Statistical Inference Course Project part 2

Anton Titov

5-April-2016

1. Overview

This report will cover the analysis of exponential distribution. We will compare parameters of sample distribution and theoretical one.

Assgiment

Load the ToothGrowth data and perform some basic exploratory data analyses Provide a basic summary of the data. 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) State your conclusions and the assumptions needed for your conclusions. Some criteria that you will be evaluated on

1.1 Basis for the study

The data for the analysis include 60 observations of length of tooth of guinea pigs. Observations differ by dose of vitamin C (0.5, 1, 2 mg) and delivety method (orange juice or ascorbic acid)

1.2 Environment

To make this report reprodusible we should inform about R enviroment.
We are loading `data.table` for fast data handling and `ggplot2` . Also we change system time display to English by `Sys.setlocale` .

```
Sys.setlocale("LC_TIME", "English")
```

```
## [1] "English_United States.1252"
```

```
library(data.table)
library(ggplot2)
library(datasets)
set.seed(1234521)

sessionInfo()
```

```
## R version 3.1.3 (2015-03-09)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 7 x64 (build 7601) Service Pack 1
##
## locale:
## [1] LC_COLLATE=Russian_Russia.1251     LC_CTYPE=Russian_Russia.1251
## [3] LC_MONETARY=Russian_Russia.1251    LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] ggplot2_2.1.0    data.table_1.9.4
##
## loaded via a namespace (and not attached):
##  [1] chron_2.3-45     colorspace_1.2-6 digest_0.6.8     evaluate_0.8.3
##  [5] formatR_1.3      grid_3.1.3       gtable_0.2.0     htmltools_0.3.5
##  [9] knitr_1.12.3     munsell_0.4.3    plyr_1.8.2       Rcpp_0.11.6
## [13] reshape2_1.4.1   rmarkdown_0.9.5  scales_0.4.0     stringr_0.6.2
## [17] tools_3.1.3      yaml_2.1.13
```

https://rpubs.com/daniambrosio/tooth_growth_exploratory_data_analysis (https://rpubs.com/daniambrosio/tooth_growth_exploratory_data_analysis)

