Statistical Inference, Assigment 2

Basic inferential data analysis

Introduction and summary of the data

Now for the second part of the assigment, we're going to analyze the ToothGrowth data in the R datasets package.

First of all we load the data from the datasets package and the an overview of the content.

```
    library(datasets)
    library(ggplot2)
    library(gridExtra)
    data("ToothGrowth")
    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 ...
```

We can see that the data consists in 3 variables (len, supp and dose) and 60 rows or obserbations. Dose is a numeric vector but it only has 3 possible values, so for the purposse of this analysis we are going to convert it to a Factor. Len is a numeric vector, and we can see some interesting data in the summary:

```
    ToothGrowth$dose <- as.factor(ToothGrowth$dose)
    summary(ToothGrowth$len)
```

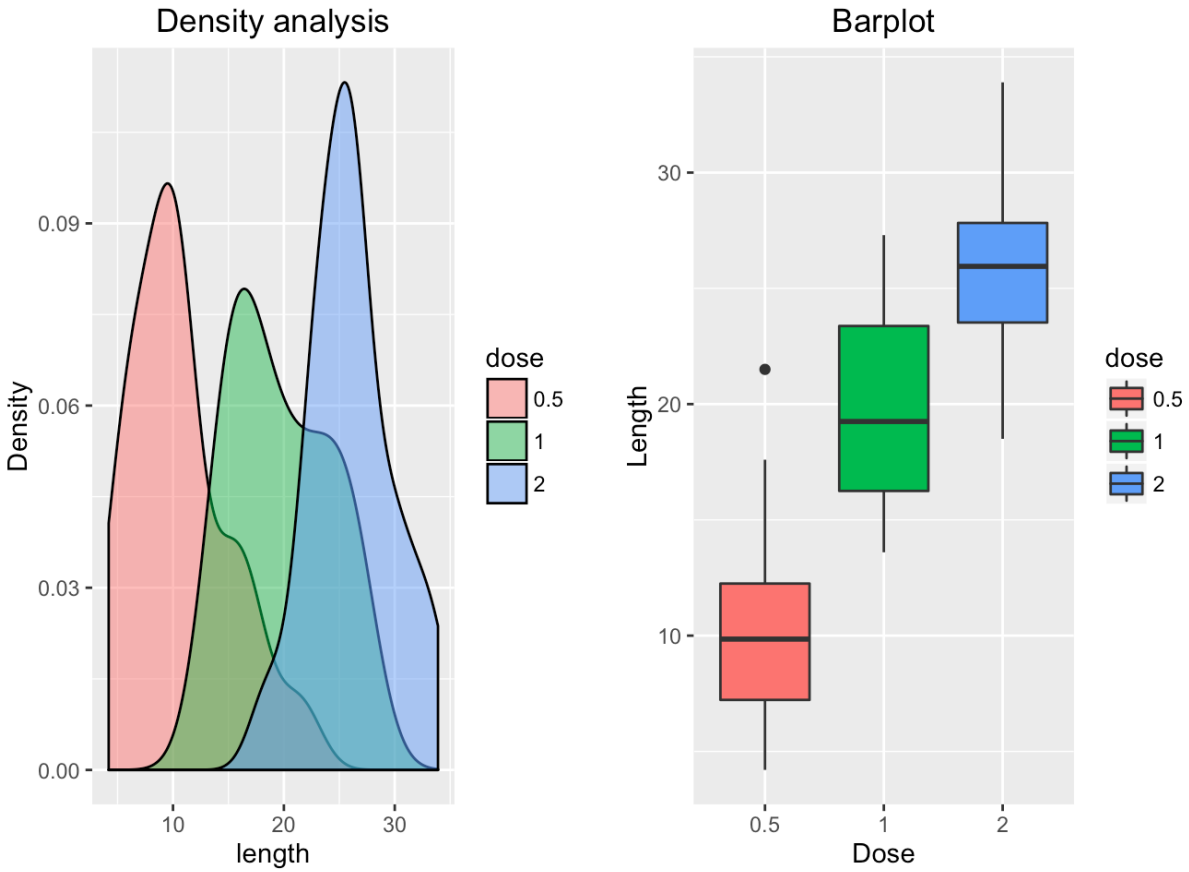
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20   13.08   19.25   18.81   25.28   33.90
```

Basic Exploratory Analysis

Len and Dose

We can see some plots to see the relation between Len and and Dose:

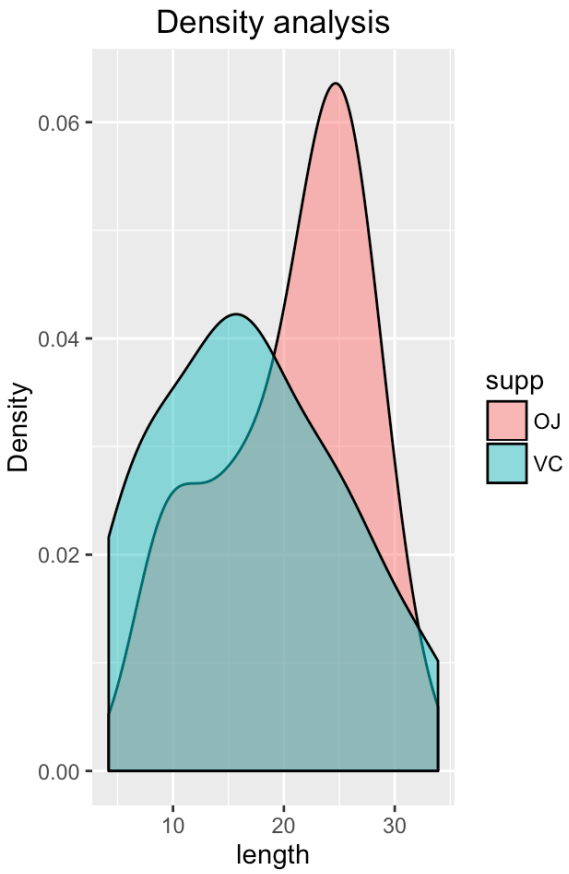
```
    plot1 <- qplot(len, data=ToothGrowth, geom="density", fill=dose, alpha=I
(.5),
            main="Density analysis", xlab="length",
            ylab="Density")
    plot2 <- qplot(dose, len, data=ToothGrowth, geom=c("boxplot"),
            fill=dose, main="Barplot",
            xlab="Dose", ylab="Length")
    grid.arrange(plot1, plot2, ncol=2)
```

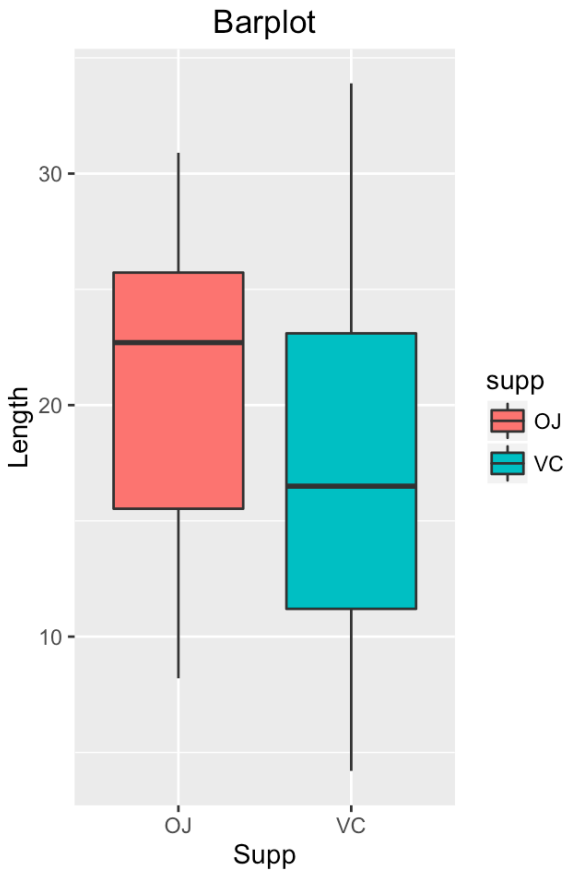


Len and supp

We can see some plots to see the relation between Len and and supp:

