Statistical Inference Project 2 Part 2

Nicolas Moreno Andrade

August 14, 2017

Overview

In this project we will study the ToothGrowth data set. This set contains data about the effect of vitamin C in tooth growth (measured by length) of guinea pigs. We'll perform some exploratory data analysis by performing quick summaries and drawing boxplots by dose level and delivery method. Additionaly t-tests will be run in order to confirm the relationships seen in the plots. At the end we will conclude that there is a positive and significant relationship between tooth growth and ingestion of vitamin C and no relationship with delivery method.

1 Basic Summary of Data

The ToothGrowth data contains 60 observations of length of odontoblasts (cells responsible for tooth growth) in guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

The following are basic summaries of the data:

```
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
table(ToothGrowth$dose, ToothGrowth$supp)
```

```
##
##        OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10

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 ...
```

2 Analysis

```
# boxplot delivery methods
ggplot(aes(x=supp, y=len), data=ToothGrowth) +
  geom_boxplot(aes(fill=supp))+
        xlab("Supplement type") +ylab("Tooth length") +
  scale_fill_discrete(name="Delivery Method")
```

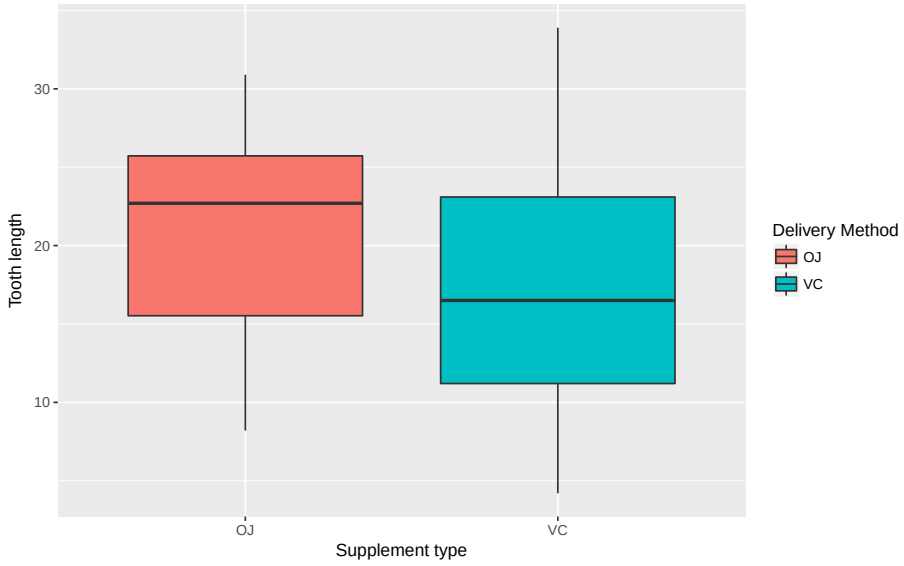

Figure 1: Boxplot comparing delivery methods by tooth length

Based on this boxplot we see that dividing the observations just by delivery method doesn't seem to show relevant differences in tooth growth length. To confirm this we perform a t-test:

```
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

The confidence interval contains 0 and the p-value is greater than 0.05 so we fail to reject the null hypothesis. There is no significant difference in tooth length between the group that received the vitamin C in orange juice and the group that received it in the form of ascorbic acid.

```
# boxplot dose
ggplot(aes(x=factor(dose), y=len), data=ToothGrowth) +
  geom_boxplot(aes(fill=factor(dose)))+
        xlab("Dose") +ylab("Tooth length") +
  scale_fill_discrete(name="Dose Level")
```

The boxplot indicates that there could be a relationship between dosage and tooth length. The bigger the dose we see longer teeth in the animals.

We now perform a t-test for each dose level:

```
dose05vsdose1 <- ToothGrowth[ToothGrowth$dose %in% c(0.5, 1.0), ]
dose05vsdose2 <- ToothGrowth[ToothGrowth$dose %in% c(0.5, 2.0), ]
dose1vsdose2 <- ToothGrowth[ToothGrowth$dose %in% c(1.0, 2.0), ]
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = dose05vsdose1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
```

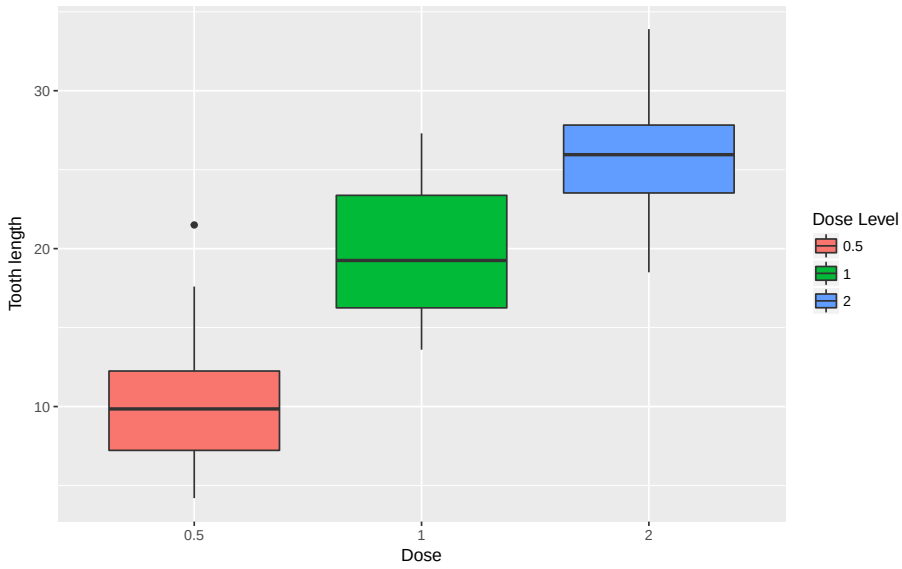

Figure 2: Boxplot comparing dose by tooth length

```
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = dose05vsdose2)

##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

```
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = dose1vsdose2)

##
##  Welch Two Sample t-test
```

```
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

The three t-tests show similar results. The confidence intervals do not contain 0 and the p-value is very small (almost 0). We conclude that the mean tooth length increases with larger dose levels.

3 Conclusions

1. The data indicates that, in guinea pigs, mean tooth length increases with larger dose levels of vitamin C.
2. The delivery method does not seem to have a relationship with mean growth length.

4 Assumptions

1. The 60 guinea pigs are a representative sample of the population of the guinea pigs.
2. We assumed unequal variance for the t-tests.