Introductory Statistics: A Problem-Solving Approach

by Stephen Kokoska

Chapter 9

Hypothesis Tests Based on a Single Sample

Hypothesis

In statistics, a **hypothesis** is a declaration, or claim, in the form of a mathematical statement, about the value of a specific population parameter (or about the values of several population characteristics).

<u>The Four Parts of a Hypothesis Test:</u>

1. The **null hypothesis**, denoted H_0, is the claim about a population parameter that is believed to be true, or the hypothesis to be tested. This claim usually represents the status quo or existing state. The null hypothesis is written in terms of a single value (with an equal sign)—for example, $\theta = 5$.
   - Says that nothing new or interesting happens here: “No effect”, “No difference”, “No change”
   - Begins with the assumption that the null hypothesis is true. Similar to the notion of innocent until proven guilty.
   - Is always about a population parameter, not about a sample statistic.

$$H_0 : \mu = 3 \qquad \cancel{H_0 : \overline{x} = 3}$$

Hypothesis

2. The **alternative hypothesis**, denoted H_a or H_1, identifies other possible values of the population parameter, or a possibility not included in the null hypothesis.
   - H_a indicates the possible values of the parameter if H_0 is false.
   - Experiments are often designed to determine whether there is evidence in favor of H_a.

3. The **test statistic**, denoted TS, is a rule, related to the null hypothesis, involving the information in a sample.
   The *value* of the test statistic will be used to determine which hypothesis is more likely to be true, H_0 or H_a.

4. The **rejection region** (RR) or **critical region** (CR), is an interval or set of numbers specified such that if the value of the test statistic lies in the rejection region, then the null hypothesis is rejected.
   There is also a corresponding *nonrejection region*: If the value of the test statistic lies in this set, then we *cannot reject* H_0.

The Test of a Statistical Hypothesis

1. The test of a statistical hypothesis is a procedure to decide whether there is evidence to suggest that the alternative hypothesis H_a is true. The ultimate objective of a hypothesis test is to use the information in a sample to decide which hypothesis is more likely to be true, H_0 or H_a. Usually, we look for evidence to reject the null hypothesis.
2. The rejection and the nonrejection regions divide the values of the test statistic into parts. These two regions are divided by the cutoff (critical) value.
3. Once the four parts are identified, the sample data are used to compute a value of the test statistic. There are only two possible decisions.
   a) If the value of the test statistic lies in the rejection region, we reject H_0.
   b) If the value of the test statistic does not lie in the rejection region, we cannot reject H_0.
4. The formal hypothesis test procedure is analogous to the four-step inference procedure used in the previous chapters. The **claim** corresponds to H_0, a claim about a population parameter. The **experiment** is equivalent to a value of the test statistic. **Likelihood** is expressed in terms of the rejection and nonrejection regions. The **decision/conclusion** is completely determined by the region in which the value of the test statistic lies.

Writing the Hypotheses

- The null hypothesis is always stated in terms of a single value of the population parameter, θ.

$$H_0: \theta = \theta_0$$

- There are three possible alternative/research hypotheses:

$$\left.\begin{array}{l} H_a: \theta > \theta_0 \\ H_a: \theta < \theta_0 \end{array}\right\} \text{One-sided alternatives}$$

$$H_a: \theta \neq \theta_0 \} \text{ Two-sided alternative}$$

- Only one alternative hypothesis is selected.
- H_a answers the question, "What is the experimenter trying to prove (detect) about θ?"

"Must you answer every question with a hypothesis?"

Examples on Writing the Hypotheses

Example: Trust Me, I'm a Doctor

According to a recent survey, only 58% of Americans trust doctors. Suppose a national advertising campaign is conducted to address confidence in doctors and medical leaders, and an experiment (new survey) is conducted to determine whether it has been effective. What null and alternative hypotheses should be used?

It is assumed that 58% of all Americans trust their doctor. Therefore,

$\boldsymbol{H_0: p = 0.58}$

The experiment (new survey) is designed to detect an increase in this proportion, for example, to answer the question, "Do Americans now have greater trust in doctors?" Researchers hope to find evidence that the proportion of Americans who trust doctors is greater than 0.58. Therefore,

$\boldsymbol{H_a: p > 0.58}$

Examples on Writing the Hypotheses

Example: Recycled Paper

The thickness (in inches) of recycled printer paper is important, because sheets that are too thick will clog the printer, and paper that is too thin will rip and bleed toner. The variance in thickness for 20-lb printer paper at a manufacturing plant is known to be 0.0007. A new process is developed that uses more recycled fiber, and an experiment is conducted to detect any difference in the variance in paper thickness. State the appropriate null and alternative hypotheses.

