Statistical Inference Course Project - Part Two

Vitor Sousa

The goal of this projet is runa simulation when first we generate 40 variables form a exponential function and repeate this 1000 times. After that the for each simulation is computed.

```
#Packages
library(dplyr, warn.conflicts = F)
library(ggplot2)

## Load the ToothGrowth data and perform some basic exploratory data analyses
ToothGrowth <- tbl_df(ToothGrowth)


ToothGrowth %>% str()
#Classes ‘tbl_df’, ‘tbl’ and '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 ...

ToothGrowth %>% summary()
 # 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
```

The dataset have 60 observations of 3 variables. **len** and **dose** numeric and one, **factor**, is factor variable.

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

Now we want to further compare teeth growth by supplement type and dose levels.

```
len_a <- ToothGrowth %>% filter(dose %in% c(0.5,1)) %>% select(len) %>% unlist()
dose_a <- ToothGrowth %>% filter(dose %in% c(0.5,1)) %>% select(dose) %>%
unlist()
test_a <- t.test(len_a~dose_a, paired = FALSE)
test_a
#   Welch Two Sample t-test

#data:  len_a by dose_a
```

```
#t = -6.4766, df = 37.986, p-value = 1.268e-07
#alternative hypothesis: true difference in means is not equal to 0
#95 percent confidence interval:
# -11.983781  -6.276219
#sample estimates:
#mean in group 0.5   mean in group 1
#           10.605            19.735
```

```
len_b <- ToothGrowth %>% filter(dose %in% c(0.5,2)) %>% select(len) %>% unlist()
dose_b <- ToothGrowth %>% filter(dose %in% c(0.5, 2)) %>% select(dose) %>%
unlist()
test_b <- t.test(len_b~dose_b, paired = FALSE)
test_b
#   Welch Two Sample t-test

#data:  len_b by dose_b
#t = -11.799, df = 36.883, p-value = 4.398e-14
#alternative hypothesis: true difference in means is not equal to 0
#95 percent confidence interval:
# -18.15617 -12.83383
#sample estimates:
#mean in group 0.5   mean in group 2
#           10.605            26.100
```

```
len_c <- ToothGrowth %>% filter(dose %in% c(1,2)) %>% select(len) %>% unlist()
dose_c <- ToothGrowth %>% filter(dose %in% c(1,2)) %>% select(dose) %>% unlist()
test_c <- t.test(len_c~dose_c, paired = FALSE)
test_c

#   Welch Two Sample t-test

#data:  len_c by dose_c
#t = -4.9005, df = 37.101, p-value = 1.906e-05
#alternative hypothesis: true difference in means is not equal to 0
#95 percent confidence interval:
# -8.996481 -3.733519
#sample estimates:
#mean in group 1 mean in group 2
#         19.735          26.100
```

The result show that is a positive relation bettwaen the **dose** level ant theeth length, the p-values are lower than the defaul significance level (.05), so I reject H0).

```
len <- ToothGrowth %>% select(len) %>% unlist()
supp <- ToothGrowth %>% select(supp) %>% unlist()
t.test(len~supp, paired=F)

#   Welch Two Sample t-test

#data:  len by supp
#t = 1.9153, df = 55.309, p-value = 0.06063
#alternative hypothesis: true difference in means is not equal to 0
#95 percent confidence interval:
# -0.1710156  7.5710156
#sample estimates:
```

```
#mean in group OJ mean in group VC
#        20.66333         16.96333
```

In the case the p-value is 0.06, greater the the significance level so the option is reject the null hypothesis.