Chapter 6
Hypothesis Testing – Single Population

L1 – Introduction to hypothesis testing

Learning Outcomes:

At the end of the lesson student should be able to

- Describe the hypothesis testing;
- define what is types of error in hypothesis testing;
- Describe critical region and p- value in hypothesis testing;
- Describe the procedure for hypothesis testing.

HYPOTHESIS TESTING

- Hypothesis: a statement about some parameters (e.g mean, variance, proportion) of one or more populations.

Example:

- the claim that $\mu = 0.75$, where μ is the true average inside diameter of a certain type of PVC pipe.
- p < 0.10 , where p is the proportion of defective items in an assembly line

- Hypothesis Testing problem:

  two contradictory hypotheses under consideration.

Example:

the claim that $\mu = 0.75$ and the other $\mu > 0.75$

Objective :

- to decide, based on sample information, which of the two hypotheses is correct.

Hypothesis testing:
decision-making procedure about the *null hypothesis*

- The Null Hypothesis (H_0):
the hypothesis that cannot be viewed as *false* unless sufficient evidence on the contrary is obtained.

- The Alternative Hypothesis (H_1):
the hypothesis *against* which the null hypothesis is tested and is viewed true when H_0 is declared false.

Example

- Consider the burning rate of a solid propellant. Suppose we are interested in deciding whether or not the mean burning rate is 50cm/sec. The hypothesis is

$$H_0 : \mu = 50 \ \ cm / \sec$$
$$H_1 : \mu \neq 50 \ \ cm / \sec$$

- The alternative hypothesis specifies values of μ that could be either greater or less than 50 cm/sec, it is called a **two-sided alternative hypothesis.**

Hypothesis Test

- Hypothesis-testing procedures rely on using the information in a random sample from the population of interest.

- If this information is *consistent* with the hypothesis (H_0), then we will conclude that the hypothesis is **true**; if this information is *inconsistent* with the hypothesis, we will conclude that the hypothesis is **false.**

Three types of Hypothesis Test:

- One sided (tailed) lower-tail test

$$\boxed{\begin{aligned} &H_0 : \mu \geq \mu_0 \quad \text{or} \quad \mu = \mu_0 \\ &H_1 : \mu < \mu_0 \end{aligned}}$$

- One sided (tailed) upper-tail test

$$\boxed{\begin{aligned} &H_0 : \mu \leq \mu_0 \quad \text{or} \quad \mu = \mu_0 \\ &H_1 : \mu > \mu_0 \end{aligned}}$$

- Two sided (tailed) test

$$\boxed{\begin{aligned} &H_0 : \quad \mu = \mu_0 \\ &H_1 : \quad \mu \neq \mu_0 \end{aligned}}$$

Note: μ_0 is the value given/assumed for the parameter μ.

Types of Error

Test RESULT	TRUE STATE OF NATURE	
	H_0 True	**H_1 True (H_0 False)**
H_0 True	Correct decision	Type II error (probability is β)
H_1 True (reject H_0)	Type I error (probability is α)	Correct decision

Type I error: rejecting H_0 when it is actually true.
Type II error: not rejecting H_0 when it is actually false.

- The probability of making type I error is called the level of significance, α,

  $\alpha = P$ (Type I error)

  $= P$ (Rejecting the null hypothesis when in fact the null hypothesis is true)

- The probability of committing a Type II error is denoted by β, which is often difficult to determine precisely.

- A test procedure is specified by the following:

  - A test statistic, a function of the sample data on which the decision (reject the null hypothesis or do not reject the null hypothesis) is to be based.
  - Rejection region. (CRITICAL REGION)

    the set of all values of the test statistics which will *reject the null hypothesis*.

Rejection area (Critical region)

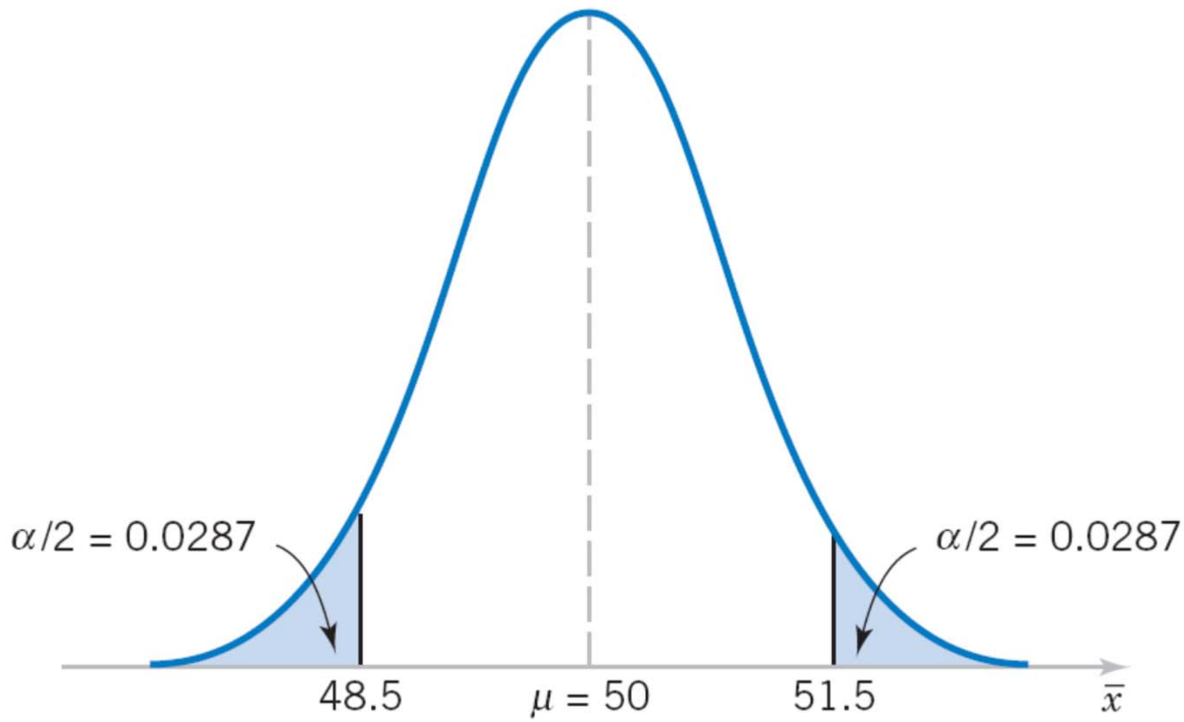


Figure 4-4 The critical region for H_0: $\mu = 50$ versus H_1: $\mu \neq 50$ and $n = 10$.

