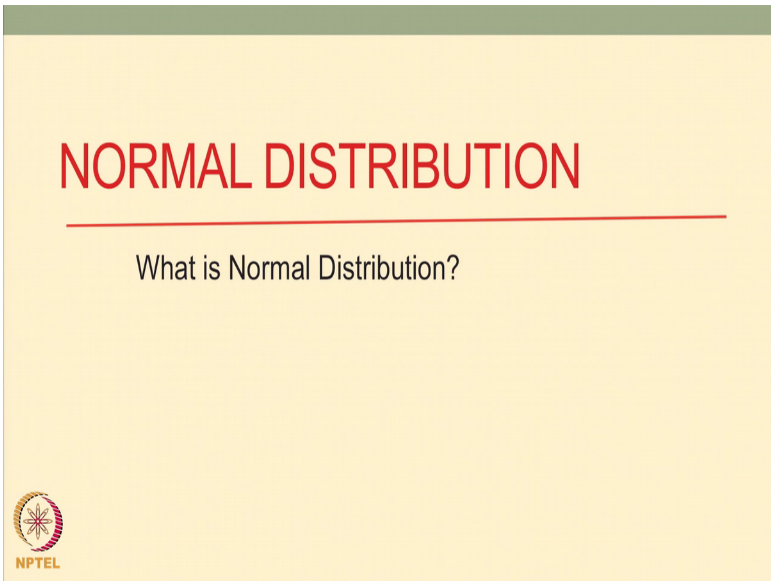
Introductory Mathematical Methods for Biologists
Prof. Ranjith Padinhateeri
Department of Biosciences & Bioengineering
Indian Institute of Technology, Bombay

Lecture – 39
Normal Distribution

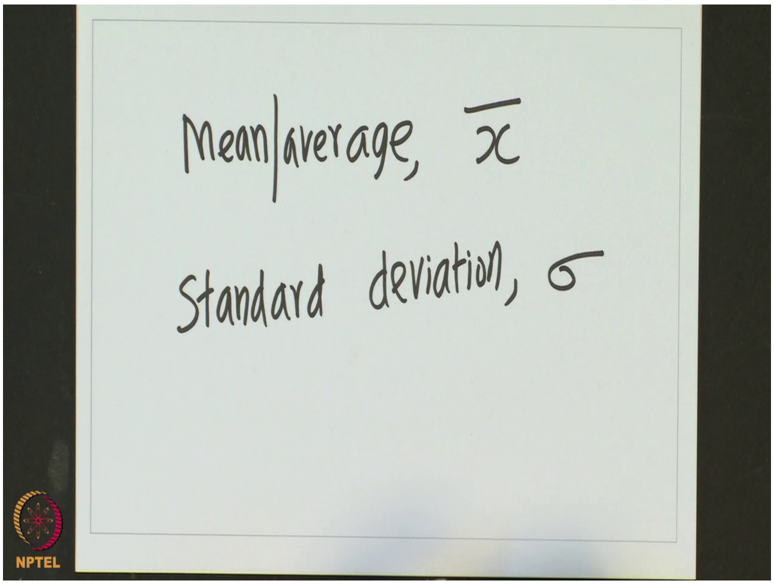
Hi, welcome to this lecture on Mathematical Methods from Biologists. We have been discussing probability distribution and statistics the topic for today's discussion is Normal Distribution.

(Refer Slide Time: 00:29)



In this lecture we will answer what is this normal distribution what is the functional form how would we use it etcetera. So, to understand this let us think of a typical situation that you would face in a lab or in real life or we do some set of measurements. And we have 2 quantities which we typically take which is mean or the average or and standard deviation.

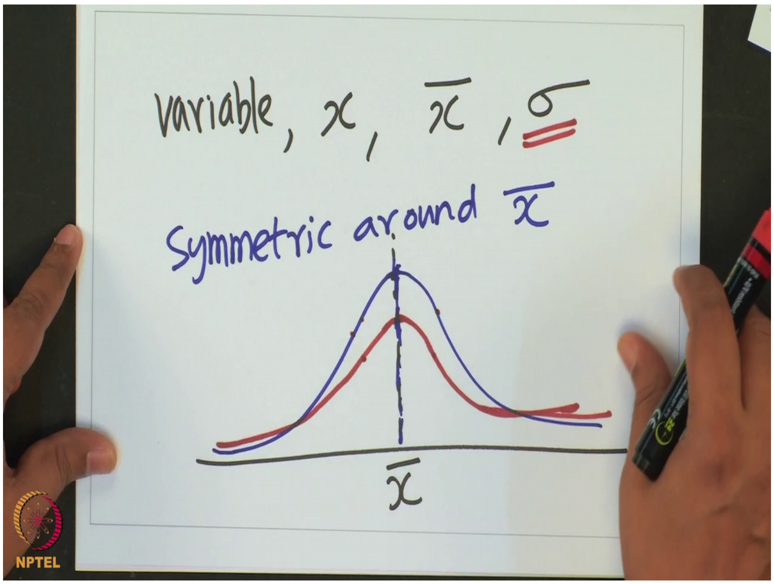
(Refer Slide Time: 01:07)



So, let us have let us think of 2 quantities the first quantity is mean the or average which denoted by x bar and if from any experiment we would also measure standard deviation which is denoted by sigma. So, if you have these 2 parameters. So, if I do an experiment and specify only 2 parameters which is the mean and the standard deviation, typically the assumption typically 1 would assume that other numbers are distributed around this mean value. If I say mean x bar is equal to 30 1 would naively think that the other numbers are distributed around 30 with a deviation of the value sigma whatever be given this by this number sigma.

