Inference Course Project - Part B

Joe Nguyen

22 Nov 2015

In this project, we use simulation to explore inference and do some simple inferential data analysis. The project consists of two parts:

- A simulation exercise
- Basic inferential data analysis

B. Inference on *ToothGrowth* Data

Information about the dataset:

The response is the length of odontoblasts (teeth) in each of 10 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 or ascorbic acid).

1. Load the *ToothGrowth* data

Explore the dataset.

```
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 a 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
```

```r
library(ggplot2)
tg <- ggplot(ToothGrowth, aes(x = dose, y = len), col = supp) +
    geom_point(shape = 1, size = 3) +
    facet_wrap(~ supp)
tg
```

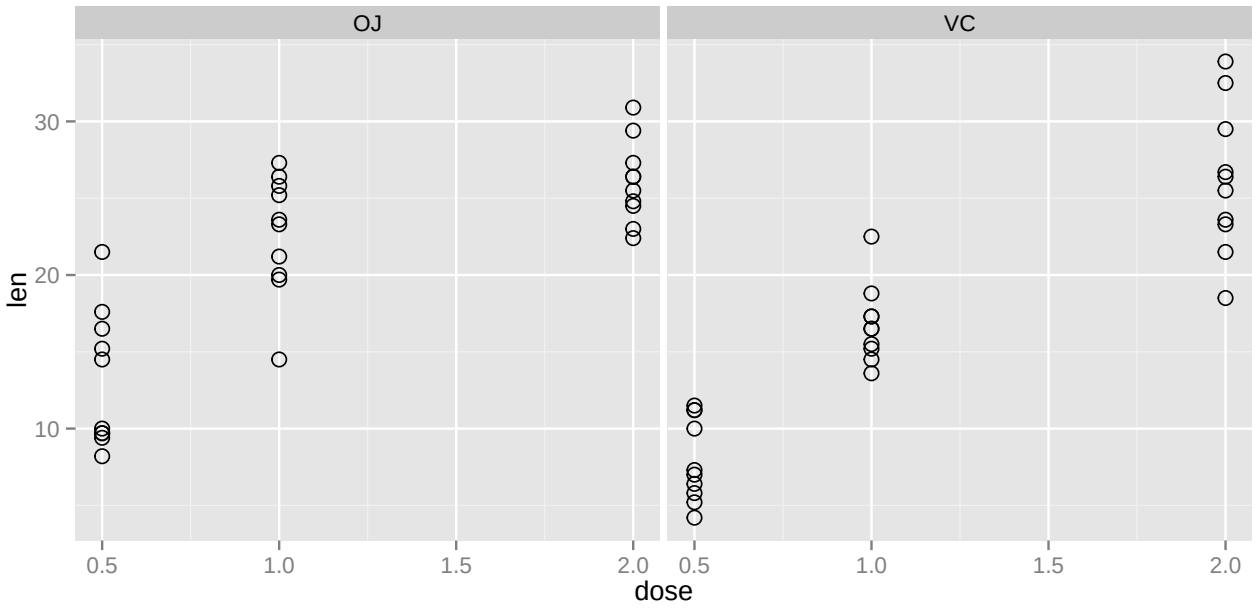

Figure 1. Response (len) based on dose (0.5, 1, 2)mg and supplement type (OJ, VC) from 10 guinea pigs.

3. Compare Tooth Growth

Tooth growth (len) is compared by dose and supplement type (supp). T-tests are used to determine the difference between pairs of dose and the supp pair. We use t-tests because the response (len) appears to by roughly symmetric and mound shaped in Fig. 1.

i. Dose 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	Dose A	Dose B
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:

```r
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
```

ii. 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.