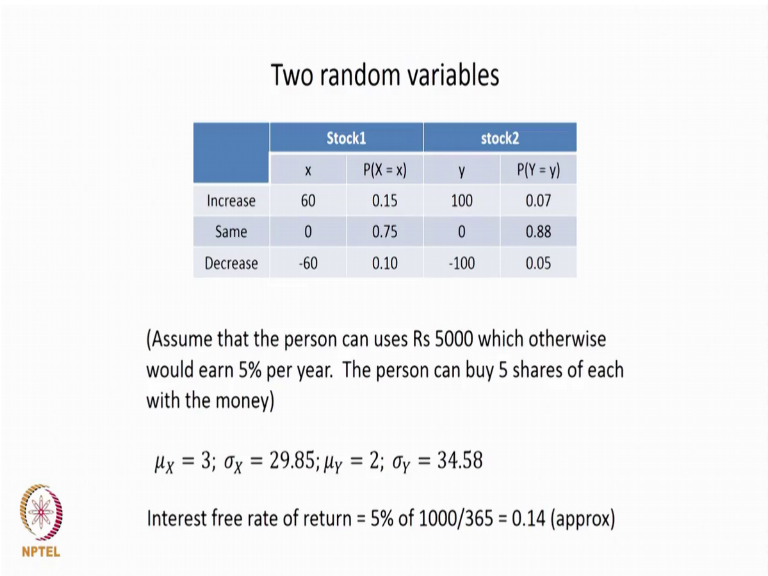
Introduction to Probability and Statistics
Prof. G. Srinivasan
Department of Management Studies
Indian Institute of Technology, Madras

Lecture – 18
Association Between Random Variables

In this lecture, we study the Association between Random Variables. In the two previous lectures, we introduced the Concept of a Random Variable and we also showed computations to find out the expected value of the random variable, which is the measure of central tendency and we also looked at equations to compute the variance and the standard deviation. Using which we compared two different random variables and use these to make decisions.

Now, we look at association between random variables just as we saw association in statistics and we looked at covariance and correlation. We would also see these measures of association and show how to compute some of these measures and make decisions based on the computation.

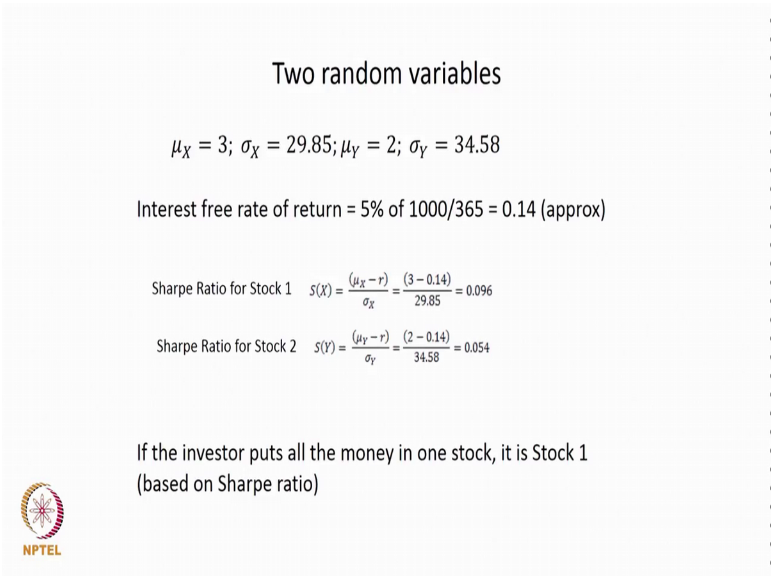
(Refer Slide Time: 01:20)



So, let us look at two random variables which are called x and y. So, x could be stock 1 and y should be stock 2. So, let us assume that the random variable x takes three values which means it can increase by 60 with the probability of 0.15. It remains the same with probability of 0.75 and decrease by 60 with the probability of 0.1. The stock second stock called y can also take three values 100, 0 and minus 100 with the associated probabilities given here. Now, mu X is equal to 3, expected value is 60 into 0.5 plus 0 into 0.75 minus 60 into 0.1. So, 60 into 0.15 is 9, plus 0 into 0.75, 0, minus 60 into 0.16. So, expected value is 3. Standard deviation based on our computation we have not shown the detail of the computation, but we have seen how to compute it in earlier lectures sigma x turns out to be 29.85.

