Part II - Basic Inferential Data Analysis

Michael Baldassaro

1/10/2018

Overview

This project aims to analyze ToothGrowth data from the R datasets package. Specifically, we will perform some basic exploratory analysis, provide a basic summary of the data, use confidence intervals and hypothesis tests to compare tooth growth by "supp" and "dose", and present preliminary conclusions.

We'll start with the null hypothesis: There is no relationship between tooth growth and Vitamin C dosage by delivery mechanism. We will set a 0.05 significance level to test our hypotheses.

Load Data

As a first step, we will load the ggplot and datasets packages and examine the ToothGrowth data in the datasets package.

```
library(datasets); library(ggplot2)
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 ...
```

The ToothGrowth dataset is a 60 x 3 data frame composed of three variables:

len: the length of the teeth of 10 different guinea pigs supp: the two delivery mechanism for providing Vitamin C to the 10 guinea pigs - orange juice (OJ) or absorbic acid (VC) dose: the amount of Vitamin C provided - 0.5, 1.0 or 2.0 milligrams

Exploratory Data Analysis

We'll perform some basic exploratory data analysis to examine if there is any obvious relationship between tooth length and Vitamin C doses:
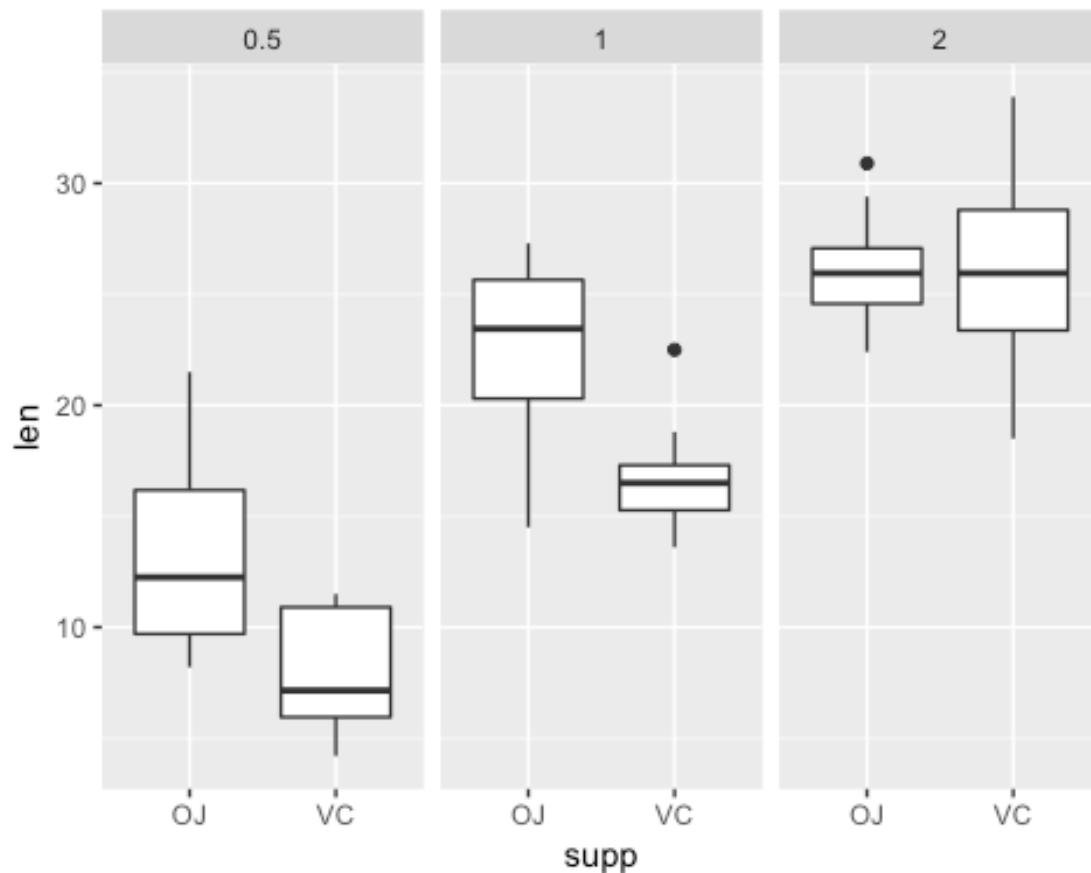
```
ggplot(ToothGrowth, aes(x=as.factor(dose), y=len)) + geom_boxplot()
```
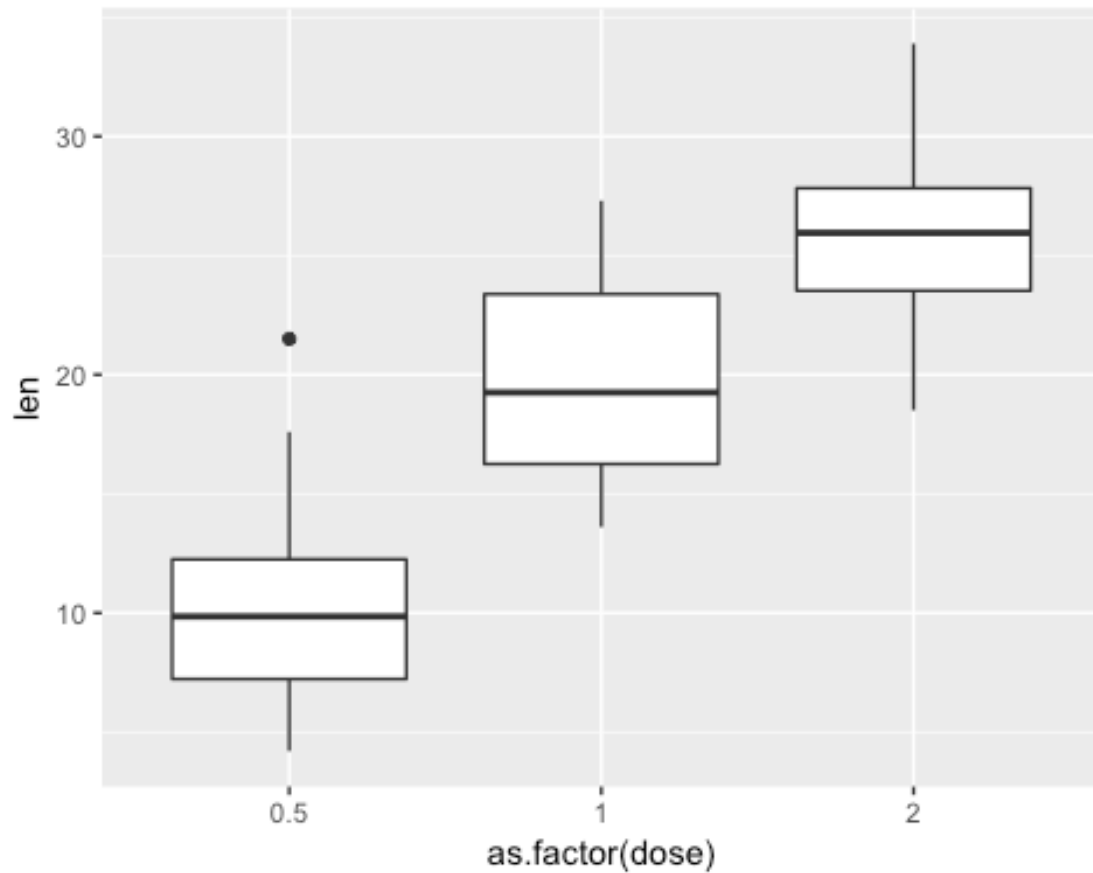
It appears cursorily that there may exist some relationship between Vitamin C dose and tooth length. Now, let's see if there is some relationship between the delivery mechanism for the dose and toothGrowth:

```
ggplot(ToothGrowth, aes(x=supp, y=len)) + geom_boxplot() + facet_grid(~
dose)
```

