Statistical Inference course project - The Effect of Vitamin C on Tooth Growth in Guinea Pigs

Evgeniy Zabrodskiy

November, 2015

Overview

The goal of this analysis is to explore the ToothGrowth data, make assumptions and apply statistical inference methods to confirm or reject those assumptions.

The dataset ToothGrowth contains the length of odontoblasts (teeth) in each of 60 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 analysis

Basic summary of the data

ToothGrowth dataframe structure:

```
## '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 there are three variables:
- len contains the length of the odontoblasts in microns;
- supp is a factor variable showing the type of the supplement containing Vitamin C ("OJ" == Orange Juice, "VC" == Ascrobic Acid);
- dose stores the dose of Vitamin C in miligrams. It will be converted to factor for convenience.

Dataframe summary:
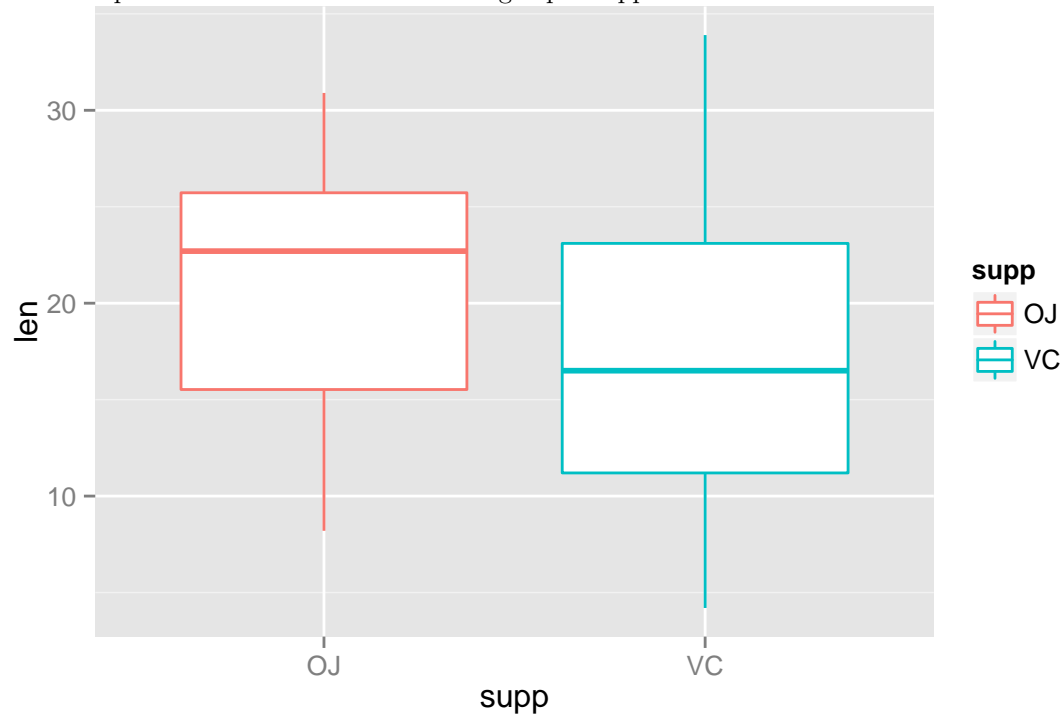
```
##       len          supp       dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

Two first lines of the data:

