Statistical Inference - Project B - Claudio Bogazzi

In this 2nd project we are going to analyze the ToothGrowth data in the R datasets package. This data comes from an essay testing the effect of Vitamin C on toothgrowth in Guinea Pigs. It consists of the measured tooth length in 6 groups of 10 animals, divided into two groups depending on the delivery method of the Vitamin C (OJ or VJ): Each of the two groups is also divided into 3 subgroups according to the dosage, 0.5, 1 or 2 mg per dose

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

```
library(datasets)
data(ToothGrowth)
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
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
```

2. Provide a basic summary of the data. We immediately provide the summary of the data set using the R function summary().

```
summary(ToothGrowth)
```

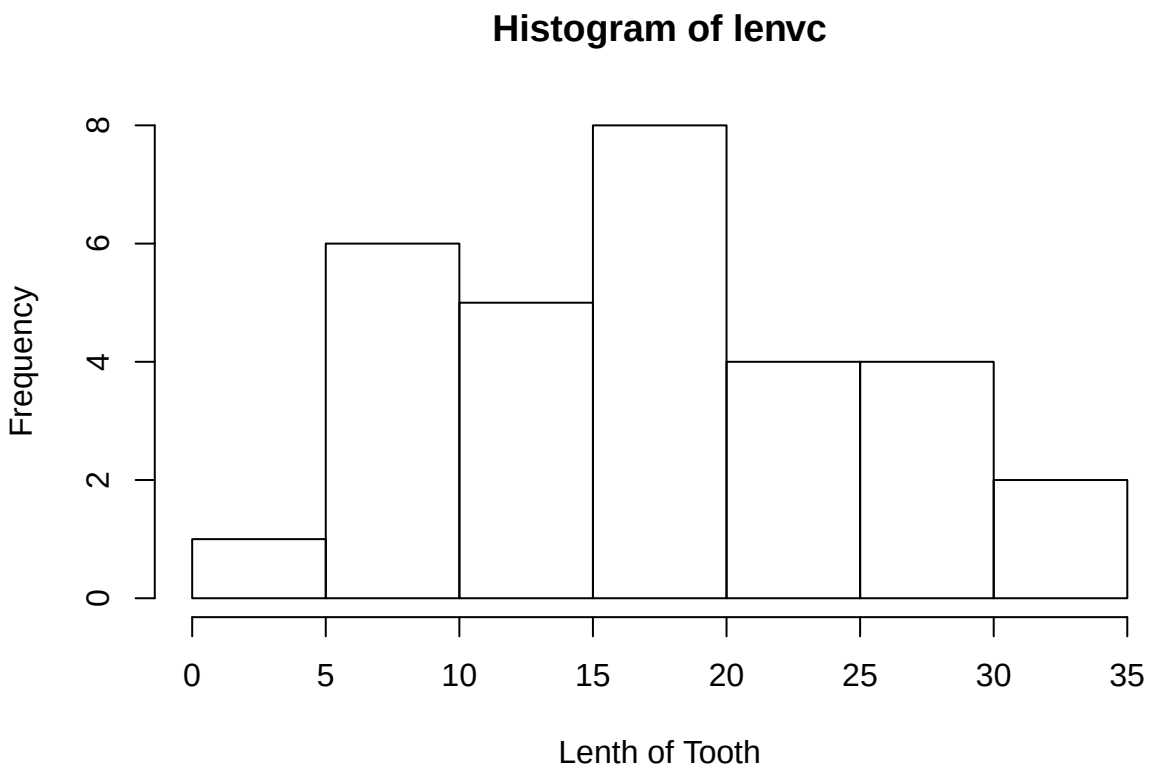
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

We now make two simple histograms showing the distribution length of the tooth for the two different delivery methods.

