Statistical Inference Course Project (Part 2)

Jeffrey M. Hunter

11 May, 2019

Contents

Overview

In this second part of a two-part project assignment, we are being asked to investigate the effect of vitamin C on tooth growth in guinea pigs.

This analysis will compare the effects of different doses of vitamin C using two delivery methods to study the tooth growth in guinea pigs. The sample will consist of 60 guinea pigs. Each guinea pig will receive one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid.

The study will be performed using the ToothGrowth dataset which is included in the R `datasets` package. The dataset is a data frame that contains 60 observations and 3 variables:

- len - A numeric vector indicating the measurement of tooth length after Vitamin C delivery
- supp - A factor vector describing the delivery method (supplement type) used. Either Orange Juice (OJ) or Ascorbic Acid (VC)
- dose - A numeric vector indicating the dosage level in milligrams (0.5, 1, or 2mg)

Further information on the ToothGrowth dataset can be found in the R documentation using `?ToothGrowth`.

Environment Setup

Load packages used in this analysis.

```
if (!require(ggplot2)) {
    install.packages("ggplot2", repos = "http://cran.us.r-project.org")
    library(ggplot2)
}
```

```
## Loading required package: ggplot2
```

```
if (!require(dplyr, warn.conflicts = FALSE)) {
    install.packages("dplyr", repos = "http://cran.us.r-project.org")
    library(dplyr, warn.conflicts = FALSE)
}
```

```
## Loading required package: dplyr
```

Load ToothGrowth dataset.

```
library(datasets)
data(ToothGrowth)
```

Basic Data Summary

After loading the ToothGrowth dataset, provide a basic summary of the data.

```
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)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
# tabulate supplement type and dosage level values
table(ToothGrowth$dose, ToothGrowth$supp)
```

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

```
# summary of tooth length data grouped by supplement type and dosage level
by(data = ToothGrowth$len, INDICES = list(ToothGrowth$supp, ToothGrowth$dose), summary)
```

```
## : OJ
## : 0.5
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.20    9.70   12.25   13.23   16.18   21.50
## ------------------------------------------------------------
## : VC
## : 0.5
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20    5.95    7.15    7.98   10.90   11.50
## ------------------------------------------------------------
## : OJ
## : 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   14.50   20.30   23.45   22.70   25.65   27.30
## ------------------------------------------------------------
## : VC
## : 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.60   15.28   16.50   16.77   17.30   22.50
```

```
## ------------------------------------------------------------
## : OJ
## : 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   22.40   24.57   25.95   26.06   27.07   30.90
## ------------------------------------------------------------
## : VC
## : 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.50   23.38   25.95   26.14   28.80   33.90
```

