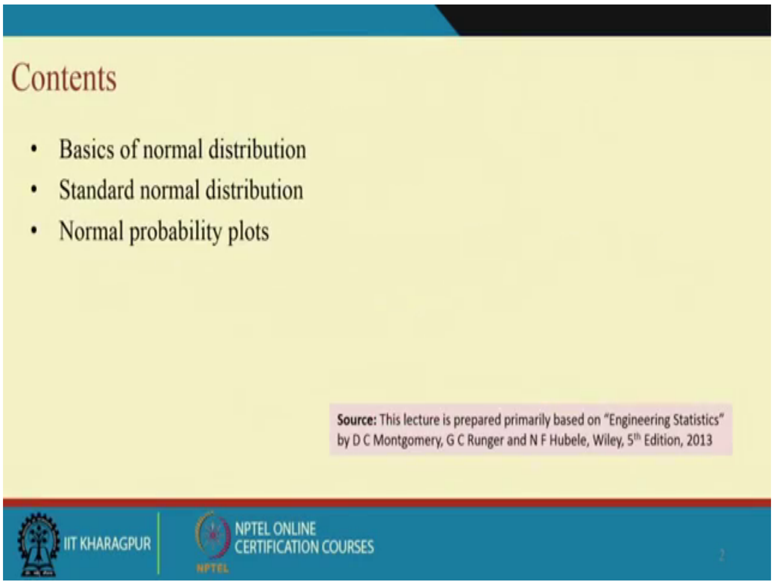
Design and Analysis of Experiments

Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture – 07
Normal distribution

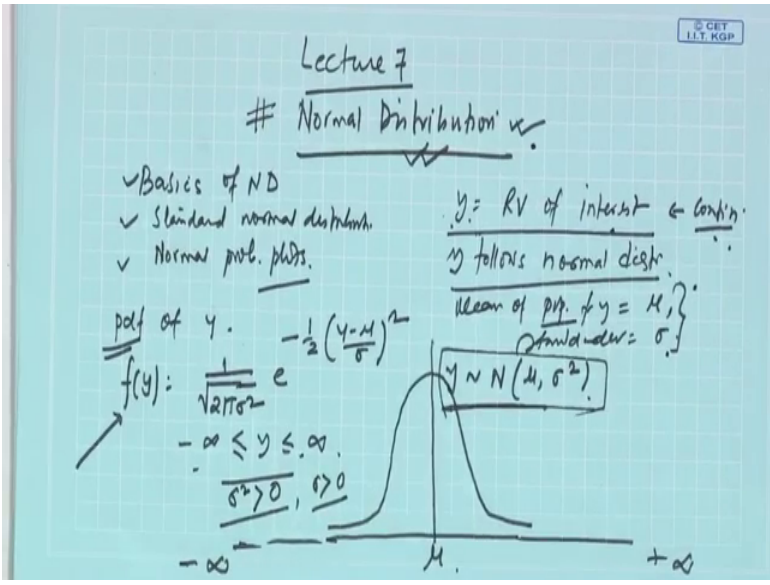
Welcome, today we discuss on Normal Distribution. I told you that normal distribution will play a big role while analyzing data and you will see later on that many a times assumption of normality is a very very important one. What does it mean? That the data coming from a normal population show around half an hour of time; today we will discuss in detail what is normal distribution?

(Refer Slide Time: 01:00)



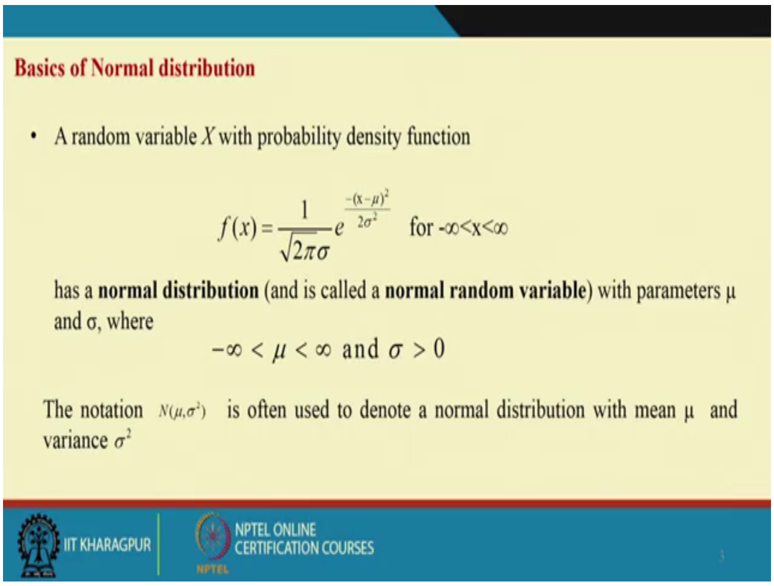
And how to use normal distribution from statistical table point of view and given a data; how do I know that the data will follow normal distribution with some plots.

(Refer Slide Time: 01:22)



So; that means, you will know the basics of normal distribution; then a second one is that standard normal distribution and finally, normal probability plots. So, again this lecture with all those lectures mostly that Montgomerie and his two books design analysis of experiment and engineering statistics books, we have developed this and this I will explain.

