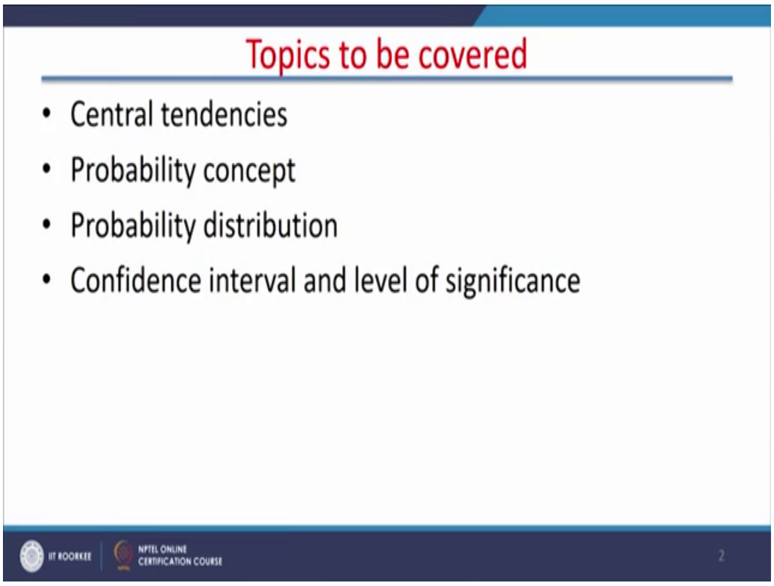
Mechanical Measurement Systems
Prof. Ravi Kumar
Department of Mechanical and Industrial Engineering
Indian Institute of Technology, Roorkee

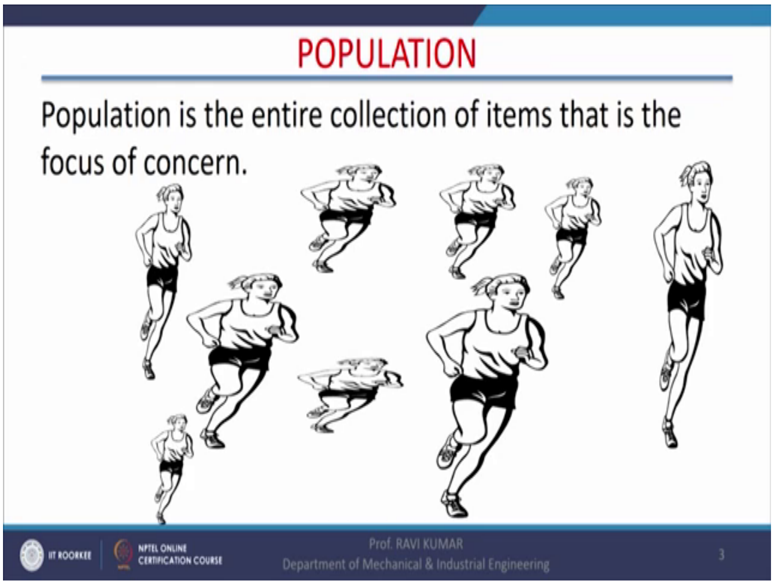
Lecture - 08
Statistical Analysis

I welcome you all in this course on mechanical measurement systems and today we will discuss the statistical analysis of experimental data. In this course, we will be covering central 10dencies probability concept probability distribution confidence interval and level of significance.

(Refer Slide Time: 00:41)

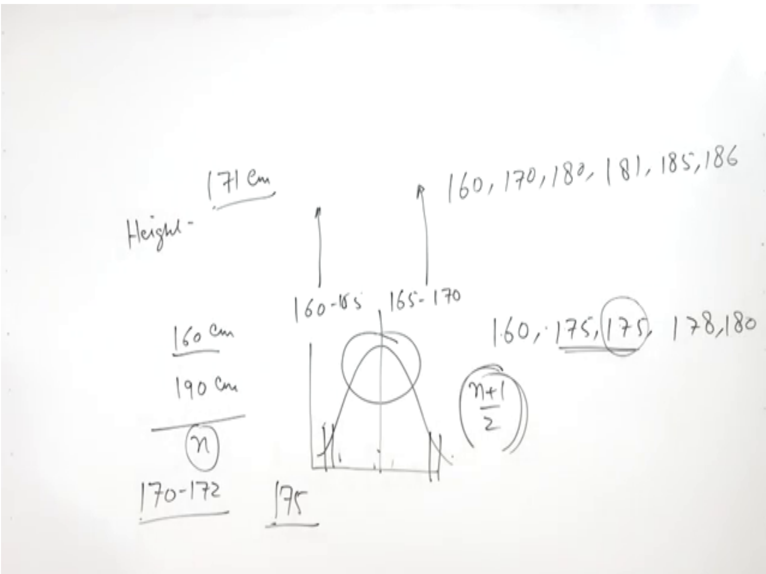


(Refer Slide Time: 00:46)



We will start with the central tendencies. Now, before going for the central 10dencies we will discuss the population. So, entire set of data population can be anything any sample of data can be a population. And population has certain characteristics for example, let us take athletes. So, athlete the characteristics of the athletes may be the height or shoe size right or many other characteristics.

(Refer Slide Time: 01:12)



So, out of these characteristics we just classify the population right height of the athletes. So, athletes the height of the athlete may vary let us say between 160 centimeter to 160 centimeters. And we have population of let’s let us say thousand athletes then what we are going to do? We will make the class intervals we will make the class interval let us say 160 165. How many athletes are there then 165 to 160? How many athletes are there and so, 4th and then we will have a diagram which is known as histogram which will depict the frequency of athletes for a particular height. This is how the scientific analysis of data a starts and in this process the first and the foremost thing which is being d1 is preparation of histogram. So, histogram gives us a fairly good idea about the population

Now, once, we have the histogram then we will like to further study the population in the sense that what is the average height of the athletes? So, in order to find average height of the athletes we will add all heights of the athletes and divided by n and then we will get average height of the athletes and average height of the athletes let us say it is 161 centimeters right

. Then we will like to know say how many athletes are there between 170 to 175 or 170 to 172-centimeter height. And how many athletes are very close to the average right. So, population distribution let us have because it is a infinite population. Let us this is the purpose suppose this is the population distribution

So, we can say we have bulk of the athletes are in this range right and this is known as or this is a part this type of part of analysis is known as central 10dencies. How close are the data towards the center?

