

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

Lecture - 15
Process Capability for Attribute data

Hello and welcome to session 15 of our course on Quality Control and Improvement using MINITAB. So, I am Professor Indrajit Mukherjee from Shailesh J. Mehta School of Management, IIT Bombay. So, last time what we are doing is that we are discussing about Process Capabilities and in case it is nonnormal what is to be done.

So, some data sets we are just exploring and trying to see and I told also there are two types of transformation that is available one is Box-Cox transformation which is a lambda transformation, optimal lambda transformation that we are doing.

So, if the CTQ is y represented as Y^{λ} , what should be the lambda values and that will transform the data into normality. And then, we can analyze process capability like that ok and the specification will also be transferred like that.




(Refer Slide Time: 01:14)

Quality Control and Improvement using MINITAB

The **assumption of normality** is critical to the usual interpretation of the **Capability Analysis**.

For **non-normal data**, option is

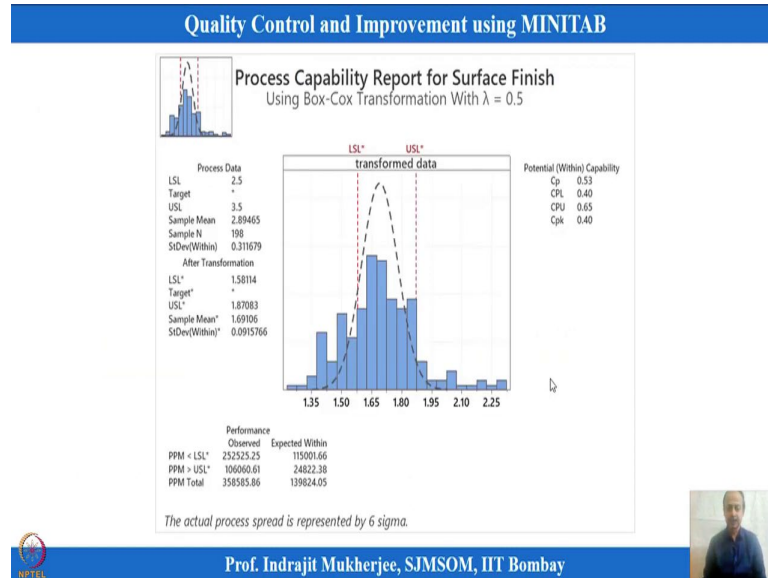
- Transform non-normal data (**Flatness, Ovality, Surface Finish, Delivery Time, Cycle times**) to normal by checking following transformation: **square root (\sqrt{Y})**, **inverse ($1/Y$)**, **log (Y)**
- Use **Box-Cox (λ) Power Transformation** for (+) ve response variables with subgroups.
- **Johnson's Transformation** for any type of response variables (+ ve or - ve)
- We must also **transform the specification(s)**

Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

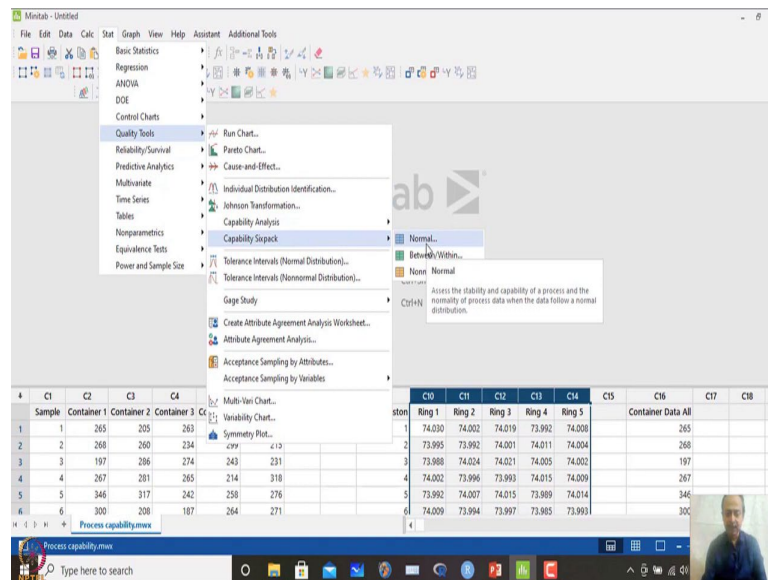
So, some examples we have taken and we will try to extend that lecture to some extent and try to see that how for the previous data what we have taken. How this

transformation works in capability analysis, when there is subgroup size is involved into that ok.

(Refer Slide Time: 01:25)

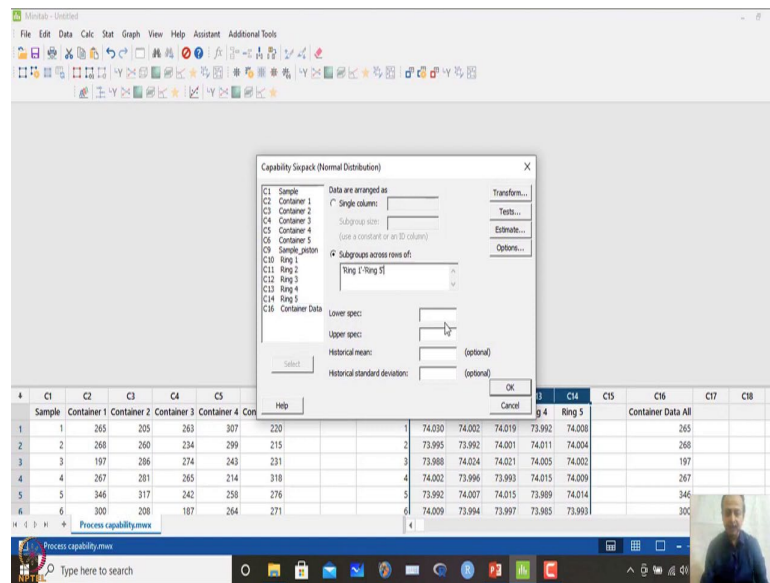


(Refer Slide Time: 01:30)



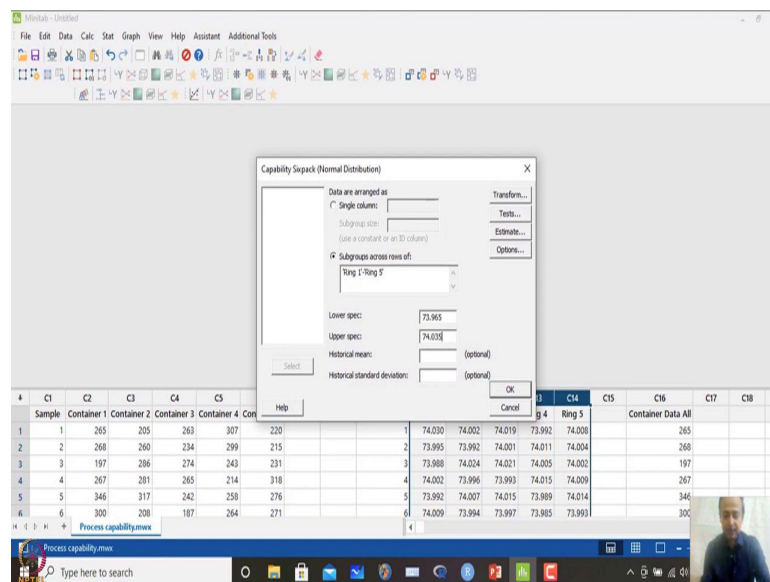
So, this is what we have done so, I will go to the examples like that. So, I will go to the examples and so this was the ring data that we are using samples so, these are the subgroups that we are having. So, what we did is that we went to stat → quality tools → capability analysis, and there is a capability analysis Sixpack also over here which we can use in case transformation is needed.

