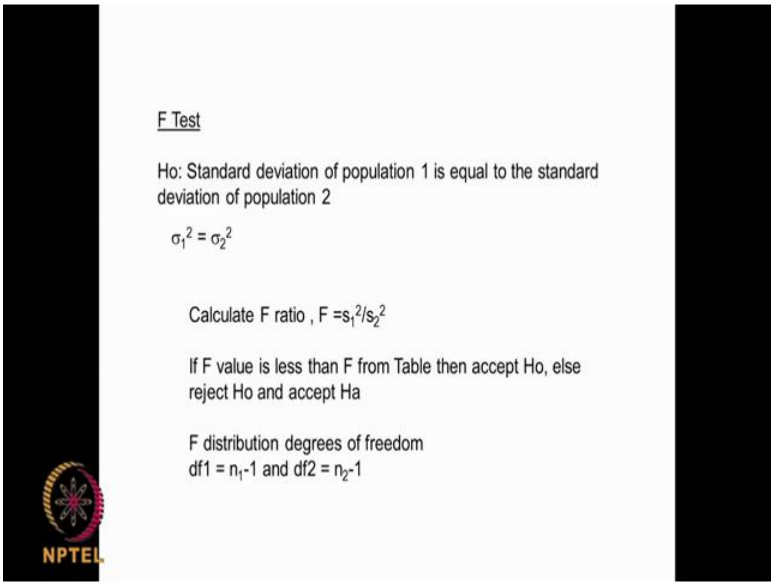
Biostatistics and Design of Experiments

Prof. Mukesh Doble
Department of Biotechnology
Indian Institute of Technology, Madras

Lecture - 13
F- tests

Hello, come to the course on Biostatistics and Design of Experiments. Today we will talk more about the F-test. Yesterday I introduced what is F-test? And how useful it is? So, we look at some more problems on this F-test.

(Refer Slide Time: 00:27)

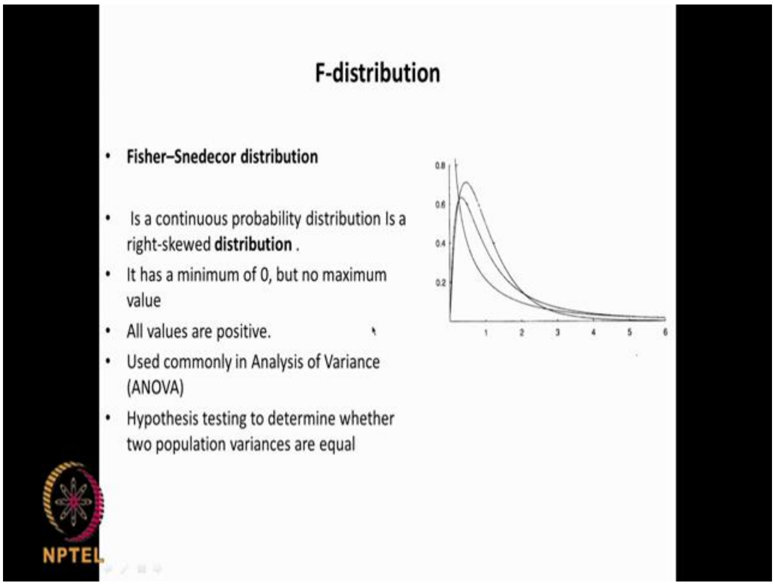


So, F-test what does it do, the Ho that is the null hypothesis the standard deviation of population 1 is equal to standard deviation of population 2, that is

$$\sigma_1^2 = \sigma_2^2$$

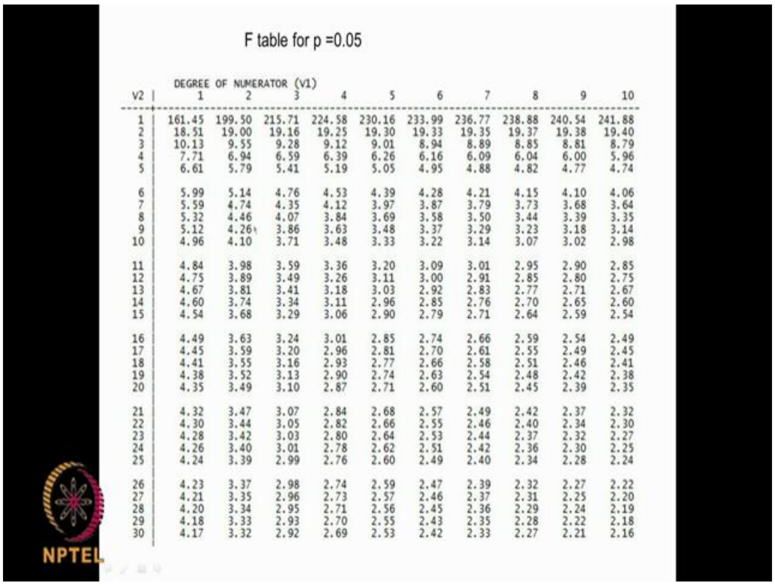
. So, the alternate hypothesis could be the standard deviations are not equal or standard deviation of population 1 is greater or standard deviation of population 2 < 1 and so on actually. Here we calculate something called F ratio, F is given by the variance of the sample 1, variance of sample 2. And if the F value is < what is reported in the table, then we do not reject the null hypothesis. The F value calculated > F table then we reject the null hypothesis and accept the alternate hypothesis. So there is a table for F here also, just like we saw table for t, we saw table for z. So here, there are 2 degrees of freedom because the numerator will have 1 degree of freedom which is n_1-1, if n 1 is the number of data points for the numerator sample and for the denominator the degrees of freedom will be n_2-1, where n_2 will be the number of samples in for the sample set 2.

