7 Hypothesis Testing

- Using Statistics
- Statistical Hypothesis Testing
- A Two-Tailed, Large-Sample Test for the Population Mean
- A Two-Tailed, Small-Sample Test for the Population Mean
- A Two-Tailed, Large-Sample Test for the Population Proportion
- One-Tailed Tests
- The *p*-Value
- The probability of a Type II Error and the Power of the Test
- Using the Computer
- Summary and Review of Terms

7-1 Introduction

- A hypothesis is a statement or assertion about the state of nature (about the true value of an unknown population parameter):
  - The accused is innocent
  - $\mu=100$
- Every hypothesis implies its contradiction or alternative:
  - The accused is guilty
  - $\mu \neq 100$
- A hypothesis is either true or false, and you may fail to reject it or you may reject it on the basis of information:
  - Trial testimony and evidence
  - Sample data

Decision-Making

- One hypothesis is maintained to be true until a decision is made to reject it as false:
  - Guilt is proven “beyond a reasonable doubt”
  - The alternative is highly improbable
- A decision to fail to reject or reject a hypothesis may be:
  - Correct
    - A true hypothesis may not be rejected
      - » An innocent defendant may be acquitted
    - A false hypothesis may be rejected
      - » A guilty defendant may be convicted
  - Incorrect
    - A true hypothesis may be rejected
      - » An innocent defendant may be convicted
    - A false hypothesis may not be rejected
      - » A guilty defendant may be acquitted

7-2 Statistical Hypothesis Testing

- A **null hypothesis**, denoted by H_0, is an assertion about one or more population parameters. This is the assertion we hold to be true until we have sufficient statistical evidence to conclude otherwise.
  - H_0: $\mu=100$
- The **alternative hypothesis**, denoted by H_1, is the assertion of all situations *not* covered by the null hypothesis.
  - H_1: $\mu \neq 100$

- H_0 and H_1 are:
  - Mutually exclusive
    - Only one can be true.
  - Exhaustive
    - Together they cover *all* possibilities, so one or the other *must* be true.

The State of Nature, the Decision, and the Two Possible Errors

A **contingency table** illustrates the possible outcomes of a statistical hypothesis test.

	State of Nature	
Decision	H_0 True	H_0 False
Do not Reject H_0	Correct	Type II Error (β)
Reject H_0	Type I Error (α)	Correct

Example 7-1

A company that delivers packages within a large metropolitan area claims that it takes an average of 28 minutes for a package to be delivered from your door to the destination. Suppose that you want to carry out a hypothesis test of this claim.

Set the null and alternative hypotheses:

H_0: $\mu = 28$

H_1: $\mu \neq 28$

Collect sample data:

$n = 100$

$x = 31.5$

$s = 5$

Construct a 95% confidence interval for the average delivery times of *all* packages:

$$\bar{x} \pm z_{.025}\frac{s}{\sqrt{n}} = 31.5 \pm 1.96\frac{5}{\sqrt{100}}$$

$$= 31.5 \pm .98 = [30.52,\ 32.48]$$

We can be 95% sure that the average time for all packages is between 30.52 and 32.48 minutes.

Since the asserted value, 28 minutes, is not in this 95% confidence interval, we may reasonably reject the null hypothesis.

Rejection Region

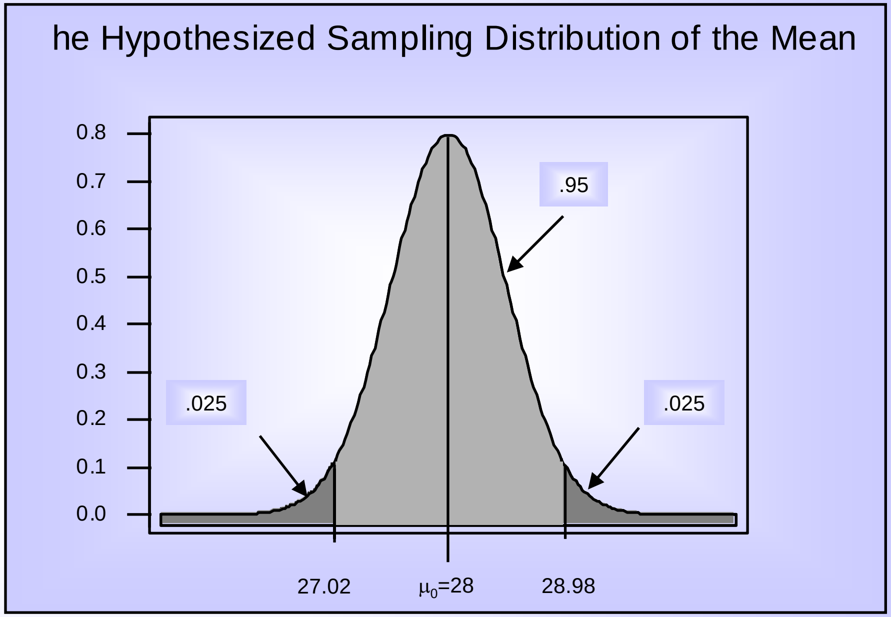
- The **rejection region** of a statistical hypothesis test is the range of numbers that will lead us to reject the null hypothesis in case the test statistic falls within this range. The rejection region, also called the **critical region**, is defined by the **critical points**. The rejection region is defined so that, before the sampling takes place, our test statistic will have a probability α of falling within the rejection region if the null hypothesis is true.

Nonrejection Region

- The **nonrejection region** is the range of values (also determined by the critical points) that will lead us not to reject the null hypothesis if the test statistic should fall within this region. The nonrejection region is designed so that, before the sampling takes place, our test statistic will have a probability 1-α of falling within the nonrejection region if the null hypothesis is true
  - In a **two-tailed test**, the rejection region consists of the values in both tails of the sampling distribution.

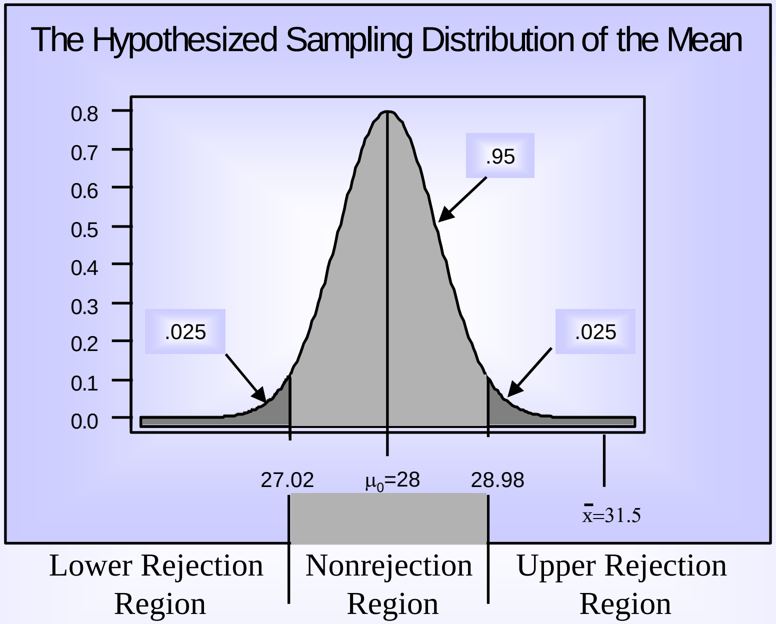
Picturing the Nonrejection and Rejection Regions

If the null hypothesis were true, then the sampling distribution of the mean would look something like this:



We will find 95% of the sampling distribution between the *critical points* 27.02 and 28.98, and 2.5% below 27.02 and 2.5% above 28.98 (a *two-tailed test*). The 95% interval around the hypothesized mean defines the *nonrejection region*, with the remaining 5% in two *rejection regions*.

The Decision Rule



- Construct a (1-α) nonrejection region around the hypothesized population mean.
  - Do not reject H_0 if the sample mean falls within the nonrejection region (between the critical points).
  - Reject H_0 if the sample mean falls outside the nonrejection region.

