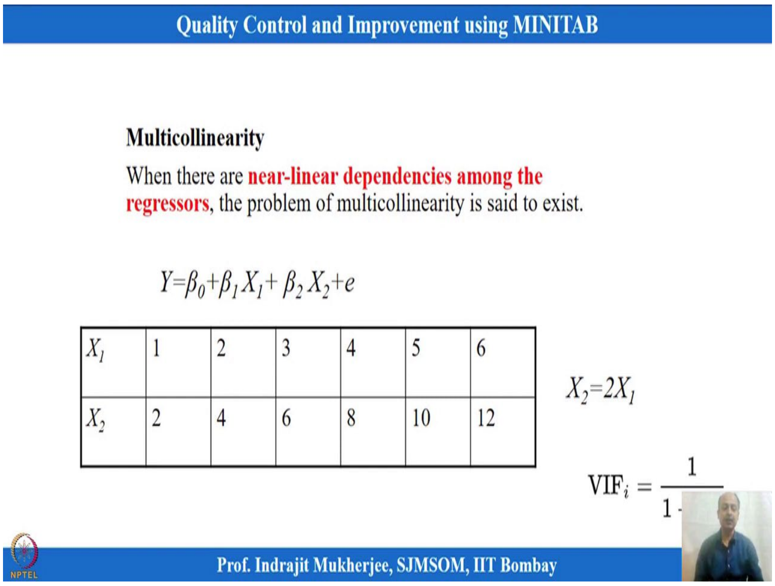
Quality Control and Improvement with MINITAB

Prof. Indrajit Mukherjee
Shailesh J. Mehta School of Management
Indian Institute of Technology, Bombay

Lecture - 27
Multicollinearity, Best Subset Regression, Multiple Regression, Basics on Design of Experiment

Hello and welcome to our session 27 on Quality Control and Improvement with MINITAB. I am Professor Indrajit Mukherjee from Shailesh J Mehta School of Management, IIT Bombay. So, previous session what we are doing is that, we are discussing about the multiple regression and multicollinearity problems like that. So, we will take another examples to again emphasize on the importance of how to tackle Multicollinearity like that ok.

(Refer Slide Time: 00:45)



Quality Control and Improvement using MINITAB

Multicollinearity

When there are **near-linear dependencies among the regressors**, the problem of multicollinearity is said to exist.

$$Y=\beta_0+\beta_1 X_1+\beta_2 X_2+e$$

X_1	1	2	3	4	5	6
X_2	2	4	6	8	10	12

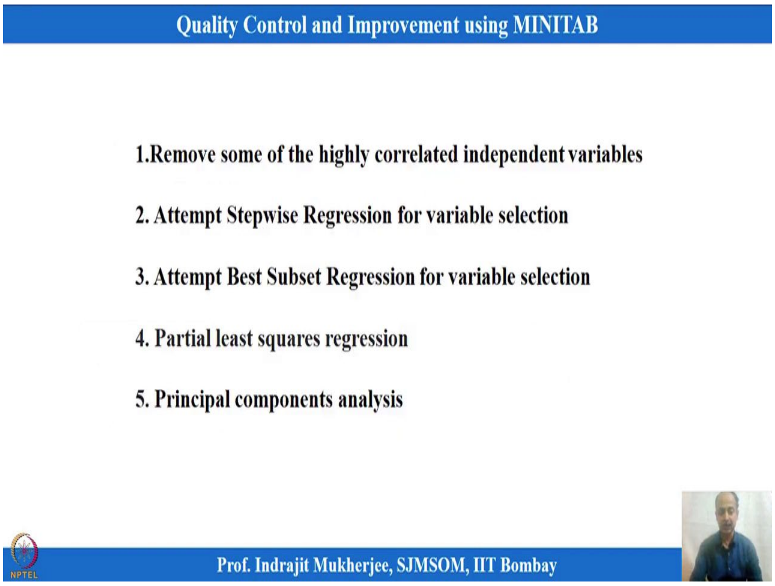
$X_2=2X_1$

$$\mathrm{VIF}_i = \frac{1}{1 -}$$

Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

So multicollinearity, what we have explained is that relationship between X variables that exist, and that can distort the relationship between what is and we are not able to generalize the equations which can be used in reality for prediction like that ok. So, we need to deal with this and variation inflation factor that we have to calculate and try to see that it is not more than 5.

(Refer Slide Time: 01:12)



Quality Control and Improvement using MINITAB

1.Remove some of the highly correlated independent variables

2. Attempt Stepwise Regression for variable selection

3. Attempt Best Subset Regression for variable selection

4. Partial least squares regression

5. Principal components analysis

Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

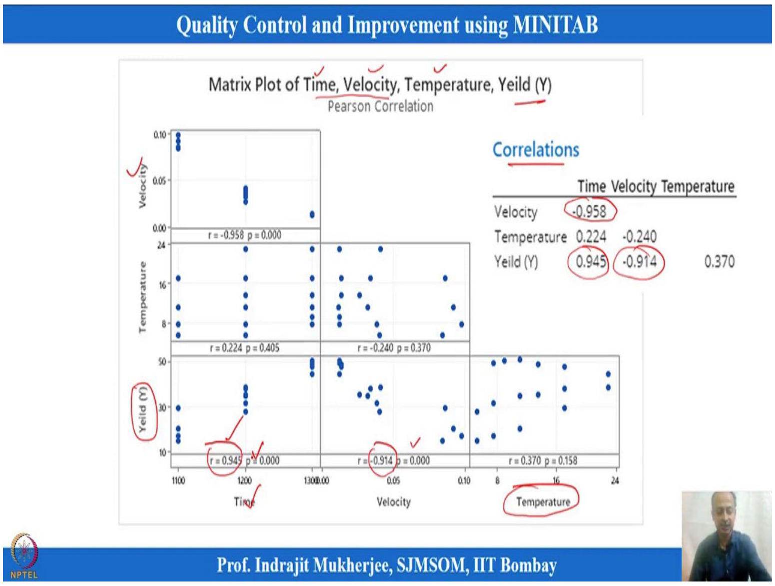
And, if it is more than 5 then we have to adopt some other means to deal with that. And how to deal with that? That we have mentioned over here. That means, highly correlated variables, one of them will we can take and the other ones can be eliminated. And, we can attempt stepwise regression method which takes care of which can suggest that which are the variables to be taken and we can reconfirm the multicollinearity exist or not in the final models like that ok.

And, we can also see best subset regression where multiple options exist because we do not want to stop only considering what stepwise regression gives. Because, some sometimes what happens say, alternate model is easy to control in real life; that means, those variables are easy to control which is suggested.

And if it is fine, if we have to sacrifice some amount of R square values or something like that, that is not a constraint for us in production processes, but variables which are very difficult to control and stepwise regression keeps those variables in the final equation then it becomes difficult for us to change that one, if we adopt that one, so in that case it becomes difficult. So, best subset regression gives you some more options to select the variables like that.

Then, appropriate methods like partial least square regression, principal component regression all these things can be adopted ok.

(Refer Slide Time: 02:27)



Quality Control and Improvement using MINITAB

Matrix Plot of Time, Velocity, Temperature, Yeild (Y)

Pearson Correlation

Correlations

	Time	Velocity	Temperature
Velocity	-0.958		
Temperature	0.224	-0.240	
Yeild (Y)	0.945	-0.914	0.370

Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

So, we will take one more examples where we have variables like, yield is the Y characteristics and this depends on time, velocity and temperature. These are the variable X that was considered to be significant or potential variables and we want to get a model out of this regression model out of this. All are continuous variable over here, so in this case.

And when we did the correlation analysis of this correlation analysis, what we observe is that velocity is highly correlated with time variables over here and then we can see that yield is highly correlated with time, velocity over here. So, but this P-Value will indicate whether which of the variable is highly correlated.

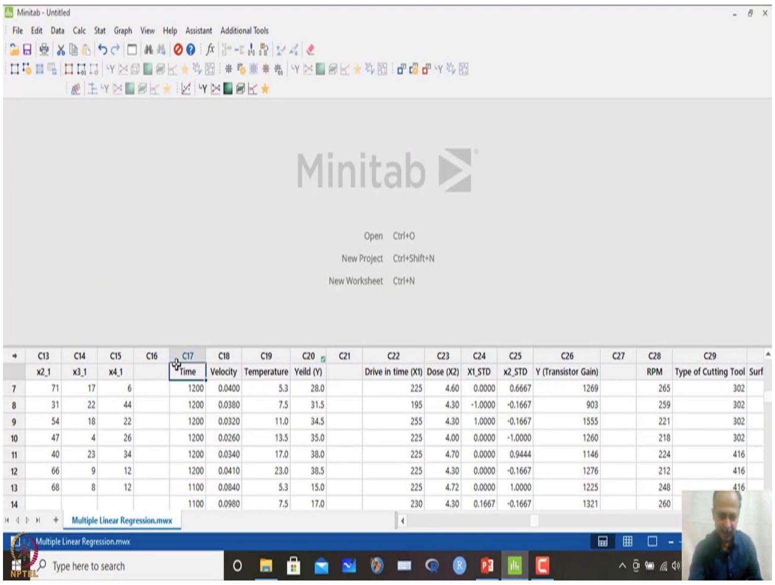
So, yield is one of the variables. So this seems to be significant over here, velocity is also significant over here, but temperature does not seem to be significantly influencing the yields over here ok. And, what we can see is that temperature with time is not significant, temperature with velocity is also not significant ok.

So, velocity with time is highly correlated, this with time velocity and time is highly correlated over here. So, this correlation matrix gives you some preliminary information, what is expected means when we run the regression analysis, what is expected in all these variables. So what we can see is that time and velocity are two variables which are highly correlated, so one we have to adopt over here. So how do we select which one to adopt over here? So, maybe when we do a trial and error basis.

If we have to select the variables the policy may be that, because yield is the correlation coefficient is 945 for time and for velocity is 914 maybe this variable we should select like that ok. Time may be the only variable and replace the velocity with the time like that.

So, yield has to be regressed with time and, if temperature is significant then we will include that one. So anyhow, this is the suggested one of the guidelines. So, which is highly correlated with the Y variable, we can select that one. And, we want to see that what happens if we select stepwise regression what happens like that.

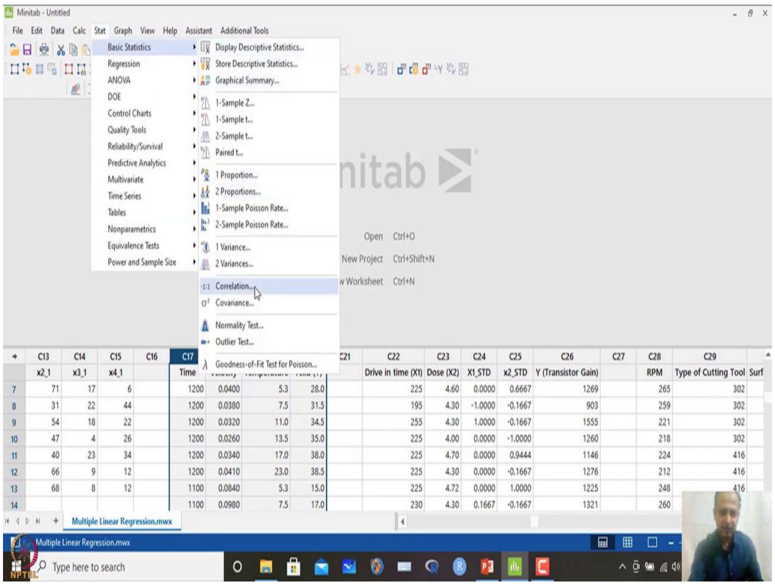
(Refer Slide Time: 04:38)



↓	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

So, this is the data set C11 to C20. This is the data set that we are having and we want to; we want to implement this.

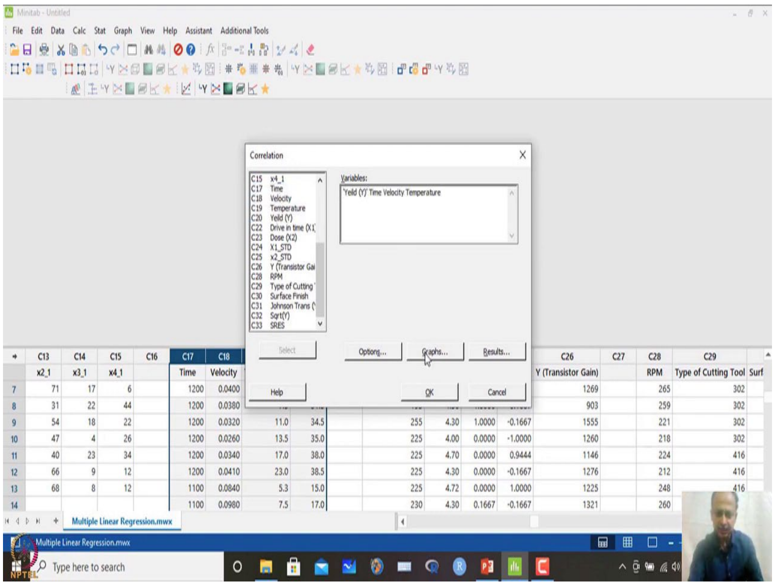
(Refer Slide Time: 04:48)



	C13	C14	C15	C16	C17				C21	C22	C23	C24	C25	C26	C27	C28	C29
	x2_1	x3_1	x4_1		Time					Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260	

So, what we have done is that basic statistics over here we have gone to correlation.

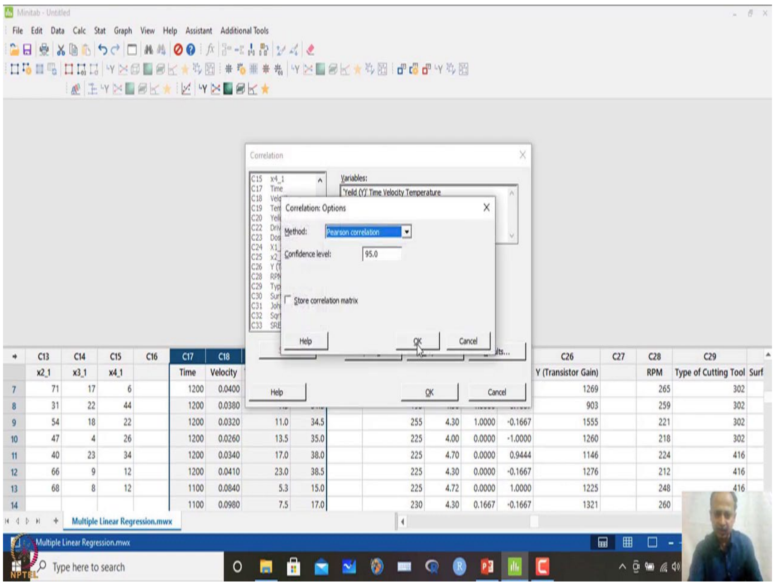
(Refer Slide Time: 04:52)



Variables: 'Yeild (Y)' Time Velocity Temperature

And then we have identified which are the variables over here. We want to understand the relationship. So, yield is the first variable time, velocity and temperature.

(Refer Slide Time: 05:02)



Correlation

Variables: 'Yeild (Y)' Time Velocity Temperature

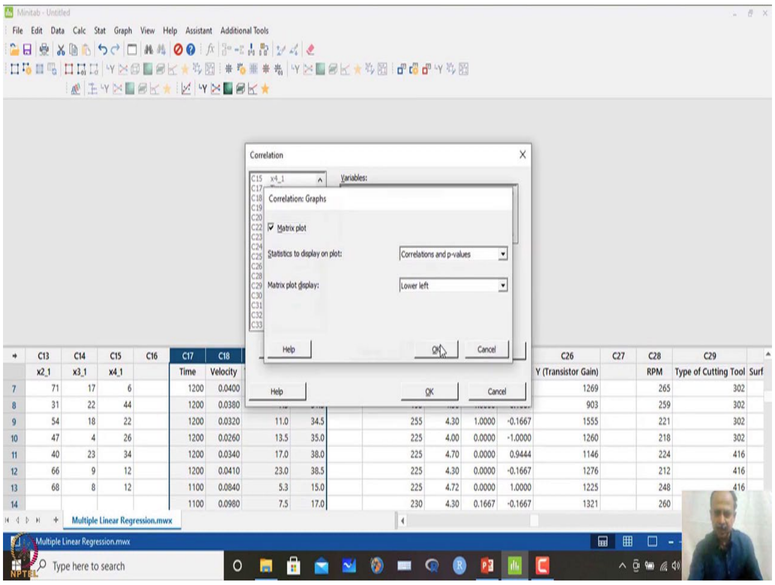
Correlation: Options

Method: Pearson correlation

Confidence level: 95.0

□ Store correlation matrix

(Refer Slide Time: 05:07)



Correlation

Correlation: Graphs

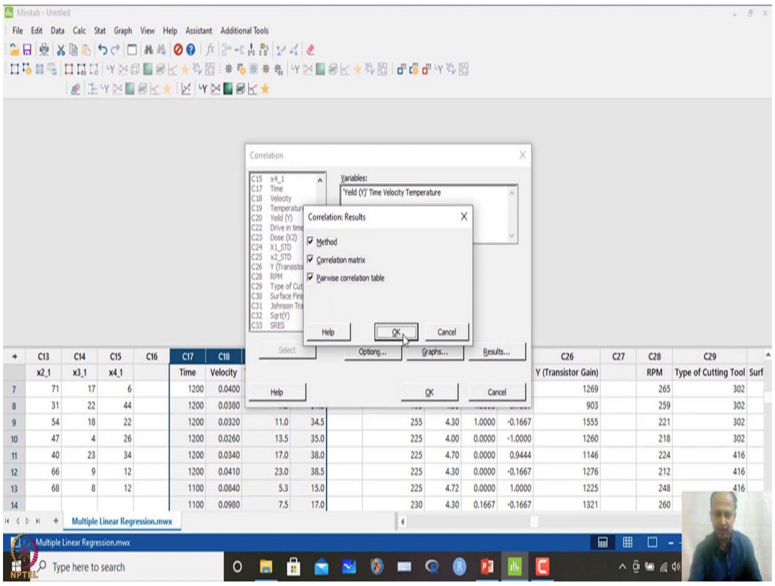
☑ Matrix plot

Statistics to display on plot: Correlations and p-values

Matrix plot display: Lower left

And, in options what we have given, Pearson correlation we want to see. And then in graphs what we want is that correlation with P-Values.

(Refer Slide Time: 05:11)



Correlation

Variables: 'Yeild (Y)' Time Velocity Temperature

Correlation: Results

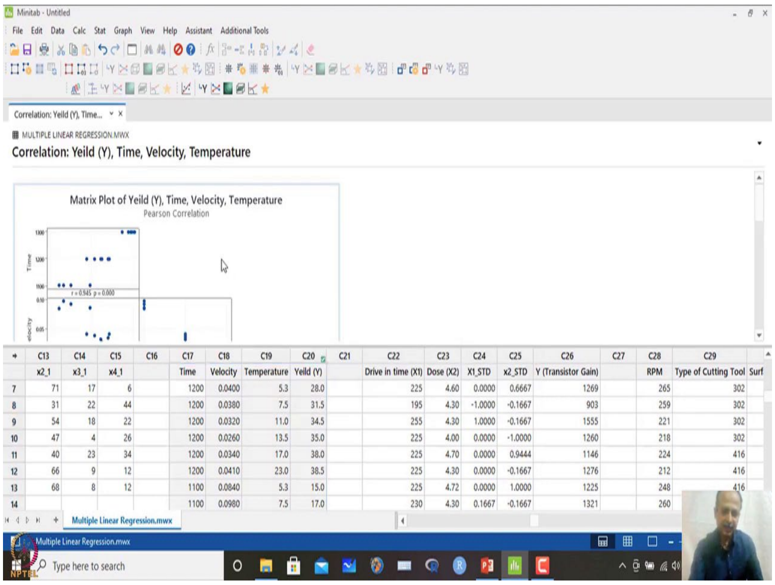
☑ Method

☑ Correlation matrix

☑ Pairwise correlation table

And if I click ok, and results what we have mentioned is we can also see pairwise correlation table.

(Refer Slide Time: 05:15)



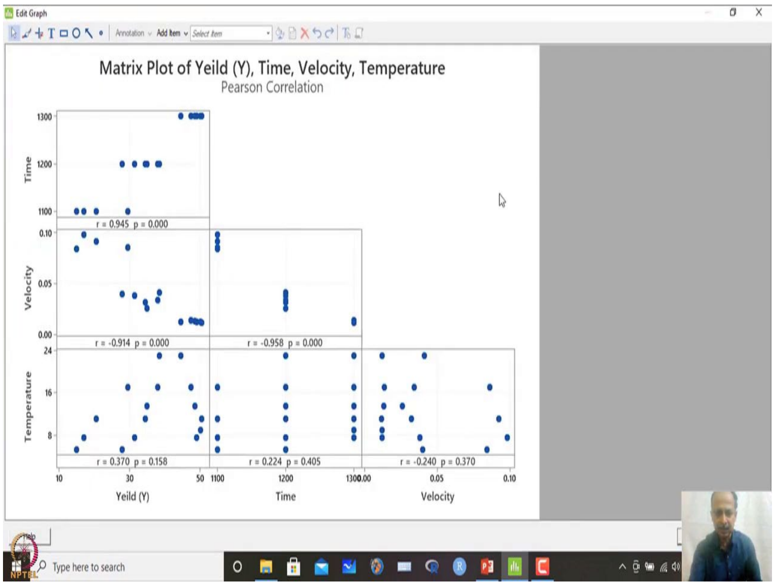
MULTIPLE LINEAR REGRESSION.MWX

Correlation: Yeild (Y), Time, Velocity, Temperature

	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260	

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Correlation: Yeild (Y), Time...
Matrix Plot of Yeild (Y), Time, Velocity, Temperature
Pearson Correlation
1300
1200
1100
Time
r = 0.945 p = 0.000
0.10
0.05
elocity
Surf
Multiple Linear Regression.mwx
Type here to search

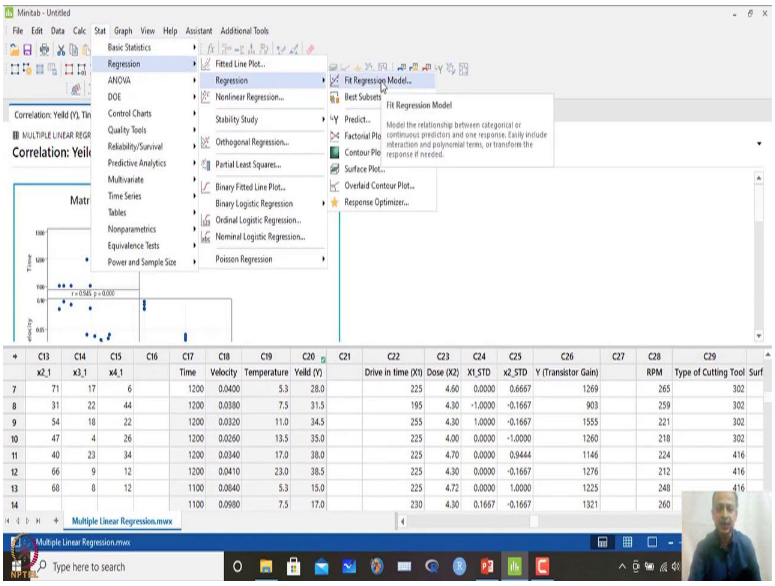
(Refer Slide Time: 05:16)



And if you click ok, this is the; this is the diagram that I have shown in the PPT slides like that. So, this shows relationship between yield and other variables. Yield is highly related with time, the P-Value is point near to 0 and velocity is also highly related with yield over here, but temperature does not seem to be significantly influencing over here.

So, this is the relationship and time is having a high time is not having a correlation with temperature over here ok, and velocity is having no correlation with temperature, but velocity is having a correlation with time over here. So that is reflected over here, temperature and time. So, this is highly 0.958 is the minus 0.958 is the.

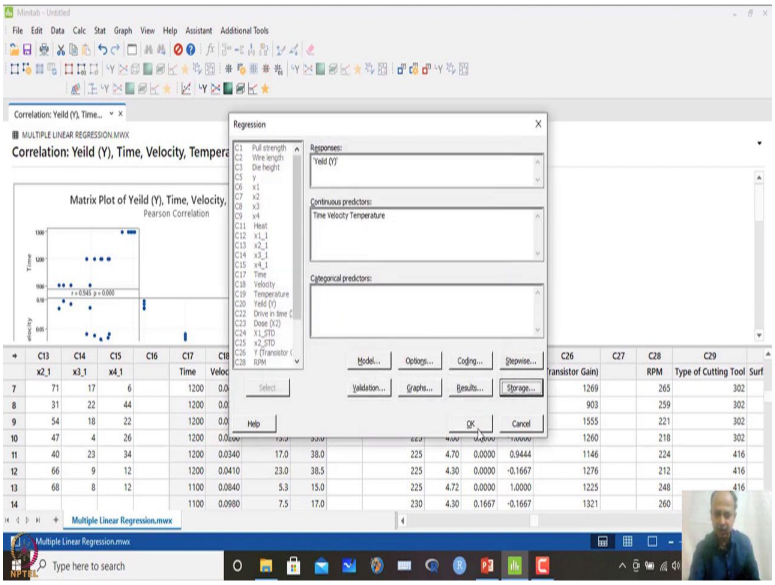
(Refer Slide Time: 06:03)



↓	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

So, this is so what we can do is that we can go to stepwise regression. So I will use stepwise regression over here.

(Refer Slide Time: 06:09)

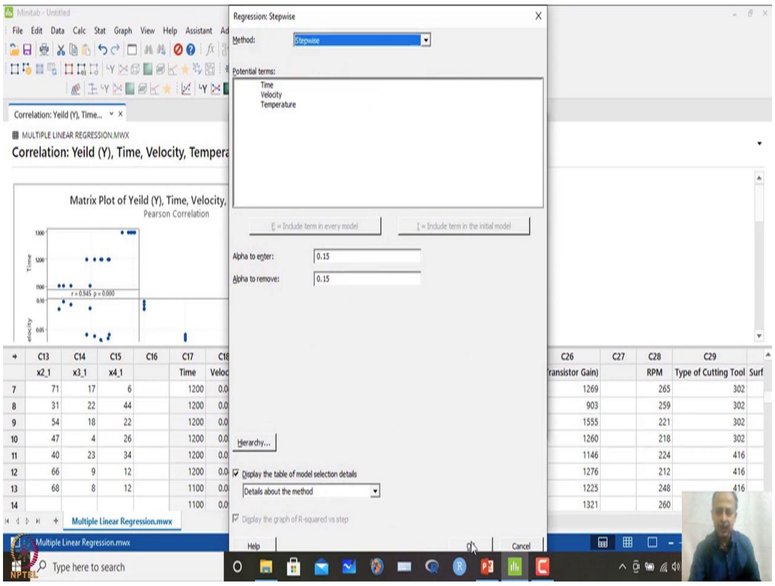


Responses: 'Yeild (Y)'

Continuous predictors: Time Velocity Temperature

Categorical predictors:

(Refer Slide Time: 06:11)



Regression: Stepwise

Method: Stepwise

Potential terms:
Time
Velocity
Temperature

Alpha to enter: 0.15

Alpha to remove: 0.15

☑ Display the table of model selection details

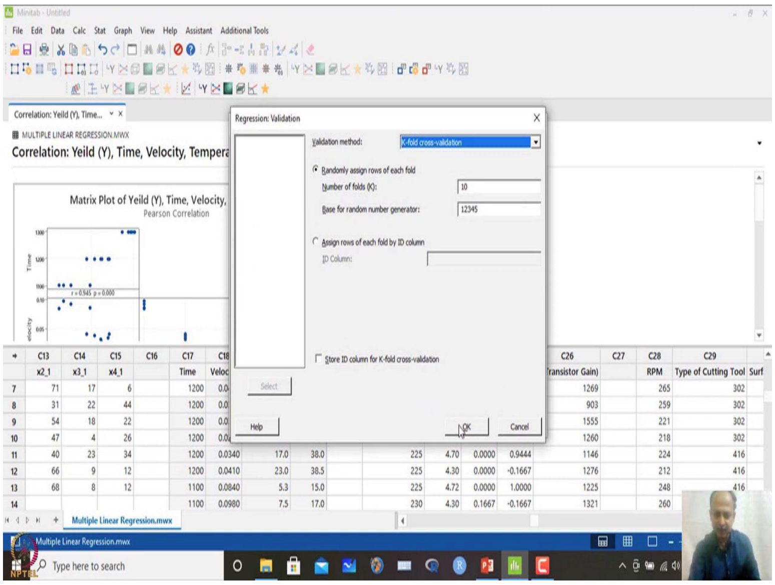
Details about the method

☑ Display the graph of R-squared vs step

E = Include term in every model
I = Include term in the initial model
Hierarchy...
Help
Cancel

So, fit regression model, and here only thing I have to do is that I have to introduce stepwise regression over here, and click ok.

(Refer Slide Time: 06:18)



Regression: Validation

Validation method: K-fold cross-validation

● Randomly assign rows of each fold

Number of folds (K): 10

Base for random number generator: 12345

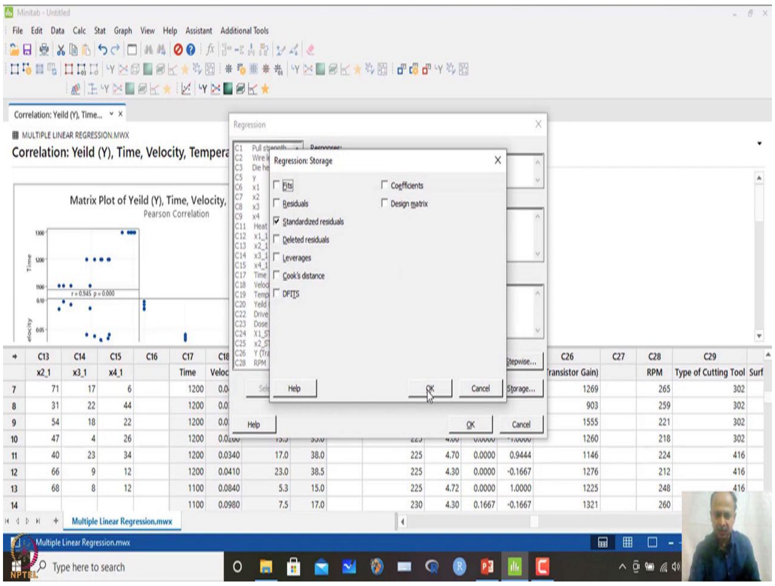
○ Assign rows of each fold by ID column

ID Column:

□ Store ID column for K-fold cross-validation

And then validation we can pair wise validation as usual.