(Refer Slide Time: 01:52)



The F distribution generally looks like this, it has got a 0, but the maximum is not fixed. It has a minimum, but the maximum is not fixed. So, it is a skewed, it is a right skewed distribution like this. It is commonly used if your comparing variances, if your performing something called analysis of variance, if you are doing curve feet that means regression analysis, you want to find out how good feet is ! Then we use this type of F ratio that means variance of 1 sample divided by the variance of the sample 2. So generally, the hypothesis, as I said null hypothesis is the variances are equal actually.

(Refer Slide Time: 02:39)

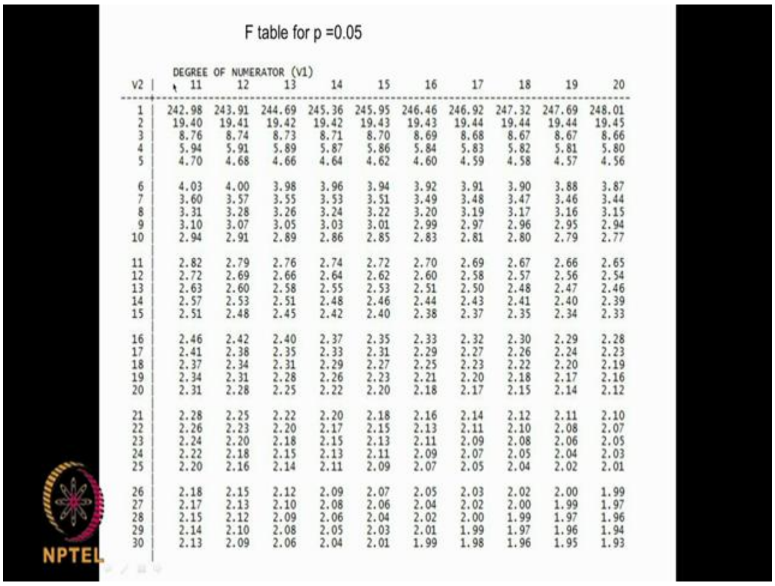

F table for p =0.05

DEGREE OF NUMERATOR (V1)

V2	1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16

That is what is F distribution is all about and I said there is a table, I showed you in the previous class also. So, for the numerator you have like this, along the column degrees of freedom and then for the denominator we have along the rows degrees of freedom. So you look at the F value and this is for p = 0.05 that means 95 % confidence interval. Now for beyond and 10 we have continuation on this table, so it goes like this 20 goes up to 30.

(Refer Slide Time: 03:02)

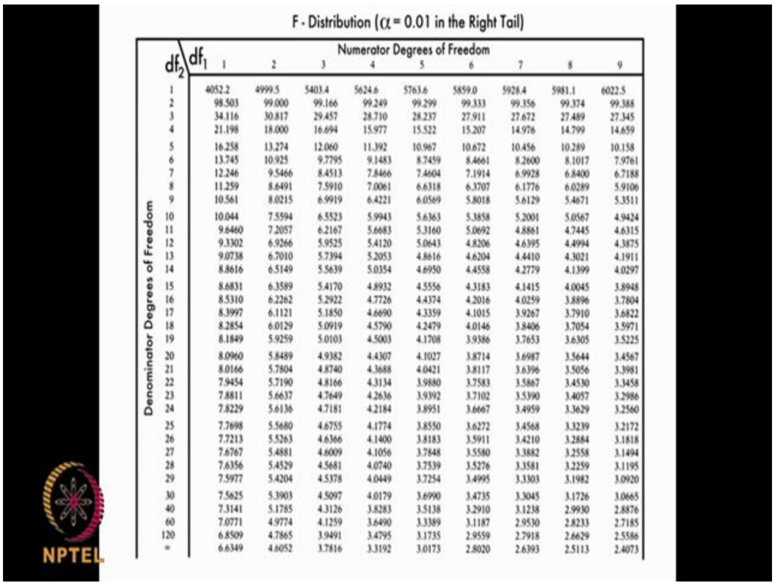

F table for p =0.05

DEGREE OF NUMERATOR (V1)

V2	11	12	13	14	15	16	17	18	19	20
1	242.98	243.91	244.69	245.36	245.95	246.46	246.92	247.32	247.69	248.01
2	19.40	19.41	19.42	19.42	19.43	19.43	19.44	19.44	19.44	19.45
3	8.76	8.74	8.73	8.71	8.70	8.69	8.68	8.67	8.67	8.66
4	5.94	5.91	5.89	5.87	5.86	5.84	5.83	5.82	5.81	5.80
5	4.70	4.68	4.66	4.64	4.62	4.60	4.59	4.58	4.57	4.56
6	4.03	4.00	3.98	3.96	3.94	3.92	3.91	3.90	3.88	3.87
7	3.60	3.57	3.55	3.53	3.51	3.49	3.48	3.47	3.46	3.44
8	3.31	3.28	3.26	3.24	3.22	3.20	3.19	3.17	3.16	3.15
9	3.10	3.07	3.05	3.03	3.01	2.99	2.97	2.96	2.95	2.94
10	2.94	2.91	2.89	2.86	2.85	2.83	2.81	2.80	2.79	2.77
11	2.82	2.79	2.76	2.74	2.72	2.70	2.69	2.67	2.66	2.65
12	2.72	2.69	2.66	2.64	2.62	2.60	2.58	2.57	2.56	2.54
13	2.63	2.60	2.58	2.55	2.53	2.51	2.50	2.48	2.47	2.46
14	2.57	2.53	2.51	2.48	2.46	2.44	2.43	2.41	2.40	2.39
15	2.51	2.48	2.45	2.42	2.40	2.38	2.37	2.35	2.34	2.33
16	2.46	2.42	2.40	2.37	2.35	2.33	2.32	2.30	2.29	2.28
17	2.41	2.38	2.35	2.33	2.31	2.29	2.27	2.26	2.24	2.23
18	2.37	2.34	2.31	2.29	2.27	2.25	2.23	2.22	2.20	2.19
19	2.34	2.31	2.28	2.26	2.23	2.21	2.20	2.18	2.17	2.16
20	2.31	2.28	2.25	2.22	2.20	2.18	2.17	2.15	2.14	2.12
21	2.28	2.25	2.22	2.20	2.18	2.16	2.14	2.12	2.11	2.10
22	2.26	2.23	2.20	2.17	2.15	2.13	2.11	2.10	2.08	2.07
23	2.24	2.20	2.18	2.15	2.13	2.11	2.09	2.08	2.06	2.05
24	2.22	2.18	2.15	2.13	2.11	2.09	2.07	2.05	2.04	2.03
25	2.20	2.16	2.14	2.11	2.09	2.07	2.05	2.04	2.02	2.01
26	2.18	2.15	2.12	2.09	2.07	2.05	2.03	2.02	2.00	1.99
27	2.17	2.13	2.10	2.08	2.06	2.04	2.02	2.00	1.99	1.97
28	2.15	2.12	2.09	2.06	2.04	2.02	2.00	1.99	1.97	1.96
29	2.14	2.10	2.08	2.05	2.03	2.01	1.99	1.97	1.96	1.94
30	2.13	2.09	2.06	2.04	2.01	1.99	1.98	1.96	1.95	1.93

