hatQuality Control and Improvement with MINITAB Prof. Indrajit Mukherjee Shailesh J. Mehta School of Management Indian Institute of Technology, Bombay

Lecture - 05 Data Visualization for Quality Control and Improvement

Hello everyone, welcome to session-5 of Quality Control and Improvement using MINITAB. So, I am Prof. Indrajit Mukherjee from Shailesh J Mehta School of Management IIT, Bombay. So, we are discussing on tools that are used for visualization, ok. So, in this context, we are just discussing about histogram where it can be used and what scenarios. Histograms are generally used whenever I have a continuous data. So, for that generally histograms are preferred.

(Refer Slide Time: 00:47)



So, there are many key quality tools which we will highlight over here. So, one of the tool is histogram, another one is box plot that is the two important tools that is used in quality so visualization of the data, so and that can be used for some interpretation out of the data. So, that we will see how it is possible.

So, here you see I will talk about box plot over here, I will talk about histogram, and then we may go to scatter plot over here. So, histogram is just representation of the data. So, how the data is spread over here. So, what is the spread of the data that can be seen? So, there can be a spread from this range to this range. And where is the peak of the data concentration of the data, cg of the data basically? CG of the data may be somewhere over here. So, but the spread of the data is showing that there are two clusters over here. So, one is this cluster over here, and one cluster is over here. So, that means, a mix of data is basically plotted over here ok. So, it will give you some sense about the data distribution. And what is the spread of the data.

Then we can see the descriptive statistics in MINITAB which can be used to see what is the mean value, what is the other measures or statistic that can be that we can also figure out, that means, what is the mean, what is the standard deviation, and the what is the range, what is the maximum-minimum, interquartile range, all things are possible over here.

> Quality Control and Improvement using MINITAB **CTQ** : Thickness 30 20 54 47 48 49 50 51 52 53 Free 10 0 405 415 425 435 445 455 465 475 485 495 Metal thickness Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

(Refer Slide Time: 02:13)

So, let us take an example of CTQ, and let us try to demonstrate that one. Let us try to see that assume that CTQ is thickness. So, we will illustrate in MINITAB how these plots are possible. So, last time also we are discussing on this. So, I will open some data sets which is available with us, and that will help us to illustrate in MINITAB.

(Refer Slide Time: 02:31)

MINITAB Data File	ie View						- a ×
to Quick Copy Paste P R Copposed	ut popy path site shortout More Copy to ' to ' Organize	New item • New folder New	Properties Open • History Open	Select all Select none Invert selection Select			
🔿 👻 🛧 🧾 « Pen driv	e Indrajit Oct 2020 > NPTEL Feb 2020 > Final Vi	deo and Other Details Dece	mber 2020 > Weekwise	Final Slides for Lecture	> Week 1 > MINITAB Data File	v õ	, Search MINITAB Data File
Pictures 💉 ^	Name	Date modified	Type	Size			
CV 2015-20	1 Pareto Defect himer	20.12.2020 10.47	Microsoft Excel 07	26 KR			
Final Video and Othe	R 2. Cause and Effect	24-12-2020 08:07	Microsoft Excel 97	26 KB			
Presynopsis Report	1. Histogram Box Plot Quality Exam Marks	31-12-2020 20:24	Microsoft Excel 97	29 KB			
Week 1	3. Histogram Motor Fuer Distana Rating	24.12.2030.08.24	Microsoft Excel W	10 KB			
	W 4. Box Plot Resturant Ex Size: 29.0 KB	el 97-2003 Worksheet	Microsoft Excel 97	29 KB			
OneDrive	S. Pie Chart Bar Chart & Date modified: 31-	12-2020 20:24	Microsoft Excel 97	22 KB			
This PC	0 5. Pie Chart	10-12-2020 10:03	Microsoft Excel W	9 KB			
3D Objects	6. Scatter plot Price and Size ofFlat	24-12-2020 08:19	Microsoft Excel 97	23 KB			
Desiton	17. Time Series oil and Gasoline	30-12-2020 11:07	Microsoft Excel 97	30 KB			
Documents	0 7. Time Series Plot	30-12-2020 11:04	Microsoft Excel W	9 KB			
Develoade							
Music							
Pintere .							
Pictures							
Videos							
Local Disk (C:)							
New Volume (D:)							
New Volume (E:)							
INDRAJIT (G:)							
INDRAJIT (G:)							
							-
P Network							
tant							

(Refer Slide Time: 02:39)

B 5- ∂-€+∓		3. Histogram Box Plo	et Quality Exam	Marks [Compatil	aiity Mode]	Excel (Produ	act Activation	Failed)					-	a >
File Home Insert Page Lay		Review View	ACROBAT	V Tell me what									Sign in	A Share
Calibri Paste of Format Painter Clipboard rs	$ \begin{array}{c c} \bullet & \Pi & \bullet & A^* & A^* \\ \hline \bullet & \bullet & \Delta & \bullet & \bullet \\ \hline \bullet & \bullet & \Delta & \bullet & \bullet \\ \hline Font & & 5 \end{array} $	Algoment	irap Text lerge & Center	General • 😨 • % 3	- % 43 er 5	Conditiona Formatting	Format as • Table • Styles	Cell Ryles •	Insert Delete	Format	∑ AutoSum ↓ Fill + ℓ Clear +	Sort & Filter *	P Find & Select *	
G11 ▼ : × ✓ fx														
A	В	c	D	E F	G	н	1	J	K	L	M	N	0	р
1 Marks in Quality Cou	irse	Layer Thickness												
2 65		3.048												
3 71		3.054												
4 66		3.079												
5 79		3.047												
6 65		3.067												
7 82		3.065												
8 80		3.083												
9 86		3.076												
67		3.043												
64		3.091												
2 62		3.094												
13 74		3.049												
67		3.070												
5 72		3.070												
68		3.056												
17 81		3.089												
18 53		3.069												
19 70		3.072												
20 76		3.027												
21 73		3.044												
22 73		3.055										-		
23 85		3.092												
Stat (+)						1	4			1			-	5
14													1	40
P Type here to search		0 🔚 🔒		M 🔞 1		8	C 1	1	X II		∧ ĝ ₩	a 41		

So, I am opening in excel. So, some other data sets, and then we will place it in MINITAB and see. So, there is a layer thickness which is an important characteristic which we want to plot over here which is a CTQ basically. There are marks in quality also we can illustrate. First let us try to see that in a manufacturing process some we are producing some thicknesses or coating over a given surface.

So, in that case, layer thickness is important. And target value we can assume as 3. So, let us try to see by plotting this one, whether some more information we can get out of

histogram. So, I have come copied this data set over here. So, I told you that when you install MINITAB, what will happen is that an icon will be coming to your desktop.

A Mir	itab - Untitled																		- 8	>
File	Edit Data Calc Stat G	raph View	Help /	Assistant	Additional 1	ools														
1	Vew	. 🗆 🕊	40	? fx	32 H	12 2/4	1. 2													
2	Open Ctrl+O	111	Y ≥6		≤★與	11 米 考	· · · · · · · · · · · · · · · · · · ·	1 'Y 🖄	Bek	*呜囵										
8	iave Project Ctrl+S		12	4 4Y 🖂	Bek	*														
1	iave Project As																			
1	iave Worksheet As																			
	Query L Save As																			
	Run an Sava the contents of th	e active oan									-									
8	Print Ctrl+P	re acure pari					M	ini	it-n	h										
200							I'I.		ιLC	U.										
	Pescription	·																		
	ventea oocumenta																			
10	Options								Open C	rl+0										
	ila Security							New F	roject C	rl+Shift+N										
	in seconda							New Wori	sheet C	ri+N										
~	oot																			
1	Recent Files	,																		
									C10	C11	612								C10	
	CI C2	G	C4	6	C6	67	C8	Ca	C10	C11	CI2	C13	C14	C15	C16	C17	C18	C19	C20	
1	3.048																			
2	3.054																			
3	3.079																			
4	3.047																			
5	3.067																	6	1	
6	3.065																	1	0	
4.0	+ + Worksheet 1									•									EL.	
4	Worksheet 1																1	100	410	
X	P Type here to searc	:h			0	a f	1	-	1					×	16	^ @ W	10 10 1	6	113	
NPT	L								-									100		

(Refer Slide Time: 03:20)

And then when you click that icon, what will happen is that we will get some interface like this. And we can just see the interface over here and C1. So, these are the columns where you have to cut paste the data set. So, first row over here where you can write the data, name of the data variables like this.

