

Engineering Econometrics
Prof. Rudra P. Pradhan
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

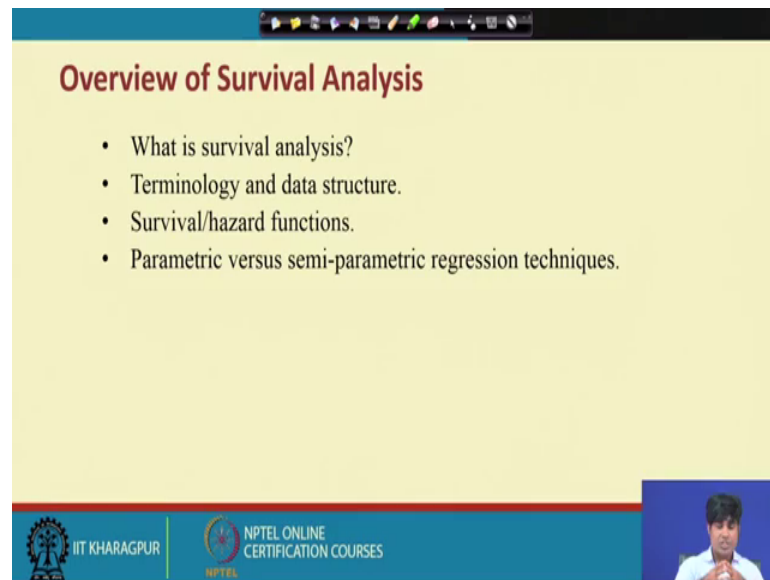
Lecture - 60
Fitting Models to Data (Contd.)

Hello everybody, this is Rudra Pradhan here. Welcome to Engineering Econometrics, and welcome to the last lecture of this particular you know series; that to the kind of you know concept called as you know; survival analysis. It is a part of actually count data modelling. And in fact, we have already discuss the concept of count data and discrete data. And we have discuss 2 typical you know models relating to count data modelling; that to the concept of you know Poisson regression modelling and the concept of you know negative binomial regression modelling.

And we have already a highlighted various, you know requirements and the structure through which we can pick up a particular you know models as per the particular requirement and the kind of you know with respect to data structure. And the again that is to whether it is you know Poisson regression modelling or negative binomial regression modelling. In response to Poisson regression modelling and negative binomial regression modelling we like to analyse a case, which is called as you know survival analysis which is actually part of this count data modelling.

And then we like to discuss what is this particular concept and how we can actually you know bring this situation and that to connect with, you know Poisson regression modelling and negative binomial regression modelling.

(Refer Slide Time: 01:55)



Overview of Survival Analysis

- What is survival analysis?
- Terminology and data structure.
- Survival/hazard functions.
- Parametric versus semi-parametric regression techniques.

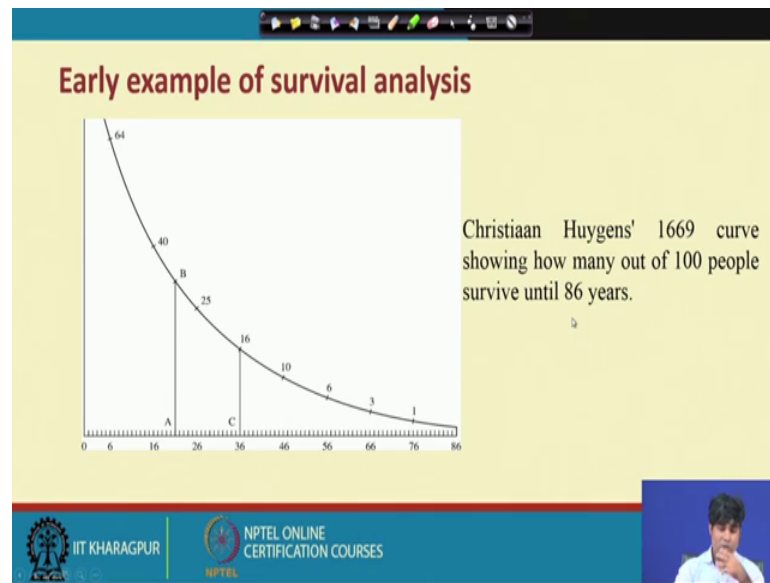
The slide is part of an NPTEL lecture from IIT Kharagpur. It features a blue header with navigation icons, a yellow main content area, and a blue footer with the IIT Kharagpur and NPTEL logos. A small video inset in the bottom right corner shows a male lecturer speaking.

So, here the overview of this particular lecture is like this. So, first of all we like to understand what is survival analysis, and what are the kind of you know data and that means, of course, since we are discussing count data modelling and discrete data modelling.

The choice of the particular model exclusively depends upon you know data structure and the features of the data. So; that means, first end we have to understand whether the data is actually count data type that to non-negative t integers type and against, we like to check actually whether mean and variance are equal or mean and variance are you now unequal. That is the first to check once through which you know the entire count data modelling will be applied.

Then we bring a kind of you know concept called as you know survival functions and which you called as you know hazard functions. And then we will list touch upon parametric versus semi parametric you know regression technique which is which will be also under the count data modelling.

(Refer Slide Time: 02:52)

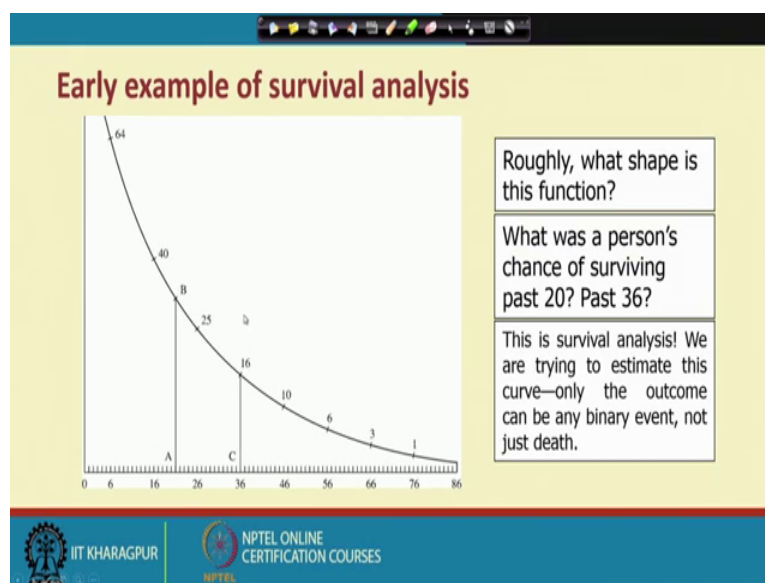


Usually survival you know theoretically you understand what is actually you know the concept you know survival.

So, that is how we have already discussed you know the typical models, you know count data models are exclusively used in a kind of, you know situation like you know that death incidents a you know kind of you know health issue. So, means see typically where you know the concept of survival is very much you know meaningful right.

So, basically you know death occurrence you know or accidents occurrence, these are the situations where you know this kind of you know modelling is very frequent. And a we can deploy you know such kind of models in these you know areas, that is why we are you know discussing this particular you know analysis. So, the usual stand of this you know survival analysis would be like this, you know it is actually over the time it is actually declining.

(Refer Slide Time: 03:48)



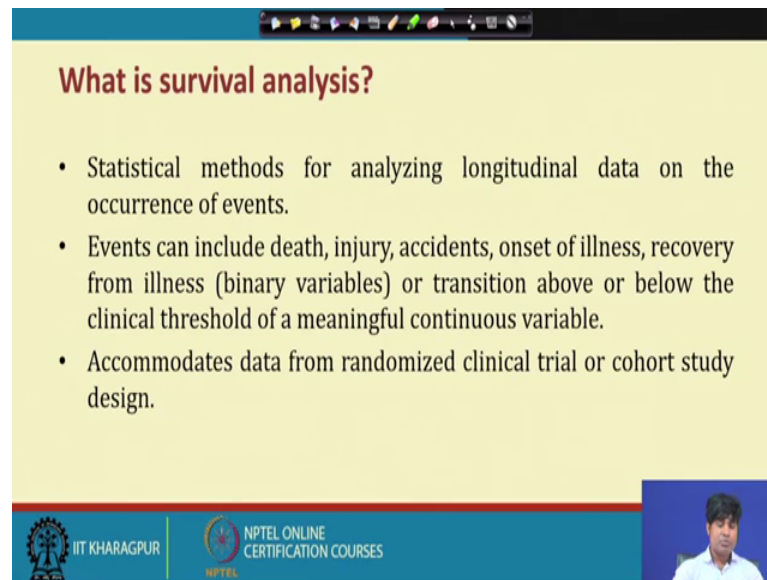
And a roughly, the question is the, what shape is this functions right? And a against what was a person's chance of you know surviving past 20 years you know past 36; that is what the prediction is about because you know somehow it is connected with you know life cycle hypothesis. And this is survival analysis we are trying to estimate this curve only the outcome can be any binary event not, just you know death.

That is what the beginning of this you know survival analysis and the kind of you know game. So, you know it is kind of you know yes no whether you know death occurs or you know death not occurs, that is how the.

So; that means, this particular you know analysis or this particular you know modelling is again a special case of you know count data modelling ok; where you know since we have started with you know count data modelling with you know 2 requirements that to non negativity and integer types, against it may be 0 1 that is the binary structure or it may be you know something different then you know you know from the 0 1.

But 0 1 type of you know you know like integer programming. So, we have a 0 1 kind you know statistical modelling or econometric modelling which you typically called as you know survival one way you called as you know survival analysis, and another way we have already discussed in you know non-linear from premo or that is called as you know binary search model. So, let us see how is that you know survival analysis structure.

(Refer Slide Time: 05:23)



What is survival analysis?

