⊛ Sampling Distribution.

A sampling distribution is a distribution of all the possible values of a sample statistics for a given size sample selected from a population.

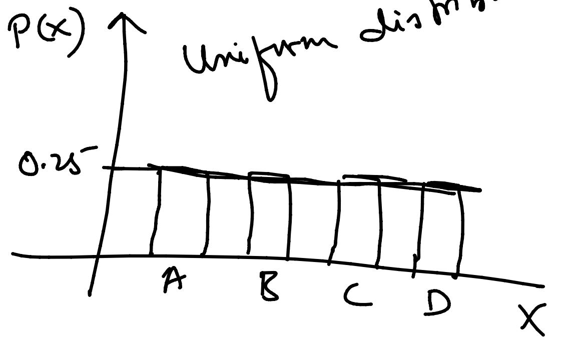
eg:- Assume there is a a population of size N - 4

Random variable X - is a age of individuals

	A	B	C	D	
X -	18	20	22	24	→ (yrs)

$$\mu = \frac{\Sigma X}{N} = 21 \rightarrow \text{Population mean}$$

$$\sigma = \left(\frac{\Sigma (x_i - \mu)^2}{N} \right)^{1/2} = 2.23$$



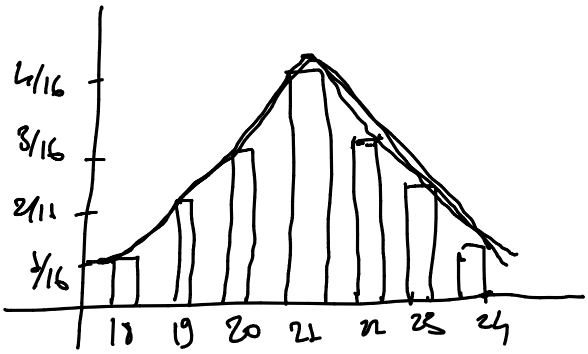
Now consider all possible samples of size n = 2.

1st obs. \ 2nd observation	18	20	22	24
18	(18,18)	(18,20)	(18,22)	(18,24)
20	(20,18)	(20,20)	(20,22)	(20,24)
22	(22,18)	-	- - .	(22,24)
24	(24,18) -	-	- -	(24,24)

16 possible samples with replacement.
(Sampling)

Mean

1st \ 2nd	18	20	22	24
18	18	19	20	21
20	19	20	21	22



18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

1 18 19 20 21 22 23 24

$$\mu_{\bar{x}} = \frac{18 + 2(19) + 3(20) + 4(21) + 3(22) + 2(23) + 24}{16}$$

$$\boxed{\mu_{\bar{x}} = 21}$$

Note:- Mean of population distribution is equal to the mean of sampling distribution.

$$\sigma_{\bar{x}} = \left[\frac{\Sigma(x_i - \mu_{\bar{x}})^2}{16}\right]^{1/2} = \left[\frac{(18-21)^2 + 2(19-21)^2 + 3(20-21)^2 + 4(21-21)^2 + 3(22-21)^2 + 2(23-21)^2 + (24-21)^2}{16}\right]^{1/2}$$

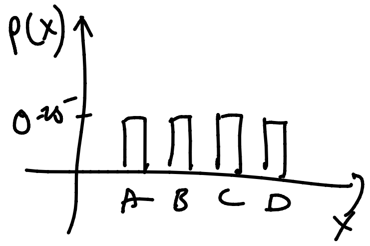
$$\boxed{\sigma = 2.23}$$

$$\sigma_{\bar{x}} = 1.581$$

In case of sampling distribution standard deviation decreases.

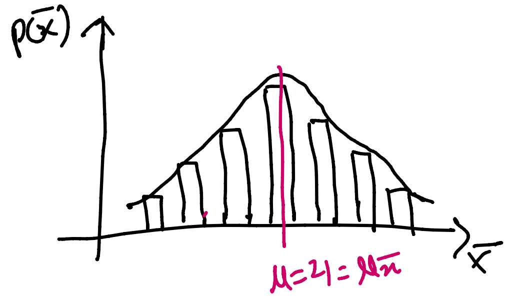
Population $N-4$

$\mu = 21$, $\sigma = 2.23$



Sampling distribution $n=2$

$\mu_{\bar{x}} = 21$, $\sigma_{\bar{x}} = 1.581$



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

... samples of the same size from the same ... means.

* Different samples of the same size from the same population will yield different sample means.

