Sampling Distributions

Sampling is a first step to draw a statistical inference about population when it is not possible or immpractical to observe the entire set of population.

For example:

1) To determine the averge life of electric bulbs of a certain brand, it would be impossible to test all such bulbs.

2) To determine the percentage of defective apples from a lot of ~~production~~. apple farm.

Imp notations in this theory

① population:

A collection of objects or individuals is said to be population.

~~Here. It may be~~

Here, the no. of observations in the population is defined as the size of the population. It may be finite or infinite.

Size of the population denoted by N.

Sample A finite subset of a population is called sample: and the selection process is known as sampling. The size of the population is denoted by n.

1) To assess the quality of a bag of rice, sugar, wheat, we examine only portion of it by taking a handful of it and decide to purchase it or not.

Here whole quantity of sugar is population and selected handful portion is sample.

ii) To estimate the portion of defective articles in a large consignment, only a portion is selected and examined.

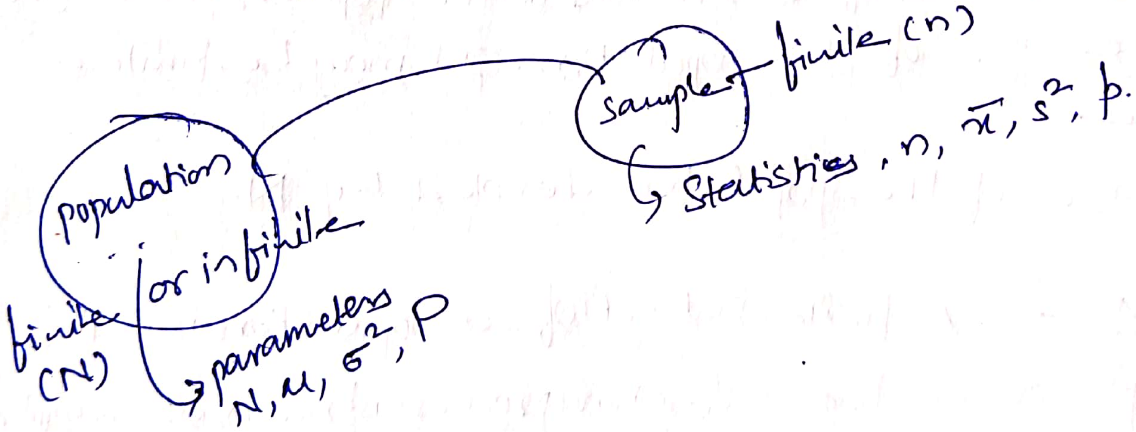
③ Sample error:

In the enumeration of population characteristics, the sample characteristics are used to estimate the population approximatly. Then the error involved in such approximation is known as sample error. and it is unavoidable.

④ parameters: The characteristics of the population such as mean, ~~s.d~~, variance, mode and median are called parameters. Denoted by N, μ, σ^2, P.

⑤ Statistics: A statistical constants obtaine from the samples are said to be statistics.

Denoted by $n, \bar{x}, s^2, p$. etc.



~~To also estimate the~~

If we wish to to draw a vaild inference about population with by sampling, The selected samples should ~~to be~~ representative of population and unbiased.

So, selection of samples is very immportant in this theory.

Scanned with CamScanner

Methods for Sampling:

I Probability sampling Methods.

1) Random sampling:

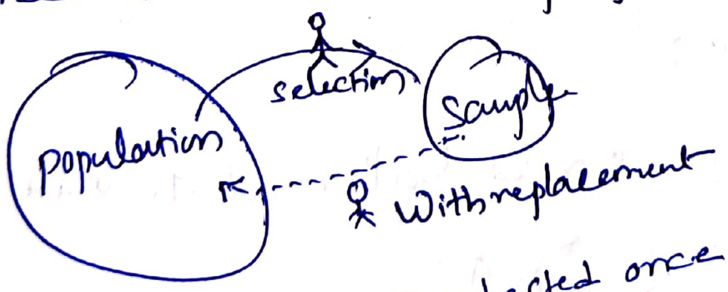
If each and every every element of the population has an equal chance of inclusion in the sample, it is called random sampling.

