EXERCISES

1. In a simple linear regression model $Y_i = \alpha + \beta X_i + u_i$ for $i = 1, 2, ..., n$ why we insert the random disturbance term 'u'?

2. State and explain the assumption of a classical linear regression model (CLRM).

3. In a simple linear regression model of the form $Y_i = \alpha + \beta X_i + u_i$ $i = 1,2,....,n$ how can you estimate the regression coefficients α and β ?

4. Describe briefly the method of moments, used in estimating the regression parameters in a two variable linear regression model.

5. Describe briefly the method of least squares used in estimating the regression parameters relating to a two variable linear regression model.

6. How can you estimate a function (two variable) whose intercept is zero ?

7. How can you estimate the elasticities from an estimated regression line ?

8. State and prove the properties of the least squares estimators relating to a two variable linear regression model (CLRM).

9. Show that in a classical linear regression model the estimated regression coefficients are unbiased.

10. Determine the mean and variance of $\hat{\alpha}$ and $\hat{\beta}$ relating to a model $Y_i = \alpha + \beta X_i + u_i$ for $i = 1, 2, ..., n$.

11. State and prove that Gauss-Markov theorem, with reference to a CLRM.

12. What is meant by the term BLUE ? Show that (i) $\hat{\alpha}$ is the BLUE of α and (ii) $\hat{\beta}$ is the BLUE of β in the CLRM.

13. How can you determine the variance of the random distrubance term in the model $Y = \alpha + \beta X + u$.

14. Show that $\sum_{i=1}^{n} e_i^2 \Big/ n-2$ is the unbiased estimator of the variance of the random disturbance term i.e., σ_u^2.

15. What is a maximum likelihood estimator (MLE) ? Show that in the model $Y_i = \alpha + \beta X_i + u_i$ (i) $\hat{\alpha}$ is the MLE of α.

(ii) $\hat{\beta}$ is the MLE of β.

(iii) $\sum_{i=1}^{n} e_i^2 \Big/ n$ is the MLE of σ_u^2.

16. Show that the least squares estimates $\hat{\alpha}$ and $\hat{\beta}$ are such that :

$$\text{(i) } \hat{\alpha} \sim N\left[\alpha,\ \sigma_u^2\left(\frac{1}{n}+\frac{\overline{X}^2}{\sum_{i=1}^{n} x_i^2}\right)\right] \quad \text{(ii) } \hat{\beta} \sim N\left[\beta,\ \frac{\sigma_u^2}{\sum_{i=1}^{n} x_i^2}\right]$$

17. Describe the testing procedure of the significance of the regression coefficients of the model $Y_i = \alpha + \beta X_i + u_i$ for i = 1, 2, ..., n.

18. What is meant by goodness of fit of the correlation coefficient R^2?

19. Show that Total sum of squares = Explained sum of squares + unexplained sum of squares.

20. What is coefficient of determination ? Show that it lies between 0 and 1 and hence show that the value of correlation coefficient between two variables lies between − 1 and +1.

21. How can you formally write the regression results of the regression model $Y_i = \alpha + \beta X_i + u_i$ where u_i (for i = 1, 2, ...,n) satisfies all the properties of CLRM ?

22. How can you use the analysis of variance in the simple classical linear regression model ?

23. What is the meaning of the term 'Prediction'? How can you incorporate the term in the CLRM ? Distinguish between point prediction and interval prediction in this regard.

24. Show that the OLS point predictor in the CLRM satisfies the BLUE property.

25. The following sums where obtained from 16 pairs of observations on X and Y : $\sum Y_i^2 = 526$, $\sum X_i^2 = 657$, $\sum X_i Y_i = 492$, $\sum Y_i = 63$, $\sum X_i = 96$, Estimate the parameters in the model : $Y_i = \alpha + \beta X_i + u_i$ and R^2.

Test the hypothesis that β = 2·0.

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

$$\sum Y_i = 21\cdot 9,\ \sum(Y_i - \overline{Y})^2 = 86\cdot 9,\ \sum(X_i - \overline{X})(Y_i - \overline{Y}) = 106\cdot 4,$$

$$\sum X_i = 186\cdot 2,\ \sum(X_i - \overline{X})^2 = 215\cdot 4.$$

Estimate α and β and calculate estimates of variances of you estimates. Estimate the conditional mean value of Y corresponding to value of X fixed at X = 20.

