Marketing Research and Analysis - II

Prof. Jogendra Kumar Nayak
Department of Management Studies
Indian Institute of Technology Roorkee

Lecture – 43
Partial & Multiple Correlation

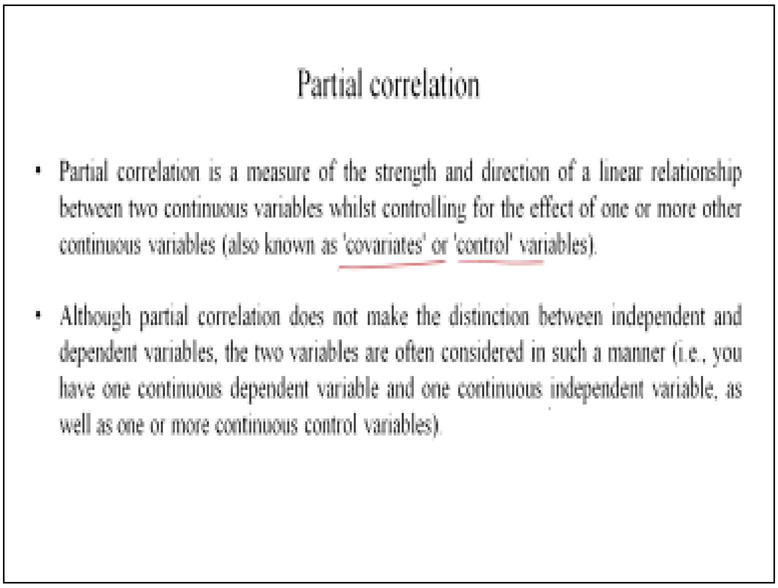
Hi Friends, I welcome you all to the course marketing research and analysis. Today, we will be continuing from our last lecture, so in the last lecture we were discussing about covariance and correlation. Where we understood the difference and similarity between the 2. So, we realise that covariance and correlation is basically more or less same but correlation being a standardized version of the covariance and it has got an advantage over covariance that it not only talks about the direction of a relationship but also the intensity of the relationship.

So, correlation has an advantage over the covariance and you can compare it with any number of variables. But sometimes comparison in a covariance case become difficult because the units of the variables will be different in nature, as I had given mentioned the case about let us say height and weight, where height is measured in centimetre and the weight is measured in kg. So, centimetre, kg makes no sense. Similarly height and years, centimetre and age of the person, so centimetre and years makes a very odd kind of combination.

So, correlation rather what it does is? It is the ratio between the covariance and the standard deviation of the X and the Y variable. So, it lies between the -1 and the +1 and anything that you do that can be compared against each other so that is the advantage. But today, we will talk about the little higher forms of correlation, which is very, very important because in real life it so happens that 2 variables are correlated but the presence of a third variable could effect the correlation.

So, similarly we need to may be sometimes there is the situation where we need to suppress the effect of one variable and find out the relationship between 2 other variables. So, this case is called a case of a partial correlation and when you have more than 2 variables you do a may be a partial correlation or a multiple correlation assuming all 3 are important. So, you take the multiple correlation or you want to suppress one, let us say X,Y, Z so you suppress Z and see the effect of X and Y, so that case it is a case of partial correlation. So, let us see what is clearly it is.

(Refer Slide Time: 03:00)

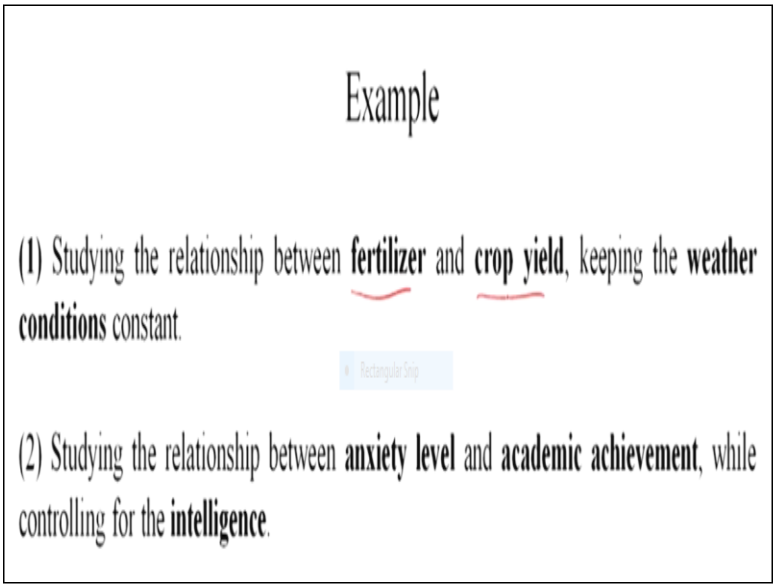


Partial correlation as it is says, it is a measure of the strength and direction of a linear relationship between 2 continuous variables, while controlling for the effect of one or more other continuous variable’s known as covariates or the control variables. Although partial correlation does not make the distinction between independent and dependent variable, although it is does not make because in correlation, we should not say but going by the spirit of the study there is actually dependent and independent variable.

The 2 variables are often considered in such a manner, that is you have one continuous dependent variable and one continuous independent variables as well as one or more continues control variables, one more reason behind this is because you see correlation and regression are very strongly connected, these 2 topics are very strongly connected.

So, when you do regression automatically we will see that we talk about the independent and dependent variable. So, that is why when you talk about correlation also in spirit you do not talk about it, but then actually there is dependent and independent variable which you understand but you do not need to mention them because we are only try to find out the relationship.

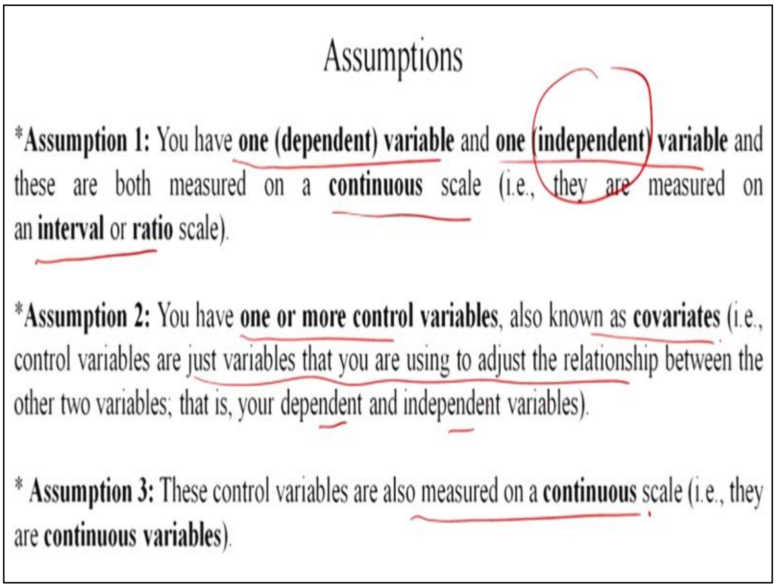
(Refer Slide Time: 04:17)



Say let us see in this case I want to study the relationship between fertilizer and the yield how much crop is produced but I am keeping the weather condition constant, so assuming that weather condition is constant, I am trying to see the effect of fertilizer in the crop yield. But in reality it might not be possible, because weather condition does play a role on crop yield and it is not only fertilizer. so, when I want to see this effect of the fertilizer in the crop yield keeping the weather constant that means different types of weathers they are suppressed and kept constant, so this case it is a case of a partial correlation.

