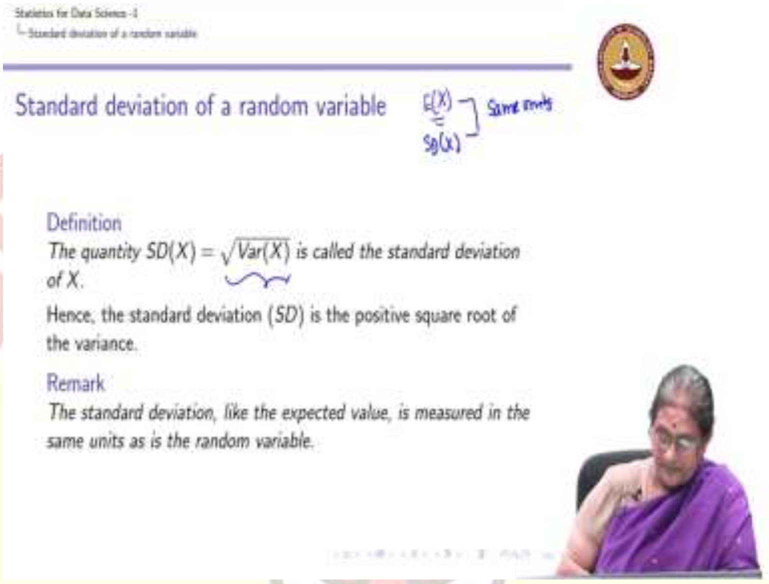
IIT Madras
ONLINE DEGREE

Statistics for Data Science – 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture 9.6
Standard deviation of a random variable

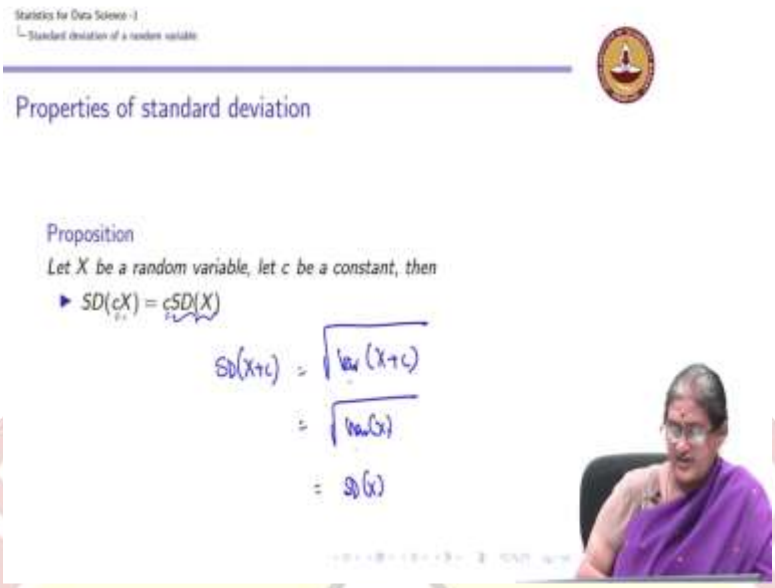
(Refer Slide Time: 0:14)