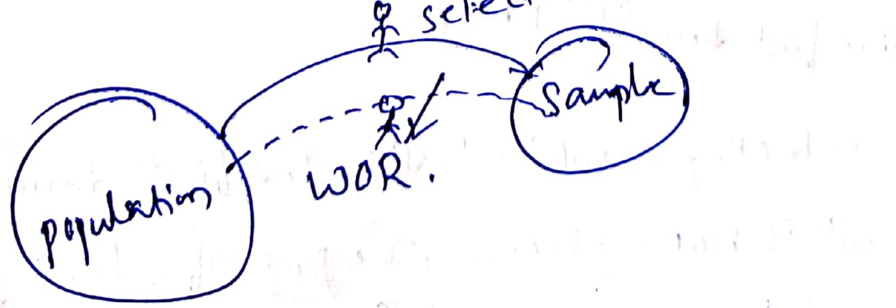
ex selection of one playing card from a well-shuffled pack.

2) choosing 10 patient from a hospital to test the ~~effi~~ efficiency of a certain drug.

Here, There is a chance to select one element more than once when it is replaced.

If each element of a population is selected more than once, then it is called sampling with replacement whereas, if the element cannot be selected more than once it is called sampling with out replacement.





If size of the population is finite N and n is the sample size, then.

i) No. of samples with replacement $= N^n$.

ii) No. of samples without replacement $= {}^N C_n$.

Simple sampling:

Simple sampling is a special case of random sampling, in which each element of the population has an equal and independent chance of being included in the sample.

ex Selection of 1 match sticks from a lot of productions with replacement is not possible.

So, Selection must be with out replacement and this selection is not simple.

~~Here~~ For a finite population, random sampling with replacement is a simple sampling while random sampling without replacement is a not a simple.

2. Stratified Sampling

If a basket contains 25 red balls and 40 white balls.

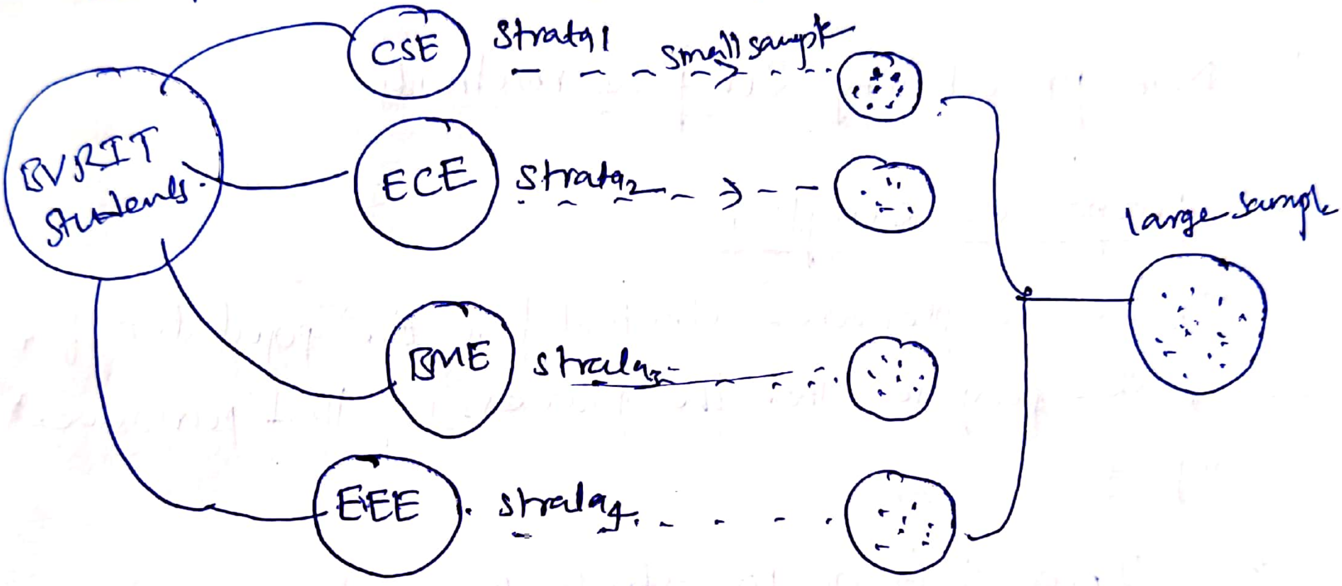
25 red, 40 white → 1 white ball in first draw → $P(W) = \frac{40}{65}$.

