

Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology – Roorkee

Lecture – 40
Hypothesis Testing: Two Sample Test- II

Hello friends. I welcome you all in this session. As you are aware, in previous session we were discussing about do sample hypothesis testing and before that we did work out a couple of examples using Minitab on a one sample hypothesis testing. So let us look at some more examples on two sample hypothesis testing and we will work out those examples using Minitab as well.

So this is the question which we worked out in previous class. We wanted to know in which share market or in which stock exchanges you would like to invest in. So we did compare mean of these two stock exchanges and we said that the null hypothesis is that the mean returns from these two stock exchanges would be same and it would be different.

(Refer Slide Time: 01:14)

Pooled-Variance t Test Example

You are a financial analyst for a brokerage firm. Is there a difference in dividend yield between stocks listed on the NYSE & NASDAQ? You collect the following data:

	NYSE	NASDAQ
Number	21	25
Sample mean	3.27	2.53
Sample std dev	1.30	1.16

Assuming both populations are approximately normal with equal variances, is there a difference in mean yield ($\alpha = 0.05$)?

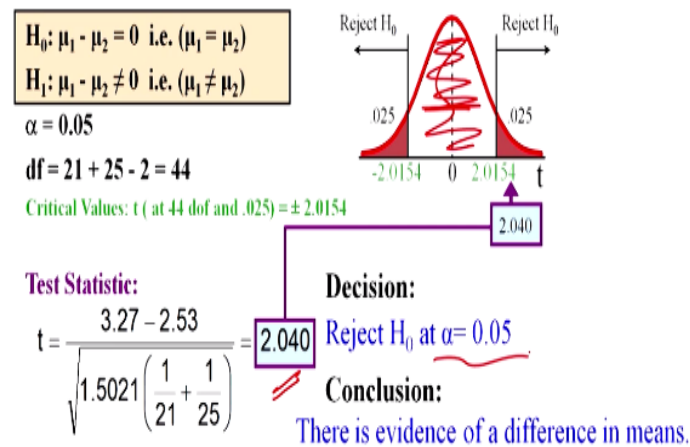


The alternative hypothesis and real said that a calculated value is outside non rejection region. This is your non rejection region. So we will reject null hypothesis. Now, we are rejecting this null hypothesis and we are saying that the mean return from these three stock exchanges, it is not

same. It is different, so different means in, in one of the stock exchanges are return is higher rate. So in which stock action in return is higher.

(Refer Slide Time: 01:53)

Pooled-Variance t Test Example: Hypothesis Test Solution



It is in New York Stock Exchange because sample mean is higher than sample mean of NASDAQ? So let us work out the same example using Minitab and let us see what is the conclusion there? So here we are rejecting the null hypothesis, so let us look at this question and you just keep in mind the number of a number over here, sample mean and standard deviation of these two.

So we will work, we are called this example using Minitab, so we will go to stat first, basic statistics and then we will go for two sample test, so in fact, there is, there is something or two sample t test. In fact, there is the, in, in, in this particular software, there is nothing like two sample z test infected calculates t test even far larger sample size. So we will go it two sample t test, since in our case, if you look at sample size, then the first sample has got size 21 and second has got 25.

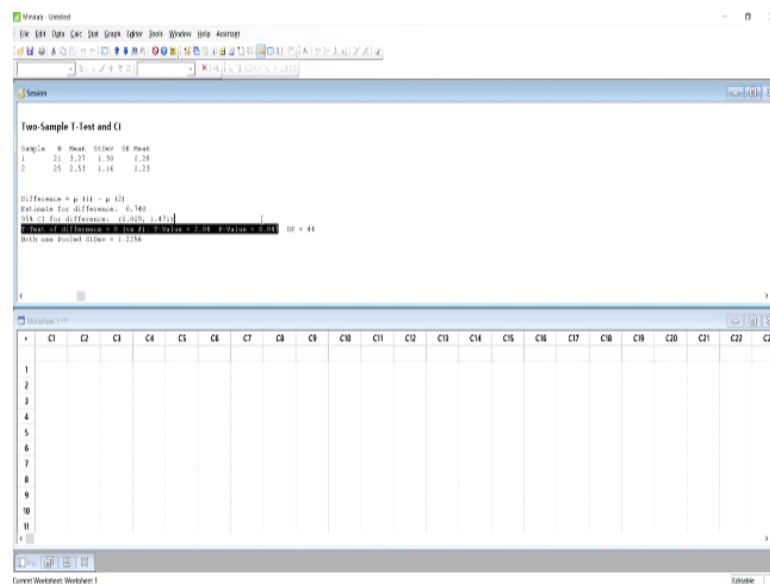
So it would be a case of t test, so you can help both Samples in column or you may have each sample in its own column and otherwise you can have summarized data, so we will click at summarize data, sample one, so we have to enter for sample size for sample one and let values 21. Then we need to write sample mean sample mean is it is 3.27 and the standard deviation is