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
```

Exploratory plots

This boxplot shows the distribution of length per supp.

This plot raises a question whether the length of odontoblasts of pigs receiving Vitamin C with orange juice is not equal to the mean length in the group of pigs which recieved Vitamin C as ascrobic acid, not taking into account different doses. But it would be better to divide the boxes by dose to make better assumptions.

The boxplot below shows the distribution of length for each supp per dose:
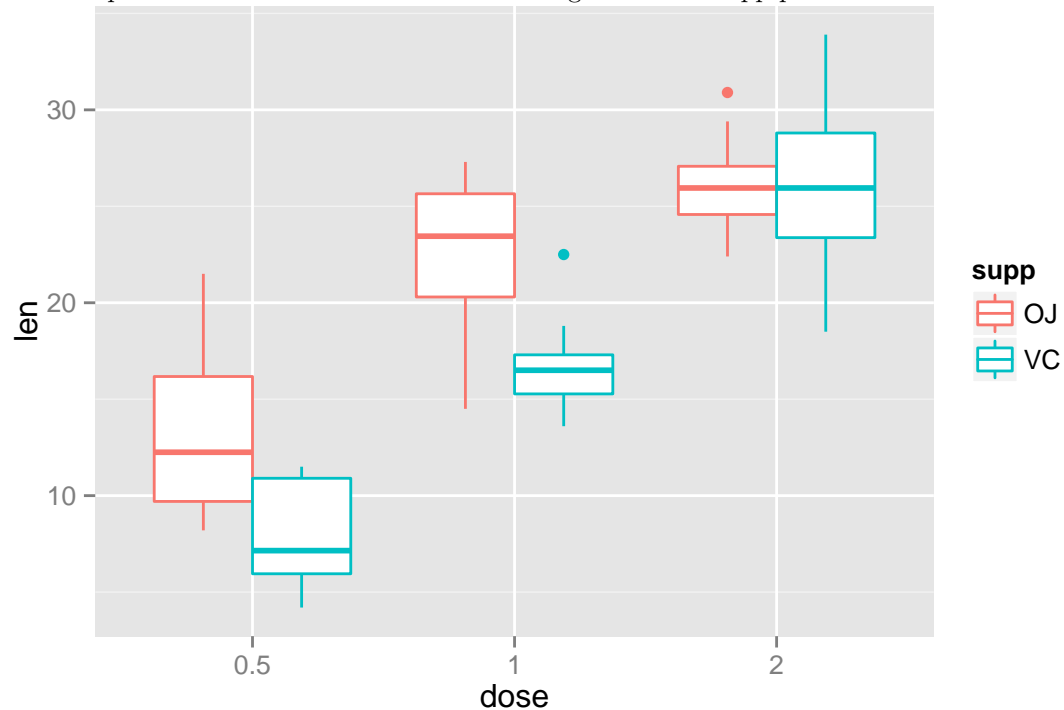
This boxplot shows some more details based on which we make the following assumptions:

Assumption 1: Mean length of odontoblasts is different for group of pigs which received Vitamin C with orange juice and group of pigs which recieved Vitamin C as ascrobic acid, dose 0.5 mg.

Assumption 2: Mean length of odontoblasts is different for group of pigs which received Vitamin C with orange juice and group of pigs which recieved Vitamin C as ascrobic acid, dose 1 mg.

Assumption 3: There is no significant difference in mean lengths for both supplements with dose 2 mg.

The boxplot below shows the distribution of length for each dose per supp:
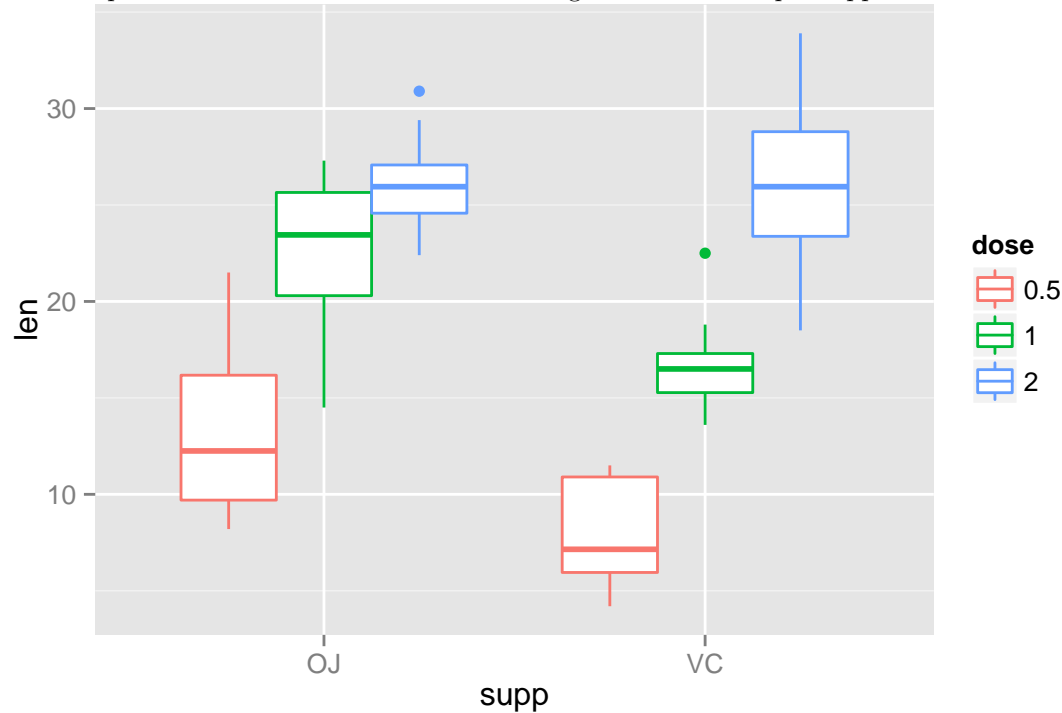
Assumption 4: Mean length of odontoblasts is different for subgroups with doses 0.5 mg and 1 mg in group which recieved Vitamin C with orange juice.

There are more assumptions that can be made but they are all similar and are tested the same way.

Testing assumptions and conclusions

Talking about lengths of odontoblasts we may assume that the lengths population is normally distributed even though we cannot clearly see that from the sample data. Based on this assumption and taking into account that the sample sizes of the groups we are going to compare is quite small (10 elements each), we should use Student's T-test to confirm or reject our assumptions. Also it is assumed that the variances of the populations of tested samples are different since all the pigs belong to the same spicies and the study is randomized.

Assumption 1: Mean length of odontoblasts is different for group of pigs which received Vitamin C with orange juice and group of pigs which recieved Vitamin C as ascrobic acid, dose 0.5 mg.

$H_0 : \mu_{oj,0.5} = \mu_{vc,0.5}$
$H_1 : \mu_{oj,0.5} \neq \mu_{vc,0.5}$

