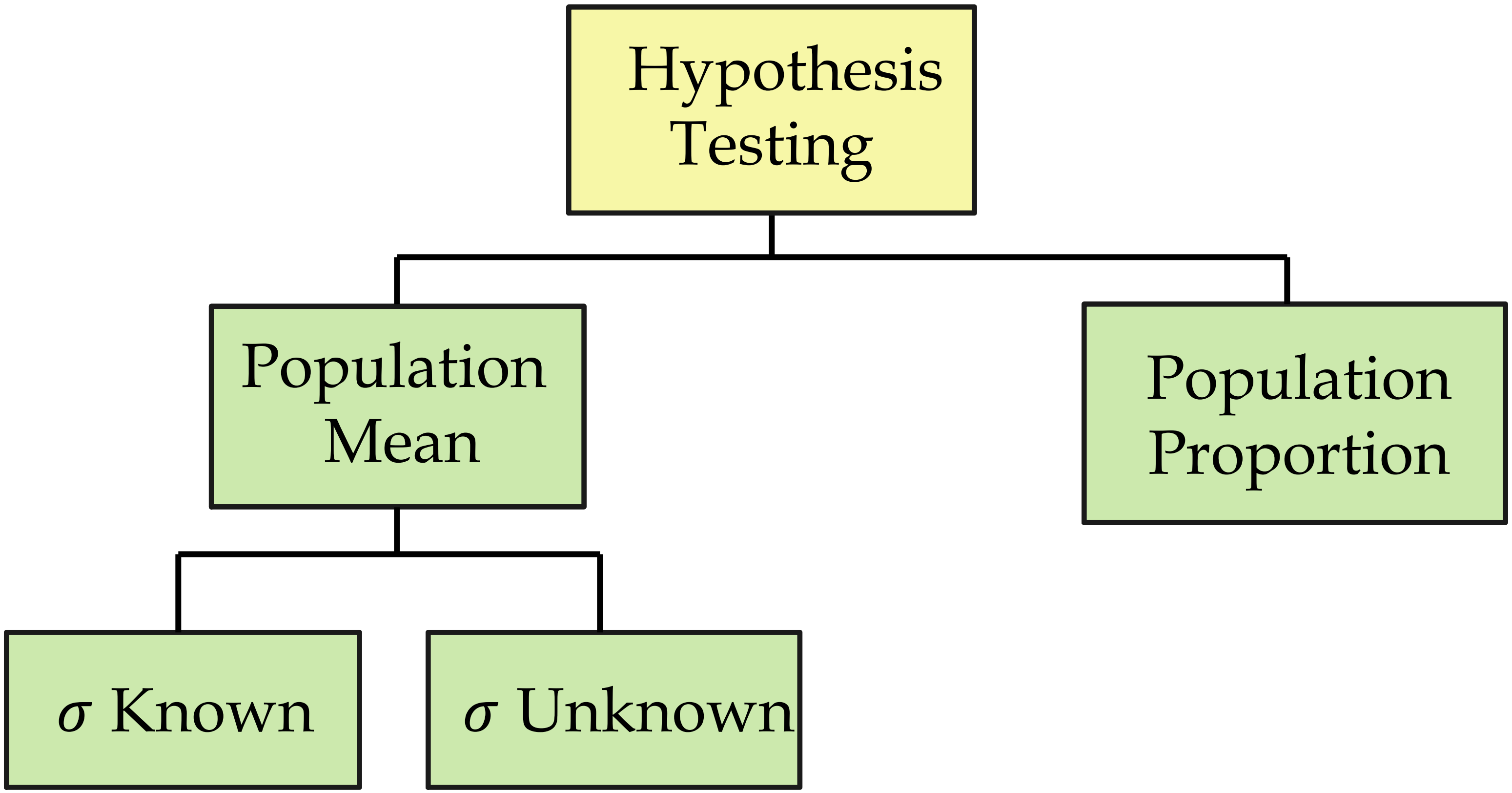
Hypothesis Testing for a Single Population

- An Introduction to Hypothesis Testing
- Developing Null and Alternative Hypotheses
- Hypothesis Testing for the Population Mean when σ is Known
- Next week
  - Hypothesis Testing for the Population Mean when σ is Unknown
  - Hypothesis Testing for the Population Proportion

- Reading: Chapter 9 (Sections 9.1-9.2)
  - Optional: using p-value in the courtroom (article on Canvas)… wait until we talk about p-value

Hypothesis Testing for a Single Population



Hypothesis Testing

A **hypothesis** is an assumption about value of a population parameter (such as mean or proportion)

Example: population mean

- The mean data use for smartphone users is μ = 1.8 gigabytes per month

Example: population proportion

- The proportion of cell phone users with 4G contracts is p = 0.62

Hypothesis Testing

- Hypothesis testing is used to determine whether a statement about the value of a population parameter should or should not be rejected

- The <u>null hypothesis</u>, denoted by H_0, is a presumed default state of nature / status quo / prior belief to be challenged in the testing procedure

- The <u>alternative hypothesis</u>, denoted by H_1, is the statement opposite (complementary) to the null hypothesis

- Hypothesis testing procedure uses data from a sample to determine whether or not sample evidence contradicts H_0

Hypothesis Testing

Every hypothesis test has both a **null hypothesis** and an **alternative hypothesis**

The **null hypothesis (H_0)** represents the status quo

- The null hypothesis is believed to be true unless there is overwhelming evidence to the contrary (H_1)
- States a belief that the population parameter is ≤, =, or ≥ a specific value (always includes an equality sign)

The **alternative hypothesis (H_1)** represents the opposite of the null hypothesis

- Is believed to be true if the null hypothesis H_0 is rejected
- The alternative hypothesis always states that the population parameter is >, ≠, or < a specific value (always opposite of what is stated in H_0)

Developing Null and Alternative Hypotheses

- As a general guideline, we use the alternative hypothesis H_1 as a vehicle to establish something new or support our belief about the value of the population parameter
- For example, many applications involve an attempt to gather evidence in support of a research hypothesis
  - In such cases, it is often best to begin with the alternative hypothesis and make it the conclusion that the researcher hopes to support
  - The conclusion that the research hypothesis is true is made if the sample data provide a sufficient evidence to show that the null hypothesis can be rejected

Developing Null and Alternative Hypotheses

- Example:

  A new teaching method is developed that is believed to be better than the current method

- Alternative Hypothesis:

  The new teaching method is better

- Null Hypothesis:

  The new method is no better than the old method

Developing Null and Alternative Hypotheses

In other situations, hypothesis tests are performed to show that a change has occurred from the status quo

⇒ The alternative hypothesis is used to represent the claim researchers wish to *support* (change occurred)

H_0: the unknown parameter *has not changed* from the status quo

H_1: there *has been a change* in the desired direction

Developing Null and Alternative Hypotheses

Stating the null and alternative hypotheses depends on the nature of the test and the motivation of the person conducting it

H_0: $\mu \leq 30$ H_1: $\mu > 30$	This test would be used by someone who thinks that the average pizza delivery time is bigger than advertised 30 minutes (rejecting the null would support the alternative that the average delivery time is > 30 min)
H_0: $\mu \geq 30$ H_1: $\mu < 30$	This would be used by someone who wants to test an assumption that the average pizza delivery time is less than advertised 30 minutes
H_0: $\mu = 30$ H_1: $\mu \neq 30$	This test would be used by someone who has no specific expectations, but wants to test the assumption that the average pizza delivery time is different from 30 minutes