The null hypothesis is given in terms of σ^2: $\boldsymbol{H_0: \sigma^2 = 0.0007}$

The experiment is designed to detect any difference in the population variance. This suggests a two-sided alternative: $\boldsymbol{H_a: \sigma^2 \neq 0.0007}$

Error Definitions

1. The value of the test statistic may lie in the rejection region, but the null hypothesis is true. If we reject H_0 when H_0 is true, this is called a **type I error**. The probability of a type I error is called the **significance level** of the hypothesis test and is denoted by α:

$$P(\text{type I error}) = P(H_0 \text{ is rejected given that } H_0 \text{ is true}) = \alpha$$

Typical values of α are 0.01, 0.05, or 0.10

2. The value of the test statistic may not lie in the rejection region, but the alternative hypothesis is true. If we do not reject the null hypothesis when H_a is true, this is called a **type II error**. The probability of a type II error is denoted by β:

$$P(\text{type II error}) = P(H_0 \text{ is not rejected given that } H_0 \text{ is false}) = \beta$$

		Decision	
		Reject H_0	**Do not reject H_0**
State of Nature	**H_0 False**	Correct decision	Type II error
	H_0 True	Type I error	Correct decision

Practical example on explaining type I and II errors

Null Hypothesis: A person does not have the disease.

Research Hypothesis: A person does have the disease.

Type 1 Error: A person tested positive for the disease s/he does not have.

Type 2 Error: A person tested negative for the disease s/he does have.

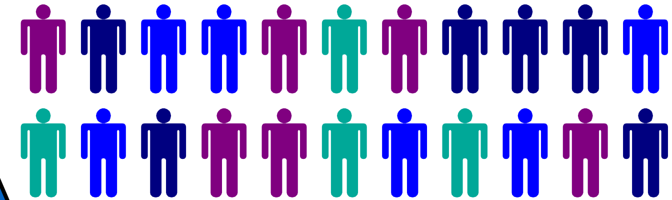
	Does Not Have Disease	Does Have Disease
Tested Negative	Correct Decision	Type II Error
Tested Positive	Type I Error	Correct Decision

		Decision	
		Reject H_0	**Do not reject H_0**
State of Nature	**H_0 False**	Correct decision	Type II error
	H_0 True	Type I error	Correct decision

- P(type I error) = P(H_0 is rejected given that H_0 is true)
- P(type II error) = P(H_0 is not rejected given that H_0 is false)

Hypothesis Testing Process

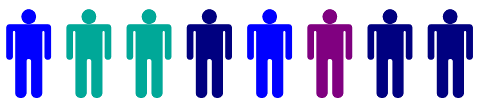
Population



Claim: **the population mean age is 50**

Null Hypothesis: H_0: $\mu = 50$

Now select a random sample:



Sample

Suppose the sample mean age is 20:
$\bar{x} = 20$

Is $\bar{x} = 20$ likely if $\mu = 50$?

If not likely,
REJECT
Null Hypothesis

One-Sample Hypothesis Tests

Three scenarios of rejection regions:

α = level of significance

Z_α = cutoff or critical value which can be found from the table or R-software

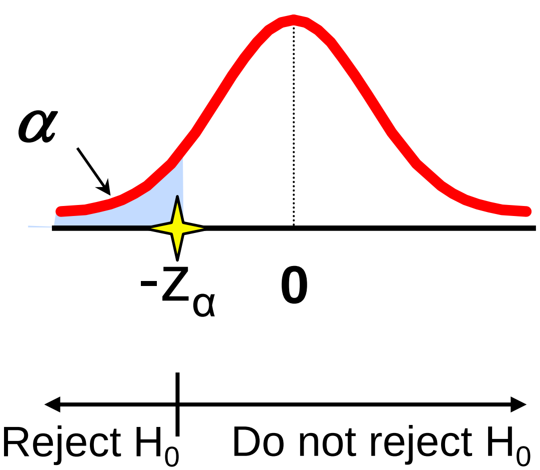
One side/tail test | **Two tailed test**