(Refer Slide Time: 02:10)

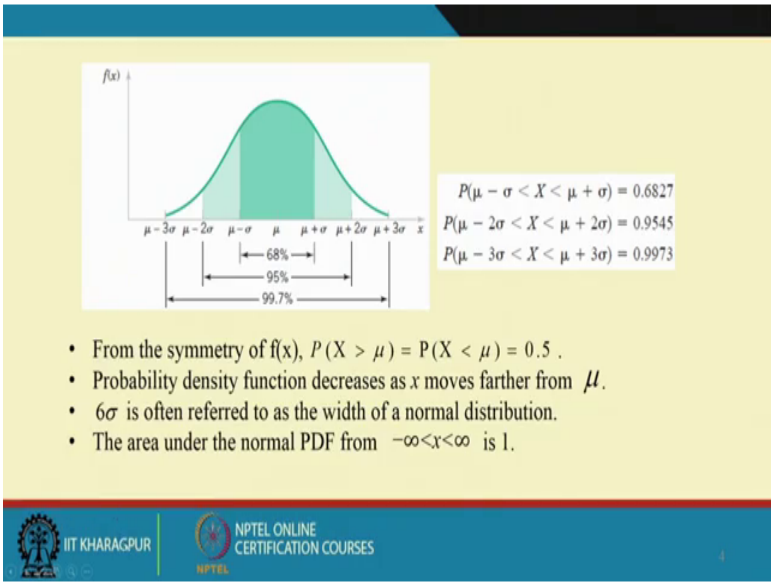


So, let Y is the random variable of interest; Y is the random variable of interest and let Y follows normal distribution and let that the mean of population for Y equal to mu and standard deviation is deviation equal to sigma. Then we can write that Y is normally distributed with parameter mu and variance mean mu and variance sigma square. If I say Y is normally distributed immediately, you will be interested to know what is the probability density function of Y? Because Y is continuous variable and normal distribution is a continuous probability distribution.

So, if Y is discrete variable; you cannot use this normal distribution for Y, there are certain transformation after that you can use normal distribution. But for the time being, you understand that normal distribution is a continuous distribution; while the rest the variable of interest must be continuous. And we are assuming that Y is continuous and Y follows normal distribution with parameters mu and sigma and we are writing; we are denoting Y in this manner.

Then the probability density function PDF of Y will be f Y; which is 1 by root over 2 pi sigma square within bracket; E to the power minus half Y minus mu by sigma to the power square. And here Y range from minus infinite to plus infinite what is this? This is the PDF; Probability Density Function of Y and this is normal probability density function. And if you plot; your plot will be like this; this will be mu and this side and this side like this; this side, this side minus infinite to plus infinite, it must be greater than 0's or other way I can say sigma must be greater than 0; sigma is not 0.

(Refer Slide Time: 05:18)



Now, see the plot this is a normal probability plot here; interestingly you see that where it is a symmetric one first of all; that means, the from the middle point left side and right side, the shape is same and at; if you go away from the mean value that the height is also same. If you go sigma away from mean both side height will be same, the density will be same.

So, this is a symmetric one and you see that there are different colors, colored zone one is mu; minus sigma 2 mu plus sigma, this is known as one sigma range means mu minus 1 sigma mu plus 1 sigma. Second one is mu minus 2 sigma mu plus 2 sigma to sigma range; third one is mu minus 3 sigma mu plus 3 sigma range. So, effectively then if I say that what is the spread from mu plus sigma to mu minus sigma; although we say that I am saying on my one sigma range, but (Refer Time: 06:36) plus 1 sigma minus; this side 1 sigma; these two sigma, then this one is 4 sigma, the range is 4 sigma mu plus minus 2 sigma and this is mu plus minus 3 sigma range is 3; 6 sigma this.

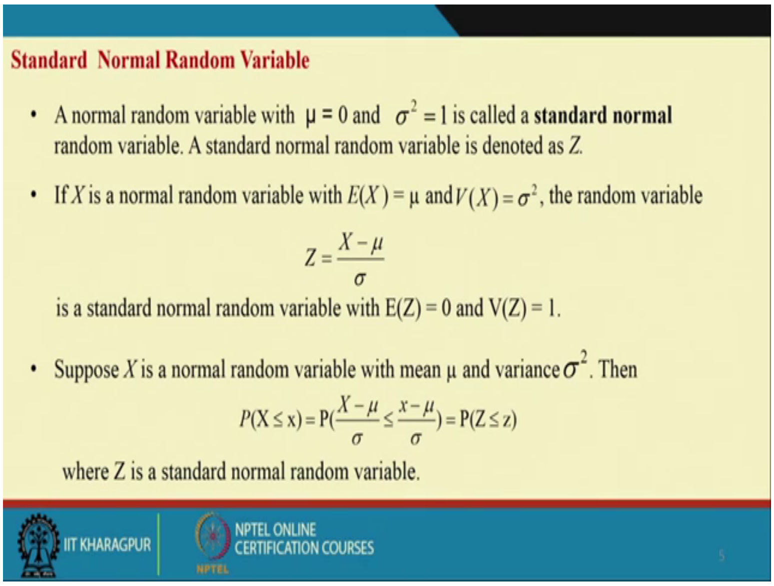
Now, there are certain other values given in the figure 68 percent, 95 percent and 99.71 percent; what does it mean? It means that, if you take the range mu minus sigma to mu plus sigma, then the probability is 0.6827.

If you take mu minus 2 to mu plus 2 sigma probability is 95.45 percent or 0.9545 percent and in this sigma 0.9973 percent. So, other what I we can say approximately 68 percent observation falls within mu plus minus 1 sigma, approximately 95 observations falls within mu plus minus 2 sigma.

Effectively 99.7 percent observation for mu plus minus 3 sigma. So, this is basically the way the normal distribution figure is also interpreted. So, another thing you can see that from the symmetry probability x greater than mu or equal to probability x less than mu also equal to 0.5 means from this side 0.5 mu left side 0.5 mu, right side is a 0.5. Probability function increases, if x moves from this side to that side.