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

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

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 α and β at (a) P = 0·95 (i.e., 5% level of significance) and (b) P = 0·99 (i.e., 1% level of significance)

In other words, we like to find the value of '*t*' that cuts off (a) 0·025 and (b) 0·005 of the area at the tail end of the distribution on both sides. From table value : $t_{0 \cdot 025}$, $n - 2 = t_{0 \cdot 025}$, $18 = 2 \cdot 101$ and $t_{0 \cdot 005}$, $n - 2 = t_{0 \cdot 005}$, $18 = 2 \cdot 878$

Therefore 95% confidence internal for α are : $\hat{\alpha} \pm t_{0 \cdot 025}, n-2. SE(\hat{\alpha})$

i.e., $P[\hat{\alpha} - t_{0 \cdot 025}, n-2\ SE(\hat{\alpha}) \leq \alpha \leq \hat{\alpha} + t_{0 \cdot 025}, n-2\ SE(\hat{\alpha})$ = 0·95 and 99% confidence interval for α are : $\hat{\alpha} \pm t_{0 \cdot 0005}, n-2\ SE(\hat{\alpha})$.

i.e., $P[\hat{\alpha} - t_{0 \cdot 005}, n-2\ SE(\hat{\alpha}) \leq \alpha \leq \hat{\alpha} + t_{0 \cdot 005}, n-2\ SE(\hat{\alpha})$ = 0·99

Therefore 95% confidence interval for α would be : $\hat{\alpha} \pm t_{0 \cdot 025}$, $n-2\ SE(\hat{\alpha})$.

$\Rightarrow -3 \cdot 505 \pm 2 \cdot 101 \times 0 \cdot 929$

or, $-3 \cdot 505 \pm 1 \cdot 9518$.

Similarly, 99% confidence interval for α would be :

$-3 \cdot 505 \pm 2 \cdot 878 \times 0 \cdot 929$

or, $-3 \cdot 505 \pm 2 \cdot 6736$.

Similarly, 95% confidence interval of β are : $\hat{\beta} \pm t_{0 \cdot 025}, n-2\ SE$

i.e., $P\left[\hat{\beta} - t_{0 \cdot 025}, n-2\ SE(\hat{\beta}) \leq \beta \leq \hat{\beta} + t_{0 \cdot 025}, n-2\ SE(\hat{\beta})\right] = 0 \cdot 95.$

and 99% confidence interval for β are : $\hat{\beta} \pm t_{0.005},\ n-2\ SE(\hat{\beta})$

i.e., $P\left[\hat{\beta} - t_{0.005},\ n-2\ SE(\hat{\beta}) \leq \beta \leq \hat{\beta} + t_{0.005},\ n-2\ SE(\hat{\beta})\right] = 0{\cdot}99$

Thus 95% confidence internal for β would be : $\hat{\beta} \pm t_{0.025}, n-2\ SE(\hat{\beta})$

or, $0{\cdot}494 \pm 2{\cdot}101 \times 0{\cdot}094$

or, $0{\cdot}494 \pm 0{\cdot}1974$

Where $\hat{\beta} = 0.494$

$t_{0.025}, n-2 = t_{0.025}, 18 = 2.101$

$SE(\hat{\beta}) = 0.094$.

(iv) **Hypothesis testing :** Suppose we like to test $H_0 : \beta = 0$ against the alternative $H_1 : \beta \neq 0$. Now on the basis of the given sample $H_0 : \beta = 0$ will be rejected at 5% lavel of significance if

$$|t_{n-2}| = \left|\frac{\hat{\beta}}{SE(\hat{\beta})} \text{ (observed)}\right| > t_{0.025},\ n-2 \text{ (table value)}$$

and will be accepted otherwise.

$$\text{Here } t_{n-2} = \frac{\hat{\beta}}{SE(\hat{\beta})} = \frac{0{\cdot}494}{0{\cdot}094} = 5{\cdot}255 \text{ (where } n = 20).$$

Thus we see that, $|t_{n-2}| = 5{\cdot}255 > t_{0.025,\ 18} (= 2{\cdot}101)$ and hence $H_0 : \beta = 0$ is rejected (alternative $H_1 : \beta \neq 0$ is accepted) at 5% level of significance. So, the hypothesis of no relationship between X and Y is to be rejected at 5% level of significance. Similarly, it can be tested for 1% level of significance.

