

Basic Inferential Data Analysis

Kenneth Fajardo

4/29/2021

Overview

We will be working on the ToothGrowth dataset, which is the result of a study by E. W. Crampton in 1947, The Effect of Vitamin C on Tooth Growth in Guinea Pigs. Quoting from the R documentation:

“The 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 (a form of vitamin C and coded as VC).”

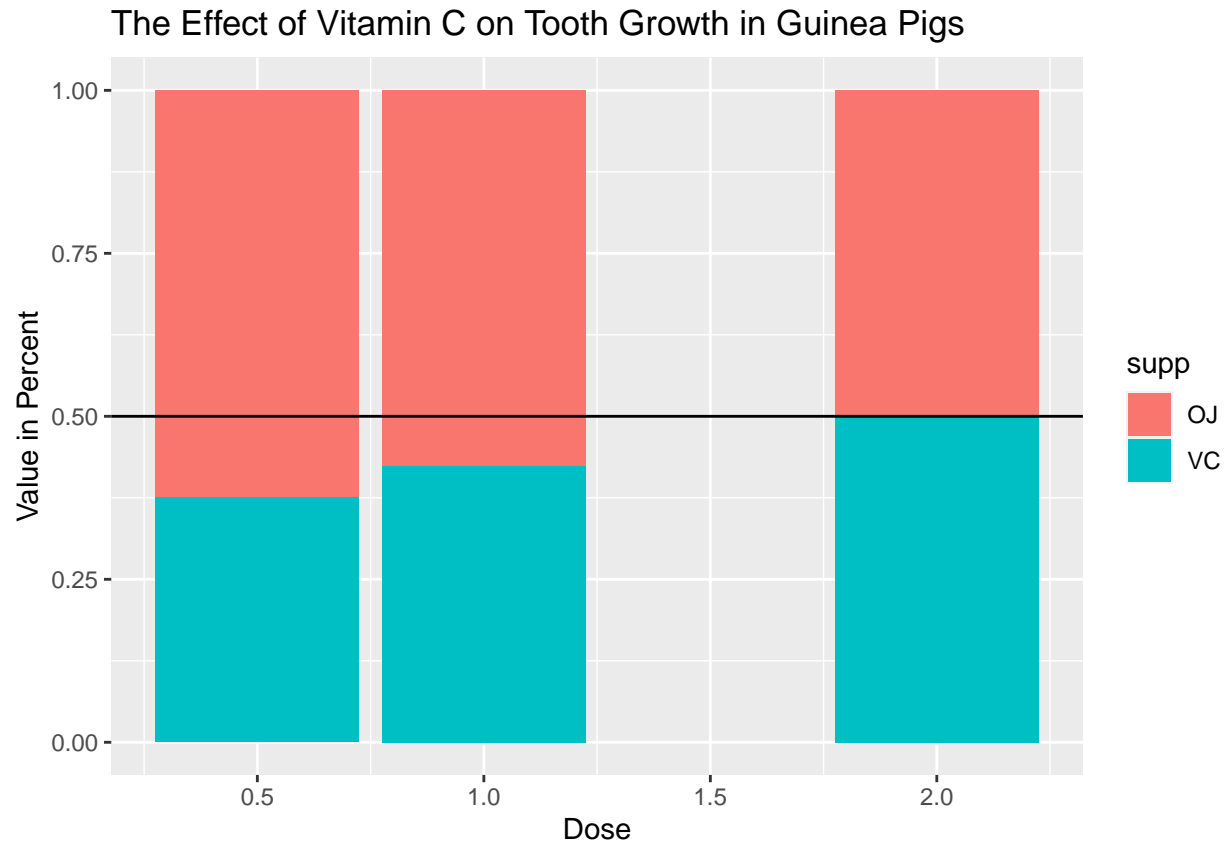
Load the ToothGrowth data and perform some basic exploratory data analyses

Load the needed libraries and then load the ToothGrowth data.

```
library(datasets)
library(ggplot2)
# load the data
data(ToothGrowth)
```

From the figure below, we can see that there is a 15 – 25% decrease in length for dosages of 0.5 and 1.0. However, this is not a basis for rejecting the null hypothesis H_0 , which suggest that Vitamin C has no effect on tooth growth ($H_0 : \mu = 0$).

```
g <- ggplot(ToothGrowth, aes(x=dose, y=len, fill=supp))
g + geom_col(position = "fill") + xlab("Dose") + ylab("Value in Percent") +
  labs(title = "The Effect of Vitamin C on Tooth Growth in Guinea Pigs") +
  geom_hline(yintercept = 0.5, color="black")
```



Provide 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 ...
```

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

We propose an alternative hypothesis that Vitamin C has an effect to tooth growth ($H_a : \mu \neq 0$). Since the sample size is not too large, we use a t-distribution for generating the confidence intervals. For convenience, we use the *t.test* function in R.

We will be testing the hypothesis according to *supp* and *dose*. The tests are divided into the following:

(1) general (regardless of dosage),

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

```
## [1] -0.1710156  7.5710156  
## attr(,"conf.level")  
## [1] 0.95
```

```
dose0.5 <- ToothGrowth[ToothGrowth$dose==0.5,]  
dose1.0 <- ToothGrowth[ToothGrowth$dose==1.0,]  
dose2.0 <- ToothGrowth[ToothGrowth$dose==2.0,]
```

Note that the tests are not paired since the subjects are not measured twice.

(2) $pair_{dose} = (0.5, 1.0)$

```
t.test(dose0.5$len, dose1.0$len, paired = FALSE)$conf
```

```
## [1] -11.983781  -6.276219  
## attr(,"conf.level")  
## [1] 0.95
```

(3) $pair_{dose} = (0.5, 2.0)$

```
t.test(dose0.5$len, dose2.0$len, paired = FALSE)$conf
```

```
## [1] -18.15617 -12.83383  
## attr(,"conf.level")  
## [1] 0.95
```

(4) $pair_{dose} = (1.0, 2.0)$

```
t.test(dose1.0$len, dose2.0$len, paired = FALSE)$conf
```

```
## [1] -8.996481 -3.733519  
## attr(,"conf.level")  
## [1] 0.95
```

Conclusions and supporting assumptions

Based from the confidence intervals, for the first case where we test regardless of dosage, we do not reject H_0 since the interval contains 0. However, when testing the ordered pair of dosages, we get negative values in the confidence interval. This implies that **the lower the dosage, the shorter the length of the teeth are**. Conversely, higher dosages yield to higher tooth growth.

NOTE: This is true only if the experiment is fair i.e. the subjects are *independently and identically distributed*, as mentioned in the documentation of the dataset.