2HSOE52 Introduction to Economics

Chapter 1 _C: Basic Statistical Concept

Theory of Distribution
Sampling Distribution

A. Theoretical Probability Distribution

Some prominent theoretical distribution are

- **Discrete Probability Distribution**
  - **Binomial Distributing**
  - **Poisson Distribution**
  - **Uniforms Distribution**

- **Continuous Probability Distribution**
  - **Normal Distribution**

Our thrust will be on Normal Distribution . As the number of observation increases (n is large), Binomial Distribution, Poisson Distribution , uniform distribution tend to normal distribution.

B.1 Normal Distribution

Definition: A continuous random variable X is said to follow normal distribution with parameters μ $(-\infty < \mu < \infty)$ and $\sigma^2 (>0)$ if it takes on any real value and its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty;$$

which may also be written as

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, -\infty < x < \infty.$$

B.1 .1 Normal Distribution : Properties

i) The probability function represented by $f(x)$ may also be written as $f(x; \mu, \sigma^2)$.

ii) If a random variable X follows normal distribution with mean μ and variance σ^2, then we may write, "X is distributed to $N(\mu, \sigma^2)$" and is expressed as $X \sim N(\mu, \sigma^2)$.

iii. The probability that a normal random variable *X* equals any particular value say 'a' is 0.

B.1 .1 Normal Distribution : Properties

iv) If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma}$ is standard normal variate having mean '0' and variance '1'. The values of mean and variance of standard normal variate are obtained as under, for which properties of expectation and variance are used (see Unit 8 of this course).

Mean of Z i.e. $E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}[E(X-\mu)]$

$$= \frac{1}{\sigma}[E(X)-\mu]$$

$$= \frac{1}{\sigma}[\mu-\mu] = 0 \qquad [\because E(X) = \text{Mean of } X = \mu\ ]$$

Variance of Z i.e. $V(Z) = V\left(\frac{X-\mu}{\sigma}\right)$

$$= \frac{1}{\sigma^2}[V(X-\mu)] = \frac{1}{\sigma^2}[V(X)]$$

$$= \frac{1}{\sigma^2}(\sigma^2) \qquad [\because \text{variance of } X \text{ is } \sigma^2]$$

$$= 1.$$

B.1 .1 Normal Distribution :Properties

v) The probability density function of standard normal variate $Z = \frac{X-\mu}{\sigma}$

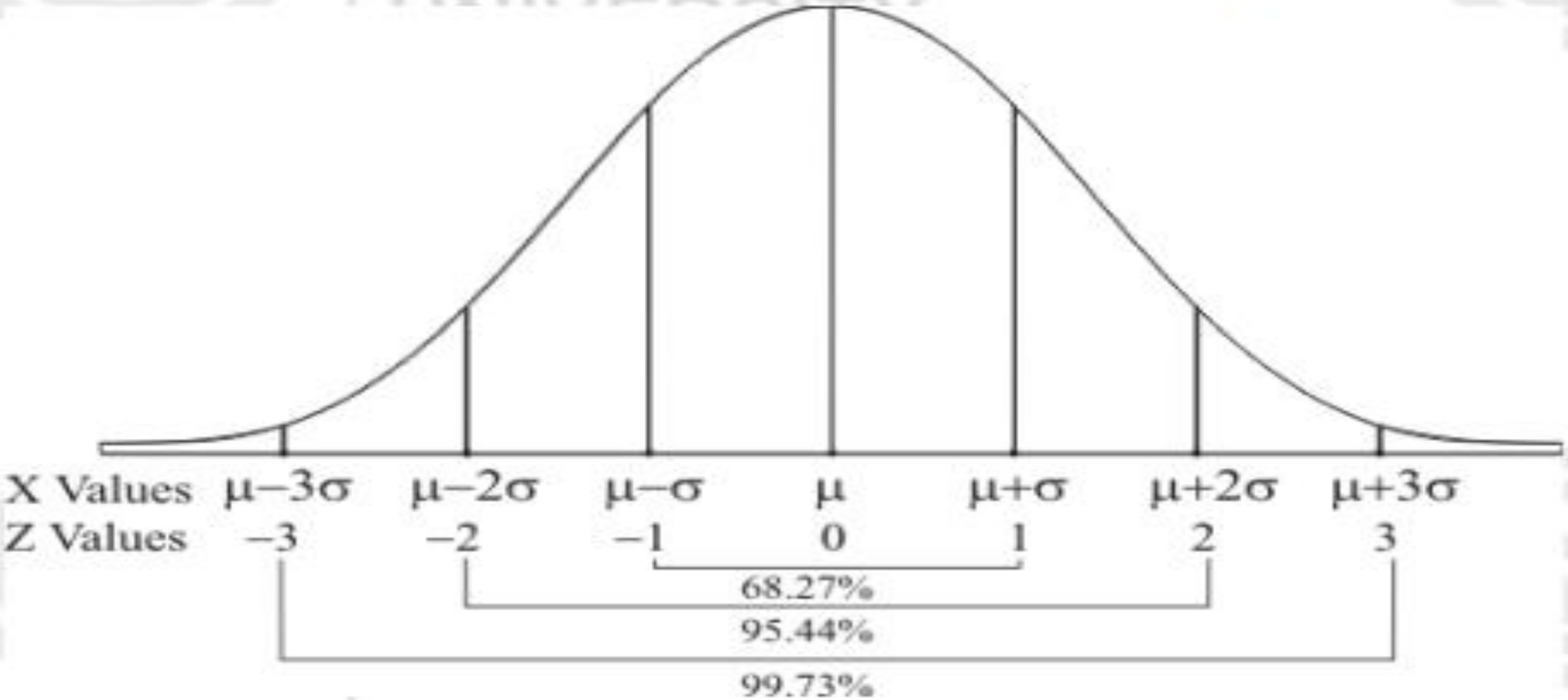
is given by $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$, $-\infty < z < \infty$.