Lower tail test

Example

$H_0: \mu = 3$

$H_A: \mu < 3$

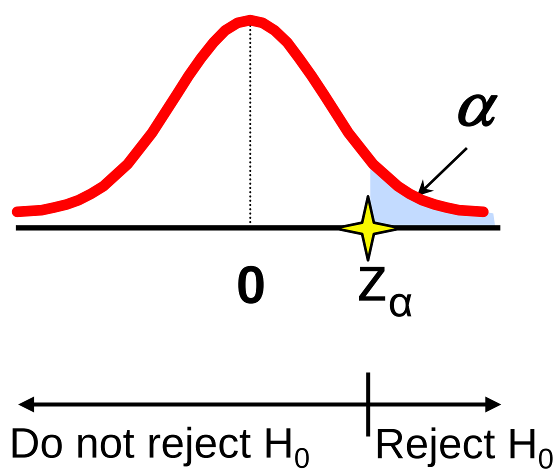


Upper tail test

Example

$H_0: \mu = 3$

$H_A: \mu > 3$

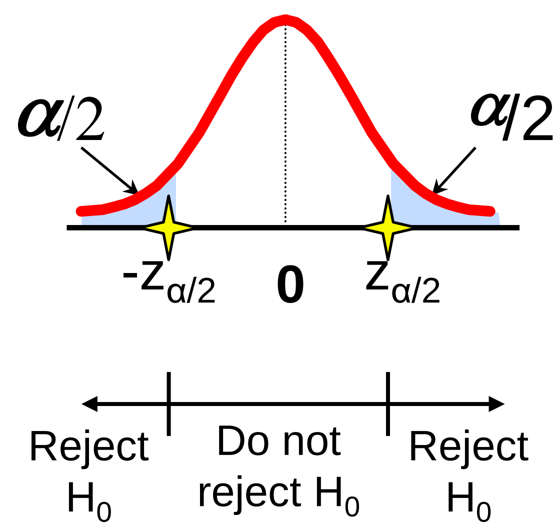


Two tailed test

Example

$H_0: \mu = 3$

$H_A: \mu \neq 3$



Steps in Hypothesis-Testing Analysis

1. Specify the parameter of interest.
2. Determine the null value and formulate the null and alternative hypotheses.
3. Calculate the test statistic.
4. Determine the rejection region (tail test) for the selected significance level
5. Decide whether H_0 should be rejected and draw a conclusion in the problem context.

Which test statistic should be used?

Hypothesis Tests for μ

σ Known

use z-statistic regardless the sample size

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

σ Unknown

$n < 30$
use t-statistic

$$T = \frac{\bar{X} - \mu_0}{S / \sqrt{n}}$$

$n \geq 30$
use z-statistic

$$Z = \frac{\bar{X} - \mu_0}{S / \sqrt{n}}$$

Example1: The Dead Zone

The long-term mean area of the Dead Zone is 5460 mi^2. As a result of recent flooding in the midwest and subsequent runoff from the Mississippi River, researchers believe that the Dead Zone area will increase. A random sample of 36 days was obtained, and the sample mean area of the Dead Zone was 6258 mi^2. Is there any evidence to suggest that the current mean area of the Dead Zone is greater than the long-term mean? Assume that the population standard deviation is 1850 and use $\alpha = 0.025$.

$n = 36$
$\sigma = 1850$
$\bar{x} = 6258$
$\alpha = 0.025$

H_0: $\mu = 5460$
H_a: $\mu > 5460$

RR: $Z \geq z_{\alpha} = z_{0.025} = 1.96$

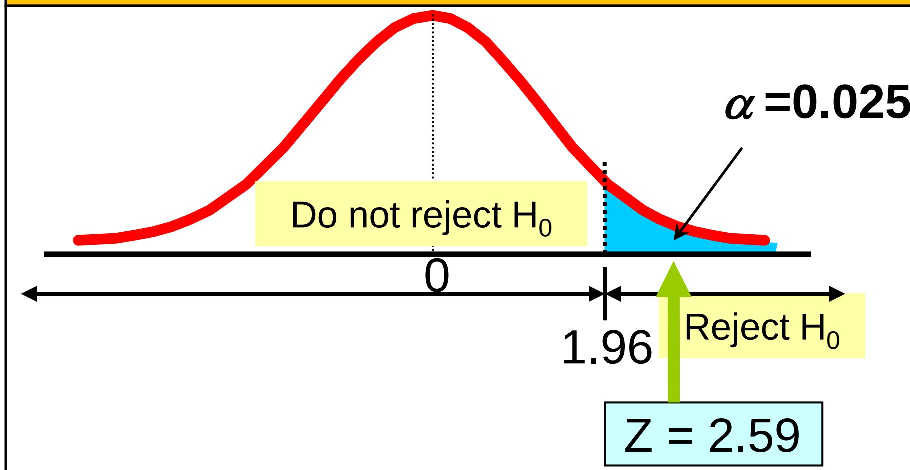
$$\text{TS: } Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{6258 - 5460}{1850/\sqrt{36}} = 2.59$$

