Statistical Inference (statinference-011) Course Project - Part 2

Esmit Pérez

February 16, 2015

Part 2

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

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

```
#load data set from R's provided datasets
data(ToothGrowth)
```

Provide a basic summary of the data.

Note: Please refer to appendix for a complete summary.

The following plot shows the growth length, per supplement at varying dosages (see R code and graph in appendix, fig 2.1):

The Effect of Vitamin C on Tooth Growth in Guinea Pigs
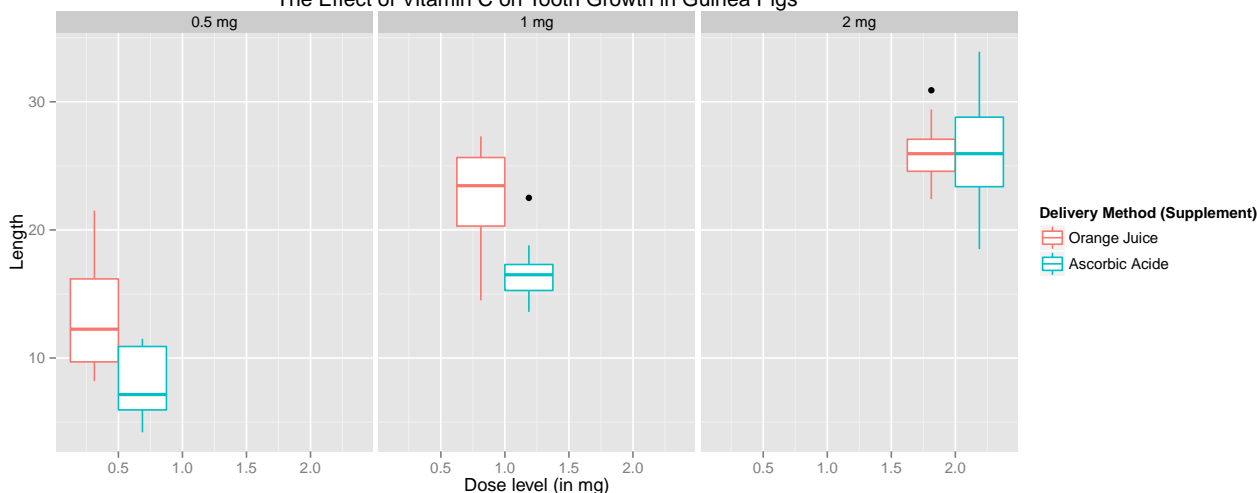
Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

H_0 : Supplement *OJ (Orange Juice)* leads to more growth.

State your conclusions and the assumptions needed for your conclusions.

Based on (Week 3 video) (https://class.coursera.org/statinference-011/lecture/243) we can perform a `t.test` to determine whether this is true, assuming variances are equal (see complete output in appendix, fig 2.2):

```
##              [,1]     [,2]
## [1,] -0.1670064 7.567006
## [2,] -0.1670064 7.567006
```

Factors in `supp` show `OJ` comes first (see appendix, fig 2.3) and in both tests the result, is the same, column `1`, correspoding to `OJ` (*Orange Juice*) is below zero, meaning it does not lead to more length.

Appendix

Part 2 - Supporting figures and exploratory analysis, ToothGrowth data.

Description

"The Effect of Vitamin C on Tooth Growth in Guinea Pigs"

The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

Summary information for `ToothGrowth`:

```r
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 ...
```

```r
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
```

```r
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
```

