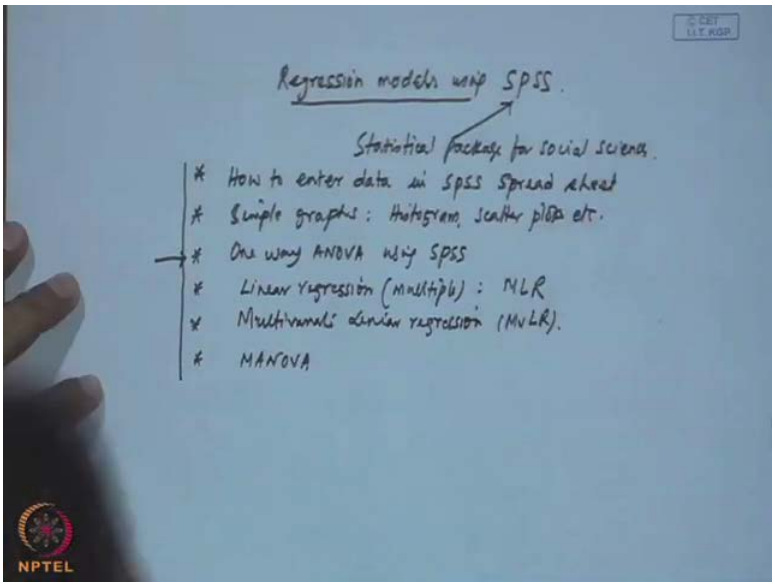
Applied Multivariate Statistical Modelling
Prof. J. Maiti
Department of Industrial Engineering and Management
Indian Institute of Technology, Kharagpur

Lecture - 32
Regression Modelling Using SPSS

(Refer Slide Time 00:23)

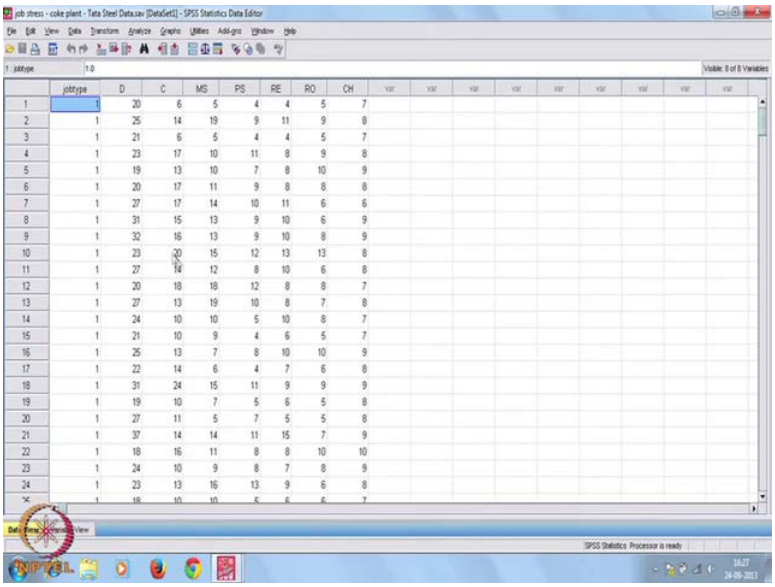


Good afternoon. Now, we will discuss regression models using SPSS, SPSS is known as statistical packagers package for social scientists, statistical package for social sciences ((Refer Time 00:58)), I think we will find out that almost all types of statistical analysis is possible using SPSS. Nowadays the determining and other advance techniques are also available in SPSS. So, what we will discuss, now I will first show you that how to enter data in SPSS spread sheet, then we will see that how the some simple graphs, simple graphs; for example, Histogram, Scatter plot, all those, etcetera, all these can be done using SPSS.

Then I will show you that one way ANOVA, although the topic is regression models using SPSS, but I will show you in one way, then how one way ANOVA, because this is the first module class we have taken, one way ANOVA using SPSS. Then I will go to linear regression, so that mean multiple linear regressions, basically linear regression in the multiple as well as simple, so you can use this one, multiple linear regression model. Then I will show you that multivariate linear regression M v L R.

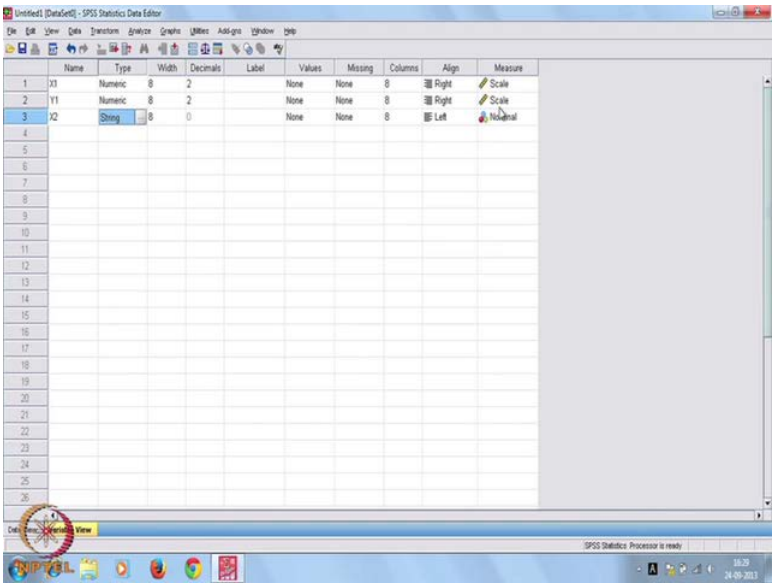
As we are discussing these one way ANOVA, so we may be also interested to know how is MANOVA possible in SPSS or not. MANOVA also using SPSS you can run MANOVA, but I am not sure that all this and this will be completed by this lecture. Later on we will go for one more that is SPSS demonstration, if requires.

(Refer Slide Time 03:43)



So, you see this one is the spread sheet for SPSS, now how I have brought data into this sheet, for example if I just I am just closing down this one. So, you go to SPSS, we are using here SPSS 17. Now, if there are earlier files what you have used, then it will ask you whether you want to open one existing file or not, for example you do not want to open any existing file, so then you put, this is what is it is asking for existing file, so you do not go for existing file.

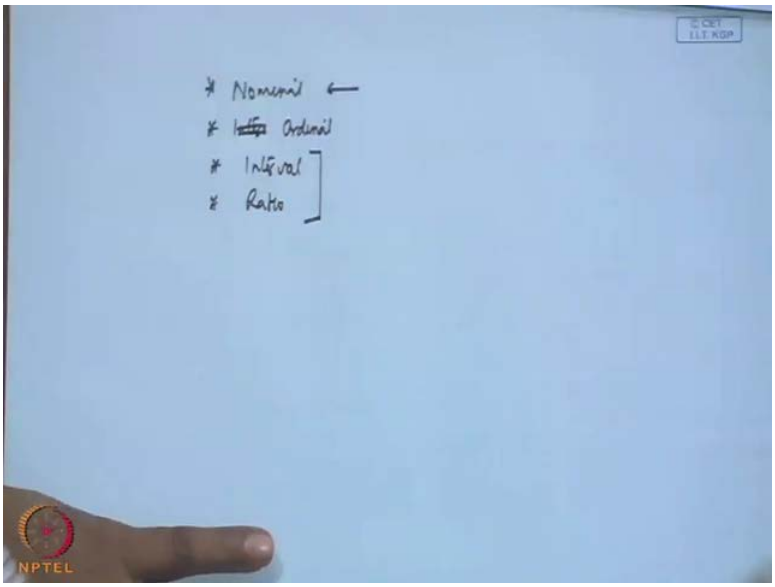
(Refer Slide Time 04:44)



Now, there are two windows, one is data view window, another one is variable view window. So, you come first variable window and create some variable, for example x 1, then what type of variable it is if you click here, you will find out that the variable type can be numeric, coma, dot, scientific notation. So, whatever we want captures the data that is given here. For example, let us this one is x 1 is a numeric variable and we are giving the width 8 and decimal place is 2, so that means up to 2 decimal places you will be storing here.

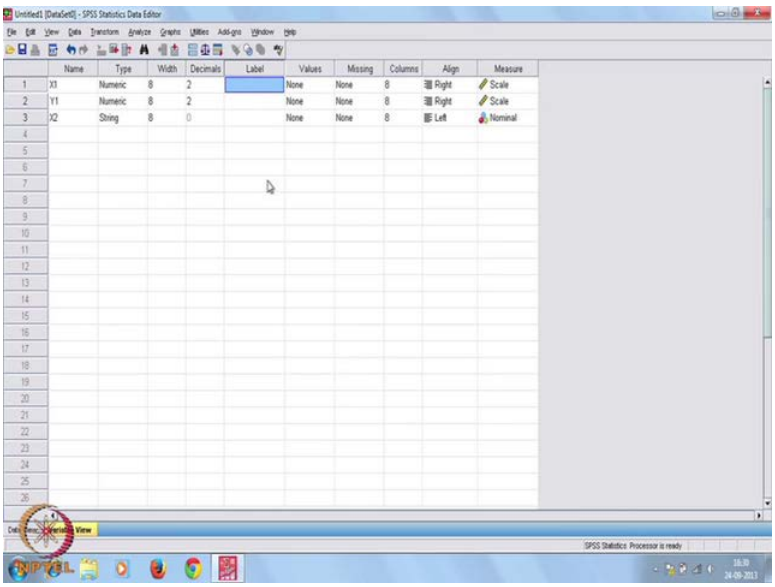
Give one more, suppose let it be y 1 and this one also your numeric with same thing and then you put x 2. Let this one is a string variable, this is a string variable, so then the symbol of string variable is like this and when the measurement is nominal that is why this string. So, I think you can you can recollect that one what we have discussed in data types.

(Refer Slide Time 06:07)



That nominal, interval, nominal then ordinal, then your interval, then your issue, so when it is said scale data. This basically coming from this two and they are also saying nominal data, so you can make this one.

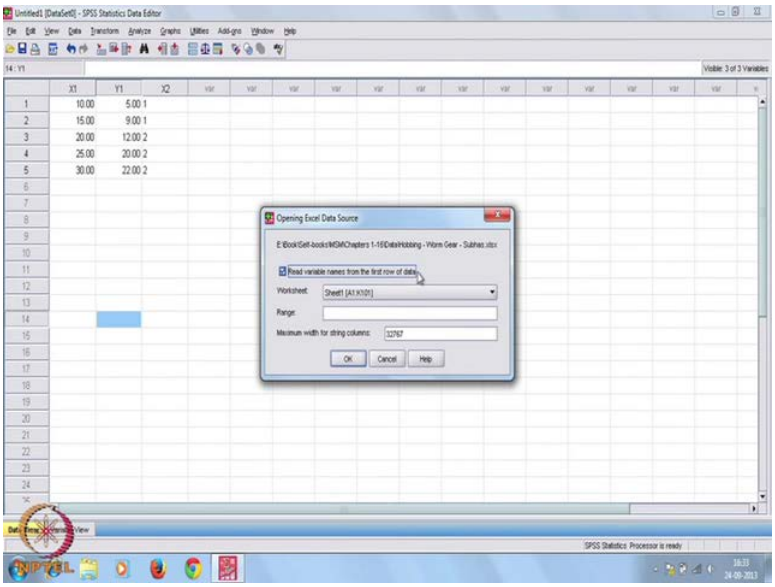
(Refer Slide Time 06:37)



Suppose, if I click this, see that scale ordinal and nominal it is available here, so if you want to make it ordinal click like this, so in case of ordinal data there will be bar chart type symbol. Suppose this one is scale data, so you want to make it nominal, make it like this. So, here also you have to change accordingly and you have to go for string and

accordingly you have to change, we will not change, so this one measurement is our scale measurement automatic, this one is nominal measurement let it be like this. So, this is what is known as you have defined the variables, you can level the variables, but for the time being we are not interested in levelling.

(Refer Slide Time 07:27)



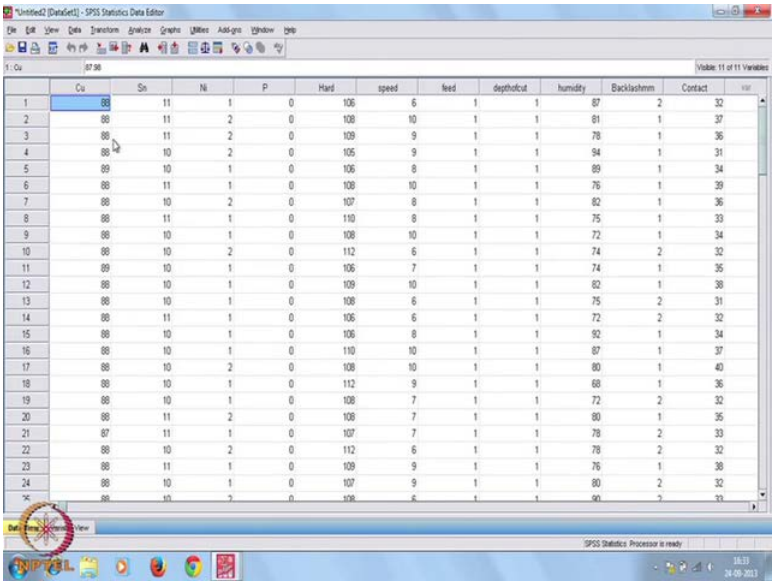
So, go to the data view, what has happened here? You see x 1 y 1 and x 2 theses three variables are created and we have said that x 2 is string variable, x 1 and y 1 are numeric variables, so put some value. For example, let this is 10 then 15 then 20 then 25, let it be 30, so y 1 let it be 5, let it be 9, let it be 12, then followed by suppose 20, follows by let it be 22.

Then x we are saying that two types, nominal that is nominal we said, for example I will keep 1 1 2 2 2, I given numerical values here 1 1 2 2 2 like this, but you can type also, you can give some another string variable. So, this is let it derivable data set, now, but it is this spread sheet is similar to excel spreadsheet, you have seen in excel also. So, instead of entering data directly in SPSS spreadsheet, you can import data copied or stored in excel file, that is also possible.

So, let us see how can you do this? Suppose, you go to open, then go to the file where the data is there, suppose I am going to the file where this data is available, this is the data folder, so see here these datas, these are all SPSS data. Now, I want to go excel data, there is excel comment also you see once you click here you will get excel here. For

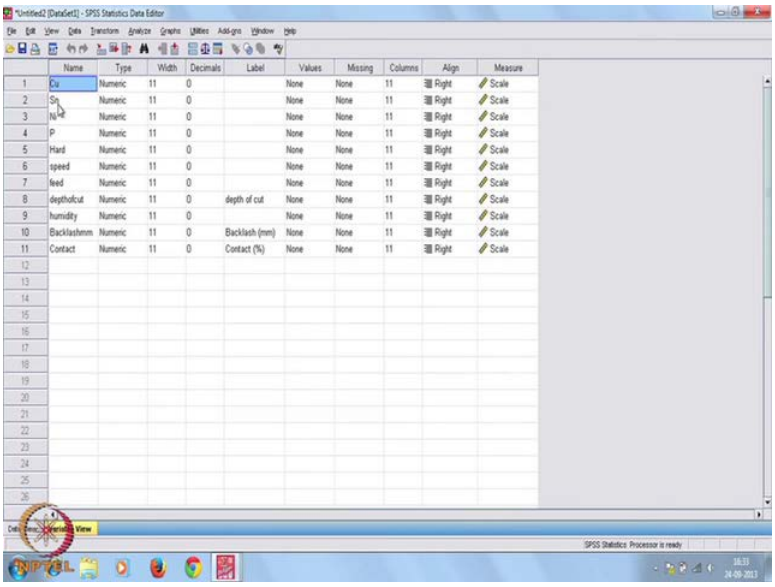
example, we will take one data set like this, for example this one click here and then make open, then one comment will come, it has a this asking read variable names from the first row of the data, variable name from the first row of the data and if you say.

(Refer Slide Time 09:57)



Then see all the variables and the related data are available in the SPSS folder, correct?

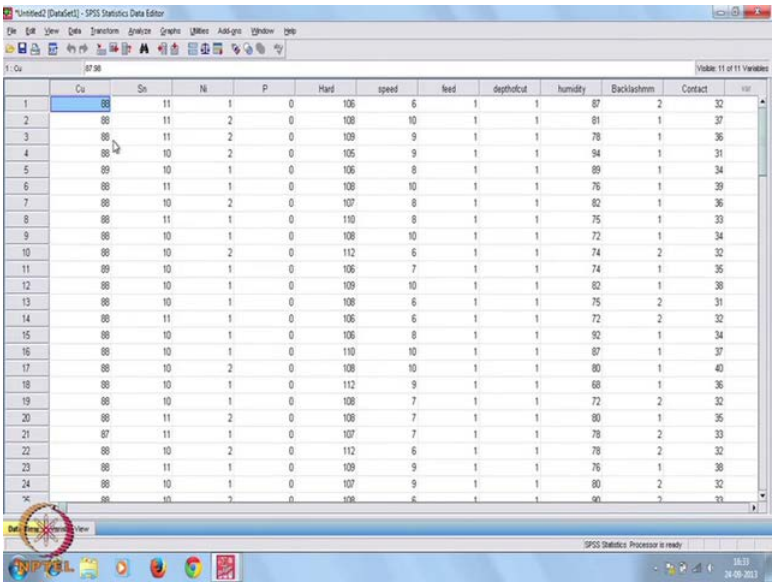
(Refer Slide Time 10:10)



So, if you go to the variable view, you see that copper, silicon, nickel, phosphorous ((Refer Time: 10:14)) speed all those variables here their data type is numeric width given and their label also given, and all are measured in ratios that scale data everything.

