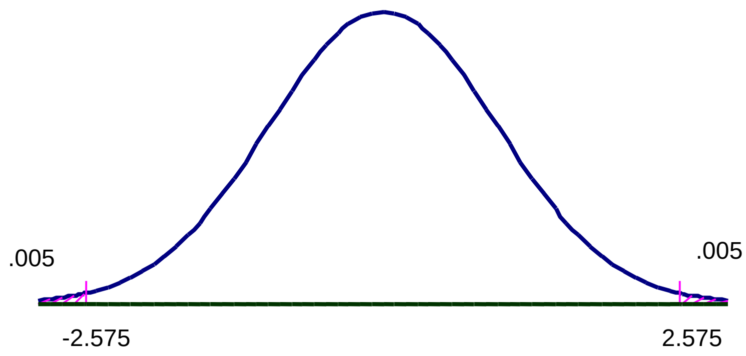
Hypotheses about μ
Two-tailed and One-tailed Z Tests
[One Sample Tests]

EXAMPLE:
A manufacturer produces bolts with a thickness of exactly 1inch (purportedly). A customer takes a random sample of 100 bolts and find that $\overline{X}$ =1.2 inches and s = .40 inches. Should the manufacturer's claim, that the bolts are exactly 1 inch (on average) be rejected? Test at α = 0.01.

[write the hypotheses]

H_0: μ = 1.00 inch
H_1: μ ≠ 1.00 inch

[choose test statistic; find critical values; draw region of rejection]



[use data to get calculated value of test statistic]

$$Z = \frac{1.20 - 1.00}{.40/\sqrt{100}} = \frac{.20}{.04} = 5$$

Therefore,

[conclusion: reject or not]

REJECT H_0

Now, Let’s try this as a 99%, Confidence Interval Estimator

1.20 ± 2.575 (0.04) = 1.20 ± .103

1.097 ←————→ 1.303 inches [NOTE: 1.00 is NOT in this interval!]

EXAMPLE:
A researcher claims that 10 year olds watch 6.6 hours of TV daily. You try to verify this with the following sample data.

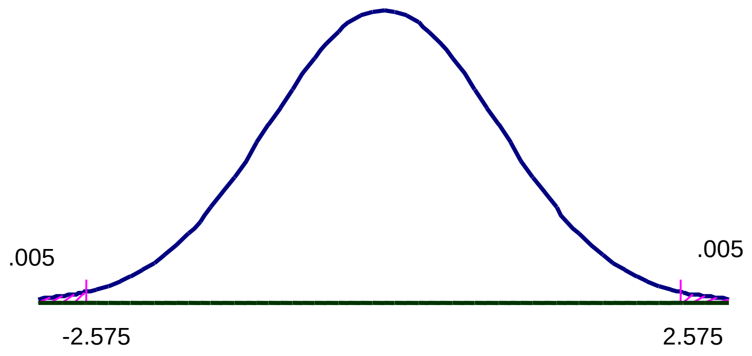
n=100

$\overline{X}$ =6.1 hours

S = 2.5 hours

Test at α= .01

H_0: $\mu = 6.6$
H_1: $\mu \neq 6.6$



$$Z = \frac{6.1 - 6.6}{\frac{2.5}{\sqrt{100}}} = \frac{-.50}{.25} = -2$$

DO NOT REJECT H_0

Therefore, we do not reject H_0 at $p < .01$

As a 99%, Confidence Interval Estimate:

$$6.1 \pm 2.575\frac{2.5}{\sqrt{100}} \quad = \quad 6.1 \pm 2.575(.25) = \; 6.1 \pm .644$$

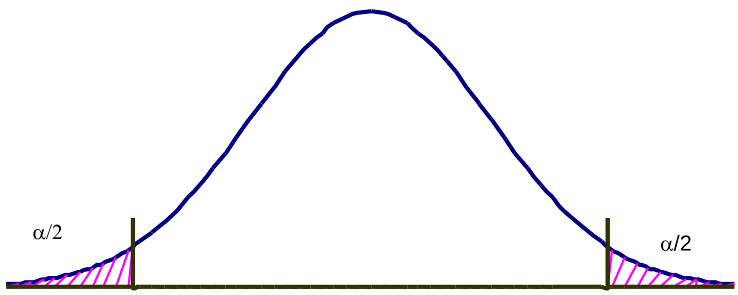
5.456 hours ←———→ 6.744 hours

Since μ_0 = 6.1 is covered by this CIE, we do not reject H_0

Of course, remember that when we are estimating a parameter using a confidence interval estimator, we do not expect to know anything about the parameter (i.e., there is no claim about the true mean).

One-Tail Tests

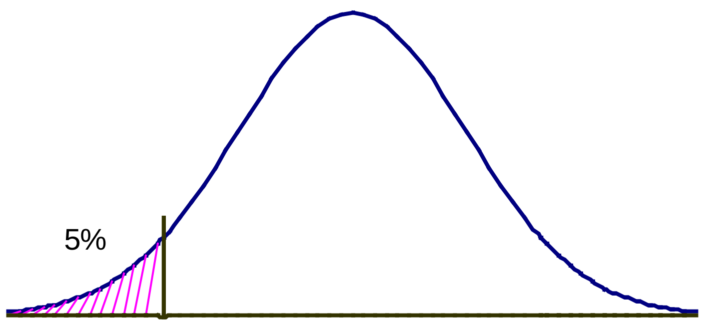
In previous examples, the tests of hypothesis were of the two-tailed variety. That is, the rejection region was divided equally into **both** tails of the sampling distribution of the statistic.



