We see if the changes in y are caused by the variable X. However, if H_0 is rejected in the lack of fit test, meaning that the linear model does not fit the data, then there is no need to test the significance of β_1 because the linear model does not fit the data, as there may be a higher degree model, for example, that is more suitable than the linear model.

How to conduct a lack of fit test

Assuming that we have a table consisting of X_i values (including frequencies in X_i values) corresponding to y_i values as follows:

X_i	y_i
$x_{11} = a$	y_1
$x_{21} = b$	y_2
$x_{31} = a$	y_3
$x_{41} = c$	y_4
$x_{51} = b$	y_5
$x_{61} = a$	y_6
$\vdots$	$\vdots$
$x_{n1} = F$	y_n

First, the values of the table above are arranged as follows:

قيم X_i المتكررة	قيم y_i المقابلة لقيم X_i المتكررة	d.f	S.S
X_i	$y_{11}, y_{12}, y_{13}, \dots, y_{1k}$	$r_1 - 1$	SS_{y_1}
X_i	$y_{21}, y_{22}, y_{23}, \dots, y_{2k}$	$r_2 - 1$	SS_{y_2}
$\vdots$	$\vdots \quad \vdots \quad \vdots \quad \vdots$ $\vdots$	$\vdots$	$\vdots$
X_i	$y_{k1}, y_{k2}, y_{k3}, \dots, y_{kk}$	$r_k - 1$	SS_{y_k}
		$\sum_{i=1}^{k}(r_i - 1)$	$\sum_{i=1}^{k} SS(pureError)$

S.O.V	d.f	S.S	M.S	$F_{cal.}$
$R(X_i)$	1	$SSR(X_i)$	$M.S.R(X_1)$	
$Error$	$n-2$	SSe	$M.S.e$	
$L.o.f$	$d.f(L.o.f) =$ $= d.f.e - d.f(p.e)$	$SS(L.o.f)$	$M.S(L.o.f)$	$F_{cal.} = \frac{M.S(L.o.f)}{M.S(p.e)}$
$P.e$	$d.f(p.e)$	$SS(p.e)$	$M.S(p.e)$	
$Total$	$n-1$			

Were:

$SS(L.o.f) = SSe - SS\ p.e.$

The tabular F is as follows:

$$F_{(\alpha\ , v1, v2)}$$

degree of freedom of $(L.o.f)$ degree of freedom $(P.e)$

Example: Using the following data, test whether the linear model **fits** the data.

X_i	y_i
35	112
25	125
40	128
35	115
64	162
25	130
50	142
67	158
69	175
50	140
70	170
50	145

Solution: Arrange the values in ascending order according to X_i.

X_i	y_i
25	125
25	130
35	112
35	115
40	128
50	142
50	140
50	145
64	162
67	158
69	175
70	170

A table is created to calculate the sum of squared net error (p.e) as follows:

قيم X_i المتكررة	قيم y_i المقابلة لقيم X_i المتكررة	d.f	S.S
$X = 25$	125 130	1	12.5
$X = 35$	112 115	1	4.5
$X = 50$	142 140 145	2	12.67
Total		4	29.67=SS(p.e)

$$\bar{X} = \frac{125+130}{2} = 127.5,$$

$$SS = \sum (X_i - \bar{X})^2$$

$$SS = (125-127.5)^2 + (130-127.5)^2$$

$$= 6.25 + 6.25$$

$$= 12.5$$

We now create an analysis of variance table to conduct the test.

S.O.V	d.f	S.S	M.S	$F_{cal.}$
$R(X_i)$	1	3963.65	3963.69	
Error	10	836.02	83.602	
$L.o.f$	10−4=6	806.35	134.39	18.12
$P.e$	4	29.67	7.4175	
Total	11			

Were:

$$S.S(L.o.f) = SSe - SS(p.e)$$
$$= 836.02 - 29.67$$
$$= 806.35$$
$$d.f(L.o.f) = d.f.e - d.f(p.e)$$
$$= 10 - 4$$
$$= 6$$

The tabular F value is:

$$F_{tab.} = F(\alpha,\ d.f(L.o.f),\ d.f(p.e))$$
$$= F(0.05, 6, 4)$$
$$= 4.16$$

By comparing the calculated F value with the table F value, we get:

$$F_{cal.} = 18.12 > F_{tab.} = 4.16$$

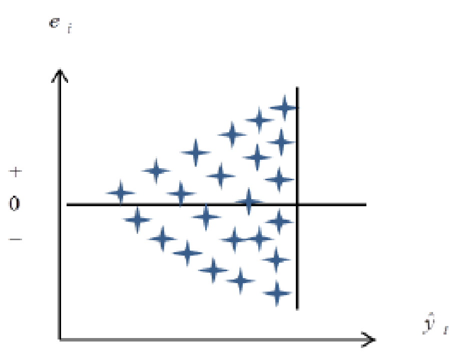
Through comparison, and since the calculated F value is greater than the tabular F value, the null hypothesis is rejected, meaning that there is a lack of fit, meaning that the linear model does not fit the data, but rather there is an equation of a second degree or higher that may fit the data.

Test whether the error variance or residual is constant and homogeneous

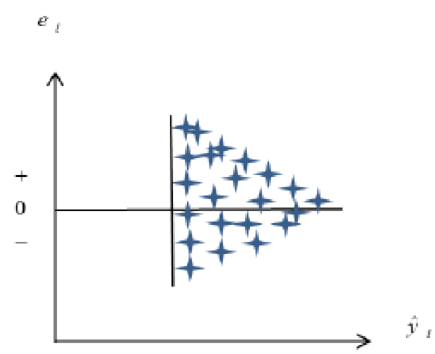
The error variance test can be performed in terms of being constant and homogeneous in two ways:

1-Using the chart:

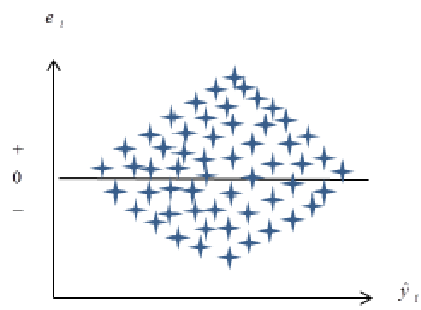
When the e_i values are plotted against the $\hat{y}_i$ or X_i values and the graph appears as one of the following shapes:



Increase in error variance



Decrease in error variance



Increase and decrease in error variance

This indicates that the error variance is not homogeneous, i.e. the error variance is not constant (heteroscedastic).

2-Using a statistical Test

arrange the data in ascending order according to the values of the independent variable X. Then, divide the arranged data into two sections (with some middle data