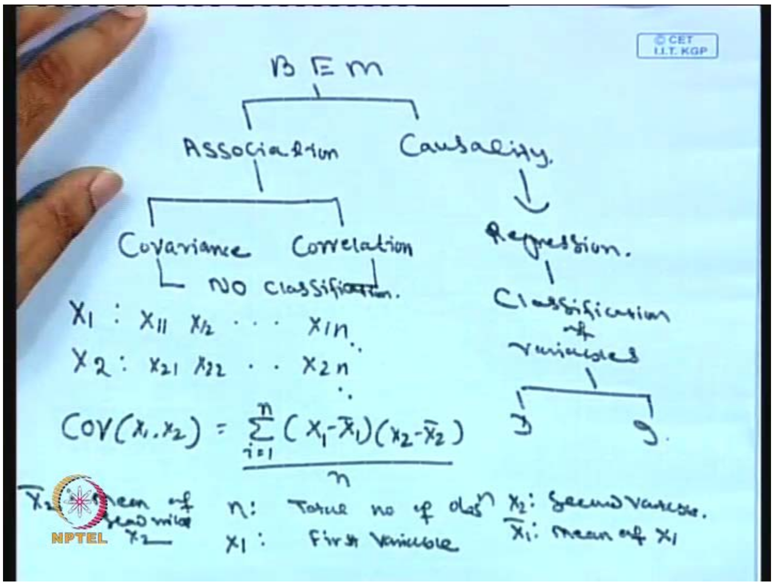
Econometric Modelling
Prof. Rudra P. Pradhan
Department of Management
Indian Institute of Technology, Kharagpur

Lecture No. # 04
Bivariate Econometric Modelling

Good afternoon. Welcome to NPTEL project on econometric modeling. Today, we will discuss the concept bivariate econometric modelling. It is a statistical tool, which deals with the relationship between two variables only. In the last class, we have discussed the entire structure of data analysis, that is, univariate modeling, bivariate modelling and multivariate modeling. Particularly in the last class, we have discussed the entire structure of univariate modeling, that is, with respect to central tendency, dispersion, skewness and photosis.

The main objective behind univariate modelling is that we have to describe the features of a particular variable, that is, with respect to its average, mid-value, frequency distributions and its variability within the system. However, in the real world lots of variables are integrated with other variables. We cannot generalize a particular problem or we cannot discuss a particular problem with respect to a single variable. The analysis of univariate modelling is very essential or it is the essential condition or you can say necessary condition for bivariate modelling and multivariate modeling. So, we are here to know details about bivariate modeling. It is the game between two variables at a time. So, let me take a case what is exactly the concept of bivariate econometric modelling.

(Refer Slide Time: 02:48)



Bivariate econometric modeling – I will call it here BEM (bivariate econometric modeling); bivariate econometric modelling basically deals with two problems: association and causality. So, we have we have two variables in a system. In a bivariate econometric modeling, our objective is to know the association between two variables; and, second objective is to know the cause and effective relationship between the two variables. In some of the problems, you may not need to know the cause and effective relationship. In some of the cases, you are also not very much interested about the association between two variables. It is a very interesting history behind the movement from univariate to bivariate and bivariate to multivariate. In this particular bivariate framework, again there is a history. So, the history is that the movement from association to causality. Both are somewhat similar in nature, but causality is little bit much higher and much better than the association, because causality is the generalized concept and association is the part of causality.

Now, in this particular stroke models here bivariate econometric modeling, we have two basic objectives: association and causality. So, we have altogether three forms of technique. In the association, we have two different techniques called as a covariance and here we have correlation. And, in the case of causality, we have a technique called as regression. So, in the bivariate framework, we have two different games: one is association between two variables and another is cause and effect relationship between two variables. So, for association is concerned, we can apply covariance and we can apply correlation. Between the two, correlation is much better and much advanced technique than covariance. We will discuss how it is advanced and how it is much better than covariance. However, the origin of bivariate econometric modelling is covariance. Correlation is an extension of covariance. Similarly, regression is also an extension of correlation. So, the movement is from covariance to correlation, then correlation to regression.

We will discuss here first what is all about the covariance; then, we like to know what is correlation; then, we have to proceed for regression. The moment you will enter to the regression, then that is the root point of real econometric modeling. Whatever components we are discussing now, it is just supporting components to econometric modelling and it is very essential. Until unless you know the concept structure of univariate modeling, bivariate modelling and its limitation or advantages, you cannot proceed further econometric modelling with a multivariate framework.

