Statistical Inference Course Project: Part II

Yifan XIA

5 Feb 2017

```
library(ggplot2); library(datasets)
```

Basic Inferential Data Analysis

Summary of the *ToothGrowth* dataset

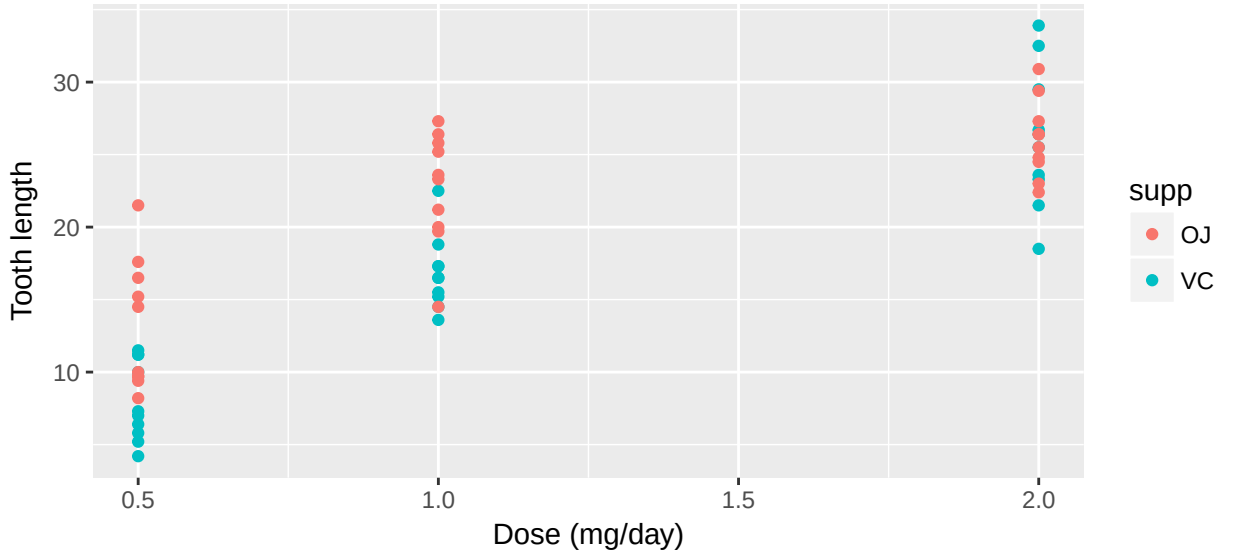
The *ToothGrowth* dataset have 60 observations of 3 variables: *len* is a numeric variable recording the tooth length, *supp* is a factor variable of two levels representing two types of supplements (*OJ* for orange juice, and *VC* for ascorbic acid, a form of vitamin C), and *dose* is a numeric variable that gives the dose of supplement. Three doses, 0.5 mg/day, 1.0 mg/day, 2.0 mg/day, are recorded in the dataset.

```
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
```

We see from the above summary of the dataset that 30 observations are made for each type of supplements.

```
ggplot(data = ToothGrowth, aes(dose, len)) + geom_point(aes(color=supp)) +
       labs(x = 'Dose (mg/day)', y = 'Tooth length')
```



This figure shows clearly that the tooth length increases with the supplement dose. However, for each value of dose, it is not clear whether one supplement has a stronger effect on the tooth growth than the other.

In order to ascertain the supplements' effect on tooth growth depending on their dose, we perform hypothesis tests for each value of dose. We will test the null hypothesis, *i.e* there is no difference between the two supplements, using a two-side t test.

Hypothesis tests

Dose = 0.5 mg/day

We choose a subset where dose = 0.5, and group the subset by *supp*.

```
OJ05 <- subset(ToothGrowth, supp == 'OJ' & dose == 0.5)
VC05 <- subset(ToothGrowth, supp == 'VC' & dose == 0.5)
t.test(OJ05$len, VC05$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ05$len and VC05$len
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

The t test gives a p-value of 0.006 for the null hypothesis, and a 95% confidence interval whose lower limit (~1.7) is larger than 0. This result strongly suggests that at a dose of 0.5 mg/day, the supplement OJ significantly increases the tooth growth, *i.e.* the null hypothesis is rejected.

Dose = 1.0 mg/day

We choose a subset where dose = 1.0, and group the subset by *supp*.

```
OJ1 <- subset(ToothGrowth, supp == 'OJ' & dose == 1.0)
VC1 <- subset(ToothGrowth, supp == 'VC' & dose == 1.0)
t.test(OJ1$len, VC1$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ1$len and VC1$len
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

The t test gives a p-value of 0.001 for the null hypothesis, and a 95% confidence interval whose lower limit (~2.8) is larger than 0. Again, the result strongly suggests that at a dose of 1.0 mg/day, the supplement OJ significantly increases the tooth growth, *i.e.* the null hypothesis is rejected.

Dose = 2.0 mg/day

We choose a subset where dose = 2.0, and group the subset by *supp*.

```
OJ2 <- subset(ToothGrowth, supp == 'OJ' & dose == 2.0)
VC2 <- subset(ToothGrowth, supp == 'VC' & dose == 2.0)
t.test(OJ2$len, VC2$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ2$len and VC2$len
## 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
```

The t test gives a p-value of 0.96 for the null hypothesis, and a 95% confidence interval that is almost centered at 0. Contrary to the previous two tests, this result suggests that at a dose of 2.0 mg/day, the two supplements have no significant difference on the tooth growth, *i.e.* the null hypothesis is kept.

Conclusion

The above inferential analysis suggests that compared with orange juice, the orange juice helps the tooth to grow when used at a relatively small dose (<0.5 mg/day). However, at large dose (>2.0 mg/day), the vitamin C and the orange juice show comparable effect on the tooth growth.

Assumptions

The following assumptions are made to draw the above conclusion:

- All measurements are independent from each other.
- The chosen guinea pigs are representative of the whole guinea pig population.
- The measurements are not paired.
- The variance of two supplement groups are different.