Chapter Two

THE CLASSICAL REGRESSION ANALYSIS

[*The Simple Linear Regression Model*]

Economic theories are mainly concerned with the relationships among various economic variables. These relationships, when phrased in mathematical terms, can predict the effect of one variable on another. The functional relationships of these variables define the dependence of one variable upon the other variable (s) in the specific form. The specific functional forms may be linear, quadratic, logarithmic, exponential, hyperbolic, or any other form.

In this chapter we shall consider a simple linear regression model, i.e. a relationship between two variables related in a linear form. We shall first discuss two important forms of relation: *stochastic* and *non-stochastic,* among which we shall be using the former in econometric analysis.

2.1. Stochastic *and Non-stochastic Relationships*

A relationship between X and Y, characterized as Y = f(X) is said to be deterministic or non-stochastic if for each value of the independent variable (X) there is one and only one corresponding value of dependent variable (Y). On the other hand, a relationship between X and Y is said to be stochastic if for a particular value of X there is a whole probabilistic distribution of values of Y. In such a case, for any given value of X, the dependent variable Y assumes some specific value only with some probability. Let's illustrate the distinction between stochastic and non-stochastic relationships with the help of a supply function.

Assuming that the supply for a certain commodity depends on its price (other determinants taken to be constant) and the function being linear, the relationship can be put as:

$$Q=f(P)=\alpha+\beta P \text{----------} (2.1)$$

The above relationship between P and Q is such that for a particular value of P, there is only one corresponding value of Q. This is, therefore, a deterministic (non-stochastic) relationship since for each price there is always only one corresponding quantity supplied. This implies that all the variation in Y is due solely to changes in X, and that there are no other factors affecting the dependent variable.

If this were true all the points of price-quantity pairs, if plotted on a two-dimensional plane, would fall on a straight line. However, if we gather observations on the quantity actually supplied in the market at various prices and we plot them on a diagram we see that they do not fall on a straight line.

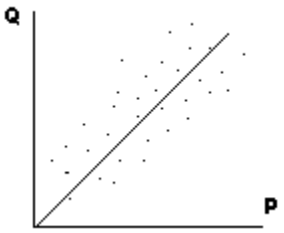


Fig 2.a The scatter diagram

The derivation of the observation from the line may be attributed to several factors.

a. Omission of variables from the function
b. Random behavior of human beings
c. Imperfect specification of the mathematical form of the model
d. Error of aggregation
e. Error of measurement

In order to take into account the above sources of errors we introduce in econometric functions a random variable which is usually denoted by the letter 'u' or 'ε' and is called error term or random disturbance or stochastic term of the function, so called be cause u is supposed to 'disturb' the exact linear relationship which is assumed to exist between X and Y. By introducing this random variable in the function the model is rendered stochastic of the form:

$$Y_i = \alpha + \beta X + u_i \quad \text{..... (2.2)}$$

Thus a stochastic model is a model in which the dependent variable is not only determined by the explanatory variable(s) included in the model but also by others which are not included in the model.

2.2. Simple Linear Regression model.

The above stochastic relationship (2.2) with one explanatory variable is called simple linear regression model.

The true relationship which connects the variables involved is split into two parts:

A part represented by a line and a part represented by the random term 'u'.

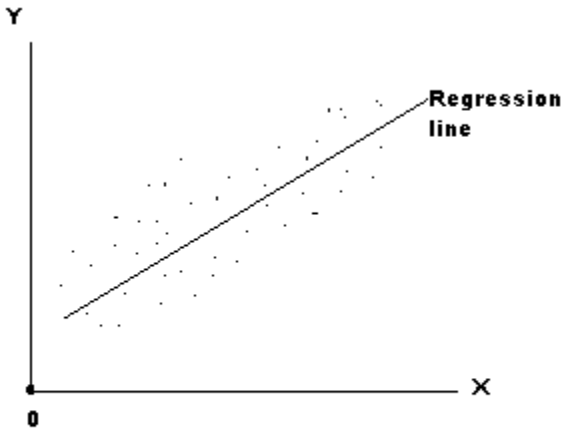


Fig 2.b the scatter diagram and the regression line

The scatter of observations represents the true relationship between Y and X. The line represents the exact part of the relationship and the deviation of the observation from the line represents the random component of the relationship.

- Were it not for the errors in the model, we would observe all the points on the line $Y'_1, Y'_2, \ldots\ldots, Y'_n$ corresponding to $X_1, X_2, \ldots, X_n$. However because of the random disturbance, we observe $Y_1, Y_2, \ldots\ldots, Y_n$ corresponding to $X_1, X_2, \ldots, X_n$. These points diverge from the regression line by $u_1, u_2, \ldots, u_n$.

$$\underbrace{Y_i}_{\text{the dependent variable}} = \underbrace{\alpha + \beta x_i}_{\text{the regression line}} + \underbrace{u_i}_{\text{random variable}}$$

- The first component in the bracket is the part of Y explained by the changes in X and the second is the part of Y not explained by X, that is to say the change in Y is due to the random influence of u_i.

