Illustration 3: Public initiative in cultural
3.6 Global public goods and creative
3.7 Culture and development
4. International trade a
4.1 Globalisation
4.2 The efficie
4.3 Marke
Illustra
4.4

8.30
9.00 15 Friday January '99

9. Estimate the variance of u_i i.e, σ_u^2

Show that unbiased estimator of σ_u^2 is unknown and constant i.e, the variance of disturbance term is unknown

Let the equation be $Y_i = \alpha + \beta X_i + u_i$

Mean value of $Y_i = \bar{Y} = \alpha + \beta \bar{X} + \bar{u}$

Estimated value of $Y_i = \hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$

We know that, Error(e_i) = Actual value − Estimated value

$$= Y_i - \hat{Y}_i$$

or, $e_i = y_i - \hat{y}_i$

Now, $Y_i - \bar{Y} = \beta(X_i - \bar{X}) + (u_i - \bar{u})$

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

$$\therefore e_i = \beta x_i + (u_i - \bar{u}) - \hat{\beta} x_i \quad [\because \hat{y}_i = \hat{\beta} x_i]$$

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

$$\Sigma e_i^2 = \Sigma [-(\hat{\beta} - \beta) x_i + (u_i - \bar{u})]^2$$

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

$$E[\Sigma e_i^2] = E\left[(\hat{\beta} - \beta)^2 \Sigma x_i^2 + \Sigma (u_i - \bar{u})^2 - 2(\hat{\beta} - \beta) \Sigma x_i (u_i - \bar{u})\right] \longrightarrow (1)$$

Now,

i) $E[(\hat{\beta} - \beta)^2 \Sigma x_i^2] = \frac{\sigma_u^2}{\Sigma x_i^2} \cdot \Sigma x_i^2 = \sigma_u^2$

ii) $E[\Sigma (u_i - \bar{u})^2] = E[\Sigma u_i^2 + n\bar{u}^2 - 2\bar{u} \Sigma u_i]$

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

$$= E\left[\Sigma u_i^2 - \frac{(\Sigma u_i)^2}{n}\right],$$

$$= n\sigma_u^2 - \frac{n\sigma_u^2}{n} \quad [\because E(u_i^2) = \sigma_u^2 \text{ by assumption}]$$

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

(iii) $2E[(\hat{\beta}-\beta)\Sigma x_i(u_i-\bar{u})]$

9.30 $= 2E\left[\frac{\Sigma u_i x_i}{\Sigma x_i^2}(\Sigma x_i u_i - \bar{u}\Sigma x_i)\right]$

10.00 $\left[\because \hat{\beta} = \beta + \Sigma w_i u_i \right.$
$\left. \alpha,\ \hat{\beta}-\beta = \frac{\Sigma u_i x_i}{\Sigma x_i^2}\right]$

10.30 $= 2E\left[\frac{(\Sigma u_i x_i)^2}{\Sigma x_i^2} - \frac{\bar{u}\Sigma x_i \Sigma u_i x_i}{\Sigma x_i^2}\right]$

11.00

11.30 $= 2E\left[\frac{(\Sigma u_i x_i)^2}{\Sigma x_i^2}\right]$ $[\because \Sigma x_i = 0]$

12.00 $= 2 \cdot \frac{\sigma_u^2 \Sigma x_i^2}{\Sigma x_i^2}$ $[\because E(\Sigma u_i x_i)^2$
$= \sigma_u^2 x_1^2 + \sigma_u^2 x_2^2 + \dots + \sigma_u^2 x_n^2$
$= \sigma_u^2 \Sigma x_i^2$

12.30 $= 2\sigma_u^2$

1.00 Hence, substructing the value of (i), (ii) and (iii) we get,

1.30

2.00 $E[\Sigma e_i^2] = \sigma_u^2 + n\sigma_u^2 - \sigma_u^2 - 2\sigma_u^2$

17 Sunday $= (n-2)\sigma_u^2$

$\alpha,\ E\left[\frac{\Sigma e_i^2}{n-2}\right] = \sigma_u^2$

$E(\hat{\sigma}_u^2) = \sigma_u^2$

where, $\hat{\sigma}_u^2 = \frac{\Sigma e_i^2}{n-2}$

$\therefore \frac{\Sigma e_i^2}{n-2}$ is an unbiased estimator of σ_u^2.

n is the number of observation.

