UNIT 2 SAMPLING DISTRIBUTION(S) OF STATISTIC(S)

Structure

2.1 Introduction
Objectives
2.2 Sampling Distribution of Sample Mean
2.3 Sampling Distribution of Difference of Two Sample Means
2.4 Sampling Distribution of Sample Proportion
2.5 Sampling Distribution of Difference of Two Sample Proportions
2.6 Sampling Distribution of Sample Variance
2.7 Sampling Distribution of Ratio of Two Sample Variances
2.8 Summary
2.9 Solutions/Answers

2.1 INTRODUCTION

In previous unit, we have discussed the concept of sampling distribution of a statistic. There are so many problems in business and economics where it becomes necessary to draw inference about population parameters such as population mean, population proportion, population variance, difference of two population means, etc. For example, one may want to estimate the average height of the students in a college, a businessman may want to estimate the proportion of defective items in a production lot, a manufacturer of car tyres may want to estimate the variance of the diameter of tyres, a pharmacist may want to estimate the difference of effect of two types of drugs, etc.

Generally, sample mean is used to draw inference about the population mean. Similarly, sample proportion and sample variance are used to draw inference about the population proportion and population variance respectively. Therefore, it becomes necessary to know the sampling distribution of sample mean, sample proportion and sample variance, etc.

This unit is divided in 9 sections. Section 2.1 is introductive in nature. One of the most important sample statistics which is used to draw conclusion about population mean is sample mean. So in Section 2.2, the sampling distribution of mean is described, whereas in Section 2.3, the sampling distribution of difference of two sample means is explored. Second useful statistic generally used in Social and Business problems is sample proportion. The sampling distribution of sample proportion is described in Section 2.4 and in Section 2.5 sampling distribution of difference of two sample proportions is explored. Sometimes, it is useful to estimate the variation of population. For this we use sampling distribution of sample variance. The sampling distribution of sample variance is described in Section 2.6 whereas the sampling distribution of ratio of two sample variances is given in Section 2.7. Unit ends by providing summary of what we have discussed in this unit in Section 2.8 and solution of exercises in Section 2.9.

Objectives

After studying this unit, you should be able to:

- describe the need of sampling distributions of different statistics;

- explain the sampling distribution of sample mean;
- explore the sampling distribution of difference of two sample means;
- describe the sampling distribution of sample proportion;
- explore the sampling distribution of difference of two sample proportions;
- explain the sampling distribution of sample variance; and
- describe the sampling distribution of ratio of two sample variances.

2.2 SAMPLING DISTRIBUTION OF SAMPLE MEAN

One of the most important sample statistics which is used to draw conclusion about the population mean is sample mean. For example, an investigator may want to estimate average income of the peoples living in a particular geographical area, a product manager may want to estimate the average life of electric bulbs manufactured by a company, a pathologist may want to estimate the mean time required to complete a certain analysis, etc. In the above cases an estimate of the population mean is required and one may estimate this on the basis of a sample taken from that population. For this, sampling distribution of sample mean is required.

We have already given you the flavor of sampling distribution of sample mean with the help of an example in Section 1.3 of previous unit in which we draw all possible samples of same size from the population and calculate the sample mean for each sample. After calculating the value of sample mean for each sample we observed that the values of sample mean vary from sample to sample. Then the sample mean is treated as random variable and a probability distribution is constructed for the values of sample mean. This probability distribution is known as sampling distribution of sample mean. Therefore, the sampling distribution of sample mean can be defined as:

The sampling distribution of sample means is a theoretical probability distribution of sample means that would be obtained by drawing all possible samples of the same size from the population.

“The probability distribution of all possible values of sample mean that would be obtained by drawing all possible samples of the same size from the population is called sampling distribution of sample mean or simply says sampling distribution of mean.”

Now, the question may arise in your mind “what is the shape of sampling distribution of mean, it is normal, exponential, binomial, Poisson, etc.?”

With the help of Fig. 1.1 of previous unit, we have described the shape of the sampling distribution of mean for different populations and for varying sample sizes. From this figure, we observed that when the parent population is normal then all the sampling distributions for varying sample sizes are also normal whereas when parent population is uniform, binomial, exponential then the shapes of the sampling distributions of mean are not in the form of specify distribution when sample size is small ($n = 2$ or 5).

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 the shapes of the sampling distribution of mean in different situations, you may be interested to know the mean and variance of the sampling distribution of mean when samples are drawn from normal population.

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 $X_1, X_2, ..., X_n$ 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

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

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

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,

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

and

$$Var(X_1) = Var(X_2) = ... = Var(X_n) = Var(X) = \sigma^2$$

Thus,

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

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

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

and variance

$$Var(\bar{X}) = Var\left[\frac{1}{n}(X_1 + X_2 + ... + X_n)\right]$$

$$= \frac{1}{n^2}\left[Var(X_1) + Var(X_2) + ... + Var(X_n)\right]$$

$$= \frac{1}{n^2}\left(\underbrace{\sigma^2 + \sigma^2 + ... + \sigma^2}_{n-\text{times}}\right) = \frac{1}{n^2}\left(n\sigma^2\right)$$

$$Var\left(\bar{X}\right) = \frac{\sigma^2}{n}$$

Hence, we conclude that if the samples are drawn from normal population with mean μ and variance σ^2 then the sampling distribution of mean $\bar{X}$ is also normal distribution with mean μ and variance σ^2/n, that is,

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

The standard error (SE) of sample mean can be obtained by the definition of SE as

$$SE\left(\bar{X}\right) = SD\left(\bar{X}\right) = \sqrt{Var\left(\bar{X}\right)} = \frac{\sigma}{\sqrt{n}}$$

Note 1: The statistical inference is based on the theoretical distribution so in whole statistical inference, we assume that the random sample is selected from the infinite population or from a finite population with replacement.

The statistical inference is based on the theoretical distribution so if we say that a random sample, say, $X_1, X_2, ..., X_n$ is taken from a population its means that the random sample is selected from the infinite population or from a finite population with replacement such that X_i's are independent and followed same population distribution.

If X and Y are two independent random variables and a & b are two constant then by the addition theorem of expectation, we have

$$E\left(aX + bY\right) = aE\left(X\right) + bE\left(Y\right)$$

and

$$Var\left(aX + bY\right) = a^2 Var\left(X\right) + b^2 Var\left(Y\right)$$

If the sampling is done without replacement from a finite normal population of size N then the sample mean is distributed normally with mean

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

and variance

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

The proof of this formula is beyond the scope of this course.

Therefore, standard error of sample mean is given by

$$SE\left(\bar{X}\right) = \sqrt{\operatorname{Var}\left(\bar{X}\right)} = \sqrt{\frac{N-n}{N-1}\,\frac{\sigma^2}{n}}$$

If the sample is drawn from any population other than normal then by central limit theorem, the sampling distribution of $\bar{X}$ tends to normal as the sample size n increases, that means, the sample mean $\bar{X}$ is normally distributed with mean μ and variance σ^2/n when the sample size is large (> 30).

Let us see an application of the sampling distribution of mean with the help of an example.

Example 1: Diameter of a steel ball bearing produced on a semi-automatic machine is known to be distributed normally with mean 12 cm and standard deviation 0.1 cm. If we take a random sample of size 10 then find

(i) Mean and variance of sampling distribution of mean.

(ii) The probability that the sample mean lies between 11.95 cm and 12.05 cm.

