Introduction to Hypothesis Testing

- Statistics, it allows us to measure behavior in samples to learn more about the behavior in populations that are often too large or inaccessible.

- For example, suppose the average score on a standardized exam in a given population is 1000.

- The sample mean as an unbiased estimator of the population mean—if we selected a random sample from a population, then on average the value of the sample mean will equal the population mean.

- In our example, if we select a random sample from this population with a mean of 1000, then on average, the value of a sample mean will equal to 1000.

- We begin by stating the value of a population mean, and then we select a sample and measure the mean in that sample. On average, the value of the sample means will equal the population mean.

- The method in which we select samples to learn more about characteristics in a given population is called hypothesis testing.

- Hypothesis testing is really a systematic way to test claims or ideas about a group or population.

- Suppose, assume that there is a statement that children in India watch an average of 10 hours of TV per week. To test whether this claim is true, we record the time (in hours) that a group of 200 children (the sample), among all children in India (the population), watch TV. The mean we measure for these 200 children is a sample mean. We can then compare the sample mean with the population mean.

- **Hypothesis testing or significance testing** is a method for testing a claim or hypothesis about a parameter in a population, using data measured in a sample.

- **The method of hypothesis testing can be summarized in four steps.**

1. To begin, we identify a hypothesis or claim that we feel should be tested. For example, we might want to test the claim that the mean number of hours that children in India watch TV is 10 hours.

2. We select a criterion upon which we decide that the claim being tested is true or not. For example, the claim is that children watch 10 hours of TV per week. Most samples we select should have a mean close to or equal to 10 hours if the claim we are testing is true. So at what point do we decide that the discrepancy between the sample mean and 10 is so big that the claim we are testing is likely not true? We answer this question in this step of hypothesis testing.

3. Select a random sample from the population and measure the sample mean. For example, we could select 200 children and measure the mean time (in hours) that they watch TV per week.

4. Compare what we observe in the sample to what we expect to observe if the claim we are testing is true. We expect the sample

mean to be around 10 hours. If the discrepancy between the sample mean and population mean is small, then we will likely decide that the claim we are testing is indeed true. If the discrepancy is too large, then we will likely decide to reject the claim as being not true.

FOUR STEPS TO HYPOTHESIS TESTING

The goal of hypothesis testing is to determine the likelihood that a population parameter, such as the mean, is likely to be true.

There are four steps of hypothesis testing.

Step 1: State the hypotheses.

Step 2: Set the criteria for a decision.

Step 3: Compute the test statistic.

Step 4: Make a decision.

Step 1: State the hypothesis. Begin by stating the value of a population mean in a null hypothesis and assume that it is true. For the children watching TV example, we state the

H_0 (null hypothesis) = "children in India watch an average of 10 hours of TV per week".

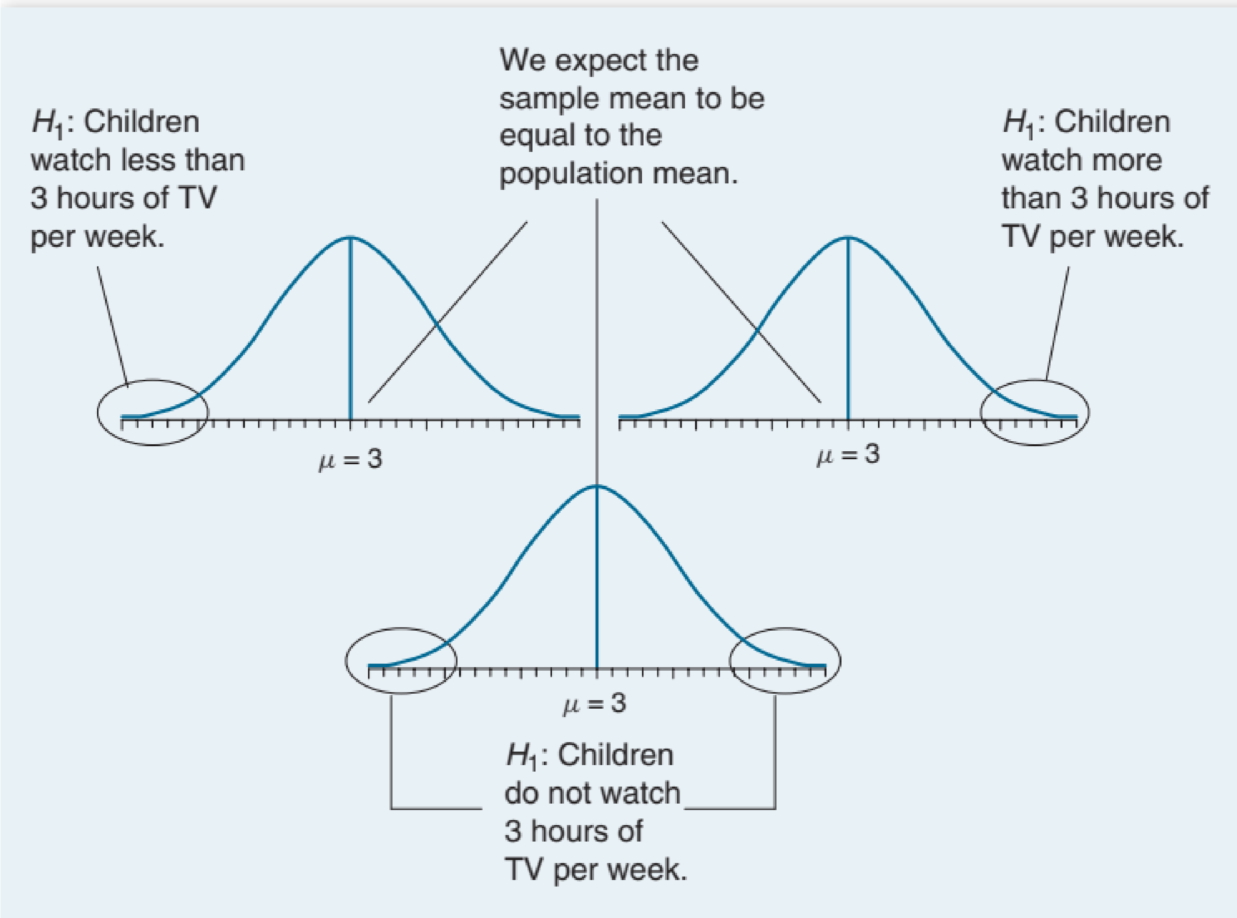
Keep in mind that the only reason we are testing the null hypothesis is because we think it may be wrong. State what is wrong about the null hypothesis as an alternative hypothesis.

