Statistical Inference Course Project - Part 2

Ken Wood

8/19/2020

Introduction

For the second portion of the course project, we're going to analyze the ToothGrowth data in the R datasets package. Specifically, we will:

1. Load the ToothGrowth data and perform some basic exploratory data analyses.
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (we will only use the techniques covered in the course, even if there are other approaches worth considering).
4. State your conclusions and the assumptions needed for your conclusions.

Load the data and perform some basic analyses...

```
tg <- ToothGrowth
str(tg)
```

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

```
names(tg) <- c("length","Supplement","dose")
unique(tg$dose)  # get list of unique values of 'dose'
```
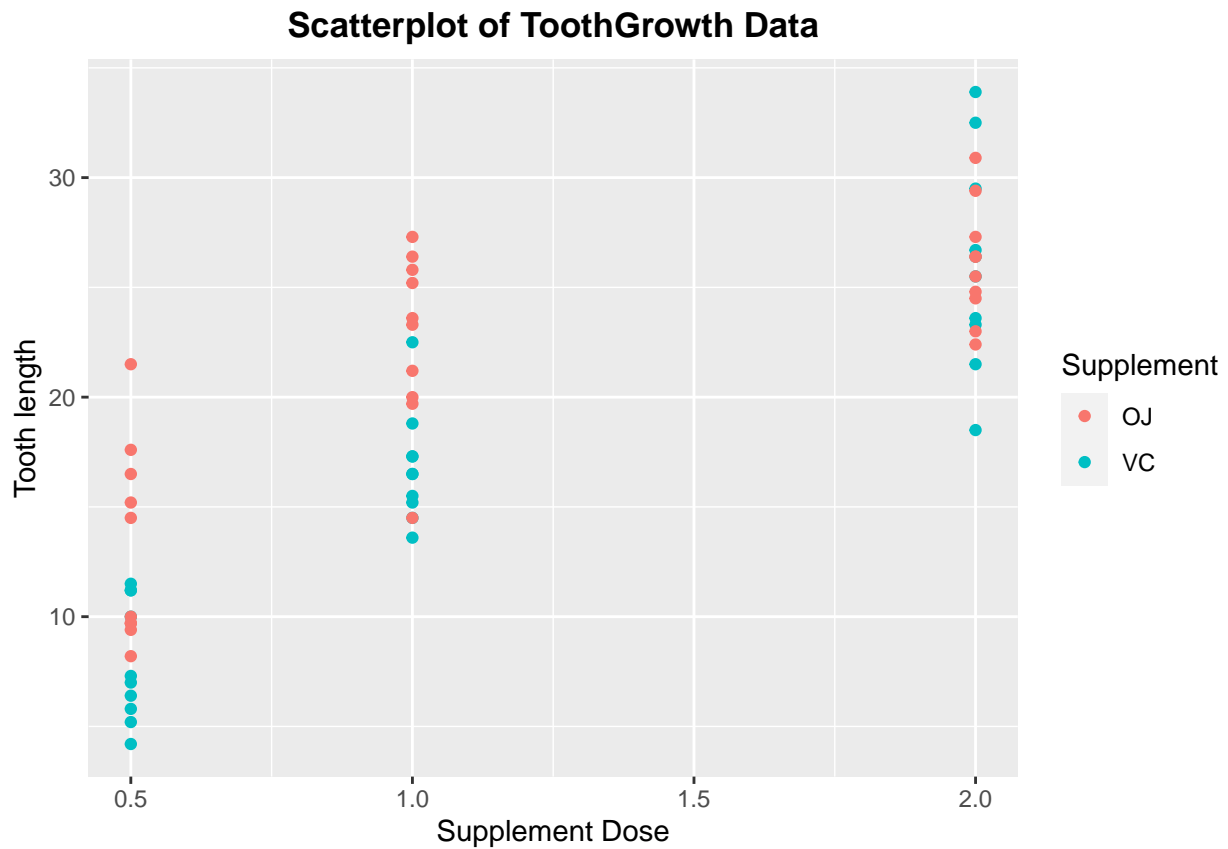
```
## [1] 0.5 1.0 2.0
```

Plots of the Data

**Create a scatter plot of the data in 'ToothGrowth'...**

```
library(ggplot2)

data_plot <- ggplot(aes(x=dose, y = length), data = tg) +
        geom_point(aes(color = Supplement)) + xlab("Supplement Dose") + ylab("Tooth length") + ggtitle(
print(data_plot)
```

Scatterplot of ToothGrowth Data

Create boxplots to show the relationships between the variables...

```
box_plt <- ggplot(aes(x = factor(dose), y = length), data = tg) +
        geom_boxplot(aes(fill = factor(dose))) + facet_wrap(~Supplement,ncol=2) + xlab("Dose") + ylab("
print(box_plt)
```