And then you can type in the data also, or you can copy paste the data from excel sheets. So, I will copy paste the data from over here. And I have already copied this one and then what I will do is that in MINITAB this worksheet, what I will do is that I will highlight on the first row, and then I will control V, I will paste this one ok.

Whenever I have pasted this one, so there are there are windows over here. So, what you see over here, this is the session window where the results will come when you do some analysis over here. And if you have to save this worksheet which is coming over here, so this worksheet this is known as worksheet over here like excel what you see worksheets.

(Refer Slide Time: 04:23)



So, this can be saved over here save worksheet as. So, what you can do is that save worksheet as, and then on the desktop maybe we can save this one. Let us say visualization of data I am going to replace this file which is existing. So, I am just replacing with a new file.

So, you can also name a new file because you know I had already created earlier so for some illustration. So, again I am redoing that when I have replaced that file. So, now, this file consists of only C1 columns with layer thickness ok.

日 (1) (1) (1) (1) (1) (1) (1) (1) (1) (1)		Eatterplot Matrix Plot Bubble Plot Marginal Plot		で言 際日 岡上	22 2/4 33 ₩ 1% ★	() 2 () () () () () () () () () () () () ()	- γ ⊠	Bek	* ¹³ E									
	PNR Hill	Histogram Dotplot Stem-and-Leaf Probability Plot Empirical CDF Probability Distributio	on Plot			Mi	ini	ta	b	>								
	豊臣国人	Boxplot Interval Plot Individual Value Plot. Line Plot					New Pr	Open Ctr oject Ctr	I+O I+Shift+N									
	11 1	Bar Chart Pie Chart				1	New Work:	iheet Ctr	I+N									
		Bar Chart Pie Chart Time Series Plot Area Graph)	New Work:	iheet Ctr	I+N									
		Bar Chart Pie Chart Time Series Plot Area Graph Contour Plot					New Work:	heet Ctr	I+N									
CI C2		Bar Chart Pie Chart Time Series Plot Area Graph Contour Plot 3D Scatterplot 3D Scatterplot		C6	C7	C8	New Works	theet Ctr	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
CI C2 Layer Thickness 3.048		Bar Chart Pie Chart Time Series Plot Area Graph Contour Plot 30 Scatterplot 30 Surface Plot		C6	C7	C8	New Works	c10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
CI C2 Layer Thickness 3.048 3.054		Bar Chart Pie Chart Time Series Plot Area Graph Contour Plot 30 Scatterplot 30 Surface Plot		C6	C7	C8	New Works	cio	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
C1 C2 Layer Thickness 3.048 3.054 3.079		Bar Chart Pie Chart Time Series Plot Area Graph Contour Plot 3D Scatterplot 3D Surface Plot		C6	C7	C8	C9	cto	C11	C12	CI3	C14	CIS	C16	C17	C18	C19	C20
C1 C2 Layer Thickness 3.048 3.054 3.057 3.057 3.047		Bar Chart Pie Chart Time Series Plot Area Graph Contour Plot 30 Scatterplot 30 Surface Plot		C6	C7	C8	C9	c10	C11	C12	C13	C14	CIS	C16	C17	C18	C19	C20
C1 C2 Layer Thickness 3.048 3.054 3.079 3.047 3.067		Bar Chart Pie Chart Time Series Plot Area Graph Contour Plot 30 Scatterplot 30 Surface Plot		C6	CT	C8	C9	c10	C11	C12	C13	C14	CIS	C16	C17	C18	C19	C20

(Refer Slide Time: 04:48)

Then I told that where from I will draw the graph. So, there is graph option over here, there are many options that you can see over here. So, out of the all the options you have histogram over here as one of the options. So, there can be other options. So, I am concentrating on histogram over here.

So, for visualization of the data, this is a continuous data, continuous scale data, so that you can understand. Thickness is measured over here. And it is a least count is up to three decimal place which was measured over here ok.

(Refer Slide Time: 05:15)



So, what I will do is that, I will plot this histogram over here. So, I clicked histogram, so then it will have simple histogram ok. I will mention it ok, I will draw a simple histogram.

(Refer Slide Time: 05:22)



Then what is the graph variable, your cursor should be over here, and double click this C1 column over here. And otherwise what you can do is that I can just delete this one, I can just highlight this one and select from here. So, that is also possible.

(Refer Slide Time: 05:36)



(Refer Slide Time: 05:40)



Then you have some scales over here keep it as default, do not want to change this one. Labels, I do not want to; if you want to name or give some titles that you can give ABC or something this is always possible. So, I am not doing this. This you can explore by yourself ok. So, data views and all these things do not do anything over here. We do not want multiple graphs to be plotted. So, data options are over here.

(Refer Slide Time: 05:58)



So, these are also not required because I have already given the options that layer thickness has to be plotted ok.

(Refer Slide Time: 06:07)



So, maybe the frequency over here which you want to; which you want to highlight over here. So, maybe that if that information is available. So, in this case, that can be provided over here that is not required over here. So, label what we can do is that maybe data labels what we can do is that use Y values to data label over here. So, in this case what will happen. When I click OK over here, I will get the frequency information also.

(Refer Slide Time: 06:30)



So, this session window will give you this window will give you all the results over here and graphs. So, when you double click this graph, you can just see that MINITAB has already done the basic analysis, and it has plotted the histogram based on the its default options that is there.

So, a number of classes what do you see over here 1, 2, 3, 4, 5, 6, 7, 8, 9 classes, it has divided the total data set and the width of the class is the interval which MINITAB automatically determines. And there are formulas to do that also. So, in that case, MINITAB does it automatically for you and gives you the frequency.

So, this over here what you can see this is the bin size over here. So, this is varying from 3.015 to 3.025. And within this bin, there are five observations. Similarly, next bin you can see how many or in that class how many observations.

So, maximum observations are within this range over here what you see 3.07. So, it is varying from this range over here. So, this is 3.055 and this will be around 3.075. So, we can assume 3.07 more or less is the average or center of gravity or accuracy or what you can see the CG of the data set right over here ok.

So, and the spread of the data varies from this range to this range. So, what you can see that peak of the data set. So, we can also plot this one by we can also what we can do is that when I draw histogram, it will ask with fits.

(Refer Slide Time: 07:59)



So, some normal distribution overlayer graphs can be placed, overlayer distribution graph can also be placed over here.

(Refer Slide Time: 08:05)



So, if you click OK, so it shows you some normal distribution plot over here. So, mean is given, standard deviation is given. Number of observation is 72 that will be given over here. So, these observations and also you can copy this graph like this. So, if you want to copy the graph and paste it in reports or in excel file.

(Refer Slide Time: 08:22)



You paste it over here and just move wherever you located. So, this is possible in board. And this is also possible in excel. So, you can save these graphs also ok. So, that is also possible. So, this gives you some basic information. Now, mean standard deviation, I may need more information over here, so that will not come in histogram over here. So, maybe maximum values, minimum values.

(Refer Slide Time: 08:47)



So, then what I have to do is that I have to go to stat, and they could go to basic statistics and maybe display descriptive statistics. So, one go you can draw histograms also using this option. So, MINITAB has this stat option where we will use basic stat, regression, ANOVA, in this course design of experiment, control chart, quality tools.

