# Statistical Inference Course Project, Part 2

*Cliff Hayes*

*February 28, 2016*

## Overview

In this report, the ToothGrowth data in the R datasets package will be explored and analyzed. Using confidence intervals and hypothesis testing, tooth growth will be compared between supplement and dose. Conclusions will be drawn from the analysis and assumptions will be explained in correlation to these conclusions.

**1. Load the ToothGrowth data and perform some basic exploratory data analyses**

```
library(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 ...
```

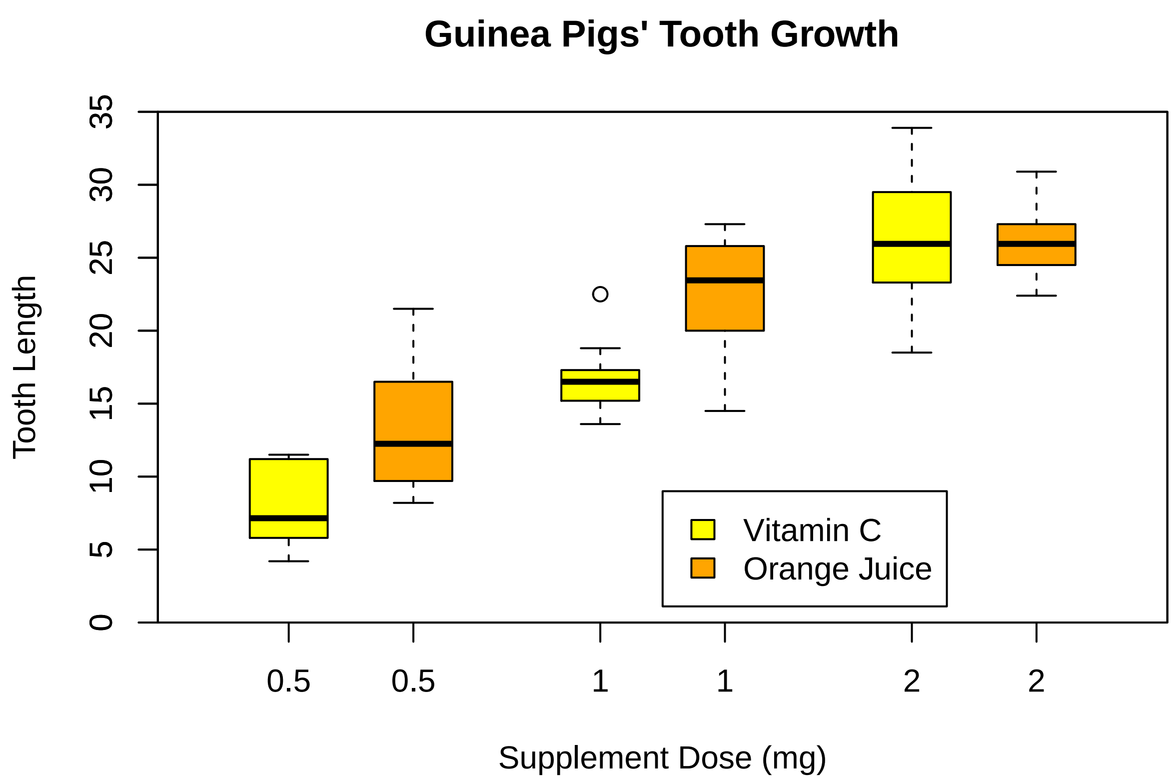
```
ToothGrowth[seq(1, 60, 10),]
```

```
##     len supp dose
## 1   4.2   VC  0.5
## 11 16.5   VC  1.0
## 21 23.6   VC  2.0
## 31 15.2   OJ  0.5
## 41 19.7   OJ  1.0
## 51 25.5   OJ  2.0
```

- Summary of ToothGrowth data
  - 60 observations of len, supp and dose
  - len = tooth length of 60 different subjects
  - supp = supplement
    - OJ = supplement of Orange Juice
    - VC = supplement of Vitamin C
  - dose = 0.5, 1.0, 2.0 mg dosage of noted supplement

```
# Lifted straight from boxplot documentation
boxplot(len ~ dose, data = ToothGrowth,
        boxwex = 0.25, at = 1:3 - 0.2,
        subset = supp == "VC", col = "yellow",
        main = "Guinea Pigs' Tooth Growth",
        xlab = "Supplement Dose (mg)",
        ylab = "Tooth Length",
        xlim = c(0.5, 3.5), ylim = c(0, 35), yaxs = "i")
boxplot(len ~ dose, data = ToothGrowth, add = TRUE,
```

```
        boxwex = 0.25, at = 1:3 + 0.2,
        subset = supp == "OJ", col = "orange")
legend(2, 9, c("Vitamin C", "Orange Juice"),
       fill = c("yellow", "orange"))
```




```
with(ToothGrowth, tapply(len, list(supp,dose), mean))
```

```
##      0.5     1     2
## OJ 13.23 22.70 26.06
## VC  7.98 16.77 26.14
```

**2. Provide a basic summary of the data.**

The boxplot above shows the tooth growth of the 60 different subjects according to the 2 different supplements which are supplied at the 3 different doses. The data appears to be telling us that the larger the supplement dosage, the more tooth growth. This is corroborated by calculating the means of each supplement by dosage.

**3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.**

Using t.test, we look for differences in the performance of the supplements. We assume a null hypothesis which claims there is no difference between the performance of the supplements.

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == .5, ])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

The t-statistic, p-value and confidence interval for the 0.5 mg doses of each supplement is shown to be statistically significant. The t-statistic is large, the p-value is small and the mean difference between OJ and VC is large. The 95% confidence interval does not contain 0. In this case, we fail to accept the null hypothesis.

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 1, ])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

The t-statistic, p-value and confidence interval for the 1.0 mg doses of each supplement is shown to be statistically significant. The t-statistic is large, the p-value is small and the mean difference between OJ and VC is large. The 95% confidence interval does not contain 0. In this case, we fail to accept the null hypothesis.

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 2, ])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

The t-statistic, p-value and confidence interval for the 2.0 mg doses of each supplement is shown to be statistically significant. The t-statistic is small, the p-value is large and the mean difference between OJ and VC is very close. The 95% confidence interval contains 0. In this case, we cannot reject the null hypothesis.

**4. State your conclusions and the assumptions needed for your conclusions.**

- Assumptions
  - The experiment is assumed to have been carried out with random assignment of subjects to the dosage categories and supplement type. This random assignment is assumed to have been carried out to minimize any unaccounted for factors that could affect tooth growth.
  - It is assumed that all 60 subjects are representative of the population of subjects.
  - The assumption made for the t-statistic is that variances are equal across the two supplements being compared.
- Conclusions
  - There is no difference between the two supplements on the effect of tooth growth. Both supplements effect tooth growth similarly.
  - The dosage of both supplements is statistically significant on tooth growth. As dosage is increased, tooth growth is increased.
  - At doses of 0.5 and 1.0 mg, there is a greater effect on tooth growth with the supplement of Orange Juice.
  - At doses of 2.0 mg, there is no difference in the effect on tooth growth with either the supplement of Orange Juice or Vitamin C.