2.15. Results of Regression Analysis

The results of regression analysis are generally presented in a conventional format. It is not sufficient merely to report the estimates of α and β. In practice we report regression coefficients together with their standard errors and the value of R^2 ...

By the way, note that the r^2 value given for the regression-through-the-origin model should be taken with a grain of salt, for the traditional formula of r^2 is not applicable for such models. *EViews*, however, routinely presents the standard r^2 value even for such models.

6.2 Scaling and Units of Measurement

To grasp the ideas developed in this section, consider the data given in Table 6.2, which refers to Indian gross domestic savings (GDS) and gross domestic product (GDP), in rupee crore as well as in rupees lakh crore measured in 1999–2000 prices.

Suppose in the regression of GDS on GDP one researcher uses data in rupee crore but another expresses data in rupee lakh crore. Will the regression results be the same in both cases? If not, which results should one use? In short, do the units in which the regressand and regressor(s) are measured make any difference in the regression results? If so, what is the sensible course to follow in choosing units of measurement for regression analysis? To answer these questions, let us proceed systematically. Let

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \tag{6.2.1}$$

where Y = GDS and X = GDP. Define

$$Y_i^* = w_1 Y_i \tag{6.2.2}$$

$$X_i^* = w_2 X_i \tag{6.2.3}$$

where w_1 and w_2 are constants, called the **scale factors;** w_1 may equal w_2 or be different.

From Equations 6.2.2 and 6.2.3 it is clear that Y_i^* and X_i^* are *rescaled* Y_i and X_i. Thus, if Y_i and X_i are measured in rupee crore and one wants to express them in rupee lakh crore, we will have $Y_i^* = 100Y_i$ and $X_i^* = 100\,X_i$; here $w_1 = w_2 = 100$.

Now consider the regression using Y_i^* and X_i^* variables:

$$Y_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^* \tag{6.2.4}$$

where $Y_i^* = w_1 Y_i$, $X_i^* = w_2 X_i$, and $\hat{u}_i^* = w_1 \hat{u}_i$. (Why?)

Table 6.2 Gross domestic savings and GDP for India, 1951–52 to 2004–05, both at 1999–2000 prices

Year	GDS (in Rs. Crore)	GDP (in Rs. Crore)	GDS (in Rs. Lakh)	GDP (in Rs. Lakh)
1951–52	969	230,034	96,900	23,003,400
1952–53	845	236,562	84,500	23,656,200
1953–54	875	250,960	87,500	25,096,000
1954–55	988	261,615	98,800	26,161,500
1955–56	1,356	268,316	135,600	26,831,600
1956–57	1,561	283,589	156,100	28,358,900
1957–58	1,356	280,160	135,600	28,016,000
1958–59	1,379	301,422	137,900	30,142,200
1959–60	1,720	308,018	172,000	30,801,800
1960–61	1,952	329,825	195,200	32,982,500
1961–62	2,074	340,060	207,400	34,006,000
1962–63	2,440	347,253	244,000	34,725,300

(Contd.)

We want to find out the relationships between the following pairs:

1. $\hat{\beta}_1$ and $\hat{\beta}_1^*$
2. $\hat{\beta}_2$ and $\hat{\beta}_2^*$
3. $\text{var}(\hat{\beta}_1)$ and $\text{var}(\hat{\beta}_1^*)$
4. $\text{var}(\hat{\beta}_2)$ and $\text{var}(\hat{\beta}_2^*)$
5. $\hat{\sigma}^2$ and $\hat{\sigma}^{*2}$
6. r_{xy}^2 and $r_{x^*y^*}^2$

From least-squares theory we know (see Chapter 3) that

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \tag{6.2.5}$$

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \tag{6.2.6}$$

$$\text{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2} \cdot \sigma^2 \tag{6.2.7}$$

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{6.2.8}$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \tag{6.2.9}$$

