StatisticalInferenceProject - Part 2

Van Wyk Burnett

03 April 2020

Part 2: Basic Inferential Data Analysis

Data Processing

In this section I will load the data and perform some basic exploratory data analysis.

```
#load data
library(datasets)
library(ggplot2)
data("ToothGrowth")
#look at data structure
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 ...
```

```
#look at head 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
```

Investigate if column "dose" should rather be a factor variable.

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

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

Convert column "dose" to factor variable.

```
#convert dose to factor
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
#look at new data structure
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 ...
```

Summary of data

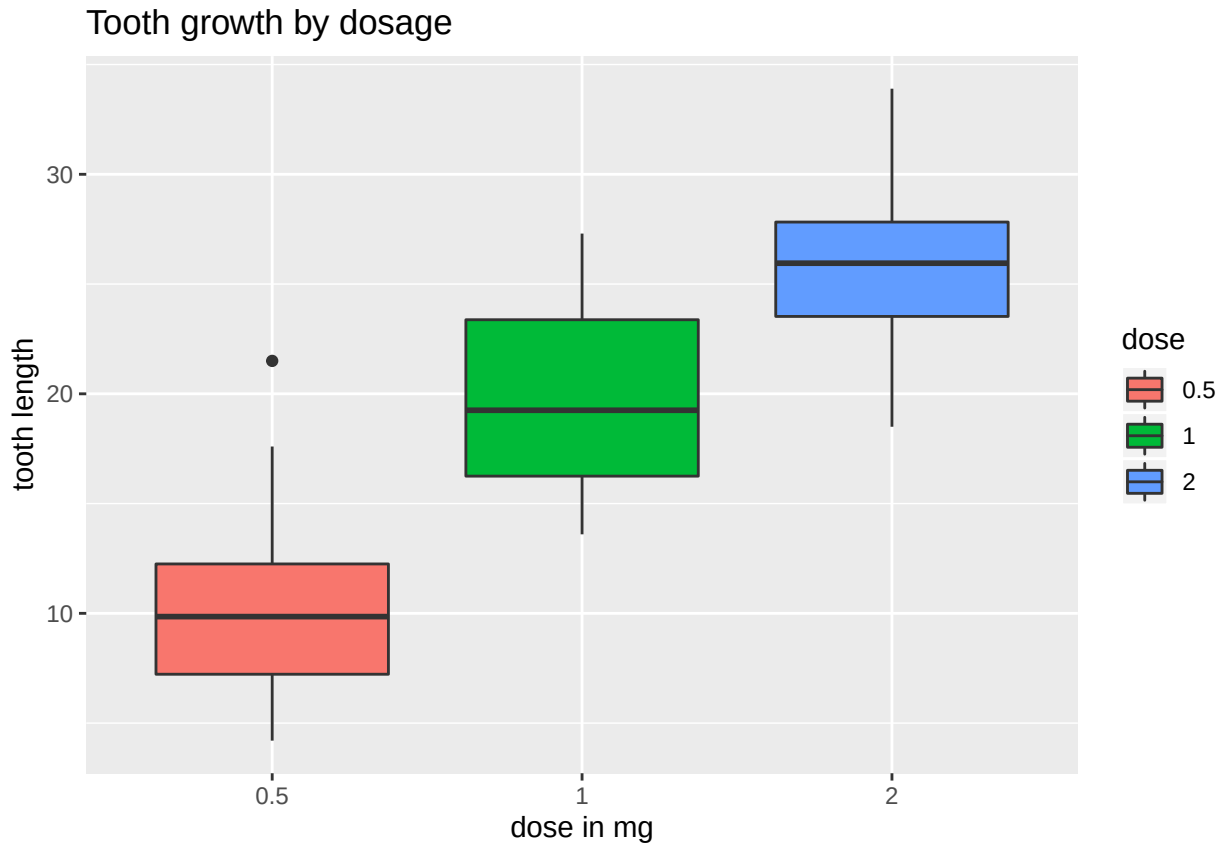
Basic summary

```
summary(ToothGrowth)
```

