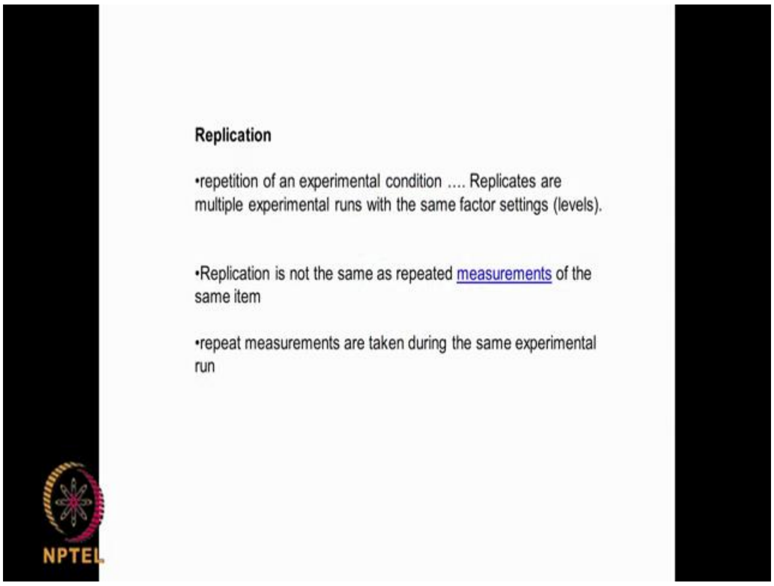
Biostatistics and Design of Experiments
Prof. Mukesh Doble
Department of Biotechnology
Indian Institute of Technology, Madras

Lecture - 17
ANOVA

Welcome to the course on Biostatistics and Design of Experiments. We will continue on ANOVA. Whereas I said ANOVA is the most important topic in the entire statistics. We can compare different operators, different process conditions, different drugs, and we can compare one-way, two-way, three-way, multi-way, and so on actually. So, it is a very powerful tool and it is a widely used tool. We were yesterday talking about something called Replication. So, what is Replication?

(Refer Slide Time: 00:43)



Replication, we completely repeat the experimental run. That means, suppose I am doing an experiment at pH = 3, temperature = 30, with one % inoculum size and some amount of carbon nitrogen. So, I complete the whole experiment, I get the biomass, I get the product metabolites. Now, I again repeat the same experiment with all these conditions, and complete the whole experiment, get the biomass and metabolites; that is called the Replication.

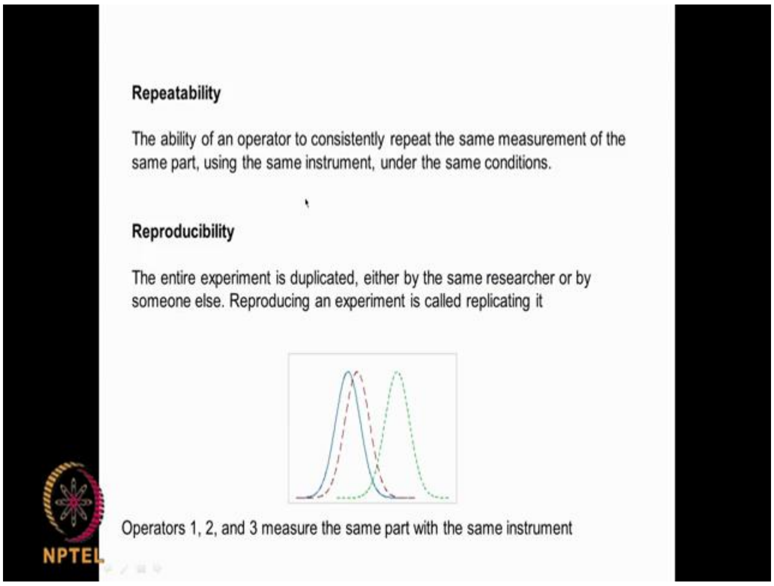
So, there is a difference between replication and repeat measurements. What is repeat measurements? Suppose I am running a fermenter, I take a sample out, and look at the biomass by measuring the OD, again I take a sample out and measure the OD, and again I may take a sample out measure the OD or I may measure the OD three times, that is called repeat measurements. But, that is not completely repeating the entire experiment from the beginning with all the variables, that is called Replication.

So, that is the difference between replication and repeated measurements in the same material. So, I may repeat the measurement. Suppose I am measuring pH, I may do it three times, four times, five times. If, I am measuring the OD, to measure the biomass, I may do the OD measurement three times, four times, five times, right? So, that is called repeated measurements. That is not Replication. So, in Replication you completely carry out the reaction once more with all the conditions, with all state variables or independent variables, complete the experiment and measure all the output variables or dependent variables. That is called Replication.

