SESSION 12: Generalized Linear Models Assignment 2

Problem Statement

1. Use the given link below:

https://archive.ics.uci.edu/ml/machine-learning-databases/communities/

Perform the below operations:

#1

#Answer the below questions:

a. Visualize the correlation between all variables in a meaningful and clear way of representing. Find out

top 3 reasons for having more crime in a city.

b. What is the difference between co-variance and correlation? Take an example from this dataset and

#show the differences if any?

#Answer1

#a)

#visualize

#using crimes dataset

```
main_data<- Crimes[,c(11,12,13,14,16,17,17,20,21)]

library(corrplot)

corrplot(cor(main_data),type = "full","pie")

corrplot(cor(main_data),type="full","number")

corrplot(cor(main_data),type="full","shade")
```

#visualize

#using mtcars dataset

```
main_mtcars<- subset(mtcars,select = c(2:12))

main_mtcars
```

#using mtcars [dataset]

```
#using correlation plot
library(corrplot)
#create a correlations matrix
#create a correlations matrix
main <- cor(main_mtcars)

# First Correlogram Example
library(corrgram)
corrgram(main, order=TRUE, lower.panel=panel.shade,
upper.panel=panel.pie, text.panel=panel.txt)

#represent correlations
corrplot(cor(main),type = "full","circle")

corrplot(cor(main),type = "full","number")

corrplot(cor(main),type = "full","pie")

corrplot(cor(main),type = "full","ellipse",
order = 'original')

corrplot(cor(main),type = "full","ellipse",
order = 'alphabet',diag = TRUE)

#2 part answer already in 1st assignment given

#b)
#A measure used to indicate the extent to which two random variables change in tandem is known as covariance. A measure used to
#represent how strongly two random variables are related known as correlation

#Covariance is nothing but a measure of correlation. On the contrary,
#correlation refers to the scaled form of covariance

#The value of correlation takes place between -1 and +1.
#Conversely, the value of covariance lies between -infi and +infi
```

#Covariance is affected by the change in scale, i.e. if all the value of one variable is multiplied

#by a constant and all the value of another variable are multiplied, by a similar or different constant, then the covariance is changed.

#As against this, correlation is not influenced by the change in scale

#Correlation is dimensionless, i.e. it is a unit-free measure of the relationship between variables. Unlike covariance,

#where the value is obtained by the product of the units of the two variables

#Covariance

#The covariance of two variables x and y in a data set measures how the two are linearly related. A positive covariance would

#indicate a positive linear relationship between the variables,

#and a negative covariance would indicate the opposite

#Correlation Coefficient

#The correlation coefficient of two variables in a data set equals to their covariance divided by the product of their individual standard deviations.

#It is a normalized measurement of how the two are linearly related.

#If the correlation coefficient is close to 1, it would indicate that the variables are positively linearly related and the scatter plot falls almost along a

#straight line with positive slope. For -1, it indicates that the variables are negatively linearly related and the scatter plot almost falls along a straight line

#with negative slope. And for zero, it would indicate a weak linear relationship between the variables.

#using mtcars dataset

#correlation test

```
mymain_data <- mtcars

res <- cor.test(mymain_data$wt, mymain_data$mpg,

method = "pearson")

res
```

#The p-value of the test is 1.29410^{-10}, which is less than the significance level alpha = 0.05. We can conclude that wt and mpg are significantly correlated

#with a correlation coefficient of -0.87 and p-value of 1.29410^{-10} .

```
# Correlations/covariances among numeric variables in
# data frame mtcars,Use listwise deletion of missing data.
cor(mtcars$mpg,mtcars$wt,method = 'spearman')
cor.test(mtcars$mpg,mtcars$wt,method = 'spearman')
cov(mtcars$mpg,mtcars$wt,method = 'spearman')

cor(mtcars$mpg,mtcars$wt,method = 'pearson')
cor.test(mtcars$mpg,mtcars$wt,method = 'pearson')
cov(mtcars$mpg,mtcars$wt,method = 'pearson')
```