
IT Helloprogrammers
Note by: Roshan Bist

Chapter-1
Sampling distribution and estimation

* **Estimation** → The statistical method of estimating unknown population parameter from the population is called estimation. The main objective of the estimation is to obtain a guess or estimate of the unknown true value from the sample data or past experience.

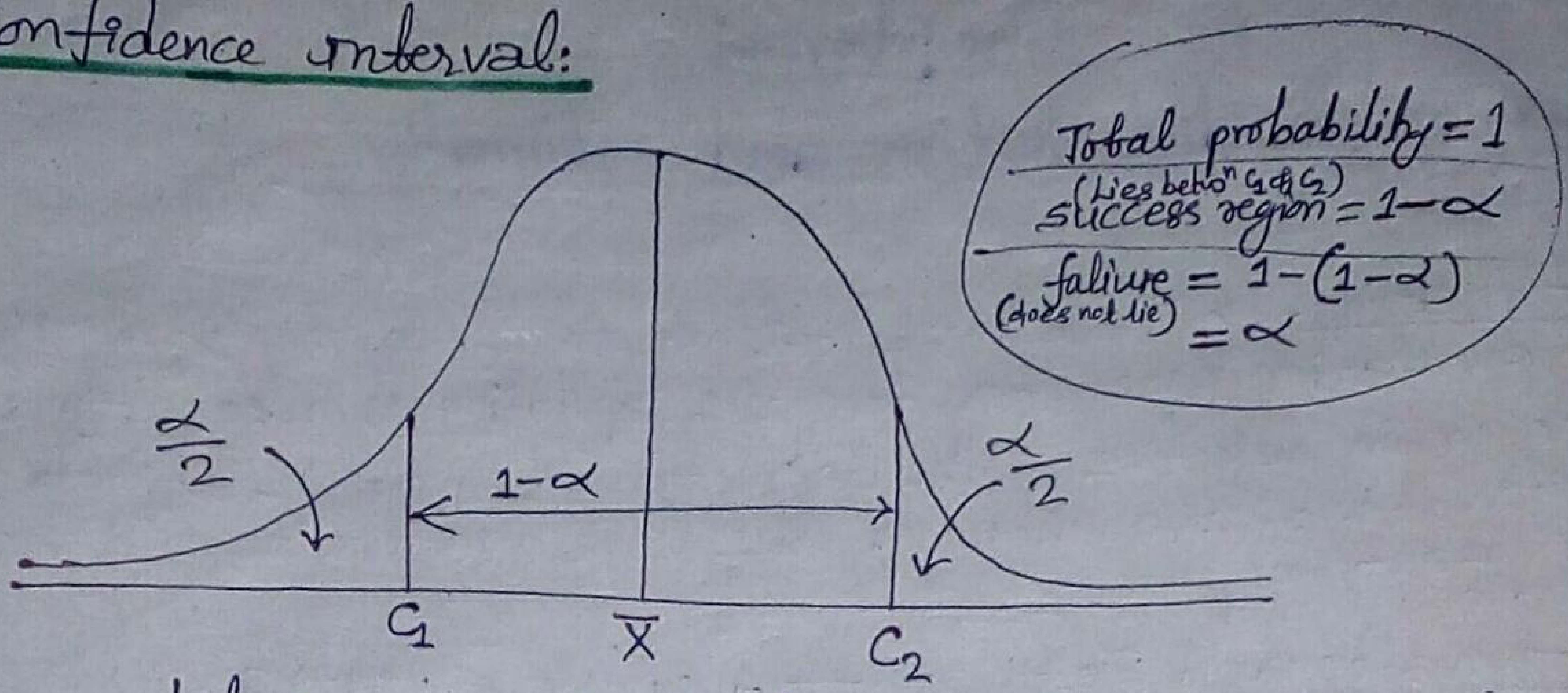
* **Estimates and Estimator** → A sample statistics which is used to estimate a population parameter is called estimator.
For example: The sample mean ($\bar{x}$) is an estimator of population mean (μ), Sample proportion (p) is the estimator of population proportion (P) and the sample standard deviation (S) is the estimator of population standard deviation (σ).
A specific numerical value of estimator is called estimate or in other words an estimate is a specific observed value of a statistic.

