UNIT 4 SAMPLING DISTRIBUTIONS

Structure **Page No.**

4.1 INTRODUCTION

In the last block we emphasised the techniques used to describe data. To illustrate these techniques we organised the data into a frequency distribution and computed various averages and measures of dispersion. Measures such as mean and standard deviation were computed to describe the central tendency of the data and the extent of its spread. In this unit we start discussing the procedures of statistical inference which is the core of statistical analysis. We have already discussed the concepts of probability and probability distributions which lays a foundation stone for statistical inference.

As we have seen in Unit 1, in statistical inference we use the sample to make inference about the unknown population characteristic. As you know, it is a common practice for every body to draw inference about a large population by examining a sample from it. This technique is used in decision making based on statistical analysis. There if we want to make decision about a large population, may be finite or infinite, we take representative samples from the population and get results for these samples which is then used to make decisions about the population. But the results we obtain from the samples may vary from sample to sample. To make valid judgements about the population, it is necessary to study this variation. In this unit we introduce you to the concept of sampling distribution. Further we discuss how to construct a sampling distribution by selecting all samples of size, say, n from a population and how this is used to make inferences about the population. To start with we study the sampling distribution of means in Sec.4.3.

The fact that different samples are likely to give different results shows that there is a possibility of error which may affect the final decision. The sampling distribution allows us to calculate this error also.

Very often, it is not easy to determine the sampling distribution exactly. In such cases we make use of a fundamental theorem in statistics known as the Central Limit Theorem. We shall only state this theorem and discuss its utility. In Sec.4.3, we also discuss sampling distribution of proportions. You will see that this distribution is useful when we take samples from a binomial population.

In Sections 4.4, 4.5 and 4.6 we shall discuss three important sampling distributions, t, χ^2 and F.

The examples and exercises in this unit are focused on how sampling techniques can assist us in making decision about various real-life problems.

Here is a list of what you should be able to do by the end of this unit.

Objectives

After reading this unit, you should be able to

- explain the need to study the sampling distribution of a statistic,
- use sampling distribution of mean and proportion to draw inferences about the population mean and population proportion,
- calculate the error in sampling using sampling distribution,
- use central limit theorem to make inferences,
- use χ^2, F and t-distributions to solve some problems of statistical inference.

4.2 POPULATION AND SAMPLES

The term population in statistics is applied to sets or collection of objects, actual or conceptual, and mainly to sets of numbers, measurements, or observations. For example, if we collect data on the number of television sets for each household in Mumbai, then a listing of number of television sets per household constitutes the population.

In some cases, such as concerning the number of television sets per household, the population is **finite**; in other cases, such as the determination of some characteristic of all units past, present, and future, that might be manufactured by a given process, it is convenient to think of the population as **infinite**. Similarly, we look upon the results obtained in a series of flips of a coin as a sample from the hypothetically infinite population consisting of all conceivably possible flips of the coin

In the last block we have seen that populations are often described by the distributions of the values, and it is a common practice to refer to a population in terms of this distribution. (For finite populations, we are referring here to the actual distribution of its values; for infinite populations, we are referring to the corresponding probability distribution or probability density.). For example, we may refer to a number of flips of a coin as a sample from a "binomial population" or to certain measurements as a sample from a "normal population." Hereafter, when referring to a "population f(x)" we shall mean a population such that its elements have a frequency distribution, a probability distribution, or a density with values given by f(x).

If a population is infinite it is impossible to observe all its values, and even if it is finite it may be impractical or uneconomical to observe it in its entirety. Thus, it is usually necessary to use a **sample**, a part of a population, and infer from it results pertaining to the entire population. Clearly, such results can be useful only if the sample is in some way "representative" of the population. It would be unreasonable, for instance, to expect useful generalisations about the population of 1984 family incomes in India on the basis of data pertaining to home owners only. Similarly, we can hardly expect reasonable generalisations about the performance of a tyre if it is tested only on smooth roads.

Here we notice that the elements in the population may not be always uniform in character. For example in the case of population of 1984, all of them may not be home owners. Like that all of them may not be having other characteristics like educated etc. Such population where the elements of the population are not having uniform characteristic is called **heterogeneous population**. Otherwise the population is called **homogeneous**. So we have to take some care while taking samples from a heterogeneous population. To assure that a sample is representative of the population from which it is obtained, and to provide a framework for the application of probability theory to problems of sampling, we shall limit our discussion to what is called a random sample. For sampling from finite populations, they are defined as follows:

Definition 1 : A set of observations $x_1, x_2, \ldots, x_n$ constitutes a random sample of size n from a finite population of size N, if it is chosen so that each subset S containing n of the N elements of the population has the same probability of being selected.

Note that this definition of randomness pertains essentially to the manner in which the samples are selected. This holds also for the following definition of a random sample from a theoretical (possibly infinite) population:

Definition 2: A set of observations $x_1, x_2, \ldots, x_n$ constitutes a random sample of size n from a population with density or mass function $f(x)$ if

1) each x_i is a realisation of a random variable whose frequency/ density function is given by $f(x)$.
2) these n random variables are independent.

There are several ways of selecting a random sample. For relatively small sample, this can be done by drawing lots, or equivalently by using a table of "random numbers" specially constructed for such purposes. We shall discuss the utility of this table in and other selection procedures in detail in Unit 12, Block 4.

Before proceeding further, why don't you try some exercises now.

E1) Suppose we want to know the average age of female students enrolled in IGNOU BDP programme. What will be the population for this study? Is this population finite or infinite?

E2) Which of the following is an appropriate sample for studying the situation given in E1.
  i) IGNOU female of BDP students selected from Delhi region.
  ii) Randomly selected female students from the list of BDP students.
  iii) Female students selected from two study centres of each regional centres.
  iv) Randomly selected students from all students registered with IGNOU.

A population is completely determined (or characterised) by certain fixed quantities (often unknown). These fixed quantities are called **parameters** and the problem consists of "inferring" about these characteristics based on a sample data.

A function of a sample observation is called a **statistic.**

For example in the case of the problem determining the average of income daily wagers in a particular city, we find average income of a sample of workers chosen. Such measures (or quantities) calculated from a sample is called a statistic.

Try this exercise, now.

E3) Give an example to show the difference between a parameter and a statistic.

Now, suppose we want to study the population using the "mean" as given in the earlier example. When we are referring to the mean of a population, we call it **population mean** and when we are referring to the mean of a sample, we call it **sample mean.**

Population mean is usually denoted by μ and sample mean is denoted by $\bar{x}$

Similarly for other measures like standard deviation, proportion etc, we have population standard deviation, sample standard deviation and population proportion, sample proportion respectively.

To understand the terms parameters and statistic in a better way, let us consider an example.

Assume that we want to draw inferences regarding the accuracy of the quantity of Milk being packed by a leading milk processing company in Western Region of India, AMUL, in 500 ml. packets.

Here the population consists of all milk packets of 500 ml. packets by AMUL company in a day. By finding out the mean of the measurements obtained from all the packets of a day's production, we get population mean. We pick up a random sample of size say n, and take the measurements. We denote the measurements as $x_1, x_2, x_3, \ldots, x_n$. Note that these measurements vary with different samples. We use $X_1, X_2, \ldots, X_n$ to denote the random variables whose particular observations are $x_1, x_2, \ldots, x_n$.
To determine the average quantity of milk being packed, we calculate the mean of the random observations, $x_1, x_2, x_3, \ldots, x_n$. This is the sample mean for this sample. We denote this mean by $\bar{x}$. $\bar{x}$ is a particular value of $\bar{X}$, where $\bar{X}$ is

$$\bar{X} = t(X_1, X_2, \ldots, X_n)$$

From the above discussion, we observe that the sample statistic, like sample mean, is a random variable. Next we shall study these random variables in detail.