Solution: Here, we are given that

$$\mu = 12,\ \sigma = 0.1,\ n = 10$$

(i) Since the population is normally distributed therefore the sampling distribution of the sample mean also follows a normal distribution with mean

$$E\left(\bar{X}\right) = \mu = 12$$

and variance

$$\operatorname{Var}\left(\bar{X}\right) = \frac{\sigma^2}{n} = \frac{(0.1)^2}{10} = 0.001$$

(ii) The probability that the sample mean lies between 11.95 cm and 12.05 cm is given by

$$P\left[11.95 < \bar{X} < 12.05\right] \qquad \left[\text{See Fig.2.1}\right]$$

To get the value of this probability, we convert $\bar{X}$ into a standard normal variate Z by the transformation

$$Z = \frac{\bar{X} - E\left(\bar{X}\right)}{\sqrt{\operatorname{Var}\left(\bar{X}\right)}} = \frac{\bar{X}-12}{\sqrt{0.001}} = \frac{\bar{X}-12}{0.03}$$

Therefore, by subtracting 12 from each term and then dividing each term by 0.03 in the above inequality, we get required probability as

If X ~
Z = X
normal
'0' and
Z = X
Simila
$\bar{X}$ ~ N
then

$$P\left[\frac{11.95-12}{0.03}<\frac{\bar{X}-12}{0.03}<\frac{12.05-12}{0.03}\right]$$

$$=P[-1.67<Z<1.67]$$

$$=P[-1.67<Z<0]+P[0<Z<1.67]$$

$$=P[0<Z<1.67]+P[0<Z<1.67]$$

$$=2P[0<Z<1.67]$$

$$=2\times 0.4525 \qquad \left[\begin{array}{l}\text{Using table for area}\\ \text{under normal curve}\end{array}\right]$$

$$= 0.9050$$

Therefore, the probability that the sample mean lies between 11.95 and 12.05 cm is 0.9050.

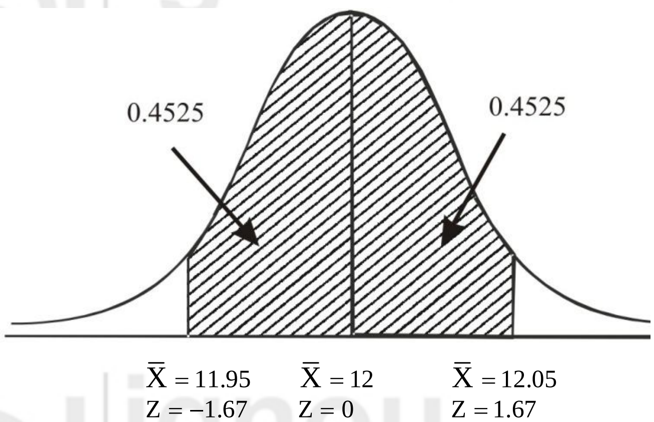

Fig. 2.1

Now, continuing our discussion about the sampling distribution of mean, we consider another situation.

In the cases described in last two pages, we assume that variance σ^2 of the normal population is known but in general it 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 σ^2 where,

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

Thus, in this case, it has been established that the sampling distribution of mean is not normal and the variate $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ is known as t-variate and follows t-distribution with (n−1) df instead of normal distribution. This distribution has wild applications in Statistics therefore it will be described in detail in Unit 3 of this course.

If n is sufficiently large (>30) then we know that almost all the distributions are very closely approximated by normal distribution. Thus in this case, t-distribution is also approximated normal distribution. So for large sample size (>30), the sampling distribution of mean also follows normal distribution with mean μ and variance S^2/n.

Now, try some exercises for your practice.

E1) If parent population is normal with mean μ and variance σ^2 then, what is sampling distribution of sample means?

E2) The weight of certain type of a truck tyre is known to be distributed normally with mean 200 pounds and standard deviation 4 pounds. A random sample of 10 tyres is selected.

(i) What is the sampling distribution of sample mean? Also obtain the mean and variance of this distribution.

(ii) Find the probability that the mean of this sample is greater than or equal to 202 pounds.

E3) The mean life of electric blubs produced by a company is 2550 hours. A random sample of 100 bulbs is selected and the standard deviation is found to be 54 hours. Find the mean and variance of the sampling distribution of mean.

2.3 SAMPLING DISTRIBUTION OF DIFFERENCE OF TWO SAMPLE MEANS

There are so many problems in business and economics where someone may be interested to draw the inference about the difference of two population means. For example, two manufacturing companies of blubs are produced same type of bulbs and one may be interested to know which one is better than the other, an investigator may want to know the difference of average income of the peoples living in a two cities, say, A and B, two different types of drugs, were tried on certain number of patients for controlling blood pressure and one may be interested to know which one has better effect on controlling blood pressure, etc. Therefore, in such situation, to draw the inference we require the sampling distribution of difference of two sample means.

The sa
of diff
means
sampli
differe
$(\bar{x}-\bar{y}$
obtain
possib
the pop

Let the same characteristic measures from two populations be represented by X and Y variables and the variation in the values of these constitute two population, say, population-I for variation in X and population-II for variation in Y. Suppose population-I having mean μ_1 and variance σ_1^2 and population-II having mean μ_2 and variance σ_2^2. Then we take all possible samples of same size n_1 from population-I and then the sample mean, say, $\bar{X}$ is calculated for each sample. Similarly, all possible samples of same size n_2 are taken from the population-II and the sample mean, say, $\bar{Y}$ is calculated for each sample. Then we consider all possible differences of means $\bar{X}$ and $\bar{Y}$. The difference of these means may or may not differ from sample to sample and so we construct the probability distribution of these differences. The probability distribution thus obtained is known as sampling distribution of the difference of sample means. Therefore, the sampling distribution of difference of sample means can be defined as:

"The probability distribution of all values of the difference of two sample means would be obtained by drawing all possible samples from both the populations is called sampling distribution of difference of two sample means."

If both the parent populations are normal, that is,

$$X \sim N\left(\mu_1, \sigma_1^2\right) \text{ and } Y \sim N\left(\mu_2, \sigma_2^2\right)$$

then as we discussed in previous section that

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \text{ and } \bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

Thus, as we have discussed in Unit 4 of MST-003 that if two independent random variables X and Y are normally distributed then the difference $(X-Y)$ of these also normally distributed. Therefore, in our case, the sampling distribution of difference of two sample means $(\bar{X}-\bar{Y})$ also follows normal distribution with mean

$$E\left(\bar{X}-\bar{Y}\right) = E\left(\bar{X}\right) - E\left(\bar{Y}\right) = \mu_1 - \mu_2$$

and variance

$$\operatorname{Var}\left(\bar{X}-\bar{Y}\right) = \operatorname{Var}\left(\bar{X}\right) + \operatorname{Var}\left(\bar{Y}\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

If X and Y are independent random variables then

$E\left(X-Y\right) = E\left(X\right) - E\left(Y\right)$

and

$\operatorname{Var}\left(X-Y\right)$

$= \operatorname{Var}\left(X\right) + \operatorname{Var}\left(Y\right)$

Therefore, standard error of difference of two sample means is given by

$$SE\left(\bar{X}-\bar{Y}\right) = \sqrt{\operatorname{Var}\left(\bar{X}-\bar{Y}\right)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

If the parent populations are not normal but n_1 and n_2 are large (> 30), then by central limit theorem, the sampling distribution of $(\bar{X}-\bar{Y})$ is very closely normally distributed with mean $(\mu_1 - \mu_2)$ and variance $\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$.