This result can be obtained on replacing $f(x)$ by $\phi(z)$, x by z, μ by 0 and σ by 1 in the probability density function of normal variate X i.e. in

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty$$

B.1 .1 Normal Distribution :Properties

vi) The graph of the normal probability function $f(x)$ with respect to x is famous 'bell-shaped' curve. The top of the bell is directly above the mean μ. For large value of σ, the curve tends to flatten out and for small values of σ, it has a sharp peak as shown in (Fig. 13.1):



B.1 .2 Normal Distribution :Characteristics/Properties

Every normal curve (regardless of its mean or standard deviation) conforms to the following "rule".

- **About 68% of the area under the curve falls within 1 standard deviation of the mean.**
- **About 95% of the area under the curve falls within 2 standard deviations of the mean.**
- **About 99.7% of the area under the curve falls within 3 standard deviations of the mean.**

Clearly, given a normal distribution, most outcomes will be within 3 standard deviations of the mean.

B.1 .2 Normal Distribution : Properties

vii) The curve of the normal distribution is bell-shaped.

viii) The curve of the distribution is completely symmetrical about x = μ i.e. if we fold the curve at x = μ , both the parts of the curve are the mirror images of each other.

ix) For normal distribution, Mean = Median = Mode

x) f (x) , being the probability, can never be negative and hence no portion of the curve lies below x-axis.

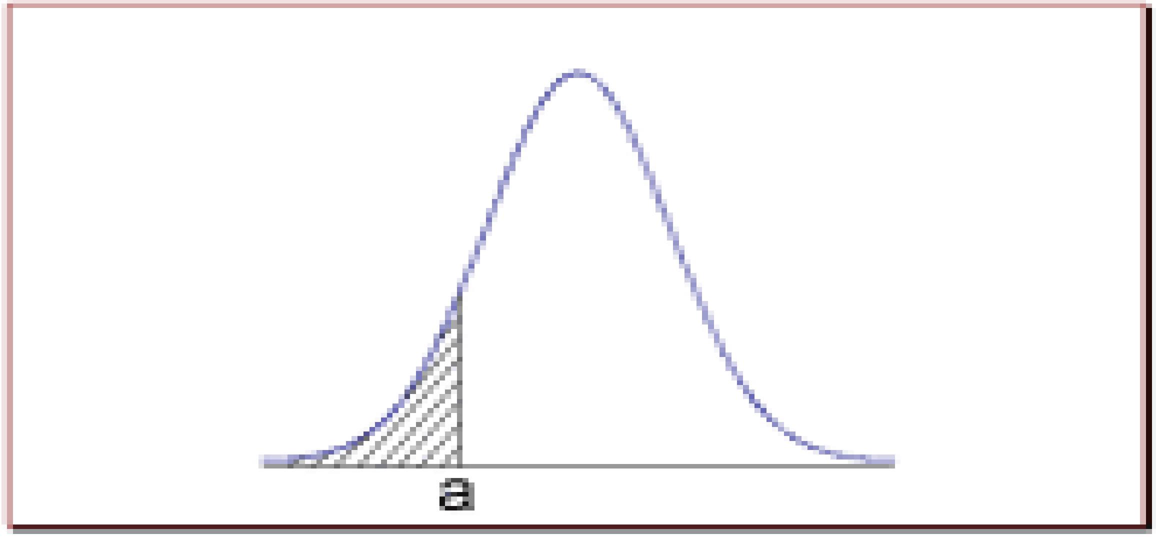
) Though x-axis becomes closer and closer to the normal curve as the magnitude of the value of *x* goes towards ∞ - ∞, yet it never touches it.

xii) Normal curve has only one mode.

xiii **The total area under the normal curve is equal to 1.**

B.1 .2 Normal Distribution :Characteristics

xiv. The probability that *X* is greater than *a* equals the area under the normal curve bounded by *a* and plus infinity (as indicated by the *non-shaded* area in the figure below).