So, that means your data is ready in SPSS spreadsheet and now you can do whatever you want to do in this data.

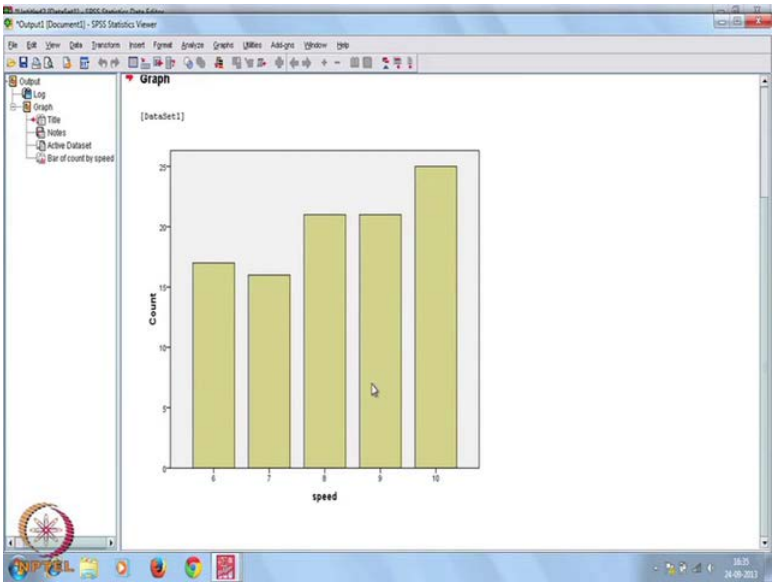
(Refer Slide Time: 10:40)



So, let us see some of the graphs, see there is one column, if you at the there is one tag is there that is graph, then analyze, then transform, so file, edit, view, data, transform, analyze, graphs, utilities, so on, so many fields are there. Now, I will go to graph there are, there are chart builder, there are some template, that we will go by the legacy dialogues here, for example bar chart you choose. So, you can choose bar chart simple clustered stacked, let us choose this one, first one that summaries for group of cases, summaries for separate variables, values of individual only, the first one you take.

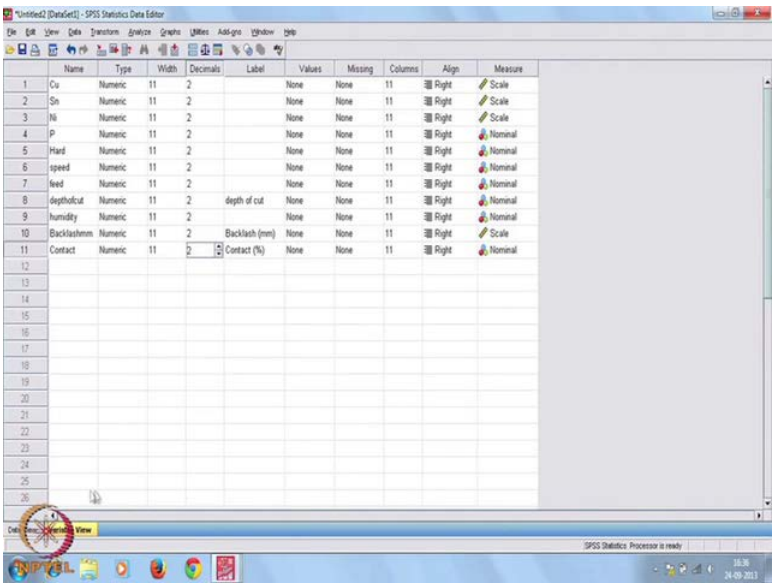
Then what it is asking it is asking, what is your category axis and what are the rows and other things. So, bars represent number of cases let it be, so let I want to give which variable we will give I think this feed this is showing 1 1 1, so let us give this one, now put.

(Refer Slide Time 11:57)



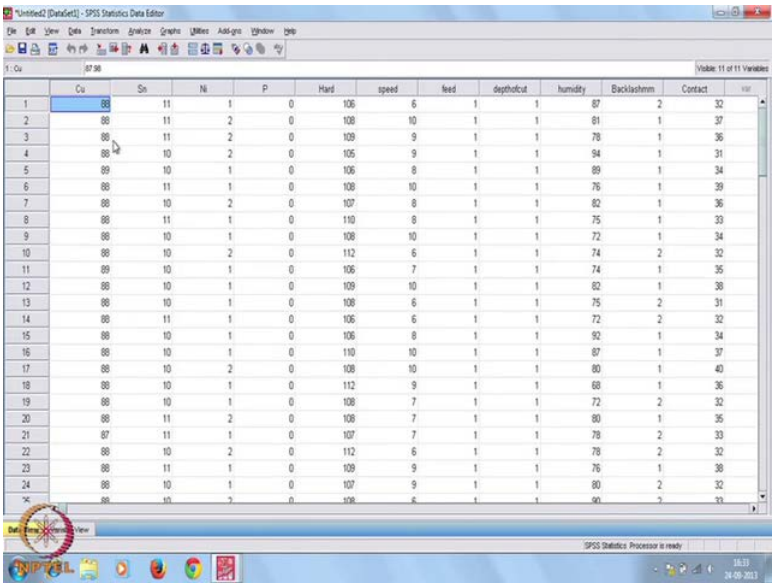
See this is the speed has 6 to 10, 6 7 8 9 10 different, actually this is not 6 7 8 9 10 these are the speed values. If you I think there is, if I go back I want to see what we have taken speed value not feed value.

(Refer Slide Time 12:21)



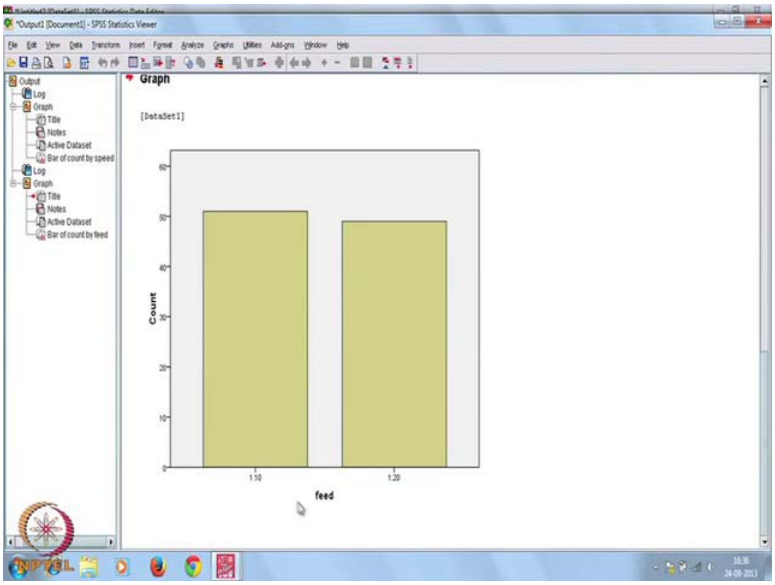
So, let me go back to the data again here decimal is, no decimal is considered I am considering second decimal up to decimal 2.

(Refer Slide Time 12:47)



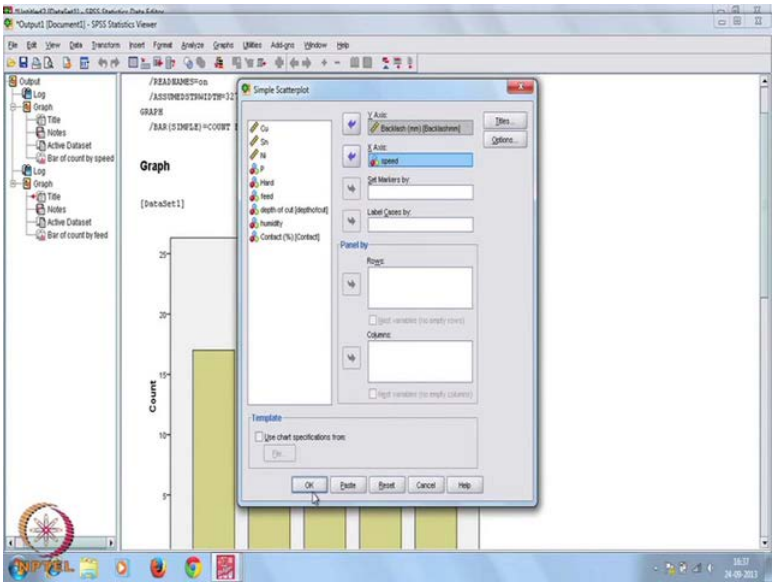
Now, go back see ultimately that all those 1 values 1.1, 1.2 like this. So, I am again going to back to your legacy graph and bar chart, we have taken the first one, and then this is fine we have taken speed, but we want to take the feed one.

(Refer Slide Time 13:03)



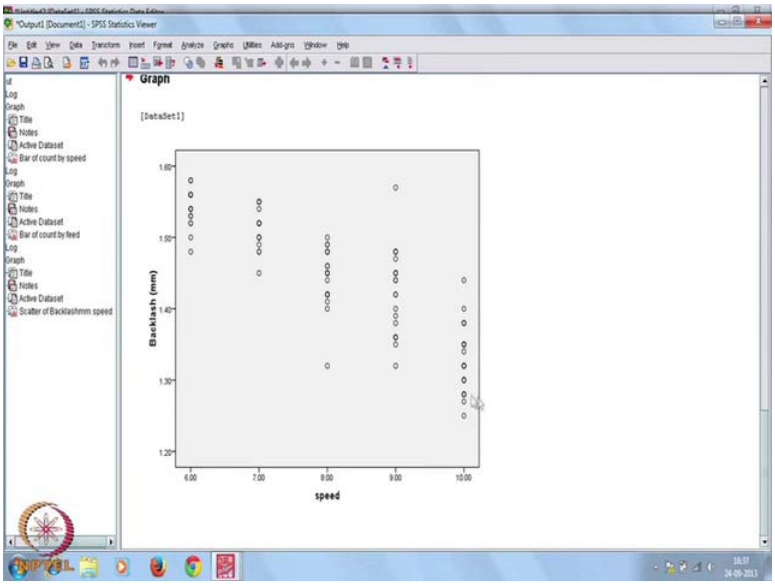
So, feed you see that what are you getting, there are two labels basically, one is 1.10 and another is 1.20. Although this 1.10 1.20 is numerical value measured in scale, but actually the process is operating under these two feed condition, duration per minute.

(Refer Slide Time 13:27)



Now, if you want to go for some other type of, for example suppose you may be interested to know scatter plot, for example simple, then this one is simple scatter, define what will be x axis? Let backlash is your x axis, what is your y axis? Let speed is your y axis. So, many things, now you can set marker labels other that option are available.

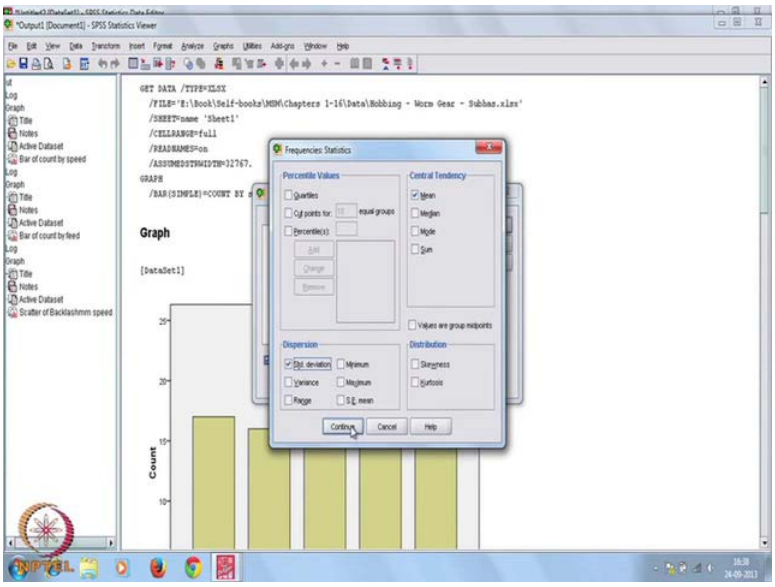
(Refer Slide Time 14:02)



We will just go by these two and see what is happening here? This is the scatter plot what I have shown you earlier under that linear regression multiple regression class,

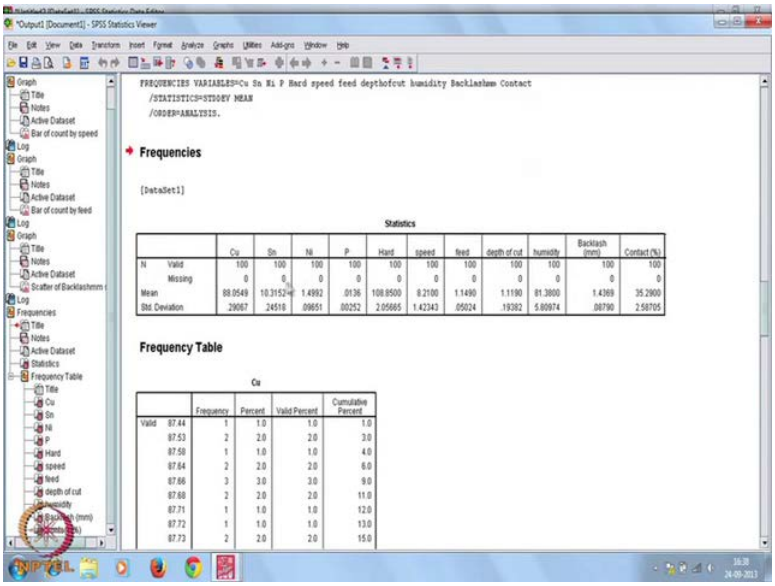
what it is happening here, it is showing that backlash decreases with increase in speed, correct?

(Refer Slide Time 14:26)



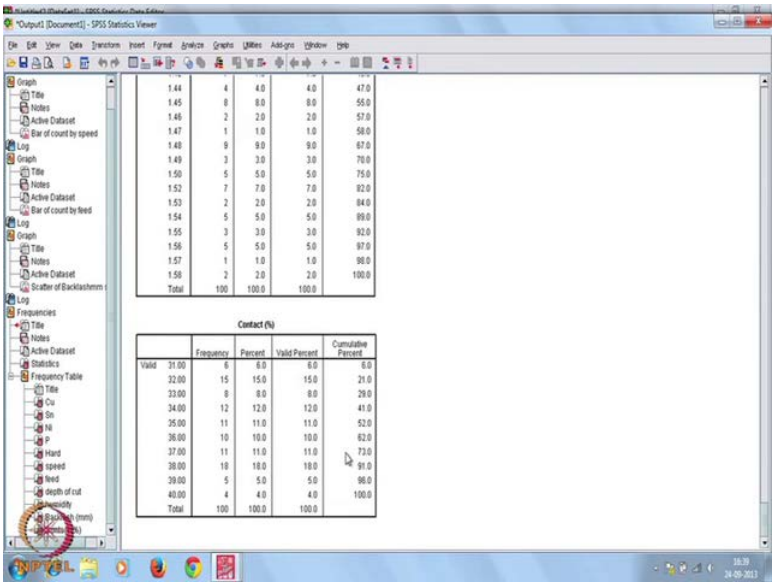
Now, you may be interested to find out some of the descriptive statistics, for example let it be frequencies, then if you can you can select, you can select one variable at a time, you can select all variable just pressing tab, first how do select all variable? First click on particular variable, then press tab, then finally press the last one, if you want to take, then this all. There are statistics what you want, quartile, some points, mean, median, mode, so many things are there. For example, I want the mean and I want standard deviation, this two only, so give make it.

(Refer Slide Time 15:10)



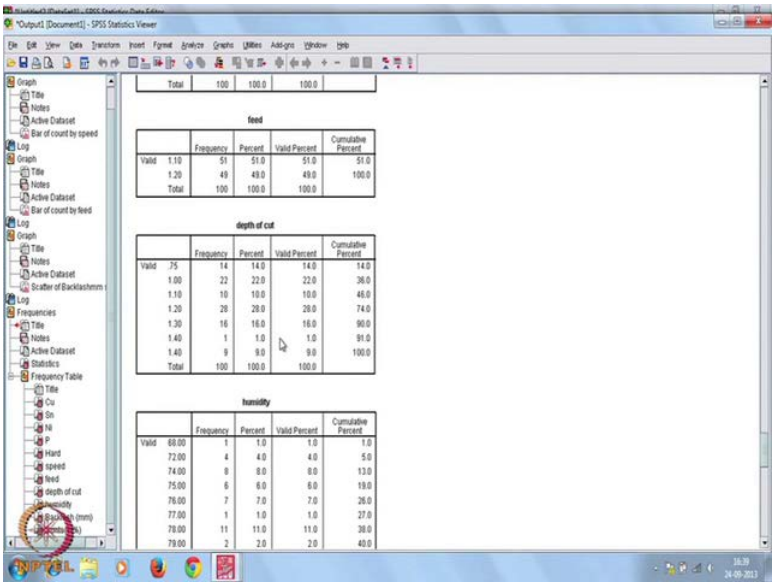
You see that for the all the variables, what are the total data points, how many missing values are there and then what is the mean value and what is the standard deviations everything coming and frequency table. Frequency table is giving basically what are the observed values under different, so it will be big one, because some cases 100 data points are there, very big.

