Analyzing Tooth growth through Statistical Inference

Avranil Ghosh

Sunday, January 25, 2015

Overview:

In this project we will analyse Tooth growth data using T-test to identify the different growth behaviour of Tooth under the effect of different doses and supp.

Loading data and analyzing summary statistics.

Here we load the data from ToothGrowth dataset in R and analyse the summary in the whole dataset as well as for the different doses and supp.

The summary for the 2 supplementary are:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.20   15.52   22.70   20.66   25.72   30.90

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20   11.20   16.50   16.96   23.10   33.90
```

The above summary illustrates that the mean for supplementary "OJ" is more than the mean growth with supplementry "VC". However the distribution with "VC" is more spread out and wider as compared to "OJ" as can be been from max and min values from the respective statistics above.

The summary for the 3 doses are:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.200   7.225   9.850  10.600  12.250  21.500

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.60   16.25   19.25   19.74   23.38   27.30

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.50   23.52   25.95   26.10   27.83   33.90
```

The above summary illustrates that the mean Tooth growth increases across doses 0.5,1 and 2 with minimum when dose 0.5 is administered and maximum when dose 2 is administered.

Hypothesis testing for growth with different supplementaries.

Here we apply t test to identify if the difference in Tooth growth under the effect of the 2 different supplementaries "oJ" nd "VC" are statistically different at 95% confidence level.

Assumption: Variances for the two datasets given different supplements are unequal and the datasets are not paired.

Lets formulate the below hypothesis:

H0 : There is no significant change in growth when supplement "OJ" over "VC".

H1 : There is a significant change in growth in the 95% confidence when supplement "OJ" is adminsitered over "VC".

The T test summary is:

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

The summary shows that T value for the sample is lower the T-statistic at 95% given by 2.0038004.Hence the null hypothesis H0 cannot be rejected.This concluson is further strengthened by the p-value statistics which is more than 0.05 and suggests that null hypothesis cannot be rejected t 95% confidence level. Hence There is no significant change in growth when supplement "OJ" is administered over "VC".

Hypothesis testing for growth with different doses.

Here we apply t test to identify if the difference in Tooth growth under the effect of the 3 different doses "0.5", "1" and "2" are statistically different at 95% confidence level.

Assumption: Variances for the 3 datasets given different doses are unequal and the datasets are not paired.

Lets formulate the below hypothesis:

H0 : There is no significant change in growth for the 3 different doses "0.5", "1" and "2".

H1 : There is a significant change in growth in the 95% confidence when dose "1" is adminsitered over "0.5".

H2 : There is a significant change in growth in the 95% confidence when dose "2" is adminsitered over "0.5".

H3 : There is a significant change in growth in the 95% confidence when dose "2" is adminsitered over "1".

The T test summary for H1:

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

The summary shows that the absolute T value for the sample is higher the T-statistic at 95% given by 2.0245696.Hence the null hypothesis H0 can be rejected.This concluson is further strengthened by the p-value statistics which is less than 0.05 and suggests that null hypothesis can be rejected t 95% confidence level. Hence There is a significant change in growth in the 95% confidence when dose "1" is adminsitered over "0.5".

The T test summary for H2:

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

The summary shows that the absolute T value for the sample is higher the T-statistic at 95% given by 2.0263778.Hence the null hypothesis H0 can be rejected.This concluson is further strengthened by the p-value statistics which is less than 0.05 and suggests that null hypothesis can be rejected t 95% confidence level. Hence There is a significant change in growth in the 95% confidence when dose "2" is adminsitered over "0.5".

The T test summary for H3:

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

The summary shows that the absolute T value for the sample is higher the T-statistic at 95% given by 2.0260081.Hence the null hypothesis H0 can be rejected.This concluson is further strengthened by the p-value statistics which is less than 0.05 and suggests that null hypothesis can be rejected t 95% confidence level. Hence There is a significant change in growth in the 95% confidence when dose "2" is adminsitered over "1".