Basic Tooth Growth Inference

Anton Hesse

9/10/2020

We will begin this data analysis by loading ggplot2 and the data, and then looking at the data structure.

```
library(ggplot2)
data("ToothGrowth")
str(ToothGrowth)
```

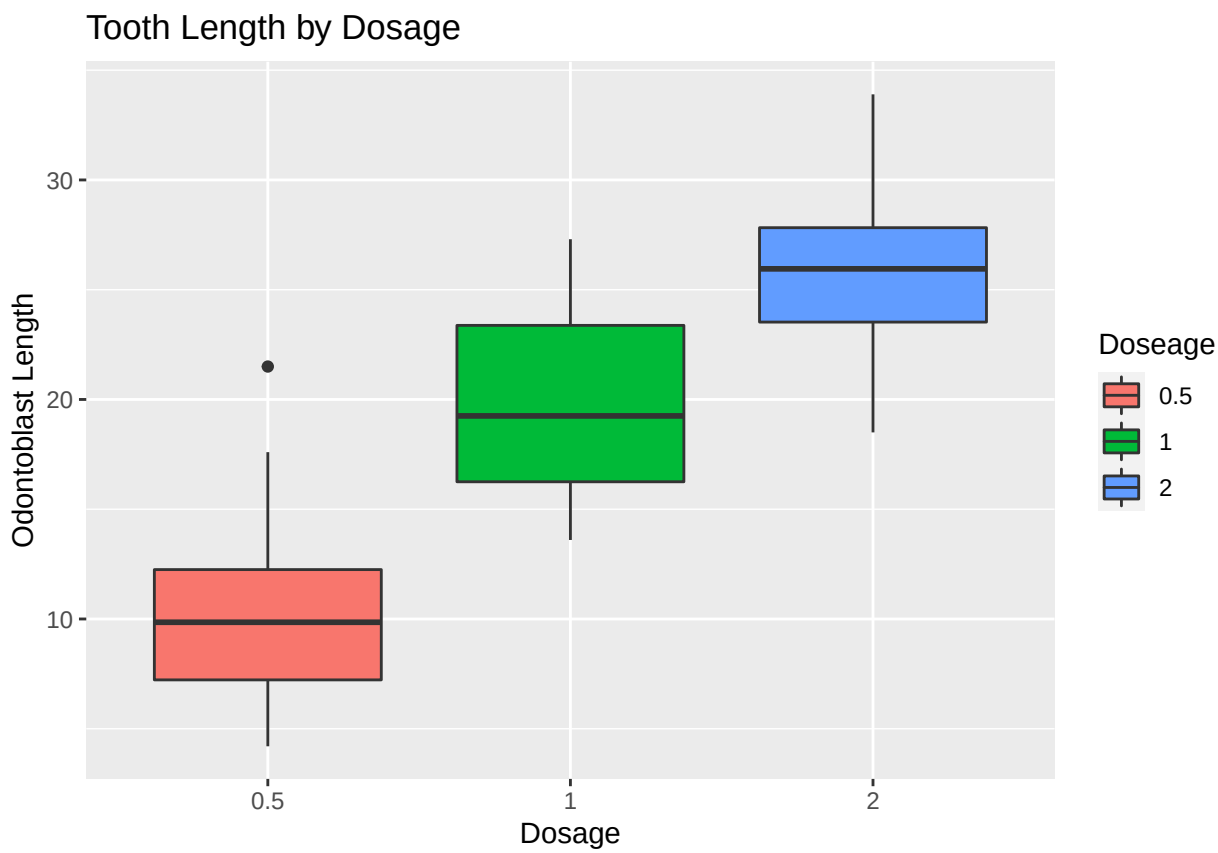
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Next we will read about the data with the help() function.

```
help(ToothGrowth)
```

