Hypothesis Tests

Introduction

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

- The hypothesis testing procedure uses data from a sample to test the two competing statements indicated by H_0 and H_a.
- The **null hypothesis**, denoted by H_0, is a tentative assumption about a population parameter.
- The **alternative hypothesis**, denoted by H_a, is the opposite of what is stated in the null hypothesis.

This chapter shows how hypothesis tests can be conducted about a population mean and a population proportion.

We begin by providing examples that illustrate approaches to developing null and alternative hypotheses.

Developing Null and Alternative Hypotheses

Care must be taken to understand the context of the situation and structure the hypotheses appropriately so that the test conclusion provides the desired information.

All hypothesis testing applications involve collecting a sample and using the sample results to provide evidence for drawing a conclusion.

Good questions to consider when formulating the null and alternative hypotheses are:

- What is the purpose of collecting the sample?
- What conclusions are we hoping to make?

In some situations, it is easier to identify the alternative hypothesis first, and then the null hypothesis. In other situations, starting with the null hypothesis is easier.

The examples in this section are intended to provide guidelines on how to formulate correct hypotheses.

The Alternative Hypothesis as a Research Hypothesis

When an attempt is being made to gather evidence in support of a research hypothesis, 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 sufficient evidence to show that the null hypothesis can be rejected.

Example: A new fuel injection unit being developed is believed to provide more than 24 miles per gallon.

So, we set the population mean miles per gallon, $\mu > 24$, as the alternative hypothesis, and the null hypothesis as $\mu \leq 24$.

$$H_0\colon \mu \leq 24$$

$$H_a\colon \mu > 24$$

The Null Hypothesis as an Assumption to be Challenged

We might begin with a belief or assumption that a statement about the value of a population parameter is true.

We then use a hypothesis test to challenge the assumption and determine if there is statistical evidence to conclude that the assumption is incorrect.

In these situations, it is helpful to develop the null hypothesis first.

Example: The label on a soft drink bottle states that it contains 67.6 fluid ounces.

We assume the label correct, provided the population mean filling weight, μ, is 67.6 fluid ounces, and set it as the null hypothesis. Then, we challenge the assumption in the alternative hypothesis.

$$H_0\text{: } \mu = 67.6$$

$$H_a\text{: } \mu \neq 67.6$$

Summary of Forms for Null and Alternative Hypotheses

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 three forms, where μ_0 is the hypothesized value of the population mean.

One-tailed (Lower-tail)	One-tailed (Upper-tail)	Two-tailed
H_0: $\mu \geq \mu_0$	H_0: $\mu \leq \mu_0$	H_0: $\mu = \mu_0$
H_a: $\mu < \mu_0$	H_a: $\mu > \mu_0$	H_a: $\mu \neq \mu_0$

In this chapter, we show how statistical inference from sample results can be used together with hypothesis testing to determine whether a statement about the value of a population parameter should or should not be rejected.

If the sample results support H_a, then we reject H_0. Otherwise, we do not reject H_0.

Type I and Type II Errors

Because hypothesis tests are based on sample data, we must allow for the possibility of errors.

There are two types of error that can be made in hypothesis tests: Type I and Type II errors.

Type I Error

A Type I error is rejecting H_0 when it is true.

The probability of making a Type I error when the null hypothesis is true as an equality is called the **level of significance**.

Applications of hypothesis testing that only control for the Type I error are often called significance tests.

Type II Error

A Type II error is accepting H_0 when it is false.

It is difficult to control for the probability of making a Type II error.

Statisticians avoid the risk of making a Type II error by using “do not reject H_0” rather than “accept H_0”.

Example of Type I and Type II Errors

Recall the hypothesis testing example about the new fuel injection unit.

H_0: $\mu \leq 24$

H_a: $\mu > 24$

In this situation, the *Type I error* of rejecting H_0 when true corresponds to claiming that the new system improves the miles-per-gallon rating ($\mu > 24$) when in fact it does not.

In contrast, the *Type II error* of accepting H_0 when false corresponds to claiming that the new system does not improve the miles-per-gallon rating ($\mu \leq 24$) when in fact it does.

		Population Condition	
		H_0 True	**H_a True**
Conclusion	**Accept H_o**	Correct Conclusion	Type II Error
	Reject H_o	Type I Error	Correct Conclusion

One-Tailed Test About a Population Mean

One-tailed tests about a population mean take one of the following two forms.

Lower Tail Test	**Upper Tail Test**
$H_0\colon \mu \geq \mu_0$	$H_0\colon \mu \leq \mu_0$
$H_a\colon \mu < \mu_0$	$H_a\colon \mu > \mu_0$

Consider the example of the Federal Trade Commission (FTC) testing a Hilltop Coffee's claim that its large can contains at least 3 pounds of coffee.

The FTC assumes Hilltop Coffee's claim to be correct ($\mu \geq 3$ pounds) and sets the alternative hypothesis to challenge the claim, using the *lower tail test*, $\mu < 3$ pounds. Thus, we write:

$$H_0\colon \mu \geq 3$$

$$H_a\colon \mu < 3$$

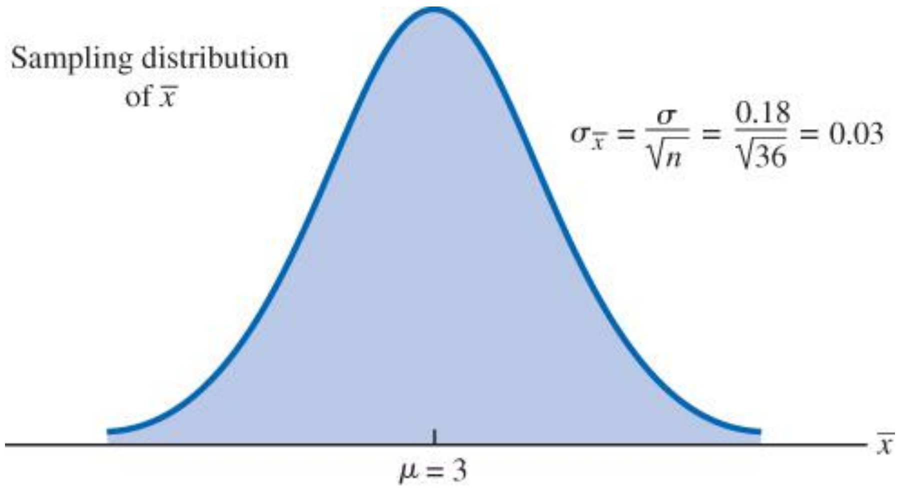
The FTC is willing to take a 1% risk of making an error on its testing of Hilltop Coffee's claim.

