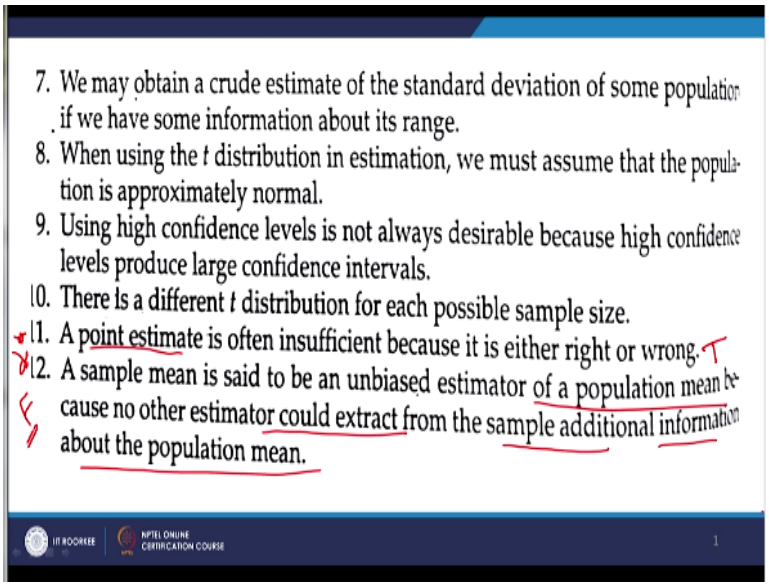
Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology-Roorkee

Lecture-29
Confidence Interval-II

I welcome you all in this session as you are aware in previous session we were working out couple of statements. And we have checked many of them work true and many of them work false, so we will continue the same exercise in this session as well.

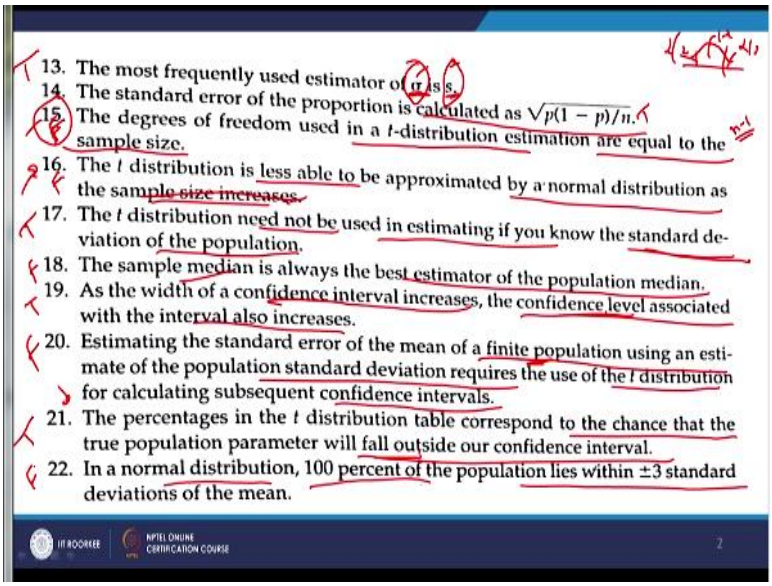
(Refer Slide Time: 00:44)



Let us look at once again the statement number 11, a point estimate is often insufficient because it is either right or wrong. And the answer to this statement was true, why it is true because point estimate either it may be right or it may be wrong. But if you view a range estimate now there is a possibility of having that right estimate more right rather than more false, so it is always good to have an interval estimate rather than point estimate for population parameter.

In fact I would like to discuss this statement once again, though we have discussed in previous session. A sample mean is set to be an unbiased estimator of a population mean, because no other estimator could extract from the sample additional information about the population mean. And we have said this statement is right but no this is not a true statement this is a false statement ok, so this is the correction you are suppose to make.

