Unit 5
Hypothesis Testing

Hypothesis Testing

Confidence intervals represent the first of two kinds of inference we will study. The second common type of inference is called **hypothesis testing**.

A hypothesis test has a different goal than confidence intervals. It helps us to assess the evidence provided by some claim concerning a population.

Hypothesis Testing

In other words, we have some claim about the value of some population parameter, and we would like to determine whether there is evidence that supports this claim. We accomplish this by looking at sample data and seeing if they are representative of the claim. Unless we can examine the entire population, we can never **prove** that a parameter has a particular value, so we try and reach our conclusions with a high probability of being correct.

Hypothesis Testing

Some examples of claims that might be made about a population mean μ:

- An economist claims the average weekly grocery expenses for Manitoba families is greater than $250.
- A bank manager claims the average time customers have to wait in line is less than five minutes.
- A real estate agent claims the average value of a house in a certain neighbourhood differs from $400,000.

Example

A nutritionist claims that the average daily vitamin C intake of Canadians is less than the recommended amount of 75 mg. Suppose it is known that daily vitamin C intake for Canadians follows a normal distribution with a standard deviation of 20 mg.

Example

To test the nutritionist's claim, we record the daily vitamin C intake for a random sample of 25 Canadians, and we find a sample average of 73 mg. Is this strong evidence to support the nutritionist's claim that the true mean is less than 75 mg?

Example

We can't just say, 73 < 75, so "yes, the true mean vitamin C intake is less than 75 mg".

Maybe the true mean **is** 75 mg, and we just happened to get a sample with a mean slightly lower.

We must ask: If the true mean daily vitamin C intake **really was** 75 mg, how likely would it be to observe a sample average as extreme as 73 mg?

Note that "extreme" in this case means "low".

Example

If the probability is **low**, then we can conclude that the true mean vitamin C intake really is lower than 75 mg. If the probability is not sufficiently low, we have no conclusive evidence to support the nutritionist's claim.

We will see shortly why we need this probability to be **low** in order to support the nutritionist's claim.

Example

The first step is to assume that the true mean vitamin C intake really is equal to 75 mg; that is, we assume $\mu = 75$. Now **if this is true, what is the probability of observing a sample mean at least as low as 73 mg?**

We have the tools to find this probability!

Assuming the true mean is in fact 75, we know that the sample mean follows a normal distribution with mean $\mu = 75$ and standard deviation $\sigma/\sqrt{n} = 20/\sqrt{25} = 4$.

Example

The probability of observing a value of $\bar{X}$ at least as low as 73 when the true mean is 75 is:

$$P(\bar{X} \leq 73) = P\left(Z \leq \frac{73 - \mu_0}{\sigma/\sqrt{n}}\right) = P\left(Z \leq \frac{73 - 75}{20/\sqrt{25}}\right)$$

$$= P(Z \leq -0.50) = 0.3085$$

Example

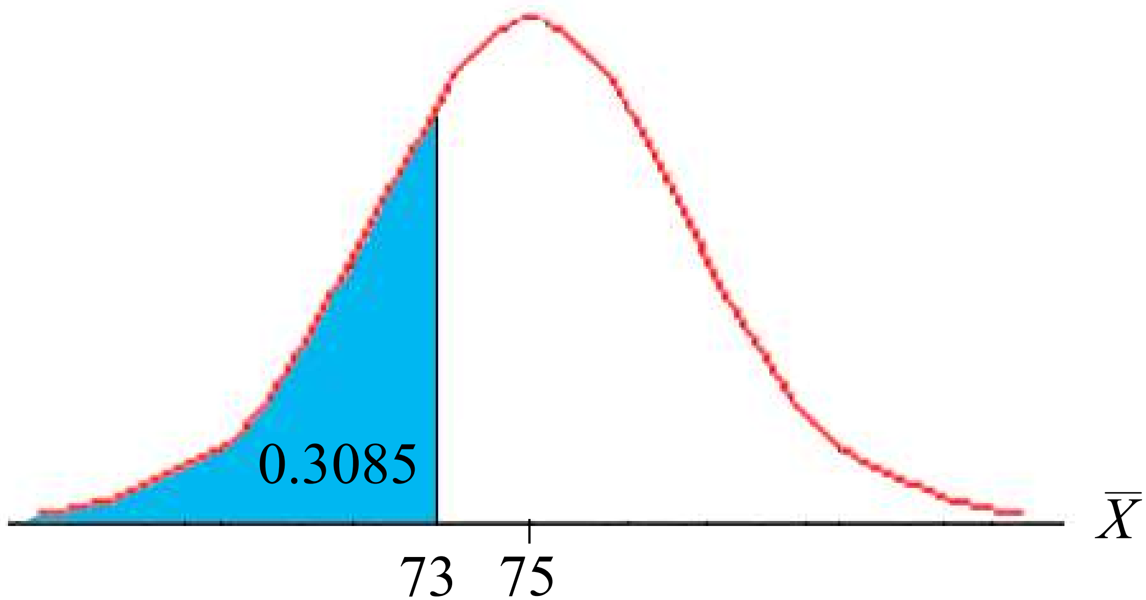
If the true mean daily vitamin C intake of all Canadians was 75 mg, the probability of observing a sample mean at least as low as 73 mg would be 0.3085.

Another way to say this:

If our original assumption was true, what we saw **would still be quite likely**. A sample mean as low as 73 is **not very surprising** if the true mean is 75.

Example

There is **insufficient evidence** to reject our initial assumption that μ = 75 and to conclude in favour of the nutritionist's claim. We would need a sample mean even lower than 73 mg in order to be convinced that the true mean was less than 75. In other words, we would need this probability to be **lower**.



Example

Note that we are **not** concluding the population mean **is** 75 and that the nutritionist's claim was wrong.

We got a sample mean of 73. This is certainly not evidence that the true mean is equal to 75.

It is just **not strong enough evidence** for us to be convinced with enough certainty that the true mean is less than 75.

Hypothesis Testing

The foundation of hypothesis testing is based on the question:

"If our initial assumption were true, how likely would it be to observe an estimate this extreme?"

In this case, our initial assumption was that the true mean is equal to 75 mg.

Example

The parent council at an elementary school appeals to the municipal government to install a red-light camera at a nearby intersection. The council claims that the average speed of motorists at the intersection is greater than the posted speed limit of 60 km/h.

Suppose the population standard deviation of speeds of vehicles at the intersection is known to be 15 km/h.

Example

A city worker is sent to measure the speeds of a random sample of 50 motorists at the intersection.

The average speed of these 50 vehicles is 66 km/h.

Is this enough evidence to conclude that the **true mean** speed μ of **all drivers** at the intersection is greater than $\mu_0 = 60$? That is, should a red-light camera be installed?

Example

We can't just say, 66 > 60, so "yes, the mean speed at the intersection is above the speed limit". We must ask: If the true mean speed of motorists at the intersection **really was** 60 km/h, how likely would it be to observe a sample average as extreme as 66 km/h?

Note that "extreme" in this case means "high".

Example

If the probability is **low**, then we can conclude that the mean speed really is higher than the limit. If the probability is not sufficiently low, we have no conclusive evidence to support the claim.

Example

The first step is to assume that the true mean speed really is equal to 60 km/h; that is, we assume $\mu = \mu_0 = 60$. Now **if this is true, what is the probability of observing a sample mean of at least 66 km/h?**

Assuming the true mean is in fact 60, we know that $\bar{X}$ follows a normal distribution with mean $\mu = 60$ and standard deviation $\sigma/\sqrt{n} = 15/\sqrt{50} = 2.12$.

Example

The probability of observing a value of $\bar{X}$ at least as high as 66 when the true mean is 60 is:

$$P(\bar{X} \geq 66) = P\left(Z \geq \frac{66-\mu_0}{\sigma/\sqrt{n}}\right) = P\left(Z \geq \frac{66-60}{15/\sqrt{50}}\right)$$

$$= P(Z \geq 2.83) = 1 - P(Z < 2.83) = 1 - 0.9977 = 0.0023$$

Example

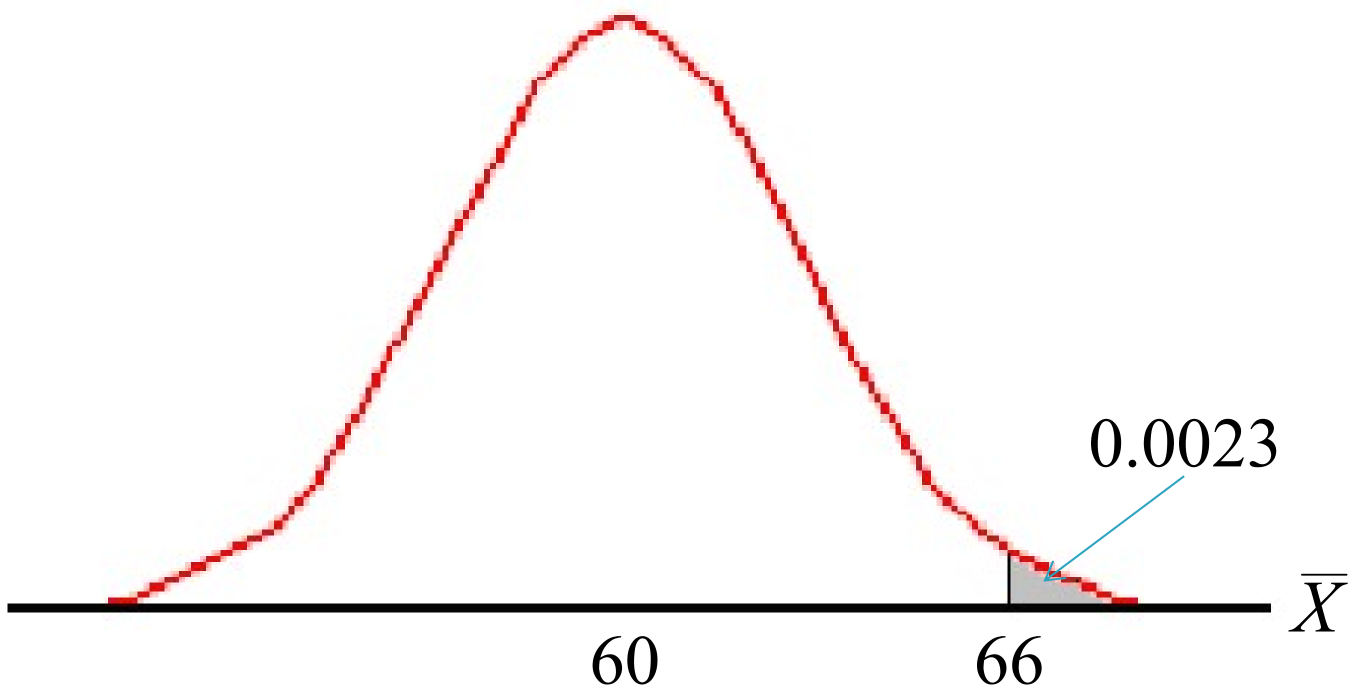
If the true mean speed of vehicles at this intersection was 60 km/h, then the probability of observing a sample mean speed at least as high as 66 would only be 0.0023.