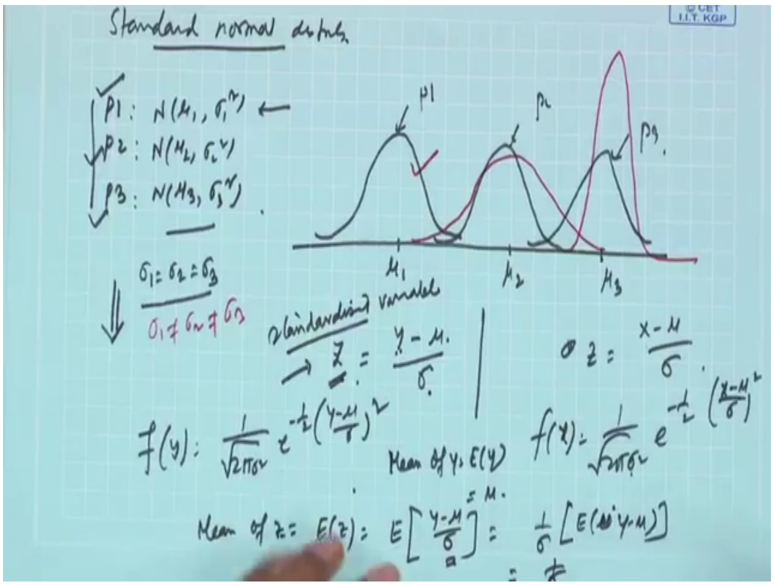
Now, here we are using f x and x and you can use Y; if your random variable is Y, then here the way I have written here Y. So, if it is Y only that in place of x; you write Y here, so it will be Y, it will be Y, it will be Y. So, in general in most of the books; they start when random variable, they write x is the random variable, but from our side as we are more interested in the responsible variable; which is an effect variable, we use the notation Y and all through in the during class; all the class will be using Y for response variable.

(Refer Slide Time: 09:20)



Now, another important concept is that is standard normal distribution, you have seen normal distribution, but standard normal distribution.

(Refer Slide Time: 09:31)



For example, suppose I have three process process 1, process 2, process 3 and everywhere it is basically normally with mu 1 and sigma 1 square; this also normally distributed with mu 2, sigma 2 square; this normally distributed mu 3, sigma 3 square let it be like this.

So, then if I want to if I draw; I will draw like this maybe this is my mu 1, this is my mu 2 and this may be mu 3 and it can be suppose sigma 1 equal to sigma 2 equal to sigma 3; let it be it, may not be let it be. So, in that case you can plot like this; so mu this is for process 1, this is for process 2, this is for process 3; so that mean for three process; three different normal distribution.

Suppose if I want to compare, you have to do like this, but I want to get something which will be use irrespective of which process, it is coming from. Hence for this process, for all those process I will be able to use that normal distribution; I want to standardize it that is what is known as standardization irrespective of which population normal population we are talking about their mean standard deviation that may vary, but I want to convert them to a standard one; that is known as standard normal.

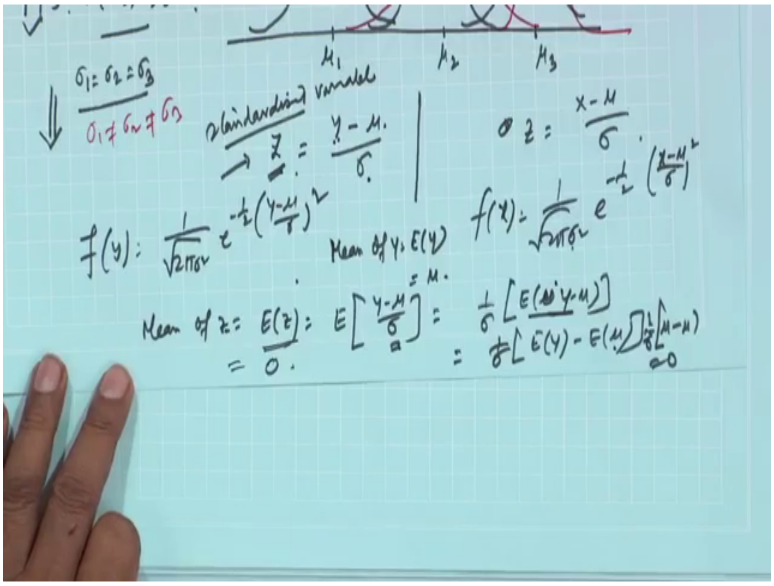
Now, if in this particular case; if sigma 1 not equal to sigma 2; not equal to sigma 3, then what may happen? This may be your sigma 1, but sigma 2 may be something like this and sigma 3; may be something like this. So, that mean not only that mean value, but the standard deviation also defined. So, that mean you have if there are N k number of process k number of such distribution will be there. So, it you want to make them standard all bring together.

So, this is possible if you do some kind of normalization or other I can say standardization, what is this standardization? Suppose, if I create a another variable which is Y minus mu by sigma; if Y is my random variable, if x is my random variable. Then you can say x minus mu by sigma; obviously, mu and sigma are the mean and standard deviation. In this case; what will happen? Then when you are subtracting the mean from the random value and then dividing by the standard deviation. So, you are getting instead of Y; you are creating another variable called Z which is known as standardized variable.

Now, if my Y; PDF is f Y is 1 by root over 2 pi sigma square E to the power minus half Y minus mu by sigma square or if you use x; then you will write the same manner 2 pi sigma square E to the power minus half x minus mu by sigma square. Please keep in mind that sigma; this sigma and mu and this sigma mu are not same, this is for Y; this is for x. If there are different random variable, different mu sigma just for the sake of simplicity; I am not giving any other notation like mu x sigma x it can be given.