* A measure of the variability in the mean from sample to sample given by the <u>standard error of the mean</u>. ($\sigma_{\bar{x}}$)

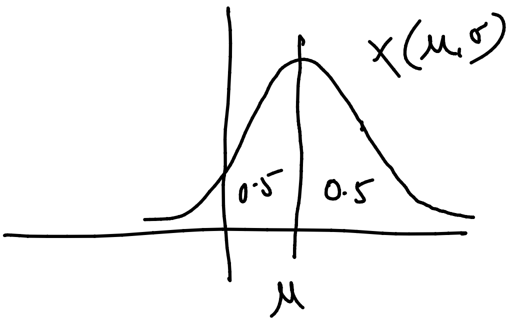
$$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

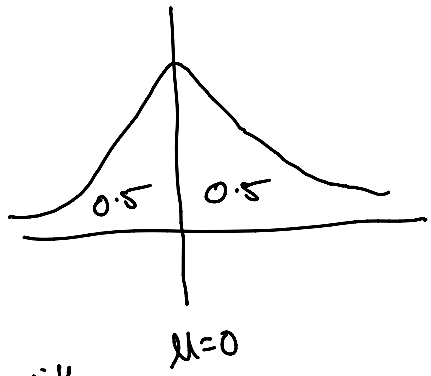
$$\sigma_{\bar{x}} = \frac{2.23}{\sqrt{2}} = 1.581$$

* Note that the standard devian of means (standard error of the means) decreases as the sample size increases.

* If the population is normal. (X, μ, σ)

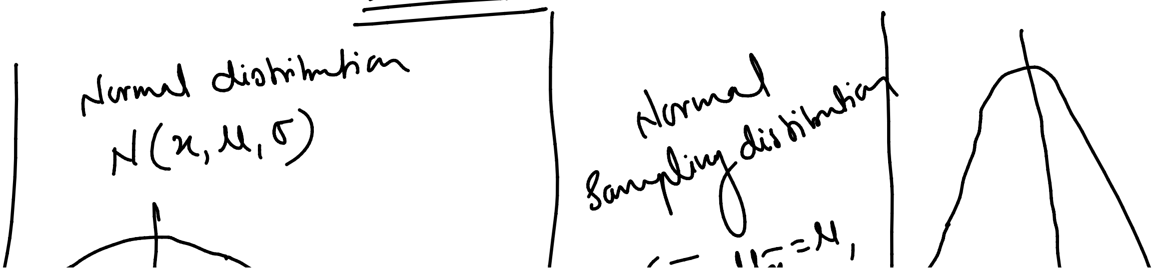
$$\boxed{Z(0,\sigma)} \qquad \boxed{Z = \frac{x-\mu}{\sigma}}$$

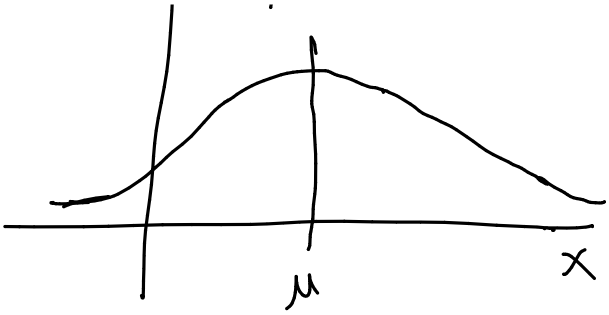


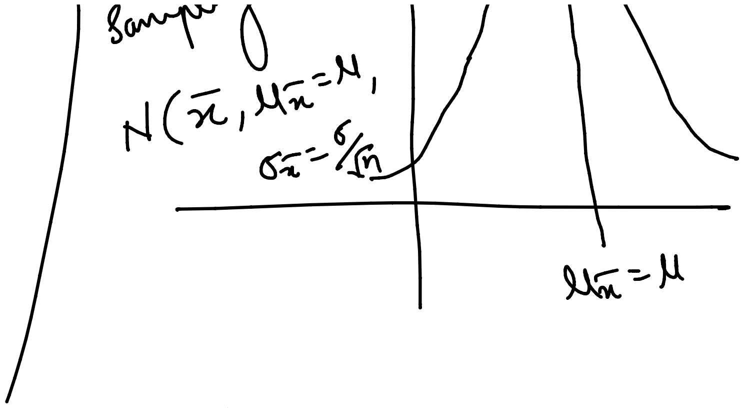


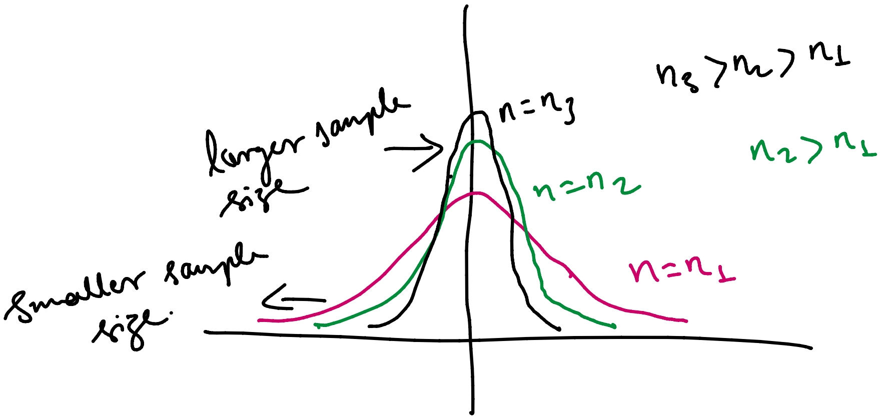
If the population is normal with mean μ and std. deviation σ. The sampling distribution of $\bar{x}$ is also normally distributed.

