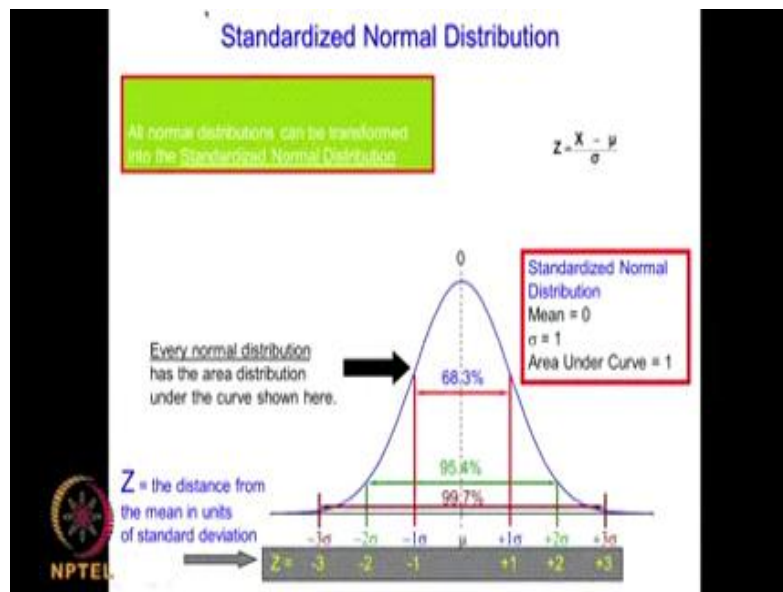


Biostatistics and Design of Experiments
Prof. Mukesh Doble
Department of Biotechnology
Indian Institute of Technology, Madras

Lecture - 06
Standardized Normal Distribution/t distribution

Hello welcome once again, we introduced the concept of a Standardized Normal Distribution in the previous class, I will continue on that again and then later on we will talk about something called t-distribution. So what is standardized normal distribution? Basically you all know what is a normal distribution. Normal distribution we call it uniform distribution, bell shaped curve and so on actually. So, the left hand side here will be equal to the right hand side exactly that is why it is called a inverted bell. But when you transform this normal distribution and that is called standardized normal distribution. Where, the mean is centered around 0, the mean is 0 and the area under the curve will become 1 and sigma that is the standard deviation will become also 1.

(Refer Slide Time: 00:45)



What is the advantage? When we convert a normal distribution into standardized it becomes very easy for us to compare one set of samples with another set of samples and so on. Otherwise what happens is any data for example, if I am measuring heights of students the mean will be in terms of some inches and the standard deviation again will be in terms of inches. If I am looking at height of plants I may

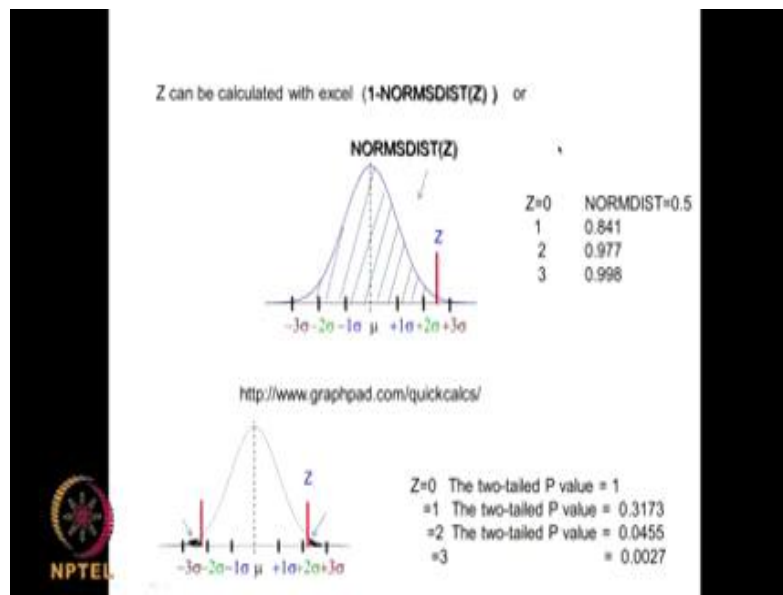
measure them in centimeters, so height will be in terms of centimeters the average height could be in terms of centimeters. If I am measuring weight of the class, then I will be measuring in terms of kilograms 60 kilograms, 70 kilograms and so on. So we have a wide range of numbers where as when you convert this normal distribution into standardized form, so that mean is always equal to 0, area under the curve is 1. It becomes easy for us to compare and also if you look at the tables that are available for determining the area under the curve, they are all assuming that the total area is 1. That is what is called standardized normal distribution. Basically what you do is we convert the data x into say for example,

$$Z = \frac{X - \mu}{\sigma}$$

σ is the standard deviation so when we do that suppose, I put X is 1σ then Z will become 1, if I put X is 2σ then Z will become 2, when I put X is equal to 3σ then Z will become 3 and that is what is shown here. We have 1 sigma, 2 sigma, 3 sigma in terms of Z it becomes 1, 2, 3 and similarly analogous -1σ will become -1 , -2σ will become -2 , -3σ will become -3 , so it is very, very nice.

