CHAPTER - 6

SAMPLING DISTRIBUTIONS

6.1 INTRODUCTION

The outcome of a statistical experiment may be recorded either as a numerical value or as a descriptive representation.

When a pair of dice are tossed and the sum of the numbers on the faces is the outcome of interest, we record a numerical value. However, if the students of a certain school are given blood tests and the type of blood is of interest, then a descriptive representation might be most useful. A person's blood can be classified in 8 ways. It must be AB, A, B or O, with a plus or minus sign, depending on the presence or absence of the *Rh* antigen.

The statistician is primarily concerned with the analysis of numerical data.

For the classification of blood types, it may be convenient to use numbers from 1 to 8 to represent the blood types and then record the appropriate number for each student.

In any particular study, the number of possible observations may be small, large but finite, or infinite. For example, in the classification of blood types we can only have as many observations as there are students in the school. The project, therefore results in a finite number of observations.

On the other hand, if we could toss a pair of dice indefinitely and record the sum that occur, we would obtain an infinite set of values, each value representing the result of a single toss of a pair of dice.

In this chapter we focus on sampling from distributions or populations and study such important quantities as the **sample mean** and **sample variance.**

6.2 POPULATION AND SAMPLE

We begin this section by discussing the notions of population and sample.

The totality of observations with which we are concerned, whether this number be finite or infinite, constitutes what we call *population*. There was a time when the word *population* referred to observations obtained from statistical studies about people. Today, the statistician uses the term to refer to groups of people, animals, or all possible outcomes from some complicated biological or engineering system or numerical data. Thus the population does not imply living beings alone, e.g., we speak of the population of births, heights, weights, prices of vegetables and so on.

Definition : Population (or universe) is the aggregate or totality of statistical data forming a subject of investigation. For example,

(*i*) the population of the heights of Indians,

(*ii*) the population of Nationalised Banks in India, etc.

Definition : The number of observations in the population is defined to be the *size of the population. It may be finite or infinite. Size of population is denoted by N.*

Examples : If there are 600 students in the school that we classified according to blood type, we say that we have a population of size 600. The numbers on the cards in a deck, the heights of residents in a certain city, and the lengths of fish in a particular lake are all examples of population with *finite* size.

The observations obtained by measuring the atmospheric pressure every day from the past on into the future, or all measurements on the depth of a lake from any conceivable position, are examples of population whose sizes are infinite.

In the field of statistical inference the statistician is interested in arriving at conclusions concerning a population when it is impossible or impractical to observe the entire set of observations that make up the population.

For example, in attempting to determine the average length of life of a certain brand of light bulb, it would be impossible to test all such bulbs if we are to have any left to sell.

Therefore, we must depend on a subset of observations from the population to help us make inferences concerning the same population. This brings us to consider the notion of sampling.

Sampling : Most of the times, study of entire population may not be possible to carry out and hence a part alone is selected from the given population. A portion of the population which is examined with a view to determining the population characteristics is called a sample. That is, a part of the population selected for study is called a sample *i.e.*, a sample is a subset (selected portion) of population and the number of objects in the sample is called the size of the sample. Size of the sample is denoted by n.

The process of selection of a sample is called sampling. It is quite often used in our day-to-day practical life. For example, (*i*) to assess the quality of a bag of rice, sugar, wheat or any other commodity, we examine only a portion of it by taking a handful of it from the bag and then decide to purchase it or not. The portion selected from the bag is called a sample, while the whole quantity of rice, sugar or wheat in the bag is the population, (*ii*) to estimate the proportion of defective articles in a large consignment, only a portion (*i.e.*, a few of them) is selected and examined. The portion selected is a sample (*iii*) cars produced in India is the population and the Nano cars is the sample.

When information is collected in respect of every individual item, the enquiry is said to be done by *Complete Enumeration* or *Census*. For example, during the *Census of Population* (which is done every ten years in India), information in respect of each individual person residing in India is collected. This method gives information for each and every unit of the population with greater accuracy. But this method involves multiplicity of causes viz., administrative and financial implications, time factor, etc. In most cases of statistical inquiry, because of limitations of time and cost, only a portion (i.e., a sample) of the available source of information is examined, and the data collected from them. This process of partial enumeration is known as *Sample Survey*. The results are then generalised and made applicable to the whole field of inquiry. This is known as *Sampling*.

Thus in estimating the characteristics of the population, instead of enumerating entire population, only the individuals in the sample are examined. Then the sample characteristics are utilised to approximately estimate the population. The error involved in such approximation is known as *sampling error* and is inherent and unavoidable in any and every sampling scheme. However the advantage of sampling process lies in considerable gains in time, cost and handling of large data.

The population is thus a universal set for the sample. The statistical constants like mean, standard deviation, correlation coefficient, etc. obtained for the population are called parameters.

Similarly, constants for the sample drawn from the given population *i.e.*, mean ($\bar{x}$), standard deviation (S) etc., are called the statistic.

If our inferences from the sample to the population are to be valid, we must obtain samples that are representative of the population. All too often, we are tempted to choose a sample by selecting the most convenient members of the population. Such a procedure may lead to erroneous inferences concerning the population. Any sampling procedure that produces inferences that consistently over estimate (or) consistently under estimate some characteristics of the population is said to be biased. To eliminate any possibility of bias in the sampling procedure it is desirable to choose a random sample in the sense that the observations are made independently and at random.

6.3 DIFFERENT METHODS OF SAMPLING

Some important methods of sampling are discussed below.

I. Probability Sampling Methods

1. Random Sampling (or Probability Sampling) [JNTU (H) Dec. 2019 (R18)]

It is the process of drawing a sample from a population in such a way that each member of the population has an equal chance of being included in the sample. The sample obtained by the process of random sampling is called a random sample.

For example :

(*i*) A handy of cards dealt from a well - shuffled pack of cards is a random sample.

(*ii*) Selecting randomly 20 words from a dictionary is a random sample.

(*iii*) Choosing 10 patients from a hospital in order to test the efficacy of a certain newly - invented drug.

If each element of a population may be 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 without replacement.**

Note : If N is the size of a population and n is the sample size, then

(*i*) The number of samples with replacement $= N^n$

(*ii*) The numberof samples without replacement $= {}^{N}C_{n}$

2. Stratified Sampling (or Stratified Random Sampling)

This method is useful when the population is heterogeneous. In this type of sampling, the population is first sub-divided into several parts (or small groups) called *strata* according to some relevant characteristics so that each stratum is more or less homogeneous. Each stratum is called a sub-population. Then a small sample (called sub-sample) is selected from each stratum at random. All the sub-samples are combined together to form the stratified sample which represents the population properly. The process of obtaining and examining a stratified sample with a view to estimating the characteristic of the population is known as *Stratified Sampling.*

For example, let us select a stratified sample of 500 families from a city having 50,000 families, with a view to studying their economic condition. For this purpose, the city area is divided into a number of strata, according to economic condition of their inhabitants, as measured by annual income (say). Thus, localities mostly inhabited by people with more or less similar annual income may be included under one stratum. A few families are then chosen at random from each so that the sum total of all the families from all the strata is 500.

3. Systematic Sampling (or Quasi - Random Sampling)

As the name suggests this means forming the sample in some systematic manner by taking items at regular intervals. In this method, all the units of the population are arranged in some order. If the population size is finite, all the units of the population are arranged in some order. Then from the first k items, one unit is selected at random. This unit and every kth 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 systematic sampling lies in the fact that in the case of a random sample all the members have to be chosen randomly, whereas in the case of a systematic sample only the first member has to be chosen at random.

II. Non-Probability Sampling Methods.

4. Purposive Sampling (or Judgement Sampling) [JNTU (H) Dec. 2019 (R18)]

When the choice of the individual items of a sample entirely depends on the individual judgement of the investigator (or sampler), it is called a *Purposive or Judgement Sampling.*

In this method, the members constituting the sample are chosen not according to some definite scientific procedure, but according to convenience and personal choice of the individual, who selects the sample. Two or more such independent purposive samples may give widely different estimates of the same population. In this type, the investigator must have a good deal of experience and a thorough knowledge of the population. Purposive selection is always subject to some kind of bias. This method is suitable when the sample is small.

For example, if a sample of 20 students is to be selected from a class of 100 to analyse the extra-curricular activities of the students, the investigator would select 20 students who, in his judgement, would represent the class.

5. Sequential Sampling

It consists of a sequence of sample drawn one after another from the population depending on the results of previous samples. If the result of the first sample leads to a decision which is not acceptable, the lot from which the sample was drawn is rejected. But if the result of the first sample is acceptable, no new sample is drawn. But if the first sample leads to no clear decision, a second sample is drawn and, as before, if required, a third sample is drawn to arrive at a final decision to accept or reject the lot. It is widely used in Statistical Quality Control in factories engaged in mass production and other areas.

6.4 CLASSIFICATION OF SAMPLES

Samples are classified in two ways.

1. Large sample : If the size of the sample, $n > 30$, the sample is said to be large sample.

2. **Small sample :** If the size of the sample $n \leq 30$, the sample is said to be small sample (or) exact sample (or) simple sample. **[JNTU (H) Dec. 2019 (R18)]**

In sampling *with replacement*, each member of the population may be chosen more than once, since the member is replaced in the population.

Thus sampling from finite population *with replacement* can be considered theoretically as sampling from *infinite population*. Whereas, in sampling *without replacement*, an element of the population cannot be chosen more than once, as it is not replaced. Therefore the sampling from finite population *without replacement* can be considered theoretically as sampling from finite population only.

