Tooth Growth in Guinea Pigs from Vitamin C Supplements

Introduction

The `ToothGrowth` data set considers the effects of Vitamin C on tooth length in guinea pigs administered through two different supplements. The supplements are administered to 10 subjects in three different doses (0.5 mg, 1 mg and 2 mg) for a total of 60 observations.

Examination of the help file for the data set reveals that the observations come from a study by C. I. Bliss published in 1952 called *The Statistics of Bioassay*. Unfortunately this document is not available online for free for further examination. Instead, we use `?ToothGrowth` for information.

The help file provides very little information. We have no information about control groups for example. We do not know the expected tooth length in guinea pigs who did not receive supplements.

Now, we'll perform our own exploratory analysis of the data, to address the following questions:

- What is the average tooth length observed in the subjects for each dosage? How are the averages different for each supplement?
- Was there enough data collected to believe that the mean observations are statistically significant at the 95% confidence level?
- Is there evidence to support the conclusion that Vitamin C delivered through ascorbic acid has the same impact as Vitamin C delivered in the form of ascorbic acid?

Exploratory Analysis

To get a sense of how the dosage level and supplement type interact, we plot the density of the sample means on an overlapped graph where the distributions and means can be easily understood visually.

```
# gather summary statistics on tooth length as it varies by dosage and
# supplement type
means <- ddply(ToothGrowth, c("supp", "dose"), summarize, sample.mean =
mean(len), sample.sd = sd(len), sample.var = var(len))
means

##   supp dose sample.mean sample.sd sample.var
## 1   OJ  0.5       13.23     4.460     19.889
## 2   OJ  1.0       22.70     3.911     15.296
## 3   OJ  2.0       26.06     2.655      7.049
```

```
## 4    VC   0.5          7.98       2.747        7.544
## 5    VC   1.0         16.77       2.515        6.327
## 6    VC   2.0         26.14       4.798       23.018
```

```
# how are the samples distributed and are the means correlated?
ggplot(ToothGrowth, aes(x = len, fill = supp)) + geom_density(alpha =
0.3) +
    geom_vline(data = means, aes(xintercept=sample.mean,  color =
supp),
                linetype = "dashed", size = 2) + facet_grid(~ dose)
```
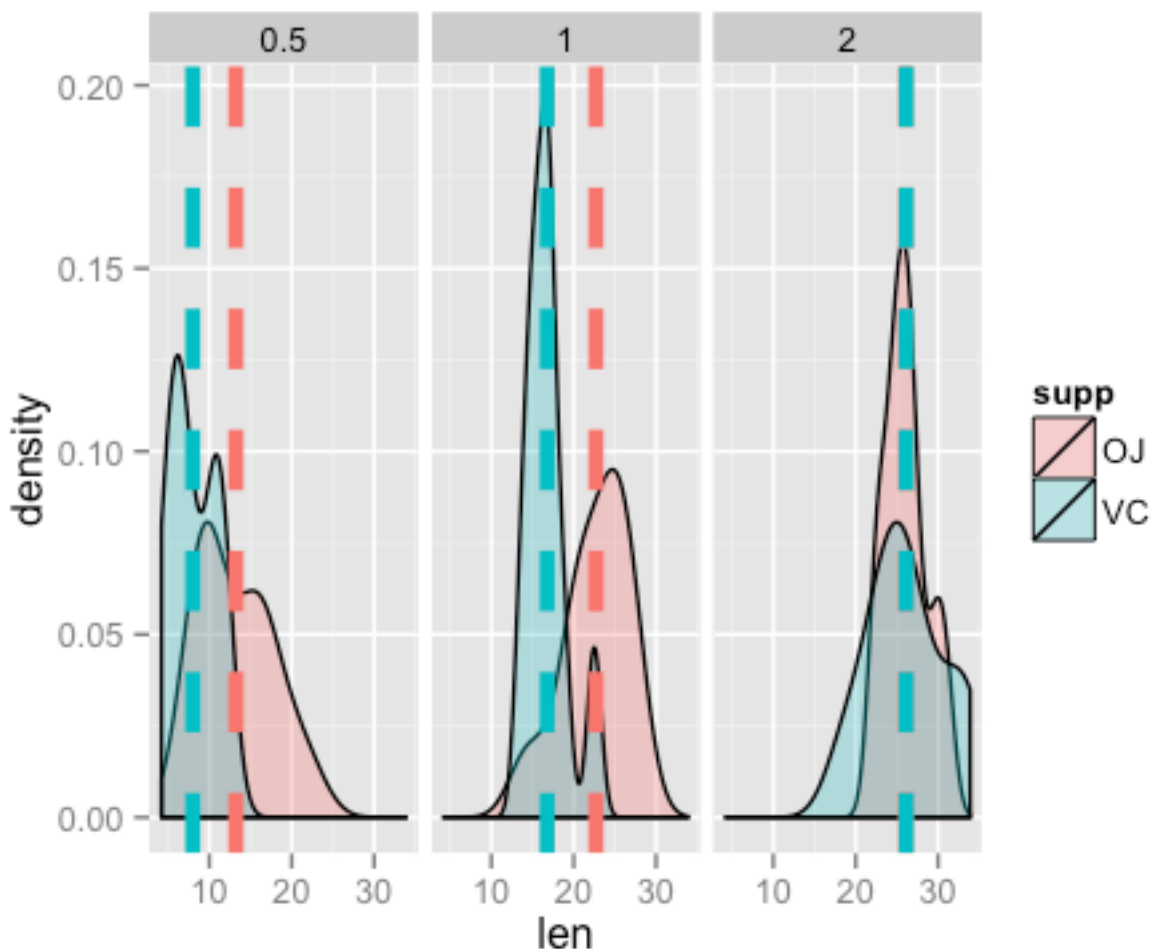

Summary of Data Set

Analysis shows that the differences in the sample means of the tooth lengths in guinea pigs appears to trend toward zero as the dosage administered increases. A simple visual analysis suggests that at the **2 mg** dosage, Vitamin C has the same effect on tooth length whether delivered through Orange Juice or Ascorbic Acid.

```
# what are the differences in the average tooth Length?
deltas <- ddply(means, c("dose"), summarize, OJ.vs.VC =
```

```
paste(abs(round(diff(sample.mean), 3)), "mg difference in sample
means"))
deltas

##   dose                          OJ.vs.VC
## 1  0.5 5.25 mg difference in sample means
## 2  1.0 5.93 mg difference in sample means
## 3  2.0 0.08 mg difference in sample means
```

However, we must demonstrate that the data collected is statistically significant enough to support this hypothesis. For that, we turn to our confidence tests.

Confidence Tests

```
# set the seed so others can reproduce easily
public.enemy <- 911
set.seed(public.enemy)
```

Conclusions

Note:

The code presented above relied on several installed packages, the list given below. If you are knitting this document yourself from the original source, use load("statinference-006-pt2.RData") to load the environment into **R**.

```
# show required packages
required.packages

## [1] "dplyr"   "ggplot2" "plyr"
```