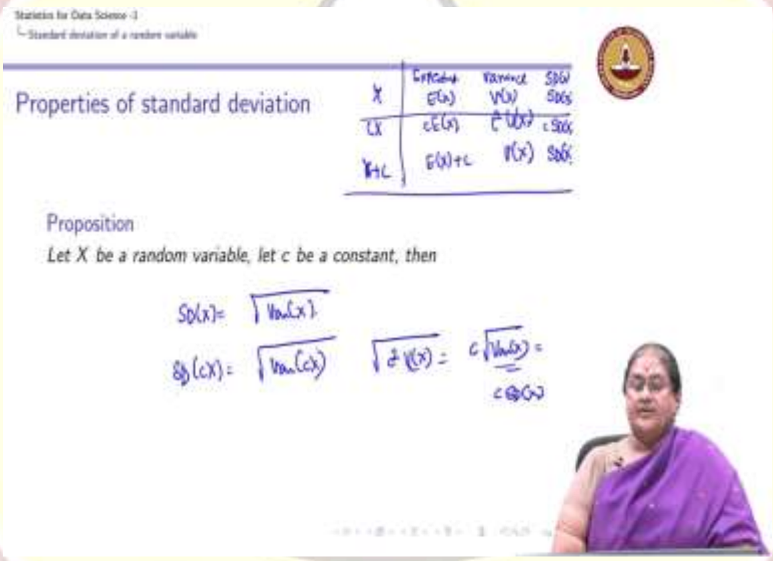


Now, we introduce just a descriptive statistics case where we talked about the variants and we said that when we have to report measure in the original units, then afterwards another measure is very useful and we introduce the notion of a standard deviation in that case. Similarly, we are going to introduce the measure of a standard deviation of a random variable.

So, I can define the standard deviation of a random variable as a square root of the variance of the random variable and this quantity is called the standard deviation of X. In other words, it is just the positive square root of the variance and like the expected value, now if I have a random variable, the expected value of the random variable would take a units and since this is the variance of square root of the variance of X, the standard deviation of the random variable and the expected value have the same units or they are measured in same units.

(Refer Slide Time: 1:25)





Now, let us again look at properties of the standard deviation. So again, what we are interested in knowing is I already have the following, if X is a random variable and c is a constant, $E[cX] = cE[X]$, and the $Var(cX) = c^2Var(X)$. Now, $E[X + c] = E[X] + c$ and $Var(c + X) = Var(X)$. This is something which we have already seen.

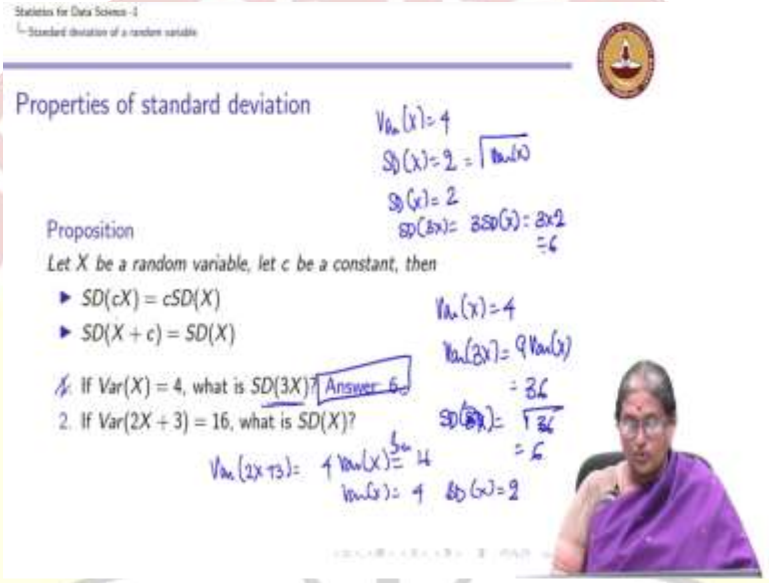
So, now let us see what happens to the standard deviation of X, so I know this standard deviation of X is given by SD(X). So, now I know that $SD(X) = \sqrt{Var(X)}$. So, what would $SD(cX)$ be? $SD(cX)$ is going to be $\sqrt{Var(cX)} = \sqrt{c^2Var(X)} = c\sqrt{Var(X)} = c\ SD(X)$. So, SD(cX) = c SD(X). That is my first property.

