[Academic Script]

Preview of Econometric Methods

Subject:	Business Economics
Course:	B. A. (Hons.), 5th Semester, Undergraduate
Paper No. & Title:	Paper – 531 Elective Paper Q1 – Advanced Econometrics
Unit No. & Title:	Unit – 1 Relaxing the Assumptions of the Classical Linear Model
Lecture No. & Title:	Lecture – 1 Preview of Econometric Methods

Academic Script

1. Introduction

Hello friends nice meeting you. We are meeting through this media in which we shall be discussing certain things connected with advance econometrics subject. In this lecture series we have designed twelve lectures which will be taken up one by one. Before going to that let me introduce something about the subject Econometrics. Econometrics means economic measurement. This subject is composed of three main branches- Economics, Mathematics and Statistics. Economic theories get there applications through this three subjects and they have immense application, immense usefulness everywhere. By studying this subject there can be number of theories which are used in economic model can be tested and applied at many places. In this lecture first of all we will give you certain preliminary things about economics subject then we will give you the fundamental basis on which it is build up that is classical two variables models and classical three variable model. By learning this in nut shell you will be able to have the fundamental basis of the subject econometrics and I am sure that this will lead to the mild stone where we want to go on. Please remember that this is our pursuit for perfection.

What is Econometrics?

The word econometrics means 'economic measurement'. As per Goldberger, econometrics may be defined as the social science in which the tools of economic theory, mathematical and statistical inference are applied to the analysis of economic phenomena. Theil describes it as the subject which is concerned with the empirical determination of economic laws. In a way it is an economist's approach to statistical analysis. It can be considered as a field based upon a theoretical-quantitative and empirical –quantitative approach to economic problems. Thus econometricians are regarded as applied economists using the tools developed by theoretical econometricians to examine and analyse economic phenomena.

While applying different tools and techniques to understand the behavior of the variables concerned for any economic phenomena, everything is based upon what is called data analysis. In general, we can have three possible types of data.

(1) Cross sectional data

This type of data consists of measurements for individual observations (persons, households, firms, states, countries etc.

(e.g. population census figures for India in 2001, 2011 etc.)

(2) Time series data

This type of data consists of one or more variables (such as GDP, interest rates, unemployment rates etc.) over time in a given space (like a specific country or state or industry etc.)

(e.g. production data for steel industry during last 30 years in India.)

(3) Panel or longitudinal data

It consists of a time series for each cross sectional unit in the sample. These data contain measurements for individual observations (persons, households, firms, states, countries, etc.) over a period of time (days, months, quarters, years).

(e.g. Data for 4 companies about their investments, value of the firm and capital stock during last 20 years).

Traditional or classical econometric methodology consists of the following steps

(1) Theoretical statement or hypothesis
(2) To specify the mathematical model of the theory
(3) To specify the statistical or econometric model
(4) To collect data
(5) To estimate the parameters of the model
(6) To state and test the hypothesis
(7) To obtain prediction based upon the fitted model
(8) To use the model developed for its policy implications.

Econometrics may be divided into two broad categories (a) Theoretical (b) applied.

In each category the subject can be approached by classical or Bayesian tradition. We want to consider for our study here the classical approach.

2. Classical Two Variables Regression Model

We illustrate this by means of family weekly consumption and income data.

Let Y = weekly consumption (in monthly units)

X = Weekly Income (in monthly units)

Since consumption depends upon income, we may post a linear model as under

$$Y = \alpha + \beta X + U \quad \ldots\ldots (1)$$

Where α and β are the parameters of the model and U is the disturbance term. If we collect information about the whole population (e.g. a region or state), the above equation (1) is called Population Regression Function (PRS). Instead, if we draw a sample from the population then we can write the above relation as in (2) which is called Sample Regression Function (SRF)

$Y = \hat{\alpha} + \hat{\beta}X + e$ . . . . …… ……………………… (2)

Where $\hat{\alpha}$ and $\hat{\beta}$ are the estimated values for α and β and e is the error term.