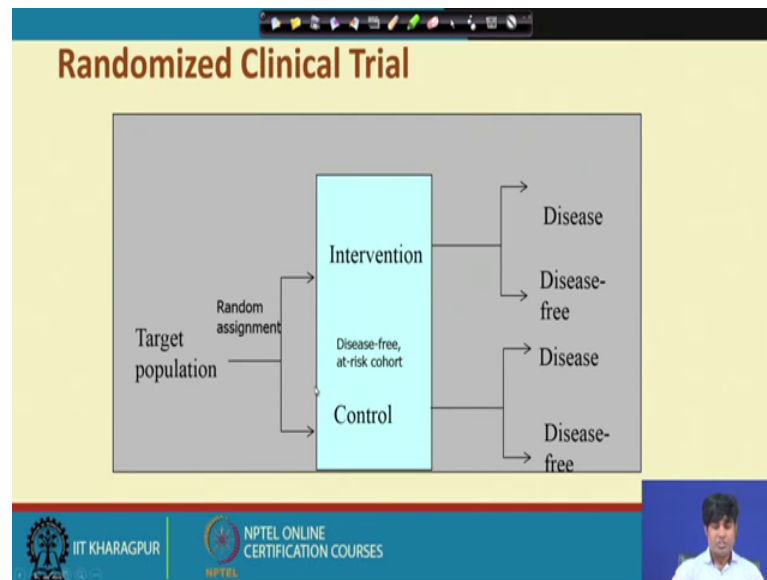
- Statistical methods for analyzing longitudinal data on the occurrence of events.
- Events can include death, injury, accidents, onset of illness, recovery from illness (binary variables) or transition above or below the clinical threshold of a meaningful continuous variable.
- Accommodates data from randomized clinical trial or cohort study design.

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, first of all what is exactly the survival analysis concept. So, basically statistical methods you know it is the method for analysing the count data on the occurrence of you know events. And what I have already highlighted the events here it can be actually include death injury accidents; onset one kind of you know illness and recovery from illness; that means, basically it is a binary type of you know situation. But must you know crucial kind of you know touch point is the question of you know survival.

So, injury means yes you know yes or no, I seriously injured not injured accident happened not happened onset of illness yes no, yes no type of things. So, then you know that I will be actually you know from the randomized you know kind of you know situation and then the story can be actually applied and then you can analyse as per the requirement.

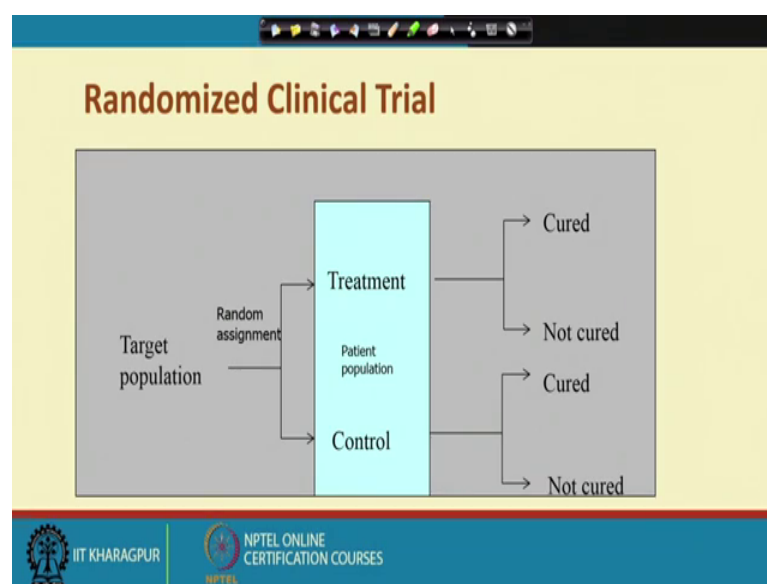
(Refer Slide Time: 06:22)



So, let us bring some kind of you know example, typical examples where you know what should be the target population and the kind of you know sampling.

So, here you see here disease free at risk a situations where we have 2 different divisions interventions you know control again disease and disease free. So, again in the control disease and disease free; that is means actually we need to bring a path diagram such a way that; this model can be apply as per the particular requirement.

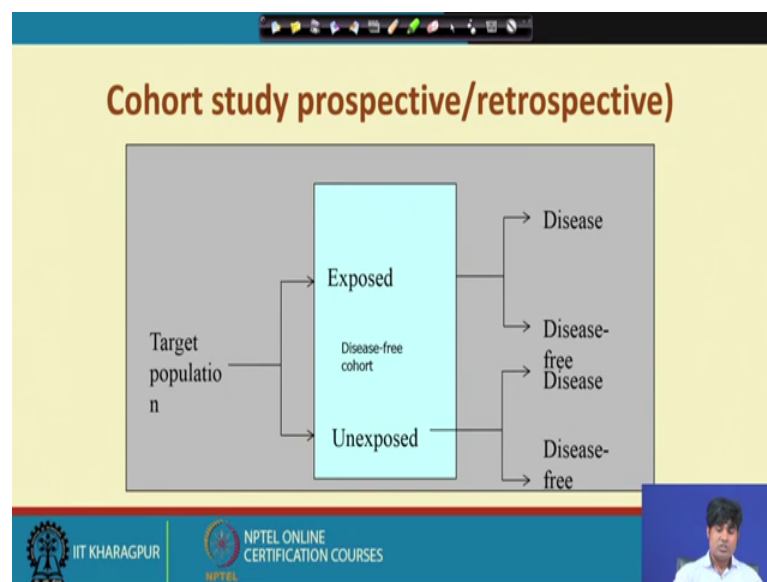
(Refer Slide Time: 06:54)



Against so the randomized assignment is actually you with two divisions of course, population related to patient and then treatment and control. So, with you know treatments we have two division cured not cured. Then again with control cured and not cured.

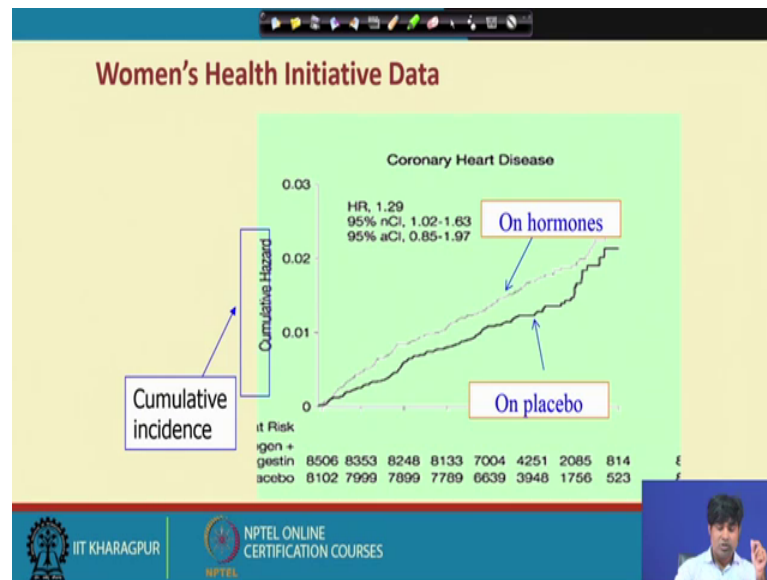
So, that how so; that means, exclusively depends you know the kind of you know count data and that to you to develop a structures. So, such that this kind of you know model can be applied. Again you know treatment controls died at the division will be dead and you know, alive and again dead and alive. So, that is how you know yes no type of you know situations right.

(Refer Slide Time: 07:37)



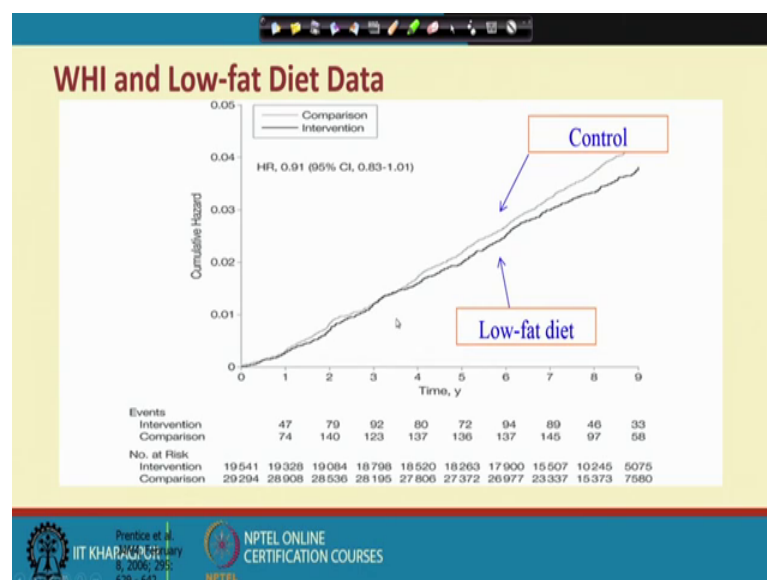
Against the 2 division should be exposed unexposed and again disease and disease free. In case of unexposed you know free and disease and then you know disease free.

(Refer Slide Time: 07:52)



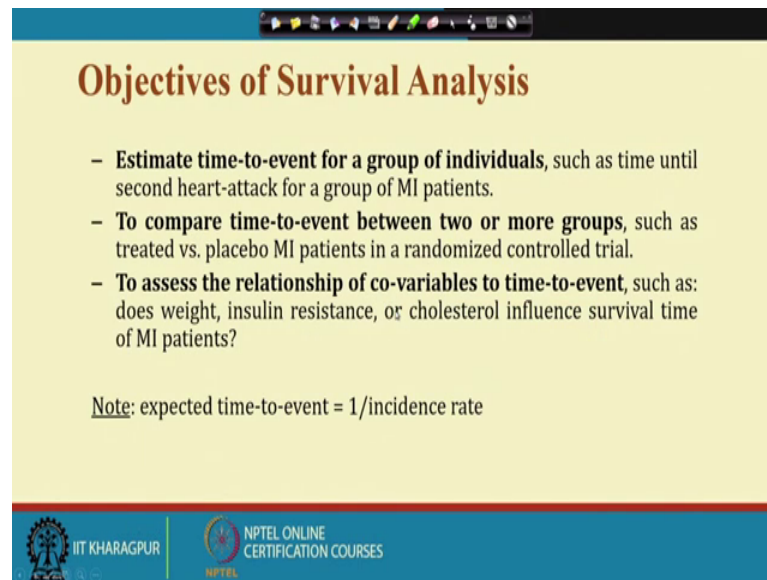
So, this is how the kind of you know case, and we are bringing such kind of you know environment here. So, this is a kind of you know plotting with respect women's health initiative data. So, will be find how this kind of you know distribution is happening. So, it is actually kind of you know 2 different kind of you know situation, and again the cumulative incidents.