(Refer Slide Time 15:37)



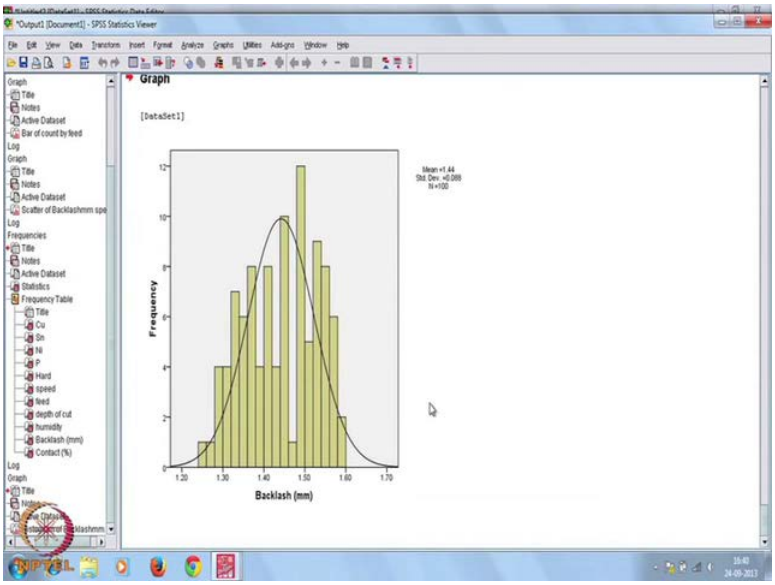
When you have continuous data, then that frequency table is not meaningful.

(Refer Slide Time 15:43)



For example, you see that these are all humidity, but depth of cut, depth of cut is measured under these 7 scales probably, 1 2 3 4 5 6 7 different categories 0.752 1.40. All these it is meaningful, because your frequency is there different, sales different level of frequency.

(Refer Slide Time 16:21)



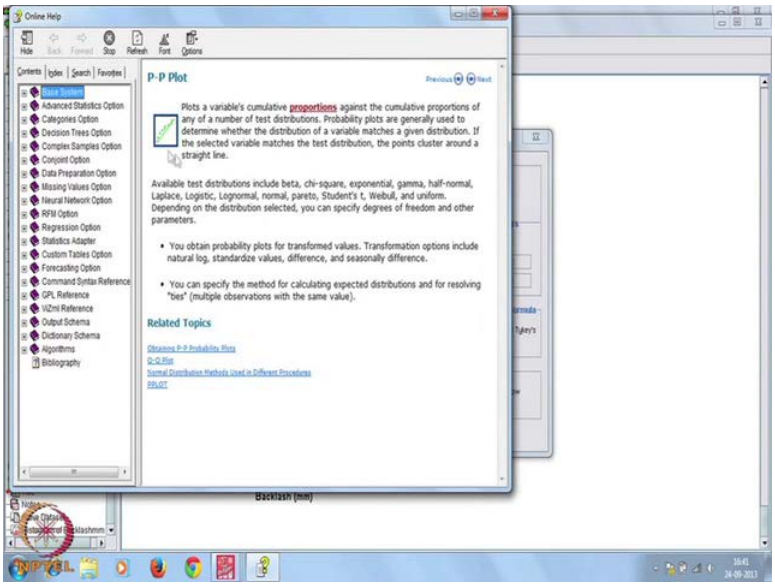
Now, let us, so if you interested further, suppose you want to develop histogram, I think this is histogram. So, you choose the variable, for which you are interested to know the histogram, if you want to display normal curve click there, you do it, so your histogram

is plotted. Actually this not a perfect normal distribution, there is certain data problem, for example here, but this type of histogram you will find out, because this is coming from the actual production shop, you can create. But for an actual production shop you have we have collected this data and as a result you are getting like this.

Now, I will show you that the analysis part, what are the options available under analysis? First one is report, in report it is basically the preliminary determining types of thing, the olap cubes is there, then you will summarize, in rows, in column, all those things are there. Then descriptive statistics starting from frequency, descriptive, explore, crosstab ratio, p p plot and q q plot, for example you want to develop p p plot, for which variable you are interested to develop p p plot?

For example, we are interested for backlash and contact and what is the test distribution you are looking for? Let it be normal. So, now there are many options you can transform the data, you can that proportion estimate is available, rank assigned to ties available and you have to understand this, for example if by chance you are not able to understand what is this you click help.

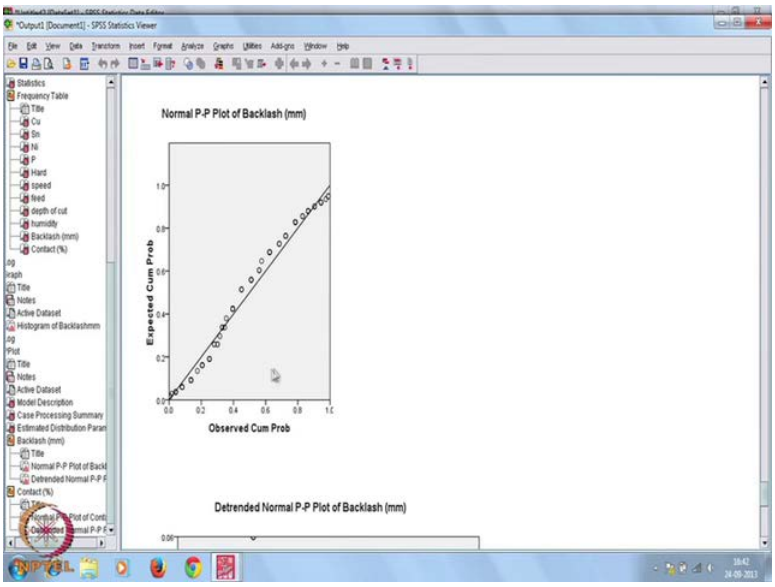
(Refer Slide Time 18:23)



Once you click help then as you have chosen p p plot, that it will explain what is p p plot. So, you may be interested to go some related topics, for example q q plot, for example how to obtain p p plot. So, these are possible and then next slowly these different tips you will be getting and which will help you while analyzing through SPSS. So, we know

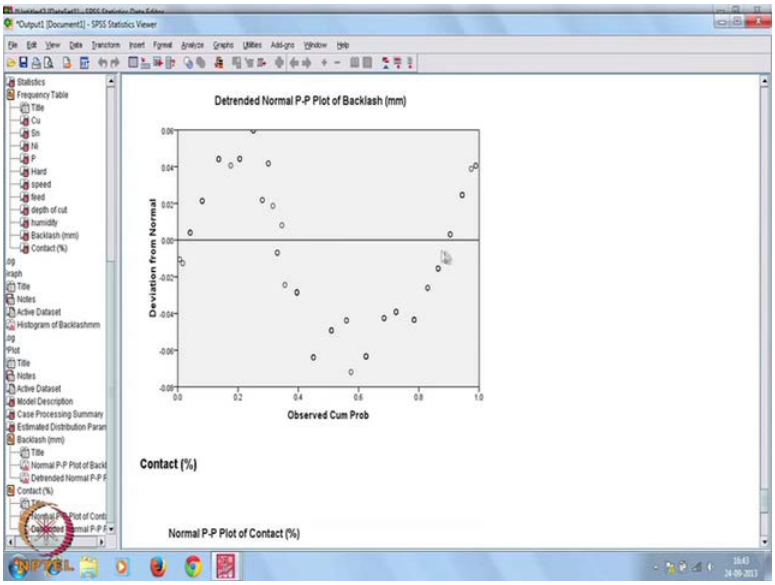
what is normal probability plot, we have discussed earlier also, so now let us click ok. What happened you see?

(Refer Slide Time 19:11)



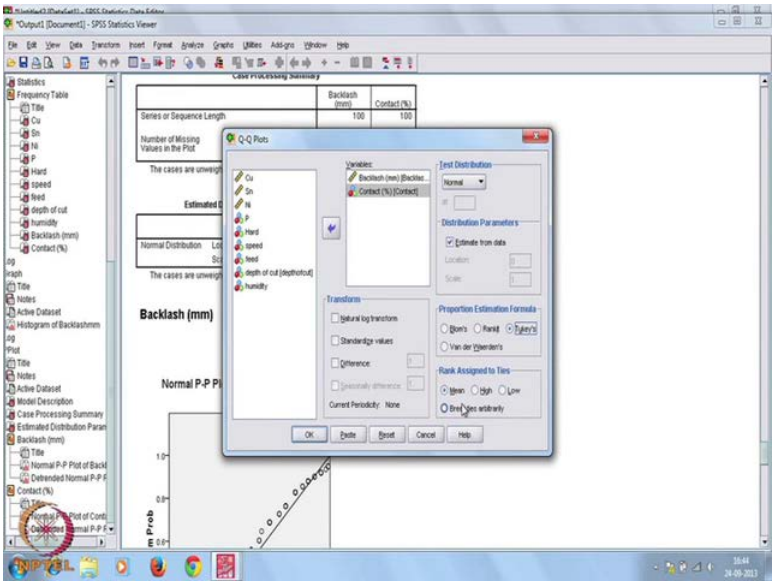
Now, this is what we have seen that the observed cumulative probability and expected cumulative probability, these things definitely will follow straight line. And you see ultimately the data seed for backlash it is on the straight line, but there is little epically a curve, some s type of curve is there, very little, but s type of curvature is there.

(Refer Slide Time 19:43)



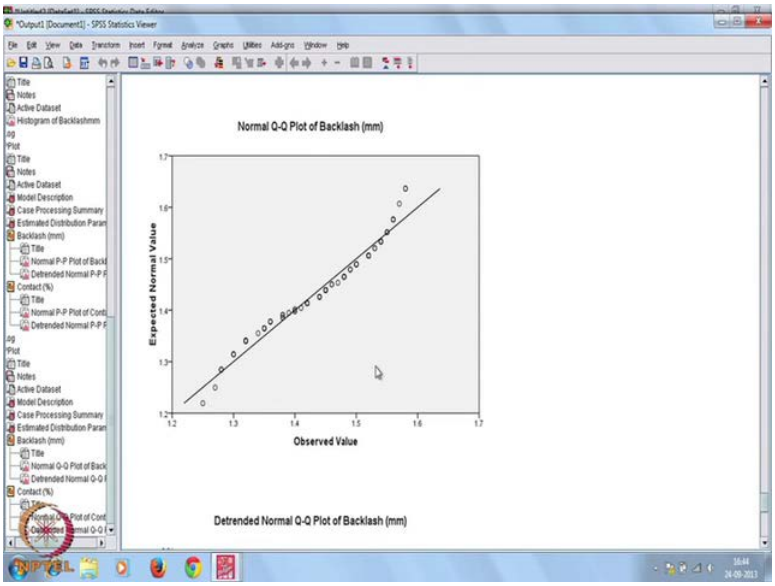
So, if you see this one, the deviation from normal and observed cumulative frequency for backlash, this much is deviation is there, this from p p plot you are getting, now whether this deviation is within the acceptable limit or not, so that you have to look into. Now, if you see normal probability of plot contact also you are getting like this and the deviation from normal is given, now what is deviation from normal, now what is deviation your for backlash, what is deviation for normal for contact? Those things you must understand properly and that theory you must know.

(Refer Slide Time 20:25)



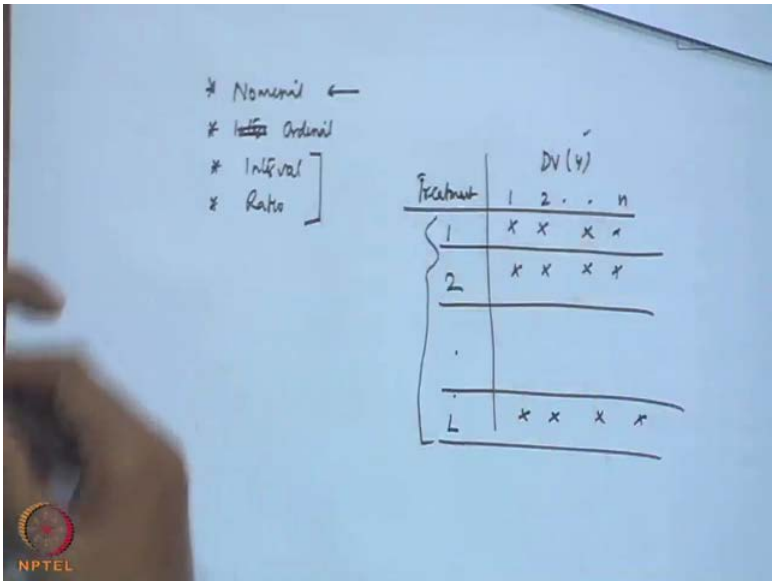
So, in addition here what you want to do suppose we are going for q q plot, then let the take the same things, same mean thing mean, same two variables, again normal. Then I think what you want here, I want to give the confidence interval, proportion, formula transform, standard value, difference, break high these, proportion, estimation formulas, it is not available here. Let it be tukey we will give, what is the difference I want to see, natural log transform.

(Refer Slide Time 20:10)



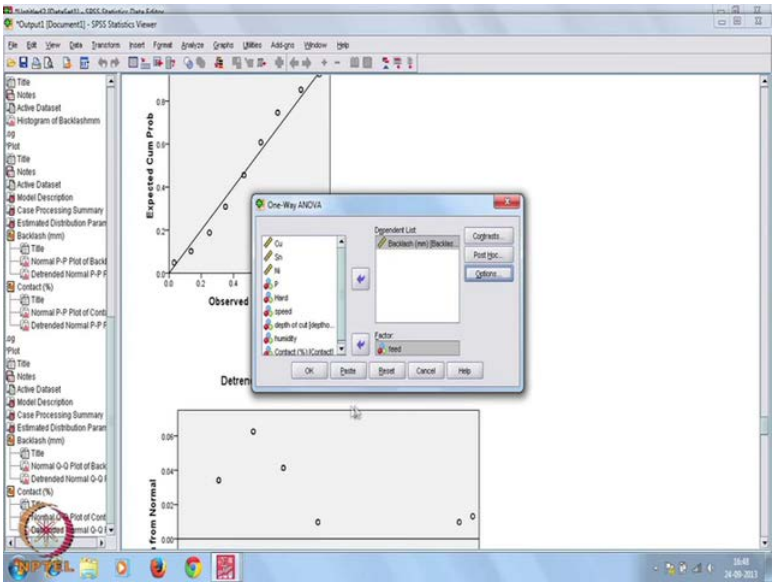
Then see this is also this is little different mean showing slight departure compared to p p plot and it is obvious departure, so this is for your contact and this what the difference found nothing else is developed here. So, fine then let us go to the, that report descriptive statistics, there are many thing, then table I can prepare r f m analysis, compare means. Now, in compare means there is one link which is one-way ANOVA, let us develop this one way ANOVA, now so what you are required to do? Go to analyze, then go to compare means, then click one-way ANOVA.

(Refer Slide Time 22:16)



So, then here you have to see what are the dependant variables and I think all of you know that ANOVA model, ANOVA we said that treatment or population, we said that 1 2 like this suppose there are L population and then there will be one dependant variable let it be y and observations 1 2 like n, so here several observation here second one several observation, so like this. By one-way ANOVA we want to test whether these treatments, different treatment or different level or different population have effect on this dependant variable or not, that is what we want to test. So, now here although dependant list is there if you put more than one variable it will basically do one-way ANOVA for repeatedly for different variables.

(Refer Slide Time 23:08)

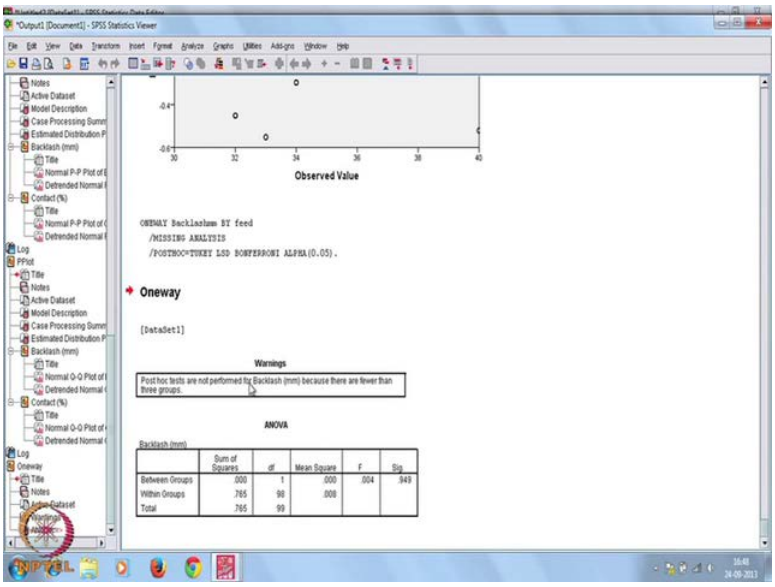


For the time being, let us give only one variable backlash, then what is your factor? Factor will be I think it is, I do not know whether these are all coming under speed nominal, anyhow let me give the feed one. Then there is contrast, under contrast polynomial degrees of freedom, coefficients, coefficients this, so here contrast nothing is shown except polynomial.

