

Analysis of the Exponential distribution using simulation

Chris Shaw

3 April 2016

Question

Remove this section in final report

Now in the second portion of the class, we're going to analyze the ToothGrowth data in the R datasets package.

Load the ToothGrowth data and perform some basic exploratory data analyses Provide a basic summary of the data. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering) State your conclusions and the assumptions needed for your conclusions. Some criteria that you will be evaluated on

Did you perform an exploratory data analysis of at least a single plot or table highlighting basic features of the data? Did the student perform some relevant confidence intervals and/or tests? Were the results of the tests and/or intervals interpreted in the context of the problem correctly? Did the student describe the assumptions needed for their conclusions?

Overview

In a few (2-3) sentences explain what is going to be reported on.

Exploratory Analysis

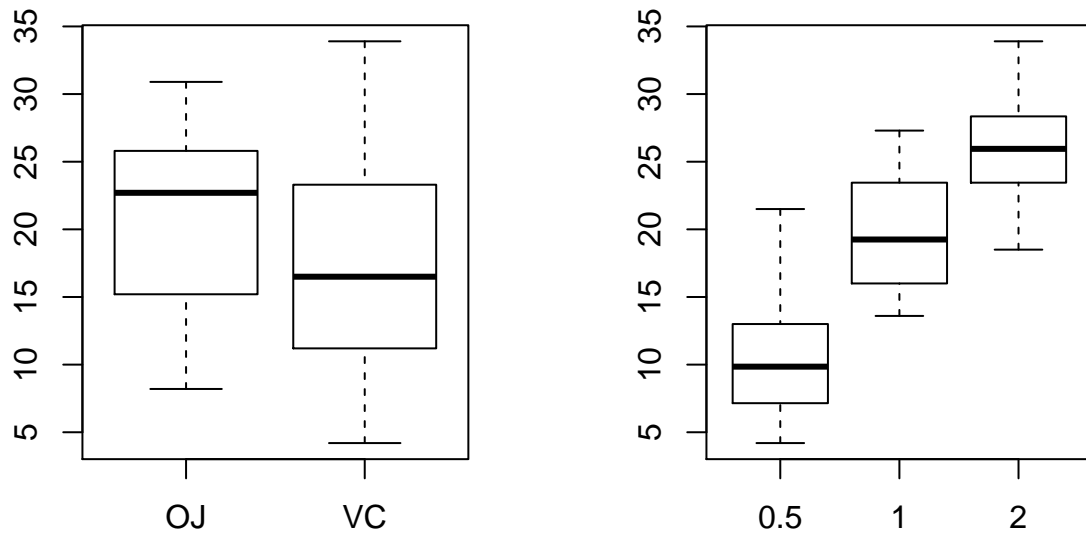
Include English explanations of the simulations you ran, with the accompanying R code. Your explanations should make clear what the R code accomplishes.

```
data(ToothGrowth)

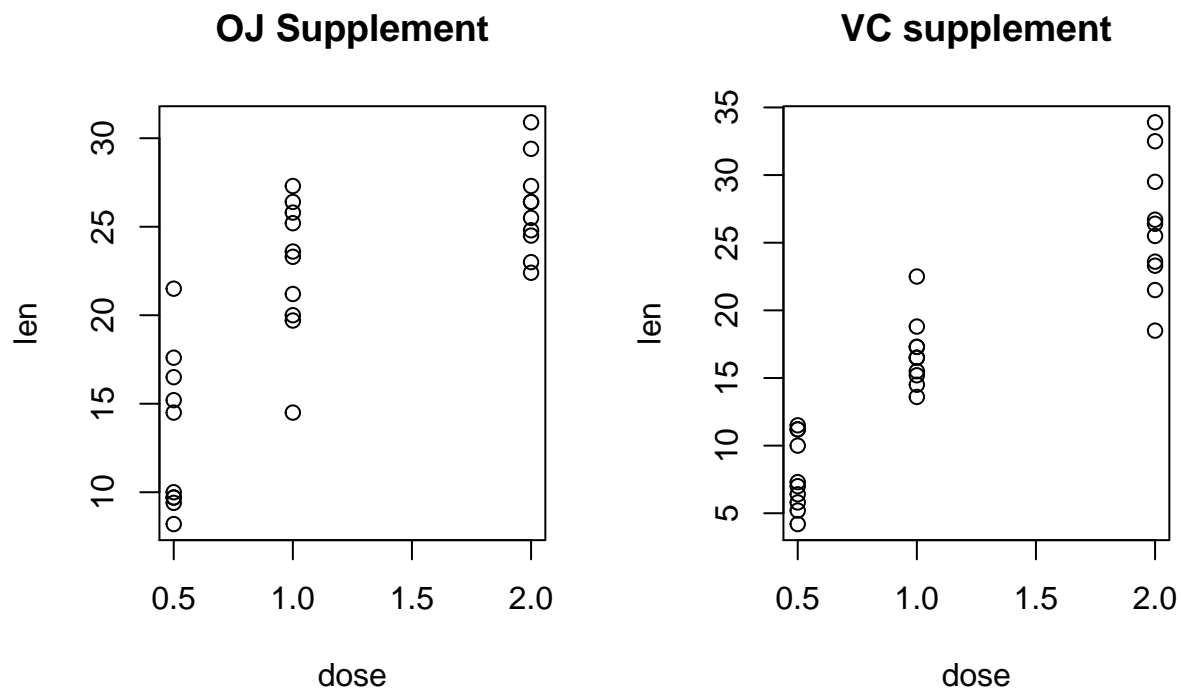
par(mfrow=c(1,2),oma = c(0, 0, 2, 0))

boxplot(len~supp, data=ToothGrowth)
boxplot(len~dose, data=ToothGrowth)
mtext("Tooth length by Supplement and Dose", outer = TRUE, cex = 1.5)
```