Similarly the relationship between level of anxiety of student and the academic achievement what is the effect of the relationship between anxieties? The level of the anxiety of the student and the academic achievement and here we are assuming that intelligence is the controlled factor. So, we are controlling for the intelligence all students are more or less the same. So, let us now after understanding what is partial correlation and the effect of partial correlation, we need to understand the assumptions that we need to follow.

(Refer Slide Time: 05:33)

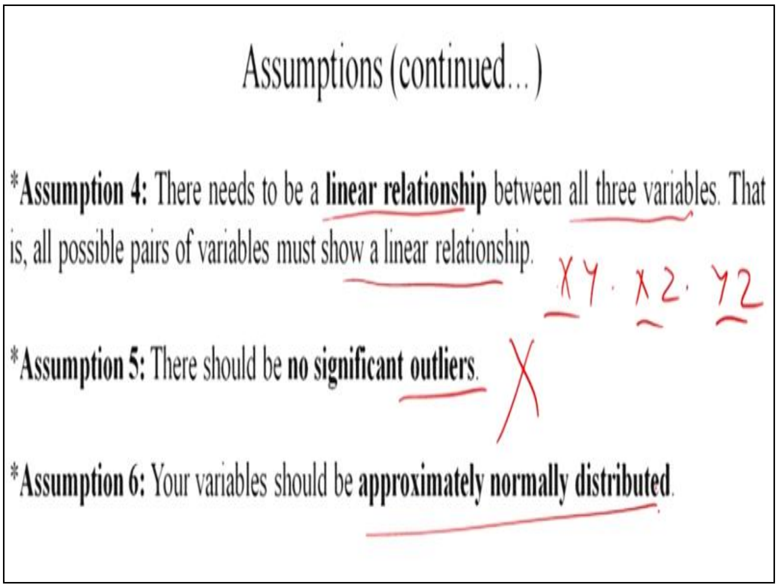


So, what are the some other assumptions, the first you have one variable which is more or less dependent variable and one which is an independent variable. And there both measured on a continuous scale, so this is I think by now you understand continuous scale which is measured as they an interval or ratio, say assumption 2 you have one or more control variables, for in the last case.

You have weather conditions may be soil quality could also be another control variable which are also known as covariates and they are, this covariates are measured in continuous again variable this is here measured on a continuous scale. This covariates basically as it is mentioned as you see are variables that you used to adjust the relationship between the 2 variables that is your dependent and the independent.

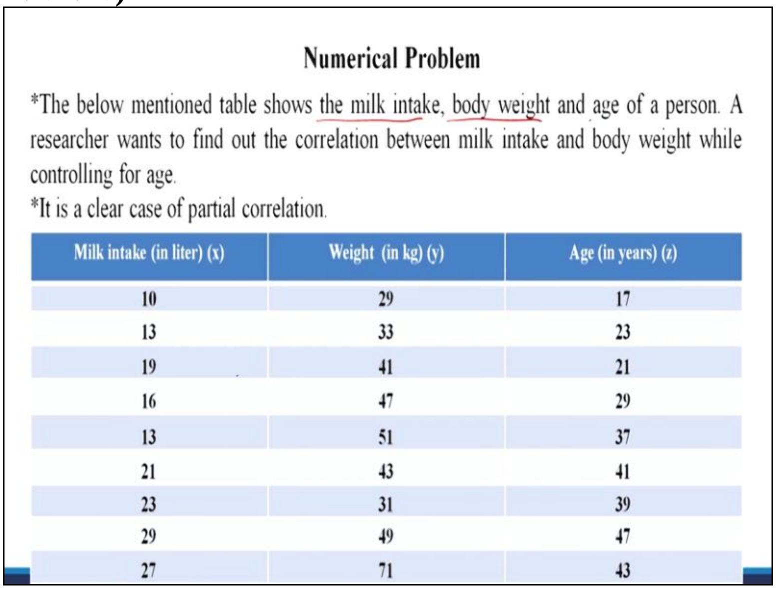
So, how it is adjusting if my whether condition changes what will happened, if my soil condition changes what will happen, so I am controlling, these are my control variables, assumption 4.

(Refer Slide Time: 06:27)



There needs to be a linear relationship between all the 3 variables the dependent, independent and the covariate that is all pairs of variables must show a linear relationship, so XY, XZ and YZ all the 3 must show a linear relationship. Outliers should not be present because we have understood by now that outliers can distort the statistical interpretations. Sixth assumption your variables should be normally distributed. So, these are the basic assumptions, which we need for any statistical or a parametric test. Let us go to a numerical problem.

(Refer Slide Time: 07:10)



Milk intake (in liter) (x)	Weight (in kg) (y)	Age (in years) (z)
10	29	17
13	33	23
19	41	21
16	47	29
13	51	37
21	43	41
23	31	39
29	49	47
27	71	43

It is says that milk intake has an effect on body weight but the age of a person has also an effect on the body weight. So, now a reseacher wants to see the relationship between intake of milk, body weight and age of a person. In the first case so what we are doing you are trying to findout the correlation between milk intake and body weight while controlling for age, just imagine this is ought to be happened because if you let us say intake milk suppose we say body weight is let us say growing, improving or increasing.

And if we take the body weight of several people and take the milk intakes, we can correlate them, we can findout some, we can draw some inference but if we omit or somehow we do not take into consideration the age of a person that would be a very wrong interpertation because the body weight of a person also changes with a growing age as you go up, as your age grows.

Generally your body weight tends to grow and untill unless you have some you know some ill some negative health effect or something you know some disease something your body weight is goes down otherwise with age the naturally the body weight is increased, so we will see how milk intake and weight are related keeping the age as a factor which needs to be controlled, so this is the value, so 10, 13, 19, 16, goes on, this the intake of milk, this is the weight of the person and this is the age of the person.

(Refer Slide Time: 08:41)

milk x	wt y	age z	$(x-\bar{x})$	$(y-\bar{y})$	$(z-\bar{z})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(z-\bar{z})^2$	$(x-\bar{x})(y-\bar{y})$	$(y-\bar{y})(z-\bar{z})$	$(z-\bar{z})(x-\bar{x})$
10	29	17	-9	-14.889	-16	81	221.679	256	134	238.222	144
13	33	23	-6	-10.889	-10	36	118.568	100	65.333	108.889	60
19	41	21	0	-2.889	-12	0	8.346	144	0	34.667	0
16	47	29	-3	3.111	-4	9	9.679	16	-9.333	-12.444	12
13	51	37	-6	7.111	4	36	50.568	16	-42.667	28.444	-24
21	43	41	2	-0.889	8	4	0.79	64	-1.778	-7.111	16
23	31	39	4	-12.889	6	16	166.123	36	-51.556	-77.333	24
29	49	47	10	5.111	14	100	26.123	196	51.111	71.556	140
27	71	43	8	27.111	10	64	735.012	100	216.889	271.111	80
171	**395**	**297**	**0**	**0**	**0**	**346**	**1336.889**	**928**	**362**	**656**	**452**

