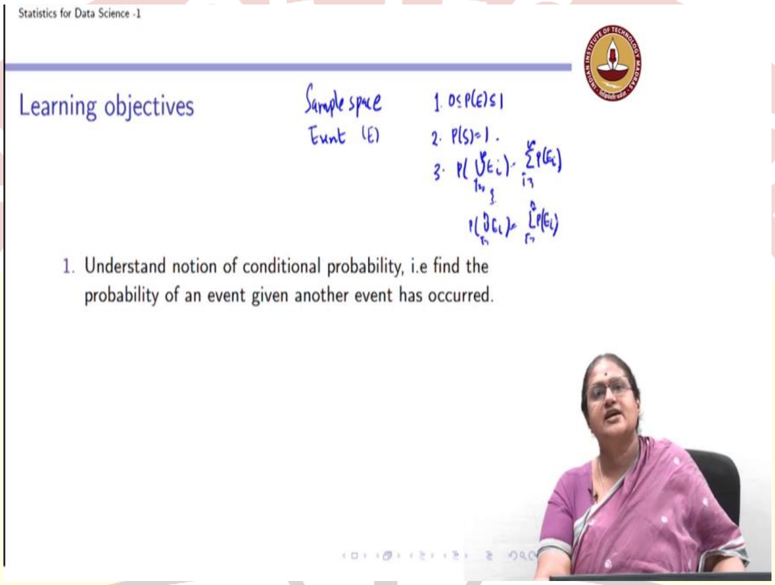
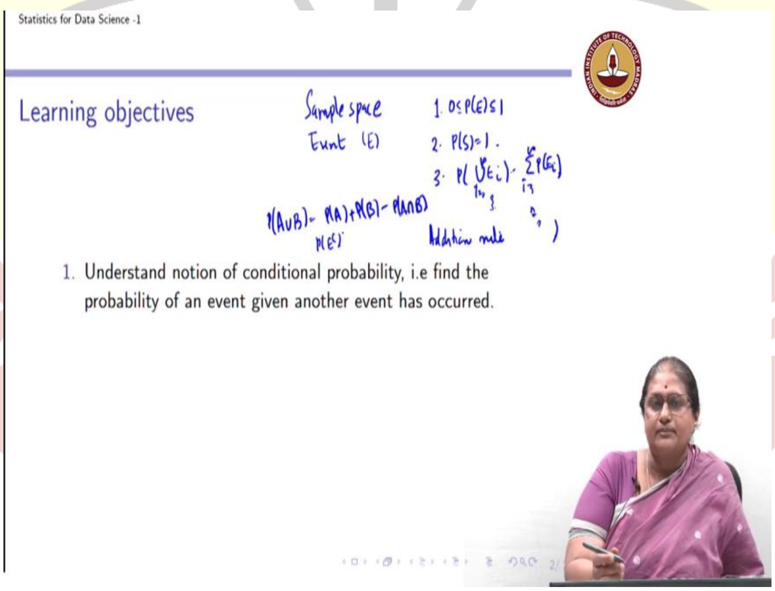
IIT Madras

ONLINE DEGREE

Statistics for Data Science – 1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture No. 7.1
Conditional Probability – Contingency Tables

(Refer Slide Time: 00:21)





So, what we have learned so far is we set up the axioms of probability, but before we set up the axioms of probability we define what was a sample space, we define what was a event and we if we have a event E, the axioms of probability said that $0 \leq P(E) \leq 1$, probability of a sample space is equal to 1 and probability of a countable union is equal to the sum of the probability when I have disjoint or mutually exclusive events.

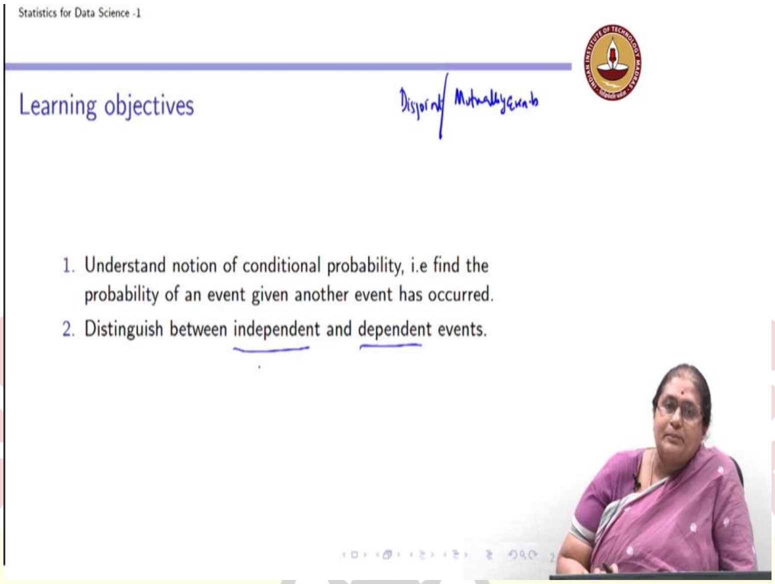
From here, we looked at a finite union and we said that the probability of a union of a disjoint set is the same as the sum of the probabilities. We also looked at what we refer to as the addition rule and we looked at when I do not necessarily have events that are mutually disjoint I come up with a probability of the union and I refer to that rule as the additional rule which says that the probability of a union of any two events is $P(A) + P(B) - P(A \cap B)$.

