Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

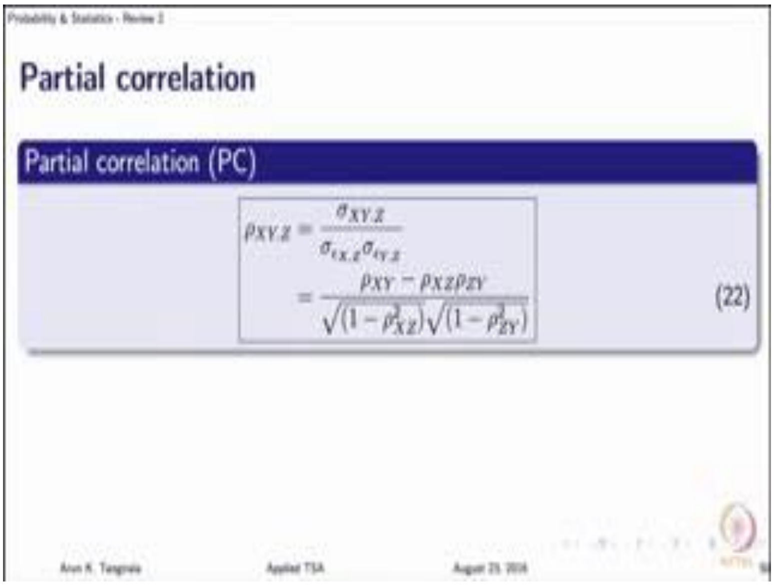
Lecture - 19
Lecture 09A - Probability and Statistics Review (Part 2)-9

Very good morning, that is what my PhD student used to tell me that, my first PhD student, even if he says, if he greets me the first time in the day, even if it is 7 o'clock in the evening, he would say good morning sir. So, I would ask him, what is the logic? He says just got nothing to do with the time of the day it is got to do whether you made the first time. So, I am meeting you for the first time today, so good morning.

I am sure you all missed the lectures and obviously, some memory loss. So, we will kind of recap where we were, we were discussing correlation and in fact, we discussed quite at length. Correlation can be viewed in many different ways and that is one quantity that we should be all comfortable with. You can look upon correlation as a measure of linear dependency which is the most important viewpoint that you want to hold on to, then you can look at it therefore, from a regression viewpoint you can also look at correlation being a measure of how much one variable explains a variability in the other and then of course, some mathematical relations and some correlation, measures the dependency in both directions and so on.

So, you should get used to different viewpoints of correlation and there are a few others also, but we do not really worried about it, we have discussed more or less the most important aspects of correlation and we also pointed out an important limitation which is that of confounding. So, we said that the amount of association, linear association that correlation measures could be through a mediator variable and we want to if the measurements of this mediating or confounding variable; sometimes its variables and available then we would like to discount the effects of those confounding variables are construct what is known as a conditioned or a partial correlation and we went through the derivation of partial covariance from which you derive the expression for the partial correlation as you see on the screen.

(Refer Slide Time: 02:37)



Of course, we have derived this expression assuming that the confounding variable is just one variable you can go through a similar process to derive expressions for partial covariance and hence partial correlation when you have more than one confounding variable.

(Refer Slide Time: 02:58)

For example if you recall we had partial covariance expression given by this sigma X Y minus sigma X Z times sigma Y Z divided by sigma square Z is that right. Now when you have more than one confounding variable any idea how this expression is modified, some intuition.

Student: (Refer Time: 03:29).

Now, yeah minus (Refer Time: 03:31) this term remains unchanged, sigma X Z times.

Student: (Refer Time: 03:38).

