Statistics for Experimentalists

Prof. Kannan. A
Department of Chemical Engineering
Indian Institute of Technology – Madras

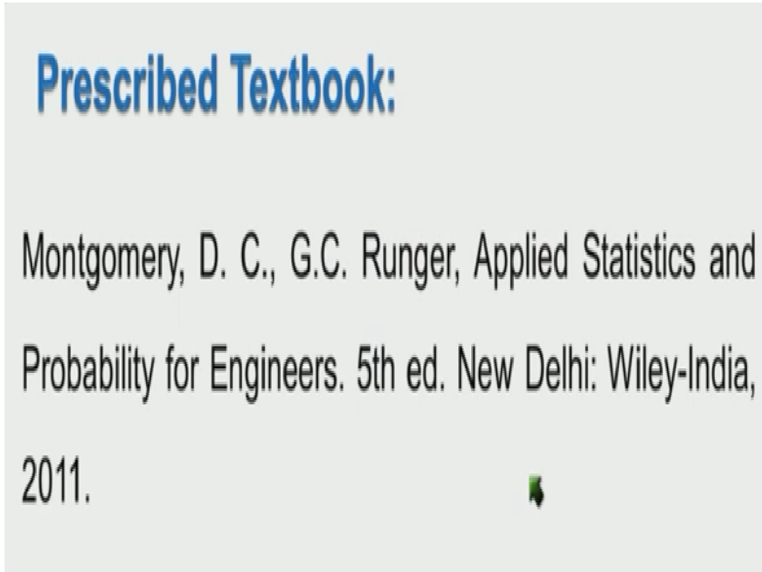
Lecture - 12
Point Estimation

Hello again, today's class, we will be looking at point estimation. In the previous lecture, we were looking at random sampling and the properties of random samples. We saw how to find the mean and variance of random samples. The appropriate degrees of freedom to be used in the calculation of the sample variance. We also saw that the random sample involved collection of random variables, x1, x2, so on to xn.

In a general case, they may not be independent, so we had to find out how to estimate the variance and mean in such cases. We also defined covariance. Now as far as the random sample goes, we simplify things somewhat by assuming that the random samples are independent so that the covariance between pairs of the random variables in that random sample vanish and also they are identically distributed.

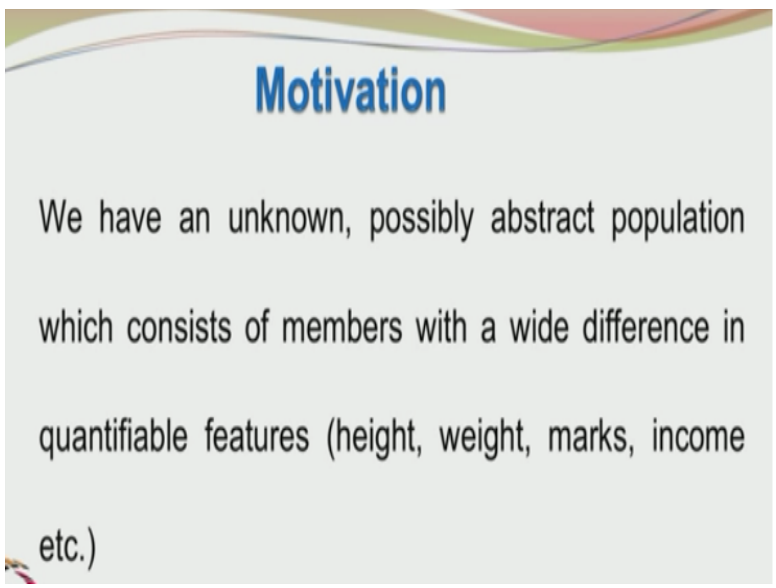
If they are identically distributed, they have the same parameters of the distribution, not only the nature of the distribution is identical for all these random variables, but the parameters are also identical, that is what I mean when I say identically distributed. Now let us look at point estimation using the sample collected.

(Refer Slide Time: 02:17)



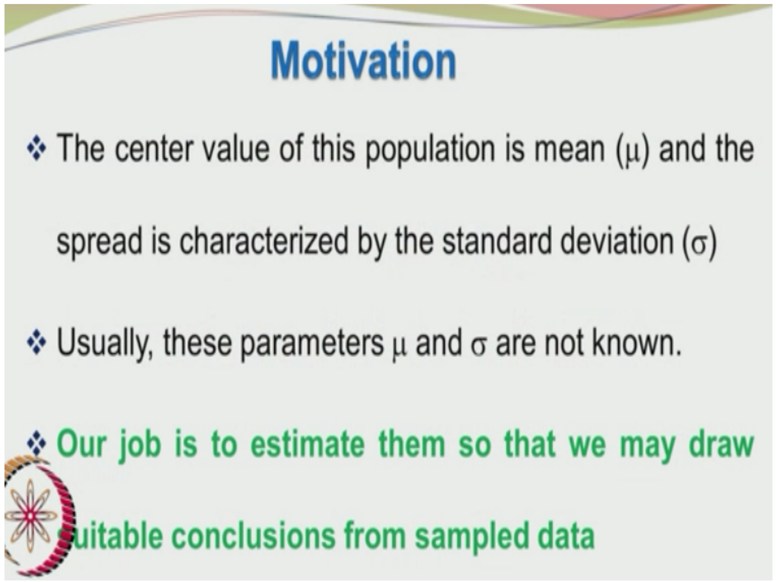
The prescribed textbook where the information regarding this topic is found is the one by Montgomery and Runger.

(Refer Slide Time: 02:26)



So the motivation for taking random samples and going in for point estimation lies in the fact that the population is an unknown entity, is a mysterious entity. We do not know the parameters of the population. All we know is the population will comprise of entities, which are having a wide difference in quantifiable features, like, height, weight, marks, income, etc. You always have entities on either extremes, but usually the majority of the entities of the population lies close to the average.

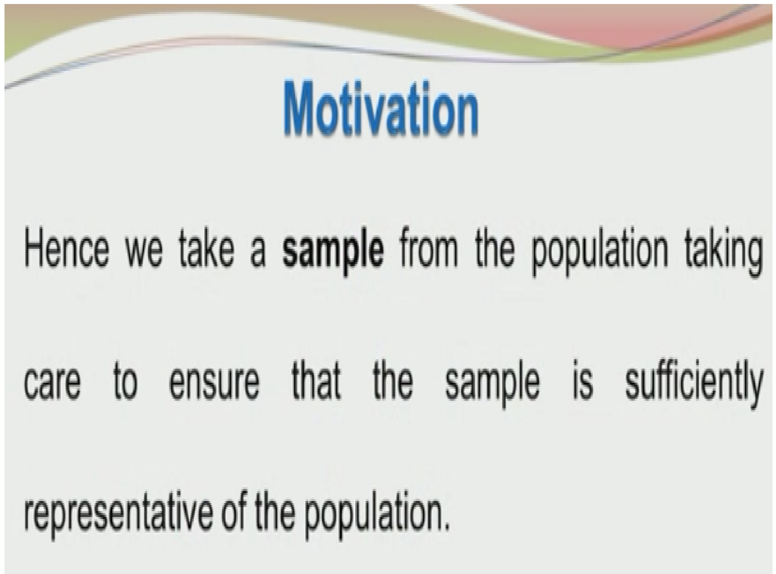
(Refer Slide Time: 03:32)



The center value of this population is mean mu and the spread is characterized by the standard deviation sigma. However, usually these parameters mu and sigma are not known. Since they are not known, it does not mean that we give up our exercise, we estimate them, so that we may draw appropriate conclusions which will help in our decision making after we have sampled the data. If you reflect many of the decisions are based on the sampling service conducted by us or by the appropriate competent authority.