$$\mu_{\bar{x}} = \mu \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$





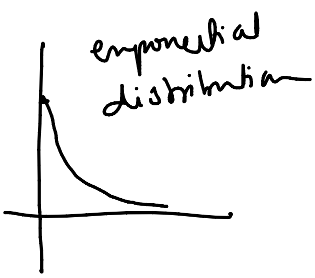


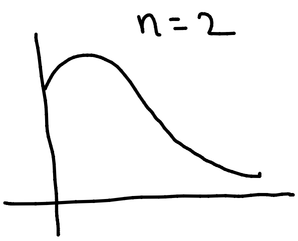


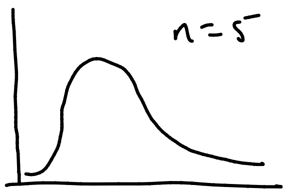
⊛ **If the Population is not Normal** then also the sampling distribution is normal provided n (sample size) is sufficiently large.

⊛ **Central limit theorem:-** As the sample size gets large enough the sampling distribution becomes almost normal regardless of shape of the proportion.

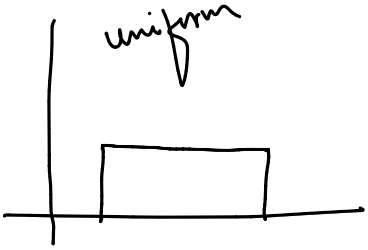
(1)

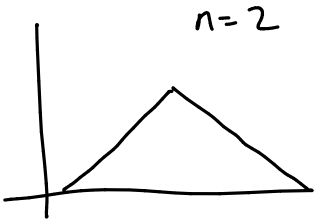


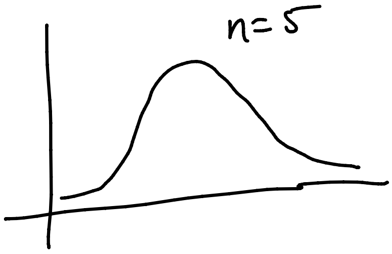


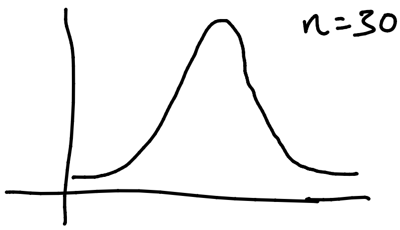


(2)



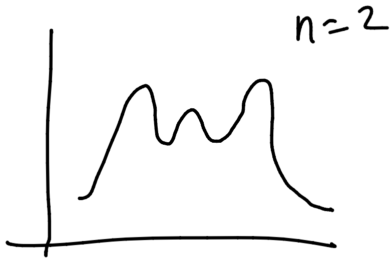


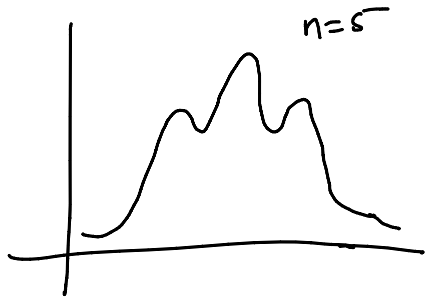


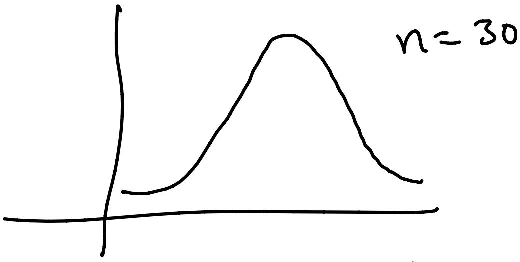


$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

(3)







(*) How large is large enough.

→ For most of the distributions, $n > 30$ will give a sampling distribution that is nearly normal.

eg:- Suppose, for example that the mean expenditure per customer at a tire store is \$85.00 with std. deviation \$9.00. If random sample of 40 customers is taken, what is the probability that the sample average expenditure per customer for this sample will be \$87.00 or more?

$$\mu = 85, \quad \sigma = 9, \quad n = 40$$

⇒ $P(\bar{x} > 87) = ?$ $\quad \mu_{\bar{x}} = \mu = 85$

$9 \quad = 1.4230$

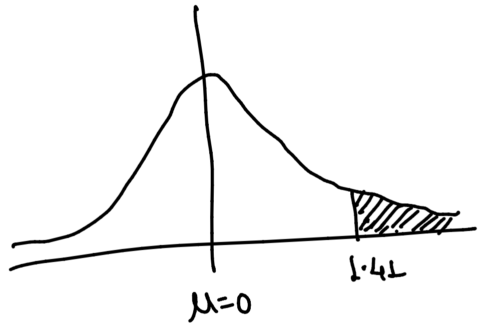
$\Rightarrow$ $\mu = 85$,

$P(x \geq 87) = ?$, $\mu_{\bar{x}} = \mu = 85$

$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{9}{\sqrt{40}} = 1.4230$

$$z = \frac{x - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$z = \frac{87-85}{9/\sqrt{40}} = \frac{2\sqrt{40}}{9} = 1.41$$