Tooth Length vs. Supplement Dose by Supplement
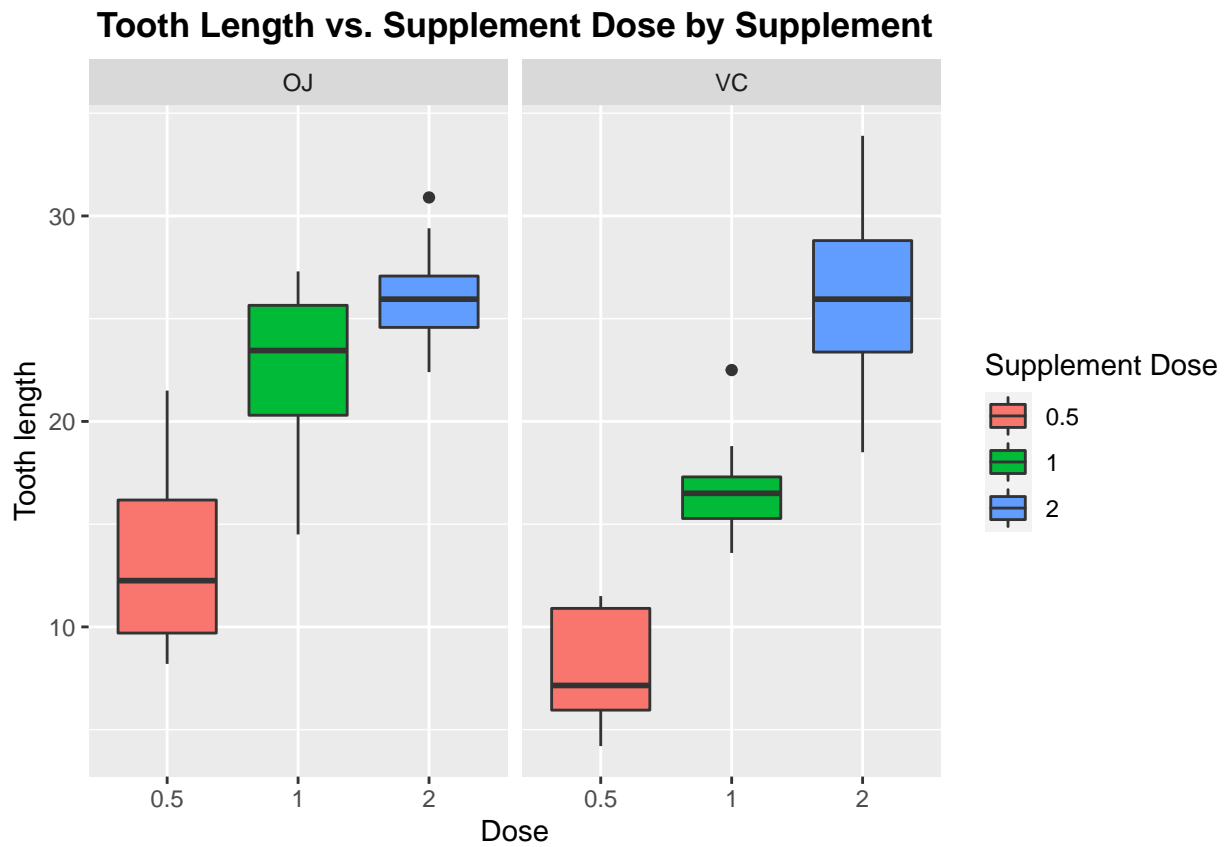

```
box_plt <- ggplot(aes(x = factor(Supplement), y = length), data = tg) +
       geom_boxplot(aes(fill = factor(Supplement))) + facet_wrap(~dose,ncol=3) + xlab("Supplement") + y
print(box_plt)
```