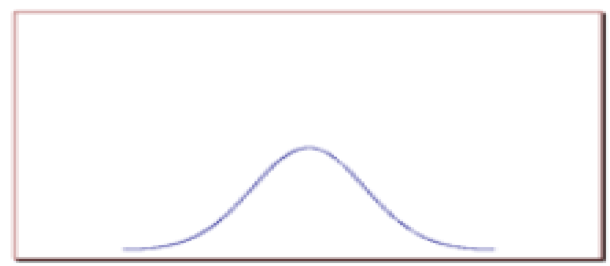
xv) The probability that *X* is less than *a* equals the area under the normal curve bounded by *a* and minus infinity (as indicated by the *shaded* area in the figure .

B.1 .3 Normal Distribution :To Remember

The graph of the normal distribution depends on two factors - the mean and the standard deviation. The mean of the distribution determines the location of the center of the graph, and the standard deviation determines the height and width of the graph. All normal distributions look like a symmetric, bell-shaped curve, as shown below.

Smaller standard deviation

Bigger standard deviation



When the standard deviation is small, the curve is tall and narrow; and when the standard deviation is big, the curve is short and wide (see above)

B. Sampling Distribution

B.1.1: BASIC TERMINOLOGY: Population

- **Population :**

  - **A population is aggregate observation of subjects grouped together by a common feature.**

  - **A population may refer to an entire group of people, objects, events, or measurements.   say ALL outgoing B.Tech Students of ITNU**

  - **The total number of elements / items / units / observations in a population is known as population size and denoted by N**

B.1.2 BASIC TERMINOLOGY: Sample

- **Sample:**

- **A sample is a random selection of members of a population. A sample is a part / fraction / subset of the population that has the characteristics of the entire population.**

- **. e.g. 100 students chosen randomly to understand their preference of higher student and job after B.Tech.**

- **The sample results are used to arrive at generalizations that are valid for entire population.**

- **The process of generalizing sample results to the population is called Statistical Inference.**

❑ Simple Random Sampling or Random Sampling

In simple random sampling, the sample is drawn in such a way that each element or unit of the population has an equal and independent chance of being included in the sample.

In simple random sampling may be classified into two types as:

- ❑ **Simple Random Sampling without Replacement (SRSWOR)**
- ❑ **Simple Random Sampling with Replacement (SRSWR)**

B.1.3 BASIC TERMINOLOGY : Simple Random Sampling or Random Sampling

- **Simple Random Sampling without Replacement (SRSWOR)**

Here, the elements or units are drawn one by one in such a way that an element or unit drawn at a time is not replaced back to the population before the subsequent draws.

In this method, the same element or unit can appear more than once in the sample.

If we draw a sample of size n from a population of size N without replacement then total number of possible samples is $^{N}C_{n}$

For example, consider a population that consists of three elements, A, B and C. Suppose we wish to draw a random sample of two elements then N = 3 and n = 2. The total number of possible random samples without replacement is $^{N}C_{n} = {}^{3}C_{2} = 3$ as (A, B), (A, C) and (B, C).

B.1.3 BASIC TERMINOLOGY : Simple Random Sampling or Random Sampling contd.

❑ Simple Random Sampling with Replacement (SRSWR)

Here, the elements or units are selected one by one in such a way that a unit drawn at a time is replaced back to the population before the subsequent draw.

In this method, the same element or unit can appear more than once in the sample and the probability of selection of a unit at each draw remains same i.e. 1/N. In this method, total number of possible samples is N^n

In above example, the total number of possible random samples with replacement is $N^n = 3^2 = 9$ as (A, A), (A, B), (A, C), (B, A), (B, B), (B, C), (C, A), (C, B) and (C, C).

B.1.4 BASIC TERMINOLOGY : Population Parameters

A population parameter is data based on an entire population.

- **Parameters are certain measures/constant which describes the characteristics of the population.**

- **For example, population mean, population variance, population, coefficient of variation, population correlation coefficient, population Proportion etc. are all parameters.**

- **Population parameter mean is usually denoted by μ and population variance denoted by σ^2**

- **In case of normal distribution, we need to know μ and σ^2 to determine the normal distribution**

B.1.5 BASIC TERMINOLOGY : Sample Statistic

- While a parameter is a characteristic of a population, a statistic is a characteristic of a sample such as sample mean, sample variance, sample proportion etc.

- A sample is drawn fro the population
- A sample statistics is a function of sample value.

- The values of statistic vary from one sample to another sample.

- Inferential statistics, we make an educated guess about a population parameter based on a sample static.

- The sample statistic is a random variable and follows a distribution called sampling distribution of the statistic.

B.1.6 BASIC TERMINOLOGY : Statistic - Sample Mean and Sample Variance

Sample Mean and Sample Variance

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population whose probability density(mass) function f(x, θ) then sample mean is defined as

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

And sample variance is defined as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

Here, we divide $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ by (n – 1) rather than n as our definition of the variance described in Unit 2 of MST-002. The reason for taking (n – 1) in place of n will become clear in the Section 5.4 of Unit 5 of this course.