Let us see an application of the sampling distribution of difference of two sample means with the help of an example.

Example 2: Electric CFL manufactured by company A have mean lifetime of 2400 hours with standard deviation 200 hours, while CFL manufactured by company B have mean lifetime of 2200 hours with standard deviation of 100 hours. If random samples of 125 electric CFL of each company are tested, find

(i) The mean and standard error of the sampling distribution of the difference of mean lifetime of electric CFLs.

(ii) The probability that the CFLs of company A will have a mean lifetime at least 160 hours more than the mean lifetime of the CFLs of company B.

Solution: Here, we are given that

$$\mu_1 = 2400, \sigma_1 = 200, \mu_2 = 2200, \sigma_2 = 100$$

$$n_1 = n_2 = 125$$

(i) Let $\bar{X}$ and $\bar{Y}$ denote the mean lifetime of CFLs taken from companies A and B respectively. Since n_1 and n_2 are large (n_1, $n_2 > 30$) therefore, by the central limit theorem, the sampling distribution of $(\bar{X}-\bar{Y})$ follows normal distribution with mean

$$E(\bar{X}-\bar{Y}) = \mu_1 - \mu_2 = 2400 - 2200 = 200$$

and variance

$$\operatorname{Var}(\bar{X}-\bar{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$= \frac{(200)^2}{125} + \frac{(100)^2}{125} = 400$$

Therefore, the standard error is given by

$$SE(\bar{X} - \bar{Y}) = \sqrt{Var(\bar{X} - \bar{Y})} = \sqrt{400} = 20$$

(ii) Here, we want to find out the probability which is given by

$$P\left[\bar{X} \geq 160 + \bar{Y}\right] = P\left[(\bar{X} - \bar{Y}) \geq 160\right] \qquad [\text{See Fig. } 2.2]$$

To get the value of this probability, we convert variate $(\bar{X} - \bar{Y})$ into a standard normal variate Z by the transformation

$$Z = \frac{(\bar{X} - \bar{Y}) - E(\bar{X} - \bar{Y})}{\sqrt{Var(\bar{X} - \bar{Y})}} = \frac{(\bar{X} - \bar{Y}) - 200}{\sqrt{400}} = \frac{(\bar{X} - \bar{Y}) - 200}{20}$$

Therefore, by subtracting 200 from each term and then dividing each term by 20 in the above inequality, we get the required probability

$$P\left[\frac{(\bar{X} - \bar{Y}) - 200}{20} \geq \frac{160 - 200}{20}\right] = P[Z \geq -2]$$

$$= P[-2 < Z < 0] + P[0 < Z < \infty]$$

$$= P[0 < Z < 2] + P[0 < Z < \infty]$$

$$= 0.4772 + 0.5 = 0.9772$$

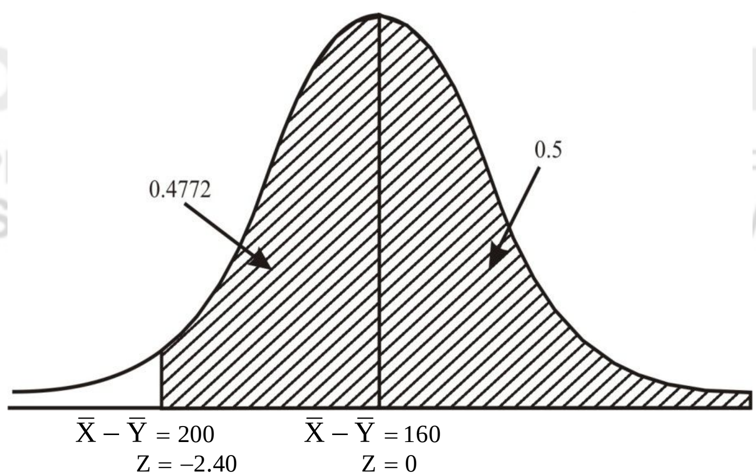

Fig. 2.2

Now, continuing our discussion about the sampling distribution of difference of two means, we consider another situation.

If population variances $\sigma_1^2 \,\&\, \sigma_2^2$ are unknown then we estimate $\sigma_1^2 \,\&\, \sigma_2^2$ by the values of the sample variances $S_1^2 \,\&\, S_2^2$ of the samples taken from the first and second population respectively. And for large sample sizes n_1 and n_2 (> 30), the sampling distribution of $(\bar{X} - \bar{Y})$ is very closely normally distributed with mean $(\mu_1 - \mu_2)$ and variance $\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)$.

If population variances $\sigma_1^2 \& \sigma_2^2$ are unknown and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ then σ^2 is estimated by pooled sample variance S_p^2 where,

$$S_p^2 = \frac{1}{n_1 + n_2 - 2}\left[n_1 S_1^2 + n_2 S_2^2\right]$$

and variate

$$t = \frac{\left(\bar{X} - \bar{Y}\right) - \left(\mu_1 - \mu_2\right)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n_1+n_2-2)}$$

follows t-distribution with $(n_1 + n_2 - 2)$ degrees of freedom. As similar as sampling distribution of mean, for large sample sizes n_1 and n_2 (> 30) the sampling distribution of $(\bar{X} - \bar{Y})$ is very closely normally distributed with mean $(\mu_1 - \mu_2)$ and variance $S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$.

Now, you can try the following exercise.

E4) The average height of all the workers in a hospital is known to be 68 inches with a standard deviation of 2.3 inches whereas the average height of all the workers in a company is known to be 65 inches with a standard deviation of 2.5 inches. If a sample of 35 hospital workers and a sample of 50 company workers are selected at random, what is the probability that the sample mean of height of hospital workers is at least 2 inch greater than that of company workers?

2.4 SAMPLING DISTRIBUTION OF SAMPLE PROPORTION

In Section 2.2, we have discussed the sampling distribution of sample mean. But in many real word situations, in business and other areas where the data collected in form of counts or the collected data classified into two categories or groups according to an attribute. For example, the peoples living in a colony may be classified into two groups (male and female) with respect to the characteristic sex, the patients in a hospital may be classified into two groups as cancer and non-cancer patients, the lot of articles may be classified as defective and non-defective, etc.

Generally, such types of data are considered in terms of proportion of elements / individuals / units / items possess (success) or not possess (failure) a given characteristic or attribute. For example, the proportion of female in the population, proportion of cancer patents in a hospital, proportion of defective articles in a lot, etc.

In such situations, we deal with population proportion instead of population mean.

When population proportion is unknown and the whole population is too large to find out the proportion. In such a situation, to draw the inference about the population proportion, we require the sampling distribution of sample proportion.

For sampling distribution of sample proportion, we draw all possible samples from the population and for each sample we calculate the sample proportion p as

$$p = \frac{X}{n} \le 1$$

where, X is the number of observations /individuals / items / units in the sample which have the particular characteristic under study and n is the total number of observations in the sample i.e. sample size.

