Exploratory Data Analysis: Tooth Growth

Nino Munoz

January 30, 2016

Synopsis

In this project, we will analyze the ToothGrowth data in the R datasets package. The data contains the length of odontoblasts (teth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5mg, 1mg, and 2mg) with each of two delivery methods (orange juice or absorbic acid). In our analysis, we will:

- Load the ToothGrowth data and perform some basic exploratory data analyses.
- Provide a basic summary of the data.
- Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.
- State conclusions and assumptions needed for those conclusions.

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

1. Load relevant packages.

```
library(datasets)
library(ggplot2)
```

2. Load data from ToothGrowth dataset.

```
## Load ToothGrowth dataset
data(ToothGrowth)

## Store data in variable
data <- ToothGrowth
```

3. Perform preliminary data exploration.

```
## Look at str summary of data
str(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 ...

## Look at first few rows of the data
head(data)
```

```
##    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
```

Provide a basic summary of the data.

1. Look at the summary data from the summary function.

```
## Convert dose to a factor variable.
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

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
```

2. Calculate the mean tooth length.

```
mean(ToothGrowth$len)
```

```
## [1] 18.81333
```

3. Calculate the standard deviation for tooth length.

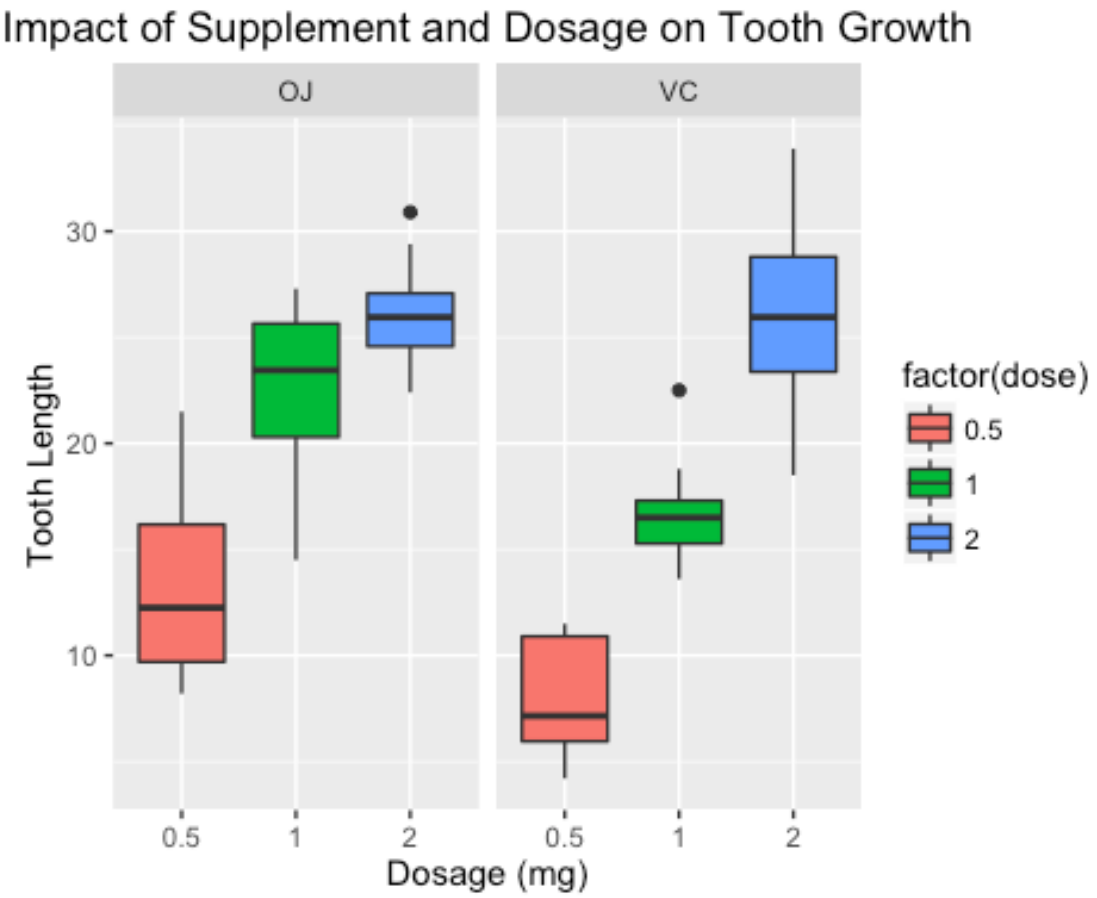
```
sd(ToothGrowth$len)
```

```
## [1] 7.649315
```

4. Generate a box plot to model the data.

```
plot <- ggplot(ToothGrowth,
               aes(x = factor(dose), y = len, fill = factor(dose)))

plot + geom_boxplot(notch = FALSE) + facet_grid(. ~ supp) +
    scale_x_discrete("Dosage (mg)") +
    scale_y_continuous("Tooth Length") +
    ggtitle("Impact of Supplement and Dosage on Tooth Growth")
```



The figure above is a boxplot modeling tooth length vs. dosage for each supplement type

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

1. Compare dosage and tooth growth, with the null hypothesis that there is no correlation between the two.

```
## Separate dosages into three intervals
dose.interval1 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
dose.interval2 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
dose.interval3 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))

## Hypothesis test dosage interval 1
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data =
dose.interval1)

##
##  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
```

```
## Hypothesis test dosage interval 2
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data =
dose.interval2)
```

```
##
##  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
```

```
## Hypothesis test dosage interval 3
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data =
dose.interval3)
```

```
##
##  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 confidence intervals allow for the rejection of the null hypothesis, and confirms that there is a significant correlation between tooth length and dosage.

---

2. Compare supplement type and tooth growth, with the null hypothesis that there is no correlation between the two.

```
## Hypothesis test tooth length vs. supp
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, 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 confidence intervals do not allow for the rejection of the null hypothesis, confirming that supplement type and tooth growth are not significantly correlated.

State conclusions and assumptions needed for those conclusions.

CONCLUSIONS:

1. Supplement type does not have a significant impact on tooth growth.
2. Increased dosage leads to increased tooth growth.

ASSUMPTIONS:

1. The guinea pigs were randomly selected and have similar characteristics.
2. Researchers were unaware of which guinea pigs were given which treatment.