STATISTICAL INFERENCE COURSE PROJECT: PART 2- ANALYZING TOOLGROWTH DATA

Lawrence Lau

Overview

This report provides basic data exploratory and statistical analysis of the ToothGrowth dataset. The dataset's infosheet (http://www.inside-r.org/r-doc/datasets/ToothGrowth) states ToothGrowth measures the length (len) of teeth in each of 30 guinea pigs after three dose (dose) levels of Vitaminc C (.5, 1, and 2 mg) with each of two delivery methods (supp), Orange Juice (OJ) and Ascorbic Acid (VC).

1.) Loading and Exploring Data

Loading and viewing the structure of the ToolGrowth 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 ...
```

We see there are 60 observations (rows) of 3 variables (columns), 2 numeric (len and dose) and 1 factor(supp).

Viewing a summary of ToothGrowth.

```
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.2   OJ:30   Min.   :0.50
##  1st Qu.:13.1   VC:30   1st Qu.:0.50
##  Median :19.2           Median :1.00
##  Mean   :18.8           Mean   :1.17
##  3rd Qu.:25.3           3rd Qu.:2.00
##  Max.   :33.9           Max.   :2.00
```

Viewing the first few lines of ToothGrowth.

```
head(ToothGrowth, n=2L)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
```

The three variables are length of teeth (len), dose levels of Vitamin C (dose), and delivery method (supp).

Let's visualize the data. (Reference Appendix for Fig 1)

```
library(ggplot2)
g <- ggplot(ToothGrowth, aes(x = dose, y = len, color = supp))
g <- g + geom_point()
g <- g + stat_summary(aes(group = 1), geom = "line", fun.y = mean, size = 1, col = "black")
g <- g + facet_grid(. ~ supp)
g <- g + ggtitle("Fig 1: ToothGrowth, length vs dose per supplement")
g
```

Fig 1 appears to show that length increases as more Vitamin C is given. OJ doses of .5 and 1.0 seem to be the more effective delivery method.

2.) Basic Summary of Data

```
library("lattice")
a <- aggregate(ToothGrowth$len, list(ToothGrowth$dose,ToothGrowth$supp),FUN=function(x) c(m
ean = mean(x), sd = sd(x)))
colnames(a)[c(1:3)] <- c("Dose", "Method", "Length")
a
```

```
##   Dose Method Length.mean Length.sd
## 1  0.5     OJ      13.230     4.460
## 2  1.0     OJ      22.700     3.911
## 3  2.0     OJ      26.060     2.655
## 4  0.5     VC       7.980     2.747
## 5  1.0     VC      16.770     2.515
## 6  2.0     VC      26.140     4.798
```

It appears as if OJ is more effective in the .5 and 1.0 dosage amounts. But let's find out if that's really the case.

3.) Compare tooth growth by supp and dose using confidence intervals and hypothesis tests

H0: Tooth growth is uneffected by supp type. (OJ mean = VC mean) Ha: Tooth growth is effected by supp type. (OJ mean != VC mean)

```
attach(ToothGrowth)
t.test(len ~ supp)$conf
```

```
## [1] -0.171  7.571
## attr(,"conf.level")
## [1] 0.95
```

The result's inconclusive, since the 95% confidence interval includes 0. However…

```
t.test(len ~ supp, conf.level=.90)$conf
```

```
## [1] 0.4683 6.9317
## attr(,"conf.level")
## [1] 0.9
```

…with a 90% confidence interval we do see that OJ is more effective than VC.

Let's see if different dosages had an effect on length. First, .5 vs 1.0.

```
Dose.5v1 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
t.test(len ~ dose, paired = FALSE, var.equal=FALSE, data=Dose.5v1)$conf
```

```
## [1] -11.984  -6.276
## attr(,"conf.level")
## [1] 0.95
```

We see a dosage of 1.0 absolutely results in a higher mean length than 0.5.

Next, 1.0 vs 2.0.

```
Dose1v2 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
t.test(len ~ dose, paired = FALSE, var.equal=FALSE, data=Dose1v2)$conf
```

```
## [1] -8.996 -3.734
## attr(,"conf.level")
## [1] 0.95
```

Here, a dosage of 2.0 absolutely results in a higher mean length than 1.0.

4.) Conclusions and Assumptions

- We have a 90% confidence interval that Orange Juice is a more effective mode of supplement for Vitamin C than Ascorbic Acid. Extrapolated to a 95% confidence interval though and the results are inconclusive.
- The higher the dose of Vitamin C, the longer the resulting length of tooth.
- These conclusions assume a constant but different variance in all groups. It is also assumed the data are not paired, since no two teeth, even in the same mouth, grow to the same exact length.

Appendix: Fig 1

```
library(ggplot2)
g <- ggplot(ToothGrowth, aes(x = dose, y = len, color = supp))
g <- g + geom_point()
g <- g + stat_summary(aes(group = 1), geom = "line", fun.y = mean, size = 1, col = "black")
g <- g + facet_grid(. ~ supp)
g <- g + ggtitle("Fig 1: ToothGrowth, length vs dose per supplement")
g
```