

Basic Inferential Data Analysis

Tony Smith

March 3, 2021

Overview

In this analysis, I will be looking at the ToothGrowth data from R datasets. I will perform some simple exploratory analysis and then compare the data by supp and dose.

Exploratory Data Analysis

```
data(ToothGrowth)
head(ToothGrowth)
```

```
##      len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

Just doing a quick analysis to see the data grouped by supplement and dose:

```
ToothGrowth %>% group_by(supp, dose) %>% summarise_all(mean)
```

```
## # A tibble: 6 x 3
## # Groups:   supp [2]
##   supp    dose  len
##   <fct> <dbl> <dbl>
## 1 OJ      0.5 13.2
## 2 OJ      1  22.7
## 3 OJ      2  26.1
## 4 VC      0.5  7.98
## 5 VC      1  16.8
## 6 VC      2  26.1
```

I see that for both supplements the tooth growth goes up as the dose goes up. It also looks like the OJ supplement is better on average across two of the doses but not the third.

Hypothesis Testing

My first null hypothesis for the supplements is that there is no difference in means. My alternative hypothesis is that they are not equal. I assume that the variances are equal and the population from which they are drawn is normally distributed. I will use a t-Test to find the p value and determine if I can reject the null hypothesis.

```
t.test(ToothGrowth$len[1:30], ToothGrowth$len[31:60], var.equal=TRUE)
```

```
##
## Two Sample t-test
##
## data: ToothGrowth$len[1:30] and ToothGrowth$len[31:60]
## t = -1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -7.5670064 0.1670064
## sample estimates:
## mean of x mean of y
## 16.96333 20.66333
```

We can see that the p value is .06 which is greater than .05, so we fail to reject the null hypothesis. We can also see that 0 is in the confidence interval.

My second null hypothesis for the doses is that there is no difference in the means between the dose of 1.0 and the dose of 2.0. My alternative hypothesis is that the 2.0 dose is greater than the 1.0 dose in terms of effect on growth. I make the same assumptions as I did above and perform another t-Test.

```
smalldose <- ToothGrowth[ToothGrowth$dose == 1.0,]
largedose <- ToothGrowth[ToothGrowth$dose == 2.0,]
t.test(smalldose$len, largedose$len, var.equal=TRUE, alternative="less")
```

```
##
## Two Sample t-test
##
## data: smalldose$len and largedose$len
## t = -4.9005, df = 38, p-value = 9.054e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
## -Inf -4.175196
## sample estimates:
## mean of x mean of y
## 19.735 26.100
```

This time the p-value comes out to be very small $\ll .05$, so this tells me that the null hypothesis must be rejected.