If you look at the area as I said the area under the curve is 1 that is 100 percent. If you look at the area spanning 1σ to -1σ it will become 68.3, if we look at the area between plus 2σ and -2σ it will be 95.4 and if you look at the area between spanning between plus 3σ and -3σ it will become 99.7 and so on actually. We can even go up to 6σ you must have heard about term call 6σ . If you are going up to 6σ then the area spanned will be 99.9999 that is, what it is. What these numbers have very, very important significance as we go along we will find that this 95 and 99 play very, very important role. So, when we say this area is 95 % that is plus or minus 2σ , then the area outside will be obviously 5 % that means, this side that is 1 side of it or one tail of it will be 2.5 % the other tail of it will have 2.5 %. Similarly, if I consider plus or minus 3σ as 99 % the outside area will be 1 % that means, one side it will become 0.5 % the other side will become 0.5 %. Now, these are very, very important as we can see later on when we talk about statistical significance. When we say 95 % significance, 99 % significance then we are talking in terms of $100 - 95$ it is 0.5 % significance, $100 - 99$ will be 1 % significance and so on. We will be more interested in the area that is outside, rather than the area that is inside, so this 95 % spans plus or minus 2σ and 99 % spans plus or minus 3σ . We will talk about it. How do you convert any data into Z ? It is very simple. I know the mean, I know the sigma I take a data X then I use this formula and convert that into Z . So, if it is 1, sigma Z will become 1, if it 2, sigma Z will become 2, if it is 3, sigma Z will become 3 and so on.

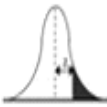
(Refer Slide Time: 05:03)




Excel also has the function that is called NORMSDIST function. But NORMSDIST function calculates this whole area, if I have a value of Z here NORMSDIST will give me the area of the whole portion. So if I want to know the area of this portion that is outside then I do $1 - \text{NORMSDIST}$. When Z is equal to 0 that means here obviously this area will be 0.5 because it is symmetric and equal and when $Z = 1$ here that means this area will be 0.841 and if Z is equal to 2 means then this area will be 0.977, so if I want to know the area outside this then obviously $1 - 0.977$ which will be equal to 0.023, if I multiply that twice that will become 0.046. The area outside two σ will be 0.046. If you if you remember in the previous case that is what it is right? The area inside 2 σ is 95.4, so the remaining is 4.6 so this portion is 2.3, this portion is 2.3. Actually we do not do it as 95.4 we call it 95 itself as a round number. This outside area will become 5, this portion will be 2.5 and this portion will be 2.5.

Generally that is what we do actually similarly for 3 σ we do not say 99.7 we say 99 % so the outside portion will be 1 σ so one side of it will be 0.5 %, another per side will be 0.5 %. We can use this NORMSDIST that is available in excel to calculate similarly we can use this GraphPad online software and that also gives you but it gives you different side it gives this marked area. When Z is equal to 0 that means here the whole area will be 1, when is Z is equal to for example 2, 2 σ that is 2 and here so obviously, the outside should be around 5 % that is what it is showing approximately 0.0455. If Z is equal to 3 when we say 99.7 the remaining portion is 3 percent that is what you are getting here. We can use either the simple and then use the Z table I am going to show you, this is an Z table.

(Refer Slide Time: 07:36)



