Video 11

HYPOTHESIS TESTING:

SINGLE SAMPLE T-TEST WITH UNKNOWN σ

BUSINESS ANALYST SALARIES

A report from 6 years ago indicated that the average gross salary for a business analyst was $69,873. Since this survey is now outdated, the Bureau of Labor Statistics wishes to test this figure against current salaries to see if the current salaries are statistically different from the old ones.

Based on this sample, we found $s = \$14,985$. We do not know σ and will therefore have to estimate it using s.

For this study, the BLS will take a sample of 12 current salaries.

Step 1: Establish Hypothesis

$$H_0\colon \mu = \$69,873 \qquad H_a\colon \mu \neq \$69,873$$

Step 2: Determine Appropriate Statistical Test and Sampling Distribution

This will be a two-tailed test. Salaries could be higher OR lower.

Since σ is unknown and n is small, we will use the t-distribution.

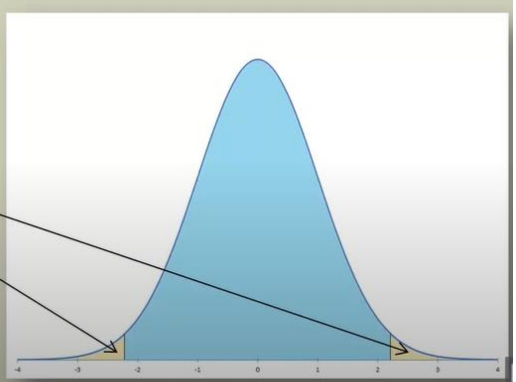
$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Step 3: Specify the Type I error rate (significance level)

$$\alpha = .05$$

Step 4: State the decision rule

For $df = 11$ $\begin{cases} \text{If } t > 2.201, \text{ reject } H_0 \\ \text{If } t < -2.201, \text{ reject } H_0 \end{cases}$



Step 5: Gather data

$$n = 12, \bar{x} = \$79{,}180$$

Step 6: Calculate test statistic

$\bar{x} = \$79{,}180$

$\mu_o = \$69{,}873$

$s = \$14{,}985$

$n = 12$

$$t = \frac{\$79{,}180 - \$69{,}873}{\frac{\$14{,}985}{\sqrt{12}}}$$

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

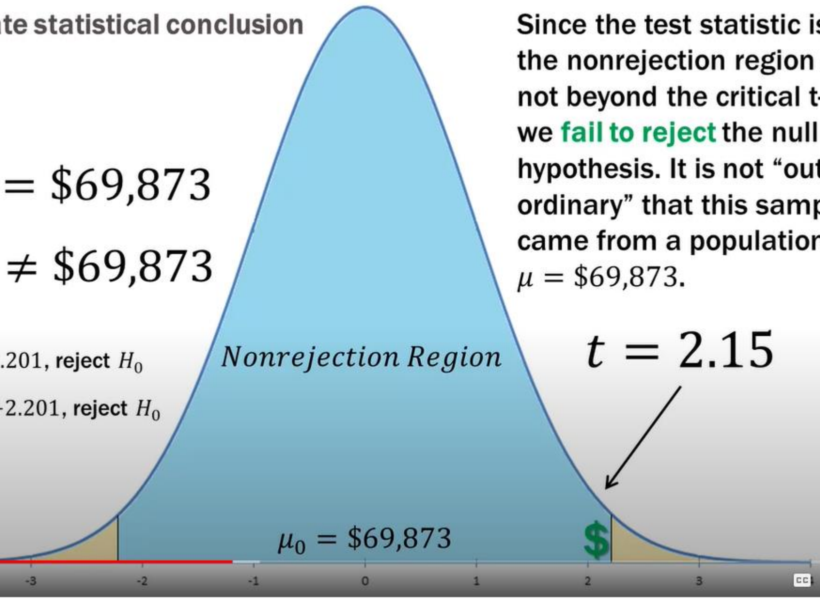
$$t = 2.15$$

Step 7 & 8: State statistical conclusion

H_0: $\mu = \$69{,}873$

H_a: $\mu \neq \$69{,}873$

For $df = 11$ $\begin{cases} \text{If } t > 2.201, \text{ reject } H_0 \\ \text{If } t < -2.201, \text{ reject } H_0 \end{cases}$



Since the test statistic is in the nonrejection region and not beyond the critical t-value, we **fail to reject** the null hypothesis. It is not "out of the ordinary" that this sample came from a population with $\mu = \$69{,}873$.

BUSINESS ANALYST SALARIES, $n = 15$

Step 6: Calculate test statistic

$\bar{x} = \$79{,}180$

$\mu_o = \$69{,}873$

$s = \$14{,}985$

$n = 15$

$$t = \frac{\$79{,}180 - \$69{,}873}{\frac{\$14{,}985}{\sqrt{15}}}$$

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$t = 2.41$$

Step 7 & 8: State statistical conclusion

$H_0\!: \mu = \$69{,}873$

$H_a\!: \mu \neq \$69{,}873$

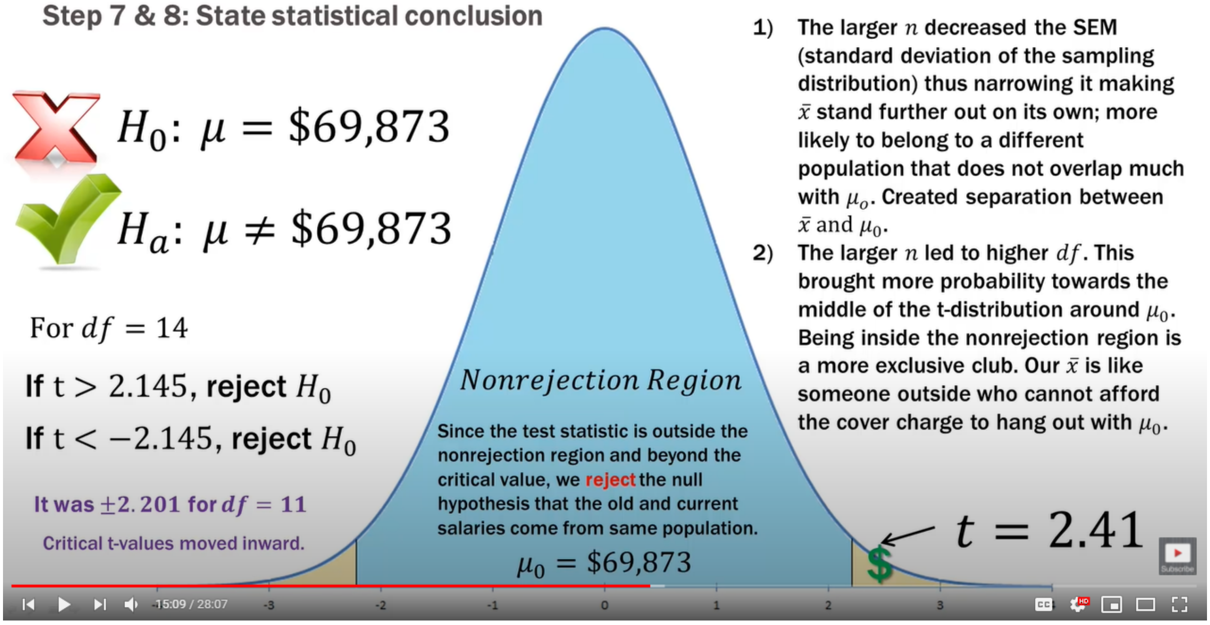
For $df = 14$

If $t > 2.145$, reject H_0

If $t < -2.145$, reject H_0

It was ± 2.201 for $df = 11$

Critical t-values moved inward.



Why?

1) The larger n decreased the SEM (standard deviation of the sampling distribution) thus narrowing it making $\bar{x}$ stand further out on its own; more likely to belong to a different population that does not overlap much with μ_o. Created separation between $\bar{x}$ and μ_0.
2) The larger n led to higher df. This brought more probability towards the middle of the t-distribution around μ_0. Being inside the nonrejection region is a more exclusive club. Our $\bar{x}$ is like someone outside who cannot afford the cover charge to hang out with μ_0.

STARBUCKS CUSTOMER SATISFACTION

Starbucks is interested in assessing customer satisfaction in the Canadian city of Toronto, Ontario. To conduct the study, Starbucks asked 25 customers in the city:

"Compared to other coffee houses in Toronto, would you say the customer service at Starbucks is much better than average (5), better than average (4), average (3), worse than average (2), or much worse than average (1)?" (Likert scale)

The mean rating was determined to be 3.50. Based on this sample, the standard deviation was found to be $s = 1.4$.