```
##
##  Two Sample t-test
##
## data:  len.VC.05 and len.OJ.05
```

```
## t = -3.1697, df = 18, p-value = 0.005304
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.729738 -1.770262
## sample estimates:
## mean of x mean of y
##      7.98     13.23
```

From the T-test output we can see that p-value = 0.0053037 which represents the probability of H_0 being true and since it is less than $\alpha = 0.05$ for a two-sided test we may reject H_0. Also the confidence interval which does not contain zero can be used to make the same conclusion. It means that with 95% probability $(1 - \alpha)$ the true difference of the means of the populations will be within the borders of the interval.

Conclusion 1: Assumption 1 is confirmed: mean length of odontoblasts is different for the groups of pigs which recieved 0.5 mg dose of Vitamin C in forms of orange juice and ascorbic acid.

Assumption 2: Mean length of odontoblasts is different for group of pigs which received Vitamin C with orange juice and group of pigs which recieved Vitamin C as ascrobic acid, dose 1 mg.

$H_0 : \mu_{oj,1} = \mu_{vc,1}$
$H_1 : \mu_{oj,1} \neq \mu_{vc,1}$

```
##
##  Two Sample t-test
##
## data:  len.VC.1 and len.OJ.1
## t = -4.0328, df = 18, p-value = 0.0007807
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.019308 -2.840692
## sample estimates:
## mean of x mean of y
##     16.77     22.70
```

From the T-test output we can see that p-value = 7.8072617×10^{-4} which represents the probability of H_0 being true and since it is less than $\alpha = 0.05$ for a two-sided test we may reject H_0.

Conclusion 2: Assumption 2 is confirmed: mean length of odontoblasts is different for the groups of pigs which recieved 1 mg dose of Vitamin C in forms of orange juice and ascorbic acid.

Assumption 3: There is no significant difference in mean lengths for both supplements with dose 2 mg.

$H_0 : \mu_{oj,2} = \mu_{vc,2}$
$H_1 : \mu_{oj,2} \neq \mu_{vc,2}$

```
##
##  Two Sample t-test
##
## data:  len.VC.2 and len.OJ.2
## t = 0.046136, df = 18, p-value = 0.9637
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.562999  3.722999
## sample estimates:
## mean of x mean of y
##     26.14     26.06
```

From the T-test output we can see that p-value $= 0.9637098$ which represents the probability of H_0 being true. This pobability is so high, so we cannot reject the H_0 for sure.

Conclusion 3: Assumption 3 is confirmed: mean length of odontoblasts is the same for the groups of pigs which recieved 2 mg dose of Vitamin C in forms of orange juice and ascorbic acid.

Assumption 4: Mean length of odontoblasts is different for subgroups with doses 0.5 mg and 1 mg in group which recieved Vitamin C with orange juice.

$H_0 : \mu_{oj,0.5} = \mu_{oj,1}$
$H_1 : \mu_{oj,0.5} \neq \mu_{oj,1}$

```
##
##   Two Sample t-test
##
## data:  len.OJ.05 and len.OJ.1
## t = -5.0486, df = 18, p-value = 8.358e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.410814  -5.529186
## sample estimates:
## mean of x mean of y
##     13.23     22.70
```

From the T-test output we can see that p-value $= 8.3575593 \times 10^{-5}$ which represents the probability of H_0 being true and since it is less than $\alpha = 0.05$ for a two-sided test we may reject H_0.

Conclusion 4: Assumption 4 is confirmed: mean length of odontoblasts difference is significant for subgroups of pigs which recieved 0.5 mg and 1 mg dose of Vitamin C with orange juice.