So, imagine that if we typically says that is what 1 would naively think very often what people assume that such distributions are normal distributions. So, what is the normal distribution it is some distribution, which is distributed symmetrically around the mean x bar value. So, let us plot this and understand the plot first. So, here first of all this is here x.

(Refer Slide Time: 02:36)



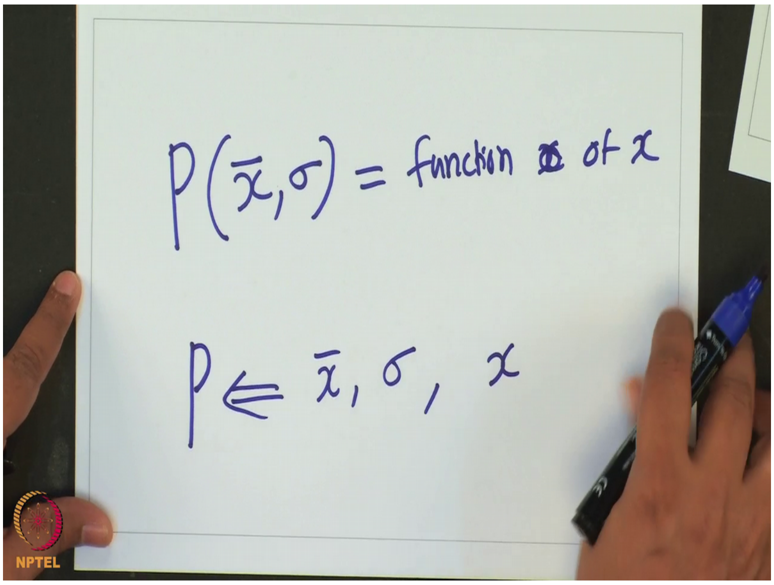
So, what you would have you would have a variable, x. So, something that you would measure variable x this is a continuous variable what does that mean continuous variable would mean that that some quantity which can take any value not discrete values. For example, the height or the gene expression it could be 3.1 micro molar the it could be any value it need not be a discrete value it is not like a this only not just like a number 1 2 3 4 it could be 3 it could be 3.1 3.7 5.3 it could be a fraction.

So, at least if you have continuous variable x and it has a mean x bar and standard deviation sigma the assumption is that around this this is my x bar around this there is a symmetric distribution. So, around this either side this is symmetric. So, if I draw something like this. So, this is symmetric around the mean value. So, this is the mean x bar and this is symmetric around this x bar value this kind of a distribution is what is normal distribution. So, first point is this is symmetric around symmetric around the mean x bar and it has a width. So, typically I could draw the same distribution symmetric around x bar in many different ways. So, I could draw this slightly differently.

So, let us say I could draw something like this this is also kind of symmetric around this, but this is different from this the blue curve is different from the red curve I could draw many curves which are symmetric around x bar, but now which of this curve will be desire is my distribution of this of our interest. So, sigma will design this curve or this curve. So, both of these curves same x bar, but this and this have different sigma standard deviation. So, that is something that we will understand as we go along, but the point is it is a curve it is symmetric around x bar.

So, the mathematical expression for such normal distribution is the following. So, as we know this normal distribution has.

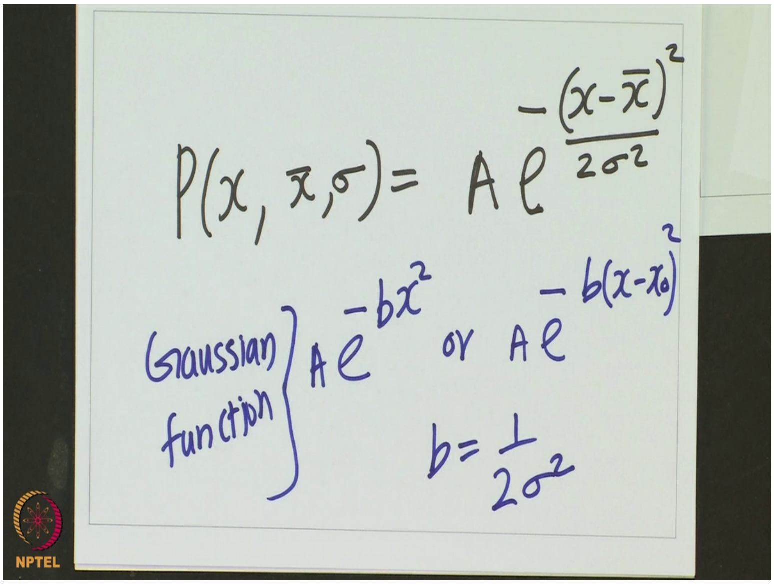
(Refer Slide Time: 05:21)



So, the probability distribution is a function of 2 parameters 1 is x bar and 1 is sigma only 2 parameters, you have to know if I specify these parameters I can get a distribution I can say this is if you specify this parameter I can get some function and this is a function of this variable x function of x.

So, basically the probability distribution P would be desired by these 2 parameters x bar and sigma and the variable is x. So, now, let us write the mathematical expression for this which is as follows.