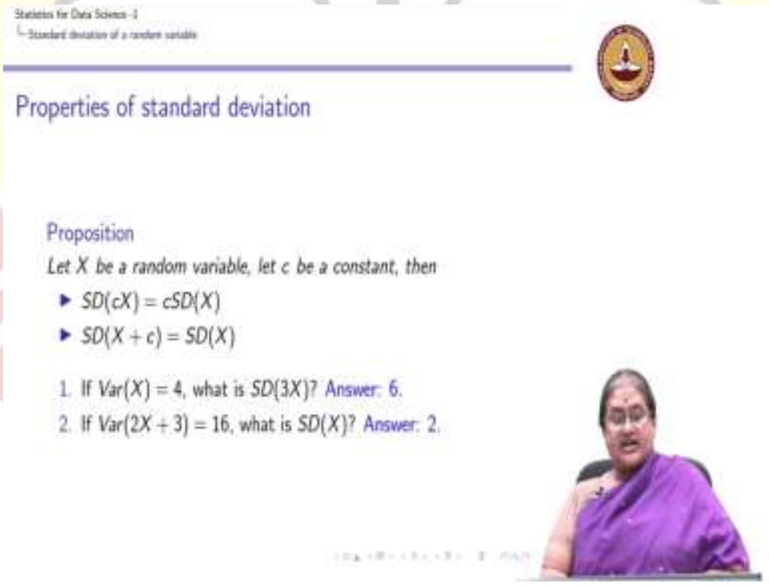
That is standard deviation of a constant multiple of a random variable is constant multiple $\times$ $SD(X)$. So, I can write here this is $c \times SD(X)$.

The next thing which we will need to see is what happens to $SD(X+c)$ this is going to be $\sqrt{Var(X+c)}$ which was same as $\sqrt{Var(X)}$ because $Var(X+c) = Var(X)$. So, I have $SD(X+c) = \sqrt{Var(X+c)} = \sqrt{Var(X)} = SD(X)$.

Now, this is very useful, these properties are very useful for me to compute certain formula. So, now let us look at an example where I am interested in knowing the following.

(Refer Slide Time: 4:01)

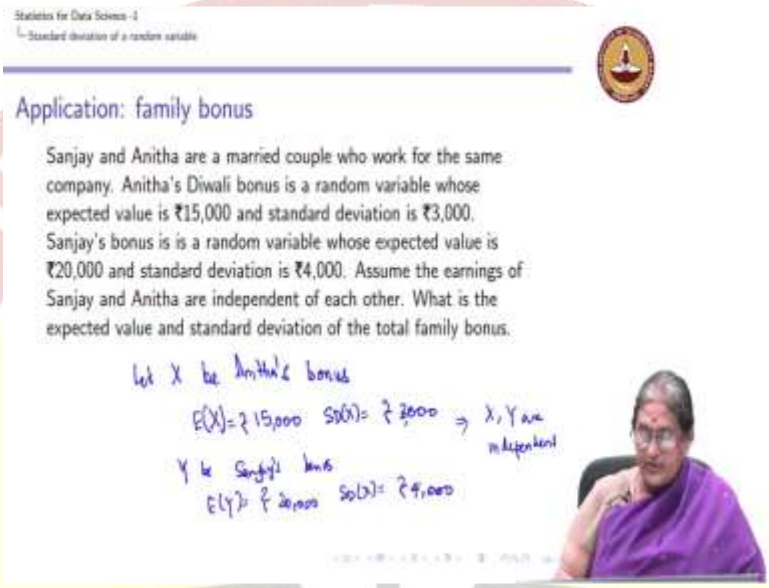




So, if $Var(X) = 4$, what is $SD(3X)$? Now, if $Var(X) = 4$, I know SD(X) = 2 which is the same as $\sqrt{Var(X)}$, now if SD(X) = 2, then I apply $SD(cX)$, c = 3 here, is $3 \times SD(X)$ which gives me $3 \times 2 = 6$ and hence the answer is 6.

Now, there is another way we can do it. If $Var(X) = 4,$ I know $Var(3X) = 9Var(X)$ which is equal to 36, hence $SD(3X) = \sqrt{36} = 6$. Both of them give you the same answer. Now, let us look at the second question $Var(2X + 3)$. I know $Var(2X + 3)$ applying my formula is $4 \times Var(X)$. Now, this is given to be 16 which gives me the $Var(X) = 4$. Hence, $SD(X) = 2$. So, $SD(X) = 2$ if $Var(2X + 3) = 16$.