Summary of Forms for Null and Alternative Hypotheses about a Population Mean

- The equality part of the hypotheses always appears in the null hypothesis
- In general, a hypothesis test about the value of a population mean μ must take one of the following forms:
  - μ_{H_0} is the hypothesized value of the population mean

One-tail (lower tail)	One-tail (upper tail)	Two-tailed
H_0: $\mu \geq \mu_{H_0}$	H_0: $\mu \leq \mu_{H_0}$	H_0: $\mu = \mu_{H_0}$
H_1: $\mu < \mu_{H_0}$	H_1: $\mu > \mu_{H_0}$	H_1: $\mu \neq \mu_{H_0}$

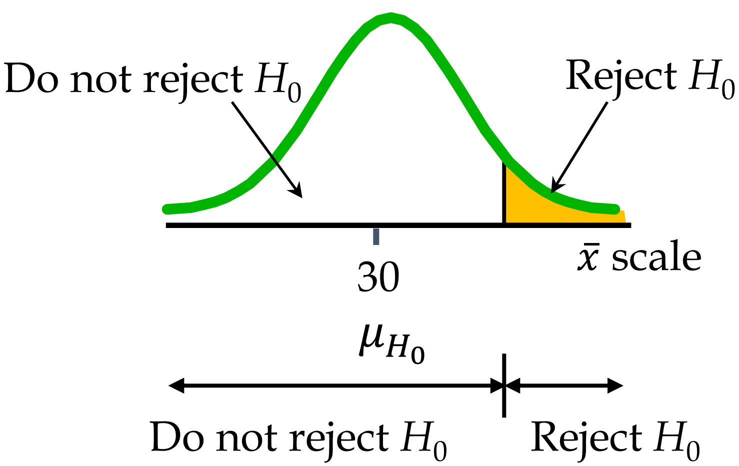
One-Tail Hypothesis Tests

A **one-tail hypothesis test** is used when the alternative hypothesis is stated as < or >

H_0: $\mu \leq 30$

H_1: $\mu > 30$

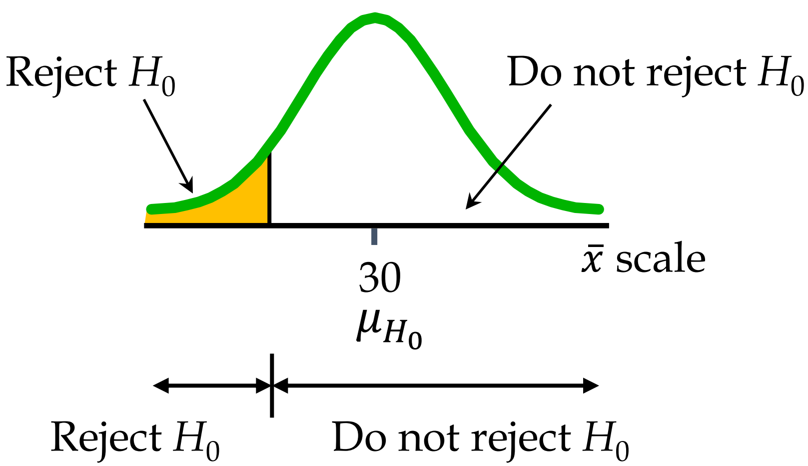
Upper tail test: We assume that μ = 30 unless the sample mean is much *higher* than 30



H_0: $\mu \geq 30$

H_1: $\mu < 30$

Lower tail test: We assume that μ = 30 unless the sample mean is much *lower* than 30

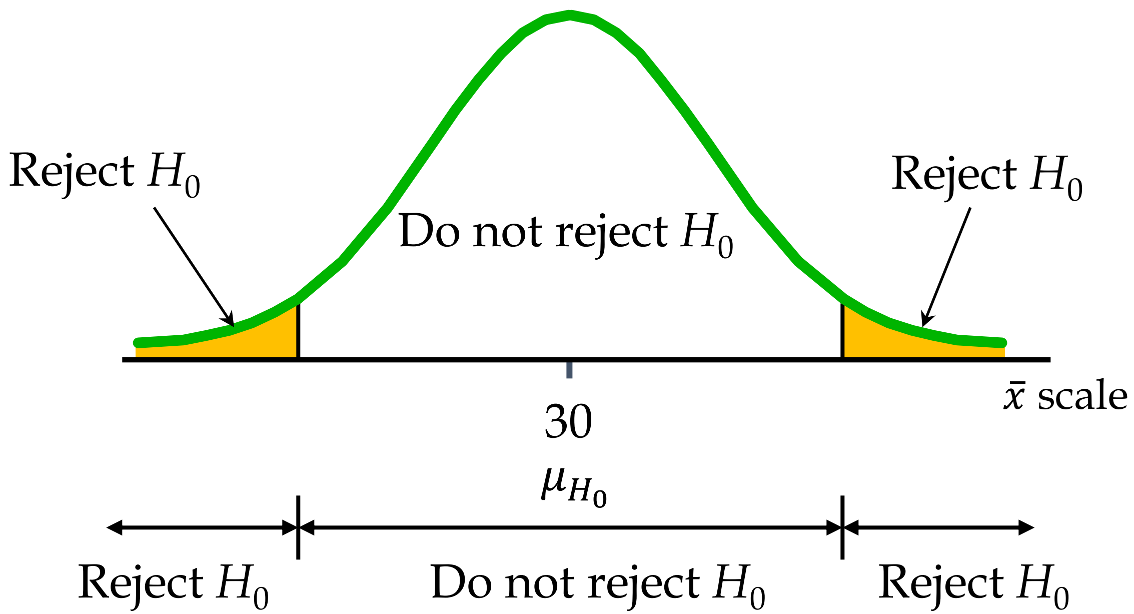


Two-Tail Hypothesis Tests

A **two-tail hypothesis test** is used whenever the alternative hypothesis is expressed as ≠

H_0: $\mu = 30$
H_1: $\mu \neq 30$

We assume that $\mu = 30$ unless the sample mean is *much higher* or *much lower* than 30



The Logic of Hypothesis Testing

- The null hypothesis can NEVER be accepted
- The only two options available are to:

  (1) **reject** the null hypothesis, or

  (2) **fail to reject** the null hypothesis

- The hypothesis testing relies on "proof by contradiction" ⇒ we start with the assumption that the null hypothesis is true and attempt to disprove it
  - Two possible conclusions:

  1) The sample result provides enough evidence to reject H_0
  2) The sample does not provide enough evidence to reject H_0

The Difference Between Type I and Type II Errors

Because we are relying on a sample, we are exposing ourselves to the risk that our conclusion about the population parameter can be wrong

A **Type I error** occurs when the null hypothesis is rejected when it is true

- The probability of making a Type I error is denoted by α, the level of significance

A **Type II error** occurs when we fail to reject the null hypothesis when it is not true