Sometimes we are interested in testing hypotheses that are to be rejected only if the sample shows significant deviation in one direction. In these cases deviations in the other direction only confirm the hypothesis.

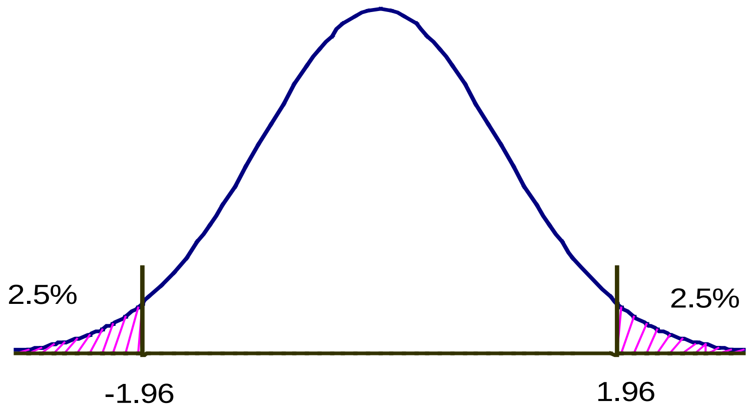
For example, in testing the life of ball bearings, the buyer wants $\mu \geq 300$ days and will only reject if $\mu < 300$ days.

H_0: $\mu \geq 300$ days
H_1: $\mu < 300$ days



[In this course, we will continue to do all confidence intervals as two-sided 2-tail CIEs.]

First - **More About Two-Tail Tests:**

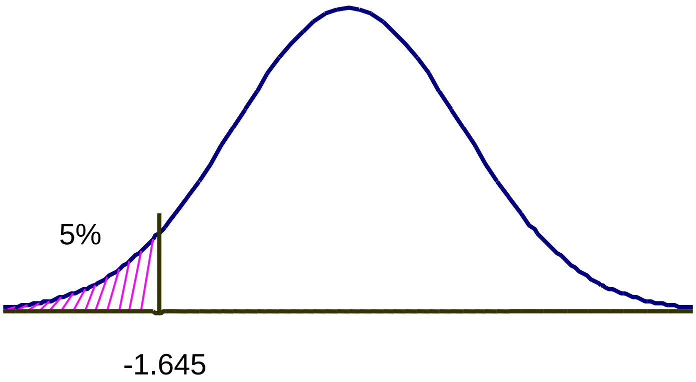


With a two tail test, the α error is split into two with $\alpha/2$ going into each tail. For instance, if you buy a watch, you want it to be accurate. Whether it is fast or slow, you have a problem with it; you want it to be exact. When there are problems with either too much or too little, you will want to do a two-tail test.

Here are some examples:

> Suppose a company makes a claim about the thickness of a bolt. The bolt must have a diameter of exactly of 10.00 centimeters. If it is thicker than 10.00 centimeters it will not fit into the hole in which it is supposed to be inserted. If it is not thick enough, then it will be loose and cause problems. We will reject whether the bolt is too thick or too thin.
>
> Suppose you manufacture coffee machines. The customer inserts two dollars and the coffee machine delivers exactly 12 ounces of premium coffee. The machine is supposed to deliver exactly 12.00 ounces of coffee. If it delivers more than 12.00 ounces, the owner of the machine will be upset since it will affect his profit margins. If it delivers less than 12.00 ounces, customers using the machine will be cheated.
>
> Suppose a single pill is supposed to contain 200 mgs. of a very powerful heart medication. Too much medication is a problem since it will kill the patient; too little and it will not work and the patient dies. The pill must have exactly 200 mgs. of the medication to work. Too much and too little are both problems.

One-Tail Tests:



When we do a one-tail test, the α error is put all into one tail of the probability distribution. This is done when you are only concerned with one side. For example, too much is a problem but too little is NOT a problem, or vice versa. For example, say you marry someone who tells you that she is worth one million dollars. It is doubtful that you will be upset if you find out she is actually worth ten million dollars.

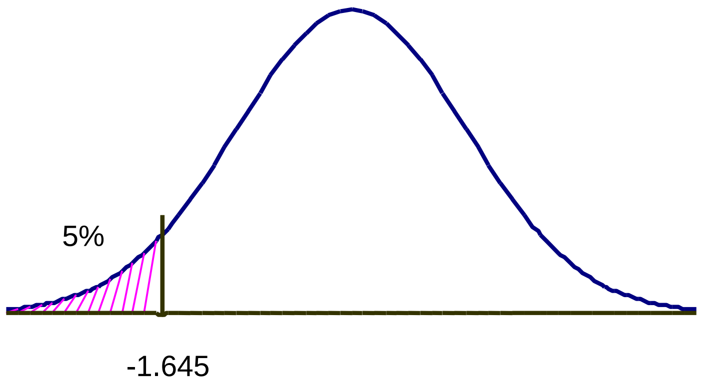
Here are some examples of one-tail tests:

> Suppose a company claims that its parts have a life of at least 10 years. A customer who buys a large number of these parts would like to test this claim. Of course, no customer will be upset if the part lasts for, say, 15 years. The problem is only in one direction. If the sample evidence indicates an average life of less than 10 years, we have to test to make sure that we are not looking at sampling error. The entire α error is on the left (the "less than") side.
>
> Suppose a peanut butter company claims that its peanut butter has no more than 100 parts per million (PPM) of impurities (I do not think you want to know what has been found in peanut butter). A consumer advocate sends a sample to an independent lab for testing. No consumer will be upset if the peanut butter has 25 PPM of impurities. The problem is only in one direction. If the sample evidence indicates impurities of more than 100 PPM, we have to test to make sure that we are not looking at sampling error. The α error is on the right side.

EXAMPLE

A company claims that a box of its raisin bran cereal contains at least 100 raisins. An inspector working for the Federal Trade Commission (FTC) takes a random sample of 324 boxes of cereals and finds that: $\overline{X} = 97$ raisins. and s = 9 raisins. Should the company's claim be rejected? Test at $\alpha = 0.05$.

H_0: $\mu \geq 100$
H_1: $\mu < 100$
[Notice that H_1 "points to" the region of rejection in the picture.]



$$Z = \frac{97 - 100}{\frac{9}{\sqrt{324}}} = \frac{-3}{.50} = -6$$

[$\frac{9}{\sqrt{324}}$ = .50 This is the standard error of the mean.]

The value of -6 is deep in the rejection region.

Therefore, reject H_0 at $P < .05$

If we took the above data and constructed a (two-sided) 95% confidence interval:

95%, Confidence Interval = 97 ± 1.96(.50)

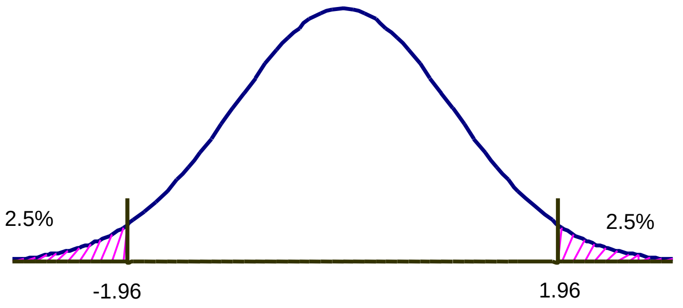
96.02 raisins ←————→ 97.98 raisins

In this case, the confidence interval uses the Z-values of ± 1.96 since it is two-sided and the alpha must be divided up into the two tails. The hypothesis test is a one-tail test so we use a critical value of -1.645. In this course, we will always construct two-sided confidence intervals.

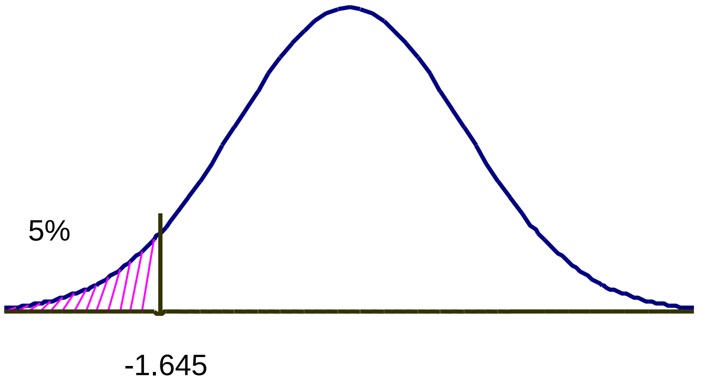
[It should be noted, however, there is such a thing as one-sided confidence intervals but we will not deal with them in this course.]

To help you understand the ramifications on the alpha level of performing a one-tail vs. two-tail test, we present the following normal curves. Note that with a two-tail test, alpha is split into two - .05 becomes .025 on either side. With a one-tail test, the entire alpha is all on one side.

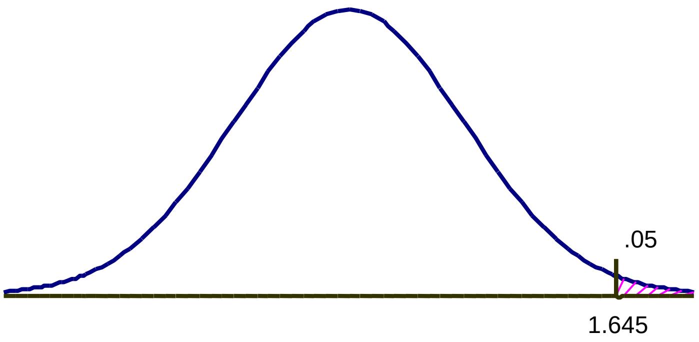
α =.05 and a two-tail test:



