statistical inference

Renukaneeli

6/2/2021

```
library(datasets)
str(ToothGrowth)
```

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

```
data(ToothGrowth)
head(ToothGrowth)
```

```
##    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
```

```
##Summary
## We check if there are NA values and display the summary information for this dataset :

sum(!complete.cases(ToothGrowth))
```

```
## [1] 0
```

```
summary(ToothGrowth)
```

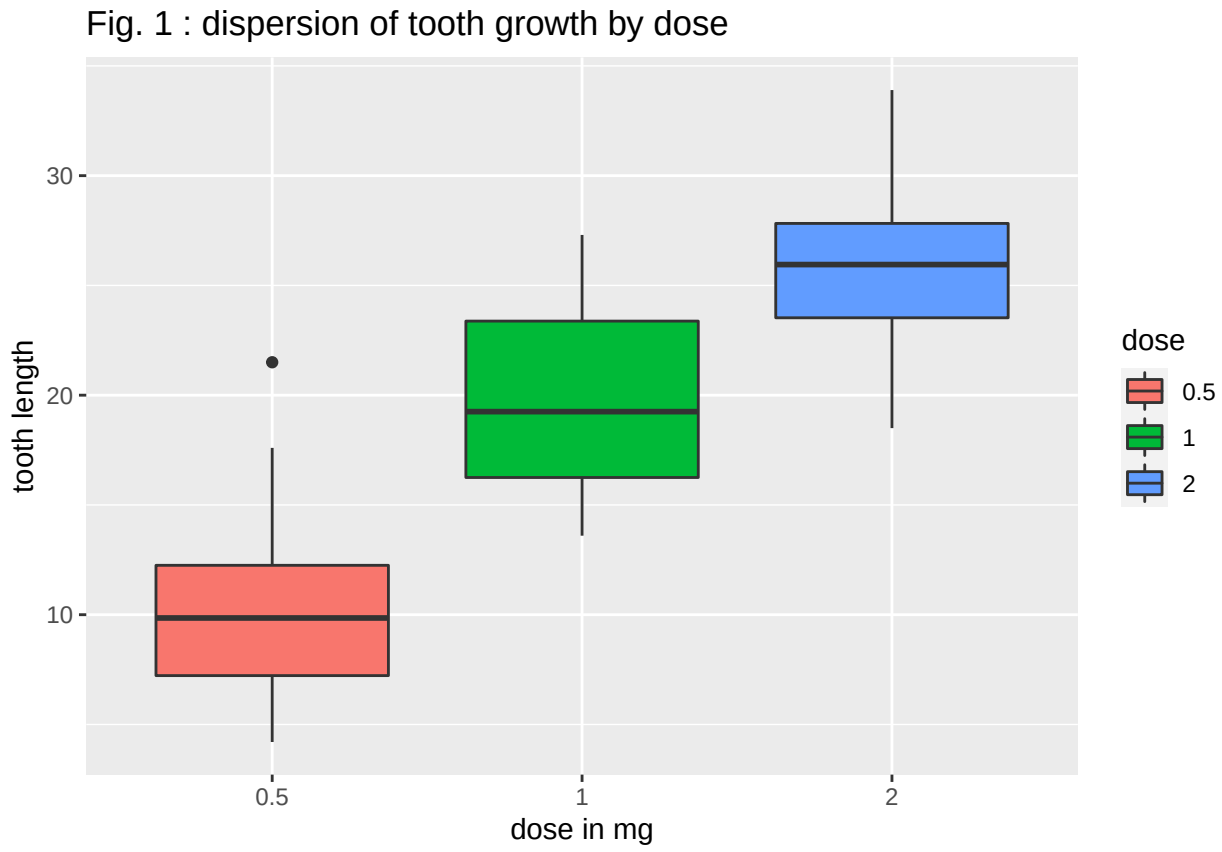
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

Exploratory Analysis In this section, we will give a summary exploratory analysis

