A Summary of the Tooth Growth Dataset

Tim Wise

May 2015

Overview and Summary

The ToothGrowth data set measures the growth of teeth in guinea pigs in response to being given varying doses of Vitamin C via two delivery methods, orange juice or ascorbic acid.

We found that . . . .

The Details

Exploratory Analysis   Let's include the package with the ToothGrowth dataset and look at the structure of the dataset:

```
require("datasets")
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Let's look at the documentation for more information on the columns:

```
?ToothGrowth
```

```
## starting httpd help server ... done
```

The description of the data set (excerpted from the documentation):

```
The response is the 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).

A data frame with 60 observations on 3 variables:

    [,1]  len   numeric  Tooth length
    [,2]  supp  factor   Supplement type (VC or OJ)
    [,3]  dose  numeric  Dose in milligrams
```

Note we don't know the units for tooth length.

Tooth length is the dependent variable. Let's look at how many samples we have for the two independent variables, supplement type and dosage:

```
table(ToothGrowth$supp, ToothGrowth$dose)
```

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

So there are 10 samples for dosage of each supplement, 20 samples for each dosage, and 30 samples for each supplement type. 10 is a small number of samples. Let's keep that in mind in later analysis.

Two obvious questions come to mind:

1. Does the supplement type have an impact on tooth growth?
2. Does the dosage have an impact on tooth growth?

We could analyze those questions for each *supplement type, dosage* pair, or we analyize a single variable at an aggregate level, say compairing supplement type and ignoring dosage levels.

For the first question, we will look at just the supplement type and ignore the dosage level. Let's look at a boxplot of tooth growth for each supplement type:

```
plot(x=ToothGrowth$supp,
     y=ToothGrowth$len,
     main='Tooth growth by supplement type',
     xlab='Supplement type',
     ylab='Tooth length')
```
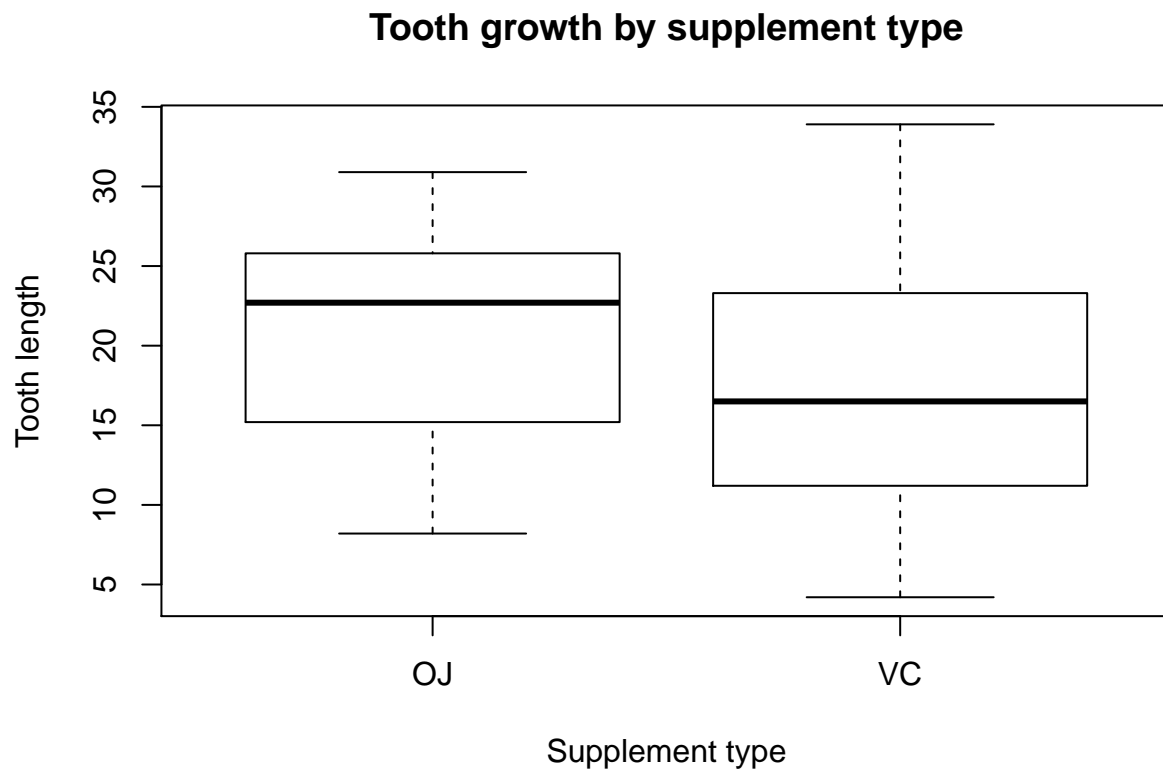
It appears that teeth growth more when the guinea pigs are given orange juice as opposed to ascorbic acid. Is this difference meaning full? This is something we will test in a moment.

For the second question, let's look at a scatterplot for each doesage level and supplement type:

