Statistical Learning for Reliability Analysis
Prof. Monalisa Sarma
Subir Chowdhury School of Quality and Reliability
Indian Institute of Technology, Kharagpur

Lecture - 47
Logistic Regression (Part Ⅲ)

Hello guys, so in continuation of our lecture on logistic regression on the module relation analysis.
(Refer Slide Time: 00:37)

We have already seen the binary logistic regressions where we have considered one explanatory variable and two categories. So, today we will be seeing the second on the case that is many explanatory variable two categories. Many explanative variables meaning like the predictors in the earlier case which you have seen for categories are two, successfully a pass fail or whatever it is presence absence whatever you can say which we designated as 1 and 0.

The categories are two but then explanative variable we have just one like though it is not a very good example, the example what I have sited is like the number of hours you have put to study and with the you the student has passed or failed. So, it is like the two categories pass fail and that is the response variable and what is the predictor variable. Predictor variable is the number of hours that a student has put on the studies.

So, that was one explanatory variable only what was the predictor, that is the number of hours of study. So, now we will be seeing the second case where the explanative variables are more than one and categories are two only. So, we call it binary basically binary means it has to be two, so many explanatory variables meaning like for the same example maybe if age is also one of the predictor.

So, number of hours and an age of the student if we have two explanatory variable two predictor which will affect the response variable then we will call it a when it is more than one then we will call it a many explanatory variable two categories but that also falls under binary logistic regression. So, we will be discussing that and then finally we will discuss multinomial logistic regression.

Multinomial logistic regression means where the number of explanatory variable is definitely

more than one and as well as number of categories also more than two.
(Refer Slide Time: 02:30)

So, now first coming to this the second case there is many explanatory variable and two categories.
(Refer Slide Time: 02:34)

So, to begin with we may consider a logistic model with M explanatory variables. So, M explanatory variables means x 1 x 2 up to say x M and two categorical variables what are the categories variable y, y = 0 and y = 1. For simple binary logistic regression model, we assumed a linear relationship is not it for the when we have considered a simple binary logistic regression model that is one explanatory variable and two response variable.

We assumed a linear relationship between the predictor variable and the log odds. Remember we wanted how we find found out the log odds. So, first we found out the log odds, what is odds success number of success by number of failures. So, that is odd and then we have done the log or we have taken the log of odds log of orders, what is that is nothing but that is the logic from logic only.

We got the logistic regression name I think you can remember that, if you do not remember kindly go through my lecture number one and two for this logistic regression, this is the lecture three logistic relation. So, and this linear relationship, now whatever linear relationship we have seen for one predictor variable for the predictor variable and the log odds the same linear relationship we can extend to the case of M explanatory variable.

If you can remember, let me write it just here remember we have used the sigmoid function. First of all, why we have used the sigmoid function, because sigmoid function what is the sigmoid function our sigmoid function is something of this sort 1 / 1 + e   where this is our logic function, we have already seen in the last class. So, now what we got logic that is t we call it t, t we got it something this form β 0 + β 1 x.

This is what we call as we got the logit function for the simple binary logistic regression model, this was our logit function remember. So, now similarly this we can extend it for M explanative variable as well, so here our logit that is t, t is nothing but log of p / 1 – p, p is 1 p is the; what is this p / 1 - p it is the odd of getting 1. What is the odd of getting 1? Odd of getting 1 in so odd of getting 1 / odd not getting 1.

There is a difference between odds and the probability, I have already mentioned. So, this is when we take log of this then we got is whatever we call it we call it a logit that is t. So, this is nothing but this if we simply extend this to the multi variable we call it we get is in this form. So, here you see the parameter β 0 to β M if there are M parameters M explanative variable x 1 to x M, M explanatory variable then we have parameters β 0 to β M that is we have parameters as well so, M + 1.

Similar here we have seen we have only one explanatory variable the number of parameters is 2 β 0 + β 1 where t is the log odds and β are parameters of the model.
(Refer Slide Time: 05:53)

So, now and here you see I have used the log base b, b means I have just want to generalize it while considering the x many explanatory variable why I had just want to generalize it in say earlier cases what we have used. We have just used log e base e is not it. Here, basically it is that if you can use either e or 2 or 10 it totally depends on the range of your data, what range your data falls into.

