Basic inferential data analysis

Akash Gupta

1 May 2019

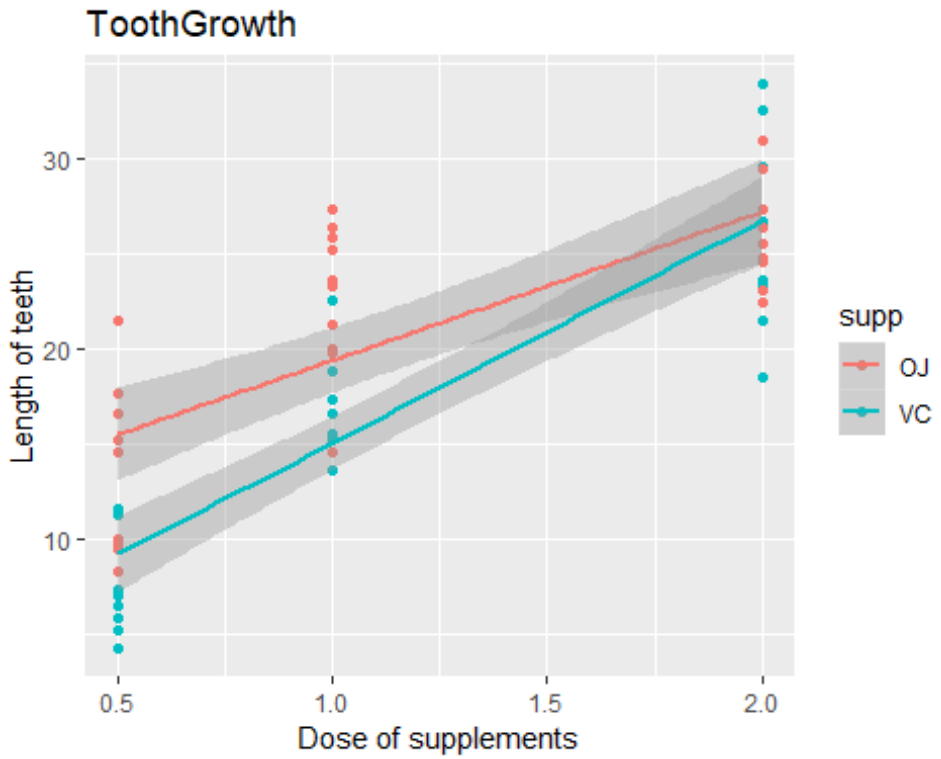
In this second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

1)Loading the ToothGrowth data and perform some basic exploratory data analyses

```
library(stats)
library(stats)
data(ToothGrowth)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.2
```

```
qplot(dose, len, data = ToothGrowth, color = supp, geom = "point") +
geom_smooth(method = "lm") + labs(title = "ToothGrowth") + labs(x = "Dose of
supplements", y = "Length of teeth")
```



From our analysis we can infer the folowing:-

a)As the amount of doses increases, the length of teeth also increases

b)With same dose of supplement “OJ” we can see that the growth of length of teeth is more

c)We can also infer that supplement “VC” has higher rate of increase of length of teeth than “OJ” as the slope of line of “VC” is more

2.)Provide a basic summary of the data.

summary of ToothGrowth 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
```

```
#sd of length
sd(ToothGrowth$len)
```

```
## [1] 7.649315
```

summary of ToothGrowth data with specefic suppliment “OJ”

```
summary(ToothGrowth[ToothGrowth$supp == "OJ", ])
```

```
##       len        supp         dose
##  Min.   : 8.20   OJ:30   Min.   :0.500
##  1st Qu.:15.53   VC: 0   1st Qu.:0.500
##  Median :22.70           Median :1.000
##  Mean   :20.66           Mean   :1.167
##  3rd Qu.:25.73           3rd Qu.:2.000
##  Max.   :30.90           Max.   :2.000
```

```
#sd of length
sd(ToothGrowth[ToothGrowth$supp =="OJ", ]$len)
```

```
## [1] 6.605561
```

summary of ToothGrowth data with specefic suppliment “VC”

```
summary(ToothGrowth[ToothGrowth$supp == "VC", ])
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ: 0   Min.   :0.500
```

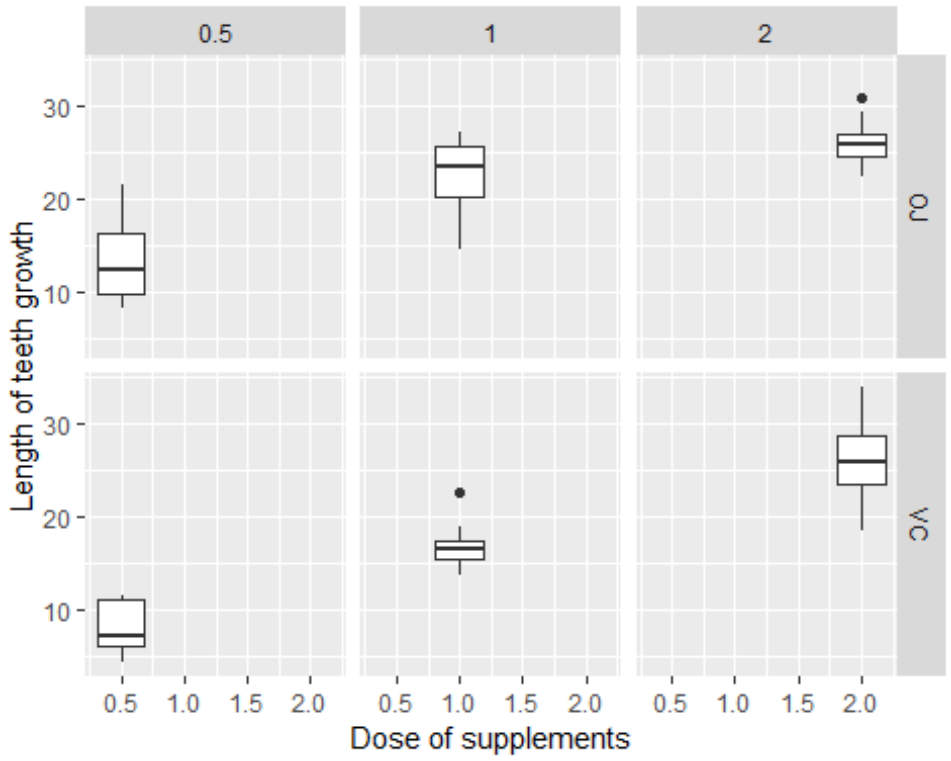
```
##  1st Qu.:11.20   VC:30   1st Qu.:0.500
##  Median :16.50           Median :1.000
##  Mean   :16.96           Mean   :1.167
##  3rd Qu.:23.10           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000

