Statistical Inference Project

Farhan Choudhary

12th November 2017

Instructions for the Project

This project consists of two parts and in this simulation we will explore inference and perform some simple inferential statistics. The project is divided into two parts

- A Simulation Exercise
- Basic Inferential Data Analysis

Problem Statement

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

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. 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)
4. State your conclusions and the assumptions needed for your conclusions.

B. Inference on ToothGrowth Data

1. Loading the dataset and basic exploratory data analysis

```
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 ...
```

```
rbind(head(ToothGrowth,2), tail(ToothGrowth,2))
```

```
##     len supp dose
## 1   4.2   VC  0.5
## 2  11.5   VC  0.5
## 59 29.4   OJ  2.0
## 60 23.0   OJ  2.0
```

2. Provide basic summary 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
```

```
library(ggplot2)

growth <- ggplot(ToothGrowth, aes(x = dose, y = len), col = supp) +
    geom_point(shape = 1, size = 3) +
    facet_wrap(~ supp)

growth
```

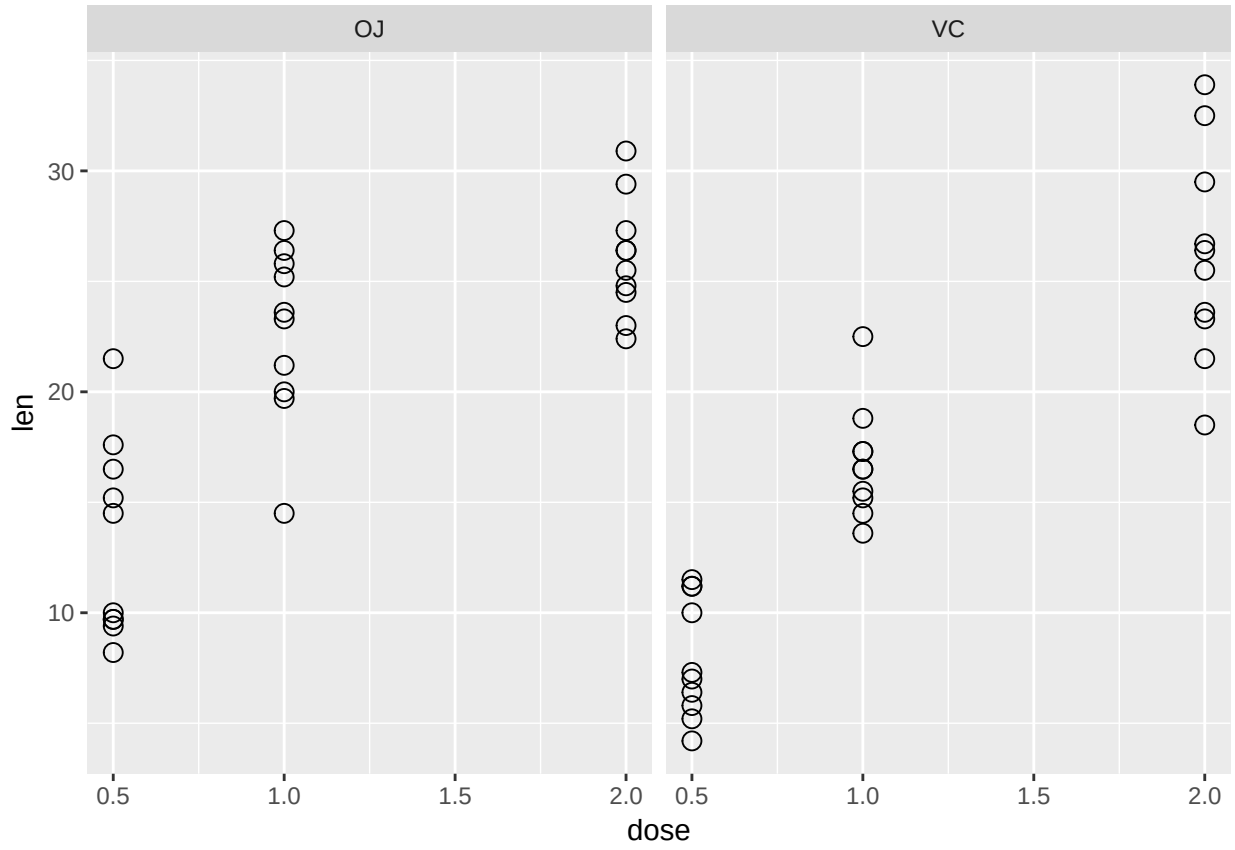

Figure 1. Response (len) based on dose (0.5, 1, 2)mg and supplement type (OJ, VC) from 10 guinea pigs.

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

First, differences in len are evaluate for pairs of dose levels = {0.5, 1, 2} mg. The data are paired by dose and so three paired t-test are performed:

t-test Index	*DoseA*	*DoseB*
1.	0.5	1
2.	0.5	2
3.	1	2

Also, group variances are assumed to be unequal as they are unknown. Assuming unequal variance results in wider confidence intervals than equal variance, which is more conservative. The 95% confidence intervals (CIs) are:

PartI : *Dose*

```
lenDose1 <- subset(ToothGrowth, dose %in% c(0.5,1))
lenDose2 <- subset(ToothGrowth, dose %in% c(0.5,2))
lenDose3 <- subset(ToothGrowth, dose %in% c(1,2))

rbind(
    t.test(len ~ dose, paired = TRUE, var.equal = FALSE, data = lenDose1)$conf,
    t.test(len ~ dose, paired = TRUE, var.equal = FALSE, data = lenDose2)$conf,
    t.test(len ~ dose, paired = TRUE, var.equal = FALSE, data = lenDose3)$conf
)
```

```
##            [,1]       [,2]
## [1,] -11.872879  -6.387121
## [2,] -18.367198 -12.622802
## [3,]  -9.258186  -3.471814
```

As the 95% CIs are entirely below zero, these t-tests suggest the smaller doses result in less tooth growth; or, increasing the dose increases tooth growth. Additionally, the p-values are small ($<< 0.05$), strongly suggesting more tooth growth with higher doses.

```
rbind(
    t.test(len ~ dose, paired = TRUE, var.equal = FALSE, data = lenDose1)$p.value,
    t.test(len ~ dose, paired = TRUE, var.equal = FALSE, data = lenDose2)$p.value,
    t.test(len ~ dose, paired = TRUE, var.equal = FALSE, data = lenDose3)$p.value
)
```

```
##              [,1]
## [1,] 1.225437e-06
## [2,] 7.190255e-10
## [3,] 1.934186e-04
```

PartII : *Supplement*

Second, a paired t-test is performed to evaluate a difference in tooth growth (len) between the two supplments (supp = {OJ, VC}). Again, unequal variance is assumed to be conservative.

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

```
##
##  Paired t-test
##
## data:  len by supp
## t = 3.3026, df = 29, p-value = 0.00255
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.408659 5.991341
## sample estimates:
## mean of the differences
##                     3.7
```

The 95% CI = (1.409, 5.991) suggests the supplment OJ provides more tooth growth than supplment VC. This is corroborated by a small p-value = 0.00255.

4. Conclusions

Confidence testing while varying dosage results that an increase in dosage from .5, 1, to 2 is proportianal to longer tooth. However, with a p-value of 0.06 and having zero in the confidence interval means we can not reject the null hypothesis that different supplement types have no effect on tooth length.