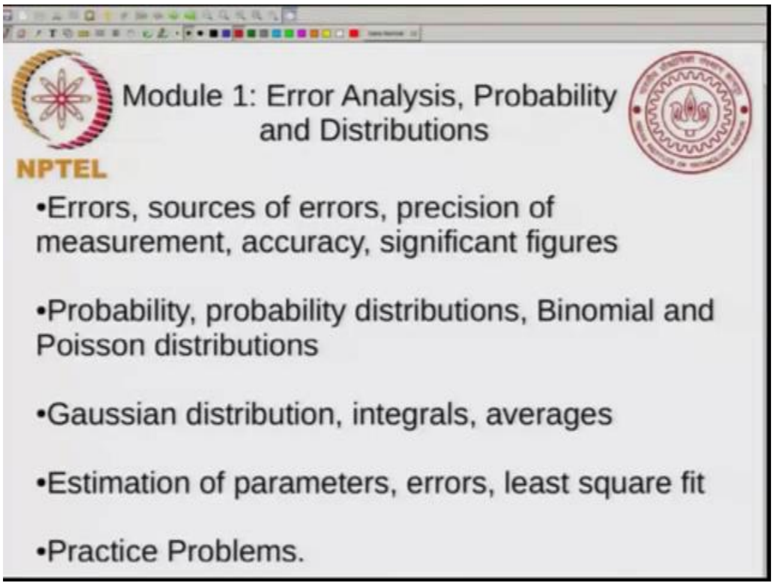
Mathematics for Chemistry
Prof. Madhav Ranganathan
Department of Chemistry
Indian Institute of Technology, Kanpur

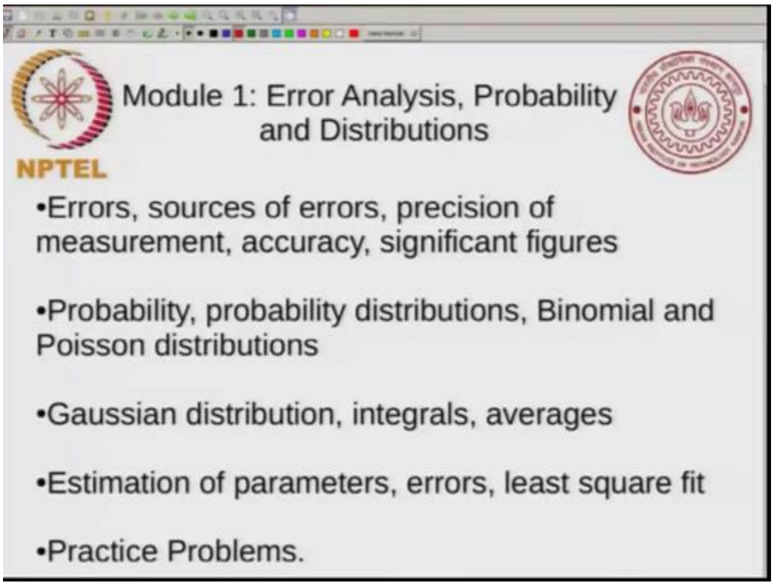
Module - 01
Lecture – 01
Errors, sources of errors, precision of measurement, accuracy, significant figures

(Refer Slide Time: 00:17)



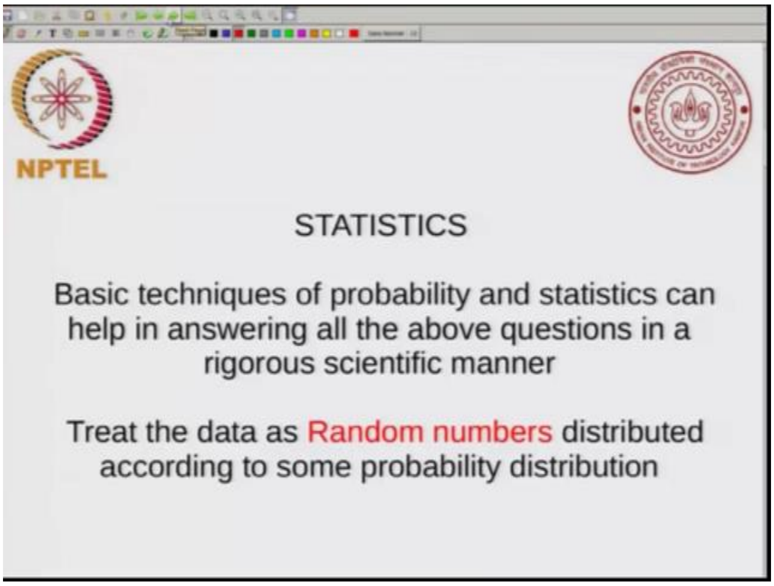
In the first part of this course, we will be talking about error analysis, probability and distributions. Now in this course we will talk about errors, the sources of errors, precision of measurement, accuracy and significant figures. Then I will talk about probability, probability distributions, the binomial and Poisson distributions. Then I will talk about Gaussian distributions, integrals involving Gaussian distributions and averages involving Gaussian distributions. Then finally, I will talk about estimation of parameters, errors and least square fit method and we will end with some practice problems. So, let us go to the first part which is errors, sources of errors, precision measurement etcetera.

