Analyzing Tooth Growth Data

Vaibhav Palve

```
library(datasets)
data("ToothGrowth")
attach(ToothGrowth)
```

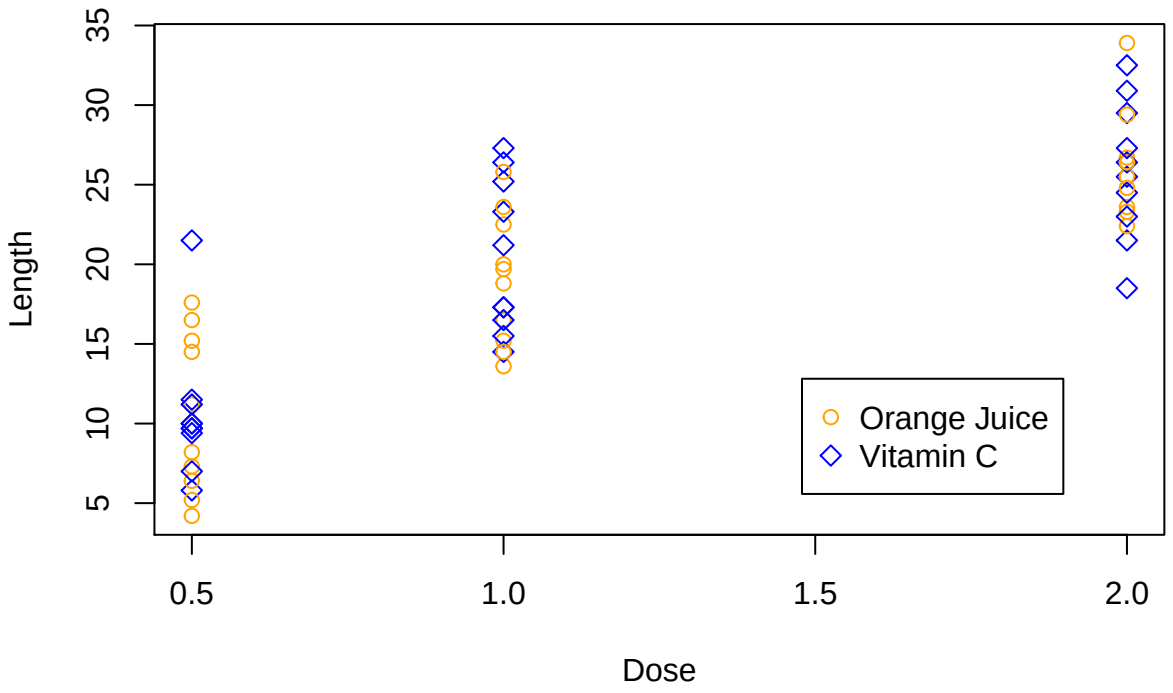
The structure of the dataset :

```
str(ToothGrowth)
```

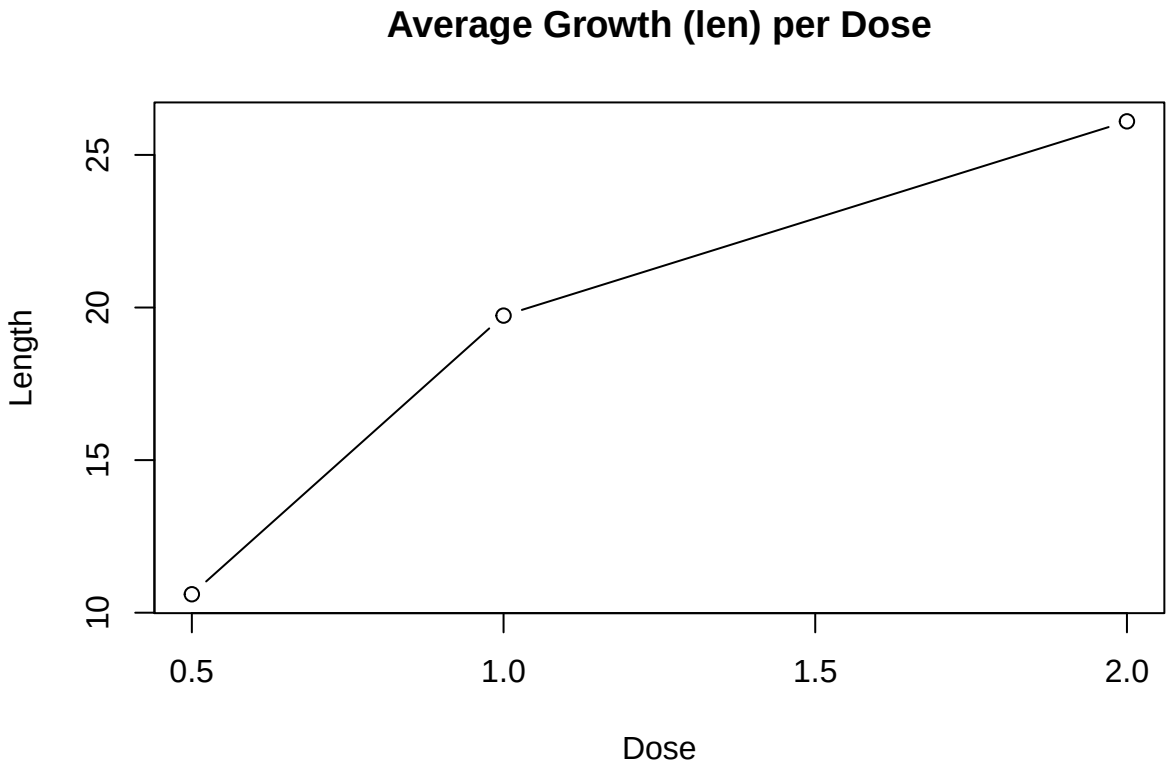
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
# Plotting the raw data