Confidence Level	80%	90%	95%	98%	99%
Alpha Two Tail	20%	10%	5%	2%	1%
Alpha One Tail	10%	5%	2.5%	1%	0.05%

Critical Value	1.28	1.65	1.96	2.33	2.58

Example1: The Dead Zone

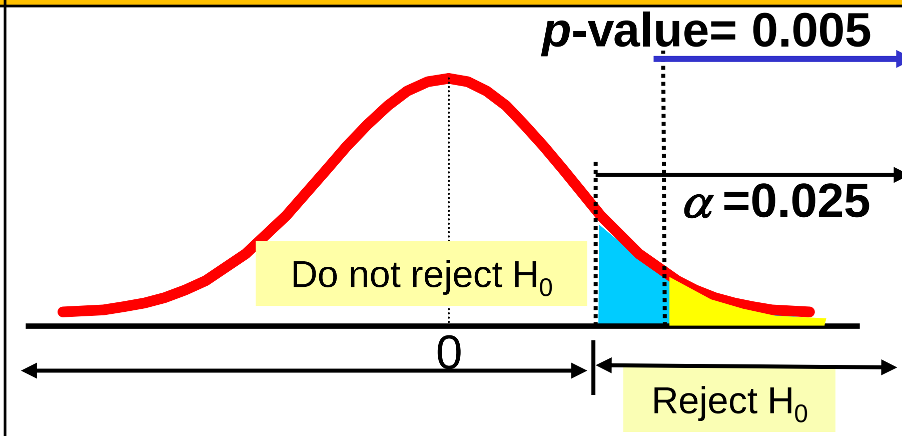
1st. approach using z-critical value



qnorm(0.975,0,1)
1.959964

Since the z-test value of 2.59 is > the critical value of 1.96, we reject the H_0

2nd. approach using *p*-value



pnorm(- 2.59,0,1)
0.004798797

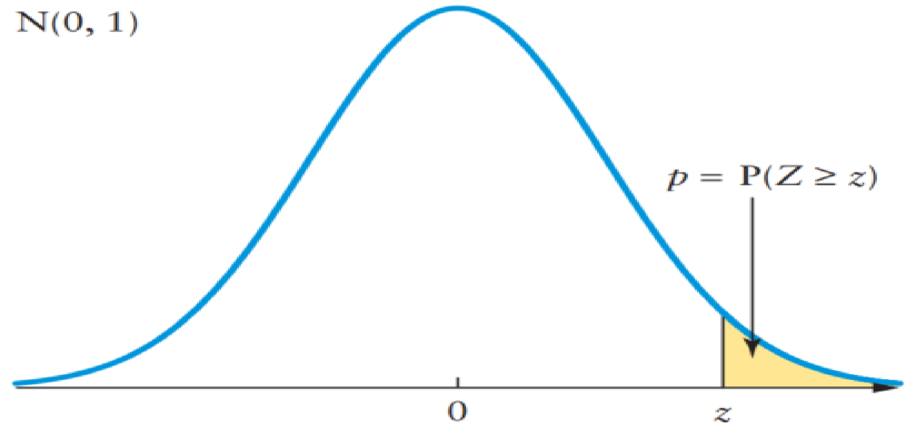
pnorm(2.59,0,1)
0.9952012 → 1 - 0.9952012 = 0.0047988

The H_0 would be rejected because the *p*-value = 0.05% is < α = 2.5%

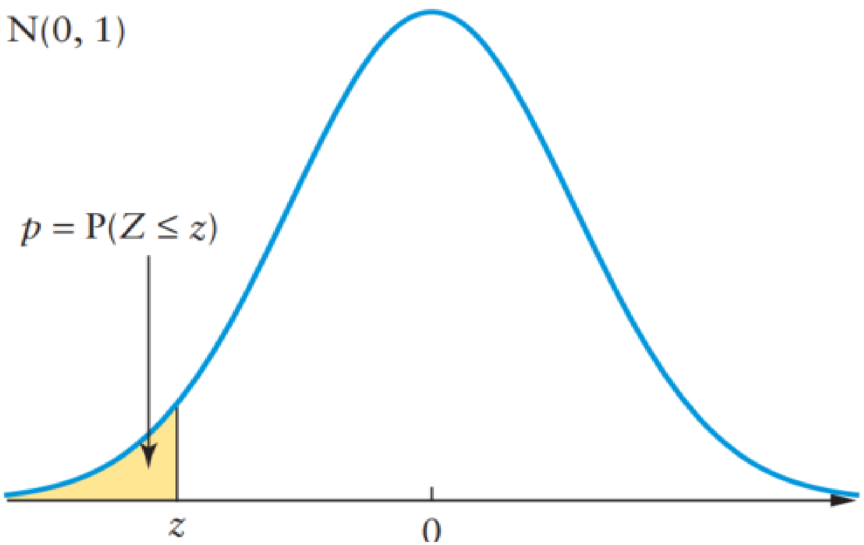
Conclusion: There is evidence to suggest that the current mean area of the Dead Zone is **greater** than 5460 mi^2.

