

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

**Lecture - 09**  
**t-tests**

Welcome to the course on Biostatistics and Design of Experiments. Today we will continue on the t-test like I had mentioned that t-test does a comparison of a means.

(Refer Slide Time: 00:21)

**Comparing Means**

The t-test allow you to determine if there is a statistically significant difference in the means of two samples, or between the mean of a sample and an established value.

- T-tests are robust to non-normal data
- The test is more sensitive if the two groups being compared have similar standard deviations.

What tool should you use?

- **1-sample t-test** : if you want to compare the mean of your data to a known value
- **2-sample t-test** : if you want to compare the means of two samples
- **Paired t-Test** : if you are comparing the difference between two paired samples to zero.

NPTEL

If the mean of one sample and you have a mean of another sample, you want to know whether they come from the same population or they come from different population and so on actually. Basically, t-test compares means, there are 3 different types of t-test 1 sample t-test, 2 sample t-test and paired t-test. So, in 1 sample t-test what to you do is, you take a sample and the mean which you get compare it with the population mean and you find out whether this sample mean comes from this population or not. In a two sample t-test, you are comparing 2 sets of samples for example if I am comparing drug a and drug b performance, then I will get some mean for sample set 1 and I will get some mean for sample set 2 and I would like to see whether there is any statistically significant difference between these 2 means or not. In the paired t-test, you use the same subject for performing the experiments for example, if I am performing clinical trials and I have 10

volunteers I may give drug a to the 10 volunteers and after a few days I may give drug b to the same 10 volunteers and I will see the change. So theoretically, if a and b perform in the same way that difference should be 0 that is called the paired t-test.

(Refer Slide Time: 01:52)



The mean heart rate of 20 rats = 282 with a cv=15%. During anaesthesia the mean heart rate = 240, but cv is same. Is the decrease in heart rate significant? Give 95% confidence limit for the change in heart rate during anaesthesia?

NPTEL

Let us continue with some problems because problems are very very important, problems give you an idea about a how to attack these type of a situations and we need to know how to use the tables, we need to know when to use a 1 sample t-test or a 2 sample t-test, when to use a 1 tailed or 2 tailed, how to create the a null hypothesis and alternate hypothesis and so on actually. The mean heart rate of 20 rats is say 282 with a coefficient variation of 15 %, what is coefficient of variation?

$$100 s / \mu$$

During anesthesia the mean rate becomes 240, it looks like it is gone down, but the CV is the same. Is the decrease in heart rate significant? We need to find out whether the decrease in the heart rate is significant. Give 95 % confidence limit for the change in the heart rate that means, it is the change that is 240 - 282 or 282 - 240. What is the confidence interval? Although we say 282 - 240 the change is 42, but then there is always be going to be a region of confidence because in any mean will always have a standard deviation. Let us go about solving 1 at a time.

(Refer Slide Time: 03:12)

The mean heart rate of 20 rats = 282 with a cv=15%. During anaesthesia the mean heart rate = 240, but cv is same. Is the decrease in heart rate significant? Give 95% confidence limit for the change in heart rate during anaesthesia?

$$cv = 15 = 100 \frac{s}{\bar{X}}$$
$$s = 15 \cdot 282 / 100 = 42.3, s^2 = 1789$$
$$s^2_y = 1789 / 20 = 89.45$$

cv is the same for the anaesthesia group also so,

$$s^2_y = 89.45$$


So coefficient of variation is given by 15 %, 15 is equal to 100 into s the standard deviation divided  $\bar{X}$ . As you can see 282 is the mean that is  $\bar{X}$  so from there I calculate s simple. I just rearrange, so s comes out to be 42.3. Now the variance

$$s^2 = 1789$$

Now I can calculate the standard error of the mean by divide by 20 here right if you remember, what is the standard error s by square root of n or the square of the standard error if the mean will be  $s^2/n$ , that is what I get it 89.45. Now CV is a same for the anaesthesia group also, the standard error will be the same for the anaesthesia group do you understand this so far. I know CV, I take 282 here  $\bar{X}$  so I calculate s that is the standard deviation. Now the standard error of the mean, I just have to divide by n and then later on I need to take square root, but right now we are not taking square root we leave it like that. Now CV is the same for the anaesthesia group also, so standard error for the anaesthesia group also will be the same ok.