Step 1: Establish Hypothesis

$$H_0: \mu \leq 3 \qquad\qquad H_a: \mu > 3$$

Step 2: Determine Appropriate Statistical Test and Sampling Distribution

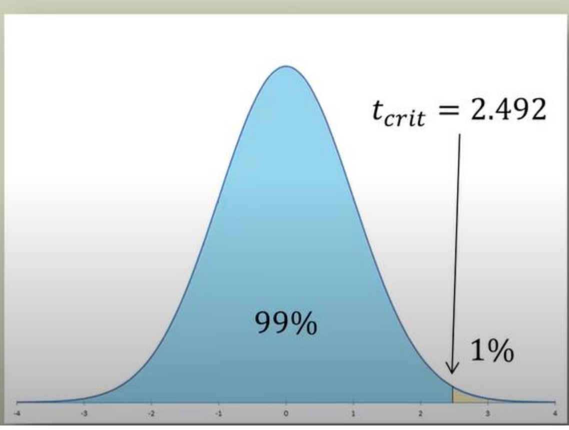
This will be a one-tailed test. Starbucks is interested in a better than average rating.

Since σ is unknown and n is small, we will use the t-distribution.

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Step 3: Specify the Type I error rate (significance level)

$$\alpha = .01$$



Step 4: State the decision rule

$For\ df = 24$

If t > 2.492, reject H_0

Step 5: Gather data

$$n = 25, \bar{x} = 3.5$$

Step 6: Calculate test statistic

$\bar{x} = 3.50$

$\mu_o = 3$

$s = 1.4$

$n = 25$

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$t = \frac{3.50 - 3}{\frac{1.4}{\sqrt{25}}}$$

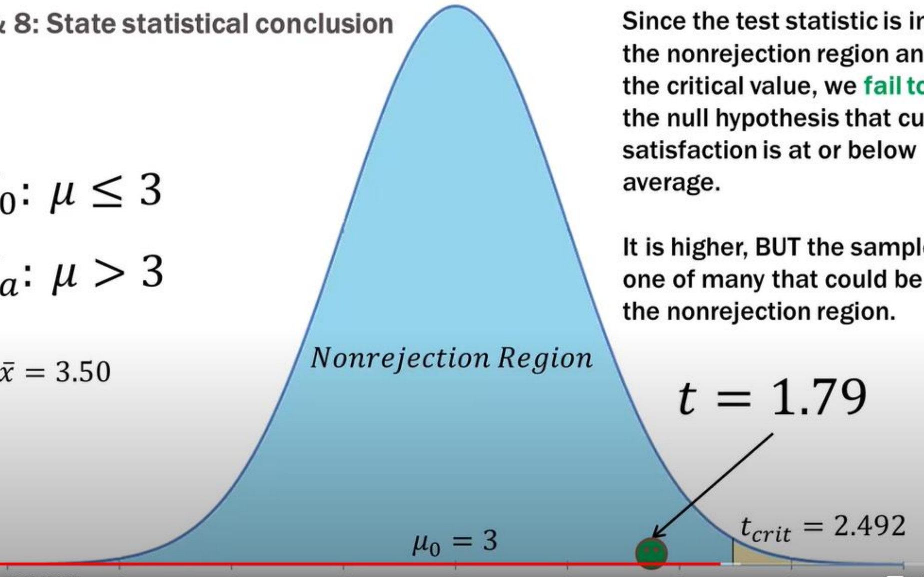
$$t = 1.79$$

Step 7 & 8: State statistical conclusion

H_0: $\mu \leq 3$

H_a: $\mu > 3$

$\bar{x} = 3.50$



Since the test statistic is inside the nonrejection region and inside the critical value, we **fail to reject** the null hypothesis that customer satisfaction is at or below average.

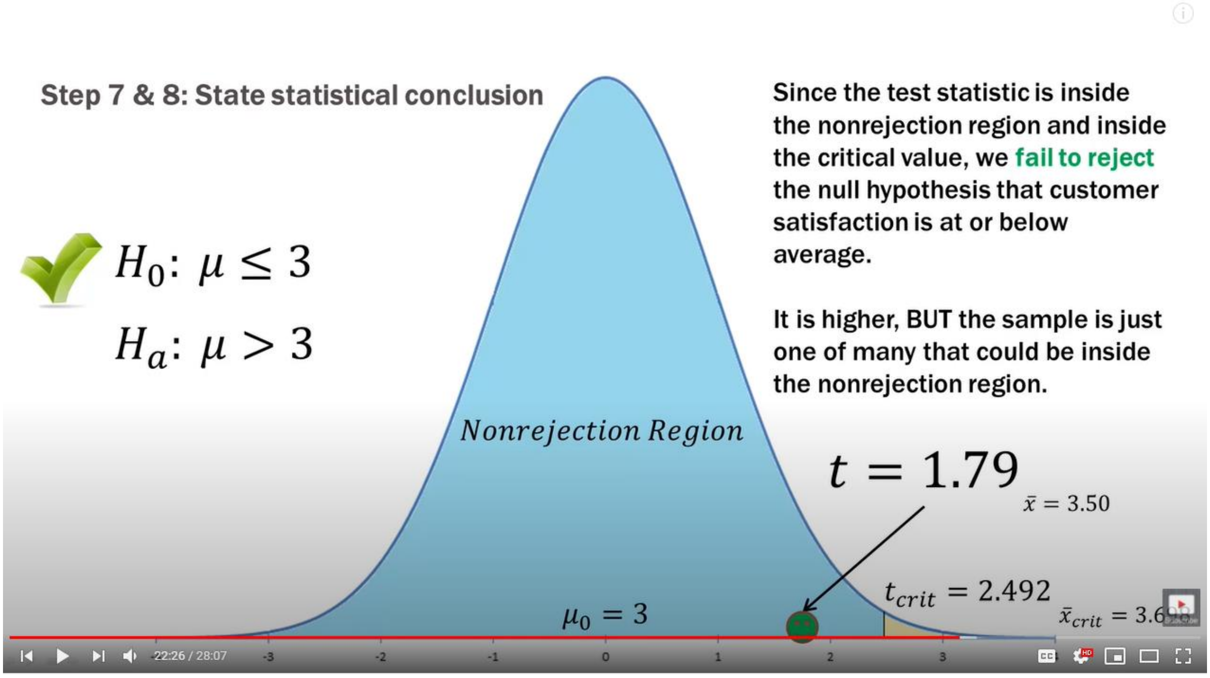
It is higher, BUT the sample is just one of many that could be inside the nonrejection region.

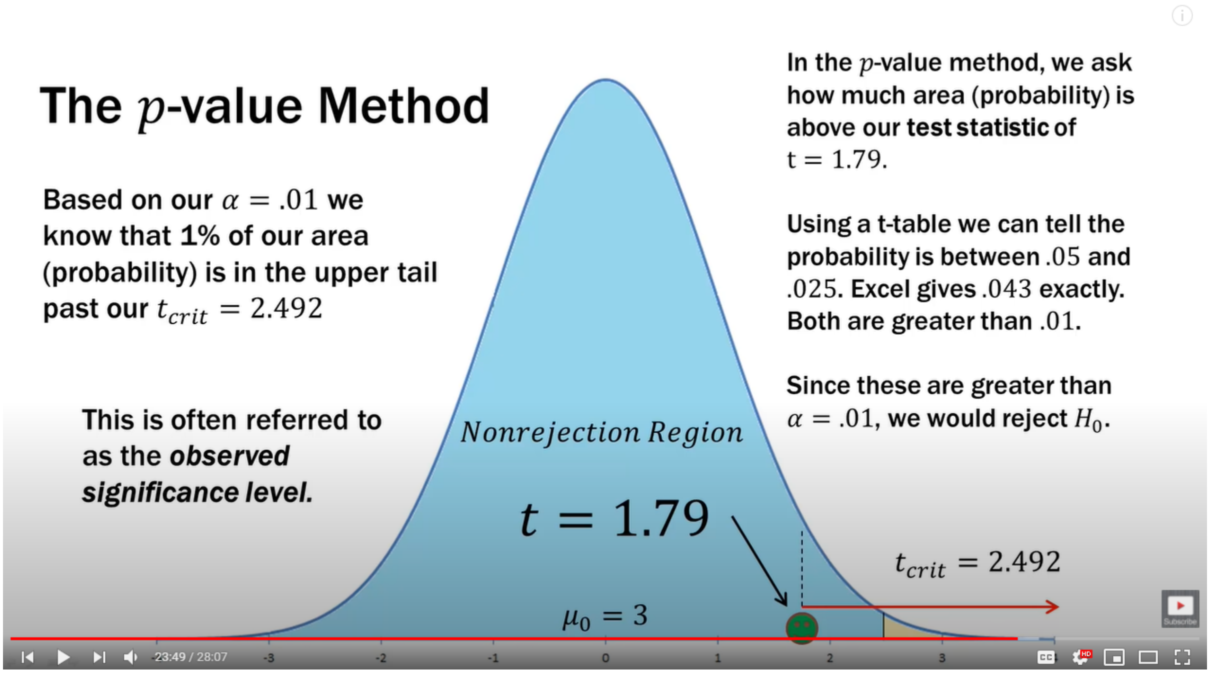
What is the mean customer satisfaction value at $t_{crit} = 2.492$?

$$t = \frac{3.50 - 3}{\frac{1.4}{\sqrt{25}}} \qquad 2.492 = \frac{\bar{x} - 3}{\frac{1.4}{\sqrt{25}}} \qquad 2.492 = \frac{\bar{x} - 3}{.28}$$

$$.698 = \bar{x} - 3 \qquad 3.698 = \bar{x}$$

Therefore any $n = 25$ sample with $\bar{x} > 3.698$ would lead to a rejection of H_0 assuming the same s (unlikely) and α.


