Tooth Growth Data Inference

Baoshi Sun

Overview

In this report, we are going to perform some exploratory analyses on the ToothGrowth data in the R datasets package and try to reach conclusions based on the results of the confidence intervals and hypothesis test.

Exploratory Analysis

```
## '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 ...
```

```
##
## OJ VC
## 30 30
```

```
##
## 0.5   1    2
##  20  20  20
```

We have 60 samples of 3 variables in the dataset, which can be splitted into 2 groups evenly on variable supp (VC and OJ) with 3 amount of doses (0.5, 1, 2) in each group.

Tooth Growth under different dose of supp
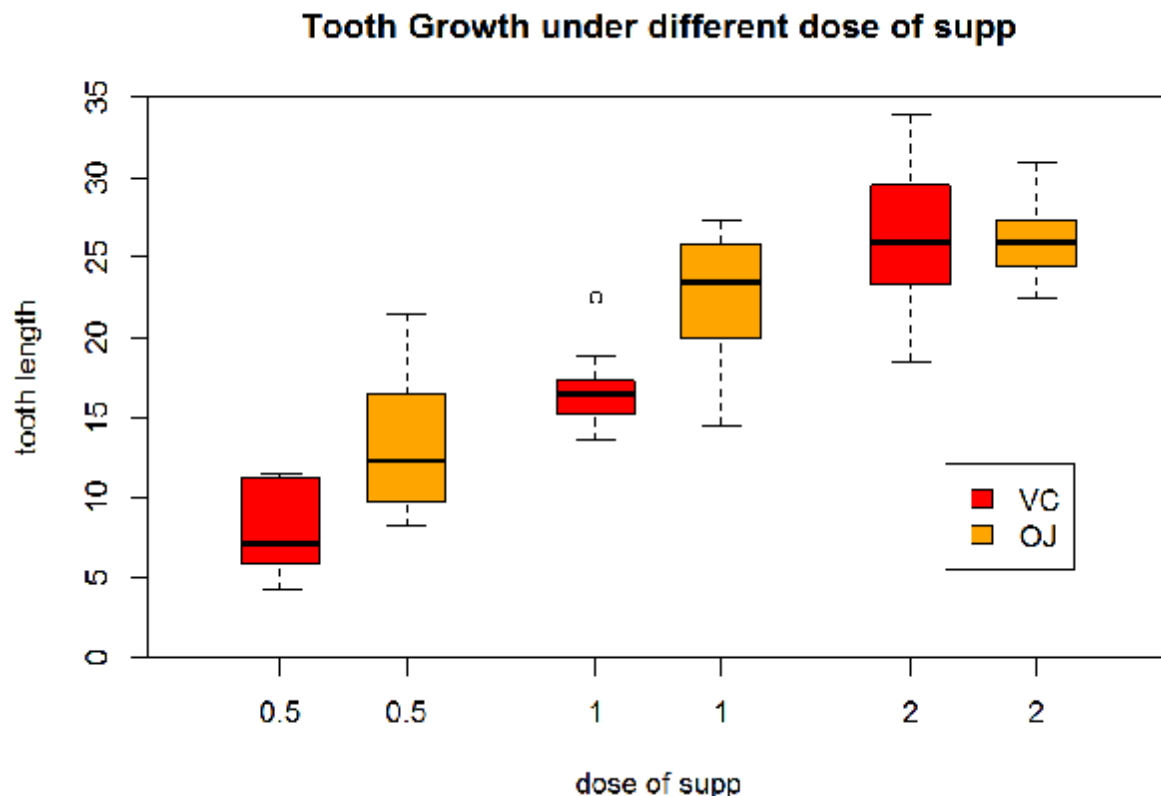
Intuitionally, OJ may lead to more tooth growth than VC given the same dose. And greater dose seems to have greater effect.

Confidence Intervals on variable supp

```
##   supp count  mean   sd
## 1   VC    30 16.96 1.51
## 2   OJ    30 20.66 1.21
```

To verify the idea of OJ having better outcome over VC, we divide the data into two small-sample groups on variable supp. Because the groups are independent, we are using indenpendent group t confidence intervals with equal variance.

```
## [1] -0.1670064  7.5670064
## attr(,"conf.level")
## [1] 0.95
```

```
## [1] 0.4708204 6.9291796
## attr(,"conf.level")
## [1] 0.9
```

Let us construct both 95% and 90% T intervals for supp comparing OJ to VC (substracting in order of OJ - VC). As can be seen, under in confidence level 95% the lower end of the T interval is below zero, whereas the interval is entirely above zero if we set the confidence level to 90%.