Now, we look at stock 2. Expected value is 100 into 0.07 which is 7, 0 in to 0.88 0 minus 100 into 0.05 is minus 5. So, expected value is 2 and standard deviation turns out to be 34.58. Now, if we assume that the interest free rate of return is 5 percent of 1000 by 365 etcetera we look per day and all these are assumed to be a random variable which represents the change between two consecutive days, then interest free rate of return will be 5 percent of 1000 by 365 which is now taken as 0.14 for our computation.

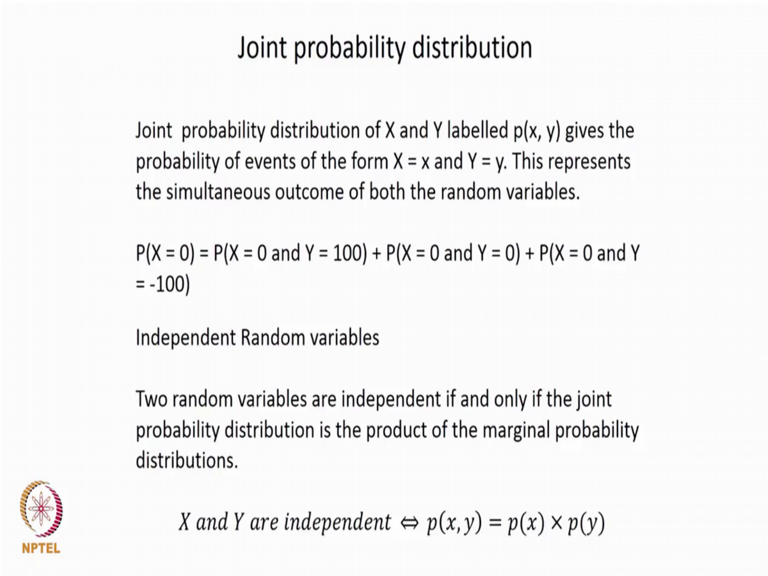
(Refer Slide Time: 03:28)



So, we compute what is called a Sharpe ratio for a stock 1 which is mu X minus r by sigma X which is 0.096, Sharpe ratio for stock 2, which is mu Y minus r by sigma Y which is 0.054. So, if the investor wishes to put all the money in one of the two stocks then the person will choose stock 1 which has a larger Sharpe ratio.

Now, we try to find out the association and see whether it is advantageous to put part of the money in stock 1 and part of the money in stock 2.

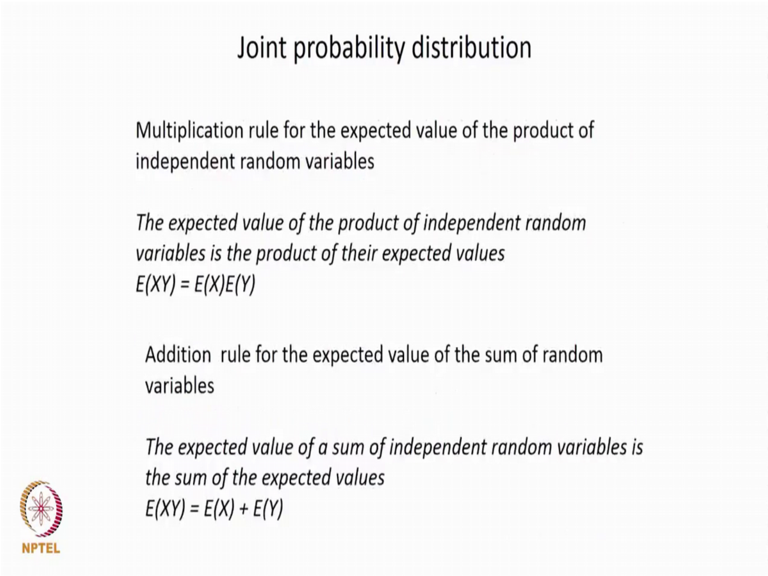
(Refer Slide Time: 04:09)



So, joint probability distribution of X and Y labeled as p x y gives the probability of events of the form X equal to x and Y equal to y. This represents the simultaneous outcome of both the random variables. So, P of X equal to 0 will be P of X equal to 0 and Y equal to 100 plus P of X equal to 0 and Y equal to 0 plus P of X equal to 0 and Y equal to minus 100. Y we know from the previous slides that Y we can take values of 100, 0 and minus 100 with the given probability.

