Investigating statistical inferences - hypothesis testing

LG

28 12 2021

Overview:

Statistical inferences is a scientific field in which we can draw inferences from a sample to a whole population. Testing statistical hypothesis to investigate whether differences e.g. between outcomes or groups are signification becomes handy in decision making or recommondations. This is depicted in this document based on an example related to vitamin C supplementation on tooth growth.

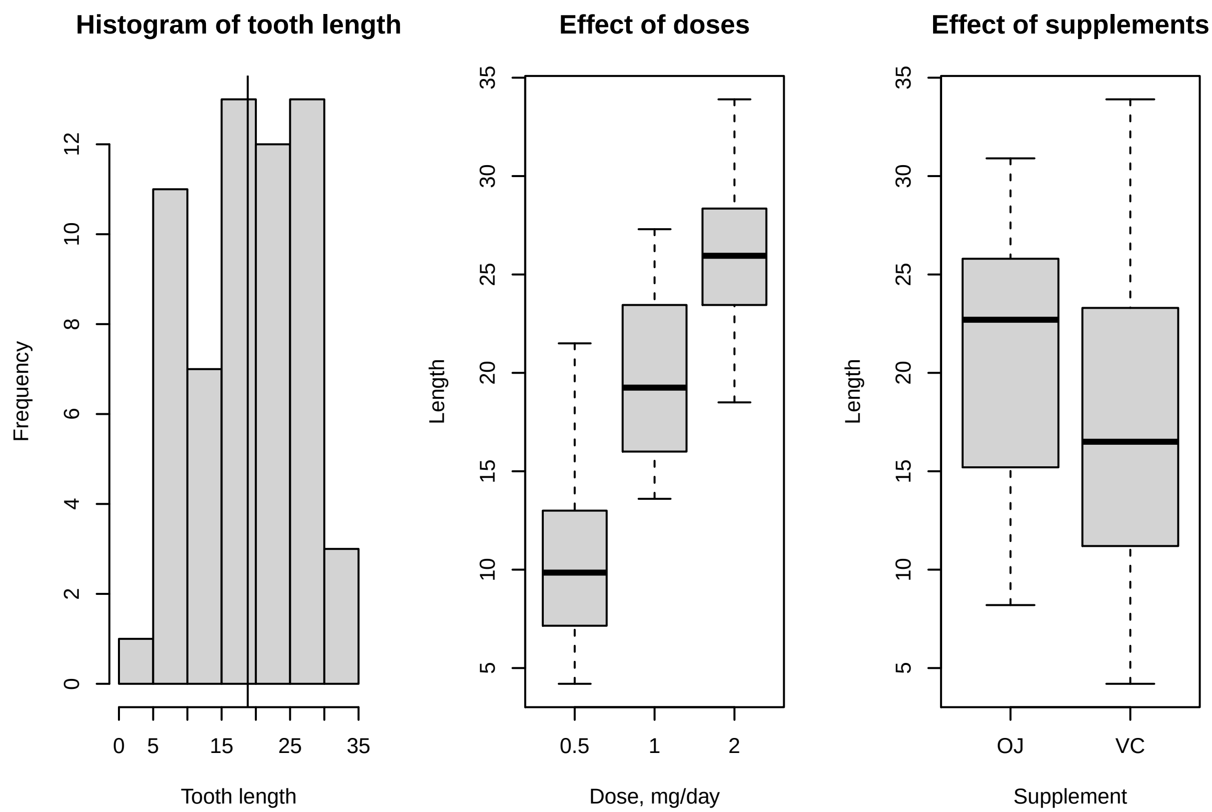
Toothgrowth Data

An experiment was conducted to examine the effects of vitamin C on tooth growths in 60 guinea pigs. The data set contains 3 variables: `len` as the length of tooth, `supp` representing the supplement, either orange juice as OJ or ascoribic acid as VC, and `dose` as the dose received in milligrams per day. Let's explore the data first.

```
library(ggplot2)
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
```

```
par(mfrow = c(1,3))
hist(ToothGrowth$len, xlab="Tooth length", main= "Histogram of tooth length")
abline(v=mean(ToothGrowth$len, col = "green", lwd =2))
with(data= ToothGrowth, plot(x=factor(dose), y=len, main = "Effect of doses"
                              , ylab= "Length", xlab= "Dose, mg/day"))
with(data= ToothGrowth, plot(x=supp, y= len, main="Effect of supplements"
                              , ylab= "Length", xlab= "Supplement"))
```



