Hypothesis Testing on ToothGrowth Data

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

Ann Crawford

September 7, 2017

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:gridExtra':
##
##     combine

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

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
```

Hypotesis 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 hypothesi 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

The violin graphs indicated varirence differs for each group.

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 probablity of observing a test statistic as large as the one calculated assuming H0 (null hypothesis) is true.

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

hoOJVCSupp <- t.test(OJ$len, VC$len, alternative = "greater", paired = FALSE,
conf.level = 0.95, var.equal = FALSE)
```

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.

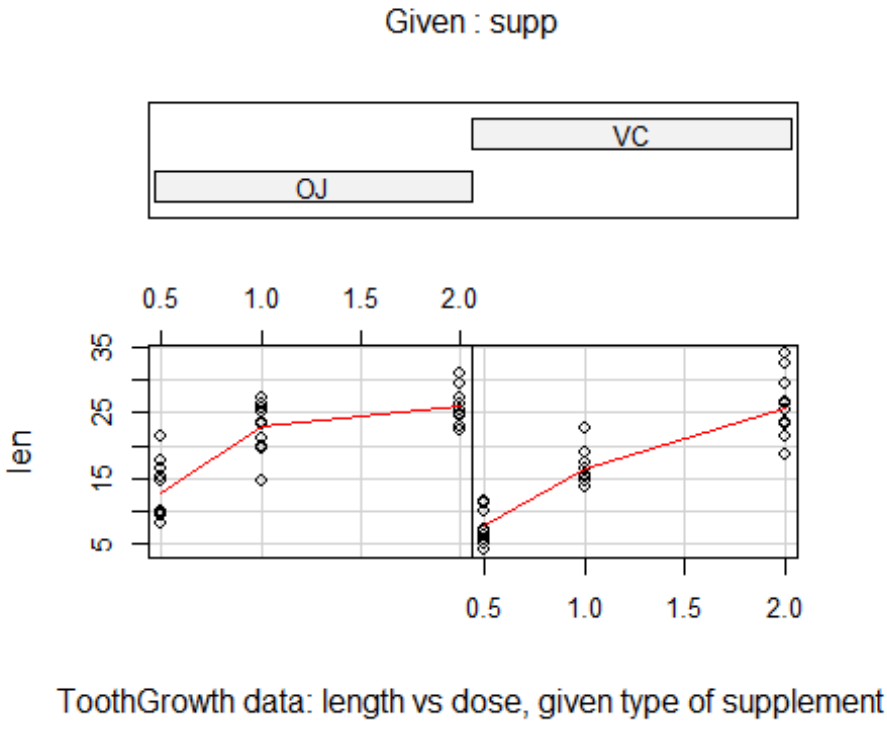
```
low     <- ToothGrowth %>% filter(dose == 0.5)
medium  <- ToothGrowth %>% filter(dose == 1)
high    <- ToothGrowth %>% filter(dose == 2)

h0highlowd <- t.test( high$len, low$len,alternative = "greater", paired =
FALSE, conf.level = 0.95,Var.equal=FALSE)
hohighmedd <- t.test(high$len, medium$len,alternative = "greater", paired =
FALSE, conf.level = 0.95,Var.equal=FALSE)
```