(Refer Slide Time: 06:21)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Correlation: Yeild (Y), Time...
MULTIPLE LINEAR REGRESSION.MWX
Matrix Plot of Yeild (Y), Time, Velocity,
Pearson Correlation
Time
velocity
r = 0.945 p = 0.000
Regression
Responses:
Stepwise...
Storage...
Help
OK
Cancel
Multiple Linear Regression.mwx
Type here to search

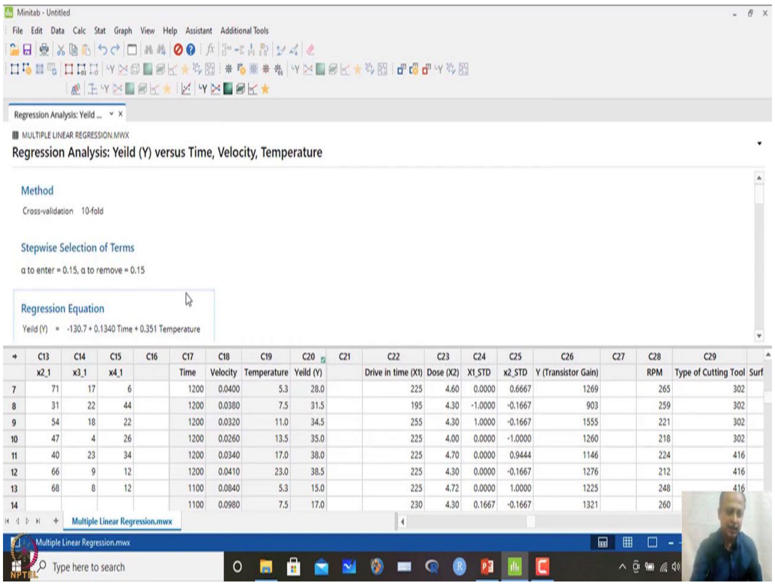
Correlation: Yeild (Y), Time, Velocity, Tempera

Regression: Storage

- ☐ Fits
- ☐ Residuals
- ☑ Standardized residuals
- ☐ Deleted residuals
- ☐ Leverages
- ☐ Cook's distance
- ☐ DFITS
- ☐ Coefficients
- ☐ Design matrix

And also storage we can save the residual standard as residual over here, and I click ok.

(Refer Slide Time: 06:29)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...
MULTIPLE LINEAR REGRESSION.MWX
Multiple Linear Regression.mwx
Type here to search

Regression Analysis: Yeild (Y) versus Time, Velocity, Temperature

Method

Cross-validation 10-fold

Stepwise Selection of Terms

α to enter = 0.15, α to remove = 0.15

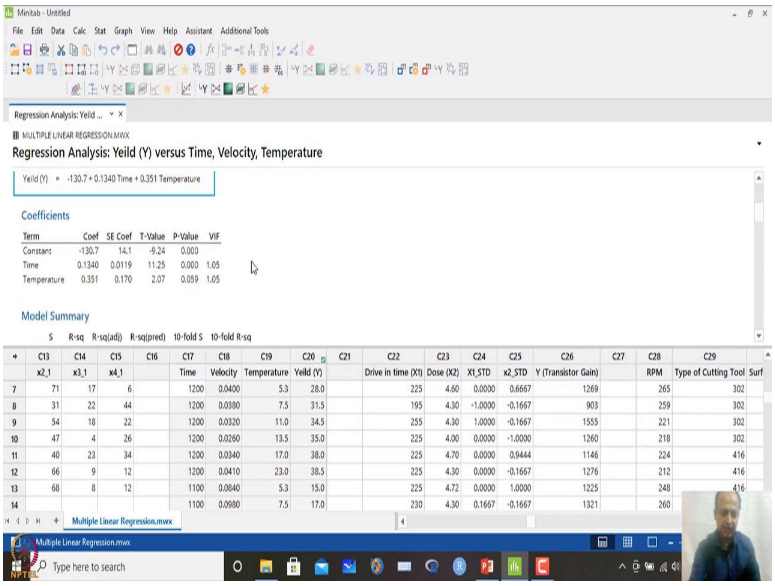
Regression Equation

Yeild (Y) = -130.7 + 0.1340 Time + 0.351 Temperature

↓	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

And let us try to see which are the variables will go in the model and which will go out of the model ok. So, what happened is that, it has identified time and temperature.

(Refer Slide Time: 06:35)



Regression Analysis: Yeild ...

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time, Velocity, Temperature

Yeild (Y) = -130.7 + 0.1340 Time + 0.351 Temperature

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-130.7	14.1	-9.24	0.000	
Time	0.1340	0.0119	11.25	0.000	1.05
Temperature	0.351	0.170	2.07	0.059	1.05

Model Summary

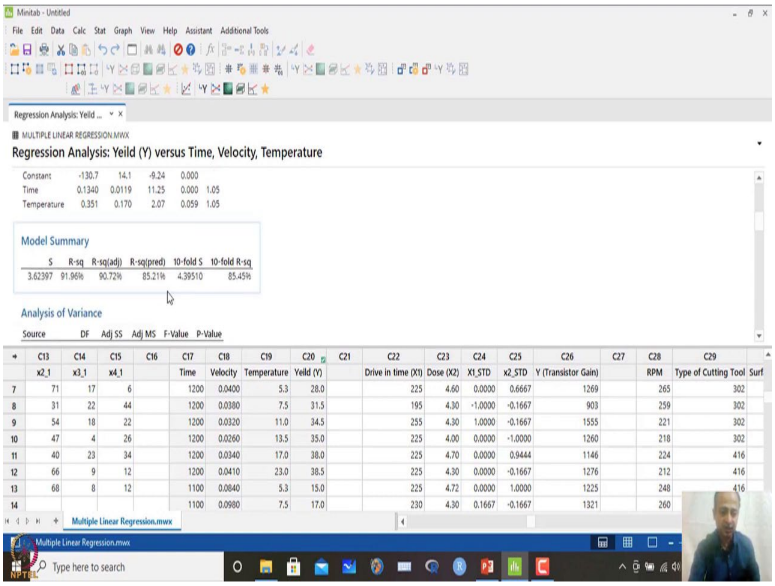
S R-sq R-sq(adj) R-sq(pred) 10-fold S 10-fold R-sq

→	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Multiple Linear Regression.mwx
Type here to search

So, initially we are thinking that it is not so highly correlated, but this is not so significant 0.05, just above the cross line that is 0.05 that is the cut off over here. So, it is the analysis over here says that, we have to take into consideration time and temperature, that is the best model that is coming out of this.

(Refer Slide Time: 06:55)



Regression Analysis: Yeild ...

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time, Velocity, Temperature

Constant	-130.7	14.1	-9.24	0.000	
Time	0.1340	0.0119	11.25	0.000	1.05
Temperature	0.351	0.170	2.07	0.059	1.05

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
3.62397	91.96%	90.72%	85.21%	4.39510	85.45%

Analysis of Variance

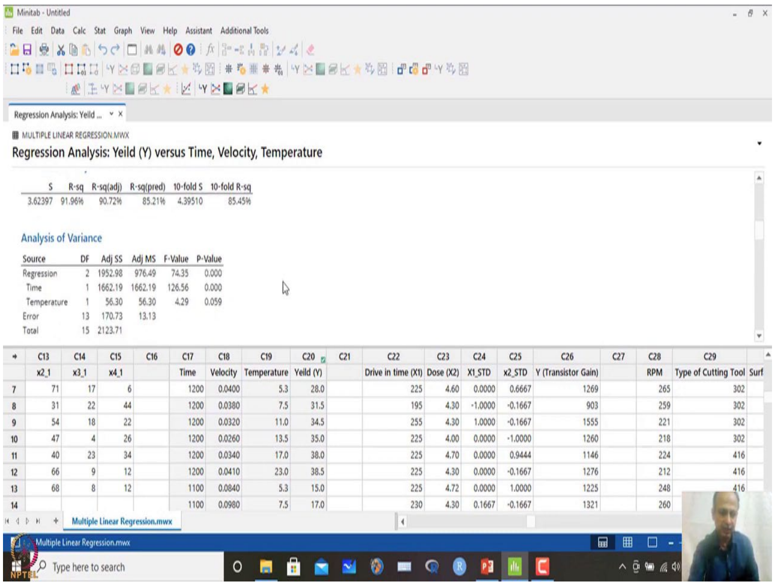
Source	DF	Adj SS	Adj MS	F-Value	P-Value

→	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Multiple Linear Regression.mwx
Type here to search

And, the R square adjusted value is 90 point something like that and 10 fold cross-validation more or less close.

(Refer Slide Time: 07:03)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time, Velocity, Temperature

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
3.62397	91.96%	90.72%	85.21%	4.39510	85.45%

Analysis of Variance

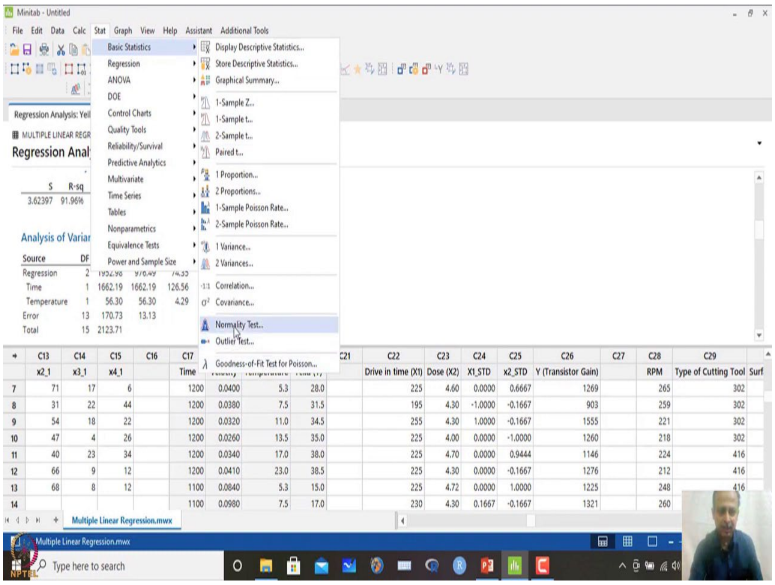
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	1952.98	976.49	74.35	0.000
Time	1	1662.19	1662.19	126.56	0.000
Temperature	1	56.30	56.30	4.29	0.059
Error	13	170.73	13.13		
Total	15	2123.71			

➔	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

Multiple Linear Regression.mwx
Multiple Linear Regression.mwx
Type here to search
NPTEL

So this model seems to be adequate, if I consider that alpha level of significance is 0.1, in that case we can also consider temperature as significant variable we may retain this one.

(Refer Slide Time: 07:13)



MULTIPLE LINEAR REGR

Regression Anal

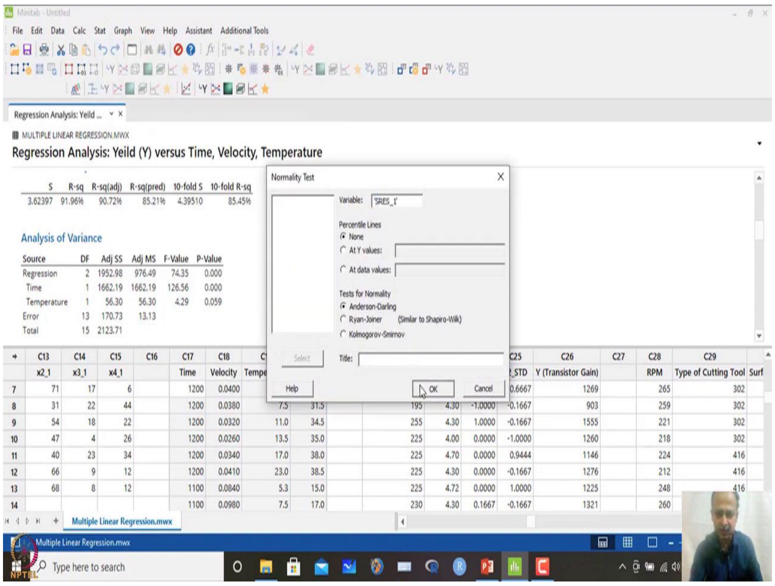
S	R-sq
3.62397	91.96%

Analysis of Varian

Source	DF			
Regression	2	1952.98	976.49	74.35
Time	1	1662.19	1662.19	126.56
Temperature	1	56.30	56.30	4.29
Error	13	170.73	13.13	
Total	15	2123.71		

➔	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time					Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

(Refer Slide Time: 07:16)



Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time, Velocity, Temperature

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
3.62397	91.96%	90.72%	85.21%	4.39510	85.45%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	1952.98	976.49	74.35	0.000
Time	1	1662.19	1662.19	126.56	0.000
Temperature	1	56.30	56.30	4.29	0.059
Error	13	170.73	13.13		
Total	15	2123.71			

Normality Test

Variable: 'SRES_1'

Percentile Lines

● None

○ At Y values:

○ At data values:

Tests for Normality

● Anderson-Darling

○ Ryan-Joiner (Similar to Shapiro-Wilk)

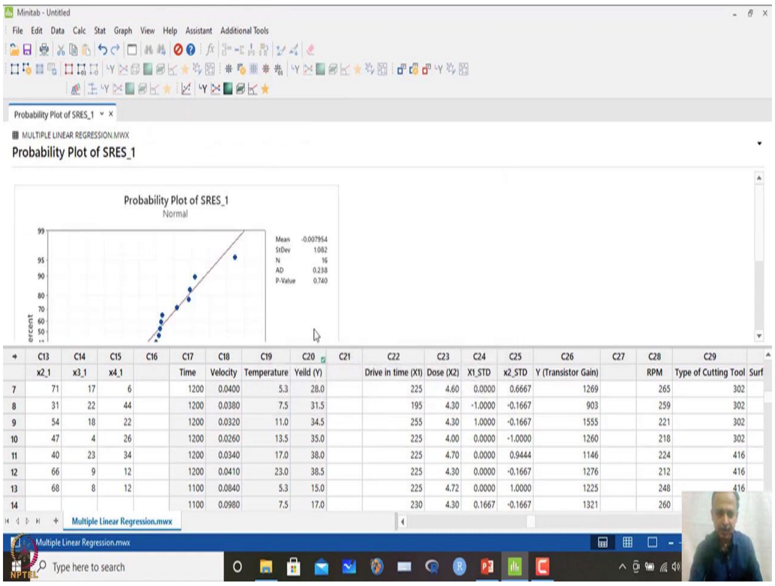
○ Kolmogorov-Smirnov

Title:

	C13	C14	C15	C16	C17	C18	C25	C26	C27	C28	C29
	x2_1	x3_1	x4_1		Time	Velocity	_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool
7	71	17	6		1200	0.0400	0.6667	1269		265	302
8	31	22	44		1200	0.0380	-0.1667	903		259	302
9	54	18	22		1200	0.0320	-0.1667	1555		221	302
10	47	4	26		1200	0.0260	-1.0000	1260		218	302
11	40	23	34		1200	0.0340	0.9444	1146		224	416
12	66	9	12		1200	0.0410	-0.1667	1276		212	416
13	68	8	12		1100	0.0840	1.0000	1225		248	416
14					1100	0.0980	-0.1667	1321		260	

So, let us try to see what about the normality distributions of this residual over here.

(Refer Slide Time: 07:20)



Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Probability Plot of SRES_1

MULTIPLE LINEAR REGRESSION.MWX

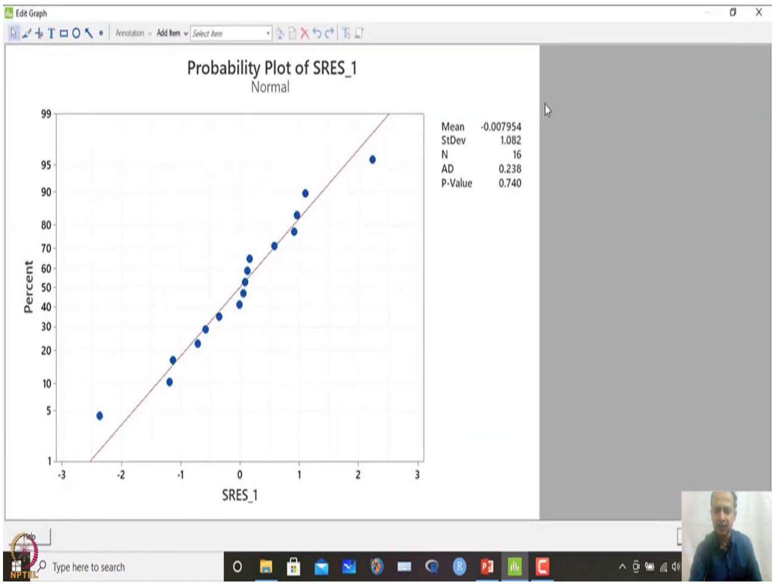
Probability Plot of SRES_1

Probability Plot of SRES_1
Normal
99
95
90
80
70
60
50
ercent

Mean	-0.007954
StDev	1.082
N	16
AD	0.238
P-Value	0.740

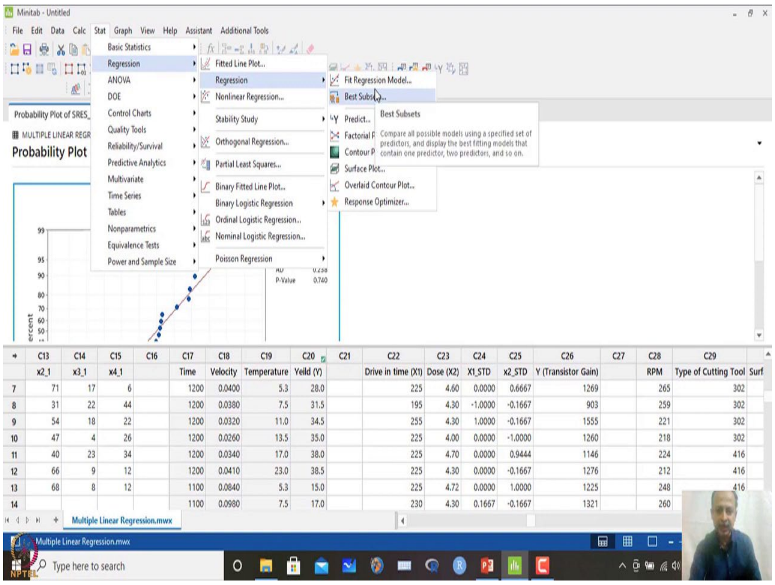
	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

(Refer Slide Time: 07:23)



So, last residual will be normality we can check, and in this case 0.74, so there is not much problem with the normality assumptions over here. And, we do not have any problem over here.

(Refer Slide Time: 07:31)

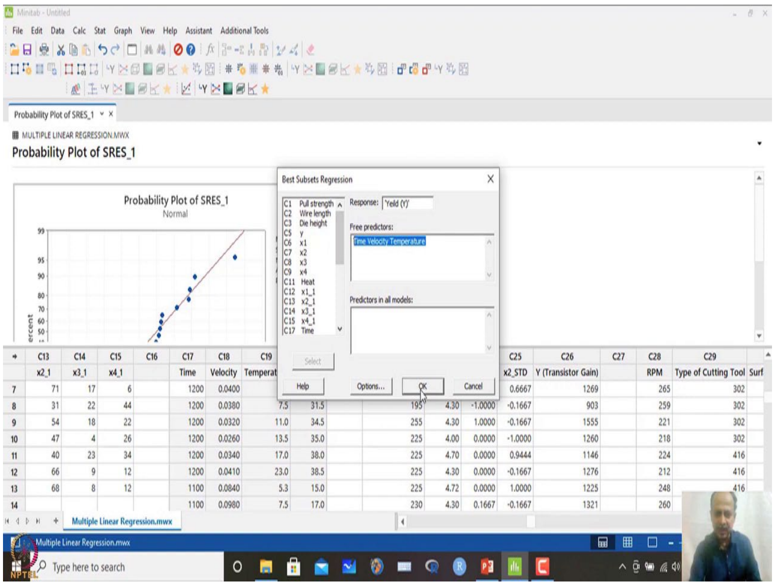


Best Subsets

Compare all possible models using a specified set of predictors, and display the best fitting models that contain one predictor, two predictors, and so on.

→	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260	

(Refer Slide Time: 07:36)



Response: 'Yeild (Y)'

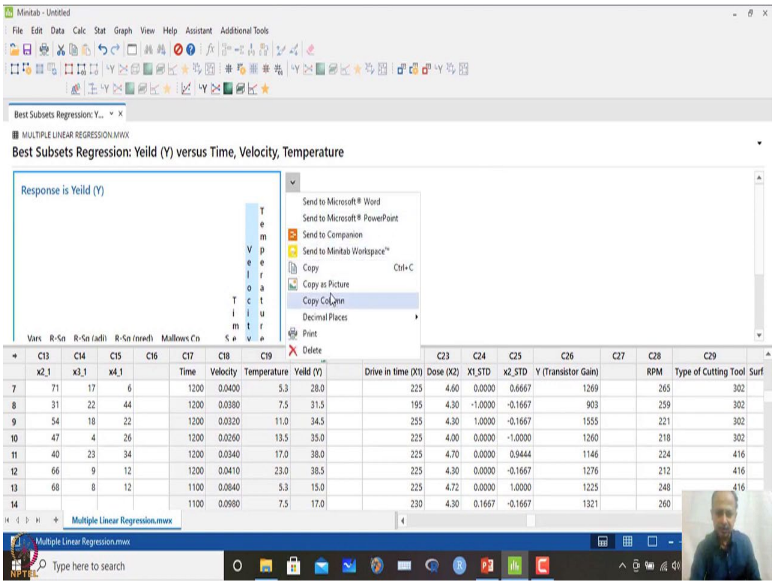
Free predictors: Time Velocity Temperature

Predictors in all models:

	C13	C14	C15	C16	C17	C18	C19							C25	C26	C27	C28	C29
	x2_1	x3_1	x4_1		Time	Velocity	Temperat							x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool
7	71	17	6		1200	0.0400								0.6667	1269		265	302
8	31	22	44		1200	0.0380	7.5	31.5			195	4.30	-1.0000	-0.1667	903		259	302
9	54	18	22		1200	0.0320	11.0	34.5			255	4.30	1.0000	-0.1667	1555		221	302
10	47	4	26		1200	0.0260	13.5	35.0			225	4.00	0.0000	-1.0000	1260		218	302
11	40	23	34		1200	0.0340	17.0	38.0			225	4.70	0.0000	0.9444	1146		224	416
12	66	9	12		1200	0.0410	23.0	38.5			225	4.30	0.0000	-0.1667	1276		212	416
13	68	8	12		1100	0.0840	5.3	15.0			225	4.72	0.0000	1.0000	1225		248	416
14					1100	0.0980	7.5	17.0			230	4.30	0.1667	-0.1667	1321		260	

So, in this case normality is not a constraint. So, in this case we can also see, when I use best subset regression which is the best model it is giving, so I have taken yield over here and we have taken all the three variables.

(Refer Slide Time: 07:44)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Best Subsets Regression: Y...
MULTIPLE LINEAR REGRESSION.MWX
Send to Microsoft® Word
Send to Microsoft® PowerPoint
Send to Companion
Send to Minitab Workspace™
Copy Ctrl+C
Copy as Picture
Copy Column
Decimal Places
Print
Delete
Multiple Linear Regression.mwx
Type here to search

Best Subsets Regression: Yeild (Y) versus Time, Velocity, Temperature

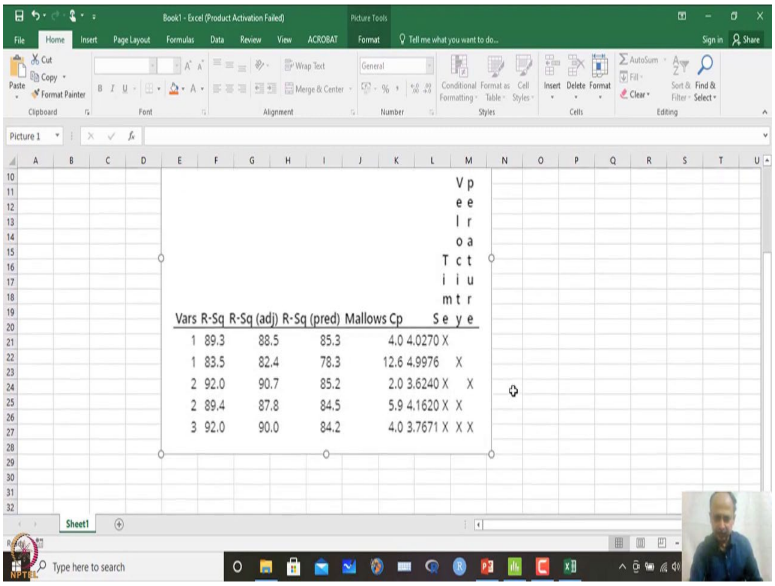
Response is Yeild (Y)

Vars R-Sq R-Sq (adj) R-Sq (pred) Mallows Cp S Time Velocity Temperature

	C13	C14	C15	C16	C17	C18	C19				C23	C24	C25	C26	C27	C28	C29
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260	

And if I click ok, then what happens let us try to see. And we can just paste it copy as a picture over here so then we can paste it in excel and try to see what happens ok. So, we can adopt, we will try to see what happens if we select, so excel we are opening one excel sheet ok. So, let us do that one and we can paste it over here.

(Refer Slide Time: 08:14)



Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	Time	Velocity	perature
1	89.3	88.5	85.3	4.0	4.0270	X		
1	83.5	82.4	78.3	12.6	4.9976		X	
2	92.0	90.7	85.2	2.0	3.6240	X		X
2	89.4	87.8	84.5	5.9	4.1620	X	X	
3	92.0	90.0	84.2	4.0	3.7671	X	X	X

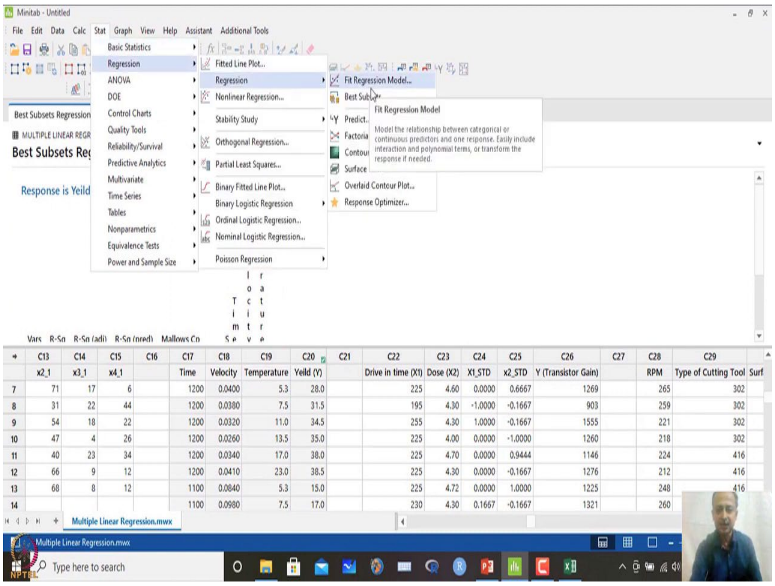
So, when we are pasting it over here. What, observations we have that we want to see. So, over here this is time variable over here, this is temperature and this is the velocity over here. With one variable what we are seeing mallow Cp is about 4, R square adjusted is 88.5.

When I take consider only time as the variables like that ok, but mallow Cp is higher than the number of variables plus 1. So this is not recommended. Second one is also not recommended with one variable, third one what we are seeing, time and temperature it is mallow Cp is less than the number of variables plus 1, so 3 is 2 is less than 3. So, this seems to be adequate, this one of the suggested model over here.

And we are getting R square value of 92 which is higher than the earlier one. And, what we are getting is that R square adjusted is also high over here. And, prediction is also more or less same what we are getting over here, only thing is that when we adopt this one tenfold cross-validation also we can check ok. 5.9 is quite high, so this goes away and last one is considering all models, we want to deduce this one.

So last one does not, we do not see the last one. So, mallow Cp suggest that this is the one when time and temperature can be considered like that. So, so that is also suggested by stepwise regression what we have seen like that.

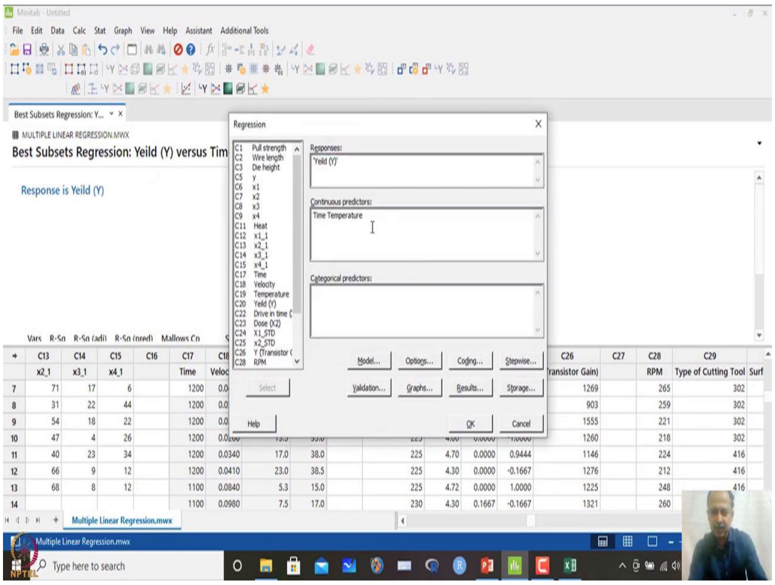
(Refer Slide Time: 09:33)



↓	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260	

So in this case, again I am doing this. So, if we select this one fit regression model over here.

(Refer Slide Time: 09:39)



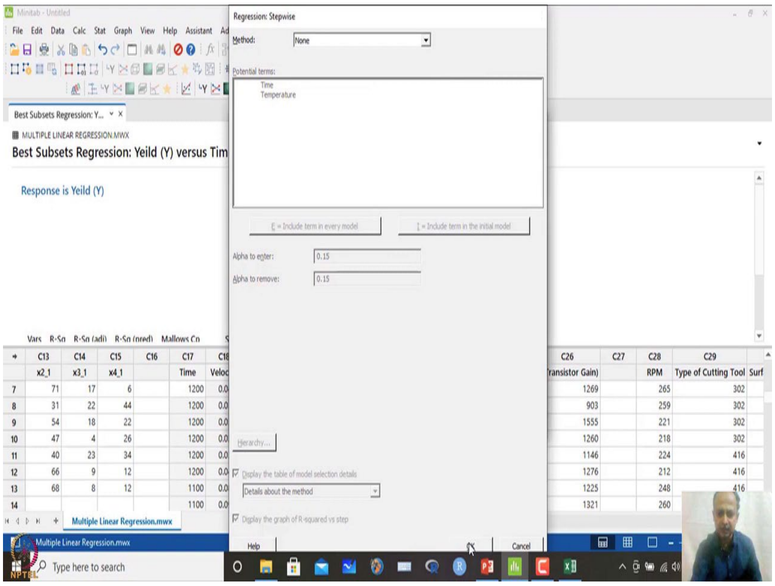
Responses: 'Yeild (Y)'

Continuous predictors: Time Temperature

Categorical predictors:

And in this case, time and temperature. So, I am removing this one and these are the two variables.

