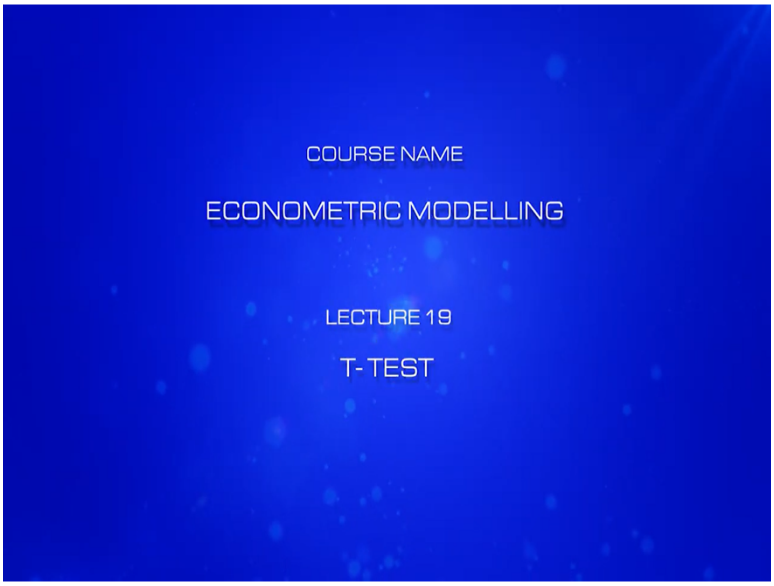
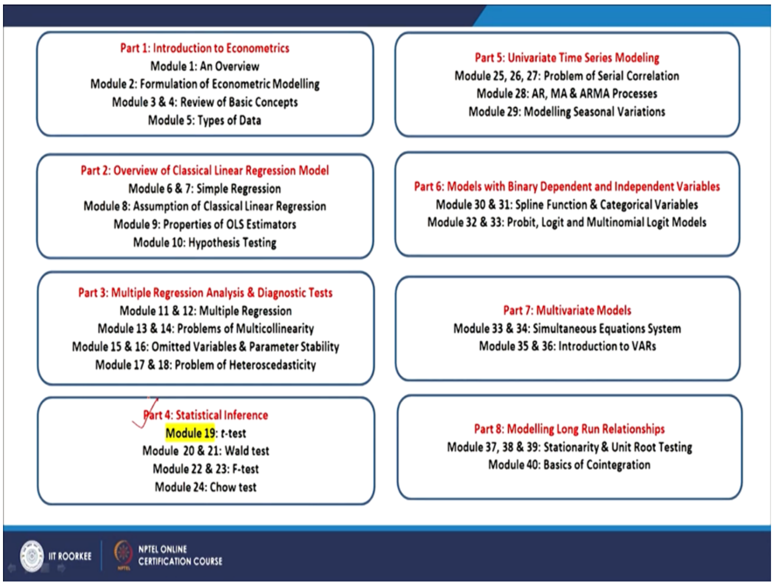
Econometric Modelling
Professor Sujata Kar
Department of Management Studies
Indian Institute of Technology, Roorkee
Lecture 19
T - Test

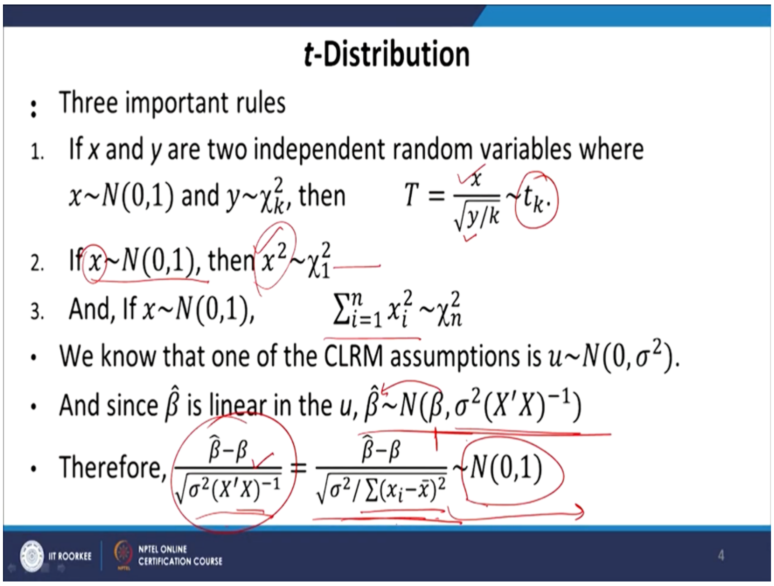
(Refer Slide Time: 00:11)





Hello, and welcome back to the course on Econometric Modelling. Today, I begin with Module 19 which is the first module under Part 4, which deals with statistical inferences. So, the first test we are going to consider is the t-test. So Module 19 on t-test.

