Analysis of ToothGrowth data

Thammanoon Kawinfruangfukul

Overview

This report provides a summary of ToothGrowth data to highlight basic features of the data. Moreover, this report shows assumptions, solutions, and conclusions of confidence intervals and hypothesis tests to compare tooth growth of 60 guinea pigs by supplement and dosage.

Exploratory Data Analysis

```r
# Load the ToothGrowth data
data(ToothGrowth)
data <- data.frame(ToothGrowth)

# Display the overview structure of the ToothGrowth 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 ...
```

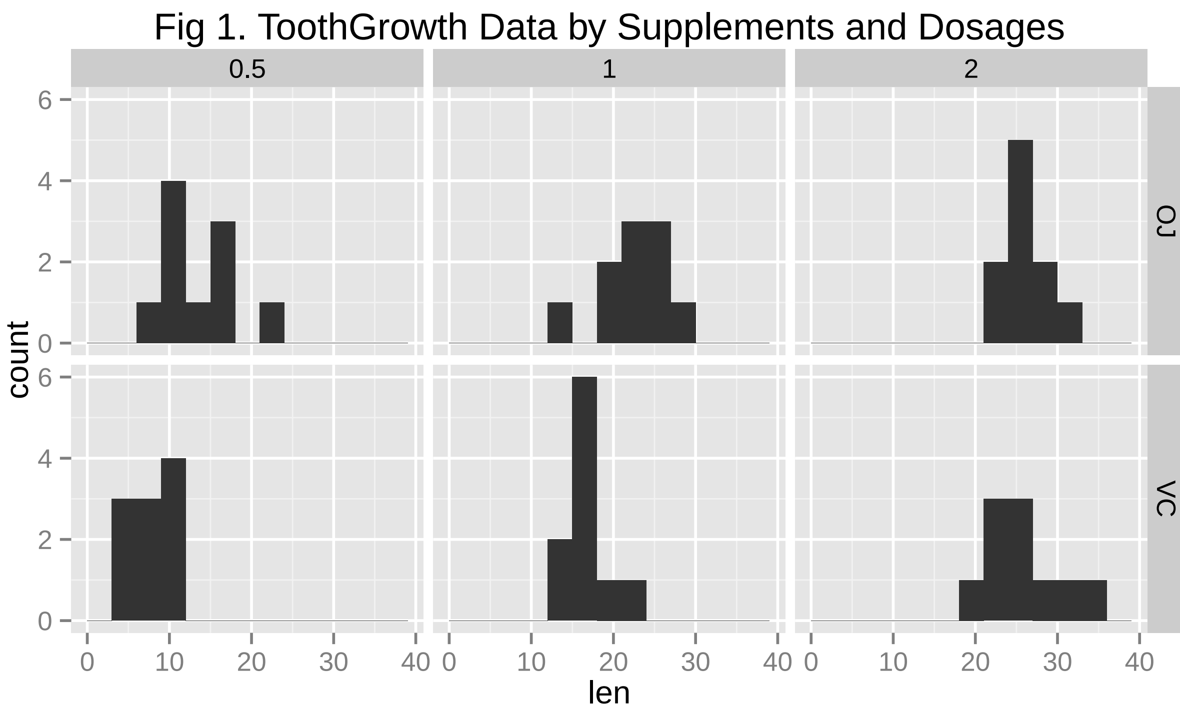
the ToothGrowth data was collected from 60 guinea pigs. The researchers measured the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with two delivery methods: ascorbic acid, orange juice coded VC, OJ.

```r
# Calculate means of the data of each of 10 guinea pigs by supplements and dosages.
means <- aggregate(data$len, list(data$supp, data$dose), mean)
colnames(means) <- c ("supp", "dose", "mean_len")
means
```

```
##   supp dose mean_len
## 1   OJ  0.5    13.23
## 2   VC  0.5     7.98
## 3   OJ  1.0    22.70
## 4   VC  1.0    16.77
## 5   OJ  2.0    26.06
## 6   VC  2.0    26.14
```

Following figure show the data of each of 10 guinea pigs by supplements and dosages.

```r
library(ggplot2)
base <- qplot(x = len, data = data, facets = supp ~ dose, binwidth = 3)
base <- base + labs(title = "Fig 1. ToothGrowth Data by Supplements and Dosages")
print(base)
```



Confidence Intervals and Hypothesis Tests

Assumptions

- ToothGrowth follows a normal distribution.
- Non-equal variances among all groups in this dataset