Step 7 & 8: State statistical conclusion
H_0: $\mu \le 3$
H_a: $\mu > 3$
Since the test statistic is inside the nonrejection region and inside the critical value, we fail to reject the null hypothesis that customer satisfaction is at or below average.
It is higher, BUT the sample is just one of many that could be inside the nonrejection region.
Nonrejection Region
$t = 1.79$
$\bar{x} = 3.50$
$t_{crit} = 2.492$
$\mu_0 = 3$
$\bar{x}_{crit} = 3.6$
22:26 / 28:07


The p-value Method
Based on our $\alpha = .01$ we know that 1% of our area (probability) is in the upper tail past our $t_{crit} = 2.492$
This is often referred to as the *observed significance level.*
In the p-value method, we ask how much area (probability) is above our test statistic of $t = 1.79$.
Using a t-table we can tell the probability is between .05 and .025. Excel gives .043 exactly. Both are greater than .01.
Since these are greater than $\alpha = .01$, we would reject H_0.
Nonrejection Region
$t = 1.79$
$t_{crit} = 2.492$
$\mu_0 = 3$
23:49 / 28:07

HYPOTHESIS TESTING PROCEDURE

1. Start with a well-developed, clear research problem or question
2. Establish hypotheses, both null and alternative
3. Determine appropriate statistical test and sampling distribution
4. Choose the Type I error rate
5. State the decision rule
6. Gather sample data
7. Calculate test statistics
8. State statistical conclusion
9. Make decision or inference based on conclusion

Video 12

HYPOTHESIS TESTING:

CALCULATING TYPE II ERROR

STARBUCKS CUSTOMER SATISFACTION

Starbucks is interested in assessing customer satisfaction in the Canadian city of Toronto, Ontario. To conduct the study, Starbucks asked 25 customers in the city:

"Compared to other coffee houses in Toronto, would you say the customer service at Starbucks is much better than average (5), better than average (4), average (3), worse than average (2), or much worse than average (1)?" (Likert scale)

The mean rating was determined to be 3.50. Based on this sample, the standard deviation was found to be $\sigma = 1.5$.

Step 1: Establish Hypothesis

$$H_0: \mu \leq 3 \qquad H_a: \mu > 3$$

Step 2: Determine Appropriate Statistical Test and Sampling Distribution

This will be a one-tailed test. Starbucks is interested in a better than average rating.

Since σ is known we will use the z-distribution.

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Step 3: Specify the Type I error rate (significance level)

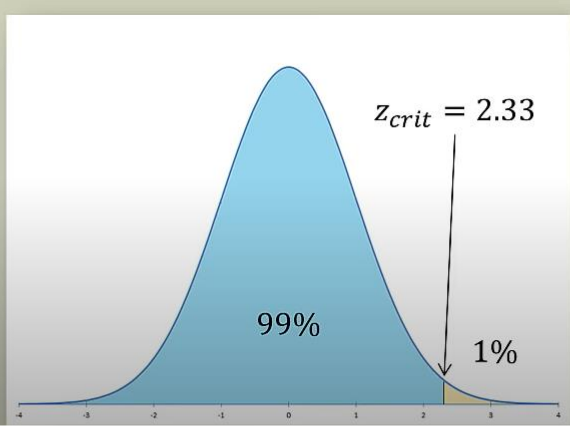
$$\alpha = .01$$

Step 4: State the decision rule

If $z > 2.33$**, reject** H_0

Step 5: Gather data

$$n = 25, \bar{x} = 3.5$$



What is the mean customer satisfaction value $\bar{x}$ at $z_{crit} = 2.33$?

$$z = \frac{3.50 - 3}{\frac{1.5}{\sqrt{25}}} \qquad 2.33 = \frac{\bar{x} - 3}{\frac{1.5}{\sqrt{25}}} \qquad 2.33 = \frac{\bar{x} - 3}{.3}$$

Instead of z being unknown, now $\bar{x}$ is unknown.

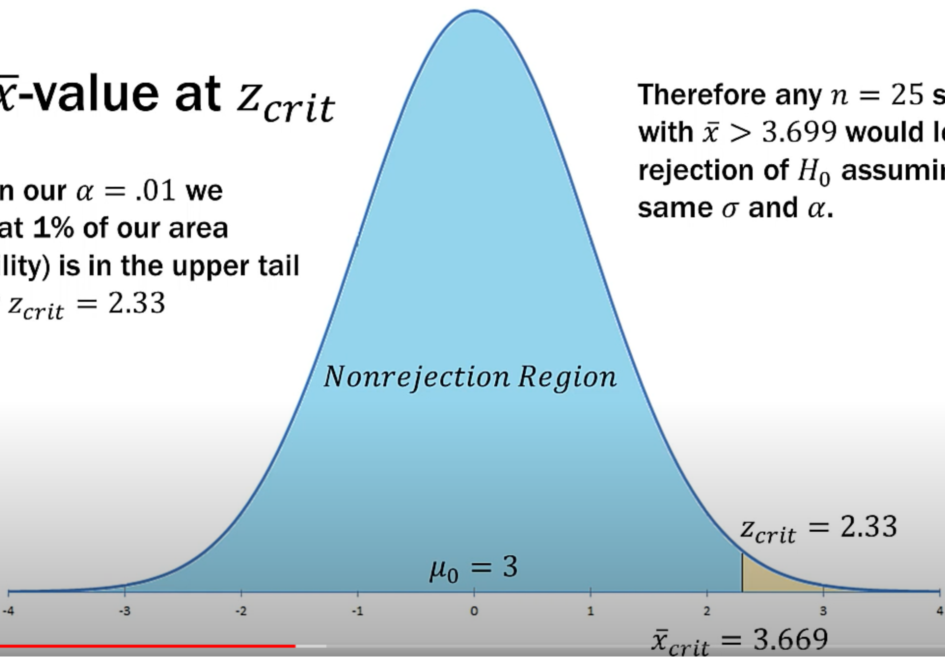
$$.699 = \bar{x} - 3 \qquad 3.699 = \bar{x}_{crit}$$

Therefore any $n = 25$ sample with $\bar{x} > 3.699$ would lead to a rejection of H_0 assuming the same σ and α.

The $\bar{x}$-value at z_{crit}

Based on our $\alpha = .01$ we know that 1% of our area (probability) is in the upper tail past our $z_{crit} = 2.33$

Therefore any $n = 25$ sample with $\bar{x} > 3.699$ would lead to a rejection of H_0 assuming the same σ and α.



CALCULATING TYPE II ERRORS

- To calculate the probability of Type II error, we have to select a μ_a that satisfies the alternative hypothesis
  - There is no "one" Type II error. It is different for every μ_a that satisfies the alternative hypothesis

$$H_0\colon \mu \leq 3 \qquad\qquad H_a\colon \mu > 3$$

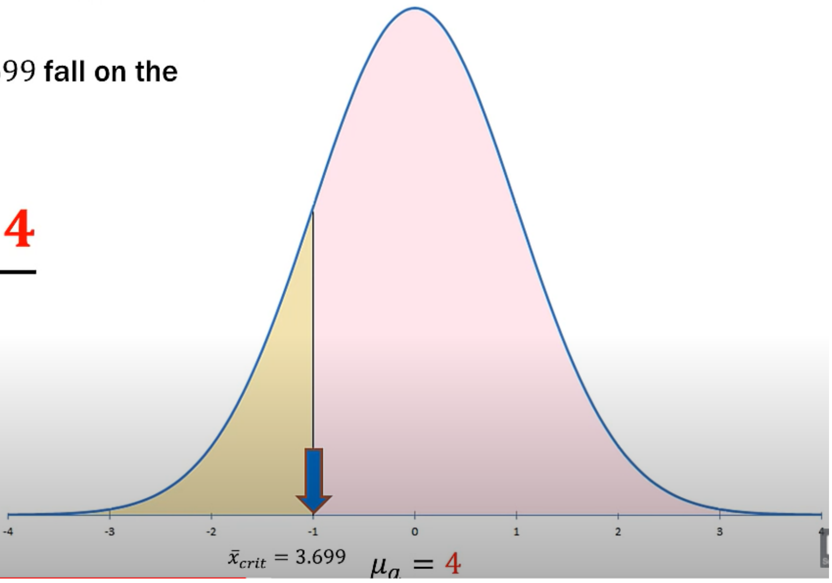
- For this example, let's choose $\mu_a = 4$
- We will begin by finding where the $\bar{x}_{crit}$ of 3.699 falls in the distribution where $\mu_a = 4$

