Project: Using the t distribution in hypothesis testing

Vineet W. Singh

19 February 2018

Introduction

The t distribution is used to test statistics on sets of data when the sample sizes are small. This brief project demonstrates the use of the t distriution in finding confidence intervals and conducting hypothesis tests on the ToothGrowth (included with R) dataset and this has 60 observations.

Analysis

The ToothGrowth data set contains data regarding Vitamin C given as a supplement (in two forms) and the resulting growth of teeth in 60 guinea pigs.

This analysis will answer the following questions:
1) Do the teeth grow the same, for all dosage of supplements or do different doses of supplements result in different growth.
2) Do both the supplements result in the same rate of growth of teeth or is the growth rate different for each supplement.

The structure of the ToothGrowth dataset can be found using the str function.

```
## '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 need to partition the data according to the supplements concerned, either “VC” for Ascorbic Acid or “OJ” for orange juice. Normalising length by dose and
storing this as a new variable “rate” will help in the tests.

Summary Statistics for each group are:
Ascorbic Acid : Summary for 10 observations in each subgroup
For 0.5(mg/day) subgroup - Mean: 16, SD: 5.49;
For 1(mg/day) subgroup - Mean: 16.8, SD: 2.52;
For 2(mg/day) subgroup- Mean: 13.1, SD: 2.40.
Orange Juice : Summary for 10 observations in each subgroup
For 0.5(mg/day) subgroup - Mean: 26.5, SD: 8.92;
For 1(mg/day) subgroup - Mean: 22.7, SD: 3.91;
For 2(mg/day) subgroup - Mean:13.0, SD: 1.33.

An interesting plot (Appendix) is included showing the growth rate pattern for each supplement type.

From the summary & plot it can be observed that for each subgroup, the growth rates vary according to dosages given, with the maximum variabilty in the 0.5 mg group for both supplements and different growth rates for each supplement.

Assumptions

We will need to make some assumptions about the data before we carry out the hypothesis tests.
We can assume that the growth rate in each animal in each subgroup is independent. Further more we can assume that the variance within each of the 6 subgroups (supplement & dosage) is the same. However, it is probably not very wise to assume that the variance accross different subgroups is the same as, clearly the dosage levels are different across each subgroup in one supplement group.

Testing and Results

Null Hypothesis No. 1
Mean growth rates for the 0.5(mg/day) group and the 1(mg/day) group of the ascorbic acid set are the same.

Since we only have 10 measurements each subgroup, we need to make confidence intervals and do tests using the t distribution. Results of the t-test (Appendix: Test 1) to check the hypothesis are t-value: -0.42396, p-value $> .50$ and the confidence interval (CI) for the difference of means contains 0. **We cannot, therefore, reject null hypothesis no. 1.**

Null Hypothesis No. 2.
Mean growth rates for the 1(mg/day) group and the 2(mg/day) group of the ascorbic acid set are the same.

Result of the t.test for this hypothesis (Appendix Test 2) shows a high t-value of 3.3662, a low p value and a confidence interval where the difference is well above 1. **We can, therefore, reject the null hypothesis no. 2** and conclude that the means for the two subgroups are different.

Null hypothesis No. 3:
Mean growth rates for the 0.5(mg/day) group and the 1(mg/day) group of the orange juice set are the same.
Null hypothesis No. 4
Mean growth rates for the 1(mg/day) group and the 2(mg/day) group of the orange juice set are the same.

We find from the results of test 3 (Appendix: Test 3/4) that the t-value is above 1 but not very high, the CI of the difference contains 0 so we cannot rule out that the means might not be the same. Results of test number 4 tell a different story, however, the t-value is very large, the p-value is negligible and the confidence interval of the difference does not contain 0.
Using this argument, we cannot reject null hypothesis no. 3, but we can reject null hypothesis no. 4.

Based on the results of the 4 hypothesis tests above, groups of 0.5(mg/day) and 1(mg/day) each supplement types are merged and further tests are carried out for to test for growth between the two supplement types. There will be 3 further tests, one between the merged 0.5-1(mg/day) subgroup of each supplement type, one between the 2 (mg/day) subgroup of each supplement type and one test that compares all the data taken together for each supplement type.

Null Hypothesis No. 5
There is no difference in the Mean growth rates between 0.5-1 (mg/day) dosage sub-group of the Ascorbic Acid and Orange Juice supplement groups.

Refering to the results of test no. 5 (Appendix: Test 5), we have large negative t-value and this implies that the means for the two supplement types are not the same. The CI does not contain 0 and is entirely negative implying that, according to this test, the average value of tooth growth in the first group (ascorbic acid) will lie below the average value of the tooth growth in the second group (orange juice). **We can, therefore reject null hypothesis no. 5**

Null Hypothesis No. 6
There is no difference in the Mean growth rates between 2(mg/day) dosage sub-group of the Ascorbic Acid and Orange Juice supplement groups.

The result for test no. 6 (Appendix: Test 6) is interesting. This can also be inferred from the plot, that the average tooth growth rate for each supplement type begins to get close to one another once the dose starts getting close to the 2(mg/day) level. The t-value is small, p-value large, The CI of the difference of means is small and does contain 0. We can conclude that based on evidence provided by this data there is no difference in the mean growth rates of the 2(mg/day) subgroup of each supplement group. **We cannot, therefore, reject null hyposthesis no. 6.**

It would be interesting to make one final test using all of the 30 observations of each supplement group.

Null hyposthesis no. 7.
There is no difference in the Mean growth rates between the Ascorbic Acid and Orange Juice supplement groups.

The result of test 7 shows that overall if we take normalised growth rates for all observations in the two groups, the growth rates for each supplement type is different as the t-value is large, p-value is negligible and the CI of the difference of the means does not contain 0. It is safe to conclude that the mean growth rates for each supplement type is different.
We can, therefore reject the null hypothesis no. 7.

Conclusion

After analysing the data and proposing seven hypothesis we can conclude that:
1) Average growth rates of teeth for the animals that were given 0.5 to 1(mg/day) of ascorbic acid as a supplement are similar. The average growth rate of the teeth of animals given 2(mg/day) ascorbic acid as a supplement is not similar to the average growth rates of the group of animals given 0.5-1(mg/day) ascorbic acid as a supplement.
2) Average growth rates of teeth for the animals that were given 0.5 to 1(mg/day) of orange juice as a supplement are similar. The average growth rate of the teeth of animals given 2(mg/day) orange juice as a supplement is not similar to the average growth rates of the group of animals given 0.5-1(mg/day) orange juice as a supplement.
3) The average growth rate of teeth of animals given 0.5-1(mg/day) ascorbic acid supplement is not similar to the average growth rate of teeth of animals given 0.5-1(mg/day) orange juice supplement.
4) The average growth rate of teeth of animals given 2(mg/day) ascorbic acid supplement is similar to the average growth rate of teeth of animals given 0.5-1(mg/day) orange juice supplement.
5) Overall, if all data points are taken together (thereby ignoring conlcusion 4 above), the average growth rate of teeth of animals given ascorbic acid supplement is not similar to the average growth rate of teeth of animals given orange juice supplement.

Appendix

Ascorbic Acid : Summary