2.2.1 Assumptions of the Classical Linear Stochastic Regression Model.

The classicals made important assumption in their analysis of regression .The most important of these assumptions are discussed below.

1. *The model is linear in parameters*.

The classical assumed that the model should be linear in the parameters regardless of whether the explanatory and the dependent variables are linear or not. This is because if the parameters are non-linear it is difficult to estimate them since their value is not known but you are given with the data of the dependent and independent variable.

Example 1. $Y=\alpha+\beta x+u$ Is linear in both parameters and the variables, so it

Satisfies the assumption

2. $\ln Y=\alpha+\beta \ln x+u$ is linear only in the parameters. Since the the classicals worry on the parameters, the model satisfies the assumption.

Dear students! Check yourself whether the following models satisfy the above assumption and give your answer to your tutor.

a. $\ln Y^2=\alpha+\beta \ln X^2+U_i$

b. $Y_i=\sqrt{\alpha+\beta X_i}+U_i$

2. U_i is a random real variable

This means that the value which u may assume in any one period depends on chance; it may be positive, negative or zero. Every value has a certain probability of being assumed by u in any particular instance.

3. ***The mean value of the random variable(U) in any particular period is zero***

This means that for each value of x, the random variable(u) may assume various values, some greater than zero and some smaller than zero, but if we considered all the possible and negative values of u, for any given value of X, they would have on average value equal to zero. In other words the positive and negative values of u cancel each other.

Mathematically, $E(U_i)=0$ .....(2.3)

4. ***The variance of the random variable(U) is constant in each period (The assumption of homoscedasticity)***

For all values of X, the u's will show the same dispersion around their mean. In Fig.2.c this assumption is denoted by the fact that the values that u can assume lie with in the same limits, irrespective of the value of X. For X_1 , u can assume any value with in the range AB; for X_2 , u can assume any value with in the range CD which is equal to AB and so on.

Graphically;

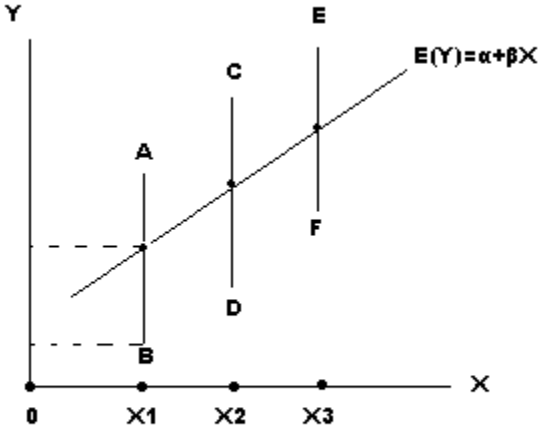


Fig. 2.c Homoscedastic Variance

Mathematically;

$Var(U_i)=E[U_i-E(U_i)]^2=E(U_i)^2=\sigma^2$ (Since$E(U_i)=0$).This constant variance is called ***homoscedasticity*** assumption and the constant variance itself is called homoscedastic variance.

5. ***The random variable (U) has a normal distribution***

This means the values of u (for each x) have a bell shaped symmetrical distribution about their zero mean and constant varianceσ^2, i.e.

$$U_i \sim N(0,\sigma^2) \quad \text{.....2.4}$$

6. ***The random terms of different observations*** (U_i,U_j)***are independent***. ***(The assumption of no autocorrelation)***

This means the value which the random term assumed in one period does not depend on the value which it assumed in any other period.

Algebraically,

$$Cov(u_i u_j)=E\left[[(u_i-E(u_i)][u_j-E(u_j)]\right]$$
$$=E(u_i u_j)=0 \quad \text{.....(2.5)}$$

7. The X_i are a set of fixed values in the hypothetical process of repeated sampling which underlies the linear regression model.

- This means that, in taking large number of samples on Y and X, the X_i values are the same in all samples, but the $u_i$values do differ from sample to sample, and so of course do the values of y_i.

8. The random variable (U) is independent of the explanatory variables.

This means there is no correlation between the random variable and the explanatory variable. If two variables are unrelated their covariance is zero.

Hence $Cov(X_i,U_i)=0$ .....(2.6)

Proof:-

$$\text{cov}(XU)=E\left[[(X_i-E(X_i)][U_i-E(U_i)]\right]$$
$$=E[(X_i-E(X_i)(U_i)] \text{ given } E(U_i)=0$$
$$=E(X_i U_i)-E(X_i)E(U_i)$$
$$=E(X_i U_i) \qquad =X_i E(U_i), \text{ given that the } x_i \text{ are fixed}$$
$$=0$$

9. The explanatory variables are measured without error

- U absorbs the influence of omitted variables and possibly errors of measurement in the y's. i.e., we will assume that the regressors are error free, while y values may or may not include errors of measurement.

