# Part 1 - Basic Inferential Data Analysis

*Peter Maneykowski*

*4/15/2018*

Load `ToothGrowth` data set and do initial transformations.

```
library(datasets); library(ggplot2)
data("ToothGrowth")
toothGrowth <- ToothGrowth
toothGrowth$dose <- as.factor(toothGrowth$dose)  # not loaded as categorical variable
```

**Basic exploratory data analyses.**

Top of data set.

```
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
```

Bottom of data set.

```
tail(toothGrowth)
```

```
##     len supp dose
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

Unique values of `supp`: supplement type, either vitamin C (`VC`) or orange juice (`OJ`)

```
unique(toothGrowth$supp)
```

```
## [1] VC OJ
## Levels: OJ VC
```

Unique values of `dose`: dosage in milligrams (of vitamin C) per day. `dose` is delivered as either a vitamin C supplement or orange juice.

```
unique(toothGrowth$dose)
```

```
## [1] 0.5 1   2
## Levels: 0.5 1 2
```

A sorted vector of unique values of `len`: length of teeth (in arbitrary units)

```
sort(unique(toothGrowth$len))
```

```
##  [1]  4.2  5.2  5.8  6.4  7.0  7.3  8.2  9.4  9.7 10.0 11.2 11.5 13.6 14.5
## [15] 15.2 15.5 16.5 17.3 17.6 18.5 18.8 19.7 20.0 21.2 21.5 22.4 22.5 23.0
```

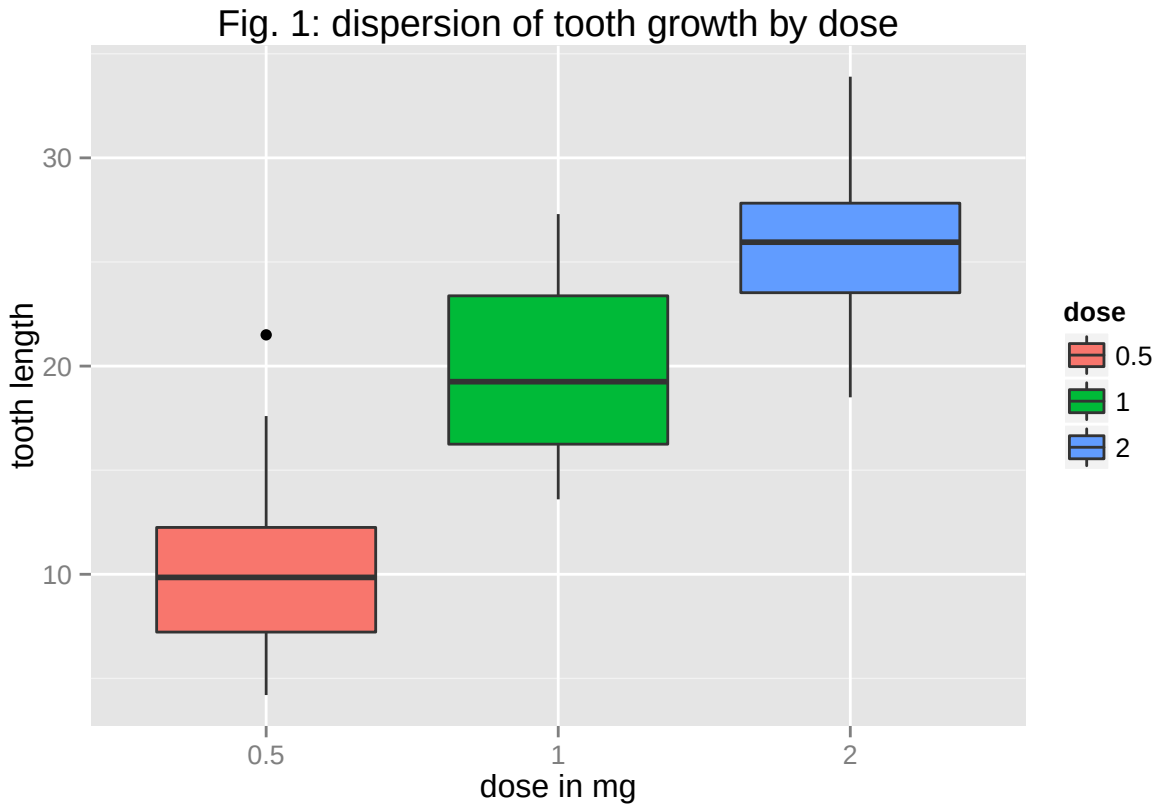
```
## [29] 23.3 23.6 24.5 24.8 25.2 25.5 25.8 26.4 26.7 27.3 29.4 29.5 30.9 32.5
## [43] 33.9
```

**Basic summary** of the data set with an accompanying boxplot demonstrating the same relationship as the summary function.

```
summary(toothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
ggplot(toothGrowth, aes(x=dose, y=len, fill=dose)) +
  geom_boxplot() +
  ggtitle("Fig. 1: dispersion of tooth growth by dose") +
  xlab("dose in mg") +
  ylab("tooth length")
```




**Confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose**:

**Hypothesis 1**: Orance juice and vitamin C produce the same tooth growth across the entire data set.

```
h1<-t.test(len ~ supp, data = toothGrowth)
h1$conf.int
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

```
h1$p.value
```

```
## [1] 0.06063451
```

`conf.int` includes `0` and `p.value` is greater than the threshold of `0.05`, so the null hypothesis that we created cannot be rejected.

**Hypothesis 2**: Orance juice and vitamin C produce the same tooth growth for a dosage of `0.5`mg/day.

```
h2<-t.test(len ~ supp, data = subset(toothGrowth, dose == 0.5))
h2$conf.int
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```
h2$p.value
```

```
## [1] 0.006358607
```

`conf.int` does not include `0`, and `p.value` is below our threshold of `0.05`, so the null hypothesis can be rejected. The alternate hypothesis that at a dosage of `0.5`mg/day orange juice produces more tooth growth than vitamin C is accepted.

**Hypothesis 3**: Oranje juice and vitamin C produce the same tooth growth for a dosage of `1.0`mg/day.

```
h3<-t.test(len ~ supp, data = subset(toothGrowth, dose == 1))
h3$conf.int
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```
h3$p.value
```

```
## [1] 0.001038376
```

`conf.int` does not include `0`, and `p.value` is below our threshold of `0.05`, so the null hypothesis can be rejected. The alternate hypothesis that at a dosage of `1.0`mg/day orange juice produces more tooth growth than vitamin C is accepted.

**Hypothesis 4**: Oranje juice and vitamin C produce the same tooth growth for a dosage of `2.0`mg/day.

```
h4<-t.test(len ~ supp, data = subset(toothGrowth, dose == 2))
h4$conf.int
```

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

```
h4$p.value
```

```
## [1] 0.9638516
```

`conf.int` does not include `0` and `p.value` is greater than the threshold of `0.05`, so the null hypothesis that we created cannot be rejected.

**Conclusions and Assumptions** By using `t.test` for hypothesis testing, we may conclude that for dosages of `0.5`mg/day and `1.0`mg/day orange juice produces more tooth growth than vitamin C. Further, we may conclude that for a dosage of `2.0`mg/day orange juice and vitamin C produce the same amount of tooth growth. Finally, when looking at the entire data set, we cannot conclude that using oranje juice is more

effective than using vitamin C. The assumptions used for testing were that there is a normal distribution of tooth lengths, and that there were no other factors that affect tooth length.