```
    plot1 <- qplot(len, data=ToothGrowth, geom="density", fill=supp, alpha=I
(.5),
            main="Density analysis", xlab="length",
            ylab="Density")
    plot2 <- qplot(supp, len, data=ToothGrowth, geom=c("boxplot"),
            fill=supp, main="Barplot",
            xlab="Supp", ylab="Length")
    grid.arrange(plot1, plot2, ncol=2)
```





Hypothesis tests to compare tooth growth by supp and dose

Hypothesis of teeth growth by supplement

To compare tooth growth by supp, we use a two sided t-test on OJ and VC. The null hypothesis (H0) is that there is no difference in Tooth Growth when using OJ or VC. The alternate hypothesis (H1) is that they are different.

So, first we subset the data and the use the t.test function.

```
len_OJ <- ToothGrowth$len[ToothGrowth$supp=="OJ"]
len_VC <- ToothGrowth$len[ToothGrowth$supp=="VC"]
t.test(len_OJ, len_VC)
```

```
##
##  Welch Two Sample t-test
##
## data:  len_OJ and len_VC
## 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:
##  -0.1710156  7.5710156
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

The P value of the test is 0.06 and the 95% confidence interval -0.171 7.571 contains 0 so we fail to reject the null hypothesis. The conclussion is that we can not say that there is an statistical difference between the two groups.

Hypothesis of teeth growth by dose

Now we repeat the process to compare the tooth growth and the dose. First we are going to compare the “0.5” and the “1” values.

```
    len_05 <- ToothGrowth$len[ToothGrowth$dose==0.5]
    len_1 <- ToothGrowth$len[ToothGrowth$dose==1]
    t.test(len_1, len_05)
```

```
##
##  Welch Two Sample t-test
##
## data:  len_1 and len_05
## 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:
##   6.276219 11.983781
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

In this case, we see that the p-value is very small 1.268e-07 and that the 95% confidence interval 6.276219 11.983781 does not contain 0 so we reject the null hypothesis H0 that both means are equal. In this case we can say that is a clear difference.

Finally, we compare the dose “1” values and the “2”.

```
    len_2 <- ToothGrowth$len[ToothGrowth$dose==2]
    t.test(len_2, len_1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len_2 and len_1
## t = 4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.733519 8.996481
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

Again, we see a very small p-value 1.906e-05 and the 95% confidence interval 3.733519 8.996481 does not contain 0 so we reject the null hypothesis that both means are the same.

Conclusions and Assumptions

The data assumed to be independent and normaly distributed. Given a small sample of data, we used a T-Distribution to do the tests.

There is no statistical evidence that proves a difference in tooth growth between sipplements “OJ” and “VC”.

The statistical evidence proves with a confidence of 95% that a dose of 1mg provides more growth that a 0.5 mg. Also a dose of 2mg provides more growth than a 1mg dose.