Another way to say this:

If our original assumption was true, what we saw would be **almost impossible**. If the true mean was 60 km/h, we would **almost never** see a sample mean as high as 66 km/h.

Example

We would rarely see such a high sample mean if the true mean speed was 60 km/h.



Example

We conclude that, since this probability is so **low**, the true mean speed of vehicles at the intersection **really is above the posted limit**. It is possible, but very unlikely, that the true mean is as low as 60.

The probability of observing such a high sample mean speed given the assumption that $\mu = 60$ is small enough that we are willing to believe the council's claim. Based on these findings, the municipal government decides to install a camera at the intersection.

Example

Note that there is always a certain probability that we will be wrong in our conclusion. Maybe this was an exceptional sample of unusually fast vehicles. However, we are able to conclude in favour of the council's claim with a high level of certainty.

Note also that there was no mention of speeds following a normal distribution. (This is almost certainly not the case, as speeds tend to be skewed to the right.) But since the sample size is high, we know that the sampling distribution of $\bar{X}$ is approximately normal by the CLT.

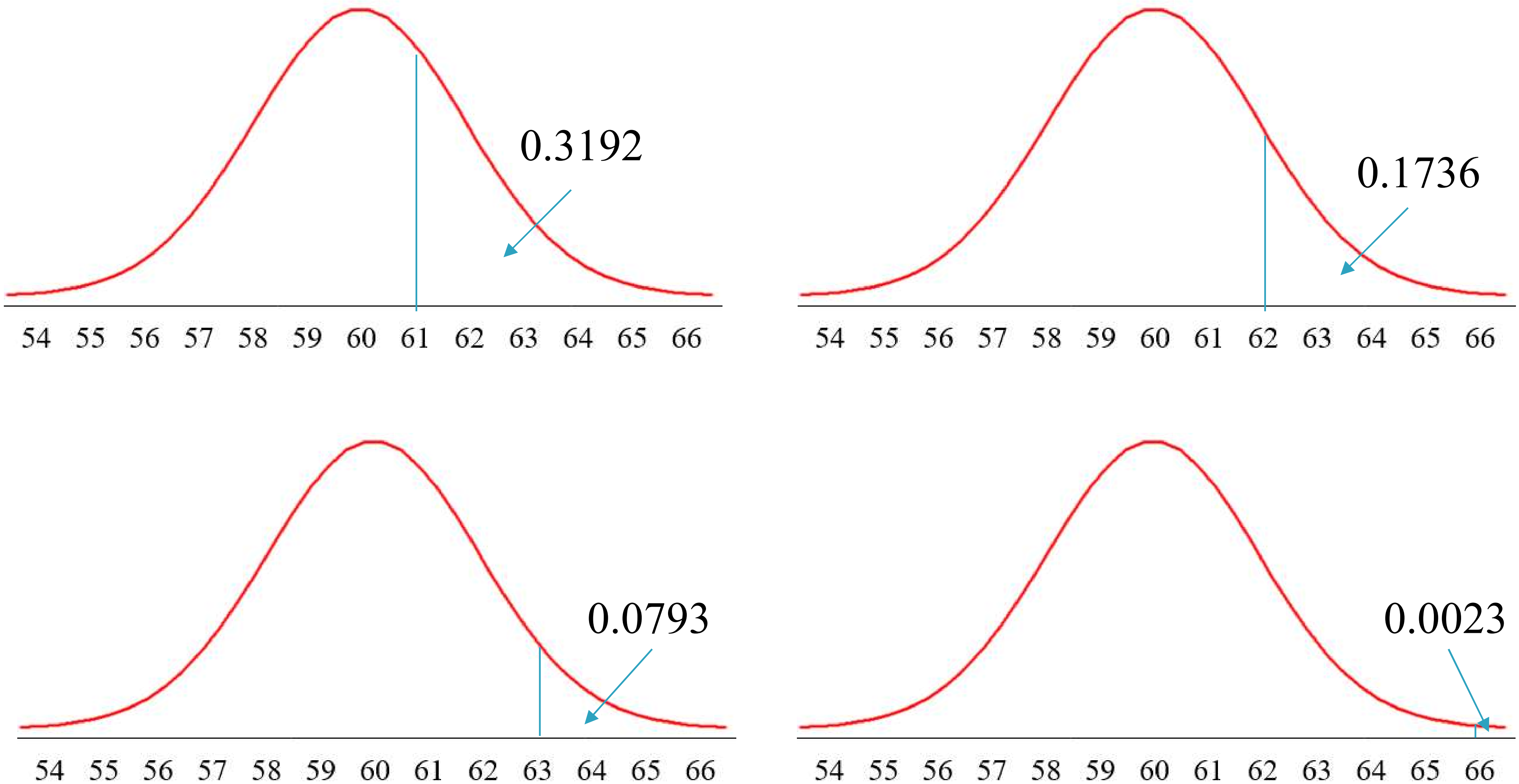
Hypothesis testing

Why does a low probability provide evidence against our original assumption that $\mu = 60$ km/h and in favour of the claim that $\mu > 60$ km/h?

To answer this question, imagine that instead of finding a sample mean of 66, we had calculated a sample mean of only 61 km/h. It should be intuitive that **a sample mean of 61 is not very convincing evidence that the true mean is greater than 60**.

You can check that if the true mean was 60, the probability of observing a sample mean greater than or equal to 61 would be 0.3192.

Hypothesis Testing



Hypothesis Testing

Obviously, the **higher** $\bar{x}$ is, the **more convinced** we become that the true mean is greater than 60. But notice from the diagrams on the previous slide what this means in terms of the probability we are calculating.

The **higher** $\bar{x}$ gets, the **lower** the probability. This is why the **lower** the probability is, the **more convinced** we are that μ must be greater than 60.

Hypothesis Testing

We have some new vocabulary in hypothesis testing, but the idea is a simple one: **An outcome that would rarely occur if an assumption were true is good evidence that the assumption is not true**.

Because we are interested in the value of a parameter for the whole population, **we always express our statements of interest in terms of population parameters (in this case, μ).**

Null Hypothesis

The opposing statements in a hypothesis test are expressed as **two hypotheses:**

The statement being tested in a hypothesis test is called the **null hypothesis**, denoted H_0. The test is designed to assess the strength of evidence **against** the null hypothesis.

H_0 is always a statement of "no difference" or "no effect". It will always be expressed in the form of an equality.

Alternative Hypothesis

The statement making the claim we are trying to support is called the **alternative hypothesis**, denoted H_a, which will always be expressed as an inequality.

The null and alternative hypotheses are precise statements of what claims we are testing. They are both given in terms of the population parameter μ.

P-value

Note that **we assume the null hypothesis is true** and calculate the probability of observing a value of the sample mean **at least as extreme as the one observed**.

This probability is called the **P-value** of the test. The **lower** the P-value, the less likely it would have been to observe a value of $\bar{x}$ as extreme as the one observed **if H_0 were true.** In other words, the lower the P-value, the stronger our evidence against the null hypothesis (and in favour of the alternative).

P-value

In our vitamin C example, the probability of 0.3085 that we calculated was in fact the P-value, which we concluded was not sufficiently low to reject our original assumption that $\mu = 75$ and to accept the nutritionist's claim.

In the speeding vehicle example, the P-value was 0.0023, which **was** sufficiently low to reject our original assumption that $\mu = 60$ and conclude that the true mean speed really is higher than 60 km/h.

P-value

This raises an obvious question:

How low must the P-value be before we are willing to reject the null hypothesis in favour of the alternative claim? For example, if the P-value is 0.08, do we consider this to be convincing enough evidence to say the null hypothesis is false?

It depends!

Level of Significance

Prior to the test, we must choose a **level of significance α**, to which we will compare the P-value.

If the P-value is less than or equal to α, we will reject H_0 in favour of H_a. If it is greater than α, we **fail to reject** H_0.

Level of Significance

As such, we can think of α as the **maximum P-value for which the null hypothesis will be rejected**.

