data-analysis

Tasks

1. Load the ToothGrowth data and perform some basic exploratory data analysis
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 the conclusions and the assumptions needed for my conclusions.

Load libraries

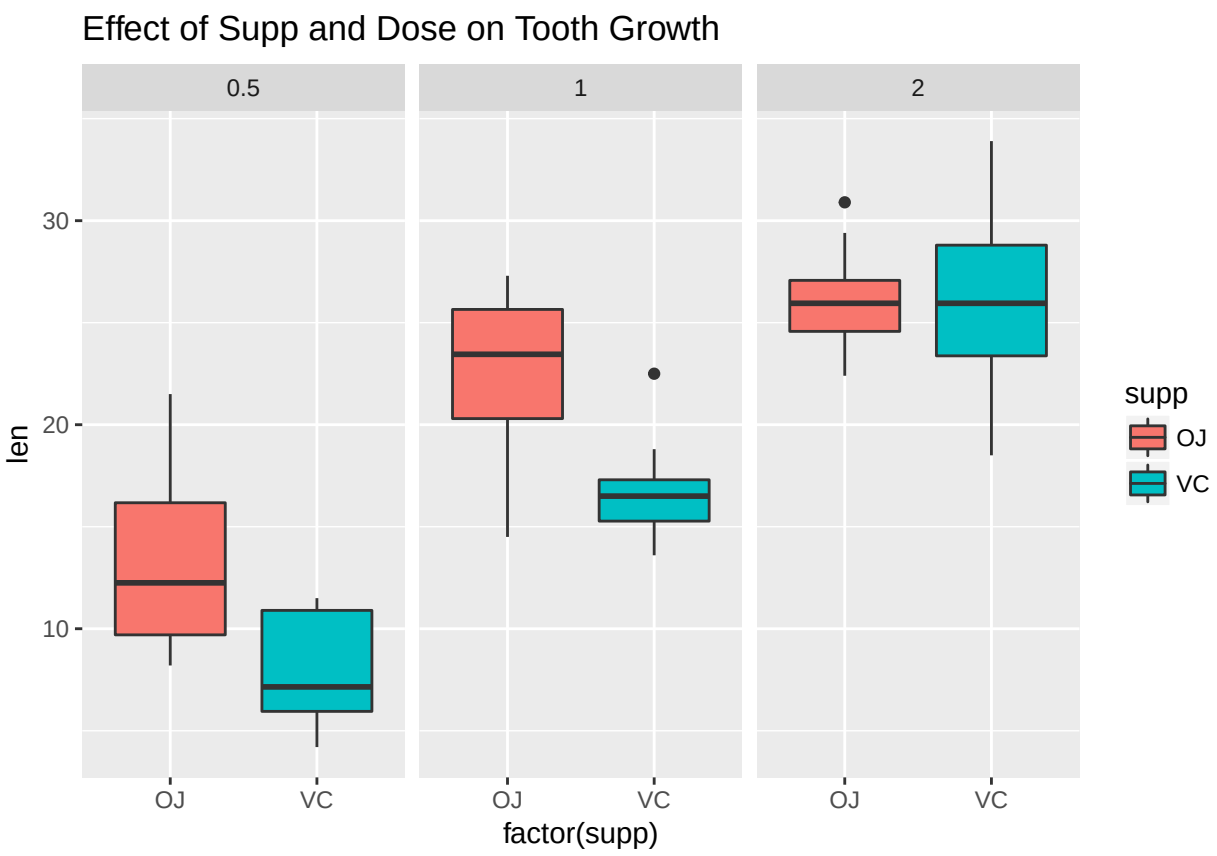
```
library(ggplot2)
library(dplyr, warn.conflicts = FALSE)
library(datasets)
```

Load data

```
data(ToothGrowth)
tooth_growth <- ToothGrowth
```

Exploratory data analysis

```
p <- ggplot(tooth_growth,aes(factor(supp),len))
p+geom_boxplot(aes(fill=supp))+facet_grid(.~dose)+labs(title="Effect of Supp and Dose on Tooth Growth")
```



Based on the graph, both supps seem to have an positive impact on tooth growth. Increasing the dose of OJ from 0.5 to 1 seems to have a much bigger impact than from 1 to 2. On the other hand, increasing the dose of VC seems to help tooth growth more steadily; its impact matches that of OJ at dose of 2. I don't see much of a pattern in distributions.

Summarize data

Summarize data

```
str(tooth_growth)
```

```
## '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(tooth_growth)
```

```
##       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
```

```
head(tooth_growth)
```

```
##    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
```

```
unique(tooth_growth$dose)
```

```
## [1] 0.5 1.0 2.0
```

Based on the summary, there are 60 observations with two variations of supp(OJ & VC) and three variations of dose(0.5, 1 & 2).

Tooth growth comparison by supp and dose

Test if there is a difference between the two supps on tooth growth impact

```
len <- tooth_growth %>% select(len) %>% unlist()
supp <- tooth_growth %>% 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
```

The p-value is higher than 0.5, so we conclude that there is no difference in impacts on tooth growth between the two supps.

Test if there is a difference between different dose levels on tooth growth impact

Between dose 1 and 2

```
len_1_2 <- tooth_growth %>% filter(dose %in% c(1,2)) %>% select(len) %>% unlist()
dose_1_2 <- tooth_growth %>% filter(dose %in% c(1,2)) %>% select(dose) %>% unlist()
(test_1_2 <- t.test(len_1_2~dose_1_2, paired = FALSE))
```

```
##
##  Welch Two Sample t-test
##
## data:  len_1_2 by dose_1_2
## 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
```

Between dose 0.5 and 1

```
len_0.5_1 <- tooth_growth %>% filter(dose %in% c(0.5,1)) %>% select(len) %>% unlist()
dose_0.5_1 <- tooth_growth %>% filter(dose %in% c(0.5,1)) %>% select(dose) %>% unlist()
(test_0.5_1 <- t.test(len_0.5_1~dose_0.5_1, paired = FALSE))
```

```
##
##  Welch Two Sample t-test
##
## data:  len_0.5_1 by dose_0.5_1
## 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
```

In both tests, the P-values are significantly lower than 0.05, so we reject the null hypothesis in both cases. We conclude that there is a significant difference in tooth growth by increasing dose level.

Conclusions and assumptions

Conclusions:

1. There is no significant difference between the impact of the two supps on tooth growth.
2. There is a significant difference between lower and high dose levels in increasing tooth growth.

Assumptions:

1. The population distribution is normal.
2. The dose and supp are randomly assigned.
3. The sample is representative of the population.