module_prob2_lecture8

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

SUMMARY KEYWORDS

helmeted, helmet, probability, mask, conditional probability, data, wearing, cyclists, stanford, risks, point, figure, derive, clicker question, hypothesis, frequency, mass, venn diagram, students, question

SPEAKERS

Sumon Majumdar

S Sumon Majumdar 00:04

Let's consider another example on conditional probability. This is based on some fun data which was collected by a Stanford Stanford professor who wanted to kind of examine the point, whether because of COVID, we put a lot of a lot more emphasis on the risks that we everyone is talking about now, which is COVID, and forgetting about some other risks which occur in day to day life. So in particular, what he did is that one day in the Stanford campus, he he studied, he counted the number of vise by cyclists on campus who are wearing helmets, versus those who are wearing masks. And of course, there were some who were wearing it. So what did he find? So, so in his data, so he found 400, Stanford by cyclists, observed 400 by cyclists of them 163 Were mass 71 were helmeted and 29, with both mask and helmet. So, so this is by this particular professor Maxwell Mayer, and he's published his article in the Stanford review earlier this year. So what what did you find? What was this point? This point was this of the 400, Stanford by cyclists, you see 163 were masked. Right? So they were wearing this mask outdoors, versus 71. Only 71 who wear helmet, right, so a lot fewer were helmeted as compared to masks. And what maximum mayor's point is that? Riding a bicycle without helmet is pretty dangerous because you can hurt your head. And Stanford, in fact, has even given away free helmets. But in spite of that, there's a considerable small fraction of people who are helmeted as compared to the fraction who are masked and wearing a mask against COVID. Outdoors. That doesn't do much in like, you know, the if you look at the CDC recommendations, it's not one of the one of the recommendations that they do. Right. So it reduces the risk only very slightly, but wearing a helmet, while by cycling, reduces considerably the injury of to one's head from afar. Right. And so what his point is that students are taking a lot more precautions against risks, which are being talked about prominently these days, and not taking as much safety against risks, which are always there, but not being that talked about entities. So what I'll do here is I'll ask a more subtle question. Right. So is there a difference in masking between helmet and non helmeted students? So the type of question that I'm interested in, is that maybe the people who are helmeted, right, so they are more safety conscious. So so they wear helmets. And maybe the people who don't wear helmets, they are not as safety conscious, they are more, they're not as risk averse, right? So what one may expect is that people who are wearing helmets, they're more likely to also wear masks versus those who are not wearing helmets, right. So what do we have to do to test this hypothesis? What we need is to look at the conditional probabilities, right? So what we are, what we need to look at is what's the probability of wearing masks among those who wear helmets versus what's the probability of masking among those who don't wear helmets? So given this data, what I would like you to do is try and see if you can figure out this probability on your own. So let me set this up as a clicker question. So you may want to stop the video at this point, and then the clicker question. So hope you got a chance to take a crack at this figuring out this conditional probability. So what we're looking for is the conditional probability of mast given that one whereas the help, right, so let's look at in this Venn diagram, right, let's look at the people who are helmeted. So there are 71 of those. Right? How many of them are masked? There's 29. So the probability the conditional probability of masking among those who wear a helmet is given by 29 over 70. So this is A so the correct answer to this clicker question is T. So, so going back to our original question, right, so, in terms of the data, we have 29, who are both masked and helmeted, right, there's 42, who are helmeted but not masked. And there's 134, who are masked, but not help. So the probability of masking given that one wears a helmet is 29 over 71. And that works out to be .4085. What about the so the remember our formula before, you can also divide derive it in the following week, again, and I'm doing this just to remind you of that formula. So the probability of mass given helmet is probability of mask and helmet on the top right, divided by the probability of helmet. So what's the probability of mask and helmet so that's 29 Out of the 400 and the probability of helmet so there are 71 of these divided by 400. Right? So if you cancel these two out, so that's 2970. Right, so now, let's compare this probability among what's the probability of masking among non helmeted students. So how would you find this? Right, so? So there are total of 71 people who are helmeted, right, so the remaining so that's 400 minus 71. So that's 329, who are non helmet? So this is over 329. And on this group, how many are arm are masked? Right? So that's this, these people here, we're not wearing a helmet, but a mass. So that's 134 over 329. And if you use your calculator, this come out to be point 4073. And so, going back to my question, is there a difference in masking between helmet and non helmeted students? See, if you compare these two probabilities, there's not much of a difference. This is .407. Maybe maybe slightly, slightly smaller, but almost virtually identical to this, which is point 408. for that. So it seems that it's not helmeted behavior that is driving maski. Right. So my hypothesis that the people who were helmeted are probably more safety conscious, and therefore more likely to wear a mask that as compared to those who don't wear helmets, that hypothesis is not borne out by the date. So this is how you use conditional probabilities to, to answer questions like the one that I just raised. But the main takeaway here is how do you figure out conditional probabilities from data like this? Now what I want to do one other thing before I move on from this example, is the idea of again, the marginal probability. Right? So what we're interested in is what's the probability of masking overall right so masking can be among the helmeted or among the non helmeted, right. So, to figure out this, this probability of masking What do I need to look at his what fraction are helmeted times the probability of masking among them? Plus, what fraction are no helmet right and What's the masking? Amanda? Okay, so she tried this out on your own right? And this should this number should come out to be .4075. Right. So this is how this what we're figuring out is the total probability or the marginal probability of mosque. So next, what I want to do is introduce you to a new way of presenting data, right. So so far, we have used the like, represent the data in in at least two different ways. One is using Venn diagrams, the other is using probability trees, I'm going to show you a third way of how data is presented and how you may encounter data in different contexts. This is what's called two way frequency tables. And these sort of tables are useful way of presenting data where there are two characteristics. So for example, here, right, this data had two characteristics. One is helmet and the other is masking. Right? So let me show you what a two way frequency table looks like. So So here is an example of such a table, right? So on this side, I put down mask and no masks. That's one characteristic. And on this side here, I put down the other characteristic, which is helmet and no helmet. So what do I have have data on? Right? So I have the data on how many people were mask and helmet, which is 29. of the total number of people who are wearing masks. So this total is, is the total of this entire row. Right, which is all the people who wear masks. So I know from the data connected by Maxwell Mayer, that's 163. Right? So from this, I can figure it out among the masked who are not wearing a helmet. So that's 163 minus 29, which is 134. Right? How do I fill out the rest of the blanks? So remember, here, I know the number of people who were wearing a helmet that 71 Right of them 29 Are mask. So who are the people who are wearing helmet but no mass that 7971 minus 29 which comes out to be 42. Okay, now, I know so here is 163. Right? So the total number of people who have surveyed is 400. So 163 are wearing a mask. So how many are not wearing a mask? That's 400 minus 163 which is let's see 237 Right. So, so 237 people are not wearing a mask, right of them. 42 are helmeted. So the remaining are non helmeted. So that's 237 minus 42. So that's 195. Alright, and now if we add these two things up, we get the total number of people without an helmet. So that's 130 for this, so that's 320. Right? So I filled out all the entries of the of the table, given the data that we have. Right, so this is a two way frequency table because what it gives is the frequency of each of these cells. So this is a cell of people who wear a mask as well, a helmet. This is the number of people who wear a mask but no helmet, right? So this is no mask and helmet. And this is no mask no health, right. So these are this group of people here. Now once we have this frequency table, now from there, we can derive probabilities. So that's what we'll be doing next. So let me stop the clip here and next we will be deriving the probabilities using this to be frequency