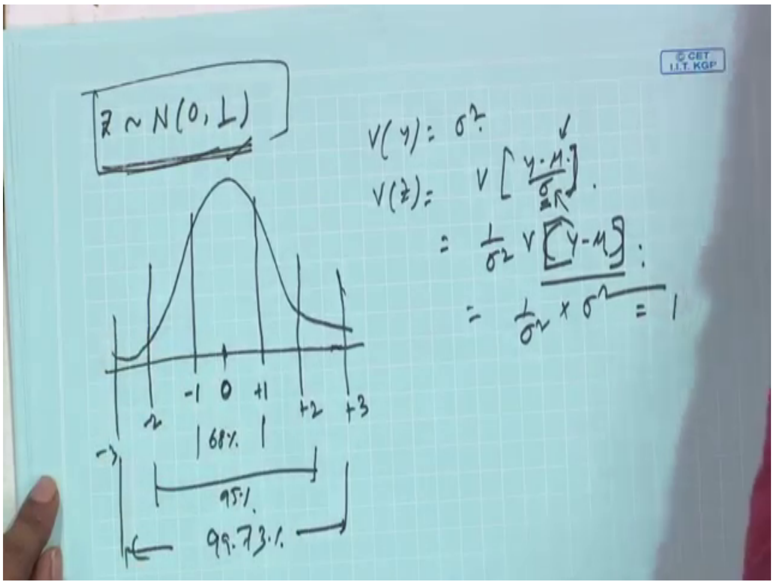
Now, we all know what is the mean? Mean of Y expected value of Y which is nothing, but mu then if I want to know what is the mean of Z? So, what you write expected value of Z? That means, expected value of Z is nothing, but Y minus mu by sigma. Now, sigma is constant; so, it will come out then expected value of mu Y minus mu. So, 1 by sigma now expected value of a; my x minus Y is expected value of x minus expected value of Y.

(Refer Slide Time: 15:04)



So, expected value of this will be expected value of Y minus expected value of mu. So, what is expected value of Y 1 by sigma will be there expected value of Y is mu and mu is constant this also mu; so it is 0. So, what happened then? The mean of Z is 0.

(Refer Slide Time: 15:29)



So, mean of Z is 0; so in standard variable, it will follow normal distribution because earlier one is normal; so, with mean 0.

Now, what will be the variance part? Now what is your variance value? Suppose, if I say variance and of Y equal to sigma square, but you want variance of Z so; that means, this is nothing, but your variance of Y minus; Y minus mu by sigma now this is a constant 1. So, it will be 1 by sigma square variance of Y minus mu variance of 1 minus mu and variance of 1 minus mu, you will find out this is nothing, but sigma square again.

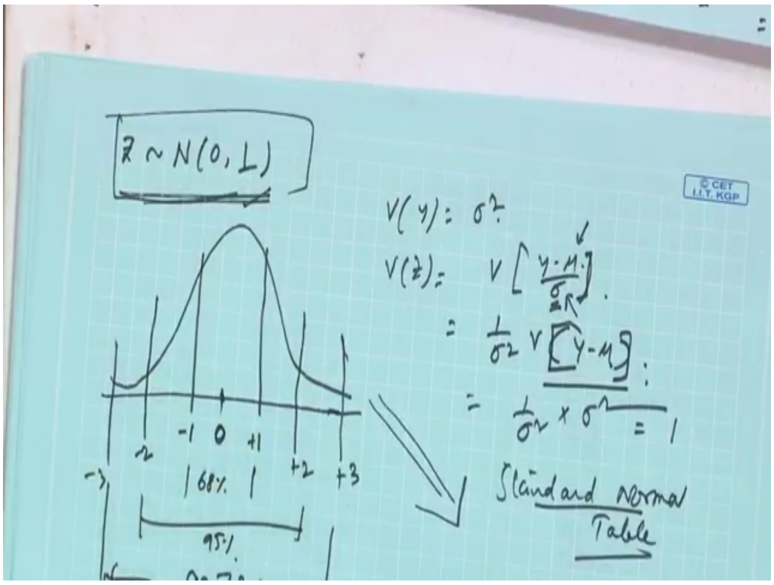
So, what you will do? You will find out because variance of mu will be sigma 0. So, that become will 0; so this will be 1 by sigma square into sigma square; it will be 1. So, now whether mu is mu 1 and sigma is; sigma 1 irrespective of mu and sigma value, when you standardize. So, all those processes earlier here what happened all those things; it is standardized to.

Only 0 1; then I have only one graph I have one density function; I have Z equal to 0 and this is my thing. This side and that mean I can write 1 sigma; 1 sigma is nothing, but minus 1, this side plus one this side; where is in sigma is 1 and the 68 percent data will fall within this. Then similarly your 2 sigma minus 2; this side minus 2; so here 95 percent approximate will fall, then 3 sigma plus 3 plus 3 minus 3 and this is the case area; 99.73 percent will form. So, what happened? You have different normal distribution.

But all can be converted to unit normal and this distribution itself is sufficient to explain behavior of this. So, if I call this is; what is your standard normal distribution? Standard normal distribution means it is a unit normal distribution; why unit? It expected value is 0 and standard deviation is 1 and this why it is standardized. Because this is a distribution, which can be normal distribution; which can be used for any normal population when it is normally distributed, whether mean is 50; standard deviation 20 or mean 100, standard deviation 5. So, that can be converted to unit normal and this distribution will be used and because of this not standard normal distribution.

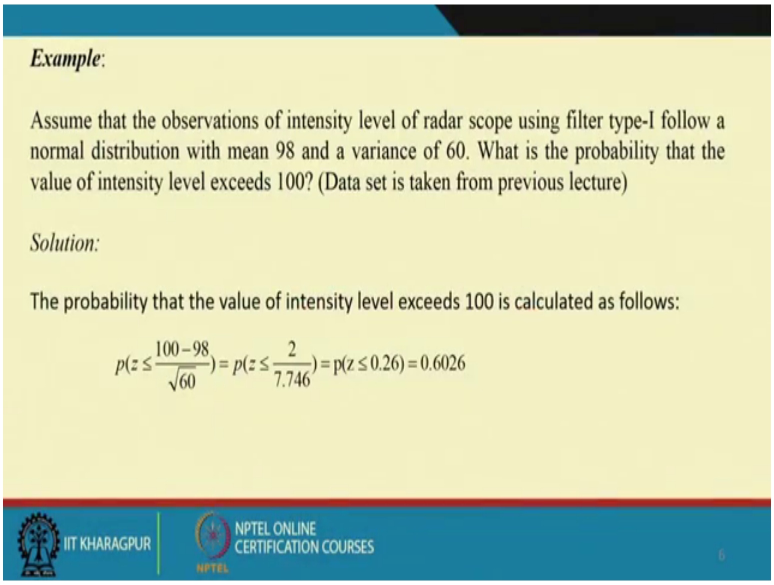
Availability of standard normal distribution, you are also having another advantage.