Definition : Let $X_1, X_2, ..., X_n$ be n independent random variables each having the same probability distribution $f(x)$. We then define $X_1, X_2, ..., X_n$ to be a random sample of size n from the population $f(x)$ and write its joint probability distribution as

$$f(x_1, x_2, ..., x_n) = f(x_1).\, f(x_2) \,........, f(x_n)$$

Our main purpose in selecting random sample is to collect information about the unknown population parameters.

Statistic : It is a real valued function of the random sample. So statistic is a function of one or more random variables not involving any unknown parameter. Therefore statistic is a random variable and has a probability or frequency distribution.

6.5 PARAMETERS AND STATISTICS

We have already discussed about measures of central tendency or averages and measures of dispersion such as mean, median, mode, standard deviation, variance, etc. These measures can be computed both from population data and sample data.

Parameter is a statistical measure based on all the units (or observations) of a population. Statistic (or sample statistic) is a statistical measure based only on all the units selected in a sample. For example, the statistical constants of the population namely mean μ, variance σ^2 are usually referred to as *parameters*, statistical measures computed from the sample observations alone *e.g.*, sample mean ($\bar{x}$), sample variance (s^2) etc., are usually referred to as statistics.

In other words, the mean, median, mode, standard deviation, variance measures of the population are called **parameters** and the measures obtained from the sample of the population are called **statistics**.

Thus *parameters* refer to population while *statistics* refer to sample.

Parameter : Any statistical measure computed from population data is known as **parameter**.

Statistic : Any statistical measure computed from sample data is known as **statistic**.

Note : As the units selected in two or more samples drawn from a population are not the same, the value of a statistic varies from sample to sample, but the parameter always remains constant (since all the units in a population remain the same). This variation in the value of a statistic is called *sampling fluctuation*. A parameter has no sampling fluctuation. Usually, statistic is used to estimate the value of an unknown parameter obtained from the sample, is a function of the sample values only.

1. The Sample Mean :

Definition : If $X_1, X_2, ..., X_n$ represent a random sample of size n, then the sample mean is defined by the statistic

$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

Note that the statistic $\bar{x}$ has the value $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$ when x assumes the values $x_1, x_2.....$ For instance, the term sample mean is applied to both the statistic $\overline{X}$ and its computed value $\bar{x}$.

2. The Sample Variance :

Definition : If $X_1, X_2, ..., X_n$ represent a random sample of size n, then the sample variance is defined by the statistic

$$s^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1} = \frac{n\sum X_i^2 - (\sum X_i)^2}{n(n-1)} \quad \text{(measure of variability of data about the mean)}$$

Note that s^2 is essentially defined to be the average of the squares of the deviations of the observations from their mean with the change that the sum of the squared deviations is divided by $n - 1$ and not n.

Ex : A comparison of coffee prices at 4 randomly selected grocery stores in a city showed increases from the previous month of 12, 15, 17 and 20 rupees for 1 kg. Find the variance of this sample of price increases.

Solution : Calculating the sample mean, we get

$$\bar{x} = \frac{12+15+17+20}{4} = \text{Rs. } 16/-$$

$$\therefore \quad s^2 = \sum_{i=1}^{4} \frac{(x_i - 16)^2}{4-1}$$

$$= \frac{(12-16)^2 + (15-16)^2 + (17-16)^2 + (20-16)^2}{3}$$

$$= \frac{(-4)^2 + (-1)^2 + (1)^2 + (4)^2}{3} = \frac{34}{3}$$

3. The Sample Standard Deviation :

The sample S. D., denoted by s, is the positive square root of the sample variance.

Note : In the above example S.D., $s = \sqrt{34/3}$.

6.6 SAMPLING DISTRIBUTION

Sampling theory is the study of relationships between a population and samples drawn from the population, and it is applicable to random samples only. We need to obtain maximum information about the population with the help of samples, i.e., to determine the true value of the population parameters (population mean, population S. D., population proportion, etc.) by using sample statistics like sample mean, sample S. D., sample proportion, etc., and to find the limits of accuracy of estimates based on samples. It helps us to determine whether the differences between two samples are actually due to chance variation or whether they are really significant. Sampling theory also useful in testing of hypothesis and significance which is important in the theory of decisions.

6.7 SAMPLING DISTRIBUTION OF A STATISTIC

Sampling distribution of a statistic is the frequency distribution which is formed with various values of a statistic computed from different samples of the same size drawn from the same population. We can draw a large number of samples of same size from a population of fixed size, each sample containing different population members. Any statistic (statistical measure of sample) like mean, median, variance, etc. may be computed for each of these samples. As a result a series of various values of that statistic may be obtained. These various values can be arranged into a frequency distribution table, which is known as the sampling distribution of the statistic.

Sampling may be done with replacement or without replacement. Sampling with replacement means that the same unit of the population may be included in each sample more than once. Sampling without replacement means that the same unit of population may not be included in each sample more than once.

Let us consider a finite population of size N and let us draw all possible random samples each of the same size n. Then we get $^{N}C_n = \dfrac{N!}{n!(N-n)!} = k$ (say) samples. Compute a statistic t (such as sample mean, S.D., etc) for each of these samples. The value of a statistic t may vary from sample to sample. Let $t_1, t_2, \ldots\ldots, t_k$ be the values of statistic k for the k samples. Each of these values occur with a definite probability. Thus we can construct a table showing the set of values $t_1, t_2, \ldots, t_k$ of t with their respective probabilities. This probability distribution of t is known as the Sampling Distribution of t. Thus sampling distribution describes how a statistic t will vary from one sample to the other of the same size. Although all the k samples are drawn from the given population, the members included in different samples are different.

If N is large, then the number k of all possible samples is also large (*i.e.*, sampling without replacement). In such a case, the values $t_1, t_2, \ldots\ldots, t_k$ of t may be arranged in the form of a relative frequency distribution and the limiting form of this relative frequency distribution when $k \to \infty$ is called the sampling distribution of statistic t. A statistic (*i.e.*, sample mean, sample S.D, etc.) has always a sampling distribution, but a parameter (*i.e.* population mean, population S.D.) has no sampling distribution.

If the statistic t is mean, then the corresponding distribution of the statistic is known as sampling distribution of means.

The sample mean can be regarded as a random variable $\bar{x}$ and each sample mean then constitute as the observed value of this new random variable $\bar{x}$. Let these values be: $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_n$. These mean values $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_n$ can be used to form a frequency distribution. Then this frequency distribution of the statistic $\bar{x}$ is known as the sampling distribution of the sample mean. Similarly the sampling distribution of standard deviation or variance may be constructed with the various values of standard deviation or variance respectively.

The main characteristic of the Sampling distribution of a statistic is that it approaches normal distribution even when the population distribution is not normal provided the sample size is sufficiently large (greater than 30). Another important feature of the sampling distribution of statistic is that the mean and the standard deviation of the sampling distribution of sample mean bear a definite relation to the corresponding parameters i.e., mean and standard deviation of parent population. These characteristics of the sampling distribution help us:

(*i*) To estimate the unknown population parameter from the known statistic

(*ii*) To set the confidence limits of the parameter within which the parameter values are expected to lie.

(*iii*) To test a hypothesis and to draw a statistical inference from it.

6.8 CENTRAL LIMIT THEOREM

[JNTU (H) May 2017]

If $\bar{x}$ be the mean of a random sample of size n drawn from a population having mean μ and S.D. σ, then the sampling distribution of the sample mean $\bar{x}$ is approximately a normal distribution with mean μ and S.D. = S.E. of $\bar{x} = \dfrac{\sigma}{\sqrt{n}}$ provided the sample size n is large $(n \geq 30)$. This is established by central limit theorem stated below (without proof).

Theorem : If $\bar{x}$ be the mean of a sample size n drawn from a population with mean μ and S. D. σ then the standardized sample mean

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

is a random variable whose distribution function approaches that of the standard normal distribution $N(z; 0, 1)$ as $n \to \infty$.

6.9 STANDARD ERROR (S.E.) OF A STATISTIC

The standard error of a statistic t (*i.e.*, S. E. of sample mean, or sample S.D.) is the standard deviation of the sampling distribution of the statistic. Thus S.E. of sample mean $\bar{x}$ is the S.D. of the sampling distribution of sample mean. It is used for assessing the difference between the expected value and observed value. It plays an important role in large sample theory and forms the basis in tests of hypothesis or tests of significance. It gives an idea about the reliability and precision of a sample. S. E. enables us to determine the confidence limits within which the parameters are expected to lie. For example, the probable limits for population proportion P are given by $p \pm 3 . \sqrt{pq/n}$.

The standard errors (S.E.) of some of the well-known statistics are given below (without proof), where μ the population mean, σ^2 the population variance, P the population proportion, n is the sample size, $\bar{x}$ the sample mean, s^2 the sample variance and p the sample proportion.

Note : S.E. of a statistic may be reduced by increasing sample size n, but this results in corresponding increase in cost, time and labour, etc.

Formulae for S.E.

1. S.E. of sample mean $\bar{x} = \frac{\sigma}{\sqrt{n}}$. It is written as S. E. $(\bar{x}) = \frac{\sigma}{\sqrt{n}}$ **[JNTU (H) Nov. 2015)]**

2. S.E. of sample proportion $p = \sqrt{\frac{PQ}{n}}$, where $Q = 1 - P$

3. S.E. of sample S. D $(s) = \frac{\sigma}{\sqrt{2n}}$

4. S.E. of the difference of two sample means $\bar{x}_1$ and $\bar{x}_2$ **[JNTU (H) Nov. 2015)]**

   i.e., S.E. of $(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ where $\bar{x}_1$ and $\bar{x}_2$ are the means of two random samples of sizes n_1 and n_2 drawn from two populations with S.D. σ_1 and σ_2 respectively.