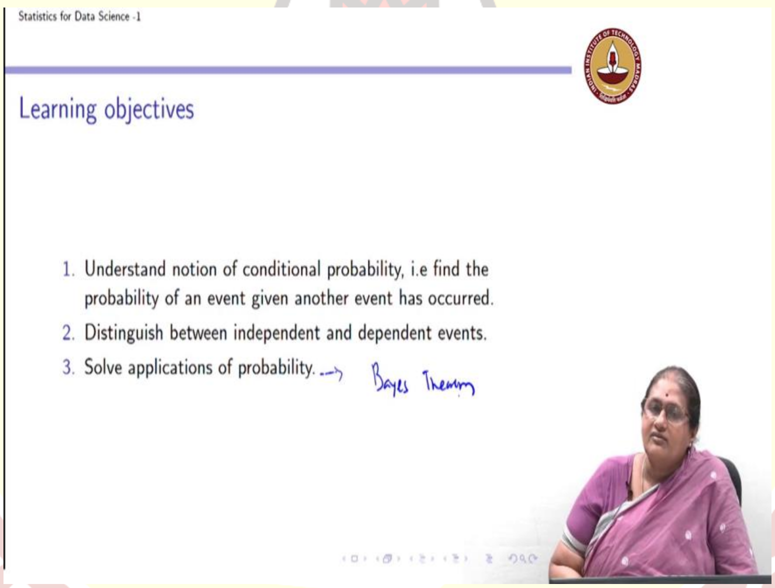
We also define what is the probability of a complement of a event and then afterwards we worked out certain examples of this theory which we develop. So, the next important thing which we are going to understand today is how do I compute probabilities when an event is conditioned on another event. Now, why is this important? Sometime I but have information or partial information of something that has happened.

Now, if I want to know what is the probability of a event happening conditioned on that event, for example I could condition I am tossing a coin twice I know that my first toss is a head, I would want to know that what is the chance of me getting ahead in the second toss conditioned on the fact that the first toss was a head.

So, I am conditioning, so I am using the information I have from the outcome of the first toss. So, this is where we introduce the notion of conditional probability, which in other words is the probability of an event occurring conditioned on or given that another event has occurred. So, we will understand what is conditional probability.

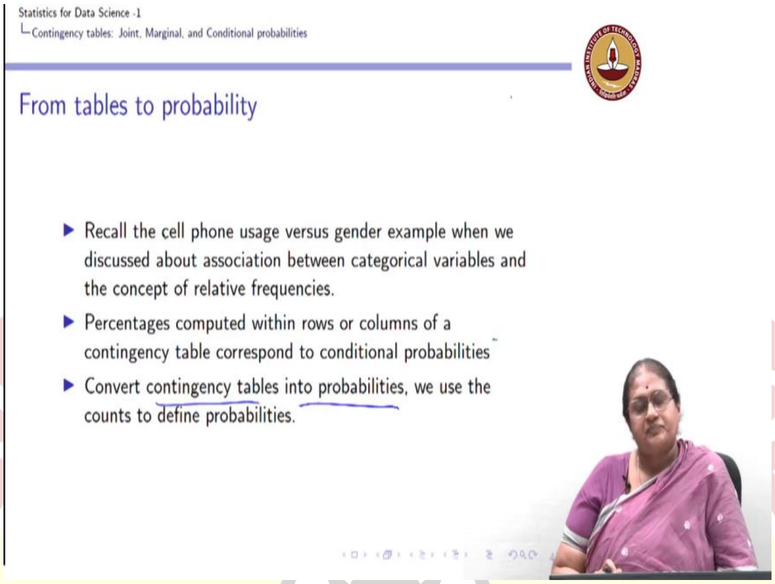
(Refer Slide Time: 03:09)