* **Types of estimation:**
i) Point estimation → It is the process in which a single sample statistics is used to estimate the population parameter is known as point estimation.
The statistics used to estimate the value of parameter is called point estimator and the numerical value taken by this point estimator is called point estimate.

* **Sample Statistics & Population Parameter:**

Sample Statistics	Population parameter
i) It represent the small portion of population	i) It represents the whole elements in population
ii) Sample mean $(\bar{x}) = \frac{\sum x}{n}$	ii) Population mean $= \mu$.
iii) Sample size $= n$.	iii) Population size $= N$.
iv) Sample standard deviation $(S) = \sqrt{\frac{1}{n-1}\sum(x-\bar{x})^2}$	iv) Population standard deviation $(\sigma) = \sqrt{\frac{1}{N}\sum(x-\bar{x})^2}$
v) Sample correlation coefficient $= r$.	v) Population correlation coefficient $= \rho$
vi) Sample population denoted by small p.	vi) Population parameter is denoted by capital P.

*. Confidence interval:



Let C_1 and C_2 be the lower and upper α limits of confidence interval and θ be the population parameter. Then probability of population parameter θ lies between C_1 and C_2 is known as confidence level or level of confidence and α is the level of significance which is the probability of θ does not lies between C_1 and C_2. So, Probability $P(C_1 \leq \theta \leq C_2) = 1-\alpha$

and α is the probability $\alpha = P$ (θ does not lies between C_1 and C_2).

* Standard Error of Sample Mean($\overline{X}$):-

Statistic	Standard error.
Mean when σ known and population size (N) is infinite.	$S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}}$ (denotes sample standard deviation when σ known)
Mean when σ is known and population size (N) is finite.	$S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$
Mean, when σ is unknown and population size is infinite	$S.E(\overline{X}) = \frac{s}{\sqrt{n}}$ (s denotes sample standard deviation when σ con
Mean when σ is unknown and population size is finite.	$S.E(\overline{X}) = \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$
Difference of means when σ's are unknown.	$S.E(\overline{X}_1 - \overline{X}_2) = \sqrt{\left(s^2\left\{\frac{1}{n_1}+\frac{1}{n_2}\right\}\right)}$
Difference of means when σ's are known	$S.E(\overline{X}_1 - \overline{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

⊛. Confidence interval estimation of population mean (μ):

$$\boxed{\bar{X} - Z_\alpha \cdot S.E \le \mu \le \bar{X} + Z_\alpha \cdot S.E}$$

(If $n \ge 30$)

replace Z_α by t_α, $n-1$ in case of $n < 30$.

Numerical Problems:

Q1. During a water storage a water company randomly sampled ~~resen~~ residensial water metres in order to monitor daily water consumption. On a particular day a sample of 30 meteres showed a sample mean of 240 gallons and sample standard deviation of 45 gallons. Find a 90% confidence interval for the mean water consumption for the population.

Sol^n

Given,

Sample size $(n) = 30$

Sample mean $(\bar{x}) = 240$ gallons.

Sample standard deviation $(S) = 45$ gallons.

level of confidence $(1-\alpha) = 90\%$

$= 0.90$

then, $\alpha = 1 - 0.90$

$= 0.10$

or, $Z_\alpha = Z_{0.10}$

$= 1.64$

i.e, 10%

book को पछाडि दिएको short cut key for Z मा हेर्ने

Now, at 95% level of confidence the confidence interval estimate of population mean μ is,

S.E according to question

$$\bar{X} - Z_\alpha \frac{S}{\sqrt{n}} \le \mu \le \bar{X} + Z_\alpha \frac{S}{\sqrt{n}}$$

$$\text{or, } 240 - 1.64 \times \frac{45}{\sqrt{30}} \le \mu \le 240 + 1.64 \times \frac{45}{\sqrt{30}}$$

$$\text{or, } 226.53 \le \mu \le 253.47$$

Hence the lower limit and upper limit for the mean water consumption for the population are 226.53 and 253.47 respectively at 90% confidence interval.

Q.2. A Random sample of 100 articles selected from a batch of 2000 articles shows that the average diameter of article is 0.354 with a standard deviation 0.048. Find 95% and 98% confidence interval for the average of this batch of 2000 students.