For the children watching TV example, we may have reason to believe that children watch more than (>) or less than (<) 10 hours of TV per week.

When we are uncertain of the direction, we can state that the value in the null hypothesis is not equal to (≠) 10 hours.

"An alternative hypothesis (H1) is a statement that directly contradicts a null hypothesis by stating that that the actual value of a population parameter is less than, greater than, or not equal to the value stated in the null hypothesis"

Step 2: Set the criteria for a decision. To set the criteria for a decision, state the **level of significance** for a test. The level of significance is typically set at **5%** in behavioral research studies. But, this is not fixed; it varies depends on the studies.



Step 3: Compute the test statistic. [Z | T | x^2 | F]

Specifically, a test statistic tells us how far, or how many standard deviations, a sample mean is from the population mean. The larger the value of the test statistic, the further the distance, or number of standard deviations, a sample mean is from the population mean stated in the null hypothesis. The value of the test statistic is used to make a decision in Step 4.

Step 4: Make a decision. Use the value of the test statistic to make a decision about the null hypothesis. The decision is based on the probability of obtaining a sample mean, given that the value stated in the null hypothesis is true. If the probability of obtaining a sample mean is less than 5% when the null hypothesis is true, then the decision is to reject the null hypothesis. If the probability of obtaining a sample mean is greater than 5% when the null hypothesis is true, then the decision is to retain the null hypothesis.

In summary, there are two decisions one can make:

1. Reject the null hypothesis. The sample mean is associated with a low probability of occurrence when the null hypothesis is true.

2. Retain the null hypothesis. The sample mean is associated with a high probability of occurrence when the null hypothesis is true.

- The probability of obtaining a sample mean, given that the value stated in the null hypothesis is true, is stated by the p value. The p value is a probability: It varies between 0 and 1 and can never be negative.

- When the p value is less than or equal to 5% ($p \leq 0.05$), reject the null hypothesis. When the p value is greater than 5% ($p > 0.05$), retain the null hypothesis.

- The decision to reject or retain the null hypothesis is called significance.
  - When the p value is less than 0.05, we reach significance; the decision is to reject the null hypothesis.
  - When the p value is greater than 0.05, we fail to reach significance; the decision is to retain the null hypothesis

MAKING A DECISION: TYPES OF ERROR

In Step 4, we decide whether to retain or reject the null hypothesis. Because we are observing a sample and not an entire population, it is possible that our conclusion may be wrong. The table given below shows that there are four decision alternatives regarding the truth and falsity of the decision we make about a null hypothesis.

1. The decision to retain the null hypothesis could be correct.
2. The decision to retain the null hypothesis could be incorrect.
3. The decision to reject the null hypothesis could be correct.
4. The decision to reject the null hypothesis could be incorrect.

■Type I Error

- Reject a true null hypothesis
- A type I error is a "false alarm"
- The probability of a Type I Error is α

■Type II Error

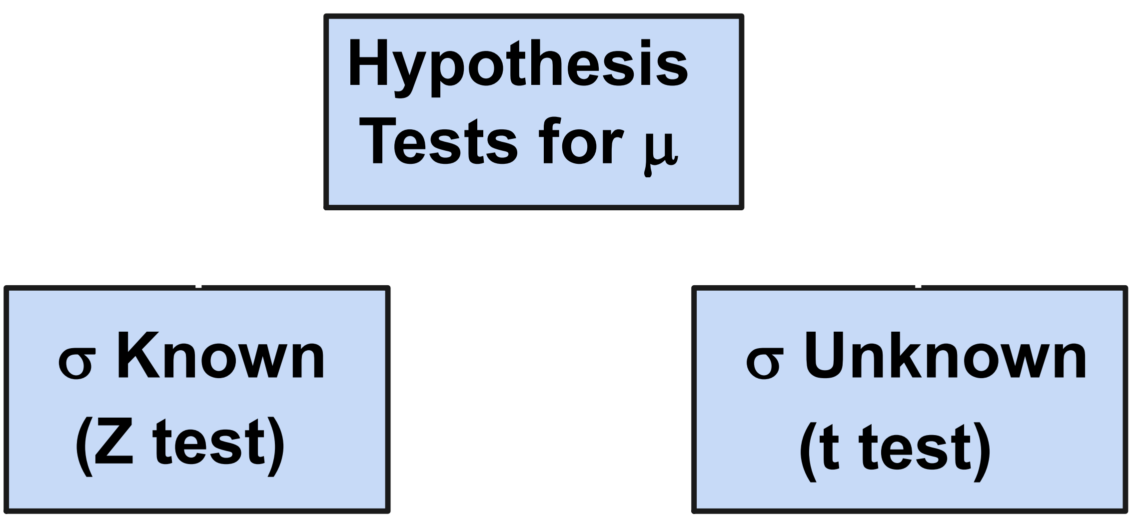
- Failure to reject a false null hypothesis
- Type II error represents a "missed opportunity"
- The probability of a Type II Error is β