(Refer Slide Time: 08:16)



Similarly, a low fat diet data so we have actually comparative kind of you know situation, that too with respect to low fat diet and then control.

(Refer Slide Time: 08:29)



Objectives of Survival Analysis

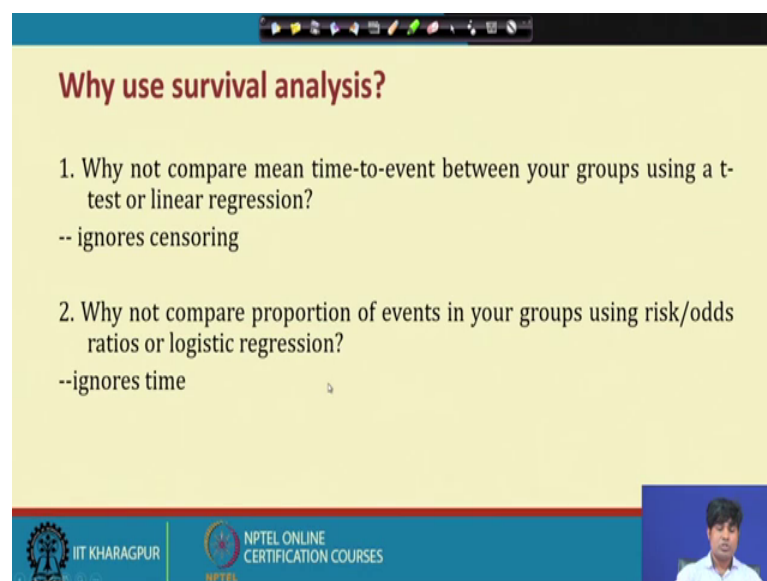
- **Estimate time-to-event for a group of individuals**, such as time until second heart-attack for a group of MI patients.
- **To compare time-to-event between two or more groups**, such as treated vs. placebo MI patients in a randomized controlled trial.
- **To assess the relationship of co-variables to time-to-event**, such as: does weight, insulin resistance, or cholesterol influence survival time of MI patients?

Note: expected time-to-event = $1/\text{incidence rate}$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So; that means, we need actually 2 different situation through which actually you can enter to this game and analyse the situation. The basic objective of survival analysis is to estimate time to event for a group of individuals, and such that the time you know second heart attack for a group of you know patients to compare time to event between 2 or more groups. And then finally, to assess the relationship of co variats of you know time to event right; that the you know check point and the expect time to event is 1 by incident rate.


(Refer Slide Time: 09:03)



Why use survival analysis?

1. Why not compare mean time-to-event between your groups using a t-test or linear regression?
-- ignores censoring
2. Why not compare proportion of events in your groups using risk/odds ratios or logistic regression?
--ignores time

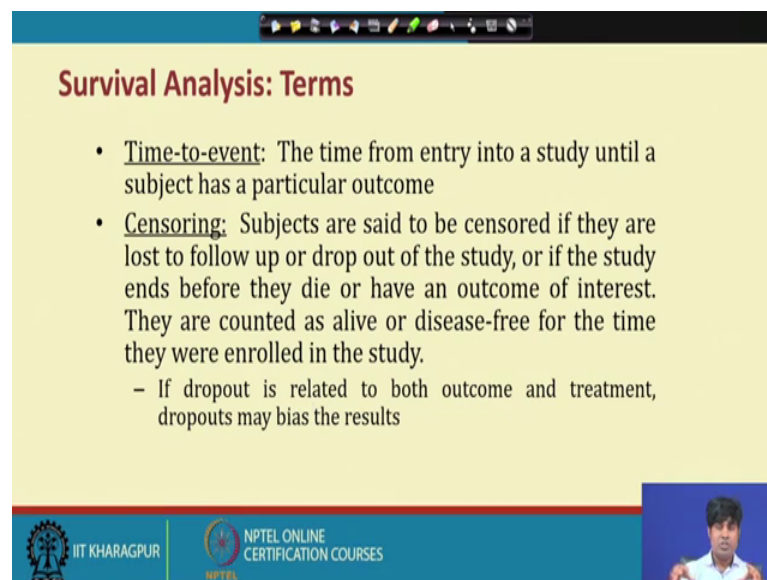
IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES



That is tells actually basically you know this is a very useful and we can a say that you know it is a kind of you know problem relating to health analytics. And in the health analytics this one of the biggest kind of you know component through which actually you frequently used, you know lots of count data moulding that to the game of you know survival analysis.

Why not this which, why not compare mean time event time to event between you know your groups using actually t test or you know linear regression, that is why how ignores you know censoring concept. And that is means actually before you start you know structure we have to actually, bring the such kind of you know check point and then think about the particular you know structure and then deploy this model.

(Refer Slide Time: 09:41)



The slide is titled "Survival Analysis: Terms" in red text. It contains two bullet points: "Time-to-event: The time from entry into a study until a subject has a particular outcome" and "Censoring: Subjects are said to be censored if they are lost to follow up or drop out of the study, or if the study ends before they die or have an outcome of interest. They are counted as alive or disease-free for the time they were enrolled in the study." A sub-bullet under Censoring states: "If dropout is related to both outcome and treatment, dropouts may bias the results". The slide footer includes the IIT Kharagpur logo, the NPTEL logo, and the text "NPTEL ONLINE CERTIFICATION COURSES". A small video inset of a presenter is visible in the bottom right corner.

Survival Analysis: Terms

- Time-to-event: The time from entry into a study until a subject has a particular outcome
- Censoring: Subjects are said to be censored if they are lost to follow up or drop out of the study, or if the study ends before they die or have an outcome of interest. They are counted as alive or disease-free for the time they were enrolled in the study.
 - If dropout is related to both outcome and treatment, dropouts may bias the results

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, 2 things are very important here, time to event and then censoring. Time to event means time from entry into a study until a subject has a particular outcome, and censoring is a subject are said to be censored if they are lost to follow means lost to follow up or you know drop out of the study, or if the study ends before they die or have an outcome of interest. They are counted as alive or disease free for the time they were enrolled in the study.

So that means, it is actually sometimes this kind of you know problem is a continuous problem and experimental type and a so we do continuously monitoring, then we use this data for this kind of you know modelling.

(Refer Slide Time: 10:40)

Data Structure: Survival Analysis

Two-variable outcome :

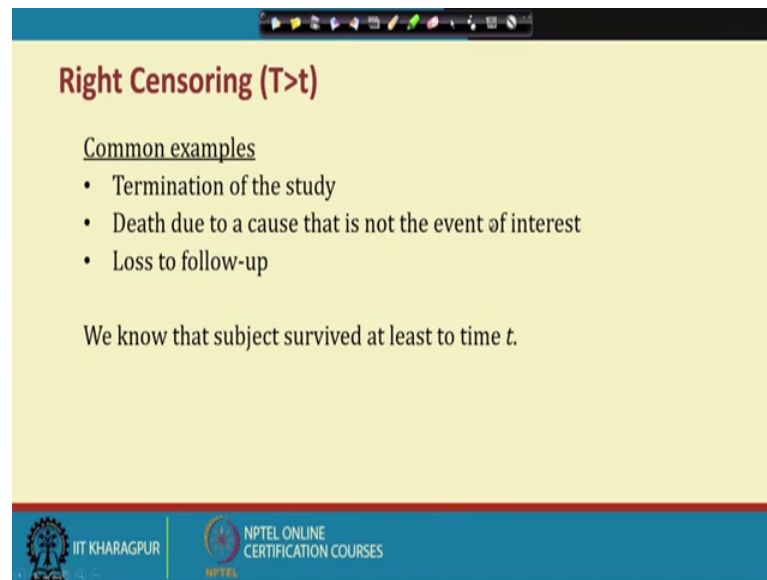
- Time variable t_i = time at last disease-free observation or time at event
- Censoring variable: $c_i = 1$ if had the event; $c_i = 0$ no event by time t_i

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, that is why you know this is very use for you know health management and you know healthcare, and to immediate issue is actually time variable and you know censoring variable. So, time variable which is called as you know t_i and then censoring variable which is called as you know c_i if had how the structure is here. So, time at least disease free observation. So, time at event and c_i equal to 1, if had the event or c_i equal to 0 not event.

That means actually it is a kind of you know dummy kind of you know structure and they that to the structure is a binary structure altogether. So, it is a what I say, what I am saying that you know it is a again specialized case under the count data modelling and a then we like to discuss the situation.

(Refer Slide Time: 11:28)



Right Censoring ($T > t$)

Common examples

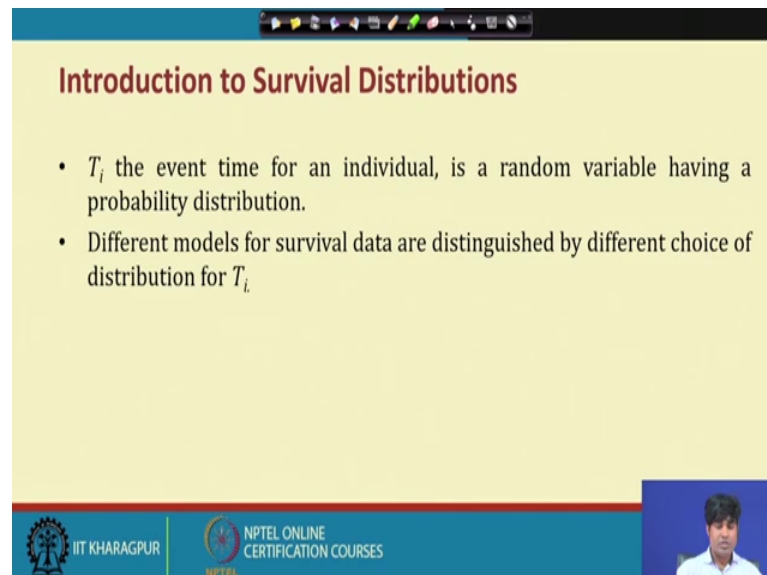
- Termination of the study
- Death due to a cause that is not the event of interest
- Loss to follow-up

We know that subject survived at least to time t .

