Statistical Inference: Course Project 2

VALAT Didier

13/06/2015

Overview

The aim of this project is to use the ToothGrowth data set and perform some basic exploratory data analyses: - Provide a basic summary of the data. - Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering). - State your conclusions and the assumptions needed for your conclusions.

Load the data

First, we will load the required libraries and define the figures size:

```
# Load libraries
library(knitr)
library(ggplot2)
library(datasets)
library(gridExtra)
library(GGally)
opts_chunk$set(fig.width=6, fig.height=3.5)
```

Then, we will load the ToothGrowth data:

```
data(ToothGrowth)
```

Basic summary of the data

We will now do a basic summary of the data:

```
str(ToothGrowth)
```

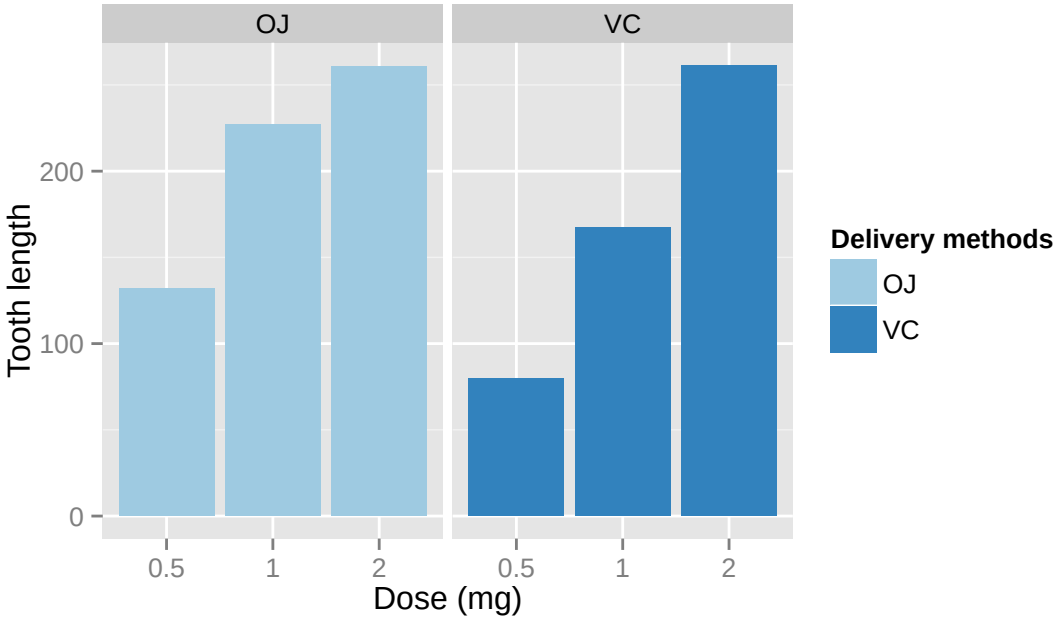
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
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
```

As we can see, the data is a set of 60 observations and 3 variables, length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice "OJ" or ascorbic acid "VC").

We will now compare the tooth length for each delivery methods:



This graph helps to see that there is a clear positive correlation between the tooth length and the dose levels of Vitamin C, for both delivery methods.

We will now use confidence intervals and/or hypothesis tests to compare tooth growth by dose and delivery methods.

Test A: does the delivery methods have an impact on tooth growth?

Null hypothesis: both delivery methods have the same effect on tooth growth

- H_o: the mean of delivery method OJ **is equal** the mean of delivery method VC.

Alternative hypothesis: the delivery methods have not the same effect on tooth growth

- H_a: the mean of delivery method OJ **is not equal** to the mean of delivery method VC.

```
t.test(data = ToothGrowth, len ~ supp, paired = FALSE, var.equal=FALSE)
```

```
##
##  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
```

Assuming a 95% confidence level the null hypothesis is **not rejected** as the p-value **is more than 0.05%**.

Test B: does the quantity of dose have an impact on tooth growth?

Null hypothesis 1: dose 0.5mg has the same effect on Tooth Growth as dose 1mg

- H_o: the mean of dose 0.5mg **is equal** to the mean of dose 1mg.

Alternative hypothesis 1: dose 0.5mg does not have the same effect on tooth growth as dose 1mg

- H_a: the mean of dose 0.5mg **is not equal** to the mean of dose 1mg.

```
t.test(dose_1$len, dose_2$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_1$len and dose_2$len
## 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:
##  -11.983781  -6.276219
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

Assuming a 95% confidence level the null hypothesis is **rejected** as the p-value is **less than 0.05%**.

Null hypothesis 2: dose 1mg has the same effect on tooth growth as dose 2mg

- H_o: the mean of dose 1mg **is equal** to the mean of dose 2mg.

Alternative hypothesis 2: dose 1mg does not have the same effect on tooth growth as dose 2mg

- H_a: the mean of dose 1mg **is not equal** to the mean of dose 2mg.

```
t.test(dose_2$len, dose_3$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_2$len and dose_3$len
## 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:
##  -8.996481 -3.733519
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

Assuming a 95% confidence level the null hypothesis is **rejected** as the p-value is **less than 0.05%**.

Conclusions

According to the hypothesis tests for delivery method ("supp" variable), there is not enough evidence to reject the null hypothesis therefore we can conlude that **the delivery method did not impact the level of tooth growth significantly**.

The hypothesis tests for dose quantity ("dose" variable) show that there is enough evidence to reject the null hypothesis so we can conclude that **the quantity of dose did impact tooth growth significantly**.

Assumptions

- The variances are different for the seperate populations.
- It seems that there is a bug on the documentation and the original experiment has been made on 60 guinea pigs: https://bugs.r-project.org/bugzilla3/show_bug.cgi?id=15953. Therefore we can assume that the data is independant.