

# Basic inferential data analysis

## Summary

This assignment is focussing on the relationship between supplements and the growth of tooth in guinea pigs. At the end of the study the result is that there is no significant difference between different types of supplements but there is a relationship between the growth of the tooth and the doses of the supplements.

## Research

### Load the ToothGrowth data and perform some basic exploratory data analyses

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).

Columns: Content: len numeric Tooth length supp factor Supplement type (VC or OJ). dose numeric Dose in milligrams/day

```
library(datasets)
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(ggplot2)
library(knitr)
library(rmarkdown)
library(reshape2)
library(cowplot)

##
## Attaching package: 'cowplot'
## The following object is masked from 'package:ggplot2':
##
##     ggsave

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

```
##      len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

## Provide a basic summary of the data.

Get a quick view on the data summerized by supplement and dose given the mean of the length of the Tooth. The bigger the dose the longer the teeth. It also looks like OJ is working better in smaller doses.

```
Growth <- ToothGrowth %>% group_by(supp, dose) %>% summarise(len = mean(len))
Growth
## Source: local data frame [6 x 3]
## Groups: supp [?]
##
##      supp  dose   len
##   (fctr) (dbl) (dbl)
## 1     OJ   0.5 13.23
## 2     OJ   1.0 22.70
## 3     OJ   2.0 26.06
## 4     VC   0.5  7.98
## 5     VC   1.0 16.77
## 6     VC   2.0 26.14
```

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

There are two hypotheses to be tested:

- There is no relation between the supplement and the length of the tooth
- There is no relation between the dose and the length of the tooth

Relation between the supplement and the length of the tooth:

```
OJ = ToothGrowth$len[ToothGrowth$supp == 'OJ']
VC = ToothGrowth$len[ToothGrowth$supp == 'VC']

t.test(OJ, VC, alternative = "greater", paired = FALSE, var.equal = FALSE,
conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ and VC
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687      Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

The null hypotheses that there is a relationship between the supplements and the length of the tooth has a p value of less than 5% (3%). The 5% interval is a rule of thumb used. This means we have to reject the null hypotheses and assume that there is a correlation.

Relation between the supplement and the length of the tooth:

```
doseHalf = ToothGrowth$len[ToothGrowth$dose == 0.5]
doseOne = ToothGrowth$len[ToothGrowth$dose == 1]
doseTwo = ToothGrowth$len[ToothGrowth$dose == 2]

t.test(doseHalf, doseOne, alternative = "less", paired = FALSE, var.equal =
FALSE, conf.level = 0.95)

##
##  Welch Two Sample t-test
##
## data:  doseHalf and doseOne
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -6.753323
## sample estimates:
## mean of x mean of y
##   10.605   19.735

t.test(doseOne, doseTwo, alternative = "less", paired = FALSE, var.equal =
FALSE, conf.level = 0.95)

##
##  Welch Two Sample t-test
##
## data:  doseOne and doseTwo
```

```
## t = -4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -4.17387
## sample estimates:
## mean of x mean of y
##      19.735      26.100
```

The null hypotheses that there is a relationship between the dose and the length of the tooth. This is tested in two ways, the relationship between the two smaller and the relationship between the two largest doses. Both have a very small p value. This means we have to reject the null hypotheses and assume that there is a correlation.

**State your conclusions and the assumptions needed for your conclusions.**

My conclusion is that based on a 5% confidence interval: 1. There is no relationship between the supplement and the length of the tooth. This means that you could use either of them. Although the basic summary suggests that one supplement is better in small doses. This is not further investigated. 2. There is a relationship between the dose and the length of the tooth. The P values are too small so the null hypotheses (difference between doses is 0) have to be rejected.