Exploring Survey Data on Health Care
Prof. Pratap C. Mohanty
Department of Humanities and Social Sciences
Indian Institute of Technology, Roorkee

Lecture - 28
Relative Risk Ratio Estimation

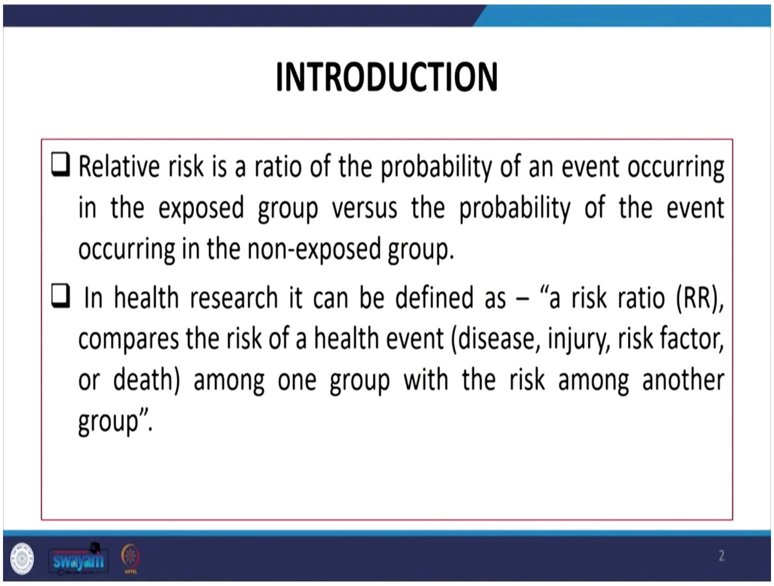
Welcome participants for this NPTEL MOOC module on exploring healthcare survey data. We are on the 6th week of our sessions. In this particular week, we have been emphasizing on influential statistics in health care. What kind of influential statistics are required in research? One of the topics were picked up for your direct use is relative risk ratio estimation.

There are number of confusions derived whether to go for an odd ratio, whether to go for a logistic regression, whether to go for a multinomial logistic regression or whether to go for any other limited dependent variable models. We have to be very careful while applying the appropriate estimation technique. We know that relative risk ratio is important that has to be assumed from the beginning.

But what are the beginning points that is going to lead us to understand the relative risk ratio to apply or to estimate. We will also give you practical handouts with example data sets that will help you to find out the estimation and its interpretation very carefully. So, let us go further. On the very introduction, I wanted to mention that the relative risk ratio is a kind of probability which is comparing the event that is occurring against the event that is not occurring.

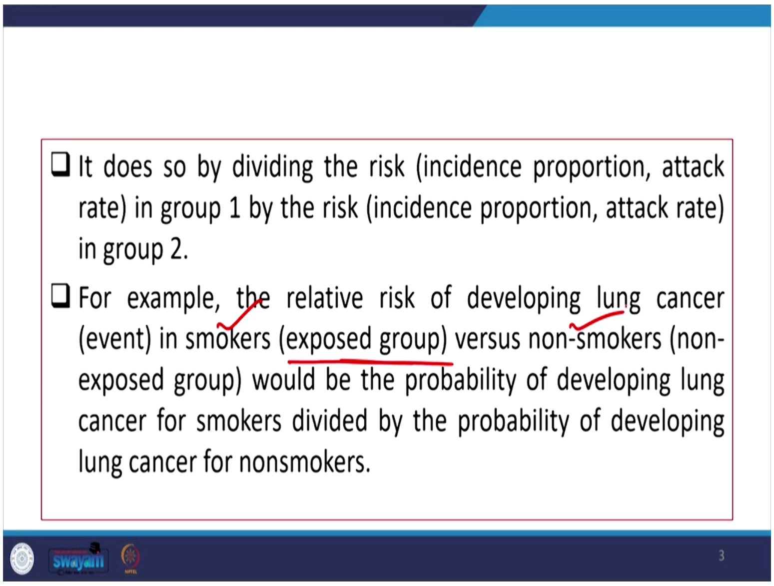
Or, more particularly, it is expressed at event occurring in the exposed group versus the probability of event occurring in the non-exposed group. Then we would certainly explain you what exposed group and non-exposed groups is?

(Refer Slide Time: 02:33)



In health research, it can be identified as a risk ratio that compares the risk of a health event maybe disease, maybe injury, may be risk factor or death among one group with the risk among another group. So, basically one group of risk is compared with another group of risk.