1.30. Similarly, you need to enter data for sample 2, sample size here is 25 25 2.53 is sample mean and standard deviation is there is look it, it is 1.16 now.

After making all these entries go to options, then you need to test this hypothesis that 95% significance level hypothesize difference is zero, because we are saying, we are saying that, the returns are same from both these stock exchanges, and ultimately a hypothesis, they are not same. So this is not equal to buy, assume equal variances. That is our assumption that this, the populations from where we are drawing samples, those populations have equal variance, then we will click okay and okay, once again, let us look at output. Here.

Output is if you look at alternative hypothesis, in fact, the difference we are saying, is it is, it is, they are same $\mu_1 = \mu_2$. In fact, here it is written $\mu_1 - \mu_2 = 0$, which is as good $\mu_1 = \mu_2$.

(Refer Slide Time: 05:48)

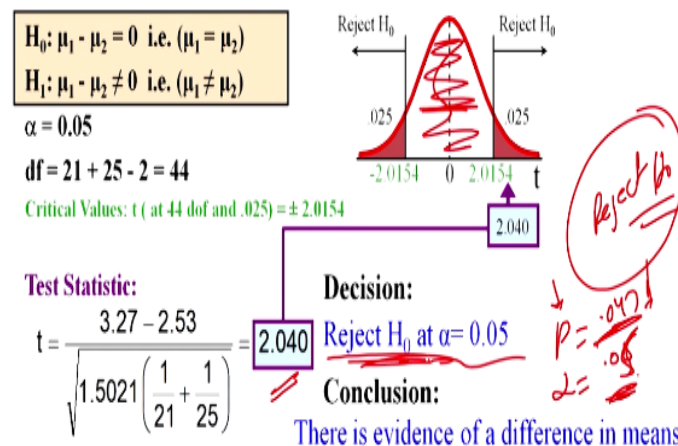


Now the p value over here is 0.045, even mines point 0.047. And what is, what is the value of Alpha p and alpha = 0.05, so is this, is p value less than alpha? No. So will will not reject null hypothesis, and this is the same conclusion which we found in critical value approach, so we will not reject null hypothesis value while looking at p value. And the similar was the conclusion. So in fact, Okay. Just look at this once again, p value here is 0.047. p is 0.047 and alpha = 0.05.

So p is less than alpha. So in fact it was my mistake. So p value is less than alpha. So we reject the null hypothesis, so the decision is rejecting null hypothesis and non rejection of null hypothesis, so we will reject null hypothesis because p is less than alpha. So either you go by p value approach, or by critical value approach will reject null hypothesis.

(Refer Slide Time: 07:20)

Pooled-Variance t Test Example: Hypothesis Test Solution



So let us look at some more examples. In fact, this is the table which we have seen in previous example as well when we solved that question using critical value approach. So let us look at one more example. Is there any difference between earnings of students of two IITs check this at 0.05 significance level? So let us say you have got mean earnings of students of IIT,. Let us say Bombay and Roorkee, and I think it would look.

So after selecting a sample of 200 stones, they are mean earning was let us say 8.95 lakh but a now, and this is after selecting one 75 students, this is 9.10 and after calculating mean, we also calculated standard deviation of these two samples, so we want to know is there any significant difference between earnings of these two students of IIT, look at IIT. So what? What should be the first step?

The first step is you need to frame null hypothesis and alternative hypothesis, so what would be the null hypothesis? The null hypothesis is this, that there is no difference so $\mu_1 = \mu_2$ and of course alternative would be $\mu_1 \neq \mu_2$. So this is what type of test, so it is one tell test or two tell

test, it is a one, it is, it is not one tell test two tell test because this is, the sign is not equal to type, so we will hope two rejection regions.