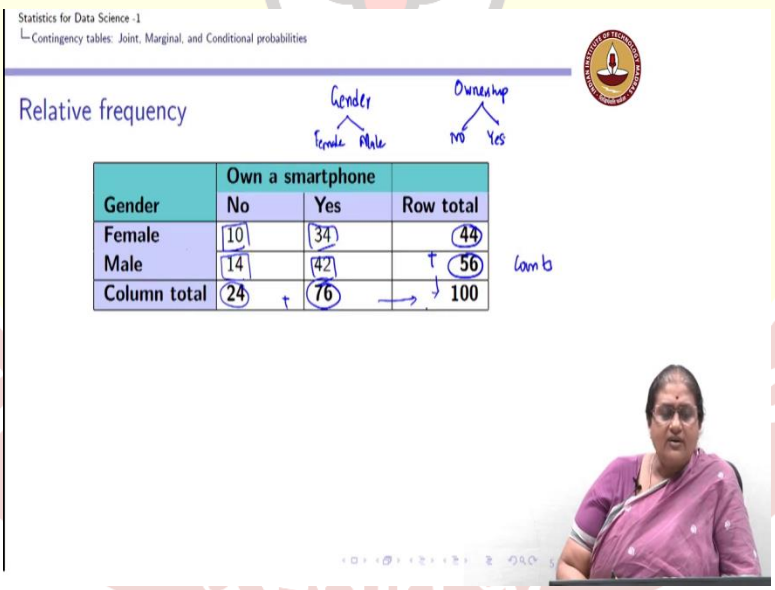
The next important concept is once we understand what is conditional probability we will introduce the notion of what are independent events and hence we will understand what are dependent events. We have already seen what are disjoint or mutually exclusive events, we are going to introduce what are independent and dependent events. And finally, we will apply all the concepts which we have learnt in for we will also introduce what is called Bayes theorem and we will apply certain concepts of whatever ever we have learnt in real-world applications. So, how do we motivate our need for conditional probability?

(Refer Slide Time: 04:01)



Recall that we looked at association between variables in particular when we looked at association between categorical variables we introduced what we refer to as contingency tables.

(Refer Slide Time: 4:16)

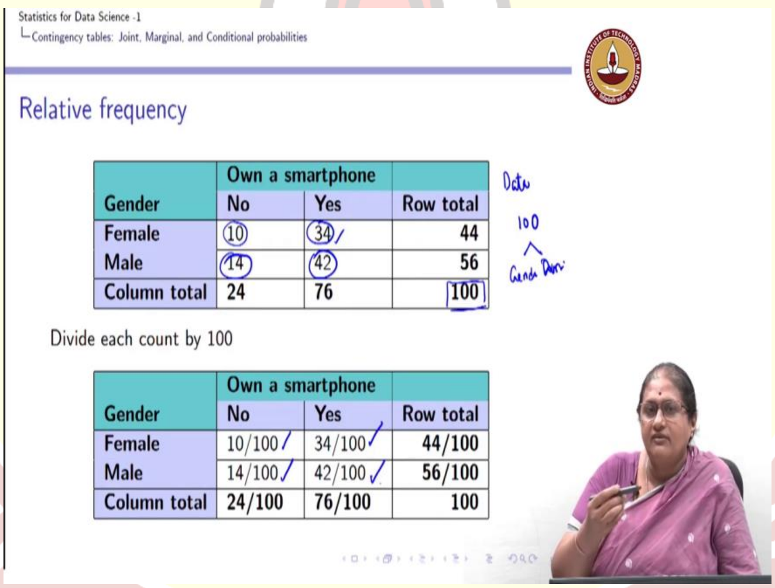


So, if you recall the cell phone versus gender usage this was what we had as a contingency table, I had two variables the first variable was gender and there are two categories of this variable which is female and male, I had the second variable which was ownership of a smartphone, which was off again two categories, no or yes and I was interested in knowing about an association between these two variables.

So, the questions we post was what is the relative frequency of a female if a person is a female what is the relative frequency of a female owning a smartphone or of a male owning a smartphone or what is the frequency of a person who owns a smartphone being a male or a female. These were the equations which we asked. So, we are going to revisit that contingency table and we are going to actually convert these contingency tables into probabilities.