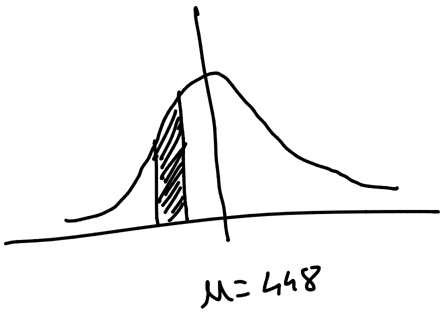
$$P(z \geq 1.41) = 0.5 - P(z = 1.41)$$
$$= 0.5 - 0.4207$$
$$= \underline{\underline{0.0793}}$$

eg:- Suppose that during any hour in a large department store, the average number of shoppers is 448, with a std. deviation of 21 shoppers. What is the probability that a random sample of 49 different shopping hours will yield a sample mean betn 441 and 446 shopers.

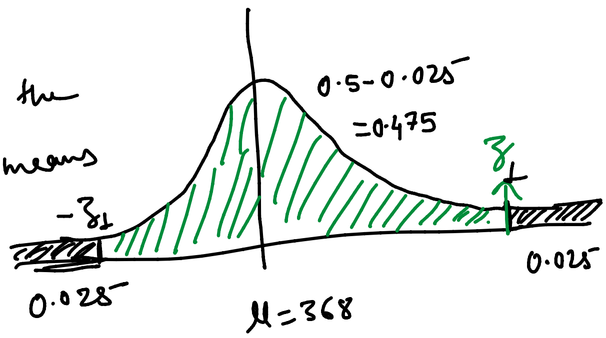
H.W.

$\Rightarrow$ $\mu = 448$, $\sigma = 21$, $n = 49$

$\mu_{\bar{x}} = 448$, $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = 3$



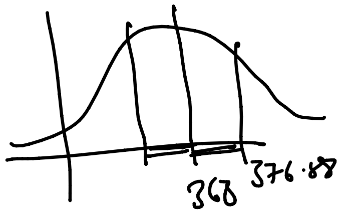
eg:- Find a symmetically distributed interval around μ that will include 95% of the sample means when μ = 368, σ = 15 and n = 25.



$\Rightarrow$ 1) Since the interval contains 95% of the sample means 5% of the sample means will be outside of the interval.

2) Since the interval is symmetric ... will be above the upper limit and

2) Since the interval ...
2.5 % will be above the upper limit and
2.5 % will be below the lower limit.



$\therefore \; z_1 = 1.96 \quad \Rightarrow \quad -z_1 = -1.96$

$$z = \frac{x-\mu}{\sigma_{\bar{x}}} \quad \Rightarrow \quad x = z\,\sigma_{\bar{x}} + \mu$$

$\therefore \quad x_1 = (1.96)(3) + 368 \;=\; 373.88$

lower limit is, $x_2 = 368 - (1.96)(3)$

$$\boxed{x_2 = 362.12}$$

⊛ Sampling distribution of Proportions :-

Population Proportions :-

(symbol)
π = the proportion of the population having some characteristics.

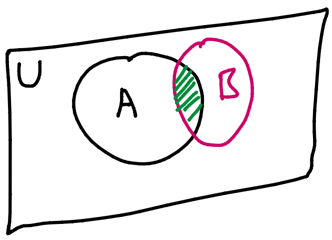
⊛ Sample proportion (p) provides an estimate of π

p success
q failure

$$p = \frac{\pi}{n} = \frac{\text{number of items in the sample having the characteristics of interest}}{\text{sample size}}$$

⊛ $0 \leq p \leq 1$

$$\frac{n(A \cap B)}{n(A)}$$



⊛ p is approximately distributed as a normal distribution when n is large.

$np \qquad q = 1-p$

$\sigma = \sqrt{npq}$

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$

$$\mu_p = \pi \quad \text{and} \quad \sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \qquad \Big| \quad \sigma = \sqrt{npq}$$

$$z = \frac{p-\pi}{\sigma_p} = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

eg:- If the true population of voters who support proposition A is $\pi = 0.4$, what is the probability that a sample size of 200 yields a sample proportion between 0.40 and 0.45.