Applying the OLS method to Equation 6.2.4, we obtain similarly

$$\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{X}^* \tag{6.2.10}$$

$$\hat{\beta}_2^* = \frac{\sum x_i^* y_i^*}{\sum x_i^{*2}} \tag{6.2.11}$$

$$\text{var}(\hat{\beta}_1^*) = \frac{\sum X_i^{*2}}{n \sum x_i^{*2}} \cdot \sigma^{*2} \tag{6.2.12}$$

$$\text{var}(\hat{\beta}_2^*) = \frac{\sigma^{*2}}{\sum x_i^{*2}} \tag{6.2.13}$$

$$\hat{\sigma}^{*2} = \frac{\sum \hat{u}_i^{*2}}{(n-2)} \tag{6.2.14}$$

From these results it is easy to establish relationships between the two sets of parameter estimates. All that one has to do is recall these definitional relationships: $Y_i^* = w_1 Y_i$ (or $y_i^* = w_1 y_i$); $X_i^* = w_2 X_i$ (or $x_i^* = w_2 x_i$); $\hat{u}_i^* = w_1 \hat{u}_i$; $\bar{Y}^* = w_1 \bar{Y}$; and $\bar{X}^* = w_2 \bar{X}$. Making use of these definitions, the reader can easily verify that

$$\hat{\beta}_2^* = \left(\frac{w_1}{w_2}\right) \hat{\beta}_2 \tag{6.2.15}$$

$$\hat{\beta}_1^* = w_1 \hat{\beta}_1 \tag{6.2.16}$$

$$\hat{\sigma}^{*2} = w_1^2 \hat{\sigma}^2 \tag{6.2.17}$$

$$\text{var}(\hat{\beta}_1^*) = w_1^2 \, \text{var}(\hat{\beta}_1) \tag{6.2.18}$$

$$\operatorname{var}(\hat{\beta}_2^*) = \left(\frac{w_1}{w_2}\right)^2 \operatorname{var}(\hat{\beta}_2) \tag{6.2.19}$$

$$r_{xy}^2 = r_{x^*y^*}^2 \tag{6.2.20}$$

From the preceding results it should be clear that, given the regression results based on one scale of measurement, one can derive the results based on another scale of measurement once the scaling factors, the w's, are known. In practice, though, one should choose the units of measurement sensibly; there is little point in carrying all those zeros in expressing numbers in lakh or crores of rupees.

From the results given in (6.2.15) through (6.2.20) one can easily derive some special cases. For instance, if $w_1 = w_2$, that is, the scaling factors are identical, the slope coefficient and its standard error remain unaffected in going from the (Y_i, X_i) to the (Y_i^*, X_i^*) scale, which should be intuitively clear. However, the intercept and its standard error are both multiplied by w_1. But if the X scale is not changed (i.e., $w_2 = 1$) and the Y scale is changed by the factor w_1, the slope as well as the intercept coefficients and their respective standard errors are all multiplied by the same w_1 factor. Finally, if the Y scale remains unchanged (i.e., $w_1 = 1$) but the X scale is changed by the factor w_2, the slope coefficient and its standard error are multiplied by the factor $(1/w_2)$ but the intercept coefficient and its standard error remain unaffected.

It should, however, be noted that the transformation from the (Y, X) to the (Y^*, X^*) scale does not affect the properties of the OLS estimators discussed in the preceding chapters.

Example 6.2 The relationship between GDS and GDP, India, 1951–52 to 2004–05

To substantiate the preceding theoretical results, let us return to the data given in Table 6.2 and examine the following results (numbers in parentheses are the estimated standard errors).

Both GDS and GDP in rupee crore:

$$\begin{aligned}\widehat{GDS_t} &= 167423.37 + 0.36\ GDP_t \\ se &= (17721.01)\quad (0.02) \qquad r^2 = 0.8891\end{aligned} \tag{6.2.21}$$

Both GDS and GDP in rupee lakh:

$$\begin{aligned}\widehat{GDS_t} &= -16742336.51 + 0.36\ GDP_t \\ se &= (1772100.74)\quad (0.02) \qquad r^2 = 0.8891\end{aligned} \tag{6.2.22}$$

Notice that the intercept and its standard error is 100 times the corresponding values in the regression (6.2.21) (note that $w_1 = 100$ is going from crore to lakhs of rupees), but the slope coefficient as well as its standard error is unchanged, in accordance with the theory.

GDS in rupee crore and GDP in rupee lakh:

$$\begin{aligned}\widehat{GDS_t} &= -167423.37 + 0.0036\ GDP_t \\ se &= (17721.01)\quad (0.0002) \qquad r^2 = 0.8891\end{aligned} \tag{6.2.23}$$

As expected, the slope coefficient as well as its standard error is 1/100 its value in Eq. (6.2.12), since only X, or GDP, scale is changed.