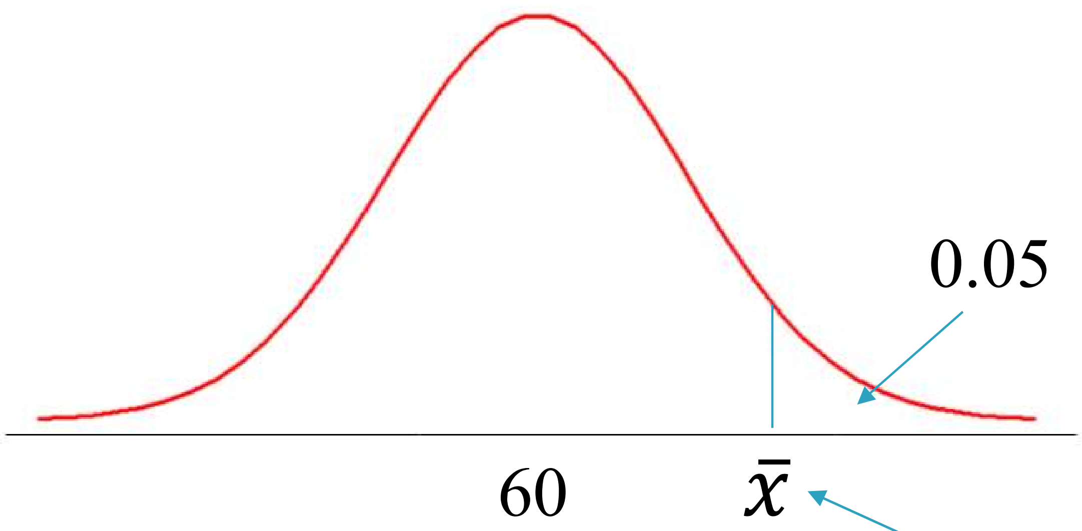
The most common values of α are 0.1, 0.05 and 0.01. We never use a value higher than 0.1 (for the same reason we almost never use a confidence level less than 90%). If a sample mean as extreme as the one we observed would occur more than 10% of the time (if H_0 is true), we will never consider it sufficient evidence to reject the null hypothesis.

Level of Significance

Note that when we ask, "how low does the P-value have to be for us to be convinced the claim is true?", it is analogous to asking, "how high does $\bar{x}$ have to be for us to be convinced the claim is true?"

For example, suppose we use a 5% level of significance. ($\alpha = 0.05$ is by far the most common level of significance used in statistical tests.)

Level of Significance



If we used a 5% level of significance, **this** would be our cutoff value (the value with area 0.05 above it). Any sample mean greater than this value would be considered sufficient evidence to reject the null hypothesis (to believe the claim with enough certainty).

Level of Significance

We now look at the previous example and do the formal hypothesis test using the new terminology we have.

Step 1: State the level of significance

Let $\alpha = 0.05$.

We will be willing to conclude in favour of the council only if the P-value is less than or equal to 0.05. (*In practice, you select the level of significance yourself, but for the purpose of this course, it will always be given.*)

Null and Alternative Hypotheses

Step 2: Statement of Hypotheses

In words:

H_0: The true mean speed at the intersection is equal to the posted limit and no red-light camera is needed.

H_a: The true mean speed at the intersection is greater than the posted limit and a red-light camera is needed.

In symbols:

$$H_0: \mu = 60 \quad \text{vs.} \quad H_a: \mu > 60$$

Decision Rule

Step 3: Statement of the Decision Rule

The ***decision rule*** *(also known as the* ***rejection rule****) is a precise statement of what must happen in order for us to reject the null hypothesis.*

Reject H_0 if the P-value $\leq \alpha = 0.05$.

Test Statistic

Step 4: Calculation of the Test Statistic

The ***test statistic*** *provides a measure of the compatibility between the null hypothesis and our data. When testing for the value of μ for a normal distribution with a known value of σ, our test statistic will be Z:*

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{66 - 60}{15 / \sqrt{50}} = 2.83$$

Note that this is calculated **assuming the null hypothesis is true** (using 60 for the mean).

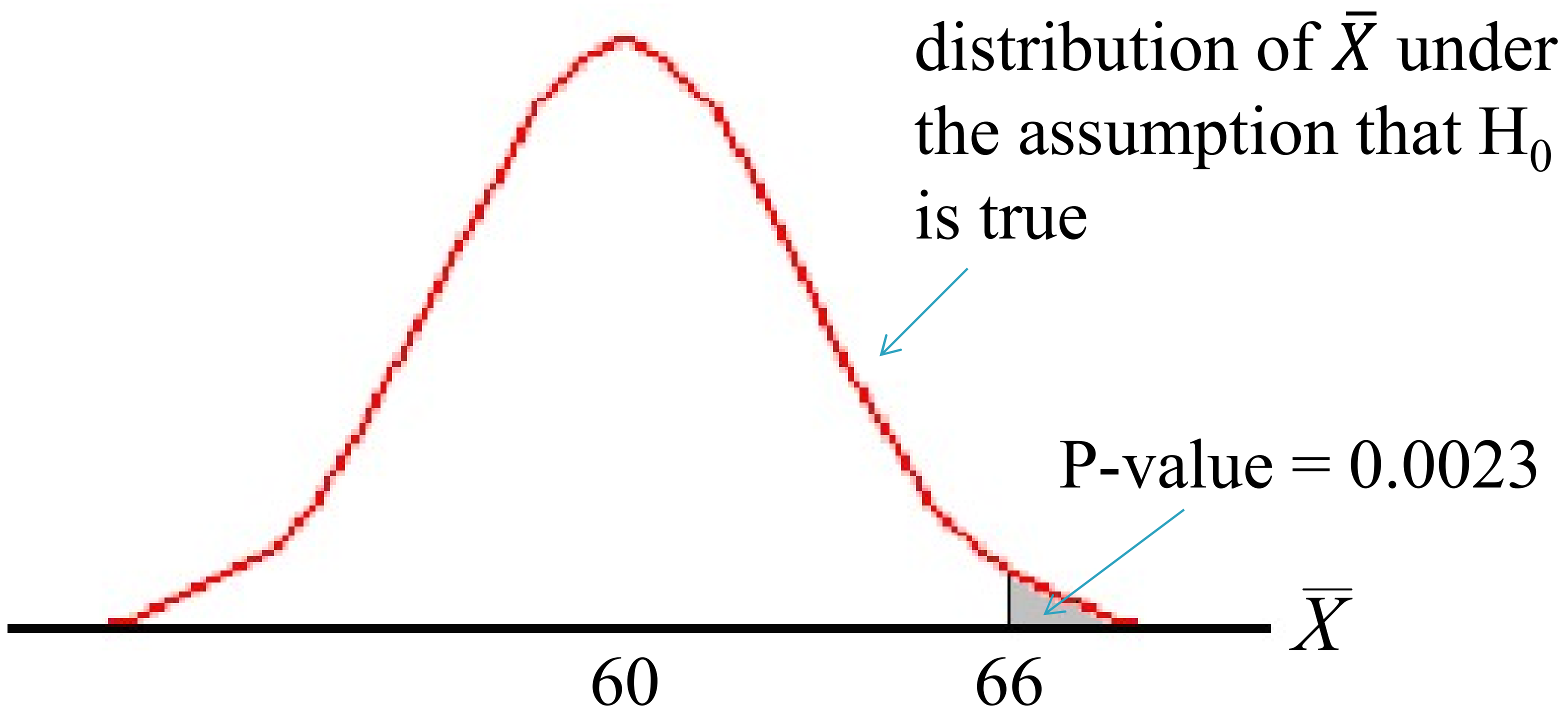
P-value

Step 5: Calculate the P-value

$$\text{P-value} = P(Z \geq 2.83) = 1 - P(Z < 2.83)$$
$$= 1 - 0.9977 = 0.0023$$

If the true mean speed of vehicles at this intersection was 60, the probability of taking a sample of 50 vehicles with a mean speed at least as high as 66 would only be 0.0023.

P-value



distribution of $\bar{X}$ under the assumption that H_0 is true
P-value = 0.0023
$\bar{X}$
60
66

Conclusion

Step 6: Conclusion

Since the P-value = 0.0023 < α = 0.05, we reject the null hypothesis in favour of the alternative. At a 5% level of significance, we have sufficient statistical evidence to conclude that the true mean speed of motorists at the intersection is greater than the posted limit of 60 km/h.

We were willing to make this conclusion for any P-value less than or equal to 0.05. Statistically speaking, the claim of the council does in fact have merit.

R Code

```
> Speed <- c(68, 53, 61, 56, 88, 78, 61, 72, 67, 64,
             40, 48, 83, 58, 50, 51, 55, 68, 60, 79,
             47, 70, 85, 65, 75, 77, 63, 54, 90, 64,
             57, 64, 73, 73, 67, 45, 61, 72, 92, 84,
             61, 72, 67, 83, 75, 68, 82, 58, 37, 59)
> z.test(Speed, mu = 60, sigma.x = 15, alternative =
         "greater")
```

```
data:  Speed
z = 2.8284, p-value = 0.002339
alternative hypothesis: true mean is greater than 60
95 percent confidence interval:
 62.51074       NA
sample estimates:
mean of x
       66
```

Practice Question

The response times of technicians of a large heating company follow a normal distribution with a standard deviation of 10 minutes. A supervisor suspects that the mean response time has increased from the target time of 30 minutes. He takes a random sample of 25 response times and calculates the sample mean response time to be 33.8 minutes. What are the null and alternative hypotheses for the appropriate hypothesis test?

(A) H_0: $\mu = 30$ vs. H_a: $\mu \neq 30$

(B) H_0: $\overline{X} = 30$ vs. H_a: $\overline{X} > 30$

(C) H_0: $\mu = 33.8$ vs. H_a: $\mu > 33.8$

(D) H_0: $\overline{X} = 33.8$ vs. H_a: $\overline{X} > 33.8$

(E) H_0: $\mu = 30$ vs. H_a: $\mu > 30$

Practice Question

The response times of technicians of a large heating company follow a normal distribution with a standard deviation of 10 minutes. A supervisor suspects that the mean response time has increased from the target time of 30 minutes. He takes a random sample of 25 response times and calculates the sample mean response time to be 33.8 minutes. What is the value of the test statistic for the appropriate hypothesis test?

(A) $z = 1.65$
(B) $z = 2.09$
(C) $z = 0.77$
(D) $z = 1.48$
(E) $z = 1.90$

Statistical Significance

Results that lead to the rejection of a null hypothesis are said to be **statistically significant**.

For this reason, statistical hypothesis tests are also referred to as **tests of significance**.

Statistical significance is an effect so large that it would rarely occur by chance alone.

Hypothesis Testing

Note that in the speeding vehicle example, we were interested in testing the claim that the mean speed of vehicles at the intersection was **greater than 60.**

Let us consider the vitamin C example where we were interested if the population was **less than** 75 mg.