$P(0.40 \le p \le 0.45) = ?$

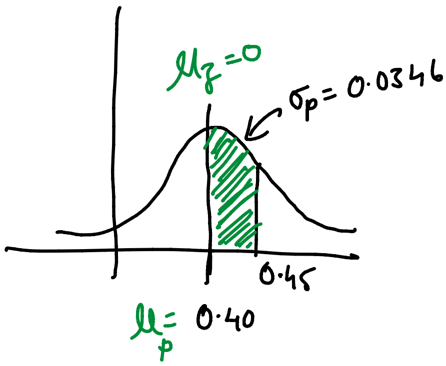
$\Rightarrow$ Here $\pi = 0.4$, $n = 200$,

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \left(\frac{0.4(0.6)}{200}\right)^{1/2} = 0.0346$$

$$P(0.40 \le p \le 0.45) = \left(0 \le z \le \frac{0.45-0.40}{0.0346}\right)$$

$$= (0 \le z \le 1.45)$$

$$\boxed{P(0.40 \le p \le 0.45) = 0.4265}$$



⊛ Testing of Hypothesis:

⊛ Hypothesis:- Quantitative statement about the population.

⊛ Null Hypothesis (H_0):- It is a claim or statement about parameter that is assumed to be true until it is declared to be false.

⊛ Alternative Hypothesis (H_1):- Any hypothesis which is complementary to null hypothesis (Research hypothesis).

500 ml.

— H_0

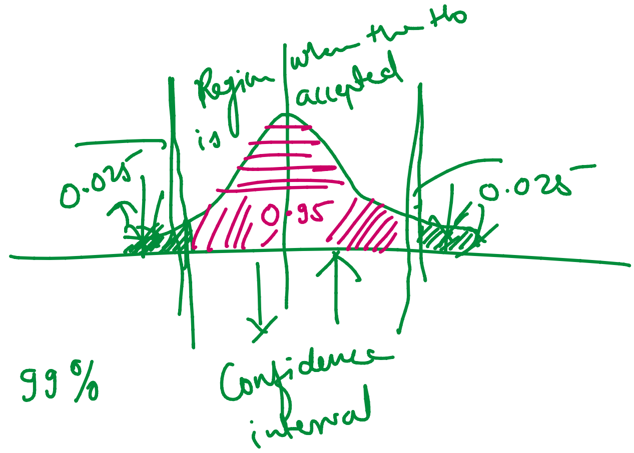
500 ml.

H_0 :- $\mu = 500$

H_1 : $500 > \mu$ false

$500 < \mu$ false

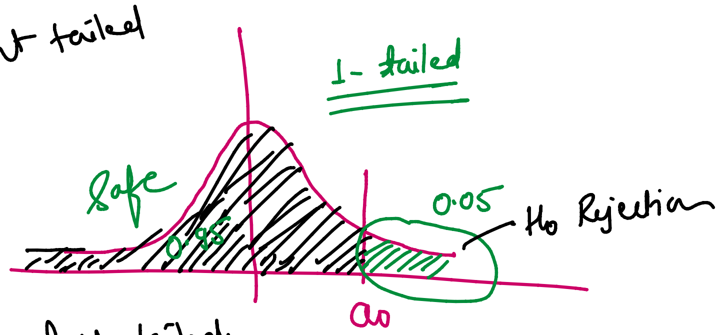
95% 5% , 1% ,



Right tailed

$x \leq a_0$

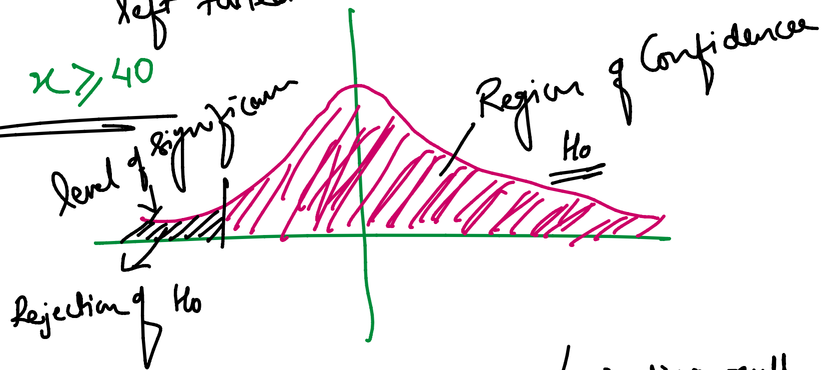
H_0: $x \leq a_0$



H_0: $x = a_0$

left tailed.

H_0: $x \geq 40$



⊛ **Level of Significance** :- (α) Prob of error in accepting / rejecting null hypothesis.