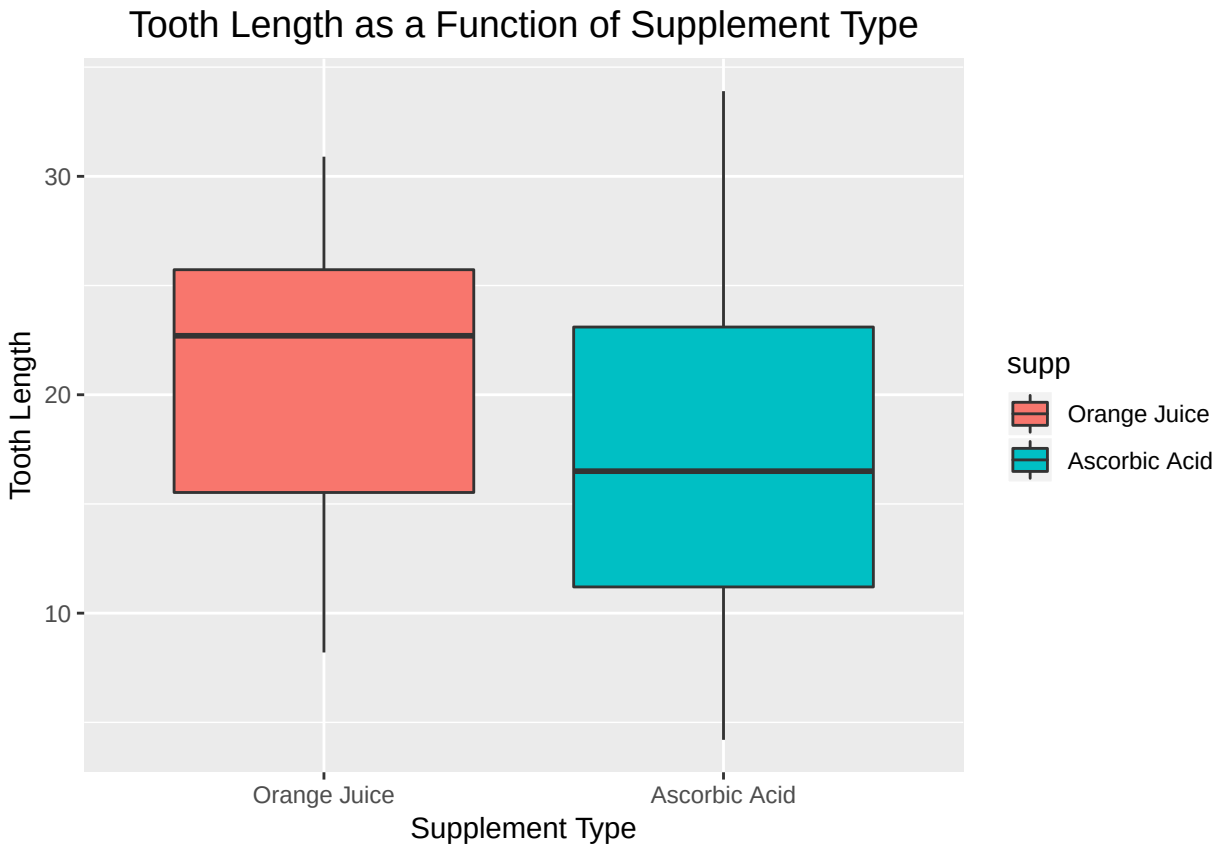
Exploratory Data Analysis

Perform some basic exploratory data analyses of the data. The analyses will explore the following relations:

1. Tooth Length (len) as a function of Supplement Type (supp)
2. Tooth Length (len) as a function of Dosage Level (dose)
3. Tooth Length (len) as a function of Supplement Type (supp) and Dosage Level (dose)

Tooth Length to Supplement Type

```
tg <- ToothGrowth
levels(tg$supp) <- c("Orange Juice", "Ascorbic Acid")
gLenSupp <- ggplot(data = tg, aes(x = supp, y = len)) +
    geom_boxplot(aes(fill = supp)) +
    xlab("Supplement Type") +
    ylab("Tooth Length") +
    theme(plot.title = element_text(size = 14, hjust = 0.5)) +
    ggtitle("Tooth Length as a Function of Supplement Type")
print(gLenSupp)
```