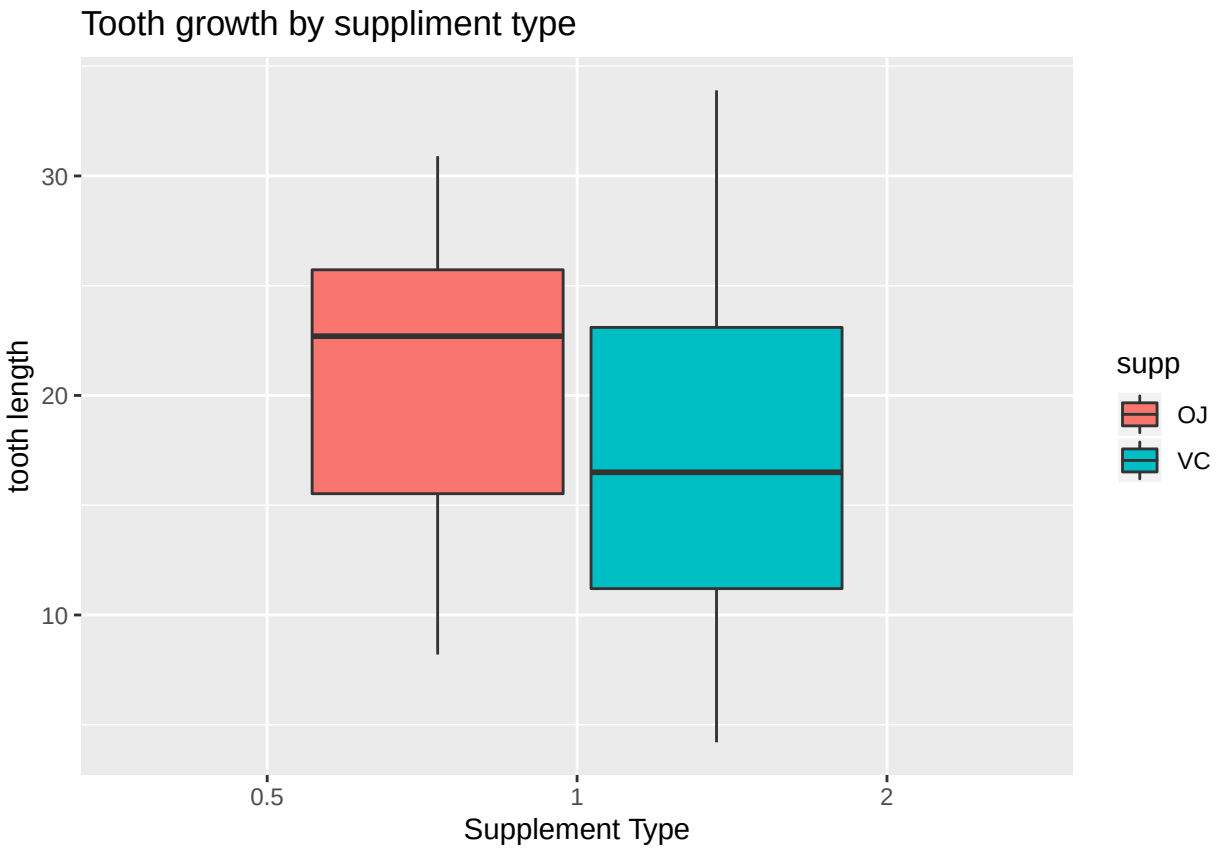
```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

Box plot to visualise the data summary.

```
p <- ggplot(data = ToothGrowth, aes(x=dose, y=len, group=dose)) + geom_boxplot(aes(fill=dose)) +ggtitle
p <- p + xlab("dose in mg") + ylab("tooth length")
p
```




```
p <- ggplot(data = ToothGrowth, aes(x=dose, y=len, group=supp)) + geom_boxplot(aes(fill=supp)) +ggtitle
p <- p + xlab("Supplement Type") + ylab("tooth length")
p
```



Hypothesis Test

This test will determine if the delivery mode has an influence on the tooth growth.

The null hypothesis: H0: The delivery mode of Vitamin C does not have any influence on the tooth growth

```
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 p-value of 0.06 at a significance level of 0.05 (alpha = 0.05) indicates that we do not reject the null hypothesis. The confidence interval include 0, so the test is not significant.

The next test will test the influence of the dose on the tooth length. There are three dose levels, thus subgroups will be neccesary to perform the t.test (can only use two levels per test).

Creating subgroups

```r
group1 <- subset(ToothGrowth, dose %in% c(0.5,1.0))
group2 <- subset(ToothGrowth, dose %in% c(0.5,2.0))
group3 <- subset(ToothGrowth, dose %in% c(1.0,2.0))
```

```r
t.test(len ~ dose, data = group1)
```

```
##
##  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
```

```r
t.test(len ~ dose, data = group2)
```

```
##
##  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
```

```r
t.test(len ~ dose, data = group3)
```

```
##
##  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 three subgroups the t.test results in a p-value of less that 0.05 and for none of the tests does the confidence interval include zero, thus we can reject the null hypothesis and conclude that an increasing dose level leads to an increase in tooth lenght.

Assumptions

1. Members of the sample population (the 60 guinea pigs) are representative of the entire population of guinea pigs. This assumption allows us to generalize the results.

Conclusion

The dose of Vitamin C does influence the tooth growth of Guinea pigs. THe delivery mode (Ascorbic Acid or Orange Juice) does not have an impact on the growth of their teeth.