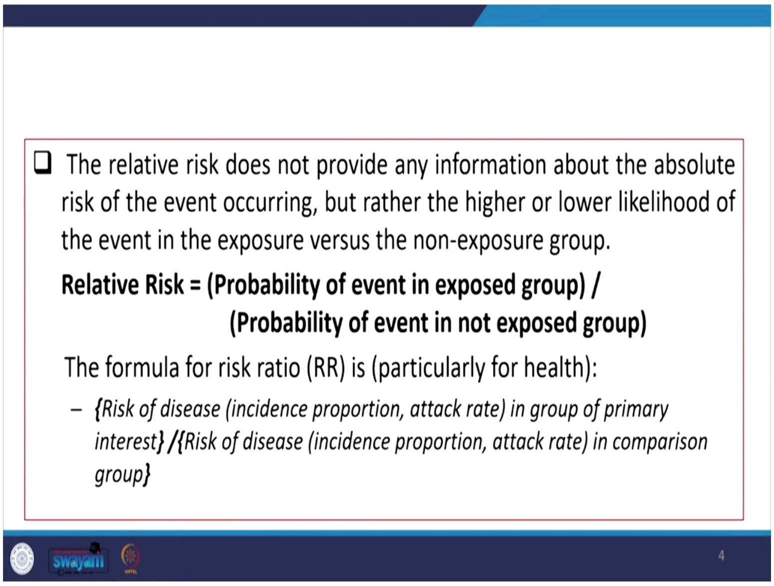
(Refer Slide Time: 02:52)



It does so by dividing the risk that is incidence proportion or attack rate in group 1 as compared to the same risk events in group 2. For example, the relative risk of developing lung cancer as such type of event in smokers, let it be the exposed group in our understanding versus non-smokers as the non-exposed. It would be the probability of developing lung cancer for smokers divided by the probability of developing lung cancer for non-smokers.

So, the comparison between exposed group and non-exposed group that we have just mentioned is basically smokers and non-smokers.

(Refer Slide Time: 03:48)

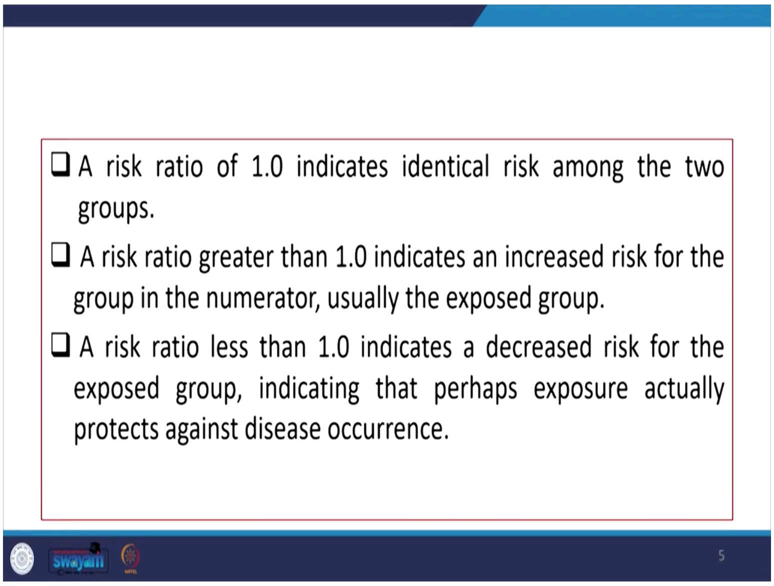


So, the probability between these two are actually compared here. The relative risk ratio does not provide any information about the absolute risk of the event occurring, but rather the higher or lower likelihood of the event in the exposure versus the non-exposure group. So, what is that? This is basically a ratio of the probability of an event in the exposed group versus the probability of an event in the non-exposed group.

The formula for this risk ratio is written as RR sometimes as RRR relative risk ratio or simply risk ratio and it is particularly applied in the context of health care i.e., it is referred as risk of disease or an incidence proportion or attack rate in group of primary interest divided by the risk of disease in comparison group, in another comparison group of those with same incidences. So, when we are taking a risk of ratio of it i.e., basically giving us the risk ratio.

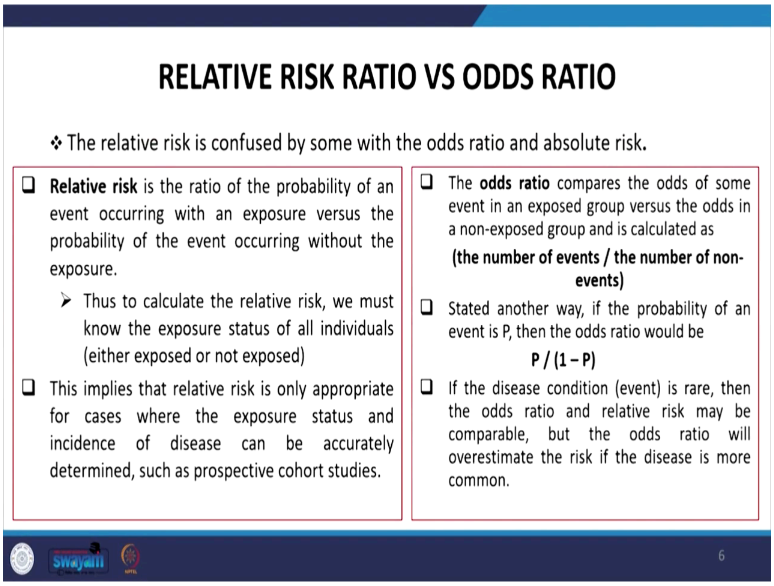
So, risk ratio is defined accordingly. So, now with regard to risk ratio-I have already clarified that, it is basically the ratio of two events. So, two events happening and not happening usually considered, but their probabilities are taken as a ratio.

