Health Economics

Dr. Pratap C. Mohanty

Department of Humanities and Social Sciences,

Indian Institute of Technology Roorkee

Week – 04

Lecture 20- Equity Measurement: A Practical

Welcome, friends, to our NPTEL MOOC module on health economics. We are now explaining equity measurement in this lecture. We will give you the practical or the hands-on. So far, we have been discussing various concepts of equity equality and also the theories that are aligned with the explanation of these concepts. Without delay, I think it's better to explain very specifically on equity measurement and practical. In this one, I think I have to recap what we have done.

Like in the previous lectures, we discussed about equity in health financing and distribution. Here, especially in the financing and distribution unit, we talked about vertical and horizontal equity. Then, we also discuss the concentration curve, concentration index, and Kakwani's

Progressivity Index. I think we have also cited the examples along with the theory, and we interpreted the coefficients based on the example data.

Today, we will take you through real data, where you can calculate the value and interpret it. Hence, the target of this is to measure equity, and hence, it is practical or hands-on. Specifically, we will be emphasizing the concentration curve and concentration index. Then, to derive this, we require two important key variables because it is plotted usually in the two-dimensional space. So, usually, for the concentration curve, we require a cumulative summation or cumulative figure as a percentage. I will discuss this.

So, the most important key variable usually taken in this context is for our case; so far, we have been emphasizing the health variable. And variable capturing living standards against which the distribution is to be assessed. So, we can compare how the standard of living is actually distributed over different quantiles. So, the health standard of the people is actually distributed over their living standards. So, these will define the concentration, whether it is pro-poor or pro-rich, which is to be discussed carefully.

The data could be of two types, which we usually face, especially when these are based on microdata. Based on

the macro, it is really not suggested that it be done. In microdata, the response is the people or the households, and how the response is based on their health and wealth. Similarly, you can relate to many other contexts, such as other social indicators. So, at the individual level, we target or collect the data from the household.

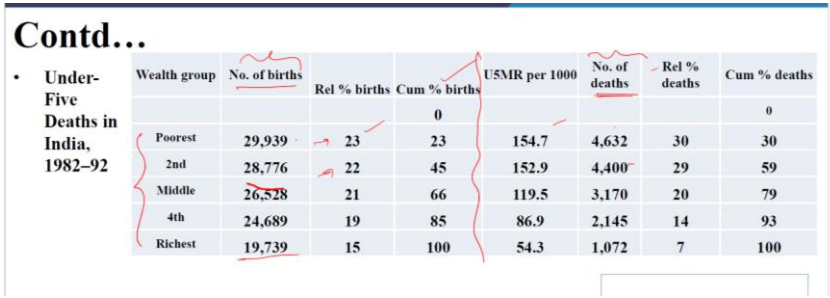
They are called raw household survey data, and these values of both the health as well as the living standard are available in different rounds of large-scale data survey data, which we have already discussed in our case. Then, in the group data, we require individual data as well as group data where the living standard is actually grouped instead of by individual information; maybe wealth or income is presented in quintiles. So, or in quintiles or maybe in deciles, maybe in triciles. So, whenever we make it in the group, then how does this group actually indicate or explain the concentration curve? Where the group data is taken, the mean value of that group variable is most important and is considered to be the unit of measurement.

The second one is the ranking of the groups, which is also important like we have to keep it lower to the upper or upper to the lower. Usually, we keep it as the poorest and then second poorest so far as living standards are concerned; we make it the poorest, second poorest, or the

middle income or the upper richest and the richest one, and so on. I will show you what it is in the data. The percentage of the sample falling into each group is also known to explain the group data. Then, the concentration curve plots shares of health variables against quintiles of the living standard variable.

In the case of group data, charts can be easily produced in Microsoft Excel. We will explain it to you through Microsoft Excel. We will also explain you through the Stata software with the help of the raw data or the individual data. But for simplicity, we can also run you can find out the concentration curve based on the raw data using Microsoft Excel as well. However, we will show you both with the help of these tools.

Below is my data as a sample; it may represent a district, a state, or a country. It is on under-five deaths in India deaths and birth, births, and deaths ranging from the period 1982 to 1992. So, here is the number of births. We have written the number of births and number of deaths. Now, in the data, as I mentioned, we require two variables. One is through their standard of living, maybe by wealth group, and it is presented with quintiles: poorest, then second level, then middle, fourth level, and the richest one.



Hence, it is in quintile, and another indicating variable is our these figures, either birth or death. So, if you have this, we can able to derive the concentration curve. The first aspect is the relative birth percentage. So, what is done this divide this is 23 percent, this is 22 percent, then and so on. This is divided by the total number; we will find out this.

Then, here is the cumulative percentage of births. Another figure from here is under 5 mortality rate per 1000 population. So, it is given, and then a number of deaths in total are written as against the poorest people, in the bracket of poorest, then similarly for second poorest, etcetera, then their relative deaths we can calculate and the cumulative death percentage we can calculate. Now, what will I do? I will use the excel.