(Refer Slide Time: 09:44)



Best Subsets Regression: Yeild (Y) versus Tim

Response is Yeild (Y)

Regression: Stepwise

Method: None

Potential terms:
Time
Temperature

E = Include term in every model | I = Include term in the initial model

Alpha to enter: 0.15

Alpha to remove: 0.15

Hierarchy...

☑ Display the table of model selection details

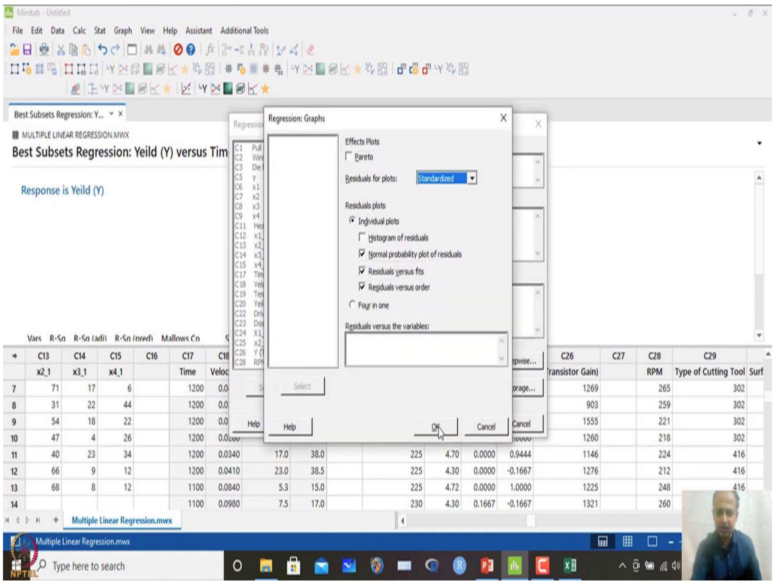
Details about the method

☑ Display the graph of R-squared vs step

Help | OK | Cancel

And stepwise regression, I will not use now. So, in this case I have already identified the variables.

(Refer Slide Time: 09:51)



Best Subsets Regression: Yeild (Y) versus Tim

Response is Yeild (Y)

Regression: Graphs

Effects Plots

□ Pareto

Residuals for plots: Standardized

Residuals plots

● Individual plots

□ Histogram of residuals

☑ Normal probability plot of residuals

☑ Residuals versus fits

☑ Residuals versus order

○ Four in one

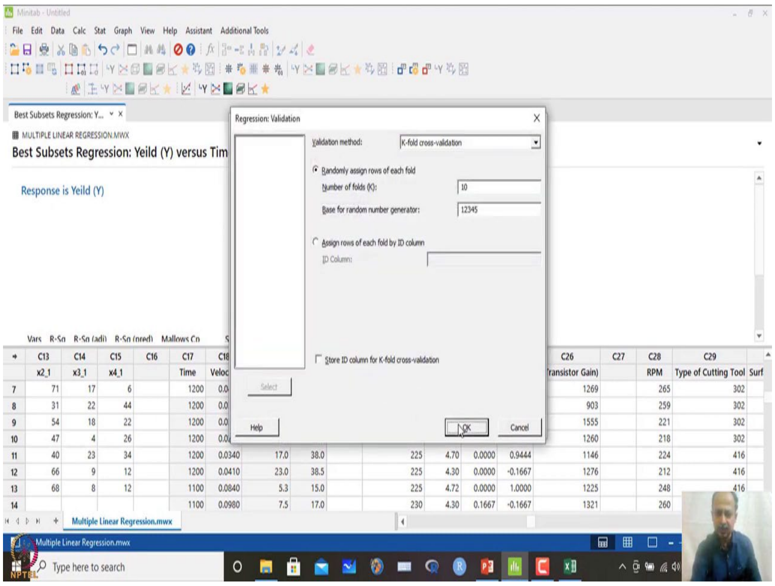
Residuals versus the variables:

Select

Help | OK | Cancel

And, I want to see all graphs, whether it is and this residual can be a standardized residual over here, so if I click ok.

(Refer Slide Time: 10:00)



Best Subsets Regression: Yeild (Y) versus Tim

Response is Yeild (Y)

Regression: Validation

Validation method: K-fold cross-validation

● Randomly assign rows of each fold

Number of folds (K): 10

Base for random number generator: 12345

○ Assign rows of each fold by ID column

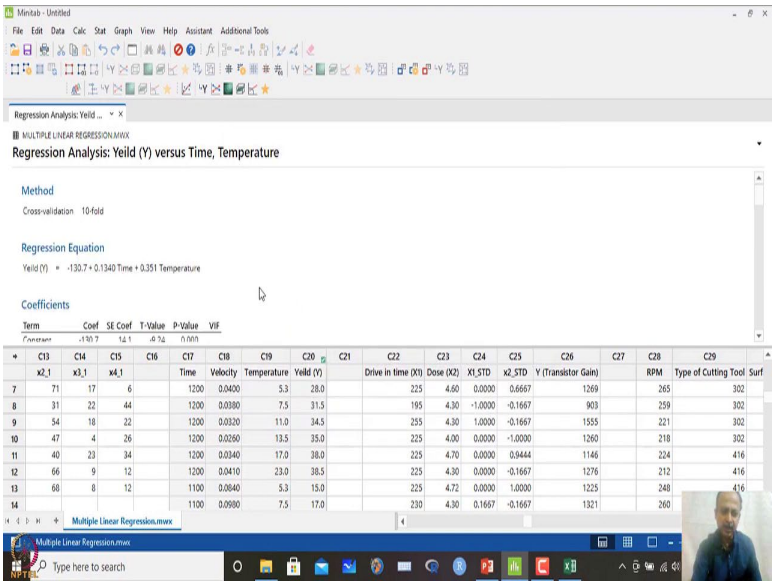
ID Column:

□ Store ID column for K-fold cross-validation

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Best Subsets Regression: Y...
MULTIPLE LINEAR REGRESSION.MWX
Vars R-Sq R-Sq (adj) R-Sq (pred) Mallows Cp
Select
Help
OK
Cancel
Multiple Linear Regression.mwx
Type here to search

And, validation we have already careful cross-validation we have given.

(Refer Slide Time: 10:05)



Regression Analysis: Yeild (Y) versus Time, Temperature

Method

Cross-validation 10-fold

Regression Equation

Yeild (Y) = -130.7 + 0.1340 Time + 0.351 Temperature

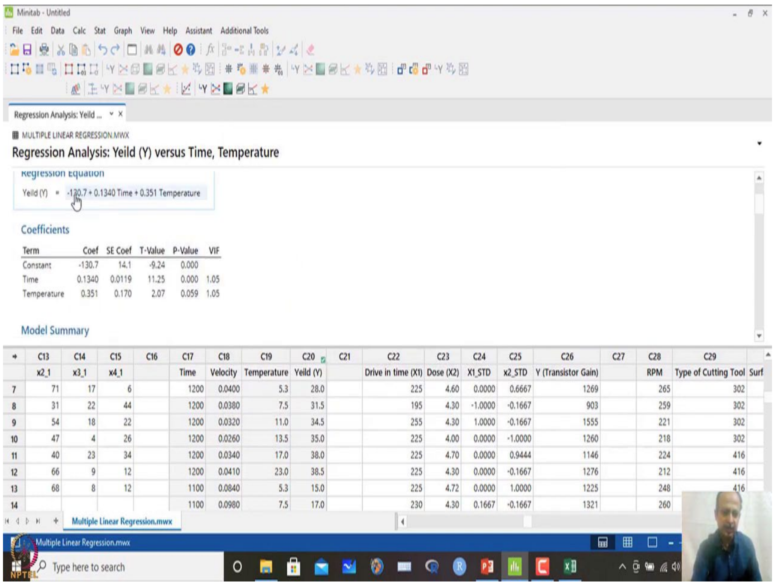
Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-130.7	14.1	-9.24	0.000	

	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260	

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...
MULTIPLE LINEAR REGRESSION.MWX
Surf
Multiple Linear Regression.mwx
Type here to search

(Refer Slide Time: 10:05)



Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time, Temperature

Regression Equation

Yeild (Y) = -130.7 + 0.1340 Time + 0.351 Temperature

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-130.7	14.1	-9.24	0.000	
Time	0.1340	0.0119	11.25	0.000	1.05
Temperature	0.351	0.170	2.07	0.059	1.05

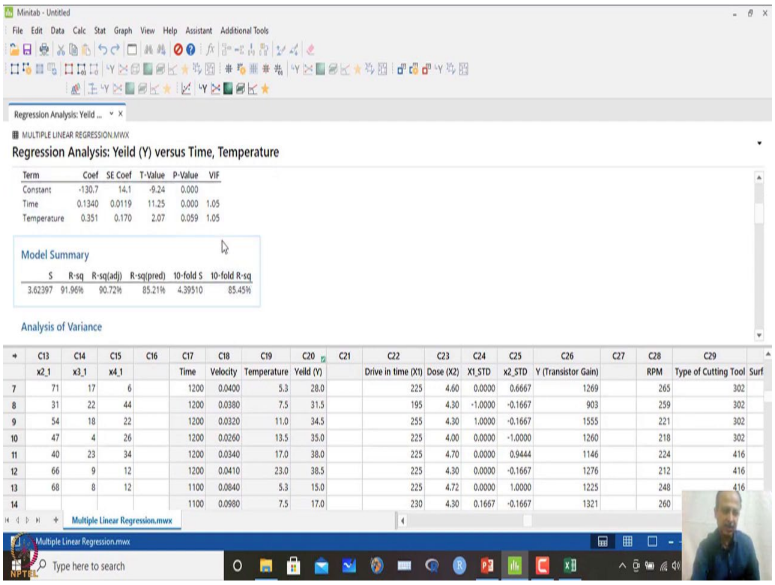
Model Summary

➔	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

Multiple Linear Regression.mwx
Multiple Linear Regression.mwx
Type here to search
NPTEL

And if you click ok, what happens is that, this is the final equation minus 130.7 and plus time it is positively correlated and second one temperature is also positively correlated like that.

(Refer Slide Time: 10:14)



Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time, Temperature

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-130.7	14.1	-9.24	0.000	
Time	0.1340	0.0119	11.25	0.000	1.05
Temperature	0.351	0.170	2.07	0.059	1.05

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
3.62397	91.96%	90.72%	85.21%	4.39510	85.45%

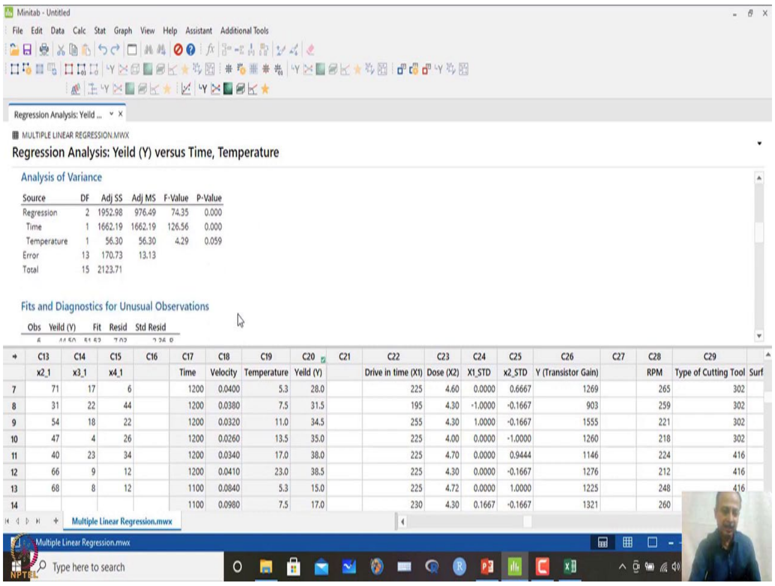
Analysis of Variance

➔	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

Multiple Linear Regression.mwx
Multiple Linear Regression.mwx
Type here to search
NPTEL

So now, temperature is retained over here, although the P-Value is not so significant over here. We can also just eliminate and see whether that model performs in any way, because we have a variable to enter and exit we have given a alpha value of 0.15, so that is why this model, it has come in the model when we have done stepwise regression. But we can retain this one, because this is very close to 0.05.

(Refer Slide Time: 10:40)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...
MULTIPLE LINEAR REGRESSION.MWX
Multiple Linear Regression.mwx
Type here to search
NPTEL

Regression Analysis: Yeild (Y) versus Time, Temperature

Analysis of Variance

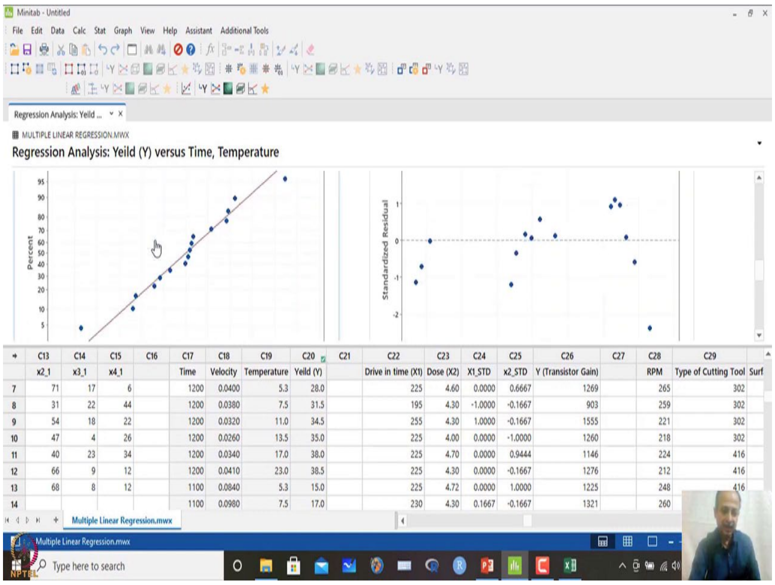
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	1952.98	976.49	74.35	0.000
Time	1	1662.19	1662.19	126.56	0.000
Temperature	1	56.30	56.30	4.29	0.059
Error	13	170.73	13.13		
Total	15	2123.71			

Fits and Diagnostics for Unusual Observations

Obs	Yeild (Y)	Fit	Resid	Std Resid
6	44.50	51.52	7.02	2.26 R

➔	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

(Refer Slide Time: 10:42)

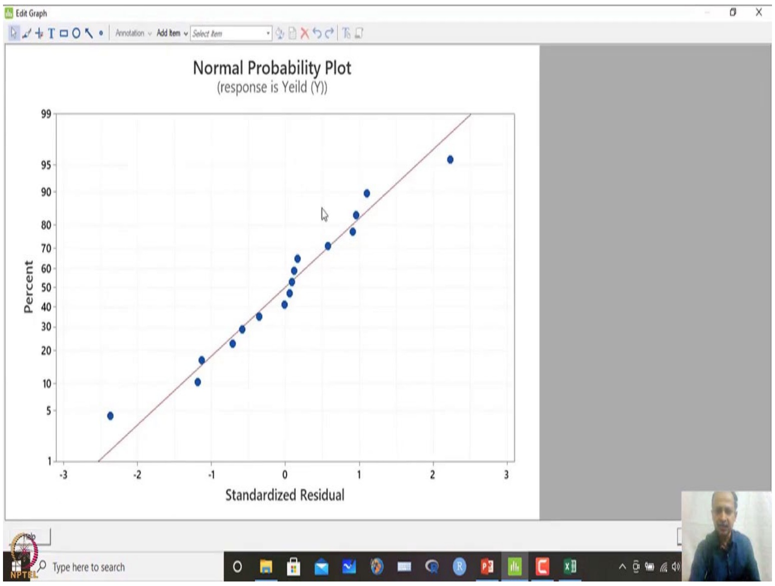

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...
MULTIPLE LINEAR REGRESSION.MWX
Percent
95
90
80
70
60
50
40
30
20
10
5
Standardized Residual
1
0
-1
-2
Multiple Linear Regression.mwx
Type here to search
NPTEL

Regression Analysis: Yeild (Y) versus Time, Temperature

➔	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

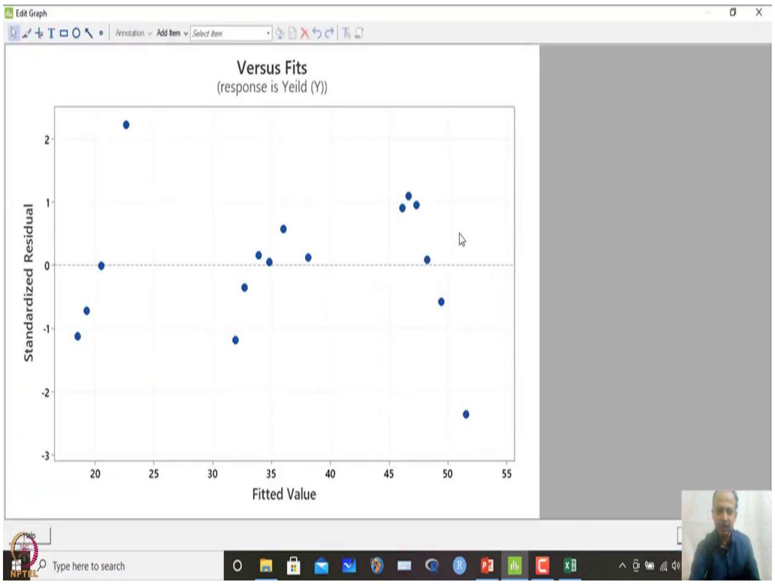
And, we can see that normal distribution assumptions over here that seems to be satisfactory as all the points.

(Refer Slide Time: 10:48)



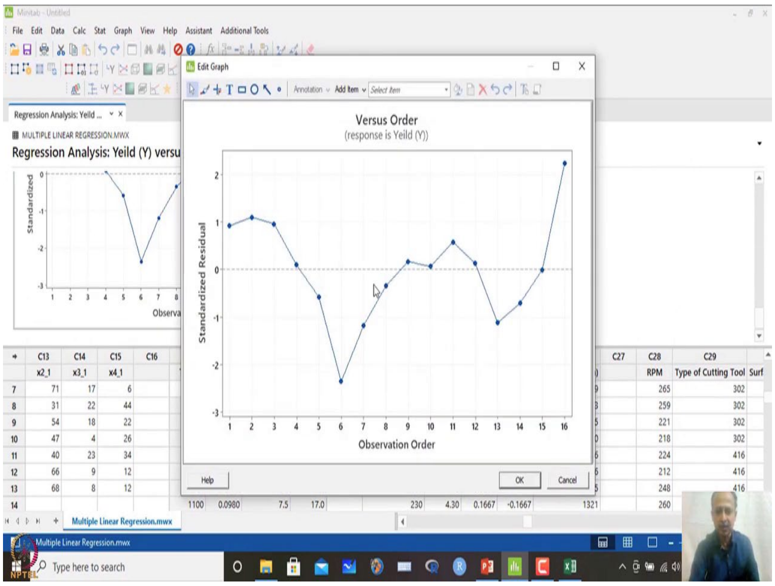
So this graphically when we see seems to be satisfactory points on the line. So, but we can cross check that one.

(Refer Slide Time: 10:55)



Even the residual versus fit plot also does not show any abnormality or patterns like that, so maybe Breusch Pagan test will also confirm this one.

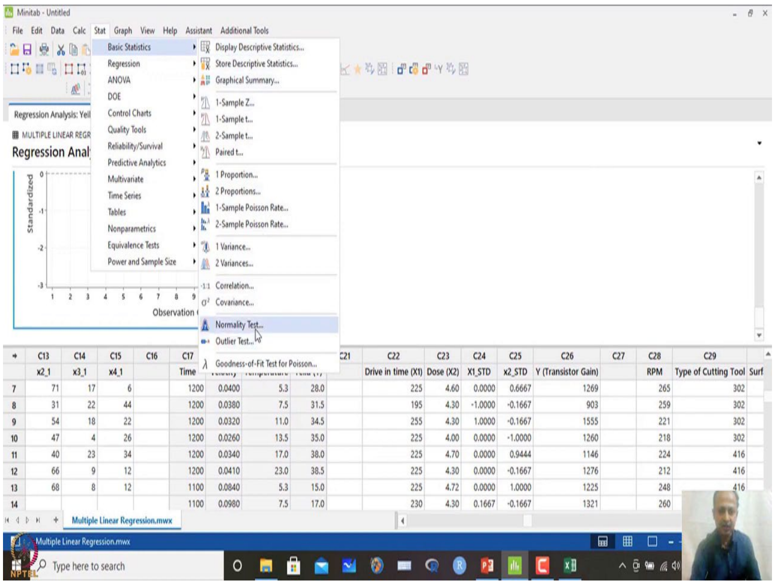
(Refer Slide Time: 11:01)



	C13	C14	C15	C16	C28	C29	
	x2_1	x3_1	x4_1		RPM	Type of Cutting Tool	Surf
7	71	17	6		265	302	
8	31	22	44		259	302	
9	54	18	22		221	302	
10	47	4	26		218	302	
11	40	23	34		224	416	
12	66	9	12		212	416	
13	68	8	12		248	416	
14					260		

And also there is no as such abnormalities in auto correlation, what we observe some trend or something is not observed it is on both side of the 000 point like that.

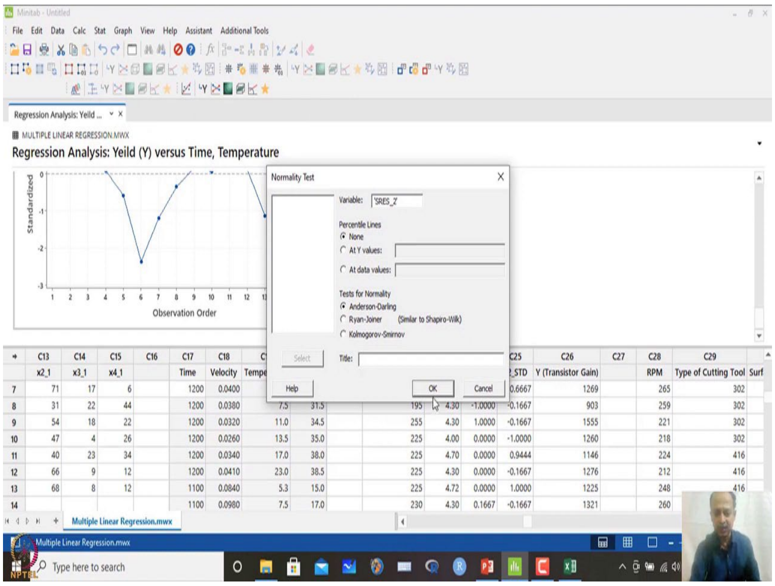
(Refer Slide Time: 11:14)



	C13	C14	C15	C16	C17				C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time					Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

So in this case, what is expected is that it should confirm the normality and other assumptions over here.

(Refer Slide Time: 11:17)



Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...
MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time, Temperature

Standardized
Observation Order

Normality Test

Variable: 'SRES_2'

Percentile Lines
● None
○ At Y values:
○ At data values:

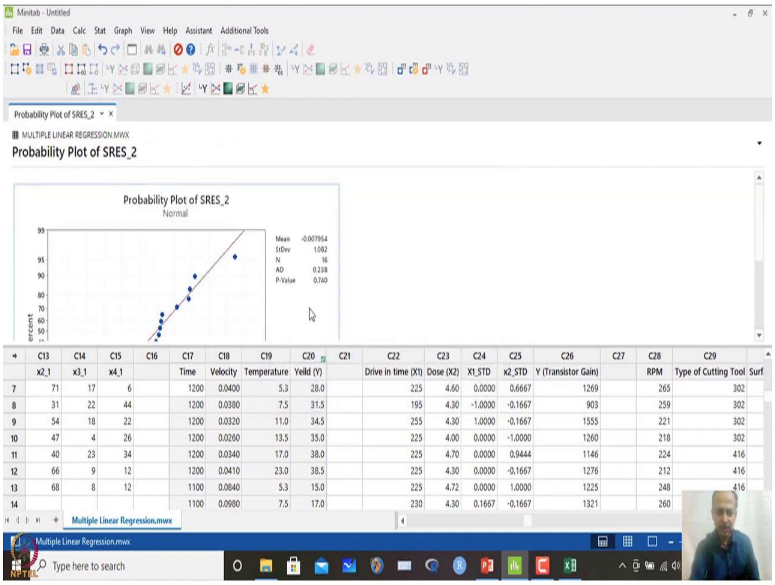
Tests for Normality
● Anderson-Darling
○ Ryan-Joiner (Similar to Shapiro-Wilk)
○ Kolmogorov-Smirnov

Title:

	C13	C14	C15	C16	C17	C18	C25	C26	C27	C28	C29
	x2_1	x3_1	x4_1		Time	Velocity	_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool
7	71	17	6		1200	0.0400	0.6667	1269		265	302
8	31	22	44		1200	0.0380	-0.1667	903		259	302
9	54	18	22		1200	0.0320	-0.1667	1555		221	302
10	47	4	26		1200	0.0260	-1.0000	1260		218	302
11	40	23	34		1200	0.0340	0.9444	1146		224	416
12	66	9	12		1200	0.0410	-0.1667	1276		212	416
13	68	8	12		1100	0.0840	1.0000	1225		248	416
14					1100	0.0980	-0.1667	1321		260	

So we can just check this one and the last row will be the last column will be the variables residuals that we have to check.

(Refer Slide Time: 11:24)



Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Probability Plot of SRES_2
MULTIPLE LINEAR REGRESSION.MWX

Probability Plot of SRES_2

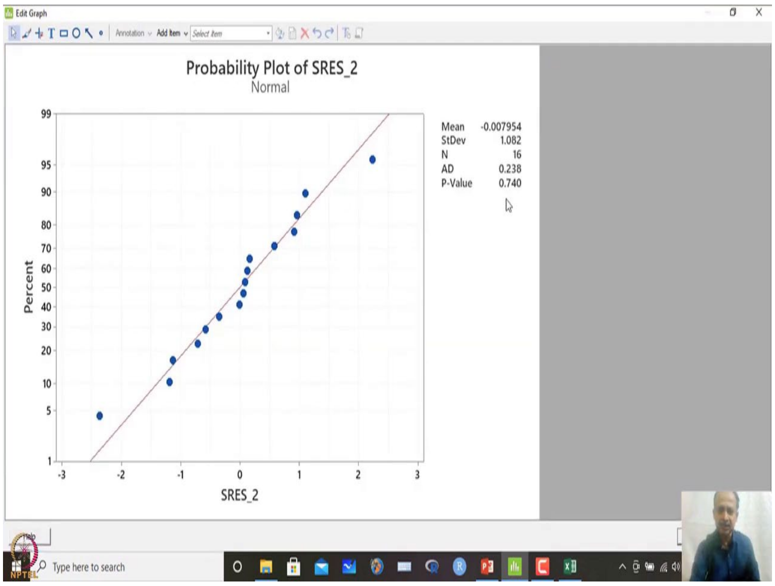
Probability Plot of SRES_2
Normal

Mean -0.007954
StDev 1.082
N 16
AD 0.238
P-Value 0.740

99
95
90
80
70
60
50
ercent

	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

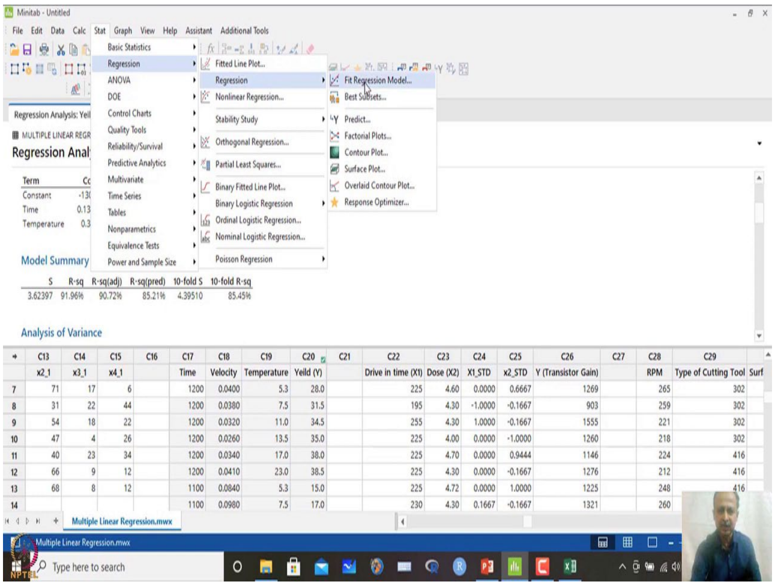
(Refer Slide Time: 11:25)



So, in this case it is coming out to be 0.7 seems to be satisfactory ok. So, in this case you have to practically also think that whether to retain the last variables or to remove, but R square adjusted as has improved what was observed.

So, R square adjusted has improved. If I consider only one variable, maybe R square adjusted will be low. So in case, I consider one variable as time as the only variable, so in this case we can do that.