GDS in rupee lakh and GDP in rupee crore:

$$\begin{aligned}\widehat{GDS_t} &= -16742336.51 + 36.33\ GDP_t \\ se &= (1772100.74)\quad (1.78) \qquad r^2 = 0.8891\end{aligned} \tag{6.2.24}$$

Again note that both the intercept and the slope coefficients as well as their respective standard errors are 100 times their values in Eq. (6.2.21), in accordance with our theoretical results.

Notice that in all the regressions presented above, the r^2 value remains the same, which is not surprising because the r^2 value is *invariant* to changes in the unit of measurement as it is pure, or dimensionless, number

(i) Larger the value of σ_u^2, the larger the variances of $\hat{\alpha}$ and $\hat{\beta}$. In other words, the greater the dispersion of the disturbance terms around the population regression line, the greater the dispersion in the values of estimated regression parameters.

(ii) $\sum_{i=1}^{n} x_i^2$ is the denominator of the variance formula of both estimators. This indicates that the more dispersed the values of the explanatory variables (i.e., larger $\sum_{i=1}^{n} x_i^2$), the smaller the variances of $\hat{\alpha}$ and $\hat{\beta}$. If $\sum_{i=1}^{n} x_i^2 = 0$ or nearer to zero ; i.e., when $X_1 = X_2 = \ldots = X_n$ both variances would be infinitely large.

(iii) The variance of $\hat{\alpha}$ is smallest when $\overline{X} = 0$ or near to zero. In particular, when $\overline{X} = 0$, $\text{var}(\hat{\alpha}) = \frac{\sigma_u^2}{n}$.

2.13. Confidence Intervals and Hypothesis Testing

It is highly essential to construct confidence intervals of the parameters in order to achieve precision of $\hat{\alpha}$ and $\hat{\beta}$. We have all the information concerning the distribution of $\hat{\alpha}$ and $\hat{\beta}$ in order to standardise them.

$$\text{Since } \hat{\alpha} \sim N\left[\alpha, \sigma_u^2\left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} x_i^2}\right)\right]$$

$$\text{and } \hat{\beta} \sim N\left(\beta, \frac{\sigma_u^2}{\sum_{i=1}^{n} x_i^2}\right)$$

$$\text{Now } \tau \text{ or } Z, = \frac{\hat{\beta} - E(\hat{\beta})}{SE(\hat{\beta})} = \frac{\hat{\beta} - \beta}{\sigma_u \sqrt{\frac{1}{\sum_{i=1}^{n} x_i^2}}}$$

$\sim N\ (0,\ 1)$, where $SE(\hat{\beta}) = \sqrt{\text{var}(\hat{\beta})}$

$$\text{and } \tau \text{ or } Z = \frac{\hat{\alpha} - E(\hat{\alpha})}{SE(\hat{\alpha})} = \frac{\hat{\alpha} - \alpha}{\sigma_u \sqrt{\dfrac{\sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n} x_i^2}}} \sim N(0,\ 1).$$

$$\text{[where } \text{var}(\hat{\alpha}) = \sigma^2 u \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} x_i^2} \right)$$

$$\therefore SE(\hat{\alpha}) = \sqrt{\text{var}(\hat{\alpha})} = \sigma_u \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} x_i^2} \right)$$

$$= \sigma_u \left(\sum_{i=1}^{n} X_i^2 \Big/ n \sum_{i=1}^{n} x_i^2 \right)$$

$$\therefore \frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} x_i^2 + n\overline{X}^2}{n \sum_{i=1}^{n} x_i^2}$$

$$= \frac{\left[\sum_{i=1}^{n} (Xi - \overline{X})^2 \right] + n\overline{X}^2}{n \sum_{i=1}^{n} x_i^2} = \frac{\left[\sum_{i=1}^{n} X_i^2 - 2\overline{X} \sum_{i=1}^{n} X_i + \overline{X}^2 \right] + n\overline{X}^2}{n \sum_{i=1}^{n} x_i^2}$$

$$= \frac{\sum_{i=1}^{n} X_i^2 - 2n\overline{X}^2 + 2n\overline{X}^2}{n \sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n} x_i^2} \Bigg]$$

Therefore 95% confidence limits for α are :

$$\hat{\alpha} \pm t_{0.025_{n-2}} . \hat{\sigma}_u \sqrt{\sum_{i=1}^{n} X_i^2 \Big/ n \sum_{i=1}^{n} x_i^2}$$

Similarly, 99% confidence limits for α are :

$$\hat{\alpha} \pm t_{0.005_{n-2}} . \hat{\sigma}_u \sqrt{\sum_{i=1}^{n} X_i^2 \Big/ n \sum_{i=1}^{n} x_i^2}$$

[The values of $t_{0.025,\, n-2}$ and $t_{0.005,\, n-2}$ corresponding to $(n - 2)$ d.f can be obtained from the table, given at the end of the book.]

