Statistical Inference Course Project, Part 2: Basic Inferential Data Analysis

Pavitra Balekai

August 22, 2015

Overiew

In this part of the project, we analyze the ToothGrowth data in the R datasets package. The data is set of 60 observations, length of odontoblasts (teeth) in each of 10 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).

```
##       supp
## dose  OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

Exploratory data analysis

The average guinea pig tooth length is 18.813 with a standard deviation of 7.649.

However, we are more interested in finding differences in tooth length across different groups of supplement types, dose levels and their respective combinations.
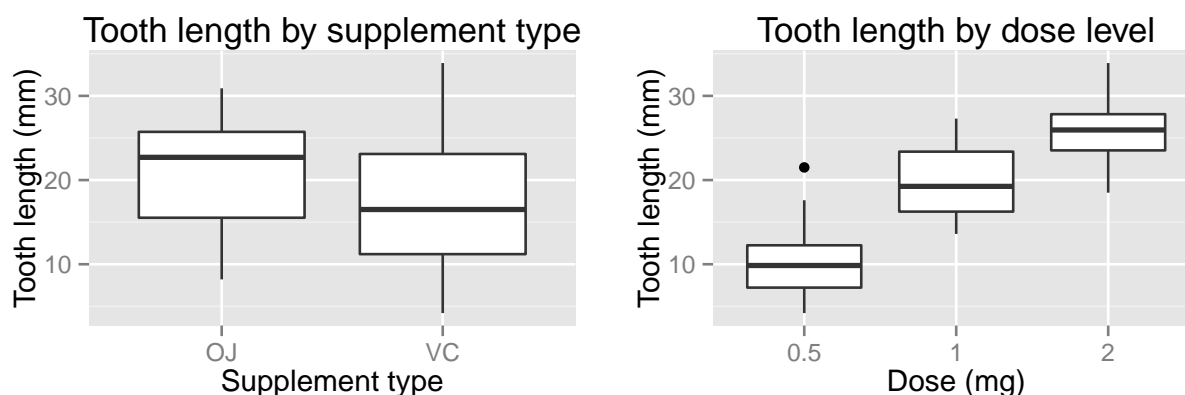

The left box plot above reveals that guinea pigs which received their dose of vitamin C using orange juice have on average a greater tooth length than those pigs which got their dose using ascorbic acid. The orange juice group distribution is skewed to the left whereas the ascorbic acid group seems to be fairly symmetric.

Grouping by dose level (right hand side box plot) reveals even greater differences in tooth length and their respective means. The differences between the dose level groups are so big that not even the interquartile range boundaries overlap:

```
##   dose    q25    q75
## 1  0.5  7.225 12.250
## 2  1.0 16.250 23.375
## 3  2.0 23.525 27.825
```

Both trends can be further observed when examining tooth length differences across both groups at the same time.

Tooth length by dose level and supplement type
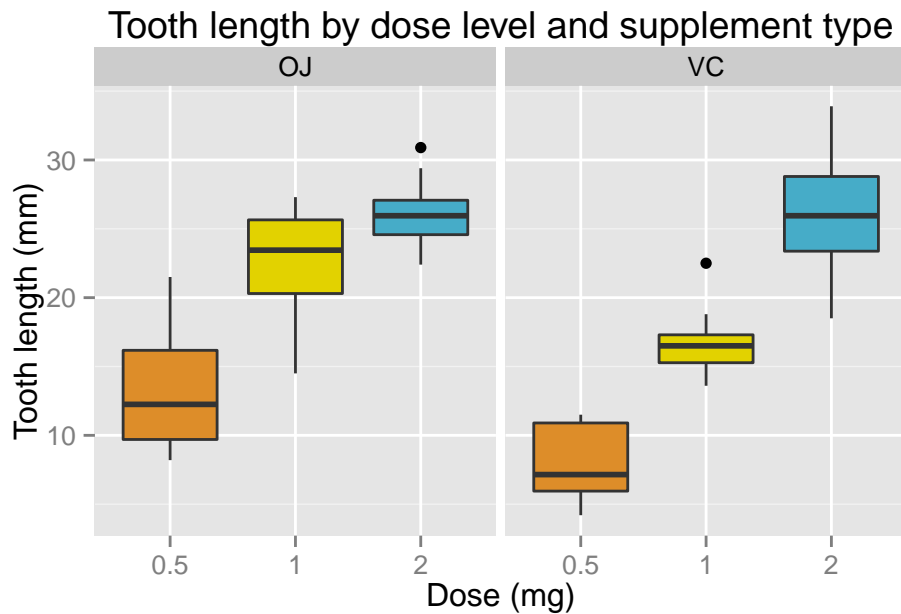
Hypothesis tests

Conditions

Since we dealing with subsets of n $<= 30$ our standard error estimate will generally not be accurate. Therefore, we will use the distribution for our hypothesis tests and when constructing confidence intervals.

Based on the current information it is hard to check if the 2 necessary prerequisites for this setup are met:

1. Independence of the observations
2. Observations come from a nearly normal distribution.

However, we assume that the guinea pigs were randomly assigned to one of the groups (guarantees independence) and that they were sampled from a nearly normal population.

Differences in supplement types

First, we will check if the observed difference in tooth length means between pigs who received their dose using orange juice and those who received their dose via ascorbic acid is statistically different.

Our conservative choice will be a t-test with unequal variances of the two samples. The table below shows the results:

```
##   estimate estimate1 estimate2 statistic    p.value parameter    conf.low
## 1      3.7  20.66333  16.96333  1.915268 0.06063451  55.30943 -0.1710156
##   conf.high
## 1  7.571016
```

With a p-value of 0.061 which is larger than the significance value, 0.05, we fail to reject the null hypothesis. The data does not provide strong evidence that the average difference in tooth length across the two supplement types is different from 0.

The hypothesis test result corresponds to the respective confidence interval -0.171, 7.571 because it includes 0.

Differences in dose levels

Since we are dealing with 3 different dose levels we need to apply 3 t-tests to cover all factor level combinations. Again, our null hypothesis in all cases is that the average difference in tooth length is 0.

The table below summarizes the results of the three 3 pairwise comparisons. Column `null_hypothesis` shows which null hypothesis was used for the respective t-test also indicating which sub group dose levels means were compared to each other.

```
##       null_hypothesis estimate estimate1 estimate2 statistic       p.value
## 1   µ2mg - µ1mg = 0     6.365    26.100    19.735  4.900484 1.906430e-05
## 2 µ2mg - µ0.5mg = 0    15.495    26.100    10.605 11.799046 4.397525e-14
## 3 µ1mg - µ0.5mg = 0     9.130    19.735    10.605  6.476648 1.268301e-07
##   parameter  conf.low conf.high
## 1  37.10109  3.733519  8.996481
## 2  36.88259 12.833833 18.156167
## 3  37.98641  6.276219 11.983781
```

In all three cases the p-value is smaller than the significance level 0.05 which means we reject the null hypothesis each time. The data provides strong evidence that the average tooth length of guinea pigs is different for each dose level.

Interpreting the first row of the table for illustriation:

We are 95% confident that the average tooth length of guinea pigs who received a 2mg dose of vitamin C is on average 3.73 to 9 millimeters higher than those who received a 1mg dose.

Conclusion

We were able to show that the observed difference of tooth length mean across supplement types is statistically not different from 0.

However, we were able to highlight the fact that tooth length indeed varies across groups of dose levels and that all of those differences are statistical and significant.