Statistical Inference Project - Part II

Evgeniia Golovina

12/03/2021

Overview

In this part of the project, we need to perform basic inferential data analysis. We analyse the ToothGrowth data to do the following:

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.
4. State your conclusions and the assumptions needed for your conclusions.

1. Loading and exploring the data

```
data(ToothGrowth); str(ToothGrowth) # 3 variables and 60 observations
```

```
## '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 ...
```

The ToothGrowth dataset contains the results of the effect of vitamin C on tooth growth in guinea pigs. The dataset has three measurments (len, dose and supp) done on 60 guinea pigs. First measurment is the tooth length, second - dose of the vitamin C (0.5, 1 or 2 mg/day) and third is delivery method (orange juice or ascorbic acid).

2. Getting basic summary of the data

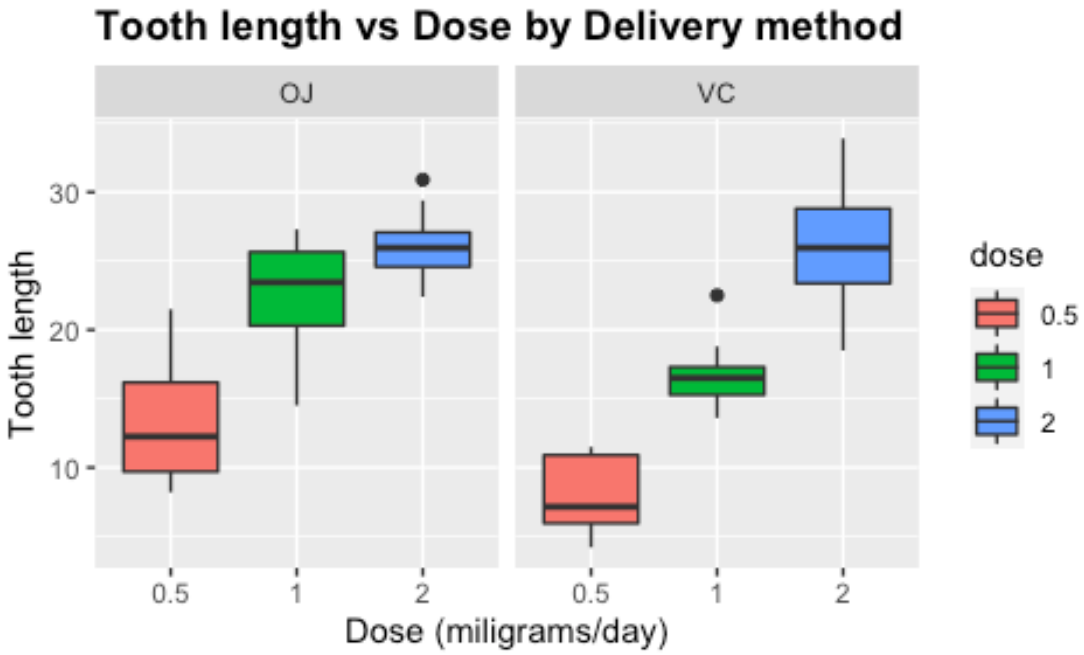
```
ToothGrowth$dose <- as.factor(ToothGrowth$dose); summary(ToothGrowth)
```

```
##       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
```

As we can see, 20 of 60 pigs received 0.5 mg/day, 20 of 60 - 1 mg/day and 20 - 2 mg/day of vitamin C. Half of the animals got vitamin C via orange juice and another half - via ascorbic acid. The maximum tooth length is 33.90, the minimum is 4.20, the mean is 18.81.

```
ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose)) +
  xlab("Dose (miligrams/day)") + ylab("Tooth length") + facet_grid(~ supp) +
```

```
  ggtitle("Tooth length vs Dose by Delivery method") +
  theme(plot.title = element_text(lineheight=.8, face="bold"))
```



The effect of vitamin C via orange juice (OJ) seems to be the same or slightly higher compared to the effect of vitamin C via ascorbic acid (VC).

```
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp)) +
  xlab("Delivery method") + ylab("Tooth length") + facet_grid(~ dose) +
  ggtitle("Tooth length vs Delivery method by Dose") +
  theme(plot.title = element_text(lineheight=.8, face="bold"))
```

The larger the dose, the higher the effect of vitamin C on the tooth growth. Delivery by oranje juice (OJ) (especially 0.5 and 1 mg/day) has more effect on tooth growth than delivery via ascorbic acid (VC).

3. Using hypothesis testing to compare tooth growth by supp and dose.

First, let's test if the tooth length depends on the delivery method. Let's perform a t test.

```
len <- ToothGrowth$len; supp <- ToothGrowth$supp; dose <- ToothGrowth$dose
t1 <- t.test(len[supp=="OJ"], len[supp=="VC"]); t1$p.value

## [1] 0.06063451
```

The p-value of this test is *0.06063* is greater than 0.05, and the confidence interval contains zero. Thus, we can't reject the null hypothesis that delivery method seems to have no effect on tooth growth.

Now, let's test the effect of dose.

```
t_2.1 <- t.test(len[dose==2], len[dose==1]); t_2.2 <- t.test(len[dose==1],
len[dose==0.5])
t_2.3 <- t.test(len[dose==2], len[dose==0.5])
p <- c(t_2.1$p.value, t_2.2$p.value, t_2.3$p.value); p

## [1] 1.906430e-05 1.268301e-07 4.397525e-14
```

As we can see, the p-value of each test is less than 0.05 and the confidence intervals doesn't contain zero. Thus, we can reject the null hypothesis and accept the alternative hypothesis that dose seems to have effect on tooth growth.

4. Conclusions

We assume that 1) the dataset represents the population and 2) the distribution of the sample means follows the Central Limit Theorem. According to our analysis, we can say that delivery method has no effect on tooth growth, while dose has a significant effect on the tooth growth.