(Refer Slide Time: 06:04)

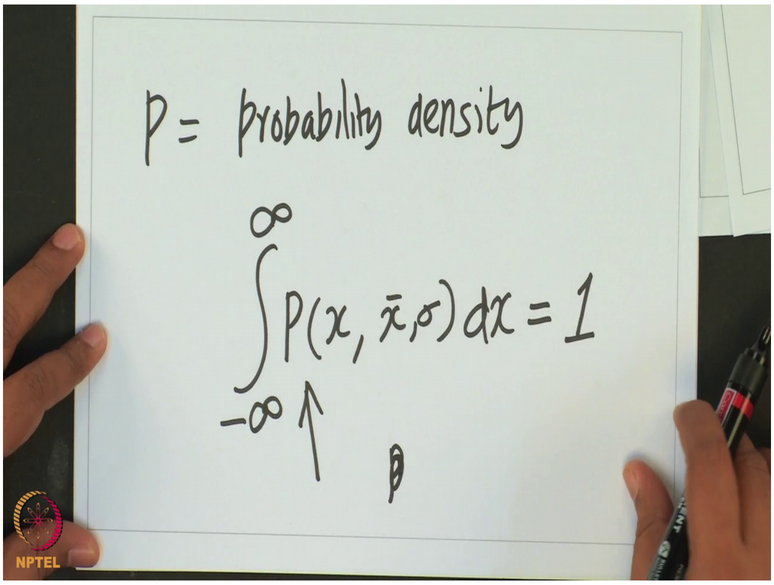


So, P of x the probability to have any value around x given an x bar and a sigma is it is a normal distribution a Gaussian function. So, e power minus x minus x bar Whole Square divided by 2 sigma square times some constant A.

So, this is the mathematical form this is a normal distribution. So, typical normal distribution is e power minus bx square or this is the Gaussian function. So, the mathematical function called the Gaussian function, which has of the form e power minus bx square or e power. So, A e power minus bx square or A e power minus bx minus x 0 whole square this is a typical functional form where b is like 1 by 2 sigma square. So, here b is like 1 by 2 sigma square

So, this is the functional form now what is A. So, we would understand what is A means, now let us make a statement for P what precisely there were P means.

(Refer Slide Time: 07:35)

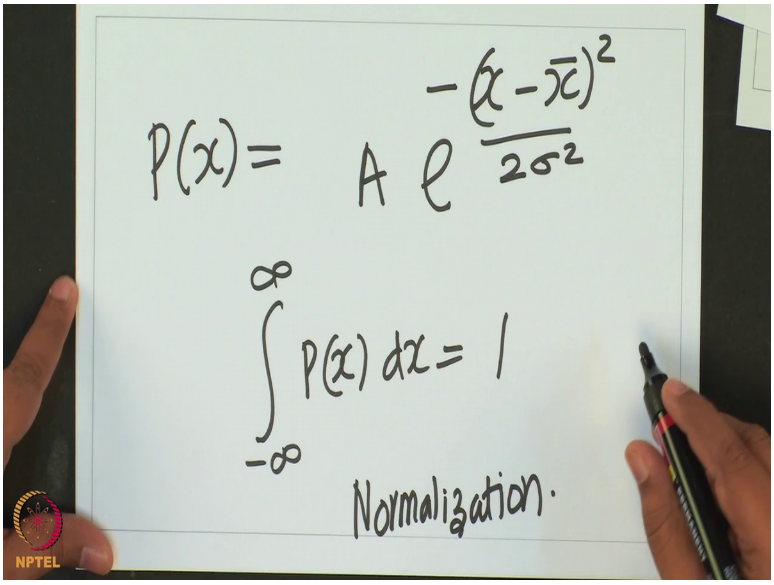


So, the P means. So, the P is nothing, but probability density probability why do we call it probability density because the integral P of x given x bar and sigma the integral of this with dx has to be 1.

So, this this is called the normalization the total probability is what does this mean the total probability is 1. Therefore, if since this integral is 1 this will have a this is like has a dimension if x is a length this will have a dimension of 1 over length. So, depending on this is P is more like a probability density this have a this is more like something divided by x. So, that you multiply with this x will be a number. So, these times this should be a number dimensionless. So, this will have a dimension of 1 over x P has a dimension of 1 over x the only then P times x will have will be a dimensionless number.

So, this is important to remember what is the meaning of P the meaning of P is the probability to find the value of x x as we just wrote.

(Refer Slide Time: 08:54)

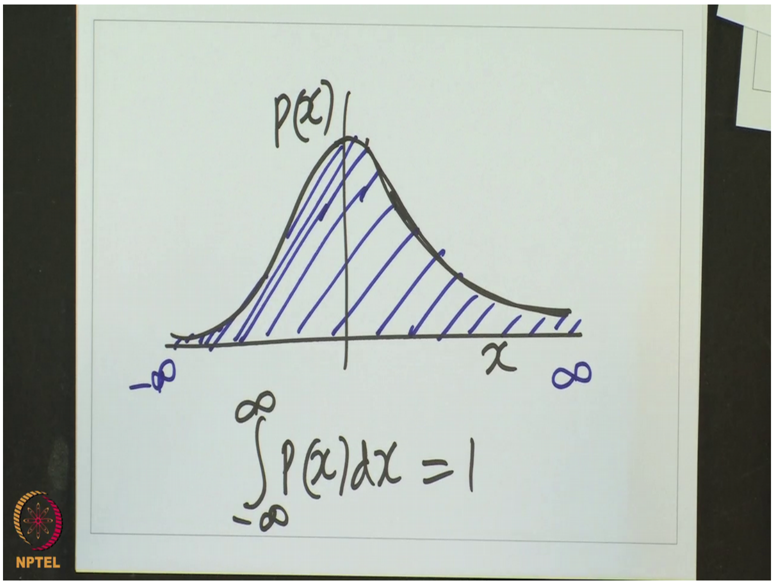


So, I would just write P of x the sum of course, sigma and x bar is a part of it. So, this is an exponential function minus x minus x bar whole square by 2 sigma square and this A value here, which would be determined by this integral as we wrote if this we would choose an a value such that this P of x dx is 1. So, to conduct we would choose an A value which would satisfy this condition.

