Ch.14 Sampling and Sampling distribution

Population
A population is defined as the aggregate or totality of all individual or objects having the same characteristics of the interest.
i.e. objects, trees etc
Statistical population
A big lot from which a few items are inspected is called statistical population or universe. The pop may be individual, objects, animals etc.The size of the pop is denoted by N it is to be noted that the term does not mean the human population
Finite Population
A pop is said to be finite pop it contains limited/Finite/countable number of items.
Example The pop of books in a library
The pop of boys in a college
Infinite Population
A pop is said to be infinite if it contains unlimited/infinite/uncountable number of items
Example The pop of points on a line
The drops of water in the sea
Target Population
A population from which information is desired is called target population.
Sampled/Studied population
A population from which information is actually taken is called sampled or studied population
Example:
Suppose, someone whish to study the problems of the students living in hostels of Bahawalpur. Due to shortage of time and money, if study is carried in "Umer hall and Ali hall" only. Then our target population is Bahawalpur hostels and studied population is "Umer hall and Ali hall".
Hypothetically population
A population which does not actually exist but we can be thought about it.
Example: All possible results of a die.
Existent population or concrete units
Populations which actually exist are called existent population or concrete units of population. Example: Such as trees, books, etc.
Sub population or domain of study
Parts or segments of a population from which the required statistical information is to be obtained separately for the specific purpose in view in order to distribute the domain of study or sub population.
Size of the population
The total numbers of units in a finite population is called the size of the population and is denoted by "N".
Population distribution:
The numerical values assigned to units of interest are created as values of a random variable "X" and the distribution of "X" is called the population distribution.
Sample: It is the part of the population under the study which is selected at random with the believe that it well represents all the characteristics of the populations
Example
a) 50 students are selected from all students at "IUB" during a season 2008-2008
b) A handful of grand's taken at random from a store etc.
Size of the sample
The number of sampling units selected from a given population is called a sample and it is denoted by "n".
Sampling units
The individual members of the population are called sampling units are simply units. A sampling units from which information is required, may be a college students, an animal, a tree, etc.
Sampling
i. A set of "n" sampling units selected from a given population is called a sample of size "n" and the process of selecting a sample is known as sampling
ii. It is a technique procedure or process of selecting a part of population under investigation or enumeration.

Sampling technique
It is a technique procedure / process of selecting a part of pop under examination.
Units of analysis
It is a unit which is used at the stage of calculation at such a unit may also be elementary units itself.
Purpose of sampling or aims of sampling
There are two basic purpose of sampling
a) To gain maximum information about population characteristic without using all the units of population.
b) To find reliable estimates of population parameters on low cost and minimum time.
Complete enumeration
If we collect information from all the units of population the study is called complete count / enumeration.
The example of complete enumerations population census Pakistan pop census was conducted in 1951, 1961, 1972, 1981, and 1998.
Reporting units
It is the unit which actually supplies the required statistical information and such units may be the elementary units or a unit representing a group of elementary units.
Example;-
The head of the family may be the reporting units as the supplies information about the number of family.
Parameter
i. Any measurement of characteristics of population is called parameter.
ii. Any characteristics such as A.M, S.D, Mode etc. which is measured directly , the population is called the parameter . It is usually denoted by a capital letters or Greek. Likes as A.M (μ); S.D (σ) etc.

The shape of distribution is depending on parameter. “Z” is slandered normal variate .any distribution based on at which things are called parameter.
Statistic
Any characteristics such as standard error, arithmetic mean etc. Which is measured from a part of pop is called statistic. It varies from sample to sample.

$t = \dfrac{\overline{X} - \mu}{S.E(\overline{X})}$ is a statistic

Standard error
The standard derivation of sampling dist. Of statistic is known as standard error of that statistic.

The S.E of mean ($\overline{X}$), when sampling with replacement is $\sigma_{\overline{X}} = S.E(\overline{X}) = S.E = \dfrac{\sigma}{\sqrt{n}}$

Sampling is with out replacement $\sigma_{\overline{X}} = S.E(\overline{X}) = S.E = \dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}$

It should be noted that S.E <S.D Or $\sigma_{\overline{X}} < \sigma$

Error used difference b/w parameter and statistics +ve skewed.
Uses of Standard error
i. To estimate efficiency of sampling method.
ii. To the estimate the sample size.
iii. The variety among the values of sample of mean
iv. Use in statistical inferences
v. To build confidence interval for pop parameter, we use standard error
vi. It is used to compare the efficiency of different estimators.
vii. Without standard error inferential statistics is blind.
viii. Standard error is used for test of consistency.
ix. Measures the reliability of sample results.
Note;-
Standard error or standard deviation of sampling distribution of sample mean measures variability among the values of sample mean standard error has great importance in statistical inference. It always less then the population standard i.e. $\sigma_{\overline{X}} < \sigma$
Standard error is also used to measure the reliability of sample results. The smaller amount of standard error indicates the greater reliability of the sample results.
Another use of Standard error is to compare the efficiency of different estimators.
In short, we can say, “Standard error is key for inferential statistics”

Sampling with replacement

Sampling is with replacement, if the selected unit is replaced to the population before selecting the next unit. In this case units are independent to each other

i. Size of population does not change

ii. A population unit may be selected more than once

iii. Total possible sample $= N^n$

iv. The sample size can be greater than population size

v. A finite population $\rightarrow$ infinite population

Sampling without replacement

Sampling is without replacement, if the selected unit is does not replaced to the population before selecting the next unit. In this case units are dependent to each other

i. Size of population decreases

ii. A population unit may not be selected again and again

iii. Total possible sample $= {}^N C_n$

iv. The sample size cannot be greater than population size

Probability sampling or random sampling

When each element in the population has a known non-zero (not necessarily equal) Probability of its being include in the sample, the sampling is said to be probability sampling. A probability sampling is also called random sampling.

The important methods or types of probability are

i. Simple random sampling
ii. Stratified random sampling
iii. Systematic sampling
iv. Cluster sampling
v. Multiphase sampling
vi. Multistage sampling
vii. Sequential samplings

Define non-probability sampling

If sampling is done on personal Judgment, then it is called non-probability sampling

The method of non-probability sampling

i. Judgment or purposive sampling
ii. Quota sampling

Sampling frame

It is complete list of all the elements in the population from which a sample is to be taken is called the sampling frame.

It is important to note that the frame should be.

i. Should cover the whole pop.

ii. Free from error and bias.

iii. Should not be insufficient.

iv. Should not be complete.

v. Should not contain inaccurate elements..

Sampling design

The complete procedure / plan regarding the size of population , size of sample , selection of sample , collection of sample , presentation of find results etc is called sampling design

Sampling distribution

Ans: The arrangement of all possible values of sample statistic with their probabilities is called sampling distribution of that statistic.

Differentiate between random sample and simple random sample?

Ans: When each unit of the population has known (not necessarily equal) probability of its selection in the sample is called random sample or probability sampling.

A sample is defined to be simple random sample. When each unit of the population has equal probability of its selection in the sample.

Sampling error or random error

i. The error which is used to the utilization of sample is called sampling error.

ii. The difference b/w the value of a statistic and true value population parameter is called sampling error. i.e.

Sampling error = statistic – parameter= $\hat{\theta} - \theta$

The sampling error can be reduced it by

i. Increasing sample size.

ii. Stratification.

iii. Multistage sampling.

It is to be noted if variance of sample statistic $\rightarrow$ minimum $\rightarrow$ greator reliability of sample.

If the difference is zero then $\overline{X}$ is an unbiased estimate of population parameter μ

The difference is may be +ve in case of over estimation.

The difference may be –ve in case of under estimation.

Non –sampling error

It is due to the non – response of the people when they are conducted during a sample survey. Non – sampling error occur at the stages of gathering and processing of data regardless. Whether a sample or complete census is taken.

These error occur due to,

i. All kind of human error.
ii. Faulty of sampling frame.
iii. Biased method of selection of units.
iv. Bias in response.
v. Errors of observation and measurements.
vi. Error editing, coding, classification.
vii. Defect in method of sampling, method of interviewing, method of measurement.

These errors can be minimized.

i. Proper selection of questionnaires
ii. Following up the non-response
iii. Proper training of investigators
iv. Correct manipulation of the collected information

Bias or Sampling bias

i. A systematic component of error which deprives the statistical results of its representiveness. Bias is different from random error, in the sense of random error can be balanced out or corrected in the long run but the bias cannot be corrected. Bias is introduced by the subjective Judgments of human beings or personal Judgments of human beings.
ii. The difference between expected value of a statistic and true value of population parameter is called sampling bias.

$$Sampling\ bias = E(Statistic) - Parameters = E(\overline{X}) - \mu$$

Sampling bias occur due to the following ways

i. Deliberate selection
ii. Substitution
iii. Incomplete coverage
iv. Haphazard selection
v. Inadequate interviewing
vi. Defect in the tool of measurement
vii. Wrong method of sampling

Deliberate selection

It is based on personal Judgments of what is representative. This personal Judgments introduced bias.

Substitution

Sometimes it becomes difficult to make contact with certain members or information is not obtained from certain members/units. In this case we substitute the members or units that are available, such substitution introduces the bias.

Incomplete coverage

Bias also enters when we fail to cover the whole of the population.

Haphazard selection

Haphazard human selection can also introduced bias, as every human being has a tendency away from randomness in his choices.

Inadequate interviewing

Bias also enters when the interviewing is incomplete and misleading.

Defect in the tool of measurement

When we fail to use an appropriate measurement bias also occur.

Wrong method of sampling

When we use wrong method of selection the samples and measurements then bias also occur.

How to avoid bias
In order to draw valid conclusions, all possible sources of sampling bias must be avoided. This end in achieved if the sampling is drawn entirely at random a well defined concepts in statistics, meanings that every unit has a known non-zero probability of being include in the sample. A sample that is free from selection and procedure of bias is called good or an unbiased sample. It is interesting to note that in some types of investigation or survays a certain amount of sampling bias, however tolerated.

Simple random sampling
A sample is defined to be simple random sample if
i. Each unit in the population has an equal probability for its selection in the sample.
ii. Each possible sample of same size has an equal probability the sample selected

Method
i) Goldfish bowl method ii) Using a random number table
iii) Using calculator/ computer (called pseudo random number)
A simple random sample can be selected above mentioned methods by with replacement and without replacement. The procedure of selecting a simple random sample is called simple random samplings

Difference between random sample and simple random sample
When each unit of the population has known (not necessarily equal) probability of its selection in the sample is called random sample or probability sampling.
A sample is defined to be simple random sample. When each unit of the population has equal probability of its selection in the sample.

Purposive sampling or Judgment sampling
A purposive sample is a non-random sample. The selection of sampling units is based on feelings of a person.
i. It is pure personal approach for selecting a sample
ii. It is subjective
iii. There exists flexibility
iv. Different person have different sample results in the same population.

Quota samplings
A quota sample is Judgment sample. The selection of sampling units is based on personal feelings of a person. Quota sampling may be considered as stratified sampling in which the selection of units within stratum is non-random.

Finite correction factor
When a sample of size "n" is drawn without replacement from a finite population of size "N" then

$$\sigma^2{}_{\overline{X}} = \frac{\sigma^2}{n}\frac{N-n}{N-1}$$

The factor $\frac{N-n}{N-1}$ is usually called finite population correction (fpc) for the variance.
fpc becomes or approaches to unity of population becomes larger and larger. So $\sigma^2{}_{\overline{X}} = \frac{\sigma^2}{n}$ and no need of fpc. fpc is dropped from the formula when $\frac{n}{N} < 5\%$ and it is used $\sigma^2{}_{\overline{X}} = \frac{\sigma^2}{n}\frac{N-n}{N-1}$ when $\frac{n}{N} > 5\%$. Here $\frac{n}{N}$ sampling fraction.

Representative sample or unbiased
A sample that is free from bias and error is called representative or unbiased sample.

Advantages of sampling
The important advantages of sampling are given below under complete enumeration
i. Need for sampling
Sometimes, sampling becomes part and parcel for us specially when some units of population are destroyed
ii. Reduced cost
Sampling saves money as it is much cheaper to collect desired information from a small sample than complete enumeration.
iii. Saving time
Sampling saves a lot of time and energy
iv. Greater speed
With the help of sampling, we can obtain our desired information's speedily.
v. Greater scope
The results obtained through sampling have a greater scope in every field of life.

vi. Reliability
The results of sampling are reliable
vii. Flexibility
Sampling has more flexible then complete enumeration.
viii. Detailed information
Detailed information can be obtained only through sampling methods.
ix. In case of infinite population
In case of infinite or inaccessible population sampling is the only way to obtain information
x. Modern techniques
There exist modern methods for sampling to solve our problems.
xi. Greater accuracy
With careful planning, there can be greater accuracy in our result.

Survey

Method of collecting detailed information relating to respective group.
Or
A study or investigation of a population usually human beings or economic, social or political institution.
The term "survey" may be taken to refer to complete coverage of the population as in a census, but it is often used to refer to a study dealing only with a sample from the population (i.e. sample survey) where the sample observations are used to make inferences or drawn conclusion about the whole population.

Sample survey

When survey is carried out with the help of sampling method then it is called sample survey i.e. in which a portion only and not the whole population are surveyed.
It may be two types
i. Descriptive survey
ii. Analytical survey

Descriptive survey

In descriptive survay objective in sampling to get information about the population
Example
The number of children and women who watch television

Analytical survey

In analytical survey comparison over made between different sub-groups of the population. In order to discover whether the difference exist in taken or not. This enables us to form all too statistical hypotheses about the population.
Principal steps of sample survey
i. State the objectives of survey
ii. Define the population we wish to survey
iii. Establish sampling frame
iv. Choose sample design
v. Organize a field work
vi. Summarize and analyze the data

Q. What is the principle steps involved in conduct of a sample survey?

Ans: The principle steps in a sample survey are grouped into following 6 headings.
i. Objectives of the survey
In a sample survey, the first step is to lay down a clear statement of objectives of the survey. it should be noted that these objectives are a cumulate with available resources in terms of money , man power and the time limit of the survey to be conducted .
ii. Definition of the population
The population from which the sample is to be drawn should be defined in clear and unambiguous terms. The sampled population (population to be sampled) should coincide with the target population (population about which information is required) the demographic, geographical, administrator and other boundaries of the population must be specified so that population must be specified so that there remains no ambiguity regarding the coverage of the survey.
iii. Determination of sampling frame and sampling units
The main requirement of sampling survey is to fix up the sampling frame i.e. (the list of all sampling units with reference to which relevant data are to be collected). It is the sampling frame which determines the sampling structure of a survey. The

determines the sampling structure of a survey. The population should be capable of division into units which are distinct non-overlapping and cover the entire population

iv. Selection of proper sampling design
If an appropriate sampling design is selected, the final estimates will be quite reliable. the size of the sample, procedure of selection and estimation of parameters along with the amount of risk involved are some of the important statistical aspects which should receive careful attention. If a number of sampling of sampling design for taking a sample is available, then the total risk i.e. the cost and precision should be considered before making a final selection of the sampling design.

v. Organization the field works
The achievement of the aims of a sample survey depends to a large extent on reliable field work. If fieldwork is done honestly, sincerely and according to the instruction laid down and if there is careful supervision of field staff. There remains no doubt about achieving aims of the survey. It is, therefore, necessary to make provision for adequate supervisory staff for inspection of fieldwork. `

vi. Summary and analyses of data
In a sample survey, the final step of analyses and drawing inferences from a sample to a population is very vital and fascinating issue.
Since the results of the survey are the basis for policy, making, it the most essential part of the sample survey and should be handled carefully. The analyses of the data collected in a survey may be broadly classified as follows.

i. Scrutiny and editing of the data.
ii. Tabulation of data.
iii. Statistical analyses.
iv. Reporting and conclusions.