5. S.E. of $(p_1 - p_2) = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$, where p_1 and p_2 are the proportions of two random samples of sizes n_1 and n_2 drawn from two populations with proportions P_1 and P_2 respectively.

6. S.E. of $(s_1 - s_2) = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$

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

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

(*ii*) S.E. of sample proportion $= \sqrt{\frac{PQ}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

For an infinite population when the sample is drawn without replacement, Formulae 1 and 2 remain the same.

Proposition (Statement without proof) : The mean and S. E. of sample mean $\bar{x}$ (*i.e.*, mean and S.D. of the sampling distribution of $\bar{x}$) when samples of the same size n are drawn from a population having mean μ and S.D. σ are given by Mean of $\bar{x} = E(\bar{x}) = \mu$ and S.E. of $\bar{x} = \frac{\sigma}{\sqrt{n}}$.

6.10 SAMPLING DISTRIBUTION OF MEANS (σ KNOWN)

The probability distribution of $\bar{X}$ is called the sampling distribution of means.

The sampling distribution of a statistic depends on the size of the population, the size of the samples, and the method of choosing the samples.

Let $X_1, X_2, X_3, \ldots, X_n$ be the n random samples drawn from a population of size N with mean μ and variance σ^2 and $\bar{X}$ is the mean of samples.

Infinite Population : Suppose the samples are drawn from an infinite population *i.e.*, $N \to \infty$ (or) sampling is done with replacement, then

(*i*) The mean of the sampling distribution of means, $\mu_{\bar{X}} = \mu$ *i.e.*, $E(\bar{X}) = \mu$

Proof: We have $\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n}$

Since $\bar{X}$ is a linear combination of the random variables $X_1, X_2, X_3, \ldots, X_n$, therefore it is also a random variable.

$$\therefore \text{ Mean of } \bar{X} = E(\bar{X}) = \frac{1}{n}E(X_1 + X_2 + X_3 + \ldots + X_n)$$

$$= \frac{1}{n}[E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_n)]$$

$$= \frac{1}{n}[\mu + \mu + \mu + \ldots + \mu] = \frac{n\mu}{n} = \mu.$$

(*ii*) Variance, $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ *i.e.*, $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$

$\therefore$ S.D of mean, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Proof: If the sample is drawn with replacement, then $X_1, X_2, X_3, \ldots, X_n$ are independent random variables. Thus we have $\text{Var}(\bar{X}) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$.

The sampling distribution of $\bar{x}$ will be approximately *normal* with mean μ and variance $\frac{\sigma^2}{n}$ provided that the sample size is large.

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

Note : 1. 95% confidence interval for μ is given by $|z| \leq 1.96$ *i.e.*, $\left|\frac{\bar{x} - \mu}{\sigma / \sqrt{n}}\right| \leq 1.96$

$$\Rightarrow \bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

Thus $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$ are known as 95% confidence limits for μ

2. Similarly, 99% confidence limits for μ are $\bar{x} \pm 2.58\frac{\sigma}{\sqrt{n}}$ and 98 % confidence limits are $\bar{x} \pm 2.33\frac{\sigma}{\sqrt{n}}$.

Finite Population :

Consider a finite population of size N with mean μ and standard deviation σ. Draw all possible samples of size n without replacement from this population. Then

(*i*) The mean of the sampling distribution of means (for $N > n$) is $\mu_{\bar{X}} = \mu$.

i.e., $E(\bar{X}) = \mu$

(*ii*) The variance of the sampling distribution of means (for $N > n$) is

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

i.e., $Var(\bar{X}) = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$ (if the sample is drawn without replacement)

and S.D is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

Here the factor $\left(\frac{N-n}{N-1}\right)$ is known as the finite population **correction factor**. We note that this term tends to become closer and closer to unity as population size becomes larger and larger.

Normal Population (Small Sample) :

Sampling distribution of $\bar{X}$ is normally distributed even for small samples of size $n < 30$ provided sampling is from normal population.

Non - Normal Population (Large Sample) :

Consider a population with unknown (non-normal) distribution. Let the population mean μ and population variance σ be both finite. Let the population be finite or infinite. In case the population is finite assume that the population size N is atleast twice the sample size n. Draw all possible samples of size n. Then the sampling distribution of $\bar{X}$ is approximately normally distributed with mean $\mu_{\bar{x}} = \mu$ and variance

$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ provided the sample size is large (*i.e.*, $n > 30$).

6.11 SAMPLING DISTRIBUTION OF PROPORTIONS

Let p be the probability of occurrence of an event (called its success) and $q = 1 - p$ is the probability of non-occurrence (called its failure). Draw all possible samples of size n from an infinite population. Compute the proportion P of success for each of these samples. Then the mean μ_P and variance σ_P^2 of the sampling distribution of proportions are given by

$$\mu_P = p \text{ and } \sigma_P^2 = \frac{pq}{n} = \frac{p(1-p)}{n}$$

while population is binomially distributed, the sampling distribution of proportion is normally distributed whenever n is large. For finite population (with replacement) of size N, we have

$$\mu_P = p \text{ and } \sigma_P^2 = \frac{pq}{n}\left(\frac{N-n}{N-1}\right).$$

6.12 SAMPLING DISTRIBUTION OF DIFFERENCES AND SUMS

Let μ_{S_1} and σ_{S_1} be the mean and standard deviation of sampling distribution of statistic S_1 obtained by computing S_1 for all possible samples of size n_1 drawn from population A. Also let μ_{S_2} and σ_{S_2} be the mean and standard deviation of sampling distribution of statistic S_2 obtained by computing S_2 for all possible samples of size n_2 drawn from another different population B.

Now compute the statistic $S_1 - S_2$, the difference of the statistic from all the possible combinations of these samples from the two populations A and B.

Then the mean $\mu_{S_1 - S_2}$ and the standard deviation $\sigma_{S_1 - S_2}$ of the sampling distribution of differences are given by

$$\mu_{S_1 - S_2} = \mu_{S_1} - \mu_{S_2} \text{ and } \sigma_{S_1 - S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}$$

assuming that the samples are independent.

Sampling distribution of sum of statistics has mean $\mu_{S_1 + S_2}$ and standard deviation $\sigma_{S_1 + S_2}$ given by

$$\mu_{S_1 + S_2} = \mu_{S_1} + \mu_{S_2} \text{ and } \sigma_{S_1 + S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}$$

For example, for infinite population the sampling distribution of sums of means has mean $\mu_{\bar{x}_1 + \bar{x}_2}$ and $\sigma_{\bar{x}_1 + \bar{x}_2}$ given by $\mu_{\bar{x}_1 + \bar{x}_2} = \mu_{\bar{x}_1} + \mu_{\bar{x}_2} = \mu_1 + \mu_2$ and

$$\sigma_{\bar{x}_1 + \bar{x}_2} = \sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

For sampling distribution of differences of proportions, we have

$$\mu_{P_1 - P_2} = \mu_{P_1} - \mu_{P_2} = p_1 - p_2$$

and $\sigma_{P_1 - P_2} = \sqrt{\sigma_{P_1}^2 + \sigma_{P_2}^2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

SOLVED EXAMPLES

Example 1 : What is the value of correction factor if n = 5 and N = 200.

Solution : Given N = Size of the finite population = 200

n = Size of the sample = 5

$\therefore$ Correction factor $= \dfrac{N-n}{N-1} = \dfrac{200-5}{200-1} = \dfrac{195}{199} = 0.98$

Example 2 : Find the value of the finite population correction factor for n = 10 and N = 1000. **[JNTU 1999, 2000S, (A) Dec. 2009, Nov. 2010, Apr. 2012 (Set No. 3), (H) Sept. 2017]**

Solution : Given N = Size of the finite population = 1000

n = Size of the sample = 10

$\therefore$ Correction factor $= \dfrac{N-n}{N-1} = \dfrac{1000-10}{1000-1} = \dfrac{990}{999} = 0.991$

Example 3 : How many different samples of size two can be chosen, from a finite population of size 25.

Solution : We can take ${}^{N}C_n$ samples of size n from the population of size N.

Here $N = 25$, $n = 2$

$\therefore$ We can take ${}^{25}C_2 = 300$ samples of size 2 from finite population of size 25.

Example 4 : In a random sample of 100 packages shipped by air freight 13 had some damage. Find the standard error of proportions. **[JNTU (H) III yr Nov. 2015]**

Solution : Proportion of damage packages, $P = \dfrac{13}{100} = 0.13$

$\therefore$ $Q = 1 - P = 1 - 0.13 = 0.87$. Also $n = 1000$

$\therefore$ Standard error of proportions = S. E. $(p) = \sqrt{\dfrac{PQ}{n}} = \sqrt{\dfrac{0.13 \times 0.87}{1000}} = 0.0106$

Example 5 : A population consists of five numbers 2, 3, 6, 8 and 11. Consider all possible samples of size two which can be drawn with replacement from this population. Find

(*a*) The mean of the population.

(*b*) The standard deviation of the population.

(*c*) The mean of the sampling distribution of means and

(*d*) The standard deviation of the sampling distribution of means (*i.e.*, the standard error of means).