Soln

Given, Sample size $(n) = 100$

Population size $(N) = 2000$

Sample mean $(\bar{X}) = 0.354$

Sample standard deviation $(\sigma) = 0.048$

(i) level of confidence $(1-\alpha) = 95\%$

$= 0.95$

$\Rightarrow \alpha = 0.05$

$\Rightarrow Z_{0.05} = 1.96$

Look at page no 309. table. Calculate $\frac{\alpha}{2}$ and see the value inside the table and find value of Z

OR. Short cut key table for Z

(ii). Also, level of confidence at 98% $(1-\alpha) = 98\%$

$= 0.98$

$\Rightarrow \alpha = 0.02$

$\Rightarrow Z_{0.02} = 2.33$

Now,

At 95% level of confidence the confidence interval estimation of population 'μ' is,

$$\bar{X} - Z_\alpha \cdot S.E \le \mu \le \bar{X} + Z_\alpha \cdot S.E$$

$$\text{or, } \bar{X} - Z_{0.05} \cdot \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \le \mu \le \bar{X} + Z_{0.05} \cdot \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$\text{or, } 0.354 - 1.96 \times \frac{0.048}{\sqrt{100}} \sqrt{\frac{2000-100}{2000-1}} \le \mu \le 0.354 + 1.96 \times \frac{0.048}{\sqrt{100}} \sqrt{\frac{2000-100}{2000-1}}$$

$$\text{or, } 0.3448 \le \mu \le 0.3632$$ //

Again,

At 98% level of confidence the confidence interval estimation of population 'μ' is $\bar{X} - Z_{0.02} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \le \mu \le \bar{X} + Z_{0.02} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

Substituting values and calculating we get,

$$0.3431 \le \mu \le 0.3649$$ //

Q3. Suppose that when a signal having value μ is transmitted from location A. The value received at location B is normally distributed with mean μ and variance 4 i.e. μ is sent, then the value received is $\mu+N$ where N is representing noise is normal with mean 0 and variance 4 to reduce error. Suppose the same value is sent 9 times, if the successive value received are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, and 10.5. Construct 95% confidence interval for μ.

Solⁿ

Given,

X: 5, 8.5, 12, 15, 7, 9, 7.5, 6.5 and 10.5

Sample size $(n) = 9$.

Population variance $(\sigma^2) = 4$

$\Rightarrow \sigma = 2$

$$\text{Sample mean } (\bar{X}) = \frac{5+8.5+12+15+7+9+7.5+6.5+10.5}{9}$$

$= 9$

level of confidence $(1-\alpha) = 95\%$

$= 0.95$

Then $\alpha = 0.05$

$\Rightarrow Z_\alpha = Z_{0.05} = 1.96$.

Now, at 95% level of confidence, the confidence interval estimate of population μ is $\bar{X} - Z_\alpha \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$

or, $9 - 1.96 \times \frac{2}{\sqrt{9}} \leq \mu \leq 9 + 1.96 \times \frac{2}{9}$

or, $7.69 \leq \mu \leq 10.36$ //

Hence the lower limits and upper limits of population mean μ are ~~52.54 and~~ 7.69 and 10.36 respectively.

Q.4. A random sample of size 25 showed a mean of 65 inches with a standard deviation of 25 inches. Determine 98% confidence interval for the mean of the population.

Soln

Given,

Sample size $(n) = 25$.

Sample mean $(\overline{x}) = 65$

Sample standard deviation $(S) = 25$

level of confidence $(1-\alpha) = 98\%$

$= 0.98$

$\Rightarrow \alpha = 0.02$

we use this $t_{\alpha, n-1}$ since till now we were solving questions of $n \geq 30$ case but this question is of type $n < 30$ so we use this method

$\Rightarrow t_{\alpha}, n-1 = t_{0.02}, 25-1$

$= t_{0.02}, 24$

$= 2.492$

Look in table given back of book Page no. 312

Now at, 95% level of confidence interval estimate of population mean μ is,

$$\overline{X} - t_{\alpha}, n-1 \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha}, n-1 \frac{S}{\sqrt{n}}$$

or, $65 - 2.492 \times \frac{25}{5} \leq \mu \leq 65 + 2.492 \times \frac{25}{2}$

or, $52.54 \leq \mu \leq 77.46$.