Just two variables of interest are required. We know that the third variable is our wealth quintile. I mean, the two variables that on the horizontal level we have for each five

indicators we have actually different wealth that is starting for here we require cumulative birth percentage, which we have just derived, and the cumulative death percentage. Now, what will I do?

So, what do we do? We will insert, then we will go to the data, and then we will insert a graph; we require a graph for it. So, from the chart, you can just see I am just going through a slow process, and here is the line diagram we need to take the help of this. Then what do we need to define here? We need to actually select this; then, we select the data. Where is our data? Now, here we have the data we need to add. The first point of explanation is that we are supposed to compare with the line of equality.

Where is the line of equality, and how does our real data diverge from the line of equality? So, let us write down that line of equality. I will just do the process once again. So, we have this data corresponding to the wealth level, and I am not skipping very quickly. So, here is my data; then, I require either the birth percentage, cumulative birth percentage, or cumulative death percentage. Then I have to go to insert, then here is the chart, and then I can pick up anything, but this is most preferred because no single scatter points are not required; a smooth line is suggested.

So, once I just selected, I need not select this; what do I do? I will just directly draw it. I am going to insert it again. Here is my chart; this is the line diagram. Then I have to select this. Here is my data, and then we need to add it. First, as I suggested, you need to write down the line of equality. We have also guided our slide on equality. Then, set the line of equality, then now when I say it is a line of equality, both the variables should actually be the same, and only then will we find out.

So, I think here I will select maybe birth or maybe death. So, let us compare to death first or let us compare through these first. Then, on the y-axis, I also have to take the same; only it indicates the line of equality. Then simple. Now, one another thing: by default, excel has taken 120 in both axes; what do I do? I have to adjust this; this has to be out of 100, and then only will it work. Yes, so 100 I have to make it 100 here.

Similarly, the y-axis is as well. So, then right-click on it; this format axis will get this and then 100. Now, we have got the required line of equality. Now, you can also change the figure if you want. Sometimes, you might have seen in a different article that the figure and this presentation should be very clear. So, you can change this figure a bit; it highlights the exact line very clearly.

Now, the next step is to add the exact line of inequity, or that is where we explain the concentration curve. So, based on the real data, so what I do, we will select this, and then we will try to add it. So, we will add and insert, then add, and we will add here. Yes, so here we are just going to add the data; here is the add one, where I am writing down the curve that we are interested in deriving, which is called the concentration curve. Then, the x variable is our this birth cumulative birth percentage, and the y variable is our death.

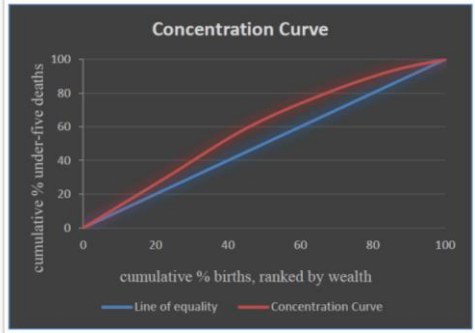
It will compare these two and find out which side it has relative bias. Now, you can see this from the diagram. Now, here is our concentration curve, and you can now see that the curve is from the line of equality; it is positioned from the left side. We have already explained earlier that when it is actually aligned at the left side, it is, in fact, explaining, pro-poor bias.

So, here is my slide. Now, so whatever the step I have said is also presented systematically; we will go to insert, then in the insert, we will select the scatter plot, then from the scatter plot, we will select the smooth line diagram either that or at the bottom whichever you take it is fine. Then rest we will add the line that I have suggested, like a line of equality, then we will plot its x variable and y variable accordingly. For the concentration curve, we are supposed

to have the cumulative percentage, which I have already shown to you. Based on all systematic guidance, we were able to find our concentration curve, which is derived based on the cumulative percentage of death as compared to birth and every time it is ranked by wealth.

So, it seems that the curve is positioned on the left side from the line of equality, as we already explained that it is pro-poor. I mean, that is, the extent of the incidence of deaths among the rich is relative to the pro-poor. Sorry, since it is on the left-hand side. So, it is pro-poor, and it is biased toward the poor, and this indicates that when the concentration is aligned at the left, it indicates that the poorest people have a higher incidence of death. So, these are all there, and we have also mentioned all these steps step by step, and we are just with that 100. So, now, from the diagram, we can just conclude that this is pro-poor in equity as it is above the line of equality. So, what we did just now is based on the group data because we presented in quintiles of the living standard.

Contd..



Now, we are explaining with the help of microdata. So, microdata will take the help of the micro responses from the survey, and graphing concentration curves in STATA can be done using the command called curve. So, glcurve, two way, or conindex. So, we will practice that is the conindex transfer concentration index. We will practice using conindex it also gives the value of the concentration index.

