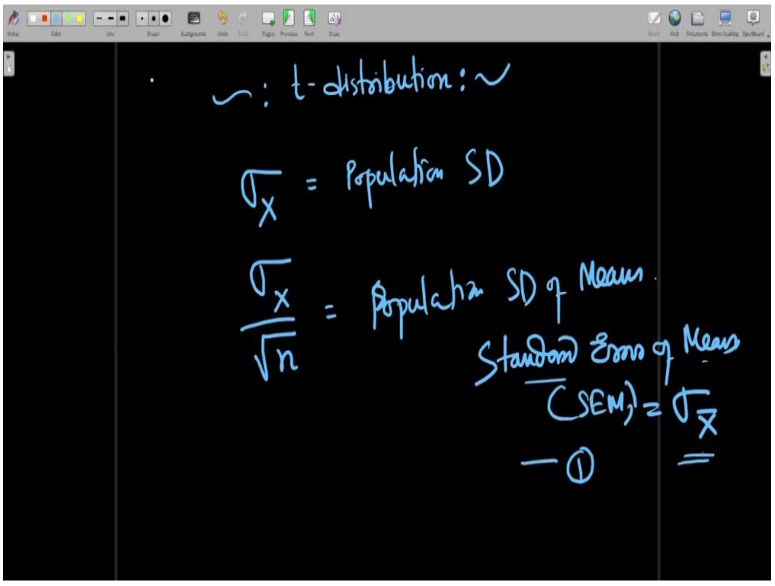
**Applied Econometrics**
**Prof. Tutan Ahmed**
**Vinod Gupta School of Management**
**Indian Institute of Technology - Kharagpur**

**Module - 3**
**Lecture - 29**
**T-Distribution**

Hello and welcome back to the lecture on Applied Econometrics. And in this lecture, we are going to talk about a very important probability distribution, that is our t-distribution. Now, t-distribution is important because, we will see in many cases, in most of the cases, we are actually going to deal with unknown parameters; we really do not have known population parameters.

And when we are actually trying to use the unknown parameters from sample, we use the t-distribution. And we will see wide-ranging application of t-distribution when you do regression and other applications. So, let us try to understand what is a t-distribution.

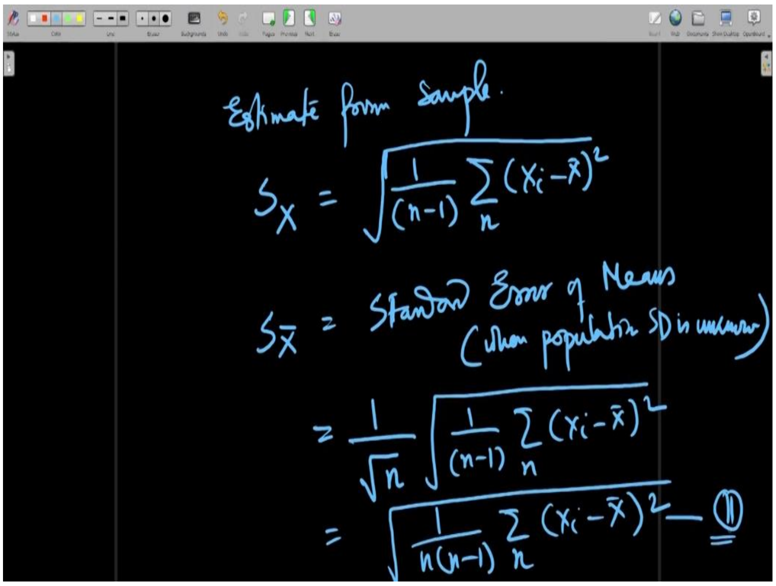
**(Refer Slide Time: 01:13)**




So, when I have a population which I have a known variance for; for example, let us say that is sigma X; and I get the say population variance, population standard deviation. Now, this, I calculate from all the observations; I do not take mean, but I just calculate from all the observations. Now, suppose I take mean, I want to estimate the population standard deviation by taking means.

And if I take means several number of times, what I get is a population standard deviation from all the means. Now, we usually do not do that; what we do as an approximation is that, we simply divide it by square root of n, and that is population SD of means. We sometimes call it standard error of means or SEM we sometimes write it; and that, sometimes we represent it as sigma X bar. Now, that is for the, when you know about the population SD; but most of the cases, as I just said, you do not know the population SD.