(Refer Slide Time: 00:43)



First of all, we talk about t-distribution. There are three important rules which we need to remember in order to understand how the t-ratios are actually formed. So, if x and y are two independent random variables, where x is a standard normal variable, so it follows a normal distribution with 0 mean and variance one, y follows a chi-square distribution with k degrees of freedom, then $\frac{x}{\sqrt{y/k}}$ follows our t-distribution with k degrees of freedom.

So this is our t-ratio, which is obtained as a ratio of a standard normal variable and a chi-square variable divided by k and the entire thing under root. So, alternatively, the other rules we need to remember are first of all if x follows a standard normal distribution, that is normal distribution with mean 0 and variance 1, then x^2 follows a chi-square distribution with 1 degree of freedom.

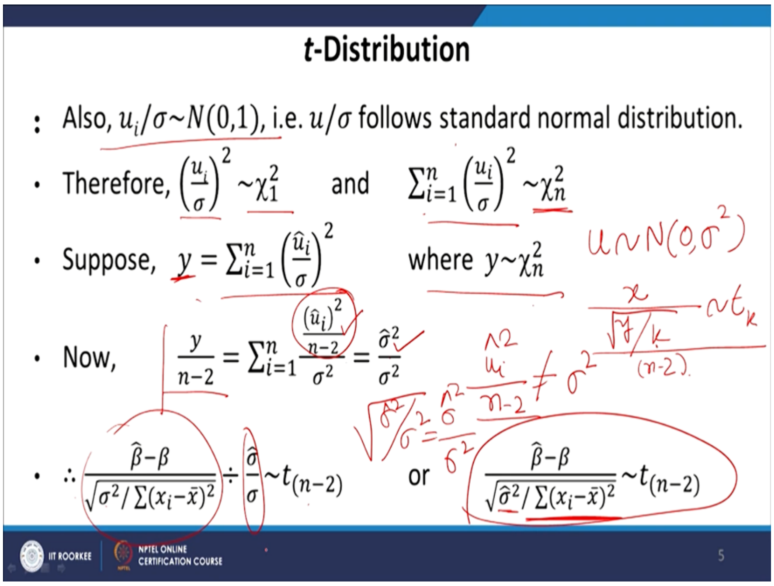
And if x follows a standard normal distribution, then (refer slide time: 2:02). So, a standard normal distribution when squared follows a chi-square distribution, and this chi-square distribution when summed up over an observation follows a chi-square distribution with n degrees of freedom. So these are the three rules.

Now, we know that one of the CLRM assumptions says (refer slide time: 2:32). We have also discussed this previously that since (refer slide time: 2:40).

Alternatively, (refer slide time 2:55- 3:23), when we do not write it in matrix form in the first place and this is also the case when we deal with simple linear regression.

So, we have only one independent variable. Suppose the independent variable is x then we would be writing it in this format. Both are the same thing. In the case of multiple regression analysis, we use it in matrix form otherwise in its simpler form and both follow a standard normal distribution with mean 0 and variance 1.

(Refer Slide Time: 03:52)



Now, we note that since (refer slide time: 3:57). Therefore, as I have just mentioned that if x follows a standard normal distribution, x^2 would follow the chi-square distribution with one degree of freedom.

So, (refer slide time: 4:26), then this follows a chi-square distribution with n degrees of freedom.

Now, suppose we call this y, where y follows a chi-square distribution with n degrees of freedom. We are denoting it by y because just in the previous slide, I mentioned that if x follows a standard normal distribution, and y follows chi-square distribution, then with k degrees of freedom, then $\frac{x}{\sqrt{y/k}}$ follows t-distribution with key degrees of freedom. So, this is something we are going to actually prove that how does it happen.

Now, if y follows a chi-square distribution with n degrees of freedom, then (refer slide time: 5:33). This is a biased estimator of the population error variance, and an unbiased estimator of the population error variance is $n - 2$, when we have only 2 independent variables, or rather 1 constant and 1 independent variable.

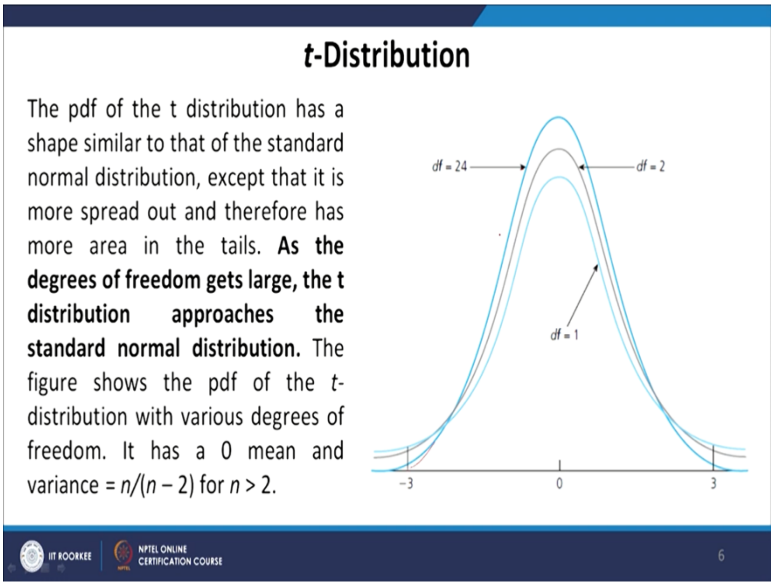
So, when we are dividing it by $n - 2$, we are replacing this thing with sigma hat square, (refer slide time 6:44).