(Refer Slide Time: 02:00)



Let us move onto next one, the most frequently used estimator of sigma is S, so what is the sigma this population standard deviation and S is sample standard deviation. So, most of the times we use sample standard deviation to estimate population standard deviation, so this statement is true. Let us look at the next one the standard error of the proportion is calculated as this, is this the correct formula under root of p*1-p/n, what do you think is this correct one yes this is true right.

So statement number 14 is true, let us look at the next one 15th number, the degree of freedom used in a t distribution estimation are equal to the sample size is it so. The degree of freedom generally we calculate by looking at sample size or something else, it is n-1 right, so this statement is false 15 number statement is a false one right. Let us look at 16th number statement, the t distribution is less able to be approximated by a normal distribution as the sample size increases, what do you think, does t distribution become normal distribution with decreasing sample size, is this correct, no it is not correct.

In fact the other way round is true, in fact when the sample size increases t distribution approaches normally right, so this is false right. Let us look at 17th number, the t distribution may not be used in estimating if you know the standard deviation of population is it, when do we

use t distribution. In fact we have talked about 2 conditions, t distribution is applicable when standard deviation of population is unknown as well as the sample size is 30 or less than 30 right.

Otherwise we will use z distribution, so this statement is t distribution need not be used if you know standard deviation of population yes, so this is true statement right ok. Let us look at the next one, the sample medium is always the best estimator of population median. In fact we have talked about estimation of population mean from sample mean we have seen estimation of population proportion through sample proportion as well.

But there is nothing like sample median or sample median is estimator population median right, so this is false statement right. As the width of confidence interval increases the confidence level associated with the interval also increases, is this statement correct, let us think over it. As the width of confidence interval increases the confidence level associated with the interval also increases, yes this statement is true right.

Again let me tell you that confidence level is the probability within which the population parameter will fall within certain limits. In confidence interval is the range or confidence interval is basically a range within which population parameter will fall right. So these two go hand in hand, so when confidence level is more confidence interval would be higher right. But in real life situation whenever you estimate something whenever you go for let us say interval estimation try to have interval estimation.

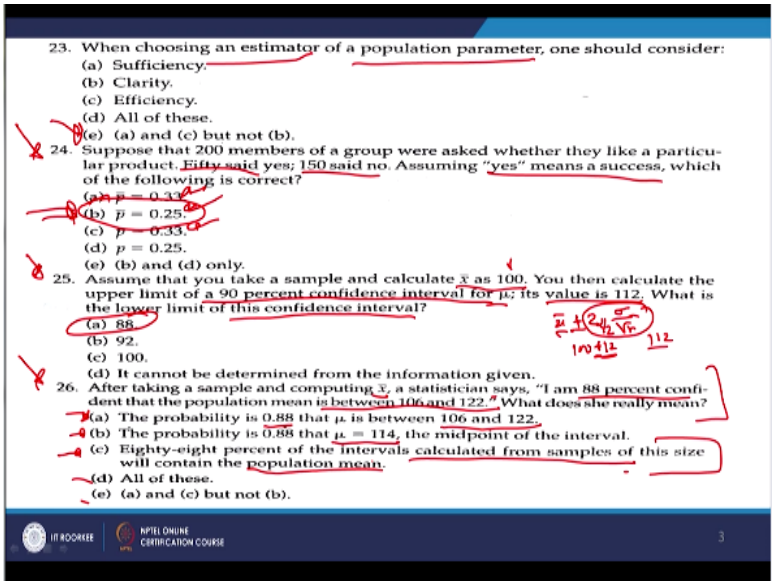
In such a way that it is range should be small and confidence level should be very high ok. So let us look at the next one estimating the standard error of mean of a finite population using an estimate of population standard deviation requires the use of the t distribution for calculating subsequent confidence interval. Though I did not talk much about this finite population parameter, so for the time being we can skip this but anyway this statement is false right.