2 is the unknown parameter in this model.

Memo

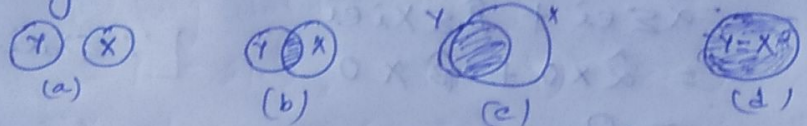
$\therefore S^2 = \frac{1}{n-2}\Sigma(y_i - \hat{\beta}x_i)^2 = \frac{1}{n-2}\Sigma(y_i - \hat{\beta}x_i)^2$

we know, $\hat{\beta} = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$ $\alpha,\ \Sigma x_i y_i = \hat{\beta}\Sigma x_i^2$ $= \frac{1}{n-2}\Sigma(y_i^2 + \hat{\beta}^2 x_i^2 - 2\hat{\beta}\, x_i y_i)$

$\therefore S^2 = \frac{1}{n-2}(\Sigma y_i^2 - \hat{\beta}^2 \Sigma x_i y_i)$

'Goodness of Fit'/ Coefficient of determination (R^2)

The coefficient of determination is a summary measure that tells how well the sample regression line fits the data.



In diagram, (Y) represent the variation of dependent variable and (X) is the variation of explanatory or independent variable. The overlap of the two circles indicates the extent to which the variation in Y is explained by the variation in X. The greater the extent of overlap, the greater the variation in Y is explained by X.

a) when there is no overlap, r^2 or $R^2 = 0$

In (b) and (c) as overlap increases explained variation of Y also increases.

(d) when the overlap is complete r^2 or $R^2 = 1$, means 100% of the variation of Y is explained by X.

The formula for, $R^2 = \frac{\text{Explained variation}}{\text{Total variation}}$.

Compute the value of R^2

If $Y_i = \alpha + \beta X_i + e_i$

and $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$

$e_i = Y_i - \hat{Y}_i$

or, $Y_i = \hat{Y}_i + e_i$

or, $Y_i - \bar{Y} = \hat{Y}_i - \bar{Y} + e_i$ (taking both sides by $-\bar{Y}$)

or, $(Y_i - \bar{Y})^2 = \{(\hat{Y}_i - \bar{Y}) + e_i\}^2$ (both sides take perfect square)

or, $y_i^2 = \hat{y}_i^2 + e_i^2 + 2\hat{y}_i e_i$

or, $\Sigma y_i^2 = \Sigma \hat{y}_i^2 + \Sigma e_i^2 + 2\Sigma \hat{y}_i e_i$ (taking both sides by Σ)

Now $\sum \hat{y}_i e_i$

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

$= \sum \hat{Y}_i e_i - \bar{Y} \sum e_i$ $\quad [\because \sum e_i = 0]$

$= \sum (\hat{\alpha} + \hat{\beta} X_i) e_i - 0$

$= \hat{\alpha} \sum e_i + \hat{\beta} \sum X_i e_i$ $\quad [\because \sum X_i e_i = 0]$

$= \hat{\alpha} \times 0 + \hat{\beta} \times 0$

$= 0$

and $\sum \hat{y}_i^2 = \sum (\hat{Y}_i - \bar{\hat{Y}})^2$

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

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

$= \hat{\beta}^2 \sum x_i^2 \Rightarrow$ Explained variation

and $\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 \Rightarrow$ Unexplaind variation

and $\sum y_i^2 = \sum (Y_i - \bar{Y})^2 \Rightarrow$ Total variation.