Test Statistic for the Hilltop Coffee Example (Population Mean: σ Known)

Suppose a sample of $n = 36$ cans of coffee is selected, and the sample mean $\bar{x}$ is computed as an estimate of the population mean μ.

Also, previous FTC tests show that the population of filling weights can be assumed normally distributed with a known standard deviation of value $\sigma = 0.18$.

We can write the sampling distribution of $\bar{x}$ as normally distributed, centered around the hypothesized mean, $\mu = \mu_0 = 3$, and with standard error, $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 0.18/\sqrt{36} = 0.03$.



For the **test statistic** of a hypothesis test about a population mean: σ known, we can use the standard normal random variable, z.

Formally, we write:

$$\boldsymbol{z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}}$$

p-Value Approach to a Lower Tail Test About a Population Mean: σ Known

A ***p*-value** is a probability, computed using the test statistic, that measures the support provided by the sample for the null hypothesis.

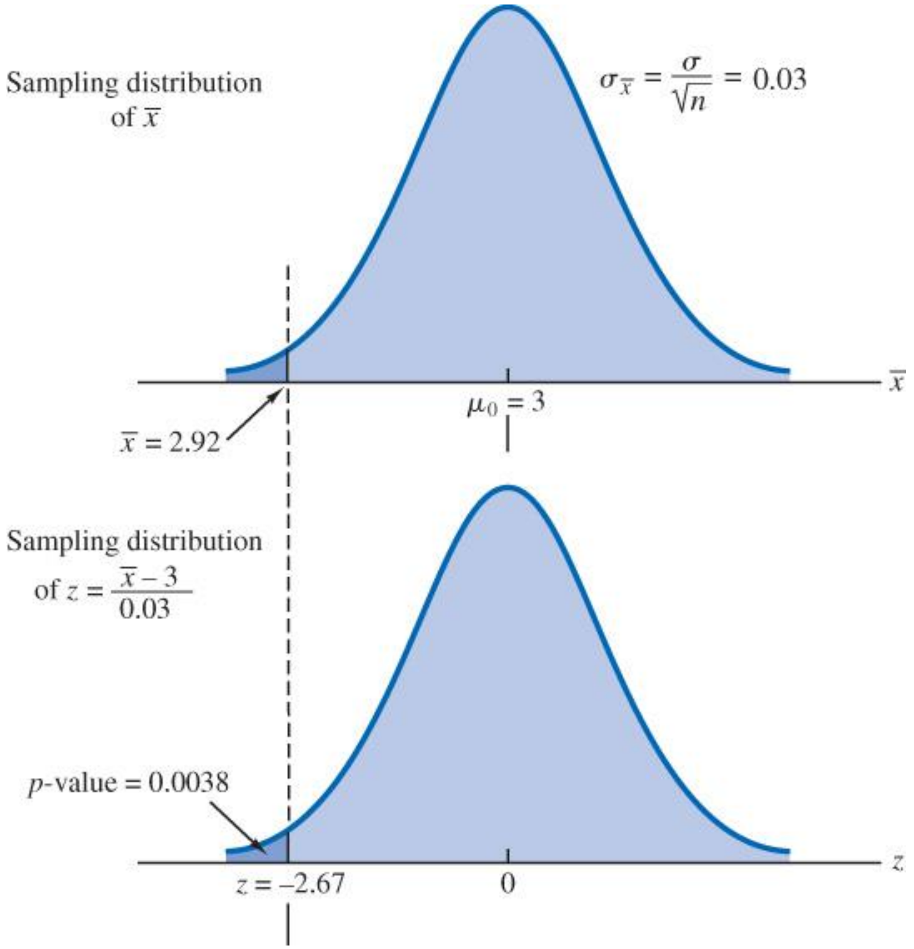
Suppose the sample of $n = 36$ coffee cans revealed $\bar{x} = 2.92$ pounds. Then, the test statistic is

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{2.92 - 3}{0.03} = -2.67$$

For a lower tail test, the *p*-value is the probability that *z* is less than or equal to −2.67.

$$p\text{−value} = P(z \leq -2.67) = 0.038$$

This *p*-value indicates a small probability of obtaining a sample mean of $\bar{x} = 2.92$ or smaller when sampling from a population with $\mu = 3$.



Rejection Rule Using *p*-Value

In the previous slide, we concluded that the calculated *p*-value does not provide much support for the null hypothesis, but is it small enough to cause us to reject H_0?

The answer depends upon the level of significance α for the test:

> If a *p*-value is less than or equal to the level of significance α, the value of the test statistic is in the rejection region.

Because the FTC stated to be willing to take a 1% risk of making a Type I error, we set the significance level $\alpha = 0.01$.

We write the rejection rule using the *p*-value as:

> **Reject $\boldsymbol{H_0}$ if *p*-value ≤ *α***

With the p−value $= 0.0038$ being less than or equal to $\alpha = \ 0.01$, we reject H_0.

Therefore, the FTC finds sufficient statistical evidence to conclude at the 0.01 level of significance that the filling weight of Hilltop Coffee cans is less than 3 pounds.