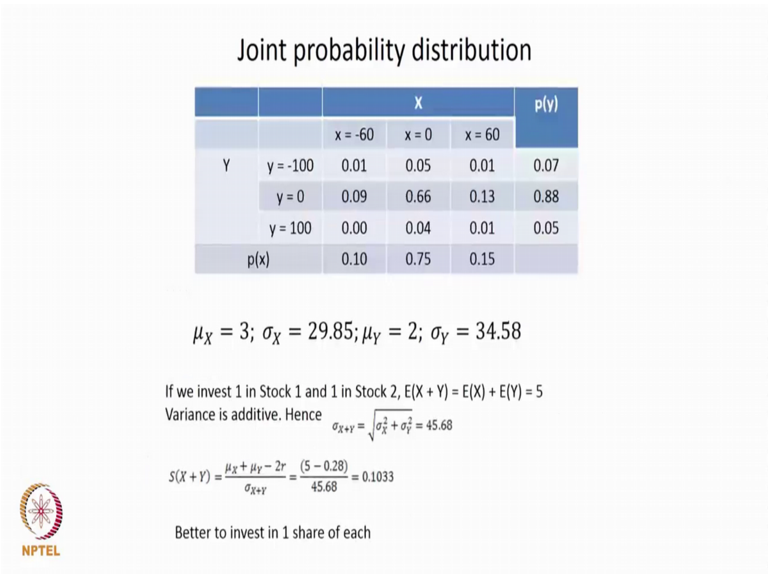
We also define independent random variables. Two random variables are independent if and only if the joint probability distribution is the product of the marginal probability distribution. X and Y are independent if p of x comma y is equal to p of x into p of y.

(Refer Slide Time: 05:15)



Now, multiply now we will define some more things multiplication rule for the expected value of the product of independent random variables the if the random variables are independent the expected value of the product of independent random variables is the product of the expected values. So, E of XY is equal to E of X into E of Y. Addition rule for the expected value of the sum. So, expected value of a sum of independent random variables is the sum of the expected values. So, E of XY is equal to E of X plus E of Y.

(Refer Slide Time: 05:54)



Now, let us find the joint probability distribution for X and Y. We show X here; x can take three values 60, 0 and minus 60, 0 and 60. You can take minus 100, 0 and 100. Please note that we have just change the order the way the order was from the earlier slide. Now, we look at the probabilities of y equal to minus 100 has 0.07. So, from this y equal to minus 100 as 0.07 and then we realize now that x equal to minus 60, x equal to 0 and x equal to plus 60. The probabilities add up to one. So, we have 0.07 here and then we multiply with the three probabilities they are rounded off suitably. So, that we get 0.01, 0.05, 0.01 which becomes 0.07.

Similarly, for y equal to 0, x equal to minus 60, they are multiplied suitably to get 0.88 multiplied and suitably shown to get 0.88 and y equal to 100 for all the x values adds up to 0.05. Similarly, the x probabilities are also 0.1, 0.75 and 0.15 as we see here 0.1, 0.75 and 0.15. So, we have completed this table just like we completed this table in an earlier lecture in statistics.

Now, we also know that mu of X is 3, sigma X is 29.85, mu of Y is 2, sigma Y is 34.58. So, if we invest something in stock 1 and something we invest 1 in stock 1 and 1 in stock 2 E of X plus Y is equal to E of X which is 3 plus mu Y is 2 which is 5 variants as additive. So, sigma is 45.68 and the ratio will be mu X plus mu Y 2r by sigma X plus Y which becomes 1.1033 and right now it makes sense to it is better to invest in one share of each together rather than put everything in either the first share or to put in the second share.

(Refer Slide Time: 08:27)

Exercise

In a sweet shop, customers buy either a 100g sweet or 200 g sweet along with 100 g of mixture or 250 g of mixture.

		X	
		100g	200g
Y	100	0.4	0.2
	250	0.3	0.1

$\mu_x = 130$; $\sigma_x^2 = 2100$;
$\mu_y = 160$; $\sigma_y^2 = 5400$

a) Find the marginal distribution of X and Y
b) What is the expected total weight of the purchase?
c) Are X and Y dependent or independent?
d) Is the variance of the total weight X + Y, equal, larger or smaller than $\sigma_x^2 + \sigma_y^2$

a) [0.7, 0.3], [0.6, 0.4]; b) 130+160 = 290;
c) dependent $0.4 \neq 0.7 \text{x} 0.6$; d) $6900 < 2100 + 5400$