Example

Step 1

Let $\alpha = 0.10$.

Step 2

We are testing the hypotheses

H_0: The true mean daily vitamin C intake for all Canadians is equal to the recommended amount of 75 mg.

H_a: The true mean daily vitamin C intake for all Canadians is less than the recommended amount of 75 mg.

Equivalently,

$$H_0 : \mu = 75 \quad \text{vs.} \quad H_a : \mu < 75$$

Example

Note that the null hypothesis is still an expression of equality, but that the alternative is now **left-sided**.

Step 3

Reject H_0 if the P-value $\leq \alpha = 0.10$.

Step 4

We calculate the test statistic:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{73 - 75}{20/\sqrt{25}} = -0.50$$

Example

<u>Step 5</u>

The P-value of the test is

$$\text{P-value} = P(Z \leq -0.50) = 0.3085$$

Notice that the P-value is now an expression of the form $P(Z \leq z)$ rather than the form $P(Z \geq z)$ that we used in the previous example. This is because the P-value is the probability of observing a value of the test statistic at least as extreme as the one observed, given the null hypothesis is true. In this case, since our alternative hypothesis was left-sided, an "extreme" value of the test statistic indicates a **low** value.

Example

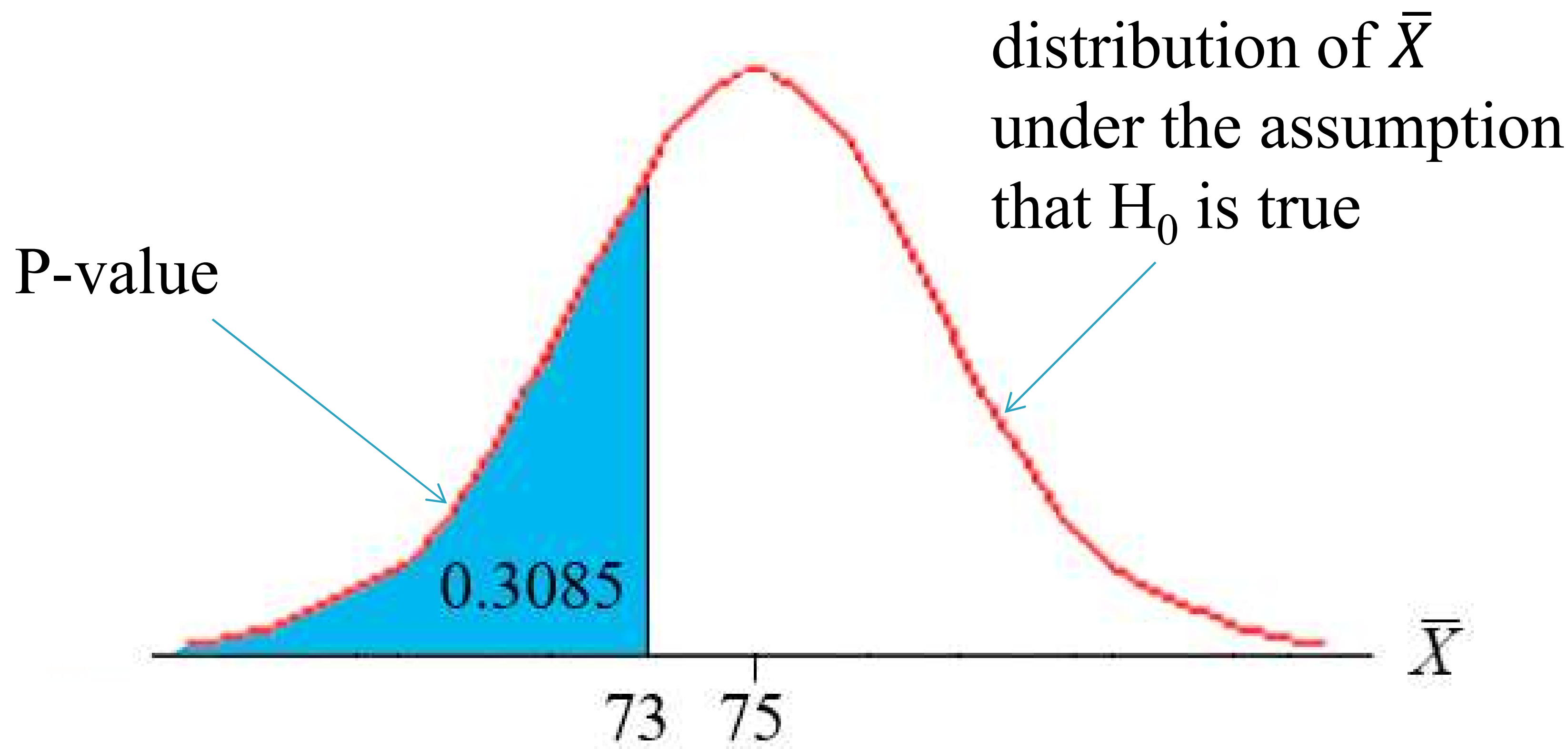
A simple rule for left- or right-sided tests:

The direction of the arrow in the P-value is the same as the direction of the arrow in the alternative hypothesis.

P-value = 0.3085 means:

If the true mean daily vitamin C intake was 75 mg, the probability of observing a sample mean as least as low as 73 mg would be 0.3085.

Example


distribution of $\bar{X}$ under the assumption that H_0 is true
P-value
0.3085
73
75
$\bar{X}$

Example

<u>Step 6</u>

Since the P-value = 0.3085 > α = 0.10, we fail to reject the null hypothesis. At the 10% level of significance, we have insufficient statistical evidence that the true mean Vitamin C intake is less than 75 grams.

Notice that we do **not** conclude that H_0 is true. A sample average of 73 is certainly not "proof" that the population mean is equal to 75. All we can say is that we do not have enough evidence to reject the null hypothesis. **Never say "accept H_0"!**

R Code

```
> VitC <- c(58, 47, 115, 63, 93, 46, 86, 77, 29, 107,
            72, 69, 94, 80, 65, 32, 84, 64, 60, 51,
            88, 93, 102, 76, 74)
> z.test(VitC, mu = 75, sigma.x = 20, alternative =
         "less")
```

```
data:  VitC
z = -0.5, p-value = 0.3085
alternative hypothesis: true mean is less than 75
95 percent confidence interval:
      NA 79.57941
sample estimates:
mean of x
       73
```

Practice Question

The average body temperature of healthy adults is commonly thought to be 37.0°C, but a physician believes the mean is actually lower. She measures the body temperatures of a random sample of 16 healthy adult patients and calculates a sample mean body temperature of 36.9°C. Suppose it is known that body temperatures follow a normal distribution with standard deviation 0.2°C. The doctor conducts a hypothesis test to test her suspicion. What is the P-value for the appropriate hypothesis test?

(A) 0.3085 (B) 0.0228 (C) 0.0075 (D) 0.1587 (E) 0.0630

Two-Sided Tests

The other possible scenario is that we may simply be interested in the possibility that the population mean is **not equal to** the value specified in the null hypothesis. In this case, the P-value is the probability of observing a value of the test statistic **at least as extreme (in either direction)**, given that the null hypothesis is true. In this case, we find the P-value by **doubling** the probability to the left or right of z (whichever is lower).

Two-Sided Tests

This is because we were originally interested in the value of the sample mean being far from μ_0 **in either direction**.

Because we are assuming that the null hypothesis is true, getting a value equally as far from the hypothesized mean in the other direction would be equally likely.

For a **two-sided test**, if z is positive, double the upper tail. If z is negative double the lower tail.

P-value Calculation for One & Two-Sided Tests

In summary, the P-value for testing H_0: $\mu = \mu_0$ versus:

$$H_a: \mu > \mu_0 \text{ is } P(Z > z)$$

$$H_a: \mu < \mu_0 \text{ is } P(Z < z)$$

i.e., double the tail area

$$H_a: \mu \neq \mu_0 \text{ is } 2P(Z > |z|)$$

These P-values are exact if the population is normal and approximate for large sample size n in other cases.

Example

The Mackenzie Valley Bottling Company distributes root beer in bottles labeled 500 ml. They routinely inspect samples of 10 bottles prior to making a large shipment, hoping to detect if the true mean volume in the shipment differs from 500 ml. If the bottles are underfilled, the company could be in trouble for false advertising. If the bottles are overfilled, the company is spending more money than they need to.

Example

Suppose it is known that fill volumes for the bottles of root beer follow a normal distribution with standard deviation 3.5 ml. One random sample of 10 bottles results in a sample average volume of 502 ml.

Does this provide convincing evidence that the true mean fill volume for the shipment differs from the advertised amount of 500 ml?

Example

Step 1

Let $\alpha = 0.05$.

Step 2

H_0: The true mean volume of root beer in all bottles in the shipment is 500 ml.

H_a: The true mean volume of root beer in all bottles in the shipment differs from 500 ml.

Alternatively, $H_0 : \mu = 500$ vs. $H_a : \mu \neq 500$

Example

<u>Step 3</u>

Reject H_0 if the P-value $\leq \alpha = 0.05$.

<u>Step 4</u>

The test statistic is

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{502 - 500}{3.5 / \sqrt{10}} = 1.81$$

Example

Step 5

The P-value is $2P(Z \geq 1.81) = 2(1 - P(Z < 1.81))$
$= 2(1 - 0.9649) = 2(0.0351) = 0.0702$.