4.3 WHAT IS A SAMPLING DISTRIBUTION

Let us start with situation of 'quantity of milk' discussed in the above section.

Now to study the average quantity of milk, suppose we decide that we take a sample of 10 packets without replacement and observe the quantity of milk in each packets. So here the sample size is **n=10**. The observed values are given in Column 2 Table 1 in the next page.

The mean of these 10 observations $\bar{x}_1$ is 496.01 for example 1. That is the average of the quantities of milk obtained from the sample is 496.01. Now, if we are using this sample to make judgements about the population, then we say that, the sample mean value is 496.01 gives an estimate of the average quantity of milk for the whole population.

Table 1

No.	Quantity of Milk in mls	
	Sample 1	Sample 2
1	502	501
2	501	493.9
3	499.5	499.6
4	501.05	490.03
5	499.05	500.09
6	497.56	500
7	501.06	500
8	459.3	499.3
9	499.6	497.5
10	500	502.09

But, if we take another sample from the population and observe the values as given in Column 3 Table 1, the mean of these values is 498.35 for sample 2. This is different from the mean of the first sample. Now, if we are using this sample to make inference about the population, then we get a different estimate of the average quantity of milk for the whole population. Like this we can take many different samples of size 10 and each case we get sample means which may or may not be distinct. From all these values we try to estimate the mean of the whole population. In this case it is the average amount of milk in any carton.

To give you a better understanding of generalising from sample statistic to the value of the parameter, let us look at the following example in which the size of the population is very small.

Example 1: Suppose we have a population of N=4 incomes of four business firms and we want to find the average return of these firms. The incomes (in Lakhs) are 100, 200, 300 and 400.

We first note that in this case the (population) mean income is 250 lakhs. Now we use this situation to illustrate how sample means differ from the population mean.

Suppose we select a sample of n $= 2$ observations in order to estimate the population mean μ. Now, there are $C(4, 2) = 6$ possible samples of size 2 and we will randomly be selecting one sample from this. We shall now calculate the means of these 6 different samples. These six different samples and their means are given in the following table.

Table 2

Sample	Sample elements X_i	Sample means $\overline{X}$
1	100,200	150
2	100,300	200
3	100,400	250
4	200,300	250
5	200,400	300
6	300,400	350

* * *

Now, from the table above, you can find that each sample has a different mean, with the exception of third and fourth samples. Therefore four of the six samples will result in some error in the estimation process. This sampling error is the difference between the population mean μ and the sample mean we use to estimate it.

Let us now consider the possible sample means and calculate with their probability. We assume that each sample is equally likely to be chosen. Then the probability of selecting a sample is $\frac{1}{6}$

Then we list every possible sample means and their respective possibilities in a table. (See Table 3 given below).

Table 3

Sample mean $\overline{X}$	Number of samples yielding $\overline{X}$	Probability $P(\overline{X})$
150	1	1/6
200	1	1/6
250	2	2/6
300	1	1/6
350	1	1/6
		Total 1

The table obtained above is called the sampling distribution of mean.

Definition 3: A list of all possible values for a sample statistic and the probability associated with each value is called a **sampling distribution of the statistic**.

In the above example if we take the sample mean $\bar{x}$ as the random variable, then Table 3 is nothing but the probability distribution of the means. That means, here the observed values are the means. Like any other list of numbers, these sample means have a mean. It is called **'mean of the sample means'** or the **grand mean**. The mean of the sample means is calculated in the usual fashion: the individual observations are summed, and the result is divided by the number of observation. Therefore if $\bar{\bar{X}}$ denotes this mean, then we have

$$\bar{\bar{X}} = \frac{\sum_{i=1}^{k} \bar{x}_i}{k}$$

where k is the number observation (samples). For the given situation of incomes, we can calculate $\bar{\bar{X}}$ from table as

$$\begin{aligned}\bar{\bar{X}} &= \frac{150 + 200 + 250 + 250 + 300 + 350}{6} \\ &= 250\end{aligned}$$

Notice that this equals the population mean $\mu = 250$. This is no coincidence. The grand mean $\bar{\bar{X}}$ will always equal the population mean. Thus we have arrived at an important observation.

If we were to take every possible sample of size n from a population and calculate each sample mean, the mean of those sample means would equal the population mean. That means $\bar{\bar{X}}$ is the population mean.

Now, why don't you calculate the mean $\bar{\bar{X}}$ for the income problem discussed, in Example 1, by taking samples of size 3 (see E4).

E4) Construct the sampling distribution for the income problem discussed in Example 1, by taking samples of size 3. Calculate the grand mean and compare it with the population mean.

After doing E4, you must have noticed that in the case of samples of size 3 also, we get that the mean of the sample means is equal to the population mean.

Note: You should not confuse n, the number of observations in a single sample, with k, the number of possible samples. In the situation of 4 incomes, the sample size is $n = 2$, while the number of possible samples is $k = 4c_2 = 6$.

Let us do a problem now.

Problem 1: A Statistics professor has given five tests. A student scored $70, 75, 65, 80$ and 95 respectively in five tests. The professor decides to determine his grade by randomly selecting a sample of three test scores. Construct the sampling distribution for this process. What observations might you make?

Solution: There are $^5C_3 = 10$ possible samples. The samples and their means are given in Table 4.

The sampling distribution is given in Table 5. Note that the population mean is $\mu = 77$. None of these samples gives the mean as 77. Five of the ten possible samples produce values of the sample mean $\overline{X}$ in excess of the population mean, while the other five samples underestimate it.

Table 4

Sample number	Sample elements X_i	Sample mean $\overline{X}$	Sample number	Sample elements X_i	Sample mean $\overline{X}$
1	70,75,65	70.0	6	70,80,95	81.7
2	70,75,80	75.0	7	75,65,80	73.3
3	70,75,95	80.0	8	75,65,95	78.3
4	70,65,80	71.7	9	75,80,95	83.3
5	70,65,95	76.7	10	65,80,95	80.0

Table 5

$\overline{X}$	$P(\overline{X})$
70.0	1/10
71.7	1/10
73.3	1/10
75.0	1/10
76.7	1/10
78.3	1/10
80.0	2/10
81.7	1/10
83.3	1/10
	1.00

×

You can try some exercises now.

E5) The ages of six executives of a company are

Name	Age
Mr. Ravi	54
Mrs. Veena	50
Mrs. Shanti	52
Mr. Suresh	48
Mr. Rajiv	50
Mr. Anil	52

i) How many samples of size 2 are possible?

ii) Construct the sampling distribution of means by taking samples of size 2 and organise the data.

iii) Calculate the mean of the sampling distribution and compare it with the population mean.

As we observed in the example above , it will be interesting to note how much these sample means vary from the population mean. From your knowledge of frequency distribution (see Unit 1, Block 1), you may think that this variance can be obtained by calculating the variance of the sample means. You are right. The sample means also have a variance. It measures the dispersion of the individual observations (sample means) around their population means. Furthermore this variance is calculated like any other variance. We can obtain this by performing the following.

1) the amount by which each of the observations (sample means) differs from the population mean.

2) squaring these deviations.