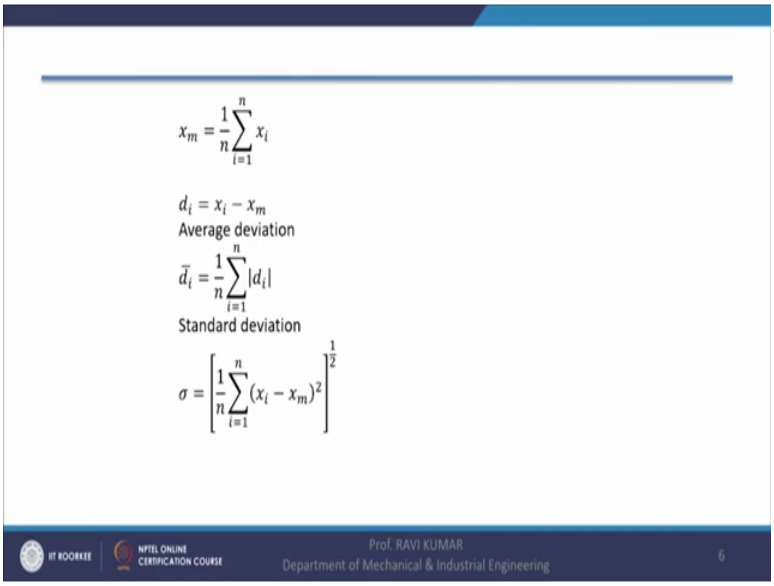
So, for central 10dencies also we should have some parameters right. So, parameters are first is mean take the average of these then another parameter is mod how frequently a particular data is appearing right for example, between 160 to 190. We have limited number of athletes at 160 limited number of athletes at 190, but we find many athletes who have 175 centimeter of height or suppose we have a sample of data. We have 160 175 175 185 180 .

So, out of these data this is the mod because it is appearing 2 times. So, this also shows the is a parameter of central 10dency another is median. Median is the middle 1 which is the middle 1. If we put athlete’s height of the athletes in ascending or descending order which is the middle term. So, middle term is always calculated as for odd number of data then it is n plus 1 by 2 or if it is even number of data then it is. So, here in this case for example, 1 2 3 4 5. So, it is going to 6 by 2 median is this right.

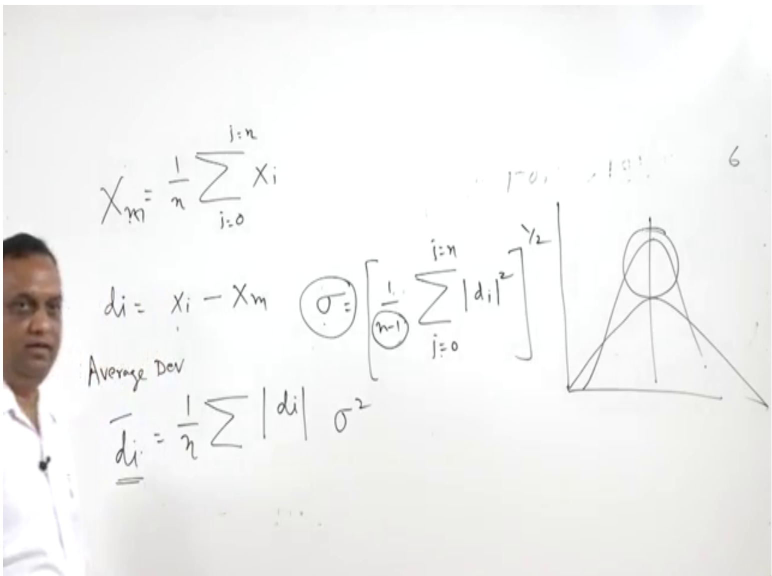
And now, for even number of data. Suppose, the data is 160 170 180 185 1 2 3 4 and let us say 184 185 1 2 3 4 5 right. In this case, the average of n plus 1 by 2 and n by 2 sorry n plus 1 by 2 and n by 2 will give the median.

Now, after central 10dencies is we will further analyze the average in details. Now if we want to depict the average.

(Refer Slide Time: 06:14)



(Refer Slide Time: 06:23)



Scientifically then average is x i average is equal to sum of x i i is equal to 0 to i is equal to n 1 by n. So, this is how statistically the average of a of given data can be expressed now deviation from the average now this average. So, all the data bellows at average value you will not find any data.

So, most of the data will deviate from the average. So, deviation from the average is depicted as di is equal to x i x i minus xm for the we will take this xm mean.

So, this is the deviation from the average and most of the time mod of the mod of the deviation from averages taking or this values taking without size for the statistical analysis. This we will discuss later on an average deviation now work is going to be the average deviation. The average deviation is going to be of di divided by n. So, all the deviations are added and it is divided by n.

Now, this shows now this shows how close are data to the center because if the data are away from the center data are like this. So, how close data are towards the center? If data are away from the center. For example, this in that case, this is going to be high right and when data are close to the center the average deviation is having the lower value. So, this is an indicator, right.