Finally, report of the finding of the survey, suggesting possible actions to be taken, should be written.

Census

It means complete enumeration of population or population study.
Or
The term “census” is used for complete enumeration or counting of a population or groups at a point in time with respect to well defined characteristics.
In general census refers to an official periodic count of as population including such information as sex, age and occupation of every individual. In Pakistan census is held every ten years, head of household is required to provide information about all these members living under their roof. The Questions asked cover names, age, family status, education and employments.

Basic aims of sampling.

Ans: **Purpose of sampling or aims of sampling**
There are two basic purpose of sampling

c) To gain maximum information about population characteristic without using all the units of population.
d) To find reliable estimates of population parameters on low cost and minimum time.

Define sampling units.

The individual members of the population are called sampling units are simply units. A sampling units from which information is required, may be a college students, an animal, a tree, etc.

Q.What do you mean by unbiased estimator?
Ans: An estimator is said to be unbiased if the mean of sampling distribution of the statistic is equal to its parameter.
$E(\bar{X}) = \mu$

Define unbiasedness.

The property of an estimator being free from bias is called unbiasedness.

Q. Differentiate between biased estimate and unbiased estimate?
Ans: An estimator is said to be unbiased if the mean of sampling distribution of the statistic is equal to its parameter. $E(\bar{X}) = \mu$ Otherwise it is biased $E(\bar{X}) \neq \mu$

Q: Why probability sampling is preferred over judgment sampling?
Ans: Probability sampling is preferred over judgment sampling because

i) In probability sampling reliability (variance) can be measured which is not possible in judgment sampling.
ii) Judgment may give rise to personal bias.

Q: Difference between random sample and simple random sample?
Ans: When each unit of the population has known (not necessarily equal) probability of its selection in the sample is called random sample or probability sampling.
A sample is defined to be simple random sample. When each unit of the population has equal probability of its selection in the sample.
Q: Why random sampling is used?
Ans: Random sampling is used to
i) Eliminate bias
ii) Provide basis for statistical inference.

Uses of sampling in daily life

Following are some uses of sampling in daily life.

i. The cook taste a bit of cooked food to find that he is cooked well or not .
ii. A food inspector take the sample of the food items , like milk , flour , ghee , wheat , & other thing to check that these things are pure or not .
iii. The customer by observation, sample the quality of vegetables or fruit which he intend to buy.
iv. While checking the blood group of a patient, the doctors take a drop of blood (as a sample) from patient's body instead of whole blood which is too dangerous for the patient.
v. We use the sampling technique while purchasing the bricks for the construction of house.
vi. The engineers take sample from different localities, while checking for oil and different types of natural minerals.

While checking the fertility of a land, the agriculture research workers take a sample of the soil for testing

Q. Why we use sample instead of population?

Ans:
Following are some points

i) It is difficult to handle a population which usually consists of a large number of units.
ii) Too much time is required to study the whole population and often the study becomes out-dated by time, it completed.
iii) Finances required to cover the whole population can hardly be made available.
iv) In a study, where individuals are killed under the observation study. The population serves no purpose. To clarify this point , we have an

Example
If all the battery cells of a manufacturing firm are put to live testing, nothing will be left for use.

i) In case the population is infinite or consists of uncountable numbers of units, its study is impossible.
ii) Some people think that the complete enumeration yields better results than the sampling study. However, this is not correct because complete enumeration adds many errors which are reduced by sampling. Hence in many cases sample study yields better results than the population study.

Q. Differentiate between sample distribution and population distribution?

Ans: If we arrange the values of a single sample in grouped data that is called sample distribution. While when we arrange all the elements of the population in grouped data that is called population distribution.

Mean square error (MSE)

In sample survey we have to get different estimate of population parameter and we want to see which estimate is very closer to the true value of population parameter. We use criteria of a mean square error, which is different from the expectation of the square differences of an estimate and the true value.

$$M.S.E(\hat{\theta}) = E(\hat{\theta}-\theta)^2$$

Q. Show that $M.S.E(\hat{\theta}) = Var(\hat{\theta}) + (Bias)^2$

Proof:
Let by definition

$$M.S.E(\hat{\theta}) = E(\hat{\theta}-\theta)^2$$

Adding and subtracting $E(\hat{\theta})$, we get

$$M.S.E(\hat{\theta}) = E\left[\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta\right]^2$$

$$M.S.E(\hat{\theta}) = E\left[(\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta)\right]^2$$
$$M.S.E(\hat{\theta}) = E\left[\left(\hat{\theta} - E(\hat{\theta})\right)^2 + \left(E(\hat{\theta}) - \theta\right)^2 + 2\left(\hat{\theta} - E(\hat{\theta})\right)\left(E(\hat{\theta}) - \theta\right)\right]$$
$$M.S.E(\hat{\theta}) = E\left(\hat{\theta} - E(\hat{\theta})\right)^2 + E\left(E(\hat{\theta}) - \theta\right)^2 + 2E\left(\hat{\theta} - E(\hat{\theta})\right)\left(E(\hat{\theta}) - \theta\right)$$
$$M.S.E(\hat{\theta}) = Var(\hat{\theta}) + \left(Bias(\hat{\theta})\right)^2 + 2(0)$$
$$M.S.E(\hat{\theta}) = Var(\hat{\theta}) + \left(Bias(\hat{\theta})\right)^2 \quad \text{Therefore} \quad E\left\{\hat{\theta} - E(\hat{\theta})\right\} = \left\{E(\hat{\theta}) - E(\hat{\theta})\right\} = 0$$

If E is an unbiased estimator then $E(\hat{\theta}) - \theta = 0$

Then

$$M.S.E(\hat{\theta}) = Var(\hat{\theta})$$

Q. What is the relation between MSE and variance of an estimate?

Ans: let by definition MSE is

$$M.S.E(\hat{\theta}) = Var(\hat{\theta}) + \left(Bias(\hat{\theta})\right)^2$$

If an estimator is an unbiased then there is no biased i.e. bias=0

$$M.S.E(\hat{\theta}) = Var(\hat{\theta}) + (0)^2$$

Then

$$M.S.E(\hat{\theta}) = Var(\hat{\theta})$$

Q. What is difference between precision and accuracy?

Ans: Precision is the size of the deviation from the repeated sample mean whereas accuracy is the size of the deviation from the overall population mean.

Q. What is the sampling fraction?

Ans: It is the ratio of the sample size “n” and the population size “N”

i.e. $Sampling\ \ fraction = \dfrac{n}{N}$

Q. What does happen with the sampling error, standard error and bias when we increase sample size?

Ans: When we increase sample size both the sampling error, standard error decreases and bias is increases.

Q. How the efficiency of the biased estimator is compared?

Ans: The efficiency of the biased estimator is compared by using their MSE(mean square error)

Q.17: Explain how would select a random sample of 10 households from a list of 250 households, by using a table of random numbers.

Ans: We are required to select a random sample of 10 household from a list of 250 household by using random number table.

i) Firstly list the household and assign the three digit number of each household from 001 to 250 (as 250 has 3 digit number)

ii) Then using first three columns from random numbers table, we have following 10 households for our sample by ignoring a number large than 250.The required household are 023,039,084,070,018,226,072,101,029,181

Note: If sampling is with replacement a household may be selected again and again but in without replacement. When without replacement no household is repeated if appear we ignore it.

Q: When we prefer different sample size allocation techniques in stratified random sampling?

Ans: If the entire stratum is of equal importance then we use equal allocation technique in which we select same sample size from from each stratum.

If the stratum sizes N_h are given then we use the proportional allocation in which n_h is proportional to the stratum size i.e. $n_h \ \alpha \ N_h$.

If the variability of the stratum is given S_h then we use Neyman allocation in which n_h is proportional $N_h S_h$ i.e. $n_h \ \alpha \ N_h S_h$

If the cost is given to each stratum along with stratum size and variability then we use optimum allocation in which n_h is proportional to $N_h S_h$ and C_h i.e. $n_h \ \alpha \ N_h S_h / \sqrt{C_h}$

Stratified random sampling

Sometimes the population highly variable material and simple random sample fail to adequately represent the population. It means that when the population is heterogeneous we cannot get more accurate results by simple random sample (SRS), so we divide the population into a numbers of mutually exclusive groups of units in such a way that the units within each group are as similar as possible.
This process of dividing the population is called stratification. The groups are called strata and the criterion by which we put sampling units in different strata is called stratifying factors.
Simple random sample from each stratum of those strata are selected and combined into a single sample. This technique is called stratified random sampling. This is also called mixed sampling as there is combined application of simple random sampling and stratified random sampling. There is probably no technique that is more widely used in sample design than stratification.

Systematic Sampling

A sample which is obtained by some systematic method, as opposed to random choice. For example sampling from a list, by taking individuals at equally spaced intervals called the sampling intervals or sampling from an area by determining a pattern of points on a map.

Example.14.1: Assume that a population consists of 5 students and the marks obtained by them in a certain statistics class are 20,15,12,16 and 18. Draw all possible random samples of two students when sampling is preformed
i) With replacement ii) without replacement. Calculate the mean marks for each sample.
Solution: Population: 20, 15,12,16,18
N=5 and n=2
Let we have a 5 students name "A, B, C, D and E"
i) Draw all possible sample of size two by with replacement $N^n = 5^2 = 25$

Sample N0.	Sample students	Sample Marks	Sample mean	Sample N0.	Sample students	Sample Marks	Sample mean
1	A,A	20,20	20	14	C,D	12,16	14
2	A,B	20,15	17.5	15	C,E	12,18	15
3	A,C	20,12	16	16	D,A	16,20	18
4	A,D	20,16	18	17	D,B	16,15	15.5
5	A,E	20,18	19	18	D,C	16,12	14
6	B,A	15,20	17.5	19	D,D	16,16	16
7	B,B	15,15	15	20	D,E	16,18	17
8	B,C	15,12	13.5	21	E,A	18,20	19
9	B,D	15,16	15.5	22	E,B	18,15	16.5
10	B,E	15,18	16.5	23	E,C	18,12	15
11	C,A	12,20	16	24	E,D	18,16	17
12	C,B	12,15	13.5	25	E,E	18,18	18
13	C,C	12,12	12				

ii) Draw all possible sample of size two by without replacement ${}^{N}C_n = {}^{5}C_2 = 10$

Sample N0.	Sample students	Sample Marks	Sample mean
1	A,B	20,15	17.5
2	A,C	20,12	16
3	A,D	20,16	18
4	A,E	20,18	19
5	B,C	15,12	13.5
6	B,D	15,16	15.5
7	B,E	15,18	16.5
8	C,D	12,16	14
9	C,E	12,18	15
10	D,E	16,18	17

Q.18: Using a random number table, select 30 samples of size 3 each with replacement from the following population distribution of heights. Find the mean of sample means.

Heights (inches)	No. of students
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8

Solution: We assign two digit numbers than we select 30 samples by using random number table of first two columns of size three

Heights (inches)	f	C.F	Assign number
60-62	5	5	00-04
63-65	18	23	05-22
66-68	42	65	23-64
69-71	27	92	65-91
72-74	8	100	92-99

Now we select 30 samples of size three along with corresponding height. We select the number within interval otherwise we ignore it

S.No	Random Numbers	Heights(X)	$\overline{X}=\frac{\sum x}{n}$	S.No	Random Numbers	Heights(X)	$\overline{X}=\frac{\sum x}{n}$
1	53,74,23	67,70,67	204/3	16	72,84,71	70,70,70	210/3
2	63,38,06	67,67,64	198/3	17	88,78,28	70,70,67	207/3
3	35,30,58	67,67,67	201/3	18	45,17,75	67,64,70	201/3
4	63,43,36	67,67,67	201/3	19	96,76,28	73,70,67	210/3
5	98,25,37	73,67,67	207/3	20	43,31,67	67,67,70	204/3
6	02,63,21	61,67,64	192/3	21	50,44,66	67,67,70	204/3
7	64,55,22	67,67,64	198/3	22	22,66,22	64,70,64	198/3
8	85,07,26	70,64,67	201/3	23	96,24,40	73,67,67	207/3
9	58,54,16	67,67,64	198/3	24	31,73,91	67,70,70	207/3
10	34,85,27	67,70,67	204/3	25	78,60,73	70,67,70	207/3
11	03,92,18	61,73,64	198/3	26	84,37,90	70,67,70	207/3
12	62,95,30	67,73,67	207/3	27	36,67,10	67,70,64	201/3
13	08,45,93	64,67,73	204/3	28	07,28,59	64,67,67	198/3
14	07,08,55	64,64,67	195/3	29	10,15,83	64,64,70	198/3
15	01,85,89	61,70,70	201/3	30	55,19,68	67,64,70	201/3

Sampling distribution of sample means

$\overline{X}$	Tally	f	$f(\overline{X})$	$\overline{X}\ f(\overline{X})$
192/3	I	1	1/30	192/90
195/3	I	1	1/30	195/90
198/3	IIIIII	7	7/30	1386/90
201/3	IIIIII	7	7/30	1407/90
204/3	IIII	5	5/30	1020/90
207/3	IIIIII	7	7/30	1449/90
210/3	II	2	2/30	420/90

$$\overline{X}=\sum \overline{X}f(\overline{X})=\frac{6069}{30}=202.3$$

Q.19: Draw with the help of random number a random sample of size 10 from a

i) Binomial distribution with parameter P=0.4 and n=5

ii) Poisson distribution with parameter $\mu=4$

Solution: As we know that the probability distribution function of binomial distribution

$P(x)={}^{n}C_{x}P^{x}q^{n-x}$

Where n=5 P=0.04 q=1-P=0.6

X	$P(x)={}^{n}C_{x}P^{x}q^{n-x}$	$P(X\le x)$	Assign number
0	0.078	0.078	000-077
1	0.259	0.337	078-336
.2	0.346	0.683	337-682
3	0.230	0.913	683-912
4	0.077	0.990	913-989
5	0.010	1.000	990-999

Now we use random number table to select 10 sample of three number digits are 537,633,353,634,982,026,645,850,585,348 corresponding values 2,2,2,2,4,0,2,3,2,2

Now we make a frequency distribution

X	Frequency
0	1
2	6
3	1
4	1

ii) As we know that the probability distribution function of binomial distribution

$$P(x)=\frac{e^{-\mu}\mu^{x}}{x!} \qquad X:0,1,2,.......$$

With Parameter $\mu=4$

X	$P(x)=\frac{e^{-\mu}\mu^{x}}{x!}$	$P(X\le x)$	Assign number
0	0.018	0.018	000-017
1	0.073	0.091	018-090
2	0.146	0.237	091-236
3	0.195	0.432	237-431
4	0.195	0.627	432-626
5	0.156	0.783	627-782
6	0.104	0.887	783-886
7	0.059	0.946	887-945
8+	$P(x\ge 8)=1-P(x<8)=0.054$	1.000	946-999

Now we use random number table to select 10 sample of three number digits are 537,633,353,634,982,026,645,850,585,348 corresponding values 4,5,3,5,8+,1,5,6,4,3

Now we make a frequency distribution

X	Frequency
1	1
3	2
4	2
5	3
6	1
8+	1

Example .14.2: The following frequency distribution gives the ages of a population of 1,000 college students:

Age(x)	14	15	16	17	18	19	20	Total
No. of students	6	61	270	491	153	15	4	1000

Using a random number table select a simple random sample of 20 students. Find the sample mean age and compare it with the population mean age.

