Coursera Statistical Inference Assignment 2

Anthony Iannolo

4/26/2018

Introduction

Within the R Library is the “ToothGrowth” data set. The tooth growth data set includes observations of guinea pigs who were given three Vitamin C dosage levels (0.5-min, 1-median, and 2-max mg) against two delivery methods of Orange Juice (OJ) or Vitamin C (VC).

The assignment consists of four key parts:

1: Load the ToothGrowth Data and perform some basic exploratory data analysis

2: Provide a basic summary of the data.

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

4: State your conclusions and assumptions needed for your conclsions.

Key Assumptions for this exercise:

- Tooth growth samples are random and Independent Identically Distributed (iid)

- Tooth growth follows the normal distribution

Part 1 - Load ToothGrowth Data and Perform Basic exploratory data analysis

- Key observations from our exploratory analysis is we have 60 observations of three variables: len (tooth length), supp (supplement type) and dose. Tooth length appears to have wide variation in size. There are two types of supplements used (OJ or VC) and they are split 30-30. We can also see the variation in dosages of 0.5, 1.0 and 2.0 mg.

Display sample of observations

```
##    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
```

Summarize structure of the DataSet

```
## '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 ...
```

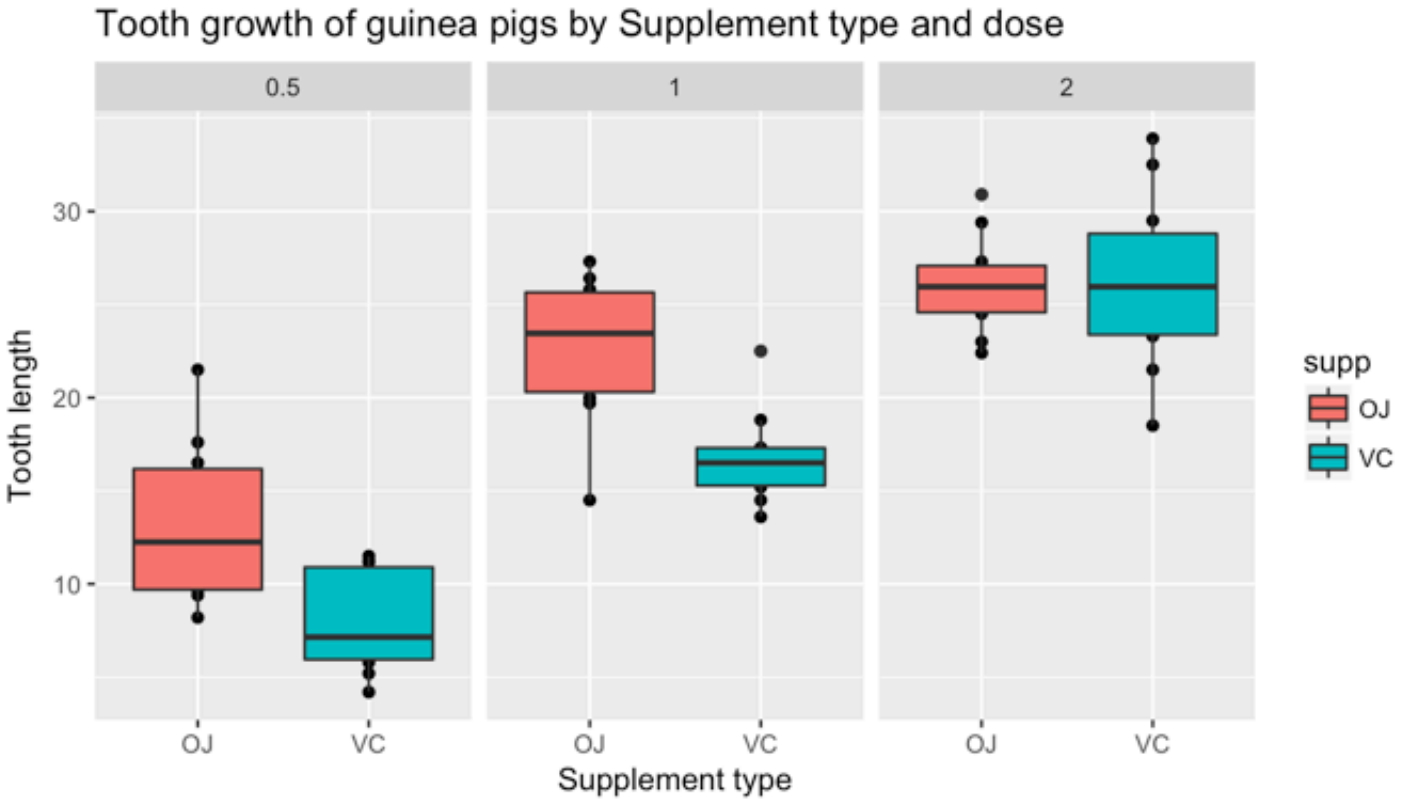
Produce summary statistics for the dataset

```
##       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
```

Plot the length of ToothGrowth by supplement and dosage

- Key observations from plot indicate it appears as dosage increases, generally the the tooth length also increases. At lower doses of 0.5 and 1.0, OJ appears to drive more tooth growth but at the 2.0 level the supplements are similar. OJ and VC seem to both drive tooth growth with OJ appearing to be a better supplement.

Next we'll use Hypothesis Testing to furhter investigate the significance of the observations.



Conduct Hypothesis Testing to compare tooth growth by supplement OJ Vs VC

Let the null hypothesis be there is no difference in tooth growth supplement OJ and VC.

