Basic inferential data analysis: tooth growth dataset

D.V.

Synopsis

We're going to analyze the ToothGrowth data in the R datasets package. Our basic steps are:

- Provide a basic summary of the data.
- Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.
- Make conclusion and the assumptions of analysis.

Summary of the data

This data frame has 60 observations and 3 variables:

len - Tooth length (numeric). supp - Supplement type (factor): VC - "vitamin C" or OJ - "orange juice". dose - Dose in milligrams (numeric).

Let's see the summary:

```
data(ToothGrowth)
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 ...
```

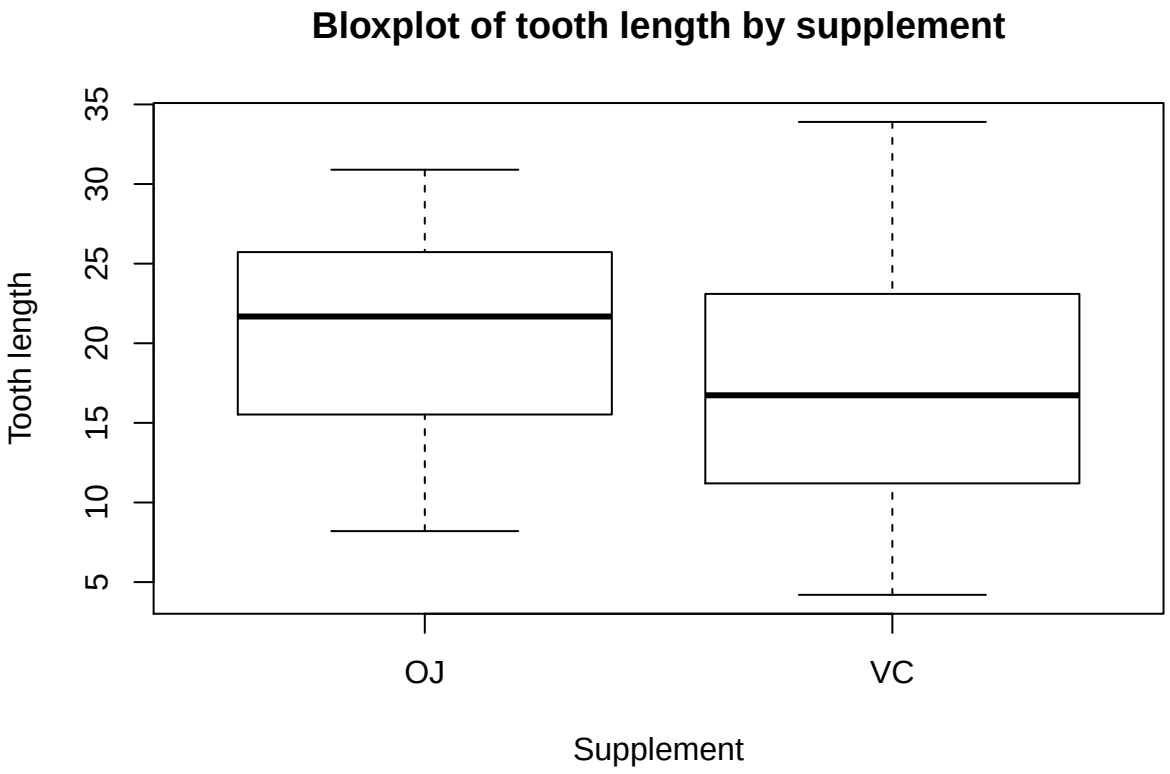
```
#summary(ToothGrowth)
sum <- tapply(ToothGrowth$len, ToothGrowth$supp, FUN=summary)
sum
```

```
## $OJ
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.20   15.53   22.70   20.66   25.73   30.90
##
## $VC
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20   11.20   16.50   16.96   23.10   33.90
```