P – values in Hypothesis Testing

The ***P*-value** is the smallest level of significance that would lead to rejection of the null hypothesis H_0.

The p - value conveys information about the weight of evidence against H_0.

The *smaller* the p – value, the greater the evidence against H_0.

When the p – value is small enough we reject H_0 .

Steps in Hypothesis Testing

1. Identify the *parameter of interest*.

2. State the null hypothesis, H_0, and alternative hypothesis, H_1.
 (determine whether one sided or two sided test)

3. Determine the appropriate *sampling distribution* of the *sample test statistic*,
 E.g. Normal (z – statistic) or t-distribution.

4. Find the critical value.

5. Define the *rejection area (critical region)* based on *level of significance, α*.

6. Compute the sample test statistic.

7. Draw conclusion based on the decision rule (rejection area).

Chapter 6
Hypothesis Testing for Single Population

L2 - Determine the **confidence interval** for hypothesis test with **variance known.**

L3 – (2).Hypothesis Test about the Normal Mean, Variance **Unknown (t-test)**

Learning Outcomes

At the end of the lesson student should be able to

- Determine the **confidence interval** for hypothesis test with **variance known.**
- Perform hypothesis test about normal mean when variance is **unknown.**
  - Calculate the P-value of the above hypothesis test.

Definition – Confidence Interval

A confidence interval (or interval estimate) is a range (or an interval) of values used to estimate the true value of a population parameter. A confidence interval is sometimes abbreviated as CI.

Definitions

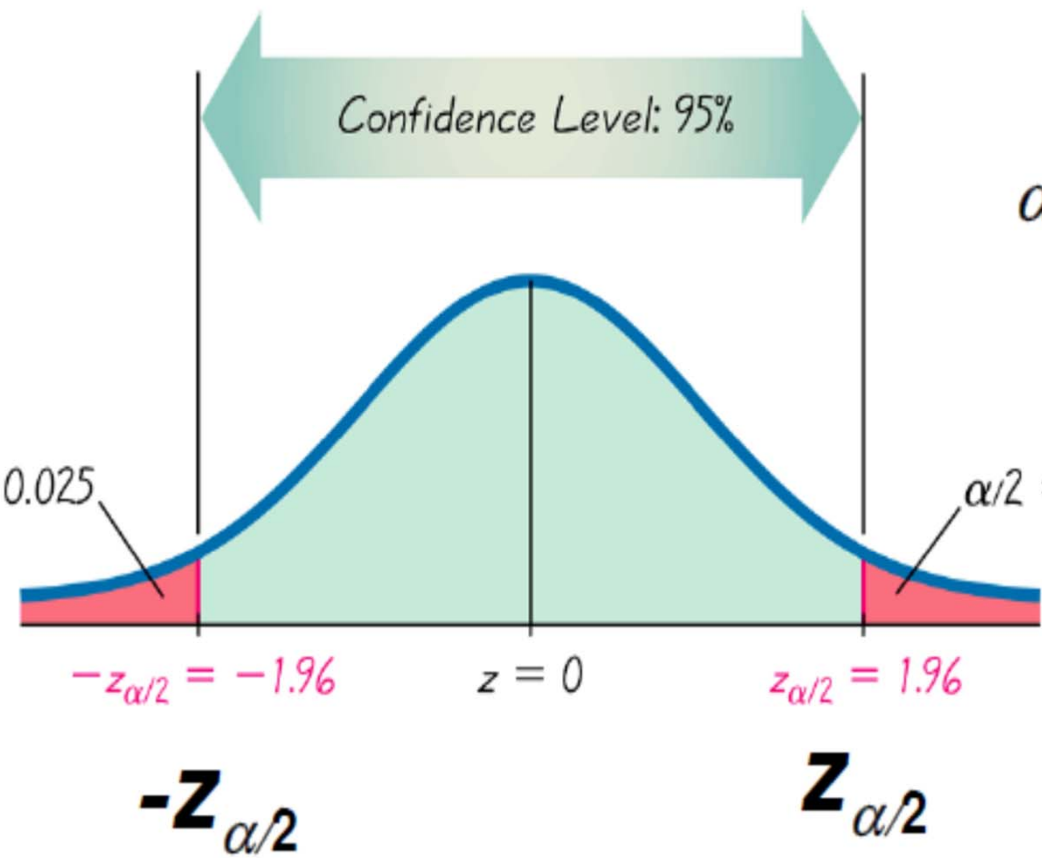
The probability that a confidence level will contain the estimated parameter is also called the confidence coefficient, designated by $1 - \alpha$ (often expressed in %).

Most common choices are 90%, 95%, or 99%.

(α = 10%), (α = 5%), (α = 1%)

A Confidence level of 95% ($\alpha = 0.05$) implies that 95% of all Samples would give an interval that includes the parameter that is being estimated, and only 5% of all sample would yield an erroneous interval.

Confidence Interval



α = 5%

$\alpha/2$ = 2.5% = .025

Next…….