Therefore we can write that

$\sum y_i^2 = \sum \hat{y}_i^2 + \sum e_i^2$

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

∴ Total variation = Explained variation + Unexplained variation

or, Total Sum of Squares (TSS) = Explained Sum Square (ESS) + Unexplained Sum Squares / Residual Sum Squares (RSS)

or, $\text{Var}(Y_i) = \text{Var}(\hat{Y}_i) + \text{Var}(e_i)$

∴ $TSS = ESS + RSS$

i) TSS Respresents the Total Sum of Squared deviations from $\bar{Y}$.

ii) $ESS = \hat{\beta}^2 \sum x_i^2$ represents the estimated effect of X on the variations in Y.

iii) $RSS = \sum e_i^2$ represents the variations in Y which unexplained by the estimated relationship.

This decomposition of total variations in Y leads to a measure of the "goodness of fit"/(R^2) coefficient of determination.

where, $R^2 = \frac{\text{Explained variation}}{\text{Total variation}} = \frac{\text{Var}(\hat{Y}_i)}{\text{Var}(Y_i)}$

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

Prove that R^2 lies between +1 and −1

Since, $\text{Var}(Y_i) = \text{Var}(\hat{Y}_i) + \text{Var}(e_i)$

and $0 \leq \text{Var}(\hat{Y}_i) \leq \text{Var}(Y_i)$

or, $0 \leq \frac{\text{Var}(\hat{Y}_i)}{\text{Var}(Y_i)} \leq 1$

or, $0 \leq R^2 \leq 1$.

$R^2 = 0$ when $\text{var}(\hat{Y}_i) = 0$ i.e, $\Sigma e_i^2 = \Sigma y_i^2$

and $R^2 = 1$ when $\text{Var}(\hat{Y}_i) = \text{Var}(Y_i)$ i.e., $\Sigma e_i^2 = 0$

Econometrics.

H.W.

1. A sample of 20 observations corresponding to regression model $Y_i = \alpha + \beta X_i + u_i$, gives the following data:

$$\Sigma Y_i = 21.9, \quad \Sigma (Y_i - \bar{Y})^2 = 86.9$$

$$\Sigma X_i = 186.2, \quad \Sigma (X_i - \bar{X})^2 = 215.4, \quad \Sigma (X_i - \bar{X})(Y_i - \bar{Y})$$

i) Obtain the estimated value of $\hat{\alpha}$ & $\hat{\beta}$.

ii) Estimate $\text{Var}(\hat{\beta})$ & $\text{Var}(\hat{\alpha})$.

iii) Find the value of R^2

$\therefore R^2 = 0 \cdot 935$.

This suggests that 93·5 per cent of the sample observations of Y can be attributed to the variations of the fitted value of Y i.e., $\hat{Y}_i$ or we say that our regression line fits the given data well.

Thus R^2 measures the proportion of variations in the dependent variable that is explained by the independent variables.

Example : 2.5. A sample of 20 observations corresponding to the regression model $Y_i = \alpha + \beta X_i + u_i$ where u_i is normally distributed with mean zero and unknown variance σ_u^2, gives the following data :

$$\sum_{i=1}^{n} Y_i = 21 \cdot 9, \quad \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = 86 \cdot 9, \quad \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}) = 106 \cdot 4$$

$$\sum_{i=1}^{n} X_i = 186 \cdot 2, \quad \sum_{i=1}^{n} (X_i - \overline{X})^2 = 215 \cdot 4, \quad n = 20$$

Obtain the usual regression results.

Solution : On the basis of the given information we have to fit a linear relation between Y (dependent variable) and X (explanatory variable).