Possible Hypothesis Test Outcomes		
	Actual Situation	
Decision	H_0 True	H_0 False
Do Not Reject H_0	No Error Probability $1 - \alpha$	Type II Error Probability β
Reject H_0	Type I Error Probability α	No Error Power $1 - \beta$

The incorrect decision is - accepting a false null hypothesis. This decision is an example of a Type II error, or β error. With each test we make, there is always some probability that the decision could be a Type II error.

The incorrect decision is - rejecting a true null hypothesis. This decision is an example of a Type I error. With each test we make, there is always some probability that our decision is a Type I error.

"The correct decision is to reject a false null hypothesis"



Z Test of Hypothesis for the Mean (σ Known)

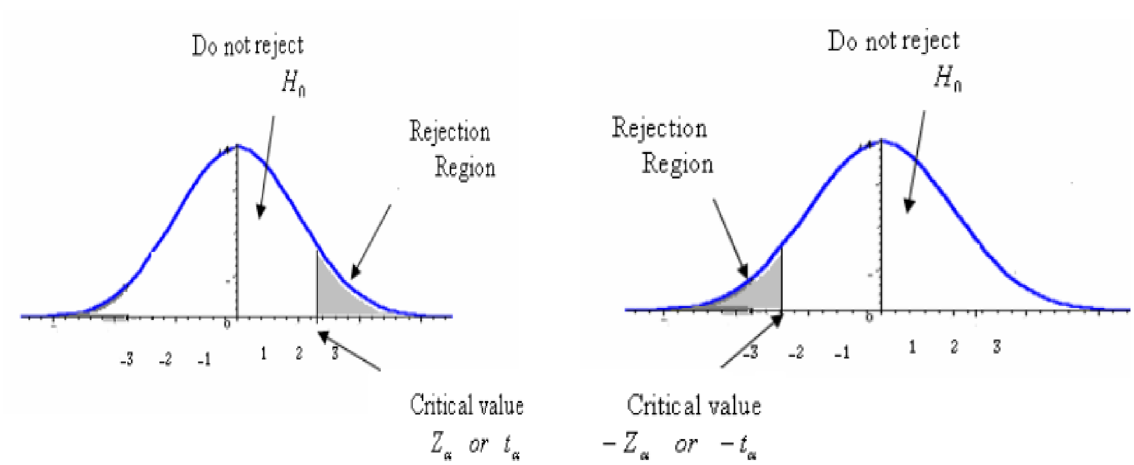
Hypothesis Tests for μ

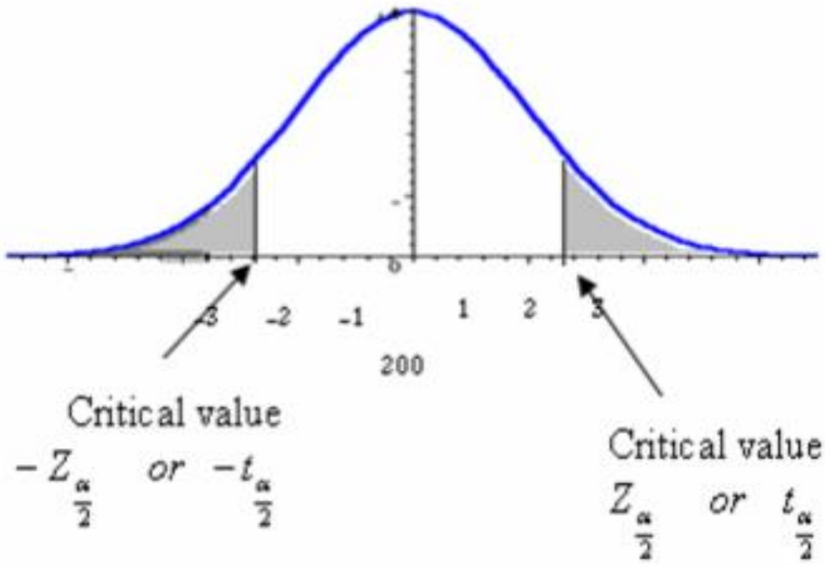
σ Known (Z test)

$$Z_{STAT} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

σ Unknown (t test)

- Convert sample statistic ($\bar{X}$) to test statistic (Z_{STAT})
- Determine the critical Z values for a specified level of significance α from a table or by using computer software
- Decision Rule: If the test statistic falls in the rejection region, reject H_0 ; otherwise do not reject H_0





Note (σ Unknown)

- If the population standard deviation is unknown, you instead use the sample standard deviation S.
- Because of this change, you use the t distribution instead of the Z distribution to test the null hypothesis about the mean.
- When using the t distribution you must assume the population you are sampling from follows a normal distribution.

