Sampling Distribution Theory

***Population and Sample:***

1. A *‘population’* is a well-defined group of individuals whose characteristics are to be studied. Populations may be finite or infinite.
	1. **Finite Population:** A population is said to be finite, if it consists of finite or fixed number of elements (i.e., items, objects, measurements or observations). For example, all the university students in Pakistan, the heights of all the students enrolled in Karachi University, etc.
	2. **Infinite Population:** A population is said to be infinite, if there is no limit to the number elements it can contain. For example, the role of two dice, all the heights between 2 and 3 meters, etc.
2. A *‘sample’* is a part of the whole selected with the object that it will represent the characteristics of the whole or population or universe. The individuals or objects of a population or a sample may be concrete things like the motor cars produced in a company, wheat produced in a farm, or abstract things like the opinion of students about the examination system. Thus all the students in schools, colleges and universities form population of students. The process of selecting the sample from a population is called *‘sampling’*. A sample may be taken with replacement or without replacement:
	1. **Sampling with Replacement:** If the sample is taken with replacement from a population finite or infinite, the element drawn is returned to the population before drawing the next element.
	2. **Sampling without Replacement:** If the sample is taken without replacement from a finite population, the element selected is not returned to the population.

***Probability Samples and Non-Probability Samples:***

1. *‘Probability samples’* are those in which every element has a known probability of being included in the sample. Following are the probability sampling designs:
2. **Simple Random Sampling:** refers to a method of selecting a sample of a given size from a given population in such a way that all possible samples of this size which could be formed from this population have equal probabilities of selection. It is a method in which a sample of n is selected from the population of N units such that each one of the NCn distinct samples has an equal chance of being drawn. This method sometimes also refers to ‘lottery method’.
3. **Stratified Random Sampling:** consists of the following two steps:
	* 1. The material or area to be sampled is divided into groups or classes called *‘strata’*. Items within each stratum are homogenous.
		2. From each stratum, a simple random sample is taken and the overall sample is obtained by combining the samples for all strata.
4. **Systematic Sampling:** is another form of sample design in which the samples are equally spaced throughout the area or population to be sampled. For e.g., in house-to-house sampling every 10th or 20th house may be taken. More specifically a systematic sample is obtained by taking every kth unit in the population after the units in population have been numbered or arranged in some way.
5. **Cluster Sampling:** One of the main difficulties in large scale surveys is the extensive area that may have to be covered in getting a random or stratified random sample. It may be very expensive and lengthy task to cover the whole population in order to obtain a representative sample. It is not possible to take a simple random or systematic sample of persons from the entire country or from within strata, since there is no such list in which all the individuals are numbered from 1 to N. Even if such a list existed, it would be too expensive to base the enquiry on a simple random sample of persons. Under these circumstances, it is economical to select groups called *‘clusters’* of elements from the population. This is called *‘cluster sampling’*. The difference between a cluster and a stratum is that a stratum is expected to be homogenous and a cluster must be heterogeneous as possible. Clusters are also known as the primary sampling units. Cluster sampling may be consisted of:
	1. Single-stage Cluster Sampling,
	2. Sub-sampling or Two-stage Sampling, and
	3. Multi-stage Sampling.
6. *‘Non-probability sampling’* designs consist of:
7. **Judgement or Purposive Sampling:** There are many situations where investigators use judgement samples to gain needed information. For example, it may be convenient to select a random sample from a cart-load of melons. The melons selected may be very large or very small. The observer may use his own judgement. This method is very useful when the sample to be drawn is small.
8. **Quota Sampling:** is widely used in opinions, market surveys, etc. In such surveys, the interviewers are simply given quotas to be filled in from different strata, with practically no restrictions on how they are to be filled in.

