Statistical Inference - Course Project: Analysis of ToothGrowth

Randy Armknecht

Saturday, August 22, 2015

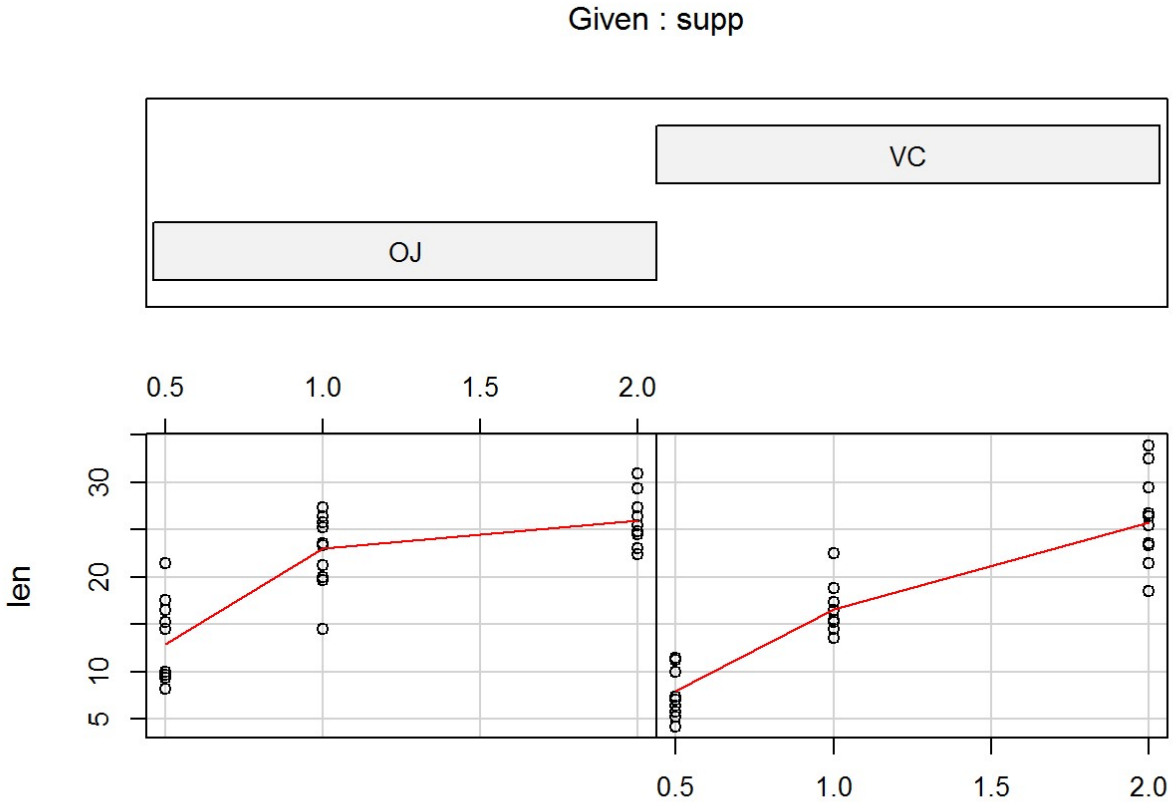
Overview

In this paper we examine the ToothGrowth dataset, using confidence intervals and hypothesis testing.