Probability Definition of p-Value

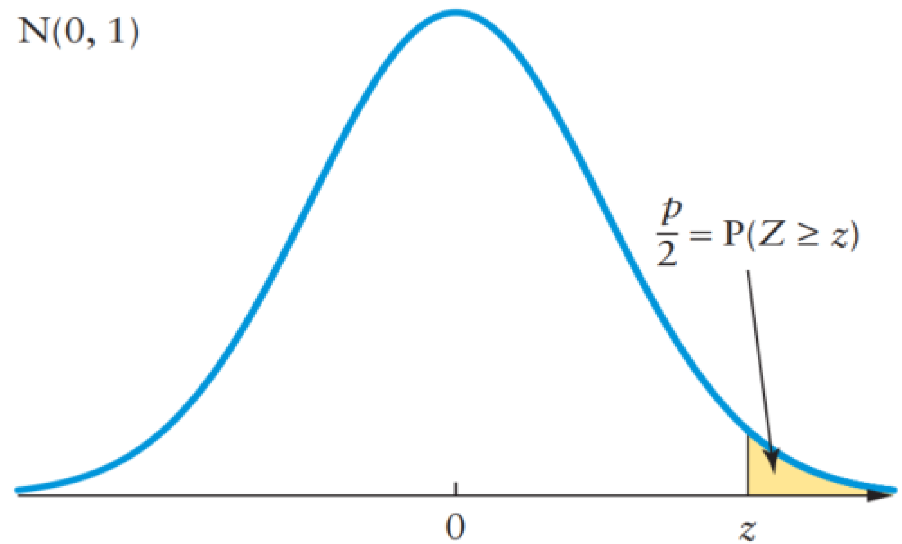
Alternative hypothesis	Probability definition
H_a: $\mu > \mu_0$	$p = P(Z \geq z)$
H_a: $\mu < \mu_0$	$p = P(Z \leq z)$
H_a: $\mu \neq \mu_0$	$p/2 = P(Z \geq z)$ if $z \geq 0$ $p/2 = P(Z \leq z)$ if $z < 0$



The p-value for H_a: $\mu > \mu_0$



The p-value for H_a: $\mu < \mu_0$

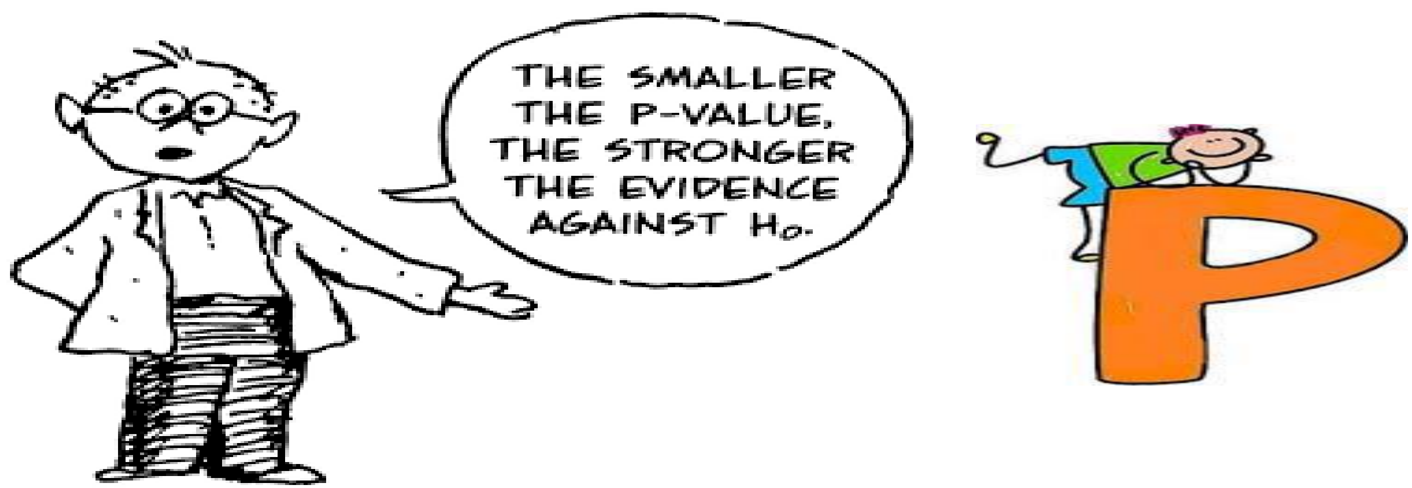


The p-value for H_a: $\mu \neq \mu_0$

Probability Definition of *p*-Value

Notes:

- Compare test statistic/value with critical value (**1st. approach**):
  - ✓ Reject the null hypothesis if the test value falls in the "reject" area.
  - ✓ Do not reject the null hypothesis if the test value falls in the "accept" area.
- Compare *p*-value with α (**2nd. approach**):
  - ✓ Reject the null hypothesis when *p*-value is $< \alpha$
  - ✓ Do not reject the null hypothesis when *p*-value is $\geq \alpha$
- *p*-value is the probability that measures the strength of evidence against a null hypothesis
- By rejecting the null hypothesis, we are concluding that the evidence is statistically significant.



16

Example2: Hyperloop One

A 500-m test tunnel, known as DevLoop, has been constructed in the Nevada desert. Suppose 20 random tests are selected, and the speed of the capsule is carefully measured for each. The sample mean is 660.1. Assume the distribution of capsule speed is normal, with $\sigma = 25$. Is there any evidence to suggest that the true mean speed is less than 670 mph? Use $\alpha = 0.05$

$n = 20$
$\sigma = 25$
$\bar{x} = 660.1$
$\alpha = 0.05$

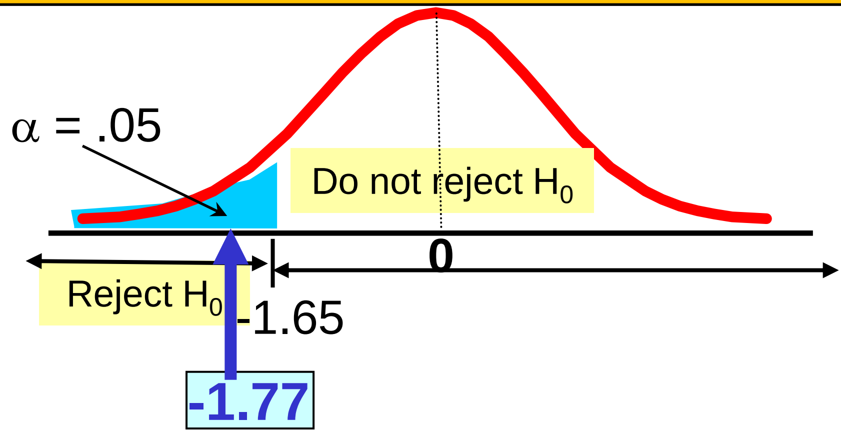
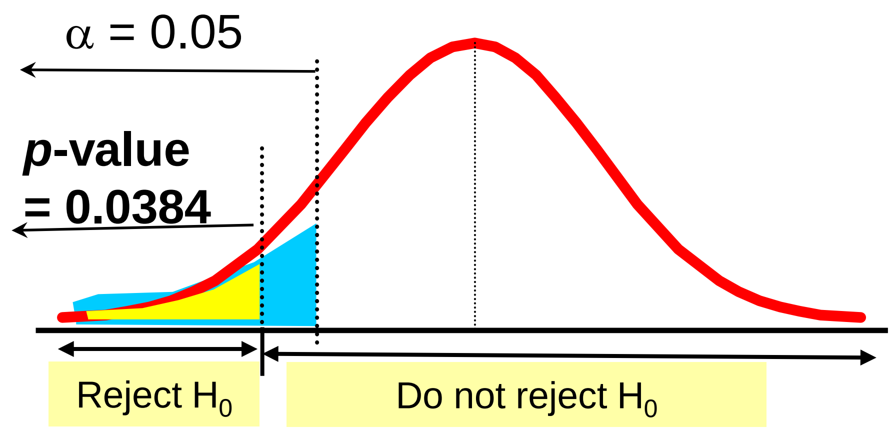
H_0: $\mu = 670$ (the true mean speed equals 670 mph)
H_a: $\mu < 670$ (the true mean speed is less than 670 mph)

$$\text{TS: } Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

$$= \frac{660.1 - 670}{25/\sqrt{20}} \approx -1.77$$

RR: $Z \le z_\alpha = z_{0.05} = -1.65$

Example2: Hyperloop One

1st. approach using z-critical value	2nd. approach using *p*-value
	
qnorm(0.05,0,1) [1] -1.644854	pnorm(-1.77,0,1) 0.03836357
	p-value = $P(Z \leq -1.77)$ $=0.0384$ (from z-table)
Since the z-test = -1.77 is < the critical value of –1.65, we reject the H_0	The H_0 would be rejected because the *p*-value = 3.84% is < α = 5%