Rejection Rule Using Critical Value for a One-Tailed Test

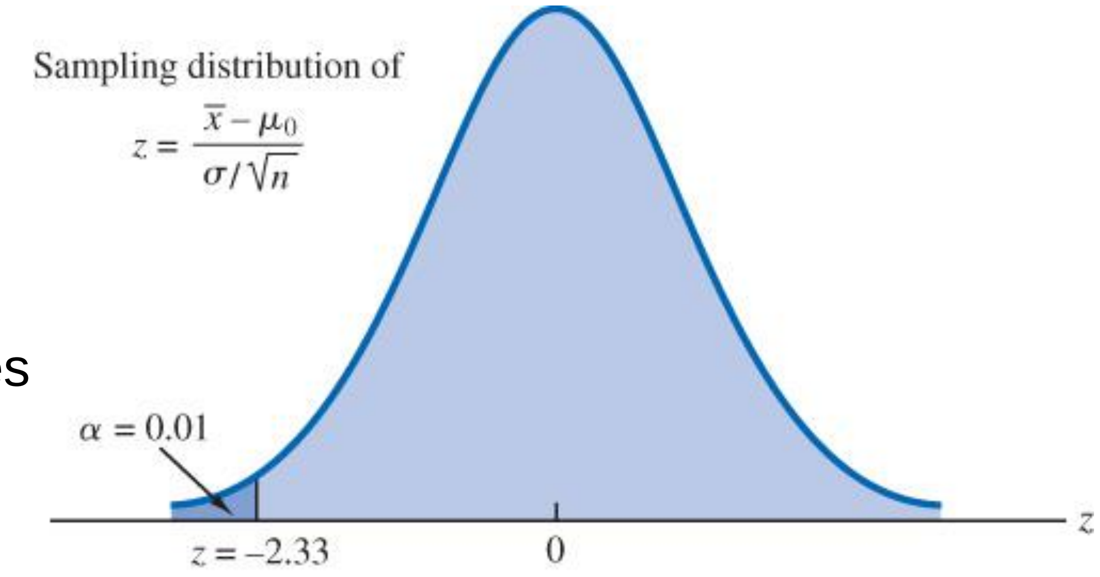
We can use the standard normal probability distribution to find the z-value with an area of α in the lower (or upper) tail of the distribution.

The value of the test statistic that establishes the boundary of the rejection region is called the **critical value** for the test.

The rejection rule for a one-tailed test using the critical value is:

Lower tail: Reject $\boldsymbol{H_0}$ if $\boldsymbol{z \leq z_\alpha}$

Upper tail: Reject $\boldsymbol{H_0}$ if $\boldsymbol{z \geq z_\alpha}$



z_α is such that $P(z \leq z_\alpha) = \alpha$. In the Hilltop Coffee case, when $\alpha = 0.01$, $z_\alpha = -2.33$.

The rejection rule for a lower tail test becomes

Reject H_0 if $z \leq -2.33$

Because $z = -2.67$, we reject H_0.

Two-Tailed Test About a Population Mean

A **two-tailed test** about a population mean takes the following form.

$$H_0: \mu = \mu_0$$

$$H_a: \mu \neq \mu_0$$

As an example, consider MaxFlight's need of testing its manufacturing process to ensure it produces golf balls with a mean driving distance of 295 yards.

If the golf balls drive less than 295 yards, MaxFlight will eventually lose sales, and if the golf balls' driving distance exceeds 295 yards, they may be rejected by the USGA.

Thus, we assume that the mean driving distance is as specified by MaxFlight, and challenge it in the alternative hypothesis, by writing:

$$H_0: \mu = 295$$

$$H_a: \mu \neq 295$$

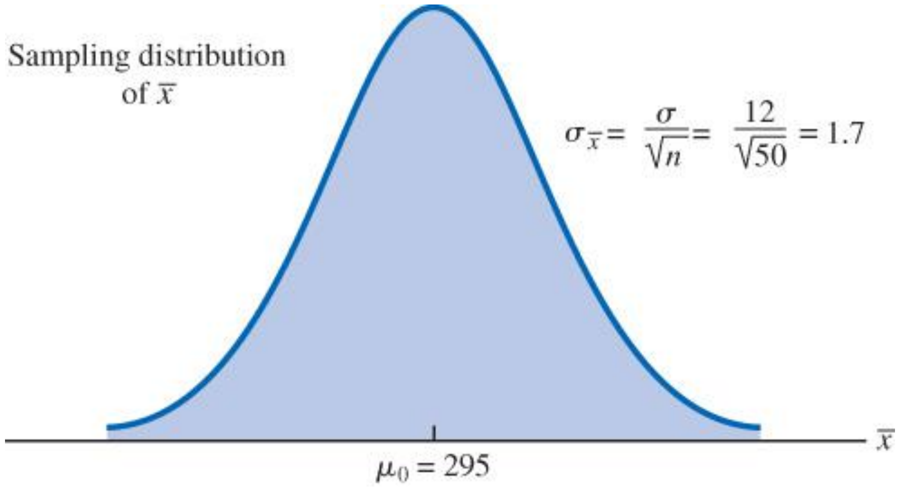
The quality control team selected $\alpha = 0.05$ as the level of significance for the test.

Test Statistic for the MaxFlight Example (Population Mean: σ Known)

Suppose a sample of $n = 50$ golf balls is selected, and the sample mean $\bar{x} = 297.6$ is computed as an estimate of the population mean μ.

Data from previous tests conducted when the process was known to be in adjustment show that the population standard deviation can be assumed known with a value of $\sigma = 12$.

We can write the sampling distribution of $\bar{x}$ as normally distributed, centered around the hypothesized mean, $\mu = \mu_0 = 295$, and with standard error, $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 12/\sqrt{50} = 1.7$.