(Refer Slide Time: 01:53)



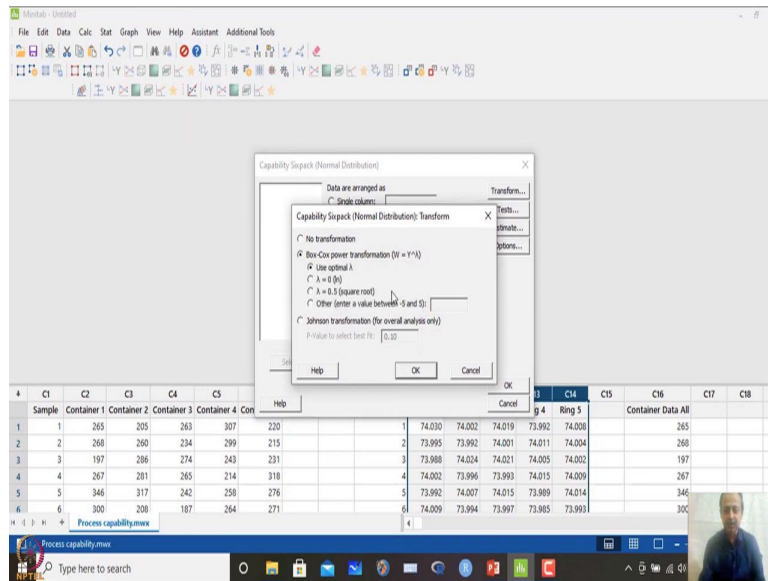
So, I will go to normal over here and then in the subgroups. I will mention that these are the subgroup information's that we are having.

(Refer Slide Time: 02:02)

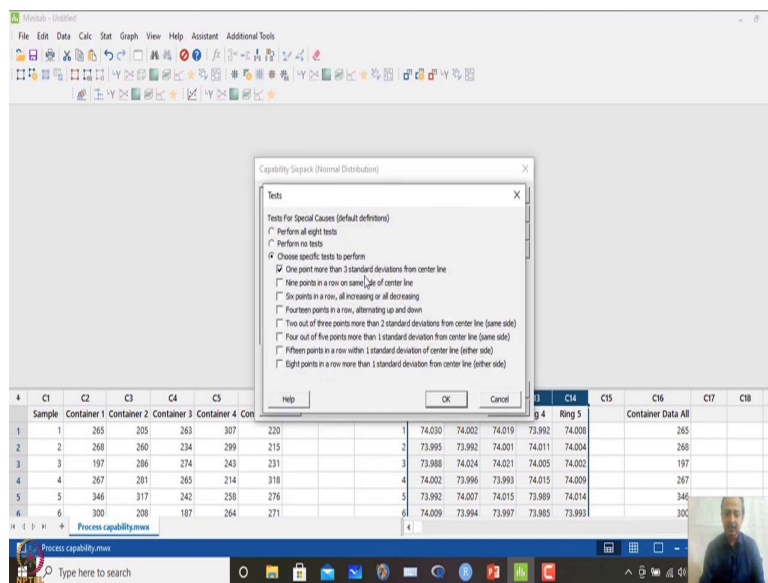


And let us select this one and lower specification what last time also we have seen is 73.965 and the upper specification was 74.035.

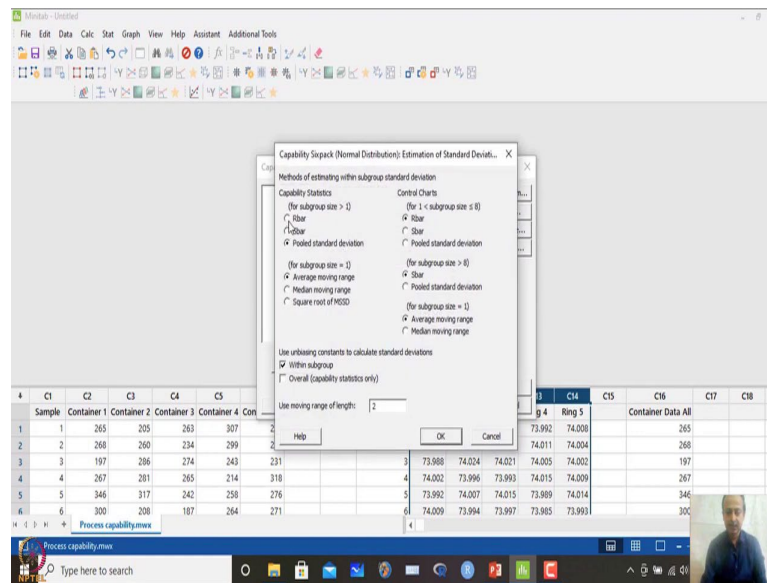
(Refer Slide Time: 02:11)



(Refer Slide Time: 02:19)

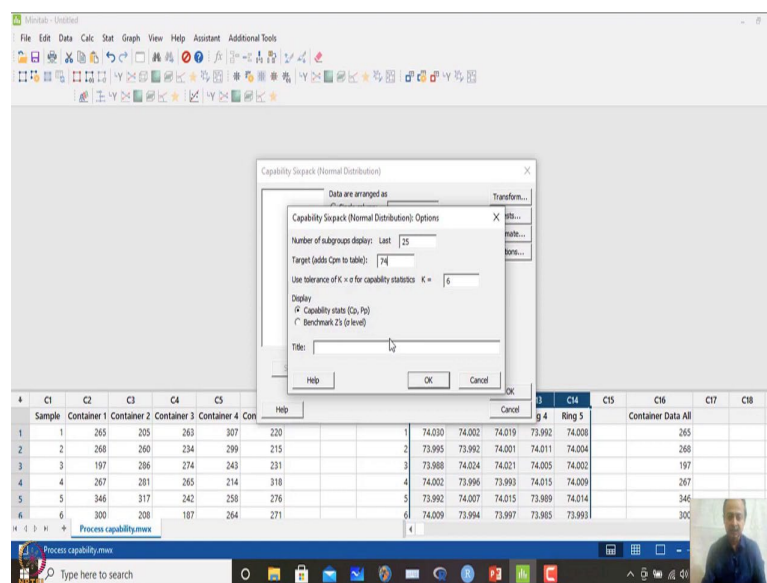


(Refer Slide Time: 02:29)



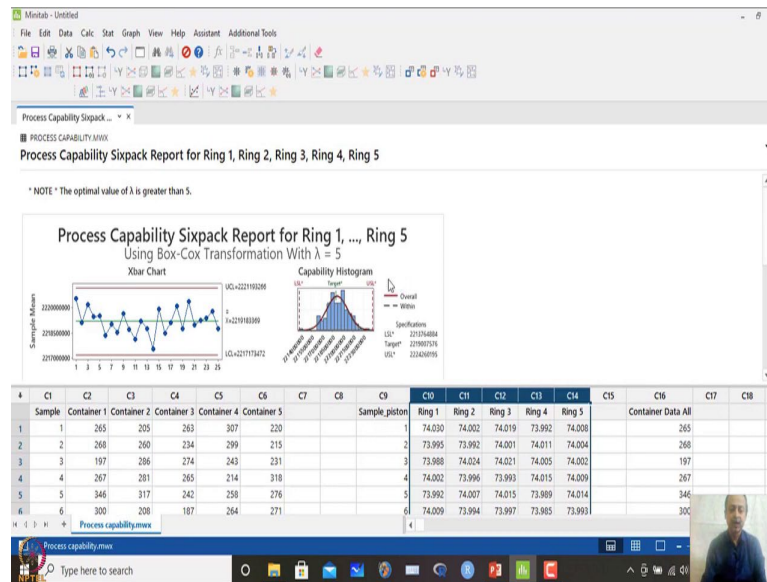
And then we went to transformation and we said that, in case transformation is required use Box-Cox transformation over here and I click OK. And then in test what I do one point going outside that will be the control that we are using over here stability criteria and then we write that if it is more than 1 subgroup. So, \bar{R}/d_2 I concept to calculate the standard deviation and based on that we will calculate capabilities short term ok.

(Refer Slide Time: 02:44)

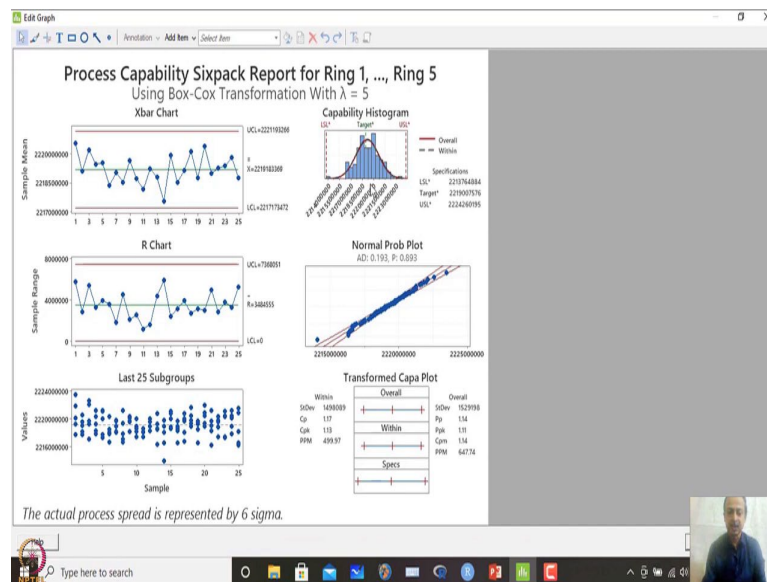


So, in options we have not mentioned any targets over here we can write 74 as the target like that ok, capability and benchmark what is the sigma labels that we have discussed last time. So, in case you want to check the sigma level you can do that.