Comparisons by Supplements

- H0: Orange juice is as effective as ascorbic acid when the guinea pigs were applied the same dosages. (Average length with orange juice is equal to average length with ascorbic acid.)

```
supp_0_5 <- t.test(len ~ supp, paired=FALSE, data=data[data$dose == 0.5,])
supp_1_0 <- t.test(len ~ supp, paired=FALSE, data=data[data$dose == 1,])
supp_2_0 <- t.test(len ~ supp, paired=FALSE, data=data[data$dose == 2,])
supps <- data.frame("p-value" = c(supp_0_5$p.value, supp_1_0$p.value, supp_2_0$p.value),
                    "Conf-Lo" = c(supp_0_5$conf[1], supp_1_0$conf[1], supp_2_0$conf[1]),
                    "Conf-Hi" = c(supp_0_5$conf[2], supp_1_0$conf[2], supp_2_0$conf[2]))
supps
```

```
##       p.value   Conf.Lo  Conf.Hi
## 1 0.006358607  1.719057 8.780943
## 2 0.001038376  2.802148 9.057852
## 3 0.963851589 -3.798070 3.638070
```

Giving orange juice and ascorbic acid at dosages 0.5 and 1.0 does not give the same effect in legth because p-values are less than 0.05 and the confidence intervals do not include 0. However, based on p-value and confidence interval, giving orange juice and ascorbic acid at dosage 2.0 provides the same effect in length.

Comparisons by Dosages of Orange Juice

- H0: Different dosages in using orange juice give the same effect.

```
oj <- data[data$supp =="OJ", ]
oj_0_5_1_0 <- t.test(oj[oj$dose == 0.5, 1], oj[oj$dose == 1, 1], paired=FALSE)
oj_0_5_2_0 <- t.test(oj[oj$dose == 0.5, 1], oj[oj$dose == 2, 1], paired=FALSE)
oj_1_0_2_0 <- t.test(oj[oj$dose == 1, 1], oj[oj$dose == 2, 1], paired=FALSE)
ojs <- data.frame("p-value" = c(oj_0_5_1_0$p.value, oj_0_5_2_0$p.value, oj_1_0_2_0$p.value),
                  "Conf-Lo" = c(oj_0_5_1_0$conf[1], oj_0_5_2_0$conf[1], oj_1_0_2_0$conf[1]),
                  "Conf-Hi" = c(oj_0_5_1_0$conf[2], oj_0_5_2_0$conf[2], oj_1_0_2_0$conf[2]))
ojs
```

```
##        p.value    Conf.Lo    Conf.Hi
## 1 8.784919e-05 -13.415634 -5.5243656
## 2 1.323784e-06 -16.335241 -9.3247594
## 3 3.919514e-02  -6.531443 -0.1885575
```

Since p-values are less than 0.05 and confidence intervals do not include 0 in all tests (differences between 0.5 and 1.0, 0.5 and 2.0, 1.0 and 2.0), it said that these comparisons do not provide the same effect.

Comparisons by Dosages of Ascorbic Acid

- H0: Different dosages in using ascorbic acid give the same effect.

```
vc <- data[data$supp =="VC", ]
vc_0_5_1_0 <- t.test(vc[vc$dose == 0.5, 1], vc[vc$dose == 1, 1], paired=FALSE)
vc_0_5_2_0 <- t.test(vc[vc$dose == 0.5, 1], vc[vc$dose == 2, 1], paired=FALSE)
vc_1_0_2_0 <- t.test(vc[vc$dose == 1, 1], vc[vc$dose == 2, 1], paired=FALSE)
vcs <- data.frame("p-value" = c(vc_0_5_1_0$p.value, vc_0_5_2_0$p.value, vc_1_0_2_0$p.value),
                  "Conf-Lo" = c(vc_0_5_1_0$conf[1], vc_0_5_2_0$conf[1], vc_1_0_2_0$conf[1]),
                  "Conf-Hi" = c(vc_0_5_1_0$conf[2], vc_0_5_2_0$conf[2], vc_1_0_2_0$conf[2]))
vcs
```

```
##        p.value   Conf.Lo    Conf.Hi
## 1 6.811018e-07 -11.26571  -6.314288
## 2 4.681577e-08 -21.90151 -14.418488
## 3 9.155603e-05 -13.05427  -5.685733
```

Since p-values are less than 0.05 and confidence intervals do not include 0 in all tests (differences between 0.5 and 1.0, 0.5 and 2.0, 1.0 and 2.0), it said that these comparisons do not provide the same effect.