2. Data Summary and Exploratory Data Analysis

Now we perform basic analysis of data

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

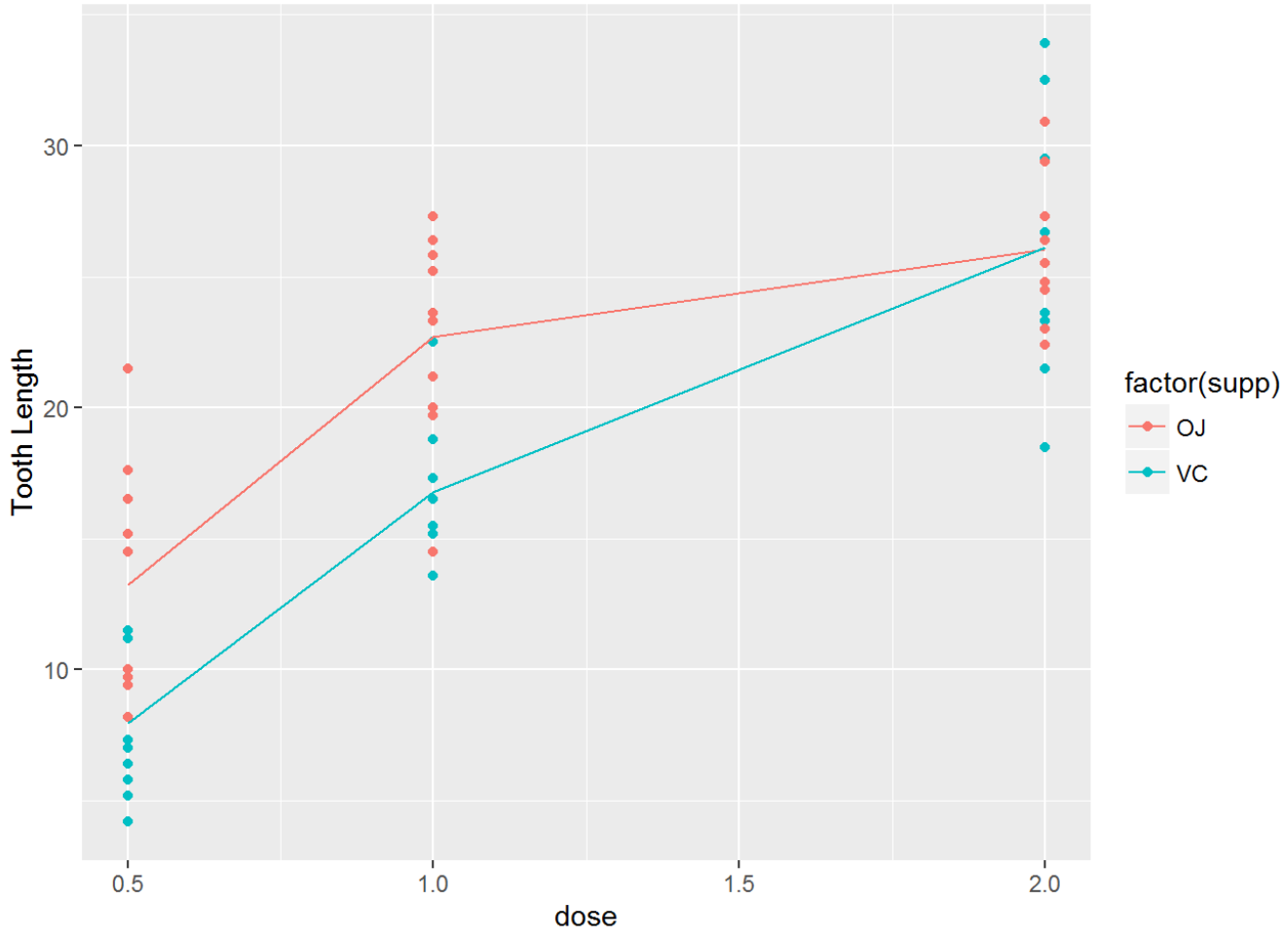
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

We can see that `supp` variable is factor of 2 values: OJ and VC.