Now, you can go for post hoc, that post hoc analysis here you can get the under equal variants, you can get the l s d, you can get bonferroni, you can get tukey, so you can get waller duncan. I think we have described l s d, bonferroni and tukey approach, if your equal variance is not assumed, then these are the different measures which are possible, so let us do l s d and tukey and bonferroni. Now, there is option, under option what are

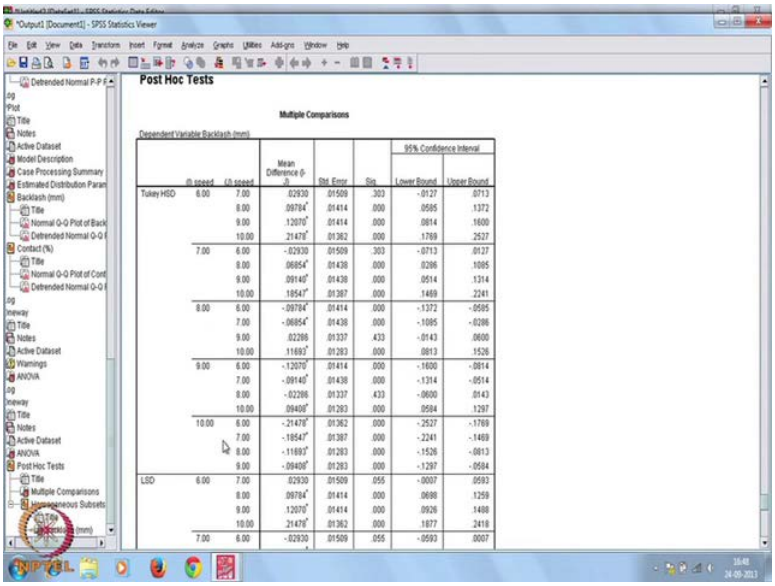
the statistics you want that is given, if there is any missing values, so how do they analyze, this analyze this missing value that is given, let us put.

(Refer Slide Time 24:37)



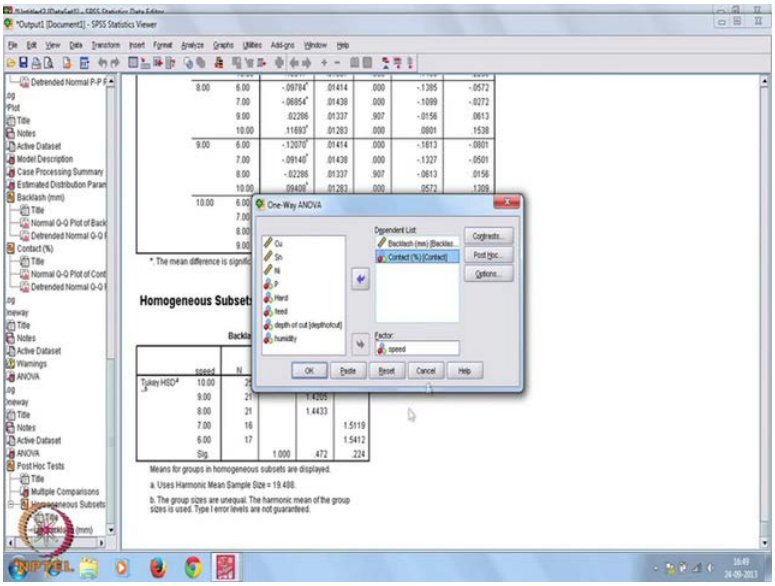
So, what you are getting here you see one way ANOVA, that post hoc test are not performed for backlash, because there are fewer than three groups. Basically only two groups are there that we have taken the feed, feed variable is having 1.1 and 1.2 data that two labels are there, so let us take some other variable. For example, compare means one way ANOVA I am not taking feed here I want to take speed, the same thing let us do what is happening here you see, because here more number of levels are there.

(Refer Slide Time 25:17)



So, 6 7 8 9 10, then this five levels are there 6 to 10, 6 7 8 9 10, five levels are there and you are testing, this is the ANOVA table, you are testing whether if your speed is that level six or level seven or level eight or up to level ten, whether there is a difference or not? ANOVA table will give you this difference, this is the f statistics, f value is 78.109 very high, so there is difference.

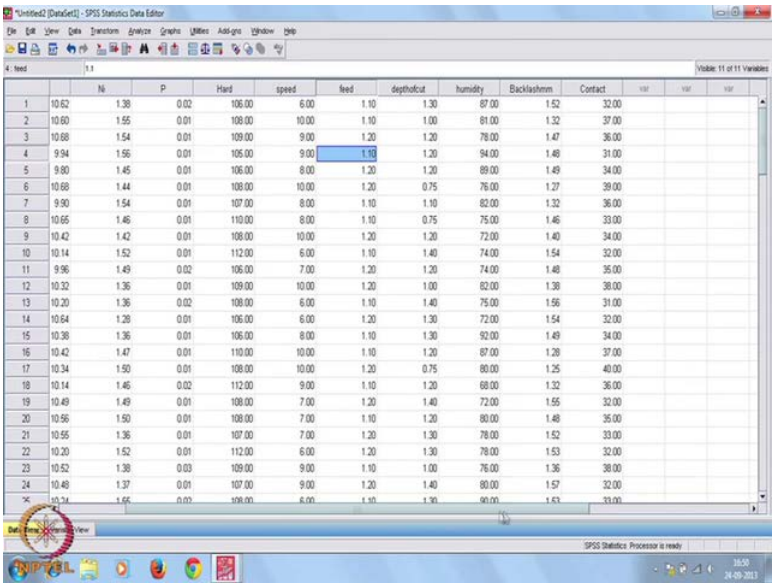
(Refer Slide Time 26:00)



So, in the same manner you can go for suppose what I do I will not, I will go for one more variable, compare means I have taken backlash, but I also want to take contact here

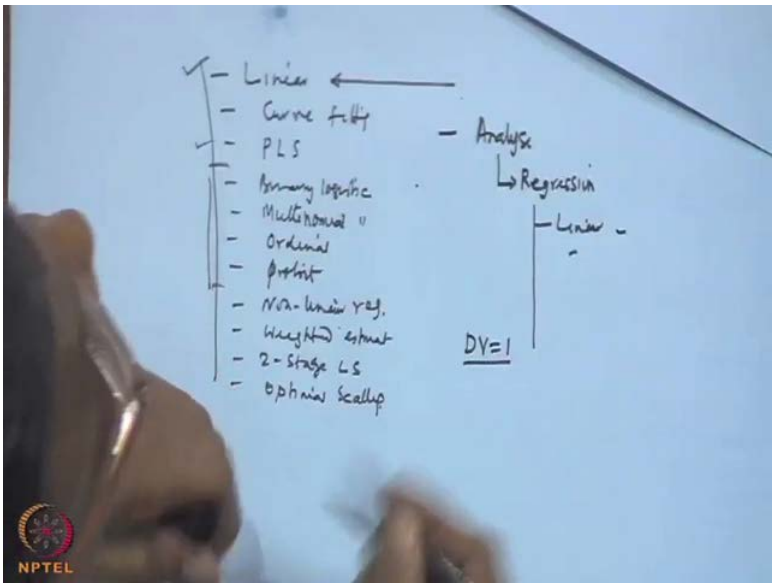
put here, then your factor is only one here, that factor speed, but you can, as it is one way ANOVA, so only one factor you have to considering. Now, if you put or click here what you are getting you are getting for backlash and contact separately. All tests everything, so it is just a one click business.

(Refer Slide Time 26:50)



Now, we will go for regression, so your data set is this and we have descried earlier that our backlash and contact, these two are dependant variable and we say others are independent variable. So, now go to analyze, I want to do regression, go to regression, under regression what are the different options available?

(Refer Slide Time 27:18)



You see there are linear regression, then there is curve estimation, that is your curve fitting, then there is partial least square P L S, there is binary logistic, multinomial logistic and you can go for ordinal regression, that ordinal logistics, then probit model, then non-linear weighted estimation, two stage estimation, then non-linear regression and then weighted estimation, two stage least square and finally this is showing optimal scaling.

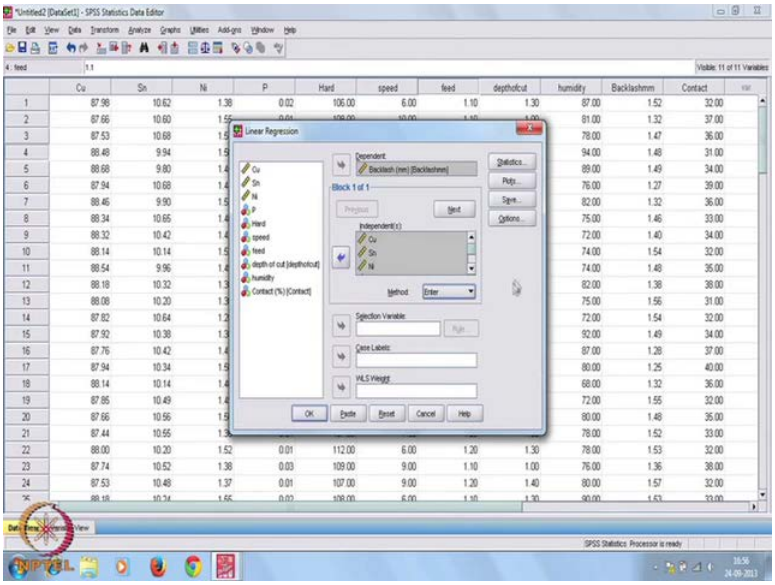
So, many under regression, so many techniques are available in SPSS, what we have discussed; we have only discussed linear regression. We have not discussed, basically linear regression is very popular, it will be partial least square regression is also very popular, all those logistic regressions including unknown probit all are very, very popular techniques.

So, that we mean, you can do linear regression, you can do non-linear regression, you can do linear regression in under different condition. For example, logistic regression is used when the dependant variable is categorical, so binary logistics means dependant variable has two categories, yes, no, 0, 1 type of things. You can go for multinomial logistic regression, when your dependant variable y has more than two categories. So, probit regression is also a special type of regression where that categories and issues are they are from the dependent variable side.

So, but we are interested in linear regression, because this one only we have computed, so then what is your what are the steps you have to follow for to use SPSS, then you go to analyze first, then click regression, if you click regression, then there will be a list, here the linear regression you click. So, in SPSS you analyze, you click regression, then click linear regression.

Now, let us click linear regression, what is happening here you see that one window coming up, that window, there is dependent, there is independent side, then there are different selection variable rules, case labels, weighted regression, weighted least square, weights everything is there. So, what you have to do first? First you have to choose your independent variable, so as it is a multiple linear regression case, so your dependent variable will be 1, D V will be 1.

(Refer Slide Time 31:54)



So, which one is our dependent variable? Our dependent variable let it be backlash, so you click on backlash, then there is a arrow under dependent variable, so when you bring mouse here, this become little larger, then you click here, your dependent variable column, now is filled with that variable, dependent variable. Suppose, you want to remove this variable, then the arrow head you see that is immediate, the arrow head takes the reverse direction, so click here, it is going now.

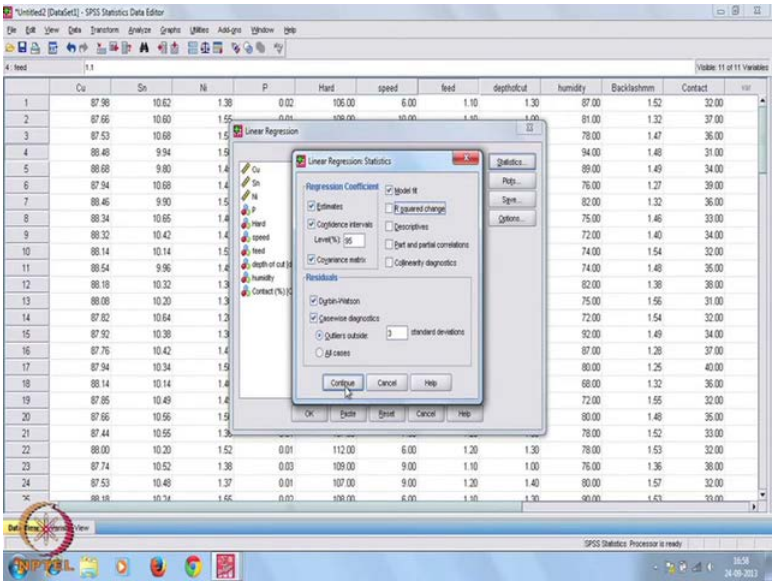
So, if you want contact, contact click contact come here like this; if you want to remove do like this, so let us take backlash. Now, what are the independent variables, when you

click on independent variables see there is no variables selected so far, so let us select all the variables, so copper to humidity, these are the independent list, so click here.

Now, question is what method you will use here? The method is enter stepwise, remove, backward, forward, these are all basically different types of treatment for the independent variables inclusion of, if you say enter all variables, means all independent variables will be taken at a time and the estimates will be given.

If you say no you want to go for stepwise regression, then depending on the certain criteria, that inclusion, exclusion, criteria, that stepwise regression will be performed. If you go for forward or backward, forward selection, backward elimination or forward selection, backward elimination, this type of model it is possible. So, let us take only enter, because we want to see first enter, once you make enter that means you are considering all the independent variables simultaneously.

(Refer Slide Time 33:15)

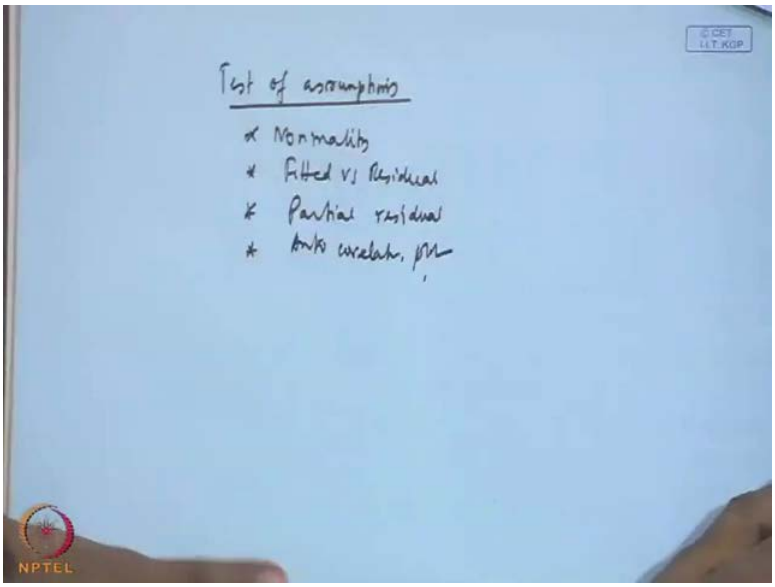


Then you want to select some of the statistics, what are the things you want, under this regression coefficient and then the residuals related things are given. So, you want the estimate, you may be interested to know the confidence interval, now what is the confidence interval level? 95 percent, 90 percent, if you want 95 percent, fine. So, under regression coefficients first you click on estimates, if you do not click on estimates it will not be displayed. So, confidence interval, if you are interested in covariance metrically,

covariance matrix, model feet, r square change, descriptive part partial correlations, all those things are there.

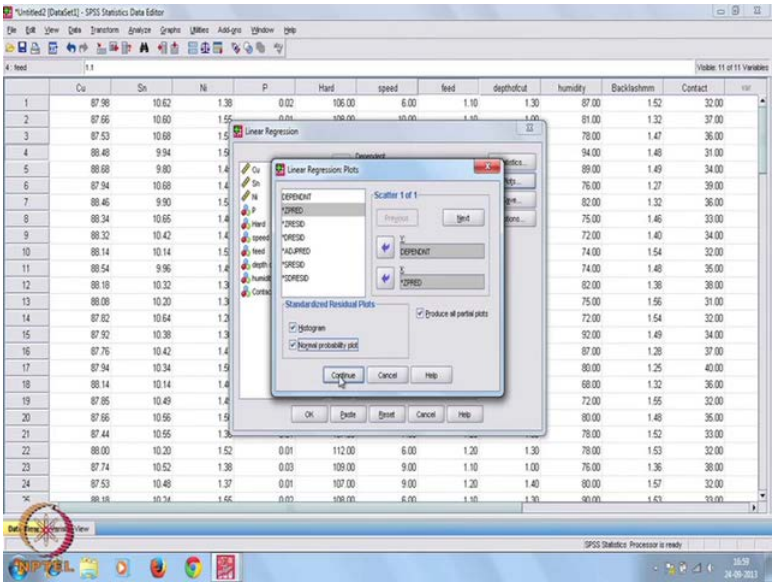
You may be interested to know what is the auto correlation, the durbin watson measure, you may be interested to know the case wise diagnostics, out layer, outside 3 standard deviation. So, now you may be interested to know the r square change also, then you click r square change, if you are not interested forget this, I think model feet estimates and confidence interval and the some out layers and residuals related information is very important, click. Then there are plots, what type of plot you want to see? In a regression, multiple regressions what are the plots I have said you?