***Parameters and Statistic:***

1. A numerical value such as mean, median or standard deviation calculated from the population is called a *‘population parameter’* or simply a *‘parameter’*. On the other hand, a numerical value such as mean, median or SD calculated from the sample is called a *‘sample statistic’* or simply a *‘statistic’*.
2. Parameters are fixed numbers, i.e., they are constants. Statistics very from sample to sample from the same population.
3. In general, corresponding to each population parameter there will be a statistic to be computed from the sample.
4. The purpose of sampling is to gather information that will be used as a basis for making generalisation about the unknown population parameters.
5. A parameter is usually denoted by a Greek letter and a statistic is usually denoted by a Roman letter. For e.g., the population mean is denoted by μ while the sample mean is denoted by . Similarly, the SD of a population is denoted by σ while the sample SD is denoted by S.

***Sampling and Non-Sampling Errors:***

**(a) Sampling Errors:**

1. The sample data deals with only a portion of the population under consideration rather than the whole population. Because of this partial information about the population, there is always a chance of *‘errors’* or *‘discrepancies’* to exist. This discrepancy or error is simply known as *‘sampling error’*. It is also known as *‘sampling variations’* and *‘chance variations’*.
2. Sampling error is present whenever a sample is drawn. Mathematically, the sampling error is defined as the difference between the sample statistic and population parameter. The conventional procedure consists of subtracting the value of parameter, θ, from that of the statistic t; that is, the sampling error, E, is:

E = t – θ

1. The sampling errors are negative if the parameter is under estimated, and positive if it is over-estimated.
2. The chance of sampling error can be reduced by increasing the size of the sample.

**(b) Non-Sampling Errors:**

1. Such errors enter into any kind of investigation whether it is a sample or a complete census.
2. Non-sampling errors arise from the following reasons:
	* Faulty interviews and questionnaires,
	* Incomplete and inaccurate responses,
	* Mistakes in recording or coding the data,
	* Errors made in processing the results, etc.
3. These errors can be controlled if the volume of data processed is small.
4. Non-sampling errors are less significant in a sample.

***Bias:***

1. It is refer to the overall or long-run tendency of the sample results to differ from the parameter in the particular way.
2. Bias should be not be confused with sampling errors. Mathematically, it is defined as below:

B = m – μ

Where μ is the true population value and m is the mean of the sample statistics of an infinity of samples.

1. The bias may be positive or negative according to as m is greater or less than μ.

***Precision and Accuracy:***

1. *‘Accuracy’* refers to the size of deviations from the true mean μ, whereas, the *‘precision’* refers to the size of deviation from the overall mean m obtained by repeated application of the sampling procedure.
2. Precision is a measure of the closeness of the sample estimates to the census count taken under identical conditions and is judged in sampling theory by the variance of the estimates concerned.

***Sampling Distribution:***

1. The value of a statistic varies from one sample to another even if the samples are selected from the same population. Thus, statistic is a random variable.
2. The distribution or probability distribution of a statistic is called a sampling distribution. For e.g., the distribution of sample mean is a sampling distribution of mean and the distribution of the sample proportion is a sampling distribution of proportion. The SD of the sampling distribution of a statistic is called the *‘standard error’* of the statistic.

**Sampling Distribution of Mean:**

From a finite population of N units with mean μ and SD σ, draw all possible random samples of size n. Find the mean  of every sample. Statistic is now a random variable. Form a probability distribution of , known as *‘sampling distribution of mean’.*

The sampling distribution of mean is one of the most fundamental concepts of statistical inference and it has the following properties:

1. The ***mean of the sampling distribution*** of mean is equal to the population mean:



1. If the sampling is done without replacement from a finite population, the ***standard error of mean*** is given by:



Where  is Finite Population Correction (f.p.c.)

 is sampling fraction

1. When f.p.c. approaches one, the ***standard error of mean*** is simplified as:

 with replacement finite

The f.p.c. approaches one in each of the following cases:

* + 1. when the population is infinite,
		2. when sampling fraction  is less than 0.05, and
		3. when the sampling is with replacement.

Whenever, the sampling is with replacement, the population is considered infinite. For e.g., a box contains 5 balls, when a sample is drawn with replacement, the sample size can be extended from n = 1 to n = 100 or whatever size is desired. Hence, the population is considered to be infinite.