(Refer Slide Time: 02:58)



(Refer Slide Time: 03:04)



So, if I click OK over here what happens is that this analysis comes up and what you see over here is that capability analysis over here and USL and LSL are transformed over here. Target is also transformed, the values you can see are different because, the MINITAB software is used a lambda transformation of 5 and after lambda

transformation what happens is that you can see this Anderson darling test over here and P value is more than 0.05.

That means, it is adheres to the normality assumption over here. And Xbar charts are shown over here on transform data and range is also shown over here on the transform data like that. And this is the overall capability analysis and Cp calculation what you can see is 1.17 and that it is saying Cpk is about 1.13. So Cp value is not changing much so, I think transformation after transformation also the data seems to be more or less and very close to normalities.

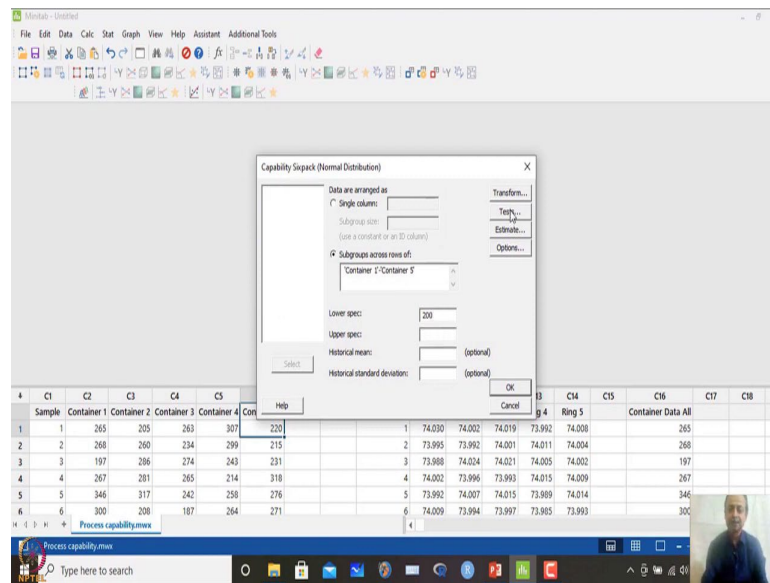
So, in this case we may not have used transformation because, it works well I am telling you that even if some small deviation moderate deviation capability analysis is not much impacted like that ok. But if it is very skewed in that case we have to think like surface finish what we have seen like that and surface finish are generally skew dimensions like that ok.

So, those CTQs needs to be converted and then according to target, we can see the process capability analysis. So, Pp and Ppk, Ppk Cpk values are given over here and those information you can get and PPM levels are also given. So, total PPM within and overall that is also given at the end ok.

So, in the graph, you will get all observations and also the randomness of the observation last 25 subgroups. So, these are the observations and it shows scattered towards and there is no such pattern what is revealed over here. So, in this case, we can consider that this analysis is quite perfect, but Cp , Cpk value indicates that we need to improve. If we have to improve beyond 1.33 so, that is the thing we need to; we need to ensure ok.

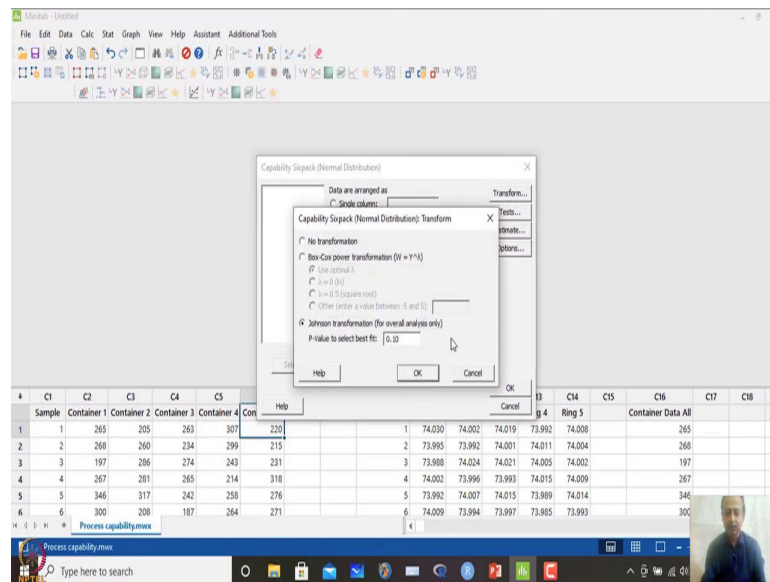
So, let us take another example over here so, we will close this one and we will take the container example where the lower specification is given as 200.

(Refer Slide Time: 05:17)



So, in this case I will use another transformation. So, quality tools maybe capability Sixpack I will go and normality analysis we are doing and container 1, 2,..., 5 over here; and I select those observations and then lower specification is 200. What was mentioned last time also this will be blank.

(Refer Slide Time: 05:28)



(Refer Slide Time: 05:32)

Capability Sixpack (Normal Distribution)

Tests

Tests For Special Causes (default definitions)

- Perform all eight tests
- Perform no tests
- Choose specific tests to perform
- One point more than 3 standard deviations from center line
- Nine points in a row on same side of center line
- Six points in a row, all increasing or all decreasing
- Fourteen points in a row, alternating up and down
- Two out of three points more than 1 standard deviation from center line (same side)
- Four out of five points more than 1 standard deviation from center line (same side)
- Fifteen points in a row within 1 standard deviation of center line (either side)
- Eight points in a row more than 1 standard deviation from center line (either side)

Help OK Cancel

Sample	Container 1	Container 2	Container 3	Container 4	Cont	g 4	C14	C15	C16	C17	C18
1	265	205	263	307	220	1	74.030	74.002	74.019	73.992	74.008
2	268	260	234	299	215	2	73.995	73.992	74.001	74.011	74.004
3	197	286	274	243	231	3	73.988	74.024	74.021	74.005	74.002
4	267	281	265	214	318	4	74.002	73.996	73.993	74.015	74.009
5	346	317	242	258	276	5	73.992	74.007	74.015	73.989	74.014
6	300	208	187	264	271	6	74.009	73.994	73.997	73.985	73.993

(Refer Slide Time: 05:35)

Capability Sixpack (Normal Distribution): Estimation of Standard Deviation

Methods of estimating within subgroup standard deviation

Capability Statistics

- (for subgroup size > 1)
 - Sbar
 - Pooled standard deviation
- (for subgroup size = 1)
 - Average moving range
 - Median moving range
 - Square root of MSD

Control Charts

- (for 1 < subgroup size ≤ 8)
 - Sbar
 - Pooled standard deviation
- (for subgroup size > 8)
 - Sbar
 - Pooled standard deviation
- (for subgroup size = 1)
 - Average moving range
 - Median moving range

Use unbiasing constants to calculate standard deviations

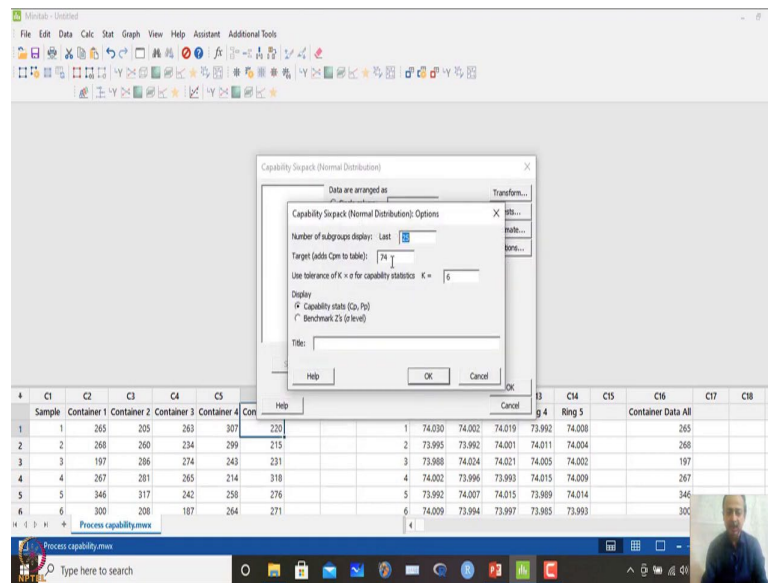
- Within subgroup
- Overall (capability statistics only)