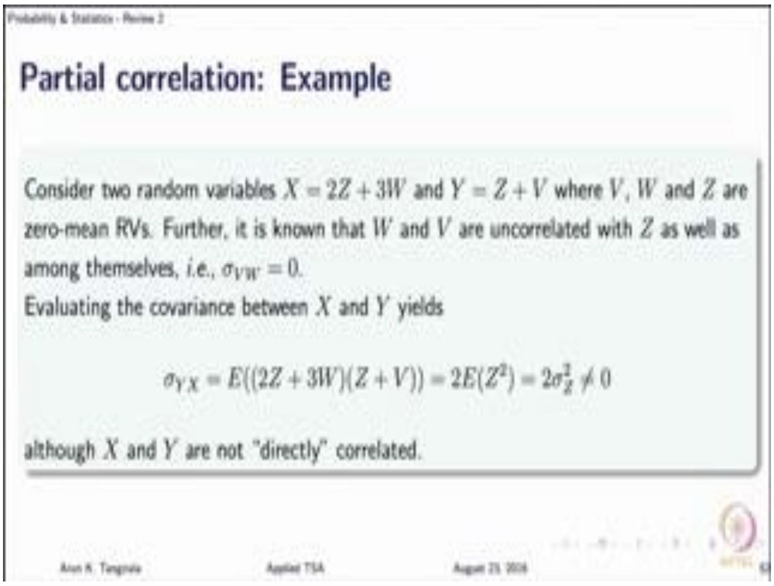
So, if Z was a vector instead then this term here would be modified to sigma X Z times sorry, sigma Z inverse times, does it make sense? Have I written things correctly or is there something I missed here? So, what is it?

Student: (Refer Time: 04:10).

So, this is also a vector now. So, let us indicate that this is a vector we already recognize this to be a matrix and we are just constructing intuitively here, assuming that this vector is a column vector right I mean in the sense sigma Y Z it is a column vector generally by vectors we mean column vectors and you can derive this expression is it correct now, at least dimensionally things look and this is indeed the expression when you have more than one confounding variable, we will use this in deriving for example, partial auto correlation functions and so on, so much for the derivation of the partial correlation. Let us go through a few more interpretations of partial correlation, we worked out an example if you recall in the last class using the fire data and we used the peak or; is a peak or package to do that. Now let us come back to partial correlation ask if we can give similar interpretations as we did for correlation.

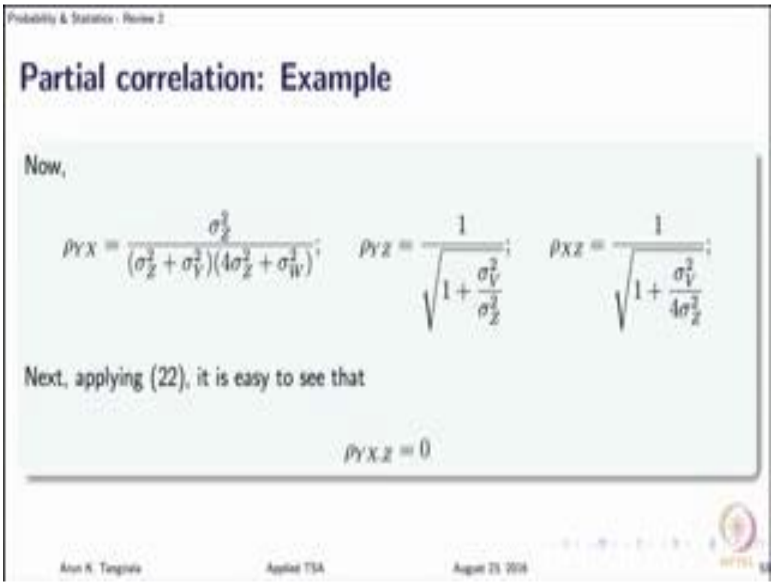
Now, one of the interpretations that one remembers always about partial correlation is that it is a correlation or it involves analysis in so called inverse domain, what we mean by inverse domain will become clear shortly, but you can think of this inverse domain always being kind of you know from a source separation viewpoint its d mixing viewpoint and so on but will not get into that, anyway let us wait until things become clear as to what we mean by inverse domain. In the last class we also had a graphical understanding of what this partial correlation is doing essentially it is severing of all the links between the X and Y that is a variables and the analysis and the confounding variable and then examining if there is a direct influence or you can say so called direct edge between X and Y.