Replication is very very important because when we carry out the same experiment completely twice, thrice, four times, we get a measure of the error that is involved in the entire process. So, the process may contain preparing the inoculum, adding the inoculum to the reactor, carrying out the fermentation, taking out samples, measuring the amount of metabolites by extraction, measuring the amount of biomass by centrifugation; so it gives you a very good measure of error.

And as you know in ANOVA, the error sum of squares is very very important; that comes in the denominator. When we are calculating F ratio, as you know, between sum of squares, that means between various groups / error sum of squares. So, error sum of squares is a measure of the repeatability of my entire process. It is not repeat but replication of the entire process, and that is very very important to know, in order to see whether there is any difference when I change temperatures or when I change pH or when I change rpm or when I change amount of carbon or when I change nitrogen and so on actually. So, Replication becomes very very important. Replication also adds to the number of degrees of freedom, and as you know, degrees of freedom are very important if you want to perform the ANOVA calculation.

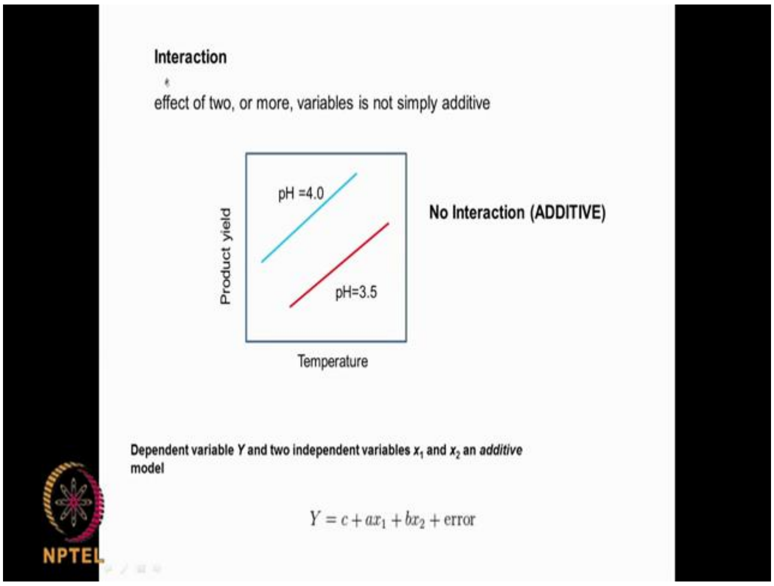
(Refer Slide Time: 03:56)



So, Repeatability is measuring the same parameter twice, thrice; like, if I measuring pH, I measure it many times, if I am measuring temperature I measure it many times, if I am measuring OD I measure it many times - that is called the Repeatability. Whereas Reproducibility, I perform the entire experiments, that is almost like, that is the replication. So, Reproducibility is very different from Repeatability. So, I may perform the whole experiment from start keeping all the independent variables and performing the entire experiment or I may ask another operator to perform the experiment assuming there is no difference between the operators. So, that is called Reproducibility.

So, Reproducibility is almost same as Replication, whereas Repeatability is repeatability of your measurement; both are different. So, suppose there are three operators and they perform some experiments. We are trying to look at how? What is the reproducibility of these three operators? So, one and two may fall very closely, whereas three may be fall away. So, that is called Reproducibility. So, both are different Repeatability and Reproducibility, and as I said Replication is the same as Reproducibility. These are terms you need to remember in statistics. Statistics has lot of terms, which cannot be used very loosely. Although, we keep using many words very loosely in statistics, we cannot do that actually.

(Refer Slide Time: 05:35)

