Page No.
Date :

ch-5

• Hypothesis Testing :

~~Types :~~

→ A hypothesis about the value of population parameter is an assertion (confident & forceful statement of fact) about its value.

→ Types of hypothesis

- null hypothesis (H_0)
- alternative hypothesis (H_a)

→ Statement about population parameter that is assumed to be true unless there is convincing evidence to the contrary. – Null hypothesis (H_0)

→ Statement about the population parameter that is contradictory to null hypothesis, and is accepted as true only if there is convincing evidence in favor of it. – Alternative hypothesis (H_a)

→ Hypothesis testing is a statistical procedure in which a choice is made between null & alternative hypothesis.

→ End result is choice of.
(1) Reject H_0 (and ∴ accept H_a)
(2) fail to reject (i.e.∴ fail to accept H_a).

→ H_0 contains '$=$'
H_a contains '$>$'/'$<$'/'$\neq$'

eg – A publisher of college textbooks claims that the average price is \$127.50. A student group believes that actual mean is higher & wishes to test their belief. State the relevant null & alternative hypothesis.

Null hypothesis : $H_0 = \mu = 127.50$
alternative ″ ″ $H_a = \mu > 127.50$

→ The null hypothesis always has the form $\mu = \mu_0$

→ In null hypothesis, initially assume H_0 is true.

→ The criterion of judging betⁿ H_0 & H_a based on the sample data is:

If value of $\bar{x}$ would be highly unlikely to occur if H_0 were true, but favors the truth of H_a, then we reject H_0 in favor of H_a. Otherwise we do reject H_0.

→ if $H_a: \mu < \mu_0$ reject H_0 if $\bar{x}$ is far to left of μ_0.

$H_a: \mu > \mu_0$ reject H_0. if $\bar{x}$ is far to right of μ_0.

$H_a: \mu \neq \mu_0$ reject H_0 if $\bar{x}$ is far away from μ_0 in either direction.

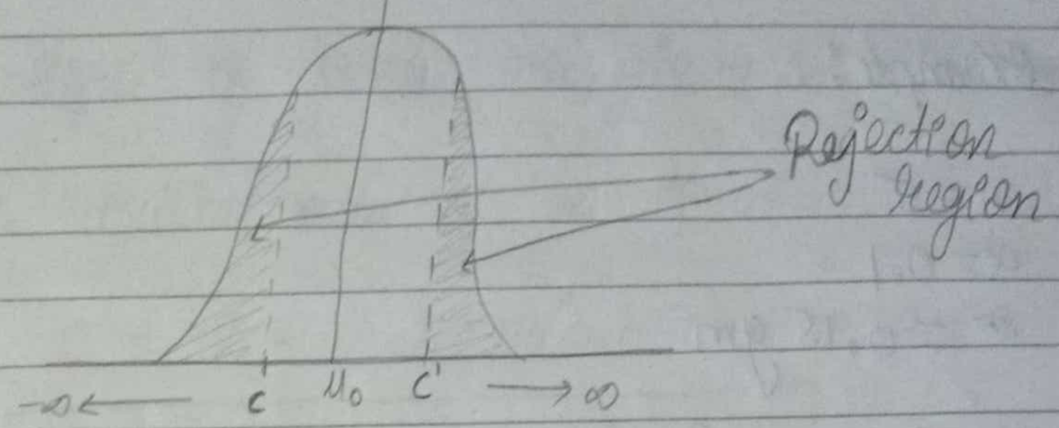
$H_0: \mu_0 = \mu$

→ <u>Rejection</u> <u>area / Region</u>:-

→ $(-\infty, c]$ $\mu < \mu_0$

$[c, \infty)$ $\mu > \mu_0$

c → some number.

