Statistical Inference Project Part 2: Basic Inferential Data Analysis Instructions

Susmitha Saripalli

Project Description

Analyze the ToothGrowth data in the R datasets package.

Loading Libraries and data

```
library(data.table)
library(ggplot2)
library(datasets)
```

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

```
data("ToothGrowth")

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
```

```
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
```

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

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

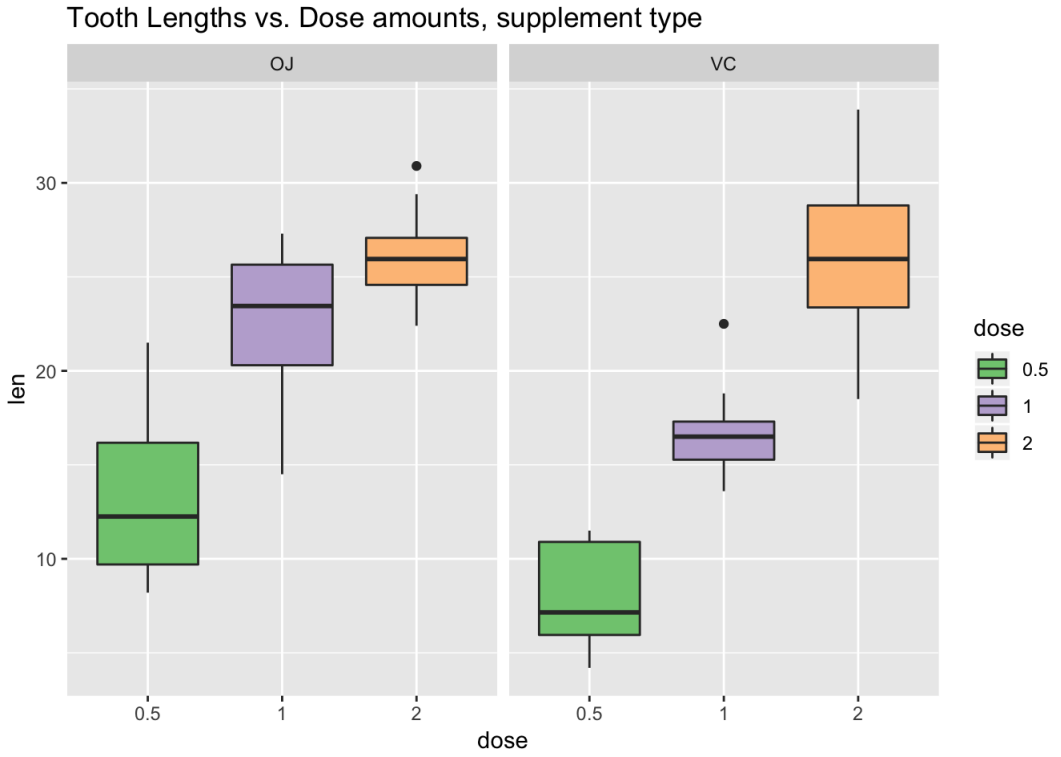
```
unique(ToothGrowth$dose)
```

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

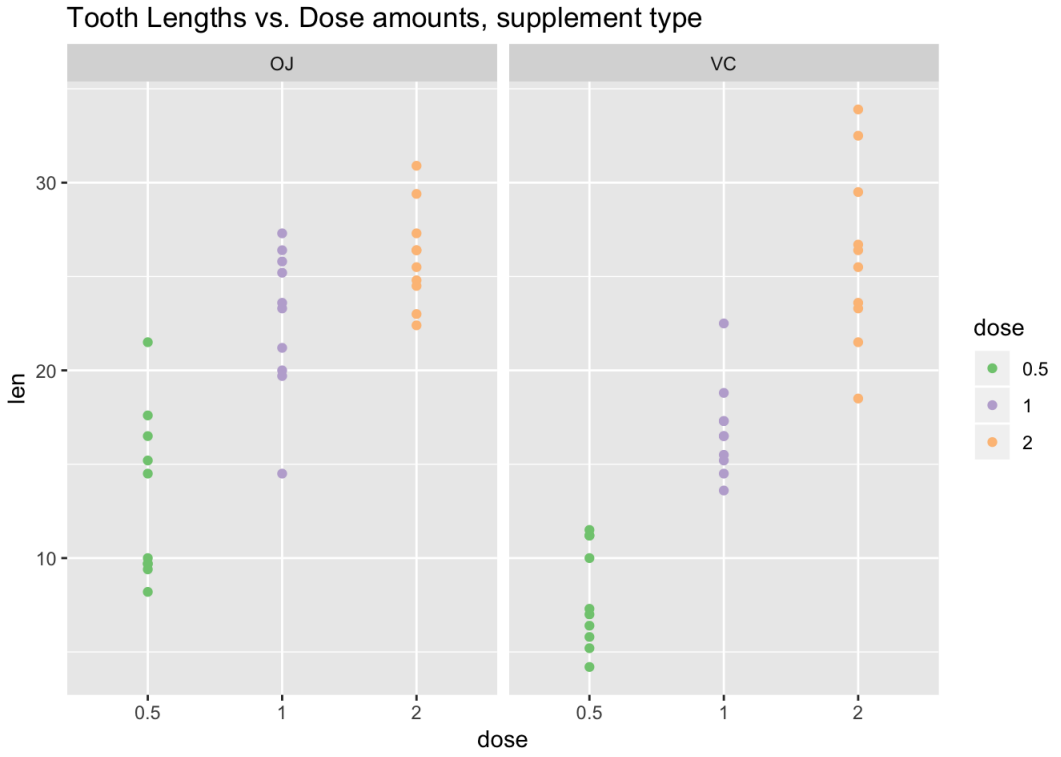
```
# looking at doses for each supplememnt type
table(ToothGrowth$supp, ToothGrowth$dose)
```

