Statistical Inference Course Project Part 2

Muhammad Imran Saeed

For this part of the project, we are going to analyze the ToothGrowth data in the R datasets package.

Tasks to be perfromed for the assignment:

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.

Installation and setup.

```
data("ToothGrowth")
```

Loading and preprocessing the data The data shows the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice (OJ) or ascorbic acid(VC) (a form of vitamin C).

The variable names and the structure of the file are given by

```
# Query names from data set
names(ToothGrowth)
```

```
## [1] "len"  "supp" "dose"
```

```
# Query structure of data set
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 ...
```

```
# Check sample data limited to 4 rows to ensure maximum of 6 pages
head(ToothGrowth,4)
```

```
##    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
```

```
# Create needed columns for analysis.
ToothGrowth$dosage <- as.factor(ToothGrowth$dose)
```

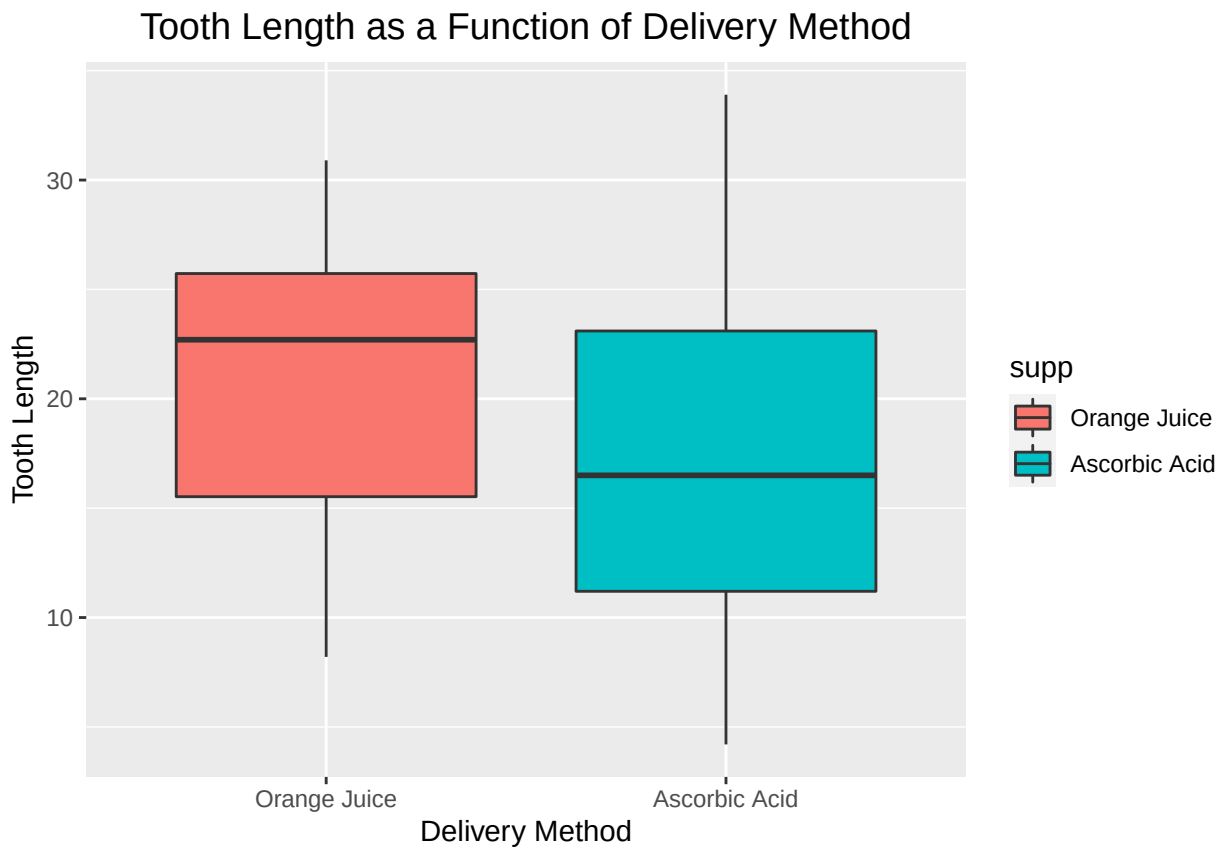
Summary of the data.

```
summary(ToothGrowth)
```

```
##       len        supp         dose       dosage
##  Min.   : 4.20   OJ:30   Min.   :0.500   0.5:20
##  1st Qu.:13.07   VC:30   1st Qu.:0.500   1  :20
##  Median :19.25           Median :1.000   2  :20
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

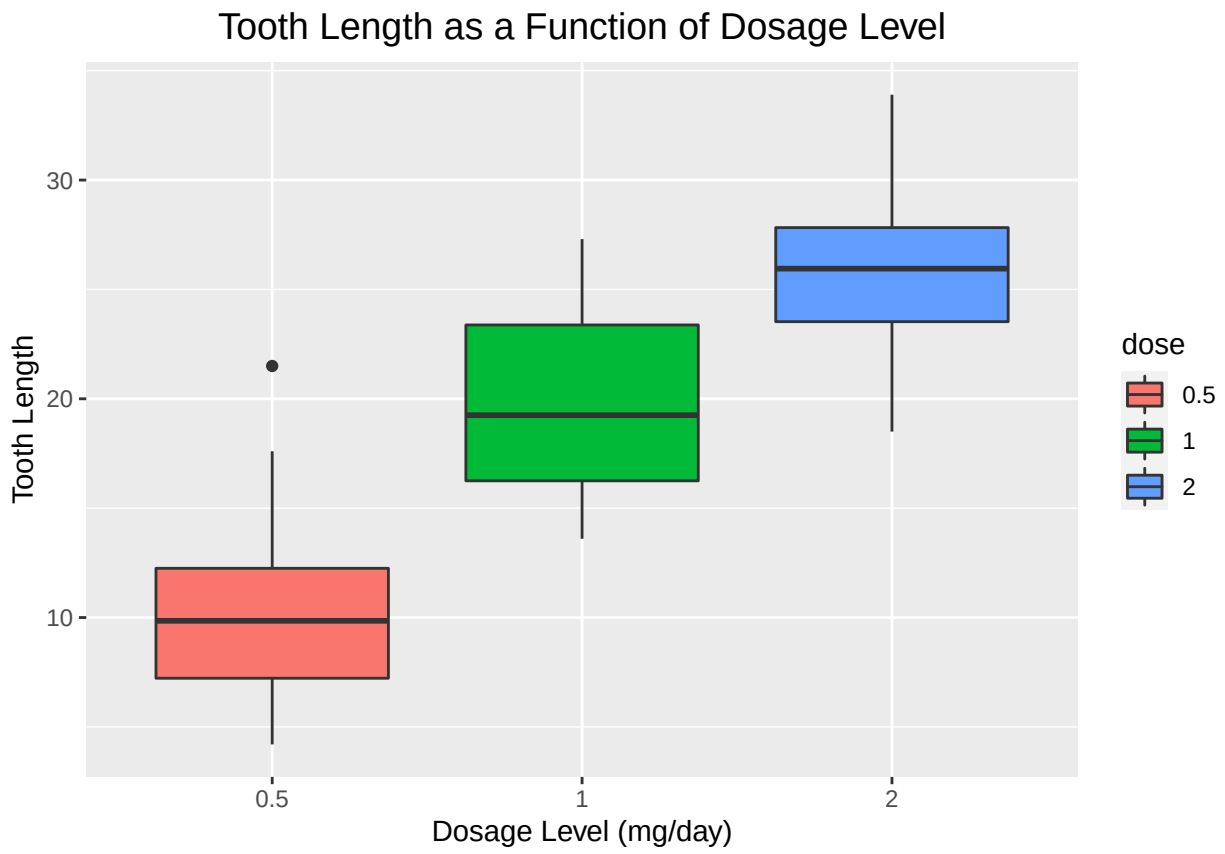
Tooth Length to Delivery Method