Hypothesis testing

From visual inspection, it seems like the dose of 2.0 mg/day is associated with greatest length of tooth and orange juice seems to be the supplement associated with longer tooth. Let's run inferential statistics: 1. is length of tooth due to delivery method different? 2. is length of tooth due to vitamin C dose different? 3. is length of tooth different for each dosage and supplement?

First, we need to check assumptions, before choosing appropriate statistical testing method. We assume that the guinea pigs are randomly sampled from and representative of the population, that the length of tooth is normally distributed and variance is equal. Due to smaller sample size, Student's T-Test is choosen at an type 1 error rate of 5%.

1 Difference in supplements

We define the null hypothesis equal to 0 whereas the alternative hypothesis is that the means differ.

```
lenOJ <- ToothGrowth$len[ToothGrowth$supp == "OJ"]
lenVC <- ToothGrowth$len[ToothGrowth$supp == "VC"]

t_supp<- t.test(lenOJ, lenVC, mu = 0, var.equal=T, alternative = "two.sided",
                paired =F, conf.level=0.95)
```

The two sample t-test reveals that the difference of tooth length is not significant between the two supplements, as we fail to reject the null hypothesis with p-value = 0.0603933712241287 and confidence intervals = -0.167006420141225, 7.56700642014122.

2 Difference in doses

We define the null hypothesis equal to 0 whereas the alternative hypothesis is that the means differ. We need to conduct three tests, as doses are either 0.5 mg/day, 1 mg/day or 2 mg/day.

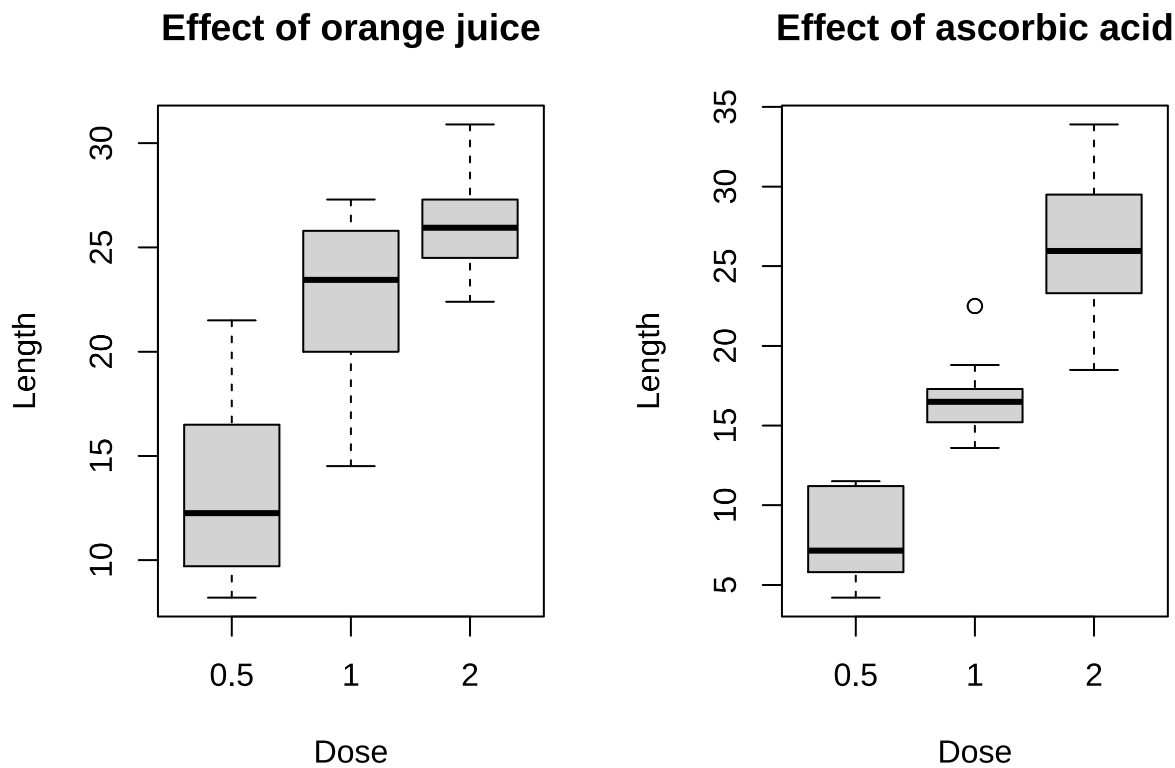
```
dose0.5 <- ToothGrowth$len[ToothGrowth$dose == 0.5]
dose1 <- ToothGrowth$len[ToothGrowth$dose == 1]
dose2 <- ToothGrowth$len[ToothGrowth$dose == 2]

t_dose0.5 <- t.test(dose0.5, dose1, mu = 0, var.equal = TRUE, alternative = "two.sided",
                    paired=FALSE, conf.level= 0.95)
t_dose1 <- t.test(dose1, dose2, mu = 0, var.equal = TRUE, alternative = "two.sided",
                  paired=FALSE, conf.level= 0.95)
t_dose2 <- t.test(dose0.5, dose2, mu = 0, var.equal = TRUE, alternative = "two.sided",
                  paired=FALSE, conf.level= 0.95)
```

