SIMPLE REGRESSION MODEL

In the simple regression cases of regression analysis is found when the dependent variable is related linearly to a single independent variable.

One of the assumpition of ordinary least square method is the X_i's are set of fixed values in the hypothetical process of ~~rotated~~ repeated sampling. Y's is dependent variable on X and Y is assumed to be random variable, while the X values are generally fixed and selected by the experimenter. The population sample regression equation is -

$$Y_i = b_0 + b_1 X + u_i$$

Where Y_i is the dependent variable, X_i is the independent variable and u_i is the error term; the only randomness in the values of Y comes the error term u_i.

The simple regression equation $Y = \beta_0 + \beta_1 X + u$ is composed of two components. They are: i) a non-random component which is explained by the regression line, ii) a purely random component - the error term u.

$$Y = \underbrace{\beta_0 + \beta_1 X}_{\text{Non-random component}} + \underbrace{u}_{\text{Random component}}$$

STOCHASTIC AND NON-STOCHASTIC RELATIONS

Assumpition of Linear Stochastic Regression Model:

Assumpition of Linear stochastic Regression model can be grouped in two catagories:

(a). Stochastic Assumpitions

(b). Other Assumpitions.

(a) Stochastic Assumpition:

These assumpitions are related to the distribution of the value of u.

1. u_i is a random real variable:

The value of u_i in any one period depends on chance, it may be positive, negative or zero.

2. Mean of u_i in any particular period is zero.

This means that for each value of x_i, u_i may assume various values, some greater than zero and some smaller than zero, but the ~~ss~~ average value of u_i is equal to zero, symbolically, $E(u_i) = 0$ i,e. Expected value of u_i is zero.

3. Variance of u_i is constant in each period.

The variance of u_i is constant, i.e $E(u_i^2) = \sigma^2$ (σ^2 is a constant) where $i = 1, 2 - - - n$.

4. The variable u_i has normal distribution

The values of u_i have a bell shaped Symmetrical distribution about their zero mean.

Symbolically, $u_i \sim N(0, \sigma_u^2)$

5. Different observations of error term are independent

This means that the covariance of any u_i with any other u_j are equal to zero.

$$E(u_i u_j) = 0 \quad i \neq j$$

This assumpition is also called as a absance of auto-correlation or non-autocorrelation.

6. u_i is independent of the explanatory variables.

The u_i is not correlated with explanatory variables

$$Cov(X_i, u_i) = E[\{X_i - E(X_i)\}\{u_i - E(u_i)\}] = 0$$

This means that $X's$ is non-stochastic.

7. $X's$ are set of fixed values in the hypothetical process of repeated sampling.

8. Explanatory variables are measured without error.

It is assumed that U absorbs the effect of omitted variables and also the error of measurement in the $Y's$.

The regression equation: $Y = \alpha + \beta X + U$ along with the given assumptions represents the Classical Linear Regression Model. These assumptions have important role to play in the sampling distributions of parameters: α and β.

The effect of first three assumption $Y_i = \alpha + \beta X_i + U_i$ Y_i is a linear function u_i. Since it is assumed that U is normally distributed, it follows that Y_i is also normally distributed.

$$Y_i = \alpha + \beta X_i + u_i$$

$$\therefore E(Y_i) = E[\alpha + \beta X_i + u_i]$$

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

$$= \alpha + \beta X_i \quad [\text{Since } E(u_i) = 0]$$

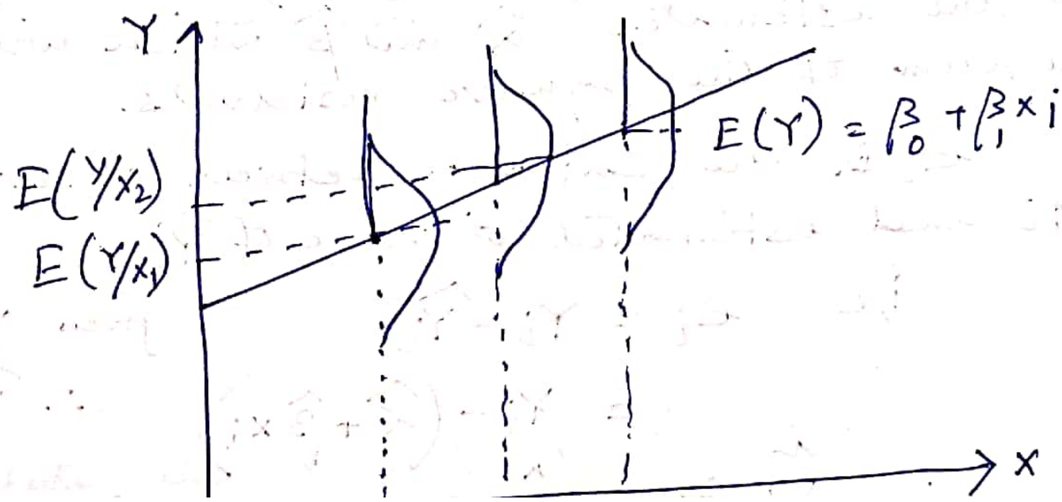
Therefore, we say that mean of Y_i is given by $(\alpha + \beta X_i)$.

Now,
$$Var(Y_i) = E[Y_i - E(Y_i)]^2$$
$$= E[\alpha + \beta x_i + u_i - (\alpha + \beta x_i)]^2$$
$$= E(u_i)^2 = \sigma^2 \quad [\text{Since } E(u_i^2) = \sigma^2]$$

Therefore, we say that variance of Y_i is equal to σ^2.

Symbolically, it can be written as
$$Y_i \sim N[\alpha + \beta x_i, \sigma^2]$$

The assumpitions about the behaviour of the values of u_i may be explained by the following diagram.