Let us look at the next one, the percentages in t distribution table corresponds to the chance that the true population parameter will fall outside our confidence interval, what do we do in t distribution. Generally we are we look at the this is that alpha/2, this is alpha/2 and this whole

area 1-alpha. So, in t table we have seen the values of t or critical value of t at a particular degree of freedom and that generally gives you the population parameter which is outside the confidence interval right, so this statement is true.

In a normal distribution 100% of the population lies within ±3 standard deviations of the mean, is this correct. We have seen properties of normal distribution it is not 100% population it is 99.7372 right, but it is not 100%, so this statement is false.

(Refer Slide Time: 08:54)



Let us look at the next one, when choosing an estimator of population parameter one should consider sufficiency, clarity, efficiency or all these, sufficiency, clarity, efficiency, all of these a and c but b, so the answer to this is a and c but not b. So, we want sufficiency and efficiency of an estimator of sample estimator for estimating population parameter. Let us look at 24 this good one suppose the 200 members of a group were asked whether they like a particular product 50 said yes, 150 said no.

Assuming yes means a success which of the following is correct, p bar 33 or 0.33, p bar 0.25, p 0.33 or p 0.25. So the correct answer for this is the b right which is 25%, so p bar is 0.25 right, so answer to this is b ok. Let us look at the next one assume that you take a sample and calculate sample mean as 100, you then calculate the upper limit of 90% confidence interval for population mean it is value is 112, what is the lower limit of this confidence interval.

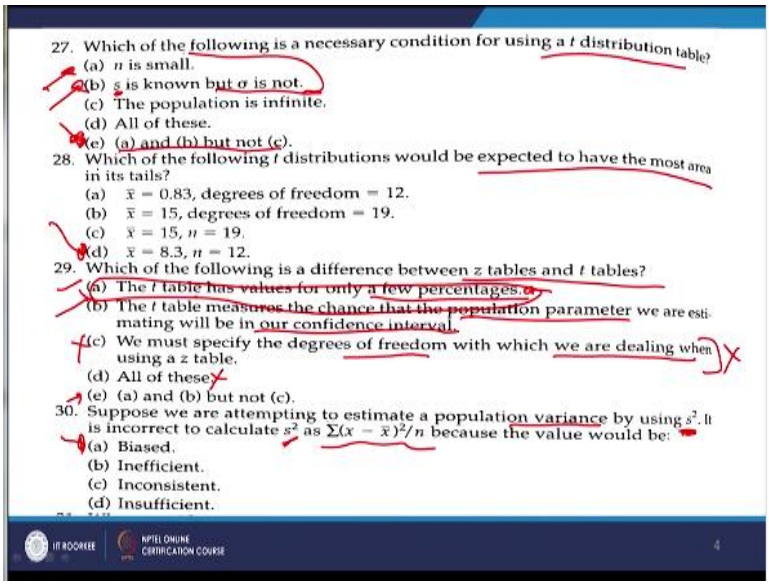
So, we know that the confidence interval is mean +/- +/- what Z* let us say alpha/2 sigma to the power divided by root n right. So, since we know that this x bar is 100, so and it is upper limit is 112 right it means this value is 12 right, this entire value is 12. So, it is upper limit is 112, 100+12 112, so the 100-12 would be 88, so this how you should be getting answer to this particular question.

Let us look at into 26th number question after taking a sample and computing mean a statistician says I am 88% confident that the population mean is between 106 to 10 sorry 122, what does he really mean. Again after taking a sample and computing sample mean a statistician says I am 88% confident that the population mean is between 106 to 122, what does she really mean, the probability is 0.88 that population mean is between this, does it mean this a part or does it mean b part.

The probability is 0.88 that mean=154, 114 the midpoint of the interval or c, 88% of the intervals calculated from samples of this size will contain population mean, is this statement true. For this particular the 1 statement all of these are e, so we have said that it is the 88% of the intervals calculated from samples of this size will contain population mean is not necessary that this sample will contain population mean.