The prob. of selecting a white ball in the first draw is $\frac{40}{65}$. If ball is not replaced, the prob of selecting red ball is changed from $\frac{25}{65}$ to $\frac{25}{64}$. So, selection of second ball is dependent on first selection. This only by WOR.

② Stratified sampling:

A hetrogeneous population may often be divided into sub population of strata or subgroups which are more homogeneous with itself. Then a small sample is selected from each stratum at random and combined together to form the stratified sample. This process of selection is called Stratified sampling.

example Find Avqurage height of BVRIT students.



Stratified sampling.

ex The people in a small village may be divided into 2 strata literates and illiterates.

③ Systematic sampling:

In this method, all the elements of the population are arranged in some order. Then from the first 'k' items, one unit is selected at random. ~~The~~ Again, this unit and every k^{th} unit of the serially listed population combined together constitute a systematic sample.

This type of sampling is known as systematic sampling.

The difference between random sampling and this lies in the fact that.

All the members have to be chosen randomly; whereas in this case only the first member has to be chosen at random from the first 'k' items ~~order~~

Ex: ~~If you which to height of 1000 BVRIT students. Arrange them according to their weigh~~

(II) Non probability sampling methods

(4) purposive sampling

If the samples are selected from ~~the~~ population for a definite purpose, Then the process is called purposive sampling.

Ex In a class of 100 students, 20 students are selected by a class teacher to analyse the extra-curricular activites.

(5) Sequential sampling:

It consists of a sequence of sample drawn one after another from the population depending on the results of previous samples.

Large and Small samples

④

If the size of the sample is equal to or greater than 30; it is refered as a large sample. Otherwise it is regarded as small sample.

Population distribution

Let ~~the~~ $x_1, x_2 \cdots x_N$ are elements of population and the charactristics of population denoted by parameters N, P, μ, σ^2 Then

population mean $= \dfrac{\Sigma x_i}{N}$ & variance $\sigma^2 = \dfrac{\Sigma (x_i - \mu)^2}{N}$.

Sample mean and variance

Let $x_1, x_2 \cdots x_n$ be the 'n' elements of random sample taken from a population of size N. Then

sample mean $\bar{x} = \dfrac{1}{n} \Sigma x_i$

Sample variance $s^2 = \dfrac{1}{n-1} \Sigma (x_i - \bar{x})^2$

Sampling distribution of mean:

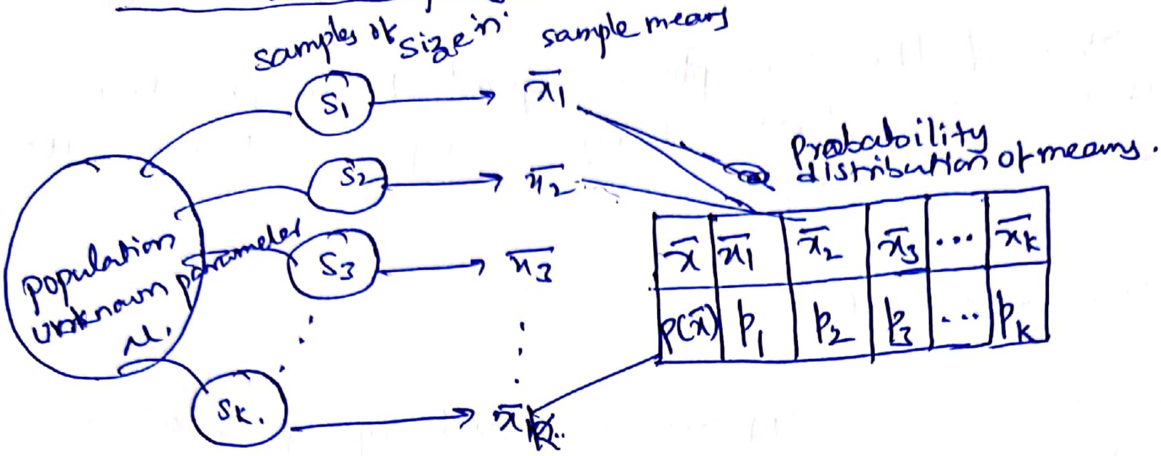
~~Let us~~ Consider a finite population of size N with mean μ and variance σ^2; ~~If~~ ~~Let us~~