For better understanding of the process, we consider the following example in which size of the population is very small:

Suppose, there is a lot of 3 cartons A, B & C of electric bulbs and each carton contains 20 bulbs. The number of defective bulbs in each carton is given below:

Table 2.1: Number of Defective Bulbs per Carton

Carton	Number of Defectives Bulbs
A	2
B	4
C	1

The population proportion of defective bulbs can be obtained as

$$P = \frac{2+4+1}{20+20+20} = \frac{7}{60}$$

Now, let us assume that we do not know the population proportion of defective bulbs. So we decide to estimate population proportion of defective bulbs on the basis of samples of size n = 2. There are $N^n = 3^2 = 9$ possible samples of size 2 with replacement. The all possible samples and their respective proportion defectives are given in the following table:

Table 2.2: Calculation of Sample Proportion

Sample	Sample Carton	Sample Observation	Sample Proportion(p)
1	(A, A)	(2, 2)	4/40
2	(A, B)	(2, 4)	6/40
3	(A, C)	(2, 1)	3/40
4	(B, A)	(4, 2)	6/40
5	(B, B)	(4, 4)	8/40
6	(B, C)	(4, 1)	5/40
7	(C, A)	(1, 2)	3/40
8	(C, B)	(1, 4)	5/40
9	(C, C)	(1, 1)	2/40

From the above table, we can see that value of the sample proportion is varying from sample to sample. So we consider all possible sample proportions and calculate their probability of occurrence. Since there are 9 possible samples therefore the probability of selecting a sample is 1/9. Then we arrange the possible sample proportion with their respective probability in Table 2.3 given in next page:

Table 2.3: Sampling Distribution of Sample Proportion

S.No.	Sample Proportion(p)	Frequency	Probability
1	2/40	1	1/9
2	3/40	2	2/9
3	4/40	1	1/9
4	5/40	2	2/9
5	6/40	2	2/9
6	8/40	1	1/9

This distribution is called the sampling distribution of sample proportion. Thus, we can define the sampling distribution of sample proportion as:

"The probability distribution of all values of the sample proportion that obtained by drawing all possible samples of same size from the population is called the sampling distribution of the sample proportion."

The sampling distribution of the sample proportion is a theoretical probability distribution of sample proportions that would be obtained by drawing all possible samples of the same size from the population.

This distribution also has its mean. Therefore, the mean of sampling distribution of sample proportion can be obtained as

$$\overline{p} = \frac{1}{K}\sum_{i=1}^{k} p_i f_i \quad \text{where, } K = \sum_{i=1}^{k} f_i$$

$$= \frac{1}{9}\left(\frac{2}{40}\times 1 + \frac{3}{40}\times 2 + ... + \frac{8}{40}\times 1\right) = \frac{7}{60}$$

Thus, we have seen that mean of sample proportion is equal to the population proportion.

As we have already mentioned in the previous unit that finding mean, variance and standard error from this process is tedious so we calculate these by another short-cut method when population proportion is known.

As we discussed in Unit 5 of course MST-003 that if a population whose elements are divided into two mutually exclusive groups− one containing the elements which possess a certain attribute (success) and other containing elements which do not possess the attribute(failure), then number of successes(elements possess a certain attribute) follows a binomial distribution with mean

$$E(X) = nP$$

and variance

$$Var(X) = nPQ \qquad \text{where } \; Q = 1 - P$$

where, P is the probability or proportion of success in the population.

Now, we can easily find the mean and variance of the sampling distribution of sample proportion by using the above expression as

$$E(p) = E\left(\frac{X}{n}\right)$$

$$= \frac{1}{n}E(X) = \frac{1}{n}nP$$

$$E(p) = P$$

and variance

$$\text{Var}(p) = \text{Var}\left(\frac{X}{n}\right)$$

$$= \frac{1}{n^2}\text{Var}(X) \qquad \left[\because \text{Var}(aX) = a^2\text{Var}(X)\right]$$

$$= \frac{1}{n^2}nPQ \qquad \left[\because \text{Var}(X) = nPQ\right]$$

$$\text{Var}(p) = \frac{PQ}{n}$$

Also standard error of sample proportion can be obtained as

$$SE(p) = \sqrt{\text{Var}(p)} = \sqrt{\frac{PQ}{n}}$$

If the sampling is done without replacement from a finite population then the mean and variance of sample proportion is given by

$$E(p) = P$$

and variance

$$\text{Var}(p) = \frac{N-n}{N-1}\frac{PQ}{n}$$

where, N is the population size and the factor (N-n) / (N-1) is called finite population correction.

If sample size is sufficiently large, such that np > 5 and nq > 5 then by central limit theorem, the sampling distribution of sample proportion p is approximately normally distributed with mean P and variance PQ/n where, Q = 1– P.

Let us see an application of the sampling distribution proportion with the help of an example.

Example 3: A machine produces a large number of items of which 15% are found to be defective. If a random sample of 200 items is taken from the population and sample proportion is calculated then find

(i) Mean and standard error of sampling distribution of proportion.

(ii) The probability that less than or equal to 12% defectives are found in the sample.

Solution: Here, we are given that

$$P = \frac{15}{100} = 0.15, \; n = 200$$

(i) We know that when sample size is sufficiently large, such that np > 5 and nq > 5 then sample proportion p is approximately normally distributed with mean P and variance PQ/n where, Q = 1– P. But here the sample proportion is not given so we assume that the conditions of normality hold, that is, np > 5 and nq > 5. So mean of sampling distribution of sample proportion is given by

$$E(p) = P = 0.15$$

and variance

$$\text{Var}(p) = \frac{PQ}{n} = \frac{0.15 \times 0.85}{200} = 0.0006$$

Therefore, the standard error is given by

$$SE(p) = \sqrt{\text{Var}(p)} = \sqrt{0.0006} = 0.025$$

(ii) The probability that the sample proportion will be less than or equal to 12% defectives is given by

$$P[p \le 0.12] \qquad [\text{See Fig. 2.3}]$$

To get the value of this probability, we can convert the random variate p into standard normal variate Z by the transformation

$$Z = \frac{p - E(p)}{\sqrt{\text{Var}(p)}} = \frac{p - 0.15}{\sqrt{0.0006}} = \frac{p - 0.15}{0.025}$$

Therefore, by subtracting 0.15 from each term and then dividing each term by 0.025 in the above inequality, we get required probability

$$P\left[\frac{p - 0.15}{0.025} \le \frac{0.12 - 0.15}{0.025}\right] = P[Z \le -1.20]$$

$$= 0.5 - P[-1.20 < Z \le 0]$$

$$= 0.5 - P[0 \le Z < 1.20]$$

$$= 0.5 - 0.3849$$

$$= 0.1151$$

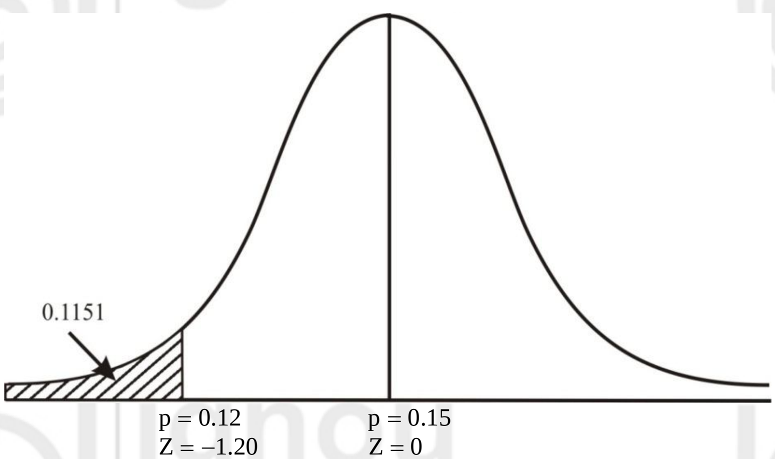

Fig. 2.3

Now, it is time for you to try the following exercise to make sure that you have learnt the sampling distribution of sample proportion.

