Analyzing the ToothGrow dataset

Sam Stites

November 23, 2014

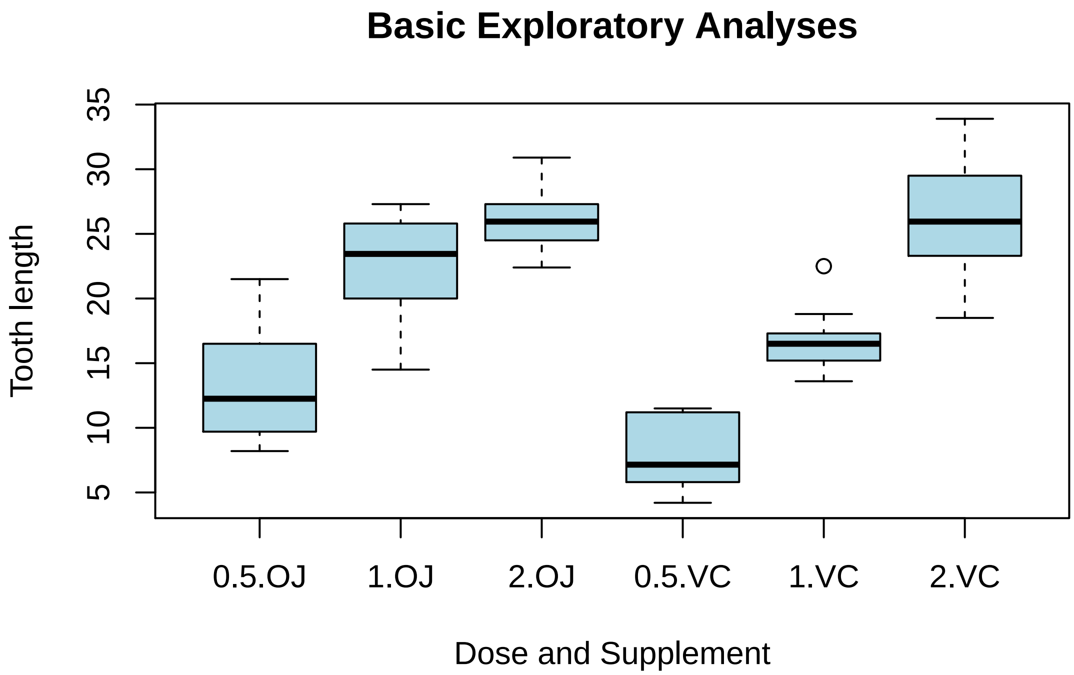
About the dataset

From the cran repository, we know that this dataset represents the response length in odontoblasts (teeth) for 10 guinea pigs with respect to a dosage level of Vitamin C through two delivery methods. Vitamin C was administersed at three dose levels of 0.5, 1, and 2 mg with two delivery methods, orange juice and ascorbic acid.

```
# Part 1
library(datasets)
data(ToothGrowth)
toothgrowth <- transform(ToothGrowth, supp = factor(supp))
summary(toothgrowth) # not great...
```

```
##       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  
```

```
boxplot(len ~ dose * supp,
        toothgrowth,
        col  = "lightblue",
        main = "Basic Exploratory Analyses",
        xlab = "Dose and Supplement",
        ylab = "Tooth length") # much better!
```



From our exploratory analysis, we can see that there is a positive correlation between tooth length and dose of supplement for both delivery methods. Orange juice seems to be more effective at lower doses, while both methods center around the same tooth length at higher doses. In addition, we can also see that ascorbic acid has greater variance at higher dose levels.

Hypothesis testing

Testing the following hypothesis:

```
# Ho = lenOJ == lenVC // no difference in length between the two treatments.
# Ha = lenOJ != lenVC //    difference in length between the two treatments.
```

We perform a two-sided t-test of the following:

```
t.test(len ~ supp, data = toothgrowth)
```

This yeilds a 95% confidence interval around `[-0.171, 7.571]`, with a p-value of 0.06063. in this scenario, we would not reject this null hypothesis. Note, however, that a slight decrease in confidence - of 90% — would yield a confidence interval of `[0.468, 6.932]` meaning that we will probably need to drill down a little deeper into the data.

Hypothesis testing: Per-dose

Given the same hypothesis from before, `Ho = lenOJ == lenVC; Ha = lenOJ != lenVC`, we run hypothesis tests at the three levels allowed in this study - 0.5, 1.0, and 2.0 mg. these yield the following results:

```
# Test of OJ vs VC at dose of 0.5
doseVal = 0.5
```

```r
t.test(len ~ supp, data = toothgrowth[toothgrowth$dose == doseVal,])
# p-val: 0.006359
# conf.int: [1.719057, 8.780943]
# reject null? yes, in positive direction (oj is more effective)

# Test of OJ vs VC at dose of 1.0
doseVal = 1.0
t.test(len ~ supp, data = toothgrowth[toothgrowth$dose == doseVal,])
# p-val: 0.001038
# conf.int: [2.802148, 9.057852]
# reject null? yes, in positive direction (oj is more effective)

# Test of OJ vs VC at dose of 2.0
doseVal = 2.0
t.test(len ~ supp, data = toothgrowth[toothgrowth$dose == doseVal,])
# p-val: 0.9639
# conf.int: [-3.79807, 3.63807]
# reject null? no
```

Conclusions and Assumptions

Upon further inquiry, we find that at high doses of 2.0mg — there seems to be no difference between the two methods. However at lower doses of 0.5 and 1.0, orange juice is more effective at increasing tooth length in guinea pigs according to our findings. Technical assumptions made include that the variances are unequal - not enough information is provided to assume otherwise, and that each observation was independent - from the description in cran it isn't entirely clear that this is the case.

Appendix

Code executed to reproduce these results:

```r
# Part 1
library(datasets)
data(ToothGrowth)
toothgrowth <- transform(ToothGrowth, supp = factor(supp))
summary(toothgrowth) # not great...
boxplot(len ~ dose * supp,
        toothgrowth,
        col  = "lightblue",
        main = "Basic Exploratory
Analyses",
        xlab = "Dose and Supplement",
        ylab = "Tooth length") # much better!
```

```r
# Hypothesis test:
# Ho = lenOJ == lenVC // no difference in length between the two treatments.
# Ha = lenOJ != lenVC //    difference in length between the two treatments.

# performing a two-sided t-test:
t.test(len ~ supp, data = toothgrowth)
```

```r
t.test(len ~ supp, data = toothgrowth, conf.level = 0.9)

# =======
# Hypothesis tests: Per-dose
# Ho = lenOJ == lenVC;  Ha = lenOJ != lenVC

# Test of OJ vs VC at dose of 0.5
doseVal = 0.5
t.test(len ~ supp, data = toothgrowth[toothgrowth$dose == doseVal,])
# p-val: 0.006359
# conf.int: [1.719057, 8.780943]
# reject null? yes, in positive direction (oj is more effective)

# Test of OJ vs VC at dose of 1.0
doseVal = 1.0
t.test(len ~ supp, data = toothgrowth[toothgrowth$dose == doseVal,])
# p-val: 0.001038
# conf.int: [2.802148, 9.057852]
# reject null? yes, in positive direction (oj is more effective)

# Test of OJ vs VC at dose of 2.0
doseVal = 2.0
t.test(len ~ supp, data = toothgrowth[toothgrowth$dose == doseVal,])
# p-val: 0.9639
# conf.int: [-3.79807, 3.63807]
# reject null? no
```