Statistical Inference Course Project

Yulong Wang

19/09/2021

```
setwd("/Users/yulong/GitHub/Statistical-Inference-Course-Project")
```

Part 2: Basic Inferential Data Analysis Instructions

Instructions

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

- Load the ToothGrowth data and perform some basic exploratory data analyses
- Provide a basic summary of the data.
- Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
- State your conclusions and the assumptions needed for your conclusions.

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
data("ToothGrowth")
data <- as.data.frame(ToothGrowth)
str(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 ...
```

2. Provide a basic summary of the data.

```
head(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
## 6 10.0   VC  0.5
```

```
summary(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
```

```
dose<-as.factor(data$dose)
table(data$supp, data$dose)
```

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

The ToothGrowth dataset explains the relation between the growth of teeth of guinea pigs at each of three dose levels of Vitamin C (0.5, 1 and 2 mg) with each of two delivery methods(orange juice and ascorbic acid).

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

```
# T Test by supplemant type
t.test(len ~ supp, 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 between group OJ and group VC 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
```

```r
# T test by dose level
Tooth.dose0.5_1.0 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
Tooth.dose0.5_2.0 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
Tooth.dose1.0_2.0 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
t.test(len ~ dose, data = Tooth.dose0.5_1.0)
```

```
##
##  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 between group 0.5 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```r
t.test(len ~ dose, data = Tooth.dose0.5_2.0)
```

```
##
##  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 between group 0.5 and group 2 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
```

```r
t.test(len ~ dose, data = Tooth.dose1.0_2.0)
```

```
##
##  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 between group 1 and group 2 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
```

```r
# T test for supplement by dose level
Tooth.dose0.5 <- subset(ToothGrowth, dose == 0.5)
Tooth.dose1.0 <- subset(ToothGrowth, dose == 1.0)
Tooth.dose2.0 <- subset(ToothGrowth, dose == 2.0)
t.test(len ~ supp, data = Tooth.dose0.5)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len ~ supp, data = Tooth.dose1.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len ~ supp, data = Tooth.dose2.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

4. State your conclusions and the assumptions needed for your conclusions.

The p.value indicators for doses of `0.5` and `1.0` are `0.0064` and `0.001` respectively, being less than 5% and the confidence intervals of the test do not contain 0. The p.value for the dose of `2.0` is `0.064`, which is greater than 5% and the confidence test contains 0. Based of the results, we can say that for doses of `0.5` and `1.0`, OJ has a greater effect on Tooth Growth than VC. For the test at `dose == 2.0` we cannot reject the Ho.

We can then conclude that to get greater tooth growth with low levels of dosage (0.5 & 1.0) one should use OJ instead of VC. At greater levels (2.0) of dosage it is uncertain whether there will be a greater effect from either OJ or VC.