(Refer Slide Time: 06:31)



So, to recap we went through this example where we had X and Y being expressed as a 2Z plus 3W and Z plus V, so Z this only confounding variable. When you compute the theoretical unconditional covariance its nonzero simply because Z is a mediating variable V and W are uncorrelated. If Z was given then I would compute the partial correlation to figure out if X and Y are truly related in a direct sense. And it turns out that they are not which is correct for this example.

(Refer Slide Time: 06:57)



Now in general is it true that the partial correlation will always turn out to be 0 when there is no direct connection at all. Well, the answer is yes and no it depends on how you look at it, look at this expression here, alright if you look at this expression carefully on the left hand side you have partial covariance on the right hand side the first term is your unconditional covariance and then of course, this term discounting for the effects of Z on X and Y suppose it turns out and it its quite likely that for some random phenomenon the covariance is 0; that means, you find the unconditional covariance to be 0.

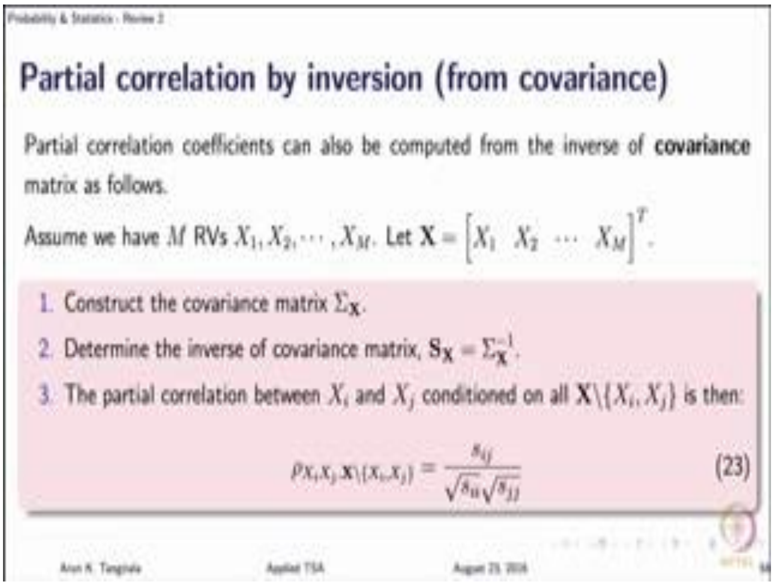
Then it is very likely when you look at the expression there that the partial covariance can be nonzero because of the presence of the second term right, indicating that your X and Y are directly related. Now that may seem a bit counterintuitive, why is it that?X and Y appear to be uncorrelated at all whereas, the moment I bring in a confounding variable X and Y appear to be directly related, but there are several examples that can be given we are not so, much interested in it, I am just discussing this example for the sake of completeness.

Suppose I find that 2 students are using completely different methods to solve a problem there is absolutely no correlation between them at all, but I bring in Z who is the instructor or the teacher who has actually taught these 2 methods in the class they are completely different methods, now you find that in presence of this teacher given that these 2 students are being taught by the same teacher you can then probably establish a correlation between them, there are numerous examples that can be given and in all such situations we say that Z is the so called suppressor variable in the sense when Z is not given X and Y appear to be uncorrelated and when Z is given they are they turn out to be directly correlated which is of course, in contrast to the example that we are looking at here. In this example in the absence of Z X and Y appear to be correlated, but the moment Z is given the correlation between X and Z vanishes.

So, there are all kinds of things that can happen when a confounding variable or a mediating variable is present. There are even more complicated situations, but we do not need to worry about that. You can actually go into the literature and read a bit more, but the fact that you should remember is your partial covariance can be nonzero even as the covariance is 0 and in any case you can say that the partial covariance or partial a partial correlation is a measure of direct association with direct being under some quotes, we will not go further on that.

