GROUP ASSIGNMENT # 03

STATISTICS AND PROBABILITY THEORY

NAMES	REGISTRATION
SUDAIS KHAN	FA20-BSE-042
ASADULLAH	FA20-BSE-028
UBAID UR REHMAN	FA20-BSE-012
SECTION	BSE-5B

- **Submitted To Sir Atta Ullah**

Question 1: Discuss correlation and regression. Explain different types of correlation and regression and its applications.

Correlation:

Correlation refers to the statistical relationship between two entities. In other words, it's how two variables move in relation to one another.

Correlation is a statistical measure that expresses the extent to which two variables are linearly related (meaning they change together at a constant rate)

<There are two types of correlation>

Positive correlation: A positive correlation would be 1. This means the two variables moved either up or down in the same direction together.

Negative correlation: A negative correlation is -1. This means the two variables moved in opposite directions.

Regression:

In Which we check the dependences of a dependent variable on another variable.

There are dependent variable and independent variable:

Dependent variable:

Whose value depends on another variable.

Independent variable:

When one or more variables are dependent.

Correlation Applications

1)Automobile Sector:

The price of fuels affects the sale of car.

2)Banking Sector:

The interest rate affects the Economic growth of country.

3)Health care Sector:

Heath recovery can be analyzed with respect to time of physiotherapy exercise.

Regression Applications

Regression analysis is the estimation of the ratio between two variables. Say you want to estimate the growth in meat sales (MS Growth), based on economic growth (GDP Growth).

Question 2: (Using Python) Find correlation coefficient and fit the regression model Y on X1, X2, X3 and predict your final marks using quiz, assignment and Mid mark.

Description: In this step will read the data from the csv file.

```
[1] from google.colab import drive
```

```
[2] drive.mount('/content/gdrive')
```

```
Mounted at /content/gdrive
```

```
[3] import pandas as pd
```

```
[10] #Importing the libraries
     import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns
```

```
[33] dataset=pd.read_csv('/content/gdrive/My Drive/assign.csv')
```

```
[84] dataset.head()
```

	S,NO	GENDER	QUIZ\n(X1)	ASSIGN\n(X2)	MID\n(X3)	FINAL\n(Y)	TOTAL	LG	GP
0	1	F	7	18	22	25	77	B	3
1	2	M	11	15	16	20	69	C+	2.3
2	3	F	27	13	27	16	51	D	1.3
3	4	M	12	16	19	16	68	C+	2.3
4	5	F	7	16	20	28	82	B+	3.3

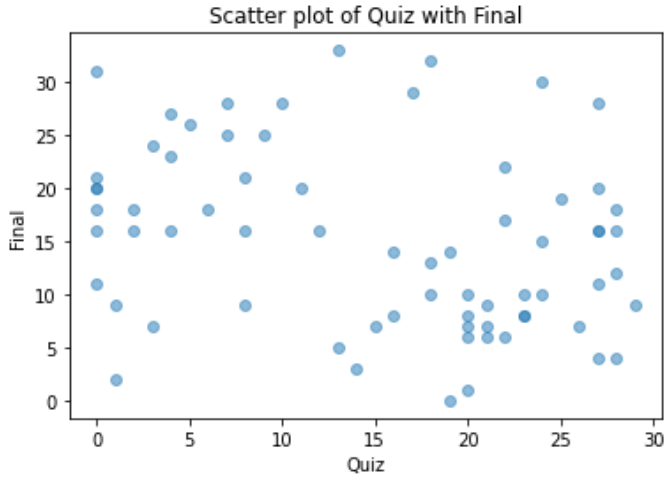
```
[85] dataset.shape
```

```
(69, 9)
```

Description: Here making the scatter plot of Quiz and Final Marks.

```
[86] plt.scatter(dataset['QUIZ\n(X1)'],dataset['FINAL\n(Y)'],alpha=0.5)
     plt.title("Scatter plot of Quiz with Final")
     plt.xlabel("Quiz")
     plt.ylabel("Final")
     plt.show
```

