y_i^2	$X_{i1}y_i$	X_{i1}^2	X_{i1}	y_i	No.
12544	3920	1225	35	112	1
16384	5120	1600	40	128	2
16900	4940	1444	38	130	3
19044	6072	1936	44	138	4
24964	10586	4489	67	158	5
26244	10368	4096	64	162	6
19600	8260	3481	59	140	7
30625	12075	4761	69	175	8
15625	3125	625	25	125	9
20164	7100	2500	50	142	10
202094	1410	491	26157	71566	Total

$$S_{Xy} = \sum X_i y_i - \frac{(\sum X_i)(\sum y_i)}{n} \Rightarrow = 71566 - \frac{(491)(1410)}{10} = 2335$$

$$S_{XX} = \sum X_i^2 - \frac{(\sum X_i)^2}{n} \Rightarrow = 26157 - \frac{(491)^2}{10} = 2048.9$$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} \Rightarrow = 202094 - \frac{(1410)^2}{10} = 3284$$

Thus, the correlation coefficient is equal to:

$$r_{Xy} = \frac{S_{Xy}}{\sqrt{S_{XX}S_{yy}}} = \frac{2335}{\sqrt{(2048.5)(3284)}} = 0.9$$

1- Create an analysis of variance table using the correlation coefficient.

S.O.V	d.f	S.S	M.S	Fcal.
R(X1)	1	$r_{Xy}^2 S_{yy} = (0.9)^2(3248) = 2660.04$	2662.04	34.1
Error	n-2 10-2 8	$(1-r_{Xy}^2)S_{yy} = (1-(0.9)^2)(3248) = 623.96$	77.99	
Total	n-1 9			

Compared $F_{cal.}$to $F_{tab.}$:

$$F_{cal.} = 34.1 > F_{tab.} = F(0.05,1,8) = 5.32$$

Therefore, the null hypothesis is rejected and the alternative hypothesis is accepted, i.e. there is a significant effect of the independent variable represented by age on the dependent variable represented by blood pressure.

Note: In this test of the ANOVA table using the correlation coefficient, the test is to test the hypothesis below, which is the same hypothesis that we usually use when creating the ANOVA table.

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

3- Find the coefficient of determination.

$$R^2 = \frac{Sum\,of\ Squares\,of\ due\,to\,Regression}{Sum\,of\ Squares\,of\ Total} = \frac{2661.04}{3284} = 0.81$$

That is, the variable X explains 0.81 of the changes in y through the linear model that describes the relationship between them.

4- Test the following hypothesis:

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

In order to conduct a t-test for this hypothesis, we use the t-test as follows:

$$t = \frac{r - r_0}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.9-0}{\sqrt{\frac{1-(0.9)^2}{10-2}}} = 5.84$$

It is the same value for the t-test when testing X. Using t-distribution tables, we obtain the tabular t-value as follows:

$$t_{(\frac{\alpha}{2}, n-2)} = t_{(0.025,8)} = 2.306$$

By comparing the calculated value with the tabular value, it is noted that:

$$t_{cal.} = 5.84 > t_{(0.025,8)} = 2.306$$

Thus, through comparison, the null hypothesis is rejected and the alternative hypothesis is accepted, i.e. the correlation coefficient is not equal to zero, and thus there is a significant correlation between the two variables X and y.

Violations or errors in the analysis assumptions of the simple regression model
How to detect and correct them

In previous lectures, the analysis assumptions of the simple regression model were explained through four points. In order to make these assumptions more compatible with the simple linear regression model, we will add a fifth assumption, which is:

5-The existence of a linear relationship between represented by the following straight-line equation:

In previous lectures, the analysis assumptions of the simple regression model were explained through four points. In order to make these assumptions more compatible with the simple linear regression model, we will add a fifth assumption, which is:

5- The existence of a linear relationship between $X, \hat{y}$ represented by the following straight-line equation:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X_{i1}$$