As you can see it starts quite high 161 F value but it keep going down, down to a lot 1.93, if the degrees of freedom for numerator and denominator are quite high. So, generally as you can see here, from a very large value it sort of comes down here. So as a ball park figure for the degrees of freedom it is good if you have about 10 to 12 for the numerator, as well as 10 to 12 for the denominator that means see here almost here. And of course, if you want to have very good comparison you need to have more degrees of freedom. But as a ball park figure just like, in key distribution also I did mention that about a 6 to 7 or 8 degrees of freedom is reasonably ok. Same way here if you are doing an F calculation, I really wish we should think about 10 degrees of freedom for the numerator as well as the denominator, so we will be somewhere here.

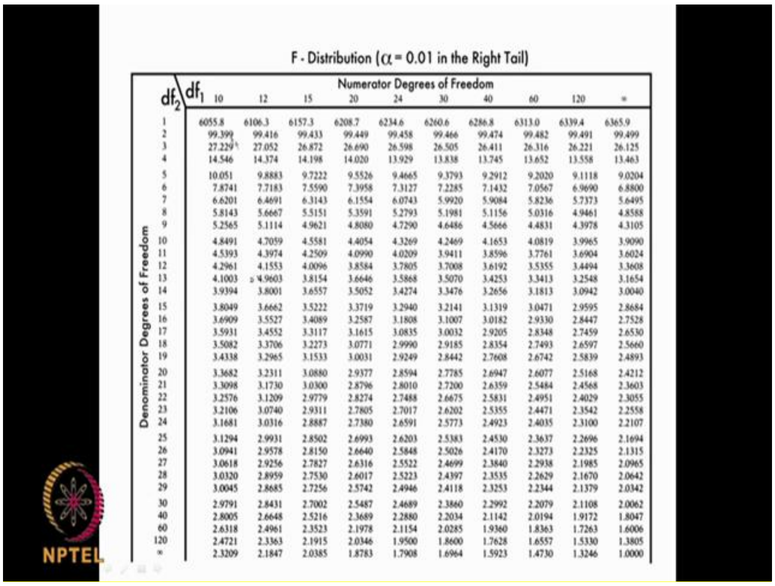
(Refer Slide Time: 04:17)



F - Distribution (α = 0.01 in the Right Tail)

df_2 \ df_1 (Denominator Degrees of Freedom \ Numerator Degrees of Freedom)	1	2	3	4	5	6	7	8	9
1	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158
6	13.745	10.925	9.7795	9.1483	8.7459	8.4661	8.2600	8.1017	7.9761
7	12.246	9.5466	8.4513	7.8466	7.4604	7.1914	6.9928	6.8400	6.7188
8	11.259	8.6491	7.5910	7.0061	6.6318	6.3707	6.1776	6.0289	5.9106
9	10.561	8.0215	6.9919	6.4221	6.0569	5.8018	5.6129	5.4671	5.3511
10	10.044	7.5594	6.5523	5.9943	5.6363	5.3858	5.2001	5.0567	4.9424
11	9.6460	7.2057	6.2167	5.6683	5.3160	5.0692	4.8861	4.7445	4.6315
12	9.3302	6.9266	5.9525	5.4120	5.0643	4.8206	4.6395	4.4994	4.3875
13	9.0738	6.7010	5.7394	5.2053	4.8616	4.6204	4.4410	4.3021	4.1911
14	8.8616	6.5149	5.5639	5.0354	4.6950	4.4558	4.2779	4.1399	4.0297
15	8.6831	6.3589	5.4170	4.8932	4.5556	4.3183	4.1415	4.0045	3.8948
16	8.5310	6.2262	5.2922	4.7726	4.4374	4.2016	4.0259	3.8896	3.7804
17	8.3997	6.1121	5.1850	4.6690	4.3359	4.1015	3.9267	3.7910	3.6822
18	8.2854	6.0129	5.0919	4.5790	4.2479	4.0146	3.8406	3.7054	3.5971
19	8.1849	5.9259	5.0103	4.5003	4.1708	3.9386	3.7653	3.6305	3.5225
20	8.0960	5.8489	4.9382	4.4307	4.1027	3.8714	3.6987	3.5644	3.4567
21	8.0166	5.7804	4.8740	4.3688	4.0421	3.8117	3.6396	3.5056	3.3981
22	7.9454	5.7190	4.8166	4.3134	3.9880	3.7583	3.5867	3.4530	3.3458
23	7.8811	5.6637	4.7649	4.2636	3.9392	3.7102	3.5390	3.4057	3.2986
24	7.8229	5.6136	4.7181	4.2184	3.8951	3.6667	3.4959	3.3629	3.2560
25	7.7698	5.5680	4.6755	4.1774	3.8550	3.6272	3.4568	3.3239	3.2172
26	7.7213	5.5263	4.6366	4.1400	3.8183	3.5911	3.4210	3.2884	3.1818
27	7.6767	5.4881	4.6009	4.1056	3.7848	3.5580	3.3882	3.2558	3.1494
28	7.6356	5.4529	4.5681	4.0740	3.7539	3.5276	3.3581	3.2259	3.1195
29	7.5977	5.4204	4.5378	4.0449	3.7254	3.4995	3.3303	3.1982	3.0920
30	7.5625	5.3903	4.5097	4.0179	3.6990	3.4735	3.3045	3.1726	3.0665
40	7.3141	5.1785	4.3126	3.8283	3.5138	3.2910	3.1238	2.9930	2.8876
60	7.0771	4.9774	4.1259	3.6490	3.3389	3.1187	2.9530	2.8233	2.7185
120	6.8509	4.7865	3.9491	3.4795	3.1735	2.9559	2.7918	2.6629	2.5586
∞	6.6349	4.6052	3.7816	3.3192	3.0173	2.8020	2.6393	2.5113	2.4073

