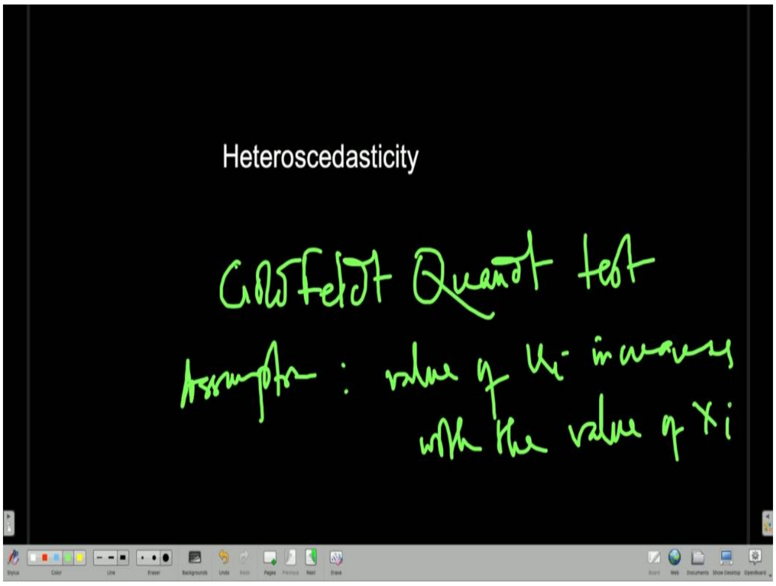
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology - Kharagpur

Module - 8
Lecture - 67
Heteroscedasticity (Contd.)

Hello and welcome back to the lecture on Applied Econometrics. We have been talking about heteroscedasticity. Now, in today's class, we are going to actually familiarise ourselves with certain tests to perform, to know if heteroskedasticity is present or not. So, the tests actually vary, and it varies depending on what assumption we take about the distribution of the error term. So, we can think that the error term is actually distributed with the values of X.

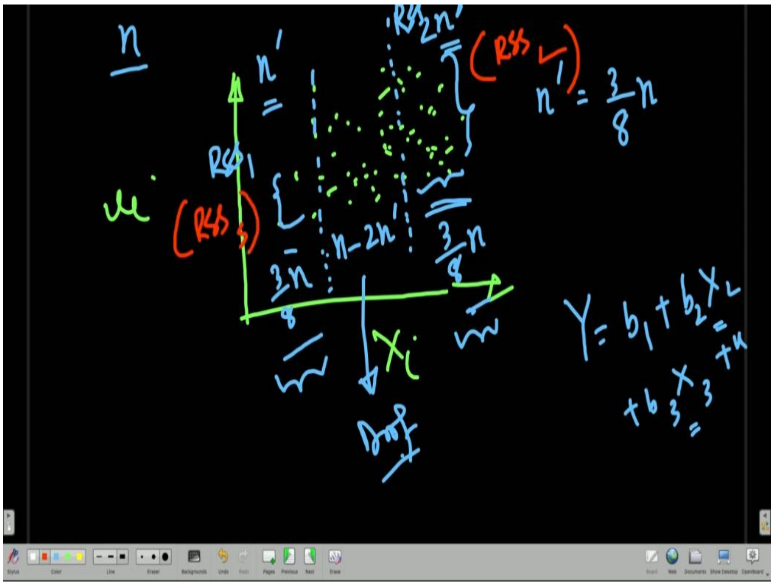
That we have seen previously, like, as X increases, the error term, the value of the error is actually also increasing. So, that we have one kind of a test. Also, we can, like have a less restrictive kind of a situation where we do not assume something like that, and we do a different test. We are going to see these different tests based on these different assumptions. The first test that we perform is called Goldfeld Quandt test or GQ test.

(Refer Slide Time: 01:19)



So, this test is somewhat restrictive. It assumes that the error term is actually distributed with the values of X. So, the assumption is, the value of u i increases with the value of X i. So, if I actually plot this, so, I would have a distribution, something like this; distribution of the error term.

(Refer Slide Time: 02:09)



So, let us say this is the value of u i; and this is the value of X i; and we will see something like this. If we have to perform this GQ test, so, we have to have something like this. We will see how in different situations, whether we can apply GQ test, Goldfeld Quandt test or not; but this is primarily the assumption that we have to make before we actually perform the Goldfeld Quandt test.

