Statistical Inference Course Project - Part 2

JLei

5/1/2020

Part 2 - Basic Inferential Data Analysis

This part allows one to analyze the ToothGrowth data in the R dataset package.

1. Load the ToothGrowth data and perform some basic exploratory data analysis.

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

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

```
head(ToothGrowth, 3)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
```

```
str(ToothGrowth)
```

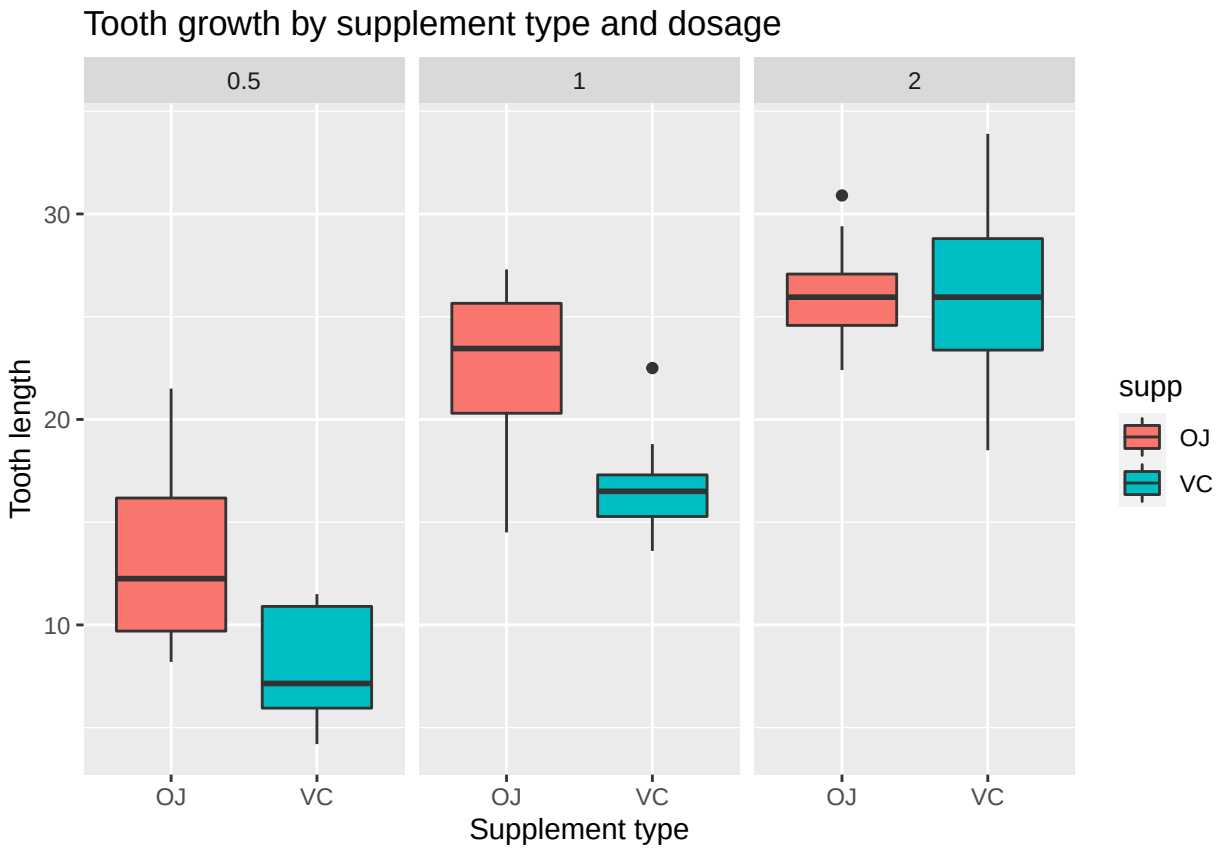
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

2. Provide a basic summary of the data

```
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
```

```
library(ggplot2)
ggplot(ToothGrowth, aes(supp, len, fill=supp)) +
        facet_grid(.~dose) +
        geom_boxplot() +
        labs(title="Tooth growth by supplement type and dosage",
             x="Supplement type",
             y="Tooth length")
```



There is a positive effect of dosage on tooth growth, such that tooth length increases as dosage increases. The positive effect of dosage appears to be linear with the supplement "VC", while the positive effect seems to plataeu with "OJ" at 2mg dosage. Generally "OJ" appears to induce higher tooth growth than "VC", except at the dosage of 2mg.