3) dividing the squared deviations by the number of sample means, k,

We denote by $\sigma_{\bar{X}}$, the standard deviation of the sampling distribution of sample means. Then we have

$$\sigma^2_{\bar{X}} = \frac{\sum(\bar{x}_i - \bar{\bar{X}})^2}{k} \tag{1}$$

Note that $\bar{\bar{X}}$ is the mean of the sampling distribution which is the same as the population mean.

Now that we have seen too many "standard deviations. One is the standard deviation of the entire population which we usually denote by σ. then the standard deviation of a single sample, and now we have the standard deviation $\sigma_{\bar{X}}$ of an entire set of sample means. **Since $\sigma_{\bar{X}}$ measures the dispersion of the sample means around μ, it gives a measure for the error in sampling. Because of this, $\sigma_{\bar{X}}$ is called sampling error or standard error. We formally define this now.**

Definition 3: The standard error is the standard deviation of the sample means around the population mean and is denoted by SE ($\bar{x}$).

Let us now calculate the standard errors of the sampling distribution of mean obtained in Problem 1.

Problem 2: Compute the standard error of the sampling distribution of sample scores obtained in Problem 1.

Solution: Here

$$\bar{\bar{X}} = \frac{\sum_{i=1}^{k} \bar{x}_i}{k} = 77$$

$$\sigma^2_{\bar{X}} = \frac{\sum_{i=1}^{k} \left(\bar{x}_i - \bar{\bar{X}}\right)^2}{k}$$

$$\sigma^2_{\bar{X}} = \frac{(70-77)^2 + (75-77)^2 + (80-77)^2 + (71.7-77)^2 + \ldots + (80-77)^2}{10}$$

$$= 17.6$$

$$\sigma_{\bar{X}} = \sqrt{17.6} = 4.20$$

This shows that the mean of all 10 possible sample means is 77. The square root of the average square deviation of the sample means from the population mean of 77 is 4.2.

×

You will study later in Unit 5 that the standard error of any statistic gives us an idea of how good a statistic is in estimating the parameters.

You may have realised that the computation of the standard deviation from the sampling distribution is a tedious process. There is an alternative method to compute standard error of the means, SE($\bar{x}$), from a single sample if we know the population standard deviation. By this method, we have the following formula for obtaining the standard error of the mean for a finite population.

$$SE(\bar{x}) = \sqrt{\frac{(N-n)}{N-1}\frac{(\sigma^2)}{n}}$$

where **N** = population size = the total number of individuals, **n** = sample size = number of individuals selected in the random sample. σ = standard deviation of the individuals in the population.

We will not derive the formula here since the process is too technical for the scope of this course. The factor $\sqrt{\frac{N-n}{N-1}}$ is called the finite population correction factor. As a rule of thumb, when $\frac{n}{N}$ is less than 0.1, this correction factor can be ignored. We use the above formula for computing $SE(\bar{x})$ when **N-n** is not very large. When **N** is large, relative to **n**, we use the formula,

$$SE(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

This formula can also be used for calculating $SE(\bar{x})$ for infinite population.

Let us see an example.

Problem 3: The U.S. Bureau of census wishes to estimate the birth rates per 1,00,000 people in the nation's largest cities. It is known that the standard deviation in the birth rates for these 100 urban centres is 12 births per 1,00,000 people. Then

(a) calculate the variance and standard error of the sampling distribution of i) **n** = 8 cities ii) **n** = 15 cities.

(b) compare the values obtained in both the cases.

Solution:

(a) i) Here **N** = 100 and n = 8 and the population variance is 12. Therefore $\frac{n}{N}$ is less than 0.1. Then we use the formula for calculating the variance as

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n} = \frac{12^2}{8} = 18$$

and the standard error is $\sigma_{\bar{X}} = \sqrt{18} = 4.24$

ii) In this case **N** = 100 and **n** = 15 and therefore $\frac{n}{N}$ is greater than 0.1. Also $\sigma = 12$. Therefore we use the formula for variance as

$$\begin{aligned}\sigma^2_{\bar{X}} &= \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n} \\ &= \left(\frac{100-15}{100-1}\right)\left(\frac{12^2}{15}\right) \\ &= 8.24 \\ \text{Hence, } \sigma_{\bar{X}} &= \sqrt{8.24} = 2.87\end{aligned}$$

(b) On comparing both the values we observe that, the larger sample has a smaller standard error and will tend to result in less sampling error in estimating the birth rates in the 100 cities.

———— × ————

Here is an exercise for you.

E6) From a population of 200 observations, a sample of n=50 is selected. Calculate the standard error if the population standard deviation equals 22.

Thus we have seen that we can compute the standard error of the sample means if we know the population standard deviation. Here you can note one thing. Usually, we do not know σ. But it is possible to estimate σ from the sample. We will talk about this later in the next section. Using that estimate we compute the standard error by the shortcut formula given earlier.

Now we summarise our discussion about mean and and the standard deviation of a sampling distribution. We state the following Theorem.

Theorem 1: If a random sample of size **n**, say $X_1, X_2, \ldots, X_n$ is taken from a population having the mean μ and variance σ^2, then $\overline{X}$ = the mean of $(X_1, X_2, \ldots X_n)$ is a random variable whose distribution has the mean μ. For samples from infinite populations the variance of the distribution is σ^2/n, for samples from finite population of size N, the variance is $\frac{\sigma^2}{n}\frac{N-n}{N-1}$

So far we have been discussing about the mean and standard deviation of the sampling distribution. Next we shall see some interesting properties regarding the shape of the sampling distribution. For that let us consider the following situation.

Suppose in an experiment 50 random samples of size $\mathbf{n} = 10$ are taken from a population having discrete uniform distribution.

Sampling is with replacement, so that we are sampling from an infinite population. The sample means of these 50 samples are given below.

4.4	3.2	5.0	3.5	4.1	4.4	3.6	6.5	5.3	4.4
3.1	5.3	3.8	4.3	3.3	5.0	4.9	4.8	3.1	5.3
3.0	3.0	4.6	5.8	4.6	4.0	3.7	5.2	3.7	3.8
5.3	5.5	4.8	6.4	4.9	6.5	3.5	4.5	4.9	5.3
3.6	2.7	4.0	5.0	2.6	4.2	4.4	5.6	4.7	4.3

We group these means into a distribution with the classes 2.0-2.9, 3.0-3.9, ..., and 6.0-6.9, then we get the following table.

Table 6

$\overline{X}$	Frequency
2.0-2.9	2
3.0-3.9	14
4.0-4.9	19
5.0-5.9	12
6.0-6.9	3
	50

It is clear from this distribution as well as its histogram shown in Figure 1 that the distribution of the means is fairly **bell-shaped** even though the population itself has a discrete uniform distribution.

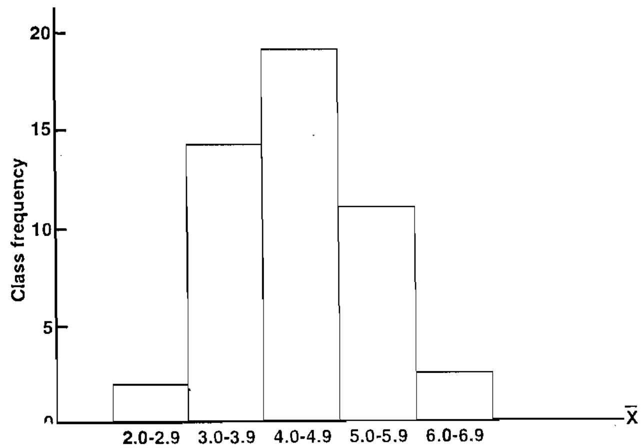

Fig. 1

This raises the question whether our result is typical of this population or any other!!