Y follows a chi-square distribution with $n - 2$, degrees of freedom, and $\frac{y}{n-2}$ would follow our t-distribution with $n - 2$ degrees of freedom. So we are having actually in the denominator $\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}$.

So, this is actually equal to (refer slide time: 7:41- 8:07).

So we are actually dealing with it with respect to a single independent variable. And the specific reason for this is that t-statistic or t-test is actually applied to test for a single hypothesis under the assumption that other things are actually not tested for.

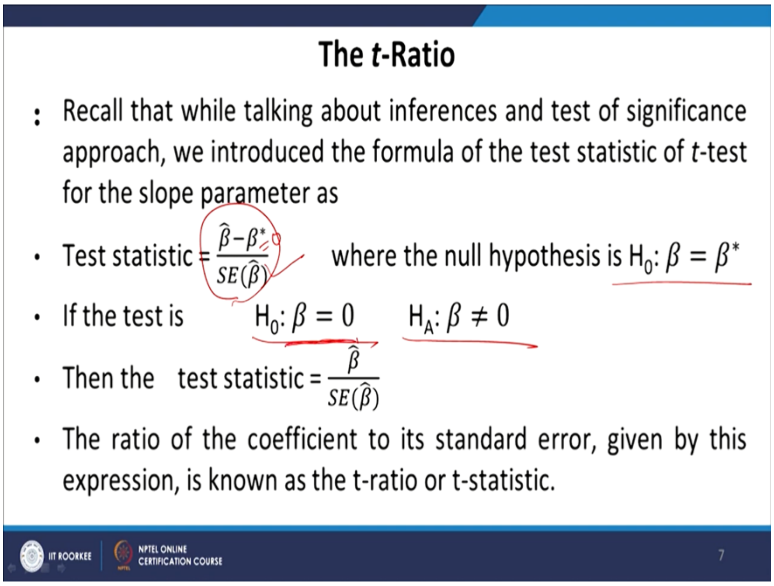
(Refer Slide Time: 08:29)



Now, talking about the pdf of the t-distribution, it has a shape similar to that of the standard normal distribution, except that it is more spread out and therefore, has more area in the tails. As the degrees of freedom get large, the t-distribution approaches the standard normal distribution.

The figure shows that the pdf of the t-distribution with various degrees of freedom, how do they move or behave. So, as you can see that the degrees of freedom increase, the t-distribution actually behaves more like that of the standard normal distribution. It has a mean 0 and variance $\frac{n}{n-2}$ for n greater than 2.

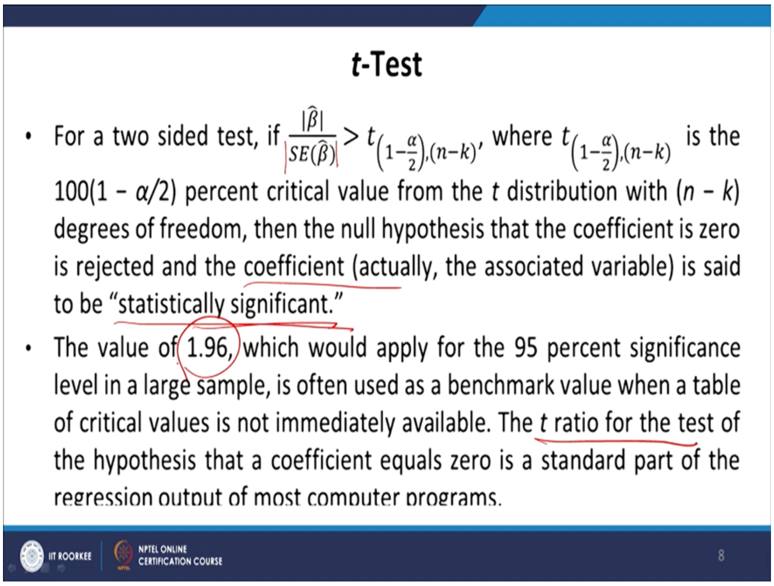
(Refer Slide Time: 09:17)



Now, talking about the t-ratio, the t-ratio that we have just derived, recall that by talking about inferences and test of significance approach, we introduce the formula of the test statistic of t-test for the slope parameter (refer slide time: 9:35- 10:03).

