module_prob2_lecture7

Fri, 1/21 12:45PM 15:48

SUMMARY KEYWORDS

conditional probability, probability, unemployment, university, education, interested, unemployed, intersection, pandemic, unemployment rate, context, fraction, formula, people, denominator, population, belong, different subsets, social sciences, numerator

SPEAKERS

Sumon Majumdar

S Sumon Majumdar 00:04

Since last clip, I introduced you to the idea of conditional probability in the context of the example of a labor force survey. And what we're interested in, is looking at unemployment rate among those with university education, or the probability of being unemployed, conditional on having a university education. So let's continue with this example. So, so out of the 1000 people surveyed, we have 110, who are unemployed, 300, with a university education, and 24, who are unemployed and have a university education. Right, so the overall probability of unemployment is the 110 people out of the 1000. So that's an unemployment rate of 11%. So social scientists, we are often interested in looking at a subset of the population and looking at how a different features of them, right, so for example, we may be interested in the differences in unemployment rate among those with a university education and those without. So what we're interested in is, what's the probability of unemployment, given that one has a university education? So what we're looking at is this subset of the population with the university education, right? And how many of them are unemployed? So if you look at the subset, right, so there are 300 people with a university education. So the size of this blue circle is 300. And in this group, how many are unemployed? It's this 24, right? Because that's the part which is the intersection of university educated and unemployment. So that's 24 over 300. And which comes out to be 0.08. So it's an 8% unemployment rate among the university educated. Now, we may be interested in what how does that compare with those without a university education. So what we would be interested in is finding the probability of unemployment conditional on having no university education. So again, in this Venn diagram, let's look at the people without the university education. So it's all the people outside this blue circle. So how many such people are there, so see, out of the 1300, of university education, so the remaining 700 Don't, and how many are unemployed in this particular group? So here, it's these cuts, right, so these are the unemployed, without a university education. So this is 86, over 700. And if you use a calculator, they'll see that this is .123. So the unemployment rate among those without a university education is 12.3%. And see, if you compare between the two, it's 8%, among the university educated, and it's 12.3%, among those without a university education, so the unemployment rate is higher among this latter group. And in the social sciences, we are very often interested in looking at such differences. So as I said before, this is data from the pandemic in from the middle of the pandemic, June 2020. And people may be interested in how has the pandemic hit employment prospects of those without university education versus those with a university education. And what people would be referring to is numbers like this, what's the probability of being unemployed if you have a university education versus if you don't? Now, let me take this forward. Right? So this is in a particular context, it's easy to see right? So now let's generalize this to me more general context. So so if you're interested in finding the probability B given A right so finding this sort of a conditional probability. So what it means is that we're going to look at the set of people with characteristic A. And see what fraction of them also have B. So what we're going to look at is, what fraction of this segment of the population also belong to B. So what we will, as we did before we compared this set, right, the intersection between A and B, that's the set of A's who also belong to B relative to the total size of A. So that's the idea of conditional probability. Now, if you want to write this in a mathematical way, so remember, earlier, we had the rule for general multiplication. So the probability of A intersection B, the probability of A and B occurring is probability of A time's the probability of B given A. And here, what we're interested in is figuring out this conditional probability, what is this? So how do we isolate out for this, we can isolate out for this by if I take this P of A to the denominator here. So what we get the formula is this. So see, the conditional probability of B given A is the probability that A, that something belongs to the intersection of A and B, right, so it's this divided by the probability that it belongs to A, so divided by this? So again, it's the fraction of a who also belong to be. So off these A's, right, what fraction also belongs to B. So that's the conditional probability of B given. Now, again, if we go back to the labor force, survey example, and use this concept, and that's what we have been doing, but like, let me show you how you can use this particular formula. There, right? So if you're looking at unemployment rate, among the university educated, so unemployment given university, so what you have in the denominator is all the is the probability of having a university education. And what do you have on the numerator is the intersection between unemployment and university? Remember, we using this formula, it's the intersection between A and B. So the intersection between unemployment and the university divided by whatever we are conditioning? Right? We're conditioning upon University. So that's the denominator. So in our particular example, right, what's the probability of having a university education to a 300? out of 1000? Right, and what fraction had unemployed meant unemployed and had university education? That was this, so 24 out of 1000? So if you divide one by the other, so see the 1000s cancel out. So that's 24 over 300, that's 0.0. Similarly, if you're looking at what's the probability of unemployment, given no university, right, so that probability of no university is in the denominator, and in the numerator is unemployment and no university. So if we look at no university, the probability of that is 700 over Townsend. And what about unemployment and no university it's these guys here. So that's eight He's six upon. So this comes out again, if you cancel these two things, it's 86 over 700, which is point 123. So this is, again, using this formula to figure out conditional probability. But the basic idea, and this is what I want you to take away from here is that finding conditional probability of B given A is that you look at a right that goes in the denominator, and whatever is the common portion. That's in the numerator. So that's this formula here, probability of A intersection B over probability of A. And this idea of conditional probability is very useful, and very widely used in the social sciences, because we are very often interested in looking at comparing between different subsets of the population. So for example, we may want to see, okay, what's, let's say? What's car ownership among people with? who live in towns versus people who live in villages? Right? Maybe car ownership among people who own an income above a certain level, versus people who earn income below a certain level? Right? We could be interested in, like in a medical context, what's the rate of lung cancer? Among people who smoke versus people who don't smoke? In the pandemic context, what's the risk? What's the probability of contracting COVID? Given you're vaccinated, or the conditional probability of contracting COVID? Given you're not vaccinate? Right? So these are all conditional probabilities because you're conditioning on being vaccinated or being unvaccinated. So this idea of conditional probability is very widely used. Let me end this clip with one. One other related concept. This is the idea of marginal probability or total probability. Okay. So coming back to this example, if you look at what's the probability of being unemployed, right, so among the unemployed, they may be people with a university education, or people without a university education. If you want to look at the totaled probability of unemployment, you have to account for unemployment among the university educated versus those among without university education. So what we're going to make use of is remember the law of total probability that we used before? Right, so the probability of unemployment is firstly, you look at the fraction of people who have university education, and look at unemployment among them. Plus, right, look at the fraction of people without University and look at unemployment among them. Together, this will give you the overall probability of unemployment. So this is sometimes this is called the total probability or the marginal probability. And the reason it's called marginal is because we are marginalizing all other events, right? We're taking account of all the various other events which could lead to employment, right? We are taking account of education, right? Whether you are university educated or not university educated and the impact on unemployment. So in this particular context, right, so if you put in the probability of university education here is 300 over 1000. So that's 0.3. We already calculated probability of unemployment given university that is 0.08. So multiply these two, add, right, then the probability of no university that 700 over 1000 So that point seven times probability of unemployment given no university, which we calculated before as .123. And if you add all of these together, this will become point 110. Right. So this is the probability of unemployment that we found before. So this again is called the marginal probability or the total probability of unemployment. So, let's stop this clip here. And in the next clip, I'll take up another example on trying to find conditional probability just to get you used to this idea