**(Refer Slide Time: 02:35)**



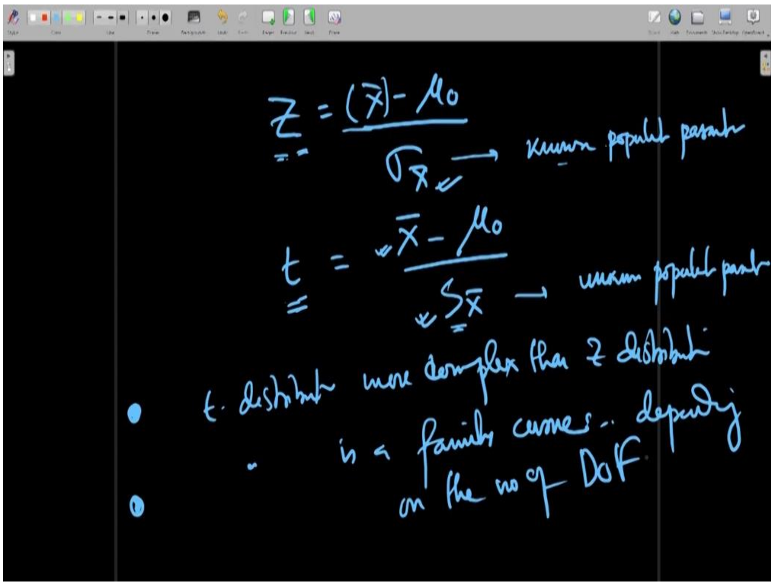

So, what you have to do is, you have to estimate from sample. Now, when you have to estimate from sample, we know we have this S X, we write it as; I can actually use square root of 1 by n - 1; then I have X i minus X bar whole square. So, I basically divide it by 1 by n - 1. And we have explained the intuitive logic, why I have to divide by n - 1 and not by n. So, that I get.

Now, that is for the; you are basically getting the standard deviation where you do not have the population standard deviation known to you. Now, if I want to get the standard deviation of means, so, you want to actually estimate the standard deviation from means, which is nothing but the standard error; we basically say it a standard error of means when population SD is unknown.

So, what we do is, we essentially divide it by, there is a same approximation procedure, 1 by root n summation 1 by n - 1 over n X i minus X bar whole square; or I can write square root of 1 by n into n - 1 summation all over n X i minus X bar whole square. So, that is how we estimate the standard error of mean for, when we are actually using sample to actually estimate the standard error of mean.

Now, in the first case; so, let me actually write down the equation. So, let us say this is my equation 1 and this is my equation 2. So, in equation 1, I estimated standard of mean, but then my sigma is known. And here, I have estimated standard of mean, but sigma is not known.

**(Refer Slide Time: 05:02)**



So, if my standard deviation is known, I write my x is equal to X bar minus mu 0 by sigma X; so, where my standard normal z is going to be this. So, here, this is known; but in case it is not known, then I use t. And when I use t, it is basically this by, I use that S X bar; it is from unknown. So, known population parameter; this is unknown population parameter. So, this is basically the difference between the normal distribution, the z-distribution; for unknown population parameter, you use the t-distribution.

Now, t-distribution is widely used, because, as I said, in most of the cases, you do not know the population parameter in advance. Now, there are some other differences. Say the difference comes in terms of the randomness; so, in this case, suppose when you talk about z, we have this X bar which is random variable; it is, you draw samples and you get the X bar value; whereas, your sigma X bar, this is given, and this is like, you do not have to bother about that.

But in case of t, your X bar, it is a random variable, at the same time, your S X bar is also random variable, because, as you draw random samples, you essentially get different values of S X bar; and that is why it keeps on changing. So, because of that, the z-distribution, it has a random variable only in numerator; and t-distribution, your random variable both the numerator and denominator.

And that is why the t-distribution is actually a little more complex than the distribution that you get for z, which is a normal distribution. Now, so, let us write down; t-distribution more complex than z-distribution. Now, another aspect, another property of t-distribution is that t-distribution is a family of curves. And what I mean by that is that, the family is actually a product of the, depending on the number of degrees of freedom. So, as the degrees of freedom are changing, your t-distribution will essentially change.