And if you look at the sample size over here and both the emphasize are more than 30. So the appropriate statistic would be what would it be? Z. So when, when we applied z statistics over here, so this how you can calculate their statistics.

(Refer Slide Time: 09:29)

Ex: Is there any difference b/w earnings of two IITs at SL= 0.05 ?????

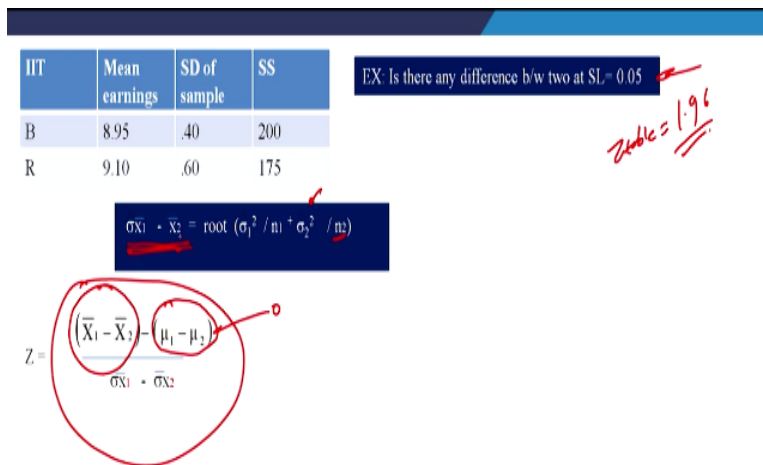
IIT	Mean earnings	SD of sample	SS
B	8.95	0.40	200
R	9.10	0.60	175

$Z_{\alpha/2}$
 $H_0: \mu_1 = \mu_2$
 $H_a: \mu_1 \neq \mu_2$

So this is the, the difference between these two samples. Sample means and we are saying $\mu_1 - \mu_2 = 0$. So this term is zero and divided by difference of mean. So this is under root of sigma. One square divided by first sample size. Then this is sigma two squared divided by sample size of the second sample. you will to taste this second, this hypo exit 0.05 significance level, and what and using critical approach method if you want to solve this question.

So for critical value approach, we know that the z double value would be what I take is 1.96. So you have got all these information over here. So you got $\bar{x}_1 - \bar{x}_2$ difference between this too, this, this time is zero. And you need to just to calculate the difference of these two standard errors of mean.

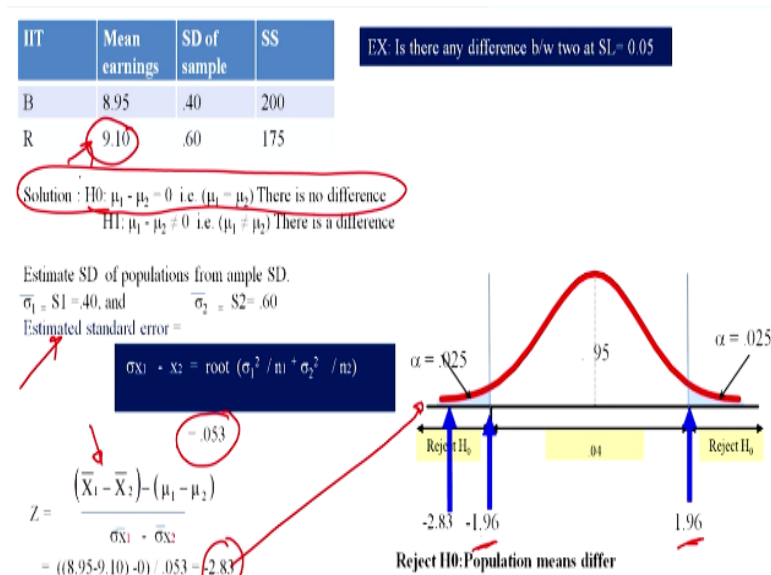
(Refer Slide Time: 10:51)



So when you solve this question this, this term is equal to 0.053 the estimated standard error is 0.053 and z value is this. So calculated z is -2.83 and the table critical values 1.96 plus and one 1.96 minus. So the calculated t value is, sorry, z value is in the rejection region? So we will reject the null hypothesis. Is not it? So we will reject the null hypothesis. What is the null hypothesis?

But there is no significant difference in the earning of these two groups of students. So we are rejecting it means there is significant difference, so when we said they did the difference it means which, which group has got her salary, this group, because means higher sample mean is higher over here. So let us look at, we will, we will work out the same example using Minitab.