So, recall what are the this 10 and 34 and 42 where the counts, by counts what I mean is of 100 people, 10 people are both female and do not own a smartphone, 34 people are female and own a smartphone, 14 are male and do not own a smartphone and 42 are male and own a smartphone. So, you can see that people who do not own a smartphone are 24 people who own a smartphone or 76, number of female candidates are 44, number of male are 56, this adds up to 100 and this also adds up to 100 people.

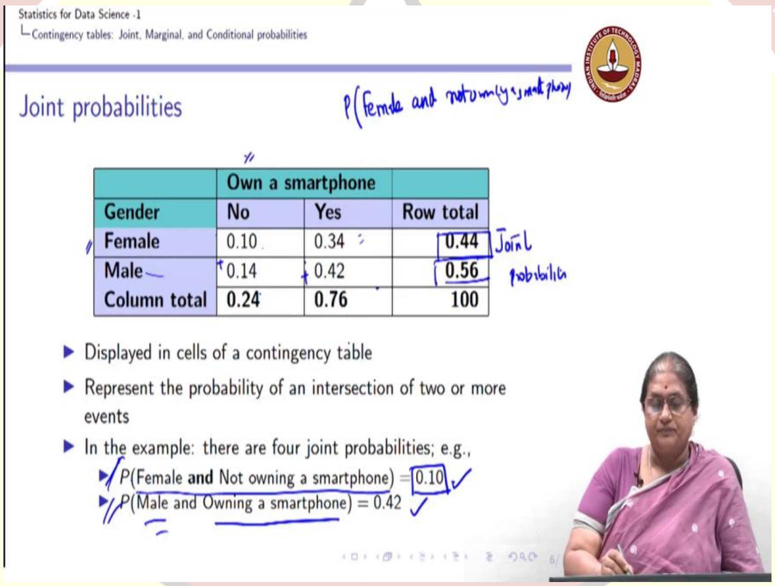
(Refer Slide Time: 06:27)



So, now let us construct a relative frequency table, we introduce what we are row relative frequencies and column relative frequencies when we studied about contingency tables. Now, we are going to introduce what is a relative frequency table. Now, what do you how do I get a relative frequency table? The total number of participants are 100, I divide each one of the entries in the table by this total number of participants, so I get a $\frac{10}{100}$, I get a $\frac{34}{100}$, I get a $\frac{14}{100}$and I get a $\frac{42}{100}$, so these are my relative frequencies.

Now, if I have to visualize or if I have to interpret this as probabilities, so first remember this is from a data I collected from 100 people where I come I actually recorded what was their gender and whether they owned a smartphone or not, ownership. Now, suppose I want to convert this into a chance, suppose I have the hundred and first person, so the questions I might ask is what is the chance that the hundred and first person is a female, what is the chance that the hundred and first person owns of smartphone or what is the chance that the hundred and first person is both female and owns a smartphone? So, these are what are the probable questions I can ask. So, I want to use these numbers which I have here to convert them into probabilities.

(Refer Slide Time: 08:17)

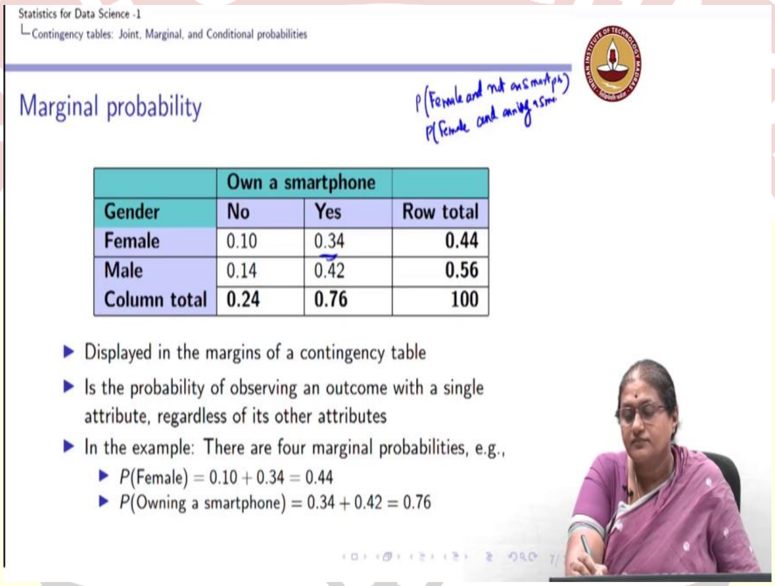


The first probability will introduce is what is referred to as the notion of joint probabilities. So, again recall this is my $\frac{10}{100}$, this is my $\frac{34}{100}$, so let us see what this entry means, this entry is equivalent to the probability of a person being a female and not owning a smartphone, that is what I have listed here, female and not owning a smartphone and that probability is given to us by 0.10, probability of a male and owning a smartphone, so male and owning a smartphone is 0.42.