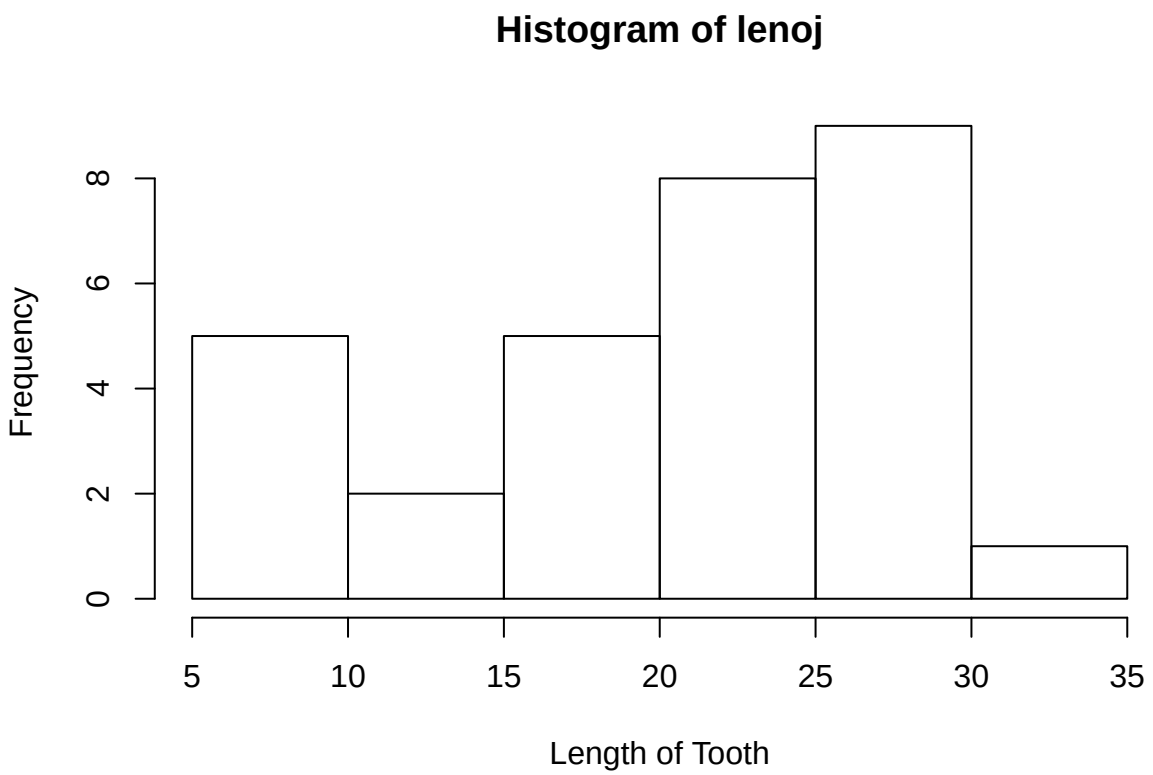
```
#a couple of simple histograms

lenvc <- ToothGrowth$len[which(ToothGrowth$supp == "VC")]
lenoj <- ToothGrowth$len[which(ToothGrowth$supp == "OJ")]

h <- hist(lenvc,xlab="Lenth of Tooth")
```




```
hh <- hist(lenoj, xlab ="Length of Tooth")
```



The next table shows the dose levels (rows) for the different delivery methods (columns)

