Tooth Grown Data Analysis

Jenhau Chen

February 17, 2016

Objective

The goal is to use ToothGrowh data set in R to performn data analysis and derive statistical inference. I will be testing on:
* Regardless of supplements (VC, Ascorbic Acid or OJ, Orange Juice) will different dosage of Vitamin C affect tooth growh in Guinea Pigs?
* For the same amount of dosage, will different delivery method yield statistically signifcant results?

The Dataset

There is some confusion on how the study was conducted on the subjects and I did some research online and found some good clarifications:

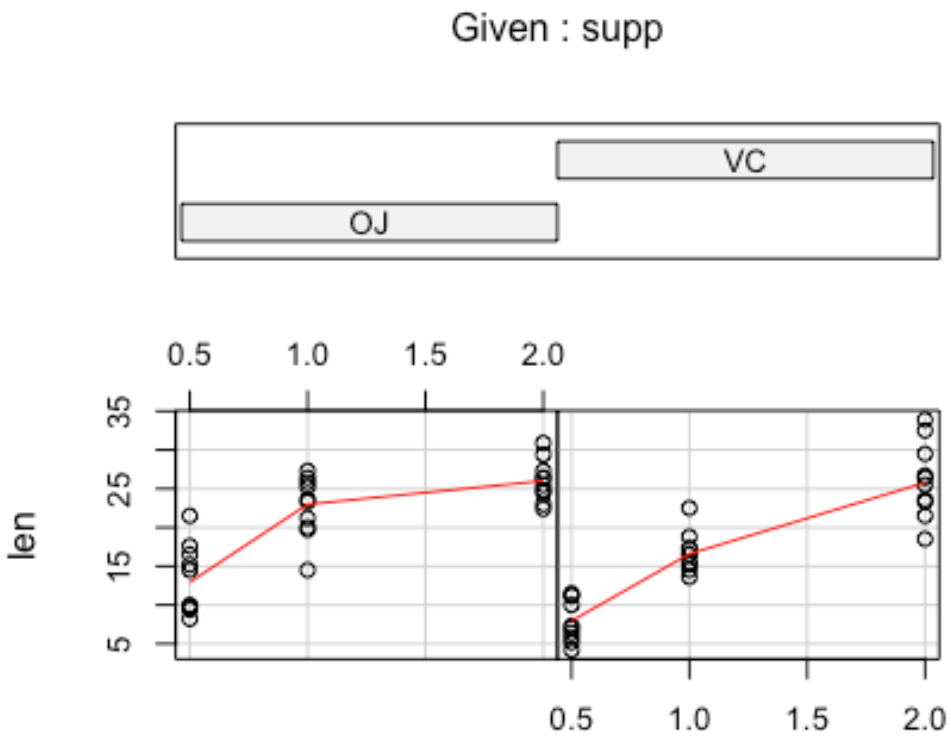
ToothGrowth description improvement

Assumptions

Based on the clarification found, this study was conducted on 60 guinea pigs. Therefore I am conducting tests assuming independent samples. I will not pair the subjects and nor will I assume equal variances

Exploratory Data Analysis

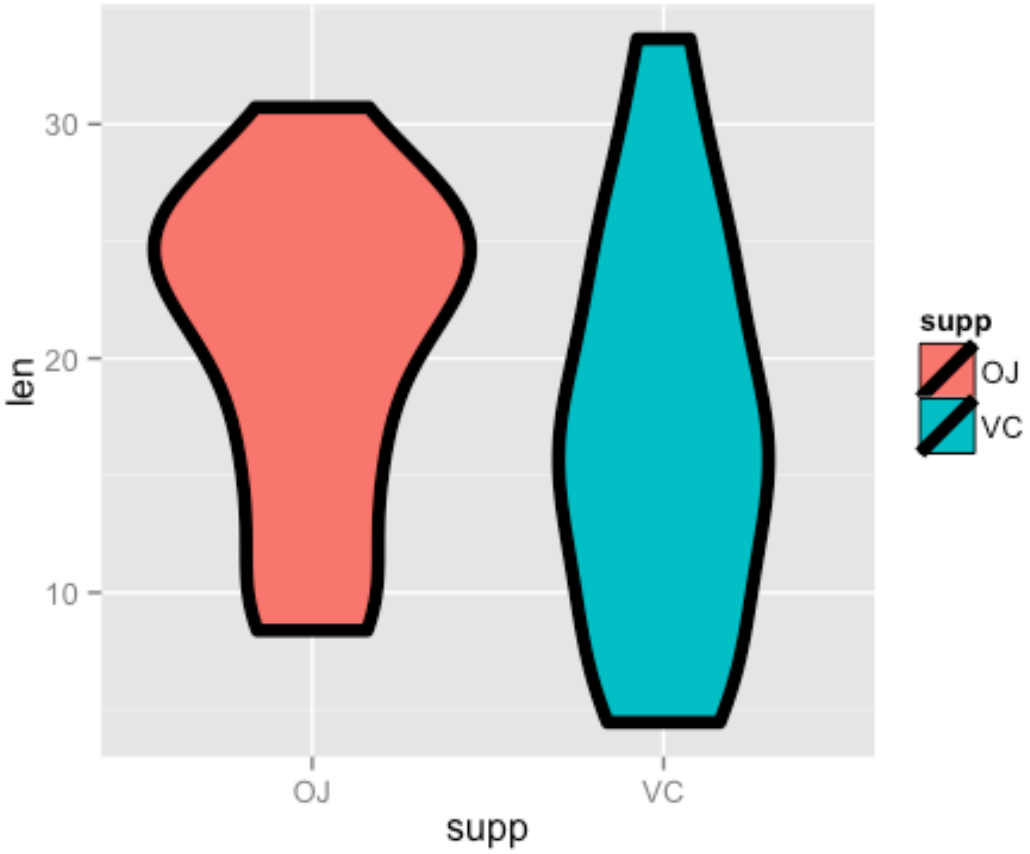
Using the example from running ?ToothGrowh in R, I have the following plots:



ToothGrowth data: length vs dose, given type of supplement

Based on these preliminary plots, it seems different delivery methods do have impact on the tooth growth. Also it is clear that amount of dosage does have an impact on tooth growth as well albeit there seems a diminishing effect while delivering through orange juice.