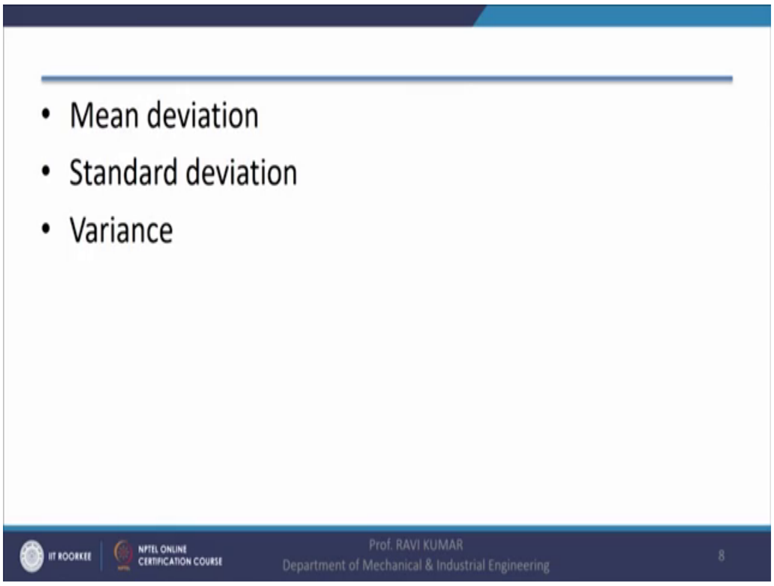
Now, if average deviation 0 in a when average deviation zero means all the data are same. They are lying on the mean and average deviation is zero means all the data are identical right. In addition to this there is a term standard deviation which is denoted by sigma and it is of10 used in statistical analysis.

Now, in case of sigma we take under root of I is equal to 0 to I is equal to n di square rise to power half or xi minus xm whole square it is added. So, this deviation is square of this a deviation is added divided by n n taken under root of that. So, this is known as the standard deviation and this is a very important parameter while doing these statistics statistical analysis of data, right.

Now, this is for the limited pop unlimited population when population is 10ding to infinity now out of this population if I select certain data certain sample. So, degree of freedom is checked right. So, now, in that case instead of taking n for a samples, we will be taking n minus 1.

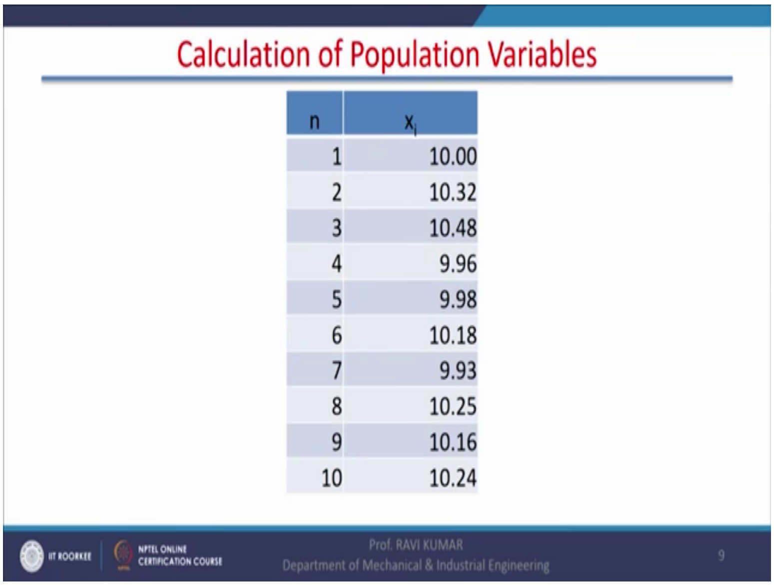
Suppose, it is a production line thousands of the parts are being manufactured and we take sample of 10 parts or sample of 20 parts it has to be less than 20. So, let us take sample of fifteen parts. So, safteen of sample of 10 or 15 parts and if you want to have a standard deviation in that case we will be taking n minus 1 not the n for the purpose of finding out the standard deviation.

(Refer Slide Time: 10:38)



So, so far, we have discussed the mean deviation a standard deviation and the variance. So, variance is nothing, but square of standard deviation. So, I have briefly introduce you with these terms with the understanding that you have already studied them in the subject mathematics.

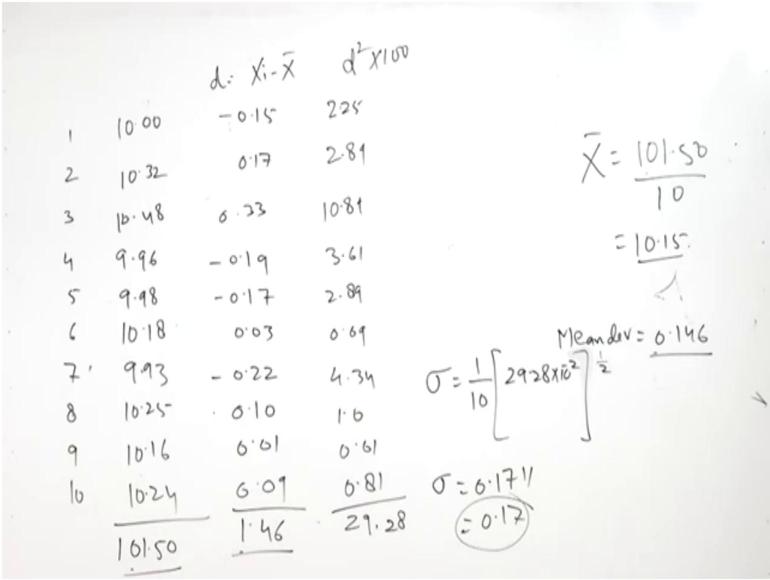
(Refer Slide Time: 11:02)



n	x_i
1	10.00
2	10.32
3	10.48
4	9.96
5	9.98
6	10.18
7	9.93
8	10.25
9	10.16
10	10.24

Now we will take 1 example when we take population variables, suppose they are 10 variables.

(Refer Slide Time: 11:06)



1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and these have values 10.00, 10.32, 10.48, 9.96, 9.98, 10.18, 9.93, 10.25, 10.16, 10.24.

Now, this can be supposed a reading of a shaft if suppose you are measuring the diameter of a shaft. So, 10 readings are taken or in the production line you have measured the diameter of the shaft 10 different shafts and the diameter of those shafts has been taken.

Now, I want to analyze these data now for the purpose of analysis first of all I will like to have mean of this what is the mean diameter of the shaft. So, mean for the purpose of finding out the mean diameter of the shaft we will simply add this. If you simply add this I have already d1 this calculation it is 100 and 1.50 right. So, your mean in this case is going to be and 101.50 divided by a number of observations this is or it is going to be 10.15.