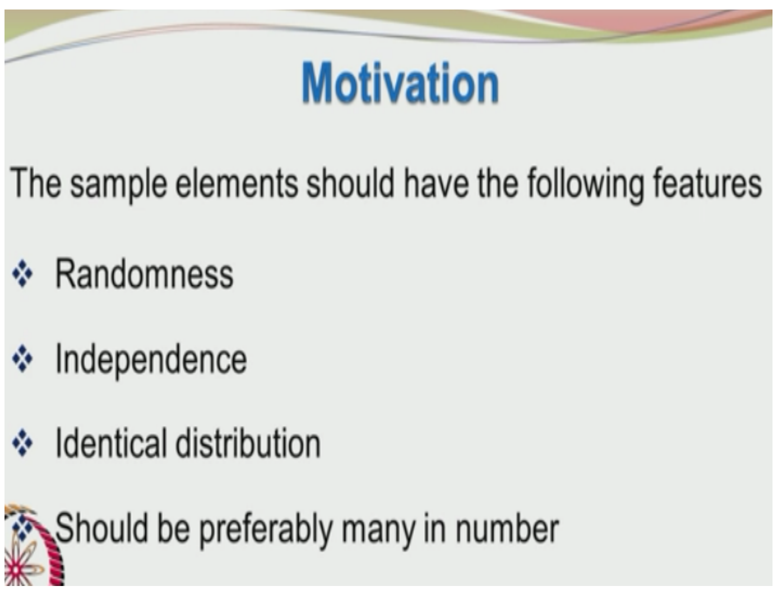
Time is not there to understand the entire population or the entire sphere of activities, so a sample survey is conducted and based on that suitable conclusions are drawn and then appropriate decisions are taken.

(Refer Slide Time: 04:54)



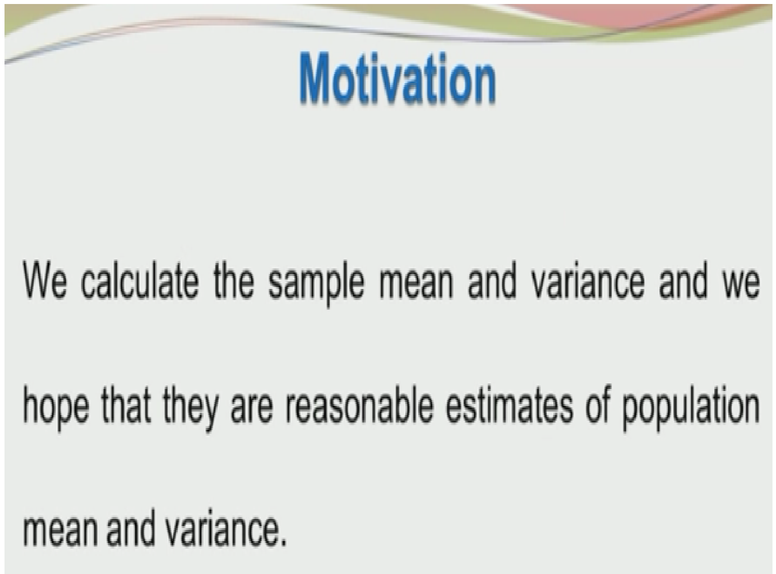
We have to make sure that whatever sample we are drawing is sufficiently representative of the population so that the decision which is being taken is affecting the entire population and not only a select portion of the population.

(Refer Slide Time: 05:25)



The sample elements should have the following features. They should be random. They should be independent. They should enjoy identical distribution and should be preferably many in number.

(Refer Slide Time: 05:44)

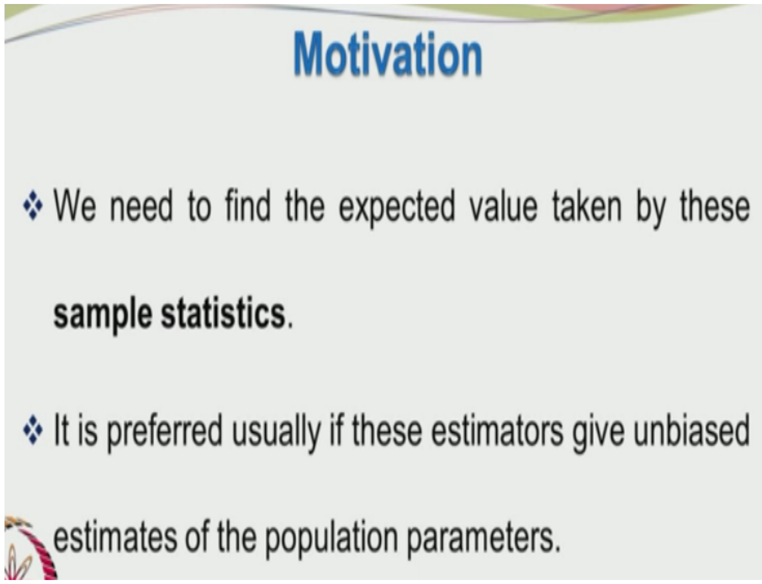


So we use the sample mean and the sample variance as surrogates for the population mean mu and population variance sigma square. We hope or we expect that these are adequate estimates or I would even modify that into adequate estimators of the population mean and population variance. So I am now introducing mu, what is mean by an estimator, what is mean by an estimate. In the previous class, we defined statistics.

Now I am introducing new terms in today's lecture, the very first new term was point estimate, then I have also introduced terms like point estimators, point estimates. So let us see how they are defined and applied. The important thing is the nomenclature or the notation for all these defined quantities. It is important that we are consistent in the notation and terminology. For this purpose, I am following the terminology given by Montgomery and Runger.

If you are following any other source of material on statistics and design of experiments, please make sure that the notation and the terminology are consistent.

(Refer Slide Time: 07:23)

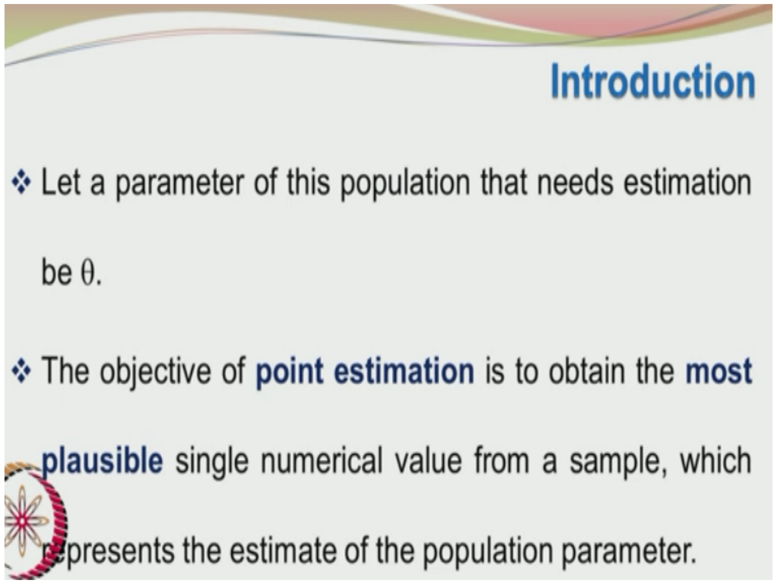


So once we have taken the sample, we do some mathematical calculations with those samples and we obtain the sample statistics. The sample statistics are used as estimators for the population and it is important that the estimators based on the sample statistics give unbiased estimates of the population parameters. They should not bloat up the population parameter or they should not unnecessarily make it very small.

If the estimator is inaccurate, for example we are looking at the gross income of the citizens of a country, if the estimators are biased, then we would get a wrong opinion about the income levels in the nation. Sometimes if the estimators are giving wrong values for the population variance, then the spread may not be accurate. It may be either too narrow or it may become too broad, in which case the decisions will also be affected by the wrong parameter estimates.

So it is important that the estimators give unbiased estimates of the population parameters. Please note that, these estimates obtained from the sample mean and sample variance are really not unique values. We live in a very fuzzy world, where nothing seems to be certain and so we need to also account for the variability in these estimators themselves. So we have to look at the variability in the sample mean. We have to look at the variability in the sample variance.

(Refer Slide Time: 09:54)

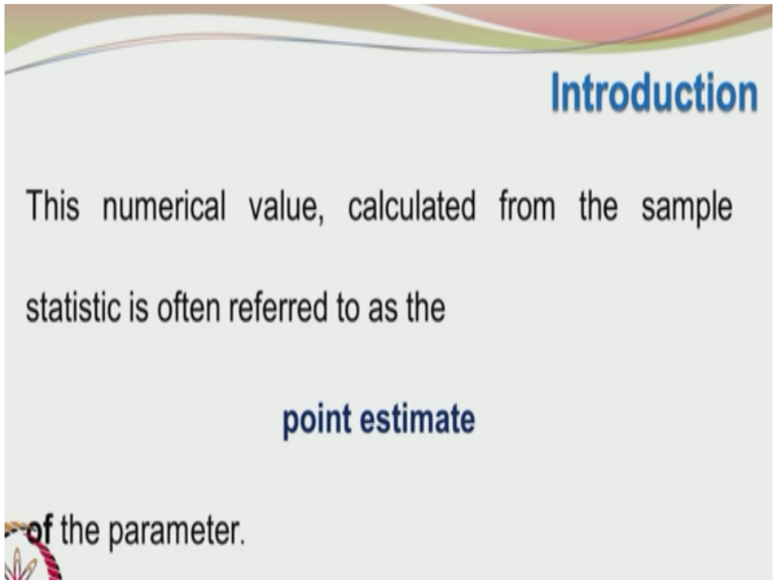


So let us denote the parameter of the population as theta. This is not a absolute parameter terminology. It is a general terminology for the population parameter. We call it as theta. If there are 2 parameters in the population, we may want to generally term it as theta1 and theta2. The next line is important. We are getting a simple value estimate of the population parameter. That is what is called as point estimation process.

