Toothgrowth data analysis exercise

Statistical Inference Class Project part 2

Ben Garthwaite

27 September 2015

Synopsis

An analysis of the ToothGrowth data in the R datasets package.

Data Processing

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

```
require(datasets)
data(ToothGrowth)
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
```

```
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
```

```
tail(ToothGrowth)
```

```
##     len supp dose
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

Data Summary

Provide a basic summary of the data.

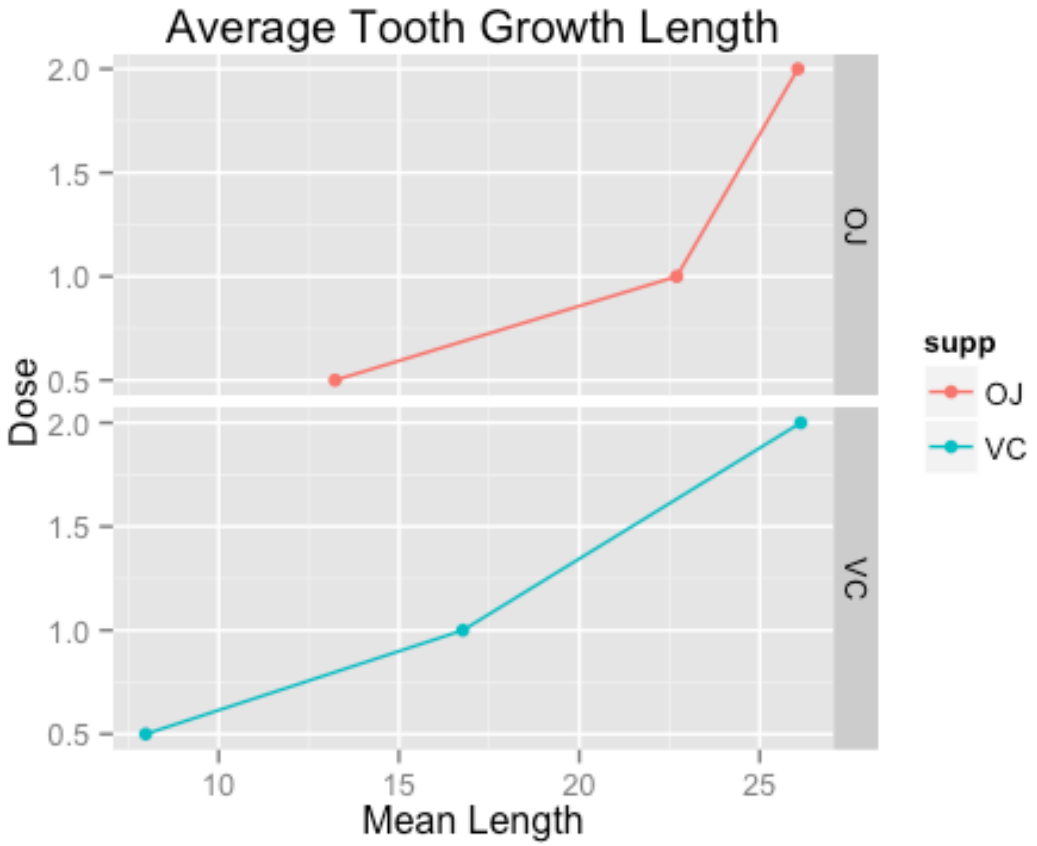
```
# Aggregate dataset
summary_mean <- setNames(
                aggregate(ToothGrowth$len,
by=list(supp=ToothGrowth$supp,dose=ToothGrowth$dose), FUN=mean)
                ,c("supp","dose","mean_len")
                )
summary_mean

##   supp dose mean_len
## 1   OJ  0.5    13.23
## 2   VC  0.5     7.98
## 3   OJ  1.0    22.70
## 4   VC  1.0    16.77
## 5   OJ  2.0    26.06
## 6   VC  2.0    26.14

summary_sd <- setNames(
                aggregate(ToothGrowth$len,
by=list(supp=ToothGrowth$supp,dose=ToothGrowth$dose), FUN=sd)
                ,c("supp","dose","sd_len")
                )
summary_sd

##   supp dose   sd_len
## 1   OJ  0.5 4.459709
## 2   VC  0.5 2.746634
## 3   OJ  1.0 3.910953
## 4   VC  1.0 2.515309
## 5   OJ  2.0 2.655058
## 6   VC  2.0 4.797731

# plot summaries
require(ggplot2)
g <- qplot(mean_len,dose,data = summary_mean,color = supp, facets =
supp ~ .)  + labs(title="Average Tooth Growth Length",x = "Mean
Length",y = "Dose")
g + geom_line()
```



Workings

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

```
# t interval for supp
rbind(
t.test(len ~ supp, paired = FALSE, var.equal = TRUE, data =
ToothGrowth)$conf, t.test(len ~ supp, paired = FALSE, var.equal =
FALSE, data = ToothGrowth)$conf
)
```

```
##             [,1]     [,2]
## [1,] -0.1670064 7.567006
## [2,] -0.1710156 7.571016
```

```
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, 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
```

Conclusions

The initial summary analysis chart shows that dose 2 has the greatest effect on tooth growth and that OJ promotes superior growth to VC.

The t interval testing shows the variances are not equal, looking at the test values the large confidence interval contains zero, so it is not possible to deny the null hypothesis.