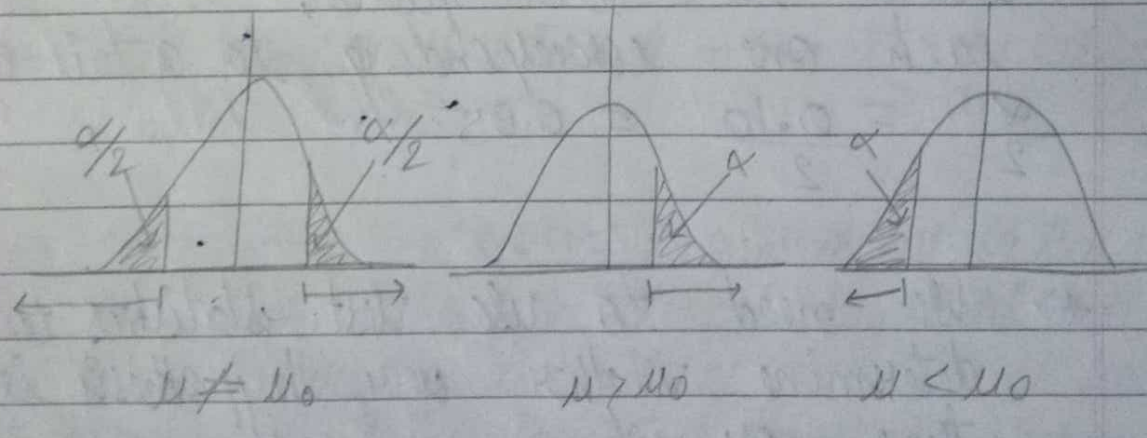


$\mu \neq \mu_0$ $\qquad$ $(-\infty, c] \cup [c', \infty$

$c, c' \rightarrow$ critical values of statistic.

→ if rejection region is single interval, select single number c.

we select small probability, denoted by α, say 1% , which we take a rare event



Shaded area → α → rejection area

Example:

$n = 5$
$\alpha = 0.1$
$\sigma = 0.15$ gm

$H_0 : \mu = 8.0$
$H_a : \mu \neq 8.0$

[Sampling distribution - $\bar{x}$]

$\mu_{\bar{x}} = \mu = 8$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.15}{\sqrt{5}} = 0.067$$

Ha contain '$\neq$' sign, for rejection region will be in two pieces, each one corresponding to a tail of area $\frac{\alpha}{2} = \frac{0.10}{2} = 0.05$.

→ We need to use test statistics to determine whether our hypothesis is true or not.

→ from table 12.3,

degree of freedom not given ∴ assume ∞.

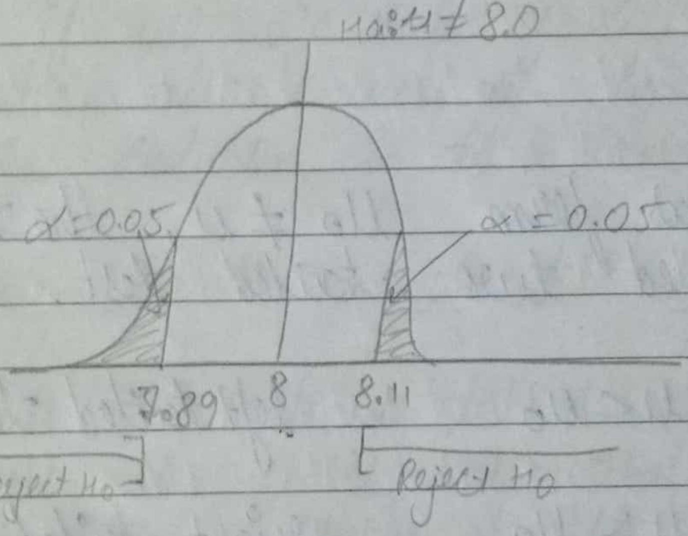
∴ critical value of '∞' df at $\alpha = 0.05$ is 1.645

$$c = 8 - (1.645)(0.067) = 7.89$$

$$\boxed{C = \mu_{\bar{x}} \pm (\text{critical value})(\sigma_{\bar{x}})}$$

± ↓
− ve for left side
+ for right side

$$c = 8 + (1.645)(0.067) = 8.11$$



population standard deviation σ

→ If compute the sample mean $\bar{x}$ by taking sample size of 5.

→ if $\bar{x}$ is either 7.89 gms or less or 8.11 gms or more then reject the hypothesis that the average amount of fat in all servings of product is 8 gm in favor of alternative that it is different from 8 gms.

Page No.
Date :

→ If Ha has form $\mu_0 \neq \mu$, the test is called two tailed test.

→ Ha : $\mu < \mu_0$ left tailed test

Ha : $\mu > \mu_0$ right tailed test

also called one - tailed test.

→ Two types of error.

		True state of Nature	
		Ho is true	Ho is false
our decision	Donot reject Ho	correct decision	Type II error
	reject Ho	Type I error	correct decision

→ There are four types of outcomes according to hypothesis.

→ to reject H_0 when it is true is more serious error than fail to reject false.

→ α is called level of significance of test.

→ if we dec value of α, chances of type I error dec, also chances of type II error inc.
$\therefore$ solution is to increase sample size.

