

Statistical Inference Project_Part 2

Moh Ahmed

06/02/2017

Overview

This report analyses the toothgrowth dataset in the R datasets packages. The aim is to perform basic exploratory data analyses and provide a basic summary of the data. Then, using confidence intervals and/or hypothesis tests to compare tooth growth by Supplement type and Dose in milligrams/day.

In this dataset, the response variable is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, either orange juice or ascorbic acid.

Summary of the data

ToothGrowth data is a data frame with 60 observations on 3 variables.

[,1] len numeric Tooth length [,2] supp factor Supplement type (VC or OJ). [,3] dose numeric Dose in milligrams/day

```
library(ggplot2)
library(datasets)
library(gridExtra)
```

```
tooth_growth_dt <- ToothGrowth
str(tooth_growth_dt)
```

```
## '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(tooth_growth_dt)
```

```
##      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
```

Exploratory Data Analyses

```

tooth_growth_dt <-
  data.frame(
    len = tooth_growth_dt$len,
    supp = tooth_growth_dt$supp,
    dose = as.factor(tooth_growth_dt$dose)
  )

plot1 <-
  ggplot(tooth_growth_dt, aes(x = dose, y = len, fill = dose)) +
  geom_boxplot() +
  facet_grid(. ~ tooth_growth_dt$supp) +
  ggtitle('Tooth Length vs Dose vs Supplement') +
  labs(fill = 'Dose (mg)') +
  xlab('Dose (mg)') +
  ylab('Tooth Length (mm)') +
  theme(legend.position = 'none') +
  theme(plot.title = element_text(size = 12, hjust = 0)) +
  theme(axis.title = element_text(size = 8))

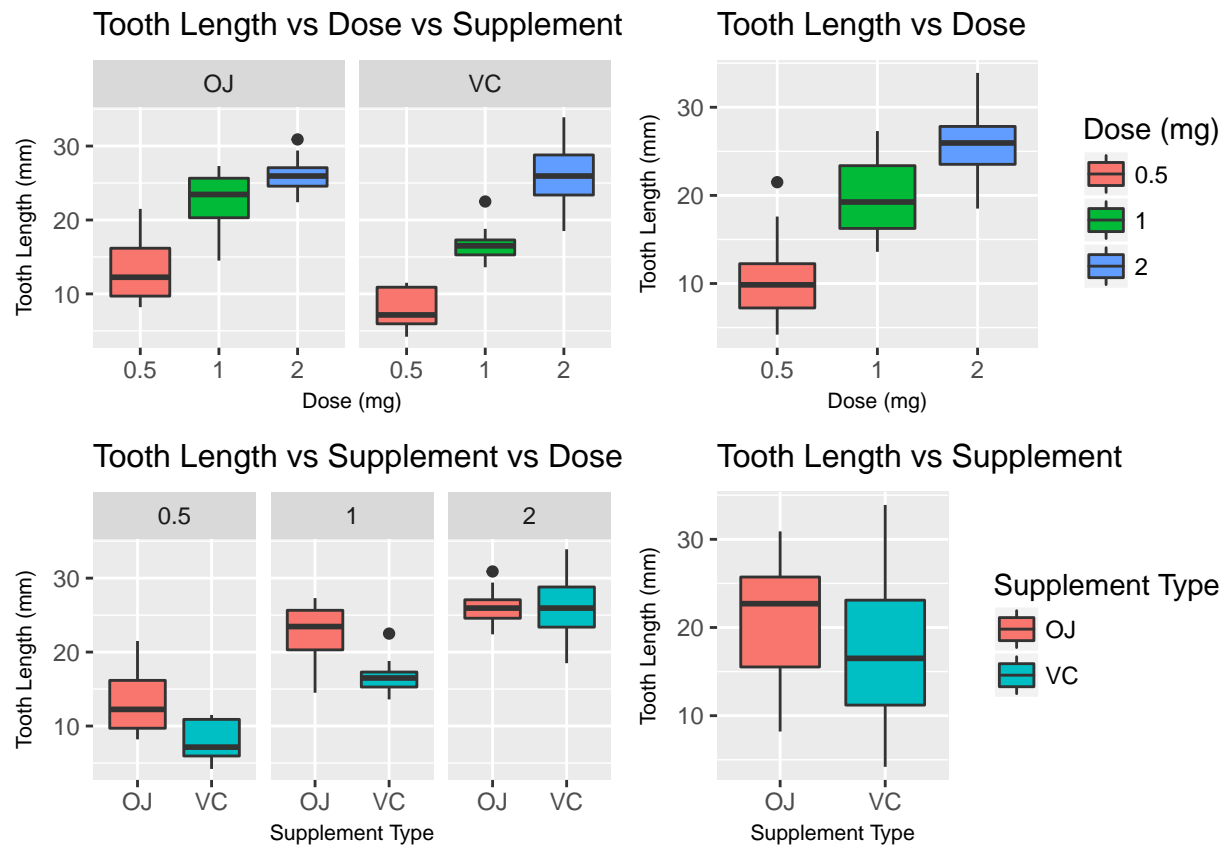
plot2 <-
  ggplot(tooth_growth_dt, aes(x = dose, y = len, fill = dose)) +
  geom_boxplot() + ggtitle('Tooth Length vs Dose') +
  labs(fill = 'Dose (mg)') +
  xlab('Dose (mg)') +
  ylab('Tooth Length (mm)') +
  theme(plot.title = element_text(size = 12, hjust = 0)) +
  theme(axis.title = element_text(size = 8))

plot3 <-
  ggplot(tooth_growth_dt, aes(x = supp, y = len, fill = supp)) +
  geom_boxplot() +
  facet_grid(. ~ tooth_growth_dt$dose) +
  ggtitle('Tooth Length vs Supplement vs Dose') +
  labs(fill = 'Supplement Type') + xlab('Supplement Type') +
  ylab('Tooth Length (mm)') +
  theme(legend.position = 'none') +
  theme(plot.title = element_text(size = 12, hjust = 0)) +
  theme(axis.title = element_text(size = 8))

plot4 <-
  ggplot(tooth_growth_dt, aes(x = supp, y = len, fill = supp)) +
  geom_boxplot() +
  ggtitle('Tooth Length vs Supplement') +
  labs(fill = 'Supplement Type') +
  xlab('Supplement Type') +
  ylab('Tooth Length (mm)') +
  theme(plot.title = element_text(size = 12, hjust = 0)) +
  theme(axis.title = element_text(size = 8))

grid.arrange(plot1, plot2, plot3, plot4, nrow = 2, ncol = 2)