**Mean and Standard Deviation of Sampling Distribution:**

Like other distribution, the sampling distribution of  has a mean and standard deviation:

 -------------------------- Mean of sampling distribution

The standard deviation of sampling distribution of  is known as *‘standard error’* (). The standard error of mean is always less than the SD of population, i.e., σ. It depends on the size of the sample drawn. If the sample size increases, the standard error of mean decreases and consequently the value of sample mean will be closer to the value of population mean.

 -------------------------- SD of sampling distribution

or alternatively

 ------------------------- SD of sampling distribution

**No. of Possible Samples:**

The number of possible samples can be calculated as below:

* 1. When sampling is done ***without replacement***, all possible samples = NCn
	2. When sampling is done ***with replacement***, all possible samples = Nn

**Example:**

A population consists of following data:

1, 2, 3, 4

Suppose that a sample of size 2 is drawn ‘with replacement’. You are required to calculate the following:

* + 1. Population mean,
		2. Population standard deviation,
		3. Mean of each sample,
		4. Sampling distribution table of sample mean with replacement, and
		5. Mean and standard deviation of sampling distribution.

**Solution:**

N = 4

n = 2

No. of samples (when sampling is with replacement) = Nn = 42 = 16

**(a) Population Mean (μ):**



**(b) Population Standard Deviation (σ):**



**(c) Mean () of Each Sample:**

*Samples (with replacement):*

|  |  |  |  |
| --- | --- | --- | --- |
| (1,1) | (2,1) | (3,1) | (4,1) |
| (1,2) | (2,2) | (3,2) | (4,2) |
| (1,3) | (2,3) | (3,3) | (4,3) |
| (1,4) | (2,4) | (3,4) | (4,4) |

*Mean ():*

|  |  |  |  |
| --- | --- | --- | --- |
| 1.0 | 1.5 | 2.0 | 2.5 |
| 1.5 | 2.0 | 2.5 | 3.0 |
| 2.0 | 2.5 | 3.0 | 3.5 |
| 2.5 | 3.0 | 3.5 | 4.0 |

**(d) Sampling Distribution:**

**Sampling Distribution of Sample Mean () with Replacement**

|  |  |
| --- | --- |
| **Frequency Distribution of**  | **Probability Distribution of**  |
|  | **Tally Marks** | ***f*** |  **=**  |  |
| 1.0 | | | 1 | 1.0 | 0.0625 |
| 1.5 | || | 2 | 1.5 | 0.125 |
| 2.0 | ||| | 3 | 2.0 | 0.1875 |
| 2.5 | |||| | 4 | 2.5 | 0.25 |
| 3.0 | ||| | 3 | 3.0 | 0.1875 |
| 3.5 | || | 2 | 3.5 | 0.125 |
| 4.0 | | | 1 | 4.0 | 0.0625 |
| Total |  | 16 |  | 1 |

**(e) Mean and standard deviation of sampling distribution:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| 1.0 | 0.0625 | 0.0625 | –1.5 | 2.25 | 0.1406 | 0.0625 |
| 1.5 | 0.125 | 0.1875 | –1.0 | 1 | 0.125 | 0.2812 |
| 2.0 | 0.1875 | 0.375 | –0.5 | 0.25 | 0.0469 | 0.75 |
| 2.5 | 0.25 | 0.625 | 0 | 0 | 0 | 1.5625 |
| 3.0 | 0.1875 | 0.5625 | 0.5 | 0.25 | 0.0469 | 1.6875 |
| 3.5 | 0.125 | 0.4375 | 1.0 | 1 | 0.125 | 1.5312 |
| 4.0 | 0.0625 | 0.25 | 1.5 | 2.25 | 0.1406 | 1 |
| Total | 1 | 2.5 |  |  | 0.625 | 6.8749 |





**Example:**

Take the data of previous example and assume sampling ‘without replacement’, and compute:

1. Population mean,
2. Population standard deviation,
3. Mean of each sample,
4. Sampling distribution table of sample mean w/o replacement, and
5. Mean and standard deviation of sampling distribution.