(Refer Slide Time: 04:38)

The slide contains the following text and formulas:

$$H_0: \mu = \mu_0$$
$$H_a: \mu < \mu_0$$
$$t = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$
$$df = n - 1$$

If t calc < than the Table t accept  $H_0$   
If t calc is > than the Table t then reject  $H_0$

One tail test

NPTEL logo is visible in the bottom left corner.

Let us formulate our hypothesis,

$$H_0 : \mu = \mu_0$$

that means there is no difference, in

$$H_a: \mu < \mu_0$$

. If you read the problem decrease in heart rate, so we are saying  $\mu < \mu_0$ . So what do you do we take our t equation,

$$t = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

you remember this equation right,

$$t = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

this call the standard error and  $\mu_0$  is the population mean and this is a sample mean  $t$  degrees of freedom will  $n - 1$ .

$$df = n - 1$$

We calculate the  $t$  and then we compare table  $t$ , if the  $t$  calculated is less than the table  $t$  we accept  $H_0$ , the  $t$  calculated is greater than the table  $t$  we reject the  $H_0$ . Here, it is a 1 tail test please remember because we are talking about this type of situation, we are not saying

$\mu_0 = \mu_0$  that means less or more, but we are specifically talking about  $\mu$  a less than  $\mu_0$ . So it is a 1 tail test simple.

(Refer Slide Time: 05:47)

$H_0 : \mu = \mu_0$   
 $H_a : \mu < \mu_0$

$$t = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$df = n - 1$

If  $t_{\text{calc}} < \text{than the Table } t$  accept  $H_0$   
 If  $t_{\text{calc}}$  is  $>$  than the Table  $t$  then reject  $H_0$

One tail test

$t_{\text{calc}} = (240-282)/\sqrt{89.45} = -4.44$   
 $t_{\text{table}}$  for one tailed test for  $p=0.05, p=1.833$

So we reject the null hypothesis and conclude the decrease in heart rate is significant at 95% confidence limit

What we do is we  $t$  calculate I know 240 the other one is 282, if you remember we got 89.45, so square root of 89.45 is what I need put here.  $240 - 282$  I will get  $- 4.44$ . Now let us go to  $t$  table, with  $t$  table if I go for 1 tail test for  $p = 0.05$ , 1 tail test  $p = 0.05$  .

(Refer Slide Time: 06:16)