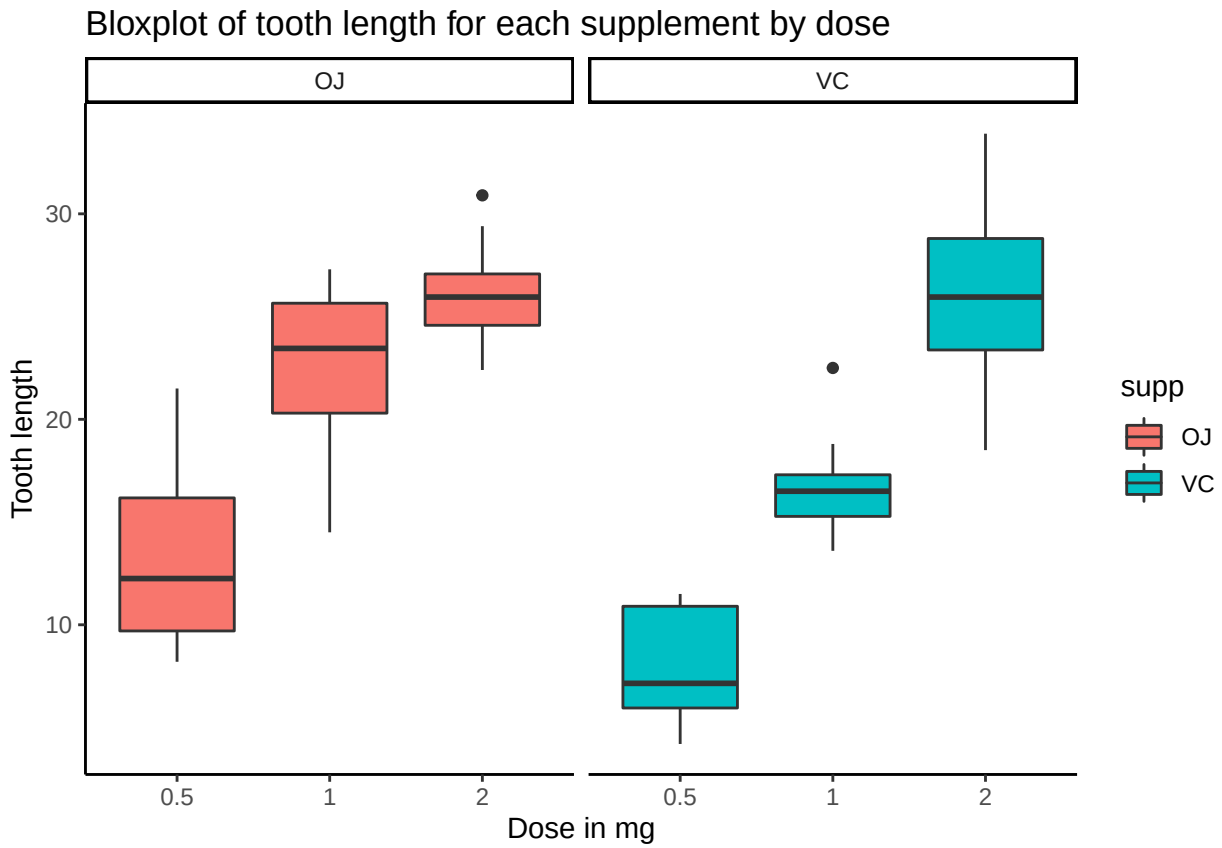
There are a bit of differences in mean and median between supp-groups. Let's compare them and check for outliers by boxplots.

Comparing tooth growth by supp and dose

```
boxplot(sum, main = "Bloxplot of tooth length by supplement",
        xlab = "Supplement",
        ylab = "Tooth length")
library(ggplot2)
```




```
ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=supp))+theme_classic()+
  geom_boxplot()+facet_grid(.~supp)+
  scale_x_discrete("Dose in mg") + scale_y_continuous("Tooth length") +
  ggtitle("Bloxplot of tooth length for each supplement by dose")
```



Data has a few outliers. In general, median effect of orange juice much larger than C-vitamin. But in higher doses (~2 mg), this effect is almost equal (however, there are unequal variance between groups, so that naive comparing isn't so well).

Let's make a statistical t-tests. The null hypothesis of t-test is that there is no difference in tooth growth between dose and supplement. If we use two sided t-test, so the alternate hypothesis is that the significant difference there is.

```
twotest <- function (x,y){
  with(ToothGrowth, paste("two.sided p-value",
    round(t.test(x, y, alternative = "two.sided",
        paired = F, var.equal = F, conf.level = 0.95)$p.value,5)))}
greattest <- function (x,y){
  with(ToothGrowth, paste("greater p-value",
    round(t.test(x, y, alternative = "greater",
        paired = F, var.equal = F, conf.level = 0.95)$p.value,5)))}
paste("OJ, VC:",
      with(ToothGrowth, twotest(len[supp=='OJ'], len[supp=='VC'])),",",
      with(ToothGrowth, greattest(len[supp=='OJ'], len[supp=='VC'])))
```

```
## [1] "OJ, VC: two.sided p-value 0.06063 , greater p-value 0.03032"
```

```
paste("OJ, VC | 0.5 mg:",
      with(ToothGrowth, twotest(len[supp=='OJ' & dose == 0.5], len[supp=='VC'& dose == 0.5])),",",
      with(ToothGrowth, greattest(len[supp=='OJ' & dose == 0.5], len[supp=='VC'& dose == 0.5])))
```

```
## [1] "OJ, VC | 0.5 mg: two.sided p-value 0.00636 , greater p-value 0.00318"
```

```
paste("OJ, VC | 1.0 mg:",
      with(ToothGrowth, twotest(len[supp=='OJ' & dose == 1.0], len[supp=='VC'& dose == 1.0])),",",
      with(ToothGrowth, greattest(len[supp=='OJ' & dose == 1.0], len[supp=='VC'& dose == 1.0])))
```

```
## [1] "OJ, VC | 1.0 mg: two.sided p-value 0.00104 , greater p-value 0.00052"
```

```
paste("OJ, VC | 2.0 mg:",
      with(ToothGrowth, twotest(len[supp=='OJ' & dose == 2.0], len[supp=='VC'& dose == 2.0])))
```

```
## [1] "OJ, VC | 2.0 mg: two.sided p-value 0.96385"
```

So, there is no significant difference (two sided, at 5% level) between OJ and VS in general and in 2.0 mg dose. The one there is - in 0.5 and 1.0 mg dose. But if we test one sided hypothesis we can confirm that we saw at boxplots: effect of orange juice much larger than C-vitamin in general.

Conclusions

We saw that vitamin C and orange juice have effect on tooth growth. And vitamin C has much larger effect. Increasing the supplement dose level to 1.0 leads to increased tooth growth in general.