Solution:

X	f	fx	Cf	Assign numbers
14	6	84	6	000-005
15	61	915	67	006-066
16	270	4320	337	067-336
17	491	8347	828	337-827
18	153	2754	981	828-980
19	15	285	996	981-995
20	4	80	1000	996-999
Total	1000	16785		

$$\mu=\frac{\sum fx}{\sum f}=\frac{16785}{1000}=16.785$$

Now we assign three digits random number table to select 15 samples and their corresponding Age(X) are 613,990,067,655,019,715,482,011,515,614,579,377,602,454,511 and values are 17,19,16,17,15,17,17,15,17,17,17,17,17,17,17

Frequency distribution

X	Tally	f	fx
15	*II*	2	30
16	*1*	1	16
17	*IIIIIIIIIII*	11	187
19	*I*	1	19
Total		15	252

$$\overline{X}=\frac{\sum fx}{15}=\frac{252}{15}=16.8$$

Hence verify the results

Example 14.3: Select a random sample of size 15 using a random number table, from a Poisson distribution with parameter $\mu = 3$.

Solution: Same as Q.19 (ii)

Example.14.4: With the help of random number draw a sample of 25 from a Norma distribution with $\mu = 60$ and $\sigma = 2.5$.

Solution: Same as Q.20

Q.20: Using a random number table, draw a sample of size 30 from a normal distribution with $\mu = 100$ and $\sigma^2 = 64$

Solution: Given that $\mu = 100$ and $\sigma^2 = 64$

As we know that the interval $(\mu \pm 3\sigma)$ contain approximately 100% observation

Then the interval is (76,124).

Now we sum up into 8 classes with 7 class interval

Classes	UCB	$Z = \frac{UCB - \mu}{\sigma}$	$P(Z \le z)$	Assign number
Under 76	76	-3.0	0.0013	0000-0012
76-83	83	-2.13	0.0166	0013-0165
83-90	90	-1.25	0.1055	0166-1054
90-97	97	-0.38	0.3520	1055-3519
97-104	104	0.50	0.6915	3520-6914
104-111	111	1.38	0.9162	6915-9161
111-118	118	2.25	0.9878	9162-9877
118- ∞	∞	∞	1.0000	9878-9999

Now we use random number table to select 30 samples of four number digits with corresponding classes

Classes	Random Numbers	f
Under 76		
76-83		
83-90	0887	1
90-97	1396,2505,2136,1214,2959	5
97-104	5190,5497,6872,4536,3941,3858,4346,5583,6356,6747,5381,5186,3591,3771	14
104-111	7391,7845,8721,3721,8301,6973,8687,	7
111-118	9167,9209,9366	3
118- ∞		

Q.21(a): Describe stratified random sampling explaining in detail the following types of allocation of sample sizes i) Proportional allocation ii) Optimum allocation

Ans: **Allocation problem**

The problem of deciding the sample size in different strata known as allocation problem.

The different types of allocations are

i) Equal allocation ii) Proportional allocation

iii) Nyman's allocation iv) Optimum allocation

Equal allocation

If the entire stratum is of equal importance then we use equal allocation technique in which we select same sample size from each stratum.

Proportional allocation

Sample sizes are chosen to be proportional to the stratum sizes.

Sample may be selected from each of strata in proportion to stratum size. This method is called proportional allocation or the stratified random sampling with a uniform sampling fraction $\left(f_h = \frac{n_h}{N_h} \right)$. The way of allocation is the simplest and most frequently used method due to the fact that the probability of an element in stratum "h" being included in the sub-sample "n_h" is equal to $\left(f_h = \frac{n_h}{N_h} \right)$

Where

N_h = Population size of "h" stratum. Hence each element in the population has the same probability of selection. We use same sampling fraction in each stratum.

n_h = "h" stratum sample size N= Size of the population

Since $\frac{n_h}{N_h} = \frac{n}{N}$

Therefore the stratum size n_h is given by $n_h = n\frac{N_h}{N}$ as we assume $\frac{N_h}{N} = W_h$,

So $n_h = nW_h$ where h=1,2,3,…,k.

Advantages

i) Proportional allocation is straight forward

ii) Its requires no knowledge about stratum variability and relative sampling cost just keeping in mind this method of allocation.

Neyman allocation

"Allocation of a sample in which cost are assume fixed in each stratum"

This method of allocation was proposed by "J.Neyman" in 1934 and finding n_h which minimizes the variance of the stratified sample mean for a fixed total sample size "n" assuming the cost of sampling units "C_h" to be same in all strata. The stratum sample size "n_h" is given by $n_h = n\frac{W_h S_h}{\sum W_h S_h} = n\frac{N_h S_h}{\sum N_h S_h}$ where h=1,2,3,…,k

Neyman allocation becomes exactly the same as the proportional allocation when all the stratum standard deviations are equal. It is used to keeping in mind only the stratum size and variability within the stratum.

Optimum allocation

We must choose the sample size "n" to satisfy certain precision or cost requirements. This method deal with different strata are likely to exhibit different degrees of variability, we must inevitably proceed beyond the choice of "n" to the allocation of the individual stratum sample size "n_h".Also to merely state our requirement in terms of the variance of some estimator will be sufficient in general. The cost of the survey also influences the variance of cost factor is not the less important. It is to imagine that the cost of survey per sampling unit in different strata cannot be the same. Different sampling costs for different strata imply that we must attempt to take some account of cost factors in determining a desirable allocation of stratum sample sizes.

In a national geographical survey, it is likely to be most convenient to sample different regions separately. Cost of sampling can also vary from region to region (Stratum to stratum) if only with respect to traveling expenses for survey workers. In one stratum, the cost of transport may be different from the other. As far as crop cutting experiments are concerned in one stratum the labour may be cheaper than the other. Hence it would not be wrong to fix the cost, if the survey in each stratum differently. Lent "C_h" be the cost per unit of the survey in the "h" stratum for which a sample of size "n_h" is stipulated. Also suppose "C_0" is the overhead fixed cost of the survey. In this way the total cost "C" of the survey comes out to be "$C = C_o + \sum C_h n_h$". Where "C_0 and C_h" are beyond our control. Hence we will determine the optimum value of "n_h" which minimizes the variance of "$\bar{y}_{st}$"

$$n_h = n\frac{W_h S_h / \sqrt{C_h}}{\sum W_h S_h / \sqrt{C_h}} = n\frac{N_h S_h / \sqrt{C_h}}{\sum N_h S_h / \sqrt{C_h}} \qquad \text{h=1,2,3,..,k}$$

It is used when

i) "n" is large

ii) S_h is large (stratum is more homogenous) iii) C_h is small

Q.21 (b): Select a stratified random sample of size n=8 by proportional allocation from the following population and find the sample mean and the estimate of the population mean.

Stratum I	$X_{11} = 3, X_{12} = 6, X_{13} = 4, X_{14} = 7$
Stratum II	$X_{21} = 10, X_{22} = 12, X_{23} = 15, X_{24} = 16, X_{25} = 16, X_{26} = 20$
Stratum III	$X_{31} = 16, X_{32} = 18, X_{33} = 21, X_{34} = 22, X_{35} = 26, X_{36} = 23$

Solution: Now we find the sample size in each stratum to use proportional allocation

$n_h = n\frac{N_h}{N}$ Where $N = 16$ $N_1 = 4$ $N_2 = 6$ $N_2 = 6$

$n_1 = 8\frac{4}{16} = 2$ $n_1 = 8\frac{6}{16} = 3$ $n_1 = 3$

Now we select a sample in each stratum by using random number table

Stratum I	$, X_{12} = 6,\ X_{14} = 7$	$\bar{\mathbf{x}}_1 = \frac{\sum x}{n_1} = \frac{13}{2} = 6.5$
Stratum II	$X_{21} = 10, X_{24} = 16, , X_{26} = 20$	$\bar{\mathbf{x}}_2 = \frac{\sum x}{n_2} = \frac{46}{3} = 15.33$
Stratum III	$X_{31} = 16, X_{33} = 21,\ X_{35} = 26$	$\bar{\mathbf{x}}_3 = \frac{\sum x}{n_3} = \frac{63}{3} = 21.0$

Now we find sample mean

$$\bar{x} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}}{n} = \frac{6+7+10+16+20+16+21+26}{8} = \frac{122}{8} = 15.25$$

To estimate the populations mean

$$\bar{X}_{st} = \sum_{i=1}^{k} W_i \bar{X}_i = \frac{\sum_{i=1}^{k} N_i \bar{x}_i}{N}$$

$$\bar{X}_{st} = \frac{N_1\bar{x}_1}{N} + \frac{N_2\bar{x}_2}{N} + \frac{N_3\bar{x}_3}{N} = \frac{4(6.5)}{16} + \frac{6(15.33)}{16} + \frac{6(21)}{16} = \frac{243.98}{16} = 15.248 = 15.25$$

Q.22 (a): At a small private college, the students are classified as follows

Classification	B.Sc	B.A	F.Sc	F.A
No. of students	150	163	195	220

If we wish to select a stratified random sample of size n=40 by proportional allocation how large a sample must we take from each stratum?

Solution: Now we find the sample size in each stratum to use proportional allocation

$n_h = n\frac{N_h}{N}$

Classification	B.Sc	B.A	F.Sc	F.A	
No. of students	150	163	195	220	728
$n_h = n\frac{N_h}{N}$	$40\frac{150}{728} = 8$	9	11	12	40

Q.22 (b): A large company has 300,000 employees the age distribution of whom is shown as follows

Age (Years)	Percentages
25 or Younger	15
26-35	30
36-45	25
46-55	20
56 or older	10

A sample of 2 percent of all the employees is desired. Design a sampling plan such that each age-group is proportionally represented.

Solution: Now we find the sample size in each stratum to use proportional allocation

$n_h = n\frac{N_h}{N}$

Age (Years)	N_h	$n_h = n\frac{N_h}{N}$
25 or Younger	300,000*0.15=45000	900
26-35	90000	1800
36-45	75000	1500
46-55	60000	1200
56 or older	30000	600

Sample of size 2% of 300000=6000=n

Sampling distribution of sample means

The arrangement of all possible values of sample means with their probabilities is called sampling distribution of sample means.

Properties

i) $E(\overline{X}) = \mu_{\overline{X}} = \mu$ In case of with and without replacement

ii) $\sigma^2_{\overline{X}} = Var(\overline{X}) = \frac{\sigma^2}{n}$ In case of with replacement

Or

$\sigma_{\overline{X}} = S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}}$

iii) $\sigma^2_{\overline{X}} = Var(\overline{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$ In case of without replacement

Or

$\sigma_{\overline{X}} = S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N-1}\right)}$

iv) Shape of sampling distribution

i) If population sampled is normal then sampling distribution of means will be normal either sample size is large or small.

ii) If population sampled is non-normal then sampling distribution of means will be approximately normal only for large sample.

Example: A finite population consists of the numbers 2, 4, 6 and 8. Calculate the sample for all possible random samples of size "n=2" that can be drawn from this population, with replacement. Assuming the 16 possible samples equally likely, make the sampling distribution of sample means and find the mean and variance of this distribution. Calculate mean and variance of population and verify the results.

i) $\mu_{\overline{X}} = \mu$ ii) $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$

Solution:

$Pop: 2,4,6,8$

$N = 4$

$n = 2$

Draw all possible sample of size "2" by with replacement $N^n = 4^2 = 16$

Sr.No	Samples	$\overline{X} = \frac{\sum x}{n}$
1	2,2	2
2	2,4	3
3	2,6	4
4	2,8	5
5	4,2	6
6	4,4	4
7	4,6	5
8	4,8	6
9	6,2	4
10	6,4	5
11	6,6	6
12	6,8	7
13	8,2	5
14	8,4	6
15	8,6	7
16	8,8	8

Sampling distribution of sample means ($\bar{X}$)

$\bar{X}$	Tally	F	$f(\bar{X})$	$\bar{X}\ f(\bar{X})$	$\bar{X}^2\ f(\bar{X})$
2	1	1	1/16	2/16	4/16
3	11	2	2/16	6/16	18/16
4	111	3	3/16	12/16	48/16
5	1111	4	4/16	20/16	100/16
6	111	3	3/16	18/16	108/16
7	11	2	2/16	14/16	98/16
8	1	1	1/16	8/16	64/16
Total		16	1	80/16	440/16

$$\mu_{\bar{X}} = E(\bar{X}) = \bar{X}f(\bar{X}) = 80/16 = 5$$

$$\sigma^2_{\bar{X}} = E(\bar{X}^2) - [E(\bar{X})]^2$$

$$Var(\bar{X}) = \sum \bar{X}^2 f(\bar{X}) - \left[\sum \bar{X}f(\bar{X})\right]^2$$

$$\sigma^2_{\bar{X}} = Var(\bar{X}) = 440/16 - [80/16]^2 = 27.5 - 25 = 2.5$$

Population distribution

X	X^2
2	4
4	16
6	36
8	64
$\sum X = 20$	$\sum X^2 = 120$

$$\mu = \frac{\sum X}{N} = \frac{20}{4} = 5$$

$$\sigma^2 = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2 = \frac{120}{4} - \left(\frac{20}{4}\right)^2 = 5$$

Verification

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

$5 = 5$ Hence verified

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

$$\sigma^2_{\bar{X}} = \frac{5}{2} = 2.5$$

$2.5 = 2.5$ Hence verified

Example: A finite population consists of the numbers 2, 4, 6 and 8. Calculate the sample for all possible random samples of size “n=2” that can be drawn from this population, without replacement and verify the results.

i) $\mu_{\bar{X}} = \mu$ ii) $\sigma^2_{\bar{X}} = \dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right)$

Solution:

$Pop: 2,4,6,8$

$N = 4$

$n = 2$

Draw all possible sample of size “2” by without replacement $= {}^N C_n = {}^4 C_2 = 6$

Sr.No	Samples	$\bar{X} = \frac{\sum x}{n}$
1	2,4	3
2	2,6	4
3	2,8	5
4	4,6	5
5	4,8	6
6	6,8	7

Sampling distribution of sample means ($\bar{X}$)

$\bar{X}$	Tally	F	$f(\bar{X})$	$\bar{X}\ f(\bar{X})$	$\bar{X}^2\ f(\bar{X})$
3	1	1	1/6	3/6	9/6
4	1	1	1/6	4/6	16/6
5	11	2	2/6	10/6	50/6
6	1	1	1/6	6/6	36/6
7	1	1	1/6	7/6	49/6
Total		6	1	30/6	155/6

$\mu_{\bar{X}} = E(\bar{X}) = \bar{X}f(\bar{X}) = 30/6 = 5$

$\sigma^2_{\bar{X}} = E(\bar{X}^2) - [E(\bar{X})]^2$

$Var(\bar{X}) = \sum \bar{X}^2 f(\bar{X}) - [\sum \bar{X}f(\bar{X})]^2$

$\sigma^2_{\bar{X}} = Var(\bar{X}) = 160/6 - [30/6]^2 = 26.667 - 25 = 1.667$

Population distribution

X	X^2
2	4
4	16
6	36
8	64
$\sum X = 20$	$\sum X^2 = 120$

$$\mu = \frac{\sum X}{N} = \frac{20}{4} = 5$$

$$\sigma^2 = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2 = \frac{120}{4} - \left(\frac{20}{4}\right)^2 = 5$$

Verification

$\mu_{\bar{X}} = \mu$ $\quad 5 = 5$ Hence verified

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

$$\sigma^2_{\bar{X}} = \frac{5}{2}\left(\frac{4-2}{4-1}\right) = 1.67$$

$1.667 = 1.667$ Hence verified

Note that: Same As

Example: 14.7

Example: 14.8

Example. 14.9

Q.26 (b) to Q.32 Do yourself

Q.33: A population consists of four numbers 2,4,6,8. Draw all possible sample of size $n = 3$ with replacement. Find the mean and the median for each sample. From the sampling distribution of means and the sampling distribution of medians. Which of these distributions has the smaller variance? How did the means of these two distributions compare with the population mean?