Or the next sample which you take up it is not necessary that would also be between a population that would also be in interval in between interval. So, it is the meaning is 88% of the times fractional 80% of the times the confidence interval forms will have the mean right and this is true for population mean as well as for population proportion, both the cases, which of the following is an necessary condition for using t distribution.

(Refer Slide Time: 14:20)



So, we have said that there are 2 conditions, first is the population standard deviation is unknown and the second one is sample size is 30 or less than 30. So, let us look at this n is small is this necessary condition, S is known but not standard deviation, S is standard deviation of sample is known but not of population. The population is infinite, all of these a and b but not c, what do you think, so answer is a and b.

Sample size is small which is 30 or less than 30 and standard deviation of sample is known not the population right, so correct answer to question number 27 is c right. Let us look at the next one which of the following t distribution would be expected to have the most area in it is tails. Generally area in tails will be for a situation where mean is high and degrees of freedom is less, so in this particular question answer number or the part d is the answer to question number 28.

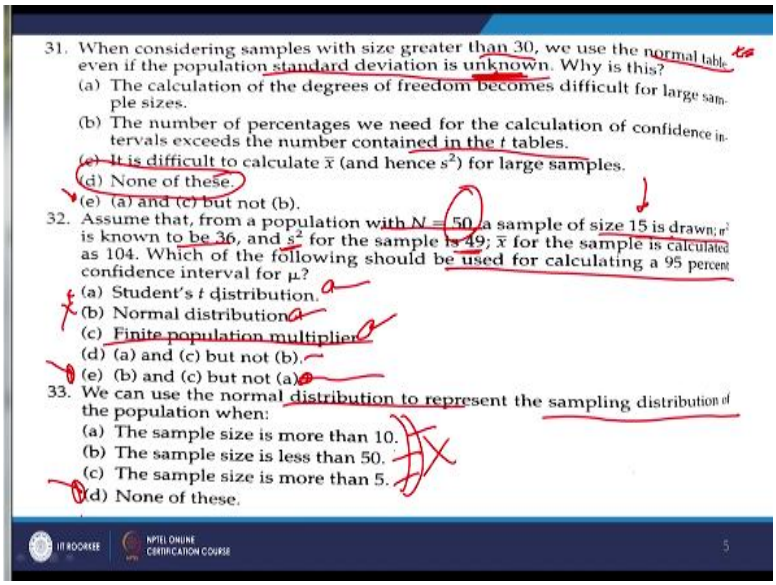
Let us look at the next one which of the following is the difference between Z table and t table let us look at these differences. The t table has values for only a few percentages, the t table measures the chance that the population parameter we are estimating will be in our confidence interval. We must specify the degrees of freedom with which we are dealing when using Z table in fact c part cannot be correct right because degrees of freedom goes with t table right.

So, you can just exclude c part from your answer list, all of these, no all of these cannot be the answer because c is not the answer. So, this is also gone from the list of possible answers right,

so you are left with a, b, a option, b option and e option, so the t value has a the t table has values for only few percentages, so for 29th this is the answer here is the correct answer right this is the answer right a part ok, by in fact b part is not correct right.

Let us look at question number 30th, suppose we are attempting to estimate a population variance though we did not calculate any question on estimation of population variance from sample variance by using sample variance it is incorrect to calculate sample variance as this because the value would be though we did not do any question on this. But that this the value would be biased right it has to be n-1 right.

(Refer Slide Time: 18:26)



Let us move onto next question when considering samples with size greater than 30, we use normal table even if the population standard deviation is unknown why. Once again let me tell you when do we use t table when standard deviation of population is unknown and when sample size is 30 or less than 30. In all other situations we will use that table, so just remember these 2 conditions.

So, when considering sample size greater than 30 we use normal table even if the population standard deviation is unknown, why is this the calculation of the degrees of freedom becomes difficult for large sample size is that the reason, no that is not the reason. The number of

percentages we need for the calculation of confidence interval exceeds the number contained in t table c part.