Tooth length by Supplement and Dose



```
plot(len~dose, data = ToothGrowth[ToothGrowth$supp=="OJ",], main="OJ Supplement")  
plot(len~dose, data = ToothGrowth[ToothGrowth$supp=="VC",], main="VC supplement")
```



```
#extract measurements for each of the supplement types and dosage levels
vc<-ToothGrowth[ToothGrowth$supp=="VC",]$len
oj<-ToothGrowth[ToothGrowth$supp=="OJ",]$len
d_05<-ToothGrowth[ToothGrowth$dose==0.5,]$len
d_1<-ToothGrowth[ToothGrowth$dose==1,]$len
d_2<-ToothGrowth[ToothGrowth$dose==2,]$len

# Calculate aggregated statistics on supp and dose (number of obs, mean and sd)
ag<-aggregate(len~supp+dose, ToothGrowth,
              function(x) c(n=as.integer(length(x)),
                             mn=round(mean(x),2),
                             sd=round(sd(x),2)))

# force the aggregated output into a printable dataframe and re-order
ag <- cbind(ag[,1:2], as.data.frame(unlist(ag$len)))
ag <- ag[with(ag, order(supp, dose)),]

kable(ag, col.names = c("Supplement", "Dose", "Num obs", "Mean", "Std dev"),
      row.names = FALSE,
      align=c("c", "c", "c", "c", "c"),
      caption = "Summary of tooth length by supplement and dosage")
```

Table 1: Summary of tooth length by supplement and dosage

Supplement	Dose	Num obs	Mean	Std dev
OJ	0.5	10	13.23	4.46
OJ	1.0	10	22.70	3.91
OJ	2.0	10	26.06	2.66
VC	0.5	10	7.98	2.75
VC	1.0	10	16.77	2.52
VC	2.0	10	26.14	4.80

Hypothesis testing

First we compare the different supplements, OJ and VC. The mean tooth length from the population using each of these is μ_{oj} and μ_{vc} respectively. The null hypothesis, H_0 is that these means are equal, i.e:

$$H_0 : \mu_{oj} = \mu_{vc}$$

$$H_a : \mu_{oj} \neq \mu_{vc}$$

We want to find evidence of whether the alternative hypothesis H_a may be true, that the means are not equal.

We run a two sided t-test and extract the p-value:

```
pv <- t.test(oj, vc, paired=FALSE, var.equal = TRUE, alternative = "two.sided")
```

The p-value is calculated at 0.0603934. This is very low, so we fail to reject the null hypothesis and conclude there is no evidence that the means are different.

```
t.test(oj, vc, paired=FALSE, var.equal = TRUE, alternative = "two.sided")
```

```
##
## Two Sample t-test
##
## data:  oj and vc
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1670064  7.5670064
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

```
t.test(d_2, d_05, paired=FALSE, var.equal = TRUE, alternative = "greater")
```

```
##
## Two Sample t-test
##
## data:  d_2 and d_05
## t = 11.799, df = 38, p-value = 1.419e-14
## alternative hypothesis: true difference in means is greater than 0
```

```
## 95 percent confidence interval:
## 13.28093      Inf
## sample estimates:
## mean of x mean of y
## 26.100      10.605
```

```
t.test(d_2, d_1, paired=FALSE, var.equal = TRUE, alternative = "greater")
```

```
##
## Two Sample t-test
##
## data: d_2 and d_1
## t = 4.9005, df = 38, p-value = 9.054e-06
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 4.175196      Inf
## sample estimates:
## mean of x mean of y
## 26.100      19.735
```

```
t.test(d_1, d_05, paired=FALSE, var.equal = TRUE, alternative = "greater")
```

```
##
## Two Sample t-test
##
## data: d_1 and d_05
## t = 6.4766, df = 38, p-value = 6.331e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 6.753344      Inf
## sample estimates:
## mean of x mean of y
## 19.735      10.605
```

Assumptions

Include figures (output from R) with titles. Highlight the variances you are comparing. Include text that explains your understanding of the differences of the variances.

Results

Conclusion

Via figures and text, explain how one can tell the distribution is approximately normal.