We write the test statistic of a hypothesis test about a population mean: σ known, using the standard normal random variable, z, *as:*

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{297.6 - 295}{1.7} = 1.53$$

p-Value Approach to a Two-Tailed Test

For the MaxFlight two-tailed test, the *p*-value corresponds to the probability of finding a value for *z* that is at least as unlikely as $z = 1.53$.

For a two-tailed test of a symmetric sampling distribution, values of $z \geq 1.53$ are just as unlikely as values of $z \leq -1.53$. Thus, we can write the *p*-value as:

$$p\text{−value} = P(z \leq -1.53) + P(z \geq 1.53)$$

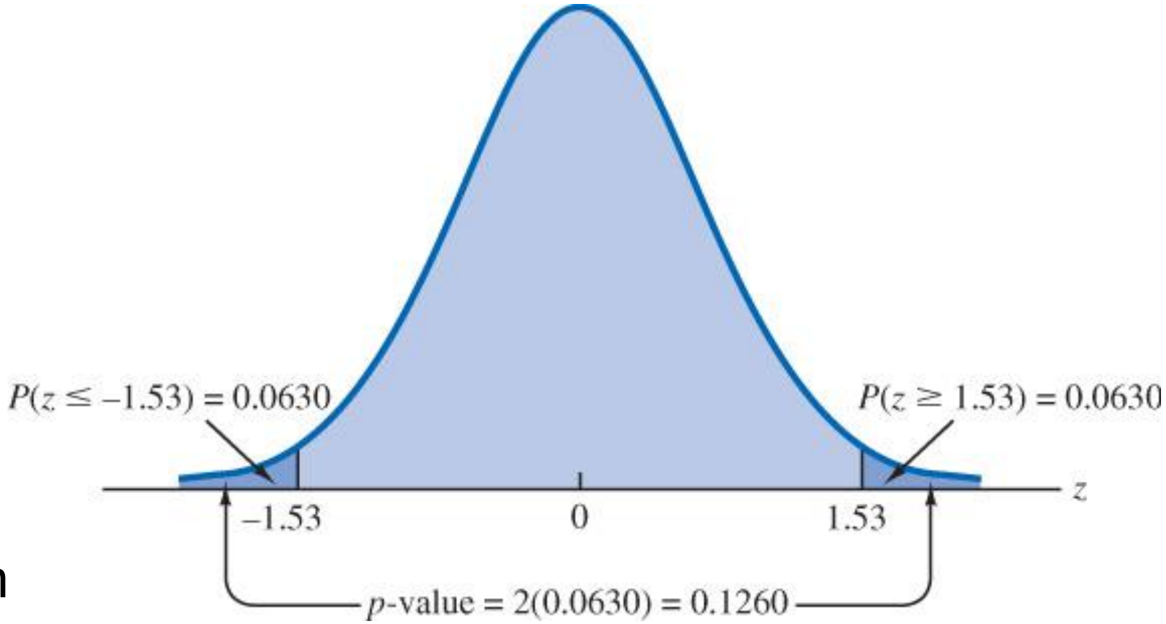
Because the normal distribution is symmetric, we can calculate the p-value as

$$p\text{−value} = 2P(z \leq -1.53) =$$

$$2(0.0630) = 0.1260$$

Next, we apply the rejection rule:

Because $p\text{−value} = 0.1260 > 0.05$, we cannot reject H_0. Thus, no action will be taken to adjust the MaxFlight's manufacturing process.



Rejection Rule Using Critical Value for a Two-Tailed Test

The rejection rule for a two-tailed test using the critical value is:

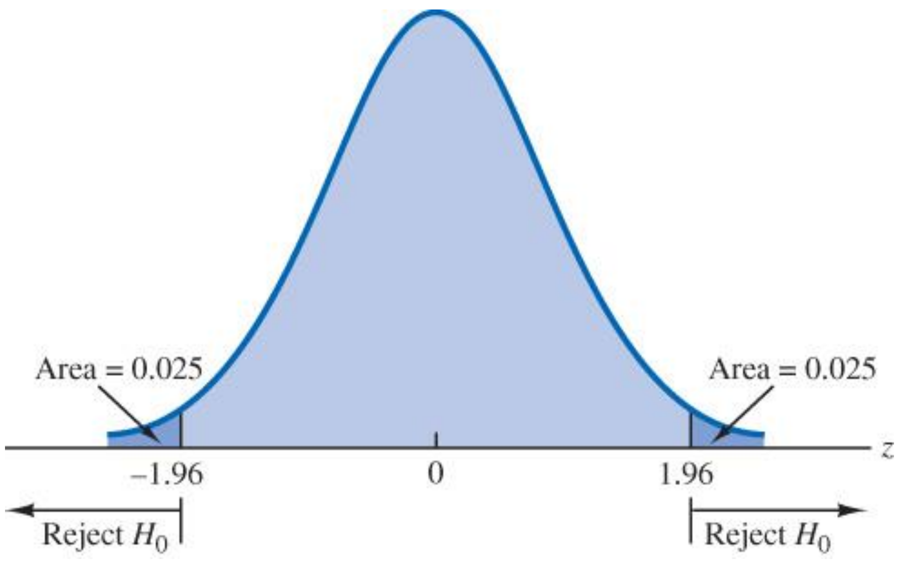
Reject $\boldsymbol{H_0}$ if $\mathbf{z} \leq -\boldsymbol{z_{\alpha/2}}$ or $\mathbf{z} \geq \boldsymbol{z_{\alpha/2}}$

In the MaxFlight case, we have $\alpha = 0.05$.

Thus, $\alpha/2 = 0.025$, and $z_{\alpha/2} = z_{.025} = 1.96$.

The rejection rule for this two-tail test becomes

Reject H_0 if $z \leq -1.96$ or $z \geq 1.96$.



Because the value of the test statistic for the MaxFlight study is $z = 1.53$, the statistical evidence will not permit us to reject the null hypothesis at the 0.05 level of significance.