Now, deviation from the mean d and d is deviation from the mean that is x bar minus x I. Now, x bar is 10.15. So, x bar minus x I is sorry it is x I minus x bar it is x I minus x bar or mean value. So, here it is going to be 0.15 here it is going to be 0.17 0.33 minus 0.19 minus 0.17 minus 0.03 this is plus;

Student: sir, second one.

This is 0.17, this is also plus this is also plus minus 0.22 0.10 0.01 and 0.01. So, 1 2 3 4 1 2 3 and 4 they are minus because they are less than the average rest are positive.

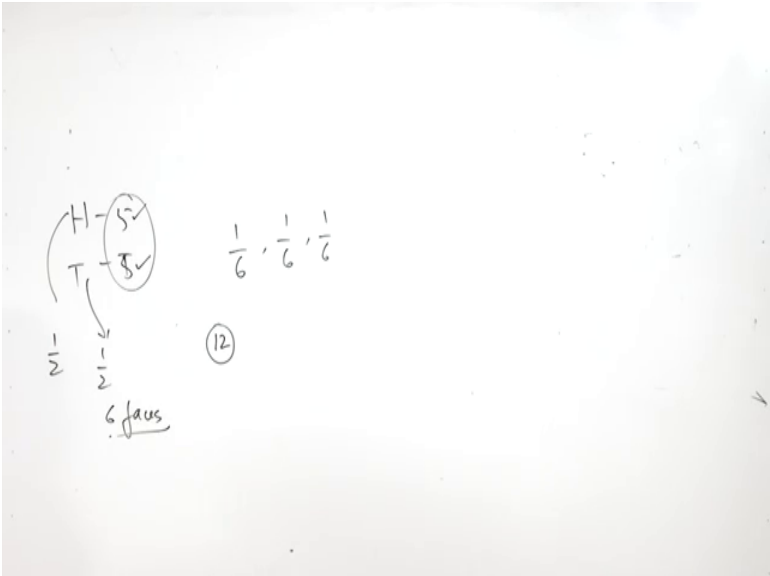
Now, when you want to have average deviation you will take mod of this irrespective of the sign. We will add all of them and we will add all of them and then we will get the value 1.46 and when we divide this 1.46. So, we are adding all these digits irrespective of the size. So, we have taken mod of this 1 point 4 6 again divided by number of samples. So, mean deviation is going to be equal to 0.146 ok.

Now, standard deviation for a for the purpose of standard deviation simply take square of d. So, when we take square signs will become irrelevant and we are going to get 2.25 multiplied by 100. Otherwise, we will have many decimal terms. So, I am just simply multiplying it by 100 it is 2.89 10.89 3.61 2.89 0.09. 4.34 1.0 0.01 0.81 and this addition gives 29.28 right and this is multiplied by 100 because it is multiplied by 100. So, 29.28.

Now, again now we have sigma of squares. So, if you want to have a standard deviation then we will have 1 by 10 sigma of squares. This is this is 29.28 into 10 to power minus 2 right. Divided by n n is 10 taking under root and the final value of a standard deviation is here is 0.1711 or we can take 0.17. So, this is how we can and this is the simplest example of or understanding of the average mean deviation and standard deviation.

Now, after this preliminary statistical analysis we will shift to the probability concept. Now probability is the likelihood of happening and even and suppose we are tossing a coin when we are tossing a coin there is a probability either heads will be there or tails will be there either it is going to be hand or it will going to be tail.

(Refer Slide Time: 17:20)



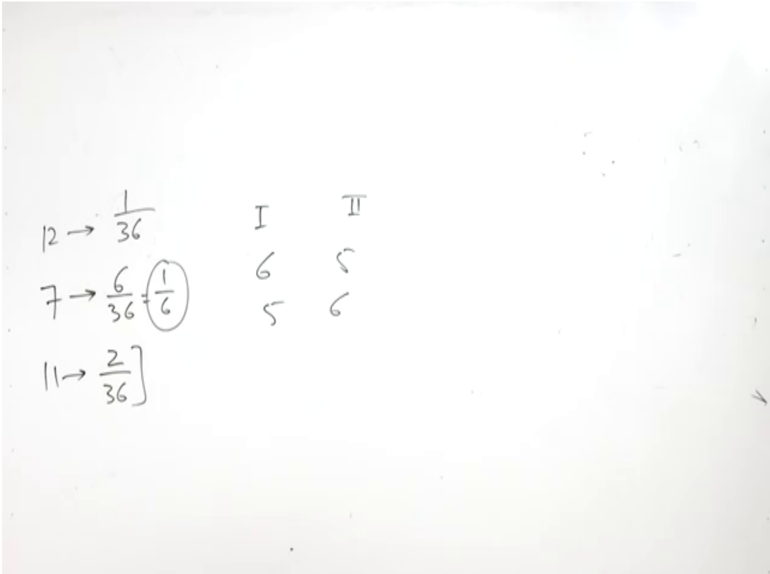
Probability is 50 percent probability is 50 percent the 50 percent chances are there it does not mean when your tossing 10 times 5 times you are getting heads and 5 times. You are getting tails if it is a bias unbiased coin I mean or w8less coin. Then if you are tossing coin in n number of times and n is tending to infinity, this is a very large sample.

You may get 50 percent of the time heads and 50 percent time of the tails, but not for the small sample let us say 5 tosses or 10 tosses. Right, but probability chances are there that head will come 50 percent of the time and tails will come 50 percent of the time this is the probability.

Now, for example, another example I will give you of the dice a dice has 6 phases and each phase is number 1 2 3 4 5 6. So, probability when you throw the dice probability of coming 1 is 1 by 6 probability of 2 is 1 by 6 probability of 3 1 by 6 and so on. So, all 6 digits have the same probability and if you add all these. So, then you will again get 1 because when you are throwing the dice you will be getting any of the point any of the number right. So, if you add all these probability 6 probabilities you will get 1 as in the case of tossing the coin.