- Null Hypothesis is Variable len of OJ = Varaible len of VC

Let the alternate hypothesis be there is more tooth growth when using supplement OJ than VC.

- Alternate Hypothesis is Variable len of OJ > Varaible len of VC

For Hypothesis Test first subset data into tooth growth by Supplement Type

Next perform a one-tailded independent t-test with unequal variance

- Key observation from the test is p is less than 0.05 indicating Null Hypothesis is not true and Alternative is true that OJ is more effective than VC.

```
t.test(OJsubset, VCsubset, alternative = "greater", paired = FALSE, var.equal = FALSE
, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJsubset and VCsubset
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

Conduct Hypothesis Testing to compare tooth growth by dosage of supplement

- Let the null hypothesis be there is no difference in tooth growth between dosages.

- Let alternate hypothesis be there is increased tooth growth at larger dosages.

For Hypothesis Test first subset data into growth by level of dosage

Next perform a one-tailded independent t-test with unequal variance

- t-test will analyze tooth growth between 0.5 mg vs 1 mg dosage

- Key observations from the two following tests are p is less than 0.05 indicating Null Hypothesis is not true and Alternative is true that there is increased growth at larger dosages.

```
t.test(subset05, subset1, alternative = "less", paired = FALSE, var.equal = FALSE, co
nf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  subset05 and subset1
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.753323
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

Next perform another one-tailded independent t-test with unequal variance

- t-test will analyze tooth growth between 1.0 mg vs 2.0 mg dosage

```
t.test(subset1, subset2, alternative = "less", paired = FALSE, var.equal = FALSE, con
f.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  subset1 and subset2
## t = -4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -4.17387
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

Compare OJ and VC tooth growth at dosage of 2.0 mg

We had observed that tooth growth is very similar at 2.0 mg and does this observation hold with two-sided t-test at 0.95 confidence level.

- Null Hypothesis is Variable len of OJ = Varaible len of VC at same dosage of 2.0 mg

- Alternate Hypothesis is that Variable len of OJ is significantly different than Varaible len of VC

For Hypothesis Test first subset data of tooth growth by Supplement Type where dose = 2.0 mg

Next perform a Two-tailed independent t-test with unequal variance.

- t-test will analyze tooth growth from OJ and VC at 2 mg dosage

- Key observations from the following two-sided t-test is t-statistic appears in center of 95% confidence interval and p value greater than 0.95. So we can't reject the Null Hypothesis that OJ and VC have equal tooth growth at 2.0 mg dosage.

```
##
##  Welch Two Sample t-test
##
## data:  OJsubset2 and VCsubset2
## 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
```

Summary

- Initial exploratory analysis indicated OJ could be a better supplement than VC and at higher dosage levels both supplements drove higher tooth growth. In addition, indications were that OJ and VC drove similar tooth growth at 2.0 mg.

By using Hypothesis Testing with t-test analysis showed the exploratory indications are statistically significant at 95% confidence interval.

Appendix including Code

- Load ToothGrowth Data and Perform Basic exploratory data analysis

```
library (ggplot2)
```

Load the R Toothgrowth DataSet, Display Observations and Summary Statistics

```
data(ToothGrowth)
head(ToothGrowth)
str(ToothGrowth)
summary(ToothGrowth)
```

Plot the length of ToothGrowth by supplement and dosage

```
qplot(supp, len, data=ToothGrowth, facets=~dose,
      main="Tooth growth of guinea pigs by Supplement type and dose",
      xlab="Supplement type",
      ylab="Tooth length") +
  geom_boxplot(aes(fill = supp))
```

- Conduct Hypothesis Testing to compare tooth growth by supplement

Create a subset of tooth growth by Supplement Type

```
OJsubset <- ToothGrowth$len[ToothGrowth$supp == 'OJ']
VCsubset <- ToothGrowth$len[ToothGrowth$supp == 'VC']
```

Next perform a one-tailded independent t-test with unequal variance

```
t.test(OJsubset, VCsubset, alternative = "greater", paired = FALSE, var.equal = FALSE
, conf.level = 0.95)
```

- Conduct Hypothesis Testing to compare tooth growth by dosage of supplement

Create subsets of tooth growth by level of dosage

```
subset05 <- ToothGrowth$len[ToothGrowth$dose == 0.5]
subset1 <- ToothGrowth$len[ToothGrowth$dose == 1]
subset2 <- ToothGrowth$len[ToothGrowth$dose == 2]
```

- Next perform a one-tailded independent t-test with unequal variance

t-test between 0.5 mg vs 1 mg dosage

```
t.test(subset05, subset1, alternative = "less", paired = FALSE, var.equal = FALSE, co
nf.level = 0.95)
```

Next perform a one-tailded independent t-test with unequal variance

t-test between 1.0 mg vs 2.0 mg dosage

```
t.test(subset1, subset2, alternative = "less", paired = FALSE, var.equal = FALSE, con
f.level = 0.95)
```

- Compare OJ and VC tooth growth at dosage of 2.0 mg

Observed that tooth growth is very similar at 2.0 mg and does this observation hold with two-sided

t-test at 0.95 confidence level

Create a subset of tooth growth by Supplement Type where dose = 2mg

```
OJsubset2 <- ToothGrowth$len[ToothGrowth$supp == 'OJ' & ToothGrowth$dose == 2]
VCsubset2 <- ToothGrowth$len[ToothGrowth$supp == 'VC' & ToothGrowth$dose == 2]
```

Two-tailed independent t-test with unequal variance.

```
t.test(OJsubset2, VCsubset2, alternative = "two.sided", paired = FALSE, var.equal = F
ALSE, conf.level = 0.95)
```