(Refer Slide Time: 05:17)



A risk ratio of 1.0 indicates identical risk; identical risk among the two groups that is being compared. Risk ratio is with greater than 1 that indicates the risk for the group in the numerator, is usually increased or higher as compared to the denominator. Whereas, the less than 1 is basically explaining the reverse that is the risk of occurrence of the particular event which is getting lesser is indicating reduction in the extent of risk.

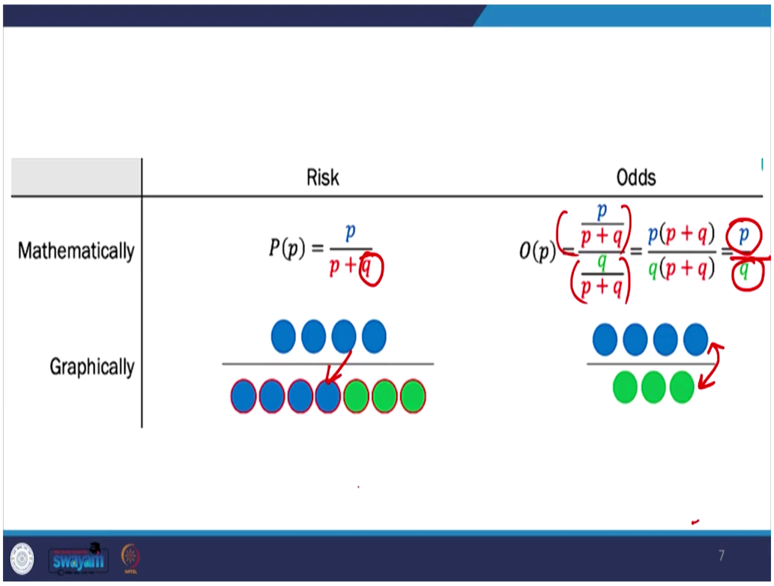
(Refer Slide Time: 05:54)



So, what do you mean by this relative risk ratio versus odds ratio? and, how these two are compared? I will just first show you this and then I will come back to first. I have already just said that relative risk ratio is the ratio of the probability of an event occurring with an exposure versus the event occurring without the exposure. That is straight away comparing with and without exposure and their probability ratios. Whereas, odd ratio compares the odds of some event in an exposed group versus the odds in a non-exposed group.

So, it is not like non-exposure, we are saying the event occurring without the exposure. So, first one is saying exposure in the risk ratio as compared to the probability of event occurring without the exposures. But here, we are actually comparing little in a different way, P versus 1 minus P is the ratio that I am just clarifying with this diagram first, then we will come back to it, with its mathematical interpretation and the diagram.

(Refer Slide Time: 07:17)



Risk is basically a probability of the total event. So, the risk is basically, q is non-occurrence for sure, then p is in fact, the occurrence probability of occurrence. So, when we are taking the ratio of the total event and it is probability of happening i.e., p out of the total possibilities, that is basically called risk, it is a kind of probability. Whereas in case of odds, it is basically p over q i.e., happening over not happening.

So, this is derived as p divided by total number of events or q divided by the total number of events. So, p plus q cancels out, then what is left is odds of happening versus not happening. So, in the diagram, you can see these two groups are not comparable and that is why it is mentioned as odds. what is called the odds in favor and odds in against. But here, it's actually representing the total events what is the probability of it.

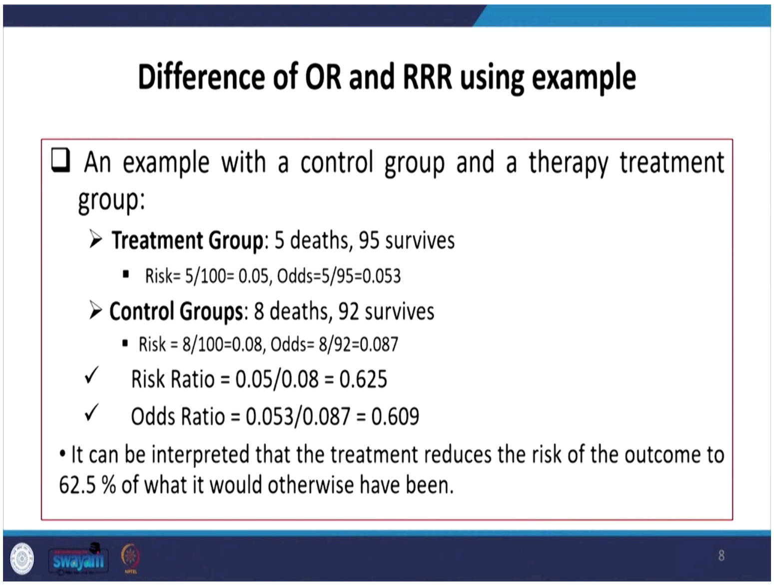
So, the risk is explaining largely about the probabilities, but odds is comparing the two event and though two events are mutually exclusive. Thus, to calculate the relative risk, we must know the exposure status of all individuals. What we are trying to say here is, all individuals and their exposure starters are actually compared whether they are exposed or not exposed. But in the case of odds ratio, all the events are not required in the denominator.