Calculates the probability of a USL defect above the given value of z

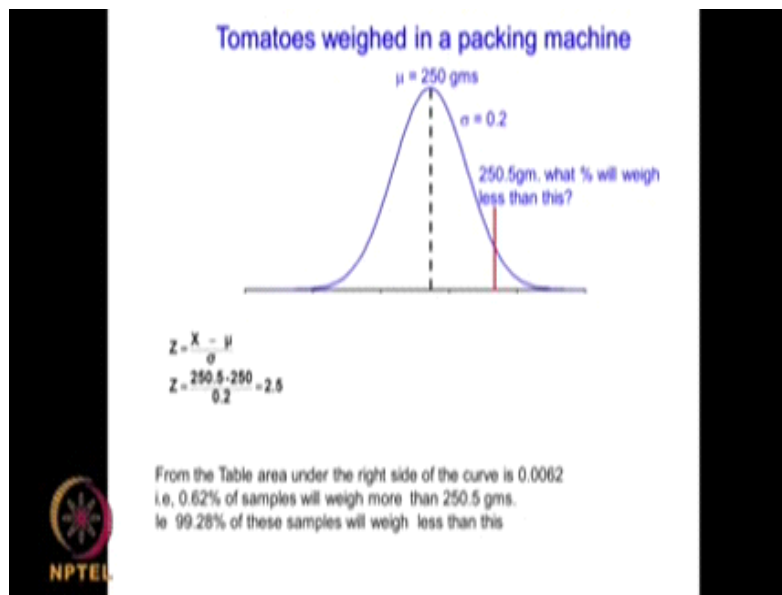


Single-Tail Z Table

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5200	0.5239	0.5279	0.5319	0.5359
0.1	0.4602	0.4641	0.4681	0.4721	0.4761	0.4801	0.4840	0.4880	0.4919	0.4959
0.2	0.4207	0.4246	0.4286	0.4325	0.4365	0.4404	0.4443	0.4483	0.4522	0.4561
0.3	0.3603	0.3643	0.3683	0.3723	0.3763	0.3803	0.3843	0.3883	0.3923	0.3962
0.4	0.3005	0.3045	0.3085	0.3125	0.3165	0.3205	0.3245	0.3285	0.3325	0.3365
0.5	0.2500	0.2539	0.2579	0.2618	0.2658	0.2698	0.2738	0.2778	0.2818	0.2858
0.6	0.1995	0.2035	0.2075	0.2115	0.2155	0.2195	0.2235	0.2275	0.2315	0.2355
0.7	0.1493	0.1533	0.1573	0.1613	0.1653	0.1693	0.1733	0.1773	0.1813	0.1853
0.8	0.1093	0.1133	0.1173	0.1213	0.1253	0.1293	0.1333	0.1373	0.1413	0.1453
0.9	0.0540	0.0580	0.0620	0.0659	0.0699	0.0738	0.0778	0.0817	0.0857	0.0897
1.0	0.0242	0.0281	0.0320	0.0359	0.0398	0.0438	0.0477	0.0517	0.0556	0.0596
1.1	0.0044	0.0083	0.0122	0.0161	0.0200	0.0239	0.0278	0.0317	0.0356	0.0395
1.2	0.0005	0.0044	0.0083	0.0122	0.0161	0.0200	0.0239	0.0278	0.0317	0.0356
1.3	0.0001	0.0040	0.0079	0.0118	0.0157	0.0196	0.0235	0.0274	0.0313	0.0352
1.4	0.0000	0.0039	0.0078	0.0117	0.0156	0.0195	0.0234	0.0273	0.0312	0.0351
1.5	0.0000	0.0038	0.0077	0.0116	0.0155	0.0194	0.0233	0.0272	0.0311	0.0350
1.6	0.0000	0.0037	0.0076	0.0115	0.0154	0.0193	0.0232	0.0271	0.0310	0.0349
1.7	0.0000	0.0036	0.0075	0.0114	0.0153	0.0192	0.0231	0.0270	0.0309	0.0348
1.8	0.0000	0.0035	0.0074	0.0113	0.0152	0.0191	0.0230	0.0269	0.0308	0.0347
1.9	0.0000	0.0034	0.0073	0.0112	0.0151	0.0190	0.0229	0.0268	0.0307	0.0346
2.0	0.0000	0.0033	0.0072	0.0111	0.0150	0.0189	0.0228	0.0267	0.0306	0.0345
2.1	0.0000	0.0032	0.0071	0.0110	0.0149	0.0188	0.0227	0.0266	0.0305	0.0344
2.2	0.0000	0.0031	0.0070	0.0109	0.0148	0.0187	0.0226	0.0265	0.0304	0.0343
2.3	0.0000	0.0030	0.0069	0.0108	0.0147	0.0186	0.0225	0.0264	0.0303	0.0342
2.4	0.0000	0.0029	0.0068	0.0107	0.0146	0.0185	0.0224	0.0263	0.0302	0.0341
2.5	0.0000	0.0028	0.0067	0.0106	0.0145	0.0184	0.0223	0.0262	0.0301	0.0340
2.6	0.0000	0.0027	0.0066	0.0105	0.0144	0.0183	0.0222	0.0261	0.0300	0.0339
2.7	0.0000	0.0026	0.0065	0.0104	0.0143	0.0182	0.0221	0.0260	0.0299	0.0338
2.8	0.0000	0.0025	0.0064	0.0103	0.0142	0.0181	0.0220	0.0259	0.0298	0.0337
2.9	0.0000	0.0024	0.0063	0.0102	0.0141	0.0180	0.0219	0.0258	0.0297	0.0336
3.0	0.0000	0.0023	0.0062	0.0101	0.0140	0.0179	0.0218	0.0257	0.0296	0.0335
3.1	0.0000	0.0022	0.0061	0.0100	0.0139	0.0178	0.0217	0.0256	0.0295	0.0334
3.2	0.0000	0.0021	0.0060	0.0099	0.0138	0.0177	0.0216	0.0255	0.0294	0.0333
3.3	0.0000	0.0020	0.0059	0.0098	0.0137	0.0176	0.0215	0.0254	0.0293	0.0332
3.4	0.0000	0.0019	0.0058	0.0097	0.0136	0.0175	0.0214	0.0253	0.0292	0.0331
3.5	0.0000	0.0018	0.0057	0.0096	0.0135	0.0174	0.0213	0.0252	0.0291	0.0330
3.6	0.0000	0.0017	0.0056	0.0095	0.0134	0.0173	0.0212	0.0251	0.0290	0.0329
3.7	0.0000	0.0016	0.0055	0.0094	0.0133	0.0172	0.0211	0.0250	0.0289	0.0328
3.8	0.0000	0.0015	0.0054	0.0093	0.0132	0.0171	0.0210	0.0249	0.0288	0.0327
3.9	0.0000	0.0014	0.0053	0.0092	0.0131	0.0170	0.0209	0.0248	0.0287	0.0326
4.0	0.0000	0.0013	0.0052	0.0091	0.0130	0.0169	0.0208	0.0247	0.0286	0.0325

We can use Z table so for given value of a Z. It gives you the area it gives you the this area it is called single tail it is because it is giving only one portion of the area where as your GraphPad gives you this plus this so I need to double this to get the results from GraphPad. Each of these softwares or Excel function or the table gives Z the area in a different way, but they are all analogous to each other and we can calculate one from the other you understand. So the GraphPad gives you this plus this area where as the table gives you one set of the area so obviously, if I want to make it appear like GraphPad I multiply by 2, so from Z equal to any value for example, when Z is equal to 2 it gives you 0.0228. If you want to multiply twice it will become 0.0228 multiplied by twice is 456 and that is what the GraphPad gives 0.0455 whereas, Excel gives the whole area when I multiply 1 minus Z will get this area which is similar to the area the table gives you. So the table gives you Z values Z is equal to 0.01 123 if you want second decimal this side we go up to 7.9, so from there we calculate this area which is equal to this area so converting a normal distribution into a standardized normal distribution has many advantages we can do many statistical calculations and that is what we are going to do. Now let us look at a problem.