- The probability of making a Type II error is denoted by β

The Difference Between Type I and Type II Errors

Decision Rules for the Two Types of Hypothesis Errors

Possible Hypothesis Test Outcomes

	Actual State of H_0	
Decision	H_0 is True	H_0 is False
Reject H_0	Type I Error P(Type I Error) = α	Correct Outcome
Do Not Reject H_0	Correct Outcome	Type II Error P(Type II Error) = β

Steps of Hypothesis Testing (General)

Step 1. State the null and alternative hypotheses

Step 2. Specify the level of significance α

Step 3. Collect the sample data and compute the test statistic

<u>Critical Value Approach</u>

Step 4. Use the level of significance α to determine the critical value

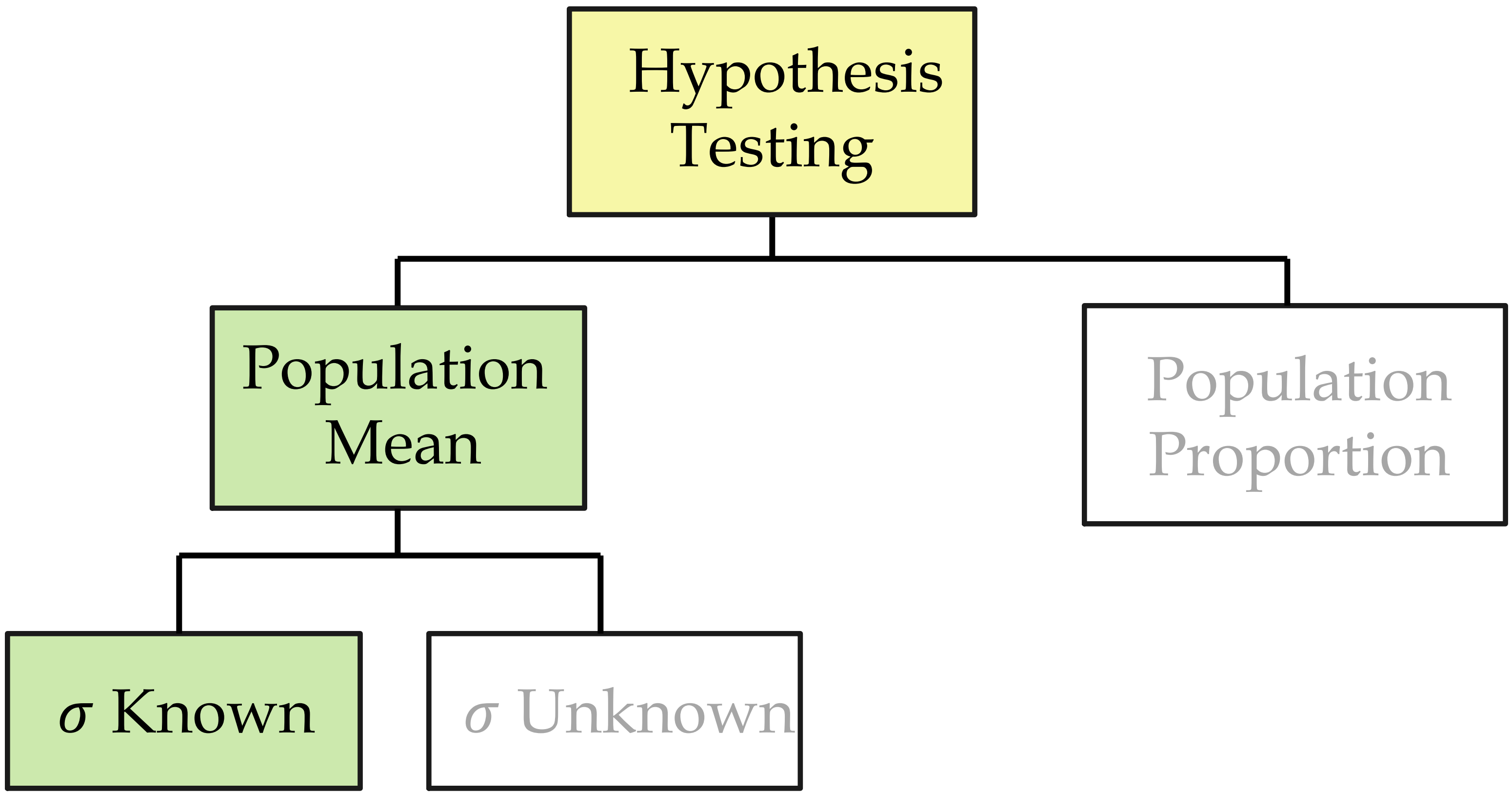
Step 5. Compare the value of the test statistic and the critical value to determine whether to reject H_0

<u>*p*-Value Approach</u>

Step 4. Use the value of the test statistic to find the p-value

Step 5. Compare the p-value and α : Reject H_0 if p-value $\leq \alpha$

Hypothesis Testing for the Population Mean when σ is Known



Hypothesis Testing for the Population Mean when σ is Known

In order to implement the test, it is essential that the sampling distribution of $\bar{x}$ is normal:

1. If the sample size is small ($n < 30$) the population must follow the normal distribution

2. If the sample size is large ($n \geq 30$) the *Central Limit Theorem* states that the sampling distribution of the mean follows the normal distribution (so, there is no restriction on the population distribution)

One-Tail Hypothesis Test for the Population Mean (σ is Known)

Example: Season's Pizza advertises that the average delivery time for their pizza is 30 minutes. You ordered a pizza from this place 10 times and the average delivery time for your orders was 36 minutes.

Based on your experience, you would like to set up a hypothesis test to check if the average pizza delivery time is higher than advertised.

- Suppose the delivery time for pizza orders at Season's Pizza follows the normal distribution
- Assume that the population standard deviation for the delivery time is 13 minutes

The following slides show **steps** to complete the test

One-Tail Hypothesis Test for the Population Mean (σ is Known)

Step 1: State the null and alternative hypotheses

H_0: $\mu \leq 30$ minutes (statement to be challenged: average delivery time is not bigger than advertised)

H_1: $\mu > 30$ minutes (average delivery time is bigger than advertised)

Step 2: Specify a value of the significance level, α

- The level of significance represents the probability of making a Type I error
- Let's take $\alpha = 0.05$
- Typically, the significance level will be given in the problem

One-Tail Hypothesis Test for the Population Mean (σ is Known)

Step 3: Calculate the appropriate test statistic

z-test statistic for a hypothesis test for the population mean when σ is known:

$$z_{\bar{x}} = \frac{\bar{x} - \mu_{H_0}}{\frac{\sigma}{\sqrt{n}}}$$

where:

$z_{\bar{x}}$ = the z-test statistic

$\bar{x}$ = the sample mean

μ_{H_0} = hypothesized value, which is assumed to be true and represents the mean of the sampling distribution

σ = the population standard deviation

n = the sample size

One-Tail Hypothesis Test for the Population Mean (σ is Known)