Solution

Q.34: Given the following population distribution

x	1	2	3	4
F(x)	1/7	3/7	2/7	1/7

Find the sampling distribution of the mean if a sample of three numbers is taken without replacement. How does the variance of the sampling distribution compare with the population variance?

Solution: First we find population mean and standard deviation from the given distribution

X	$f(X)$	$Xf(X)$	$X^2 f(X)$
1	1/7	1/7	1/7
2	3/7	6/7	12/7
3	2/7	6/7	18/7
4	1/7	4/7	16/7
Total	1.0	17/7	47/7

$\mu = \sum Xf(X) = 17/7 = 2.43$

$\sigma^2 = \sum X^2 f(X) - (\sum Xf(X))^2 = 47/7 - (2.43)^2 = 0.815$

Population: 1,2,2,2,3,3,4 N=7 and n=3

Draw all possible sample of size “3” by without replacement ${}^{N}C_n = {}^{7}C_3 = 35$

Sr.No	Samples	$\overline{X}=\frac{\sum x}{n}$	Sr.No	Samples	$\overline{X}=\frac{\sum x}{n}$
1	1,2,2	5/3	19	2,2,4	8/3
2	1,2,2	5/3	20	2,2,3	7/3
3	1,2,3	6/3	21	2,2,3	7/3
4	1,2,3	6/3	22	2,2,4	8/3
5	1,2,4	7/3	23	2,3,3	8/3
6	1,2,2	5/3	24	2,3,4	9/3
7	1,2,3	6/3	25	2,3,4	9/3
8	1,2,3	6/3	26	2,2,3	7/3
9	1,2,4	7/3	27	2,2,3	7/3
10	1,2,3	6/3	28	2,2,4	8/3
11	1,2,3	6/3	29	2,3,3	8/3
12	1,2,4	7/3	30	2,3,4	9/3
13	1,3,3	7/3	31	2,3,4	9/3
14	1,3,4	8/3	32	2,3,3	8/3
15	1,3,4	8/3	33	2,3,4	9/3
16	2,2,2	6/3	34	2,3,4	9/3
17	2,2,3	7/3	35	3,3,4	10/3
18	2,2,3	7/3			

Sampling distribution of sample means $\overline{X}$

$\overline{X}$	Tally	$\overline{X}$	$f(X)$	$Xf(X)$	$X^2 f(X)$
5/3	*III*	3	3/35	15/105	75/315
6/6	*IIIIII*	7	7/35	42/105	252/315
7/3	*IIIIIIII*	10	10/35	70/105	490/315
8/3	*IIIIII*	8	8/35	64/105	512/315
9/3	*IIIII*	6	6/35	54/105	486/315
10/3	*I*	1	1/35	10/105	100/315
		Total	35/35=1	225/105	1915/315

$$\mu_{\overline{X}}=\sum \overline{X}f(\overline{X})=225/105=2.43$$

$$\sigma^2_{\overline{X}}=\sum \overline{X}^2 f(\overline{X})-\left(\sum \overline{X}f(\overline{X})\right)^2=1915/315-(2.43)^2=0.181$$

Verification

$$\mu_{\overline{X}}=\mu$$

$2.43=2.43$ Hence Proved

$$\sigma^2_{\overline{X}}=\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)=\frac{0.815}{3}\left(\frac{4}{6}\right)=0.181$$ Hence Proved

Example.14.10: The weights of 1500 ball bearing are normally distributed with mean of 22.40 ounces and a standard deviation of 0.048 ounces. If 300 random samples of size 36 are drawn from this population

a) Determine the expected mean and standard deviation of the sampling distribution of mean if sampling is done i) with replacement ii) Without replacement

Solution: $N=1500$ $n=36$ $\mu=22.4$ $\sigma=0.048$

i) with replacement

As we know that

$$\mu_{\overline{X}}=\mu=22.4$$

$$\sigma_{\overline{X}}=\sqrt{\frac{\sigma^2}{n}}=\sqrt{\frac{(0.048)^2}{36}}=0.008$$

ii) Without replacement

As we know that

$$\mu_{\overline{X}}=\mu=22.4$$

$$\sigma_{\overline{X}}=\sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}$$

First we check by $\left(\frac{n}{N}\times 100\right)$ if it is less than 5% or not if yes than we drop FPC otherwise we use FPC. So, $\left(\frac{36}{1500}\times 100\right)=2.4\%$ it is less than then we drop the factor $\left(\frac{N-n}{N-1}\right)$ and we use

$$\sigma_{\overline{X}}=\sqrt{\frac{\sigma^2}{n}}=0.008$$

b) How many of the random samples would have their means between 22.39 and 22.42 oz?

Solution: Given that $\mu_{\overline{X}}=22.4$ $\sigma_{\overline{X}}=0.008$

$P(22.39<\overline{X}<22.42)=?$

By Standard normal variate (S.N.V)

$$Z=\frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{x}}}$$

$$Z_1=\frac{22.39-22.4}{0.008}=-1.25$$

$$Z_2=\frac{22.39-22.4}{0.008}=2.50$$

$P(22.39<\overline{X}<22.42)=P(-1.50\le Z\le 2.50)=P(-1.50\le Z\le 0)+P(0\le Z\le 2.50)=0.3944+0.4938=0.8882$

Expected number of samples between 22.39 and 22.42=300*0.8882=267

Example.14.11: A construction company has 310 employees who have an average annual salary of Rs.24, 000. The standard deviation of annual salaries is Rs.5, 000. In a random sample of 100 employees, what is the probability that the average salary will exceed Rs.24, 500?

Solution: Given that $N=310$ $n=100$ $\mu=24000$ $\sigma=5000$

As we know that

$\mu_{\overline{X}}=\mu=24000$

Here the sample size greater than 5% so we use FPC

$$\sigma_{\overline{X}}=\sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}=412.20$$

$P(\overline{X}\ge 24500)=?$

By Standard normal variate (S.N.V)

$$Z=\frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{x}}}$$

$$Z=\frac{24500-24000}{412.20}=1.21$$

$P(\overline{X}\ge 24500)=P(Z\ge 1.21)=P(0\le Z\le\infty)-P(0\le Z\le 1.21)=0.5-0.3869=0.1131$

Q.35: A random sample of size $n=100$ is taken from a population having a mean of 20 and a standard deviation of 5.0. The shape of the population distribution is unknown.

a) What can you say about the sampling distribution of the sample mean $\overline{X}$?

b) Find the probability that $\overline{X}$ will be exceeds 20.75.

Solution:

a) As sample size is large then by central limit theorem then the sampling distribution of sample mean becomes approximately normal.

b) Given that $n=100$ $\mu=20$ $\sigma=5.0$

As we know that

$\mu_{\overline{X}}=\mu=20$ $\sigma_{\overline{X}}=\sqrt{\frac{\sigma^2}{n}}=\sqrt{\frac{(5)^2}{100}}=0.50$

$P(\overline{X}\ge 20.75)=?$

By Standard normal variate (S.N.V)

$$Z=\frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{x}}}$$

$$Z = \frac{20.75-20}{0.50} = 1.50$$

$P(\bar{X} \geq 1.50) = P(Z \geq 1.50) = P(0 \leq Z \leq \infty) - P(0 \leq Z \leq 1.50) = 0.5 - 0.4332 = 0.0668$

Q.36: In a local agriculture reporting area the average wheat of yields is known to be 60 bushels per acre with a standard deviation of 10 bushels. If a random sample of 64 acres is selected and the wheat yields recorded, what is the probability that the sample mean will lie between 59 and 61 bushels?

Solution: Given that $n = 64$ $\mu = 60$ $\sigma = 10$

As we know that

$\mu_{\bar{X}} = \mu = 60$

Sampling is done with or without replacement first we check by $\left(\frac{n}{N}\times 100\right)$ if it is less than 5% than we drop FPC otherwise we use FPC. So, $\left(\frac{25}{1000}\times 100\right) = 2.5\%$ it is less than then we use

$$\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{(10)^2}{64}} = 1.25$$

$P(59 \leq \bar{X} \leq 61) = ?$

By Standard normal variate (S.N.V)

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{x}}}$$

$$Z_1 = \frac{59-60}{1.25} = -0.80$$

$$Z_2 = \frac{60-60}{1.25} = 0.80$$

$P(59 \leq \bar{X} \leq 61) = P(-0.80 \leq Z \leq 0.80) = P(-0.80 \leq Z \leq 0) + P(0 \leq Z \leq 0.80) = 0.2881 + 0.2881 = 0.5762$

Q.37: The heights of 1000 students are approximately normally distributed with a mean of 68.5 inches and a standard deviation of 2.7 inches. If 200 random samples size 25 are drawn from this population and the means recorded to the nearest tenth of an inch, determine

a) The expected mean and standard deviation of the sampling distribution of the mean

b) The number of sample means that fall between 67.9 and 69.2 inclusive.

Solution: Given that $N = 1000$ $n = 25$ $\mu = 68.5$ $\sigma = 2.7$

a) As we know that

$\mu_{\bar{X}} = \mu = 68.5$

Sampling is done with or without replacement first we check by $\left(\frac{n}{N}\times 100\right)$ if it is less than 5% than we drop FPC otherwise we use FPC. So, $\left(\frac{25}{1000}\times 100\right) = 2.5\%$ it is less than then we use

$$\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{(2.7)^2}{25}} = 0.54$$

b) $P(67.9 \leq \bar{X} \leq 69.2) = ?$

By Standard normal variate (S.N.V)

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{x}}}$$

$$Z_1 = \frac{67.9-68.5}{0.54} = -1.11$$

$$Z_2 = \frac{69.2-68.5}{0.54} = 1.30$$

$P(-1.11 \leq \bar{X} \leq 1.30) = P(-1.11 \leq Z \leq 1.30) = P(-1.11 \leq Z \leq 0) + P(0 \leq Z \leq 1.30) = 0.3665 + 0.4032 = 0.7697$

So, the required no. of samples 200*0.7697=154

Q.38: The heights of a large number of shrubs of the same kind produced for sale by a horticultural nursery are normally distributed with mean 1.14m and standard deviation 0.25m. Fifty samples each consisting of 100 shrubs are selected. In how many of these samples would expect to find the mean samplers being to be

i) Greater than 1.16m ii) Between 1.13m and 1.18m

Solution: Given that $n=100$ $\mu=1.14$ $\sigma=0.25$

As we know that

$\mu_{\overline{X}}=\mu=1.14$

$$\sigma_{\overline{X}}=\sqrt{\frac{\sigma^2}{n}}=\sqrt{\frac{(0.25)^2}{100}}=0.025$$

i) $P(\overline{X}_1>1.16)=?$

By Standard normal variate (S.N.V)

$$Z=\frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{x}}}$$

$$Z=\frac{1.16-1.14}{0.025}=0.80$$

$P(\overline{X}>1.16)=P(Z>0.80)=P(0<Z<\infty)-P(0<Z<0.80)=0.5-0.2881=0.2119$

The required no. of these samples would expect to find the mean samplers being to be greater than 1.16m is 50*0.2119=11

ii) $P(1.13<\overline{X}<1.18)=?$

By Standard normal variate (S.N.V)

$$Z=\frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{x}}}$$

$$Z_1=\frac{1.13-1.14}{0.025}=-0.40$$

$$Z_2=\frac{1.18-1.14}{0.025}=1.60$$

$P(1.13<\overline{X}<1.18)=P(-0.40<Z<1.60)=P(-0.40\le Z<0)+P(0\le Z<1.60)=0.1554+0.4452=0.6006$ T

he required no. of these samples would expect to find the mean samplers being to be between 1.13m and 1.18m is 50*0.6006=30.0

Q.39 (a): The following table shows the distribution of 14-year old schoolboy intelligence test markings

I.Q	80-89	90-99	100-109	110-119	120-129	130-139	140-149
Numbers	30	52	75	109	65	42	27

On the assumption that this group is a random sample estimate the standard error of the mean

Solution:

X	f	fX	fX^2
84.5	30		
94.5	52		
104.5	75		
11405	109		
124.5	65		
134.5	42		
144.5	27		

$$\mu=\frac{\sum fX}{\sum f}=-=113.53$$

$\mu_{\overline{X}}=\mu=1.14$

Here σ is unknown and sample size is large so, we find

$$S^2=\frac{\sum fX^2}{\sum f}-\left(\frac{\sum fX}{\sum f}\right)^2=16.01$$

As we know that

$\mu_{\overline{X}}=\mu=113.53$ $\sigma_{\overline{X}}=\sqrt{\frac{S^2}{n}}=\sqrt{\frac{16.01}{400}}=0.80$

Example.14.12: Same as Q.39 (a)

Q.39 (b): The random variable "X" has the following probability distribution

x	4	5	6	7
P(x)	0.2	0.4	0.3	0.1

i) Find the mean $\mu_{\overline{X}}$ and variance $\sigma^2_{\overline{X}}$ of the $\overline{X}$ for a random sample of 36

ii) Find the probability that the mean of 36 items will be less than 5.5

Solution:

x	P(x)	$xP(x)$	$x^2P(x)$
4	0.2	0.8	3.2
5	0.4	2.0	10.0
6	0.3	1.8	10.8
7	0.1	0.7	4.9

$$\mu = \sum xP(x) = 5.3$$

$$\sigma^2 = \sum x^2P(x) - \left(\sum xP(x)\right)^2 = 28.9 - (5.3)^2 = 0.81$$

i) As we know that

$$\mu_{\overline{X}} = \mu = 5.3$$

$$\sigma_{\overline{X}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{0.81}{36}} = 0.15$$

ii) $P(\overline{X} < 5.5) = ?$

By Standard normal variate (S.N.V)

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}}$$

$$Z = \frac{5.5 - 5.3}{0.15} = 1.33$$

$$P(\overline{X} < 5.5) = P(Z < 1.33) = P(-\infty < Z < 0) + P(0 < Z < 1.33) = 0.50 + 0.4082 = 0.9082$$

Example.14.13: Given the population 1,1,1,3,4,5,6,6,6,7

a) Find the probability that a random sample of size 36 selected with replacement will yield a sample mean between 3.26 and 4.74

b) Find the mean and standard deviation for the sampling distribution of means for a sample of size 4 selected at random without replacement, between what two values you expect at least $\frac{3}{4}$ of the sample means to fall?