Now, if you are throwing 2 co dices; now if you are throwing 2 dices you can have combinations up to 12. Now, I ask you; what is the probability of getting 12. So, for getting 12 for both the dices you should have 6 and 6 when you are throwing the dices then for both the dices you have you should have you should get 6 and 6 and there are 36 possible combinations.

(Refer Slide Time: 19:23)

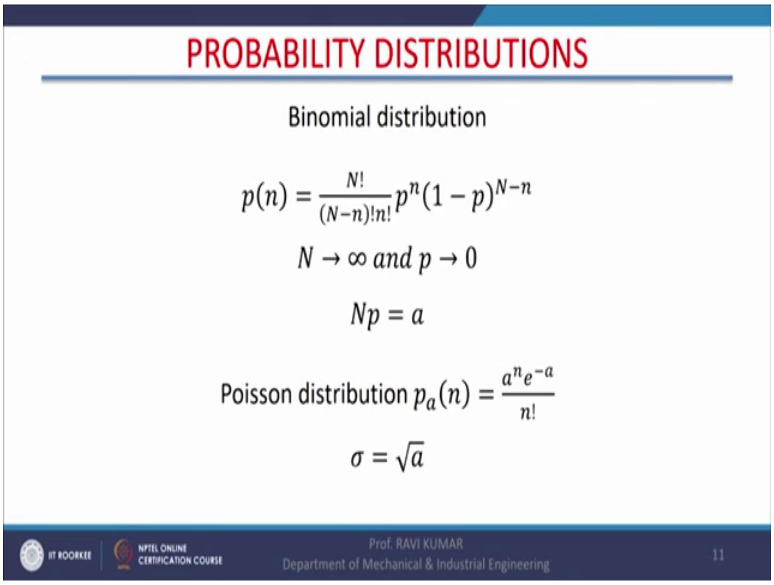


So, probability of getting 12 is 1 by 36. What is the probability of getting 7 the probability of getting 7 is higher than this I mean this is die 1 and this is a die 2. So, either you get 4 3 3 4 3 4 5 2 2 5 right and then 1 6 6 1. So, there are 5 probabilities. So, 5 by 30 6 there are 5 combinations out of the 6 where you can have the value 7. So, probability of 10 the 2 dices are thrown probability of getting 7 is greater than the probability of getting 12.

Now, probability of 11. So, when you are throwing dice the probability is now. So, there are 6 combinations there are 6 combinations which will produce the digit 7 right. So, we will have 6 by 36 or 1 by 6 the probability of getting 7 is 6 by 36. Now 11 I want to have 11 when I throw 2 dices then for getting 11 either it has to be 6 and 5 or 5 and 6 that is it.

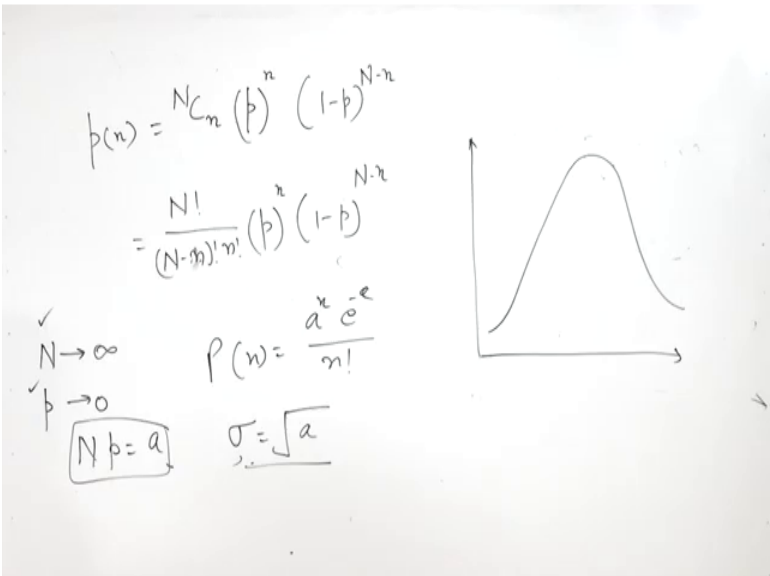
So, there are 2 combinations. So, the probability is 36 2 by 36. So, in case of dice if we are using 1 dice in that case for all the values the probability is 1 by 6, but if you are using 2 dice the probability will change and it will depend upon the number.

(Refer Slide Time: 21:20)



Now, probability distribution probability distribution the most pop popular probability distribution is binomial distribution. Now, binomial distribution; suppose there is a player and the player is player wants to put ring here at a distance x.

(Refer Slide Time: 21:39)



It is a player and he is throwing the ring and he is a very good player the player is champion is a very good player. So, what is going to happen most of the rings will fall close to this many of the rings will fall in this. If it is a good player most of the ring will fall in this many of the rings will fall in the vicinity of this right and a few rings will be far away from this object.

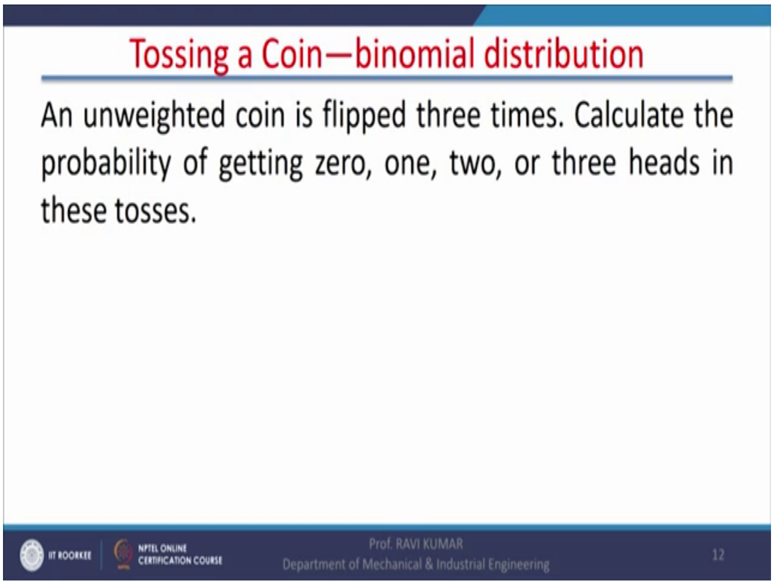
So, the population entire pop not entire population majority of the population will be concentrated on this or in the vicinity of this and their rear chances their rare chances that the ring is falling and sufficiently short of this or falling apart from this. Sufficiently, apart from because he is a very good player.