Tooth Length vs. Supplement by Supplement Dose
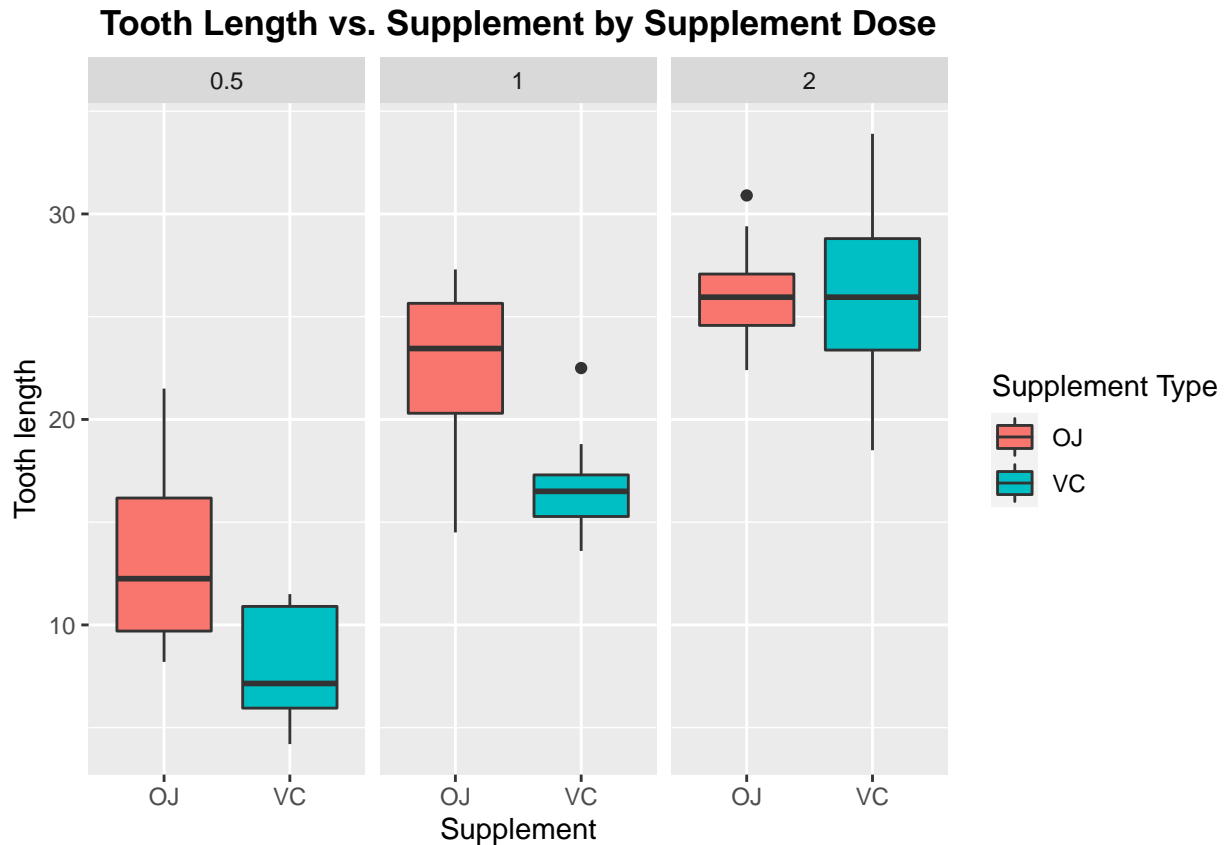
T-Tests and Confidence Intervals

Now we will compare tooth growth by supplement doses using a series of t-tests. Our hypotheses for each dose level are as follows:

- H_0 : there is no difference in the mean tooth growth length between the supplements VC and OJ
- H_a : there is a statistically significant difference in the mean tooth growth length between the supplements VC and OJ

Dose = 0.5

```
t.test(length~Supplement,data=tg[tg$dose==0.5,])
```

```
##
##  Welch Two Sample t-test
##
## data:  length by Supplement
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

The p-value of this test is 0.006. Since the p-value < 0.05 and the confidence interval of the test does not contain a mean difference = 0, we can say that, when compared to each other, the two supplement types (at a dose of 0.5) seem to have an impact on toothgrowth length based on this test. In other words, we reject H_0.

Dose = 1.0

```
t.test(length~Supplement,data=tg[tg$dose==1.0,])
```

```
##
##  Welch Two Sample t-test
##
## data:  length by Supplement
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

The p-value of this test is 0.001. Since the p-value < 0.05 and the confidence interval of the test does not contain a mean difference $= 0$, we can say that, when compared to each other, the two supplement types (at a dose of 1.0) seem to have an impact on toothgrowth length based on this test. In other words, we reject H_0.

Dose = 2.0

```
t.test(length~Supplement,data=tg[tg$dose==2.0,])
```

```
##
##  Welch Two Sample t-test
##
## data:  length by Supplement
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

The p-value of this test is 0.96. Since the p-value > 0.05 and the confidence interval of the test contains a mean difference $= 0$, we can say that, when compared to each other, the two supplement types (at a dose of 2.0) do not have an impact on toothgrowth length based on this test. In other words, we fail to reject H_0.

Conclusion

Based on the above analysis, if the supplement OJ or VC were to be independently and identically administered among a population of guinea pigs, we can conclude that OJ, when administered in a moderate dosage ($<$ 2.0), would have a significant impact on the tooth growth.