**Assignment-IV**

Jitesh Rawat

2/26/2021

**library**(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3 v purrr 0.3.4  
## v tibble 3.0.4 v dplyr 1.0.2  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

**library**(lattice)  
**library**(car)

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

## The following object is masked from 'package:purrr':  
##   
## some

**library**(funModeling)

## Loading required package: Hmisc

## Loading required package: survival

## Loading required package: Formula

##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':  
##   
## src, summarize

## The following objects are masked from 'package:base':  
##   
## format.pval, units

## funModeling v.1.9.4 :)  
## Examples and tutorials at livebook.datascienceheroes.com  
## / Now in Spanish: librovivodecienciadedatos.ai

data <- datasets**::**ToothGrowth  
**glimpse**(data)

## Rows: 60  
## Columns: 3  
## $ len <dbl> 4.2, 11.5, 7.3, 5.8, 6.4, 10.0, 11.2, 11.2, 5.2, 7.0, 16.5, 16...  
## $ supp <fct> VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC...  
## $ dose <dbl> 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.0, 1.0, 1....

data**$**dose <- **as.factor**(data**$**dose)

*#EDA*  
*#Trying to see relation betewwn supplementaires and length*  
**ggplot**(data, **aes**(x=supp, y=len))**+geom\_point**()



*#Length b*  
**require**(ggplot2)  
**ggplot**(ToothGrowth, **aes**(x=**factor**(dose), y=len, fill=supp)) **+**   
 **geom\_boxplot**() **+**  
 **ggtitle**('Tooth Length by Supplement Type and Dosage') **+**  
 **xlab**('dosage') **+**  
 **ylab**('tooth length') **+**  
 **guides**(fill=**guide\_legend**(title='Supplement Type'))



**ggplot**(data, **aes**(x=**factor**(dose),y=len))**+geom\_point**()



*#We see that not much difference in variances of length w.r.t. different doses visually*

*#One way annova*  
  
tdata <- data **%>%**   
 **select**(len,supp)

**summary**(tdata)

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

**ggplot**(tdata)**+**  
 **aes**(supp,len,color=supp)**+**  
 **geom\_jitter**()**+**  
 **theme**(legend.position="none")**+**  
 **labs**(title="supp vs len")



res\_aov <- **aov**(len **~** supp,data= tdata)  
  
*#we check normality with visualistions*  
  
**par**(mfrow=**c**(1,2)) *# combine plots*  
  
*# histogram*  
**hist**(res\_aov**$**residuals)  
  
**qqnorm**(res\_aov**$**residuals,xlab = "norm quantiles")



*#Now we test statistically with shapiro wiks test*  
  
**shapiro.test**(res\_aov**$**residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: res\_aov$residuals  
## W = 0.96949, p-value = 0.1378

*#we see with p-value that data is normal*

**boxplot**(len **~** supp, data=tdata)



**dotplot**( len **~** supp, data =tdata)



**leveneTest**(len **~** supp, data=tdata)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 1 1.2136 0.2752  
## 58

*#From this p-value we can say that we do not reject the null hypothesis that is vairances are equal between supplementaries*

*#One way ANNOVA*  
  
**ggplot**(tdata)**+**  
 **aes**(supp,len,color=supp)**+**  
 **geom\_boxplot**()**+**  
 **theme**(legend.position="none")**+**  
 **labs**(title="Box plots of supplementaries")



*#Some more staistics*  
**group\_by**(tdata,supp) **%>%**   
 **summarise**(  
 mean = **mean**(len,na.rm=TRUE),  
 sd = **sd**(len,na.rm=TRUE)  
 )

## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 2 x 3  
## supp mean sd  
## <fct> <dbl> <dbl>  
## 1 OJ 20.7 6.61  
## 2 VC 17.0 8.27

*#we see that mean and sd of lengths are different w.r.t. supplemenaries*

*#1st method*  
**oneway.test**(len **~** supp,data = tdata,var.equal = TRUE)

##   
## One-way analysis of means  
##   
## data: len and supp  
## F = 3.6683, num df = 1, denom df = 58, p-value = 0.06039

*#2nd method*  
res\_aov <- **aov**(len **~** supp,  
 data = tdata)  
**summary**(res\_aov)

## Df Sum Sq Mean Sq F value Pr(>F)   
## supp 1 205 205.35 3.668 0.0604 .  
## Residuals 58 3247 55.98   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

*#so we observe that the p-value through both tests are same so even in case of equal variances results would not change*

*#Two way ANNOVA*  
  
data <- datasets**::**ToothGrowth

**ggplot**(data)**+**  
 **aes**(supp,len,color=dose)**+**  
 **geom\_jitter**(na.rm = TRUE)**+** *# geom\_point(position = "jitter")*  
 **theme**(legend.position="right")**+**  
 **labs**(title="supp vs len")



*#so we see that visually there is not much difference in lengths w.r.t supplementaries*

twa<- **aov**(len **~** supp**+**dose,data=data)  
  
**par**(mfrow=**c**(1,2)) *# combine plots*  
**plot**(twa,2)



**shapiro.test**(twa**$**residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: twa$residuals  
## W = 0.95538, p-value = 0.02809

*#with shapiro test we see that p-value and we can conclude that data is normal*

**leveneTest**(len **~** supp**\*factor**(dose), data=data)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 5 1.7086 0.1484  
## 54

*#With levene test also we do not reject null hypothesis that means variances are similar*

**summary**(twa)

## Df Sum Sq Mean Sq F value Pr(>F)   
## supp 1 205.4 205.4 11.45 0.0013 \*\*   
## dose 1 2224.3 2224.3 123.99 6.31e-16 \*\*\*  
## Residuals 57 1022.6 17.9   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1