Now, if this is how the error term looks like, and if we want to understand if heteroskedasticity is present or not, so, what we need to do is, simply, we need to see, let us say, we can divide the whole range of observations into different parts. Let us say I divide it here. Let me use a different colour. I create one wall here. I create another divider here. And what I will try to do is that, I will try to see if the variance of this part and variance of this part are significantly different.

So, if the variances are significantly different, say at the lower value of X and at the higher value of X, so, then we can say that, well, there is heteroscedasticity. And how we do it? There are certain rules we will follow. Usually, we take this part as; let us say this is called, let us say the total number of observations is n; and I take some n prime from here; and I take some n prime from here.

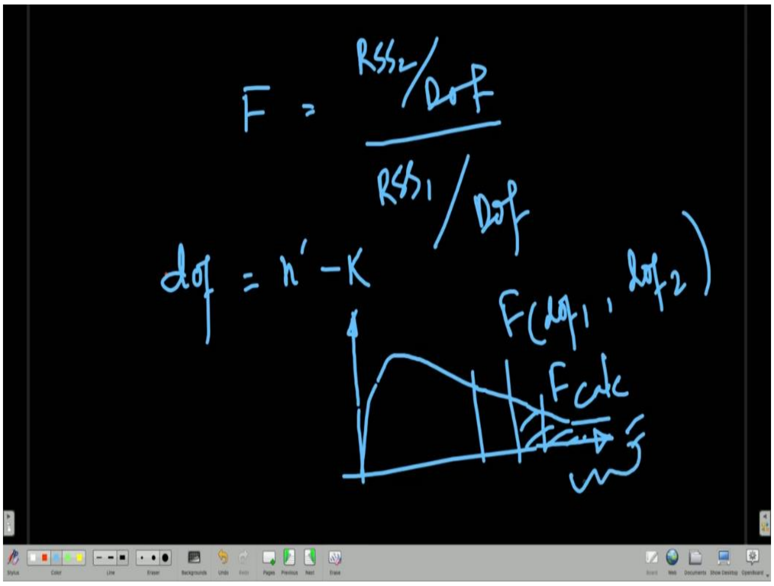
So, these are same; the number of observations here and number of observations there, we take same. And this n prime usually is equal to 3 by 8 of n; so, three-eighth of the total number of observation here, and three-eighth of the total number of observations at the end.

So, these two we take. So, what we have left out is this; 3 by 8 n here, 3 by 8 n there. Now, how we actually do the test?

So, we actually perform a F-test here to understand if the variances are actually significantly different or not. And to do that, we run a regression. So, let us say, in original regression equation, we have something like beta 2 X 2, beta 3 X 3 and some error term. Now, how we do it? We basically run the regression first for this group here, and second for this group here. We run the same regression, we have the same explanatory variable.

So, that way, we will see the degrees of freedom that we calculate for both the regression equations are going to be same. So, how we will actually see if the two, the variances of the error are similar or not?

(Refer Slide Time: 05:33)

