Statistical Inference Course Project - part 2

Andreia Carlos

14 October 2016

Overview

This project consists of two parts:

1. A simulation exercise
2. **Basic inferential data analysis**

This report presents part 2. Basic inferential data analysis are used to analyse the ToothGrowth data (available in R datasets package). The report starts with summaries of the data, exploratory analysis, and inferential statistics to make assumptions about the data and support conclusions.

Part 2: Basic Inferential Data Analysis

1. Load the ToothGrowth data and perform some basic exploratory data analysis

Load required libraries

```
library(pander)
library(dplyr)
library(ggplot2)
library(tidyr)
```

Load the data

```
data("ToothGrowth")
```

Explore the data - Learn about the data

```
help("ToothGrowth") # to learn more about this data
pander(head(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

```
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 ...
```

This data has 60 observations and 3 variables. len is the tooth `length`, `supp` is the supplement type (VC or OJ) and `dose` is the dose of vitamins the guinea pigs took in milligrams/day.

2. Provide basic summary of the data

```
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
```

From the summary of the data, we can se that tooth length (represented by the quantity of odontoblasts) has a mean of `18.81` in a range between `4.20` and `33.90`. Half of the guinea pigs have a 19.25 tooth length. The supplement was equally given to the two groups of pigs (30 each), and the doses of it varied between 0.5 and 2.

```
# How many pigs were in each dose group
pander(summary(as.factor(ToothGrowth$dose)))
```

0.5	1	2
20	20	20

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

The pigs were equally divided per each group.

Exploratory data analysis

Exploratory question 1: Is the tooth length different in the two supplement groups?

Determine the mean of tooth length for each group and compare these means with exploratory analysis

```
SuppMeans <- ToothGrowth %>%
        group_by(supp) %>%
        summarise(meanLen = mean(len)) %>%
        pander()
```

The plot in appendix shows the tooth length by groups of supplements: OJ or VC

```
ggplot(ToothGrowth, aes(x = supp, y = len)) +
        labs(title = "Mean tooth growth by supplement type",
             x = "Supplement type", y = "Tooth length") +
        geom_boxplot(aes(fill = supp)) + scale_fill_brewer(palette="PuOr")
```

Exploratory question 2: Is the tooth length different in the three doses group? Does it increase as the dose increases?

```
DoseMeans <- ToothGrowth %>%
        group_by(dose) %>%
        summarise(meanLen = mean(len))
```

The following code produces a plot (available in appendix) that shows the tooth length by the three doses and in the two delivery methods groups

```
ggplot(ToothGrowth, aes(x = supp, y = len)) +
        labs(title = "Tooth growth by dose",
             x = "dose", y = "Tooth length") +
        geom_boxplot(aes(fill = supp)) + facet_grid( ~ dose) + scale_fill_brewer(palette="PuOr")
```

** Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose**

Inferential Statistics

Considering a null hypothesis that there is no difference between the means of tooth length by the type delivery method of the supplement, we test the difference in the means with an independent two-sided T test.

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

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

Table 3: Welch Two Sample t-test: `len` by `supp`

Test statistic	df	P value	Alternative hypothesis
1.915	55.31	0.06063	two.sided

Since the p-value is greater than 0.05, we accept the null hypothesis that the tooth growth (tooth length) is not significantly different between the two delivery methods of supplements. However better understanding of the area is advised to make clinically relevant assumptions.

```
pander(t.test(len ~ supp, data = ToothGrowth)$conf.int)
```

-0.171 and *7.571* The confidence interval contains zero which supports the assumption that there is no difference in the mean of tooth growth between the two supplement groups.

Considering a null hypothesis that there is no difference between the means of tooth length and the different doses, we test these difference in the means using a T-test as well.

Comparing dose 0.5 with dose 1.0

```
Group1 <- ToothGrowth %>%
        filter(dose %in% c(0.5, 1.0))
pander(t.test(len ~ dose, data = Group1))
```

Table 4: Welch Two Sample t-test: `len` by `dose`

Test statistic	df	P value	Alternative hypothesis
-6.477	37.99	1.268e-07 * * *	two.sided

```
Group2 <- ToothGrowth %>%
        filter(dose %in% c(1.0, 2.0))
pander(t.test(len ~ dose, data = Group2))
```

Table 5: Welch Two Sample t-test: `len` by `dose`

Test statistic	df	P value	Alternative hypothesis
-4.9	37.1	1.906e-05 * * *	two.sided