B.1.6 BASIC TERMINOLOGY : Estimator and Estimate

Generally, population parameters are unknown and the whole population is too large to find out the parameters. Since the sample drawn from a population always contains some or more information about the population, therefore we guess or estimate the value of the parameter under study based on a random sample drawn from that population.

- **Estimator**

Any statistic used to estimate an unknown population parameter is known as estimator

- **Estimate**

the particular value of the estimator is known as estimate of parameter.

B.2 SAMPLING DISTRIBUTION : Example

Table 1.2: Number of Error per Typist

Typist	Number of Errors
A	4
B	2
C	3
D	1

The population mean (average number of errors) can be obtained as

$$\mu = \frac{4+2+3+1}{4} = 2.5$$

Now, let us assume that we do not know the average number of errors made by typists. So we decide to estimate the population mean on the basis of sample of size n = 2. There are N^n = 4^2 = 16 possible simple random samples with replacement of size 2.

B.2 SAMPLING DISTRIBUTION : Example contd.

All possible samples of size n = 2 and for each sample the sample mean

Sample Number	Sample in Term of Typist	Sample Observation	Sample Mean (X)
1	(A, A)	(4, 4)	4.0
2	(A, B)	(4, 2)	3.0
3	(A, C)	(4, 3)	3.5
4	(A, D)	(4, 1)	2.5
5	(B, A)	(2, 4)	3.0
6	(B, B)	(2, 2)	2.0
7	(B, C)	(2, 3)	2.5
8	(B, D)	(2, 1)	1.5
9	(C, A)	(3, 4)	3.5
10	(C, B)	(3, 2)	2.5
11	(C, C)	(3, 3)	3.0
12	(C, D)	(3, 1)	2.0
13	(D, A)	(1, 4)	2.5
14	(D, B)	(1, 2)	1.5
15	(D, C)	(1, 3)	2.0
16	(D, D)	(1, 1)	1.0

B.2 SAMPLING DISTRIBUTION : Example contd.

below.

Table 1.4: Sampling Distribution of Sample Means

S. No.	$\bar{X}$	Frequency(f)	Probability(p)
1	1.0	1	1/16 = 0.0625
2	1.5	2	2/16 = 0.1250
3	2.0	3	3/16 = 0.1875
4	2.5	4	4/16 = 0.2500
5	3.0	3	3/16 = 0.1875
6	3.5	2	2/16 = 0.1250
7	4.0	1	1/16 = 0.0625

So the arrangement of all possible values of sample mean with their corresponding probabilities is called the sampling distribution of mean.

B.3 SAMPLING DISTRIBUTION OF SAMPLE MEAN

Generally, when samples are drawn non-normal populations then it is not possible to specify the shape of the sampling distribution of mean when the sample size is small.

Although when sample size is large (30) then we observed that sampling distribution of mean converges to normal distribution whatever the form of the population i.e. normal or non-normal.

After knowing means of different samples, we may be interested to know the mean and variance of the sample means.

B.3.1 MEAN of Sample Distribution of Mean

In practice, only one random sample is actually selected and the concept of sampling distribution is used to draw the inference about the population parameters. If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a normal population with mean μ and variance σ^2 then it has also been established that sampling distribution of sample mean $\bar{X}$ is also normal. The mean and variance of sampling distribution of $\bar{X}$ can be obtained as

$$\text{Mean of } \bar{X} = E(\bar{X}) = E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right] \quad [\text{By defination of } \bar{X}]$$

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

Since $X_1, X_2, \ldots, X_n$ are randomly drawn from same population so they also follow the same distribution as the population. Therefore,

$$E(X_1) = E(X_2) = \ldots = E(X_n) = E(X) = \mu$$

and

$$\text{Var}(X_1) = \text{Var}(X_2) = \ldots = \text{Var}(X_n) = \text{Var}(X) = \sigma^2$$

Thus,

$$E(\bar{X}) = \frac{1}{n}\left(\underbrace{\mu + \mu + \ldots + \mu}_{n-\text{times}}\right)$$

$$= \frac{1}{n}(n\mu) = \mu$$

$$E(\bar{X}) = \mu$$

B.3.1 MEAN of Sample Distribution of Mean : example

- **Mean /Expected value of Sample Means**

$$\text{Mean of sample means} = \bar{\bar{X}} = \frac{1}{K}\sum_{i=1}^{k}\bar{X}_i f_i \quad \text{where, } K = \sum_{i=1}^{k} f_i$$

$$= \frac{1}{16}\left(1.0\times1 + 1.5\times2 + \ldots + 4.0\times1\right) = 2.5 = \mu$$

The mean of sample means can also be calculated as

$$E(\bar{X}) = \bar{\bar{X}} = \sum_{i=1}^{k}\bar{X}_i p_i = 1.0\times\frac{1}{16} + 1.5\times\frac{2}{16} + \ldots + 4.0\times\frac{1}{16} = 2.5$$

Thus, we have seen for this population that mean of sample means is equal to the population mean, that is, $\bar{\bar{X}} = \mu = 2.5$. The fact that these two means are

B.3.1 Standard Error of Sample Mean

- **STANDARD ERROR**

The standard deviation of a sampling distribution of a statistic is known as standard error and it is denoted by SE.