(Refer Slide Time: 11:47)



MULTIPLE LINEAR REGR

Regression Anal

Term	Co
Constant	-13
Time	0.13
Temperature	0.3

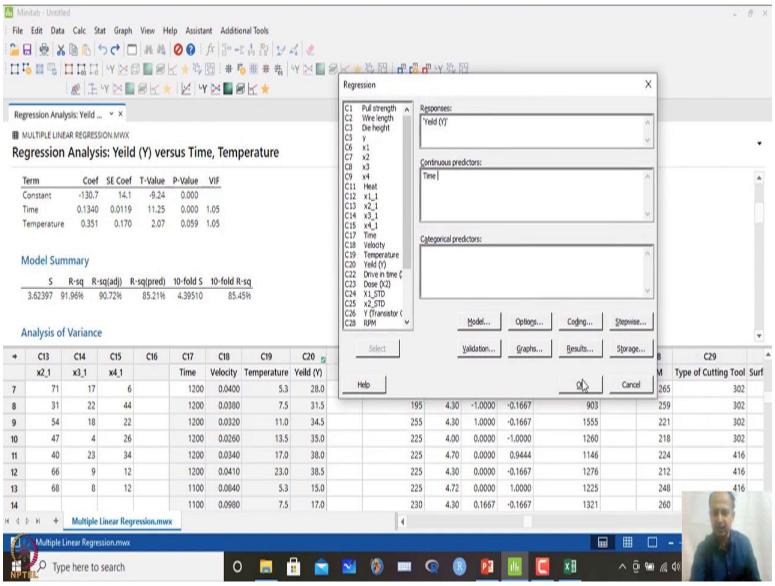
Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
3.62397	91.96%	90.72%	85.21%	4.39510	85.45%

Analysis of Variance

↓	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

(Refer Slide Time: 11:52)



MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time, Temperature

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-130.7	14.1	-9.24	0.000	
Time	0.1340	0.0119	11.25	0.000	1.05
Temperature	0.351	0.170	2.07	0.059	1.05

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
3.62397	91.96%	90.72%	85.21%	4.39510	85.45%

Analysis of Variance

Responses: 'Yeild (Y)'

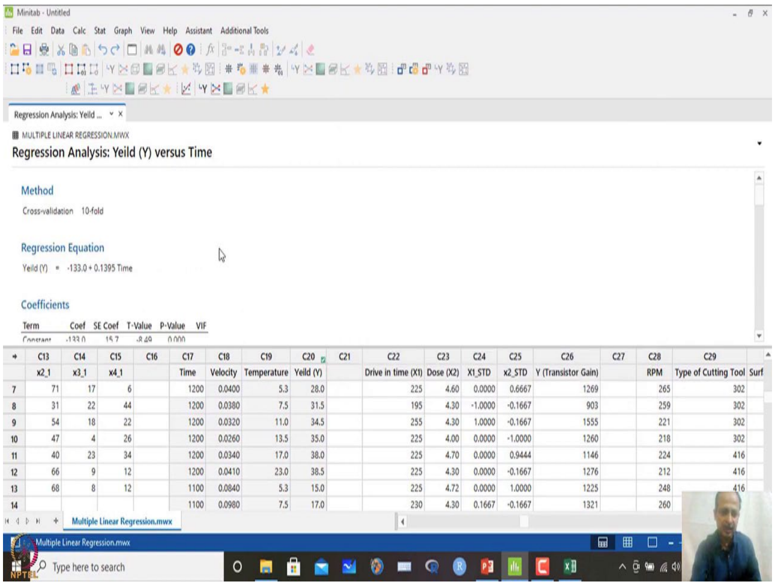
Continuous predictors: Time

Categorical predictors:

	C13	C14	C15	C16	C17	C18	C19	C20
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)
7	71	17	6		1200	0.0400	5.3	28.0
8	31	22	44		1200	0.0380	7.5	31.5
9	54	18	22		1200	0.0320	11.0	34.5
10	47	4	26		1200	0.0260	13.5	35.0
11	40	23	34		1200	0.0340	17.0	38.0
12	66	9	12		1200	0.0410	23.0	38.5
13	68	8	12		1100	0.0840	5.3	15.0
14					1100	0.0980	7.5	17.0

So, we can just consider time as the only variable, so fit regression model instead of temperature I will remove this one. So, I just keep a note of this R square adjusted is around 90, this is around 85 R square predicted and tenfold cross-validation earlier was around 85.

(Refer Slide Time: 12:07)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...
Multiple Linear Regression.mwx
Type here to search

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time

Method

Cross-validation 10-fold

Regression Equation

Yeild (Y) = -133.0 + 0.1395 Time

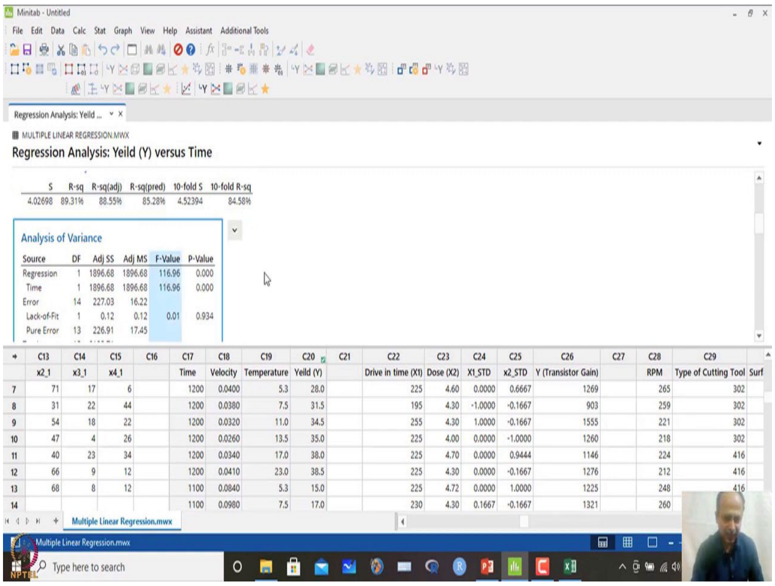
Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-133.0	15.7	-8.49	0.000	

	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260	

So, 90 and 85 approximately that is the range what we are getting.

(Refer Slide Time: 12:09)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...
Multiple Linear Regression.mwx
Type here to search

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
4.02698	89.31%	88.55%	85.28%	4.52394	84.58%

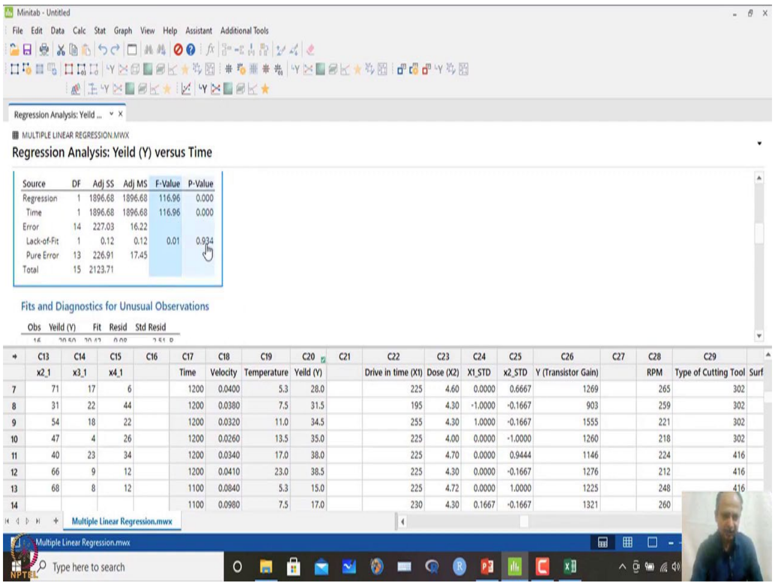
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1896.68	1896.68	116.96	0.000
Time	1	1896.68	1896.68	116.96	0.000
Error	14	227.03	16.22		
Lack-of-Fit	1	0.12	0.12	0.01	0.934
Pure Error	13	226.91	17.45		

➔	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

And, when we do this here you see R square adjusted is also low lower than the earlier one. So, it depends on the practical sense of your, whether to adopt the variable or whether to remove the variable which is not significant like that that depends on the process engineers or somebody who is knowledgeable about that whether to retain that one or to eliminate that one. And, there is no lack of fit as such that is observed over here.

(Refer Slide Time: 12:31)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...
Multiple Linear Regression.mwx
Type here to search

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1896.68	1896.68	116.96	0.000
Time	1	1896.68	1896.68	116.96	0.000
Error	14	227.03	16.22		
Lack-of-Fit	1	0.12	0.12	0.01	0.934
Pure Error	13	226.91	17.45		
Total	15	2123.71			

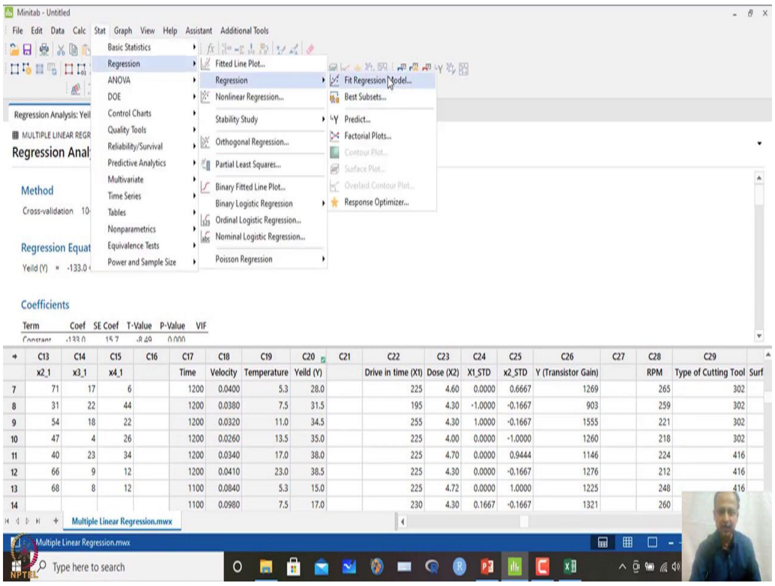
Fits and Diagnostics for Unusual Observations

Obs	Yeild (Y)	Fit	Resid	Std Resid

➔	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

So this is lack of fit. So, linear model seems to be adequate so we can adopt this one. So, you have to think from practical aspects like that. So, multicollinearity problem when we are considering, both the variables over here, so if I consider both the variables, and in that case we are not getting any multicollinearity issue over here.

(Refer Slide Time: 12:43)



Cross-validation 10-

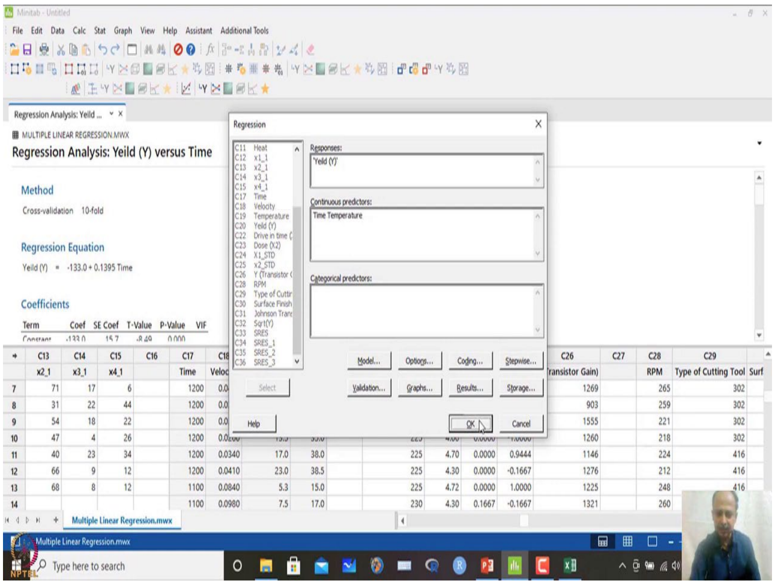
Yeild (Y) = -133.0

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-133.0	15.7	-8.49	0.000	

	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260	

So, this time and temperature if we consider both of them.

(Refer Slide Time: 12:46)



MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time

Method

Cross-validation 10-fold

Regression Equation

Yeild (Y) = -133.0 + 0.1395 Time

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-133.0	15.7	-8.49	0.000	

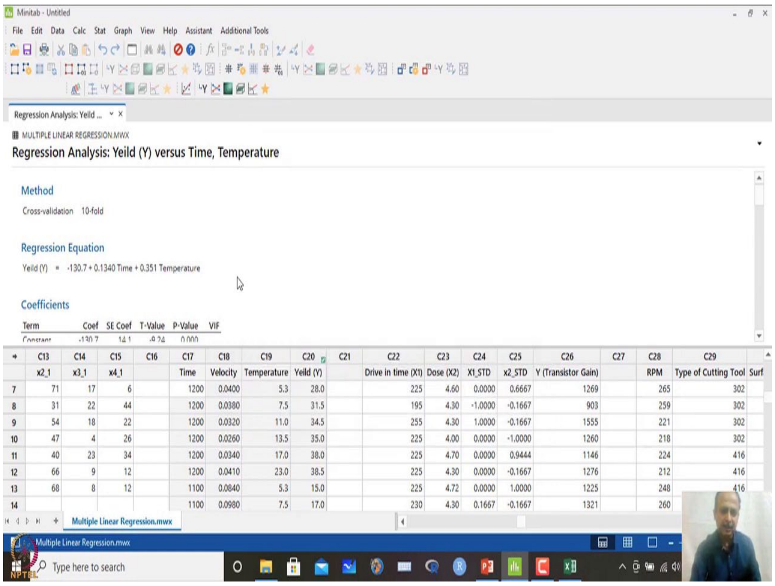
Responses: 'Yeild (Y)'

Continuous predictors: Time Temperature

Categorical predictors:

And variation inflation factor is around 1.05, so which is quite satisfactory.

(Refer Slide Time: 12:56)



MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time, Temperature

Method

Cross-validation 10-fold

Regression Equation

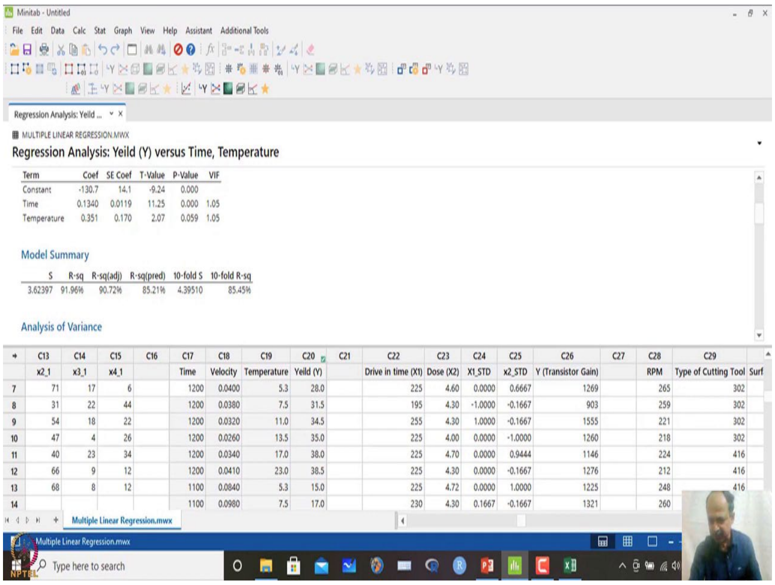
Yeild (Y) = -130.7 + 0.1340 Time + 0.351 Temperature

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-130.7	14.1	-9.24	0.000	

↓	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

(Refer Slide Time: 12:56)



MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time, Temperature

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-130.7	14.1	-9.24	0.000	
Time	0.1340	0.0119	11.25	0.000	1.05
Temperature	0.351	0.170	2.07	0.059	1.05

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
3.62397	91.96%	90.72%	85.21%	4.39510	85.45%

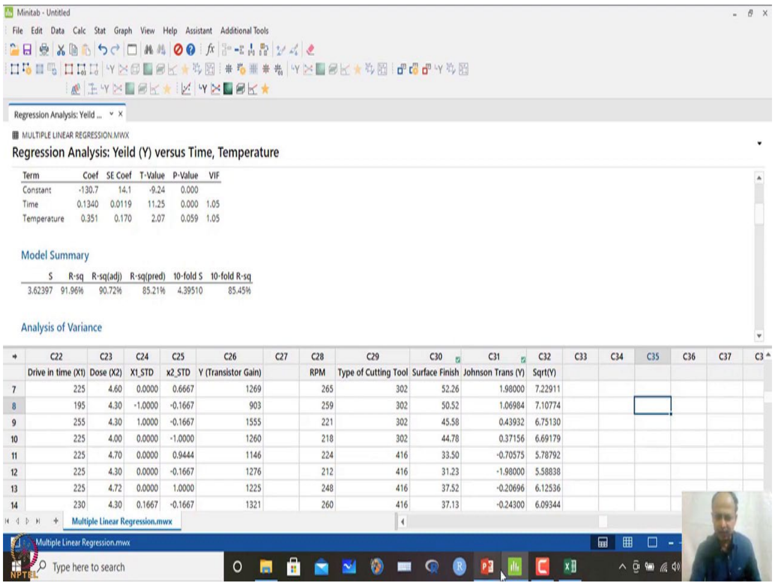
Analysis of Variance

↓	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	
	x2_1	x3_1	x4_1		Time	Velocity	Temperature	Yeild (Y)		Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surf
7	71	17	6		1200	0.0400	5.3	28.0		225	4.60	0.0000	0.6667	1269		265	302	
8	31	22	44		1200	0.0380	7.5	31.5		195	4.30	-1.0000	-0.1667	903		259	302	
9	54	18	22		1200	0.0320	11.0	34.5		255	4.30	1.0000	-0.1667	1555		221	302	
10	47	4	26		1200	0.0260	13.5	35.0		225	4.00	0.0000	-1.0000	1260		218	302	
11	40	23	34		1200	0.0340	17.0	38.0		225	4.70	0.0000	0.9444	1146		224	416	
12	66	9	12		1200	0.0410	23.0	38.5		225	4.30	0.0000	-0.1667	1276		212	416	
13	68	8	12		1100	0.0840	5.3	15.0		225	4.72	0.0000	1.0000	1225		248	416	
14					1100	0.0980	7.5	17.0		230	4.30	0.1667	-0.1667	1321		260		

And, so whether to include both the variables or not to include both the variables, that is the judgment that you have to take. But, I will suggest to retain this one because this is improving R square adjusted and also tenfold cross-validation and R square predicted is also improving so we can retain this one ok.

So, but there is no black and white scenario using regression like that. So it depends on the process engineers and then see the predictive behaviors and then we try to adopt whichever model is very close ok. Suggested model over here is, we can include time and temperature both the variables. But if you go by significance, in that case temperature may be dropped and we stick to time only that is the only variable ok.

(Refer Slide Time: 13:46)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...
MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time, Temperature

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-130.7	14.1	-9.24	0.000	
Time	0.1340	0.0119	11.25	0.000	1.05
Temperature	0.351	0.170	2.07	0.059	1.05

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
3.62397	91.96%	90.72%	85.21%	4.39510	85.45%

Analysis of Variance

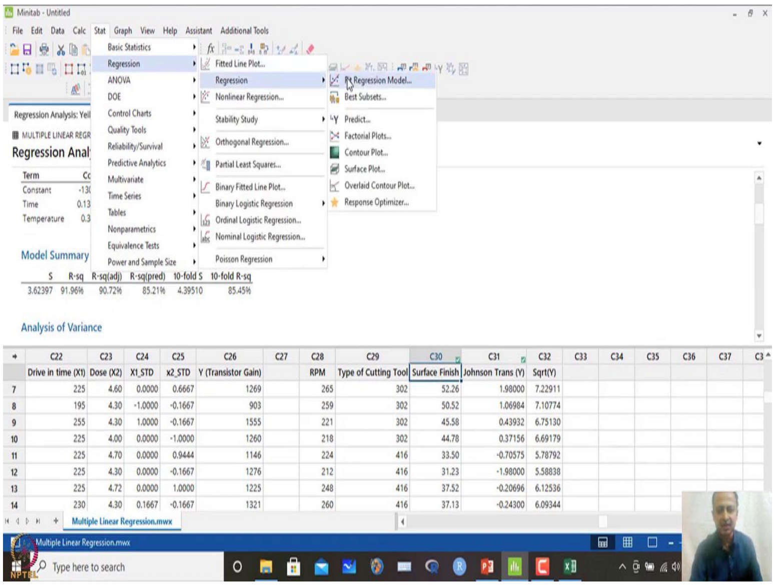
➜	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)					
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911					
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774					
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130					
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179					
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792					
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838					
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536					
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344					

Multiple Linear Regression.mwx
Multiple Linear Regression.mwx
Type here to search
NPTEL

So, this is one scenario. And, we can have in regression some other scenarios like that we can place this is the example, I want to place over here. Maybe, we will delete this one residual over here. So, this is one example that we are taking, so RPN type of cutting tool and surface finish this is the variable surface finish is Y and RPM is the continuous variable and there can be a categorical variable, that means, or type of the cutting tool.

So, here the number is categorical. It can take two values, 302 and 416 like that. So this is the two types of tool. So this is categorical variable basically, so this is no I cannot say 302 is greater than 416 like that. So this cannot be arranged in ascending order, descending order. So this is like color, different types of color. So, this is categorical variable we have to treat, and regression has an option to deal with categorical variable also.

(Refer Slide Time: 14:34)

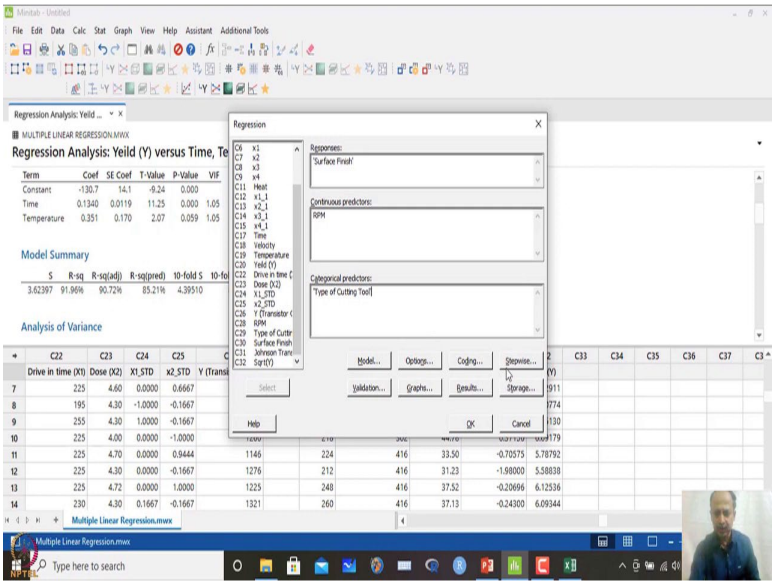


S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
3.62397	91.96%	90.72%	85.21%	4.39510	85.45%

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344

So, surface finish is Y over here. So what we can do is that stat go to stat regression, fit regression and fit regression model over here.

(Refer Slide Time: 14:38)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...
MULTIPLE LINEAR REGRESSION.MWX
Model Summary
Analysis of Variance
Regression
Model... Options... Coding... Stepwise...
Select Validation... Graphs... Results... Storage...
Help OK Cancel
Multiple Linear Regression.mwx
Type here to search

Regression Analysis: Yeild (Y) versus Time, Te

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-130.7	14.1	-9.24	0.000	
Time	0.1340	0.0119	11.25	0.000	1.05
Temperature	0.351	0.170	2.07	0.059	1.05

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S
3.62397	91.96%	90.72%	85.21%	4.39510

C6 x1, C7 x2, C8 x3, C9 x4, C11 Heat, C12 x1_1, C13 x2_1, C14 x3_1, C15 x4_1, C17 Time, C18 Velocity, C19 Temperature, C20 Yeild (Y), C22 Drive in time (, C23 Dose (X2), C24 X1_STD, C25 x2_STD, C26 Y (Transistor (, C28 RPM, C29 Type of Cuttir, C30 Surface Finish, C31 Johnson Trans, C32 Sqrt(Y)

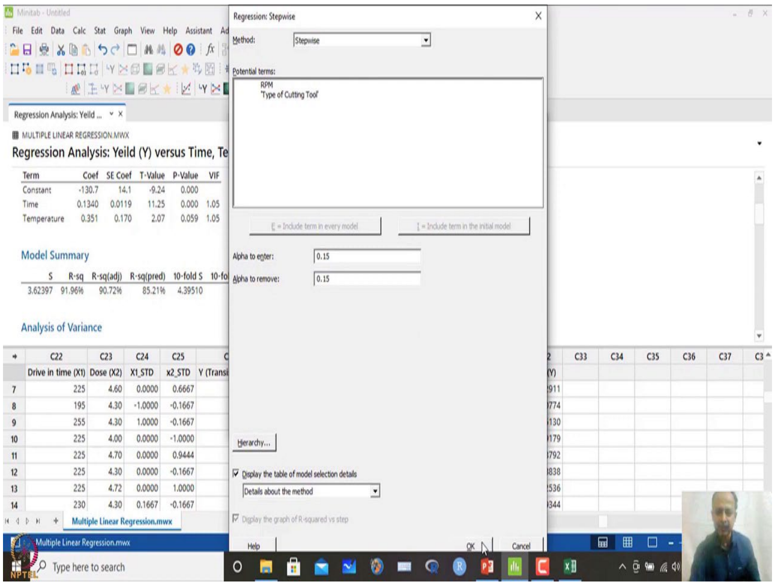
Responses: 'Surface Finish'

Continuous predictors: RPM

Categorical predictors: 'Type of Cutting Tool'

And, instead of this variables over here I can just use continuous variable and categorical predictor also we can include in the regression model over here. So in this case, what we will do is that, we will we will incorporate the response as surface finish which is the actual variable and then continuous predictor over here is RPN and then categorical predictor over here is will be type of cutting tool over here.

(Refer Slide Time: 15:07)



Regression Analysis: Yeild (Y) versus Time, Te

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-130.7	14.1	-9.24	0.000	
Time	0.1340	0.0119	11.25	0.000	1.05
Temperature	0.351	0.170	2.07	0.059	1.05

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S
3.62397	91.96%	90.72%	85.21%	4.39510

Analysis of Variance

	C22	C23	C24	C25
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD
7	225	4.60	0.0000	0.6667
8	195	4.30	-1.0000	-0.1667
9	255	4.30	1.0000	-0.1667
10	225	4.00	0.0000	-1.0000
11	225	4.70	0.0000	0.9444
12	225	4.30	0.0000	-0.1667
13	225	4.72	0.0000	1.0000
14	230	4.30	0.1667	-0.1667

Regression: Stepwise

Method: Stepwise

Potential terms:
RPM
'Type of Cutting Tool'

Alpha to enter: 0.15

Alpha to remove: 0.15

☑ Display the table of model selection details

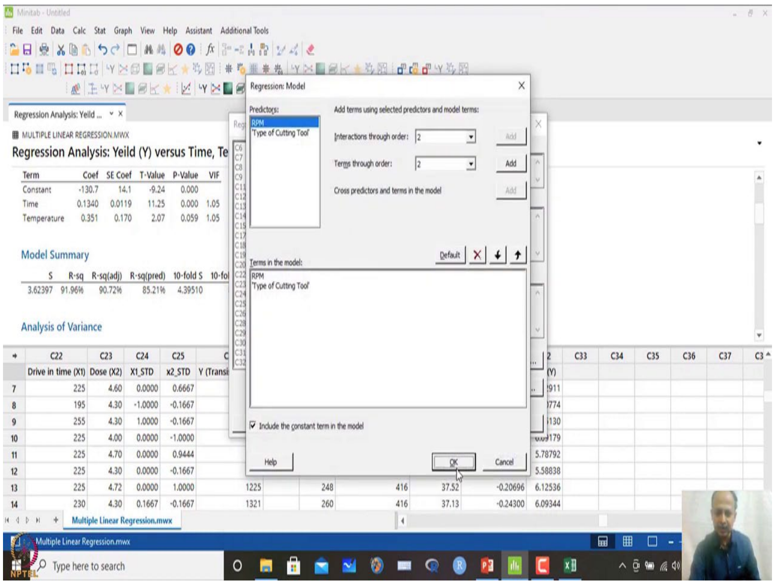
Details about the method

☑ Display the graph of R-squared vs step

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant
Regression Analysis: Yeild ...
MULTIPLE LINEAR REGRESSION.MWX
E = Include term in every model
I = Include term in the initial model
Hierarchy...
Help
OK
Cancel
Multiple Linear Regression.mwx
Type here to search

So, in this case stepwise regression we can use and we can see which variables will go in and which will go out. So same significance level we have used.

(Refer Slide Time: 15:14)



Regression: Model

Predictors:
RPM
'Type of Cutting Tool'

Add terms using selected predictors and model terms:

Interactions through order: 2

Terms through order: 2

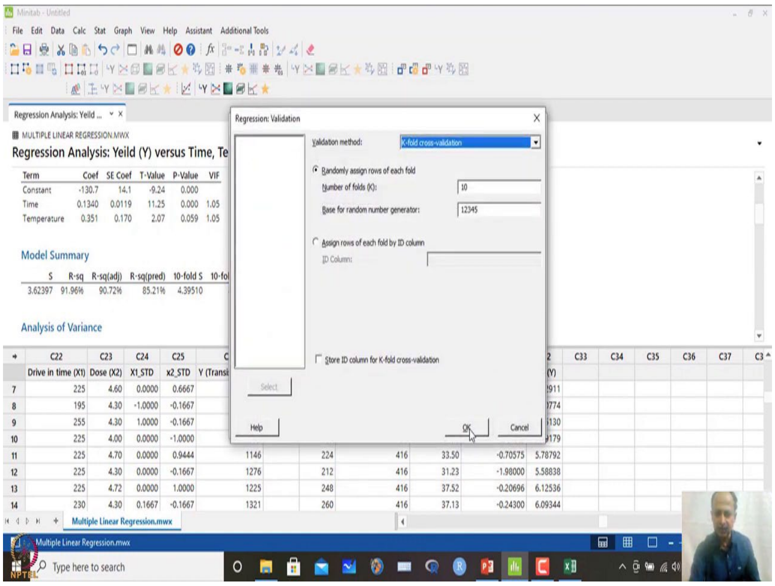
Cross predictors and terms in the model

Terms in the model:
RPM
'Type of Cutting Tool'

☑ Include the constant term in the model

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...
MULTIPLE LINEAR REGRESSION.MWX
Regression Analysis: Yeild (Y) versus Time, Te
Model Summary
Analysis of Variance
Add
Default
Help
OK
Cancel
Multiple Linear Regression.mwx
Type here to search

(Refer Slide Time: 15:18)



MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time, Te

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-130.7	14.1	-9.24	0.000	
Time	0.1340	0.0119	11.25	0.000	1.05
Temperature	0.351	0.170	2.07	0.059	1.05

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fol
3.62397	91.96%	90.72%	85.21%	4.39510	

Analysis of Variance

	C22	C23	C24	C25
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD
7	225	4.60	0.0000	0.6667
8	195	4.30	-1.0000	-0.1667
9	255	4.30	1.0000	-0.1667
10	225	4.00	0.0000	-1.0000
11	225	4.70	0.0000	0.9444
12	225	4.30	0.0000	-0.1667
13	225	4.72	0.0000	1.0000
14	230	4.30	0.1667	-0.1667

Regression: Validation

Validation method: K-fold cross-validation

● Randomly assign rows of each fold

Number of folds (K): 10

Base for random number generator: 12345

○ Assign rows of each fold by ID column

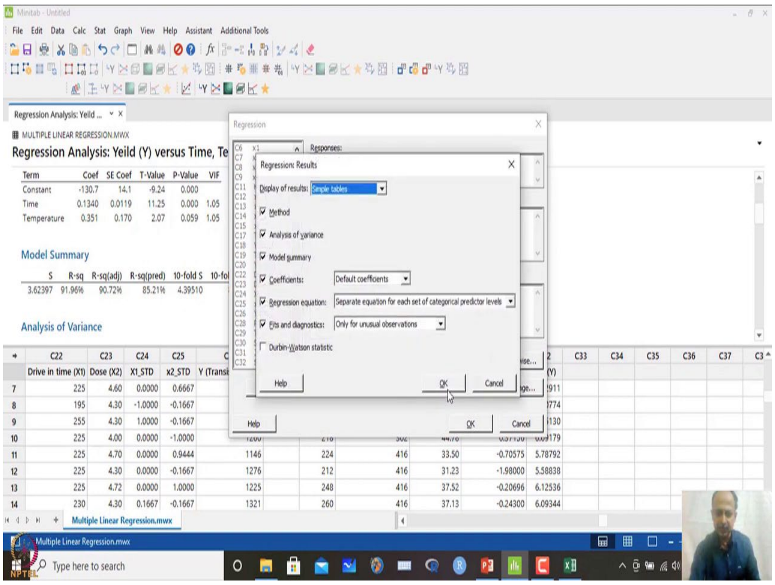
ID Column:

□ Store ID column for K-fold cross-validation

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...
Select
Help
OK
Cancel
Multiple Linear Regression.mwx
Type here to search

And, in models we have included both the variables and included the constant term also, validation also we have taken, cross-validations over here.