#sd of length
sd(ToothGrowth[ToothGrowth$supp =="VC", ]$len)

## [1] 8.266029
```

Effect of different dose of specific supplement of length of teeth

```
g<-ggplot(ToothGrowth,aes(dose,len,group=supp))
g+geom_boxplot()+facet_grid(supp ~ dose)+labs(x = "Dose of supplements", y =
"Length of teeth growth")
```



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

If we assume the data is normally distributed, we have a null hypothesis that there is no difference between the mean under each kind of supplements, or each dose of the supplements:

a.)

```
t.test(ToothGrowth$len,paired=F)$conf
```

```
## [1] 16.83731 20.78936
## attr(,"conf.level")
## [1] 0.95
```

we can say that 95% of time our true mean lie in between this interval

Now ,we calculate the mean under each kind of supplements:

```
mean(ToothGrowth[ToothGrowth$supp == "OJ", ]$len)
```

```
## [1] 20.66333
```

```
mean(ToothGrowth[ToothGrowth$supp == "VC", ]$len)
```

```
## [1] 16.96333
```

we can see that both mean lie within the interval so we fail to rejec the null hypothesis

b.)

Now ,we calculate the mean under each kind of dose:

```
mean(ToothGrowth[ToothGrowth$dose == 0.5, ]$len)
```

```
## [1] 10.605
```

```
mean(ToothGrowth[ToothGrowth$dose == 1.0, ]$len)
```

```
## [1] 19.735
```

```
mean(ToothGrowth[ToothGrowth$dose == 2.0, ]$len)
```

```
## [1] 26.1
```

Thus the mean of teeth growth after taking dose of 0.5 is 10.6; the mean of teeth growth after taking dose of 1.0 is 19.74; the mean of teeth growth after taking dose of 2.0 is 26.1. We are able to reject the null hypotheis, and there is a difference in teeth growth between each dose of supplements.

c.)

Now ,we calculate the mean under each kind of dose and each kind of supplement:

```
#mean growth of teeth under dose of 0.5 for both supplement
dose05 <- ToothGrowth[ToothGrowth$dose == 0.5, ]
t.test(x = dose05$len, paired = FALSE, conf.level = 0.95)$conf

## [1]  8.499046 12.710954
## attr(,"conf.level")
## [1] 0.95

mean(dose05[dose05$supp == "VC", ]$len)

## [1] 7.98

mean(dose05[dose05$supp == "OJ", ]$len)

## [1] 13.23
```

Thus under the dose of 0.5, there are 95% of the time that a confidence interval between 8.50 and 12.71 will contain the true population mean. We also know that the mean of teeth growth after taking 0.5 dose of VC is 7.98 and the mean of teeth growth after taking 0.5 dose of OJ is 13.23. We rejected the null hypo.

```
#mean growth of teeth under dose of 1.0 for both supplement
dose10 <- ToothGrowth[ToothGrowth$dose == 1.0, ]
t.test(x = dose10$len, paired = FALSE, conf.level = 0.95)$conf

## [1] 17.66851 21.80149
## attr(,"conf.level")
## [1] 0.95

mean(dose10[dose10$supp == "VC", ]$len)

## [1] 16.77

mean(dose10[dose10$supp == "OJ", ]$len)

## [1] 22.7
```

Thus under the dose of 1.0, there are 95% of the time that a confidence interval between 17.67 and 21.80 will contain the true population mean. We also know that the mean of teeth growth after taking 1.0 dose of VC is 16.77 and the mean of teeth growth after taking 1.0 dose of OJ is 22.7. We rejected the null hypo.

```
#mean growth of teeth under dose of 2.0 for both supplement
dose20 <- ToothGrowth[ToothGrowth$dose == 2.0, ]
t.test(x = dose20$len, paired = FALSE, conf.level = 0.95)$conf

## [1] 24.33364 27.86636
## attr(,"conf.level")
## [1] 0.95

mean(dose20[dose20$supp == "VC", ]$len)

## [1] 26.14

mean(dose20[dose20$supp == "OJ", ]$len)

## [1] 26.06
```

Thus under the dose of 2.0, there are 95% of the time that a confidence interval between 24.33 and 27.87 will contain the true population mean. We also know that the mean of teeth growth after taking 2.0 dose of VC is 26.14 and the mean of teeth growth after taking 2.0 dose of OJ is 26.06. Both of them are within the confidence interval. We fail to reject the null hypothesis.

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

The conclusion is when the dose is 0.5 or 1.0 there is a difference between the teeth growth after taking OJ and VC, while when the dose is 2.0, there is no difference between the teeth growth after taking OJ and VC. The assumption needed is we first assumed the whole population is normally distributed, then we assumed the population is normally distributed under each dose