Summary of Hypothesis Tests About a Population Mean: σ Known Case

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0: \mu \geq \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu \leq \mu_0$ $H_a: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$
Test Statistic	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
Rejection Rule: *p*-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $z \leq -z_\alpha$	Reject H_0 if $z \geq z_\alpha$	Reject H_0 if $z \leq -z_{\alpha/2}$ or $z \geq z_{\alpha/2}$

A Confidence Interval Approach to the Testing of a Hypothesis

First, let us use the sample statistics for the MaxFlight example to develop a confidence interval for the population mean (σ known case)

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 297.6 \pm 1.96\frac{12}{\sqrt{50}} = 297.6 \pm 3.3$$

We are 95% confident that the mean distance for the population of golf balls is between $297.6 - 3.3 = 294.3$ and $297.6 + 3.3 = 300.9$ yards.

Recollect how the MaxFlight hypothesis test takes the following form:

$$H_0\colon \mu = 295$$

$$H_a\colon \mu \neq 295$$

Because the confidence interval contains the hypothesized population mean, $\mu_0 = 295$, the null hypothesis cannot be rejected.

A One-Tailed Test Example About a Population Mean

DATAFile: *AirRating*

A business travel magazine wants to classify London's Heathrow Airport according to the mean rating for the population of business travelers.

A rating scale with a score from 0 to 10 is used, and if the population mean rating is greater than 7, Heathrow Airport will be designated as a superior service airport.

We set the alternative hypothesis to $\mu > 7$, so that the rejection of the null hypothesis leads to the conclusion that Heathrow should be designated as a superior service airport.

$$H_0\colon \mu \leq 7$$

$$H_a\colon \mu > 7$$

We use $\alpha = 0.05$ as the level of significance for the test.

Test Statistic for the Air Ratings Example (Population Mean: σ Unknown)

In Section 8.2 we learned that when we use the sample standard deviation, s, as an estimate of the population standard deviation, σ, the sampling distribution of $\bar{x}$ follows a t distribution with $n - 1$ degrees of freedom.

The test statistic for hypothesis tests about a population mean when σ is unknown is

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

The surveyed sample of $n = 60$ business travelers at Heathrow Airport provided sample mean rating $\bar{x} = 7.25$, and sample standard deviation $s = 1.052$. Thus, we have

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{7.25 - 7}{1.052/\sqrt{60}} = 1.84$$

The *p*-value is in the upper tail of a t distribution with $n - 1 = 59$ degrees of freedom.

$$p\text{−value} = P(t \geq 1.84)$$

p-Value Approach to a One-Tailed Test About a Population Mean: σ Unknown

The t distribution table included in most textbooks will only provide information about a range of *p*-values.

For instance, using Table 2 in Appendix B, the t distribution with 59 degrees of freedom provides the following information.

Area in upper tail	**0.20**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
***t* Value ($df = 59$)**	0.848	1.296	1.671	2.001	2.391	2.662

Because $t = 1.84$ is between 1.671 and 2.001, the values in the "Area in Upper Tail" row show that the *p*-value range must be: $0.025 \le p\text{–value} \le 0.05$.

We can also use Excel to compute the exact *p*-value as: $=\text{T.DIST.RT}(1.84{,}59) = 0.0354$

Either approach allows us to make the decision to reject the null hypothesis and conclude that Heathrow should be classified as a superior service airport.

Critical Value Approach to a One-Tailed Test About a Population Mean: σ Unknown

The decision of whether to reject the null hypothesis in the σ unknown case can also be made using the critical value approach.

We use the same information from Table 2 in Appendix B we just considered for the *p*-value approach.

Area in upper tail	**0.20**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
***t* Value ($df = 59$)**	0.848	1.296	1.671	2.001	2.391	2.662

The critical value corresponding to an area of $\alpha = 0.05$ in the upper tail of a t distribution with 59 degrees of freedom is $t_{0.05} = 1.671$.

We can also use Excel to compute the critical value as: $=\mathrm{T.INV}(1 - 0.05{,}59) = 1.671$

Thus, because the rejection rule in this example is $t = 1.84 \geq 1.671 = t_{0.05}$, we can conclude that Heathrow should be classified as a superior service airport.

Two-Tailed Test Example About a Population Mean

DATAFile: *Orders*

Serrano Toys' marketing director is expecting demand for this year's most important toy to average 40 units per retail outlet.

Serrano's marketing director decided to survey a sample of 25 retailers to obtain information about the demand for the new product before making the final production decision.

We set the null hypothesis to $\mu = 40$, so that its rejection leads to the conclusion that the expectation about the toy's demand is incorrect.

$$H_0\colon \mu = 40$$

$$H_a\colon \mu \neq 40$$

We use $\alpha = 0.05$ as the level of significance for the test.

The surveyed sample of $n = 25$ retail stores provided a sample mean rating $\bar{x} = 37.4$, and a sample standard deviation $s = 11.79$.

p-Value Approach to a Two-Tailed Test About a Population Mean: σ Unknown

The test statistic for the population mean, σ unknown case is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{37.4 - 40}{11.79/\sqrt{25}} = -1.10$$

Because we have a two-tailed test, the *p*-value is two times the area under the curve of the t distribution for $t \leq -1.10$, and with $n - 1 = 24$ degrees of freedom.

Using Table 2 in Appendix B, the *t* distribution table for 24 degrees of freedom provides

Area in upper tail	**0.20**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
***t* Value ($df = 24$)**	0.857	1.318	1.711	2.064	2.492	2.797

Because the t distribution is symmetric, the upper tail at $t = 1.10$ is the same as the lower tail at $t = -1.10$. Thus, the *p*-value is twice the 0.10 to 0.20 range, or $0.20 \leq p\text{–value} \leq 0.40$.