The true relation between X and Y is given by

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

The estimated regression line is -

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

The estimated relation is

$$Y_i = \hat{\beta}_0 + \beta_1 X_i + e_i$$

where e_i is the estimate of true value of random term ~~of~~ u_i

ESTIMATION OF REGRESSION PARAMETERS

Estimation of α and β by least square method (OLS) or classical least square (CLS) involves finding value for the estimates $\hat{\alpha}$ and $\hat{\beta}$ which will minimise the sum of the square residuals.

e_i is the differance between observed value of Y_i and estimated ~~o~~ value of $\hat{Y}_i$,

i.e $\quad e_i = Y_i - \hat{Y}_i$ $\qquad$ given $Y_i = \alpha + \beta X_i + u_i$

$\quad = Y_i - (\hat{\alpha} + \hat{\beta} X_i)$ $\qquad \therefore \hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$

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

To find the values of $\hat{\alpha}$ and $\hat{\beta}$ that minimise their sum ~~differance~~ differentiate the sum of the squares of the residuals with respect to $\hat{\alpha}$ and $\hat{\beta}$ and equate the partial derivatives equal to zero.

$$\frac{\partial}{\partial \hat{\alpha}}\left[\sum e_i^2\right] = 2\sum (Y_i - \hat{\alpha} - \hat{\beta} X_i)(-1) = 0$$

$$\Rightarrow \sum (Y_i - \hat{\alpha} - \hat{\beta} X_i) = 0 \quad \text{--- (1)}$$

$$\cancel{\frac{\partial}{\partial \hat{\beta}}\left[\sum e_i^2\right] = 2\sum (Y_i - \hat{\alpha} - \hat{\beta} X_i)(-\hat{\beta}) = 0}$$

Scanned by CamScanner

From equation (i) we get,

$$\sum Y_i - n\hat{\alpha} - \hat{\beta}\sum X_i = 0$$

$$\text{or} \quad \sum Y_i = n\hat{\alpha} + \hat{\beta}\sum X_i \quad \text{———} \quad (2)$$

Again,

$$\frac{\partial}{\partial \hat{\beta}}\left[\sum e_i^2\right] = 0$$

$$\Rightarrow \quad 2\sum\left(Y_i - \hat{\alpha} - \hat{\beta}X_i\right)(-X_i) = 0.$$

$$\Rightarrow \quad \sum\left(X_iY_i - \hat{\alpha}X_i - \hat{\beta}X_i^2\right) = 0$$

$$\Rightarrow \quad \sum X_iY_i - \hat{\alpha}\sum X_i - \hat{\beta}\sum X_i^2 = 0 \quad \text{———} \quad (3)$$

From equation (2) $\Rightarrow$

$$\frac{\sum Y_i}{n} = \hat{\alpha} + \hat{\beta}\sum X_i/n.$$

$$\text{or} \quad \bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X}$$

Since $\bar{Y} = \sum Y_i/n$

$\bar{X} = \sum X_i/n.$

$$\therefore \quad \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

Substitute the value of $\hat{\alpha}$ in eqn-(3) we get,

$$\sum X_iY_i - \left(\bar{Y} - \hat{\beta}\bar{X}\right)\sum X_i - \hat{\beta}\sum X_i^2 = 0$$

$$\text{or} \quad \sum X_iY_i - \bar{Y}\sum X_i + \hat{\beta}\bar{X}\sum X_i - \hat{\beta}\sum X_i^2 = 0$$

$$\sum x_i y_i - \bar{Y}\sum x_i = \hat{\beta}\sum x_i^2 - \hat{\beta}\bar{x}\sum x_i$$

$$\therefore \hat{\beta} = \frac{\sum x_i y_i - \bar{Y}\sum x_i}{\sum x_i^2 - \bar{x}\sum x_i}$$

$$= \frac{\sum x_i y_i - \frac{\sum y_i}{n}\sum x_i}{\sum x_i^2 - \frac{\sum x_i}{n}\sum x_i} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$= \frac{n\sum x_i y_i - \bar{x}\bar{Y}}{\sum x_i^2 - n\left(\frac{\sum x_i}{n}\right)^2}$$

$$= \frac{n\sum x_i y_i - \bar{x}\bar{Y}}{\sum x_i^2 - n(\bar{x})^2}$$

~~Now,~~ $\cancel{n\sum x_i y_i - \bar{x}\bar{Y}}$

Now, $n\sum x_i y_i - \sum x_i \sum y_i$

$$= \left(n\sum x_i y_i - \sum x_i \sum y_i\right) + \left(\sum x_i \sum y_i - \sum x_i \sum y_i\right)$$

Since $\sum x_i = n\bar{x}$, $\sum y_i = n\bar{Y}$

$$= n\sum x_i y_i - n\bar{x}\sum y_i - n\bar{Y}\sum x_i + \sum x_i \sum y_i$$

$$= n\sum x_i y_i - n\bar{x}\sum y_i - n\bar{Y}\sum x_i + n^2\bar{x}\cdot\bar{Y}$$

$$= n\left[\sum x_i y_i - \bar{x}\sum y_i - \bar{Y}\sum x_i + n\bar{x}\bar{Y}\right]$$

$$= n\sum_i (x_i - \bar{x})(y_i - \bar{Y})$$

Scanned by CamScanner

Again,

$$n\sum x_i^2 - \left(\sum x_i\right)^2$$

$$= n\sum x_i^2 - 2\left(\sum x_i\right)^2 + \left(\sum x_i\right)^2$$

$$= n\sum x_i^2 - 2n\bar{x}\sum x_i + n^2\bar{x}^2$$

$$= n\left[\sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2\right]$$

$$= n\sum (x_i-\bar{x})^2$$

$$\therefore \hat{\beta} = \frac{n\sum(x_i-\bar{x})(Y_i-\bar{Y})}{n\sum(x_i-\bar{x})^2}$$

Let $x_i - \bar{x} = x_i$ and $Y_i - \bar{Y} = y_i$

$$\therefore \hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

Statistical Properties of Least Squares Estimation

(i) Linearity :-

Least squares estimators are linear : Now recall the equation for $\hat{\beta}$.

$$\hat{\beta} = \frac{\sum(x_i-\bar{x})(Y_i-\bar{Y})}{\sum(x_i-\bar{x})^2}$$

$$= \frac{\sum Y_i(x_i-\bar{x}) - \bar{Y}\sum(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}$$

$$= \frac{\sum Y_i(x_i-\bar{x})}{\sum(x_i-\bar{x})^2} \quad \left[\text{Since } \bar{Y}\sum(x_i-\bar{x}) = 0\right]$$

$$\therefore \hat{\beta} = \frac{\sum Y_i x_i}{\sum(x_i-\bar{x})^2} \quad \text{where } x_i = (x_i-\bar{x})$$

Let us suppose that,

$$K_i = \frac{\sum x_i}{\sum x_i^2} \qquad (i = 1, 2 \cdots n)$$

$$\therefore \hat{\beta} = \sum_{i=1}^{n} K_i Y_i$$

from the least square formula we get,

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$= \frac{1}{n}\sum Y_i - \bar{X}\sum K_i Y_i$$

$$= \frac{1}{n}\sum Y_i - \bar{X}. K_i \sum Y_i$$

$$= \sum\left[\frac{1}{n} - \bar{X}K_i\right]\sum Y_i \qquad \text{(i)}$$

Thus both $\hat{\alpha}$ and $\hat{\beta}$ are expressed as linear function of Y_i's.