(Refer Slide Time: 19:10)



There is when from this a general standard normal table is standard normal table is created, otherwise if you consider in the original scale without standardize the data; then how many normal table is required for every mu and standard deviation combination 1 table? So, billion infinite number of tables required, but here because of this standard 1 table is required. Because every normal variable can be converted to standard normal; then using standard normal table, you will be able to know; what is the probability of happening not happening or happening less than right to that all those things.

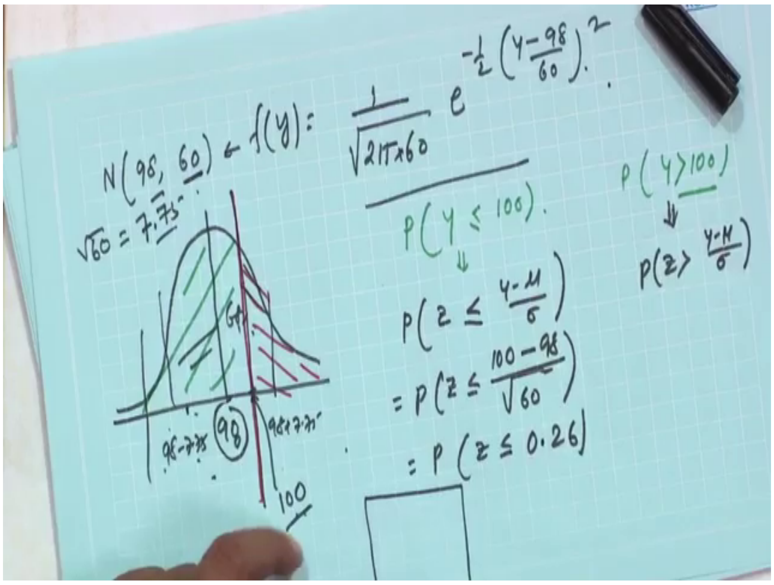
(Refer Slide Time: 19:54)



Let us see the use of standard normal distribution; assume that the observation of intensity level of radar scope using filter type 1; follow a normal distribution with mean 98 and variance of 60; it is basically bi variance we are talking about. The standard deviation of sequences variance 60; we have to use variance 60.

So, we have not used the unit so, but any how in some units what is the probability that the value of intensity level exceeds 100? So, what is given then if a process where basically observing through radar scope, the target this is the process you are coming and observing and at certain intensity level only; you will be able to see the target.

(Refer Slide Time: 20:53)



So, now this is we are seeing normally distributed with mean intensity level 98. And variance 60 above then if I want to know what is the PDF for this? If Y is the response variable then this is nothing, but 1 by root 2 over 2 pi sigma square is 60 into 60 E to the power half Y minus mu is 98 by sigma is 60, this square; this is the PDF for this process.

So, you can summarize the; what I can say you right down tone 20 and this some value will come and ultimately this exponent, this will be there. And may be your distribution normal distribution; suppose if it is like this, where this one is nothing, but 98 and if I go to 1 sigma that root 60 means how much root 60 means? Seven point something seven point suppose 7 5; let it be.

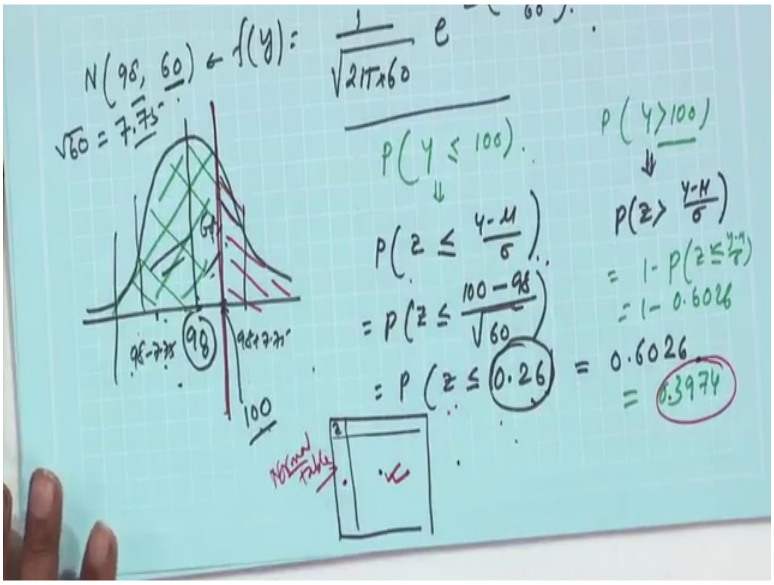
So, then 1 sigma means 98 minus 7.75; this side 98 plus 7.75; then 2 sigma like this. This is my 1 sigma, then 2 sigma 98 minus 2 into this; like this 68 percent will fall under what is my question? Question is what is the probability that the value of the intensity level exceeds 100? Exceeds 100 value of the intensity level does not exceed 100; that is what we are interested not exceed 100 that does not exceed 100; that is what I want so; that means, if it is 98 somewhere; here is 100 is there, please do not think that the top one; what I mean to say the top one means; what I mean to say that not this vertical axis.

We are talking about this value only; suppose this is 100; so, there can be two possibilities. Suppose, if you say that what is the probability that the Y value exceed 100 means you are interested to know the probability of this set; this set suppose I told you that mine interest is to know what is the probability that Y value does not exceed 100? Then probability this side, the area under this curve; that means, what happened here? That means your; let us start with this suppose probability that Y value does not exceeds 100.