So, what are the entries here, these entries are referred to as joint probabilities, the reason why they are referred to as joint probability it is the probability of two events happening together and what are these events that are happening together either female with owning a smartphone or female with either female with not owning a smartphone or female with owning a smartphone, male not owning a smartphone and male owning a smartphone.

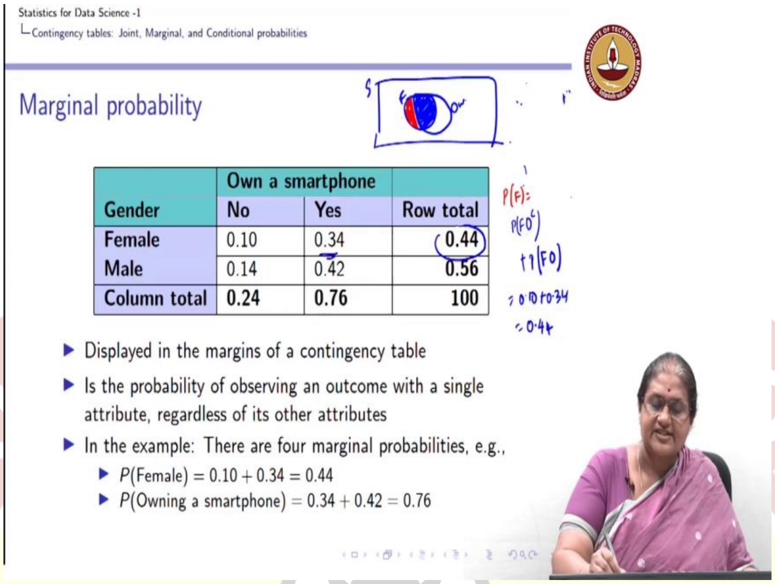
So, the entries inside this contingency table are referred to as joint probabilities. Now, we can see that this 0.44 is 0.10+ 0.34 and 0.56 is 0.14+0.42, similarly you can see that 0.10+ 0.14 is a 0.24, 0.34+0.42 is a 0.76. Now, let us understand what is this 0.10.

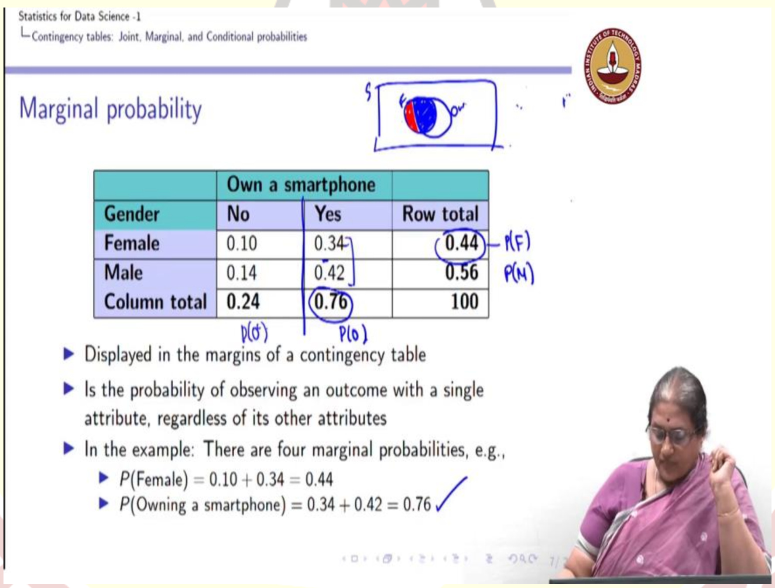
(Refer Slide Time: 10:29)



0.10 is nothing but the probability of a female being a female and not owning a smartphone, whereas 0.34 is the probability of being a female and owning a smartphone.

(Refer Slide Time: 11:05)





So, if I want to know what is the chance that you had a female let us understand this using a Venn diagram, so this is my sample space, this is my event of a person being a female, let me have this as event of owning a smartphone, so this region which is being shaded in blue is the region of being a female and owning a smartphone. The blue region is the region of being a female and owning a smartphone.