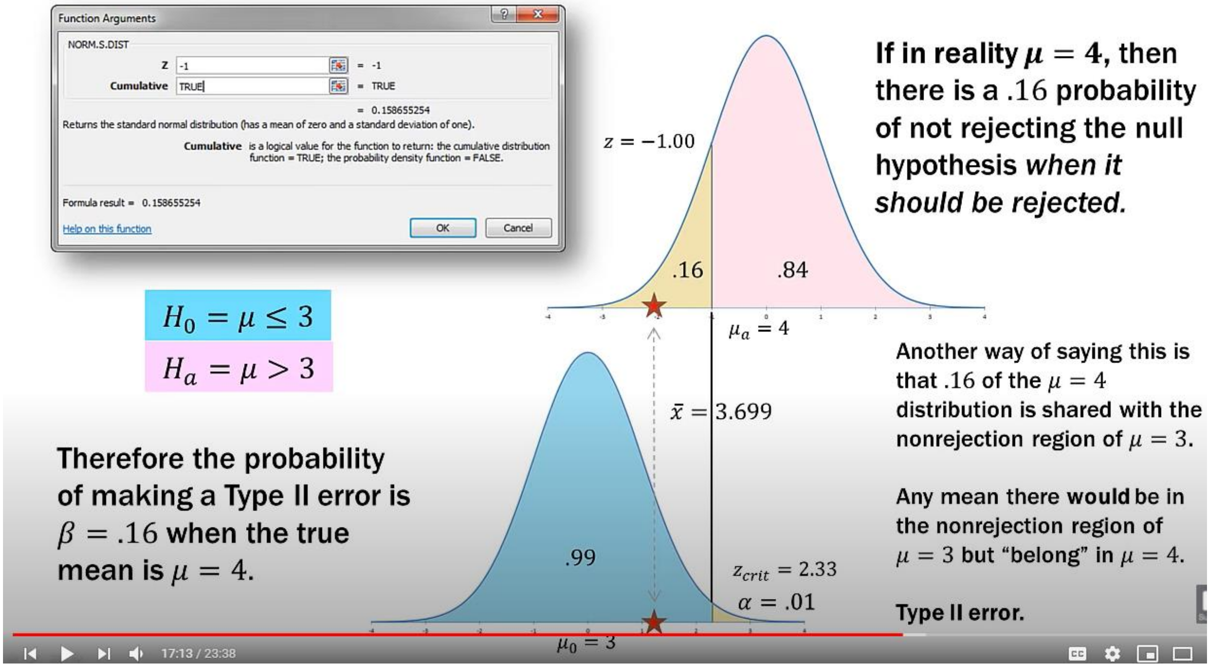
What if the true population mean is $\mu = 4$?

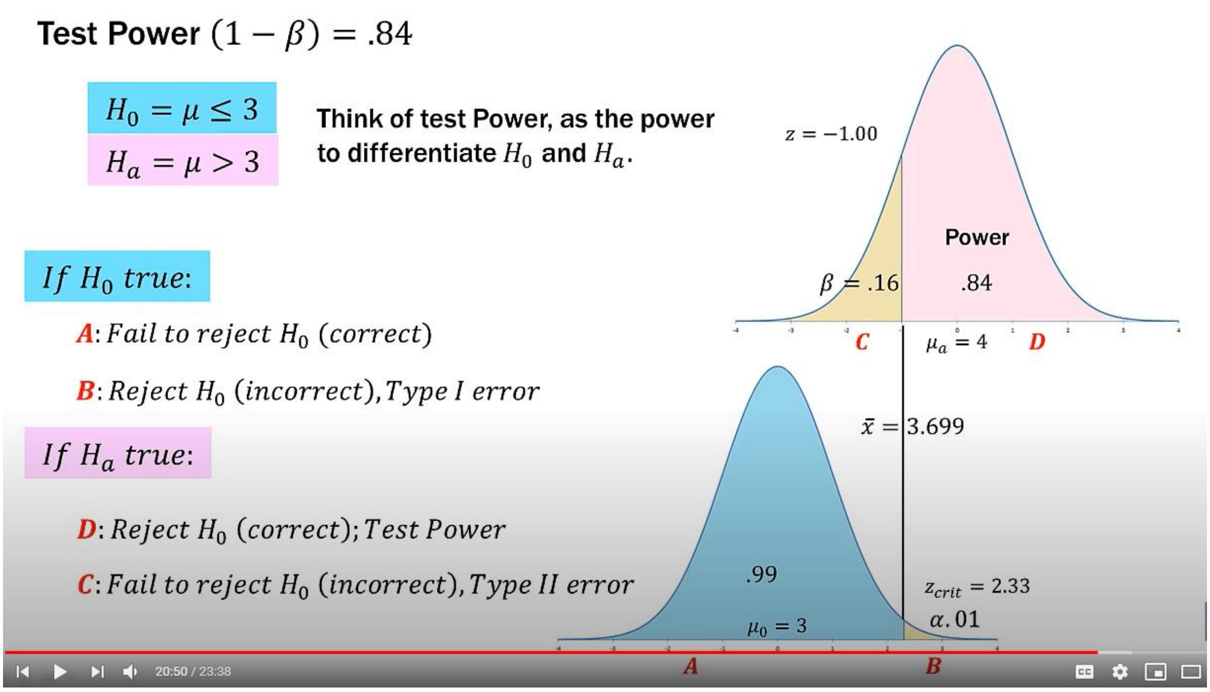
Where does $\bar{x}_{crit} = 3.699$ fall on the distribution if $\mu_a = 4$?

$$z = \frac{3.699 - \mathbf{4}}{\frac{1.5}{\sqrt{25}}}$$

$$z = -1.00$$




Function Arguments
NORM.S.DIST
Z -1 = -1
Cumulative TRUE = TRUE
= 0.158655254
Returns the standard normal distribution (has a mean of zero and a standard deviation of one).
Cumulative is a logical value for the function to return: the cumulative distribution function = TRUE; the probability density function = FALSE.
Formula result = 0.158655254
Help on this function
OK
Cancel
$H_0 = \mu \leq 3$
$H_a = \mu > 3$
Therefore the probability of making a Type II error is $\beta = .16$ when the true mean is $\mu = 4$.
$z = -1.00$
.16
.84
$\mu_a = 4$
$\bar{x} = 3.699$
.99
$z_{crit} = 2.33$
$\alpha = .01$
$\mu_0 = 3$
If in reality $\mu = 4$, then there is a .16 probability of not rejecting the null hypothesis *when it should be rejected.*
Another way of saying this is that .16 of the $\mu = 4$ distribution is shared with the nonrejection region of $\mu = 3$.
Any mean there would be in the nonrejection region of $\mu = 3$ but "belong" in $\mu = 4$.
Type II error.
17:13 / 23:38

