For cited: *Al-Khaiat, S. S. J., Noori, M. Z. & Cengiz, M. A. (2022). Application of the Regression Analysis in Python, SPSS and Microsoft Excel Programs. Journal of Current Researches on Educational Studies, 12 (2), 27-46.*

Journal of Current Researches on Educational Studies (JoCuRES)
www.jocures.com

Year: 2022
Volume: 12
Issue: 2

Crossref doi: 10.26579/jocures.12.2.3

Research Article

Application of the Regression Analysis in Python, SPSS and Microsoft Excel Programs

Shubar Sahib Jawad AL-KHAİAT[1], Mohammed Zaki NOORİ[2] & Mehmet Ali CENGİZ[3]

Keywords
Machine Learning, Regression Analysis, Python, Spss, Microsoft Excel.

Article History
Received
16 May, 2022
Accepted
30 Dec, 2022

Abstract
Regression analysis has been one of the most used statistical methods during the past fifty years to analyze relationships between variables. Due to its flexibility, usefulness, applicability, theoretical and technical analysis, regression analysis has become an essential statistical tool for solving problems in the real world. In order to apply these scientific techniques successfully and effectively, one needs a vision and an understanding of both basic theory and its practical application in the various fields of engineering, physical sciences, mathematical sciences and statistics. So it blends theory and practice so it has a sufficiently deep understanding of the basic principles needed to apply regression building techniques.

First, we will present a comprehensive study, but rather simple, study to basic regression knowledge by discussing how regression analysis works, the requirements and assumptions it relies on, and how we can define a regression analysis model that allows us to make critical business, customer, or project decisions. Each step of regression analysis is related to its implementation in several advanced and various computer programs such as Python, SPSS, and Microsoft Excel. We will show how to use a set of easy-to-learn statistical procedures for these programs that underlie regression analysis, which will allow us to analyze, map, validate regression results and assess the strength of the analysis. Interpreting the outputs may be somewhat difficult, but we will try to make this easier by studying an annotated case within the use, application, and interpretation of regression analysis.

1. Introduction

Regression analysis is a set of statistical methods that serve as a basis for drawing basic information about the relationships between variables. These techniques are applicable in every field of study, including social, physical, biological and business sciences. Regression analysis may be now, the statistical scientific method used for all methods of data analysis. Hence, its aim is to develop the basic theory of this important statistical method and to clarify the theory with a variety of selected

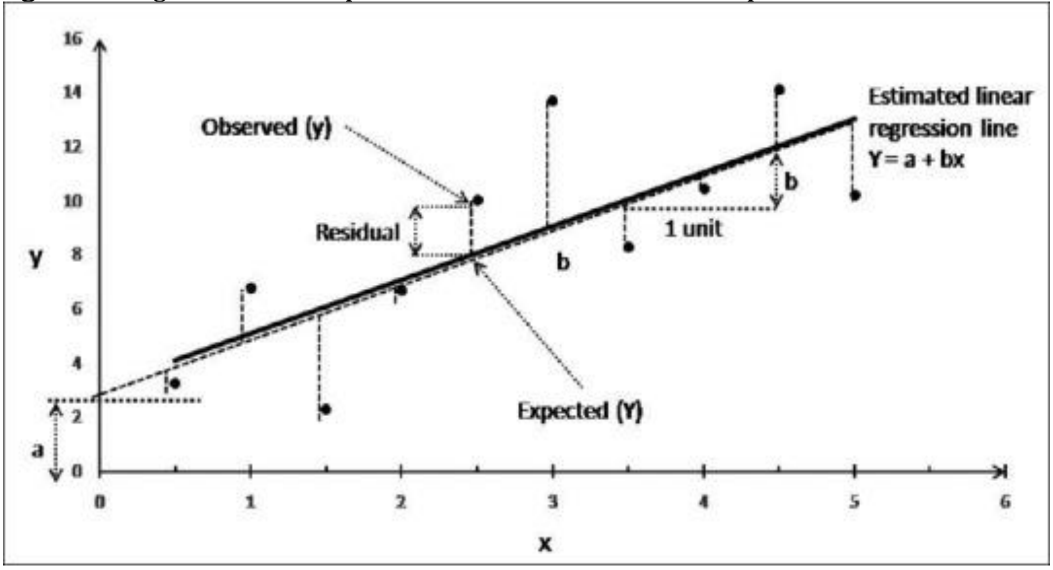
[1] Corresponding Author. ORCID: 0000-0002-3113-4322. Ondokuz Mayis University, Samsun, Turkey, shub712000@yahoo.com
[2] ORCID: 0000-0002-8892-7171. Ondokuz Mayis University, Samsun, Turkey, mahammadzeki97@gmail.com
[3] ORCID: 0000-0003-4972-167X. Ondokuz Mayis University, Samsun, Turkey, macengiz@omu.edu.tr

examples from economics, demography and engineering. Golberg, M. A., & Cho, H. A. (2004).

In statistics, the regression is the representation of the relationship between two variables in an analytical algebraic form. This concept has been called regression because it is regressive in estimating the values of the variables towards averages. The regression analysis is used to estimate relationships between two or more variables. The variable that can be expected is called the dependent variable or response variable (Y). The variable that we use to predict the value of the other variable is called the independent variable or predictor variable (X). Regression analysis helps to understand how the dependent variable changes when one of the independent variables changes and allows identifying which of these variables actually has an effect. The relationship between a dependent variable and one or more independent variable can be expressed using a linear function. Linear regression models are fairly simple models and provide an easy-to-interpret mathematical formula that can make predictions. Linear regression can be applied to various fields of business and academic study. David A. Freedman (2009).

Regression analysis can be defined as a graph of data and drawing the best fit line through it. This "most appropriate" line is chosen so that the sum of the squares of all the remaining values is the minimum - the so-called "least squares line" as in the figure below, and this line can be defined mathematically by the equation of the model (Fig. 1)

Figure 1. Regression line representation of the observed, expected, and residual data



$$\mathbf{Y = a + b\,x + e}$$

whereas:

y: the expected value (or dependent variable, output, or response).

x: the independent variable (or predictor or explanatory).

a: the intersection of the regression line with the y-axis (constant) and represents the value of y when x = 0.

b: the regression coefficient (slope of the regression line).

e: random error (the effect of random factors on the dependent variable (y)). The sum of squares of (e) is also called the residual, which is an important indicator for judging the appropriateness of a linear regression

Also note the value of "b" is positive when the value of (y) increases with an increase in each unit in x and it is negative if the value of (y) decreases with the increase in each unit in x. (Aggarwal, & Ranganathan, 2017).