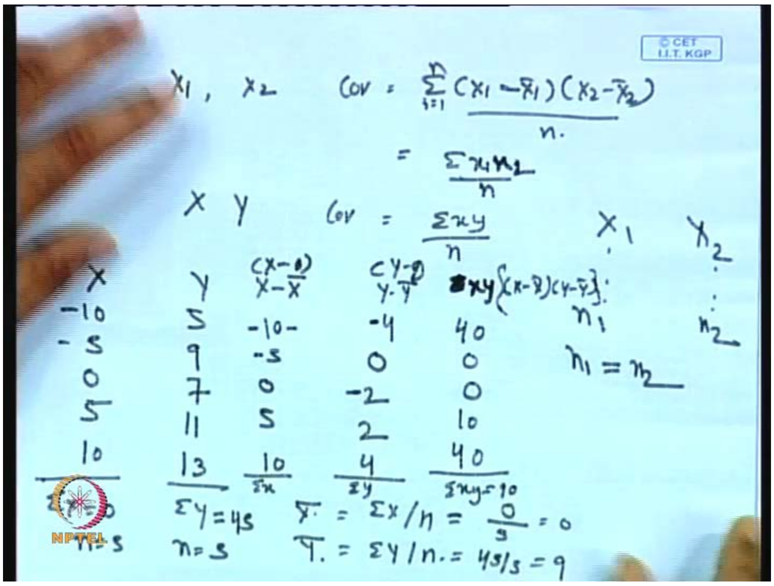
Now, we will start with the issue of the bivariate game; that to covariance analysis. So, what is all about this covariance analysis? Let me explain here covariance analysis. Let us take a case; here (Refer Slide Time: 07:44) is two variables say X 1, which represents the component X 1 1, X 1 2 up to X 1 n. And, another variables we have X 2 to X 2 1, X 2 2 up to X 2 n. So, now, the moment you will say bivariate econometric modeling, the boundary must be in between two variables. That is first condition. And, second condition is that since we like to trans the association between two variables, the sufficient condition is that the sample information must be uniform, must be similar. For instance, if X 1 contains n number of sample points, then X 2 must have n number of points. If X 1 is n minus 1 and X 2 is n or vice versa, then the system is inconsistent. To apply covariance technique or correlation technique or regression technique, the first prime requirement is that you must have uniform sample distribution. So, the observation for both the variables should be similar and unique. If the observations are not similar or unique, the system itself is inconsistent. So, now, with the inconsistent system, you cannot apply none of the techniques; neither correlations nor you can say regression or covariance.

Now, the starting point here (Refer Slide Time: 09:38) is that the classification of variables. In the last class, I have discussed variable classification and that too in a multivariate modelling and that too entire structure of data analysis. When the system is univariate, then the classification of variable is not at all matter, because there is only one variable; we have no clue to make the classification. The moment you enter to the bivariate econometric modeling, you must have the problem about classification of variables. In a bivariate framework, classification of variable is sometimes important, sometimes may not be important. And, it depends upon the technique, which we use in the particular process. If we handle the technique covariance or correlation, then the classification of variables are not at all important. However, if you go for regression technique, then classification of variable is very important; until and unless you classify the variable, then you cannot apply the regression technique; that means, what is all about this classification? Classification here we mean the dependent classification and independent classification, which we have discussed earlier in detail.

Here the issue is (Refer Slide Time: 11:44) the classification of variables. Now, in the case of covariance and correlation, we need not require any classifications or no classification of variables; that too dependent and independent. However, in the case of regression, you need to have a classification of variables, that is, with respect to dependent structure and independent structure. Here in the case of covariance, you need not require anything to describe the classification of dependent variable and independent variable. Before we go for this regression or all about this dependent structure and independent structure, it is better we first know the exact issue of covariance; then, we have to move correlation and regression. So, question is what is covariance?

For this particular system, X 1 contains n items; X 2 contains n items; then, covariance is represented as Cov upon X 1 and X 2, represents summation X 1 minus X 1 bar into X 2 minus X 2 bar i upon 1 to n divided by n. So, this is the standard formula we like to apply to get the covariance. So, covariance is simply represented as the sum of X 1 deviation and X 2 deviation. And, it is the total number of observation. Here n represents total number of observations (Refer Slide Time: 13:49). And, X 1 represents a particular variable – first variable; and, X 2 represents second variable. Then, X 1 bar represents mean of X 1 variables. Similarly, X 2 bar represents mean of second variables, that is, X 2. Now i is the class size or class interval.