df	Probability in Two-Tail Distribution					
	0.10	0.05	0.025	0.01	0.005	0.001
1	1.638	1.638	1.638	1.638	1.638	1.638
2	1.617	1.617	1.617	1.617	1.617	1.617
3	1.605	1.605	1.605	1.605	1.605	1.605
4	1.594	1.594	1.594	1.594	1.594	1.594
5	1.583	1.583	1.583	1.583	1.583	1.583
6	1.573	1.573	1.573	1.573	1.573	1.573
7	1.563	1.563	1.563	1.563	1.563	1.563
8	1.554	1.554	1.554	1.554	1.554	1.554
9	1.545	1.545	1.545	1.545	1.545	1.545
10	1.536	1.536	1.536	1.536	1.536	1.536
11	1.527	1.527	1.527	1.527	1.527	1.527
12	1.518	1.518	1.518	1.518	1.518	1.518
13	1.509	1.509	1.509	1.509	1.509	1.509
14	1.500	1.500	1.500	1.500	1.500	1.500
15	1.491	1.491	1.491	1.491	1.491	1.491
16	1.482	1.482	1.482	1.482	1.482	1.482
17	1.473	1.473	1.473	1.473	1.473	1.473
18	1.464	1.464	1.464	1.464	1.464	1.464
19	1.455	1.455	1.455	1.455	1.455	1.455
20	1.446	1.446	1.446	1.446	1.446	1.446
21	1.437	1.437	1.437	1.437	1.437	1.437
22	1.428	1.428	1.428	1.428	1.428	1.428
23	1.419	1.419	1.419	1.419	1.419	1.419
24	1.410	1.410	1.410	1.410	1.410	1.410
25	1.401	1.401	1.401	1.401	1.401	1.401
26	1.392	1.392	1.392	1.392	1.392	1.392
27	1.383	1.383	1.383	1.383	1.383	1.383
28	1.374	1.374	1.374	1.374	1.374	1.374
29	1.365	1.365	1.365	1.365	1.365	1.365
30	1.356	1.356	1.356	1.356	1.356	1.356
31	1.347	1.347	1.347	1.347	1.347	1.347
32	1.338	1.338	1.338	1.338	1.338	1.338
33	1.329	1.329	1.329	1.329	1.329	1.329
34	1.320	1.320	1.320	1.320	1.320	1.320
35	1.311	1.311	1.311	1.311	1.311	1.311
36	1.302	1.302	1.302	1.302	1.302	1.302
37	1.293	1.293	1.293	1.293	1.293	1.293
38	1.284	1.284	1.284	1.284	1.284	1.284
39	1.275	1.275	1.275	1.275	1.275	1.275
40	1.266	1.266	1.266	1.266	1.266	1.266
41	1.257	1.257	1.257	1.257	1.257	1.257
42	1.248	1.248	1.248	1.248	1.248	1.248
43	1.239	1.239	1.239	1.239	1.239	1.239
44	1.230	1.230	1.230	1.230	1.230	1.230
45	1.221	1.221	1.221	1.221	1.221	1.221
46	1.212	1.212	1.212	1.212	1.212	1.212
47	1.203	1.203	1.203	1.203	1.203	1.203
48	1.194	1.194	1.194	1.194	1.194	1.194
49	1.185	1.185	1.185	1.185	1.185	1.185
50	1.176	1.176	1.176	1.176	1.176	1.176

Table B.1.1. A. Area under Standard Normal Curve. Probability in Two-Tail Distribution. Copyright 2010, Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

http://www.nptel.com/school/Statistics/Tables/

I need to look into this region. I get the 9 degrees of freedom because there are 10 rats, 10 - 1 is 9, it is 1.833. P from the table is 1.833, t calculated is - 4.44. It is greater, so we reject the null hypothesis at  $p = 0.05$ . We need to accept the alternative hypothesis that means, the heart rate has decreased because of anesthesia and this decrease is significant at 95 confidence interval. So you understand this it is very simple. We need to use this table as I said if you are talking about 1 tail test use the top numbers, if you are using 2 tail test we use the bottom number. When I said 0.05 for 2 tail I mean sorry 1 tail I used this, where has if I had said for a 2 tail I would have used this. As I said before also, 2 tail includes the outside area on both sides of my normal distribution curve, if I want to consider only one side obviously  $0.05 / 2$  that is way here it is 0.025.

(Refer Slide Time: 07:47)

$$\mu = \bar{X} \pm \frac{t_{df} S}{\sqrt{n}}$$

Change in heart rate 282-240=42

$S^2_{xx} = 89.45+89.45 = 178.9$

$S_{xx} = 13.4$

For 18 degrees of freedom , two tail,  $p=0.05$ ,  $t=2.1$  (from table)

Confidence limit for change in heart beat =  $42 \pm 2.1*13.4 = 42 \pm 28.14 = 14$  to 70

NPTEL

So you know how to use. Now I want to no change in the heart rate. What is the change? 42 is a change in the heart rate. So, I need to calculate the overall standard error understood. The overall standard error how do I do, both are same right 89.45 this is this variance of that 89.45 / 89.45 that comes out to be 178.9, I take square root of that I will get 13.4. Now I need to consider all the data sets. I have 1 set is 10, I have 9 degrees of freedom other set is 10, I have 9 degrees of freedom now I 18 degrees of freedom and I need to use 2 tail test, 2 tail table because we are talking about plus or minus here, you remember these equation right

$$\mu = \bar{X} \pm \frac{t_{df} S}{\sqrt{n}}$$

. So, I use the both the regions  $p = 0.05$  with 18 degrees of freedom, I read out this 2.101 I take 2.1 and 13.4 comes from here and 42 is the change in heart. The confidence limit for the change in heart at 95 % confidence is 14 to 70.