7-3 A Two-Tailed, Large Sample Test for the Population Mean

The test statistic (the sample mean) can be *standardized*, expressed as a number of standard errors from the hypothesized population mean:

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

In the same way, the critical points in a hypothesis test can be expressed in terms of values of the standard normal random variable:

$$\pm z_{\frac{\alpha}{2}}$$

Elements of a Two-Tailed, Large-Sample, Standardized Test for the Population Mean

Null Hypothesis	H_0: $\mu = \mu_0$
Alternative Hypothesis	H_0: $\mu \neq \mu_0$
Significance Level of the Test	α (often 0.05 or 0.01)
Test Statistic	$z = \dfrac{\bar{x} - \mu_0}{\dfrac{s}{\sqrt{n}}}$ (assuming σ is unknown, otherwise substitute σ for s)
Critical Points	The bounds $\pm z_{\frac{\alpha}{2}}$ that capture an area of (1-α)
Decision Rule	Reject the null hypothesis if either $z > z_{\frac{a}{2}}$ or $z < -z_{\frac{a}{2}}$

Critical Points of *z*

α	$\frac{\alpha}{2}$	$z_{\frac{\alpha}{2}}$
0.01	0.005	2.576
0.02	0.010	2.326
0.05	0.025	1.960
0.10	0.050	1.645
0.20	0.100	1.282

Equivalence of Testing Methods

The critical points, nonrejection region, rejection region, and test statistic can all be expressed in terms of values of the *standard normal random variable, z*.

For a 5% test, the critical values of z are ± 1.96

The nonrejection region is:

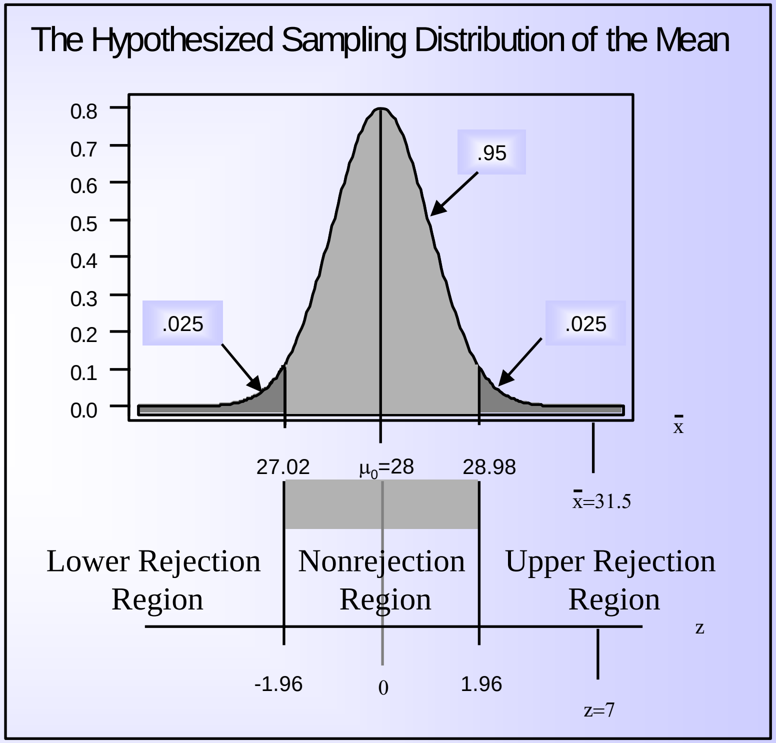
$[-1.96 \leq z \leq 1.96]$

The two rejection regions are:

$[z < -1.96]$ and $[z > 1.96]$

In this example, the test statistic is:

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{31.5 - 28}{\frac{5}{\sqrt{100}}} = \frac{3.5}{.5} = 7$$



The test statistic falls in the upper rejection region, so the null hypothesis is rejected.

Example 7-2

As part of a survey to determine the extent of required in-cabin storage capacity, a researcher needs to test the null hypothesis that the average weight of carry-on baggage per person is $\mu_0 = 12$ pounds, versus the alternative hypothesis that the average weight is not 12 pounds. The analyst wants to test the null hypothesis at $\alpha = 0.05$.

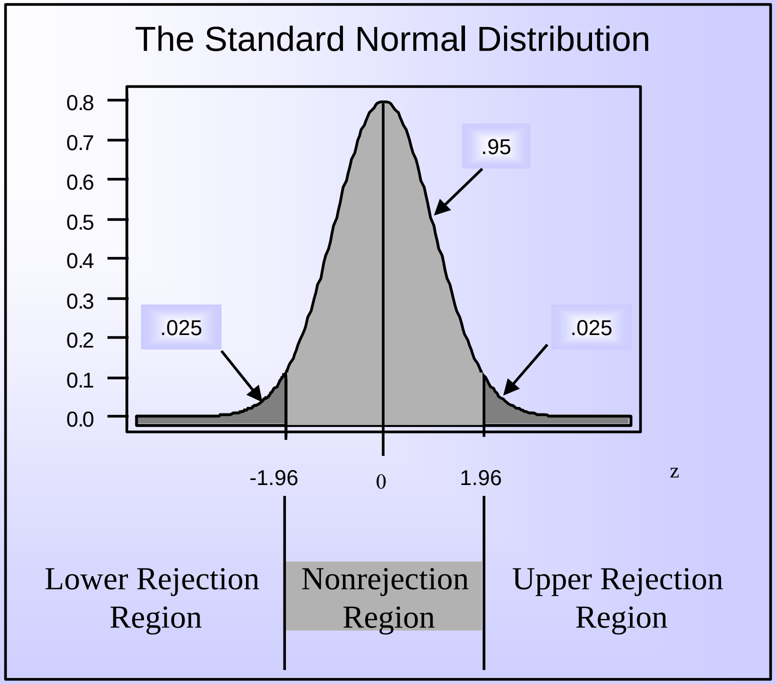
H_0: $\mu = 12$
H_1: $\mu \neq 12$

For $\alpha = 0.05$, critical values of z are ± 1.96

The test statistic is: $$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Do not reject H_0 if: $[-1.96 \leq z \leq 1.96]$

Reject H_0 if: $[z < -1.96]$ or $[z > 1.96]$



Example 7-2: Solution

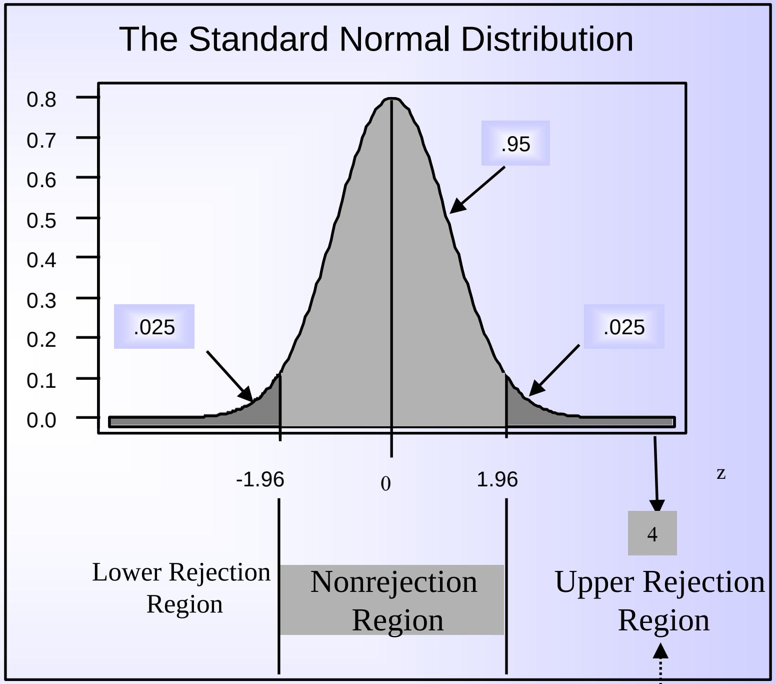
n = 144

x = 14.6

s = 7.8

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{14.6\text{-}12}{\frac{7.8}{\sqrt{144}}}$$

$$= \frac{2.6}{0.65} = 4$$



Since the test statistic falls in the upper rejection region, H_0 is rejected, and we may conclude that the average amount of carry-on baggage is more than 12 pounds.

Example 7-3

An insurance company believes that, over the last few years, the average liability insurance per board seat in companies defined as "small companies" has been \$2000. Using $\alpha = 0.01$, test this hypothesis using Growth Resources, Inc. survey data.