So, in the relative risk ratio, relative risk is only appropriate for cases where the exposure status and the incidence of disease can be accurately determined such as prospective cohort studies etc. When accurately determining, any event like cohort studies, then relative risk issue is more applied, but in case of odds ratio, we take probability of probability and then, non-probabilities.

If the disease condition is rare, then the odds ratio and relative risk ratio may be comparable. When it is rare, this may be comparable, but the odd ratio will overestimate the risk if the disease is more common. When the disease is very frequent or common in that case, odd ratio is going to overestimate. So, usually in cohort studies for specific diseases, risk ratio is more applied that is going to give you the right estimate.

Because this is going to find out the probabilities. But if you find out the odds ratio, since your probabilities are happening. It is very high in case of more common diseases so, this is going to derive probability over non-probability, since its very high. So, that is going to overestimate your result that is going to be much higher than that of the risk ratio. That is why these two has to be taken down very correctly while comparing. This is what we already explained.

(Refer Slide Time: 10:45)



Now, we are further differentiating with certain examples like an example with a control group and a therapy treatment, a therapy treatment group is like 5 deaths as compared to 95 survives. So, risk ratio is equal to 5 divided by the total number of events that is 100. So, 5 divided by total number event it is actually 0.05. But the odd ratio is in fact, 5 divided by 95 because i.e., the case p is for 5 events happen in favor against it is 95.

So, 5 divided by 95 is against 5 divided by 100, 5 divided by 100 is your risk ratio, it is a kind of probabilities, but odds ratio are not probabilities, these are coefficient. And coefficient comparing they have probability of an event as against probability of not happening an event. So, what do you mean by control groups cases like 8 deaths in 92 survives if it is the case, in that case, risk is of course, 0.08 and the odds are 8 divided by 92 that is 0.087.

So, a risk ratio is equal to as compared to these two comparison, treatment and control groups, if you have treatment groups and control groups where in the treatment group some sort of treatment is given, and control is the idealistic condition i.e., there is no treatment is given. We just want to compare these two results with certain interjection in the treatment group. So, we can compare the result of, we can find risk ratio in these two groups as well.

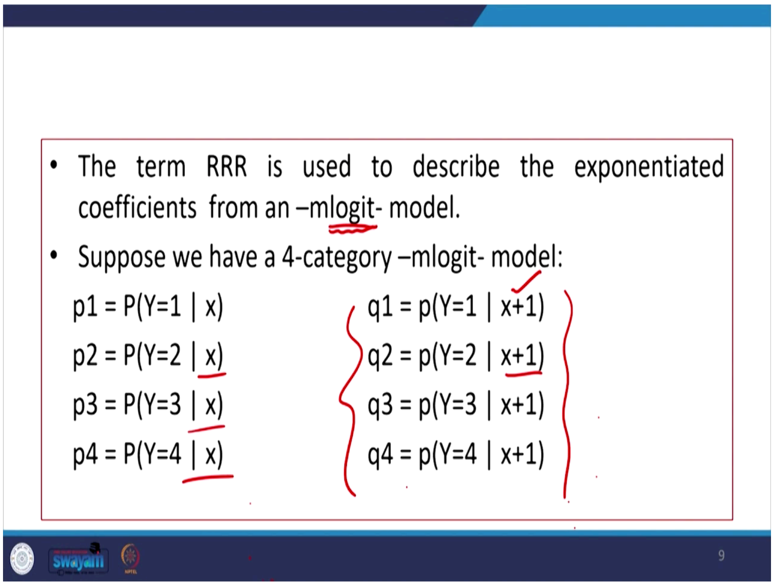
So, in one case, it is 0.05 that is 5 percent, in another case, 8 percent. So, 5 percent divided 8 percent can be also compared. So, that is also going to give you a risk ratio with treatment group and risk ratio. So, the treatment group is going to give you 62.5 percent result or 0.625

whereas, the odds ratio of these two events like control and treatment group is going to give you is lesser i.e., 0.609.

Now, what is the take away in this case that- though in both the cases, odd ratios where were very high, odd ratio are higher than that of the than that of the risk ratio. But when we take the comparison between the treatment and control groups, now we can see the risk ratio actually captured better. So, the odd ratio is not capturing the right aspect.

So, it can be interpreted that the treatment reduces the risk of the outcome to 62.5 percent of what it would otherwise have been. So, the treatment group actually acts as change significantly to the risk now of a particular event or of treatment therapy if you have given some therapy that has actually impacted 62.5 percent in the event.