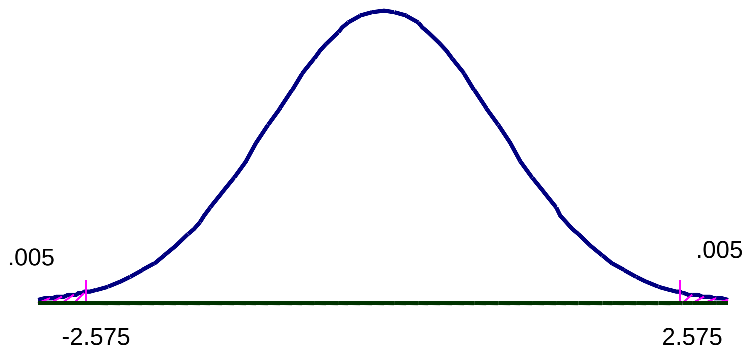
α =.05 and a one-tail test:



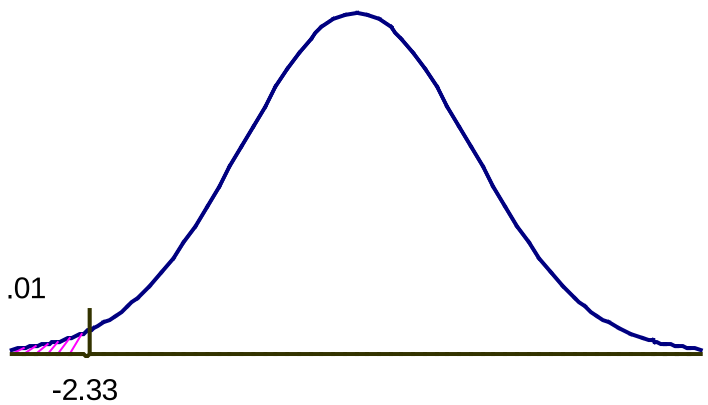
OR



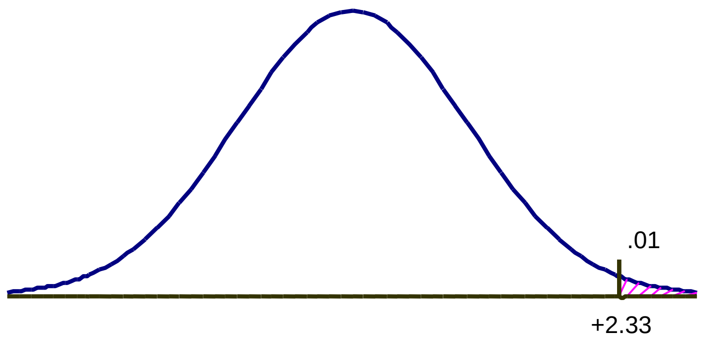
α =.01 and a two-tail test:



α =.01 and a one-tail test:



OR



DIGRESSION - SETTING UP THE NULL AND ALTERNATE HYPOTHESES

Question: Is it easier to prove something is false or that something is true?

Answer: It is much easier to prove something is false than something is true.

[Suppose someone states that all geese are white. To refute this statement, you only need to produce one black goose. To prove it is true, you will need to check the color of every goose in the world.]

This is why it is very important to know how to set up the null and alternate hypotheses.

Sometimes, getting H_0 and H_1 right will depend on who is doing the testing.

- The testing might be done by the company itself to *substantiate* its own claim. Substantiation of a claim may be necessary in order to demonstrate compliance with government standards, e.g., for advertising claims or claims of product purity.
- On the other hand, some testing might be done by a government agency (acting on behalf of consumers) or even by by a competitor. In this case, the testing is done to see whether the claim can be *refuted*.

It is interesting to see how the direction of the hypotheses changes depending on whether we are trying to substantiate or refute a claim.

SUBSTANTIATING A CLAIM

When it comes to substantiating a claim, a company wants to set up the hypotheses, H_0 and H_1, in such a way that there is strong evidence that its claim is accurate. The cost of being mistaken is very high. The company wants to demonstrate that its claim is true. It does not want the sample evidence to be inconclusive (i.e., having no evidence to reject the claim). This is why the company's claim belongs in the alternative hypothesis, H_1.

In addition to ensuring compliance with government standards against making false or misleading claims, this approach is very important in quality control. The cost of being wrong when it comes to quality is very high. A company will lose its reputation and even get sued. Imagine how angry consumers get when they find that a product they just purchased does not work as expected.

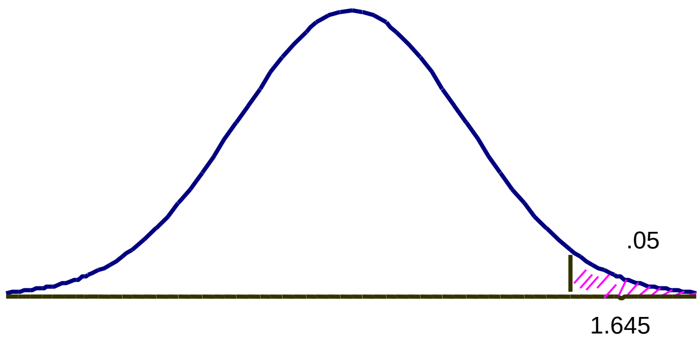
Example: Your company, Company A, produces cables for a particular task that requires them to have a breaking strength of at least 200 lbs. Since customers will not purchase cables with a lower breaking strength, your firm works hard to make sure that its cables meet customers' needs. Failure can result in mass returns of unused cables or, worse, litigation. Your firm uses the following statistical test to substantiate its claim (to customers and, perhaps, to government agencies):