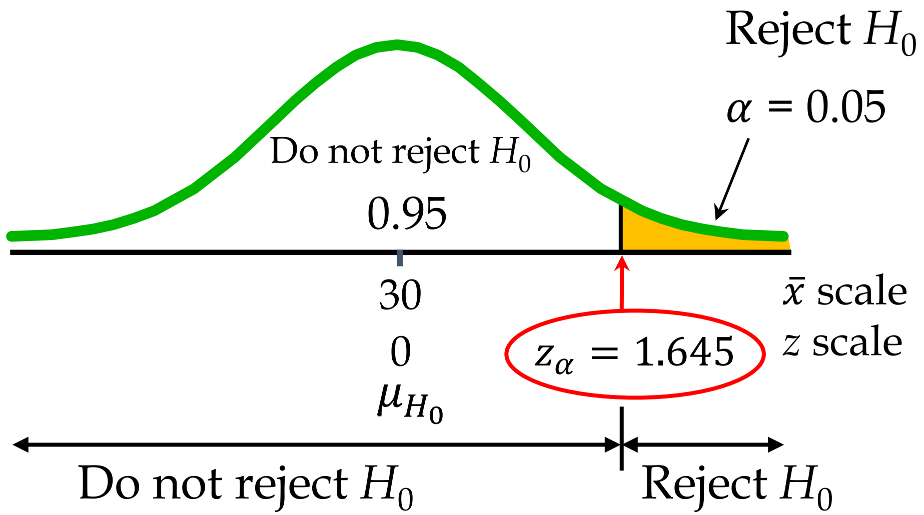
Step 3: Calculate the appropriate test statistic

$$z_{\bar{x}} = \frac{\bar{x} - \mu_{H_0}}{\frac{\sigma}{\sqrt{n}}} = \frac{36 - 30}{\frac{13}{\sqrt{10}}} = \frac{6}{4.11} = 1.46$$

One-Tail Hypothesis Test for the Population Mean (σ is Known): Critical Value Approach

Step 4: Determine the appropriate **critical value**

- σ is known $\Rightarrow$ use a critical z-score
- The critical z-score identifies the rejection region
- Since this is a one-tail (upper tail) test, the entire area for $\alpha = 0.05$ is placed on the right side (upper tail) of the z distribution:



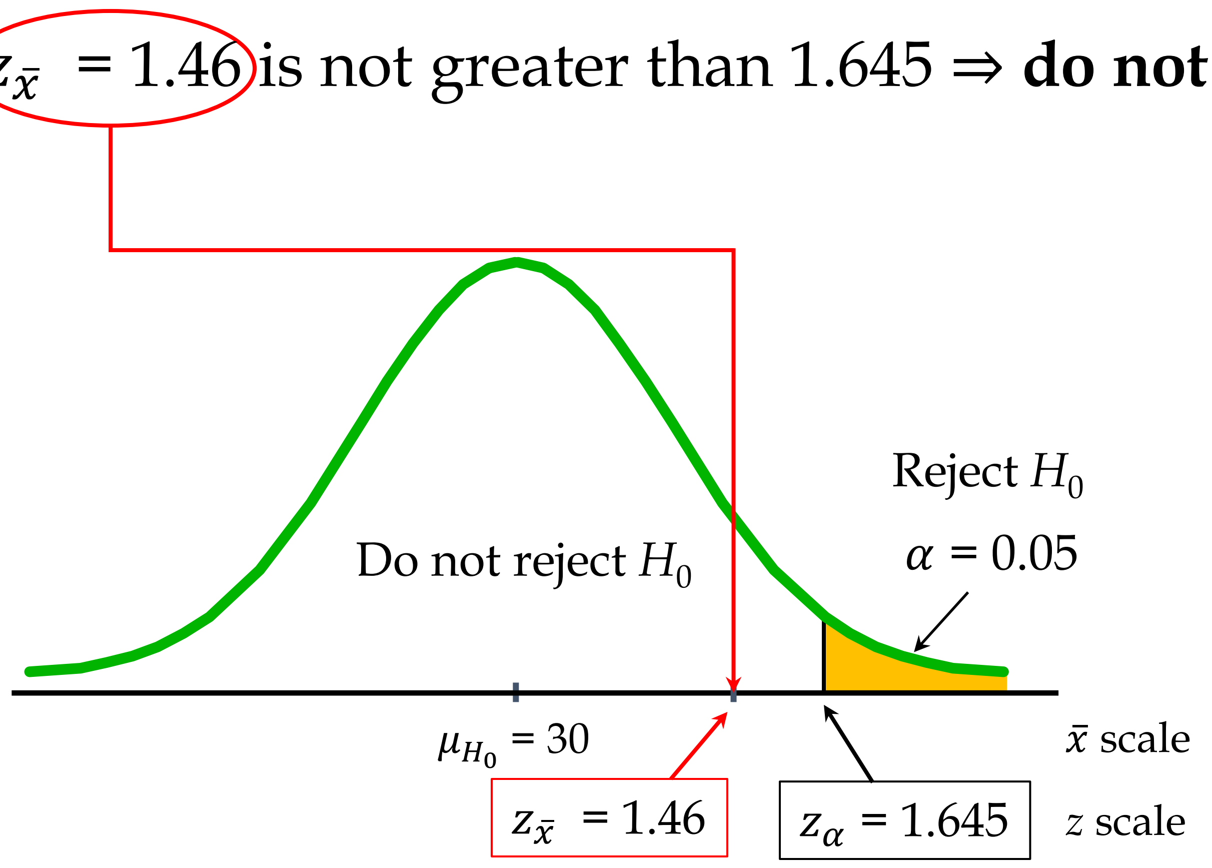
One-Tail Hypothesis Test for the Population Mean (σ is Known): Critical Value Approach

Step 5: Compare the test statistic $z_{\bar{x}}$ and the critical value z_α

- For a one-tail upper tail test, reject the null hypothesis if

$$z_{\bar{x}} > z_\alpha$$

- Here, $z_{\bar{x}} = 1.46$ is not greater than 1.645 ⇒ **do not reject H_0**



One-Tail Hypothesis Test for the Population Mean (σ is Known): Conclusion

Step 5: State the conclusion

Since the value of the test statistic $z_{\bar{x}}$ = 1.46 is smaller than the critical value z_{α} = 1.645, we can conclude that…

According to the sample of 10 pizza deliveries from Season's Pizza, there is not enough evidence to conclude that the average pizza delivery time exceeds advertised 30 minutes.