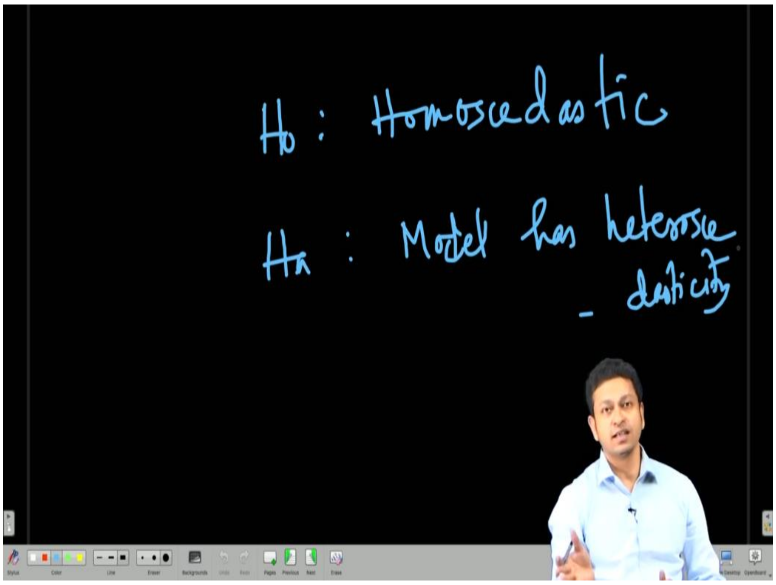


So, what we will do is, simply we will do a F-test, variant of F-test, where we will do F is equal to, say residual sum of square RSS 2 by corresponding Dof and by RSS 1 residual sum of square by corresponding Dof. So, how we do for F-test? We take the larger one in the numerator. So, this is the part where we have larger variance; so, let us call it RSS 2, or you can also call it, say RSS large. Let me use a different colour, it looks a little cluttered.

Let us say you also write RSS large. And the previous one, the smaller one, we call RSS 1, or we can also call it RSS small, residual sum of square small. Now, how we do it is, we basically divide this corresponding residual sum of square with the degrees of freedom, and then we get the F-statistic. Now, what is the degrees of freedom here? So, degrees of freedom here is also pretty straightforward, since we have in both the cases we have the number of observations same, which is n prime; and the K is basically, the number of explanatory variable is also same.

So, the Dof here is, Dof in both the cases is going to be n prime minus K. Now, that is how we calculate the F-statistic. And then, we simply do a F-test. And my F calculated, depending on whether it is here or here, we take a decision whether to reject or not reject the null hypothesis.

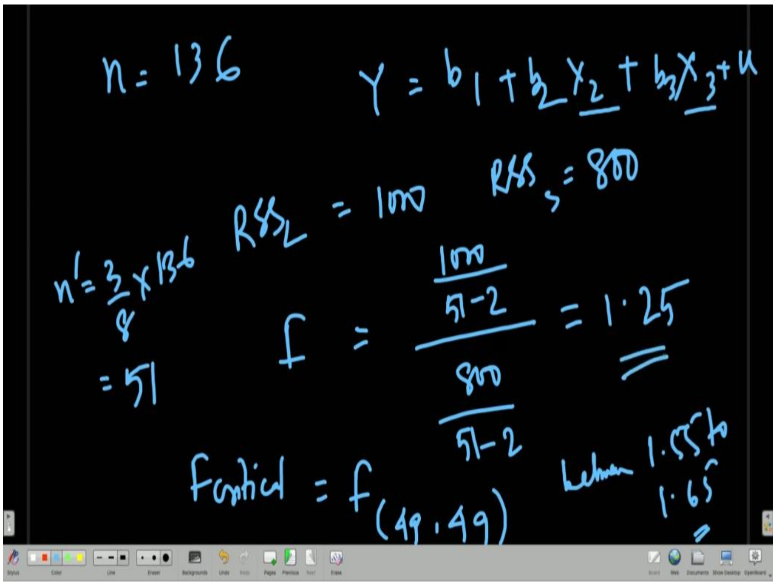
(Refer Slide Time: 07:27)



So, null hypothesis here is, H nought is going to be that RSS 1 and RSS 2; or basically, we can simply say the model is homoskedastic; or there is no heteroskedasticity. Whereas H 1 is that the model has heteroskedasticity. So, if I have my F calculated; let us say F calculated is here. So, that means it is on the right side of the F critical. So, if it is on the right side of my F critical, so, then I basically reject the null that the models are, basically, it is homoskedastic.

So, the variances are basically equal. So, that is not the case if my F calculated is going to be on the right side of my F critical. So, that is about it. And let us do a small problem, small hands-on.

(Refer Slide Time: 08:52)