(Refer Slide Time: 14:46)

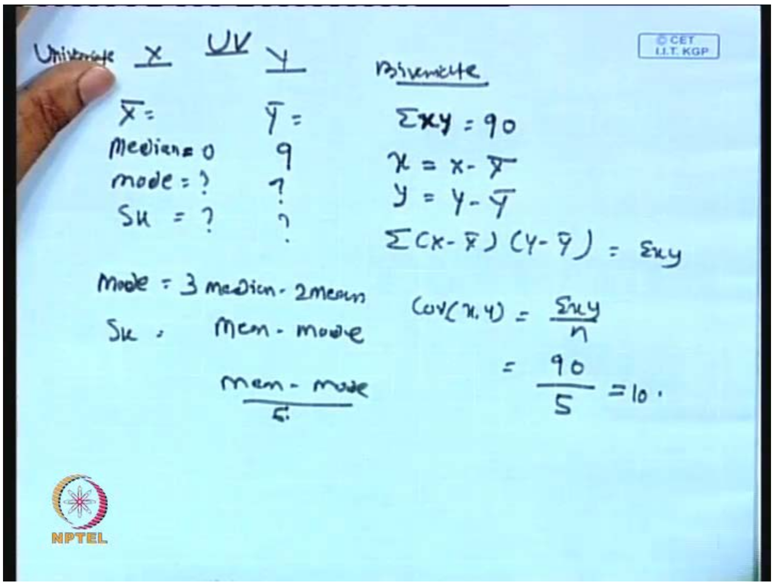


Let me write in a different way. Now, for X 1 and X 2, covariance is equal to summation X 1 minus X 1 bar into X 2 minus X 2 bar; i equal to 1 to n divided by n. So, now, I can write this structure into summation, simply x 1 and x 2 divided by n. And, for simplicity, you can write like this first: X and Y. Then, covariance of X Y equal to summation x y divided by n. So, n is the number of observations. However, for X and Y if you put X 1 and X 2, then obviously, the observation is n 1 and n 2. So, these are the observations. So, now, I have already mentioned that for covariance, the essential condition is that sample observation must be same. If it is same or uniform, then you can proceed further; that means n 1 must be equal to n 2. If n 1 not equal to n 2, then the system itself is inconsistent. So, now, you cannot apply here covariance if the sample observations are completely different. So, in order to have sample observation different, then you have to adjust the system, that is, with respect to simplicity of this picture.

Now, to highlight all these things, you need to have examples. Let us take a case of two examples (Refer Slide Time: 16:45) here: X and Y. X here is minus 10, minus 5, 0, 5 and 10. Now, here you have 5, 9, 7, 11, then 13. So, now, you need to calculate covariance. So, covariance is simply the summation of X minus X bar and Y minus Y bar divided by total number of observations; that means, we like to know what is sum of X, what is sum of Y. Then, we like to know what is X bar, what is Y bar. X bar is nothing but sum X by n and Y bar is nothing but sum Y divide by n. Corresponding X bar and Y bar, we must have the component X minus X bar and Y minus Y bar.

Now, X minus X bar – that means, (Refer Slide Time: 17:57) minus 10 here and minus this X bar. So, what is X bar here? Now, sum of X is equal to here 0; sum of Y is here equal to 45. So, now, corresponding to X bar, sum of X equal to 0 divide by number of observations; number of observation is here; n equal to 5 and this is also n equal to 5. So, 0 by 5 is equal to 0. And, summation Y by n is nothing but 45 by 5; it is equal to 9. So, obviously, it is X minus X bar; it is otherwise known as X minus 0; and, it is nothing but Y minus 9. So, now, the corresponding component is nothing but minus 10, then minus 5, then 0, then 5, then 10. So, in the case of Y, it is minus 4; then 0; then it is minus 2; then it is 2; it is 4. So, now, this is this sum of x and sum of y. So, we need to have X into Y. Then, we have to get summation XY. So, XY means here small xy; that is nothing but X minus X bar into Y minus Y bar. So, this is the sum case. So, xy is nothing but minus 10 into minus 4. So, it is 40; then, this is 0, this is 0, this is 10, this is 40. So, sum of xy is equal to simply 90. Now, I will summarize it.

(Refer Slide Time: 20:06)



Now, remember here; we have two variables: X variables and Y variables. Let us start with the univariate structure. Let us we start with the component univariate structure. So, what is univariate structure? Now, you like to know what is X bar, what is Y bar; this means arithmetic mean. Then, what is median, what is mode, what is skewness? Similarly, in the case of X and in the case of Y. Now, you know median is nothing but middle must below this sequence. So, if we arrange it in proper sequence, then obviously, in the first case, the median structure is here; so, 0. In the second case, you have to arrange (Refer Slide Time: 21:11) it in ascending order. And, if you apply the technique, then obviously, the size of median is the 9 also. Now, in the case of X, median (Refer Slide Time: 21:23) is 0; and, in the case of Y, it is 9. So, to get mode, mode equal to 3 median minus 2 mean. Accordingly, you will fill the gap from mode and you can fill the gap of mode. And, for skewness, skewness is equal to mean minus mode or mean minus mode by standard deviation, so that we have to calculate this skewness component. So, this is what the univariate structure is concerned.