(Refer Slide Time: 5:42)

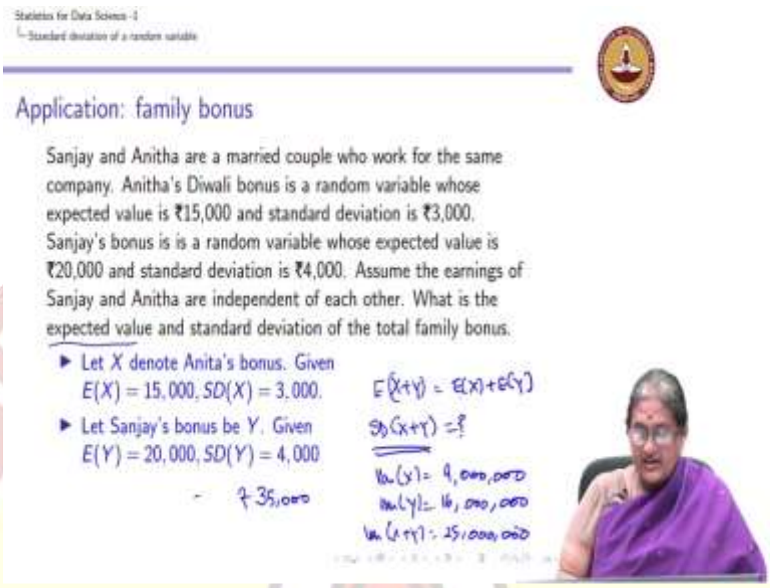


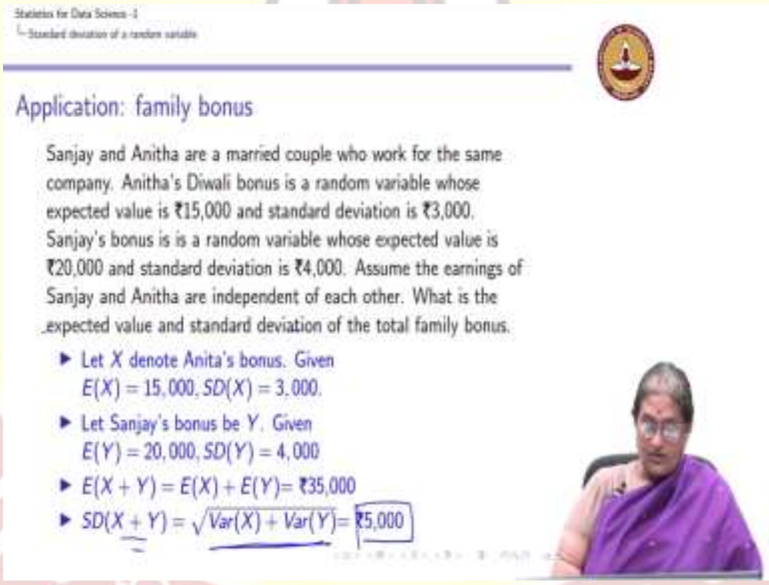
Now, let us look at a few applications of standard deviation. Remember, standard deviation takes the same units as that of the expected value. Recall we talked about Sanjay and Anitha who were two people. Now assume that they are a married couple that is, I define a family, they are a married couple, this is how I am defining it. They work for the same company. Now, I am giving Anitha's bonus as a random variable. I know the expected value and standard deviation of Anitha's bonus.

So, let X be Anitha's bonus, it is a random variable and bonus, X is a random variable, I given that $E[X] = ₹15{,}000$, it takes the values and standard deviation of X also takes the same value, it is ₹3,000. Now, let Y be Sanjay's bonus. What is given to us is $E[Y] = ₹20{,}000$ and standard deviation, again it takes the same units, ₹4,000. Again I am assuming earnings of Sanjay and Anitha are independent of each other, this implies that X and Y are independent random variables.

That the earnings of Sanjay and Anitha are independent of each other, so if I have the bonus they are independent of each other, hence X and Y are independent random variables. This is an assumption. I want to know what is the expected value and standard deviation of the total family bonus where family I am just assuming Sanjay and Anitha form a family.

