Analyzing the ToothGrowth data in R

Sunday, August 24, 2014

Data: http://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html

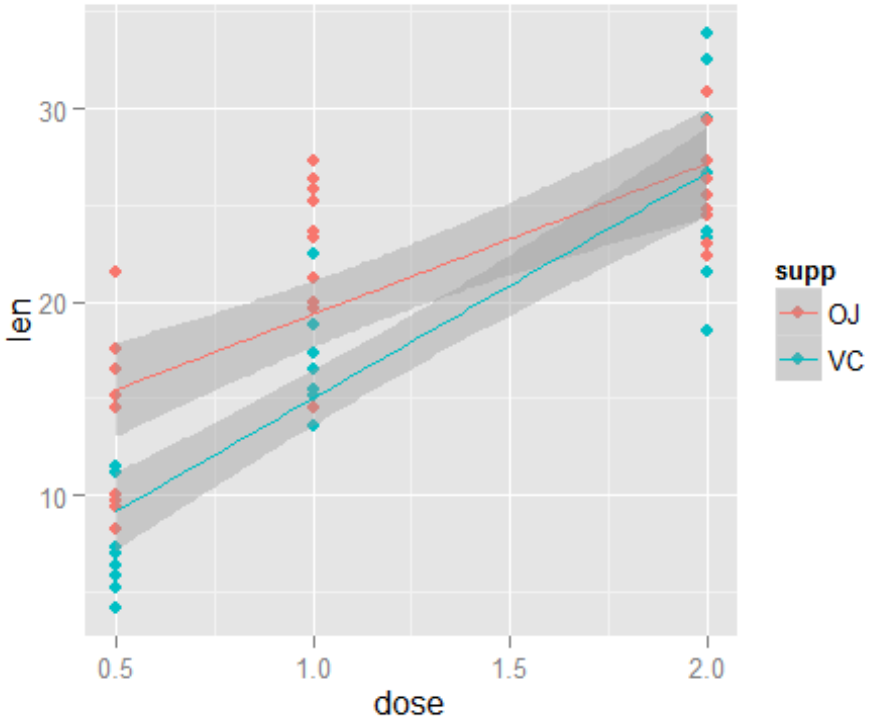
Basic exploratory data analysis

Comparing the effect of orange juice and vitamin C dosage on tooth lengths. The plot below shows that at a lower dosage, orange juice has a higher effect. However, as the dosage becomes higher vitamin C performs better.

```
data(ToothGrowth); library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.0.3

g <- ggplot(data=ToothGrowth, aes(x=dose, y=len, colour=supp))
g + geom_point() + geom_smooth(method="lm")
```



Summary of the data

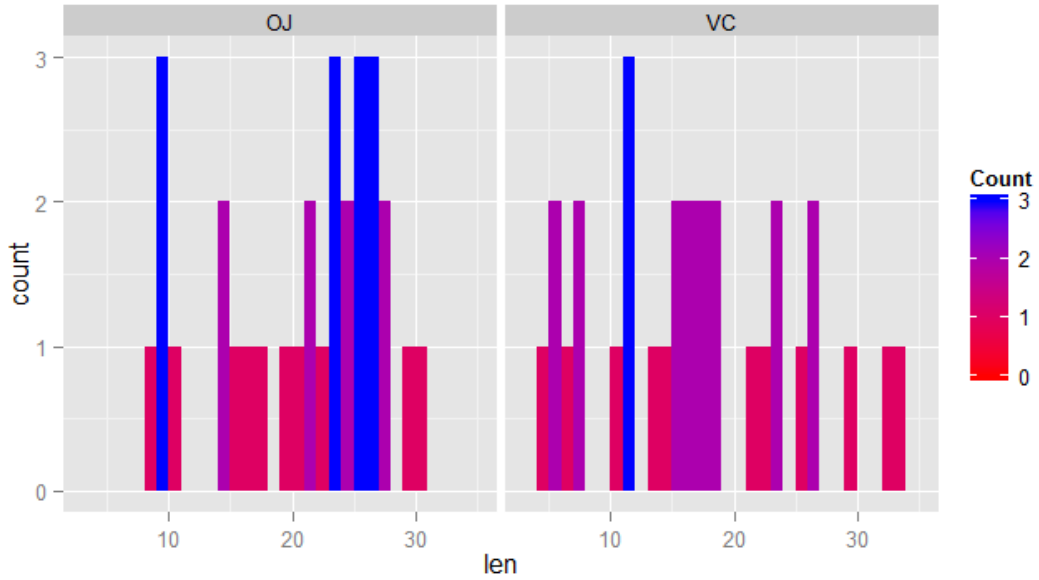
This dataset includes 3 of guinea pig tooth length observations as a function of 3 dose levels of Vitamin C (0.5, 1, or 2 mg) using 2 delivery methods (orange juice or ascorbic acid.). The Tooth lengths vary from 4.2 to 33.9, with a mean of 18.8133. The variance in the observed length data is relatively high: var=58.512, sd=7.6493.