(Refer Slide Time: 00:58)



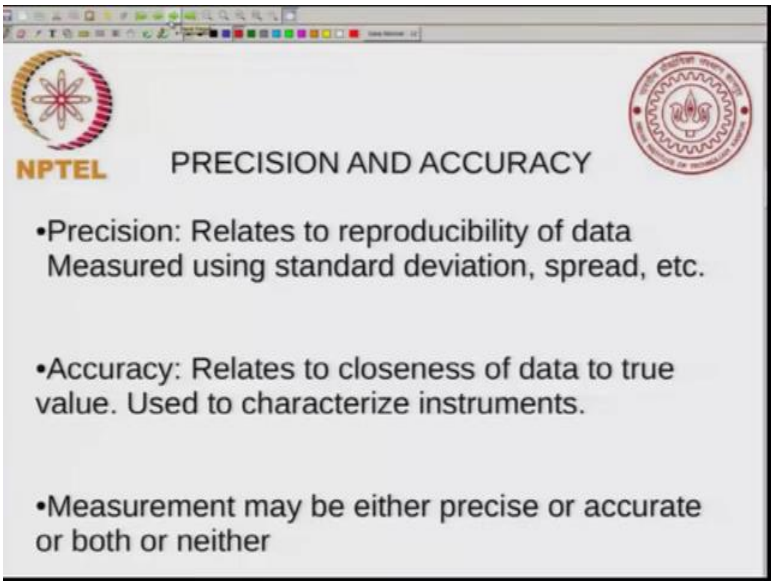
So, here you consider the following. Say, you are given some experimental data and you want decide the quality of the data. You want to say something about how good or how accurate or how precise the data is. Alternatively, you might have several experiments that all calculate some quantity and based on all those experiments you are required to report value. A third case might have, you might have a large amount of frequency data and you want calculate statistics and find and another case you can have is several data points for a dependent variable against an independent variable and you may be required to fit a graph through that data. In all these cases you will encounter the use of statistics.

(Refer Slide Time: 01:55)



So, you will be using basic techniques of probability and statistics to answer all the questions about in a rigorous scientific manner. And essentially underlying all these methods is that you are treating the data as random numbers distributed according to some probability distribution and over the next 2 and half hours you will see how we think of these random numbers and how probability distributions are involved.

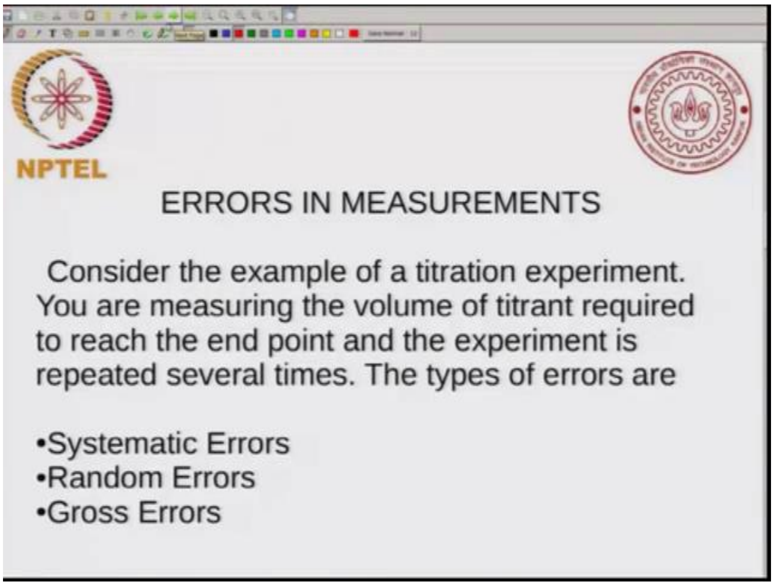
(Refer Slide Time: 02:24)