(Refer Slide Time: 7:43)

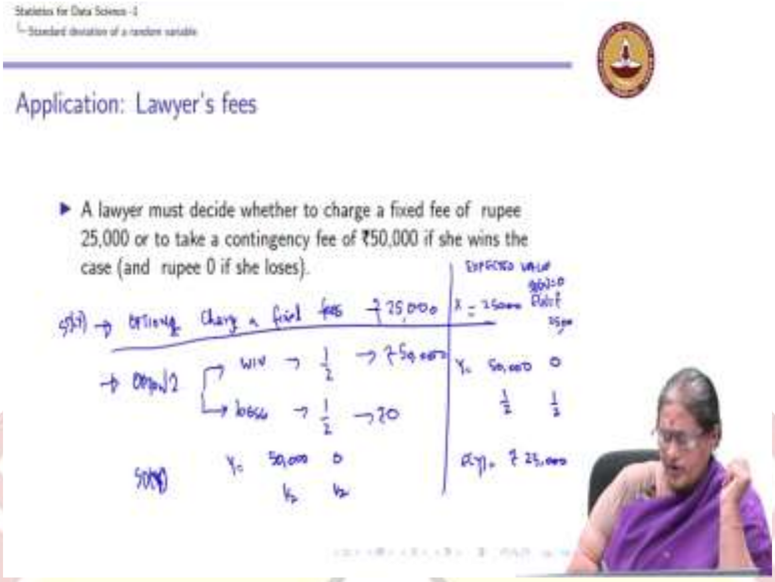




So, my total family bonus is going to be X + Y is my total family bonus, I want to know what is expectation of $E[X+Y]$ and I want to know what is SD(X+Y). X and Y are any random variable, we know $E[X+Y] = E[X] + E[Y]$ that is expectation of sum is sum of expectation. Hence, expectation expected value of the total family bonus is going to be 15000 + 20000 which is going to give me a ₹35, 000.

Now to compute the $SD(X+Y)$. I know the $Var(X) = 9{,}000{,}000, Var(Y) = 16{,}000{,}000$, $Var(X+Y) = 25{,}000{,}000$ which will give me the $SD(X+Y)$ = $\sqrt{Var(X+Y)}$ = ₹5,000. Hence, I have computed the expected value and standard deviation of the total family income.

(Refer Slide Time: 9:15)



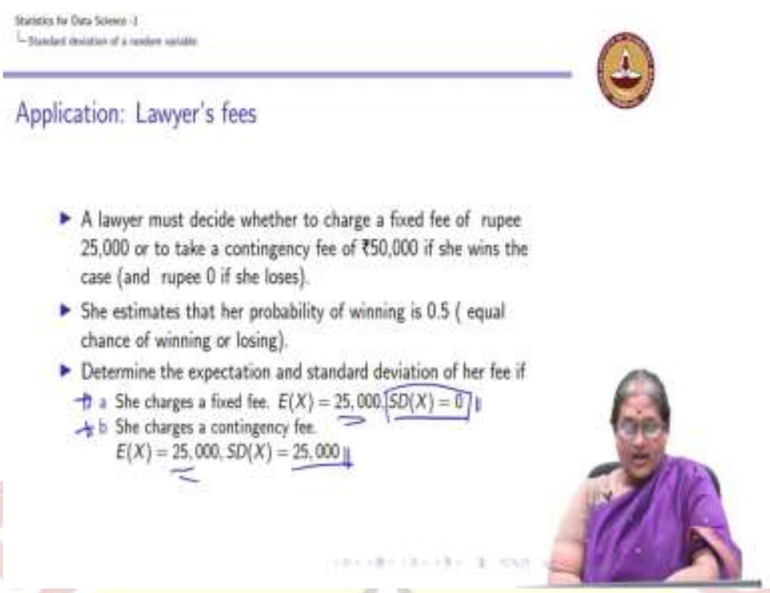
Now let us look at another application which is Lawyer's fees. Suppose I have a lawyer who has to decide about two options. The first option of the lawyer is charge a fixed fees whether I win or lose, if I am the lawyer, I am going to just charge a fixed fee and what is my fixed fee? It is ₹25,000. And the second option, this is option 1, the second option is I am going to gamble.