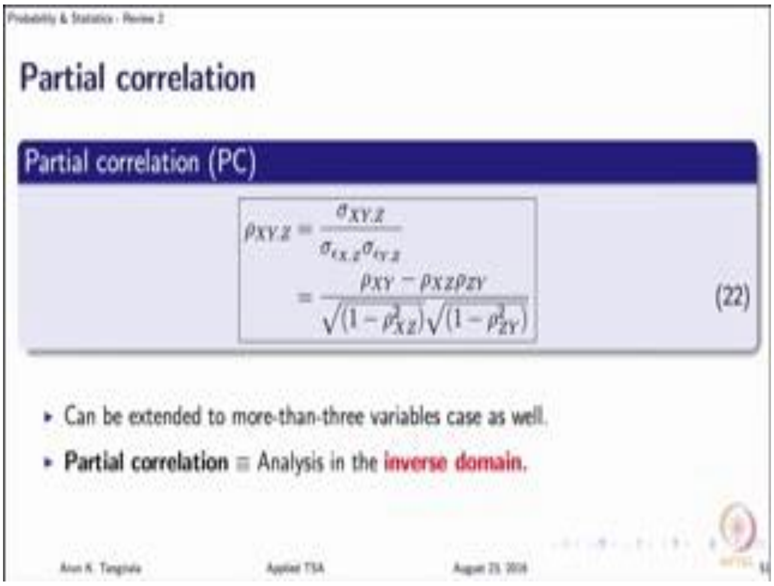
Let me talk now about the inverse domain that we mentioned earlier we said that computing power shell correlation amounts to performing analysis in the inverse domain. For now we will understand what that means, using this relation between partial correlation and covariance as shown on the slide.

(Refer Slide Time: 10:42)



We have learned how to compute partial correlation through the standard expression that we saw earlier using this expression.

(Refer Slide Time: 10:50)



But there is now an alternative way of doing that using the relation as I said equivalence between partial analysis and inverse domain analysis and the relation is as follows. You construct the covariance matrix between the 3 variables for example, in the previous example we had X Y and Z.

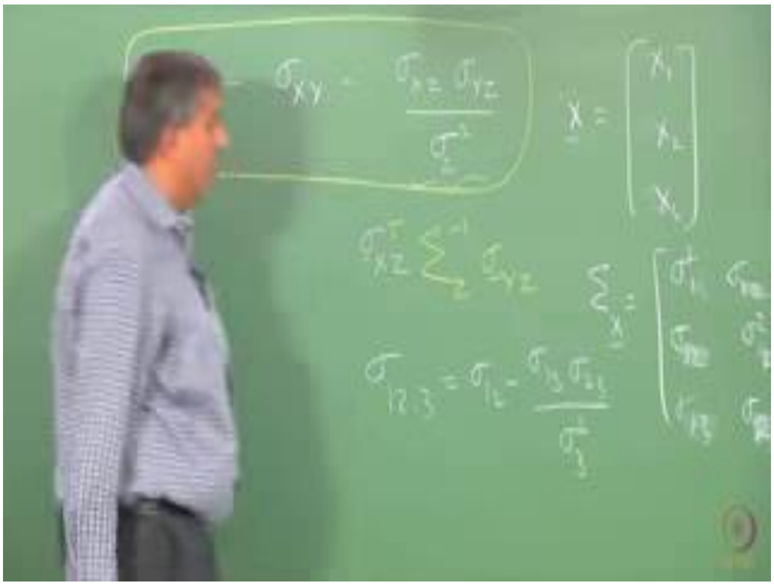
(Refer Slide Time: 11:15)

These are the vector of random variables that we looked at, club them all together in maybe a single variable vector of random variables call this as the big X I apologize for the confusion the notation, but club them all together as a vector of random variables.

Now construct your covariance matrix right, how does the covariance matrix look like? Sigma X would have sigma square X, sigma square Y, sigma square Z and then you would have sigma X Y sigma X Z and then you would have here sigma Y Z of course, it is a symmetric matrix, so you would be able to write it this way, Y Z. Of course, you do not like this confusion of notation you can replace this with X 1, X 2, X 3. That is your covariance matrix and then you invert this covariance matrix.

