The ToothGrowth data Analysis

Kiattisak Chaisomboon

07/06/2020

Overview

We are going to analyze the ToothGrowth data in the R datasets package. We will study the change of the tooth growth in pigs depends on supplement type and dose quantity.

Exploratory Data Analysis

Loading the dataset and check its dimension

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
tg <- ToothGrowth
dim(tg)
```

```
## [1] 60  3
```

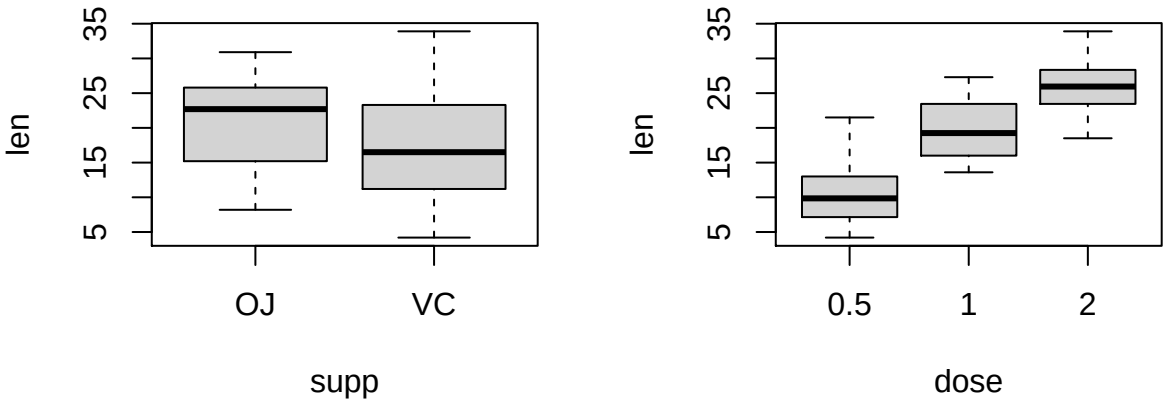
The dataset contains 60 observations and 3 variables are:
len = Tooth length
supp = Supplement type (VC or OJ)
dose = Dose in milligrams/day

Summarizing the basic information from the dataset.

```
summary(tg)
```

```
##       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
```

```
par(mfrow=c(1,2))
boxplot(len ~ supp, data=tg)
boxplot(len ~ dose, data=tg)
```



The results show that:
- Using the orange juice give a tooth growth rate higher than using the acid.
- Increasing dosage give a higher tooth growth rate.

We'll set the hypothesis testing related with this information.

The Change of Tooth Length based on the Type of Supplement

Create variables based on the type of supplement.

```
OJ <- subset(tg, supp == 'OJ')
VC <- subset(tg, supp == 'VC')
```

We hypothesize that the mean of tooth length when using an orange juice (OC) is higher than the mean of tooth length when using an ascorbic acid (VC), then we set:
$H_0 : \mu_{oc} = \mu_{vc}$
$H_\alpha : \mu_{oc} > \mu_{vc}$
Set the confidence interval = 95% ($\alpha = 0.05$)

```
t.test(OJ$len, VC$len, alternative='greater', paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ$len and VC$len
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

The output reports that the p-value = 0.03032, is less than $\alpha = 0.05$ then we'll reject the null hypothesis. We are 95% confident that the mean of tooth length when using the orange juice is higher than the mean of tooth length when using the ascorbic acid.

The Change of Tooth Length based on Level of Dose

Create variables based on level of dose

```
half.dose <- subset(tg, dose == 0.5)
one.dose <- subset(tg, dose == 1)
two.dose <- subset(tg, dose == 2)
```

We hypothesize that the mean of tooth length when using an one dosage is higher than the mean of tooth length when using a half dosage, then we set:
$H_0 : \mu_{1.0D} = \mu_{0.5D}$
$H_\alpha : \mu_{1.0D} > \mu_{0.5D}$
Set the confidence interval = 95% ($\alpha = 0.05$)

```
t.test(one.dose$len, half.dose$len, alternative='greater', paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  one.dose$len and half.dose$len
## t = 6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.753323      Inf
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

The output reports that the p-value = 6.342e-08, is less than $\alpha = 0.05$ then we'll reject the null hypothesis. We are 95% confident that the mean of tooth length when using an one dosage is higher than the mean of tooth length when using a half dosage.

Next, we hypothesize that the mean of tooth length when using a double dosage is higher than the mean of tooth length when using an one dosage, then we set:
$H_0 : \mu_{2.0D} = \mu_{1.0D}$
$H_\alpha : \mu_{2.0D} > \mu_{1.0D}$
Set the confidence interval = 95% ($\alpha = 0.05$)

```
t.test(two.dose$len, one.dose$len, alternative='greater', paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  two.dose$len and one.dose$len
## t = 4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  4.17387     Inf
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

The output reports that the p-value = 9.532e-06, is less than $\alpha = 0.05$ then we'll reject the null hypothesis. We are 95% confident that the mean of tooth length when using a double dosage is higher than the mean of tooth length when using an one dosage.