- Now time for you do find the confidence interval → (…… , …)
- Find the Confidence Interval (CI) for a normal mean (μ) when variance (σ^2) is **known** – **(Z test)**

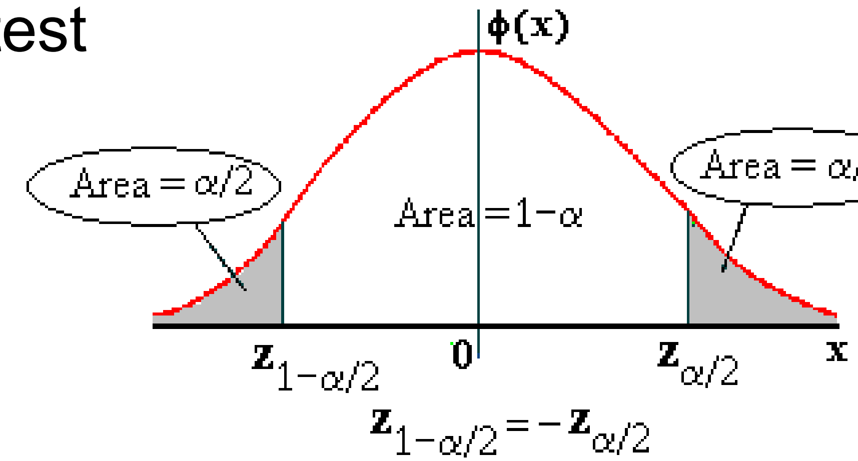
CI for normal mean when variance is **KNOWN**

A 100 (1- α)% confidence level for the mean μ of a normal Population when the value of σ is known is given by

Alternative Hypothesis, H_1	Reject H_0 IF μ_0 falls outside of the interval
Two tailed test, $\mu \neq \mu_0$	$\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
Upper tailed test, $\mu > \mu_0$	$-\infty \leq \mu \leq \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
Lower tailed test, $\mu < \mu_0$	$\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \infty$

CI FOR NORMAL MEAN WHEN VARIANCE IS **KNOWN**

Example – for two-tailed test

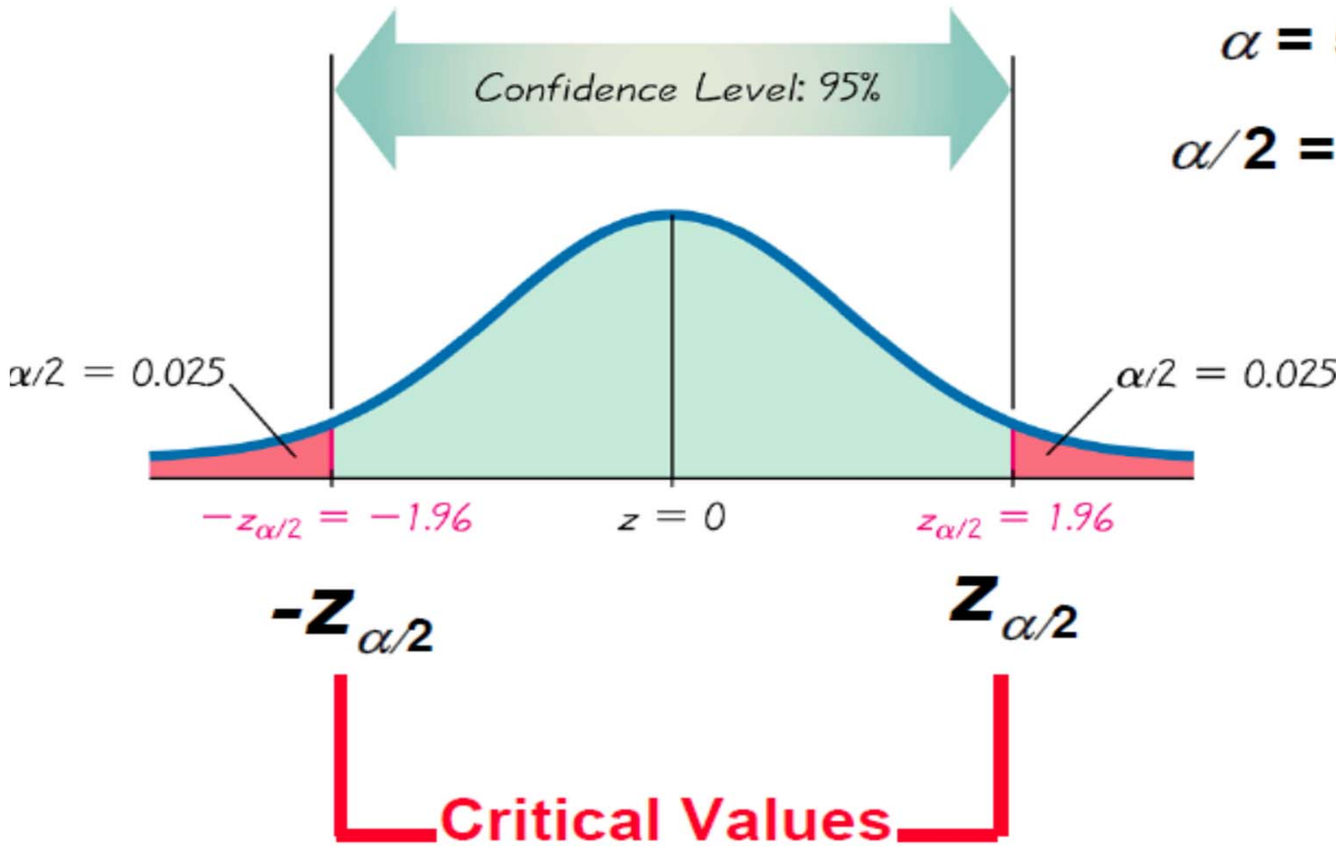


$$[\bar{x} - (z_{\alpha/2})(\frac{\sigma}{\sqrt{n}}), \bar{x} + (z_{\alpha/2})(\frac{\sigma}{\sqrt{n}})]$$

Or equivalently

$$\bar{x} - (z_{\alpha/2})(\frac{\sigma}{\sqrt{n}}) \le \mu \le \bar{x} + (z_{\alpha/2})(\frac{\sigma}{\sqrt{n}})$$

Finding $z_{\alpha/2}$ for a 95% Confidence Level



α = 5%

$\alpha/2$ = 2.5% = .025

Example 1

Standards set by government agencies indicate that Malaysians should not exceed the average daily sodium intake of 3300 milligrams (mg). The daily sodium intake is assumed to be normally distributed with standard deviation of 1100 mg. To find out whether Malaysians are exceeding this limit, a sample of 100 Malaysians is selected, and the mean is found to be 3400mg. At 5% significance level,

a) Is there evidence to support the claim the mean daily sodium intake exceeds 3300mg?
b) Test for conformance to the health requirement using p-value.
c) Construct 95% CI on the mean daily sodium intake.