Use moving range of length: 2

Help OK Cancel

Sample	Container 1	Container 2	Container 3	Container 4	Cont	g 4	C14	C15	C16	C17	C18
1	265	205	263	307	220	1	74.030	74.002	74.019	73.992	74.008
2	268	260	234	299	215	2	73.995	73.992	74.001	74.011	74.004
3	197	286	274	243	231	3	73.988	74.024	74.021	74.005	74.002
4	267	281	265	214	318	4	74.002	73.996	73.993	74.015	74.009
5	346	317	242	258	276	5	73.992	74.007	74.015	73.989	74.014
6	300	208	187	264	271	6	74.009	73.994	73.997	73.985	73.993

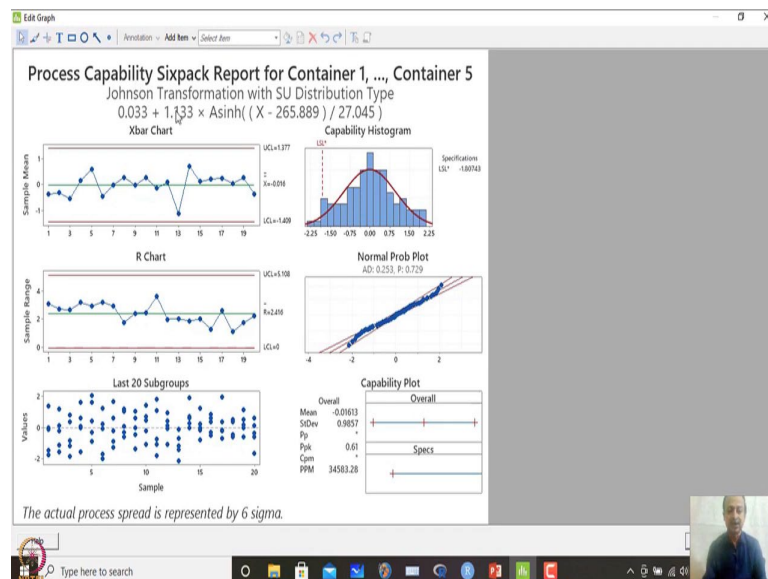
(Refer Slide Time: 05:39)



Transformation let us go for Johnson's transformation, now and click ok and then we test this one point going outside and then estimation over here. The same estimation process we will follow, and in options target value is 74. We can delete and otherwise capability analysis we want to see ok.

One-sided so, Cp values will not come. So, in this case, any also we have changed this data set we have given options for transformation as Johnson's transformation over here and we want to see what happens ok.

(Refer Slide Time: 05:57)



So, when you click this diagram over here, what you see is that. It has used its family of transformation. So, in this case, what happens is that this is the transformation that is done on the Y characteristics you can see X.

So, X will be replaced let us say by Y you can think about the data that is given is Y information or CTQs generally we expressed like effect that is the CTQ so, final outcomes that is the CTQ and that is Y. So, a MINITAB says that we have transformed with a symbol X like that the data set is X over here ok.

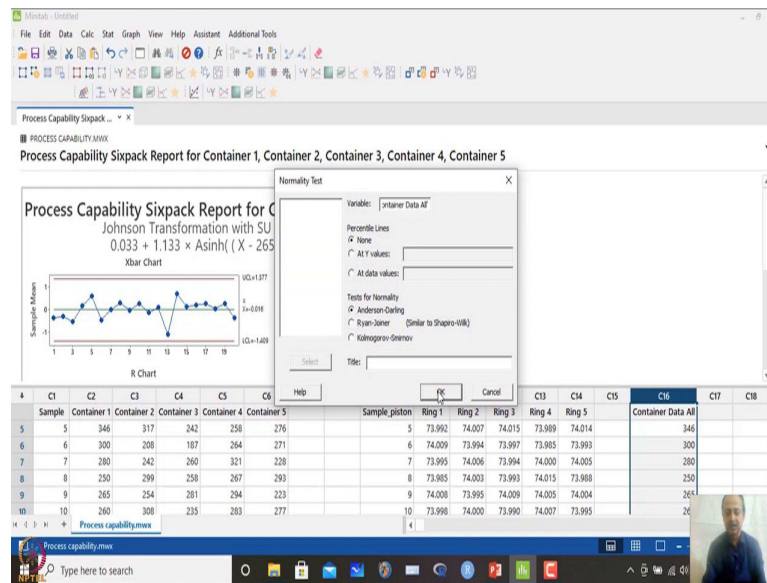
So, then what you see is that, after transformation of the data set, so, y is transformed to $0.033 + 1.333 * \text{Asinh}((X - 265.889) / 27.045)$ and this is the family of transformation that is used to calculate the parameters over here parameters are calculated, and then the final outcomes what you see. So, data was transformed, and why it has done transformation because the data does not follow normal distribution.

So, how do we ensure that one that we will see and after transformation what happens the P-value over here, what do you see Anderson darling test P-value is more than 0.05 and I told that if it is more than 0.5. We have not explained on that aspect we will do next time in our subsequent lectures and we will try to see what is the P value, but at present, we assume that a P is more than 0.05, in that case, it adheres to normality like that.

So, the data set was plotted and P-value was calculated as 0.729 and that adheres to our normality assumptions. So, there is no problem with normality after transformation like that, and the index was calculated over here Ppk index was calculated as 0.61 like that ok.

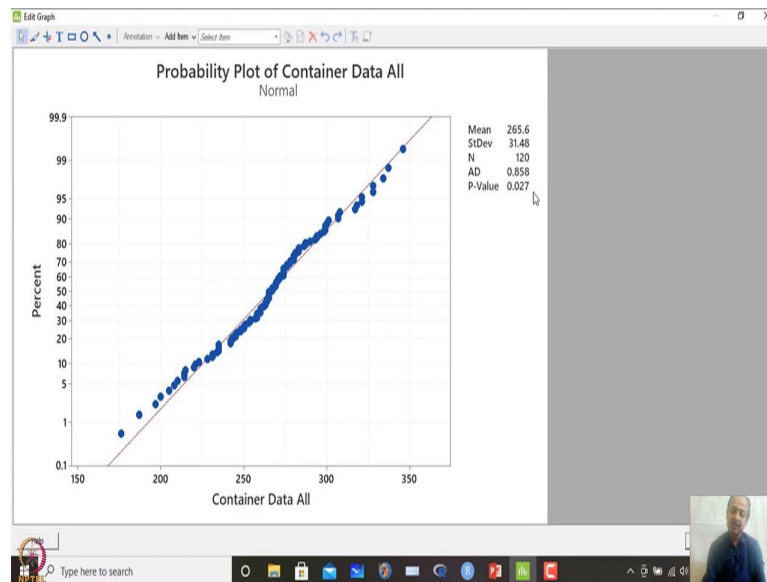
So, this is already given over here so, this is what I wanted to show over here and this is the transform data on which the analysis was done. So, I will click this one.

(Refer Slide Time: 07:53)



Now, this data set a is in subgroup so, how to ensure that this is normality. So, what I did is that I took all the data set into one column C16 and then I want to went to stat basic stat and I wanted to check normality assumption over here, I gave container data all and in this case, I wanted to do Anderson darling test and I click OK.

(Refer Slide Time: 08:12)



And what I found is that P-value that I am getting is less than 0.05 so, less than 0.05 and that indicates that data is nonnormal scenario exist over here. So, nonnormality is there that is why Box-Cox or Johnson's transformation is used. But, both the transformation

will work everywhere it does not you cannot ensure that one, sometimes it works and sometimes both can fail also so, that is also possible like that.

So, in case everything fails then we have to see some other ways to calculate nonnormal process capabilities. So, there are fitting distribution, and based on that you will get options in MINITAB to do that ok. So, this is the data set that we are talking about over here.