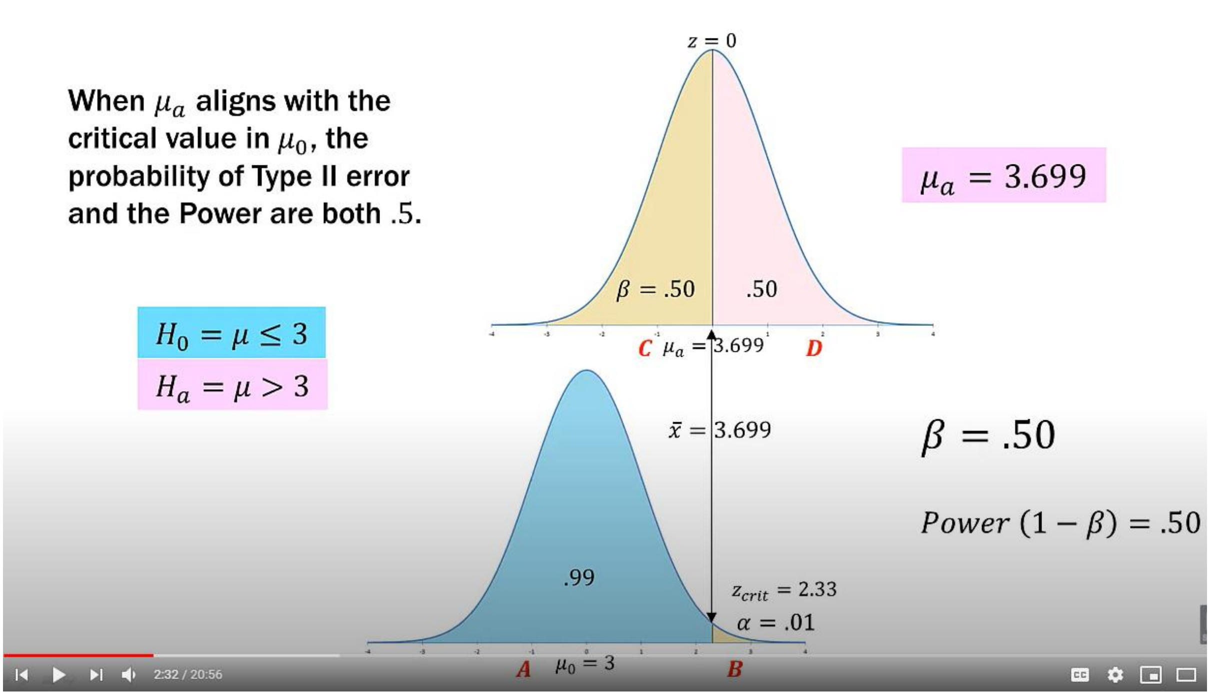

Test Power $(1 - \beta) = .84$
$H_0 = \mu \leq 3$
$H_a = \mu > 3$
Think of test Power, as the power to differentiate H_0 and H_a.
$If\ H_0\ true$:
A: $Fail\ to\ reject\ H_0\ (correct)$
B: $Reject\ H_0\ (incorrect), Type\ I\ error$
$If\ H_a\ true$:
D: $Reject\ H_0\ (correct); Test\ Power$
C: $Fail\ to\ reject\ H_0\ (incorrect), Type\ II\ error$
$z = -1.00$
Power
$\beta = .16$
.84
C
$\mu_a = 4$
D
$\bar{x} = 3.699$
.99
$z_{crit} = 2.33$
$\mu_0 = 3$
$\alpha .01$
A
B
20:50 / 23:38

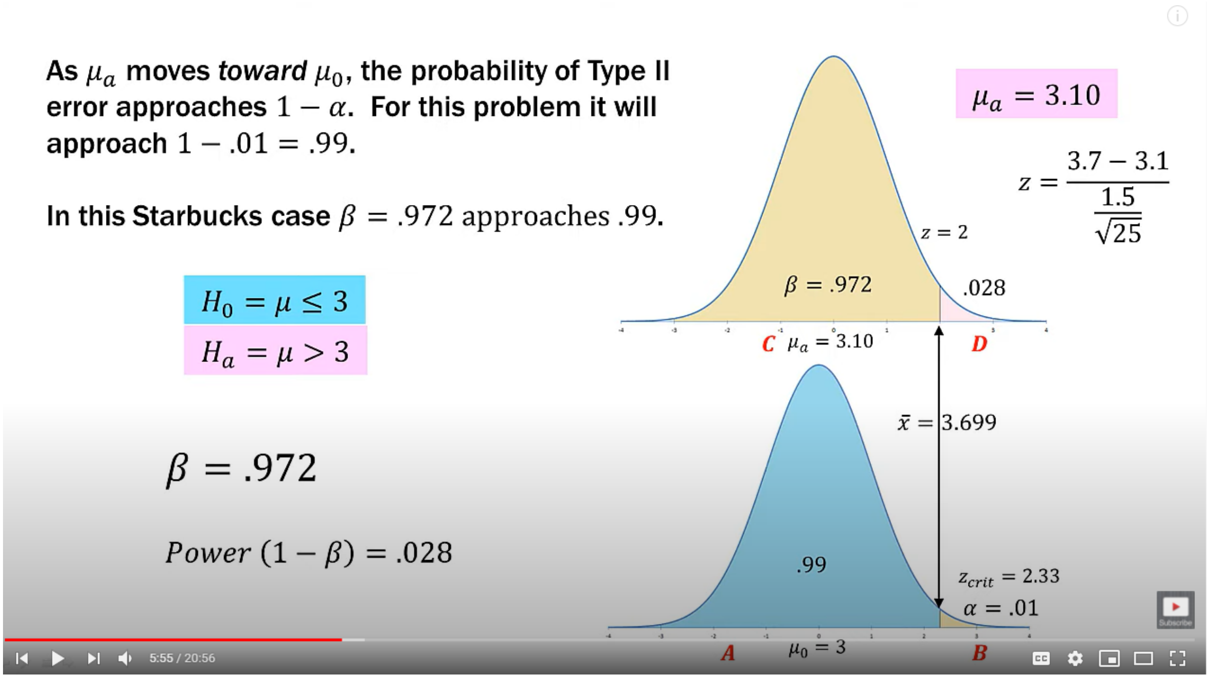
CALCULATING TYPE II ERRORS

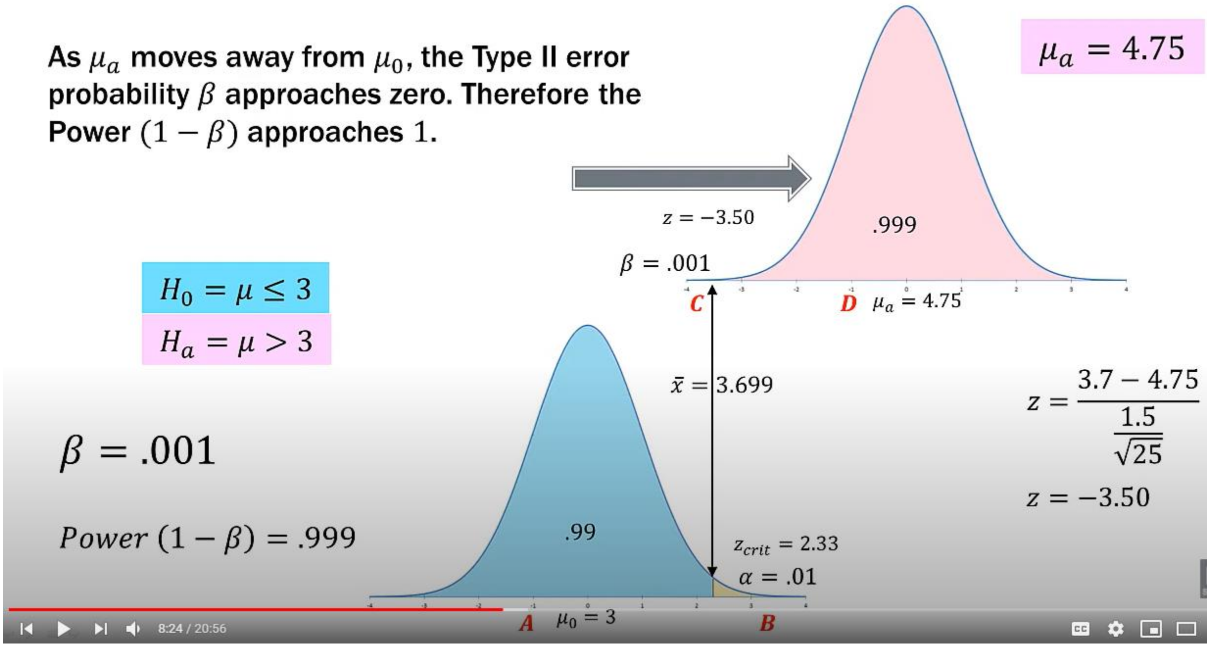
- There will be a unique Type II error for every value in the alternative hypothesis.
- The H_a value on which the Type II error is based is simply a choice. In this case we chose an H_a value of $\mu = 4$.
- As μ_a moves away from μ_0, the Type II error probability approaches zero. Therefore the Power $(1 - \beta)$ approaches 1.
  - Another way of saying this is that the test "gets better" at correctly rejecting H_0 as μ_a distances itself from μ_0.
- As μ_a moves *toward* μ_0, the probability of Type II error approaches $1 - \alpha$. In this Starbucks case, $.99$.
- When μ_a aligns exactly with the critical value in μ_0, the probability of Type II error and the Power are both $.5$.