For example, let us say I want to compute this sigma X Y dot Z, the first thing that you do is you inverse is covariance matrix. Assume that the inverse exists which will generally exist unless under some very special cases I presume you know how to compute inverse of 3 by 3 matrices and call that as S; the big S. Now to compute the partial correlation between let us say to stick to the notation that have given on the slide you can actually call this as X 2, X 3 if you. So, I can replace here with 1 2, 1 3; notice that when we use a square, we do not use a double subscript that is a standard notation, that is used with subscripts. So, here, as well you have 1 3 when you have sigma 2 3 and then you have here sigma 2 3. So, now, if you want to compute the partial correlation between 1 and 2 conditioned on 3, my P S have asked me to be well behaved at the board I said please do not go far away becomes very difficult for the students in particular in MSP 241.

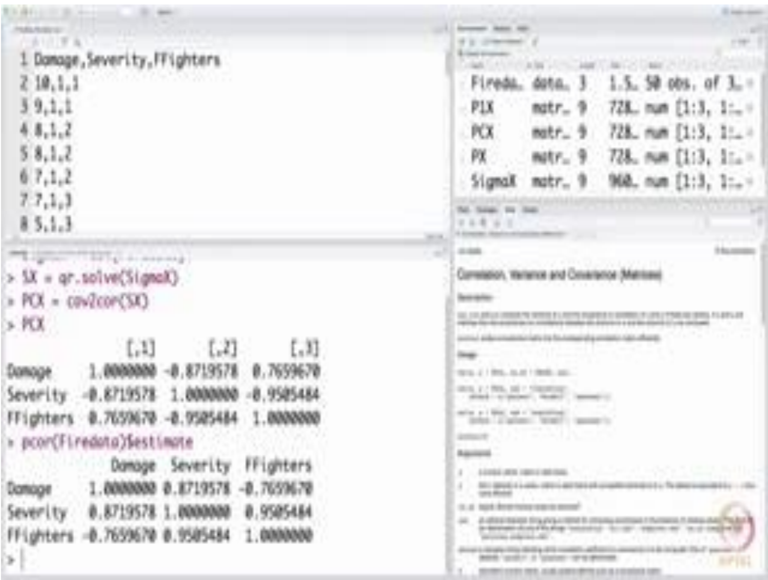
(Refer Slide Time: 14:01)



I am trying to be good boy here. So, sigma 1 2 dot 3 is the partial covariance between 1 and 2 conditioned on 3, if you have to use this formula you would obtain here sigma 1 2 minus here sigma 1 3 times sigma 2 3 by sigma square 3.

Do you see that? In fact, you can see here that when I look at this element here the minor of this element, what would I get? I would get sigma 1 2 time sigma square 3 minus the product of these 2. In fact, that is what you see in the numerator here of course, we go a step further to compute the partial correlation, we are talking of partial correlation, here straight away we are not talking of partial covariances, we are talking of partial correlations. So, the partial correlation now between 1 and 2 conditioned on 3 is first obtained by inverting this and then taking the particular element that you are looking at and then standardizing that with the corresponding partial, the diagonal elements square root of the diagonal elements that is what your equation 23 tells you, as an example here if I were to go to r, I can do that.

(Refer Slide Time: 15:31)



Let me actually again read the fire data which is already there, this is the same data set that we used in the previous lecture remember we computed partial correlation using the package; I can do it using without the package also, using this relation here. So, the first thing that I would do is compute sigma X using the standard cov routine and then let us actually compute here the inverse and in r inverses are computed using this q r dots all; are you able to see?

Student: (Refer Time: 16:12).

Then you should tell me, the benefit of doubt always goes to me if you do not speak that is the default for any teacher, if you do not speak I do not know if you know that, silence means agreement. So, please do ask like somebody asked sir is this a makeup class.

Let us change that, does this help, yes or no?

Student: Yes.

Thank you, anybody has difficulty, I can increase the font size for them which means the font I mean almost the font will come to you Q R dots solve sigma X. So, this is my inverse, now to compute the partial correlation. In fact, to compute the correlation how do you compute the correlation matrix in general keep aside partial correlation if I give you covariance matrix how do you compute correlation. Suppose I would I were to ask you to compute the correlation forget about partial correlation for this vector X, given this covariance matrix how would you do it?