One-Tail Hypothesis Test for the Population Mean: *p*-value Approach

The ***p*-value** is the probability of observing a sample mean *at least as extreme* as the one selected for the hypothesis test, *assuming the null hypothesis is true with equality*

- Intuitively, the *p*-value is a probability that measures the evidence against the null hypothesis provided by the sample
- Smaller *p*-value indicates the value of the test statistic is unusual given that H_0 is true which provides more evidence against H_0

One-Tail Hypothesis Test for the Population Mean: *p*-value Approach

Let's go back to our example and use the *p*-value approach

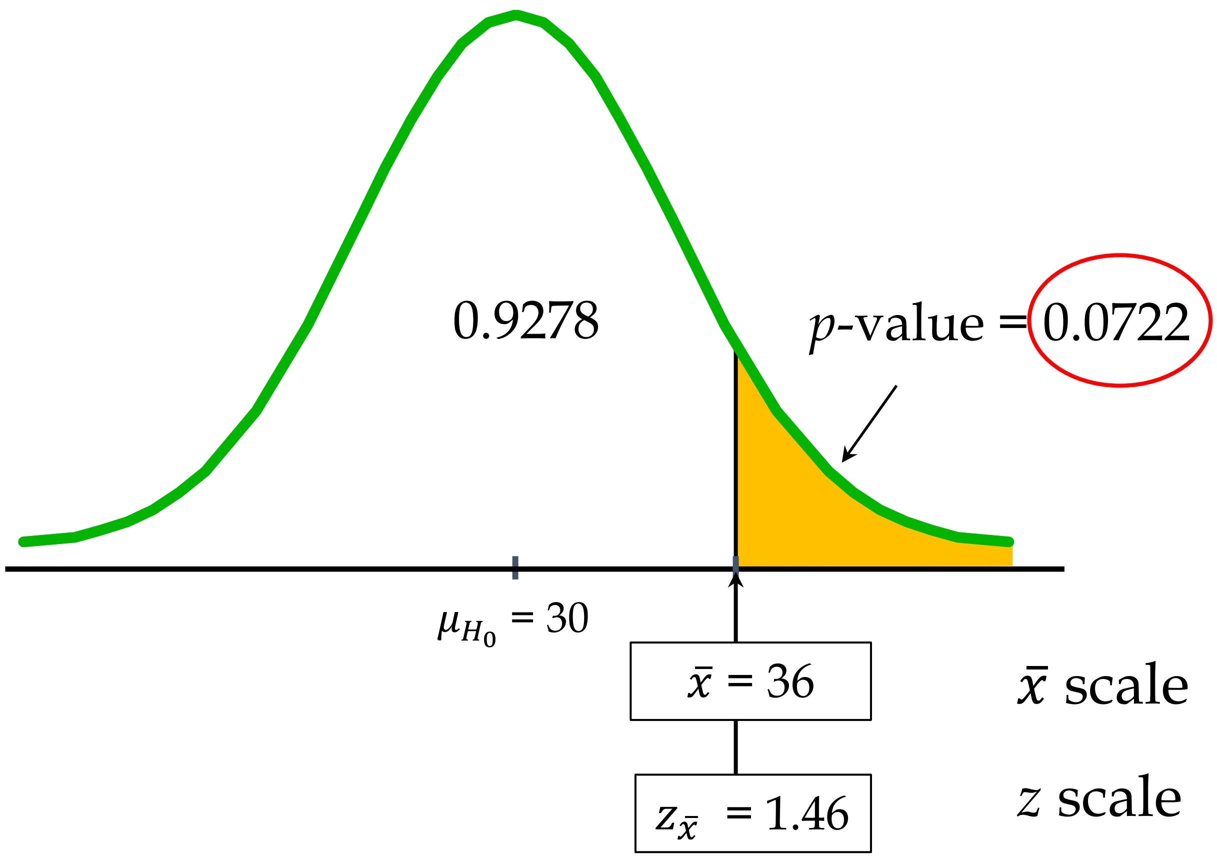
Step 4: Use the value of the test statistic to compute the *p*-value

The *p*-value represents the probability of observing an **average delivery time in the sample of 10 orders equal to 36 minutes or greater** if the overall average delivery time is 30 minutes

$$P(\bar{x} > 36) = P(z > 1.46) = 1 - 0.9278 = 0.0722$$

Note: the sign is important because this is an upper tail test and we are interested in the values more extreme than $\bar{x}$ (in the direction of the alternative) $\Rightarrow$ our interest is the upper tail of the distribution

One-Tail Hypothesis Test for the Population Mean: *p*-value Approach



Meaning of the *p*-value

Recall the definition…

The *p*-value is the probability of observing a sample mean *at least as extreme* as the one selected for the hypothesis test, *assuming the null hypothesis is true with equality*

In the direction of H_1

the sample mean

For our example…

p-value = 0.0722 is the probability of observing an average delivery time in the sample equal to 36 minutes **or greater** if the overall average delivery time is 30 minutes

One-Tail Hypothesis Test for the Population Mean: *p*-value Approach

Step 5: Compare the p-value and α

Decision Rule for Hypothesis Tests Using the p-value

$$p\text{-value} \geq \alpha \quad \Rightarrow \quad \text{Do not reject } H_0$$

$$p\text{-value} < \alpha \quad \Rightarrow \quad \text{Reject } H_0$$

Since p-value = 0.0722 > $\alpha = 0.05$, we **do not reject H_0**

Summary of Hypothesis Tests for the Population Mean (σ is Known)

- In all tests (one tail or two tail):
  - Testing procedure follows the same steps
  - The test statistic is computed in the same way
- However, depending on the type of the test:
  - Critical values and *p*-values are determined differently
  - Rejection rules are different for the critical value approach

Decision Rules for Hypothesis Tests: The Critical Value Approach

Test	**Hypothesis**	**Condition**		**Conclusion**
Two-tail	$H_0: \mu = \mu_{H_0}$	$\lvert z_{\bar{x}} \rvert > \lvert z_{\alpha/2} \rvert$	$\rightarrow$	Reject H_0
	$H_1: \mu \neq \mu_{H_0}$	$\lvert z_{\bar{x}} \rvert \leq \lvert z_{\alpha/2} \rvert$	$\rightarrow$	Do not reject H_0
One-tail	$H_0: \mu \leq \mu_{H_0}$	$z_{\bar{x}} > z_{\alpha}$	$\rightarrow$	Reject H_0
(upper)	$H_1: \mu > \mu_{H_0}$	$z_{\bar{x}} \leq z_{\alpha}$	$\rightarrow$	Do not reject H_0
One-tail	$H_0: \mu \geq \mu_{H_0}$	$z_{\bar{x}} < -z_{\alpha}$	$\rightarrow$	Reject H_0
(lower)	$H_1: \mu < \mu_{H_0}$	$z_{\bar{x}} \geq -z_{\alpha}$	$\rightarrow$	Do not reject H_0

p-value for Hypothesis Tests

Test	**Hypothesis**	**Calculation of the *p*-value**
One-tail *(upper)*	$H_0: \mu \leq \mu_{H_0}$ $H_1: \mu > \mu_{H_0}$	$P(z > z_{\bar{x}})$
One-tail *(lower)*	$H_0: \mu \geq \mu_{H_0}$ $H_1: \mu < \mu_{H_0}$	$P(z < z_{\bar{x}})$
Two-tail	$H_0: \mu = \mu_{H_0}$ $H_1: \mu \neq \mu_{H_0}$	$2 \times \; P(z > \lvert z_{\bar{x}} \rvert)$ or $2 \times \; P(z < -\lvert z_{\bar{x}} \rvert)$

Two-Tail Hypothesis Test for the Population Mean (σ Is Known)

Example: The mean data use for smartphone users is claimed to be $\mu = 1.8$ gigabytes per month

- Suppose data use is recorded for 49 randomly selected smartphone users and the average use is found to be 1.86 gigabytes per month
- Assume that the population standard deviation is 0.20 gigabytes per month
- Suppose that we are interested to test if there was **any** change in the data use

The following slides show **steps** to complete this test

Two-Tail Hypothesis Test for the Population Mean (σ Is Known)