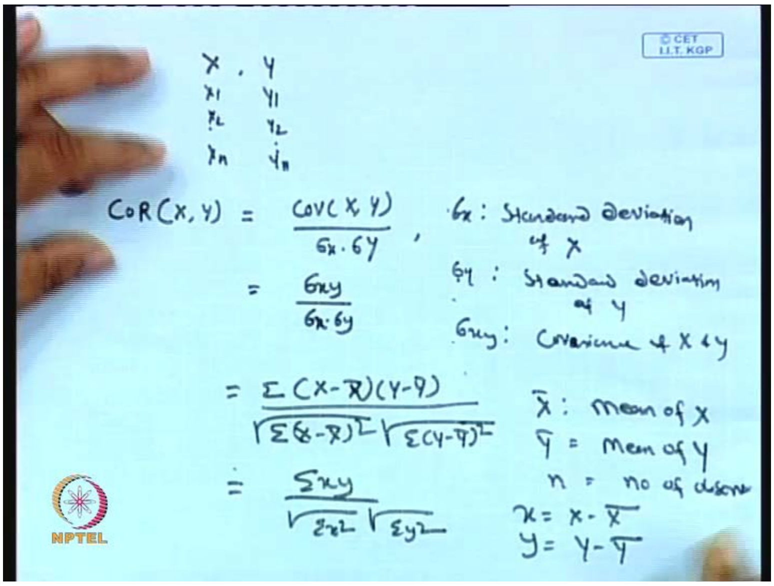
Now, before explaining the issue of covariance, you must have a clear-cut understanding about univariate statistics, because the univariate output will give you the path for bivariate structure. So, bivariate results depends upon the univariate results. So, you must have complete information about univariate statistic; then, you have to proceed further. Further, you can say (Refer Slide Time: 22:29) bivariate structure.

Now, coming to bivariate (Refer Slide Time: 22:35) structure, we need to know what is summation XY. So, summation xy is here – 90; this summation xy is the small xy, not capital XY. I like to clarify one thing here. x is a deviation, which represents X minus X bar; and, y is also deviation, which represents Y minus Y bar. So, obviously, sum of X minus X bar into Y minus Y bar is nothing but summation xy. So, now, covariance upon x, y represents summation xy by n. That is nothing but 90 divide by number of observation is 5, which is nothing but equal to 18. So, now, this structure will give you the concept of covariance. So, covariance is nothing but… So, the value of covariance is 18; that means, the association between X and Y is nothing but 18. So, now, we have complete information about univariate structure and complete information about the bivariate structure. So, that means… We like to know what is the X 1 in the system or X in the system and what is the nature of Y in the system. And, we like to know what the association between the two is. Now, that is possible under the structure of covariance.

Now, with the help of covariance, we like to know degree of associations, but sometimes there is a problem. That is how it is another technique called as a correlation. Now, one thing is very clear here is, because covariance will give you similar results, correlation also gives you similar results, because the objective of correlation and covariance is that it measures the degree of association between two variables. So, now, once you calculate the degree of association through covariance, then correlation not at all matters; or, if you calculate through correlation, then covariance does not matters. But, there is sometimes issue. The issue is that correlation is much better technique, much advanced technique than the covariance. Why? Because in some of the cases, covariance has limitation.

For instance, if you will go for comparative analysis… last class, we have discuss the issue between the US dollar and Japanese yen. Now, when you will go for comparative analysis, then obviously, unity of measurement always matters. So, in that case, the relative measure sometimes is much handy for the analysis or for a particular problem. Like in the case of univariate analysis, covariance is much better than standard deviation. In the case of bivariate modeling, correlation is much better than the covariance, because it is unitless measurement. So, now, we would like to know what is the structure and setup of correlations. So, let me highlight here what is the entire structure of correlation here.

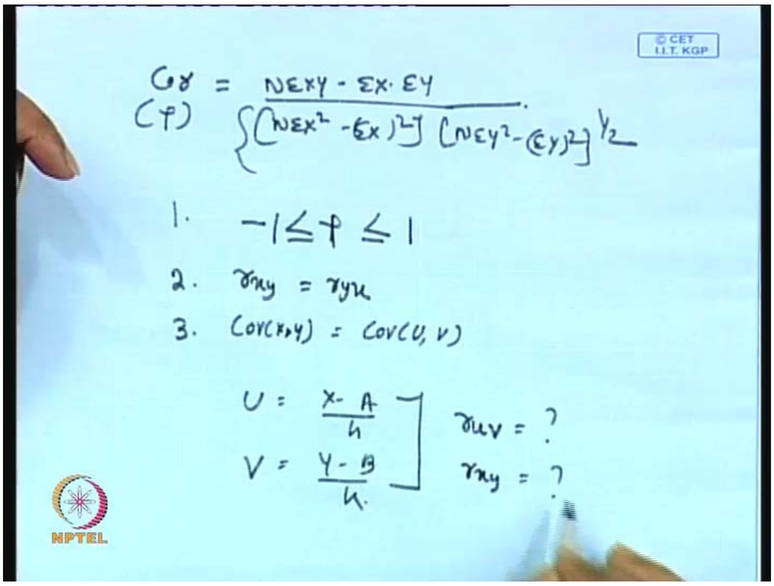
(Refer Slide Time: 26:28)