Hence the lower limits and upper limits of population mean μ are 52.54 and 77.46 respectively.

⊛. Interval Estimation of Population Proportion:

i) Standard error of sample proportion $(p) = \sqrt{\frac{PQ}{n}}$ (When population size is infinite i.e. N & PQ are known)

where,
P = Population proportion.
$Q = 1 - P$
& n = sample size.

ii) Standard error of sample proportion $(p) = \sqrt{\frac{pq}{n}}$ (If p and q are known & population size (N) is infinite)

where, p = sample proportion $= \frac{x}{n}$

$\left[\begin{array}{l} \text{Since, } p = \hat{P} \\ \& \ q = \hat{Q} \end{array}\right]$

iii) Standard error of sample proportion $(p) = \sqrt{\frac{PQ}{n}} \sqrt{\frac{N-n}{N-1}}$ (If Population size is finite i.e. N & PQ known)

iv) Standard error of sample proportion $(p) = \sqrt{\frac{pq}{n-1}} \cdot \sqrt{\frac{N-n}{N}}$ (If p and q are known & N is finite)

$\therefore$ Interval Estimation of population proportion (P) is

$$\boxed{p - Z_\alpha \cdot S.E \leq P \leq p + Z_\alpha \cdot S.E}$$

Q5. It is observed that 28 successes in 70 independent Bernoulli trial. Compute 90% confidence interval for population proportion.

Soln

Given, no. of independent trial $(n) = 70$.
no. of sucess $(X) = 28$
So, proportion of sucess $(p) = \frac{X}{n} = \frac{28}{70} = 0.4$

$q = 1 - \frac{28}{70}$
$= 0.6$

level of confidence $(1-\alpha) = 90\% = 0.90$
$\Rightarrow \alpha = 0.10$
$\Rightarrow Z_\alpha = Z_{0.10}$
$= 1.65$

Now, at 90% level of confidence the confidence interval estimate of population proportion 'p' is,

$$p - Z_\alpha \cdot S.E \leq P \leq p + Z_\alpha \cdot S.E$$

or, $0.4 - 1.65 \times \sqrt{\frac{pq}{n}} \leq 0.4 + 1.65 \times \sqrt{\frac{pq}{n}}$

or, $0.4 - 1.65 \times \sqrt{\frac{0.4 \times 0.6}{70}} \leq 0.4 + 1.65 \times \sqrt{\frac{0.4 \times 0.6}{70}}$

or, $0.304 \leq P \leq 0.496$.

Hence, the lower and upper limit of population proportion are 0.304 and 0.496 respectively.

Q.6. A random sample of 80 people form a community of 300 showed that 30 were smoker. Find 90% of confidence limit for the proportion of smoker.

Sol^n

Given, Population size $(N) = 300$

Sample size $(n) = 80$

no. of smoker $(x) = 30$.

Sample proportion of smoker $(p) = \frac{x}{n} = \frac{30}{80} = 0.375$

$q = 1 - 0.375 = 0.625$

level of confidence $(1-\alpha) = 99\% = 0.99$

$\Rightarrow \alpha = 0.01$

$\Rightarrow Z_\alpha = Z_{0.01} = 2.58$

Now, at 99% level of confidence, the confidence interval estimate of pop^n proportion p is

$$p - Z_{0.01} \times \sqrt{\frac{pq}{n-1}} \cdot \sqrt{\frac{N-n}{N}} \leq P \leq p + Z_{0.01} \sqrt{\frac{pq}{n-1}} \sqrt{\frac{N-n}{N}}$$

or, $0.375 - 2.58 \times \sqrt{\frac{0.375 \times 0.625}{80-1}} \sqrt{\frac{300-80}{300}} \leq P \leq 0.375 + 2.58 \times \sqrt{\frac{0.375 \times 0.625}{80-1}} \sqrt{\frac{300-80}{300}}$

or, $0.2546 \leq P \leq 0.4953$

Hence the lower limit and upper limit of population proportion of smoker are 0.2546 and 0.4953 respectively.

यो। probability value
अर्को number find गर्न N ले multiply गर्ने.

