Basic inferential data_analysis

Zhibin Guan

Overview

This is the second portion of the statistical interference course project (coursera).

The basic information of ToothGrowth data:

1. `len`: Tooth length in millimeters (numeric variable);
2. `supp`: Supplement type (factor variable with levels ascorbic acid (VC) and orange juice (OJ);
3. `dose`: Dose in miligrams (numeric variable) – 0.5, 1, or 2 mg/day.

Demands

1. Load the ToothGrowth data and perform some basic exploratory data analysis;
2. Provide a basic summary of the data;
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

Answers

Pre-processing for this second portion of the project

Load the needed libraries:

```
library(datasets)
library(dplyr)
library(stringr)
library(broom)
library(ggplot2)
```

Answer of demand 1

Recall demand 1: load ToothGrowth data and perform some basic EDA.

```
str(dat)
```

```
## '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 ...
```
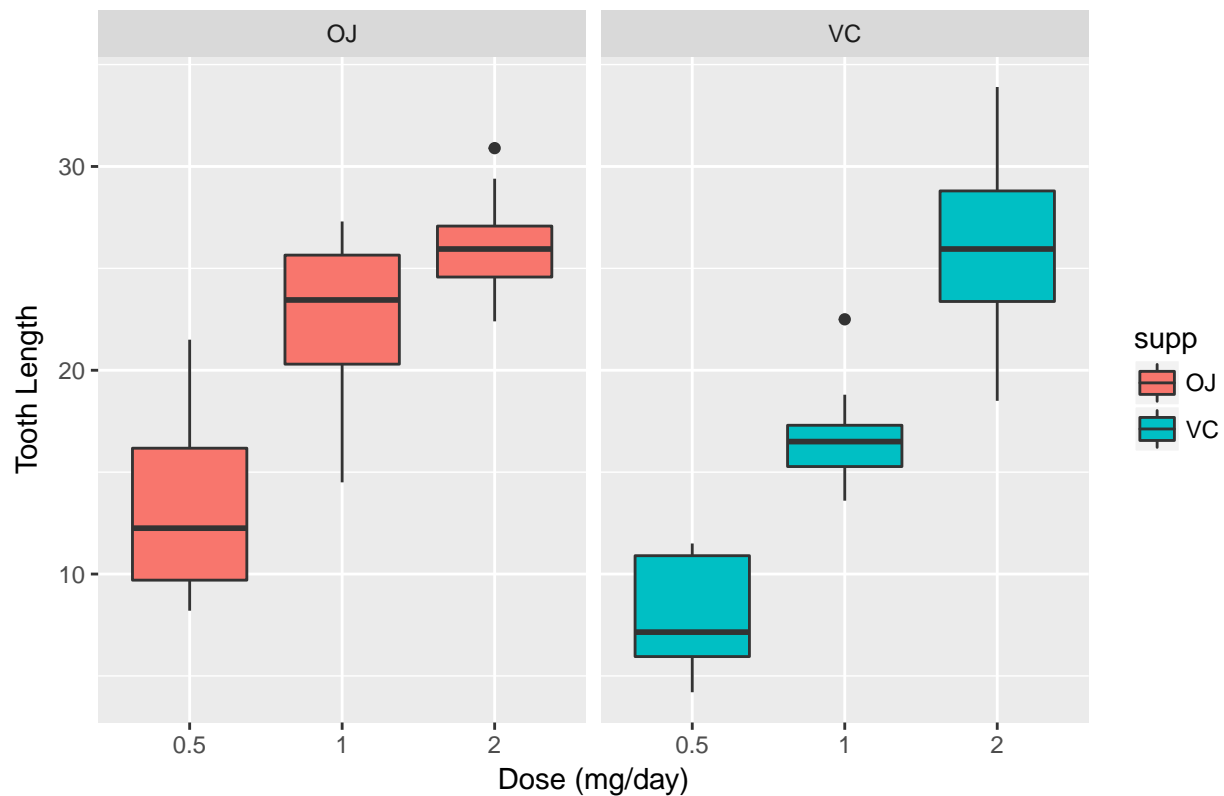
```
summary(dat)
```

```
##       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
```

Plot the ToothGrowth data according level supp – OJ and VC:

Tooth length by dose level (group VC and group OJ)

Answer of demand 2 – Basic summary of the ToothGrowth data.

The box plots show that the tooth length increases as dose increases in both group (OJ and VC). Differently, OJ is more effective than VC for tooth growth when dose equal 0.5 mg/day and 1 mg/day. And the effective of OJ and VC for tooth growth is equal when dose is 2 mg/day.

Answer of demand 3 – Hypothesis tests

Recall demand 3: use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

Since the number of samples of each group (OJ and VC) is less than 30, the standard error estimate is not accurate. Therefore, the T distribution is a well option.

Hypothese #1: OJ and AC generate the same effect for tooth growth accordind to the ToothGrouth data.

The confidence interval of hypothese 1:

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

The p-value of hypothese 1:

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

Conclusion of test_1: we can observe that the confidence intervals is (-0.1710156, 7.571016) which contains 0, and the p-value is about 0.06 which is greater than the alpha 0.05. Hence, we fail to reject the null hypothesis 1.

Hypothese #2: OJ and AC generate the same effect for tooth growth when doss is 0.5 mg/day.

The confidence interval of hypothese 2:

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

The p-value of hypothese 2:

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

Conclusion of test_2: the confidence intervals does not contains 0, and the p-values is about 0.006 which is less than alpha 0.05. Hence, we can reject the null hypothesis 2.

Hypothese #3: OJ and AC generate the same effect for tooth growth when doss is 1 mg/day.

The confidence interval of hypothese 3:

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

The p-value of hypothese 3:

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

Conclusion of test_3: the confidence intervals does not contains 0, and the p-values is about 0.001 which is less than alpha 0.05. Hence, we can reject the null hypothesis 3.

Hypothese #4: OJ and AC generate the same effect for tooth growth when doss is 2 mg/day.

The confidence interval of hypothese 4:

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

The p-value of hypothese 4:

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

Conclusion of test_4: the confidence intervals contains 0, and the p-values is about 0.964 which is greate than alpha 0.05. Hence, we fail reject the null hypothesis 4. What's more, the p-value is too big, we can accept the null hypothesis 4.

Answer of demand 4:

Recall demand 4: state your conclusions and the assumptions needed for your conclusions

Conclusions and assumptions OJ and AC generate the same effect for tooth growth when dose is 2 mg/day. Different, OJ generate more effect than AC when dose is 0.5 and 1 mg/day. Therefore,

Assumptions: * The ToothGrowth data is approximately normal distribution * When the number of dose is greater than 2 mg/day, the OJ and AC generate similar effect for tooth growth.