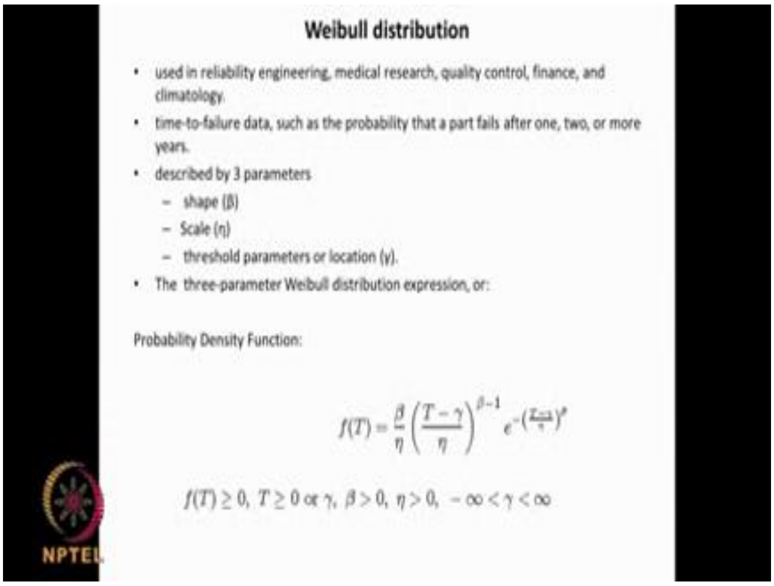
Biostatistics and Design of Experiments
Prof. Mukesh Doble
Department of Biotechnology
Indian Institute of Technology, Madras

Lecture - 26
Weibull distribution

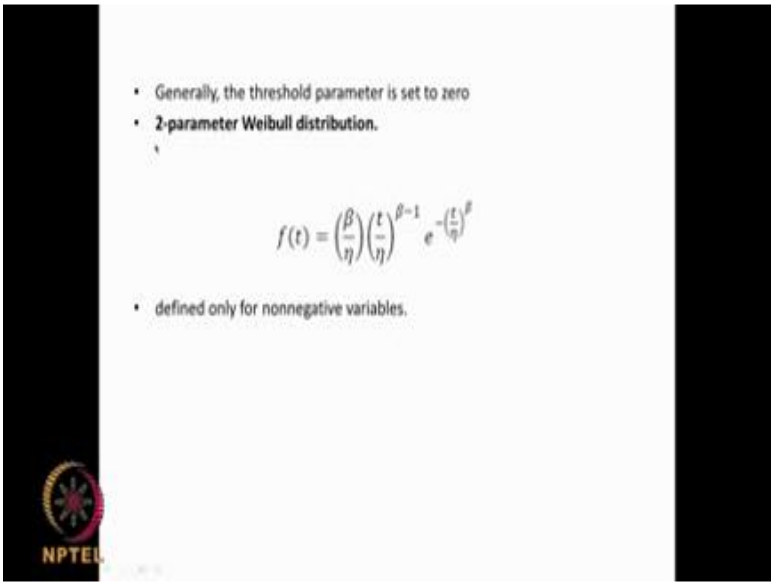
Welcome to the course on Biostatistics and Design of Experiments. We introduced new distribution in the previous lecture it is called the Weibull distribution, we will continue on that.

(Refer Slide Time: 00:23)



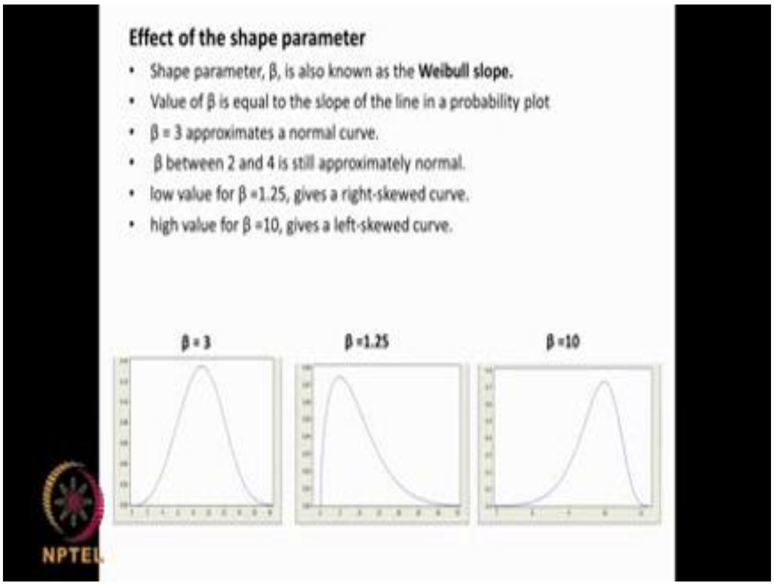
Weibull is a very useful distribution, it gives in reliability of a part, bio material for example. It can be used in quality control, you are manufacturing some material I want to know what is the failure rate, I want to know when each material will fail on average. It is a time to failure data, how long it will take to fail. You can do this type of calculation. So, will you take 1 year, 2 years or more than that? There are 3 parameters in Weibull distribution function or sorry, Weibull probability density function. • β that is the shape, scale is the • η and other one is the location or threshold • γ and the probability density function looks like this.