So, the first 2 terms that I want to define are precision and accuracy. Precision relates to reproducibility of data. So, your data is set to be precise if all the measurements give you approximately the same value. It says nothing about the accuracy; the accuracy on the other hand is related to the closeness of the data to the true value. So, if you make a measurement you might have a measurement that is very precise, but not accurate. Accuracy is related to how close the experimental data is to the true value. So, if you know the true value then you can tell the accuracy. So, accuracy typically is used to characterize instruments. So, if there is an instrument then you make measurements on that instrument and then you say that this is the accuracy of that instrument. You take some standard measurements and you compare those measurements with what is observed in the experiments and you characterize accuracy.

However, on the other hand precision is related to the reproducibility of data and this can be measured just using the data. So, just using the data you can tell how precise it is by calculating things like standard deviation and square. And again I am emphasize at a measurement may be either precise or accurate or both or neither and all things are possible. In this part of the course we will be mostly interested in quantifying the precision of the data.

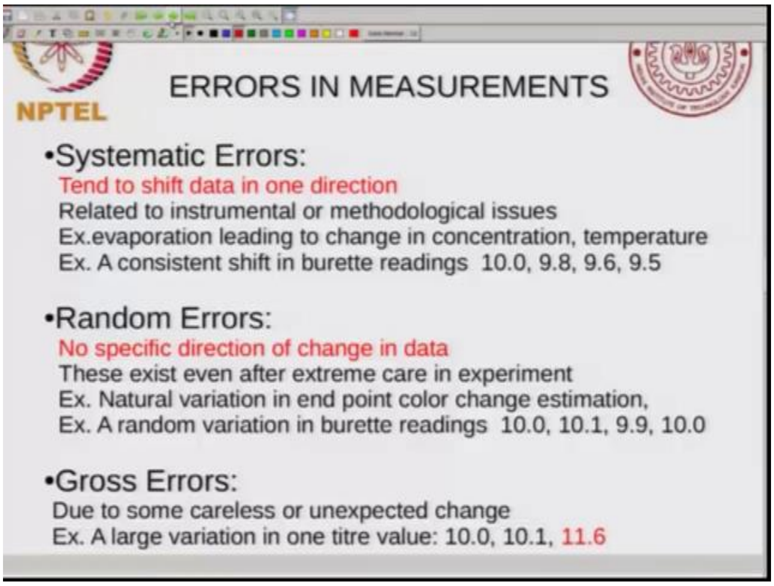
(Refer Slide Time: 04:05)



So, what are the typical errors that are seen in any measurements? So, let us take for convenience, let us take an example of a titration experiment that most of you would have done at some point and what we are measuring in the titration experiment is a volume of the titrate required to reach the end point and we repeat the experiments several times.

So, the types of errors that come up in this measurement or in any typical measurements or can be categories into 3 types one is systematic, the other is random and the other is gross errors and we will talk about each of these in more detail.

(Refer Slide Time: 04:44)