Now, what is this P of x dx 1 actually meant it is the normal it is normalization as we said we used y as we use in every probability distribution this is the normalization. So, which says that essentially the total probability is 1 that is what is means. So, what is P of x means P of x means probability to find to have a value x around x x between x and x plus dx.

So, now, the total probability to get any value between minus infinity to infinity is essentially 1, let us a little bit more think about this integral of this a little bit. So, let us plot the distribution first. So, this let us plot a normal distribution.

(Refer Slide Time: 10:27)

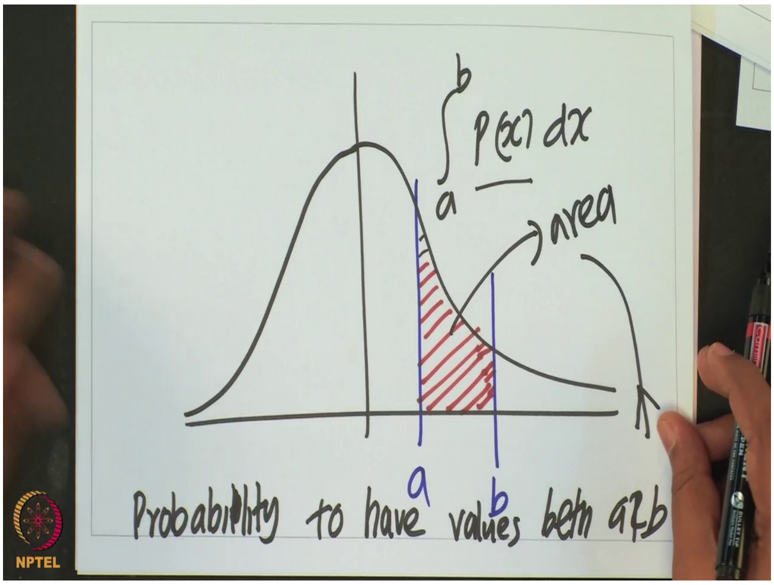


So, I have a distribution which is should be symmetric a nice, now if I just take this integral P of x dx though this is P of x versus x and this integral is equal to 1 would means that the total area under this the whole area under this is 1.

So, if I just mark this whole region from minus infinity to infinity this area is 1 that is what it means so the total probability. So, this also physically means to probability to get any x value between minus infinity to infinity this is minus infinity somewhere here infinity here is 1 you would surely find some of this x values between minus infinity and infinity.

Now, if I take a small part here which is a small region let us take the same distribution.

(Refer Slide Time: 11:44)

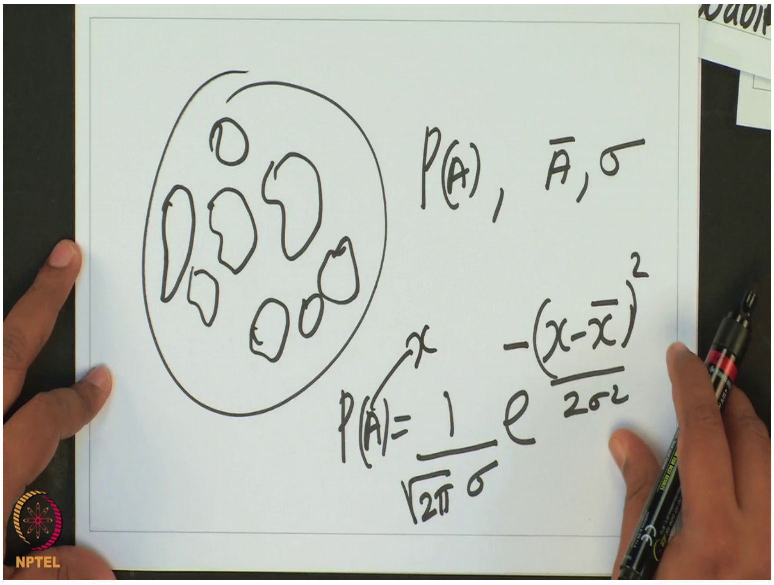


And look at a small region. So, let us look at a region between this and this. So, this is a and this is b and this area under this area of this shaded region mathematically will be essentially written as integral a to b P of x dx. So, this is the if you do this integral which is basically the area of this shaded region and what does this mean this essentially means probability to get any value between a and b probability to have any value between a and b.

So, if I have some data which is normally distributed and the probability to have any value between a and b is basically shown calculated by computing this area between a and b. So, this is important remember this means probability; probability to have values between a and b values between a and b, probability to have values between a and b is basically obtained by this area. So, this area gives the probability to have values with a and b in the distribution.

So, let us take an example imagine that you have a bunch of cells. So, again we will let us go back to experiment.

