Inferential Exercise

bbneo

03/18/2015

The Effect of Vitamin C on Tooth Growth in Guinea Pigs

Introduction

The `ToothGrowth` dataset in `R` is a set of measurements on:

"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)."

```
data(ToothGrowth)
```

See:

```
help(ToothGrowth)
```

In this exercise, we will perform some basic exploratory and inferential data analysis on the `ToothGrowth` dataset.

Basic Summary of the ToothGrowth Data

```
data(ToothGrowth)
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
head(ToothGrowth)
```

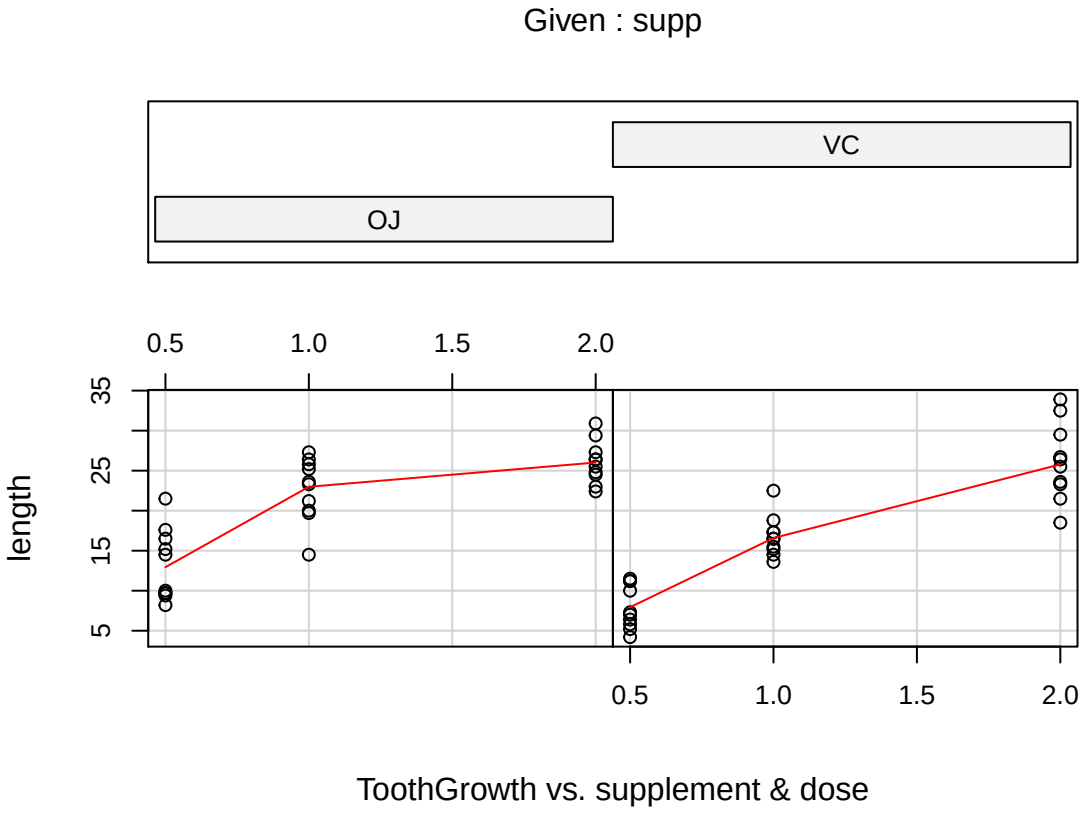
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```r
with(ToothGrowth, table(dose,supp))
```

```
##      supp
## dose  OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

The `coplot` example from the `help()` file gives a nice overview of the results. In this, you can see a series of length measurements from each of the 10 animals at 3 dose levels for each of the two supplement forms. A dose-response trend seems clear in the data.

