module_prob2_lecture9

Fri, 1/21 12:45PM 11:46

SUMMARY KEYWORDS

probability, marginal probability, intersection, helmet, table, mask, conditional probabilities, frequency, survey, party, cyclists, data, young voters, joint probability, events, voters, age, observations, cell, divided

SPEAKERS

Sumon Majumdar

S Sumon Majumdar 00:04

So in the last clip, I introduced you to the idea of two way frequency tables, and depicted the, the data from the observation of cyclists on Stanford campus using a two way frequency table. So we had a table like this, which depicted the number of people with masks number of people with mask and helmet, no helmet, and so on. So how do we interpret this data in terms of the probabilities that we were talking about before? So, you will look at this two way frequency table, right? And look at a cell like here. So this is the set of people who are both helmeted and wearing a mask. So in our language before, this is the probe this, if you look at this cell, this is the set of people with helmets, intersection mask. So if you want to debate, what's the probability of the event of helmet intersection mask, so that means people wearing helmet and mask, so this will be the number of people here divided by the total number of observations, which is 400. So if I go back to the table, the number there is 29. So this will be 29, over 400. And if you calculate that using a calculator, this will be 0.724. So that is how you can convert the observations in the frequency table into probabilities of joint events, right? The probability of the intersection of event H and M. Similarly, for example, if you look at at this cell here, so this is no helmet intersection, mask, again, if we go to the numbers here, right, so there's 134 observations in that cell. So the so the probability in this sample of encountering someone who has no helmet and mask, that's going to be 134, over 400. And if you calculate that, that's going to be 0.33. So this week, you can interpret this cell and this cell as well. Now even more interestingly, if you look at the cells in the margins, right? So remember, when we go back to the original table, right, if you look at this 71. So this is adding up the helmet and mask, and helmet and normals. Great. So adding up these two numbers gave you the total number of people wearing helmets. So if we go back here, and look at that 71 over 400 That's the probability of wearing helmet among Stanford cyclists, right. So, this here turns out to be 0.1775 Right. And in the language that I had introduced before, this is the marginal probability, this is a marginal probability of wearing helmet among Stanford cyclists. And similarly, if you look at here, right, so, in the previous table, this is the total so if I go back here, right. So, this 163 is the total number of people who are wearing masks in this date, right. So, if I depict this here, right, this 163 over 400 This is the probability of wearing a mask in the sample right? Again, this here is the marginal probability this is the marginal probability of right. And see, this gives you another inspiration for the name marginal probability because these are things which occur at the margins, right? This is a this margin this is that this margin, but what it does give you is the probability of age Right, and the probability of H comes, comes in, through looking at people. So those who are wearing helmet and wearing masks, and some who are not most, right, together, this gives you the total probability or the marginal probability of H. Now, once you have these, you can also derive the conditional probabilities. So suppose you want to look at the probability of mass given helmet, m given H. So let's use the formula that we had before. So this is probability of M intersection H divided by probability of H. Right? So, again, remember that what's on the denominator is what you're conditioning on the probability of that, right? So probability of age because you're conditioning on age, and on top is the intersection of M and H. So now, let's go back to the table, right? So the probability of age is this here. So that's .1775. What's the probability of M intersection H, this is this here. So that's 0.0725. And again, if you use your calculator to figure this out, this will be point four zeros. Sorry, that'll be .408. Similarly, you can look at em given an H, that is no helmet, right? So what we'll have to do is look at it. So that will be on the numerator will be the intersection of these two events, M intersection and h divided by probability of NH, right? So it will be the probability here divided by the probability here. So this is how you can figure out conditional probabilities from this two way frequency table. So just to solidify that idea, so let me ask you to find another conditional probability from the same table. Right? So you may want to stop the clip at this point, and try out this clicker question. So this clicker question, I'm asking you to find a conditional probability. So this is probability of H given. And again, we are going to use the formula that we were discussing just a second ago. So this is going to be probability on the numerator is going to be the intersection of H intersection M, and M, the denominator is going to be the probability of the conditioning event, that is probability of M. Right? So if I go to the stable, what's the probability of H intersection M? That's this? So that's five divided by the probability of M. That's this? That's .4075. Okay. And if I remove the decimals, that 725 over four zeros. So the correct answer here is this. So this is how you make use of two way frequency tables to figure out the different probabilities, the joint probabilities, that means the intersection of two events, the marginal probabilities, which are the total probabilities of particular events, and also conditional probabilities, which come about by dividing the joint probability by the marginal so will now just again, to make sure you have this idea about two way frequency tables and how to use them. Let's do one more example. So this example is, is like is based on an on a survey among groups of voters about which parties do they support so here, so 1000 voters were were surveyed about the support for three parties, party A, party B and party C and the two characteristics here, right one is the party that they support. And secondly is the age of the voters, right? So there is some group of voters who are under 30 years. So these are the young voters, then those with age between 30 and 64. These are the middle aged voters, and those with ages 65 Plus, so these are the older voters, right? And what this survey found is that, so if we look at the younger voters, right, so most of them, right, 180, support party 120 support B and 100 support C among the middle age voters, but it's different. Most of them support party C or more of them support party C than the other two parties. And among the older voters, it's more larger support for party a, then comes party C and party B. Right. Now, the question that will probably arise, like, you know, this sort of surveys of orders are done all the time, and particularly so before elections. Now, the question is, how can we use the data from the survey to extrapolate what might happen in the total population? Right? So from this data, can we see can we say something about which party is more likely to win the election? So let's stop here, and I'll pick up on this in the next