(Refer Slide Time: 04:26)

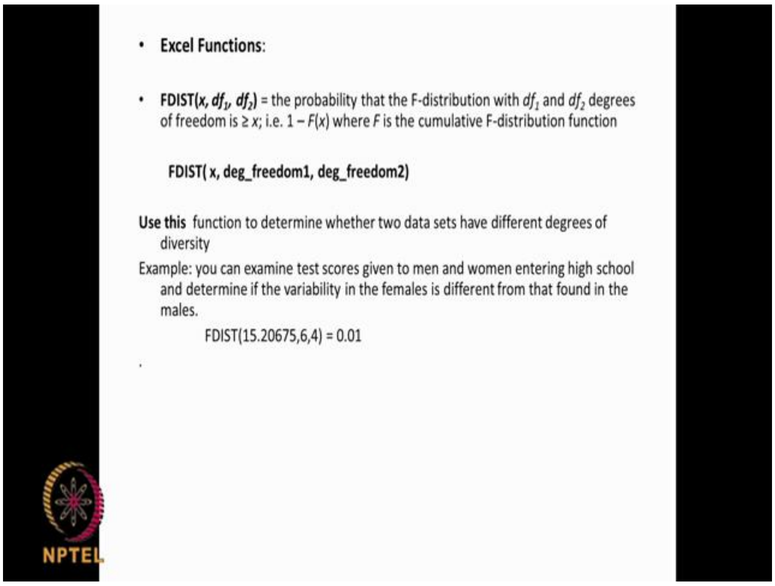


F - Distribution (α = 0.01 in the Right Tail)

df_2 \ df_1 (Denominator Degrees of Freedom \ Numerator Degrees of Freedom)	10	12	15	20	24	30	40	60	120	∞
1	6055.8	6106.3	6157.3	6208.7	6234.6	6260.6	6286.8	6313.0	6339.4	6365.9
2	99.399	99.416	99.433	99.449	99.458	99.466	99.474	99.482	99.491	99.499
3	27.229	27.052	26.872	26.690	26.598	26.505	26.411	26.316	26.221	26.125
4	14.546	14.374	14.198	14.020	13.929	13.838	13.745	13.652	13.558	13.463
5	10.051	9.8883	9.7222	9.5526	9.4665	9.3793	9.2912	9.2020	9.1118	9.0204
6	7.8741	7.7183	7.5590	7.3958	7.3127	7.2285	7.1432	7.0567	6.9690	6.8800
7	6.6201	6.4691	6.3143	6.1554	6.0743	5.9920	5.9084	5.8236	5.7373	5.6495
8	5.8143	5.6667	5.5151	5.3591	5.2793	5.1981	5.1156	5.0316	4.9461	4.8588
9	5.2565	5.1114	4.9621	4.8080	4.7290	4.6486	4.5666	4.4831	4.3978	4.3105
10	4.8491	4.7059	4.5581	4.4054	4.3269	4.2469	4.1653	4.0819	3.9965	3.9090
11	4.5393	4.3974	4.2509	4.0990	4.0209	3.9411	3.8596	3.7761	3.6904	3.6024
12	4.2961	4.1553	4.0096	3.8584	3.7805	3.7008	3.6192	3.5355	3.4494	3.3608
13	4.1003	3.9603	3.8154	3.6646	3.5868	3.5070	3.4253	3.3413	3.2548	3.1654
14	3.9394	3.8001	3.6557	3.5052	3.4274	3.3476	3.2656	3.1813	3.0942	3.0040
15	3.8049	3.6662	3.5222	3.3719	3.2940	3.2141	3.1319	3.0471	2.9595	2.8684
16	3.6909	3.5527	3.4089	3.2587	3.1808	3.1007	3.0182	2.9330	2.8447	2.7528
17	3.5931	3.4552	3.3117	3.1615	3.0835	3.0032	2.9205	2.8348	2.7459	2.6530
18	3.5082	3.3706	3.2273	3.0771	2.9990	2.9185	2.8354	2.7493	2.6597	2.5660
19	3.4338	3.2965	3.1533	3.0031	2.9249	2.8442	2.7608	2.6742	2.5839	2.4893
20	3.3682	3.2311	3.0880	2.9377	2.8594	2.7785	2.6947	2.6077	2.5168	2.4212
21	3.3098	3.1730	3.0300	2.8796	2.8010	2.7200	2.6359	2.5484	2.4568	2.3603
22	3.2576	3.1209	2.9779	2.8274	2.7488	2.6675	2.5831	2.4951	2.4029	2.3055
23	3.2106	3.0740	2.9311	2.7805	2.7017	2.6202	2.5355	2.4471	2.3542	2.2558
24	3.1681	3.0316	2.8887	2.7380	2.6591	2.5773	2.4923	2.4035	2.3100	2.2107
25	3.1294	2.9931	2.8502	2.6993	2.6203	2.5383	2.4530	2.3637	2.2696	2.1694
26	3.0941	2.9578	2.8150	2.6640	2.5848	2.5026	2.4170	2.3273	2.2325	2.1315
27	3.0618	2.9256	2.7827	2.6316	2.5522	2.4699	2.3840	2.2938	2.1985	2.0965
28	3.0320	2.8959	2.7530	2.6017	2.5223	2.4397	2.3535	2.2629	2.1670	2.0642
29	3.0045	2.8685	2.7256	2.5742	2.4946	2.4118	2.3253	2.2344	2.1379	2.0342
30	2.9791	2.8431	2.7002	2.5487	2.4689	2.3860	2.2992	2.2079	2.1108	2.0062
40	2.8005	2.6648	2.5216	2.3689	2.2880	2.2034	2.1142	2.0194	1.9172	1.8047
60	2.6318	2.4961	2.3523	2.1978	2.1154	2.0285	1.9360	1.8363	1.7263	1.6006
120	2.4721	2.3363	2.1915	2.0346	1.9500	1.8600	1.7628	1.6557	1.5330	1.3805
∞	2.3209	2.1847	2.0385	1.8783	1.7908	1.6964	1.5923	1.4730	1.3246	1.0000

