Coursera Statistical Inference Class Project - Part II - Tooth Growth Data Analysis

CRR

Saturday, June 20, 2015

OVERVIEW: In this part of the project, the ToothGrowth data in the R datasets package will be analyzed. This will include some basic exploratory data analysis and summarization of the data, as well as using confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. Conclusions will be drawn accordingly and the assumptions needed for the conclusions will be stated.

ANALYSIS: The analysis was done using the R code given in the Appendix. Comments within the code detail what is being done within each section.

RESULTS/CONCLUSIONS:

See Appendix below for figures/plots and values discussed here.

Exploration/Summarization. Exploring the basic structure of the data reveals that there are two supplements tested, three dosages are used for each, and tooth length is the final outcome determined. This indicates how to structure plots that might give further insights.

Scatter plots were done for the various doses of VC and OJ versus tooth length. They indicate greater variability in results for lower doses of OJ compared to VC, but the opposite for the highest dose; tooth length seems to increase with increased dosage of either supplement. Similar conclusions can be drawn from the corresponding box plots, though these may be easier to read.

Means/Confidence Intervals. The results of the calculation of means and their 95% confidence intervals (based on normal and t) indicate tooth length (growth) is much higher at the lower doses for OJ than for VC, but they are very close at the highest dosage tested; tooth length increases as dosage increases with either supplement. The confidence intervals for VC are wider than for OJ at the highest dosage, where the means are very close; this may indicate a break point of some sort and may warrant further testing at higher dosages to elucidate.

Assumptions. The confidence intervals used here are based on the assumption of iid variables and that the normal and t distributions are applicable as required for the respective interval calculations. Intertwined with this is the assumption of the applicability of the Central Limit Theorem.

[Please skip the rest of this page and go to the Appendix starting on page 2.

Please skip the rest of this page and go to the Appendix starting on page 2.

Please skip the rest of this page and go to the Appendix starting on page 2.

Please skip the rest of this page and go to the Appendix starting on page 2.

Please skip the rest of this page and go to the Appendix starting on page 2.

Please skip the rest of this page and go to the Appendix starting on page 2.

Please skip the rest of this page and go to the Appendix starting on page 2.

Please skip the rest of this page and go to the Appendix starting on page 2.

Please skip the rest of this page and go to the Appendix starting on page 2.

Please skip the rest of this page and go to the Appendix starting on page 2.

Please skip the rest of this page and go to the Appendix starting on page 2.

Please skip the rest of this page and go to the Appendix starting on page 2.]

APPENDIX: R code (with comments) and values calculated and figures produced by it.

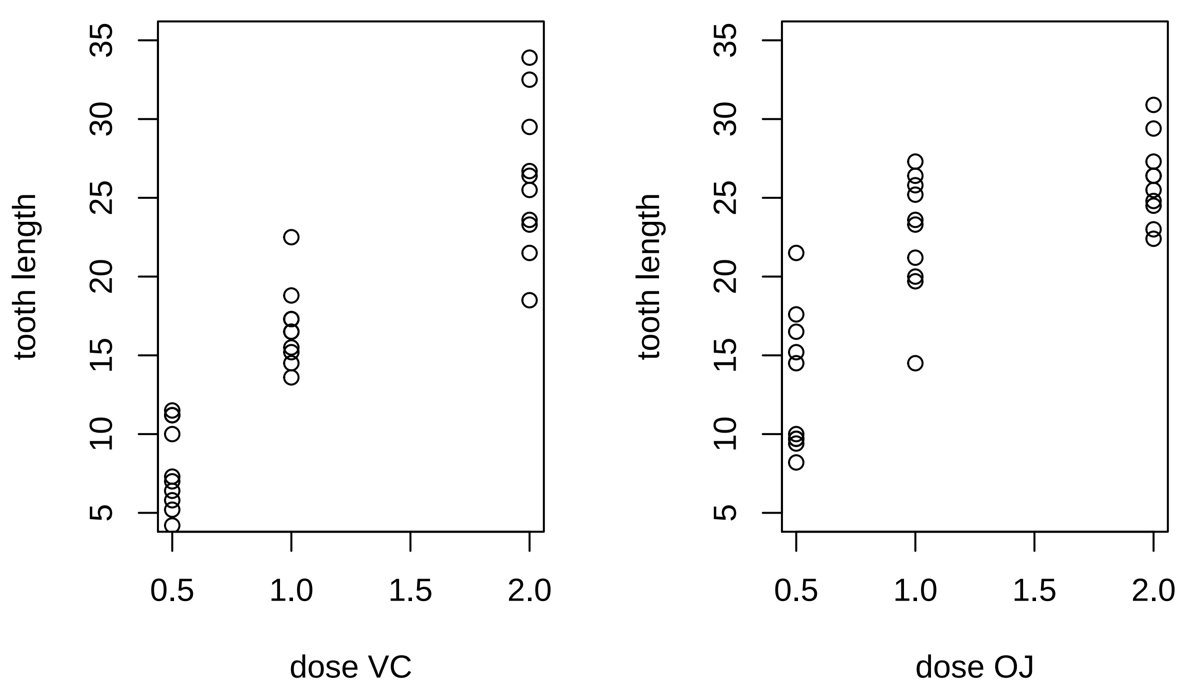
```r
## ToothGrowth Data Analysis
data(ToothGrowth)
## Exploring the basic structure of the data
str(ToothGrowth)
```

```
## '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 ...
```