So the objective of the point estimation is to get the most plausible single numerical value from a sample, which represents the estimate of the population parameter. So we have a sample. We use the sample to get the most likely or the most believable single numerical value and we then proclaim that it is the reasonable estimate of the population parameter. So there will be skeptics who will question, how can you confidently say that the estimate you have taken from the sample is truly reflective of the population parameter.

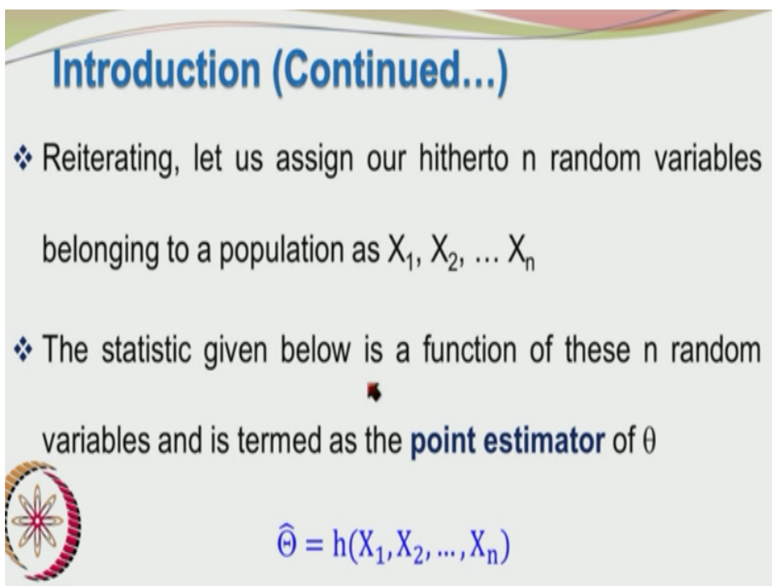
So we need to understand about this point estimation process further.

(Refer Slide Time: 11:51)



So this numerical value calculated from the sample statistic is often referred to as the point estimate of the parameter. So you have obtained the point estimate, a single value of the population parameter and you call it as the point estimate.

(Refer Slide Time: 12:20)

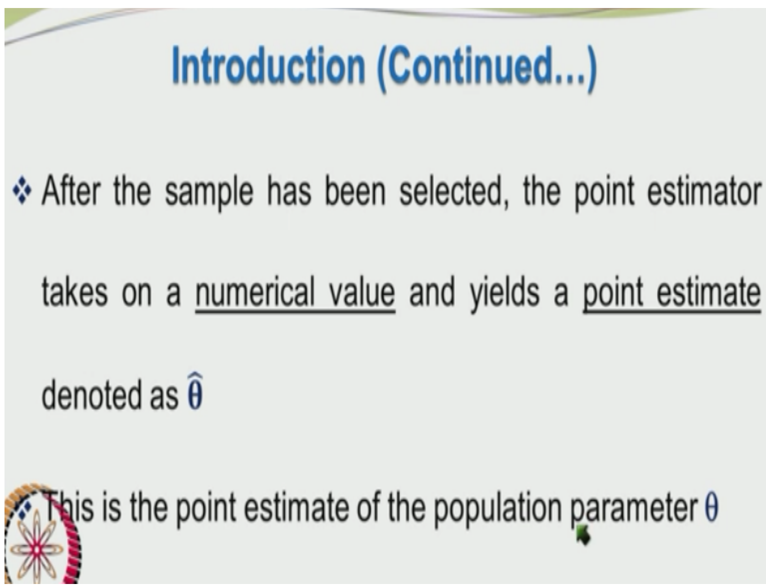


So a sort of summarizing what we have done up to now, we have n random variables picked up from a population x1, x2, so on to xn. The statistic given below is a function of these n random variables and is termed as the point estimator of theta. So we have a function which will manipulate the n random variables in a suitable manner and create a new random variable. When the collection of random variables are mathematically manipulated, added, subtracted, multiplied whatever, they finally yield a function relating all these random variables.

That itself is a random variable. It is also based on the sample so we call it as the sample statistic and a suitably chosen sample statistic is used as the point estimator of the population parameter. So coming to the slide, the population parameter is represented by theta and you have a statistic which is based on a functional relationship between the n random variables and that statistic or a suitably chosen statistic is used as the point estimator theta.

We denote the point estimator of theta using the hat and this symbol. This h represents the functional relationship involving the n-th random variables.

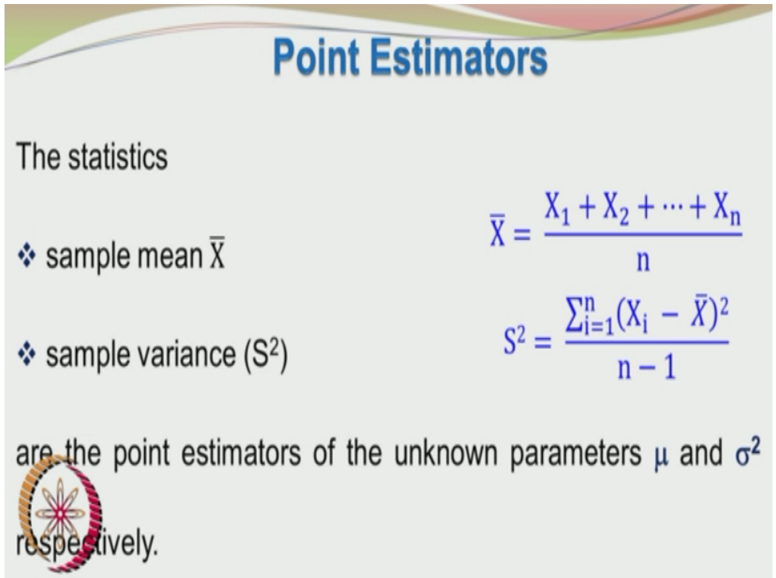
(Refer Slide Time: 14:31)



So once you have chosen the sample, let us say that you are having a population of people with varying heights and we really do not know the average height of the population. For example, let us say the population is described as the height of soldiers in the army. So we have absolutely no idea on the average height of soldiers in the army. So we have to take a random sample. Once you have taken a random sample, then you know the heights of all the army people, you have chosen during your sampling.

So the values of the random variables are now known. So based on these, you can use the defined estimator based on the statistic to obtain a numerical value and that is the point estimate of the required population parameter and we call it as theta hat. Theta is the actual terminology for a population parameter. The point estimate of the theta is denoted by theta hat. So theta hat is the point estimate of the population parameter theta.

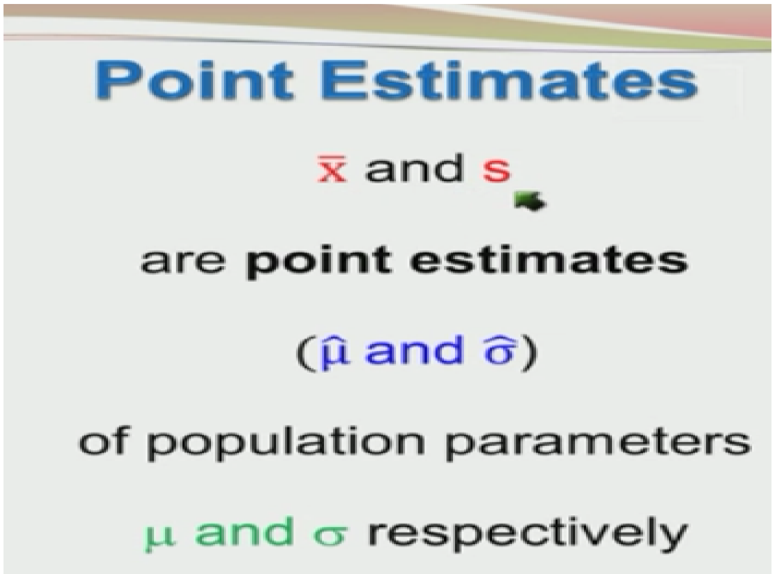
(Refer Slide Time: 16:27)



So what are the point estimators which are available to us and which are also usually encountered or commonly encountered and they are not surprisingly the sample mean x bar. It is a point estimator. The sample variance s square is also a point estimator. The sample mean x bar is an estimator for the population mean mu, a sample variance s square is an estimator, a point estimator at that for the population variance sigma square.

