Coursera's Statistical Inference - Data analysis

Thomas Guenther

Sunday, September 14, 2014

Introduction

First i must apologize for the bad style of this report where figures and tables are not on their right place within this document. Instead they are appended to the Appendix. Personally i'am unhappy with this approach. But i have no other choice to keep the **2+3 pages limitation**. Due to this page limit i've setup a GitHub repository with all my files: https://github.com/dreammaster38/coursera_statistical_inverence_project_2

We have to analyze the Tooth Growth data set which comes with R. It contains measurments of the length of ordonblasts (teeth) of ten guinea pigs at each of three dose levels of Vitamin C ascorbic acid with each of two delivery methods: orange juice (OJ) or Vitamin C (VC).

1. Exploratory data analysis

First we will take a look at the ToothGrowth data set and make some plots to see if we can see differences in the tooth length for the two delivery methods (pure ascorbic acid and for orange juice). Therefore we make two box plots. The first one shows the length of teeth for each delivery method without condidering the dose levels.

In Figure 1.1 you can see that the middle 50% of data reaches from ≈ 15.2 to ≈ 26.0 for orange juice and ≈ 11.5 to ≈ 22.7 for ascorbic acid. Furthermore there is a slight left skewness for orange juice and a slight right skewness for ascorbic acid. It looks like we have a difference in teeth length for different supplements. The plot on Figure 1.2 shows the teeth length for different doses of vitamin c. It seems that the dose has an influence of the teeth length. The higher the dose the longer the teeth. There is also an indication that the length of teeth is influenced by the type of ascorbic acid delivered.

2. Basic summary of data

Let's see what we have within our data set. We have 60 observations with 4 variables (one added by me: suppDose). There are 30 observations for the supplement of orange juice (OJ) and also 30 observations for the supplement of pure ascorbic acid (VC) as can be seen by the summary of Table 2.1.

The table Table 2.2 gives us a view about the overall mean and the standard deviation of the teeth length for the two delivery methods.

Here we get an idea how the length of the teeth is influenced by the doses 0.5 mg, 1.0 mg and 2.0 mg for each of the two delivery methods.

What we can say **for each delivery method** at this point is: **the higher the dose of ascorbic acid the longer the teeth.**

3. Compare tooth groth by Hypothesis testing

Test the length of teeth for the two delivery methods In the first test we try to find out if the length of the teeth is different either for vitamin c or orange juice. We will use a two tailed non-paired t-test. We will try to find an answer for the following questions:

- H_0 : The mean of the teeth length is the same for booth delivery methods

- H_A : The mean of the teeth length for VC is different from the length of teeth for OJ

For the R output of this test please refer to Table 3.1.

We will not reject the H_0 hypothesis because the confidence interval at the 95%-level contains 0 and the p-value=0.06 is slightly greater than $\alpha = 0.05$.

Test teeth length for the three given doses of the overall supplements This test should show us if for different doses of ascorbic acid the length of the teeth is also different. This will be shown with an two-sided t-test as well. First we will test the 0.5 mg dose and the 1.0 mg dose and afterwards test the the 1.0 mg dose and the 2.0. Our hypothesis for the means of the teeth is as follows:

- H_0 : The mean of the teeth length is the same for each dose of ascorbic acid
- H_A : The mean of the teeth length is different for each dose of ascorbic acid

For the R output of this test please see Table 3.2 and Table 3.3.

Comparing the 0.5 mg and 1.0 mg doses as stated in Table 3.2 we can see that the p-Value is much smaller than $alpha = 0.05$ and the confidence intervall does not contain 0. So we can to reject our null hypothesis. The same conclusion can be made for the comparison of the 1.0 mg and 2.0 mg samples. There we have to reject also due to very small p-Value and the confidence intervall which not contains 0 (Table 3.3).

4. Conclusion and assumptions

Under the assumption that the samples are random, independent, and come from normally distributed population we can make the following conclusions and assumptions:

- As the first test showed there is no clear difference if orange juice has a better effect on tooth growth than vitamin c. Booth delivery methods have (statistically) the same effect. It looks like that our assumption about the relation between the two dekivery methods made in section one cannot been accepted.
- The second test examined if the dose of an ascorbic acid supplement has an influence on the mean of the teeth length. There we could show that the dose has an influence. A look on the boxplot shows that the higher the dose of ascorbic acid the longer the teeth.

We could do some further testing which dose of which supplement has a better effect but this is lead to the reader...

5. Appendix

In this section sokme R outputs are collected to clarify things.