(Refer Slide Time 34:50)



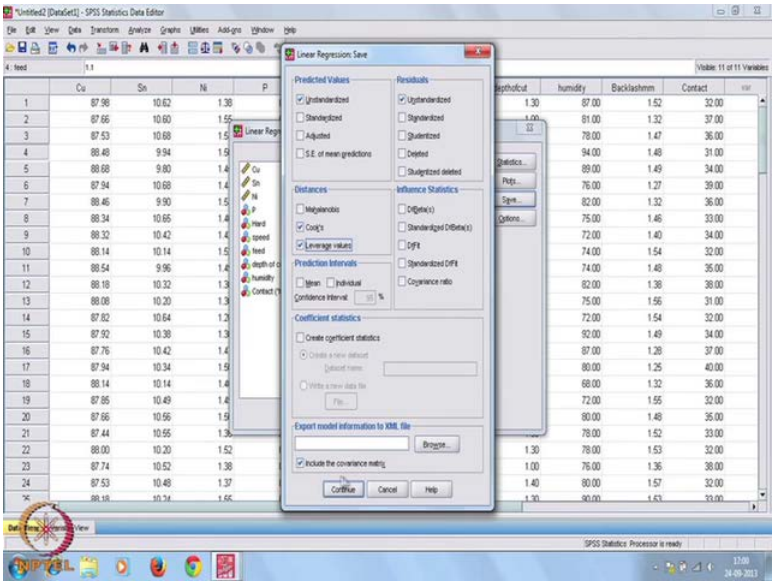
I said you that test of assumptions if you can remember, test of assumptions, then there are many plots we said, normality plot, normality of error terms, then we said the fitted versus residual, then we said partial residual plot, then we also said that auto correlation plot, so similarly many plots you have seen.

(Refer Slide Time 35:29)



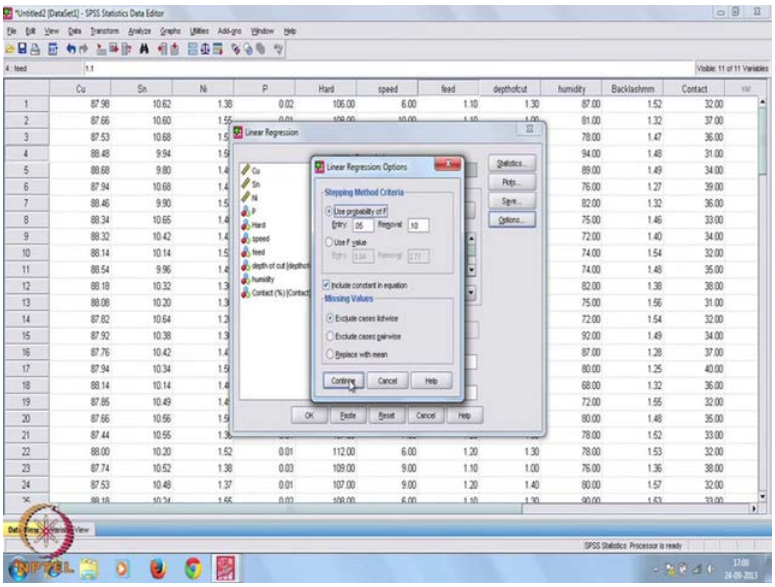
Now, let us see that we want to find out that linear regression plots here, what are the things given the dependent and independent, basically what you want dependent variable, now in this independent side which one you want? Predicted, residual, adjusted residual, student that student residual, so many things are there, let it be the predicted one. Now, you go for produce all partial plots, then you may be interested in histogram, you may be interested in normal probability plot, so you continue.

(Refer Slide Time 36:14)



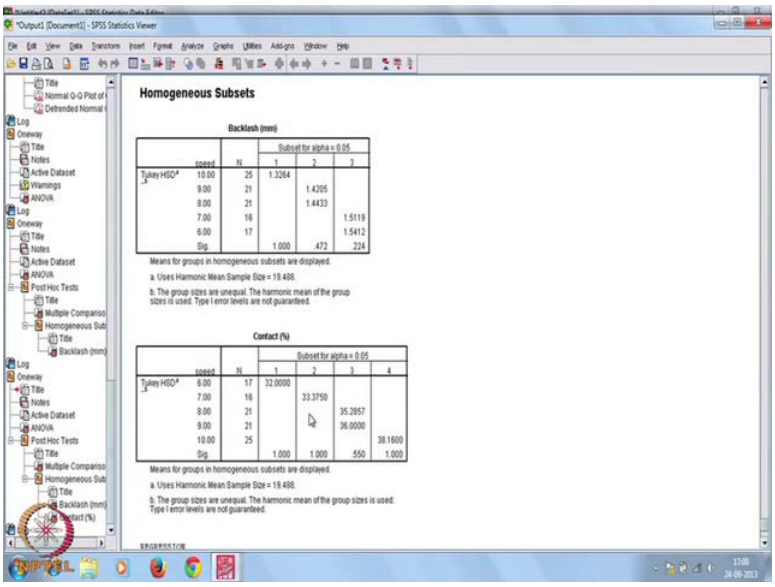
Now, there is save option, in save option means in what will happen? Suppose you I am saving unstandardized predicted values, let it be that residual, unstandardized residuals we want to save, model diagnostics. Also suppose cooks distance you are interested to know that leverage values also you are interested to know, ok this is fine.

(Refer Slide Time 36:40)



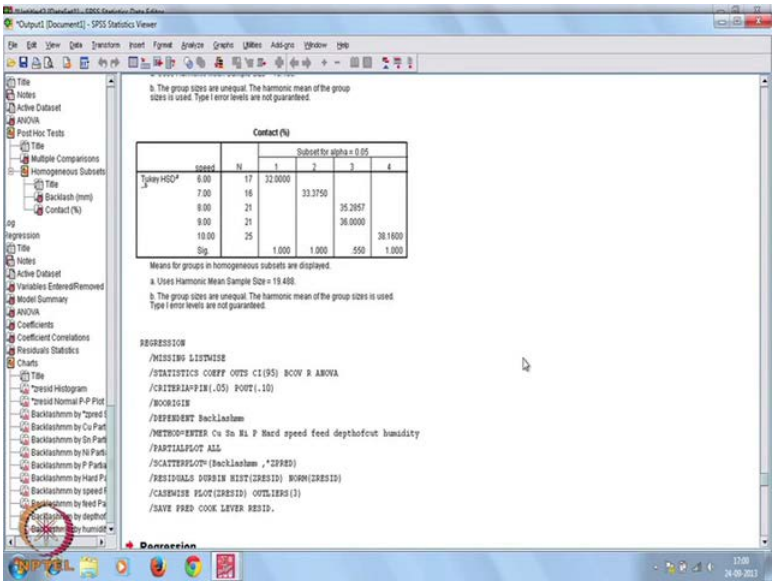
Now, options there are large number of all options also use probability f for entry and removal, then how to handle missing values these things are there, whether you want a constant in the equation or not? That is also there, ok click.

(Refer Slide Time 36:56)



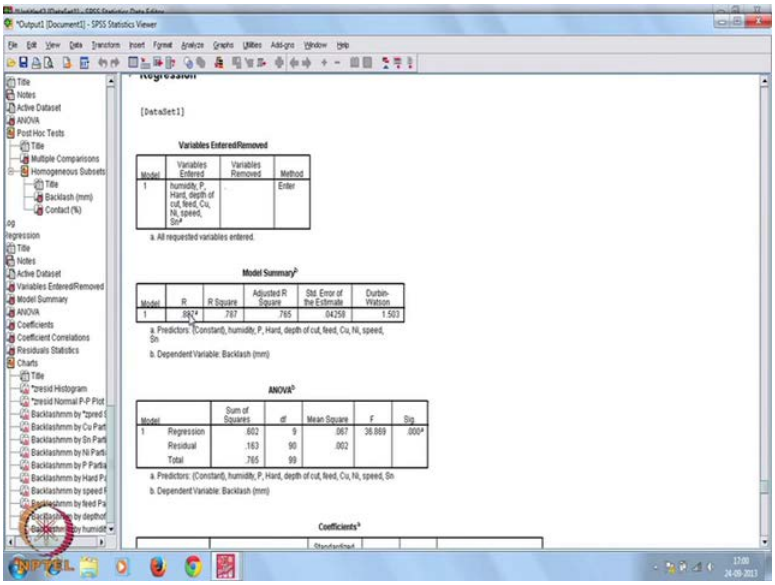
Then once you click, what is happening you see it is running I think ok.

(Refer Slide Time 37:07)



So, now regression, variables we say entered, so all independent variables are entered here.

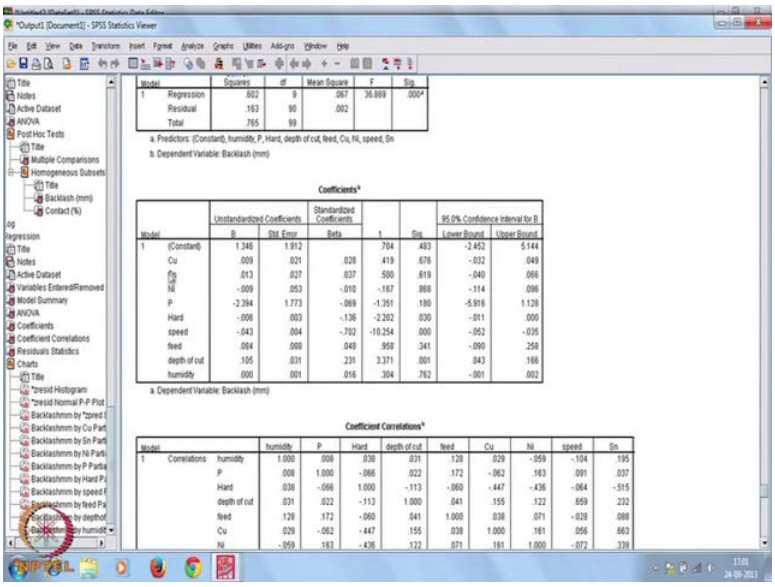
(Refer Slide Time 37:15)



What is the r square value? That is 78.7 percent, I think you can recall my class there I said that for backlash78.7 percent variability each explained by the regression model and adjusted r square is 76.5 percent. Now, see the analysis of variance table is coming here, this table is very important, because this will tell you that to the hypothesis h 0, beta j is

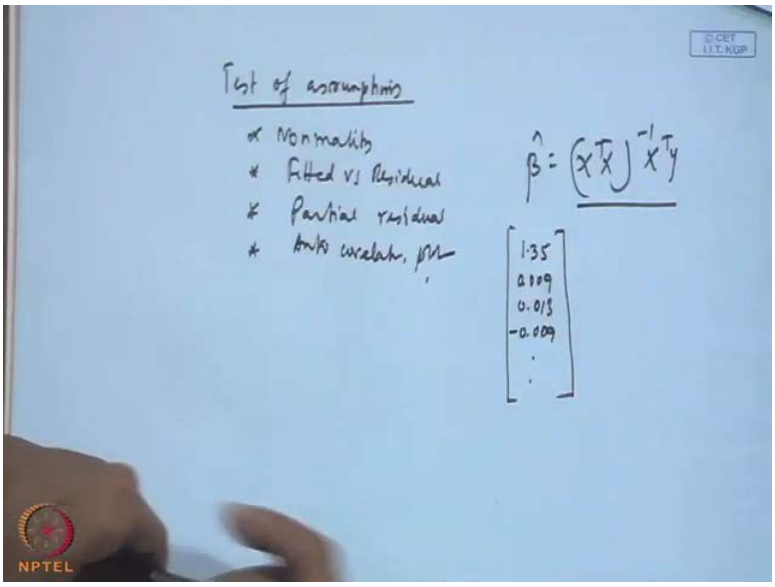
equal to 0 and h 1 beta j not equal to 0 for at least one of the regression coefficients, one of the variable influence is not 0.

(Refer Slide Time 37:57)



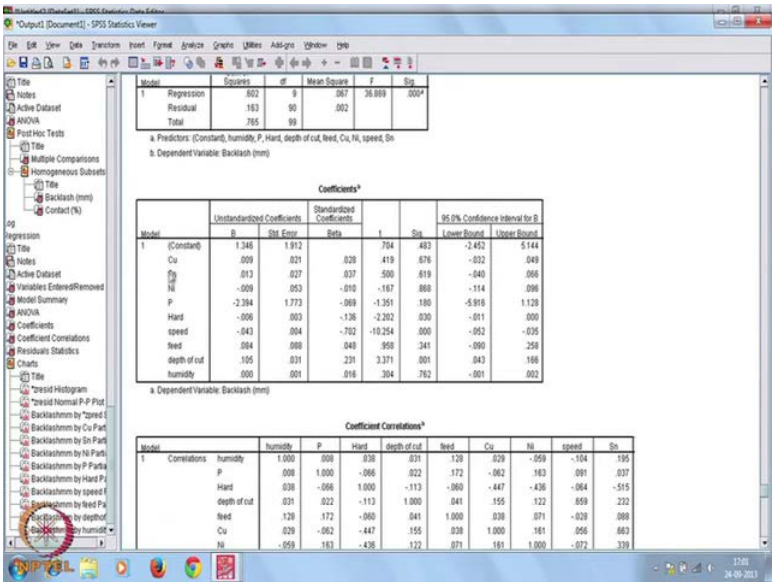
Then you see this is the parameter estimates, so constant, copper, silicon, all those things and these are the beta values.

(Refer Slide Time 38:04)



These are beta values means these values only beta cap equal to X transpose X inverse X transpose y, so this beta cap these values, these values you are getting here 1.35 then 0.009, 0.013, then minus 0.009, like this values you have got.

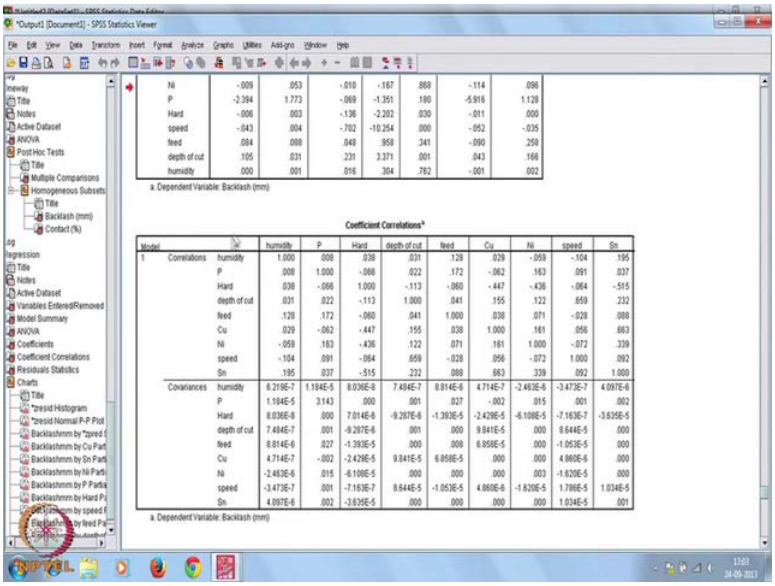
(Refer Slide Time 38:32)



Now, come back to this SPSS output, so see this is the beta, then standard error, then standardized coefficients, mean you have taken original variable values. You can standardize the variables and then the standardized regression coefficients you will get that ((Refer Time: 38:53)) value also you will get and significance, if you see the significance level and if we consider the 0.05 level is considered to be acceptable to reject the null hypothesis.

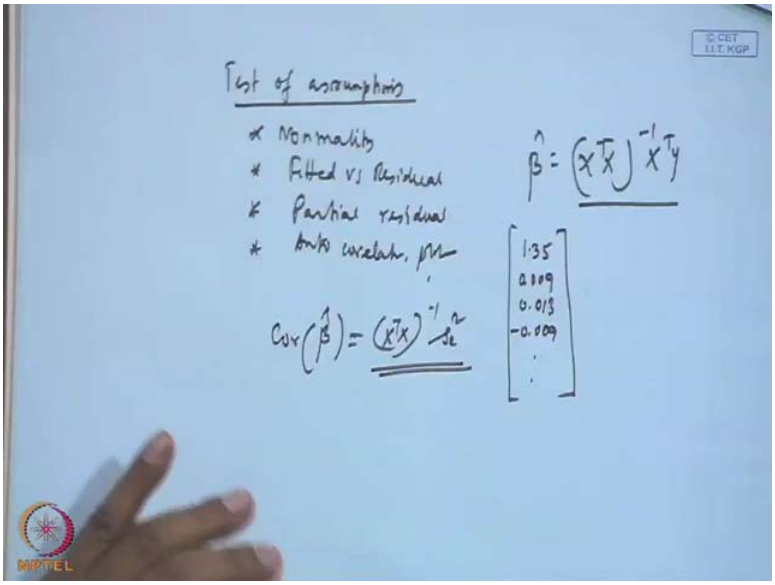
So, then you see that hardness this one here this hardness, if I click this see this one is 0.30, so this is significant. Second one speed is significant and depth of cut is significant, speed depth of cut, then speed and hardness these are significant, so we have seen earlier also that is that has happened.