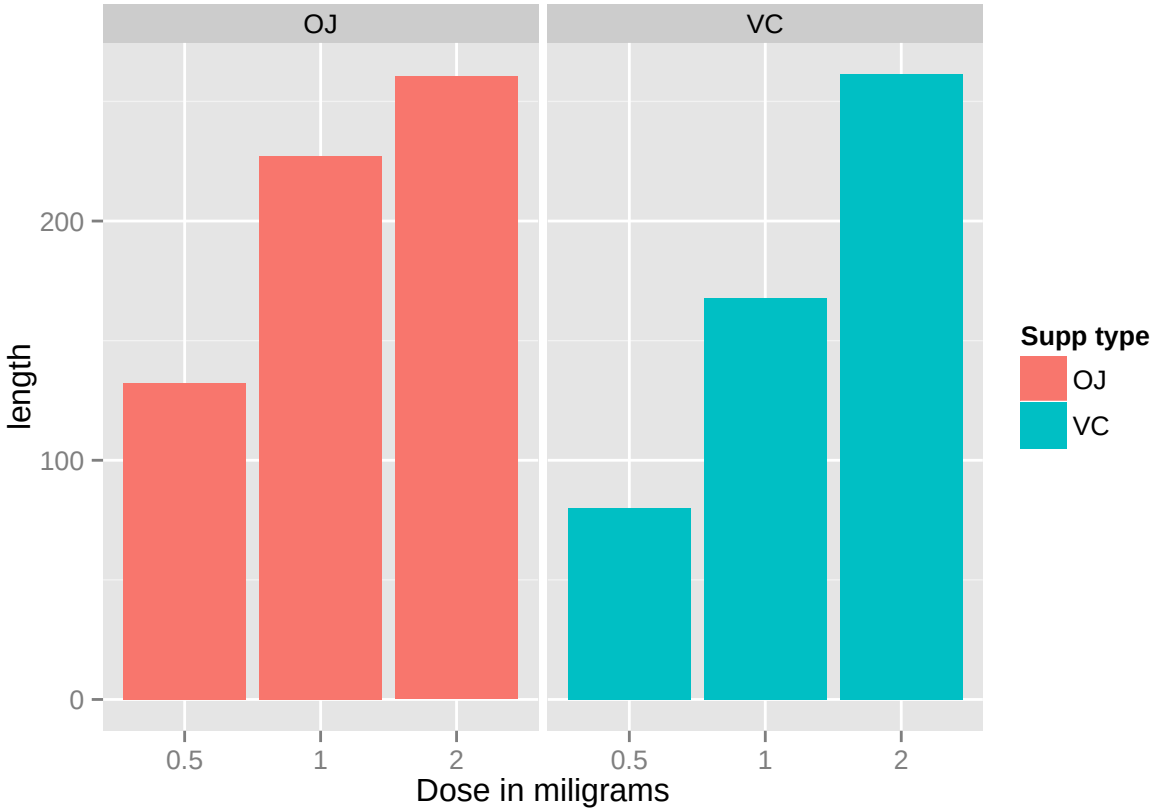
```r
# table to show dose levels and delivery methods
table(ToothGrowth$dose, ToothGrowth$supp)
```

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

Finally, we plot the dose levels as a function of the length of the tooth for the two methods.

```r
#dose vs tooth lenth

library(ggplot2)
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
    geom_bar(stat="identity",) +
    facet_grid(. ~ supp) +
    xlab("Dose in miligrams") +
    ylab("length") +
    guides(fill=guide_legend(title="Supp type"))
```



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

We now compute the different confidence intervals by using the t.test() function in R. First we consider the case where the dose is equal to 0.5. Let's see if there is a statistically significant difference in the length of tooth associated with the two different delivery methods.

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

data1 <- len[which(supp=="OJ" & dose==0.5)]
data2 <- len[which(supp=="VC" & dose==0.5)]

t.test(data1,data2)
```

```
##
##  Welch Two Sample t-test
##
## data:  data1 and data2
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

Indeed there is a difference. Moreover, since both limits of the confidence intervals are positive we can say that the mean of the length when the delivery method is "OJ" is greater than the alternative.

Let's now have a look at the case where the dose is equal to 1.0 and 2.0.

```
data3 <- len[which(supp=="OJ" & dose==1.0)]
data4 <- len[which(supp=="VC" & dose==1.0)]

t.test(data3,data4)
```

```
##
##  Welch Two Sample t-test
##
## data:  data3 and data4
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

```
data5 <- len[which(supp=="OJ" & dose==2.0)]
data6 <- len[which(supp=="VC" & dose==2.0)]

t.test(data5,data6)
```

```
##
##  Welch Two Sample t-test
##
## data:  data5 and data6
```

```
## t = -0.0461, 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
```

The results obtained for the case dose = 1.0 are similar to dose = 0.5. A different result is obtained for dose = 2.0 where we can see that the means for the two delivery methods are similar. Indeed, the confidence interval contains 0.

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

The mean value of the length of the tooth when the delivery method "OJ" is is greater than when the delivery method is "VC" for both doses of 0.5 and 1.0.

When the dose is 2.0 there are no evidence that the means of the length of the tooth for the two delivery methods is different.