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

We know that, $$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n} (X_i - \overline{X})^2}$$

$$\therefore \hat{\beta} = \frac{106 \cdot 4}{215 \cdot 4} = 0 \cdot 494$$

Where $\overline{Y} = \dfrac{\sum_{i=1}^{n} Y_i}{n} = \dfrac{21 \cdot 9}{20} = 1 \cdot 095$

and $\hat{\alpha} = \overline{Y} - \hat{\beta}\,\overline{X}$

$= 1 \cdot 095 - 0 \cdot 494 \times 9 \cdot 31$

$= 1 \cdot 095 - 4 \cdot 60 = -3 \cdot 505$

and $\overline{X} = \sum_{i=1}^{n} \dfrac{X_i}{n} = \dfrac{186 \cdot 2}{20} = 9 \cdot 31$

Thus we have, $\hat{\alpha} = -3 \cdot 505$ and $\hat{\beta} = 0 \cdot 494$ and our estimated regression line is :

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i \Rightarrow \hat{Y}_i = -3 \cdot 505 + 0 \cdot 494 X_i$$

BLUE'S

(ii) **Estimation of variances :**

Since we know that, $\text{var}(\hat{\alpha}) = \sigma_u^2 \left(\frac{\sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n} x_i^2} \right)$ and $\text{var}(\hat{\beta}) = \frac{\hat{\sigma}_u^2}{\sum_{i=1}^{n} x_i^2}$.

Here we see that σ_u^2 is not known and hence we replace it by its unbiased estimator $\hat{\sigma}_u^2 = \sum_{i=1}^{n} e_i^2 / n-2$.

Thus we have, $\text{var}(\hat{\alpha}) = \hat{\sigma}_u^2 \left(\frac{\sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n} x_i^2} \right)$ and $\text{var}(\hat{\beta}) = \frac{\hat{\sigma}_u^2}{\sum_{i=1}^{n} x_i^2}$

Again we know that, $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} y_i^2 - \hat{\beta} \sum_{i=1}^{n} x_i^2$

$\therefore \sum_{i=1}^{n} e_i^2 = 86 \cdot 9 - (0 \cdot 494)^2 \times 215 \cdot 4$

$= 86 \cdot 9 - 52 \cdot 56 = 34 \cdot 34$

Now $\hat{\sigma}_u^2 = \sum_{i=1}^{n} e_i^2 \Big/ n-2$

$= \frac{34 \cdot 34}{20-2} = \frac{34 \cdot 34}{18} = 1 \cdot 908$

Where $\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$

and $\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2$

Now $\text{var}(\hat{\alpha}) = \hat{\sigma}_u^2 \left(\frac{\sum_{i=1}^{n} X_i^2}{n. \sum_{i=1}^{n} x_i^2} \right) = \frac{1 \cdot 908 \times 1948 \cdot 922}{20 \times 215 \cdot 4} = 0 \cdot 8631$

$\left[\therefore \sum_{i=1}^{n} (X_i - \bar{X})^2 = 215 \cdot 4 \text{ or, } \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = 215 \cdot 4 \right.$

Thus we conclude that OL

$$\text{or, } \sum_{i=1}^{n} X_i^2 = 215 \cdot 4 + n\bar{X}^2 = 215{\cdot}4 + 20 \times (9{\cdot}3)^2$$

$$= 215{\cdot}4 + 1733{\cdot}522$$

$$= 1948{\cdot}922 \ ]$$

$$\therefore \text{var}(\hat{\alpha}) = 0 \cdot 8631$$

$$\text{Similarly, } \text{var}(\hat{\beta}) = \frac{\hat{\sigma}_u^2}{\sum_{i=1}^{n} x_i^2} = \frac{1 \cdot 908}{215 \cdot 4} = 0 \cdot 0089$$

$$\text{Now, } SE(\hat{\alpha}) = \sqrt{\text{var}(\hat{\alpha})} = \sqrt{0 \cdot 8631} = 0 \cdot 929$$

$$SE(\hat{\beta}) = \sqrt{\text{var}(\hat{\beta})} = \sqrt{0 \cdot 0089} = 0 \cdot 094 .$$

(iii) **Construction of confidence intervals**

Now we like to set up a confidence interval for
= 0·95 (i.e., 5% level of significance) and (b) P = 0
of significance)