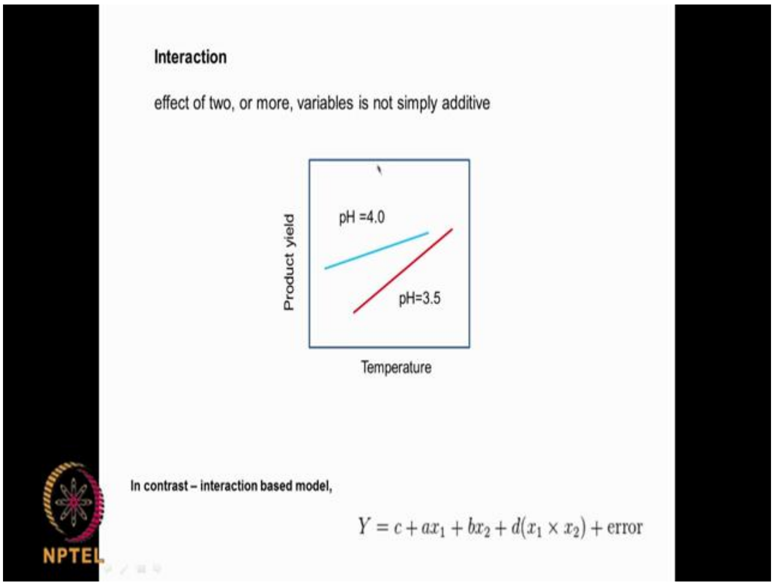


Now, when we have more than one variable, say two variables, three variables, four variables. You may have a situation, where they could be interacting with each other. For example, suppose I am changing pH, and I am changing the temperature, and running a fermentation process and measuring the product yield. At pH = 3.5 I change temperature, so the yield keeps going up like this. Now at pH = 4, again I perform experiments by changing temperature, again the yield goes up like this. So, it looks like an additive, right? So, I could develop a simple regression model, we will talk about regression much more in detail later. So, I could develop a regression model like this – some

$$Y = c + ax_1 + bx_2 + d(x_1 \times x_2) + \text{error}$$

constant + a X x_1, x_1 could be temperature, b X x_2, x_2 could be pH + some error. So, these two parameters x1, x2 which are called the independent variables - in this particular case pH and temperature - they seem to be independent of each other, so they are adding to each other. So, there is no interaction.

(Refer Slide Time: 06:48).



But on the contrary you could have situation like this where at pH = 3.5, you are changing temperature, your product yield goes up; but, at pH = 4 your product yield does not go up in the same way; it goes up and all most remains same. So, there is interaction; there is an interaction between pH and temperature. So, the performance of the fermentation process seems to be not linearly changing in with temperature, but it also depends on at what pH you are running at. So, at pH = 3.5, you are having like this, whereas at pH = 4, it is not going up in the same fashion, but it seems to be sort of falling down. So, there is an interaction between pH and temperature.

If you want to develop a simple linear regression model in this particular case, what will you have? You may have constant + ax1 that could be temperature; bx2, x2 could be pH and also there could be a relationship equation which has d X x1 and x2. So, x1 could be temperature; x2 could be pH; so obviously they are interacting here; a, b, c, d are constants here. So, on the contrary to the previous equation, where we have ax1 + bx2. So, this seem to be additive. So, the effect of temperature, effect of pH are additive towards yield. Whereas, in this particular equation you have an additive term plus also you have an interacting term between these two independent variables. So you may have like this; understand? Or you may also have like this at pH = 3.5, as I change temperature product yield goes up, but at pH = 4 when I change temperature product yield goes up very drastically. So, you see each one of these curves, these are almost parallel. So, this is more like additive, these sort of converging, and these are diverging tremendously.

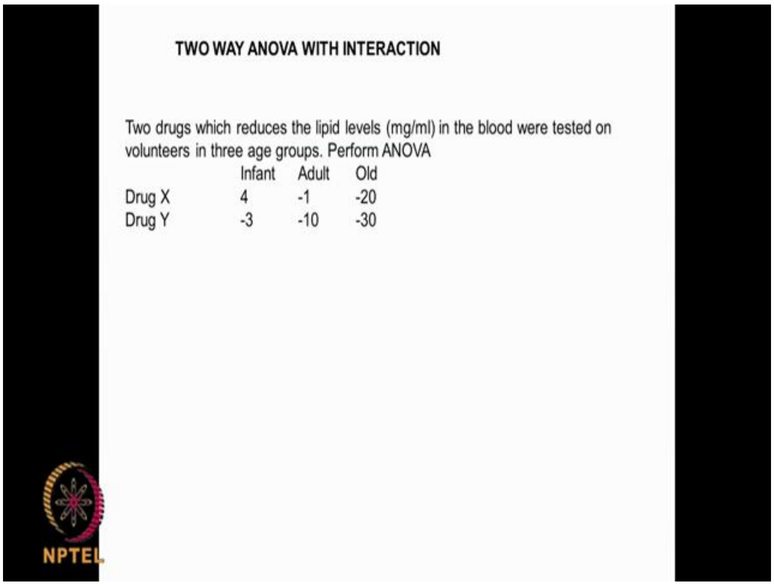
Again, you could develop a regression relation of this form, where you say

$$Y = c + ax_1 + bx_2 + d(x_1 \times x_2) + \text{error}$$

. So, here again we can say pH is interacting with temperature. Interactions are very common especially in drug discovery, you are testing some drugs on a Caucasian population vis-a-vis African population, they may perform very differently. If you are testing some drugs with Asian and Chinese, and the Americans, the drug may be performing differently. So, it is well known because of the genetic makeup of the people from different continents, drugs may be acting on different path ways and different genes, and hence they may be performing very differently.