suppose all possible 'K' no. of samples of size 'n' drawn from population randomly and $\bar{x}$ is the mean of samples.

ie $\bar{x}_1, \bar{x}_2 \cdots \bar{x}_k$ are ~~samples of~~ respective means of k samples.

Then set of values $\bar{x}_1, \bar{x}_2, \cdots \bar{x}_k$ of statistic mean $\bar{x}$ is known as sampling distribution of mean. It is ~~used~~ used to estimate unknown population mean. or to draw statistical inference about mean.

Mean of sampling distribution of means:



$\bar{x}$	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	...	$\bar{x}_k$
$P(\bar{x})$	p_1	p_2	p_3	...	p_k

~~Then~~ Here $\bar{x}$ is a sample mean and it is defined as random variable with probability function $P(\bar{X}=\bar{x}_i)=p_i, i=1,2\ldots k$.

Then mean of sampling distribution of means is defined and denoted as

$$\mu_{\bar{x}} = E(\bar{x}) = E\left[\frac{x_1+x_2+\cdots+x_n}{n}\right] \quad (\because \bar{x} \text{ is sample mean.})$$

$$= \frac{1}{n}\left[E(x_1)+E(x_2)+\cdots+E(x_n)\right] \quad \because x_1, x_2 \cdots x_n \text{ are also elements of population.}$$

$$= \frac{1}{n}\left[\mu+\mu+\cdots+\mu\right]$$

$$= \frac{\cancel{n}\mu}{\cancel{n}} = \mu. \quad \therefore \boxed{\mu_{\bar{x}} = \mu}$$

Variance of sampling distribution of means denoted by $\sigma^2_{\bar{x}}$ and defined as

$$\sigma^2_{\bar{x}} = var(\bar{x}) = var\left[\frac{x_1+x_2+\cdots+x_n}{n}\right]$$

$$= \frac{1}{n^2}\left[var(x_1)+\cdots+var(x_n)\right]$$

$$= \frac{1}{n^2}\left[\sigma^2+\cdots+\sigma^2\right] \text{ (n times)}$$

$$= \frac{1}{n^{\cancel{2}}}\cancel{n}\sigma^2$$

$$\boxed{\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}}$$

(5)

Note i) Suppose the samples are drawn from an infinite population i.e. $N \to \infty$ or sampling is done with replacement then $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$

ii) If samples are drawn without replacement from a finite population of size N.

Then $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

Here $\left(\frac{N-n}{N-1}\right)$ is called correction factor.

iii) The sampling distribution of $\bar{x}$ will be approximately "normal" with mean $\mu_{\bar{x}} = \mu$ and variance $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ provided that the sample size is large.

