UNIT 4 NORMAL DISTRIBUTION: DEFINITION, CHARACTERISTICS AND PROPERTIES

Structure

4.1   INTRODUCTION

The word probability is a part of our daily lives. We use it quite frequently in our day to day life. We ask such questions. How likely it is that I will get an A grade in this exam ? It is likely to rain heavily this evening. How likely it is that a price of equity shares of company will increase in the next few days?

While common man answers these questions in a vague and subjective way, the researcher attempts to give the answer to these questions in a more objective and precise way. There are different types of probability distributions. Normal distribution is a kind of probability distribution,. If we look around us we will see that persons differ in terms of attributes like intelligence, interest, height, weight etc. Take for example intelligence, it can be seen that majority of us possess average intelligence i.e. IQ between 85-110 and the persons who have above average i.e. IQ 145 and above or below average IQ i.e. less than 70 etc., will be very few. Similarly if we see the height of the persons we will find that the height of the maximum number of persons range between 5.2 to 6 feet. The number of persons having height less than 5 feet and more than 6 feet is relatively very few. Similar type of trend we will find in the biological field, Anthropometrical data, social and economic data. If we plot these variations or data in the form of a distribution

or put it in the form of graph, we would get distribution known as normal curve or normal distribution. In this unit we will discuss, what is probability, different types of probabilities and the concept of normal distribution.

4.2 OBJECTIVES

After completion of this unit, you will be able to understand:

- Define probability and explain different types of probability;
- Describe meaning of normal distribution;
- Identify the characteristics of the normal distribution;
- Analyse the properties of the normal distribution; and
- Apply the normal distribution.

4.3 DEFINITIONS OF PROBABILITY

In the previous units you have learned the way to describe variables. We generate hypothesis, collect the data, categorise the data and summarise the data by computing measures of central tendency and variability.

Our interpretation and conclusions about variables are based on what we observed. But here our approach will be some what different. We will first suggest certain theories, propositions or hypothesis about variables, which will then be tested using the data we observe. The process of testing hypothesis through analysis of data is probability.

According to Beri (2007), **"Probability is the chance that a particular event will occur."** (What is the chance of getting a head when a coin is tossed.). To take another example, A company has launched new product what is the chance that it will be successful?

According to Levin and Fox (2006): "The term probability refers to the relative likelihood of occurrence of any given outcome or event."

Probability associated with an event is the number of times an event can occur relative to the total number of times any event can occur.

Probability of an outcome or event = The total number of times the occurrence of the event / the total possible times an event can occur.

For example, if in a room there are three women and seven men, the probability that the next person coming out of the room is a woman would be 3 in 10.

Probability of a women coming out next = number of women in the room / total number of men and women in the room

= 3/10 = .30

The probability of an event not occuring is known as converse rule of probability.

4.3.1 Types of Probability

There are two types of probability, one based on theoretical mathematics and the other based on systematic observation.

Theoretical probabilities reflect the operation of chance along with certain assumption we make about the events. For example the probability of getting

a head on a coin flip is .5 (1/2 = .5). The probability of guessing the correct answer of five item multiple choice question is .20 (1/5).

Empirical probabilities are those for which we depend on observation to determine their value. For example, the probability that Indian team wins a cricket match is about .6 (6 out of 10 matches) a 'fact' we know from observing hundreds of games with various countries over a year.

In both form probabilities (P) varies from 0 to 1.0. In most situations, the percentage and not the decimal is used to express the level of measurement. For example 0. 5 probability means 50% chance.

A zero probability means impossible and 1.00 probability means certainty.

4.3.2 Probability Distribution

A probability distribution is directly analogous to a frequency distribution. The only difference is probability distribution is based on the theory (probability theory), while frequency distribution is based on empirical data. In a probability distribution, first we specify the possible values of a variable and calculate the probability associated with each.

There are three types of probability distribution: the Binomial distribution, the poisson distribution and the normal distribution. Here we are interested in normal distribution.

4.4 THE NORMAL DISTRIBUTION

The concept of normal distribution is very important in statistical theory and practice.

In 1973 the French mathematician Abraham de Moivere discovered the formula of the normal curve. In the 19th Century Gauss and Laplace rediscovered the normal curve independently Gauss was primarily interested in the problem of astronomy which led to the consideration of a theory of error of observation. In the middle of the 19th Century Quetelet promoted the applicability of the normal curve. He believe that the normal cure could be extended to apply to problem of anthropology sociology and human affair.