2. Objectıves

The objective of the regression study is useful in:

- Describe the degree of the relationship between the variables, whether it is a positive or an inverse relationship
- Deduce the algebraic mathematical relationship (functional relationship) between variables
- Draw graphical lines that show the relationship between the variables
- Predicting the values of one of the variables if the other variable is known and its relationship to that variable. Chatterjee, S., & Hadi, A. S. (2015).

3. Methods and Materials

In this scientific paper we will discuss the use of linear regression application in a group of various scientific computer programs such as Microsoft Excel and the use of statistical software packages such as IBM SPSS® Statistics and Python that greatly simplify the process of using linear regression equations, linear regression models and linear regression formula. These programs can also be used in techniques such as simple linear regression and multiple linear regression.

3.1. Application of Regression with Python

In this paper, we will focus on implementing regression models and mainly its application in Python. Linear regression is a statistical model that examines the linear relationship between two variables (simple linear regression) or (multiple linear regression) - dependent variable and independent variables. The linear relationship basically means that when the value of the independent variables increases or decreases, the dependent variable increases or decreases also. The linear relationship can also be positive (the independent variable rises, the dependent variable rises) or negative (the independent variable rises, the dependent variable decreases).

There are two main ways to perform linear regression in Python - using Statmodels and scikit-learn.

Linear Regression in Statmodels:

Statmodels is a " important Python module that provides classes and functions for estimating many different statistical models, as well as for performing statistical tests, and for exploring statistical data." As with NumPy and Pandas, the easiest way to get or install *Statmodels is through the Anaconda package.*

Linear Regression in Sk-learn

Sk-learn is pretty much the standard when it comes to machine learning in Python. It has several learning algorithms, for regression, classification, clustering and dimension reduction.

In practice, we have not used the entire data set, but will divide the data into training data to train the model on, and test data to test the model (prediction). (Bronshtein, 2017).

4. Results and Descutions

For the following reasons, using a linear regression model is critical:

- Descriptive: It aids in determining the strength of the relationship between the dependent variable and the predictor factors.
- Adjustment: It takes into account the impact of variables and confounders.
- Predictors: It aids in the estimation of significant elements that influence the dependent variable.
- The amount of prediction: It aids in determining how a one-unit change in the independent variable affects the dependent variable.
- Prediction: It aids in the quantification of new cases.

The link between the independent and technically dependent variables is described in regression analysis. The value of a dependent variable can be predicted using regression analysis if at least one independent variable is known.

The correlation coefficient "r" is a dimensionless number that ranges from -1 to +1 in correlation analysis. A value closer to 1 denotes a negative or inverse association, whereas a value closer to +1 denotes a positive relationship. The Pearson correlation is used when the data has a normal distribution, while the Spearman rank correlation is used when the data has an aberrant distribution. Kumari, (2018)

In regression analysis, the P-values and coefficients work together to inform you whether correlations in your model are statistically significant and what those relationships are like. The coefficients show how each independent variable and the dependent variable are related mathematically. The coefficients' p-values show whether or not these associations are statistically significant.

Check the remaining plots first after fitting the regression model to ensure we have unbiased estimates. It's now time to interpret the statistical results. Linear regression analysis can yield a plethora of outcomes, which I will assist you in navigating.

Example problem statement for simple linear regression:

We will take a dataset with two variables: sales value (dependent variable) and Advertising costs (independent variable). The objectives of this problem are:

We want to see if there is any correlation between these two variables

We will find the best suitable line for the data set.

Notice how the dependent variable is changing by changing the dependent variable?

We will implement a simple linear regression model in machine learning using Python. Jaiswal, S. (2020).

Example: The following data are the advertising costs for marketing a specific commodity and the dollar value of sales for one of the companies producing this commodity.

Table 1. Table of Advertising costs and Sales value.

Index	Advertising costs (x)	Sales value (y)
1	25	1015
2	6	296
3	8	300
4	17	688
5	2	28
6	13	506
7	23	914
8	30	1186
9	28	1169
10	14	604
11	19	757
12	4	174
13	24	938
14	1	16
15	5	184

When we apply these algorithms in Python with regression analysis, we will write the hypothesis first and then apply in Python:

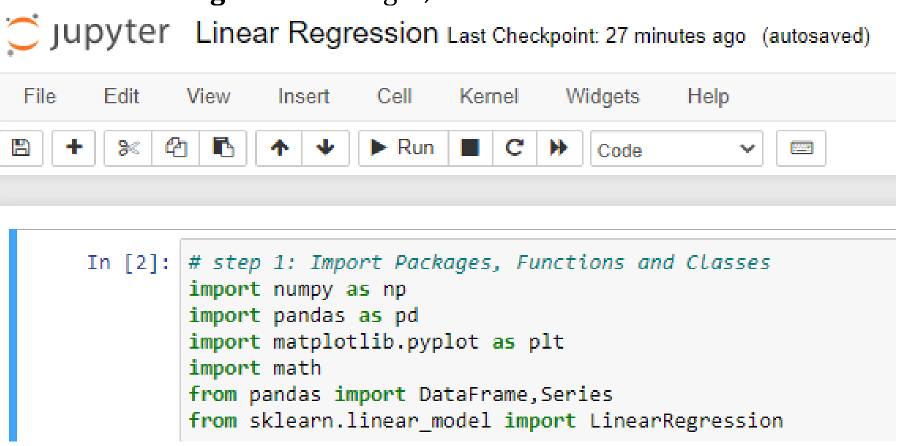
$$\mathbf{Y} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\, \mathbf{x} + \mathbf{e}$$

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

Step 1: Import Packages, Functions and Classes.

Figure 2. Packages, Functions and Classes



Step 2: Get Data.

Figure 3. The mean values of (x, y).

