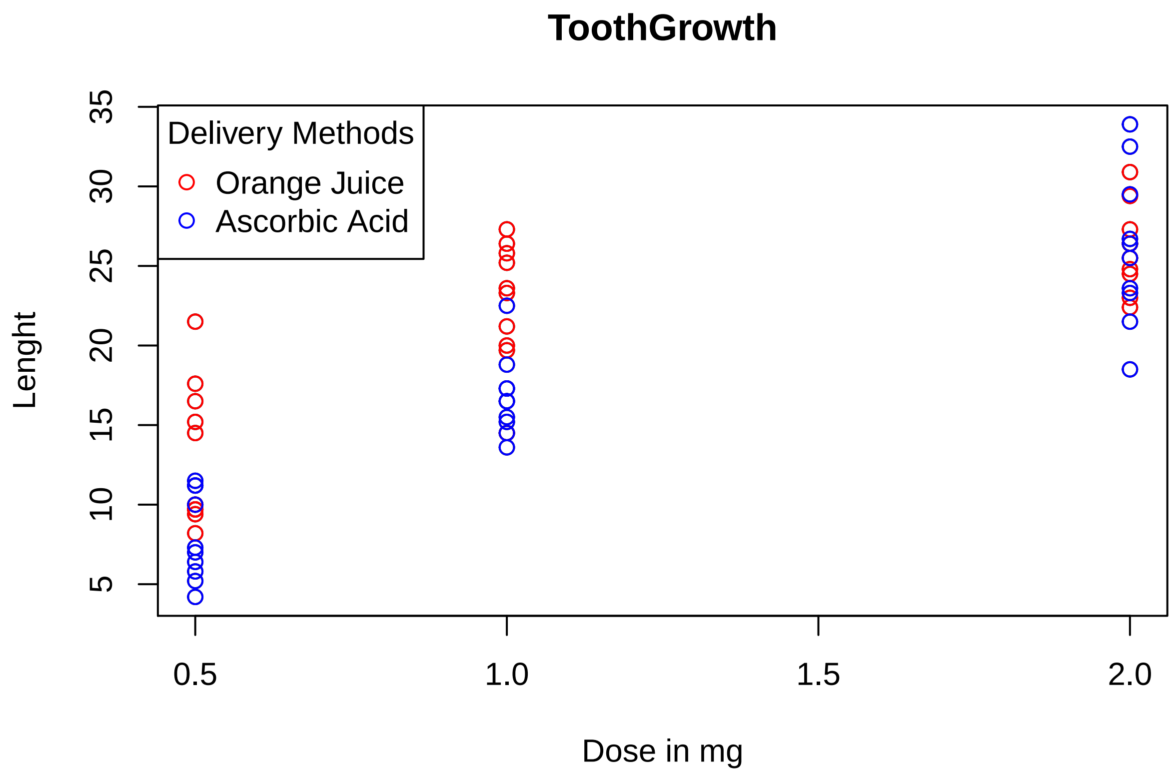
Basic inferential data analysis.

1 Load the ToothGrowth data and perform some basic exploratory data analyses

The ToothGrowth database measures the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

The following plot illustrates the ToothGrowth databe. It shows the length of teeth by dose a delivery methods.

```
data <- ToothGrowth
plot(data$dose,data$len,xlab="Dose in mg",ylab="Lenght",main="ToothGrowth")
points(data$dose[data$supp == "OJ"],data$len[data$supp == "OJ"], col = "red")
points(data$dose[data$supp == "VC"],data$len[data$supp == "VC"], col = "blue")
legend("topleft", legend = c("Orange Juice","Ascorbic Acid"), col = c("red","blue"), pch = 1,title="Del
```



2 Provide a basic summary of the data.

The following codes give a basic summary of the database.

```
summary(ToothGrowth)
```

```
##       len        supp         dose     
##  Min.   : 4.2   OJ:30   Min.   :0.50  
##  1st Qu.:13.1   VC:30   1st Qu.:0.50  
##  Median :19.2           Median :1.00  
##  Mean   :18.8           Mean   :1.17  
##  3rd Qu.:25.3           3rd Qu.:2.00  
##  Max.   :33.9           Max.   :2.00
```

```
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 ...
```

This table shows that there is a 10 pig sample for every combination of dose and delivery method. The description of the database says that the 10 pigs are the same for every sample.

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

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

The following code groups the database by dose and delivery method and computes the mean of the 10 pig sample for every combination of parameters.

```
library(dplyr)
```

```
## 
## Attaching package: 'dplyr'
## 
## The following object is masked from 'package:stats':
## 
##     filter
## 
## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union
```

```
data <- group_by(data,supp,dose)
sum <- summarize(data,mean(len))
sum
```

```
## Source: local data frame [6 x 3]
## Groups: supp
## 
##   supp dose mean(len)
## 1   OJ  0.5     13.23
```

```
## 2   OJ  1.0     22.70
## 3   OJ  2.0     26.06
## 4   VC  0.5      7.98
## 5   VC  1.0     16.77
## 6   VC  2.0     26.14
```

3 Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

The sample of 10 pigs is the same for each combination of dose and delivery methods. So we will use paired T-test to compare the effect of dose and delivery methods in the lenght of teeth.