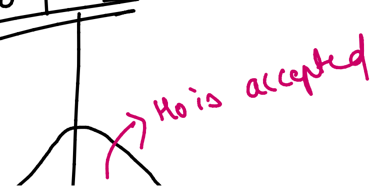
$\alpha = 5\%$ or 1%

$\boxed{95\%}$ $\boxed{99\%}$

Level of Confidence :- (C) $\boxed{C = 1 - \alpha}$

Level of Significance of 5%

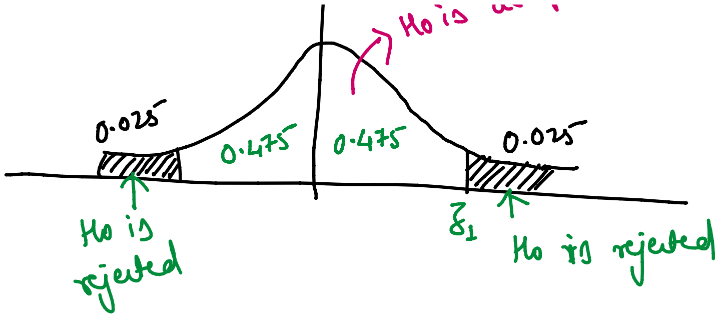
Two tailed



Level of significance of

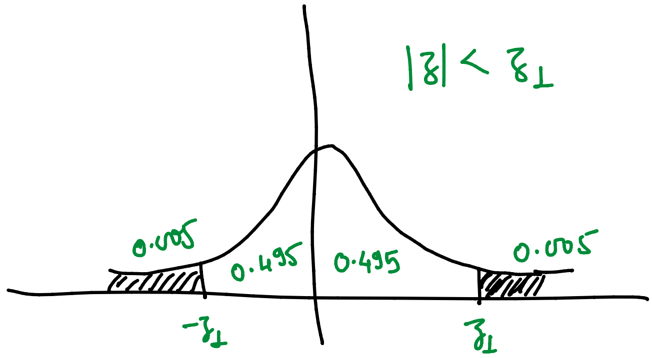
$H_0 =$

$H_1 \neq$



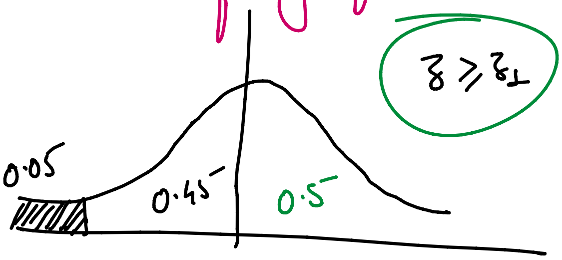
Level of Significance of 1%

$|z| < z_1$

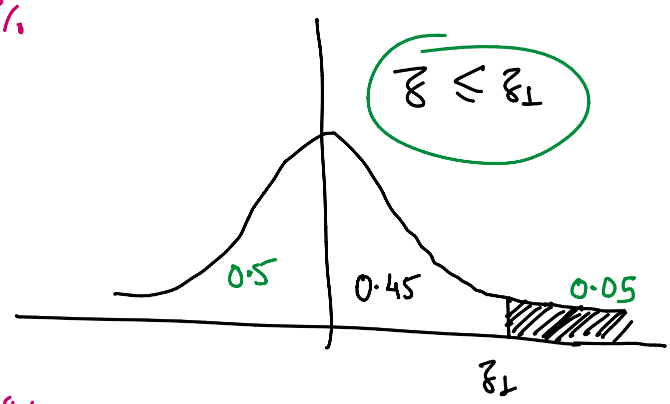


Level of Significance of 5%

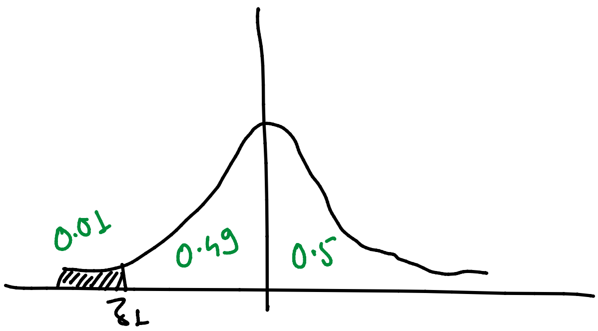
$z \geq z_1$

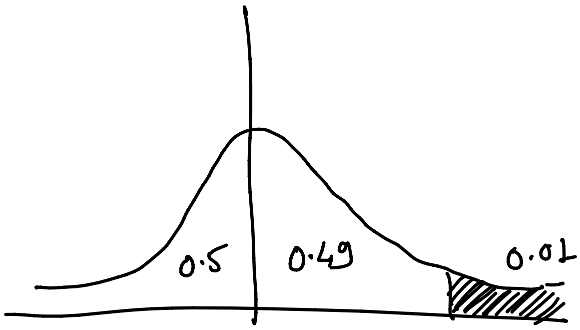


$z \leq z_1$



Level of Significance of 1%





⊛ Computation of Test statistics

$\boxed{\text{Z-test}, \ \text{t-test}}$, F-test,

Annova Test

Z-Test:- ① When $n \geq 30$, $Z = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

std. of sample

Z-Test:-

1. When $n \geq 30$, ... $\sigma_{\bar{x}}$ ...

$\bar{x} \rightarrow$ sample mean, $\sigma_{\bar{x}} \rightarrow$ std. of sample

$\mu \rightarrow$ population mean

2. Z-test is based the standard normal distributions.
3. Z-test is also called as Large sample test.

* T-test.