II) Unbiasedness :-

We know that, $\hat{\beta} = \sum K_i Y_i$

$$= \sum K_i(\alpha + \beta X_i + u_i)$$

$$= \alpha\sum K_i + \beta\sum K_i X_i + \sum K_i u_i \quad \text{(ii)}$$

Where $K_i = \frac{\sum x_i}{\sum x_i^2}$

and $K_i = \frac{\sum x_i}{\sum x_i^2} = \frac{\sum(X_i - \bar{X})}{\sum x_i^2} = 0.$

Now Putting the value $x_i = X_i - \bar{X}$ Since $\sum(X_i - \bar{X}) = 0$

$$\therefore X_i = x_i + \bar{X}.$$

Then, $\sum K_i X_i = \cancel{\sum K_i(x_i + \bar{X}) = \sum K_i x_i + \bar{X}\sum K_i}$

$$= \frac{\sum x_i}{\sum x_i^2}.X_i$$

$$= \frac{\sum x_i(x_i + \bar{x})}{\sum x_i^2}$$

$$= \frac{\sum x_i^2}{\sum x_i^2} + \frac{\bar{x}\sum x_i}{\sum x_i^2}$$

$$= 1 + 0 = 1$$ since $\sum x_i = \sum(X_i - \bar{X}) = 0$.

Substituting the value of $\sum K_i X_i$ in eqn-(ii)

$\hat{\beta} = \alpha \cdot 0 + \beta \cdot 1 + \sum K_i u_i$ since $\sum K_i = 0$, $\sum K_i X_i = 1$.

$$\therefore \hat{\beta} = \beta + \sum K_i u_i \quad \text{——— (3)}$$

$$\therefore E(\hat{\beta}) = E(\beta) + \sum K_i E(u_i)$$

$$= \beta + 0 = \beta$$ since $E(u_i) = 0$ and $E(\beta) = \beta$.

From eqn (i) we get,

$$\hat{\alpha} = \sum\left[\frac{1}{n} - \bar{X}K_i\right] Y_i$$

$$= \sum\left[\frac{1}{n} - \bar{X}K_i\right](\alpha + \beta X_i + u_i)$$

$$= \alpha + \beta \cdot \frac{1}{n}\sum X_i + \frac{1}{n}\sum u_i - \alpha\bar{X}\sum K_i - \beta\bar{X}\sum K_i X_i - \bar{X}\sum K_i u_i$$

$$= \alpha + \beta\bar{X} + \frac{1}{n}\sum u_i - 0 - \beta\bar{X}\cdot 1 - \bar{X}\sum K_i u_i$$

since $\sum K_i = 0$ and $\sum K_i X_i = 1$.

$$= \alpha + \beta\bar{X} + \frac{1}{n}\sum u_i - \beta\bar{X} - \bar{X}\sum K_i u_i$$

$$= \alpha + \frac{1}{n}\sum u_i - \bar{X}\sum K_i u_i$$

$$\therefore E(\hat{\alpha}) = \alpha + \frac{1}{n}\sum E(u_i) - \bar{X}\sum K_i E(u_i)$$

$$= \alpha \quad \text{Since } E(u_i) = 0$$

Thus, it proves that $\hat{\alpha}$ and $\hat{\beta}$ are the unbiased estimator of α and β respectively.

(iii) <u>Minimum Variance of $\hat{\alpha}$ and $\hat{\beta}$.</u>

We know that,

$$Var(\hat{\beta}) = E[\hat{\beta} - E(\hat{\beta})]^2$$

$$= E[\hat{\beta} - \beta]^2 \quad \text{Since } E(\hat{\beta}) = \beta$$

From equation (3), we get,

$$\hat{\beta} - \beta = \sum(K_i u_i)$$

Now

$$var(\hat{\beta}) = E[\sum(K_i u_i)^2]$$

$$= E[K_1^2 u_1^2 + K_2^2 u_2^2 + \dots + K_n^2 u_n^2 + 2K_1K_2u_1u_2 + \dots + 2K_{n-1}K_n u_{n-1}u_n]$$

$$= [\sum K_1^2 E(u_1^2) + \sum K_2^2 E(u_2^2) + \dots + \sum K_n^2 E(u_n^2) + 2E[\sum_{i \neq j} K_i K_j u_i u_j]$$

$$= \sum K_i^2 E(u_i^2) + 2\sum_{i \neq j} K_i K_j E(u_i u_j)$$

$$= \sum K_i^2 \cdot \sigma^2 + 0 \quad \text{Since } E(u_i^2) = \sigma^2 \text{ and } E(u_i u_j) = 0.$$

$$= \sigma^2 \sum K_i^2$$

Scanned by CamScanner

$\therefore \ Var(\hat{\beta}) = \sigma^2 \sum k_i^2$

We know that $k_i = \dfrac{\sum x_i}{\sum x_i^2}$

$$\therefore \sum k_i^2 = \frac{\sum x_i^2}{(\sum x_i)^2}$$

$$= \frac{1}{\sum x_i^2}$$

$$\therefore \ Var(\hat{\beta}) = \sigma^2 \sum k_i^2$$

$$= \sigma^2 \cdot \frac{1}{\sum x_i^2} = \frac{\sigma^2}{\sum x_i^2} \quad \text{--- (4)}$$

Now

$$Var(\hat{\alpha}) = E\left[\hat{\alpha} - E(\hat{\alpha})\right]^2$$

$$= E\left[\hat{\alpha} - \alpha\right]^2 \quad \text{Since } E(\hat{\alpha}) = \alpha.$$