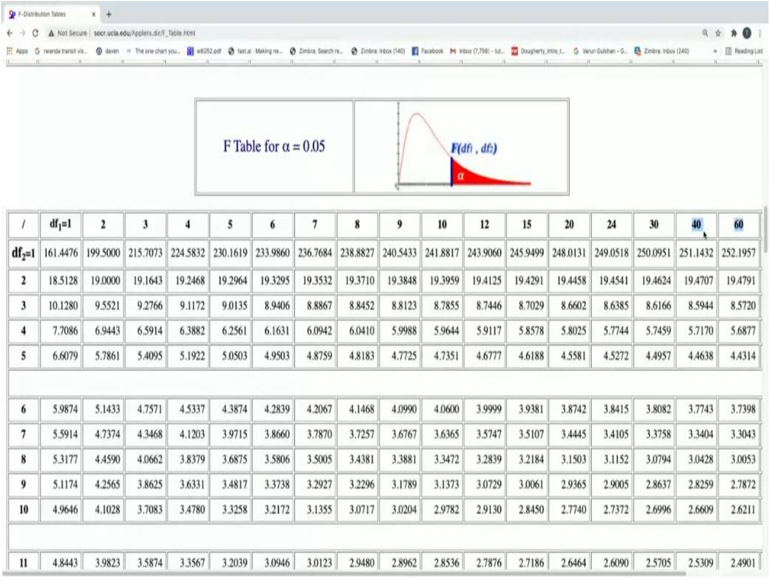


I will give you this problem. Let us say I am running a regression with n = 136. And let us say I have this model. I do not need to specify the model; I only need the degrees of freedom from here; b 1, b 2 X 2; just the same; b 3 X 3 plus the error term. And then I have, let us say the residual sum of square, RSS large is going to be 1000 or RSS 2; or RSS small or RSS 1 is going to be 800. Now, I have to say if this model has heteroskedasticity or not.

So, how do I do that? I simply get the F-statistic. My F-statistic calculated is going to be RSS 2. And my n is say 136; so, my n prime is going to be three-eighth of my n. So, I basically took this number for my own convenience. So, r is going to be; if I multiply it, it is going to be, I think 51. So, n prime is 51. So, I have 51 and my K here; my degrees of freedom. So, basically, to get my degrees of freedom, the number of explanatory variables, which is K = 2.

So, I have 51 - 2. And here also, I have 800 by 51 - 2. We know that the degrees of freedom are basically same. So, essentially, what I get is, I get 1.25. So, that is my F calculated. Now, I have to get my F critical. And my F critical is basically; what is the degrees of freedom for F critical? Is f 49, 49. So, in both cases, I have same degrees of freedom, which is 49 here. Now, how do I obtain this? I have to basically go get the F-table.

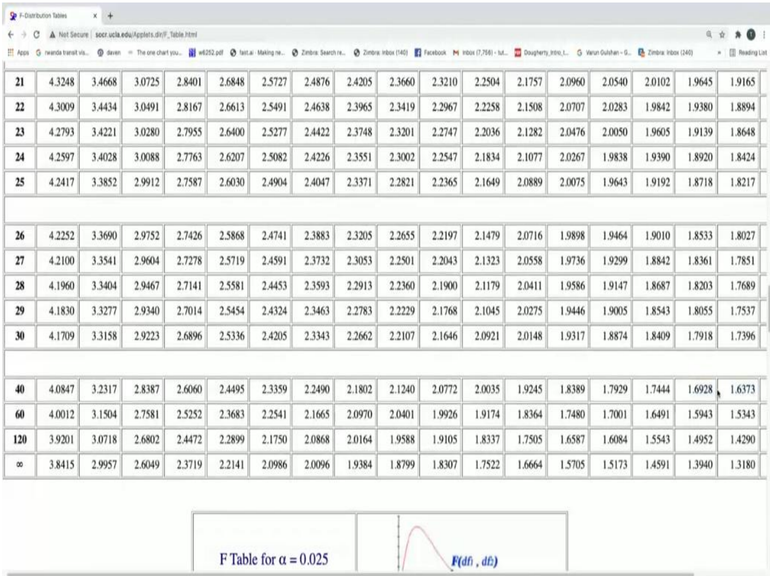
(Refer Slide Time: 11:13)



/	df_1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60
df_2=1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817	243.9060	245.9499	248.0131	249.0518	250.0951	251.1432	252.1957
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959	19.4125	19.4291	19.4458	19.4541	19.4624	19.4707	19.4791
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901

So, let me actually get the F-distribution table, and I usually take a significance level alpha is equal to 0.05. So, you can see that I do not have exact degrees of freedom is equal to 49, but I have the range 40 to 60. So, we can get something like in between, close to 50.