So, here I say that I can either win or lose, I assume I can win and lose with equal probability. I say that you pay me but you pay me 50000 if I win, you do not have to pay me anything if I lose. So, the question is, is option 1 or option 2 favourable to me if I say I am lawyer. So, let us look at the expected value in both the cases. Now, here if X is my earning when I am charging a fixed fee, $P(X = 25000) = 1$, hence $E[X] =$ ₹25,000

Now, $P(Y = 50000) = \frac{1}{2}$ and $P(Y = 0) = \frac{1}{2}$. $E[Y] =$ ₹25,000. So, if only expected value was of concern to me, then both the options return the same expected value namely, 25000 rupees. But there is a risk. Whereas option 1 always gives me 25000 surely whether I win or lose, here option 2 gives me a high return if I win and I could lose money, so I lose money that is I do not get anything at all if I lose. So, there is a risk of me not getting anything at all. The expected value does not reflect this risk.

So, how can I compute that? Now let us look at the standard deviation of the X values, $SD(X)$ and $SD(Y)$. X takes only 1 values, hence the $SD(X) = 0$, whereas when I look at $SD(Y)$, Y takes the value 50000 and 0 with equal probabilities.

(Refer Slide Time: 12:22)

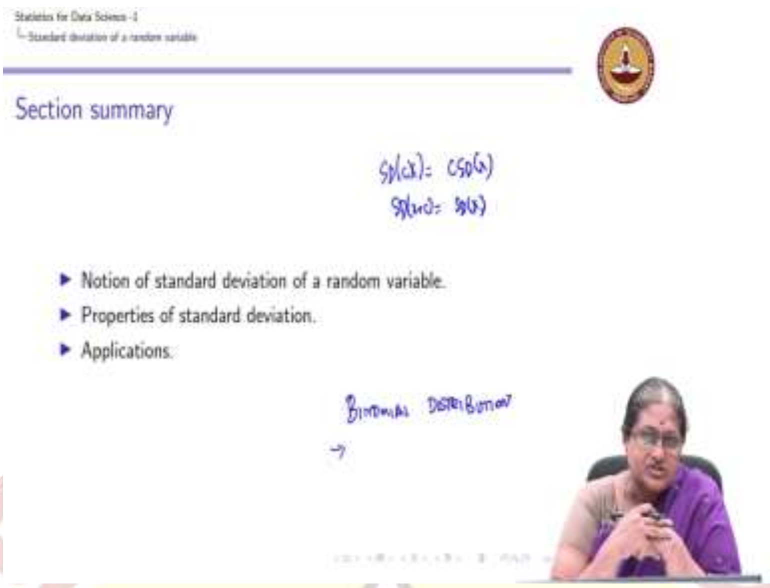


I can check that the $SD(Y) = 25{,}000$. So, you can see that whereas here if in a sense I am measuring the risk of me losing money with standard deviation, then I can see that this option 2 has a much higher risk than option 1. Though in expectation both of them return the same value.

Hence, the notion of standard deviation plays an important role in applications of this kind where you want to measure in addition to expected value or the average, you want to have a measure of dispersion, dispersion can measure volatility or spread or risk in some way, I am loosely defining this terms but I just want you to give an idea of where you could use measures of standard deviation and he that can help you or aid you in decision making.

So, if you are a person from whom this risk is important, then you might just go with option a, but for if you are a person who feels that it is okay to take this level of risk, then option b is a better option for you because you have a chance of getting 25000 extra which is you get 50000 rupees.

(Refer Slide Time: 14:03)



So, in summary what we have learned so far is the notion of a standard deviation of a random variable. We looked at again properties of the standard deviation namely we saw that $SD(cX) = cSD(X)$ and $SD(X + c) = SD(X)$ and we looked at how to apply concepts of standard deviation to answer certain real time applications.

With this our discussion on general discrete random variable is complete. The next thing which you are going to look at is a very special and specific random variable which is called the binomial distribution. So, again we will start by looking at where our instances were a binomial distribution make sense, we look at the expectation and the variance of a binomial random variable and we look at a lot of applications wherein the random variable can be modelled as a binomial random variable and how we can use that information to reach to conclusions. Thank you.