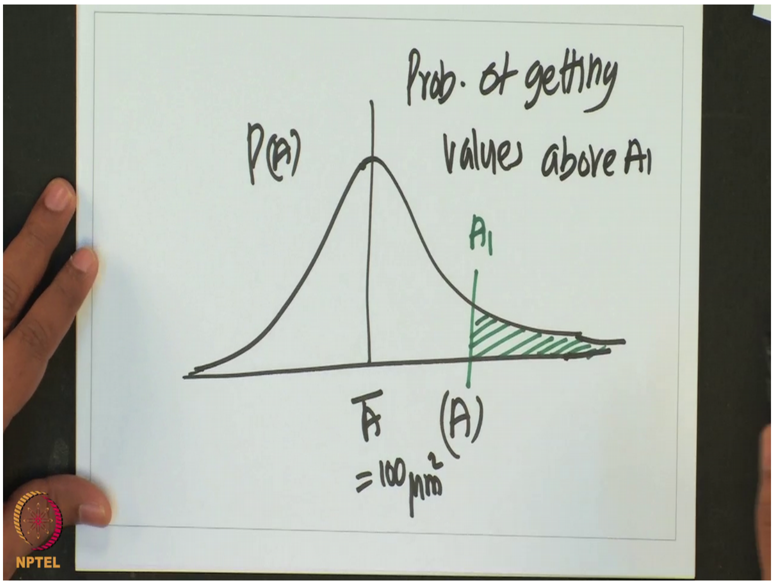
(Refer Slide Time: 13:35)



We have a bunch of cells and let us say we are measuring the area of each of this cell and if we plot this area P of A, that is what we are plotting and the mean is A bar and the standard deviation of sigma the expression for this P of a would be some constant which I would write now here called normalization. So, it will be e power minus x minus x bar whole square here x is A by 2 sigma square times the normalization constant which turn out to be 1 by root pi sigma always this a the normalization concern will turn out to be 1 by root 2 pi sigma this is the form of A.

So, this A is actually x. So, you know so if I have this measurement where I have lot of area data given.

(Refer Slide Time: 14:47)

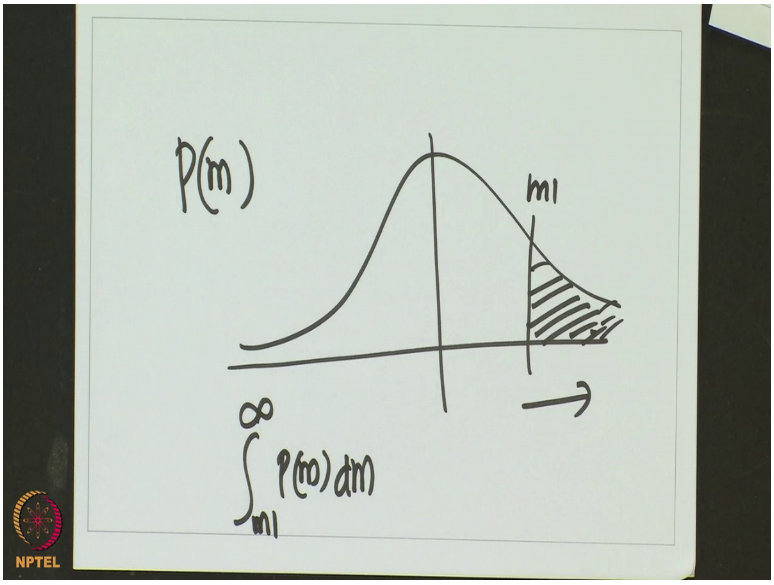


So, I have lot of area data and from this area data I plot this distribution, which is a normal distribution P of A versus A and this is my A bar. So, does a A bar is 100 micrometer square and the probability to get very small area, what does this curve mean this curve means probability to get very small area is small probability to get very large area is also very small probability to get area around this hundred micrometer square is very large that is what this essentially means.

Now, if I just look at this part alone what does this mean. So, let us say I call let me that let me call this number as A 1 A 1 is this x axis number the area corresponding to the line is A 1 this area means, this area under this curve means probability to have the data above A 1 probability to have data probability of getting values above A 1 that is what this means.

Now, this could be in x here or area way you could have any other variable of your interest it could be a height it could be size it would be gene expression. So, if the gene expression is in some unit m there is a number of.

(Refer Slide Time: 16:35)



If let us say you are measuring the amount of mrna P of m or the it could be some protein level all of this could have a normal distribution it could have mean and a curve around this.

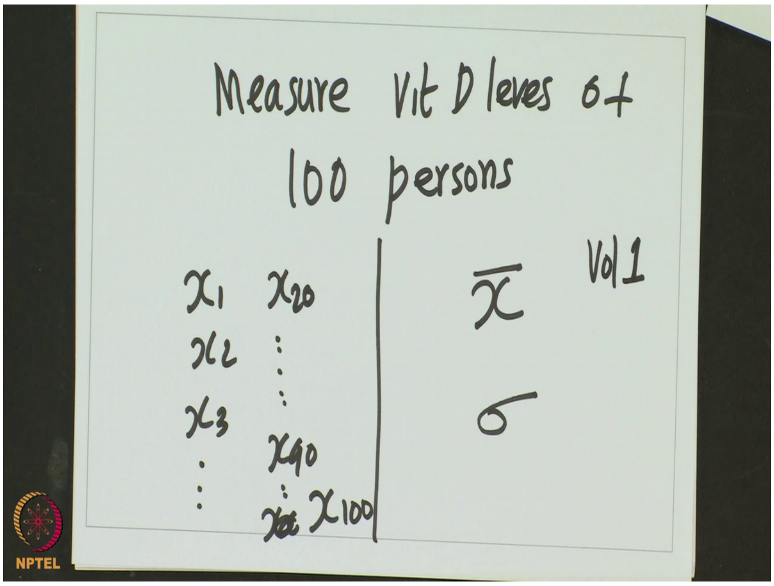
So, this is the first 0.1 should remember that this has this meaning and the area means the. So, here this area would mean the probability to get values beyond this that is what this mean and this is given by integral. So, if this is let me call this m 1 this integral is given by m 1 to infinity P of m dm if I do this this would give me this area and this has A meaning that the probability to get this number above this value m 1.