Correlation – For X and Y, where X contains X 1 up to X n and Y contains Y 1, Y 2, Y n. Now, we like to know what is the structure of correlations. Now, correlation is nothing but we will write CoR upon X and Y; then, correlation of X, Y is nothing but covariance of X and Y divided by sigma X into sigma Y. So, what is sigma X and what is sigma Y? That is the issue here. Now, sometimes, it is represented as sigma xy by sigma x and sigma y; where, sigma x stands for standard deviation of X and sigma y represents standard deviation of Y and sigma xy represents covariance of X and Y. So, sigma x, sigma y is the product of univariate modelling and sigma xy is the product of bivariate modelling. So, now, correlation is nothing but the ratio between covariance of XY by standard deviation of X and standard deviation of Y. So, put it in explicitly format. Then, the correlation of XY is nothing but summation X minus X bar into Y minus Y bar divided by summation X minus X bar whole square into summation Y minus Y bar whole square.

Now, the issue is here; so, you like to know what is X bar here; X bar is the mean of X; Y bar is the mean of Y; and, n represents number of observations. Now, n does not matter here, because upper side n and lower side n is cancelled. Now, if I simplify this structure, then it is simply represented as summation xy by summation x squares and summation y squares. Now, x is nothing but X minus X bar and y is nothing but Y minus Y bar; both are in deviation format.

(Refer Slide Time: 29:54)

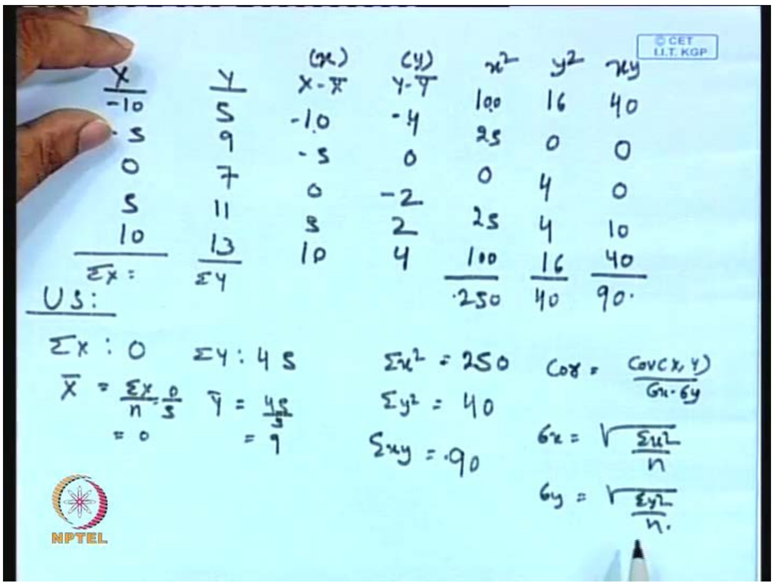


Now, with further simplicity, the correlation coefficient can be calculated as N summation XY minus summation X into summation Y divided by N summation X square minus sum X whole square into N summation Y squares minus sum Y whole squares whole to the power 1 by 2. So, this is the complete structure of correlations.

Now, the interesting (( )) is here – the value and nature of correlations. Now, before we proceed to put it in a real example format, we like to know what is the features and specialty of correlation. The feature and specialty of correlation is that first, the value of correlation coefficient, which is usually denoted as rho; so, minus 1 less than 1 rho less than equal to 1; that means, the value of correlation coefficient lies between minus 1 to plus 1. So, this is the standard techniques; mathematically, there is a proof. So, it is always true that the value of correlation coefficient should be in between minus 1 to plus 1. So, if it is minus value, then it is represented as negative correlation; and, if it is towards close, then it is called as positive correlation.

Now, second property is that correlation coefficient is symmetric in nature; r xy is equal to r yx. Then, covariance of X, Y is sometimes represented as covariance of U and V; U and V is treated as another set of variables; that means, the most important trick is that correlation coefficient is independent of change of original scale. For instance, if you need to take a case here, U equal to X minus A by h and V equal to Y minus B by k, then we will simplify this; then, we will get r uv. Whatever r uv you will get, same (Refer Slide Time: 33:03) structure you will be get it through r xy. Sometimes this origin and scale are very important when we will go for higher order problem and complex problem. So, it is very important for that particular angle. So, now, before proceeding further, I like to highlight here one thing that one of the condition of this correlation is that the variable structure or sample observations must be uniform from both end X and Y. So, let me here take an example – so, how this example can be applied to calculate the correlation coefficient?