Sample evidence:
n = 64
$\bar{X}$ = 220 lbs
s = 48 lbs

H_0: μ ≤ 200 lbs
H_1: μ > 200 lbs

Note that to substantiate the claim, the test is more stringent than what is typical. Instead of substantiating the claim that the true population breaking strength is at least 200 lbs, as in the narrative, we are actually trying to reject the hypothesis that the breaking strength is not greater than 200 lbs. The claim being substantiated by this test is that the cables have a breaking strength of greater than 200 lbs. The claim is in H_1. The company wants to reject the null hypothesis and thus show that its cables have a mean breaking strength of at least 200 pounds. This is also important if a government agency demands proof that the company has substantiated its claim.

With an alpha of .05, this test requires a critical value for Z of +1.645:



Computing the value of Z using the sample data above, we get:

$$Z = \frac{220 - 200}{48/\sqrt{64}} = \frac{20}{6} = 3.33$$

Conclusion: REJECT H_0. The company's claim is substantiated (confirmed).

REFUTING A CLAIM

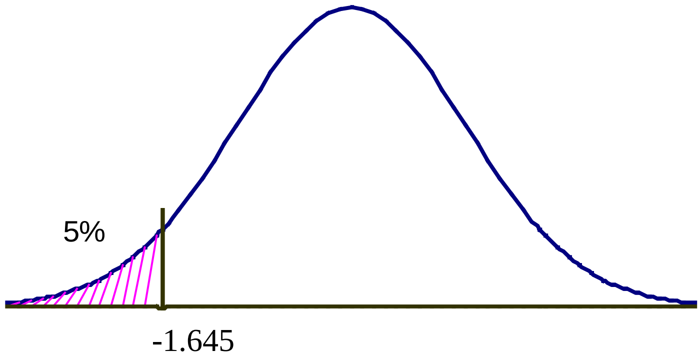
Example. Same as above, but this time the research is done by a competing firm, which wishes to refute company A's claim.

Company B is trying to use the sample evidence to refute Company A's claim that its cables have a mean breaking strength of at least 200 lbs. Thus the hypothesis test is set up so that the claim in contained in the null hypothesis. If the null hypothesis is rejected, then the alternative hypothesis is taken to be true and Company A's claim is refuted. For the competitor to win its case, it will need to produce a sample mean breaking strength that is significantly less than 200 pounds. In fact, any sample evidence that results in a mean that is 200 pounds or more automatically makes it impossible to reject H0. It is not even necessary to do any statistical test.

They know that the company is claiming that the population mean breaking strength is at least 200 pounds. The null and alternative hypotheses for this test would be:

H_0: $\mu \geq 200$ lbs
H_1: $\mu < 200$ lbs

With an alpha of .05, this test requires a critical value for Z of -1.645:



If we are using the same sample evidence as above, with $\bar{X}$ = 220 lbs, there is no reason to continue. There is no way H0 will be rejected unless the sample mean is whon to be significantly less than 200 lbs. Clearly the sample evidence supports the claim that $\mu \geq 200$ lbs.

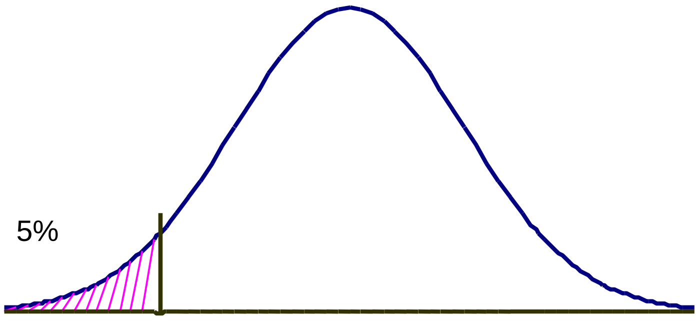
Now let us continue with more examples of two-tail and one-tail tests of hypothesis using the z statistic.

EXAMPLE:
A manufacturer purchases bulbs that are supposed to burn for a mean life of at least 3,000 hours, with a standard deviation of 500 hours. A sample of 100 bulbs is taken, with:

$\overline{X}$ = 2,800 hours

Test at $\alpha = .05$

H_0: $\mu \geq 3{,}000$ hours
H_1: $\mu < 3{,}000$ hours



$$Z = \frac{2800 - 3000}{500/\sqrt{100}} = \frac{-200}{50} = -4$$

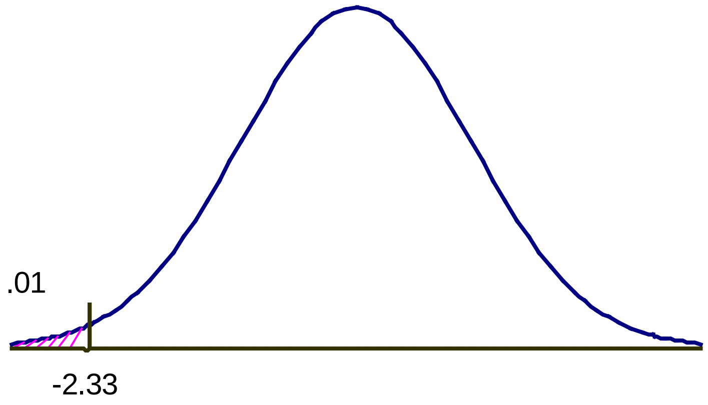
REJECT H_0

Therefore, we reject H_0 at $p < .05$

Using above data, we could construct a 2-sided 95% Confidence Interval Estimate: 2800 ± 1.96(50)
2702 hours ↔ 2898 hours