Solution: Given that

Population: 1,1,1,3,4,5,6,6,6,7

N=10

$$\mu = \frac{\sum X}{N} = \frac{40}{10} = 4.0$$

$$\sigma = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2} = 2.24$$

a) As we know that

$$\mu_{\overline{X}} = \mu = 4$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2.24}{\sqrt{36}} = 0.373$$

$P(3.26 < \overline{X} < 4.74) = ?$

By Standard normal variate (S.N.V)

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}}$$

$$Z_1 = \frac{3.26 - 4.0}{0.373} = -1.98$$

$$Z_2 = \frac{4.74 - 4.0}{0.373} = 1.98$$

$$P(3.26 < \overline{X} < 4.74) = P(-1.98 < Z < 1.98) = P(-1.98 \le Z < 0) + P(0 \le Z < 1.98) = 0.4762 + 0.4762 = 0.9524$$

b) n=4 And N=10

As we know that

$\mu_{\bar{X}} = \mu = 4$

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\left(\frac{N-n}{N-1}\right) = \frac{2.24}{\sqrt{364}}\left(\frac{6}{9}\right) = 0.912$

The Chebyshev's inequality says "at least $\left(1-\frac{1}{k^2}\right)$ fraction of the data lies in the interval $(Mean \pm k\,S.D)$ but in this question says at least ¾ so, we get

$\left(1-\frac{1}{k^2}\right) = \frac{3}{4}$

$\frac{1}{k^2} = 1 - \frac{3}{4} = \frac{1}{4}$

$k = 2$

Hence we expect at least ¾ of the sample means to fall in the interval $\mu_{\bar{X}} \pm k\,\sigma_{\bar{X}}$ that is (2.2, 5.8).

Example.14.14: A random sample of size 25 is selected from a Poisson distribution with $\mu = 3$. Find, using central limit theorem, the probability that the sample mean will be greater than 4.

Solution: As we know that in Poisson distribution mean and variance are same so, we get

$\mu = 3 \qquad \sigma^2 = 3 \qquad n = 25$

By central limit theorem it becomes approximately normal distribution

$\mu_{\bar{X}} = \mu = 3.0$

$\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{3}{25}} = 0.35$

$P(\bar{X}_1 > 4) = ?$

By Standard normal variate (S.N.V)

$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$

$Z = \frac{4-3}{0.35} = 2.86$

$P(\bar{X} > 4) = P(Z > 2.86) = P(0 < Z < \infty) - P(0 < Z < 2.86) = 0.5 - 0.4979 = 0.0021$

Q.40 (a): The mean of a certain normal distribution is equal to S.E (Standard error) of mean of samples of 100 from that distribution. Find the probability that the mean of a sample of 25 from the distribution will be negative.

Solution: Given that

$\mu = S.E = \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{100}} = \frac{\sigma}{10}$ At $n = 100$

As we know that

$\mu_{\bar{X}} = \mu = \frac{\sigma}{10}$

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{25}} = \frac{\sigma}{5}$

$P(Sample\ mean\ will\ be\ negative) = P(\bar{X} < 0) = ?$

By Standard normal variate (S.N.V)

$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$

$Z = \frac{0 - \frac{\sigma}{10}}{\frac{\sigma}{5}} = -\frac{5}{10} = -0.50$

$P(\bar{X} < 0) = P(Z < -0.50) = P(-\infty < Z < 0) - P(-0.50 < Z < 0) = 0.5 - 0.1915 = 0.3085$

Q.40 (b): A normal population has a mean of 0.1 and a standard deviation of 2.1. Find the probability that the mean of a simple random sample of 900 members will be negative.

Solution: Given that

$\mu = 0.10$ $\sigma = 2.1$ $n = 900$

As we know that

$\mu_{\bar{X}} = \mu = 0.10$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2.1}{\sqrt{900}} = 0.07$$

$P(Sample\ mean\ will\ be\ negative) = P(\bar{X} < 0) = ?$

By Standard normal variate (S.N.V)

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{x}}}$$

$$Z = \frac{0 - 0.10}{0.07} = -1.43$$

$P(\bar{X} < 0) = P(Z < -1.43) = P(-\infty < Z < 0) - P(-1.43 < Z < 0) = 0.5 - 0.4236 = 0.0.0764$

Q.41 (a): A random sample of size 100 is taken from a binomial distribution with parameters $P = 0.5$ and $n = 40$. Find, using the central limit theorem the approximate probability that $\bar{X}$ is

i) Greater than 20.5 ii) Less than 19.3 and iii) between 19.3 and 20.5.

Solution: Given that $P = 0.5$ $n_1 = 40$

As we know that in binomial distribution

$\mu = nP = 0.5(40) = 20.0$

$\sigma = \sqrt{npq} = \sqrt{40(0.5)(0.5)} = 3.162$

Then

$\mu_{\bar{X}} = \mu = 20$

$$\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{(3.162)^2}{40}} = 0.50$$

i) $P(\bar{X} > 20.5) = ?$

By Standard normal variate (S.N.V)

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{x}}}$$

$$Z = \frac{20.5 - 20}{0.5} = 1.0$$

$P(\bar{X} > 20.5) = P(Z > 1.0) = P(0 < Z < \infty) - P(0 < Z < 1.0) = 0.50 - 0.3413 = 0.1587$

ii) $P(\bar{X} < 19.3) = ?$

By Standard normal variate (S.N.V)

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{x}}}$$

$$Z = \frac{19.3 - 20}{0.5} = -1.40$$

$P(\bar{X} < 19.3) = P(Z < -1.40) = P(-\infty < Z < 0) - P(-1.40 < Z < 0) = 0.50 - 0.4192 = 0.0808$

iii) $P(19.3 < \bar{X} < 20.5) = ?$

By Standard normal variate (S.N.V)

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{x}}}$$

$$Z_1 = \frac{19.3 - 20}{0.5} = -1.40$$

$$Z_2 = \frac{20.5 - 20}{0.5} = 1.0$$

$P(19.3 < \bar{X} < 20.5) = P(-1.40 < Z < 1.0) = P(-1.40 \le Z < 0) + P(0 \le Z < 1.0) = 0.4192 + 0.3413 = 0.7605$

Q.41 (b): A sample of 36 cases is drawn from a negatively skewed population with mean of 2 and a standard deviation of 3. What is the probability that the sample mean obtained will be negative? How many points must we go from the mean to include 50 percent of all sample means?

Solution: Given that $n = 36$ $\mu = 2$ $\sigma = 3$

As we know that

$\mu_{\bar{X}} = \mu = 2.0$

$$\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{9}{36}} = 0.50$$

$P(Sample\ mean\ will\ be\ negative) = P(\bar{X} < 0) = ?$

By Standard normal variate (S.N.V)

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{x}}}$$

$$Z = \frac{0 - 2.0}{0.5} = -24.0$$

$P(\bar{X} < 0) = P(Z < -4.0) = P(-\infty < Z < 0) - P(-4 < Z < 0) = 0.5 - 0.5 = 0$

Now 50% Area of normal distribution lies between $\bar{X} \pm 0.6745\sigma_{\bar{X}}$

So, the interval between Samples mean.

$\bar{X} \pm 0.6745(0.5)$

$\bar{X} \pm 0.33725$

Mean must go the interval in 50% observation ± 0.33725

Sampling distribution of difference between two sample means $(\bar{X}_1 - \bar{X}_2)$

The arrangement of all possible difference between two sample means $(\bar{X}_1 - \bar{X}_2)$ with their probabilities is called sampling distribution of difference between two sample means $(\bar{X}_1 - \bar{X}_2)$.

Properties

i) $E(\bar{X}_1 - \bar{X}_2) = \mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$ In case of with and without replacement

ii) $\sigma^2_{\bar{X}_1 - \bar{X}_2} = Var(\bar{X}_1 - \bar{X}_2) = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$ In case of with replacement

Or

$$\sigma_{\bar{X}_1 - \bar{X}_2} = S.E(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

iii) In case of without replacement

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1}\left(\frac{N_1 - n_1}{N_1 - 1}\right) + \frac{\sigma_2^2}{n_2}\left(\frac{N_2 - n_2}{N_2 - 1}\right)$$

Or

$$\sigma_{\bar{X}_1 - \bar{X}_2} = S.E(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1}\left(\frac{N_1 - n_1}{N_1 - 1}\right) + \frac{\sigma_2^2}{n_2}\left(\frac{N_2 - n_2}{N_2 - 1}\right)}$$

iv) Shape of sampling distribution

i) If population sampled is normal then sampling distribution of $(\bar{X}_1 - \bar{X}_2)$ will be normal either sample size is large or small.

ii) If population sampled is non-normal then sampling distribution of $(\bar{X}_1 - \bar{X}_2)$ will be approximately normal only for large sample.

Example: Draw all possible random sample of size $n_1 = 2$ with replacement from the population "4 and 6". Similarly, draw all possible samples of size $n_2 = 2$ with replacement from the population "3 and 4". Construct the sampling distribution of difference between two sample means $(\bar{X}_1 - \bar{X}_2)$ and show that

i) $E(\bar{X}_1 - \bar{X}_2) = \mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$ ii) $\sigma^2_{\bar{X}_1 - \bar{X}_2} = Var(\bar{X}_1 - \bar{X}_2) = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$

Solution: 1st population

$Pop: 4,6$

$N_1 = 2$

$n_1 = 2$

Draw all possible sample of size "2" by with replacement $= N^n = (2)^2 = 4$

Sr.No	Samples	$\bar{X}_1 = \frac{\sum x}{n}$
1	4,4	4
2	4,6	5
3	6,4	5
4	6,6	6

2nd population

$Pop: 3,4$

$N_2 = 2$

$n_2 = 2$

Draw all possible sample of size "2" by with replacement $= N^n = (2)^2 = 4$

Sr.No	Samples	$\bar{X}_2 = \frac{\sum x}{n}$
1	3,3	3
2	3,4	3.5
3	4,3	3.5
4	4,4	4

Possible difference between two sample means $(\bar{X}_1 - \bar{X}_2)$

	$\bar{X}_2$			
$\bar{X}_1$	3	3.5	3.5	4
4	1	0.5	0.5	0
5	2	1.5	1.5	1
5	2	1.5	1.5	1
6	3	2.5	2.5	2

Sampling distribution of difference between two sample means $(\bar{X}_1 - \bar{X}_2)$

$\bar{X}_1 - \bar{X}_2$	Tally	F	$f(\bar{X}_1 - \bar{X}_2)$	$(\bar{X}_1 - \bar{X}_2) f(\bar{X}_1 - \bar{X}_2)$	$(\bar{X}_1 - \bar{X}_2)^2 f(\bar{X}_1 - \bar{X}_2)$
0	1	1	1/16	0/16	0/16
0.5	11	2	2/16	1/16	0.5/16
1	111	3	3/16	3/16	3/16
1.5	1111	4	4/16	6/16	9/16
2	111	3	3/16	6/16	12/16
2.5	11	2	2/16	5/16	12.5/16
3	1	1	1/16	3/16	9/16
Total		16	1	24/16	46/16

$$\mu_{\bar{X}_1 - \bar{X}_2} = E(\bar{X}_1 - \bar{X}_2) = (\bar{X}_1 - \bar{X}_2) f(\bar{X}_1 - \bar{X}_2) = 24/16 = 1.5$$

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = E(\bar{X}_1 - \bar{X}_2)^2 - \left[E(\bar{X}_1 - \bar{X}_2)\right]^2$$

$$Var(\bar{X}) = \sum (\bar{X}_1 - \bar{X}_2)^2 f(\bar{X}_1 - \bar{X}_2) - \left[\sum (\bar{X}_1 - \bar{X}_2) f(\bar{X}_1 - \bar{X}_2)\right]^2$$

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = 46/16 - [24/16]^2 = 2.875 - 2.25 = 0.625$$

Population distribution

1st population

X_1	X_1^2
4	16
6	36
$\sum X_1 = 10$	$\sum X_1^2 = 52$

$$\mu_1 = \frac{\sum X_1}{N_1} = \frac{10}{2} = 5$$

$$\sigma_1^2 = \frac{\sum X_1^2}{N_1} - \left(\frac{\sum X_1}{N_1}\right)^2 = \frac{52}{2} - \left(\frac{10}{2}\right)^2 = 1$$

2nd Population

X_2	X_2^2
3	9
4	16
$\sum X_1 = 7$	$\sum X_2^2 = 25$

$$\mu_2 = \frac{\sum X_2}{N_2} = \frac{7}{2} = 3.5$$

$$\sigma_2^2 = \frac{\sum X_2^2}{N_2} - \left(\frac{\sum X_2}{N_2}\right)^2 = \frac{25}{2} - \left(\frac{7}{2}\right)^2 = 0.25$$

Verification

$E(\bar{X}_1 - \bar{X}_2) = \mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2$

$E(\bar{X}_1 - \bar{X}_2) = 5 - 3.5 = 1.5$

$1.5 = 1.5$ Hence verified

$\sigma^2_{\bar{X}_1-\bar{X}_2} = Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

$\sigma^2_{\bar{X}_1-\bar{X}_2} = \frac{1}{2} + \frac{0.25}{2}$

$0.625 = 0.625$ Hence verified

Note: Same as Example14.15 and Q.43

Q.42 (b): Random sample of size 100 are drawn with replacement from two populations and their means $\bar{X}_1$ *and* $\bar{X}_2$ computed. If $\mu_1 = 10, \sigma_1 = 2, \mu_2 = 8$ *and* $\sigma_2 = 1$, find the probability that the difference between a given pair of sample means is

i) Less than 1.5 ii) Greater than 1.75 but less than 2.5

Solution: Given that $n_1 = 100$ $\mu_1 = 10$ $\sigma_1 = 2.0$ $n_2 = 100$ $\mu_2 = 8.0$ $\sigma_2 = 1$

As we know that

$\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2 = 10 - 8 = 2.0$

$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{4}{100} + \frac{1}{100}} = 0.224$

i) $P(\bar{X}_1 - \bar{X}_2 < 1.5) = ?$

By Standard normal variate (S.N.V)

$Z = \frac{\bar{X}_1-\bar{X}_2 - \mu_{\bar{X}_1-\bar{X}_2}}{\sigma_{\bar{X}_1-\bar{X}_2}}$

$Z = \frac{1.5 - 2.0}{0.224} = -2.23$

$P(\bar{X}_1 - \bar{X}_2 < 1.5) = P(Z < -2.23) = P(-\infty < Z < 0) - P(-2.23 < Z < 0) = 0.5 - 0.5 - 0.4871 = 0.0129$

ii) $P(1.75 < \bar{X}_1 - \bar{X}_2 < 2.5) = ?$

By Standard normal variate (S.N.V)

$Z = \frac{\bar{X}_1-\bar{X}_2 - \mu_{\bar{X}_1-\bar{X}_2}}{\sigma_{\bar{X}_1-\bar{X}_2}}$

$Z_1 = \frac{1.75 - 2}{0.224} = -1.12$

$Z_2 = \frac{2.5 - 2}{0.224} = 2.23$

$P(1.75 < \bar{X}_1 - \bar{X}_2 < 2.5) = P(-1.12 < Z < 2.23) = P(-1.12 \le Z < 0) + P(0 \le Z < 2.23) = 0.3686 + 0.4871 = 0.8557$

Q.44: The Television picture tubes of manufacturer A have a mean lifetime of 6.5 years and standard deviation of 0.9 while those of manufacturer B have a mean lifetime of 6.0 years and a standard deviation of 0.8 years. What is the probability that a random sample of 36 tubes from manufacturer A will have a mean lifetime that is at least year 1 year more than the mean lifetime of a sample of 49 tubs from manufacturer B?