(Refer Slide Time: 01:12)



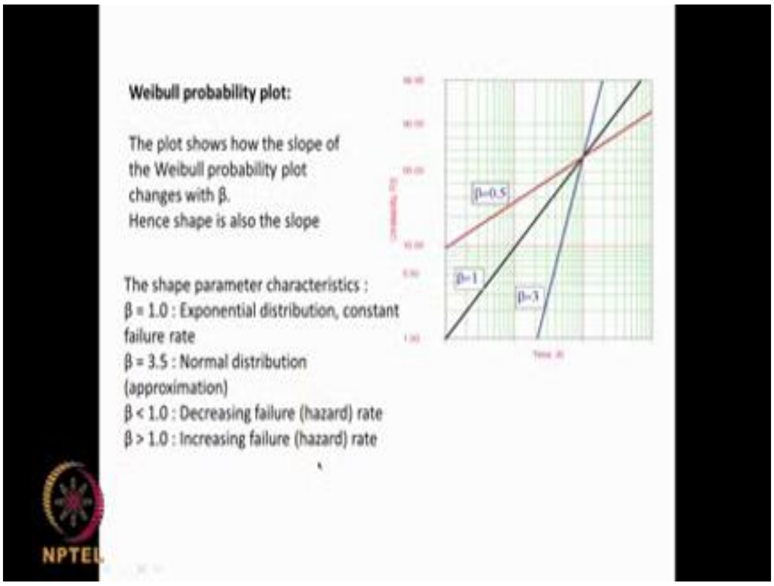
We can make γ = 0. Then it becomes a 2-parameter density function, β and η. β is your shape and η is your scale, both are non-negative that means they cannot be, they are > 0.

(Refer Slide Time: 01:33)



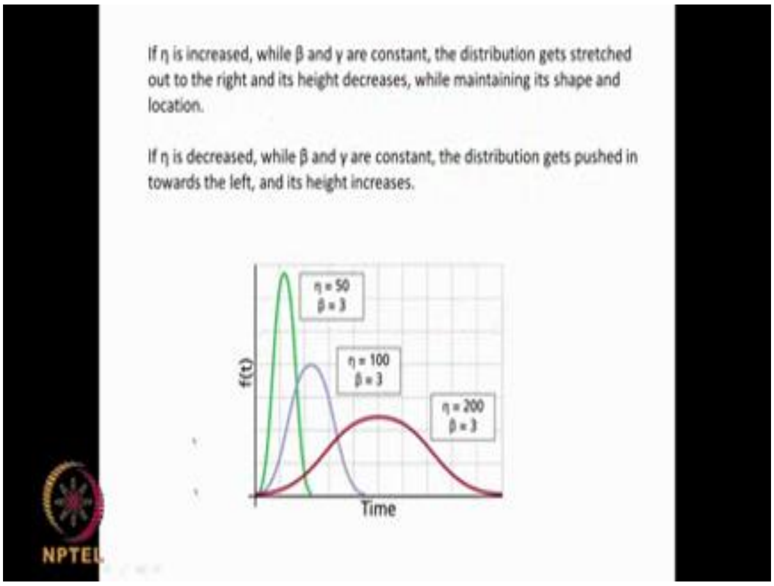
The shape parameter tells you how the curve looks like. Approximately around 3, the Weibull distribution if β = 3, Weibull distribution looks like a normal distribution. If it is 1.25, it is sort of right skewed and if it is large number like 10, it is left skewed. So, it is leaning towards the right-hand side, leaning towards the left-hand side.