Solution part (a) :

1. Parameter of Interest

   Two sided test about the mean, variance known.

2. Hypothesis

   Null hypothesis H_0: $\mu = 3300$

   Alternative hypothesis H_1: $\mu > 3300$

3. Test statistics:

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

4. Critical value:

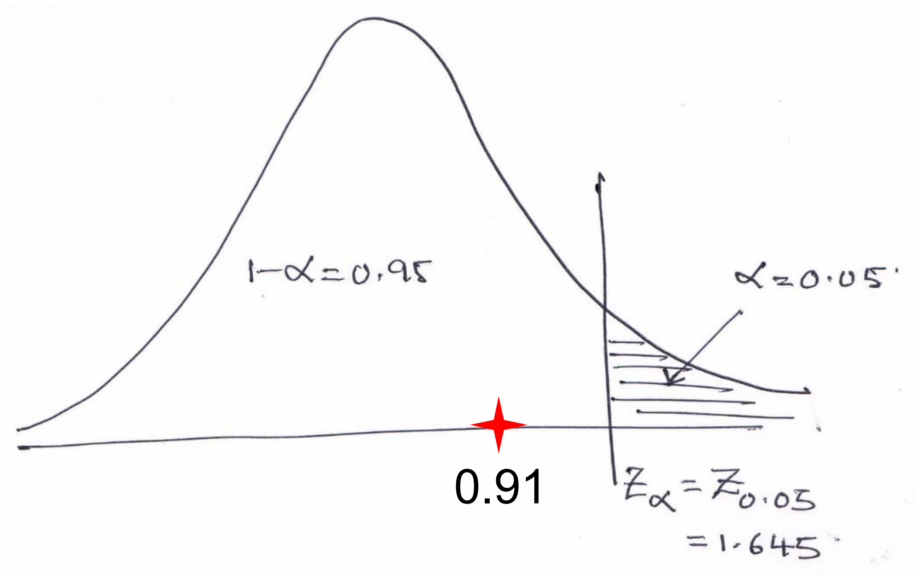
$$\alpha = 0.05$$

$z_\alpha = z_{0.05} = 1.645$ (from the normal tables)

Solution part (a) Cont….

5. Reject region:

$$Z > z_{\alpha} = 1.645$$



6. Compute test statistic :

$$\bar{x} = 3400 \text{ and } \sigma = 1100,$$

$$z = \frac{3400 - 3300}{1100/\sqrt{100}} = 0.91$$

7. Conclusion:

Since z < 1.645, we fail to reject reject H_0: μ = 3300mg

There is not enough evidence that the average daily sodium Intake exceeds 3300mg.

Solution part (b)

Test for conformance to the health requirement using p-value.

	Normal Distribution
Two-tailed test	P-value $= 2[1-\Phi(\lvert z \rvert)]$
Upper-tailed test	P-value $= 1- \Phi(z)$
Lower-tailed test	P-value $= \Phi(z)$

$$p\text{-}value = \mathrm{P}(\ z > 0.91) = 1 - 0.8186 = 0.1814$$

Since p –value (0.1814) > α (=0.05), we fail to reject H_0

Based on sample of 100 measurements, the average daily sodium intake does conforms with health requirement.

Solution part (c) Construct 95% CI on the mean daily sodium intake (One-sided CI)

$$\bar{x} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}} \le \mu \le \bar{x} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}$$

$$-\infty \le \mu \le \bar{x} + \frac{(1.96)1100}{\sqrt{100}}$$

$$-\infty \le \mu \le 3400 + 215.6$$

$$-\infty \le \mu \le 3615.6$$

The 95% CI: (-∞, 3615.6]

Situation or Question for you to think!!!!!!!

In conducting experiment to evaluate a new but very costly process for producing synthetic diamonds, you are able to study **only six** diamonds generated by the process.

How can you **use these 6 measurements** to make inferences about the average weight μ of diamonds from this process?

We use Normal Distribution….

