

Analyzing the ToothGrowth Dataset and the Impact of Delivery Methods

Michael Connolly

Saturday, May 23, 2015

The ToothGrowth dataset describes the length of 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 and ascorbic acid). That turns into 60 different observations on 3 variables: length (len), supplement type (supp), and the dose size (dose).

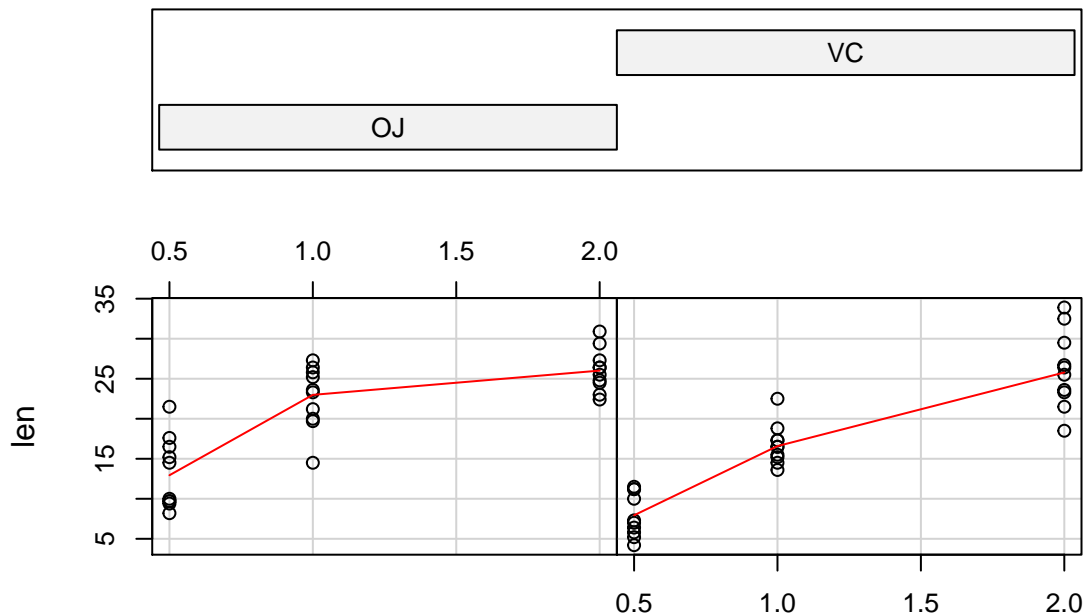
```
head(ToothGrowth)
```

```
##      len supp dose
## 1   4.2   VC  0.5
## 2  11.5   VC  0.5
## 3   7.3   VC  0.5
## 4   5.8   VC  0.5
## 5   6.4   VC  0.5
## 6  10.0   VC  0.5
```

A quick and dirty analysis of the data shows a correlation between doseage and length, as well as the OJ delivery method having a larger impact than ascorbic acid at lower dose levels, such as 1.0.

```
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth, xlab = "ToothGrowth data: length vs
```

Given : supp



ToothGrowth data: length vs dose, given type of supplement

One interesting question we explore is if the difference in delivery method is statistically interesting. We can use T Tests to do this analysis, and the answer is, yes, it's statistically interesting. Below, through the results of a T test (both done with the `t.test` method, and also calculated manually), you can see the confidence interval is clearly completely above zero. This results points to the OJ delivery method has a statistically relevant positive impact of tooth growth, as compared to the alternative delivery method, ascorbic acid, at the 95% confidence level.

Conversely, if the interval had overlapped zero, based on this sample size, we would not have confidence that there was an impact difference due to delivery method.

```
teeth <- ToothGrowth
group_oj <- teeth[teeth$supp=="OJ",]
group_vc <- teeth[teeth$supp=="VC",]
group_diff <- group_oj$len - group_vc$len
group_diff_mean <- mean(group_diff)
group_diff_sd <- sd(group_diff)
group_diff_count <- length(group_diff)
t.test(group_diff)
```

```
##
## One Sample t-test
##
## data: group_diff
## t = 3.3026, df = 29, p-value = 0.00255
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
```

```
## 1.408659 5.991341
## sample estimates:
## mean of x
##      3.7

group_diff_mean + c(-1, 1) * qt(.975, group_diff_count - 1) * group_diff_sd / sqrt(group_diff_count)

## [1] 1.408659 5.991341
```