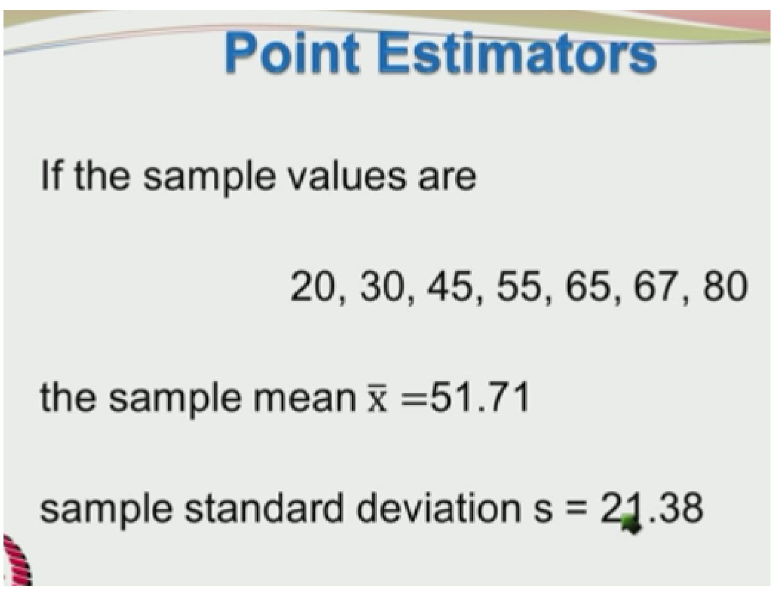
Sample mean x bar is also a point estimator and these are the definitions for the sample mean and sample variance. These are point estimators of the unknown parameters mu and sigma square respectively.

(Refer Slide Time: 17:33)



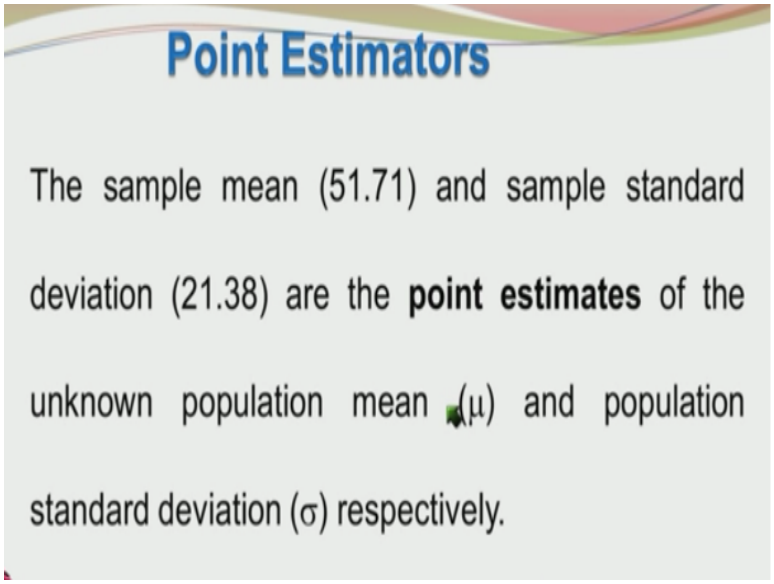
X bar and s are point estimates mu hat and sigma hat of population parameters mu and sigma respectively. So from the sample variance, we can find the sample standard deviation and that will be denoted by s. So x bar and s are point estimates mu hat and sigma hat of the population parameters mu and sigma respectively. We use this for general notation, the theta hat, theta 1 hat and theta 2 hat or mu hat and sigma hat are used in the terminology and we will stick to it.

(Refer Slide Time: 18:28)



So you have a sample comprising of 7 entities. Obviously this is a small sample. However, in life, you may have to work with what you get and maybe there are certain reasons why you are unable to collect a large sample. So we have to use a small sample and draw or try to draw the appropriate conclusions. The sample mean x bar is 51.71, you may want to take up a calculator or a spread sheet and verify that it is in deed, so the sample standard deviation is rather high at 21.38.

(Refer Slide Time: 19:20)



So these are actual numbers and hence the sample mean 51.71 and sample standard deviation 21.38 or the point estimates of the unknown population mean mu and population standard deviation sigma respectively. So let us say that we have measured the required attribute from our particular sample. We have calculated the sample mean and sample standard deviation and we use them as point estimates of the unknown population mean mu and standard deviation sigma.

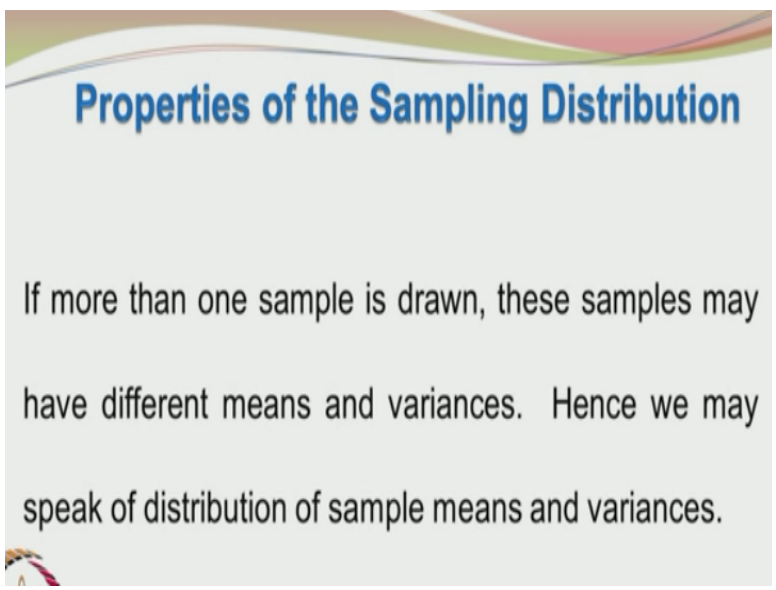
For example, we are interested in finding the marks in a particular subject. That involves a huge population that are students belonging to a particular board who are taking the particular subject, let us call it as mathematics and we want to find the average of this population and also the standard deviation. We want to know the average mark and the average standard deviation. To do that, we have to either look at the records of the students who have been writing the maths exams for the last 30-40 years or we can take a particular sample and see the marks.

Obviously, the sample has to be carefully chosen. The sample, which is being chosen based on the current performances may not be adequately reflective of the performances over the last 30-40 years. So we may have to draw a sample of adequate size across the years. So there will be a lot of other issues involved in random sampling to ensure it is truly random. It is beyond the scope of this course to get into these issues.

So let us assume that we have collected a random sample and the sample is indeed random and it is obeying the required attributes. So the value we get from the sample are the mean and standard deviation usually and we can get those pretty easily and let us say in this particular case we have numbers like 51.71 and 21.38 for sample mean and sample standard deviation respectively. So we go even far to say that the population mean and population standard deviation sigma are pretty much close to these values.

We are not claiming that they are indeed 51.71 and 21.38. We say that they would be close to these values. So these values are estimates of mu and sigma, estimates are numbers which are considered to be close to the actual values. How close they are, how far they are, how identical they are, we really cannot say.