So, that is what and this is going to be very useful in understanding the normal distribution normal distribution has various uses, where are we going to uses when will a distribution be normal. So, typically when you have been experiment we cannot tell what the distribution would be a normal distribution would it be a some other distribution would be, what would be the nature of the distribution we cannot tell a priori before doing an experiment in most of the cases .

So, the variable that we are measuring that will be area, that will be length, in some other case it could be some medical quantity as we said the distribution of vitamin d levels in a population. If we measure some quantities like this we cannot tell what would be the distribution of this numbers that we get, but we can do 1 thing we can say 1 thing about the mean which is very often used in normal distribution what are we what are we measure for example, let us say we do a survey and measure vitamin d levels of population in different pockets of different parts of a city or different parts of a country.

So, let us say each so there are many people thousand volunteers going around and measuring vitamin D level of people in different parts of the country. So, let us say I am a volunteer my job is to meet hundred people and measure their vitamin D levels and so what I will have at the end of the day.

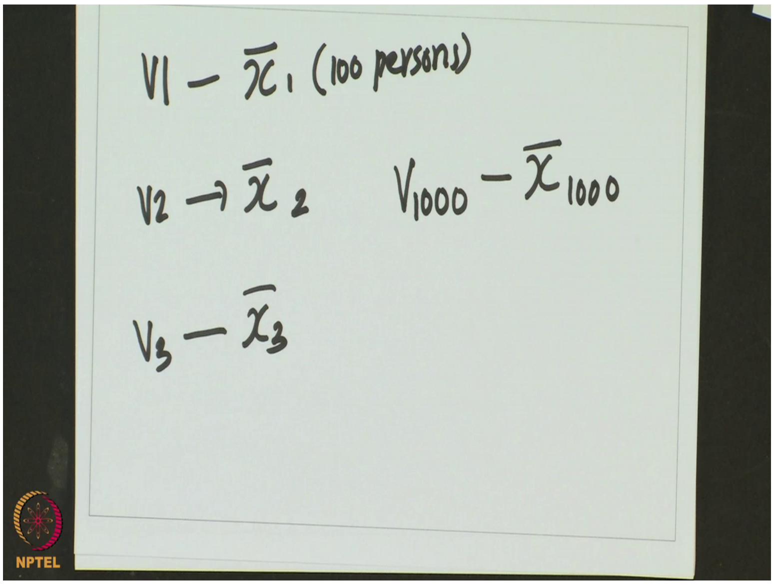
(Refer Slide Time: 19:19)



So, what are we doing I am going to. So, measure vitamin d levels of 100 persons, thus what I would do at the end of the day I will have data x 1 x 2 x 3 up 2 and so on and so forth, x 20 to 30 40 x 90 finally, x 100. So, finally, will have x 100 you will have 100 data points.

So, from this data point I can calculate in x bar and of course, a standard deviation also I can calculate. Now if I have this x bar and the standard deviation. So, I am a 1 I am volunteer one. So, I have 1 volunteer 1 I have an x bar. Similarly there are 1000 volunteers going around in different parts of the country. So, let us say volunteer 1 gets a each volunteer reports the mean value of vitamin d level of 100 people. So, what I would have at the end of the day.

(Refer Slide Time: 20:34)

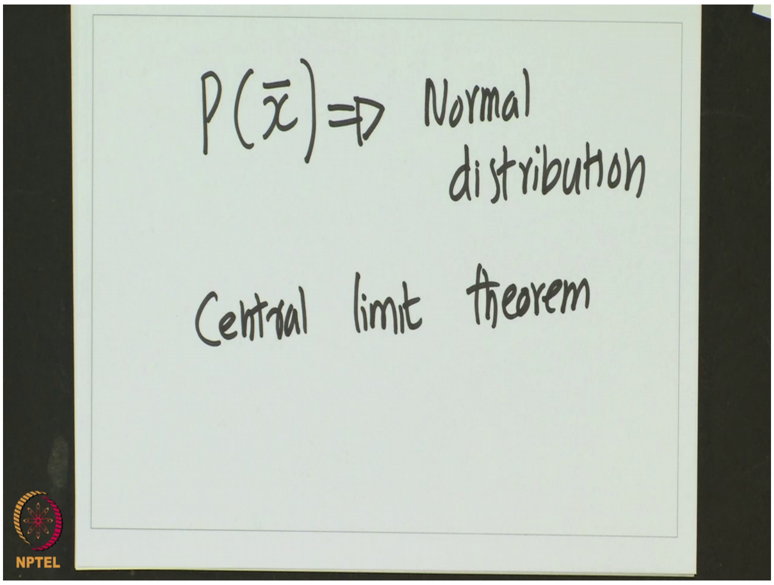


If all the volunteers come together the volunteer 1 would have x bar 1, the volunteer 2 would have an average value which is a different average, volunteer 3 will have another mean value because this volunteer 3 met different set of people.

