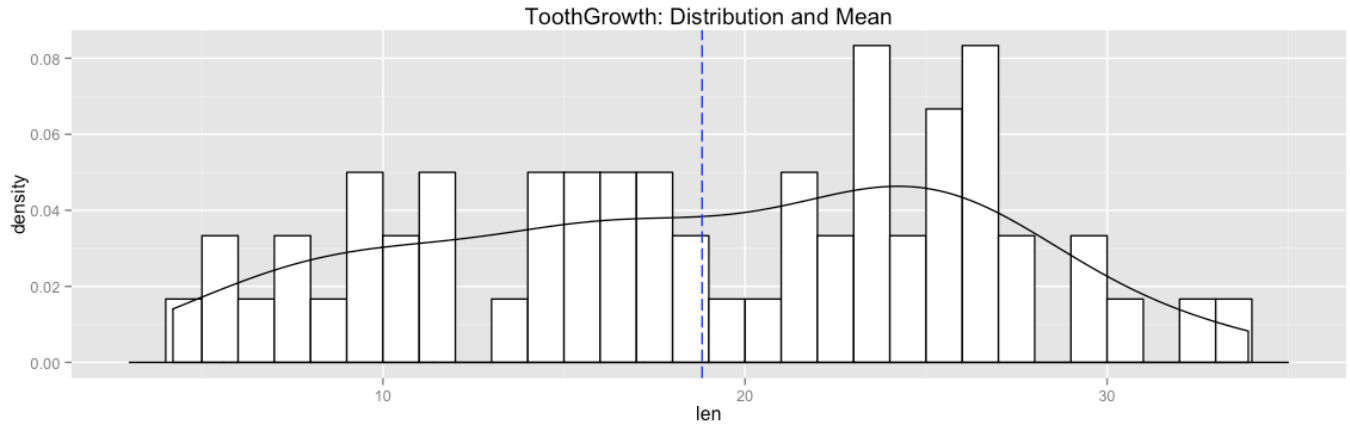
Statistical Inference: Tooth Grow

Ariel Lev

26. April 2015

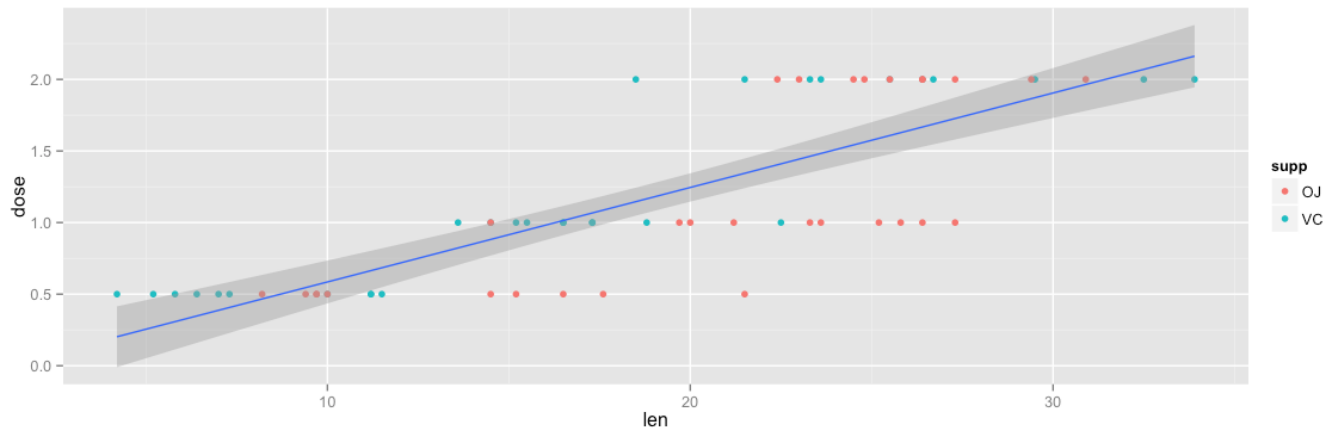
Basic summary and exploratory analysis

```
data(ToothGrowth)
```



We can see that the population is continuous, fairly distributed in a mound shape, unskewed, and that the number of observations is relatively small. That satisfies the basic prerequisites for using t tests. Because the observations are subject independent, so we cannot directly say which observation refers to which subject, unpaired test will be conducted. Because the data was collected from the same exact population, equal variance are assumed. The tests will be performed with a confidence interval of 95%.

To gather a bit more intution, lets explore the dependency of the variables.



Intuitively, it seems like **dose level** and **thooth length** are positively correlated. Lets see what the tests have to say regrading that assumption.

What is the difference in tooth length by supplement, holding the dose constant?