Do you understand the second part of the problem? The second part of the problem is you are suppose to find out the confidence limit for the change in the heart rate, change in the heart rate is originally it was 282 now it became 240. Now to find the confidence limit I use this equation if you remember this so, it is a

$$\pm \frac{t_{df} S}{\sqrt{n}}$$

. Now we have 89.45 for one set 89.45 for the second set, so total 178.9. I take a square root here, it becomes 13.4, that is the standard this whole portion. Now t I use 18 degrees of freedom do you understand why because, the first set has a 10 rats 9 degrees freedom, second has 10 rats 9 degrees of freedom totally 18 degrees of freedom. I used 2 tail because this is a  $\pm$  we are doing so and then t comes out form the table 0.05 18 degrees of freedom 2.101, so I substitute here I put 13.4. So when I do that I get the change in the heart rate, the confidence limit for the change in heart rate at 95 % confidence is 40 into 70. It is a interesting problem very very useful problem, we do quite a lot of these type of a studies in clinical trails we have a drug, you do give the drug and you see changes you know the original population and you are looking at change then you are trying to see whether the change is statistically significant or no. So this is varying.

(Refer Slide Time: 11:01)



The concentration of a drug in the brain of 6 rabbits is determined to 5 mins after administering the drug. Data is 40.2, 38.7, 41.6, 40.5, 43.2, 39.4 (mg/g). Calculate the 99% confidence interval for the true mean concentration. The mean concentration of the same drug in rats are reported to be 45 mg/g. Is there a statistical difference between these two data

NPTEL

Let us look at another problem, the concentration of a drug in the brain of 6 rabbit's is determined after 5 minutes, after administrating the drug. You are injecting a drug on 6 rabbit's then after 5 minutes we are measuring their concentration, it comes out to be 40.2, 38.7, 41.6, 40.5, 43.2, 39.5. Now calculate the 99 % confidence interval for the true mean concentration that we can do, the mean concentration of the same drug in rats are reported to be 45 milligram per gram now this is the sample data from rabbit's these the

data which they got from rats is there a statistical significance difference between these 2 data. So 2 parts, 1 is calculating the 95, 99 % confidence is very simple. We can calculate

$\bar{X}$  from this data then a for 5 degrees of freedom with 95 %, 99 % confidence from the

table we can calculate t and then we can calculate the standard error that is

$$\frac{S}{\sqrt{n}}$$

that will give you my confidence limit right.

(Refer Slide Time: 12:19)

The concentration of a drug in the brain of 6 rabbits is determined to 5 mins after administering the drug. Data is 40.2, 38.7, 41.6, 40.5, 43.2, 39.4 (mg/g). Calculate the 99% confidence interval for the true mean concentration. The mean concentration of the same drug in rats are reported to be 45 mg/g. Is there a statistical difference between these two data

$\bar{X} = 40.6, s = 1.6$

$cv = 100 \cdot 40.6 / 1.6 = 3.96\%$

$$t = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$df = n - 1$

If  $t <$  than the Table  $t$  accept  $H_0$   
 If  $t >$  than the Table  $t$  then reject  $H_0$

Two tail test

Then the second part of the problem. Let us look at this average so, the average comes out to be 40.6 and  $S = 1.6$ . Of course, we can calculate the CV also in this case, but mainly we are more interested in this. So, we take it as a 40.6, we take this as 45 then you know the

$$\frac{S}{\sqrt{6}}$$

and then degrees of freedom is 5, if  $t$  is  $<$  that the  $t$  calculated is less than the table  $t$  accept  $H_0$ ,  $t >$  table  $t$  reject  $H_0$ . We are doing 2 tail test because, the question is, in rats it

is 45 milligram per gram is there a statistical difference between these 2 data, is there are a statistical difference. The  $H_0$  will be  $\mu = \mu_0$ , h alternate will be  $\mu \neq \mu_0$ .