(Refer Slide Time: 02:05)



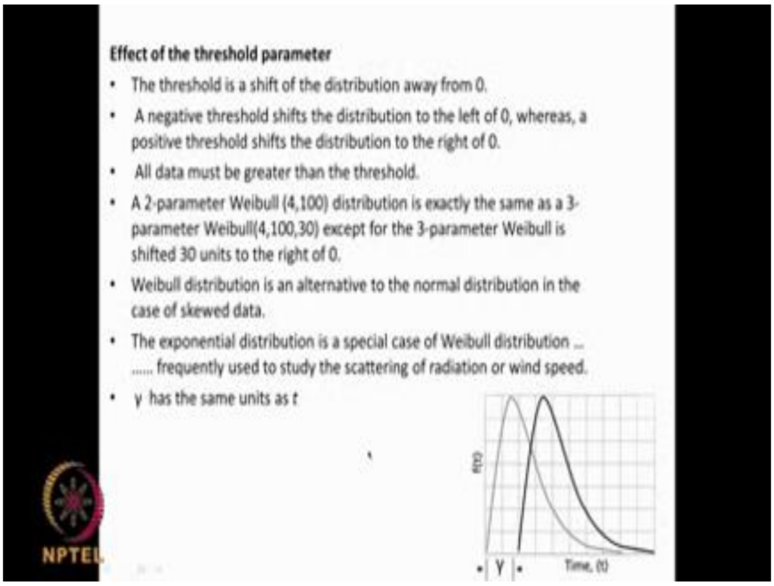
Then if we look at the probability it is called the unreliability and the time here. For different values of β, β = 1, goes like this is called the exponential distribution that is constant failure rate, $\beta < 1$ like this you know decreasing failure rate that is with time the failure rate goes down and $\beta > 1$ and say for β = 3 increasing failure rate that means failure rate will increase with time. Generally, we will observe in many situations like this actually failure rate keeps increasing with time because of wear, tear and so many other factors, age we sometimes call it age right?, because of the age the probability of the equipment failing also keeps going up and up.

(Refer Slide Time: 02:54)



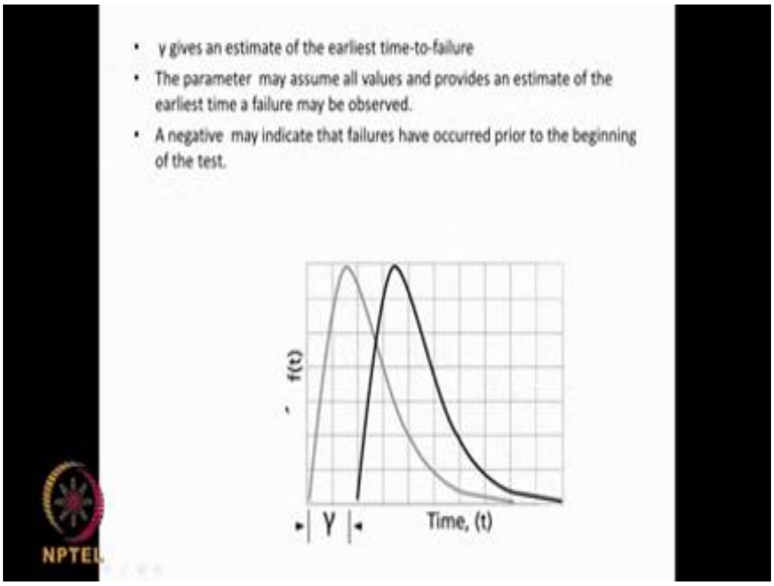
If you look at keeping β constant if we keep changing η, we see the graph gets pushed to the right, so the mean value gets pushed to the right and the peak also keeps going down, you can see this very clearly. So, the peak also goes down and the mean gets pushed to the right. Larger the η, we are trying to prolong the failure rate that is what it means actually.

(Refer Slide Time: 03:39)