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

- **Hypothesis to compare OJ and VC**
  Let the null hypothesis be that there is no difference in tooth growth between the use of the supplement of OJ and VC.
  $\mathbf{len_{OJ} = len_{VC}}$

Let the alternative hypothesis be that there is more growth in tooth with the use of the supplement of OJ than VC.
$\mathbf{len_{OJ} > len_{VC}}$

Now let's subset the data by supplement.

```
OJ <- ToothGrowth$len[ToothGrowth$supp=="OJ"]
VC <- ToothGrowth$len[ToothGrowth$supp=="VC"]
```

Then perform a one-sided t-test to test our hypothesis.

```
t.test(OJ, VC, alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ and VC
## 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 p-value of 0.03032 can be interpreted as there is about 3% of chance for an extreme value in the difference of the mean in tooth growth. Since it is lower than alpha=0.5, then we reject the null hypothesis. In other words, we can conclude that the supplement OJ is likely to induce greater tooth growth than VC.

- **Hypothesis to compare dosage** Let the null hypothesis be that there is no difference in tooth growth between dosage.
  $\mathbf{len_{0.5}} = \mathbf{len_1}$ and $\mathbf{len_1} = \mathbf{len_2}$

Let the alternative hypothesis be that there is more growth in tooth with higher dosage than lower dosage.
$\mathbf{len_{0.5}} < \mathbf{len_1}$ and $\mathbf{len_1} < \mathbf{len_2}$

First, let's subset the data by dosage.

```
halfdose <- ToothGrowth$len[ToothGrowth$dose==0.5]
onedose <- ToothGrowth$len[ToothGrowth$dose==1]
twodose <- ToothGrowth$len[ToothGrowth$dose==2]
```

Next perform a one-sided t-test to compare tooth growth by half dosage and one dosage.

```
t.test(halfdose, onedose, alternative = "less", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

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

We can conclude that the dosage of 1mg is likely to stimulate greater tooth growth than 0.5mg since the p-value of 6.342e-08 is very much less than alpha=0.5 (thus we rejected the null hypothesis of $\mathbf{len_{0.5}} = \mathbf{len_1}$).

Then perform a one-sided t-test to compare tooth growth by one dosage and two dosages.

```
t.test(twodose, onedose, alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  twodose and onedose
## 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
```

Similarly, we can conclude that the dosage of 2mg is likely to stimulate greater tooth growth than 1mg since the p-value of 9.532e-06 is very much less than alpha=0.5 (thus we rejected the null hypothesis of $\mathbf{len_1 = len_2}$).

Although t-test was not performed to compare 2mg vs 0.5mg, the above two tests suggests that we would very likely to reject $\mathbf{len_{0.5} = len_2}$.

Finally, we can draw a conclusion that the tooth growth increases as the dosage increases.

- **Hypothesis to compare OJ and VC at dosage 2mg** Let the null hypothesis be that there is no difference in tooth growth between the use of the supplement of OJ and VC at dosage of 2mg.
  $\mathbf{len_{OJ2} = len_{VC2}}$

Let the alternative hypothesis be that there is more growth in tooth with the use of the supplement of OJ than VC at dosage of 2mg.
$\mathbf{len_{OJ2} > len_{VC2}}$

Now let's subset the data by supplement at 2mg dosage.

```
OJ2 <- ToothGrowth$len[ToothGrowth$supp=="OJ" & ToothGrowth$dose==2]
VC2 <- ToothGrowth$len[ToothGrowth$supp=="VC" & ToothGrowth$dose==2]
```

Then perform a one-sided t-test to test our hypothesis.

```
t.test(OJ2, VC2, alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ2 and VC2
## t = -0.046136, df = 14.04, p-value = 0.5181
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -3.1335     Inf
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

Contrary to the t-test comparing OJ and VC, the t-test comparing OJ and VC at the dosage of 2mg yields a p-value of 0.5181, which is much greater than alpha=0.05, thus we fail to reject the null hypothesis. In other words, there is insufficient evidence to show any difference in tooth growth using the supplement OJ and VC at the dosage of 2mg.

4. State the conclusions and the assumptions needed for the conclusions.

The conclusions are embeded in the result interpretations for #3.

The assumptions are as follows:

- The variabels are I.I.D (independent, identically distributed)
- Unequal variance of tooth growth with different supplements and dosages
- A normal distribution for the tooth growth