Statistical Inference Course Project 2

hiroo.ishikawa

2019/7/31

```
library(tidyverse)
```

Basic inferential data analysis

Use the buildin datasets 'ToothGrowth', perform some basic exploratory data analyses.

The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC).

First, Check the characteristics of the data.

```
head(ToothGrowth,4)
```

```
##    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
```

```
summary(ToothGrowth)
```

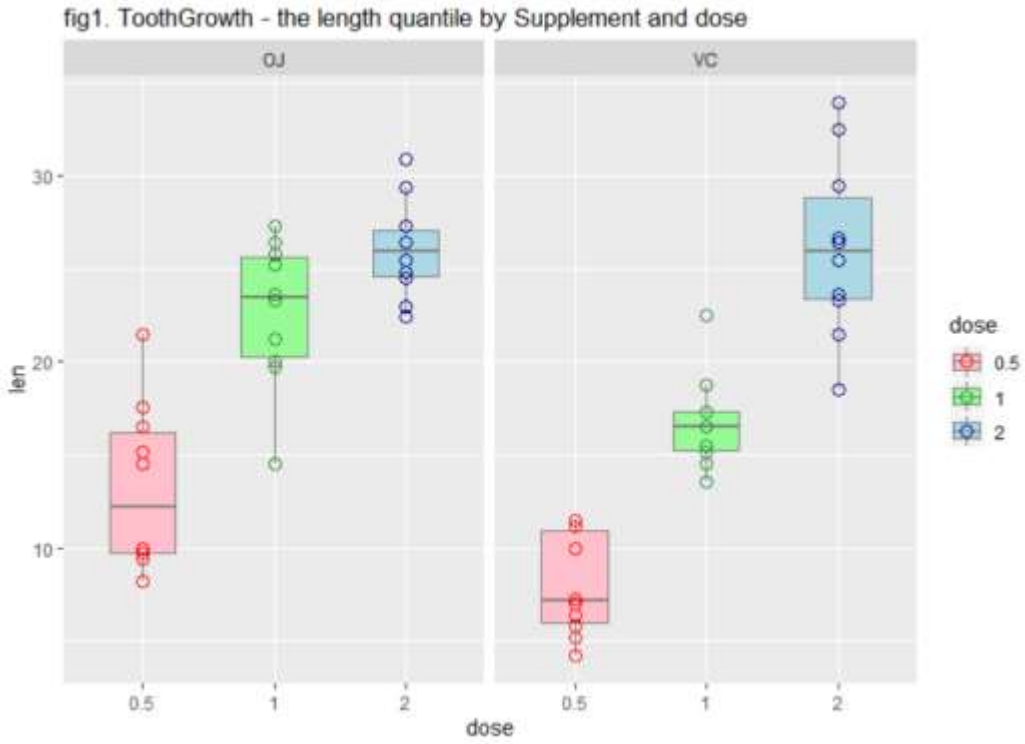
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
table(ToothGrowth$supp,ToothGrowth$dose)
```

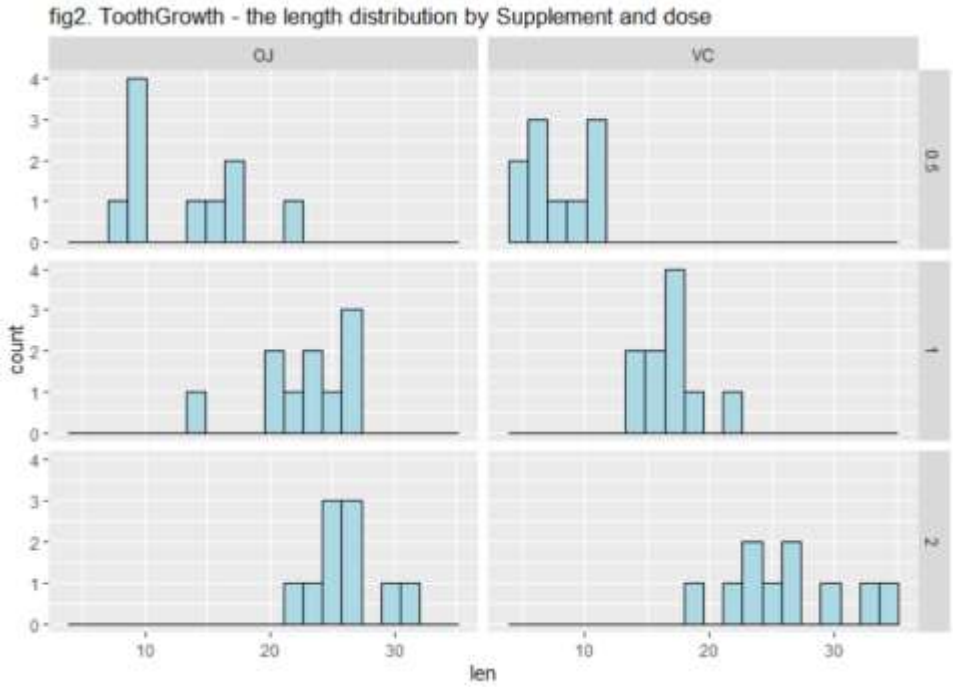
```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

```
ToothGrowth <- ToothGrowth %>% mutate(dose=factor(dose))

ggplot(ToothGrowth, aes(x=dose,y=len)) +
    geom_boxplot(aes(fill=dose), color='gray50', width=0.5, outlier.shape=NA) +
    geom_point(  aes(color=dose), shape=1, size=3 ) +
    facet_wrap( ~supp) + ggtitle('fig1. ToothGrowth - the length quantile by Supplement and dose') +
    scale_color_manual(values=c('red','forestgreen','navyblue')) +
    scale_fill_manual( values=c('pink','palegreen','lightblue'))
```




```
ggplot(ToothGrowth, aes(x=len)) +
    geom_histogram(bins=20,color='black',fill='lightblue') +
    facet_grid(dose~supp) + ggtitle('fig2. ToothGrowth - the length distribution by Supplement and d
ose')
```