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, right censoring that to common example is you know termination of the study, death due to a cause that is not event of interest, loss to follow up then again we know that the subject survived at least you know to time t is at one instance is it you know t greater than 2 small t .


(Refer Slide Time: 11:48)



Introduction to Survival Distributions

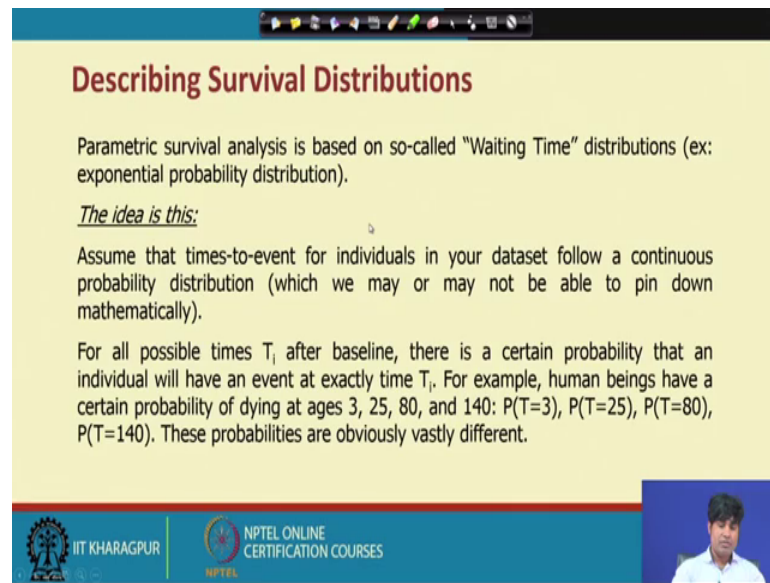
- T_i the event time for an individual, is a random variable having a probability distribution.
- Different models for survival data are distinguished by different choice of distribution for T_i .

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES



And t_i the event time for an individual is a random variable having priority distribution. And second different models for survival data distinguish by different choices of you know distributions for t_i .

(Refer Slide Time: 12:06)



Describing Survival Distributions

Parametric survival analysis is based on so-called "Waiting Time" distributions (ex: exponential probability distribution).

The idea is this:

Assume that times-to-event for individuals in your dataset follow a continuous probability distribution (which we may or may not be able to pin down mathematically).

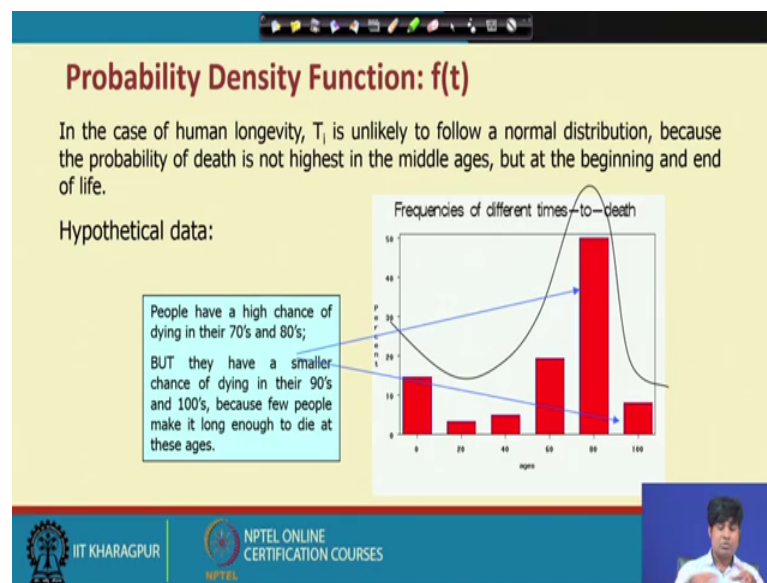
For all possible times T_i after baseline, there is a certain probability that an individual will have an event at exactly time T_i . For example, human beings have a certain probability of dying at ages 3, 25, 80, and 140: $P(T=3)$, $P(T=25)$, $P(T=80)$, $P(T=140)$. These probabilities are obviously vastly different.

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, that is how the game so, so far as a describe; I mean say survival distribution is a concerned. So, we have actually parametric survival analysis and is based on so called you know waiting, you know time that is waiting line distribution. And exclusively it is a kind of you know exponential type of you know probability distribution.

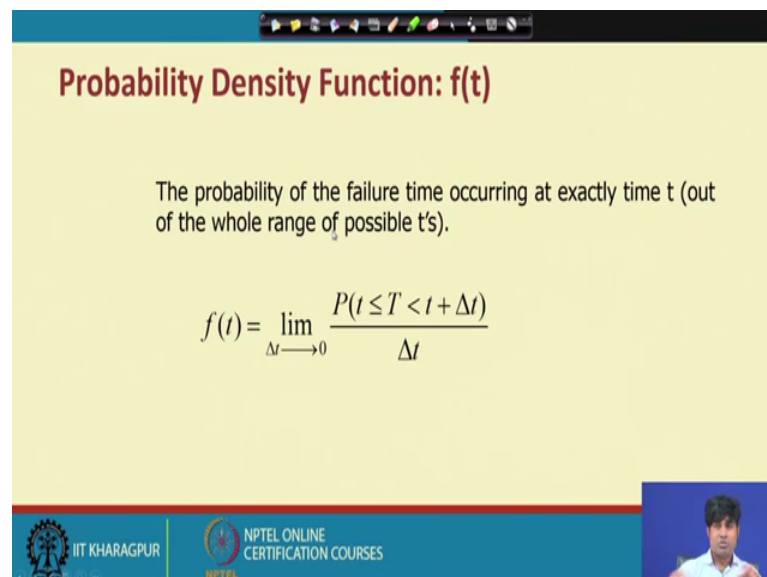
And the idea is that you know assume that the time to event for individual in our data set follow a continuous priority distribution. And if not then it will follow the different kind of you know distribution. Of course, the dealing is actually with respect to either you know Poisson regression modelling and negative binomial regression modelling, but subject to condition is that you know the kind of you know survival structure will be bring into the picture.

(Refer Slide Time: 12:57)



So, here is actually the kind of you know probability density functions in case of you know human longevity. So, we like to check actually t_i is unlikely to follow a normal distribution because the probability of the death is not actually highest in the middle age, but at the beginning and at the end in the end of life.

(Refer Slide Time: 13:21)



So, this is a actually hypothetical examples, but we can means typically it is a called question of you know life cycle hypothesis. So, one of the most important actually

component in the survival analysis is the density function and like you know Poisson regression modelling and negative binomial regression modelling.

Because it is actually special kind of you know problem relating to count data modelling; that means, it is a conditional kind of you know problems. Since we are putting actually 0 1 in the count data. So, that by default it is a conditional problem and it is more specialised and slightly you know different, and maybe little bit different because the data will be experimental type and you may not so, easily you can have this kind of you know structure.

So, the function will be you know with respect to limit Δt equal to 0. And probability of t less than 2 capital t and less than t Δt divide by Δt that is how the move we have to create a you know; that means, the data structure or you know data transformation or data restructuring will be follow with respect to these function so far as you know survival analysis is concerned.

(Refer Slide Time: 14:31)

Survival function: 1-F(t)

The goal of survival analysis is to estimate and compare survival experiences of different groups.

Survival experience is described by the cumulative survival function:

$$S(t) = 1 - P(T \leq t) = 1 - F(t)$$

F(t) is the CDF of f(t), and is "more interesting" than f(t).

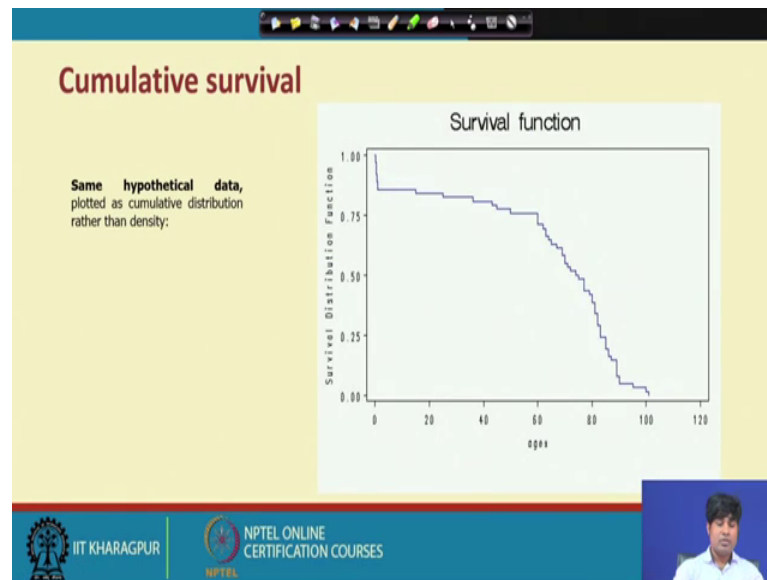
Example:

If $t=100$ years, $S(t=100)$ = probability of surviving beyond 100 years.