We know that $\hat{\alpha} = \alpha + \frac{1}{n}\sum u_i - \bar{X} \sum k_i u_i$

$$\therefore (\hat{\alpha} - \alpha) = \frac{1}{n}\sum u_i - \bar{X} \sum k_i u_i$$

$$\therefore \ Var(\hat{\alpha}) = E\left[\left(\frac{1}{n}\sum u_i - \bar{X}\sum k_i u_i\right)^2\right]$$

$$= E\left[\sum\left(\frac{1}{n} - \bar{X} k_i\right)^2 u_i^2\right]$$

$$= \sum\left(\frac{1}{n} - \bar{X} k_i\right)^2 E(u_i^2)$$

$$= \sigma^2 \sum\left(\frac{1}{n} - \bar{X} k_i\right)^2 \quad \left[\text{Since } E(u_i^2) = \sigma^2\right]$$

$$= \sigma^2 \sum\left[\frac{1}{n^2} - \frac{2}{n}\bar{X} k_i + \bar{X}^2 k_i^2\right]$$

$$= \sigma^2\left[\frac{1}{n} - \frac{2\bar{X}}{n}\sum k_i + \bar{X}^2 \sum k_i^2\right]$$

Scanned by CamScanner

$$= \sigma^2\left[\frac{1}{n} + \bar{x}^2 \cdot \frac{1}{\sum x_i^2}\right]$$

Since $\sum k_i = 0$

and $\sum k_i^2 = \frac{1}{\sum x_i^2}$

$$= \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum x_i^2}\right]$$

Again, $\frac{1}{n} + \frac{\bar{x}^2}{\sum x_i^2}$

$$= \frac{\sum x_i^2}{n\sum x_i^2} + \frac{n\bar{x}^2}{n\sum x_i^2}$$

$$= \frac{\sum x_i^2 + n\bar{x}^2}{n\sum x_i^2}$$

$$= \frac{\sum (x_i - \bar{x})^2 + n\bar{x}^2}{n\sum x_i^2}$$

$$= \frac{\sum (x_i^2 - 2\bar{x}\sum x_i + \bar{x}^2) + n\bar{x}^2}{n\sum x_i^2}$$

$$= \frac{\sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2 + n\bar{x}^2}{n\sum x_i^2}$$

$$= \frac{\sum x_i^2 - 2\bar{x}\cdot n\bar{x} + n\bar{x}^2 + n\bar{x}^2}{n\sum x_i^2}$$

$$= \frac{\sum x_i^2}{n\sum x_i^2}$$

$$\therefore var(\hat{\alpha}) = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum x_i^2}\right] = \frac{\sum x_i^2}{n\sum x_i^2}$$

Scanned by CamScanner

In order to establish that $\hat{\beta}$ possesses the minimum variance property, we compared its variance with that of some alternative unbiased estimator of β, say β^*.

Let, $\beta^* = \sum w_i Y_i$ where constant $w_i \neq K$
but $w_i = K_i + c$.

$$= \sum w_i [\alpha + \beta x_i + u_i]$$

$$= \alpha \sum w_i + \beta \sum w_i x_i + \sum w_i u_i \quad \text{—(5)}$$

$$\therefore E(\beta^*) = \alpha \sum w_i + \beta \sum w_i x_i + \sum w_i E(u_i)$$

$$= \alpha \sum w_i + \beta \sum w_i x_i \quad [\text{Since } E(u_i) = 0]$$

Since β^* is assumed to be an unbiased estimator which implies that $\sum w_i = 0$ and $\sum w_i x_i = 1$ in the above equation.

But $\sum w_i = \sum (K_i + c_i) = \sum K_i + \sum c_i$

$\therefore \sum c_i = 0 \quad \because \sum K_i = \sum w_i = 0$

Again, $\sum w_i x_i = \sum (K_i + c_i) x_i = \sum K_i x_i + \sum c_i x_i$

$\therefore \sum c_i x_i = 0 \quad \because \sum w_i x_i = 1$ and $\sum K_i x_i = \sum K_i x_i = 1$.

Also $\sum c_i x_i = \sum c_i x_i + \bar{X} \sum c_i = 0$.

Thus we have shown that if β^* is to be unbiased estimator then the following result must hold true.

$\sum w_i = 0, \quad \sum w_i x_i = 1, \quad \sum c_i = 0, \quad \sum c_i x_i = \sum c_i x_i = 0$

The variance of this assumed estimator β^* is then

$$Var(\beta^*) = E[\beta^* - \beta]^2$$
$$= E[(\sum w_i u_i)^2]$$
$$= \sigma^2 \sum w_i^2$$ by following exactly the same arguments that we used in obtaining $Var(\hat{\beta})$.

$$\therefore \cancel{Var(\beta^*)} \quad Var(\beta^*) = \sigma^2 \sum w_i^2$$

But $\sum w_i^2 = \sum (k_i + c_i)^2$

$$= \sum k_i^2 + 2\sum k_i c_i + \sum c_i^2$$

$$\therefore \sum k_i c_i = \sum c_i \cdot \frac{x_i}{\sum x_i^2} = \frac{\sum c_i x_i}{\sum x_i^2} = 0$$

[Since $\sum c_i x_i = 0$]

$\therefore \sum w_i^2 = \sum k_i^2 + \sum c_i^2$ so that

$$Var(\beta^*) = \sigma^2 [\sum k_i^2 + \sum c_i^2]$$
$$= \sigma^2 \sum k_i^2 + \sigma^2 \sum c_i^2$$
$$= Var(\hat{\beta}) + \sigma^2 \sum c_i^2$$

Where $\sum c_i^2$ must be positive, so that,

$$Var(\beta^*) > Var(\hat{\beta})$$

In case $\sum c_i^2 = 0$, then $Var(\beta^*) = Var(\hat{\beta})$

This proves that $\hat{\beta}$ possesses minimum variance property.

IMPORTANCE OF THE BEST LINEAR UNBIASED (BLU) PROPERTIES

(a) Linearity: This property is desirable because it facilitates the computations of the estimates.

(b) Unbiasedness: This property is not useful. The only assurance it gives is that if we have a very large number of samples, the estimators of the parameters obtained from these samples will ~~not~~ on the average give the true value of β's.

(c) Best: Also in this case, least variance property by itself is not particularly desirable, because an estimate may have zero variance and yet have an enormous bias.

VARIANCE OF THE RANDOM VARIABLE (U_i).

It has been shown that variance of $\hat{\alpha}$ and $\hat{\beta}$ are determined by the variance of the disturbance term in the equation, σ_u^2.