E5) A state introduced a policy to give loan to unemployed doctors to start own clinic. Out of 10000 unemployed doctors 7000 accept the policy and got the loan. A sample of 100 unemployed doctors is taken at the time of allotment of loan. What is the probability that sample proportion would have exceeded 60% acceptance.

2.5 SAMPLING DISTRIBUTION OF DIFFERENCE OF TWO SAMPLE PROPORTIONS

In Section 2.3, we have discussed the sampling distribution of difference of two sample means. In some cases, we are interested to comparative study that the proportions of an attributes in the two different populations or group are equal or not. For example, one may wish to test whether the proportions of alcohol drinkers in the two cities are same, one may wish to compare the proportions of literates between two groups of people, one may estimate the decrease in the proportions of consumption of tea after the increase in excise duty on it, etc. Therefore, we require the sampling distribution of difference of two sample proportions for getting the inference about two populations.

Suppose, there are two populations, say, population-I and population-II under study and the population-I having population proportion P_1 and population-II having population proportion P_2 according to an attribute. For describing the sampling distribution of difference of two sample proportions, we consider all possible samples of same size n_1 taken from population-I and for each sample calculate the sample proportion p_1 of success. Similarly, determine the sample proportion p_2 of success by considering all possible sample of same size n_2 from population-II. Then we consider all possible differences of proportions p_1 and p_2. The difference of these proportions may or may not be differ so we construct the probability distribution of these differences. The probability distribution thus obtained is called the sampling distribution of the difference of sample proportions. Therefore, the sampling distribution of difference of two sample proportions can be defined as:

"The probability distribution of all values of the difference of two sample proportions that have been obtained by drawing all possible samples of same sizes from both the populations is called sampling distribution of difference between two sample proportions."

The sa
differe
propor
sampli
differe
propor
would
drawin
from b

As we have seen in case of single proportion described in previous section that if sample size is sufficiently large, such that np > 5 and nq > 5 then by central limit theorem, the sampling distribution of sample proportion p is approximately normally distributed with mean P and variance PQ/n where, Q = 1– P. Therefore, if n_1 and n_2 are sufficiently large, such that $n_1p_1 > 5$, $n_1q_1 > 5$, $n_2p_2 > 5$ and $n_2q_2 > 5$ then

$$p_1 \sim N\left(P_1, \frac{P_1Q_1}{n_1}\right) \text{ and } p_2 \sim N\left(P_2, \frac{P_2Q_2}{n_2}\right)$$

where, $Q_1 = 1 - P_1$ and $Q_2 = 1 - P_2$.

Also, by the property of normal distribution described in Unit 13 of MST 003, the sampling distribution of the difference of sample proportions follows normal distribution with mean

$$E(p_1 - p_2) = E(p_1) - E(p_2) = P_1 - P_2$$

and variance

$$Var(p_1 - p_2) = Var(p_1) + Var(p_2) = \frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}$$

That is,

$$p_1 - p_2 \sim N\left(P_1 - P_2, \frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}\right)$$

Thus, standard error is given by

$$SE(p_1 - p_2) = \sqrt{Var(p_1 - p_2)} = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$$

Let us see an application of the sampling distribution of difference of two sample proportions with the help of an example.

Example 4: In one population, 30% persons had blue-eyed and in second population 20% had the same blue-eye. A random sample of 200 persons is taken from each population independently and calculate the sample proportion for both samples, then find the probability that the difference in sample proportions is less than or equal to 0.02.

Solution: Here, we are given that

$P_1 = 0.30$, $P_2 = 0.20$,

$n_1 = n_2 = 200$

Let p_1 and p_2 be the sample proportions of blue-eye persons in the samples taken from both the populations respectively. We know that when n_1 and n_2 are sufficiently large, such that $n_1p_1 > 5$, $n_1q_1 > 5$, $n_2p_2 > 5$ and $n_2q_2 > 5$ then sampling distribution of the difference between two sample proportions is approximately normally distributed. But here the sample proportions are not given so we assume that the conditions of normality hold. So mean of sampling distribution of the difference between two sample proportions is given by

$$E(p_1 - p_2) = P_1 - P_2 = 0.30 - 0.20 = 0.10$$

and variance

$$Var(p_1 - p_2) = \frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}$$

$$= \frac{0.30 \times 0.70}{200} + \frac{0.20 \times 0.80}{200}$$

$$= 0.0019$$

The probability that the difference in sample proportions is less than or equal to 0.02 is given by

$$P\left[p_1 - p_2 \le 0.02\right] \qquad \left[\text{See Fig.2.4}\right]$$

To get the value of this probability, we convert variate (p_1-p_2) into a standard normal variate Z by the transformation

$$Z = \frac{(p_1 - p_2) - E(p_1 - p_2)}{\sqrt{Var(p_1 - p_2)}} = \frac{(p_1 - p_2) - 0.10}{\sqrt{0.0019}} = \frac{(p_1 - p_2) - 0.10}{0.04}$$

Therefore, by subtracting 0.10 from each term and then dividing each term by 0.04 in the above inequality, we get the required probability as

$$P\left[\frac{(p_1 - p_2) - 0.10}{0.04} \le \frac{0.02 - 0.10}{0.04}\right] = P[Z \le -2.0]$$

$$= 0.5 - P[-2.0 \le Z \le 0]$$

$$= 0.5 - P[0 \le Z \le 2.0]$$

$$= 0.5 - 0.4772 = 0.0228$$

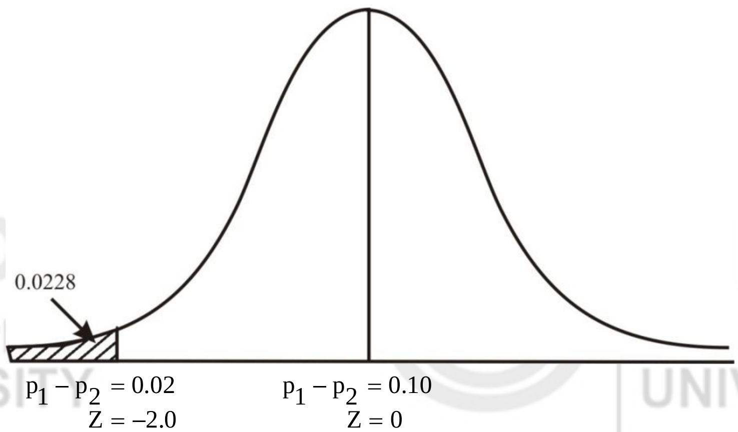

Fig. 2.4

Now, try to solve the following exercises to ensure that you understand this section properly.

E6) In city A, 12% persons were found to be alcohol drinkers and in another city B, 10% persons were found alcohol drinkers. If 100 persons of city A and 150 persons of city B are selected randomly, what is the probability that the difference in sample proportions lies between 0.01 and 0.08?

2.6 SAMPLING DISTRIBUTION OF SAMPLE VARIANCE

In Sections 2.2 and 2.4, we have discussed the sampling distributions of sample mean and sample proportion respectively. But many practical situations concerned with the variability. For example, a manufacturer of steel ball bearings may want to know about the variation of diameter of steel ball bearing, a life insurance company may be interested in the variation of the number of polices in different years, etc. Therefore, we need information about the sampling distribution of sample variance.