(Refer Slide Time: 14:10)

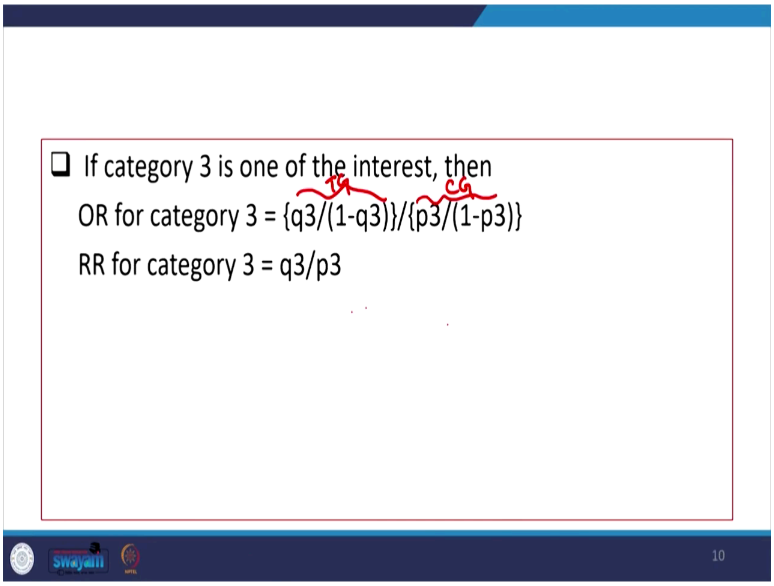


The term relative risk ratio is used to describe the exponentiated coefficients from an mlogit model. So, mlogit likewise we have discussed in this week about logistic regression, probability regression, they are binary regression that we have already discussed. We are discussing at this moment on the multinomial logistic regression, we have also discussed a bit about multiple logit regression.

Now, we are saying the coefficients that is going to be useful in the relative risk ratio is through the mlogit model only that is the multinomial logit only. Suppose we have 4 categories in the mlogit model, 4 categories whose probability are mentioned, the outcome is also mentioned here. The p1, p2, and p3, p4 are mentioned here. p1 of an event is basically Y. It is the probability of Y equal to 1 given the x control variables, then p2 is equal to P(Y= 2) given the control variables, similarly here in all the cases, its probabilities are determined.

In case of the treatment group, the change in the model can also be estimated within, with change in this one, x1+ 1 event, then x1+ 1 in each of the cases we can also find out their probabilities with the change in the control variables.

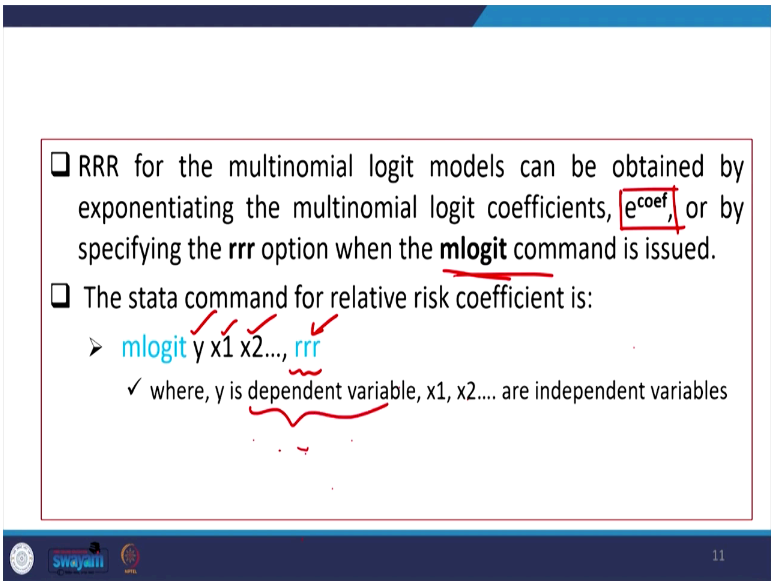
(Refer Slide Time: 15:42)



If the category 3 is one of our interest, then the odds ratio in that case, in that category 3 is equal to q3 which we have just said q3 divided by (1- q3). What is odd ratio? Odd ratio is basically 1- q3 whole divided by p3 -1 by p3. So, basically this is your treatment group TG, and this is your control group, note it.

So, odds ratio between the probabilities of the treatment group with the t+1 event and a t or a t0 event. So, it is an odd ratio for category 3 can be defined, but in case of relative risk ratio or risk ratio for category 3, it is simply q3 by p3. Now, q3 is our probability, simply we can take the ratio of these two. It is going to give us the result correctly. So, these two would be compared, alright.

