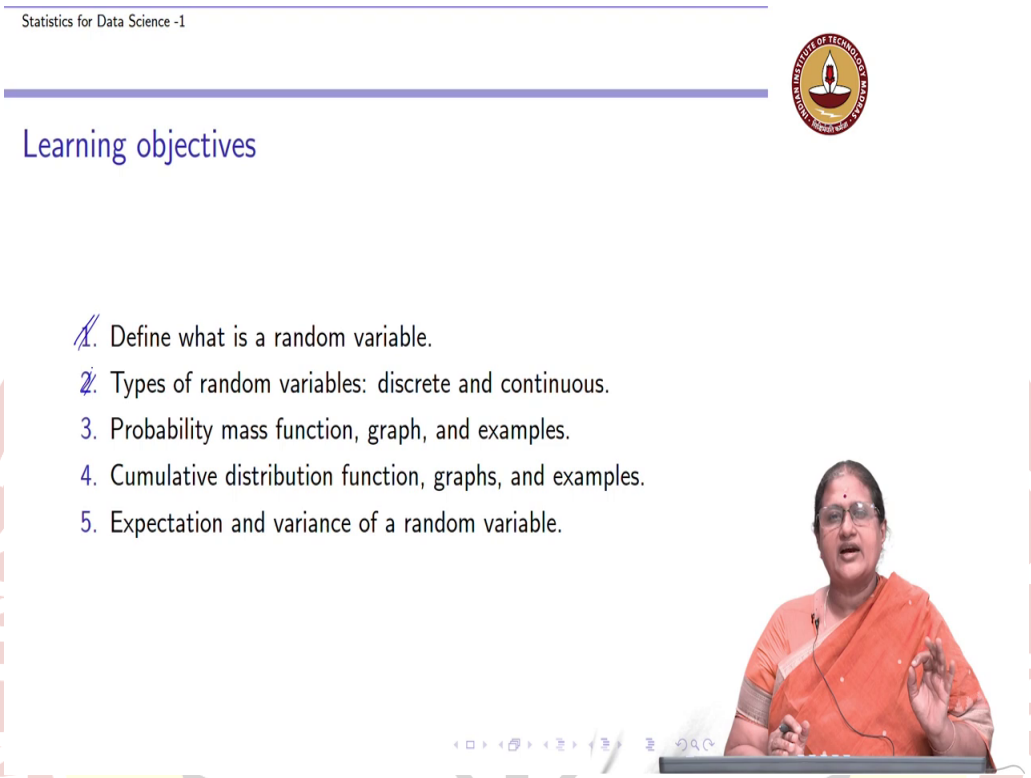
IIT Madras

ONLINE DEGREE

Statistics for Data Science 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Random variables - Discrete and continuous random variable

(Refer Slide Time: 00:14)



So, in the last lecture, we understood what was a, we defined what was a random variable. So, now we are going to look at, and towards the end of last lecture, we defined what was a discrete and continuous random variable.

(Refer Slide Time: 00:35)

Discrete and Continuous random variables

Definition

A random variable that can take on at most a countable number of possible values is said to be a discrete random variable.

- Thus, any random variable that can take on only a finite number or countably infinite number of different values is discrete.
- There also exist random variables whose set of possible values is uncountable.

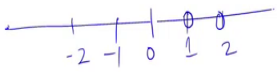
Definition

When outcomes for random event are numerical, but cannot be counted and are infinitely divisible, we have continuous random variables.

So, we will spend some time to understand more about what is a discrete and continuous random variable. So, these were the definitions which we check, which we gave in the last session, which was we define the discrete random variable as a random variable that can take on at most a countable number of possible values. Whereas we said a continuous random variable is it cannot be counted, or it is infinitely divisible. Now, let us understand what is a discrete random variable and a continuous random variable through a example.

(Refer Slide Time: 01:25)

