Stat Inference Project Part II

Jillian Katz

12/28/2020

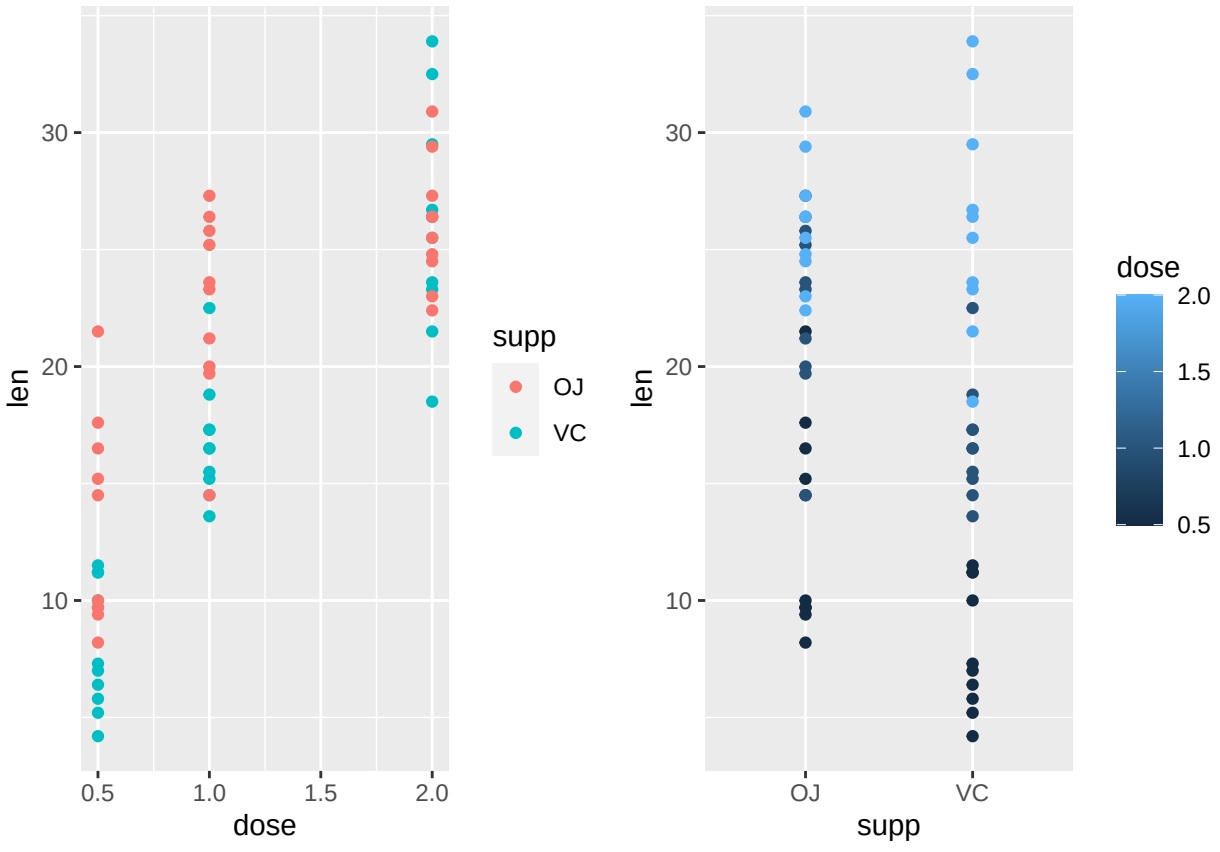
Part 2

Exploration

To begin investigating the ToothGrowth dataset, I have plotted toothgrowth (length, or "len") against both the supplement type ("supp") and the dose quantity ("dose").

After an initial glance at these plots, I can see a couple trends worth investigating quantitatively:

1. Tooth growth seems to increase with vitamin C dose.
2. OJ and VC seem to have an similar mean, though VC appears to have a greater variance.
3. At lower doses, OJ seems to outperform VC.