1. T-test is small sample test. $(n \leq 30)$
2. $$t = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$
3. T-test is used for test of significance of regression coefficient in regression model.
4. We used t-statistics when parameter of population are normal.
5. Population variance is not known

eg:- A manufactures supplies the rear axles for U.S. Postal Service mail trucks. These axles must be able to withstand 80,000 pounds per square inch in stress tests, but an excessively strong axle raises production costs significantly. Long experience indicates that the std. deviation of the strength of its axles is 4,000 pounds per sq. inch. The manufacture selects a sample of 100 axles from production, tests them, and finds that the mean stress capacity of the sample is 79,600 pounds per sq. inch. If the axle manufacturer uses a significance level (α) of 5%, will the axles meet his stress requirement?

$\Rightarrow \quad \mu = 80,000 \;; \quad \sigma = 4,000 \;, \quad n = 100$

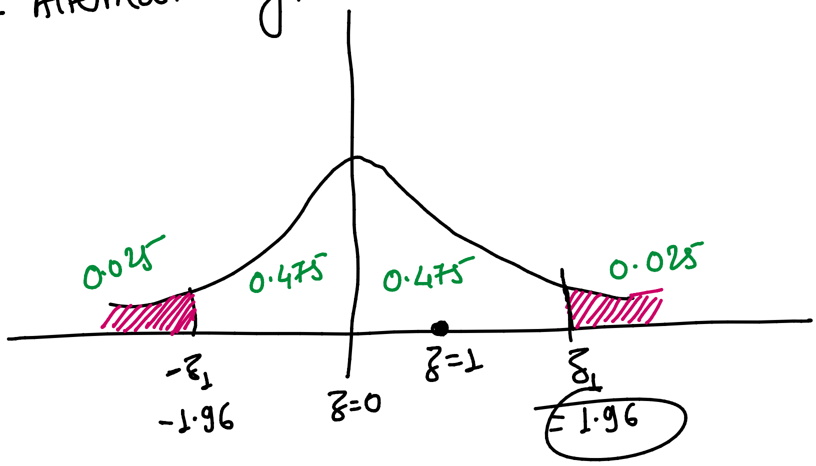
$\bar{x} = 79,600$

$H_0: \mu = 80,000$ ← null hypothes.

$H_1: \mu \neq 80,000$ ← Alternative hypothesis
$(>, <)$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4000}{\sqrt{100}} = 400$$

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{79,600 - 80,000}{400}$$



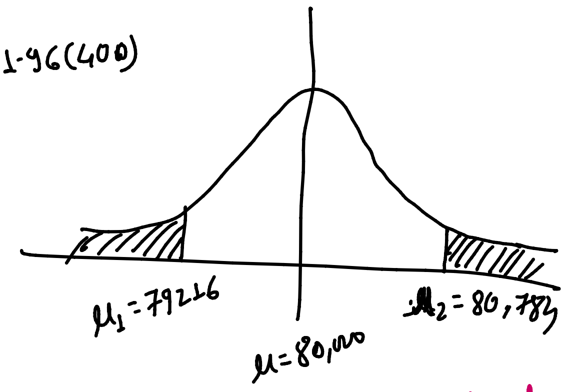
$$\boxed{z = 1}$$

$$-1.96 \leq 1 \leq 1.96$$

Null hypothesis is accepted, alternative hypothesis is rejected.

$$\mu + 1.96\,\sigma_{\bar{x}} = 80,000 + 1.96\,(400) = 80784$$

$$\mu - 1.96\,\sigma_{\bar{x}} = 80,000 - 1.96(400)$$
$$= 79216$$



eg:: Suppose a hospital uses large quantities of packaged doses of a particular drug. The individual dose of this drug is 100 cubic centimeters (100 cc). The action of the drug is such that the body will harmlessly pass off excessive doses. On the other hand, insufficient doses do not produce the desired medical effect, and they interfere with patient treatment. The hospital has purchased ... from the same manufacturer for a number of years and ... 2 cc. The

they interfere with ... a number of
this drug from the same manufacturer for a number of ... known that the population std. deviation is 2 cc. The hospital inspects 50 doses of this drug at random from a very large shipment and finds the mean of these doses to be 99.75 cc. If the hospital sets 1% significance level and ask us whether the dosages in this shipment are too small?

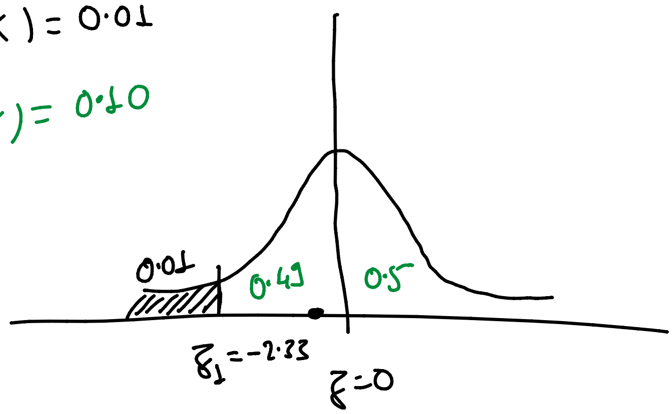
$\Rightarrow$ $\mu = 100$, $\bar{x} = 99.75$, $n = 50$, $\sigma = 2$

$H_0: \mu = 100 \leftarrow$ null hypothesis

$H_1: \mu < 100$

level of significance $(\alpha) = 0.01$

HW. level of significance $(\alpha) = 0.10$



$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

$$= \frac{99.75 - 100}{2/\sqrt{50}} = \frac{-0.25}{0.28}$$

$= -0.892$ $\therefore$ The null hypothesis is accepted.