(Refer Slide Time 39:41)



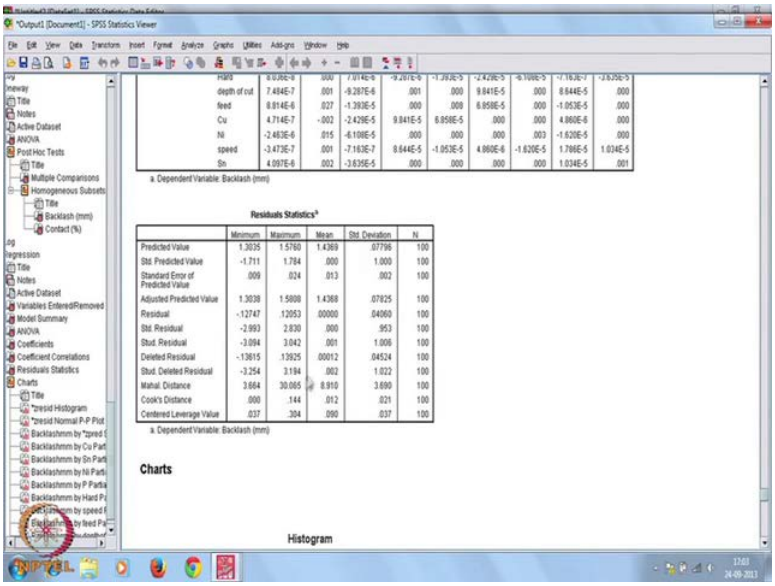
Now, this coefficient correlation, so when each coefficient whether they are related or not that humidity to humidity like this, this coefficient matrix also given, the covariance matrix between the coefficients also given. I think you know that covariance each covariance of beta.

(Refer Slide Time 40:00)



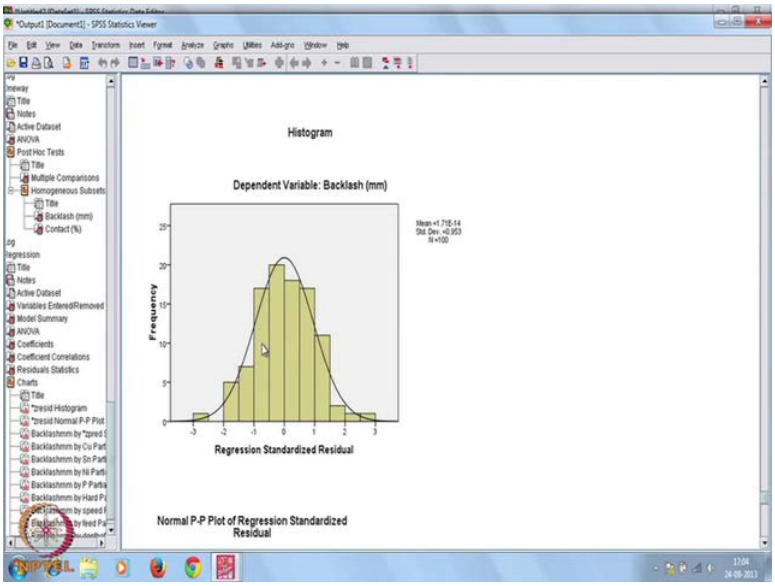
We have discussed covariance of beta cap, this one is X transpose X inverse, what? S square this one, so ultimately you are getting these values.

(Refer Slide Time 40:17)



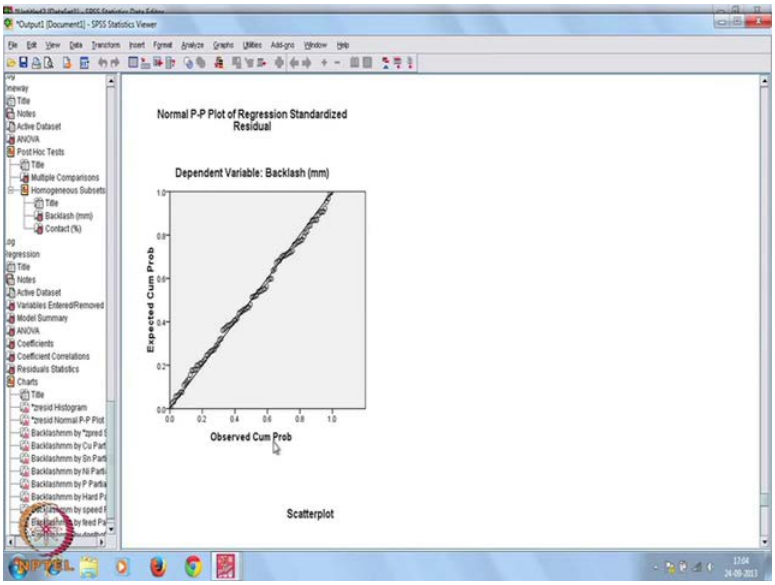
Now, residual statistics you see that predicted value, standard predicted value, standard error, all those things has given here, mean standard deviation, everything is given here. What you want almost everything will be available from this place is output and whatever the options available and the outputs available here that is sufficient for your work.

(Refer Slide Time 40:50)



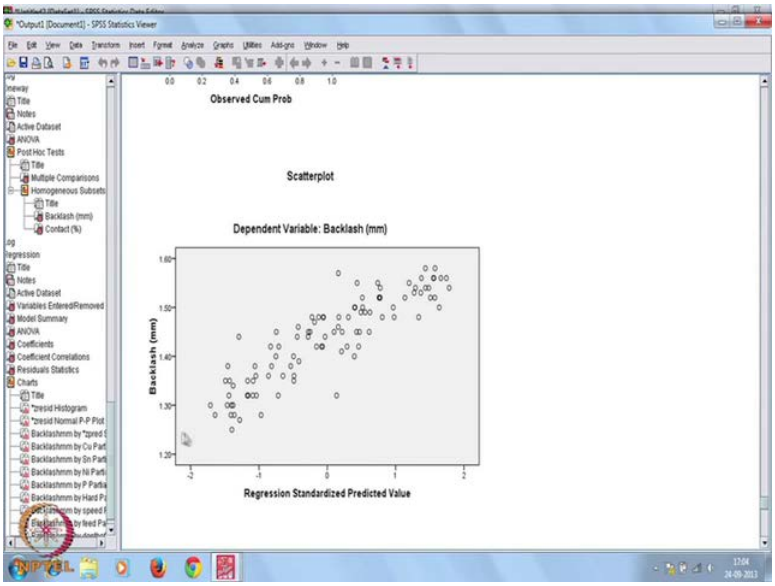
Then go to the charts, this is the error histogram residual, correct? So, the residual histogram part and it is with the probability distribution, also that normal distribution curve that is almost normal.

(Refer Slide Time 41:10)



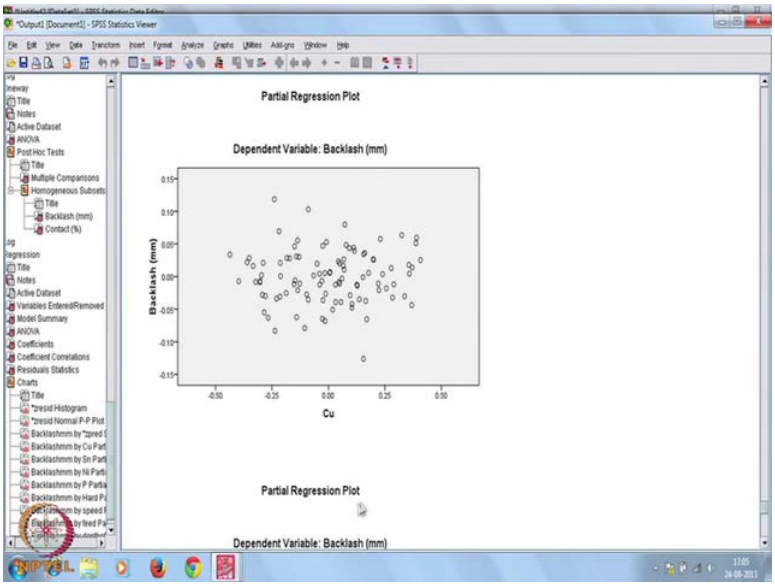
Now, you see the object probability versus expected cumulative probability versus object cumulative probability that is p p plot, this is also across the straight line, so that is fine.

(Refer Slide Time 41:25)



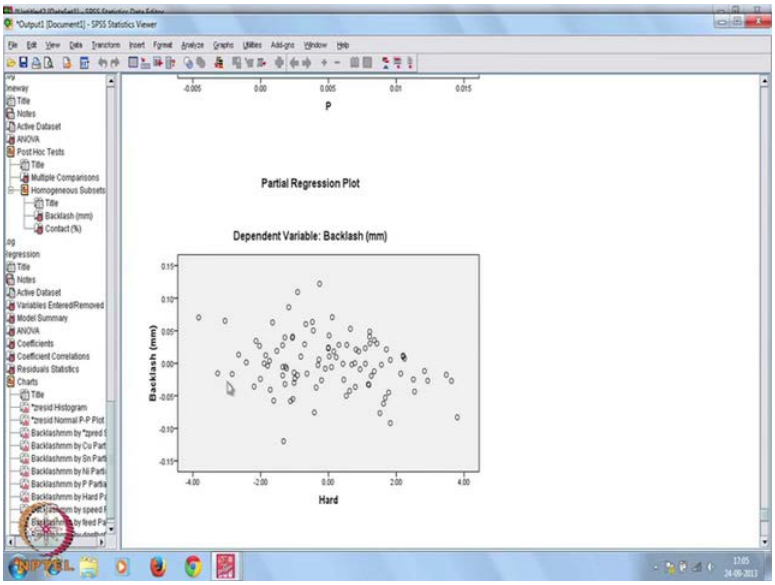
Now, you see backlash versus regression predicted values it is a linear, clear linear regression.

(Refer Slide Time 41:37)



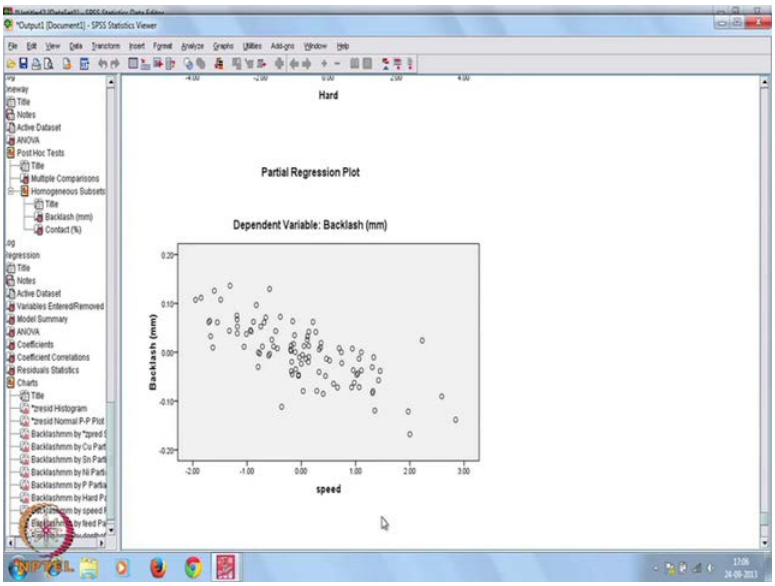
Now, you see that backlash versus copper that is basically dependent variable backlash that partial plot, if copper is significant what will happen? That means, either this copper will contributes positively or negatively, then there will be a regression line, there will be a straight line, if it is positive it will go up and negative it will come down like this, this or this, but the slope will be as per the beta value. So, here copper is not significant and that we have seen also.

(Refer Slide Time 42:16)



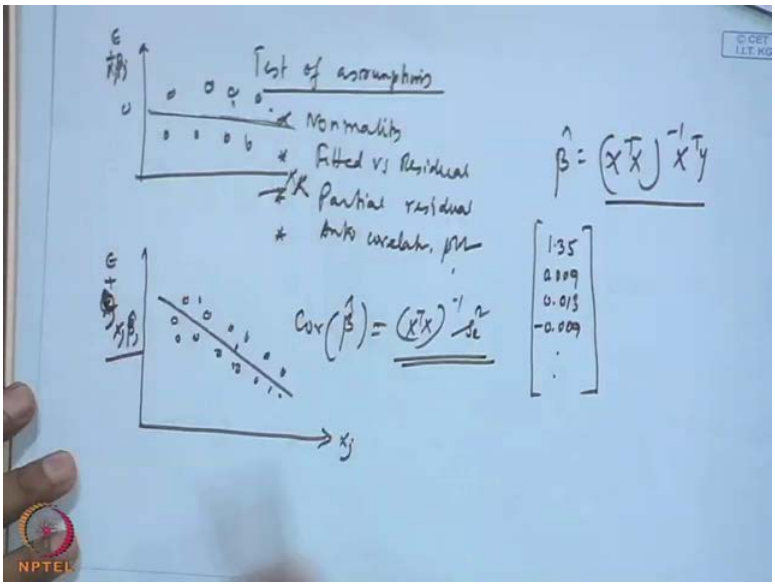
Similarly your backlash versus silicon similarly, your backlash versus nickel, then backlash versus phosphorous, then it is your hardness, hardness has little, you have seen that it is hardness is negatively contributing, you see that one that all the values that residual values, that in the partial, in this partial plot what is happening? It is one straight line can be framed here, this way, getting me?

(Refer Slide Time 42:52)



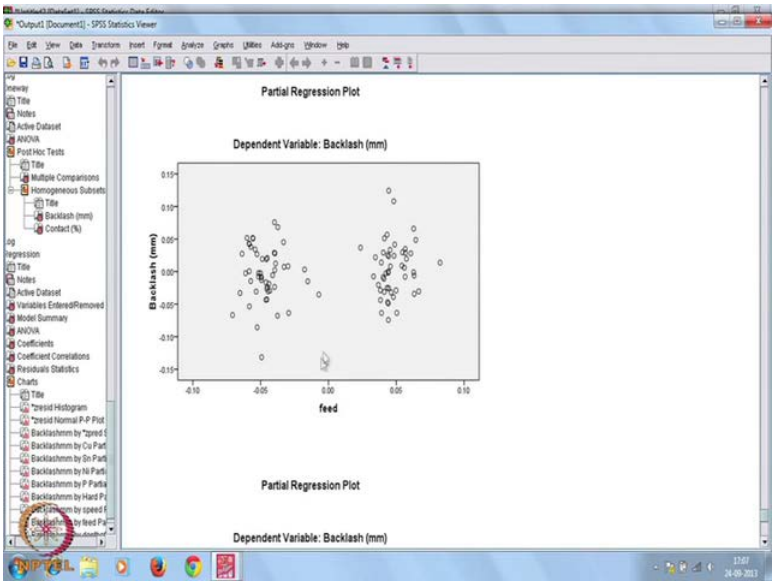
Now, you see that we say we have seen that feed also, sorry speed have significance, influencing variable, so this residual plot also that.

(Refer Slide Time 43:08)



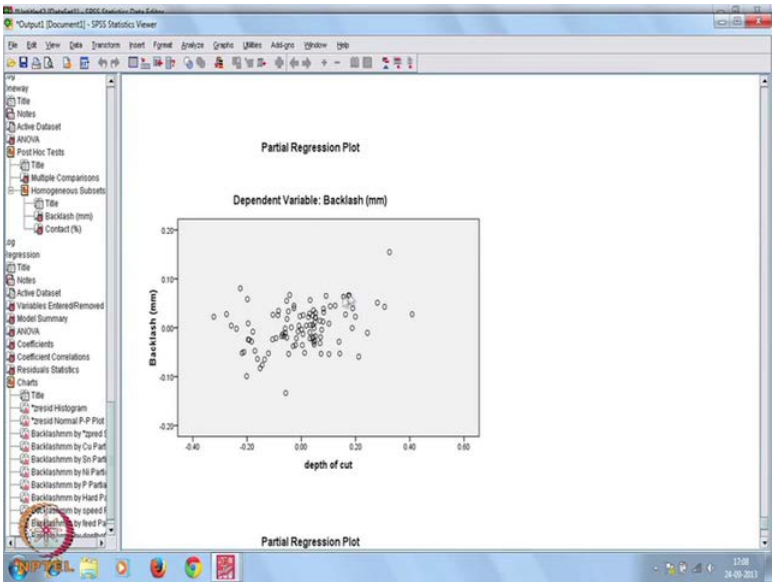
We have seen in residual plot, we said that when you go for residual plot this is basically X j, this side is error plus, no X j beta j cap. So, then as speed is having negative effect see the plot is like this, this is partial plot. When there is no effect you will be finding out like this, there is suppose one variable that epsilon plus beta j, X j, then there will be mean that 0 and it will be like this, if that X j variable has no effect, for example X K no effect, so that we are seen earlier.