Central limit theorem:

If $\bar{x}$ be the mean of a sample size 'n' drawn from a population with mean μ and S.D σ then standardized sample mean $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ is a random variable whose distribution function approaches that of standard normal distribution ~~with~~ $N(z; 0,1)$ as $n \to \infty$. i.e when n is large, $\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}})$

Q1 A population consists of the four numbers 1, 5, 6, 8. Consider all possible samples of size two that can be drawn without replacement from this population. Find

i) The population mean $\rightarrow \mu = 5$

ii) The population standard deviation $\rightarrow \sqrt{6.5} = \sigma$

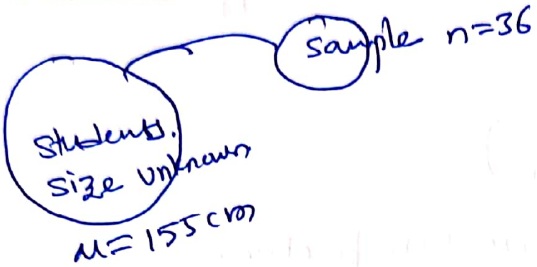
iii) Mean of sampling distribution of means $\rightarrow \mu_{\bar{x}} = 5$

iv) standard deviation of sampling distribution of means.

$\sigma_{\bar{x}} = 1.612$

Q2: The mean height of students in a college is 155 cms and standard deviation is 15. What is the probability that the mean ~~of~~ height of 36 students is less than 157 cms.

Sol: A r.v is mean height of students in every sample of 36 students.



mean of population $= 155 = \mu$

S.D of population $= 15 = \sigma$

$n = 36$ = sample size.

$$P(\bar{x} \leq 157) = P\left[\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq \frac{157-155}{\frac{15}{\sqrt{36}}}\right] = P(z < 0.8)$$

$$= 0.7881.$$

Thus prob that the mean height of 36 students is less than 157 is 0.7881

Q3 A population consists of the four numbers 3, 7, 11, 15. Consider all possible sample of size 2 that can be drawn with replacement from this population. Find

i) μ ii) σ iii) $\mu_{\bar{x}}$ iv) $\sigma_{\bar{x}}$ v) verify the results.

Sol Given population is 3, 7, 11, 15

i) mean of population $= \mu = \frac{3+7+11+15}{4} = 9$

ii) variance of population $= \sigma^2 = \frac{1}{4}\left[(3-9)^2 + (7-9)^2 + (11-9)^2 + (15-9)^2\right]$

$\sigma^2 = 20$, and $\sigma = \sqrt{20} = 4.4721$.

List of possible samples of size 2 ~~that~~ by with replacement are.

$= \{(3,3), (3,7), (3,11), (3,15), (7,3), (7,7), (7,11), (7,15)$
$(11,3)\ (11,7)\ (11,11)\ (11,15), (15,3)\ (15,7)\ (15,11)\ (15,15)\}$.

Now list of sample means for every sample are

$= \{3, 5, 7, 9, 5, 7, 9, 11, 7, 9, 11, 13, 9, 11, 13, 15\}$.

so, sampling distribution of means is

$\bar{x}$	3	5	7	9	11	13	15	total
f_i	1	2	3	4	3	2	1	16

on counting $\mu_{\bar{x}} = 9 = \mu$ (iii)

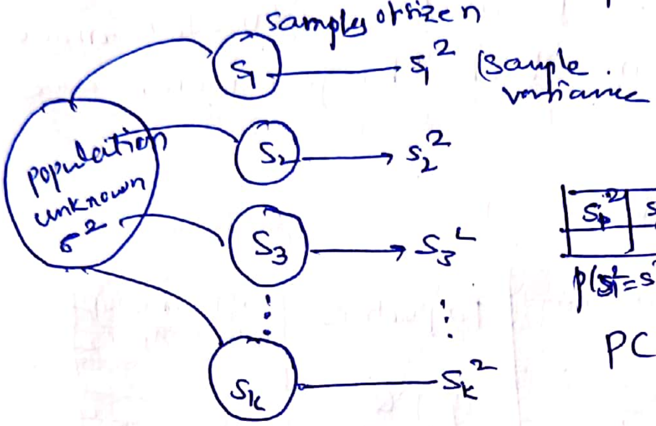
(iv) $\sigma_{\bar{x}} = 3.1622 = \frac{\sigma}{\sqrt{n}}$

verified the results.

Sampling distributions of variances:

As like sampling distribution of means, collect the all possible samples to estimate population variance σ^2 by sample variance s^2.