(Refer Slide Time: 09:36)

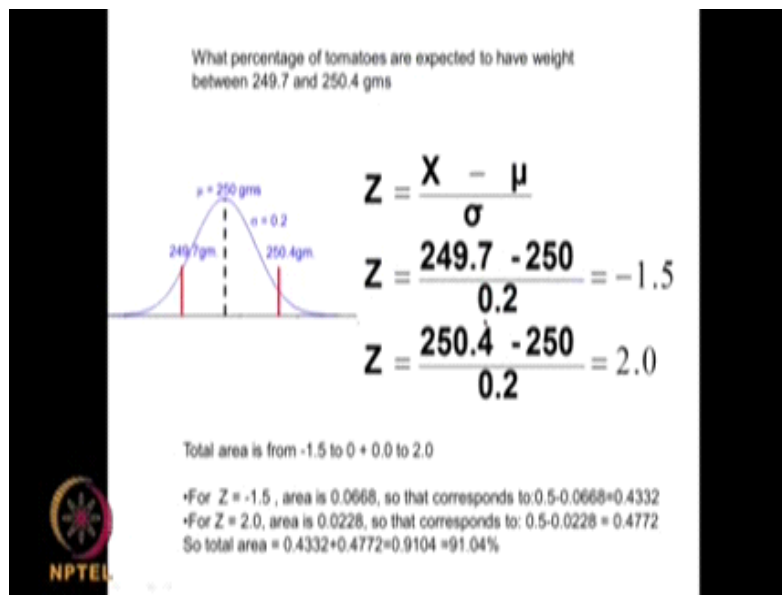


Now, tomatoes weight in a packing machine the mean is given as 250 grams and σ is given as 0.2, so this like a population that is why I put μ and σ . It is collected over a very long period of time so it gives you this and this. Now I have a tomato which is 250.5 grams. What % will weighed less than this, so I have taken 1 sample 250.5 grams. What will be the weight of tomatoes? What will be the % of tomatoes which will weighed less than 250.5 grams. What I do,

$$z = \frac{X - \mu}{\sigma}$$

250 σ is 0.2. So Z is equal to 2.5. I go to my table 2.5 I get 0.0062. 0.062. So 0.062 % of the sample will may more than 250 because the table gives you this side right so table gives you this side. If you want to know this side all obviously, I subtract by 100, so 100 minus 0.062 is 99.28 so 99.28 % of the tomatoes samples if you pick up will weigh less than 250.5 grams do you understand? How to do? Now mu 250, I know σ 0.02, 250.5 substitute here I get Z as 2.5, so I go to table Z is 2.5 so which is 0.0062 so you should note that it is giving the outside area. So I want to know tomato weighing less than 250.5 obviously, I need to get this area. I can subtract 1 minus 0.0062 or 100 minus 0.062%. That is what I have done here. So 99.28 % of the tomatoes will weigh less than this particular number very interesting. So it is a very useful set of data.

(Refer Slide Time: 11:41)



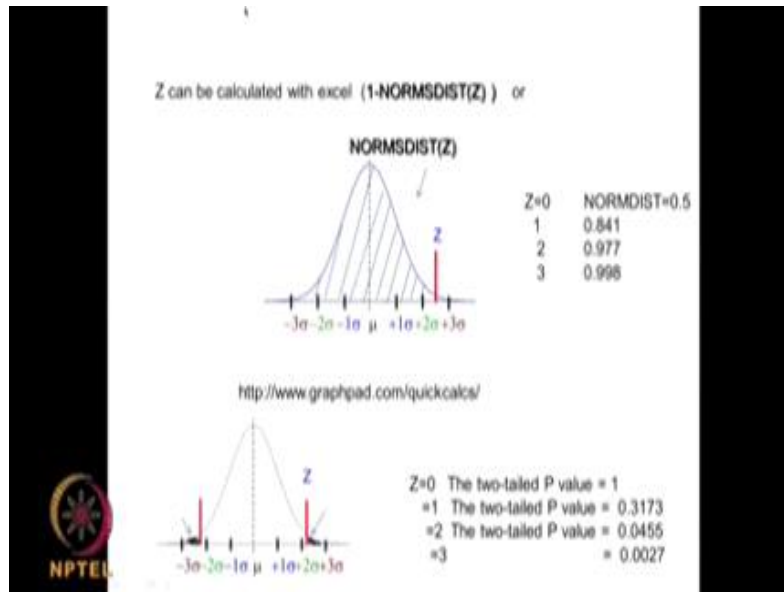
Now, let us look at another problem so again tomatoes the mean is 250 grams and σ 0.2. What percentages of tomatoes are expected to have weight between 249.7 and 250.4? I want to know this area, this area understand. I want to know this area. What do I put same formula

$$Z = \frac{X - \mu}{\sigma}$$

I will put 249.7 -250 I will get - 1.5 then I will get 250.4 - 250 that will get give me 0.2. I need to get the area for - 1.5 to 2. The this area that is between - 1.5 and 2 is what I am interested in. Now Z is equal to - 1.5 the table will give me this whole area so obviously, I need to subtract that from 0.5 because this area is 0.5. Z = -1.5 or Z = 1.5 it is does not matter. So I take Z = 1.5 that is 0.0668 but this whole area, right. I need to subtract 0.5 that this portion then I will know this area, so I will subtract 0.5 - 0.0668 that is 4332. That is this area do you understand this portion now Z is equal to 2. I go again to the table Z is equal to 0.0228.