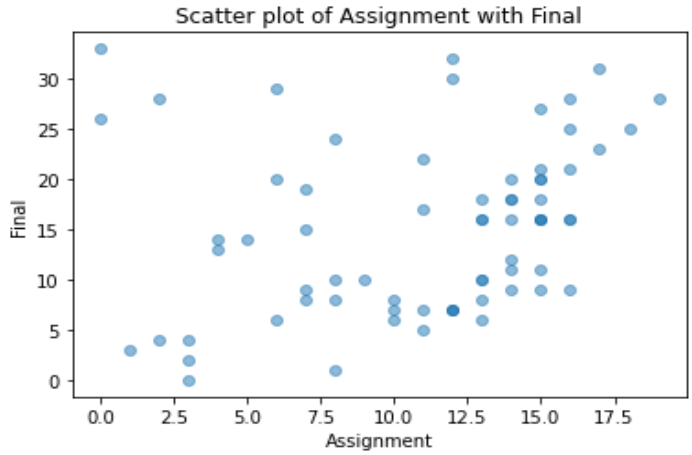
```
<function matplotlib.pyplot.show(*args, **kw)>
```



Description: Here making the scatter plot of Assign and Final Marks.

```
[87] plt.scatter(dataset['ASSIGN\n(X2)'],dataset['FINAL\n(Y)'],alpha=0.5)
     plt.title("Scatter plot of Assignment with Final")
     plt.xlabel("Assignment")
     plt.ylabel("Final")
     plt.show
```

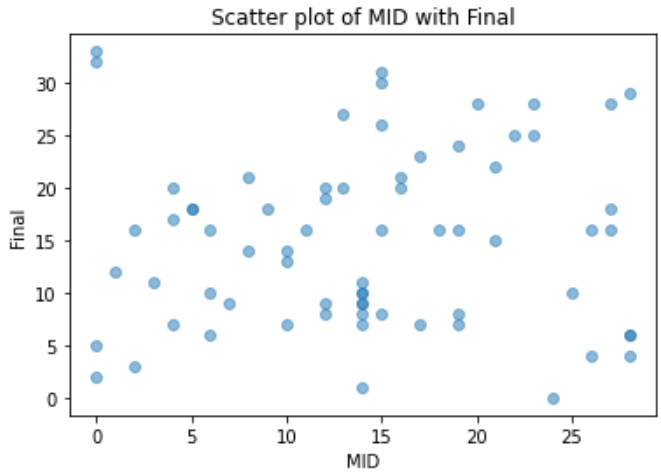
```
<function matplotlib.pyplot.show(*args, **kw)>
```



Description: Here making the scatter plot of Mid Marks and Final Marks.

```
[88] plt.scatter(dataset['MID\n(X3)'],dataset['FINAL\n(Y)'],alpha=0.5)
     plt.title("Scatter plot of MID with Final")
     plt.xlabel("MID")
     plt.ylabel("Final")
     plt.show
```

```
<function matplotlib.pyplot.show(*args, **kw)>
```



Description: Here Setting the value to Dependent Variable and Independent Variable.

```
[89] #Setting the value for X and Y
     x = dataset[['QUIZ\n(X1)', 'ASSIGN\n(X2)', 'MID\n(X3)']]
     y = dataset['FINAL\n(Y)']
```

```
[90] from sklearn.model_selection import train_test_split
```

```
[91] X_train,X_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=0)
```

```
[92] from sklearn.linear_model import LinearRegression
     #Fitting the Multiple Linear Regression model
     mlr = LinearRegression()
     mlr.fit(X_train, y_train)
```

```
LinearRegression()
```

Description: Here Predicted the Quiz marks Assignment Marks and Mid Marks.

```
[94] #Intercept and Coefficient
     print("Intercept: ", mlr.intercept_)
     print("Coefficients:")
     list(zip(x, mlr.coef_))
```

```
Intercept:  13.320188543465171
Coefficients:
[('QUIZ\n(X1)', -0.16422901836164294),
 ('ASSIGN\n(X2)', 0.27782467798180455),
 ('MID\n(X3)', 0.14783414288449956)]
```

```
#Prediction of test set
y_pred_mlr= mlr.predict(X_test)
#Predicted values
print("Prediction for test set: {}".format(y_pred_mlr))
```

```
Prediction for test set: [13.83519655 13.4819315  13.55998907 15.8677386  13.87741777 16.78898194
 19.49165721 15.71700699 13.68301621 14.20587581 13.1211131  13.85883508
 17.80107061 19.11373428]
```