(Refer Slide Time: 13:00)

$$t = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$$df = n - 1$$
 If  $t_{calc} <$  than the Table  $t$  accept  $H_0$   
 If  $t_{calc}$  is  $>$  than the Table  $t$  then reject  $H_0$

Two tail test

NPTEL

So we do that do that two tail test. For a 5 degrees of freedom 99 % confidence.

(Refer Slide Time: 13:34)

$$t = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$$df = n - 1$$
 If  $t_{calc} <$  than the Table  $t$  accept  $H_0$   
 If  $t_{calc}$  is  $>$  than the Table  $t$  then reject  $H_0$

Two tail test

Table  $t$  value for 5 df , two tail, 99% confidence interval = 4.032

$t_{calc} = (40.6 - 45) / (1.6 / \sqrt{6}) = -6.73$

The data is significantly different

The difference in mean is = -4.4  
 CI on the mean (99%) = -1.74 to -7.05

NPTEL

Let us look at the table two tail test 99 % confidence means it is 0.01 for a 5 degrees of freedom you get this here 4.032. So this is the equation here,

$$\bar{X} - \mu_0$$

, we do  $40.6 - 45$  and then you are standard deviation is 1.6 for the data. So, divided by

$\sqrt{6}$

that will come to  $- 6.73$  the data is significantly different. Data is significantly different so that means, we can reject the null hypothesis. That means we accept the alternative hypothesis. Now the difference in the mean  $40.6 - 45$  is 4.4 right. So confidence interval on the mean so obviously

$t \times s / \text{square root } n$ ,  $t$  is 5 degrees of freedom. We get this

$$\frac{S}{\sqrt{n}}$$

So we get a confidence interval on the mean as this actually, the difference in the mean is this, confidence interval on the mean is this. Now we can use the GraphPad also to do the same calculation, we can use not only the GraphPad we can also use our excel also to do the calculation. Shall we look at excel? Quite simple.

(Refer Slide Time: 15:13)

The concentration of a drug in the brain of 6 rabbits is determined to 5 mins after administering the drug. Data is 40.2, 38.7, 41.6, 40.5, 43.2, 39.4 (ng/g). Calculate the 90% confidence interval for the true mean concentration. The mean concentration of the same drug in rats are reported to be 45 ng/g. Is there a statistical difference between these two data.

$\bar{X} = 40.6, s = 1.6$

$cv = 100 \times 40.6 / 1.6 = 3.96\%$

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$\alpha = 0.1$

If  $t <$  than the Table  $t$  accept  $H_0$   
If  $t >$  than the Table  $t$  then reject  $H_0$

Two tail test

Let us go back to excel, so the data is 40.2, 38.7, 41.6, 40.5, 43.2, 39.4 so we have a

sorry 38.7 and this data is not there. We have 6 data sets, now the command if you remember we go to this and there is something called a t-test. So when you look at the t-test what are the various things array 1. So we give the data set here and then we give the array 2. So here we give the 45, we give two tail test and then here we give pair or two sample t-test of course, as you can see we cannot do unfortunately we cannot do in excel this type of a calculations unfortunately. But we can use the other software here as you can see descriptive statistics, t-test here the GraphPad, here you have the 1 sample t-test. We can give the data here 40.2 then 38.7 then 41.6 and then 40.5, 43.2, 39.4 then we say we are comparing with 45. So, we have a 6 and then  $\mu_0$  is here so now when you say calculate now of course, this gives you only at a 95 % confidence interval. This is doing at the p value comes out to be 0.0011 and it is a statistically significant difference as you can see here. Even at 95 % confidence interval also it will come out to be statistically significant difference because if you look at the table here.

