Statistical Inference Course Project - Basic Inferential Data Analysis

Gary Lu

January 7, 2021

Contents

In this part, we will analyze the ToothGrowth data in the R datasets package.

Data Exploration

First, let’s load the data and explore the data.

```
data("ToothGrowth") # The Effect Of Vitamin C On Tooth Growth In Guinea Pigs
```

```
str(ToothGrowth) # Display the structure of the data
```

```
## ’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 ...
```

```
head(ToothGrowth, 5) # Explore the first 5 rows of the data
```

```
##    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
```

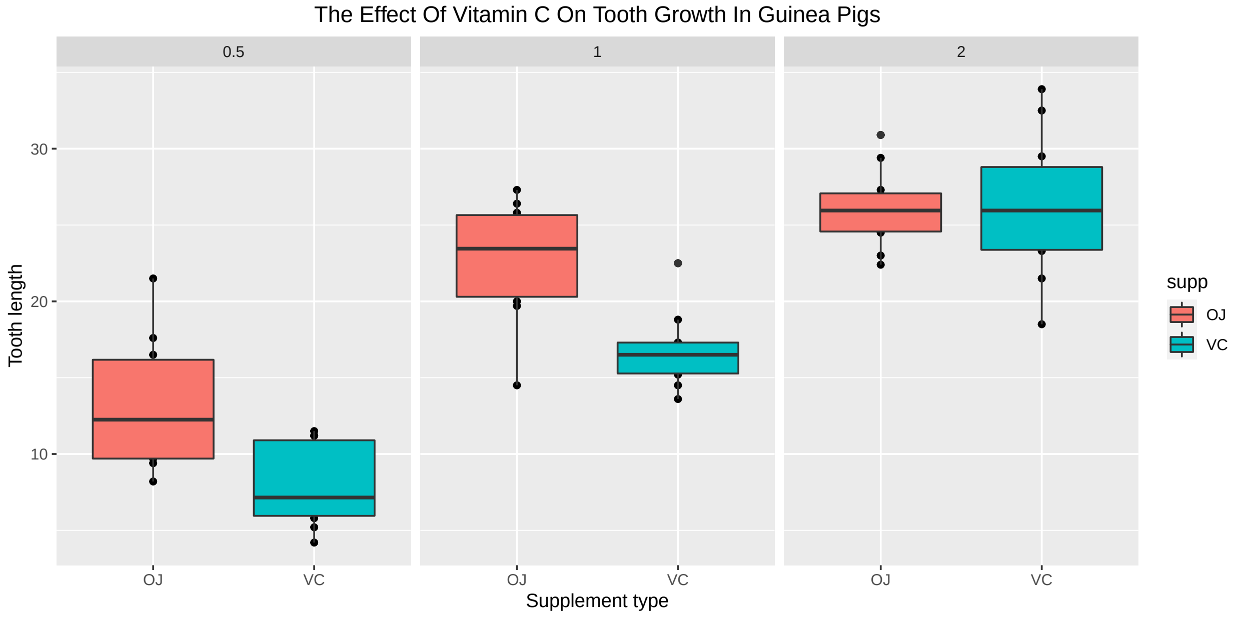
```
summary(ToothGrowth) # Produce summaries of the data
```

```
##       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  
```

ToothGrowth dataset has 60 observations and 3 variables:

- len: numeric, Tooth length
- supp: factor, Supplement type (VC or OJ).
- dose: numeric, Dose in milligrams/day

```
qplot(supp, len,
      data=ToothGrowth,
      facets=~dose,
      main="The Effect Of Vitamin C On Tooth Growth In Guinea Pigs",
      xlab="Supplement type",
      ylab="Tooth length") +
  geom_boxplot(aes(fill = supp)) +
  theme(plot.title = element_text(hjust = 0.5))
```



From the plot, we can observe the following 2 hypothesis that we will test in the next step:

1. Increasing the dosage has a positive effect on tooth growth.
2. In general, the OJ type supplement has a better effect than the VC type, but under 2.0 dosage, two supplement types may have similar effect.

Hypothesis Testing

Before we carry out the hypothesis testing, we have to make some assumptions:

1. The random variables are independent and identical distributed (i.i.d.);
2. The tooth growth roughly follow a normal distribution;
3. The variances of tooth growth is different under different supplement and dosage.

And through all the following testing, we will use 5% as the tolerance limit of errors.

Dosage

In order to test the hypothesis that increasing the dosage has a positive effect on tooth growth, we will split the data into 3 group `dose_05`, `dose_10`, and `dose_20` according to different dosage.

```
dose_05 <- ToothGrowth$len[ToothGrowth$dose == 0.5]
dose_10 <- ToothGrowth$len[ToothGrowth$dose == 1]
dose_20 <- ToothGrowth$len[ToothGrowth$dose == 2]
```

First, we will compare `does_05` and `dose_10`. The null hypothesis H_0 is $len(dose\_05) = len(dose\_10)$, and the alternative hypothesis H_a is $len(dose\_05) < len(dose\_10)$. We will have one-sided t-test with unequal variance.

```
t.test(dose_05, dose_10,
       alternative = "less", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_05 and dose_10
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.753323
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

As the p value (6.342×10^{-8}) is lower than 5%, we reject the null hypothesis.

Next, we will carry out the similar t-test for `dose_10` and `dose_20`. The null hypothesis H_0 is $len(dose\_10) = len(dose\_20)$, and the alternative hypothesis H_a is $len(dose\_10) < len(dose\_20)$.

```
t.test(dose_10, dose_20,
       alternative = "less", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_10 and dose_20
## t = -4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -4.17387
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

As the p value (9.532×10^{-6}) is lower than 5%, we reject the null hypothesis.

With the testings above, we conclude that it is very likely that a higher dosage has positive effect on tooth growth.

Supplement Type

First, let’s test the hypothesis about two different supplement types in general. We will split the data into two groups `oj` and `vc`.

```
oj <- ToothGrowth$len[ToothGrowth$supp == ’OJ’]
vc <- ToothGrowth$len[ToothGrowth$supp == ’VC’]
```

The null hypothesis H_0 is $len(oj) = len(vc)$, and the alternative hypothesis H_a is $len(oj) < len(vc)$. We will have one-sided t-test with unequal variance.

```
t.test(oj, vc,
       alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  oj and vc
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

As the p value (0.03032) is lower than 5%, we reject the null hypothesis.

Next, we will test the hypothesis about two different supplement types when dosage equals to 2.0. We will define the following two groups `oj_20` and `vc_20`.

```
oj_20 <- ToothGrowth$len[ToothGrowth$supp == ’OJ’ & ToothGrowth$dose == 2]
vc_20 <- ToothGrowth$len[ToothGrowth$supp == ’VC’ & ToothGrowth$dose == 2]
```

The null hypothesis H_0 is $len(oj\_20) = len(vc\_20)$, and the alternative hypothesis H_a is $len(oj) \neq len(vc)$. We will have two-sided t-test with unequal variance.

```
t.test(oj_20, vc_20,
       alternative = "two.sided", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  oj_20 and vc_20
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

As the p value (0.9639) is greater than 5%, we fail to reject the null hypothesis.

With the testings above, we conclude that it is likely that supplement type OJ has a better effect on tooth growth than supplement type VC in general. However, under dosage 2.0 milligrams/day, there is not enough evidence to show that there is a difference between the two types.