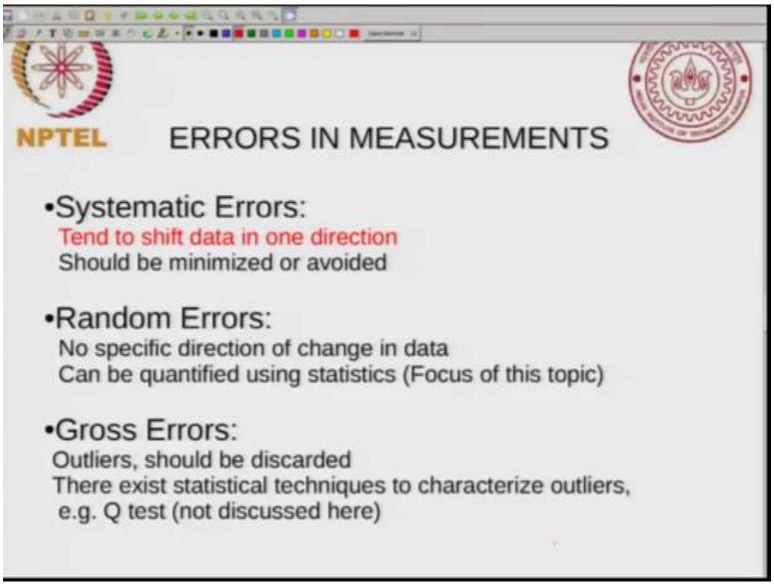
So, what are systematic errors? So, systematic errors they tend to shift the data in one direction. They are related to either instrumental or methodological issues. For example, in the case of titration, you could have evaporation of water that leads to change in concentration or you could have a change in temperature and what this will do is; it will give a consistent shift in the burette readings. For example, if you see successive burette readings. So, first time you might get 10, second time you might get 9.8, 9.6, 9.5 and we see a consistent shift in these readings and this is typical of a systematic error that is made in this in measurements. And in general what you want to do is you want to minimize these systematic errors. So, you want to do your experiments, you want to improve your methodology, so that these systematic errors are minimized. You could also have random errors where there is no specific direction of change in data. This exists even after extreme care is taken in experiments.

So, even if you do the experiments very carefully, there will always be some random errors. For example, if you are doing the titration there is a natural variation in end point color change estimation. So, you estimate the end point by looking at where the color changes and you might see that you know that exact point of noting where the color change happens might not be very accurate and there might be small random errors and this will be reflected in a random variation in the burette readings. For example, you might get a burette reading of 10.0, the first time 10.1, the second point, 9.9 the second time and 10 again the 4th time.

So, basically there is no unlike the systematic errors there is no specific direction of these errors and the nice thing about random errors is that these actually can be quantified. Random errors are things that can be quantified and we will see that in the next half an hour or so. The third kind of errors which I am just mentioning is what are known as gross errors and these are due to some careless or unexpected change and what will typically happen is that you will see a very large variation in one of the values. For example, if you do 3 measurements might get 10.0, 10.1 and then one of the values might look completely different 11.6 and this is typically due to either some spill or some sudden change that happened and gross errors should be eliminated.

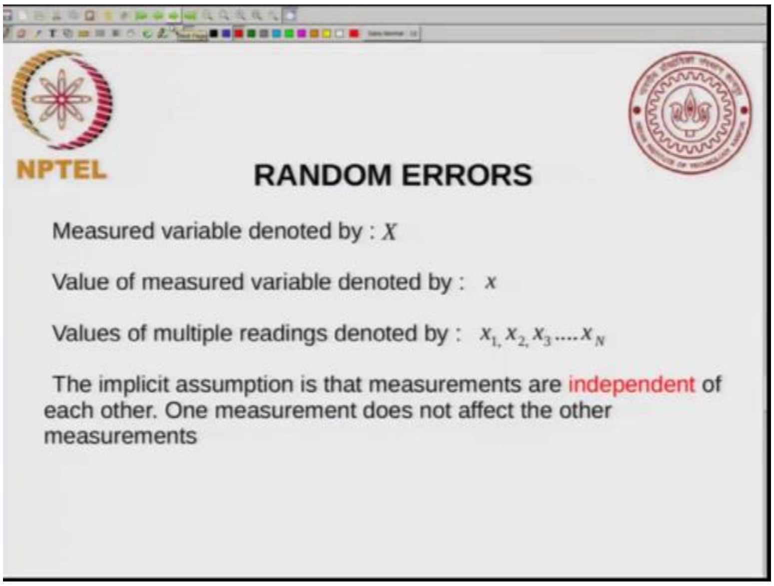
So, if you want to eliminate these gross errors and there is an actual formal way to do that, but we would not be talking about that much, we will be talking mainly about random errors and how to deal with them.