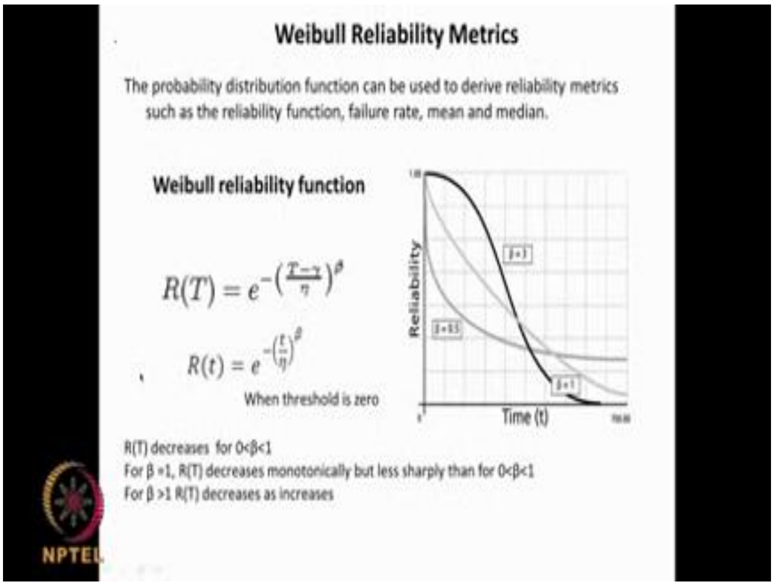
Next one is a threshold that is the η, what η does is it shifts the entire graph to the right-hand side I mean to the right, it shifts. As you can see if γ = 0 the graph will start from the origin whereas any other γ it will start from that particular place and go, γ does only that just shift the whole thing. So, γ has a units of a time your β also has units of time, both are units of time. Time could be hours or week or minutes or years and so on actually. A 2 parameter Weibull like 4,100 distribution is exactly same as a 3 parameter 4,100,30 where this 30 all it does is it shifts the graph from 0 that is the origin into the 30th point. Weibull distribution is an alternate to the normal distribution in the case of skewed data. If you have a skewed data then we can use a Weibull distribution for that actually and the exponential distribution is a special case of Weibull distribution because exponential distribution we say β = 1, we have constant failure rate, γ has the same unit as time.

(Refer Slide Time: 05:03)



It tells you γ gives an estimate of the earliest time to failure. Suppose I say if γ is 10 then we can say your one of the parts will fail it will start failing from the 10th hour. It is the earliest time for failure. A negative may indicate that failures have already occurred prior to the beginning of the test also that is, if γ is negative so it the graph is starting from some other place here right below the origin, obviously that means failure have already occurred before we start our test. These are the 3 parameters in Weibull distribution and these are the meaning of these 3 parameters. We will look at some more term terminologies.

(Refer Slide Time: 05:54)



There is something called Weibull reliability function is given by this,

Weibull reliability function

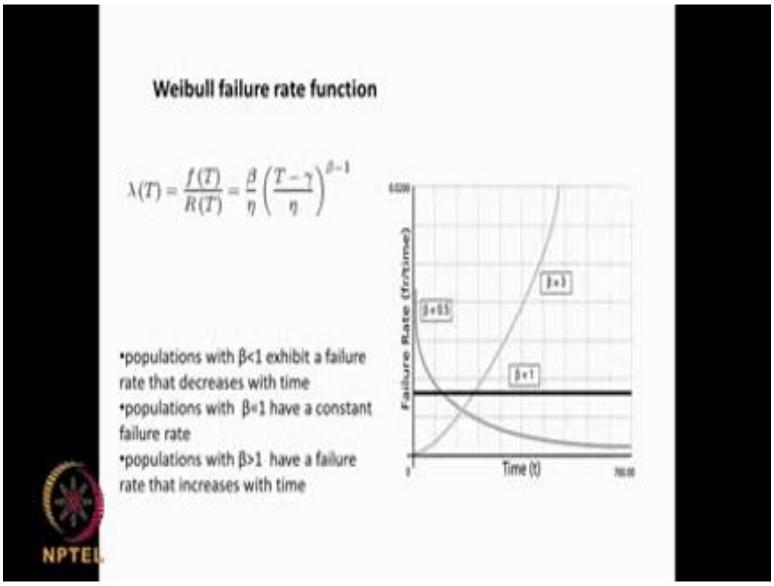
$$R(T) = e^{-\left(\frac{T-\gamma}{\eta}\right)^{\beta}}$$

whereas as usual, if I remove the threshold I put it as 0 then you will end up having

$$R(t) = e^{-\left(\frac{t}{\eta}\right)^{\beta}}$$

When do use it? It can be used to derive reliability metrics such as reliability function, failure rate, mean, median and that sort of situation. R (T) decreases for • β lying between 0 to 1, for • β = 1, R (T) decreases monotonically like this with time. So the reliability of any part will keep coming down with time that is obvious, right. Even in your house you have light bulb, new light bulb reliability is very high, after 1 or 2 years you are very scared because you think the bulb may fail that is true. Even with your mobile phones, with any material, generally the reliability will keep going down monotonically when β = 1. When β > 1 the reliability decreases. So β > 1, the reliability R (T) decreases and then it can also increase. If β, if we have a β is say 0.5 that is < 1 it goes down but later on it sort of stabilizes. So sometimes we see very very old television keeps working after sometime you will not know how long it has been working for, that could be β is equal to this type of number β < 0.1.