**Solution:**

**(a) and (b) Population mean and SD:**

As calculated above

**(c) Mean of each sample:**

No. of possible samples = NCn = 4C2 = 6 samples

*Samples (without replacement):*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| (1,2) | (1,3) | (1,4) | (2,3) | (2,4) | (3,4) |

*Mean:*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1.5 | 2 | 2.5 | 2.5 | 3 | 3.5 |

**(d) Sampling Distribution:**

**Sampling Distribution of Sample Mean () without replacement**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | *f*(**)** |  |  |  |  |  |  |
| 1.5 | 1/6 | 0.25 | –1 | 1 | 0.17 | 2.25 | 0.375 |
| 2 | 1/6 | 0.33 | –0.5 | 0.25 | 0.04 | 4 | 0.666 |
| 2.5 | 2/6 | 0.84 | 0 | 0 | 0 | 6.25 | 2.082 |
| 3 | 1/6 | 0.5 | 0.5 | 0.25 | 0.04 | 9 | 1.5 |
| 3.5 | 1/6 | 0.58 | 1 | 1 | 0.17 | 12.25 | 2.042 |
| Total | 1 | 2.5 |  |  | 0.42 |  | 6.665 |

**(e) Mean and SD of Sampling Distribution:**





**Sampling Distribution of the Differences of Means:**

1. Suppose we have two infinite populations I and II with means μ1 and μ2, and SD σ1 and σ2 respectively.
2.  is the sample mean of n1 from population I and  of n2 from population II with SDs  and  respectively.
3. From the two finite populations, we can obtain a distribution of differences of means.  is called *‘Sampling Distribution of Differences of the Means’*:



Provided that  and  = 0.05

The distribution of  is normal if:

1. the samples are drawn from Normal (or Symmetrical) populations, or
2. n1 and n2 both are at least 30.

The distribution of *‘z’* will be standard normal:



**Example:**

Population I = {1, 2, 3, 4}

Population II = {3,4,5}

Samples drawn from each population with replacement:

n1 = 2

n2 = 2

Compute means of each samples, possible differences between  and , sampling distribution of , and mean and SD of sampling distribution of .

**Solution:**

No. of possible samples from Population I = Nn = 42 = 16 samples

Samples I:

|  |  |  |  |
| --- | --- | --- | --- |
| 1,1 | 1,2 | 1,3 | 1,4 |
| 2,1 | 2,2 | 2,3 | 2,4 |
| 3,1 | 3,2 | 3,3 | 3,4 |
| 4,1 | 4,2 | 4,3 | 4,4 |



|  |  |  |  |
| --- | --- | --- | --- |
| 1.0 | 1.5 | 2.0 | 2.5 |
| 1.5 | 2.0 | 2.5 | 3.0 |
| 2.0 | 2.5 | 3.0 | 3.5 |
| 2.5 | 3.0 | 3.5 | 4.0 |

No. of possible samples from Population II = Nn = 32 = 9 samples

Samples II:

|  |  |  |
| --- | --- | --- |
| 3,3 | 3,4 | 3,5 |
| 4,3 | 4,4 | 4,5 |
| 5,3 | 5,4 | 5,5 |

:

|  |  |  |
| --- | --- | --- |
| 3.0 | 3.5 | 4.0 |
| 3.5 | 4.0 | 4.5 |
| 4.0 | 4.5 | 5.0 |