For practice, we have taken the data that is the latest health survey data or NSS 75th round, which was conducted in 2017-18, and we are keeping two variables for deriving the concentration curve that is out of pocket expenditure on healthcare and the living standard, that is, household income. So, we are supposed to load the data in STATA and then some of the operation we need to do. We need to

install, you know, then like install conindex, then you know Lorenz curve packages and follow the next steps accordingly. So, these are all there.

So, we just need to move to STATA. So, here are the packages we have already done. However, you need to from the screen itself you need to install this thing SSC install conindex like if I just do it. So, here is my STATA window I think we have already explained in our lectures. So, I will just do the install packages. You can also search or take the help install conindex.

So, these are the packages that have been installed. So, now, similarly, we are supposed to also install SSC install Lorenz. So, we have also installed this, but it has to be just taken checking Lorenz's consistency and verifying that it is not already installed. All files exist and are up to date; it is fine, I think. The next step is to sort the variables. Now, where is the data? I just need to mention here the data I have kept for our record example data.

So, I will just take that example data from here from my data set. I will also suggest how you can do it. So, once the data is loaded, what the data stands for is called out-of-pocket expenditure on health care, and the second one is the household income quintile. However, in this quintile, these values are actually plotted against each

individual. So, the most important aspect is that these are actually each individual data. So, we might be wrong; hence, we are saying these are called microdata.

You can just check how many observations it is 5,56,756 observations are there in total in the survey data as per the 75th round. Now, the next step is to sort the variable, which is a household income quintile. So, it has to be kept as poorest, then second poorest, then middle, then accordingly the richest one. Then, we need to go for the concentration index command. So, what I do is write it down as it is (given below). I have also shown it in the conindex.

Contd..

- ssc install conindex
- ssc install lorenz
- sort hh_income_quintile
- conindex oophe, rankvar(hh_income_quintile) robust truezero graph ytitle(cumulative proportion of oophe) xtitle(Rank of hh income quintile)
- The result will be shown in table along with figure.

```
. conindex oophe, rankvar( hh_income_quintile ) robust truezero graph ytitle(cum
> Rank of hh income quintile)
```

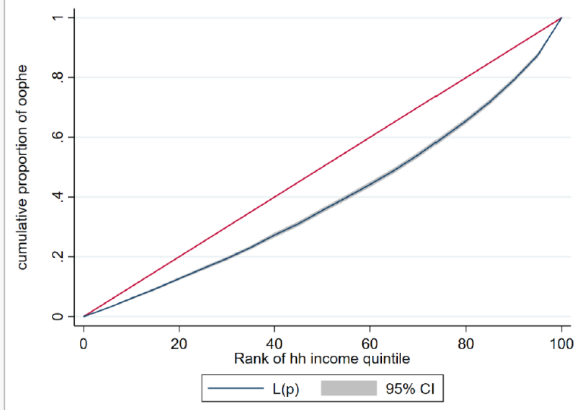
Index:	No. of obs.	Index value	Robust std. error	p-value
CI	93150	.20511516	.00551949	0.0000

Then, this variable is the first variable of interest out-of-pocket expenditure, and there has to be a comma. Then, we need to write down as rank var. This is as per the command, and then we should have the second variable by ranking. So, we need to close the bracket and then robust actually, the standard error, etcetera, is measured through robust, then truezero, graph ytitle. In the Y axis, the Y title has to be mentioned as a cumulative proportion of OOPHE. You can, in short, write down the cumulative proportion of the variable you have taken OOPHE. So, the X title is alright then in the bracket rank of household income quintile.

So, these are this is the standard approach we usually use to derive, and then with enter, we will get our desired result. So, now, you can just see this is our desired result. You can just see The X-axis and Y-axis, both of which we have already derived. And now, as per the microdata, you can just come back to the slide, and as per the microdata, we have got a reverse result. The result is also shown in the STATA window, and the confidence interval and the effective number of observation final is 93,150. The index value is also given as 0.20511516 standard error, and so on. I will just explain from my slide, which will be more comfortable for you.

The details of what we have derived based on the command, but make sure that you are actually running these and keeping them as per this command. And as I just said, our index value is positive, and it is significant. Now, it is significant because the p-value explains the significance it is as per the 95 percent confidence interval. Now, our concentration curve is derived to be right from the equal line of equality. This explains that since on the right-hand side or on the X axis we have taken the income quintile, that means your cumulative proportion of the spending is actually pro-rich and is biased towards the rich people (shown below).

Contd..



The index value is 0.2051; hence it is pro-rich inequity, and also we have explained. Weights are also required for proper and whole analysis. Sometimes, we need to represent the entire district or so. In that case, the respective weight in that round can be taken. Usually, since it is individual data, iweight can be in the bracket. So, this is just for your primary introduction, and I think the rest you can follow from this reading by Xu K.

It is one of the standard paper published in this journal, Bull Health Organization 2008. So, thank you, that is all I think you will enjoy with these practical directions. Thank you.