Statistical Inference - Tooth Growth Data

Markus Divjak

Overview

The following report analysis the tooth growth data that is delivered as part of the datasets package of R. The content of the dataset is the length of teeth of 60 guinea pigs who have received a supplement. The supplement was received in two different types with three different dosages each. More information about the dataset can be found under https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html. The following document shows some exploratory data analysis followed by a test on whether the tooth length can be related to the supplement type or dosage.

Tooth Growth Data in R

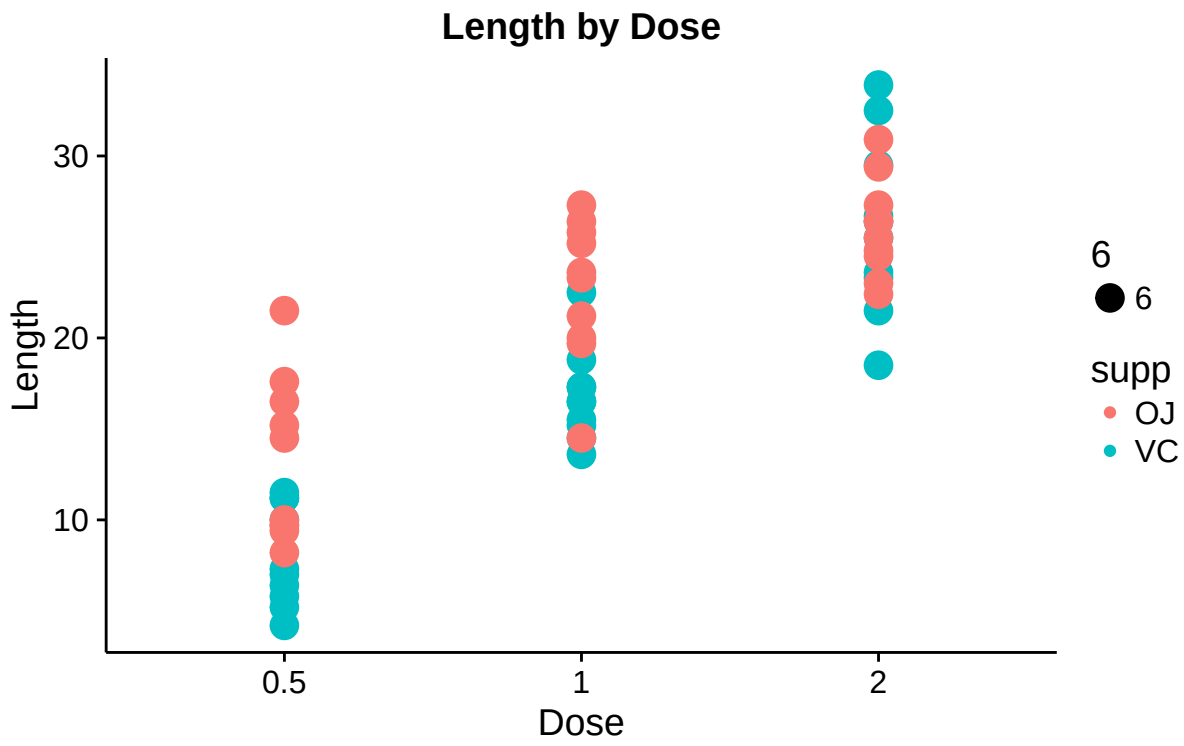
After loading the tooth growth data we run some basic exploratory analysis.

```
data("ToothGrowth")
ToothGrowth$dose <- factor(ToothGrowth$dose)
summary(ToothGrowth)
```