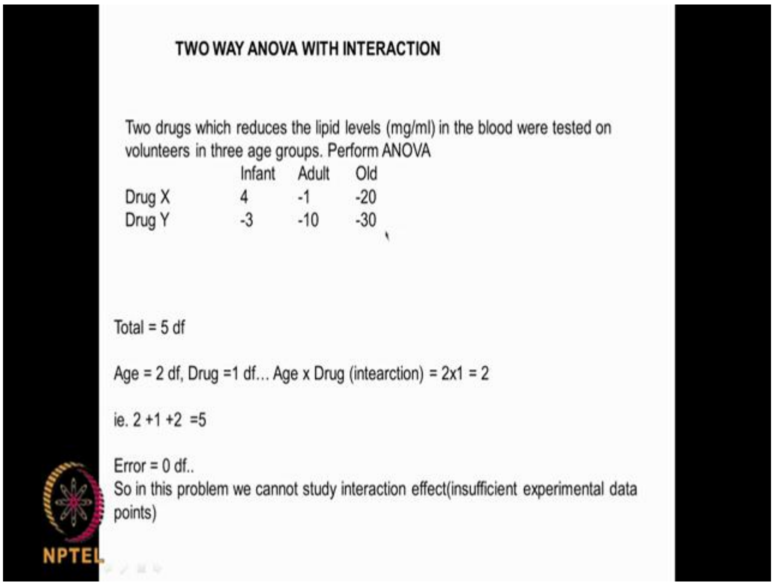
A drug which may be very active in some part of the world and some part of the population, may be even toxic or detrimental in some other part. They are called interaction. And that is why now-a-days most of the clinical trials are carried out in multi-country or multi-continents, so that you get, you are able to capture this type of effect. Suppose you carry out clinical trial only in say Europe, and you tried to extend this drug in a market in Asia or Africa, you could get into trouble because the drug may be acting very differently on the other type of population as against the European population. So, that is called Interaction. And it is very very important to know interaction because it is very common, as I showed you in the fermentation example, drug discovery it is very very common; in many situations you will find these type of problems. For example, some drug may work in one way on male population, whereas it may work in a different way on female population. Some drugs may work differently on infants, as against on aged people. So, these type of things are called interactions. The way some food works on aged people as against adult people and so on actually. That is called interaction and in statistics using ANOVA, we will be able to find out some these interaction effects also.

(Refer Slide Time: 11:28)



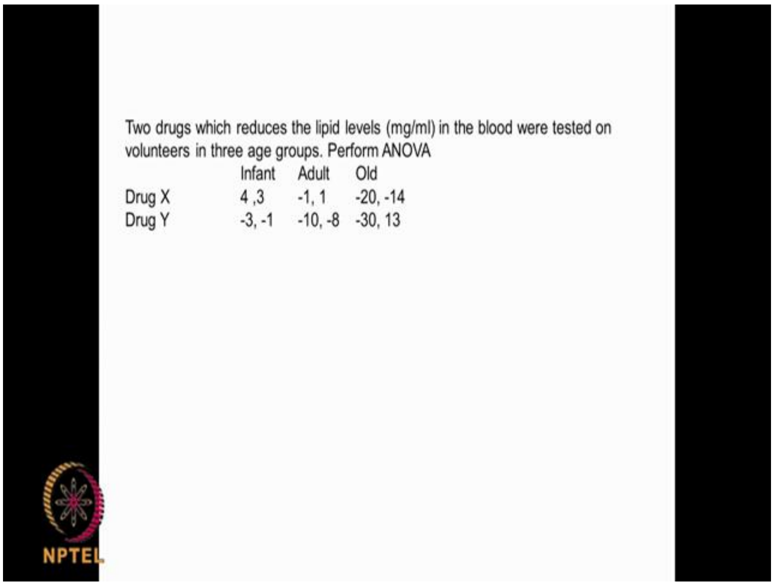
So let us look at this problem. You have two drugs, I think you remember we saw this problem in the previous class. We have two drugs, drug X and drug Y. It is supposed to be lowering the lipid levels. On the adult, it is lowering little bit drug X, on old people its lowering a large amount. When you take drug Y, drug X is increasing little bit on the infant, it is lowering on drug Y when drug Y is given and it is almost constant on the aged people. So obviously, there appears to be an interaction between drug and the age of the subjects or age of the volunteers.

