



FACULTY OF ENGINEERING

SCHOOL OF COMPUTING

SECI2143-03 PROBABILITY & STATISTICAL DATA ANALYSIS

CODES OF PROJECT 2

NO	NAME	MATRIC ID
1	ARIF AMIRUDDIN BIN SADIRAN	A19EC0022

LECTURER'S NAME: DR. ARYATI BINTI BAKRI

DATE OF SUBMISSION: 27 JUNE 2020

```

install.packages("stringr")
library(stringr)
install.packages("ggpubr")
library("ggpubr")
install.packages("ggplot2")
library(ggplot2)

#Correlation
#12-16
A<-0
A[1]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`12_lt16_mf`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])
A[2]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...10`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])
A[3]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...11`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])
A[4]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...12`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])
A[5]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...13`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])
A[6]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...14`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])
A[7]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`16_lt18_mf`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])
A[8]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...16`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])
A[9]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...17`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])
A[10]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...18`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])
A[11]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...19`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])
A[12]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...20`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])
A[13]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`18_lt22_mf`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])
A[14]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...22`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])
A[15]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...24`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=="beer")]))[[2]])

```

```

A[16]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...25`[(datasets_2494_4172_
alcohol_german_teens_allGender$`...1`=='beer')]))[[2]])
A[17]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...26`[(datasets_2494_4172_
alcohol_german_teens_allGender$`...1`=='beer')]))[[2]])
A[18]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`22_lt26_mf`[(datasets_2494_
4172_alcohol_german_teens_allGender$`...1`=='beer')]))[[2]])
A[19]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...28`[(datasets_2494_4172_
alcohol_german_teens_allGender$`...1`=='beer')]))[[2]])
A[20]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...30`[(datasets_2494_4172_
alcohol_german_teens_allGender$`...1`=='beer')]))[[2]])
A[21]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...31`[(datasets_2494_4172_
alcohol_german_teens_allGender$`...1`=='beer')]))[[2]])
A[22]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...32`[(datasets_2494_4172_
alcohol_german_teens_allGender$`...1`=='beer')]))[[2]])

for(i in 1:22){A[i]<-(str_replace(A[i], ",", "."))}
Y1<-as.numeric(c(A[1],A[2],A[3],A[4],A[5],A[6]))
Y2<-as.numeric(c(A[7],A[8],A[9],A[10],A[11],A[12]))
Y3<-as.numeric(c(A[13],A[14],A[15],A[16],A[17]))
Y4<-as.numeric(c(A[18],A[19],A[20],A[21],A[22]))
X1<-(c(1,1,1,1,1,1))
X2<-(c(2,2,2,2,2,2))
X3<-(c(3,3,3,3,3))
X4<-(c(4,4,4,4,4))

# Get the input values.
input <- c(Y1,Y2,Y3,Y4)
input2<-c(X1,X2,X3,X4)
TotalX=sum(input2)
TotalY=sum(input)
TotalXY=sum(input2*input)
TotalX2=sum(input2*input2)
TotalY2=sum(input*input)
Correlationr<-((TotalXY-((TotalX*TotalY)/22))/sqrt((TotalX2-
((TotalX*TotalX)/22))*(TotalY2-(TotalY*TotalY)/22)))
Correlationr
# Give the chart file a name.
png(file = "Correlation.png")

plot(x = input2,y = input,
      xlab = "Ages",
      ylab = "Beer Consumption for atleast once a month",
      xlim = c(1,5),
      ylim = c(1,100),
      main = "Beer Consumption for atleast once a month with ages"
)
# Save the file.
dev.off()

