Part 2

Carlos Barco

5/01/2017

Part 2: 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.

1 Loading the data

```
# 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
```

```
tail(ToothGrowth)
```

```
##      len supp dose
## 55  24.8   OJ    2
## 56  30.9   OJ    2
## 57  26.4   OJ    2
## 58  27.3   OJ    2
## 59  29.4   OJ    2
## 60  23.0   OJ    2
```

```
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 ...
```

2 Basic data summary

Provide a basic summary of the data.

```
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
```

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)

```
library(ggplot2)
t = ToothGrowth
g <- ggplot(ToothGrowth, aes(x = dose, y = len, dose = factor(supp)))
g <- g + geom_line(size = 1, aes(colour = supp)) + geom_point(size =10, pch = 21,  alpha = .5)
g
```
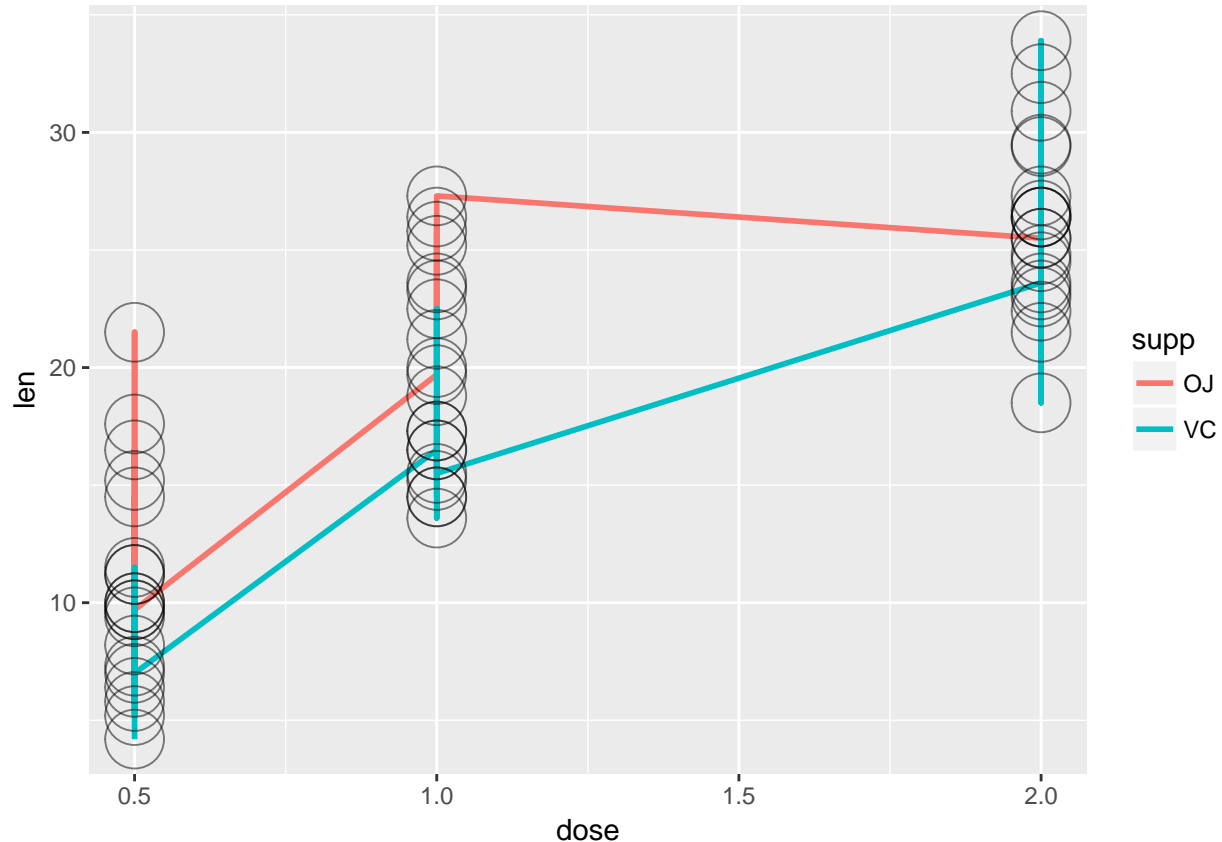

3 Hypothesis testing

Hypothesis 1: Orange juice and ascorbic acid delivers the same tooth growth across the data set.

```
hypo1 <- t.test(len~supp, data=t)
hypo1$conf.int
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

```
hypo1$p.value
```

```
## [1] 0.06063451
```

Because confidence interval includes 0 and p-value is greater than threshold .05, then we cannot reject hypothesis 1.

Hypothesis 2: If the dosage is .5 mg/day, both supplements delivers the same tooth growth.

```
hypo2 <- t.test(len~supp, data=subset(t, dose==.5))
hypo2$conf.int
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```
hypo2$p.value
```

```
## [1] 0.006358607
```

Because confidence interval does not include 0 and the p-value is below threshold .05, then we can reject hypothesis 2.

Hypothesis 3: If the dosage is 1.0 mg/day, both supplements delivers the same tooth growth.

```
hypo3 <- t.test(len~supp, data=subset(t, dose==1.0))
hypo3$conf.int
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```
hypo3$p.value
```

```
## [1] 0.001038376
```

Because confidence interval does not include 0 and the p-value is below threshold .05, then we can reject hypothesis 3.

Hypothesis 4: If the dosage is 2.0 mg/day, both supplements delivers the same tooth growth.

```
hypo4 <- t.test(len~supp, data=subset(t, dose==2.0))
hypo4$conf.int
```

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

```
hypo4$p.value
```

```
## [1] 0.9638516
```

Because confidence interval includes 0 and the p-value is greater than threshold .05, then we cannot reject hypothesis 4.

State your conclusions and the assumptions (needed for your conclusions).

Conclusions: By using confidence intervals to compare Tooth Growth by supplement type -Orange Juice(OJ) vs Vitamin C(VC) and dose size, we can saw that higher the dose applied, longer the tooth grows over time. But, according to the data set, we cannot yet affirm that OJ is more effective with regards to Tooth Growth than VC. **Assumptions:** The tooth lengths follows a normal distribution and no other unmeasured factors are affecting it. The analysis were based on a 'two.sided' approach.