In the same way, for testing β, we have,

$$\tau = \frac{\hat{\beta} - \beta}{\sigma_u \sqrt{\dfrac{1}{\sum_{i=1}^{n} x_i^2}}} = \frac{(\hat{\beta} - \beta)\sqrt{\sum_{i=1}^{n} x_i^2}}{\sigma_u}$$

when σ_u is not known then it is replaced by its unbiased estimator $\hat{\sigma}_u$, then we have :

$$t = t_{n-2} = \frac{(\hat{\beta} - \beta)\sqrt{\sum_{i=1}^{n} x_i^2}}{\hat{\sigma}_u}$$ with d.f = n – 2. Now rearranging we may get,

$$\hat{\beta} - \beta = t_{n-2} . \frac{\hat{\sigma}_u}{\sum_{i=1}^{n} x_i^2} \quad \text{or,} \quad \beta = \hat{\beta} \pm t_{n-2} . \frac{\hat{\sigma}_u}{\sqrt{\sum_{i=1}^{n} x_i^2}}$$

Therefore 95% confidence limits for β would be :

$$\hat{\beta} \pm t_{0.025,\, n-2} . \frac{\hat{\sigma}_u}{\sqrt{\sum_{i=1}^{n} x_i^2}}$$

and 99% confidence limits for β would be :

$$\hat{\beta} \pm t_{0.005,\, n-2} . \frac{\hat{\sigma}_u}{\sqrt{\sum_{i=1}^{n} x_i^2}}$$

Usual test procedure : One hypothesis of usual interest is that we hypothesise that there is no relationship (linear) between the explanatory variable X and the dependent variable Y in the regression model : $Y = \alpha + \beta X$.

As such the null hypothesis of no relationship between X and Y is $H_0 : \beta = 0$ and we have to test it against the alternative hypothesis $H_1 : \beta \neq 0$.

Now the appropriate test statistic under $H_0 : \beta = 0$ would be

$$t = t_{n-2} = \frac{\hat{\beta} - 0}{SE(\hat{\beta})} = \frac{\hat{\beta}}{\hat{\sigma}_u \Big/ \sqrt{\sum_{i=1}^{n} x_i^2}} = \frac{\hat{\beta}\sqrt{\sum_{i=1}^{n} x_i^2}}{\hat{\sigma}_u}$$

which has t-distribution with d.f = $(n - 2)$.

Now at 5% level of significance the null hypothesis will be rejected for the given sample if $|t_{n-2}|$ (observed) $> t_{0.025,\, n-2}$ and will be accepted otherwise (i.e., if $-t_{0.025,\, n-2} \le t \le t_{0.025,\, n-2}$). Similarly, at 1% level of significance the null hypothesis will be rejected for the given sample if $|t_{n-2}|$ (observed) $> t_{0.005,\, n-2}$ and will be accepted otherwise (i.e., if $-t_{0.005,\, n-2} \le t \le t_{0.005,\, n-2}$). The confidence limits for β (acceptance region in a two tailed test) at 5% and 1% levels of significance with $(n - 2)$ degrees of freedom will be given by,

$$-t_{0.025,\, n-2} . SE(\hat{\beta}) \le \beta \le +t_{0.025,\, n-2} . SE(\hat{\beta})$$

and $-t_{0.005,\, n-2} . SE(\hat{\beta}) \le \beta \le +t_{0.005,\, n-2} . SE(\hat{\beta})$

where $SE(\hat{\beta}) = \dfrac{\hat{\sigma}_u}{\sqrt{\sum_{i=1}^{n} x_i^2}}$ (σ_u is not known, and replaced by $\hat{\sigma}_u$)

2.14. Goodness of Fit of the Multiple Correlation Coefficient (R^2)

So far we were concerned with the estimation and precision