Let, $y_i = \alpha + \beta X_i + u_i$

$\sim \sum Y_i = \sum \alpha + \beta \sum X_i + \sum u_i$

$$\sim \frac{\sum Y_i}{n} = \frac{n\alpha}{n} + \beta \cdot \frac{\sum X_i}{n} + \frac{\sum u_i}{n}$$

$\sim \bar{Y} = \alpha + \beta \bar{X} + \bar{u}$

$\therefore \quad Y_i - \bar{Y} = \beta[X_i - \bar{X}] + (u_i - \bar{u})$

Let $y_i = Y_i - \bar{Y}$

$x_i = X_i - \bar{X}$

$\therefore \quad y_i = \beta x_i + (u_i - \bar{u})$

Therefore, $e_i = y_i - \hat{y}$ $\cancel{\quad} = \hat{\beta}_1 x_i$

$= \cancel{\beta \quad \beta x_i}$

Now $\cancel{\quad \beta x}$ $\cancel{\hat{Y}_i = \hat{\alpha} + \hat{\beta}} x_i$

$\therefore \cancel{\hat{Y}_i} =$

$= \beta x_i + (u_i - \bar{u}) - \hat{\beta} x_i$

(Substituting in $e_i = y_i - \hat{\beta} x_i$)

$$\therefore \quad e_i = (u_i - \bar{u}) - (\hat{\beta} - \beta) x_i$$

$$\therefore \quad \sum e_i^2 = \sum \left[(u_i - \bar{u}) - (\hat{\beta} - \beta) x_i \right]^2$$

$$= \sum (u_i - \bar{u})^2 - 2(\hat{\beta} - \beta) \sum (u_i - \bar{u}) x_i + (\hat{\beta} - \beta)^2 \sum x_i^2$$

$$\therefore \quad E\left[\sum e_i^2\right] = \sum E(u_i - \bar{u})^2 - 2(\hat{\beta} - \beta) \sum \cancel{E(\hat{\beta}_i - \beta) x_i} (u_i - \bar{u}) + E\left[(\hat{\beta} - \beta)^2\right] \sum x_i^2$$

Taking each term of R.H.S separately:-

$$\cancel{\sum E(u_i - \bar{u})^2 = E\left[\sum u_i^2 + \frac{1}{n} \sum u_i^2\right] + 2 \cdot \sum u_i \cdot \frac{\sum u_i}{n}}$$

$$= \cancel{\quad}$$

$$= \cancel{E\left[\sum u_i^2 - \frac{1}{n} \sum u_i^2\right]}$$

$$= \cancel{(n-1)} \sigma^2$$

<u>1st term:</u>

$$\sum E(u_i - \bar{u})^2 = E\left[\sum (u_i - \bar{u})^2\right]$$

$$= E\left[\sum u_i^2 - 2\bar{u} \cdot \sum u_i + n\bar{u}^2\right]$$

$$= E\left[\sum u_i^2 - 2n\bar{u}^2 + n\bar{u}^2\right]$$

Scanned by CamScanner

$$= E\left[\sum u_i^2 - \bar{u}^2\right]$$

$$= E\left[\sum u_i^2 - \frac{\sum u_i^2}{n}\right]$$

$$= E\left[\sum u_i^2 - \frac{(u_1 + u_2 + \cdots + u_n)^2}{n}\right]$$

$$= E\left(\sum u_i^2\right) - \left[\frac{E(u_i^2) + 2E\sum u_i u_j}{n}\right]$$

$$= n\sigma_u^2 - \frac{n\sigma^2}{n} \qquad \left(\text{Since } E\left(\sum u_i^2\right) = n\sigma^2\right)$$

$$= (n-1)\sigma^2$$

$$\therefore\ E\sum(u_i - \bar{u})^2 = (n-1)\sigma^2$$

<u>2nd term:</u>

$$E(\hat{\beta} - \beta)\sum(u_i - \bar{u})x_i$$

$$= E$$

The Sampling distribution of the Least square Estimates:

Since the least square estimators are linear combination of independent ~~va~~ normal variables $Y_1, Y_2, Y_3 --- Y_n$; $\hat{\alpha}$ and $\hat{\beta}$ must also normally distributed with the following characteristics:-

1) $\hat{\alpha}$ and $\hat{\beta}$ are unbiased estimators, their mean being equal to their values α and β.

11) Variance of each estimator is known.

Both these result may be stated in -

$$\hat{\alpha} \sim N\left[\alpha, \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right)\right]$$

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum x_i^2}\right)$$

Variance of the parameters are directly related to the variances of the disturbances, thus following points are very important:

1) Larger the value of σ^2, the larger the variances of $\hat{\alpha}$ and $\hat{\beta}$.

11) $\sum x_i^2$ is ~~on~~ the denominator of the variance formula of both estimators. This means the more dispersed the values of the explanatory variables (i,e. larger $\sum x_i^2$), the smaller the variances of $\hat{\alpha}$ and $\hat{\beta}$. If $\sum x_i^2 = 0$ or nearer to zero, i,e, when $x_1 = x_2 = x_3 --- = x_n$ both variances would be infinitely large.

111) The variance of $\hat{\alpha}$ is smallest when $\bar{X} = 0$ or nearer to zero. i,e when $\bar{X} = 0$, $var(\hat{\alpha}) = \sigma^2/n$.

Confidence intervals and Hypothesis Testing

Construction of confidence intervals is important in order to achieve precision of $\hat{\alpha}$ and $\hat{\beta}$.

So the standarised form will be.

$$Z = \frac{\hat{\beta} - \beta}{\sqrt{\sigma^2/\sum x_i^2}} = \frac{\hat{\beta} - \beta}{\sigma . \sqrt{1/\sum x_i^2}}$$

$$\text{i,e} \quad Z = (\hat{\beta} - \beta) . \frac{\sqrt{\sum x_i^2}}{\sigma}$$

Scanned by CamScanner

and $$Z^* = \frac{\hat{\alpha} - \alpha}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\Sigma x_i^2}\right)}}$$

$$= \frac{(\hat{\alpha} - \alpha)}{\sigma\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\Sigma x_i^2}}}$$

$$= \frac{(\hat{\alpha} - \alpha)}{\sigma\sqrt{\frac{1}{n} + \frac{n\bar{X}^2}{n\Sigma x_i^2}}}$$

$$= \frac{(\hat{\alpha} - \alpha)}{\sigma\sqrt{\frac{1}{n} + \frac{\Sigma X_i^2}{n\Sigma x_i^2}}}$$

$$= \frac{(\hat{\alpha} - \alpha)}{\sigma\sqrt{\frac{\Sigma X_i^2}{n\Sigma x_i^2}}}$$

$$Z^* = \frac{(\hat{\alpha} - \alpha) \cdot \sqrt{n} \cdot \sqrt{\Sigma x_i^2}}{\sigma\sqrt{\Sigma X_i^2}}$$

where $Z \sim N(0,1)$.