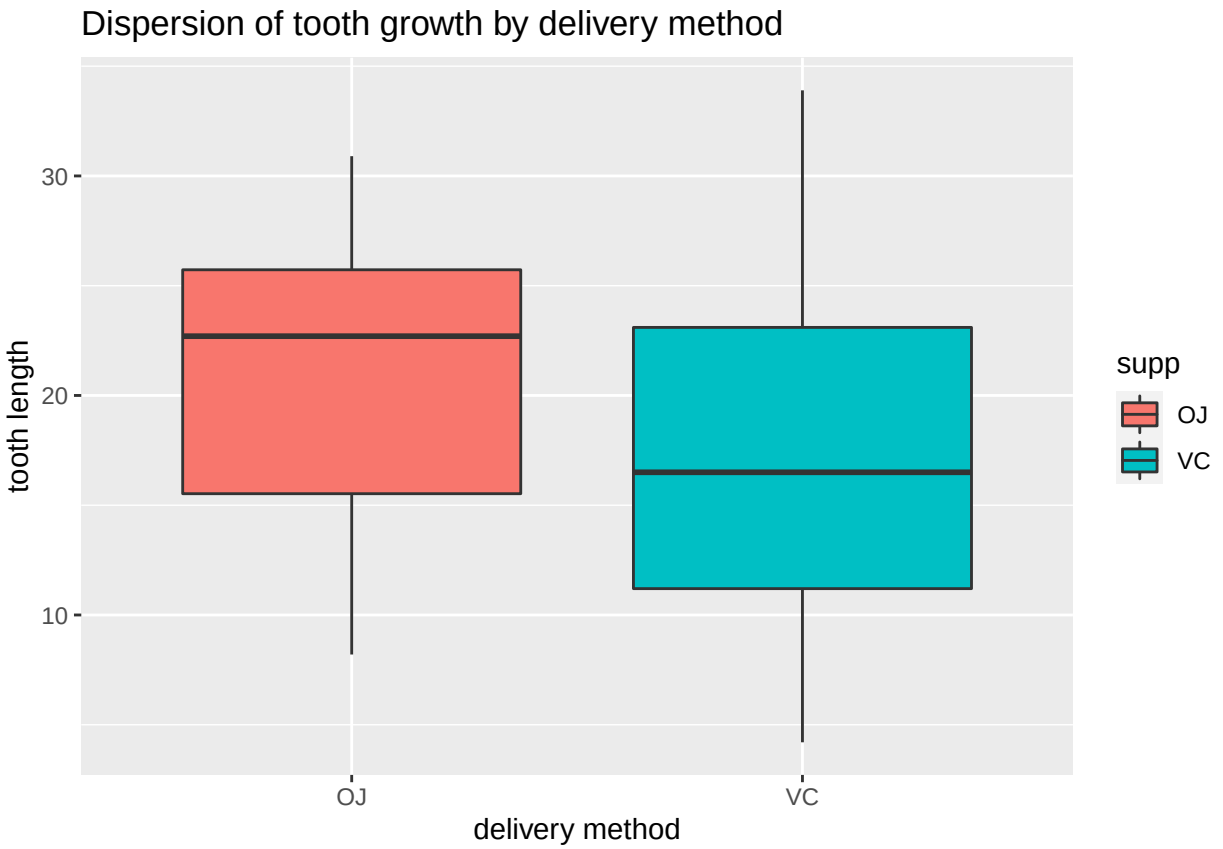
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
p <- ggplot(ToothGrowth, aes(x=dose, y=len,fill=dose)) +  geom_boxplot() +  ggtitle("Fig. 1 : dispersio
p
```




```
p1 <- ggplot(ToothGrowth, aes(x=supp, y=len,fill=supp)) +  geom_boxplot() +  ggtitle("Dispersion of too
p1
```




```
dose <- ToothGrowth$dose
supp <- ToothGrowth$supp
len <-  ToothGrowth$len

t.test(len[supp == "VC"],len[supp == "OJ"], paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len[supp == "VC"] and len[supp == "OJ"]
## 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:
##  -7.5710156  0.1710156
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

```
t.test(len[dose == 0.5],len[dose == 1], paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len[dose == 0.5] and len[dose == 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 of x mean of y
##    10.605    19.735
```