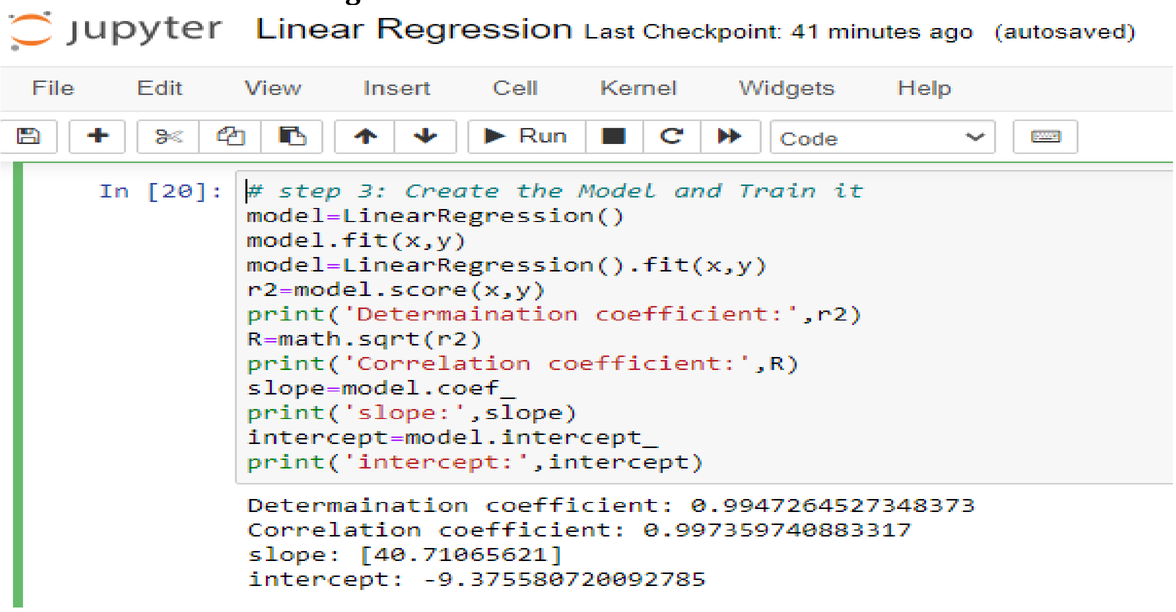
```
jupyter   Linear Regression Last Checkpoint: 31 minutes ago  (autosaved)

In [5]: # step 2: Get Data
x=np.array([25,6,8,17,2,13,23,30,28,14,19,4,24,1,5]).reshape(-1,1)
y=np.array([1015,296,300,688,28,506,914,1186,1169,604,757,174,938,16,184])
print(x,y)
mean_x=np.mean(x)
mean_y=np.mean(y)
print('mean_x:',mean_x)
print('mean_y:',mean_y)

[[25]
 [ 6]
 [ 8]
 [17]
 [ 2]
 [13]
 [23]
 [30]
 [28]
 [14]
 [19]
 [ 4]
 [24]
 [ 1]
 [ 5]] [1015  296  300  688   28  506  914 1186 1169  604  757  174  938   16
  184]
mean_x: 14.6
mean_y: 585.0
```

From the above figure, we get the mean values of (x) = 14.5

And the mean value of (y) = 585

Step 3: Create the Model and Train it.

Figure 4. Create the Model and Train it.



```
jupyter   Linear Regression Last Checkpoint: 41 minutes ago  (autosaved)

In [20]: # step 3: Create the Model and Train it
model=LinearRegression()
model.fit(x,y)
model=LinearRegression().fit(x,y)
r2=model.score(x,y)
print('Determaination coefficient:',r2)
R=math.sqrt(r2)
print('Correlation coefficient:',R)
slope=model.coef_
print('slope:',slope)
intercept=model.intercept_
print('intercept:',intercept)

Determaination coefficient: 0.9947264527348373
Correlation coefficient: 0.997359740883317
slope: [40.71065621]
intercept: -9.375580720092785
```

From the figure 4, We get the following:

Coefficient of Determination = 0.995, it means that the value of the independent variable (x) in the simple linear regression equation can explain the amount 0.995 of the change in the dependent variable (y).

Correlation Coefficient = 0.997, it means that there is a very strong correlation between the independent variable (x) and the dependent variable (y).

$\beta_1 = 40.71$, it means the Slope coefficient of the regression line equation

$\beta_0 = -9.37$, it means the intercept coefficient of the regression line equation

Step 4: Evaluate the Model:

Figure 5. Estimating the Model

```
In [10]: # step 4: Evaluate the Model
         pred=model.predict(x)
         print('y_pred:',pred)

y_pred: [1008.39082462  234.88835656  316.30966899  682.70557491   72.04573171
  519.86295006  926.9695122  1211.94410569 1130.52279326  560.57360627
  764.12688734  153.46704413  967.68016841   31.33507549  194.17770035]

In [12]: # step 4: Evaluate the Model (other way)
         pred_y=model.intercept_+model.coef_*x
         print('pred_y:',pred_y)

pred_y: [[1008.39082462]
 [ 234.88835656]
 [ 316.30966899]
 [ 682.70557491]
 [  72.04573171]
 [ 519.86295006]
 [ 926.9695122 ]
 [1211.94410569]
 [1130.52279326]
 [ 560.57360627]
 [ 764.12688734]
 [ 153.46704413]
 [ 967.68016841]
 [  31.33507549]
 [ 194.17770035]]
```

From the figure 5, We get the Predictive values of the regression line equation as explained above.