It is difficult to calculate mean for large sample, none of these and e option is a and c but not b, so the answer to this is none of these ok. So, let us move onto next question assume that from a population with sample size 50, a sample of size 15 is drawn, so how much percentage is this approximately 30% right sample size is 30% of the population size. So, assume that from a population mean with n=50 a sample of size 15 is drawn standard deviation of population is known which is 36.

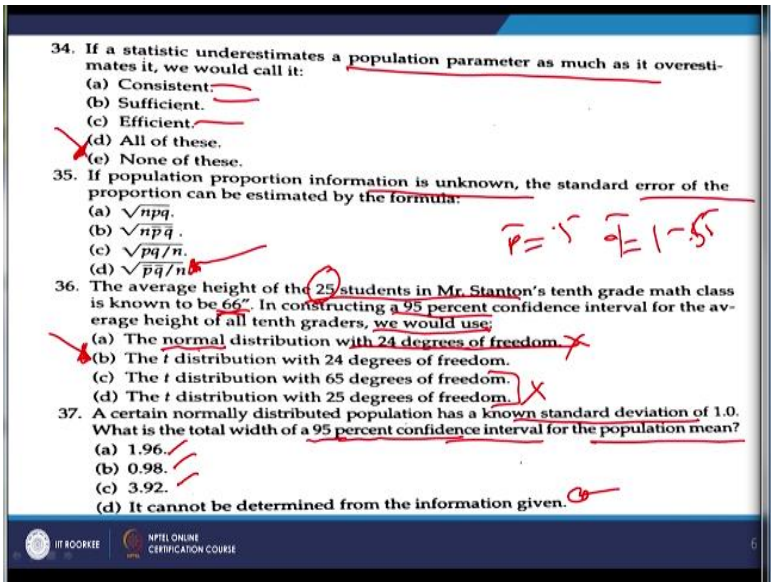
And standard deviation for the sample is 49, sample mean for the sample is calculated as this. So, when we took a sample size of 15 come a population of 50 it is variance was this standard variance of sample is this and sample mean is this which of the following should be used for calculating a 95% confidence interval for mu, which of the following should be used for calculating a 95% confidence interval for mean.

The student's t distribution, normal distribution, finite population multiplier a and c**,** but not b and b and c but not a ok. So, the answer for this is e, so b and c but not a, so b and c is normal distribution why normal distribution because you have been given what standard deviation of the population. And what is that value is 6 because variance is 36, so standard deviation would be 6, so will be using normal distribution, finite population multiplier is needed.

Because this sample size is approximately 30% of population, so generally we use finite population multiplier if it is sample size is 10% or more than 10% of the population right, so the answer to question number 32nd is e ok. Let us move onto next one we can use the normal distribution to represent the sampling distribution of the population when do we use normal distribution.

Sample size more than 10, less than 50, more than 5, none of these, we can use normal distribution to represent sampling distribution of the population when in fact these are just few statements written over here to confuse you, none of these is the answer right.

(Refer Slide Time: 23:34)



Let us look at next one if a statistic underestimates a population parameter as such as it overestimates it. We conclude or we would call it consistent, sufficient, efficient, if a statistic underestimate population parameter as much as it overestimates it, so it is none of these, the answer is none of these. In fact you know if a statistic underestimate a population parameter as may face it is overestimates then you would get the correct population parameter, that is the property of getting correct population parameter.

Let us look at 35th one if population proportion information is unknown, the standard error of the proportion can be estimated by the formula. If population proportion information is unknown, so in situations like this what we take, we take 8+0.5 right, the standard error or proportion can be estimated by this formula and the answer is d right, so we will assume p=0.5 and q would be 1-0.5 right 0.5 right.