Let us check the null hypothesis, stating there is no significant difference between the two methods (orange juice or ascorbic acid), vs the alternative hypothesis, stating that there is such.

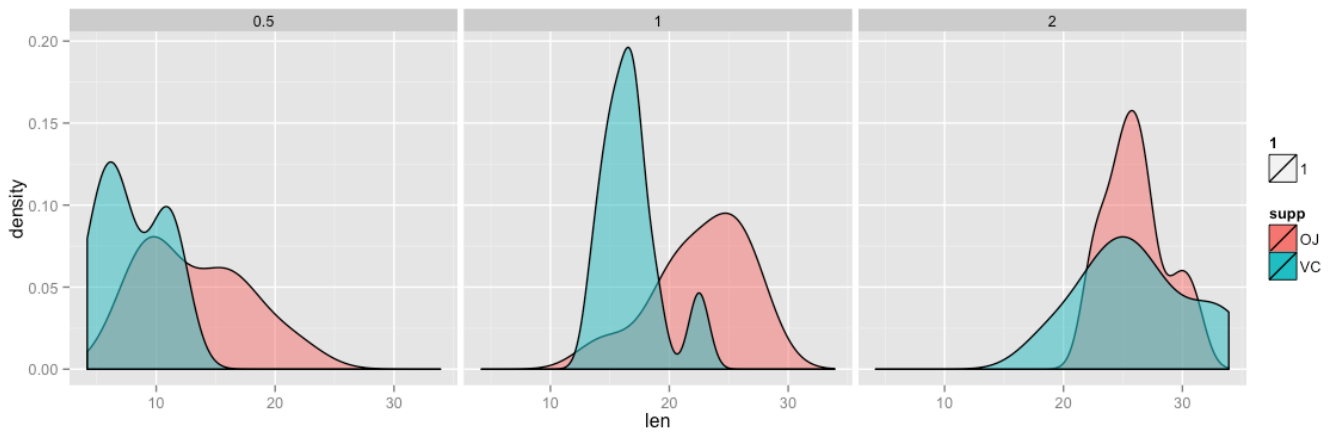
```
# splitting across dose level
dose_split = split(ToothGrowth, ToothGrowth$dose)

# performing tests
# extracting t-statistic and p values out
p_values <- sapply(dose_split, function(x) { round(t.test(len ~ supp, paired =
F, var.equal = T, data = x)$p.value, 4)})
p_values
```

```
##    0.5      1      2
## 0.0053 0.0008 0.9637
```

From the results of the p-values above we can conclude, that **it is only for dose level 2.0 that we fail to reject the null hypothesis**, i.e the supplement method indeed yields **significantly longer teeth for dose level 0.5 or 1**.

Finally, we can confirm the results of the p-values by summarizing the sample means, and by taking a look of how the population is distributed among the different dose level.

```
## Source: local data frame [6 x 3]
## Groups: dose
##
##   dose supp mean(len)
## 1  0.5   OJ     13.23
## 2  0.5   VC      7.98
## 3  1.0   OJ     22.70
## 4  1.0   VC     16.77
## 5  2.0   OJ     26.06
## 6  2.0   VC     26.14
```



What is the difference in tooth length by dose, holding the supplement type constant?

```
# splitting across supplement methods
supp_split = split(ToothGrowth, ToothGrowth$supp)

# all-pairs of possible dose combinations
comb <- combn(c(0.5,1,2), 2)
comb
```

```
##      [,1] [,2] [,3]
## [1,]  0.5  0.5    1
## [2,]  1.0  2.0    2
```

```
p_values <- sapply(supp_split, function(x) {
  tests = NULL
  for (i in 1:ncol(comb)) {
      y <- x[x$dose %in% comb[,i], ]
      tests <- c(tests, round(t.test(len ~ dose, paired = F, var.equal = T, dat
a = y)$p.value, 7))
  }
  tests
  })

row.names(p_values) <- c("[0.5, 1]", "[0.5, 2]", "[1, 2]")
p_values
```

```
##                  OJ       VC
## [0.5, 1] 0.0000836 6.0e-07
## [0.5, 2] 0.0000003 0.0e+00
## [1, 2]   0.0373628 3.4e-05
```

From the results of the p-values above we can conclude, that **the null hypothesis should be rejected for any dose level**, i.e there are significant differnces between subjects who received different amounts of vitamin C. Finally, we can confirm the results of the p-values by summarizing the sample means, and by taking a look of how the population is distributed among the different dose level.

```
## Source: local data frame [6 x 3]
## Groups: supp
##
##   supp dose mean(len)
## 1   OJ  0.5     13.23
## 2   OJ  1.0     22.70
## 3   OJ  2.0     26.06
## 4   VC  0.5      7.98
## 5   VC  1.0     16.77
## 6   VC  2.0     26.14
```