2.1 Visual Analysis

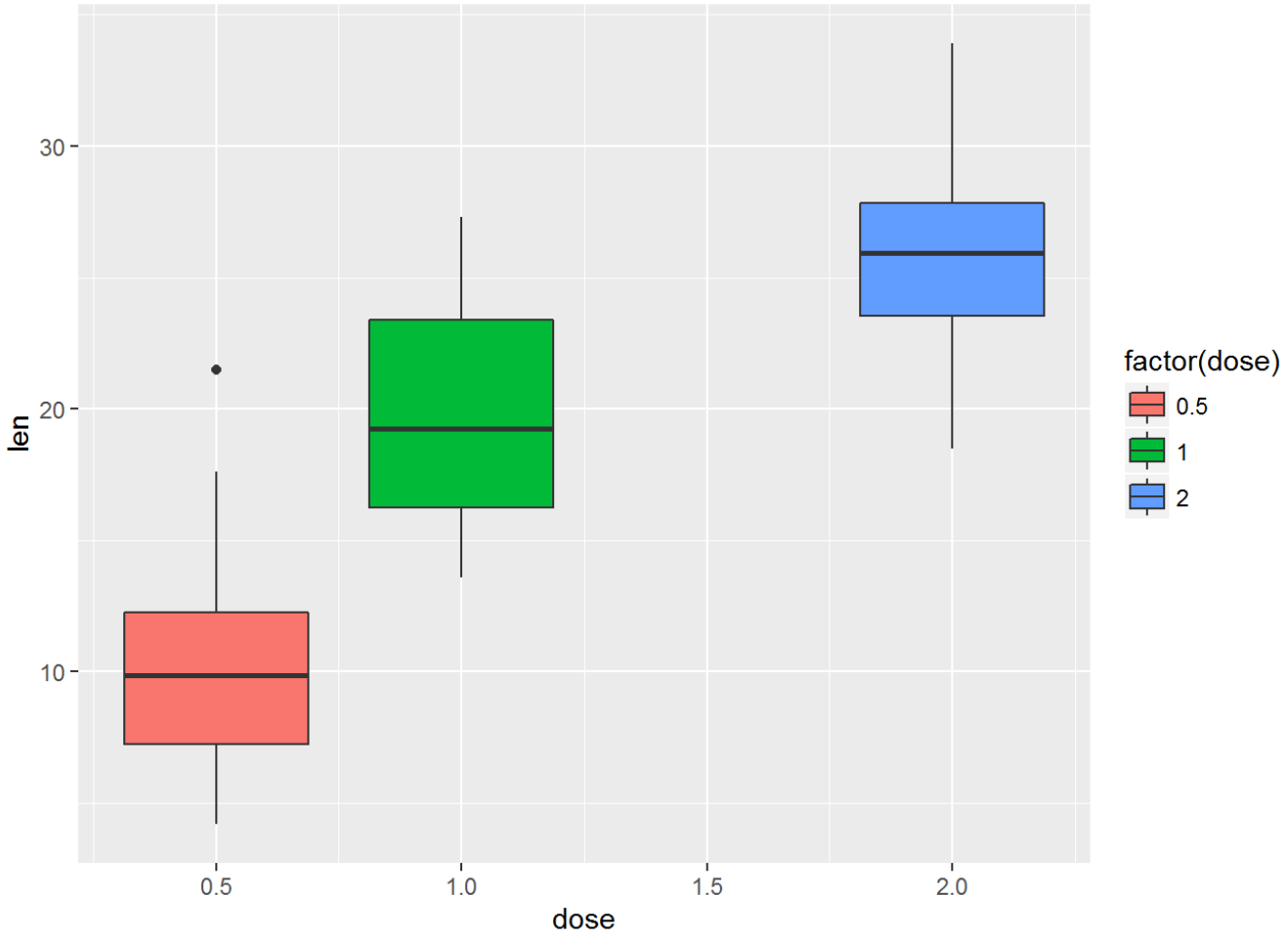
Let us plot mean values of length by dosage.

```
means <- aggregate(len~., ToothGrowth, mean)
g = ggplot(data=data.frame(ToothGrowth), aes(x=dose,y=len))
g = g + geom_point(aes(color=factor(supp))) + geom_line(data=means, aes(color = supp))
g = g + labs(y="Tooth Length")

print(g)
```