(Refer Slide Time: 08:58)

Quality Control and Improvement using MINITAB







Process Capability from Attribute Data (Attribute Data)

- Process capability for count data is assessed differently from continuous data
- Summary of process capability using count data
 - DPU, DPMO and Sigma Level metrics
 - Ability to meet customer requirement
- Process must be stable
- **Attribute Capability Example**
- Customer spec. is sigma quality level ≥ 4
- Last month 16,810 units were produced over 18 days
- 231 were defective
- **Did we meet customer requirements?**

When dealing with attribute data, **defects per unit (DPU)** is used as a statistic to measure the capability

$$DPU = \frac{\text{Total number of defects}}{\text{Total number of units produced}}$$

$$DPU = \frac{231}{16810} = 0.0137418$$


Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay






And then go to the slides what we wanted to cover next. So, this is about process capability. Now, in the case of attribute data we have defects or defectives like that so, in this case, how do we calculate the sigma levels or process capability that way is there any conversion that we can do ok.

So, for continuous data, we have seen that earlier method works well, and in case of process using count data that is defects over here. Some other units are used over here defects per units, defects per million opportunity, and then based on that we can signify we can tell what is the sigma level like that.

Although, normal distribution assumption is taken to convert this one so, but basic assumption is that we can make a sigma level transformation from the defects per units, defect per opportunities. And from there we can calculate the sigma; we can go back to

sigma level and based on that we can also calculate process capability like that. So, there is an interrelationship between all these matrices that we are using ok.

But we are assuming over here process stability that is the basic assumptions that we are adhering and customer let us say specification given as sigma level should be greater than 4 or something like that in this particular case.

But last month what happened is I have manufactured 16,000 approximately units and over a period of 18 days. And I want to check what is the sigma level of the process; and we found out that 231 defectives were reported like that or we can say defects or something like that so, you have to define.

So, let us say it is defects over here, 231 defects because Six Sigma methodology nowadays connects everything to defects, not defectives. So, in this case, let us assume this is defective so, 231 defects over here reported out of 16,000 units like that. So, do we meet the specification can we check the sigma levels of this ok and then accordingly we will take the actions like that.

So, then the definition is that how do we do that we calculate defects per unit,ok. So, 16,000 units and 231 was defects let us say. So, defects per unit will be 231/16810, that is around 0.0137 like that ok. You can go up to two place of decimal or three-place of decimal like the defects per unit so, how many defects per unit like that ok.

(Refer Slide Time: 11:18)

Quality Control and Improvement using MINITAB


The DPU measure **does not** directly take the complexity of the unit into account. A widely used way to do this is the **defect per million opportunities (DPMO) measure**

DPO=DPU/Opportunity in 1 UNIT


$$DPMO = \frac{\text{Total number of defects}}{\text{Total number of units produced} \times \text{Number of opportunities}} \times 1,000,000$$

$$DPMO = \frac{231}{16810} \times 1,000,000 = 13741.8$$

Sigma Level = (NORMSINV(1-DPO))+1.5



Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay



When we have done defects per unit then we can convert into defects per unit per opportunity in one units like that so, one important terminology that comes into over here is opportunity ok. So, this term opportunity is explained in any Six Sigma methodology course like that. So, I will only tell that this opportunity has to be defined, and let us assume the opportunity in one over here.

So, in this case what we are saying is that, the defects per opportunity remain same so, defects per unit is about same fraction and then divided by 1 also. Defects per opportunity also comes out to be same ok. So, then defects per million opportunities so in 1 unit one opportunity this much so, in million what we 10 to the power 6 multiplied by 10 to the 6 simple mathematics over here.

So, 10 to the power 6 multiplication over here defects per million opportunity this comes around 13741 like that ok. And now, how to convert that into sigma levels what we have in excel this conversion formulas like that, norms inverse.


And 1 minus 1 minus DPU that values that we have got and plus 1.5 that is added over here; and that will give you the short term sigma levels like that. So, that will give you the short term sigma level, this function is available in excel and you can do that in excel only ok.

(Refer Slide Time: 12:45)


Quality Control and Improvement using MINTAB

Sigma	DPMO	Sigma	DPMO	Sigma	DPMO	Sigma	DPMO	Sigma	DPMO	Sigma	DPMO	Sigma	DPMO	Sigma	DPMO	Sigma	DPMO
2.01	305026	2.26	223627	2.51	156248	2.76	103835	3.01	65522	3.26	39204	3.51	22216	3.76	11911		
2.02	301532	2.27	220650	2.52	153864	2.77	102042	3.02	64256	3.27	38364	3.52	21692	3.77	11604		
2.03	298056	2.28	217695	2.53	151505	2.78	100273	3.03	63008	3.28	37538	3.53	21178	3.78	11304		
2.04	294598	2.29	214764	2.54	149170	2.79	98525	3.04	61780	3.29	36727	3.54	20675	3.79	11011		
2.05	291160	2.30	211855	2.55	146859	2.80	96801	3.05	60571	3.30	35930	3.55	20182	3.80	10724		
2.06	287740	2.31	208970	2.56	144572	2.81	95098	3.06	59380	3.31	35148	3.56	19699	3.81	10444		
2.07	284339	2.32	206108	2.57	142310	2.82	93418	3.07	58208	3.32	34379	3.57	19226	3.82	10170		
2.08	280957	2.33	203269	2.58	140071	2.83	91759	3.08	57053	3.33	33625	3.58	18763	3.83	9903		
2.09	277595	2.34	200454	2.59	137857	2.84	90123	3.09	55917	3.34	32884	3.59	18309	3.84	9642		
2.10	274253	2.35	197662	2.60	135666	2.85	88508	3.10	54799	3.35	32157	3.60	17864	3.85	9387		
2.11	270931	2.36	194894	2.61	133500	2.86	86915	3.11	53699	3.36	31443	3.61	17429	3.86	9137		
2.12	267629	2.37	192150	2.62	131357	2.87	85344	3.12	52616	3.37	30742	3.62	17003	3.87	8894		
2.13	264347	2.38	189430	2.63	129238	2.88	83793	3.13	51551	3.38	30054	3.63	16586	3.88	8656		
2.14	261086	2.39	186733	2.64	127143	2.89	82264	3.14	50503	3.39	29379	3.64	16177	3.89	8424		
2.15	257846	2.40	184060	2.65	125072	2.90	80757	3.15	49471	3.40	28716	3.65	15778	3.90	8198		
2.16	254627	2.41	181411	2.66	123024	2.91	79270	3.16	48457	3.41	28067	3.66	15386	3.91	7976		
2.17	251429	2.42	178786	2.67	121001	2.92	77804	3.17	47460	3.42	27429	3.67	15003	3.92	7760		
2.18	248252	2.43	176186	2.68	119000	2.93	76359	3.18	46479	3.43	26803	3.68	14629	3.93	7549		
2.19	245097	2.44	173609	2.69	117023	2.94	74934	3.19	45514	3.44	26190	3.69	14262	3.94	7344		
2.20	241964	2.45	171056	2.70	115070	2.95	73529	3.20	44565	3.45	25588	3.70	13903	3.95	7143		
2.21	238852	2.46	168528	2.71	113140	2.96	72145	3.21	43633	3.46	24998	3.71	13553	3.96	6947		
2.22	235762	2.47	166023	2.72	111233	2.97	70781	3.22	42716	3.47	24419	3.72	13209	3.97	6756		
2.23	232695	2.48	163543	2.73	109349	2.98	69437	3.23	41815	3.48	23852	3.73	12874	3.98	6569		
2.24	229650	2.49	161087	2.74	107488	2.99	68112	3.24	40929	3.49	23295	3.74	12545	3.99	6387		
2.25	226627	2.50	158655	2.75	105650	3.00	66807	3.25	40059	3.50	22750	3.75	12224	4.00	6210		

3.7 0.8 64



Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay



So, we do not need it, but there is a chart conversion what is also possible over here there is a complete chart where you can found that. If DPMO is this much what is the sigma level like that. So, our case is approximately equals to 13741, so, let us try to see what is the sigma level 13 if 13741 is DPMO over here.

So, if you go over here in this graphically so, what you can see is that this is approximately defects around 13741 approximately over here. So, in between somewhere over here 3.07 and 3.71 so, it is approximately near to 3.705 or something like that we can assume like that.

So, is it less than 4 yes it is less than 4 so, we have not reached the sigma level which was earlier defined that customer specification is saying that sigma level for this specific process should be at least equals to 4. And here what we are seeing is that we have not reached that sigma level over here ok.

So, this way also attribute data we can figure out and using this chart over here and there are many every many of the books will give you this chart of conversion of the sigma level short term sigma with DPMO.