Some more with respect to time we can plot also which is known as time series we will not cover, but we will show you some illustration of graphical illustration how it is done. So, a MINITAB has options of non-parametric test. Those are the tests we will also illustrate in certain scenarios where it is required ok. So, what we will do is that now in this case, so, I want to see the basic statistics.

(Refer Slide Time: 09:32)



So, in this case display descriptive statistics I will mention which is the so cursor will be here and double click C1.

(Refer Slide Time: 09:36)

Histogram of Laver T			Z Y N	B 8K												
unite Brann an enfer in	nickn × ×															
INSUAL DATA MWX	ayer Thickr	less			Disp	Display Day	rolection Statistics itatistics: Statistics		×							
Frequency 5 8	Histo	gram of La	yer Thickr	ness		jean E of megn Zandard deviation (ariance icefficient of varia inst quartile teglian tyrid quartile nterguartile range tode Help	F Trimmed m Sym IF Sym IF Majmum IF Magmum IF Bange IF Sum of squ IF Semperss IF Symbols IF	an F N norr F N mass F N total F Qumuk F Dencer F Qumuk res Oheck start © Defaul C None C All	nissing 199 Ative N A Ative percent Istics t Cancel							
	C2 C1	C4	CS	C6				_(;		C14	C15	C16	C17	C18	C19	C20
ci Ci																
C1 Layer Thickness 2.048																
C1 Layer Thickness 3.048 3.054																
 C1 Layer Thickness 3.048 3.054 3.079 																
 C1 Layer Thickness 3.048 3.054 3.079 3.047 																
 C1 Layer Thickness 3.048 3.054 3.079 3.047 3.067 																

So, what are the statistics that will be required? So, I have gone to statistics and I have clicked that one. I want mean information, I want standard deviation information, variance information, coefficient of variance information, first quartile median, these are all information what I want to see; maximum, minimum, range, these are not required. So, maybe these are the basic things one wants to see in the data set.

(Refer Slide Time: 09:58)



So, in this case, I will click OK. And if you want to see histogram over here also with normal fits is also possible.

(Refer Slide Time: 10:04)

Minitat File Ed	it Data Ca	ic Stat	Graph Vie C □ 0	tw Help € # ⊘ -Y ⊠ €	Assistant	Additional Te 2= -2 ↓ ≤ ★ \$\} {	ools 設 2/ 4 図 非 均	(≪ ∭ # 指	Y 🔀	8	< ★ 号 圖									- 8	×
Descript	tive Statistics:	Laye `	×																		
	iptive Sta	tistics	: Layer '	Thickne	SS																•
Stati Varia Løyer Varia Løyer	stics ble Thickness ble Thickness H	Mean 3.0613 Range 0.0873 istograr	<u>StDev</u> V 0.0209 0.0292 n (with N	ariance C 0.00044	0.68 0.68 Irve) of Li	inimum 3.0173 3 ayer Thicl	Q1 Me 3.0468 3. kness st0	dian Q 0540 3.076 m 3.061 ev 0.02090	3 Maximu 0 3.10	10 10 10 10 10 10 10 10 10 10 10 10 10 1	Send to P Send to P Send to O Send to O Send to P Copy Co Decimal Print	ficrosoft® ficrosoft® fompanion finitab Wo licture umn Places	Word PowerPoint rkspace ^{to} C	я+С ,							
,	4						N	72			X Delete										¥
	CI	C 2	C3	C4	CS	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	P
Laye	r Thickness																				
	3,048																				
	3.079																				
	3.047																				
	3.067																				
	3.065																		6	0	
400	+ Vis	ual Data.n	WX								4							-		8	
AY	isual Data.mw																⊞ C	1 1	100	10	
NPTEL	O Type he	re to sea	irch			0	i		2	۲		8	C	1	×B	16	∧ ĝ ₩	<i>(</i> , 4)	1		

So, this when you click this one, so this comes automatically in MINITAB. So, now, these things can also be, the results can also be copied as a picture, and you can just paste it wherever it is required. So, if you want to paste it over here, you just paste it over here, and then save that one or in reports anywhere ok.

(Refer Slide Time: 10:17)



So, what information we get over here mean value is 3.0613. So, up to fourth place of decimal data sets, we are getting standard deviation. Coefficient of variation, that means, what is the magnitude of variance with respect to mean. So, variance information is also given over here minimum, maximum, inter quartile, quartile 1, quartile 3, median value that 50 percent observation are less than this 50 above this one.

So, range information is also given inter quartile range is also given. So, all information we can get out of the MINITAB software which is necessary over here. So, mean is 3.06, but I told you initially that our target is 3, that means, CTQ that is generated thickness layer that is generated over here should lie between 2.9 to 3.1 let us say, and the target value is 3.

(Refer Slide Time: 11:11)



So, what we can see from this graph is we are basically off-centered you see. So, 3 is around on this axis on the left-hand side. And our average what is coming out of the process is around 3.07. So, we are accuracy part we are missing over here. And also, we are getting some information which is beyond 3.1 that is out of spec also data we are generating over here ok.

So, that is also an alarming situation in quality, that means, our process is off-centered, variability is high, and also, we are producing which is creating rejections also. So, this gives you an immediate impression of the data set without much statistics I can make this kind of interpretation out of the data ok.

So, another important thing what can be done when whenever I have the spread of the data, we can also see another important graphical illustration over here which is known as box plot.

(Refer Slide Time: 12:01)



So, here also box plot is sometimes used to define some of the. So, this what it will show is that what is the width means within which we have the complete data set. So, this is the range which is known as interquartile range within which most of the data sets are concentrated over here.

So, if it is more spread, that means, we have a much variability in the data. If this spread is very less or we have a small spread of this. So, in that case, data is squeezed more or less over here. And in that case, density of the data is high. So, variability is less we can also interpret way. So, it is like spring more it is spread more, more less, that means, more data spread of the data is quite high.

And if this is less squeezed, that means, we have a much less variability in the data. So, we can think about that. And these are the stretchable extension what you see over here. So, these are known as whisker over here. So, this line what you see over here extended up to 1.5 into interquartile range what you see over here definition. And anything lying outside this is basically outlier over here that is indicative measures that it takes.

Anything beyond this is an outlier there are other tests for in identifying outliers, but this is in box plot also we can see that which data is very unusual. So, that we can also see from the box plot. So, this on this the first start of the box over here this is known as first quartile Q1, and this is the second quartile which is known as Q2 or the median value of the data set where 50 percent of the observation will be on the right hand side of this

value and 50 percent on the left hand side; left hand side of this value and right hand side of this.

And this is known as third quartile values over here. And difference between this first quartile and third quartile, this area is known as interquartile range ok. Here also Whisker is extended to the smallest value 1.5 of that. And anything beyond that is a outlier basically ok. So, this is also possible in MINITAB. So, we will use the same data sets to illustrate box plot where we have drawn histogram.

Here we want to see the interquartile and where the data is most concentrated. So, what we will do is that is also possible over here. So, that is also possible over here. So, this same data set we are using. So, I will go to MINITAB over here.

(Refer Slide Time: 14:24)



(Refer Slide Time: 14:26)



(Refer Slide Time: 14:30)



And then what we can do is that we can just see the box plot, and then we go to graph, and there is a box plot option over here. Then there is simple box plot over here. I click OK. And I give the variable a C1, double click that one. And do not go, I do not need to see much options over here.

(Refer Slide Time: 14:37)



So, we can place the values over here also. So, just click on this, and you will get all information quartile 1 is 3.04 that is a 25 percent of the observation is less than this; 3.06 is the median value. So, 50 percent of the observations are below this, and 50 above this one. So, this is the interquartile range.