(Refer Slide Time: 15:21)



MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time, Te

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-130.7	14.1	-9.24	0.000	
Time	0.1340	0.0119	11.25	0.000	1.05
Temperature	0.351	0.170	2.07	0.059	1.05

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fol
3.62397	91.96%	90.72%	85.21%	4.39510	

Analysis of Variance

Regression: Results

Display of results: Simple tables

☑ Method

☑ Analysis of variance

☑ Model summary

☑ Coefficients: Default coefficients

☑ Regression equation: Separate equation for each set of categorical predictor levels

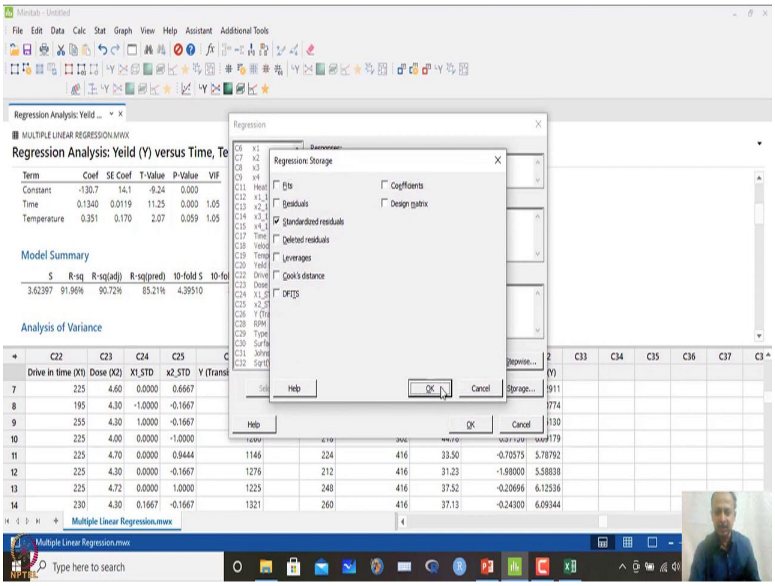
☑ Fits and diagnostics: Only for unusual observations

□ Durbin-Watson statistic

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression
Responses:
Help
OK
Cancel
Multiple Linear Regression.mwx
Type here to search

And results, what we can do is that we can see all results over here. And here, there is a options of Durbin Watson statistics which has to be compared with tabulated value, then we can use this one. So, I am not using because I can convert into R I can go to R and model that one and see Breusch Pagan test and also the DW test and corresponding P-Values.

(Refer Slide Time: 15:41)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...
Regression
Multiple Linear Regression.mwx
Type here to search

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Yeild (Y) versus Time, Te

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-130.7	14.1	-9.24	0.000	
Time	0.1340	0.0119	11.25	0.000	1.05
Temperature	0.351	0.170	2.07	0.059	1.05

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S
3.62397	91.96%	90.72%	85.21%	4.39510

Analysis of Variance

Regression: Storage

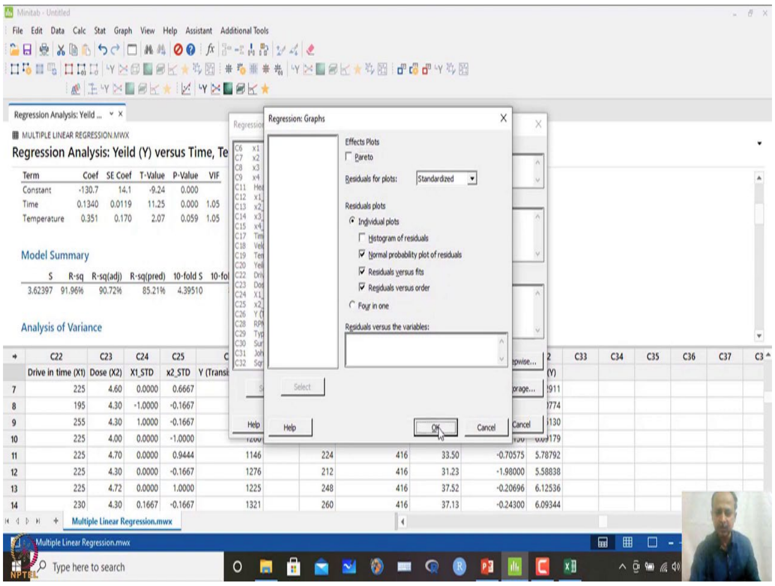
- ☐ Fits
- ☐ Residuals
- ☑ Standardized residuals
- ☐ Deleted residuals
- ☐ Leverages
- ☐ Cook's distance
- ☐ DFITS
- ☐ Coefficients
- ☐ Design matrix

Help | OK | Cancel

	C22	C23	C24	C25
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD
7	225	4.60	0.0000	0.6667
8	195	4.30	-1.0000	-0.1667
9	255	4.30	1.0000	-0.1667
10	225	4.00	0.0000	-1.0000
11	225	4.70	0.0000	0.9444
12	225	4.30	0.0000	-0.1667
13	225	4.72	0.0000	1.0000
14	230	4.30	0.1667	-0.1667

What I will do is that, standardized residual I will save over here again.

(Refer Slide Time: 15:45)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Yeild ...
Regression
Select
Multiple Linear Regression.mwx
Type here to search

Regression: Graphs

Effects Plots

- ☐ Pareto

Residuals for plots: Standardized

Residuals plots

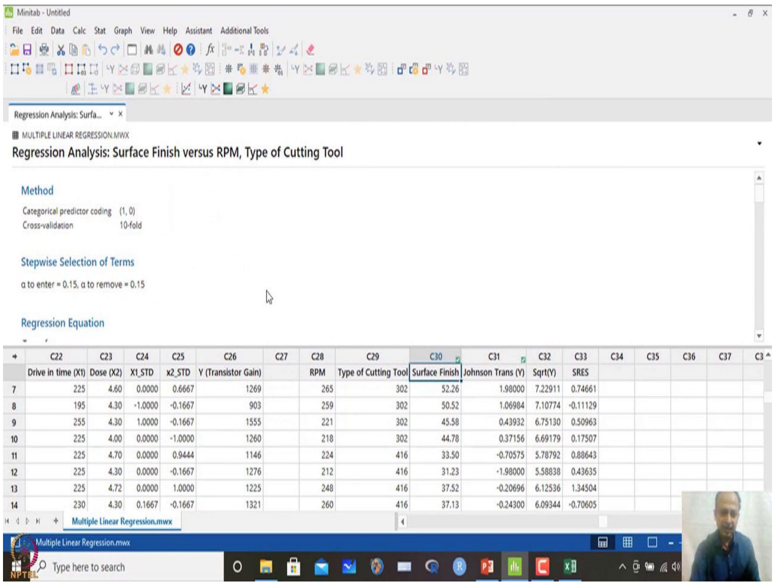
- ● Individual plots
  - ☐ Histogram of residuals
  - ☑ Normal probability plot of residuals
  - ☑ Residuals versus fits
  - ☑ Residuals versus order
- ○ Four in one

Residuals versus the variables:

Help | OK | Cancel

And then graphically what we want to see, normal plot, residual plot, and order of the data. Plots like that.

(Refer Slide Time: 15:52)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Surfa...
Multiple Linear Regression.mwx
Type here to search

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Surface Finish versus RPM, Type of Cutting Tool

Method

Categorical predictor coding	(1, 0)
Cross-validation	10-fold

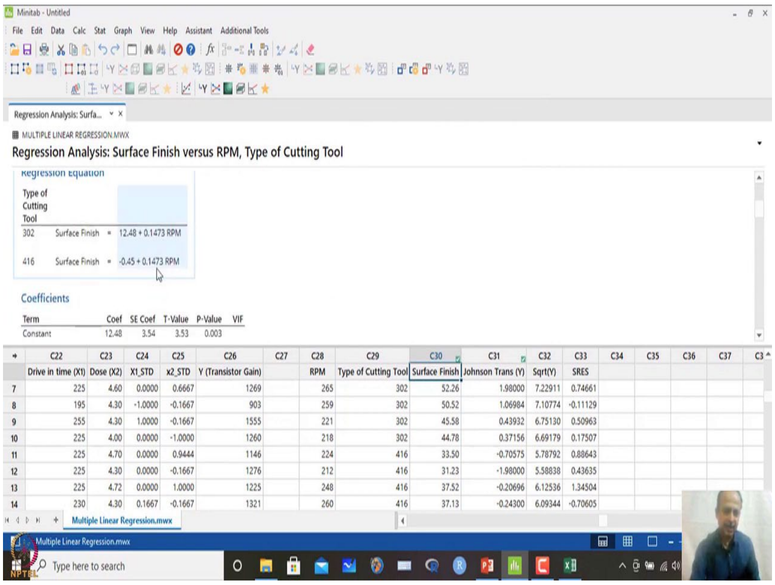
Stepwise Selection of Terms

α to enter = 0.15, α to remove = 0.15

Regression Equation

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES				
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.74661				
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.11129				
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.50963				
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.17507				
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.88643				
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.43635				
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.34504				
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.70605				

(Refer Slide Time: 15:53)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Surfa...
Multiple Linear Regression.mwx
Type here to search

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Surface Finish versus RPM, Type of Cutting Tool

Regression Equation

Type of Cutting Tool			
302	Surface Finish	=	12.48 + 0.1473 RPM
416	Surface Finish	=	-0.45 + 0.1473 RPM

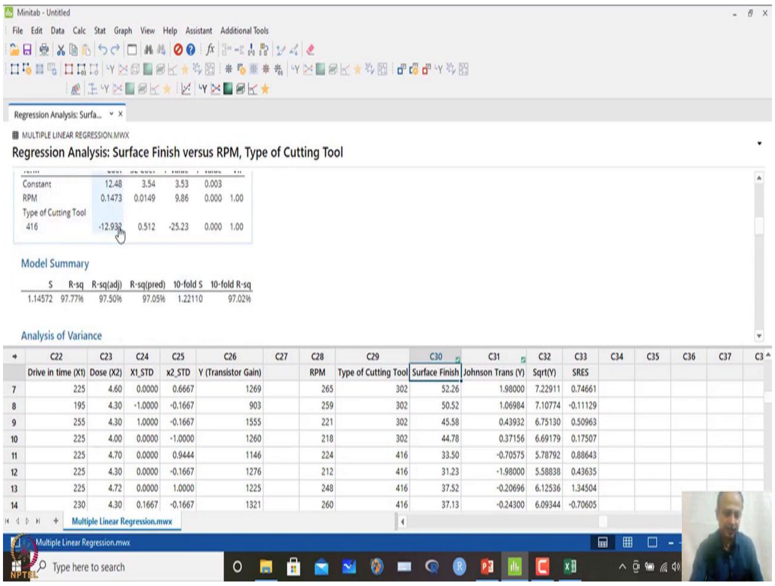
Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	12.48	3.54	3.53	0.003	

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES				
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.74661				
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.11129				
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.50963				
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.17507				
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.88643				
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.43635				
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.34504				
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.70605				

So, if you click ok what happens is that, it its suggest selection process and then what we have selected like that by default. And, these are the two for 302 types of cutting tool, this is the surface finish is related to RPM, this is the equation, and for 416 it will give a different equation.

(Refer Slide Time: 16:15)



Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Surfa...
MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Surface Finish versus RPM, Type of Cutting Tool

Constant	12.48	3.54	3.53	0.003	
RPM	0.1473	0.0149	9.86	0.000	1.00
Type of Cutting Tool					
416	-12.932	0.512	-25.23	0.000	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
1.14572	97.77%	97.50%	97.05%	1.22110	97.02%

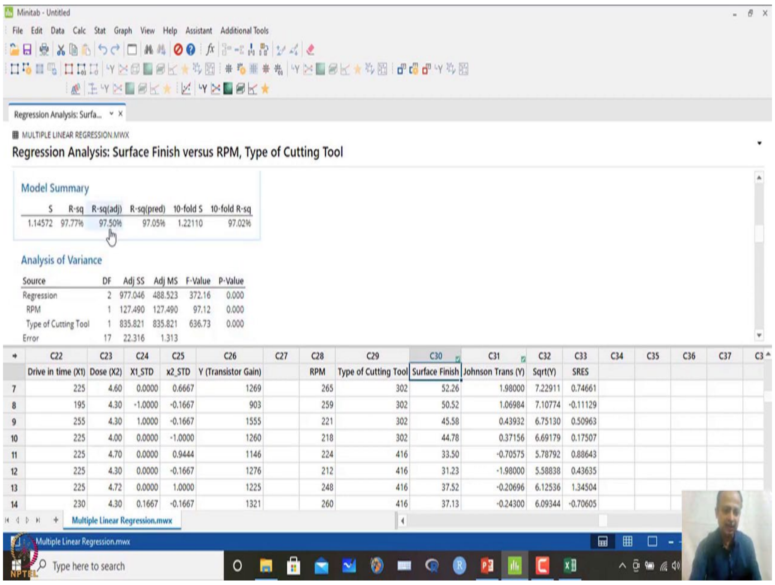
Analysis of Variance

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES				
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.74661				
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.11129				
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.50963				
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.17507				
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.88643				
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.43635				
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.34504				
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.70605				

Multiple Linear Regression.mwx
Multiple Linear Regression.mwx
Type here to search

So, categorical variables, different levels that we have selected, each levels we will have a different equation with the continuous variable. So, this is shown over here.

(Refer Slide Time: 16:30)



Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Surfa...
MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Surface Finish versus RPM, Type of Cutting Tool

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
1.14572	97.77%	97.50%	97.05%	1.22110	97.02%

Analysis of Variance

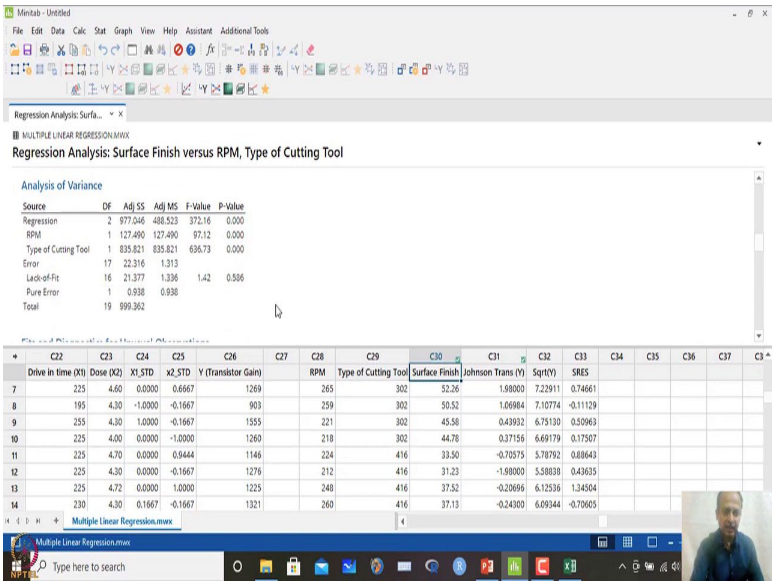
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	977.046	488.523	372.16	0.000
RPM	1	127.490	127.490	97.12	0.000
Type of Cutting Tool	1	835.821	835.821	636.73	0.000
Error	17	22.316	1.313		

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES				
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.74661				
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.11129				
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.50963				
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.17507				
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.88643				
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.43635				
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.34504				
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.70605				

Multiple Linear Regression.mwx
Multiple Linear Regression.mwx
Type here to search

And, the coefficients are also given over here, so based on which we have developed the regression equation. And, VIC, variation inflation factors is approximated, so that is not an issue. So, these two variables type of cutting tool and also RPN both are significant over here. And, about R square adjusted value is 97.5 and cross-validation is 97.02, so very close. So, this model seems to be very accurate.

(Refer Slide Time: 16:38)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Surfa...
Multiple Linear Regression.mwx
Type here to search

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Surface Finish versus RPM, Type of Cutting Tool

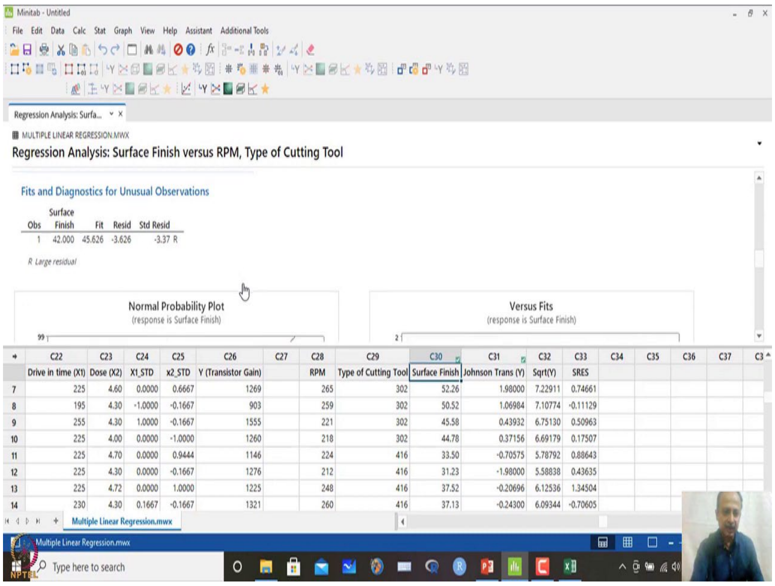
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	977.046	488.523	372.16	0.000
RPM	1	127.490	127.490	97.12	0.000
Type of Cutting Tool	1	835.821	835.821	636.73	0.000
Error	17	22.316	1.313		
Lack-of-Fit	16	21.377	1.336	1.42	0.586
Pure Error	1	0.938	0.938		
Total	19	999.362			

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C3
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES					
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.74661					
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.11129					
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.50963					
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.17507					
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.88643					
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.43635					
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.34504					
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.70605					

And, what we are getting is that no lack of fit is observed over here.

(Refer Slide Time: 16:42)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Surfa...
Normal Probability Plot
(response is Surface Finish)
99
Versus Fits
(response is Surface Finish)
2
Multiple Linear Regression.mwx
Type here to search

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Surface Finish versus RPM, Type of Cutting Tool

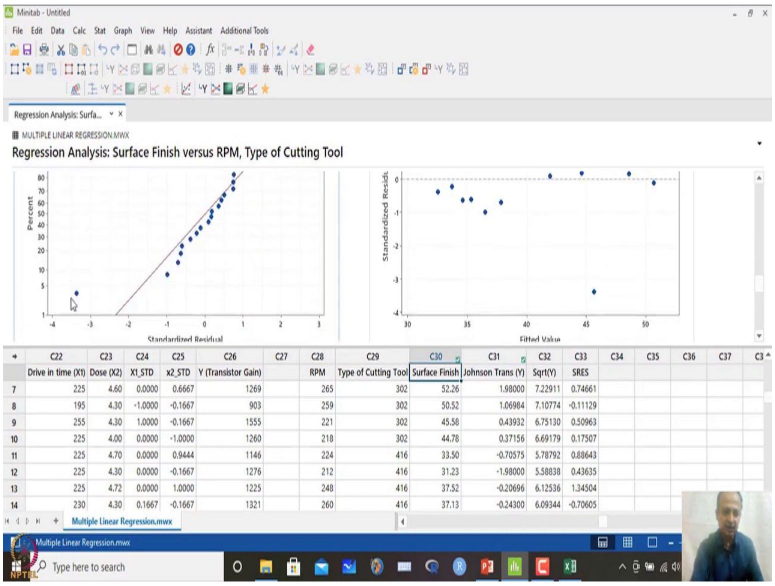
Fits and Diagnostics for Unusual Observations

Obs	Surface Finish	Fit	Resid	Std Resid
1	42.000	45.626	-3.626	-3.37 R

R Large residual

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C3
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES					
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.74661					
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.11129					
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.50963					
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.17507					
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.88643					
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.43635					
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.34504					
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.70605					

(Refer Slide Time: 16:44)



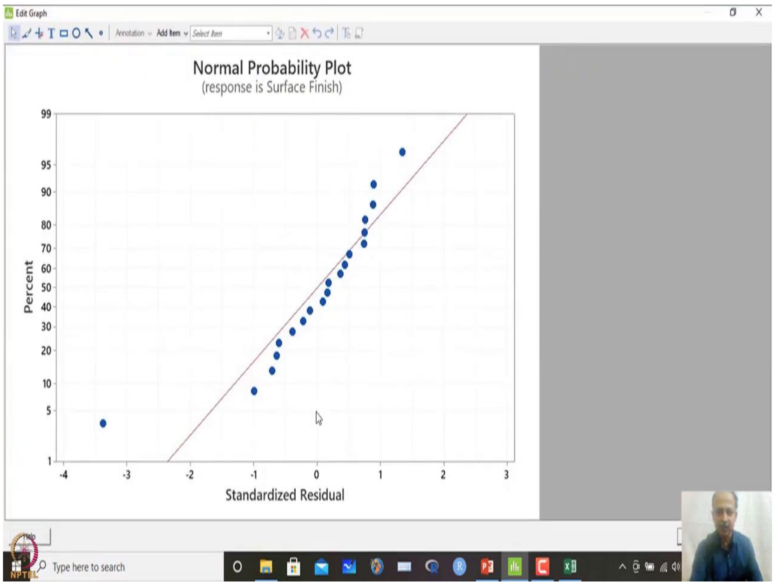
Regression Analysis: Surface Finish versus RPM, Type of Cutting Tool

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Surfa...
MULTIPLE LINEAR REGRESSION.MWX
Percent
Standardized Residual
Standardized Resids
Fitted Value
Multiple Linear Regression.mwx
Type here to search

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES				
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.74661				
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.11129				
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.50963				
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.17507				
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.88643				
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.43635				
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.34504				
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.70605				

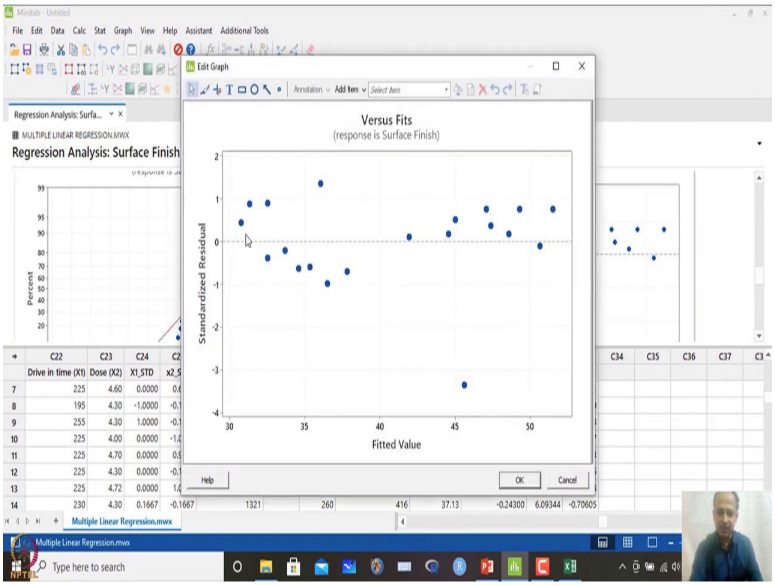
Only thing is that, we have to check whether normality assumptions is violated.

(Refer Slide Time: 16:47).



So, here what you see, is maybe there is a problem with normality over here. One point is over here, some has gone outside like that, so we have to test that one.

(Refer Slide Time: 16:55)

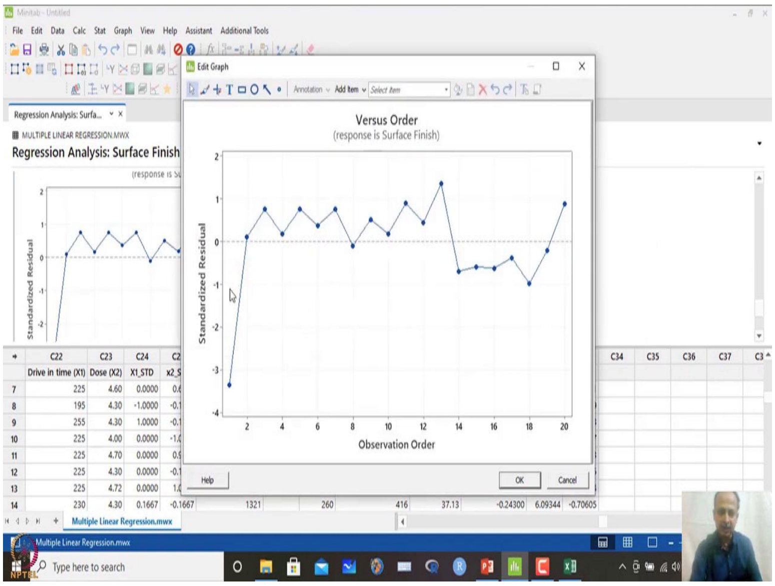

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Edit Graph
Annotation
Add Item
Select Item
Regression Analysis: Surfa...
MULTIPLE LINEAR REGRESSION.MWX
Regression Analysis: Surface Finish
Percent
Versus Fits
(response is Surface Finish)
Standardized Residual
Fitted Value
2
1
0
-1
-2
-3
-4
30
35
40
45
50
Help
OK
Cancel
Multiple Linear Regression.mwx
Type here to search

	C22	C23	C24	C2
	Drive in time (X1)	Dose (X2)	X1_STD	x2_S
7	225	4.60	0.0000	0.(
8	195	4.30	-1.0000	-0.1
9	255	4.30	1.0000	-0.1
10	225	4.00	0.0000	-1.(
11	225	4.70	0.0000	0.5
12	225	4.30	0.0000	-0.1
13	225	4.72	0.0000	1.(
14	230	4.30	0.1667	-0.1667

1321 260 416 37.13 -0.24300 6.09344 -0.70605

And, heteroscedasticity does not seem to be a problem because it seems to be random.

(Refer Slide Time: 17:01)

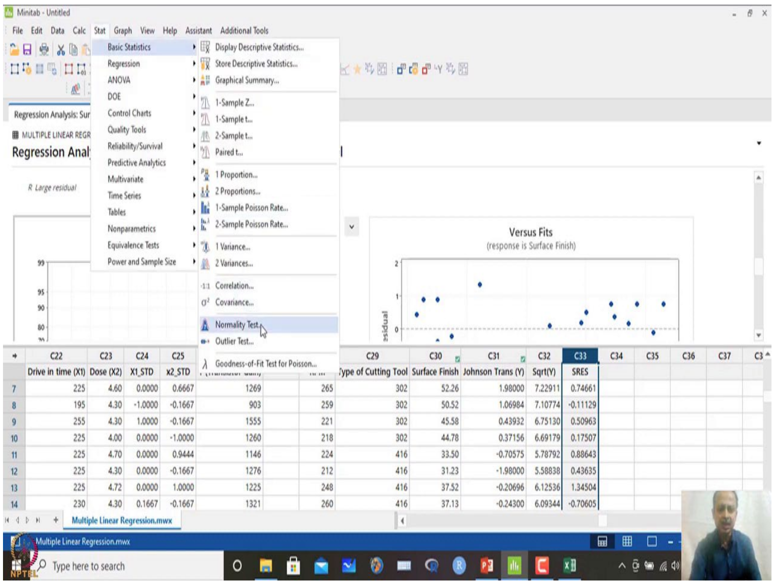

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Edit Graph
Annotation
Add Item
Select Item
Regression Analysis: Surfa...
MULTIPLE LINEAR REGRESSION.MWX
Regression Analysis: Surface Finish
Versus Order
(response is Surface Finish)
Standardized Residual
Observation Order
2
1
0
-1
-2
-3
-4
2
4
6
8
10
12
14
16
18
20
Help
OK
Cancel
Multiple Linear Regression.mwx
Type here to search

	C22	C23	C24	C2
	Drive in time (X1)	Dose (X2)	X1_STD	x2_S
7	225	4.60	0.0000	0.(
8	195	4.30	-1.0000	-0.1
9	255	4.30	1.0000	-0.1
10	225	4.00	0.0000	-1.(
11	225	4.70	0.0000	0.5
12	225	4.30	0.0000	-0.1
13	225	4.72	0.0000	1.(
14	230	4.30	0.1667	-0.1667

1321 260 416 37.13 -0.24300 6.09344 -0.70605

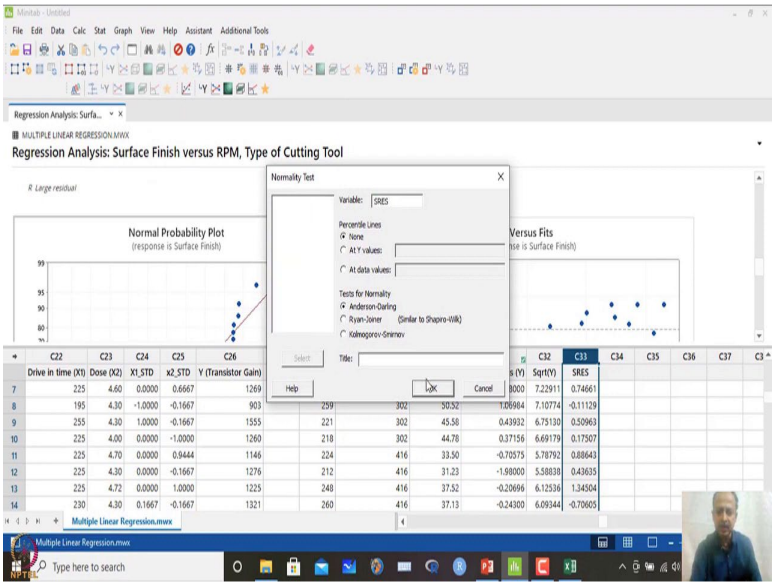
And, also we can see that this one may not be auto correlation is not so significant. So, these two checks can be done, but what we want to see normality, so this C 33 column will tell me whether normality assumptions is violated or not.