For describing the sampling distribution of the sample variance, we consider all possible sample of same size, say, n taken from the population having variance σ^2 and for each sample we calculate sample variance S^2. The values of S^2 may vary from sample to sample so we construct the probability distribution of sample variances. The probability distribution thus obtained is known as sampling distribution of the sample variance. Therefore, the sampling distribution of sample variance can be defined as:

"The probability distribution of all values of the sample variance would be obtained by drawing all possible sample of same size from the parent population is called the sampling distribution of the sample variance."

The
the

Let the characteristic measures from the population be represented by X variable and the variation in the values of X constitute a population. Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from normal population with mean μ and variance σ^2. W.S. Gosset first described the sampling distribution of sample variance S^2. Here, it is to be noted that S^2 always be positive, therefore S^2 does not follow normal distribution because the normal variate takes positive as well as negative values. The distribution of sample variance can not be obtained directly therefore in this case, some transformation is made by multiplying S^2 by $(n-1)$ and then dividing the product by σ^2. The obtained new variable is denoted by the symbol χ^2 read as chi-square and follows the chi-square distribution with $(n-1)$ degrees of freedom as,

For degrees of freedom, please see Section 3.2 of next unit.

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

where, S^2 is called sample variance and given below

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \quad \text{for all } i = 1, 2, ..., n$$

The chi-square distribution has wild applications in statistics, therefore, it will be described in detail in the next unit of this course.

For solving the problems relating to sampling distribution of variance, we convert the variate S^2 in a χ^2-variate by above transformation. For better understanding, we explain it by taking an example given below.

Example 5: A manufacturer of steel ball bearings has found the variance of diameter of ball bearings 0.0018 inches^2. What is the probability that a random sample of 25 ball bearings will result in a sample variance at least 0.002 inches^2?

Solution: Here, we are given that

$\sigma^2 = 0.0018$, n = 25

The probability that the sample variance is at least 0.002 inches^2 is given by

$$P\left[S^2 \geq 0.002\right]$$

To get the value of this probability, we convert variate S^2 into χ^2-variate which follows the chi-square distribution with n - 1 = 24 degrees of freedom by the transformation

$$\chi^2_{(n-1)} = \frac{(n-1)S^2}{\sigma^2} = \frac{24S^2}{0.0018}$$

Therefore, multiplying each term by 24 and then dividing each term by 0.0018 in the above inequality. We get the required probability as

$$P\left[\frac{24S^2}{0.0018} \geq \frac{24 \times 0.002}{0.0018}\right] = P\left[\chi^2_{(24)} \geq 26.67\right]$$

From the chi-square table, we see that $P\left[\chi^2_{(24)} \geq 15.66\right] = 0.90$ and

$P\left[\chi^2_{(24)} \geq 33.20\right] = 0.10.$

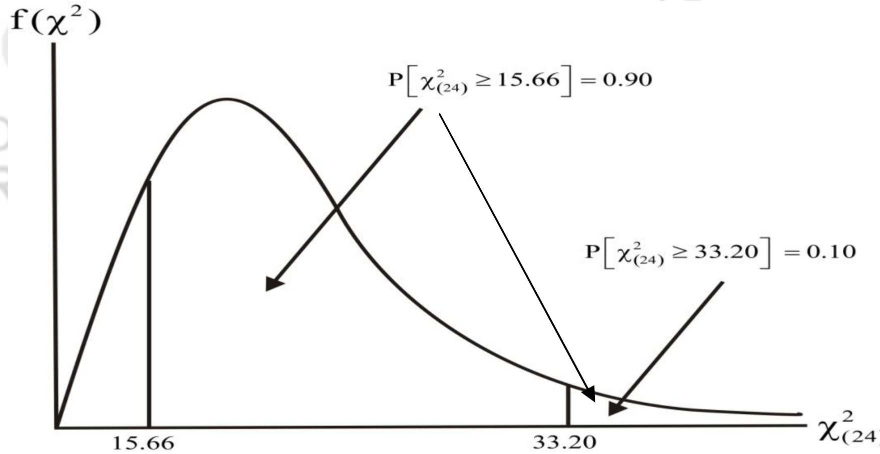

Fig. 2.5

Therefore, the above probability must lie between 0.10 and 0.90, that is,

$$0.10 < P\left[\chi^2 \geq 26.67\right] = P\left(S^2 \geq 0.002\right) < 0.90$$

Now, you can try the following exercises.

E7) The weight of certain type of truck tyres has a variance of 11 pounds^2. A random sample of 20 tyres is selected. What is the probability that the variance of this sample is greater than or equal to 16 pounds^2?

2.7 SAMPLING DISTRIBUTION OF RATIO OF TWO SAMPLE VARIANCES

In previous section, we have discussed about the sampling distribution of sample variance where we were dealing with one population. Now, one may be interested to persue for the comparative study of the variances of the two normal populations as we have done for the means and the proportions. For example, suppose a quality control engineer wants to determine whether or not the variance of the quality of the product is changing over the time, an economist may wish to know whether the variability in incomes differs across the two populations, etc. For solving such types of problems we require the sampling distribution of the ratio of sample variances.

Suppose there are two normal populations, say, population-I and population-II under study, the population-I has variance σ_1^2 and population-II has variance σ_2^2. For describing the sampling distribution for ratio of population variances, we consider all possible samples of same size n_1 from population-I and for each sample we calculate sample variance S_1^2. Similarly, calculate sample variance S_2^2 from each sample of same size n_2 drawn from population-II.

Then we consider all possible values of the ratio of the variances S_1^2 and S_2^2, The values of S_1^2/S_2^2 may vary from sample to sample so we construct the probability distribution of the ratio of the sample variances. The probability distribution thus obtained is known as sampling distribution of the ratio of sample variances. Therefore, the sampling distribution of ratio of sample variances can be defined as:

"The probability distribution of all values of the ratio of two sample variances would be obtained by drawing all possible samples from both the populations is called sampling distribution of ratio of sample variances."

The sampling distribution of ratio of variances is given by Prof. R. A. Fisher in 1924. According to Prof. R. A. Fisher, the ratio of two independent chi-square variates when divided by their respective degrees of freedom follows F-distribution as

$$F = \frac{\chi^2_{(n_1-1)}/(n_1-1)}{\chi^2_{(n_2-1)}/(n_2-1)}$$

Since $\chi^2 = \frac{(n-1)S^2}{\sigma^2}$ therefore,

$$F = \frac{(n_1-1)S_1^2/\sigma_1^2(n_1-1)}{(n_2-1)S_2^2/\sigma_2^2(n_2-1)}$$

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{(n_1-1,\, n_2-1)}$$

If $\sigma_1^2 = \sigma_2^2$ then

$$F = \frac{S_1^2}{S_2^2} \sim F_{(n_1-1,\, n_2-1)}$$

Therefore, the sampling distribution of ratio of sample variances follows F-distribution with $(n_1 - 1, n_2 - 1)$ degrees of freedom. The F-distribution is described in detail in fourth unit of this block.

With this we end this unit. We now summarise our discussion.

2.8 SUMMARY

In this unit, we have discussed following points:

1. The need of sampling distributions of different statistics.
2. The sampling distribution of sample mean.
3. The sampling distribution of difference between two sample means.
4. The sampling distribution of sample proportion.
5. The sampling distribution of difference of two sample proportions.
6. The sampling distribution of sample variance.
7. The sampling distribution of ratio of two sample variances.

