

Hypothesis Testing on ToothGrowth Data

Statistical Inference Project Part II, Class 6 in data science series

Ann Crawford

September 7, 2017

Dependencies

```
##install.packages("dplyr")
##install.packages("ggplot")
##install.packages("gridExtra")
library(ggplot2)
library(grid)
library(gridExtra)
```

```
# Clear our workspace
rm(list=ls())
```

Tooth Growth Data

The standard R data set, ToothGrowth, measures the effect of Vitamin C on tooth growth in guinea pigs. The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC)).

```
summary(ToothGrowth)
```

##	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

Hypothesis Testing using T Test.

Visually examine the sample data graphs (see Appednix). The graphs lend evidence to the t test assumptions. Based on the graphs the following Null hypothesis will be tested: 1. Supply using orange juice as the delivery method induces longer tooth growth. 2. Increase of vitamin C does induces longer tooth growth. * Assumptions for t tests: + Data includes trials from 60 independent guinea pigs. + Data are iid normal. + The distribution of the data is roughly symmetric and mound shaped

Testing the affect of supp, delivery method

Assume Null hypothesis the tooth growth is greater for OJ., Divide the data into two groups, OJ and VC.

P value is the probability of observing a test statistic as large as the one calculated assuming H is true.

```
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
library(magrittr) # for %$% extraction

OJ <- ToothGrowth %>% filter(supp=="OJ")
VC <- ToothGrowth %>% filter(supp=="VC")

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

##
## Welch Two Sample t-test
##
## data:  OJ$len and VC$len
## 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
```

Testing the affect of dose

There are 3 groups of doses, 0.5, 1, 2. Divide the data into 3 groups, OJ and VC and run two t tests. Test that more dosage leads to longer tooth growth.

```
low    <- ToothGrowth %>% filter(dose == 0.5)
medium <- ToothGrowth %>% filter(dose == 1)
high   <- ToothGrowth %>% filter(dose == 2)

t.test( high$len, low$len, alternative = "greater", paired = FALSE, conf.level = 0.95, Var.equal=FALSE)

##
## Welch Two Sample t-test
##
## data:  high$len and low$len
## t = 11.799, df = 36.883, p-value = 2.199e-14
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  13.27926      Inf
## sample estimates:
## mean of x mean of y
##   26.100    10.605
```

```
t.test(high$len, medium$len, alternative = "greater", paired = FALSE, conf.level = 0.95, Var.equal=FALSE)

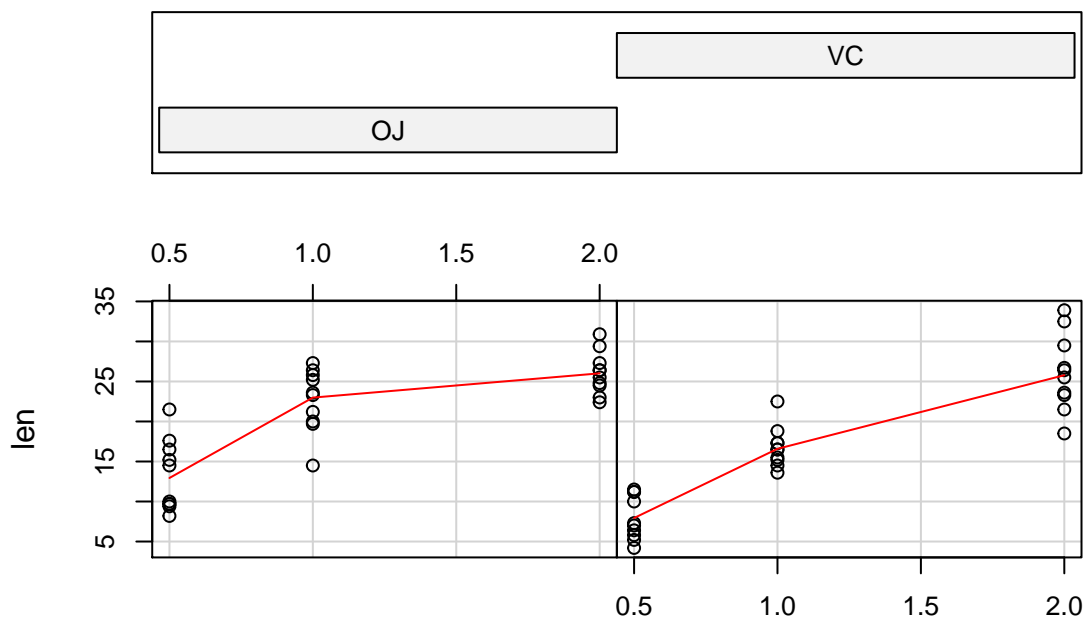
##
## Welch Two Sample t-test
##
## data: high$len and medium$len
## t = 4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  4.17387      Inf
## sample estimates:
## mean of x mean of y
##  26.100    19.735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

Conclusion

The p values and confidence intervals returned from the above tests return low p-values conclude: Accept Null hypotheses that mean tooth length of OJ is statistically significant greater than VC. Accept Null hypotheses that mean tooth length of higher doses is statistically significant greater than lower doses.

Appendix

Given : supp



ToothGrowth data: length vs dose, given type of supplement

```
## Warning: package 'ggplot2' was built under R version 3.3.3
```

```
## Warning: package 'gridExtra' was built under R version 3.3.3
##
## Attaching package: 'gridExtra'
## The following object is masked from 'package:dplyr':
##
##   combine
```