Basic summary of the data

The conditions by combination of 2 Supplement type (OJ and VC) and 3 dose levels(0.5, 1, 2). The total 60 length data from 10 samples for each conditions. the samples distribution of each condition's looks like nomal distribution.

Compare the effection of dose for each supplement

Apply hypothesis tests(t-test) to confirm relation the teeth length growth with dose levels.

A. supplement: Orange Juice

1) H_0: no difference between mean of dose=0.5 and dose=0.1 H_a: mean(dose=0.5) < mean(dose=1)

```
g1 <- ToothGrowth %>% filter(supp=='OJ',dose==0.5) %>% pull(len)
g2 <- ToothGrowth %>% filter(supp=='OJ',dose==1  ) %>% pull(len)
p  <- t.test(g2,g1,paired=FALSE,var.equal=TRUE,alternative='greater')$p.value
cat('T test: p-value=', p, ' < 0.05 is reject the null hypothesis.\n')

## T test: p-value= 4.17878e-05  < 0.05 is reject the null hypothesis.
```

It is statistically significant that length at dose=1 is more greater than dose=0.5.

2) H_0: no difference between mean of dose=1 and dose=2 H_a: mean(dose=0.5) < mean(dose=1)

```
g1 <- ToothGrowth %>% filter(supp=='OJ',dose==1  ) %>% pull(len)
g2 <- ToothGrowth %>% filter(supp=='OJ',dose==2  ) %>% pull(len)
p  <- t.test(g2,g1,paired=FALSE,var.equal=TRUE,alternative='greater')$p.value
cat('T test: p-value=', p, ' < 0.05 is reject the null hypothesis.\n')

## T test: p-value= 0.0186814  < 0.05 is reject the null hypothesis.
```

It is statistically significant that length at dose=2 is more greater than dose=1.

B. supplement: Vitamine C

1) H_0: no difference between mean of dose=0.5 and dose=0.1 H_a: mean(dose=0.5) < mean(dose=1)

```
g1 <- ToothGrowth %>% filter(supp=='VC',dose==0.5) %>% pull(len)
g2 <- ToothGrowth %>% filter(supp=='VC',dose==1  ) %>% pull(len)
p  <- t.test(g2,g1,paired=FALSE,var.equal=TRUE,alternative='greater')$p.value
cat('T test: p-value=', p, ' < 0.05 is reject the null hypothesis.\n')

## T test: p-value= 3.246132e-07  < 0.05 is reject the null hypothesis.
```

It is statistically significant that length at dose=1 is more greater than dose=0.5.

2) H_0: no difference between mean of dose=1 and dose=2 H_a: mean(dose=0.5) < mean(dose=1)

```
g1 <- ToothGrowth %>% filter(supp=='VC',dose==1  ) %>% pull(len)
g2 <- ToothGrowth %>% filter(supp=='VC',dose==2  ) %>% pull(len)
p  <- t.test(g2,g1,paired=FALSE,var.equal=TRUE,alternative='greater')$p.value
cat('T test: p-value=', p, ' < 0.05 is reject the null hypothesis.\n')

## T test: p-value= 1.698789e-05  < 0.05 is reject the null hypothesis.
```

It is statistically significant that length at dose=2 is more greater than dose=1.

Conclusion

In both supplimant type OJ and VC, there is a statistically significant as the dose increases, the mean of teeth length is longer, by one can reject the null hypothesis that the true no difference in means.