EXAMPLE:
A company claims that its weight-reducing drug will cause a weight loss of at least 10 pounds within one month. A random sample of 64 subjects is taken and the average weight loss is 7 lbs. with s = 4 lbs.
Test at $\alpha = .01$

H_0: $\mu \geq 10$
H_1: $\mu < 10$



$$Z = \frac{7-10}{4/\sqrt{64}} = \frac{-3}{.5} = -6$$

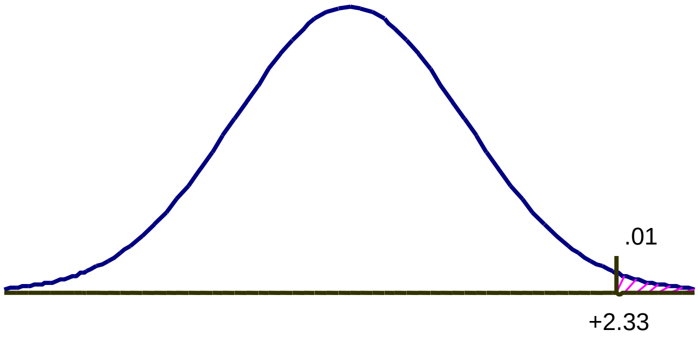
REJECT H_0

Therefore, reject H_0 at $p < .01$

EXAMPLE:
A wine manufacturer claims that his wine has at most 9 ppm (parts per million) impurities in a barrel of wine.
Test at $\alpha = .01$

n=64
$\bar{X}$ = 9.06 ppm impurities
s = 0.12 ppm

H_0: $\mu \leq 9$ ppm
H_1: $\mu > 9$ ppm



$$Z = \frac{9.06 - 9.00}{.12/\sqrt{64}} = \frac{.060}{.015} = 4$$

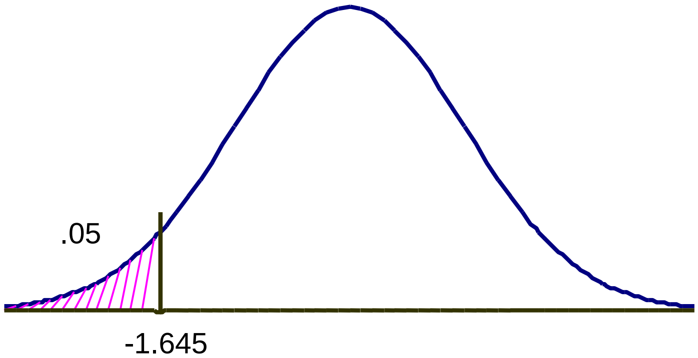
Therefore, reject H_0 at $p < .01$

EXAMPLE:
A company claims that its batteries have a life of at least 100 hours. You sample:

n=121
$\overline{X}$ = 97 hours
s = 3 hours

Test at α = .05 level

H_0: $\mu \geq 100$ hours
H_1: $\mu < 100$ hours



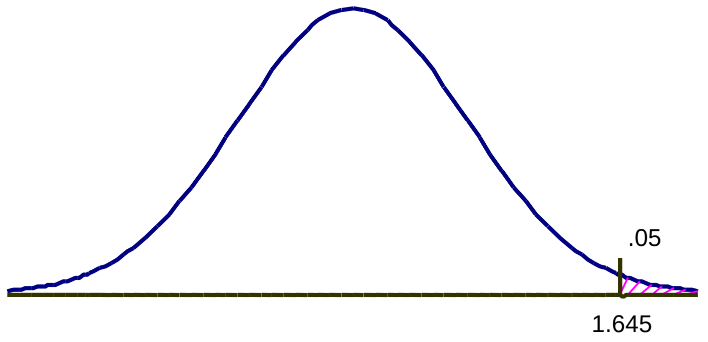
$$Z = \frac{97 - 100}{3/\sqrt{121}} = \frac{-3}{3/11} = -11$$

Therefore, reject H_0 at $p < .05$

EXAMPLE:
The Just-Like-Evian Pure Water Company claims that there are at most 1 ppm units of urine in its highly overpriced (er, regarded) water. Use the following sample data to test at $\alpha = .05$.