σ represents the variance of the unobservable ~~distributi~~ disturbances which is known. If we substitute unbiased estimator of σ^2 in the standard normal variable Z, the resulting variable $\left(\frac{Z\sqrt{V-2}}{V}\right)$ follow the t-distribution with $(n-2)$ degree of freedom.

In case of $\hat{\alpha}$, $Z = \frac{(\hat{\alpha} - \alpha)\sqrt{n\Sigma x_i^2}}{\sigma\sqrt{\Sigma X_i^2}}$, $V^2 = \frac{\Sigma e_i^2}{\sigma^2} = \frac{(n-2)\sigma^{*2}}{\sigma^2}$

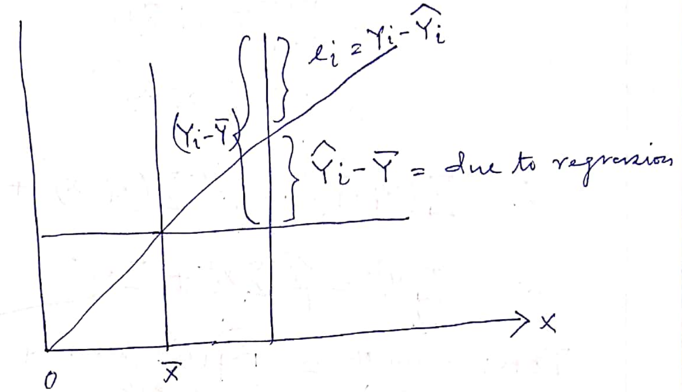
$$\therefore t = \frac{(\hat{\alpha} - \alpha)\sqrt{n\Sigma x_i^2}}{\sigma\sqrt{\Sigma X_i^2}} \cdot \sqrt{n-2} \cdot \frac{\sigma}{\sqrt{n-2} \cdot \sigma^*}$$

$$= \frac{(\hat{\alpha} - \alpha) \cdot \sqrt{n\Sigma x_i^2}}{\sigma^* \cdot \sqrt{\Sigma X_i^2}}$$

Thus the transforming the Z variable to t-variable the unknown true disturbance σ^2 has disappeared which depends solely on the sample observations and the [illegible]

GOODNESS OF FIT (R^2)

A measure of goodness of fit is the square of the correlation coefficient (R^2), which shows the percentage of the total variation of the dependent variable that can be explained by the independent variable (X).



The above diagram shows the breakdown of the variation of Y_i into two components: explained and unexplained. To compute R^2; we proceed as below:

$$e_i = y_i - \hat{y}$$

$$y_i = \hat{y} + e_i$$

~~For all values of y_i~~, $Y_i = \hat{Y}_i + e_i$

$$\text{or } Y_i - \bar{Y} = (\hat{Y}_i - \bar{Y}) + e_i$$

$$\text{or } \sum(Y_i - \bar{Y}) = \sum(\hat{Y}_i - \bar{Y}) + \sum e_i$$

$$\text{or } \sum(Y_i - \bar{Y})^2 = \sum(\hat{Y}_i - \bar{Y})^2 + 2\sum(\hat{Y}_i - \bar{Y})e_i + \sum e_i^2$$

But $\sum(\hat{Y}_i - \bar{Y})e_i$

$$= \sum[\hat{\alpha} + \hat{\beta}X_i - \bar{Y}]e_i$$

$$= \hat{\alpha}\sum e_i + \hat{\beta}\sum X_i e_i - \bar{Y}\sum e_i$$

$$= 0 \quad [\because \text{ Since } \sum e_i = 0 \text{ and } \sum X_i e_i = 0]$$

Again, $\sum(Y_i-\bar{Y})^2 = \sum(\hat{\alpha}+\hat{\beta}x_i-\bar{Y})^2$

$$= \sum[(\bar{Y}-\hat{\beta}x_i)+\hat{\beta}x_i-\bar{Y}]^2$$

since $\hat{\alpha} = \bar{Y}-\beta x_i$

$$= \sum[\hat{\beta}(x_i-\bar{x})]^2$$

$$= \hat{\beta}^2\sum(x_i-x)^2$$

$$\therefore \sum(Y_i-\bar{Y})^2 = \hat{\beta}^2\sum(x_i-x)^2 + \sum e_i^2$$

or $\sum y_i^2 = \hat{\beta}^2\sum x_i^2 + \sum e_i^2$

Total sum of squares (TSS) = Explain sum of squares (ESS) + Unexplained sum of squares. (RSS)

$\sum(Y_i-\bar{Y})^2$ represents the total sum of squared deviations from $\bar{Y}$, which we may as a measure of the total variations in Y which are required to be explained.

The decomposition of the total variations in Y leads to a measure of goodness of fit, it is also called the coefficient of determination which is given by:

$$R^2 = \frac{\text{Explained sum of squares (ESS)}}{\text{Total sum of squares (TSS)}}$$

$$= \frac{\hat{\beta}^2\sum x_i^2}{\sum y_i^2}$$

As $ESS = TSS - RSS$

$$R^2 = \frac{TSS-RSS}{TSS}$$

$$\therefore R^2 = \frac{\sum y_i^2 - \sum e_i^2}{\sum y_i^2} = 1 - \frac{\sum e_i^2}{\sum y_i^2}$$

Scanned by CamScanner

Properties of R^2:

(a) It is ~~an~~ a non-negative quantity $R^2 \geq 0$. It is calculated with the assumption that there is an intercept term in the regression equation of Y on X's.

(b). Its limit are $0 \leq R \leq 1$

When $R^2 = 0$, this means that there is no relationship between dependent variable and the explanatory variable.

When $R^2 = 1$, this means that there is a perfect fit.

