

Statistical Inference project 2

Balogun Stephen Taiye

November 22, 2015

Contents

Overview	1
Basic Exploratory Analysis	1
Conclusions and Assumptions	3
<i>Conclusions:</i>	3
<i>Assumptions:</i>	3

Overview

This second project aims to analyse the *Toothgrowth* data in R `datasets` package, perform some basic exploratory analysis on it, then run some statistical inference with respect to calculation of confidence interval/hypothesis testing

Basic Exploratory Analysis

Glancing through the data to have a general look at the data, check if there are missing entries and some other basic exploration.

```
Toothdata <- ToothGrowth ## loads the "toothgrowth" data in the datasets package
dim(Toothdata)           ## shows the number of rows and columns of the data
```

```
## [1] 60 3
```

```
summary(Toothdata) ## summarises the data
```

len	supp	dose
Min. : 4.20	OJ:30	Min. :0.500
1st Qu.:13.07	VC:30	1st Qu.:0.500
Median :19.25	NA	Median :1.000
Mean :18.81	NA	Mean :1.167
3rd Qu.:25.27	NA	3rd Qu.:2.000
Max. :33.90	NA	Max. :2.000

```
any(is.na(Toothdata)) ## checks if there are any uncompleted records in the data
```

```
## [1] FALSE
```

```
head(Toothdata)    ## checks the first few rows to give an idea how the data looks like
```

len	supp	dose
4.2	VC	0.5
11.5	VC	0.5
7.3	VC	0.5
5.8	VC	0.5
6.4	VC	0.5
10.0	VC	0.5

```
tail(Toothdata)    ## checks the last few rows to be sure that there are same number of entries for each
```

	len	supp	dose
55	24.8	OJ	2
56	30.9	OJ	2
57	26.4	OJ	2
58	27.3	OJ	2
59	29.4	OJ	2
60	23.0	OJ	2

Good, all columns are of equal length, no missing entries and the data format is appropriate for further analysis. Now that basic exploration has been done, we try to format the data to allow for our hypothesis testing, creating subset for each of the `dose`.

```
library(dplyr)    ## seems to be my favourite for getting and cleaning data
sub0.5 <- filter(Toothdata, dose == .5)  ## creates a subset of dose 0.5
sub1 <- filter(Toothdata, dose == 1.0)   ## creates a subset of dose 1.0
sub2 <- filter(Toothdata, dose == 2.0)   ## creates the subset with dose 2.0
```

Next, we run `t.test()` to compare the t-test for the two different types of supplements at the three doses given.

```
t0.5 <- t.test(len~supp, data = sub0.5)$conf.int  ## t.test at dose 0.5
t1.0 <- t.test(len~supp, data = sub1)$conf.int    ## t.test at dose 1.0
t2.0 <- t.test(len~supp, data = sub2)$conf.int    ## t.test at dose 2.0
rbind(t0.5, t1.0, t2.0)  ## for direct comparison of the different t.test() for the three doses
```

t0.5	1.719057	8.780943
t1.0	2.802148	9.057852
t2.0	-3.798071	3.638070

The formula above runs `t.test()` on each of the subsets, then binds them so they appear on different rows of the same column to allow for comparison.

Conclusions and Assumptions

Conclusions:

The result shows that the `t.test()` result for doses 0.5 and 1.0 does not include 0 while the result appears to be balanced at 0 when the dose is 2.0. This means that Supplement OJ have a better growth response on the teeth causing longer length compared to supplement VC at doses less than 2.0. At dose of 2.0, there does not seem to be a significant difference in the tooth length when either of the supplements is used.

Assumptions:

The analysis assumes that the variables are identical and independently distributed. that is, the results are from samples drawn randomly and the outcome of one does not influence the outcome of the other.