[JNTU Nov 2004, April 2005 (Sets 3, 4), (A) Dec. 2009 (Set No. 4)]

Solution :

(*a*) Mean of the population is given by

$$\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$$

(*b*) Variance of the population (σ^2) is given by

$$\sigma^2 = \sum \frac{(x_i - \bar{x})^2}{n}$$

$$= \frac{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}{5}$$

$$= \frac{16+9+0+4+25}{5} = 10.8$$

$\therefore$ $\sigma = \sqrt{10.8}$ *i.e.* $\sigma = 3.29$

(*c*) Sampling with replacement (Infinite population) :

The total no. of samples with replacement is

$N^n = 5^2 = 25$ samples of size 2

Here N = population size and n = sample size listing all possible samples of size 2 from population 2, 3, 6, 8, 11 with replacement we get 25 samples

$$\begin{Bmatrix} (2,2) & (2,3) & (2,6) & (2,8) & (2,11) \\ (3,2) & (3,3) & (3,6) & (3,8) & (3,11) \\ (6,2) & (6,3) & (6,6) & (6,8) & (6,11) \\ (8,2) & (8,3) & (8,6) & (8,8) & (8,11) \\ (11,2) & (11,3) & (11,6) & (11,8) & (11,11) \end{Bmatrix}$$

Now compute the arithmetic mean for each of these 25 samples. The set of 25 means $\bar{x}$ of these 25 samples, gives rise to the distribution of means of the samples known as sampling distribution of means.

The samples means are

$$\begin{Bmatrix} 2 & 2.5 & 4 & 5 & 6.5 \\ 2.5 & 3 & 4.5 & 5.5 & 7 \\ 4 & 4.5 & 6 & 7.0 & 8.5 \\ 5 & 5.5 & 7 & 8 & 9.5 \\ 6.5 & 7 & 8.5 & 9.5 & 11 \end{Bmatrix} \quad \ldots\ldots (1)$$

and the mean of sampling distribution of means is the mean of these 25 means.

$$\mu_{\bar{x}} = \frac{\text{Sum of all sample means in (1)}}{25} = \frac{150}{25} = 6$$

Illustrating that $\mu_{\bar{x}} = \mu$.

(*d*) The variance $\sigma_{\bar{x}}^2$ of the sampling distribution of means is obtained by subtracting the mean 6 from each number in (1) and squaring the result, adding all 25 members thus obtained, and dividing by 25.

$$\sigma_{\bar{x}}^2 = \frac{(2-6)^2 + \ldots + (11-6)^2}{25} = \frac{135}{25} = 5.40 \quad \text{and thus} \quad \sigma_{\bar{x}} = \sqrt{5.40} = 2.32.$$

Clearly, for finite population involving sampling with replacement (or infinite population)

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \text{ or } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.29}{\sqrt{2}} = 2.32$$

Example 6 : Solve the above example (*i.e.* Ex. 4) without replacement.

Solution : (*a*) $\mu = 6$ (*b*) $\sigma = 3.29$

(*c*) **Sampling without replacement (finite population) :**

The total no.of samples without replacement is ${}^{N}C_n = {}^{5}C_2 = 10$ samples of size 2

The 10 samples are

$$\begin{Bmatrix} (2,3) & (2,6) & (2,8) & (2,11) \\ (3,6) & (3,8) & (3,11) & \\ (6,8) & (6,11) & & \\ (8,11) & & & \end{Bmatrix}$$

The selection (2, 3) is considered same as (3, 2)

The corresponding sample means are

$$\begin{Bmatrix} 2.5 & 4 & 5 & 6.5 \\ 4.5 & 5.5 & 7 & \\ 7 & 8.5 & & \\ 9.5 & & & \end{Bmatrix}$$

The mean of the sampling distribution of means is

$$\mu_{\bar{x}} = \frac{(2.5+4+5+6.5+4.5+5.5+7+7+8.5+9.5)}{10} = 6$$

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

(*d*) The variance of sampling distributions of means is

$$\sigma_{\bar{x}}^2 = \frac{(2.5-6)^2+(4-6)^2+........+(9.5-6)^2}{10} = 4.05$$

and $\sigma_{\bar{x}} = 2.01$

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

$$= \frac{10.8}{2}\left(\frac{5-2}{5-1}\right) = 4.05$$

for sampling without replacement.

Example 7 : A population consists of 5,10,14,18,13,24. Consider all possible samples of size two which can be drawn without replacement from the population. Find

(*a*) The mean of the population

(*b*) The standard deviation of the population

(*c*) The mean of the sampling distribution of means

(*d*) The standard deviation of sampling distribution of means.

[JNTU (H) Nov. 2010 (Set No.3)]

Solution : (*a*) The mean of the population μ is given by

$$\mu = \frac{\Sigma x}{n} = \frac{5+10+14+18+13+24}{6} = \frac{84}{6} = 14$$

(*b*) Variance of the population σ^2 is given by

$$\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2}{n}$$

$$= \frac{1}{6}[(5-14)^2+(10-14)^2+(14-14)^2+(18-14)^2+(13-14)^2+(24-14)^2]$$

$$= \frac{1}{6}[81+16+0+16+1+100]$$

$$= \frac{214}{6} = 35.67$$

(*c*) All possible samples of size two without replacement (the no. of samples $= {}^{6}C_2 = 15$) and their means are shown in the following table.

Sample No.	Sample values	Total of Sample values	Sample mean
1	5,10	15	7.5
2	5,14	19	9.5
3	5,18	23	11.5
4	5,13	18	9
5	5,24	29	14.5
6	10,14	24	12
7	10,18	28	14
8	10,13	23	11.5
9	10,24	34	17
10	14,18	32	16
11	14,13	27	13.5
12	14,24	38	19
13	18,13	31	15.5
14	18,24	42	21
15	13,24	37	18.5
Total			210

$\therefore$ Mean of sample means $= \frac{210}{15} = 14$

i.e., The Mean of the sampling distribution of means is $\mu_{\bar{x}} = 14$

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

(*d*) The variance of sampling dstribution of means is

$$\sigma_{\bar{x}}^2 = \frac{1}{15}[(7.5-14)^2+(9.5-14)^2+(11.5-14)^2+(9-14)^2 + (14.5-14)^2+....+(21-14)^2+(18.5-14)^2]$$

$$= \frac{1}{15}[42.25 + 20.25 + 6.25 + 25 + 0.25 + 4 + 0 + 6.25 + 9 + 4 + 0.25 + 25 + 2.25 + 49 + 20.25]$$

$$= \frac{214}{15} = 14.2666$$

$\therefore$ Standard deviation of sampling distribution of means is $\sigma_{\bar{x}} = \sqrt{14.2666} = 3.78$

Example 8 : Let $u_1 = (3, 7, 8)$, $u_2 = (2, 4)$. Find

(*a*) μ_{u_1} (*b*) μ_{u_2}

(*c*) Mean of the sampling distribution of the difference of means $\mu_{u_1-u_2}$

(*d*) σ_{u_1} (*e*) μ_{u_2}

(*f*) the standard deviation of the sampling distribution of the diffrences of means $(\sigma_{u_1-u_2})$.

[JNTU Nov. 2008 (Set No. 3)]

Solution : Given $u_1 = \{3, 7, 8\}$ and $u_2 = \{2, 4\}$

$u_1 - u_2 = \{1, 5, 6, 4, 3, -1\}$

Now,

(*a*) $\mu_{u_1} = \frac{3+7+8}{3} = 6$

(*b*) $\mu_{u_2} = \frac{2+4}{2} = 3$

(*c*) $\mu_{u_1-u_2} = \frac{1+5+6+4+3-1}{6} = 3$

(*d*) $\sigma_{u_1} = \sqrt{\frac{(6-3)^2+(6-7)^2+(6-8)^2}{3}} = \sqrt{\frac{14}{3}}$.

(*e*) $\sigma_{u_2} = \sqrt{\frac{(2-3)^2+(3-4)^2}{2}} = 1$

(*f*) $\sigma_{u_1-u_2} = \sqrt{\frac{(1-3)^2+(5-3)^2+(6-3)^2+(4-3)^2+(3-3)^2+(-1-3)^2}{6}}$

$$= \sqrt{\frac{34}{6}} = \sqrt{\frac{17}{3}}$$

Example 9 : Find the mean and Standard deviation of sampling distribution of variances for the population 2,3,4,5 by drawing samples of size two (*a*) with replacement (*b*) without replacement.

[JNTU (A) 2009, Nov. 2011 (Set No. 2,4)]

Solution : Population size, N = 4, sample size, $n = 2$.

(*a*) Number of samples of size two with replacement is $N^n = 4^2 = 16$.

They are

(2,2), (2,3), (2,4), (2,5), (3,2), (3,3), (3,4), (3,5), (4,2), (4,3), (4,4), (4,5), (5,2), (5,3), (5,4), (5,5).

We now compute the statistic variance for each of these 16 samples.

Variance for the sample (2,2) with mean 2 is $\frac{1}{2}[(2-2)^2+(2-2)^2] = 0$

Similarly, the variance for sample (2,3) with mean 2.5 is

$$\frac{1}{2}[(2-2.5)^2+(3-2.5)^2] = 0.25.$$

Thus the variances of these 16 samples are

$$\begin{Bmatrix} 0 & 0.25 & 1 & 2.25 \\ 0.25 & 0 & 0.25 & 1 \\ 1 & 0.25 & 0 & 0.25 \\ 2.25 & 1 & 0.25 & 0 \end{Bmatrix}$$

Thus the sampling distribution of variances (with replacement) is

S^2	0	0.25	1	2.25
Frequency	4	6	4	2

$\therefore$ Mean of S.D. of variances $= \frac{4(0)+6(0.25)+1(4)+2(2.25)}{16} = \frac{10}{16} = 0.625$

Variance of S.D. of variances

$$= \frac{1}{16}[4(0-0.625)^2+6(0.25-0.625)^2+4(1-0.625)^2+2(2.25-0.625)^2]$$

$$= \frac{8.25}{16} = 0.5156$$

(*b*) This is left as an exercise to the reader.