The series of values $s_1^2, s_2^2, \dots s_k^2$ of statistic sample variance s^2 is known as sampling distribution of variance.



s^2	s_1^2	s_2^2	s_3^2	...	s_k^2
$p(s^2=s_i^2)$	p_1	p_2	p_3	...	p_k

$P(s^2 = s_i^2) = p_i. \quad i = 1, 2 \dots k.$

Then $\mu_{s^2} = E(s^2) = \sigma^2$ and $\sigma_{s^2}^2 = var(s^2) = \sigma^2\sqrt{\frac{2}{n}}$.

Q1 Find the mean and standard deviations of sampling distributions of variances for the population 2, 3, 4, 5 by drawing samples of size two ⓐ without replacement ⓑ with ~~out~~ replacement.

Sol population size, $N=4$, sample size, $n=2$

ⓐ N.o. of samples of size 2 without replacement $^4C_2 = 6$.

They are, $\underset{s_1}{(2,3)}, \underset{s_2}{(2,4)}, \underset{s_3}{(2,5)}\ \underset{s_4}{(3,4)}, \underset{s_5}{(3,5)}, \underset{s_6}{(4,5)}$

compute variances ~~for~~ each of these 6 samples.

Variance of (2, 3) with mean 2.5 is $s_1^2 = \frac{1}{2-1}(2-2.5)^2 + (3-2.5)^2 = 0.5$

$s_2^2 = \frac{1}{2-1}[(2-3)^2 + (4-3)^2] = 2.$

$s_3^2 = \frac{1}{2-1}[(2-3.5)^2 + (5-3.5)^2] = 4.5$

$s_4^2 = 0.5,\ s_5^2 = 2,\ s_6^2 = 0.5.$

sample variance $= \begin{Bmatrix} 0.5, 2, 4.5 \\ 0.5, 2, 4.5 \end{Bmatrix}$.

sampling distribution of variance

s^2	0.5	2	4.5
f_i	3	2	1

$E(s^2) = \frac{1}{6}[0.5\times3 + 2\times2 + 4.5\times1]$

$= 1.6666 = \sigma^2$

Standard Error of a Statistic:

~~Standard error of a stati~~

The standard deviation of the sampling distribution of statistic is called the standard error (S.E). It is used to assess the difference between expected and observed values. It plays an important role in large sample theory and forms basis in tests of significance. S.E enables us to determine the confidence limits within which the parameters are expected to lie.

A table of S.E of Statistics

statistic	Mean $\bar{x}$	S.D s	variance	proportions	Difference of means $\bar{x}_1 - \bar{x}_2$
S.E	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	$\frac{\sigma}{\sqrt{2n}}$	$\sigma^2\sqrt{\frac{2}{n}}$	$\sqrt{\frac{PQ}{n}}$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

For a finite population of size N, when sample is drawn without replacement, we have

i) S.E of sample mean $= \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

ii) S.E of sample variance $= \sigma^2\sqrt{\frac{2}{n}}\sqrt{\frac{N-n}{N-1}}$

iii) S.E of sample proportion $= \sqrt{\frac{PQ}{n}}\sqrt{\frac{N-n}{N-1}}$

Questions.

1) A sample of size 400 is taken from population whose S.D is 16. Find the S.E Ans 0.8.

Sol. Sample size $n = 400$ and population S.D $\sigma = 16$

So, S.E of sample means for large samples. $= \frac{\sigma}{\sqrt{n}}$

$$= \frac{16}{\sqrt{400}} = 0.8.$$

2) The mean weekly wages of workers are with S.D of Rs 4. A sample of 625 is selected. Find the S.E of the mean.

Ans. 0.16. Given $\sigma = 4$, $n = 625$.

S.E $= \frac{4}{\sqrt{625}}$

(3) When we draw a sample from an infinite population, what (7) happens to the S.E of the mean if the sample size is (a) Increased from 50 to 200 (b) Decreased from 640 to 40.