So, expecting that for many explanatory variable it may have a very higher range if it is a very high range then, we will go for base 2 accordingly what if it is in the range of 2 to power 10 then we will go to the for the base of 2. So, likewise I have kept a base as b so as to reflect that it can use any base 2 e or 10 depending totally depending on the range of the data values.
(Refer Slide Time: 06:50)

So, now what as I told you our explanative variables are total M explanatory variable x 1 + x 1 x 2 x 2 x 3 up to x M total M explanative variable and for M explanatory variable how many parameters we need we need M + 1 parameters that is β 0 to β M. So, just to generalize it we to make it more compact notation here I have introduced one more predictor. We call it a predictor or explanatory variable whatever it is and the output we call it a response variable.

So, we have introduced one more that is x 0 and x 0 is nothing but x 0 = 1. So, it will not make any change because you see simple equation when we have t 0 = β 0 x 0 + β 1 x 1 where x 0 = 1 it is just equal to β 0 + β 1 x 1. So, it does not make any difference that is why just to make it compact we have introduced one more explanatory variable x 0 which is the which is we have assumed it to be 1 and so the logit will take a very compact form.

So, what is the now logit is M summation of all β M x M say this function, what is t is this one. So, to make it compact we have just introduced x 0 nothing no great rocket (()) (08:13) science here just then our t = β x.

(Refer Slide Time: 08:20)

Now, solving for the probability p that y = 1 given a particular value of the explanatory variables. So, what is the probability that we it falls in the category 1. So, how we get that we probability we get just by using the sigmoid function is not it. So, let us go in a backward reverse direction. So, this is what is my this is my sigmoid function is not it. So, from this basically I can write it in this pattern we have already seen that.

So, basically p x is this it is nothing but the sigmoid function. So, where S b is the sigmoid function with base b see earlier, I have used e now to just to generalize it I have used it e b do not get confused with that, the above formula shows that once the b m's are fixed. We can easily compute the log odds that y = 1 for a given observation. So, once we know the value of the parameter β 0 b 1 b M.

Once you know the value of the parameter, we can easily compute what is p x is not it what is p x probability that it falls in there. What is the probability that it falls in the category 1 is not it, it is taking that there are two category p x is the probability that it falls in a category 1. Then 1 - p x is false in the category 0. So, the main use of the velocity model is given an observation x estimate the probability p x that y = 1.

So, why we can make the model, so we make the model so that given an observation x we can estimate what is the probability that it will fall in the category 1 like the example what I have given for in a fair when you try to hit the target. Given that I try to hit the target, I am trying to should from a particular fit say 4.5 feet what is the probability that I will hit the target. So, if the regression model is there, I will be able to do that.

So, now the question what we will have to find out the values of the parameters what the parameters different parameters what the parameters will take what values once we find that values will be able to given any explanatory variable given any value of the any predictor will be able to give the what will be the response, we will be able to give the probability of the different response variable.

So essentially, we need to estimate the values of the parameters. So, for estimating the values of the parameters whatever we have seen for single explanatory variable two categories same applies here also. What we have done there we have used MLE method.
(Refer Slide Time: 11:01)

So, see here also we will be using MLE method, same nothing no difference here the same MLE method just that here explanative variable is that it is instead of one here our explanative variable has M classes total sorry, M explanatory variables not one explanative variable but more than one explanatory variable that is the only difference that it has nothing other. So, we will have to find out the M will be using MLE estimation that is maximum likelihood estimation.

So, you I am sure you can remember what is MLE? So, once you know basically from that first we find out what is the likelihood and, in this likelihood, basically, when we try to multiply by the probabilities it becomes unstable so we take the log of that. So, it is called log likelihood log likelihood nothing but the maximum likelihood estimation and I am not explaining going to more details here.

Please if you cannot remember I request you please go to my second lecture of logistic regression which I have covered in details what is MLEM. So, after that once we find the log likelihood expression done to optimize the; which is the optimized value of β what we will do? We will have to differentiate it or differentiate it for different parameters and usually how do we find at optimum value.