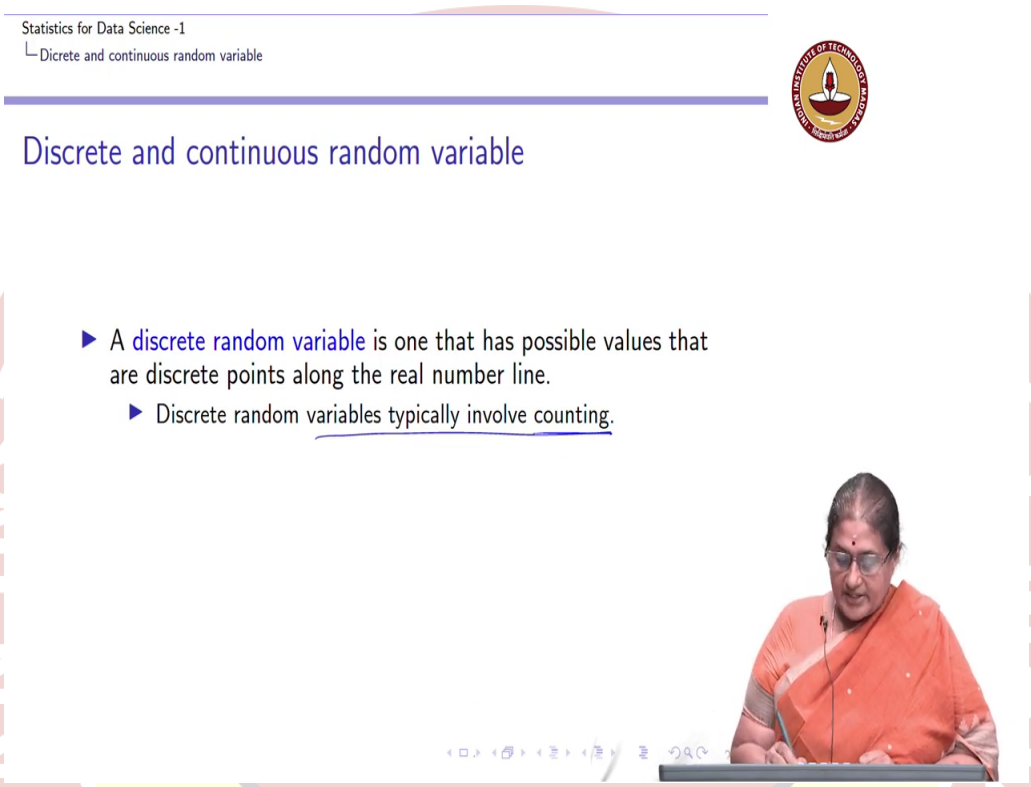
Discrete and continuous random variable



- A discrete random variable is one that has possible values that are discrete points along the real number line.

So, instead of going and understanding a definition, I can also say a discrete random variable is a random variable that can take possible values that are discrete points, I am again, emphasising that are discrete points. So, if I have my real number line here, so I have a 0 here, I have a 1 here, I have a 2 here, it could take a -1, it could take a -2. So, these are isolated discrete points along the real number line. I repeat, a discrete random variable is a random variable that can take possible values, along a real number line that are discrete points, the key idea is they are discrete points.

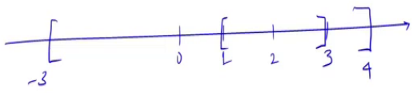
(Refer Slide Time: 02:22)



So, typically, when we talk about discrete random variables, you involve the operation of counting number of heads, you are counting.

(Refer Slide Time: 02:38)

Discrete and continuous random variable



- A discrete random variable is one that has possible values that are discrete points along the real number line.
  - Discrete random variables typically involve counting.
- A continuous random variable is one that has possible values that form an interval along the real number line.

Whereas a continuous random variable is the random variable that has possible values. So, if I again go back to my real number line, it will not take. So, I have it will not take isolated discrete values, but it would take values that might form an interval along the real number line, can it be this interval, it could be, it could be -3 to $+4$, it could be, but it takes the values in an interval.

A more rigorous definition of what is a continuous random variable would be offered in advanced courses. But at this point of time, you need to understand a discrete random variable, will take discrete points along a real number line, whereas a continuous random variable would take values in an interval along a real number.

(Refer Slide Time: 03:41)

Discrete and continuous random variable

- A discrete random variable is one that has possible values that are discrete points along the real number line.
  - Discrete random variables typically involve counting.
- A continuous random variable is one that has possible values that form an interval along the real number line.
  - Continuous random variables typically involve measuring.

So typically, a continuous random variable would involve a measurement or something whereas a discrete random variable involves a counting.

