Statistical Inference Course Project: Distribution Simulation and Basic Inferential Data Analysis - Part 2

Alejandro Montoya

May 8, 2017

Part 2: Basic Inferential Data Analysis

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

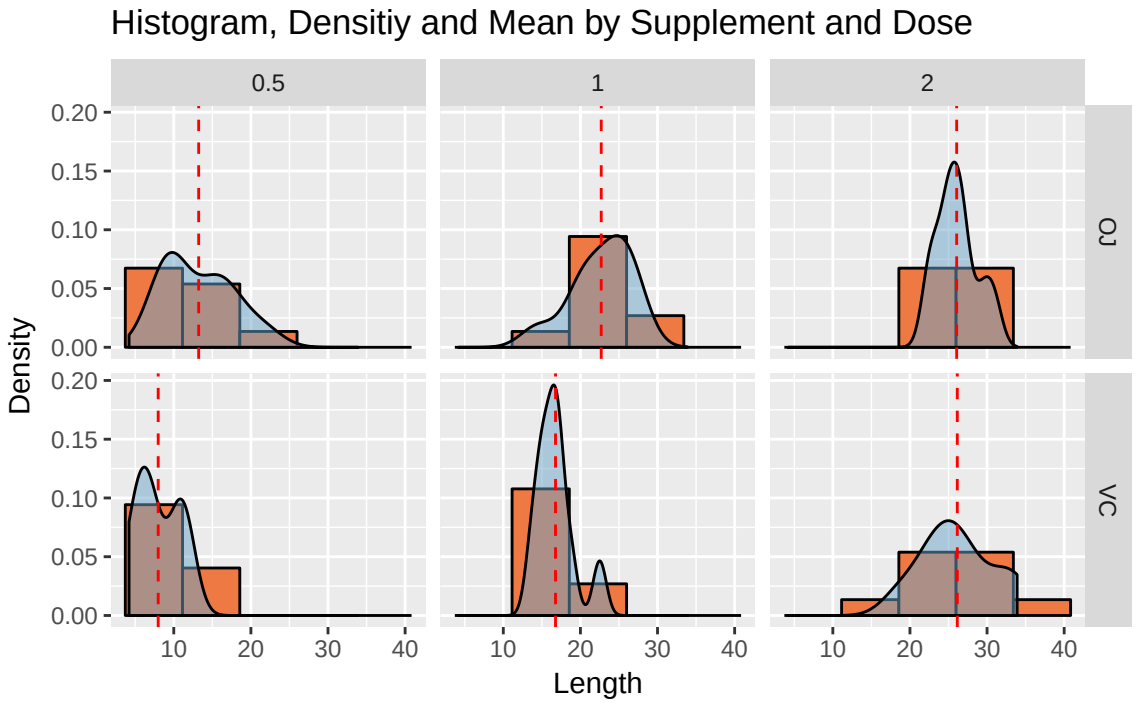
Loading the data

The data frame has 60 observations with 3 variables:

- len: Shows the total length of "odontoblasts" (cells responsible for tooth growth)
- supp: Vitamin C supplement used in the subject (OJ: Orange Juice, VC: Ascorbic Acid)
- dose: Amount of vitamins given to the subject (0.5, 1 and 2 mg/day)

Visual Exploration and Analysis

Now, let's explore the data to give us an idea of what we are dealing with. Let's use a chart where we can easily see the distribution of each sample, by looking at the histogram, density plot and mean of each one of the different groups that were sampled, i.e., we'll group by Supplement and Dose against the Length



This chart clearly shows us the distribution of each test group, and how the mean of the length varies from one group to another.

Numeric Exploration and Analysis

Now, let's try to do the same but looking at the numbers:

Table 1: Summary by Supplement and Dose

Supplement	Dose	Mean Length	Max Length	Min Length	Variance	Std Dev	Observations
OJ	0.5	13.23	21.5	8.2	19.889	4.460	10
OJ	1.0	22.70	27.3	14.5	15.296	3.911	10
OJ	2.0	26.06	30.9	22.4	7.049	2.655	10
VC	0.5	7.98	11.5	4.2	7.544	2.747	10
VC	1.0	16.77	22.5	13.6	6.327	2.515	10
VC	2.0	26.14	33.9	18.5	23.018	4.798	10

Confidence Intervals / Hypothesis Testing: Comparison of tooth growth by supp and dose

Now, let's compare the different tests to identify if there's any difference in the growth of the teeth when using different supplements in different dosage. These will be our hypothesis:

- H_0: There's NO difference in tooth growth when using different Supplements (i.e. Orange Juice or Ascorbic Acid)
- H_1: There's a difference in tooth growth when using different Supplements (i.e. Orange Juice or Ascorbic Acid)

We also have to take into account the fact that the dosage is different, so we will run the hypothesis analysis with every different dose:

Table 2: Hypothesis Testing: Confidence Intervals by Dose

Dose	P Value	Conf. Interval	Decision
0.5 mg/day	0.0064	1.72, 8.78	Reject null hypothesis
1 mg/day	0.001	2.8, 9.06	Reject null hypothesis
2 mg/day	0.9639	-3.8, 3.64	Failed to reject null hypothesis

Conclusions

By the analysis seen above, we can conclude that:

- When applied a dose of 0.5 mg/day of vitamin C, depending on which Supplement was used, the tooth growth varies, given bigger growth when using Orange Juice (OJ) than when using Ascorbic Acid (VC)
- When applied a dose of 1 mg/day of vitamin C, depending on which Supplement was used, the tooth growth varies, given bigger growth when using Orange Juice (OJ) than when using Ascorbic Acid (VC)
- When applied a dose of 2 mg/day of vitamin C, we CAN NOT say that there's no difference in tooth growth, independently of which Supplement was used

Assumptions

In order for these conclusions to be valid we have to assume that:

- The Guinea pigs are representative of the population and were randomly selected
- The dosage and supplement were assigned randomly to the Guinea pigs

Appendix

The R Code used to generate all these results is located at Tooth Growth Analysis