Dear students! We can now use the above assumptions to derive the following basic concepts.

A. The dependent variable Y_i is normally distributed.

$$\text{i.e } \underline{\underline{\therefore Y_i \sim N\left[(\alpha+\beta x_i), \sigma^2\right]}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.7)$$

Proof:

Mean: $E(Y) = E(\alpha+\beta x_i + u_i)$

$= \alpha+\beta X_i$ Since $E(u_i) = 0$

Variance: $Var(Y_i) = E\left(Y_i - E(Y_i)\right)^2$

$= E\left(\alpha+\beta X_i + u_i - (\alpha+\beta X_i)\right)^2$

$= E(u_i)^2$

$= \sigma^2$ (since $E(u_i)^2 = \sigma^2$)

$$\underline{\underline{\therefore \text{var}(Y_i) = \sigma^2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.8)$$

The shape of the distribution of Y_i is determined by the shape of the distribution of u_i which is normal by assumption 4. Since $\alpha \text{ and } \beta$, being constant, they don't affect the distribution of y_i. Furthermore, the values of the explanatory variable, x_i, are a set of fixed values by assumption 5 and therefore don't affect the shape of the distribution of y_i.

$$\underline{\underline{\therefore Y_i \sim N(\alpha+\beta x_i, \sigma^2)}}$$

B. successive values of the dependent variable are independent, i.e

$$Cov(Y_i, Y_j) = 0$$

2.2.2 Methods of estimation

Specifying the model and stating its underlying assumptions are the first stage of any econometric application. The next step is the estimation of the numerical values of the parameters of economic relationships. The parameters of the simple linear regression model can be estimated by various methods. Three of the most commonly used methods are:

1. Ordinary least square method (OLS)
2. Maximum likelihood method (MLM)
3. Method of moments (MM)

But, here we will deal with the OLS and the MLM methods of estimation.

2.2.2.1 The ordinary least square (OLS) method

The model $Y_i=\alpha+\beta X_i+U_i$ is called the true relationship between Y and X because Y and X represent their respective population value, and $\alpha \ and \ \beta$ are called the true parameters since they are estimated from the population value of Y and X But it is difficult to obtain the population value of Y and X because of technical or economic reasons. So we are forced to take the sample value of Y and X. The parameters estimated from the sample value of Y and X are called the estimators of the true parameters $\alpha \ and \ \beta$ and are symbolized as $\hat{\alpha} \ and \ \hat{\beta}$.

The model $Y_i=\hat{\alpha}+\hat{\beta} X_i+e_i$, is called estimated relationship between Y and X since $\hat{\alpha} \ and \ \hat{\beta}$ are estimated from the sample of Y and X and e_i represents the sample counterpart of the population random disturbance U_i .

Estimation of $\alpha \ and \ \beta$ by least square method (OLS) or classical least square (CLS) involves finding values for the estimates $\hat{\alpha} \ and \ \hat{\beta}$ which will minimize the sum of square of the squared residuals ($\sum e_i^2$).

From the estimated relationship $Y_i = \hat{\alpha} + \hat{\beta} X_i + e_i$, we obtain:

$$e_i = Y_i - (\hat{\alpha} + \hat{\beta} X_i) \quad \text{.....(2.6)}$$

$$\sum e_i^2 = \sum (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2 \quad \text{.....(2.7)}$$

To find the values of $\hat{\alpha} \ and \ \hat{\beta}$ that minimize this sum, we have to partially differentiate $\sum e_i^2$ with respect to $\hat{\alpha} \ and \ \hat{\beta}$ and set the partial derivatives equal to zero.

1. $\frac{\partial \sum e_i^2}{\partial \hat{\alpha}} = -2\sum (Y_i - \hat{\alpha} - \hat{\beta} X_i) = 0 \quad \text{.....(2.8)}$

Rearranging this expression we will get: $\sum Y_i = n\alpha + \hat{\beta} \Sigma X_i$ **.....(**2.9)

If you divide (2.9) by 'n' and rearrange, we get

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} \quad \text{.....(2.10)}$$

2. $\frac{\partial \sum e_i^2}{\partial \hat{\beta}} = -2\sum X_i (Y_i - \hat{\alpha} - \hat{\beta} X) = 0 \quad \text{.....(2.11)}$

Note: at this point that the term in the parenthesis in equation 2.8and 2.11 is the residual, $e = Y_i - \hat{\alpha} - \hat{\beta} X_i$. Hence it is possible to rewrite (2.8) and (2.11) as $-2\sum e_i = 0$ and $-2\sum X_i e_i = 0$. It follows that;

$$\sum e_i = 0 \text{ and } \sum X_i e_i = 0 \quad \text{.....(2.12)}$$

If we rearrange equation (2.11) we obtain;

$$\sum Y_i X_i = \hat{\alpha} \Sigma X_i + \hat{\beta} \Sigma X_i^2 \quad \text{.....(2.13)}$$