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

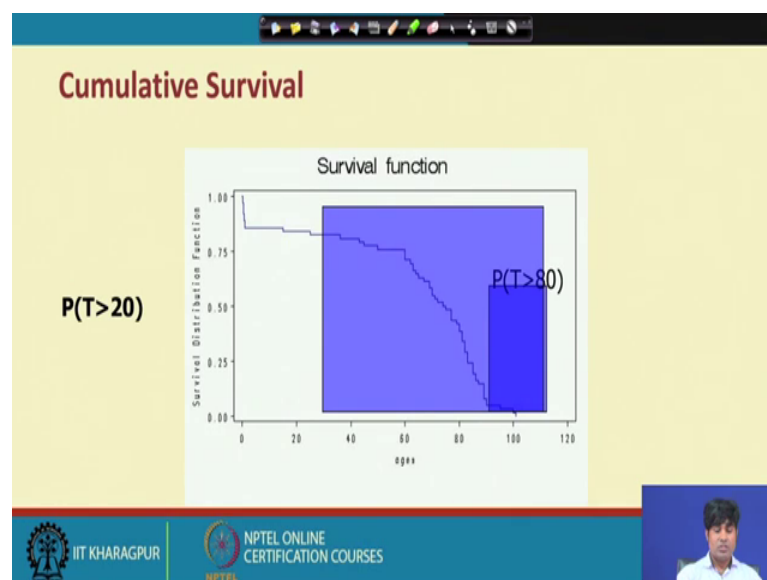
And then the density function cumulative density function will be $S(t)$ equal to 1 minus $P(T \leq t)$ that is what we called as you know $1 - F(t)$.

(Refer Slide Time: 14:51)

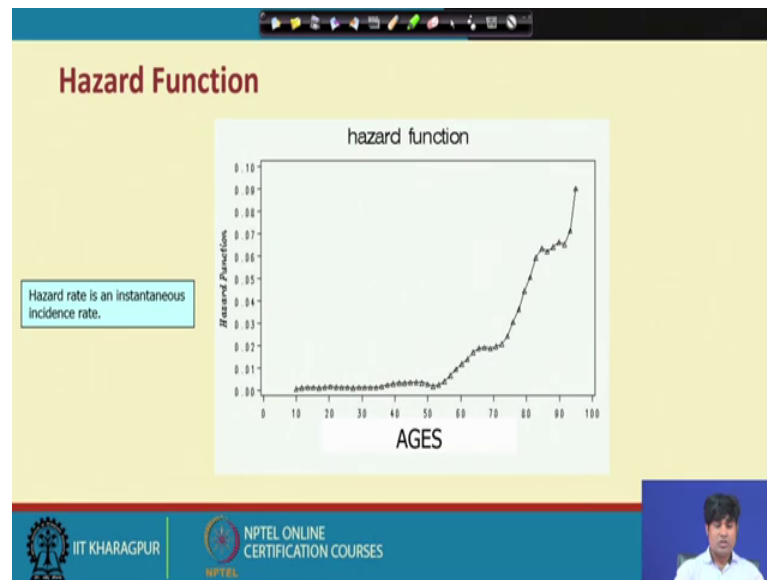


So, somehow it is actually connected to you know normal density functions. And then you see here the cumulative survival you know, you know situations where you know we have a survival functions. And the hypothetical data will be you know like you know you are plotting with here with respect to you know cumulative distribution rather than you know simple density.

(Refer Slide Time: 15:09)



(Refer Slide Time: 15:12)



This is another kind of you know examples and this what actually called as you know hazard functions. And hazard rate is actually a is a kind of you know incidents what is happening, and with respect to age how is the kind of you know occurrence that we have to first predict.

(Refer Slide Time: 15:29)

Hazard Function

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t / T \geq t)}{\Delta t}$$

In words; the probability that **if you survive to t**, you will succumb to the event in the next instant.

Hazard from density and survival: $h(t) = \frac{f(t)}{S(t)}$

Derivation (Bayes' rule):

$$h(t)dt = P(t \leq T < t + dt / T \geq t) = \frac{P(t \leq T < t + dt \& T \geq t)}{P(T \geq t)} = \frac{P(t \leq T < t + dt)}{P(T \geq t)} = \frac{f(t)dt}{S(t)}$$

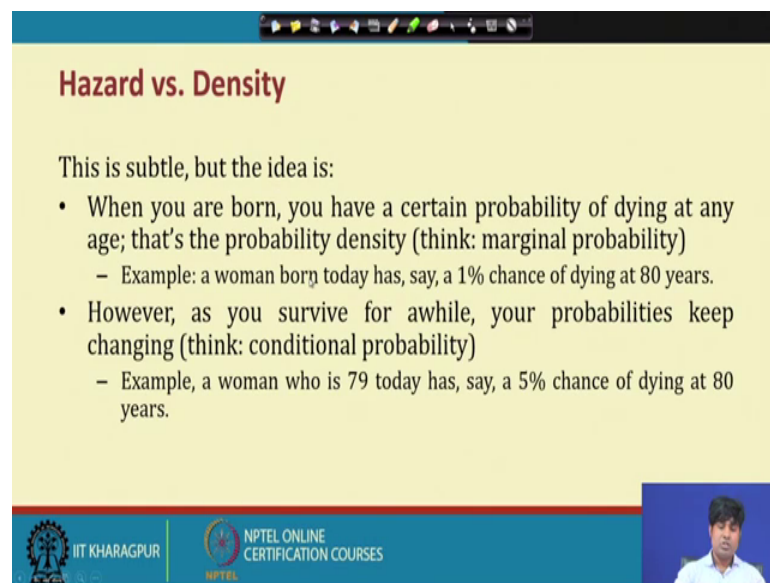
IIT KHARAGPUR NPTEL ONLINE CERTIFICATION COURSES

That means technical the as usual you have to go for you know plotting and you know check what is a exactly happening. And as in the corresponding to whatever we have discuss the hazard functions instead of S t we can call it is a h t. So, same structures and

a the probability that if you survive to T . You will be having a you know kind of you know event, and then hazard function density and survival rate equal to $h(t)$ equal to $f(t)$ divide by $S(t)$.

And then the a then we can apply actually some kind of you know Bayes theorem or Bayes rules to you know give such kind of you know scenario.

(Refer Slide Time: 16:08)



Hazard vs. Density

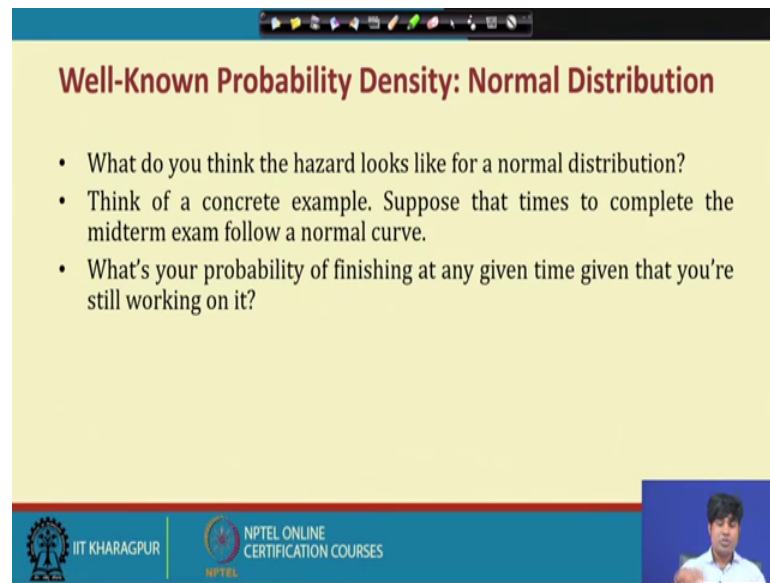
This is subtle, but the idea is:

- When you are born, you have a certain probability of dying at any age; that's the probability density (think: marginal probability)
 - Example: a woman born today has, say, a 1% chance of dying at 80 years.
- However, as you survive for awhile, your probabilities keep changing (think: conditional probability)
 - Example, a woman who is 79 today has, say, a 5% chance of dying at 80 years.

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

Because it is a conditional problem you know problem altogether. So, far is a conditional probability is concerned and a Bayes theorem is this the solid theorem which you can bring and connect and then we can discuss you know some way to analyse the situation. So, the density is actually into the pictures and with respect to probability density function that too this kind of you know conditional situations.

(Refer Slide Time: 16:38)



Well-Known Probability Density: Normal Distribution

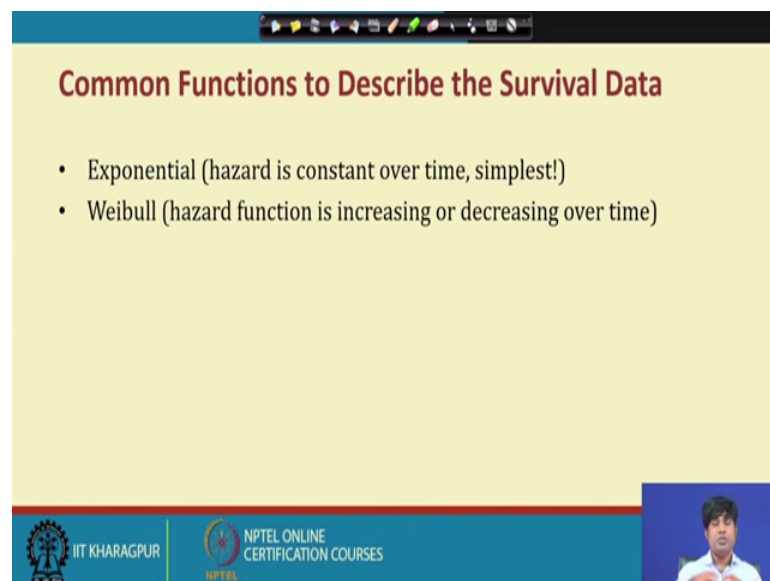
- What do you think the hazard looks like for a normal distribution?
- Think of a concrete example. Suppose that times to complete the midterm exam follow a normal curve.
- What's your probability of finishing at any given time given that you're still working on it?

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

The slide features a yellow background with a blue header and footer. A small video inset in the bottom right corner shows a man in a white shirt speaking. The footer includes the IIT Khargapur logo and the NPTEL Online Certification Courses logo.