(Refer Slide Time: 12:12)



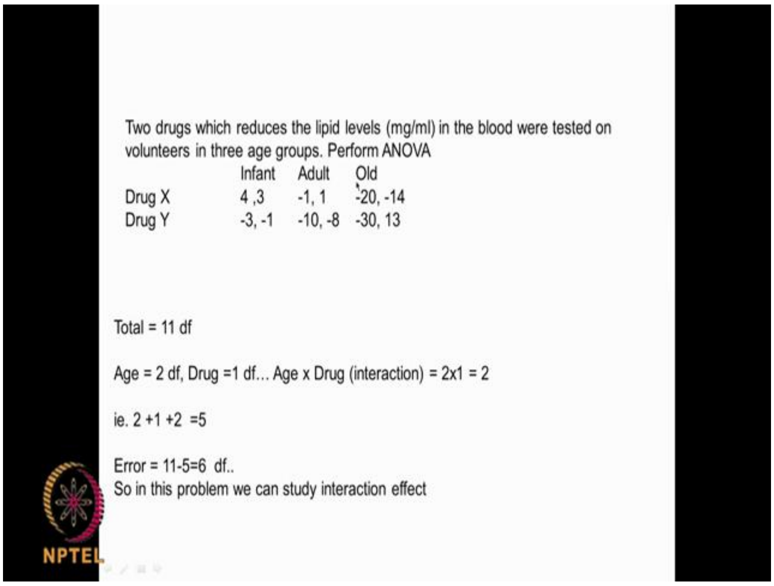
Now let us look the various degrees of freedom. Now, there are 6 data points. So, you have 5 degrees of freedom, right? 6 - 1. Now age there are three groups obviously, 2 degrees of freedom. Drug is 1 degree of freedom because you have 2 drugs. So, if you want to look at age into drug that is interaction, then you need to multiply 2 x 1 that comes 2. So, if I add these 2 age, if I add this one drug, and if I add this age drug interaction two, they all are add up to 5. So, 5 - 5 is what error will be so obviously, we cannot use this data set to find out interaction. So, we can use this data set to find out effect of age, mean effect will call it effect of drug mean effect. So, 2 + 1 is 3, 5 -3 is 2 and we will call 2 degrees of freedom for error, but if you want to bring in interaction age X drug, then error will have only zero degrees of freedom. So, we cannot solve this problem. So, what does that mean? It means we have insufficient experimental data points. That means, the degrees of freedom is zero. So, we need to have more data points, that means we should have done a replication of all these. So, if you had a replicates, set of replicates, then obviously, we will have more degrees of freedom for error. In fact, the next example, look at this next example.

(Refer Slide Time: 13:36)



So, you have taken two infants and carried out the experiment. You took two adults carried out; two aged people. So now you have 12 data points. That means you have totally 11 degrees of freedom.

(Refer Slide Time: 13:56)

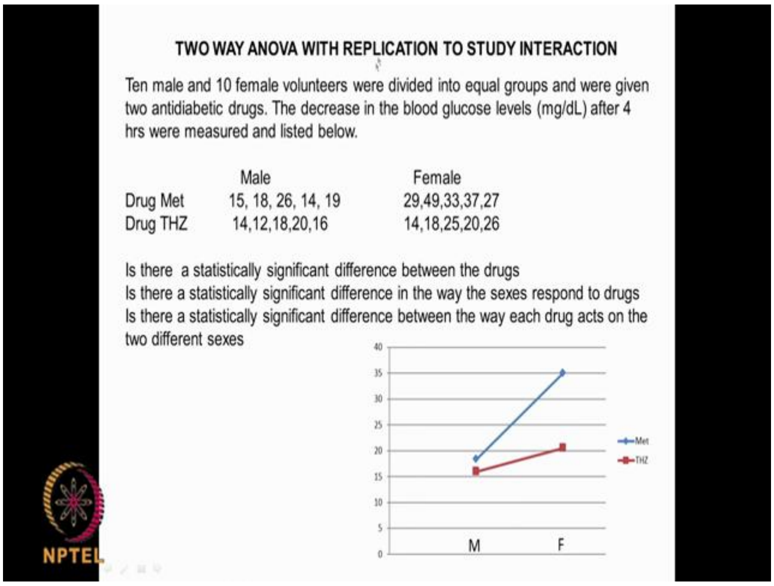


Now, age wise three age groups; so, you have 2 degrees of freedom. Drug wise you have two drugs; so 1 degree freedom. So age X drug is 2. So, 2 + 2 + 1 is 5. So, 11 - 5 will be 6 degrees of freedom for error. So, you will be able to determine the sum of squares of error and we can also determine the mean sum of squares of error; because, how do we calculate mean? We divide by the degrees of freedom. So, in this particular problem, we can we can calculate both interaction error, because we have enough data points. Whereas in the previous case what happened? The error degrees of freedom was 0, if you want to calculate interaction. So, what we did? We said we will not calculate interaction. So, the error degrees of freedom will become 2; so we can calculate only the main effects.