There seems to be some relationship between tooth growth and delivery mechanism in smaller doses (0.5) and (1), but the difference between means appears non-existent when the dosage is increased (2).

Hypothesis Testing - T Tests, Confidence Intervals & P-Values

We'll see if we can reject the null hypothesis that there is no relationship between tooth growth and dosage by delivery mechanism by performing multiple T tests, finding the 95 confidence interval and p-value:

```
t.test(len~supp, ToothGrowth[ToothGrowth$dose == .5,])

##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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
```

Here we find that the p-value is below the 0.05 significance level we set. Since it is below the significance level, we can reject the null hypothesis that there is no relationship between tooth growth and Vitamin C dosage by delivery mechanism at 0.5mg.

```
t.test(len~supp, ToothGrowth[ToothGrowth$dose == 1.0,])

##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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
```

Here we find the p-value is again below the 0.05 significance level, allowing us to reject the null hypothesis that there is no relationship between tooth growth and Vitamin C dosage by delivery mechanism at 1.0mg.

```
t.test(len~supp, ToothGrowth[ToothGrowth$dose == 2.0,])

##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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
```

Here we find the p-value is well above the 0.05 significance level, which means we cannot reject the null hypothesis that there is no relationship between tooth growth and Vitamin C dosage by delivery mechanism at 2.0mg.

Conclusion

Overall, we cannot reject the null hypothesis that there is no relationship between tooth growth and Vitamin C dosage by delivery mechanism. We will set a 0.05 significance level to test our hypotheses. We have credible evidence to believe that there may be a difference if you consider lower doses and delivery mechanism, but that was not the null hypothesis that we established.