So, all of this all got from hundred persons similarly there will be 1000 volunteers. So, v 1000 volunteer will have a mean value x bar 1000 this is also got from 1000 persons. So, 1000 people when then met 100 100 100 100 individuals and collected the vitamin D levels from them and each volunteer calculated a mean value.

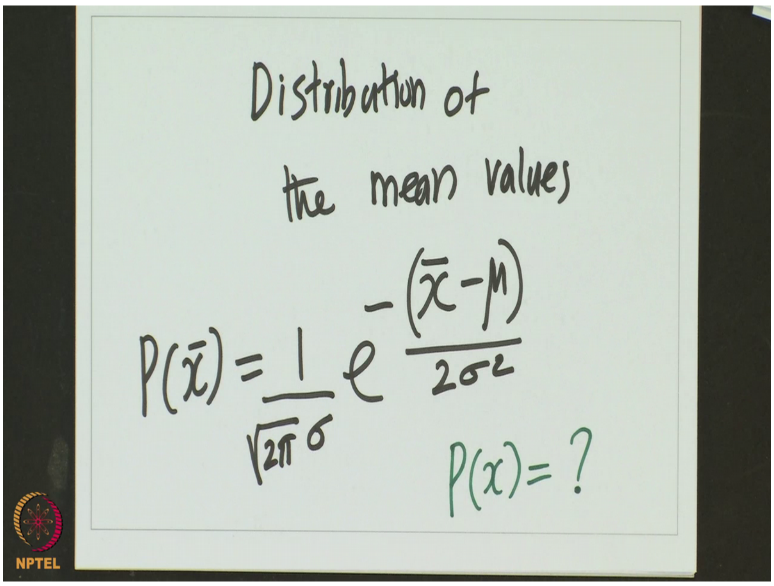
So, we have thousand mean values now we can plot the distribution of the mean value that is we can plot the distribution of the mean value itself.

(Refer Slide Time: 21:38)



So, P of x bar we can do this and it turns out that this always would be a normal distribution. So, there is a theorem in mathematics which says the distribution of the means would be a normal distribution; this is called a central limit theorem. So, it is using central element theorem 1 can argue that if I do many many experiments separately and each experiment we could get the mean value of some n samples, the dist if I collect all the mean values from different people and plot the distribution of the mean itself that would be a normal distribution.

(Refer Slide Time: 22:45)

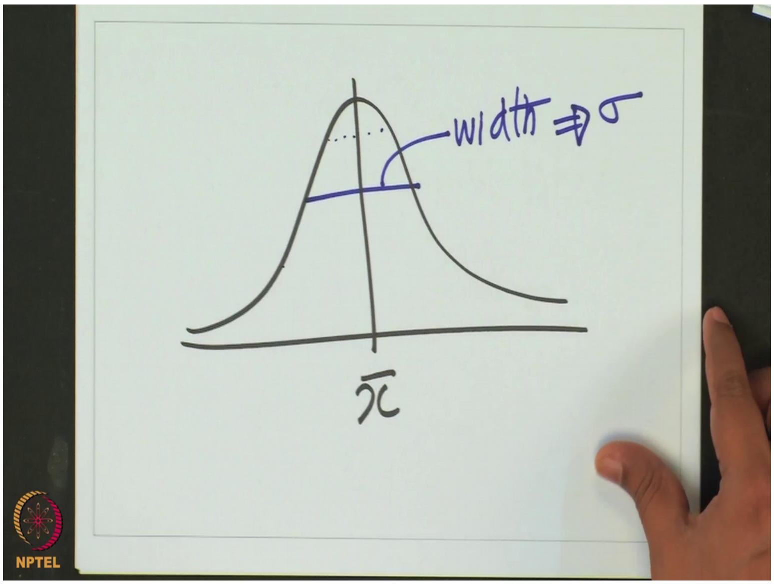


Now, how would we write a formula for the distribution of the means? So, let me write here distribution of the mean values distribution of the mean values I can right P of x bar of course, this also has to have a mean of the mean. So, I would write x bar minus mu divided by 2 sigma square 1 by route 2 pi sigma.

So, this would be the distribution of the mean itself now this is important remember in this case we do not know what is the distribution of the x. So, this is if I meet 100 individuals and measure vitamin D level what is the distribution of the vitamin D level I do not know irrespective of this the distribution of the means would be a normal distribution.

So, this is the case where normal distribution automatically emerges. So, whenever you have a set of measurements and you come you sum them and compute the mean the result would be a normal distribution. So, this is given to us and this is something that resists a very surely a normal distribution this is the case where we usually will get a normal distribution, also to remember that whenever we have 2 variables whenever we have 2 variables that is 2 parameters, which is mean and standard deviation and if x is a continuous variable then typically the distribution would be a normal distribution. It normally distributions only 2 parameters x bar and sigma which is the mean and the standard deviation, 1 more thing to remember is what is the physical meaning of sigma this is very important remember.

