(0.5),]
t.test(len ~ supp, data = df, var.equal=FALSE, paired = FALSE, alternative="less")
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.9968
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##  -Inf 8.15396
## sample estimates:
## mean in group OJ mean in group VC
##      13.23      7.98
```

```
# Does Tooth Length vary with supplement type for Dose of 1.0
df = data[data$dose %in% c(1.0),]
t.test(len ~ supp, data = df, var.equal=FALSE, paired = FALSE, alternative="less")
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.9995
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##  -Inf 8.503842
## sample estimates:
## mean in group OJ mean in group VC
##      22.70      16.77
```

```
# Does Tooth Length vary with supplement type for Dose of 2.0
df = data[data$dose %in% c(2.0),]
t.test(len ~ supp, data = df, var.equal=FALSE, paired = FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.79807 3.63807
## sample estimates:
## mean in group OJ mean in group VC
## 26.06 26.14
```

p-values for all the tests are greater than 0.50. The null hypothesis that the Tooth Length is greater in the case of Orange Juice for a dose of 0.5 and 1.0 mg/day cannot be rejected. Similarly, the null hypothesis that mean Tooth Length is equal for Orange Juice and Vitaminc C for a dose of 2.0 mg/day cannot be rejected.

Assumptions And Conclusions

We have used the t-test for hypothesis testing we have assumed that the data comes from a normal distribution and also that the data is a random sample from the actual population. For each supplement individually, increase in dosage increases the tooth length. Overall, there is no statistical significant difference between orange juice and vitamin C. However, for 0.5 mg/day and 1.0 mg/day dose orange juice gives longer tooth length but for 2.0 mg/day there is no statistically significant difference.