Solution: Given that $n_A = 36$ $\mu_A = 6.5$ $\sigma_A = 0.90$ $n_B = 49$ $\mu_B = 6.0$ $\sigma_B = 0.80$

As we know that

$\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2 = 6.5 - 6.0 = 0.5$

$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{(0.9)^2}{36} + \frac{(0.8)^2}{49}} = 0.189$

$P(\bar{X}_1 - \bar{X}_2 \ge 1.0) = ?$

By Standard normal variate (S.N.V)

$Z = \frac{\bar{X}_1-\bar{X}_2 - \mu_{\bar{X}_1-\bar{X}_2}}{\sigma_{\bar{X}_1-\bar{X}_2}}$ $Z = \frac{1.0 - 0.5}{0.189} = 2.65$

$P(\bar{X}_1 - \bar{X}_2 \ge 1.0) = P(Z \ge 2.65) = P(0 \le Z \le \infty) - P(0 \le Z < 2.65) = 0.5 - 0.4960 = 0.004$

Q.45: A random sample of size 25 is taken from a normal population having a mean of 80 and a standard deviation of 5. A second random sample of size 36 is taken from a different normal population having a mean of 75 and a standard deviation of 3.0. Find the probability that the sample mean computed from the 25 measurements will exceed the sample mean computed from the 36 measurements by at least 3.4 but less than 5.9. Assume the means to be measured nearest 10th.

Solution: Given that $n_1 = 25 \quad \mu_1 = 80 \quad \sigma_1 = 5 \; n_2 = 36 \quad \mu_2 = 75 \quad \sigma_2 = 3$

As we know that

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2 = 80 - 75 = 5$$

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{25}{25} + \frac{9}{36}} = 1.12$$

$$P(3.4 \le \bar{X}_1 - \bar{X}_2 < 5.9) = ?$$

By Standard normal variate (S.N.V)

$$Z = \frac{\bar{X}_1-\bar{X}_2 - \mu_{\bar{X}_1-\bar{X}_2}}{\sigma_{\bar{X}_1-\bar{X}_2}}$$

$$Z_1 = \frac{3.4-5}{1.12} = -1.43$$

$$Z_2 = \frac{5.9-5}{1.12} = 0.80$$

$$P(3.4 \le \bar{X}_1 - \bar{X}_2 < 5.9) = P(-1.43 \le Z < 0.80) = P(-1.43 \le Z < 0) + P(0 \le Z < 0.80) = 0.4236 + 0.2881 = 0.7117$$

Sampling distribution of sample proportion $(\hat{P})$

The arrangement of all possible values of sample proportions $(\hat{P})$ with their probabilities is called sampling distribution of sample proportions $(\hat{P})$.

Properties

i) $E(\hat{P}) = \mu_{\hat{P}} = P$ In case of with and without replacement

ii) $\sigma_{\hat{P}}^2 = Var(\hat{P}) = \frac{pq}{n}$ In case of with replacement

Or

$$\sigma_{\hat{P}} = S.E(\hat{P}) = \sqrt{\frac{pq}{n}}$$

iii) $\sigma_{\hat{P}}^2 = Var(\hat{P}) = \frac{pq}{n}\left(\frac{N-n}{N-1}\right)$ In case of without replacement

Or

$$\sigma_{\hat{P}} = S.E(\hat{P}) = \sqrt{\frac{pq}{n}\left(\frac{N-n}{N-1}\right)}$$

iv) Shape of sampling distribution

a) Sampling distribution of sample proportion $\hat{P}$ is binomial for ample size is small.

b) Sampling distribution of sample proportion $\hat{P}$ is approximately normal with mean "P" and variance $\frac{pq}{n}$ for large sample sizes.

Example: A finite population consists of the numbers 1, 2, 3 and 4. Calculate the proportion of even numbers for all possible random samples of size "n=2" that can be drawn from this population, without replacement and make the sampling distribution of proportion verify the results.

i) $\mu_{\hat{P}} = P = E(\hat{P})$ ii) $Var(\hat{P}) = \sigma_{\hat{P}}^2 = \frac{pq}{n}\left(\frac{N-n}{N-1}\right)$

Solution:
$Pop:1,2,3,4$
$N=4$
$n=2$
Draw all possible sample of size "2" by without replacement $={}^{N}C_{n}={}^{4}C_{2}=6$

Sr.No	Samples	$\hat{P}=\frac{X(even\ digit)}{n}$
1	1,2	1/2
2	1,3	0/2
3	1,4	1/2
4	2,3	1/2
5	2,4	2/2
6	3,4	1/2

Sampling distribution of sample means ($\overline{X}$)

$\hat{P}$	Tally	f	$f(\hat{P})$	$\hat{P}f(\hat{P})$	$\hat{P}^2\ f(\hat{P})$
0/2	1	1	1/6	0/12	0/24
1/2	1111	4	4/6	4/12	4/24
2/2	1	1	1/6	2/12	4/24
Total		6	1	6/12	8/24

$\mu_{\hat{P}}=E(\hat{P})=\hat{P}f(\hat{P})=6/12=0.5$

$\sigma^2_{\hat{P}}=E(\hat{P}^2)-\left[E(\hat{P})\right]^2$

$Var(\hat{P})=\sum\hat{P}^2 f(\hat{P})-\left[\sum\hat{P}f(\hat{P})\right]^2$

$\sigma^2_{\hat{P}}=Var(\hat{P})=8/24-\left[6/12\right]^2=0.333-0.25=0.083$

Population distribution

X
1
2
3
4

$P=\frac{X(Even\ Digit)}{N}=\frac{2}{4}=0.5$

$q=1-P=1-0.5=0.5$

Verification

$\mu_{\hat{P}}=E(\hat{P})=P$

$0.5=0.5$ Hence verified

$\sigma^2_{P}=\frac{pq}{n}\left(\frac{N-n}{N-1}\right)$

$\sigma^2_{\hat{P}}=\frac{(0.5)(0.5)}{2}\left(\frac{4-2}{4-1}\right)=0.083$

$0.083=0.083$ Hence verified

Note: Same as Example.14.17, Q.47 and Q.48

Q.49 (a): Two percent of the trees in a population are known to have a certain disease. What is the probability that, in a sample of 250 trees

i) Less than 1% ii) More than 4% are diseased?

Solution:

i) Given that n=250 $P=2\%=0.02$ $q=1-P=0.98$

As we know that

$\mu_{\hat{P}}=P=0.02$

$\sigma_{\hat{P}_1}=\sqrt{\frac{pq}{n}}=\sqrt{\frac{0.02\times 0.98}{250}}=0.0.0088$

$P(\hat{P}<0.01)=?$

By Standard normal variate (S.N.V)

$$Z = \frac{\hat{P} - \mu_{\hat{P}}}{\sigma_{\hat{P}}}$$

$$Z = \frac{0.01 - 0.02}{0.0088} = -1.14$$

$P(\hat{P} < 0.01) = P(Z < -1.14) = P(-\infty < Z < 0) - P(-1.14 < Z < 0) = 0.50 - 0.3729 = 0.1271$

ii) $P(\hat{P} > 0.04) = ?$

By Standard normal variate (S.N.V)

$$Z = \frac{\hat{P} - \mu_{\hat{P}}}{\sigma_{\hat{P}}}$$

$$Z = \frac{0.04 - 0.02}{0.0088} = 2.27$$

$P(\hat{P} > 0.04) = P(Z > 2.27) = P(0 < Z < \infty) - P(0 < Z < 2.27) = 0.50 - 0.4884 = 0.0116$

Q.49 (b): Suppose that 60% of a city population favours public finding for a proposed recreational facility. If 150 persons are to be randomly selected and interviewed. What is the probability that the sample proportion favouring this issue will be less than 0.52?

Solution: Given that n=150 $P = 60\% = 0.60$ $q = 1 - P = 0.40$

As we know that

$$\mu_{\hat{P}} = P = 0.60$$

$$\sigma_{\hat{P}_1} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.60 \times 0.40}{150}} = 0.04$$

$P(\hat{P} < 0.52) = ?$

By Standard normal variate (S.N.V)

$$Z = \frac{\hat{P} - \mu_{\hat{P}}}{\sigma_{\hat{P}}}$$

$$Z = \frac{0.52 - 0.60}{0.04} = -2.0$$

$P(\hat{P} < 0.52) = P(Z < -2.0) = P(-\infty < Z < 0) - P(-2.0 < Z < 0) = 0.50 - 0.4772 = 0.0228$

Q.50: A small professional society has $N = 4500$ members. The president has mailed $n = 400$ questionnaires to a random sample of members asking whether they wish to affiliate with a larger group. Assuming that the proportion of the entire membership favouring consolidation is $P = 0.7$, find that the sample proportion $\hat{P}$ differs from this by no more than 0.05.

Solution: Given that $N = 4500$ $n = 400$ $P = 0.7$ $q = 1 - P = 0.30$

As we know that

$$\mu_{\hat{P}} = P = 0.70$$

$$\sigma_{\hat{P}_1} = \sqrt{\frac{pq}{n}\left(\frac{N-n}{N-1}\right)}$$

$$\frac{n}{N} \times 100 = \frac{400}{4500} \times 100 = 8.8\%$$

So we use FPC because it's greater than 5%

$$\sigma_{\hat{P}_1} = \sqrt{\frac{0.70 \times 0.30}{400}\left(\frac{4500 - 400}{4500 - 1}\right)} = 0.022$$

$P(\hat{P} \textit{ differ by more than } 0.05) = P(0.65 \le \hat{P} \le 0.75) = ?$

By Standard normal variate (S.N.V)

$$Z = \frac{\hat{P} - \mu_{\hat{P}}}{\sigma_{\hat{P}}}$$

$$Z_1 = \frac{0.65 - 0.70}{0.022} = -2.27$$

$$Z_2 = \frac{0.75 - 0.70}{0.022} = 2.27$$

$P(\hat{P} \textit{ differ by more than } 0.05) = P(0.65 \le \hat{P} \le 0.75) = P(-2.27 \le Z \le 2.27)$

$= P(-2.27 < Z < 0) + P(0 < Z < 2.27) = 0.4884 + 0.4884 = 0.9768$

Sampling distribution of difference between two sample proportions $(\hat{P}_1-\hat{P}_2)$

The arrangement of all possible difference between two sample proportions $(\hat{P}_1-\hat{P}_2)$ with their probabilities is called sampling distribution of difference between two proportions $(\hat{P}_1-\hat{P}_2)$

Properties

i) $E(\hat{P}_1-\hat{P}_2)=\mu_{\hat{P}_1-\hat{P}_2}=P_1-P_2$ In case of with and without replacement

ii) $\sigma^2_{\hat{P}_1-\hat{P}_2}=Var(\hat{P}_1-\hat{P}_2)=\frac{p_1q_1}{n_1}+\frac{p_1q_1}{n_2}$ In case of with replacement

Or

$$\sigma_{\hat{P}_1-\hat{P}_2}=S.E(\hat{P}_1-\hat{P}_2)=\sqrt{\frac{p_1q_1}{n_1}+\frac{p_1q_1}{n_2}}$$

iii) In case of without replacement

$$\sigma^2_{\hat{P}_1-\hat{P}_2}=Var(\hat{P}_1-\hat{P}_2)=\frac{p_1q_1}{n_1}\left(\frac{N_1-n_1}{N_1-1}\right)+\frac{p_2q_2}{n_2}\left(\frac{N_2-n_2}{N_2-1}\right)$$

Or

$$\sigma_{\hat{P}_1-\hat{P}_2}=S.E(\hat{P}_1-\hat{P}_2)=\sqrt{\frac{p_1q_1}{n_1}\left(\frac{N_1-n_1}{N_1-1}\right)+\frac{p_2q_2}{n_2}\left(\frac{N_2-n_2}{N_2-1}\right)}$$

iv) Shape of sampling distribution

i) If population sampled is normal then sampling distribution of $\hat{P}_1-\hat{P}_2$ will be normal either sample size is large or small.

ii) Sampling distribution of $\hat{P}_1-\hat{P}_2$ will be approximately normal for large sample sizes.