So it gives you this whole area but again I need to subtract the 0.5 portion. It will give me the whole area according to from the table but I need to subtract 0.5 from that because of this because I need to know only this area. So 0.5 - 0.0228 is 0.04772. This area this area is 4772 this area is 4332, so I add all I get 0.91 that means 91 % of the tomatoes will have weight between 250.4 and 249.7. It is a very useful way of calculating things if I know the μ and the σ . As you can see here how powerful it is now. Of course, you can do the same thing using Excel NORMDIST function or you can also use the GraphPad.

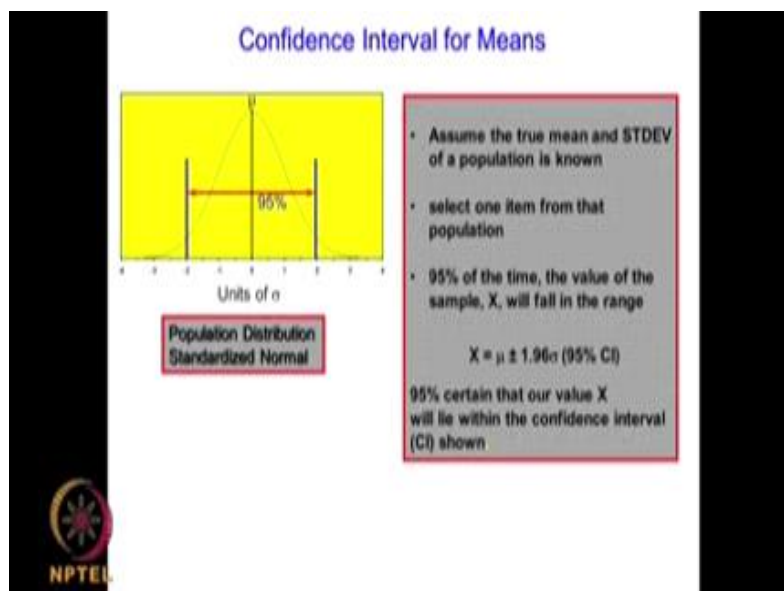
(Refer Slide Time: 14:23)



Because you have to remember that the GraphPad results or area of this and this where as the single tail t table gives you this area, so if you multiply this by 2 I will get the GraphPad results. Where as the Excel software gives you the inside area so I need to subtract 1 minus to get the outside area. So that is the relationship between the 3 different approaches.

Now you can as a homework take up and see whether you get the same answer as this using the NORMSDIST function as well as the GraphPad function. I do not want to go into that.

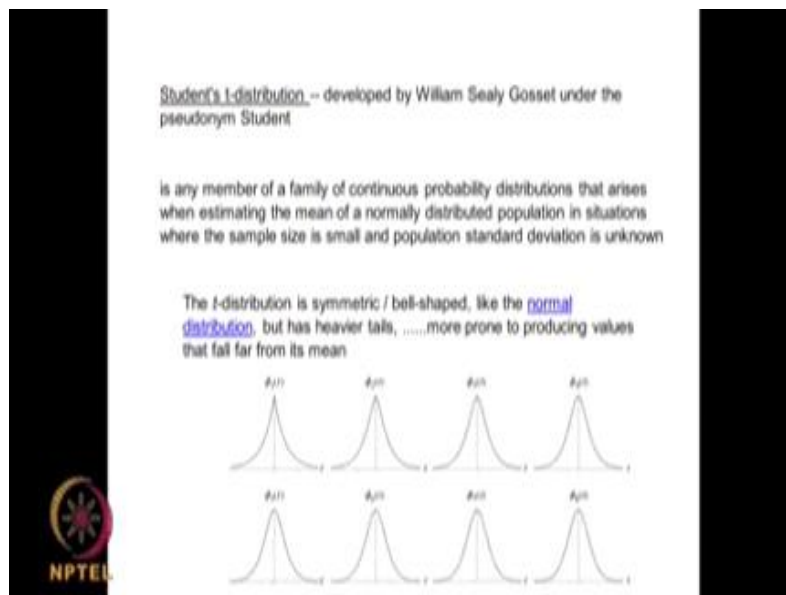
(Refer Slide Time: 15:13)



Now, there is something called the confidence interval I talked about it right 2σ it is approximately 95 % actually 95.4 we take it as 95 %, 3σ is 99.7 but we will take as 99 %. When we have a data set well I know the global population mean and standard deviation and then I suppose I take 1 sample in it as I mentioned long time back it is very difficult to take the entire population for sampling, but normally we will take only 1 or 2 samples, so I take only 1 out of this and 95 % of the time 95 % of the time it will fall within this region. That means, if I know the average mean 95 % of the time it will be within this average plus or minus 1.96σ , how would I get this 1.96 that is the 2σ as you know $0.2 = 2.28\%$. So multiply both sides that will come to approximately 2.28 multiply both sides will be approximately 0.045 % So that is 95 % here. So 95 % of the time the data will lie between this region the $+ \text{ or } - 2\sigma$. So if I take a sample 95 % of the time it will lie between $\mu + \text{ or } - 1.96\sigma$. So if you want to know 99 % obviously, I need to take this 3σ right that will be the formula is 2.58 because for 2σ as you know it will become 2 this is 0.0028 for 3σ the outside area is 0.0013 right.

