Practical 11:
Multiple Regression- Prediction of salary

Question:
Mike Wilde is president of the teachers' union for Otsego School District. In preparing for upcoming negotiations, he would like to investigate the salary structure of classroom teachers in the district. He believes there are three factors that affect a teacher's salary: years of experience, a rating of teaching effectiveness given by the principal, and whether the teacher has master's degree. A random sample of 20 teachers resulted in the data below.

Salary ($ thousands) (Y)	Years of experience (X1)	Principals rating (X2)	Master's degree (X3)
21.1	8	35	0
23.6	5	43	0
19.3	2	51	1
33	15	60	1
28.6	11	73	0
35	14	80	1
32	9	76	0
26.8	7	54	1
38.6	22	55	1
21.7	3	90	1
15.7	1	30	0
20.6	5	44	0
41.8	23	84	1
36.7	17	76	0
28.4	12	68	1
23.6	14	25	0
31.8	8	90	1
20.7	4	62	0
22.8	2	80	1
32.8	8	72	0

o   Develop a correlation matrix.
o   Which independent variable has the strongest correlation with the dependent variable?
o   Identify the best set of predictors using as measure of goodness of fit.
o   Define the regression equation for the variables.
o   How much the predictors are effective to explain variation in the salary (use to answer this question)?

Objective:
To predict the salary structure of classroom teachers in the district

Methodology:
- o Scatter plot: A scatter plot (also called a scatterplot, scatter graph, scatter chart, scattergram, or scatter diagram) is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data
- o Correlation
- o
- o Multiple regression is a statistical technique that can be used to analyse the relationship between a single dependent variable and several independent variables. The objective of **multiple regression** analysis is to use the independent variables whose values are known to predict the value of the single dependent value.

- o $Y_i = \beta_o + \beta_1 X_{i1} + \beta_2 X_{i2} + .. + \epsilon$
Where, i = n observations

$$Y_i = dependent\ variable$$
$$X_i = expanatory\ variable$$
$$\beta_o = y\ intercept$$

- o $a = \dfrac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$
- o $b = \dfrac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$

Analysis:
We analyzed the given data to predict the salary structure of classroom in the district using multiple regression on MS Excel 2019.

Using regression formulas, we calculated,

Intercept	10.11570487
Years of experience (X1)	0.892648174
Principals Rating (X2)	0.146380408

Y, X1	0.7527	R_(Y.1)^2	0.7527
Y, X2	0.2994	R_(Y.2)^2	0.2994
Y, X3	0.9644	R_(Y.3)^2	0.0964
Y, X1,X2	0.9063	R_(Y.12)^2	0.9063
Y,X1,X3	0.7705	R_(Y.13)^2	0.9063
Y,X2,X3	0.3039	R_(Y.23)^2	0.7705
Y,X1,X2,X3	0.9081	R_(Y.123)^2	0.30339

Thus, we get regression equation:
$$\hat{Y} = 10.1157 + 0.8926 X_1 + 0.1463 X_2$$