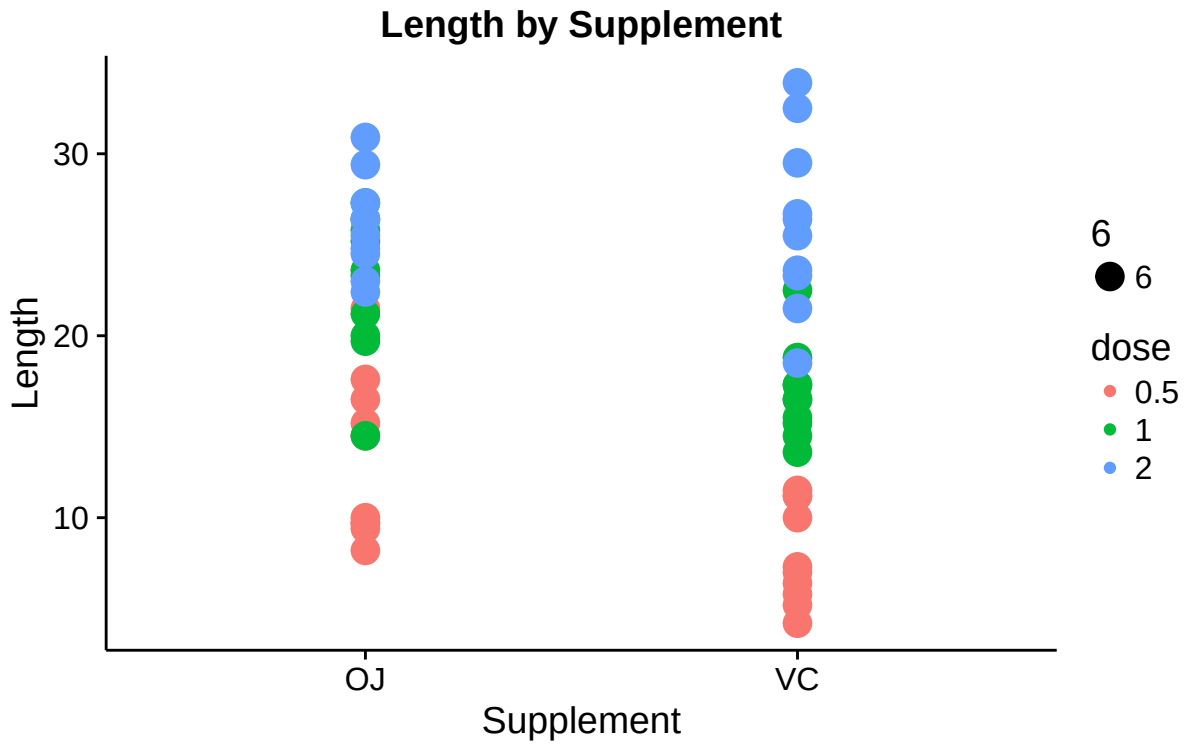
```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

What we can see from the summary of the data is that the mean of the tooth length is 19, that there are two different supplements and that the there are 3 discrete levels of doses: 0.5, 1.0 and 2.0. Below charts show the tooth size when compared to the the supplements and the different levels of dose.

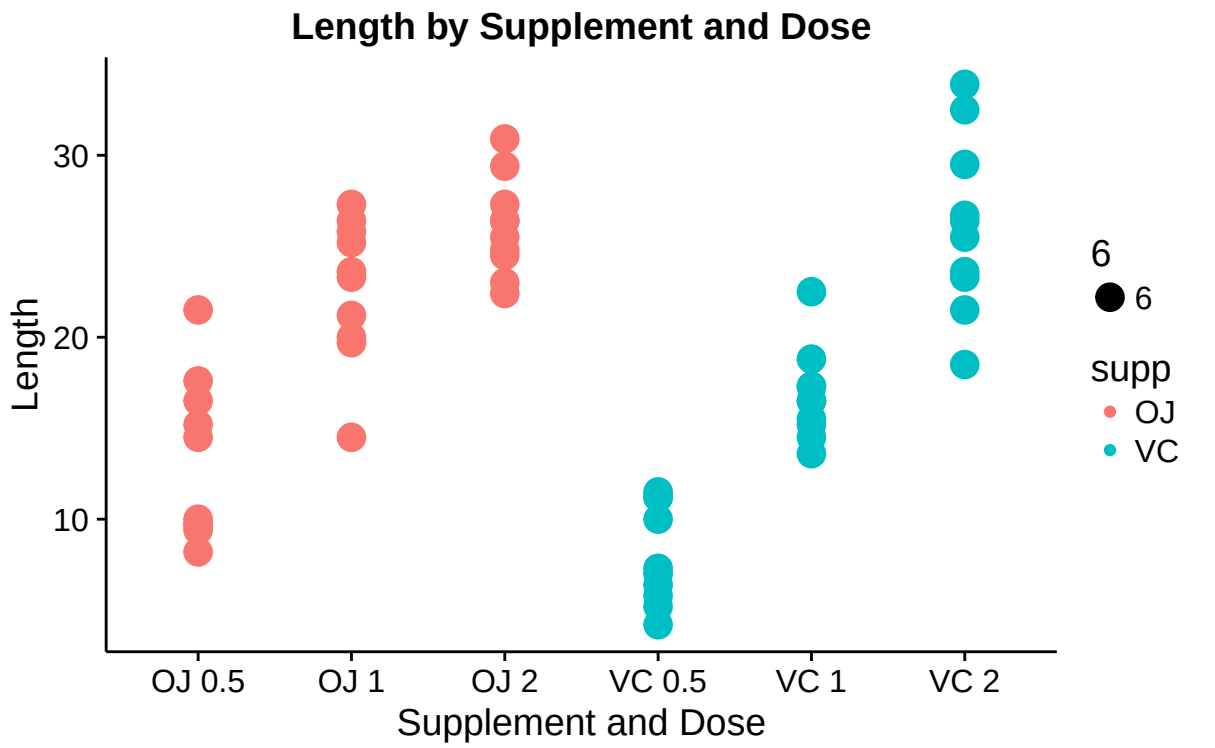
```
ToothGrowth <- mutate(ToothGrowth,combine=paste(supp,dose,sep=" "))
p1 <- qplot(dose, len, data=ToothGrowth, color=supp, size= 6,xlab="Dose",ylab="Length",main="Length by
p2 <- qplot(supp,len, data=ToothGrowth, color=dose, size=6,xlab="Supplement", ylab="Length", main="Leng
p3 <- qplot(combine,len,data=ToothGrowth, color=supp, size=6, ylab="Length", xlab="Supplement and Dose"
p4 <- ggplot(ToothGrowth, aes(x=len, fill=combine)) + geom_histogram(colour="black", binwidth=5) + face
p1
```


Length by Dose
Length
30
20
10
0.5
1
2
Dose
6
6
supp
OJ
VC

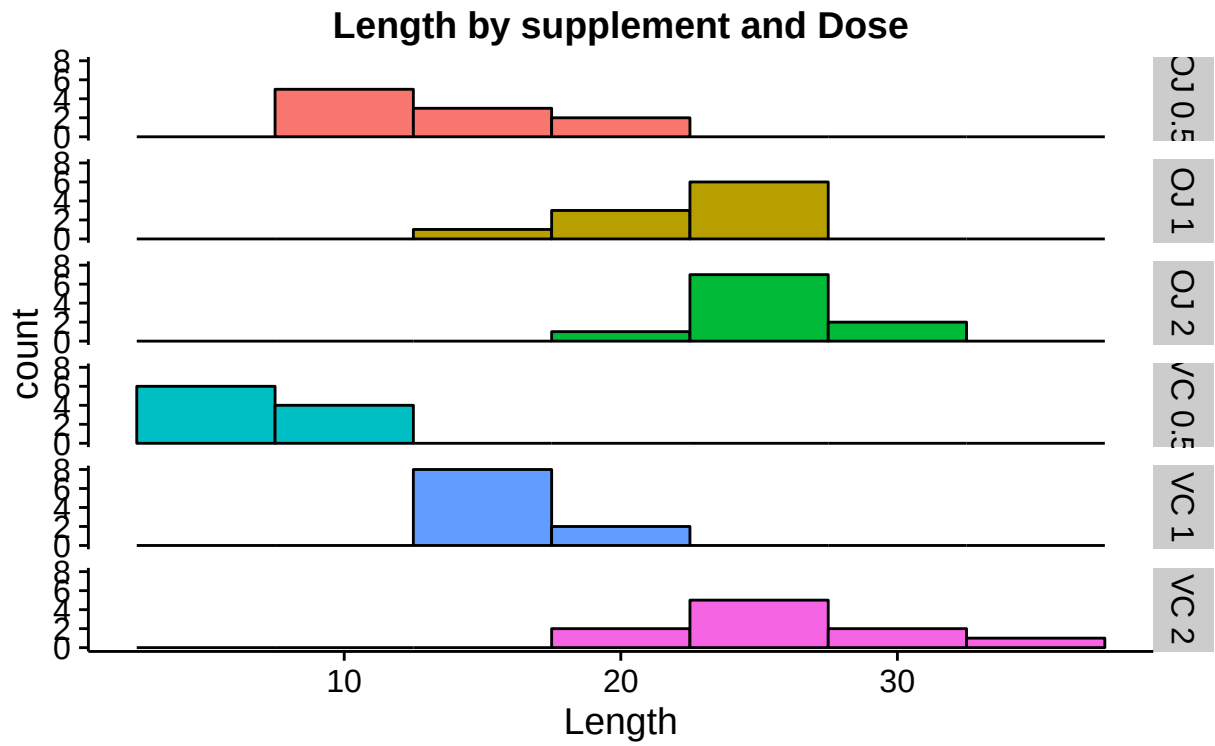
```
p2
```


Length by Supplement
Length
30
20
10
OJ
VC
Supplement
6
6
dose
0.5
1
2