These estimates of α and β are obtained by the method of least squares (hence they are called OLS estimators – ordinary least squares estimators)

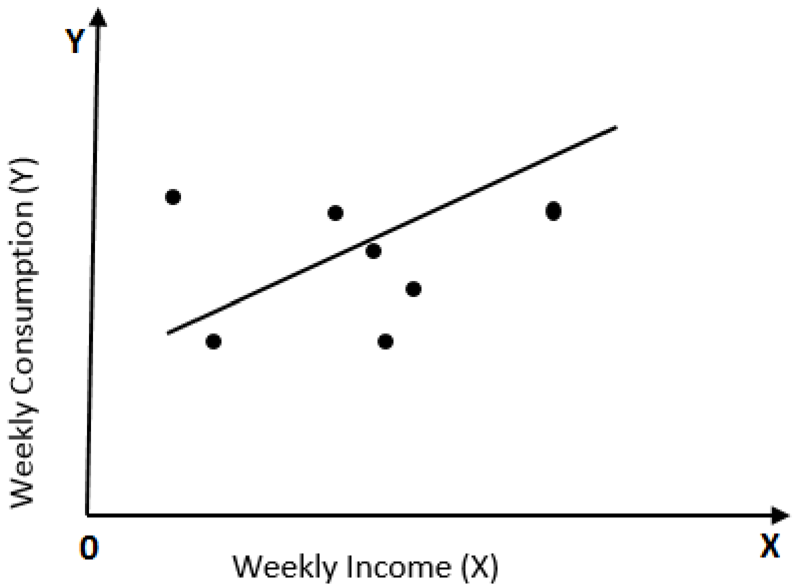
This is done by least squares method.

Analytically $Y_i = \hat{\alpha} + \hat{\beta} X_i + e_i$ (i = 1, 2 ……., n)…………………………..…. (3)

Then $e_i = Y_i - \hat{\alpha} - \hat{\beta} X_i$ = error

Hence $\sum_1^n e_i^2 = \sum_1^n (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2 = Error\ sum\ of\ squares$ … … … … … … …(4)

We find $\hat{\alpha}$ and $\hat{\beta}$ by minimizing the error sum of squares.



Diagrammatically, we are searching for a line in the plotted graph for (X_i, Y_i) such that the sum of squares of the distances from the points to the line is minimum.

Theoretically, we can obtain OLS estimators by the formula

$\hat{\beta} = \frac{\sum_1^n x_i y_i}{\sum_1^n x_i^2}$ . .. . (5)

$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$ . . . . (6)

Where $x_i = X_i - \bar{X}$, $y_i = Y_i - \bar{Y}$, Where $\bar{X}$ =Mean of X and $\bar{Y}$ = Mean of Y

If we insert these estimators in the model then the prediction for Y_i for given X_i can be obtained as $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$ (i= 1, 2,3,…..,n)

The formula for variances and standard errors of the estimators are as under

V $(\hat{\beta}) = \frac{\hat{\sigma}^2}{\sum_1^n x_i^2}$ . . . . (7)

V $(\hat{\alpha}) = \left[\frac{\sum X_i^2}{n\sum x_i^2}\right]\hat{\sigma}^2$ . . . . (8)

Where σ^2 is population variance and it is estimated by $\hat{\sigma}^2 = \frac{\sum_1^n e_i^2}{(n-2)}$ . . .. (9)

Where $e_i = Y_i - \hat{Y}_i$ (i = 1,2 , …. ,n)

Positive square root of the estimated parameters are their standard errors.

Thus S.E. of $\hat{\beta} = +\sqrt{V(\hat{\beta})}$ . . . . (10)

S.E. of $\hat{\alpha} = +\sqrt{V(\hat{\alpha})}$ . . . . (11)

The correlation coefficient between X and Y is computed from the formula

$r = \frac{\sum x_i y_i}{\sqrt{(\sum x_i^2)(\sum y_i^2)}}$ $x_i = X_i - \bar{X}$ and $y_i = Y_i - \bar{Y}$ . ... (12)

Theoretically r lies between -1 to 1 ($-1 \le r \le 1$)

r^2 (Usually denoted by R^2) is called Coefficient of determination.