I am using the violin plot to review the dataset again and we can observe the wider variance for ascorbic acid, VC, delivery method versus orange juice, OJ.



Data Preparation I slice the dataset in various ways. First I create two data frames. Each has the same delivery method but different dosage. I then create three subsets for each delivery method with different dosage: 0.5 vs 1.0, 1.0 vs. 2.0, and 0.5 vs. 2.0. I will be using these six data frames to test the effect of dosage

```
t_supp_vc = ToothGrowth[ToothGrowth$supp == "VC",]
t_supp_oj = ToothGrowth[ToothGrowth$supp == "OJ",]
t_supp_vc_05_10 = t_supp_vc[1:20,]
t_supp_vc_10_20 = t_supp_vc[11:30,]
t_supp_vc_05_20 = rbind(t_supp_vc[1:10,], t_supp_vc[21:30,])
t_supp_oj_05_10 = t_supp_oj[1:20,]
t_supp_oj_10_20 = t_supp_oj[11:30,]
t_supp_oj_05_20 = rbind(t_supp_oj[1:10,], t_supp_oj[21:30,])
```

Again I create three data frames from the orignal dataset by slicing the amount of dosage this time. I will be using these three datasets to test whether it is statistical significant when using different delivery methods

```
t_supp_05 = ToothGrowth[ToothGrowth$dose == 0.5,]
t_supp_10 = ToothGrowth[ToothGrowth$dose == 1.0,]
t_supp_20 = ToothGrowth[ToothGrowth$dose == 2.0,]
```

The Analysis

I am planning on using $\alpha = 5\%$ for all my analysis. My null hypothesis for all the analysis is assuming there is no impact ($H_0: \mu = 0$).

The Effect from Dosage

I decide to use t-test due to the small sample size. For VC delivery method, I have:

```
t.test(len ~ factor(dose), t_supp_vc_05_10, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by factor(dose)
## t = -7.4634, df = 17.862, p-value = 6.811e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.265712  -6.314288
## sample estimates:
## mean in group 0.5   mean in group 1
##              7.98             16.77
```

```
t.test(len ~ factor(dose), t_supp_vc_10_20, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by factor(dose)
## t = -5.4698, df = 13.6, p-value = 9.156e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.054267  -5.685733
## sample estimates:
## mean in group 1 mean in group 2
##           16.77           26.14
```

```
t.test(len ~ factor(dose), t_supp_vc_05_20, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by factor(dose)
## t = -10.388, df = 14.327, p-value = 4.682e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -21.90151 -14.41849
## sample estimates:
## mean in group 0.5   mean in group 2
##              7.98             26.14
```

For OJ delivery, I have:

```
t.test(len ~ factor(dose), t_supp_oj_05_10, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by factor(dose)
## t = -5.0486, df = 17.698, p-value = 8.785e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.415634  -5.524366
## sample estimates:
## mean in group 0.5   mean in group 1
##             13.23             22.70
```

```
t.test(len ~ factor(dose), t_supp_oj_10_20, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by factor(dose)
## t = -2.2478, df = 15.842, p-value = 0.0392
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.5314425 -0.1885575
## sample estimates:
## mean in group 1 mean in group 2
##           22.70           26.06
```

```
t.test(len ~ factor(dose), t_supp_oj_05_20, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by factor(dose)
## t = -7.817, df = 14.668, p-value = 1.324e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -16.335241  -9.324759
## sample estimates:
## mean in group 0.5   mean in group 2
##             13.23             26.06
```

Based on the p-value, clearly the amount of dosage does have an impact on tooth growth regardless of the delivery method although for OJ, we can see it is less effective when comparing dosage 1.0 vs. 2.0, which is consistent to our exploratory plot showing diminishing effect at higher dosage.

Using confidence interval or hypothesis test, we reach the same conclusion as well.

The Effect from Delivery Methods

By controlling the dosage let's compare the effect of delivery methods

```
t.test(len ~ supp, t_supp_05, paired=FALSE, var.equal=FALSE)
```

```
##
##  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
```

```
t.test(len ~ supp, t_supp_10, paired=FALSE, var.equal=FALSE)
```

```
##
##  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, t_supp_20, paired=FALSE, var.equal=FALSE)
```

```
##
##  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
```

This time it is not that clear whether different delivery methods have any impact on tooth growth when higher dosage is being applied. The same conclusion can be derived through p-value, confidence interval or hypothesis test.

Now just to see the impact across the board, I run another test on the original dataset to compare the impact from different delivery methods:

```
t.test(len ~ supp, 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
```

If you review the first plot I draw in **Exploratory Data Analysis**, it gives us the impression that delivery methods may have a statistically significant impact but the test result clearly shows otherwise.

Conclusion

Due to the small sample size and the result from the analysis, I am less confident to argue which delivery methods are more effective although it does show certain level of difference for OJ vs. VC when the dosage is low.

However I am more confident in saying that the amount of vitamin dosage does have an impact on tooth growth given the consistent result from the analysis.