Therefore, the standard error of sample mean is given by

$$SE(\bar{X}) = \sqrt{\frac{1}{K}\sum_{i=1}^{k} f_i(\bar{x}_i - \mu)^2} \quad \text{where, } K = \sum_{i=1}^{k} f_i$$

In the previous example, we can calculate the standard error of sample mean as

$$SE(\bar{X}) = \sqrt{\frac{1}{16}\left[1\times(1.0-2.5)^2 + 2\times(1.5-2.5)^2 + \ldots + 1\times(4.0-2.5)^2\right]}$$

$$= \sqrt{\frac{1}{16}(2.25 + 2 + \ldots + 2.25)} = \sqrt{\frac{10}{16}} = 0.791$$

B.2 SAMPLING DISTRIBUTION : Example contd.

- **STANDARD ERROR contd.**

The computation of the standard error is a tedious process. There is an alternative method to compute standard error of the mean from a single sample as:

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population with mean μ and variance σ^2 then the standard errors of sample mean $(\bar{X})$ is given by

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

- **The standard error is used to express the accuracy or precision of the estimate of population parameter**

- **Standard error also confidence limits within which the population parameter may be expected to lie with certain level of confidence.**

- **Standard error is also applicable in testing of hypothesis**

C. CENTRAL LIMIT THEOREM

According to the central limit theorem, the sampling distribution of the sample means tends to normal distribution as sample size tends to large . n> 30 .First introduced by De Movers in the early eighteenth century

According to the central limit theorem, if $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population with mean μ and variance σ^2 then the sampling distribution of the sample mean tends to normal distribution with mean μ and variance σ^2/n as sample size tends to large $(n > 30)$ whatever the form of parent population, that is,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

and the variate

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

follows the normal distribution with mean 0 and variance unity, that is, the variate Z follows standard normal distribution.

C. STANDARD SAMPLING DISTRIBUTIONS

χ^2-DISTRIBUTION

t-DISTRIBUTION

F-DISTRIBUTION

Chi-square Distribution

The chi-square distribution is first discovered by Helmert in 1876 and later independently explained by Karl- Pearson in 1900. The chi-square distribution was discovered mainly as a measure of goodness of fit in case of frequency.

If a random sample $X_1, X_2, \ldots, X_n$ of size n is drawn from a normal population having mean μ and variance σ^2 then the sample variance can be defined as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

or

$$\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = (n-1)S^2 = \nu S^2$$

where, $\nu = n-1$ and the symbol ν read as 'nu'.

Thus, the variate $\chi^2 = \frac{\nu S^2}{\sigma^2}$, which is the ratio of sample variance multiplied by its degrees of freedom and the population variance follows the χ^2-distribution with ν degrees of freedom.

If $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$ and

$$Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2_{(1)}$$

where, $\chi^2_{(1)}$ read as chi-square with one degree of freedom.

In general, if X_i's (i =1, 2,…, n) are n independent normal variates with means μ_i and variances $\sigma_i^2\,(i = 1, 2, \ldots, n)$ then the sum of squares of n standard normal variate follows chi-square distribution with n df i.e.

If $X_i \sim N(\mu_i, \sigma_i^2)$, then $Z_i = \dfrac{X_i - \mu_i}{\sigma_i} \sim N(0,1)$

Therefore, $\chi^2 = \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n}\left(\dfrac{X_i-\mu_i}{\sigma_i}\right)^2 \sim \chi^2_{(n)}$

1. The probability curve of the chi-square distribution lies in the first quadrant because the range of χ^2 variate is from 0 to ∞.
2. Chi-square distribution has only one parameter n, that is, the degrees of freedom.