R^2 lies between 0 and 1.

Value of R^2 near to 1 indicates very strong linear relation between the concerned variables.

(e. g. R^2 = 0.87 shows that about 87% of the variation is explained by the model thus leaving only 13% which is not explained).

If R^2 = 1, it is a perfect positive correlation between X and Y. Here all the points will fall on the fitted line.

If R^2 = 0 , there is no linear relationship.

(However there may be some sort of nonlinear relationship).

2. Testing the Model

(A) Testing the parameters

We can use student's t tests as under

(1) Test for β

$H_0 : \beta = \beta_0$ Where β_0 = specified value of β

Vs. H_1 : β ≠ β_0

Then t $= \frac{\hat{\beta} - \beta_0}{S.E\ of\ \hat{\beta}}$ .... (13)

Which has student's t distribution with (n-2) degrees of freedom.

We can compare the calculated value t_c with tabulated value t_T from t tables for

(n-2) d .f. and decide

If $t_c > t_T$: we reject H_O (significant)

If $t_c \leq t_T$: we may accept H_O (not significant)

We can test at 5% or 1% level of significance. If the hypothesis may be accepted then we have confidence interval for ß given as

$$\left[\beta_O - t_{\frac{\alpha}{2}}(\text{S.E. of } \hat{\beta}) \leq \beta \leq \beta_O + t_{\frac{\alpha}{2}}(\text{S.E. of } \hat{\beta})\right] \text{.................... ..(14)}$$

This gives the interval in which β will fall with probability $1-\alpha$, for given β_O.

(e. g. α = 0.05, gives 95% C.I for $\hat{\beta}$ and α = 0.01, gives 99 % C.I for $\hat{\beta}$)

(2) Test for α

$H_O : \alpha = \alpha_O$ Where α_O = specified value of α

Vs. H_1 : α ≠ α_O

Then t = $\frac{\hat{\alpha}-\alpha_0}{S.E \text{ of } \hat{\alpha}}$ .... (15)

which has student's t distribution with (n-2) degrees of freedom, then the decision can be taken as above.

Here also, we get the confidence interval for α as under

$$\left[\alpha_0 - t_{\frac{\alpha}{2}}(\text{S.E. of } \hat{\alpha}) \leq \alpha \leq \alpha_O + t_{\frac{\alpha}{2}}(\text{S.E. of } \hat{\alpha})\right] \text{................................ (16)}$$

Hence as usual we get 95% and 99% C.I as shown above.

(3) Test for ρ

We put hypothesis $H_O : \rho = 0$ Where ρ is the population correlation coefficient. Vs. $H_1 : \rho \neq 0$

Then apply t test by

t = $\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$.. (17)

which has student's t distribution with (n-2) degrees of freedom.

For suitability of the model, we expect t to be significant.

(4) Test for σ^2

We put hypothesis $H_0 : \sigma^2 = {\sigma_o}^2$

Vs. $H_1 : \sigma^2 \neq {\sigma_o}^2$

Where ${\sigma_o}^2$ is specified value of σ^2, then we apply χ^2 test by the formula

$$\chi^2 = \frac{(n-2)\hat{\sigma}^2}{{\sigma_o}^2} \quad \text{with (n-2) d. f.}$$ . . . . (18)

If $\chi_c^2 \leq \chi_T^2$: We may accept H_0

If $\chi_c^2 > \chi_T^2$: We reject H_0

$100(1-\alpha)$ % confidence interval for σ^2 is given by

$$\left[\frac{(n-2)\hat{\sigma}^2}{\chi^2_{\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-2)\hat{\sigma}^2}{\chi^2_{1-\frac{\alpha}{2}}}\right]$$ . ...(19)

$\chi^2_{\frac{\alpha}{2}}$ and $\chi^2_{1-\frac{\alpha}{2}}$ are the tabulated values for χ^2 for significance level 100(1-α) %

(e. g. for 95% confidence limits, $\chi^2_{\frac{\alpha}{2}} = \chi^2_{0.025}$ and $\chi^2_{1-\frac{\alpha}{2}} = \chi^2_{0.975}$)