Going from here to correlation matrix just correlation no partial correlation what would you do? You would actually normalize this matrix right, what would you normalize with? You pick the diagonals, take the diagonal square root of the diagonals and pre and post multiply with the inverse of those. So, you pick the diagonals of this covariance matrix and then take the square root of that and scale this covariance matrix such that you pre and post one to play with that and that is what is called covariance to correlation conversion, it is a very standard expression. Fortunately, in r you have a routine that does; that means, now we can switch off our thinking and then go to that routine now let us ask for the partial correlation.

Here what we are doing is we are actually, let me go back to the slide and explain what is happening. You can see step 3 is nothing, but actually converting a covariance to correlation and of conversion exactly that. So, you can use that same cov to cor, but not on sigma X, but on inverse of sigma X to compute the partial correlation. So, we use that here and I have here sorry SX and that is it. So, now, I have my partial correlations with me, how do I know it is correct? Well let us ask the package, assuming that the package has been well tested and so on. So, we can ask for the partial correlation estimate and you see it is exactly that.

Any questions, I am sorry, I am sorry

Student: Negative sign is (Refer Time: 19:34).

Sorry, sorry, right there is a negative sign issue sorry I think that also in the formula there should be. So, there is a sign reversal issue and I will just fix that, thank you I will do that. In fact, the formula is actually correctly specified in the next slide there is a negative sign here. So, I will go back and correct this slide; so that negative sign comes about because when you take the minus here remember the there is a sign associated with the minor of this element and there is a negative sign associate when you are actually computing inverses.

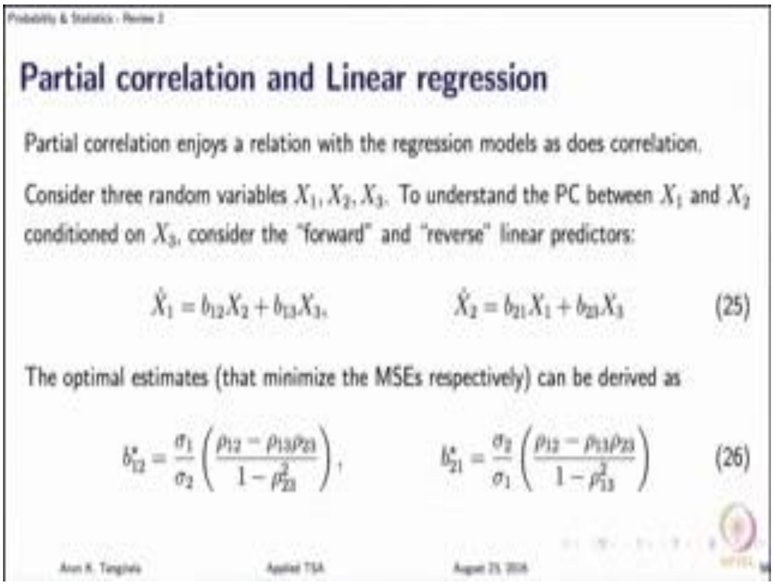
You would, because the minor of this would be minus of sigma 1 to time sigma square 3 minus a product of this whereas, the partial covariance is without the negative sign. So, you have to adjust for that account for that negative sign and then I will fix that. So, this is how you can compute partial correlations without the inversion and that is probably it is not the best way because you are computing inverses, there are more efficient computationally efficient ways of computing the partial correlations and more computationally efficient way of doing it is not exactly doing the inversion as we did, but using what is known SQR factorization which is a computationally efficient way of computing in verses we have just used. In fact, the QR dots all exactly does that; if you have to do by hand and try to implement exactly the inverse as is in the computer that is not numerically or computationally efficient and robust.

The QR dots all is actually doing a computationally efficient and a numerically robust method a way of inverting the matrix, fine. So, there is a proof which I am avoiding here there is a proof to show that the partial correlation is indeed the way we have computed; that means, you invert and then you get the relation partial correlation. There are several ways of proving this, but I am not going to prove in any one way either. The proof will gradually become clear when we understand now and a bit later as to how partial measures are computed particularly covariances or correlations are computed through linear regression. So, once linear regression becomes clear you can develop your own proof it is a fairly straightforward proof, all right.