- <u>Standardizing the test statistics</u>:

→ A standardized test statistic for a hypothesis test is the statistic that is formed by subtracting from the statistic of interest its mean & dividing by its standard deviation.

$$\frac{\overline{X} - 8.0}{0.067}$$

When the test statistic has the standard normal distribution:

Symbol of Ha	Terminology	Rejection region
$<$	Left tailed test	$(-\infty, -Z_\alpha]$
$>$	right tailed test	$[Z_\alpha, \infty)$
$\neq$	two tailed test	$(-\infty, -Z_{\alpha/2}] \cup [Z_{\alpha/2}, \infty)$

When the test statistic has student's t - distribution:

Symbol of Ha	Terminology	Rejection Region
$<$	left tailed test	$(-\infty, -t_\alpha]$
$>$	right tailed test	$[t_\alpha, \infty)$
$\neq$	two tailed test	$(-\infty, -t_{\alpha/2}] \cup [t_{\alpha/2}, \infty)$

→ Systematic hypothesis testing procedure:

★ Critical value approach.

1. Identify null & alternative hypothesis.
2. Identify the relevant test statistic & its distribution.

3. Compute from the data the value of the test statistic.

4. Construct the rejection region.

5. Compare the value computed in step 3. to the rejection region constructed in step 4 and make a decision. formulate the decision in the context of problem, if applicable.

Problems

1. a) The avg july temprature in a region historically has been 74.5 °F. perhaps it is higher now.

$H_0 : \mu_0 = 74.5°F$

$H_a = \mu > \mu_0$

b) The avg wt. of female airline passenger with luggage was 145 pounds ten years ago. The FAA believes it to be higher now.

$H_0 : \mu_0 =$ ~~74.5~~ 145 pounds

$H_a : \mu > \mu_0$

ie $\mu > 145$ pounds

format: — hypothesis $H_{-} : \mu =$ ___

Page No.
Date :

(c) The avg stipend for doctoral students in a particular discipline at a state university is \$14,756. The department chairman believes that the national avg is higher.

$H_0 = \mu_0 = \$14756$
$H_a : \mu_0 > \$14756$

d) The avg room rate in hotels in a certain region is \$82.53. A travel agent believes that the average in particular resort area is diff.

$H_0 : \mu_0 = \$82.53$
$H_a : \mu_0 \neq \$82.53$

e) The avg farm size in predominately rural state was 69.4 acres. The secretary of agriculture of that state asserts that it is less today.

Null hypothesis $H_0 : \mu_0 = 69.4$
Alternative hypothesis $H_a : \mu_a < \mu_0$

2. a) Null hypothesis $H_0 : \mu = 38.2$ mins
Alternative hypothesis $H_a : \mu_a < 38.2$

b) Null hypothesis $H_0 : \mu = \$58,291$
Alternative hypothesis $H_a : \mu_a \neq \$58,291$.

c) Null hypothesis $H_0 : \mu = 133$ mg
Alternative hypothesis $H_a : \mu_a > 133$ mg

d) Null hypothesis $H_0 : \mu = 161.9$
Alternative hypothesis $H_a : \mu_a \neq 161.9$

e) $H_0 : \mu = 42.8$ yrs.
$H_a : \mu_a > 42.8$

Sampling distribution

→ Sample is random, every statistic is a random variable.

→ Probability distribution of a statistic is called its sampling distribution.

population mean μ
Sample mean $\bar{x}$

$\bar{x}$
$\bar{X}$ → Sample mean, random variable.
Standard deviation $\mu_{\bar{x}}$
→ mean of sample mean.

Que 152, 156, 160, 164.

random Sample of size 2.

152, 156 ⇒ $\frac{152+156}{2} = 154$

152, 160 $\frac{152, 160}{2} = 156$

152, 164 $\frac{152+164}{2} = 158$

Samples	$\overline{x}$
152, 156	154
152, 160	156
152, 164	158
156, 160	158
156, 164	160
160, 164	162
152, 152	152
156, 156	156
156, 152	154
160, 160	160
160, 152	156
160, 156	158
~~160,~~	
164, 164	164
164, 152	158
164, 156	160
164, 160	162

$\cancel{\overline{x} = 154+156+158+158+160+162+152+15}$

Sampling distribution.