(4) The mean of certain normal population is equal to S.E of mean of the samples of 64 from the distribution. Find the probability that the mean of the sample size 36 will be negative

Sol The S.E of mean $= \frac{\sigma}{\sqrt{n}}$. and Sample size $= 64$

Given that mean $\mu =$ S.E of means.

ie $\mu = \frac{\sigma}{\sqrt{64}} = \frac{\sigma}{8} \rightarrow$ population mean

we know sampling distribution of mean will follow normal distribution with Statistic z when sample is large.

So, $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{\bar{x} - (\sigma/8)}{\sigma/6}$ (here sample size 36)

$z = \dfrac{6\bar{x}}{\sigma} - \dfrac{3}{4}$. and if $z < 0.75$, $\bar{x}$ will be $-$ve.

so, $P(z < 0.75) = P(-\infty < z < 0.75) = \underline{0.7734}$

(5) A random sample of size 100 is taken from an infinite population having mean 76 and variance 256. What is the probability that $\bar{x}$ will be between 75 and 78.

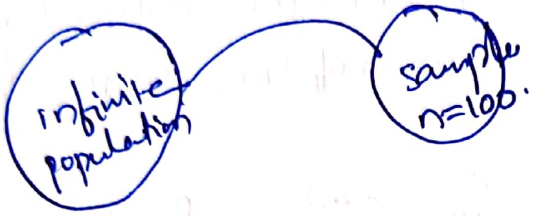
(6) The guaranteed average life of a battery is 700 days with S.D of 60 days. It is required to sample output so as to ensure that 95% of the batteries do not fall short of guaranteed average life by more than 2.5%. What is the minimum sample size?

Solutions of Q.5 and (6)

Q5 Sol Given a random sample of size 100 is taken from infinite population whose mean is 76 and variance is 256.

ie $\mu = 76$

$\sigma^2 = 256$, $\sigma = 16$.



$\mu = 76$, $\sigma = 16$.

Let $\bar{x}$ be the sample mean. and it denote: random variable of sampling distribution of means.

Also, the sampling distribution means will be approximated by normal distribution with statistic z.

Since, the sample size $n = 100$, and it is large sample. z It obeys central limit theorem

i.e $\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}})$

Question is, what is the probability that $\bar{x}$ will be between 75 and 78

ie $P(75 \leq \bar{x} \leq 78) = P\left(\frac{75-76}{16/\sqrt{100}} \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq \frac{78-76}{16/\sqrt{100}}\right)$ (1).

Since statistic z is a standard normal variable of normal distribution of means $\bar{x}$. and it defined as

$$z = \frac{\bar{x} - \text{mean of sampling}}{\text{S.D of sampling}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

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

$\leq X \leq 18) = P(-0.625 \leq z \leq 1.25)$ 8

$= \underline{0.628}.$

Q6

Let 'n' be the sample size.

Given that the guaranteed average life a battery is 700 days with S.D of 60 days.

ie ~~population mean~~

ie population mean $\mu = 700$ and

Population S.D $\sigma = 60$.

We do not want the mean of sample to be less than 2.5% of 700, i.e 17.5 days.

So, sample mean must be more than of 700 − 17.5

ie $\bar{x} > 682.5$ — ①

For this, the percentage of ~~a~~ guaranteed average life is 95%.

ie $P(\bar{x} > 682.5) = 0.95$ — ②.

By central limit theorem, $\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}})$ when ~~the~~ sample size is large.

Here n is assumed to be large,

then $P(\bar{x} > 682.5) = 0.95$

$\Rightarrow P\left(\frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} > \frac{682.5 - \mu_{\bar{x}}}{\sigma_{\bar{x}}}\right) = 0.95$

$\Rightarrow P\left(z > \frac{682.5 - \mu}{\sigma/\sqrt{n}}\right) = 0.95$

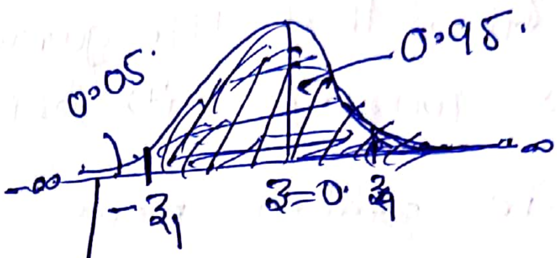
or $P(z > z_1) = 0.95$ — ③

~~Here~~

9

Here $-z_1 = \dfrac{682.5 - 700}{60/\sqrt{n}}$ ——— (4)

From Eq (3), one can notice that the area under standard normal curve to the left of z_1 should be 0.95.