(Refer Slide Time: 14:12)

Quality Control and Improvement using MINTAB

Sigma	DPMO	Sigma	DPMO	Sigma	DPMO	Sigma	DPMO	Sigma	DPMO	Sigma	DPMO	Sigma	DPMO	Sigma	DPMO
4.01	6037	4.26	2890	4.51	1306	4.76	557	5.01	224	5.26	85	5.51	30.4	5.76	10.2
4.02	5868	4.27	2803	4.52	1264	4.77	538	5.02	216	5.27	82	5.52	29.1	5.77	9.8
4.03	5703	4.28	2716	4.53	1223	4.78	519	5.03	208	5.28	78	5.53	27.9	5.78	9.4
4.04	5543	4.29	2635	4.54	1183	4.79	501	5.04	200	5.29	75	5.54	26.7	5.79	8.9
4.05	5386	4.30	2555	4.55	1144	4.80	483	5.05	193	5.30	72	5.55	25.6	5.80	8.5
4.06	5234	4.31	2477	4.56	1107	4.81	467	5.06	185	5.31	70	5.56	24.5	5.81	8.2
4.07	5085	4.32	2401	4.57	1070	4.82	450	5.07	179	5.32	67	5.57	23.5	5.82	7.8
4.08	4940	4.33	2327	4.58	1035	4.83	434	5.08	172	5.33	64	5.58	22.5	5.83	7.5
4.09	4799	4.34	2256	4.59	1001	4.84	419	5.09	165	5.34	62	5.59	21.6	5.84	7.1
4.10	4661	4.35	2186	4.60	968	4.85	404	5.10	159	5.35	59	5.60	20.7	5.85	6.8
4.11	4527	4.36	2118	4.61	936	4.86	390	5.11	153	5.36	57	5.61	19.8	5.86	6.5
4.12	4397	4.37	2052	4.62	904	4.87	376	5.12	147	5.37	54	5.62	19.0	5.87	6.2
4.13	4269	4.38	1988	4.63	874	4.88	362	5.13	142	5.38	52	5.63	18.1	5.88	5.9
4.14	4145	4.39	1926	4.64	845	4.89	350	5.14	136	5.39	50	5.64	17.4	5.89	5.7
4.15	4025	4.40	1866	4.65	816	4.90	337	5.15	131	5.40	48	5.65	16.6	5.90	5.4
4.16	3907	4.41	1807	4.66	789	4.91	325	5.16	126	5.41	46	5.66	15.9	5.91	5.2
4.17	3793	4.42	1750	4.67	762	4.92	313	5.17	121	5.42	44	5.67	15.2	5.92	4.9
4.18	3681	4.43	1695	4.68	736	4.93	302	5.18	117	5.43	42	5.68	14.6	5.93	4.7
4.19	3573	4.44	1641	4.69	711	4.94	291	5.19	112	5.44	41	5.69	14.0	5.94	4.5
4.20	3467	4.45	1589	4.70	687	4.95	280	5.20	108	5.45	39	5.70	13.4	5.95	4.3
4.21	3364	4.46	1538	4.71	664	4.96	270	5.21	104	5.46	37	5.71	12.8	5.96	4.1
4.22	3264	4.47	1489	4.72	641	4.97	260	5.22	100	5.47	36	5.72	12.2	5.97	3.9
4.23	3167	4.48	1441	4.73	619	4.98	251	5.23	96	5.48	34	5.73	11.7	5.98	3.7
4.24	3072	4.49	1395	4.74	598	4.99	242	5.24	92	5.49	33	5.74	11.2	5.99	3.6
4.25	2980	4.50	1350	4.75	577	5.00	233	5.25	88	5.50	32	5.75	10.7	6.00	3.4



Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay



So, if you see the last DPMO over here, which is a Six Sigma process this is around 3.4 that after shifting the distribution this is 3.4 sigma level is 6 over here. So, in this case, it is a Six Sigma process we can say. So, this is a sigma level what we are getting 6, if

defect is 3.4 here it is 3.07 something and that is below 4. So, we have not reached that sigma level yet what is required by the customer like that ok.

So, this is all about process capability what we have discussed, and we will now require some amount of statistics which will help us to deal with experimentation. So, when we are dealing with experimentation some basic idea of statistics is required.

And I will give you some brief on statistics and some of the things that is really important for design of experiments, and those things I will highlight not I will highlight huge statistics over here, but whatever is required at least we should cover over here ok.

So, that I understand basics of hypothesis, although you can do some other courses where basic statistics are covered and I will give a brief introduction to that. And because that will give some idea what we are doing basic next time ok subsequent lectures over here. So, that way we will try to deliver something ok.


(Refer Slide Time: 15:29)

Quality Control and Improvement using MINITAB

Statistics is the **Science of Data**.

```
graph TD; A([Statistics]) --> B([Descriptive]); A --> C([Inferential]);
```

Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay



So, what is statistics we want to understand because everywhere there is stat you see quality over here. So, and everywhere one of the values I have mentioned over here that means, basic statistics \bar{x} I have mentioned then sigma I have mentioned. So, this is coming what is the estimation of sigma all these things we have seen and now, we have blindly assumed that whatever is told and we are assuming that is the correct thing and we will do like that ok.

But, what is the basis of that what is the foundation of that so, we want to understand that one. So, statistics can be divided into two categories over here and this is all about data science basically ok. So, over here what we are doing is that, we do not have ideas of population so in that case, we need to take samples because a machine is running for many years and many many samples have gone out so, countably although it is finite, but it is countably infinite basically.

So, we do not have all information all data sets over here. So, what we do is that we take samples out of that we take a screenshot and we say process capabilities this much like that ok.

So, how we are making that inference basically. So, what we do is that we can see statistics in two ways one is known as descriptive statistics that means, visualization of data what we have told then summary statistics what we have seen mean standard deviation all these locations.

And that is one area of statistics which helps us to visualize, which helps us to make some initial interpretation about the data; and the second part of that is known as inferential statistics. In inferential statistics what happens is that, we will make conclusions based on the data the availability data availability like that ok. So, decision-making happens in inferential statistics and descriptive statistics shows visualization of the data over here ok.



(Refer Slide Time: 17:20)

Quality Control and Improvement using MINTAB

Types of Data

Data fall into three main groups:

- Interval Scale** Data, e.g. temperature, heights, weights
- Ratio Scale** Data, e.g. sales, number of customers
- Ordinal Scale** Data, College e.g. course rating system (1-5)
- Nominal Scale** Data, marital status, Colour of a Pen

 Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay 

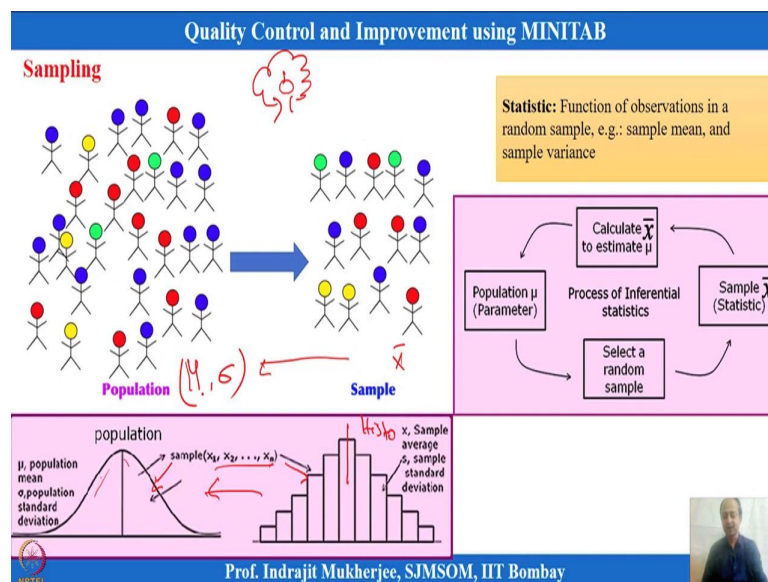
And for that using statistics it depends on what type of data I am having like what we are mentioning attribute data and the continuous data like that so, the category of data needs to be identified so we need to know what are the different types of data categories. So, the highest level of data category is ratio scale data or interval data like that.

So, these are the data's which are having the highest level of quality where we can do many analysis of statistics like that. So, interval data means, this is the data where we have other than 0 numbers can take any infinite numbers in a scale like that ok.