Example 10 : A population consists of six numbers 4,8,12,16,20,24. Consider all possible samples of size two that can be drawn without replacement from this popuation. Find

a) The population mean

b) The population standard deviation

c) The mean of the sampling distribution of means

d) The standard deviation of the sampling distribution of means

[JNTU 2000, (H) Nov. 2009 (Set No. 1)]

Solution : (*a*) The mean of the population,

$$\mu = \frac{\Sigma x_i}{N} = \frac{4+8+12+16+20+24}{6} = \frac{84}{6} = 14$$

(*b*) The variance of the population, $\sigma^2 = \frac{1}{N}\sum (x_i - \bar{x})^2$

$$= \frac{1}{6}[(4-14)^2 + (8-14)^2 + (12-14)^2 + (16-14)^2 + (20-14)^2 + (24-14)^2]$$

$$= \frac{1}{6}[100+36+4+4+36+100] \quad = \frac{280}{6} = 46.67.$$

∴ The population standard deviation, $\sigma = \sqrt{46.67} = 6.83$

The number of samples that can be drawn from the population of size 6 without replacement is ${}^6C_2$ *i.e.*, $\frac{6!}{2!4!}$ or 15.

The (15 samples are) sampling distribution is

$$\left\{\begin{matrix} (4,8), (4,12), (4,16), (4,20), (4,24), (8,12), (8,16), (8,20), \\ (8,24), (12,16), (12,20), (12,24), (16,20), (16,24), (20,24). \end{matrix}\right\}$$

The means are

$$\left\{\begin{matrix} 6, 8, 10, 12, 14, \\ 10, 12, 14, 16, \\ 14, 16, 18, \\ 18, 20 \\ 22 \end{matrix}\right\}$$

(*c*) The mean of the sampling distribution of means is

$$\mu_{\bar{x}} = \frac{1}{15}[6+8+10+12+14+10+12+14+16+14+16+18+18+20+22]$$

$$= \frac{210}{15} = 14$$

(*d*) The variance of the sampling distribution of means is

$$\sigma_{\bar{x}^2} = \frac{1}{N}\sum (x_i - \bar{x})^2$$

$$= \frac{1}{15}[(6-14)^2 + (8-14)^2 + (10-14)^2 + (12-14)^2 + (14-14)^2 + (10-14)^2 +$$

$$(12-14)^2 + (14-14)^2 + (16-14)^2 + (14-14)^2 + (16-14)^2 + (18-14)^2 +$$

$$(18-14)^2 + (20-14)^2 + (22-14)^2]$$

$$= \frac{1}{15}[64+36+16+4+0+16+4+0+4+0+4+16+16+36+64]$$

$$= \frac{280}{15} = 18.67$$

Hence the standard deviation of the sampling distribution of means is

$$\sigma_{\bar{x}} = \sqrt{18.67} = 4.32$$

Note : $\mu = \mu_{\bar{x}}$ and $\sigma^2 \neq \sigma_{\bar{x}}^2$

Example 11 : Samples of size 2 are taken from the population 1,2,3,4,5,6 (*i*) with replacement and (*ii*) without replacement. Find

(*a*) The mean of the population

(*b*) Standard deviation of population

(*c*) The mean of the sampling distribution of means

(*d*) The standard deviation of the sampling distribution of means.

Verify that means of sampling distribution is equal to the mean of population and standard deviations of the means of sampling distribution are not equal to the standard deviation of the population. **[JNTU (H) Nov. 2009, (A) Nov. 2011 (Set No. 3)]**

Solution :

(*i*) (*a*) The mean of the population, $\mu = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$

(*b*) The variance of the population is $\sigma^2 = \frac{1}{n}\sum (x_i - \mu)^2$

$$= \frac{1}{6}[(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2]$$

$$= \frac{1}{6}[6.25+2.25+0.25+0.25+2.25+6.25]$$

$$= \frac{17.50}{6} = 2.917$$

∴ The standard deviation of the population is, $\sigma = \sqrt{2.917} = 1.71$.

(*c*) Number of samples of size two with replacement is $N^n = 6^2 = 36$.

They are

$$\left\{\begin{matrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{matrix}\right\}$$

∴ The number of samples, $n = 36$

Their corresponding means are

$$\begin{Bmatrix} 1, & 1.5, & 2, & 2.5, & 3, & 3.5 \\ 1.5, & 2, & 2.5, & 3, & 3.5, & 4 \\ 2, & 2.5, & 3, & 3.5, & 4, & 4.5 \\ 2.5, & 3, & 3.5, & 4, & 4.5, & 5 \\ 3, & 3.5, & 4, & 4.5, & 5, & 5.5 \\ 3.5, & 4, & 4.5, & 5, & 5.5, & 6 \end{Bmatrix}$$

The mean of the sampling distribution of means is

$$\mu_x = \frac{1}{36}[1+1.5+2+2.5+3+3.5+2.5+\ldots+6]$$

$$= \frac{126}{36} = 3.5.$$

(*d*) The standard deviation of the sampling distribution of means is

$$\sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{36}(x_i - \mu_{\bar{x}})^2}{n}$$

$$= \frac{1}{36}[(1-3.5)^2 + (1.5-3.5)^2 + (2-3.5)^2 + (2.5-3.5)^2 + (3-3.5)^2$$
$$+(3.5-3.5)^2 + (1.5-3.5)^2 + \ldots + (6-3.5)^2]$$

$$= \frac{52.5}{36} = 1.46$$

$$\therefore \sigma_{\bar{x}} = \sqrt{1.46} = 1.2083$$

It is clear that the mean of the sampling distribution of means is equal to the mean of the population and standard deviation of the means of sampling distribution is not equal to the standard deviation of the population.

(*ii*) This is left as an exercise to the reader.

Example 12 : Samples of size 2 are taken from the population 3,6,9,15,27 with replacement. Find
a) The mean of the population
b) The standard deviation of the population
c) Mean of the sampling distribution of means
d) The standard deviation of the sampling distribution of means.

[JNTU (H) Nov. 2009 (Set No.4)]

<u>**Solution**</u> : (*a*) Mean of the population, $\mu = \frac{3+6+9+15+27}{5} = \frac{60}{5} = 12$

(*b*) Standard deviation of the population, $\sigma = \sqrt{\frac{1}{n}\sum_i (x_i - \mu)^2}$

$$= \sqrt{\frac{1}{5}[(3-12)^2 + (6-12)^2 + (9-12)^2 + (15-12)^2 + (27-12)^2]}$$

$$= \sqrt{\frac{1}{5}(81+36+9+9+225)} = \sqrt{\frac{360}{5}} = 8.4853$$

(*c*) The sampling distribution with replacement is

$$\begin{Bmatrix} (3,3) & (3,6) & (3,9) & (3,15) & (3,27) \\ (6,3) & (6,6) & (6,9) & (6,15) & (6,27) \\ (9,3) & (9,6) & (9,9) & (9,15) & (9,27) \\ (15,3) & (15,6) & (15,9) & (15,15) & (15,27) \\ (27,3) & (27,6) & (27,9) & (27,15) & (27,27) \end{Bmatrix}$$

The means are

$$\begin{Bmatrix} 3 & 4.5 & 6 & 9 & 15 \\ 4.5 & 6 & 7.5 & 10.5 & 16.5 \\ 6 & 7.5 & 9 & 12 & 18 \\ 9 & 10.5 & 12 & 15 & 21 \\ 15 & 16.5 & 18 & 21 & 27 \end{Bmatrix}$$

The mean of the sampling distribution of means,

$$\bar{x} = \frac{3+4.5+6+9+15+4.5+\ldots+18+21+27}{25} = \frac{300}{25} = 12.$$

(*d*) The standard deviation of the sampling distribution of means $= \sqrt{\frac{\sum(x_i - \bar{x})^2}{25}}$

$$= \sqrt{\frac{1}{25}[(3-12)^2 + (4.5-12)^2 + (6-12)^2 + (9-12)^2 + (15-12)^2 + \ldots + (18-12)^2 + (21-12)^2 + (27-12)^2]}$$

$$= \sqrt{\frac{1}{25}[81+56.25+36+9+9+56.25+36+20.25+2.25+20.25+9+20.25+9+0+36+9+2.25+0+9+81+9+56.25+36+81+225]}$$

$$= \sqrt{\frac{909}{25}} = 6.03$$

Example 13 : If the population is 3, 6, 9, 15, 27

(*a*) List all possible samples of size 3 that can be taken without replacement from the finite population.

(*b*) Calculate the mean of each of the sampling distribution of means.

(*c*) Find the standard deviation of sampling distribution of means.

[JNTU Nov 2004, (A) Dec. 2009, (K) II Sem. June 2015 (Set 3)]

(or) A population consists of five numbers 3,6,9,15,27. Consider all possible samples of size three that can be drawn without replacement from this population. Find

(*i*) The population mean. (*ii*) The population standard deviation.

(*iii*) The mean of the sampling distribution of he means .

(*iv*) The standard deviation of the sampling distribution of the means.

[JNTU (K) Dec. 2015 , (H) May 2017]

Solution : Mean of the population, $\mu = \frac{3+6+9+15+27}{5} = \frac{60}{5} = 12$

Standard deviation of the population,

$$\sigma = \sqrt{\frac{(3-12)^2+(6-12)^2+(9-12)^2+(15-12)^2+(27-12)^2}{5}}$$

$$= \sqrt{\frac{81+36+9+9+225}{5}} = \sqrt{\frac{360}{5}} = 8.4853$$

(*a*) **Sampling without replacement (finite population) :**

The total number of samples without replacement is ${}^{N}C_n = {}^{5}C_3 = 10$

The 10 samples are

(3, 6, 9), (3, 6, 15), (3, 9, 15), (3, 6, 27), (3, 9, 27), (3, 15, 27), (6, 9, 15)

(6, 9, 27), (6, 15, 27), (9, 15, 27).