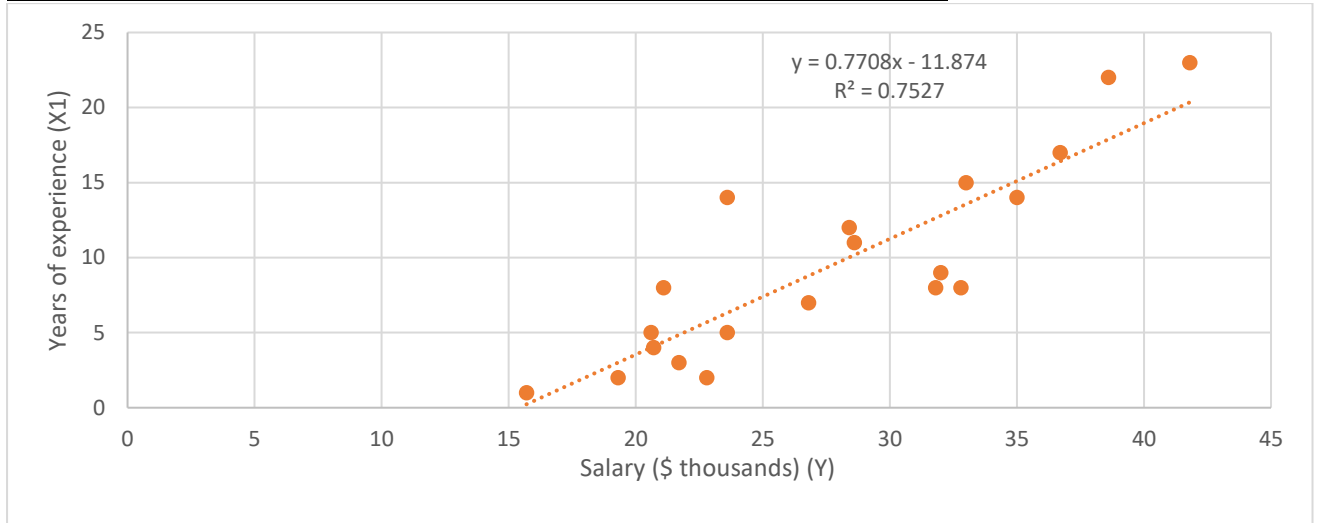Figure 1: Scatter plot Years of experience (X1) v/s Salary ($ thousands) (Y)

$y = 0.7708x - 11.874$
$R^2 = 0.7527$

Figure 2: Scatter plot Principals rating (X2) v/s Salary ($ thousands) (Y)

$y = 1.4917x + 21.036$
$R^2 = 0.2994$
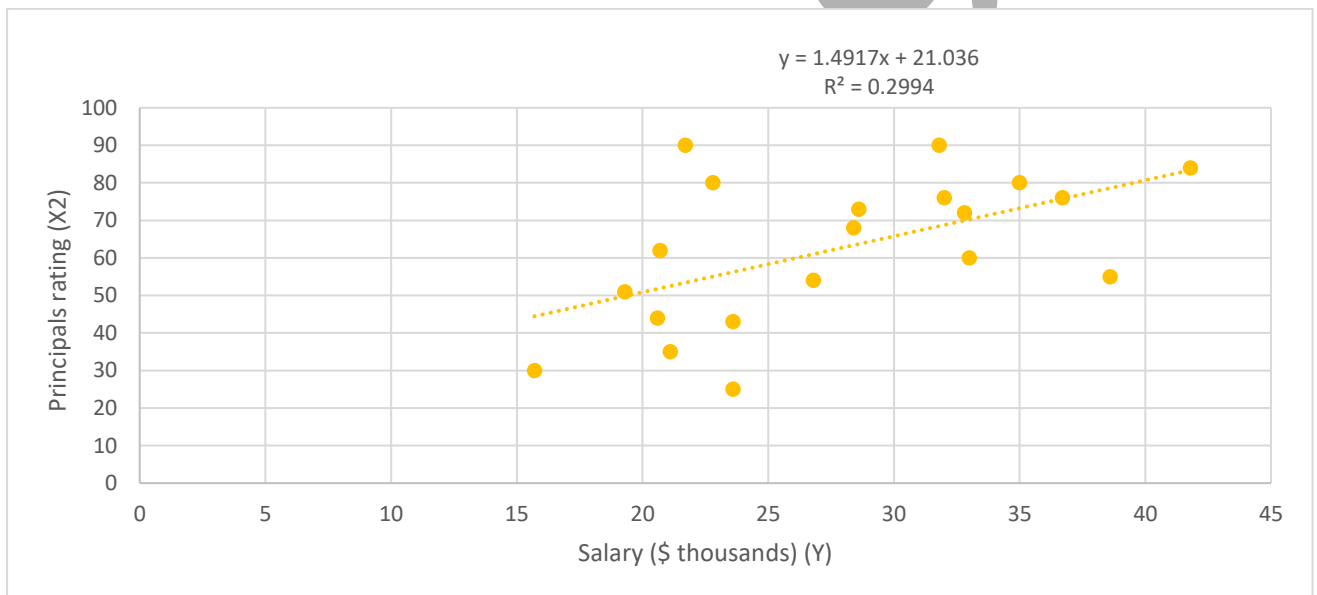
Figure 3: Scatter plot Master's degree (X3) v/s Salary ($ thousands) (Y)