So, now I am introducing another term that is called main effect. Main effect in this particular problem is drug and age. The interaction effect is drug X age, then you have error. So, in this problem also main effect is drug and age is another main effect. The interaction effect is drug X age. So, in this particular case where we have 12 experiments performed; that means, we have replicates here in each of the case. So, we have 11 degrees of freedom, so there is no problem; age is 2 degrees of freedom; drug is 1 degree of freedom. Age X drug, you multiply this by this; so get 2 degrees of freedom. So, they all add up to 5; 11 - 5 is 6 error degrees of freedom. So, we can determine, what is the error sum of squares; then we can determine what is that mean error sum of squares; then we can determine the F for the interaction also. Do you understand?

Let us look at problems using interactions also. Because as I said interaction is very very important. So, what does this mean? It means that replication is very very essential if you want to look at interactions, if you want to look at errors and so on, because you need to have sufficient degrees of freedom, because error is what we use, the mean sum of squares of error is what we use for dividing the group sum of squares to get the F value. Let us look at this problem. This two-way ANOVA with replications and also we are trying to study interaction.

(Refer Slide Time: 16:24).



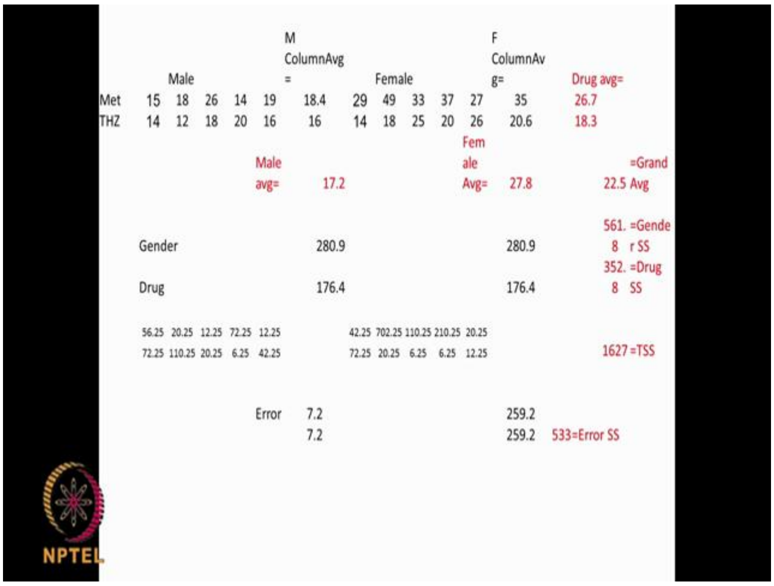
So, we have ten volunteers. We have five males, who were given a drug called Metformin. You all know what is Metformin? It is meant for a lowering glucose. You had a another five male volunteers given other glucose controlling drug THZ. Similarly, you had five female volunteers who were given Metformin and you had five more female volunteers who were given THZ. Now the question is - is there a statistically significant difference between drugs? You know how to do it very simple. Is there a statistically significant difference between the sexes or the genders? Is there a statistically significant difference between the way these drugs act with each other? That is very interesting. So that means, there might be some interactions.

So, how do you do this? For example, if I take an average of all these now, because I have replicated with 5 males, I take an average of all these. So, it may come up to around 16 point something. For Metformin in male here. If I take average of all these. So, this may come to around, that is male THZ, that may come around 19, that is here. Now if I take average of all the females on Metformin, these numbers are very large, right? So, you will get around almost 21, 22 and so on actually. So, sorry Metformin is here. So, if we take the average of all these it may come to almost 35 here. And if we take average of all THZ you may get around 21, that is here.