Objective 1: Data Load and Exploratory Analysis

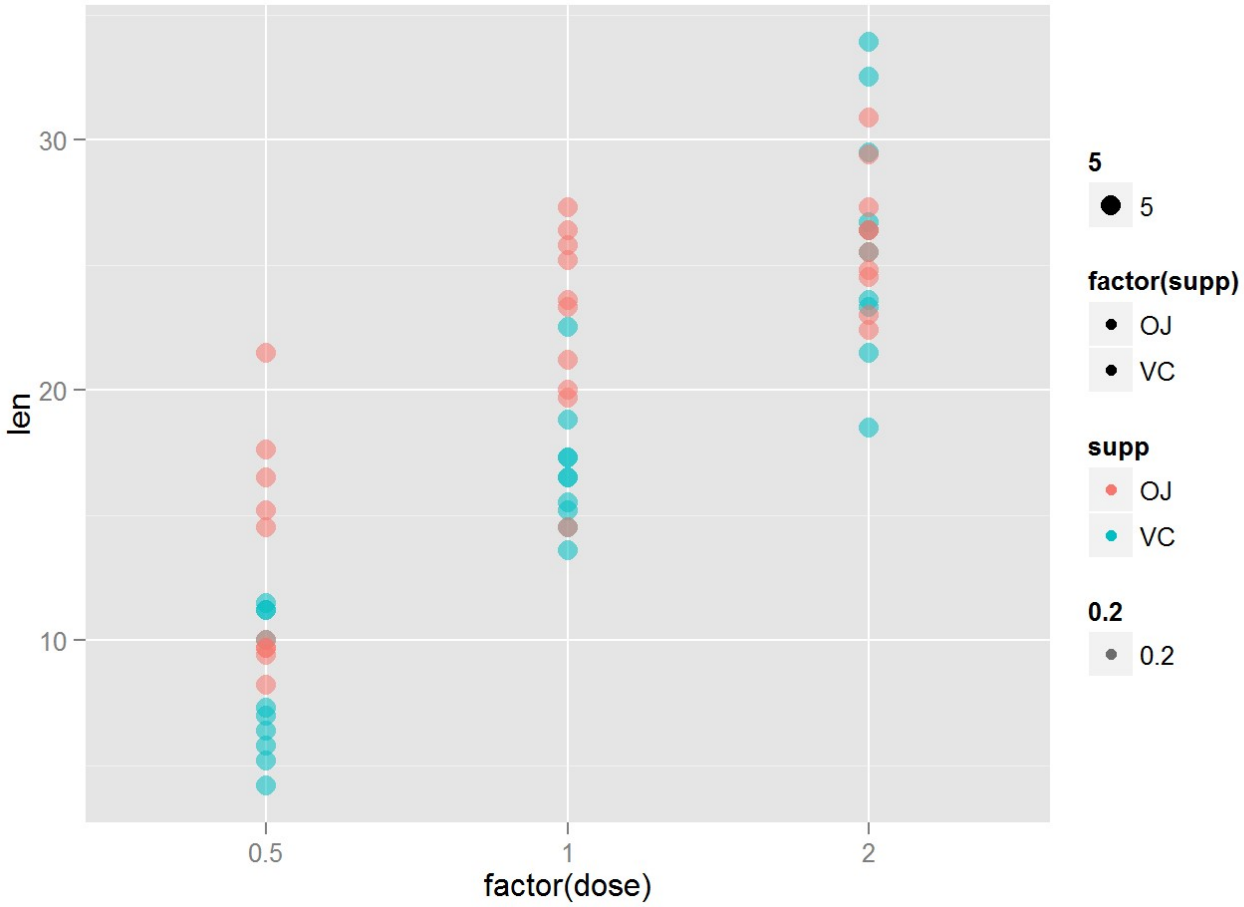
The first step is to load the ToothGrowth data, and conduct some exploratory analysis. I find that a nice way of doing this is by visualizing the data. I used *summary* and *head* as well as a description of the dataset[1].

```
data("ToothGrowth")

# How the data is displayed in the ToothGrowth documentation
require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")

# How I like to display the data
library(ggplot2)
```



ToothGrowth data: length vs dose, given type of supplement

```
ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=factor(supp), color=supp, size=5,
alpha=0.2)) + geom_point()
```



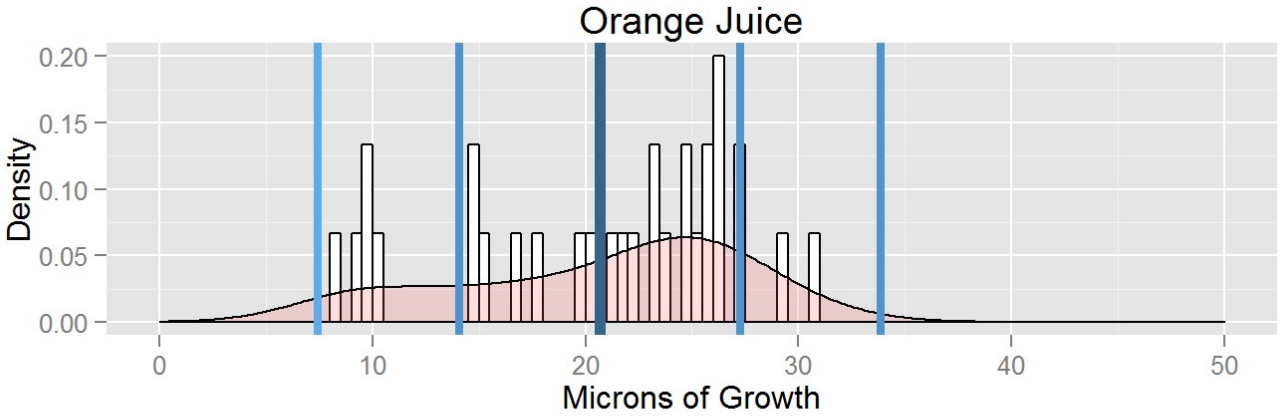
Objective 2/3: Provide a Basic Summary, Comparison by Supplement