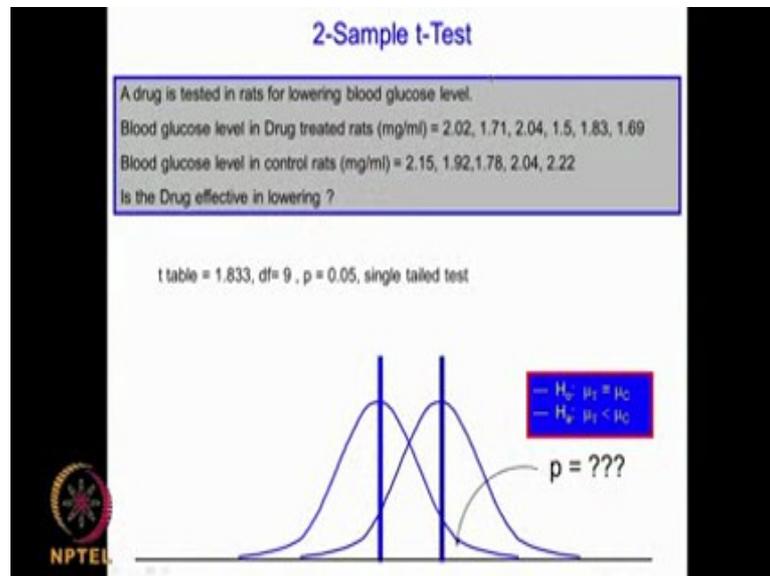
(Refer Slide Time: 18:25)

df	Probability in Two Tails (Combined)					
	0.99	0.95	0.90	0.80	0.70	0.50
1	6.381	6.965	7.625	8.163	8.602	9.000
2	6.985	7.709	8.453	9.000	9.400	9.800
3	7.457	8.155	8.933	9.500	9.900	10.300
4	7.879	8.602	9.388	10.000	10.400	10.800
5	8.256	9.000	9.779	10.400	10.800	11.200
6	8.591	9.388	10.136	10.800	11.200	11.591
7	8.898	9.779	10.469	11.200	11.591	11.969
8	9.181	10.136	10.781	11.591	11.969	12.331
9	9.441	10.469	11.074	11.969	12.331	12.679
10	9.679	10.781	11.350	12.331	12.679	13.015
15	10.247	11.591	12.101	13.015	13.409	13.747
20	10.691	12.101	12.601	13.409	13.850	14.191
25	11.081	12.601	12.991	13.850	14.241	14.581
30	11.451	13.015	13.381	14.241	14.631	14.971
40	11.981	13.747	14.011	14.850	15.081	15.421
50	12.401	14.331	14.501	15.081	15.471	15.811
60	12.751	14.850	14.951	15.471	15.861	16.201
70	13.041	15.331	15.381	15.861	16.251	16.591
80	13.291	15.781	15.781	16.251	16.641	16.981
90	13.511	16.201	16.151	16.641	17.031	17.371
100	13.701	16.591	16.501	17.031	17.421	17.761
120	14.081	17.331	17.201	17.421	17.811	18.151
140	14.411	17.971	17.781	17.811	18.201	18.541
160	14.701	18.501	18.281	18.201	18.591	18.931
180	14.961	19.031	18.701	18.591	18.981	19.321
200	15.191	19.561	19.051	18.981	19.371	19.711
250	15.631	20.401	19.631	19.371	19.761	20.101
300	16.001	21.081	20.081	19.761	20.151	20.491
400	16.491	22.321	20.801	20.151	20.541	20.881
500	16.891	23.331	21.331	20.541	20.931	21.271
600	17.231	24.151	21.781	20.931	21.321	21.661
700	17.531	24.811	22.181	21.321	21.711	22.051
800	17.801	25.351	22.541	21.711	22.101	22.441
900	18.051	25.801	22.871	22.101	22.491	22.831
1000	18.281	26.271	23.181	22.491	22.881	23.221

Yeah if you look at the table here for a 5 degrees of freedom, for a 99 % confidence interval it is 4.032, for 95 % confidence interval it is 2.571. So your t value which we calculate is quite high here 6.6892 so obviously that number is very high, your t value which your calculating is pretty high so obviously, that number is much higher than the table t whether it is at 95 % or whether it is at 99 %. You see that with the excel we do not have way of calculating the 1 sample t-test, where as with GraphPad we can do this type of a calculation with is. Now let us go to next type of tests that is called a two

sample t-tests.