Similarly just like p = 0.05, we can have another F table for p = 0.01. You can see here the numerator goes like this, denominator goes like this and again continuation of numerator goes like this actually. This is for the right tail as I did show you, so this will be the area outside for both the t table gives you the area outside. So, do not forget that.

(Refer Slide Time: 04:43)



There is an Excel function. In fact, there are 2 Excel functions, one is called the FDIST Excel function, other is called FINV Excel function.

Insert the slide of FINV from the video

- **FINV(α, df_1, df_2)** = the value x such that FDIST(x, df_1, df_2) = $1 - \alpha$;
- The value x such that the right tail of the F-distribution with area α occurs at x. This means that $F(x) = 1 - \alpha$, where F is the cumulative F-function.

- Returns the inverse of the F probability distribution. If p = FDIST(x,...), then FINV(p,...) = x.

FINV(probability,degrees_freedom1,degrees_freedom2)

Probability is a probability associated with the F cumulative distribution

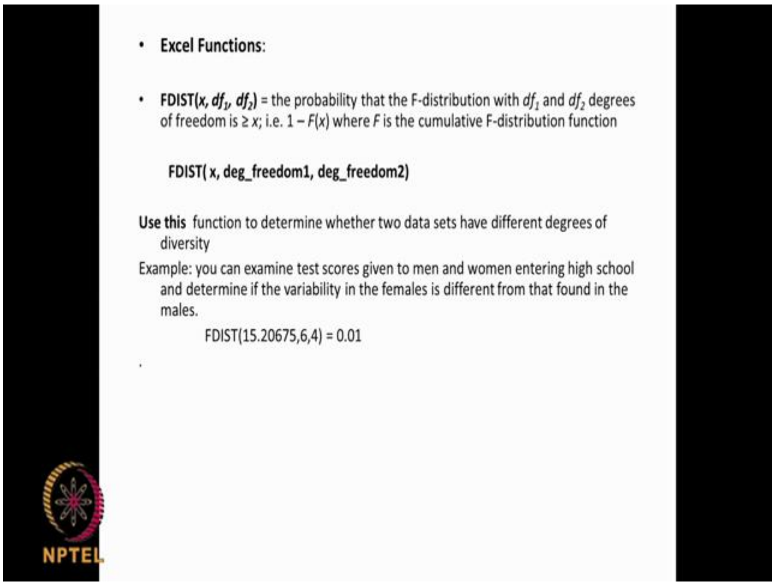
FINV(0.01,6,4) = 15.20675

FINV uses an iterative technique for calculating the function.

Given a probability value, FINV iterates until the result is accurate to within ± 3×10^{-7}.

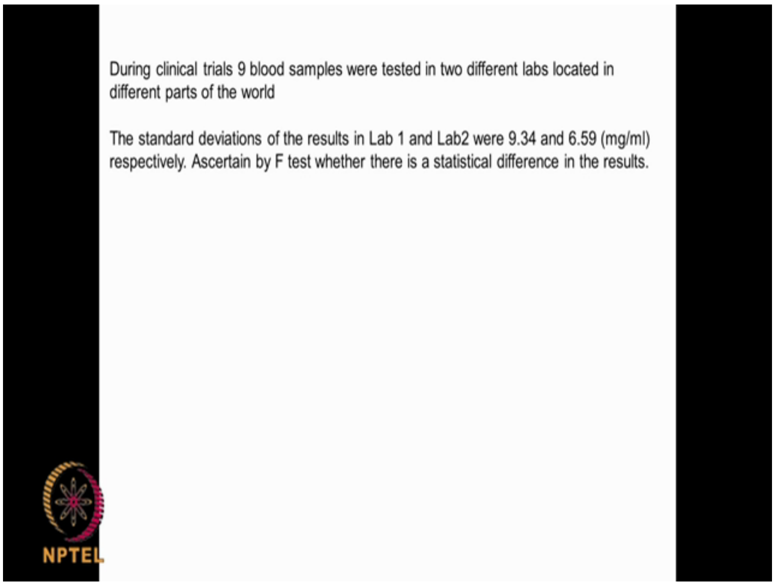
If FINV does not converge after 100 iterations, the function returns the #N/A error value.