We have to bring such environment so that we can apply the kind of you know models. Means the question is actually a if you are using actually when you know this kind of you know models. So now, how is the approximation to the you know normal distributions which is actually a considered to be a well-known distribution so far as you know, the entire econometric analysis or you know statistical analysis is concerned. What I have already mentioned we have to know when ends turns to infinity and most of the distributions will be follow the normal distributions.

(Refer Slide Time: 17:09)



Common Functions to Describe the Survival Data

- Exponential (hazard is constant over time, simplest!)
- Weibull (hazard function is increasing or decreasing over time)

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

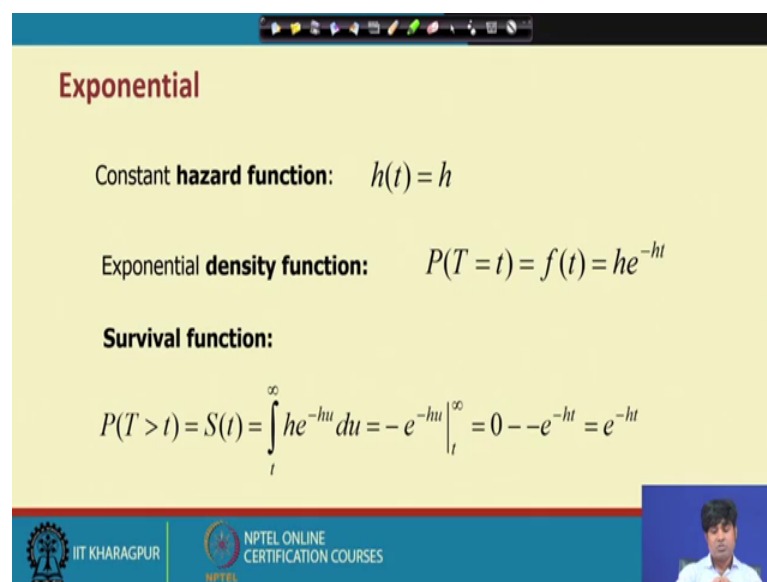
The slide features a yellow background with a blue header and footer. A small video inset in the bottom right corner shows a man in a white shirt speaking. The footer includes the IIT Khargapur logo and the NPTEL Online Certification Courses logo.

And then we have already discussed the case where you know the exponential distributions you know the kind of you know Poisson distribution, negative binomial distribution can close each others, when ends turns to infinity. The similar situation will be also here so; that means, every time 2 things very important so; that means, you if you increase the sample size indefinitely and a take care the outlier issue then most of the distributions will be follow normal patterns.

And behave normally and then you can predict the particular you know problems and the particular requirement very easily as per the particular you know engineering requirement.

So, the common 2 functions are you know exponential type and which is actually means where the hazard is constant of a time that is the simple. And the other one is actually hazard function is increasing or decreasing over times that is like, you know exactly the kind of you know poisons structure and the negative binomial structure one case. There is no question of dispersions in another case there is a kind of you know over dispersions.

(Refer Slide Time: 18:27)



Exponential

Constant **hazard function**: $h(t) = h$

Exponential **density function**: $P(T = t) = f(t) = he^{-ht}$

Survival function:

$$P(T > t) = S(t) = \int_t^{\infty} he^{-hu} du = -e^{-hu} \Big|_t^{\infty} = 0 - (-e^{-ht}) = e^{-ht}$$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So the hazard function $h(t)$ equal to h and exponential density function will be he^{-ht} to the power minus ht , that same thing we have to bring again then you have to you know structure the model.

(Refer Slide Time: 18:35)

Relating to Survival Analysis

Hazard from density and survival : $h(t) = \frac{f(t)}{S(t)}$

Survival from density : $S(t) = \int_t^{\infty} f(u) du$

Density from survival : $f(t) = -\frac{dS(t)}{dt}$

Density from hazard : $f(t) = h(t)e^{(-\int_0^t h(u) du)}$

Survival from hazard : $S(t) = e^{(-\int_0^t h(u) du)}$

Hazard from survival : $h(t) = -\frac{d}{dt} \ln S(t)$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

And so, far as a survival analysis concerned we have couple of items which you like to understand then connect as per the particular requirement. So, first thing actually to understand the hazard and hazard rate, then hazard from the density and survival and survival from the density and then density from survival $f(t)$.

And again density from a hazard so; that means, typically these are all somewhat we can called as you know different mathematics or you know mathematical modelling is are suppose to work out and know before you start you know doing the modelling, and that to do this kind of you know count data modelling. So, without knowing this once you cannot actually a do this analysis.

(Refer Slide Time: 19:21)

Getting density from hazard...

$h(t) = .01 * t$
 $h(5) = .05$
 $h(10) = .1$

Example: Hazard rate increases linearly with time.

Density from hazard: $f(t) = h(t)e^{(-\int_0^t h(u)du)}$

$f(t) = .01 * t e^{(-\int_0^t 0.01u du)} = .01(t) e^{-\int_0^t 0.01u du} = .01(t) e^{-.005t^2}$

$f(t=5) = .01(5) e^{-.005(25)} = .05 e^{-.125} = .044$

$f(t=10) = .1(10) e^{-.005(100)} = .1 e^{-.5} = .06$

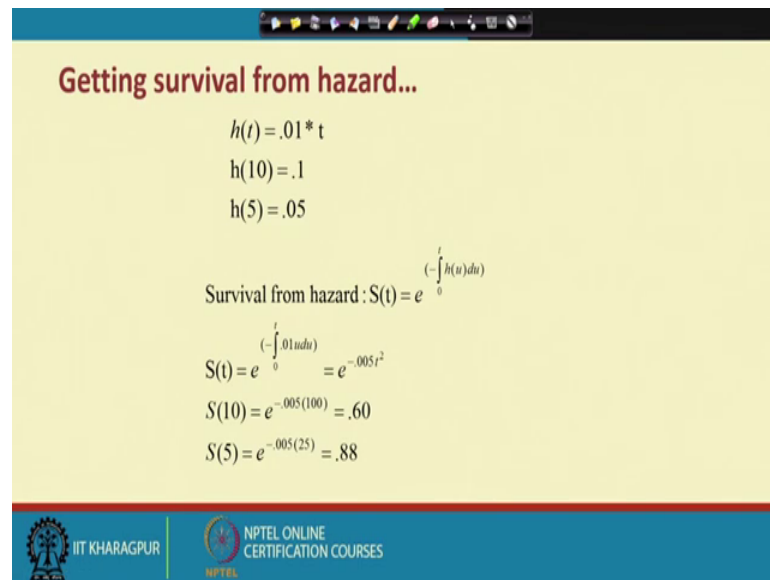
IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

That means actually what is more important here you know since, it is a specialised type of you know problems. So, of course, data will be like that and you need the restructuring of the data and you to somehow bring such as kind of you know situations that you know this particular model can be applied to analyse the situation.

Sometimes you know outliers may create a problems because when the outlet should be there in the system there is a high chance; you know mean you know over dispersion and where mean, and you know variance may not equal. But by the way with you know data transformation increasing sample size and restructuring of data. And somehow if you can remove the out layer then the problem can be solved.

So now corresponding to this you know structure here requirement. So, we have sample examples. So, we have to bring S t then S t then somehow some kind of energy structuring then finally, you to find out the hazard rate and the it is model it is a another kind of an example.

(Refer Slide Time: 20:19)



Getting survival from hazard...

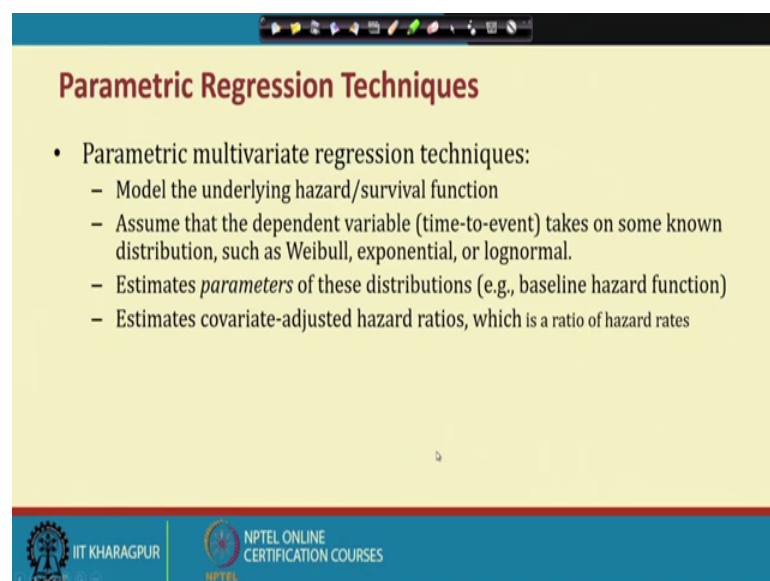
$$h(t) = .01 * t$$
$$h(10) = .1$$
$$h(5) = .05$$

Survival from hazard : $S(t) = e^{(-\int_0^t h(u) du)}$

$$S(t) = e^{(-\int_0^t .01 u du)} = e^{-.005 t^2}$$
$$S(10) = e^{-.005(100)} = .60$$
$$S(5) = e^{-.005(25)} = .88$$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

(Refer Slide Time: 20:23)



Parametric Regression Techniques

- Parametric multivariate regression techniques:
 - Model the underlying hazard/survival function
 - Assume that the dependent variable (time-to-event) takes on some known distribution, such as Weibull, exponential, or lognormal.
 - Estimates *parameters* of these distributions (e.g., baseline hazard function)
 - Estimates covariate-adjusted hazard ratios, which is a ratio of hazard rates

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

And corresponding to the hazard type of you know solutions. So, which actually somehow connected to you know non parametric versus you know nonparametric regression techniques.

In the parametric versions that too parametric multivariate regression techniques so, model the you know underlying hazard or survival functions. Assume that the dependent variable that too what you know time to event takes on some known distribution whether it is actually you know exponential type or log normal type.