(Refer Slide Time: 07:38)



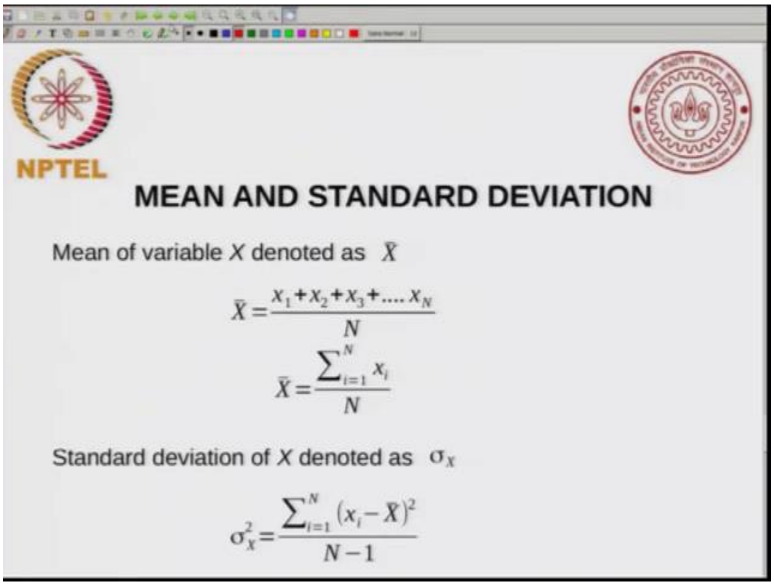
So, let us go to the errors in measurements again. So, just summarize systematic errors should they tend shift data in one direction and they should be minimized or avoided. Random errors on the other hand have no specific direction in change of change in data. These can be quantified using statistics and this will be the focus of this lecture and then there are gross errors which lead to what are called as outliers or you know in some data that is very different from all the others and these actually should be discarded and. In fact, there exist statistical techniques to characterize outliers. For example, the Q test is one very well known statistical test that is used to characterize outliers into identify which of the data are completely off and should be thrown away, but we will not be discussing that this lecture. We will be focusing mainly on the random errors.

(Refer Slide Time: 08:25)



So, coming to random errors; so suppose you are measuring a variable and that variable that you are measuring is denoted by X. So, might be present the volume of the titrate. So, X is the name of the variable, name of the quantity. The value of a measure variable is denoted by little x and if you make multiple readings then, you see different values like x 1, x 2, x 3 and so on up to x n. Now the implicit assumption in all that we are doing is that the different measurements are independent of each other. In other words, one measurement does not affect the value of the other measurements.

(Refer Slide Time: 09:19)

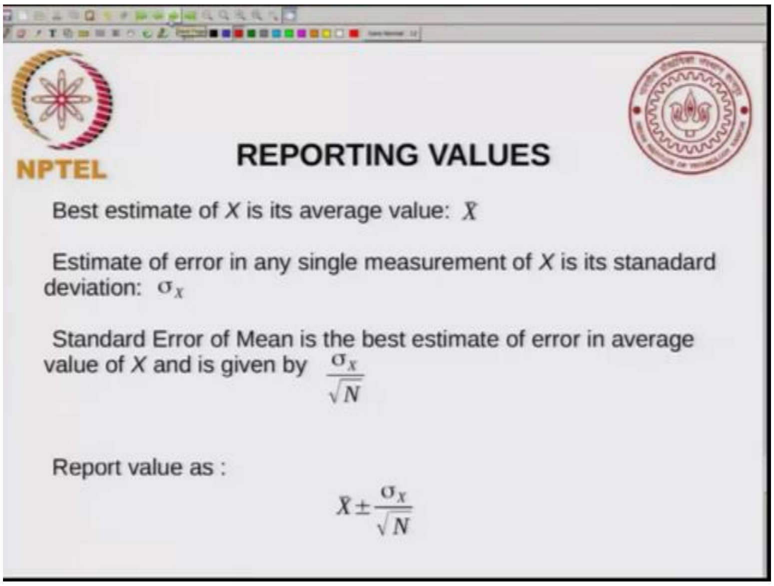


So, again we emphasize that x 1, x 2. So, when you do a titration once you do a titration in the second time the value of what you got from the first titration does not affect what happens in the next titration. Now in such cases we can define things like mean and standard deviation. So, the mean of variable X is denoted as X bar and it is given by the sum of the values divided by the total number of measurements. So, N is the number of measurements and mean is defined as x 1 plus x 2 plus x 3 up to x N divided by N. You can write that in short notation as sum over i equal to 1 to N, x i divided by N.