The ratio of the coefficient to its standard error given by this expression is known as the t-ratio or t-statistic. So, what we have just derived in the previous slide is called the t-ratio. And depending on the value of β, under the null hypothesis, we will be forming our t-ratio or t-statistic.

(Refer Slide Time: 10:25)



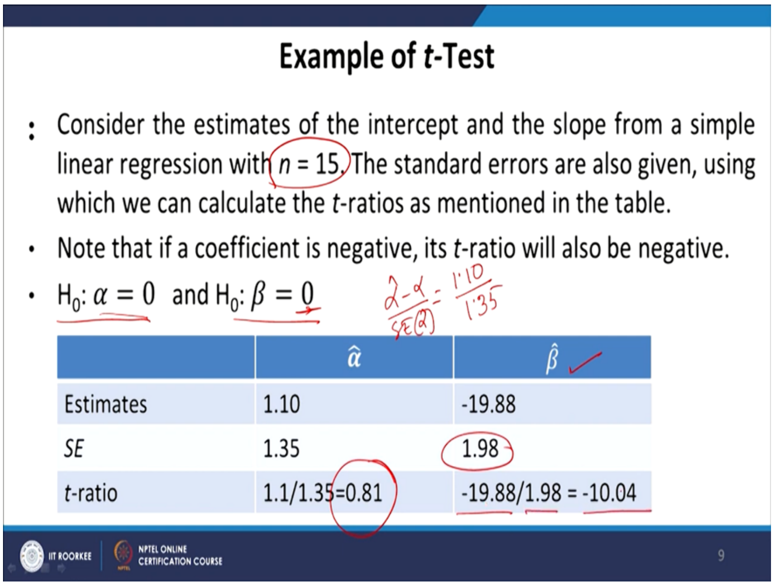
For a two-sided test, (refer slide time: 10:30) should always be a non-negative number, because variance is always a non-negative number. So that is why we put the mod value of the only $\hat{\beta}$ if it is (refer slide time: 10:48), greater than, then the null hypothesis that the coefficient is 0 is rejected and the coefficient, actually the associated variable, is said to be statistically significant.

So what we mean by this is that if we reject the null hypothesis, then first of all the coefficient is statistically significant. And having a statistically significant coefficient implies that the explanatory variable is statistically significant, which implies that the contribution of that explanatory variable in explaining variations in the dependent variable is statistically significant, significantly different from 0.

The value of 1.96, which would apply for the 95 percent significance level in a large sample is often used as a benchmark value when a table of critical values is not immediately available. The t-ratio for the test of the hypothesis that a coefficient equals 0 is a standard part of the regression output of most computer programs.

So, most computer programs provide the regression results with their t-ratios and for large samples since their t-test approaches the standard normal distribution. So, we can use 1.96, which actually corresponds to the 5 percent significance level of the standard normal distribution as also the 5 percent significance level for the t-distribution provided my sample is large enough.

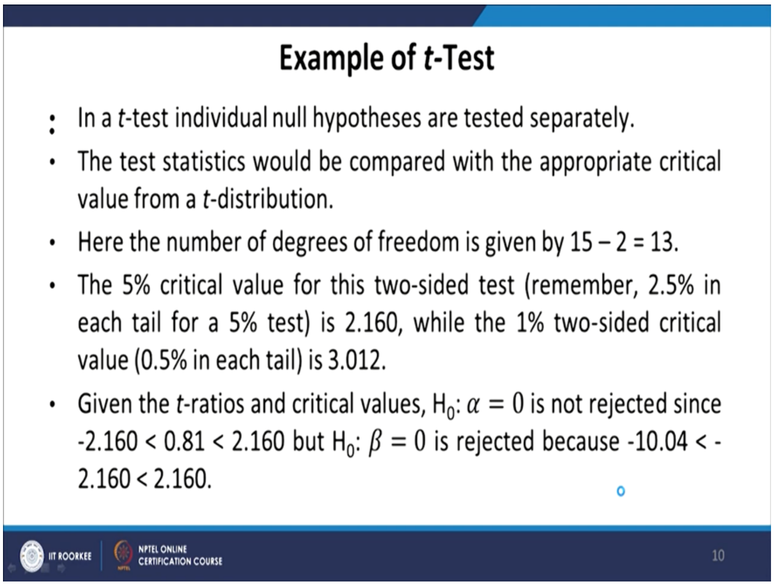
(Refer Slide Time: 12:45)



Now, we take an example of a t-test in a simple linear regression format. Consider the estimates of the intercept and the slope from a simple linear regression with n equals 15. So, we have 15 observations. The standard errors are also given using which we can calculate the t-ratios as mentioned in the table.