Types of test	**Z**	**t**
The one – tailed test (Right)	Z_{α}	$t_{(\alpha,\ n-1)}$
The one – tailed test (left)	$-Z_{\alpha}$	$-t_{(\alpha,\ n-1)}$
The two – tailed test	$\pm Z_{\frac{\alpha}{2}}$	$\pm t_{\left(\frac{\alpha}{2},\ n-1\right)}$

Sample Problem - 1

A government agency reported that the population mean score on the quantitative portion of the Graduate Record Examination (GRE) General Test for students taking the exam between 2014 and 2017 was 558 ± 139 ($\mu \pm \sigma$). Suppose we select a sample of 100 participants ($n = 100$). We record a sample mean equal to 585 ($M = 585$). Compute the one–independent sample z test for whether or not we will retain the null hypothesis ($\mu = 558$) at 0.05 level of significance ($\alpha = 0.05$).

Step 1: State the hypotheses. The population mean is 558, and are testing whether the null hypothesis is (=) or is not (≠) correct:

$H0$: $\mu = 558$ Mean test scores are equal to 558 in the population.

$H1$: $\mu \neq 558$ Mean test scores are not equal to 558 in the population.

Step 2: Set the criteria for a decision. The level of significance is 0.05, which makes the alpha level $\alpha = 0.05$. To locate the probability of obtaining a sample mean from a given population, use the standard normal distribution.

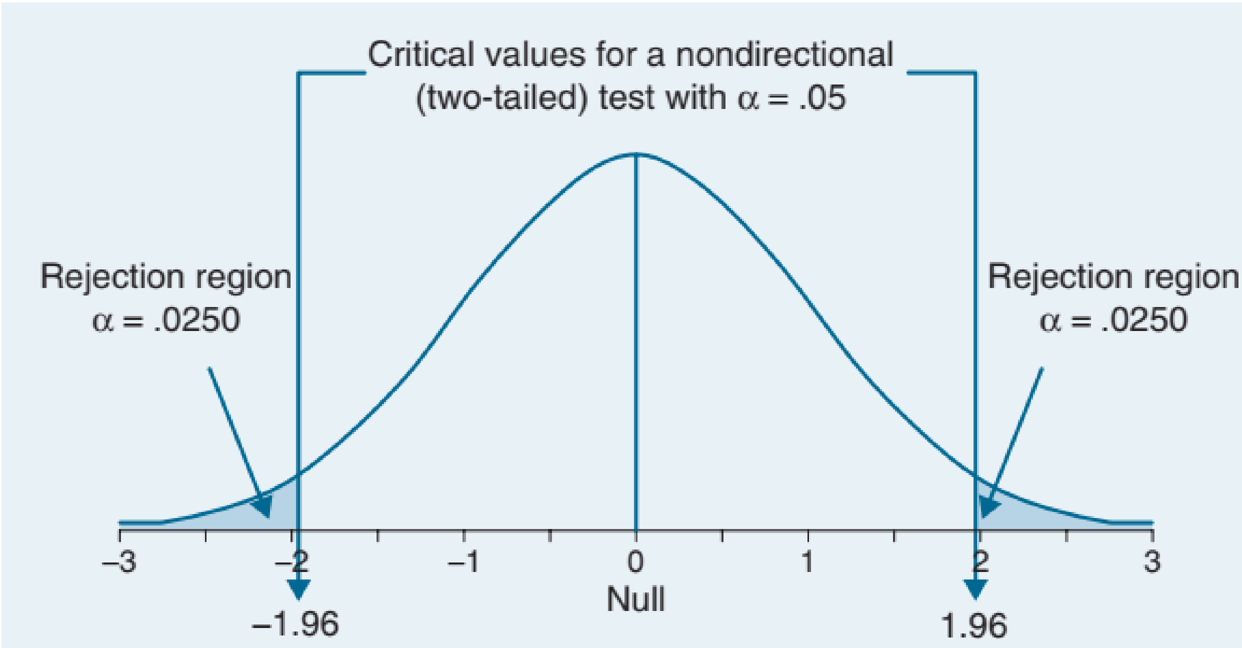
Locate the z scores in a standard normal distribution that are the cutoffs, or critical values, for sample mean values with less than a 5% probability of occurrence if the value stated in the null ($\mu = 558$) is true.

In a non-directional [two-tailed test], divide the alpha value in half so that an equal proportion of area is placed in the upper and lower tail.

The following table gives the critical values for one- and two-tailed tests at a 0.05, 0.01, and 0.001 level of significance.

	Type of Test	
Level of Significance (α)	One-Tailed	Two-Tailed
0.05	+1.645 or −1.645	±1.96
0.01	+2.33 or −2.33	±2.58
0.001	+3.09 or −3.09	±3.30

The following graph shows the critical values for one- and two-tailed tests.



The value, 0.025, is listed for a z-score equal to z= 1.96. This is the critical value for the upper tail of the standard normal distribution. Since the normal distribution is symmetrical, the critical value in the bottom tail will be the same distance below the mean, or z = –1.96. The regions beyond the critical values, displayed in the above graph, are called the rejection regions. If the value of the test statistic falls in these regions, then the decision is to reject the null hypothesis; otherwise, retain the null hypothesis.

Step 3: Compute the test statistic.

The z statistic formula is the sample mean minus the population mean stated in the null hypothesis, divided by the standard error of the mean.