Let us consider the other situation which was discussed in the earlier section (where we want to estimate the average quantity of Milk).

Suppose in that experiment we are taking 10 random samples having size n=10 (In Block 4 we shall see how we can select random samples). The following table illustrates the 10 samples.

Table 7

1	2	3	4	5	6	7	8	9	10
502.0	501	493.9	497.8	502.09	502.09	502.9	493.9	459.3	497.3
501.0	493.9	499.6	499.3	499.6	501	501.9	493.09	497.56	493
499.5	499.6	502	499	490.9	500.09	501	499	493.0	499
501.05	490.03	501.3	501	496.8	500	500	499.6	499.05	499
499.05	500.09	500	502.9	500	500	500	499.9	499.5	499.3
497.56	500	499.9	501.9	503	499.3	499.3	500	499.6	500
501.06	500	493.99	500	502	499.6	499	500	500.0	500
459.3	499.3	500	499	501	493.9	499	501.3	501.0	501
499.6	497.5	502	500	493.9	497.5	493	502	501.03	501.9
500.0	502.09	499	493	503	490.03	497.8	502	501.06	502.9

If we plot these samples means, then we get the following graph.

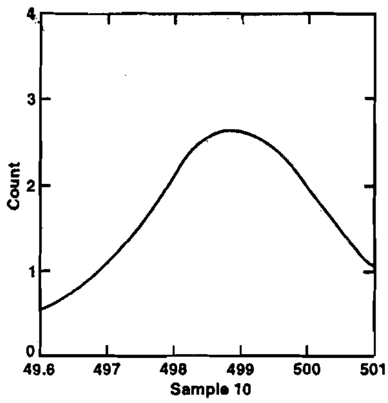

Fig.2

Here also you can observe that it is approximately bell-shaped.

Note that unlike in the earlier situation, in this case we do not know the distribution of the population.

Does the above observations give an indication that irrespective of the nature of the distribution of the population, the sampling distribution of means is approximately normal? This idea will be more clear to you if you look at the figure given at end of the unit (see Appendix 1). The figure shows graphs of the distribution function of 4 different populations see Fig. (a). The graphs in each of the figures in (b), (c) and (d) shows the sampling distribution for the sample sizes $n = 2, n = 5$ and $n = 30$ for the respective population. You note that the parent distribution for population 4 is normal and sampling distributions are also normal. This is not surprising. But the surprising fact is that for population 1, 2 and 3 parent distributions are not normal, ~~still, as~~ the sample size increases, the sampling distribution in each case approaches a normal distribution. The remarkable normality of the distribution of the sample mean is the substance of the famous central limit theorem. We state a simple version of the theorem.

Theorem 2: Central Limit Theorem : - $X_1, X_2, \ldots, X_n$ be n independent and identically distributed random variables having the same mean μ and variance σ. Let $S_n = X_1 + X_2 + \ldots + X_n$. Then $\frac{S_n - n\mu}{\sqrt{n}\sigma}$ has approximately the standard normal distribution $(0, 1)$.

Now we shall illustrate this theorem in the case of sampling distribution of means $(\overline{X})$.

Suppose we take a sample. Let $X_1, X_2, \ldots X_n$, are independent random variables by the sampling scheme and also they are identically distributed. X_i's have the same mean $\mu < \infty$ and variance $\sigma^2 < \infty$ (also called the population mean and population variance) and therefore we can apply the central limit theorem to $X_1, X_2, \ldots, X_n$ and conclude that $\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sqrt{n}\sigma}$ is distributed approximately as standard normal. But note that $\frac{X_1 + \ldots + X_n}{n}$ is nothing but the sample mean $\overline{X}$, therefore according to the central limit theorem we have $\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ follows standard normal distribution $N(0, 1)$.

So far we have been discussing the sample means and their distribution. Our interest has been in some variable which might be measured and averaged. However there are many instances where we may need some other statistic, other than mean, to make inference about the population. In the next section we shall discuss such a statistic known as 'proportion' and discuss its sampling distribution.

Sampling Distribution of Proportion

There are many instances where we are interested in determining the number or proportion of observations falling in a particular category. For example, a doctor wants to know how many of his patients can survive after administering a particular drug. A politician may want to know how many voters will not vote for him. In all these situations there are two possible outcomes of the observation - whether an observation falls in a particular category or not. You might recall that such situations are related to the problems dealing with binomial distribution. The politician may not be interested in the actual number of people who are going to vote for him, she is interested on what percentage of the people will do so. In such a situation we deal with sample proportions instead of sample means. The population parameter in this case is the proportion. We denote this by π. In general, for a finite population, we define the population proportion π as

$$\pi = \frac{k}{n}$$

where **k** is the number of observations that fall in a particular category and **n** is the total number of observation. When the population is very large, we may take samples to study the population and for each sample we calculate the sample proportion, **p**, as

$$p = \frac{s}{n}$$

where **s** denotes the number of observations in the sample which meet the particular characteristic, under study and **n** is the sample size.

For example, assume that a politician surveys 1,000 voters and finds that only 350 are going to vote for him. Then

$$p = \frac{350}{1000} = 0.35$$

You note that, a different sample of n = 1,000 voters may yield a different p. If we calculate the possible sample proportions then a **list** of these observations is called the **sampling distribution of proportions.**

Let us consider an example. As, in the case of sample means, we are considering a simple situation where the population size is very small.

Example 2: In our discussion about the quantification of milk packets, let us put the condition that if the quantity in any packet measured is less than 495ml, it will be rejected by the consumer, hence call them as defective and other wise they are non-defective.

For the data given in Table 7 we consider the first five columns (five cartons) as population then the population has 50 milk packets.

From the population we get the proportion of defective as

$$P(\text{Defective}) = \pi = \frac{7}{50} = 0.14$$

and the proportion of non-defective is

$$P(\text{non-defective}) = 1 - \pi = 0.86$$

Let us try to use the sampling method to find this proportion. Assume that random samples of size 2 out of 5 (that is two cartons out of five cartons) have been selected and the possible samples and their defective proportions are as given below

Table 8

Sample	Sample proportions (p_i)
C_1, C_2	2/20
C_1, C_3	3/20
C_1, C_4	2/20
C_1, C_5	3/20
C_2, C_3	3/20
C_2, C_4	2/20
C_2, C_5	3/20
C_3, C_4	3/20
C_3, C_5	4/20
C_4, C_5	3/20

Let us calculate the mean of the sample proportion $\bar{p}$

$$\bar{p} = \frac{\Sigma p_i}{k}$$

$$= \frac{28}{20 \times 10}$$

$$= .14$$

This is same as the population proportion.

* * *

From the earlier example we conclude the following fact:

The mean of the sampling distribution of proportions is equal to the population proportion

Now the standard error of the sample proportions, by definition, is the standard deviation of this sampling distribution. As we have mentioned earlier, in the case of mean, the computation of standard deviation from the table is a tedious process. So we make use of a short cut formula by which we can compute the **standard error** which we **denote by SE(p)**. If we know the population proportion, population size and sample size, the formula is

$$SE(p) = \sqrt{\frac{(N - n)\pi(1 - \pi)}{(N - 1).n}}$$

where π is the population proportion and N and n are the population size and sample size, respectively. If the population in infinite or if the size of the population, relative to the sample size is extremely large, then

$$SE(p) = \sqrt{\frac{\pi(1 - \pi)}{n}}.$$

Let us now calculate the standard error for the situation in Example 3. In this case $\pi = 0.4, \mathrm{N} = 5$ and n=2. Therefore,

$$\begin{aligned}\mathrm{SE(p)} &= \sqrt{\frac{(0.14)(0.86)}{2}}\sqrt{\frac{5-2}{5-1}}\\ &= 0.2454 \times 0.866\\ &= 0.2125.\end{aligned}$$

Here is an exercise for you.