$\bar{x} = \sum x/n = 171/9 = 19$

$\bar{y} = \sum y/n = 395/9 = 43.889$

$\bar{z} = \sum z/n = 297/9 = 33$

(Refer Slide Time: 08:46)

$$r_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 * \sum(y-\bar{y})^2}} \qquad r_{yz} = \frac{\sum(y-\bar{y})(z-\bar{z})}{\sqrt{\sum(y-\bar{y})^2 * \sum(z-\bar{z})^2}} \qquad r_{zx} = \frac{\sum(z-\bar{z})(x-\bar{x})}{\sqrt{\sum(z-\bar{z})^2 * \sum(x-\bar{x})^2}}$$

$$= 362/\sqrt{346 * 1336.889} \qquad = 656/\sqrt{1336.889 * 928} \qquad = 452/\sqrt{928 * 346}$$

$$= 0.5323 \qquad = 0.589 \qquad = 0.7977$$

0

So how do you find out, so if you see these are the formulas,we are finding out 3 different formulas, so first what values we need let us see so first findout this you need a relationship between X and Y, you need to find out relationship between Y and Z and Z and X. 3 relationships can happened, so what is the formula for let us say,

$$r_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 * \sum(y-\bar{y})^2}}$$

This is for X and Y, now taking for Y and Z, XY, YZ so you see what is happening instead of here what would $(x - \bar{x})$ you are having $(y - \bar{y})$ and insead of $(y - \bar{y})$ you have a $(z - \bar{z})$ divided by instead of this you having a Y and instead of Y you having a Z similarlly you can making it for Z and X, so this will be Z, this will be X and this will be Z and this will be X other things remain the same. So, now we are taken this let us say $(x - \bar{x})$, so X is my this much, so X bar is 171 upon 1,2,3,4,5,6,7,8,9 respondents values are there so divided by 9.

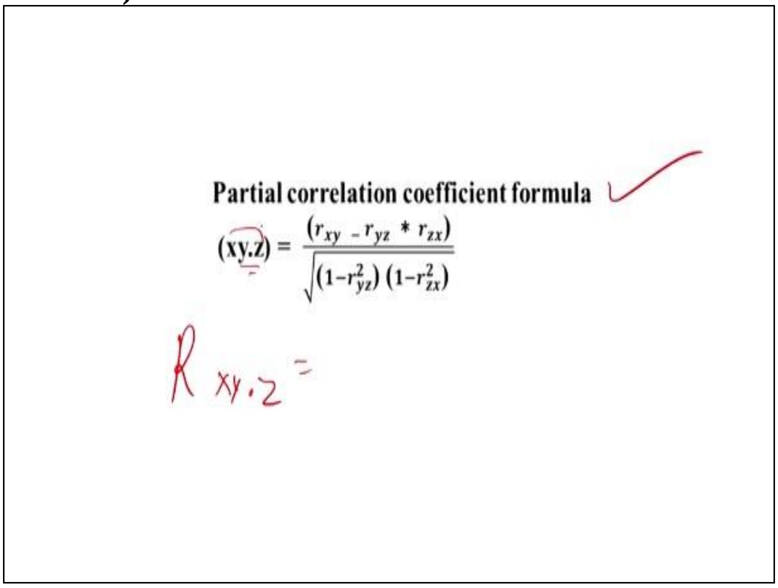
So 19 X – X bar, so 10 – 19 – 9 ,13 – 19 – 6, 19 - 19 – 0 it goes on, similarlly $(y - \bar{y})$ what is the Y bar 43.889, so we are done it, similarlly Z age, so we have seen the 297 and the mean is 33, so we have found out $(x - \bar{x})$, $(y - \bar{y})$, and $(z - \bar{z})$. Now you need the $(x - \bar{x})$ square, so $(x - \bar{x})$ square is 81, 36, and goes on and this to is total, similarlly $(y - \bar{y})$ the total is this much and $(z - \bar{z})$ the total is this much one more product we require that is look at this formula.

This $(x - \bar{x}) * (y - \bar{y})$ similarlly for others, so $(x - \bar{x}) * (y - \bar{y})$ so how much is this -9 * $(y - \bar{y})$ is 14 point wehatever it is. So, if you multiplied this it will be roughly it is 134 let us say

-6 * 10.889, so this is comes in something around 65.3 could be little more also we can 65, so this is 0, so we have found out for $(x - \bar{x}) * (y - \bar{y})$ similarlly $(y - \bar{y}) * (z - \bar{z})$ and $(z - \bar{z}) * (x - \bar{x})$ so all the required values are there with us in our hand right.

Now let us calculate the individual correlations, so for the first case the relationship between X and Y it is equal to 0.5323, relationship between the Y and Z is this much and relationship between the X and Z is this much.

(Refer Slide Time: 11:52)



So the Partial correlation coefficient formula is now, so what we are saying , we need to find out the relationship between let us say X Y controlling Z, so the formula is this much is this one now, this is equal to the

$$(xy.z) = \frac{(r_{xy} - r_{yz} * r_{zx})}{\sqrt{(1-r_{yz}^2)\,(1-r_{zx}^2)}}$$

(Refer Slide Time: 12:33)

$$P(xy.z)= \frac{(r_{xy} - r_{yz} * r_{zx.})}{\sqrt{(1-r_{yz}^2)*(1-r_{zx}^2)}}$$

$$= \frac{(0.5323 - 0.589*0.7977)}{\sqrt{(1-0.3469)(1-0.6363)}}$$

$$= \frac{(0.5323 - 0.4698)}{\sqrt{(0.6531)(0.3637)}}$$

$$= \frac{0.0625}{\sqrt{0.2375}}$$

$$= \frac{0.0625}{0.4873}$$

$$= \mathbf{0.1283}$$

So when we use this formula we will find that the partial correlation which we get is equal to by using that formula you see we have calculated here 0.1283, so this is the partial correlation, so that means when we are trying to find out the change in weight due to the intake of milk by keeping the by suppressing or by controlling the age factor we see that the partial correlation is coming .1283

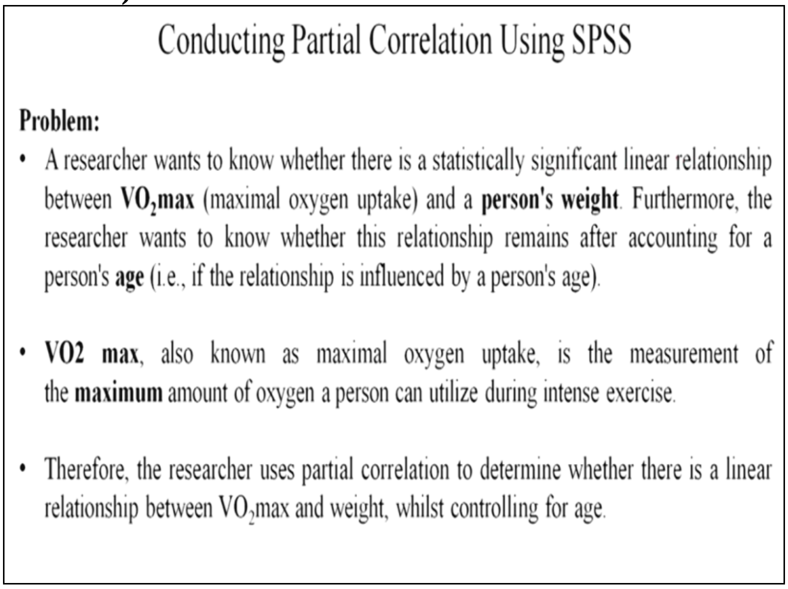
(Refer Slide Time: 13:01)

***Interpretation:** The analysis revealed that there is a moderately positive correlation between milk intake and body weight while controlling for age.