n=121 bottles
$\bar{X} = 1.1$ ppm
s = .33 ppm

H_0: $\mu \leq 1$ ppm
H_1: $\mu > 1$ ppm

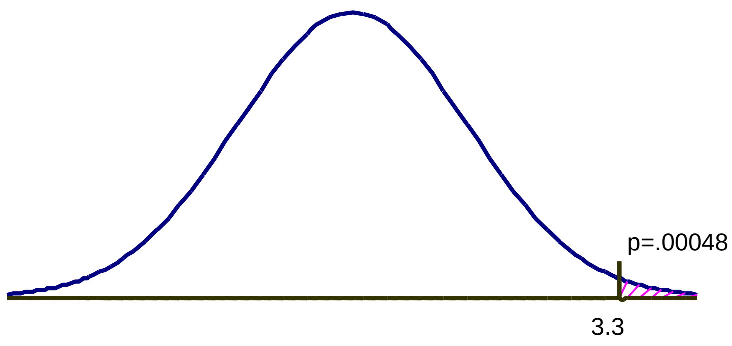


$$Z = \frac{1.1 - 1.0}{\frac{.33}{\sqrt{121}}} = \frac{.10}{.03} = 3.3$$

REJECT H_0 at $p < .05$

About p-value

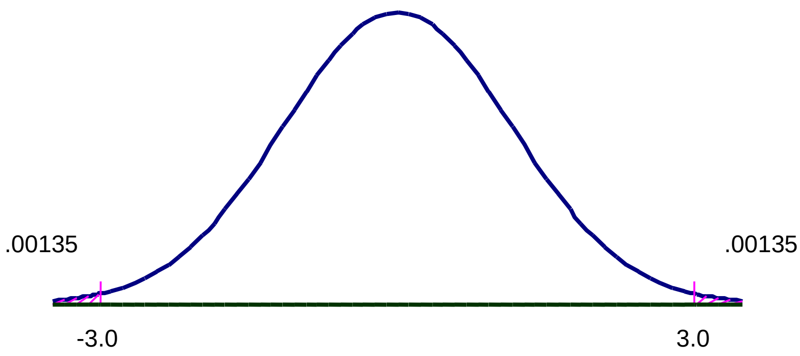
What is the “p-value”? It is the probability that we get the data value we got (the $\overline{X}$, as converted to a Z_{calc}), or ***worse***. We can get that probability from the Z table. In this case, What is the probability of getting a Z_{calc} value of 3.3 or higher (more extreme)?



p = .5 - .49952 = .00048

This is <.05 (the “nominal” α level).

p-value for a two-tailed test:
then we want the probability of getting the Z_{calc} from the data or worse (symmetrically, on both sides of the distribution). For example, suppose we got a Z_{calc} = -3.



$p = .00135 + .00135 = .00270$

that's $p < .05$ (if α is .05) and even $p < .01$ (if α is .01)

ADDITIONAL PROBLEMS

A toothpick manufacturer wants every box to contain exactly (on average) 500 toothpicks. Suppose you took a random sample of n = 81 boxes, and found:

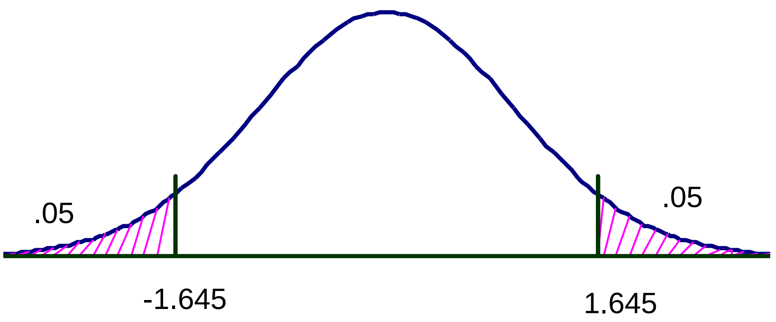
$\bar{X}$ =498 toothpicks

S = 9 toothpicks

Test at $\alpha = .10$

H_0: $\mu = 500$

H_1: $\mu \neq 500$



$$Z = \frac{498 - 500}{\frac{9}{\sqrt{81}}} = \frac{-2}{1} = -2$$

REJECT H_0

Therefore, we reject H_0 at $p < .1$

As a 90%, Confidence Interval:

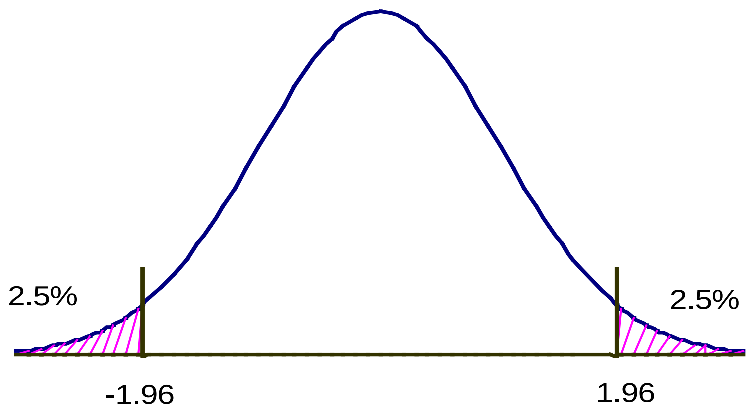
$$498 \pm 1.645\left[\frac{9}{\sqrt{81}}\right] = 498 \pm 1.645(1)$$

496.36 ←——————→ 499.65 toothpicks per box

[REJECT H_0 : 500 is NOT in the interval.]