Investigation & Analysis:

1. Is there an actual difference in tooth growth based on vitamin C dose?

Null Hypothesis (H0): mean(for dose = 0.5) = mean(for dose = 1) = mean(for dose = 2)

This can be investigated with a set of Gosset's t tests with 95% confidence intervals. The results below show each confidence interval to exclude zero, which allows us to reject the null hypothesis: The data suggests a true difference in tooth growth with different doses of Vitamin C.

```
##   Dose Comparisons 95% Confidence Interval
## 1            0.5:1              -11.983781  -6.276219
## 2            0.5:2              -18.156167 -12.833833
## 3              1:2               -8.996481  -3.733519
```

2. Do OJ and Vitamin C perform differently?

First, let's consider a basic center and spread of the data, divided by supplement type.

The data shows that, for this sample, OJ outperformed VC, and the VC data was somewhat more spread out.

```
##                 Data        OJ        VC
## 1               mean 20.663333 16.963333
## 2 standard deviation  6.605561  8.266029
```

A t test will determine if we can expect OJ to outperform VC in a different sample.

A two-sided 95% confidence t test interval gives a confidence interval ranging from -.17 to 7.57, as this includes zero, we fail to reject the null hypothesis. Thus, we expect there to be no difference in tooth growth considering supplement type alone.

```
## [1] -0.1710156  7.5710156
```

3. Does OJ outperform VC at a lower dose?

Even though the whole sample does not suggest a difference due to supplement type, perhaps there is a difference considering dose levels independently.

When we perform the t tests comparing supplement types for each dose level, the result matches what we can infer on the scatterplot: At lower doses (0.5mg and 1mg) of Vitamin C, guinea pig tooth growth is greater when the vitamin C is provided in orange juice. With a dose of 2mg, we fail to reject the null hypothesis and suppose there is no difference in the effectiveness of either form of Vitamin C.

```
##   Supplement Comparisons 95% Confidence Interval
## 1             dose = 0.5                1.719057 8.780943
## 2               dose = 1                2.802148 9.057852
## 3               dose = 2               -3.798070 3.638070
```

Conclusion:

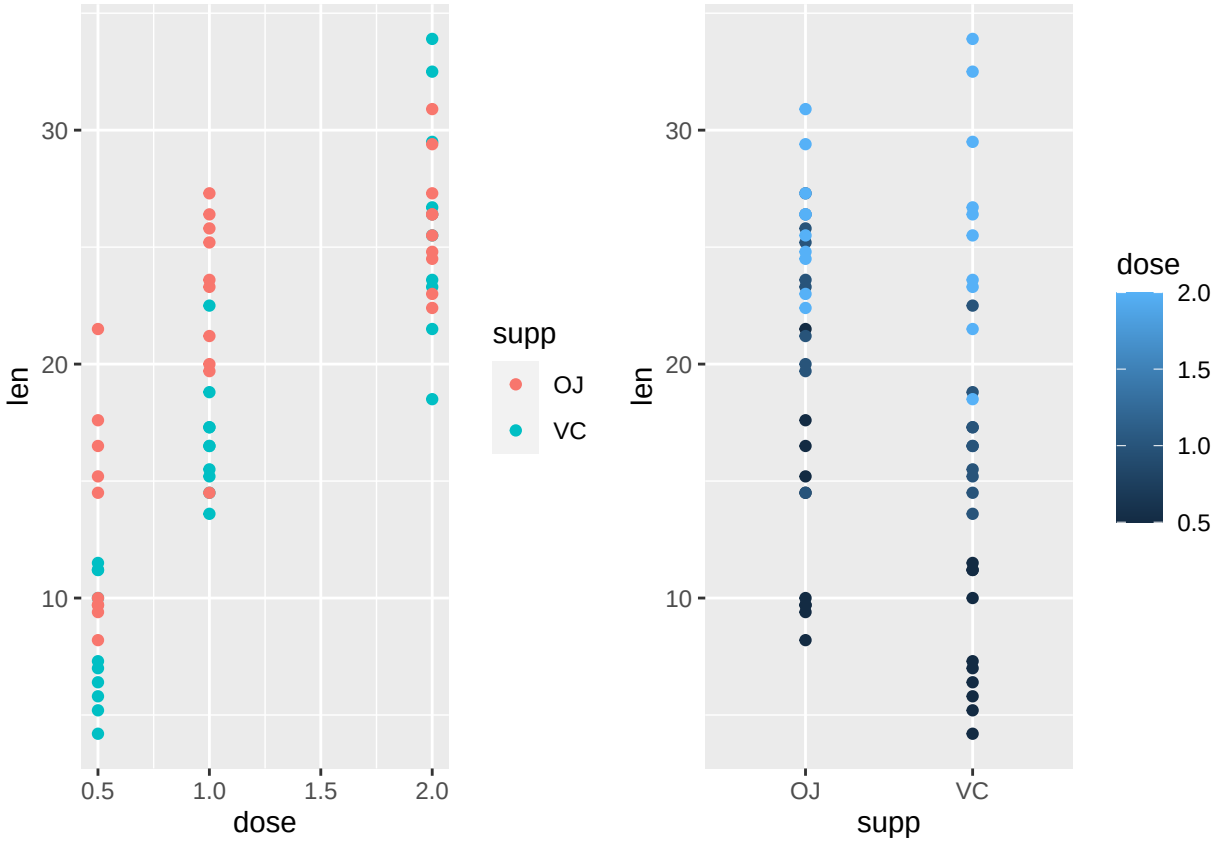
Assuming the ToothGrowth dataset comes from a random, representative sample of the guinea pig population, the data suggests that for doses between 0.5 and 2 mg of Vitamin C per day, higher doses of Vitamin C lead to greater tooth growth. Additionally for Vitamin C supplements of 0.5 or 1mg/day, orange juice leads to more tooth growth compared to ascorbic acid, however there is no difference in result with a dose of 2mg of vitamin C supplement.

Part 2 Code

```
library(ggplot2)
library(dplyr)
library(cowplot)
library(tidyr)

#Data Exploration
data("ToothGrowth")

gds <-ggplot(ToothGrowth, aes(x = dose, y = len)) +
  geom_point(aes(color = supp))
gsd <-ggplot(ToothGrowth, aes(x = supp, y = len)) +
  geom_point(aes(color = dose))
plot_grid(gds, gsd, nrow = 1, ncol=2)
```




```
#Comparing Doses

dhalf <- filter(ToothGrowth, dose ==.5)$len
d1<- filter(ToothGrowth, dose ==1)$len
d2<- filter(ToothGrowth, dose ==2)$len
i1<- t.test(dhalf, d1)$conf[1:2]
i2<- t.test(dhalf, d2)$conf[1:2]
i3<-t.test(d1, d2)$conf[1:2]
```

```r
ints <- data.frame("Dose Comparisons" = c("0.5:1", "0.5:2", "1:2"),
                   c(i1[1], i2[1], i3[1]),
                   c(i1[2], i2[2], i3[2]))
colnames(ints) <- c("Dose Comparisons","95% Confidence Interval", "")
ints

#Comparing Supplement Types
oj <- filter(ToothGrowth, supp == "OJ")$len
vc <- filter(ToothGrowth, supp == "VC")$len
result <- data.frame("Data" = c("mean", "standard deviation"),
                     "OJ" = c(mean(oj), sd(oj)),
                     "VC" = c(mean(vc), sd(vc)))
result

#Comparing Supplement Types at Each Dose
ojhalf <- filter(ToothGrowth, dose ==.5, supp == "OJ")$len
oj1<- filter(ToothGrowth, dose ==1, supp == "OJ")$len
oj2<- filter(ToothGrowth, dose ==2, supp == "OJ")$len

vchalf <- filter(ToothGrowth, dose ==.5, supp == "VC")$len
vc1<- filter(ToothGrowth, dose ==1, supp == "VC")$len
vc2<- filter(ToothGrowth, dose ==2, supp == "VC")$len

inthalf<- t.test(ojhalf, vchalf)$conf[1:2]
int1<- t.test(oj1, vc1)$conf[1:2]
int2<-t.test(oj2, vc2)$conf[1:2]

ints2 <- data.frame("Supplement Comparisons" = c("dose = 0.5", "dose = 1",
                                                 "dose = 2"),
                    c(inthalf[1], int1[1], int2[1]),
                    c(inthalf[2], int1[2], int2[2]))
colnames(ints2) <- c("Supplement Comparisons","95% Confidence Interval", "")
ints2
```