(Refer Slide Time: 04:49)



In the FINV excel function, if α is the probability value, so if I am talking about 95 % confidence I will give α as 0.05, when df 1, df 2, I give the degrees of freedom of numerator and denominator respectfully. It tells you, what is the F value, exactly it is like that table. So given the α that is 0.05 or 0.01, given the df 1 and df 2 it tells you what is the F value exactly like the table. Whereas if you look at the FDIST, here it gives you the probability that mean it gives the p value. So df 1 is a degrees of freedom, df 2 is the degrees of freedom and x is that ratio which we calculate from our calculation and it will give the probability. So if I am interested in knowing the probability I use this function that means I put in the F ratio here, then it gives you the probability and you can see whether the probability is less than 0.05 or less than 0.01 for a 95 % or a 99 % confidence respectively. So for example, here it is shown so the F ratio is 15.2, numerator degrees of freedom is 6, denominator is 4. So the probability is given as 0.01. So FINV is exactly like the table, given the probability α it gives you the F value like 0.01, it gives you the 15.2 the F ratio, so we can use both these functions. And similarly the GraphPad software also has the option of calculating the probability, I showed you last time.

(Refer Slide Time: 06:41)

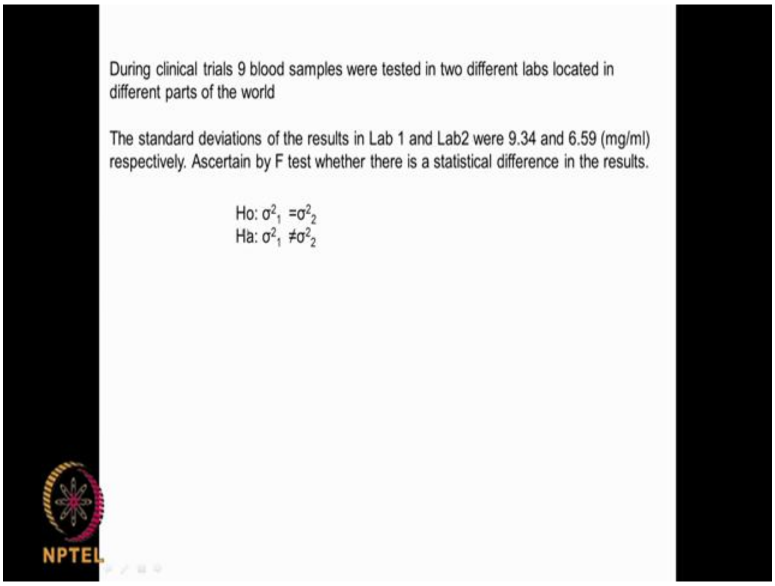


Let us look at another problem, during clinical trials as you know clinical trials are carried out on human volunteers, it could be a multicentric, multi with different locations, different labs, it can even go different continents also. So sometimes samples are sent to 2 or 3 different labs and cross checked to see whether the results are consistent and we say consistent, that mean variances are minimal, that means can I say there is no difference between the variance from lab 1 to lab 2. So that I can be confident that irrespective of where I send my sample, I will get good results or is there a difference in the variance in lab 1 and lab 2. I can use this type of a F test for comparing instruments also, I take the same set of samples and then I inject in 2 different instruments and I can get variance for instrument 1, variance for instrument 2, calculate F ratio and then I can tell, is there a statistically significant difference or there is no statistical significant difference.

Let us come back to our problem. So during clinical trials 9 blood samples were tested in 2 different labs located in 2 different parts of the world. This is very common actually because clinical trials sometimes done in different continents, may be America's, may be Europe's, may be Asia's, Africa's and so on. So you want to be sure whether the results you get because some of these sample blood samples for example, are analyzed using high pressure liquid chromatography. So you will not know the samples analyzed in say Asian lab are consistent or similar to the samples analyzed in European or an American lab. This is a typical study. So same 9 blood samples they were tested in 2 different labs located in 2 different parts of the world.

So the standard deviation, so 9 blood samples. How do you get standard deviation? You take the mean and then you take the difference from the mean of each values, square it up, sum it up, divide by n - 1, here in n is 9 so 8 take a square root, that gives you standard deviation. So you find that 9.34 is the standard deviation of these samples from lab 1 and 6.59 samples from lab 2. Ascertain by F test whether there is a statistical difference in the results, simple.

(Refer Slide Time: 09:10)