Now, the standard deviation of data is denoted as sigma x and sigma x square is given by the sum of squares of deviation you take x i minus average value of X. So, you need to calculate the average value of X in ordered to calculate sigma x. This is the square of sigma x which is given in this form and notice that the denominator of sigma x has an N minus 1 and not an N and usually if N is large then it does not make much difference, but for small values of N this does make a difference.

But formally the right quantity here should be N minus 1 in the definition of standard deviation and there is a mathematical reason for this, but we would not get into that, but we will just keep in mind that when you calculating the standard deviation you should use N minus 1 in the denominator.

(Refer Slide Time: 10:47)



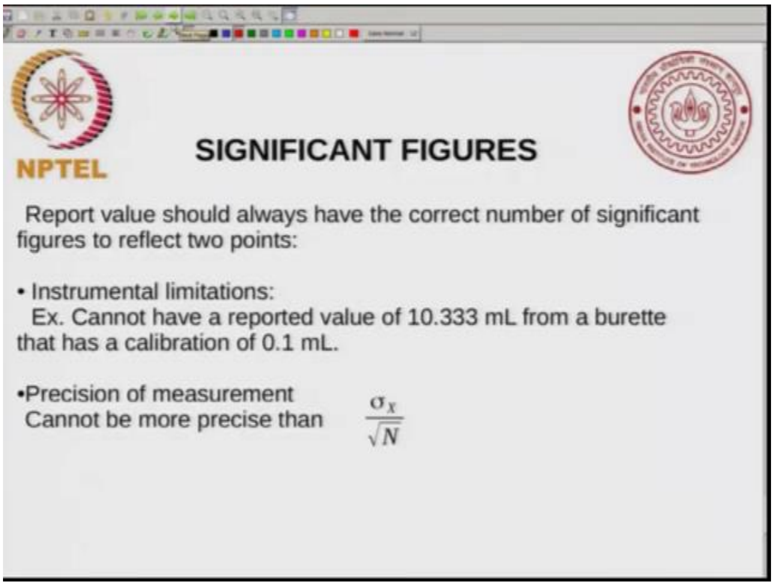
So, now the question is suppose you make number of measurements and you want to report the value. So, you make a number of measurements for X. As we saw before we made a number of we made N measurements and we got values x 1, x 2 or up to x N and what you are asked to do? At the end of the experiment you want to report one single value. So, how do you report one single value? Now the obvious thing to do is to report the average and in turns out that mathematically the best estimate of X, if the different measurements are independent of each other is X bar which is average of value of X.

We can also estimate the error in any single measurement of X. So, the error in any single measurement, any one measurement of X is the standard deviation. So, if you make a single measurement and you ask, what is the estimate of the error in this measurement, in one single measurement? Again this is very important that it is only for one single measurement the error is estimated using the standard deviation. Now there is another quantity called the standard error of mean which is the best estimate for the error in an average value of X.

So, suppose you make N measurements and you report the value of X as the average value of X. What is the estimate of the error in this average value of X? And that is given by sigma x divided by square root of N. So, sigma x is the error in one single measurement, but the error in the average value is actually sigma x is divided by square root of N and important to notice is that as N increases even if sigma x does not change, your error estimate, your standard error of mean goes down; that means, as N increases then your X bar, the error in X bar goes down. So, if you make more measurements then you say that your X bar becomes more precise. So, the correct way to report the value is X bar plus minus sigma x divided by square root of N.

So, this is the estimation of the error. So, you say X bar and it is a random error. So, it can be plus or minus. So, what we mean is that the value of X bar can have variations of approximately this range. So, this is again an estimate and later on we will see why the standard deviation enters into this expression. So, the point is you report a value of X bar and you say that this is the estimate of error. So, the correct way to report the value is X bar plus minus sigma x divided by square root of N.