```

```

#-----
#-----
#Regression
ObservedY<-mean(input)
SST<-sum((input-ObservedY)*(input-ObservedY))
Ytopi<-c((11.123*input2)+18.239)
SSE<-sum((input-Ytopi)*(input-Ytopi))
SSR<-sum((Ytopi-ObservedY)*(Ytopi-ObservedY))
Determinant<-SSR/SST
SSR<-sum((input))
SS3<-sqrt(SSE/17)
sb<-SS3/(sqrt(TotalX2-((TotalX*TotalX)/22)))
Tregression<-11.123/sb
model <- lm(input~input2)
summary(model)
# Give the chart file a name.
png(file = "Regression.png")

plot(x = input2,y = input,
      xlab = "Ages",
      ylab = "Beer Consumption for atleast once a month",
      xlim = c(1,5),
      ylim = c(1,100),
      main = "Beer Consumption for atleast once a month with ages"
)
abline(model)
# Save the file.
dev.off()
#-----
#-----
#ANOVA(6 Types of Alcohol x Consumption)
#Beer
Y[1]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`12_It26_mf`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='beer')]))[[2]])
Y[2]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...4`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='beer')]))[[2]])
Y[3]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...6`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='beer')]))[[2]])
Y[4]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...7`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='beer')]))[[2]])
Y[5]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...8`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='beer')]))[[2]])
for(i in 1:5){Y[i]<-(str_replace(Y[i], ",", "."))}
AnovaBeer<-as.numeric(c(Y[1],Y[2],Y[3],Y[4],Y[5]))

#Wine
Y[1]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`12_It26_mf`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='wine_sparkWine')]))[[4]])
Y[2]<-

```

```

c((datasets_2494_4172_alcohol_german_teens_allGender$`...4`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='wine_sparkWine'))][[4]])
Y[3]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...6`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='wine_sparkWine'))][[4]])
Y[4]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...7`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='wine_sparkWine'))][[4]])
Y[5]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...8`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='wine_sparkWine'))][[4]])
for(i in 1:5){Y[i]<-(str_replace(Y[i], ",", "."))}
AnovaWine<-as.numeric(c(Y[1],Y[2],Y[3],Y[4],Y[5]))
AnovaWine

#Hard_Liquor
Y[1]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`12_lt26_mf`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='hard_liquor'))][[6]])
Y[2]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...4`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='hard_liquor'))][[6]])
Y[3]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...6`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='hard_liquor'))][[6]])
Y[4]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...7`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='hard_liquor'))][[6]])
Y[5]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...8`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='hard_liquor'))][[6]])
for(i in 1:5){Y[i]<-(str_replace(Y[i], ",", "."))}
AnovaLiquor<-as.numeric(c(Y[1],Y[2],Y[3],Y[4],Y[5]))
AnovaLiquor

#Alcopops
Y[1]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`12_lt26_mf`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='alcopops'))][[8]])
Y[2]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...4`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='alcopops'))][[8]])
Y[3]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...6`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='alcopops'))][[8]])
Y[4]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...7`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='alcopops'))][[8]])
Y[5]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...8`[(datasets_2494_4172_alcohol_german_teens_allGender$`...1`=='alcopops'))][[8]])
for(i in 1:5){Y[i]<-(str_replace(Y[i], ",", "."))}
AnovaAlcopops<-as.numeric(c(Y[1],Y[2],Y[3],Y[4],Y[5]))
AnovaAlcopops
#AlcopopsBeer