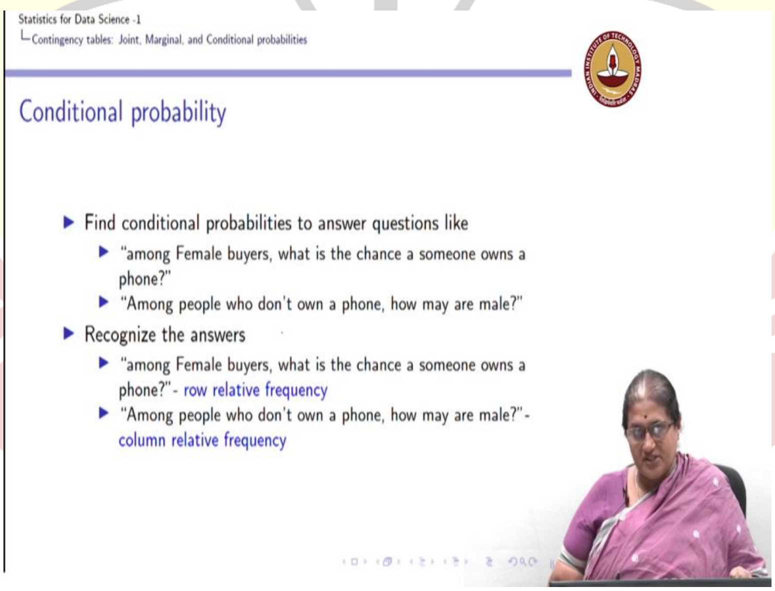
The region of red is being a female and not owning a smartphone, so you can see that if I want to know what is probability of F, this probability of F is nothing but the union of the red region, the red region is being a female and not owning a smartphone that is why I am putting o compliment union, probability of being a female and owning a smartphone.

Now, this is given to us to be 0.10, this is given to us to be 0.34, so you can see that this works out to be 0.44, which is precisely what I have here and these are what we refer to as marginal probabilities and this is what how I have converted the entries of a contingency table to represent A joint probabilities and B now marginal probability.

Similarly, if you look at the variable of owning a smartphone so you can look at this column which is a 0.34, so I can own a smartphone either, so the probability of owning a smartphone is probability of being a male and owning a smartphone and probability of being a female and owning a smartphone. So, if I add up these two entries I get 0.76 and that is again my marginal probability.

So, I have 4 marginal probabilities, this is probability of being a female, this is probability of being a male, this is probability of owning a smartphone and this is probability of not owning a smartphone which is probability of O^C.

(Refer Slide Time: 13:38)

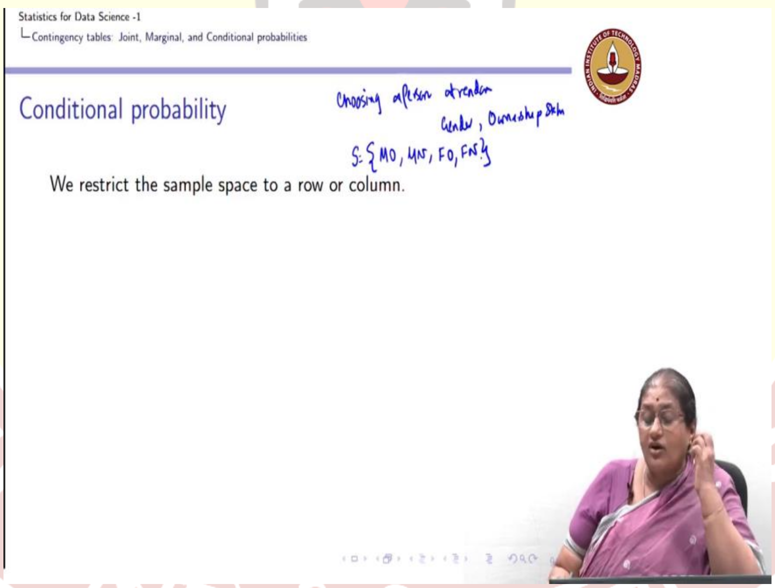


So, using these three let us introduce now the notion of conditional probability. So, when we talk about conditional probability, we are seeking answers to questions like among female buyers, what is the chance that someone owns a smartphone? So, what is the partial information that is given to us? The partial information that is given to us here is among or I am conditioning it on female buyers or among female buyers, I know that the buyer set I am looking at now as the female buyer set. So, among these buyers, so how many females do we have? We have 44 females I am interested I know how many among them or smartphone? Or do not own a smartphone.

The second question is among the people who do not own the first smartphone, so how many of them are there? 24 people. So, these are the total number of people who do not own a smartphone, I know there are 24 people, I am interested in knowing what are the chance that I have a male or how many are male. Remember we again introduced the concept of a row relative and column relative frequency when we discussed about the contingency table.