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

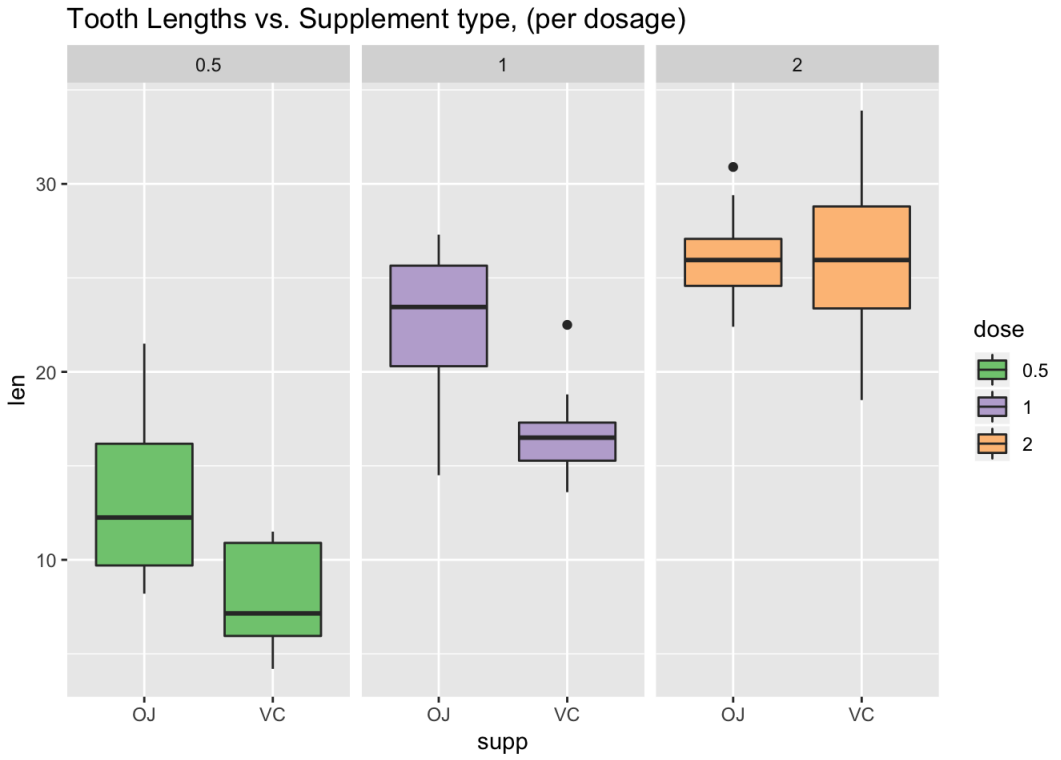
```
# By Supplement Type
ToothGrowth$dose = as.factor(ToothGrowth$dose)
g <- ggplot(ToothGrowth, aes(x=dose, y= len))
g + geom_boxplot(aes(fill=dose)) +
  facet_grid(.~supp) +
  scale_fill_brewer(palette="Accent") +
  ggtitle("Tooth Lengths vs. Dose amounts, supplement type")
```



