Statistical Inference: Part 1

Paul Degtyariov

02 05 2020

Overview

This is a part 2 for project in Statistical Inference cource by JHU. It's dedicated to examining the tooth grow dataset.

Tasks

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.

Loading data and packages

```
library(dplyr)
library(ggplot2)
library(knitr)

data(ToothGrowth)
```

Tasks 1 and 2

```
#Looking at data
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 ...
```

```
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
```

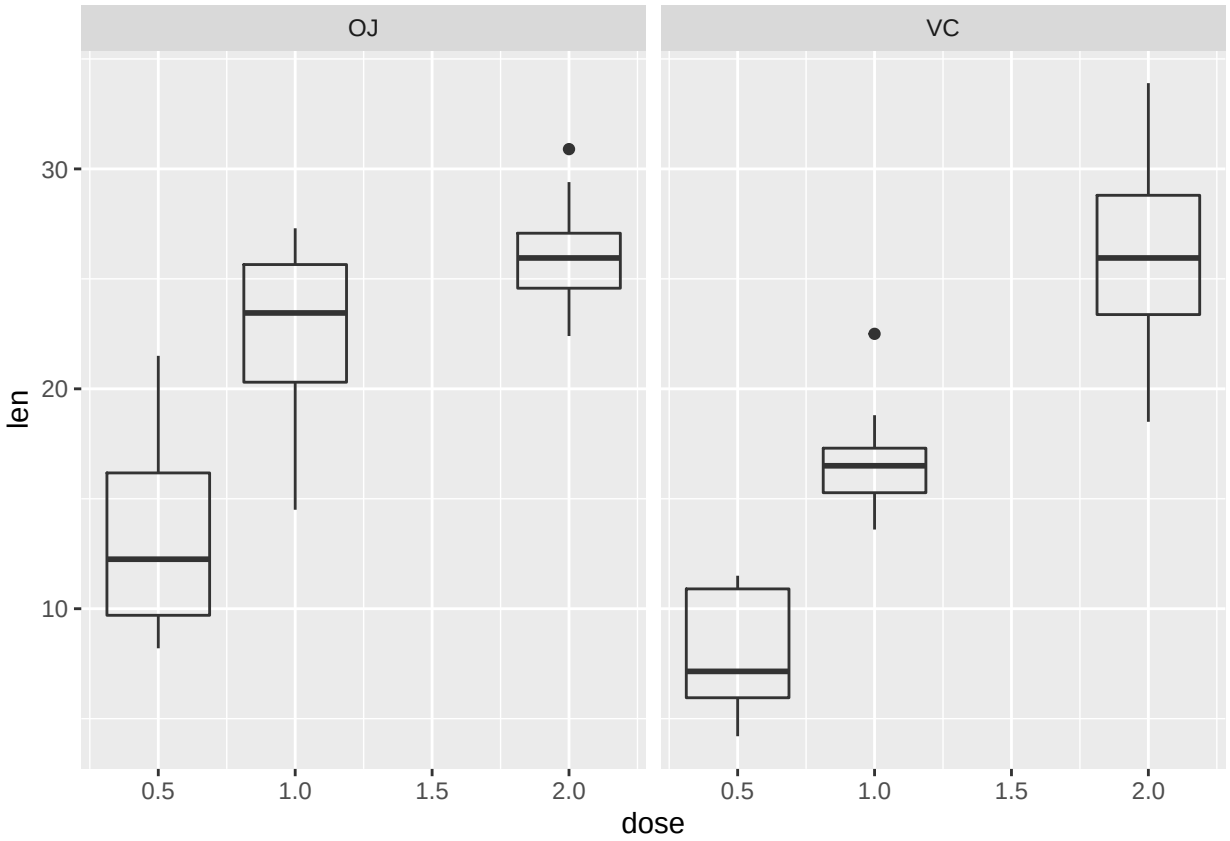
```
#As there is a factor varible that describes the type of summplemet, let's look on some summaries stati

summ <- ToothGrowth %>%
    group_by(supp,dose) %>%
    summarise(count=n(),Mean=mean(len),SD=sd(len),Min=min(len),Med=median(len),Max=max(len)) %>%
    ungroup()

kable(summ)
```

supp	dose	count	Mean	SD	Min	Med	Max
OJ	0.5	10	13.23	4.459708	8.2	12.25	21.5
OJ	1.0	10	22.70	3.910953	14.5	23.45	27.3
OJ	2.0	10	26.06	2.655058	22.4	25.95	30.9
VC	0.5	10	7.98	2.746634	4.2	7.15	11.5
VC	1.0	10	16.77	2.515309	13.6	16.50	22.5
VC	2.0	10	26.14	4.797731	18.5	25.95	33.9

```
#At this point it seems like there is a connection between the dosage and length for both supplemets. L
ggplot(ToothGrowth,aes(x=dose,y=len,group=dose))+
    geom_boxplot(fill=NA)+
    facet_wrap(~supp,nrow=1)
```



Task 3

First let's explore the effect of supplement on tooth length. We will use a two-sample t-test with a confidence level of 95% and unequal variances.

```
supp_test <- t.test(len ~ supp, data= ToothGrowth, var.equal = FALSE, paired=FALSE ,conf.level = .95)

supp_test
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

As it's shown in the summary, the null hypothesis is proven, meaning there is no significant difference in teeth length in relation to used supplement - the p-value is higher then 0.05

Now let's look at the effect of dosage using the same method

```r
#Doses 0.5 and 1
dose_test1 <- t.test(len ~ dose, data= ToothGrowth %>% filter(dose==0.5|dose==1), var.equal = FALSE, pa

dose_test1
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```r
#Doses 1 and 2
dose_test2 <- t.test(len ~ dose, data= ToothGrowth %>% filter(dose==1|dose==2), var.equal = FALSE, pair

dose_test2
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

Finnally, let's explore the effect of dosage of each supplement on teeth length using the same method

```r
#0.5 dose
supp_05_test <-  t.test(len ~ supp, data= ToothGrowth %>% filter(dose==0.5), var.equal = FALSE, paired=

supp_05_test
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##            13.23             7.98
```

```r
#1 dose
supp_1_test <-  t.test(len ~ supp, data= ToothGrowth %>% filter(dose==1), var.equal = FALSE, paired=FALS

supp_1_test
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##            22.70            16.77
```

```r
#2 dose
supp_2_test <-  t.test(len ~ supp, data= ToothGrowth %>% filter(dose==2), var.equal = FALSE, paired=FALS

supp_2_test
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##            26.06            26.14
```

Task 4 - Conclusions

Conluding from this analysis, we can state that increased doses of any supplement corresponds to higher teeth length as on booth tests p.value is close to zero. It's also to be noted that increase of a dose from 0.5 to 1 corresponds to bigger discrepency mean length then increse of a dose from 1 to 2.

However, there are two points of dosage - 0.5 and 1,- where OJ is shown to be corresponding to bigger values of teeth length then VC (as in both cases the p.value is smaller then 0.01. For the dosage of 2 there is no difference between supplements, and the p,value for the T test is 0.96