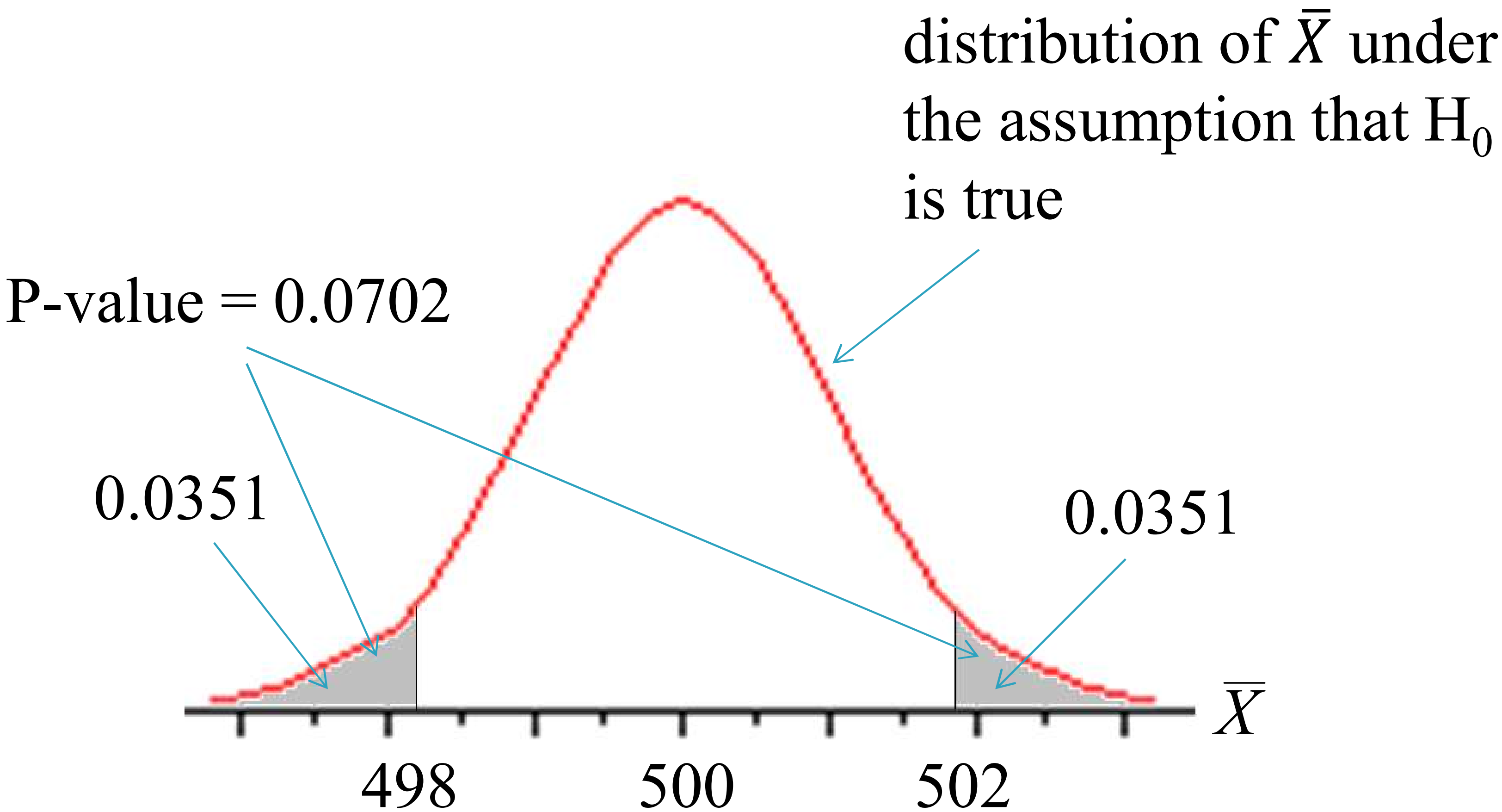
Step 6

Since the P-value = 0.0702 > α = 0.05, we fail to reject H_0. At the 5% level of significance, we have insufficient evidence that the true mean volume of all bottles in the shipment differs from 500 ml.

P-value Interpretation

If the true mean fill volume was 500 ml, the probability of observing a sample mean as least as extreme as 502 ml would be 0.0702.

Example



distribution of $\bar{X}$ under the assumption that H_0 is true
P-value = 0.0702
0.0351
0.0351
498
500
502
$\bar{X}$

R Code

```
> rootbeer <- c(507, 499, 503, 505, 496, 495, 506,
                505, 498, 506)
> z.test(rootbeer, mu = 500, sigma.x = 3.5,
         alternative = "two.sided")
```

```
data:  rootbeer
z = 1.807, p-value = 0.07076
alternative hypothesis: true mean is not equal to 500
95 percent confidence interval:
 499.8307 504.1693
sample estimates:
mean of x
      502
```

Hypothesis Testing

Note that in this case, the company doesn't want to see the null hypothesis rejected. That would indicate that the true mean fill volume differs from the amount printed on the label.

If the null hypothesis was rejected, they would likely need to stop the production process to investigate why the bottles are being overfilled (or underfilled).

Hypothesis Testing

However, in this case, we were very close to rejecting H_0. Our P-value of 0.0702 is only slightly higher than our (arbitrarily chosen) level of significance of 0.05.

Remember, the fact that we didn't reject H_0 doesn't mean we have evidence that the mean **is** 500 ml. It just means we didn't quite have enough evidence to conclude with enough certainty that μ differs from 500.

Hypothesis Testing

Because we got a value of $\bar{x}$ that was very close to rejecting H_0, the company might want to take another sample of bottles and conduct another test to verify the results.

As mentioned, in this case, the company doesn't want to reject H_0. Sometimes, the person conducting the hypothesis test has an interest in the outcome, and they often do want to reject H_0.

Hypothesis Testing

Suppose we wanted to reject H_0 in the previous hypothesis test.

We might say "our P-value of 0.0702 was almost lower than our level of significance (α = 0.05), which would allow us to reject H_0. So why don't we just change α to 0.10? Then our P-value would be lower than α, and we can reject H_0 as we wanted."

Hypothesis Testing

While increasing α to 10% would give us the desired result, it is unethical to make changes after a test has been completed.

We decided before the test that we would only consider a result to be significant if the P-value was less than 0.05. It is **not appropriate** to adjust our intentions after the fact.

Hypothesis Testing

If we wanted to reject H_0, we might also have reasoned as follows:

"We were testing whether μ **differs** from 500, but we got a sample mean $\bar{x} = 502$, which is **greater** than 500. So it stands to reason that if μ differs from 500, it is probably higher, so why don't we just conduct an upper tailed test of H_0: $\mu = 500$ vs. H_a: $\mu > 500$? Then we would not have doubled the tail area. Our P-value would be 0.0351, which is less than $\alpha = 0.05$. We would therefore reject H_0, which is the result we wanted."

Hypothesis Testing

Again, it would not be appropriate to do this. We were originally interested in determining whether μ differs from 500 (in either direction). The fact that $\bar{x}$ was greater than 500 doesn't mean we can change our alternative hypothesis, as though suddenly we are only interested in μ being greater than 500.

The level of significance and hypotheses are determined **before** the test is conducted, and cannot be changed to get a certain desired outcome.

P-value Interpretation Template

If the null hypothesis was true, the probability of

observing a value of the sample mean at least as

high
low
extreme

________________ as we did would be **P-value**______________.

Template for Writing a Conclusion

Since the P-value $\underline{\overset{<}{>}}$ α, we ___**reject** / **fail to reject**___ H_0.

At the α level of significance, we have ___**sufficient** / **insufficient**___

evidence that ___**H_a is true**___.

Practice Question

The prosecutor in a criminal trial introduces a key piece of evidence – a letter written by the criminal that was left at the scene of the crime. The defendant in the case is left-handed, and handwriting experts can often distinguish between the writing of left-handed and right-handed people. The prosecutor asks an expert to analyze the letter. The expert tests the hypotheses

H_0: The letter writer is right-handed.

H_a: The letter writer is left-handed.

Practice Question

H_0: The letter writer is right-handed.
H_a: The letter writer is left-handed.

The expert analyzes the letter and determines a P-value of 0.20. At a 5% level of significance, we conclude that:

(A) The probability the letter writer is left-handed is 0.20.
(B) The probability the letter writer is right-handed is 0.20.
(C) there is sufficient evidence that the letter writer is left-handed.
(D) there is insufficient evidence that the letter writer is right-handed.
(E) there is insufficient evidence that the letter writer is left-handed.

Practice Question

We would like to conduct a hypothesis test to determine if the true mean length of all commercials on a certain TV network differs from 30 seconds. The standard deviation of lengths of commercials on the network is known to be 9 seconds. A random sample of 36 commercials on the network has a mean length of 32 seconds. Conduct an appropriate test at the 5% level of significance.

Confidence Interval Method

For a two-sided test, we have one additional method to conduct the hypothesis test – the confidence interval method.

In order to use a confidence interval to conduct a hypothesis test, the following two conditions must be satisfied:

1) The test must be two-sided, **and**
2) The confidence level and level of significance must add up to one.

Confidence Interval Method

In other words, if we are conducting a two-sided test with a 1% level of significance, we need to use a 99% confidence interval. For a 5% level of significance, we need a 95% confidence interval, etc.

A two-sided test with significance level α rejects the null hypothesis if μ_0 falls outside the $100(1 - \alpha)\%$ confidence interval. If μ_0 falls inside the interval, we fail to reject H_0.

Example

Consider again the root beer example. We conduct the hypothesis test again, this time using the confidence interval method.

Step 1

Let $\alpha = 0.05$.

Step 2

$H_0 : \mu = 500$ vs. $H_a : \mu \neq 500$

Example

<u>Step 3</u>

Reject H_0 if $\mu_0 = 500$ is not in the 95% confidence interval for μ.

<u>Step 4</u>

The 95% confidence interval for μ is:

$$\bar{x} \pm z^* \left(\frac{\sigma}{\sqrt{n}}\right) = 502 \pm 1.96 \left(\frac{3.5}{\sqrt{10}}\right)$$

$$= 502 \pm 2.17 = (499.83, 504.17)$$

Example

Since $\mu_0 = 500$ falls within the 95% confidence interval, we fail to reject H_0. At the 5% level of significance, we have insufficient evidence that the true mean volume of all bottles in the shipment differs from 500 ml.

Note that we got the same conclusion using the P-value method and the confidence interval method. **This will always be the case**.