Hypothesis Testing on variable dose

```
##    dose count  mean   sd
## 1  0.5     20 10.61 1.01
## 2  1.0     20 19.73 0.99
## 3  2.0     20 26.10 0.84
```

Hypothesis Testing 1: dose 1 on does 0.5

- $H_0 : \mu = 10.61$
- $H_a : \mu > 10.61$
  When tested under 95% confidence level, we got p-value 6.331484810^{-8}, which is less than $\alpha = 0.05$. Therefore, we can reject the null hypotheis.

Hypothesis Testing 1: dose 2 on does 1

- $H_0 : \mu = 19.73$
- $H_a : \mu > 19.73$
  When tested under 95% confidence level, we got p-value 9.054142710^{-6}, which is less than $\alpha = 0.05$ too. Thus we reject the null hypotheis again.

Conclusions

From above analysis, concusions can be made under the following assumptions:
* Samples in ToothGrowth dataset are randomly drawn from the same population, which means the 30 samples are independent and identically distributed (iid)
* Although with limited amount of samples in each subset, we assume the data are normally distributed

- Conclusion 1: With 90% confidence level, the OJ has better effect than VC on the growth of tooth.
- Conclusion 2: With 95% confidence level, dose 2 has greater effect than dose 1. The same result is observed for dose 1 on dose 0.5.

Appendix: source code for the report

```
data(ToothGrowth)
str(ToothGrowth)
table(ToothGrowth$supp); table(ToothGrowth$dose)

#with(ToothGrowth, {
#    boxplot(len ~ dose, boxwex = 0.25, at = 1:3 - 0.2,
#        subset = supp == "VC", col = "red",
#        main = "Tooth Growth under different dose of supp",
#        xlab = "dose of supp", ylab = "tooth length",
#        xlim = c(0.5, 3.5), ylim = c(0, 35), yaxs = "i")
#    boxplot(len ~ dose, add = TRUE, boxwex = 0.25, at = 1:3 + 0.2,
#        subset = supp == "OJ", col = "orange")
#    legend(2.91, 12.1, c("VC", "OJ"), fill = c("red", "orange"))
#})

g_vc <- ToothGrowth$len[ToothGrowth$supp == "VC"]
g_oj <- ToothGrowth$len[ToothGrowth$supp == "OJ"]
n_vc <- length(g_vc); n_oj <- length(g_oj)
m_vc <- round(mean(g_vc), 2); m_oj <- round(mean(g_oj), 2)
s_vc <- round(sd(g_vc) / sqrt(n_vc), 2); s_oj <- round(sd(g_oj) / sqrt(n_oj)
, 2)

sum_supp <- data.frame(rbind(c("VC", n_vc, m_vc, s_vc), c("OJ", n_oj, m_oj,
s_oj)))
names(sum_supp) <- c("supp", "count", "mean", "sd")
#sum_supp

t.test(g_oj, g_vc, paired=FALSE, var.equal=TRUE)$conf
t.test(g_oj, g_vc, paired=FALSE, var.equal=TRUE, conf.level = 0.9)$conf

g_dose1 <- ToothGrowth$len[ToothGrowth$dose == 0.5]
g_dose2 <- ToothGrowth$len[ToothGrowth$dose == 1]
g_dose3 <- ToothGrowth$len[ToothGrowth$dose == 2]
n_dose1 <- length(g_dose1); n_dose2 <- length(g_dose2); n_dose3 <- length(g_
dose3)
m_dose1 <- round(mean(g_dose1), 2)
m_dose2 <- round(mean(g_dose2), 2)
m_dose3 <- round(mean(g_dose3), 2)
s_dose1 <- round(sd(g_dose1) / sqrt(n_dose1), 2)
s_dose2 <- round(sd(g_dose2) / sqrt(n_dose2), 2)
s_dose3 <- round(sd(g_dose3) / sqrt(n_dose3), 2)

sum_dose <- data.frame(rbind(c(0.5, n_dose1, m_dose1, s_dose1), c(1, n_dose2
, m_dose2, s_dose2), c(2, n_dose3, m_dose3, s_dose3)))
names(sum_dose) <- c("dose", "count", "mean", "sd")
#sum_dose

ht_d1 <- t.test(g_dose2, g_dose1, paired=FALSE, var.equal=TRUE, alternative=
"greater")
al_d1 <- ht_d1$alternative; pv_d1 <- ht_d1$p.value
ht_d2 <- t.test(g_dose3, g_dose2, paired=FALSE, var.equal=TRUE, alternative=
```

```
"greater")
al_d2 <- ht_d2$alternative; pv_d2 <- ht_d2$p.value
#ht_d1; ht_d2
```