

# Statistical Inference Project Part 2

Emanuel Chalela

26/12/2020

## Part 2: Basic Inferential Data Analysis Instructions

1. Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

```
library("ggplot2")
library("datasets")
library("tinytex")
```

2. Load the ToothGrowth data and perform some basic exploratory data analyses

Provide a basic summary of the data.

```
data(ToothGrowth)
summary(ToothGrowth)
```

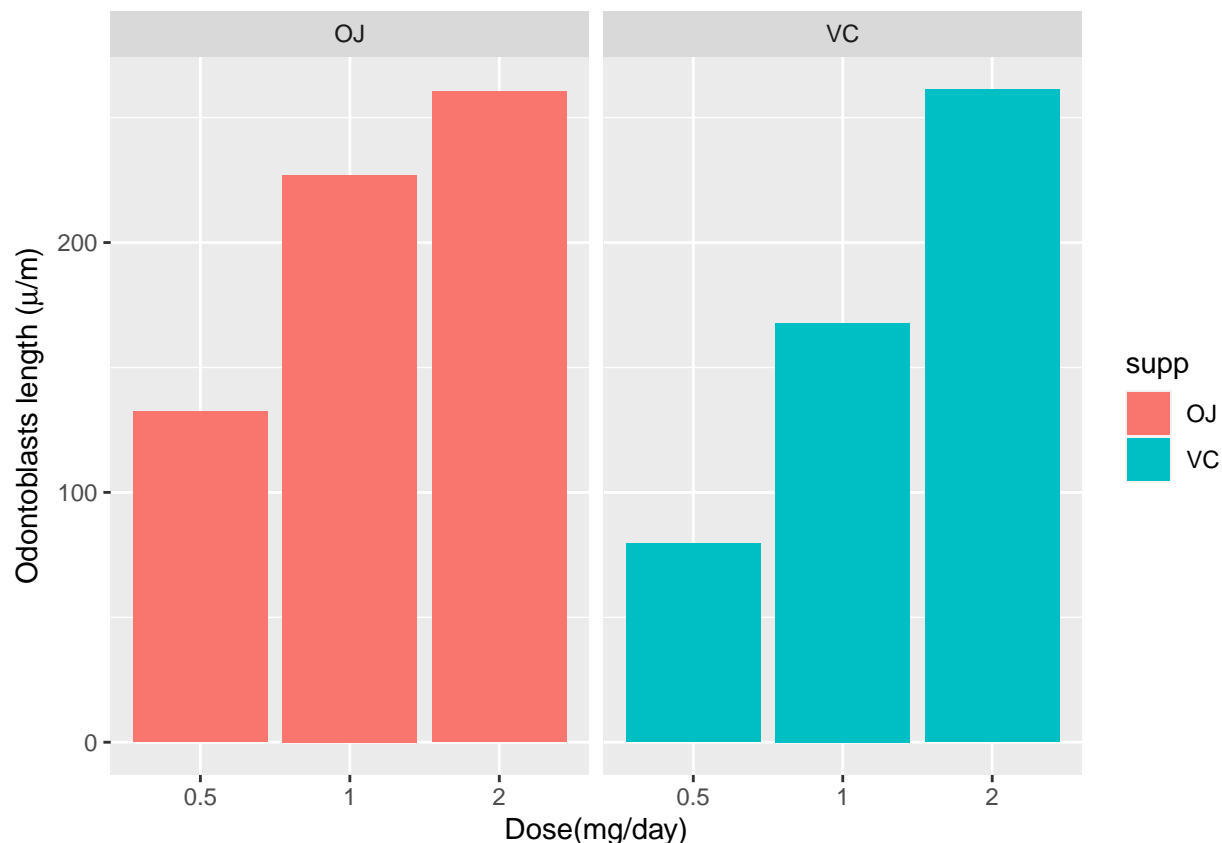
```
##      len      supp      dose
##  Min.   : 4.20   OJ:30   Min.    :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean    :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.    :2.000
```

ToothGrowth has 60 observations

3. 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.

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
ggplot(data=ToothGrowth, aes(x=dose, y=len, fill=supp)) +
  geom_bar(stat="identity") +
  facet_grid(. ~ supp) +
  xlab("Dose(mg/day)") +
  ylab(expression(paste("Odontoblasts length (" ,mu, "/m)")))
```



### Hypothesis tests Test if the delivery mode has an influence on the tooth growth. ##### H0 could be formulated as follows :

H0 : The delivery mode of Vitamin C does not have any influence on the tooth growth

```
dose <- ToothGrowth$dose
supp <- ToothGrowth$supp
len <- ToothGrowth$len

t.test(len[supp == "VC"], len[supp == "OJ"], paired=FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: len[supp == "VC"] and len[supp == "OJ"]
## 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:
## -7.5710156 0.1710156
## sample estimates:
## mean of x mean of y
## 16.96333 20.66333
```

This first test shows the following :

the p-value is 0.061, i.e. nearly the significance level  $\alpha = 0.05$ . So we do not reject the null hypothesis, but as  $0.05 \leq p\text{-value} \leq 0.1$ , it is not clearly obvious that we can reject the null hypothesis. the confidence

interval contains 0, so the test is not really significant. Now let's try to test the influence of the dose on the tooth growth

```
t.test(len[dose == 0.5], len[dose == 1], paired=FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: len[dose == 0.5] and len[dose == 1]
## 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
```

In this test, we can clearly see that the quantity of Vitamin C has an impact on the tooth growth :

the p-value is nearly 0, so we can obviously reject  $H_0$  the confidence interval does not contain 0. An identical conclusion can be taken comparing the length of dose = 1 and dose = 2. This could already be detected from the boxplot above.

## Conclusion

The dose of Vitamin C is clearly a factor of growth of the teeth for Guinea pigs the delivery mode (Ascorbic Acid or Orange Juice) does not have any obvious impact on the teeth growth for Guinea pigs.