Statistical Inference Course Project: Part 2

Dongying

8/20/2020

Basic Inferential Data Analysis Instructions

Now in the second portion of the project, I'm going to analyze the ToothGrowth data in the R datasets package.

First, let's load the ToothGrowth data and perform some basic exploratory data analyses Provide a basic summary of the data.

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

```
head(ToothGrowth)
```

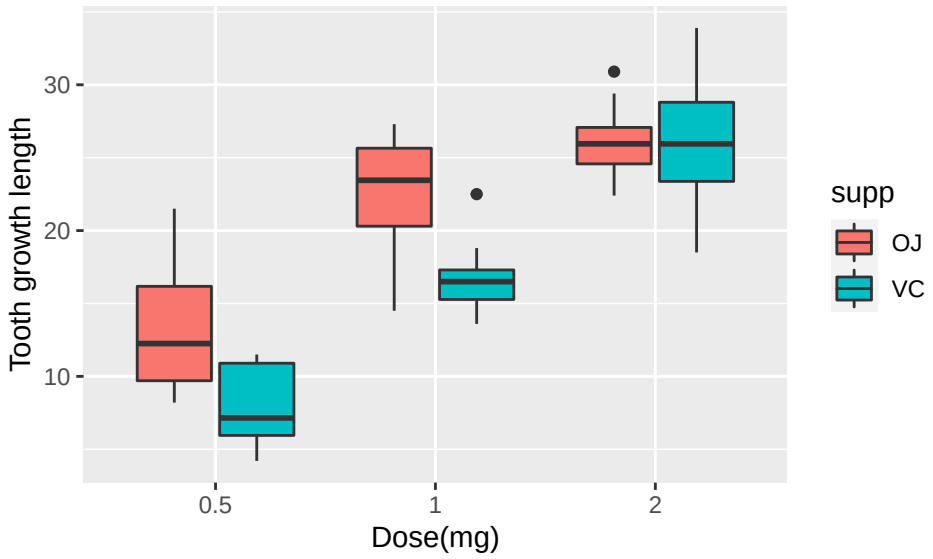
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
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
```

We can also make a box plot to see how the tooth growth length is related to supp and dose:

```
library(ggplot2)
ggplot(data = ToothGrowth, aes(x = as.factor(dose), y = len, fill = supp)) +
    geom_boxplot() +
    xlab("Dose(mg)") +
    ylab("Tooth growth length")
```



Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. ### Hypothesis 1: Two supplies have the same influence to the tooth growth over the whole dose range.

```
test_supp1 <- t.test(len ~ supp, data = ToothGrowth)
test_supp1$conf.int
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

```
test_supp1$p.value
```

```
## [1] 0.06063451
```

The confidence interval includes 0 and p value is greater than 0.05. Fail to reject hypo 1. Can't tell the difference.

Hypothesis 2:

Two supplies have the same influence to the tooth growth with 0.5 mg dosage.

```
test_supp2 <- t.test(len ~ supp, data = subset(ToothGrowth, dose == 0.5))
test_supp2$conf.int
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```
test_supp2$p.value
```

```
## [1] 0.006358607
```

The confidence interval is greater than 0 and p value is smaller than 0.05. Reject hypo 2. OJ performs better than VC with 0.5 mg dosage.

Hypothesis 3:

Two supplies have the same influence to the tooth growth with 1.0 mg dosage.

```
test_supp3 <- t.test(len ~ supp, data = subset(ToothGrowth, dose == 1.0))
test_supp3$conf.int
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```
test_supp3$p.value
```

```
## [1] 0.001038376
```

The confidence interval is greater than 0 and p value is smaller than 0.05. Reject hypo 3. OJ performs better than VC with 1.0 mg dosage.

Hypothesis 4:

Two supplies have the same influence to the tooth growth with 2.0 mg dosage.

```
test_supp4 <- t.test(len ~ supp, data = subset(ToothGrowth, dose == 2.0))
test_supp4$conf.int
```

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

```
test_supp4$p.value
```

```
## [1] 0.9638516
```

The confidence interval includes 0 and p value is greater than 0.05. Fail to reject hypo 4. Can't tell the difference.

Conculsion

OJ performs better than VC for dosages 0.5 & 1.0 mg. OJ and VC gives the same amount of tooth growth for dose amount 2.0 mg.