SI Course Project (Part 2)

Cecilia Cruz-Ram, MD DPCOM

4/7/2018

Basic Inferential Data Analysis

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

A. Setwd and load package

```
setwd("/Users/sexybaboy/Documents/Files/Zetch/Online Courses/Data Science Specialization Feb18/R/Statistical I
require(ggplot2)
```

```
## Loading required package: ggplot2
```

B. Instructions

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

Load ToothGrowth data

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

Show Unique Values

```
unique(ToothGrowth$len)
```

```
##  [1]  4.2 11.5  7.3  5.8  6.4 10.0 11.2  5.2  7.0 16.5 15.2 17.3 22.5 13.6
## [15] 14.5 18.8 15.5 23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5 23.3 29.5 17.6
## [29]  9.7  8.2  9.4 19.7 20.0 25.2 25.8 21.2 27.3 22.4 24.5 24.8 30.9 29.4
## [43] 23.0
```

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

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

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

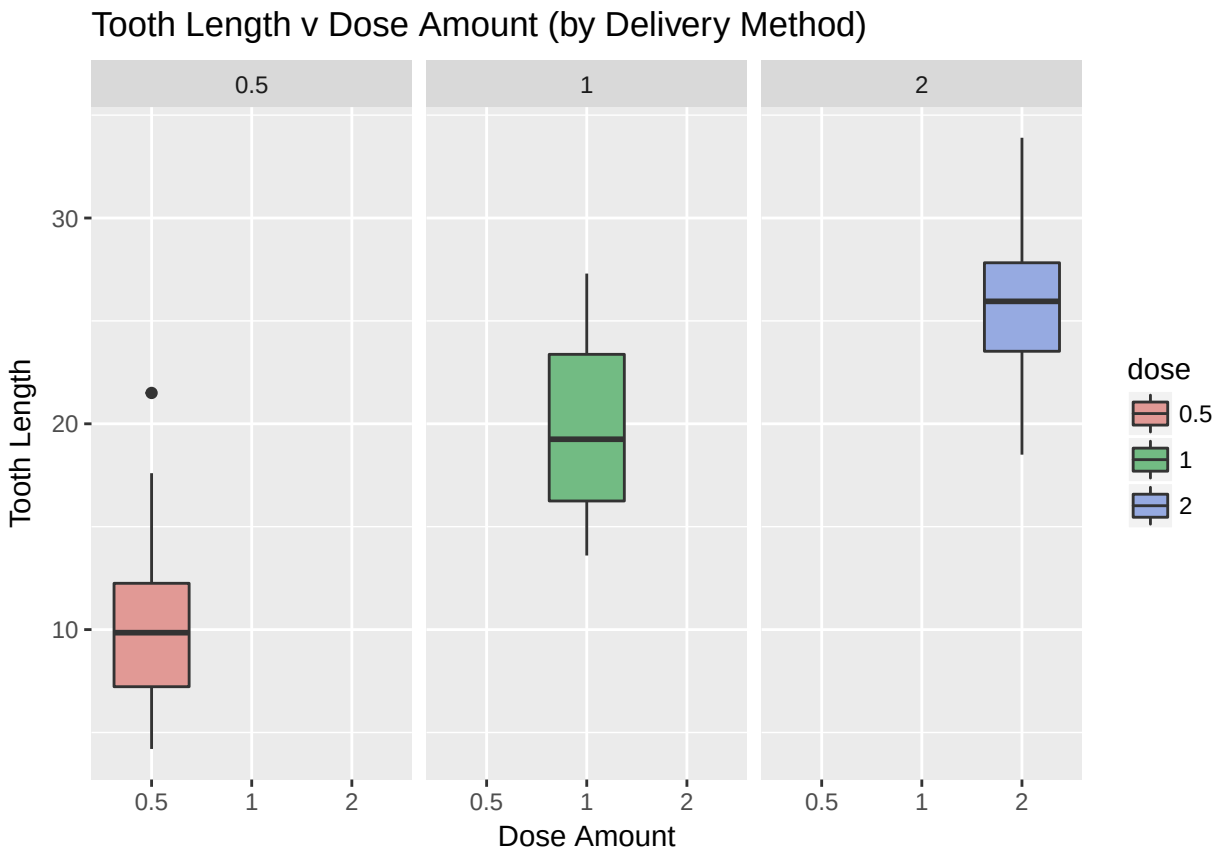
```
## [1] 0.5 1.0 2.0
```

Analyze and explore data via plots Convert dose to a factor

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

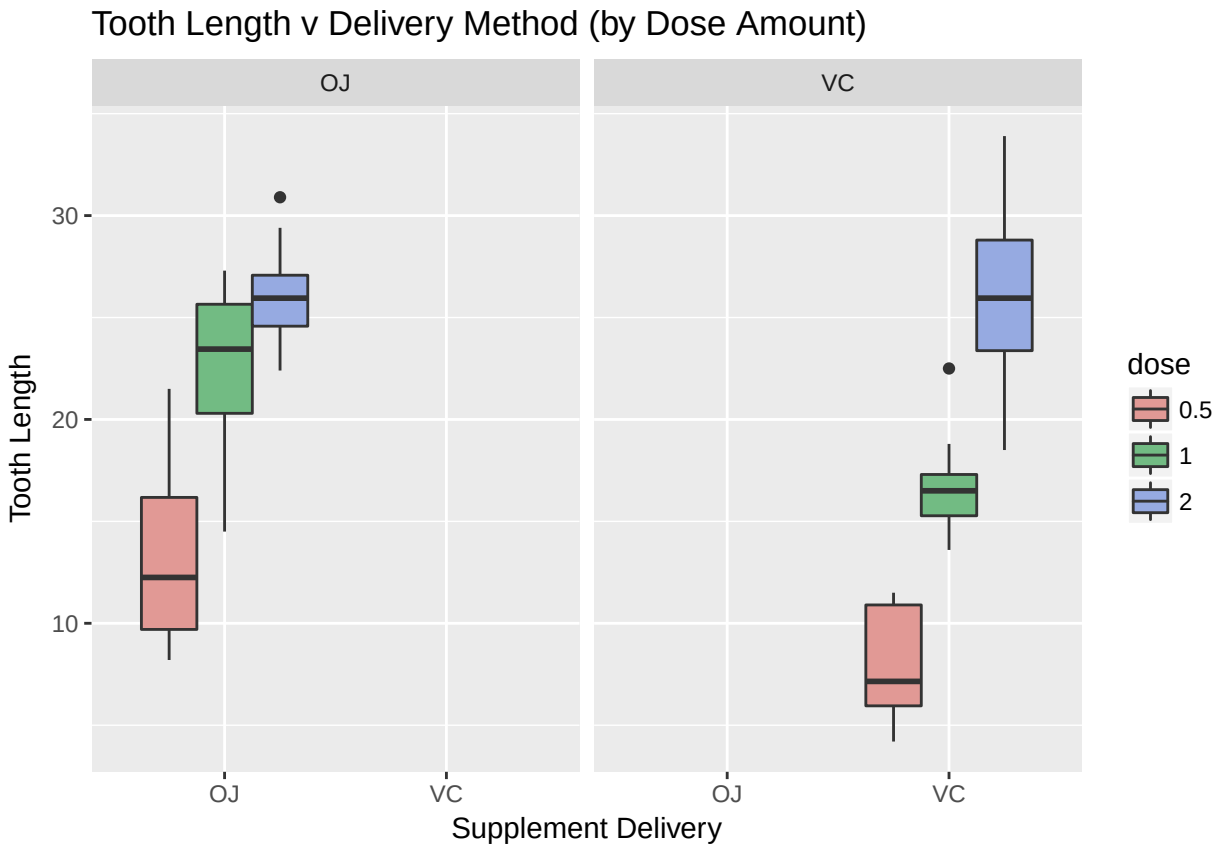
Plot tooth length ('len') vs. the dose amount ('dose'), broken out by supplement delivery method ('supp')

```
bp <- ggplot(ToothGrowth, aes(x = dose, y = len, fill = dose)) +
  geom_boxplot(position = "dodge", varwidth = TRUE) + xlab("Dose Amount") + ylab("Tooth Length") + facet_grid(
  ggtitle("Tooth Length v Dose Amount (by Delivery Method)") +
  theme(plot.title = element_text(lineheight = .3))
bp + scale_fill_hue(l = 70, c = 50)
```



Plot tooth length ('len') vs. supplement delivery method ('supp') broken out by the dose amount ('dose')

```
bp <- ggplot(ToothGrowth, aes(x = supp, y = len, fill = dose)) +
  geom_boxplot(position = "dodge", varwidth = TRUE) + xlab("Supplement Delivery") + ylab("Tooth Length") + fac
  theme(plot.title = element_text(lineheight = .3))
bp + scale_fill_hue(l = 70, c = 50)
```



2. Provide a basic summary of the data.

```
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
```

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)

Run 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
```

Run t-test using dose amounts 0.5 and 1.0

```
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
```

Run t-test using dose amounts 0.5 and 2.0

```
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5, 2.0))
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
```

Run t-test using dose amounts 1.0 and 2.0

```
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0, 2.0))
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
```

4. State your conclusions and the assumptions needed for your conclusions.

Assuming the sample is representative of the population and the distribution of the sample means follows the Central Limit Theorem, we found out that the higher the dosage, the longer the tooth grows over time. Supplement delivery method, however, has no effect on tooth growth/length.