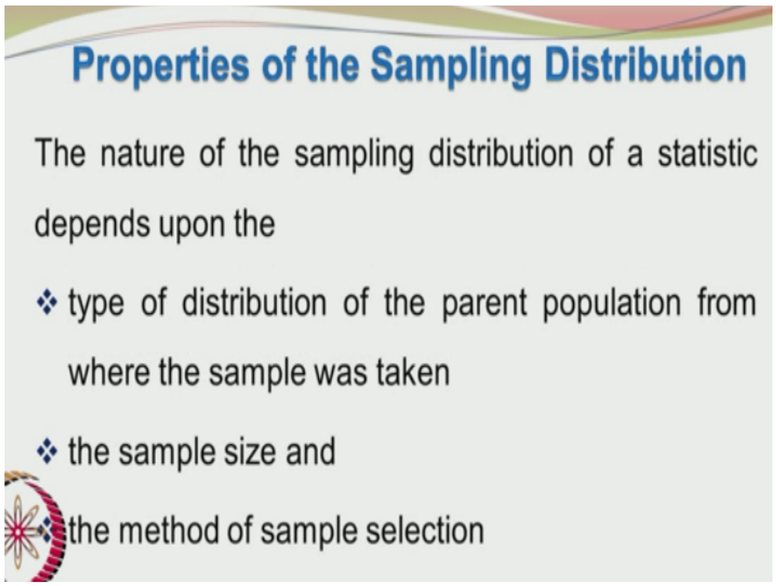
(Refer Slide Time: 23:16)



We need a bit more understanding in this course to get to those issues. I will come to that. The spread of the data and how far the estimate is expected to be from the actual value. So these issues we will address shortly after proceeding a bit further in this course. As the opinion polls experience shows that many sources conduct their own opinion polls, many agencies conduct their own opinion polls and so different samples are taken and the results are varying.

They are not identical. The sample surveys are not identical in their prediction, which means that the attributes of the samples drawn from a population can themselves be different. So we have to understand this difference in order to know the properties of their estimates. Since the samples can have different variances specifically, we may speak of a distribution of sample means and sample variances. This we saw in the previous class. I am just reiterating that point once again.

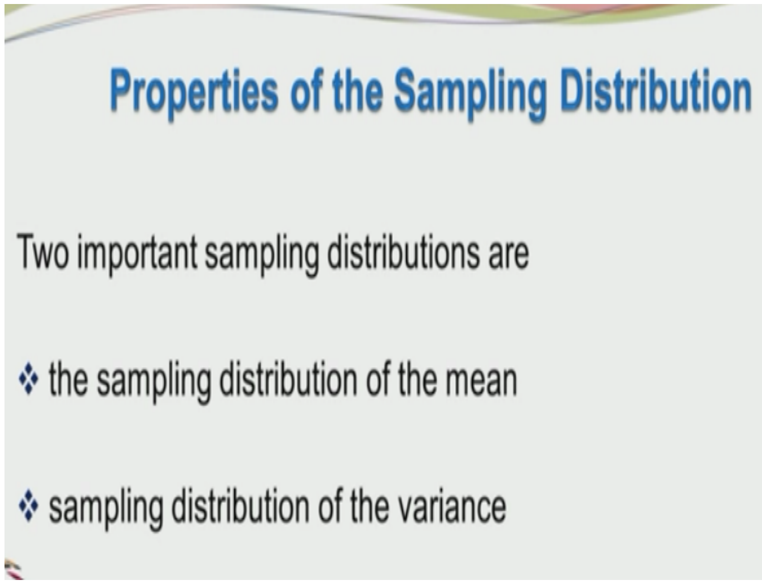
(Refer Slide Time: 25:10)



So the statistic itself is a random variable. It has a probability distribution associated with it and the nature of the sampling distribution of the statistic depends upon the type of distribution of the parent population from where the samples were taken. The sample size and the method sample selection, we saw that the distribution is narrow or the spread is less, when the sample size was larger. The spread denotes uncertainty and when we take a sample of a larger size, we reduce the uncertainty if not completely eliminated.

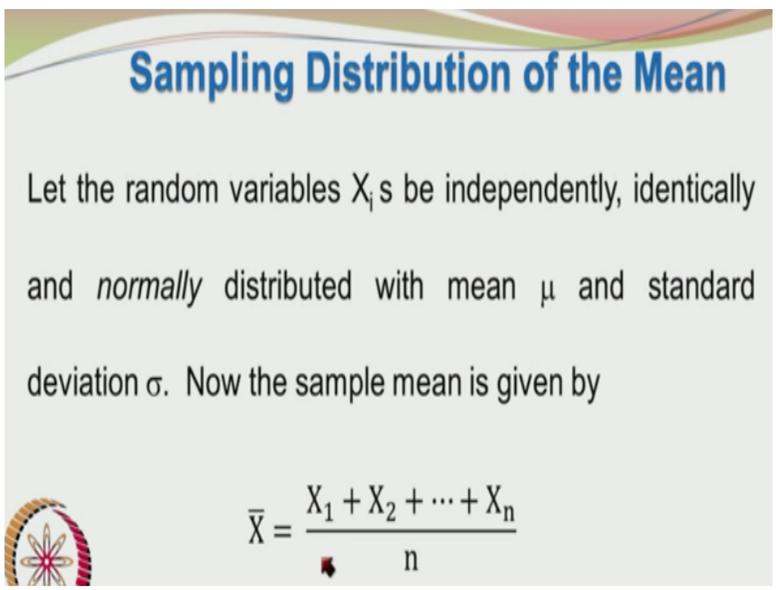
If you want to completely eliminate the uncertainty, the sample size should be pretty close to infinity, in other words we are sampling the entire population, which of course is not practical.

(Refer Slide Time: 26:18)



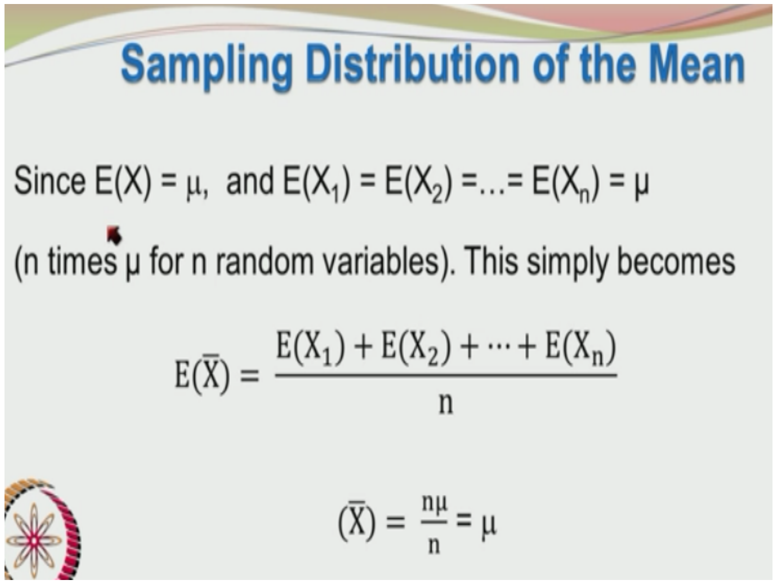
There are 2 important sampling distributions and these are the sampling distribution of the mean and the sampling distribution of the variance. We will be first focusing our attention on the sampling distribution of the mean.

(Refer Slide Time: 26:38)



So the statistic, the sample mean is defined as x1+x2+ so on to xn/sample size n. We have come across this definition several times during the course of these lectures and by now we should be familiar with the sample mean.

(Refer Slide Time: 27:05)

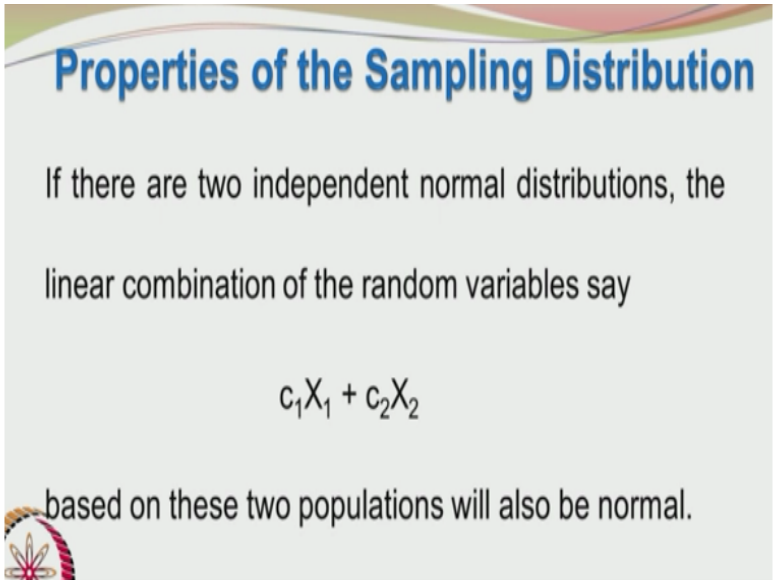


I also showed yesterday that the expected value of x bar is = population mean mu and that proof is very straight forward, expected value of x bar=E(x1)+E(x2)+ so on to E(xn)/n. Since all these random variables are taken from identical populations, which are not only identical in their shape, but also in their parameters. So all of them share the same parameters mu and sigma for the mean and standard deviation.