(c) $R^2 = r^2$

Proof: From the defination, r can be written as—

$$r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} \cdot \sqrt{\sum y_i^2}} \quad \text{where} \quad x_i = X_i - \bar{X}, \quad y_i = Y_i - \bar{Y}.$$

$$\therefore R^2 = \hat{\beta}^2 \frac{\sum x_i^2}{\sum y_i^2} \quad \text{where } \hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$= \left[\left(\frac{\sum x_i y_i}{\sum x_i^2}\right)^2 \cdot \frac{\sum x_i^2}{\sum y_i^2}\right]$$

$$= \left[\frac{\sum x_i y_i}{\sum x_i^2 \cdot \sum y_i^2}\right]^2$$

$$= r^2$$

$\therefore$ The correlation coefficient $r = \pm\sqrt{R^2}$

While R^2 varies between 0 and 1, the correlation coefficient r varies between -1 and $+1$, indicating negative correlation and possitive linear correlation respectively, at the two extreme values.

Example-1: The following table gives data on weekly family consumption expenditure (Y) and weekly family income (X).

Y:	70	65	90	95	110	115	120	140	155	150
X:	80	100	120	140	160	180	200	220	240	260

i) Estimate the consumption function of the family

$$Y = \beta_0 + \beta_1 X + u.$$

ii) Test the significance of the parameter at 5 percent level of significance.

iii) Find R^2

iv) Write the economic interpretation of the estimated results.

Solⁿ: From the following table :- we get,

$\sum Y_i = 1110$, $\sum X_i = 1700$ $\sum X_i Y_i = 205500$

$\sum X_i^2 = 322000$ $\sum Y_i^2 = 132100$

$\therefore n = 10$, $\bar{Y} = \sum Y_i / n = 111$

$\bar{X} = \sum X_i / n = 170$

i) Estimation of $\hat{\alpha}$ and $\hat{\beta}$:

$$\therefore \hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$= \frac{\sum X_i Y_i - \bar{Y} \sum X_i}{\sum X_i^2 - \bar{X} \sum X_i}$$

$$= \frac{20550 - 111(1700)}{32000 - 170(1700)}$$

$$= 0.509$$

$$\therefore \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$= 111 - 0.509(170) = 24.47$$

The estimated Consumption function is-

$$\hat{Y} = 24.47 + 0.509X$$

Scanned by CamScanner

(ii) Estimation of variances:

$$\sum x_i^2 = \sum X_i^2 - \frac{(\sum X_i)^2}{n}$$

$$= 32200 - \frac{(1700)^2}{10} = 33000$$

$$\sum y_i^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n}$$

$$= 132100 - \frac{(1110)^2}{10} = 8890$$

$$\sum e_i^2 = \sum y_i^2 - \hat{\beta}_1^2 \sum x_i^2 \quad \text{or} \quad \sum y_i^2 - \hat{\beta} \sum x_i y_i$$

$$= 8890 - (0.509)^2 . 3300$$

$$= 340.33$$

variance of error term

$$\sigma_u^2 = \frac{\sum e_i^2}{n-2} = \frac{340.33}{10-2} = 42.54$$

$$Var(\hat{\beta}) = \frac{\sigma_u^2}{\sum x_i^2} = \frac{42.54}{33000} = 0.0013$$

$$S.E(\hat{\beta}_1) = \sqrt{0.0013} = 0.036$$

$$Var(\hat{\alpha}) = \frac{\sigma^2 \sum X_i^2}{n \sum x_i^2} = \frac{42.54(322000)}{10(33000)} = 41.51$$

$$S.E(\hat{\alpha}) = \sqrt{Var(\hat{\alpha})}$$

$$= \sqrt{41.51} = 6.44$$

iii) Hypothesis of Testing:

Since sample is small therefore we apply 't' test.

Suppose, $H_0 : \beta = 0$

$H_1 : \beta \neq 0$

$$t^* = \hat{\beta} / SE(\hat{\beta}) = \frac{0.509}{0.036} = 14.14$$

$$t_{0.025(n-2)} = t_{0.025, 8} = 2.31$$

Scanned by CamScanner

Since the value of 14.14 lies outside the acceptance region, the hypothesis of no relation between X and Y is to be rejected. Therefore, β_1 is statistically significant.

Similarly, $H_0: \alpha = 0$

$H_1: \alpha \neq 0$

$$t^{**} = \hat{\alpha} / S.E(\hat{\alpha})$$

$$= \frac{24.47}{6.44} = 3.799$$

Where $t_{0.025,(n-2)} = t_{0.025,8} = 2.31$

Since $t^{**} > t_{0.025,8}$, ~~the~~ reject the null hypothesis and we accept that α coefficient is also statistically significant.

iv) $$R^2 = \frac{\hat{\alpha} \sum x_i^2}{\sum y_i^2} = 1 - \frac{\sum e_i^2}{\sum y_i^2}$$

$$= \frac{(0.509)^2 \cdot 33000}{8890} = 0.9617$$

Hence, 96.17 present of the variations are explained by the variable X, Hence, it is good.

v) Reporting the result of regression anylsis:-

$$\hat{Y} = 24.47 + 0.509X$$

$$S.E = (6.44) \quad (0.036)$$

Interpretation of the estimated consumpition equation,
The estimated consumpition equation shows that the autonomus consumpition is Rs 24.47, ie, if income is zero, the average level of consumpition expenditure of the family is Rs. 24.47. The coeffieent of X is $(\hat{\beta}) = 0.509$ which measures the slope of the compution function. ~~which~~ This measures the marginal propensity to consume and which is less than 1.

(*) The least squares estimator of α in $Y = \alpha + \beta X + U$ is $\hat{\alpha} = \sum\left(\frac{1}{n} - \bar{X} w_i\right) Y_i$ where $w_i = x_i / \sum x_i^2$ with

$$var(\hat{\alpha}) = \sigma_u^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right]$$

Show that no other linear unbiased estimate of α can be constructed with smaller variance.