E quation (2.9) and (2.13) are called the ***Normal Equations***. Substituting the values of $\hat{\alpha}$ from (2.10) to (2.13), we get:

$$\sum Y_i X_i = \Sigma X_i(\bar{Y} - \hat{\beta}\bar{X}) + \hat{\beta}\Sigma X_i^2$$

$$= \bar{Y}\Sigma X_i - \hat{\beta}\bar{X}\Sigma X_i + \hat{\beta}\Sigma X_i^2$$

$$\sum Y_i X_i - \bar{Y}\Sigma X_i = \hat{\beta}(\Sigma X_i^2 - \bar{X}\Sigma X_i)$$

$$\Sigma XY - n\bar{X}\bar{Y} = \hat{\beta}(\Sigma X_i^2 - n\bar{X}_2)$$

$$\hat{\beta} = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X_i^2 - n\bar{X}2} \quad \text{.....(2.14)}$$

Equation (2.14) can be rewritten in somewhat different way as follows;

$$\Sigma(X - \bar{X})(Y - \bar{Y}) = \Sigma(XY - X\bar{Y} - \bar{X}Y + \bar{X}\bar{Y})$$

$$= \Sigma XY - \bar{Y}\Sigma X - \bar{X}\Sigma Y + n\bar{X}\bar{Y}$$

$$= \Sigma XY - n\bar{Y}\bar{X} - n\bar{X}\bar{Y} + n\bar{X}\bar{Y}$$

$$\Sigma(X - \bar{X})(Y - \bar{Y}) = \Sigma XY - n\bar{X}\bar{Y} \text{----------(2.15)}$$

$$\Sigma(X - \bar{X})^2 = \Sigma X^2 - n\bar{X}^2 \text{----------(2.16)}$$

Substituting (2.15) and (2.16) in (2.14), we get

$$\hat{\beta} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(X - \bar{X})^2}$$

Now, denoting $(X_i - \bar{X})$ as x_i, and $(Y_i - \bar{Y})$ as y_i we get;

$$\hat{\beta} = \frac{\Sigma x_i y_i}{\Sigma x_i^2} \quad \text{..... (2.17)}$$

The expression in (2.17) to estimate the parameter coefficient is termed is the formula in *deviation form*.

2.2.2.2 Estimation of a function with zero intercept

Suppose it is desired to fit the line $Y_i=\alpha+\beta X_i+U_i$, subject to the restriction $\alpha=0$. To estimate $\hat{\beta}$, the problem is put in a form of restricted minimization problem and then Lagrange method is applied.

We minimize: $\Sigma e_i^2=\sum_{i=1}^{n}(Y_i-\hat{\alpha}-\hat{\beta}X_i)^2$

Subject to: $\hat{\alpha}=0$

The composite function then becomes

$Z=\sum(Y_i-\hat{\alpha}-\hat{\beta}X_i)^2-\lambda\hat{\alpha}$, where λ is a Lagrange multiplier.

We minimize the function with respect to $\hat{\alpha},\hat{\beta},\ and\ \lambda$

$$\frac{\partial Z}{\partial\hat{\alpha}}=-2\Sigma(Y_i-\hat{\alpha}-\hat{\beta}X_i)-\lambda=0--------(i)$$

$$\frac{\partial Z}{\partial\hat{\beta}}=-2\Sigma(Y_i-\hat{\alpha}-\hat{\beta}X_i)(X_i)=0--------(ii)$$

$$\frac{\partial z}{\partial\lambda}=-2\alpha=0--------------------(iii)$$

Substituting (iii) in (ii) and rearranging we obtain:

$$\Sigma X_i(Y_i-\hat{\beta}X_i)=0$$

$$\Sigma Y_iX_i-\hat{\beta}\Sigma X_{i^2}=0$$

$$\hat{\beta}=\frac{\Sigma X_iY_i}{\Sigma X_i^2} \quad \text{.....(2.18)}$$

This formula involves the *actual values* (observations) of the variables and not their deviation forms, as in the case of unrestricted value of $\hat{\beta}$.

2.2.2.3. Statistical Properties of Least Square Estimators

There are various econometric methods with which we may obtain the estimates of the parameters of economic relationships. We would like

to an estimate to be as close as the value of the true population parameters i.e. to vary within only a small range around the true parameter. *How are we to choose among the different econometric methods, the one that gives 'good' estimates?* We need some criteria for judging the 'goodness' of an estimate.

'Closeness' of the estimate to the population parameter is measured by the mean and variance or standard deviation of the *sampling distribution* of the estimates of the different econometric methods. We assume the usual process of repeated sampling i.e. we assume that we get a very large number of samples each of size 'n'; we compute the estimates $\hat{\beta}$'s from each sample, and for each econometric method and we form their distribution. We next compare the *mean (expected value)* and the *variances* of these distributions and we choose among the alternative estimates the one whose distribution is concentrated as close as possible around the population parameter.