(Refer Slide Time: 03:55)

Example: Apartment complex

12 Apartments
4 Floors – 3 houses
1BR 2BR 3BR

Apartment complex data:

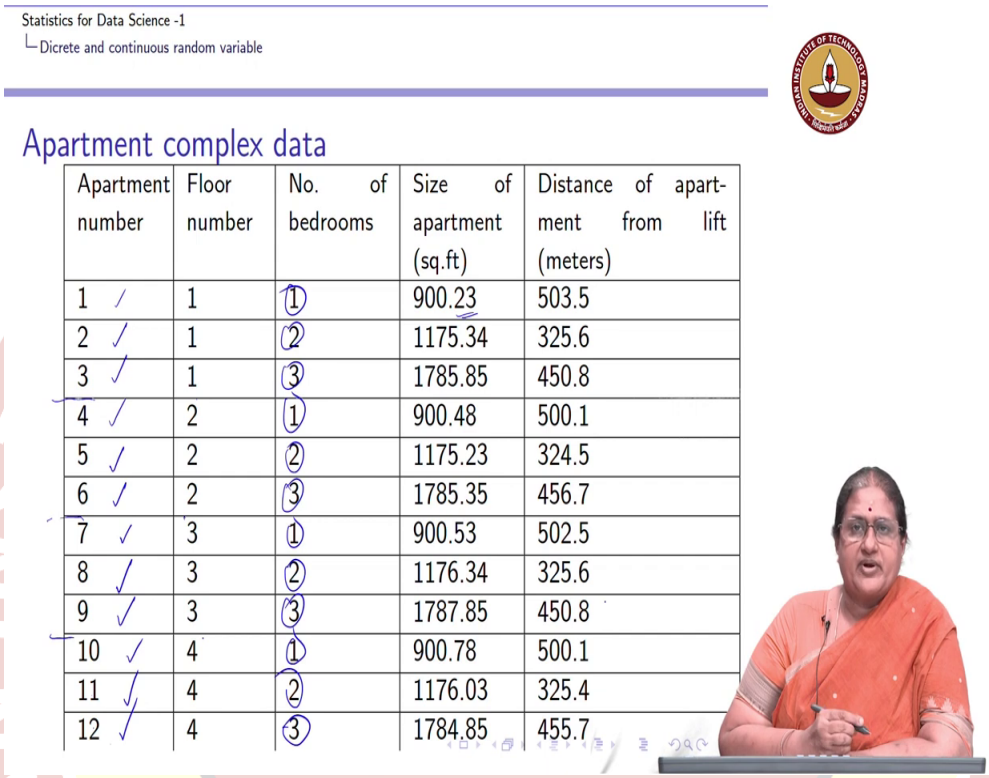
- There are four floors in the apartment complex.
- Each floor has three apartments: a one bedroom, a two bedroom and a three bedroom apartment.
- The data on the apartments is summarized in the table

So, now let us motivate and understand this through an example. Let us look at an apartment complex. So, what is this apartment complex? In apartment complex there are 4 floors in an apartment complex. Each floor has 3 apartments. So, if I am looking at this apartment complex,

each floor has a 3 apartments in so it has a 1-bedroom apartment, a 2 bedroom apartment and a 3 bedroom apartment.

So, there are totally 12 apartments. There are 4 floors, there are 3 houses per floor. And in these 3 houses or 3 apartments. First is a 1 bedroom I have a 2 bedroom and a 3 bedroom apartment. This is how this apartment complex has been structured.

(Refer Slide Time: 04:51)



Apartment number	Floor number	No. of bedrooms	Size of apartment (sq.ft)	Distance of apartment from lift (meters)
1	1	1	900.23	503.5
2	1	2	1175.34	325.6
3	1	3	1785.85	450.8
4	2	1	900.48	500.1
5	2	2	1175.23	324.5
6	2	3	1785.35	456.7
7	3	1	900.53	502.5
8	3	2	1176.34	325.6
9	3	3	1787.85	450.8
10	4	1	900.78	500.1
11	4	2	1176.03	325.4
12	4	3	1784.85	455.7

So, let us look at the data which is summarised in the following table. So, if you look at the data, I have 12 apartments that is given by the cases 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 these are my 12 apartments. So, you see the first 3 apartments are in the first floor 4, 5, 6 are in second floor 7, 8, 9 are in third floor and 10, 11, 12 are in fourth floor. When you look at the number of bedrooms, the first apartment, fourth apartment, seventh and tenth have 1 bedroom each.

The second, fifth, eighth and eleventh have 2 bedrooms each. Whereas the third, 6, 9 and 12 have 3 bedrooms each. So, this is the data I have collected. Now, even though typically you might expect the sizes of apartments to be the same, but there is some error and I am allowing I am correct, computing the sizes of the apartment, correct to 2 decimal points in square feet.

So, the first apartment is 900.23 $sq.feet.$, the fourth is 900.48, the seventh is 900.53 $sq.feet$. So, all the measurement of the sizes of the apartment in $sq.feet$ for each of the apartment is given. In addition, we also are capturing how many metres from the left is each one of the apartment. This is just fictious data, this is just data from an apartment complex. And we are capturing it in this following way. Now, I can look at this as a random experiment.