So expected value of x1 would be mu, expected value of x2 will also be mu, expected value of xn will also be mu. We have n such entities, so you have n mu/n, just mu. So there is a correction here. Earlier it was x bar, but it should not be x bar, expected value of x bar=n mu/n, which is =mu, very nice. It is not mu+0.3 mu or whatever, it is precisely mu. We expect that the x bar distribution will have mu as its average.

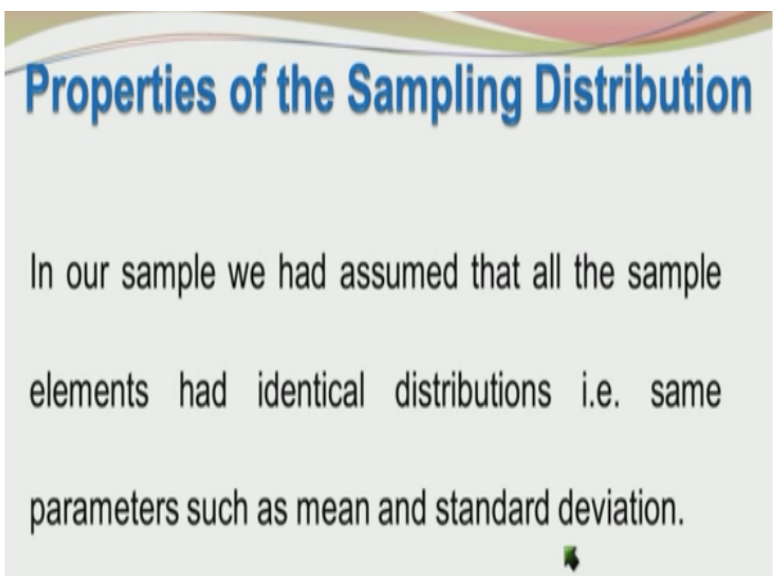
We know the expected value of a distribution is its mean, x bar is a distribution. That is a distribution of sample means and the mean of the distribution of the sample means=mu. So understanding this is important. You have a distribution of the samples means. If there are many samples taken, their means or averages would be different. They would form a distribution, but the average of this distribution of sample means will be = population mean mu. So that is what we have to keep in mind.

(Refer Slide Time: 29:35)



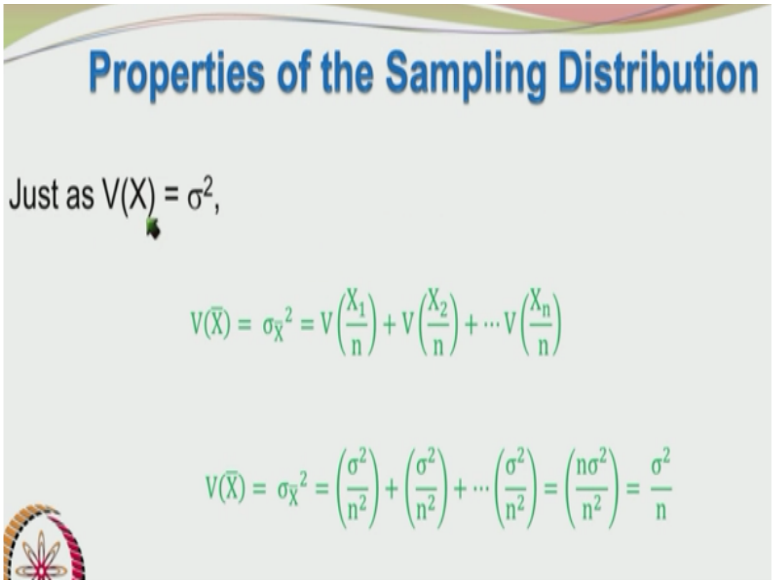
If the random variables for simplicity, let us say that we have only 2 random variables, x1 and x2, then we combine them in a linear fashion. For example, c1x1+c2x2, then the resulting random variable will also be normal. So you have 2 random variables x1 and x2, a linear combination is c1x1+c2x2. This will definitely be a random variable. So it will have its own distribution. If x1 and x2 were normal distributions, the random variable formed by the linear combination of x1 and x2 will also be a normal distribution.

(Refer Slide Time: 30:35)



So this x1 and x2 are coming from normal distributions. They are independent and we also assumed that they have identical parameters. So mu and sigma for both x1 and x2 are the same.

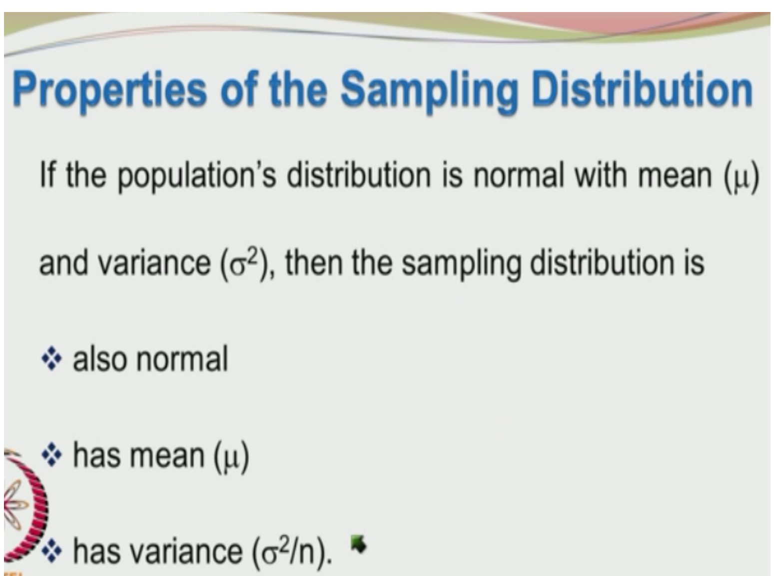
(Refer Slide Time: 30:57)



We know that the variance of x=sigma square, variance of x bar will be the variance of the spread of the distribution of the sampling means that will be=sigma square/n. So even if you have n random variables, you will have sigma square/n square because when you take variance of this quantity, it becomes 1/n square*variance of x1, the random variable x1, 1/n square*variance of the random variable x2+ so on to 1/n square*variance of xn.

So since all of these are identically distributed, you have sigma square, sigma square everywhere and so you have n sigma square/n square, which is sigma square/n. Hence the variance of the sampling distribution of the mean would be sigma square/n.

(Refer Slide Time: 32:02)



So if the population distribution is normal with mean mu and variance sigma square, then the sample distribution of the mean is also normal as the mean mu, as the parameter of the population itself and it has a variance sigma square/n. So it is opportunity for us to reflect a bit on this. Rather than taking these at their face value, what do they really tell us. We are making the assumption that the population is normal.

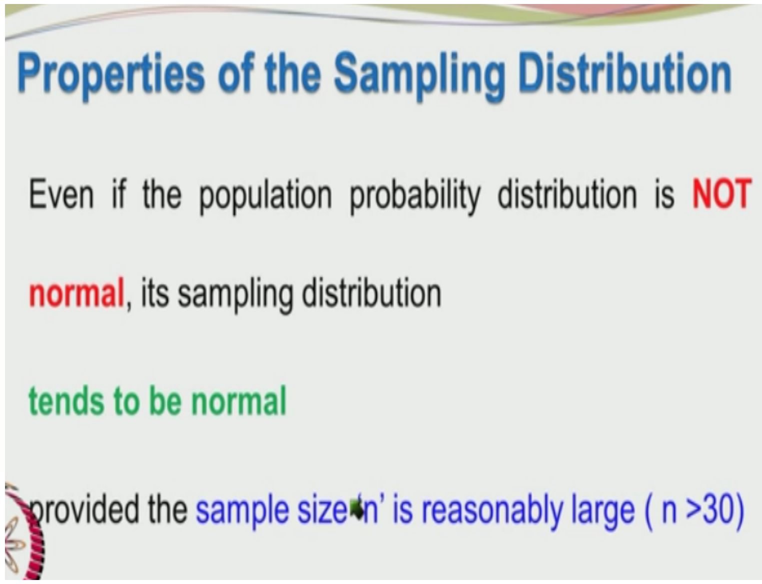
If there are 2 identical normal populations and we are taking x1 and x2, the random variables from the first population and the second population, then you form a linear combination, c1x1+c2x2. When you do that you also get a normal distribution. What are the parameters of such a distribution, resulting distribution is what we would like to know. What happens is, first we will assume that x1 and x2 are belonging to identical distributions.