Conclusion: There is evidence to suggest that the true mean capsule speed is less than 670 mph

Example3: Attention Span

According to a study by Microsoft, the mean attention span of an adult has decreased from the mean of 12 sec found in 2000—some would say largely due to technology. A random sample of 18 adults was obtained, and the attention span of each was measured using a standardized test. The sample mean was 10.85 sec and s = 2.1 sec. Is there any evidence to suggest that the true mean attention span is less than 12 sec? Assume the underlying distribution is normal and use $\alpha = 0.05$.

$n = 18$
$s = 2.1$
$\bar{x} = 10.85$
$\alpha = 0.05$

H_0: $\mu = 12$ (The mean attention span of an adult equals 12 sec)

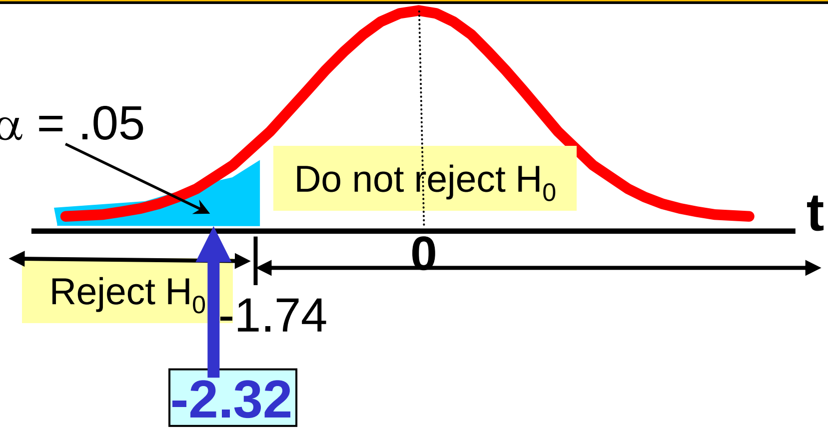
H_a: $\mu < 12$ (The mean attention span of an adult has decreased from the mean of 12 sec)

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{10.85 - 12}{2.1/\sqrt{18}} \approx -2.32$$

RR: $t \geq -t_{\alpha,n-1} = t_{0.05,17} = -1.7396$

Example3: Attention Span

1st. approach using z-critical value



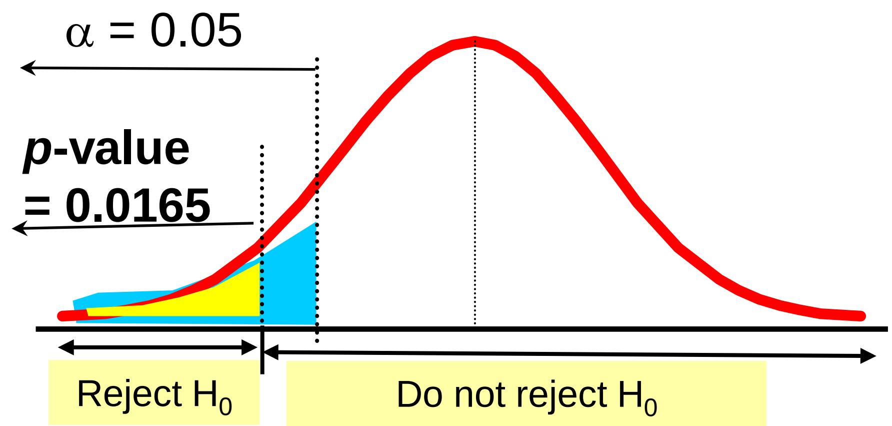
$v = n - 1 = 18 - 1 = 17$

qt(0.05, 17)

-1.739607

Because the t-test = -2.32 is < the critical value of –1.74, we reject the H_0

2nd. approach using *p*-value



pt(-2.32,17)

0.01651968 ≈ 0.0165

Finding the *p*-value from Table 5 is in the next slide

The H_0 would be rejected because the *p*-value ≈ 1.65% is < α = 5%

Conclusion: There is evidence to suggest that the mean attention span of adults has decreased from 12 sec.

p-Value Bounds

How to Bound the *p*-Value for a *t* Test

Suppose t is the value of the test statistic in a one-sided hypothesis test.