So the interpretation is that the analysis revealed that there is a moderately positive relationship correlation between milk intake and body weight while controlling for age. Now just imagine this is very important because sometimes we need to when we are looking at several relationships correlations in our own research study we need to partial out or suppress sometimes some of the variables that might impact.

But then we want to suppress it that we want to say we do not want to considered it and see the effect of the relationship between other 2 variables. So, this is the case of the partial correlation. Now how you can conduct the partial correlation in SPSS, I will show that also, so otherwise what happens many times students face the difficulty that they understand through ok now by hand how to do it? But now how to do it on the SPSS it is a problem, so this the case

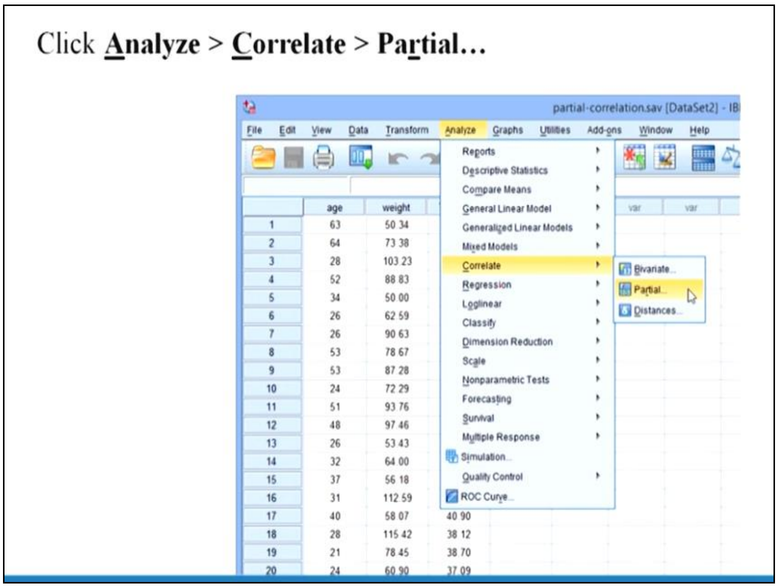
(Refer Slide Time: 13:57)



A researcher wants to know whether there is the statistical significant relationship between VO2 max, what is this VO2 max? I will explain here, maxmium oxygen uptake and a person's weight. Futher more researcher wants to know whether this relationship remains same or not after accounting for a person's age that is if this relationship is influnced by a person's age or not, now what is the VO2 max you see it is the maximum oxygen carrying capacity or it is the measurement of the maximum amount of oxygen a person can utilize during intense exercise.

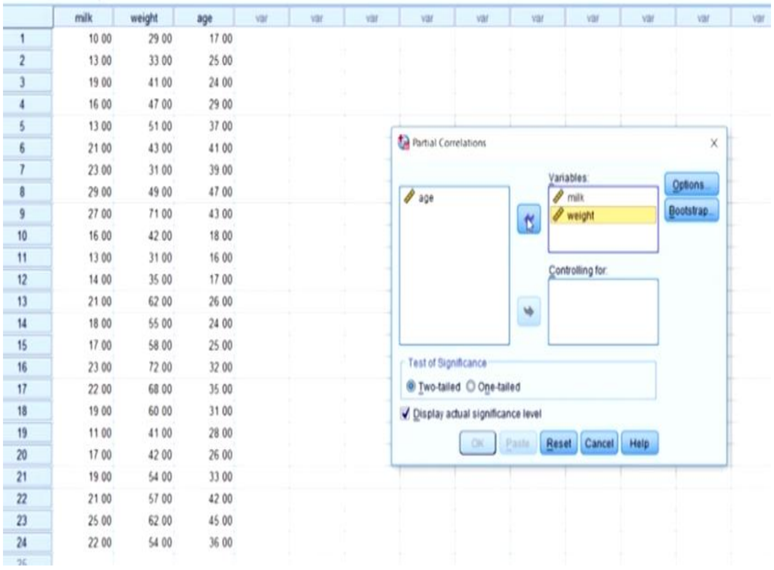
So you see the utilizing ability of the oxygen varies from person to person, it has been seen the people who lived in the mountains regions they have a higher efficiency using this oxygen.Therefore the researcher uses a partial correlation obviously we have understood now well I want to check by effect of this 2 by keeping the age as a factor controlling factor, so while controlling for age.

(Refer Slide Time: 15:00)



Now how do we do it? So there is steps are shown to you here. But I will directly go and show you on the dataset that I brought.

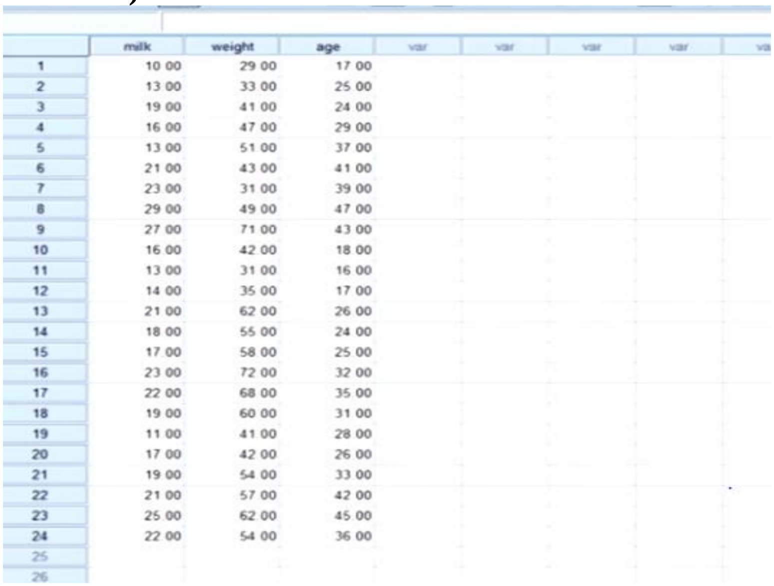
(Refer Slide Time: 15:08)



So, this is the dataset, same data that is have been talking about. Now how do I go? First I go to analyze and now I go to correlate, now you see this bivariate, partial and distance. But distance is a case which is not necessary to be discuss now, will be talk about, when we talk about cluster and all we will talk about it. But Now we want talk about bivariate and partial. So what is bivarite? When 2 variables we want to find out. But here our case is a special case where we want to findout the relationship between not 2, but the relationship between 2 keeping third as a control variable.

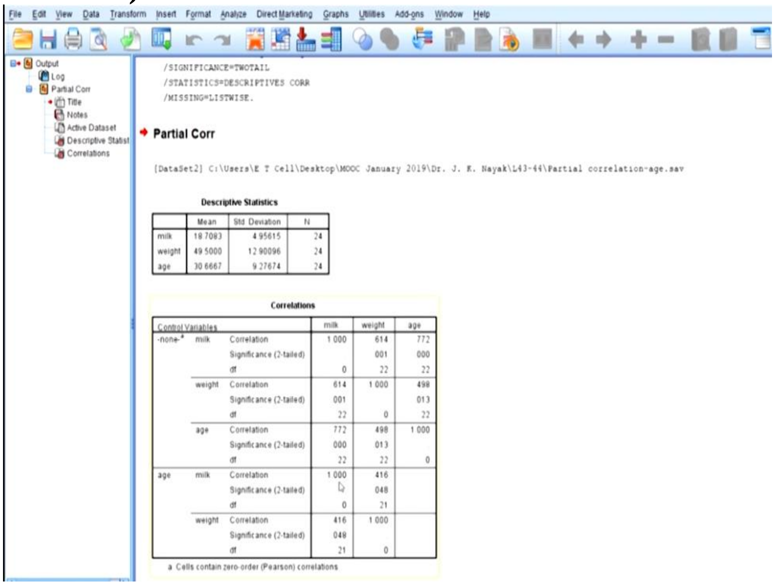
So, what is my 2 variables? Now variable is milk and weight. But what do I want to partial out? I want to partial out the controling for the age. So I want to control for the age factor. So age factor which generally we know that with age weight increses that we are taking as a control varaible. Now we are taking this 0 order correlation and mean. So these 2 we will take and see what is the effect now?