If we look at the probability distribution, if you draw the distribution of the success then we will be getting this curve if he does infinite number of throws. So, we will be getting a distribution which is known as binominical distribution in by analytical binominical binomial distribution. The probability of success of an event the probability of success of event is total number of event combination of total number of event and number of success. Probability of success raised to power number of success they are 2 different things probability of success and getting success and it is 1 minus p raise to power capital N minus 1.

Now, NCn as you know if it is a combination then it can always be taken as factorial n divided by factorial N minus small n N factorial small n. Then again pro probability of success raised to power N 1 minus probability of failure raised to power N minus N. when N 10ds to infinity when N 10ds to infinity if the player does infinite number trial 1000, 10,000, 1 lakh trials, right.

And probability of success is revert probability of success is revert it means the object is placed far off from the thrower and just he is throwing the rings he has given infinite attempts probability of getting success is bleak right. In that case, or there is a rider when n p is equal to a constant a. So, number of throws and the property probability their product is a, then it becomes Poisson distribution. And the Poisson distribution is again reflected as a raise to power n e raise to power minus a divided by factorial n right. We will take 1 example here to understand further understand the probability we will take 1 example. Now, 1 thing more in Poisson distribution the standard deviation is under root a this is the condition for this is how we can identify it is a Poisson distribution.

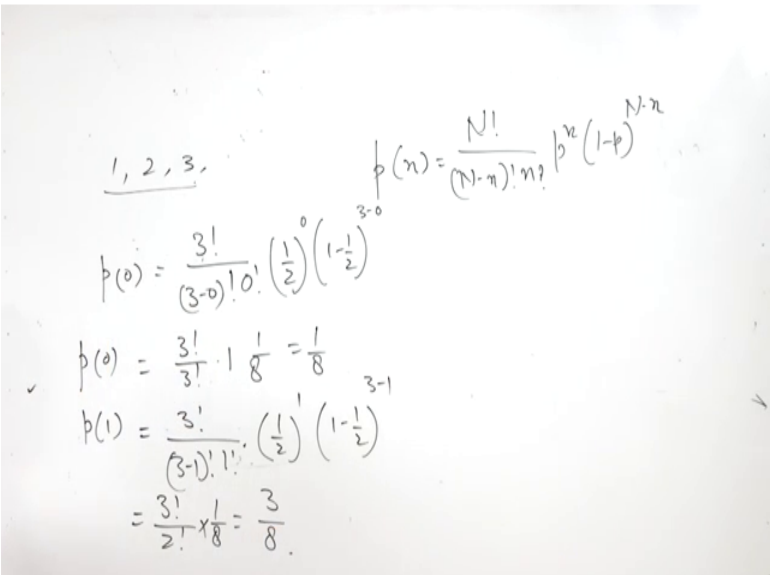
(Refer Slide Time: 25:31)



Now, in order to under for the understand this we will take the example of tossing of coin. Now, if we take 2 coins, if we take 2 coins, then the probability of getting head in 1 just a minute. Now, if you are tossing 2 coins the probability of getting head in 1 head 2 head 3 head and 4 heads. Let us confine to 3 heads only. We make 3 tosses right I will just show the textbook this is not given here.

So, if you make 3 tosses what is the probability of getting 1 head and 2 head and 3 heads I mean you are getting heads in all tosses or getting tails in all tosses. So, this can be simply dealt with this binomial theorem binomial distribution.

(Refer Slide Time: 26:38)



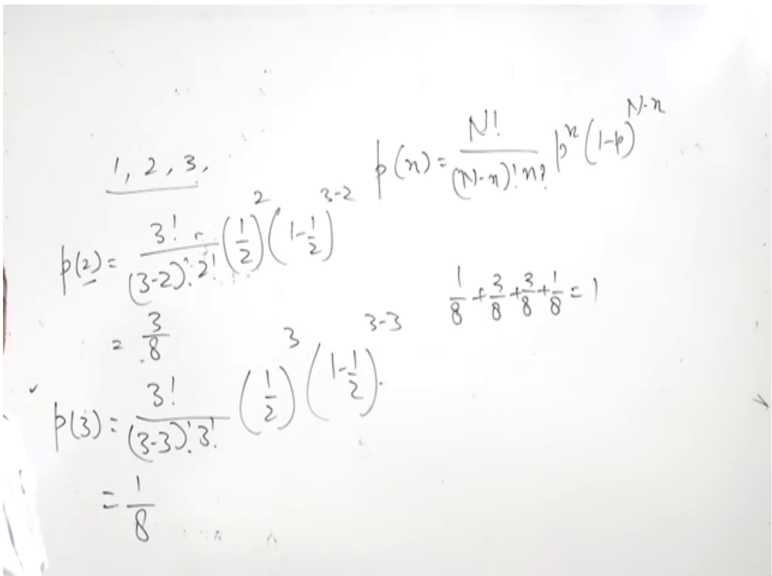
Factorial; so, probability of binomial distribution I will write again pn is equal to n divided by n minus small n factorial. Factorial n p raise to power n 1 minus p raise to power n minus n probability is 1 by 2. So, p is always 1 by 2.

So, getting 0 no heads at all is going to be n is 3. So, factorial 3 capital n minus n is equal to 3 minus 0 because success is 0 success is 0 factorial 0 half raised to power 0 1 minus half raised to power 3 minus 0. Now, here we simplify then we will be getting factorial 3 divided by factorial 3 this is 1 and this is 1 by 8 is equal to 1 by 8.