And we can also obtain the values of the coefficients from the package (scipy & stats) as follows:

Figure 6. Calculate some statistical values.

```
In [22]: # Another way for the same example:
         from scipy import stats
         x=(25,6,8,17,2,13,23,30,28,14,19,4,24,1,5)
         y=(1015,296,300,688,28,506,914,1186,1169,604,757,174,938,16,184)
         slope,intercept,r_value,p_value,std_err=stats.linregress(x,y)
         print('slope:',slope)
         print('intercept:',intercept)
         print('r_value:',r_value)
         print('r2:',r_value**2)
         print('p_value:',p_value)
         print('std_err:',std_err)

slope: 40.71065621370499
intercept: -9.375580720092898
r_value: 0.997359740883317
r2: 0.9947264527348373
p_value: 3.3985057892769943e-16
std_err: 0.8221215160305974
```

From all the above figures (3,4,5,6), We will get some values for the regression equation like:

The values of the dependent variable (y) = (1015,296,300,688,28,506,914,1186,1169,604,757,174,938,16,184)

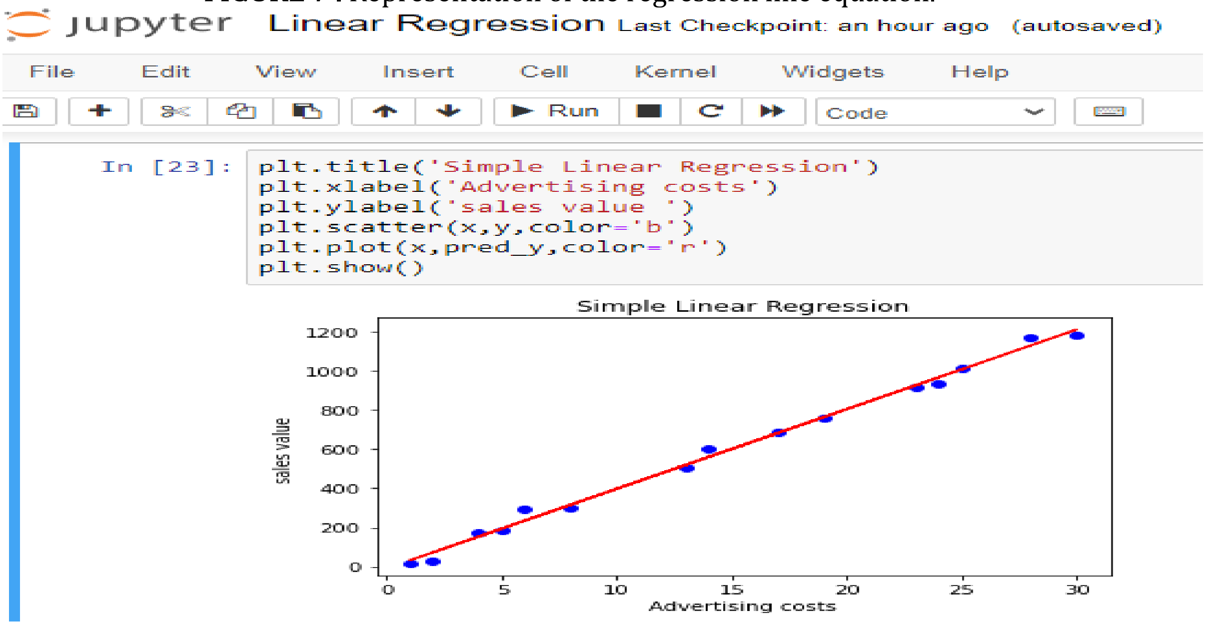
The values of the independent variable (x) = (25,6,8,17,2,13,23,30,28,14,19,4,24,1,5)

Mean (x) = 14.6, Mean (y) = 585.0, Correlation coefficient = 0.99735, Determination coefficient = 0.99472,

Intersection = -9.3755, The slope of the regression line = 40.710, Standard error = 0.82212, P. value = 3.3985

In the case of drawing the equation of the regression line, we will obtain the following:

FIGURE 7 . Representation of the regression line equation.



And from all of the above, we note that there is a strong direct correlation between the independent variable (Advertising costs (x)) and the dependent variable (Sales value (y)).

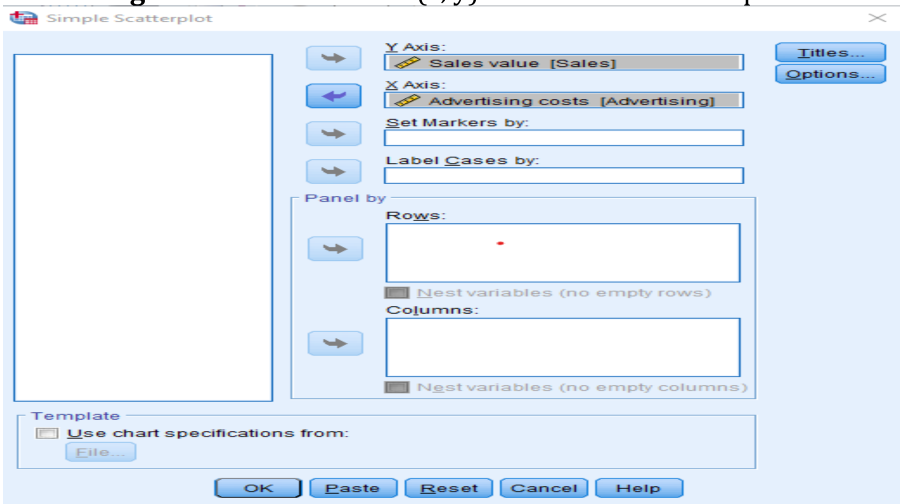
5. Application of Regression with SPSS

SPSS stands for Statistical Package for Social Sciences. It is a statistical software that makes it easy to collect and analyze data and perform various statistical analyzes with this package. It has its own menu bar which is useful for analyzing the input data easily. Many researchers like to use regression and correlation analysis, as many of them have applied this analysis in their theses and articles. The main objective of this study is to share knowledge about regression and correlation analysis and the conditions required to be used in their research paper, and the calculation of the regression and correlation equation is very complicated for big data, so it requires a program via SPSS which is very easy and faster. The results also showed that correlation and regression are two main tools for quantitative data analysis, and to study regression and correlation analysis in SPSS, we determine the type of variables whether they are dependent or independent.

Correlation analysis output table, Model summary, ANOVA table, Coefficients table. The following example represents the application of regression analysis with SPSS.

Sarad Chandra, (2019)

Figure 8. Determine the (x, y) values of the scatter plot.