(Refer Slide Time 44:12)



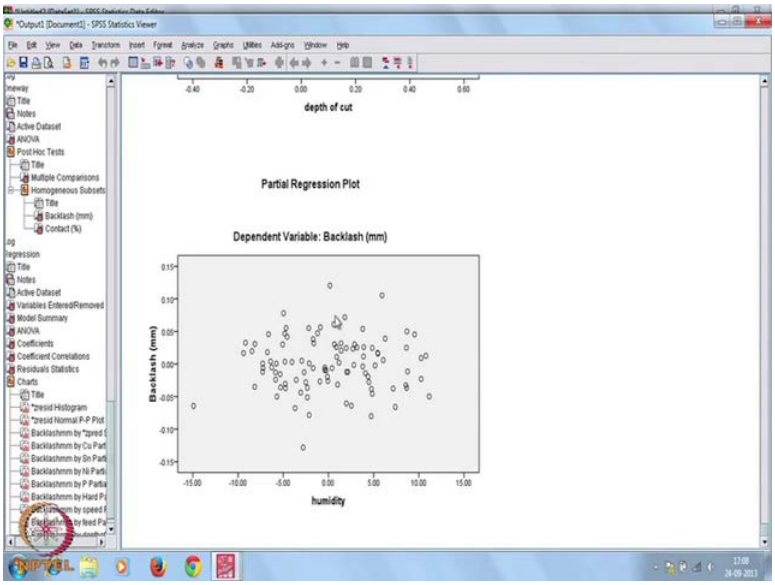
Now, let us see other plots what happen feed versus backlash, because basically that had only 1.1 and 1.2, that to in this two places feed is that errors are also scattered under two different groups. That means, when you feed will change from all levels to another level, so what is happening ultimately the backlash property also changing here and I think the way it is represented here, it raise some questions.

(Refer Slide Time 45:02)



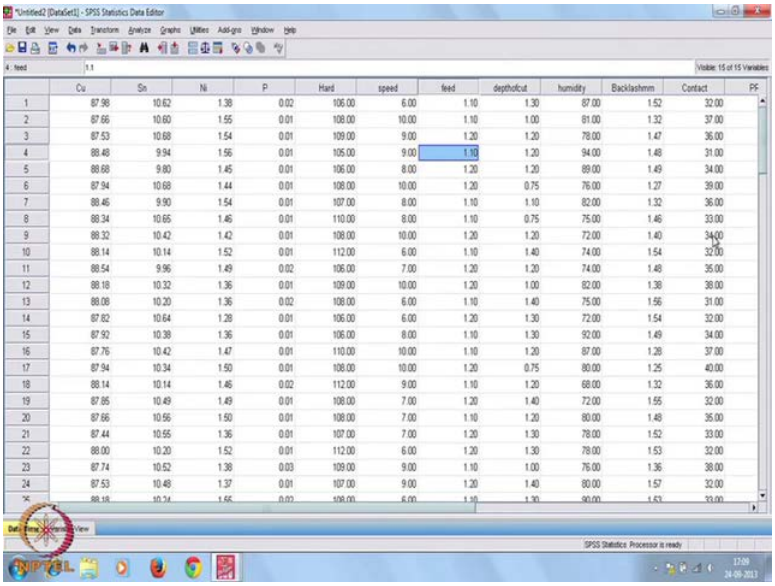
Then you go to backlash ((Refer Time: 45:03)) depth of cut, depth of cut has I think a positive relationship, so you are getting a positive line here, a positive slope will be there if you put a regression line here.

(Refer Slide Time 45:15)



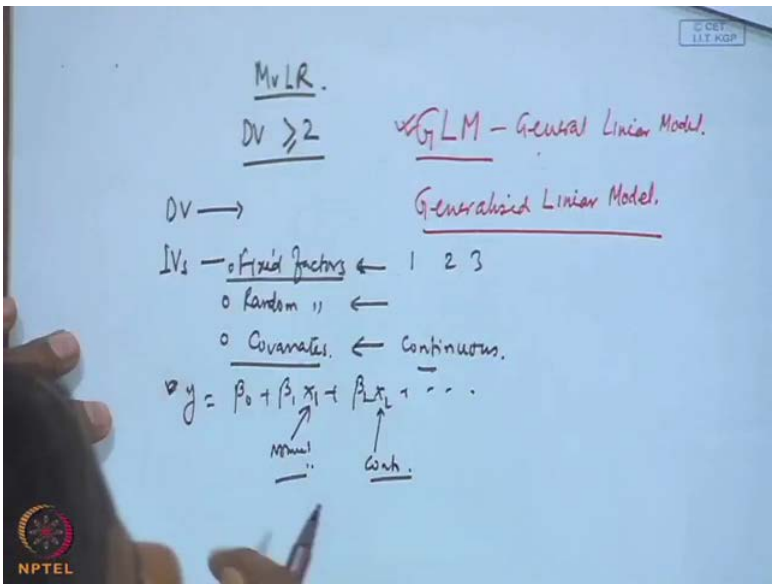
Then humidity has no effect on backlash, so you are getting straight way what you are getting here, it is a random one. So, there is no systematic part, so these are the things what you want from this regression.

(Refer Slide Time 45:40)



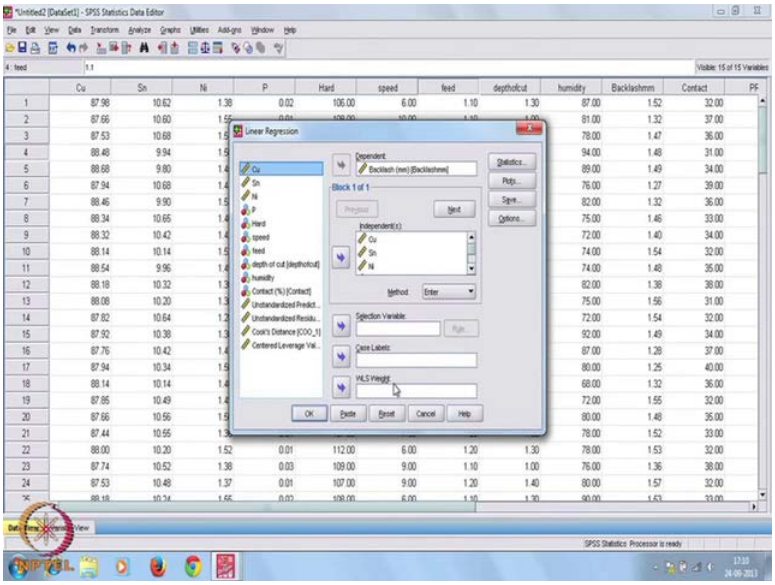
Now, then what happen to your data matrix you see that we have stored some of the something we have stored here, you see that our original variable was up to contact, then predicted value, residual ((Refer Time: 45:55)) distance and leverage points, these are stored here, getting me? So, SPSS will give you the output, as the data vector, for every output you will be getting stored into this spread sheet and which can be used later on.

(Refer Slide Time 46:37)



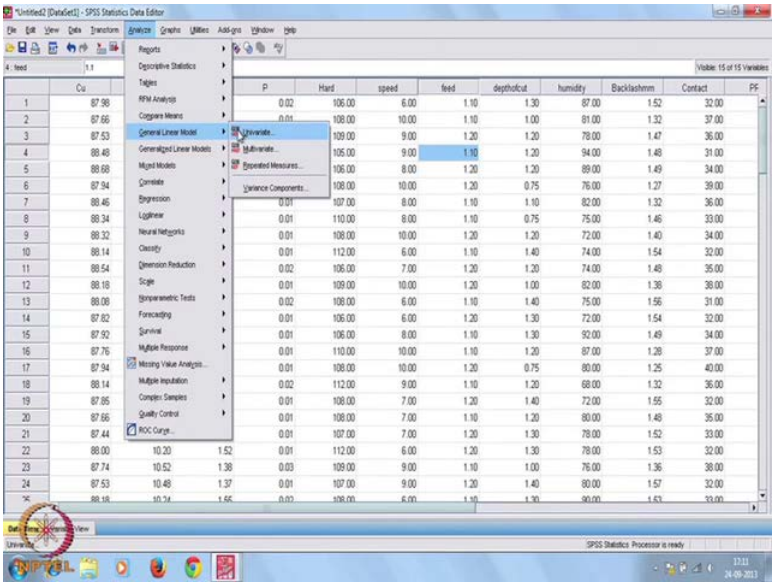
Now, we want to see that we want to do multivariate linear regression using this SPSS, in M v L R you have more than 1 d v, so it is basically greater than, d v greater than equal to 2 this case.

(Refer Slide Time 46:54)



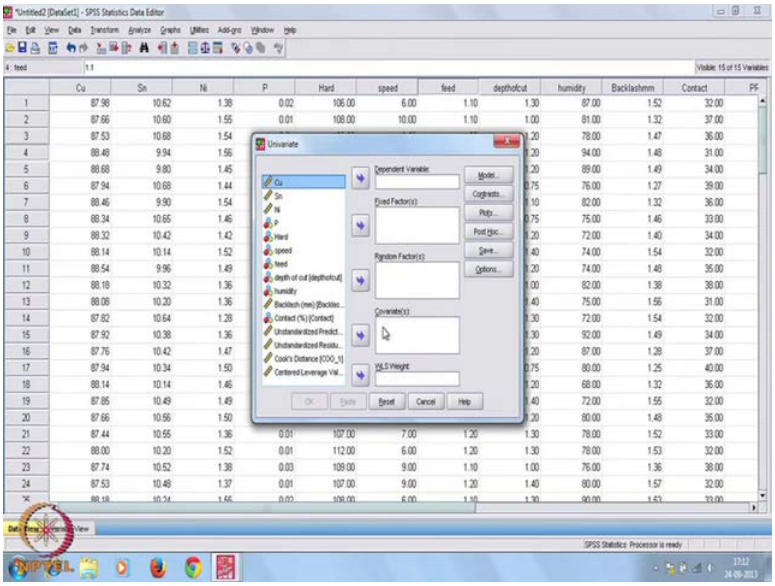
What you have seen here, you have seen here in regression, suppose go to analyze, then obviously as it is multivariate regression you will click on regression. Then you go to that linear, but there is no where multivariate part is there, so if you click linear, here option is only one variable can be given, under dependent, it is said the dependent only. So, that mean you cannot do under this regression, that regression link or regression option.

(Refer Slide Time 47:27)



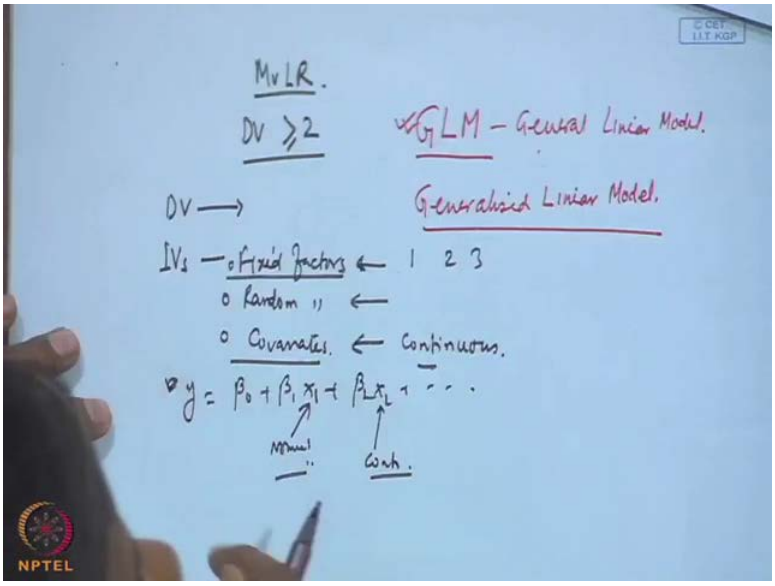
You have to go other place what is known as general linear model, if you go to general linear model, then there is univariate, G L M, general linear model is known as, probably known as G L M. So, general linear model, there is another linear model which is known as generalized linear model, so general linear model, then generalized linear model, so this one is general linear model, general linear model. We will go for general linear model not generalized linear model, there difference is generalized linear model that in this case mainly that when the dependent variable is not normal, that time that logistic regression or your Poisson regression those things are come under generalized linear model.

(Refer Slide Time 48:42)



For example, let us first do this general linear model univariate case, you click what is happening? When you are saying univariate, then dependent variable will be one, so under this there are fixed factors, there are random factors, there are covariates, getting me?

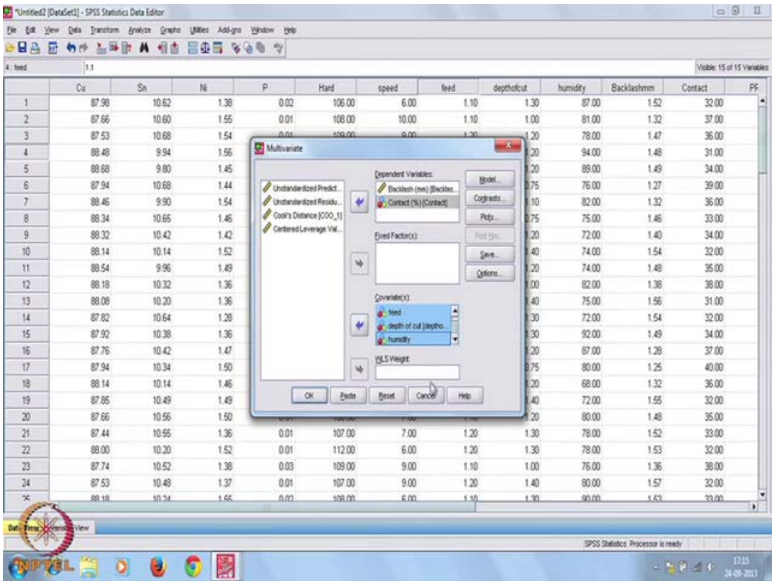
(Refer Slide Time 49:08)



So, three things are there, one is dependent variable, then there are under independent side you have fixed factors, you have random factors and you have covariates, what does it mean? Fixed factors basically the factor when it will be has different levels definitely,

but these values are fixed. Random means what will happen that value may change, suppose if I say that for level is 1 2 3 here, these are 1 2 3is fixed, but in case of random it is not like this, so random effect model is there. We are basically considering fixed factor model and covariates is, if the factor is continuous, getting me? For example, my dependent variable is y d v is y, suppose this is beta 0 plus beta 1 X 1 plus beta 2 X 2 like this, suppose this one is continuous measurement and this is different level, that means categorical or nominal difference, correct?

(Refer Slide Time 50:37)

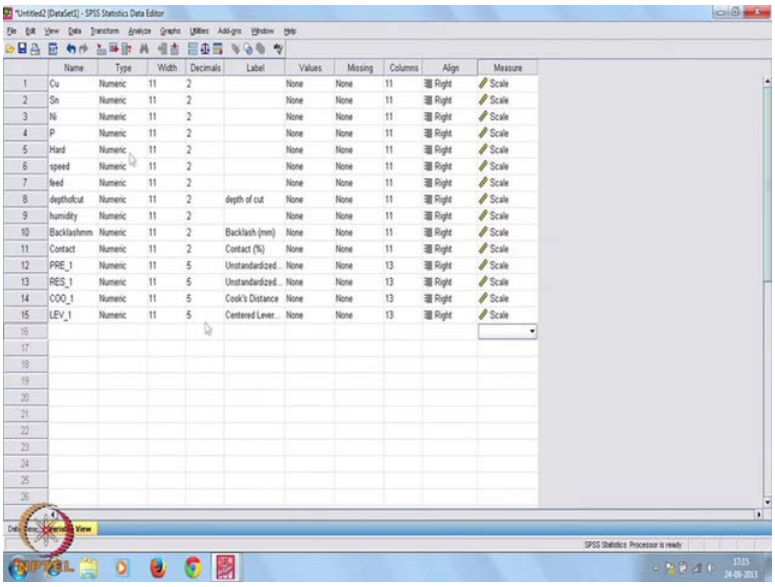


So, then you accordingly you had to put, where you want to put, so our case is dependent variable is backlash put here and we are interested in the covariate side, because all other variables, our all the variables here if I take all up to humidity under covariates what will happen you see. Then the model is there pool model and custom model, there will be different interactions, two ways, three way, all interactions are there I think. Let us go to the multivariate side, the similar nature, because I am interested to show you the multivariate part. Then what you will do? You cancel it, you go to analyze, then go to general model multivariate, the difference is everything remain same difference is dependent variables, more than one variable.

So, let us do like this, backlash, one variable and your another one is contact, second variable. Now, let me take all the variables in dependent side up to humidity under

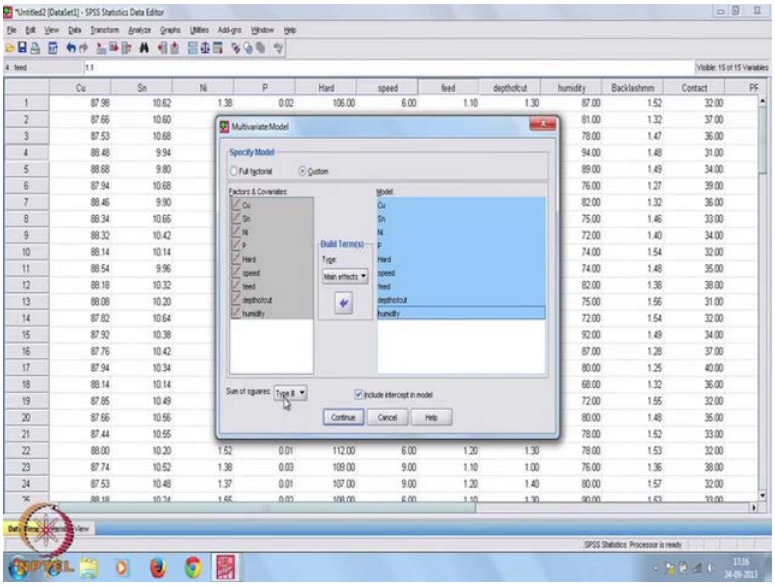
covariates, it is coming everything, but the symbol is giving here as if they are measured in nominal scale.