So, probability of getting 1 head out of 3 tosses is 1 by 8 probability of getting zero heads is 1 by 8. Now, probability of getting 1 head now probability of getting 1 head is again factorial 3 divided by 3 minus 1 factorial 1 multiplied by 1 by 2 raised to power 1 1 minus 1 by 2 raised to power 3 minus 1.

Now, in this case we will get factorial 3 divided by factorial 2 multiplied by 1 by 8 or it is going to be 3 by 8. So, getting 1 head is 3 by 8 similarly for getting 2 heads 2 heads is again factorial 3.

(Refer Slide Time: 28:35)

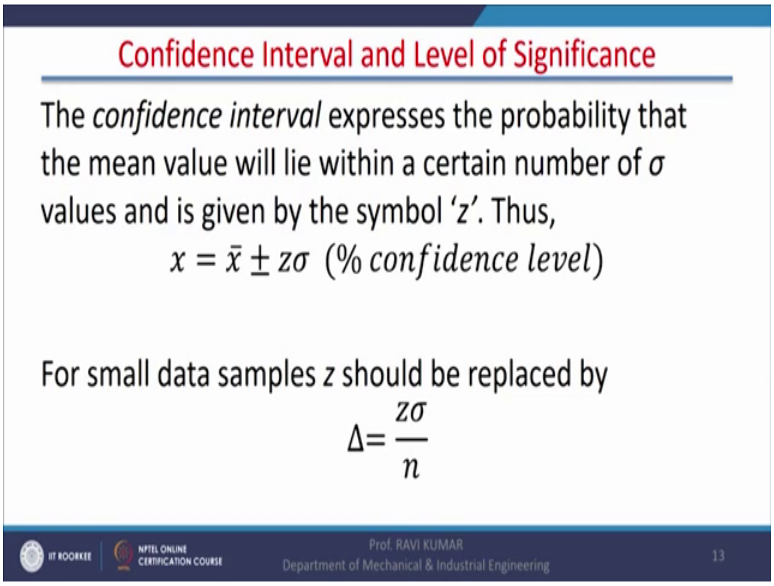


3 minus 2 factorial 2 factorial 1 by 2 raised to power 2 1 minus 1 by 2 3 minus 2, again you will find that it is 3 by 8. Now, again probability of getting 3 3 heads is again factorial 3 3 minus 3 all success. So, capital n is equal to small n 3 minus 3 factorial factorial 3 1 by 2 raised to power 3 1 minus 1 by 2 raised to power 3 minus 3. Now, here

also if you simplify you will be getting 1 by 8. So, probability of getting 1 head or 1 tail is 1 by 8 sorry zero head is 1 by 8 probability of getting 1 or 2 heads is 3 by 8 and probability of getting all 3 heads is. Again 1 by 8 and if you add all these because they are the total probabilities if you add all these it is going to be 1 by 8 plus 3 by 8 plus 3 by 8 plus 1 by 8 is equal to 1 because there is no other probability at all ok.

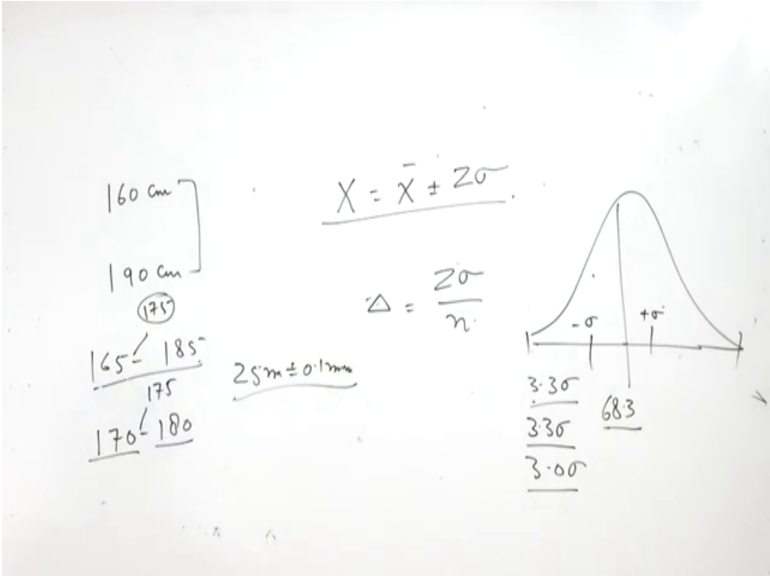
So, after this discussions on the topic on the binomial distribution.

(Refer Slide Time: 30:12)



We will discuss about the conscious confidence interval or confidence level. Now, confidence interval or confidence level is speaks about how confident about we are about the data. Now, sometimes it is misleading also. For example, let us take example of athletes. So, the ring is 1 60-centimeter height of athletes 2 1.

(Refer Slide Time: 30:42)



Let us say 1 90 centimeters this is the range of the height of athletes if I randomly pick the athletes if I say the height of the athlete is in the range of 165 to 185. Another option is height of the athlete is the range of 170 to 180, I will be more confident. In this case data may not be correct, but I will be more confident in this a case that this I am very confident that the height of the athlete is in this range in comparison to in this range.

So, high confidence level does not always reflect that the there is no error in suppose we are doing some measurement. So, order of error so height this confidence interval does not reflect the order of error, but here the things are different. We are randomly picking a data we are namely picking athlete and we if we have given a choice. So, in this case we are going to be more confident, but error in this case may be high error may be suppose the height of the athlete is 175. So, here error we are taking an error bind of plus minus 10 centimeters suppose height of the error is athlete is 175, we here we are taking error bind of plus minus 5 centimeters.

But confidence is low confidence is low, but significance of this data is more than this data right. So, we make number of observations and the average plus minus z sigma suppose we measure a diameter of a shaft 10 centimeters plus minus 0.1 centimeter. We write like this the normally we do not give the exact dimensions we say diameter shaft is 25 mm plus minus 0.1 mm.

Now, here we will also see that we can have we can improve the accuracy of measurement by taking repetitive observations. Now, for a small data sample suppose in the production line we pick certain data. So, for a small data sample z is reps sorry z is replaced by there is a term delta and delta is z sigma by n. Sigma is the deviation right, now what is z? Z simplices the percentage of confidence interval. So, there is a table.