Example `coplot` as per `help(ToothGrowth)`:

```r
require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```
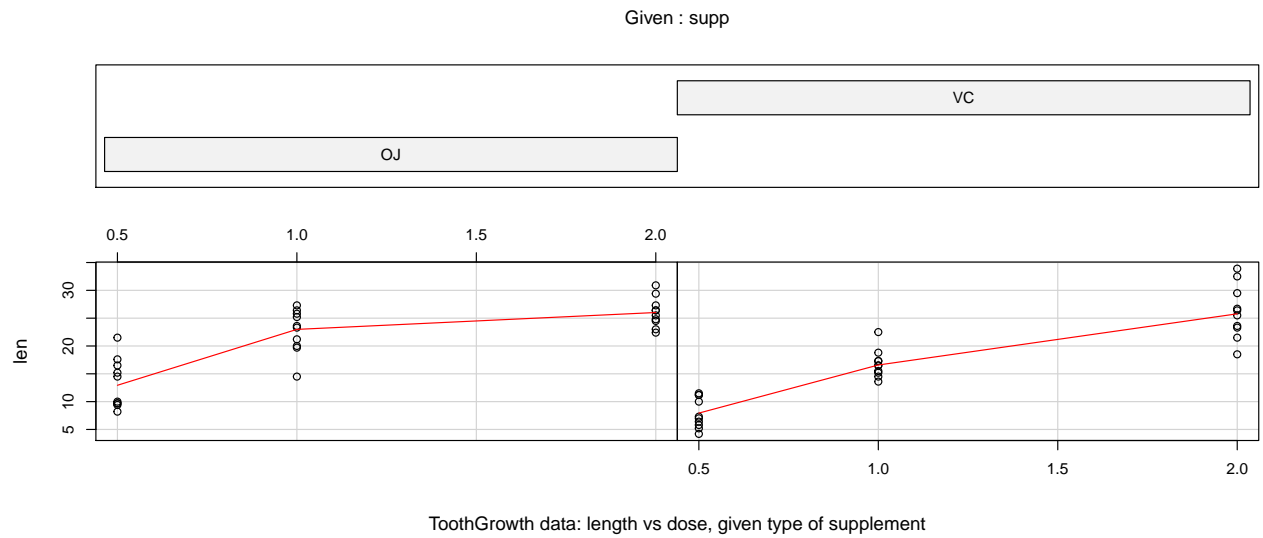
Given : supp

ToothGrowth data: length vs dose, given type of supplement

Figure 2.1

```
#plot and customize graph
g <- ggplot(ToothGrowth,aes(x = dose, y = len, color=supp))
g <- g + ggtitle("The Effect of Vitamin C on Tooth Growth in Guinea Pigs")
g <- g + xlab("Dose level (in mg)") + ylab("Length")
g <- g + scale_color_discrete(name="Delivery Method (Supplement)",
                    breaks=c("OJ", "VC"),
                    labels=c("Orange Juice", "Ascorbic Acide"))
g <- g + geom_boxplot()
g <- g + facet_grid(. ~ dose , labeller = function(variable, value){ paste0(value," mg") })
g
```
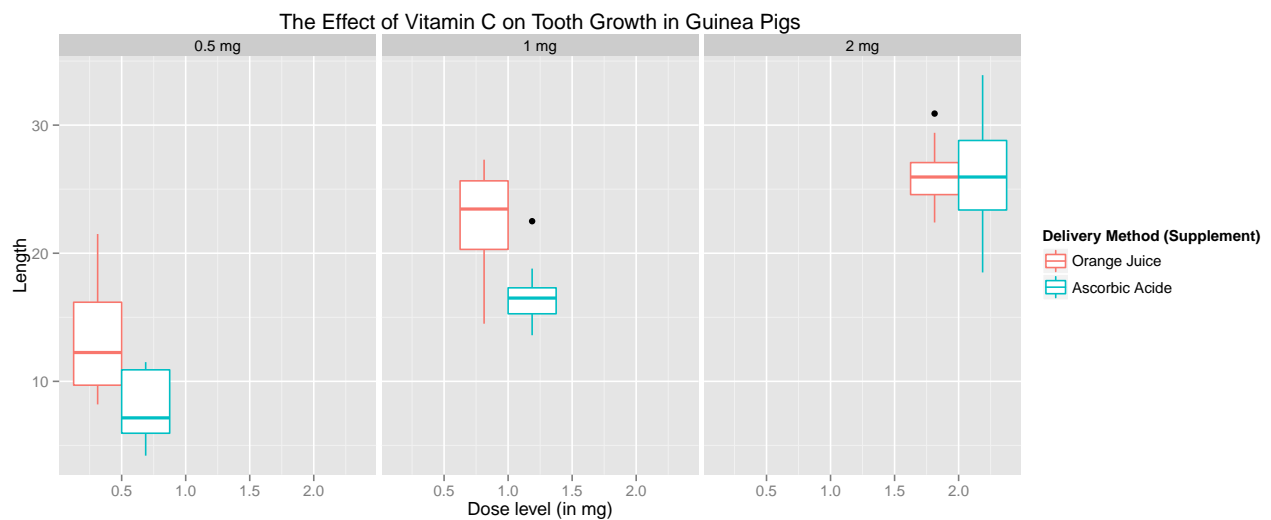

Figure 2.2

```
# Assume variances ARE equal
rbind(t.test(len ~ supp, paired=FALSE, var.equal=T, data=ToothGrowth)$conf,
      t.test(len ~ supp, paired=FALSE, var.equal=T, data=ToothGrowth)$conf
      )
```

```
##              [,1]     [,2]
## [1,] -0.1670064 7.567006
## [2,] -0.1670064 7.567006
```

```
factor(ToothGrowth$supp)
```

Figure 2.3

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