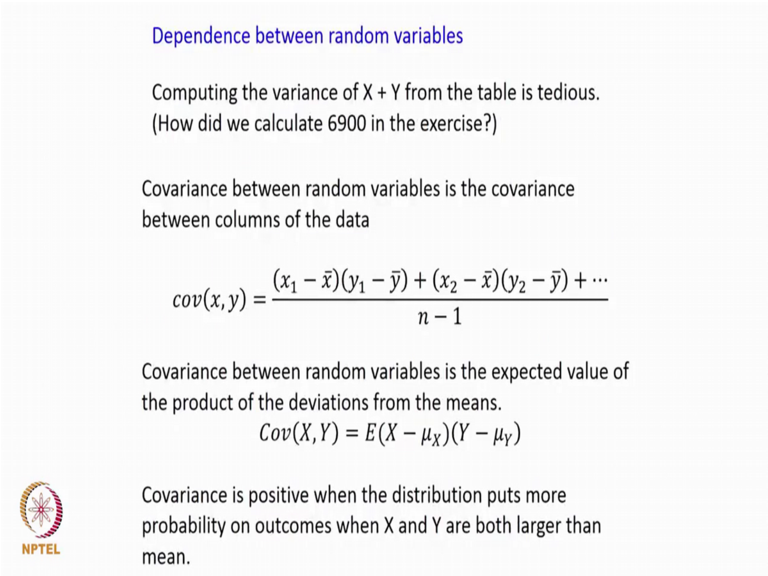
Now, let us look at a similar example to understand this further. In a sweetshop customers by either a 100 gram sweet or a 200 gram sweet along with the 100 gram mixture or a 250 gram mixture. Now, the probabilities are given. So, probability of buying a 100 gram sweet and a 100 gram mixture is 0.4 and so on. So, find the marginal distribution of X and Y, the table is shown here. What is the expected total weight of the purchase? So, marginal distribution would be 0.7 and 0.3, 0.6 and 0.4. So, the expected weight of the purchase will be 130 plus 160 which is 290.

Now, how do we get this? So, this 130 and 160 come as 100 gram sweet with 0.7 X is 70, 200 grams with 0.3 60 giving us 130, 100 gram with 0.6 is 60, 250 grams with 0.4 is 100. So, we get 160 and 290 is the expected weight E of X plus E of Y. Are X and Y dependent or independent they are dependent because we also from the; if we do this, this is 0.6 here and this is 0.7 here. So, we would have multiplied them and got 0.42.

But, since we have only 0.4 they are now dependent on each other. If they were independent then the number here would have been 0.6 into 0.7 the number here would have been 0.6 into 0.3 which is 0.1 8, but since it is 0.2 and not equal to the multiplication of the probabilities we say that they are dependent, this is the variance of the total weight X plus Y equal or larger or smaller than sigma X square plus sigma Y square.

Now, we realize that sigma X square is 2100, sigma Y square is 5400 and the variance of the total weight X plus Y we can do that as well we can say X plus Y can now take 200, can take 300, can take 250 and take 450 with the probabilities that are given, we can now find the expected value, and the variance and if we do that we will get 6900 which is smaller than sigma X square plus sigma Y square.

(Refer Slide Time: 11:08)



Now, let us also try to find out the dependence between random variables, just as we introduced the covariance earlier in statistics, now we try to find the covariance between random variables. So, computing the variance of X plus Y from the table is a little difficult. How did we calculate 6900, we will see is there a way to do that. Now, covariance between random variables is the covariance between the columns of the data.

So, covariance between x and y is x 1 minus x bar into y 1 minus y bar plus x 2 minus x bar plus y 2 minus y bar etcetera divided by n minus 1. So, covariance between random variables is the expected value of the product of the deviations from the mean. So, covariance X, Y is equal to E of X minus mu X into Y minus mu Y. So, covariance is positive when the distribution puts more probability and outcomes when X and Y are both larger than the mean.

(Refer Slide Time: 12:11)

Covariance

		X			p(y)
		x = -60	x = 0	x = 60	
Y	y = -100	0.01	0.05	0.01	0.07
	y = 0	0.09	0.66	0.13	0.88
	y = 100	0.00	0.04	0.01	0.05
p(x)		0.10	0.75	0.15	