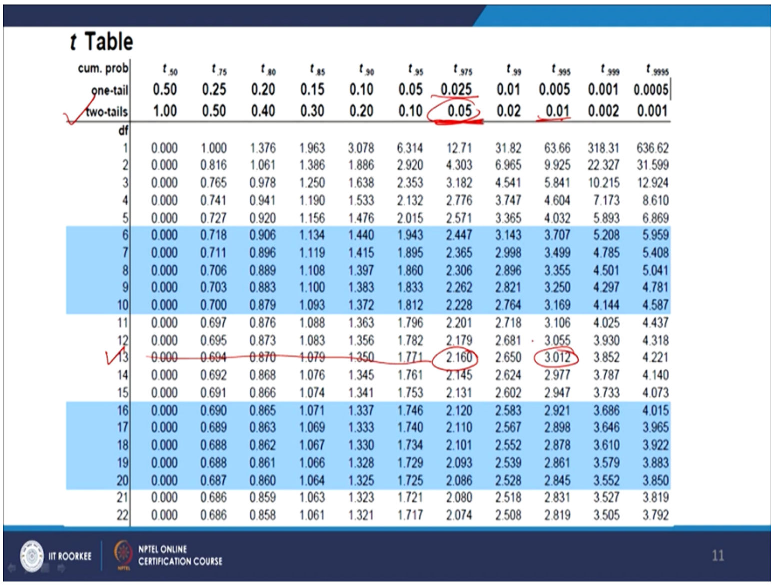
Note that if a coefficient is negative, the t-ratio will also be negative because standard errors are always non-negative numbers. So my hypotheses are, we test them separately. (Refer slide time: 13:23- 14:07).

(Refer Slide Time: 14:06)



So in a t-test, individual null hypotheses are tested separately as I have already mentioned. The test statistics would be compared with the appropriate critical values from a t-distribution. I have a t-table just in the next slide.

(Refer Slide Time: 14:36)



t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792

Here the number of degrees of freedom is given by 15 minus 2 that is 13 and the 5 percent critical value for this two-sided test is actually 2.5 percent in each tail and you can see that from this table, this is my 2 tails.

I have a 5 percent significance level corresponding to 0.025 one tail value, that is 2.5 percent in a particular tail, and as shown in the previous slide that my degrees of freedom are 13. So corresponding to this, I have 2.160 at a 5 percent level of significance at individual tails. And if I consider a 1 percent significance level, then my critical value is 3.012 for individual tails.

(Refer Slide Time: 15:20)

Example of *t*-Test

- In a *t*-test individual null hypotheses are tested separately.
- The test statistics would be compared with the appropriate critical value from a *t*-distribution.
- Here the number of degrees of freedom is given by 15 – 2 = 13.
- The 5% critical value for this two-sided test (remember, 2.5% in each tail for a 5% test) is 2.160, while the 1% two-sided critical value (0.5% in each tail) is 3.012.
- Given the *t*-ratios and critical values, $H_0: \alpha = 0$ is not rejected since -2.160 < 0.81 < 2.160 but $H_0: \beta = 0$ is rejected because -10.04 < -2.160 < 2.160.

Example of *t*-Test

- Consider the estimates of the intercept and the slope from a simple linear regression with *n* = 15. The standard errors are also given, using which we can calculate the *t*-ratios as mentioned in the table.
- Note that if a coefficient is negative, its *t*-ratio will also be negative.
- $H_0: \alpha = 0$ and $H_0: \beta = 0$

	$\hat{\alpha}$	$\hat{\beta}$
Estimates	1.10	-19.88
SE	1.35	1.98
t-ratio	1.1/1.35=0.81	-19.88/1.98 = -10.04

t Table

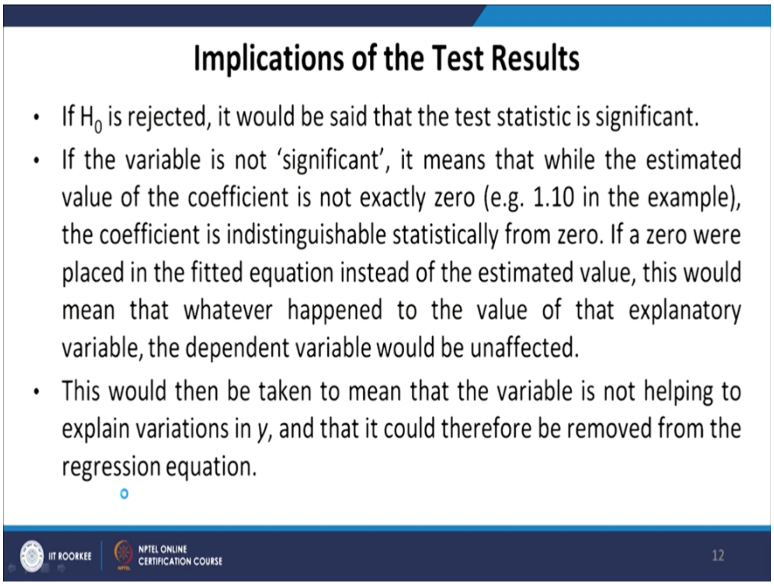
cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792