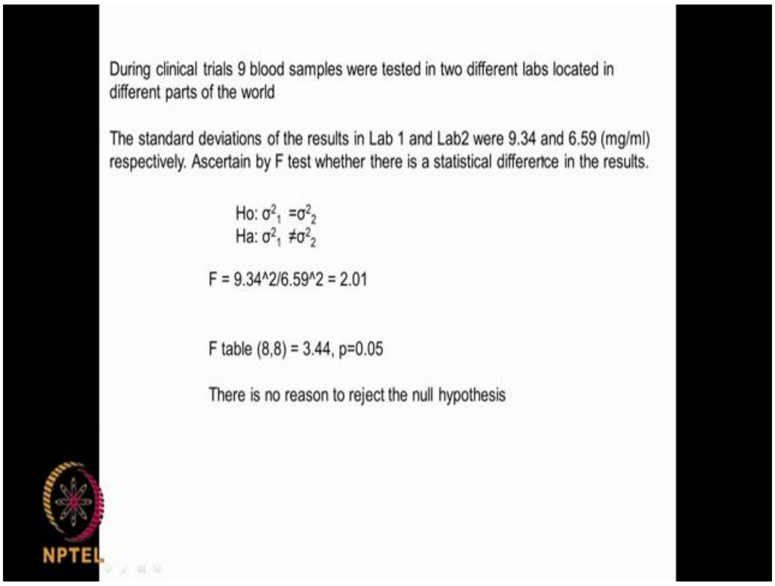
So you calculate their F ratio. Of course the

$$Ho: \sigma^2_1 = \sigma^2_2$$,

then $$Ha: \sigma^2_1 \neq \sigma^2_2$$

, because we are here we are looking at only not equal or different. So like that.

(Refer Slide Time: 09:27)

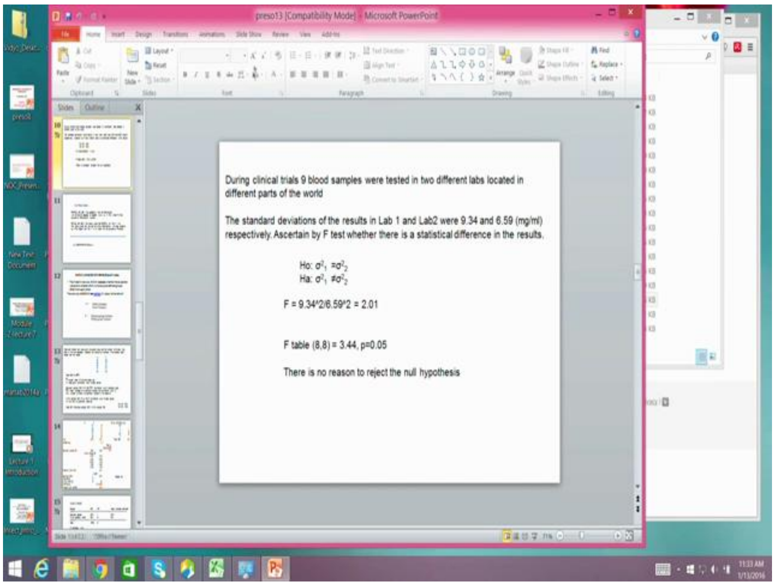


Now of course we can put a p value of 0.05. So F table for 8 degrees of freedom, let me go back to the F table for 8 degrees of freedom each 8 and 8, I get 3.44. So the table t is 3.44 for 8 and 8 degrees of freedom, when I take $9.34^2 / 6.59^2$, I get 2.01. So obviously, the F calculated is less than the F table. There is no reason to reject the null hypothesis. We need to accept the null hypothesis that means, there is no statistical significant difference between the two labs results from both the labs.

(Refer Slide Time: 10:14)

We can use the same thing using the Excel like I said we can use any one of them. So let us use FDIST, FINV.

(Refer Slide Time: 10:29)

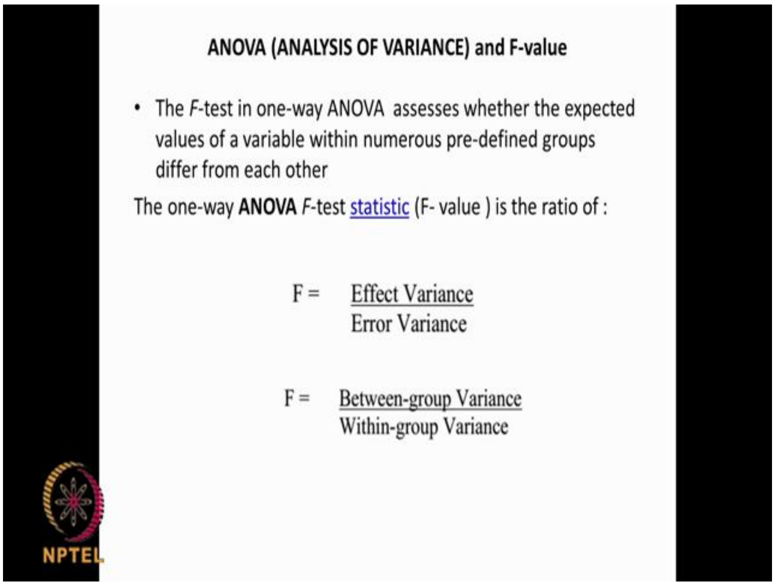


I go to excel here then I will say F distribution. I give the value of F is equal to 2.01, 8 degrees of freedom, 8 degrees of freedom. It calculates p as 0.17, so obviously p is very large, it should have been less than 0.05 for a 95 %. So we accept the null hypothesis. If we look at FINV, we give the probability and it will calculate the F value. So, FINV the probability 95 %, 8 degrees of freedom so it gives you 3.438, so exactly FINV function here in your excel is exactly like your table. So FINV function is exactly like your table. If you do not have a table we can use FINV. So we give the 0.05 as the probability for a 95 %, give the 2 degrees of freedom and calculate 3.438 that is the F table and F value is 2.01 so obviously, there is no reason for us to reject the null hypothesis.

