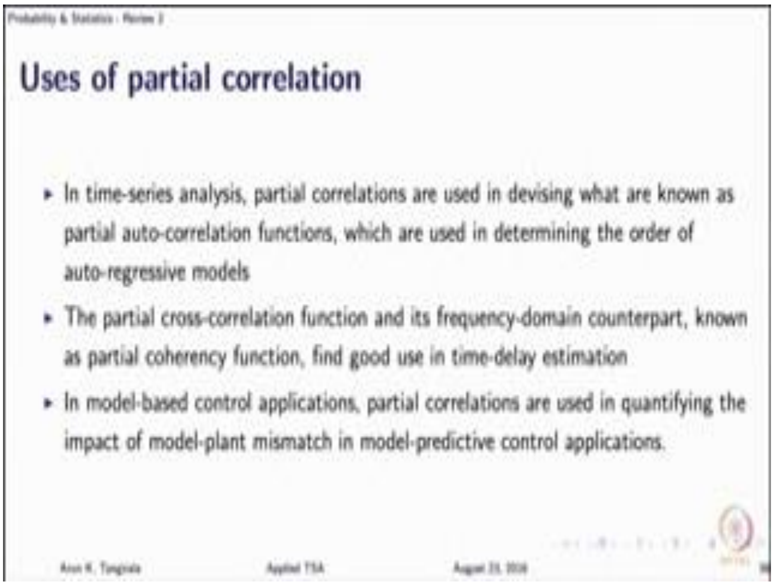
Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 20
Lecture 09B - Probability and Statistics Review (Part 2)-10

Getting back to our discussion, where do we use partial correlation? We used very widely in time series we use this to compute partial auto correlation functions which we shall learn shortly.

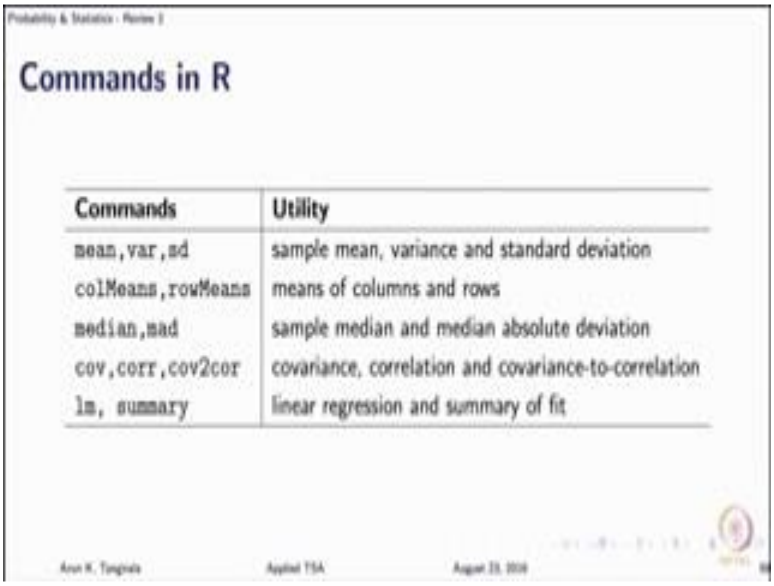
(Refer Slide Time: 00:24)



As I said in many control applications or wherever you have multivariable processes partial measures are used to analyze the effects between a pair of variables, conditioned on all the other variables, so essentially to isolate the effect in a particular channel twice to quantify the strength of a particular channel, regardless of all the other channels.

There are frequency domain versions of correlation partial correlation and so on, but it is too early to talk about it as I keep approving 1 or 2 every day. So, if I start talking of frequency domain now then I am afraid that I will have to scroll several pages. So, let us not get into that.

(Refer Slide Time: 01:14)

