Basic Inferential Data Analysis On Tooth Growth In Vitamin C Supplemented Guinea Pigs

Jon Taylor

September 20, 2016

Overview

We are going to explore a dataset where a scientist compared how dosage (0.5, 1.0, 2.0) and form of Vitamin C (supplement, orange juice) affected tooth length in guinea pigs.

Summary

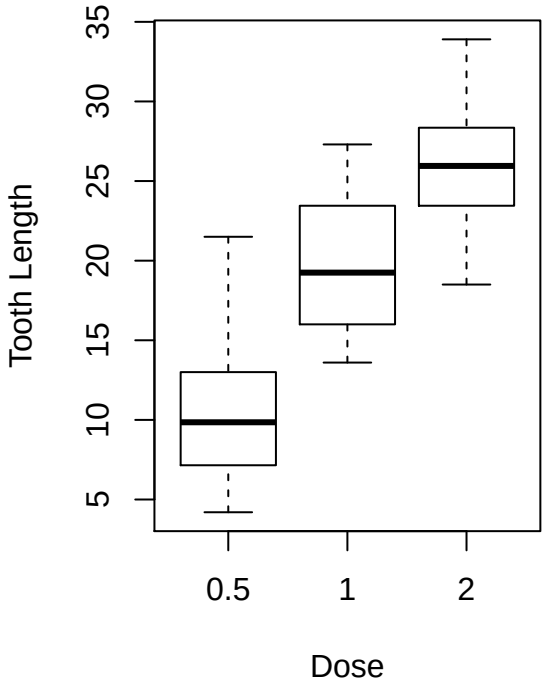
First lets load the data and see how Tooth length is ditributed based on dose and form of Vitamin C.

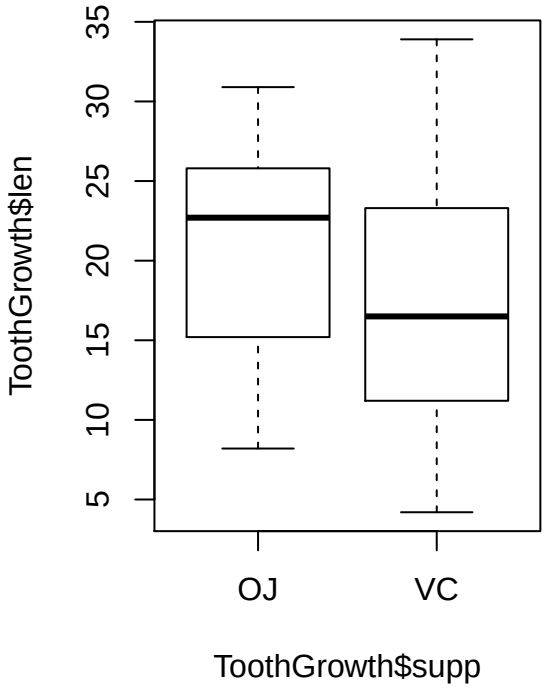
```
#1. Load the ToothGrowth data and perform some basic exploratory data analyses

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
```

```
#Plot length vs dose and length vs supp
par("mfcol"=c(1,2))
boxplot(ToothGrowth$len ~ ToothGrowth$dose, xlab="Dose", ylab="Tooth Length")
plot(ToothGrowth$len ~ ToothGrowth$supp)
```






```
#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
```

```
mean(ToothGrowth$len[ToothGrowth$supp == "OJ"])
```

```
## [1] 20.66333
```

```
sd(ToothGrowth$len[ToothGrowth$supp == "OJ"])
```

```
## [1] 6.605561
```

```
mean(ToothGrowth$len[ToothGrowth$supp == "VC"])
```

```
## [1] 16.96333
```

```r
sd(ToothGrowth$len[ToothGrowth$supp == "VC"])
```

```
## [1] 8.266029
```

T-Tests

The box plots show that tooth length may be affected by both dose and form of Vitamin C. We should use a t-test to see if thats true. Let's start with the form of Vitamin C.

```r
#3. Use confidence intervals and/or hypothesis tests to compare tooth growth #by supp and dose. (Only u

t.test(ToothGrowth$len~ToothGrowth$supp)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len by ToothGrowth$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
```

The t-test shows a p-value of just over 0.06 which means we cannot reject the null hypothesis at a 95% confidence interval. However, it is close, so further experimentation may be worth pursuing.