(5) Testing the model

For appropriateness of the model we can carry out analysis of variance (ANOVA) as under

Source	**D. F.**	**S.S**	**M.S. S**	**F**
Due to regression (ESS)	1	$\sum \hat{y}_i^2 = \hat{\beta}^2 \sum_1^n {x_i}^2$	$\frac{A}{1}$ = A′	$\frac{A'}{B'}$

		= A		
Due to residuals(RSS)	n-2	$\sum_1^n e_i^{\,2}$ =B	$\frac{B}{n-2}$ = B′	–
Total (TSS)	n-1	$\sum_1^n y_i^{\,2}$	–	–

Here F = $\frac{A'}{B'}$ = $\frac{\frac{(ESS)}{1}}{\frac{(RSS)}{(n-2)}}$ which has F distribution with d. f. υ_1 and υ_2 = n-2 degrees of freedom.

We expect F to be significant for deciding about the suitability of the model.

Thus if $F_c > F_T$: F is significant

And if $F_c \leq F_T$: F is not significant.

(ANOVA also determines for significance of R^2)

<u>Note</u>

1) The above discussed classical two variables linear regression model is based upon the following assumption (i) E (U) = 0 (ii) V (U) = σ^2 (iii) Disturbance term U has normal distribution (this is needed for the testing of hypothesis problems).
2) The given model (assumed linear relationship) can be tested for normality assumption by means of tests of normality. (Refer to D. N. Gujarati p.141-144)
3) For hypothesis testing, the decisions can also be taken by means of p values .p value is the lowest significance level at which a null hypothesis can be rejected. (Refer to D.N. Gujarati p.132-133)

4) Once linear model is posed between the variables concerned (note that these are causal variables – some cause for relationship) only estimation process is not sufficient, Hypothesis testing part must be done as indicated above.
5) Also model will be useful if predictions are given based upon the fitted model. This can also help for policy planning decisions.
6) In the above model, α is the intercept and β is interpreted as the marginal propensity to consume-MPC). β is the slope coefficient.
7) Standard error of a statistic is defined as the positive square root of the sampling variance of the statistic.
8) We have two measures $S^2 = \frac{1}{n}\sum_1^n (x_i - \bar{x})^2$ and

$$s^2 = \frac{1}{n-1}\sum_1^n (x_i - \bar{x})^2.$$

s^2 is unbiased estimator for σ^2 , i.e. E(s^2) $=\sigma^2$

Thus (n-1) s^2 = nS^2 $\Rightarrow$ $s^2 = \left(\frac{n}{n-1}\right) S^2$

Illustration

Ex.1 The following data give the advertising expenditure and sales of a company during 1990 to 1997.

Year	1990	1991	1992	1993	1994	1995	1996	1997
Adv.exp.(thousands Rs.)	12	15	15	23	24	38	42	48
Sales(millions)Rs.	5.0	5.6	5.8	7.0	7.2	8.8	9.2	9.5

Estimate linear regression model for sales on advertising expenditure and estimate sales when the advertising expenditure is 60,000 Rs.

We consider the model $Y = \alpha + \beta X + U$

Where X = Advertising expenditure (000 Rs)

Y = Sales (Millions Rs.)

$\hat{Y} = \hat{\alpha} + \hat{\beta} X$

Using relation $\hat{\beta} = \frac{\sum_1^n x_i y_i}{\sum_1^n x_i^2}$ $x_i = X_i - \bar{X}$, $y_i = Y_i - \bar{Y}$

We get $\hat{\beta}$ = 0.125

Then $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = 3.87$

Hence the estimated linear regression is

$\hat{Y}$ = 3.87 + 0.125X

Now when X = 60, $\hat{Y}$ = 3.87+0.125(60) = 11.37

Hence the estimated sales will be 11.37 million Rs.