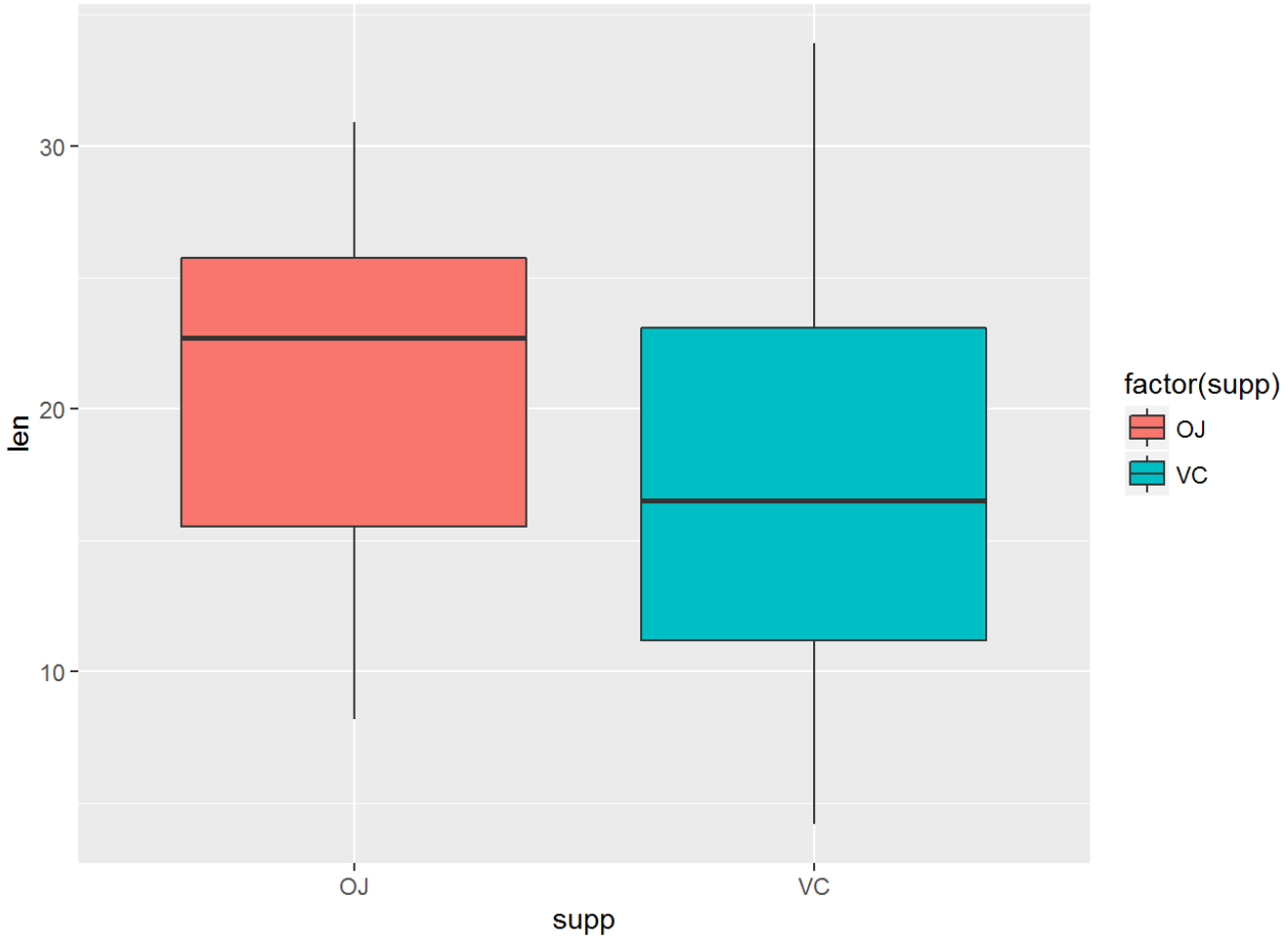
Next the same data will be shown as boxplots.

```
g = ggplot(data=data.frame(ToothGrowth), aes(x=dose,y=len))
g = g + geom_boxplot(aes(fill=factor(dose)))

print(g)
```