=x=

Hypothesis testing of Proportions : Large Samples.

Two - tailed tests of Proportions :-

eg:- A company that is evaluating the promotability of its employees, that is determining the promotions whose experience quality them ability, training and supervisory ... management.

for promotion to the next higher level of management.
The human resources director tells the president that roughly 80%, of the employees in the company are promotable. The president assembles a special committee to the promotability of all employees. This committee conducts in depth interviews with 150 employees and finds that in its judgment only 70% of the samples are qualified for promotion. The president wants to test at the 5% significant level the hypothesis that 80% of the employees are promotable.

$\Rightarrow$ $\pi_{H_0} = 0.8 \leftarrow$ Hypothesized value of the population proportion of success

$(1-\pi)_{H_0} = 0.2 \leftarrow$ -,,- of failure

$n = 150$

$\pi_s = 0.7 \leftarrow$ Sample proportion of promotables

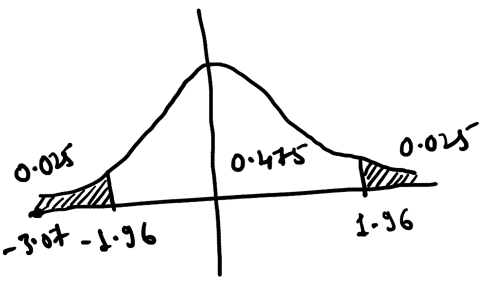
$(1-\pi)_s = 0.3 \leftarrow$ Sample proportions of not promotable

$H_0 = \pi = 0.8 \leftarrow$ Null hypothesis is 80% of the employees are promotable

$H_1 = \pi \neq 0.8$ The proportion of promotable employees is not 80%.

$\alpha = 0.05$ (level of significance)

$$z = \frac{\pi_s - \pi_{H_0}}{\sigma_s}$$



$$\sigma_s = \sqrt{\frac{\pi(1-\pi)}{n}}$$

$$\sigma = \left[\frac{0.8(0.2)}{150}\right]^{1/2} = 0.0325$$

$$z = \frac{0.7 - 0.8}{0.0325} = \frac{-0.1}{0.0325}$$

$$\boxed{z = -3.07}$$

$|z| < 1.96$

Hence the null hypothesis is rejected.

$\therefore$ Alternative hypothesis is accepted.

One-Tailed tests of Proportions

eg:-

A member of a public interest group concerned with environmental pollution asserts at a public hearing that, " fewer than 60 % of the industrial plants in this area are complying with air-pollution standards." Attending this meeting is an official of the environmental ~~production~~ protection agency who believes that 60 % of the plants are complying with the standards; she decides to test that hypothesis at 0.02 significance level. The officials makes a through search of the records in her office. She samples 60 plants from a population of over 10000 plants and finds that 33 are complying with air-pollution standards. Is is assertion by the member of the public interest group a valid one?

$\Rightarrow$ $\pi_{H_0} = 60\% = 0.6$ $\qquad$ $n = 60$

$(1-\pi)_{H_0} = 0.4$ $\qquad$ $\pi_s = 33/60 = 0.55$

$\alpha = 0.02$ $\qquad$ $(1-\pi)_s = 0.45$

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{60}} = \left[\frac{(0.6)(0.4)}{60}\right]^{1/2}$$

$$= \sqrt{0.004}$$

$$\boxed{\sigma_p = 0.0632}$$

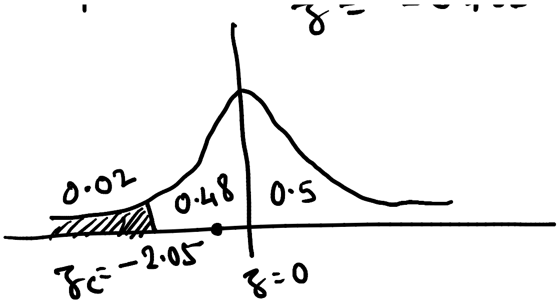
$$z = \frac{\pi_s - \pi_{H_0}}{\sigma_p} = \frac{0.55 - 0.60}{0.0632}$$

$$z = -0.791$$

$H_1 : \pi \geq 0.6$

$H_0: \pi \geq 0.6$

$H_1: \pi < 0.6$



The null hypothesis is accepted

$$t = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

$$\frac{\sigma}{\sqrt{n}}$$

$\boxed{n < 30}$