So, build from difference first order differentiation and then we try to equalize it to 0 and once we equalize it to 0 then we get the then we will get a set of equations then from this equation we find out the values of the for particular value for which we can say that function is optimum. Here, but as I told you this the expression here that we get it is not an algebraic equation so, here it is if so, we just cannot directly equalize it to 0.
(Refer Slide Time: 12:52)

Though, I have mentioned it is putting I as well so putting this equal to 0 but because that is the standard format of standard wave when we try to find out the optimum value of a particular thing when first order differentiation then equalize it to 0 and find out the values of the parameter that is the standard form that is a do I have written it to 0. But actually, you cannot just equalize it to 0 because, it is not a finite series it is not a finite series of some algebraic operations.

So, we will have to use approximation method but you will use Newton Raphson method. Newton Raphson method is the most suitable method you can use any there are other numerical techniques also but Newton Raphson method is suitable. So, with that you will be able to find out the different β, 0 β 1 β up to β M different β values. Once you found out the β values then you will be able to find out the t that is the logit once you find out the logit then you will be able to easily find out the probability.
(Refer Slide Time: 13:45)

So, here an example is given. So, consider an example with M = 2 explanatory variable and suppose base is 10 and so we got by MLE method we found β 0 is also t β 1 is also β which have been determined. Now, here I want to tell you this MLE method is that the thing is that you do not have to do it by hand. Of course, you should know how to it is done but you as I have already mentioned before you do not have to do it by hand.

There are many software’s available you will be finding many Pascal code which does all this thing you just have to feed the data and but then you should know the theory behind. So, you know when β 0 β 1 β 2 values are given then what is my logic this is my logit, this is my logit expression, 2 explanatory variables. So, β 0 + 1 x 1 + 2 x 2 this is my logit once I know my logit what is my p p of x p of x is nothing but what is p of x.

We have seen what is p of x this is nothing but e of x + 1 + e of x is not it x sorry not x t logit, p of x is $e^{t / 1 + e t}$ we have already seen that. So, here what is we have considered what is t t is nothing but the β x this is our t. So, and we have considered base = 10, so this is the formula this is how we got it. Now, we have this now for any value given any given any value of x will be able to find out what is the probability that y = 1.

We have this this is our model now this is the model from this model if we have our value x because there are 2 parameters, we will we need to know the values of both the 2 parameters x 1 and x 2. And both the parameters are given then we will be able to; suppose in case of shooting the target the parameters may be distance from the target. This is one parameter and a way direction of the wind may be another one parameter.

So, accordingly both the parameters values when both the parameters values are given we will just put it here and we will get the probability. So, now see here just for this value what is given β 0 = - 3 what is β 0, β 0 is nothing but the intercept and what is β 1 and β 2 β 1 and β 2 basically it is it implies how the different parameters are affecting the response variable.
(Refer Slide Time: 16:17)

So, when we have β 1 = 1 means that increasing x 1 / 1 increases log odds by 1, how it is affecting the response variable. B 2 means increasing x 2 by 1 increases log odds by 2 when the parameter values is more does explanatory variable has more effect on the response variable.
(Refer Slide Time: 16:41)

So, now that is all from the binary logistic regressions, now we will consider multinomial logistic regression.

(Refer Slide Time: 16:51)

Multinomial logistic regressions I can tell it is a case tree where many explanatory variables and many categories. So, many categories like we can just take a simple example like based on some concentration of certain thing element in our body we can say whether there is a whether it will lead to a benign tumour cancerous tumour or it is not a tumour at all basically I can say normal.

So, there may be 3 categories, so here my explanatory variable may be there may be 2 maybe 1 explanatory variable it has 1 or many basically, 1 explanatory variable that is basically the concentration or maybe 2 explanatory variable concentration of a particular element and maybe the a's, a's and concentration will have an effect. So, these are the 2 explanatory variable.

And the different categories maybe it is a benign tumour or it is a benign or cancerous or normal means it is maybe it is not a tumour at all. So, there are three different categories.
(Refer Slide Time: 17:58)

So, now and so when the above cases of two categories the categories were indexed at 0 and 1 and we had 2 probability distribution. We had just 2 probabilities 1 probability distribution for 0 another probability distribution for 1. By now I am sure all of you know what is probably distribution if at this point if you are forgotten what is probabilistic, please go back to my lecture where I have explained very clearly what is probably distribution that is in the beginning.