Step 1: Identify the null and alternative hypotheses

H_0: $\mu = 1.8$ (status quo: average use is 1.8 G/month)

H_1: $\mu \neq 1.8$ (average use is not equal to 1.8 G/month)

Step 2: Set a value for the significance level, α

- Suppose that $\alpha = 0.05$ is chosen

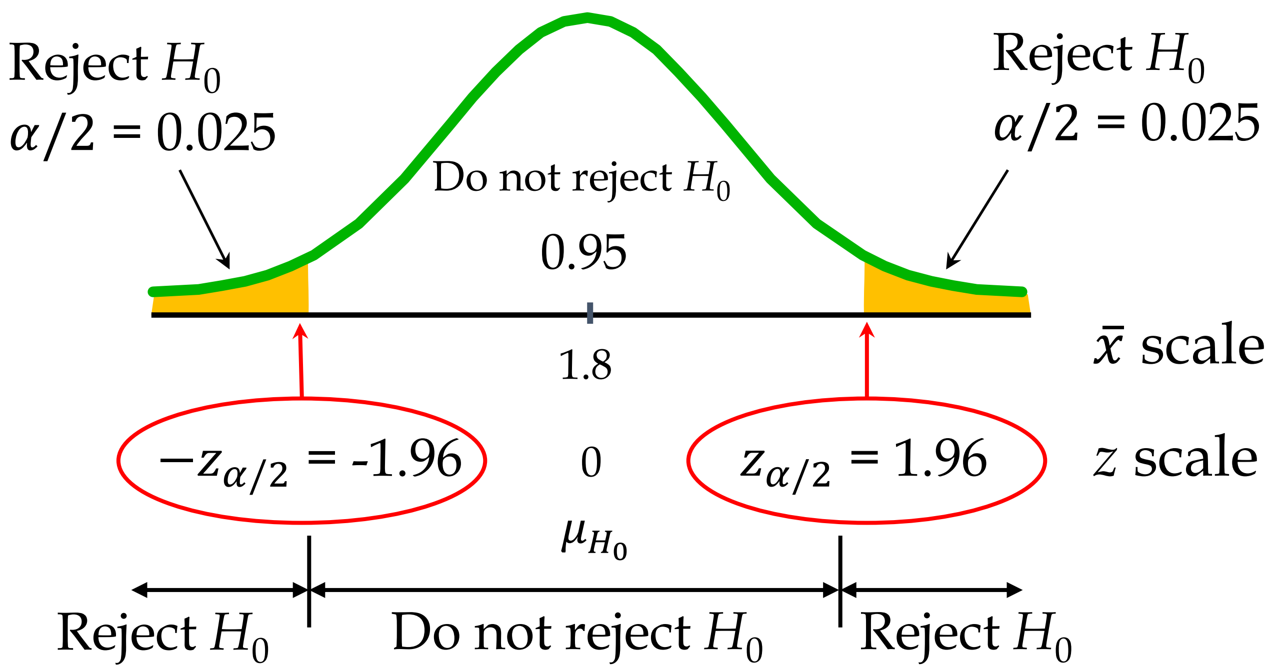
Step 3: Calculate the appropriate test statistic:

$$z_{\bar{x}} = \frac{\bar{x} - \mu_{H_0}}{\frac{\sigma}{\sqrt{n}}} = \frac{1.86 - 1.8}{\frac{0.20}{\sqrt{49}}} = \frac{0.06}{0.0286} = 2.10$$

Two-Tail Hypothesis Test for the Population Mean (σ Is Known)

Step 4: Determine the appropriate **critical value**

- σ is known $\Rightarrow$ use the **critical *z*-score**
- Since this is a two-tail test, $\alpha = 0.05$ is split evenly between two tails:



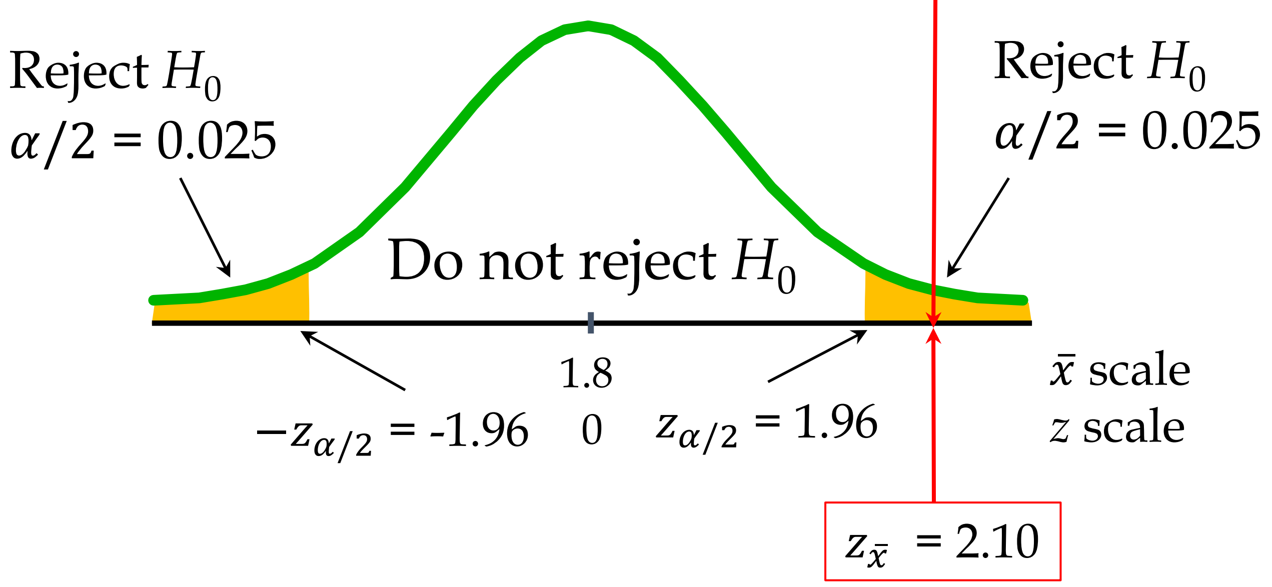
Two-Tail Hypothesis Test for the Population Mean (σ Is Known)