3. Chi-square probability curve is highly positive skewed.

4. Chi-square-distribution is a uni-modal distribution, that is, it has single mode.

5. The mean and variance of chi-square distribution with n df are n and 2n respectively

Probability Curve of χ^2-distribution

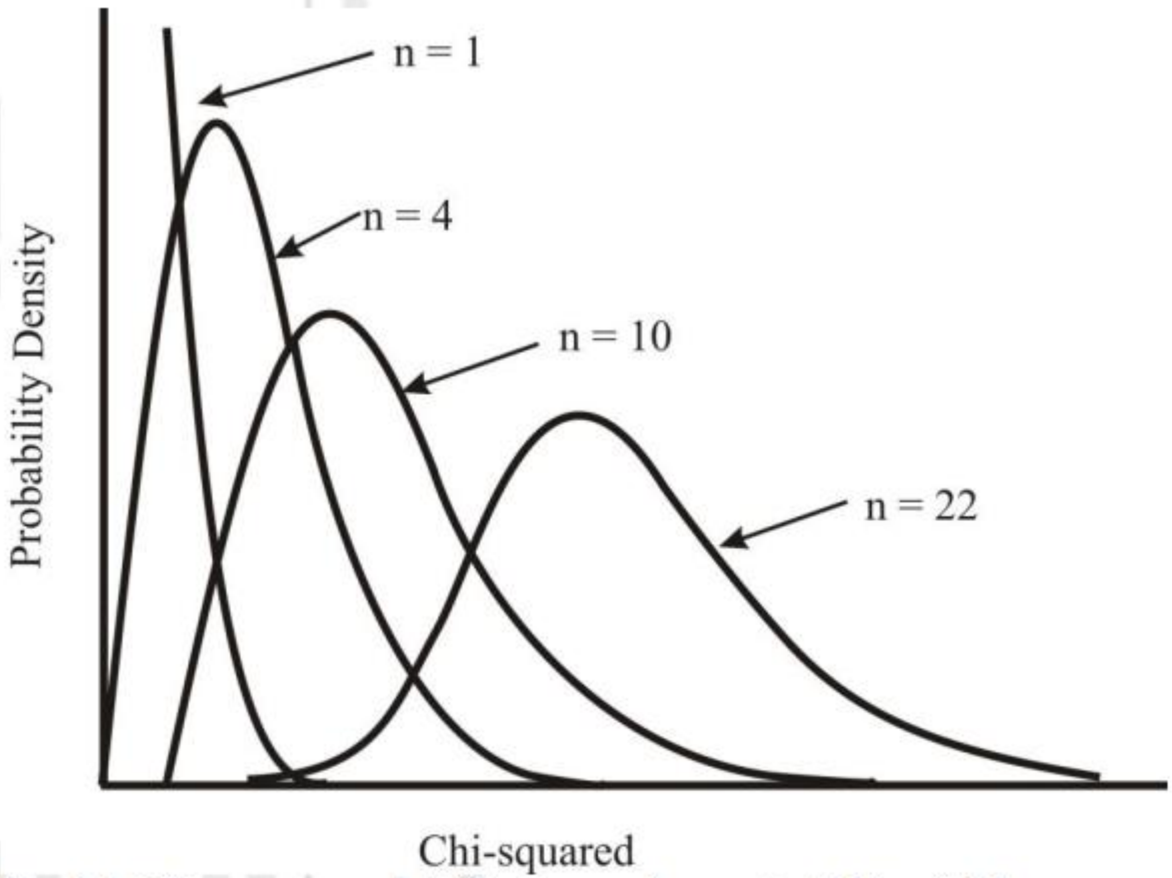


Fig. 3.1: Chi-square probability curves for n = 1, 4, 10 and 22

APPLICATIONS OF χ^2-DISTRIBUTION

The chi-square distribution is used:

1. To test the hypothetical value of population variance.

2. To test the goodness of fit, that is, to judge whether there is a discrepancy between theoretical and experimental observations.

3. To test the independence of two attributes, that is, to judge whether the two attributes are independent.

t-DISTRIBUTION

The t-distribution was discovered by W.S. Gosset in 1908. He was better known by the pseudonym **'Student'** and hence t-distribution is called **'Student's t-distribution'**.

If a random sample $X_1, X_2, \ldots, X_n$ of size n is drawn from a normal population having mean μ and variance σ^2 then we know that the sample mean $\bar{X}$ is distributed normally with mean μ and variance σ^2/n, that is, if $X_i \sim N(\mu, \sigma^2)$ then $\bar{X} \sim N(\mu, \sigma^2/n)$, and also the variate

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

is distributed normally with mean 0 and variance 1, i.e. $Z \sim N(0, 1)$.

In general, the standard deviation σ is not known and in such a situation the only alternative left is to estimate the unknown σ^2. The value of sample variance (S^2) is used to estimate the unknown σ^2 where,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2$$

Thus, in this case the variate $\frac{\bar{X} - \mu}{S/\sqrt{n}}$ is not normally distributed whereas it follows t-distribution with (n−1) df i.e.

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)} \qquad \ldots (4)$$

where, $S^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$

The t-variate is a widely used variable and its distribution is called student's t-distribution on the pseudonym name 'Student' of W.S. Gosset. The

PROPERTIES OF t-DISTRIBUTION

The t-distribution has the following properties:

1. The t-distribution is a uni-modal distribution, that is, t-distribution has single mode.
2. The mean and variance of the t-distribution with n df are zero and $\frac{n}{n-2}$ if $n > 2$ respectively.
3. The probability curve of t-distribution is similar in shape to the standard normal distribution and is symmetric about $t = 0$ line but flatter than normal curve.
4. The probability curve is bell shaped and asymptotic to the horizontal axis.