**Differences of Independent Sample Means **

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **1** | **1.5** | **2** | **2.5** | **1.5** | **2** | **2.5** | **3** | **2** | **2.5** | **3** | **3.5** | **2.5** | **3** | **3.5** | **4** |
| **3** | -2 | -1.5 | -1 | -0.5 | -1.5 | -1 | -0.5 | 0 | -1 | -0.5 | 0 | 0.5 | -0.5 | 0 | 0.5 | 1 |
| **3.5** | -2.5 | -2 | -1.5 | -1 | -2 | -1.5 | -1 | -0.5 | -1.5 | -1 | -0.5 | 0 | -1 | -0.5 | 0 | 0.5 |
| **4** | -3 | -2.5 | -2 | -1.5 | -2.5 | -2 | -1.5 | -1 | -2 | -1.5 | -1 | -0.5 | -1.5 | -1 | -0.5 | 0 |
| **3.5** | -2.5 | -2 | -1.5 | -1 | -2 | -1.5 | -1 | -0.5 | -1.5 | -1 | -0.5 | 0 | -1 | -0.5 | 0 | 0.5 |
| **4** | -3 | -2.5 | -2 | -1.5 | -2.5 | -2 | -1.5 | -1 | -2 | -1.5 | -1 | -0.5 | -1.5 | -1 | -0.5 | 0 |
| **4.5** | -3.5 | -3 | -2.5 | -2 | -3 | -2.5 | -2 | -1.5 | -2.5 | -2 | -1.5 | -1 | -2 | -1.5 | -1 | -0.5 |
| **4** | -3 | -2.5 | -2 | -1.5 | -2.5 | -2 | -1.5 | -1 | -2 | -1.5 | -1 | -0.5 | -1.5 | -1 | -0.5 | 0 |
| **4.5** | -3.5 | -3 | -2.5 | -2 | -3 | -2.5 | -2 | -1.5 | -2.5 | -2 | -1.5 | -1 | -2 | -1.5 | -1 | -0.5 |
| **5** | -4 | -3.5 | -3 | -2.5 | -3.5 | -3 | -2.5 | -2 | -3 | -2.5 | -2 | -1.5 | -2.5 | -2 | -1.5 | -1 |

**Sampling Distribution of  with Replacement**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Tally Marks** | ***f*** |  |  |  |
| –4 | | | 1 | 0.00694 | –0.02776 | 0.11104 |
| –3.5 | |||| | 4 | 0.02778 | –0.09723 | 0.340305 |
| –3 |  | 10 | 0.06945 | –0.20835 | 0.62505 |
| –2.5 | ||| | 18 | 0.125 | –0.3125 | 0.78125 |
| –2 |  | 25 | 0.17361 | –0.34722 | 0.69444 |
| –1.5 | ||| | 28 | 0.19444 | –0.29166 | 0.43749 |
| –1 |  | 25 | 0.17361 | –0.17361 | 0.17361 |
| –0.5 | ||| | 18 | 0.125 | –0.0625 | 0.03125 |
| 0 |  | 10 | 0.06945 | 0 | 0 |
| 0.5 | |||| | 4 | 0.02778 | 0.01389 | 0.006945 |
| 1 | | | 1 | 0.00694 | 0.00694 | 0.00694 |
| Total |  | 144 | 1 | –1.5 | 3.20832 |



**Shape of the Sampling Distribution of :**

The Central Limit Theorem describes the shape of the sampling distribution of mean. The theorem states that the sampling distribution of mean is normal distribution either if the population is normal or if the sample size is more than 30.

Central limit theorem also specifies the relationship between μ and  and the relationship between σ and .

If the sampling distribution of mean is normal, we would expect 68.27%, 95.45% and 99.73% of the sample means to lie within the intervals ,  and  respectively.

**Sampling Distribution of Proportion:**

1. The sampling distribution of proportion is defined as:



Where x is the number of successes (values with a specified characteristic) in a sample of size n.

1. If the sampling procedure is simple random, with replacement, x is recognised as Binomial Random Variable with parameters n and π, π is the probability of success. π can also be interpreted as the population proportion, since:



1. To determine the mean and variance of p:

|  |  |
| --- | --- |
| *Infinite Population with Replacement:* | or alternatively |
| *Finite Population without Replacement:* |

**Example:**

A coordination team consists of seven members. The education of each member as follows: (G = Graduate, PG = Post Graduate)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Members | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| **Education** | G | PG | PG | PG | PG | G | G |

1. Determine the proportion of post-graduates in the population.
2. Select all possible samples of two members from the population without replacement, and compute the proportion of post-graduate members in each sample.
3. Compute the mean (μp) and the SD (σp) of the sample proportion computed in (ii).

