Statistical Inference Course Project - Part 2

Giovanny Jimenez Gomez

November 25, 2020

1. Basic inferential data analysis instructions.

1. Load the ToothGrowth data and perform some basic exploratory data analyses.

```
set.seed(2018)
data("ToothGrowth")       # Load Data
str(ToothGrowth)           # Structure of data
```

```
## '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 ...
```

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

2. Provide a basic summary of the data.

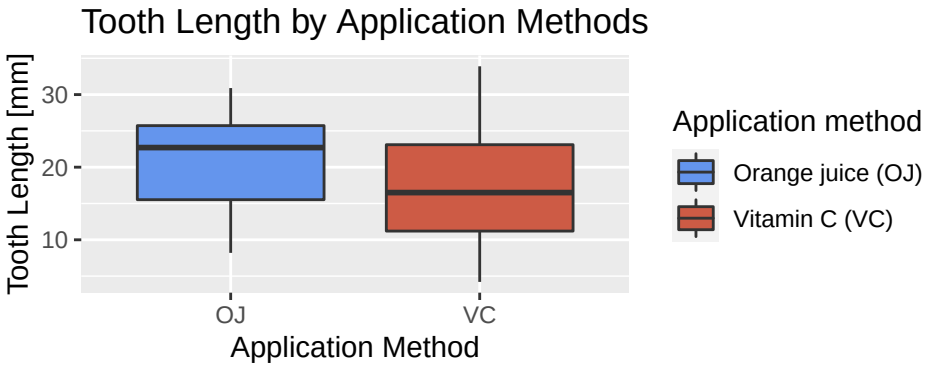
```
summary(ToothGrowth)           # Structure of data
```

```
##       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
```

```
library(ggplot2)              # In order to provide visualization
mean.supp <- split(ToothGrowth$len, ToothGrowth$supp) # Means by supp
sapply(mean.supp, mean)
```

```
##       OJ       VC
## 20.66333 16.96333
```

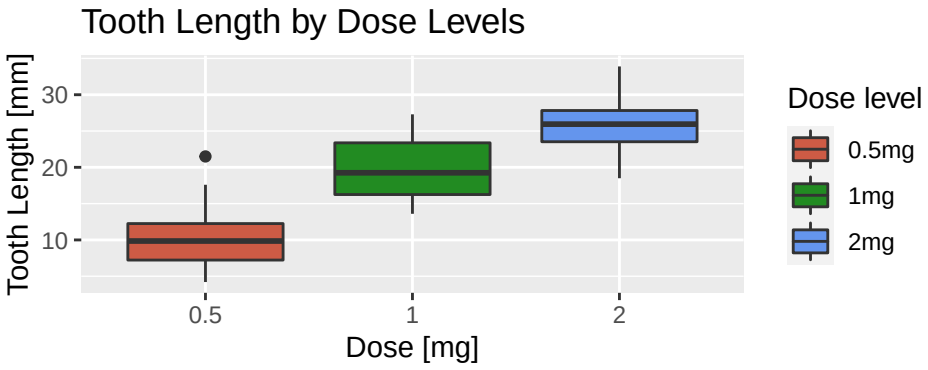
```
ggplot(aes(x = supp, y = len), data = ToothGrowth) + geom_boxplot(aes(fill = supp)) + xlab("Application
```




```
mean.dose <- split(ToothGrowth$len, ToothGrowth$dose) # Means by dose
sapply(mean.dose, mean)
```

```
##    0.5      1      2
## 10.605 19.735 26.100
```

```
ggplot(aes(x = dose, y = len), data = ToothGrowth) + geom_boxplot(aes(fill = dose)) + xlab("Dose [mg]")
```



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).

We are comparing tooth growth by supplement using a $t-test$.

```
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
```

Since the $p - value = 0.06 > 0.05$ and the confidence interval $[-0.17, 7.57]$ contains zero, we can say that supplement types seems to have no impact on Tooth growth based on this test.

Now, in order to compare tooth growth by dose, we need to look at the different pairs of dose values.

```
# t-test using dose amounts 0.5 and 1.0 [a 0.5 increase in dosage]
ToothGrowth.sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,0.5))
t.test(len~dose,data=ToothGrowth.sub)
```

```
##
##  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
```

```
# t-test using dose amounts 1.0 and 2.0 [a 1 increase in dosage]
ToothGrowth.sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(2.0,1.0))
t.test(len~dose,data=ToothGrowth.sub)
```

```
##
##  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
```

```
# t-test using dose amounts 0.5 and 2.0 [a 1.5 increase in dosage]
ToothGrowth.sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(2.0,0.5))
t.test(len~dose,data=ToothGrowth.sub)
```

```
##
##  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
```

In all three cases:

1. The $p - value$ was approximately zero, and < 0.05.
2. The confidence interval does not include zero.
3. Furthermore, the higher the increase in dosage, the smaller the $p - value$ of the test.

As a result, we can assume that the average tooth length increases with an inceasing dose, and therefore the three null hypothesis can be rejected.

4. Conclusions and assumptions.

Given the following assumptions:

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

By observing the t-test analysis above, we can conclude that: (a) supplement delivery method has no effect on tooth $growth/length$, but (b) increased dosages do result in increased tooth length.