Now let's do the same for the dose levels.

```r
t.test(ToothGrowth$len[ToothGrowth$dose==2], ToothGrowth$len[ToothGrowth$dose==0.5])
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose == 0.5]
## 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:
##  12.83383 18.15617
## sample estimates:
## mean of x mean of y
##    26.100    10.605
```

```r
t.test(ToothGrowth$len[ToothGrowth$dose==2], ToothGrowth$len[ToothGrowth$dose==1])
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose == 1]
```

```
## 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:
##  3.733519 8.996481
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

```
t.test(ToothGrowth$len[ToothGrowth$dose==1], ToothGrowth$len[ToothGrowth$dose==0.5])
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 1] and ToothGrowth$len[ToothGrowth$dose == 0.5]
## 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:
##   6.276219 11.983781
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

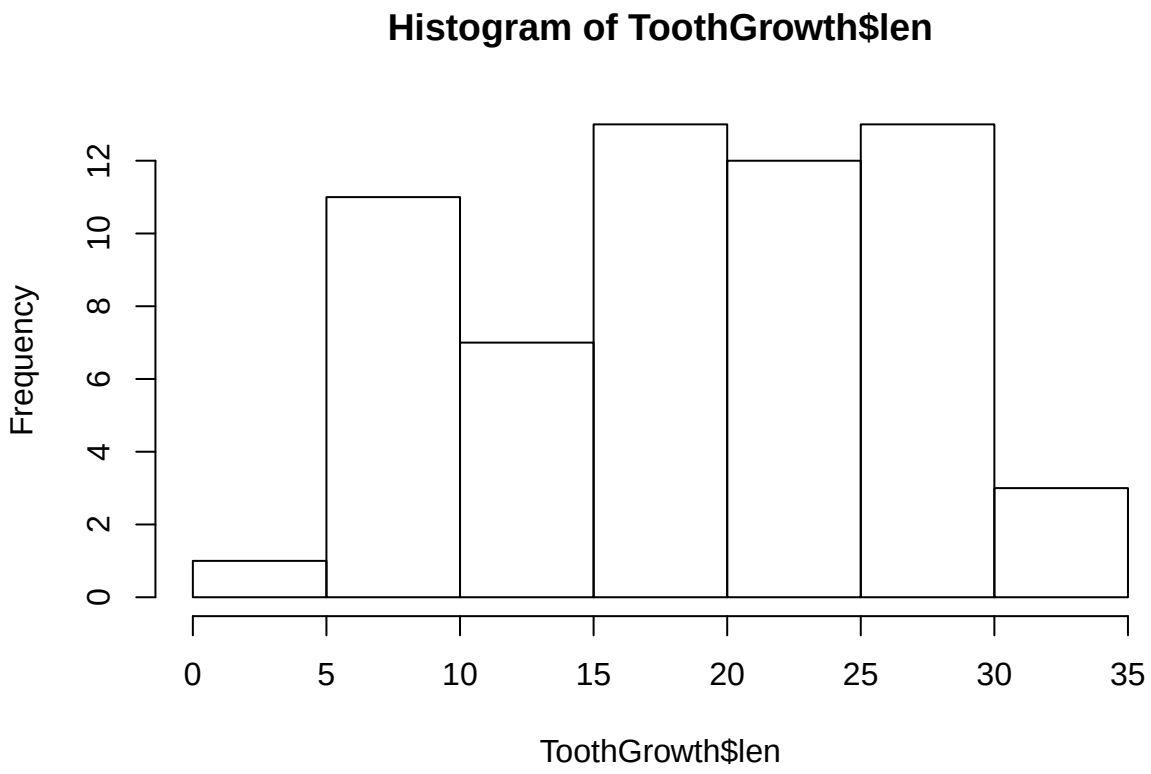
The results show that all of the dosage levels are different than one another. This leads us to reject the null hypothesis that Vitamin C dosage has no effect on tooth length at the 95% confidence interval.

