Statistical Inference - Tooth Growth

Ron Ferens

September 26, 2015

Overview

The **ToothGrowth** dataset describes 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).

Exploratory Data Analysis

Loading the Data

As first step, we will load the **ToothGrowth** database and make sure the data is properly formatted. We will also examine the column headers to make sure that they are properly formated for R data frames:

```
data(ToothGrowth)
str(ToothGrowth)
```

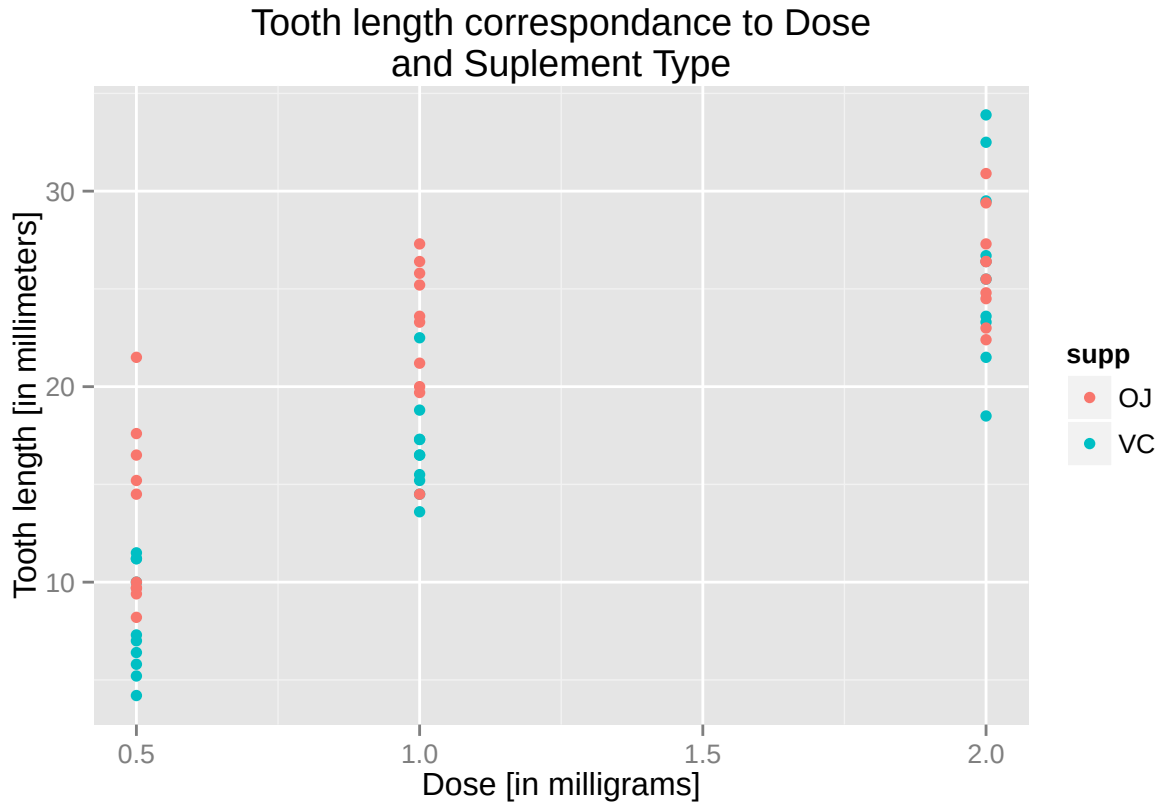
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

As shown above, the **ToothGrowth** database conssist of 60 observations of three variables:
1. `len`: Tooth length in millimeters (numeric variable)
2. `supp`: Supplement type (factor variable with levels VC and OJ)
3. `dose`: Dose in milligrams (numeric variable)

Examining the Data

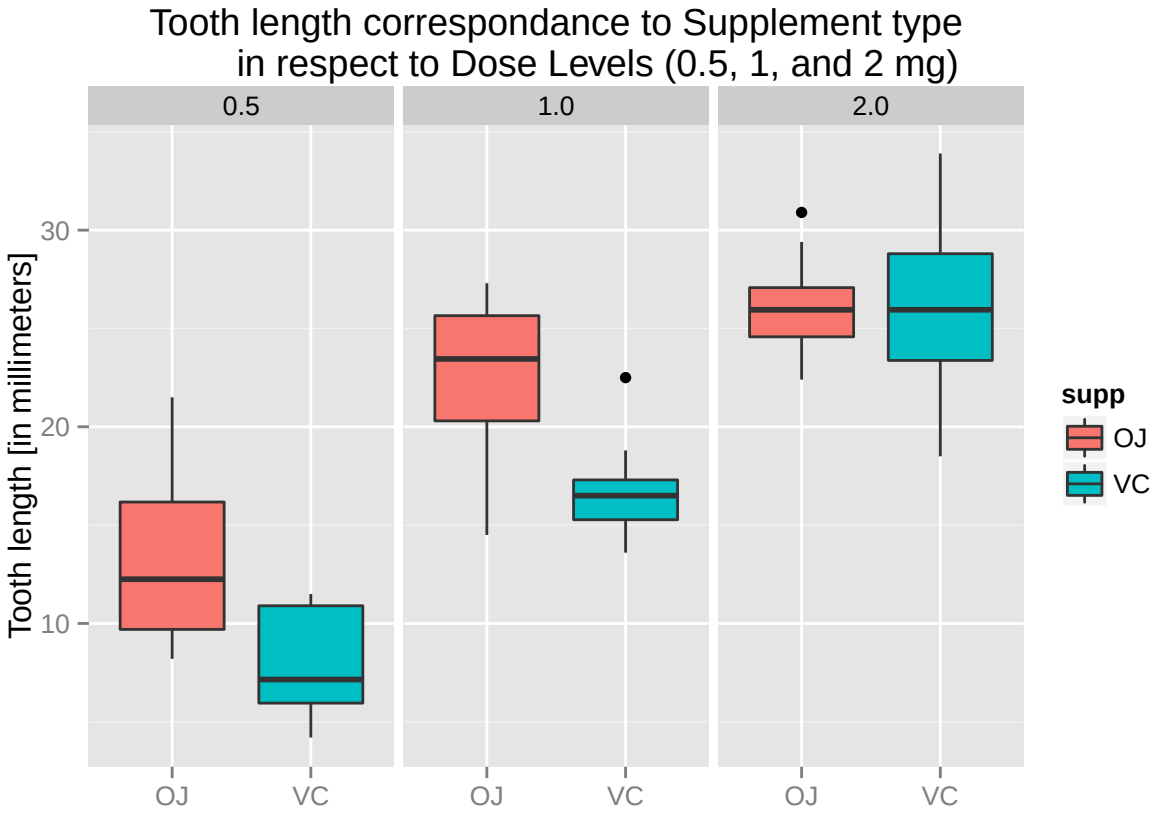
Once the data have been loaded properly, we can start performing an exploratory data analysis. Since the **ToothGrowth** dataset describes the effect of Vitamin C on Tooth Growth in Guinea Pigs, as first step, we would to plot the tooth length across different groups of supplement types, dose levels.

```
ggplot(aes(x=dose, y = len), data = ToothGrowth) +
  geom_point(aes(color = supp)) +
  labs( x="Dose [in milligrams]",
        y="Tooth length [in millimeters]",
        title="Tooth length correspondance to Dose\nand Suplement Type")
```



