Statistical Inference Course Project Part 2: Inferential Data Analysis

Massimiliano Figini

October 9, 2016

Analysis of the ToothGrowth data in the R datasets package.
The data are about 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).

```r
# Basic settings and libraries
knitr::opts_chunk$set(echo = TRUE, tidy.opts=list(width.cutoff=60), tidy=TRUE)
library(pander)
library(ggplot2)
library(dplyr)
library(datasets)    # load the data
```

QUESTION 1 and 2: Load the ToothGrowth data, perform some basic exploratory data analyses and provide a basic summary of the data.

```r
# Show the table
pandoc.table(ToothGrowth)
```

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	VC	0.5
11.2	VC	0.5
11.2	VC	0.5
5.2	VC	0.5
7	VC	0.5
16.5	VC	1
16.5	VC	1
15.2	VC	1
17.3	VC	1
22.5	VC	1
17.3	VC	1
13.6	VC	1
14.5	VC	1
18.8	VC	1
15.5	VC	1
23.6	VC	2
18.5	VC	2
33.9	VC	2
25.5	VC	2

len	supp	dose
26.4	VC	2
32.5	VC	2
26.7	VC	2
21.5	VC	2
23.3	VC	2
29.5	VC	2
15.2	OJ	0.5
21.5	OJ	0.5
17.6	OJ	0.5
9.7	OJ	0.5
14.5	OJ	0.5
10	OJ	0.5
8.2	OJ	0.5
9.4	OJ	0.5
16.5	OJ	0.5
9.7	OJ	0.5
19.7	OJ	1
23.3	OJ	1
23.6	OJ	1
26.4	OJ	1
20	OJ	1
25.2	OJ	1
25.8	OJ	1
21.2	OJ	1
14.5	OJ	1
27.3	OJ	1
25.5	OJ	2
26.4	OJ	2
22.4	OJ	2
24.5	OJ	2
24.8	OJ	2
30.9	OJ	2
26.4	OJ	2
27.3	OJ	2
29.4	OJ	2
23	OJ	2

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)  # dose is factor
str(ToothGrowth)  # Table variables
```

```
## '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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

The variables are:
- len: the tooth length
- supp: the supplement type of vitamin C (VC is ascorbic acid, OJ is orange juice)
- dose: dose in milligrams/day of vitamin C

```
# Basic summary
summary(ToothGrowth)
```

```
##       len          supp        dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
# Means for supplement type and dose
TG <- tbl_df(ToothGrowth)
TG2 <- summarize(group_by(TG, supp, dose), mean_of_len = mean(len))
TG2
```

```
## Source: local data frame [6 x 3]
## Groups: supp [?]
##
##       supp   dose mean_of_len
##     <fctr> <fctr>       <dbl>
## 1      OJ    0.5       13.23
## 2      OJ      1       22.70
## 3      OJ      2       26.06
## 4      VC    0.5        7.98
## 5      VC      1       16.77
## 6      VC      2       26.14
```

```
# Basic exploratory graph
qplot(supp, data = TG2, geom = "bar", weight = mean_of_len, facets = . ~
    dose, fill = supp, main = "Mean of tooth lenght for supplement type and quantity",
    xlab = "", ylab = "Tooth length")
```
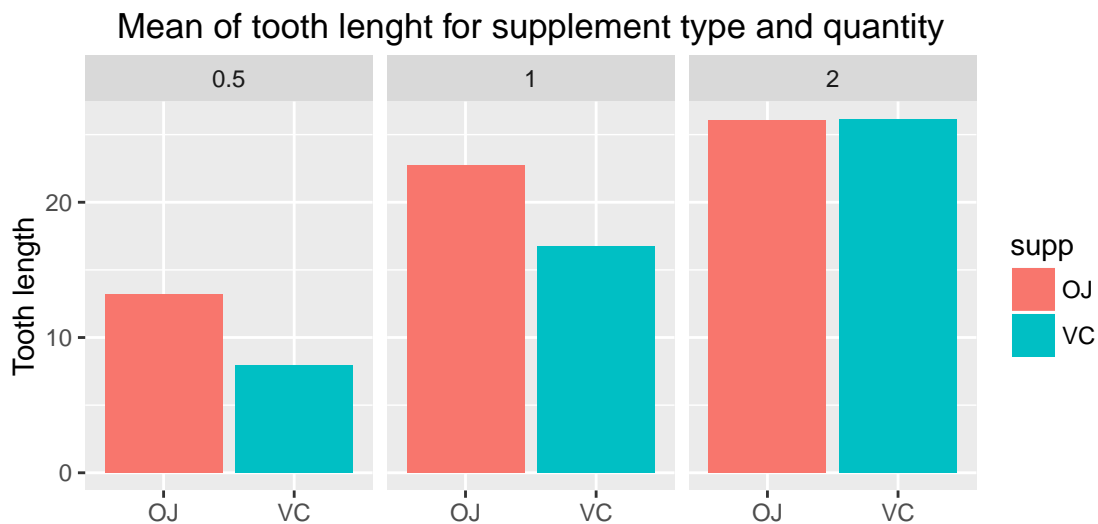

The lenght seems correlated with dose quantity and, in part, with supplement type.

QUESTION 3: Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

We perform a first T-test with significance level of 5%, with this Null Hypotesis: *different supplement type hasn't effect with tooth growth.*

```
t.test(len ~ supp, data = ToothGrowth)
```

```
##
##  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
```

We can't reject null hypotesis because confidence interval contains zero. **We can't say that supplement types have significant impact on tooth growth**.

Now we perform three tests with significance level of 5%, with this Null Hypotesis: *different dose hasn't effect with tooth growth.*

```
# dose = 2 vs dose = 0.5
t.test(ToothGrowth$len[ToothGrowth$dose == 2], ToothGrowth$len[ToothGrowth$dose ==
    0.5])
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose == 0.5]
## 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:
##  12.83383 18.15617
## sample estimates:
## mean of x mean of y
##    26.100    10.605
```

```
# dose = 2 vs dose = 1
t.test(ToothGrowth$len[ToothGrowth$dose == 2], ToothGrowth$len[ToothGrowth$dose ==
    1])
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose == 1]
## 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:
##  3.733519 8.996481
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

```
# dose = 1 vs dose = 0.5
t.test(ToothGrowth$len[ToothGrowth$dose == 1], ToothGrowth$len[ToothGrowth$dose ==
    0.5])
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 1] and ToothGrowth$len[ToothGrowth$dose == 0.5]
## 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:
##    6.276219 11.983781
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

All three tests have confidence intervals that don't contains 0. This means that we can accept the null hypotesis.
Increasing the dose of vitamin C have impact on tooth growth.

QUESTION 4: State your conclusions and the assumptions needed for your conclusions.

We assume that the experiment was done with random assignment of guinea pigs and the sample is representative of the entire population. The variances are assumed to be different for the groups being compared.
If all the assumptions are true, we can say that increasing the dose of vitamin C have a significant impact on tooth growth of the pig: a higher dose level of vitamin C correspond to longer teeth.
We also can say that supplement types of vitamin C (orange juice or ascorbic acid) don't have significant impact on tooth growth.