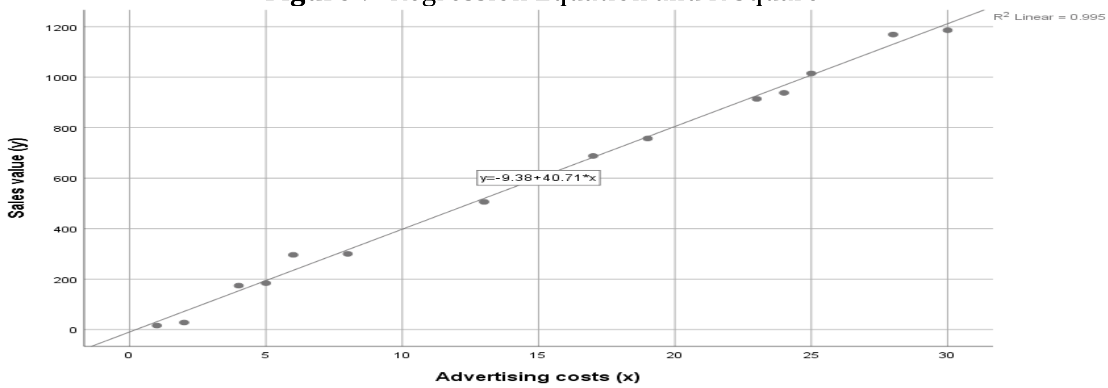


And move the (Sales) to Y Axis, the (Advertising) to X Axis > Ok. we obtain on the Diagram below and we see also the Regression Equation and R Square Coefficient:

The Regression Equation: $y = -9.38 + 40.71 * x$

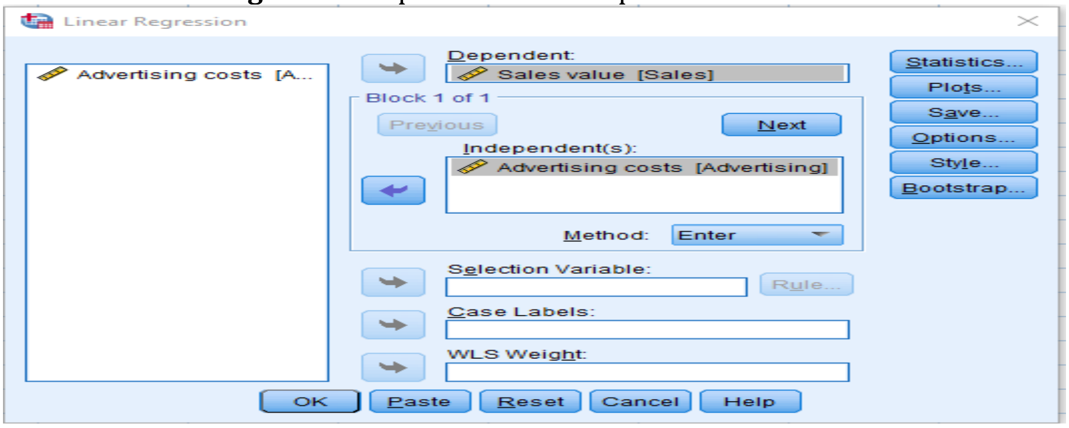
R Square Coefficient: $R^2 = 0.995$

Figure 9. Regression Equation and R Square.



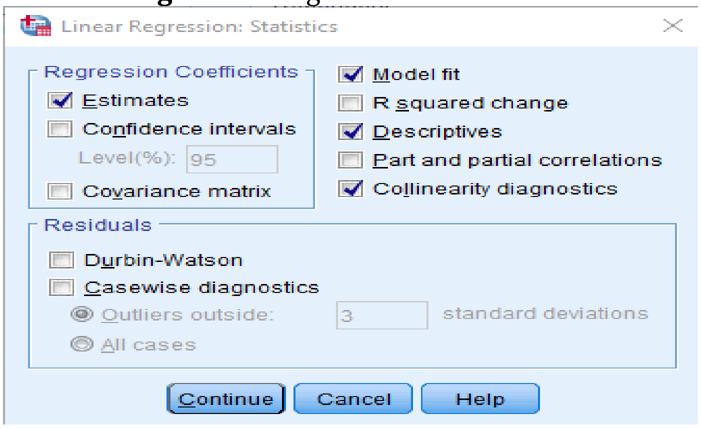
And then choosing Analyze > Regression > Linear > and then Moving to choose the Dependent variable and the independent variable as shown in the figure below:

Figure10. Dependent and Independent variables.



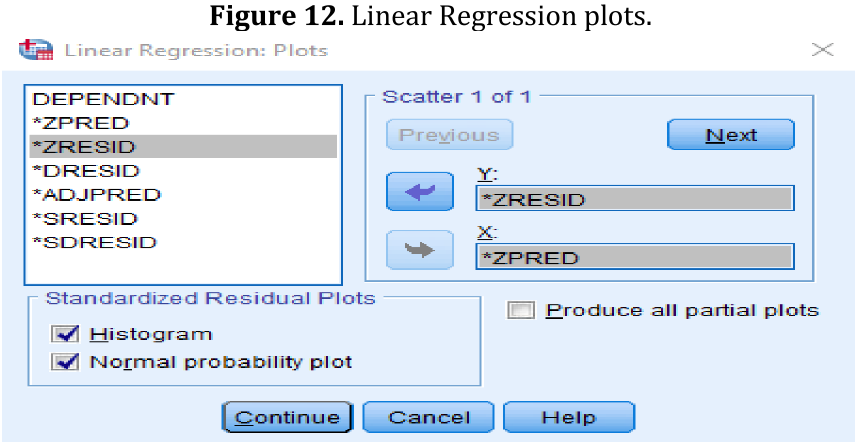
And click on Statistics > Check on the following options > Continue as shown on the figure below.

Figure 11. Regression Coefficients.



and click on the Plots > Check Histogram and Normal Probability plot under Standardized Residual Plots > Move *ZRESID to Y and Move *ZPRED to X > Continue we will get the figure below:

Figure 12. Linear Regression plots.