Ex.2 Following is the consumption-Income relationship for 10 families

Consumption =12.35 + 046(Income) S.E. = (3.95) (0.87) R^2 = 0.7854

Interpret this model

Here $\hat{\alpha}$= 12.35, $\hat{\beta}$ = 0.46 = MPC

S.E. of $\hat{\alpha}$ = 3.95, S.E of $\hat{\beta}$ =0.87

For an increase in income level by one unit more than 46% of the income goes into consumption.

78.54% of the variation is explained by the model.

3. Tri variate Linear Regression Model

For an industrial set up in a region, Let us define cob-Douglass production Function as under

Q = $AL^{\alpha}K^{\beta}$ . . . . (1)

Where Q = Output,

L =Labor force participation and

K = Capital Invested

α and β are the partial elasticies of labour and capital respectively. We can write double-log model relationship from the above which is

$\log Q = \log A + \alpha \log L + \beta \log K + U$ . . .. (2)

And define $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$ . . . . (3)

Where $Y_i = \log Q_i$, $X_{2i} = \log L_i$, $X_{3i} = \log K_i$

$\beta_1 = \log A$, $\beta_2 = \alpha$, $\beta_3 = \beta$

For the above PRF, based upon a sample of size n, SRF can be stated as under

$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + e_i$ . . . . (4) (i = 1, 2,.., n)

This is trivariate linear Regression model as given in (4) above

Note that $\beta_1, \beta_2, \beta_3$ are the partial regression coefficients and $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ are their estimated values. Using again least square principles, we can find OLS estimators by minimizing error sum of squares.

$\sum_1^n e_i^2 = \sum(Y_i - \widehat{\beta_1} - \widehat{\beta_2} X_{2i} - \widehat{\beta_3} X_{3i})^2$ . . .. (5)

Equation (3) above is the equation for plane of regression for Y on the variables X_2 and X_3.

For the above model in (3), we make the following assumptions

(1)E $(U_i | X_{2i}, X_{3i})$ = 0 for all i

(2)V (U_i) = σ^2 for all i (homoscedasticity)

(3)COV $(U_i, U_j) = 0$, i = 1, 2 , . . . , n,   j = 1 , 2 , . . . , n $(i \neq j)$

(4) Variables X_2 and X_3 are not linearly related. (No Multicollinearity)

OLS estimators for β_1 ,β_2 and β_3 are given by the relations as under

Write $x_{2i} = X_{2i} - \overline{X_2}$ = Deviation from mean

$x_{3i} = X_{3i} - \overline{X_3}$ = Deviation from mean

$y_i = Y_i - \bar{Y}$ = Deviation from mean

Then we have the following relations

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}{}^2)-(\sum y_i x_{3i})(\sum x_{2i}x_{3i})}{(\sum x_{2i}{}^2)(\sum x_{3i}{}^2)-(\sum x_{2i}x_{3i})^2} \quad \ldots\ldots (6)$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}{}^2)-(\sum y_i x_{2i})(\sum x_{2i}x_{3i})}{(\sum x_{2i}{}^2)(\sum x_{3i}{}^2)-(\sum x_{2i}x_{3i})^2} \quad \ldots\ldots (7)$$

Then $\hat{\beta}_1$ is computed from

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 \quad \ldots\ldots (8)$$

Let us define r_{23} = correlation coefficient between X_2 and X_3 given in (9)

$$r_{23}^2 = \frac{(\sum x_{2i}x_{3i})^2}{(\sum x_{2i}^2)(\sum x_{3i}^2)} \quad \ldots\ldots (9)$$

We have following relations for variance of estimators

$$V(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2(1-r_{23}^2)} \quad \ldots\ldots (10)$$

$$V(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2(1-r_{23}^2)} \quad \ldots\ldots (11)$$

S.E. of $\hat{\beta}_2$ and $\hat{\beta}_3$ are the positive square root of their variances, and covariance between $\hat{\beta}_2$ and $\hat{\beta}_3$ is given by

$$COV(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23} \cdot \sigma^2}{(1-r_{23}^2)\left(\sqrt{x_{2i}^2}\right)\left(\sqrt{\sum x_{3i}^2}\right)} \quad \ldots\ldots (12)$$