PROPERTIES OF OLS ESTIMATORS

The ideal or optimum properties that the OLS estimates possess may be summarized by well known theorem known as the ***Gauss-Markov Theorem.***

Statement of the theorem: "***Given the assumptions of the classical linear regression model, the OLS estimators, in the class of linear and unbiased estimators, have the minimum variance, i.e. the OLS estimators are BLUE.***

According to the this theorem, under the basic assumptions of the classical linear regression model, the least squares estimators are linear, unbiased and have minimum variance (i.e. are best of all linear unbiased estimators). Some times the theorem referred as the BLUE

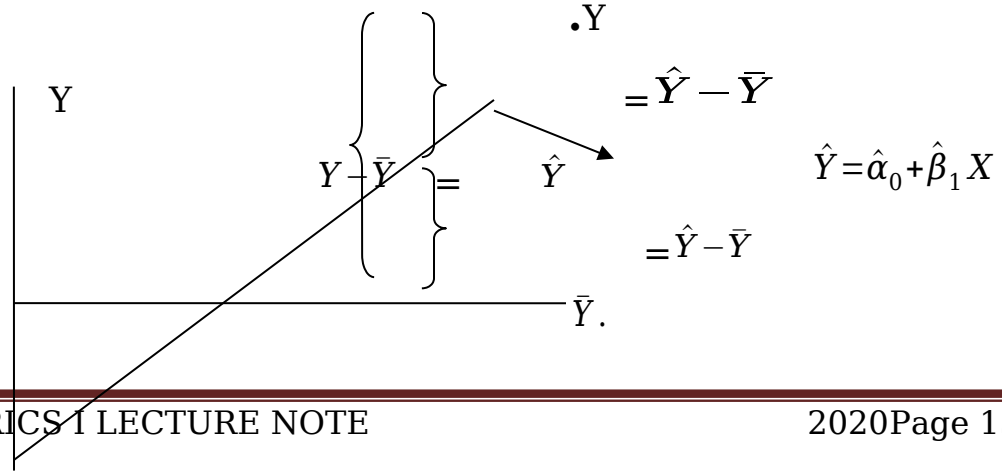
theorem i.e. Best, Linear, Unbiased Estimator. An estimator is called BLUE if:

a. ***Linear*****:** a linear function of the a random variable, such as, the dependent variable Y.
b. ***Unbiased***: its average or expected value is equal to the true population parameter.
c. ***Minimum variance***: It has a minimum variance in the class of linear and unbiased estimators. An unbiased estimator with the least variance is known as an efficient estimator.

According to the Gauss-Markov theorem, the OLS estimators possess all the BLUE properties.

1. TESTS OF THE '*GOODNESS OF FIT*' WITH R^2

r_2 shows the percentage of total variation of the dependent variable that can be explained by the changes in the explanatory variable(s) included in the model. To elaborate this let's draw a horizontal line corresponding to the mean value of the dependent variable $\bar{Y}$. (see figure'd' below). By fitting the line $\hat{Y}=\hat{\alpha}_0+\hat{\beta}_1 X$ we try to obtain the explanation of the variation of the dependent variable Y produced by the changes of the explanatory variable X.



X

Figure 'd'. Actual and estimated values of the dependent variable Y.

As can be seen from fig.(d) above, $Y-\bar{Y}$ represents measures the variation of the sample observation value of the dependent variable around the mean. However the variation in Y that can be attributed the influence of X, (i.e. the regression line) is given by the vertical distance $\hat{Y}-\bar{Y}$. The part of the total variation in Y about $\bar{Y}$ that can't be attributed to X is equal to $\hat{Y}-\bar{Y}$ which is referred to as the residual variation.

In summary:

$e_i=Y_i-\hat{Y}$ = deviation of the observation Y_i from the regression line.

$y_i=Y-\bar{Y}$ = deviation of Y from its mean.

$\hat{y}=\hat{Y}-\bar{Y}$ = deviation of the regressed (predicted) value ($\hat{Y}$) from the mean.