(Refer Slide Time: 14:57)



So, this is the spread of the data set what we can see. So, most of the data set are concentrated in this zone. So, data set concentration can be seen over here, so ok. So, this is the width of the box. So, most of the data are concentrated within this, ok. So, this is

the interquartile range. So, this information we can get and how this is useful I will use an example to illustrate that one.

So, let me also take some examples, like let us say the quality marks and which I have not discussed or what we have started. So, I am taking these marks in quality. I have to make a decision let us say whether to attend the course this is the last year marks whatever what was given.

(Refer Slide Time: 15:39)



So, I want to select whether to go for this course or not.

(Refer Slide Time: 15:53)



(Refer Slide Time: 15:57)



So, what we can do is that just we can see the basic statistics of this. So, display descriptive statistics. I will display the marks. And statistics what we have mentioned is already given over here. So, and graphical illustration I want to see the box plot of this rather than histogram. And, so I will click that one OK and I will click OK over here and then I will get the box plot over here.

(Refer Slide Time: 16:07)



So, on an average the 50 percent median value is approximately equals to 72, so that means, on an average people are getting work. What we can say is that the middle value

50 percent of the people are below this, but 50 percent are above 72. So, this is a favourable situation what we can see over here.

And also, we can see that 72 is the average value from here. This can also be copied, results can also be copied from here, copy as a picture. And you can paste the results what I have shown earlier also over here. So, statistics can be pasted. So, that is also possible. So, in this case, what we are seeing is that on an average 72 marks are there.

(Refer Slide Time: 16:45)



(Refer Slide Time: 16:49)



So, I can also see the histogram of the data set. And if you want to see the histogram of the data set, then some more information we can get. Just click, double click that variable graph variable, and click OK over here.



(Refer Slide Time: 16:57)

And then what you see is that you will get the histogram of the data. So, what you can see is that on the left-hand side much less observations are there. And on an average, 70 is the we can say center of gravity of the data. So, most of the people are more than 70 what we can see over here.

So, this is a favourable situation just by drawing what I am seeing is that number of people getting more than 70 is much higher as compared to the number of people getting less than 70. So, that is also reflected over here. So, it is more skewed on the right-hand side as compared to the left-hand side.

So, on the higher side of marks basically. So, in this case, I can make an immediate judgment ok, this course in that case I can take a chance because people are getting more than 70, most of the people are getting more than 70. It is skewed on the right, so that is why this is a favorable situation. So, that also is possible.

(Refer Slide Time: 17:47)



(Refer Slide Time: 17:50)



(Refer Slide Time: 17:55)



But if I want to make a box plot and interpretation also, I can see that one, and I can also see the marks information, and I can draw the box plot. And when I see that one also, I see the same information, but there is some variability over here, and but favourable is more than 70, many observations are more than 70.

So, you do not see symmetry over here. So, more observations on the other side, so density of that is more. So, colour blue is on the, so it is not symmetrically what do you see. So, there is this more values on the other side. So, skewness will be represented also in box plot.

You can directly see which side it is more skewed. Data is concentration, data concentration is more on one side as compared to the other side. So, that gives you an immediate impression that what we can do out of this data that we are. So, some inference we can draw, some decision we can make out of this ok. So, box plot can also be used we can draw multiple plots over here.

So, let us take another example like I have some data set where restaurant information is given. And I want to see which restaurants to go which has minimum service time. So, some data is provided over here. So, like a Burger King, Pizza Hut, McDonalds, Subways, Domino's Pizza and all are providing burgers let us say, and I want to take wherever there is a least time. So, some random sample information is given over here.

So, this information was collected by some researchers. And based on that, I have to make a decision which restaurants I should go. So, what I can do is that I can just copy this information. I do not know much about statistics, but I want to place that one. And I want to compare these data values.



(Refer Slide Time: 19:30)

So, in this case, in one go, I want to analyze this one. So, Burger King this data I have pasted from there copy pasted from excel, and then I want to see that how its box plot, can I see multiple graphs in one. So, what I will do is that, I will go to box plot over here, then I will go to multiple Y options over here, I click OK over here.

(Refer Slide Time: 19:50)



Then I, what I do is that Burger King from here, and I press shift and I select up to C10. So, then I select this one and it will automatically; or otherwise, what you can do is that just double click one, then second one, third one, fourth one and fifth one. And keep it as default and whatever conditions.

(Refer Slide Time: 20:09)



So, you will get this graph in the session window. So, if you enlarge this graph, what happens is that you can see in one scale in one graph you can see the variability of the dataset. So, the Pizza Hut condition what you see is that it is this is the median value

over here. And the median value is coming out of 143 which is very less as compared to any of the median that you see that central line that you see any of the median.

So, this is the lowest, it is giving you the lowest median, that means, you can also see that this information helps you that what is the median value of the observations which can be taken as the location of the data central tendency of the data set also we can consider. Sometimes medians can be useful for that interference also, but mean also can be calculated and we can see the values of means ok.

But what it says is that the median value is very low is lower than any of these options. So, immediately I can discard Domino's Pizza because this is very high. And then subway it is also high and variability is also high. This bandwidth of the box that you see over here the bandwidth is quite high. So, this is also not favourable for me, and this is also quite high, Burger King is also quite high.

So, I have options between Pizza Hut and McDonald. But if you see Pizza Hut and McDonald, what comes into my mind is bandwidth of the Pizza Hut that interquartile range is quite low as compared to McDonalds over here. So, I will select Pizza Hut because that is the service time of Pizza Hut is less as compared to McDonald over here. Although you can see some outliers over here, there is some service disruption may have happened in one or two cases, but I may be willing to take the risk.

But if it is on the lower side, you see that one star is over here on the lower side of this. This is also an outlier, but this is a favourable situation. If you see delivery time less means it is a favourable situation. So, in this case, I am not worried about this. And these two I can take it because I can see that most of the observations are having less variability in service time.

So, variation in service time is also less and the median value is also quite less. So, I will take a decision to go to Pizza Hut, and take the deliveries. So, if I have to choose between any of these restaurants based on the data, I will choose that one. So, a multiple box plot can be shown in single graph, that is also possible in MINITAB.

So, that is a quick understanding of the data I will have visualization of the data that is possible in many situations we will have such kind of plots which will be helpful in quality also. So, this is box plot what we have illustrated. So, you can just practice that one, and see some of the data sets of CTQ. And plot in single one or plot in multiple plots in single graph, so that is also possible ok.



(Refer Slide Time: 23:11)

So, then we can do is that, we can also understand and plot like dependency between variables which is known as scatter plot. Sometimes what happens is that this is extensively used in regression. So, this can be used in regression, and we have regression analysis.

So, I will so this is one variable. So, this is one X variable on this axis over here. So, and the other axis is Y over here. So, this is another axis what we can see ok. So, here what is important is that, whether there is any relationship between the data set or not, so that I want to check.

So, whether there is any relationship that exists between the two data set X and Y. So, there I will see this is extensively used in when I am trying to develop a function between Y and X. So, that is where we will use. So, if I am interested to illustrate that this is a X variable, Y is a function of X, I want to see that one.

And whether linear function, non-linear function, so this scatter plot will help me to understand that whether one increases, other also increases or not, so what is the correlation that exists between the data set, strong relationship or weak relationship; linear, non-linear, many things are possible over here ok. So, this will help in regression analysis. And regression is an important tool in quality. So, that is used in design of experiments also, ok. And that is also used in prediction. Many of the times predictive models are used to control the process to get the best CTQs, so that is also an important aspect what we cover in quality.

So, that is, but you can see any regression analysis in many other courses like statistical course. So, we will give a brief idea of regression also in this course ok. So, what I wanted to illustrate that this gives you a demonstration between relationship between two variables. So, this is known as scatter plot.