So, to answer the first question basically I am looking at a row relative frequency whereas to answer the second question I am looking at a column relative frequency. In other words to put it in the framework of a conditional probability I am restricting my sample space.

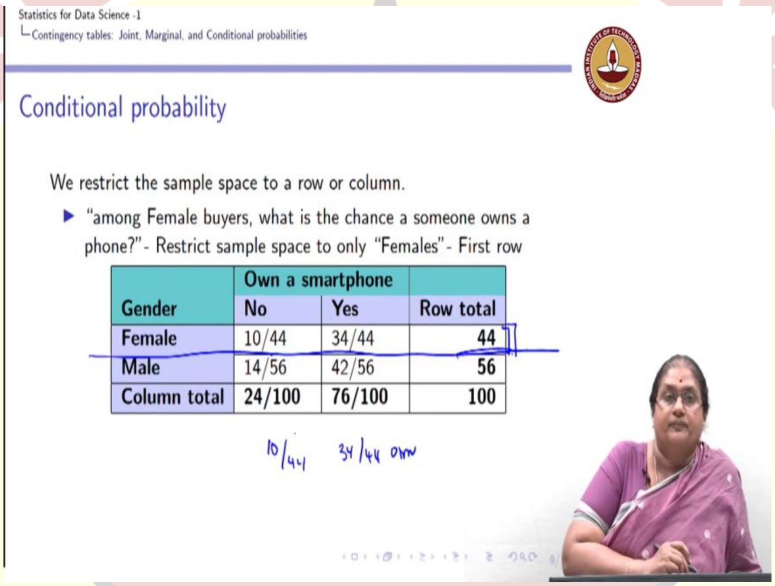
(Refer Slide Time: 15:34)

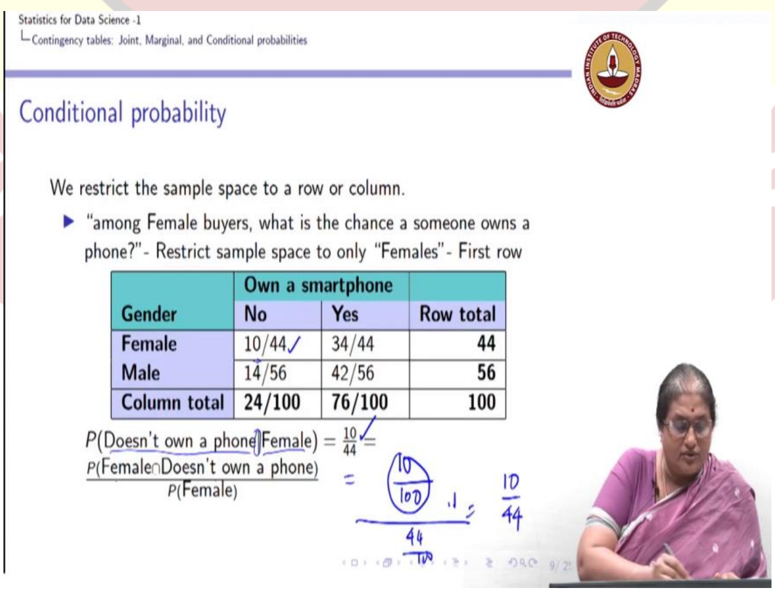


So, if my sample space earlier was having so if I am one way I can so first I need to understand what is my experiment here. So, in the experiment I am choosing a person at random, I am recording their gender and their ownership status, by ownership status I mean yes or no. So, the outcomes possible outcomes are I could have a male and I could have a person who owns I could have a male and who does not own M with does not own a female who owns and female who does not own.

But here remember there are not equally likely. So, I need to find out a way of how to compute these probabilities. So, the way to compute this probabilities I first I work on what I call a restricted sample space. So, by restricted sample space if my original sample space is male owner male does not own female owns and female does not own and I know some information about the person, suppose the information is given that the person is female, then I restrict that sample space to only female owns and female does not own, I restrict the sample space. So, how do we compute it you restrict a sample space.