Practice Question

The manager at a grocery store would like to estimate the true mean amount of money spent by customers in the express lane. She selects a simple random sample of 50 receipts and calculates a 98% confidence interval for μ to be **($15.50, $20.75)**. Suppose we wish to conduct a hypothesis test to determine whether there is evidence that the true mean amount spent by customers in the express lane differs from $20. Which of the following statements is true?

(A) At a significance level of $\alpha = 0.01$, we have sufficient evidence that $\mu \neq 20$.
(B) At a significance level of $\alpha = 0.01$, we conclude that $\mu = 20$.
(C) At a significance level of $\alpha = 0.02$, we have sufficient evidence that $\mu \neq 20$.
(D) At a significance level of $\alpha = 0.02$, we have insufficient evidence that $\mu \neq 20$.
(E) At a significance level of $\alpha = 0.04$, we have insufficient evidence that $\mu \neq 20$.

Practice Question

The half-life of a drug is defined as the amount of time required for the concentration of the drug in the body to be reduced by one-half. A pharmaceutical company has made some adjustments in its manufacturing process and is interested in determining if the true mean half-life of a certain drug has changed. Prior to these adjustments, the mean was 48 hours. It is known that the half-life of the drug follows a normal distribution with standard deviation 3.4 hours. The drug is tested in clinical trials using 75 participants and the sample mean half-life is calculated to be 47.5 hours. Conduct a hypothesis test at the 5% level of significance.

Practice Question

In the half-life example, we tested

$$H_0: \mu = 48 \text{ vs. } H_a: \mu \neq 48$$

We calculated $\bar{x} = 47.5$ and obtained a P-value of 0.2040. What is the correct interpretation of this P-value?

(A) If the true mean half-life was 48 hours, the probability of getting a sample mean as extreme as 47.5 hours would be 0.2040.
(B) The probability that the true mean half-life differs from 48 hours is 0.2040.
(C) The probability that the true mean half-life is equal to 48 hours is 0.2040.
(D) 20.4% of similar tests will result in the rejection of H_0.
(E) If the true mean half-life was 48 hours, the probability of rejecting the null hypothesis would be 0.2040.

Practice Question

The time it takes people to drive from Winnipeg to Grand Forks, North Dakota follows a normal distribution with standard deviation 14 minutes. A random sample of 32 people driving from Winnipeg to Grand Forks had a mean travel time of 146 minutes.

(a) Conduct a hypothesis test at the 3% level of significance to determine whether the true mean travel time between Winnipeg and Grand Forks differs from 140 minutes.

(b) Based on the data, a 97% confidence interval for the true mean travel time between the two cities is calculated to be **(140.63, 151.37)**. Could this confidence interval have been used to conduct the test in (a)? Why or why not? If it could have been used, what would the conclusion be, and why?

Critical Value Method

We now examine an alternate method for conducting a hypothesis test.

Consider again the speeding vehicle example. Suppose that, instead of asking “What is the probability that the value of Z is at least as high as the value we observed?”, we ask:

How high does z have to be in order for us to reject the null hypothesis?

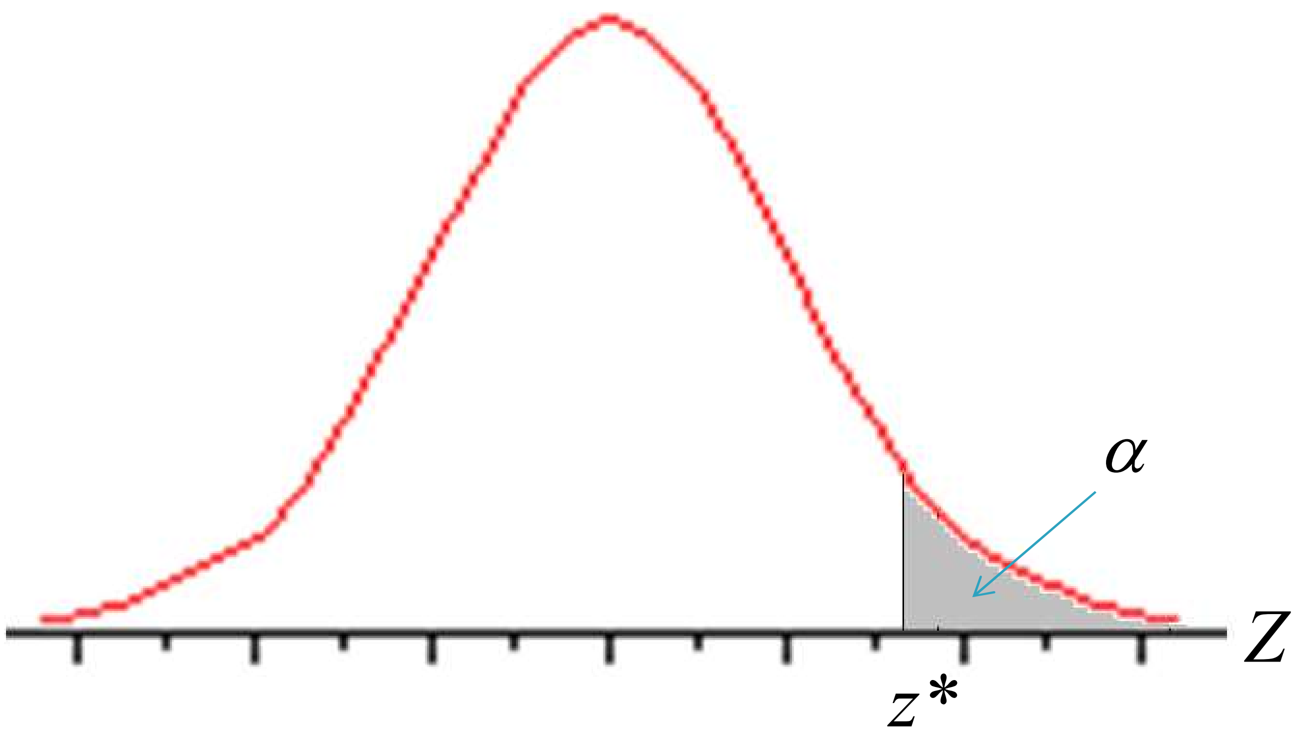
Critical Value Method

The level of significance for this test was given as $\alpha = 0.05$, and so we reject H_0 if the P-value ≤ 0.05.

In other words, we will reject the null hypothesis for any $z \geq z^*$, where

$$P(Z \geq z^*) = 0.05$$

Critical Value Method



$z \geq z^* \Leftrightarrow$ P-value $\leq \alpha \Leftrightarrow$ Reject H_0

Critical Value Method

From Table 2 we find this value to be $z^* = 1.645$. In other words, we will reject the null hypothesis if $z \geq 1.645$, and we will fail to reject H_0 if $z < 1.645$. (Instead of looking at the bottom of the table as we did for confidence intervals, we find α in the top row, then go to the bottom of that column to get the value of z^*.)

The value z^* is called the **critical value** of the test, and so the new test method is called the **critical value method**. The area beyond the value z^* under the normal curve is called the **critical region** or the **rejection region**.

We now revisit the speeding vehicle example and conduct the test using the critical value approach.

Example

Step 1: State the level of significance

Let $\alpha = 0.05$.

Step 2: Statement of Hypotheses

In words:

H_0: The true mean speed at the intersection is equal to the posted limit and no red-light camera is needed.

H_a: The true mean speed at the intersection is greater than the posted limit and a red-light camera is needed.

In symbols:

$$H_0: \mu = 60 \quad \text{vs.} \quad H_a: \mu > 60$$

Example

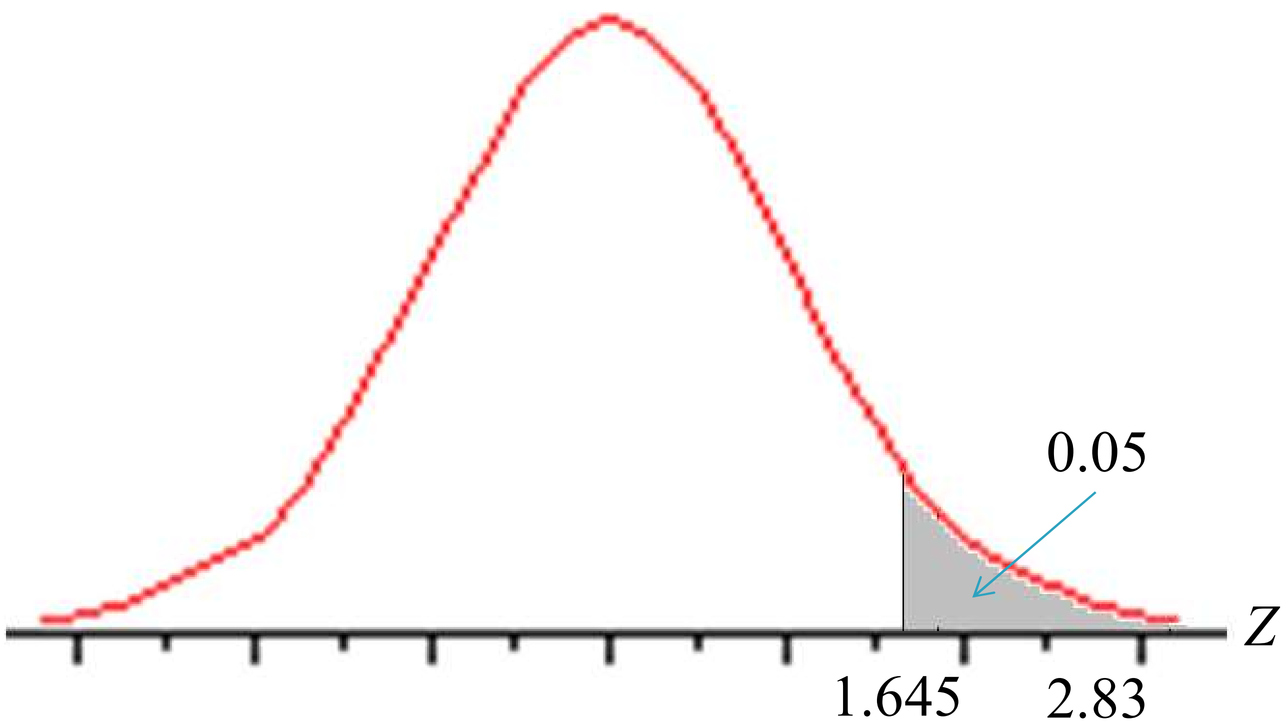
Step 3: Statement of the Decision Rule

Reject H_0 if $z \geq z^* = 1.645$.