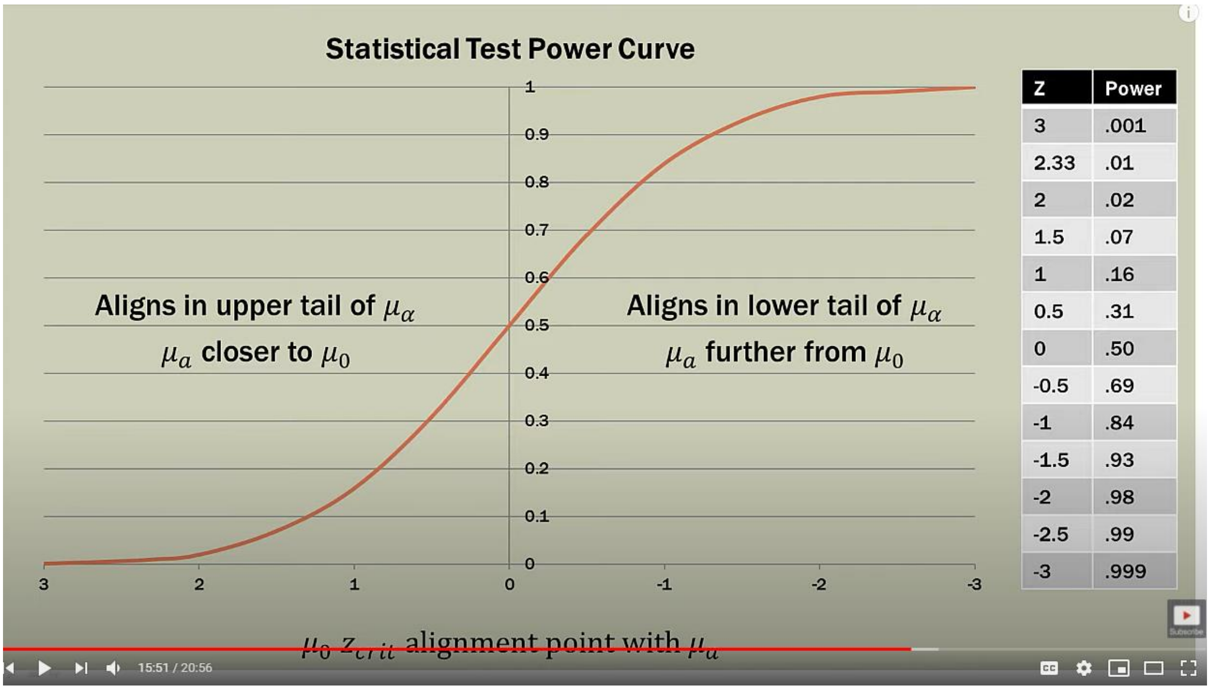
Video 13

Calculating Type 2 error, Test Power Curve









Z	Power
3	.001
2.33	.01
2	.02
1.5	.07
1	.16
0.5	.31
0	.50
-0.5	.69
-1	.84
-1.5	.93
-2	.98
-2.5	.99
-3	.999

HYPOTHESIS TESTING PROCEDURE

1. Start with a well-developed, clear research problem or question
2. Establish hypotheses, both null and alternative
3. Determine appropriate statistical test and sampling distribution
4. Choose the Type I error rate
5. State the decision rule
6. Gather sample data
7. Calculate test statistics
8. State statistical conclusion
9. Make decision or inference based on conclusion

CALCULATING TYPE II ERRORS

- There will be a unique Type II error for every value in the alternative hypothesis.
- The H_a value on which the Type II error is based is simply a choice. In this case we chose an H_a value of $\mu = 4$.
- As μ_a moves away from μ_0, the Type II error probability approaches zero. Therefore the Power $(1 - \beta)$ approaches 1.
  - Another way of saying this is that the test "gets better" at correctly rejecting H_0 as μ_a distances itself from μ_0.
- As μ_a moves *toward* μ_0, the probability of Type II error approaches $1 - \alpha$. In this Starbucks case, $.99$.
- When μ_a aligns with the critical value in μ_0, the probability of Type II error and the Power are both $.5$.

Video 14

HYPOTHESIS TESTING:

CONTROLLING TYPE II ERROR USING n

IMPORTANCE OF STANDARD ERROR

- The importance of understanding standard error can not be underestimated
- The standard error of the mean (the standard deviation of a sampling distribution) is foundational to almost everything in inferential statistics

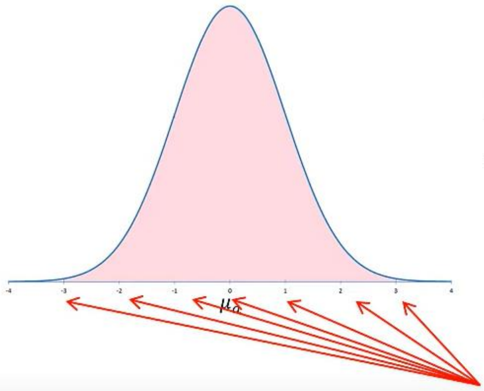
$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}} \qquad z = \frac{\bar{x} - \mu_0}{\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}}$$

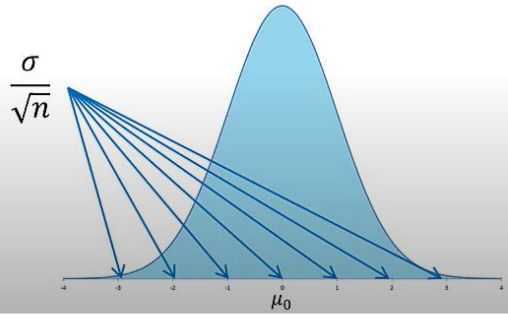
- By looking at the standard error component in these formulas, what happens to the overall standard error as n increases assuming σ remains constant?
- It DECREASES. Therefore it NARROWS the sampling distribution.

SETTING TYPE I AND TYPE II ERROR

- It is possible to establish ahead of time the acceptable amount of Type II error in the same fashion the Type I error is established
- To achieve this however we must locate the two distributions in the proper location / alignment
- Since the population means and variances are given in the problem, we will be forced to "manipulate" the sample size to achieve a certain standard error
- We will find the appropriate sample size to shape the variance in a manner that brings the distributions into alignment at the proper Type I and Type II error locations

By manipulating/changing the sample size n, we are changing the values (literally) at the standard error marks.





A larger sample size, given the same σ, will decrease the standard error making the interval between the marks smaller.

Starbucks is interested in assessing customer satisfaction in the Canadian city of Toronto, Ontario. To conduct the study, Starbucks asked 25 customers in the city:

“Compared to other coffee houses in Toronto, would you say the customer service at Starbucks is much better than average (5), better than average (4), average (3), worse than average (2), or much worse than average (1)?” (Likert scale)

The mean rating was determined to be 3.50. Based on this sample, the standard deviation was found to be $\sigma = 1.5$.

Step 1: Establish Hypothesis

$$H_0\colon \mu \leq 3 \qquad\qquad H_a\colon \mu > 3$$

Step 2: Determine Appropriate Statistical Test and Sampling Distribution

This will be a one-tailed test. Starbucks is interested in a better than average rating.

Since σ is known we will use the z-distribution.

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Step 3: Specify the Type I error rate (significance level)

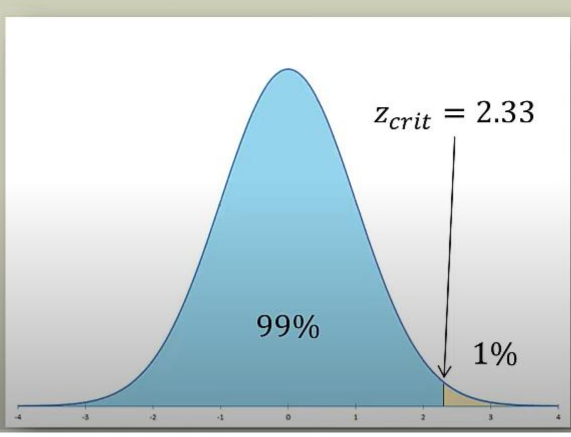
$$\alpha = .01$$

Step 4: State the decision rule

If $z > 2.33$, reject H_0

Step 5: Gather data

$$n = 25, \bar{x} = 3.5$$



What is the mean customer satisfaction value $\bar{x}$ at $z_{crit} = 2.33$?

$$z = \frac{3.50 - 3}{\frac{1.5}{\sqrt{25}}} \qquad 2.33 = \frac{\bar{x} - 3}{\frac{1.5}{\sqrt{25}}} \qquad 2.33 = \frac{\bar{x} - 3}{.3}$$

Instead of z being unknown, now $\bar{x}$ is unknown.

$$.699 = \bar{x} - 3 \qquad 3.699 = \bar{x}_{crit}$$

Therefore any $n = 25$ sample with $\bar{x} > 3.699$ would lead to a rejection of H_0 assuming the same σ and α.

