Statistical Inference Assesment Part 2

Sergey Chigrinov

November 14, 2015

This part is aimed at analysis of the ToothGrowth data in the R datasets package

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

```
library(datasets)
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 ...
```

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

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

```
suppressWarnings(library(ggplot2)); ggplot(ToothGrowth, aes(x = dose, y = len, color = supp)) +
    geom_point(shape = 1) + geom_smooth(method=lm) +  ylab("Tooth Length") + xlab("Vitamin C Dose") + gg
```
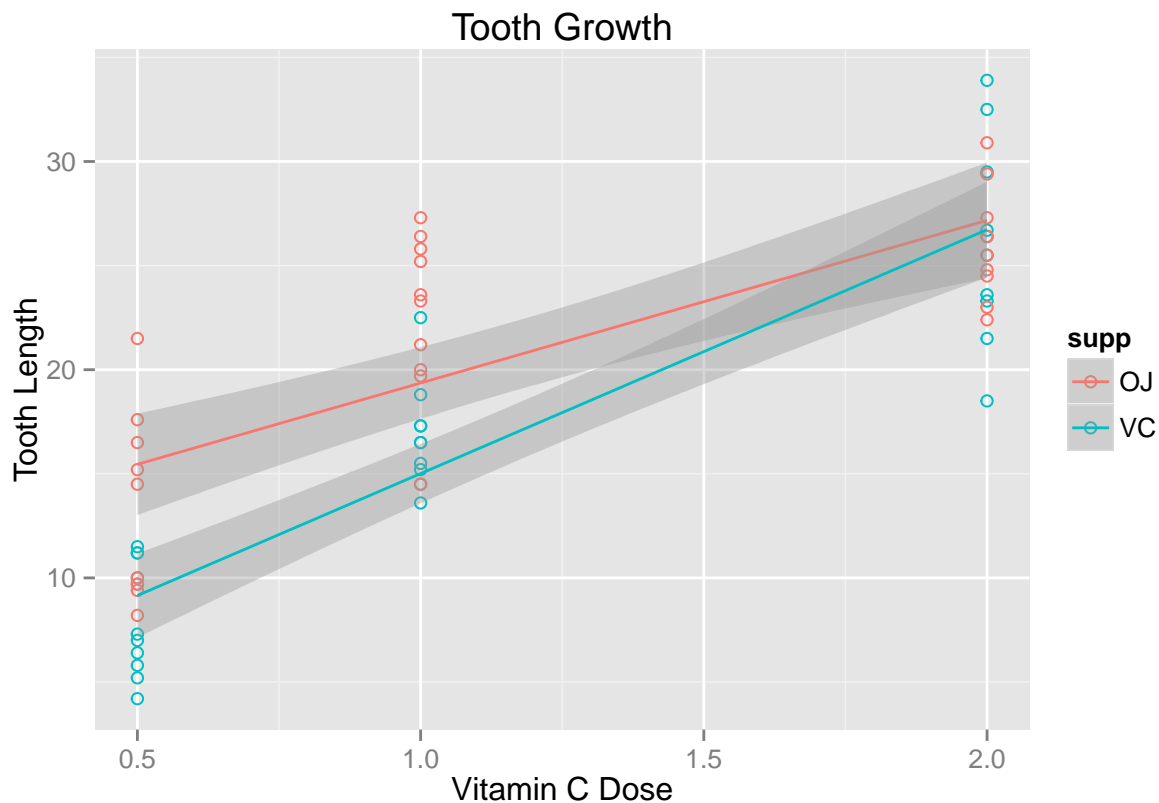
The ToothGrowth variable contain 60 observations, and consists of 3 fields - len, supp and dose. "len" stands for tooth length, "supp" stands for the delivery method, and "dose" shows one of the three dose levels.

2. 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
```

We see that the mean length of the teeth is 18.81, there are two delivery methods and dose is between 0.5 and 2.0

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

In this part we'll check whether we can conclude that there are significant differences in mean length for different subroups (different dose levels and delivery methods). To make paper shorter we show only confidence intervals.

```
t.test(len ~ dose, paired= FALSE, var.equal = FALSE, data = ToothGrowth[ToothGrowth$dose %in% c(0.5,1),]
```

```
## [1] -11.983781  -6.276219
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(len ~ dose, paired= FALSE, var.equal = FALSE, data = ToothGrowth[ToothGrowth$dose %in% c(1,2),])
```

```
## [1] -8.996481 -3.733519
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(len ~ dose, paired= FALSE, var.equal = FALSE, data = ToothGrowth[ToothGrowth$dose %in% c(0.5,2),]
```

```
## [1] -18.15617 -12.83383
## attr(,"conf.level")
## [1] 0.95
```

We can reject the null hypotesys in all above cases and reasonably assume that there is a difference in tooth growth for different doses.

```r
t.test(len ~ supp, paired= FALSE, var.equal = FALSE, data = ToothGrowth[ToothGrowth$dose==0.5,])$conf.in
```

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

```r
t.test(len ~ supp, paired= FALSE, var.equal = FALSE, data = ToothGrowth[ToothGrowth$dose==1,])$conf.int
```

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

```r
t.test(len ~ supp, paired= FALSE, var.equal = FALSE, data = ToothGrowth[ToothGrowth$dose==2,])$conf.int
```

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

We can note that there is no significant evidence that the mean of length of teeth is different for different vitamin C delivery methods when the dose equals 2. For all other cases we can reject the null hypotesys and conclude that the impact of the delivery method is significant.

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

From the graph of Tooth Growth and from the results of t-tests we can conclude that there are evidences that tooth growth depends on the dose of vitamin C. However, the delivery method is important only for small doses, while when dose=2 we can't see the strong evidences that mean tooth growth is different.