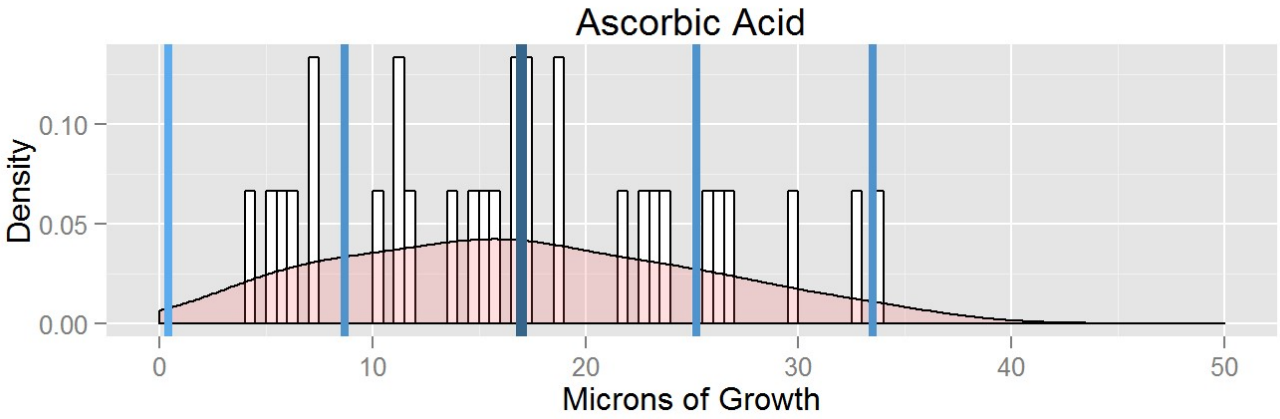
By examining the chart below, we can see a comparison of the growth results by supplement, regardless of dose. The dark blue line is the mean, and the lighter blue lines are the Standard Diviation markings. With the x-axis lined up, we can see that the second sd of the orange juice data is near the first standard diviation marker of the ascorbic acid. The mean is also greater for the orange juice.

```
library(gridExtra)
mean(ToothGrowth$len)
```

```
## [1] 18.81333
```

```
oj <- ToothGrowth$len[ToothGrowth$supp=="OJ"]
ojm <- mean(oj)
ojs <- sd(oj)

vc <- ToothGrowth$len[ToothGrowth$supp=="VC"]
vcm <- mean(vc)
vcs <- sd(vc)

p.oj <- ggplot(data.frame(oj), aes(x=oj)) + xlim(0,50) + labs(x="Microns of Growth",y=
"Density", title="Orange Juice") +
  geom_histogram(aes(y=..density..), binwidth=.5, color="black", fill="white") +
  geom_density(alpha=.2, fill="#ff6666") + geom_vline(x=ojm, size=2, color="steelblue4
") +
  geom_vline(x=ojm-ojs, size=1.5, color="steelblue3") + geom_vline(x=ojm+ojs, size=1.5
, color="steelblue3") +
  geom_vline(x=ojm-(2*ojs), size=1.5, color="steelblue2") + geom_vline(x=ojm+(2*ojs),
size=1.5, color="steelblue3")

p.vc <- ggplot(data.frame(vc), aes(x=vc)) + xlim(0,50) + labs(x="Microns of Growth",y=
"Density",
                                                              title="Ascorbic Acid") +
  geom_histogram(aes(y=..density..), binwidth=.5, color="black", fill="white") +
  geom_density(alpha=.2, fill="#ff6666") + geom_vline(x=vcm, size=2, color="steelblue4
") +
  geom_vline(x=vcm-vcs, size=1.5, color="steelblue3") + geom_vline(x=vcm+vcs, size=1.5
, color="steelblue3") +
  geom_vline(x=vcm-(2*vcs), size=1.5, color="steelblue2") + geom_vline(x=vcm+(2*vcs),
size=1.5, color="steelblue3")

grid.arrange(p.oj,p.vc)
```





Objective 4: Conclusions

After my analysis I would conclude that the Orange Juice had a great impact on tooth growth. My assumptions are mainly that this extremely small sample is representative of the population.

References

1. http://www.inside-r.org/r-doc/datasets/ToothGrowth (http://www.inside-r.org/r-doc/datasets/ToothGrowth)