Conclusion

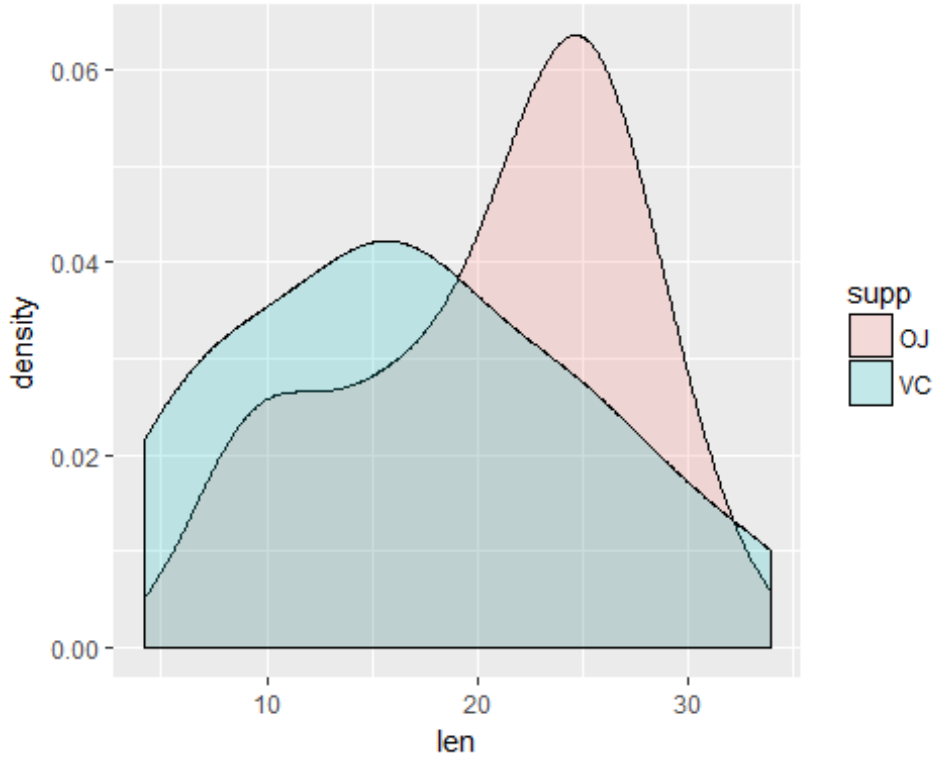
The p values and confidence intervals returned from the above tests return low p-values conclude: Accept Null hypothises that mean tooth length of OJ is statisically significant greater than VC. Accept Null hypothises that mean tooth lenght of higher doses is statistically significant greater than lower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

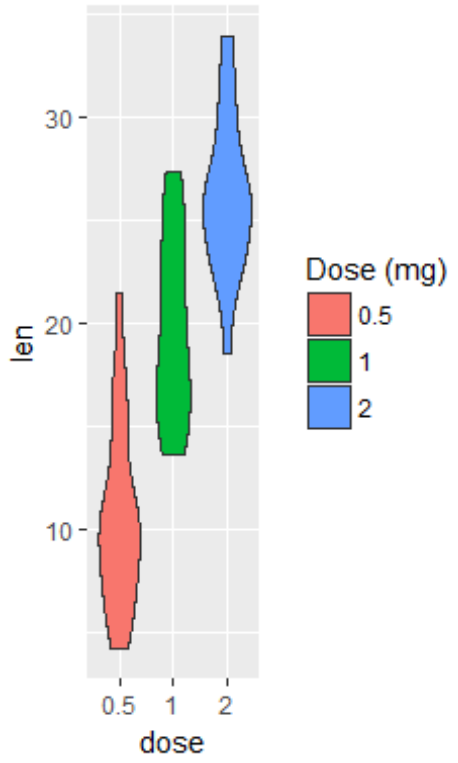
H0	P-Value	Conf. Interval	accept/reject
mean OJ - mean VC > 0	0.0303173	0.4682687,	accept
mean dose(2) - mean dose(0.5) > 0	2.198762510^{-14}	13.2792571,	accept
mean dose(2) - mean dose(1.0) > 0	9.532147610^{-6}	4.1738697,	accept

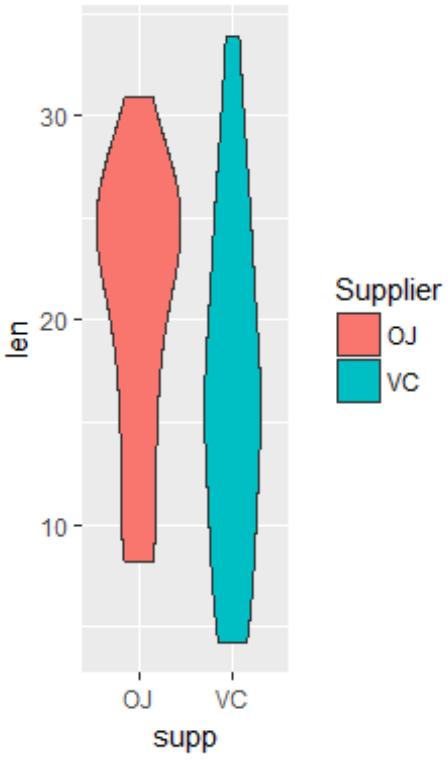
Appendix



ToothGrowth data: length vs dose, given type of supplement


0.06
0.04
0.02
0.00
density
10
20
30
len
supp
OJ
VC


30
20
10
len
0.5
1
2
dose
Dose (mg)
0.5
1
2


30
20
10
len
OJ
VC
supp
Supplier
OJ
VC