(Refer Slide Time: 06:52)

Apartment complex

- Random experiment: Randomly selecting an apartment in an apartment complex of 12 apartments.
- $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

Now, what is the random experiment in this case? Now, let me define the random experiment is I am randomly selecting an apartment from these 12 apartments, I am selecting one apartment from these 12 apartments. So, the sample space of my random experiment is any one of the 12 apartments. So, the outcome is I could either have selected the first or the second or the third, so forth up to the 12th.

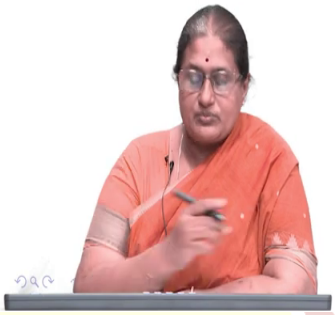
So, this constitutes the sample space of my experiment. The outcomes is one, one represents the first apartment, two represents second, third and so forth, 12 represents the twelfth apartment. Now, recall when we talked about a random variable, we said that a random variable associates some quantity with every outcome of the sample space.

(Refer Slide Time: 08:05)

Statistics for Data Science -1
└Dicrete and continuous random variable

Questions

1. Let the random variable be number of bedrooms, what are the possible values that might be observed?



Statistics for Data Science -1
└Dicrete and continuous random variable

Apartment complex data

Apartment number	Floor number	No. of bedrooms	Size of apartment (sq.ft)	Distance of apartment from lift (meters)
1	1	1	900.23	503.5
2	1	2	1175.34	325.6
3	1	3	1785.85	450.8
4	2	1	900.48	500.1
5	2	2	1175.23	324.5
6	2	3	1785.35	456.7
7	3	1	900.53	502.5
8	3	2	1176.34	325.6
9	3	3	1787.85	450.8
10	4	1	900.78	500.1
11	4	2	1176.03	325.4
12	4	3	1784.85	455.7

Now, suppose I am asking the following questions. Let a random variable be the number of bedrooms. So, let us go back to our data, I can see that the random variable that is associated with my first outcome is 1, the random variable that is associated with the second is 2, third is 3, fourth is again 1, fifth is 2.

(Refer Slide Time: 08:38)

Apartment complex

- Random experiment: Randomly selecting an apartment in an apartment complex of 12 apartments.
- $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

1 2 3 1 2 3 1 2 3 1 2 3

X = number of bedrooms

X = 1, 2, 3

So, if I am looking at random variables associated with each outcome, I have a 1, 2, 3; a 1, 2, 3; a 1, 2, 3; a 1, 2, 3. These are the random variables. Where what is the random variable? The random variable X is the number of bedrooms in the apartment. And I see that this random variable, every outcome of my sample space has a random variable or a quantity associated with it. What are the possible values this random variable can take now? It can take the possible values the random variable can take are 1, 2, and 3.

(Refer Slide Time: 09:30)

Questions

1. Let the random variable be number of bedrooms, what are the possible values that might be observed?
   Answer: 1,2,3
2. Let the random variable be floor number of the apartment. What are the possible values that might be observed?

Apartment complex data

Apartment number	Floor number	No. of bedrooms	Size of apartment (sq.ft)	Distance of apartment from lift (meters)
1	1	1	900.23	503.5
2	1	2	1175.34	325.6
3	1	3	1785.85	450.8
4	2	1	900.48	500.1
5	2	2	1175.23	324.5
6	2	3	1785.35	456.7
7	3	1	900.53	502.5
8	3	2	1176.34	325.6
9	3	3	1787.85	450.8
10	4	1	900.78	500.1
11	4	2	1176.03	325.4
12	4	3	1784.85	455.7

So, you can see that the possible values the random variable can take are 1, 2, and 3. Now, let us go to the second question. The second question says that let the random variable be the floor number of the apartment. Now, again let us go back to our data. The first for with the first apartment. I know that the floor number associated with my first number is 1 with a second apartment is 1, third is 1.