Now, we may write the observed Y as the sum of the predicted value ($\hat{Y}$) and the residual term (e_i).

$$\underbrace{Y_i}_{Observed\ Y_i} = \underbrace{\hat{Y}}_{predicted\ Y_i} \underbrace{+e_i}_{\text{Residual}}$$

Fromequation (2.34) we can have the above equation but in deviation form

$y=\hat{y}+e$. By squaring and summing both sides, we obtain the following expression:

$$\Sigma y^2=\Sigma(\hat{y}^2+e)^2$$

$$\Sigma y^2 = \Sigma(\hat{y}^2 + e_i^2 + 2\,yei)$$

$$= \Sigma y_{i^2} + \Sigma e_i^2 + 2\,\Sigma\,\hat{y}\,ei$$

But $\Sigma\,\hat{y}ei = \Sigma e(\hat{Y} - \bar{Y}) = \Sigma e(\hat{\alpha} + \hat{\beta}\,x_i - \bar{Y})$

$$= \hat{\alpha}\,\Sigma e_i + \hat{\beta}\,\Sigma e x_i - \hat{Y}\,\Sigma e_i$$

(but $\Sigma e_i = 0\,,\,\Sigma\,ex_i = 0$)

$$\Rightarrow \sum \hat{y}\,e = 0 \quad \text{.....(2.46)}$$

Therefore;

$$\underbrace{\Sigma y^2}_{\text{variation}} = \underbrace{\Sigma \hat{y}^2}_{\text{variation}} + \underbrace{\Sigma e^2}_{\text{variation}} \quad \text{.....(2.47)}$$

OR,

$$\underbrace{\text{Total sum of square}}_{TSS} = \underbrace{\text{Explained sum of square}}_{ESS} + \underbrace{\text{Residual sum of square}}_{RSS}$$

i.e

$$TSS = ESS + RSS \quad \text{.....(2.48)}$$

Mathematically; the explained variation as a percentage of the total variation is explained as:

$$\frac{ESS}{TSS} = \frac{\Sigma\,\hat{y}^2}{\Sigma y^2} \quad \text{.....(2.49)}$$

From equation (2.37) we have $\hat{y} = \hat{\beta}\,x$. Squaring and summing both sides give us

$$\Sigma\,\hat{y}^2 = \hat{\beta}^2\,\Sigma x^2 \text{----------(2.50)}$$

We can substitute (2.50) in (2.49) and obtain:

$$ESS/TSS = \frac{\hat{\beta}^2\,\Sigma x^2}{\Sigma y^2} \quad \text{..... .....(2.51)}$$

$$=\left(\frac{\Sigma xy}{\Sigma x^2}\right)^2\frac{\Sigma x_i^2}{\Sigma y^2}, \quad \text{Since } \hat{\beta}=\frac{\Sigma x_i y_i}{\Sigma x_i^2}$$

$$=\frac{\Sigma xy}{\Sigma x^2}\frac{\Sigma xy}{\Sigma y^2} \quad \text{.....(2.52)}$$

Comparing (2.52) with the formula of the correlation coefficient:

$r = Cov(X,Y) / \sigma_x^2\sigma_x^2 = \Sigma xy / n\sigma_x^2\sigma_x^2 = \Sigma xy / (\Sigma x^2 \Sigma y^2)^{1/2}$ .....(2.53)

Squaring (2.53) will result in: $r^2 = (\Sigma xy)^2 / (\Sigma x^2 \Sigma y^2)$. .....(2.54)

Comparing (2.52) and (2.54), we see exactly the expressions. Therefore:

$$\text{ESS/TSS} = \frac{\Sigma xy}{\Sigma x^2}\frac{\Sigma xy}{\Sigma y^2} = r^2$$

From (2.48), RSS=TSS-ESS. Hence R^2 becomes;

$$R^2=\frac{TSS-RSS}{TSS}=1-\frac{RSS}{TSS}=1-\frac{\Sigma e_i^2}{\Sigma y^2} \quad \text{.....(2.55)}$$

From equation (2.55) we can drive;

$$RSS=\Sigma e_i^2=\Sigma y_i^2(1-R^2)\text{----------(2.56)}$$

The limit of R^2: The value of R^2 falls between zero and one. i.e. $0 \le R^2 \le 1$.

- ***Interpretation of R^2***

Suppose $R^2=0.9$, this means that the regression line gives a good fit to the observed data since this line explains 90% of the total variation of the Y value around their mean. The remaining 10% of the total variation in Y is unaccounted for by the regression line and is attributed to the factors included in the disturbance variable u_i.

2. TESTING THE SIGNIFICANCE OF OLS PARAMETERS

To test the significance of the OLS parameter estimators we need the following:

- Variance of the parameter estimators
- Unbiased estimator of σ^2
- The assumption of normality of the distribution of error term.
- $\text{var}(\hat{\beta})=\frac{\hat{\sigma}^2}{\Sigma x^2}$
- $\text{var}(\hat{\alpha})=\frac{\hat{\sigma}^2 \Sigma X^2}{n\Sigma x^2}$
- $\hat{\sigma}^2=\frac{\Sigma e^2}{n-2}=\frac{RSS}{n-2}$

For the purpose of estimation of the parameters the assumption of normality is not used, but we use this assumption to test the significance of the parameter estimators; because the testing methods or procedures are based on the assumption of the normality assumption of the disturbance term. Hence before we discuss on the various testing methods it is important to see whether the parameters are normally distributed or not.