Assumptions

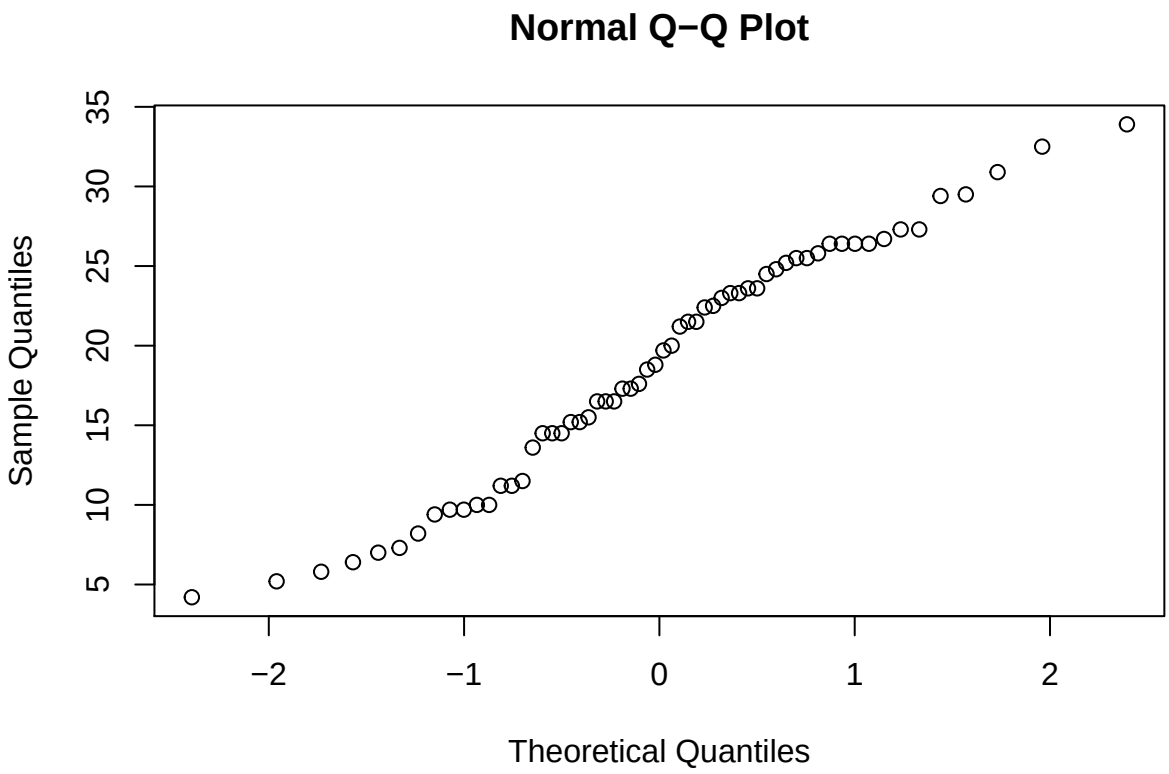
The independent two group t-test relies on assumptions that the dependent variable is normally distributed which we will test below.

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

#Assumption of normality of the dependent variable
par("mfcol"=c(1,1))
hist(ToothGrowth$len)
```




```
qqnorm(ToothGrowth$len) #looks normal
```



The results show that the distribution is appears to be roughly normally distributed.

The independent two group t-test also relies on the assumption that the independent variables have a homogeneaity of variances. We can test this using bartlett's test.

```
bartlett.test(len ~ interaction(supp,dose), data=ToothGrowth)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  len by interaction(supp, dose)
## Bartlett's K-squared = 6.9273, df = 5, p-value = 0.2261
```

The results of the bartlett test show that the independent variables have similar variances.

Conclusions

Having tested our assumptions of the independent two group t-test, we can conclude that Tooth Length is affected by Vitamin C dosage, but not by the form in which the Vitamin C is consumed.