(Refer Slide Time: 10:07)

Apartment complex

- Random experiment: Randomly selecting an apartment in an apartment complex of 12 apartments.
- $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

1 2 3 1 2 3 1 2 3 1 2 3

1 1 1 2 2 2 3 3 3 4 4 4

X = number of bedrooms

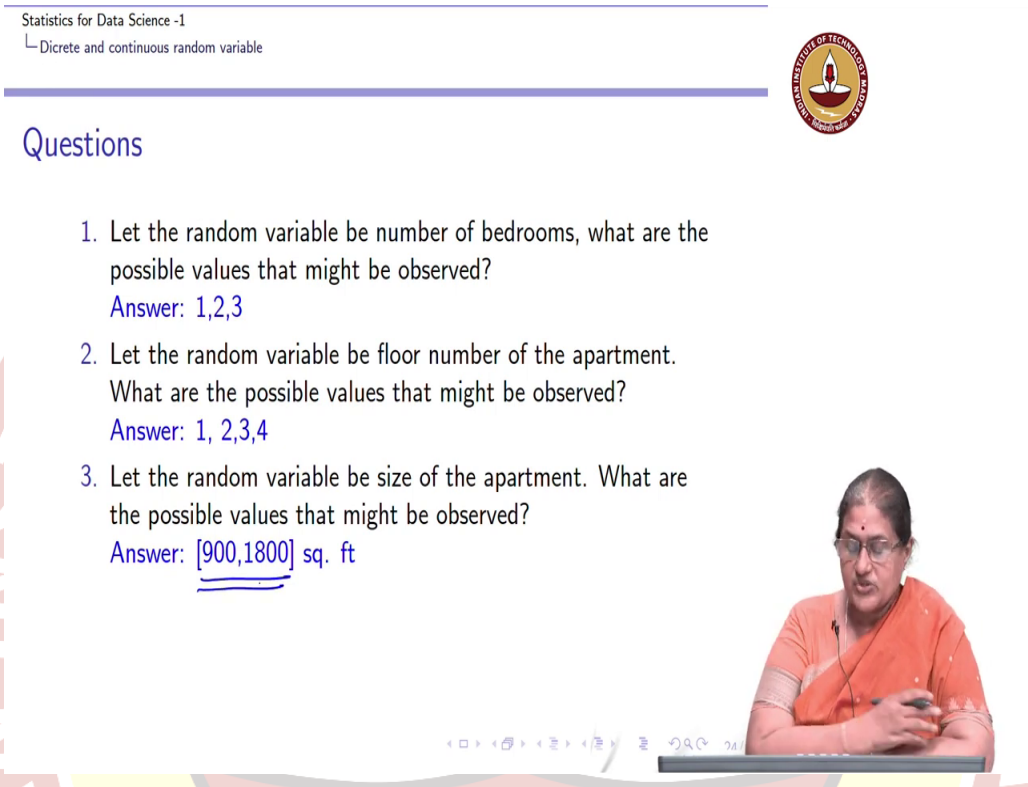
X = 1, 2, 3

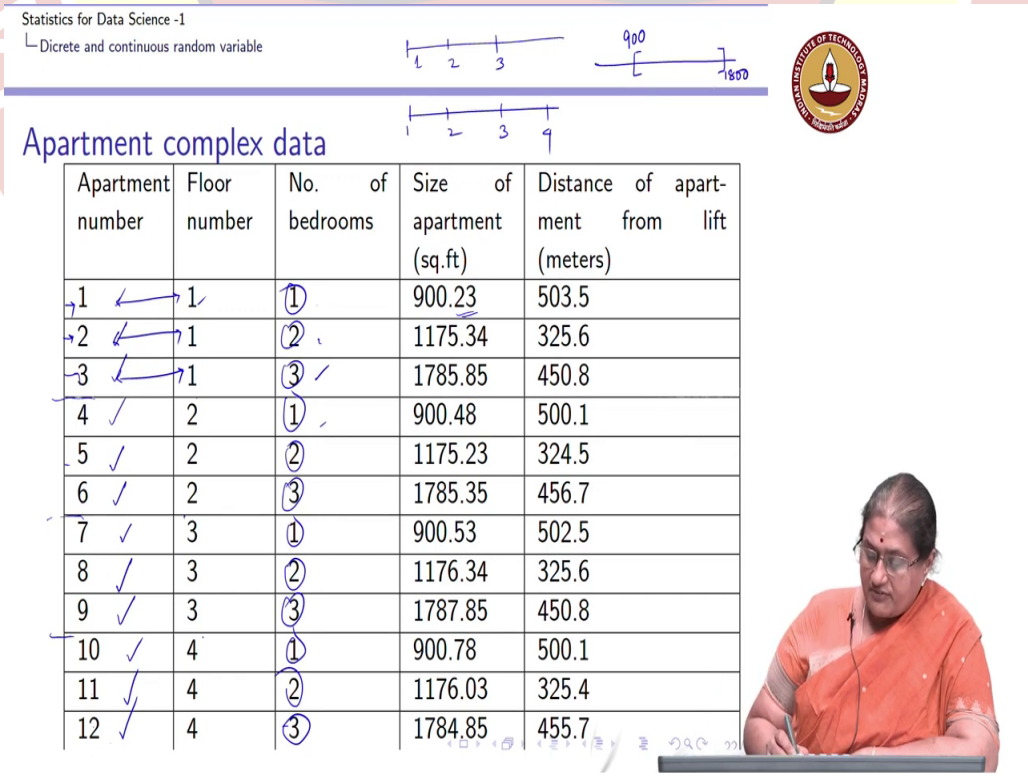
Y = number of floor

Y = 1, 2, 3, 4

So, in terms of my sample space, when I come back here, and I am defining say Y to be the number of the floor or the floor number with 1 I am associating variable 1 with 2 again 1 with 3 again 1, 4 it is a 2, 5 it is a 2, 6 is a 2, 7 is a 3, 8 is a 3, 9 is a 3, 10 is a 4, 11 is a 4, and 12 is a 4. So, the possible values the random variable, which is the floor number, every outcome is associated with a quantity, the possible values, this random variable can take are 1, 2, 3 and 4.