(Refer Slide Time: 08:29)



There is another term which is called Weibull failure rate, it is called Weibull failure rate function it is given like this,

$$\lambda(T) = \frac{f(T)}{R(T)} = \frac{\beta}{\eta}\left(\frac{T-\gamma}{\eta}\right)^{\beta-1}$$

. In the previous case we looked at the reliability which is given by this relationship whereas this is Weibull failure rate which is given by this relationship failure rate function. It is a combination of *f (T)* and *R (T), R (T)* is your reliability function and *f (T)* is your original Weibull distribution. So *f (T)* / *R(T)* is your failure rate function that is failures per time, as a function of time here. It is called the failure rate function. When • β < 1 obviously it will keep going down and down and down like that, the failure rate will decrease its time. Initially it may fail much more, failure rate is much higher but after a very long time the failure rate is much lower actually but • β = 1, the failure rate is constant this called a constant failure rate, • β = 1 because when we put • β = 1 this become 0, so the whole term becomes 1. We have only • β ÷ • η, • β = 1 it goes like this. Now if • β > 1 then • β > 1 the fail here you have a say • β = 3 means this terms become square obviously it takes up this type of shape. It is increasing dramatically that means the failure rate increases with time. Initially if you expect it to have a failure rate of something 2 per time ,2 per second or 2 per week or something, with time it may go dramatically very high. It may go from 2 failures per minute to 100 failures per minute because you get a sharp increase in the failure rate. This is called the failure rate function. This is also very important and the previous one is called the reliability function how reliable the material is, *R(T)*. We can also calculate unreliable by subtracting this from 1 whereas the previous one this particular graph is called the Weibull 2 parameter or 3 parameter Weibull distribution function, Weibull distribution probability function.

(Refer Slide Time: 11:45)

All these are useful when we start looking at problems and then there is another term that is called the mean time to failure MTTF, mean time to failure. What is the mean time? Like in clinical when you are doing drug discovery we say concentration at which 50 % of cancer cells die, what is we call it as IC_{50}, concentration at which 50 % of bacterial growth is inhibited we call it as IC_{50}, concentration at which 50 % of the animals die we call it ID_{50}lethal dose 50. So similar to that we can calculate also 50 % life from this formula but this particular equation gives you the mean time to failure. We have a • γ, we have • η, we have • β and so if you want see what is the failure rate what is a failure for 50 then you end up having this type of relationship, right?. Now this is called a average time, mean time = • γ + • η this is a • γ function and • γ function is defined like this • γ

$$\Gamma(n) = \int_0^{\infty} e^{-x} x^{n-1} dx$$

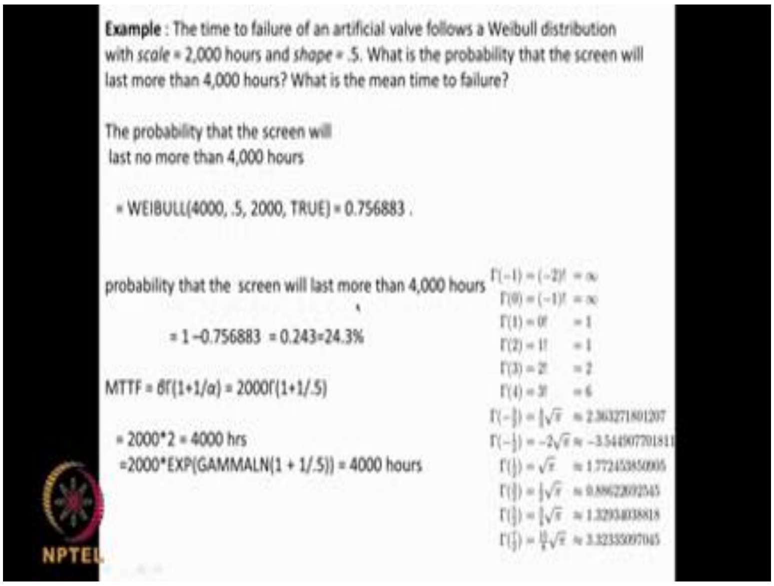
. So you do not have to get tense about it because there is a table which gives you for • γ 1 is 0 factorial that is 1, • γ 2 is 1 factorial that is 1, • γ 3 is 2 factorial that is 2, • γ 4 is 3 factorial that is 3 * 2, 6 and so on actually. • γ 0 is - 1 factorial which is infinity and • γ 1 is - 2 factorial again this is infinity and so on actually. We can use this particular table and do our calculations, nothing to worry about it because we are going to look at some problems also as we move along.