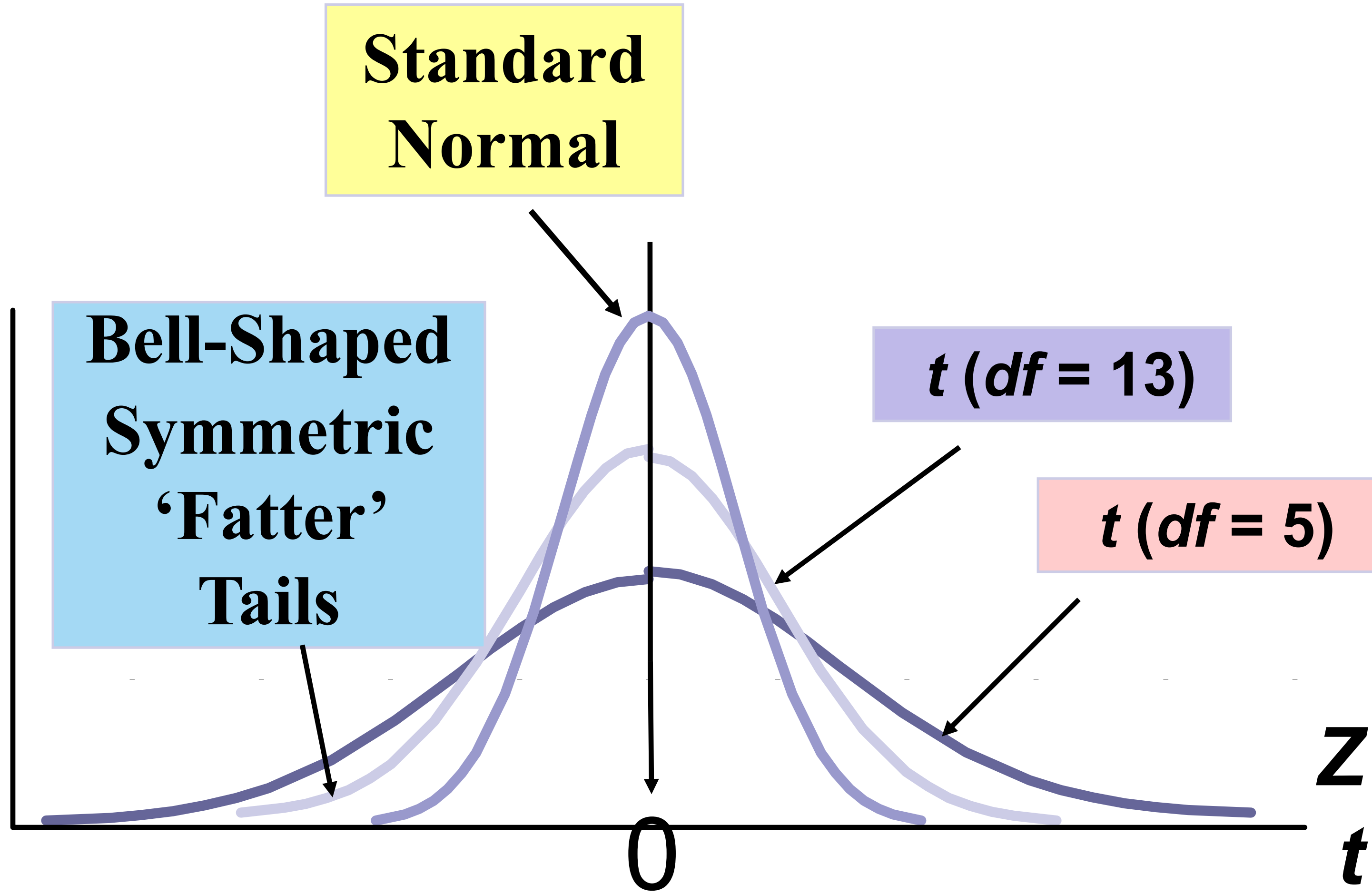
i. If original sampled population is normal, then $\bar{X}$ have a normal population for any sample size.
ii. If original sampled population is not normal, then $\bar{X}$ have a normal population if $n \geq 30$ (CLT)

We use the **student t-distribution ….**

i. When sample size n is small, $\bar{X}$ does not have a normal Distribution
ii. When sample size n is small (n<30).

Student's *t* Distribution

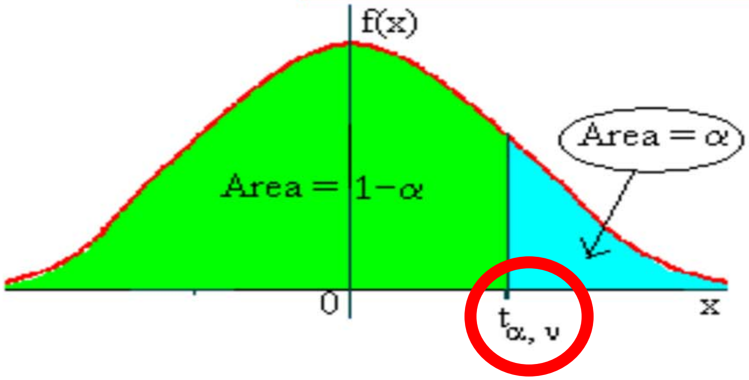
$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+x^2\right)^{-\frac{\nu+1}{2}}$$

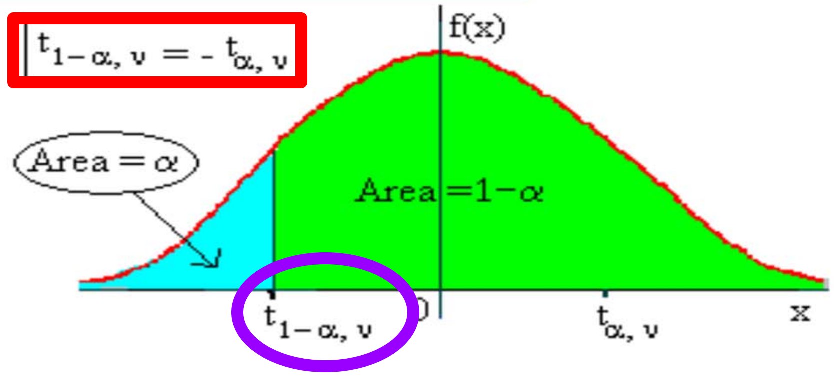


Critical Values of t-DISTRIBUTION

$$\boxed{T \sim t(\nu) \Rightarrow P\left(T \geq t_{\alpha,\nu}\right) = \alpha}$$

$$\boxed{t_{1-\alpha,\nu} = -t_{\alpha,\nu}}$$

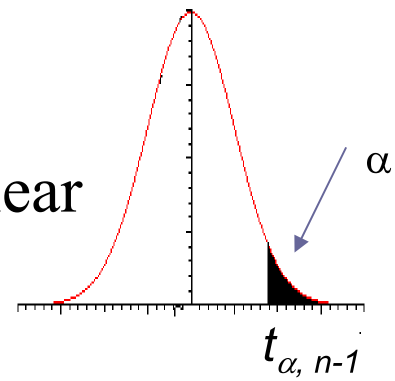




ν \ α	0.975	0.95	0.50	0.05	0.025
5	-2.571	-2.015	0	2.015	2.571
120	-1.9809	-1.658	0	1.658	1.9809
∞	-1.960	-1.645	0	1.645	1.960