(Refer Slide Time: 12:11)

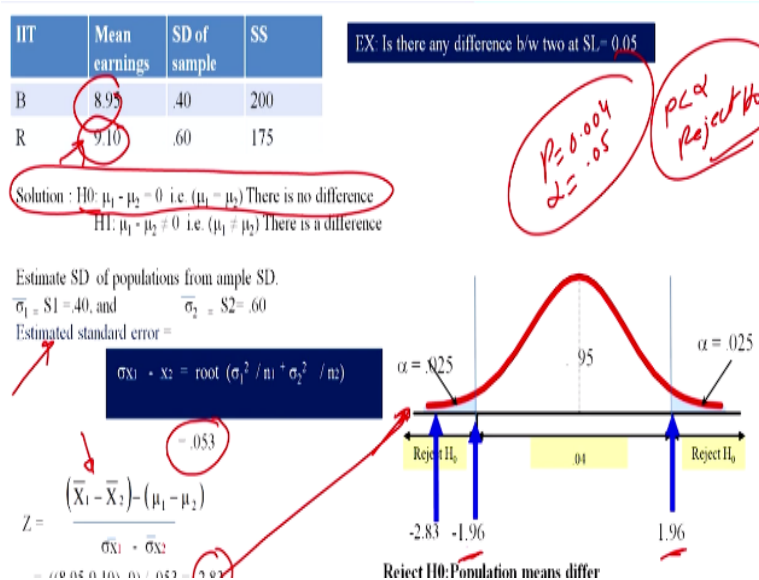


Okay. So let us look at Minitab and workout. We will work on this example. So this is the case of a, again, it is a two sample. We do not have any option of z test, so we will click it. Two sample t test, so we have got two samples to the first sample. It is emphasize is 200 sample mean is 8.95 which is here, is not it? Then Standard Deviation is 0.40 0.40. Similarly, will make entries for sample2. So sample size is 175, 175 and 9.10.

Sample mean is 9.10. There is look at standard deviation, I think at this 0.60. yes, 0.60. We will go to options. Options we have got significance level is 0.05 so this is 95% confidence level. Hypothetically, the difference is zero again. And alternative hypotheses, again, same. There is no difference between these two groups? And we will do assume equal variances. So we will click it OK. Again OK.

So p value is this 0.004. So p values 0.004 alpha is 0.05 it is $p < \alpha$ so $p < \alpha$. So we will reject null hypothesis, this is the conclusion we are drawing from probability, very probably devalue approach. And this was the result using critical value approach, the results are same, so you can work out, an example like this using Minitab as well and you will get the same answer, let us look at one more question.

(Refer Slide Time: 14:43)



Two independent samples of observations for a collected, for the first sample of 60 elements, the mean was 86 and the standard deviation was 6. The second sample of 75 elements had a mean of

82. So if you look at this in this case, sample size is 60 mean is 86 sample size 75 mean is 82 and standard deviation is 9, so what we have to do, first of all, compute the standard error for difference between these two means.

So this is what we calculated in previous example as well. This is what we have to calculate, so this is what we are supposed to calculate and then using $\alpha = 0.01$ test whether the two samples can reasonably be considered to have come from populations with the same mean. So our null alternative hypothesis, what would be what? So your null hypothesis is that these two means are same.

(Refer Slide Time: 16:34)

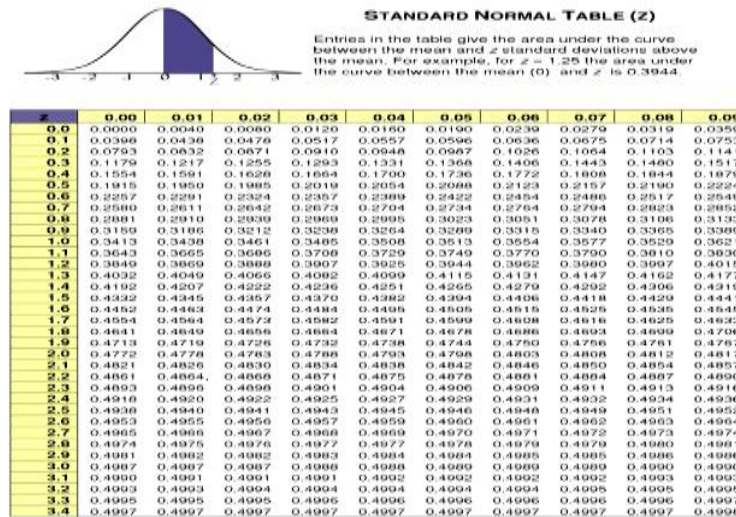
Ex: Two independent samples of observations were collected. For the first sample of 60 elements, the mean was 86 and the standard deviation 6. The second sample of 75 elements had a mean of 82 and a standard deviation of 9.