We can use both this commend FDIST, knowing the F ratio it calculates the probability, FINV knowing probability, it calculates the F ratio. Now we can do the same thing, using the GraphPad software also let me show you that. So GraphPad as you can see here F value is done here. So we go here, continue we say. So again this also has both the options given probability, it can calculate F value, given F value it can calculate p value. So let us look from here, calculate p value so we say continue, go to F value so F is 2.01. So degrees of freedom numerator is 9, denominator is 9. So we compute, it gives you 0.1565 as in the Excel or we can use this, given from the probability it will calculate from the probability it will calculate your F value so 0.05, 8 degrees of freedom, 8 degrees of freedom, compute F, here you can see F = 3.44 so again if you look here 3.4. So we see in both the cases, we can get both these using these 2 different commands that is available. So here from the probability, that means I give 0.05 it will give me the F value, that means, exactly that table that is FINV. Whereas here given the ratio it will give the probability that is that F test command, that is the FDIST command actually. Whereas in our calculations what we do is, in our calculations we calculate F value and then we go to the table 8 , 8 degrees of freedom for p = 0.05 we get 3.44. So there is no reason for us to reject the null hypothesis. So I taught you so many different approaches by which we can calculate the F distribution.

So just numerically with the help of a table, we can do this or we can use the excel function FINV and FDIST function are there or we can use the graph pad software also there. So, both can be done.

(Refer Slide Time: 14:50)



Now let us look at something much more advanced it is extremely useful that is called the ANOVA, analysis of variance. ANOVA also makes use of this F test basically. So it also calculates some variance divided by some other variance, and then it looks at the F table and either accepts the null hypothesis or rejects the null hypothesis. So ANOVA is extremely useful in that in that way, we can use it for comparing the large number of sample size. Whereas if I have only 2 sample sets I can use 2 sample t test, but 2 sample t test compares means. I can use an ANOVA which compares variances using the F test, but the main advantages if I have many samples in 1 shot, I can use the analysis of variance and I can compare and see whether there are differences in the variances of each of these population or not.

So there is something called one - way ANOVA, two way ANOVA, three way ANOVA and so on actually. Suppose if I am comparing 4, 5 drugs for diabetes then that is called a one - way ANOVA that means, I am comparing only drugs. So I will see variations in the performance of each of the drug. But suppose I am comparing not only say drug a, b, c but I am also comparing between male and female, I want to know how the drug perform between male and female. So I have drug as 1 set of group and I also have the gender as another set of the group, because you can always have a drug performing differently on a male and differently on a female. There are many examples where drugs perform very well on male, it does not perform very well on female. So we can have 2 sets of a groups here right, I can group them based on drugs a, b, c or I can group them based on the gender male and female. So that is called a two way ANOVA.

Like that we can have a three way ANOVA for example, if I am comparing 3 drugs a, b, c, I am comparing male and female gender male and female and I am comparing Europe and U.S because now a days it is well known that drugs behave very differently between Asian population and African and Europeans and so on actually. So imagine I am performing clinical trials in USA and Europe simultaneously, on 3 different drugs on male and female. So what we have? 1 set of groups should be drugs a, b, c, another set of could be the male and female gender, the 3 rd one could be the continent Europe and U.S. So we have a three way ANOVA.

So like that we can have multi way ANOVA also, one way ANOVA, two way ANOVA, three way ANOVA, right. So we are going to look at each one of them slightly in more detail. So here also we perform F test that means, we divide 1 variance by another variance. So what are the variances we divide by? So in a one way ANOVA for example, suppose there is a drug a, b, c we are comparing. So between group variance that means, between the drug variance, that is a, b, c and within the group variance that means, when I am doing multiple same drug given to many patient will be some variance inside right, because same drug will perform differently. So you may have different sets of answers there that is called within group variance. So F is calculated based on between group variance / within group variance and that is what is compared with the table F. This is called within group variances error because if I am repeating a sample many times on an instrument I will not get same answer I will get different answers. So that is called an error or within the group variance.

The other one is the between group variance. For example, I am testing 5 samples on 2 different machines I could have between the 2 machine variance and if am doing within the instrument. So I will have 4 or 5 samples that is called the Error variance. So I am dividing the between the group variance / error variance. So if the error variance is small, I will get large F value that means, I will be able to differentiate between say drugs or I will be able to differentiate between instruments. I can use it for operators also, suppose I am giving 5 previous problem like I give 5 drugs to 2 operators and ask them to analyze and then the results. So there will be between group, that means between the 2 operators and within group, within group is more like error. Whatever operator 1 does within is like any error variance or within group variance. So I divide the between group with within group and if I get large F value, then I will be able to reject the null hypothesis. So what it means is, it also means if the errors variances are small, then I will be able to really differentiate between groups it could be drugs, it could be operators, it could be instruments, it could be anything. So the most important thing in experimental design strategy is consistency that is repeatability has to be good. If am doing a sample 10 times, if the variance of these readings are very large, obviously if am going to compare with something else with another group then I will not be able to get a good F value, F value will be small because denominator will be very large.

I will not be able to really reject null hypothesis, there is no reason for that. So it is a very important lesson and that is taught by this concept of ANOVA, analysis of variance. And again here also we come up with the F value here as I said it could be between different drugs, between different operators, between different instruments, between different assets, between different animals. We are looking at the variances and then trying to make a comparative study actually. So there are many problems which one can do and we will talk about this problem as we go along and ultimately, the same statistics is used. Variance of 1 / variance of 2, where 1 could be between groups variance, within group variance. Within group variance is nothing but the Error variance. So as you can see it is a very important concept one needs to understand when you are doing ANOVA. ANOVA is very powerful because as I said you can use it for 1 group like drugs, comparison of drugs or it can be 2 different groups comparison of drugs, as well as on different genders male or female. Advantage here, when you do that we can even look at interactions between drug a with male, interaction with drug a with b, and it is well known nowadays some drug work very differently on different types of population, different types of genders, different types of age groups. So that interactions can be well studied using this concept of ANOVA and we are going to look at in more detail as we go long in the next few classes.

Thank you very much for your time.

Key words: Standard Deviation, Variance, Degrees of Freedom, Probability, F-tests, Analysis of Variance