$Z_{Stat} = \frac{M - \mu}{\sigma M}$ Where $\sigma M = \frac{\sigma}{\sqrt{n}} = \frac{139}{\sqrt{100}} = 13.9$

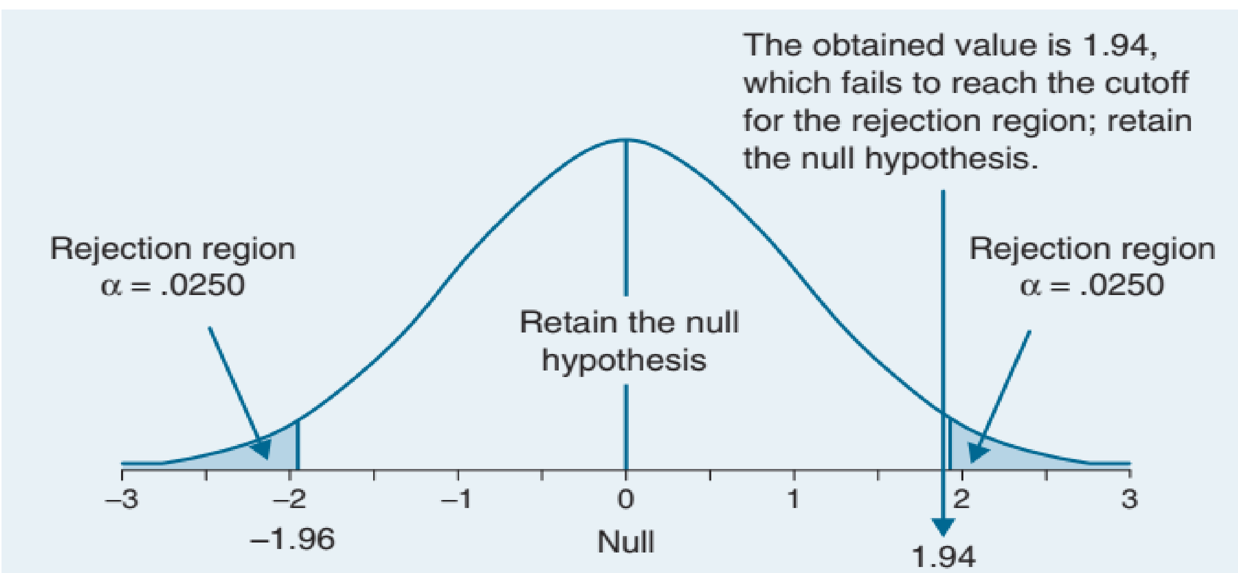
$Z_{Stat} = \frac{585 - 558}{13.9} = 1.942$

Step 4:

Compare the obtained value to the critical values. Reject the null hypothesis if the obtained value exceeds a critical value.

But, referring the figure, it shows that the obtained value (Zstat = 1.94) is less than the critical value; it does not fall in the rejection region.

The decision is to accept the null hypothesis.



Therefore, conclude that the mean score on the GRE General Test in this population is 558 (the value stated in the null hypothesis).

Sample problem – 2

A government agency reported that the population mean score on the quantitative portion of the Graduate Record Examination (GRE) General Test for students taking the exam between 2014 and 2017 was 558 ± 139 ($\mu \pm \sigma$). Suppose, we select a sample of 100 students enrolled in an elite private school ($n = 100$). We hypothesize that students at this elite school will score higher than the general population. We record a sample mean equal to 585 ($M = 585$), same as measured in the above problem. Compute the one–independent sample z test [single tailed] at a 0.05 level of significance.

Step 1: State the hypotheses. The population mean is 558, and we are testing whether the alternative is greater than (>) this value.

H0: $\mu = 558$ Mean test scores are equal to 558 in the population of students at the elite school.

H1: $\mu > 558$ Mean test scores are greater than 558 in the population of students at the elite school.

Step 2: Set the criteria for a decision. The level of significance is 0.05.

Step 3:

$Z_{Stat} = \frac{M-\mu}{\sigma M}$ Where $\sigma M = \frac{\sigma}{\sqrt{n}} = \frac{139}{\sqrt{100}} = 13.9$

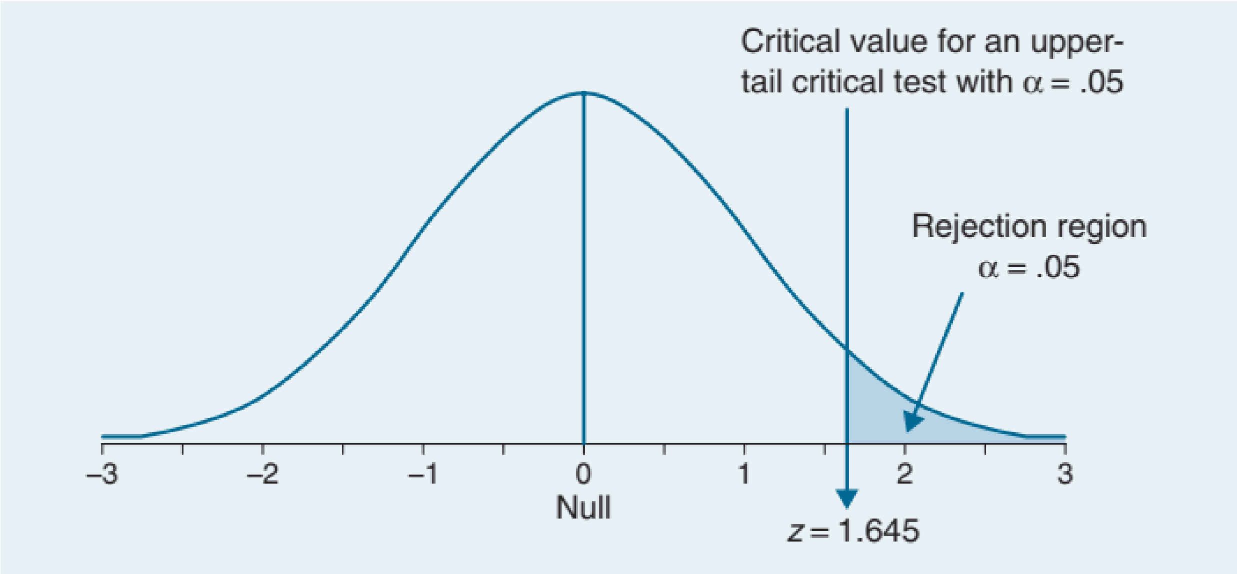
$Z_{Stat} = \frac{585-558}{13.9} = 1.942$

Step 4:

Compare the obtained value to the critical values. Reject the null hypothesis if the obtained value exceeds a critical value.