(Refer Slide Time: 16:51)

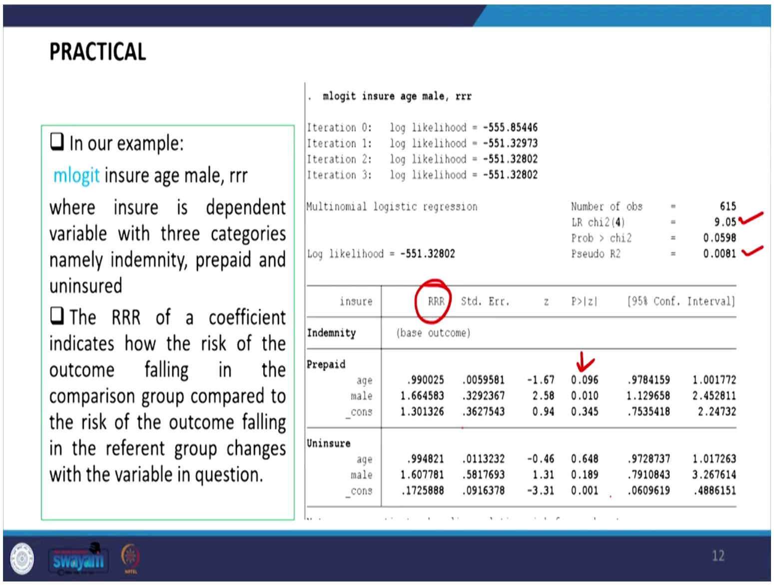


Relative risk ratio for the multinomial logit models can be obtained by and exponentiating the multinomial logit coefficients i.e., e to the power coefficient if you do it that is basically the RRR that can be determined or by specifying the RRR option when the mlogit command is used. This mlogit command is given and at the same time, we are also adding with relative risk ratio command rrrr, then that will be giving us the relative risk ratio coefficients or relative risk ratio values.

Otherwise, if you have multinominal logit coefficient and if simply you take the exponentiated values of each coefficient, then you will certainly able to get the relative risk ratio values. So, mlogit command which we have already given here is our dependent variable mentioned, then q1, q2 and its all-control variables are mentioned, given with the options we have cited, we have already added to the command is rrr.

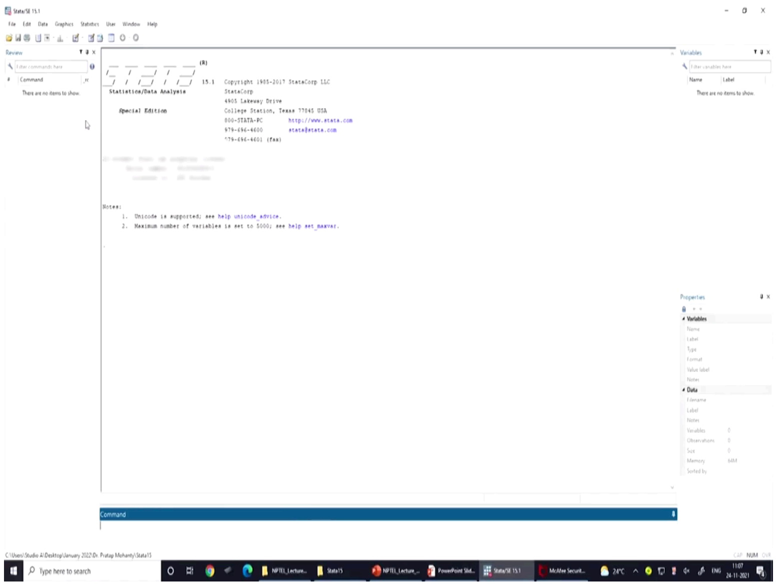
So, y is the dependent variable and we have assumed that in the dependent variable, we have more than binary responses, it is having multinomial i.e., more than two categories. And we have one practical example, we will also show it to you here, then we have couple of interpretation then we will come back to the conclusion.

(Refer Slide Time: 18:30)

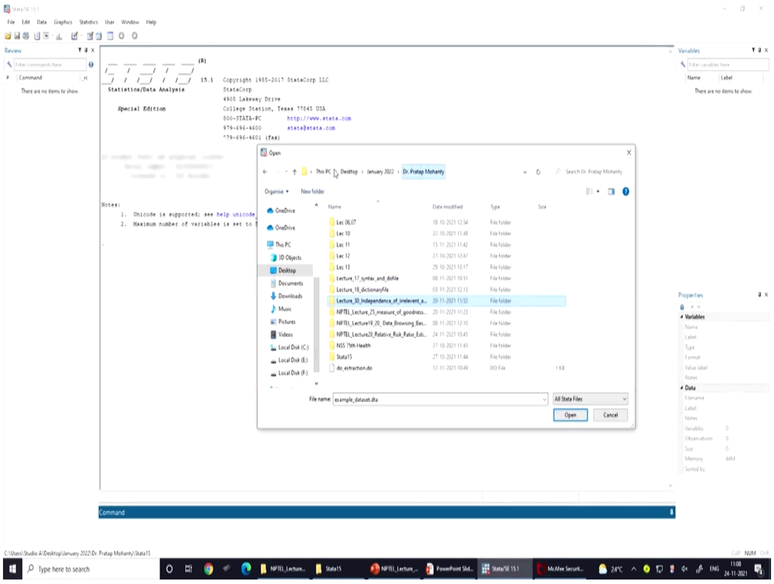


So, in the example data set that is what we are going to operate is basically on two control variables and three categories. So, that those categories we have considered as indemnity, prepaid and uninsured in the persons who are having some reply related to insurance. So, that is what we are going to operate here, and I am just going to show it, we are going to operate with the help of Stata.