```
#calculate summaries for sub group
tg1 %>% group_by(dose) %>% summarise(mean=mean(rate),sd=sd(rate),n=n())
```

```
## # A tibble: 3 x 4
##    dose  mean    sd     n
##   <dbl> <dbl> <dbl> <int>
## 1 0.500  16.0  5.49    10
## 2 1.00   16.8  2.52    10
## 3 2.00   13.1  2.40    10
```

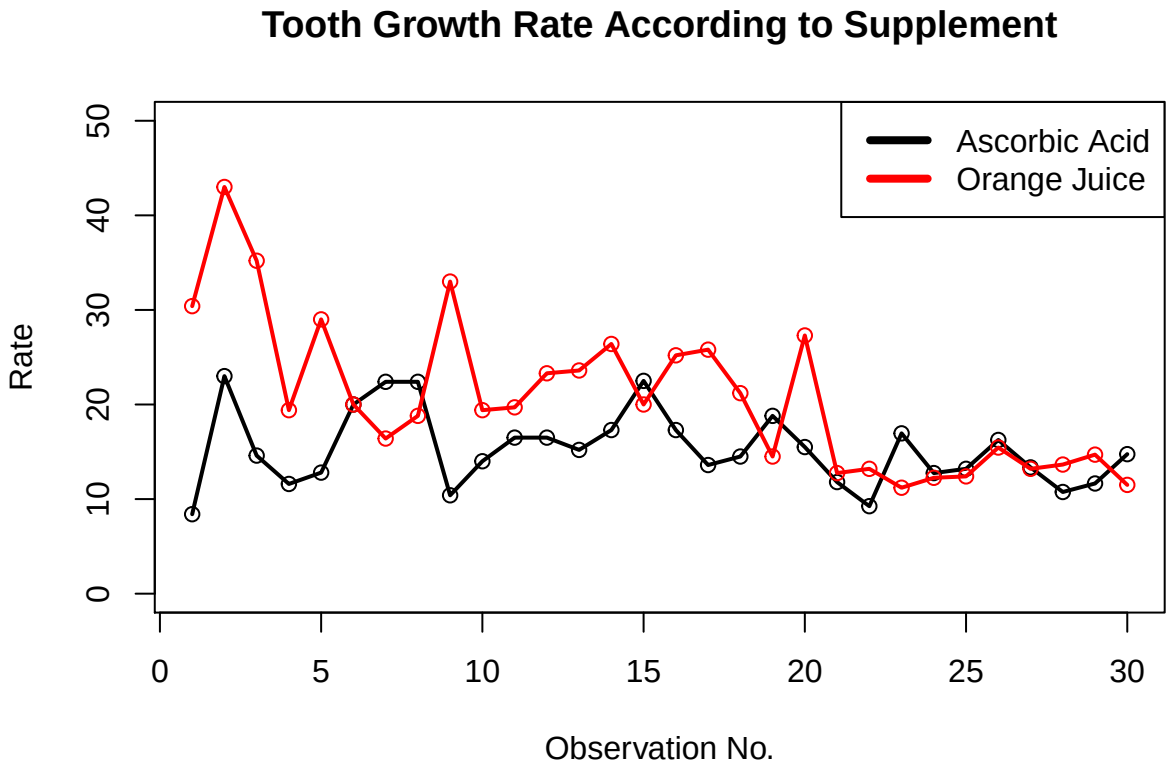
Orange Juice : Summary