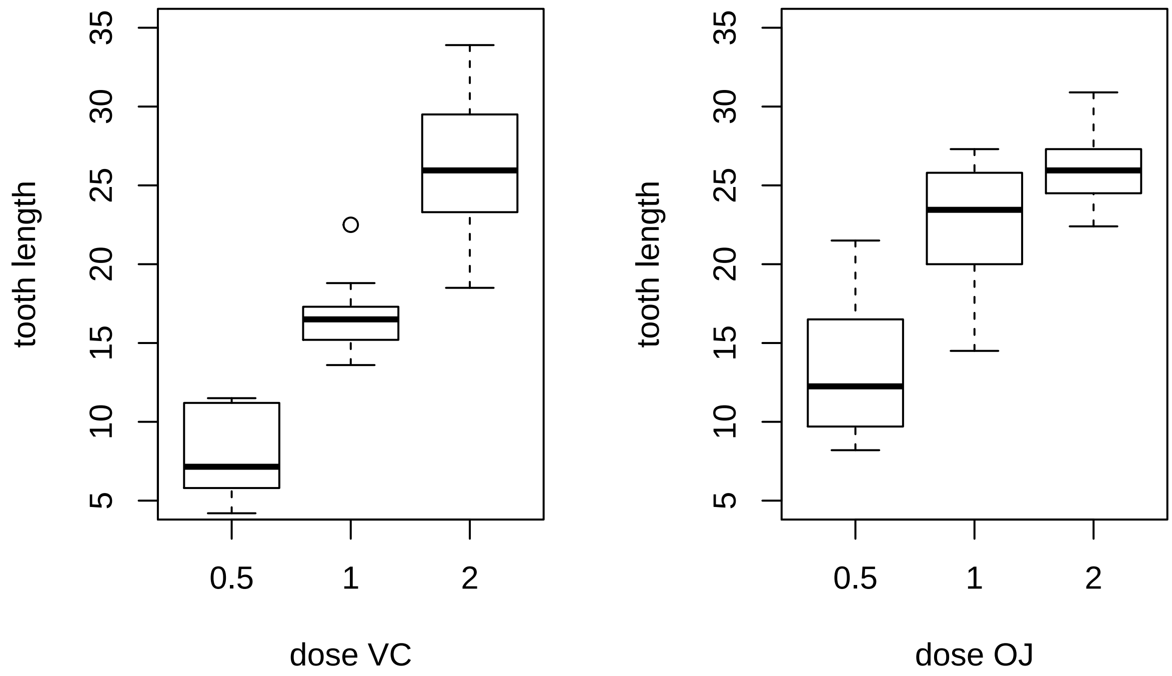
```r
table(ToothGrowth$dose)
```

```
##
## 0.5   1   2
##  20  20  20
```

```r
## Making some basic plots of the data - scatterplots and boxplots
## for further insights
par(mfrow=c(1,2))
with(subset(ToothGrowth, supp=="VC"), plot(dose, len, ylim=c(5,35), xlab="dose VC", ylab="tooth length")
with(subset(ToothGrowth, supp=="OJ"), plot(dose, len, ylim=c(5,35), xlab="dose OJ", ylab="tooth length")
```



```
par(mfrow=c(1,2))
with(subset(ToothGrowth, supp=="VC"), boxplot(len~dose, ylim=c(5,35), xlab="dose VC", ylab="tooth lengt
with(subset(ToothGrowth, supp=="OJ"), boxplot(len~dose, ylim=c(5,35), xlab="dose OJ", ylab="tooth lengt
```




```
## Subsetting the data based on previous findings
## vc and oj are the types of supplements tested
vc<-subset(ToothGrowth, supp=="VC")
oj<-subset(ToothGrowth, supp=="OJ")
vch<-subset(vc, dose==0.5)
vc1<-subset(vc, dose==1.0)
vc2<-subset(vc, dose==2.0)
ojh<-subset(oj, dose==0.5)
oj1<-subset(oj, dose==1.0)
oj2<-subset(oj, dose==2.0)
## Calculating mean tooth length for VC, and the corresponding 95% normal and 95% t confidence interval
tvc<-data.frame(dose=c(0.5,1.0,2.0), MeanLenVC=c(1:3), CNorm95L=c(1:3), CNorm95U=c(1:3),Ct95L=c(1:3), C
tvc[1,2]<-mean(vch$len)
tvc[1,3:4]<-mean(vch$len) + c(-1,1)*qnorm(0.975)*sd(vch$len)/sqrt(length(vch$len))
tvc[1,5:6]<-mean(vch$len) + c(-1,1)*qt(0.975,length(vch$len)-1)*sd(vch$len)/sqrt(length(vch$len))
tvc[2,2]<-mean(vc1$len)
tvc[2,3:4]<-mean(vc1$len) + c(-1,1)*qnorm(0.975)*sd(vc1$len)/sqrt(length(vc1$len))