$\mu_X = 3;\ \sigma_X = 29.85;\ \mu_Y = 2;\ \sigma_Y = 34.58$

$(\sigma^2_{X+Y} = 2235) \neq (\sigma^2_X + \sigma^2_Y = 2087)$

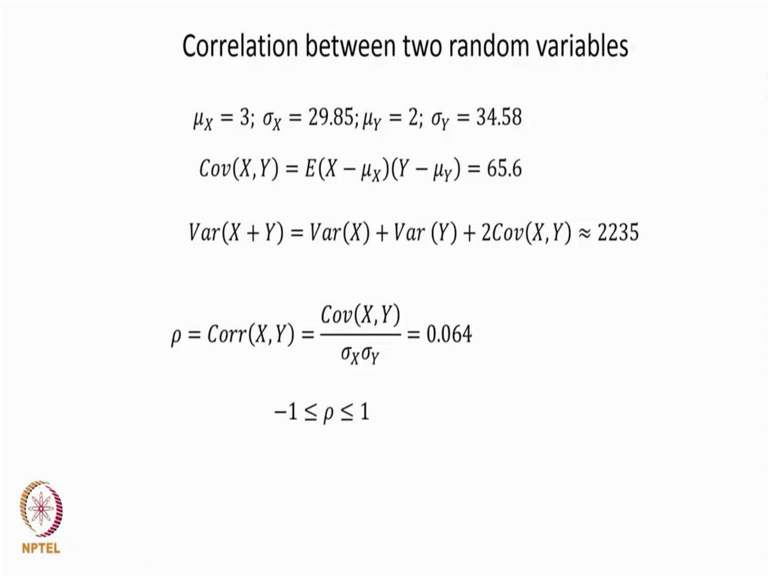
$$Cov(X,Y) = E(X-\mu_X)(Y-\mu_Y) = 65.6$$

$$Var(X+Y) = Var(X) + Var\,(Y) + 2Cov(X,Y) \approx 2235$$

So, let us now find out the covariance for these and from here we can know this. So, x is minus 60, 0 and plus 60; y is minus 100, 0 and plus 100. We have the individual probabilities mu X is 3 sigma X is 29.85, mu Y is 2, sigma Y is 34.58 and we also found out that sigma square x plus Y is 2235. It was just not equal to sigma x square plus sigma Y square.

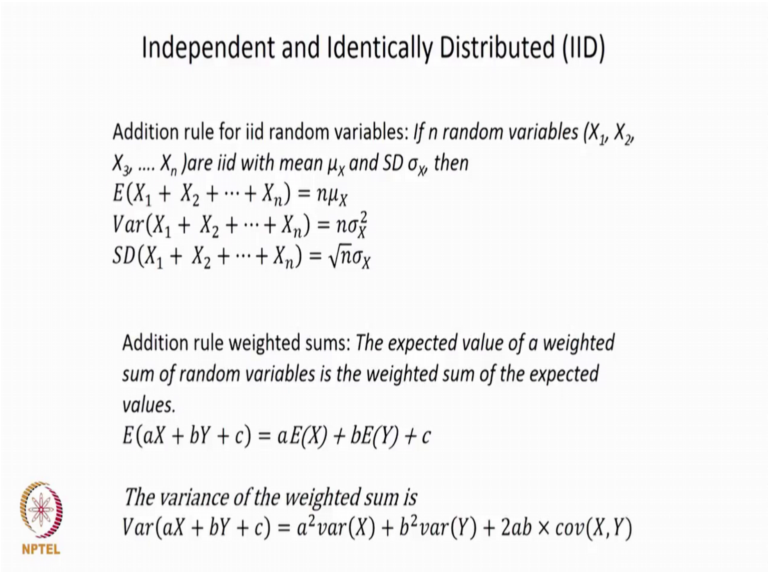
Now, covariance we can now calculate expected value of X minus mu X into Y minus mu Y. We know mu X is 3, x takes these three values, Y mu Y is 2, y takes these values. So, covariance is 65.6. Now, the variance of X plus Y is equal to variance of X plus variance of Y plus 2 times covariance of X, Y which we get 2235 in this example. That is a way to find out the variance of X plus Y rather than try to take each of these cases and individually try to find out it is easy to find the covariance and from the covariance go back and calculate variance of X plus Y.

(Refer Slide Time: 13:33)



Now, we can also find the correlation between two random variables. So, correlation of X comma Y is equal to covariance of X, Y divided by sigma X sigma Y and in this case it turns out to be 0.064, because we already know the values of sigma X, sigma Y we know the covariance of X, Y and correlation is 0.064. As usual the correlation coefficient is between minus 1 and plus 1.

(Refer Slide Time: 14:01)



Now, addition rule for independent and identically distributed iid as they are called addition rule for iid random variables. So, if n random variables X 1, X 2, X 3, to X n are iid independent and identically distributed with mean mu X and sigma standard deviations sigma X then the expected value of the sum of them are X 1 plus X 2 to X n is equal to n times mu X, variance is n times sigma X square, standard deviation is root n into sigma X.