(Refer Slide Time 52:02)



So, I will go back and I want to see here, why, but all are numeric and why suddenly that symbol is given in this side type of things, I want scale, this one should be scale, this should be scale, this should be scale, this should be scale and this should be scale.

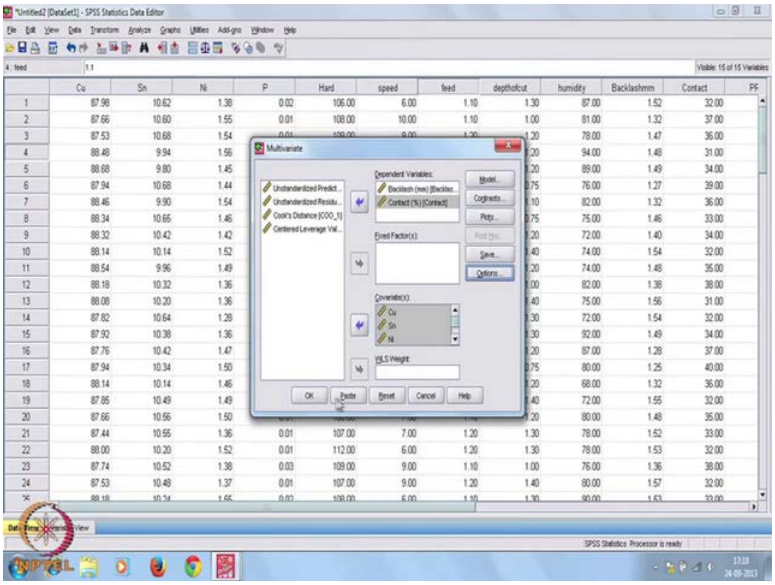
(Refer Slide Time 52:29)



Now, go to data, then go to analyze, go to general linear model multivariate, then ok fine, now backlash is your first dependent, contact is second dependent, then copper to

humidity put under covariates, what model you want? We do not want interaction here, we want custom model, here we want only main effects, then you click all the variables here, this, getting me? What happen you have taken all the independent variable, you have taken main effects only. That means, there is no interaction effect between the independent variables, if there is interactional effect between the amongst the independent variables, they are not truly independent, so that we are going for multivariate regressions where i vs are truly i v.

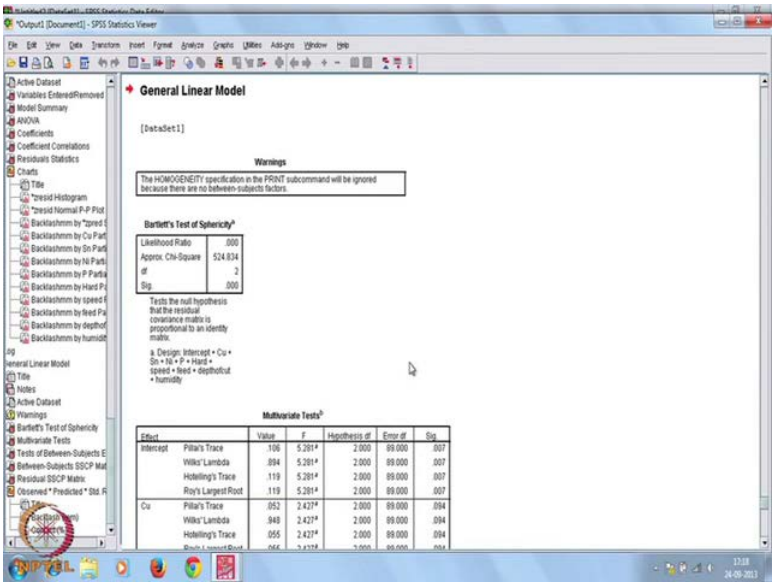
(Refer Slide Time 53:36)



So, then you continue and usually we used type three sum up square, include intercept continue, then contrast we do not require to have any contrast in this case it is that, then you go for plots, what plot you want? If you want I think here no factors we have considered, so plots are not that is why not given. Now, your save options are there, you can save unstandardized predicted values, unstandardized residuals, cook's distance, leverage values.

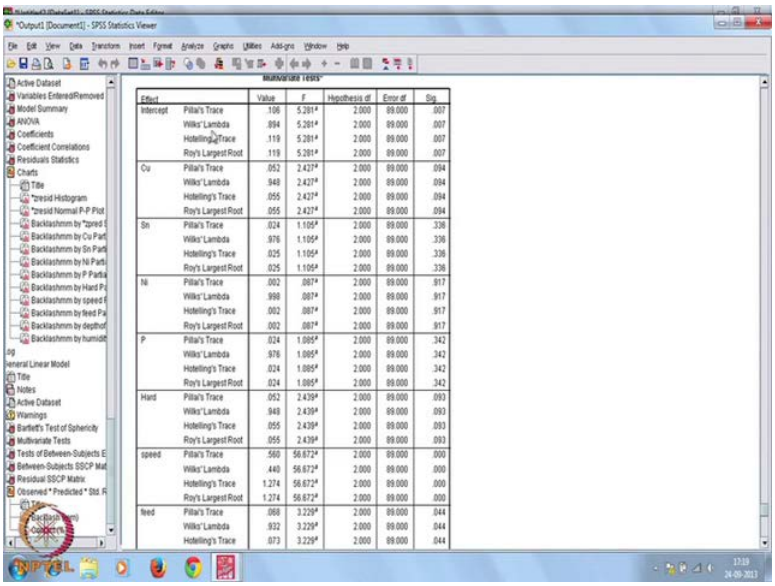
And you can create coefficient statistics like in regression that everything is possible here, there is option also, you may be interested to know the factor factor interaction is nothing is there, so there are so many things. For example, I am interested to know the s s c p matrices, residual s s c p matrices, then homogeneity tests might be interested to know, then residual plots, let it be like this, so click.

(Refer Slide Time 54:35)



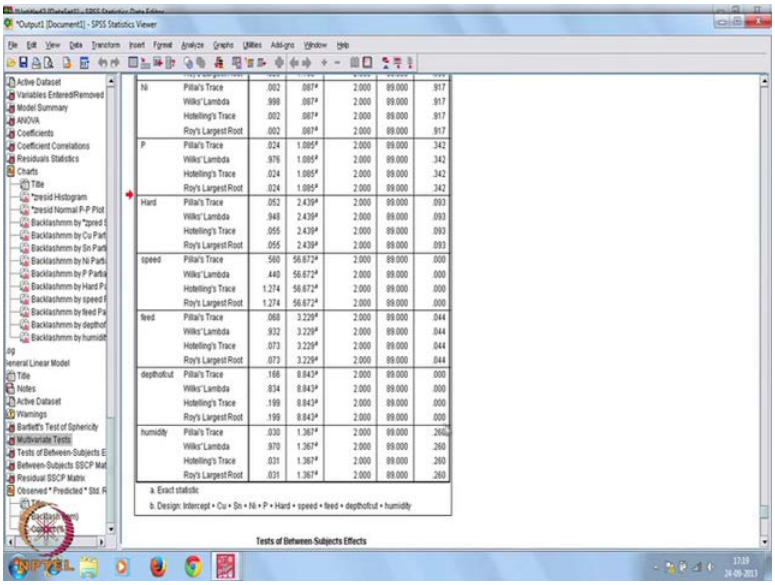
Now, first is Bartlett's test of sphericity, it says that this is 0 that hypothesis test, what is this that there is what is the Bartlett's test of sphericity? That data is, that why you are able to here that they are not deleted here. So, we are basically it is it is 1 0 0, then 0 is this tests the null hypothesis that the residual covariance matrix is proportional to an identity matrix, this one is rejected. Now, design intercept this is our model beta 0 plus beta 1 x 1 like this, this is our model.

(Refer Slide Time 55:32)



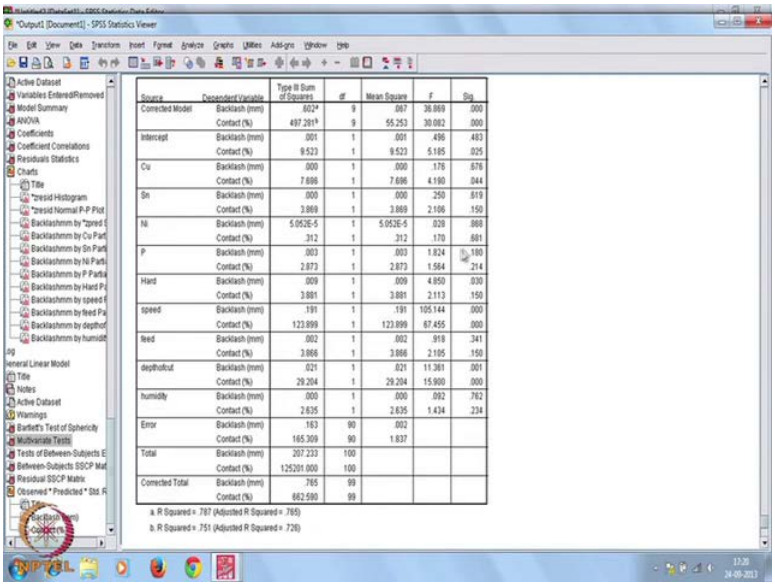
You can remember I said in multivariate regression Pillai's trace, Wilks lambda, Hotelling's trace, Roy's largest root, these all values, for every variable it is computed and they are significant level tested 0.05 level copper is not significant, silicon is not significant, nickel not significant, you see this is the case. So, ultimately your finding speed is significant, speed then your speed 0.0004, then your speed is also significant here, then your depth of cut is significant, but immediately not significant.

(Refer Slide Time 56:21)



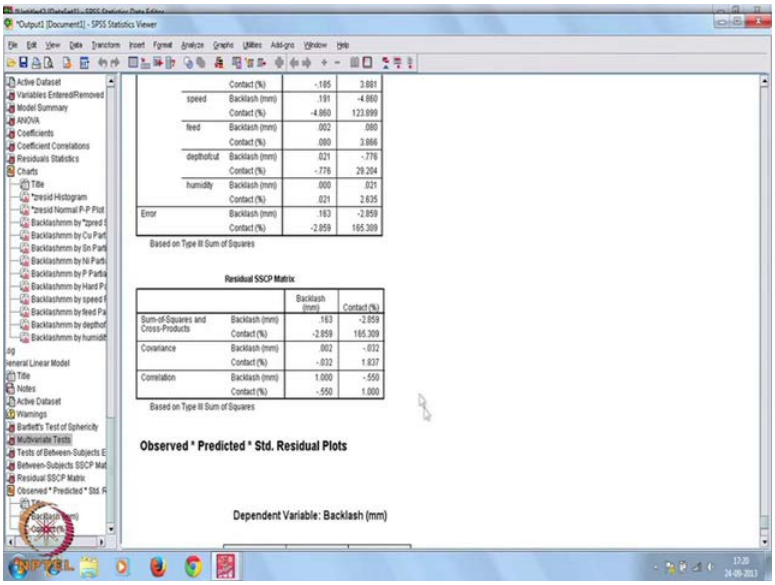
So, this is for what? This is for which variable they are considering? In totality it is consider, then design excel statistics like this and this is test between subjects, mean square, then your sum squares, mean squares like this.

(Refer Slide Time 56:44)



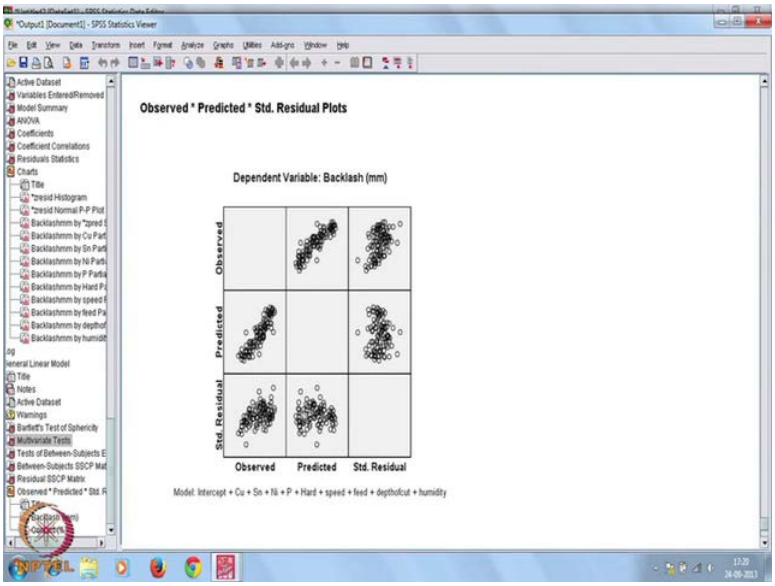
That sum square degrees of freedom individual variable that contributions it is also found out and their significance label is there, then you see that s s c p matrix between subject s s c p matrix is related to regression.

(Refer Slide Time 57:00)



Then residual s s c p matrix they show the error one. So, now for backlash, for contact, this is cross products, then covariance and correlation it is given here ok.

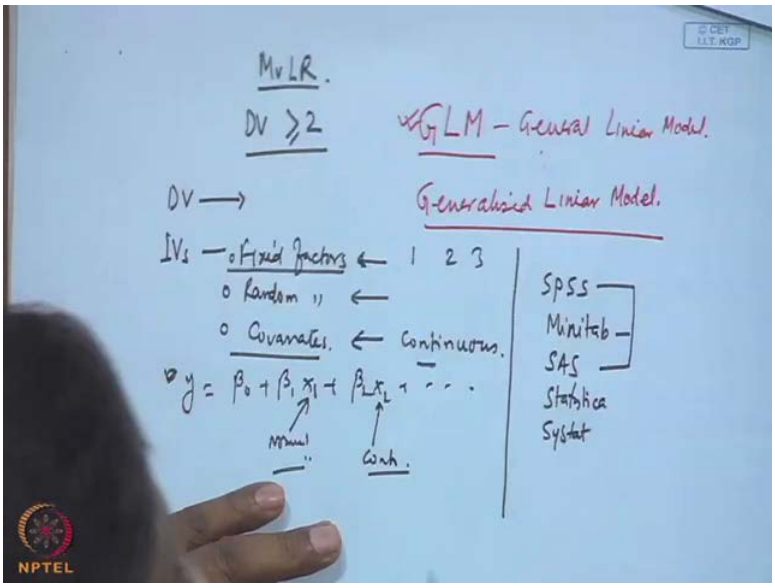
(Refer Slide Time 57:17)



Now, see that dependant variable (Refer Time: 57:18) is predicted, ultimately what we found out that observed versus predicted if you see there is relationship, so observed predicted observed nothing is there, that residual versus observed, when residual versus predicted that should we random, now that is what is happening here. So, similarly that for, this is for backlash, other one for contact and these are the things what you want, you are getting everything, so you are getting the statistics from adequacy point of view.

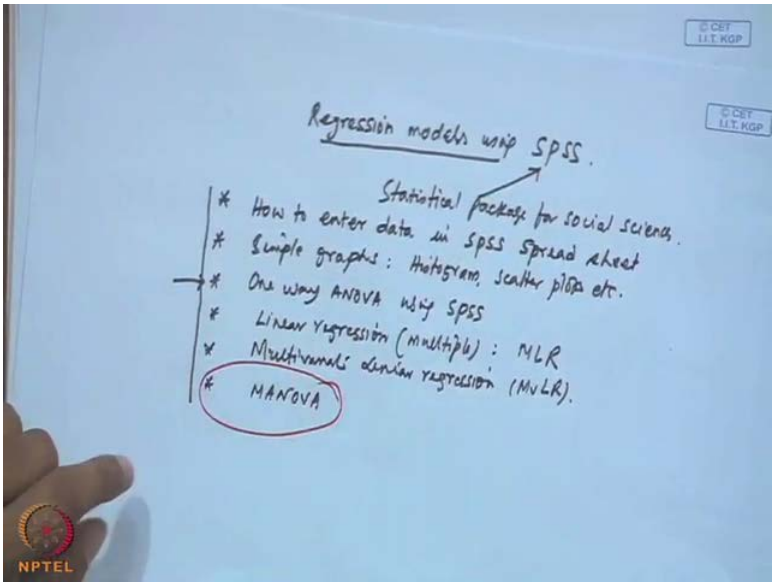
You are getting the individual regression coefficients and they are test which one is significant or not and you are also getting the model diagnostics related values like residual plots. So, if you know the theory and if you are really applying multivariate statistics, then SPSS is very good too, so you can use SPSS, you can use minitab as there, you can use as there.

(Refer Slide Time 58:33)



So, there are many statistical sub tires SPSS, Minitab, S A S, Statistica, Systat. So, I know these many stat graphics earlier few I was using long back, but I do not know the present status, but usually this SPSS, Minitab and S A S, these are available in our department and you can use those things in the laboratory. So, this is what is our multivariate linear regression using SPSS.

(Refer Slide Time 59:27)



So, ultimately we have covered many things, but only one thing we could not cover to, now this MANOVA part, I think in some other class I will show you that how

MANOVA will be used using SPSS and Minitab also we can think some class, some any time. Any question?

Thank you very much.