EXAMPLE:
A company wishes to determine if the average salary of its clerks is really \$340. The company researcher takes a sample of 64 clerks and finds that $\overline{X}$ =\$300 and s = \$80. Test at $\alpha = 0.05$.

H_0: μ = \$340
H_1: $\mu \neq$ \$340



$$Z = \frac{300 - 340}{80/\sqrt{64}} = \frac{-40}{10} = -4$$ REJECT H_0

Therefore, reject H_0 at $p < .05$

As a 95%, CIE:

$$300 \pm \frac{80}{\sqrt{64}} = 300 \pm 1.96(10) = 300 \pm 19.6$$

\$280.40 ⟵⟶ \$319.60

[NOTE: 340 is not in this interval.]

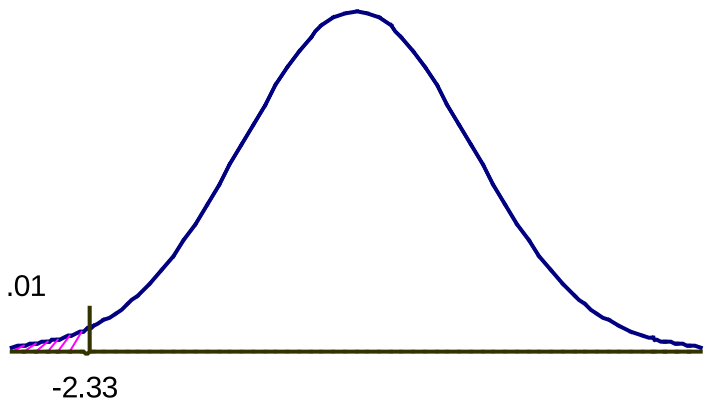
EXAMPLE:
The FTC wishes to determine whether 9 oz candy bars really are 9 oz. They take a sample of 49 candy bars:

$\bar{X}$ = 8.94 oz

S = .12 oz

Test at α = .01 level

H_0: $\mu \geq 9$ oz
H_1: $\mu < 9$ oz



$$Z = \frac{8.94 - 9.00}{.12/\sqrt{49}} = \frac{-.060}{.017} = -3.5$$

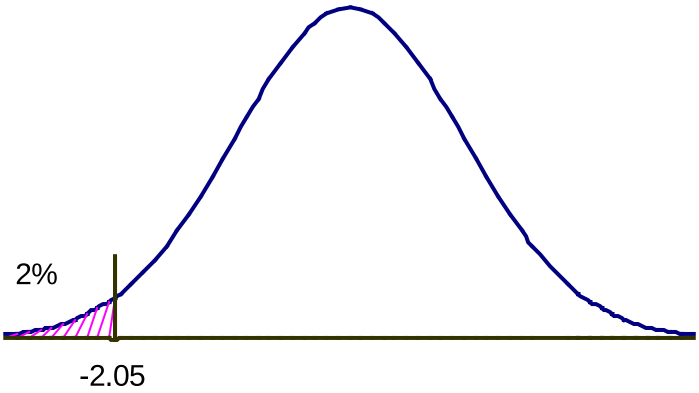
REJECT H_0

Therefore, reject H_0 $p < .05$

EXAMPLE:
A company claims that its toasters have an average life of at least 10 years. Use the following sample to test at $\alpha = .02$.

n=100
$\bar{X}$ = 9.4 years
s = 1.2 years

H_0: $\mu \geq 10$ years
H_1: $\mu < 10$ years



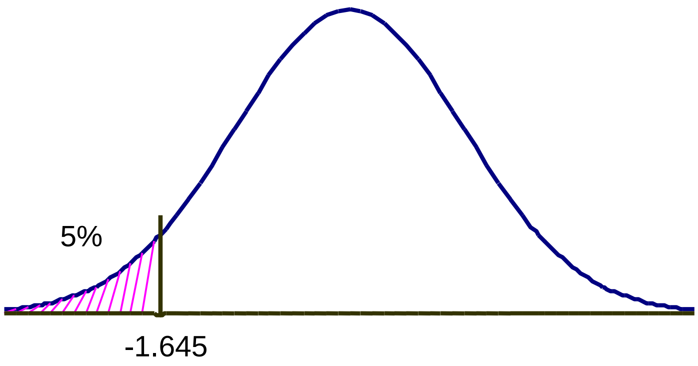
$$Z = \frac{9.4 - 10}{1.2/\sqrt{100}} = \frac{-.60}{.12} = -5$$

REJECT H_0

Therefore, reject H_0 at $p < .02$

EXAMPLE:
A manufacturer produces drill bits with an intended life of at least 580 hours and a standard deviation of 30 hours. A quality control scientist draws a sample of 100 bits and finds $\overline{X}$=577. Test at α=.05 to see if the machinery needs adjusting.

H_0: $\mu \geq 580$ hours
H_1: $\mu < 580$ hours



$$Z= \frac{577-580}{\frac{30}{\sqrt{100}}} = \frac{-3}{3} = -1.00$$

DO NOT REJECT H_0

P > .05

Do NOT adjust the machinery.