Example: Draw all possible random sample of size $n_1=2$ with replacement from the population "4 and 6". Similarly, draw all possible samples of size $n_2=2$ with replacement from the population "3 and 4". Construct the sampling distribution of difference between two sample proportions $(\hat{P}_1-\hat{P}_2)$ of odd numbers and show that

i) $E(\hat{P}_1-\hat{P}_2)=\mu_{\hat{P}_1-\hat{P}_2}=P_1-P_2$ ii) $\sigma^2_{\hat{P}_1-\hat{P}_2}=Var(\hat{P}_1-\hat{P}_2)=\frac{p_1q_1}{n_1}+\frac{p_1q_1}{n_2}$

Solution: 1st population

$Pop:1,2$

$N_1=2$

$n_1=2$

Draw all possible sample of size "2" by with replacement $=N^n=(2)^2=4$

Sr.No	Samples	$\hat{P}_1=\frac{x(odd\ digit)}{n}$
1	1,1	2/2
2	1,2	1/2
3	2,1	1/2
4	2,2	0/2

2nd population

$Pop:3,4$

$N_2=2$

$n_2=2$

Draw all possible sample of size "2" by with replacement $=N^n=(2)^2=4$

Sr.No	Samples	$\hat{P}_2=\frac{x(odd\ digit)}{n}$
1	3,3	2/2
2	3,4	1/2
3	4,3	1/2
4	4,4	0/2

Possible difference between two sample means $(\hat{P}_1 - \hat{P}_2)$

$\hat{P}_2$

$\hat{P}_1$	2/2	1/2	1/2	0/2
2/2	0/2	1/2	1/2	2/2
1/2	-1/2	0/2	0/2	1/2
1/2	-1/2	0/2	0/2	1/20
0/2	-2/2	-1/2	-1/2	0/2

Sampling distribution of difference between two sample proportions $((\hat{P}_1 - \hat{P}_2))$

$(\hat{P}_1 - \hat{P}_2)$	Tally	F	$f(\hat{P}_1 - \hat{P}_2)$	$(\hat{P}_1 - \hat{P}_2)f(\hat{P}_1 - \hat{P}_2)$	$(\hat{P}_1 - \hat{P}_2)^2 f(\hat{P}_1 - \hat{P}_2)$
-2/2	1	1	1/16	-2/32	4/64
-1/2	1111	4	4/16	-4/32	4/64
0/2	111111	6	6/16	0/32	0/64
1/2	1111	4	4/16	4/32	4/64
2/2	1	1	1/16	2/32	4/64
Total		16	1	0/32	16/64

$$\mu_{\hat{P}_1-\hat{P}_2} = E(\hat{P}_1 - \hat{P}_2) = (\hat{P}_1 - \hat{P}_2)f(\hat{P}_1 - \hat{P}_2) = 0/32 = 0$$

$$\sigma^2_{\hat{P}_1-\hat{P}_2} = E(\hat{P}_1 - \hat{P}_2)^2 - \left[E(\hat{P}_1 - \hat{P}_2)\right]^2$$

$$Var(\hat{P}_1 - \hat{P}_2) = \sum(\hat{P}_1 - \hat{P}_2)^2 f(\hat{P}_1 - \hat{P}_{22}) - \left[\sum(\hat{P}_1 - \hat{P}_2)f(\hat{P}_1 - \hat{P}_2)\right]^2$$

$$\sigma^2_{\hat{P}_1-\hat{P}_2} = 16/64 - [0/32]^2 = 0.25 - 0 = 0.25$$

Population distribution

1st population

X_1
1
2

$$P_1 = \frac{X_1(odd\ Digit)}{N_1} = \frac{1}{2} = 0.5$$

$$q_1 = 1 - P_1 = 1 - 0.5 = 0.5$$

2nd Population

X_2
3
4

$$P_2 = \frac{X_2(odd\ Digit)}{N_2} = \frac{1}{2} = 0.5$$

$$q_2 = 1 - P_2 = 1 - 0.5 = 0.5$$

Verification

$$E(\hat{P}_1 - \hat{P}_2) = \mu_{\hat{P}_1-\hat{P}_2} = P_1 - P_2$$

$$E(\hat{P}_1 - \hat{P}_2) = 0.5 - 0.5 = 0$$

$0 = 0$ Hence verified

$$\sigma^2_{\hat{P}_1-\hat{P}_2} = Var(\hat{P}_1 - \hat{P}_2) = \frac{p_1 q_1}{n_1} + \frac{p_1 q_1}{n_2}$$

$$\sigma^2_{\hat{P}_1-\hat{P}_2} = \frac{(0.5)(0.5)}{2} + \frac{(0.5)(0.5)}{2} = 0.25$$

$0.25 = 0.25$ Hence verified

Note:

Q.51 (b): Two random samples of sizes $n_1 = 40\ and\ n_2 = 45$ are drawn from a binomial population with $P = 0.70$. What is the sample probability that $-0.10 < \hat{P}_1 - \hat{P}_2 < 0.10$?

Solution: Given that $n_1 = 40\ and\ n_2 = 45$ $P_1 = P_2 = 0.70$ And $q_1 = q_2 = 0.70$

As we know that

$$\mu_{\hat{P}_1-\hat{P}_2} = P_1 - P_2 = 0.70 - 0.70 = 0$$

$$\sigma_{\hat{P}_1-\hat{P}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} = \sqrt{\frac{(0.70 \times 0.30)}{40} + \frac{(0.70 \times 0.30)}{45}} = 0.0995 = 0.10$$

By Standard normal variate (S.N.V)

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - \mu_{\hat{P}_1 - \hat{P}_2}}{\sigma_{\hat{P}_1 - \hat{P}_2}}$$

$$Z_1 = \frac{-0.10 - 0}{0.10} = -1.0$$

$$Z_2 = \frac{0.10 - 0}{0.10} = 1.0$$

$$P(-0.10 < \hat{P}_1 - \hat{P}_2 < 0.10) = P(-1.0 < Z < 1.0) = (P(-1.0 < Z < 0) + P(0 < Z < 1.0) = 0.3413 + 0.3413 = 0.682$$

Sampling distribution of sample of biased variances $S^2 = \frac{\sum(X - \overline{X})^2}{n}$

The arrangement of all possible values of sample variances $S^2 = \frac{\sum(X - \overline{X})^2}{n}$ with their probabilities is called sampling distribution of variances (S^2).

Properties

i) $E(S^2) = \mu_{S^2} = \left(\frac{n-1}{n}\right)\sigma^2$ In case of with replacement

ii) $E(S^2) = \mu_{S^2} = \left(\frac{N}{N-1}\right)\left(\frac{n-1}{n}\right)\sigma^2$ In case of without replacement

Example: A finite population consists of the numbers 2, 4, 6 and 8. Calculate the sample for all possible random samples of size "n=2" that can be drawn from this population, without replacement and make sampling distribution of sample variances if

$S^2 = \frac{\sum(X - \overline{X})^2}{n}$ and verify the results.

$$E(S^2) = \mu_{S^2} = \left(\frac{N}{N-1}\right)\left(\frac{n-1}{n}\right)\sigma^2$$

Solution:

$Pop: 2,4,6,8$

$N = 4$

$n = 2$

Draw all possible sample of size "2" by without replacement $= {}^N C_n = {}^4 C_2 = 6$

Sr.No	Samples	$\overline{X} = \frac{\sum x}{n}$	$S^2 = \frac{\sum(X - \overline{X})^2}{n}$
1	2,4	3	1
2	2,6	4	4
3	2,8	5	9
4	4,6	5	1
5	4,8	6	4
6	6,8	7	1

Sampling distribution of sample means ($\overline{X}$)

S^2	Tally	f	$f(S^2)$	$S^2 f(S^2)$
1	111	3	3/6	3/6
4	11	2	2/6	8/6
9	1	1	1/6	9/6
Total		6	1	20/6

$$\mu_{S^2} = E(S^2) = S^2 f(S^2) = 20/6 = 3.33$$

Population distribution

X	X^2
2	4
4	16
6	36
8	64
$\sum X = 20$	$\sum X^2 = 120$

$$\sigma^2 = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2 = \frac{120}{4} - \left(\frac{20}{4}\right)^2 = 5$$

Verification

$$\mu_{S^2} = E(S^2) = \left(\frac{n-1}{n}\right)\left(\frac{N}{N-1}\right)\sigma^2$$

$$E(S^2) = \left(\frac{2-1}{2}\right)\left(\frac{4}{4-1}\right)5 = 3.33$$

$$3.33 = 3.33$$

Hence verified

Note: Same as Q.52

Sampling distribution of sample of unbiased variances $s^2 = \frac{\sum (X-\bar{X})^2}{n-1}$

The arrangement of all possible values of sample variances $s^2 = \frac{\sum (X-\bar{X})^2}{n-1}$ with their probabilities is called sampling distribution of variances (s^2).

Properties

i) $E(s^2) = \mu_{S^2} = \sigma^2$ In case of with replacement

ii) $E(s^2) = \mu_{S^2} = \left(\frac{N}{N-1}\right)\sigma^2$ In case of without replacement

Example: A finite population consists of the numbers 2, 4, 6 and 8. Calculate the sample for all possible random samples of size "n=2" that can be drawn from this population, without replacement and make sampling distribution of sample variances if $s^2 = \frac{\sum (X-\bar{X})^2}{n-1}$ and verify the results. $E(s^2) = \mu_{S^2} = \frac{\sigma^2}{n}\left(\frac{N}{N-1}\right)$

Solution:

$Pop: 2,4,6,8$

$N = 4$

$n = 2$

Draw all possible sample of size "2" by without replacement $= {}^N C_n = {}^4 C_2 = 6$

Sr.No	Samples	$\bar{X} = \frac{\sum x}{n}$	$s^2 = \frac{\sum (X-\bar{X})^2}{n-1}$
1	2,4	3	2
2	2,6	4	8
3	2,8	5	18
4	4,6	5	2
5	4,8	6	8
6	6,8	7	2

Sampling distribution of sample means ($\bar{X}$)

S^2	Tally	f	$f(S^2)$	$S^2 f(S^2)$
2	111	3	3/6	6/6
8	11	2	2/6	16/6
18	1	1	1/6	18/6
Total		6	1	40/6

$$\mu_{S^2} = E(s^2) = s^2 f(s^2) = 40/6 = 6.67$$

Population distribution

X	X^2
2	4
4	16
6	36
8	64
$\sum X = 20$	$\sum X^2 = 120$

$$\sigma^2 = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2 = \frac{120}{4} - \left(\frac{20}{4}\right)^2 = 5$$

Verification

$$\mu_{S^2} = E(S^2) = \left(\frac{N}{N-1}\right)\sigma^2$$

$$E(S^2) = \left(\frac{4}{4-1}\right)5 = 6.67$$

$6.67 = 6.67$ Hence verified

Q: Prove that the sample mean ($\bar{y}$) is an unbiased estimate of the population mean ($\bar{Y}$)

Proof:

Let by definition

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{1}{n}\left[y_1 + y_2 + y_3 + ... + y_n\right] \qquad \text{Where } i = 1,2,3,...,n$$

Taking expectation on both sides

$$E(\bar{y}) = \frac{1}{n}E\left[y_1 + y_2 + y_3 + ... + y_n\right] = \frac{1}{n}\left[E(y_1) + E(y_2) + E(y_3) + ... + E(y_n)\right]$$

$$E(\bar{y}) = \frac{1}{n}E\left[y_1 + y_2 + y_3 + ... + y_n\right] = \frac{1}{n}\left[E(y_1) + E(y_2) + E(y_3) + ... + E(y_n)\right]$$

For specified value taken from the population could have anyone of the "N: values with an equal probability of $\frac{1}{N}$ as all the values are equally likely.

$$E(\bar{y}) = \frac{1}{n}\left[\bar{Y}_1 + \bar{Y}_2 + \bar{Y}_3 + ... + \bar{Y}_n\right] = \frac{1}{n}\left[n\bar{Y}\right] = \bar{Y} \text{ Proved}$$

Q: Prove that the sample mean ($\bar{y}$) is an unbiased estimate of the population mean ($\bar{Y}$) in case of without replacement.

Proof:

Let by definition

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad \text{Where } i = 1,2,3,...,n$$

Suppose that there are "k" possible samples with mean $\bar{y}_1, \bar{y}_2, \bar{y}_3, ..., \bar{y}_k$. Here all possible samples $k={}^{N}C_n$

Then taking averaging all possible samples

$$E(\bar{y}) = \frac{1}{k}\left[\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + ... + \bar{y}_k\right] = \frac{1}{k}\left[\frac{\sum_{i=1}^{n} y_{1i}}{n} + \frac{\sum_{i=1}^{n} y_{1i}}{n} + \frac{\sum_{i=1}^{n} y_{1i}}{n} + ... + \frac{\sum_{i=1}^{n} y_{ki}}{n}\right]$$

$$E(\bar{y}) = \frac{1}{k}\left[\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + ... + \bar{y}_k\right] = \frac{1}{nk}\left[\sum_{i=1}^{n} y_{1i} + \sum_{i=1}^{n} y_{2i} + \sum_{i=1}^{n} y_{3i} + ... + \sum_{i=1}^{n} y_{ki}\right]$$

We have to find out the no. of samples that contain any specified value y_i . The no. of such sample is ${}^{N-1}C_{n-1}$. i.e. the no. of ways in which the (n-1) other units in the sample to be selected from all remaining (N-1) units.

$$E(\bar{y}) = \frac{1}{nk}\left[{}^{N-1}C_{n-1}(y_{1i}) + {}^{N-1}C_{n-1}(y_{2i}) + {}^{N-1}C_{n-1}(y_{3i}) + ... + {}^{N-1}C_{n-1}(y_{ki})\right]$$

$$E(\bar{y}) = \frac{1}{n({}^{N}C_n)}\,{}^{N-1}C_{n-1}\left[y_{1i} + y_{2i} + y_{3i} + ... + y_{ki}\right] = \frac{1}{n\,(\frac{N}{n})\,{}^{N-1}C_{n-1}}\,{}^{N-1}C_{n-1}\left[y_{1i} + y_{2i} + y_{3i} + ... + y_{ki}\right]$$

$$E(\bar{y}) = \frac{1}{N}\left[\sum_{i=1}^{N} y_i\right] = \bar{Y}$$

Q: Prove that the sample mean ($\bar{y}$) is an unbiased estimate of the population mean ($\bar{Y}$) in case of with replacement.