Ans:

given,

$$Var(\hat{\alpha}) = \sigma_u^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right]$$

$$= \sigma_u^2\left[\frac{\sum x_i^2 + n\bar{X}^2}{n\sum x_i^2}\right]$$

$$= \sigma_u^2\left[\frac{\sum X^2 - n\bar{X}^2 + n\bar{X}^2}{n\sum x_i^2}\right]$$

Since $\sum x_i^2 = \sum X^2 - n\bar{X}^2$

$$= \sigma_u^2\left[\frac{\sum X^2}{n\sum x_i^2}\right]$$

To show that no other linear unbiased estimator of α can be constructed with smaller variance,

(*) The following information is given that the percapita household expenditure of food (Y) and per capita household expenditure (X)

$\sum X^2 = 3737500$, $\sum X = \cancel{2800}$ 6750, $\sum Y = \cancel{6750}$ 2800, $\sum Y^2 = 584850$

$n = 15$, $\sum XY = 1463750$

given the information:-

$$\sum x_i^2 = \sum X_i^2 - \frac{(\sum X_i)^2}{n}$$

$$= 3737500 - \frac{(6750)^2}{15} = 70000$$

$$\sum x_i y_i = \sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}$$

$$= 1463750 - \frac{6750 \times 2800}{15}$$

$$= 203750$$

$$\sum y_1^2 = \sum Y_1^2 - \frac{(\sum Y_1)^2}{n}$$

$$= 584850 - \frac{(2800)^2}{15}$$

$$= 62183.34$$

$$\therefore \hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{203750}{700000} = 0.2910$$

We know that,

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$= 186.67 - 0.2910(450)$$

$$= 55.72$$

Hence, the estimated equation is –

$$\hat{Y} = 55.72 + 0.2910X$$

The marginal propensity to consume is 0.2910 and autonomous consumption is Rs 55.72.

Testing the parameter estimates,

$$TSS = \sum y_i^2 = 62183.34$$

$$ESS = \hat{\beta}^2 \sum x_i^2 = (0.2910)^2 \cdot 700000 = 59276.7$$

$$\therefore R^2 = \frac{\hat{\beta}^2 \sum x_i^2}{\sum y_i^2} = \frac{59276.7}{62183.34} = 0.9533$$

$$\therefore \sum e_i^2 = (1-R^2)\sum y_i^2$$

$$= (1-0.9533) \times 62183.34 = 2903.96$$

$$RSS = TSS - ESS$$

$$= 62183.34 - 59276.7$$

$$= 2906.3$$

$$\therefore \sigma_u^2 = \frac{\sum e_i^2}{n-2} = \frac{2906.3}{15-2} = 223.38$$

$$Var(\hat{\beta}) = \frac{\sigma_u^2}{\sum x_i^2}$$

$$= \frac{223.38}{700000}$$

$$= 0.00032$$

$$\therefore SE(\hat{\beta}) = \sqrt{0.00032}$$

$$= 0.01786.$$

~~Var(β~~ $$Var(\hat{\alpha}) = \frac{\sigma_u^2 \sum X_i^2}{n \sum x_i^2}$$

$$= \frac{223.38(3737500)}{15(700000)}$$

$$= 28.1961.$$

$$\therefore SE(\hat{\alpha}) = 5.31.$$

Hypothesis testing:

~~The~~ Null Hypothesis $H_0: \alpha = 0$

Alternative hypothesis $H_1: \alpha \neq 0$

$$\therefore t^*_{\hat{\alpha}} = \frac{\hat{\alpha}}{SE(\hat{\alpha})} = \frac{55.72}{5.31} = 10.4934$$

From table $t_{0.025} = 2.160$ with degree of freedom $(n-2) = 15-2 = 13$

Here $t^*_{\hat{\alpha}} > 2.160$

Here Rejected the null hypothesis.

Again Null hypothesis $H_0: \beta = 0$

Alternative hypothesis $H_1: \beta \neq 0$

$$\therefore t^*_{\hat{\beta}} = \frac{\hat{\beta}}{SE(\hat{\beta})} = \frac{0.2910}{0.01783} = 16.32$$

$\therefore t^*_{\hat{\beta}} > t_{0.025}$ with degree of freedom $(n-2) = 13$.

Here Rejected the null hypothesis —

Both the parameters are statistically significant.

Task:

1. Given $\sum X_i = 50$, $\sum X_i^2 = 304$, $\sum X_i Y_i = 353$, $\sum Y_i = 6$
$\sum Y_i^2 = 428$, $n = 10$
Estimate the parameters of the model: $Y_i = \alpha + \beta X_i + u_i$ and R^2. Test the hypothesis that $\alpha = 0$.

2. Given the following data: $\sum X_i Y_i = 200$, $\sum X_i^2 = 100$
$\sum Y_i^2 = 500$, $\sum X_i = 27$, $\sum Y_i = 40$ and $n = 27$
Estimate the parameters of the model and test the significance of the parameters.

3. The following table gives the number of T.V sets (Y) and its price.

No. of T.V (Y):	543	580	618	695	724	812	887	991	1186
~~sets (Y)~~ Price (X):	61	50	43	38	36	28	23	19	10

i) Estimate the demand function for T.V sets.
ii) Estimate the price elasticity of demand.
iii) Test the statistical significance of the parameter at 5% level of significance.

4. Prove that the relationship between R^2 and the slope of the regression line is given by $R^2 = \hat{\beta} \frac{\sum x_i y_i}{\sum y_i^2}$

5. Given that the following
$\sum X_i = 250$, $\sum Y_i = 300$, $n = 25$
$\sum x_i^2 = 350$, $\sum x_i y_i = 600$, $\sum e_i^2 = 120$
Where x_i and y_i are deviations from mean. Estimate the regression line $Y_i = \alpha + \beta X_i + u_i$. Also find 't' statistics for $\hat{\beta}$ and R^2.

6. The least squares estimate of α in $Y = \alpha + \beta X + U$ is

$$\hat{\alpha} = \sum\left(\frac{1}{n} - \bar{X} w_i\right) Y_i \quad \text{where } w_i = \frac{x_i}{\sum x_i^2} \text{ with}$$

$$\text{Variance}(\hat{\alpha}) = \sigma_u^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right]$$

Show that no other linear unbiased estimate of α can be constructed with ~~similar~~ smaller variance.

Proof:

given

$$Var(\hat{\alpha}) = \sigma_u^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right] \quad ?$$

$$= \sigma_u^2\left[\frac{\sum x_i^2 + n\bar{X}^2}{n\sum x_i^2}\right]$$

<u>Questions</u>

1. State the assumption of U_i.
2. State and prove the stastical properties of the estimators.
3. Derive R^2.
4. What are the reasons for the inclusion of the random variable in an econometric model.