Coursera Statistical Inference Project 2

Seth Dobrin

February 9, 2016

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.

```
library(knitr)
library(ggplot2)
```

knitr settings

```
knitr::opts_chunk$set(echo=TRUE, tidy = TRUE, message=FALSE, warnings=FALSE, fig.width=8, fig.height=5,
```

Part 1: Load 'ToothGrowth' data and perform some exploratory analysis

Load data and examine the strucure of ToothGrowth data frame

```
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 ...
```

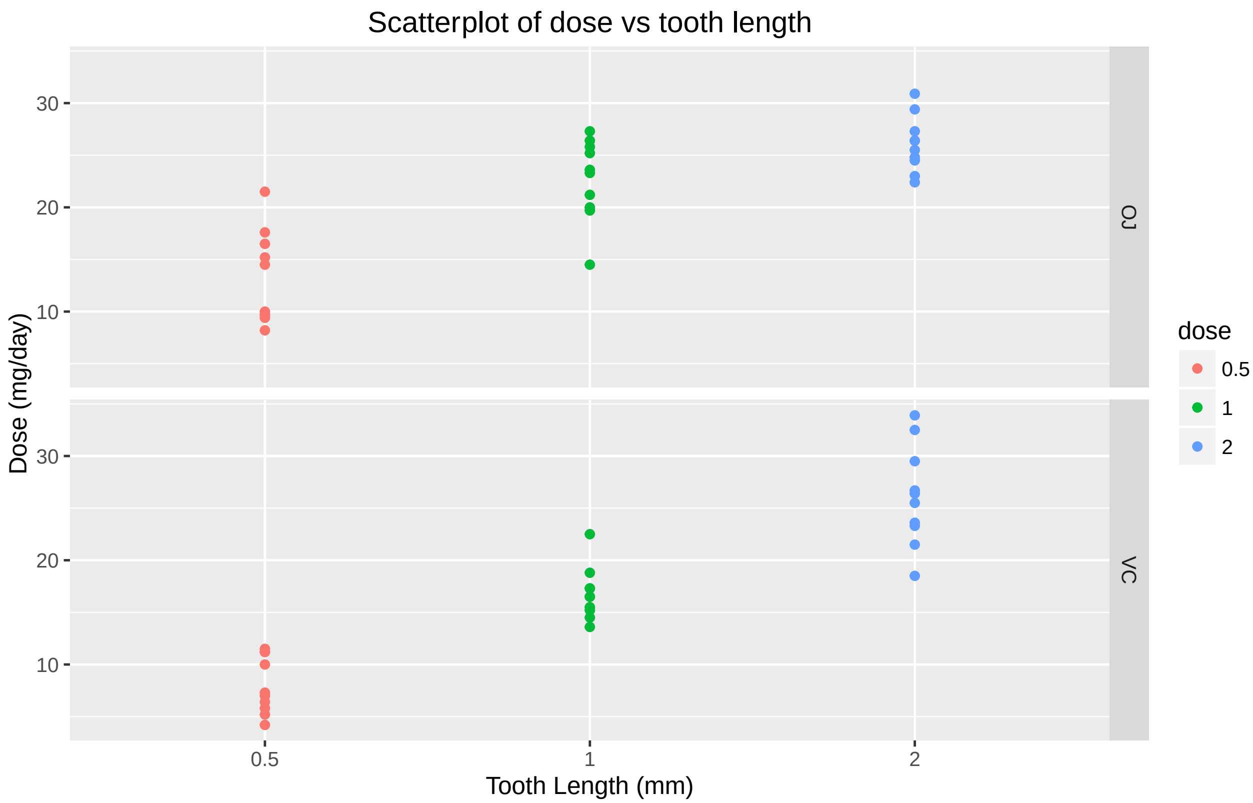
Convert dose from numeric to factor

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

Plot the data

```
qplot(dose, len, data = ToothGrowth, facets = supp ~ ., color = dose, xlab = "Tooth Length (mm)",
    ylab = "Dose (mg/day)", main = "Scatterplot of dose vs tooth length")
```