3.6.1 Probability Curve of t-distribution

The probability curve of t-distribution is bell shaped and symmetric about $t = 0$ line. The probability curves of t-distribution is shown in Fig. 3.2 at two different values of degrees of freedom as at $n = 4$ and $n = 12$.

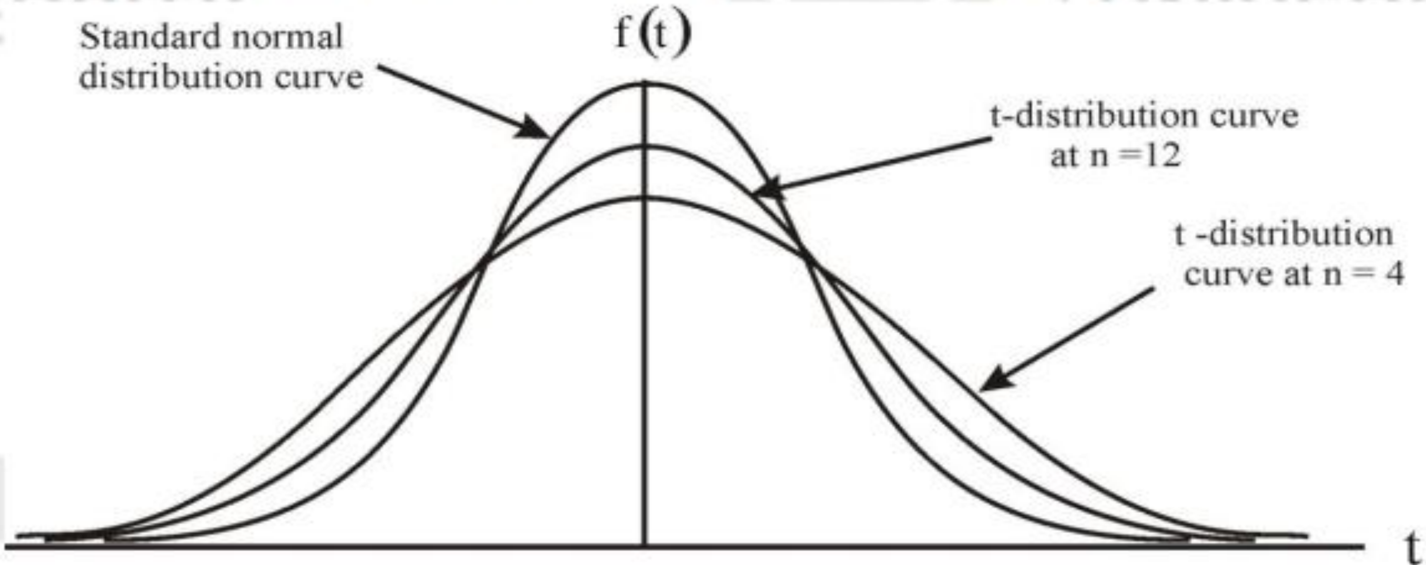


Fig. 3.2: Probability curves for t-distribution at n = 4, 12 along with standard normal curve

APPLICATIONS OF t-DISTRIBUTION

The t-distribution
is used:

1. To test the hypothesis about the population mean.
2. To test the hypothesis about the difference of two population means of two normal populations.
3. To test the hypothesis that population correlation coefficient is zero.

F-DISTRIBUTION

As we have said in previous unit that F-distribution was introduced by Prof. R. A. Fisher and defined as the ratio of two independent chi-square variates when divided by their respective degrees of freedom. If we draw a random sample $X_1, X_2, ..., X_{n_1}$ of size n_1 from a normal population with mean μ_1 and variance σ_1^2 and another independent random sample $Y_1, Y_2, ..., Y_{n_2}$ of size n_2 from another normal population with mean μ_2 and variance σ_2^2 respectively then as we have studied in Unit 3 that $\nu_1 S_1^2 / \sigma_1^2$ is distributed as chi-square variate with ν_1 df i.e.

$$\chi_1^2 = \frac{\nu_1 S_1^2}{\sigma_1^2} \sim \chi^2_{(\nu_1)} \qquad \ldots (1)$$

where, $\nu_1 = n_1 - 1$, $\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$ and $S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}\left(X_i - \bar{X}\right)^2$

Similarly, $\nu_2 S_2^2 / \sigma_2^2$ is distributed as chi-square variate with ν_2 df i.e.

$$\chi_2^2 = \frac{\nu_2 S_2^2}{\sigma_2^2} \sim \chi_{(\nu_2)}^2 \qquad \ldots (2)$$

where, $\nu_2 = n_2 - 1$, $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ and $S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}\left(Y_i - \bar{Y}\right)^2$

Now, if we take the ratio of the above chi-square variates given in equations (1) and (2), then we get

$$\frac{\chi_1^2}{\chi_2^2} = \frac{\nu_1 S_1^2/\sigma_1^2}{\nu_2 S_2^2/\sigma_2^2}$$

$$\Rightarrow \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2} \sim F_{(\nu_1, \nu_2)} \qquad \ldots (3)$$

By observing the above form given in equation (3) we reveal that the ratio of two independent chi-square variates when divided by their respective degrees of freedom follows F-distribution where, ν_1 and ν_2 are called the degrees of freedom of F-distribution.