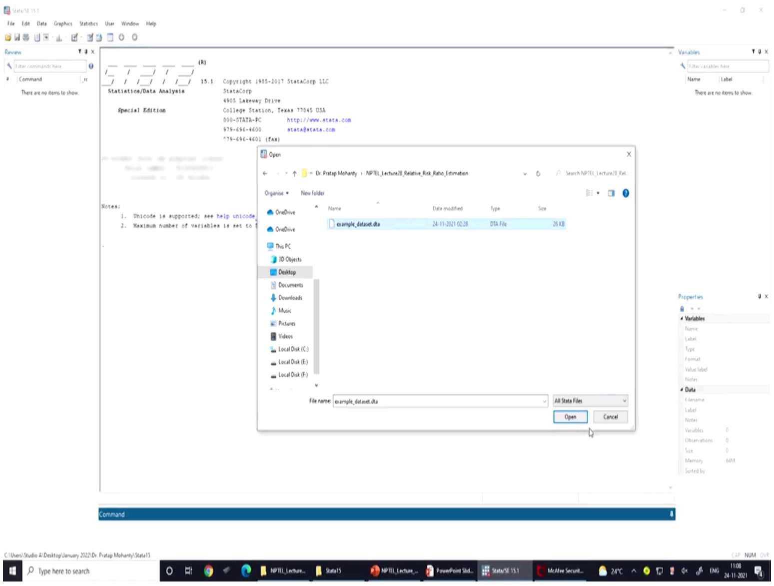
(Refer Slide Time: 19:14)



(Refer Slide Time: 19:22)

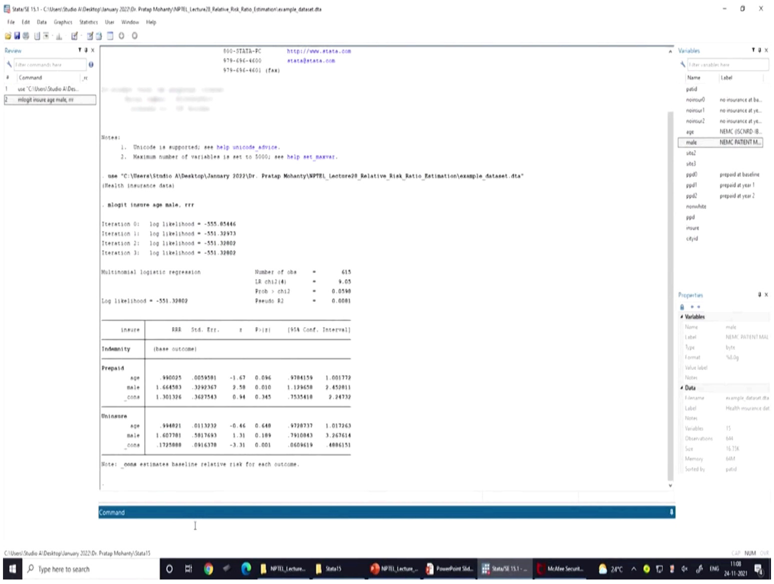


(Refer Slide Time: 19:33)



So, here is the stata window in front of us. We will go to the data and we have opened the data here and here is the no, this is not the one, there is the another one we will open in the respective lecture 128th lecture here is our data set.

(Refer Slide Time: 19:37)



Now, I am going to apply the command mlogit, mlogit is the command that we used to say it is the multinomial logistic regression model, then this is the one, then we have to apply our variables. At the end, please make sure that you are giving the command as triple rrr. And if you are not giving rrr command, then those are resulted with the coefficients only, you have to again convert with their exponent exponential values. Once we have given this command, we are coming up with the relatively risk ratio values.

Now, relative risk ratios (rrr) being displayed here. Here, since we have already said that in the dependent variable, we have three categories, indemnity, prepared and uninsured as the person reported regarding their protection on for the health policies. Those who have said prepaid, those have said uninsured, we are comparing for the base categories, it has already given basic outcome as the indemnity here, indemnity as compared to prepaid, indemnity as compared to uninsured.

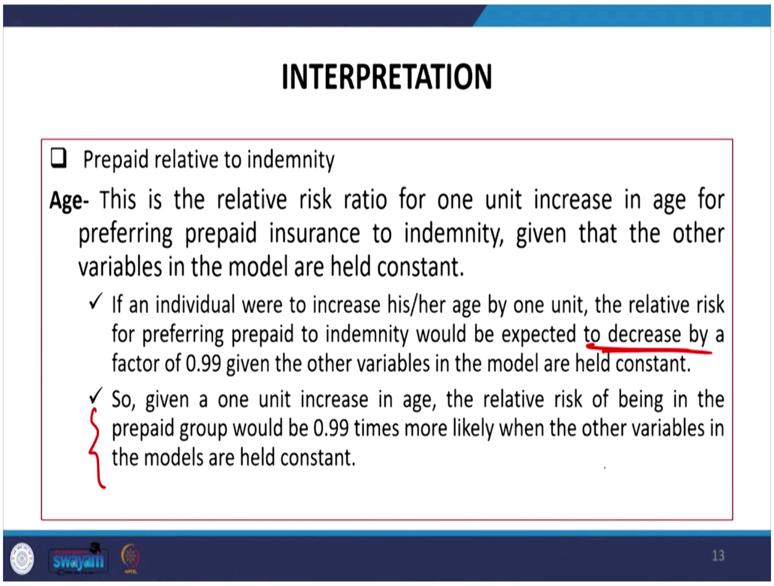
Now, you will get a very clear picture here that these are not coefficient, these are basically comparison to one ratio. Since you are saying ratio, if it is more than one, if it is less than one accordingly, you can have your interpretation. Like the age category here is having 0.99, those who have opted for prepaid insurance, as the age increases, they their chances to be as compared to prepaid as compared to the indemnity is having 99 percent.