Some time we write

$t_{\alpha,n-1}$ as t_{α} when the degrees of freedom is clear



- Values of $t_{\alpha,\ n\text{-}1}$ *can be obtained from the tables book*

Example Determine the following

$$t_{0.025,14}, \quad t_{0.05.15}, \quad t_{0975.15}, \quad t_{0.9,14}$$

- 1. $t_{0.025,14}$ =? **2.145** 2. $t_{0.05,15}$ =? **1.753**
- 3. $t_{0.975,15}$ =? **- 2.131** 4. $t_{0.9,14}$ =? **- 1.345**

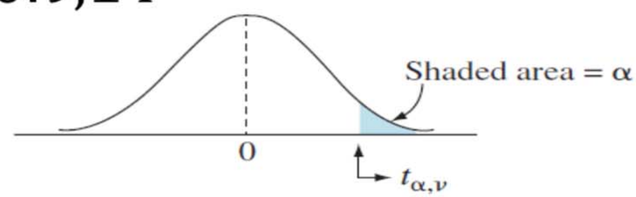


TABLE 2

Percentage points of Student's t distribution

df/α =	.40	.25	.10	.05	.025	.01	.005	.001	.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.646	3.965

2.Hypothesis test claim about a normal mean (μ) when variance (σ^2) is **unknown (t-test)**

- Three Test Problems about Normal mean μ:
  1. **H_0: $\mu = \mu_0$ (or $\mu \le \mu_0$) against H_1: $\mu > \mu_0$**
  2. **H_0: $\mu = \mu_0$ (or $\mu \ge \mu_0$) against H_1: $\mu < \mu_0$**
  3. **H_0: $\mu = \mu_0$ against H_1: $\mu \ne \mu_0$**
- **The Key: Test Statistic**

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

- $T \sim t(n-1)$ **if** $\mu = \mu_0$

- **<u>Critical Region</u>: (Respective regions: Reject H_0)**
  1. $T > t_{\alpha, n-1}$
  2. $T < -t_{\alpha, n-1}$
  3. $|T| > t_{\alpha/2, n-1}$: $T < -t_{\alpha/2, n-1}$ or $T > t_{\alpha/2, n-1}$

Steps in Hypothesis Testing – By Traditional Method

1. Identify the *parameter of interest*.
2. State the null H_0 and alternative H_1 hypothesis.
3. Determine the appropriate *sampling distribution* of the *sample test statistic* (z or t-distribution).
4. Find the critical value.
5. Define the *rejection area (critical region)* based on *level of significance (α)*.
6. Compute the sample test statistic.
7. Draw conclusion based on the decision rule (rejection area).

P- values in Hypothesis Testing

Calculate P-value :

	Normal Distribution	t- Distribution
Upper-tailed test	$P\text{-value} = 1 - \Phi(z)$	$P\text{-value} = P(T_{n-1} > t)$
Lower-tailed test	$P\text{-value} = \Phi(z)$	$P\text{-value} = P(T_{n-1} < t)$
Two-tailed test	$P\text{-value} = 2[1-\Phi(\lvert z \rvert)]$	$P\text{-value} = 2P(T_{n-1} > \lvert t \rvert)$

When $P < \alpha$, reject H_0

When $P \geq \alpha$, fail to reject H_0

Steps in Hypothesis Testing - By p-Value Method

1. Identify the *parameter of interest*.
2. State the null H_0, and alternative H_1 hypothesis.
3. Determine the appropriate *sampling distribution* of the *sample test statistic*. (z or t-distribution)
4. Compute the sample test statistic.
5. Compute the p-values
6. Draw conclusion based on the p-values.

Chapter 6 Hypothesis Testing – Single Population

L4 - Confidence Interval for Mean,
Variance Unknown

L5 - Hypothesis Test about a Proportion
and Hypothesis Test for variance

- **Learning Outcomes**:

At the end of the lesson student should be able to

- construct CI for mean when variance unknown and relate the CI with the hypothesis test;
- Perform hypothesis test about a proportion
- Perform hypothesis test for variance

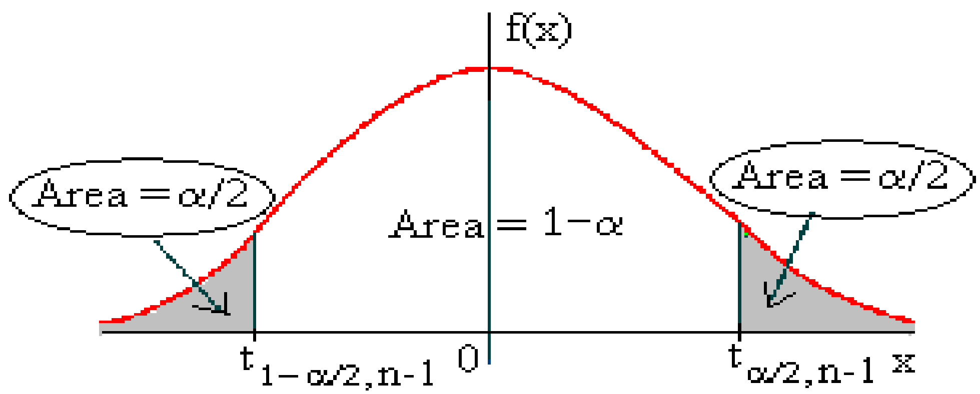
WHEN VARIANCE IS UNKNOWN

A 100 (1- α)% confidence level for the mean μ of a normal Population when the value of σ is not known is given by

$$[\bar{x} - (t_{\alpha/2, n-1})(\frac{s}{\sqrt{n}}), \bar{x} + (t_{\alpha/2,n-1})(\frac{s}{\sqrt{n}})]$$

Or equivalently

$$\bar{x} - (t_{\alpha/2,n-1})(\frac{s}{\sqrt{n}}) \le \mu \le \bar{x} + (t_{\alpha/2,n-1})(\frac{s}{\sqrt{n}})$$



$$\boxed{t_{1-\alpha/2,n-1} = -t_{\alpha/2,n-1}}$$

Example 1

The flow discharge of Perak River (measured in m^3/s) was obtained at random. 20 readings were collected and the mean flow discharge was found to be $3.85 m^3/s$ with a standard deviation of $0.5 m^3/s$.