Referring the table and the graph, that the obtained value (Zstat = 1.942) is greater than the critical value [1.645]; it falls in the rejection region. Therefore, the decision is to reject the null hypothesis.

Level of Significance (α)	Type of Test	
	One-Tailed	Two-Tailed
0.05	+1.645 or −1.645	±1.96
0.01	+2.33 or −2.33	±2.58
0.001	+3.09 or −3.09	±3.30



Based on this, it is concluded that the mean score on the GRE General Test in this population is not 558.

Sample problem – 3

A government agency reported that the population mean score on the quantitative portion of the Graduate Record Examination (GRE) General Test for students taking the exam between 2014 and 2017 was 558 ± 139 ($\mu \pm \sigma$). Suppose we select a sample of 100 students enrolled in a school with low funding and resources ($n = 100$). We hypothesize that students at this school will score lower than the general population. We record a sample mean equal to 585 ($M = 585$), same as measured in the above problem. Compute the one-independent sample z test [single tailed] at a 0.05 level of significance.

Step 1: State the hypotheses. The population mean is 558, and we are testing whether the alternative is less than (<) this value.

H0: μ = 558 Mean test scores are equal to 558 in the population of students at the normal school.

H1: μ < 558 Mean test scores are less than 558 in the population of students at the normal school.

Step 2: Set the criteria for a decision. The level of significance is 0.05.

Step 3:

$Z_{Stat} = \frac{M-\mu}{\sigma M}$ Where $\sigma M = \frac{\sigma}{\sqrt{n}} = \frac{139}{\sqrt{100}} = 13.9$

$Z_{Stat} = \frac{585-558}{13.9} = 1.942$

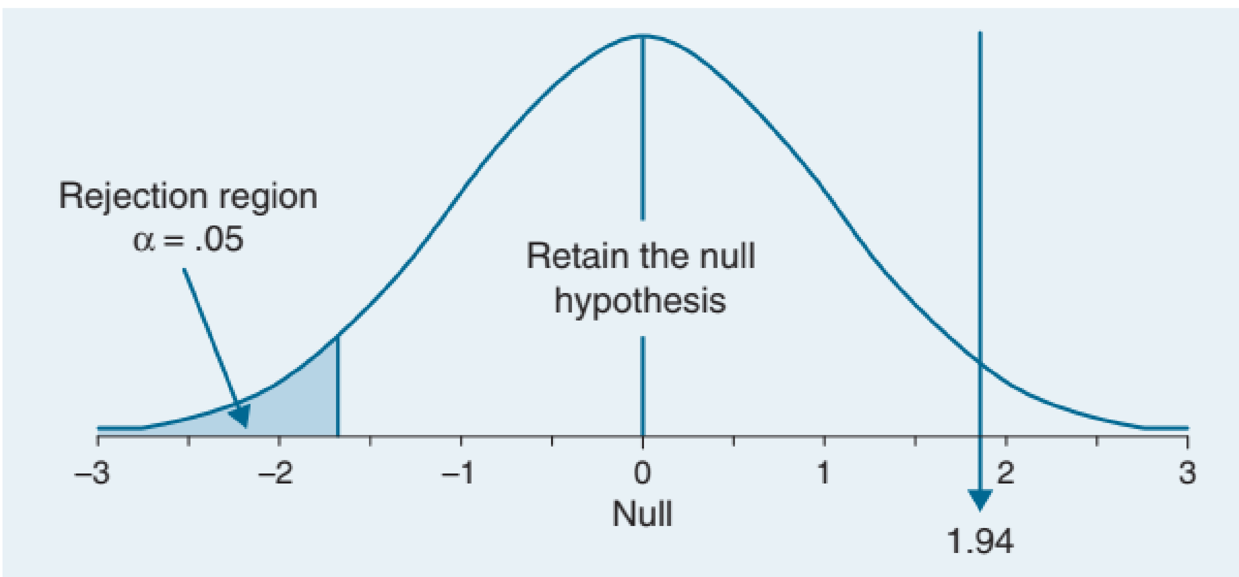
Step 4:

Compare the obtained value to the critical values. Reject the null hypothesis if the obtained value exceeds a critical value.

Referring the table and the graph, that the obtained value (Zstat = 1.942) is greater than the critical value [1.645]; The graph shows that the obtained value

(Zobt = +1.94) does not exceed the critical value. Instead, the value obtained is located in the opposite tail. The decision is to retain the null hypothesis. Therefore, the decision is to accept the null hypothesis.

Level of Significance (α)	Type of Test	
	One-Tailed	**Two-Tailed**
0.05	+1.645 or −1.645	±1.96
0.01	+2.33 or −2.33	±2.58
0.001	+3.09 or −3.09	±3.30



Therefore, conclude that the mean score on the GRE General Test in this population is 558 (the value stated in the null hypothesis).