We get like this. So this also very useful, if I am taking a sample, if I take 10 samples from a population, so I will know the mean of those 10 samples but now I want to know, what is an estimate of the confidence interval for the population because the 10 samples which we take will give me some mean but that will not be the exact population mean right because if I had taken another 10 samples I would have got another mean, if I take another 10 samples I will get another mean and so on but they are not real representation of my population mean. I can use this type of approach, should tell what will be the range in which my population will lie that is the advantage of getting something called the confidence interval. Now let us go to the next topic that is called students t-distribution.

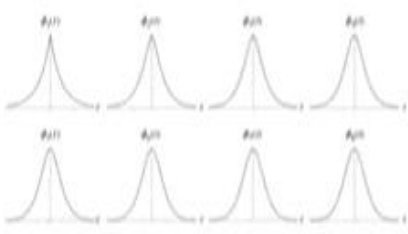
(Refer Slide Time: 18:25)



Student's t-distribution -- developed by William Sealy Gosset under the pseudonym Student

is any member of a family of continuous probability distributions that arises when estimating the mean of a normally distributed population in situations where the sample size is small and population standard deviation is unknown

The t-distribution is symmetric / bell-shaped, like the normal distribution, but has heavier tails,more prone to producing values that fall far from its mean



NPTEL

What is this student's t-distribution? This was developed by somebody called William Sealy Gosset. He wrote some very important paper in statistics under the pseudo name of student that is how it stuck on actually, he said normal distribution is a bell shaped uniform, but then the **N** should be very, very large but if I have very, very small n. Now if I have **N** equal to 1, 2, 3, 4, 5, 6, 30, 25 and so on that also will be symmetric but it will have heavier tails not like a nice looking normal distribution this is a nice looking normal distribution whereas, when I take **N** is equal to say 1 or 2 or 3 will have very tail, it will be uniform just like your normal distribution, it will be symmetric and all that but it will never be a bell shaped. A bell shaped should have been like this, right, like this, as you can see when I take much smaller numbers it will be like this. When I take a very large number as you see ideally if you have more 30 or 30 then you can almost get-distribution almost looking like a normal distribution that is what it is actually. When the sample size is small then the population, so then we are going to get it is skewed. Although it is uniform symmetric just like a normal distribution but it will have longer tails in real life. We are dealing only with samples we are not dealing with the population so obviously, student t-distribution becomes very important because whatever I do I am collecting 10 students and measuring their height, so that is a sample, I am testing drug on 20 rats that is a small, small sample, I am taking 10 tomato plants and measuring their weight, that is a sample. So student t-distribution becomes very important and we might not collect 50 samples, 60 samples, 100 samples we may collect 8, 9, 10 and so on, so that is how this t-distribution will become important and you will come across this term **T, T, T** many times everywhere in bio statistics. What are the properties of the t-distribution?

The mean of the distribution equal to 0 right because it is also symmetric just like a normal distribution Z.

(Refer Slide Time: 21:03)



Properties of the t Distribution

- The mean of the distribution is equal to 0 .
- The variance is $= v / (v - 2)$, where v is the degrees of freedom and $v > 2$.
- The variance is always > 1
- With ∞ degrees of freedom, the t distribution is = the standard normal distribution.

NPTEL

<http://stattrek.com/probability-distributions/t-distribution.aspx>

The variance that is the spread is given by

$$= v / (v - 2),$$

where v is the degrees of freedom and v is > 2 , the variance is always > 1 so the degrees of freedom reaches infinity the t-distribution reaches a standard normal distribution of course, infinity but even a 30, 35. I would say it looks almost like a Normal Distribution. If you have 30 samples then I think it is almost like a normal.

(Refer Slide Time: 21:36)



When to Use the t-Distribution

- with any statistic having a bell-shaped distribution (i.e., approximately normal).
- The population distribution is normal.
- The population distribution is **symmetric, unimodal**, without **outliers**, and the sample size is at least 30.
- The population distribution is moderately **skewed**, unimodal, without outliers, and the sample size is at least 40.
- The sample size is greater than 40, without outliers.
- The t distribution should *not* be used with small samples from populations that are not approximately normal.

NPTEL

<http://stattrek.com/probability-distributions/t-distribution.aspx>

When to use this so any statistics having a bell shaped, that means approximately normal the population distribution is normal, it is symmetric, uni modal. You should not have bi model, I taught you what is mode? That is the central tendency it should not have outliers. If we have a samples of 30 it should not have outliers, it should be symmetric and unimodal. If the sample size is 40 it can be little bit skewed but still it should be unimodal no outliers. If it is greater than 40 then it should not have outliers, but otherwise it is ok. The t-distribution should not be used with small samples that are not approximately normal, that means if the samples are very small and it is not normal then be careful do not use the t-distribution. So there is something called test for normality we will talk about that later in the course of time to see whether your data is normal and if it is reasonably normal then we can use the t -distribution. As I said why do we use t-distribution because we always collect samples, we never collect entire population data set samples will be having a degrees of freedom of 5, 6, 10, 20 and so on. In most of the situations we use the concept of t-distribution and perform the analysis.