(a) Test the hypothesis that mean flow discharge at Perak River is not equal to $4 m^3/s$. Use $\alpha=0.05$;

(b) Use the *P-value* approach to test the hypothesis null.

(c) Construct a 95% two-sided CI on mean flow discharge. What is conclusion?

Solution: (a)

1. Problem: To test about the mean, variance unknown.
2. Hypothesis : $H_0 : \mu = 4 \quad \text{vs} \quad H_1 : \mu \neq 4$
3. Test statistics: $$\boxed{T = \frac{\overline{X} - \mu_0}{s / \sqrt{n}}}$$
4. Critical value: Critical value: $\alpha = 0.05$

   $t_{?/2,\, n-1} = t_{0.025,\, 19} = 2.093$
5. Rejection region: Reject Ho IF

   $$T > 2.093 \quad \text{or} \quad T < -2.093$$
6. Calculation: $$T = \frac{3.85 - 4}{0.5 / \sqrt{20}} = -1.34$$
7. Conclusion:

   Since -1.34 > -2.093, so we fail to reject the null hypothesis and conclude the true mean flow discharge is not significantly different from 4m^3/s at α = 0.05.

Solution: (b)

From a *t*-distribution table, for a *t* – distribution with 19 degree of freedom, that $T=1.34$ *is* falls between two values: 1.328 for which α=0.1 and 1.729 for which α=0.05. So the *P*-value is :

$$2(0.05 < P < 0.1) = 0.1 < P < 0.2$$

Since $P > 0.05$, thus we fail to reject H_0 and conclude that the mean flow discharge is not significantly different from 4m^3/s. Same result as in (a).

Solution: (c)

A 95% two-sided CI flow discharge is

$$\bar{x} = 3.85,\ s = 0.5,\ n = 20,\ \ t_{\alpha/2,n-1} = t_{0.025,19} = 2.093$$

$$\bar{x} - t_{\alpha/2,n-1}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{\alpha/2,n-1}\left(\frac{s}{\sqrt{n}}\right)$$

$$3.85 - (2.093)\left(\frac{0.5}{\sqrt{20}}\right) \le \mu \le 3.85 + (2.093)\left(\frac{0.5}{\sqrt{20}}\right)$$

$$\text{so the 95\% two sided CI is } 3.616 \le \mu \le 4.084$$

Since $\mu = 4$ is falls inside of the CI, so we fail to reject the null hypothesis and conclude the true mean flow discharge is not significantly different from $4 m^3/s$ at $\alpha = 0.05$.

Same results as in (a) and (b).

Example 2:

A practical brand of diet margarine was analyzed to determine the level of polyunsaturated fatty acid (in percent). A sample of six packages resulted in the following data: 16.8, 17.2, 17.4, 16.9, 16.5 and 17.1.

i. Test the hypothesis that the mean is less than to 17.0 at $\alpha = 0.01$.

ii. Find the P-value of this test. What is your comment?

iii. Construct 95% two-sided CI on the mean and what is the conclusion?.

Solution: (i)

1. Problem: To test about the mean, variance unknown.
2. Hypothesis : $H_0 : \mu = 17.0 \quad \text{vs} \quad H_1 : \mu < 17.0$
3. Test statistics: $\boxed{T = \dfrac{\overline{X} - \mu_0}{s / \sqrt{n}}}$
4. Critical value: Critical value: $\alpha = 0.01$

   $t_{?, n-1} = t_{0.01, 5} = 3.365$
5. Reject region: Reject Ho IF $T < -3.365$
6. Calculation: $\overline{x} = 16.98, \quad s = 0.3188$

   $$\Rightarrow T = \frac{16.98 - 17}{0.3188 / \sqrt{6}} = -0.1537$$
7. Conclusion:

   Since - 0.1537 > - 3.365, so we fail to reject the null hypothesis and conclude the true mean is 17.0 at α = 0.01.

Solution: (ii)

From a *t*-distribution table, for a *t* – distribution with 5 degree of freedom, that $T = 0.1537$ is falls at less then 0.267 for which α= 0.4, so the *P*-value > 0.4

Since $P > 0.01$, thus we fail reject H_0 and conclude that the mean is 17. Same result as in (a).

Solution: (c)

A 95% two-sided CI flow discharge is

$$\bar{x} - t_{0.025,5}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,5}\left(\frac{s}{\sqrt{n}}\right)$$

$$16.98 - 2.571\left(\frac{0.3188}{\sqrt{6}}\right) \le \mu \le 16.98 + 2.571\left(\frac{0.3188}{\sqrt{6}}\right)$$

Thus the 95% CI is $16.645 \le \mu \le 17.3146$

Since $\mu = 17$ is falls inside of the CI, so we fail to reject the null hypothesis and conclude the true mean is 17.
Same results as in (a) and (b).

Example 3(Large Sample Size)

The flow discharge of Perak River (measured in m^3/s) was obtained at random. 100 readings were collected and the mean flow discharge was found to be $3.85 m^3/s$ with a standard deviation of $0.5 m^3/s$.