Computations :

Sample means	6	8	12	9	13	15	10	14	16	17
x_i	6	8	9	10	12	13	14	16	15	17
f_i	1	1	1	1	1	1	1	1	1	1

(*b*) Mean of the sampling distribution of means is

$$\mu_{\bar{x}} = \frac{6+8+9+10+12+13+14+15+16+17}{10} = \frac{120}{10} = 12$$

(*c*) $$\sigma_{\bar{x}}^2 = \frac{1}{9}[(6-12)^2+(8-12)^2+(9-12)^2+(10-12)^2+(12-12)^2$$

$$+(13-12)^2+(14-12)^2+(15-12)^2+(16-12)^2+(17-12)^2] = \frac{120}{9} = 13.3$$

$\therefore \sigma_{\bar{x}} = \sqrt{13.3} = 3.651$

Example 14 : Let S = {1, 5, 6, 8}, find the probability distribution of the sample mean for random sample of size 2 drawn without replacement.

(OR) 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 (*ii*) The population standard deviation (*iii*) The mean of the sampling distribution of means, (*iv*) The standard deviation of the sampling distribution of means.

Solution : Let $S = \{1,5,6,8\}$

(*a*) Mean of the population, $\mu = \frac{1+5+6+8}{4} = \frac{20}{4} = 5$

(*b*) S. D. of the population, $\sigma = \sqrt{\frac{1}{n}\sum_i (x_i - \mu)^2}$

$$= \sqrt{\frac{1}{4}(1-5)^2+(5-5)^2+(6-5)^2+(8-5)^2}$$

$$= \sqrt{\frac{1}{4}[16+0+1+9]} = \sqrt{\frac{26}{4}} = \sqrt{\frac{13}{2}} = \sqrt{6.5} = 2.5495$$

(*c*) We have $S = \{1, 5, 6, 8\}$

Here size 2 is to be drawn

$$\begin{Bmatrix} (1,5) & (1,6) & (1,8) \\ (5,6) & (5,8) & \\ (6,8) & & \end{Bmatrix}$$

i.e., 6 samples

Sampling distribution of means are 3, 3.5, 4.5, 5.5, 6.5, 7

$\therefore$ Mean of sampling distribution, $\bar{x} = \frac{3+3.5+4.5+5.5+6.5+7}{6} = \frac{30}{6} = 5$

(*d*) S.D. of sampling distribution of means,

$$\sigma = \sqrt{\frac{(3-5)^2+(3.5-5)^2+(4.5-5)^2+(5.5-5)^2+(6.5-5)^2+(7-5)^2}{5}}$$

$$= \sqrt{\frac{4+2.25+0.25+0.25+2.25+4}{5}} = \sqrt{\frac{13}{5}} = 1.612$$

Example 15 : The variance of a population is 2. The size of the sample collected from the population is 169. What is the standard error of mean.

Solution : n = The size of the sample = 169

σ = S.D of population = $\sqrt{\text{variance}} = \sqrt{2}$

Standard Error (S.E.) of mean = $\frac{\sigma}{\sqrt{n}} = \frac{\sqrt{2}}{\sqrt{169}} = \frac{1.41}{13} = 0.10846$

Example 16 : What is the effect on standard error, if a sample is taken from an infinite population of sample size is increased from 400 to 900.

Solution : S.E of the mean = $\frac{\sigma}{\sqrt{n}}$; Sample size = n

Let $n = n_1 = 400$

$\therefore$ S. E$_1 = \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{400}} = \frac{\sigma}{20}$

$\therefore P(75 \le \bar{x} \le 78) = P(z_1 \le z \le z_2)$

$= P(-0.625 \le z \le 1.25)$

$= P(-0.625 \le z \le 0) + P(0 \le z \le 1.25)$

$= 0.2334 + 0.3944 = 0.628$

Example 21 : A random sample of size 64 is taken from an infinite population having the mean 45 and the Standard deviation 8. What is the probability that x will be between 46 and 47.5.

[JNTU (H) Dec. 2011 (Set No.1)]

Solution : We are given

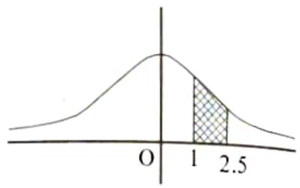
n = Sample size = 64

μ = Mean of infinite population = 45

σ = S. D. of infinite population = 8

The standard normal variate corresponding to $\bar{x}$ is

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



When $\bar{x} = 46$, $z = \dfrac{46-45}{8/\sqrt{64}} = 1$

and when $\bar{x} = 47.5$, $z = \dfrac{47.5-45}{8/\sqrt{64}} = 2.5$

$\therefore$ Required probability, $p = P(46 < \bar{x} < 47.5)$

$= P(1 < z < 2.5)$

$= A(2.5) - A(1)$

$= 0.4938 - 0.3413$ [From Normal distribution tables]

$= 0.1525$

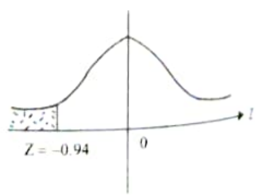
Example 22 : A normal population has a mean of 0.1 and standard deviation of 2.1. Find the probability that mean of a sample of size 900 will be negative.

[JNTU (H) Nov. 2010, (K) May 2013, Dec. 2015, II Sem. June 2015, (H) Sept. 2017]

Solution : Given $\mu = 0.1$, $\sigma = 2.1$ and $n = 900$

The standard normal variate is

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{x}-0.1}{2.1/\sqrt{900}} = \frac{\bar{x}-0.1}{2.1/30} = \frac{\bar{x}-0.1}{0.07}$$



$\Rightarrow \bar{x} = 0.1 + 0.07\,z$ where $z \sim N(0, 1)$

The required probability, that the sample mean is negative is given by

$P(\bar{x} < 0) = P((0.1 + 0.07\,z) < 0)$

$= P(0.07\,z < -0.1)$

$= P\left(z < \dfrac{-0.1}{0.07}\right) = P(z < -1.43)$

$= 0.50 - P(0 < z < 1.43)$

$= 0.50 - 0.4236 = 0.0764$

Example 23 : If a 1-gallon can of paint covers on an average 513 square feet with a standard deviation of 31.5 square feet, what is the probability that the mean area covered by a sample of 40 of these 1-gallon cans will be anywhere from 510 to 520 square feet?

[JNTU (K) Nov. 2009 (Set No. 2)]

Solution : Given $n = 40, \mu = 513$ and $\sigma = 31.5$ sq.ft

The test statistic is $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

When $\bar{x} = 510$, $z = \dfrac{510-513}{31.5/\sqrt{40}} = -0.6$

When $\bar{x} = 520$, $z = \dfrac{520-513}{31.5/\sqrt{40}} = 1.4$

$\therefore$ Required probability $= P(-0.6 < z < 1.4)$

$= P(-0.6 < z < 0) + P(0 < z < 1.4)$

$= P(0 < z < 0.6) + P(0 < z < 1.4)$

$= 0.2258 + 0.4192 = 0.645$

Example 24 : A random sample of size 64 is taken from a normal population with $\mu = 51.4$ and $\sigma = 68$. What is the probability that the mean of the sample will (*a*) exceed 52.9 (*b*) fall between 50.5 and 52.3 (*c*) be less than 50.6. **[JNTU (K) Nov. 2011, Dec. 2015 (Set No. 3)]**

Solution :

Given n = Size of the sample = 64

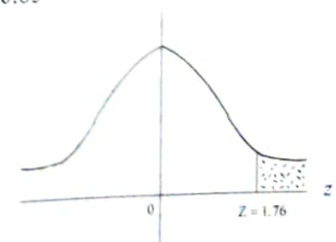
μ = Mean of the population = 51.4

σ = S.D of the population = 6.8

Standard error, $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{6.8}{\sqrt{64}} = \dfrac{6.8}{8} = 0.85$

(*a*) $P(\bar{x}$ exceed $52.9) = P(\bar{x} > 52.9)$

$\therefore z = \dfrac{\bar{x}-\mu}{\sigma_{\bar{x}}} = \dfrac{52.9-51.4}{0.85} = 1.76$



$\therefore P(\bar{x} > 52.9) = P(z > 1.76)$

$= 0.5 - P(0 < z < 1.76)$

$= 0.5 - 0.4608 = 0.0392$

$\therefore P(75 \le \bar{x} \le 78) = P(z_1 \le z \le z_2)$

$= P(-0.625 \le z \le 1.25)$

$= P(-0.625 \le z \le 0) + P(0 \le z \le 1.25)$

$= 0.2334 + 0.3944 = 0.628$

Example 21 : A random sample of size 64 is taken from an infinite population having the mean 45 and the Standard deviation 8. What is the probability that x will be between 46 and 47.5.

[JNTU (H) Dec. 2011 (Set No.1)]

Solution : We are given

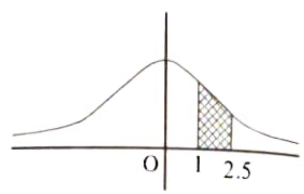
n = Sample size = 64

μ = Mean of infinite population = 45

σ = S. D. of infinite population = 8

The standard normal variate corresponding to $\bar{x}$ is

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



When $\bar{x} = 46$, $z = \frac{46 - 45}{8 / \sqrt{64}} = 1$

and when $\bar{x} = 47.5$, $z = \frac{47.5 - 45}{8 / \sqrt{64}} = 2.5$

$\therefore$ Required probability, $p = P(46 < \bar{x} < 47.5)$

$= P(1 < z < 2.5)$

$= A(2.5) - A(1)$

$= 0.4938 - 0.3413$ [From Normal distribution tables]

$= 0.1525$

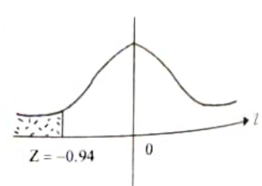
Example 22 : A normal population has a mean of 0.1 and standard deviation of 2.1. Find the probability that mean of a sample of size 900 will be negative.