Step 4: Calculation of the Test Statistic

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{66 - 60}{15 / \sqrt{50}} = 2.83$$

Example



$z > z^* = 1.645 \Leftrightarrow$ P-value $< \alpha \Leftrightarrow$ Reject H_0

Example

<u>Step 5: Conclusion</u>

Since z = 2.83 > z* = 1.645, we reject H_0. At a 5% level of significance, we have sufficient statistical evidence to conclude that the true mean speed of motorists at the intersection is greater than the posted limit of 60 km/h.

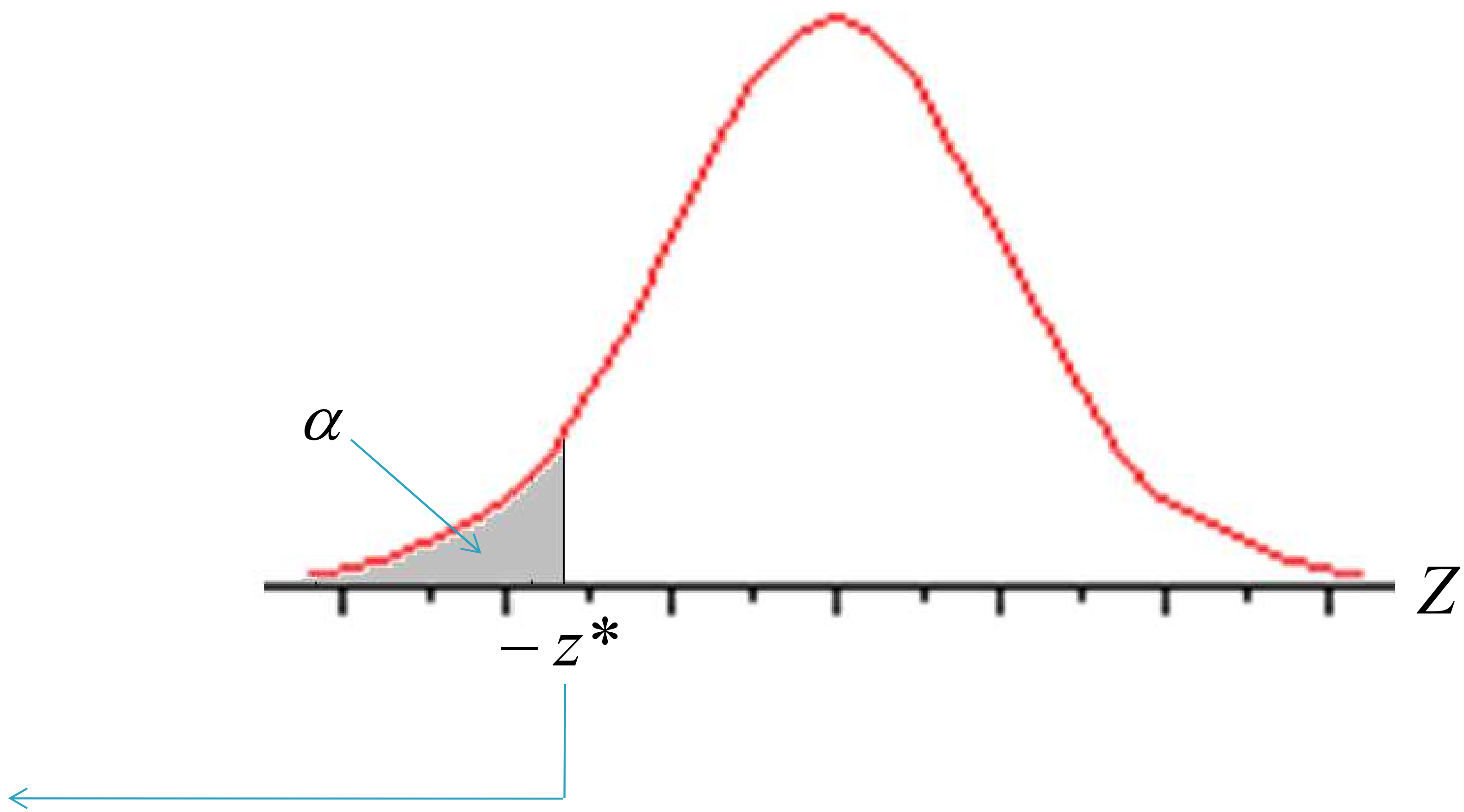
Critical Value Method

Note that there are only five steps in conducting a hypothesis test using the critical value approach, since the calculation of a P-value is no longer necessary.

In the case of a left-sided test, we reject the null hypothesis if $z \leq -z^*$, where $-z^*$ is the value of Z such that

$$P(Z \leq -z^*) = \alpha$$

Critical Value Method



$z < -z^* \Leftrightarrow$ P-value $\leq \alpha \Leftrightarrow$ Reject H_0

Example

We consider again the Vitamin C example using the critical value approach.

Step 1

Let $\alpha = 0.10$.

Step 2

H_0: The true mean daily vitamin C intake for all Canadian is equal to the recommended amount of 75 mg.

H_a: The true mean daily vitamin C intake for all Canadian is less than the recommended amount of 75 mg.

Equivalently,

$$H_0 : \mu = 75 \quad \text{vs.} \quad H_a : \mu < 75$$

Example

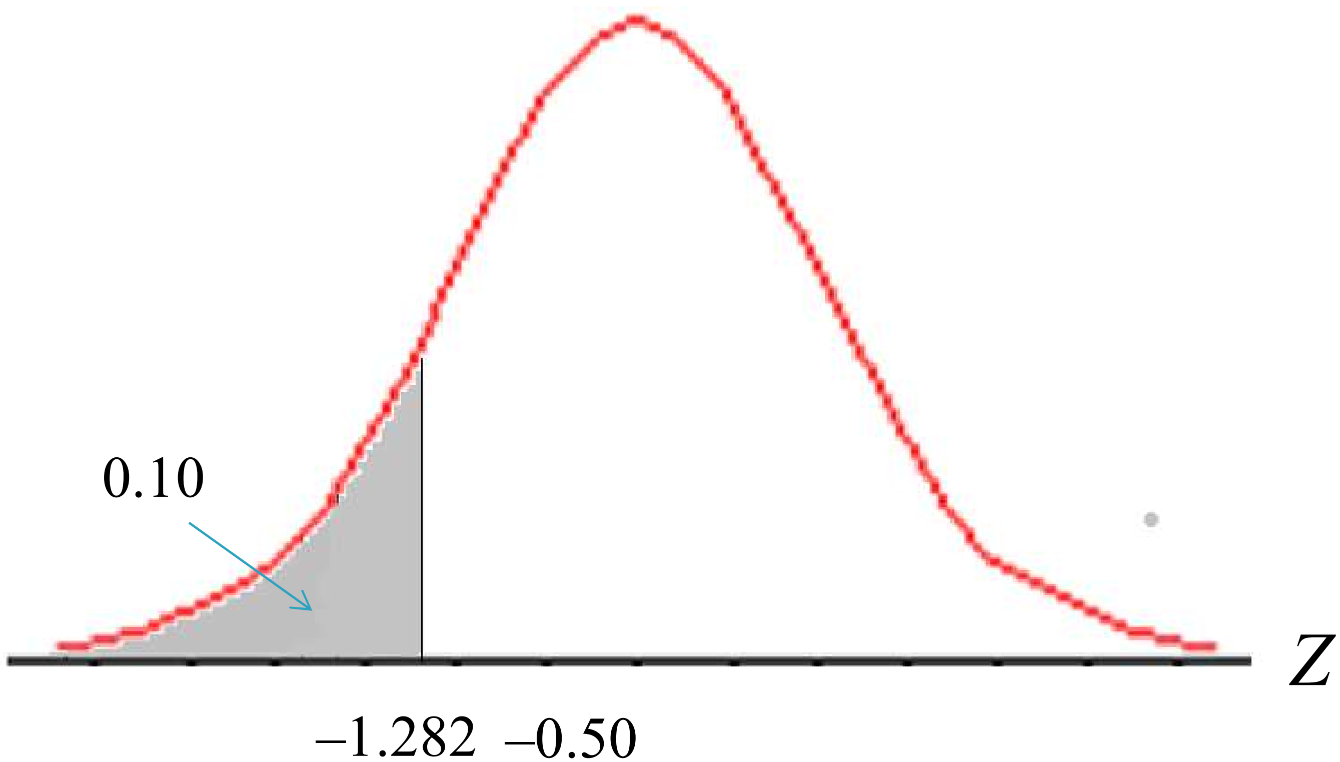
Step 3

Reject H_0 if $z \leq -z^* = -1.28$.

Step 4

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{73 - 75}{20/\sqrt{25}} = -0.50$$

Example



$z > -z^* \Leftrightarrow$ P-value $> \alpha \Leftrightarrow$ Fail to Reject H_0

Example

<u>Step 5</u>

Since $z = -\,0.50 > -\,z^* = -\,1.282$, we fail to reject the null hypothesis. At the 10% level of significance, we have insufficient statistical evidence that the true mean vitamin C intake is lower than 75 mg.

Practice Question

We would like to conduct a hypothesis test to determine whether the true mean height of Canadian women is less than 165 cm. We take a random sample of six women, and their heights are as follows:

157.4 168.8 159.9 162.3 160.0 158.7

Heights of Canadian women are known to follow a normal distribution with standard deviation 5.2 cm.

Practice Question

The test statistic is calculated to be $z = -1.80$. At a 5% level of significance, we should:

(A) reject H_0, since $z > -1.960$.
(B) fail to reject H_0, since $z > -1.960$.
(C) reject H_0, since $z < -1.645$.
(D) fail to reject H_0, since $z < -1.645$.
(E) reject H_0, since $z < -1.282$.