From all the above figures (10,11,12), Finally, we will obtain on Results Output SPSS as shown below:

$$\mathbf{Y} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{x} + \mathbf{e}$$

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

Table 2. Descriptive Statistics

	Mean	Std. Deviation	N
Sales value (y)	585.00	404.906	15
Advertising costs (x)	14.60	9.920	15

Table 3. Correlations

		Sales value (y)	Advertising costs (x)
Pearson Correlation	Sales value (y)	1.000	.997
	Advertising costs (x)	.997	1.000
Sig. (1-tailed)	Sales value (y)	.	.000
	Advertising costs (x)	.000	.
N	Sales value (y)	15	15
	Advertising costs (x)	15	15

Table 4. Model Summery

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.997[a]	.995	.994	30.514	.995	2452.134	1	13	.000

Table 5. Analysis of Variance

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	2283175.732	1	2283175.732	2452.134	.000[b]
	Residual	12104.268	13	931.098		
	Total	2295280.000	14			

Table 4. Coefficients values

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	-9.376	14.358		-.653	.525			
	Advertising costs (x)	40.711	.822	.997	49.519	.000	.997	.997	.997

From the above table, we notice that the value of the regression coefficient (β_1) is (40.711) to and that the value of the standard error is (.822), and if we divide the value of (β_1) by the value of the standard error, we will get the value of t, as shown in the above table:

$t - Statistics = \frac{40.711}{0.822} = 49.52$ and the Sig value = .000, This indicates that the coefficient of (x) has an effect and is statistically significant. We can also use the value of significance. If this value is less than 0.05, we can judge that the coefficient of (x) has an effect, or we can say that we reject the null hypothesis.

Table 6. Residuals Statistics.

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	31.34	1211.94	585.00	403.837	15
Residual	-44.046	61.112	.000	29.404	15
Std. Predicted Value	-1.371	1.552	.000	1.000	15
Std. Residual	-1.443	2.003	.000	.964	15

Figure 13. Regression standardized residuals.

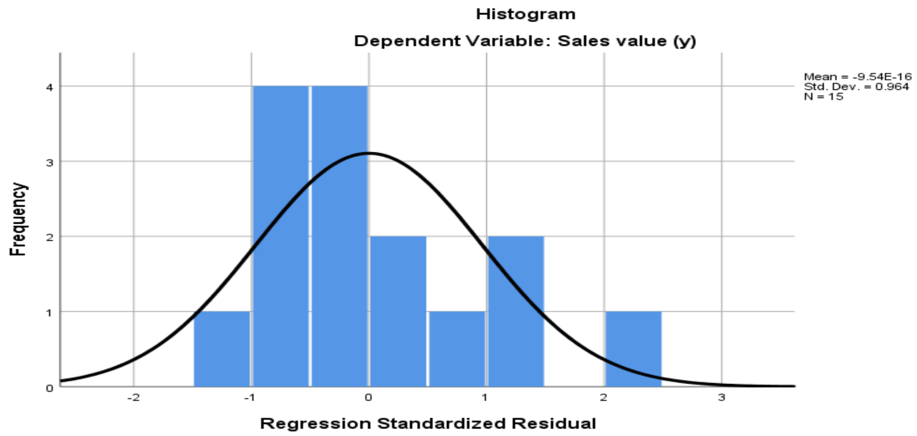
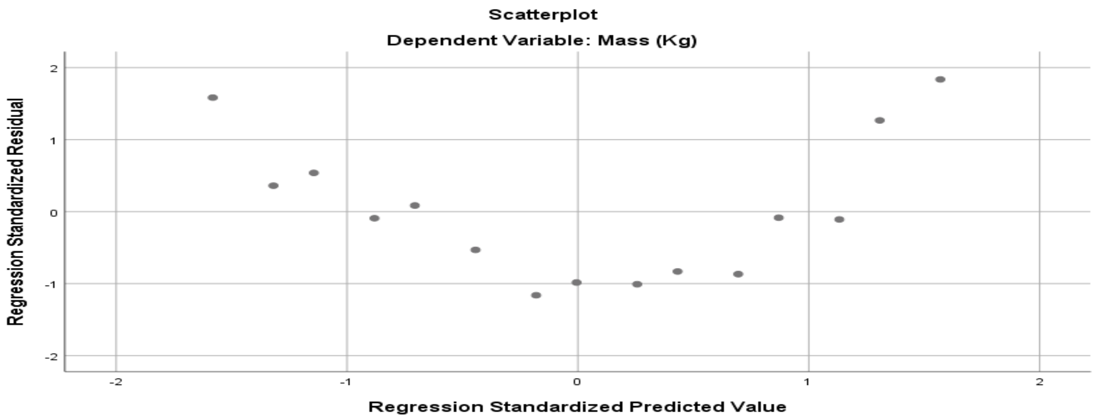


Figure 14. regression standardized predicted values.



From the tables above (**2**,**3**,**4**,**5**,**6**) and Figure (9), we found that there is a strong direct relationship between the dependent variable (y) and the independent variable (x), and we got the regression equation and the following results, as follows: Hasan, (2020) The Regression Equation: $y = -9.38 + 40.71 * x$

The Mean of the Mass (Kg), (y): 585.00

The Mean of the Height (m), (x): 14.60

Regression coefficient: 40.71

Intersection constant: -9.38

Correlation coefficient: 0.997

Determination of Coefficient: 0.995

Standard error for regression coefficient: 0.822

We can also use the value of significance. If this value is less than 0.05, we can judge that the coefficient of (x) has an effect, or we can say that we reject the null hypothesis.

From the foregoing, it was found that there is a strong correlation between the independent variable(x) and the dependent variable (y), and after looking at the value of the significant, which showed us that it is less than 0.05 we can reject the null hypothesis and accept the alternative hypothesis, which says $\beta_1 \neq 0$

6. Application of Regression with Microsoft Excel

In statistical modeling, as we knew earlier, regression analysis is used to estimate the relationships between two or more variables, and regression analysis helps us understand how the dependent variable changes when one of the independent variables changes and allows us to determine which of these variables has a clear effect. Technically, the regression analysis model is based on sum of squares, which is a mathematical method for finding the dispersion of data points and getting the smallest possible sum of squares and drawing a line closer to the data.

In statistics, they differentiate between simple and multiple linear regression. If the relationship between a dependent variable and one independent variable uses a linear function, then it is a simple linear regression model. If we use two or more explanatory variables to predict the dependent variable, we are dealing with multiple linear regression. The focus of our research will be on simple linear regression.