$y = 0.022x - 0.1106$
$R^2 = 0.0964$
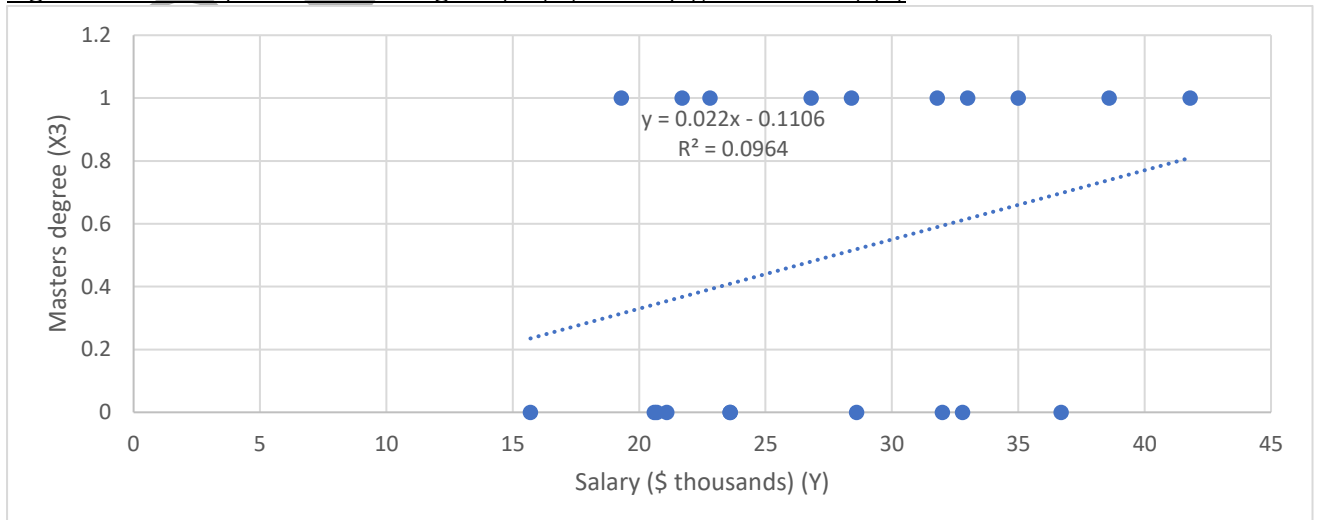
Table 1: Correlation Matrix

	Salary ($ thousands) (Y)	Years of experience (X1)	Principals rating (X2)	Master's degree (X3)
Salary ($ thousands) (Y)	1			
Years of experience (X1)	0.867597028	1		
Principals rating (X2)	0.547196687	0.186823696	1	
Master's degree (X3)	0.310558329	0.20750259	0.457774867	1

Interpretation:

- Regression equation
$$\hat{Y} = 10.1157 + 0.8926X_1 + 0.1463X_2$$
$$R^2 = 0.9064$$
- Intercept Term = 10.1157
$$X_1 = 0; \ X_2 = 0$$
- The average salary is $10115.7 of an individual with no experience and 0 principals rating is $10115.7.
- Regression coefficient of Y on X_1 keeping X_2 fixed = 0.8926.
  This tells that the average salary of a person will increase by $892.6 with a year of increase in the experience, keeping the principal rating fixed.
- Regression coefficient of Y on X_2 keeping X_1 fixed =0.1463.
  For the persons having same year of experience, their salary increases by $146.3 as the rating by principal increases by a unit.

$$R^2 = 0.9064$$
- The value of R^2 says that years of experience and principal's rating explain 90.64% of variation in the salary.

Excel Options:
- To make scatter plot, select the given data and go to insert. Select the desired graph and add titles. Your graph is created.
- To add trend line, select the intersection points on the graph and right click and select add trendlines and the type of trendline (linear). Select Display equation on chart and display R-squared value on chart.
- For data analysis, go to data analysis, select correlation and then select the data. Your correlation matrix is created.

Submitted by:-Tanya
Smart id-abbsm21198
Roll no-2111793
Submitted to-Dr.Gargi tyagi