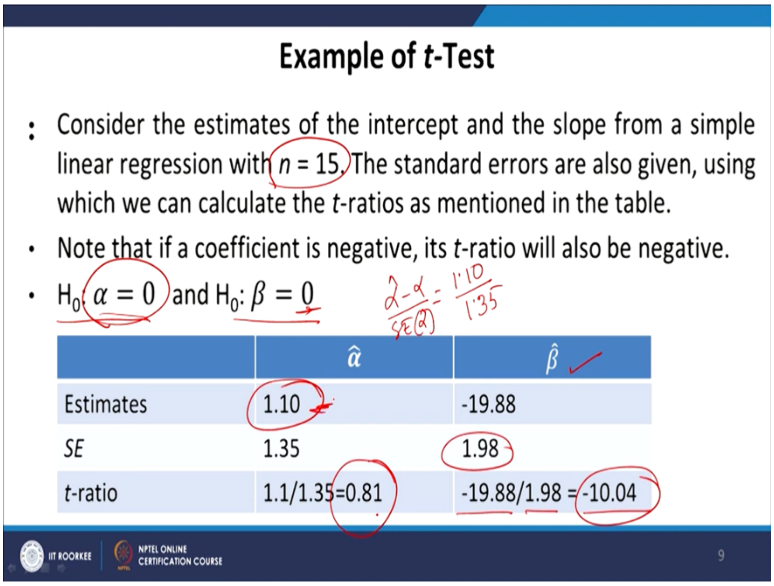
So you can see that this is 2.160 for 5 percent and while the 1 percent two-sided critical value is 3.012, just as we have observed from the table. Given the t-ratios and critical values, $H_0: \alpha = 0$ is not rejected since our calculated value of 0.81 actually lies in the acceptance region or non-rejection region, rather, I must say.

So this is my distribution and this is at 2.5 percent, 2.5 percent. The values are -2.16 and 2.16. Since it takes a value of 0.81, it actually is somewhere here. So it falls in the non-rejection region. We do not reject the null hypothesis, but $H_0: \beta = 0$ is rejected because the calculated value -10.04 is actually less than -2.16, and of course less than plus 2.16 or -10.04 will be somewhere here in the rejection region.

And that is why the second null hypothesis or the null hypothesis concerning β is rejected. So this is the t-table that I had just used. I just want you to note that this t-table has degrees of freedom up to 22. The rest of the degrees of freedom of this table will be presented in a later slide.

(Refer Slide Time: 17:00)

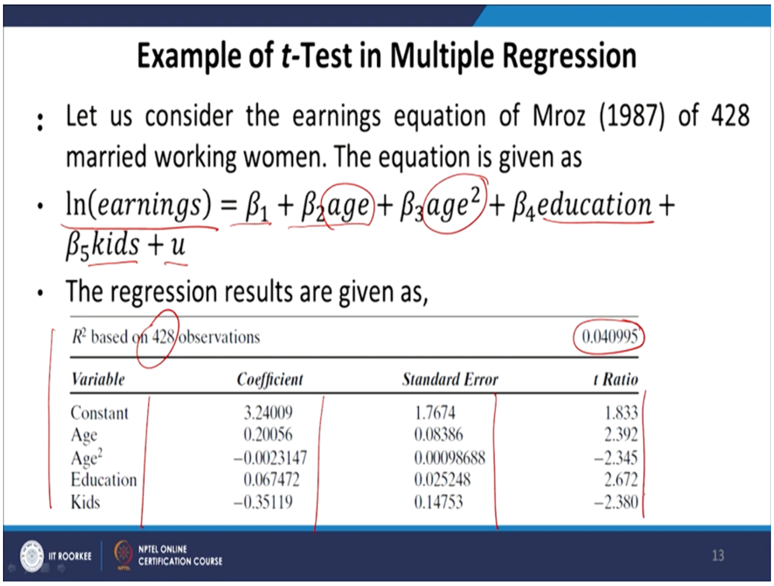




Now, the implications of the t-test results, if H_0 is rejected, it would be said that the test statistic is significant. So this implies that $\hat{\beta}$ is significantly different from 0. If the variable is not significant, it means that while the estimated value of the coefficient is not exactly 0. For example, (refer slide time: 17:23).

So, if the variable is not significant or the estimated coefficient is not significant, it means that while the estimated value of the coefficient is not exactly 0, the coefficient is indistinguishable statistically from 0. So statistically, it is not significantly different from 0. If a 0 were placed in the fitted equation instead of the estimated value, this would mean that whatever happened to the value of that explanatory variable, the dependent variable would be unaffected. This would then be taken to mean that the variable is not helping to explain variations in y and that it could be therefore removed from the regression equation.