(a) Test the hypothesis that mean flow discharge at Perak River is not equal to $4 m^3/s$. Use $\alpha=0.05$;

(b) Use the *P-value* approach to test the hypothesis null.

(c) Construct a 95% two-sided CI on mean flow discharge. What is conclusion?

Solution: (a)

1. Problem: To test about the mean, variance unknown (Large Sample).
2. Hypothesis : $H_0 : \mu = 4 \quad \text{vs} \quad H_1 : \mu \neq 4$
3. Test statistics: $\boxed{Z = \dfrac{\overline{X} - \mu_0}{s / \sqrt{n}}}$
4. Critical value: Critical value: $\alpha = 0.05$

   $z_{?/2} = Z_{0.025} = 1.96$
5. Rejection region: Reject Ho IF

   $$Z > 1.96 \quad \text{or} \quad Z < -1.96$$
6. Calculation: $Z = \dfrac{3.85 - 4}{0.5 / \sqrt{100}} = -3.0$
7. Conclusion:

   Since -3.0 < -1.96, so we reject the null hypothesis and conclude the true mean flow discharge is significantly different from $4m^3/s$ at $\alpha = 0.05$.

Solution: (b)

P-value is :

$$2(1-\Phi(3)) = 2(1-0.998) = 0.004$$

Since $P < 0.05$, thus we to reject H_0 at $\alpha = 0.05$ and conclude that the mean flow discharge is significantly different from $4 m^3/s$. Same result as in (a).

Solution: (c)

A 95% two-sided CI flow discharge is

$$\bar{x} = 3.85,\ s = 0.5,\ n = 100,\ Z_{\alpha/2} = 1.96$$

$$\bar{x} - Z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + Z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

$$3.85 - (1.96)\left(\frac{0.5}{\sqrt{100}}\right) \le \mu \le 3.85 + (1.96)\left(\frac{0.5}{\sqrt{100}}\right)$$

so the 95% two sided CI is $3.752 \le \mu \le 3.948$

Since $\mu = 4$ is falls outside of the CI, so we reject the null hypothesis and conclude the true mean flow discharge is significantly different from $4 m^3/s$ at $\alpha = 0.05$.

Same results as in (a) and (b).

TEST ABOUT A PROPORTION p

- **Three Test Problems about a proportion p:**
  1. H_0: $p = p_0$ (or $p \le p_0$) against H_1: $p > p_0$
  2. H_0: $p = p_0$ (or $p \ge p_0$) against H_1: $p < p_0$
  3. H_0: $p = p_0$ against H_1: $p \ne p_0$

- **Key: Test Statistic:**

$$\boxed{Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}, \ \hat{p} = \frac{X}{n}}$$

  - $Z \approx N(0,1)$ when $p = p_0$

- **Critical Region:** (The region for Rejecting H_0)
  1. $Z > z_\alpha$
  2. $Z < -z_\alpha$
  3. $Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$, $|Z| > z_{\alpha/2}$

Example 4:

Regardless of age, about 20% of Malaysian adults participate in fitness activities at least twice a week. In a local survey of 100 adults over 40 years old, a total of 30 people indicated that they participated in a fitness activity at least twice a week. Do these data indicate that the participation rate for adults over 40 years of age is larger than 20%? Carry out a test at 10% significance level and draw appropriate conclusion.

Solution: Problem: Test about a proportion p

$$H_0 : P = 0.2 \quad \text{vs} \quad H_1 : P > 0.2$$

Test statistic used:

$$\boxed{\mathbf{Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \approx N(0,1), \text{ if } p = p_0, \text{ where } \hat{p} = \frac{X}{n}}}$$

Critical region: **Reject Ho IF $Z > z_{\alpha}$**

Compute the value of the test statistic:

X =30 , n = 100, $\alpha = 0.1$, $z_{\alpha} = z_{0.1} = 1.28$ $\hat{p} = 30/100 = 0.3$

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \approx N(0,1),\ if\ \ p = p_0,\ where\ \ \hat{p} = \frac{X}{n}$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.30 - 0.2}{\sqrt{(0.2)(0.8)/100}} = 2.5$$

Decision: since Z = 2.5 > 1.28 , then we reject H_0 , a strong evidence say that the participation rate for adults over 40 years of age is significantly larger than 20%.

CI FOR PROPORTION p

- Key: X: number of Success-items in an n- sample

$$\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}} \approx N(0,1)\ , \quad \text{where} \quad \hat{p}=\frac{X}{n}$$

- **Condition:** $n\hat{p} \geq 5,\ \ n(1-\hat{p}) \geq 5$

- Therefore the Approximate (1 - α) CI for p is:

$$\hat{p}-(z_{\alpha/2})(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}) \leq p \leq \hat{p}+(z_{\alpha/2})(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}})$$

UNIVERSITI TEKNOLOGI PETRONAS

Example 5:

Find the 90% two-sided CI for the proportion of age, about 20% of Malaysian adults participate in fitness activities at least twice a week if a total of 30 out of 100 people indicated that they participated in a fitness activity.

Solution:

Problem: Two-side CI for the proportion of age, about 20% of Malaysian adults participate in fitness activities at least twice a week.

$$\hat{p} - (z_{\alpha/2})\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \le p \le \hat{p} + (z_{\alpha/2})\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

Given Sample Data: $n = 100$, $X = 30$, $\alpha / 2 = 0.05$, $z_{0.05} = 1.65$

Calculations:

$$\hat{p} = X / n = 30/100 = 0.3$$

$$z_{0.05}\sqrt{\hat{p}(1-\hat{p})/n} = (1.65)(\sqrt{(0.3)(0.7)/100} = 0.0756$$

$$0.3 - 0.0756 \le p \le 0.3 + 0.0756$$

$$0.2244 \le p \le 0.3756$$

Since $p = 0.2$ is not in the interval, then we reject H_0 ,at $\alpha = 0.1$. A strong evidence say that the participation rate for adults over 40 years of age is significantly greater than 20%.