Statistical Inference Course Project – Part 2

Author: Venkat Ram Rao

Overview: Analyze the ToothGrowth dataset in R.

Exploratory Analysis and summary: The Toothgrowth dataset consists of a 60x3 dataframe comprising of 60 observations of tooth length in Guinea Pigs by Supplement and dosage.

dim(ToothGrowth)

    60 3

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 ...

There are two supplements – "OJ" and "VC" and 3 dosage levels of 0.5, 1.0 and 2.0.

Data can be split either by Dosage or by Supplement

By Supplement – "OJ" & "VC" (R code in Appendix Section 1a):

The teeth lengths with OJ has a mean of 20.66333 and a standard deviation of 6.605561

The teeth lengths with VC has a mean of 16.96333 and a standard deviation of 8.266029

By Dosage – "0.5" , "1.0", "2.0" (R code in Appendix Section 1b)::

The dosage = 0.5 group has a mean of 10.605 and a standard deviation of 4.499763

The dosage = 1.0 group has a mean of 19.735 and a standard deviation of 4.415436

The dosage = 2.0 group has a mean of 26.1 and a standard deviation of 3.77415

HYPOTHESIS 1: Mean Tooth Growth is greater with OJ than VC (i.e. OJ is a more effective supplement):

This can be tested by a one sided t test: (full R code in Appendix Section 2)

t.test(OJ, VC, alternative = "greater", paired = FALSE, conf.level = 0.95)

This yields a P-Value of **0.03032.** The t- statistic is **1.9153** (df = **55.309**) The 95% confidence interval is **0.4682687 – Infinity –** vs a difference in means of **1.66**

Conclusion: There is a 97% chance that the mean tooth growth with the supplement OJ is better that the supplement VC. For an 95% confidence we would accept the Hypothesis.

HYPOTHESIS 2: Mean Tooth Growth is greater with A dosage of 2 than 1 regardless of supplement *: (full R code in Appendix Section 3)*

One sided test: t.test(two, one, alternative = "greater", paired = FALSE, conf.level = 0.95)

This yields a P-Value of **9.532e-06.** The t- statistic is **4.9005** (df = **37.101**) The 95% confidence interval is **4.17387– Infinity –** vs a difference in means of **6.365**

Conclusion: There is very likely that the mean tooth growth with the dosage of 2 is better than the dosage of 1. For an 95% confidence we would accept the Hypothesis.

HYPOTHESIS 3: Mean Tooth Growth is greater with A dosage of 1 than 0.5 regardless of supplement *: (full R code in Appendix Section 4)*

One sided test: t.test(one, half, alternative = "greater", paired = FALSE, conf.level = 0.95)

This yields a P-Value of **6.342e-08.** The t- statistic is **6.4766** (df = **37.986**) The 95% confidence interval is **6.753323– Infinity–** vs a difference in means of **9.13**

Conclusion: There is very likely that the mean tooth growth with the dosage of 1 is better than the dosage of 0.5. For an 95% confidence we would accept the Hypothesis.

Note: A similar test can be run for a dosage of **2** vs **0.5** but the above two indicate that a dosage of 2 would be more effective than 0.5

Assumptions:

The key assumption here is that the samples are representative of the entire population and were chosen appropriately.

We are also assuming that they test subjects were relatively homogeneous and extraneous factors were controlled

APPENDIX

1) **Break up the ToothGrowth Dataset and calculate Mean and Standard Deviation**

a) By Supplement:

```
OJ = ToothGrowth$len[ToothGrowth$supp == 'OJ']
VC = ToothGrowth$len[ToothGrowth$supp == 'VC']
mean(OJ)

## [1] 20.66333

sd(OJ)

## [1] 6.605561

mean(VC)

## [1] 16.96333

sd(VC)

## [1] 8.266029
```

b) By Dosage:

```
half = ToothGrowth$len[ToothGrowth$dose == 0.5]
one = ToothGrowth$len[ToothGrowth$dose == 1]
two = ToothGrowth$len[ToothGrowth$dose == 2]
mean(two)

## [1] 26.1

sd(two)

## [1] 3.77415

mean(one)

## [1] 19.735

sd(one)

## [1] 4.415436

mean(half)

## [1] 10.605

sd(half)

## [1] 4.499763
```

2) **One sided T test – OJ is better than VC for tooth growth**

```r
t.test(OJ, VC, alternative = "greater", paired = FALSE, conf.level = 0.95)
##  Welch Two Sample t-test
## 
## data:  OJ and VC
## 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
```

3) **One sided T test – Dosage of 2 is better than dosage of 1 for tooth growth**

```r
t.test(two, one, alternative = "greater", paired = FALSE, conf.level = 0.95)
##  Welch Two Sample t-test
## 
## data:  two and one
## t = 4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  4.17387     Inf
## sample estimates:
## mean of x mean of y 
##    26.100    19.735
```

4) **One sided T test – Dosage of 1 is better than dosage of 0.5 for tooth growth**

```r
t.test(one, half, alternative = "greater", paired = FALSE, conf.level = 0.95)
##  Welch Two Sample t-test
## 
## data:  one and half
## t = 6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.753323      Inf
## sample estimates:
## mean of x mean of y 
##    19.735    10.605
```