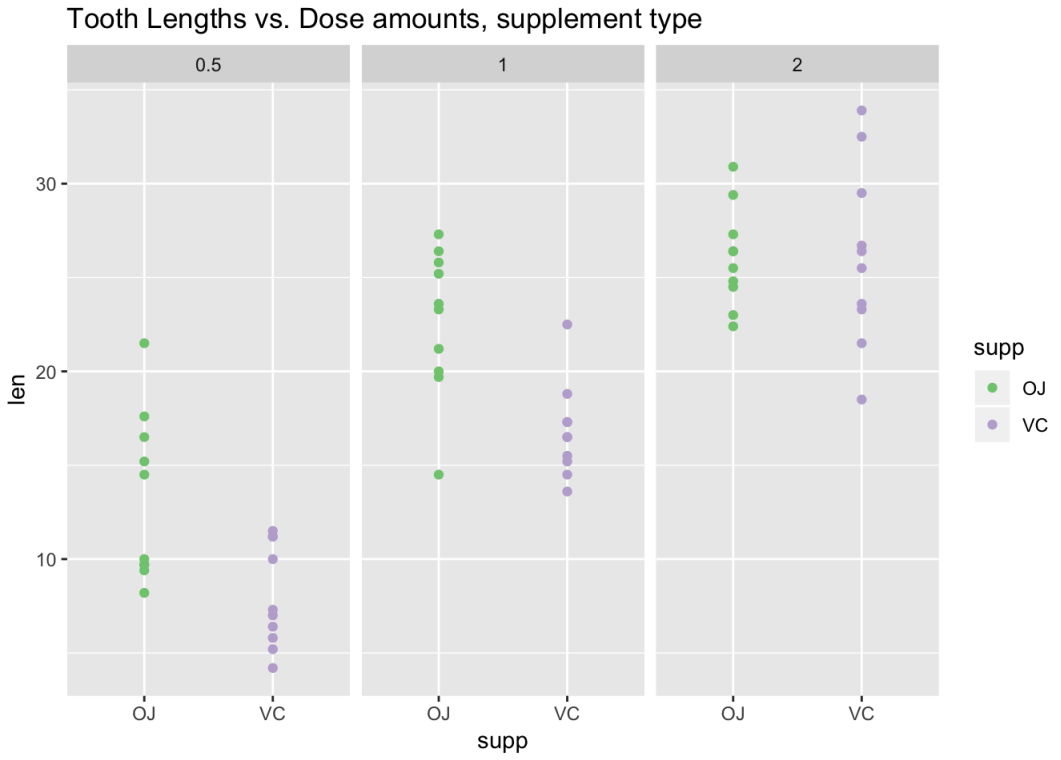

```
s <-ggplot(ToothGrowth,  aes(x=dose, y= len, color = dose))
s + geom_point() +
  facet_grid(.~supp) +
  scale_color_brewer(palette="Accent") +
  ggtitle("Tooth Lengths vs. Dose amounts, supplement type")
```



```
# By Dosage
g <- ggplot(ToothGrowth, aes(x=supp, y= len))
g + geom_boxplot(aes(fill=dose)) +
  facet_grid(.~dose) +
  scale_fill_brewer(palette="Accent") +
  ggtitle("Tooth Lengths vs. Supplement type, (per dosage)")
```




```
s <-ggplot(ToothGrowth,  aes(x=supp, y= len, color = supp))
s + geom_point() +
  facet_grid(.~dose) +
  scale_color_brewer(palette="Accent") +
  ggtitle("Tooth Lengths vs. Dose amounts, supplement type")
```



2. Provide a basic summary of the data.

The general trend is that tooth length seems proportional to dosage, meaning longer teeth require a higher dosage, which is to be expected. The data seems collected in a way that gives us 10 of each supplement type per dose amount so we can't make any conclusions about the usage of one supplement type over another. Alternatively, there doesn't seem to be a visual correlation between supplement type and tooth length.

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

t-test for group differences due to different supplement type,

(assume unequal variances between the two groups)

```
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 is 0.06063 and zero is in the confidence interval. This means we can't reject the null hypothesis: "different supplement types have no effect on tooth length"

t-test for group differences due to different dosages

splitting the doses into subsets of different pairs of dosages

```
doses_0.5_1.0 <- subset (ToothGrowth, dose %in% c(0.5, 1.0))
doses_0.5_2.0 <- subset (ToothGrowth, dose %in% c(0.5, 2.0))
doses_1.0_2.0 <- subset (ToothGrowth, dose %in% c(1.0, 2.0))
```

dose levels (0.5, 1.0), assuming unequal variances between the two groups

```
t.test(len ~ dose, data = doses_0.5_1.0)
```

```
##
##  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
```

dose levels (0.5, 2.0), assuming unequal variances between the two groups

```
t.test(len ~ dose, data = doses_0.5_2.0)
```

```
##
##  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
```

dose levels (1.0, 2.0), assuming unequal variances between the two groups

```
t.test(len ~ dose, data = doses_1.0_2.0)
```

```
##
##  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
```

The p-value is less than 0.05 for all the dosage pairs, and the confidence interval does not contain zero. The mean tooth length of the higher dosage is always higher than the mean length of the lower dosage. This means we can reject the null hypothesis and conclude that increasing the dosage leads to an increase in tooth length.

Assumptions

To do t tests, we assumed unequal variances between the groups we were testing. In the case of supplement type, the groups were the two supplements (VC or OJ). In the case of dose levels, the groups were the three different combinations of dosage level pais (0.5 & 1.0, 0.5 & 2.0, 1.0 & 2.0).

Conclusions

t tests confirmed that supplement type has no effect on tooth but inreasing the dosage level leads to an increased tooth growth.