Based on the graph above, is seems that mainly the dosage affect tooth length. The higher the supplement, the longer the tooth length. When we try to examine the effect of the supplement type it is not clear whether the differences are statistically significant. In order to get a better understanding, we will evaluate the effect of the supplement type for each one of the of the three dose levels of Vitamin C (0.5, 1, and 2 mg).

```
ggplot(aes(x = supp, y = len), data = ToothGrowth) +
  geom_boxplot(aes(fill = supp)) + facet_wrap(~ dose) +
  labs( x=NULL,
        y="Tooth length [in millimeters]",
        title="Tooth length correspondance to Supplement type
        in respect to Dose Levels (0.5, 1, and 2 mg)")
```



It appears that for lower dosage levels: 0.5 and 1 mg, supplement type does affects tooth length. In those cases, orange joice (OJ) have larger effect than ascorbic acid (VC). However, at higher dosages no significant differences can be found for the two supplement type.

Confidence Intervals Testing

To validate our initial conclusions from the exploratory data analysis we have performed, we will run several confedence interval tests. Since our dataset is conssit of two parameters ('supp' and 'dose') run the follwoing confidence interval testing:

1. Dosage Level Alone
2. Supplement Type Alone
3. Supplement Type in respect to Dosage Level

Confidence Interval Testing for Dosage Level

In this section we will run cofidence interval test for all three dosage levels. We will try to examine whether indeed higher dosage result in longer tooth length.

```
Tooth.dose_05_to_10 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
Tooth.dose_05_to_20 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
Tooth.dose_10_to_20 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))

t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = Tooth.dose_05_to_10)
```

```
##
##  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
```

```
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = Tooth.dose_05_to_20)
```

```
##
##  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
```

```
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = Tooth.dose_10_to_20)
```

```
##
##  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
```

Based on the result of the three confedence interval testings, we can conclude that increase in the dosage level does affect the tooth length. We can determine since in all three tests, the *p-value* is smaller than the significance level 0.05. This mean that we can reject the *null hypothesis* that increacing the dose does not increase tooth length.

Confidence Interval Testing for Supplement Type

In this section we will run cofidence interval test for the two different supplement types - orange juice or ascorbic acid.

```r
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
```

For the confidence interval testing for *Supplement Type*, we got *p-value* which is **not** smaller than the significance level 0.05. This mean that we can not reject the *null hypothesis*. Hence, we **can not** determine that a differnet type of supplement does indeed increase tooth length.

Confidence Interval Testing for Supplement Type in respect to Dosage Level

In this third and final test, we will investige, whether a different supllement type have an affect of the pigs tooth length in each one of the dosage levels.

```r
Tooth.dose_05 <- subset(ToothGrowth, dose == 0.5)
Tooth.dose_10 <- subset(ToothGrowth, dose == 1.0)
Tooth.dose_20 <- subset(ToothGrowth, dose == 2.0)

t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = Tooth.dose_05)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```r
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = Tooth.dose_10)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
```

```
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = Tooth.dose_20)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

We can see that the latest confidence interval testing confirm our initial conclusions we got in the exploratory data analysis phase. For the two lower dosage levels (0.5 and 1.0 mg) we got *p-value* with is smaller than the significance level 0.05. However, for dosage level of 2.0 mg, we got *p-value* which is **not** smaller than the significance level 0.05. This means, that differnet suplement type have an effect on the pigs' tooth lenght only when lower dosage levels are used.

Conclusions

1. Dosage levels of Vitamin C have more significant effect on pigs' tooth length in respect to the suplement type. The higher the dosgae level the longer the tooth grows.

2. Suplement type have a notable effect to the tooth length only when smaller dosage level are used (0.5 and 1.0 mg).

3. When dosage level of 2.0 mg, the suplement type have no effect of the tooth length.

Assumptions

- The distribution of the given dataset are approximately normal
- For the CLT to apply, we assume that the given samples are i.i.d
- The variances between the separate tested populations are different (we used *var.equal = FALSE* for all the t tests)