- (a) Compute the estimated standard error of the difference between the two means.
- (b) Using $\alpha = 0.01$, test whether the two samples can reasonably be considered to have come from populations with the same mean.

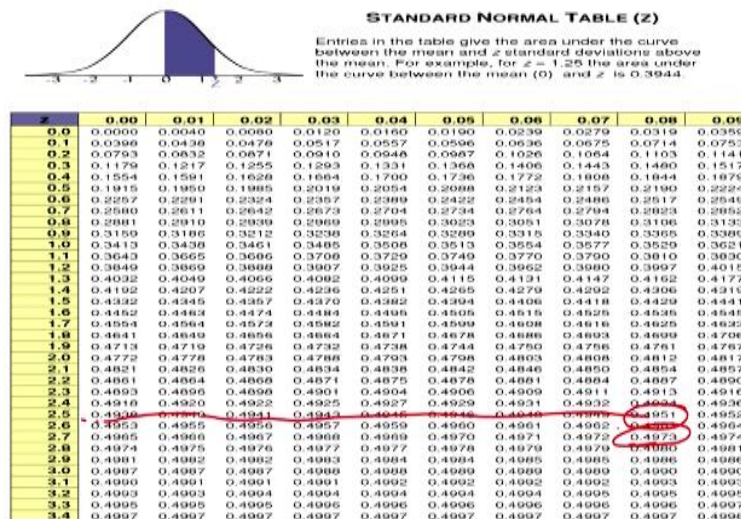
*Ans: $H_0: \mu_1 = \mu_2$
 $H_a: \mu_1 \neq \mu_2$*

This null hypothesis and the alternative hypothesis that they are not same, so you can solve this question using a critical value approach.

(Refer Slide Time: 16:44)



So far for this Alpha 0.01 we need to look at the table, which is actually 2.58. So it is 2.58. So 2.5 and 8 this is the one, let me delete this first. So again, let me tell you where that value is, so first of all, 0.01 is the alpha, so what you are supposed to do it is $1 - 0.01$, $8.99/2$. So whatever is that value, you do find out z value, that probability, so that value is at 2.58; 2.5, this is the one, 2.58 I think. No, this is the 0.4951. So 0.4951 is basically nothing but $0.99/2$ is in it. (Refer Slide Time: 17:45)



So that is your table critical value, and but what we have to find out is this right in the first part of the question. So this is very simple. You just put, values of n_1 and n_2 and sample standard variance. Variance of these two samples, since we have been given a standard division, you just

take a square of this, so s_2 is, this is s_2 square is 81, s_1 is 6 s_1 is our variance is 36, so it 1.29. Now when you look at z value, when you calculate the z value, this is the difference.

You can easily calculate. So you know, they say it be 86-82 so this is, this term is 4 and this term is zero. So $4 / 1.29$ is $4/1.296$ you will get 3.09 so if you draw distribution for this particular example, you have got these two rejection regions, this is also rejection region and, this is 2.58 - this 2.58 plus and your calculated z values is what 3.09. 3.09 means the somewhere here, three one zero nine so what will you do? Will you reject it? We will do not reject the null hypothesis.

You will reject null hypothesis because z value is in the rejection region. Okay? So we are saying that will reject null hypothesis and we will conclude that it is reasonable to conclude that the two samples come from different populations, they do not come from same population. That is our conclusion.