Figure 1.1

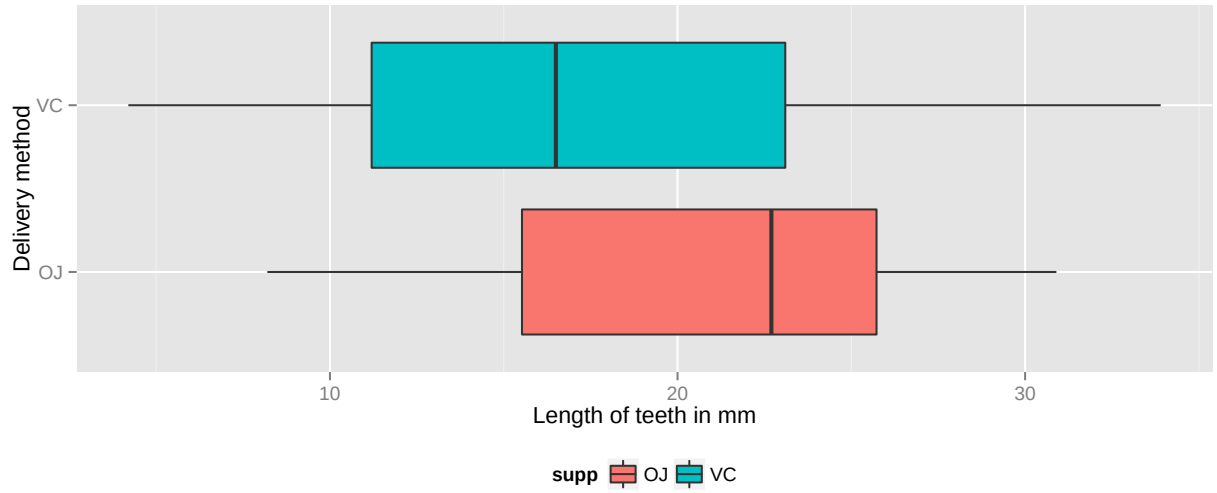

Figure 1.2

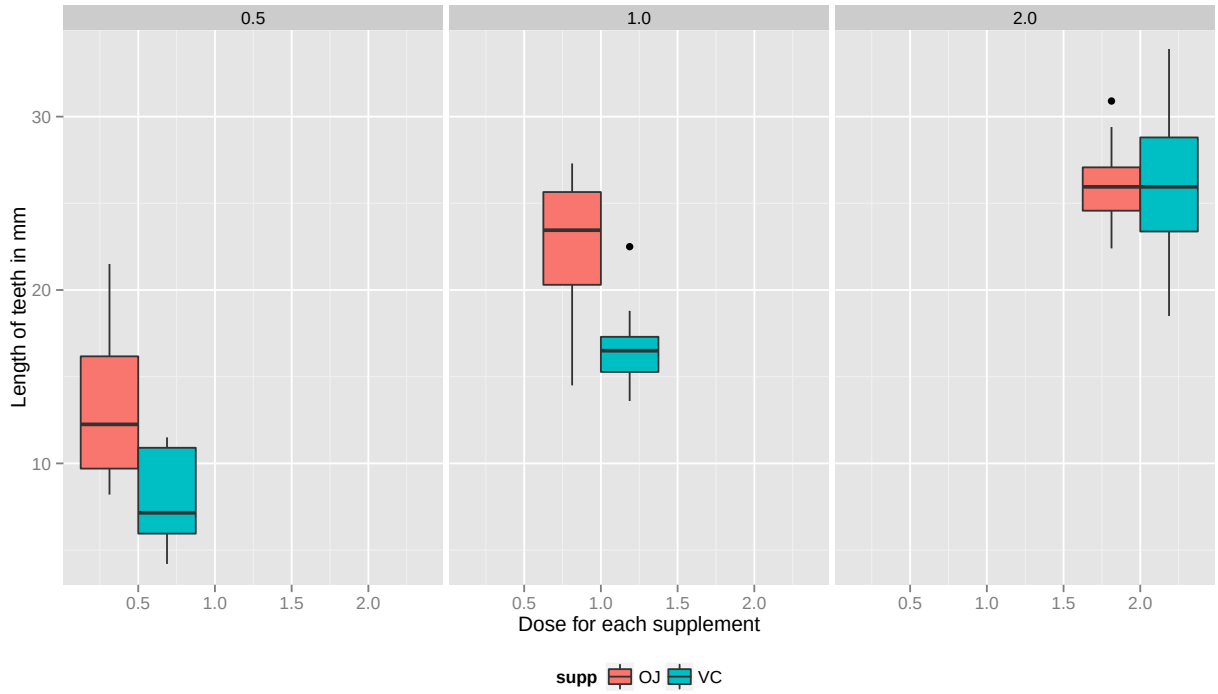

Table 2.1

```
##       len        supp         dose         suppDose
##  Min.   : 4.20   OJ:30   Min.   :0.500   OJ.0.5:10
##  1st Qu.:13.07   VC:30   1st Qu.:0.500   VC.0.5:10
##  Median :19.25           Median :1.000   OJ.1  :10
##  Mean   :18.81           Mean   :1.167   VC.1  :10
##  3rd Qu.:25.27           3rd Qu.:2.000   OJ.2  :10
##  Max.   :33.90           Max.   :2.000   VC.2  :10
```

```
##   supp meanTeethLength sdTeethLength
## 1   OJ        20.66333      6.605561
## 2   VC        16.96333      8.266029
```

Table 2.2

```
##   supp dose meanTeethLength sdTeethLength
## 1   OJ  0.5           13.23      4.459709
## 2   OJ  1.0           22.70      3.910953
## 3   OJ  2.0           26.06      2.655058
## 4   VC  0.5            7.98      2.746634
## 5   VC  1.0           16.77      2.515309
## 6   VC  2.0           26.14      4.797731
```

Table 3.1

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

Table 3.1

Table 3.2

```
##
##  Two Sample t-test
##
## data:  len[dose == 0.5] and len[dose == 1]
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983748  -6.276252
## sample estimates:
```

```
## mean of x mean of y
##    10.605    19.735
```

Table 3.2

Table 3.3

```
##
##  Two Sample t-test
##
## data:  len[dose == 1] and len[dose == 2]
## t = -4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.994387 -3.735613
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```