So that means, actually it is the game between you know what we can call as you know a dependent variable so; that means, the entire count data modelling specifically actually you know what I can call as you know dummy type of you know situation where the specific target is actually to dependent variable. So; that means, all the restriction will go to the dependent variable like what we have already discussed the concept called as you know, domain dependent modelling.

So, where actually, we have no issue about the dependent variables, but we have issue about the dependent variable, like we have already discussed 3 different types of you know models in the case of you know domain dependent modelling that too starting you know linear poverty model logic model and prohibit model.

And more or less the same you know similar structure and you know same structure we are bringing here in the case of you know count data. There the only difference is that you know in the case of you know domain dependent modelling so; we have no strict guidelines about the you know data where we can use actually scaling data.

So, the relativity will be will be allowed there for instance the degree of difference like you know good bad you know best like that kind of you know situation, but here you know the specific restriction is that you know we need actually some kind of you know numeric kind of you know structuring where, fractional values will not be allowed. And the information should be integer type and that too sometimes it may be 0 1 type sometimes, you know you know different, but ultimately it should be integer type.

So, that is how it is you know specialized kind of you know situation where you have to apply then connect the model, means particular model as per the particular you know data structure and the problem requirement and the kind of you know engineering requirement.

So, the third point actually estimates parameter of these distributions and corresponding to baseline hazard function. And like what we have already discussed here, this (Refer Time: 23:12), these are all you know various parameters related to hazard functions and. These are for the entire you know these are all mathematical model related to survival analysis and then you know hazard type of you know modelling. And with the sample examples we have highlighted only 2 cases.

So, one of the important things is you know to estimate this parameters and then connect with you know the regression modelling where the you know the subsequent objective is to study the impact of independent variables to dependent variable. That means, actually here we have little bit you know complex and 2 types of you know requirement that too; specification about the dependent variable how you have to bring the dependent variable you know in a kind of you know structure where we can apply the survival analysis. Or you know hazard type bring the hazard type of you know situation, without any restriction to the independent variable.

But whatever the restriction and the you know you know structural requirement, you have to do and then restructure the data as per the requirement of a particular modelling. And then last, but not the least estimate covariate adjust hazard ratio which is actually ratio of you know hazards rates.

(Refer Slide Time: 24:32)

Parametric Regression Model

Components:

- A baseline hazard function (which may change over time).
- A linear function of a set of k fixed covariates that when exponentiated gives the relative risk.

Exponential model assumes fixed baseline hazard that we can estimate.

$$\log h_i(t) = \mu + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

Weibull model models the baseline hazard as a function of time. Two parameters (shape and scale) must be estimated to describe the underlying hazard function over time.

$$\log h_i(t) = \mu + \alpha \log t + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

The slide includes logos for IIT KHARAGPUR and NPTEL ONLINE CERTIFICATION COURSES at the bottom. A small video inset of a person is visible in the bottom right corner.

So, likewise you know this is what actually parametric a regression modelling. And we start with you know simply what I have mentioned you know $\log h_i(t)$ equal to μ plus you know $\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$ of course, i represent the different kind of you know sample structuring looks like you know baseline model. You know similarly if you know destructure little bit then you will find you know different kind of you know modelling structure all together.

So that means, technically these are all know different corresponding to the models which we have discussed earlier, this is similar kind of you know models, but only you know important thing is here the type of you know data, the structure of the data and the typical problem which you can actually apply. So that means, this kind of you know models mostly apply to transportation kind of you know problems, and you know medical science problems, where this kind of you know incidents are you know very frequent of course this problems are very rare, but still the frequency is very high in the areas.

And depending upon a particular you know problem and the kind of you know requirement whether it is a organisational requirement, or you know the kind of you know country requirement, or you know situation requirement, you can just pick up the problems and arrange the data or you know you can generate the data on a experimental bases depending upon the requirement of a particular you know model. Then you analyse the situation as per the requirement so; that means, technically.

(Refer Slide Time: 26:15)

The model

Components:

- A baseline hazard function
- A linear function of a set of k fixed covariates that when exponentiated gives the relative risk.

When exponentiated, risk factor coefficients from both models give hazard ratios (relative risk).

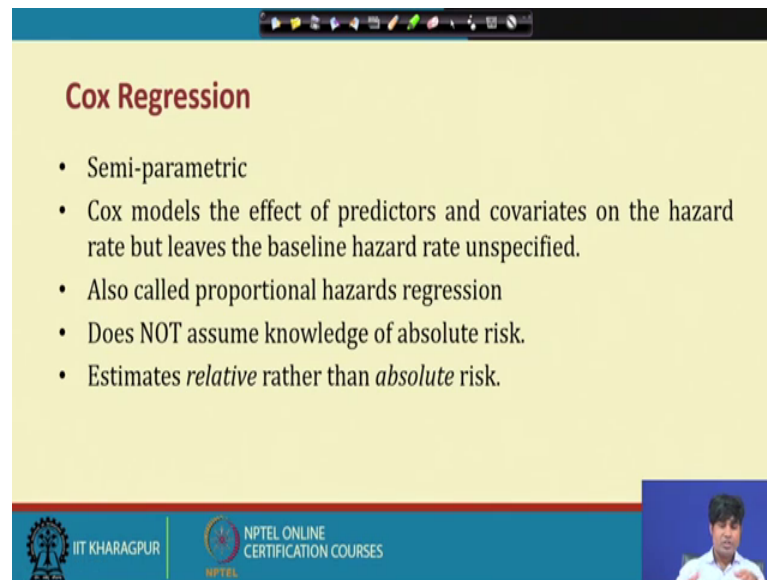
$$\log h_i(t) = \mu + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

$$\log h_i(t) = \mu + \alpha \log t + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, there are various ways you can actually analyse so this is how another way of you know bringing this kind of you know hazard functions.

(Refer Slide Time: 26:21)



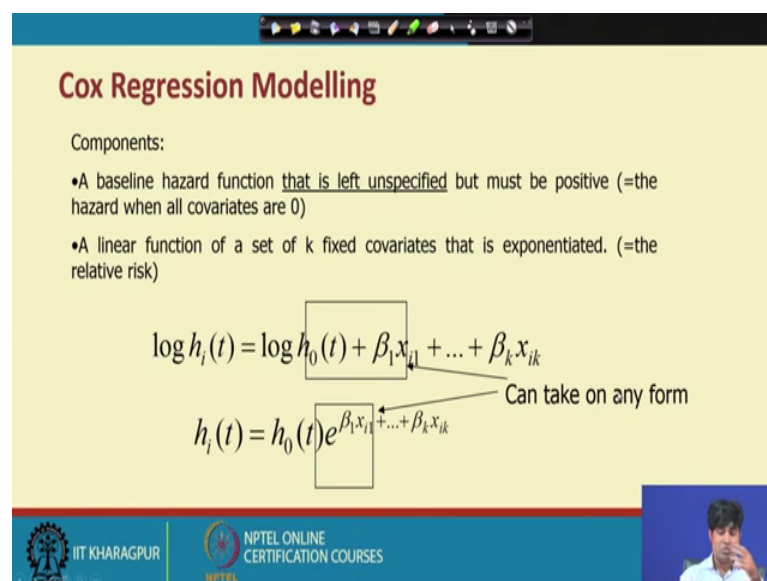
Cox Regression

- Semi-parametric
- Cox models the effect of predictors and covariates on the hazard rate but leaves the baseline hazard rate unspecified.
- Also called proportional hazards regression
- Does NOT assume knowledge of absolute risk.
- Estimates *relative* rather than *absolute* risk.

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

And there is a again you know corresponding to this clusters there is a concept as you know Cox regression, it is a slightly you know in this basket which is called as you know semi parametric. And Cox models the effect of predictors and covariates on the hazard rate, but leaves the baseline hazard rate unspecified. It is also called as a you know professional hazard regression. And here we does not means we do not actually assume knowledge of you know absolute risk and then finally, estimates relative rather than you know absolute risk.

(Refer Slide Time: 26:57)



Cox Regression Modelling

Components:

- A baseline hazard function that is left unspecified but must be positive (=the hazard when all covariates are 0)
- A linear function of a set of k fixed covariates that is exponentiated. (=the relative risk)

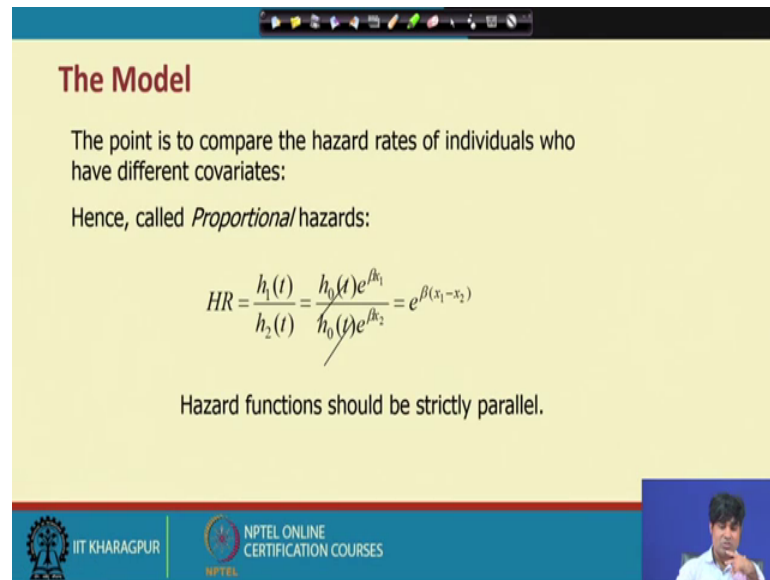
$$\log h_i(t) = \log h_0(t) + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$
$$h_i(t) = h_0(t) e^{\beta_1 x_{i1} + \dots + \beta_k x_{ik}}$$

Can take on any form

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

That is what kind of you know structure this is how the components of you know Cox regression moulding. And it is more or less you know similar kind of you know survival analysis.