(Refer Slide Time: 19:55)

$$\begin{aligned}
 & s_1 = 6 \quad n_1 = 60 \quad \bar{x}_1 = 86 \quad s_2 = 9 \quad n_2 = 75 \quad \bar{x}_2 = 82 \\
 & (a) \quad \hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{36}{60} + \frac{81}{75}} = 1.296 \\
 & (b) \quad H_0: \mu_1 = \mu_2 \quad H_1: \mu_1 \neq \mu_2 \quad \alpha = 0.01 \\
 & \quad \text{The limits of the acceptance region are } z = \pm 2.58, \text{ or} \\
 & \quad \bar{x}_1 - \bar{x}_2 = 0 \pm z \hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \pm 2.58(1.296) = \pm 3.344 \\
 & \quad \text{Because the observed z value} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} \\
 & \quad = \frac{(86 - 82) - 0}{1.296} = 3.09 \\
 & \quad = 3.09 > 2.58 \text{ (or } \bar{x}_1 - \bar{x}_2 = 86 - 82 = 4 > 3.344), \text{ we reject } H_0. \text{ It is} \\
 & \quad \text{reasonable to conclude that the two samples come from different} \\
 & \quad \text{populations.}
 \end{aligned}$$

Now we will look at the same example using Minitab. Let us look at data first. So this is 60, 86 and 6. So we will go to stat, basic statistics. Yes, two sample t test 60, 60, 86 and 6. So this reasonable lead information given about sample one 60, 86 and 6. Let us look at sample2. It is 75, 82 and 9. Now you need to click adoption since our alpha is 0.01 so this is at 99.0, hypothesize differences is zero again and of course alternative hypotheses also given as.

There is no significant difference and we are assuming we are the answers. Let us click it, click it. Okay. And again, okay, so p values 0.40 is this less than alpha. 0.004 is over p value and alpha value is 0.01. So we will say that we will lose less than alpha. So we will reject null hypothesis, which is the decision we will taken here as well when we used critical value approach.

Now in this example, let us look at this. In 2013, the financial accounting standards board was considering a proposal to require companies to report the potential effect of employee stock options on earnings per share EPS. So what FASB has to did this? They collected 41 high technology firms later revealed that the new proposal would reduce EPS by an average of 13.8% with a standard deviation of 18.9.

So FASB collected data from high technology firms and the mean was 13.8%. It is the standard deviation was 13.9%. A random sample of 35 producers of consumer goods, these are, these are some different set of companies, consumer goods slowed that the proposal would reduce EPS by 9.1% on average with a standard deviation of 8.7 percentage on the basis of these samples. Is it reasonable to conclude that FASB proposal will cause a greater reduction in EPS for high technology firms then producers of consumer goods?

So here, let us called μ_1 is high technology firms and μ_2 is consumer goods companies and what is our null hypothesis is? It would be $\mu_1 = \mu_2$. And alternative hypothesis would be what proposal we will cause a greater reduction in EPS first. So $\mu_1 < \mu_2$ this is our alternative hypothesis.

(Refer Slide Time: 24:22)

Ex. In 2013, the Financial Accounting Standards Board (FASB) was considering a proposal to require companies to report the potential effect of employees' stock options on earnings per share (EPS). A random sample of 41 high-technology firms revealed that the new proposal would reduce EPS by an average of 13.8 percent, with a standard deviation of 18.9 percent.

A random sample of 35 producers of consumer goods showed that the proposal would reduce EPS by 9.1 percent on average, with a standard deviation of 8.7 percent. On the basis of these samples, is it reasonable to conclude (at $\alpha = 0.10$) that the FASB proposal will cause a greater reduction in EPS for high-technology firms than for producers of consumer goods?

μ_1 μ_2
 $\mu_1 \neq \mu_2$

Now, we will work out this question right now. If you will see this question then he has emphasizes are more than 13 both cases, you know what, 41 and yogurt 35. So we will be applying again, z test. So we need to calculate the critical value approach first. So for critical value approach at 0.10 so at this $0.90/2 = 0.45$ so look at the value of z when area under the curve is 0.45. 0.45 area under cover is this is here.

So this is somewhere here, 1.64 is not it? So we will say that the critical values here are + 1.64 and - 1.64. Okay. It is not one point in it is 1.28 actually. So let us look at what is their 1.2, 1.28, this is, it is 0.4 and in fact this is, this is the one tell test. That is why this had it been a two tell test. This would have been the value of z. But since this is a case of one tell test, and we just, again, I left tell tests.

So the distribution is like this, so this is your rejection regenerate, and this year mean, so that is why we know that this area is 0.5, this areas 0.5, so why we are looking, I said when, when area under the curve is 0.4 because this is our significance level is 0.10, so this is 0.5, $0.5 - 0.1$ becomes 0.40. So look at 0.40 probability in this table, which is here, 1.28. So the first part will remain same. This is the difference in the standard error.