They enjoy the same mean and same variance sigma square and they are also normal. When you combine them, you also have a normal distribution. This is very important to us. Next, when you combine x1 and x2 and then divide by 2, we get a mean. The sample is of size 2 and we get the mean based on the 2 random variables x1 and x2. Then, the sampling distribution of such samples of size 2 would be normal. It would have a mean mu and it would have a variance, sigma square/2.

So a variance of sigma square/2 is quite large for the distribution of the sampling means. Suppose you have taken n entities in each sample and you combine them to define the sample mean and you take several such samples, they will have a sampling distribution, which is also normal, because all the n random variables we have chosen came from identically distributed normal distributions and it would have mean mu.

The sampling distribution of the means would have a mean mu and it would have a variance sigma square/n.

(Refer Slide Time: 35:12)



So the next question to address at this point would be what would happen if the population from where the random samples were drawn is not normal. I will let the cat out of the bag even now by saying that if you have a large samples size, say n>30, then even if the parent population from where the random variables were chosen, where the random variables x1, x2 so on to xn were chosen, even if those were not from a normal distribution, the sampling distribution of the mean would tend to be normal.

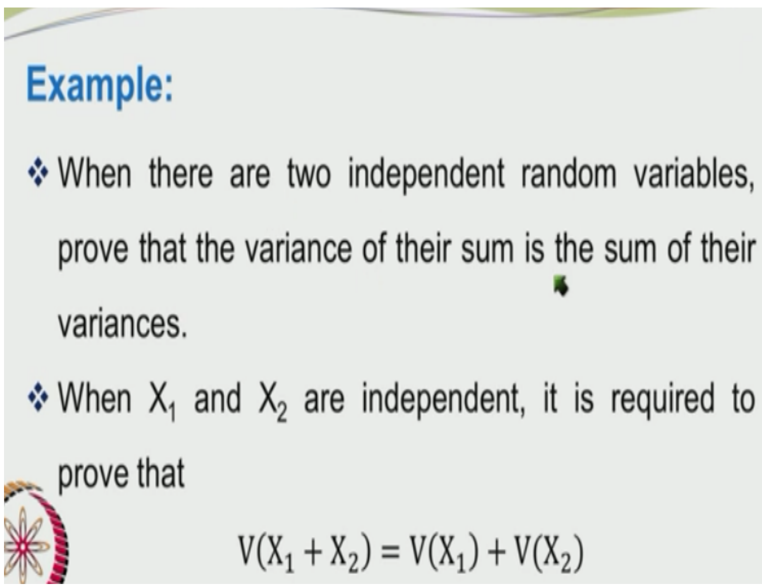
So this is very interesting and very useful. Because normal distribution is very well known and its properties are well tabulated. It is a simple distribution, nice symmetrical mean median=mode. The properties of the sampling distribution can be found from statistical tables. You can even use your spread sheet to find the probabilities. So it is very easy and we also are quite familiar with it. We know the bounds for mu+/-sigma.

We know the bounds for mu+/-2sigma. How much percentage of the population mu+/-2sigma will encompass, all these things are quite familiar to us. So the normal distribution is a very familiar and friendly distribution and very conveniently if you take an adequately large sample, then the distribution of the sampling means would tend to be normal. If I choose several samples from a population, each of size >30.

For example, all of them are having size of 35, even if the parent distribution was not normal. let us say it is gamma distribution or it is some other kind of distribution, some arbitrary distribution, but we are taking samples from such a distribution and those samples are of size >30. So let us say 35. Now each sample you have taken would have its own sample mean and it would have its own sample variance.

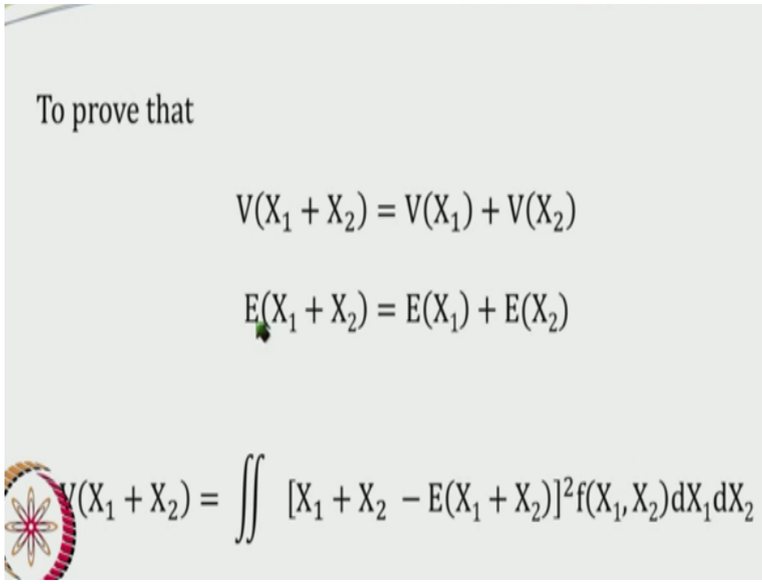
So there is a distribution of the sample means. The sample mean itself is a random variable. It is going to have a probability distribution. What is the probability distribution? if the sample size is > 30. If the sample size is >30 or it is a large sample, the sampling distribution of the mean tends towards normal behavior.

(Refer Slide Time: 38:34)



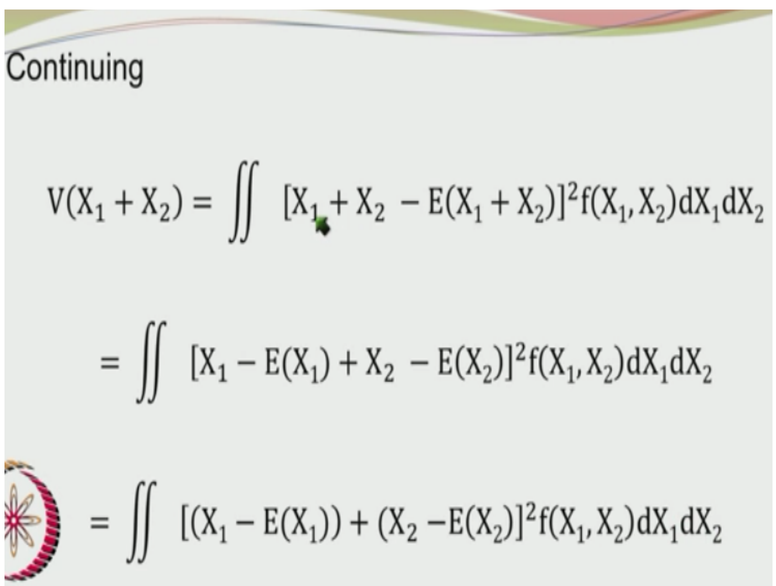
Let us look at a small example, if you are having 2 independent random variables, prove that the variance of their sum is the sum of their variances. So what we have to show is variance of x1+x2=variance of x1+x2.

(Refer Slide Time: 38:52)



The expected value of x1+x2, please note = expected value of x1+expected value of x2. Now going to the definition of the joint probability distribution function. We know that variance of x1+x2=x1+x2-expected value of (x1+x2) whole square F(x1, x2) dx1 dx2. I am talking about a general case first. Because we are having a combined joint distribution function.