The first T-test (t1) tests the diference between the sample with a 1mg dose and the sample with a 0.5mg dose.

```
t1 = t.test(data$len[data$dose == 1],data$len[data$dose == 0.5],paired = TRUE)
```

The second T-test (t2) tests the diference between the sample with a 2mg dose and the sample with a 1mg dose.

```
t2 = t.test(data$len[data$dose == 2],data$len[data$dose == 1],paired = TRUE)
```

The third T-test (t3) tests the diference between the sample with a 0.5mg dose delivered with Orange Juice and the sample with a 0.5mg dose delivered with ascorbic acid.

```
t3 = t.test(data$len[data$supp == "OJ" & data$dose == 0.5],data$len[data$supp == "VC" & data$dose == 0.
```

The fourth T-test (t4) tests the diference between the sample with a 1mg dose delivered with Orange Juice and the sample with a 1mg dose delivered with ascorbic acid.

```
t4 = t.test(data$len[data$supp == "OJ" & data$dose == 1],data$len[data$supp == "VC" & data$dose == 1],pa
```

The last T-test (t5) tests the diference between the sample with a 2mg dose delivered with Orange Juice and the sample with a 2mg dose delivered with ascorbic acid.

```
t5 = t.test(data$len[data$supp == "OJ" & data$dose == 2],data$len[data$supp == "VC" & data$dose == 2],pa
```

4 State your conclusions and the assumptions needed for your conclusions.

A Test t1

I assume that the difference between the sample with a 1mg dose and the sample with a 0.5mg dose is equal to 0.

```
t1
```

```
##
##  Paired t-test
##
## data:  data$len[data$dose == 1] and data$len[data$dose == 0.5]
```

```
## t = 6.967, df = 19, p-value = 1.225e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.387 11.873
## sample estimates:
## mean of the differences
##                    9.13
```

We can reject this assumption. The confidence interval is entirely above zero so the 1mg dose appears to be more effective than the 0.5mg dose.

B Test t2

I assume that the difference between the sample with a 2mg dose and the sample with a 1mg dose is equal to 0.

```
t2
```

```
##
##  Paired t-test
##
## data:  data$len[data$dose == 2] and data$len[data$dose == 1]
## t = 4.605, df = 19, p-value = 0.0001934
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.472 9.258
## sample estimates:
## mean of the differences
##                   6.365
```

We can reject this assumption. The confidence interval is entirely above zero so the 2mg dose appears to be more effective than the 1mg dose.

C Test t3

I assume that the difference between the sample with a 0.5mg dose delivered with Orange Juice and the sample with a 0.5mg dose delivered with ascorbic acid is equal to 0.

```
t3
```

```
##
##  Paired t-test
##
## data:  data$len[data$supp == "OJ" & data$dose == 0.5] and data$len[data$supp == "VC" & data$dose == 0
## t = 2.979, df = 9, p-value = 0.01547
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.263 9.237
## sample estimates:
## mean of the differences
##                    5.25
```

We can reject this assumption. The confidence interval is entirely above zero so for a 0.5mg dose the Orange Juice delivery method appears to be more effective than the Ascorbic Acid delivery method.

D Test t4

I assume that the difference between the sample with a 1mg dose delivered with Orange Juice and the sample with a 1mg dose delivered with ascorbic acid is equal to 0.

```
t4
```

```
##
##  Paired t-test
##
## data:  data$len[data$supp == "OJ" & data$dose == 1] and data$len[data$supp == "VC" & data$dose == 1]
## t = 3.372, df = 9, p-value = 0.008229
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.952 9.908
## sample estimates:
## mean of the differences
##                    5.93
```

We can reject this assumption. The confidence interval is entirely above zero so for a 1mg dose the Orange Juice delivery method appears to be more effective than the Ascorbic Acid delivery method.

E Test t5

I assume that the difference between the sample with a 2mg dose delivered with Orange Juice and the sample with a 2mg dose delivered with ascorbic acid is equal to 0.

```
t5
```

```
##
##  Paired t-test
##
## data:  data$len[data$supp == "OJ" & data$dose == 2] and data$len[data$supp == "VC" & data$dose == 2]
## t = -0.0426, df = 9, p-value = 0.967
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.329  4.169
## sample estimates:
## mean of the differences
##                   -0.08
```

We can't reject this assumption. There is no evidence that a delivery method is better than the other for a 2mg dose.