(Refer Slide Time: 16:12)



	milk	weight	age	var	var	var	var	var
1	10.00	29.00	17.00					
2	13.00	33.00	25.00					
3	19.00	41.00	24.00					
4	16.00	47.00	29.00					
5	13.00	51.00	37.00					
6	21.00	43.00	41.00					
7	23.00	31.00	39.00					
8	29.00	49.00	47.00					
9	27.00	71.00	43.00					
10	16.00	42.00	18.00					
11	13.00	31.00	16.00					
12	14.00	35.00	17.00					
13	21.00	62.00	26.00					
14	18.00	55.00	24.00					
15	17.00	58.00	25.00					
16	23.00	72.00	32.00					
17	22.00	68.00	35.00					
18	19.00	60.00	31.00					
19	11.00	41.00	28.00					
20	17.00	42.00	26.00					
21	19.00	54.00	33.00					
22	21.00	57.00	42.00					
23	25.00	62.00	45.00					
24	22.00	54.00	36.00					
25								
26								

(Refer Slide Time: 16:13)



Descriptive Statistics

	Mean	Std Deviation	N
milk	18.7083	4.95615	24
weight	49.5000	12.90096	24
age	30.6667	9.27674	24

Correlations

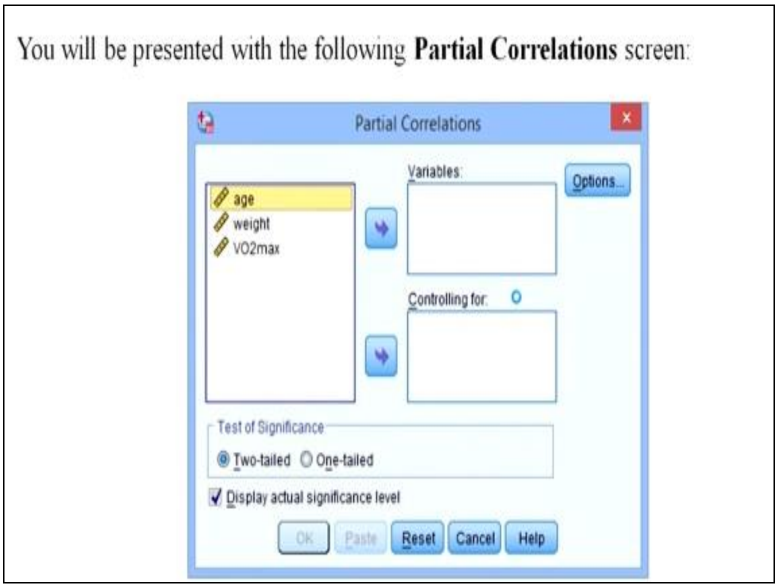
Control Variables			milk	weight	age
-none-[a]	milk	Correlation	1.000	.614	.772
		Significance (2-tailed)		.001	.000
		df	0	22	22
	weight	Correlation	.614	1.000	.498
		Significance (2-tailed)	.001		.013
		df	22	0	22
	age	Correlation	.772	.498	1.000
		Significance (2-tailed)	.000	.013	
		df	22	22	0
age	milk	Correlation	1.000	.416	
		Significance (2-tailed)		.048	
		df	0	21	
	weight	Correlation	.416	1.000	
		Significance (2-tailed)	.048		
		df	21	0	

a. Cells contain zero-order (Pearson) correlations

So let us go to the output file. 0 order correlation means understand there is no correlation any other variable you know effecting it, so age, the relationship between correlation between milk and age is how much? You see is .416 . When we take, this is a partial factor, so we are controlling this age so you want to see the realtionship between the milk and weight, milk and milk will obviously be 1, milk and weight you see is 0.416, similarly the weight and milk also will be the same because it is 0.416. Now the significance value, look at the significance now it is saying 0.048.

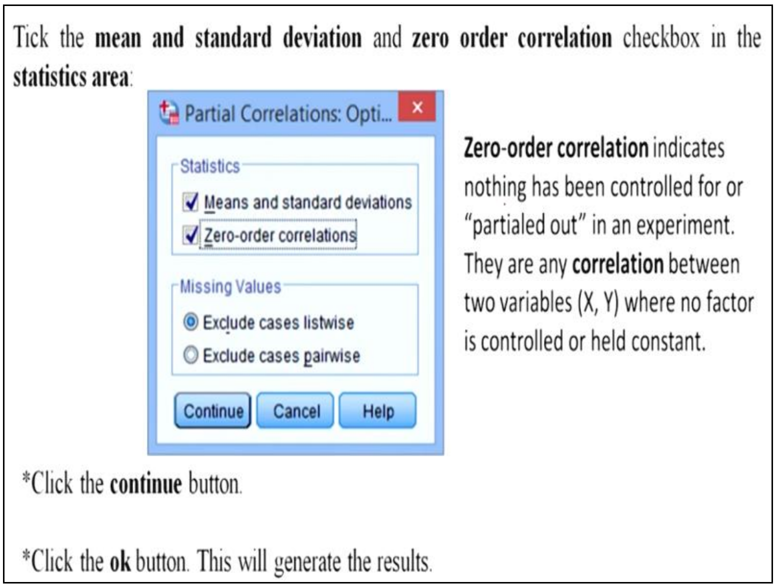
So that means what, we are saying that means it is a significant relationship, that means the relationship is when it is less than 0.05 or 5% then we say that the null hypothesis is rejected. So, the null hypothesis is that there is no effect on age, the controlling factor age on the intake of milk over the weight of a person. So, this hypothysis is now rejected and we will say, now well age does play a role, has a role in effecting the relationship between milk and weight which are positively related. So this is what actually it means.

(Refer Slide Time: 17:40)



Now let us go back to the slide and see, so this is how we are done.

(Refer Slide Time: 17:46)



So, we are taken the varibles and so I have written here, it indicates that nothing has been controlled, 0 order means nothing has controlled or partialed out in an experiment.They are any correlation between 2 variables where no factor is controlled it is a normal correlation.

(Refer Slide Time: 18:00)

Interpretation

Descriptive Statistics

	Mean	Std. Deviation	N
Weight	79.6603	15.08984	100
VO2max	43.6298	8.57131	100
Age	31.1000	9.14253	100

The descriptive statistics show that we had no missing data since the recorded sample size, *N* = 100, is the same as the number of participants that took part in the study. We can also see that the mean value of the dependent variable, VO_2max, was 43.63 ml/min/kg (with a standard deviation of 8.57 ml/min/kg), whilst the mean weight of participants was 79.7 kg (with a standard deviation of 15.1 kg), and finally, the mean age of participants was 31.1 years (with a standard deviation of 9.1 years). This suggests that the sample of participants was slightly on the younger side rather than representing the population as a whole, which is useful to know when discussing the generalizability of the findings in your report.