E7) a) In a manufacturing process of electric bulbs, company gives an output of 250 electric bulbs per day. To test the defectiveness of the manufactured bulbs, a random sample of 50 has selected in a particular day's production. Inspection of the samples showed 4 as defective and the rest as non-defective. Calculate the standard error of sample proportion of defective.

b) If the process is considered to be a continuous one (observed for many days) and a random sample of 60 has been selected from the produced electric bulbs and observed that 7 are defective. What will be the standard error of sample proportion of defective.

So far we have discussed sampling distributions of two statistics $\overline{\mathrm{x}}$ and $\overline{\mathrm{p}}$. We also emphasised the role of Central limit theorem and the idea that distributions of $\overline{\mathrm{x}}$ and p, will be approximately normal even when the data in the parent population are not distributed normally, provided the sample sizes are large.

But there are many situations where we may have to deal with small samples, may be due to limited availability of items or by other factors such as time of cost. We may also have to deal with certain statistics other than $\overline{\mathrm{x}}$ and p. For example in some situations we may be interested in the distribution of $\overline{\mathrm{x}}/\mathrm{s}$, where s is the sample standard deviation. In some other situations, we may be interested in the distribution of $\sum_{i=1}^{n}(x_i - \overline{x})^2$. The important exact sampling distributions are chi-square, student t- and F-distributions. These distributions are widely used in statistical analysis. In the following sections we shall discuss these distributions.

4.4 t-DISTRIBUTION

William G. Gosset, a brew master for Guiness Brewarier, developed a family of distributions each of which corresponds to a parameter ν, (called nu-a positive integer). The density function of a random variable whose distribution belongs to this family is given by

$$f_\nu(y) = \frac{1}{\Gamma(\pi\nu)}\frac{\Gamma(\nu + 1/2)}{\Gamma(\nu/2)}\left(1 + \frac{y^2}{\nu}\right)^{-\nu/2}$$

where $\Gamma(x)$ for any $x > 0$ is called the gamma function. The values of $\Gamma(x)$ for different values of $x > 0$ are tabulated. The three members of the family of this distribution for $\nu = 9$, $\nu = 14$ and $\nu > 30$ are shown in the accompanying figure.(See Fig. 3 in the next page.)

Now we shall illustrate the importance of this distribution as distribution in the case of sampling distribution.

Suppose we take a sample from a population having normal distribution $N(\mu, \sigma^2)$, where μ and σ^2 are unknown. If $X_1, X_2, \ldots, X_n$ are n observations in the sample, then $X_1, \ldots, X_n$ are independently and identically distributed(i.i.d) random variables. and we have $\frac{\overline{X} - \mu}{s/\sqrt{n}}$ has t–distribution with parameter $\nu = n - 1$, where s is the sample variance. The parameter ν is called the number of degrees of freedom (in short d.f) of the distribution.

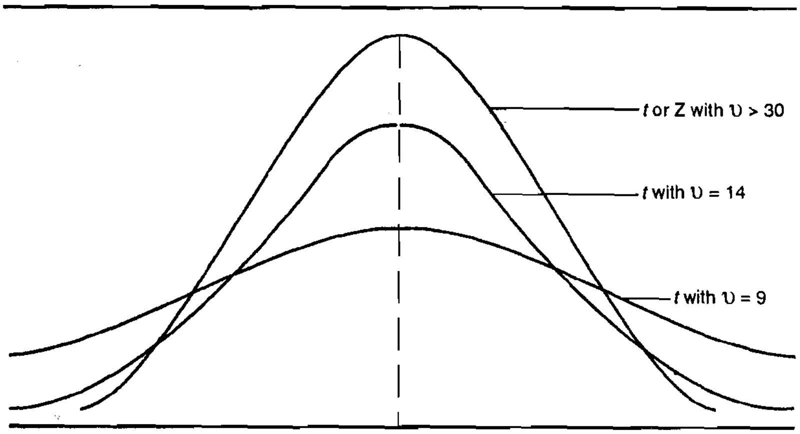

Fig. 3 t-distributions

Now that, you have been introduced to the distribution, we shall see how we use this distribution to make judgements about the population.

As we have seen in the case of normal distribution (see Unit 3, Block 1), there are tables available using which we can calculate probability for this distribution. One such table is given in the appendix (see Table 2 in appendix of this block). The table contains selected values of t_α for various values of ν, where t_α is such that the area under the t-distribution to its right is equal to α. In this table the left-hand column contains values of ν, the column headings are areas α in the right-hand tail of the t-distribution, and the entries are the values of t_α(See Fig. 4 below)

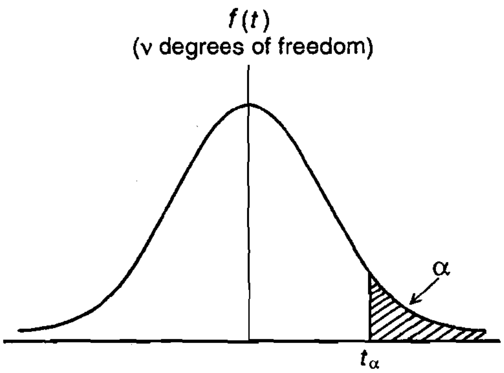

Fig. 4: Tabulated values of t

It is not necessary to tabulate the values of t_α for $\alpha > 0.50$, as it follows from the symmetry of the distribution that $t_{1-\alpha} = t_\alpha$; thus, the value of t that corresponds to a left-hand tail area of α is t_α. You note that the bottom row of the table the entries are the same as the values of the standard normal variate 2. For example $t_{0.025} = 1.96 = z_{0.025}$ for large values of ν i.e. $\nu \geq 30$.

Remark: Please note that the tables for the distribution can vary from book to book because of the definitions of parameters involved. Therefore, for the purposes of this course, please only refer to the table given at the end of the block.