(Refer Slide Time: 17:15)



	C22	C23	C24	C25				C29	C30	C31	C32	C33	C34	C35	C36	C37
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD				Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES				
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.74661				
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.11129				
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.50963				
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.17507				
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.88643				
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.43635				
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.34504				
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.70605				

(Refer Slide Time: 17:17)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Surfa...
MULTIPLE LINEAR REGRESSION.MWX
R Large residual
Normal Probability Plot
(response is Surface Finish)
Versus Fits
(response is Surface Finish)
Select
Help
OK
Cancel
Multiple Linear Regression.mwx
Type here to search

Regression Analysis: Surface Finish versus RPM, Type of Cutting Tool

Normality Test

Variable: SRES

Percentile Lines
● None
○ At Y values:
○ At data values:

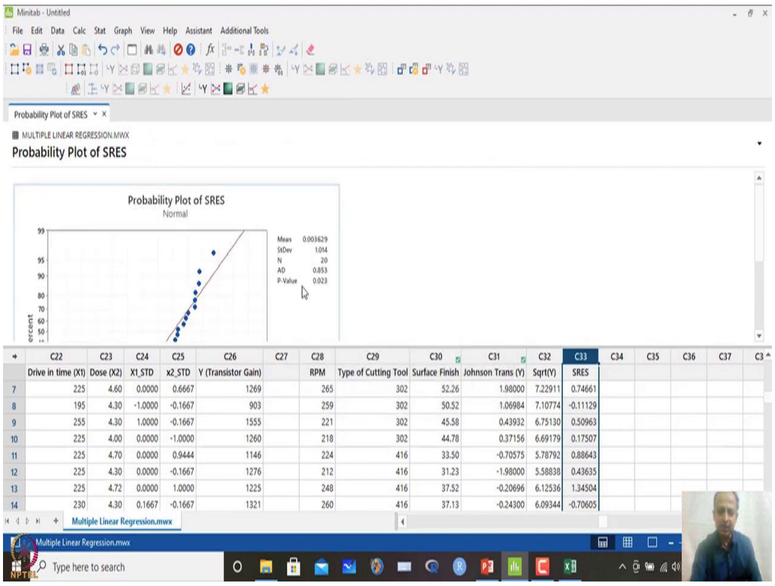
Tests for Normality
● Anderson-Darling
○ Ryan-Joiner (Similar to Shapiro-Wilk)
○ Kolmogorov-Smirnov

Title:

	C22	C23	C24	C25	C26						C32	C33	C34	C35	C36	C37
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)					s (Y)	Sqrt(Y)	SRES				
7	225	4.60	0.0000	0.6667	1269					8000	7.22911	0.74661				
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.11129				
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.50963				
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.17507				
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.88643				
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.43635				
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.34504				
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.70605				

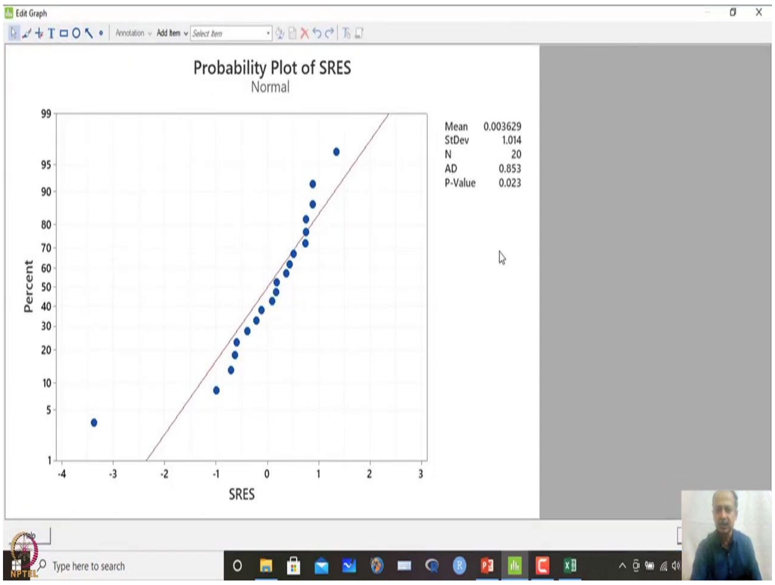
So what I will do is that, basic stat and normality test and what we will do is that, normality residual we can check and try to see what happens.

(Refer Slide Time: 17:22)



↓	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C3
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES					
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.74661					
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.11129					
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.50963					
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.17507					
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.88643					
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.43635					
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.34504					
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.70605					

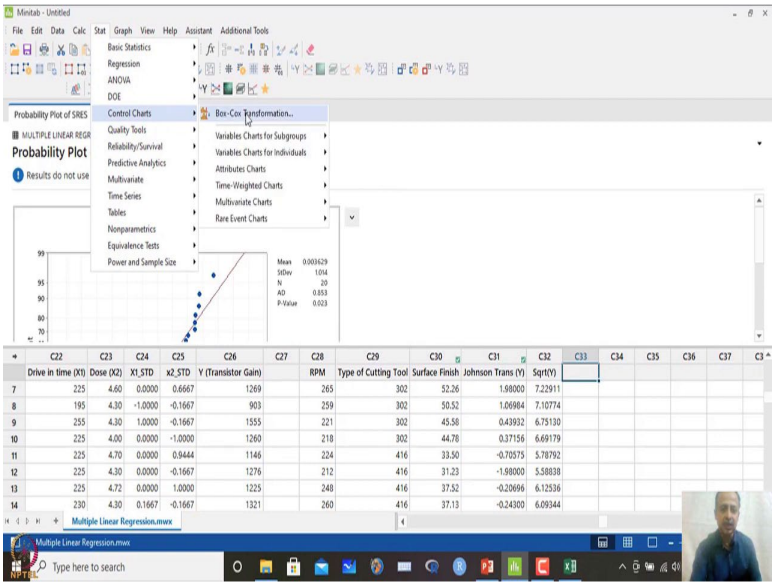
(Refer Slide Time: 17:24)



Normality test and what we observe over here for the residuals is that P-Value is less than 0.05. That indicates that there is a problem of normality issue over here of the residuals.

So, in this case, directly the Y characteristics needs some transformation and we have Box-Cox transformation and also Johnson transformation, both the options are there.

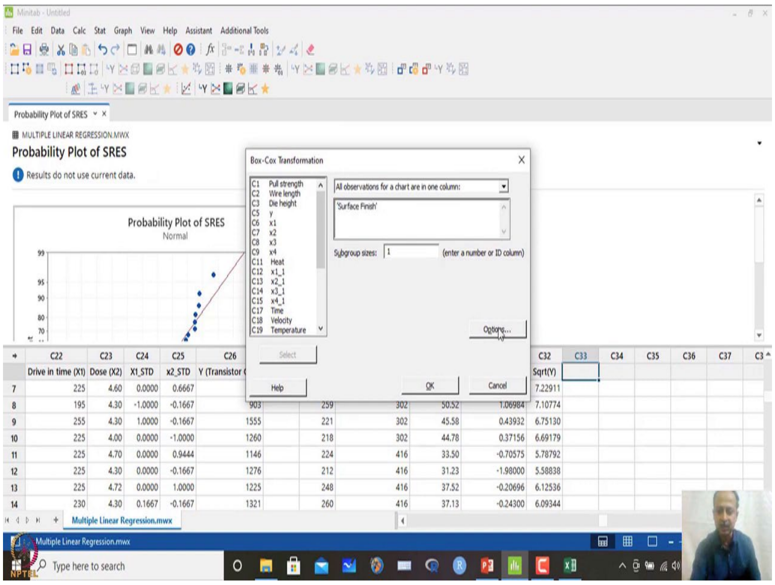
(Refer Slide Time: 17:46)



	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344

So, in this case what we will do is that, first we will see whether Box-Cox transformation works for this, for the Y variable. So what we will do, control chart Box-Cox transformation.

(Refer Slide Time: 17:51)



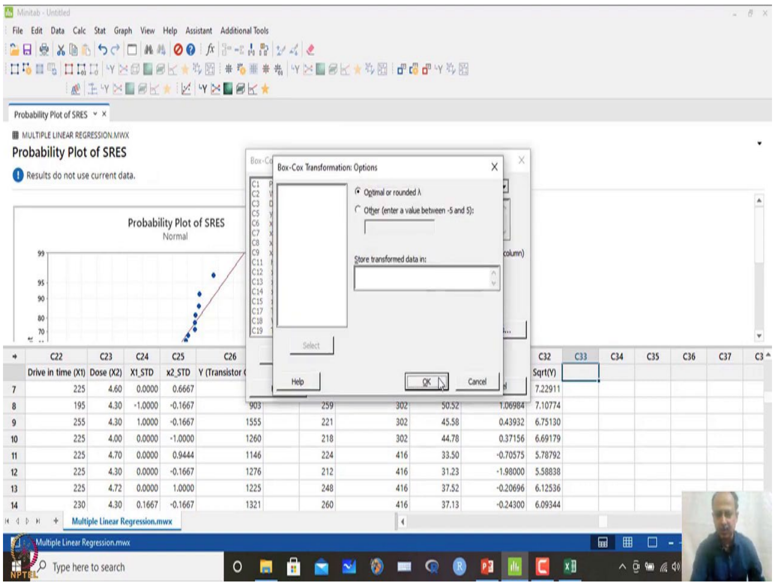
All observations for a chart are in one column:

'Surface Finish'

Subgroup sizes: 1 (enter a number or ID column)

And, for this surface finish we have taken this variable over here. Subgroup size is 1 selected.

(Refer Slide Time: 17:57)



Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Probability Plot of SRES

MULTIPLE LINEAR REGRESSION.MWX

Probability Plot of SRES

Results do not use current data.

Box-Cox Transformation: Options

● Optimal or rounded λ

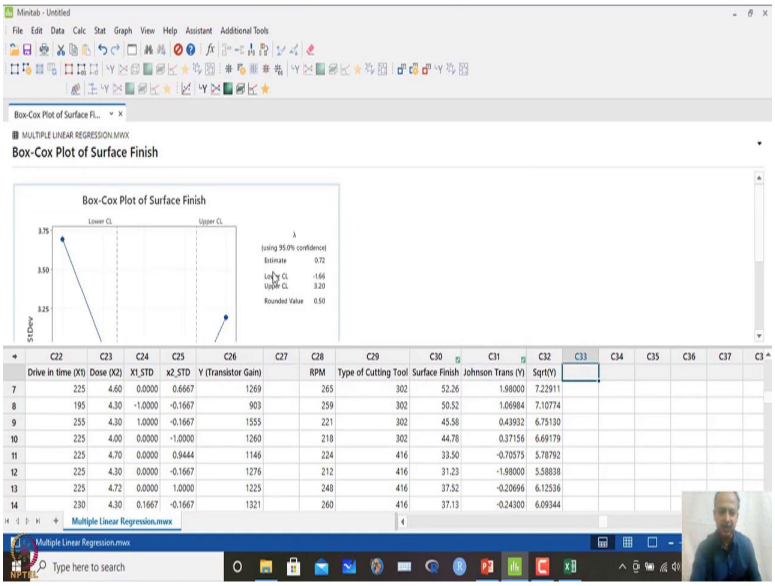
○ Other (enter a value between -5 and 5):

Store transformed data in:

	C22	C23	C24	C25	C32
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Sqrt(Y)
7	225	4.60	0.0000	0.6667	7.22911
8	195	4.30	-1.0000	-0.1667	7.10774
9	255	4.30	1.0000	-0.1667	6.75130
10	225	4.00	0.0000	-1.0000	6.69179
11	225	4.70	0.0000	0.9444	5.78792
12	225	4.30	0.0000	-0.1667	5.58838
13	225	4.72	0.0000	1.0000	6.12536
14	230	4.30	0.1667	-0.1667	6.09344

And in options optimal or rounded value of alpha just mention that one.

(Refer Slide Time: 18:00)



Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Box-Cox Plot of Surface Fi...

MULTIPLE LINEAR REGRESSION.MWX

Box-Cox Plot of Surface Finish

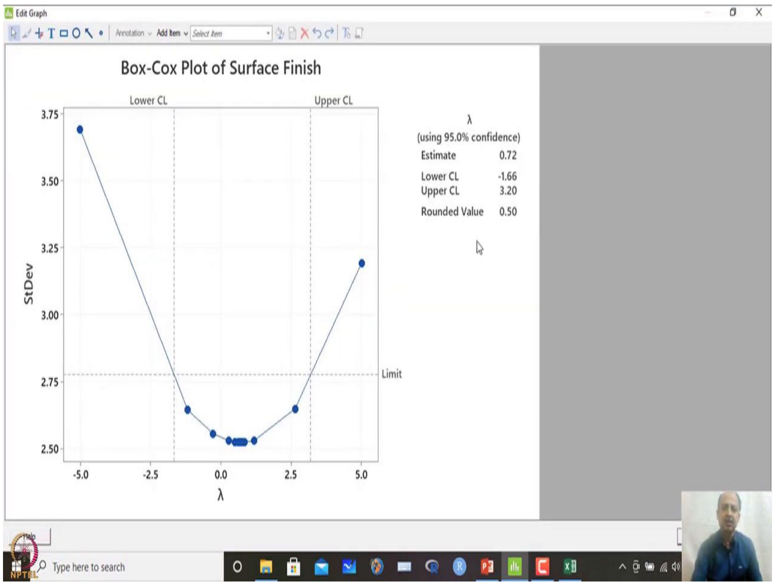
Box-Cox Plot of Surface Finish
Lower CL
Upper CL
3.75
3.50
3.25
StDev

λ
(using 95.0% confidence)

Estimate	0.72
Lower CL	-1.66
Upper CL	3.20
Rounded Value	0.50

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344

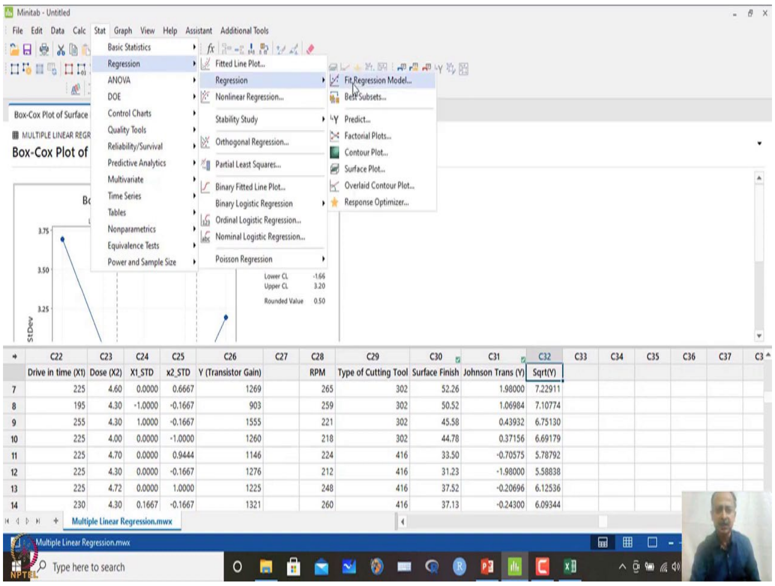
(Refer Slide Time: 18:01)



And when we when you just click that option, what happens is that it suggest you that rounded value is 0.5; that means, lambda is approximately we can take as 0.5, although the actual value is 0.72, but we can take a rounded value of which is lying within the confidence interval over here.

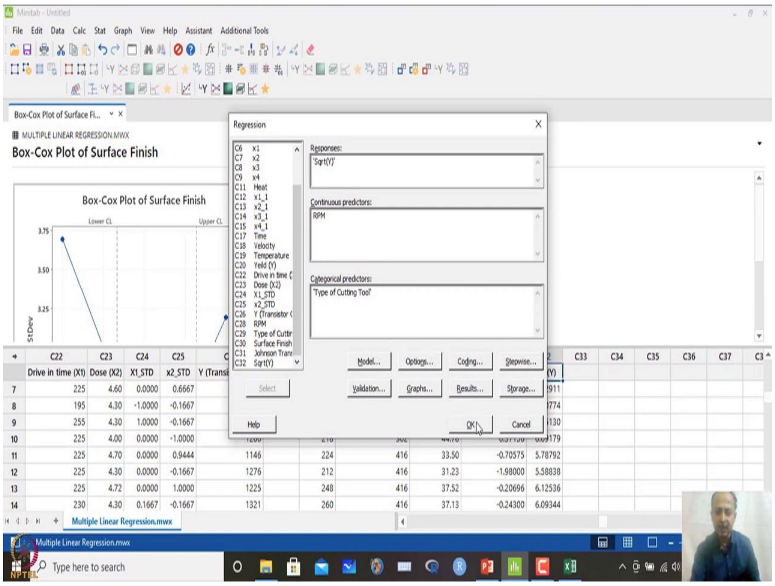
So, 0.5 we are selecting over here as rounded value. So, I have taken a square root transformation over here. So, 0.5 means Y to the power 0.5 means, square root transformation that is suggested ok.

(Refer Slide Time: 18:30)



	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C3
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)						
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911						
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774						
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130						
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179						
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792						
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838						
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536						
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344						

(Refer Slide Time: 18:33)



Regression

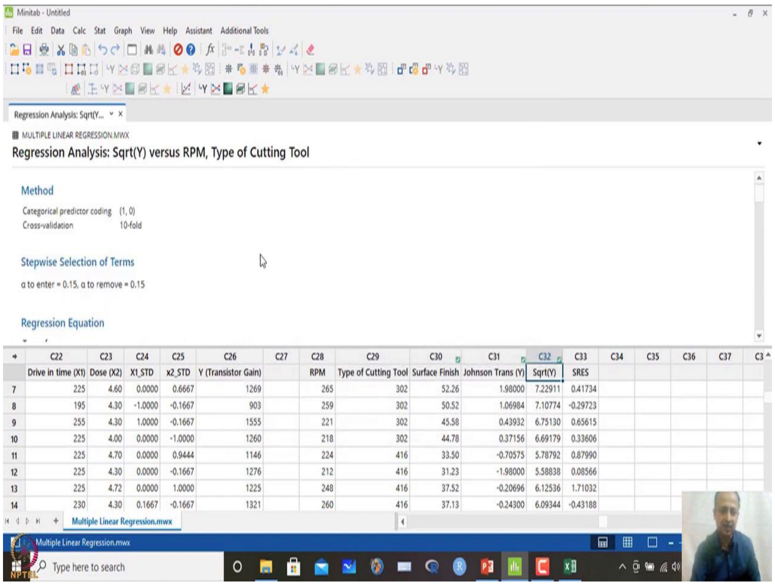
Responses: 'Sqrt(Y)'

Continuous predictors: RPM

Categorical predictors: 'Type of Cutting Tool'

With square root transformation, then what have what we have to do is, that regress variables fit. Instead of surface finish what we will do is that, we will mention that let us go for square root transformation of Y that values and do all the analysis same analysis over here. So, regression analysis over here.

(Refer Slide Time: 18:43)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Sqrt(Y...
MULTIPLE LINEAR REGRESSION.MWX
Multiple Linear Regression.mwx
Type here to search

Regression Analysis: Sqrt(Y) versus RPM, Type of Cutting Tool

Method

Categorical predictor coding (1, 0)
Cross-validation 10-fold

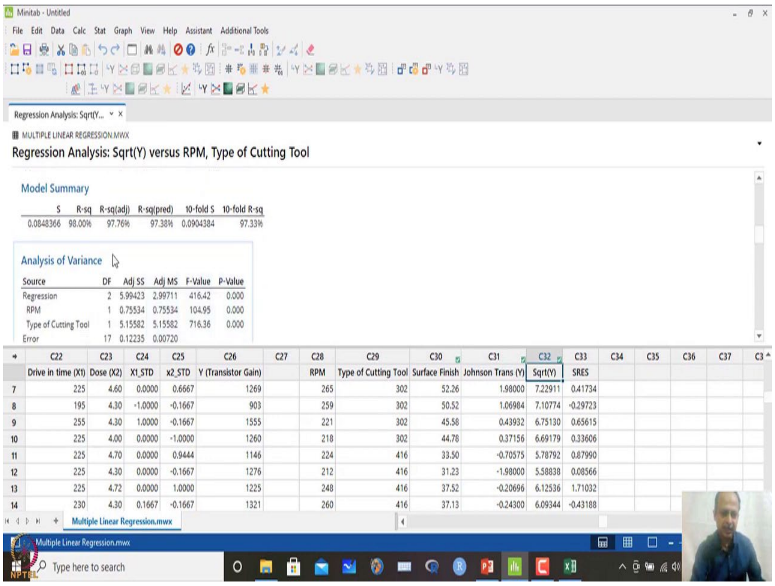
Stepwise Selection of Terms

α to enter = 0.15, α to remove = 0.15

Regression Equation

↓	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C3
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES					
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.41734					
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.29723					
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.65615					
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.33606					
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.87990					
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.08566					
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.71032					
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.43188					

(Refer Slide Time: 18:46)



MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Sqrt(Y) versus RPM, Type of Cutting Tool

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
0.0848366	98.00%	97.76%	97.38%	0.0904384	97.33%

Analysis of Variance

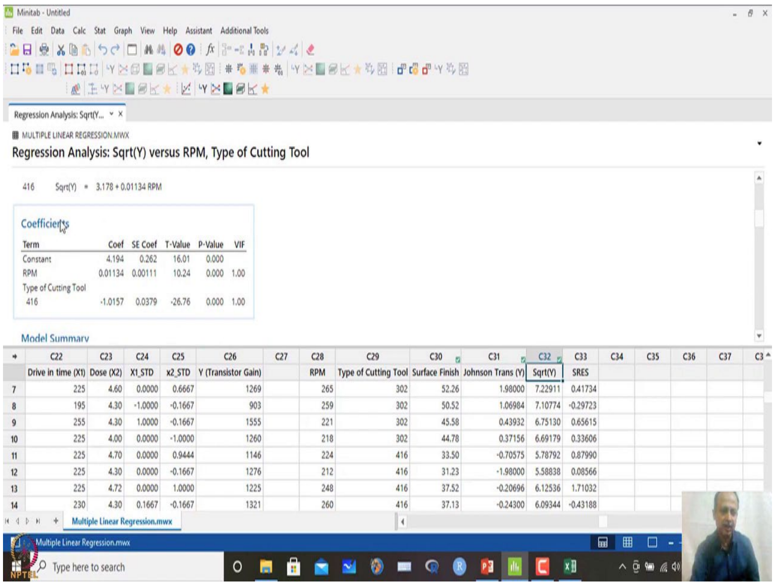
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	5.99423	2.99711	416.42	0.000
RPM	1	0.75534	0.75534	104.95	0.000
Type of Cutting Tool	1	5.15582	5.15582	716.36	0.000
Error	17	0.12235	0.00720		

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.41734
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.29723
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.65615
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.33606
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.87990
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.08566
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.71032
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.43188

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Sqrt(Y...
Multiple Linear Regression.mwx
Multiple Linear Regression.mwx
Type here to search

And, here also we see that R square adjusted 97.

(Refer Slide Time: 18:50)



MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Sqrt(Y) versus RPM, Type of Cutting Tool

416   Sqrt(Y) = 3.178 + 0.01134 RPM

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	4.194	0.262	16.01	0.000	
RPM	0.01134	0.00111	10.24	0.000	1.00
Type of Cutting Tool					
416	-1.0157	0.0379	-26.76	0.000	1.00

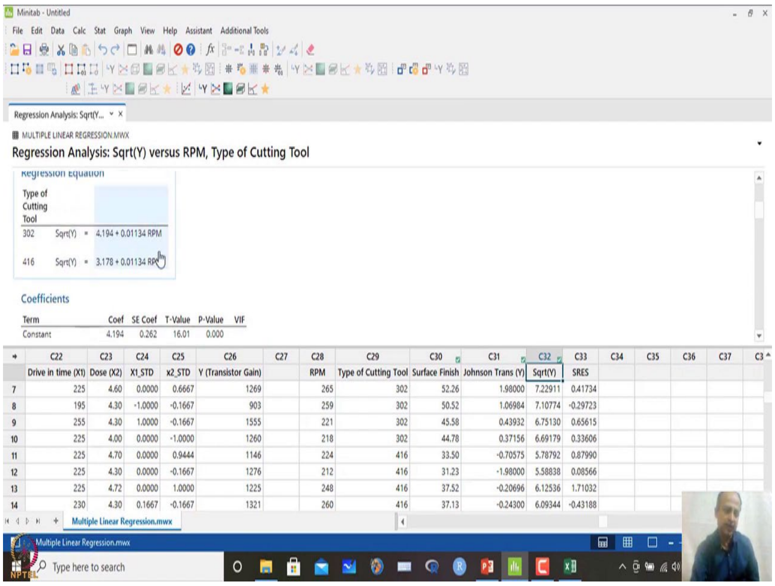
Model Summary

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.41734
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.29723
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.65615
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.33606
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.87990
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.08566
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.71032
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.43188

Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Sqrt(Y...
Multiple Linear Regression.mwx
Multiple Linear Regression.mwx
Type here to search

And so it has regressed square root of Y with the variable.

(Refer Slide Time: 18:53)



Minitab - Untitled

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Sqrt(Y) versus RPM, Type of Cutting Tool

Regression Equation

Type of Cutting Tool			
302	Sqrt(Y)	=	4.194 + 0.01134 RPM
416	Sqrt(Y)	=	3.178 + 0.01134 RPM

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	4.194	0.262	16.01	0.000	

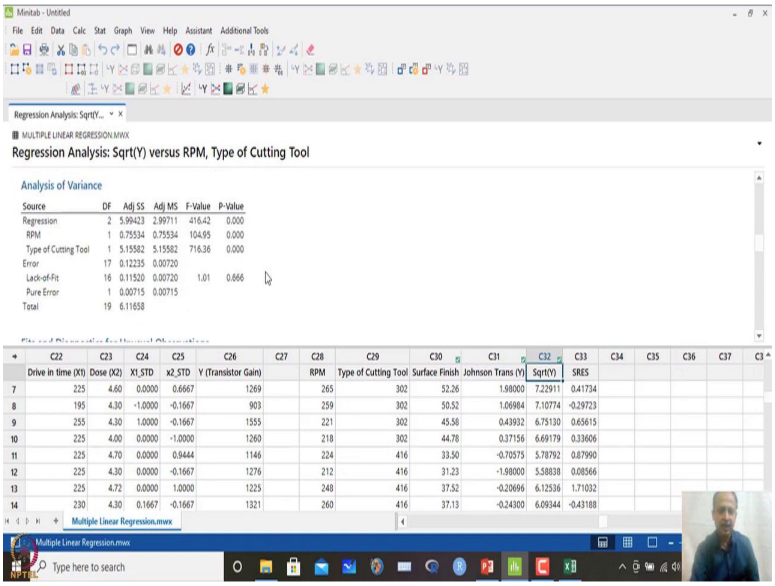
	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.41734
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.29723
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.65615
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.33606
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.87990
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.08566
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.71032
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.43188

Multiple Linear Regression.mwx

File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Sqrt(Y...
Type here to search
NPTEL

And only one with the variable RPN over here. So this both the variables are significant.

(Refer Slide Time: 18:58)



Minitab - Untitled

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Sqrt(Y) versus RPM, Type of Cutting Tool

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	5.99423	2.99711	416.42	0.000
RPM	1	0.75534	0.75534	104.95	0.000
Type of Cutting Tool	1	5.15582	5.15582	716.36	0.000
Error	17	0.12235	0.00720		
Lack-of-Fit	16	0.11520	0.00720	1.01	0.666
Pure Error	1	0.00715	0.00715		
Total	19	6.11658			

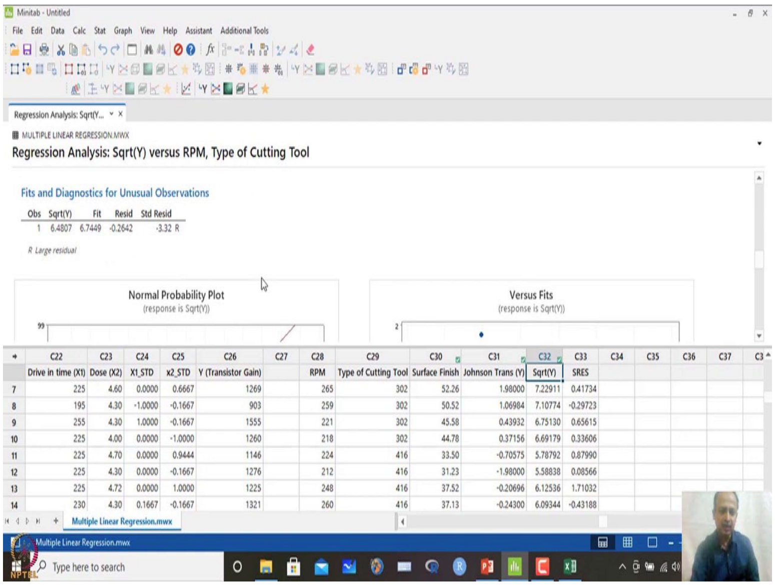
	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.41734
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.29723
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.65615
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.33606
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.87990
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.08566
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.71032
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.43188

Multiple Linear Regression.mwx

File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Sqrt(Y...
Type here to search
NPTEL

And, the equations is given, lack of fit is not prominent.

(Refer Slide Time: 19:01)



Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Sqrt(Y...
MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Sqrt(Y) versus RPM, Type of Cutting Tool

Fits and Diagnostics for Unusual Observations

Obs	Sqrt(Y)	Fit	Resid	Std Resid
1	6.4807	6.7449	-0.2642	-3.32 R

R Large residual

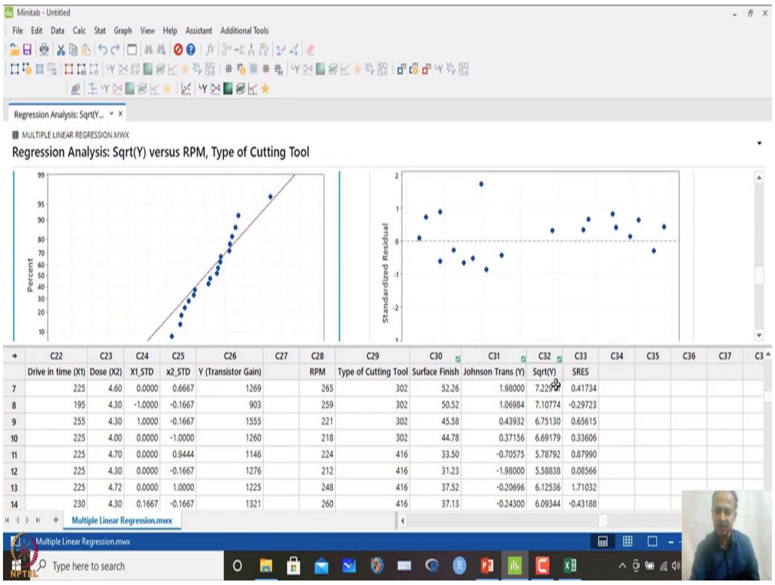
Normal Probability Plot
(response is Sqrt(Y))
99
Versus Fits
(response is Sqrt(Y))
2

↓	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C3
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES					
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.41734					
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.29723					
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.65615					
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.33606					
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.87990					
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.08566					
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.71032					
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.43188					

Multiple Linear Regression.mwx
Multiple Linear Regression.mwx
Type here to search
NPTEL

So in this case, one out layer is recorded over here as it is minus 3 point more than plus or minus 2.