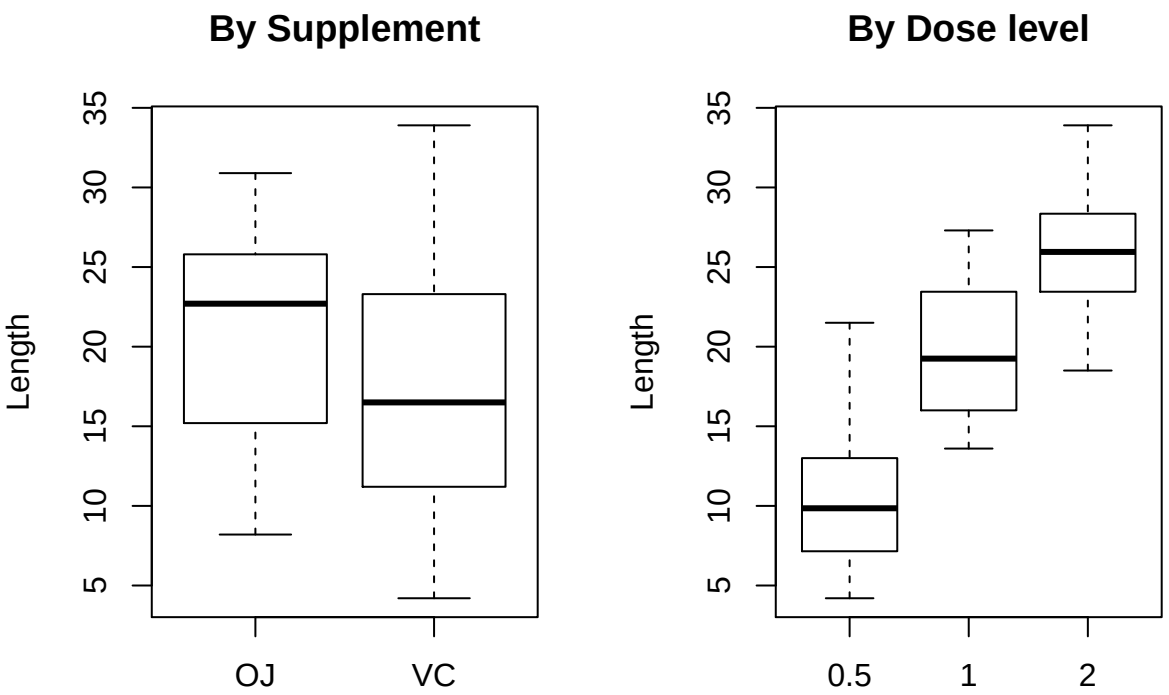
To be clear, *independent* groups of 10 animals were used for each of 6 different conditions for 60 animals in all. (i.e., the groups were not paired.)



Statistical Effects of Dose & Supplement

Overview

A summary boxplots from the data suggest inferential questions to test.