(Refer Slide Time: 11:02)





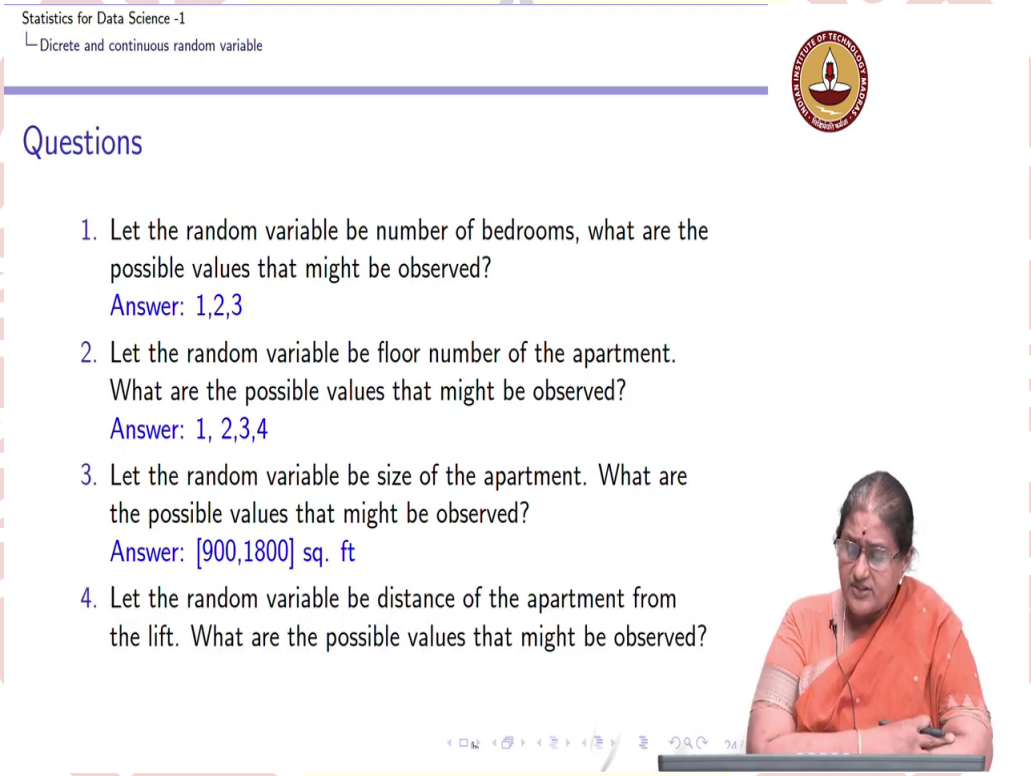
Apartment number	Floor number	No. of bedrooms	Size of apartment (sq.ft)	Distance of apartment from lift (meters)
1	1	1	900.23	503.5
2	1	2	1175.34	325.6
3	1	3	1785.85	450.8
4	2	1	900.48	500.1
5	2	2	1175.23	324.5
6	2	3	1785.35	456.7
7	3	1	900.53	502.5
8	3	2	1176.34	325.6
9	3	3	1787.85	450.8
10	4	1	900.78	500.1
11	4	2	1176.03	325.4
12	4	3	1784.85	455.7

So, now you can see that the possible values this random variable can take are 1, 2, 3 and 4. Now, let me define the random variable to be the size of the apartment. Earlier, we looked at number of bedrooms, then we looked at the floor number, now I am looking at the size of the apartment. Again, let us go back.