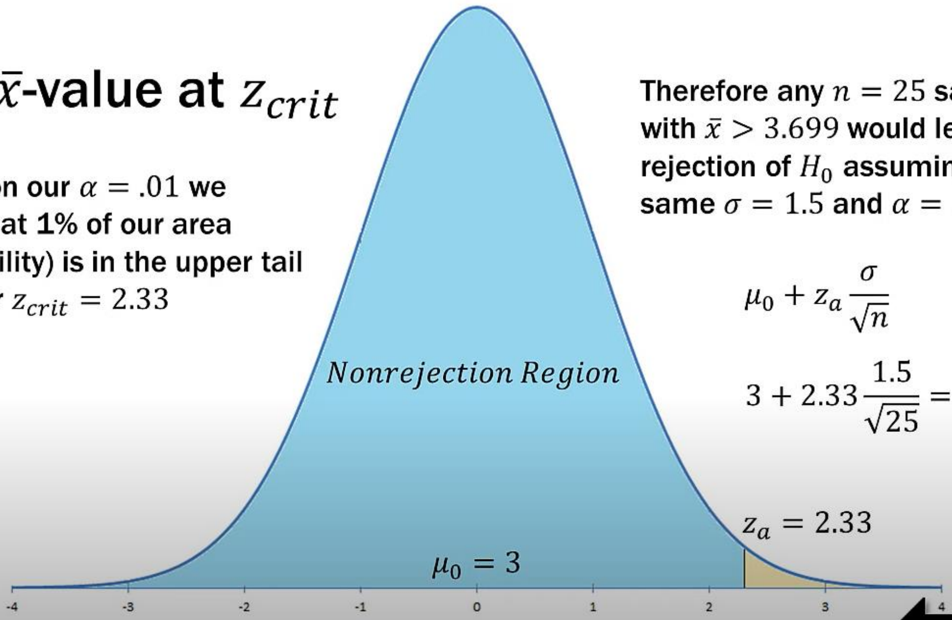
The $\bar{x}$-value at z_{crit}

Based on our $\alpha = .01$ we know that 1% of our area (probability) is in the upper tail past our $z_{crit} = 2.33$

Therefore any $n = 25$ sample with $\bar{x} > 3.699$ would lead to a rejection of H_0 assuming the same $\sigma = 1.5$ and $\alpha = .05$.

$$\mu_0 + z_a \frac{\sigma}{\sqrt{n}}$$

$$3 + 2.33 \frac{1.5}{\sqrt{25}} = 3.699$$



16:28 / 38:09

CONTROLLING TYPE II ERRORS

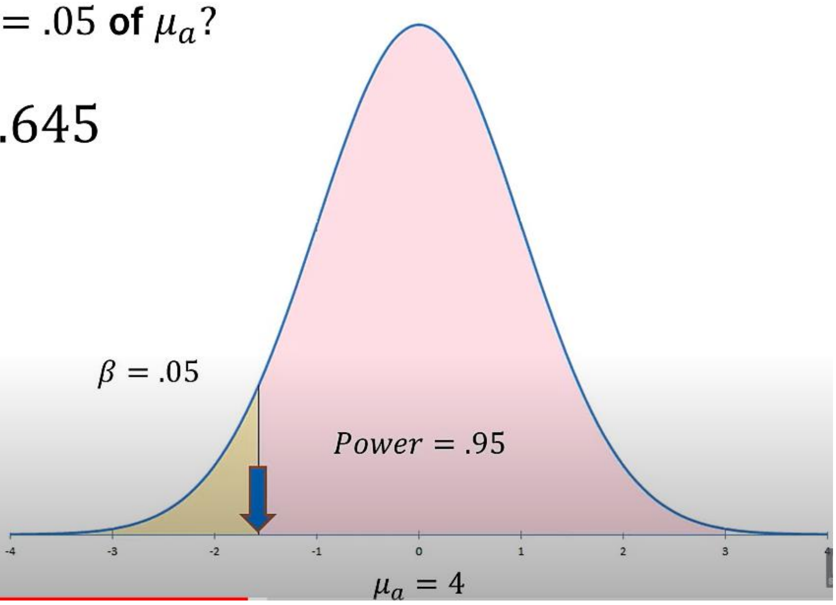
- To calculate the probability of Type II error, we have to select a μ_a that satisfies the alternative hypothesis
  - There is no "one" Type II error. It is different for every μ_a that satisfies the alternative hypothesis

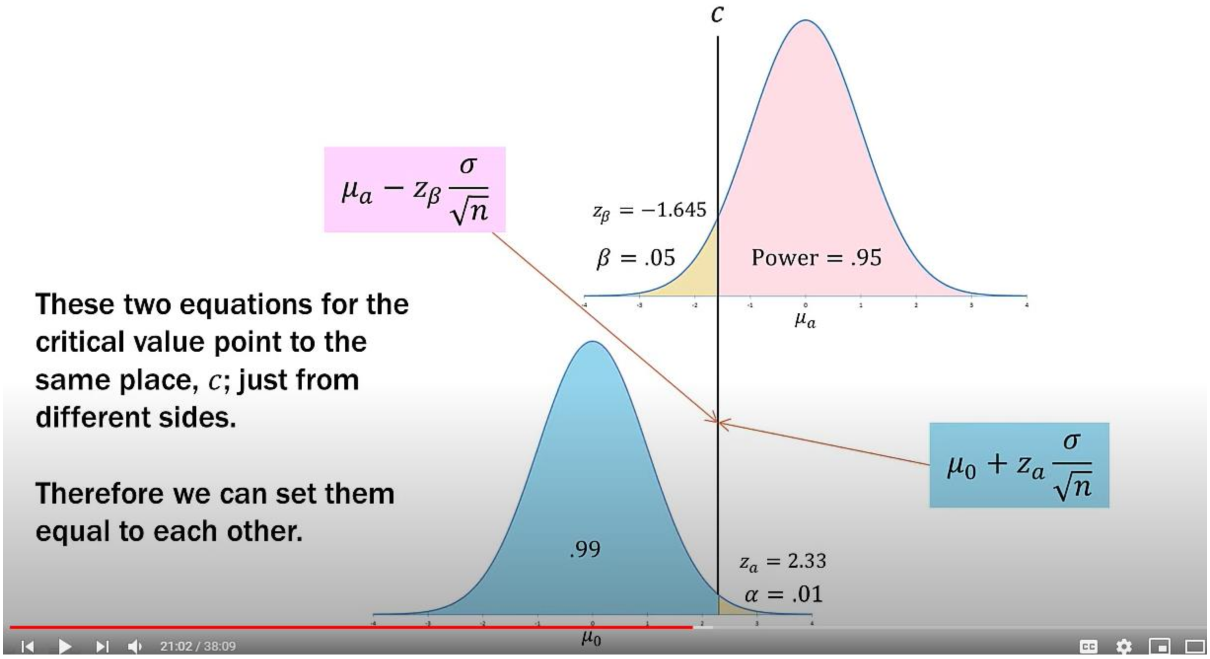
$$H_0: \mu \leq 3 \qquad\qquad H_a: \mu > 3$$

- For this example, let's choose $\mu_a = 4$
- Let's stipulate a Type II error rate of 5% or $\beta = .05$

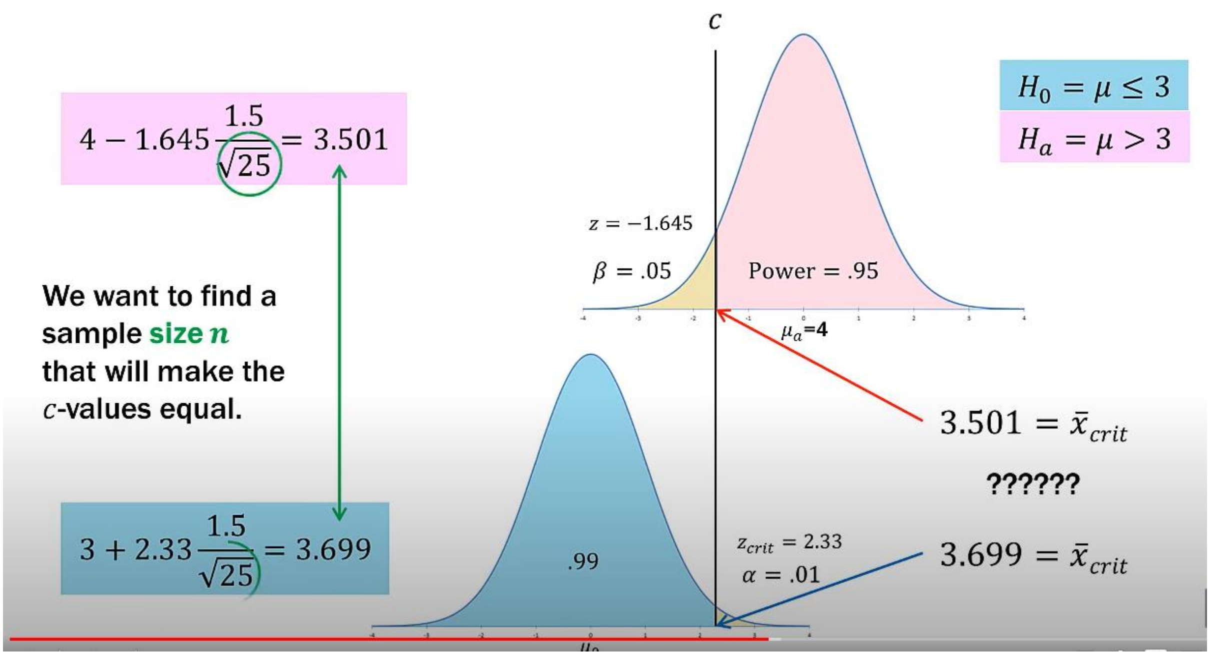
What region is $\beta = .05$ of μ_a?