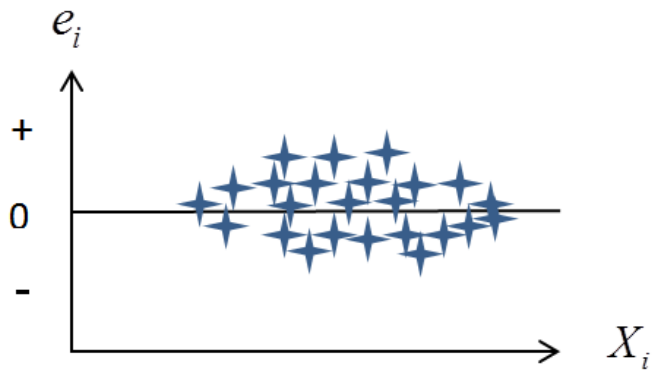
Conducting a test of these assumptions and discovering whether there are violations or errors in one or more of them and explaining how to correct them and the method used for this purpose is what is called error or residual analysis.

The availability of the analysis hypotheses can be tested through the drawing as follows:

1- Using the graph of the values of e_i against $\hat{y}_i$ or X_i, which generally explains to us whether the analysis hypotheses are available or not.

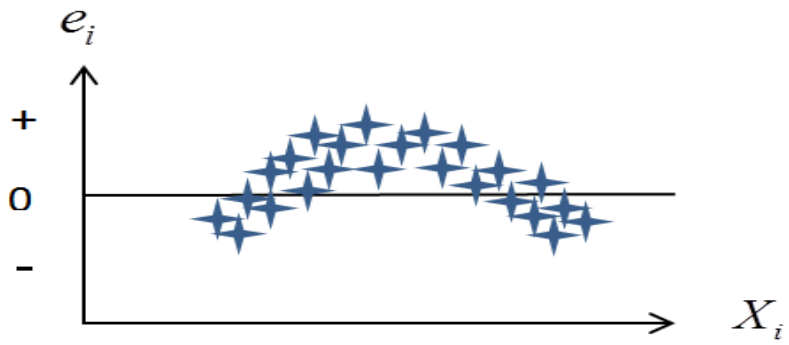
2-

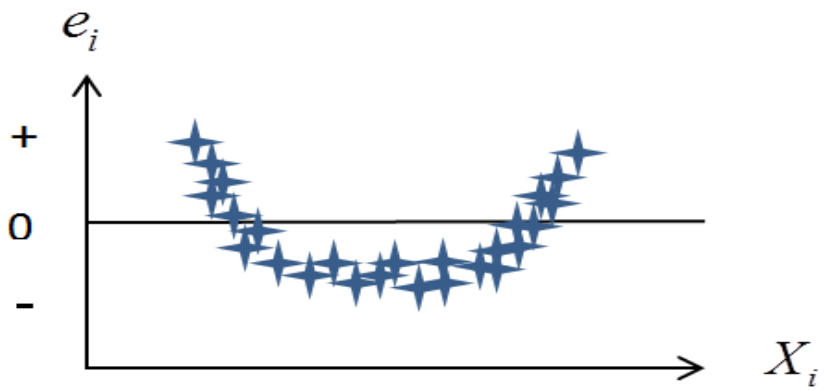
If the points are centered in an organized manner around zero, this indicates that the analysis hypotheses are available as in the following drawing:



If the graph differs from this, this indicates that there is a defect in the analysis assumptions.

The drawing can also be in the two forms below:





The previous two graphs test the relationship between X and Y, as the two graphs indicate that there is an equation with a higher degree than the straight-line equation, which is more appropriate for the data.

2- Using a statistical Test

There are two cases:

A- When there is no repetition of the values of the independent variable X_i, in this case we continue to add other terms to the equation to become of second or third degree...etc., until we find the appropriate equation.

B- When there is a repetition of values X_i, we test the lack of fit, which can be described as follows:

A test of lack of fit

The lack of fit test means the extent to which the linear model fits the data, provided that there is repetition in the values X_i.

If the error resulting from the model not fitting is significant, we say that the model does not fit the data. Thus, we will have two types of errors.

1- Pure Error

2- Lack of fit error

The purpose of the lack of fit test can be explained, which is to test whether the linear model is suitable for the data or not.

In order to conduct this test, there are two hypotheses:

H_0: There is no lack of fit, i.e. the linear model fits the data.

H_1: There is a lack of fit, i.e. the linear model does not fit the data.

If the lack of fit test is done and it appears that the linear model fits the data (accepting the null hypothesis), then we conduct a significance test of β_1, i.e. we conduct the following hypothesis test:

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$