H_0: $\mu = 2000$
H_1: $\mu \neq 2000$

For $\alpha = 0.01$, critical values of z are ± 2.576

The test statistic is: $z = \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$

Do not reject H_0 if: $[-2.576 \leq z \leq 2.576]$

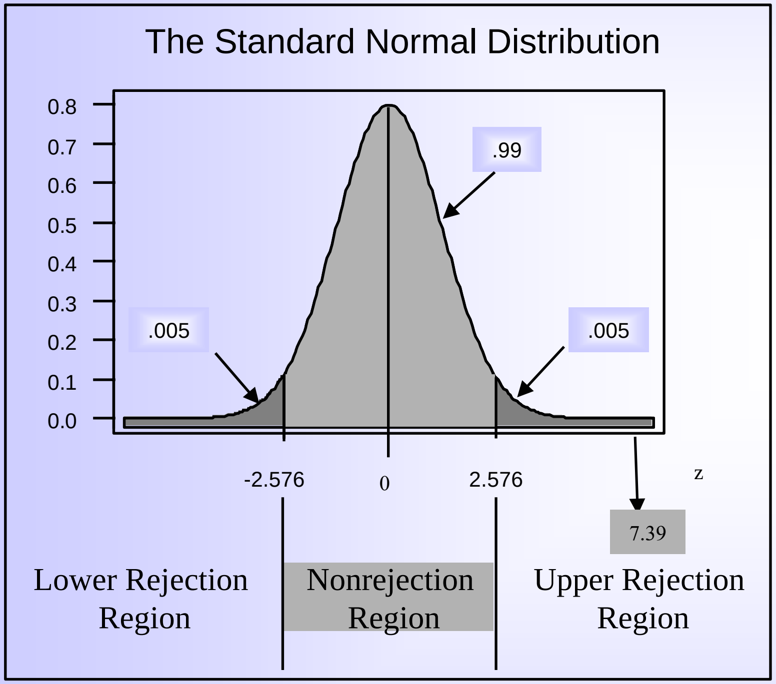
Reject H_0 if: $[z < -2.576]$ or $[z > 2.576]$

$n = 100$
$\bar{x} = 2700$
$s = 947$

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{2700 - 2000}{\frac{947}{\sqrt{100}}}$$

$$= \frac{700}{94.7} = 7.39 \Rightarrow \text{Reject } H_0$$

Example 7-3: Continued



Since the test statistic falls in the upper rejection region, H_0 is rejected, and we may conclude that the average insurance liability per board seat in "small companies" is more than $2000.

Example 7-4

The average time it takes a computer to perform a certain task is believed to be 3.24 seconds. It was decided to test the statistical hypothesis that the average performance time of the task using the new algorithm is the same, against the alternative that the average performance time is no longer the same, at the 0.05 level of significance.

H_0: $\mu = 3.24$
H_1: $\mu \neq 3.24$

For $\alpha = 0.05$, critical values of z are ± 1.96

The test statistic is: $z = \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$

Do not reject H_0 if: $[-1.96 \leq z \leq 1.96]$

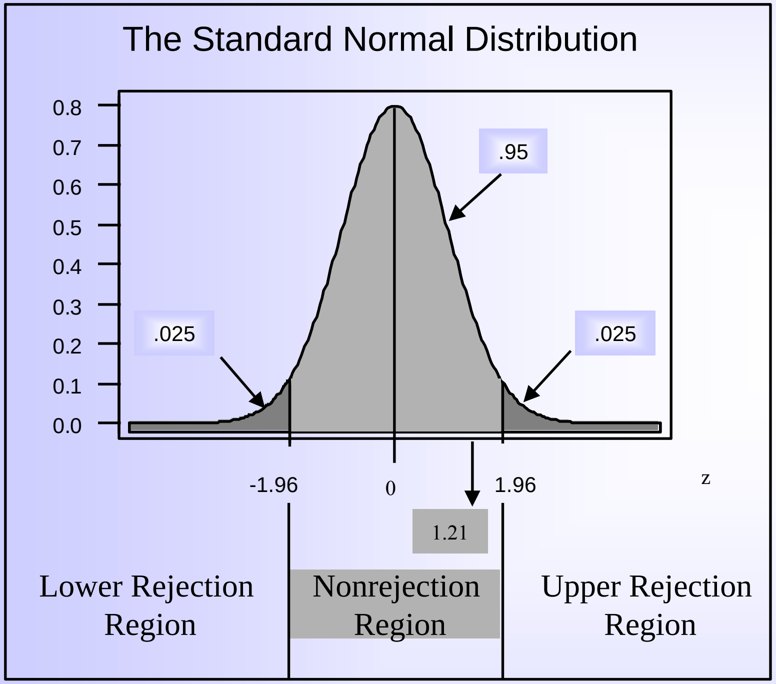
Reject H_0 if: $[z < -1.96]$ or $[z > 1.96]$

$n = 200$
$\bar{x} = 3.48$
$s = 2.8$

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{3.48 - 3.24}{\frac{2.8}{\sqrt{200}}}$$

$$= \frac{0.24}{0.20} = 1.21 \Rightarrow \text{Do not reject } H_0$$

Example 7-4: Continued



Since the test statistic falls in the nonrejection region, H_0 is not rejected, and we may conclude that the average performance time has not changed from 3.24 seconds.

7-4 A Two-Tailed, Small-Sample, Standardized Test for the Population Mean

When the population is normal, the population standard deviation, σ, is unknown and the sample size is small, the hypothesis test is based on the *t* distribution, with (*n*-1) degrees of freedom, rather than the standard normal distribution.

Small - sample test statistic for the population mean, μ:

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

When the population is normally distributed and the null hypothesis is true, the test statistic has a *t* distribution with $n - 1$ degrees of freedom

Example 7-5

According to the Japanese National Land Agency, average land prices in central Tokyo soared 49% in the first six months of 1995. An international real estate investment company wants to test this claim against the alternative that the average price did not rise by 49%, at a 0.01 level of significance.

H_0: $\mu = 49$
H_1: $\mu \neq 49$
n = 18
For $\alpha = 0.01$ and (18-1) = 17 df , critical values of t are ± 2.898

The test statistic is: $t = \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$

Do not reject H_0 if: $[-2.898 \le t \le 2.898]$

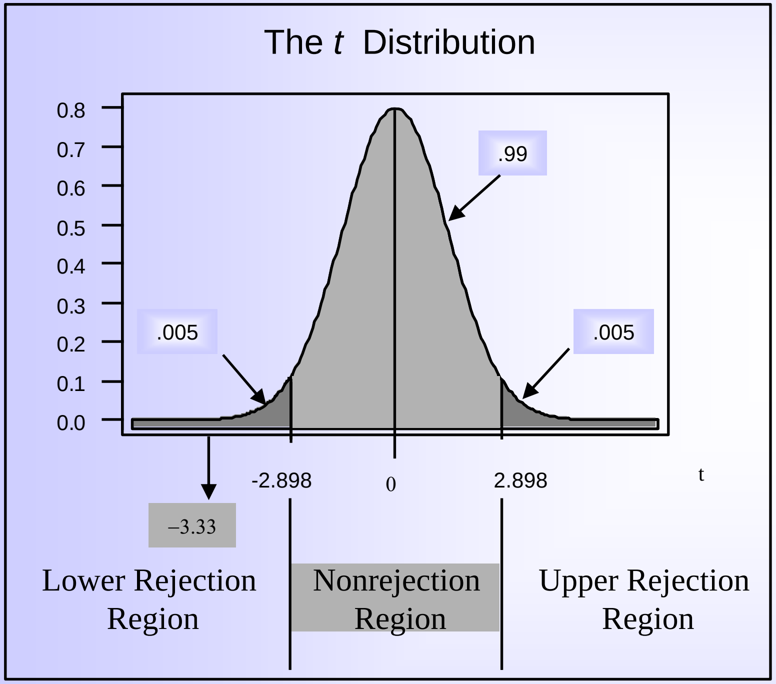
Reject H_0 if: $[t < -2.898]$ or $[t > 2.898]$

n = 18
$\bar{x}$ = 38
s = 14

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{38 - 49}{\frac{14}{\sqrt{18}}}$$

$$= \frac{-11}{3.3} = -3.33 \Rightarrow \text{Reject } H_0$$

Example 7-5: Continued



Since the test statistic falls in the rejection region, H_0 is rejected, and we may conclude that the average price has not risen by 49%. Since the test statistic is in the lower rejection region, we may conclude that the average price has risen by less than 49%.

Example 7-6

Canon, Inc,. has introduced a copying machine that features two-color copying capability in a compact system copier. The average speed of the standard compact system copier is 27 copies per minute. Suppose the company wants to test whether the new copier has the same average speed as its standard compact copier. Conduct a test at an $\alpha = 0.05$ level of significance.