Step 5: Compare the test statistic $z_{\bar{x}}$ with the critical value

- For a two-tail test, reject the null hypothesis if

$$|z_{\bar{x}}| > |z_{\alpha/2}|$$

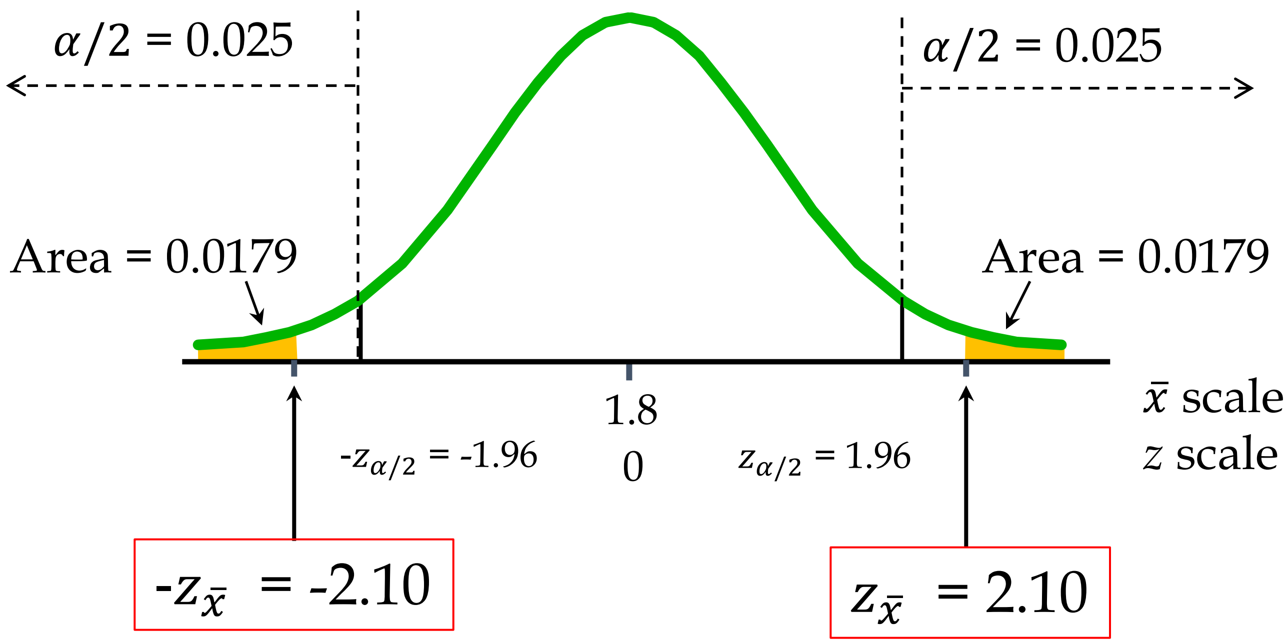
- Here, $z_{\bar{x}}$ = 2.10 is greater than 1.96 $\Longrightarrow$ **reject H_0**



Two-Tail Hypothesis Test for the Population Mean (σ Is Known)

Step 4: Determine ***p*-value**

- For a two-tail test, the p-value is a sum of two tail areas
- Double the area in one tail



$$p\text{-value} = 2 \times P(\bar{x} > 1.86) = 2 \times P(z > 2.10) = 2 \times 0.0179 = 0.0358$$

Two-Tail Hypothesis Test for the Population Mean (σ Is Known)

Step 5: Compare p-value and α

$$p\text{-value} < \alpha \Rightarrow \text{Reject } H_0$$

Conclusion

Our sample of 49 smartphone users provides a sufficient evidence to reject the null hypothesis. Therefore, we can support the alternative hypothesis that the average data use is different from 1.8 gigabytes per month

Excel Exercise 1 >>

Excel Exercise 2 >>

Excel Time: Exercise 9.5

Suppose the University of Memphis advertises that its average class size is 35 or fewer students. A student organization is concerned that budget cuts have led to increased class sizes and would like to test this claim. A random sample of 38 classes was selected, and the average class size was found 36.9 students. Assume that the standard deviation for the class size in the college is 8 students.

Using $\alpha = 0.01$, test the student organization's claim. To test the claim:

- Use the critical value approach
- Use the *p*-value approach
- Explain in your own words how Type I and Type II errors can occur in this hypothesis test

Back <<

Excel Time: Exercise 9.6

Suppose the coffee industry claimed that the average US adult drinks 1.7 cups of coffee per day. To test this claim, a random sample of 34 adults was selected. And their average coffee consumption was found to be 1.95 cups per day. Assume that the population standard deviation of daily coffee consumption is 0.5 cups.

Using $\alpha = 0.1$, can you test the claim of the coffee industry? To test the claim:

- Use the critical value approach
- Use the *p*-value approach

Back <<

Excel Time: *p*-value for One-Tail Tests

Two Excel's functions can be used to determine the *p*-value.
For the lower tail test:

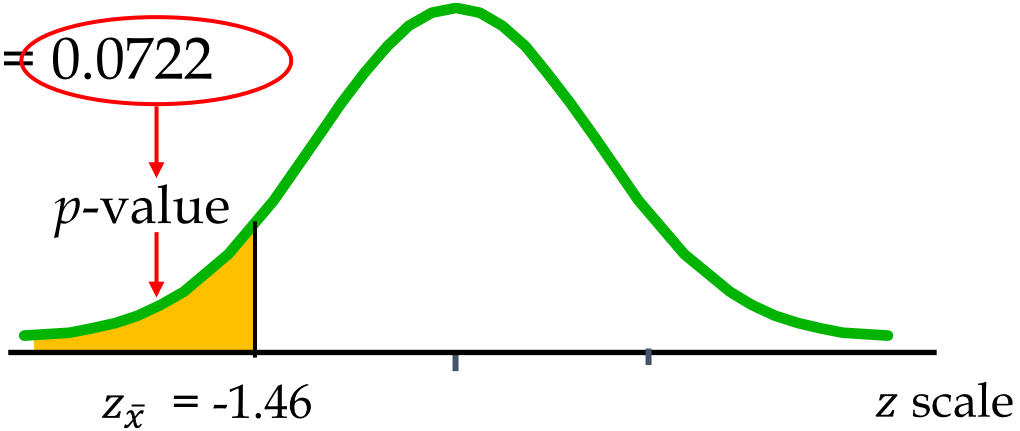
=NORM.S.DIST(*z*, cumulative)

or

=NORM.DIST(*z*, 0, 1, cumulative)

Use the test statistic for *z* and set cumulative = TRUE to get the area **to the left** of the test statistic
For example,
=NORM.S.DIST(-1.46, TRUE) = 0.0722

Because the *p*-value is the *area to the left* for the *lower tail test*, no further calculation is needed



Excel Time: *p*-value for One-Tail Tests

For the upper tail test:

= 1 - NORM.S.DIST(*z*, cumulative)

or

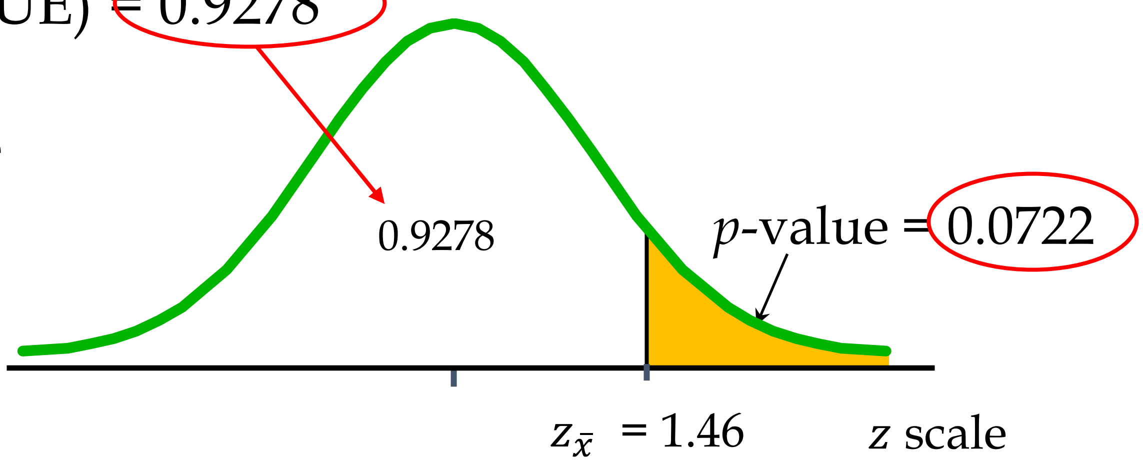
= 1 - NORM.DIST(*z*, 0, 1, cumulative)

Use the test statistic for *z* and set cumulative = TRUE to get the area **to the left** of the test statistic.