Example 3: The graph of a t-distribution with 9 degrees of freedom is given below. Let us find the values of t_1, for which

i) the shaded area of the right = 0.05

ii) the total shaded area = 0.05

iii) the total unshaded area = 0.99

iv) the shaded area on the left = 0.01

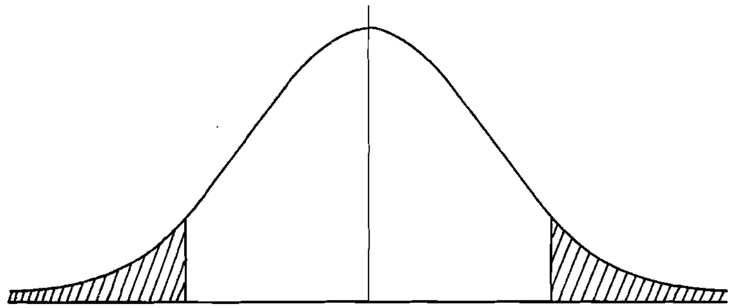

Fig. 5

Let us try (i) to (iv) one by one.

i) If the shaded area on the right is 0.05, comparing the figure above with Fig.4, we get that $\alpha = 0.05$. It is already given that $\nu = 9$. Referring to the table given in the Appendix, we proceed downward under column headed ν until entry 9 is reached. Then proceed right to the column headed 0.05. and get the required value of t as 1.833.

ii) Given that the total shaded area is 0.05. Then by symmetry, the shaded area on the right is $\frac{0.05}{2} = 0.025$ i.e. $\alpha = 0.025$. Also $\nu = 9$. Therefore from the table we get the required value as 2.262.

iii) If the total unshaded area is 0.99, then the total shaded area is $1 - 0.99 = 0.01$. Therefore the shaded area on the right of t_1 is $\frac{0.01}{2} = 0.005$. Then we get the required value from the table as 3.250.

iv) If the shaded area on the left is 0.01, then by symmetry the shaded area on the right is also 0.01. Then from the table we get the value as 2.821

* * *

Problem 4: For the given sample sizes and the t-values, find the corresponding probability α

i) $n = 26$, $t = 2.485$

ii) $n = 14$, $t = 1.771$

Solution:

i) Referring to the t-table given in the Appendix, we find the row corresponding to $\nu = 25$ and look for the entry $t = 2.485$. Then we observe that this t-value lies in the column headed $t_{0.01}$. This gives that the corresponding α-value is 0.01.

ii) Similarly you can see that α-value corresponding to $\nu = 13$ and $t = 1.771$ is 0.05.

———— × ————

Example 4: A manufacturer of fuses claims that with a 20 % overload, his fuses will blow in 12.40 minutes on the average. To test this claim, a sample of 20 of the fuses was subjected to a 20 % overload, and the times it took them to blow had a mean of 10.63 minutes and a standard deviation of 2.48 minutes. if it can be assumed that the data constitute a random sample from a normal population, do they tend to support or refute the manufacturer's claim? Let us see how we can make use of the t-distribution to find out an answer for this.

We first note that the sample is taken from a normal population and the size is small, n = 20. Then as we have stated in the discussion above, the random variable $\frac{\overline{x} - \mu}{s/\sqrt{n}}$ follows t-distribution with n − 1 degrees of freedom.

Here $\overline{x} = 10.63$, $\mu = 12.40$ and s = 2.48 and n = 20. Therefore,

$$t = \frac{10.63 - 12.40}{2.48/\sqrt{20}} = -3.19$$

Now we compare this t-value with the values of t given in the table for the parameter, $\nu = 20 - 1 = 19$. We look for the t-value which is nearest to 3.19 and less than 3.19. Then we find that the required t-value is 2.861. Also from the table, we find that the α-value corresponding to 2.861 is 0.005. That means the probability that t will exceed 2.861 is 0.005. Now we note that −3.19 is less than −2.861 therefore the α-value corresponding to −3.19 will be less than 0.005, which is very small (see Fig.6).

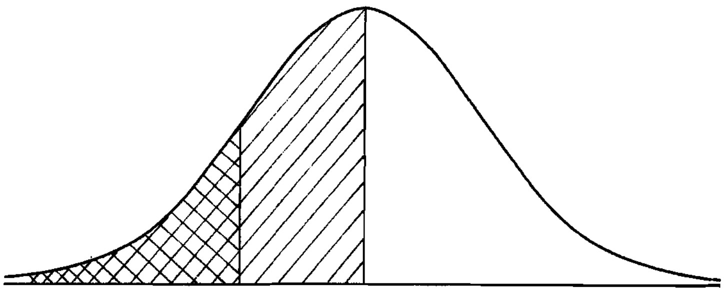

Fig. 6

Hence, we can conclude that the data tend to refute the manufacturer's claim.

* * *

Try these exercises now.

E8) Find the values of t for given values of ν and α

i) $\alpha = 0.01$ and $\nu = 19$

ii) $\alpha = 0.05$ and $\nu = 6$

iii) $\alpha = 0.025$ and $\nu = 23$

iv) $\alpha = 0.1$ and $\nu = 10$

v) $\alpha = 0.005$ and $\nu = 29$

E9) A process for making certain bearings is under control if the diameters of the bearings have a mean of 0.5 cm. What can we say about this process if a sample of 10 of these bearings has a mean diameter of 0.5060 cm. and a standard deviation of 0.0040 cm. Assume that the data constitute a random sample from a normal population.

E10) Find the values of t for which the area of the right-hand tail of the t-distribution is 0.05 if the number of degrees of freedom is equal to
(a) 16, (b) 27, (c) 200

Note that for applying t-distribution we made the assumption that the samples should come from a normal population and population standard deviation is not known. Studies have been shown later that the first assumption can be relaxed. It has been shown that the distribution of random variable with the values $t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$ is fairly close to a t-distribution even for samples from certain non-normal population. In practice, it is necessary to make sure primarily that the population from which we are sampling is approximately bell-shaped and too skewed.

Next we shall consider another exact sampling distribution.

4.5 CHI-SQUARE DISTRIBUTION

Another exact sampling distribution which is very useful in statistical problems is the chi-square distribution. In this section we shall introduce you to this distribution.

We say that a random variable Y has χ^2 (called chi-square) distribution with ν degrees of freedom if the density function $f_Y(y)$ of Y is given by

$$\begin{aligned} f_Y(y) &= \frac{y^{(\nu/2-1)}e^{-y}}{\Gamma(\nu/2)}, \text{if } y > 0 \\ &= 0, \text{if } y \leq 0 \end{aligned}$$

where $\Gamma(x)$ is the gamma function.

Note that this family of distributions is parametrised by ν, called the degrees of freedom. The graph of the density function for some members of this family ($\nu = 1, 3, 8, 10$) are shown in Fig.7 below

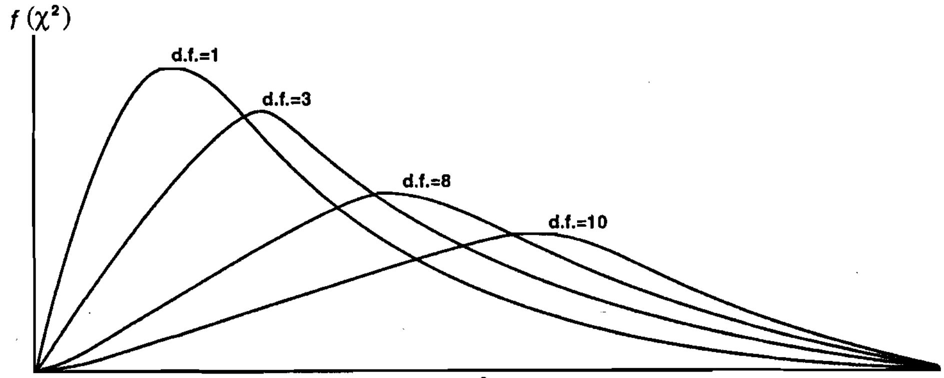

Fig.7 Various Chi-square distributions

As in the case of normal and t-distributions, a table containing selected values of χ^2_α for various values of υ, again called the number of degrees of freedom, is given in the Appendix (see Table 3). χ^2_α is the value such that the area under the chi-square distribution to its right is equal to α. (i.e. the probability that any value is greater than equal to or χ^2_α is α). In this table the left-hand column contains values of υ, the column headings are areas α in the right-hand tail of the chi-square distribution, and the entries are the values of χ^2_α. (see Fig.8 below).