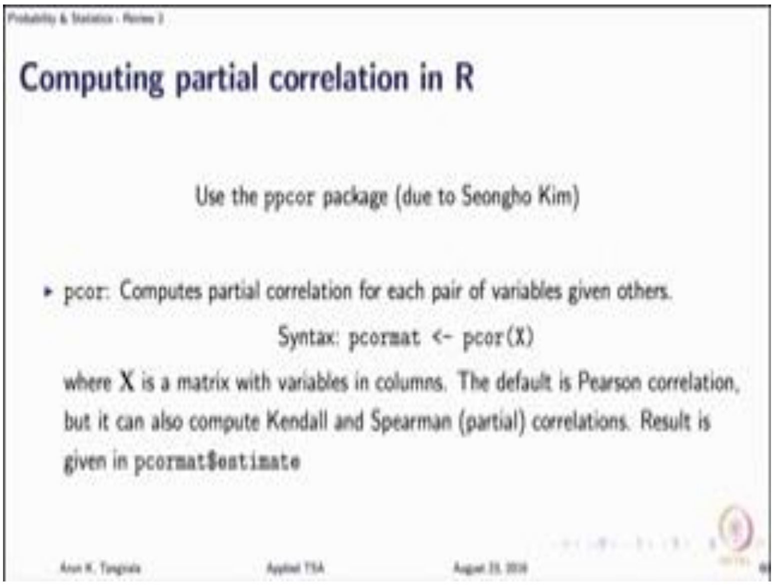


Let me conclude this discussion on partial correlation with just a quick summary of the commands relevant commands in our, that will allow you to compute sample version; that means, estimates as you know mean, var and sd, they all give you estimates of the mean variance and standard deviation.

When you have you know many variables then you can use and you assemble them in a matrix then you can use call means or row means the depending on how you arrange your data and then you have median and mad, it is mad, but let us actually talk read it as mad, mad stands for median absolute deviation.

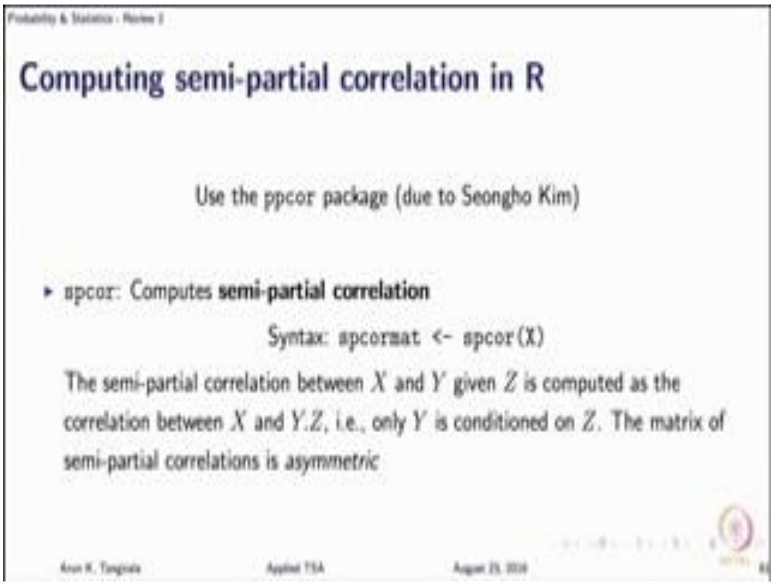
Now, median is a robust measure or a robust estimate of the mean it is still an estimate only computes an estimate there is a theoretical definition of median, we know what it is. You can use this median to estimate the mean and you can use mad to estimate the standard deviation; sd is one way of estimating the standard deviation, mad is another way of estimating the standard deviation, both are estimating the same theoretical one and of course, we have used cov, corr and cov2cor and I think it should be cor not probably cor and then I have shown you already how to use lm for performing linear regression and summary to display the summary of the regression that you perform.

(Refer Slide Time: 02:48)



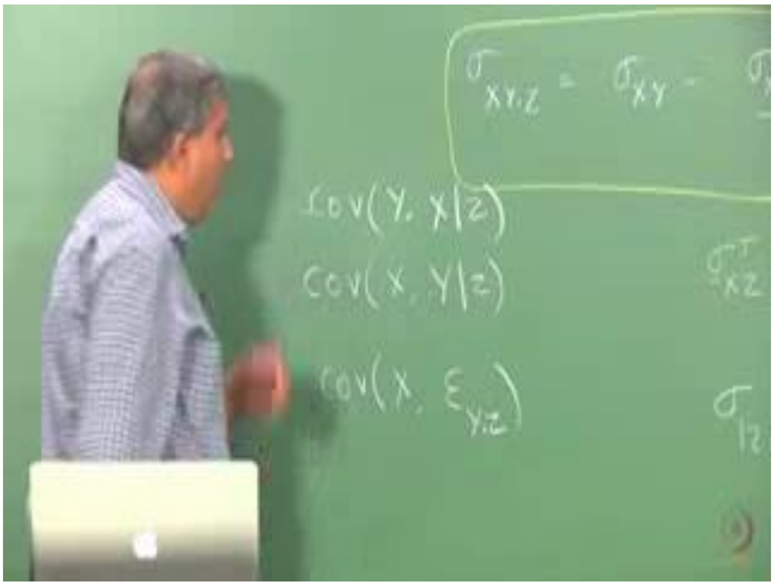
Now, we have also illustrated to you the how to compute partial correlations in R using this ppcor package.