(Refer Slide Time: 34:02)

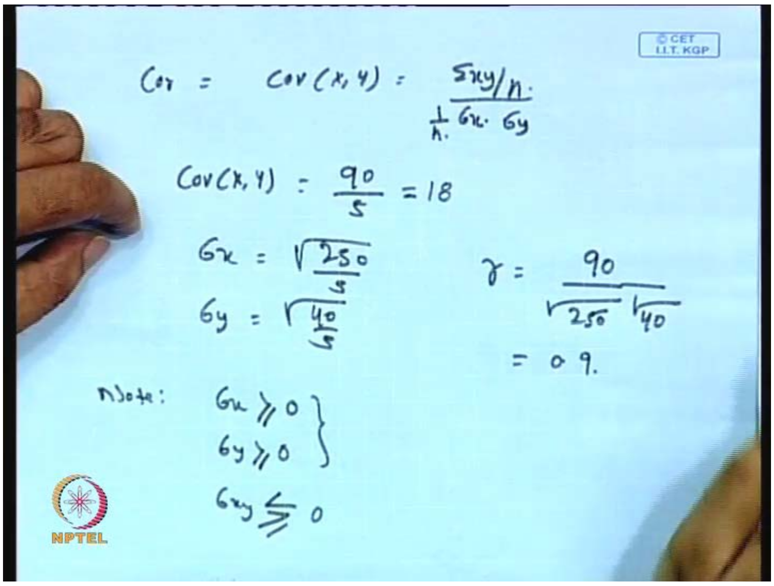


Now, we can cite the same example. In fact, we can cite the same example here; the example is here – X represents minus 10, minus 5, 0, 5, 10; and, Y represents 5, 9, 7, 11 13. We need to calculate correlation coefficient. As usual you must have some univariate statistics first; that is, descriptive situation. So, the (( )) will come automatically. Let me highlight here. This is, you need summation X you need summation Y.

Now, univariate statistic structure – you would like to know what is summation X, what is summation Y. Summation X is here 0 and summation Y is here 45. So, corresponding to summation X, we have X bar, which is nothing but summation X by n and n represents here 5. So; obviously, X bar is equal to summation X by n; that is nothing but 0 by 5 and which is equal to 0. So, now, summation Y case – it is Y bar is equal to again 45 by 5; it is equal to 9. Now, this particular information is very much required for further analysis. So, now, we need to have X minus X bar and Y minus Y bar. So, now, I am directly writing here; so, minus 10, minus 5, 0, 5, 10. Then, here minus 4, 0, 2, minus 2, 2, then 4. So, then, this represents X component (Refer Slide Time: 35:55) – small x; and, this represents small y; so, now, small x and small y.

We need to have information about small x square and we need to have information about small y square; and, we need to have information about small x into small xy. Now, x square represents here 100, then 25, 0, then 25, and 100. Then, y square is equal to 16, 0, then 4, then 4, then16. So, now, we like to know what is sum x square. So, now, come down to here sum x square and sum y square. So, then, we must have summation xy. Now, for summation xy, this is (Refer Slide Time: 36:55) minus 10 into minus 4 – this is 40; then, minus 5 into 0 – 0; then 0 minus 2 – 0; then, 5 into 2 – 10; then, 10 into 4 – it is 40. Now, if we simplify further, then summation x square is equal to 250; then, summation y square is 40; and, summation xy is nothing but 90. So, the sum total is 250, 40 and 90. Now, we have to know what the value of correlation coefficient is. Having such information here, we like to know the correlation statistics. So, now, correlation for XY is nothing but covariance of X, Y into sigma x into sigma y. So, now, sigma x is equal to summation x square by n and sigma y is equal to summation y square by n.