$\overline{x}$	152	154	156	158	160	162	164
$P(\overline{x})$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

★ $\boxed{\text{Sample mean}: \mu_{\bar{x}} : \Sigma \bar{x}\, P(\bar{x})}$

$$= \frac{152 + 308 + 468 + 632 + 480 + 324 + 164}{16}$$

$\boxed{\mu_{\bar{x}} = 158}$

$$\sigma_{\bar{x}} = \Sigma \bar{x}^2\, P(\bar{x})$$

$$= \frac{152 + 616 + 1404 + 2528 + 1440 + 648 + 164}{(16)^2}$$

$\boxed{\sigma_{\bar{x}} = 27.156}$

$$\sigma_{\bar{x}} = \frac{23104 + 23716\times2 + 24336\times3 + 24964\times4 + 25600\times3 + 26244\times2 + 26896}{16}$$

$$= 24974$$

★ $\boxed{\sigma_{\bar{x}} = \sqrt{\Sigma \bar{x}^2\, P(\bar{x}) - \mu_{\bar{x}}^2}}$

$$= \sqrt{24974 - (158)^2}$$

$$= \sqrt{10}$$

$$\sigma = \sqrt{\frac{\Sigma (x-\mu)^2}{N}} \qquad \mu = 158$$

$$= \sqrt{\frac{36+4+4+36}{4}} = \sqrt{\frac{80}{4}} = \boxed{\sqrt{20}}$$

★ $$\boxed{\mu_{\bar{x}} = \mu \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

population standard deviation σ
sample " " $\sigma_{\bar{x}}$
sample mean $\mu_{\bar{x}}$
population " μ

8/9

Que $\mu = \$13,525 \qquad \sigma = \$4180$
$n = 100$

$\mu_{\bar{x}} = \mu = \$13525$
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\cancel{13525}\ 4180}{10} = \cancel{1352.5}\ 418$

• Standard Normal distribution

→ normally distributed random variable with mean $\mu = 0$ & standard deviation $\sigma = 1$.

→ denoted by 'Z'.

eg- finding probability

a) $P(Z < 1.48)$ b) $(Z > -0.85)$ $(Z > -1.02)$

table 12.2

a) → 0.9306
b) → 0.4013

	0.1	0.2	0.3	...	0.8
1.1					
1.2					
1.3					
1.4					

$P(Z < -1.02) = P(Z \leq -1.02)$
$= 0.1539$
$P(Z > -1.02) = 1 - P(Z \leq 1.02)$
$= 1 - 0.1539$

find probabilities:

1) $P(0.5 < Z < 1.57)$ →
b) $P(-2.55 < Z < 0.09)$ → $P(Z < 0.09) - P(Z < -2.55)$
$= 0.5359 - 0.0054 = 0.5305$

→ Probabilities for standard normal variable is computed by table 12.2.

