ToothGrowth Data Analysis

Dongyuan Wu

2019-05-09

Overview

This project analyzes the ToothGrowth data in the R datasets package. The following things will be done:

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

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.

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

```
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 ...
```

We can find that ToothGrowth data has 60 observations and 3 variables. The tooth length (len) is a numeric variable, the supplement type (supp) is a factor variable, and the dose in milligrams per day (dose) is a numeric variable.

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
```

The mean and median of the tooth lengths is 18.8133 and 19.25. There are both 30 pigs receive VC or OJ, respectively.

1

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

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

As we can see, the dose in milligrams per day only has 3 values, so we consider it as a factor variable.

Then draw a boxplot to investigate the tooth lengths.

```
library(ggplot2)

ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=factor(dose))) +
    geom_boxplot() + facet_grid(. ~ supp) +
    labs(x="Dose in Milligrams Per Day", y="Tooth Length",
        title="Effect of Dose and Supplement Type on Tooth Growth") +
    guides(fill=guide_legend(title="Dose"))
```
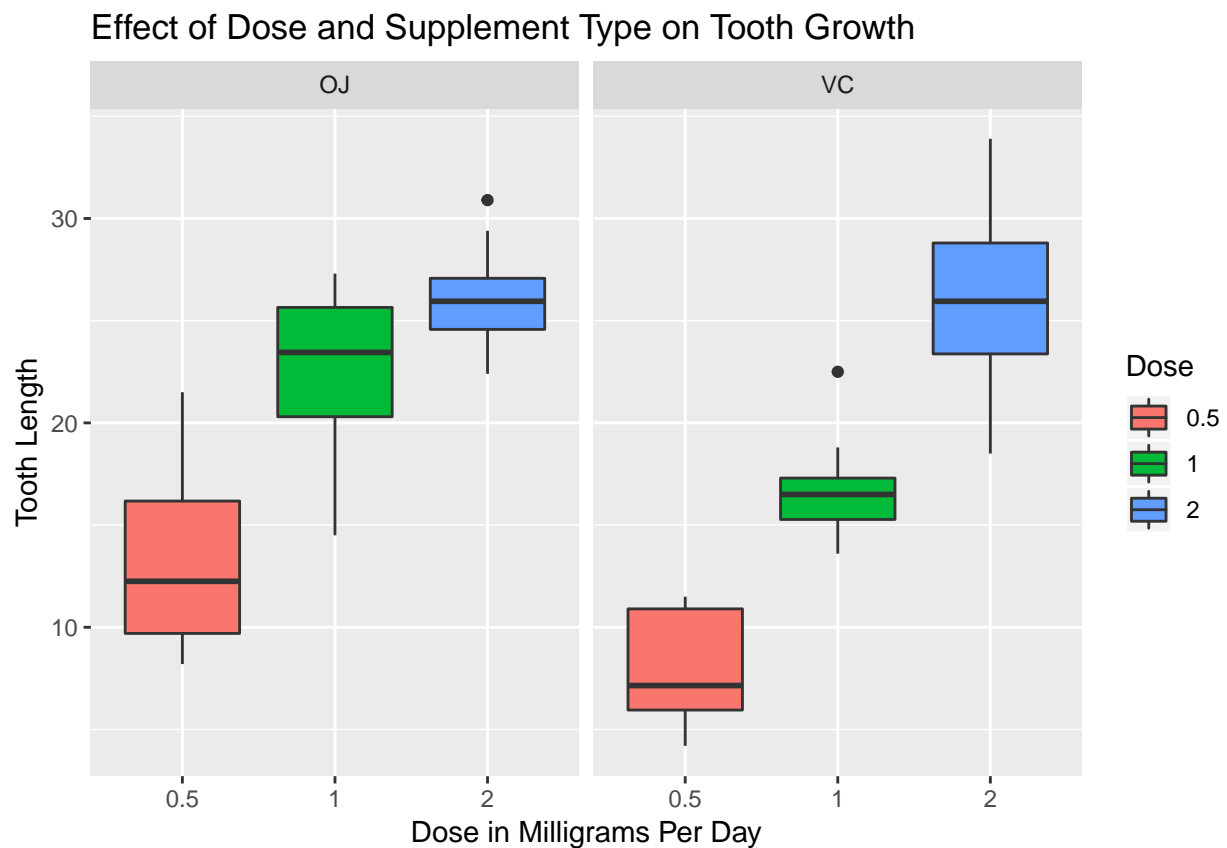

We can find tooth length grows up as the dose increases. When dose is less than 2 milligrams/day, OJ will let tooth length longer than VC.

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

3.1 Comparation between tooth growth by different supplement types