(Refer Slide Time: 14:03)

There is a Excel function which is available it is called Weibull $x, \alpha, \beta,$ cumulative. So α and β are the shape and scale parameters here they use different name shape and scale parameters. In our original problem we used a shape and scale parameters β and • η but in Excel say you call it α, β. No problem. So x is the value at which you want to calculate the Weibull should be $\geq$ 0 cumulative true or false, if we give it as false then gives you the value of the probability density function $f(x)$ at x, if we give it as true then it gives you the Weibull cumulative distribution function at x. If you give true it gives cumulative, if you give false it gives the actual value.

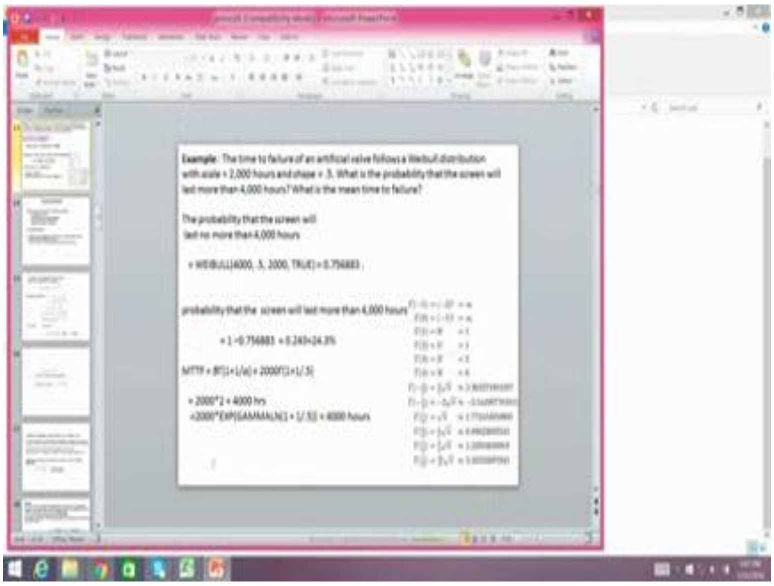
We look at some problems as we go long. We have defined quite a lot of terminologies here one is called the mean time to failure and other is called the Weibull failure rate function and other one is called the Weibull reliability function.

(Refer Slide Time: 15:24)



Let us look at some problems that will make things much clearer, the time to failure of an artificial valve follows a Weibull distribution with scale = 2000 and shape = 0.5, shape is 0.5, scale is 2000. What are they, scale is η, shape is β, this given as 2000 and 0.5. What is the probability that these artificial valves will last more than 4000 hours and now what is the mean time to failure? So the probability that the artificial valve will last no more than 4000 hours we can calculate Weibull, x value is 4000 and as you can see first is the shape, second is the scale, the shape is 0.5, scale is 2000 then if you give true it gives the total probability for failure. If you want to calculate probability that it will last more than, we have to subtract from 1, so the probability is 24.3 minutes sorry 24.3 %. Probability that the screen will last more 4000 hours is 0.243.

(Refer Slide Time: 16:54)



Let us look at the Excel function also, we say Weibull. First is we want know what is a probability that it will last probably the screen last more than 4000 hours. So 4000 is the first term comma the shape we have to give the shape first and then we have to give the scale next and then we are giving it as true because we want to get the total that gives you 0.7563. The probability that the problem is giving last, probability of failure for less than is given by this because where if we look at the Weibull it gives you this value cumulative distribution. We want to know the last thing that means we have to do 1 - this, 1 - this will give you 0.243 that means probability that these artificial valves will last at least 4000 hours is given by 0.243. What is the mean time to failure? How do you calculate? There is again a formula here mean time to failure is given by this formula

$t = \cdot\ \gamma + \cdot\ \eta\ 1 / \beta + 1$. So we will say we can give 1 ÷1 + 0.5 is the $\cdot\ \gamma$ function of this and then this is the valve we will neglect this particular $\cdot\ \gamma$ here and then we will put this 2000 here. So

2000*EXP(GAMMALN(1 + 1/.5)) = 4000 hours

. We need to remember that this β is different from that β because this β is different from our β here actually, so do not get confused that η is same as this β. What we will do is, we have 2000 here and then γ function because the equation is given like this, so here we will neglect this small γ. So we will have this into gamma function of 1 by β +1, γ function of a 1 + 1 / by 0.5 and as you know 1 + 1 / 0.5 is nothing but 2 + 1 is 3, γ of 3 is given by 2 and then on this side here 2000, so 2000 * 2 is 4000 hours, mean time to failure is 4000 hours actually.