So, ratio scale means it includes 0 over here and then we have ordinal scale where like rating systems when you go do customer service like that. So, in that case rate in a scale of 1-5, 1-7 like that, that is ordinal scale and there is nominal scale like colour or where we cannot arrange the data into orders like that. But we can say that they are different like that although ordering is not possible.

So, ordinal scale ordering is possible interval ratio possible, but nominal scale like marriage, marital status, or male or female like that so, we cannot say which is bigger than which one like that which will come first, which will come second. So, we cannot order that one like colours also we cannot do that ok.

(Refer Slide Time: 18:42)



So, most of the analysis what we do in quality is assuming that CTQ is continuous because, we are assuming normal distribution and for that continuous data assumption is

required or it is interval scale or ratio scale where we are applying this concept of statistics basically ok. Because we have to make inferences after doing experimentation so, after experimentation we need to let us say select which variable is important which is not and which is to be changed like that.

So, that analysis and to make an inference is only possible using statistics like that ok. So, statistics what we are doing basically is that, we do not have idea of the population so, what we are doing is that we are taking samples out of the population like that and based on the sample information we are extending our idea about the population basically ok.

So, we get an estimation from the samples and based on the estimation what we say is that population parameters will be this much like, if I get a mean over here. So, if I get a mean over here I want to extend that one what is the population mean like that.

So, I take only one sample I cannot do it several times like that. So, with one sample I want to; I want to predict what should be the μ values or the population over here I do not have population information over here, like in a tree. If you are; if you are trying to predict what is the apple weights like that and in that case what will happen is that it will take some apples and calculate the weights and then because, this is destructive testing so, I cannot take all apples from the tree then we cannot those is unnecessary over here.

So, we will take some samples, and based on that we will try to estimate about the population over here. So, we will make an estimation about the population. This is what we are doing in inferential statistics we do not have information on population that is why we are taking samples that is why statistics is coming into picture.

And we take some samples and based on that we try to estimate population parameters like that, whether it is standard deviation whether it is mean because mean and standard deviation is the only thing about a CTQ we want to infer about ok.

So, where is the mean, so, where is the standard deviation, how much is the standard deviation like that of the population basically have I controlled when I have taken I have done experimentation let us say simple experimentation one time can I infer it in population so, in population also behaviour of the CTQ will be like that.

So, what we are doing is that we will do experiments with small samples, and based on that we will try to infer that if I keep this setting over here it will work for the population entire population whatever inputs you are giving; and this is the process setting which we can adopt in future like that ok.

So, that is why we are using statistics, we will not do experiments each and every time and then fix the parameters like that. We will freeze it one go and then maybe after certain time points, if we feel that again experimentation is required we will do that ok. So, what we are doing is that we are talking about population estimation over here and from sample information that we are getting over here.

So, some sample information and this is the histogram what do you see so, this is a histogram what we are doing over here and from this information, I want to predict about the population. So, sample information is collected from here, and I want to predict the behaviour of the populations like that.

Population can have normal distribution has two parameters one is μ and one is σ over here so, these are the two parameters that define normal distribution a complicated probability function, which defines the probability density function of normal distribution.

But, what we are doing over here is that assuming that is known and in this case, I take some sample assuming that is normal and then can I predict the population parameters. What should be the μ value, what should be the σ value like that can we do that so, that is what we are doing in statistics ok. So, from sample information making inference.

(Refer Slide Time: 22:31)

Quality Control and Improvement using MINITAB

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	7	10	49	98	66	03	50	34	80	93	44	22	22	45	53	86	23	51		
2	0	36	41	52	30	13	97	46	27	30	27	16	35	34	47	30	36	00		
3	4	23	50	21	03	03	25	79	07	80	54	53	41	12	15	03	08	56		
4	7	2	01	00	33	25	19	16	23	58	03	78	47	43	77	88	15	02	55	67
5	0	06	06	49	47	32	58	08	75	29	63	66	89	09	23	35	97	74	39	80
6	65	76	34	11	33	60	95	03	53	72	06	78	28	14	51	78	76	45	26	45
7	83	76	95	25	70	60	13	32	52	11	87	38	49	03	82	84	99	02	64	00
8	58	80	07	84	20	80	87	93	36	65	10	71	83	93	42	46	34	41	44	01
9	54	74	67	11	15	78	21	96	43	14	11	22	71	17	02	54	51	78	76	76
10	56	81	92	73	40	87	20	05	26	63	97	86	48	51	59	12	46	89	75	64
11	34	89	06	21	22	38	22	32	85	26	37	00	62	27	74	46	02	61	59	81
12	02	26	92	27	85	87	59	18	18	39	93	38	38	73	23	20	19	68	68	90
13	43	84	25	36	00	45	73	80	02	63	33	10	86	72	39	02	00	47	06	96
14	92	56	51	22	11	06	86	81	77	86	59	87	66	13	82	33	97	23	33	41
15	67	42	43	20	60	84	18	68	48	83	00	00	48	35	48	57	63	38	84	

1 - 10

Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

So, how many ways we can do sampling there are different ways of doing sampling when we pick the apples from the tree how we can do that, and you have a huge target population like that and from that how to select that one. There are probabilistic way there are nonprobabilistic sampling also method that means, which does not follow basic randomness in that case like convenience sampling.

And reference sampling like that. Convenience means, based on my convenience I pull any of the people and ask that what is the what is what do you think about the restaurant food or something like that. I am standing just outside and whatever customer is coming and I will do the survey that is convenience because, convenience means I will stand in one restaurant and try to infer about all the restaurant like that.

So, that is not recommended, but that is non probabilistic sampling ok. And there can be referral sampling that means I am trying to understand the philosophy of criminals in that case, I will ask a criminal and he will give me some other reference of criminals like that.

So, that will be our referral sampling that means, I go from one point and then the person I am serving will give me hints who is the next one I can survey like that. So, that is also nonprobabilistic type of sometimes it gives you some lead, but that is not probabilistic sampling. There are different methods of probabilistic sampling; one is simple random sampling, stratified sampling, systematic sampling, cluster sampling.

So, simple random sampling what we do generally when we are doing experimentation and we call it as randomization, we call it as randomization. Systematic sampling when we are doing like in control chart processes what we are doing is systematic sampling at a given time point we visit the process. And that is assumed to be random. So, the first starting point and then, maybe after half an hour or something like that I am going to the process.

So, there is systematic sampling and stratified sampling you understand that age group wise I am doing sampling like that, mostly in survey we use this stratified concept. And cluster sampling like pooling and all these things exit pools or something like that there you can use cluster sampling like that.

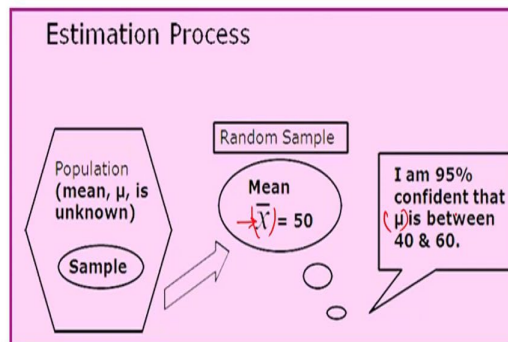
So, most relevant to quality over here, simple random sampling and systematic sampling that is the only thing I can suggest over here which is of importance to us in this course simple random sampling randomization is; and how do we ensure randomization. We ensure randomization by random number generation like that so, samples will be selected based on random number.

So, if you have let us say 10 samples over here, which one to be selected first maybe the second digit of this number, we can take. 7 the next one 0 we will skip then 9 will be taken like that then 8 will be taken like this, either I go row wise or I go column wise or row wise like that. Anyway you go it is basically random sampling that we are doing. So, we select the samples randomly and we do experimentation.

Why we do that, there is a proper reason for that ok why randomization is important like that in experimentation and simple random sampling is like that and systematic means which samples I will select that is random, but what time point I will go so, at a defined interval we visit the process and take the data set like, what we have done in control chart techniques like that ok. So, that is systematic sampling ok.

(Refer Slide Time: 25:46)

Quality Control and Improvement using MINITAB



Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay



So, these are the two types of sampling and what we are doing is that from the samples we are estimating some parameters \bar{X} let us say average values over here. So, can I extend that where is the μ values over here, and for that the concept of confidence interval is used so, that we will discuss in the next session like that.

So, here what we are doing is that, we are taking some samples, we are estimating some means and we are trying to define that what is the population parameters like that. What is the population parameters like that ok. So, up to this point we will try to extend after this lecture after this one we will extend.