```
##                        pvalue   CI.lower CI.upper
## OJ vs. VC at 0.5 dose 0.006358607  1.7190573 8.780943
## OJ vs. VC at 1 dose   0.001038376  2.8021482 9.057852
## OJ vs. VC at 2 dose   0.963851589 -3.7980705 3.638070
## OJ vs. VC             0.060634508 -0.1710156 7.571016
```

We can find if we just compare tooth growth by different supplement types without considering dose, OJ and VC seem to be the same (pvalue is larger than 0.05 and the confidence interval contains zero). However, the tooth growth by supplement types are different if we just consider 0.5 dose or 1 dose (pvalue is smaller than 0.05 and the confidence interval does not contain zero).

3.2 Comparation between tooth growth by different doses

```
##                         pvalue   CI.lower    CI.upper
## dose .5 vs. 1 at OJ 8.784919e-05 -13.415634  -5.5243656
## dose .5 vs. 1 at VC 6.811018e-07 -11.265712  -6.3142880
## dose .5 vs. 1       1.268301e-07 -11.983781  -6.2762187
## dose 1 vs. 2 at OJ  3.919514e-02  -6.531443  -0.1885575
## dose 1 vs. 2 at VC  9.155603e-05 -13.054267  -5.6857333
## dose 1 vs. 2        1.906430e-05  -8.996481  -3.7335195
## dose .5 vs. 2 at OJ 1.323784e-06 -16.335241  -9.3247594
## dose .5 vs. 2 at VC 4.681577e-08 -21.901512 -14.4184880
## dose .5 vs. 2       4.397525e-14 -18.156167 -12.8338335
```

We can find there are significant differences among doses, whichever supplement type.

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

4.1 Assumptions

1. Variances of tooth growth are different (var.equal=FALSE).

2. Observations are independent.

4.2 Conclusions

1. For lower doses (0.5 and 1 mg), OJ will let tooth length be longer than VC.

2. For high doses (2 mg), tooth growth is the same for different supplement types.

3. Higher doses will provide more tooth growth, whichever supplement type.

Appendix

1. Code for comparation between tooth growth by different supplement types

```r
t.dose.05 <- t.test(len ~ supp, data = ToothGrowth[ToothGrowth$dose == 0.5, ])
t.dose.1 <- t.test(len ~ supp, data = ToothGrowth[ToothGrowth$dose == 1, ])
t.dose.2 <- t.test(len ~ supp, data = ToothGrowth[ToothGrowth$dose == 2, ])
t.dose <- t.test(len ~ supp, data = ToothGrowth)

ttest1 <- data.frame(pvalue=c(t.dose.05$p.value, t.dose.1$p.value,
                             t.dose.2$p.value, t.dose$p.value),
                    CI.lower=c(t.dose.05$conf.int[1],t.dose.1$conf.int[1],
                             t.dose.2$conf.int[1], t.dose$conf.int[1]),
                    CI.upper=c(t.dose.05$conf.int[2],t.dose.1$conf.int[2],
                             t.dose.2$conf.int[2], t.dose$conf.int[2]))
rownames(ttest1) <- c("OJ vs. VC at 0.5 dose", "OJ vs. VC at 1 dose",
                     "OJ vs. VC at 2 dose", "OJ vs. VC")
ttest1
```