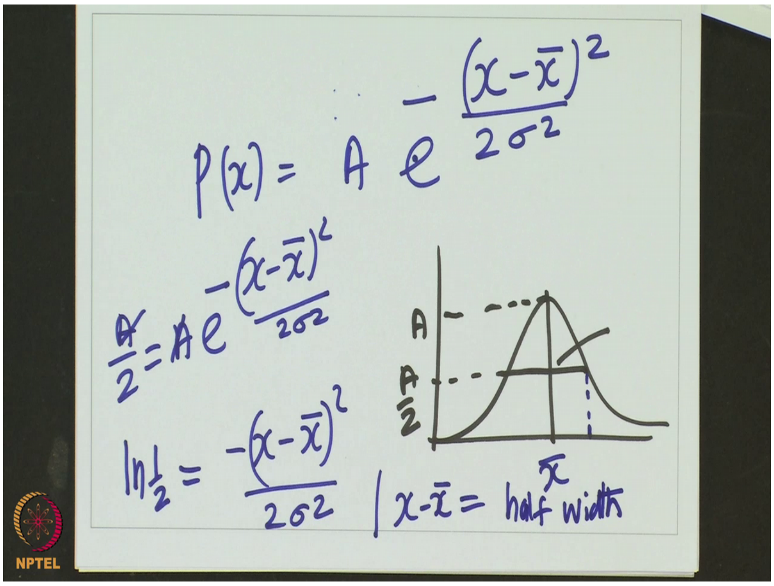
(Refer Slide Time: 24:41)



So, if I have a normal distribution this is my x bar. So, if I have normal distribution this is my x bar what is the physical meaning of sigma it turns out the sigma is the width. So, this width is related to sigma this width is related to sigma so, but this width we can measures at different places. So, where will we measure the width again measure the width here or I can measure the width here. So, if I measure the width of the distribution at some point I would get the sigma.

So, let us think about measuring the width of this distribution. So, the functional form that we have.

(Refer Slide Time: 25:34)

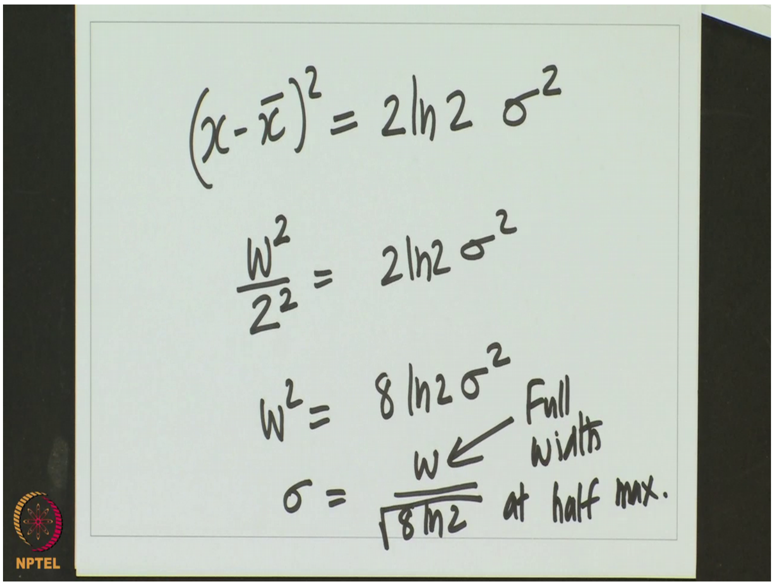


So, I would write the functional form which is P of x is a e power a e power minus x minus x bar by 2 sigma square this is the this is my normal distribution function now if I measure the width at half of it is peak. So, when x equal to 0 of course, when A when x equal to x bar when x equal to x bar this is 0 e power 0 is 1 and P of x will be A. So, this would be my when x equal to x bar this is my A if I measure the width at some point half A somewhere half the maximum.

So, if I measure the width at half maximum. So, let me draw that here I have this this values A if I measure the width that A by 2, if I measure the width here which is A by 2. So, if P is A by 2 let us substitute A by 2 here. So, what I would have P is my A by 2. So, a by 2 is A e power minus x minus x bar whole square by 2 sigma square. So, you would get this A cancels I can take log on both sides . So, log of half will be log of half on the left hand side which will be log of e power minus something which is minus x minus x bar Whole Square by 2 sigma square.

Now what is x minus x bar this is x minus x bar this is my x bar and this is my x. So, x minus x bar is this. So, half width is x minus x bar. So, half width is x minus x bar. So, x minus x bar is half width. So, if I take if I simplify this, what would I get?

(Refer Slide Time: 27:52)



What I would get is that x minus x bar whole square is equal to log 2 log 2 sigma square that is what I would get 2 log 2 sigma square, this is my half width. So, I would write w by 2 whole square is equal to 2 log 2 sigma square, which would mean that if I take everything to 1 side what do I get the full width which is w w square is 2 square which is. So, 8 log 2 sigma square in other words sigma is w by 8 log 2.

So, this is root of this of course, root of this and this is what I would get and what can calculate if I know the full width at half maximum. So, the w is full width at half maximum. So, if I know the full width at half maximum I could get sigma by using this formula. So, it is basic basically sigma is related to the width and if I know the full width at half maximum I can get the sigma.

So, to summarize we have a distribution which has a mean and a standard deviation and given these mean and standard deviation we can construct a distribution which is called normal distribution typically and this distribution has a few properties. Of course, peak at x bar the mean and it is symmetric and the width is essentially the sigma or width is related to sigma and typically when you provide the distribution of the means it will always be a normal distribution these are the things to remember with this we will stop this lecture and continue in the next lecture. Bye.