```
tg <- ToothGrowth
levels(tg$supp) <- c("Orange Juice", "Ascorbic Acid")
gLenSupp <- ggplot(data = tg, aes(x = supp, y = len)) +
    geom_boxplot(aes(fill = supp)) +
    xlab("Delivery Method") +
    ylab("Tooth Length") +
    theme(plot.title = element_text(size = 14, hjust = 0.5)) +
    ggtitle("Tooth Length as a Function of Delivery Method")
print(gLenSupp)
```



Observation Above indicates using OJ as delivery method,independent of dosage, has favorable effect on tooth growth than ascorbic acid. ### Tooth Length to Dosage Level

```
gLenDose <- ggplot(data = ToothGrowth, aes(x = factor(dose), y = len)) +
    geom_boxplot(aes(fill = factor(dose))) +
    xlab("Dosage Level (mg/day)") +
    ylab("Tooth Length") +
    guides(fill=guide_legend(title="dose")) +
    theme(plot.title = element_text(size = 14, hjust = 0.5)) +
    ggtitle("Tooth Length as a Function of Dosage Level")
print(gLenDose)
```

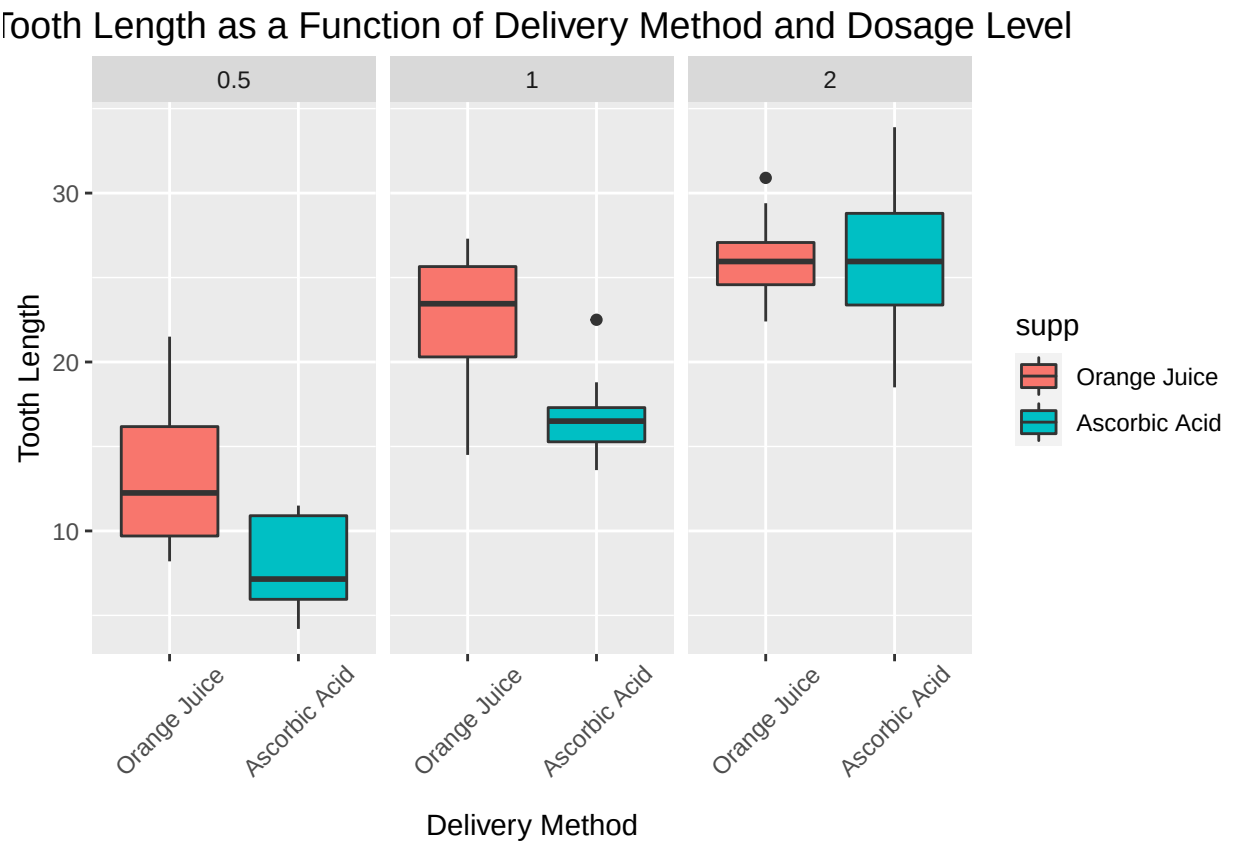


Observation

Above indicates higher dosage of VC irrespective of deliver method have favorable outcome on tooth growth. The best results are achieved by high dosage i.e 2mg/day followed by dosage in descending order(1mg ,0,5 mg/ day).

Tooth Length to Delivery Method and Dosage Level

```
tg <- ToothGrowth
levels(tg$supp) <- c("Orange Juice", "Ascorbic Acid")
gLenSuppDose <- ggplot(data = tg, aes(x = supp, y = len)) +
    geom_boxplot(aes(fill = supp)) +
    facet_wrap(~ dose) +
    xlab("Delivery Method") +
    ylab("Tooth Length") +
    guides(fill=guide_legend(title="supp")) +
    theme(plot.title = element_text(size = 14, hjust = 0.5, vjust = 0.5),
          axis.text.x = element_text(angle = 45,
                                     hjust = 0.5,
                                     vjust = 0.5,
                                     margin = margin(b = 10))) +
    ggtitle("Tooth Length as a Function of Delivery Method and Dosage Level")
print(gLenSuppDose)
```



Observation Above chart indicates higher dosage i.e mg/day is more effective than the lower dosage(s); however, all delivery methods have similar effectiveness.

```
Lower <- subset(tg, dose %in% c(0.5, 1.0))
Middle <- subset(tg, dose %in% c(0.5, 2.0))
Upper <- subset(tg, dose %in% c(1.0, 2.0))
```

Perform T test on basis of dosage for Lower.

```
t.test(len ~ dose, paired = F, var.equal = F, data = Lower)
```

```
##
##  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
```

Perform T test on basis of dosage for Middle.

```r
t.test(len ~ dose, paired = F, var.equal = F, data = Middle)
```

```
##
##  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
```

Perform T test on basis of dosage for Upper.

```r
t.test(len ~ dose, paired = F, var.equal = F, data = Upper)
```

```
##
##  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
```

Perform T test on basis of supply

```r
t.test(len ~ supp, paired = F, var.equal = F, data = tg)
```

```
##
##  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 Orange Juice mean in group Ascorbic Acid
##                    20.66333                    16.96333
```

Conclusions and the assumptions needed for your conclusions

1. Supplement type has no effect on tooth growth as we can't reject null hypothesis here
2. Increasing dosage results in positive tooth growth, the clear growth of confidence interval allow us to reject null hypothesis