(Refer Slide Time: 17:12)



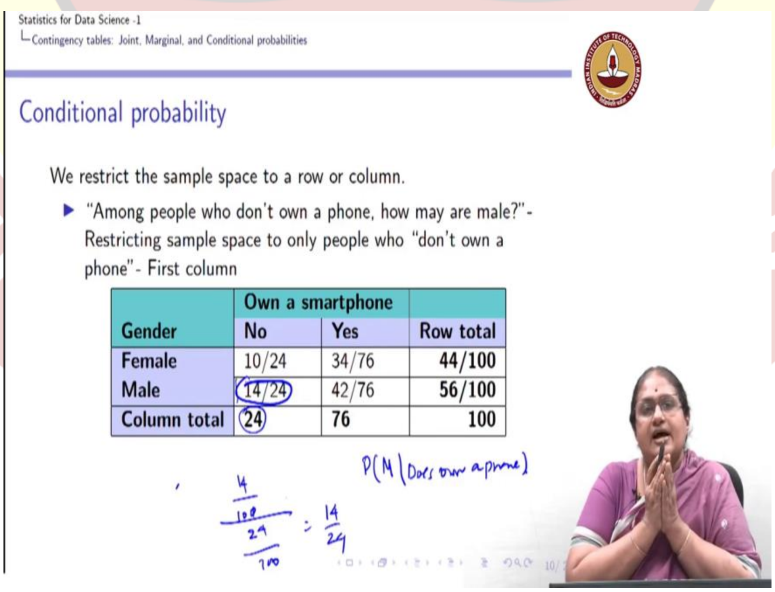


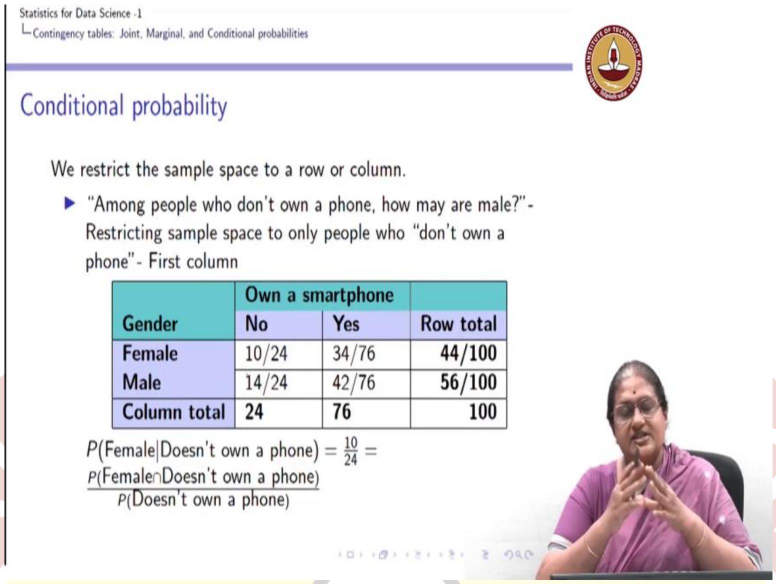
So, now if I restrict the sample space to only females, so I am not bothered about the sample space of perhaps the person is a male among these females, so earlier I computed $\frac{10}{100}$because I had the expanded sample space, now I am only finding what is the relative frequency, related to what? Related to the person being a female, I know 10 out of 44 do not own and 34 out of 44 own a smartphone.

Similarly, 10 out of 56 again 56 is the number of males, 14 out of 56 do not own a phone and 42 out of 56 own a phone. So, I am restricting the sample to females only or males only which is my first row. And I know that the chance that someone owns a phone is $\frac{34}{44}$. And the chance that a person, so if I am looking at probability of does not own a phone given a female this is we will explain this notation in a while, but if I am looking at a chance that I know a person is a female what is the probability that they do not own a phone, I see that it is $\frac{10}{44}$.

And we can see that this is nothing but 10 was $\frac{10}{100}$ was a probability that I had a female and does not own a phone, probability of a female is $\frac{10}{44}$ and I can see that this $\frac{10}{44}$ can be given by $\frac{10}{100}$ divided by $\frac{44}{100}$which is $\frac{10}{44}$, this is something which I observe.

(Refer Slide Time: 19:21)





Similarly, if I restrict among the people who do not own a phone, who are the people who do not own a phone, I have 24 people who do not own a phone, how many are males? So, I know 14 people in this, so jointly I know 14 people out of a 100 people do not 14 males out of 100 males do not own a phone. But given so if I am want restricted to the sample space of not owning a phone I know only 24 do not own a phone.

So, if I want to find out the conditional probability, so if I want to find out the conditional probability of what is the chance of being male given a person does not own a phone that would reduce to $\frac{14}{24}$ divided by my $\frac{14}{100}$ divided by $\frac{24}{100}$ which is $\frac{14}{24}$ and that is what I have here. In other words when we are talking about conditional probability we are actually restricting our sample space to include all those events conditioned on the event that has occurred.