```r
plot(x=ToothGrowth$dose,
     y=ToothGrowth$len,
     col=c('red','blue')[ToothGrowth$supp],
     pch=19,
     main='Tooth growth by dosage',
     xlab='Supplement type',
     ylab='Tooth length')

legend('bottomright',
       legend=levels(ToothGrowth$supp),
       pch=19,
       col=c('red','blue'),
       text.col = 'black')
```

Tooth growth by dosage
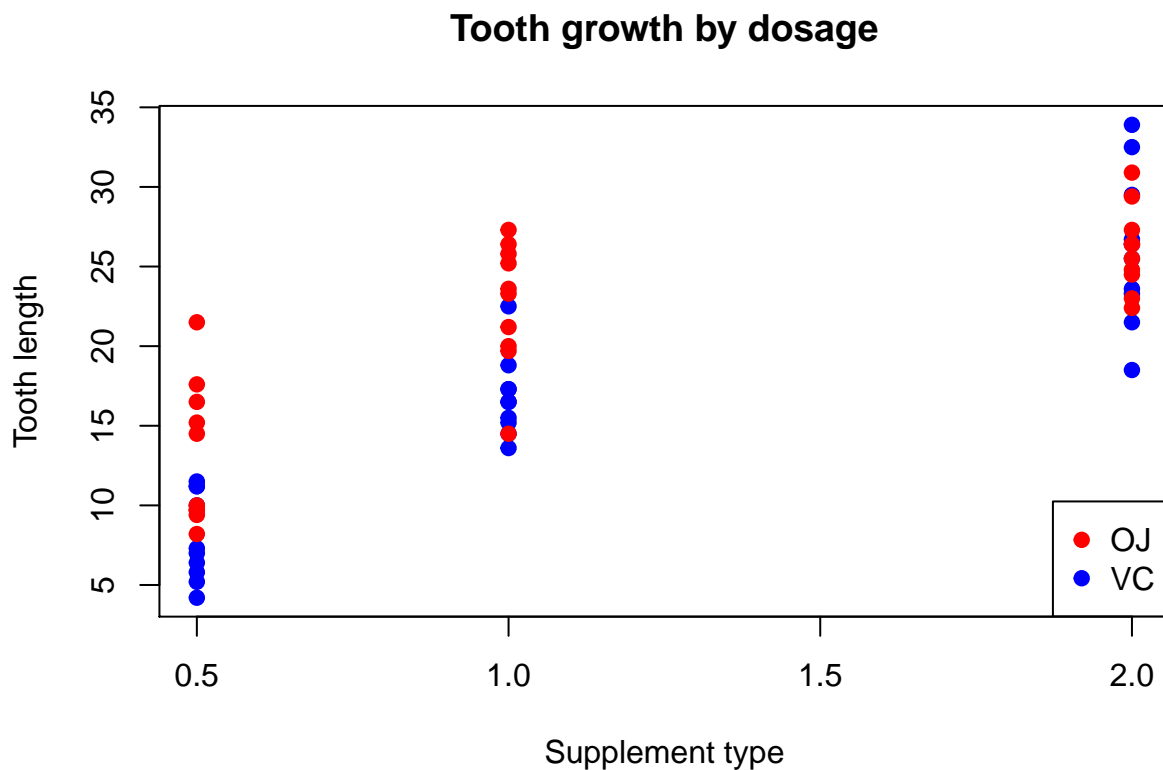
Does supplement type have an impace on tooth growth?

Does dosage have an impact on tooth growth?