Now, addition rules like we did before. So, E of a X plus b Y plus c is a of E X plus b of E Y plus c. Variance is a square into variance of X plus b square and a variance of Y plus 2 ab into covariance of X, Y the constant c goes.

(Refer Slide Time: 14:57)

Match the following

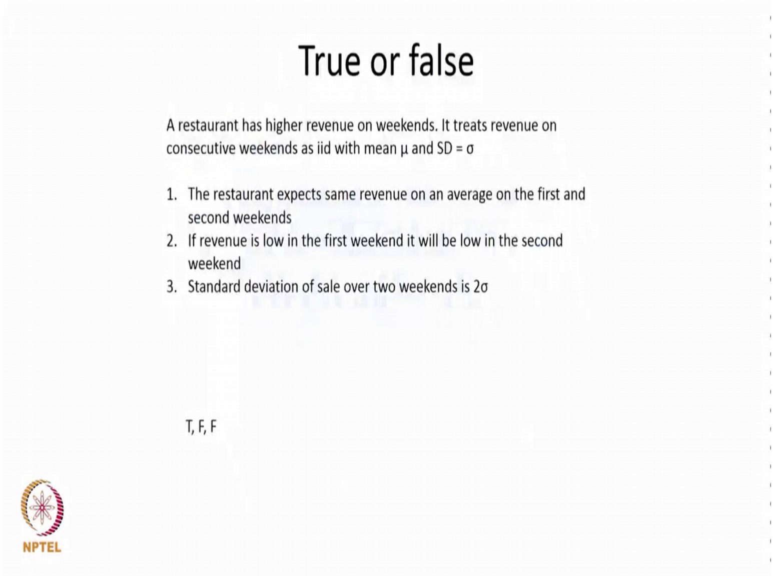
Number	Column A	Column B	
1	Positive covariance	ρ	6
2	X and Y are identically distributed	Var (X+Y) = Var (X) + Var (Y)	3
3	Uncorrelated random variables	X + 3Y	5
4	Covariance	Var (X+Y) > Var (X) + Var (Y)	1
5	Weighted sum of two random variables	$\rho\sigma_x\sigma_y$	4
6	Correlation coefficient	p(x) = p(y)	2

NPTEL

So, we now have six items from 1 to 6. So, you look at this. So, we have positive covariance. So, when the covariance is positive variance of X plus Y is greater than variance of X plus variance of Y which is something we saw. If X and Y are identically distributed p of X is equal to p of Y. 3 – uncorrelated random variables; so, variance of X plus Y is equal to variance of X plus variance of Y. So, covariance does not act therefore, no correlation.

Item – 4, covariance. Covariance is correlation coefficient into sigma X and sigma Y the definition of correlation is covariance by sigma X sigma Y. Therefore, covariance is correlation which is rho into sigma X sigma Y. Weighted sum of two random variables is something like X plus 3y, x is a random variable Y is another random variable. 6 – correlation coefficient is given by the symbol rho.

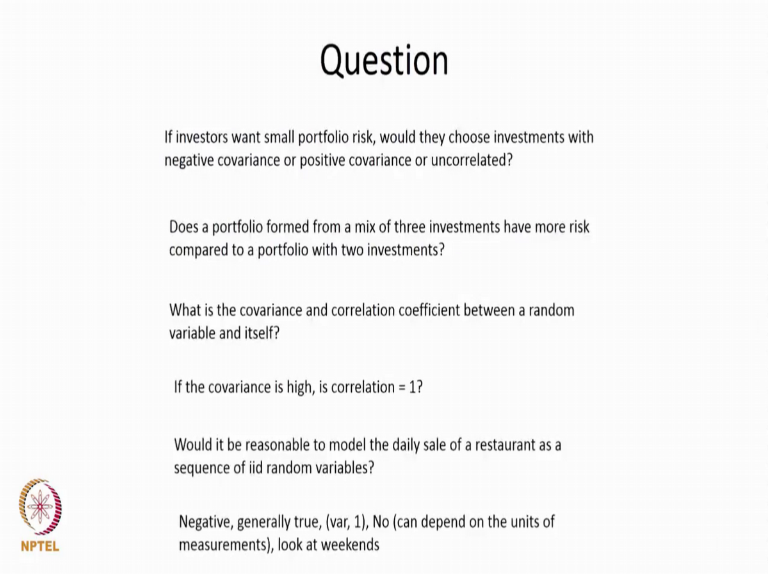
(Refer Slide Time: 16:01)



Look at some true or false. A restaurant has higher revenue on weekends it treats the revenue on considered a weekends as iid with mean mu and standard deviation sigma. So, we will just check the restaurant expects the same revenue on an average on the first and second weekends. Yes, because they are iid, independent and identically distributed you can expect the same average.