So, when you plot the average, these five values, average of this, average of this, average of this into two graphs. So, this red indicates THZ that is this and this - average of this and average of this, and here it indicates male, here it indicates female. Look at Metformin its going like this; THZ is there is going like this. Obviously, there is an interaction. If you remember the old pictures I showed you, ideally if they are parallel to each other, we can say both the drugs act in the similar way between both the genders. But in this particular case, both the drugs do not act in the same way with both the genders because they are not really parallel they are diverging. That means, the Metformin is acting much more on female when compared to male. Whereas, THZ is acting only marginally on female, when compared to male. So, it is because it is diverging, right? So obviously, there is an interaction and we are interested in knowing the interaction. There are many drugs which have such effects. They may act differently on female and they may act differently on male. Let us go back to our usual ANOVA table and look at the problem.

(Refer Slide Time: 19:19)

.



So, we write down the male data for Metformin, male data for THZ, and then we have the female data for Metformin, and female data. So, male this particular average; so if I take the entire male average 15, 18, 26, 14, 19 - whole lot - I will get 17.2; that means, irrespective of whether it is Metformin or THZ, 10 data points. Similarly, if I take all the 10, I will get the female average 27.8. Now, if I add up all this, and if I add up all these, and then I get these. This is called the drug average; this is for Metformin, this is for the THZ, you understand? This for Metformin. So just like male average, there is male average; I take all these 10 data points to get 17.2. Female average; I will take all these 10 data points and get the female average. Drug average, I add all the Metformin and take 10, divide by 10, all the THZ and divide by 10. Now, the grand average will be average of these two or average of these two, this is the grand average.

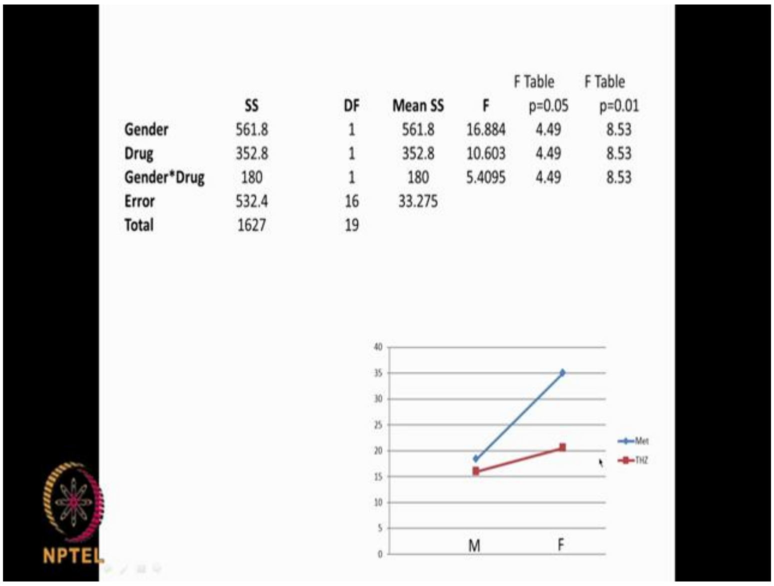
Now, how do you get each of these sum of squares? So, we have the male average 17.2, female average 27.2. So, how do you get the gender sum of squares? Simple. This is the male average $17.2 - 22.5^{2}$ X 10, because there are 10 data sets here. So, and then similarly for the female, what you do? $27.8 - 22.5^{2}$ X 10 because there are 10 data points; do you understand this? So, we take the male average, all these 10 data points; female average, all these; over all grand average is this. So, for gender, 10 into this minus this square, plus 10 into this minus this square. Then if you add up, you get 561.8 that will be called the gender sum of squares.

Now, how do we do for the drug? So, we have the drug average here 26.7 - 22.5 that is the grand average $\bar{X}$, if you remember square 10, because there are 10 points here right?

And then same thing 18.3 -22.5 ^{2} X 10. Then you add up these two you will get 352.8; that is called the drug sum of squares. So, we got gender sum of squares, we got drug sum of squares. How do you get the total sum of squares? You all know right? 15 - 22.5 ^{2}, then 14 - 22.5 ^{2}, 18 - 22.5, like that. All the 10 data points, we do like that. So, how did you get this? 29 - 22.5 ^{2}, 49 - 22.5 ^{2} like that. So, if you add up all these, you will get the total sum of squares. So, you know how to do total sum of squares, you know how to do drugs sum of squares, you know how to gender sum of squares. Now, error; how do you get the error sum of squares error? Error, So, this is 18.4. That is the column average. Now this is 17.2 is the overall average, 17.2 -18.4 ^{2} X 5; because there are 5 data points here, that is called the error sum of squares.