So, you can also compute partial correlation by working with the correlation matrix, it is the same story except that instead of inverting the covariance matrix you invert the correlation matrix and then you account for the sign, it does not make a big difference now. In fact, maybe it is better to actually work with the correlation matrix sometimes for numerical reasons and so on, but and the end theoretically they are the same. So, I am going to skip this method, everything is the same except that you replace a covariance matrix with the correlation matrix.

So, before we conclude this discussion on partial correlation like we did for correlation there is a relationship between the computation or the concept of partial correlation and linear regression. In the case of correlation we showed that calculating correlation essentially is an essential or a critical part of linear regression remember the optimal estimate of the coefficients in the regression model depends on the correlation and that correlation captures effects in both directions, the so called forward and the reverse models.

(Refer Slide Time: 23:26)

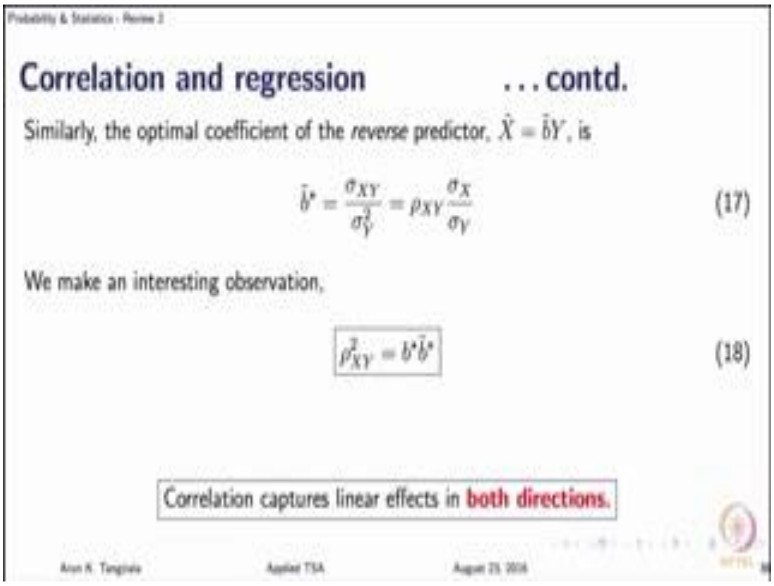


Likewise here suppose I have 3 variables X 1 X 2 and X 3 and I want to compute the partial correlation between X 1 and X 2 conditioned on X 3, same story. Once again we consider the forward and reverse models, what do we mean by forward here? Predict X 1 using X 2 and X 3 and if you want to call the other model as X predict X 2 using X 1 and X 3 then again you can come up with the optimal estimates like we did in the univariate case you can obtain the optimal estimates of the coefficients b 12 and b 21 and I have given those expressions for you, I do not spend time in the classes, but deriving this, but it is fairly straightforward the same story you minimize expectation of X 2 minus X 2 hat square or X 1 minus X 1 hat square as the case may be and then you arrive at these expressions.

You see that when you multiply straight away you see that when you multiply both these coefficients, which coefficients are we looking at? The relation between X 1 and X 2 with the conditioning variable still being X 3 and then X 2 and X 1 with the conditional variable being still being X 3 when you take the product of those 2 then you can straightaway see that you get this squared partial correlation.

Do you see that? When I multiply b 12 star with b 21 star, the factors in front of the brackets cancel out and we are left with squared partial correlation. It is very much analogous to the relation that we had derived for correlation if you recall.