(Refer Slide Time: 23:04)

Population and sample

standard error (SE) is the standard deviation of the sampling distribution of a set of means taken from a population.

sample mean is the usual estimator of a population mean.


different samples drawn from the same population would have different values of the sample mean, so there is a distribution of sampled means (with its own mean and variance).

The **standard error of the mean (SEM)** is the standard deviation of the **sample-mean's estimate of a population mean**

$$SE_x = \frac{s}{\sqrt{n}}$$

s is the **sample standard deviation** (i.e., the sample-based estimate of the standard deviation of the population),

n is the size of the sample



We also need to know the population and sample I am coming back to the same thing again. If I draw a sample from a population and take the mean, I will get some mean. If I take another set of samples, I will get slightly different mean then if I take another set of samples, I will get slightly different mean. So, this means will have it is own mean and variance. The mean of all these means is a good representation of the population mean but the variance is not really exact representation of the standard deviation of the population. There is something called standard error of the mean that is the standard deviation of the sample means which is an estimate of population standard error of all these means that is given by

$$SE_x = \frac{s}{\sqrt{n}}$$

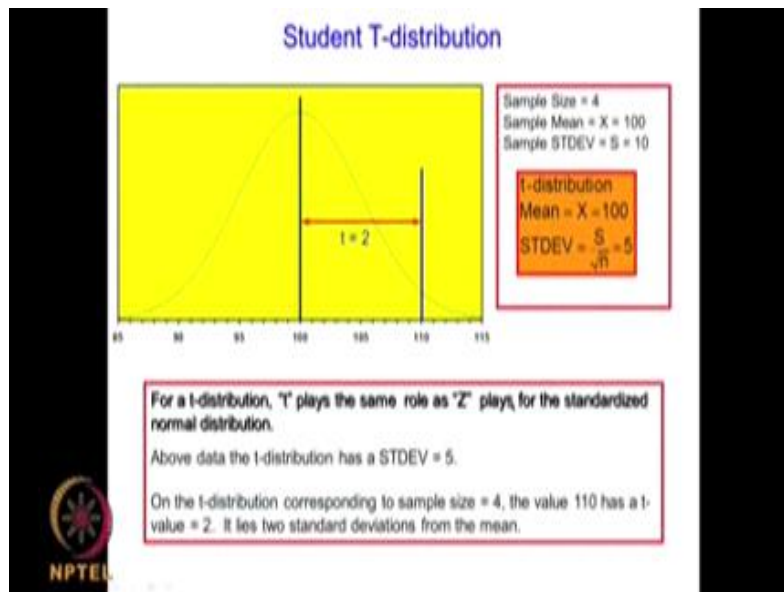
s is your sample standard deviation, n is your number of data points. This is a standard error of the mean. In fact, this is a real representation of the standard deviation of the population so you understand. I have a population I take a few samples and then calculate a mean then I take put it back, then I take another set of samples and get a mean, I will definitely get something different then I will take another set of samples and then get a mean. I will have a large set of mean, these means will have its own mean and some standard deviation. The means of the means which is a good representation of the population mean but the standard deviation of these means are not and it is given like this, the standard error of the mean is the standard deviation of the sample mean is estimate of a population mean.

So, s is your sample standard deviation and n is a number of samples you take square root of n . This is how it is and this standard error is very, very important. Standard error is a standard deviation of the sampling distribution of a set of means taken from a population. So, standard error is a standard deviation of all these means which you have taken with different samples you have taken from the same population do you understand. The mean of all these means is a good representation of the population mean but the if you want represent standard deviation so, if we get something called standard error, that is the standard deviation of the sampling distribution of a set means taken from a population and that is a good representation of the standard deviation of the population and that is given by

$$\frac{s}{\sqrt{n}}$$

where S is the sample standard deviation and n is your number of data or samples you pick up and this is very, very important we use this in many calculations as we go along.

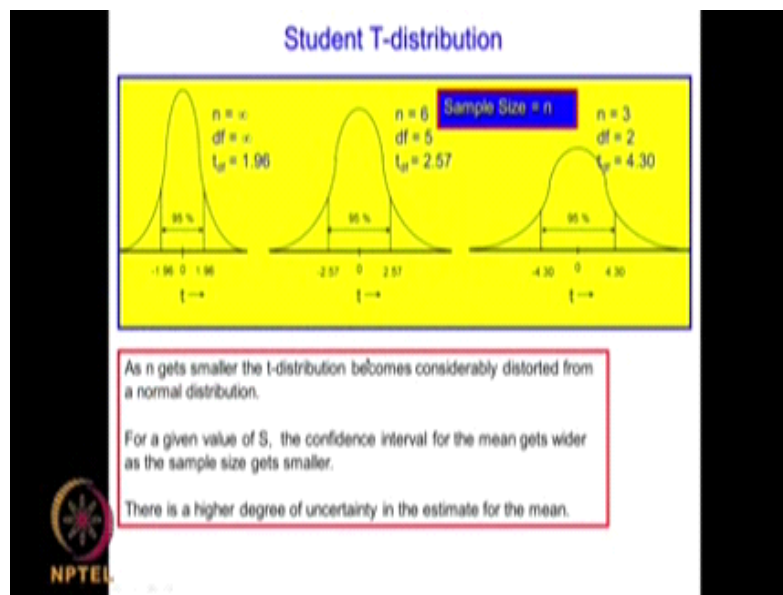
(Refer Slide Time: 26:18)