2.9 SOLUTIONS / ANSWERS

E1) Since parent population is normal so sampling distribution of sample means is also normal with mean μ and variance σ^2/n.

E2) Here, we are given that

$$\mu = 200, \sigma = 4, n = 10$$

(i) Since parent population is normal so sampling distribution of sample means is also normal. Therefore, the mean of this distribution is given by

$$E\left(\bar{X}\right) = \mu = 200$$

and variance

$$\operatorname{Var}\left(\bar{X}\right) = \frac{\sigma^2}{n} = \frac{(4)^2}{10} = \frac{16}{10} = 1.6$$

(ii) Now, the probability that the sample mean $\bar{X}$ is greater than or equal to 202 pounds is given by

$$P\left[\bar{X} \geq 202\right] \qquad \left[\text{See Fig. 2.6}\right]$$

To get the value of this probability, we convert $\bar{X}$ into a standard normal variate Z by the transformation

$$Z = \frac{\bar{X} - E\left(\bar{X}\right)}{\sqrt{\operatorname{Var}\left(\bar{X}\right)}} = \frac{\bar{X} - 200}{\sqrt{1.6}} = \frac{\bar{X} - 200}{1.26}$$

Therefore, by subtracting 200 from each term and then dividing each term by 1.26 in the above inequality, we get required probability as

$$P\left[\frac{\bar{X} - 200}{1.26} \geq \frac{202 - 200}{1.26}\right] = P\left[Z \geq 1.59\right]$$

$$= 0.5 - P\left[0 \leq Z \leq 1.59\right]$$

$$= 0.5 - 0.4441 = 0.0559$$

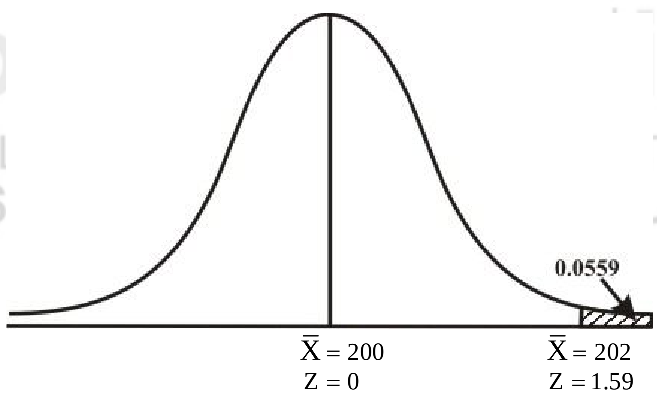

Fig. 2.6

E3) Here, we are given that

$$\mu = 2550,\ n = 100,\ S = 54$$

First of all, we find the sampling distribution of sample mean. Since sample size is large (n = 100 > 30) therefore, by the central limit theorem, the sampling distribution of sample mean follows normal distribution. Therefore, the mean of this distribution is given by

$$E\left(\bar{X}\right) = \mu = 2550$$

and variance

$$\operatorname{Var}\left(\bar{X}\right) = \frac{S^2}{n} = \frac{(54)^2}{100} = \frac{2916}{100} = 29.16$$

E4) Here, we are given that

$$\mu_1 = 68,\ \sigma_1 = 2.3,\ n_1 = 35$$

$$\mu_2 = 65,\ \sigma_2 = 2.5,\ n_2 = 50$$

To find the required probability, first of all we find the sampling distribution of difference of two sample means. Let $\bar{X}$ and $\bar{Y}$ denote the mean height of workers taken from hospital and company respectively. Since n_1 and n_2 are large (n_1, n_2 > 30) therefore, by the central limit theorem, the sampling distribution of $(\bar{X}-\bar{Y})$ follows normal distribution with mean

$$E\left(\bar{X}-\bar{Y}\right) = \mu_1 - \mu_2 = 68 - 65 = 3$$

and variance

$$\mathrm{Var}\left(\bar{X}-\bar{Y}\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{(2.3)^2}{35} + \frac{(2.5)^2}{50}$$

$$= 0.1511 + 0.1250 = 0.2761$$

The probability that the sample mean of height of hospital workers is 2 inch less than that of company workers is given by

$$P\left[\left(\bar{X}-\bar{Y}\right) < 2\right] \qquad [\text{See Fig. 2.7}]$$

To get the value of this probability, we convert variate $\left(\bar{X}-\bar{Y}\right)$ into a standard normal variate Z by the transformation

$$Z = \frac{\left(\bar{X}-\bar{Y}\right) - E\left(\bar{X}-\bar{Y}\right)}{\sqrt{\mathrm{Var}\left(\bar{X}-\bar{Y}\right)}} = \frac{\left(\bar{X}-\bar{Y}\right)-3}{\sqrt{0.2761}} = \frac{\left(\bar{X}-\bar{Y}\right)-3}{0.53}$$

Therefore, by subtracting 3 from each term and then dividing each term by 0.53 in the above inequality, we get required probability

$$P\left[\frac{\left(\bar{X}-\bar{Y}\right)-3}{0.53} < \frac{2-3}{0.53}\right] = P[Z < -1.89] = 0.5 - P[-1.89 \le Z \le 0]$$

$$= 0.5 - P[0 \le Z \le 1.89]$$

$$= 0.5 - 0.4706 = 0.0294$$

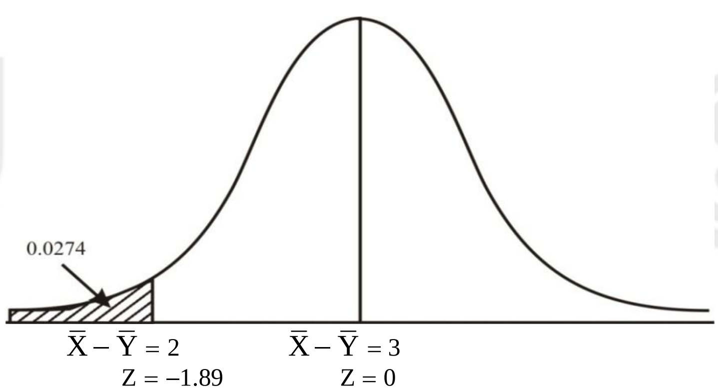

Fig. 2.7

E5) Here, we are given that

$$N = 10000,\ X = 7000 \Rightarrow P = \frac{X}{N} = \frac{7000}{10000} = 0.70\ \ \&\ n = 100$$

First of all, we find the sampling distribution of sample proportion. Here, the sample proportion is not given and n is large so we can assume that the conditions of normality hold. So the sampling distribution is approximately normally distributed with mean

$$E(p) = P = 0.70$$

and variance

$$Var(p) = \frac{PQ}{n} = \frac{0.70 \times 0.30}{100} = 0.0021 \quad [\because Q = 1 - P]$$

Since the sample size n is less than 10% of the population size N, therefore, correction factor $\sqrt{\frac{N-n}{N-1}}$ is ignored.