For example,

=NORM.S.DIST(1.46, TRUE) = 0.9278

Because the *p*-value is the *area to the right* of the test statistic for the *upper tail test*, we subtract 0.9278 from 1



Excel Time: *p*-value for Two-Tail Tests

To determine the *p*-value, we use either of two functions:

=NORM.S.DIST(*z*, cumulative)

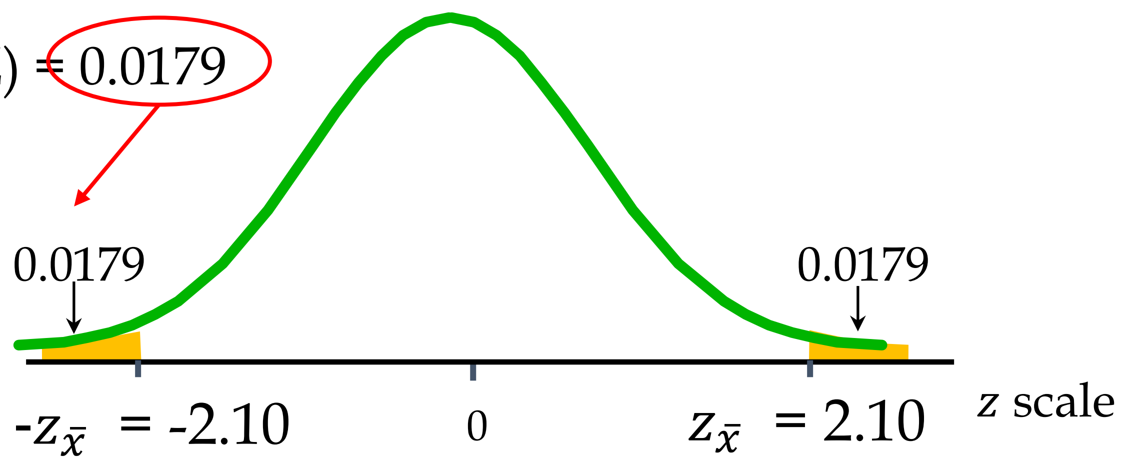
=NORM.DIST(*z*, 0, 1, cumulative)

Use the test statistic for *z* and set cumulative = TRUE to get the area to the left of the test statistic.

For example,

=NORM.S.DIST(-2.10, TRUE) = 0.0179

Because we need the ***area in both tails*** for a ***two-tail test***, if **$z < 0$**, multiply the result by 2. Here, multiply 0.0179 by 2 to get the *p*-value = 0.0358



Excel Time: *p*-value for Two-Tail Tests

p-value requires ***area in both tails*** for a ***two-tail test***:

- If $z > 0$, we
  1. Find the area on the right of the z test statistic subtracting the result of NORM.DIST() or NORM.S.DIST() from 1;
  2. Multiply the result by 2.

  For example, if $z_{\bar{x}} = 2.10$:

  - *p*-value = 2 × (1 – NORM.S.DIST(2.10)) = 0.0358

- If $z < 0$, we
  1. Find the area on the left of the z test statistic using NORM.DIST() or NORM.S.DIST();
  2. Multiply the result by 2.

  For example, if $z_{\bar{x}} = -2.10$:

  - *p*-value = 2 × NORM.S.DIST(-2.10) = 0.0358

Summary Excel Functions: Critical Value, I

Test	**Distribution**	**Value**	**Excel Function**
Two-tail	z	$z_{\alpha/2}$	= ABS(NORM.S.INV($\alpha/2$))
			= ABS(NORM.INV($\alpha/2$, 0, 1))
			= NORM.S.INV(1 - $\alpha/2$)
			= NORM.INV(1 - $\alpha/2$, 0, 1)
One-tail	z	z_{α}	= ABS(NORM.S.INV(α))
			= ABS(NORM.INV(α, 0, 1))
			= NORM.S.INV(1 - α)
			= NORM.INV(1 - α, 0, 1)

Summary of Excel Functions: *p*-value, I

Test	**Distribution**	***p*-Value**
Two-tail	z	= 2 × NORM.S.DIST(z, 1) if $z < 0$
		= 2 × (1-NORM.S.DIST(z, 1)) if $z > 0$
		= 2 × NORM.DIST(z, 0, 1, 1) if $z < 0$
		= 2 × (1-NORM.DIST(z, 0, 1, 1)) if $z > 0$
One-tail	z	= NORM.S.DIST(z, 1)
(*lower*)		= NORM.DIST(z, 0, 1, 1)
One-tail	z	= 1-NORM.S.DIST(z, 1)
(*upper*)		= 1-NORM.DIST(z, 0, 1, 1)

Excel Time: Exercise 9.7 (Extra Practice)

Season's Pizza recently hired additional drivers and as a result now claims that its average delivery time for the orders is under 45 minutes. A sample of 30 customer deliveries was examined, and the average delivery time was found to be 42.3 minutes. Historically, the standard deviation for delivery time was 11.6 minutes.

Using $\alpha = 0.03$, can you support the claim made by Season's Pizza? To test the claim:

- Use the critical value approach
- Use the *p*-value approach

Excel Time: Exercise 9.8 (Extra Practice)

Bob's Sporting Goods believes the average age of its customer is 40 or less. A random sample of 60 customers was surveyed, and the average customer ages was found to be 42.7 years. Assume the population standard deviation for customer age is 8.0 years.

Using $\alpha = 0.02$, does this sample provides enough evidence to **challenge** Bob's Sporting Goods belief? To test the claim:

- Use the critical value approach
- Use the *p*-value approach

Excel Time: Exercise 9.43 (Extra Practice)

According to statista.com, the average room rate for a New York City hotel was \$337 in 2016. Suppose the Chamber of Commerce of New York City would like to test if this rate has changed recently by randomly sampling 40 room rates. The mean of this sample was found to be \$351.20. Assume the population standard deviation is \$55.

Using $\alpha = 0.08$, perform a hypothesis test to help the Chamber of Commerce of New York City answer their question:

- Use the critical value approach
- Use the *p*-value approach

Excel Time: Exercise 9.59 (Extra Practice)

The average wait time on the phone for taxpayers calling the IRS in 2015 was 1,380 seconds. Suppose that the IRS made operational changes in an effort to reduce wait times. To test the effectiveness of these changes, a random sample of 50 phone calls was selected and the wait time of each call was recorded. The average wait time in the sample was calculated to be 1,250 seconds. Assume the population standard deviation for the wait time is 250 seconds.

Using $\alpha = 0.01$, perform a hypothesis test to determine if the changes introduced by the IRS were effective:

- Use the critical value approach
- Use the *p*-value approach