Analyze the ToothGrowth data in the R datasets package

Shivam Mishra

07/08/2020

Overview

We're going to analyze the ToothGrowth data in the R datasets package.
Load the ToothGrowth data and perform some basic exploratory data analyses Provide a basic summary of the data.

```
library(datasets)
data(ToothGrowth)
```

Data Exploration

```
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
```

```
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
```

Hypothesis Testing

Assumptions:
1. The variables must be independent and identically distributed (i.i.d.). 2. Variances of tooth growth are different when using different supplement and dosage. 3. Tooth growth follows a normal distribution.

Hypothesis for the supplement OJ vs VC
Let our null hypothesis to be there is no difference in tooth growth when using the supplement OJ and VC.

lenOJ=lenVC

Let our alternate hypothesis to be there are more tooth growth when using supplement OJ than VC.

lenOJ>lenVC

Then, we obtain the tooth growth by supplement type from the data

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

We will perform a t-test following the indications of the work to be evaluated.

One-tailed independent t-test with unequal variance.

```
t.test(OJ, VC, alternative = "greater", paired = FALSE, var.equal = FALSE, co
nf.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
```

As the p-value (0.03032) is lower than 0.05 (the default value for the tolerance of the error alpha), then, we reject the null hypothesis. That can be interpreted as there is aproximately 3% of chance of obtain an extreme value for the difference in mean of tooth growth.

Finally, based on this low p-value, we can conclude that it is very likely that supplement OJ, the greater the effect on tooth growth than supplement VC.

Hypothesis for the dossage

The null hypothesis is that there is no difference in tooth growth between dosage. Our alternate hypothesis is that there are more tooth growth when the dosage increases.

Extract the tooth growth by dosage.

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

One-tailed independent t-test with unequal variance.

```
t.test(doseHalf, doseOne, alternative = "less", paired = FALSE, var.equal = F
ALSE, 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
```

As the p-value (6.342e-08) is lower than 0.05 (the default value for the tolerance of the error alpha), then, we reject the null hypothesis. That can be interpreted as there is almost null chances of obtain an extreme value for the difference in mean of those dossages (doseHalf < doseOne) on the tooth growth.

```
t.test(doseOne, doseTwo, alternative = "less", paired = FALSE, var.equal = FA
LSE, 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 conclusion is similar than the previous, the p-value is 9.532e-06, close to 0. Then we reject the null hypothesis. That can be interpreted as there is almost null chances of obtain an extreme value for the difference in mean of those dossages (doseOne < doseTwo) on the tooth growth. The value is extreme (that's what we reject the null hypothesis).

Finally, based on these low p-values, we can conclude that it is very likely that dossage has effect, and a higher dossage higher tooth growth.