Now from the rough sketch and S.N.D table

$-z_1 = -1.64$ ——— (5) $P(z < -z_1) = 0.05$

since $P(z < -1.64) = 0.05$.

Now, use the value of $-z_1$ in Eq (4), we get

Eq (4) $\Rightarrow -1.64 = \dfrac{682.5 - 700}{60/\sqrt{n}}$ $\sigma = 60$

$\Rightarrow \cancel{-}1.64 = \cancel{-}\dfrac{17.5}{60/\sqrt{n}}$

$\Rightarrow 1.64 = \dfrac{17.5}{60} \times \sqrt{n} \Rightarrow \sqrt{n} = \dfrac{60 \times 1.64}{17.5}$

$\Rightarrow \sqrt{n} = 5.63$

$\Rightarrow n = 31.6969$

$\Rightarrow n \simeq \underline{\underline{32}}$

$\therefore$ The minimum sample size is $\underline{\underline{32}}$

Additional questions:

(7) How many different samples of size $n=2$ can be choosen from a finite population of size $N=25$

sol $N=25$, and $n=2$

No. of different samples of size 2 is $^{25}C_2 = 30$.

8) Find the value of the finite population correction factor for $n=5$ and $N=200$. (9)

Sol Correction factor $= \frac{N-n}{N-1} = \frac{200-5}{200-1} = 0.9799$.

(9) A population consist of six numbers 4, 8, 12, 16, 20, 24. Consider all samples of size 3 which can be drawn with out replacement from this population. Find (a) μ (b) σ (c) $\mu_{\bar{x}}$ (d) $\sigma_{\bar{x}}$. Verify the results.

(10) A random sample of size 64 is taken from a normal population with $\mu = 51.4$ and $\sigma^2 = 68$. What is the probability that the mean of the sample will (a) fall between 50.5 and 52.3 (b) exceed 52.9 (c) be less than 50.6.

(11) A normal population has mean of 0.1 and S.D of 2.1 Find the probability that mean of a sample of size 900 will be negative.

Sol Given $\mu = 0.1$ $\sigma = 2.1$ and $n = 900$.

$P(\bar{x} < 0) = ?$

$= P(z < 0)$ here $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

$\Rightarrow z = \frac{\bar{x}-0.1}{2.1/\sqrt{900}}$

so $P(\bar{x} < 0)$

$\Rightarrow z = \frac{\bar{x}-0.1}{0.07}$

$P(0.1 + 0.07z < 0)$

$\Rightarrow \bar{x} = 0.1 + 0.07z$.

$P(0.07z < -0.1)$

$P(z < \frac{-0.1}{0.07})$

$\Rightarrow P(z < -1.43) = 0.0764$ from S.N.D table.

Sampling distributions of proportions:

Let P be the population proportion of success (event) in a population.

Assume that p be the probability of success and $q = 1 - p$ probability of ~~fail~~ failure in a sample of size n. To estimate population proportion P. Draw all possible samples of size n from population and compute proportion in each sample say $p_1, p_2 \ldots p_n$.

Then mean μ_p and variance σ_p^2 of sampling distributions of proportions are given by.

$\mu_p = P$ and $\sigma_p^2 = \frac{PQ}{n}$ when n is large or population is infinite.

S.E of sample proportions $= \sigma_p = \sqrt{\frac{PQ}{n}}$

For a finite population (with out replacement) of size N, we have

$$\mu_p = P \text{ and } \sigma_p^2 = \frac{PQ}{n}\left(\frac{N-n}{N-1}\right).$$