Now, the probability that the outcome was in category 1 is given by p x and outcome is in category 0 is 1 - p x and the sum of the both the probabilities equal to unity. Of course, there are 2 classes and probability of 1 class is p x another class is 1 - p x and both has to equal to 1 and that was we have seen in case of binary logistic regression. In general, if we have M + 1 explanatory variable including x 0 we have seen why we have taken x 0.

So, as to make the form compact if we have M + 1 explanative variable and N + 1 categories the example recently what I have given that can benign cancel as a normal. So, there are N + 2 + 1 categories, so there are total 3 categories. We will needs N + 1 separate probability distribution. So, we will be needing if there are 3 categories we will be needing 3 probability distribution.

If there are 4 categories, we will be needing 4 probability distribution. One for each category indexed by M, which describe the; probability that the category categorical outcome y for explanatory variable x will be in category y n. So, if there are 4 categorical variables that is four response variables and four classes response classes, then we will be having four probability distributions and what this probability distribution described.

It describes the probability that the categorical outcome y for the explanatory variable. If it is one explanatory variable 2 or 3 whatever it is explanative variable at one point will be in category y n.

(Refer Slide Time: 20:13)

So, and now if there are 3 categories the sum of all this probability has to be 1. If there are 4 categories sum of all these 4 categories has to be 1. So, it is also required that the sum of these probabilities over all categories be equal to unity that is obvious. So, now using the mathematical convenient base e, these probabilities are same format what we have used for binary logistic equation. What we have written for p and f?

So, what we have written for p x p x we have written is e^t e^t 1 + e^t, what is t? t is nothing but the logit what is the logit is t = β 0 t is nothing but β 0 + β 1 x. So, I have to make it compact let me write it β 0 x 0 β 1 x ,1 so t is nothing just simple we can compare form we can write as β x, t is β x, so this is p x. Similarly, I can write it for 1 - p x.

So, now here there is when there was only two category, we could write it this way. Now, here more than 2 categories, so what will be my p x p x again I will have to write it in this is my p x p and x just for suppose one category p and x. so, this is nothing but it will be in this format where for n = 1 to n so here I have taken 1 + 1 n + 1 / n + 1 you will see the difference here, now y instead I could have easily considered n just to make the things more understandable I have used n + 1.

So, now this is for p of n x here I have just considered p x and otherwise rest one is 1 - p x. Now I cannot do that there are an N + 1 category so it will be p 0 p 1 p 2 p 3 up to P n. So, now I am first let me consider from 1 to N so p of n x p of n x is nothing but $e^{\beta n x}$ + 1 + summation of all these values for all the different categories. So, this is the value for p n x and I have separately I have considered p 0 what is p 0, p 0 is nothing but 1 minus of this.

Just to give reflect this only I have separately used the p 0 I can use p 1 also p 2 also. Just why I have used means if I want to find out the probability of one class it is nothing but 1 - summation of all other probabilities it gives me p 0. So, that is why I have just written it in this way p 0 format and there is one more necessary I will come to that. Now, this is how I have defined p n x p 0 x.

(Refer Slide Time: 23:12)

So, now it can be seen that as required the sum of p n x over all categories is unity. Of course, of all categories it is the, note that the selection of p 0 x is defined in terms of other policies artificial. Here, we have used p 0 x you can use any one p 1 x in terms of the other probabilities. Why I have used p 0, when I use p 0 it becomes easier for me to explain it summation of because 1 minus it will be 1 - we did not back at p 1 + p 2 + p 3 up to p n.

So, I can write it in very compact from this way. So, I have taken p 0, now one more thing like when I use the odds there is one more requirement why I have separately used one classes where I am again, I am repeating you can consider any 1 p 0 p 1 p 0 any 1. So, when we have seen when we say when we calculate the odds of a particular class when we calculate the odds of a particular class how did you calculate the odds of the particular class. So, we will just see how we have calculated the log odds.

