Part 2 ToothGrowth data

Joaquin Arango

12/24/2020

Part 2: Load ToothGrowth data and perform basic exploratory analysis, and provide basic statistics

```
knitr::opts_chunk$set(echo = TRUE)
#Load packages
library(datasets)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(ggthemes)
data("ToothGrowth")

#Load data and convert to tbl format

ToothGrowth <- tbl_df(ToothGrowth)
```

```
## Warning: `tbl_df()` is deprecated as of dplyr 1.0.0.
## Please use `tibble::as_tibble()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

Review structure of dataset and summarize

```
#Sructure of the dataframe
ToothGrowth %>% str()
```

```
## tibble [60 x 3] (S3: tbl_df/tbl/data.frame)
##  $ len : num [1:60] 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 [1:60] 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```r
#summary
ToothGrowth %>% summary()
```

```
##       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 dataset has 60 observations and 3 variables

len: tooth length, numeric variable supp: supplement type (VC: vitamin C or OJ: Orange Juice), factor variable dose: dose(in milligrams), numeric variable

```r
# Unique values in the dose vector
ToothGrowth %>% select(dose) %>% unique()
```

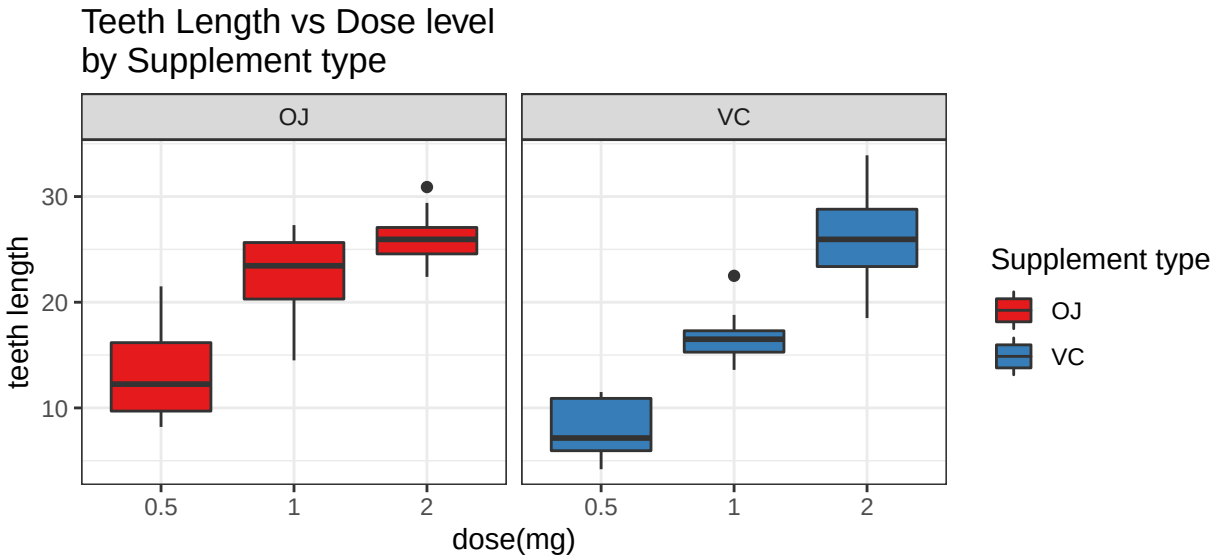
```
## # A tibble: 3 x 1
##    dose
##   <dbl>
## 1   0.5
## 2   1
## 3   2
```

Dose contians just 3 unique values: 0.5, 1, 2. Convert this ino a factor varible with three levels

```r
# covert to factor
ToothGrowth <- ToothGrowth %>% mutate(dose = as.factor(dose))
```

Now we plot

```r
ToothGrowth %>%
  ggplot(aes(x=dose, y=len, fill=supp)) +
  geom_boxplot() +
  facet_grid(. ~supp) +
  scale_fill_brewer(palette="Set1") +
  theme_bw() +
  ggtitle("Teeth Length vs Dose level \nby Supplement type") +
  labs(x="dose(mg)", y="teeth length") +
  guides(fill=guide_legend(title = "Supplement type"))
```



This plot shows the relationship between teeth length and dose level for each supplement type. There is a positive relationship for both supplement types. As the amount of supplement type increases, so does teeth length.

Now we want to further compare teeth growth by supplement type and dose levels. We will use the t-test. In our dataset we have two supp: OJ and VC and three levels for dose: 0.5, 1, 2. Thus we'll have to run one hypothesis test for factor *supp* and one for each possible pair of the 3 levels in the factor *dose*, so we will run a total of 4 tests.

Test 1 dose = 0.5 and dose = 1

```
# Extract the len and dose vectors from the df ToothGrowth
len_a <- ToothGrowth %>% filter(dose %in% c(0.5,1)) %>% select(len) %>% unlist()
dose_a <- ToothGrowth %>% filter(dose %in% c(0.5,1)) %>% select(dose) %>% unlist()
(Test.a <- t.test(len_a~dose_a, paired=FALSE))
```

```
##
##  Welch Two Sample t-test
##
## data:  len_a by dose_a
## 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
```

Test 2 dose = 0.5 and dose = 2

```
#Exract the len and dose vectors from the df ToothGrowth
len_b <- ToothGrowth %>% filter(dose %in% c(0.5,2)) %>% select(len) %>% unlist()
dose_b <- ToothGrowth %>% filter(dose %in% c(0.5,2)) %>% select(dose) %>% unlist()
#Test
(Test.b <- t.test(len_b~dose_b, paired = FALSE))
```

```
##
##  Welch Two Sample t-test
##
## data:  len_b by dose_b
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

Test 3 dose = 1 and dose = 2

```
#Exract the len and dose vectors from the df ToothGrowth
len_c <- ToothGrowth %>% filter(dose %in% c(1,2)) %>% select(len) %>% unlist()
dose_c <- ToothGrowth %>% filter(dose %in% c(1,2)) %>% select(dose) %>% unlist()
#Test
(Test.c <- t.test(len_c~dose_c, paired = FALSE))
```

```
##
##  Welch Two Sample t-test
##
## data:  len_c by dose_c
## 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
```

We went through all the combinations of levels from the factor variable dose and in all cases the p-value is lower than the default significance level 0.05. Thus, we reject Ho. There appears to be a positive relationship between dose level and teeth length

Test by supplement

```
#Exract the len and dose vectors from the df ToothGrowth
len <- ToothGrowth %>% select(len) %>% unlist()
supp <- ToothGrowth %>% select(supp) %>% unlist()
t.test(len~supp, paired = FALSE)
```

```
##
##  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
```

We can see that the p-value of the test is 0.06. Since the p-value is greater than 0.05 and the confidence interval of the test contains zero, we can reject the null hypothesis and say that supplement types don't seem to have any impact on teeth growth. There is no significant statistical difference between them.

Conclusions

There is a statistically significant difference between teeth length and dose levels across both delivery methods, as the dose increases so does teeth length.

There doesn't seem to be a statistically significant difference between delivery methods, with OJ apparently more effective at dose levels 0.5 and 1, and VC slightly more effective at dose level 2.