

Statistical Inference Project Part 2: Basic Inferential Data Analysis

P.Y.Fong

12/24/2020

Overview:

In the second part of the project, we analyze the ToothGrowth dataset from the R datasets package.

Loading Libraries and Data.

```
library(ggplot2)
library(datasets)
data(ToothGrowth)
attach(ToothGrowth)
```

Exploration of dataset:

```
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: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

ToothGrowth is a dataframe with 60 observations relating 3 variables which explores the effect of Vitamin C on tooth growth in Guinea Pigs.

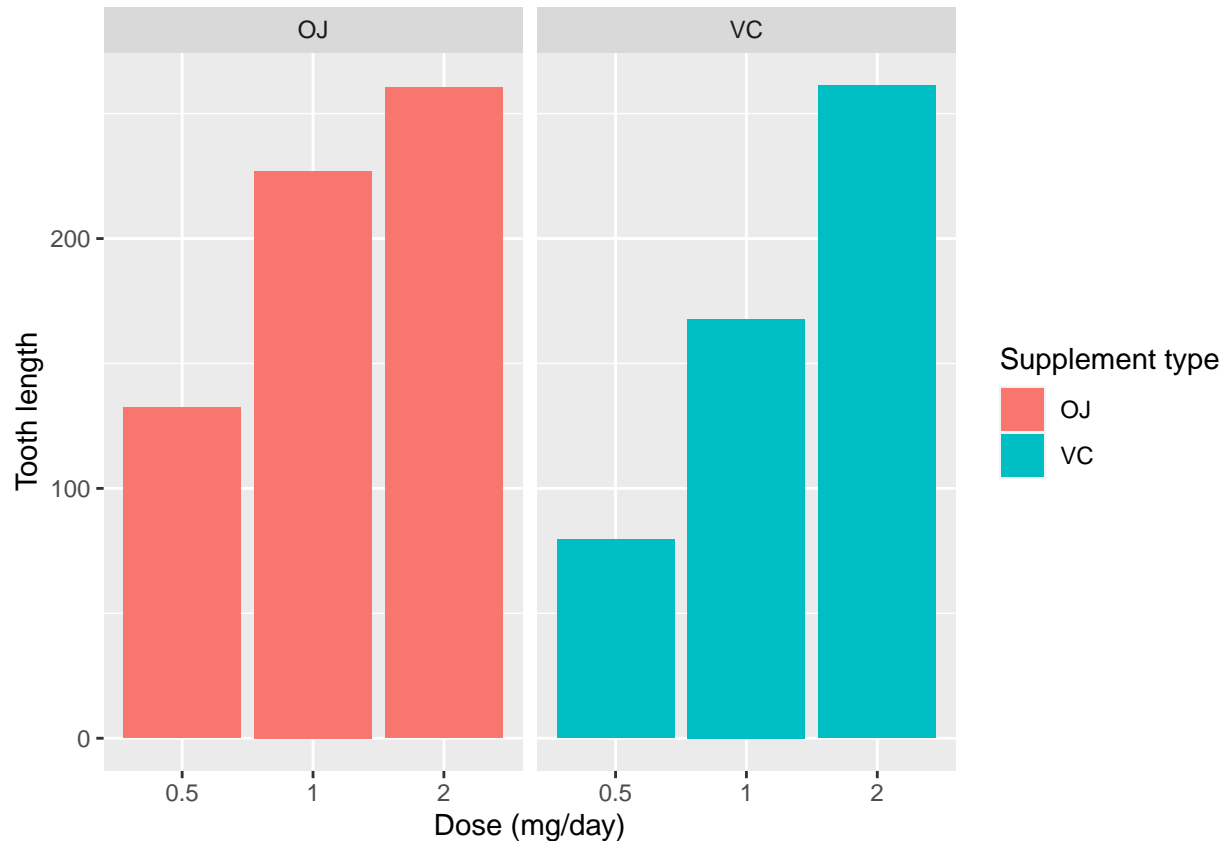
Variables:

1. len - numeric Tooth length. (response variable).
2. supp - factor Supplement type. OJ(Orange Juice) or VC (Ascorbic Acid).
3. dose - numeric Dose in milligrams/day.

Information about the dataset can be found at

<https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html>

```
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
  geom_bar(stat="identity") +
  facet_grid(. ~ supp) +
  xlab("Dose (mg/day)") +
  ylab("Tooth length") +
  guides(fill=guide_legend(title="Supplement type"))
```



From the above plots, there is a clear positive correlation between tooth length and daily Vitamin C dosage for both supplement types.

The question to be addressed then should be, whether the supplement type (i.e. OJ-orange juice or VC-ascorbic acid) has any effect on tooth length.

Analysis

Hypothesis evaluation using the Welch Two Sample t-test:

- Null Hypothesis: No significant difference between the means of **Supplement Type** and **Tooth Length**.
 - H_0 : $\text{mean}(\text{VC}) = \text{mean}(\text{OJ})$.
- Alt Hypothesis: Significant difference between **Supplement Type** and **Tooth Length**.
 - H_a : $\text{mean}(\text{VC}) <> \text{mean}(\text{OJ})$.

Assumptions:

1. The sample population, i.e. the 60 guinea pigs, are representative of the population of guinea pigs in order to generalize the results.
2. Observations were made based on random assignment of guinea pigs.
3. Variances in the samples are assumed to be different for the groups differentiated by supplement Type.

Testing:

```
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth)

##
## Welch Two Sample t-test
##
## data: len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1710156 7.5710156
## sample estimates:
## mean in group OJ mean in group VC
## 20.66333 16.96333
```

The 95% confidence interval is [-0.1710156, 7.5710156].

The p value = 0.06063, > 0.05 (alpha = 5%).

Decision: Fail to reject H_0 .

Conclusions:

1. Dosage is positively correlated to increase in Tooth Length.
2. Supplement Type‘ has no significant impact on Tooth Length with 95% confidence.