General

Normal distribution :-

$$P(a < x < b) = P\left(\frac{a-\mu}{\sigma} < z < \frac{b-\mu}{\sigma}\right)$$

range $(-\infty, \infty)$

mean μ, Standard deviation σ

X - general normally distributed random variable

Z - Standard

eg - $\sigma = 2.5$ $\quad \mu = 10$

a) $P(X < 14)$

$$\frac{b-\mu}{\sigma} = \frac{14-10}{2.5} = \frac{4}{2.5} = 1.6$$

$P(z < 1.6)$

$= \underline{0.9452}$

b) $P(8 < X < 14)$

$$\frac{a-\mu}{\sigma} = \frac{8-10}{2.5} = \frac{-2}{2.5} = -0.8.$$

$P(-0.8 < z < 1.6)$

$= P(z < 1.6) - P(z < -0.8)$

$= 0.9452 - 0.2119$

$= \underline{0.7333}$

eg - $\mu = 37500$ $\quad \sigma = 4500$

$P(30,000 < X < 40,000)$

$P \quad \frac{a-\mu}{\sigma} = \frac{30,000 - 37500}{4500} = -1.6$

$\frac{b-\mu}{\sigma} = \frac{40,000 - 37500}{4500} = 0.555$

$P(-1.6 < Z < 0.55)$

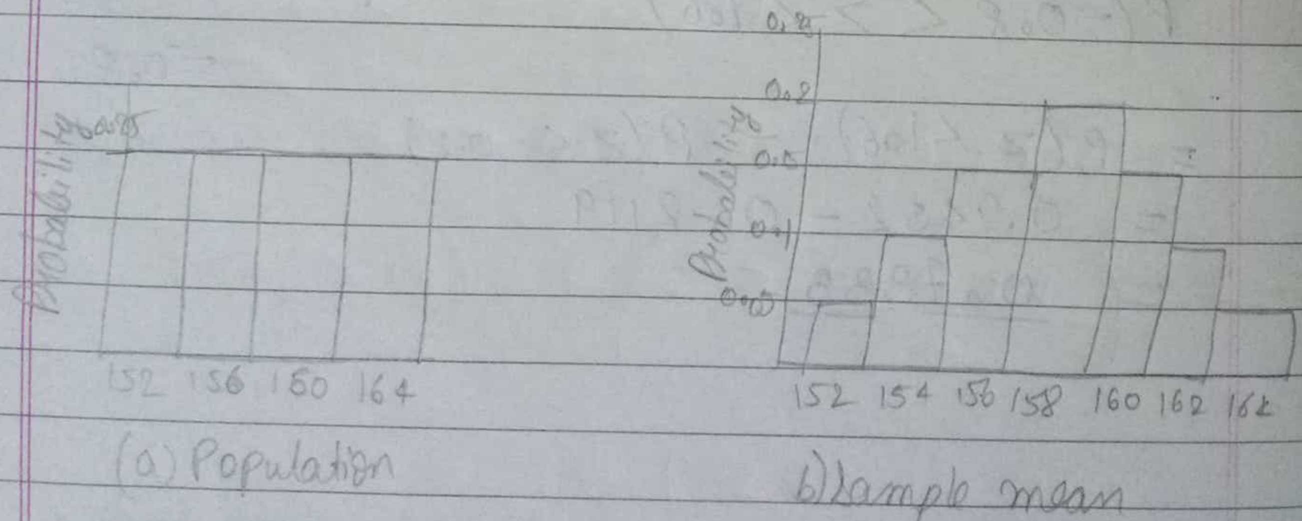
$= -P(-1.67 < Z) + P(Z < 0.55)$

$= 0.7088 - 0.0548 \rightarrow 0.56.$

$= 0.6548$

The sampling distribution of sample mean

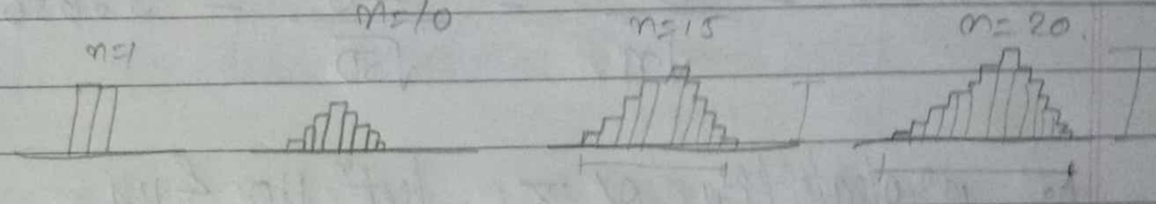
- <u>Central</u> <u>limit</u> <u>theorem</u>:



(a) Population

b) sample mean

→ as n inc., sampling distribution $\bar{x}$ evolves in interesting way: probabilities of lower & upper shrinks at at middle it incs its height.



→ larger the sample size, better is the approximation.

→ ~~Importance of Central limit theorem:~~

→ x, the measurement of single element selected at random from the population, the distribution of x is the distribution of the population, with mean the population mean μ and standard deviation the population standard deviation σ.

→ $\bar{x}$, the mean of measurements in a sample of size n, the distribution of $\bar{x}$ of its sampling distribution, with mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \sigma/\sqrt{n}$

Page No.
Date :

Q) $n = 50$
$\mu = 112 \quad \sigma = 40.$

a. $\mu_{\bar{x}} = \mu = 112$
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{50}} = 5.656$

b. probability of $\bar{x}$ betⁿ 110 & 114.

$P(110 < \bar{x} < 114)$

$P\left(\frac{110 - \mu_{\bar{x}}}{\sigma_{\bar{x}}} < \cancel{Z} < \frac{114 - \mu_{\bar{x}}}{\sigma_{\bar{x}}}\right)$

$P\left(\overset{-0.357}{\cancel{-0.05}} < Z < \overset{0.357}{\cancel{0.05}}\right)$

$P(Z<0.05) - P(Z < -0.05)$

$= \cancel{-0.4801 + 0.5199}$
$= \cancel{0.}$
$= 0.6368 - 0.3632 = 0.2736$

c) $P(\bar{x} > 113)$ $P\left(Z > \frac{113 - \mu_{\bar{x}}}{\sigma_{\bar{x}}}\right) = P(Z > 0.18$

$= 1 - P(Z < 0.18)$
$= 1 - 0.5714 = 0.4286$

Page No.
Date :