Part 2: Provide a basic summary of the data

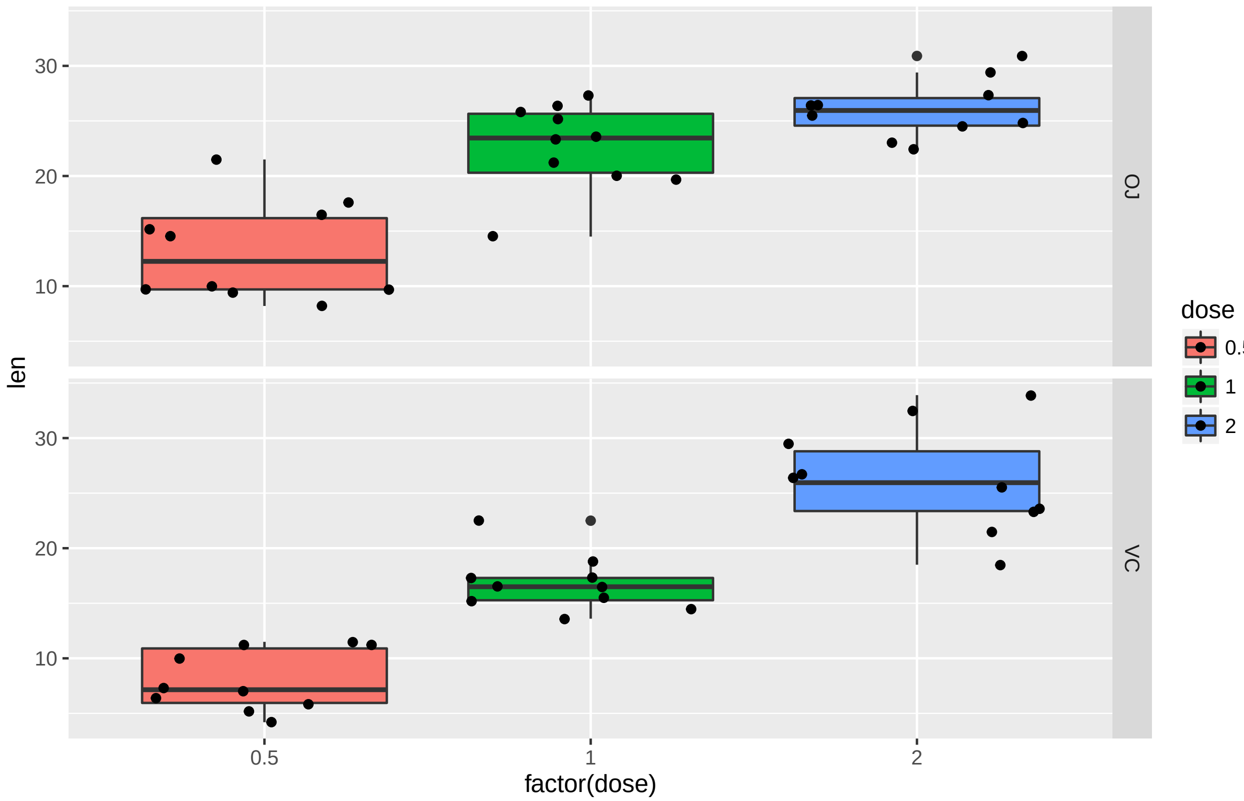
Summary statistics

```r
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
```

Graphically represent the summary statistics using a boxplot

```r
qplot(factor(dose), len, data = ToothGrowth, facets = supp ~ ., geom = c("boxplot",
    "jitter"), fill = dose)
```



Part3: Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

Using t.test, we check if there's difference in the performance of the treatments, looking if the p value>0.05 and if the confidence interval contains 0

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

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowth$supp == "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 this test is not less than 0.05. Based on this we cannot reject the null hypothesis and we canot assume that the form the vitamin is delivered effect tooth growth

```r
t.test(ToothGrowth$len[ToothGrowth$dose == 2], ToothGrowth$len[ToothGrowth$dose ==
    1])
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose == 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
```

Since the p-value is well below 0.05, it is highly likely that there is an effect on tooth length between 1 mg and 2 mg/day of vitamin C indepent of delivery method.

```r
t.test(ToothGrowth$len[ToothGrowth$dose == 1], ToothGrowth$len[ToothGrowth$dose ==
    0.5], var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 1] and ToothGrowth$len[ToothGrowth$dose == 0.5]
## t = 6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276252 11.983748
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

Again, since the p-value is well below 0.05, it is highly likely that there is an effect on tooth length between 0.5 mg and 1 mg/day of vitamin C indepent of delivery method.

```r
t.test(ToothGrowth$len[ToothGrowth$dose == 0.5], ToothGrowth$len[ToothGrowth$dose ==
    1], var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 0.5] and ToothGrowth$len[ToothGrowth$dose == 1]
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983748  -6.276252
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

Since the preivious two tests showed significance it is not surprising that there is a significant effect on tooth growth between 0.5 and 2 mg/day of vitamin C indepent of delivery method.

Part4: State your conclusions and the assumptions needed for your conclusions.

Our asumptions for this experiment are as follows:

1. Guniea pig populations have a common vairance (var.equal=TRUE)
2. Guniea pigs were randomly assigned a dose and treatment type allowing us to treat samples as independent.

Conclusion

Higher levels of vitamin C directly impact tooth growth independent of delivery method when comparing orange juice and ascorbic acid. This is demonstrated with statistical significance.