Tooth Data Inferential Analysis

Neel

9/18/2020

Load dataframe and basic expanatory data alaysis

Load dataframe from R datasets into new dataframe teeth and plotting system

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

```
library(datasets)
teeth <- ToothGrowth
```

Look at the types of information in the columns of the dataframe and shape.

```
str(teeth)
```

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

Look at the first 5 rows of the dataframe.

```
head(teeth, 5)
```

```
##    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
```

Summary of data

Summary of dataframe from summary function

```
summary(teeth)
```

```
##       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
```

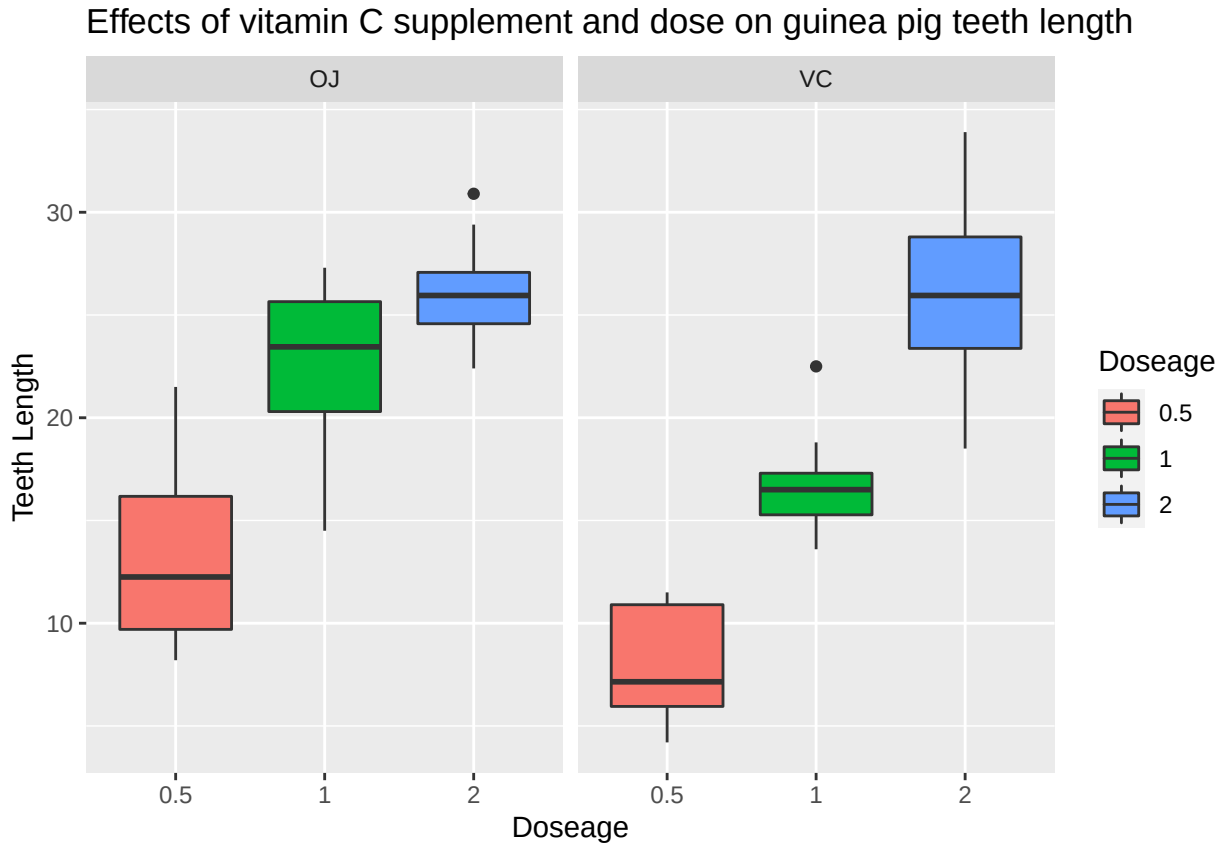
Aggregated data by dose and supplement given

```
aggregate_teeth <- aggregate(len~dose + supp, teeth, mean)
aggregate_teeth
```

```
##   dose supp   len
## 1  0.5   OJ 13.23
## 2  1.0   OJ 22.70
## 3  2.0   OJ 26.06
## 4  0.5   VC  7.98
## 5  1.0   VC 16.77
## 6  2.0   VC 26.14
```

