Statistical Inference: Inferential Analysis

Seth Dimick

12/24/2017

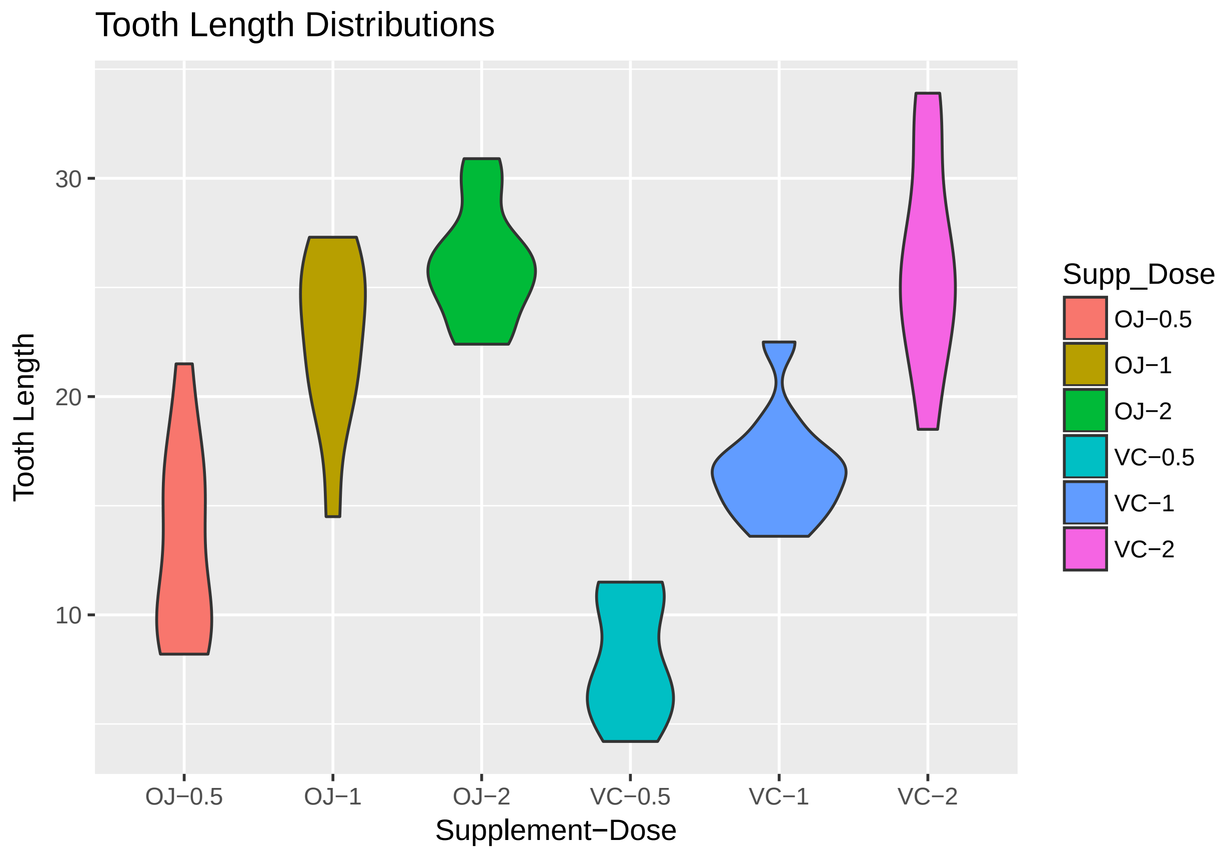
Here we'll analyze the assigned data set and see if we can draw any statistically significant conclusions. The data set is called *ToothGrowth* and includeds the variables len, supp, and dose. Let's assume that supp is the type of supplement intended to promote tooth growth, dose is the dosage prescribed to a subject, and len is the observed lenth of a specific tooth of each subject after completing the prescribed supplement-dosage regiment.

Let's take a look at the data numerically and visually.

```
data("ToothGrowth")
ToothGrowth$Supp_Dose <- factor(paste(ToothGrowth$supp, ToothGrowth$dose, sep = '-'))
ToothGrowth %>%
  select(len,Supp_Dose) %>%
  group_by(Supp_Dose) %>%
  summarise(mean = mean(len), sdev = sd(len), n = n()) %>%
  as.data.frame()
```

```
##   Supp_Dose  mean     sdev  n
## 1    OJ-0.5 13.23 4.459709 10
## 2      OJ-1 22.70 3.910953 10
## 3      OJ-2 26.06 2.655058 10
## 4    VC-0.5  7.98 2.746634 10
## 5      VC-1 16.77 2.515309 10
## 6      VC-2 26.14 4.797731 10
```

```
ToothGrowth %>%
  ggplot(aes(x = Supp_Dose, y = len, fill = Supp_Dose)) +
  geom_violin() +
  ggtitle("Tooth Length Distributions") +
  xlab("Supplement-Dose") +
  ylab("Tooth Length")
```



With our 10 observations of each supplement-dosage combinations, it appears our big difference in tooth length comes from the dosage levels, and not the supplement type. Let's test the difference of means between dosage levels within supplement groups, between dosage of 1 and 0.5 and between 2 and 1. Our hypotheses are that in each group, a dosage of 1 has a greater mean than a dosage of 0.5, and a dosage of 2 has a greater mean than a dosage of 1.

T-Test Results:

```
#Set up our length measurments for each group
OJ0.5 <- ToothGrowth[ToothGrowth$Supp_Dose=='OJ-0.5',]$len
OJ1 <- ToothGrowth[ToothGrowth$Supp_Dose=='OJ-1',]$len
OJ2 <- ToothGrowth[ToothGrowth$Supp_Dose=='OJ-2',]$len
VC0.5 <- ToothGrowth[ToothGrowth$Supp_Dose=='VC-0.5',]$len
VC1 <- ToothGrowth[ToothGrowth$Supp_Dose=='VC-1',]$len
VC2 <- ToothGrowth[ToothGrowth$Supp_Dose=='VC-2',]$len

#Test OJ1 vs OJ0.5
t.test(OJ1, OJ0.5)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ1 and OJ0.5
## t = 5.0486, df = 17.698, p-value = 8.785e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   5.524366 13.415634
## sample estimates:
```

```
## mean of x mean of y
##     22.70     13.23
```

```r
#Test OJ2 vs OJ1
t.test(OJ2, OJ1)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ2 and OJ1
## t = 2.2478, df = 15.842, p-value = 0.0392
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1885575 6.5314425
## sample estimates:
## mean of x mean of y
##     26.06     22.70
```

```r
#Test VC1 vs VC0.5
t.test(VC1, VC0.5)
```

```
##
##  Welch Two Sample t-test
##
## data:  VC1 and VC0.5
## t = 7.4634, df = 17.862, p-value = 6.811e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.314288 11.265712
## sample estimates:
## mean of x mean of y
##     16.77      7.98
```

```r
#Test VC2 vs VC1
t.test(VC2, VC1)
```

```
##
##  Welch Two Sample t-test
##
## data:  VC2 and VC1
## t = 5.4698, df = 13.6, p-value = 9.156e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   5.685733 13.054267
## sample estimates:
## mean of x mean of y
##     26.14     16.77
```

Conclusion:

Eeach of our t-tests produce 95% confidence intervals greater than 0 in entirety, so we can reject our null hypotheses and say that within each supplement group, a dosage of 1 results in longer teeth than a dosage of 0.5 and a dosage of 2 results in longer teeth than a dosage of 1.