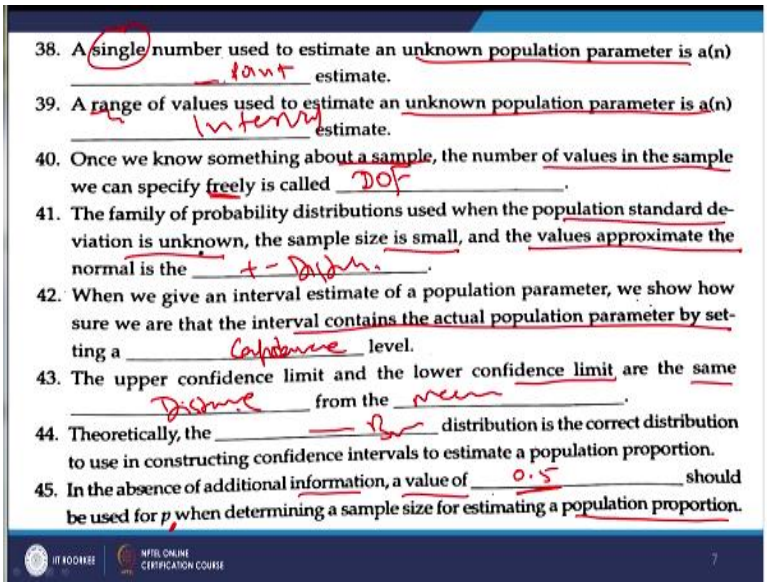
Let us look at next question the average height of 25 students in Mr. Stanton's 10 grade math class is known to be 66 inches average height is given constructing a 95% confidence interval for the average height of all 10 grade we would be using what, what we would use. The normal distribution with 24 degrees of freedom you see normal distribution is are Z distribution. So, this cannot be correct because degrees of freedom goes with t distribution right, t distribution with 24 degrees of freedom.

So what is the sample size 25, so n-1 is 24, so this answer is correct, in fact out of all these 4 options you can easily identify a right answer because when n is 25 you will have to have 24 degrees of freedom. So, you can easily skip these 2 options from your answer list only the option left would be a and b, a cannot be because degrees of freedom is associated with t distribution, so only b is the right answer right, so this how you can answer some of the questions.

A certain normally distributed population has a known standard deviation of 1, so there is a normally distributed population standard deviation is 1, mean is not given right, what is the total width of a 95% confidence interval for the population mean 1.96, 0.98, 3.92 and nothing. So, it cannot be determine from this information why because we have not been given mean right ok, so what is the total width of the of 95% confidence interval.

Because we have not been given population mean, so either you will have to have population mean or you will have to have sample mean right, then only you can have confidence interval.

(Refer Slide Time: 28:01)



Let us look at this slide, the last one the single number use to estimate an unknown population parameter is a dash estimate. A single number use to estimate an unknown population parameter is point estimate. A range of values use to estimate a range of values right not a single number, a range of values use to estimate an unknown population parameter is interval estimate.

Let us look at the next one once we know something about sample the number of values in which sample the number of values in the sample we can specify freely is called degrees of freedom. Next one the family of probability distribution used when population standard deviation is unknown, the sample size is small and the values approximate the normal is called what student's t distribution or just t distribution.

A family of probability distribution used when the population standard deviation is unknown, so we use simple standard deviation the sample size is small and the values approximate the normal is t distribution right. Let us look at the next one when we give an interval estimate of a population parameter we show how sure we are get the interval content the actual population parameter by setting a what confidence level confidence level ok.

The upper confidence limit and the lower confidence limit are the same dash from dash are the same distance from the mean right. Theoretically dash distribution is the correct distribution to use in constructing confidence interval to estimate a population proportion so this binomial. Let us look at the last one last question in the absence of additional information a value of dash should be used for p when determining a sample size for estimating a population proportion 0.5 should be used.

So, in today's session we have worked out couple of examples and I hope that with these couple of questions the fundamentals of estimation would have more known to you, will have next session on hypothesis testing, thank you very much.