For the age, age is giving 0.99 value, male is in favor, male is in fact, in higher probabilities, those who are under prepaid, those who are uninsured similarly, we can have different values.

Uninsured with indemnity as the base category, then prepaid with in indemnity with their base categories. Rest you also need to check whether they are significant or not. If they are significant, then that value can be considered for the analysis. If they are not significant, then accordingly, we can take a different structure of the model.

So, I am just going back to the slide we have been discussing. This is what the result we derived and we are trying to discuss this, we are trying to also discuss this relative risk ratio. And the chi square values are determined whether the model is significant or not, this has been identified with the pseudo R square value. Now, the relative risk ratio of a coefficient indicates how the risk of the particular outcome falling in the comparison group compared to the risk of the outcome falling in the reference group change changes with the variable in question.

(Refer Slide Time: 22:58)



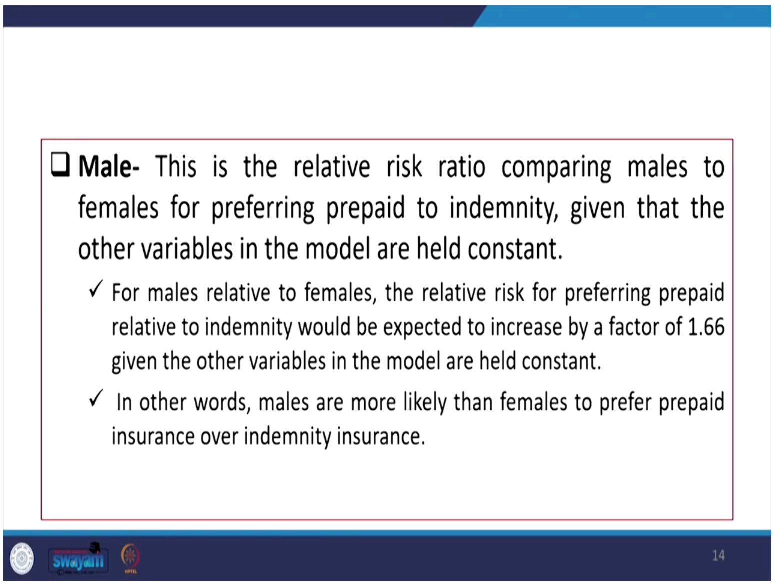
Like in this case, if you take age as the control variable in that case, the relative risk ratio for one unit increase in age with the change in age by one unit for preferring prepaid insurance to indemnity given that the other variables in the model are held constant. We are basically comparing like if all other are considered to be constant with the change of age by one year, its comparison is interpreted accordingly.

Like as per the value, it was 0.99, if an individual were to increase his or her age by one unit or one year, the relative risk for preferring preferred to indemnity would be expected to decrease since it is less than one. If it is more than one, the probability is actually higher.

Here, the values are by probability, probability since ratios are taken, probabilities are actually divided so, finally, we come up with the final ratio.

So, the person with increase in age by year, it the probability to go for prepaid option is lesser as compared to the base category that is indemnity. So, there will be a decrease in value because it is less than one. So, the value here is 0.99. Similarly, in other case, like , the interpretation is explained here once again. Given a one unit increase in age, the relative risk of being in the prepaid group would be 0.99 times more likely than the variables in the models; variables in the model keeping all thing constant or other factors constant.

(Refer Slide Time: 25:00)



Similarly, male is the category, it has the value with 1.19; 1 point male has 1.664583, here it is 1.607781. The interpretation is that for male relative to females, the relative risk of preferring prepaid relative to the indemnity would be expected to be higher by a factor of 1.66 given other control factors and those are considered to be constant. So, similarly, you can interpret any other type of variables on your interest, but I think we have already clarified about it.

So, at the conclusion, I just wanted to mention that this is most useful in case of cohort studies, this is most useful in case of health care issues, in case of health risk context.

And most importantly when you have more of dependent qualitative variables or dependent categorical variables. And it is little advance than that of the multinomial logit because this is not just going to give you the odds ratio, this is also going to give you the ratio that can be interpreted as a kind of risk, risk for a person's to be attached for a particular event.

So, these are all for this particular lecture and we will carry forward the discussion in the particular week accordingly. I hope you enjoyed and try to go through the data and go by the command which you have already sited, I am sure you will get the result and enjoy reading.

Thank you here, we will see in another class.