(Refer Slide Time: 25:16)

And the relationship can be like this. So, you can see that X and Y is on this axis. So, in this case, the relationship between these two variables name you can change it. So, one axis may be X_1 , other axis maybe X_2 , but X and Y is that I am trying to say that one is the CTQ, and one is the control variable. So, in this case, control level variable. So, that I want to illustrate over here. And for that what I am doing is that I am showing. So, this is a linear relationship.

So, what we can see is that, one increases, X increases on this side, Y also increases like this. So, here you can see some curvilinear relationship exists. So, this is X increases, initially it increases, and then it becomes steady, and then it becomes this down. So, it is a non-linear kind of scenario. This can also be thought of as non-linear relationship means this is a polynomial equation basically. So, maybe X square is prominent over here. So, in this case some square equation will be prominent over here.

And here it can be negative relationship that means one increases other goes down. So, one is decreasing. So, one is increasing this can be positive, and this can be negative relationship what you see over here ok. So, these are the types of relationship when you plot that one, you will immediately you can visualize that one and immediately you can make an interpretation out of that ok.

(Refer Slide Time: 26:43)



So, visualization of the data set. So, then here you can see that if the density of the data are more concentrated towards this, then that means, strong relationship. So, bandwidth of this is very less and all the data points are within this. So, this is a strong relationship positive one.

And when positive means, one increases, other also increases, and linear relationship that we have to take into mind to consider. So, but the width is very high over here. So, in this case, not very strong relationship. So, how do we say strong weak?

So, we have a correlation coefficient to understand the relationship between variables ok. Pearson correlation coefficient can be seen to see linear relationship between Y and X if X increases, Y increase. What is the strength of the relationship that can be seen linear relationship I am saying. So, in this case, there are options ok. (Refer Slide Time: 27:30)



So, this can be seen. So, this is now there is no relationship. You see random ups and downs no trend as we can see increasing, decreasing, nothing is possible over here ok.

(Refer Slide Time: 27:49)



So, and this relationship what we see over here can be expressed in mathematical terminologies which is known as correlation coefficient over here which is given by covariance a measure that association between two variables divided by standard deviation of x variable, standard deviation of y variables. This is the covariance formulation over here what you see.

And covariance divided by these two will give you information about the strength of relationship. So, r lies between minus 1 and plus 1. So, this will be the minus 1 or plus 1 relationship that you will see ok.



We will take an example to illustrate that one. So, let us take one examples to illustrate this one relationship between both variables, how it can be done in MINITAB. So, I will take one examples over here. And let us take a simple example scatter plot example.

(Refer Slide Time: 28:49)



(Refer Slide Time: 28:32)

And where we have size of the flat and price of the flat, and I want to see whether relationship exist or not. So, these are coded variables. So, I am just copy pasting the information. So, and I am pasting it over here. So, in this case, I will paste the information over here. And then what I will do is that I want to see the relationship. And first I will draw the scatter plot, and then figure out what is the correlation between these two data points.

(Refer Slide Time: 29:13)



(Refer Slide Time: 29:16)



So, what I will do is that I will first option is scatter plot over here. So, when you click this one, what will happen is that it will show do you want a simple scatter plot? Yes, I want a simple scatter plot. There are many options with regression and all these things, I am not interested. I want to see.



(Refer Slide Time: 29:26)

So, which is a Y variable, I am taking arbitrarily price as a Y variable and size of the flat as X variable over here. There are other options that is available over here. I am not going to the options. You can always explore that one. I will click OK over here and it will give me a visual impact over here.

(Refer Slide Time: 29:44)



So, if you see the graphical illustration that indicates that size of the flat increases, price also increases. So, this gives you an impression that maybe they are related; maybe they are related over here ok. So, and positively related over here, linear relationship may exist but the strength of the relationship I need to check.

(Refer Slide Time: 30:12)

	X III III	East Graph View Help Basic Statistics Regression ANOVA	Assistant Add	ational tools ay Descriptive Descriptive St hical Summary	Statistics atistics	44 (× 1 8k	* 15	6						
Scatterplot o UISUAL DA Scatterp	of Price for ti TA.MWX Iot of Pr	DOE Control Charts Quality Tools Reliability/Survival Predictive Analytics	1-Sar 1-Sar 2-Sar 1 Paire	nple Z nple t nple t d t											•
S 340	catterplo	Multivariate Time Series Tables Nonparametrics	, 1 Pro ↓ 2 Pro ↓ 1-Sar ↓ 2-Sar	portion portions nple Poisson F nple Poisson F	Rate Rate	~									*
(p) 320		Equivalence Tests Power and Sample Size	• 11 Corre	iance											
at (Co			O ² Cova	niance.											
Price for the Flat (Coo		•	o ² Cova A Norm - Outli λ Good	riance. hality Test er Test fness-of-Fit Te	st for Poisson										¥
00 300	1 C	• • 2 C3 Marke in Queller C	α ² Cova Norm Outli λ Good C4	riance. [5 nality Test er Test Iness-of-Fit Te CS	st for Poisson C6	C7	C8	C9 Subarau	C10	C11	C12	CI3 Relia for the Elst (Carled)	C14	C15	*
00 300 - 11 230 - 12 240 - 240 - 240 - C Layer Th	1 C ickness 3.048	e 2 C3 Marks in Quality Co	α² Cova Norm Outli λ Good C4 Ourse 65	riance.	st for Poisson C6 Burger King 150	C7 Pizza Hut 173	C8 McDonalds 198	C9 Subway 158	C10 Domino's Pizza 289	C11	Ct2 Size of Flat (Coded) 23	C13 Price for the Flat (Coded) 315	C14	C15	•
0 0 00 00 00 00 00 00 00 00 00 00 00 00	1 C ickness 3.048 3.054	e 2 C3 Marks in Quality Co	α² Cova Δ Norm → Outli λ Good C4 C4 000056 71	riance.	c6 Burger King 150 197	C7 Pizza Hut 173 95	C8 McDonalds 198 131	C9 Subway 158 195	C10 Domino's Pizza 289 273	C11	C12 Size of Flat (Coded) 23 18	C13 Price for the Flat (Coded) 315 229	C14	CIS	•
 C C C Layer The C Layer Th C C	1 C ickness 3.048 3.054 3.079	e 2 C3 Marks in Quality Co	α² Cova Δ Norm → Outli λ Good c4 C4 ourse 65 71 66	niance [] naity Test er Test dness-of-Fit Tes CS	c6 Burger King 150 197 186	C7 Pizza Hut 173 95 154	C8 McDonalds 198 131 158	C9 Subway 158 195 177	C10 Domino's Pizza 289 273 254	C11	C12 Size of Flat (Coded) 23 18 26	C13 Price for the Flat (Coded) 315 229 235	C14	C15	•
• C Layer Th L L L L L L L L L L L L L L L L L L L	1 C ickness 3.048 3.054 3.079 3.047	e 2 C3 Marks in Quality Co	α² Cova Δ Norm •• Outli λ Good 65 71 66 79	riance.[nality Test er Test Iness-of-Fit Tes CS	c6 Burger King 150 197 186 166	C7 Pizza Hut 173 95 154 104	C8 McDonalds 198 131 158 123	C9 Subway 158 195 177 147	C10 Domino's Pizza 289 273 254 282	C11	C12 Size of Flat (Coded) 23 18 26 20	C13 Price for the Flat (Coded) 315 229 355 261	C14	CIS	•
00) 100 - 101 11 200 - 415 200 - 92 - 92 200 - 92 -	1 C ickness 3.048 3.054 3.079 3.047 3.067	• • 2 C3 Marks in Quality Ce	α² Cova Δ Norm α² Cova Δ Norm α² Cova Δ Good δ5 C4 65 71 66 79 65 S	riance.[.] nality Test er Test thess-of-Fit Te	t for Poisson C6 Burger King 150 197 186 166 196	C7 Pizza Hut 173 95 154 104 183	C8 McDonalds 198 131 158 123 139	C9 Subway 158 195 177 147 181	C10 Domino's Pizza 289 273 254 282 282 235	C11	C12 Size of Flat (Coded) 23 18 26 20 20 22	C13 Price for the Flat (Coded) 315 229 355 235 234 234	C14	CIS	•

So, in this case, strength of the relationship, how do we see the strength of the relationship over here? Then I have an option to see that one. So, basic statistics then I have a correlation matrix, where the formula that I have shown covariance divided by

standard deviation of X and standard deviation of Y, that formulation what we have shown over here. This formulation will be used.