Again we come back to student t-distribution, so t student is almost a Z you know they play the same role Z you have in the nor standardized normal distribution where as t value you have in a normal distribution with very small sample size. If I have a sample size of 4, mean is 100, sample standard

deviation is 10, then when I do a t-distribution mean is still 100, the standard deviation will be 10 divided by 4 that is 10 by 2 which is 5. So this is what it is and as I said, the t here plays a same role as Z in the standardized normal distribution. If you take a data as 100 here and this is 110 and 110 and value of 110 what will be the t value here obviously, t value will be if we take the sample standard deviation as 10 t value will be equal to 2 actually, that means it lies 2 standard deviations from the mean. In fact, if you remember exactly in the z in the normal standard deviation also we had the same numbers coming out. So, 2σ becomes 2 and similarly here 2 times standard deviation become $t = 2$.

(Refer Slide Time: 27:56)



As I mentioned, if n is equal to infinity data points, then 95 % confidence will be instead of 2σ , if you remember here 95 % confidence 2σ but then we are talking about t where the number of data points are suppose to be less than the population. There will be instead of 2σ , it will become 1.96 + or - σ . So you understand this 1.96 that is how it came here if you take a population then this will become 2σ but we are not taking a population here we are talking t-distribution so it is 1.96, so if n is equal to 6, that means number of data is very as reduced dramatically degrees of freedom n - 1 then for 95 % confidence t is 2.57. How do you get this? There is a table which gives you that. I will show you the table. If n is equal to 3 data is still less as you can see its got a long tail its symmetric uniform but the tail has become longer and longer none of them look like my beautiful normal distribution were here we are talking in terms of n is equal to infinity. We are talking about t-distribution. Where n is smaller than infinity, obviously it will not look like a beautiful bell shaped curve, uniform curve it will have a longer tail so when n equal to infinity in student-distribution for a 95 percent it gives you t of

1.96 and 1.96 and for n is equal to 6 it becomes 2.57, n is equal to 3 is 4.3, do you understand this? If the data points keep going down the uncertainty also goes up because from 1.96 gone 2.57 is gone to 4.3. So that is there is a higher degree of uncertainty in the estimate of the mean. So if I take only 3 data points and take a mean the uncertainty on the population mean is higher because this is much larger.

Whereas, if I take very large data point the uncertainty is much less because the graph also looks smaller I mean more sharper, so the area also is much less. You understand. This 1.96, if we remember I introduced that some time back right here you can see this 1.96 same thing. So it is mean plus or minus it gives an idea about the confidence interval mean plus or minus 1.96 sigma, so this 1.96 corresponds to 95 percent confidence interval. It is not exactly 2 σ that is when you use population when you use samples and try to get an estimate of population obviously, it is 1.96, this 1.96 will come quite often if for 95 percent. Similarly if I take 99 % instead of getting 3 σ 3, I will get 2.58 so that is from the table I will show you a table called t table and that is very useful the t table comes here.

(Refer Slide Time: 31:39)

df	Confidence Level		
	0.90	0.95	0.99
1	1.638	1.963	3.078
2	1.601	1.925	2.920
3	1.577	1.893	2.767
4	1.554	1.860	2.624
5	1.532	1.828	2.501
6	1.511	1.797	2.398
7	1.491	1.767	2.308
8	1.472	1.738	2.228
9	1.455	1.710	2.157
10	1.439	1.683	2.093
11	1.424	1.657	2.035
12	1.410	1.632	1.982
13	1.397	1.608	1.933
14	1.385	1.585	1.888
15	1.374	1.562	1.846
16	1.364	1.540	1.806
17	1.355	1.519	1.768
18	1.347	1.498	1.732
19	1.339	1.478	1.698
20	1.332	1.458	1.666
25	1.319	1.437	1.626
30	1.308	1.418	1.589
40	1.299	1.401	1.554
50	1.292	1.385	1.521
60	1.286	1.371	1.490
70	1.281	1.358	1.463
80	1.277	1.346	1.440
90	1.273	1.335	1.420
100	1.270	1.326	1.403
∞	1.267	1.318	1.389

That is called the t table as you can see here so you can see for a 1.96 you can see for infinity data points for a 95 % confidence interval. If you want take a 99 % confidence interval it is about 2.58, these 2 numbers play very important role. If I want to know a confidence interval 95 % using t-distribution 9, 8, I will multiply by 1.96. If I want to know the confidence interval 99 I multiply by 2.58. These 2 terms are very, very important, do you understand? We go back again in the t-distribution so when the data points are less, it will have very long tail. As you can see as the data points increase

the tail goes down which means we get more confidence in estimating the population mean when I have more data points. Now we will continue this student t-distribution in the next class also.

Thank you very much.

Key words- Student T-distribution, Standardized Normal Distribution, Single-Tail Z Table, NORMSDIST(Z), Confidence Interval for Means, continuous probability distributions