

Statistical Inference: Course Project

Created by H.Wang on November 19, 2015

Basic Settings

```
echo= TRUE # make scripts visible to others
```

Description of the Problem

Analyze the ToothGrowth data in the R datasets package, provide data summary and use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

Load Data and Perform Basic Exploratory Data Analysis

```
# load the dataset
library(datasets)
data("ToothGrowth")

# convert column dose to factor (same as supp)
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

# Table Check
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 ...
## $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 ...

nrow(ToothGrowth)

## [1] 60
```

Provide a basic summary of the data

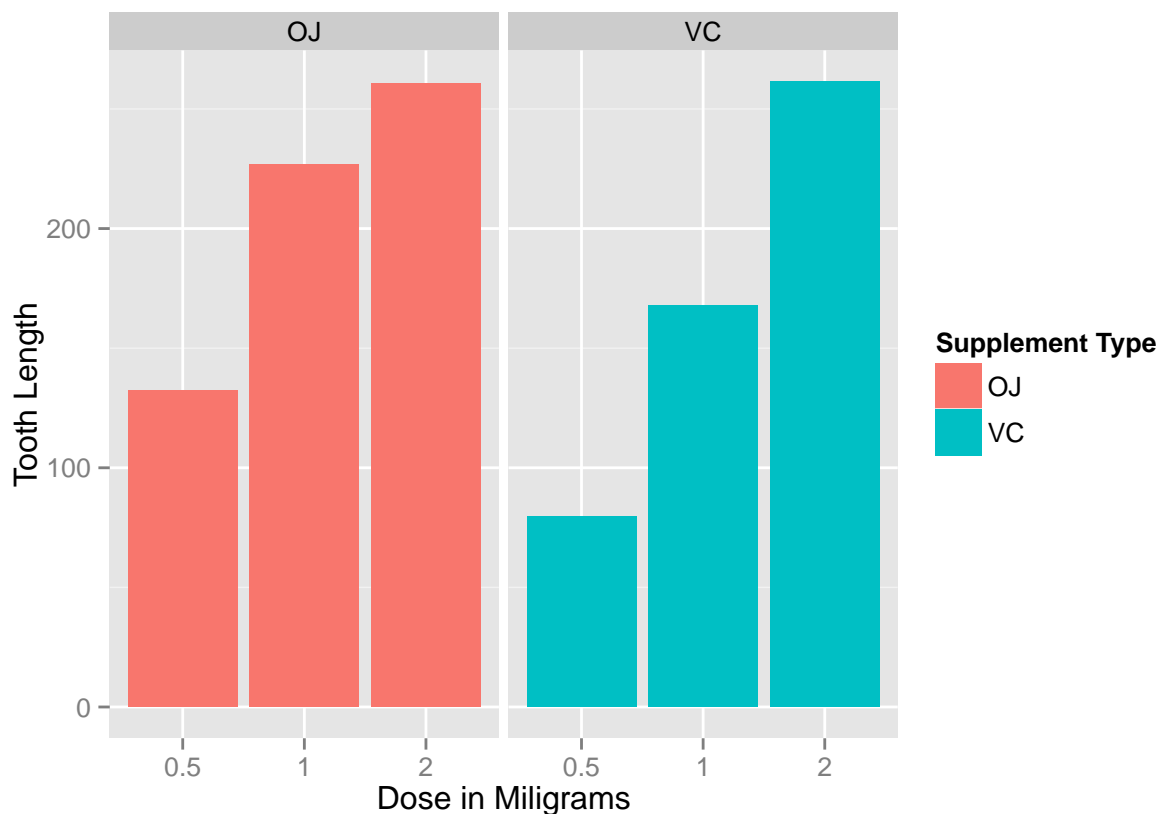
```
# summary for the dataset
summary(ToothGrowth)

##           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
```

```
# split of cases between different dose and supp level
table(ToothGrowth$dose, ToothGrowth$supp)
```

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

```
library(ggplot2)
ggplot(data=ToothGrowth, aes(x=dose, y=len, fill=supp)) +
  geom_bar(stat="identity",) + facet_grid(. ~ supp) + xlab("Dose in Miligrams") + ylab("Tooth Length")
```



There is a positive correlation between tooth length and the dose level of VC and OJ.

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

Use 95% confidence intervals for two variables and the intercept are as follows:

```
# linear regression for len vs dose and supp
fit <- lm(len ~ dose + supp, data = ToothGrowth)
# confidence intervals
confint(fit)
```

```
##           2.5 %    97.5 %
## (Intercept) 10.475238 14.434762
## dose1       6.705297 11.554703
## dose2      13.070297 17.919703
## suppVC      -5.679762 -1.720238
```

The confidence level is 95% which means 95% of time the simulation results(coefficient estimation) will be in these ranges. All p-values are less than 0.05, rejecting the null hypothesis and suggesting that each variable explains a variability in tooth length.

```
summary(fit)
```

```
##
## Call:
## lm(formula = len ~ dose + supp, data = ToothGrowth)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.085 -2.751 -0.800  2.446  9.650
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   12.4550     0.9883   12.603 < 2e-16 ***
## dose1         9.1300     1.2104    7.543 4.38e-10 ***
## dose2        15.4950     1.2104   12.802 < 2e-16 ***
## suppVC        -3.7000     0.9883   -3.744 0.000429 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.828 on 56 degrees of freedom
## Multiple R-squared:  0.7623, Adjusted R-squared:  0.7496
## F-statistic: 59.88 on 3 and 56 DF,  p-value: < 2.2e-16
```

The intercept is 12.455, meaning without supp the average tooth length will be 12.455 units. The coefficient of dose is 9.13 . As long as dose is increased by 1 mg (no change in supp) the tooth length would increase 9.13 units. The last one is for supp VC and it is negative (-3.7), which could be interpreted as with delivery a given supp VC (without changing dose) the tooth length will drop 3.7 units. We could also conclude that on average, delivery of supp OJ would increase the tooth length by 3.7 units due to two category under supp.