In other words, we like to find the value of 't'
0·025 and (b) 0·005 of the

In terms of our earlier example (Example 2.5) the estimate regression results can be written as :

$$Y_i = -3{\cdot}505 + 0{\cdot}494X_i \quad R^2 = 0{\cdot}6048.$$
$$(0{\cdot}929) \quad (0{\cdot}094)$$

Here $\hat{\alpha} = -3{\cdot}505$, $\hat{\beta} = 0{\cdot}494$, $SE(\hat{\alpha}) = 0{\cdot}929$, $SE(\hat{\beta}) = 0{\cdot}094$,

$$\text{and } R^2 = \frac{\hat{\beta}^2 \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} y_i^2} = \frac{(0{\cdot}494)^2 \times 215{\cdot}4}{86{\cdot}9} = \frac{52{\cdot}5653}{86{\cdot}9} = 0{\cdot}6048$$

This suggests that 60·48 per cent of the sample observations c can be attributed to the variations of the fitted value of Y or we that our regression line fits the given data moderately (not very w

Some econometricians report the t-ratios of the estima coefficients in place of standard errors. This way of presenta makes the testing of hypothesis easier and direct.

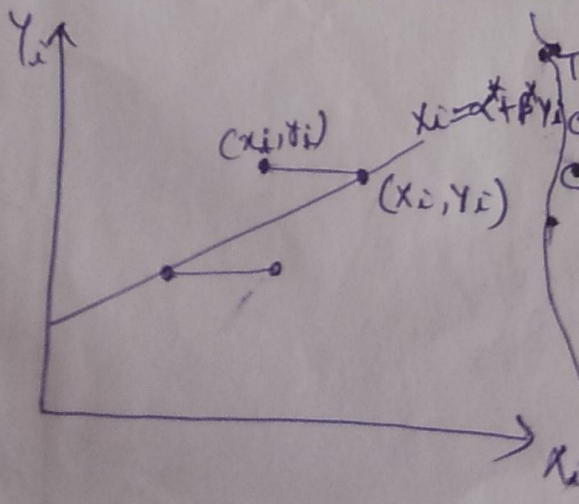
Thus the other form of presentation of

Reverse Regression

In some cases where the usual regression provides biased estimates and in its place suggested an alternative procedure for obtaining an unbiased estimate of the parameters on the basis of related econometric research (Vanhonacker and Day, 1987). This estimation procedure is referred to as "reverse regression".

In this method the role of endogenous and exogenous variables are reversed, i.e, exogenous variable are regressed on endogenous variables. The method of least squares estimates the parameters α and β by minimizing the sum of squares of difference between observations and the line in the scatter diagram.

The reverse (or inverse) regression approach minimizes the sum of squares of horizontal distance between the observed data points and the line in the following scatter diagram to obtain the estimates of regression parameters



The regression equation in case of reverse regression can be written as

$$x_i = \alpha^* + \beta^* y_i + \varepsilon_i$$

$(i = 1, 2, \cdots, n)$

The regression equation in case of reverse regression can be written as

$$x_i = \alpha^* + \beta^* y_i + \delta_i \quad (i = 1, 2, \cdots, n)$$

where δ_i's are the associated random error components and satisfy the assumptions as in the case of usual simple regression model.

The reverse regression estimates $\hat{\alpha}^*$ of α^* and $\hat{\beta}^*$ of β^* for the model are obtained by interchanging the x and y in the direct regression estimators of α and β. The estimates are obtained as

$$\hat{\beta}^* = \frac{\sum y_i x_i}{\sum y_i^2} = \frac{s_{xy}}{s_{yy}}$$

and $\hat{\alpha}^* = \bar{x} - \hat{\beta}^* \bar{y}$

The residual sum of squares in this case is

$$RSS^* = s_{xx} - \frac{s_{xy}^2}{s_{yy}}$$

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

Coefficient of determination,

and $R^2 = \hat{\beta}^* \cdot \beta$

where β is the direct regression estimator of slope parameter.