(Refer Slide Time: 18:28)

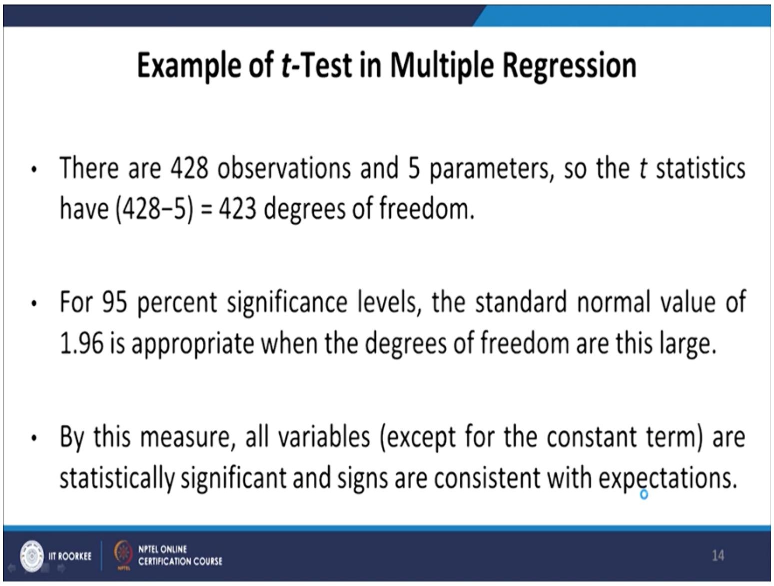


Now, we take another example of a t-test from a multiple regression framework. So, let us consider the earnings equation of Mroz, the study was published in 1987, where he considered 428 married working women, and the equation he estimated with a large number of data. So, the data actually consisted of more women, but out of them 428 were working.

So, this equation specifically pertains to only those women who are working and their number was 428. So, the equation is (refer slide time: 19:08). Then we have education and whether the woman has kids or not, plus an error term.

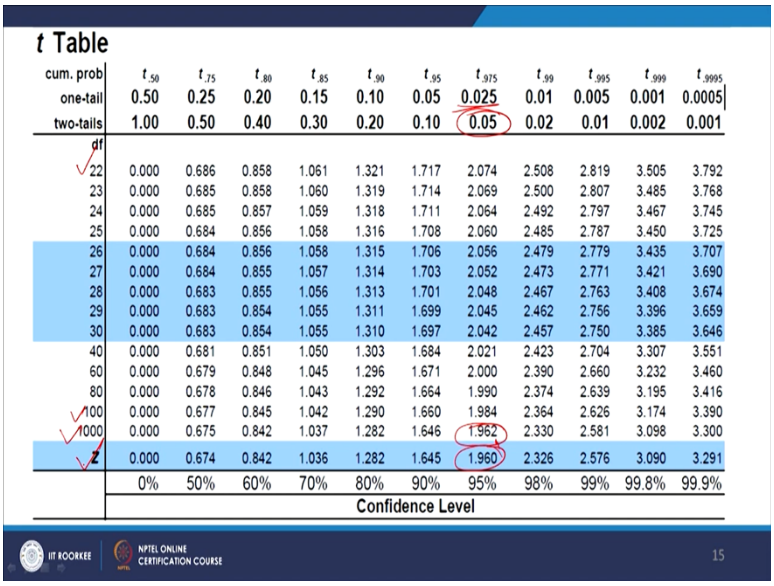
This is the framework of the equation. The regression results are given as shown in the table. So R^2 value is observed to be 0.04, which is actually very small and this R^2 is based on 428 observations. And these are the variables, their coefficients, their standard errors, and the t-ratios that are being calculated. So what do we observe?

(Refer Slide Time: 19:59)



There are 428 observations and 5 parameters, so the t-statistic have $(428 - 5 = 423)$ degrees of freedom. For 95 percent significance levels, their standard normal value of 1.96 is appropriate when the degrees of freedom are this large.

(Refer Slide Time: 20:21)