**(Refer Slide Time: 07:49)**

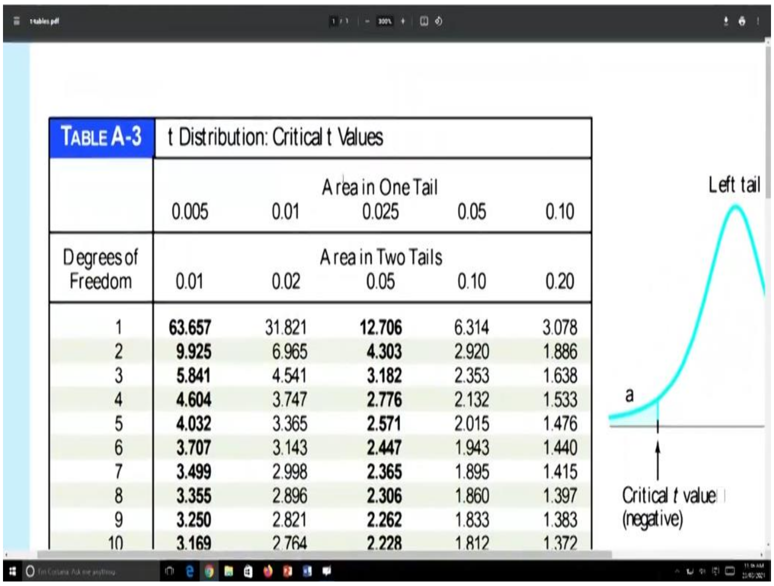
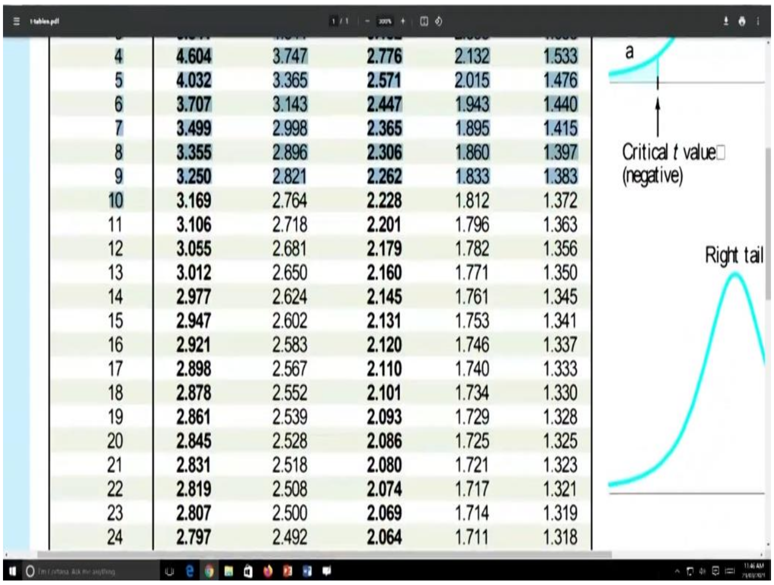


TABLE A-3 t Distribution: Critical t Values

| Degrees of Freedom | 0.005 (one tail) / 0.01 (two tails) | 0.01 / 0.02 | 0.025 / 0.05 | 0.05 / 0.10 | 0.10 / 0.20 |
|---|---|---|---|---|---|
| | Area in One Tail 0.005 | 0.01 | 0.025 | 0.05 | 0.10 |
| | Area in Two Tails 0.01 | 0.02 | 0.05 | 0.10 | 0.20 |
| 1 | **63.657** | 31.821 | **12.706** | 6.314 | 3.078 |
| 2 | **9.925** | 6.965 | **4.303** | 2.920 | 1.886 |
| 3 | **5.841** | 4.541 | **3.182** | 2.353 | 1.638 |
| 4 | **4.604** | 3.747 | **2.776** | 2.132 | 1.533 |
| 5 | **4.032** | 3.365 | **2.571** | 2.015 | 1.476 |
| 6 | **3.707** | 3.143 | **2.447** | 1.943 | 1.440 |
| 7 | **3.499** | 2.998 | **2.365** | 1.895 | 1.415 |
| 8 | **3.355** | 2.896 | **2.306** | 1.860 | 1.397 |
| 9 | **3.250** | 2.821 | **2.262** | 1.833 | 1.383 |
| 10 | **3.169** | 2.764 | **2.228** | 1.812 | 1.372 |