You want to know; it or may be another one probability that Y exceed 100. So, let us do this one first; then what do you have to use a table and a normal standard; normal table which is available. So, you convert this to Z; so that mean if I want this in terms of Z, this is nothing but probability that Z less than or equal to Y minus mu by sigma or here it will be probability that Z greater than Y minus mu by sigma.

So, what is Y value here? This Y value is probability Z less than or equal to Y is 100. What is your mean value; mu value is 98, what is the sigma value root over 60 and this value will ultimately comes like this. P Z less than or equal to 0.26; so there is normal distribution table a, I will show you in the class of sampling distribution time normal table.

(Refer Slide Time: 25:19)

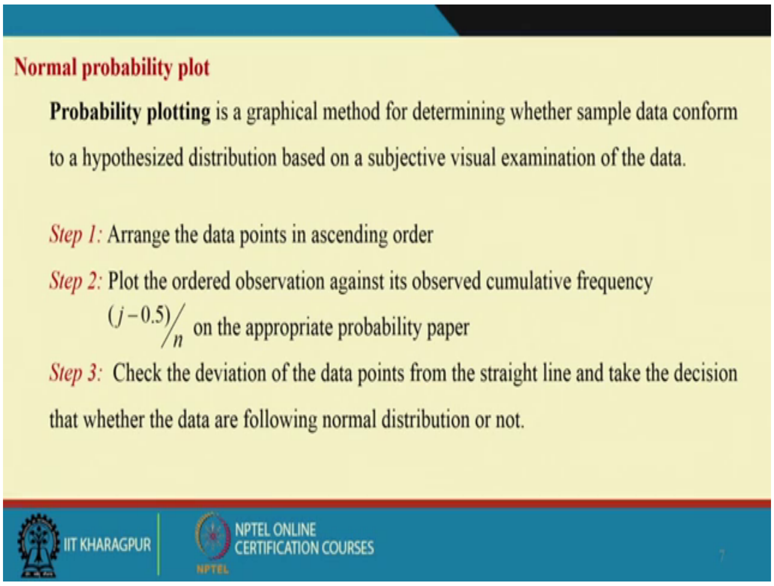


So, there the probability Z values; Z values with probabilities will be given; now for this Z value, suppose this is 0.26; what is the probability value? It will be, you can get from the table and that value is 0.6026 so; that means, this side the green color sided area this side; this value is 0.69; that means, the probability that the intensity level in the experiment intensity level will be 100 or less.

Is this then what will be this? This will be the total is 100. So, this is nothing, but 1 minus probability Z less than equal to Y minus mu by sigma. So, that is 1 minus 0.6026 this is nothing, but 4 7 9 3. So, red colored area this is 0.3974; so, this is the use of unit normal you are bringing to unity normal first converting the original normal to unit normal. Then from unit normal, you are getting Z value there will be normal table and using the normal distribution table. Once you know the Z value, you will be having the probability value.

Other end; if you have the probability value, you can get also Z value; how to see normal table is very important and you all must practice that how to read normal say unit normal table as such other no table also you must required to know.

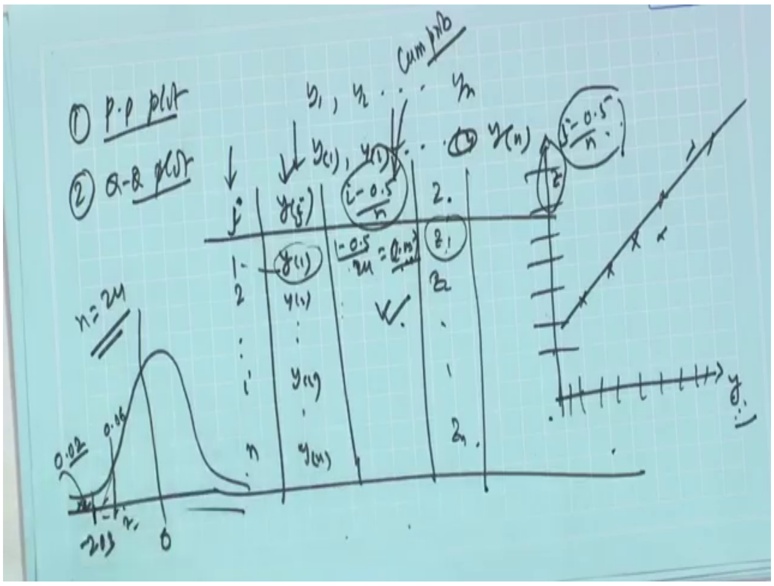
(Refer Slide Time: 27:21)



But fine; this is fantastic way show along data is coming from normal distribution, but you are assuming that it is normal, but what is the guarantee that things are normal or population is normal. So, as you do not know; have a completed if you have some earlier experience that you know; that the data coming from the normal the population is normal fine, but if you do not know that whether the population is normal or not. So, in that case when you do take some sample as such you do some experiment.

And from that experimental data, you can very easily do some plotting or some kind of quantitative study also through which you can say that whether my population is normal or not normal. So, we will see such plot will show you two plot; two different plots; one is P P plot, that is probability probability plot, another one is Q-Q plot that is quantile quantile plot.