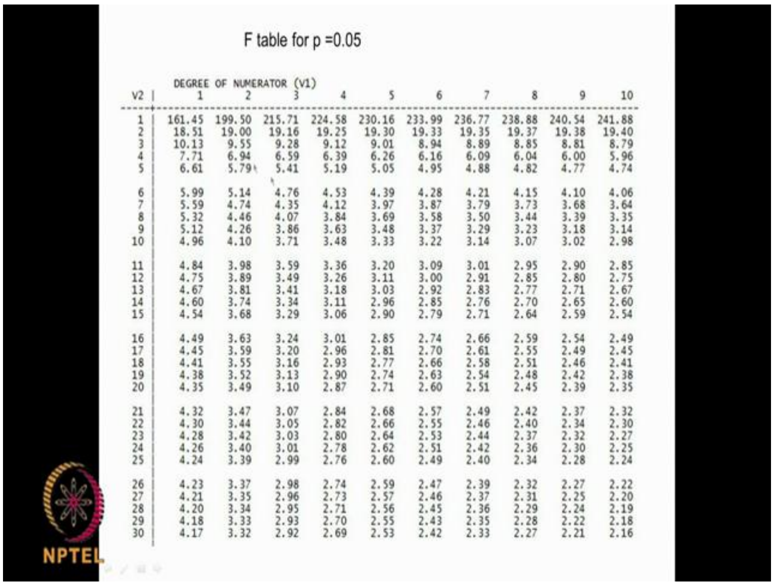
And, similarly we can do for the other one also; we will get 29 - 35 ^{2}, 35 - 27.8 ^{2} X 5. So, like that you keep on doing, then again for this 16 - 17.2 ^{2} X 5. So that if you add up all these, that will give you the error sum of squares. So, the error sum of squares, so we know the grand average, we know the gender sum of squares, we know the drug sum of squares, we know the total sum of squares. So, the error sum of squares is also known. So, if I subtract error sum of squares, drug sum of squares, gender sum of squares, I should be able to get the interaction sum of squares. Do you understand?

(Refer Slide Time: 24:29)



So, we go to ANOVA table. So, we have the gender sum of squares 561.8, drug 352.8, total is 1627; 533 is error; and then, errors gender into drug that is an interaction. You add these these terms and subtract from 1627. Now, in total we have 20 data points. So, 20 - 1 will give you 19 degrees of freedom for total gender; there are two genders male and female. So, DF is 1, drug we have two drugs. So, DF is 1, gender into drug is 1. So, the error is 19 - 3 is 16. So, mean sum of squares, what I do? I just divide with the DF. So, I will get 561.8 as the gender mean sum of squares. 352.8 as the drug sum of squares, gender into drug as 180, and of course, the error is there actually. Now error is 532 / 1633. Now, how do you calculate F for gender? This divided by error. How do you calculate for drug? This divided by 33, gender into drug 180 / 33.

(Refer Slide Time: 26:04)


F table for p =0.05

V2	DEGREE OF NUMERATOR (V1) 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

Now, let us go to F table for 1 and 16 of freedom. 4.49 at 95 %; 4.49. So, at 95 % confidence, all these three are significant; that means, there is significant difference the way the drug acts on gender, the way the performance of the two drugs, and there appears to be a gender into drug interaction, at 99 % confident interval for 1 X 16 for 1 , 16 is 8.53.

(Refer Slide Time: 26:34)

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

Denominator Degrees of Freedom (df_2) / Numerator Degrees of Freedom (df_1)

df_2 \ df_1	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

NPTEL

So, only gender and drug become significant and the gender into drug - which is called interaction - is not at all significant. But at 95 %, it is significant and at 99. And as I showed you here in this picture, so if you look at THZ, male it performs in certain fashion, female it performs there is an increase. Whereas, we take Metformin, male it performs, but in female, it is a very drastic increase. It is not like this type of increase with the slope, but the slope is very very high. So, if you come across such a situation, we can very confidently tell that, there is an interaction between the way the drug responds on male as against on female. Otherwise, if there is no interaction you should get approximately a parallel line, understand? So, these are very important problem, where you are looking at a two-way ANOVA with replication to study interaction. So, the main point is you need to have replication of the data here; that is very very important. You need to have replication of this data, otherwise as I showed in one simple example, if we do not replicate there are no degree, if you take interaction, then there are no degrees of freedom for error. So, if you consider error, then we cannot measure the interaction effect.

So, will continue further on this ANOVA with replication and interaction in the next class.

Thank you very much.

Key words: Reproducibility, Repeatability, repeat measurements, Replication,

Interaction, Mean sum of squares, error sum of squares, total sum of squares, Interaction effect, Main effect, degrees of freedom