H_0: $\mu = 27$
H_1: $\mu \neq 27$
n = 24
For $\alpha = 0.05$ and (24-1) = 23 df, critical values of t are ± 2.069

The test statistic is: $$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

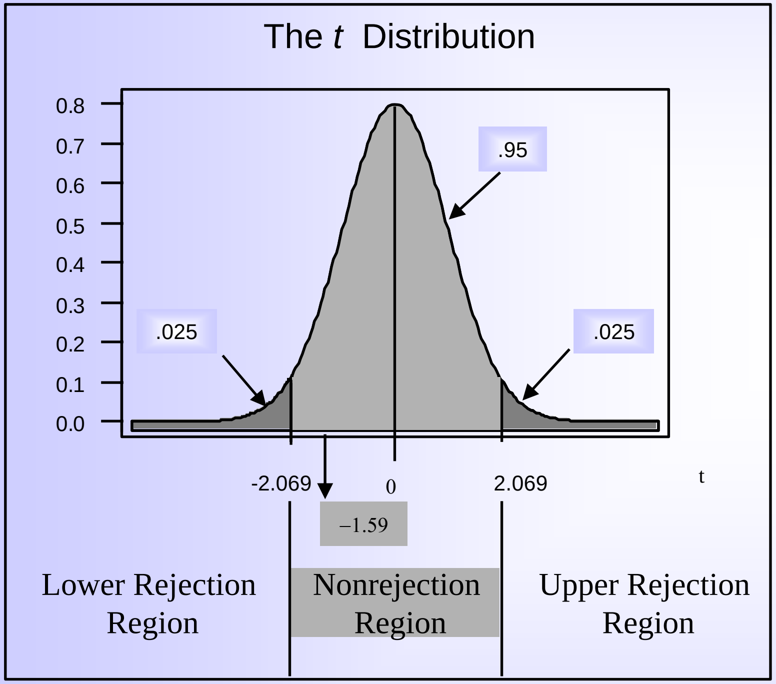
Do not reject H_0 if: $[-2.069 \leq t \leq 2.069]$
Reject H_0 if: $[t < -2.069]$ or $[t > 2.069]$

$n = 24$
$\bar{x} = 24.6$
$s = 7.4$

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{24.6 - 27}{\frac{7.4}{\sqrt{24}}}$$

$$= \frac{-2.4}{1.51} = -1.59 \Rightarrow \text{Do not reject } H_0$$

Example 7-6: Continued



Since the test statistic falls in the nonrejection region, H_0 is not rejected, and we may not conclude that the average speed is different from 27 copies per minute.

Statistical Significance

While the null hypothesis is maintained to be true throughout a hypothesis test, until sample data lead to a rejection, the aim of a hypothesis test is often to disprove the null hypothesis in favor of the alternative hypothesis. This is because we can determine and regulate α, the probability of a Type I error, making it as small as we desire, such as 0.01 or 0.05. Thus, when we reject a null hypothesis, we have a high level of confidence in our decision, since we know there is a small probability that we have made an error.

A given sample mean will not lead to a rejection of a null hypothesis unless it lies in outside the nonrejection region of the test. That is, the nonrejection region includes all sample means that are not significantly different, in a statistical sense, from the hypothesized mean. The rejection regions, in turn, define the values of sample means that are significantly different, in a statistical sense, from the hypothesized mean.

7-5 A Two-Tailed, Large-Sample Test for the Population Proportion

When the sample size is large (both *np* > *5* and *nq* > *5*), the distribution of the sample proportion may be approximated by a normal distribution with mean *p* and variance *pq*.

Large-sample test statistic for the population proportion, *p*:

$$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$$

where $q_0 = (1 - p_0)$

Example 7-7

An investment analyst for Goldman Sachs and Company wanted to test the hypothesis made by British securities experts that 70% of all foreign investors in the British market were American. The analyst gathered a random sample of 210 accounts of foreign investors in London and found that 130 were owned by U.S. citizens. At the $\alpha = 0.05$ level of significance, is there evidence to reject the claim of the British securities experts?

H_0: $p = 0.70$
H_1: $p \neq 0.70$
n = 210
For $\alpha = 0.05$ critical values of z are ± 1.96
The test statistic is: $$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

Do not reject H_0 if: $[-1.96 \leq z \leq 1.96]$
Reject H_0 if: $[z < -1.96]$ or $[z > 1.96]$

n = 210

$$\hat{p} = \frac{130}{210} = 0.619$$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.619 - 0.70}{\sqrt{\frac{(0.70)(0.30)}{210}}}$$

$$= \frac{-0.081}{0.0316} = -2.5614 \Rightarrow \text{Reject } H_0$$

Elements of a Large-Sample, Two-Tailed Test for the Population Proportion

Null Hypothesis	H_0: $p = p_0$
Alternative Hypothesis	H_0: $p \neq p_0$
Significance Level of the Test	α (often 0.05 or 0.01)
Test Statistic	$z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$ [where $q_0 = (1-p_0)$]
Critical Points	The bounds $\pm z_{\frac{\alpha}{2}}$ that capture an area of (1-α)
Decision Rule	Reject the null hypothesis if either $z > z_{\frac{\alpha}{2}}$ or $z < -z_{\frac{\alpha}{2}}$

Critical Points of *z*

α	$\frac{\alpha}{2}$	$z_{\frac{\alpha}{2}}$
0.01	0.005	2.576
0.02	0.010	2.326
0.05	0.025	1.960
0.10	0.050	1.645
0.20	0.100	1.282

7-6 One-Tailed Tests

In a **one-tailed test**, the question of interest is whether the population parameter is greater than (or less than) a hypothesized value. In quality control, it is more meaningful to test the null hypothesis that the proportion of defective items produced is less than or equal to 0.10 versus the alternative that the proportion of defective items is more than 0.10.

$$H_0: p \leq 0.10$$
$$H_1: p > 0.10$$

This leads to a **right-tailed test**, since the entire rejection region is in the right tail of the distribution.

Left, Right, and Two-Tailed Tests

The tails of a statistical test are determined by the need for an action. If action is to be taken if a parameter is greater than some value *a*, then the alternative hypothesis is that the parameter is greater than *a*, and the test is a **right-tailed** test.

$H_0: \mu \leq 50$
$H_1: \mu > 50$

If action is to be taken if a parameter is less than some value *a*, then the alternative hypothesis is that the parameter is less than *a*, and the test is a **left-tailed** test.

$H_0: \mu \geq 50$
$H_1: \mu < 50$

If action is to be taken if a parameter is either greater than or less than some value *a*, then the alternative hypothesis is that the parameter is not equal to *a*, and the test is a **two-tailed** test.

$H_0: \mu = 50$
$H_1: \mu \neq 50$

Elements of a Right-Tailed, Large-Sample, Standardized Test for the Population Mean

Null Hypothesis	H_0: $\mu \leq \mu_0$
Alternative Hypothesis	H_0: $\mu > \mu_0$
Significance Level of the Test	α (often 0.05 or 0.01)
Test Statistic	$z = \dfrac{\bar{x} - \mu_0}{\dfrac{s}{\sqrt{n}}}$ (assuming σ is unknown, otherwise substitute σ for s)
Critical Points	The bound z_α that captures an area of α to its right
Decision Rule	Reject the null hypothesis if $z > z_\alpha$

Critical Points of *z*
(One-Tailed Test)

α	z_α
0.005	2.576
0.010	2.326
0.025	1.960
0.050	1.645
0.100	1.282

Picturing Right-Tailed Tests

In a right-tailed test, there is a single **positive** critical value, z_{α}, which places the entire rejection region in the upper tail.

For a 5% right-tailed test, the critical value of z is 1.645

The nonrejection region is:

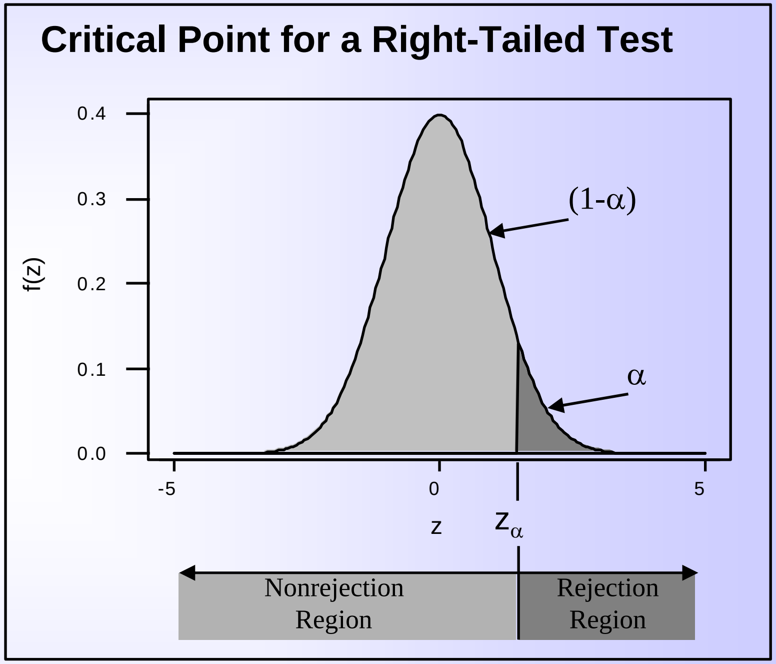
$[z \leq 1.645]$

The rejection region is:

$[z > 1.645]$

The test statistic is:

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



Example 7-8

The EPA sets limits on the concentrations of pollutants emitted by various industries. Suppose that the upper allowable limit on the emission of vinyl chloride is set at an average of 55 ppm within a range of two miles around the plant emitting this chemical. To check compliance with this rule, the EPA collects a random sample of 100 readings at different times and dates within the two-mile range around the plant. The findings are that the sample average concentration is 60 ppm and the sample standard deviation is 20 ppm. Is there evidence to conclude that the plant in question is violating the law?

