Basic Inferential Data Analysis

Dipanshu

08/12/2020

Synosis

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. lambda = 0.2 for all of the simulations. We will investigate the distribution of averages of 40 exponentials.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.
4. State your conclusions and the assumptions needed for your conclusions.

Loading the Libraries

```
library(datasets)
library(ggplot2)
```

Loading the data and having a look

```
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
```

Let's see what this data entails

```
str(ToothGrowth)
```

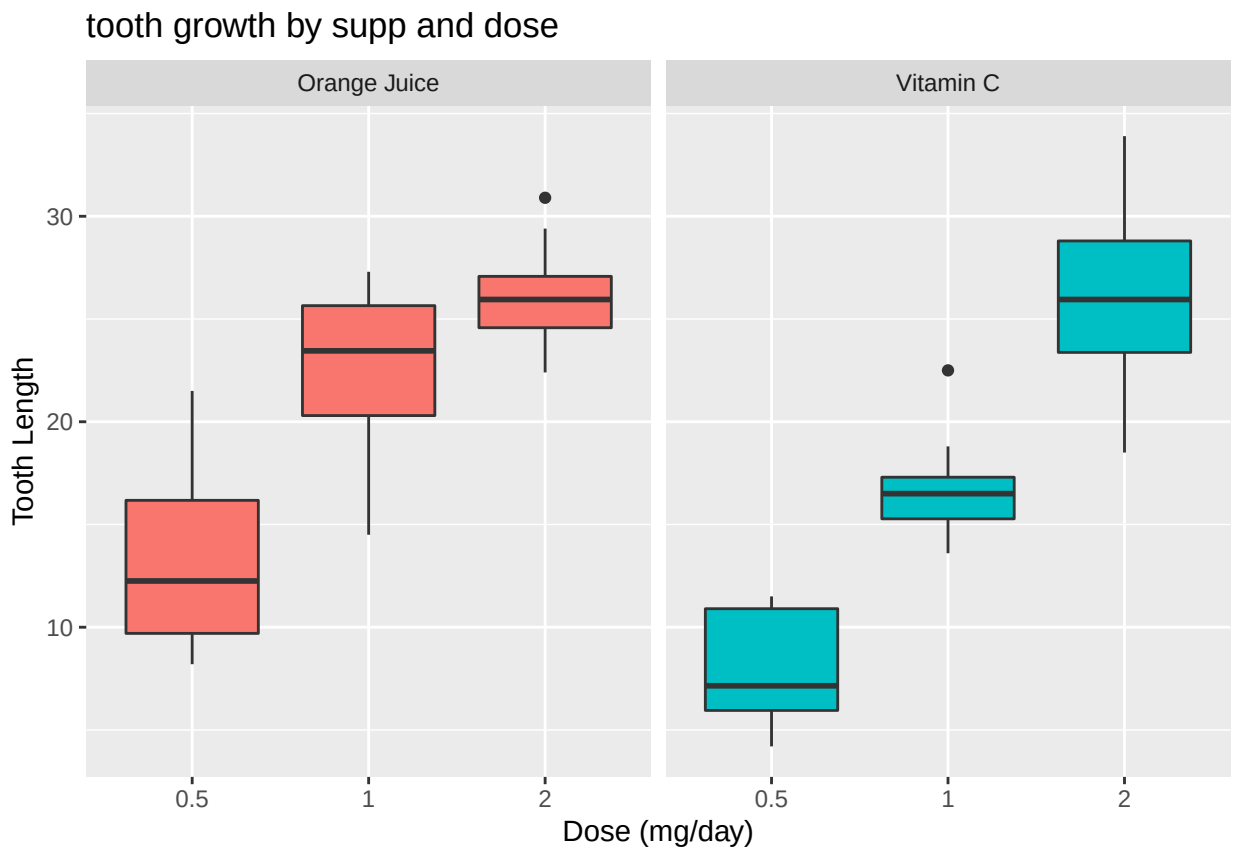
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
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
```

Let's first compare tooth growth by plotting it.

```
levels(ToothGrowth$supp) <- c("Orange Juice", "Vitamin C")
ggplot(ToothGrowth, aes(x=factor(dose), y=len)) +
  facet_grid(.~supp) +
  geom_boxplot(aes(fill = supp), show.legend = FALSE) +
  labs(title="tooth growth by supp and dose",
    x="Dose (mg/day)",
    y="Tooth Length")
```



Plot

From plot we can see, 1. Increasing the dosage increases the tooth growth. 2. Orange juice is more effective than vitamin C till 1 mg/day.

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

Hypothesis 1 H0 = Orange juice & Vitamin C deliver the same tooth growth across the data set.

```
hypoth1<-t.test(len ~ supp, data = ToothGrowth)
hypoth1$conf.int
```

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

```
hypoth1$p.value
```

```
## [1] 0.06063451
```

- The confidence intervals includes 0 and the p-value is greater than the threshold of 0.05.
- H0 can't be rejected

Hypothesis 2 H0 = For the dosage of 0.5 mg/day, the two supplements deliver the same tooth growth. HA= 0.5 mg/day dosage of orange juice delivers more tooth growth than Vitamin C is accepted.

```
hypoth2<-t.test(len ~ supp, data = subset(ToothGrowth, dose == 0.5))
hypoth2$conf.int
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```
hypoth2$p.value
```

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

- The confidence interval does not include 0 and the p-value is below the 0.05 threshold.
- H0 is rejected.

Hypothesis 3 H0 = For the dosage of 1 mg/day, the two supplements deliver the same tooth growth HA = 1 mg/day dosage of orange juice delivers more tooth growth than Vitamin C is accepted.

```
hypoth3<-t.test(len ~ supp, data = subset(ToothGrowth, dose == 1))
hypoth3$conf.int
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```
hypoth3$p.value
```

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

- The confidence interval does not include 0 and the p-value is below the 0.05 threshold.
- H0 is rejected.

Hypothesis #4 H0 = For the dosage of 2 mg/day, the two supplements deliver the same tooth growth HA = 2 mg/day dosage of orange juice delivers more tooth growth than Vitamin C is accepted.

```
hypoth4<-t.test(len ~ supp, data = subset(ToothGrowth, dose == 2))
hypoth4$conf.int
```

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

```
hypoth4$p.value
```

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

- The confidence interval includes 0 and the p-value is below the 0.05 threshold.
- H0 can't be rejected.

Conclusions

we can conclude that * Supp has no impact of tooth growth * Increasing Dose results in increasing tooth growth

Assumptions

- The sample is representative of the population
- The distribution of the sample means follows the Central Limit Theorem
- All the observations are independent