(Refer Slide Time: 19:09)



Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Sqrt(Y...
MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Sqrt(Y) versus RPM, Type of Cutting Tool

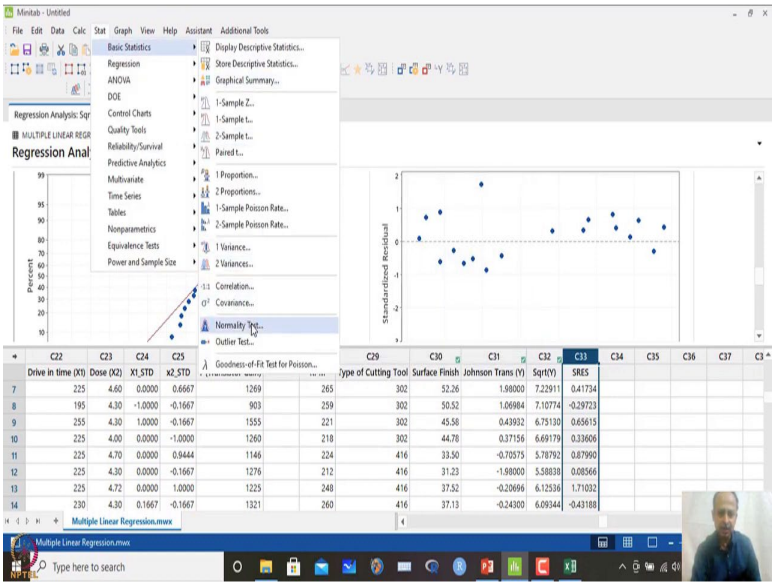
Percent
99
95
90
80
70
60
50
40
30
20
10
Standardized Residual
2
1
0
-1
-2

↓	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C3
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES					
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.229	0.41734					
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.29723					
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.65615					
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.33606					
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.87990					
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.08566					
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.71032					
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.43188					

Multiple Linear Regression.mwx
Multiple Linear Regression.mwx
Type here to search
NPTEL

So in this case, what we have to see is that, again there is a after this transformation there is a residual that is generated over here.

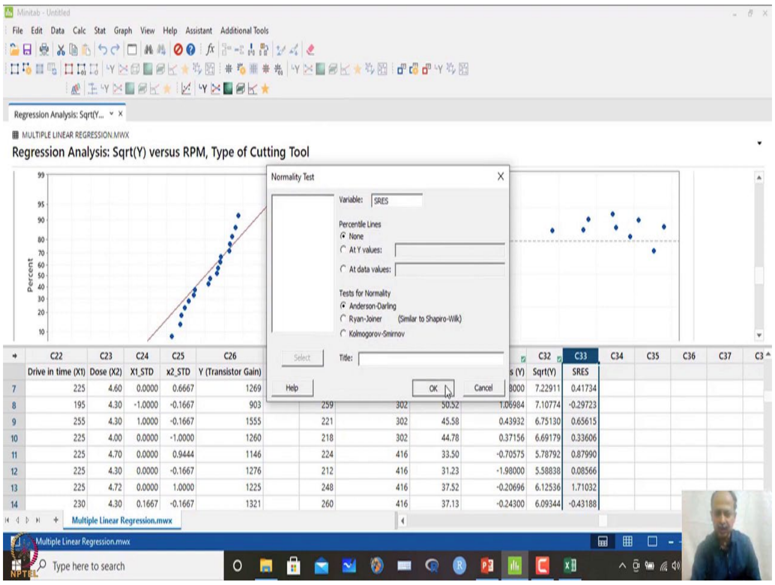
(Refer Slide Time: 19:15)



	C22	C23	C24	C25				C29	C30	C31	C32	C33	C34	C35	C36	C37	C3
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD				Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES					
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.41734					
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.29723					
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.65615					
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.33606					
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.87990					
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.08566					
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.71032					
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.43188					

Let us try to check whether the correction has happened with Box-Cox transformation or this is not adequate. So, in this case what we will do is that, we will see the residual over here and try to see ok.

(Refer Slide Time: 19:32)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Sqrt(Y...
MULTIPLE LINEAR REGRESSION.MWX
Percent
Normality Test
Select
Help
OK
Cancel
Multiple Linear Regression.mwx
Type here to search

Regression Analysis: Sqrt(Y) versus RPM, Type of Cutting Tool

Variable: SRES

Percentile Lines
● None
○ At Y values:
○ At data values:

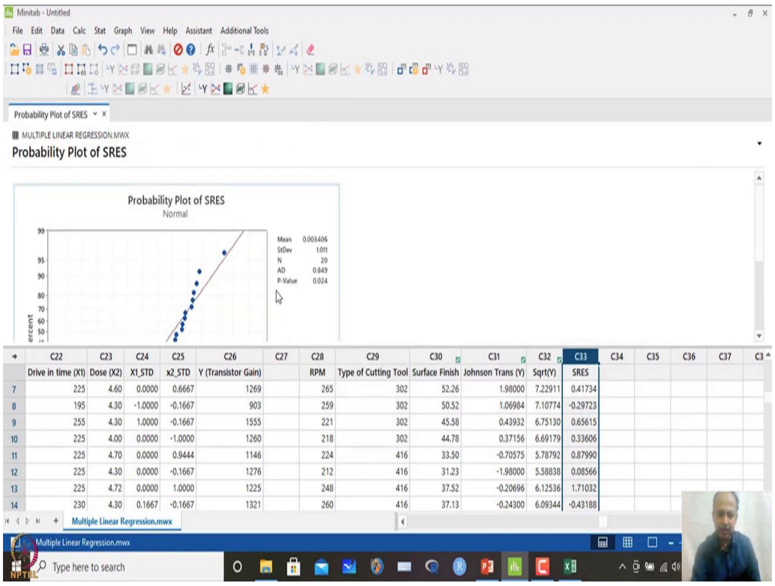
Tests for Normality
● Anderson-Darling
○ Ryan-Joiner (Similar to Shapiro-Wilk)
○ Kolmogorov-Smirnov

Title:

	C22	C23	C24	C25	C26						C32	C33	C34	C35	C36	C37	C3
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)					s (Y)	Sqrt(Y)	SRES					
7	225	4.60	0.0000	0.6667	1269					8000	7.22911	0.41734					
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.29723					
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.65615					
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.33606					
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.87990					
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.08566					
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.71032					
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.43188					

So, let me just check this is stat basic stat, normality test over here, so I will go to the last variable, that is recorded.

(Refer Slide Time: 19:36)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Probability Plot of SRES
MULTIPLE LINEAR REGRESSION.MWX
Multiple Linear Regression.mwx
Type here to search
NPTEL

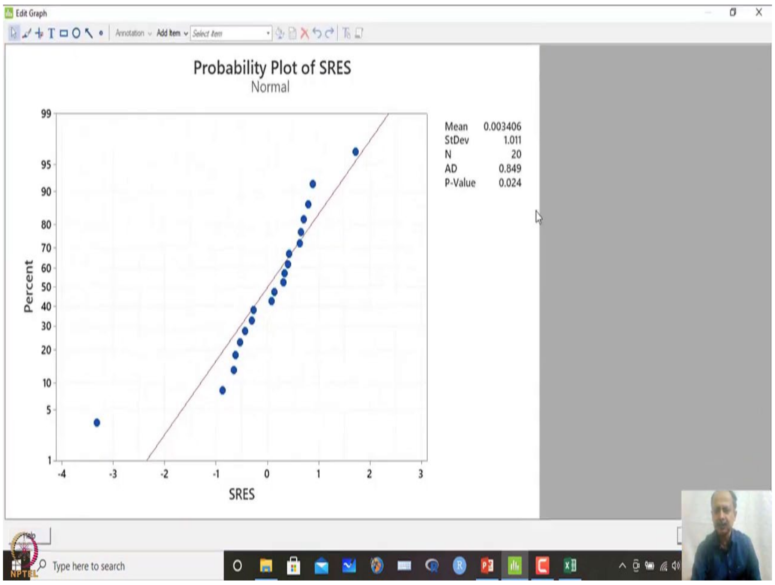
Probability Plot of SRES

Probability Plot of SRES
Normal

Mean 0.003406
StDev 1.011
N 20
AD 0.849
P-Value 0.024

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES				
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	0.41734				
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-0.29723				
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130	0.65615				
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179	0.33606				
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792	0.87990				
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838	0.08566				
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536	1.71032				
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344	-0.43188				

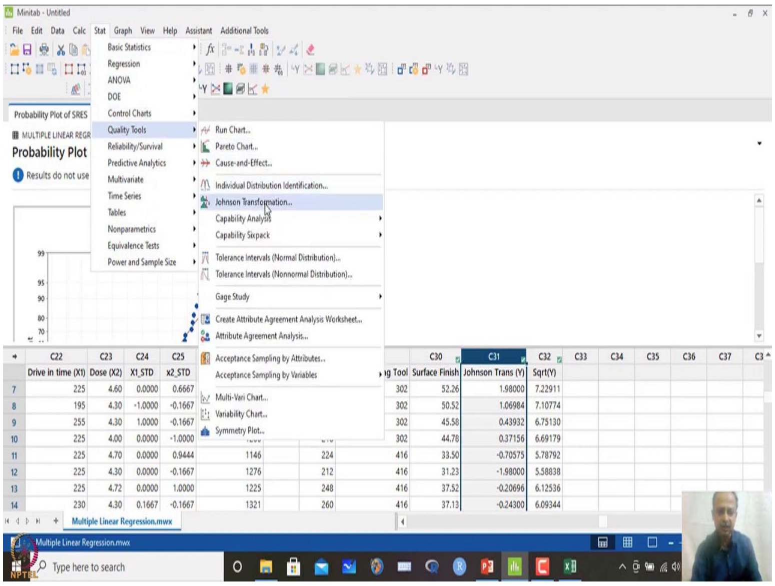
(Refer Slide Time: 19:37)



Mean 0.003406
StDev 1.011
N 20
AD 0.849
P-Value 0.024

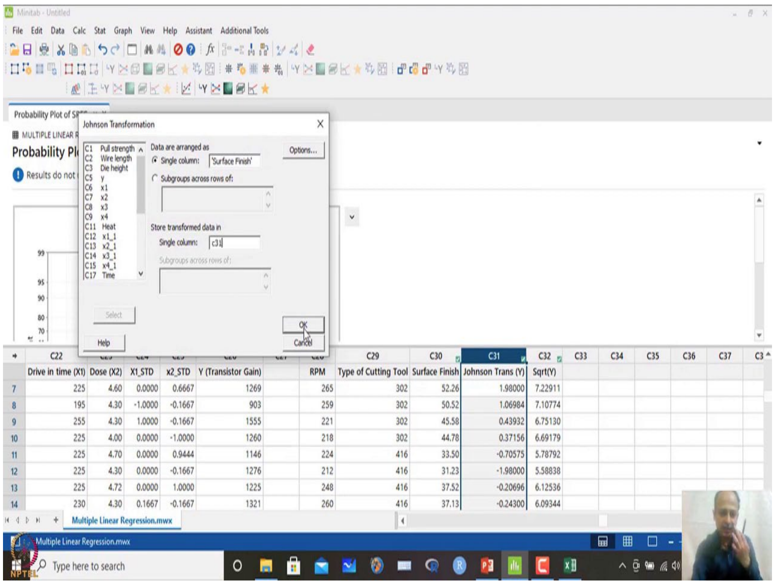
And then I see what is the value of the P-Value. So, P-Value is again less than 0.05. So again there is a problem and that the problem is not resolved, so error is not coming out to be normal, so in this case again it is not white noise. So, in this case what we have to do is that we have to so then what I have done is that I have gone for Johnson's transformation, family of transformation so Johnson's transformation.

(Refer Slide Time: 20:00)



So, what I have done is that basic stat, sorry this is quality tools, and in that case Johnson's transformation is there.

(Refer Slide Time: 20:06)



Data are arranged as
● Single column: 'Surface Finish'
○ Subgroups across rows of:

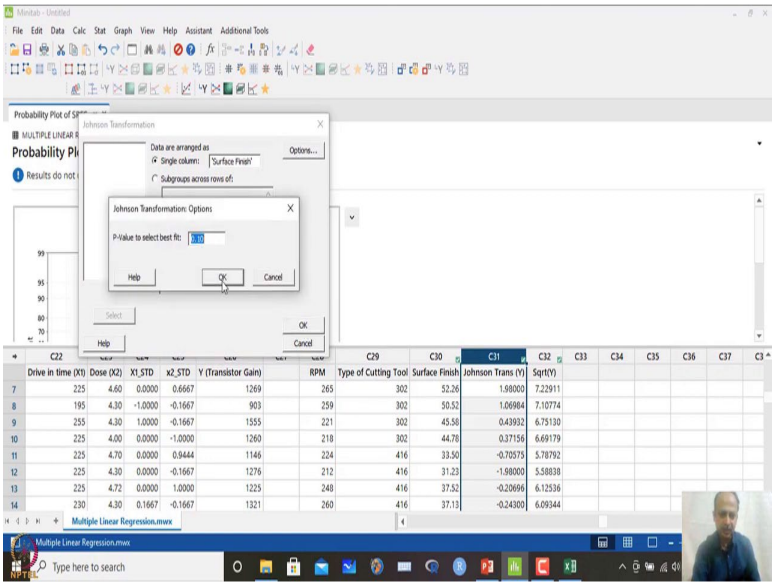
Store transformed data in
Single column: c31

	C22							C29	C30	C31	C32
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344

And in this case, I have reported that place it into single column data are arranged where the data is. So, I will say surface finish is the data and store in which column. So, I have mentioned over here as C31.

So, when you click that one C31 over here, so in this case what happens is that, if I click ok.

(Refer Slide Time: 20:25)



Minitab - Untitled

Johnson Transformation

Data are arranged as

● Single column: 'Surface Finish'

○ Subgroups across rows of:

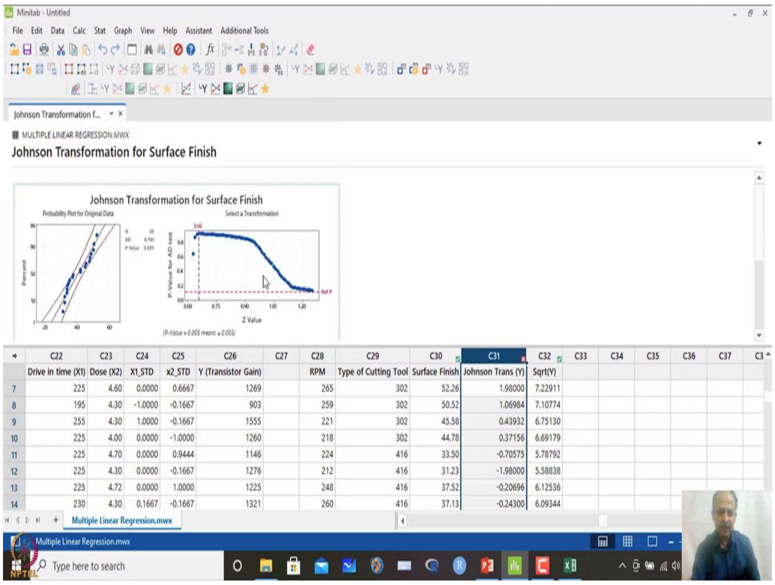
Johnson Transformation: Options

P-Value to select best fit: 0.10

	C22							C29	C30	C31	C32	C33	C34	C35	C36	C37	C3
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)						
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911						
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774						
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130						
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179						
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792						
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838						
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536						
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344						

And options what we have given is that 0.1 is the value to select the best fit, so in this case ok.

(Refer Slide Time: 20:30)



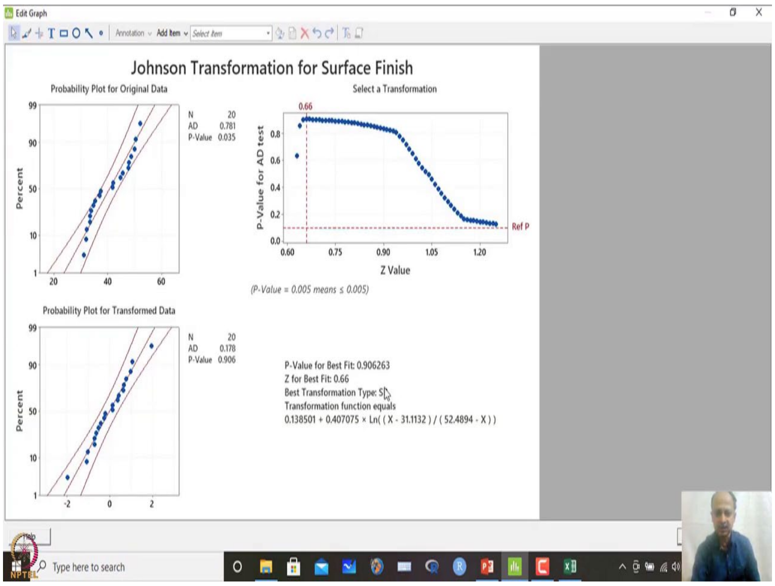
Minitab - Untitled

MULTIPLE LINEAR REGRESSION.MWX

Johnson Transformation for Surface Finish

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C3
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)						
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911						
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774						
9	255	4.30	1.0000	-0.1667	1555		221	302	45.58	0.43932	6.75130						
10	225	4.00	0.0000	-1.0000	1260		218	302	44.78	0.37156	6.69179						
11	225	4.70	0.0000	0.9444	1146		224	416	33.50	-0.70575	5.78792						
12	225	4.30	0.0000	-0.1667	1276		212	416	31.23	-1.98000	5.58838						
13	225	4.72	0.0000	1.0000	1225		248	416	37.52	-0.20696	6.12536						
14	230	4.30	0.1667	-0.1667	1321		260	416	37.13	-0.24300	6.09344						

(Refer Slide Time: 20:32)



N 20
AD 0.781
P-Value 0.035

(P-Value = 0.005 means ≤ 0.005)

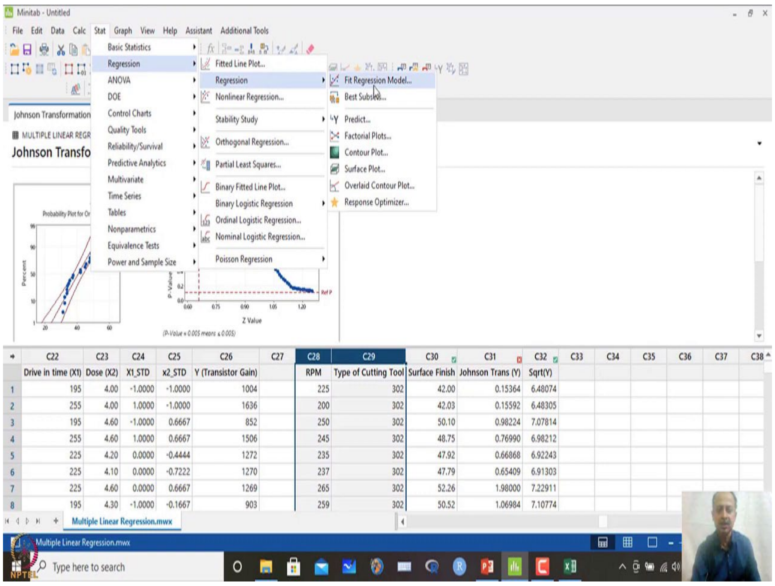
N 20
AD 0.178
P-Value 0.906

P-Value for Best Fit: 0.906263
Z for Best Fit: 0.66
Best Transformation Type: S
Transformation function equals
0.138501 + 0.407075 × Ln((X - 31.1132) / (52.4894 - X))

So, it will give you the transformation that is required. So, this is the transformed equation transform function over here that is given L n what you see the last over here and this is transforming the variable. So initially P-Value is less than point the original data is less than 0.05 and after transformation P-Value is coming out to be 0.906; that means it has done a rightful transformations.

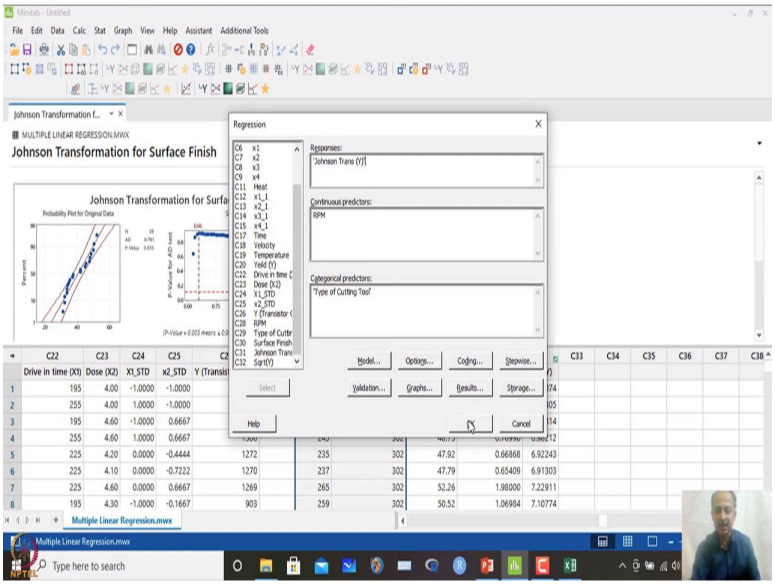
Now we have to only confirm. And, this is saved over here Johnson's transformation is saved over here in C31 column. Now, what we will do is that, with this column we will regress with RPN and type of over here.

(Refer Slide Time: 21:02)



	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)
1	195	4.00	-1.0000	-1.0000	1004		225	302	42.00	0.15364	6.48074
2	255	4.00	1.0000	-1.0000	1636		200	302	42.03	0.15592	6.48305
3	195	4.60	-1.0000	0.6667	852		250	302	50.10	0.98224	7.07814
4	255	4.60	1.0000	0.6667	1506		245	302	48.75	0.76990	6.98212
5	225	4.20	0.0000	-0.4444	1272		235	302	47.92	0.66868	6.92243
6	225	4.10	0.0000	-0.7222	1270		237	302	47.79	0.65409	6.91303
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774

(Refer Slide Time: 21:05)



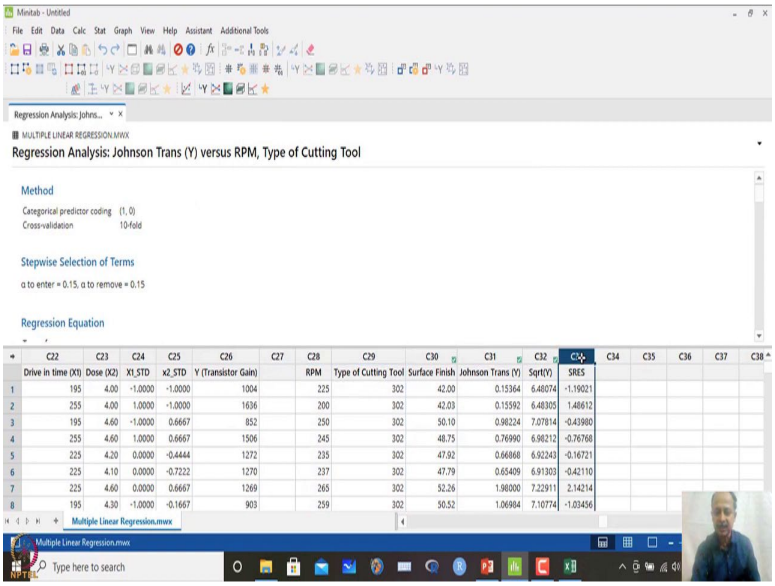
Responses: 'Johnson Trans (Y)'

Continuous predictors: RPM

Categorical predictors: 'Type of Cutting Tool'

So, what we will do is that we will go to regression, regression analysis, fit regression, instead of this square root what we will do is that we will use Johnson's transform variable and other things remain same. So, I will click ok.

(Refer Slide Time: 21:14)



Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Johns...
Multiple Linear Regression.mwx
Type here to search

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Johnson Trans (Y) versus RPM, Type of Cutting Tool

Method

Categorical predictor coding (1, 0)
Cross-validation 10-fold

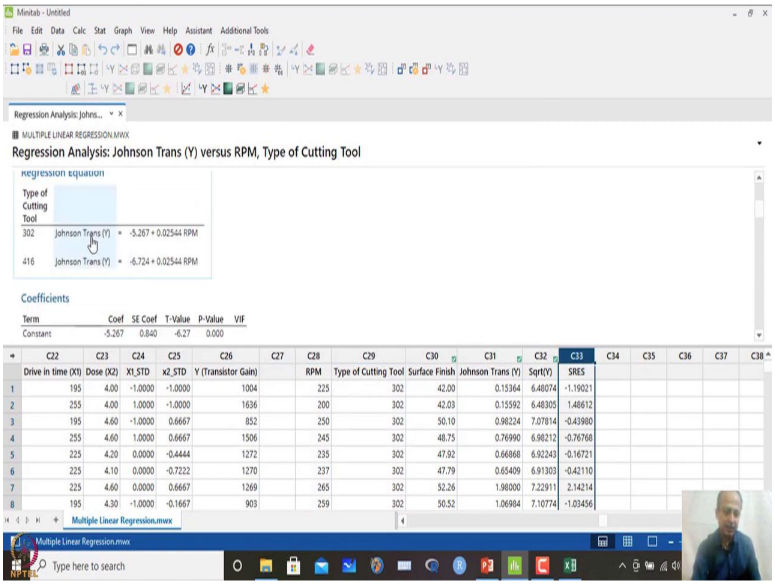
Stepwise Selection of Terms

α to enter = 0.15, α to remove = 0.15

Regression Equation

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C38
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES					
1	195	4.00	-1.0000	-1.0000	1004		225	302	42.00	0.15364	6.48074	-1.19021					
2	255	4.00	1.0000	-1.0000	1636		200	302	42.03	0.15592	6.48305	1.48612					
3	195	4.60	-1.0000	0.6667	852		250	302	50.10	0.98224	7.07814	-0.43980					
4	255	4.60	1.0000	0.6667	1506		245	302	48.75	0.76990	6.98212	-0.76768					
5	225	4.20	0.0000	-0.4444	1272		235	302	47.92	0.66868	6.92243	-0.16721					
6	225	4.10	0.0000	-0.7222	1270		237	302	47.79	0.65409	6.91303	-0.42110					
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	2.14214					
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-1.03456					

(Refer Slide Time: 21:17)



Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Johns...
Multiple Linear Regression.mwx
Type here to search

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Johnson Trans (Y) versus RPM, Type of Cutting Tool

Regression Equation

Type of Cutting Tool			
302	Johnson Trans (Y)	=	-5.267 + 0.02544 RPM
416	Johnson Trans (Y)	=	-6.724 + 0.02544 RPM

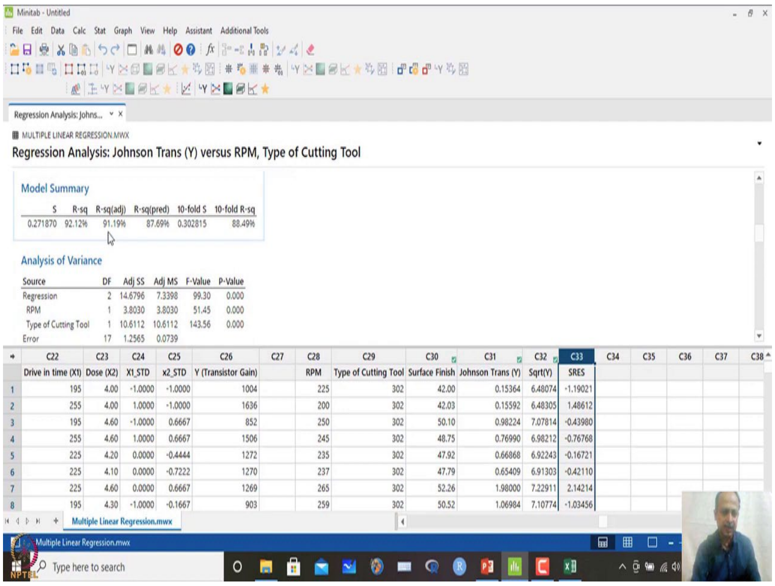
Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-5.267	0.840	-6.27	0.000	

	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C38
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES					
1	195	4.00	-1.0000	-1.0000	1004		225	302	42.00	0.15364	6.48074	-1.19021					
2	255	4.00	1.0000	-1.0000	1636		200	302	42.03	0.15592	6.48305	1.48612					
3	195	4.60	-1.0000	0.6667	852		250	302	50.10	0.98224	7.07814	-0.43980					
4	255	4.60	1.0000	0.6667	1506		245	302	48.75	0.76990	6.98212	-0.76768					
5	225	4.20	0.0000	-0.4444	1272		235	302	47.92	0.66868	6.92243	-0.16721					
6	225	4.10	0.0000	-0.7222	1270		237	302	47.79	0.65409	6.91303	-0.42110					
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	2.14214					
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-1.03456					

And we will get a residual over here and we will get the equation. So, after Johnston's transformed variable with RPN, this is given for 302 and 416 types of tools like that.