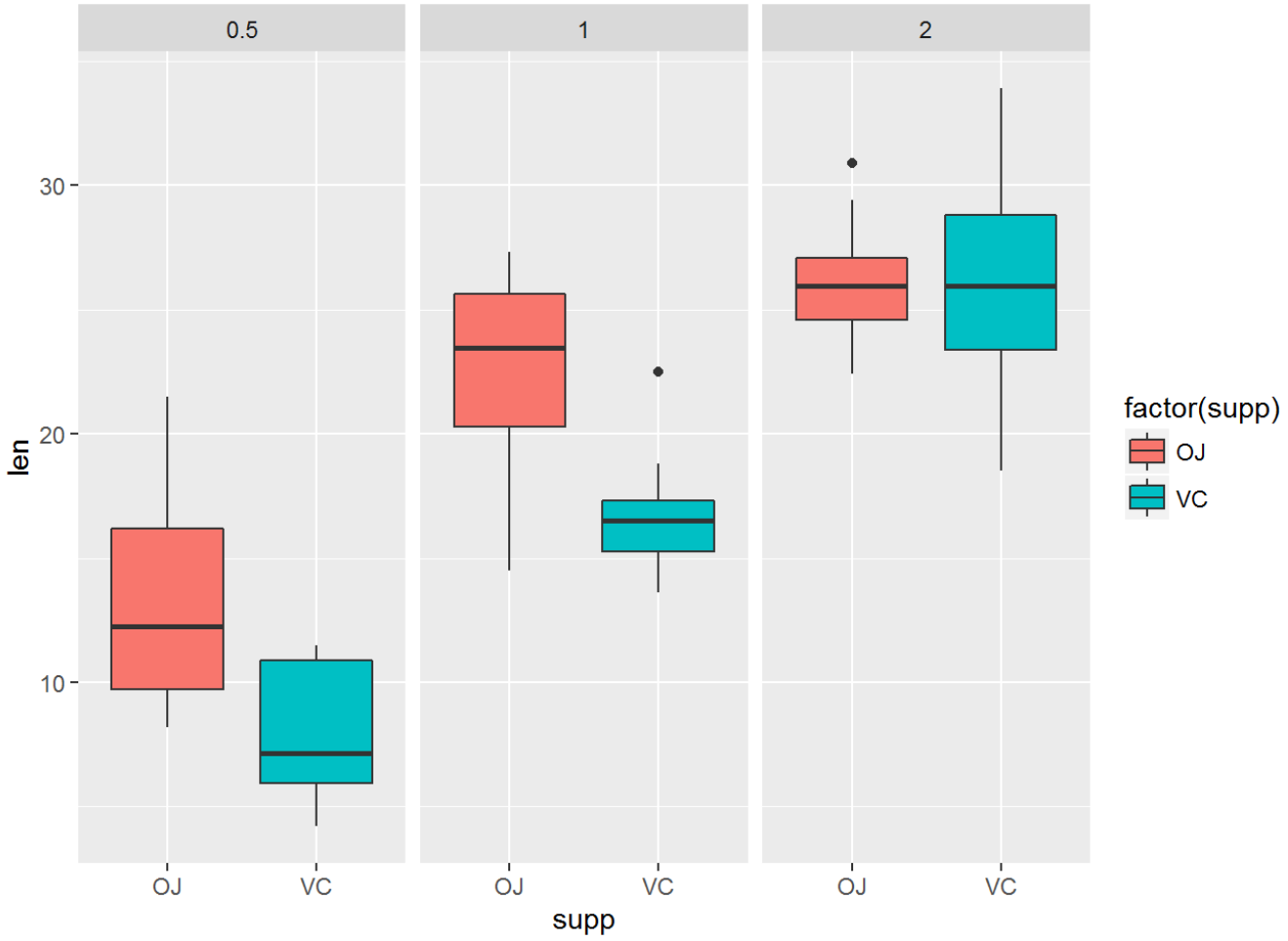
Then we explore possible relation between tooth length and delivery method.

```
g = ggplot(data=data.frame(ToothGrowth), aes(x=supp,y=len))
g = g + geom_boxplot(aes(fill=factor(supp)))

print(g)
```



Then we explore possible relation between tooth length and delivery method considering dosage.

```
g = ggplot(data=data.frame(ToothGrowth), aes(x=supp,y=len))
g = g + geom_boxplot(aes(fill=factor(supp))) + facet_wrap(~dose)

print(g)
```



Let us conclude visual explaratory analysis. We can see some correlation between dosage and toothgrowth: the more dosage the more tooth length. However we can not say the same about delivery method. Depending on dosage delivery method affects to tooth length in opposite ways.