```
p3
```


Length by Supplement and Dose
Length
30
20
10
OJ 0.5
OJ 1
OJ 2
VC 0.5
VC 1
VC 2
Supplement and Dose
6
6
supp
OJ
VC

```
p4
```


Length by supplement and Dose
count
Length
10
20
30
OJ 0.5
OJ 1
OJ 2
VC 0.5
VC 1
VC 2

The plots seem to suggest that there is a relationship between the dose and the length while there seems to be no (or a much weaker) relationship between the supplement and the length. Therefore we will test the relationship between length and dose in the following chapter.

Effects of Dose on Length

To investigate the effect of dose on length we run three t-tests comparing the different levels of dose to each other.

```
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,1))
t.test(len ~ dose,data=ToothGrowth_sub)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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 in group 0.5   mean in group 1
##            10.605            19.735
```

```
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1,2))
t.test(len ~ dose,data=ToothGrowth_sub)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

```
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,2))
t.test(len ~ dose,data=ToothGrowth_sub)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

As you can see the p-values of those three tests are all essentially zero and the confidence intervals do not contain 0. Therefore we can safely reject the null hypothesis and conclude that the dose has a positive effect on the length of a tooth.

For completeness sake we also run a t-test for the relationship between length and supplement

```
t.test(len ~ supp,data=ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##         20.66333         16.96333
```

The p-value of 0.06 or 6 % suggests that we should not reject the null hypothesis that the supplement does not have any influence on the length.

Conclusion

Under the following assumptions

1. The sample dataset is representative of the population
2. The distribution of the sample means follows the CLT We can conclude that the supplement ttype does not have significant influence on the length of tooth. However, the tooth length increases with an increase in dose.