⊛ Determination of Sample Size (n):

(a) By using mean:- Let a population of size N and drawing a sample from population with size n and population standard deviation is σ, then by using central limit theorem (CLT).

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Value of Z at α level of significance

$$\text{or, } Z_\alpha = \frac{E}{\sigma/\sqrt{n}} \quad (\text{where Error } (E) = |\bar{x} - \mu|)$$

$$\text{or, } Z_\alpha = \frac{E\sqrt{n}}{\sigma}$$

$$\text{or, } \sqrt{n} = \frac{Z_\alpha \cdot \sigma}{E}$$

$$\text{or, } \boxed{n = \left(\frac{Z_\alpha \cdot \sigma}{E}\right)^2}$$

(b) By using proportion:- Let a sample size (n) is drawn from the population with population proportion (P) and the sample proportion is (p) then, we have

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$\text{or, } Z_\alpha = \frac{E}{\sqrt{\frac{PQ}{n}}} = \frac{E\sqrt{n}}{\sqrt{PQ}} \quad (\because E = |p - P|)$$

$$\text{or, } \sqrt{n} = \frac{Z_\alpha}{E}\sqrt{PQ}$$

$$\text{or, } \boxed{n = \left(\frac{Z_\alpha}{E}\right)^2 \cdot PQ}$$

Note: If P and Q are unknown/not given then take $P = Q = 0.5$

Q.7. Assuming population standard deviation 3, how large should a sample be to estimate population mean with margin of error not exceeding 0.5?

Solⁿ Given, Population standard deviation $(\sigma) = 3$.

Error $(E) = 0.5$

level of significance$(\alpha) = 0.05$

Now,

$$\text{sample size } (n) = \left(\frac{Z_\alpha \cdot \sigma}{E}\right)^2$$

$$= \left(\frac{Z_{0.05} \times 3}{0.5}\right)^2$$

$$= \left(\frac{1.96 \times 3}{0.5}\right)^2$$

$$= 138.29$$

$$\approx 138$$

Q.8. The principle of a college wants to estimate the proportion of students who were interested to develop startup. What size of a sample should he select so as to have the difference of ~~po~~ proportion of interested students with true mean not to exceed by 10% with almost certainity? It is believed from previous records that the proportion of interested students was 0.30?

Question मा almost certainity भने Z_α को value 3 हुन्छ।

Solⁿ Given, Error $(E) = 10\%$

$= 0.10$

Population proportion $(P) = 0.30$

then, $Q = 1 - 0.30 = 0.70$

Since this is the case of almost certainity, so we take $Z_\alpha = 3$.

Now, $$n = \left(\frac{Z_\alpha}{E}\right)^2 \cdot PQ$$

$$= \left(\frac{3}{0.10}\right)^2 \times 0.30 \times 0.70$$

$$= 189$$

Scanned by TapScanner

Q.N.9. A random sample of size 64 has been drawn from a population with standard deviation 20. The mean of sample is 80. Calculate 95% confidence limit for the population mean. How does the width of confidence interval change if sample size is 256 instead?

Soln Given,

Sample size $(n) = 64$

Population standard deviation $(\sigma) = 20$

Sample mean $(\bar{x}) = 80$

level of confidence $(1-\alpha) = 95\%$

$= 0.95$

$\alpha = 0.05$

$\Rightarrow Z_\alpha = Z_{0.05} = 1.96$

Now, at 95% level of confidence, the confidence interval estimate of population mean 'μ' is,

$$\bar{x} - Z_{0.05}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + Z_{0.05}\frac{\sigma}{\sqrt{n}}$$

or, $80 - 1.96 \times \frac{20}{\sqrt{64}} \le \mu \le 80 + 1.96 \times \frac{20}{\sqrt{64}}$

or, $75.1 \le \mu \le 84.9$

Here the width of confidence interval is $84.9 - 75.1$

$= 9.8$

If $n = 256$ then,

$$80 - 1.96 \times \frac{20}{\sqrt{256}} \le \mu \le 80 + 1.96 \times \frac{20}{\sqrt{256}}$$

or, $77.55 \le \mu \le 82.45$

Here, width of confidence interval is $82.45 - 77.55$

$= 4.90$

Hence, if we increase the size of sample then width of confidence interval decreases.

Scanned by TapScanner