```r
#calculate summaries for sub group
tg2 %>% group_by(dose) %>% summarise(mean=mean(rate),sd=sd(rate),n=n())
```

```
## # A tibble: 3 x 4
##    dose  mean    sd     n
##   <dbl> <dbl> <dbl> <int>
## 1 0.500  26.5  8.92    10
## 2 1.00   22.7  3.91    10
## 3 2.00   13.0  1.33    10
```

PLOT

```r
plot(x=row.names(tg1),y=tg1$rate,type="p",pch=1,ylab="Rate",
     main="Tooth Growth Rate According to Supplement",
     xlab="Observation No.", ylim=c(0,50),col=1)
lines(x=row.names(tg1),y=tg1$rate,type="l",lty=1,lwd=2,col=1)
points(x=row.names(tg2),y=tg2$rate,pch=1,col=2)
lines(x=row.names(tg2),y=tg2$rate,type="l",lty=1,lwd=2,col=2)
legend("topright", c("Ascorbic Acid","Orange Juice"),col=c(1,2),lwd=4)
```




```r
tg11<- tg1 %>% filter(dose==0.5)
tg12<- tg1 %>% filter(dose==1)
tg13<- tg1 %>% filter(dose==2)

tg21<- tg2 %>% filter(dose==0.5)
tg22<- tg2 %>% filter(dose==1)
tg23<- tg2 %>% filter(dose==2)
```

RESULTS OF TEST 1

```
t.test(tg11$rate,tg12$rate,paired=F,var.equal = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  tg11$rate and tg12$rate
## t = -0.42396, df = 12.615, p-value = 0.6787
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.950377  3.330377
## sample estimates:
## mean of x mean of y
##     15.96     16.77
```

RESULTS OF TEST 2

```
t.test(tg12$rate,tg13$rate,paired=F,var.equal = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  tg12$rate and tg13$rate
## t = 3.3662, df = 17.96, p-value = 0.003448
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.390398 6.009602
## sample estimates:
## mean of x mean of y
##     16.77     13.07
```

RESULTS OF TEST 3

```
t.test(tg21$rate,tg22$rate,paired=F,var.equal = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  tg21$rate and tg22$rate
## t = 1.2209, df = 12.337, p-value = 0.245
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.93001 10.45001
## sample estimates:
## mean of x mean of y
##     26.46     22.70
```

RESULTS OF TEST 4

```
t.test(tg22$rate,tg23$rate,paired=F,var.equal = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  tg22$rate and tg23$rate
## t = 7.404, df = 11.047, p-value = 1.321e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##   6.796869 12.543131
## sample estimates:
## mean of x mean of y
##     22.70     13.03
```

RESULTS OF TEST 5

```
#check rates and hypothesis test for groups according to dose & supplement
t.test(tg11$rate,tg21$rate,var.equal=F,paired=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  tg11$rate and tg21$rate
## t = -4.5183, df = 31.083, p-value = 8.455e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.922742  -4.507258
## sample estimates:
## mean of x mean of y
##    16.365    24.580
```

RESULTS OF TEST 6

```
#check rates and hypothesis test for groups according to dose & supplement
t.test(tg12$rate,tg22$rate,var.equal=F,paired=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  tg12$rate and tg22$rate
## 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:
##  -1.819035  1.899035
## sample estimates:
## mean of x mean of y
##     13.07     13.03
```

RESULTS OF TEST 7

```
#check rates and hypothesis test for groups according to supplement
t.test(tg1$rate,tg2$rate,var.equal=F,paired=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  tg1$rate and tg2$rate
## t = -3.3707, df = 42.61, p-value = 0.001603
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.732894 -2.193773
## sample estimates:
## mean of x mean of y
##  15.26667  20.73000
```