In fact, this question, in fact, if you look at once again, then this is in fact what we are in with. What you need to do this is a consumer goods. This be a high tech one and this be a consumer

goods, so we have to just reverse it, rather than, having it, less than type, you can get it, her turning the hypothesis like this, a greater than type, because we all do see that the proposal will cause a greater reduction in EPS for high technology, so reduction is more in high tech compared to consumer goods.

So you need to reframe your null alternative hypothesis like this, so μ_1 and μ_2 are equal and $\mu_1 > \mu_2$, so this is an upper tail test. Of course, Instead of this, now this is your rejection agent, right with me. Repeat it. Let me raise everything first and let me redraw distribution. How this tuition all looked look like. So this is your distribution. This would be your rejection agent. One tail test, so $\mu_1 > \mu_2$.

(Refer Slide Time: 28:15)

Ex. In 2013, the Financial Accounting Standards Board (FASB) was considering a proposal to require companies to report the potential effect of employees' stock options on earnings per share (EPS). A random sample of 41 high-technology firms revealed that the new proposal would reduce EPS by an average of 13.8 percent, with a standard deviation of 18.9 percent.

A random sample of 35 producers of consumer goods showed that the proposal would reduce EPS by 9.1 percent on average, with a standard deviation of 8.7 percent. On the basis of these samples, is it reasonable to conclude (at $\alpha = 0.10$) that the FASB proposal will cause a greater reduction in EPS for high-technology firms than for producers of consumer goods?

$$\frac{13.8 - 9.1}{3.29}$$

So after calculating standard error difference, which is 3.92 find out value of. So this term is zero. $\bar{x}_1 - \bar{x}_2$ which is this $13.8 - 9.1 / 3.29$. So this is equal to 1.43 and z values 1.28, so this is how you should be drawing your distribution. So this is your 1.28 and this is your rejection region and calculators and values somewhere here, which is 1.43 so you reject the null hypothesis.

And when you reject the null hypothesis, that means you are accepting alternative hypothesis and it means that the proposal will cause a significant greater reduction in EPS for high technology firm. So whatever you are assuming in this question is correct, so this is how you can solve this

type of question using Minitab as well. So let us look at how to solve this question using Minitab. So we will go to state first basic statistics, then two sample t test will enter data.

So let us look at our data. So you have got 41, 13.8 and 18.9 so 41, 13.8 and the third one is 18.9 let us look at data for a sample 2 this 35 9.1 and 8.7, 35 9.1 and 8.7, let us look at options. So options this wheel to taste it, 0.1 it means 99 90%, that is 0.10, it is not 0.01, so the alternative hypothesis is greater than type.

So it is, it is an upper tell test. Okay. And again, okay. Let us look at p value p value is 0.091, 0.091 this is p value and alpha is 0.10. So p is less than alpha. So p is less than alpha. So we reject the null hypothesis, this is what we concluded in critical value approach as well as.

(Refer Slide Time: 31:37)

Sample 1 (HT firms): $s_1 = 18.9$ $n_1 = 41$ $\bar{x}_1 = 13.8$ ✓

Sample 2 (CG producers): $s_2 = 8.7$ $n_2 = 35$ $\bar{x}_2 = 9.1$ ✓

$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$ $\alpha = 0.10$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(18.9)^2}{41} + \frac{(8.7)^2}{35}} = 3.298 \text{ percent}$$

The upper limit of the acceptance region is $z = 1.28$, or

$$\bar{x}_1 - \bar{x}_2 = 0 + z\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = 1.28(3.298) = 4.221 \text{ percent}$$

 Because the observed z value = $\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} = \frac{(13.8 - 9.1) - 0}{3.298}$

$= 1.43 > 1.28$ (or $\bar{x}_1 - \bar{x}_2 = 4.7 > 4.221$), we reject H_0 and conclude that the FASB proposal will cause a significantly greater reduction in EPS for high-tech firms.

Handwritten notes: A red circle around '1.43' with 'p = 0.091' written next to it. A red circle around '3.298' in the denominator of the z-value formula. A red checkmark next to the conclusion.

So with this, let me stop here and in next session, we will work out couple of more examples on do sample test for proportion as well, not only for me. Thank you very much.