tvc[2,5:6]<-mean(vc1$len) + c(-1,1)*qt(0.975,length(vc1$len)-1)*sd(vc1$len)/sqrt(length(vc1$len))
tvc[3,2]<-mean(vc2$len)
tvc[3,3:4]<-mean(vc2$len) + c(-1,1)*qnorm(0.975)*sd(vc2$len)/sqrt(length(vc2$len))
tvc[3,5:6]<-mean(vc2$len) + c(-1,1)*qt(0.975,length(vc2$len)-1)*sd(vc2$len)/sqrt(length(vc2$len))
```

```r
tvc
```

```
##   dose MeanLenVC CNorm95L CNorm95U     Ct95L     Ct95U
## 1  0.5      7.98  6.27765  9.68235  6.015176  9.944824
## 2  1.0     16.77 15.21102 18.32898 14.970657 18.569343
## 3  2.0     26.14 23.16639 29.11361 22.707910 29.572090
```

```r
## Calculating mean tooth length for OJ, and the corresponding 95% normal and 95% t confidence intervals
toj<-data.frame(dose=c(0.5,1.0,2.0), MeanLenOJ=c(1:3), CNorm95L=c(1:3), CNorm95U=c(1:3),Ct95L=c(1:3), Ct
toj[1,2]<-mean(ojh$len)
toj[1,3:4]<-mean(ojh$len) + c(-1,1)*qnorm(0.975)*sd(ojh$len)/sqrt(length(ojh$len))
toj[1,5:6]<-mean(ojh$len) + c(-1,1)*qt(0.975,length(ojh$len)-1)*sd(ojh$len)/sqrt(length(ojh$len))
toj[2,2]<-mean(oj1$len)
toj[2,3:4]<-mean(oj1$len) + c(-1,1)*qnorm(0.975)*sd(oj1$len)/sqrt(length(oj1$len))
toj[2,5:6]<-mean(oj1$len) + c(-1,1)*qt(0.975,length(oj1$len)-1)*sd(oj1$len)/sqrt(length(oj1$len))
toj[3,2]<-mean(oj2$len)
toj[3,3:4]<-mean(oj2$len) + c(-1,1)*qnorm(0.975)*sd(oj2$len)/sqrt(length(oj2$len))
toj[3,5:6]<-mean(oj2$len) + c(-1,1)*qt(0.975,length(oj2$len)-1)*sd(oj2$len)/sqrt(length(oj2$len))
toj
```

```
##   dose MeanLenOJ CNorm95L CNorm95U    Ct95L    Ct95U
## 1  0.5     13.23 10.46589 15.99411 10.03972 16.42028
## 2  1.0     22.70 20.27601 25.12399 19.90227 25.49773
## 3  2.0     26.06 24.41441 27.70559 24.16069 27.95931
```