(Refer Slide Time: 11:46)



21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8894
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180

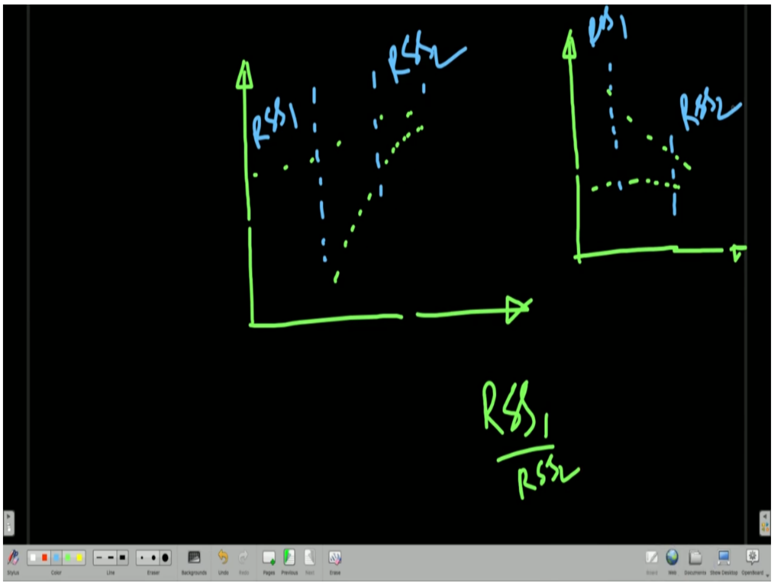
And here also, I will see that we do not have exact value for 49, but we have the value for 40 and 60. So, basically, our F-statistic would lie somewhere in between these values. So, this is for 40 and 60. So, one could be in between these 2; so 1.69 or 1.63. So, let us say it is going to be 1.65. And this one, 1.59 and 1.53; let us say it is going to be 1.55. So, my F critical is going to be somewhere in between 1.55 to 1.65.

So, we can just get an approximate value. And that is enough for us, because our F calculated is 1.25, which is actually smaller than my F critical. So, if my F calculated is smaller than my

F critical, so, then, that means my F calculated is actually going to lie here. So, if my F calculated is going to lie here, so, that means, I actually reject the null hypothesis of homoscedasticity.

So, this is always confusing, heteroscedasticity and homoscedasticity; but yeah, basically heteroscedasticity means that the, sort of disproportion and variance, and that is basically the alternative hypothesis. So, essentially, we say that the models are not of equal variance. And this is how we actually conduct this Goldfeld Quandt test. So, there are few more things to talk about when we talk about Goldfeld Quandt test. So, one thing is that; so, we kind of said that the error term distribution is going to look like this, but there could be other situations also.

(Refer Slide Time: 13:34)



So, there could be a situation where you can have your error term inversely distributed with X. So, let us say this is looking like this; it is more like this or it could be something like this. So, there could be different types of distribution of the error term. And in case, suppose your variance is actually decreasing with the increasing value of X, so, what we do is, we simply, basically change RSS 1 and RSS 2.

Initially we did RSS 2 by RSS 1, but in these cases, we will do RSS 1 by RSS 2. So, essentially, if this is your; here it is; this is your RSS 1; so, initially, it was in the denominator, but now, it will go in the numerator. So, essentially, you remember that when you calculate the F-statistic, you need to have the higher value on the numerator here. So, basically, you do RSS 1 by RSS 2. Here also, you do RSS 1 by RSS 2.

So, RSS 2 goes in the denominator. So, this is how we do the Goldfeld Quandt test. So, another type; so, this is a numerical variable; we kind of got all the values of the variable; but it can also happen for qualitative variable, which is basically the dummy variable. So, let us say you are trying to see the abortion rate for different religion groups; like religion actually influences abortion and if the kind of heteroscedasticity is present for that.

Similarly, if you want to see the wage for male, females; so, basically use up dummy variables. So, even then, you can use this GQ test, and it is a good test to do in that kind of situations. So, what you have to do? You again basically get the residual sum of squares for different dummy categories, and you do the F-test. So, in that case, you do not have to follow that 3 by 8 rule.

So, that is basically the Goldfeld Quandt test with certain restrictions. So, in the next lecture, we are going to see the test for heteroscedasticity where we do not have any restriction. Thank you.