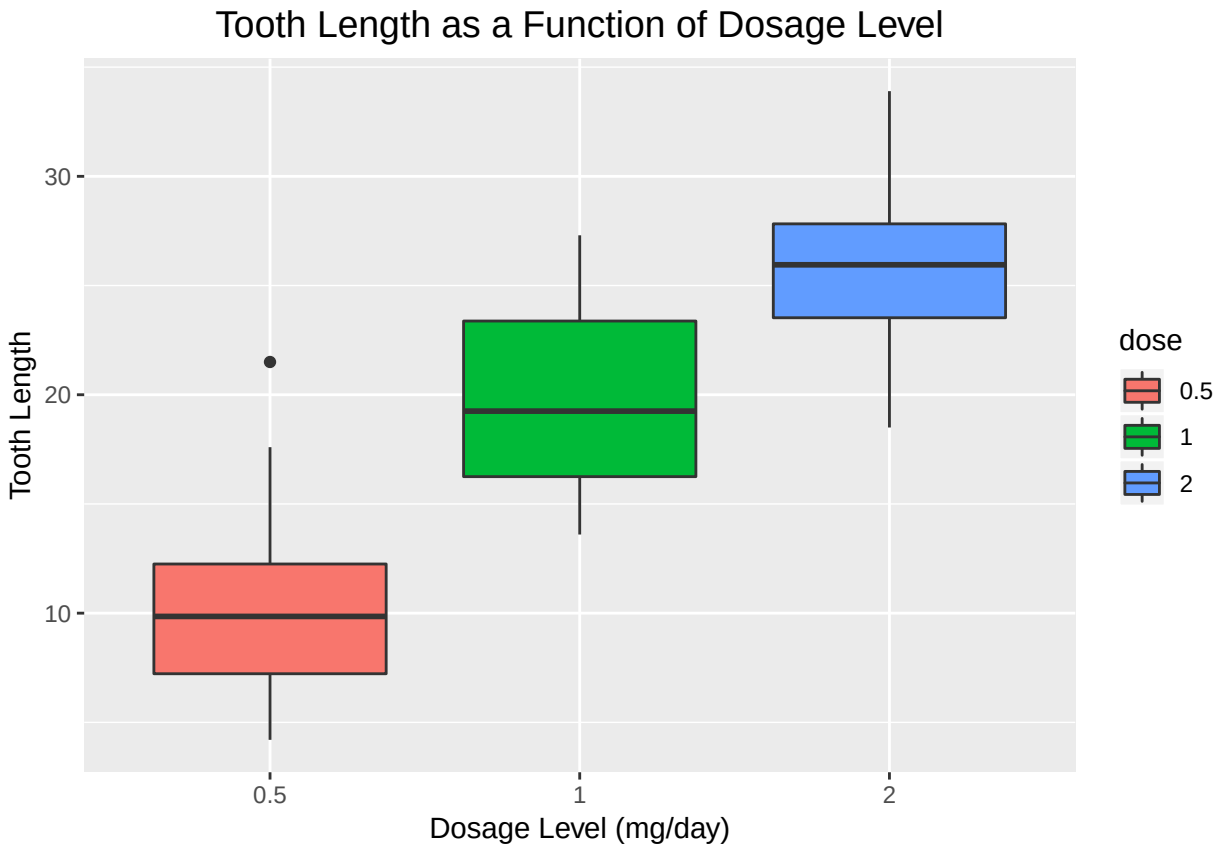


Observation

The above chart shows that using Orange Juice (OC) as the delivery method, independent of dosage level, had a more favorable effect on tooth growth than Ascorbic Acid (VC) .

Tooth Length to Dosage Level

```
gLenDose <- ggplot(data = ToothGrowth, aes(x = factor(dose), y = len)) +
    geom_boxplot(aes(fill = factor(dose))) +
    xlab("Dosage Level (mg/day)") +
    ylab("Tooth Length") +
    guides(fill=guide_legend(title="dose")) +
    theme(plot.title = element_text(size = 14, hjust = 0.5)) +
    ggtitle("Tooth Length as a Function of Dosage Level")
print(gLenDose)
```