(Refer Slide Time: 25:23)



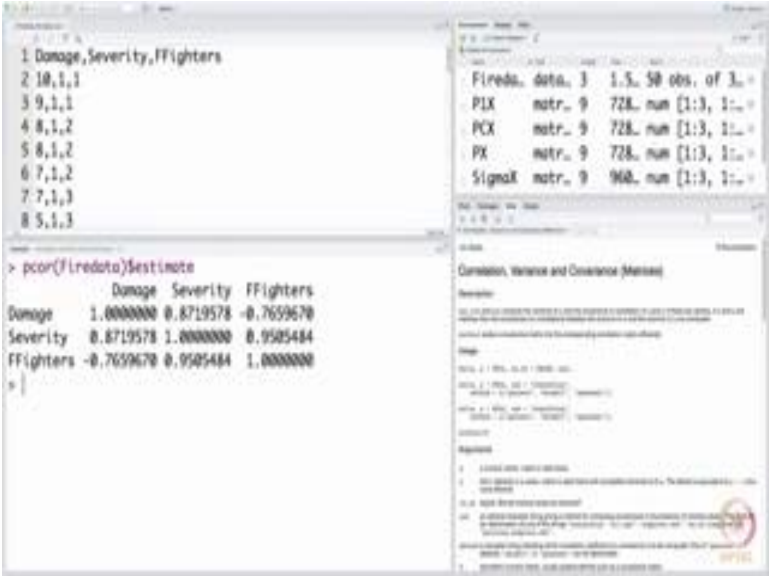
It is exactly the relation that we derived for the correlation where we considered only 2 variables, we built the forward and reverse models and showed that squared correlation is a product of the coefficients in the forward and reverse directions. The only difference is now we are conditioning on X 3, that is all, now somebody pointed out, one of the students pointed out at the end of the previous lecture that all that you are doing is regression; yes, that is true and that is exactly the point correlation in essence is no different from linear regression.

But the concepts are kind of different I mean in the in the sense that in linear regression you are building a model in correlation you are only looking, you are not looking at a model, but underneath there is a model and that is a message that you should carry with you whenever you are computing correlation do not think that it is a model free kind of measure it means that you are building a linear model, except that explicitly you do not see that linear model appearing on your paper on the board. When we define covariance or correlation, we just work with expectations and when we work with linear regressions we write out an equation explicitly, but these relations tell us clearly that computing correlation amounts to carrying out linear regression and vice versa performing linear regression amounts to computing correlations.

Linear regression explicitly tells you how much each variable is contributing to the predicted variable that is a very or the dependent variable. So, that is very important. Many people have the notion that working with correlations means that I am not really working with models; that is not correct, it is not a model free kind of measure.

This is something to remember and there are a few other relations also that you can derive for partial correlation, but that is not so much of interest to us, it is pretty typically of interest in a statistics course or in a psychometric scores or some other matrix code statistics based matrix scores and so on. For us whatever we have learnt suffices and once again you should remember partial correlation is after all a correlation measure therefore, it does not know the direction of the effect and it cannot tell you the physics for example, here in the fire data example that we solved.

(Refer Slide Time: 28:02)



Let me actually clear the screen and bring up this. So, we noticed that there is a negative correlation between the damage and the number of firefighters conditioned on severity which makes sense more the firefighters lesser the damage and someone had pointed out look there is a positive correlation between severity and firefighters, what does it mean? More firefighters means more severity, more severity will call for more firefighters, but you can very quickly come to that interpretation that more firefighters can mean more severity, but that is wrong because we do not know which is a cause which is a effect either we have to rely on physics that is our understanding of the phenomenon or turn to more advanced measures of causality and so on which can tell you even without the physics they can sometimes tell you which is a cause which is a effect, but here it is fairly clear. So, the positive it says for a given damage, there is a very strong correlation between the number of firefighters and severity, but it is a severity which is a cause and firefighters which is the effect it is not the other way round.

One has to be able to still figure out what is the cause, what is the effect, typical in many of the processes we do know these measures are used to assess the extent of influence, sometimes we do not know if there exist an influence, again we do ask turn to these measures, Now one of the things that very important things that we are not doing is assessing so called significance like we talked about significance test for correlation when I showed you the linear regression, we talked about the null hypothesis that a particular coefficient this estimate is significant and so on, we are not talking about that, but that is exactly what hypothesis testing talks about and this is not the moment to talk about you can go through the videos. When we discuss estimation talk about estimation theory we will discuss this so called significance of the estimates at that time.