- Normally distributed Populations

Sample size ≥ 30 Sample mean normally distributed.

eg. Population μ 38500 $\sigma = 2500$

$n = 5$

Sample mean will be less than 36000.

$$\sigma_{\bar{x}} = \frac{2500}{\sqrt{5}} = 1118.0339$$

$P(\bar{x} < 36000)$

$$P\left(Z < \frac{36000 - 38500}{1118.0339}\right) = P(-2.236) = P(-2.24)$$

$= \cancel{0.0091}$ $\underline{0.0125}$

eg $\mu = 50$ $\sigma = 6$

a) $P(\bar{x} < 48) = P\left(Z < \frac{48-50}{6}\right)$

$= P(Z < -0.333)$

$= \cancel{0.6293}$ $\cancel{0.4880}$ 0.3707

b) $n = 36$ $P(\bar{x} < 48) = P\left(Z < \frac{48-50}{1}\right)$

$= P(Z < -2) = \underline{0.0228}$

Page No.
Date :

Large sample test for population mean

$n \geq 30$

Statistic $\rightarrow \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

is σ is unknown, replace with sample standard deviation s.

Tests concerning a single population mean

σ known $\quad Z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

σ unknown $\quad Z = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

Test statistic has the standard normal distribution.

→ here critical value is Z_α

eg

Eg 4

It is hoped that a newly developed pain reliever will more quickly produce perceptible reduction in pain to patients after minor surgeries than a standard pain reliever.

average = 3.5 min $\sigma = 2.1$ min

sample $\bar{x} = 3.1$ $s = 1.5$

$n = 50$

Solution $H_0 : \mu = \mu_0 = 3.5$

$H_a : \mu < 3.5$ @ $\alpha = 0.05$

Should give relief within 3.5 min.

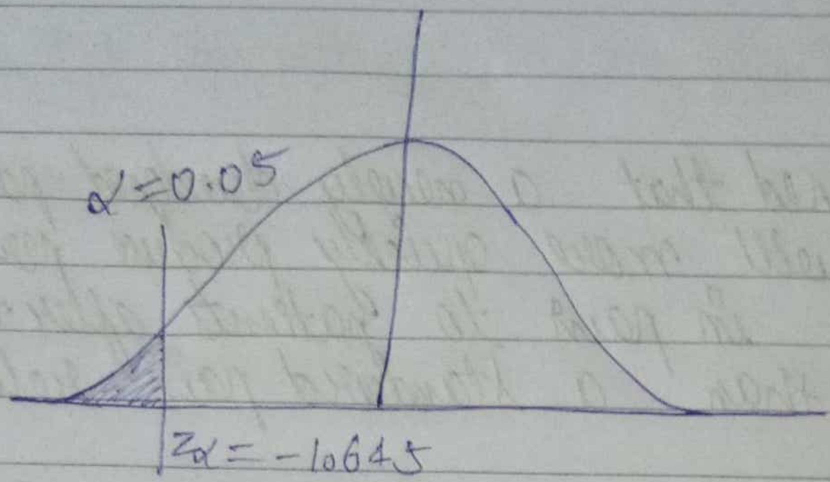
$$Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{3.1 - 3.5}{1.5/\sqrt{50}}$$

$$= -1.886.$$

$$-Z_\alpha = -Z_{0.05} = -1.645$$

H_a has "<" so it is left tailed test.

—



Z	A
-1.64	0.0505
Z_α	0.05
-1.65	0.049

through interpolation,

$$\frac{Z_1 - Z_2}{A_1 - A_2} = \frac{Z_1 - Z_\alpha}{A_1 - A_\alpha}$$

$$\frac{-1.64 + 1.65}{0.0505 - 0.049} = \frac{-1.64 - Z_\alpha}{0.0505 - 0.05}$$

$$Z_\alpha = -1.645$$

Example:

$\mu = 8.1 \qquad n = 30$

$\sigma = 0.22$

$\bar{x} = 8.2 \qquad s = 0.25$

$\alpha = 1\% = 0.01$

Solution:-

$H_0 = \mu = \mu_0 = 8.1$

$H_a = \mu \neq \mu_0 \quad \text{ie} \quad \mu \neq 8.1$

$$Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{8.2 - 8.1}{0.25/\sqrt{30}}$$

$\boxed{Z = 2.19}$

here, In H_a '$\neq$' is used, therefore

$$\frac{\alpha}{2} = \frac{0.01}{2} = 0.005$$

$Z \qquad A$

0.0102