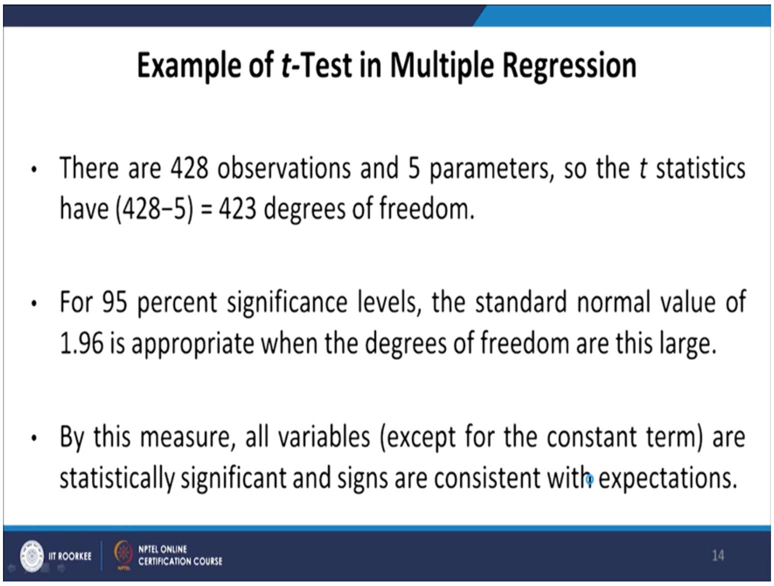
t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Now, I show you the table with degrees of freedom 22 or more. So we are having here sample size which is 428. Degrees of freedom pertaining to 100 are mentioned here. After that, it goes to 1000. Now, again if I consider 2 tail significance levels of 5 percent and 1 tail significance level of 2.5 percent then you can see that pertaining to 1000, this is actually

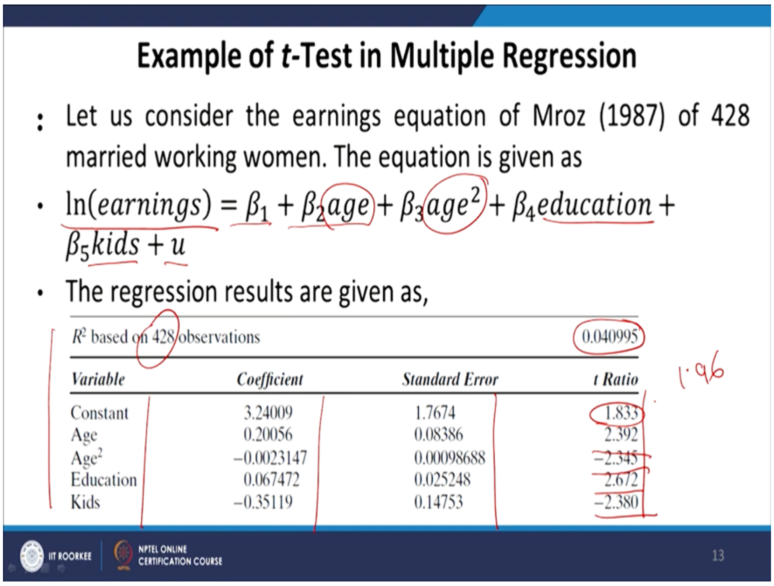
1.962 for 1000 observations. And this is so close to the standard normal critical value which is 1.96.

(Refer Slide Time: 21:17)



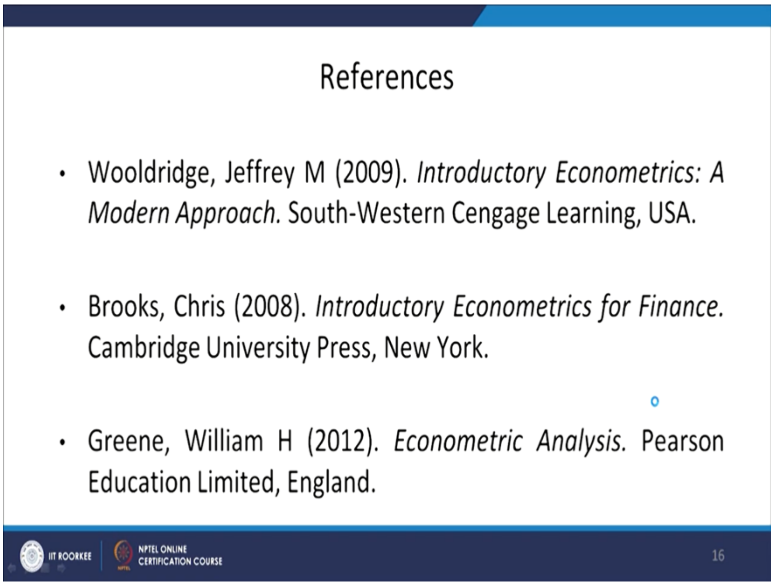
So that is why we can consider it to be 1.96. When the sample goes large, the t-distribution actually becomes equivalent to the standard normal distribution. And these, of course, represent the confidence level. Now, we can see that for 95 percent significance levels, the standard normal value of 1.96 is appropriate when the degrees of freedom are this large.

(Refer Slide Time: 21:39)



By this measure, all variables except for the constant term are statistically significant and signs are consistent with expectations. So we will not get into the signs as we are not actually discussing the economic implications of the variables included in the equation, but provided that my critical value is 1.96. So, you can see that all these estimated t-values or t-statistics are greater than 1.96, except for the constant term. All the estimated coefficients are statistically significant. Statistically significant implies that they are significantly different from 0. So this is how we understand t-ratios and t-test in a multiple regression framework.

(Refer Slide Time: 22:17)



These are the references that I have followed. Thank you.