(Refer Slide Time: 02:57)



And finally, there is also the semi partial correlation where you would not you condition only one variable on the other one. So, we have here in partial covariance, where conditioning both X and Y on Z.

(Refer Slide Time: 03:12)



But in semi partial correlation, you would compute the covariance between X and Y dot Z or Y given Z, in other words in terms of the residuals you would compute covariance between X in X and epsilon Y dot Z, this would be the semi partial correlation between X and conditioned Y. Likewise you can compute semi partial correlation between Y and conditioned X.

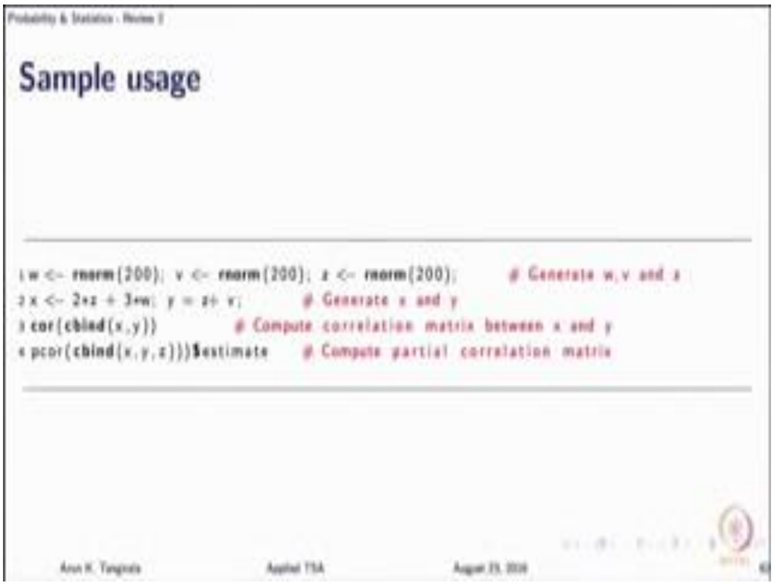
Obviously, when you ultimately compute the correlations, they turn out to be asymmetric; you should expect them to be asymmetric and the package that is given to you computes that. You may wonder why one computes semi partial, partial and so on. Well there are reasons depending on what you want to do we do not use semi partial correlations in our time series analysis at all, but if you want to know more you can read, but in all of the literature as you read one thing that will help you understand the difference between semi partial and partial correlation is this perspective of correlation measuring the amount of variability one variable explains about the other.

If I think of Y as a dependent variable and X as the independent variable the correlation between Y and X will tell me how much variability in Y, X is able to explain the conditions or the partial correlation is also an explanation of variability, but after discounting the effects of Z on both Y and X. So, it tells us what is unique information that X contains about Y that cannot be explained by any other variable.

Whereas the semi partial correlation in suppose you look at covariance between for example, Y and X given Z, will tell me how much X can that is when you discount the effects of Z, whatever is left out how much that can uniquely explain, that can explain the entire variability in Y, it depends on what you want to do, do not worry about where in the sense whether we will use it in time since we do not, but maybe it will be of interest to some of you depending on the kind of analysis that you are doing. These are used quite a lot in social data analysis in sometimes in economic data analysis and so on we do not we generally work with partial correlation functions.

The package that you are looking at allows you to compute semi partial correlation. In fact, if you look at the expression for semi partial correlation, you will find that for example, the semi partial correlation between Y and conditioned X will exactly looks the same, this here is the same there is no difference, but the correlation looks different. The semi partial covariance looks a same, but the correlation would look different because of the way you are normalizing that is all. In that sense the semi partials and partials are different and; obviously, you should expect semi partial correlation to be asymmetric that is; so I think that there should be good enough, let us not spend more time on this and you know this is some of the ways that I have already demonstrated more than the sample usage. There is no point in going over this set of commands, so that is it.

(Refer Slide Time: 06:50)



So, we now draw our curtains on the review of random variables, covariances, expectations, conditional expectations, independence and so on.