H_0: $\mu \leq 55$
H_1: $\mu > 55$
n = 100
For $\alpha = 0.01$, the critical value of z is 2.326

The test statistic is: $$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

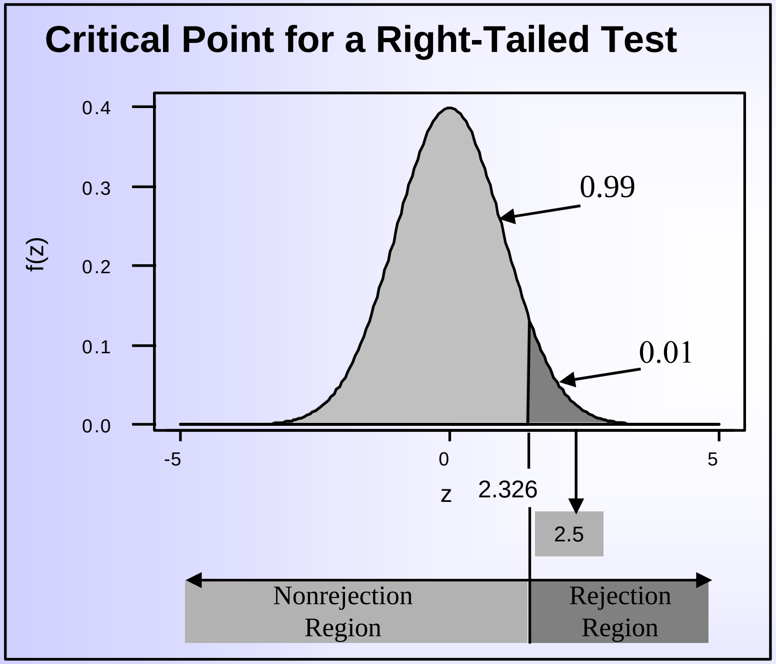
Do not reject H_0 if: $[z \leq 2.326]$
Reject H_0 if: $[z > 2.326]$

n = 100
$\bar{x}$ = 60
s = 20

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{60 - 55}{\frac{20}{\sqrt{100}}}$$

$$= \frac{5}{2} = 2.5 \quad \Rightarrow \quad \text{Reject } H_0$$

Example 7-8: Continued



Since the test statistic falls in the rejection region, H_0 is rejected, and we may conclude that the average concentration of vinyl chloride is more than 55 ppm.

Elements of a Left-Tailed, Large-Sample, Standardized Test for the Population Mean

Null Hypothesis	H_0: $\mu \geq \mu_0$
Alternative Hypothesis	H_0: $\mu < \mu_0$
Significance Level of the Test	α (often 0.05 or 0.01)
Test Statistic	$z = \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ (assuming σ is unknown, otherwise substitute σ for s)
Critical Points	The bound $-z_\alpha$ that captures an area of α to its left
Decision Rule	Reject the null hypothesis if $z < -z_\alpha$

Critical Points of *z*
(One-Tailed Test)

α	$-z_\alpha$
0.005	-2.576
0.010	-2.326
0.025	-1.960
0.050	-1.645
0.100	-1.282

Picturing Left-Tailed Tests

In a left-tailed test, there is a single **negative** critical value, $-z_{\alpha}$, which places the entire rejection region in the lower tail.

For a 5% left-tailed test, the critical value of *z* is -1.645

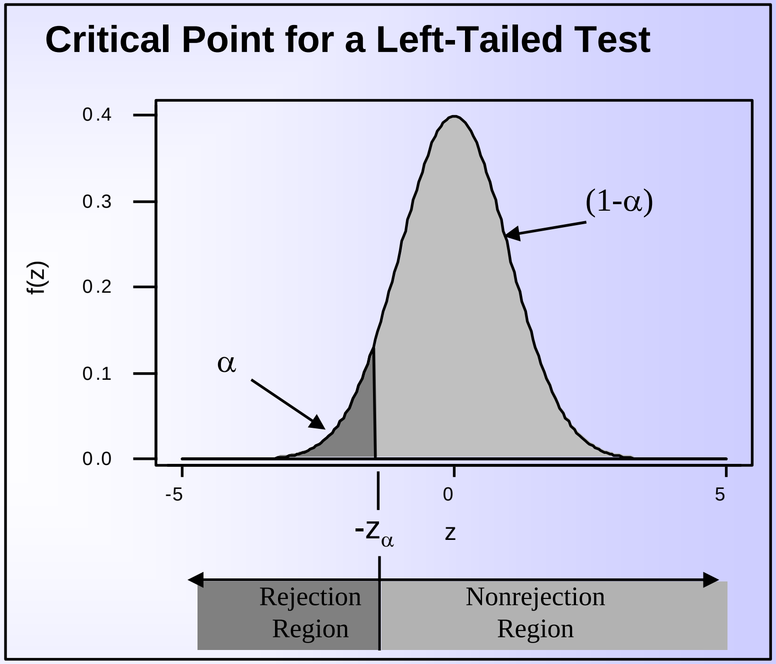
The nonrejection region is:

$[z \geq -1.645]$

The rejection region is:

$[z < -1.645]$

The test statistic is: $$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$



Example 7-9

A certain kind of packaged food bears the following statement on the package: "Average net weight 12 oz." Suppose that a consumer group has been receiving complaints from users of the product who believe that they are getting smaller quantities than the manufacturer states on the package. The consumer group wants, therefore, to test the hypothesis that the average net weight of the product in question is 12 oz. versus the alternative that the packages are, on average, underfilled. A random sample of 144 packages of the food product is collected, and it is found that the average net weight in the sample is 11.8 oz. and the sample standard deviation is 6 oz. Given these findings, is there evidence the manufacturer is underfilling the packages?

H_0: $\mu \geq 12$
H_1: $\mu < 12$
n = 144
For $\alpha = 0.05$, the critical value of z is -1.645

The test statistic is: $$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Do not reject H_0 if: [$z \geq -1.645$]
Reject H_0 if: [$z < -1.645$]

n = 144

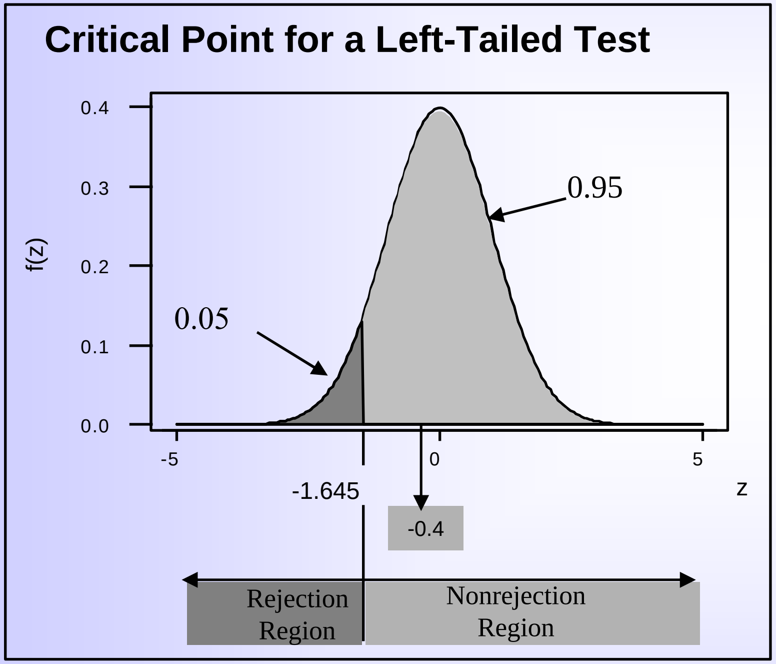
$\bar{x}$ = 11.8

s = 6

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{11.8 - 12}{\frac{6}{\sqrt{144}}}$$

$$= \frac{-.2}{.5} = -0.4 \quad \Rightarrow \text{ Do not reject } H_0$$

Example 7-9: Continued

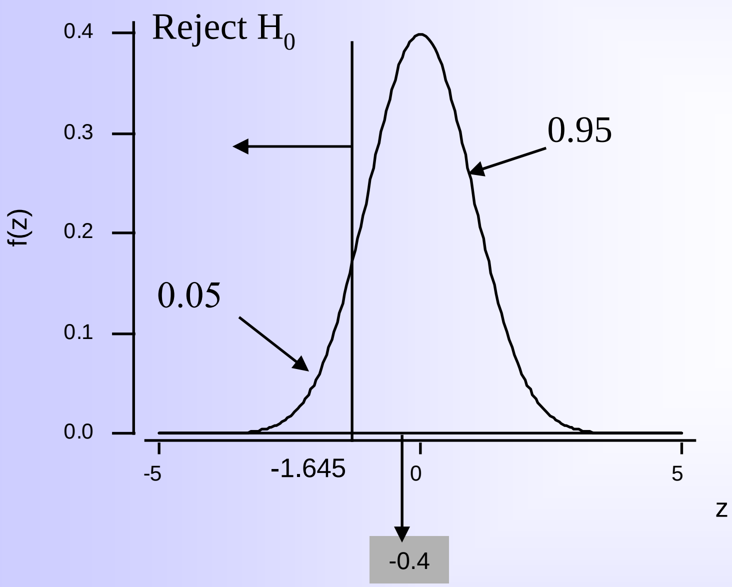


Since the test statistic falls in the nonrejection region, H_0 is not rejected, and we may not conclude that the manufacturer is underfilling packages on average.