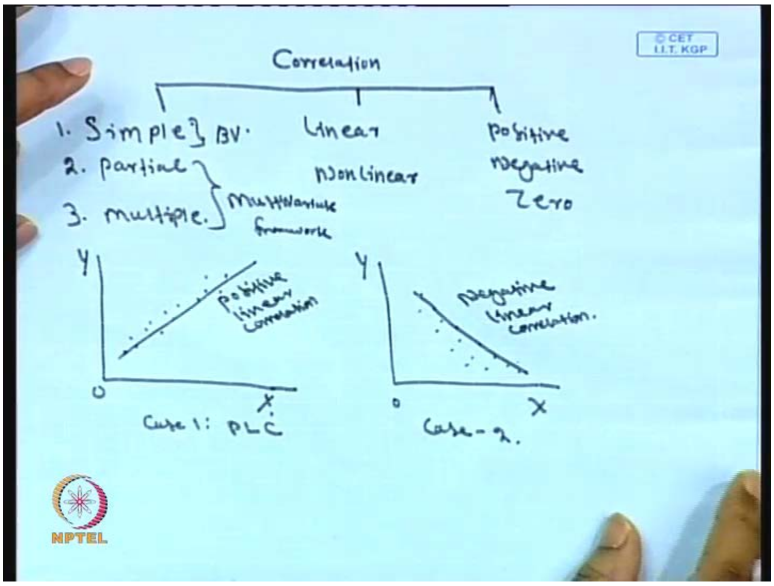
(Refer Slide Time: 38:17)



Now, the moment you will put it here, then obviously, correlation coefficient is equal to covariance of XY, which is nothing but summation xy by n divided by sigma x and sigma y upon 1 by n. So, now, we like to know what is covariance of X, Y. Covariance of X, Y is nothing but summation xy by n, which is nothing but 90 by 5. So, it will be around 18. So, now, we have already sigma x; sigma x equal to summation x square – 250 by 5 square root; and, sigma y is equal to 40 by 5 square root. So, the structure of correlation coefficient is that r equal to 90 by square root of 250 and square root of 40. So, n, n automatically cancels. Now, this is around 0.9

But, one thing is very clear here, we know the correlation coefficient is always between minus 1 and plus 1; and, the natural correlation depends upon the movement of covariance. The reason is that standard deviation of X must be always positive, because it is the square root of variance. Similarly, standard deviation of Y is also always positive. So, the value of correlation whether it is negative or positive depends upon the value of covariance. Now, if the covariance is negative, then we must have negative correlation; and, if the value of covariance is positive, then we must have positive correlation. So, now, with these structures we can cite here (Refer Slide Time: 40:46) the essential condition for this structure is that sigma x is always greater than 0; sigma y is always greater than 0; and, sigma x y is either greater than or less than equal to 0. Sigma x sometimes can be also equal to 0. So, that is why the condition is that sigma x greater than equal to 0; sigma y greater than equal to 0; and, sigma xy is greater than equal to 0.

(Refer Slide time: 41:41)



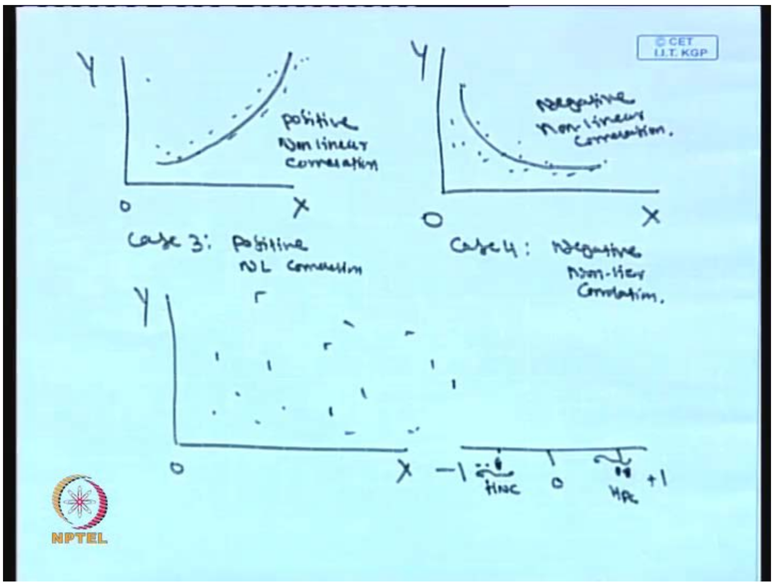
Now, in order for better simplicity, we like to know the nature of the correlation coefficient. Now, I will explain here the detailed structure of correlation. Correlation basically has three different formats: the first format represents the simple correlation; then, second is called as a partial correlation; third is called as a multiple correlation. So, correlation again can be linear and can be non-linear. Then, correlation can be positive, can be negative or can be zero. So, correlation – the basic framework is simple correlation, partial correlation and multiple correlation. It can be linear in nature; it can be non-linear in nature; it can be positive; it can be negative. Sometimes, the value of correlation coefficient can be also 0. If the value of correlation coefficient equal to 0, then there is no association between these two variables.

For instance, we have a relationship between pen and paper, but we may not have a relationship between pen and chair, because we do not have any link between pen and chair. So, one interesting issue is here that before entering to the correlation statistic, there must be sound theory behind it, because anything you try to integrate, you will get some value, because it is all about mathematical calculation. But, the interpretation, the utility, the usefulness depends upon its theory only. Theory will give you support or you can say sound structure, so that you can establish the problem setup. If there is no theory behind the correlation approach, then this term is called as simply nonsense correlation; it is sometimes called as a nonsense correlation.