Visualize data using ggplot2 boxplots

```
ggplot(teeth, aes(x = factor(dose), y = len, fill = factor(dose))) + geom_boxplot() +
facet_grid(.~supp) +
labs(title = "Effects of vitamin C supplement and dose on guinea pig teeth length",
x = "Doseage", y = "Teeth Length") + scale_fill_discrete(name="Doseage")
```



Test effectiveness of treatment and doseage

Our null hypothesis is that there is no difference in tooth length by supplement given regardless of the doseage. The alternative hypothesis is that the Vitamin C supplement should be greater than that of regular orange juice.

```
oj <- teeth[teeth$supp == "OJ", 1]
vc <- teeth[teeth$supp == "VC", 1]

oj_vs_vc <- t.test(oj, vc, alternative = "greater", conf.level = 0.95)
```

Our null hypothesis for the three tests are that between all combinations of doseages the increase of dose does not increase tooth length. Our alternative hypothesis is that the greater dose leads to an increase in tooth length.

```
half_dose = teeth[teeth$dose == 0.5, 1]
one_dose = teeth[teeth$dose == 1, 1]
two_dose = teeth[teeth$dose == 2, 1]

half_vs_one <- t.test(half_dose, one_dose, alternative = "less", conf.level = 0.95)
half_vs_two <- t.test(half_dose, two_dose, alternative = "less", conf.level = 0.95)
one_vs_two <- t.test(one_dose, two_dose, alternative = "less", conf.level = 0.95)
```

Our null hypothesis for the three tests are that between dose and supplement there is no difference between

dose by supplement. Our alternative hypothesis is that there is a difference between dose by choice of supplement.

```
supp_half <- teeth[teeth$dose == 0.5, ]
supp_one <- teeth[teeth$dose == 1, ]
supp_two <- teeth[teeth$dose == 2, ]

half_vs_supp <- t.test(len ~ supp, data = supp_half, alternative = "two.sided")
one_vs_supp <- t.test(len ~ supp, data = supp_one, alternative = "two.sided")
two_vs_supp <- t.test(len ~ supp, data = supp_two, alternative = "two.sided")
```

Data frame of p values and lower and upper bounds of the 95% confidence interval.

```
results <- data.frame("p_val" = c(oj_vs_vc$p.value, half_vs_one$p.value, half_vs_two$p.value,
one_vs_two$p.value, half_vs_supp$p.value, one_vs_supp$p.value,
two_vs_supp$p.value),
"lower interval" = c(oj_vs_vc$conf.int[1], half_vs_one$conf.int[1], half_vs_two$conf.int[1],
one_vs_two$conf.int[1], half_vs_supp$conf.int[1], one_vs_supp$conf.int[1], two_vs_supp$conf.int[1]),
"upper interval" = c(oj_vs_vc$conf.int[2], half_vs_one$conf.int[2], half_vs_two$conf.int[2],
one_vs_two$conf.int[2], half_vs_supp$conf.int[2], one_vs_supp$conf.int[2], two_vs_supp$conf.int[2]),
row.names = c("oj_vs_vc", "half_vs_one", "half_vs_two",
"one_vs_two", "half_vs_supp", "one_vs_supp", "two_vs_supp"))
results
```

```
##                     p_val lower.interval upper.interval
## oj_vs_vc     3.031725e-02      0.4682687            Inf
## half_vs_one  6.341504e-08           -Inf      -6.753323
## half_vs_two  2.198762e-14           -Inf     -13.279257
## one_vs_two   9.532148e-06           -Inf      -4.173870
## half_vs_supp 6.358607e-03      1.7190573       8.780943
## one_vs_supp  1.038376e-03      2.8021482       9.057852
## two_vs_supp  9.638516e-01     -3.7980705       3.638070
```

Conclusions

We are 95% confident that we reject the null hypothesis such that there is a significant increase in tooth length for guinea pigs with Vitamin C instead of orange juice. We are 95% confident that we reject the null hypothesis in all three cases in favor of the alternative hypothesis such that increasing the doseage of vitamin C regardless of the supplement increases the length of guinea pig teeth. We are 95% confident that we reject the null hypothesis such that there is a significant increase in tooth length for guinea pigs when the doseage is 0.5 or 1mg between supplements. At the 95% confidence interval we fail to reject the null such that there is not a significant difference in guinea pig teeth length in a dose of 2mg between supplements.