In the above formula,σ^2 is estimated by

$\hat{\sigma}^2 = \frac{\sum e_i^2}{(n-3)}$ ……………………………………………………………… (13)

Note that when the parameters $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ are estimated, the value of Y_i for given values of X_{2i} and X_{3i} can be estimated (prediction) by

$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$ (i = 1, 2, 3,. . . ,n)

We can define product moment (simple) correlation coefficient as under

(Note that Y ⇒1, X_2 ⇒2, X_3 ⇒3)

r_{12} = correlation coefficient between Y and X_2

r_{13} = correlation coefficient between Y and X_3

r_{23} = correlation coefficient between X_2 and X_3

Then we have the following correlation coefficient matrix P, defining the above correlation coefficient

$$P = \begin{array}{c} \\ Y \\ X_2 \\ X_3 \end{array} \begin{array}{c} \begin{array}{ccc} Y & X_2 & X_3 \end{array} \\ \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix} \end{array} \quad : 3\times 3$$ . . . . ………………………………………...

(14)

$r_{ij} = r_{ji}$ (i , j = 1,2,3, i≠j)

Which is a real symmetric matrix.

1. <u>Partial correlation coefficients</u>

We consider (pair wise) the correlation coefficient of two variables, when the third one is kept constant. Then we can define the partial correlation coefficients as under

$r_{12.3}$ = Partial correlation coefficient between Y and X_2 keeping X_3 as fixed (constant)

$r_{13.2}$ =Partial correlation coefficient between Y and X_3 keeping X_2 as fixed (constant)

$r_{23.1}$ =Partial correlation coefficient between X_2 and X_3 keeping Y as fixed (constant)

(Variable after subscript shows omitted category)

The relations for the above partial correlation coefficients are as under in terms of the simple (product moment) correlation coefficients

$$r_{12.3} = \frac{r_{12} - r_{13}.r_{23}}{\sqrt{(1-r_{13}{}^2)(1-r_{23}{}^2)}} \quad \ldots\ldots\ldots\ldots\ldots (15)$$

$$r_{13.2} = \frac{r_{13} - r_{12}.r_{23}}{\sqrt{(1-r_{12}{}^2)(1-r_{23}{}^2)}} \quad \ldots\ldots\ldots\ldots\ldots (16)$$

$$r_{23.1} = \frac{r_{23} - r_{12}.r_{13}}{\sqrt{(1-r_{12}{}^2)(1-r_{13}{}^2)}} \quad \ldots\ldots\ldots\ldots\ldots (17)$$

These are also called <u>first order correlation coefficients</u>. All product moment and partial correlation coefficients lie between -1 and +1.

4. Multiple correlation Coefficient and Multiple Coefficient of determination

To examine the combined effect of both X_2 and X_3 on the dependent variable Y, we measure <u>Multiple Correlation Coefficient</u> denoted by $R_{1.23}$ or R. It is defined as

$$R = R_{1.23} = \text{Corr}(Y, \hat{Y}) = \frac{COV(Y, \hat{Y})}{\sqrt{V(Y).V(\hat{Y})}} \quad \ldots\ldots\ldots\ldots (18)$$

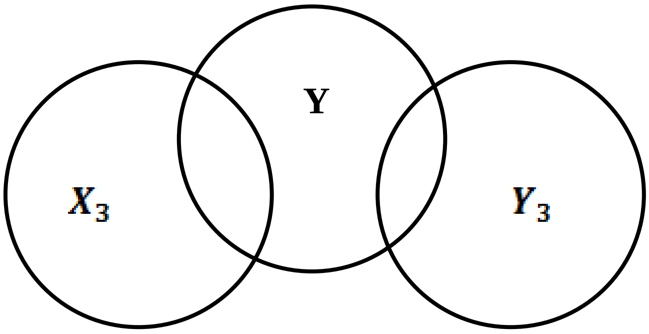
We define R^2 (i.e. $R^2{}_{1.23}$) as the Multiple Coefficient of Determination and it can be computed by R^2 = $\frac{Explained\ sum\ of\ squares}{Total\ sum\ of\ squares} = \frac{EV}{TV}$ . . . . (19)