Now, before we go into detail structure about that issue, let me explain how the accurate structure here is. Now, within the particular setup, this (Refer Slide Time: 45:17) particular structure is called as a multivariate framework. It cannot be with two variables here; it is with respect to more than two variables. But, simple correlation is a bivariate game. However, partial and multiple correlations are multivariate game. So, here we will not discuss about this partial correlation coefficient and multiple correlation coefficient, because we will discuss details when we will go for multivariate modeling.

Now, if we integrate all these structures here, then we have various forms. Let me explain here what these forms are. I will give you indication. This is (Refer Slide Time: 46:18) one way we can represent the correlation coefficient. Here X information and Y information. Now, one step of correlation structure is like this. Now, I will just draw this. This particular setup is called as a positive linear correlation. This is case 1. Case 2 – 0, X and Y; now, the structure may be like this. I am just highlighting what are the possibilities under the correlation modeling. Now, this particular structure is called as a negative linear correlation; this is situation of case 2.

(Refer Slide Time: 47:45)



I will put another structure here. I will represent like this 0, X, Y. Then, this structure may be like this; I will put like this; then, I will join like this. So, this is called as a positive non-linear correlation. This is case 3. So, case 3 is a positive non-linear correlation. Then, I represent another case here. Now, that case is like this; then, I will draw like this. So, this is X and Y. So, this particular structure is called as a negative non-linear correlation. So, now, case 4 is negative non-linear correlation. So, now, we have four different games: first is positive linear correlation; negative linear correlation, positive non-linear correlation, negative non-linear correlation. So, that means, altogether four different sets of correlation we can find.

If the structure is in between 0 to minus 1 and 0 to plus 1; if the structure or the value of correlation coefficient is exactly at that 0 level, that means the structure is completely different, which we call it zero correlation. So, that means there is another case, where we have no correlation between two variables. So, that means the setup is like this (Refer Slide Time: 50:21). So, we have no relationship between these two variables. So, that is why, you must have sound logic, sound theory, sound structure; then, you can apply the correlation technique or covariance technique.

Without any theory, logic and structure if you will apply correlation coefficient, then obviously, sometimes, either you may get zero correlation or you may get simply nonsense correlation; nonsense correlation means the value may not be equal to 0, but it does not support any theory. For example, if I will just plot one cite, this room, number of chairs are there; first room, number of chairs – 10; second room, number of chairs is 20; third room, number of chairs is 40; then, first room, the number of availability of pen is 5; second room, number of pens available – 20; third, number of pen is 40; then obviously, I do not find any theory between number of chairs in particular structure and number of pens in a another particular structure, where we will get the correlation. Now, correlation can be positive correlation, can be negative correlation; that is, with linear structure; and that is, with non-linear structure. And, in between, there may be the case called as a zero correlation. So, this is the entire setup of the structure of correlation.

Now, the correlation is a very important tool for bivariate econometric modeling. Specially, it is the middle part in between covariance and regressions. It is much better than covariance and less better than the regression. The advantage of correlation is that it brings the degree of association between these two variables. Since we have already mentioned, the value of correlation coefficient is in between minus 1 and plus 1, then obviously, the nature of association will be very divergent, also interesting. If the value of correlation is exactly 1, then it is called as a perfect positive correlation. If the value of correlation is exactly minus 1, then it is called as a perfectly negative correlation. If it is very close to minus 1, then it is highly correlated, negatively correlated correlation. If it is very close to plus 1, then it is called as a highly positive correlation.

However, if it is very close to in betweens like this – so, the structure is, this is (Refer Slide Time: 53:33) minus 1 and this is plus 1. So, this is 0. Now, if I will say here minus 0.8 and this is here 0.8, then if this is the (( )) then this is called as a high correlation; and, this is also high correlation; high negative correlation and this is high positive correlation. So, now, if it is in between 4 to 6, then it is called as a moderate correlation; it may be moderate positive; it may be moderate negative. So, now, if the value of correlation is less than 4, 3 or very close to 0, then it is called as a very low correlation. So, now, the association of the two variables can be very strong if the value of correlation is very high; if the value of correlation is low, then the association is also very low.

This is very interesting component and very useful for multivariate technique and it is a component regression technique. However, the essential part of this bivariate modelling is that you must have a thorough knowledge and complete information about the univariate modeling. Until and unless you have complete information, complete setup, you cannot handle the game of correlation. Now, it is not possible for us to discuss the detail about regression modelling here; again, within the setup of bivariate modeling, which we will discuss in the next class. With this, we can conclude this session.

Thank you very much. Have a nice day.