(Refer Slide Time: 19:23)



That means, I have 2 drugs I am comparing the performance of 2 drugs and I am trying tell whether there is a statistically significant difference or not or I am comparing a control and drug A which I have introducing in the market or I am comparing already existing drug and another drug or I am controlling, I am comparing the performance of a drug on a male population and a female population. We have 2 sets of samples, so you will have a  $\bar{x}_1$  and  $\bar{x}_2$  that is two means and then we are trying to see whether each of this sample means comes from the same population or they come from different population. For example, we are telling a drug is tested in rats for lowering blood glucose level this is drug treated rats, this is the glucose level and where as this is the controlled rats this is the glucose level. So we have 2 sets of samples, 6 samples for a drug treated rats, 6 for the controlled. Now is the drug effective in lowering the blood glucose level. For example imagine this sample comes from one population like this, imagine this sample comes from another population like this, now are these populations or overlapping so much that we can consider it as one single population or they are different that is a whole idea of this and is the drug effective in lowering that means, this is a single tailed test, for totally we have t value so we can find out and we can tell whether there is a statistically significant difference by saying whether the drug is lowering or there is no different. The null hypothesis will be

$$\mu_T = \mu_C$$

that means a test and control we say drug treated we call it test, c means control whereas alternate could be mu t less than mu c.

(Refer Slide Time: 21:34)

**2- Sample t-Test**

$$t = \frac{\bar{X}_t - \bar{X}_c}{S_p \sqrt{\frac{1}{n_t} + \frac{1}{n_c}}}$$

$$S_p = \sqrt{\frac{(n_t-1)s_t^2 + (n_c-1)s_c^2}{(n_t-1) + (n_c-1)}}$$

$$df = (n_t-1) + (n_c-1)$$

If t calc < than the Table t accept Ho  
If t calc is > than the Table t then reject Ho

This is the equation by which these things are calculated. Suppose, we have a number of data points for one set of sample, number of data points of another set of sample so the degrees of freedom will be this + this - 2, that is why in this particular problem so may have 3, 4, 5, 6; 3, 4, 5, 6 + 5 is 11, 11 - 2 is 9 degrees of freedom do you understand. So here, we have taken 6 data points for a drug treated here we have taken 5 data point. The advantage of 2 sample t-test I could have different sets of samples also in fact this particular equation if you look at this equation very carefully I can have any value of n h, I can have any value of n l so it did not be the same. So t value is given by this equation

$$\bar{X}_L$$

$$X_H -$$

that means, mean of one data set - mean of another data set or sample set

$$\sqrt{\frac{1}{n_H} + \frac{1}{n_L}}$$

at then we have something called  $S_p$  here, this is the standard error. If you look it very carefully this for a one sample we used to have  $S_p / \sqrt{n}$ . This almost analogous to that here we have square root of  $1/n_H + 1/n_L$ , but here  $S_p$  it looks slightly different

$$S_p = \frac{(n_H - 1)s_H^2 + (n_L - 1)s_L^2}{(n_H - 1) + (n_L - 1)}$$

the equation for  $S_p$  here. There are some similarities if you look at it this equation and the normal t equation for one sample t-test and so on. So, it is quite simple. We have 2 data sets I get the mean of this, I get the standard deviation of this  $n_H = 6$ ,  $n_L = 5$ . I substitute the standard deviations here square it here it will become  $6 - 1$ ,  $5 - 1$  so we here  $1/6$   $1/5$  and then here you can get the mean of both the data sets and you get the answer. So we can use this simple excel to calculate means and standard deviations and that is what I will do. That is the whole idea is we try to calculate t using fundas it is easy to use software's whether excel command like t-test it is easy to use the GraphPad like software or any commercial software, but then you need to know what is the underline equation that is very, very important and that is what I would like to emphasize on.