So, we have done it and suppose you want to write, how do you write it? Now interpretation, this is way to write. So whatever you got that is you can write that also. The descriptive statistics shows that we had no missing data, 100 cases where there and N = 100, the same number of participants that took part in the study . We can also see that the mean value of the dependent variable was 43.63 with standard deviation of 8.57.

While the mean weight of participants was 79.7 with standard deviation of 15.1. And finally the mean age of the participants was 31.1 with the standard deviation of 9.1. This suggests that the sample of participants was slightly on the younger side rather than representing the population as a whole because it is 31 which is useful to know when discussing the generalizability of findings in your report.

(Refer Slide Time: 19:01)

Interpretation (continued...)

Correlations

Control Variables			Weight	VO2max	Age
-none-[a]	Weight	Correlation	1.000	-.307	-.004
		Significance (2-tailed)	.	.002	.972
		df	0	98	98
	VO2max	Correlation	-.307	1.000	-.191
		Significance (2-tailed)	.002	.	.057
		df	98	0	98
	Age	Correlation	-.004	-.191	1.000
		Significance (2-tailed)	.972	.057	.
		df	98	98	0
Age	Weight	Correlation	1.000	-.314	
		Significance (2-tailed)	.	.002	
		df	0	97	
	VO2max	Correlation	-.314	1.000	
		Significance (2-tailed)	.002	.	
		df	97	0	

a. Cells contain zero-order (Pearson) correlations.

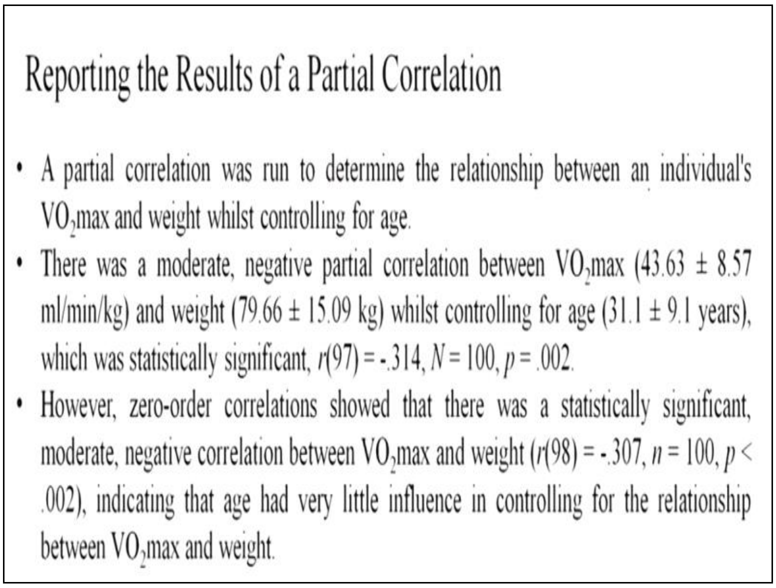
The results of the partial correlation highlighted by the red rectangle show that there was a moderate, negative partial correlation between the dependent variable, "VO_2max", and independent variable, "weight", whilst controlling for "age", which was statistically significant ($r(97) = -.314$, $n = 100$, $p = .002$). However, when we refer to the Pearson's product-moment correlation – also known as the zero-order correlation – between "VO_2max" and "weight", without controlling for "age", as highlighted by the blue rectangle, we can see that there was also a statistically significant, moderate, negative correlation between "VO_2max" and "weight" ($r(98) = -.307$, $n = 100$, $p = .002$). This suggests that "age" had very little influence(9%) in controlling for the relationship between "VO_2max" and "weight".

Now you see this is what we are checking there also now how I am writing, The results of the partial correlation highlighted by the red rectangle show that there was a moderate, negative partial correlation, why moderate negative? This value is -3.14. So, a negative partial correlation between the dependent variable VO2 max and weight, while controlling for age, so this is how you write in your research paper or your report which was statistically significant, now see how I am writing r97 why 97 you see here the degrees of freedom is 97 is equal to -0.314 , n = 100, no missng values are there at p = .002 significance value.

However when we refer to the Pearson’s product moment correltion also known as the 0 order or the simple correlation. Without controlling for a age as highlighted by the blue rectangle, so here this is the blue rectangle where the 0 order correlation is taken there is no effect without controlling anything age or anything. we can see that there was also a statistically significant moderate negative relationship between VO2 and weight, so look at it 98 degree of freedom and the correlation value is -.307 and it is also significant at .002, VO2 versus weight.

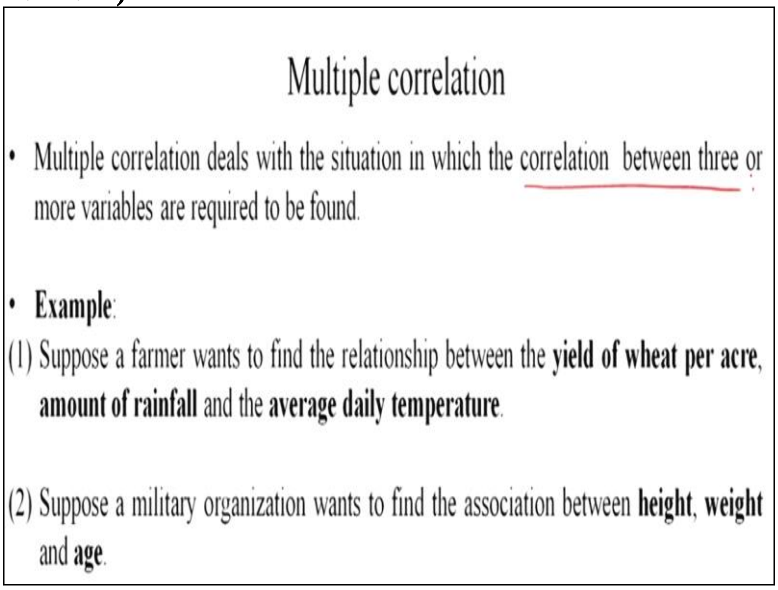
This suggests that age had very little influence 9% why I am saying 9% if you see .3 is the relationship so if r is my .3 what is my coefficient of determination r^2 which is equal to .3 * .3 that is equal to 9%, so only 9% you did not write this part but just you should understand so had very little influence controlling for the relationship between VO2 max and weight, so we will suggest that it had a influence but the influence is very little, age had a very little relationship influence on the relationship between weight and VO2 utilization, oxygen utiliation, so again writing the same.

(Refer Slide Time: 21:26)



A partial correlation was run to determine the relationship between an individuals VO2 max and weight while controlling for age. There was a moderate relationship between VO2 max so this is how you write the data from there and weight this much while controlling for age like this, which was statistically significant at this, however this is neccessary, 0 order correlate showed that there was a statistically significant moderate negative relationship between this and indicating that age had very little influence in controlling for the relationship, so when I see the difference between by taking the age and by not taking the age and I compare obviously I can see the effect, so this is what we do.

