Title: “Statistical Inference Course Project Part 2”

author:“Sohair Zaki” **date:**“May 12, 2017”

Part 2: Basic Inferential Data Analysis Instructionsless Now in the second portion of the project, we’re going to analyze the ToothGrowth data in the R datasets package.

Load the ToothGrowth data and perform some basic exploratory data analyses Provide a basic summary of the data. 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) State your conclusions and the assumptions needed for your conclusions.

```
#Load the ToothGrowth data and perform some basic exploratory data analyses
# let load the library datasets

library(datasets)
# let us load the dataset toothgrowth
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 have 60 obss of 3 variables len--numeric, supp--factor, dose --numeric
# let us display the data frame head
head(ToothGrowth)
```

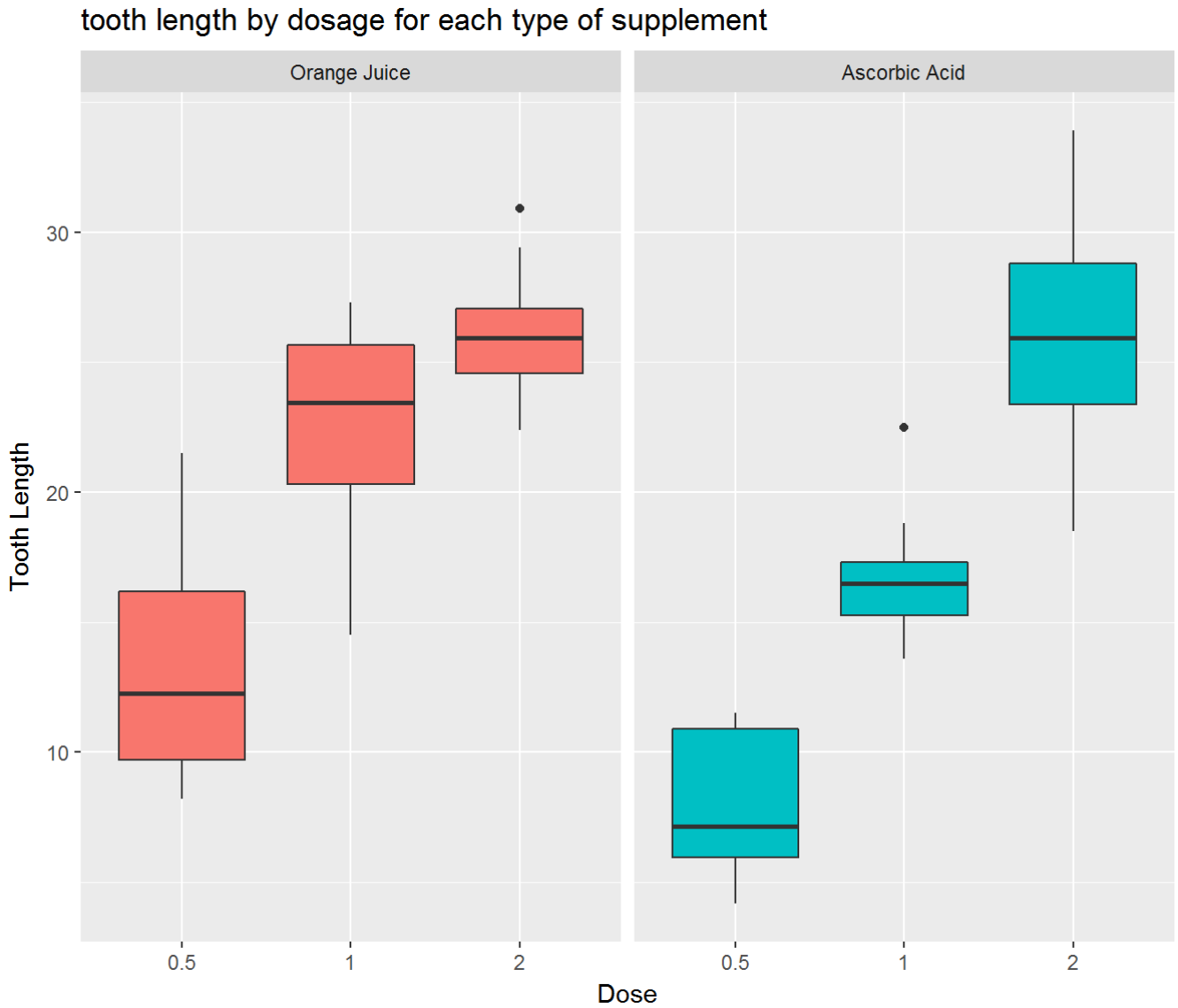
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
# let us summaries the data # mean, median, quartiles, max, min, etc
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
```