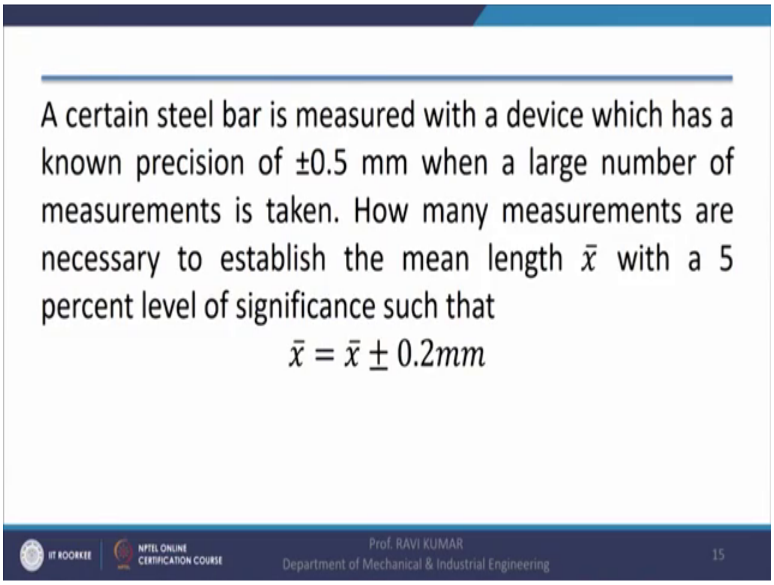
(Refer Slide Time: 33:47)

Confidence Interval	Confidence Level %	Level of Significance
3.30	99.9	0.1
3.0	99.7	0.3
2.57	99.0	1.0
2.0	95.4	4.6
1.96	95.0	5.0
1.64	90.0	10.0
1.0	68.3	31.7

There is a confidence interval confidence level in percentage level of significance.

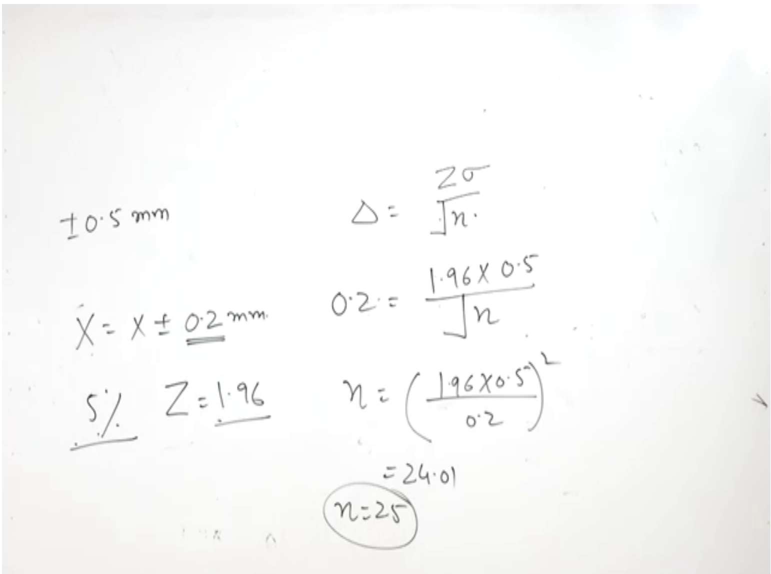
Now, here the confidence interval is 3 point 3 it means it means from the average if we are going either side 3-point 3 sigma sigma is the standard deviation. So, in that case 90.9 99 99.9 percent of the data will be covered or confidence level will be high. Suppose, this is the distribution this is the mean if you are going either side 3.36 sigma sorry 3.3 sigma 3.3 sigma if you are going on either side 99.9 percent data of we will be covered for the case of Poisson distribution and if we are going 3.0 sigma on both the sides. In that case 99.7 percent of data will be covered. So, it even is displayed on the screen and you can say when we are plus minus 1 sigma it is 68.3 when the sigma is plus minus sigma right.

(Refer Slide Time: 35:16)



Now, how this is relevant how this information is relevant? We will take one example. Now, this example says that a certain steel bar is measured with a device which has a known precision of 0.5 millimeter. Now, with these statistical methods, we can improve the precision also that will also be shown here. So, this numerical reads says states that a certain steel bar is measured with a device which has a known precision of plus minus 0.5 mm. So, precision is plus minus 0.5 mm, right when a large number of measurement is taken when the sample is of large number.

(Refer Slide Time: 35:53)



How many measurements are necessary to replenish establish the mean length x bar with the percentage level of significance such that x it is x bar is equal to sorry x is equal to x bar plus minus how much it is 0.2 mm right.

Now, for the solution of this the significance level is 5 percent. Now, for 5 percent significance level, we can see here for the 5 percent of significance level. It is confidence level is 95 percent and for 95 percent of con confidence level. The confidence interval is 1.96. So, z is this is confidence interval that is 1.96, right. And now, if we put this value here because it is a sample small sample; so, for a small sample we can always take delta as z sigma by under root n.

Now, [vocalized-noise here this delta is 0.02 sorry 0.2 z is 1.96 sigma is 0.5 deviation from the mean confidence level for 5 percent z is taken 1.96 divided by under root n and the n is here is we have to find n right. So, n is going to be equal to 1.96 into 0.5 divided by 0.2 whole square and if we calculate the final value of n it is going to be it is 24.01. So, it is 24.01 and then n has to be 25 n has to be 25. So, we have to take n 25 number of observations.

So, here this is this is a very in interesting conclusion that we have improved the precision of the measurement, right. If we have 25 number of observations right we can have the measurements in the range of this right; however, as stated in this problem the precision is plus minus 0.5 mm. So, we have by dating taking repeated measurements we have improved the precision of measurement. So, by increase or we can have more number of samples more number of samples and we can have we can we can further improve the precision of measurement right. So, this is how this is being d1 I think this is enough for today

Thank you very much.