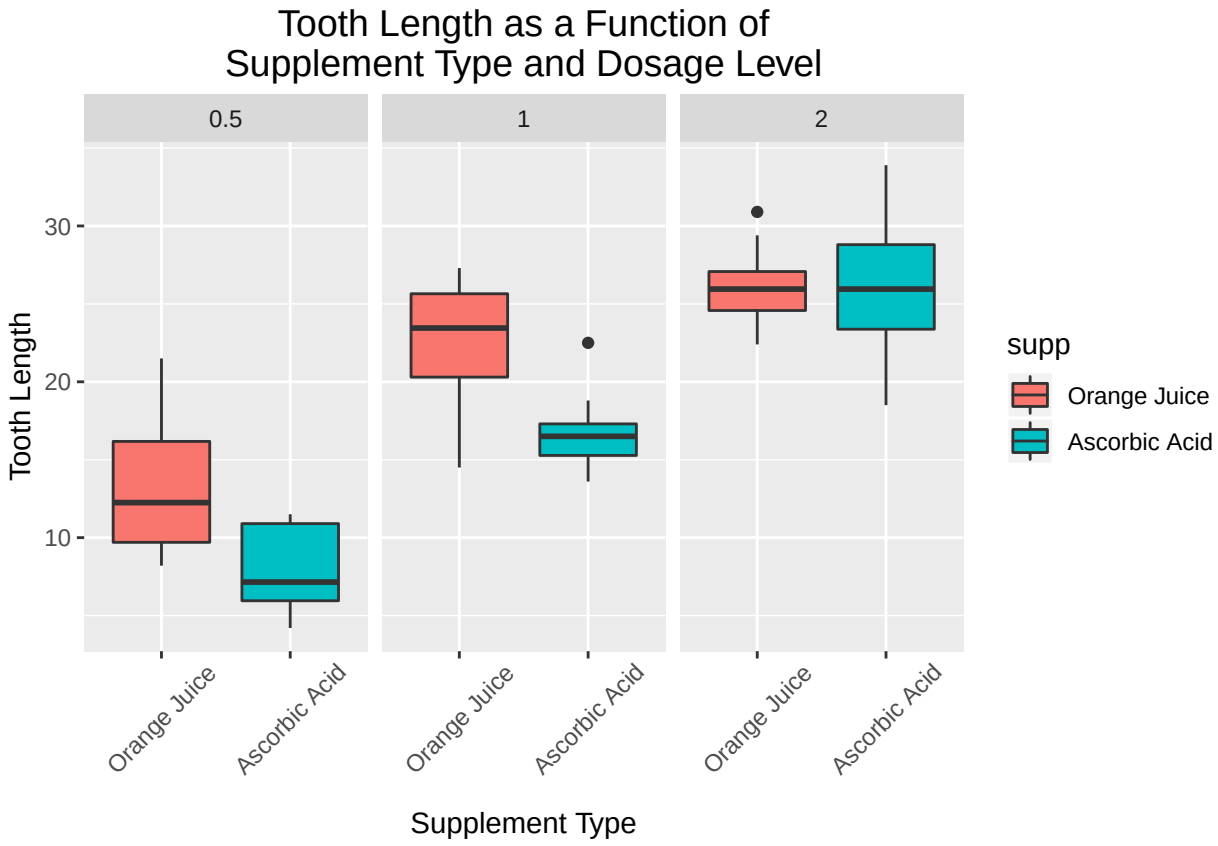


Observation

The above chart shows a positive relationship that higher dosage levels of vitamin C, independent of delivery method, had a more favorable effect on tooth growth than lower dosages of vitamin C. The 2.0 mg/day dosage level had the best effect on tooth growth followed by 1.0 mg/day and then 0.5 mg/day.

Tooth Length to Supplement Type and Dosage Level

```
tg <- ToothGrowth
levels(tg$supp) <- c("Orange Juice", "Ascorbic Acid")
gLenSuppDose <- ggplot(data = tg, aes(x = supp, y = len)) +
    geom_boxplot(aes(fill = supp)) +
    facet_wrap(~ dose) +
    xlab("Supplement Type") +
    ylab("Tooth Length") +
    guides(fill=guide_legend(title="supp")) +
    theme(plot.title = element_text(size = 14, hjust = 0.5, vjust = 0.5),
          axis.text.x = element_text(angle = 45,
                                     hjust = 0.5,
                                     vjust = 0.5,
                                     margin = margin(b = 10))) +
    ggtitle("Tooth Length as a Function of\nSupplement Type and Dosage Level")
print(gLenSuppDose)
```



Observation

Looking at the above chart which shows tooth growth as a function of delivery method and dosage level, it appears that Orange Juice is more effective than Ascorbic Acid as the delivery method when the dosage level is 0.5 to 1.0 mg/day. The higher dosage level of 2.0 mg/day is more effective than the lower dosages, but both delivery methods are equally as effective.

Confidence Intervals

Confidence Interval test that the two supps are equal. Null hypothesis is that there is no difference in length which would imply that the 95% confidence interval should contain 0.

Hypothesis Test

Conclusion