```
pander(t.test(len ~ dose, data = Group2)$conf.int)
```

-8.996 and *-3.734*

```
Group3 <- ToothGrowth %>%
        filter(dose %in% c(0.5, 2.0))
pander(t.test(len ~ dose, data = Group3))
```

Table 6: Welch Two Sample t-test: `len` by `dose`

Test statistic	df	P value	Alternative hypothesis
-11.8	36.88	4.398e-14 * * *	two.sided

```
pander(t.test(len ~ dose, data = Group3)$conf.int)
```

-18.16 and *-12.83*

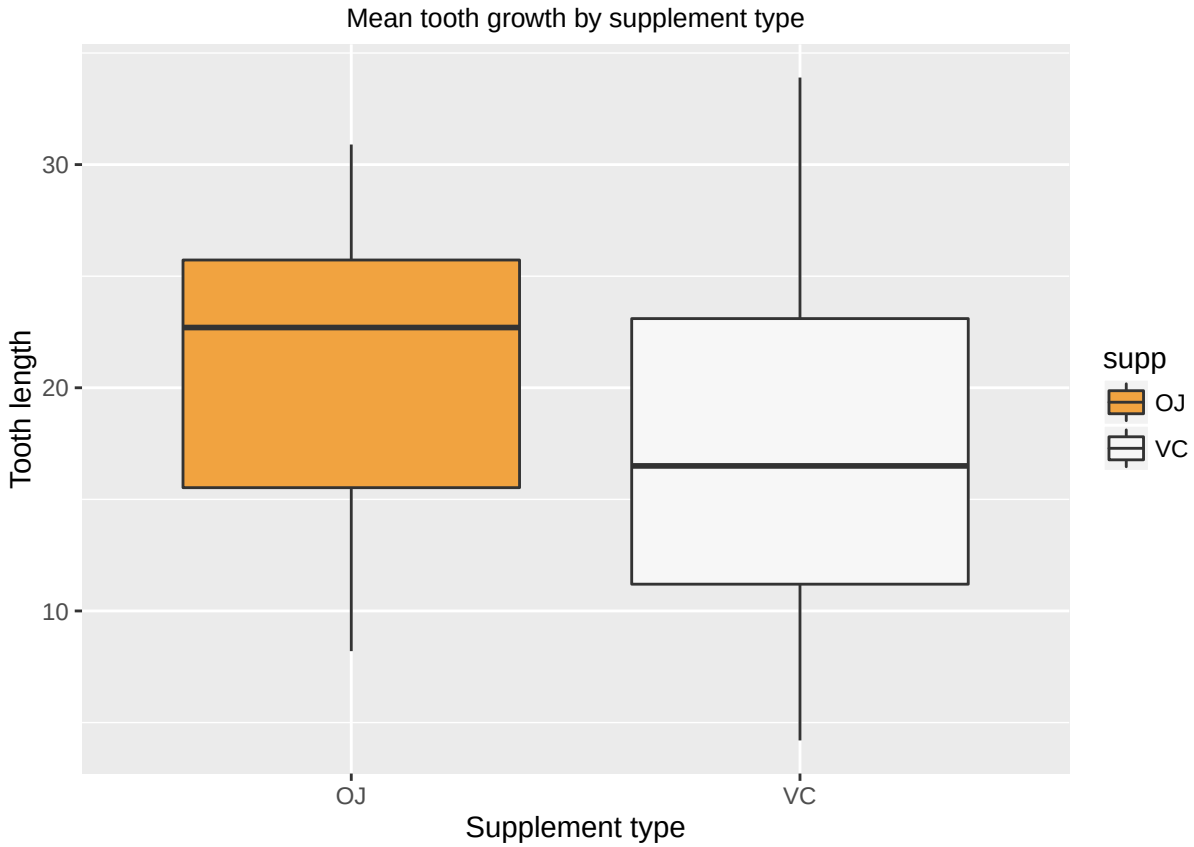
In these two sample T tests, the null hypothesis can be rejected as the confidence intervals do not contain zero.

4. Conclusions and assumptions to support these conclusions

Considering the exploratory analysis and inferential statistics used, we can make assumptions in relation to the supplement effect in the tooth growth. Therefore the supplement doesn't seem to have a different effect with the method delivery (VO or OJ), but there seems to be higher tooth growth in guinea pigs when comparing the administered doses. The mean tooth length is significantly different from lower to higher doses of supplement.

Appendix

The plot below shows the tooth length by groups of supplements: OJ or VC



We can see the tooth length mean represented with the dark horizontal line in each group.

The following plot shows the tooth length by the three doses and in the two delivery methods groups