(Refer Slide Time: 24:22)

So, when we calculated the log odds so, log odd of say p 1 x p 1 x was what? p 1 x was number of occurrence of 1th class by number of occurrence of 0th class remember. So, that was my that is how I calculated my p 1 x. So, basically now if I should not write it this way, this way it is becoming very complicated. So, like my p x I will tell it. Here, this is p 1 x / p 0 x, so this was my log odds.

Log odds that it falls in the category 1 given a variable given the explanative variable x it falls in this category 1 this is my log odds. That is the number of occurrences that falls in the category in the one divided by the number of occurrences that falls in a category 0, this gives me log odds. Now, when I have more than two categories then what I will do is that I will use one of the classes one of the response classes as a Pivot class.

Or we can call it regress class basically other class other classes, I will regress it based on that class. The other class I will regress it based on that class. Now, suppose I have 3 categories what are the sub the example what I have given that based on the concentration to there are three different types of it can be 3 different type of cancers and 3 different type of classes maybe it is a benign tumour maybe it is a cancerous or maybe it is normal it is not a tumour at all so there are three different classes.

So, when there are 3 different classes then let me find out how will I find out the log odds of whether it is a log odds of benign, log odds of cancers, log odds of normal. So, 3 different log odds I have to find out. So, in that case how do I find out is like in the when there are 2 categories I found out the log odds of just one category that is p 1 and other, I did not find that is 1 minus of that and that only gives me that.

So, now when there is three categories I will find out the log what is up 2 categories, when there are n categories I will found out the logouts of n - 1 category. How do I find? For finding out the log odds from any category, I will be using a one reference class that I call it a p both class. So, based on that class I will be basically regressing my log odds whichever class I trying to find out I will regress that based on that reference class or Pivot class whatever it is.

Now, here let me con use that I am using p 0 as my Pivot class. So, if I use p 0 as my pivot class then what happens you see here.
(Refer Slide Time: 27:05)

The special value of n is termed as a pivot index and the log odds t n are expressed in terms of pivot probability and are again expressed as a linear combination of explanative variable. So, t n and for that it belongs to the log odds that it belongs to the nth category this I am writing it in this way. Log of p n f y / p 0 x where my p 0 x I am considering it as a pivot index, that I am considering that as a reference class my 0th class I am considering as a reference class.

Let me consider in here may be my benign class and maybe I am considering as a pivot class. Now, when similarly suppose let me see if I have such a way if I have three 3.
(Refer Slide Time: 27:47)

So, how will I do and let me again consider my 0th class as the 0 what to say 3rd class by normal class if I consider as a pivot class then my I will be having three differences. That is log of 1 is p 0 of x p 0 of x by if I am considering, sorry I am not considering p 0 this is p 1 p 2 p 3. So, p 1 x / p 3 x, so this is my logit of t 1 log it of t 1 is this. Now, my logit of t 2 = log of p 2 x / p 3 x, so I got log it of t 1 and t 2.

Here basically you see I have compared p 1 with p 3 and p 2 with p 3, see if compared p 1 with p 3 p 2 with p 3 which comparison is left here from here p 1 and p 2 combination is still not there. What I want to show here is that whichever whatever reference class you use either here I have used reference class as a p 3. Whichever reference class you use it is immaterial, it will be getting the same value.

Because, see here I have used the reference plus but I have done also using just 2 logits I will be able to get all the logics of all other classes because see here in t 1 I have compared p 1 with p 3 and in t 2 I have compared p 2 with p 3. Now, what is remaining is my remaining is p 1 and p 2, now if I do t 1 – t 2 what would I get you see t 1 – t 2 what will I get des log of this - log of this what will I get log of p 1 x / p 2 x.

So, see using this I got this comparison as well. So, whichever may be the logit reference class whichever maybe the pivot class you consider that is immaterial in a way you will be basically trying to compare or while finding out the odds basically you will what to say. It will compare the odds for all the classes you will find out the odd odds for all the classes. So, if there are three the categories.

So, I will find log it is 2 logics if there are four categories I will find three logics so basically if there are k categories so we will total find k - 1 log it is. So, when I find k - 1 logit that means I will be getting k - 1 expression. So, now from the logit is from this logged in term is in terms of the parameter you can see β 1 β n x here is not it logit is always in terms of the different parameters.