Let's take same example above, we do a linear regression using the least squares method and look for the coefficients a and b such that:

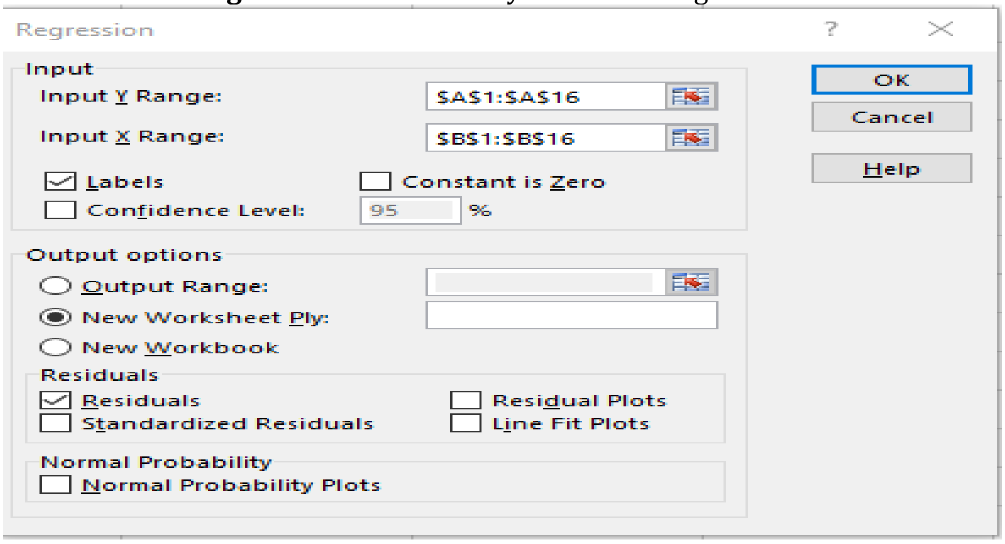
Y = b x + a

There are a few different ways to find "a" and "b" to perform linear regression analysis in Excel are: Svetlana,(2021).

How to the linear regression graph in Excel with Data Analysis?

We perform these steps to run a regression analysis in Excel: On the Data tab, click the Data Analysis button, select Regression, click OK, and select the Input Y Range, Input X Range, Labels Box, Choose Output Options, Select Residuals. We will get on the figure below (15):

Figure 15. x-values and y-values for Regression.



When we click OK we will obtain on the results for the Analysis:

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

Table 7. Output Summary of Regression Analysis.

Regression Statistics	
Multiple R	0.99735974
R Square	0.99472645
Adjusted R Square	0.9943208
Standard Error	30.5138904
Observations	15

We'll get four main tables from the output of the regression analysis.

Regression Analysis Output: Output Summary, this part tells you how suitable the computed linear regression equation is.

- R coefficient = 0.99735974, The correlation coefficient that measures the strength of the linear relationship between two variables. It can be between -1 and 1

1 means a strong positive relationship

-1 means a strong negative relationship

0 means no relationship at all

- R square = 0.99472645, The coefficient of determination that is used as an indicator of the quality of the fit. The value for R2 is calculated from the grand total of squares, it is the sum of the squared deviations of the original data from the mean. In our example, R2 equals 0.91, this means that 91% of our values fit the regression analysis model. That is, 91% of the dependent variables (y-values) are explained by the independent variables (x-values).

- Square R adjustment = 0.9943208, The square of R adjusted for the number of independent variable in the model. We will need to use this value in place of the R square for multiple regression analysis.
- Standard Error = 30.5138904, Another good metric that demonstrates the accuracy of the regression analysis - the smaller the number, the greater the confidence in the accuracy of the regression equation.
- Observations = 15, Number of observations or sample size

Regression analysis output: ANOVA

Table 8. ANOVA Table.

ANOVA					
	Df	SS	MS	F	Significance F
Regression	1	2283175.73	2283175.73	2452.13387	3.3985E-16
Residual	13	12104.2676	931.097505		
Total	14	2295280			

- df: The number of degrees of freedom.
- SS: The sum of squares. The smaller the residual SS size compared to Total SS, the better the model fits the data.
- MS: The mean of the square.
- F: The F statistic, or F-test for the null hypothesis. It is used to test the general significance of the form.
- F significance: The P-value of F.

The significance value F gives an idea of the reliability of the results (statistical significance). If the Significance F is less than 0.05, the model is OK. If it is greater than 0.05, it is better to choose another independent variable.

Regression analysis output: coefficients

Table 9. Coefficients Values.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-9.3755807	14.357735	-0.6529986	0.52513787	-40.393581	21.64242
X	40.7106562	0.82212152	49.5190254	3.3985E-16	38.9345707	42.4867418

The most important table in this section is the table of coefficients. Enables you to construct a linear regression equation in Excel:

$$y = -9.37 + 40.71 * x$$

From the above tables (8,9,10), we notice that the value of the regression coefficient (X) is (40.711) to and that the value of the standard error is (0.822), and the value of t- Stat, as shown in the above table:

$t - Statistics = 49.52$ and the P-value = 3.3985E-16, This indicates that the coefficient of (x) has an effect and is statistically significant. We can also use the P-value. If this value is less than 0.05, we can judge that the coefficient of (x) has an effect, or we can say that we reject the null hypothesis.

Regression analysis output: residuals

For this table to compare the estimated and actual number of Sales value corresponding to the Advertising costs, we will see that these figures are somewhat different:

Estimated number = 1008.39 , Actual number = 1015

This difference is because the independent variables are not ideal predictors of the dependent variables. This helps us to know the extent to which the actual values are distorted from the expected.

Table 10. Residuals Outputs.