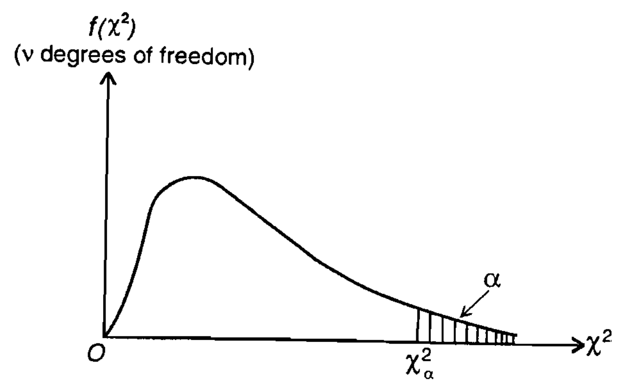

Fig.8

Note that unlike t-distribution, chi-square distribution is not symmetrical. Therefore, it is necessary to tabulate values of χ^2_α for $\alpha > 0.50$ also.

Let us see some examples

Example 5: Let us find the values of χ^2_α of the χ^2-distribution with 5 degrees of freedom for $\alpha = 0.05$ and 0.01.

Let us first consider $\alpha = 0.05$. To find the values of χ^2_α for 5 degrees of freedom and $\alpha = 0.05$, we refer to Table 3 in the Appendix. We proceed downward under column headed ν until entry 5 is reached, then proceed right to column headed $\alpha = 0.05$ which gives the required value as $\chi^2_\alpha = 11.070$.

Similarly for $\alpha = 0.01$, we get the required value as $\chi^2_\alpha = 15.056$.

* * *

Now we shall illustrate the importance of χ^2-distribution in the case of sampling distribution.

Suppose we take a random sample of size n when n is small ($n \leq 30$) from a normal population having the variance σ^2 and if s^2 is the variance of the random sample, then

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

is a value of a random variable having χ^2-distribution with $\nu = n - 1$.

Let us see how we use this result for solving the following problem.

Problem 5: An optical firm purchases glass to be ground into lenses, and it knows from past experience that the variance of the refractive index of this kind of glass is $1.26.10^{-4}$. As it is important that the various pieces of glass have nearly the same index of refraction, the firm rejects such a shipment if the sample variance of 20 pieces selected at random exceeds $2.00.10^{-4}$. Assuming that the sample values may be looked upon as a random sample from a normal population, what is the probability that a shipment will be rejected even though $\sigma^2 = 1.26.10^{-4}$?

Solution: We first note that the sample is taken from normal population and the size is n small, n = 20. Then as (we stated in the discussion above), the random variable $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ follows χ^2-distribution with $(n - 1)$ degrees of freedom.

Here $\sigma^2 = 1.26 \times 10^{-4}, s^2 = 2 \times 10^{-4}$ and n = 20. Therefore

$$\chi^2 = \frac{(10-1) \times 2 \times 10^{-4}}{1.26 \times 10^{-4}} = 30.2.$$

Now we compare this χ^2-value with the values of χ^2 given in the table for the parameter $\nu = 20 - 1 = 19$. Then we find that the χ^2-value 30.1 is close to the calculated value 30.2 and less than the calculated value. The corresponding α-value is 0.05. That means that the probability that χ^2-value exceeds 30.1 is 0.05. Therefore we can conclude that the probability that a good shipment will erroneously be rejected is less than 0.05.

———— × ————

Here is an exercise for you.

E11) Find the values of χ^2 for which the area of the right-hand tail of the χ^2-distribution is 0.5, if the number of degrees of freedom ν is equal to (a) 15, (b) 21.

E12) A random sample of 10 observations is taken from a normal population having the variance $\sigma^2 = 42.5$. Find the approximate probability of obtaining a sample standard deviation between 3.14 and 8.94.

In the next section we shall discuss another important distribution

4.6 F-DISTRIBUTION

R.A. Fisher, a British statistician, developed this distribution in the early 1920's. It is defined as follows.

Definition : If a random variable Y_1 has a χ^2-distribution with ν_1 degrees of freedom, and if a random variable Y_2 has a χ^2-distribution with ν_2 degrees of freedom and if Y_1 and Y_2 arc independent, then $Y_1/\nu_1 \div Y_2/\nu_2$ has an F-distribution with ν_1 and ν_2 degrees of freedom.

In other words, the ratio of two independent χ^2 random variables, each divided by its number of degrees of freedom, is an F random variable.

Like the t and χ^2 distributions, the F distribution is in reality a family of probability distributions, each corresponding to certain numbers of degrees of freedom. But unlike the t and χ^2 distributions, the F distribution has two numbers of degrees of freedom, not one. Figure shows the F distribution with 2 and 9 degrees of freedom. As you can see, the F distribution is skewed to the right. However, as both numbers of degrees of freedom become very large, the F distribution tends toward normality. As in the case of the χ^2 distribution, the probability that an F random variable is negative is zero. This must be true since an F random variable is a ratio of two nonnegative numbers. (Y_1/ν_1 and Y_2/ν_2 are both nonnegative.) Once again, it should be emphasised that any F random variable has two numbers of degrees of freedom. Be careful to keep these numbers of degrees of freedom in the correct order, because an F distribution with ν_1 and ν_2 degrees of freedom is not the same as an F distribution with ν_2 and ν_1 degrees of freedom.

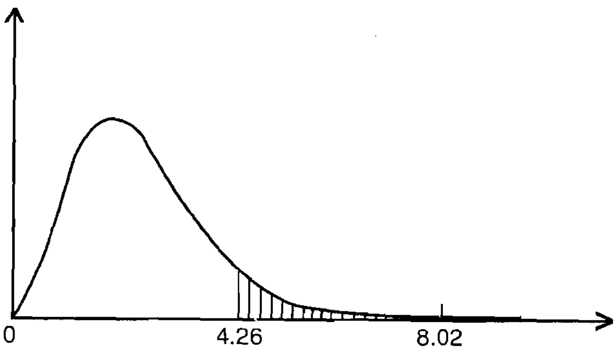

Fig.9

Tables are available which show the values of F that are exceeded with certain probabilities, such as .05 and .01. Appendix Table 9 shows, for various numbers of degrees of freedom, the value of F that is exceeded with probability equal to .05. For example, if the numbers of degrees of freedom are 2 and 9, the value of F that is exceeded with probability equal to .05 is 4.26.(See Fig.9.) Similarly, Appendix Table 10 shows, for various numbers of degrees of freedom, the value of F that is exceeded with probability equal to .01. For example, if the numbers of degrees of freedom are 2 and 9, the value of F exceeded with probability equal to .01 is 8.02.(See Fig.9.)

Let us see some examples.

Example 6: A random variable has F distribution with 40 and 30 degrees of freedom, let us find the probability that it will exceed a) 1.79 b) 2.30.

We first consider the value 1.79. We look at table of F-distribution, Table 4 for 0.05. We proceed downward under column headed ν_2 until entry 30 is reached, then proceed right to the column headed $\nu_1 = 40$ which shows the given value 1.79. Therefore we get that the probability is 0.05.

Since the value cannot be in Table 4 we look at Table 5 which is for 0.01. We proceed similarly and find the value 2.30 in the table. Therefore we get that the probability is 0.01.

* * *

We shall now illustrate the importance of this distribution in the case of sampling distribution.

Suppose that we have two independent random samples of sizes n_1 and n_2 respectively, taken from two normal population having the same variance. If s_1^2 and s_2^2 are sample variances, then the function

$$F = \frac{s_1^2}{s_2^2}$$

has F distribution with ν_1 and ν_2 degrees of freedom.