(Refer Slide Time: 28:24)

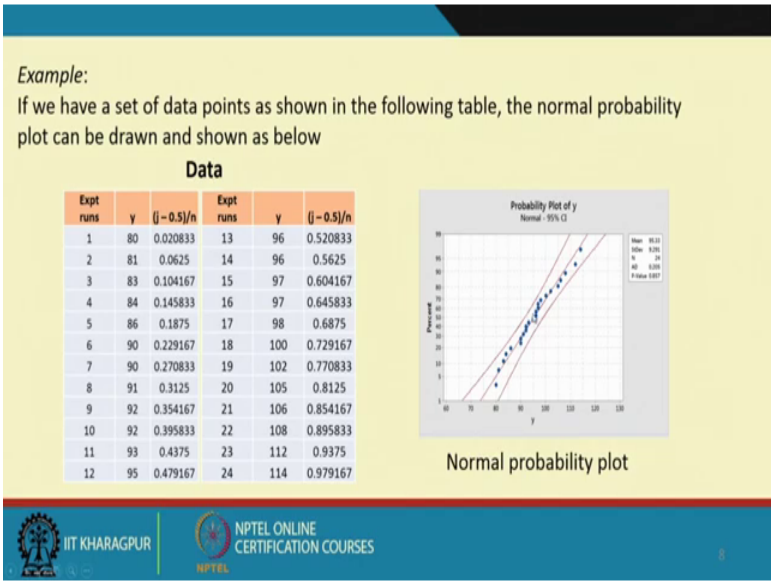


So, in P P plot, when you have data from Y 1 to Y n; what you do? You arrange them in ascending order and then what happen? I 1, 2 like I to n; these are the data points and your ascending data ordered data Y 2, Y I, Y n fine and this side; you write down I minus 0 0.5 by n. So, here it will be 0.1 minus 0.5 by n; means if n equal to 24; in our case then 24; so like this.

So, you will get some probability, some percentage values you will get these are nothing, but this is nothing, but cumulative probability value or percentage probability values. So, once you get this one; if you plot this with this order that what are observations mean? This side your Y; which is minimum to maximum order 1 and this side, if you plot this I minus 00.5 by n and if you get a straight line kind if you plot (Refer Time: 29:53) plot is like this so, but if you draw them its resembling a straight line then your data is coming from normal probability distribution are written there, population is normal sometimes instead of I; if we if I use it is j then I will write here j minus; these are all notation only.

So, Y j like this; so let me read out here arrange data points in ascending order first one and then plot the order observation against its observed frequence cumulative frequency. So, I want to use the slides.

(Refer Slide Time: 30:39)



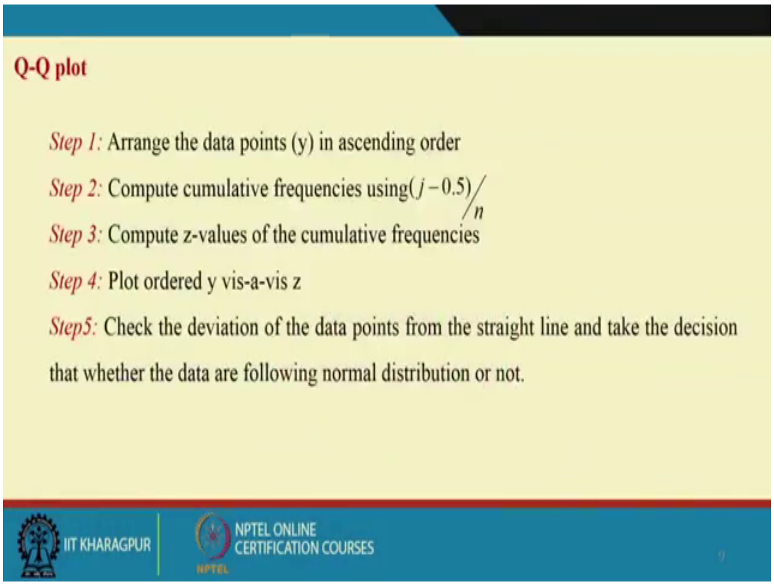
So, now with the given example you see what we have done and then here we are showing the probability plot. This is a normal probability; that normal probability paper you have to use this side the percentage which is nothing, but I minus 0.5 by N or j minus 0.5 by N; this side and this axis is the Y values and we have use many term and that is why and we got this kind of.

Now, you see the middle one; middle one is nothing, but what is the line joining them; this is a straight line kind of things. So, we assume that it is probability normally distributed data come from normal population, but there are two other line, this side; this slightly slanted line curve line; these are nothing, but the 95 percent confidence interval of this data. Any how those confidence interval part you do not know now, but later on you will be able to understand this confidential report. This is one what is Q-Q plot in Q-Q plot; you do all those things. This first; this one; this one and this three you do; so instead of plotting like this Y versus this cumulative probability; what you do? You find out the Z value.

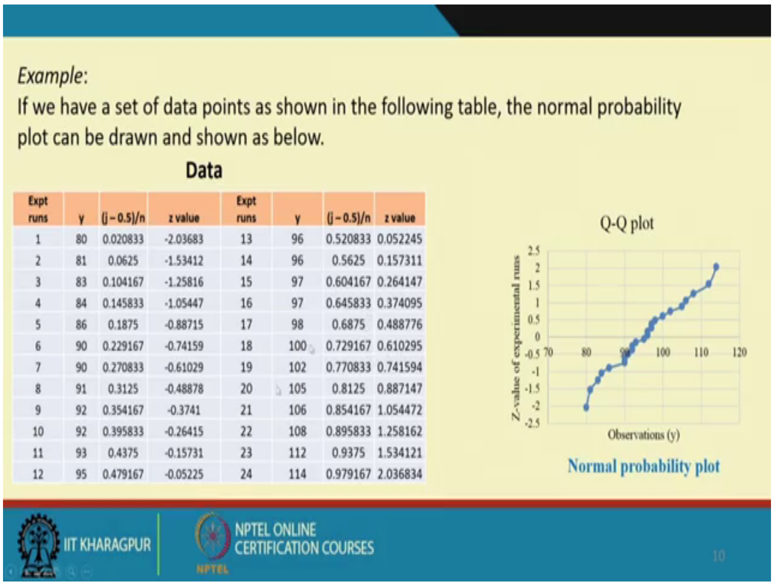
Means suppose, if this is 0.005; then what is the Z value corresponding to this probability? Suppose, it is Z 1; then it is Z 2; so like this it is Z N; this Z value. So, what you do you basically plot Z versus Y? This is also quantile in terms of Z and this is the quantile interms of Y both quantile; quantile means when you divide the data into N segments every segment is a quant this is known as quantile.