**Solution:**

**(i)** Proportion of PG in the population:

N = 7

No. of PG = 4

π = 4/7 = 0.57

**(ii)** No. of possible samples (without replacement) = NCn = 7C2 = 21 samples.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1,2 | 1,3 | 1,4 | 1,5 | 1,6 | 1,7 |
|  | 2,3 | 2,4 | 2,5 | 2,6 | 2,7 |
|  |  | 3,4 | 3,5 | 3,6 | 3,7 |
|  |  |  | 4,5 | 4,6 | 4,7 |
|  |  |  |  | 5,6 | 5,7 |
|  |  |  |  |  | 6,7 |

The corresponding sampling proportions are:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 0.5 | 0.5 | 0.5 | 0.5 | 0 | 0 |
|  | 1 | 1 | 1 | 0.5 | 0.5 |
|  |  | 1 | 1 | 0.5 | 0.5 |
|  |  |  | 1 | 0.5 | 0.5 |
|  |  |  |  | 0.5 | 0.5 |
|  |  |  |  |  | 0 |

# Sampling Distribution of Proportion

|  |  |  |  |
| --- | --- | --- | --- |
| ***p*** | **Tally Marks** | ***f*** | ***P(p)*** |
| 0 | ||| | 3 | 3/21 = 1/7 = 0.143 |
| 0.5 | || | 12 | 12/21 = 4/7 = 0.571 |
| 1 | | | 6 | 6/21 = 2/7 = 0.286 |
| Total |  | 21 | 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***p*** | ***P(p)*** | ***p.P(p)*** |  |  |  | ***p2.P(p)*** |
| 0 | 0.143 | 0 | –0.5715 | 0.32661 | 0.04671 | 0 |
| 0.5 | 0.571 | 0.2855 | –0.0715 | 0.00511 | 0.00292 | 0.14275 |
| 1 | 0.286 | 0.286 | 0.4285 | 0.18361 | 0.05251 | 0.286 |
| Total |  | 0.5715 |  |  | 0.10214 | 0.42875 |

**(iii)** Mean () and SD () of sample proportion distribution:



or alternatively



The results are verified as below:



**Shape of the Sampling Distribution of Proportion p:**

The central limit theorem also holds for the random variable p, which states that:

1. The sampling distribution of proportion p approaches a normal distribution with mean  and SD (with replacement)
2. If the random sampling is without replacement and the sampling fraction , the f.p.c. must be used as below in the formula of SD:



1. When n ≥ 50 and both n.π and n(1 – π) are greater than 5, the sampling distribution can be considered *‘normal’*.
2. When the distribution of p is normal, the following statistic will be standard normal variable:



**Sampling Distribution of Difference between Two Proportions:**

1. If two random samples of size n1 and n2 are drawn independently from two populations with proportions π1 and π2 the sampling distribution of (p1 – p2) the difference between two sample proportions, approaches normal distribution with:



as n1 and n2 increase.

Moreover:



will be standard normal variable.

1. For unknown π1 and π2, sample estimates p1 and p2 are used thus:



1. When the two unknown population proportions can be assumed equal, an estimated  is obtained as below:



and the estimated standard error as below:



**Sampling Distribution of t:**

1. If a random sample of size n is drawn from a known Normal Population with mean μ and SD σ, the sampling distribution of the sample mean  is a normal distribution with mean  and standard error , and hence z would be a standard normal variable:



1. But when the population is unknown with unknown SD σ, the value of σ is replaced the sample SD *‘S’*, as given below:



Therefore, the standard error is equal to :



1. According to W.S. Gossett, the following statistics is denoted by *‘t’* instead of *‘z’*, which follows another distribution known as *‘students’ t-distribution’* or simply *‘t-distribution’*.
2. The sample standard deviation is given by:



In the above equation the (n – 1) is called *‘Degree of Freedom’* or simply d.f., through which we can obtain *‘t-value’* from *‘t-table’*.