```
#load library ggplot2
library(ggplot2)
ourData = ToothGrowth
# specify levels
levels(ourData$supp) <- c("Orange Juice", "Ascorbic Acid")
# box plot specifying labels using labs for the title and x axis and y axis
ggplot(ourData, aes(x=factor(dose), y=len)) +
facet_grid(.~supp) +
geom_boxplot(aes(fill = supp), show_guide = FALSE) +
labs(title="tooth length by dosage for each type of supplement",
x="Dose",
y="Tooth Length")
```

```
## Warning: `show_guide` has been deprecated. Please use `show.legend`
## instead.
```



Conclusion: Basic summary of the data Looking at the we can conclude that as the dose increases the tooth growth increases. Also orange juice is more effective than ascorbic acid.increasing the dosage increases the tooth growth. Orange juice is more effective than ascorbic acid for tooth growth But both are equally effecive if the daily dosage is 2.0 milligrams.

Step 2: Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. Hypothesis #1 Orange juice & ascorbic acid deliver the same tooth growth across the data set.

```
library(broom)


#****
hypoth1 <- t.test(len ~ supp, data = ourData)

tidy(hypoth1)
```

```
##   estimate estimate1 estimate2 statistic    p.value parameter   conf.low
## 1      3.7  20.66333  16.96333  1.915268 0.06063451  55.30943 -0.1710156
##   conf.high                 method alternative
## 1  7.571016 Welch Two Sample t-test   two.sided
```

```
# note the p value is 0.06063451, confidence interval -0.1710156 and 7.571016 and confidence lev
el 0.95
# conclusion
# The confidence intervals includes 0 and the p-value is greater than the threshold of 0.05.
#The null hypothesis cannot be rejected.

#Hypothesis #2
#For the dosage of 0.5 mg/day, the two supplements deliver the same tooth growth.

hypoth2<-t.test(len ~ supp, data = subset(ourData, dose == 0.5))

tidy(hypoth2)
```

```
##   estimate estimate1 estimate2 statistic     p.value parameter conf.low
## 1     5.25     13.23      7.98  3.169733 0.006358607  14.96875 1.719057
##   conf.high                 method alternative
## 1  8.780943 Welch Two Sample t-test   two.sided
```

```
# note the confidence interval is  1.719057 8.780943 and p alue is 0.006358607

# conclusion The confidence interval does not include 0 and the p-value is below the 0.05 thresh
old.
# The null hypothesis can be rejected. The alternative hypothesis that is accepted.

#*******
#Hypothesis #3
#For the dosage of 1 mg/day, the two supplements deliver the same tooth growth
# let us use the t test
hypoth3 <-t.test(len ~ supp, data = subset(ourData, dose == 1))

tidy (hypoth3)
```

```
##   estimate estimate1 estimate2 statistic     p.value parameter conf.low
## 1     5.93      22.7     16.77   4.03277 0.001038376  15.35767 2.802148
##   conf.high                 method alternative
## 1  9.057852 Welch Two Sample t-test   two.sided
```

```
hypoth3$conf.int
```

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

```
## note Confidence interval is  2.802148 9.057852, confidence level 0.97 and p value is 0.001038
376



# conclusion: The confidence interval does not include 0 and the p-value is smaller than the 0.0
5 threshold.
# The null hypothesis can be rejected. The alternative hypothesis is accepted.

# ****
#Hypothesis #4
#For the dosage of 2 mg/day, the two supplements deliver the same tooth growth
# use the t test
hypoth4<-t.test(len ~ supp, data = subset(ourData, dose == 2))

tidy (hypoth4)
```

```
##   estimate estimate1 estimate2  statistic   p.value parameter conf.low
## 1    -0.08     26.06     26.14 -0.0461361 0.9638516  14.03982 -3.79807
##   conf.high                 method alternative
## 1   3.63807 Welch Two Sample t-test   two.sided
```

```
#Confidence interval is  -3.79807 3.63807, confidendence level is 0.95, p value is 0.9638516

# conclusion : The confidence interval does include 0 and the p-value is larger than the 0.05 th
reshold.
#The null hypothesis cannot be rejected.

#***
#Conclusions & assumptions
# conclusion
#Orange juice delivers more tooth growth than ascorbic acid for dosages 0.5 & 1.0. Orange juice
 and
#ascorbic acid deliver the same amount of tooth growth for dose amount 2.0 mg/day.
#but For the entire data set we cannot conclude orange juice is more effective that ascorbic aci
d.
#Assumptions
# we are assuming Normal distribution of the tooth lengths and No other  factors are affecting t
ooth length
```