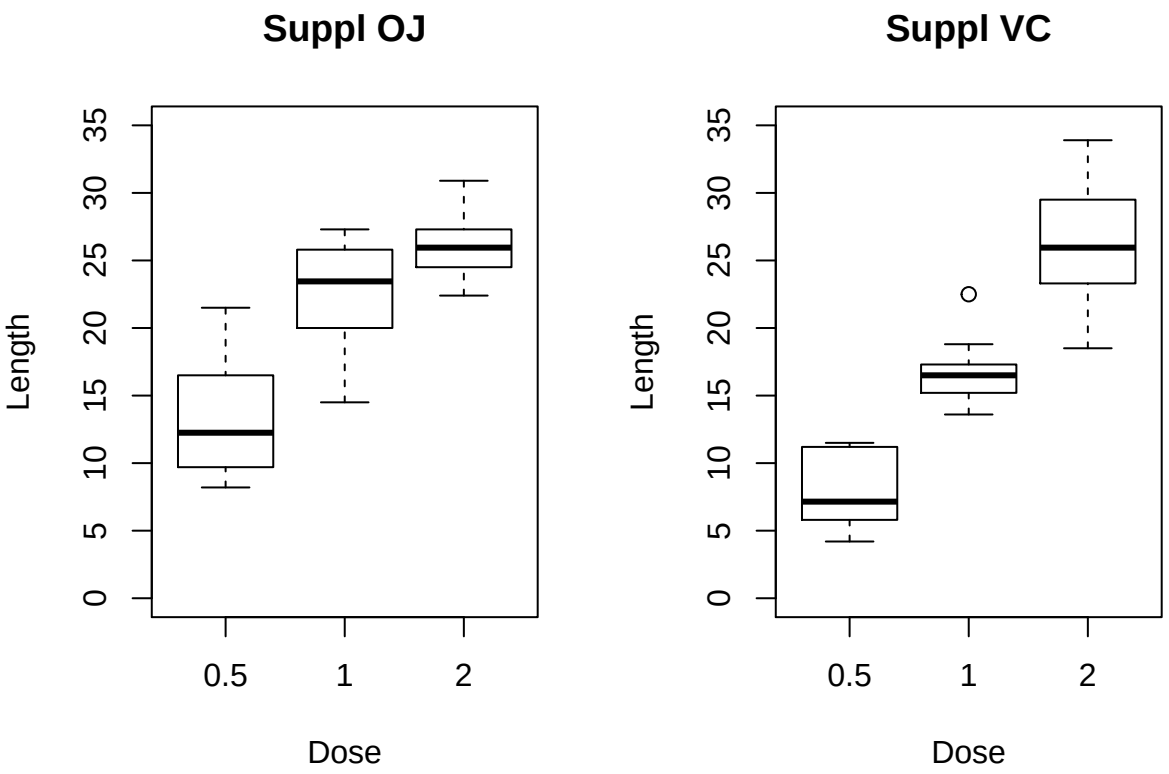
(The lower and upper ends of the box-and-whisker plot represent the 25th and 75th quartiles of the data. The dashed lines represent the extent to the minimum and maximum values of the data, and the midline across the box represents the *median* of the data.)

There seems to be a clear dose response relationship between the supplements and toothgrowth, but no clear difference between supplement types.

Subsetting Growth by Dose and Supplement

Breaking the dose response relationship out by supplement type highlights some interesting differences between the supplement types at different doses.

Although there does not seem to be a significant difference in growth between the supplements at the highest dose level, at the low and middle dose levels, the `OJ` supplement demonstrates more growth. I will present the statistical test results below.



What is the significance of effects of dose and/or supplement on toothgrowth?

1. dose levels for a given supplement, or
2. supplements at a given dose level?

Hypothesis tests of length vs. supplement

Over the data for all dose levels, we cannot reject the hypothesis that the difference in tooth growth between the supplements is zero, because the 95% CI includes zero.

```
t.test( len ~ supp, ToothGrowth )
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

Tooth growth between supplements at the *same* dose levels, does show some significant differences.

The underlying assumption to this analysis is that the growth measurements at given dose and supplement are iid random variables.

The R code below will build up a result dataframe with t-test results for the tooth growth between supplements, by dose levels:

```
dose <- c(0.5,1.0,2.0)

results <- data.frame()
for (d in dose) {
  tdose <- t.test( len ~ supp,
                   subset(ToothGrowth,
                          dose==d))
  row <- with(tdose,
              cbind(d,statistic,parameter,
                    conf.int[1],conf.int[2],
                    p.value)
              )
  results <- rbind(results, row)
}

names(results) <- c("Dose","t.stat","df",
                    "95CI.lo", "95CI.hi",
                    "pval")

round(results,3)
```

```
##    Dose t.stat     df 95CI.lo 95CI.hi  pval
## t   0.5  3.170 14.969   1.719   8.781 0.006
## t1  1.0  4.033 15.358   2.802   9.058 0.001
## t2  2.0 -0.046 14.040  -3.798   3.638 0.964
```