```

```

Y[1]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`12_It26_mf`[(datasets_2494_
4172_alcohol_german_teens_allGender$`...1`=='alcopops_beer')]))[[10]])
Y[2]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...4`[(datasets_2494_4172_al
cohol_german_teens_allGender$`...1`=='alcopops_beer')]))[[10]])
Y[3]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...6`[(datasets_2494_4172_al
cohol_german_teens_allGender$`...1`=='alcopops_beer')]))[[10]])
Y[4]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...7`[(datasets_2494_4172_al
cohol_german_teens_allGender$`...1`=='alcopops_beer')]))[[10]])
Y[5]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...8`[(datasets_2494_4172_al
cohol_german_teens_allGender$`...1`=='alcopops_beer')]))[[10]])
for(i in 1:5){Y[i]<-(str_replace(Y[i], ",", "."))}
AnovaAlcopopsBeer<-as.numeric(c(Y[1],Y[2],Y[3],Y[4],Y[5]))
AnovaAlcopopsBeer

#AlcopopsLiquor
Y[1]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`12_It26_mf`[(datasets_2494_
4172_alcohol_german_teens_allGender$`...1`=='alcopops_liquor')]))[[12]])
Y[2]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...4`[(datasets_2494_4172_al
cohol_german_teens_allGender$`...1`=='alcopops_liquor')]))[[12]])
Y[3]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...6`[(datasets_2494_4172_al
cohol_german_teens_allGender$`...1`=='alcopops_liquor')]))[[12]])
Y[4]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...7`[(datasets_2494_4172_al
cohol_german_teens_allGender$`...1`=='alcopops_liquor')]))[[12]])
Y[5]<-
c((datasets_2494_4172_alcohol_german_teens_allGender$`...8`[(datasets_2494_4172_al
cohol_german_teens_allGender$`...1`=='alcopops_liquor')]))[[12]])
for(i in 1:5){Y[i]<-(str_replace(Y[i], ",", "."))}
AnovaAlcopopsLiquor<-as.numeric(c(Y[1],Y[2],Y[3],Y[4],Y[5]))
AnovaAlcopopsLiquor

Anova<-
c(AnovaBeer,AnovaWine,AnovaLiquor,AnovaAlcopops,AnovaAlcopopsBeer,AnovaAlcopo
psLiquor)
Anova

#calculate Mean
meanB<-mean(AnovaBeer)
meanW<-mean(AnovaWine)
meanL<-mean(AnovaLiquor)
meanA<-mean(AnovaAlcopops)
meanAB<-mean(AnovaAlcopopsBeer)
meanAL<-mean(AnovaAlcopopsLiquor)
#calculate Standard Deviation
StandardDeviationB<-sd(AnovaBeer)
StandardDeviationW<-sd(AnovaWine)
StandardDeviationL<-sd(AnovaLiquor)

```

```

StandardDeviationA<-sd(AnovaAlcopops)
StandardDeviationAB<-sd(AnovaAlcopopsBeer)
StandardDeviationAL<-sd(AnovaAlcopopsLiquor)
#Find N
N<-table(c(Y[1],Y[2],Y[3],Y[4],Y[5]))
N<-sum(N)
#Mean Sample
MeanSample<-mean(c(meanB,meanW,meanL,meanA,meanAB,meanAL))
#Standard Deviation Sample
StandardDeviationSample<-sd(c(meanB,meanW,meanL,meanA,meanAB,meanAL))
#Variance between Sample
Var1<-N*StandardDeviationSample
Var1
#Variance within samples
Var2<-
var(c(StandardDeviationB,StandardDeviationW,StandardDeviationL,StandardDeviationA,
StandardDeviationAB,StandardDeviationAL))
#TestStatistic
F<-Var1/Var2
F
#df
M<-
table(c(StandardDeviationB,StandardDeviationW,StandardDeviationL,StandardDeviationA
,StandardDeviationAB,StandardDeviationAL))
M<-sum(M)
num<-(M-1)
denom<-M*(N-1)
#-----
#-----
#Hypothesis Testing
D1<-sum(AnovaBeer)/500
D2<-sum(AnovaWine)/500
D3<-sum(AnovaLiquor)/500
D4<-sum(AnovaAlcopops)/500
D5<-sum(AnovaAlcopopsBeer)/500
D6<-sum(AnovaAlcopopsLiquor)/500
MeanHT<-mean(c(D1,D2,D3,D4,D5,D6))
StandardDeviationHT<-sd(c(D1,D2,D3,D4,D5,D6))
TestStatistic<-(MeanHT-0.12)/(StandardDeviationHT*sqrt(6))

```