(Refer Slide Time: 30:25)



So, this is $\frac{S_{xy}}{S_x S_y}$. So, this is the relationship it will indicate what is the value of r. So, r is

on the positive side, plus 1 side, so that is positive strong positive relationship; on the negative side towards minus 1, and close to minus 1 strong negative relationship linear relationship basically ok. So, then what I can do is that I want to see this one. So, in this case MINITAB, I will use this one stat, basic stat, calculation will be done based on the data set X and Y over here which I have defined.

(Refer Slide Time: 30:58)



So, correlation between two variables, so I want to see price and size, let us say. So, you can just. So, sequence of this is not important whether price is placed first, size is placed second, it does not matter.

(Refer Slide Time: 31:11)

		1110	⁶ う	● ● # # □ 5 ■ ■ = = = = = = = = ■ = = = = = = = = =	fx }• - 8 k ★ 'Y ≥ 1 8	は日 9月日 1日	₩ ⁴ 4 (2) 非 ¹⁶ (1) # 4	N Y ≥ I	8K)	e 45 🔛					
icatt	verplot o	of Price for	r the	▼ X											
ica!	tterp	lot of I	Price	for the Flat (Codeo) vs Size	Correla	rtion					×			
	s	catterp	lot of I	Price for the Flat (Cor	ed) vs Size	C1 C3 C6	ayer Thickness Harks in Quality Cour	Variables: Price for	the Flat (Code	ed)' 'Size of Flat	(Coded)'	~			
	360				•	C7 C8 C9 C10 C12	Azzi Correlation: Op AcD Don Suby Method:	vtiens learson correla	iton	•	×	~			
						C13	Pric Confidence level	: 29	5.0						
de de	(p 320			•											
The Prestor	120 Jon 120 Jo			•	•		☐ Store correla	tion matrix							
being fast than films after dealer	700 000 000 000 000 000 000 000 000 000			•	•		F Store correla	tion matrix		Ūk					
Notes for the flat of the sec-	120 320	1	2	•	•		F Store correla	ton matrix		QK QK	Cancel Its	C12	CI3	C14	C15
The factor for the first of the	(page 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	1 ickness	62	C3 Marks in Quality Cours	•		Help	ton matrix		QK QK	Cancel Its Cancel	C12 Laize of Flat (Coded)	C13 Price for the Flat (Coded)	C14	C15
La Contra California California	(popo) 100 -	1 ickness 3.048	62	C3 Marks in Quality Cours 6	•		Help Uniger King Pla	ton matrix	contaitus 3 198	<u>QK</u> <u>QK</u> 158	Cancel Its Cancel Its Cancel 289	C12 Jalze of Flat (Coded) 23	C13 Price for the Flat (Coded) 315	C14	C15
T Nice for the first of the state	(popo) 300 300 240 240 C1 Nyer Thi	1 ickness 3.048 3.054	C2	• • C3 Marks in Quality Cours 6 7	•		Help Help 150 197	ton matrix are not mot 173 95	198 131	QK QK 158 195	Cancel its Cancel 289 273	C12 alze of Flat (Coded) 23 18	C13 Price for the Flat (Coded) 315 229	C14	C15
T Non far the far the given on the far	(popo) 320 300 240 240 C1 ayer Thi	1 ickness 3.048 3.054 3.079	62	• • • • • • • • • • • • • • • • • • •	•		Help Help 150 197 186	ton matrix 	2011a103 5 198 131 158	0K 0K 158 195 1777	Cancel Its Cancel 289 273 254	ct2 size of Flat (Coded) 23 18 26	C13 Price for the Flat (Coded) 315 229 355	C14	C15
The factor for the state	(paper) 300 1113 200 200 200 240 Cl ayer Thi	1 ickness 3.048 3.054 3.079 3.047	C2	• C3 Marks in Quality Cours 6 7 6 7 7	•		Help Help 150 197 186 166	ton matrix 173 95 154 104	198 131 158 123	0K 0K 158 195 177 147	Cancel ts Cancel 289 273 254 282 282	C12 pize of Flat (Coded) 23 18 26 20 20	C13 Price for the Flat (Coded) 315 2290 355 261	C14	C15
The factor of the second secon	(page 1 200	1 ickness 3.048 3.054 3.079 3.047 3.067	2	C3 Marks in Quality Cours 6 7 6 7 7 6 0	•		Heb Heb 150 197 186 166 196	ton metrix 244 million medrix 173 95 154 104 183 130	198 131 158 123 139	0K 0K 158 195 177 147 181	Cancel ts., Cancel 209 273 254 202 235 245	C12 28e of Flat (Coded) 28 26 20 20 22 22	C13 Price for the Flat (Coded) 315 229 355 281 281 281	C14	CIS

(Refer Slide Time: 31:17)



(Refer Slide Time: 31:31)



So, option is that I want to see Pearson correlation over here. So, in this case, and graphical matrix plot over here. So, what you want to see over here, I want to see correlations between the variables. So, only there are many options, so statistical interpretation. So, I will use the simplest one over here.

So, I will click OK over here. Then what you want to do you want to see the tables. So, whatever default also you can keep ok. So, correlation coefficient is coming out to be r, 0.727.

(Refer Slide Time: 31:48)