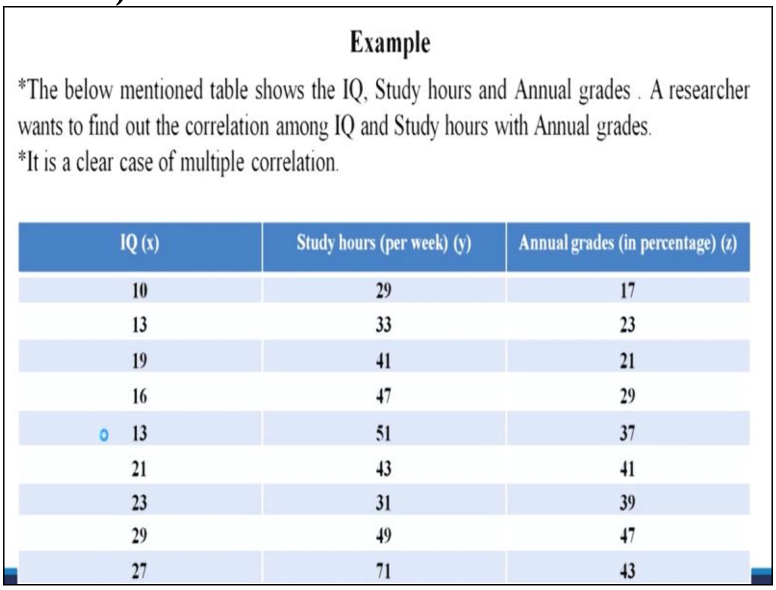
(Refer Slide Time: 22:20)



So now from here we will move in to the next which is the multiple correlation. So, now we are not talking about partialing out or supperising or controlling anything so we are moving in to the third condition. Multiple correlation deals with the situation in which the correlation between 3 or more variables are required to be found, atleast 3 or more.

Example, a farmer wants to find the relationship between yield of wheat per acre, amount of rainfall and the average daily temperature, so he is not partializing out and so he is not controling anything, so he wants to take all the 3 variables and want to see what is the relationship at one time see at one point of time, what is the relationship, similarly, a military organization wants to find the association between height, weight and age.

(Refer Slide Time: 22:54)



Example

*The below mentioned table shows the IQ, Study hours and Annual grades . A researcher wants to find out the correlation among IQ and Study hours with Annual grades.

*It is a clear case of multiple correlation.

IQ (x)	Study hours (per week) (y)	Annual grades (in percentage) (z)
10	29	17
13	33	23
19	41	21
16	47	29
13	51	37
21	43	41
23	31	39
29	49	47
27	71	43

Now take this example, a study shows that IQ, Study Hours and Annual grades are related. A researcher wants to find out the correlation amoung IQ, study hours with Annual grades. So it is a clear case of multiple correlation, however sombody would have said let us take the study hours as a control factors something then we could have been a partial correlation, but here we are not taking anything controlling, anything we are taking all the 3 at the same time, my IQ is given of some students, the number of hours they put per week is given and there grades are percentages are given, now how do you do that.

(Refer Slide Time: 23:46)

Multiple correlation coefficient formula

$$(R_{xyz}) = \sqrt{\frac{(r_{xz}^2 + r_{yz}^2 - 2\ r_{xy}\ r_{yz} r_{zx})}{(1 - r_{xy}^2)}}$$

$$= \sqrt{\frac{\{(0.6363 + 0.3469) - 2(0.5323 * 0.5890 * 0.7977)\}}{(1 - 0.2833)}}$$

$$= \sqrt{\frac{0.9832 - 2*(0.2501)}{0.7167}}$$

$$= \sqrt{\frac{0.482}{0.7167}}$$

$$= \sqrt{0.6725}$$

= 0.82

Interpretation: Study found a strongly positive correlation among IQ and Study hours with Annual grades.

Same way, let us find out, so I am findig r xy, r yz, r zx and then we use this formula

$$(R_{xyz}) = \sqrt{\frac{(r_{xz}^2 + r_{yz}^2 - 2\ r_{xy}\ r_{yz} r_{zx})}{(1 - r_{xy}^2)}}$$

$$= \sqrt{\frac{\{(0.6363 + 0.3469) - 2(0.5323 * 0.5890 * 0.7977)\}}{(1 - 0.2833)}}$$

$$= \sqrt{\frac{0.9832 - 2*(0.2501)}{0.7167}}$$

$$= \sqrt{\frac{0.482}{0.7167}}$$

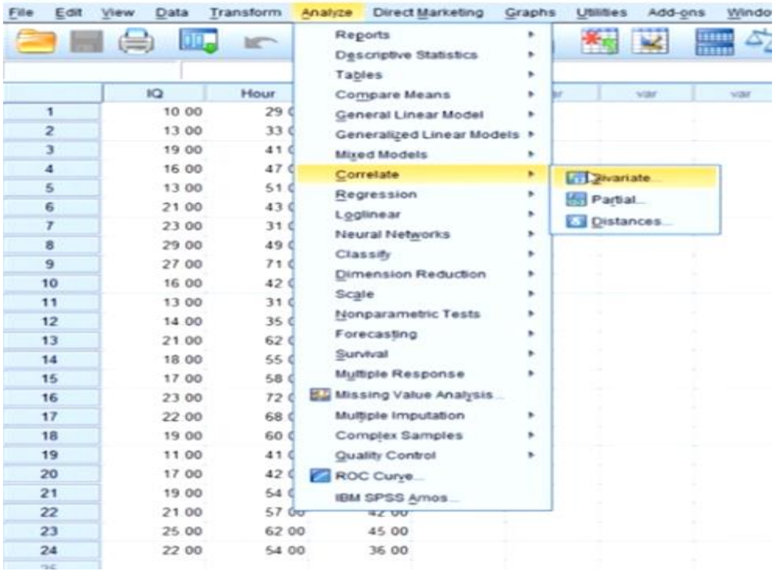
$$= \sqrt{0.6725}$$

= 0.82

so now using this formula we will comeback and see, we have done it earlier also, so let us just look at it, so same if you remember, these values are more or less the same, so 346, 1336 are same, so we are use the same that means the problem again.

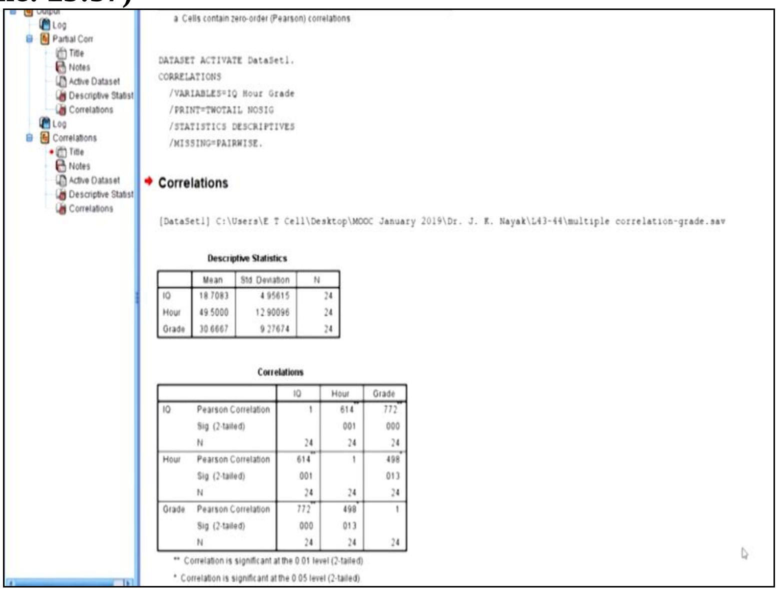
So by calculating it we have found out r square the multiple correlation coefficient is equal to 0.82, so if my correlation is 0.82, what is my interpretation the study found a strong positive correlation amoung IQ and study hours with Annual grades, so what was it you see X is IQ, Y is my study hours and this is my grade, so what I am saying study found a strongly positive correlation among IQ and study hours with Annual grades, so this is all we have now let us solve the same in SPSS also, so I brought one more dataset for you

(Refer Slide Time: 25:08)



Now this is the same case, so IQ, hour and grade now you want to do a multiple correlation,so let us go to correlation, so if you do correlation then you can do it let us say I want to go to bivariate, now I will take all the 3. Since I am not controlling anything, I take all, you can see this is the pearson correlation, why Pearson's correlation? By this time we should be aware it is normally distributed data, you can even check the normal distribution individually of the variables and then you should move ahead and please kindly check all the assumptions that are required they should be intact, so I do not need anything else, I am going with this, now Let us look the output.