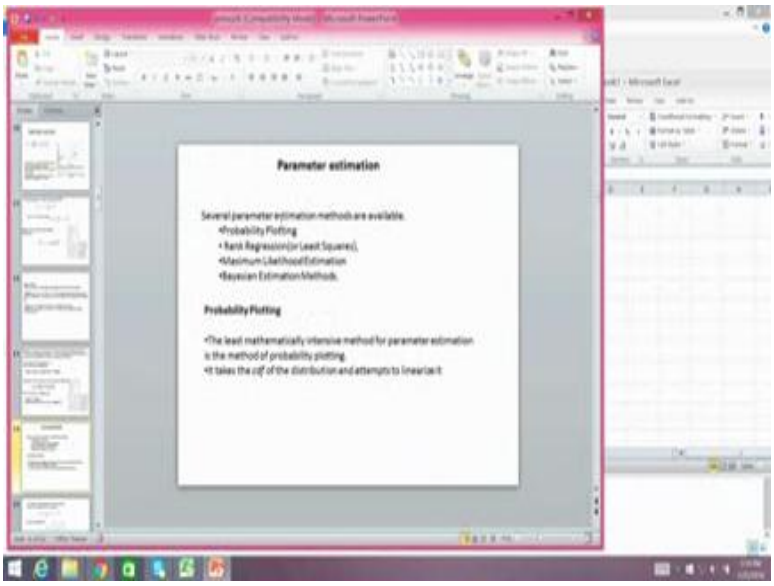
We can also use a command called γ l n in Excel that also can be used to calculate this same thing actually. Let me once more repeat, we have the mean time to failure is given by this formula we can neglect this γ if you want to have a 2 parameter Weibull. This is η as I said is nothing but scale and β is a shape. In our problem scale is given by 2000 hours, shape is given by 0.5. What you do? You take 2000 here γ function of 1 / 0.5 + 1 that will become γ (3) and γ (3)is given by 2 factorial that is 2 and your scale is 2000, so 2000 * 2 that will give you 4000 hours, the mean time to failure is 4000 hours. Do you understand and as I said there is a command also called a GAMMALN which can be also be used in Excel combination when you combine it x exponent of gamma l n you will get the same answer. In this particular problem we are trying to look at probability that the artificial valve will last more than 4000 hours and as you can see here probability that artificial valves fail less than 4000 hours will be 75 percent. So 1 - such 75 % will be about 24.3 % and then mean time to failure we used this particular equation we remove this γ call it 0 we have the scale factor and we have the shape factor. So this scale factor is given by 2000 hours, shape factor is given by 0.5 we say 2000 multiplied by γ function of 1 + 1 /0.5 and γ function of 3 is 2 factorial. So 2000 *2 that is the 4000 hours is the mean m t t f, mean time to failure.

I understand how to go about doing these types of problems they are quite useful actually to address, you do not forget that. In the Excel also we can do using this Weibull but Weibull Excel cannot do the MTTF type of calculation. For MTTF we can combine the 2 commands x and GAMMALN together to get the m t t f but it is not very difficult to do the MTTF because by using this table the γ function of various numbers are given γ of 1 is 1, γ

$$
\begin{aligned}
\Gamma(1) &= 0! &&= 1 \\
\Gamma(2) &= 1! &&= 1 \\
\Gamma(3) &= 2! &&= 2 \\
\Gamma(4) &= 3! &&= 6
\end{aligned}
$$

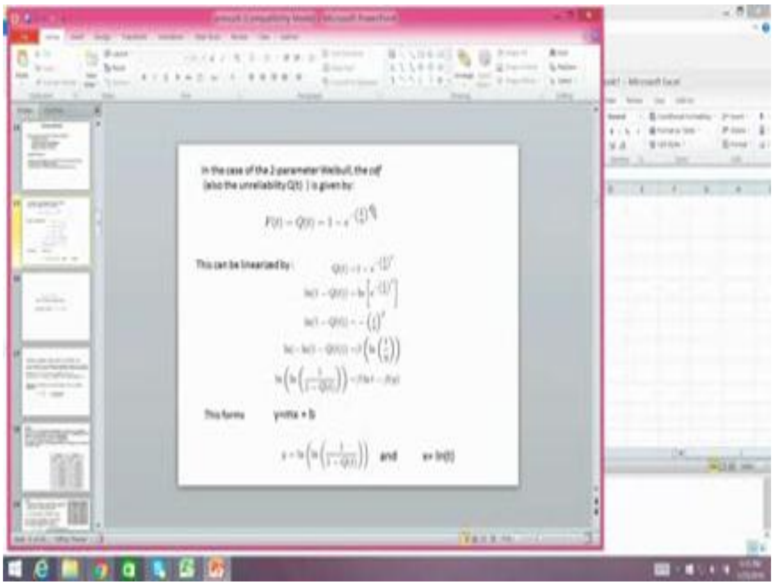
and so on actually. It is very simple and straight forward for you to look at the mean time to failure and also to calculate probability that the artificial valve will last for more than 4000 hours. As you can see it is a very useful type of distribution which can be applied to many different situations actually.