(Refer Slide Time: 27:06)



The Model

The point is to compare the hazard rates of individuals who have different covariates:

Hence, called *Proportional* hazards:

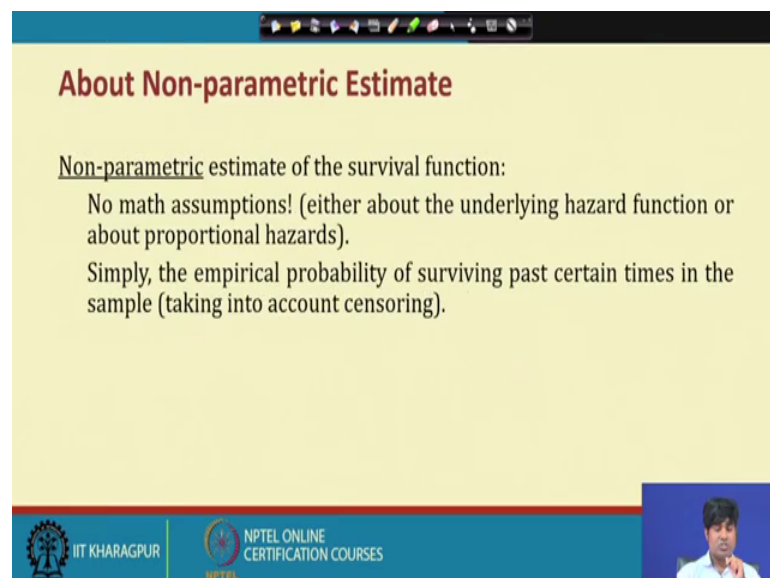
$$HR = \frac{h_1(t)}{h_2(t)} = \frac{h_0(t)e^{\beta_1}}{h_0(t)e^{\beta_2}} = e^{\beta(x_1 - x_2)}$$

Hazard functions should be strictly parallel.

The slide features a presentation toolbar at the top and logos for IIT Kharagpur and NPTEL Online Certification Courses at the bottom. A small video inset of the presenter is visible in the bottom right corner.

And not so much you know different and hazard function should be actually you know exclusively which is called as you know HR you know proportional hazards. And the point is to compare the hazard rates of you know individual wave actually different covariates.

(Refer Slide Time: 27:25)



About Non-parametric Estimate

Non-parametric estimate of the survival function:

No math assumptions! (either about the underlying hazard function or about proportional hazards).

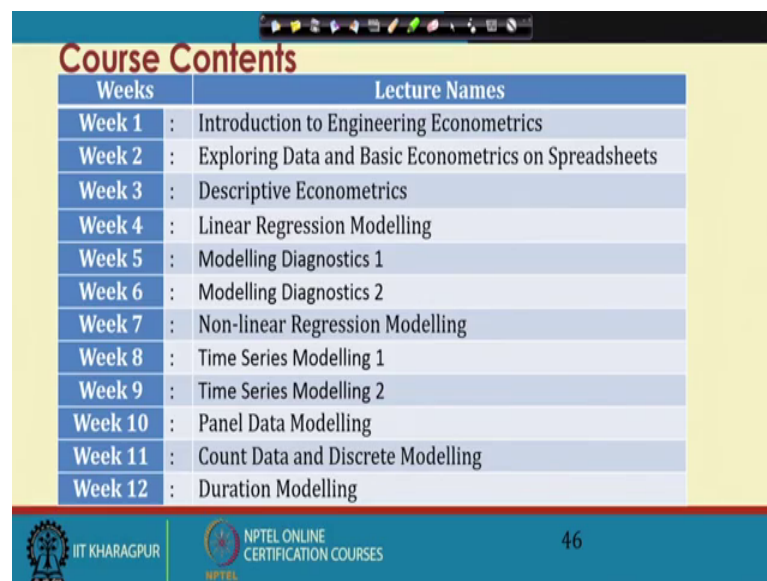
Simply, the empirical probability of surviving past certain times in the sample (taking into account censoring).

The slide features a presentation toolbar at the top and logos for IIT Kharagpur and NPTEL Online Certification Courses at the bottom. A small video inset of the presenter is visible in the bottom right corner.

So, this is another kind of you know instance and then last, but not the least actually we start with the parametric then semi parametric and then the kind of you know non parametric. And I am not going in details about all these you know classifications again. So, non-parametric estimation of the survival function is like that you know, no mathematical assumptions either about the underlying hazard functions or about proportional hazards.

And it is simply brings the situation where the empirical probability of surviving past certain times in the samples taking into account called as you know censoring samples. That is how the kind of you know situation, and ultimately the requirement is that you know we have discussed various types of you know models relating to the you know some of the engineering problems relating to do data. And then you know various kind of you know situations and these kind of you know requirement.

(Refer Slide Time: 28:30)



Weeks	Lecture Names
Week 1	: Introduction to Engineering Econometrics
Week 2	: Exploring Data and Basic Econometrics on Spreadsheets
Week 3	: Descriptive Econometrics
Week 4	: Linear Regression Modelling
Week 5	: Modelling Diagnostics 1
Week 6	: Modelling Diagnostics 2
Week 7	: Non-linear Regression Modelling
Week 8	: Time Series Modelling 1
Week 9	: Time Series Modelling 2
Week 10	: Panel Data Modelling
Week 11	: Count Data and Discrete Modelling
Week 12	: Duration Modelling

IIT KHARAGPUR NPTEL ONLINE CERTIFICATION COURSES 46

Ultimately I am bringing to the actually the entire you know structure about this you know engineering econometrics before I you know conclude this particular session and this you know the conclude this particular you know lecture series.

So, I would like to once again you know bring the extract of this particular you know course. So, starting with you know introduction to engineering econometrics then we have discuss the concept about the exploring data. So; that means, data is one of the biggest component which we are you know in the process of you know; solving

engineering problems in that to do the modelling. Which we have actually highlighted you know in kind of you know special category in the last lecture that too count data modelling.

But whether it is a count data modelling or any kind of you know modelling, like you not dummy modelling or time series modelling penal data modelling so data is actually big deal every times. So, until unless you know the data structure and the behaviour of the data or the kind of you know structure of data you may not actually do better modelling. And if you could not do better modelling you cannot actually solve the engineering problems as per the particular requirement.

Sometimes you know having the data you visualize it to know what type of you know function you can apply, then go for the estimation and the kind of you know process through which you can analyse the problem.

Then we have already discussed various descriptive econometrics as per the you know engineering econometrics requirement and you know problems requirement. And then we have discussed you know various types of you know models in one hand we have classified into 2 different structures, linear regression modelling and non-linear regression modelling.

Under linear regression modelling we have discussed various types of you know models starting with the simple, one the bivariate one that is the bivariate one and the multivariate one and various you know special cases, and the diagnostic tests, then the specification test goodness speed test; all these kind of you know reliabilities we have discussed.

And same thing we apply to the non-linear regression modelling whether it is a type of you know functional forms or the kind of you know as per the data requirement or something like that, but we have discussed various types of you know non-linear regression modelling. Then we have discussed various models relating to time series data and as a result we bring the situation you know call a situation called time series modelling.

Where we have discussed various models relating to you know some of the engineering problems; that means, where the time series data can help to solve sum of the

engineering problems. Starting with you know autoregressive models, moving average models, ARIMA models then we have discussed various types of you know volatility modelling. Then again we have discussed various types of you know VAR modelling starting with you know all the you know time series you know requirements unit root co-integrations and that causality.

So, these are the concepts which we have actually highlighted and these are also requirements of you know some of the engineering problems as per the need and the requirements. Then we have discussed you know something called as you know panel data modelling just you know connecting cross sectional modelling whether it is a linear structure or non-linear structure and then times you know time series modelling.

So, it is a concept of you know where a data is actually the kind of you know big deal, we adjust clubbing time series data with the cross sectional data and bringing the situation as per the particular you know requirement.

Under this umbrella we have discussed you know 3 types of you know models, fixed effect models, random effect model and generalised methods of moments. And again in the last unit we have discussed the concept called as you know count data modelling and including the duration modelling.

So, in the count data modelling the count data modelling it is somewhat not different, but it is under the umbrella of you know all these cross sectional modelling, time series modelling and panel data modelling. And again under the umbrella of you know both linear and non-linear modelling.

What is more important in this unit is the type of you know data which is actually a special kind of you know structures and again it is applied under a special kind of you know problems for example, like you know survival analysis. So, here the problems relating to this you know models count data modelling is the data structure where data it will be non-negativity in nature and integer type.

And again we have a special kind of you know problems in this count data is called as you know, survival analysis that too just like you know 0 1 integer programming.

So, all you know if you summarise we have covered you know almost all you know pool of you know various engineering economics tools to solve some of the engineering problems, as per the particular you know requirement. You know and whether it is organisational requirement corporate requirement or country requirement.

But we have highlighted you know various models you know depending upon the situation, and depending upon the data structure, depending upon the problem environment. And now my suggestion is that you know you go through all these you know models and then depending upon your you know area you know specialisation.

You pick up the problem because until unless you understand the problem you cannot actually pickup any model to solve that problem. So, understanding problems is one aspect and picking up the problem and understanding the problem in another aspect. But this kind of you know paper or this kind of you know modelling requires you know both sides of you know knowledge, you know theoretical knowledge and then the kind of you know econometrics knowledge.

So, once you are you know familiar with all these you know items then you can come with you know situation and where you can you know apply these models to solve the engineering problems as per the particular requirement. But there is no such restriction whether to apply only in you know kind of civil engineering or you know kind of navel science or some kind of you know make a other engineering problems. It can be applied to any engineering problems subject to the problem requirement and whether the model can be fitted there

If your problem is identified well data is available. So, just you check the problem and connect a particular model and then analyse as per the particular requirement. With this will stop here and thank you very much for you know choosing the subject, and I hope you have enjoyed all these lectures. And if you have any doubt or any kind of you know clarifications and how to do practice and how to connect with you know application you can keep touch with me, and my mobile and my emails are always with you.

So, hope to see you soon again by you know applying these models in a kind of you know real life situation and you know real life requirement. With you know this wish you all the best all of you.

Thank you very much and have a nice day.