Exercises

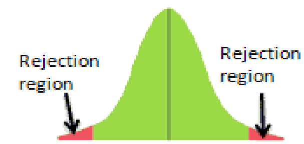
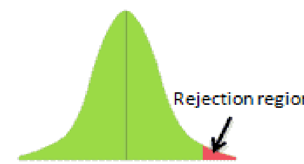
1. A researcher conducts a one–independent sample z test. The z statistic for the upper-tail critical test at a .05 level of significance was $Z_{Stat} = 1.84$. What is the decision for this test?

2. A researcher conducts a hypothesis test and finds that the probability of selecting the sample mean is p = .0689; if the value stated in the null hypothesis is true. What is the decision for a hypothesis test at a .05 level of significance?

3. A coaching institute claims that the students' mean scores in their institute are greater than the 82 marks with a standard deviation of 20. A sample of 81 students is selected, and the mean score is 90 marks. At 95% confidence level, is there enough evidence to support the claim?

4. A sample of 100 students is taken from a large population. The mean height of the students in this sample is 160 cm. Can it be reasonably regarded that, in the population, the mean height is 165 cm and the SD is 10 cm?

5. The mean breaking strength of the cables supplied by a manufacturer is 1800 with an SD of 100. By a new technique in the manufacturing process, it is claimed that the breaking strength of the cable has increased. To test this claim, a sample of 50 cables is tested and it found that the mean breaking strength is 1850. Can we support this at 1% level of significance?

Z- Test	Null Hypothesis (H_0)	Alternative Hypothesis (H_1)	Statistical conclusion
Two-tailed	$\mu = \mu_0$	$\mu \neq \mu_0$	
Left-tailed	$\mu \geq \mu_0$	$\mu < \mu_0$	
Right-tailed	$\mu \leq \mu_0$	$\mu > \mu_0$	

Critical values for Z.
A two-tailed test at the 5% level has a critical boundary Z score of +1.96 and -1.96
A one-tailed test at the 5% level has a critical boundary Z score of +1.64 or -1.64
A two-tailed test at the 1% level has a critical boundary Z score of +2.58 and -2.58
A one-tailed test at the 1% level has a critical boundary Z score of +2.33 or -2.33.

T – Test Statistic – Hypothesis Testing

$$T_{stat} = \frac{M-\mu}{\sigma M} \quad \text{where} \quad \sigma M = \frac{\sigma}{\sqrt{n}}$$

M = Sample Mean

μ = Population Mean

n = Sample Size

σ = Population Standard Deviation

- If the population standard deviation is not known; use sample standard deviation if known. Because of this change use T – test statistic instead of Z – test statistic.
- If both population and sample standard deviations are not known; then compute the sample standard deviation using the following formula

Population	Sample
$\sigma = \sqrt{\frac{\Sigma(x_i-\mu)^2}{n}}$	$S = \sqrt{\frac{\Sigma(x_i-\overline{x})^2}{n-1}}$
μ - Population Average **xi - Individual Population Value** **n - Total Number of Population**	**$\overline{x}$ - Sample Average** **x_i - Individual Population Value** **n - Total Number of Sample**

Critical values for t – test statistic
[Level of significance, Degrees of Freedom]

Types of test	Z	t
The one – tailed test (Right)	Z_{α}	$t_{(\alpha,\ n-1)}$
The one – tailed test (left)	$-Z_{\alpha}$	$-\ t_{(\alpha,\ n-1)}$
The two – tailed test	$\pm Z_{\frac{\alpha}{2}}$	$\pm t_{\left(\frac{\alpha}{2},\ n-1\right)}$

Sample Problem - 1

Suppose you are buyer of large supplies of light bulbs. You want to test at the 5% significant level, the manufacturer's claim that his bulbs last more than 800 hours; after testing 36 bulbs and found that the sample mean is 816 hours, with standard deviation of 70 hours. Should you accept the claim or not?

Solution

Here Population Mean, Sample Mean and Sample Standard Deviation are given. So, **t test** is used to find the result of the claim.

Ho: $\mu = 800$

H1: $\mu > 800$ [one tailed (right) test]

T_{stat} = $\frac{M-\mu}{\sigma M}$ where σM = $\frac{\sigma}{\sqrt{n}}$ = $\frac{70}{\sqrt{36}}$ = 11.67

T_{stat} = $\frac{816-800}{11.67}$ = 1.371

Level of Significance is 5%; i.e., t at 0.05 [LoS], 35 [DoF])

$t_{(35,\ 0.05)}$ = 1.690 [from the t – table]

T_{stat} = 1.371 < $t_{(35,\ 0.05)}$ = 1.690

Ho is accepted

Therefore, the manufacturer's claim that his bulbs last more than 800 hours is wrong.

Sample Problem – 2

The McFarland Insurance Company Claims Department reports the mean cost to process a claim is $60. An industry comparison showed this amount to be larger than most of the other insurance companies. The company instituted cost-cutting measures. To evaluate the effect of the cost-cutting measures, the Supervisor of the Claims Department selected a random sample of 26 claims processed last month. The sample information is reported as: $\bar{X}$ = 56.42, S = 10.04; n = 26. At 0.01 significance level is it reasonable that a claim is now less than $60?

Solution

Here Population Mean = 60, Sample Mean = 56.42 and Sample Standard Deviation = 10.04 and n =26 are given. So, **t test** is used to find the result of the claim.