If the revenue is low in the first weekend it will be low in the second weekend, not necessarily only the expected values are equal. The random variable can still take different values and therefore, it can be false. Standard deviation of sale over two weekends as 2 sigma. It will be root 2 times sigma because we found out that variance is are additive, standard deviation is not. So, for two weeks it will be sigma square plus sigma square which is 2 sigma square and the standard deviation will be root 2 times sigma.

(Refer Slide Time: 16:59)

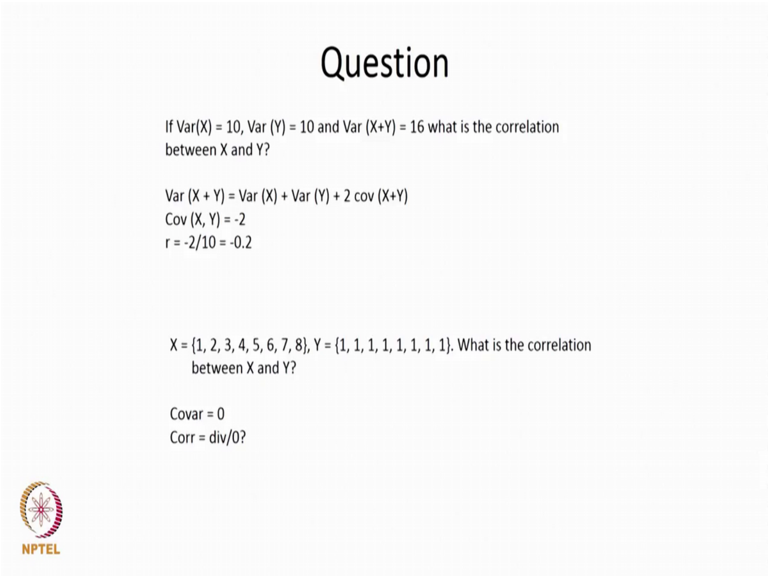


If investors want small portfolio risk would they choose investments with negative covariance or positive covariance or uncorrelated? So, they would choose something with a negative covariance, so that the risk which is the variance of X plus Y would reduce with negative covariance. So, it is good to choose investments which have a negative covariance. So, that the portfolio risk comes down. Does a portfolio formed from a mix of three investments have more risk compared to portfolio with two investments? I would say it is generally true because more the diversification the less would be the risk, but then one also has to look at returns and in this question we are only looking at risk. Therefore, we would say generally true, but then the return can come down and so on.

What is the covariance and correlation coefficient between a random variable and itself? So, between the random variable and itself the covariance is the variance and the correlation coefficient is 1. If the covariance is high is the correlation equal to 1? So, one would generally get a feeling that since correlation is equal to covariance divided by sigma X sigma Y high covariance can lead to a correlation closer to 1, but the actual answer is covariance by itself having a large value can also depend on the unit of measurement of the covariance.

We have already seen that when things were measured in rupees there was a certain value and when they were or when they are measured in paisa then the covariance becomes different and becomes much larger. So, the answer would depend on the unit of measurement of the random variable. Would it be reasonable to model the daily sale as a sequence of radically distributed and independent random variables? Not necessary, because we have to look at weekends before we do that it might follow two different kinds of things with weekend sales being different as seen in the previous question.

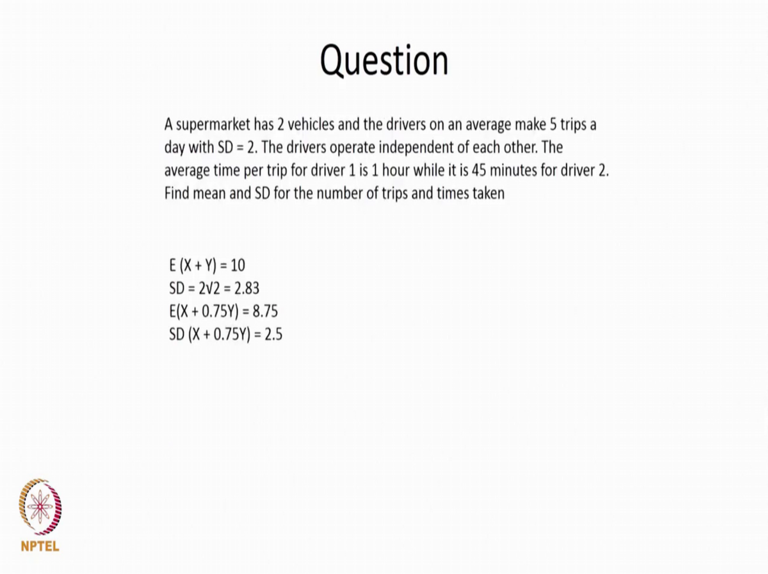
(Refer Slide Time: 19:07)