2.2 Numeric Analysis

```
by(ToothGrowth, INDICES = list(ToothGrowth$supp, ToothGrowth$dose), summary)
```

```
## : OJ
## : 0.5
##       len       supp         dose
##  Min.   : 8.2   OJ:10   Min.   :0.5
##  1st Qu.: 9.7   VC: 0   1st Qu.:0.5
##  Median :12.2           Median :0.5
##  Mean   :13.2           Mean   :0.5
##  3rd Qu.:16.2           3rd Qu.:0.5
##  Max.   :21.5           Max.   :0.5
## ------------------------------------------------------------
## : VC
## : 0.5
##       len        supp         dose
##  Min.   : 4.20   OJ: 0   Min.   :0.5
##  1st Qu.: 5.95   VC:10   1st Qu.:0.5
##  Median : 7.15           Median :0.5
##  Mean   : 7.98           Mean   :0.5
##  3rd Qu.:10.90           3rd Qu.:0.5
##  Max.   :11.50           Max.   :0.5
## ------------------------------------------------------------
## : OJ
## : 1
##       len       supp         dose
##  Min.   :14.5   OJ:10   Min.   :1
##  1st Qu.:20.3   VC: 0   1st Qu.:1
##  Median :23.4           Median :1
##  Mean   :22.7           Mean   :1
##  3rd Qu.:25.6           3rd Qu.:1
##  Max.   :27.3           Max.   :1
## ------------------------------------------------------------
## : VC
## : 1
##       len       supp         dose
##  Min.   :13.6   OJ: 0   Min.   :1
##  1st Qu.:15.3   VC:10   1st Qu.:1
##  Median :16.5           Median :1
##  Mean   :16.8           Mean   :1
##  3rd Qu.:17.3           3rd Qu.:1
##  Max.   :22.5           Max.   :1
## ------------------------------------------------------------
## : OJ
## : 2
##       len       supp         dose
##  Min.   :22.4   OJ:10   Min.   :2
##  1st Qu.:24.6   VC: 0   1st Qu.:2
##  Median :25.9           Median :2
##  Mean   :26.1           Mean   :2
##  3rd Qu.:27.1           3rd Qu.:2
##  Max.   :30.9           Max.   :2
## ------------------------------------------------------------
## : VC
## : 2
##       len       supp         dose
```

```
##  Min.   :18.5   OJ: 0   Min.   :2
##  1st Qu.:23.4   VC:10   1st Qu.:2
##  Median :25.9           Median :2
##  Mean   :26.1           Mean   :2
##  3rd Qu.:28.8           3rd Qu.:2
##  Max.   :33.9           Max.   :2
```

3. Confidence Intervals and Hypothesis Testing

Since there are three different type of dosage. But T-Test is applied to two group at a time. Therefore, three data set has been created with two different dosage at a time. Our Null hypothesis for each dataset there is no significant effect of dosage on lenght of theth. Which can be interpreted there sample means are equal.

3.1 Dosage as a Factor

```
Tooth051 <- subset(ToothGrowth, dose %in% c(0.5,1) )
t.test(len~dose,data = Tooth051 , paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -6.477, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.984  -6.276
## sample estimates:
## mean in group 0.5   mean in group 1
##             10.61             19.73
```

Confidense interval between dose `0.5` and `1` equals to -11.9837, -6.2763. That leads to the conclusion that dosage increase in dosage from `0.5` to `1` affects directly to tooth length.

```
Tooth12 <- subset(ToothGrowth, dose %in% c(1,2) )
t.test(len~dose, data = Tooth12, paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -4.901, df = 38, p-value = 0.00001811
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.994 -3.736
## sample estimates:
## mean in group 1 mean in group 2
##           19.73           26.10
```

Confidense interval between dose `1` and `2` equals to -8.9944, -3.7356. That leads to the conclusion that dosage increase in dosage from `1` to `2` affects directly to tooth length.

```
Tooth052 <- subset(ToothGrowth, dose %in% c(0.5,1) )
t.test(len~dose, data = Tooth052, paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -6.477, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.984  -6.276
## sample estimates:
## mean in group 0.5   mean in group 1
##             10.61             19.73
```

Confidense interval between dose `0.5` and `2` equals to -11.9837, -6.2763. That leads to the conclusion that dosage increase in dosage from `0.5` to `2` affects directly to tooth length.

3.2 Supplement as a Factor

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

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.915, df = 55.31, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.171  7.571
## sample estimates:
## mean in group OJ mean in group VC
##            20.66            16.96
```

Confidence interval for difference of the means of different delivery methods includes zero: -0.171, 7.571. So we conclude no correlation between delivery method and tooth growth.

4. Conclusions and Assumptions

4.1 Assumptions

1. The poplulations are independent, the variances between populations are different and a random population was used
2. The population was comprised of similar guinea pigs, measurement error was accounted for with significant digits, and double blind research methods were used. This means researchers were unaware which pigs were taking specific dosage using specific delivery method.
3. For the populations to be independent, 60 guinea pigs would have to be used so each combination of dose level and delivery method were not affected by the other methods.

4.2 Conclusions

1. Increasing the dosage leads to increased tooth growth.
2. Delivery method has no relation on tooth growth.