plot(dose,len,xlab="Dose",ylab= "Length",pch = c(1, 5), col = c("orange", "blue"))
legend("bottomright", inset = c(0.1, 0.08), c("Orange Juice", "Vitamin C"),
    pch = c(1, 5), col = c("orange", "blue"))
```



```r
#Plotting the Average Length per dose
ToothAggr <- aggregate(len ~ dose, , mean)
plot(ToothAggr, type = "b", xlab = "Dose", ylab = "Length", main = "Average Growth (len) per Dose")
```



The basic summary of the data :

```r
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
```

Hypothesis Testing :

The Assumptions are: 1. The variables (len, dose, supp) must be independent and identically distributed (i.i.d.).

2. Variances of tooth growth are different when using different supplement and dosage.
3. Tooth growth follows a normal distribution.

Hypothesis 1 :

Let our null hypothesis to be there is no difference in tooth growth when using the supplement OJ and VC i.e len_OJ = len_VC

Let our alternate hypothesis to be there are more tooth growth when using supplement OJ than VC i.e len_OJ > len_VC

Now Extracting the tooth growth by supplement type

```
OJ = ToothGrowth$len[ToothGrowth$supp == 'OJ']
VC = ToothGrowth$len[ToothGrowth$supp == 'VC']
```

t-test : One-tailed independent t-test with unequal variance.

```
t.test(OJ, VC, alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ and VC
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

As the p-value is 0.03032 which is less than 0.05, we reject the null hypothesis - there is only a 3% chance that the difference in mean tooth growth is due to random chance.

Hence, we can conclude that it is very likely that supplement OJ, the greater the effect on tooth growth than supplement VC.

Hypothesis 2 :

Let our null hypothesis to be there is no difference in tooth growth between dosage. len_0.5 = len_1.0 = len_2.0

Let our alternate hypothesis to be there are more tooth growth when the dosage increases. len_0.5 > len_1.0 > len_2.0

Extracting the tooth growth by dosage:

```
# split data set
doseHalf = ToothGrowth$len[ToothGrowth$dose == 0.5]
doseOne = ToothGrowth$len[ToothGrowth$dose == 1]
doseTwo = ToothGrowth$len[ToothGrowth$dose == 2]
```

```r
t.test(doseHalf, doseOne, alternative = "less", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

t-test :

```
##
##  Welch Two Sample t-test
##
## data:  doseHalf and doseOne
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.753323
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

```r
t.test(doseOne, doseTwo, alternative = "less", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  doseOne and doseTwo
## t = -4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -4.17387
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

As the p-values are 6.342e-08 and 9.532e-06, which are less than 0.05, we reject the null hypothesis - there is only less than 0.1% chance that the differences in mean tooth growth is due to random chance.

Hence, we can conclude that it is very likely that higher the dosage, the greater the effect on tooth growth.

Hypothesis 3 :

Let our null hypothesis to be there is no difference in tooth growth when using the supplement OJ and VC at dosage 2.0 mg.

And let our alternate hypothesis to be there is difference tooth growth when using supplement OJ and VC at dosage 2.0 mg.

```r
OJ2 = ToothGrowth$len[ToothGrowth$supp == 'OJ' & ToothGrowth$dose == 2]
VC2 = ToothGrowth$len[ToothGrowth$supp == 'VC' & ToothGrowth$dose == 2]
```

```r
t.test(OJ2, VC2, alternative = "two.sided", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

t-test :

```
##
##  Welch Two Sample t-test
##
## data:  OJ2 and VC2
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

As the p-value is 0.9639 which is more than 0.05, we cannot reject the null hypothesis.

Hence, we can only conclude that there is insufficient evidence to show that there is a difference in tooth growth when using supplement OJ and VC at dosage 2.0 mg.