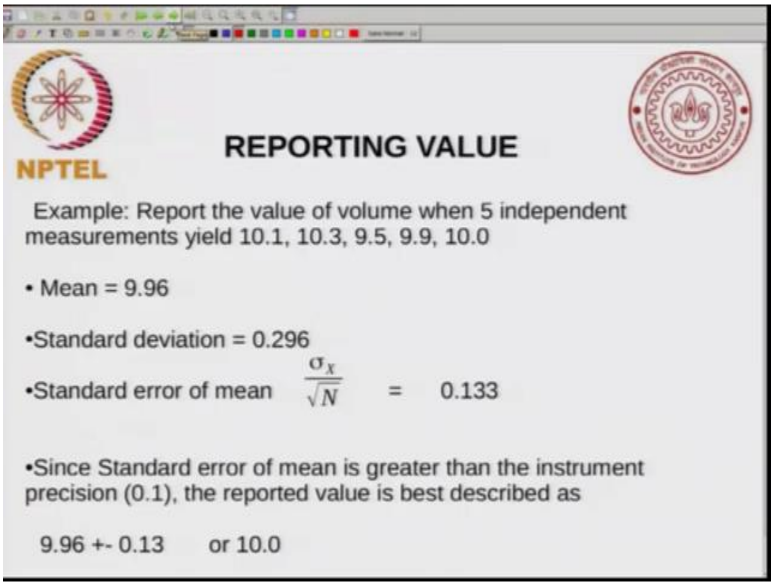
(Refer Slide Time: 13:43)



Now, there is something that you have seen which is, which you may have seen before which is significant figures that relates to the value of the real number that is used to denote the data. So, the reported value should always have the correct number of significant figures to reflect 2 points. So, how many digits do you use after decimal? For example, if you are making a titration measurement, how many digits should you use after decimal? And this is related to significant figures and it should reflect 2 points; first is the instrumental limitations. You cannot have a reported value of 10.333 mL from a burette that has calibration of 0.1 mL.

So, you might make multiple readings, you might make 3 readings and you might take and you might get an average of 10.333 mL, but if your burette has a calibration of only 0.1 mL then you cannot have this kind of accuracy. The second point is that your number of significant figures should reflect the precision of the measurement. So, you cannot have a measurement more precise than this sigma x divided by square root of N. So, we say that whenever you report an average value there is always this much error. So, you cannot have a reported value that is more precise than this. So, what we mean by that? Let us just take an example.

(Refer Slide Time: 15:08)



So, suppose you are asked to report the volume when you make 5 independent measurements, that yield the following numbers and that yield let us say; 10.1, 10.3, 9.5, 9.9, 10. So, these might be volume in milliliters or etcetera. So, now, if you want to report the value of volume and you make 5 independent measurements and you get these values. So, you calculate the mean and you get exactly 9.96, you calculate the standard deviation and you get 0.296. So, next you calculate the standard error of mean and this is nothing, but point 0.296 divided by square root of N and this is about 0.133.

Now, what we notice is that the standard error of mean, this 0.133 that you have is more precise than the instrumental precision. So, the instrument gives you only up to 0.1 and it is less precise than the instrumental position. In other words the standard error of mean, the value of the standard error of mean is greater than the instrumental precision and in such a case you use the standard error of mean to describe the value. So, what you would say is the reported value will be 9.96 plus minus 0.13. Notice that, notice I have kept 9.96 and 0.13. This is actually the value of 9.96 is more precise than the instrumental value, but I have clearly stated the error of 0.13. I have clearly stated that you know you this can only be the error in the measurement is 0.3.

Now, often you know since the instrumental precision is only 0.1 and if you round off 0.133, you get about 0.1. If you round off 9.96 you get about 10, you get 10. So, you can simply report the value as 10.0. So, these are the 2 ways you can report, you can either say 9.96 plus minus 0.13 or you can say 10.0 and both are acceptable. What is important is that your significant figures should reflect both the precision of the instrument and the precision of the measurement. So, if you take lot more measurements then your standard error of mean will go down; however, your precision of the instrument will always be 0.1. So, you can never go below 0.1 even though your standard error of mean goes becomes lots smaller than 0.1.

So, with this I will conclude this discussion on errors, means and standard deviations and what I want you to take home from this lecture is that any measurement whether you do in a lab or you look at data from a table should always be thought of in terms of averages and standard deviations provided the different measurements are independent of each other. So, you only use standard deviations and error estimates and average values if the different measured values are completely independent of each other. If your values are dependent on each other there might be systematic errors, then you cannot use things like averages or means etcetera. So, again it is very important that before you calculate average or standard deviation or standard error of mean, you should make sure that the data that you have is completely independent of each other.

Thank you.