Now, when we go back to the data and look at the size of the apartment, we do not see, with that it can be mapped to discrete points on the real number line. Whereas here I could have, so in the first case, I could map it to 1, 2, and 3, these are discrete isolated points. In the second case, it is 1, 2, 3, 4.

Whereas in the third case, when I talk about a real number line, I see that I can have that the data can take any value between 900 square feet to say 1780 or 1800 square feet. So, in a sense, I can give the values that possible values, the size of the apartment can take, the possible value, the size can be in an interval, and this interval is 900 to 1800 square feet.

(Refer Slide Time: 12:35)



Apartment complex data

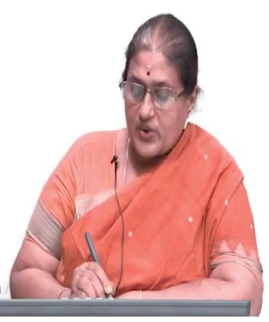
Apartment number	Floor number	No. of bedrooms	Size of apartment (sq.ft)	Distance of apartment from lift (meters)
1	1	1	900.23	503.5
2	1	2	1175.34	325.6
3	1	3	1785.85	450.8
4	2	1	900.48	500.1
5	2	2	1175.23	324.5
6	2	3	1785.35	456.7
7	3	1	900.53	502.5
8	3	2	1176.34	325.6
9	3	3	1787.85	450.8
10	4	1	900.78	500.1
11	4	2	1176.03	325.4
12	4	3	1784.85	455.7

Now, let us define the fourth random variable to be the distance of the apartment from the left. Again, we go back to our data, when we go back to our data, we see that again here there are not discrete points, but it is again a continuum. And when you look at a continuum, I can again write that the distance of the apartment from the feet.

(Refer Slide Time: 13:01)

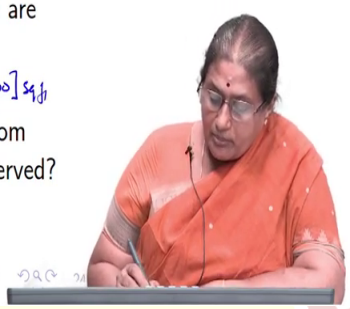
Questions

1. Let the random variable be number of bedrooms, what are the possible values that might be observed?
   Answer: 1,2,3
2. Let the random variable be floor number of the apartment. What are the possible values that might be observed?
   Answer: 1, 2,3,4
3. Let the random variable be size of the apartment. What are the possible values that might be observed?
   Answer: [900,1800] sq. ft



Questions

1. Let the random variable be number of bedrooms, what are the possible values that might be observed?
   Answer: 1,2,3 $X = 1, 2, 3$
2. Let the random variable be floor number of the apartment. What are the possible values that might be observed?
   Answer: 1, 2,3,4 $Y = 1, 2, 3, 4$
3. Let the random variable be size of the apartment. What are the possible values that might be observed?
   Answer: [900,1800] sq. ft $Z \in [900, 1800]$ sq ft
4. Let the random variable be distance of the apartment from the lift. What are the possible values that might be observed?
   Answer: [324,505] meters $Z_1 \in [324, 505]$ metres



So here, if I define this random variable in the earlier case, the size to be some Z, I know that this random variable Z would belong to 900 to 1800 square feet. Now, I could call this say Z_1 and I know that this belongs to 324 to 505 metres. So, we have seen the example of a random variable here, which takes values 1, 2, 3 here it takes the values 1, 2, 3, 4, here it belongs to a interval and in the last case, it also belongs to a interval.

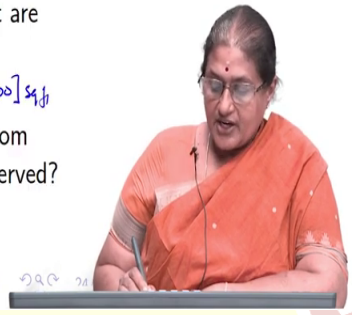
(Refer Slide Time: 14:05)

Discrete and continuous random variable

- A discrete random variable is one that has possible values that are discrete points along the real number line.
  - Discrete random variables typically involve counting.
- A continuous random variable is one that has possible values that form an interval along the real number line.
  - Continuous random variables typically involve measuring.