(Refer Slide Time: 25:57)

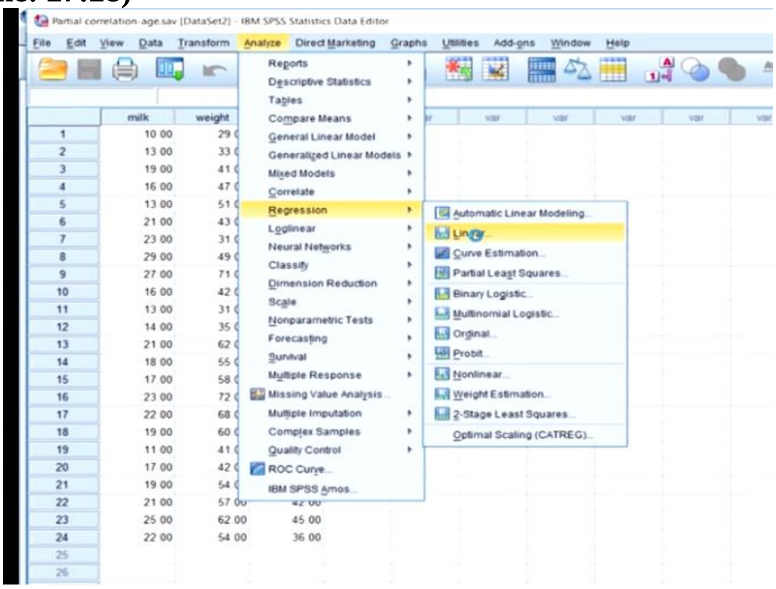


So what is the output, now here we have the correlation first look at the descriptive statistics, the mean of IQ is 18.7, hours is 49.5, grade is 30.6 and standard deviations are given, so there are 24 cases total, now the correlation between IQ and IQ will be 1, IQ and hour is how much .614 and IQ and grade is .772 and you can see the star mark.

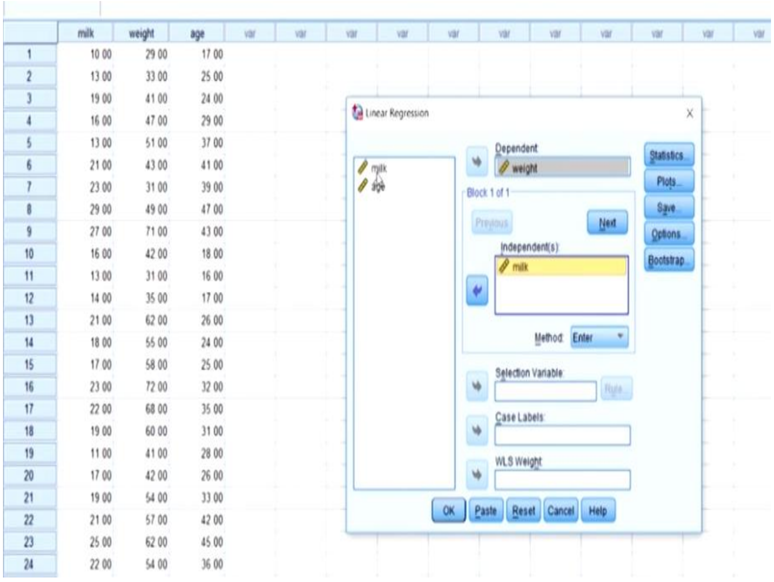
Now it is star marked means what it is saying, if it is 2 star double star the correlation is significant at 1 % or 0.01 level for 2- tailed test, similarlly the correlation is significant if it is 1 star, is there any 1 stared here? This one is 1 stared, so hour and grade the relationship it is 1 stared that means significant at the 5, this is the way of only marking something coding something there is 5% or 1%, so now look at it.

So IQ and hour .614 it is significant. IQ and grade the relationship also significant at 1%, hour and grade is also significant but it is significant at 5% level of significant so when you write this, so this the way you interpret the data and you write the inferrence in your research report, now but we want to suppose find out the overall multiple correlation value, what we do.

(Refer Slide Time: 27:28)

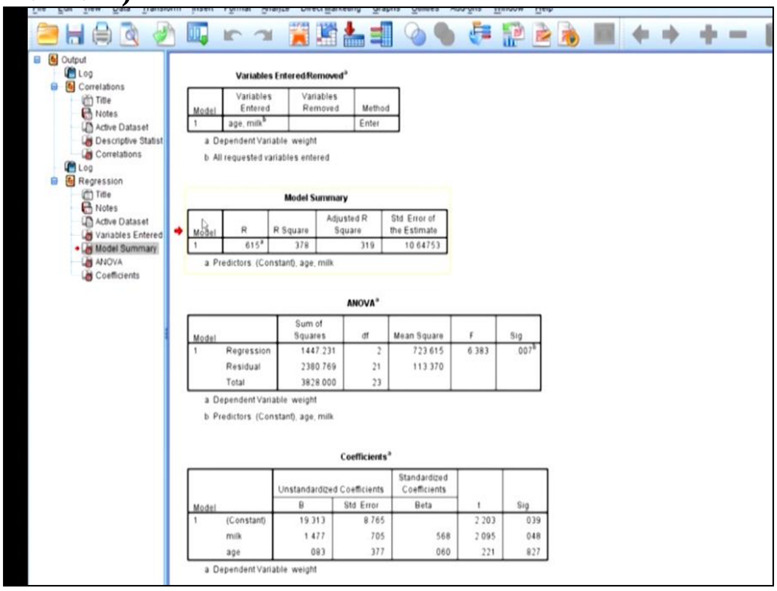


(Refer Slide Time: 27:48)



So let us go back to the file and go to analyze and go to regression and linear, now what I will do is I will take my dependent variable as what now weight and milk and age are my independent variable, so I do not need anything else so I will keep things constant I will just go for it.

(Refer Slide Time: 28:30)



Now let us look at the output file now if you look at the output file you can see that when we have taken this r tells us overall relationship between the age, milk and weight, this r value is important to us, this r basically in some we explained it regression as the explained variance but in other term we say this the correlation we are talking about the overall correlation is 615, so the model has the correlation of 615 that means 61% of the variances explained and this is the overall model the correlation coefficient, so this is what we had today. I hope you are clear with the concept of partial correlation and multiple correlation.

So partial correlation you will use when you want to suppress or control some factor, some variable and multiple correlation you will do it when you want to find out the relationship between 3 or more variables at the same time without controlling any one of them. So, this the over all way you have to understand and then this is how you have to write the interpretation in results paper. I think this the clear to you and this is very vital because this is highly helpful for getting a good publication and for making the readers understand what you have actually done in an experiment. So, Thank You very much.