Example 7-9: Continued Using Excel

- The NORMSINV function is used to determine the critical z-value. The NORMSDIST function is used to determine the p-value from the test statistic. The graph shows the standard normal (or z) distribution and the range of test statistic values that would result in a decision to reject the null hypothesis.

Example 7-9: Continued Using Excel



Ho: μ >=	12
H_1: μ <	12
α	0.05
CRITICAL VALUE z_α	-1.64485
DECISION RULE:	
SAMPLE SIZE	144
SAMPLE MEAN	11.8
SAMPLE STANDARD DEVIATION	6
STANDARD ERROR	0.5
TEST STATISTIC z	-0.4
P-VALUE	0.344578
CONCLUSION: DO NOT REJECT H_0	

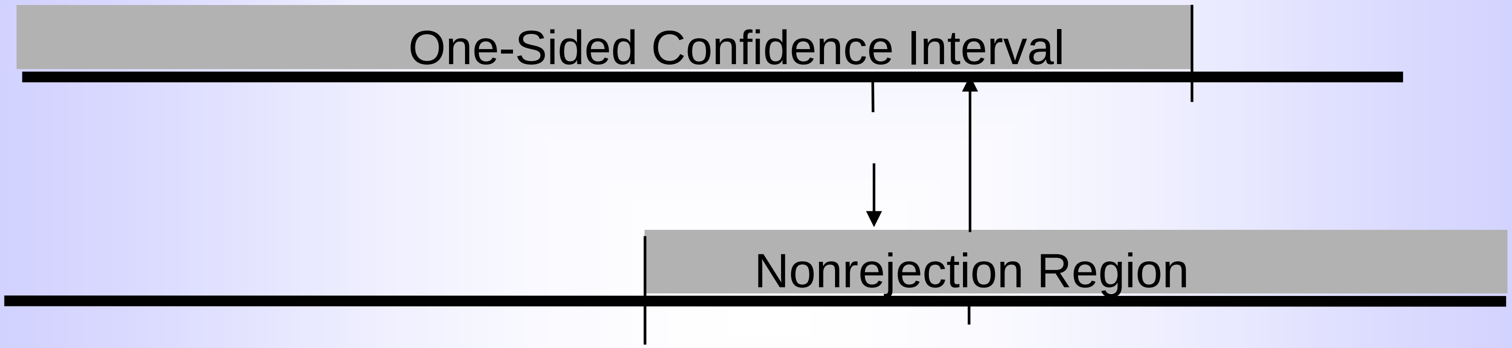
Example 7-10: Using Excel

- The state newspaper believes it takes a newspaper 5 seconds to travel from the driver's car to the doorstep. This is a cause for concern because competitors claim their drivers throw the paper in 4.58 seconds, speeding up overall delivery time. If there is such a difference, management has decided to hire a former quarterback to assist its drivers with their throwing technique. The paper has a statistician accompany 40 different drivers to find out the real time, which ended up being 4.91 seconds. At the 0.05 level of significance, is there a problem?
- Use the STDEVP Function to find the population standard deviation of the data set.
- Use the AVERAGE Function to find the mean.
- Use the ZTEST Function to find the test statistic.

Example 7-10: Using Excel (continued)

Data	Set		Data Characteristics	
2.9	5.7		Number	40
6.2	3.9		Pop. St. Dev.	1.05
4.7	6.5		Sample Mean	4.91
4.9	5.3		Rejection Region	>1.645
5.2	5.8		area	0.023516
3.2	5.4		z-value	1.99
7.5	3.9			
5.4	4.6		**Reject Null Hypothesis**	
3.3	3.8			
3.3	5.9			
7.2	4.9			
5.4	4.8			
3.2	4.9			
4.9	4.4			
3.9	4.7			
4.9	4.5			
4.5	6.2			
4.3	5.5			
4.1	5.5			
5.9	5.3			

Equivalence of One-Sided Confidence Intervals and One-Tailed Hypothesis Tests



A one-sided confidence interval contains the hypothesized value of a parameter if an only if a one-tailed test (in direction opposite to the confidence interval, using the corresponding level of significance, α) would lead to nonrejection of the null hypothesis.

A Small-Sample, One-Tailed Test for the Population Mean: Example 7-11

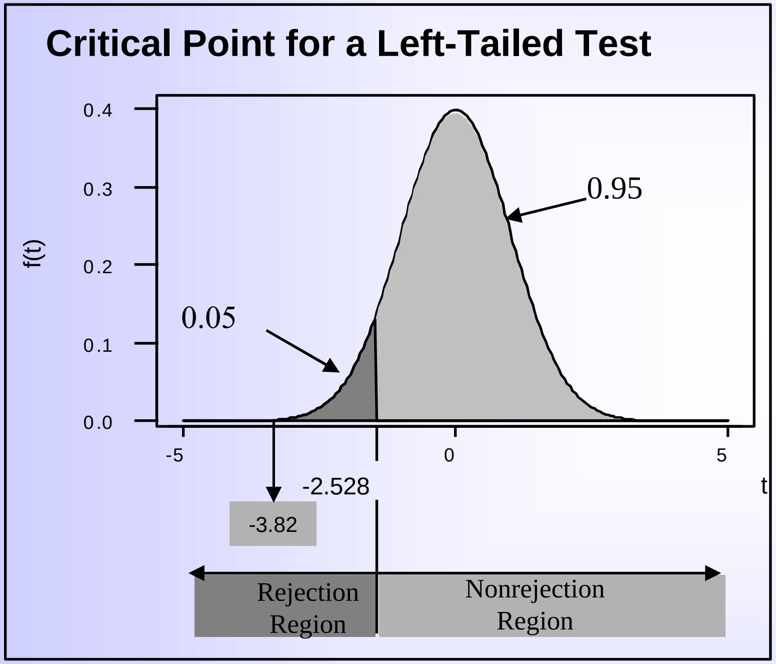
A floodlight is said to last an average of 65 hours. A competitor believes that the average life of the floodlight is less than that stated by the manufacturer and sets out to prove that the manufacturer's claim is false. A random sample of 21 floodlight elements is chosen and shows that the sample average is 62.5 hours and the sample standard deviation is 3. Using α=0.01, determine whether there is evidence to conclude that the manufacturer's claim is false.

H_0: $\mu \geq 65$
H_1: $\mu < 65$
n = 21
For α = 0.01 an (21-1) = 20 df, the critical value -2.528

The test statistic is:

Do not reject H_0 if: [$t \geq$-2.528]
Reject H_0 if: [$z < -2.528$]

Example 7-11: Continued



Since the test statistic falls in the rejection region, H_0 is rejected, and we may conclude that the manufacturer's claim is false, that the average floodlight life is less than 65 hours.

A Large-Sample, One-Tailed Test for the Population Proportion: Example 7-12

"After looking at 1349 hotels nationwide, we've found 13 that meet our standards." This statement by the Small Luxury Hotels Association implies that the proportion of all hotels in the United States that meet the association's standards is 13/1349=0.0096. The management of a hotel that was denied acceptance to the association wanted to prove that the standards are not as stringent as claimed and that, in fact, the proportion of all hotels in the United States that would qualify is higher than 0.0096. The management hired an independent research agency, which visited a random sample of 600 hotels nationwide and found that 7 of them satisfied the exact standards set by the association. Is there evidence to conclude that the population proportion of all hotels in the country satisfying the standards set by the Small Luxury hotels Association is greater than 0.0096?