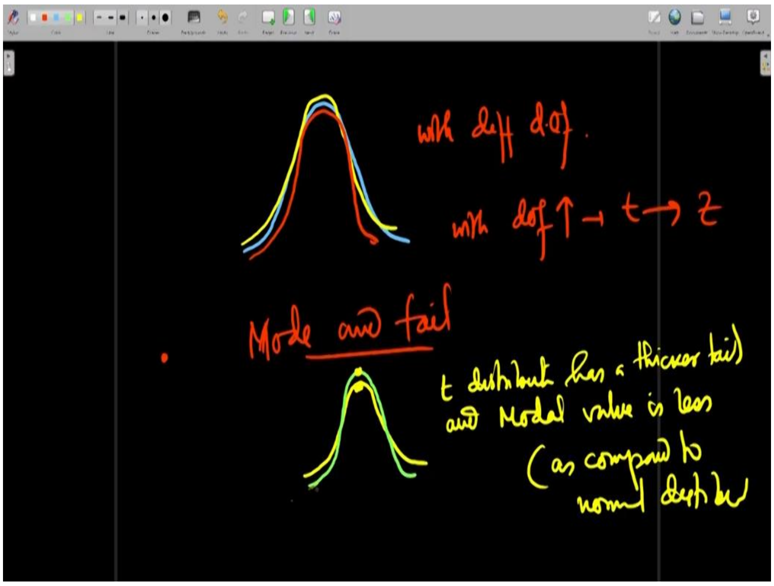




| Degrees of Freedom | 0.01 | 0.02 | 0.05 | 0.10 | 0.20 |
|---|---|---|---|---|---|
| 4 | **4.604** | 3.747 | **2.776** | 2.132 | 1.533 |
| 5 | **4.032** | 3.365 | **2.571** | 2.015 | 1.476 |
| 6 | **3.707** | 3.143 | **2.447** | 1.943 | 1.440 |
| 7 | **3.499** | 2.998 | **2.365** | 1.895 | 1.415 |
| 8 | **3.355** | 2.896 | **2.306** | 1.860 | 1.397 |
| 9 | **3.250** | 2.821 | **2.262** | 1.833 | 1.383 |
| 10 | **3.169** | 2.764 | **2.228** | 1.812 | 1.372 |
| 11 | **3.106** | 2.718 | **2.201** | 1.796 | 1.363 |
| 12 | **3.055** | 2.681 | **2.179** | 1.782 | 1.356 |
| 13 | **3.012** | 2.650 | **2.160** | 1.771 | 1.350 |
| 14 | **2.977** | 2.624 | **2.145** | 1.761 | 1.345 |
| 15 | **2.947** | 2.602 | **2.131** | 1.753 | 1.341 |
| 16 | **2.921** | 2.583 | **2.120** | 1.746 | 1.337 |
| 17 | **2.898** | 2.567 | **2.110** | 1.740 | 1.333 |
| 18 | **2.878** | 2.552 | **2.101** | 1.734 | 1.330 |
| 19 | **2.861** | 2.539 | **2.093** | 1.729 | 1.328 |
| 20 | **2.845** | 2.528 | **2.086** | 1.725 | 1.325 |
| 21 | **2.831** | 2.518 | **2.080** | 1.721 | 1.323 |
| 22 | **2.819** | 2.508 | **2.074** | 1.717 | 1.321 |
| 23 | **2.807** | 2.500 | **2.069** | 1.714 | 1.319 |
| 24 | **2.797** | 2.492 | **2.064** | 1.711 | 1.318 |

So, for example, let us see, if I actually use a t-table, you will see that your t-distribution, all the different; this is a t-distribution of critical t values; you will see that for different degrees of freedom, you are actually getting all the different critical values. Now, how it will look like if I actually try to draw this t-distribution?

**(Refer Slide Time: 08:11)**




So, it will be like; t-distribution actually looks like a normal distribution, but it is going to be like this; I will use a different colour; so, many different curves, you may just have. So, it all depends on what is the number of degrees of freedom. And we will see how these distributions actually vary with different degrees of freedom. Now, as we see that t-distribution is kind of similar to normal distribution, is an approximation to normal distribution.

So, as we improve the degrees of freedom, as the degrees of freedom increases, the t-distribution approximates; with dof increasing, the t-distribution approximates to z-distribution. Now, what happens here is that, for t-distribution, in general, so, the difference between t and normal distribution is that, the mode and the tail; so, we need to remember this; the mode and the tail, where we see some differences between t and z-distribution.

So, what I mean by that is that, if this is a t-distribution, essentially it means that my t-distribution has a thicker tail and the modal value is less as compared to normal distribution. So, we see that; so, the the modal value for t-distribution is here, whereas, that for normal distribution is here. And the tail is, for normal distribution is a thinner tail, and for t-distribution is a thicker tail.

So, these are very important when we are actually going to do the significance testing. How it will be like? So, if we have a thicker tail, so, it will be like, the more number of observation will be on the extremes. So, that is what, given a t-distribution or normal distribution, you would prefer a normal distribution, because it has a thinner tail; but t-distribution, but nonetheless, you have to use t-distribution when you have the population parameters not known.

So, in the next lecture, we are going to see how this tail, if it is a thicker tail or if it is a thinner tail, how it actually matters when you are actually doing the significance testing. Thank you.