simulation

Mohamed

24/12/2020

```
library(tinytex)
```

```
## Warning: package 'tinytex' was built under R version 4.0.3
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.3
```

```
library(ggplot2)

set.seed(28) ##to ensure reproducibility, I am setting seed arbitrary on 28.

lambda <- 0.2
```

Part 2 Basic Inferential Data Analysis Instructionsless

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses. The dataset contains data from a study on the Effect of Vitamin C on Tooth Growth in Guinea Pigs.

```
sampleCI <- round (mean(exp_means) + c(-1,1)*1.96*sd(exp_means)/sqrt(1000),3)
cat ("95% confidence interval of my sample : ",sampleCI)
```

```
## 95% confidence interval of my sample :  4.978 5.075
```

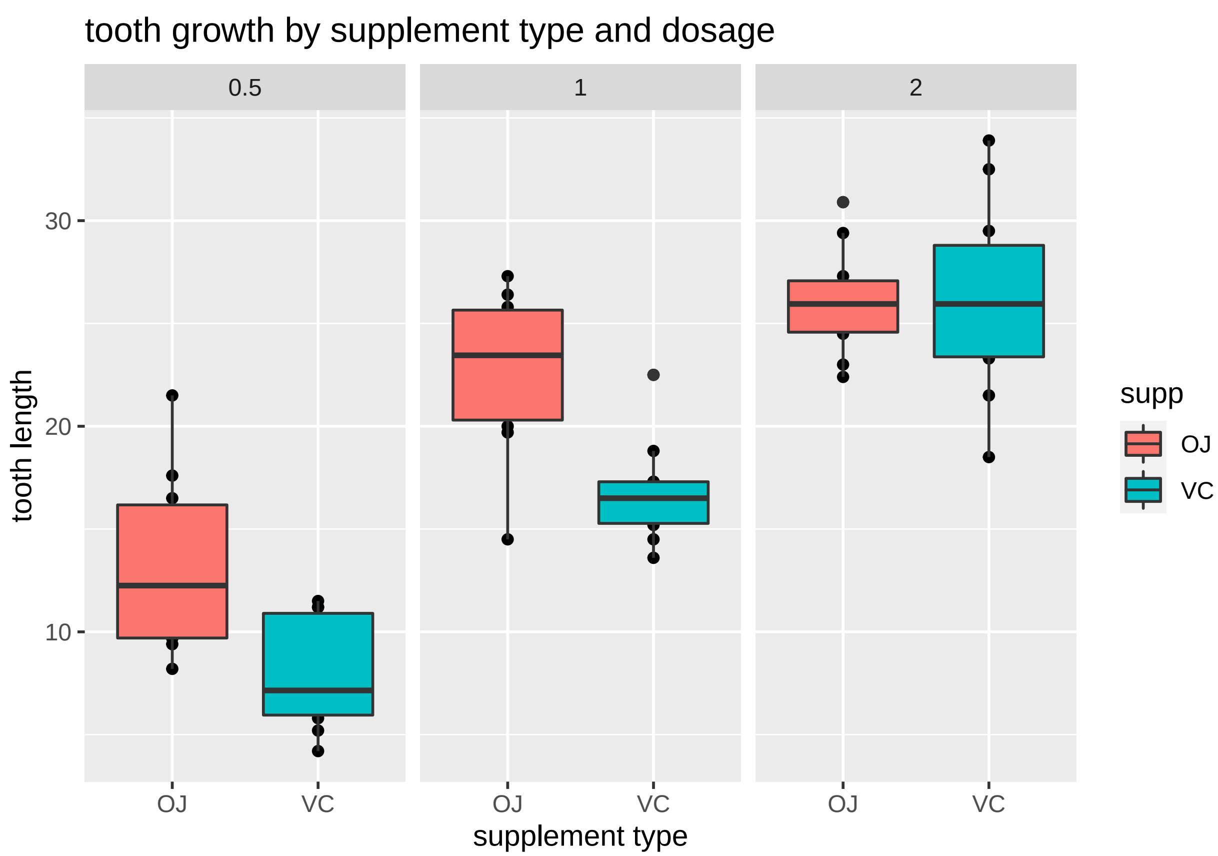
```
theoryCI <- round (5 + c(-1,1)*1.96*0.79/sqrt(1000),3)
cat ("     ; 95% confidence interval in theory : ",theoryCI)
```

```
##      ; 95% confidence interval in theory :  4.951 5.049
```

```
data(ToothGrowth)
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
summary(ToothGrowth)
```

```
##       len         supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```r
qplot(x=supp,y=len,data=ToothGrowth, facets=~dose,
      main="tooth growth by supplement type and dosage",xlab="supplement type", ylab="tooth length") +
      geom_boxplot(aes(fill = supp))
```



2- Provide a basic summary of the data. The summary revelas the dataset consists of 3 variables and 60 observations:

2 numeric variables: lenght (?) and dosage (mg/day)

1 factor variable supp (OJ = Orange Juice or VC = Vitamin C).

3- Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. Before we can do some 2 sample t-testing on the dataset we need to split the data into groups with a level of 2 by supplement OJ and VC:

```r
OJ<-subset(ToothGrowth, ToothGrowth$supp == "OJ")
VC<-subset(ToothGrowth, ToothGrowth$supp == "VC")
dose5<-subset(ToothGrowth, ToothGrowth$dose == 0.5)
dose1<-subset(ToothGrowth, ToothGrowth$dose == 1)
dose2<-subset(ToothGrowth, ToothGrowth$dose == 2)


cat("variance for OJ supp. :",var(OJ$len))
```

```
## variance for OJ supp. : 43.63344
```

```r
cat(";    variance for VC supp. :",var(VC$len))
```

```
## ;    variance for VC supp. : 68.32723
```

```r
cat(";    variance for dose 0.5 :",var(dose5$len))
```

```
## ;    variance for dose 0.5 : 20.24787
```

```r
cat(";   variance for dose 1 :",var(dose1$len))
```

```
## ;   variance for dose 1 : 19.49608
```

```r
cat(";    variance for dose 2 :",var(dose2$len))
```

```
## ;    variance for dose 2 : 14.24421
```

Then we can test whether OJ or VC per similar dosis of x mg/mL have statistical significant differences in mean length (tooth growth):

Dosis of 0.5 mg/mL have a p-value lower than 0.05 which means there is a difference in means. The zero hypothesis can be rejected (when p is low H0 must go…) and there is a significant difference in supplement type with the chosen dosis

```r
t.test(OJ$len,VC$len, var.equal = F, paired = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ$len and VC$len
## 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 of x mean of y
##  20.66333  16.96333
```

```r
t.test(dose5$len,dose1$len, var.equal = T, paired = F)
```

```
##
##  Two Sample t-test
##
## data:  dose5$len and dose1$len
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983748  -6.276252
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

```r
t.test(dose5$len,dose2$len, var.equal = F, paired = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose5$len and dose2$len
## 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 of x mean of y
##    10.605    26.100
```

```r
t.test(dose1$len,dose2$len, var.equal = F, paired = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose1$len and dose2$len
## 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 of x mean of y
##    19.735    26.100
```

```r
OJ5<-subset(ToothGrowth, ToothGrowth$supp == "OJ" & ToothGrowth$dose == 0.5)
VC5<-subset(ToothGrowth, ToothGrowth$supp == "VC" & ToothGrowth$dose == 0.5)
```

Dosis of 1.0 mg/mL have a p-value lower than 0.05 which means there is a difference in means. The zero hypothesis can be rejected (when p is low H0 must go…) and there is a significant difference in supplement type with the chosen dosis:

```r
cat("variance for OJ supp. :",var(OJ5$len))
```

```
## variance for OJ supp. : 19.889
```

```r
cat(";    variance for VC supp. :",var(VC5$len))
```

```
## ;    variance for VC supp. : 7.544
```

```r
t.test(VC5$len, OJ5$len, paired=F, var.equal = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  VC5$len and OJ5$len
## t = -3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.780943 -1.719057
## sample estimates:
## mean of x mean of y
##      7.98     13.23
```

```r
OJ1<-subset(ToothGrowth, ToothGrowth$supp == "OJ" & ToothGrowth$dose == 1)
VC1<-subset(ToothGrowth, ToothGrowth$supp == "VC" & ToothGrowth$dose == 1)
```

Dosis of 2.0 mg/mL have a p-value greater than 0.05 which means there is NOT a difference in means. The zero hypothesis can NOT be rejected and there is NOT a significant difference in supplement type with the chosen dosis:

```r
cat("variance for OJ supp. :",var(OJ1$len))
```

```
## variance for OJ supp. : 15.29556
```

```r
cat(";    variance for VC supp. :",var(VC1$len))
```

```
## ;    variance for VC supp. : 6.326778
```

```r
t.test(VC1$len, OJ1$len, paired=F, var.equal = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  VC1$len and OJ1$len
## t = -4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.057852 -2.802148
## sample estimates:
## mean of x mean of y
##     16.77     22.70
```

```r
OJ2<-subset(ToothGrowth, ToothGrowth$supp == "OJ" & ToothGrowth$dose == 2)
VC2<-subset(ToothGrowth, ToothGrowth$supp == "VC" & ToothGrowth$dose == 2)
```

All types of dosis (0.5 - 2.0 mg/mL) have a p-value lower than 0.05 which means there is a difference in means. The zero hypothesis can be rejected (when p is low H0 must go...) and there is a significant difference in supplement type:

```r
cat("variance for OJ supp. :",var(OJ2$len))
```

```
## variance for OJ supp. : 7.049333
```

```r
cat(";    variance for VC supp. :",var(VC2$len))
```

```
## ;    variance for VC supp. : 23.01822
```

```r
t.test(VC2$len, OJ2$len, paired=F, var.equal = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  VC2$len and OJ2$len
## t = 0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.63807  3.79807
## sample estimates:
## mean of x mean of y
##     26.14     26.06
```

4- State your conclusions and the assumptions needed for your conclusions. The t-test assumes random and independent sampling (paired = FALSE), normality of data distribution, adequacy of sample size, and equality of variance (var.equal = TRUE). From the tests it seems that supplement type have a significant difference in mean tooth length (growth) except when dosis is high (2.0 mg/mL).

A brief conclusion on part2

I have observed that dose and treatments had an effect. However, in the context of this course, I only used really basic tests. A much more correct approach would have been to test properly normality of groups compared and to use a correction as I used multiple comparisons.