Now, to finalize my model I will have to find out the best values of these betas. So, how do I find out the best values of betas. So, I will have to again the same technique I will have to basically find out what to say I will have to do the first order derivations of my problem likelihood function.
(Refer Slide Time: 31:14)

So, what is my likelihood function here the likelihood function is the joint probability distribution of here the likelihood function is the joint probability distribution of all the categories all the different stores in the since there are and probabilities and different categories. So, for n different categories I will be having and different probabilities, so it is the joint probability distribution of all categories. So, I will be finding my; that is my total function.

So, once my total function is there, I will have to find out the log likelihood function. Once you find out do the derivative of this function likelihood, log likelihood function mind it. We first find out the likelihood function, likelihood function is nothing but the probability distribution of all the different classes. So, then we will find out the log likelihood and then we will have to optimize this log likelihood function.

So, to optimize this log likelihood function how we will do will first do derivation first order derivation and first order derivation based on the different parameters.
(Refer Slide Time: 32:05)

And once we found out the first order derivation on different parameters then we will find out the and we will equalize it to 0 and find out the different probabilities, different values for the different parameter. Similarly, the same technique is applied here also, well so what is basically, we use a function we use a special term this is called an indicator function which is equals to unity if y k is = n and 0 otherwise, so now this is my log likelihood function.

So, this function I will use the derivative of that and I will try to find out the parameters of that. So, again for this it is better to use by software, software code is available pascal code is available.
(Refer Slide Time: 32:47)

So, basically you will have to derive this term using all the parameters. So, what does B nm see I am differentiating with respect to β nm what is β nm β nm is the mth coefficient of the m β and vector. See we saw here β is a vector because there are different categories so β is a vector and what is m even this is also a vector, we have m variable. So, β mm is the mth coefficient of the β n vector and x m k is the mth explanative variable of the kth measurement.

So, differentiate with respect to β and m for different values of n different values of m and then we can find out by using Newton Raphson method you can find out the values of the different betas. Once you find out the values of β then finding out the probability is nothing just put the values in the logic function and once you know the logic function, we can very easily find out the properties of the different categories.
(Refer Slide Time: 33:50)

So, now coming to the different examples of logistic regression and logistic regression it has applications in various domain. So, in medical domain it has tremendous applications, so once this application we can see in the tree school that is trauma and injury severity score which is widely used to predict the mortality in injured patients was originally developed by Boyd et al using this is used logistic regression, this is a very well-known process.

Then many other medical scales used to access severity of a patient have been developed using logistic regressions. Again, logistic regression may be used to predict the risks of developing a given disease like diabetes, coronary, heart disease based on different parameters what is the risks of the developing diabetes then it may fall into the different category. So, based on the observed characteristics of the patient like age, gender, body mass index results of various blood tests etcetera.
(Refer Slide Time: 34:46)

Then, in social sciences also there are lots of examples of logistic regressions. Then in engineering and engineering basically, this is I have specially kept an example from the reliability engineering prospect. This technique can be used in especially for predicting the probability of failure of a given process system or product given the values of the different variables.

Variables means different conditions, maybe for a system maybe for a system if we see the values of the different like the system may have capacitor inductor resistor many other components. So, it is different values at a certain point of time it is different values, how this will affect the you know what to say failure of the system, failure we can keep it in different categories prone to failure very much prone to failure that way we can keep in different categories.

And these different values of this will defined in which probability of system being in which state at that instant of time, similarly, it has application in marketing domain.
(Refer Slide Time: 35:47)

Then, it has also application in economics, in economies it can be used to predict the likelihood of a person ending them in a labour force and business application would be to predict the likelihood of a homeowner faulting on mortgage. So, and here I forgot to mention one thing do I have that. So, like in bank transaction whether it is a customer is a fraudulent customer, a genuine customer they say logic revision is used to find that out. So, in natural language processing also logistic regression has it is applications.
(Refer Slide Time: 36:20)

Then, coming to the conclusion we learned a multinomial logistic regression. So, now we have from in three classes we have learned logistic regression. In the next lecture we will cover a tutorial on the logistic regressions.
(Refer Slide Time: 36:34)

So, this is the reference and thank you guys.