F-distribution is used to compare the equality of two sample variances.

Let us consider the following problem.

Problem 6: If two independent random samples of size $n_1 = 7$ and $n_2 = 13$ are taken from a normal population, what is the probability that the variance of the first sample will be at least three times as large as that of the second sample?

Solution: In this we have to compare the variances of the samples. Therefore we apply F-distribution. From Table we find that $F_{.05} = 3.00$ for $\upsilon_1 = 7 - 1 = 6$ and $\upsilon_2 = 13 - 1 = 12$; thus, the desired probability is 0.05.

———— × ————

Why don't you try these exercises now.

E13) A random variable has the F distribution with 15 and 12 degrees of freedom. What is the value of this random variable that is exceeded with a probability of 0.5 with a probability of .01?

E14) If a random variable F has 15 and 23 degrees of freedom, what is the probability that it will exceed 2.93.

E15) If two independent random samples of size $n_1 = 9$ and $n_2 = 16$ are taken from a normal population, what is the probability that the variance of the first sample will be at least four times as large as the variance of the second sample?

E16) If independent random samples of size $n_1 = n_2 = 8$ come from normal populations having the same variance, what is the probability that either sample variance will be at least seven times as large as the other?

In this unit you have only been introduced to these three specific distributions t, χ^2 and F. You will realise the utility of these distribution as you study the following units.

This brings us to the end of this unit. We have discussed the sampling distribution of a statistic at length. Let's now briefly recall the various concepts which we have covered here.

4.7 SUMMARY

In this unit we have explained the following concepts

1\) i) Parameter - Values that describe the characteristic of a population.

ii) Statistic - Measures describing the characteristic of a sample.

iii) Random sample - A sample from a population in which all the items in the population have an equal chance of being in the sample.

iv) Sampling distribution of a statistic - For a given population, a probability distribution of all the possible values of a statistic may taken as for a given sample size.

v) Standard error - The standard deviation of the sampling distribution of a statistic.

2) We have discussed the central limit theorem.

3) We have introduced the following three specific distributions

i) t-distribution

ii) χ^2-distribution

iii) F-distribution

and explained the use of these distributions as sampling distribution.

4.8 SOLUTIONS/ANSWERS

E1) Female students enrolled in IGNOU BDP. Finite population

E2) (ii) and (iii)

E3) Suppose there are 500 employees working in a company and we are interested to find the average salary paid to them in a month, the average amount calculated from the enter employees is called parameter. On the other hand it is possible to estimate the average based on selected few employees. The average income calculated from sample is called statistic.

E4) The sampling distribution is given in the following table:

Table 9

Sample No.	Sample elements	Sample Means
1	100, 200, 300	200
2	100, 200, 400	233 (approximately)
3	100, 300, 400	267 (approximately)
4	200, 300, 400	300

The grand mean is 250 and is equal to the population mean.

E5) i) There are

$$^6C_2 = \frac{6 \times 5}{1 \times 2} = \frac{30}{2} = 15$$

ii) samples of size 2 are given in the following table.

Table 10

Sample	$\overline{X}_i$	Sample	$\overline{X}_i$	Sample	$\overline{X}_i$
54, 50	52	50, 52	51	52, 50	51
54, 52	53	50, 48	49	52, 52	52
54, 48	51	50, 50	50	48, 50	49
54, 50	52	50, 52	51	48, 52	50
54, 52	53	52, 48	50	50, 52	51

iii) The population mean, the grand mean,

$\mu = 51.00$

$\overline{X} = \frac{\Sigma \overline{x}_i}{k} = 51.000$

The grand mean is equal to the population mean.

E6) N = 200

n = 50

$\sigma = 22$

$$SE(\overline{X}) = \sigma/\sqrt{n} \cdot \sqrt{\frac{N-n}{N-1}}$$

$$22/\sqrt{50}\sqrt{\frac{200-50}{199}}$$

$= 2.7012$

E7) a) $$SE(p) = \sqrt{\frac{250-50}{250-1} \times \frac{\frac{4}{50} \times \frac{46}{50}}{50}}$$

$= 0.03438.$

b) $$SE(p) = \sqrt{\frac{\frac{7}{60} \times \frac{53}{60}}{60}}$$

$= 0.04144.$

E8) $t_{0.01} = 3.143$

$t_{0.025} = 2.069$

$t_{0.1} = 1.372$

$t_{0.005} = 2.756$

E9) We find the value of the random variable

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

where $\overline{x} = 0.5060$, $\mu = 0.5$ and $s = 0.0040$ and $n = 10$. Substituting, we get the value of t as t = 4.7434. From the table we get that the t-value for the parameter $\nu = 9$ and $\alpha = .05$ is 2.26 which is less than the calculate value t = 4.7434. Since the probability is less i.e. 0.05, we conclude that the process is out of control.

E10) From the table we find that corresponding to $\alpha = 0.05$ and the parameter

a) $\nu = 16$ the value is 1.75

b) $\nu = 27$ the value is 1.70

c) $\nu = 200$ is 1.645 (note that this value corresponds to the entry to the last row marked ∞).

E11) From the table in the appendix, we find that the value of χ^2_α for $\alpha = 0.5$ is

a) 24.996 for $\nu = 15$ and

b) 32.671 for $\nu = 21$.

E12) We make use of the formula

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}.$$

Note that we are given that $\sigma^2 = 42.5$ and n = 10. Let s_1 and s_2 denote the two sample standard deviations, then $s_1^2 = 3.14$ and $s_2^2 = 8.94$ corresponding χ^2 values are

$$\chi_1^2 = \frac{(n-1) \times 3.14}{42.5}$$

$$= \frac{9 \times 3.14}{42.5} = 0.664$$

$$\text{and } \chi_2^2 = \frac{9 \times 8.94}{42.5} = 1.893$$

To find the required approximate probability, it is enough to find the area α_0 given below.

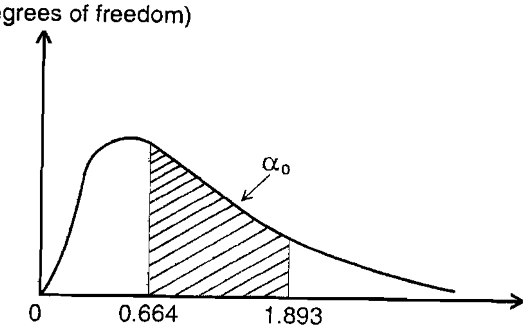

Fig. 10

From the table we find those values of χ^2 for 9 degrees freedom which are close to 0.664 and 1.893.

E13) Table 4 shows that the answer to the first question is 2.62 and Table 5 shows that the answer to the second question is 3.84.

E14) We first look at Table 4 and find that the given value is not there. Then it must be in Table 5 (check it!). Therefore the probability is 0.01.

E15) Solution: Here $n_1 = 9$ and $n_2 = 16 \therefore \nu_1 = 8$ and $\nu_2 = 15$. We have

$$F = \frac{s_1^2}{s_1^2} = 4$$

To find the probability we look for value F = 4 in both the tables and find that it is in Table 5. Therefore the probability is 0.01.

E16) $n_1 = n_2 = 8 \therefore \nu_1 = \nu_2 = 7$ and given that

$$F = \frac{s_1^2}{s_2^2} = 7$$

Proceeding similarly we find the value 6.99 in Table 5 which is almost equal to 7. Therefore the probability is 0.01.