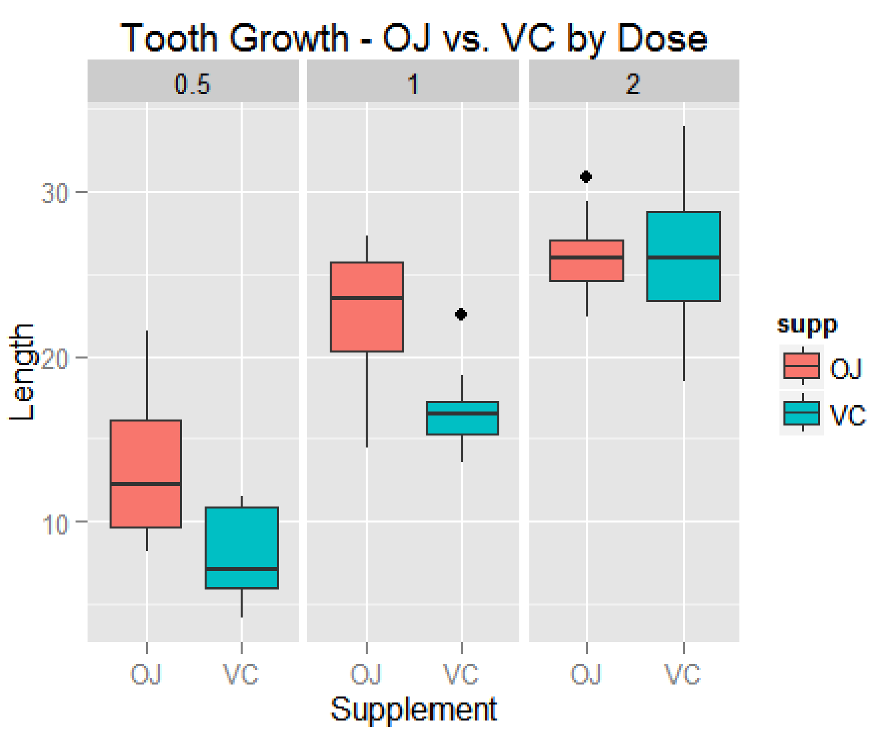
Analysis of Tooth Growth Data

Overview

In this report I will exam the `ToothGrowth` data set from the `datasets` R library. According to the R Documentation, the `ToothGrowth` data set contains the results from an experiment in which "the response is the 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)." After a brief exploratory data analysis, I perform multiple hypothesis tests to see if the two factors, dose and delivery method, have a significant effect on tooth growth.

Exploratory Data Analysis

```
library(ggplot2)
ggplot(aes(supp, len), data = ToothGrowth) +
  geom_boxplot(aes(fill = supp)) + facet_grid(. ~ dose) +
  labs(x = "Supplement", y = "Length", title = "Tooth Growth - OJ vs. VC by Dose")
```



After an initial exploration of the data using the box plot shown above, the statements listed below appear to be true.

- At the .5 mg and 1.0 mg doses, ascorbic acid delivered via OJ promote greater growth, than when delivered via vitamin C
- For the dose of 2.0 mg, VC is the superior delivery method, i.e. it promote greater growth
- For a given supplement, each incremental increase in dose promotes greater growth

Hypothesis Tests and Results

Hypothesis Tests

I will use the `t.test` function, with a 95% confidence level, to test whether the observations outlined above are true. In other words, I will use hypothesis tests, in which the null hypothesis is stated as $H_0: \mu_1 = \mu_2$, to see if there is statistically significant difference between each corresponding groups.

```
library(dplyr)
# Compare delivery method @ .5 mg
d05 <- filter(ToothGrowth, dose == .5)
t05 <- t.test(len ~ supp, data = d05)
# Compare delivery method @ 1.0 mg
d10 <- filter(ToothGrowth, dose == 1)
t10 <- t.test(len ~ supp, data = d10)
# Compare delivery method @ 2.0 mg
d20 <- filter(ToothGrowth, dose == 2)
d20oj <- d20$len[d20$supp == "OJ"]; d20vc <- d20$len[d20$supp == "VC"]
t20 <- t.test(d20vc, d20oj)
# Compare doses for OJ: .5 mg to 1.0 mg, .5 mg to 2.0 mg, 1.0 mg to 2.0 mg
oj <- filter(ToothGrowth, supp == "OJ")
oj05 <- filter(oj, dose == .5)
oj10 <- filter(oj, dose == 1)
oj20 <- filter(oj, dose == 2)
tOj10_05 <- t.test(oj10$len, oj05$len)
tOj20_05 <- t.test(oj20$len, oj05$len)
tOj20_10 <- t.test(oj20$len, oj10$len)
# Compare doses for VC: .5 mg to 1.0 mg, .5 mg to 2.0 mg, 1.0 mg to 2.0 mg
vc <- filter(ToothGrowth, supp == "VC")
vc05 <- filter(vc, dose == .5)
vc10 <- filter(vc, dose == 1)
vc20 <- filter(vc, dose == 2)
tVc10_05 <- t.test(vc10$len, vc05$len)
tVc20_05 <- t.test(vc20$len, vc05$len)
tVc20_10 <- t.test(vc20$len, vc10$len)
```

Delivery Method Results

The results from the first 3 tests are shown in the table below. Based on these results, we can reject H_0 for doses 0.5 mg and 1.0 mg, i.e. the true difference in means is not equal to 0; in both cases, the OJ delivery method promotes greater tooth growth than VC. However, we cannot reject H_0 for the 2.0 mg dose.

Dose	p-value	Confidence Interval	Mean (OJ)	Mean (VC)
0.5 mg	0.006	[1.72, 8.78]	13.23	7.98
1.0 mg	0.001	[2.8, 9.06]	22.7	16.77
2.0 mg	0.964	[-3.64, 3.8]	26.14	26.06

Dose Level Results

The results from the remaining 6 tests are shown in the table below. Based on these results, we can reject H_0 for all of the test.

Delivery	X	Y	p-value	Confidence Interval	Mean (X)	Mean (y)
OJ	1.0 mg	0.5 mg	8.78e-05	[5.52, 13.42]	22.7	13.23
OJ	2.0 mg	0.5 mg	1.32e-06	[9.32, 16.34]	26.06	13.23
OJ	2.0 mg	1.0 mg	3.92e-02	[0.19, 6.53]	26.06	22.7
VC	1.0 mg	0.5 mg	6.81e-07	[6.31, 11.27]	16.77	7.98
VC	2.0 mg	0.5 mg	4.68e-08	[14.42, 21.9]	26.14	7.98
VC	2.0 mg	1.0 mg	9.16e-05	[5.69, 13.05]	26.14	16.77

Conclusions and Assumptions

At the .05 mg and 1.0 mg doses, OJ is promotes greater tooth growth than VC; however, at a 2.0 mg dose we cannot say that one delivery method results in more growth than the other. Furthermore, when we look at incremental doses, using the same delivery method, higher doses promote more growth than lower doses. These conclusions, and the test performed to arrive at these conclusions, are based on the following assumptions:

- The guinea pigs were chosen using an accepted method for randomization
- The sample groups are independent
- The variances of each group, in any given test, are not equal