```
##                          pvalue    CI.lower CI.upper
## OJ vs. VC at 0.5 dose 0.006358607  1.7190573 8.780943
## OJ vs. VC at 1 dose   0.001038376  2.8021482 9.057852
## OJ vs. VC at 2 dose   0.963851589 -3.7980705 3.638070
## OJ vs. VC             0.060634508 -0.1710156 7.571016
```

2. Code for comparation between tooth growth by different doses

```r
# compare dose 0.5 and dose 1
t.d1.OJ <- t.test(len ~ factor(dose),
                data = ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose != 2, ])
t.d1.VC <- t.test(len ~ factor(dose),
                data = ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose != 2, ])
t.d1 <- t.test(len ~ factor(dose),
              data = ToothGrowth[ToothGrowth$dose != 2, ])

# compare dose 2 and dose 1
t.d2.OJ <- t.test(len ~ factor(dose),
                data = ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose != 0.5, ])
t.d2.VC <- t.test(len ~ factor(dose),
                data = ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose != 0.5, ])
t.d2 <- t.test(len ~ factor(dose),
              data = ToothGrowth[ToothGrowth$dose != 0.5, ])

# compare dose 0.5 and dose 2
t.d3.OJ <- t.test(len ~ factor(dose),
                data = ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose != 1, ])
t.d3.VC <- t.test(len ~ factor(dose),
                data = ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose != 1, ])
t.d3 <- t.test(len ~ factor(dose),
              data = ToothGrowth[ToothGrowth$dose != 1, ])
```

```
ttest2 <- data.frame(pvalue=c(t.d1.OJ$p.value, t.d1.VC$p.value, t.d1$p.value,
                              t.d2.OJ$p.value, t.d2.VC$p.value, t.d2$p.value,
                              t.d3.OJ$p.value, t.d3.VC$p.value, t.d3$p.value),
                    CI.lower=c(t.d1.OJ$conf.int[1], t.d1.VC$conf.int[1], t.d1$conf.int[1],
                               t.d2.OJ$conf.int[1], t.d2.VC$conf.int[1], t.d2$conf.int[1],
                               t.d3.OJ$conf.int[1], t.d3.VC$conf.int[1], t.d3$conf.int[1]),
                    CI.upper=c(t.d1.OJ$conf.int[2], t.d1.VC$conf.int[2], t.d1$conf.int[2],
                               t.d2.OJ$conf.int[2], t.d2.VC$conf.int[2], t.d2$conf.int[2],
                               t.d3.OJ$conf.int[2], t.d3.VC$conf.int[2], t.d3$conf.int[2]))
rownames(ttest2) <- c("dose .5 vs. 1 at OJ", "dose .5 vs. 1 at VC", "dose .5 vs. 1",
                      "dose 1 vs. 2 at OJ", "dose 1 vs. 2 at VC", "dose 1 vs. 2",
                      "dose .5 vs. 2 at OJ", "dose .5 vs. 2 at VC", "dose .5 vs. 2")
ttest2
```

```
##                          pvalue    CI.lower     CI.upper
## dose .5 vs. 1 at OJ 8.784919e-05 -13.415634   -5.5243656
## dose .5 vs. 1 at VC 6.811018e-07 -11.265712   -6.3142880
## dose .5 vs. 1       1.268301e-07 -11.983781   -6.2762187
## dose 1 vs. 2 at OJ  3.919514e-02  -6.531443   -0.1885575
## dose 1 vs. 2 at VC  9.155603e-05 -13.054267   -5.6857333
## dose 1 vs. 2        1.906430e-05  -8.996481   -3.7335195
## dose .5 vs. 2 at OJ 1.323784e-06 -16.335241   -9.3247594
## dose .5 vs. 2 at VC 4.681577e-08 -21.901512  -14.4184880
## dose .5 vs. 2       4.397525e-14 -18.156167  -12.8338335
```