[JNTU (H) Nov. 2010, (K) May 2013, Dec. 2015, II Sem. June 2015, (H) Sept. 2017]

Solution : Given $\mu = 0.1$, $\sigma = 2.1$ and $n = 900$

The standard normal variate is

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - 0.1}{2.1/\sqrt{900}} = \frac{\bar{x} - 0.1}{2.1/30} = \frac{\bar{x} - 0.1}{0.07}$$



$\Rightarrow \bar{x} = 0.1 + 0.07\, z$ where $z \sim N(0, 1)$

The required probability, that the sample mean is negative is given by

$P(\bar{x} < 0) = P((0.1 + 0.07\, z) < 0)$

$= P(0.07\, z < -0.1)$

$= P\left(z < \frac{-0.1}{0.07}\right) = P(z < -1.43)$

$= 0.50 - P(0 < z < 1.43)$

$= 0.50 - 0.4236 = 0.0764$

Example 23 : If a 1-gallon can of paint covers on an average 513 square feet with a standard deviation of 31.5 square feet, what is the probability that the mean area covered by a sample of 40 of these 1-gallon cans will be anywhere from 510 to 520 square feet?

[JNTU (K) Nov. 2009 (Set No. 2)]

Solution : Given $n = 40, \mu = 513$ and $\sigma = 31.5$ sq.ft

The test statistic is $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

When $\bar{x} = 510$, $z = \frac{510 - 513}{31.5/\sqrt{40}} = -0.6$

When $\bar{x} = 520$, $z = \frac{520 - 513}{31.5/\sqrt{40}} = 1.4$

$\therefore$ Required probability $= P(-0.6 < z < 1.4)$

$= P(-0.6 < z < 0) + P(0 < z < 1.4)$

$= P(0 < z < 0.6) + P(0 < z < 1.4)$

$= 0.2258 + 0.4192 = 0.645$

Example 24 : A random sample of size 64 is taken from a normal population with $\mu = 51.4$ and $\sigma = 68$. What is the probability that the mean of the sample will (*a*) exceed 52.9 (*b*) fall between 50.5 and 52.3 (*c*) be less than 50.6. **[JNTU (K) Nov. 2011, Dec. 2015 (Set No. 3)]**

Solution :

Given n = Size of the sample = 64

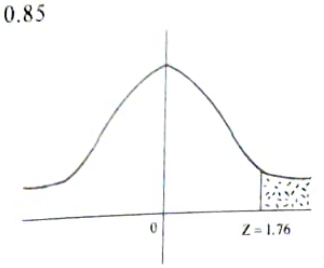
μ = Mean of the population = 51.4

σ = S.D of the population = 6.8

Standard error, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6.8}{\sqrt{64}} = \frac{6.8}{8} = 0.85$

(*a*) $P(\bar{x}$ exceed $52.9) = P(\bar{x} > 52.9)$

$\therefore z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{52.9 - 51.4}{0.85} = 1.76$



$\therefore P(\bar{x} > 52.9) = P(z > 1.76)$

$= 0.5 - P(0 < z < 1.76)$

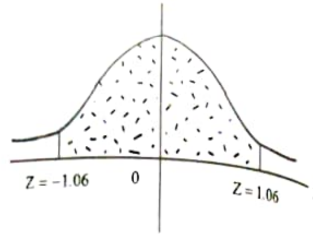
$= 0.5 - 0.4608 = 0.0392$

(b) $P(\bar{x}$ fall between 50.5 and 52.3)

i.e. $P(50.5 < \bar{x} < 52.3) = P(\bar{x}_1 < \bar{x} < \bar{x}_2)$

$$\therefore z_1 = \frac{\bar{x}_1 - \mu}{\sigma_{\bar{x}}} = \frac{50.5-51.4}{0.85} = -1.06$$

$$z_2 = \frac{\bar{x}_2 - \mu}{\sigma_{\bar{x}}} = \frac{52.3-51.4}{0.85} = 1.06$$



$\therefore P(50.5 < \bar{x} < 52.3)$

$= P[-1.06 < z < 1.06]$

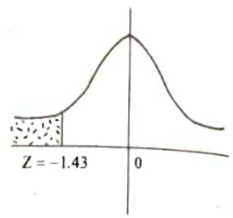
$= P[-1.06 < z < 0] + P[0 < z < 1.06]$

$= P[0 < z < 1.06] + P(0 < z < 1.06] = 2P[0 < z < 1.06]$

$= 2(0.3554) = 0.7108$

(c) $P(\bar{x}$ will be less than 50.6) $= P(\bar{x} < 50.6)$

$$= P(z < -0.94)\left[\because z = \frac{50.6-51.4}{0.85} = -0.94\right]$$



$= 0.50 - P(0.94 < z < 0)$

$= 0.50 - P(0 < z < 0.94) = 0.50 - 0.3264$

$= 0.1736$

Example 25 : If the mean of breaking strength of copper wire is 575 lbs, with a standard deviation of 8.3 lbs. How large a sample must be used in order that there will be one chance in 100 that the mean breaking strength of the sample is less than 572 lbs ?

[JNTU (A) Apr. 2012, (K) June 2015 (Set No.2)]

Solution : Given $\bar{x}$ = Mean of the sample = 572 lbs.

μ = Mean of the population = 575 lbs.

σ = Standard deviation of population = 8.3 lbs

n = ?

$$\therefore z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{572-575}{8.3/\sqrt{n}} = \frac{3\sqrt{n}}{8.3}$$

The probability that $\bar{x} < 572$ is $\frac{1}{100} = 0.01$

$\therefore$ We have to find the value of z for which 0.01 = area to the right at the variate z

So the area to the left is 0.99

The corresponding value of z is 2.33 (from table)

$$\therefore z = \frac{3}{8.3}(\sqrt{n}) = 2.33 \Rightarrow n = 42 \text{ (nearly)}$$

Example 26 : The guaranteed average life of a certain type of electric bulbs is 1500 hrs with a S.D of 120 hrs. It is decided to sample the output so as to ensure that 95% of bulbs do not fall short of the guaranteed average by more than 2%. What will be the minimum sample size ?

[JNTU (A) Nov. 2010, (K) June 2015 (Set No. 3)]

Solution : Let n be the size of the sample.

The guaranteed mean is 1500

We do not want the mean of the sample to be less than 2% of (1500) i.e., 30 hrs.

So $1500 - 30 = 1470$

$\therefore \bar{x} > 1470$

$$\therefore |z| = \left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right| = \left|\frac{1470-1500}{120/\sqrt{n}}\right| = \frac{\sqrt{n}}{4}$$

From the given condition, the area of the probability normal curve to the left of $\sqrt{n}/4$ should be 0.95

$\therefore$ The area between 0 and $\sqrt{n}/4$ is 0.45

We do not want to know about the bulbs which have life above the guaranteed life.

When the area is 0.45, $z = 1.65$

$$\therefore \frac{\sqrt{n}}{4} = 1.65 \quad i.e., \sqrt{n} = 6.6 \quad \therefore n = 44$$

Example 27 : Determine the expected number of random samples having their means (a) Between 22.39 and 22.41 (b) Greater than 22.42 (c) Less than 22.37 (d) Less than 22.38 or more than 22.41 for the following data.

N = Size of the population = 1500

n = Size of the sample = 36, Number of samples (N_s) = 300

σ = Population S.D = 0.48, μ = Population mean = 22.4

Solution :

(a) $P(22.39 < \bar{x} < 22.41) = P(-1.26 < z < 1.26)$

$$\text{Since } z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{x}-22.41}{0.0079} = \frac{2}{0.3962} = 0.7924$$

$\therefore$ Expected no. of samples = (Total no. of samples) × Probability

$= N_s \times P(\bar{x})$

$= 300 \times 0.7924 \cong 238$

(b) $P(\bar{x} > 22.42) = P(z > 2.53) = 0.00057$

$\therefore$ Expected no. of samples $= 300\ (0.00057) \cong 2$

(c) $P(\bar{x} > 22.37) = P(z < -3.8) = 0.0001$

Expected no. of samples $= (300)\ (0.0001) \cong 0$

(d) $P\,[(\bar{x} < 22.38)$ and $(\bar{x} > 22.41)] = P(z < -2.53$ and $z > 1.26)$

$= 0.0057 + 0.1038 = 0.1095$

$\therefore$ Expected no. of samples $= (300)(0.1095) \approx 33$

Example 28 : The mean voltage of a battery is 15 and S.D is 0.2. Find the probability that four such batteries connected in series will have a combined voltage of 60.8 or more volts.