(Refer Slide Time: 21:24)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Johns...
Multiple Linear Regression.mwx
Type here to search
NPTEL

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Johnson Trans (Y) versus RPM, Type of Cutting Tool

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
0.271870	92.12%	91.19%	87.69%	0.302815	88.49%

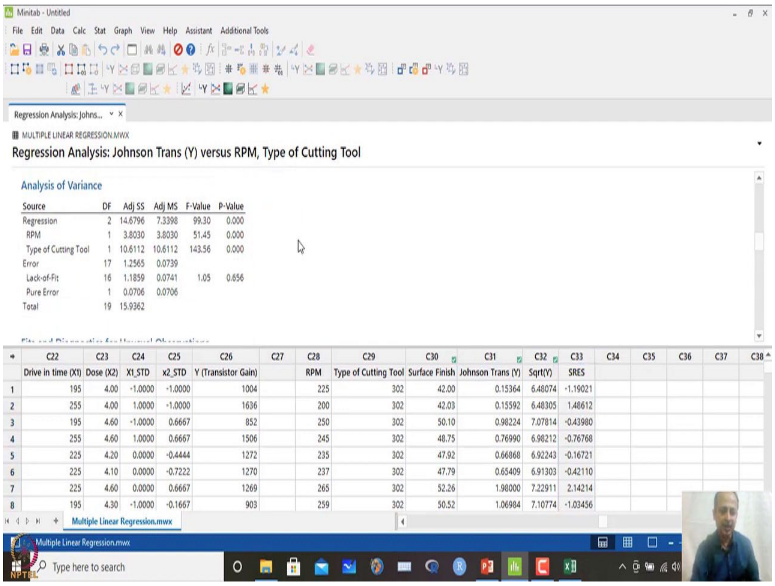
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	14.6796	7.3398	99.30	0.000
RPM	1	3.8030	3.8030	51.45	0.000
Type of Cutting Tool	1	10.6112	10.6112	143.56	0.000
Error	17	1.2565	0.0739		

↓	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C38
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES					
1	195	4.00	-1.0000	-1.0000	1004		225	302	42.00	0.15364	6.48074	-1.19021					
2	255	4.00	1.0000	-1.0000	1636		200	302	42.03	0.15592	6.48305	1.48612					
3	195	4.60	-1.0000	0.6667	852		250	302	50.10	0.98224	7.07814	-0.43980					
4	255	4.60	1.0000	0.6667	1506		245	302	48.75	0.76990	6.98212	-0.76768					
5	225	4.20	0.0000	-0.4444	1272		235	302	47.92	0.66868	6.92243	-0.16721					
6	225	4.10	0.0000	-0.7222	1270		237	302	47.79	0.65409	6.91303	-0.42110					
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	2.14214					
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-1.03456					

And, what we observe is that 91 percent R square adjusted 88, so very close.

(Refer Slide Time: 21:28)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Johns...
Multiple Linear Regression.mwx
Type here to search
NPTEL

MULTIPLE LINEAR REGRESSION.MWX

Regression Analysis: Johnson Trans (Y) versus RPM, Type of Cutting Tool

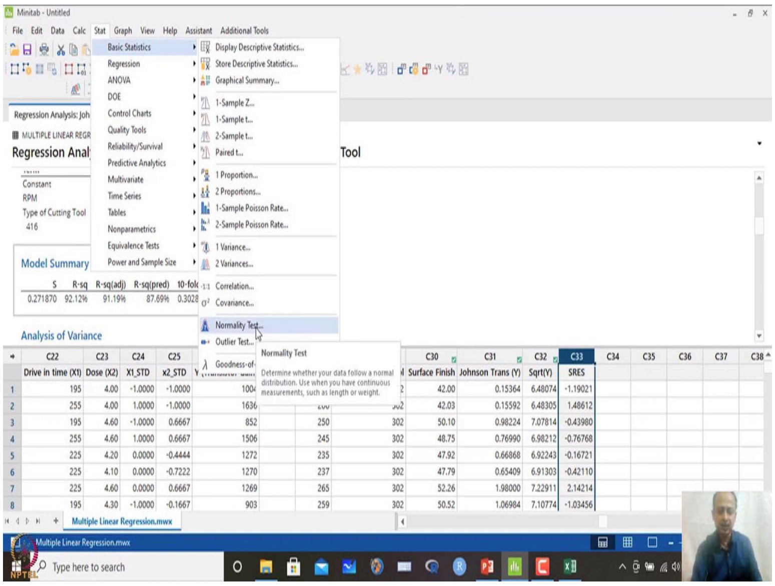
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	14.6796	7.3398	99.30	0.000
RPM	1	3.8030	3.8030	51.45	0.000
Type of Cutting Tool	1	10.6112	10.6112	143.56	0.000
Error	17	1.2565	0.0739		
Lack-of-Fit	16	1.1859	0.0741	1.05	0.656
Pure Error	1	0.0706	0.0706		
Total	19	15.9362			

↓	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C38
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES					
1	195	4.00	-1.0000	-1.0000	1004		225	302	42.00	0.15364	6.48074	-1.19021					
2	255	4.00	1.0000	-1.0000	1636		200	302	42.03	0.15592	6.48305	1.48612					
3	195	4.60	-1.0000	0.6667	852		250	302	50.10	0.98224	7.07814	-0.43980					
4	255	4.60	1.0000	0.6667	1506		245	302	48.75	0.76990	6.98212	-0.76768					
5	225	4.20	0.0000	-0.4444	1272		235	302	47.92	0.66868	6.92243	-0.16721					
6	225	4.10	0.0000	-0.7222	1270		237	302	47.79	0.65409	6.91303	-0.42110					
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	2.14214					
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-1.03456					

So, this is quite acceptable and in this case there is no lack of fit and both the variables are significant that is observed over here. Categorical as well as RPM over here.

(Refer Slide Time: 21:39)



Regression Anal... Tool

Constant
RPM
Type of Cutting Tool
416

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fol
0.271870	92.12%	91.19%	87.69%	0.3028

Analysis of Variance

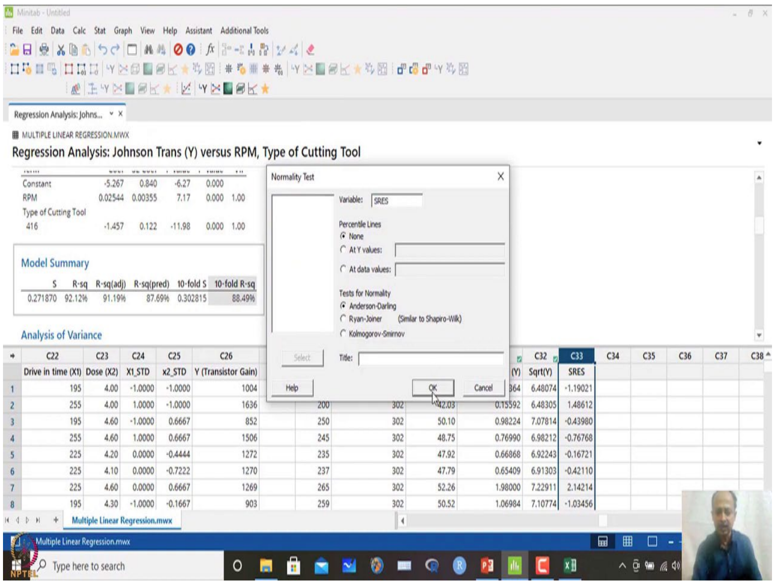
Normality Test

Determine whether your data follow a normal distribution. Use when you have continuous measurements, such as length or weight.

	C22	C23	C24	C25					C30	C31	C32	C33
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y			l	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES
1	195	4.00	-1.0000	-1.0000	1004			2	42.00	0.15364	6.48074	-1.19021
2	255	4.00	1.0000	-1.0000	1636		200	302	42.03	0.15592	6.48305	1.48612
3	195	4.60	-1.0000	0.6667	852		250	302	50.10	0.98224	7.07814	-0.43980
4	255	4.60	1.0000	0.6667	1506		245	302	48.75	0.76990	6.98212	-0.76768
5	225	4.20	0.0000	-0.4444	1272		235	302	47.92	0.66868	6.92243	-0.16721
6	225	4.10	0.0000	-0.7222	1270		237	302	47.79	0.65409	6.91303	-0.42110
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	2.14214
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-1.03456

So, in this case residual we can check, whether the correction has happens some positive things has happened over here.

(Refer Slide Time: 21:42)


Minitab - Untitled
File Edit Data Calc Stat Graph View Help Assistant Additional Tools
Regression Analysis: Johns...
MULTIPLE LINEAR REGRESSION.MWX
Select
Help
OK
Cancel
Multiple Linear Regression.mwx
Type here to search
NPTEL

Regression Analysis: Johnson Trans (Y) versus RPM, Type of Cutting Tool

Constant	-5.267	0.840	-6.27	0.000	
RPM	0.02544	0.00355	7.17	0.000	1.00
Type of Cutting Tool					
416	-1.457	0.122	-11.98	0.000	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)	10-fold S	10-fold R-sq
0.271870	92.12%	91.19%	87.69%	0.302815	88.49%

Analysis of Variance

Normality Test

Variable: SRES

Percentile Lines
● None
○ At Y values:
○ At data values:

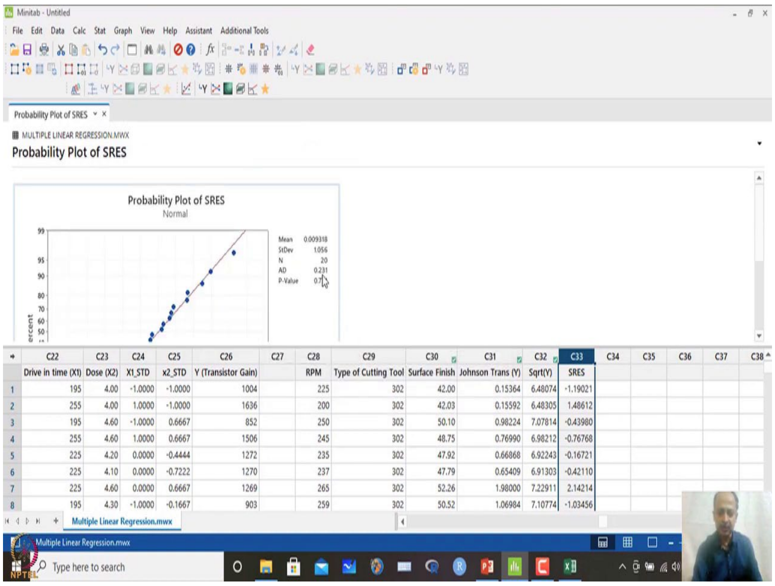
Tests for Normality
● Anderson-Darling
○ Ryan-Joiner (Similar to Shapiro-Wilk)
○ Kolmogorov-Smirnov

Title:

	C22	C23	C24	C25	C26
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)
1	195	4.00	-1.0000	-1.0000	1004
2	255	4.00	1.0000	-1.0000	1636
3	195	4.60	-1.0000	0.6667	852
4	255	4.60	1.0000	0.6667	1506
5	225	4.20	0.0000	-0.4444	1272
6	225	4.10	0.0000	-0.7222	1270
7	225	4.60	0.0000	0.6667	1269
8	195	4.30	-1.0000	-0.1667	903

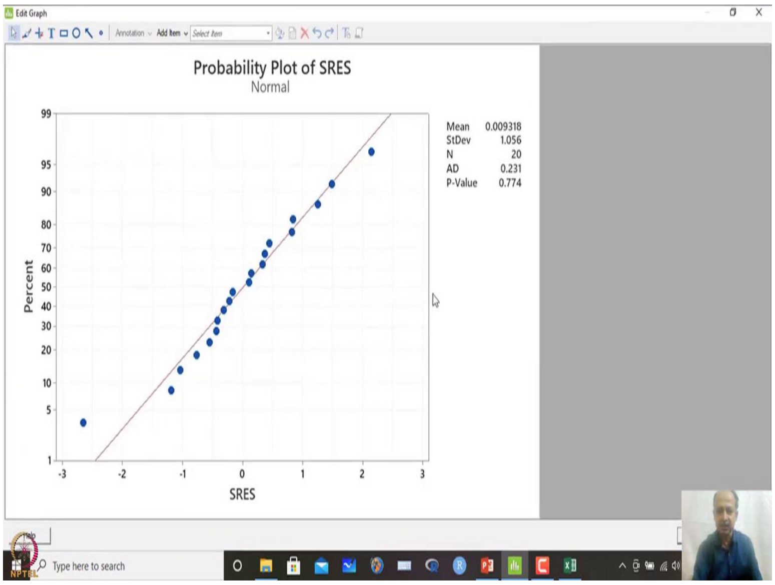
So, I am going to normality test over here, and what I do is that I go to the last residual values and I click ok.

(Refer Slide Time: 21:48)



	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C38
	Drive in time (X1)	Dose (X2)	X1_STD	x2_STD	Y (Transistor Gain)		RPM	Type of Cutting Tool	Surface Finish	Johnson Trans (Y)	Sqrt(Y)	SRES					
1	195	4.00	-1.0000	-1.0000	1004		225	302	42.00	0.15364	6.48074	-1.19021					
2	255	4.00	1.0000	-1.0000	1636		200	302	42.03	0.15592	6.48305	1.48612					
3	195	4.60	-1.0000	0.6667	852		250	302	50.10	0.98224	7.07814	-0.43980					
4	255	4.60	1.0000	0.6667	1506		245	302	48.75	0.76990	6.98212	-0.76768					
5	225	4.20	0.0000	-0.4444	1272		235	302	47.92	0.66868	6.92243	-0.16721					
6	225	4.10	0.0000	-0.7222	1270		237	302	47.79	0.65409	6.91303	-0.42110					
7	225	4.60	0.0000	0.6667	1269		265	302	52.26	1.98000	7.22911	2.14214					
8	195	4.30	-1.0000	-0.1667	903		259	302	50.52	1.06984	7.10774	-1.03456					

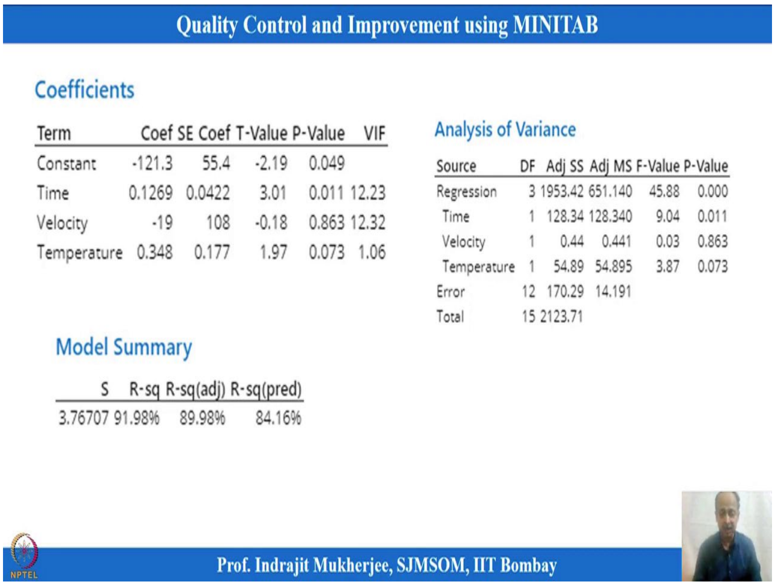
(Refer Slide Time: 21:50)



And, what I observe is that, when I have done this transformation suddenly this problem has gone so; that means the normality problem of the error residual has gone. So, when I am using Johnson's transformation this has happened. When I have used Johnson's transformation this is giving results like that.

So, which one Box-Cox or Johnson you have to try out and figure out that whichever gives you error as white noise, so that has to be adopted like that. So, this is one of the example, when categorical variable is also considered in the model and we are able to address that one.

(Refer Slide Time: 22:22)



Quality Control and Improvement using MINITAB

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-121.3	55.4	-2.19	0.049	
Time	0.1269	0.0422	3.01	0.011	12.23
Velocity	-19	108	-0.18	0.863	12.32
Temperature	0.348	0.177	1.97	0.073	1.06

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	1953.42	651.140	45.88	0.000
Time	1	128.34	128.340	9.04	0.011
Velocity	1	0.44	0.441	0.03	0.863
Temperature	1	54.89	54.895	3.87	0.073
Error	12	170.29	14.191		
Total	15	2123.71			

Model Summary

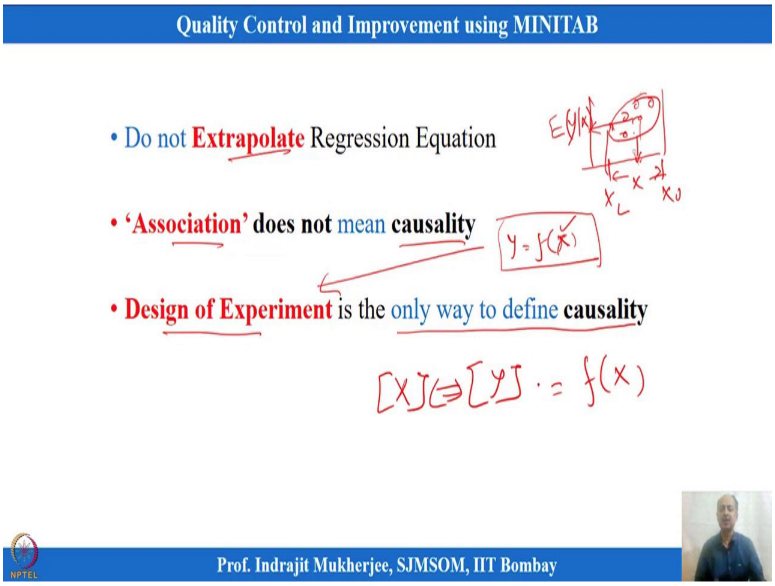
S	R-sq	R-sq(adj)	R-sq(pred)
3.76707	91.98%	89.98%	84.16%

Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

So, categorical variable and how to address in case there is a non normal situation in multiple regression, how it is to be addressed that we can see ok.

And, how to select the variables in case we are in dilemma which variable. So we have talked about stepwise regression. So, this regression is an important aspects which can suggest that which are the variables or potential variables can be considered in experimentation.

(Refer Slide Time: 22:46)



Quality Control and Improvement using MINITAB

- Do not Extrapolate Regression Equation
- 'Association' does not mean causality
- Design of Experiment is the only way to define causality

Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

So, but we have to understand that regression does not say we cannot extrapolate regression equation. So, whatever regression that experiment is happening; that means, this is y and this is the range of x domain within which we have got the information; that means, data are collected over here and these are the observations over here.

So, I can only restrict to this region and predict. So, for a given value of x, at a given value of x what is the predicted value of expected value of y. So, expectation of y for a given x can be only calculated within the domain or within the range of x over here. So, x we can say upper bound and x lower bound within this, so no extrapolation is generally preferred in regression equation ok.

And, association does not mean, that means, y is related with x function of x, does not mean that it is a real variable that impacts y. So, certain scenarios can be, there is a relationship between two variables which are not physically any way connected but there can be high correlation.

So, many examples can be cited like that. So, it does not mean that there is a causality. So regression does not prove causality. So, design of experiment is the only way to define causalities like that. So, to understand causalities between the variables we have to intentionally induce variation by changing, the factors changing the factors or variables that we are interested into and try to understand what is the functional relationship.

So, the best appropriate way to develop the functional relationship is by design of experiments. So, there is no other alternative. So, regression with historical data does not give you proper association or causality cannot be proved based on that ok.

So, then what is the option, then what is the option for quality improvements. Then the options we have identified, few variables based on regressions. So there are variables X, X matrix over here, and I have collected also y variable information. So, how do I connect these two and this connection and develop the function. So, I need a function over here, that will explain the variability of y over here. So, what is required is that systematic way of variation is required, systematic way.

(Refer Slide Time: 24:40)

Quality Control and Improvement using MINITAB

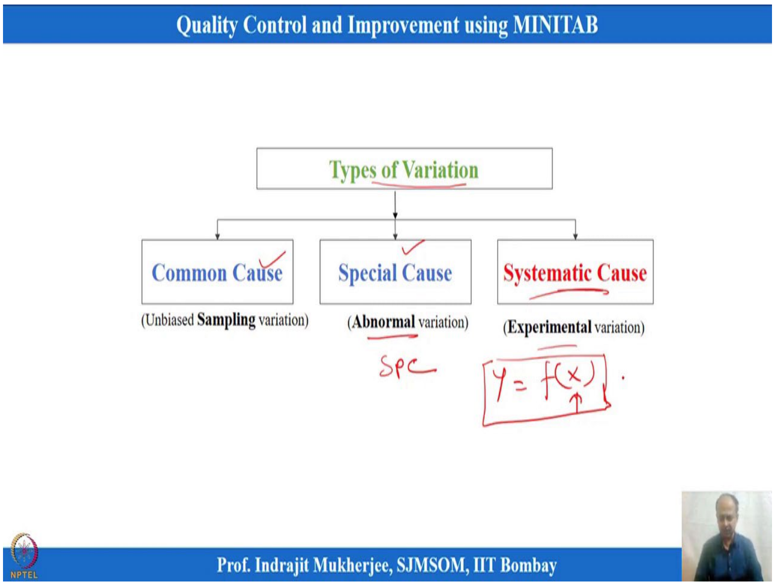
Improvement Phase:

Design of Experiments (DOE)

Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

And then what comes is statistical experimentation which is the improvement phase. Until and unless I have a right function I cannot improve. I cannot reach to the global optimal point like that or setting points like that ok. So, design of experiment is the stepping stone for improvement phase like that.

(Refer Slide Time: 25:01)



Quality Control and Improvement using MINITAB

Types of Variation

Common Cause

(Unbiased **Sampling** variation)

Special Cause

(**Abnormal** variation)

SPC

Systematic Cause

(**Experimental** variation)

$\boxed{Y = f(x)}$

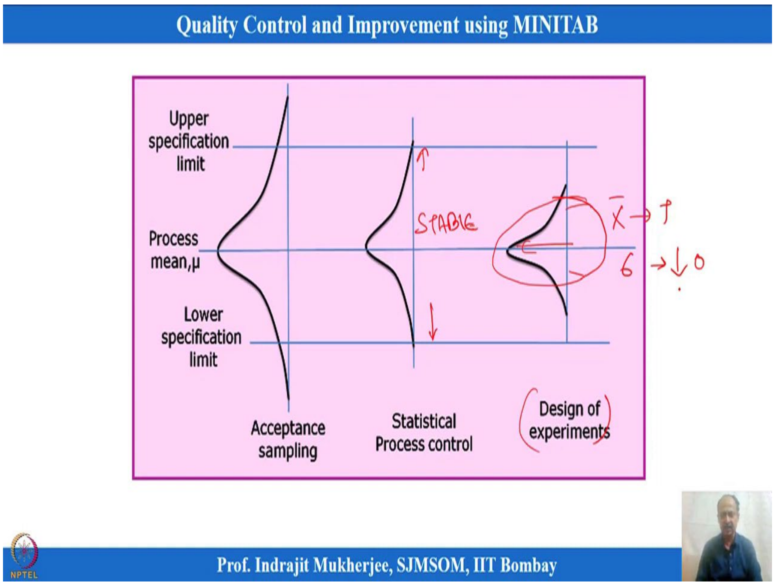
Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

So, basic things that our understanding over here is that, there are types of variability types of variation we have already spoken about it and some basic information I am providing over here which we can think of as recap. So, there are common cause variability in the process, there are special cause abnormal variability which are easy to detect based on SPCprocess control chart.

And, then what we do is that, to understand the functional relationship between y and x, we have to do it a systematic change in we have to induce variability we have to induce variability by changing the factor, by changing the factor x over here. And, if we can do that systematically what happens is that we can generate a function and the function can be optimized.

That function, we know this is the real equation between the in the process. This is the empirical relationship that exists and this is based on systematic theoretically strong models that we have developed over here. Then we can use optimization technique just to reach to the global optimal point like that. So, we are interested into and for that design of experiment is most preferred like that ok.

(Refer Slide Time: 26:05)



Quality Control and Improvement using MINITAB

Upper specification limit
Process mean, μ
Lower specification limit
Acceptance sampling
Statistical Process control
Design of experiments
STABLE
$\bar{X} \to T$
$\sigma \to \downarrow 0$

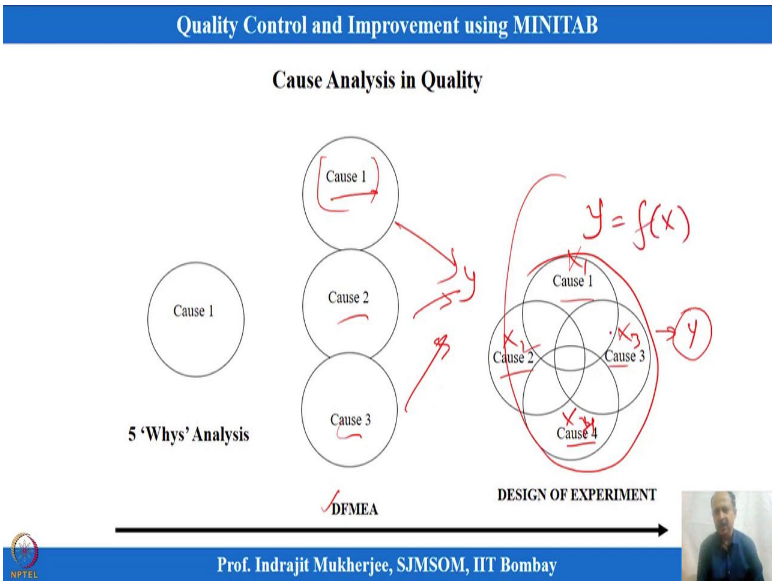
Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

So, design of experiment what we are seeing over here is basically reducing the common cause variability. So, this can be applied when the process is in statistical control. So, when the process is in stable process, so whenever it is stable process we say and it is in statistical control then only we can go for design of experiment, because we want to reduce the common cause variability further, because more and more I reduce the variability, so accuracy and precision both improves over here.

So, we are interested in being and variance both aspects like that, what we have seen in capability analysis also. So, for that what is required reduced for the variability requires design of experiments. So, for that we need to, so that helps, statistical design of experiment helps us to bring the mean to target value and also the reduced variability near to 0.

So, we want minimum variability over here. So, various design of experiment techniques are proposed and huge amount of resource is available for design of experiments. So you can see. But, preliminary book that you can see is Montgomery's book which is which can be a good resource of learning initial steps like that, but there are other books also you can see, so there are many other books which you can see like that. Amitava Mitra is one of the books where design of experiment is detailed ok.

(Refer Slide Time: 27:22)



Quality Control and Improvement using MINITAB

Cause Analysis in Quality

Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

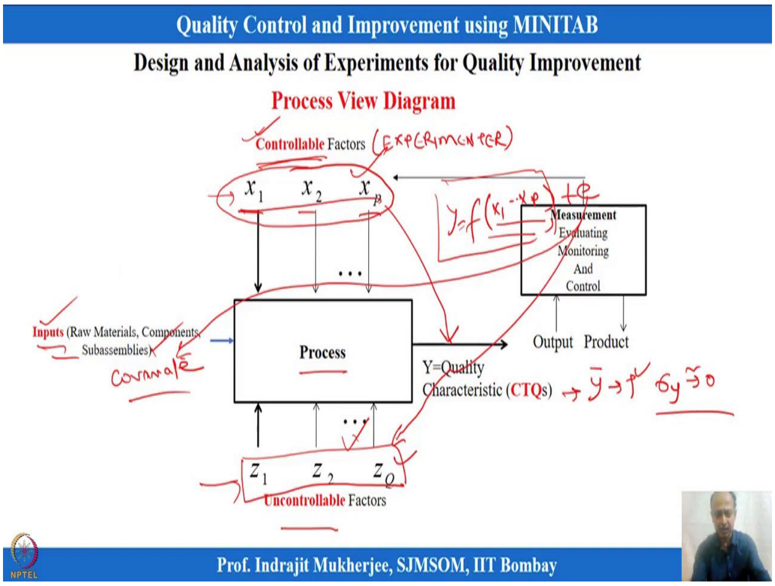
And, so this is one of the techniques. And also, I want to recap that one of the concepts that if you have variability, that means, causes and this is impacting my y and causes are interlinked also somewhat they are interlinked with each other and they together can impact the y like this. So this is the scenario like this.

This is one of the variable X1 this is one of the variable X2, X3 and this is X4 over here, and they are interrelated with their in this what you are seeing and they are not completely independent over here what is observed over here.

So, if they are completely independent in that case, we can understand that what is the level of X that will deliver the optimal y over here. So, this is possible, but this is not possible when we are having a complex relationship between multiple X's over here and together they can impact the y, that can be scenario.

So, it can be scenarios which is not possible to identify. So, by simple other analysis like, design failure mode and effect analysis or process failure mode and effect analysis that is not possible by that ok. So, complex relationship and developing the function, when we are developing the functions over here only design of experiments can help.

(Refer Slide Time: 28:33)



Quality Control and Improvement using MINITAB

Design and Analysis of Experiments for Quality Improvement

Process View Diagram

Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

So, design of experiments can help. And, before we go into details of design of experiment this process p information is again I am recapping over here. There will be certain control variables which is in the hand of experimenters. So, experimenter will change this control variable basically.

So when you go to a process what you observe is that, you will see the operator is changing some of the variables, not all variables, some of the key variables which is possible to control. So, these are the variables X1 to Xp which is known as controllable variables.

Design of experiment is all about controllable variable, most of the time we deal with controllable variables. There will be some inputs and I also talked about covariates which can influence the outcomes over here, and that can also be considered in the model when we are doing experimentation we can also consider the covariates as a variable which will vary.

We cannot control that one, but influence of that has to be considered when we are developing the mathematical models or something like that. Or we can deal with that, so covariates also we can see ok. And, there will be some uncontrollable factor; that means, it is uneconomical to control.

And sometimes we do not have any information enough information about the some of the variables, because you cannot build a perfect model like that y is a function of all Xp variables X1 to Xps like that. There will be some amount of error in the model because of this noise variables over here, because of this noise variable. So, we do not get a perfect function.

Because all the variables that is impacting the process is some of the variables are unknown to us. So, in that case, these unknown as noise variables or and the influence of that is very less because we have identified most of this. So, these are the variables and some of them are controllable, some of them are covariates over here, so we have knowledge about that.

So, and there will be some noise variable which we do not have any control or we do not have enough information about these variables like that ok. So, we want to; we want to get a setting so in design of experiment, what we are trying to do is that this is control variable, so it is in our hand to control. So, we want to get a combination of this combination of this that will optimize my Y CTQ.

And y will be the average value of y should be close to the target value, and the variability of y that is that we will generate is very close to 0 values like that. That means there is no as such variability we want to develop. So, our objective of design of experiment is twofold objective; one is which are the variables impacting y, and then if these are the variables which is significantly impacting y, what can be the combination or setting conditions of this Xp variables that is impacting over here, which I can control.

So that I get the y exactly to the target value defined by the designer and also the variability near the target value is near to 0 like that. So, that is the overall objective of design of experiments. So, and another objective is to develop the functional relationship so that we can optimize like that.

So, whenever I have a function, I can optimize. So, I need a function. So, function to need a function I need what are the Xp variables which is controllable which is significantly contributing to the variability in y; that means, this is and there will be some error that can be because of this noise variables which are not considered and which is difficult to control and uneconomical to control.

This can also be impact with these errors may also be impacted by these covariates which are influencing the , but we do not have any control on that. So, I can only control this Xp variables, I do not have any control on the inputs over here and the uncontrollable variables over here. So, most of the time we try to get the best combination of X1 to Xp that in presence of this noise and in presence of covariates or inputs variability that we are experiencing in the process like that ok.

So this is all about what we will try to understand in design of experiments and how to do the experiments that we will try to understand over here. There are different ways of doing experimentation, different designs are available, and we will see some very few of them. So that that is our objective in experimentation, so that we will cover in our next session ok.

Thank you for listening.