Although you can say I will reject the null hypothesis is a 95 % confidence, I would not go too much I may try to do more experiments so that we get a better difference between the t calculated and the t table do you understand. This is a very close call actually.

(Refer Slide Time: 27:26)

**TTEST(array1,array2,tails,type)**  
Returns probability associated with a Student's t-Test

Array1 is the first data set.  
Array2 is the second data set.  
Tails specifies the number of distribution tails. If tails = 1, TTEST uses the one-tailed distribution. If tails = 2, TTEST uses the two-tailed distribution.  
Type is the kind of t-Test to perform.

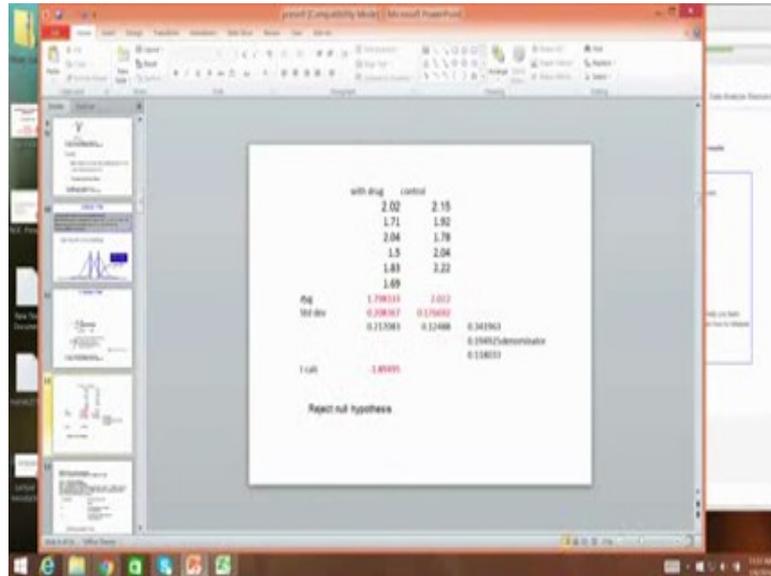
If type equals	This test is performed
1	Paired
2	Two-sample equal variance (homoscedastic)
3	Two-sample unequal variance (heteroscedastic)

From Excel probability = 0.043



The excel also can do this and I showed you last time just few minutes back excel cannot do one sample t-test, but it can do a two sample t-test, it can do a paired two sample t-test, it can do a two sample t-test when equal variance, two sample t-test when the unequal variance so the commands are array 1, array 2 tails and type. Array 1 is 1 first data set, array 2 is second data set, tails is one tail, 2 means two tail. So, from excel we get 0.043 that is the probability, so it is less. We can say we will reject the null hypothesis, but it is very close call. How do you do it in excel? We just put these data inside and then we do the calculations.

(Refer Slide Time: 28:22)



We say 2.02, 1.71, 2.04, 1.5, 1.83, 1.69 then here 1.71, 2.04 then 2.15, 1.92, 1.78, 2.04 2.2 so use the same command t test. So we give the first data set and then comma we give the second data set then we comma it is a 1 sample t-test and we can give equal variance or unequal variance so if I put 3 .043. Basically I have given it as 3 that means unequal variance so I can even give it as equal. For unequal variance I am getting 0.43 yes the variances are unequal as you can see standard deviations are unequal here. It is giving 0.43 so we reject the null hypothesis hence, accept the alternate hypothesis. Same thing we can do with the, a GraphPad software also.

But you see it is going to be very close in GraphPad software. So what do we do? How do we do this in GraphPad? Let us go back there so here we have the t-test command compare 2 means. So we go there and then we continue.

(Refer Slide Time: 30:38)



Because the data the t calculated and t table are very close to each other some software's may say it is not statistically significance, some software may say statistically significant. Especially you need to watch out with the data is very close to each other that is t calculated and t table is better to do more experiments to have more confidence in your conclusion. So we will continue on these t-tests in the next class also.

Thank you very much