Now, let us look at a few more simple questions a variance of X is 10 and variance of Y is 10 and variance of X plus Y is 16, what is the correlation between X and Y? So, variance of X plus Y is equal to variance of X plus variance of Y plus 2 times covariance of X plus Y. Therefore, covariance of X Y is minus 2 and correlation is minus 0.2.

X is a random variable 1, 2, 3, 4, 5, 6, 7, 8, Y is 1, 1, 1, 1, 1, 1, 1, 1. What is the correlation? So, coefficient will be covariance will be 0 and we cannot find a correlation because we could standard deviation of one of them would be 0 and therefore, we would not be able to find the correlation.

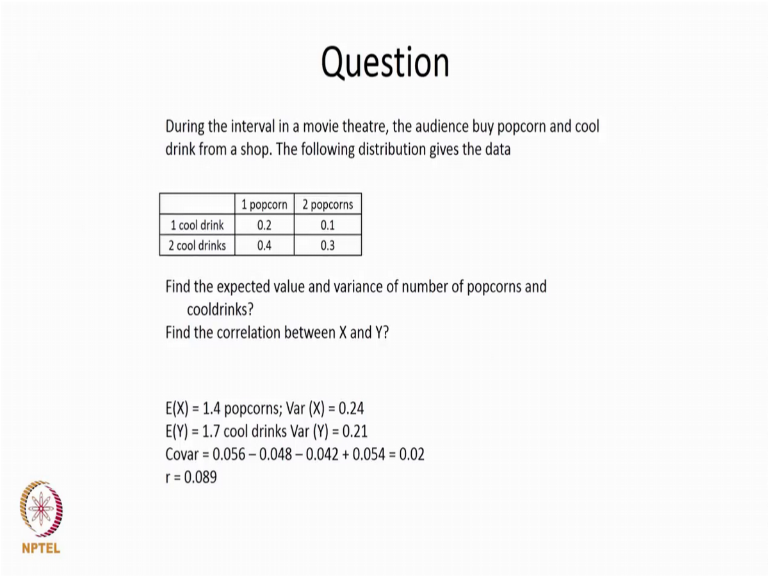
(Refer Slide Time: 19:53)



Now, a supermarket has 2 vehicles and the drivers on an average make 5 trips a day with standard deviation equal to 2. The drivers operate independent of each other. The average time per trip for driver 1 is 1 hour, while it is 45 minutes for driver 2. Find the mean and standard deviation of the number of trips and time taken.

So, they make 5 trips on and average. So, E of X plus Y is 10, 5 plus 5. Standard deviation is 2 times root 2, the standard deviation is 2. So, root 2 times 2 there are two drivers so, 2.83. The time taken is X plus 0.75Y, so, 8.75 and standard deviation of X plus 0.75Y turns out to be 2.5 when we do use the equations to find out the value.

(Refer Slide Time: 20:43)



Another example in a during an interval in a movie theater the audience buy popcorn and cool drink from a shop. So, the distribution is given. So, we could assume that the they buy one cool drink or two cool drinks with one popcorn or two popcorns distributions are given. So, find the expected value and the variance of the number of popcorns and cool drinks, find the correlation.

So, expected value for X is 1.4 popcorn, because 1 into 0.6 plus 2 into 0.4 is 1.4, variance of X is 0.24. For Y it is one cool drink into 0.3 plus two cool drinks into 0.7, which is 1.7 and that variance is 0.21. Covariance we can separately find as X minus mu X into Y minus mu Y, which gives 0.02 and the correlation is 0.089.

So, with this we come to the end of the discussion on the topic association between random variables. In the remaining lectures in this course we would concentrate a little more on known distributions and we will take up the binomial distribution and the normal distribution has two examples and continues our discussion on these distributions in the remaining part of this course.