$$z_\beta = -1.645$$

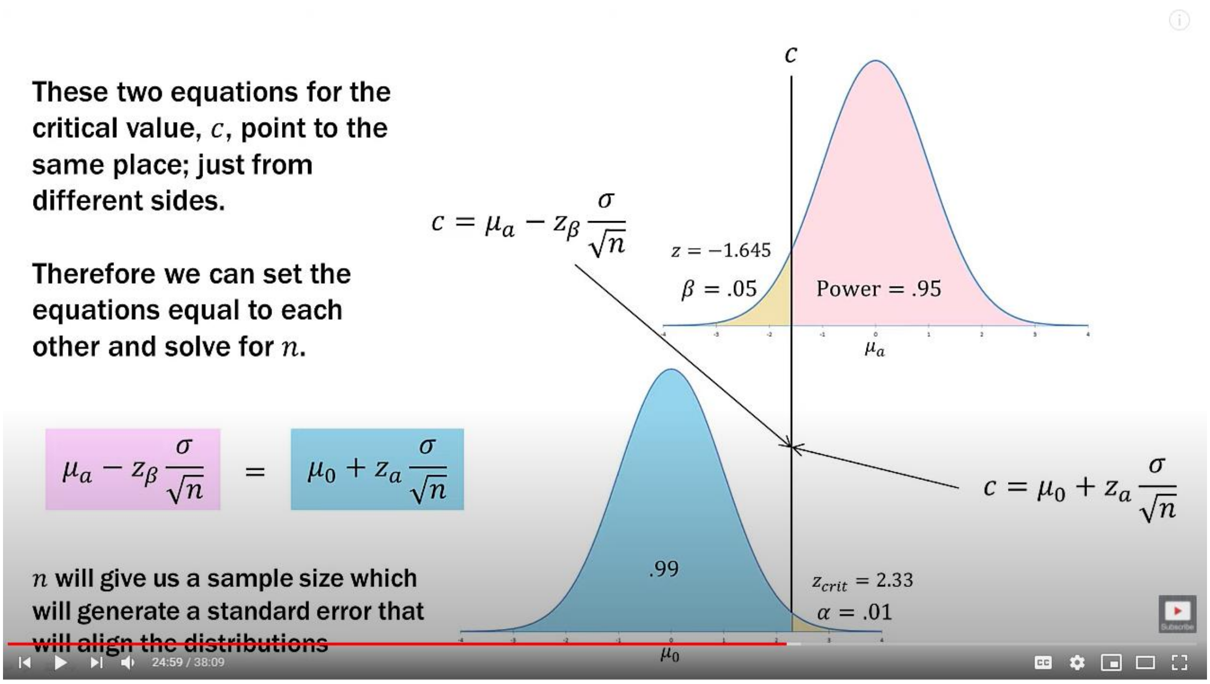




c
$\mu_a - z_\beta \frac{\sigma}{\sqrt{n}}$
$z_\beta = -1.645$
$\beta = .05$
Power = .95
μ_a
These two equations for the critical value point to the same place, c; just from different sides.
Therefore we can set them equal to each other.
$\mu_0 + z_a \frac{\sigma}{\sqrt{n}}$
.99
$z_a = 2.33$
$\alpha = .01$
μ_0
21:02 / 38:09



c
$H_0 = \mu \leq 3$
$H_a = \mu > 3$
$4 - 1.645 \frac{1.5}{\sqrt{25}} = 3.501$
$z = -1.645$
$\beta = .05$
Power = .95
μ_a=4
We want to find a sample size n that will make the c-values equal.
$3.501 = \bar{x}_{crit}$
??????
$3 + 2.33 \frac{1.5}{\sqrt{25}} = 3.699$
.99
$z_{crit} = 2.33$
$\alpha = .01$
$3.699 = \bar{x}_{crit}$



SOLVING FOR n TO CONTROL β

$$\mu_0 + z_a \frac{\sigma}{\sqrt{n}} = \mu_a - z_\beta \frac{\sigma}{\sqrt{n}}$$

$$\mu_0 - \mu_a = -z_a \frac{\sigma}{\sqrt{n}} - z_\beta \frac{\sigma}{\sqrt{n}}$$

$$\mu_0 - \mu_a = z_a \frac{\sigma}{\sqrt{n}} + z_\beta \frac{\sigma}{\sqrt{n}}$$

$$\mu_0 - \mu_a = \frac{(z_a + z_\beta)\sigma}{\sqrt{n}}$$

$$\sqrt{n} = \frac{(z_a + z_\beta)\sigma}{(\mu_0 - \mu_a)}$$

$$n = \frac{(z_a + z_\beta)^2 \sigma^2}{(\mu_0 - \mu_a)^2}$$

$$n = \frac{\left(z_a + z_\beta\right)^2 \sigma^2}{(\mu_0 - \mu_a)^2}$$

z_a = z value giving area of α in the upper tail of the standard z−distribution

z_β = z value giving area of β in the upper tail of the standard z−distribution

σ = the population standard deviation

μ_0 = the value of the population mean in the null hypothesis

μ_α = the value of the population mean used/chosen for the Type II error

Note: If the hypothesis is a two-tailed test, substitute $z_{a/2}$ for z_a

$$n = \frac{\left(z_a + z_\beta\right)^2 \sigma^2}{(\mu_0 - \mu_a)^2}$$

$z_a = 2.33$

$z_\beta = 1.645$

$\sigma = 1.5$

$\mu_0 = 3$

$\mu_\alpha = 4$

$$n = \frac{(2.33 + 1.645)^2 1.5^2}{(3-4)^2} = 35.6 \cong 36$$

Therefore to control Type II error at 5%, a sample size of 36 must be taken. This n will produce the correct standard error to align distributions at a common c. What c?

FINDING THE COMMON c

$$\mu_0 + z_a \frac{\sigma}{\sqrt{n}} = \mu_a - z_\beta \frac{\sigma}{\sqrt{n}}$$

$$3 + 2.33 \frac{1.5}{\sqrt{35.55}} = 4 - 1.645 \frac{1.5}{\sqrt{35.55}}$$

$$3 + 2.33(.25) = 4 - 1.645(.25)$$

$$3 + 0.586 = 4 - 0.411$$

$$3.59 \cong 3.59$$

Test Power $(1-\beta) = .95$

$H_0 = \mu \leq 3$

$H_a = \mu > 3$

The means did not move. The distributions narrowed due to a larger n decreasing the standard error.

$If\ H_0\ true$:

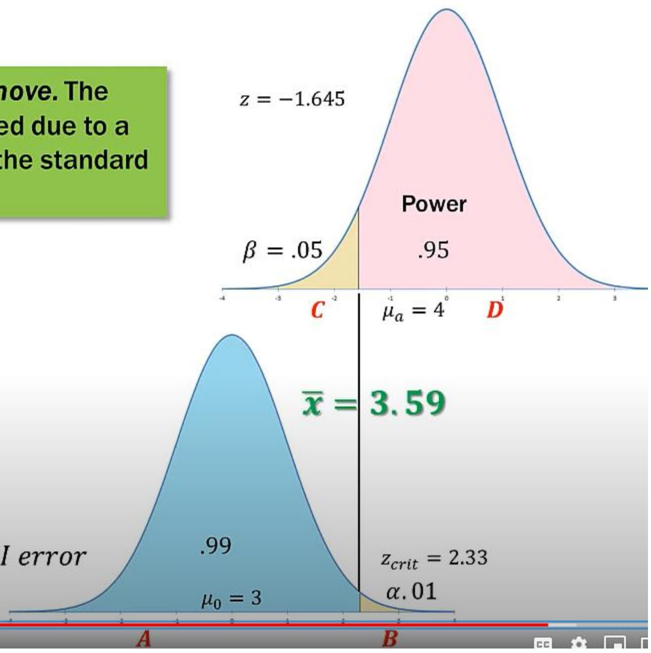
A: $Fail\ to\ reject\ H_0\ (correct)$

B: $Reject\ H_0\ (incorrect), Type\ I\ error$

$If\ H_a\ true$:

D: $Reject\ H_0\ (correct); Test\ Power$

C: $Fail\ to\ reject\ H_0\ (incorrect), Type\ II\ error$



CONTROLLING β REVIEW

- Controlling α (Type I error) is easy…we chose it
- Controlling β (Type II error) is a bit more complicated
- The goal is to align the α and β regions in the μ_0 and μ_a distributions respectively
- Since the means, σ, and critical values are set, we can "manipulate" the sample size, n, to generate a standard error that brings α and β into alignment at c
- This new n will create a new $\bar{x}_{crit}$ value for our decision rule
- Since Type II error is controlled, we CAN use the phrases "reject H_0" or "accept H_0"
- We would have to redo our study with an n of at least 36 to control Type II error at 5%