R^2 Always lies between 0 and 1. ($0 \leq R^2 \leq 1$)

(e.g. R^2 = 0.73 means 73% of the variation is explained by the model, thus leaving only 27% unexplained)

For our 3 variables linear model, formula for R^2 is given by

$$R^2 = \frac{\widehat{\beta_2}\sum y_i x_{2i} + \widehat{\beta_3}\sum y_i x_{3i}}{\sum {y_i}^2} \quad \ldots\ldots \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (20)$$



(Diagram representing the situation for multiple correlation of X_2 and X_3 on Y)

R^2 (i.e. $R^2_{1\cdot 23}$) can be expressed in terms of zero order correlation coefficients by the formula

$$R^2_{1\cdot 23} = \frac{r^2_{12} + r^2_{13} - 2r_{12}\, r_{23}\, r_{13}}{1 - {r_{23}}^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (21)$$

Similarly it can be expressed in terms of zero order and first order correlation coefficients, given by

$$R^2{}_{1.23} = 1 - (1 - r^2_{12})(1 - r^2_{13\cdot 2}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots. (22)$$

2. R^2 and Adjusted R^2

To compare two R^2 values, one must take into account the number of X variables present in the model.

We have $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum {e_i}^2}{\sum {y_i}^2}$

We define Adjusted R^2 given by $\bar{R}^2$ (R^2_{adj}) by

$$\bar{R}^2 = 1 - \frac{\sum {e_i}^2 / (n-3)}{\sum {e_i}^2 / (n-1)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (23)$$

It is called adjusted R^2 for d.f. associated with the sum of squares .This is useful for comparing two regression models. In each of them, dependent variable is the same, but explanatory variables can be different.

Here n = Number of observations (sample size)

k = Number of parameters in the model including the intercept term=3

Relation between $\bar{R}^2$ and R^2 is given by

$$\bar{R}^2 = 1 - \left(\frac{(n-1)}{(n-k)}\right)(1 - R^2) \quad \text{. .... (24)}$$

(i) For k$> 1, \bar{R}^2 < R^2$.

(ii) R^2 is positive but $\bar{R}^2$ can be negative and its value is taken as zero.

(iii) We can choose model with greater $\bar{R}^2$ as more suitable model.

3. Testing Problems

Test procedures for ordinary and first order partial correlation coefficients are similar to those for two variables (note that d.f. = n-3) .Also confidence intervals can be computed accordingly.

4. <u>ANOVA Approach</u>

For testing the suitability of the model we have ANOVA table as under

Source	D.F.	S.S.	M.S.S.	F
Due to regression(ESS)	2	$\widehat{\beta_2} \sum y_i x_{2i}$ + $\widehat{\beta_3} \sum y_i x_{3i}$	A′ = A/2	A′/B′

		= A		
Due to residual(RSS)	n-3	$\sum e_i^2$ = B	B′ = B/(n-3)	-
Total	n-1	$\sum y_i^2$	-	-

To test the hypothesis

$H_0 : \beta_2 = \beta_3 = 0$

Vs $H_1 : \beta_2 \neq \beta_3 \neq 0$

We expect F to be significant for validity of the model

Here F = $\frac{A'}{B'}$ has F distribution with ν_1 = 2 and ν_2 = n-3 d.f.

Hence it can be tested as usual.

5. Testing for significance of R^2

$H_0 : \rho^2 = 0 \Rightarrow \rho = 0$ ρ = population multiple coefficient of determination.

Vs $H_1 : \rho^2 \neq 0 \Rightarrow \rho = 0$

We have F = $\frac{R^2/2}{(1-R^2)/(n-3)}$ which has ν_1 = 2 and ν_2 = n-3 d.f. and

we expect F to be significant for validity of the model. (that is for significance of R^2)

<u>Note</u> (1) Only estimation of parameters is not sufficient.

(2) One must go for testing problems as indicated above.

(3) Once the model is found to be suitable (best fitted) we can obtain prediction of Y based upon given (or forecasted) values of the explanatory variables.

(4) These predictions can be useful for future planning purposes.