So, what we have done is that, in this complete lectures that we have covered what we have said is that process capability analysis can be can have non normal scenarios in that case what can be done is that. We can transform the data set is in normality and then transform the USL, LSL like that.

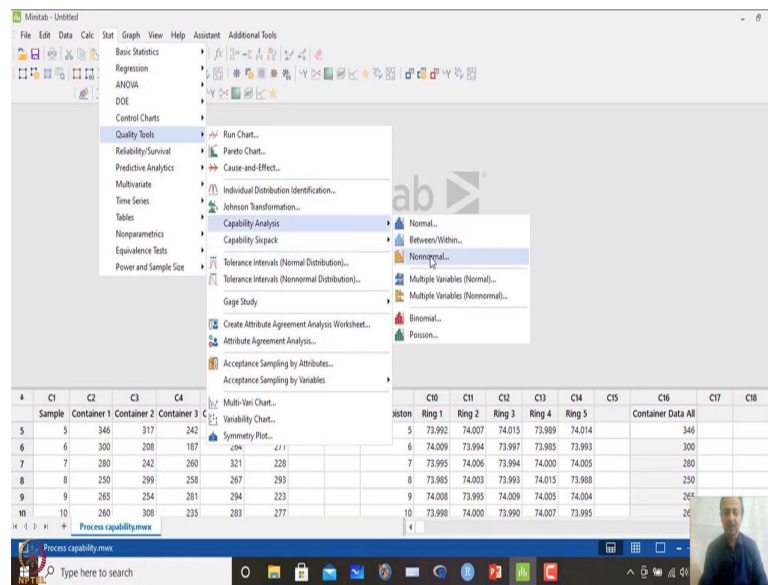
And then what we can do is that we can calculate the capabilities and that is the best way of doing and but if there is a moderate deviation in that case. We can assume that normality assumptions are we can do that and without transformation also we can make inference.

But plotting the data is always helpful and seeing that how much skewed it is ok. If it is not much skewed and we do not want to fit another distribution and try to increase the

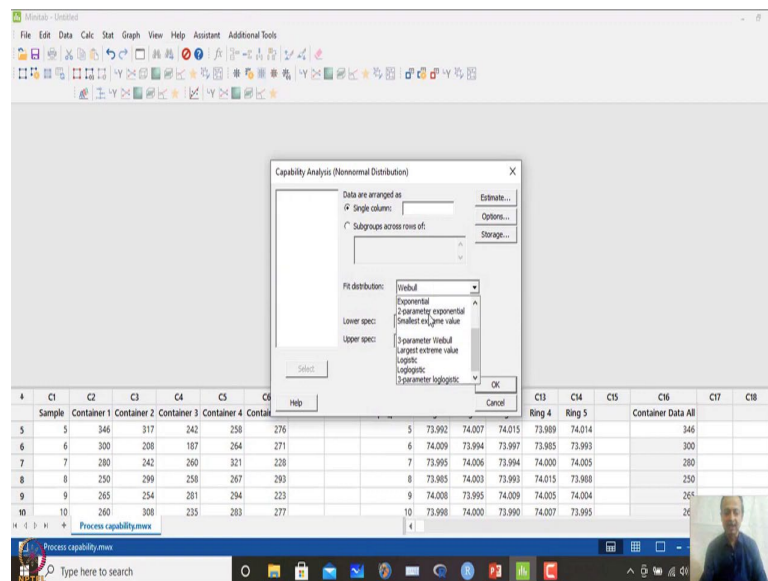
complexity because shock load people will not understand so much complexities like that.

So, what we do is that, in case it is required we will say process capability is this much and in case transformation works we will just show that with this transformation this is the current capability do not do it assuming normality over here. Because, another distribution may be quite correct over here and this is the nonnormal situation we can address that one using MINITAB.

(Refer Slide Time: 27:47)

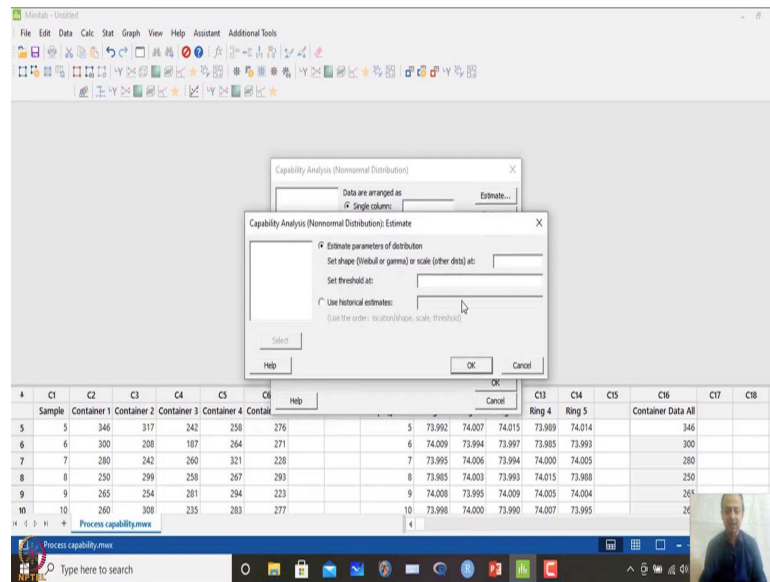


(Refer Slide Time: 27:49)

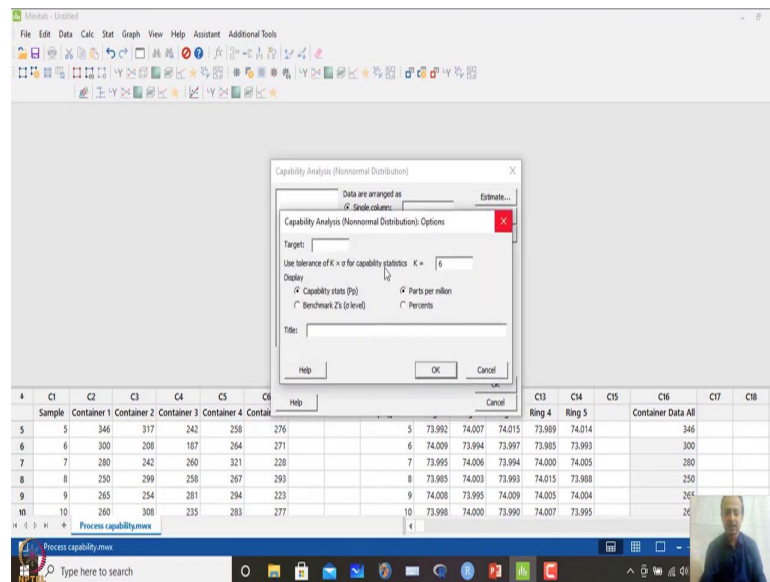


And there is option in Minitab. So, we have options over here to fit distribution, and then based on that we can do quality analysis. So, if I go to capability flag or capability analysis there will be nonnormal scenario, but you have to assign some distribution over here so, distribution has to be fitted.

(Refer Slide Time: 27:55)



(Refer Slide Time: 27:58)



And based on that specification you give and it will use some estimation over here for the parameters and then, options will be given the target same capability benchmark like

that. So, I need to fit distributions like that, if I can fit the distribution perfect distribution what it follows like that and it is q distribution that we are fitting over here.

There are different types of distribution and distributions can be fit and seen which is closest to that in Minitab also it is possible; and based on that what we can do is that then we can do capability analysis like that. So, that was the starting point where and then we went to the discussion that statistics.

Then we talked about attribute data and how to calculate sigma labels like that, in case it is defects in that case how it is done so many other scenarios can happen that this is the assembly operations and what is the sigma level of this. So, everywhere we want to implement the concept of sigma levels of the process like that so, can I define that one? So, I have seen you have seen the dp defects per unit and based on that we can calculate, what is the sigma level of the process ok.

So then, we talked about that whatever we are inferring in quality and all these things is based on statistics and we want to understand some amount of statistics over here. So, that will help us for making inference when we are doing experimentation. Basically, when we are doing experimentation and for the basic idea of hypothesis is required and one of the value that we have used P value we need to understand what is that P-value.

So, that a brief idea of statistics will help us to introduce in some of the topics which are very relevant to quality like regression what will be used and in design of experiments.

So, how to make inference in design of experiment after doing experimentation how to make inference and for that some background is required. So, that background I will highlight, but you can always do some course on statistics and to understand more about statistics like that ok.

So, thank you for listening this lecture, we will stop over here and we will continue from here ok on statistics more about statistics.

Thank you.