So, one segment, one quantile like this; so similarly Z is here in data points; these are split into N quantiles. Here also we found out the corresponding Z and you have also split into N quantile; then this quantile means this one first one, this one is compared with this you have made a plot and this plot also will resemble a straight line. So, let us see this that this Q plot.

(Refer Slide Time: 33:08)



(Refer Slide Time: 33:09)



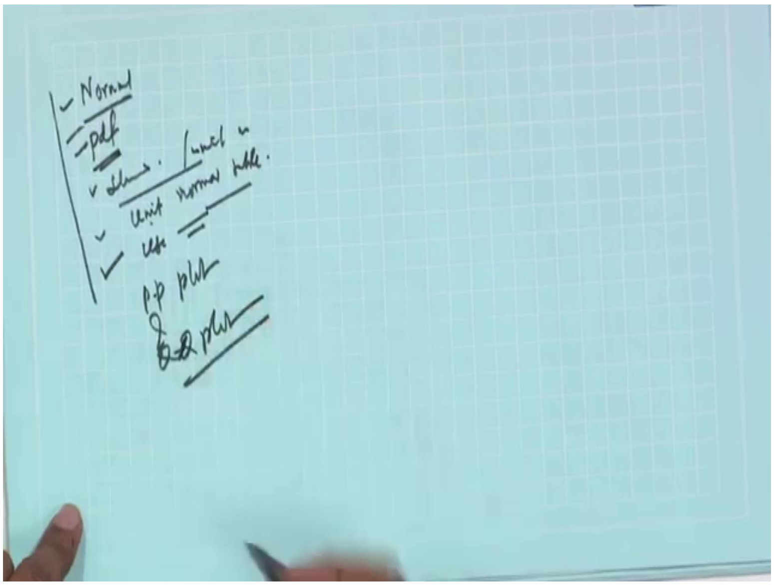
Expt runs	y	(j – 0.5)/n	z value	Expt runs	y	(j – 0.5)/n	z value
1	80	0.020833	-2.03683	13	96	0.520833	0.052245
2	81	0.0625	-1.53412	14	96	0.5625	0.157311
3	83	0.104167	-1.25816	15	97	0.604167	0.264147
4	84	0.145833	-1.05447	16	97	0.645833	0.374095
5	86	0.1875	-0.88715	17	98	0.6875	0.488776
6	90	0.229167	-0.74159	18	100	0.729167	0.610295
7	90	0.270833	-0.61029	19	102	0.770833	0.741594
8	91	0.3125	-0.48878	20	105	0.8125	0.887147
9	92	0.354167	-0.3741	21	106	0.854167	1.054472
10	92	0.395833	-0.26415	22	108	0.895833	1.258162
11	93	0.4375	-0.15731	23	112	0.9375	1.534121
12	95	0.479167	-0.05225	24	114	0.979167	2.036834

And see for this data set what happened is Y part remain same this part also remain same. So, from this we found out the Z value; what is this minus 2.03 minus 2.03 means?

Suppose, if I go to Z naught Z; this is 0 somewhere here minus 2.03. So, the probability value is 0.02; that mean if it is Z normal, so this portion this side this value is 0.02. So, similarly next probability is 0.06; left hand side 0.06; so 0.06 and you found out the Z two value. So, like this you know the probability get the corresponding Z value and this Z value using x l that norm inverse we found out this values.

Now, this Y versus Z otherwise; I can write Z versus Y, this is Y observation Z observation plot and you see that when I join all the points; it is almost resembling a straight line. So, we can say the data coming from a normal distribution or I can say the population considered is normal population. So, I hope that you have understood this fully; so today, what are the take home for you? That what is normal distribution? How to denote normal distribution?

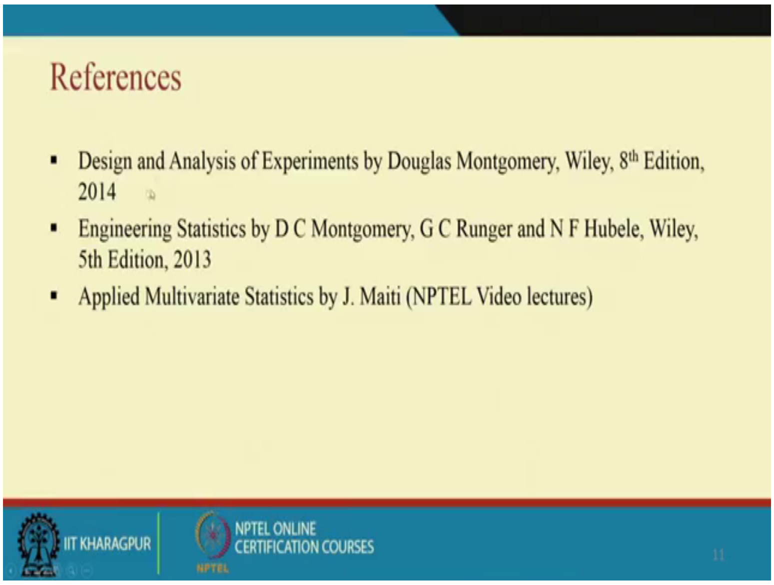
(Refer Slide Time: 34:34)



What is the PDF for normal distribution? How to read a PDF of a normal distribution? What is standardized normal or unit normal unit normal and then how to read unit normal table unit; normal table and although I have not shown you any table, but I hope that you will be able to find out other I will show you in next class; next probably.

And then example or use of unit normal table now as you assume that the population is normal. Suppose you have some data; may be from the experiment, so using that experimental data; can I say the population is normal or not then the P P plot and Q-Q plot; Q-Q plot is told to you, thank you very much.

(Refer Slide Time: 35:34)



Again I acknowledge the references; please see the references Design Analysis of Experiments by Montgomery and again Montgomery Statistics and Multivariate lecture of mine; thanks a lot; see you next class.