	initab - Unobled														- 6	
File	Edit Data Cal	lc Stat	Graph View Help Assista	nt Additi	onal Tools											
		15	A A O O	fx 30 .	e 1 3	400										
		11 11		ak	14.69	****	丧 Ly b	x Rek	- NL (P							
<i>c</i> .	valation: Drice for	the f														
.0	relation, Price for	UPE Pas														
	relation: Dr	ico foi	the Elet (Coded) C	To of I	lat (Co	dad)										0
.c	rrelation: Pr	ice io	the Flat (Coded), 5	ize of r	iat (Co	ueu)										
	p: pairwise Pearson	n correlati	ion													1
1	Correlations															
		Price	for													
		(Co	Flat													
1	Size of Flat (Coded)	0	727													
			ir ar													
-	Pairwise Pearso	on Corr	elations													
-	Pairwise Pearso	on Corr	elations	rrelation	95% CI fr	ro P-Value										
	Pairwise Pearso Sample 1 Size of Flat (Coded)	Samp Price f	elations le 2 N juge	rrelation 0.727	95% CI fo (0.262, 0.5	er p P+Value 18) 0.007										
	Pairwise Pearso Sample 1 Size of Flat (Coded) C1	Samp Price t	elations le 2 N pC br the Fiat (Coded) 12	rrelation 0.727 C4	95% CI fc (0.262, 0.5 CS	r p P-Value 18) 0.007 C6	67	C8	C9	C10	C11	C12	C13	C14	CIS	
	Pairwise Pearsc Sample 1 Size of Flat (Coded) C1 Layer Thickness	Samp Price t	elations le 2 N pc for the Flat (Coded) 12 C3 Marks in Quality Course	c4	95% CI fc (0.262, 0.5 C5	r p P-Value 18) 0.007 C6 Burger King	C7 Pizza Hut	C8 McDonalds	C9 Subway	C10 Domino's Pizza	C11	C12 Size of Flat (Codec	C13) Price for the Flat (Coded)	C14	C15	A DECEMBER OF A
	Pairwise Pearso Sample 1 Size of Flat (Coded) C1 Layer Thickness 3,048	Samp Samp Price t	elations le 2 N pc for the Flat (Coded) 12 C3 Marks in Quality Course 65	0.727 C4	95% CI fc (0.262, 0.5 CS	r p P-Value 18) 0.007 C6 Burger King 150	C7 Pizza Hut 173	C8 McDonalds 198	C9 Subway 158	C10 Domino's Pizza 289	C11	C12 Size of Flat (Codec 2	C13 Price for the Flat (Coded) 3 315	C14	C15	
	C1 C1 C1 C1 C3 C1 C3 C1 C3 C3 C3 C3 C3 C3 C3 C3 C3 C3 C3 C3 C3	Samp Price t	elations le 2 N pC for the Flat (Coded) 12 C3 Marks in Quality Course 65 71	rrelation 0.727 C4	95% CI fc (0.262, 0.5 CS	rp P-Value 18) 0.007 C6 Burger King 150 197	C7 Pizza Hut 173 95	C8 McDonalds 198 131	C9 Subway 158 195	C10 Domino's Pizza 289 273	C11	C12 Size of Flat (Codec 2 1	C13) Price for the Flat (Coded) 3 315 8 229	C14	C15	
	C1 C1 Layer Thickness 3.048 3.054 3.079	Samp Price t	elations le 2 N jpc or the Flat (Coded) 12 C3 Marks in Quality Course 65 71 66	C4	95% CI fc (0.262, 0.5 CS	r ρ P-Value 18) 0.007 C6 Burger King 150 197 186	C7 Pizza Hut 173 95 154	C8 McDonalds 198 131 158	C9 Subway 158 195 177	C10 Domino's Pizza 289 273 254	C11	C12 Size of Flat (Codec 2 1 2	C13 Price for the Flat (Coded) 3 315 8 229 6 355	C14	C15	
	Ct Ct Layer Thickness 3.048 3.054 3.079 3.047	Samp Price f	elations le 2 N ISC or the Flat (Coded) 12 C3 Marks in Quality Course 65 71 66 79	C4	95% CI fc (0.262, 0.5 C5	r p P-Value 18) 0.007 C6 Burger King 150 197 186 166	C7 Pizza Hut 173 95 154 104	C8 McDonalds 198 131 158 123	C9 Subway 158 195 177 147	C10 Domino's Pizza 289 273 254 282	C11	C12 Size of Flat (Codec 1 2 2	0 Price for the Flat (Coded) 3 0 76 ef or the Flat (Coded) 5 0 209 6 355 0 261	C14	CIS	
	Ci Ci Ci Layer Thickness 3.048 3.054 3.079 3.047 3.067	Samp Price 1 C2	elations le 2 N jjCc or the Flat(Coded) 12 Marks in Quality Course 65 71 66 79 65 65	C4	95% CI fc (0.262, 0.5 CS	r p P-Value 18) 0.007 C6 Burger King 150 197 186 166 196	C7 Pizza Hut 173 95 154 104 183	C8 McDonalds 198 131 158 123 139	C9 Subway 158 195 177 147 181	C10 Domino's Pizza 289 273 273 254 282 235	C11	C12 Size of Flat (Codec 1 2 2 2 2 2	C13 0 Price for the Flat (Coded) 3 315 5 3259 5 3259 2 2 24 2 24	C14	C15	
	Cit Cit Size of Flat (Coded) 3.048 3.054 3.054 3.056 3.067 3.065 3.065	C2	elations te 2 N (cc r the Fue (cosed) 12 Marks in Quality Course 65 71 66 79 65 82	C4	95% CI fc (0.262, 0.5 C5	r p P-Value 18) 0.007 C6 Burger King 150 197 186 166 196 192	C7 Pizza Hut 173 95 154 104 183 120	C8 McDonalds 131 158 123 139 160	C9 Subway 158 195 177 147 181 192	C10 Domino's Pizza 209 203 224 282 282 282 282 282 283 283 284	C11	C12 Size of Flat (Codec 1 2 2 2 2 1	C3 Price for the Flat (Coded) 5 225 5 255 2 2224 4 216	CI4	CIS	
4	2airwise Pearsc Sample 1 С1 Layer Thickness 3.048 3.054 3.059 3.047 3.067 3.065 3.065	Samp Samp Price t C2	elations le 2 N 100 or the Plat Coded) 12 Marks in Quality Course 63 71 66 79 65 82 82	C4	95% CI fc (0.262, 0.5 C5	r p P-Value 18) 0.007 C6 Burger King 150 197 186 166 196 192	C7 Pizza Hut 173 95 154 104 183 120	C8 McDonalds 198 131 158 123 139 160	C9 Subway 158 195 177 147 181 192 4	C10 Domin's Fizza 289 273 254 282 282 285 285 285 285 285 285	C11	C12 Size of Flat (Codec 1 2 2 2 2 1	C13 Price for the Flat (Caded) Price for the Flat (Caded) 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	CI4	CIS	
	2airwise Pearsc Sample 1 С1 Layer Thickness 3.048 3.054 3.047 3.047 3.047 3.047 3.045 2.048 3.048 3.048 3.049 3.047 3.045	C2	elations e 2 N jcc ar the Flat (code) 12 Marks in Quality Course 65 71 66 79 65 82 ww	C4	95% CI fc (0.262, 0.5 C5	r p P-Value 18) 0.007 Burger King 150 197 186 166 166 196 192	C7 Pizza Hut 173 95 154 104 183 120	C8 McDonalds 198 131 158 123 139 160	C9 Subway 158 195 177 147 181 192 4	C10 Domino's Pizza 289 273 284 282 285 245	C11	C12 Size of Flat (Codec 2 2 2 2 2 2 1	C13 Price for the Flat (Codel), 3 3 3 4 2 2 2 2 2 2 2 4 2 7	C14	CIS	

And so, this is the correlation coefficient that you see over here. So, this is the Pearson correlation coefficient that you see over here ok.

(Refer Slide Time: 32:01)