The difference between dose of 0.5 mg/day and 1 mg/day is p = 1.26629696132167e-07 with confidence intervals of -11.984, -6.276. The difference between dose of 1 and 2 mg/day is p = 1.81082853618171e-05 with confidence intervals of -8.994, -3.736. The difference between dose 0.5 and 2 mg/day is p = 2.83755316767472e-14 with confidence intervals of -18.154, -12.836. These results show, that no matter what difference in dose is compared, the difference is significant and we can in all cases reject the null hypothesis.

3 Difference of dosage per supplement

Let's see if we can detect a certain combination of supplement and dose to make a recommendation which alternative in fact will support long teeth. From visual inspection, we can assume differences.

```
par(mfrow= c(1,2))
with(ToothGrowth[ToothGrowth$supp== "OJ",], plot(x= factor(dose), y = len,
                                                 ylab= "Length", xlab="Dose",
                                                 main="Effect of orange juice"))
with(ToothGrowth[ToothGrowth$supp== "VC",], plot(x= factor(dose), y = len,
                                                 ylab= "Length", xlab="Dose",
                                                 main="Effect of ascorbic acid"))
```




```r
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.5

##
## 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
```

```r
t_supp_dose0.5 <- t.test(len~supp,data=filter(ToothGrowth, dose=="0.5"))
t_supp_dose1 <- t.test(len~supp,data=filter(ToothGrowth, dose=="1"))
t_supp_dose2 <- t.test(len~supp,data=filter(ToothGrowth, dose=="2"))

t_supp_dose0.5$p.value
```

```
## [1] 0.006358607
```

```
t_supp_dose1$p.value
```

```
## [1] 0.001038376
```

```
t_supp_dose2$p.value
```

```
## [1] 0.9638516
```

These results show that at a dose of 2 mg/day it doesn't matter what supplement is delivered, as p-value is 0.964 greater than type 1 error of 0.05. However, in case of 0.5 mg/ day or 1 mg/day Orange juice should be preferred over the alternative as in both cases the p-value < 0.05. Comparing the means, its obvious that 0.5 mg/day of Orange Juice results on average in longer teeth than with a dose of 1 mg/day in guinea pigs.