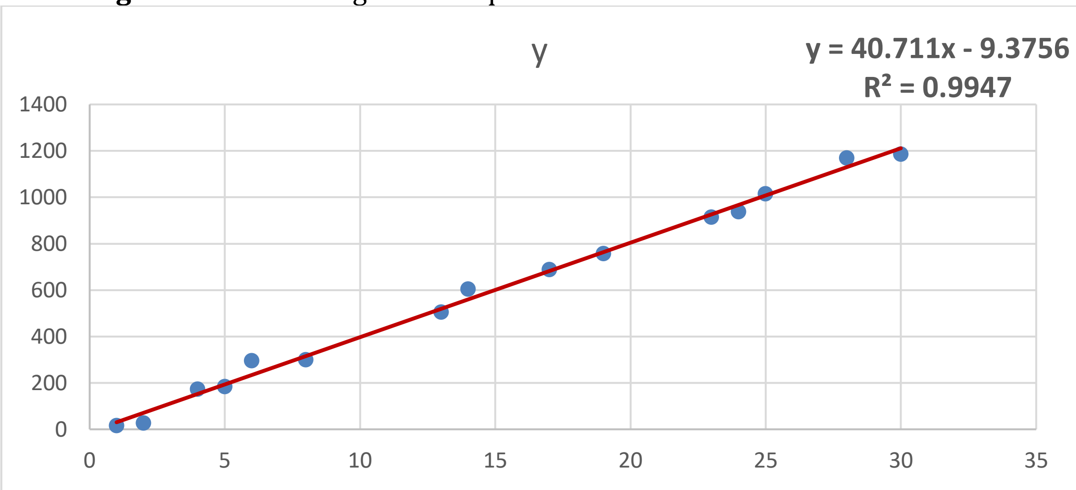
Observation	Predicted y	Residuals
1	1008.39082	6.60917538
2	234.888357	61.1116434
3	316.309669	-16.309669
4	682.705575	5.29442509
5	72.0457317	-44.045732
6	519.86295	-13.86295
7	926.969512	-12.969512
8	1211.94411	-25.944106
9	1130.52279	38.4772067
10	560.573606	43.4263937
11	764.126887	-7.1268873
12	153.467044	20.5329559
13	967.680168	-29.680168
14	31.3350755	-15.335075
15	194.1777	-10.1777

- How to the linear regression graph in Excel with chart?

Here, draw a linear regression chart using a few simple steps:

Select the two columns for the data, including Title, and in the tab, we click Chart > Scatter Plot, and select the first scatter plot

Figure 16. Linear Regression equation and coefficient of Determination.



7. Conclusion

Through these applications, you will find that linear regression is used in everything from medical, biological, behavioral, environmental and, social sciences to business, as linear regression models have become a proven method of scientifically and reliably predicting the future. Because linear regression is a well-established statistical procedure, the properties of linear regression models are well understood and can be trained very quickly.

Correlation and linear regression are two strategies for determining the relationship between two variables. The strength of a linear link between two variables is measured by correlation, whereas regression expresses the relationship as an equation. We've used basic examples with, Python, SPSS, and Excel in this research to demonstrate linear regression analysis and encourage users to use these approaches to examine their own data.

In our example we used the sales value as the dependent variable and the advertising expenditures as the independent variable to test the linear regression.

The correlation association ($r = 0.997$) between the value of sales and advertising costs was given.

Only 99.5 % of the sales value is explained by advertising costs, according to R Square = 0.995.

The 'utility' of a linear regression model with ($P < 0.05$) is shown in the ANOVA table. This gives a numerical representation of the link between advertising costs and sales value. If more than one independent variable is added, the linear regression model will correct for the effect of additional dependent variables when assessing the effect of one variable when testing the effect of one variable.

The following is a set of advantages and disadvantages for each of the programs under our study, as in the table below:

Table 11. Advantages and Disadvantages

Index	Name of Program	Advantages	Disadvantages
1	Python	1. Python is one of the languages that allow it to be used in many computer operating systems. 2. It is one of the languages that are easy to learn and easy to read, and this is the most important advantage, because it does not make learners find difficulties while learning it. 3. It is an open source language, so you can modify this language. 4. This language deals very easily with all the various databases. 5. Help programmers create applications that are made on the web. 6. Programmers can use it in all areas that work on digital protection.	1. It is considered one of the slow languages if compared to other programming languages where tasks are executed line by line unlike the rest of the languages. 2. This language has not been able to enter the mobile apps race yet. 3. This language led to many programmers relying on this language in their way of recording data and writing, so they are now unable to learn new languages and understand them well.
2	SPSS	1. Relatively easy to use. 2. Compatibility with other programs such as Microsoft Office package. 3. Ability to save files with several extensions. The results of data analysis in the program are in the form of easy-to-read charts and graphs.	1. SPSS is not free, and it is rather expensive. 2. The program does not provide programming codes format and modification for those who need advanced procedures. 3. The program is only intended for statistical operations of the social sciences, and you will need to purchase more other packages for various purposes. 4. The available default designs and graphics are of low quality.
3	Excel	1. Instant Calculations. 2. Easy to Edit. 3. Multiple Spreadsheets in one file. 4. Easy and Effective Comparisons. 5. Running on a low-efficiency computer: 6. Graphical display of data.	1. hard to use 2. Easy to install viruses through this program 3. With the growth and increase of data, the problems increase

References

1. Golberg, M. A., & Cho, H. A. (2004). Introduction to regression analysis. WIT press.

2. David A. Freedman (2009). Statistical Models: Theory and Practice. Cambridge University Press. p. 26. A simple regression equation has on the right hand side an intercept and an explanatory variable with a slope coefficient.

3. Aggarwal, R., & Ranganathan, P. (2017). Common pitfalls in statistical analysis: Linear regression analysis. Perspectives in clinical research, 8(2), 100.

4. Chatterjee, S., & Hadi, A. S. (2015). Regression analysis by example. John Wiley & Sons.

5. Bronshtein, A., 2017. Simple and Multiple Linear Regression in Python. [online] Medium. Available at:https://towardsdatascience.com/simple-and-multiple-linear-regression-in-python-c928425168f9.

6. Jaiswal, S. (2020). Retrieved 10 February 2021, from https://www.javatpoint.com/linear-regression-in-machine earning#:~:text=Linear%20regression%20is%20one%20of,age%2C%20product%20price%2C%20etc

7. Sarad Chandra, K., 2019. (PDF) Correlation and Regression Analysis Using SPSS. [online] Research Gate. Available at: https://www.researchgate.net/publication/343282545_Correlation_and_Regression_Analysis_Using_SPSS.

8. Hasan, N., 2020. [online] Sites.education.miami.edu. Available at: https://sites.education.miami.edu/statsu/wp-content/uploads/sites/4/2020/06/Regression-Webinar.pdf.

9. Svetlana, C., 2021. Linear regression analysis in Excel. [online] Ablebits.com. Available at: https://www.ablebits.com/office-addins-blog/2018/08/01/linear-regression-analysis-excel/.

10. Kumari, K., & Yadav, S. (2018). Linear regression analysis study. Journal of the practice of Cardiovascular Sciences, 4(1), 33.