```



Hypotheses Testing

Hypothesis 1: Orange Juice vs Ascorbic Acid effect on Tooth growth

Our null hypothesis states that the effect supplement type is the same on the tooth growth length.

Our alternative hypothesis states that the effect supplement type is not the same on the tooth growth length.

```
t.test(len~supp, data = tooth_growth_dt)
```

```
##
##  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
```

Performing T test, **we fail to reject the null hypothesis** since our p-value > 0.05, hence the difference is not statistically significant.

Hypothesis 2: Dosage of 0.5 mg/day of Orange Juice vs Ascorbic Acid effect on Tooth growth

Our null hypothesis states that at dosage of 0.5 mg/day, the effect of supplement type is the same on the tooth growth length.

Our alternative hypothesis states that at dosage of 0.5 mg/day, the effect of supplement type on the tooth growth length is not the same.

```
t.test(len~supp, data = subset(tooth_growth_dt, dose ==0.5))
```

```
##
##  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
```

Performing T test, **we reject the null hypothesis** since our p-value < 0.05, hence the difference is statistically significant.

Hypothesis 3: Dosage of 1.0 mg/day of Orange Juice vs Ascorbic Acid effect on Tooth growth

Our null hypothesis states that at dosage of 1.0 mg/day, the effect of supplement type is the same on the tooth growth length.

Our alternative hypothesis states that at dosage of 1.0 mg/day, the effect of supplement type on the tooth growth length is not the same.

```
t.test(len~supp, data = subset(tooth_growth_dt, 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
```

Performing T test, **we reject the null hypothesis** since our p-value < 0.05, hence the difference is statistically significant.

Hypothesis 4: Dosage of 2.0 mg/day of Orange Juice vs Ascorbic Acid effect on Tooth growth

Our null hypothesis states that at dosage of 2.0 mg/day, the effect of supplement type is the same on the tooth growth length.

Our alternative hypothesis states that at dosage of 2.0 mg/day, the effect of supplement type on the tooth growth length is not the same.

```
t.test(len~supp, data = subset(tooth_growth_dt, dose ==2.0))

##
##  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
```

Performing T test, **we fail to reject the null hypothesis** since our p-value > 0.05 , hence the difference is not statistically significant.

Conculusions

From Hypothesis 1, we cannot decide whether Orange Juice or Ascorbic Acid is more effective on tooth growth. Since the difference in their effect is not statistically significant.

In addition from Hypothesis 4, we see that at dosage 2.0 mg/day, the effect of the supplement type on tooth growth length is not statistically significant.

However from Hypothesis 2 and 3, we can deduce that Orange Juice at dosage of 0.5 mg/day and 1 mg/day is more effective on tooth growth length than its Ascorbic Acid. This is because the difference in their effectiveness is statistically significant.

Assumptions

1. The data follows a normal distribution.
2. The samples are randomly distributed.