[JNTU (A) Nov. 2010 (Set No. 2), (K) May 2010S (Set No. 1)]

Solution : Let mean voltage of batteries A, B, C, D be $\bar{X}_A, \bar{X}_B, \bar{X}_C, \bar{X}_D$. The mean of the series of the four batteries connected is

$$\mu_{\bar{X}_A+\bar{X}_B+\bar{X}_C+\bar{X}_D} = \mu_{\bar{X}_A} + \mu_{\bar{X}_B} + \mu_{\bar{X}_C} + \mu_{\bar{X}_D}$$
$$= 15 + 15 + 15 + 15 = 60$$

$$\sigma_{A+B+C+D} = \sqrt{\sigma_A^2+\sigma_B^2+\sigma_C^2+\sigma_D^2} = \sqrt{4\,(0.2)^2} = 0.4$$

Let X be the combined voltage of the series

When $x = 60.8$, $z = \dfrac{x-\mu}{\sigma} = \dfrac{60.8-60}{0.4} = 2$

Then probability that the combined voltage is more than 60.8 is given by

$P(X \geq 60.8) = P(z \geq 2) = 0.5 - 0.4772 = 0.0228.$

Example 29 : Three masses are measured as 62.34, 20.48, 35.97 kgs with S.D 0.54, 0.21, 0.46 kgs. Find the mean and S.D of the sum of the masses.

Solution :

Let the three masses measured for A, B, C be $\bar{X}_A, \bar{X}_B, \bar{X}_C$

The mean of the sum of the masses is

$$\mu_{\bar{X}_A+\bar{X}_B+\bar{X}_C} = \mu_{\bar{X}_A} + \mu_{\bar{X}_B} + \mu_{\bar{X}_C}$$
$$= 62.34 + 20.48 + 35.97 = 118.79$$

and $$\sigma_{A+B+C} = \sqrt{\sigma_A^2+\sigma_B^2+\sigma_C^2} = \sqrt{(0.54)^2+(0.21)^2+(0.46)^2}$$
$$= 0.74$$

Example 30 : The mean life time of light bulbs produced by company is 1500 hours and S.D of 150 hours. Find the probability that lighting will take place for (a) atleast 5000 h (b) at most 4200h if three bulbs are connected such that when one bulb burns out, another bulb will go on. Assume that life times are normally distributed.

Solution :

Let the mean life time of light bulbs L_1, L_2 and L_3 be $\bar{X}_{L_1}, \bar{X}_{L_2}, \bar{X}_{L_3}$

Then mean of these light bulbs is $\mu_{\bar{X}_{L_1}+\bar{X}_{L_2}+\bar{X}_{L_3}}$

$$= \mu_{\bar{X}_{L_1}} + \mu_{\bar{X}_{L_2}} + \mu_{\bar{X}_{L_3}} = 1500 + 1500 + 1500 = 4500$$

and S.D is $\sigma_{L_1+L_2+L_3} = \sqrt{\sigma_{L_1}^2+\sigma_{L_2}^2+\sigma_{L_3}^2}$

$$= \sqrt{3\,(150)^2} = 260.$$

(a) The probability that lighting will take place at least 5000 h is

$$P(X > 5000) = P\left(z > \frac{x-\mu}{\sigma}\right) = P\left(z > \frac{5000-4500}{260}\right) = P(z > 1.92)$$
$$= 0.5 - 0.4726 = 0.0274$$

(b) The probability that the lighting will take place at most 4200 h is

$$P(X < 4200) = P\left(z < \frac{4200-4500}{260}\right)$$
$$= P\,(z < -1.15)$$
$$= 0.5 - 0.3749 = 0.1251.$$

Example 31 : The diameter of motor shafts in a lot has a mean of 0.249 inch and a S.D of 0.003 inch. The inner diameter of bearings in another lot have a mean of 0.255 inch and a S.D of 0.002 inch.

(i) What are the mean and the S.D of the clearances between shafts and bearings selected from those lots ?

(ii) If a shaft and a bearing are selected at random, what is the probability that the shaft will not fit inside, the bearing ? Assume that both dimensions are normally distributed. **[JNTU April 2003]**

Solution :

Let $\bar{X}_1$ be the mean diameter of bearing and $\bar{X}_2$ be the mean diameter of shaft.

Given $\bar{X}_1 = 0.255$, $\bar{X}_2 = 0.249$ and $\sigma_1 = 0.002$, $\sigma_2 = 0.003$

(i) Let $\bar{X}_d$ be the mean diameter of the difference in the two diameters.

$$\therefore \bar{X}_d = \bar{X}_1 - \bar{X}_2 = \mu_1 - \mu_2 = 0.255 - 0.249 = 0.006$$

$$\sigma_d = \sqrt{\sigma_1^2+\sigma_2^2} = \sqrt{(0.002)^2+(0.003)^2} = 0.0036$$

(ii) Shaft will not fit inside bearing if $d < 0$.

Now $z = \dfrac{0-0.006}{0.0036} = -1.6666$

Probability that shaft will not fit inside bearing $= P(d < 0) = P(z < -1.6666)$

$= 0.5 - 0.4515 = 0.0485.$

Example 32 : If the distribution of the weights of all men travelling by air between Delhi and Mumbai has a mean of 163 pounds and a standard deviation of 18 pounds. What is the probability that the consigned gross weight of 36 men travelling between these two cities is more than 6000 pounds. **[JNTU (A) Dec. 2009 (Set No. 2)]**

Solution : Given μ = The mean of the population = 163 pounds

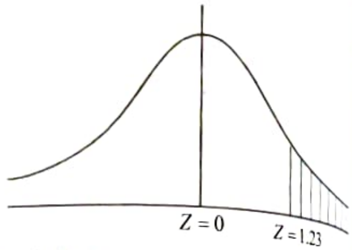
σ = The standard deviation of the population = 18 pounds

Gross weight of 36 men = 6000 pounds

$\therefore \bar{x}$ = Average weight of a man $= \frac{6000}{36} = 166.7$ pounds

Also we have n = sample size = 36

$$\therefore z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{166.7-163}{18/\sqrt{36}} = \frac{3.7}{3} = 1.23$$



Hence $P(\bar{x} \geq 166.7) = P(Z \geq 1.23)$

$= 0.5 - P(Z < 1.23)$

$= 0.5 - 0.3907$ (from normal tables)

$= 0.1093$

Thus the probability that the consigned gross weight of 36men travelling between these two cities is more than 6000 pounds is 0.1093.

REVIEW QUESTIONS

1. Define population, sample and sampling. Give an example each. **[JNTU (K) Dec. 2013 (Set No.1)]**
2. Define (*i*) simple sample (*ii*) random sample (*iii*) purposive sample. **[JNTU (H) Dec. 2019 (R18)]**
3. Define parameter and statistic.
4. What is a statistic ? Give an example. **[JNTU (H) Dec. 2014]**
5. Define standard error of a statistic. **[JNTU (H) Dec. 2014]**
6. Explain sampling distribution and sampling distribution of a statistic.
7. State Central limit theorem. **[JNTU (H) May 2017]**
8. Write the standard error of
   (*i*) sample mean (*ii*) difference of two sample means **[JNTU (H) III yr.Nov. 2015]**
9. Write the standard error of (*i*) sample S.D. (*ii*) difference of two sample S.Ds

EXERCISE 6(A)

1. Find the values of the finite population correction factor for
   (*a*) $n = 5$ and $N = 200$ (*b*) $n = 50$ and $N = 300$
2. How many different samples of size $n = 2$ can be chosen from a finite population of size $N = 25$.
3. (*i*) A sample of size 400 is taken from a population whose S.D is 16. Find the standard error of means.
   (*ii*) A sample of size 80 is taken from a population whose S.D. is 15. Find the standard error of means. **[JNTU (H) Dec. 2019 (R18)]**

4. The mean weekly wages of workers are with S.D of rupees 4. A sample of 625 is selected, find the standard error of the mean.
5. When we draw a sample from an infinite population, what happens to the standard error of the mean if the sample size is
   (*a*) Increased from 50 to 200 (*b*) Decreased from 640 to 40
6. (*i*) A random sample of size 81 is taken from an infinite population having the mean 65 and S.D 10. What is the probability that $\bar{x}$ will lie between 66 and 68 ?
   (*ii*) Determine the probability that $\bar{X}$ will be between 22.39 and 22.41 if a random sample of size 36 is taken from an infinite population having the mean $\mu = 22.4$ and $\sigma = 0.048$.
   (*iii*) Determine the probability that $\bar{X}$ will be between 66.8 and 68.3 if a random sample of size 25 is taken from an infinite population having the mean $\mu = 68$ and $\sigma = 3$.
   (*iv*) Find $P(\bar{X} > 66.75)$ if a random sample of size 36 is drawn from an infinite population with mean $\mu = 63$ and $\sigma = 9$.
7. Find the probabilities that a random variable having the standard normal distribution will take on a value.
   (*a*) between 0.87 and 1.28 (*b*) between – 0.34 and 0.62 **[JNTU 1999]**
8. The average cost of a studio condominium in the cedar lakes development is Rs. 62,000 and the S.D is Rs. 4200. What is the probability that a condominium in this development will cost at least Rs. 65,000.
9. Determine the probability that the sample mean area covered by the sample of 40 of 1 litre paint boxes will be between 510 and 520 square feet, given that a 1 litre of such paint box covers on the average 513.3 square feet with S.D of 31.5 sft. **[JNTU April 2004 (Set No.3)]**
10. Calculate the probability that a random sample of 16 computers will have an average life of less than 775 hours assuming that length of life of computers is approximately normally distributed with mean 800 hours and S.D 40 hours.
11. (*i*) Construct S.D of means for the population 3, 7, 11, 15 by drawing samples of size two with replacement. Determine (*a*) μ (*b*) σ (*c*) S.D.M (*d*) $\mu_{\bar{x}}$ (*e*) $\sigma_{\bar{x}}$. Verify the results.
    (*ii*) A population consists of the four numbers 3,7,11,5. Consider all possible samples of size two that can be drawn with replacement from this population.
    Find (*a*) The population mean,
    (*b*) The population standard deviation,
    (*c*) The mean of the sampling distribution of means,
    (*d*) The standard deviation of the sampling distribution of means.