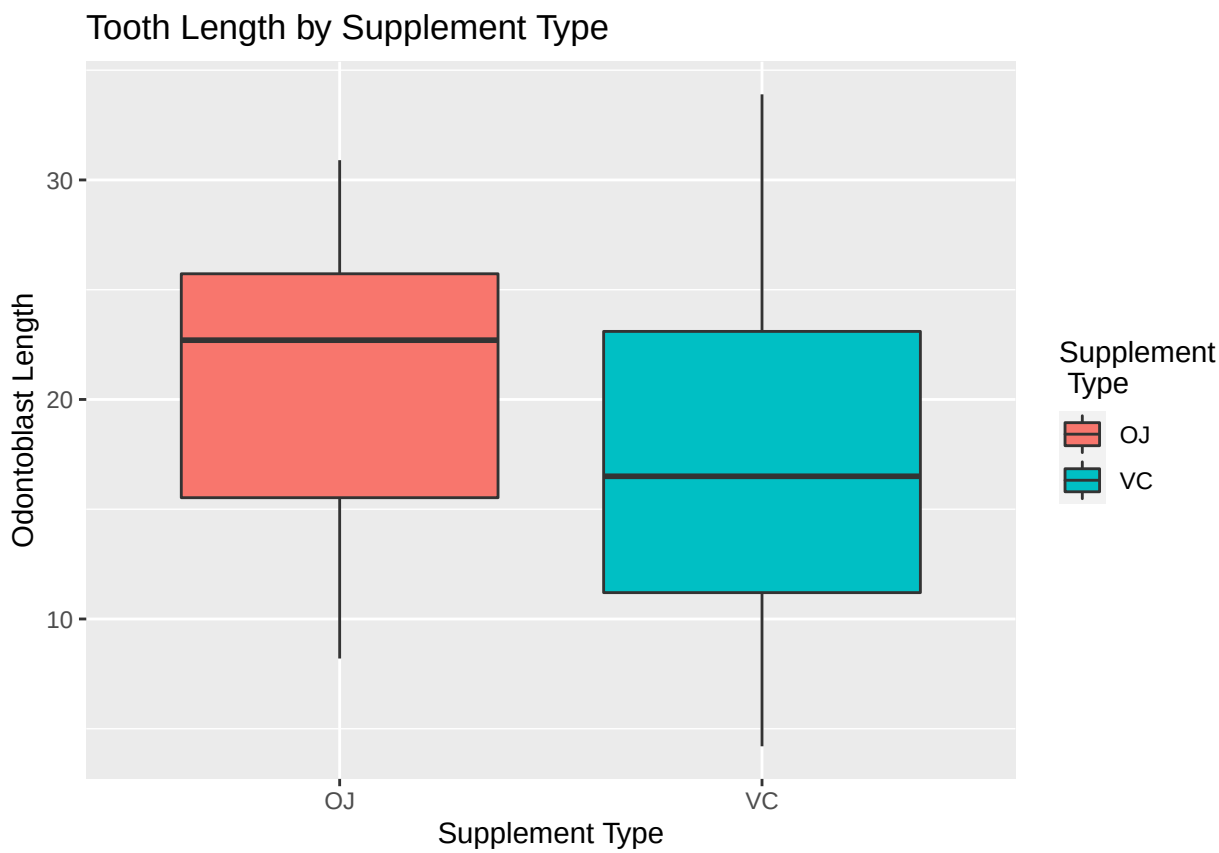
The description tells us that there were 3 dosages of vitamin C in two different supplement forms. The effects of these variables were applied to the tooth length of guinea pigs. Thus, we will examine our data by considering dose as a factor. We will first examine tooth growth vs. dosage.

```
ggplot(ToothGrowth) +
    geom_boxplot(aes(x = as.factor(dose), y = len, fill = as.factor(dose))) +
    ggtitle( "Tooth Length by Dosage") +
    xlab("Dosage") +
    ylab("Odontoblast Length") +
    labs(fill = "Doseage")
```



Visual assessment suggests that all three groups have different means. Now we will assess the supplementation method.

```
ggplot(ToothGrowth) +
    geom_boxplot(aes(x = supp, y = len, fill = supp)) +
    ggtitle( "Tooth Length by Supplement Type") +
    xlab("Supplement Type") +
    ylab("Odontoblast Length") +
    labs(fill = "Supplement\n Type")
```



The difference between tooth growth by supplementation type do not appear as stark as compared to dosage, but direct vitamin C supplementation may result in lower tooth growth.

For simplicity, we will only examine two of the dosages against each other. We will compare the 1 and 2 mg/day with an independent t-test because those are closer together and thus harder to tell if there is a difference between them.

We will use an independent t-test because separate guinea pigs received different doses. We will also assume that each group equal variance. We can compute the variance for each group as follows:

```
byDose <- split(ToothGrowth, ToothGrowth$dose)
sapply(byDose, function(x) var(x[,1]))
```

```
##      0.5        1        2
## 20.24787 19.49608 14.24421
```

Although they aren't exactly the same, this is a balanced design because the number of subjects in each group are:

```
sapply(byDose, function(x) length(x[,1]))
```

```
## 0.5   1   2
##  20  20  20
```

Therefore, we have some robustness against heterogeneity of variance. Our final assumption is that our sample size is sufficiently large for the central limit theorum to apply. The general rule of thumb is that at

least 30 observations is required, and we have 1000. Given that we saw how the 2 mg/day dose was higher, we will use a one-sided hypothesis test.

```
t.test(byDose[2][[1]][1], byDose[3][[1]][1],
       alternative = "two.sided", paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  byDose[2][[1]][1] and byDose[3][[1]][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:
##  -8.994387 -3.735613
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

This shows we have sufficient evidence to reject the null hypothesis ($p < 0.001$) that there is no difference between the 1 mg/day and 2mg/day doses. Furthermore, we are 95% confident that the true mean difference between the 1 and 2 mg/day doses lies between 19.74 and 26.10.

We will now test the difference between the two supplementation methods. We will make the same assumptions regarding equal variance between groups and normality. Given that we saw how the orange juice method was higher, we will use a one-sided hypothesis test.

```
bySupp <- split(ToothGrowth, ToothGrowth$supp)
t.test(bySupp[1][[1]][1], bySupp[2][[1]][1],
       alternative = "two.sided", paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  bySupp[1][[1]][1] and bySupp[2][[1]][1]
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

We do not have sufficient evidence to reject the null hypothesis ($p > 0.05$) that there is no difference in tooth growth length between the vitamin C and ascorbic acid supplementation methods. Also, the 95% confidence interval includes 0.