1. The t-distribution approaches standard normal distribution as n increases. Typically when n > 30, the t-distribution is considered approximately standard normal.

**Properties of t-distribution:**

1. The t-distribution, like the standard normal, is bell shaped, unimodal and symmetrical about the mean,
2. There is a different t-distribution for every possible sample size,
3. The exact shape of t-distribution, depends on the parameter, the number of degrees of freedom, denoted by ν.
4. As the sample size increases, the shape of t-distribution becomes approximately equal to the standard normal distribution:



z-distribution

t-distribution (n = 8)

t-distribution (n = 5)

1. The mean and standard error of t-distribution are:



**Sampling Distribution of Variances:**

*Population Variance:*



or alternatively



*Mean of sampling distribution of S2 ():*



**Example:**

A population consists of the following numbers: 1,3,5,7. Find the population variance (σ2) and the mean of sampling distribution of variances (), if all samples are drawn with replacement of size 2 from the population.

**Solution:**



No. of possible samples (with replacement) = Nn = 42 = 16 samples

*Samples:*

|  |  |  |  |
| --- | --- | --- | --- |
| 1,1 | 1,3 | 1,5 | 1,7 |
| 3,1 | 3,3 | 3,5 | 3,7 |
| 5,1 | 5,3 | 5,5 | 5,7 |
| 7,1 | 7,3 | 7,5 | 7,7 |

*Means of samples:*

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | 2 | 3 | 4 |
| 2 | 3 | 4 | 5 |
| 3 | 4 | 5 | 6 |
| 4 | 5 | 6 | 7 |

*Variances of samples:*

|  |  |  |  |
| --- | --- | --- | --- |
| 0 | 1 | 4 | 9 |
| 1 | 0 | 1 | 4 |
| 4 | 1 | 0 | 1 |
| 9 | 4 | 1 | 0 |

*Sampling Distribution of S2:*

|  |  |  |  |
| --- | --- | --- | --- |
| ***S2*** | **Tally Marks** | ***f*** | ***f.S2*** |
| 0 | |||| | 4 | 0 |
| 1 | | | 6 | 6 |
| 4 | |||| | 4 | 16 |
| 9 | || | 2 | 18 |
| Total |  | 16 | 40 |



**Pooled Estimate of Variance:**

1. If random samples of size n1 and n2 are drawn independently from two normal populations with means μ1 and μ2 and variances σ12 and σ22, the sampling distribution of the difference between the sample means  follows a normal distribution with mean and standard error given as below:



Thus, the π will be equal to:



and it will be a standard normal variable.

1. But if σ12 and σ22 are unknown and equal, their estimators S12 and S22 are defined as:



When the σ12 and σ22 are replaced by the estimators S12 and S22 the distribution of  can be standardised provided that the samples are large (n1 and n2 > 30).

1. But when samples are small, i.e., less than 30 (n1 and n2 ≤ 30), σ12 and σ22 are replaced by a single estimator known as *‘pooled variance’* denoted by Sp2:

*Weighted Average of S12 and S22:*



Where (n1 + n2 – 2) is the degree of freedom.

1. With same size of samples n1 and n2, the estimator Sp2 is the simple average of S12 and S22:



1. The pooled variance Sp2 assumes that the population variance is unknown and equal. However, the same Sp2 is used to replace σ12 and σ22 for slightly unequal population variances provided that the samples are of equal size, i.e., n1 = n2.
2. In both of the above situations, i.e., equal population variance and slightly unequal population variance with equal samples (i.e., n1 = n2), the statistic t is calculated as below:



Where Sp is pooled SD.

1. Now consider the situation where σ12 and σ22 are considerably different (both unknown) and it is impossible to draw samples of equal size, the statistics used in this case would be:



Where the degree of freedom ν is as follows:

