Stat Inference Project Part 2

Joe Logan

Statistical Inference Final Poject

Part 1: Basic Inferential Data Analysis Instructions

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

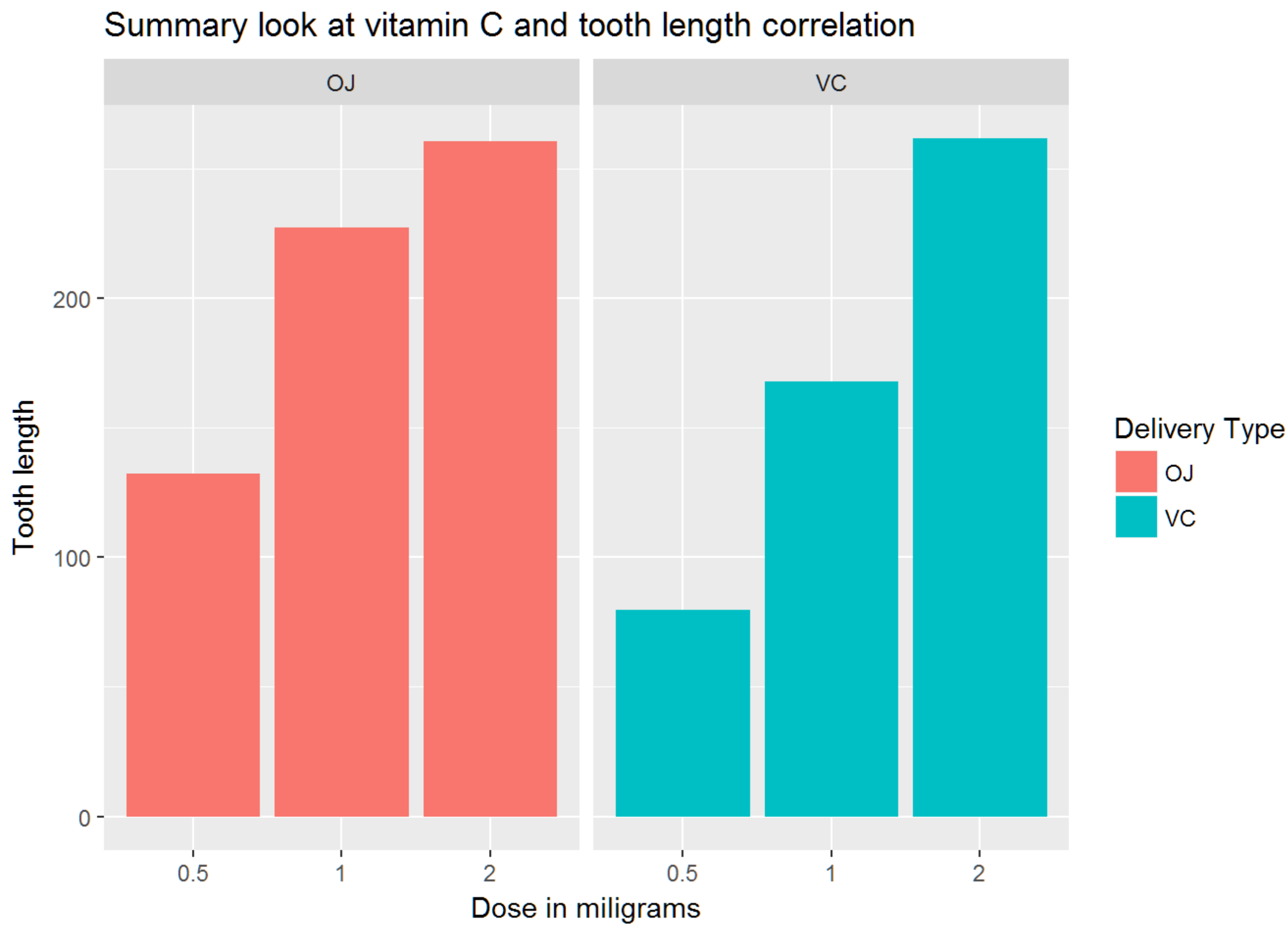
1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)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

Initialization of required libraries and set Options

```
# Set the data
data("ToothGrowth")
toothDataDF <- data.frame(ToothGrowth)
# Here is a quick look a few lines of the data...
head(toothDataDF)
# Here is a look at some key statistics.
summary(toothDataDF)
```

Now lets take a quick look at a plot of the data

```
ggplot(data=toothDataDF, aes(x=as.factor(dose), y=len, fill=supp)) +
    geom_bar(stat="identity") + facet_grid(. ~ supp) +
    labs(x="Dose in miligrams", y="Tooth length", title="Summary look at vitamin C and tooth len
gth correlation") +
    guides(fill=guide_legend(title="Delivery Type"))
```



Hypothesis testing

Hypothesis for the impact of SUPPLIMENT TYPE (Delivery Method)

Looking at the chart above it does seem that there is a correlation in both the dosage and delivery method of vitamin C in guinea pigs. Our null hypothesis is that the tooth growth is zero. (H0)

Let's run a t-test for the type of suppliment (Orange Juice vs. Vitamin C)

```
mean_supp = split(toothDataDF$len, toothDataDF$supp)
sapply(mean_supp, var)

t.test(toothDataDF$len[toothDataDF$supp == "VC"], toothDataDF$len[toothDataDF$supp == "OJ"], pai
red=FALSE, var.equal=FALSE)
```

Here we can see a p-value of .06 which is low which means we cannot in confidence reject the null hypothesis.

Hypothesis for the impact of DOSAGE QUANTITY

Let's run a t-test on this as well using both the .5. > 1 and 1 > 2 dosages

```
t.test(toothDataDF$len[toothDataDF$dose==1], toothDataDF$len[toothDataDF$dose==0.5], paired = FA
LSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  toothDataDF$len[toothDataDF$dose == 1] and toothDataDF$len[toothDataDF$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
```

```
t.test(toothDataDF$len[toothDataDF$dose==2], toothDataDF$len[toothDataDF$dose==1], paired = FALS
E, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  toothDataDF$len[toothDataDF$dose == 2] and toothDataDF$len[toothDataDF$dose == 1]
## t = 4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.735613 8.994387
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

```
t.test(toothDataDF$len[toothDataDF$dose==2], toothDataDF$len[toothDataDF$dose==0.5], paired = FA
LSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  toothDataDF$len[toothDataDF$dose == 2] and toothDataDF$len[toothDataDF$dose == 0.5]
## t = 11.799, df = 38, p-value = 2.838e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  12.83648 18.15352
## sample estimates:
## mean of x mean of y
##    26.100    10.605
```

All 3 tests show a large confidence interview as well as a low P-value which indicates that the dosage increases do in fact correlate to greater tooth length in guinea pigs.