Critical Value Method

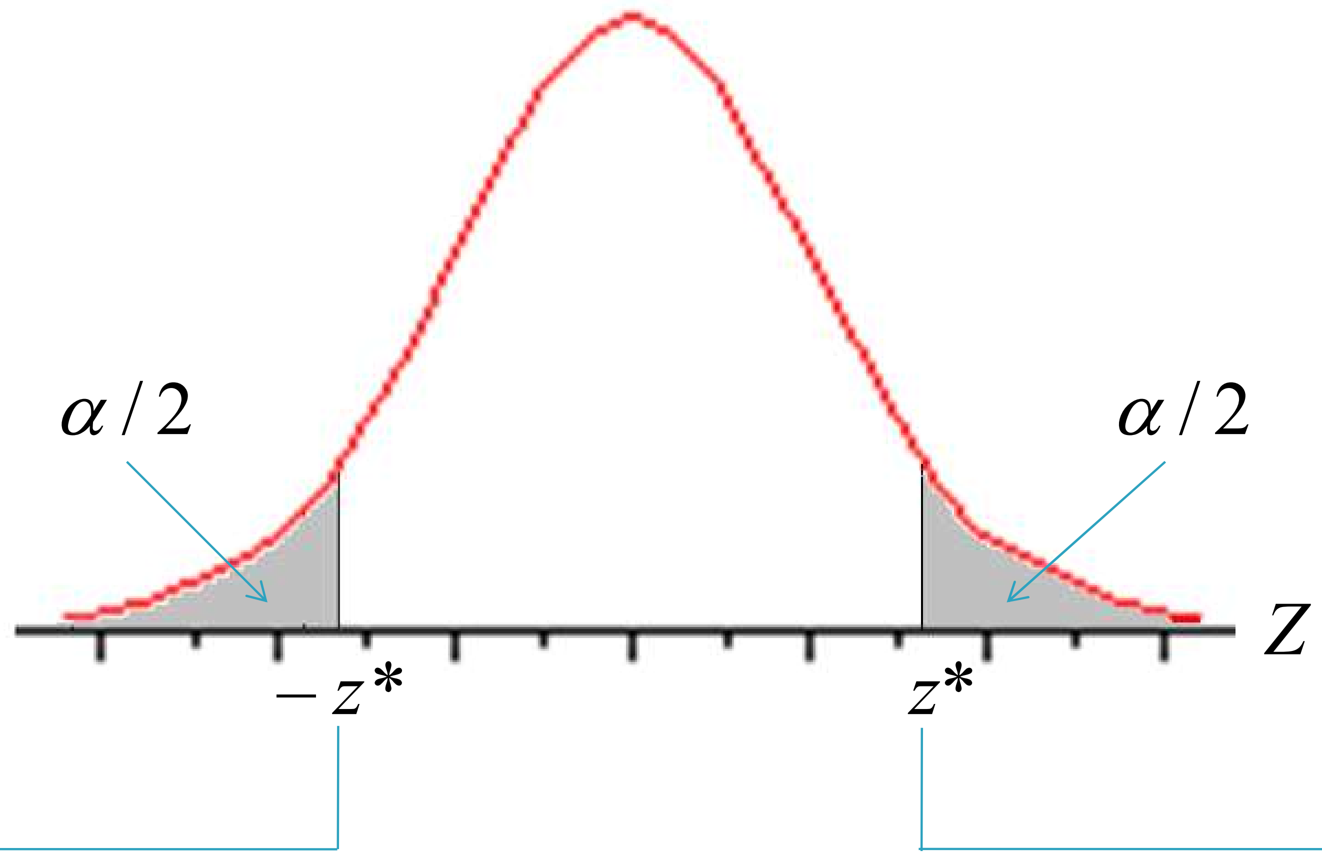
If we use the critical value approach for a two-sided test, we reject the null hypothesis if $|z| \geq z^*$, where z^* is the value of z such that

$$P(Z \geq z^*) = \alpha/2$$

i.e., we reject H_0 if

$$Z \leq -z^* \quad \text{or} \quad Z \geq z^*$$

Critical Value Method



$Z \le -z^*$ or $Z \ge z^* \Leftrightarrow$ P-value $\le \alpha \Leftrightarrow$ Reject H_0

Example

Consider again the root beer example, this time using the critical value approach.

Step 1

Let $\alpha = 0.05$.

Step 2

H_0: The true mean volume of root beer in all bottles in the shipment is 500 ml.

H_a: The true mean volume of root beer in all bottles in the shipment differs from 500 ml.

Alternatively, $H_0 : \mu = 500$ vs. $H_a : \mu \neq 500$

Example

<u>Step 3</u>

Reject H_0 if $|z| \geq z^* = 1.960$, i.e., if $z \leq -1.960$ or $z \geq 1.960$.

<u>Step 4</u>

The test statistic is

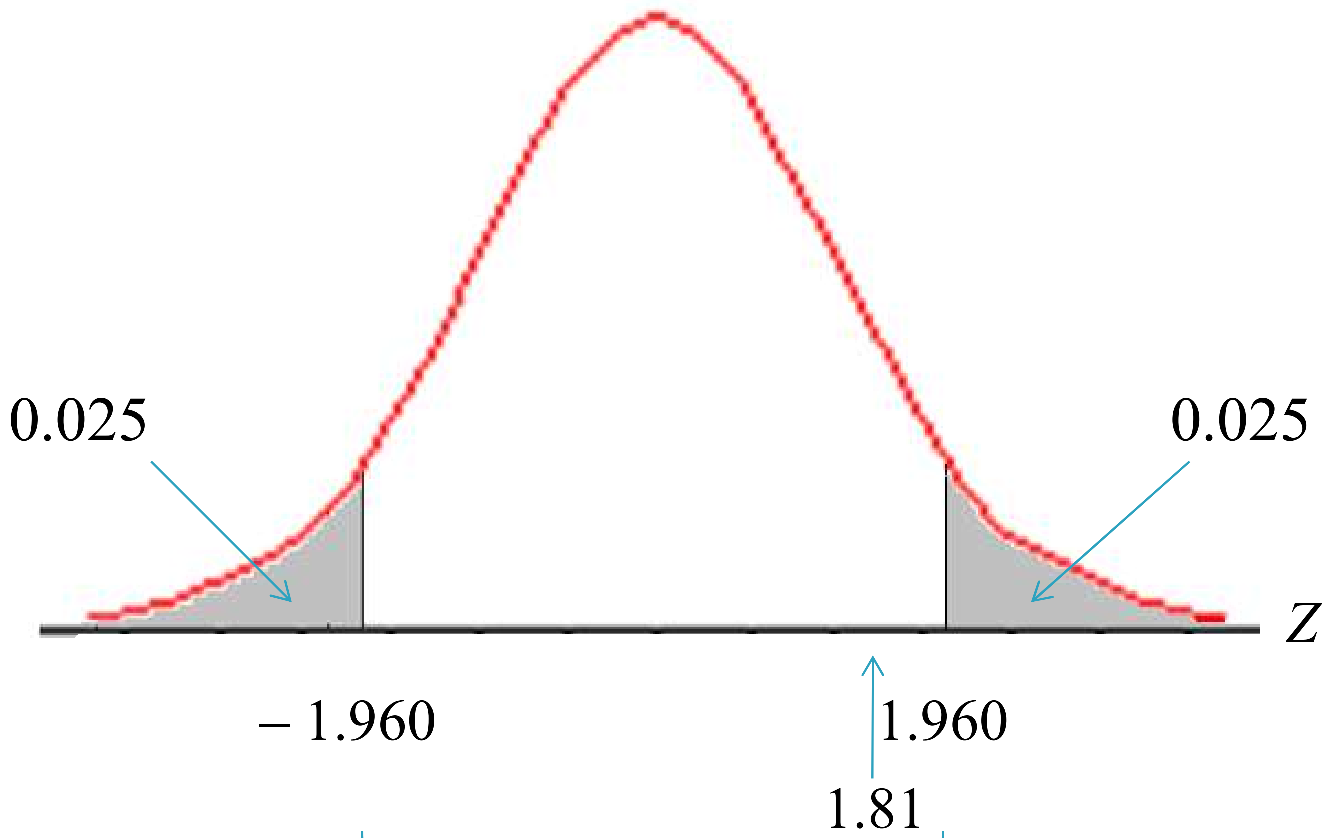
$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{502 - 500}{3.5 / \sqrt{10}} = 1.81$$

Example

<u>Step 5</u>

Since $-1.960 < z = 1.81 < 1.960$, we fail to reject H_0. At the 5% level of significance, we have insufficient evidence that the true mean volume of all bottles in the shipment differs from 500 ml.

Example



$$-z^* \leq Z \leq z^* \;\Leftrightarrow\; \text{P-value} > \alpha \;\Leftrightarrow\; \text{Fail to Reject } H_0$$

Summary

For an upper-tailed test: Find the value of α in the top row of Table 2, then go down that column to the z^* row. Reject H_0 if $z \geq z^*$.

For a lower-tailed test: Find the value of α in the top row of Table 2, then go down that column to the z^* row and put a negative sign in front of z^*. Reject H_0 if $z \leq -z^*$.

For a two-sided test: Find the value of $\alpha/2$ in the top row of Table 2, then go down that column to the z^* row. Reject H_0 if $z \leq -z^*$ or $z \geq z^*$. (Instead of looking up $\alpha/2$ in the top row, you could also look up the corresponding confidence level in the bottom row.)

Practice Question

Lengths of giraffes' necks are known to follow a normal distribution with standard deviation 17 cm. We will use the critical value approach to conduct a hypothesis test at the 2% level of significance to determine if the true mean length of all giraffes' necks differs from 180 cm. The decision rule is to reject H_0 if:

(A) $|z| > 2.054$
(B) $z > 2.054$
(C) $|z| > 2.326$
(D) $z > 2.326$
(E) $|z| > 2.576$