Ho: $\mu \geq 60$

H1: $\mu < 60$ [one tailed (left) test]

$$T_{stat} = \frac{M-\mu}{\sigma M} \text{ where } \sigma M = \frac{\sigma}{\sqrt{n}} = \frac{60}{\sqrt{26}} = 1.968$$

$$T_{stat} = \frac{56.42-60}{1.968} = -1.819$$

Level of Significance is 1%; i.e., t at 0.01 [LoS], 25 [DoF])

$t_{(25,\ 0.01)} = 2.485$ [from the t – table]

Because it is left tailed test, because of the symmetry; $t_{(25,\ 0.01)} = -2.485$.

$T_{stat} = -1.819 < t_{(25,\ 0.01)} = -2.485$

Ho is accepted

It is concluded that, the company claim of mean cost to process an insurance claim is correct.

Sample Problem – 3

The following data represents hemoglobin values in gm/dl for 10 patients: 10.5, 9, 6.5, 8, 11, 7, 7.5, 8.5, 9.5, 12.

Is the mean value for patients significantly differ from the mean value of general population (12 gm/dl). Evaluate the role of chance.(α = 0.05).

Here, data collected for 10 patients and population mean is given. Standard Deviations are not known. But, still it is possible to compute Sample Mean and Standard Deviation.

Sample

$$S = \sqrt{\frac{\Sigma(x_i-\overline{x})^2}{n-1}}$$

$\overline{x}$ - Sample Average
x_i - Individual Population Value
n - Total Number of Sample

Sample mean $\bar{x}$ =

$$\frac{\sum x}{n} = \frac{10.5 + 9 + 6.5 + 8 + 11 + 7 + 7.5 + 8.5 + 9.5 + 12}{10} = 8.95$$

x	$\bar{x}$	$(x-\bar{x})^2$
10.5	8.95	2.4025
9		0.0025
6.5		6.0025
8		0.9025
11		4.2025
7		3.8025
7.5		2.1025
8.5		0.2025
9.5		0.3025
12		9.3025
Total		29.225

Sample Standard Deviation (s) = $\sqrt{\frac{29.225}{9}}$ = 1.802

Ho: μ = 12

H1: μ $\neq$ 12

Therefore, consider it as two tailed test.

$$T_{stat} = \frac{\bar{x}-\mu}{\sigma M} \text{ where } \sigma M = \frac{s}{\sqrt{n}} = \frac{1.802}{\sqrt{10}} = \mathbf{0.5698}$$

$$T_{stat} = \frac{\overline{8.95}-12}{0.5698} = \mathbf{-5.3528}$$

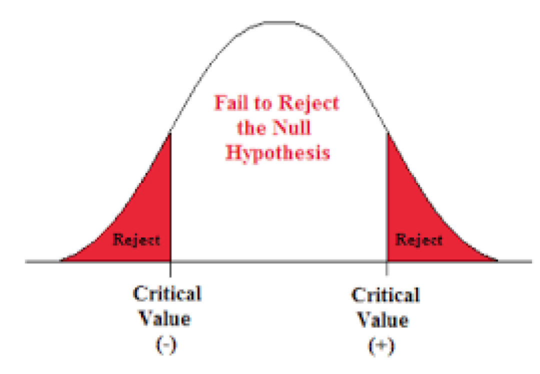
Because of two tailed test, t at $\frac{0.05}{2}$[LoS], 9 [DoF])

Level of Significance is 0.025%; i.e., t at 0.025 [LoS], 9 [DoF])

$t_{(9,\ 0.025)}$ = 2.262 [from the t – table]

T_{stat} = -5.3528 > $t_{(9,\ 0.025)}$ = -2.262   [two tailed]

The calculated value > $t_{(9,\ 0.025)}$ = -2.262



Therefore, Ho is rejected.

It is concluded that there is a statistically significant difference between the mean of sample and population mean.

Exercise Problem:

Based on field experiments, a new variety green gram is expected to give a yield of 12.0 quintals per hectare. The variety was tested on 10 randomly selected farmers' fields. The yield (quintals/hectare) were recorded as 14.3, 12.6, 13.7, 10.9, 13.7, 12.0, 11.4, 12.0, 12.6, 13.1. Do the results conform the expectation?

Assume that LoS = 0.05

Population Mean = 12

N = 10

Ho: $\mu = 12$

H1: $\mu \neq 12$ [two tailed test]

Calculated Values:

Sample Mean = 12.63

Sample SD = 1.0853

σM = 0.3432

t_{stat} = 1.836

Because of two tailed test, t at $\frac{0.05}{2}$[LoS], 9 [DoF])

Level of Significance is 0.025%; i.e., t at 0.025 [LoS], 9 [DoF])

$t_{(9,\ 0.025)}$ = 2.262 [from the t – table]

Since, t_{stat} = 1.836 < $t_{(9,\ 0.025)}$ = 2.262

Ho is accepted

It is conclude that the new variety of green gram will give an average yield of 12 quintals/hectare.

Exercise Problems

1. The mean lifetime of a sample of 25 bulbs is found as 1550 hours, with a SD of 120 hours. The company manufacturing the bulbs claims that the average life of their bulbs is 1600 hours. Is the claim acceptable at 5%?

2. A random sample of 10 boys had the following IQs: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Does the data support the assumption of a population mean IQ of 100?

3. A certain injection administered to each of the 12 patients resulted in the following increase of Blood Pressure: 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4. Can it be concluded that the injection will be, in general accompanied by an increase in BP?