0	Edit Data Cal	c Stat	Graph View Help Assists C I A A A O O II I I I I I I I I I I I I I I I I I I	nt Addition fx 2= -± @ k_ ★ % (≥ ■ @	al Tools . 2 2 -4 2 # % 2 * %	€ ■ # 哉	₩ ≿∎ øk	* ⁴ } M								
CO	rrelation: Price for	the F	Y X													
=	VISUAL DATA MWX	ico foi	the Elat (Coded)	ize of Ela	t (Coded)											•
-	or opinize Parce	correlation	in the riat (coded), a	ize of Fia	it (coded)	Covariance			×							
	Correlations Size of Flat (Coded)	Price the (Col	e for Flat ded) .727			C1 Layer C3 Marks C6 Burgt C7 Pizza C8 McDo C9 Subw C10 Dom C12 Size C13 Price	Thickness Variables in Quality Size of er King for the hut nalds ay no's Pizza of Flat (Coc for the Fla	: Flat (Coded)' 'Pric Flat (Coded)'	20 0							
	Pairwise Pearso Sample 1 Size of Flat (Coded)	Samp Price f	elations le 2 N Co for the Flat (Coded) 12	0.727 (0	5% Ci for p P 262, 0.918)	- 	□ Store	matrix (display n	othing)							*
•	Pairwise Pearso Sample 1 Size of Flat (Coded) C1	Samp Price f	elations le 2 N Cr for the Flat (Coded) 12 C3	0.727 (0 C4	CI for p P 262, 0.918) C5 C	6 Help	et Ot	matrix (display n	ncel	C11	C12		C13	C14	C15	
*	Pairwise Pearso Sample 1 Size of Flat (Coded) C1 Layer Thickness	Samp Price f	elations le 2 N Co for the Flat (Coded) 12	0.727 (0 C4	6% CI for p P 262, 0.918) C5 C Burge	6 Help r King Pizza	C Store	matrix (display n	ncel mino's Pizza	C11	C12 Size of Flat	(Coded) Pric	C13 e for the Flat (Coded)	C14	C15	
•	Pairwise Pearso Sample 1 Size of Flat (Coded) C1 Layer Thickness 3.048	Samp Price t	elations le 2 N Cr or the Flat (Coded) 12 C3 Marks in Quality Course 65	0.727 (0 C4	CS C GUIDE CS C	6 Help r King Pizza	Line to the store of the store	Car Subway Dor 158	ncel mino's Pizza 289	C11	C12 Size of Flat	(Coded) Pric	C13 e for the Flat (Coded) 315	C14	C15	v
+ 1 2	Pairwise Pearso Sample 1 Size of Flat (Coded) C1 Layer Thickness 3.048 3.054	Samp Price f	elations le 2 N Cr for the Flat (Codeo) 12 C3 Marks in Quality Course 65 71	0.727 (0 C4	C5 C G Burge	6 Help r King Pizza 150 197	← Store → OW Hut McDonalds 173 198 95 131	subway Dor 158 195	noel mino's Pizza 289 273	C11	C12 Size of Flat	(Coded) Price 23 18	C13 e for the Flat (Coded) 315 229	C14	C15	
* 1 2 3	Pairwise Pearso Sample 1 Size of Flat (Coded) C1 Layer Thickness B.048 B.054 B.054 B.054	Samp Price 1 C2	elations le 2 N Cr for the Flat (Coded) 12 C3 Marks in Quality Course 65 71 66	0.727 (0 C4	5% CI for p P 262, 0.918) C5 C Burge	6 Help 150 197 186	← Store → → → → → → → → → → → → → → → → → → →	Subway Dor 158 195 177	noel mino's Pizza 289 273 254	C11	C12 Size of Flat	(Coded) Price 23 18 26	C13 e for the Flat (Coded) 315 229 355	C14	CIS	¥
* 1 2 3 4	Pairwise Pearso Sample 1 Site of Flat (Coded) C1 Layer Thickness 3.048 3.054 3.079 3.047	Samp Price f	elations le 2 N Cr pr the Flat (Coded) 12 C3 Marks in Quality Course 65 71 66 71 66 79	C4	6% CI for p P 262, 0.918) C5 C Burge	6 <u>Help</u> King Pizza 150 197 186 166	Mut McDonalds 173 198 95 131 154 158 104 123	Car Subway Dor 158 195 177 147	nothing) mino's Pizza 289 273 254 282	C11	C12 Size of Flat	(Coded) Prio 23 18 26 20	C13 e for the Flat (Coded) 315 229 355 261	C14	CIS	
+ 1 2 3 4 5	Pairwise Pearso Sample 1 Size of Flat (Coded) C1 Layer Thickness 3.048 3.054 3.054 3.079 3.047 3.067	Samp Price t	elations le 2 N C or the Flat[Coded] 12 Marks in Quality Course 65 71 66 79 65 79	C4	5% CI for p P 262, 0.918) CS C Burge	6 Help 6 Help 7 King Pizza 150 197 186 166 196	Mut McDonalds 173 198 95 131 154 158 104 123 183 139	Car Car Subway Dor 158 195 177 147 181	ncel mino's Pizza 289 273 254 282 282 235	C11	C12 Size of Flat	(Coded) Pric 23 18 26 20 22	C13 8 for the Flat (Coded) 315 229 355 261 224	C14	CIS	

(Refer Slide Time: 32:06)

		12 ×	(20	ek i		Calculator	d 59)
ovariances: Size of VISUAL DATA MWX OVARIANCES:	Flat (C • >	t (Coded), Price t	for the	Flat (C	oded)	= standar	72154	1/(252.2399 • 55519	651125888) × 191 =	History Memory 1/(252.2399651125888) × 191 0.7572154551906394
Size of Flat (Co Price for the Fl	Size of Flat (27.1) 191.0	Co Price for the FI 15 100 2545.4773				МС	мя. М+	M·	MS	v(2545) × 5 252.239965112588
					l⊋	%	CE	с	٩	
						½x	x ²	$\sqrt[3]{x}$	÷	
	0	0			~	7	8	9	×	
Layer Thickness 3.048	N	tarks in Quality Course	64	6	Burger King 150	4	5	6	-	
3.034 3.079 3.047		66 79			197 186 166	1	2	3	+	
3.067		65			196					

So, variance also what we can see is that basic stat we can also see the covariance of this we can calculate. So, in this case, covariance of this will be given like when you do that one. So, this is the covariance matrix what you see over here. So, in this case, so covariance what you can do is that price and size covariance is 191 basically. So, what I can do is that, I can open a calculator also.

And I can just see the calculations whether it is rightly calculating whether MINITAB is doing it correctly. So, 191 has to be divided by standard deviation. So, size and size this is 27 approximately. This is 2545. So, square root of these two values when you multiply, and then make an inverse, and then multiply with 191 you should get the correlation coefficient.

So, I am making a square root of 27 let us say, and when you do that one around 5, I can assume 2545, I will take a square root of that and it is about 50. So, 50 multiplied by 5, 250; 50 multiplied by I am doing 5 let us say and 250. And make a inverse of this and then multiply it with 191 that is the information that we have. And what we get is 0.75 approximation over here. So, in this case, and what we got earlier was a 0.75. So, earlier what we have got correlation coefficient.

So, in this case, what we have gone is that approximately 0.72. So, some calculation error is coming because of square root and all, but it is perfect. So, 727, if we go by exact

calculation method, so it will be same, value will be 0.7527. So, that is possible, so that is possible. So, we will ignore the last part of this. So, we will consider that one.

So, what is important over here strength of the relationship. So, this is 0.72, what we have seen over here what we have seen. So, graphically what we have seen and positive, so 0.727 that is positive. And generally more than 0.7 can be considered as a strong relationship. And this is on the positive side because this is plus r equals to plus 0.727, so that means, I have a somewhat strength of the relationship is quite high, linear relationship is quite high over here ok.

So, that can be also seen over here. So, this type of analysis how it will help is that it will try to identify. So, when I am defining a variable between relationship between two variables over here. So, y and x over here, and in that case whether they are related or not that will help me to select the x which will be used in design of experiments. And experiment on those things to develop the functional relationship y and x.

So, this can be one of the x that I am trying to see the scatter plot because there is a 2D plot and if you have more variables, we can add more variables over here 3D up to 3D plots are possible ok that can be visualized basically more than 3 dimensions cannot be visualized so in that case.

And it gives you some common sense that this is a potential variable which can be considered when I am modelling that one y with x or which variable can be considered in experimentation to understand when I change the x, it will have impact on y. So, we do not select arbitrarily we select those x, and try to enumerate those x at different levels and try to experiment on using statistical means ok. So, to understand the relationship between y and x.

So, selection can be initial selection can be based on correlation information. So, with this illustration, so we have we will stop over here for this session. So, what we have done is that we have shown you histogram where the data can be plotted and CTQ and we can see the variance of the data and we can also see the central tendency of the data. Then with MINITAB and that is quite convenient.

Then we have a box plot whether straight other data like spring, I can see that concentration of the data is more between these values and this one is first quartile and third quartile. Interquartile range can also be we can think about the spread of the data ok. Box plot helps you, it can be used for multiple scenario comparison so which is the best option like restaurants what we have taken. So, that is one important plot.

So, in quality I am demonstrating three important plots in this session. So, one is histogram, one is box plot, and the last one what we have demonstrated is relationship between variables because this will help me to select the X variables or control variables which impacts Y basically. So, whether there is any relationship between X and Y, and I am showing linear relationship.

So, non-linear relationship also can exist. And in that case those variables can be selected for experimentation, and finding out the mapping relationship between Y and X which can be used for optimization also. So, this is the overall idea. And we will stop over here. And see you in session 6, and where we will discuss more about quality visualization of data and which are preferably used in quality control and improvement.

Thank you.