Questions

1. Let the random variable be number of bedrooms, what are the possible values that might be observed?
   Answer: 1,2,3 $X = 1, 2, 3$
2. Let the random variable be floor number of the apartment. What are the possible values that might be observed?
   Answer: 1, 2,3,4 $Y = 1, 2, 3, 4$
3. Let the random variable be size of the apartment. What are the possible values that might be observed?
   Answer: [900,1800] sq. ft $Z \in [900, 1800]$ sq ft
4. Let the random variable be distance of the apartment from the lift. What are the possible values that might be observed?
   Answer: [324,505] meters $Z_1 \in [324, 505]$ meters



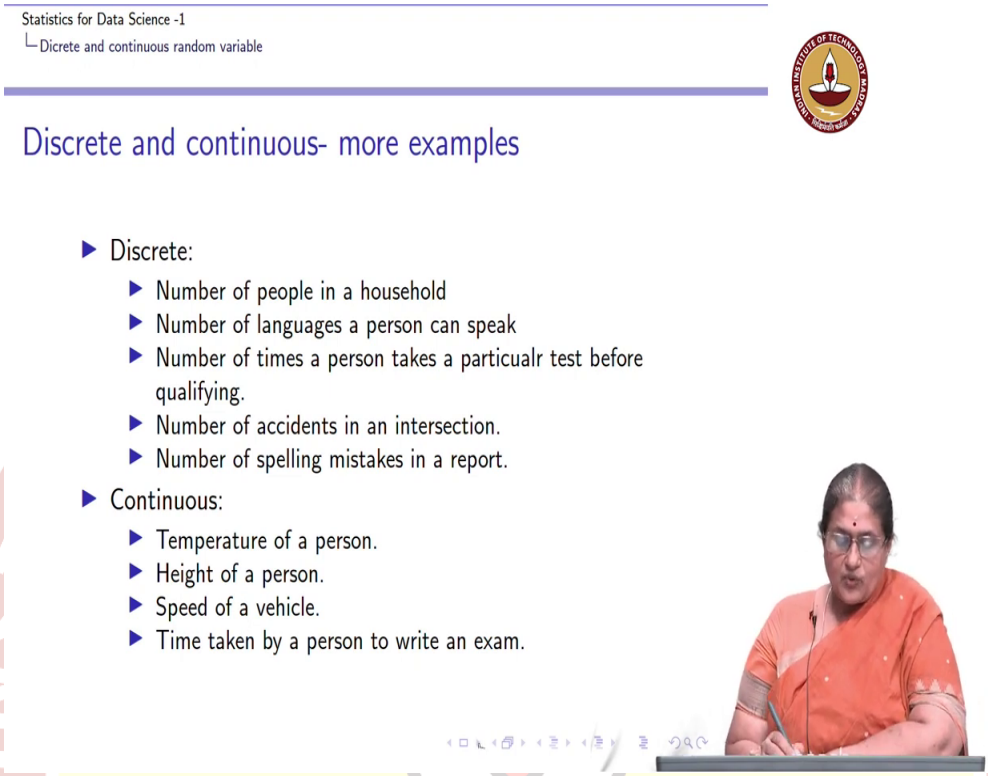
Discrete versus continuous

- Which variables are discrete random variables?
  - Number of bedrooms, floor number.
- Which variables are continuous random variables?
  - Size, distance to the lift.

So, going back to the definition, which we proposed is of definition the understanding which we proposed for a random variable, the discrete and continuous, we said a discrete random variable typically would take discrete points along the real number line, and here it forms an interval, I can now classify the 2 random, the 4 random variables which we have given as examples here as the first 2 that is the discrete random variables in the data are nothing but the number of bedrooms and floor number, whereas the continuous random variables in my

example data set, the apartment data set are size of the apartment, and distance of the apartment to the left.

(Refer Slide Time: 14:48)



So, let us look at a few more examples where we naturally come across discrete random variables. Again, remember discrete random variables involves counting. So, whenever you are having instances where for example, you count the number of people in a household, a number of languages a person can speak, number of times a person takes a particular test before qualifying or before passing, number of accidents that happen in an intersection, number of spelling mistakes in a report all these count, you can see that everything represents the operation of counting, you can discuss or you can think of this as a discrete random variable.

Whereas the instances where I have, I am computing or measuring the temperature of a person, measuring the height of a person, measuring the speed of a vehicle, or measuring the time taken by a person to write an exam, time taken by a person to reach office. All these instances which involve the measure of a quantity, I can think of the variable or the random variable that is involved as a continuous random variable.

(Refer Slide Time: 16:16)

Section summary

- Definitions of Discrete random variable versus continuous random variable
- Identify discrete and continuous random variables.

So, in summary, what we have learnt in this section was, we first defined formally what was a discrete random variable and a continuous random variable, a discrete random variable takes countable values, and it can be mapped onto discrete points on a real number line. Whereas, a continuous variable, it actually can be mapped on to an interval on a real number line. For now, we are going to focus on discrete random variables.