Using Excel, the exact *p*-value is $2P(t \leq -1.10) = 2 \times \text{T.DIST.RT}(-1.10{,}24) = 0.2822$

Critical Value Approach to a Two-Tailed Test About a Population Mean: σ Unknown

For a two-tail test, σ unknown case, Table 2 in Appendix B provides the critical value with 24 degrees of freedom as

Area in upper tail	**0.20**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
***t* Value ($df = 24$)**	0.857	1.318	1.711	2.064	2.492	2.797

Thus, the two critical values for a two-tailed test are $-t_{.025} = -2.064$ and $t_{.025} = 2.064$.

We can also use Excel to compute the critical values as: $=\mathrm{T.INV}(1-.025,24) = 2.064$

Using either approach, we have the following rejection rules:

p-value approach: Reject H_0 if $p-value \leq 0.05$

critical value approach: Reject H_0 if $t \leq -2.064$ or $t \geq 2.064$

Thus, H_0 cannot be rejected, and Serrano should continue its production planning for the coming season based on an expected demand of 40 toys per retail store.

Summary of Hypothesis Tests About a Population Mean: σ Unknown Case

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0\colon\ \mu \geq \mu_0$ $H_a\colon\ \mu < \mu_0$	$H_0\colon\ \mu \leq \mu_0$ $H_a\colon\ \mu > \mu_0$	$H_0\colon\ \mu = \mu_0$ $H_a\colon\ \mu \neq \mu_0$
Test Statistic	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
Rejection Rule: *p*-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $t \leq -t_\alpha$	Reject H_0 if $t \geq t_\alpha$	Reject H_0 if $t \leq -t_{\alpha/2}$ or $t \geq t_{\alpha/2}$

Hypothesis Test About a Population Proportion

A hypothesis test about a population proportion, *p*, takes one of the following three forms.

Lower Tail Test	**Upper Tail Test**	**Two-Tailed Test**
$H_0: p \geq p_0$	$H_0: p \leq p_0$	$H_0: p = p_0$
$H_a: p < p_0$	$H_a: p > p_0$	$H_a: p \neq p_0$

As an application of a hypothesis test about a population proportion, consider the following.

The park manager at Knoebels Amusement Park wants to determine whether a targeted online marketing campaign has increased non-local participation from a historical 20%.

Thus, we write:

$$H_0: p \leq 0.20$$

$$H_a: p > 0.20$$

The park manager specified that a level of significance of $\alpha = 0.05$ be used for this test.

Test Statistic for a Population Proportion

Suppose a random sample of $n = 400$ park visitors was selected, and that $x = 100$ of the visitors were non-local. The proportion of non-local visitors in the sample is

$$\bar{p} = \frac{x}{n} = \frac{100}{400} = 0.25$$

For the test statistic of a hypothesis test about a population proportion, we can use the standard normal random variable, z, to describe the sampling distribution of $\bar{p}$ so long as we have at least five expected favorable outcomes, $np_0 \geq 5$, and five expected unfavorable outcomes, $n(1 - p_0) \geq 5$.

In this example, we have: $np_0 = 100(0.2) = 20$, and $n(1 - p_0) = 100(1 - 0.2) = 80$

Thus, we can write the test statistic as:

$$\boldsymbol{z = \frac{\bar{p} - p_0}{\sigma_{\bar{p}}} = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}} = \frac{0.25 - 0.20}{\sqrt{\frac{0.20(1 - 0.80)}{400}}} = \frac{0.05}{0.02} = 2.50$$

p-Value Approach to a One-Tailed Test About a Population Proportion

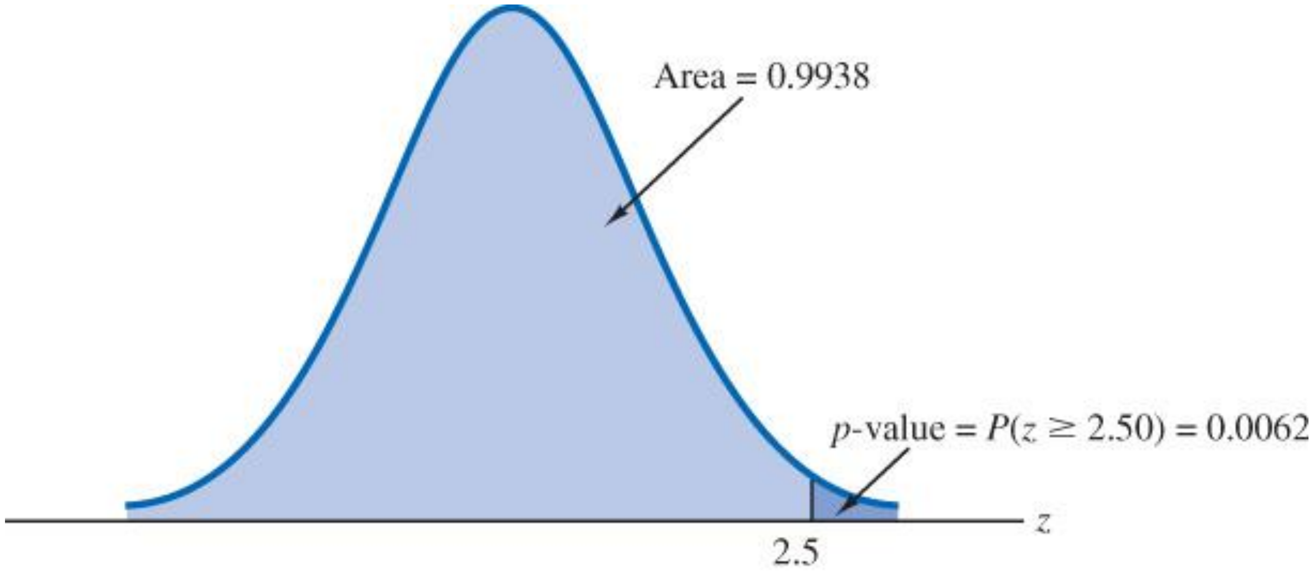
We can use a statistical software package or the standard normal probability table to calculate the *p*-value for an upper tail test as

$$p\text{−value} = P(z \geq 2.50) = 1 - P(z < 2.50) = 1 - 0.9938 = 0.0062$$

Recollect the rejection rule for the *p*-value approach:

Reject H_0 if $p\text{−value} \leq \alpha$

Because $0.0062 \leq 0.05$, we reject H_0 and conclude that the proportion of non-local visitors to Knoebels Amusement Park has increased.



Critical Value Approach to a One-Tailed Test About a Population Proportion

We can use a statistical software package or the standard normal probability table to also calculate the critical value for an upper tail test, z_α, such that

$$P(z \geq z_\alpha) = \alpha$$

Recollect how, for this example, the park manager set the significance level to $\alpha = 0.05$.

Thus, we have

$$z_{0.05} = 1.645.$$

Recollect also the rejection rule for the critical value approach to an upper-tail test:

Reject H_0 if $z \geq z_\alpha$

Because the test statistic, $z = 2.50 \geq 1.645$, we reject H_0 and conclude that the proportion of non-local visitors to Knoebels Amusement Park has increased.

Summary of Hypothesis Tests About a Population Proportion

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0: p \geq p_0$ $H_a: p < p_0$	$H_0: p \geq p_0$ $H_a: p < p_0$	$H_0: p \geq p_0$ $H_a: p < p_0$
Test Statistic	$z = \dfrac{\bar{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$	$z = \dfrac{\bar{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$	$z = \dfrac{\bar{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$
Rejection Rule: *p*-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $z \leq -z_\alpha$	Reject H_0 if $z \geq z_\alpha$	Reject H_0 if $z \leq -z_{\alpha/2}$ or $z \geq z_{\alpha/2}$

Hypothesis Testing and Decision-Making

In the previous sections, we learned to compare the p-value to α, the level of significance for the test, which is the controlled probability of a Type I error.

If p-value $\leq \alpha$, we made the conclusion "reject H_0" and declared the results significant.

If p-value $> \alpha$, we made the conclusion "do not reject H_0" and declared the statistical evidence not conclusive.

However, with a significance test, we control the probability of a Type I error, but not the probability of a Type II error. Hence the reason why we recommended the conclusion "do not reject H_0" rather than "accept H_0".

When two different decisions must be made depending on the test outcome, statisticians recommend controlling the probabilities of committing a Type I as well as a Type II error.

With the probability of a Type II error under control, the conclusion "accept H_0" rather than "do not reject H_0" can be made.

A Lot-Acceptance Sampling Example

Consider this application of hypothesis testing for decision-making to lot-acceptance sampling.

A quality control manager must decide to accept a shipment of batteries from a supplier if the batteries have a mean useful life of at least 120 hours or to return the shipment if they do not.

The null and alternative hypotheses can be set as follows.

$$H_0\colon \mu \geq 120$$

$$H_a\colon \mu < 120$$

Where μ denotes the mean number of hours of useful life for batteries in the shipment.

To evaluate the quality of a new shipment, a sample of $n = 36$ batteries is selected and tested.

If H_0 is rejected, H_a is declared true and the shipment of batteries is returned to the supplier.

But if H_0 is not rejected, the quality control manager needs to control the probability of a Type II error before the decision is made to declare H_0 true and accept the new shipment of batteries.

Calculating the Sample Mean as a Rejection Rule

Suppose a level of significance, $\alpha = 0.05$ is used to conduct the hypothesis test.

If based upon previous testing, the population standard deviation can be assumed known with a value of $\sigma = 12$ hours, the test statistic for the sample of $n = 36$ batteries is

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 120}{12/\sqrt{36}} = \frac{\bar{x} - 120}{2}$$

The rejection rule for the critical value approach to a lower tail test is to reject H_0 if $z \leq -1.645$.

Thus, the rejection rule indicates that we will reject H_0 if

$$z = \frac{\bar{x} - 120}{2} \leq -1.645$$ Solving for $\bar{x}$ yields: $\bar{x} \leq 120 - 1.645(2) = 116.71$

If we reject H_0 when $\bar{x} \leq 116.71$, it means that we will accept the shipment when $\bar{x} > 116.71$.

We are now ready to compute probabilities associated with making a Type II error.

Calculating the Probability of Type II Errors

Suppose the shipment is of poor quality if the batteries have a mean life of $\mu = 112$ hours.

Because we make a Type II error when we accept a false H_0, let us calculate the probability that $\bar{x} > 116.71$ and conclude that $\mu \geq 120$ when the true mean battery life is $\mu = 112$.

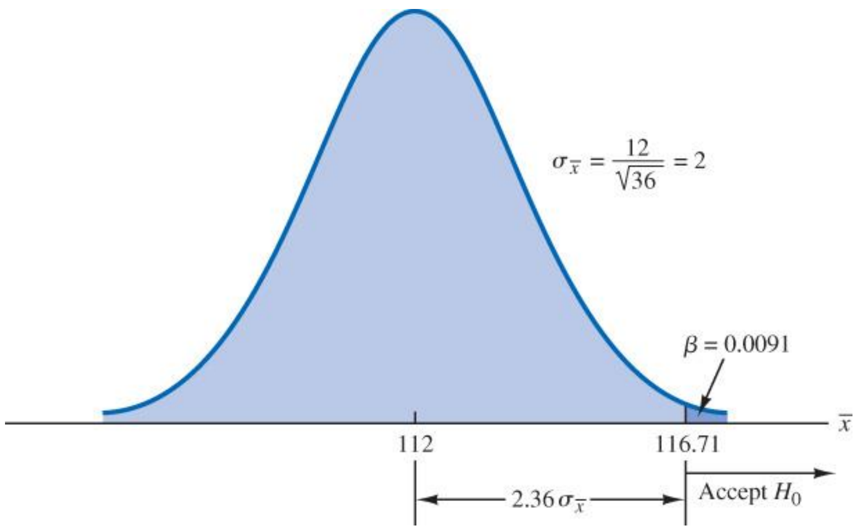
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{116.71 - 112}{12/\sqrt{36}} = 2.36$$

When $z = 2.36$, the area in the upper tail is:

$$P(z \geq 2.36) = 1 - P(z < 2.36) = 1 - 0.9909 = 0.0091$$

Denoting the probability of making a Type II error as β, we see that when $\mu = 112$, $\beta = 0.0091$.

Therefore, if the mean of the population is 112 hours, the probability of making a Type II error is only 0.0091.



Conversely, the probability of correctly rejecting H_0 when false is called the **power** of a test:

$$1 - \beta = 1 - 0.0091 = 0.9909$$

The Power Curve for the Lot-Acceptance Test

If we repeat the calculations for other values of μ less than 120, we can show in a graph the relationship between μ and the power of the test. We call such graph a **power curve**.

For example, when the true mean battery life is $\mu = 115$, we have

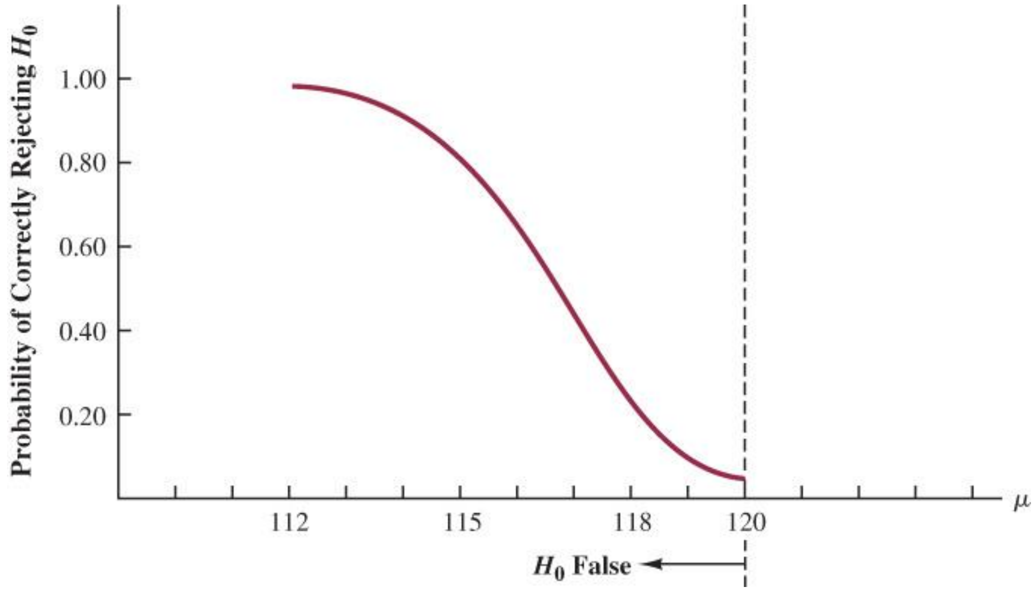
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{116.71 - 115}{12/\sqrt{36}} = 0.86$$

Thus, the probability of making a Type II error for a lower tail test increases to:

$$\beta = P(z \geq 0.86) = 1 - P(z < 0.86) = 1 - 0.8051 = 0.1949$$

And the power of the test decreases to

$$1 - \beta = 1 - 0.1949 = 0.8051$$



The power curve at any value of μ indicates the probability of correctly rejecting H_0 when false.

Determining the Sample Size for a Hypothesis Test About a Population Mean

We learned that the specified level of significance, α, for a hypothesis test to be conducted about a population mean, determines the probability of making a Type I error.

By controlling the sample size, we can also control the probability, β, of making a Type II error.

It can be shown that the minimum sample size requirement for a one-tailed hypothesis test about a population mean with probabilities of making a Type I and Type II error controlled to α and β, respectively, is

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

Where, z_α and z_β are the z values providing an area of α and β in the upper tail of a standard normal distribution, respectively (*see notes); σ Is the population standard deviation, μ_0 is the value of the population mean in the null hypothesis, and μ_a is the value of the population mean used for the Type II error.

Determining the Sample Size for the Lot-Acceptance Sampling Example

In the lot-acceptance example, shipments of batteries were returned if H_0: $\mu \geq 120$ was rejected. Recollect that the population standard deviation was assumed to be $\sigma = 12$ hours.

Suppose the quality control manager made the following statements about α and β:

Type I error: "If $\mu = 120$, I am willing to risk an $\alpha = 0.05$ probability of returning the shipment."

Type II error: "If $\mu = 115$, I am willing to risk an $\beta = 0.10$ probability of accepting the shipment."

Using the standard normal distribution, we have $z_\alpha = 1.645$ and $z_\beta = 1.28$.

The recommended sample size for the lot-acceptance example is

$$n = \frac{\left(z_\alpha + z_\beta\right)^2 \sigma^2}{(\mu_0 - \mu_a)^2} = \frac{(1.645 + 1.28)^2 (12)^2}{(120 - 115)^2} = 49.3$$

Rounding up, we recommend a sample size of 50 batteries. The quality control manager is now justified in using the accept and reject H_0 statements for the hypothesis test.

Statistical Significance and Practical Significance

The results of a hypothesis test on a mean or a proportion are statistically significant when we have sufficient evidence to reject H_0.

However, we learned that the standard error of the associated sampling distributions decreases as the sample size increases.

Thus, for increasingly larges sample sizes, we will reject $H_0\colon\ \mu = \mu_0$ for increasingly smaller differences between the sample mean $\bar{x}$ and the hypothesized population mean μ_0, or between the sample proportion $\bar{p}$ and the hypothesized population proportion p_0.

Care should be taken when interpreting the results of extremely large samples on hypothesis tests of the mean and proportion to ensure that any statistically significant differences between the point estimates and the hypothesized values of the parameters being tested are of **practical significance**.