We have already assumed that the error term is normally distributed with mean zero and variance σ^2, i.e. $U_i \sim N(0,\sigma^2)$. Similarly, we also proved that $Y_i \sim N[(\alpha+\beta x),\sigma^2]$. Now, we want to show the following:

1. $\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\Sigma x^2}\right)$
2. $\hat{\alpha} \sim N\left(\alpha, \frac{\sigma^2 \Sigma X^2}{n\Sigma x^2}\right)$

To show whether $\hat{\alpha}$ *and* $\hat{\beta}$ are normally distributed or not, we need to make use of one property of normal distribution. *"..... any linear*

function of a normally distributed variable is itself normally distributed."

$$\hat{\beta}=\Sigma k_i Y_i=k_1 Y_1+k_2 Y_{2i}+....+k_n Y_n$$

$$\hat{\alpha}=\Sigma w_i Y_i=w_1 Y_1+w_2 Y_{2i}+....+w_n Y_n$$

Since $\hat{\alpha}\ and\ \hat{\beta}$ are linear in Y, it follows that

$$\hat{\beta}\sim N\left(\beta,\frac{\sigma^2}{\Sigma x^2}\right) \quad ; \quad \hat{\alpha}\sim N\left(\alpha,\frac{\sigma^2\Sigma X^2}{n\Sigma x^2}\right)$$

The OLS estimates $\hat{\alpha}\ and\ \hat{\beta}$ are obtained from a sample of observations on Y and X. Since sampling errors are inevitable in all estimates, it is necessary to apply test of significance in order to measure the size of the error and determine the degree of confidence in order to measure the validity of these estimates. This can be done by using various tests. The most common ones are:

i) Standard error test ii) Student's t-test iii) Confidence interval

All of these testing procedures reach on the same conclusion. Let us now see these testing methods one by one.

i) Standard error test

This test helps us decide whether the estimates $\hat{\alpha}\ and\ \hat{\beta}$ are significantly different from zero, i.e. whether the sample from which they have been estimated might have come from a population whose true parameters are zero. $\alpha=0\ and/or\ \beta=0$.

Formally we test the null hypothesis

$H_0:\beta_i=0$ against the alternative hypothesis $H_1:\beta_i\neq 0$

The standard error test may be outlined as follows.

First: Compute standard error of the parameters.

$$SE(\hat{\beta})=\sqrt{\text{var}(\hat{\beta})}$$

$$SE(\hat{\alpha})=\sqrt{\text{var}(\hat{\alpha})}$$

Second: compare the standard errors with the numerical values of $\hat{\alpha}$ *and* $\hat{\beta}$.

Decision rule:

- If $SE(\hat{\beta}_i)>{}^{1}\!/_{2}\hat{\beta}_i$, accept the null hypothesis and reject the alternative hypothesis. We conclude that $\hat{\beta}_i$ is statistically insignificant.
- If $SE(\hat{\beta}_i)<{}^{1}\!/_{2}\hat{\beta}_i$, reject the null hypothesis and accept the alternative hypothesis. We conclude that $\hat{\beta}_i$ is statistically significant.

The acceptance or rejection of the null hypothesis has definite economic meaning. Namely, the acceptance of the null hypothesis $\beta=0$ (the slope parameter is zero) implies that the explanatory variable to which this estimate relates does not in fact influence the dependent variable Y and should not be included in the function, since the conducted test provided evidence that changes in X leave Y unaffected. In other words acceptance of H_0 implies that the relation ship between Y and X is in fact $Y=\alpha+(0)x=\alpha$, i.e. there is no relationship between X and Y.

ii) Student's t-test

Like the standard error test, this test is also important to test the significance of the parameters. From your statistics, any variable X can be transformed into t using the general formula:

$t = \frac{X - \mu}{s_x}$, with n-1 degree of freedom.

Where $\mu_i =$ value of the population mean

$s_x =$ sample estimate of the population standard deviation

$$s_x = \sqrt{\frac{\Sigma (X - \bar{X})^2}{n-1}}$$

$n =$ sample size

We can derive the t-value of the OLS estimates

$t_{\hat{\beta}} = \frac{\hat{\beta}_i - \beta}{SE(\hat{\beta})}$ ¿¿¿¿ with n-k degree of freedom.

Where:

SE = is standard error

k = number of parameters in the model.

Since we have two parameters in simple linear regression with intercept different from zero, our degree of freedom is n-2. Like the standard error test we formally test the hypothesis: $H_0 : \beta_i = 0$ against the alternative $H_1 : \beta_i \neq 0$ for the slope parameter; and $H_0 : \alpha = 0$ against the alternative $H_1 : \alpha \neq 0$ for the intercept.

To undertake the above test we follow the following steps.

