Tooth Growth

hoaihuongbk

June 26, 2019

Overview

We will use ToothGrowth in dataset package for exploring data analyses, use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

Preparation 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 ...
```

Basic summary

```
summary(ToothGrowth)
```

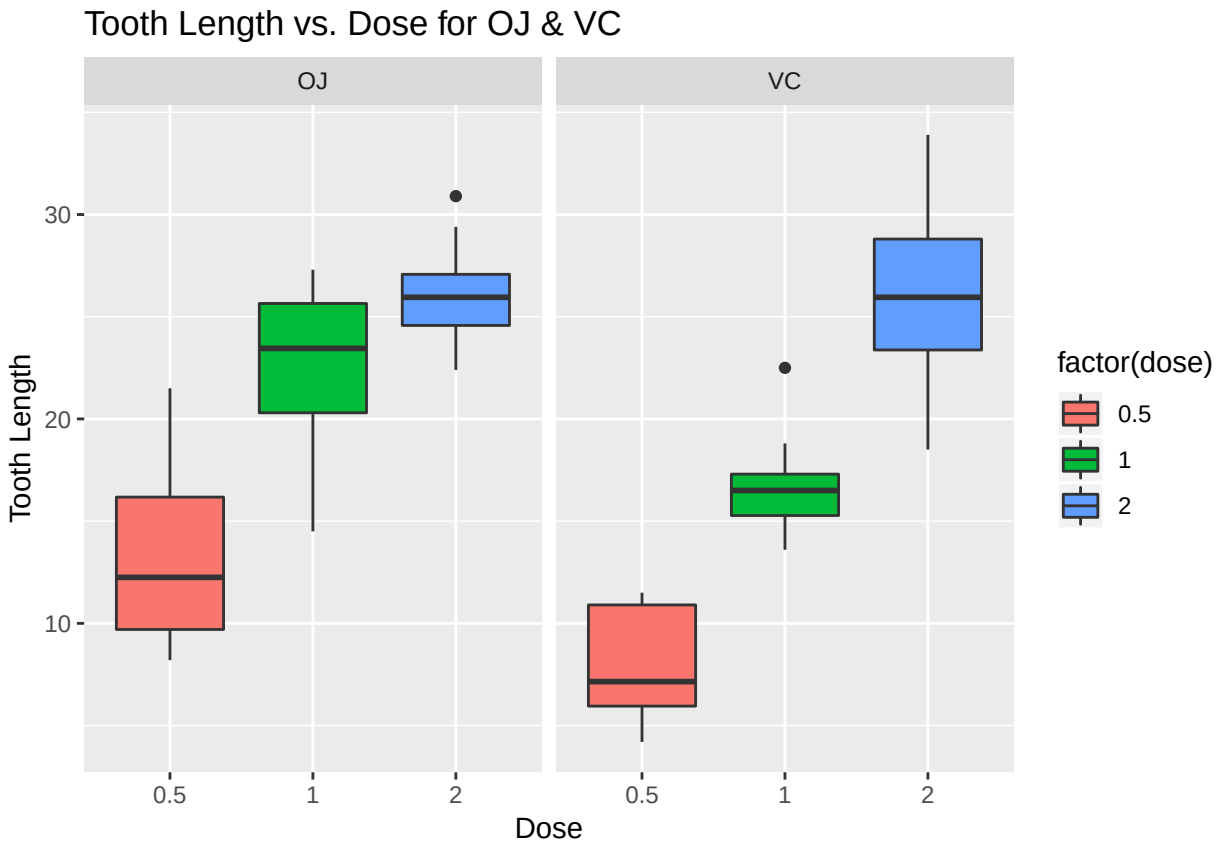
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
table(ToothGrowth$supp,ToothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

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

```
library(ggplot2)
g <- ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=factor(dose))) +
        geom_boxplot() +
        facet_grid(.~supp) +
        labs(title="Tooth Length vs. Dose for OJ & VC", x="Dose", y="Tooth Length")
g
```



The confidence intervals for the difference in means provide a range of likely values for $(\mu_1 - \mu_2)$. It is important to note that all values in the confidence interval are equally likely estimates of the true value of $(\mu_1 - \mu_2)$. If there is no difference between the population means, then the difference will be zero $(\mu_1 - \mu_2 = 0)$. Zero is the null value of the parameter (in this case the difference in means). If a 95% confidence interval includes the null value, then there is no statistically meaningful or statistically significant difference between the groups. If the confidence interval does not include the null value, then we conclude that there is a statistically significant difference between the groups.

Dose 0.5

```
library(dplyr)
OJ_0.5 <- ToothGrowth %>% filter(supp == "OJ" & dose == 0.5)
VC_0.5 <- ToothGrowth %>% filter(supp == "VC" & dose == 0.5)
t.test(OJ_0.5$len, VC_0.5$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ_0.5$len and VC_0.5$len
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

95% confidence that limits from 1.719057 to 8.780943 does not contain zero, this confidence interval suggests that two population means are not equal.

Dose 1

```
OJ_1 <- ToothGrowth %>% filter(supp == "OJ" & dose == 1)
VC_1 <- ToothGrowth %>% filter(supp == "VC" & dose == 1)
t.test(OJ_1$len, VC_1$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ_1$len and VC_1$len
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

95% confident that the limits from 2.802148 to 9.057852 does not contain zero, this confidence interval also suggests that two population means are not equal.

Dose 2

```
OJ_2 <- ToothGrowth %>% filter(supp == "OJ" & dose == 2)
VC_2 <- ToothGrowth %>% filter(supp == "VC" & dose == 2)
t.test(OJ_2$len, VC_2$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ_2$len and VC_2$len
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

95% confident that the limits from -3.79807 to 3.63807 contain zero, this confidence interval suggests that it is very possible two population means are equal.

State your conclusions and the assumptions needed for your conclusions

There are 95% confidence that dose 0.5 and dose 1 of OJ result in longer tooth length than dose 0.5 and dose 1 of VC. With highest dose of 2, there is no statistically significant difference between the effects of OJ and VC.

Appendix

Source code: https://github.com/hoaihuongbk/Statistical_Inference