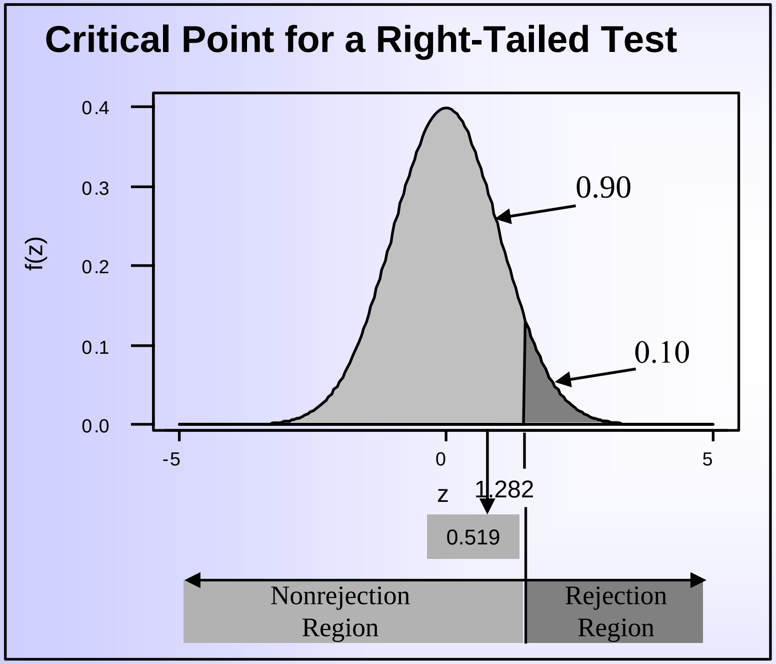
H_0: $p \leq 65$
H_1: $p > 65$
n = 600

For $\alpha = 0.10$ the critical value 1.282

The test statistic is:

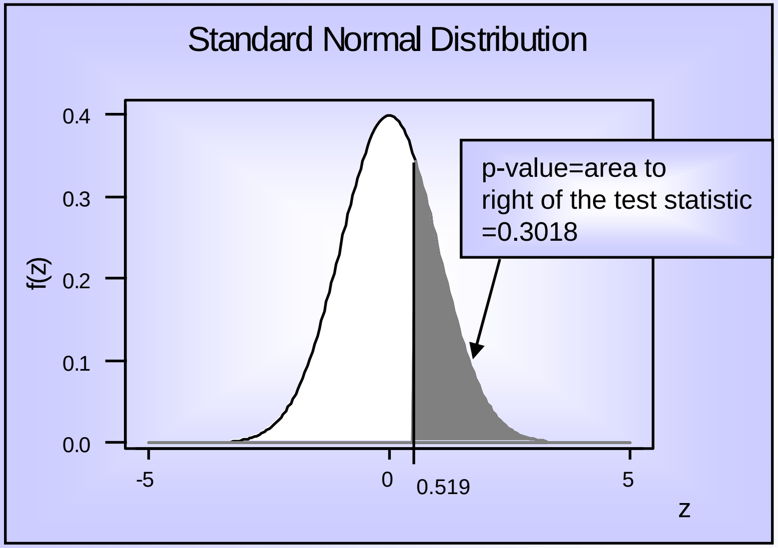
Do not reject H_0 if: $[z \leq 1.282]$
Reject H_0 if: $[z > 1.282]$

Example 7-12: Continued

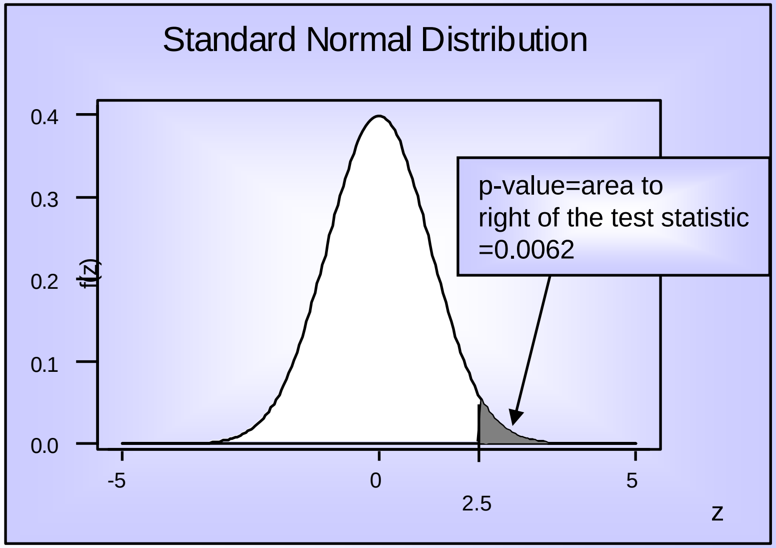


Since the test statistic falls in the nonrejection region, H_0 is not rejected, and we may not conclude that proportion of all hotels in the country that meet the association's standards is greater than 0.0096.

7-7 The p-Value



Example 7-12



Example 7-8

The ***p-value*** is the probability of obtaining a value of the test statistic as extreme as, or more extreme than, the actual value obtained, when the null hypothesis is true.

The p-value is the smallest level of significance, α, at which the null hypothesis may be rejected using the obtained value of the test statistic.

The p-Value: Rules of Thumb

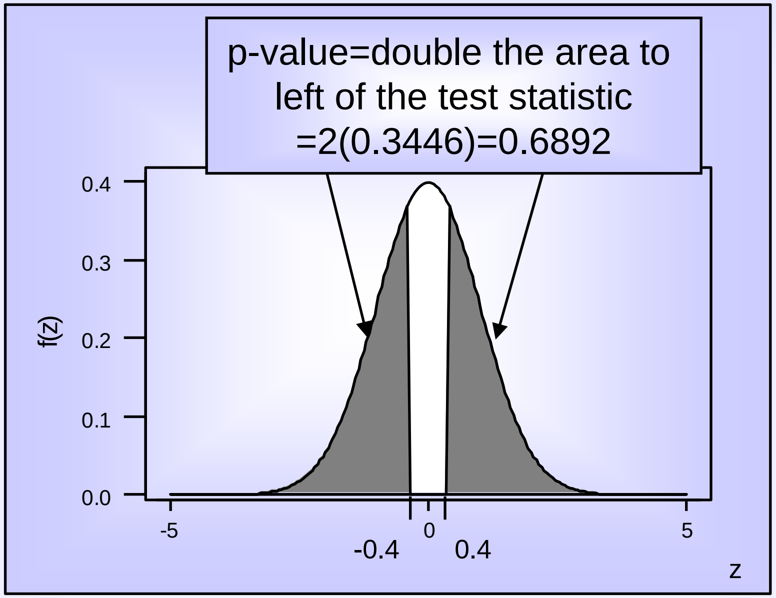
When the p-value is *smaller than 0.01*, the result is called *very significant*.

When the p-value is *between 0.01 and 0.05*, the result is called *significant*.

When the p-value is *between 0.05 and 0.10*, the result is considered by some as *marginally significant* (and by most as not significant).

When the p-value is *greater than 0.10*, the result is considered *not significant*.

p-Value: Two-Tailed Tests



Example 7-9

In a two-tailed test, we find the p-value by *doubling* the area in the tail of the distribution beyond the value of the test statistic.

The p-Value and Hypothesis Testing

The further away in the tail of the distribution the test statistic falls, the smaller is the p-value and, hence, the more convinced we are that the null hypothesis is false and should be rejected.

In a right-tailed test, the p-value is the area to the right of the test statistic if the test statistic is positive.

In a left-tailed test, the p-value is the area to the left of the test statistic if the test statistic is negative.

In a two-tailed test, the p-value is twice the area to the right of a positive test statistic or to the left of a negative test statistic.

For a given level of significance, α:

Reject the null hypothesis if and only if $\alpha \geq$ p-value

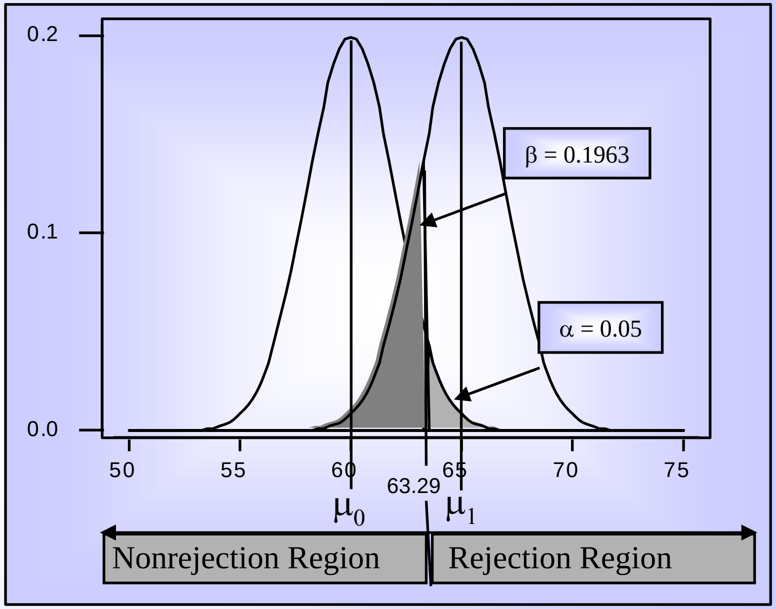
7-8 The Probability of a Type II Error and the Power of the Test

Consider the following two (unusual) hypotheses:

H_0: $\mu = 60$
H_1: $\mu = 65$

and a test of these hypotheses with: $\alpha = 0.05$, n=100, and $\sigma = 20$.