(Refer Slide Time: 24:05)



Now let us proceed, you have these parameters we have the scale parameter, the shape parameter and also the other third small γ parameter. How do you estimate all these? There are different approaches by which you can estimate one is called the probability plotting,other is called the rank regression or least square approach, maximum like estimation, Bayesian estimation method and so many methods are there but this is the most simplest probability plotting least mathematically intensive method for parameter estimation. It is quite simple and it tries to linearize this function and then it tries to calculate. So from the data how do I calculate the scale parameter and the shape parameter that is the question actually. There are, this particular approach by which one could do that. If you have 2 parameter Weibull then we can calculate the unreliability matrix. What is unreliability matrix? 1 - that is called the unreliability matrix.

(Refer Slide Time: 25:20)



As I mentioned before that 1 -, so if you have the reliability

$$F(t) = Q(t) = 1 - e^{-\left(\frac{t}{\eta}\right)^{\beta}}$$

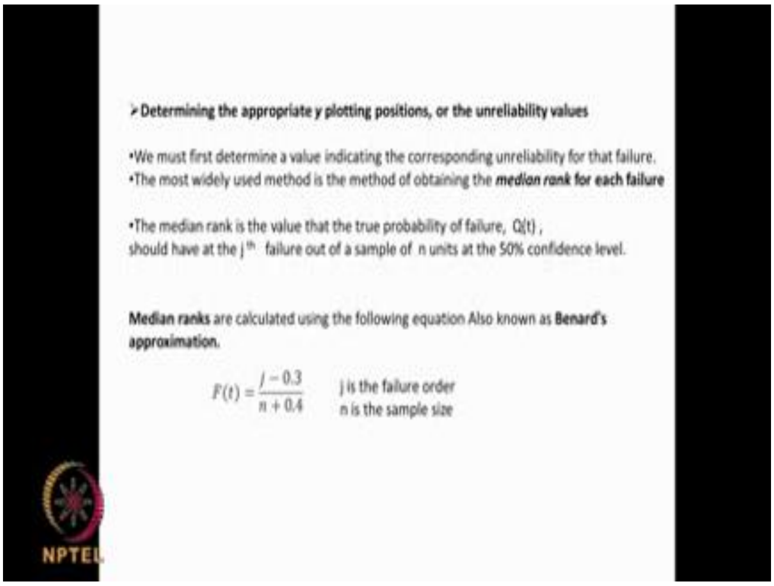
of that is called the unreliability matrix, this can be linearized. If I linearized this, this, this all of you know how to linearize this you end up with term like this,

$$\ln\left(\ln\left(\frac{1}{1-Q(t)}\right)\right) = \beta \ln t - \beta(\eta)$$

$Q(t)$ is this unreliability matrix $F(t) = Q(t) = 1 - e^{-\left(\frac{t}{\eta}\right)^{\beta}}$. Now this looks like y is equal to m x plus b type of equation where your x is equal to time. If I have data of $Q(t)$ versus t all I have to do is plot this versus this and the slope will give me, sorry plot this versus ln (t), do not forget, the slope will give me β and the intersect should give me the other one the - β · η that is how you do it. So we have the unreliability data q t as a function of time, we calculate this by taking 1 / 1 - Q(t) ln then again ln and then plot that against logarithm of t and the slope should give you β and intercept should give you this particular term. So that is another approach by which we can calculate these parameters in the model.

Let us look at some problems actually later on as we go along because I think like doing problems which give you a lot of ideas about how to go about.

(Refer Slide Time: 27:25)



We can determine by plotting these or the unreliability values and then we can achieve this type of calculations and results actually. So we will continue more in the next class.

Thank you very much for your time.

Key Words: Weibull distribution, Reliability, Reliability function, Mean time to failure, probability, weibull failure rate function, parameters, unreliability matrix, parameter estimation