1. Select the row in Table 5 in the Appendix that corresponds to $n - 1$, the number of degrees of freedom associated with the test.
2. Place $|t|$ in this ordered list of critical values.
3. To compute p:
   a) If $|t|$ is between two critical values in the $n - 1$ row, then the p-value is bounded by the corresponding significance levels.

   b) If $|t|$ is greater than the largest critical value in the $n - 1$ row, then $p < 0.0001$ (the smallest significance level in the table).

   c) If $|t|$ is less than the smallest critical value in the $n - 1$ row, then $p > 0.20$ (the largest significance level in the table).

$|t| = |-2.32| = 2.32$

In Table 5 in the Appendix, row $n - 1 = 18 - 1 = 17$, place 2.32 in the ordered list of critical values. $2.1098 \leq 2.32 \leq 2.5669$

$t_{0.025,17} \leq 2.32 \leq t_{0.01,17}$

Therefore, $0.010 \leq p \leq 0.025$ from R: $p \approx 0.0165$

Example4:

A random sample of **25** boxes of painkiller pills from a Normal population distribution showed that the sample mean is 372.5 mg. Does the painkiller box contain 368 mg pills? The **σ = 15 mg.** Use α= 0.05.

$n = 25$
$\sigma = 15$
$\bar{x} = 372.5$
$\alpha = 0.05$

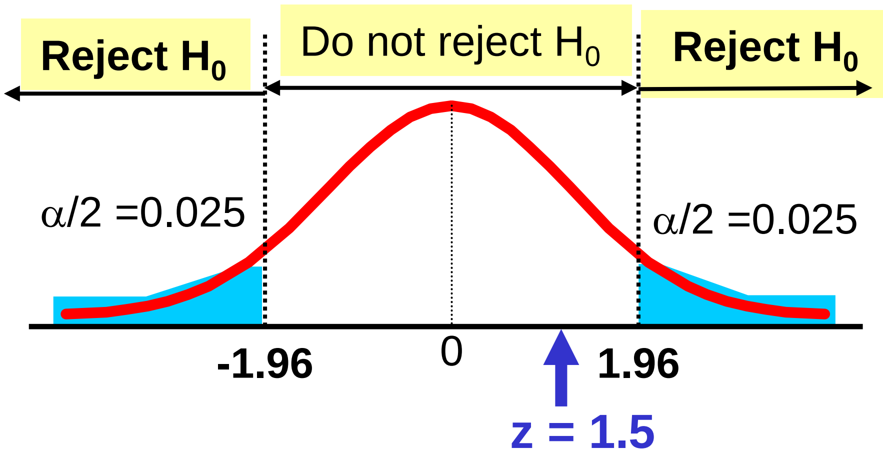
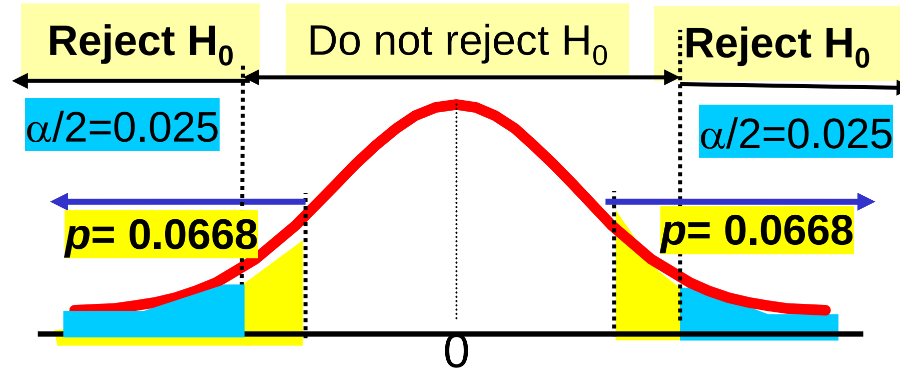
H_0: $\mu = 368$ (The painkiller box contains 368 mg pills.)

H_1: $\mu \neq 368$ (The painkiller box does not contain 368 mg pills.)

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{372.5 - 368}{15 / \sqrt{25}} = 1.5$$

RR: $|Z| \geq z_{\alpha/2} = z_{0.025} = 1.96$

Example4:

1st. approach using z-critical value	2nd. approach using *p*-value
	

1st. approach using z-critical value

qnorm(0.025,0,1)
-1.959964

Since the z-test value = 1.5 is < the critical value =1.96, we do not reject H_0

2nd. approach using *p*-value

pnorm(1.5,0,1)
0.9331928

pnorm(-1.5,0,1)
0.0668072

p-value = 2(1 - 0.9332) = 0.1336

The H_0 would be not rejected because the *p*-value = 13.36% is > α = 5%

Conclusion: The box of painkiller pills contains 368mg pills based on the observed data