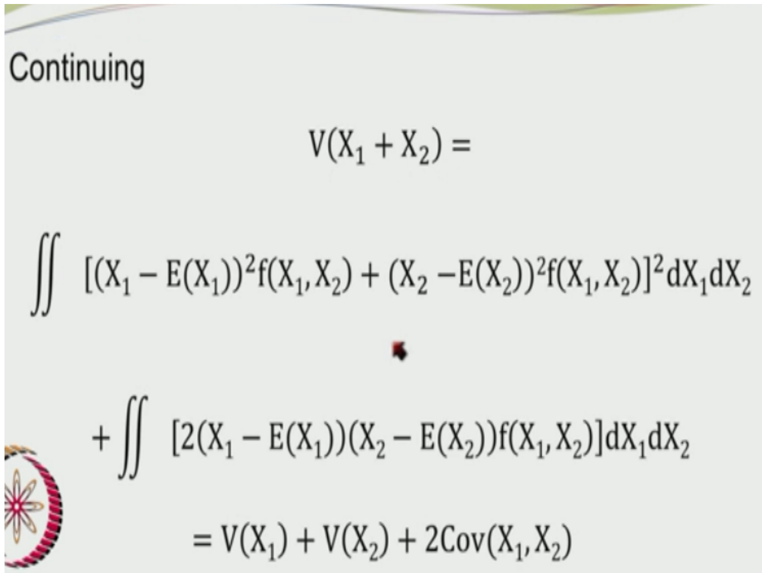
(Refer Slide Time: 39:35)



So what we do is we have to expand the term within the parenthesis, after taking the square, but before that we will collect the deviation terms. What I mean is, we will write it as x1-E(x1)+x2-E(x2) and then we will square that expression. We have x1+x2-E(x1+x2) square F(x1, x2) dx1 dx2. So we can write that as x1-E(x1)+x2-E(x2) and that is being squared to get F(x1, x2) dx1 dx2. So we have x1-E(x1)+x2(Ex2) whole square F(x1, x2) dx1 dx2.

The next job is to expand them. You can see that this will become x1-E(x1) square + 2*x1-E(x1)*x2-E(x2)+x2-e(x2) whole square and that we will multiply with these terms.

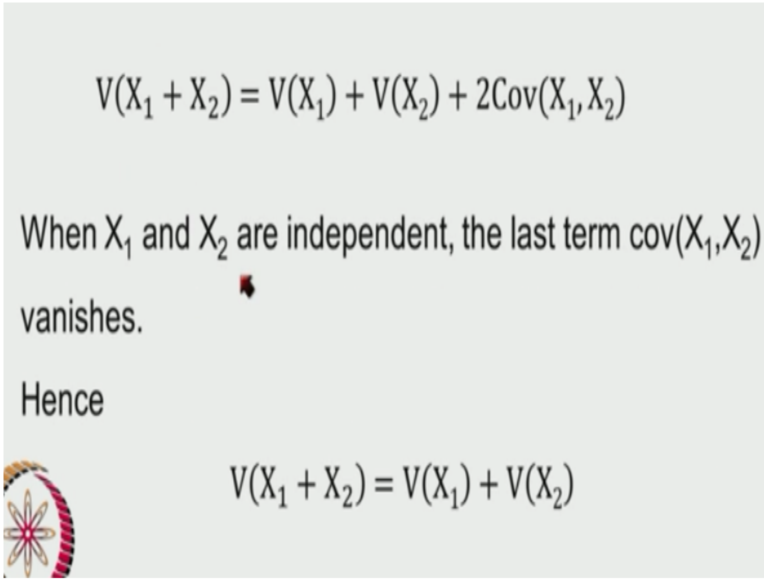
(Refer Slide Time: 41:01)



So this is the expression. I am splitting this into 2 terms, the 1 involving the square of the deviations with the probability distribution function, the joint probability distribution function. Similarly we have x2-E(x2) whole square F(x1, x2). There is a typo, I will just correct the typo. So you have these square of the deviation times F(x1, x2)+square of the deviation*F(x1, x2) and these terms should be now familiar to you.

Because they are the square of the deviation with respect to x1*the probability distribution function. Similarly this is the square of the deviation with respect to x2, then multiplied by the probability distribution function + the cross product terms of the deviations 2*x1-E(x1) * x2-E(x2) F(x1, x2) dx1 dx2. So we have this as the variance of x1+variance of x2 and this term is the covariance between x1 and x2.

(Refer Slide Time: 42:31)

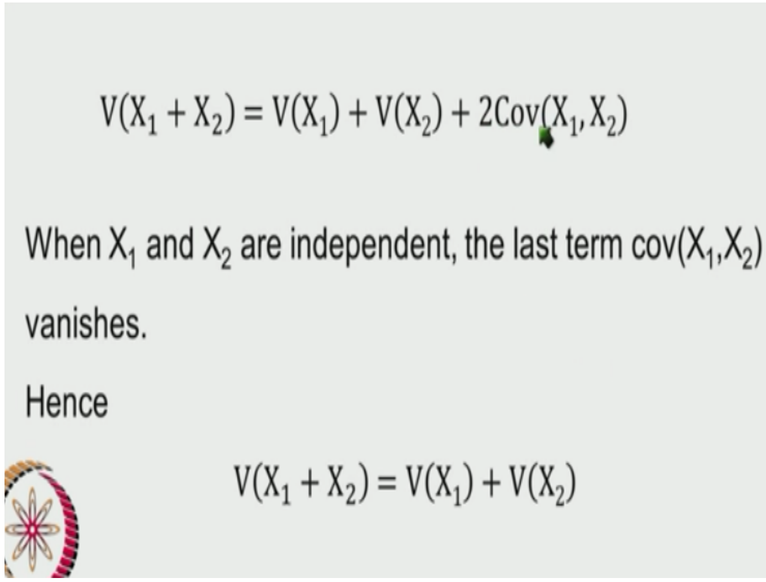


Then x1 and x2 are independent, the covariance term will vanish and you will get V(x1+x2) =V(x1) + V(x2). This is a very interesting result. Let us see what would have happened if we had variance of x1-x2. The immediate answer we may hastily write would be variance of x1-variance of x2. It does not somehow seem correct. If variance of x2 was higher than variance of x1, we will join function based on the combination of the 2 random variables or the difference of the 2 random variables, can they have a negative variance.

Actually, we have to go and do the mathematics properly rather than speculating what would be the sign of the resulting variance, so if you look at it, if you put variance of x1-x2, then it would be –x2-E(x1)-x2. So if you carry through with the mathematics, what you will find is, you will be finding variance of x1+variance of x2-2*the covariance of x1, x2. For example, if you are having a+b whole square and a-b whole square, both of them will have a square + b square term.

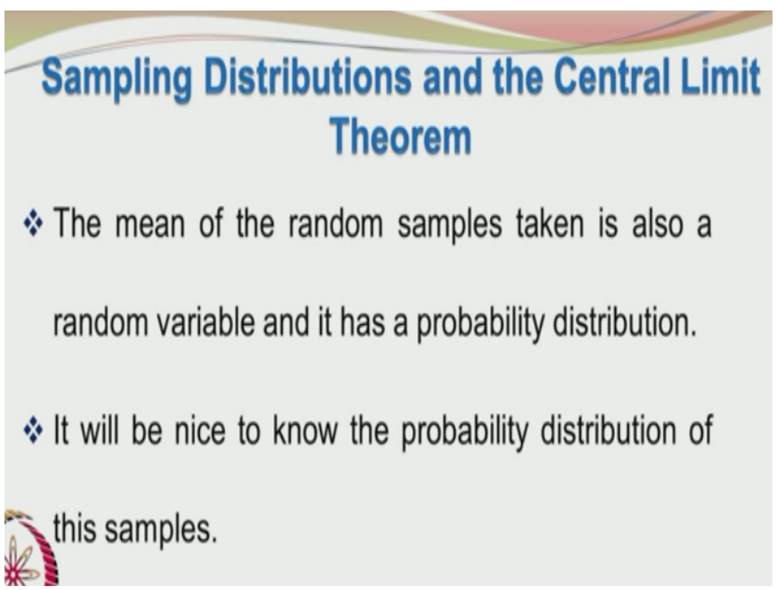
Only in the cross product term between a and b, you will have +2ab for a+b whole square and you will have -2ab for a-b whole square. So that negative sign depending upon the difference in the 2 variances, variance of x1-x2, the negative sign would actually arise come in the covariance coefficient. So the variance terms would be still having the positive sign relating them.

(Refer Slide Time: 45:11)



Anyway if the covariance vanishes, because x1 and x2 are independent, then we show that variance of x1+x2=V(x1) + V(x2), variance of the sum of 2 random variables=the variance of random variable 1+variance of random variable 2. If you had variance of x1-x2, then you still have the variance of x1+variance of x2 for the case where x1 and x2 are independent. For 2 independent random variables x1 and x2, the variance of their sum as well as the variance of their difference are both identical and they are given by variance of x1+variance of x2.

(Refer Slide Time: 46:15)



In the next part, we will be looking at the central limit theorem. We will continue after a small break.