Now, if variances of both the populations are equal i.e. $\sigma_1^2 = \sigma_2^2$ then F-variate is written in the form of ratio of two sample variances i.e.

$$F = \frac{S_1^2}{S_2^2} \sim F_{(\nu_1, \nu_2)} \qquad \ldots (4)$$

PROPERTIES OF F-DISTRIBUTION

The F-distribution has wide properties in Statistics. Some of them are as follow:

1. The probability curve of F-distribution is positively skewed curve. The curve becomes highly positive skewed when ν2 is smaller than ν1.

2. F-distribution curve extends on abscissa from 0 to ∞.

3. F-distribution is a uni-modal distribution, that is, it has single mode.

4. The square of t-variate with ν df follows F-distribution with 1 and ν deg

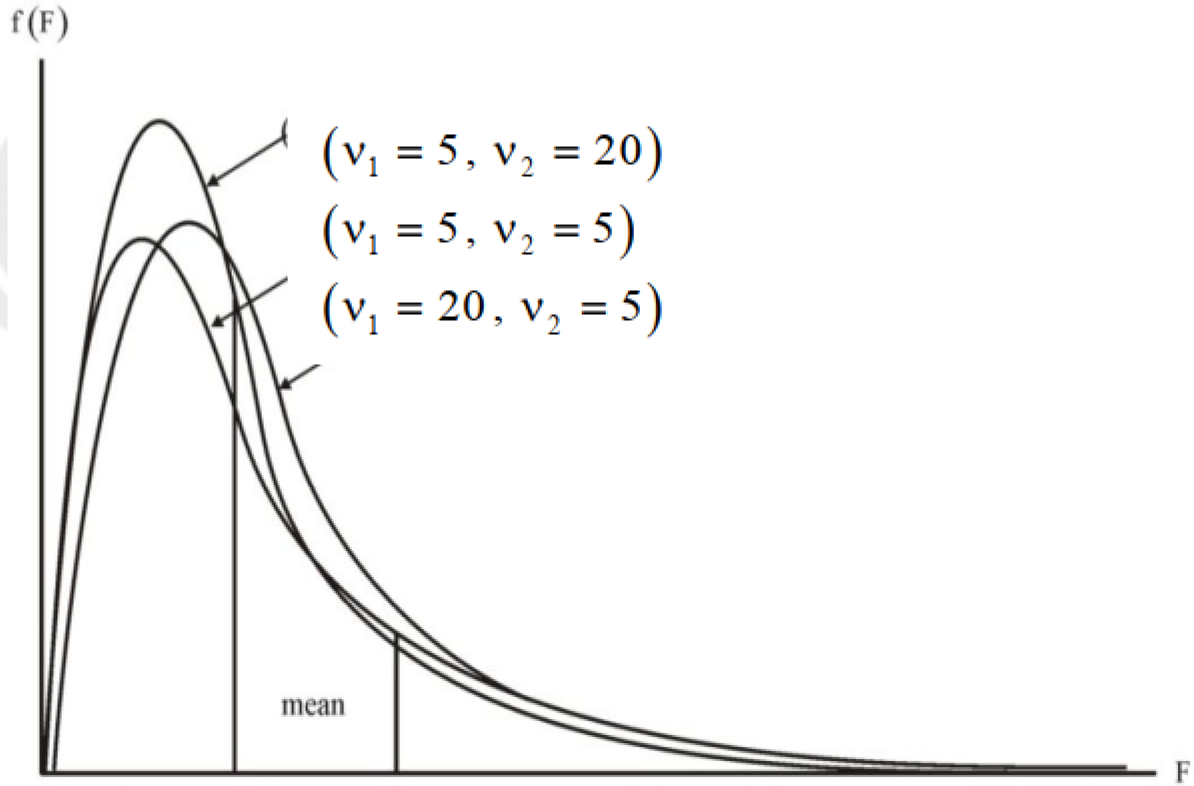


Fig. 4.1: Probability curves of F-distribution for (5, 5), (5, 20) and (20, 5) degrees of freedom.

APPLICATIONS OF F-DISTRIBUTION

The F-distribution has the following applications:
1. F-distribution is used to test the hypothesis about equality of the variances of two normal populations.

2. F-distribution is used to test the hypothesis about multiple correlation coefficients.

3. F-distribution is used to test the hypothesis about correlation ratio.
4. F-distribution is used to test the equality of means of k-populations, when one characteristic of the population is considered i.e. F-distribution is used in one-way analysis of variance.

5. F-distribution is used to test the equality of k-population means for two characteristics at a time i.e. F-distribution is used in two-way analysis of variance.

IGNOU Books

Inputs in these slides are exclusively collected from above sources.

Samir K Mahajan