Do not reject H_0 if: $[x \leq 63.29]$
Reject H_0 if: $[x > 63.29]$



If the null hypothesis is true, there is an $\alpha = 0.05$ chance of committing a Type I error (rejecting a true null hypothesis).

If the null hypothesis is not true (if the alternative hypothesis is true), then there is a $\beta = 0.1963$ chance of committing a Type II error (accepting a false null hypothesis).

Probabilities of Type I and Type II Errors

Probability of a Type I error:

$$\alpha = P(\bar{X} > C \mid \mu = \mu_0)$$

Probability of a Type II error:

$$\beta = P(\bar{X} < C \mid \mu = \mu_1)$$

Probabilities of Type I and Type II Errors

In this example:

$$\beta = P(\bar{x} < C \mid \mu = \mu_1)$$

$$= P\left[\frac{\bar{x} - \mu_1}{\frac{\sigma}{\sqrt{n}}} < \frac{C - \mu_1}{\frac{\sigma}{\sqrt{n}}}\right] = P\left(z < \frac{63.29 - 60}{\frac{20}{\sqrt{100}}}\right)$$

$$= P(z < -0.855) = 0.1963$$

The Power of a Test

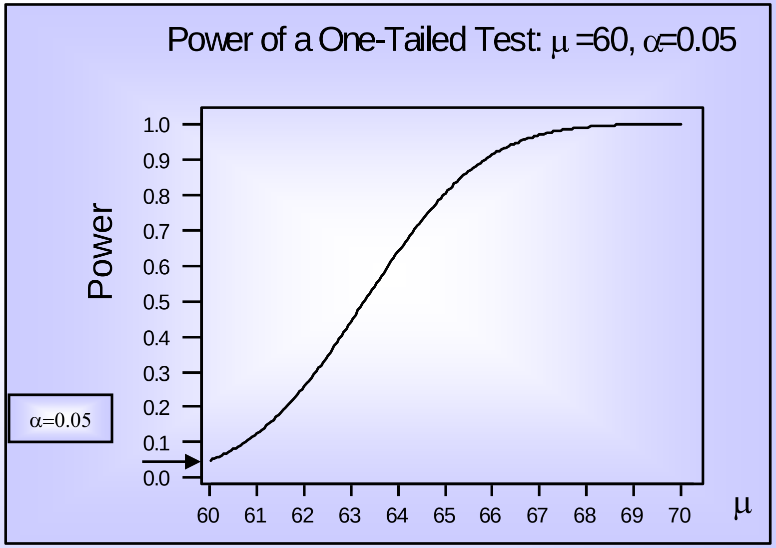
The **power** of a statistical hypothesis test is the probability of rejecting the null hypothesis when the null hypothesis is false.

Power = (1 - β)

The Power Function

The probability of a type II error, and the power of a test, depends on the actual value of the unknown population parameter. The relationship between the population mean and the power of the test is called the **power function**.

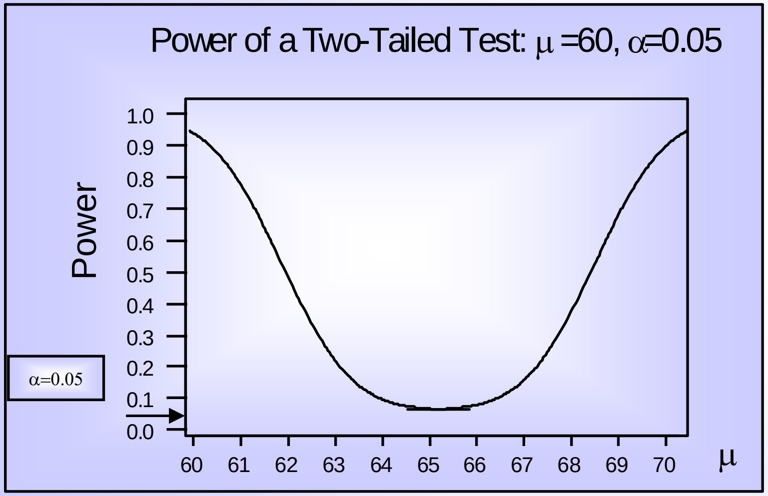
Value of μ	β	Power = $(1 - \beta)$
61	0.8739	0.1261
62	0.7405	0.2695
63	0.5577	0.4423
64	0.3613	0.6387
65	0.1963	0.8037
66	0.0877	0.9123
67	0.0318	0.9682
68	0.0092	0.9908
69	0.0021	0.9972

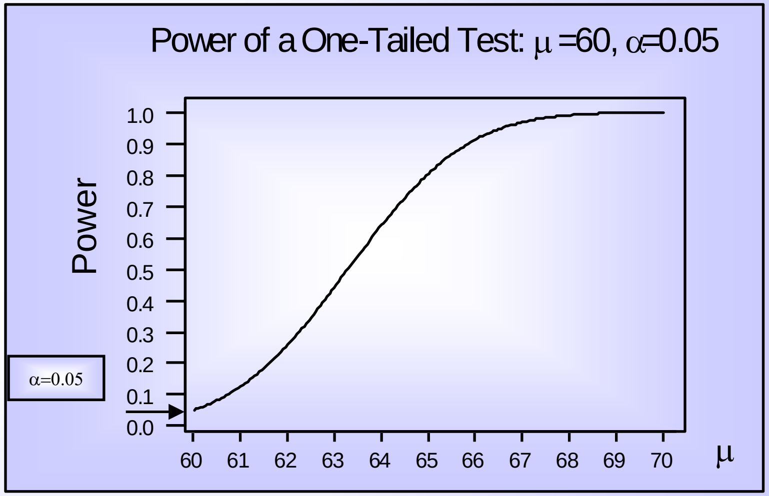


Factors Affecting the Power Function

- The power depends on the distance between the value of the parameter under the null hypothesis and the *true value* of the parameter in question: *the greater this distance, the greater the power*.
- The power depends on the population standard deviation: *the smaller the population standard deviation, the greater the power.*
- The power depends on the sample size used: *the larger the sample, the greater the power*.
- The power depends on the level of significance of the test: *the smaller the level of significance,* α *, the smaller the power.*

Power Functions of One-Tailed and Two-Tailed Tests





Example 7-13

The makers of a competing automobile want to test the Saab 9000 and carry out the following hypotheses test: $H_0: \mu \le 7.6$ and $H_1: \mu > 7.6$. The firm is interested in finding the power of the test evaluated at $\mu_1 = 7.7$

$H_0: \mu \le 7.6$
$H_1: \mu > 7.6$
n = 150
$\sigma = 0.4$
The critical point for this test is:

$$C = \mu_0 + 1.645\frac{\sigma}{\sqrt{n}}$$
$$= 7.6 + 1.645\frac{0.4}{\sqrt{150}}$$
$$= 7.6 + 0.0537 = 7.6537$$

$$Power = P(\bar{x} > C | \mu = 7.7)$$
$$= P\left(z > \frac{7.653 - 7.7}{\frac{0.4}{\sqrt{150}}}\right)$$
$$= P(z > -1.42) = 0.9222.$$

Using the Computer: Example 7-14

Federal Express claims in an advertisement that customers can get a report on the exact status of their package within 30 minutes. A competitor wants to test this claim, trying to prove that the average time for an exact status report by Federal Express is over 30 minutes. The null hypothesis is therefore H_0: $\mu \leq 30$, and the alternative hypothesis is H_1: $\mu > 30$. A random sample of 20 status reports is collected.

```
MTB > set c1
DATA> 39 35 37 28 40 25 27 24 42 45
DATA> 20 38 39 25 42 26 48 51 48 41
DATA> end.
MTB > ttest of MU = 30 Alternative = 1 on data in C1

T-Test of the Mean


Test of mu = 30.00 vs mu > 30.00

Variable    N    Mean   StDev   SE Mean      T   P-Value
C1         20   36.00    9.23      2.06   2.91    0.0045

MTB >
```