Step 1: Compute t*, which is called the computed value of t, by taking the value of β in the null hypothesis. In our case $\beta=0$, then t* becomes:

$$t* ¿ \frac{\hat{\beta}-0}{SE(\hat{\beta})}=\frac{\hat{\beta}}{SE(\hat{\beta})}$$

Step 2: Choose level of significance. Level of significance is the probability of making 'wrong' decision, i.e. the probability of rejecting the hypothesis when it is actually true or the probability of committing a type I error. It is customary in econometric research to choose the 5% or the 1% level of significance. This means that in making our decision we allow (tolerate) five times out of a hundred to be 'wrong' i.e. reject the hypothesis when it is actually true.

Step 3: Check whether there is one tail test or two tail test. If the inequality sign in the alternative hypothesis is ¿, then it implies a two tail test and divide the chosen level of significance by two; decide the critical rejoin or critical value of t called t_c. But if the inequality sign is either > or < then it indicates one tail test and there is no need to divide the chosen level of significance by two to obtain the critical value of to from the t-table.

iii) Confidence interval

Rejection of the null hypothesis doesn't mean that our estimate $\hat{\alpha}$ *and* $\hat{\beta}$ is the correct estimate of the true population parameter α *and* β. It simply means that our estimate comes from a sample drawn from a population whose parameter β is different from zero.

In order to define how close the estimate to the true parameter, we must construct confidence interval for the true parameter, in other words we must establish limiting values around the estimate with in

which the true parameter is expected to lie within a certain "degree of confidence". In this respect we say that with a given probability the population parameter will be with in the defined confidence interval (confidence limits).

We choose a probability in advance and refer to it as confidence level (interval coefficient). It is customarily in econometrics to choose the 95% confidence level. This means that in repeated sampling the confidence limits, computed from the sample, would include the true population parameter in 95% of the cases. In the other 5% of the cases the population parameter will fall outside the confidence interval.

In a two-tail test at α level of significance, the probability of obtaining the specific t-value either –tc or tc is $\alpha/2$ at n-2 degree of freedom. The probability of obtaining any value of t which is equal to $\frac{\hat{\beta}-\beta}{SE(\hat{\beta})}$ at n-2 degree of freedom is $1-\left(\alpha/2+\alpha/2\right)\ i.e.\ 1-\alpha$.

i.e. $\Pr\left[-t_c<t*¿t_c\right]=1-\alpha$ .....(2.57)

but $t*¿\frac{\hat{\beta}-\beta}{SE(\hat{\beta})}$ .....(2.58)

Substitute (2.58) in (2.57) we obtain the following expression.

$$\Pr\left[-t_c<\frac{\hat{\beta}-\beta}{SE(\hat{\beta})}<t_c\right]=1-\alpha \quad \text{.....(2.59)}$$

$$\Pr\left[-SE(\hat{\beta})t_c<\hat{\beta}-\beta<SE(\hat{\beta})t_c\right]=1-\alpha \text{-----by multiplying } SE(\hat{\beta})$$

$$\Pr\left[-\hat{\beta}-SE(\hat{\beta})t_c<-\beta<-\hat{\beta}+SE(\hat{\beta})t_c\right]=1-\alpha \text{-----by subtracting } \hat{\beta}$$

$$\Pr\left[+\hat{\beta}+SE(\hat{\beta})>\beta>\hat{\beta}-SE(\hat{\beta})t_c\right]=1-\alpha \text{-----by multiplying by } -1$$

$$\Pr\left[\hat{\beta}-SE(\hat{\beta})t_c<\beta<\hat{\beta}+SE(\hat{\beta})t_c\right]=1-\alpha \text{-----int erchanging}$$

The limit within which the true β lies at $(1-\alpha)\%$ degree of confidence is:

$[\hat{\beta}-SE(\hat{\beta})t_c, \hat{\beta}+SE(\hat{\beta})t_c]$; where t_c is the critical value of t at $\alpha/2$ confidence interval and n-2 degree of freedom.

The test procedure is outlined as follows.

$$H_0: \beta=0$$

$$H_1: \beta\neq 0$$

Decision rule: If the hypothesized value of β in the null hypothesis is within the confidence interval, accept H_0 and reject H_1. The implication is that $\hat{\beta}$ is statistically insignificant; while if the hypothesized value of β in the null hypothesis is outside the limit, reject H_0 and accept H_1. This indicates $\hat{\beta}$ is statistically significant.

2.2.3 Reporting the Results of Regression Analysis

The results of the regression analysis derived are reported in conventional formats. It is not sufficient merely to report the estimates of β's. In practice we report regression coefficients together with their standard errors and the value of R^2. It has become customary to present the estimated equations with standard errors placed in parenthesis below the estimated parameter values. Sometimes, the estimated coefficients, the corresponding standard errors, the p-values, and some other indicators are presented in tabular form.

These results are supplemented by R^2 on (to the right side of the regression equation).

Example: $\underset{(38.2)\ (0.85)}{Y=128.5+2.88X}$, $R^2 = 0.93$. The numbers in the parenthesis below the parameter estimates are the standard errors. Some econometricians report the t-values of the estimated coefficients in place of the standard errors.