Confidence intervals and hypotheses

Two hypotheses are examined 2: 1. Orange Juice (OJ) and Vitamin C (VC) have the same effect on tooth length. 1. OJ and VC have the same effect on tooth length when the dosage is 2.0mg

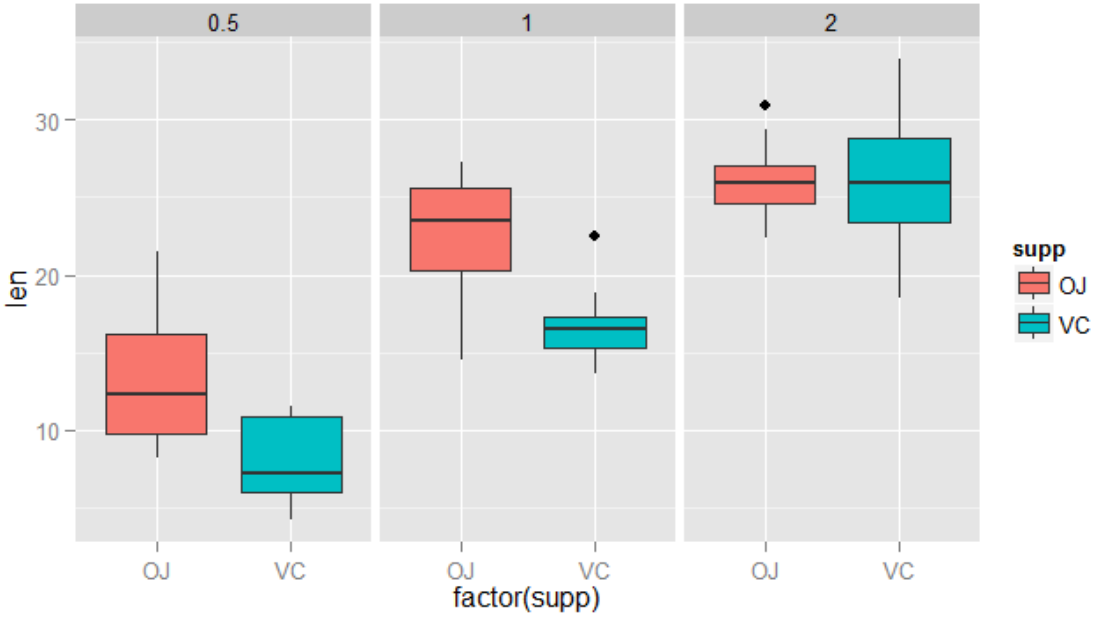
Hypothesis 1:

By comparing the means of each effect, the mean of OJ (20.6633) > mean of VC (16.9633), it seems that orange juice has a higher impact. Nevertheless, tests are run to prove the hypothesis. The histogram below shows that OJ is concentrated more at higher lengths compared to VC. Visually, it seems that the hypothes is false. The results of the t.test are shown in the conclusion.



Hypothesis 2:

By looking only at the 2.0mg dosage, we come at a different conclusion. The mean of OJ (26.06) is close to VC (26.14). The results of the t.test are shown in the conclusion. The boxplot below shows that when we segment the data according to dosage, the 2.0mg dosage seems to have very similar results regardless of whether using OJ or VC.



Conclusions and Assumptions

Hypothesis #1 is false. The Test Statistic (TS) (1.9153) is above the 95% confidence interval of (1.6716). Hypothesis #2 is true. The Test Statistic (TS) (-0.0461) is below the 95% confidence interval of (1.7341). This indicates that the dosage amount of 2.0mg is more important than the method of the delivery used.