Proof:
Let by definition

$$\bar{y}=\frac{\sum_{i=1}^{n} y_i}{n} \qquad \text{Where } i=1,2,3,...,n$$

Suppose that there are "k" possible samples with mean $\bar{y}_1,\bar{y}_2,\bar{y}_3,...,\bar{y}_k$. Here all possible samples $k=N^n$

Then taking averaging all possible samples

$$E(\bar{y})=\frac{1}{k}\left[\bar{y}_1+\bar{y}_2+\bar{y}_3+...+\bar{y}_k\right]=\frac{1}{k}\left[\frac{\sum_{i=1}^{n} y_{1i}}{n}+\frac{\sum_{i=1}^{n} y_{1i}}{n}+\frac{\sum_{i=1}^{n} y_{1i}}{n}+...+\frac{\sum_{i=1}^{n} y_{ki}}{n}\right]$$

$$E(\bar{y})=\frac{1}{k}\left[\bar{y}_1+\bar{y}_2+\bar{y}_3+...+\bar{y}_k\right]=\frac{1}{nk}\left[\sum_{i=1}^{n} y_{1i}+\sum_{i=1}^{n} y_{2i}+\sum_{i=1}^{n} y_{3i}+...+\sum_{i=1}^{n} y_{ki}\right]$$

We have to find out the no. of samples that contain any specified value y_i. The no. of such sample is nN^{n-1}

$$E(\bar{y})=\frac{1}{nk}\left[nN^{n-1}(y_{1i})+nN^{n-1}(y_{2i})+nN^{n-1}(y_{3i})+...+nN^{n-1}(y_{ki})\right]$$

$$E(\bar{y})=\frac{nN^{n-1}}{nN^{n}}\left[y_{1i}+y_{2i}+y_{3i}+...+y_{ki}\right]=\frac{1}{N}\left[y_{1i}+y_{2i}+y_{3i}+...+y_{ki}\right]$$

$$E(\bar{y})=\frac{1}{N}\left[\sum_{i=1}^{N} y_i\right]=\bar{Y}$$

Q.53: Show that the variance of the sample mean, $\bar{Y}$ from a simple random sample of size "n" is given by $V(\bar{y})=\sigma^2_{\bar{y}}=\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

Proof: Let by definition

$$Var(\bar{y})=E(\bar{y}-E(\bar{y}))^2 \qquad E(\bar{y})=\bar{Y}$$

$$Var(\bar{y})=E(\bar{y}-\bar{Y})^2$$

$$Var(\bar{y})=E\left(\frac{\sum_{i=1}^{n} y_i}{n}-\bar{Y}\right)^2=\frac{1}{n^2}E\left(\sum_{i=1}^{n} y_i-n\bar{Y}\right)^2=\frac{1}{n^2}E\left(\sum_{i=1}^{n} y_i-\sum_{i=1}^{n}\bar{Y}\right)^2=\frac{1}{n^2}E\left[\sum_{i=1}^{n}\left(y_i-\bar{Y}\right)\right]^2$$

$$Var(\bar{y})=\frac{1}{n^2}E\left[\sum_{i=1}^{n}\left(y_i-\bar{Y}\right)^2+\sum_{i\neq j}^{n(n-1)}\left(y_i-\bar{Y}\right)\left(y_j-\bar{Y}\right)\right]=\frac{1}{n^2}\left[\sum_{i=1}^{n}E\left(y_i-\bar{Y}\right)^2+\sum_{i\neq j}^{n(n-1)}E\left(y_i-\bar{Y}\right)\left(y_j-\bar{Y}\right)\right]$$

$$Var(\bar{y})=\frac{1}{n^2}\left[nE\left(y_i-\bar{Y}\right)^2+n(n-1)E\left(y_i-\bar{Y}\right)\left(y_j-\bar{Y}\right)\right] \qquad E\left(y_i-\bar{Y}\right)^2=\sigma^2$$

$$Var(\bar{y})=\frac{1}{n^2}\left[n\sigma^2+n(n-1)E\left(y_i-\bar{Y}\right)\left(y_j-\bar{Y}\right)\right]\text{.................................. (A)}$$

Now we consider

$$\sum_{i=1}^{N}\left(Y_i-\bar{Y}\right)=0$$

Squaring both side

$$\left[\sum_{i=1}^{N}\left(Y_i-\bar{Y}\right)\right]^2=0$$

$$\left[\sum_{i=1}^{N}\left(Y_i-\bar{Y}\right)+\sum_{i\neq j}^{N(N-1)}\left(Y_i-\bar{Y}\right)\left(Y_j-\bar{Y}\right)\right]=0$$

Taking expectation

$$E\left[\sum_{i=1}^{N}\left(Y_i-\bar{Y}\right)+\sum_{i\neq j}^{N(N-1)}\left(Y_i-\bar{Y}\right)\left(Y_j-\bar{Y}\right)\right]=0$$

$$\left[\sum_{i=1}^{N} E\left(Y_i-\bar{Y}\right)^2+\sum_{i\neq j}^{N(N-1)} E\left(Y_i-\bar{Y}\right)\left(Y_j-\bar{Y}\right)\right]=0$$

$$\left[N\sigma^2+N(N-1)E\left(Y_i-\bar{Y}\right)\left(Y_j-\bar{Y}\right)\right]=0$$
$$\left[N(N-1)E\left(Y_i-\bar{Y}\right)\left(Y_j-\bar{Y}\right)\right]=-N\sigma^2$$

$$E\left(Y_i-\bar{Y}\right)\left(Y_j-\bar{Y}\right)=\frac{-N\sigma^2}{N(N-1)}=\frac{-\sigma^2}{(N-1)}$$

Put in equation (A)

$$Var(\bar{y})=\frac{1}{n^2}\left[n\sigma^2+n(n-1)\frac{-\sigma^2}{N-1}\right]=\frac{n\sigma^2}{n^2}\left[1-(n-1)\frac{1}{N-1}\right]=\frac{\sigma^2}{n}\left[\frac{N-1-n+1}{N-1}\right]\frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right]$$

As we know that $\sigma^2=\frac{N-1}{N}S^2$

$$Var(\bar{y})=\frac{N-n}{N-1}\frac{\sigma^2}{n}$$

Required result

Q.Show that the variance of the sample mean, $\bar{Y}$ from a simple random sample of size "n" is given by $Var(\bar{y})=\frac{\sigma^2}{n}$

Proof:

Let by definition

$$Var(\bar{y})=E\left(\bar{y}-E(\bar{y})\right)^2 \qquad E(\bar{y})=\bar{Y}$$

$$Var(\bar{y})=E\left(\bar{y}-\bar{Y}\right)^2$$

$$Var(\bar{y})=E\left(\frac{\sum_{i=1}^{n} y_i}{n}-\bar{Y}\right)^2=\frac{1}{n^2}E\left(\sum_{i=1}^{n} y_i-n\bar{Y}\right)^2=\frac{1}{n^2}E\left(\sum_{i=1}^{n} y_i-\sum_{i=1}^{n}\bar{Y}\right)^2=\frac{1}{n^2}E\left[\sum_{i=1}^{n}\left(y_i-\bar{Y}\right)\right]^2$$

$$Var(\bar{y})=\frac{1}{n^2}E\left[\sum_{i=1}^{n}\left(y_i-\bar{Y}\right)^2+\sum_{i\neq j}^{n(n-1)}\left(y_i-\bar{Y}\right)\left(y_j-\bar{Y}\right)\right]=\frac{1}{n^2}\left[\sum_{i=1}^{n}E\left(y_i-\bar{Y}\right)^2+\sum_{i\neq j}^{n(n-1)}E\left(y_i-\bar{Y}\right)\left(y_j-\bar{Y}\right)\right]$$

$$Var(\bar{y})=\frac{1}{n^2}\left[nE\left(y_i-\bar{Y}\right)^2+n(n-1)E\left(y_i-\bar{Y}\right)\left(y_j-\bar{Y}\right)\right] \qquad E\left(y_i-\bar{Y}\right)^2=\sigma^2$$

$$Var(\bar{y})=\frac{1}{n^2}\left[n\sigma^2+n(n-1)E\left(y_i-\bar{Y}\right)\left(y_j-\bar{Y}\right)\right]$$

As sampling done with replacement .so each unit independent to each other and crossproduct term vanish or we can write it equal to zero

$$Var(\bar{y})=\frac{1}{n^2}\left[n\sigma^2+n(n-1)(0)\right]=\frac{1}{n^2}\left[n\sigma^2\right]=\frac{\sigma^2}{n}$$

$$\sigma^2=\frac{\sum_{i=1}^{N}\left(Y_i-\bar{Y}\right)^2}{N} \qquad \text{In case of infinite population ....(i)}$$

$$S^2=\frac{\sum_{i=1}^{N}\left(Y_i-\bar{Y}\right)^2}{N-1} \qquad \text{In case of finite population ....(ii)}$$

From (i) and (ii)

$$\sigma^2=\frac{N-1}{N}S^2$$

$$Var(\bar{y})=\frac{N-1}{N}\frac{S^2}{n}=\frac{N-1}{nN}S^2=(1-\frac{1}{N})\frac{S^2}{n}$$

Required result

Q. Prove that in simple random sampling $E(s^2)=\sigma^2$ in case of with replacement.
Proof:
Let by definition

$$s^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}$$

$$(n-1)s^2=\sum_{i=1}^{n}(y_i-\bar{y})^2$$

Adding and subtracting $\bar{Y}$ inside the square bracket

$$(n-1)s^2=\sum_{i=1}^{n}(y_i-\bar{Y}+\bar{Y}-\bar{y})^2$$

$$(n-1)s^2=\sum_{i=1}^{n}\left[(y_i-\bar{Y})-(\bar{y}-\bar{Y})\right]^2$$

$$(n-1)s^2=\sum_{i=1}^{n}\left[(y_i-\bar{Y})^2+(\bar{y}-\bar{Y})^2-2(y_i-\bar{Y})^2(\bar{y}-\bar{Y})^2\right]^2$$

$$(n-1)s^2=\sum_{i=1}^{n}(y_i-\bar{Y})^2+\sum_{i=1}^{n}(\bar{y}-\bar{Y})^2-2(\bar{y}-\bar{Y})\sum_{i=1}^{n}(y_i-\bar{Y})^2$$

$$(n-1)s^2=\sum_{i=1}^{n}(y_i-\bar{Y})^2+n(\bar{y}-\bar{Y})^2-2(\bar{y}-\bar{Y})\sum_{i=1}^{n}(y_i-\bar{Y}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(A)}$$

$$\sum_{i=1}^{n}(y_i-\bar{Y})=\sum_{i=1}^{n}y_i-n\bar{Y}=n\bar{y}-n\bar{Y}=n(\bar{y}-\bar{Y})$$

Put in A

$$(n-1)s^2=\sum_{i=1}^{n}(y_i-\bar{Y})^2+n(\bar{y}-\bar{Y})^2-2(\bar{y}-\bar{Y})n(\bar{y}-\bar{Y})$$

$$(n-1)s^2=\sum_{i=1}^{n}(y_i-\bar{Y})^2+n(\bar{y}-\bar{Y})^2-2n(\bar{y}-\bar{Y})^2$$

$$(n-1)s^2=\sum_{i=1}^{n}(y_i-\bar{Y})^2-n(\bar{y}-\bar{Y})^2$$

Taking expectation on both sides

$$E\left[(n-1)s^2\right]=E\left[\sum_{i=1}^{n}(y_i-\bar{Y})^2-n(\bar{y}-\bar{Y})^2\right]$$

$$(n-1)E(s^2)=\left[\sum_{i=1}^{n}E(y_i-\bar{Y})^2-nE(\bar{y}-\bar{Y})^2\right]$$

$$(n-1)E(s^2)=\left[nE(y_i-\bar{Y})^2-nVar(\bar{y})\right]$$

Therefore $E(y_i-\bar{Y})^2=\sigma^2$ $\quad Var(\bar{y})=\dfrac{\sigma^2}{n}$ In case of with replacement

$$(n-1)E(s^2)=\left[n\sigma^2-n\frac{\sigma^2}{n}\right]=\sigma^2\left[n-\frac{n}{n}\right]=\sigma^2[n-1]$$

$$(n-1)E(s^2)=\sigma^2(n-1)$$

$$E(s^2)=\sigma^2$$

Q: Prove that in simple random sampling $E(s^2)=S^2$ in case of without replacement.
Proof:
Let by definition

$$s^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}$$

$$(n-1)s^2=\sum_{i=1}^{n}(y_i-\bar{y})^2$$

Adding and subtracting $\bar{Y}$ inside the square bracket

$$(n-1)s^2=\sum_{i=1}^{n}(y_i-\bar{Y}+\bar{Y}-\bar{y})^2$$

$$(n-1)s^2 = \sum_{i=1}^{n}\left[\left(y_i - \bar{Y}\right) - \left(\bar{y} - \bar{Y}\right)\right]^2$$

$$(n-1)s^2 = \sum_{i=1}^{n}\left[\left(y_i - \bar{Y}\right)^2 + \left(\bar{y} - \bar{Y}\right)^2 - 2\left(y_i - \bar{Y}\right)^2\left(\bar{y} - \bar{Y}\right)^2\right]^2$$

$$(n-1)s^2 = \sum_{i=1}^{n}\left(y_i - \bar{Y}\right)^2 + \sum_{i=1}^{n}\left(\bar{y} - \bar{Y}\right)^2 - 2\left(\bar{y} - \bar{Y}\right)\sum_{i=1}^{n}\left(y_i - \bar{Y}\right)^2$$

$$(n-1)s^2 = \sum_{i=1}^{n}\left(y_i - \bar{Y}\right)^2 + n\left(\bar{y} - \bar{Y}\right)^2 - 2\left(\bar{y} - \bar{Y}\right)\sum_{i=1}^{n}\left(y_i - \bar{Y}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots \text{(A)}$$

$$\sum_{i=1}^{n}\left(y_i - \bar{Y}\right) = \sum_{i=1}^{n} y_i - n\bar{Y} = n\bar{y} - n\bar{Y} = n(\bar{y} - \bar{Y})$$

Put in A

$$(n-1)s^2 = \sum_{i=1}^{n}\left(y_i - \bar{Y}\right)^2 + n\left(\bar{y} - \bar{Y}\right)^2 - 2\left(\bar{y} - \bar{Y}\right)n(\bar{y} - \bar{Y})$$

$$(n-1)s^2 = \sum_{i=1}^{n}\left(y_i - \bar{Y}\right)^2 + n\left(\bar{y} - \bar{Y}\right)^2 - 2n\left(\bar{y} - \bar{Y}\right)^2$$

$$(n-1)s^2 = \sum_{i=1}^{n}\left(y_i - \bar{Y}\right)^2 - n\left(\bar{y} - \bar{Y}\right)^2$$

Taking expectation on both sides

$$E\left[(n-1)s^2\right] = E\left[\sum_{i=1}^{n}\left(y_i - \bar{Y}\right)^2 - n\left(\bar{y} - \bar{Y}\right)^2\right]$$

$$(n-1)E(s^2) = \left[\sum_{i=1}^{n} E\left(y_i - \bar{Y}\right)^2 - nE\left(\bar{y} - \bar{Y}\right)^2\right]$$

$$(n-1)E(s^2) = \left[nE\left(y_i - \bar{Y}\right)^2 - nVar(\bar{y})\right]$$

Therefore $E\left(y_i - \bar{Y}\right)^2 = \sigma^2$ $\quad Var(\bar{y}) = \dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right)$ In case of with replacement

$$(n-1)E(s^2) = \left[n\sigma^2 - n\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)\right] = \sigma^2\left[n - \left(\frac{N-n}{N-1}\right)\right] = \sigma^2\left[\frac{nN - n - N + n}{N-1}\right]$$

$$(n-1)E(s^2) = \sigma^2\left[\frac{nN - N}{N-1}\right] = \sigma^2\left[\frac{N(n-1)}{N-1}\right]$$

$$E(s^2) == \sigma^2\left[\frac{N}{N-1}\right]$$

As we know that $\sigma^2 = \dfrac{('N-1)S^2}{N}$

$$E(s^2) == \frac{('N-1)S^2}{N}\left[\frac{N}{N-1}\right] = S^2$$

$$E(s^2) = S^2$$

Proved

Prepared by: Muhammad Riaz
Lecturer: Statistics
The Islamia University Bahawalpur
RYK