The probability that the sample proportion would have exceeded 60% acceptance is given by

$$P[p > 0.60] \qquad\qquad [\text{See Fig. 2.8}]$$

To get the value of this probability, we can convert the random variate p into standard normal variate Z by the transformation

$$Z = \frac{p - E(p)}{\sqrt{Var(p)}} = \frac{p - 0.70}{\sqrt{0.0021}} = \frac{p - 0.70}{0.046}$$

Therefore, by subtracting 0.70 from each term and then dividing each term by 0.046 in the above inequality, we get the required probability as

$$P\left[\frac{p - 0.70}{0.046} > \frac{0.60 - 0.70}{0.046}\right] = P[Z > -2.17]$$

$$= P[-2.17 < Z < 0] + 0.5$$

$$= P[0 < Z < 2.17] + 0.5$$

$$= 0.4850 + 0.5$$

$$= 0.9850$$

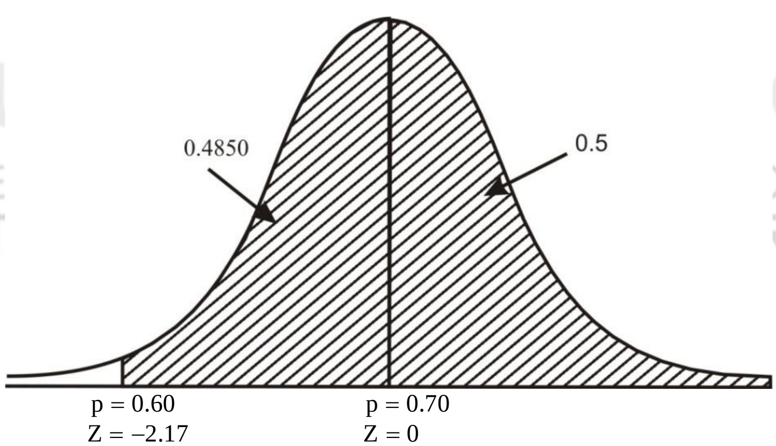

Fig. 2.8

E6) Here, we are given that

$$P_1 = \frac{12}{100} = 0.12,\; P_2 = \frac{10}{100} = 0.10,\; n_1 = 100,\; n_2 = 150$$

Let p_1 and p_2 be the sample proportions of alcohol drinkers in two cities A and B respectively. Here the sample proportions are not given and n_1 and n_2 are large $(n_1, n_2 > 30)$ so we can assume that conditions of normality hold. So the sampling distribution of difference of proportions is approximately normally distributed with mean

$$E(p_1 - p_2) = P_1 - P_2 = 0.12 - 0.10 = 0.02$$

and variance

$$\begin{aligned} Var(p_1 - p_2) &= \frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2} \\ &= \frac{0.12 \times 0.78}{100} + \frac{0.10 \times 0.90}{150} \qquad [\because Q = 1 - P] \\ &= 0.0009 + 0.0006 = 0.0015 \end{aligned}$$

The probability that the difference in sampling proportions lies between 0.01 and 0.08 is given by

$$P[0.01 < (p_1 - p_2) < 0.08] \qquad [\text{See Fig. 2.9}]$$

To get the value of this probability, we convert variate $(p_1 - p_2)$ into a standard normal variate Z by the transformation

$$Z = \frac{(p_1 - p_2) - E(p_1 - p_2)}{\sqrt{Var(p_1 - p_2)}} = \frac{(p_1 - p_2) - 0.02}{\sqrt{0.0015}} = \frac{(p_1 - p_2) - 0.02}{0.039}$$

Therefore, by subtracting 0.02 from each term and then dividing each term by 0.039 in the above inequality, we have required probability

$$\begin{aligned} P\left[\frac{0.01 - 0.02}{0.039} < \frac{(p_1 - p_2) - 0.02}{0.039} < \frac{0.08 - 0.02}{0.039}\right] &= P[-0.26 < Z < 1.54] \\ &= P[-0.26 < Z < 0] + P[0 < Z < 1.54] \\ &= P[0 < Z < 0.26] + P[0 < Z < 1.54] \\ &= 0.1026 + 0.4382 = 0.5408 \end{aligned}$$

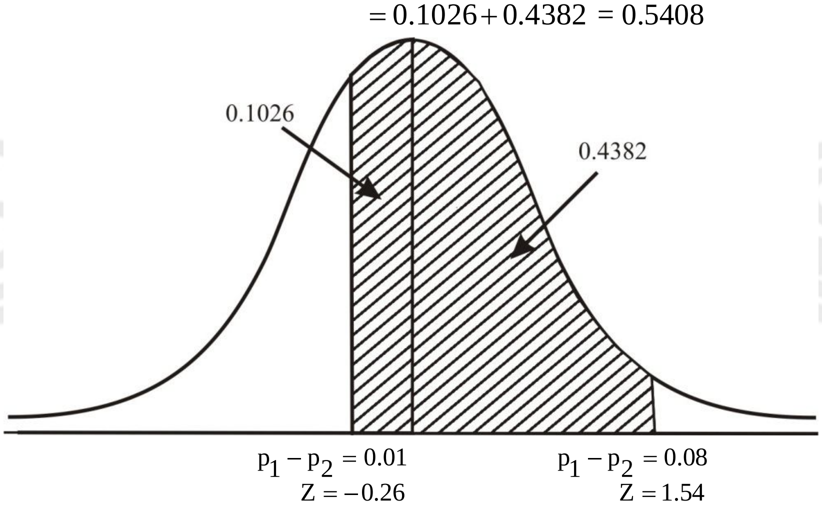

Fig. 2.9

E7) Here, we are given that

$$\sigma^2 = 11, n = 20$$

The probability that the variance of the given sample is greater than 16 pounds^2 is given by

$$P\left[S^2 \geq 16\right]$$

To get the value of this probability, we convert variate S^2 into χ^2-variate which follows the chi-square distribution with n - 1 = 19 degrees of freedom by the transformation

$$\chi^2_{(n-1)} = \frac{(n-1)S^2}{\sigma^2} = \frac{19S^2}{11}$$

Therefore, multiplying each term by 19 and then dividing each term by 11 in the above inequality, we get required probability as

$$P\left[\frac{19S^2}{11} \geq \frac{19 \times 16}{11}\right] = P\left[\chi^2_{(19)} \geq 27.64\right]$$

From the chi-square table, we see that $P\left[\chi^2_{(19)} \geq 27.20\right] = 0.10$ and $P\left[\chi^2_{(19)} \geq 30.14\right] = 0.05$.

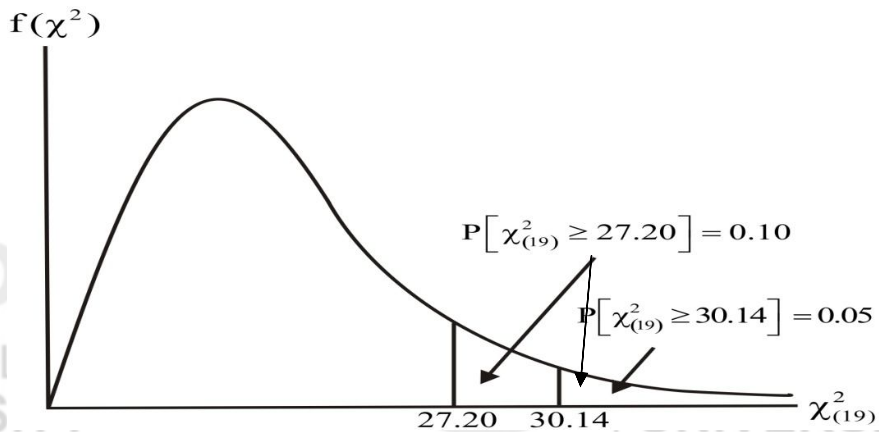

Fig. 2.10

Therefore, the above probability must lie between 0.05 and 0.10, that is,

$$0.05 < P\left[\chi^2 \geq 29.09\right] = P\left(S^2 \geq 16\right) < 0.10.$$