In the latter part of the 19th century Sir Francis Galton began the first serious study of individual differences and during his systematic study he found that most of the physical and psychological traits of human being conformed reasonably well to the normal curve. In this way he extended the applicability of the normal curve. Normal curve is also known as *Gaussion Curve and bell shaped curve*.

A normal curve is one which graphically represents normal distribution. A normal distribution is one in which majority of the cases falls in the middle of the scale and small number of cases are located at both extremes of the scale. In psychology most of the traits are normally distributed, for example, if we administer an intelligence test on randomly selected large sample, we will find that the greatest proportion of IQ scores fall between 85 and 115. We would see a gradual falling off of scores on either side with few 'geniuses' who score higher than 145 and equally few who score lower than 55.

So far as physical human characteristic is concerned, most adults would fall within the 5 to 6 feet range of height, with far fewer being either very short (less than 5 feet) or very tall (more than 6 feet).

Normal probability distribution is a continuous probability distribution. It represents the frequency with which a variable occurs when the occurrence of that variable is governed by the laws of chance.

The normal curve is a theoretical or ideal model that was obtained from a mathematical equation rather than from actually conducting research and gathering data.

The normal curve takes into account the law which states that greater is the deviation of an event from the mean value in a series the less frequently it occurs.

In social sciences we conduct the study on representative sample and not on the entire population. Therefore, in actual practice the slightly deviated or distorted bell shaped curve is also accepted as the normal curve.

Self Assessment Questions

1) Given below are statements, indicate in each statement whether it is true or false.

   i)   A distribution where mean and median have different value is a normal distribution.   T/F

   ii) The right and left tail of the normal curve touch the horizontal axis.
                                                                          T/F

   iii) A probability distribution is based on actual observation.   T/F

   iv) Probability varies from 0 to 1.   T/F

   v)   The distribution are said to be skewed negatively when there are many individuals in a group with their scores higher than the average score of the group.   T/F

2) Fill in the blanks :

   i)   The mean median mode are _____ in normal curve.

   ii) A _____ indicates how for an individual raw scores falls from the mean of a distribution.

   iii) The _____ indicates how the scores in general scattee around the mean.

   iv) _____ cases lie between the mean and ±1s on the base line.

   v)   To find the deviation from the point of departure (i.e.) mean _____ of the distribution is used as a unit of measurement.


4.5 DEVIATION FROM THE NORMALITY

Although most of the variables in social sciences approximate the theoretical notion of normal distribution but some variables in social science do not conform to the theoretical notion of the normal distribution and they deviate from the normal distribution. This deviation from normality tends to vary in two ways.

4.5.1 Skeweness

Skeweness refers to lack of symmetry. A normal curve is perfectly symmetrical, there is a perfect balance between the right and left halves of the curve. For this curve, mean, median and mode are at the same point. A distribution is said to be 'skewed' when the mean and median fall at different

points in the distribution and the balance is shifted to one side or the other – that is to the left or to the right (Garrete 1981). Look at the figure given below.
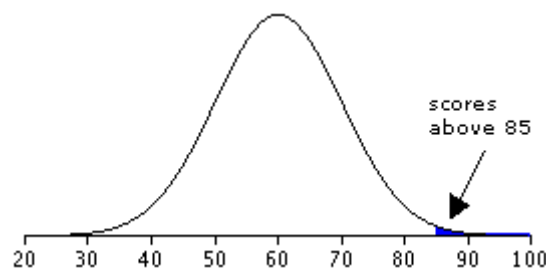
Properties of a normal curve

- The normal curve is one of a number of possible models of probability distributions.

- The normal curve is not a single curve, rather it is an infinite number of possible curves, all described by the same algebraic expression:

Similarity of Normal Curves of varied data include the following:

1) Shape

2) Symmetry

3) Tails approaching but never touching the X-axis, and

4) Area under the curve.

5) Bilaterally symmetrical

6) Most of the area under normal curve falls within a limited range of the number line.

7) All normal curves have a total area of 1.00 under the curve. This implies that the area in each half of the distribution is .50 or one half.

Drawing a normal curve

The standard procedure for drawing a normal curve is to draw a bell-shaped curve and an X-axis.

1) A tick is placed on the X-axis corresponding to the highest point (middle) of the curve.

2) Then, three ticks are placed to both the right and left of the middle point. These ticks are equally spaced and include all but a very small portion under the curve.

3) The middle tick is labeled with the value of $\square$

4) Sequential ticks to the right are labeled by adding the value of σ .

5) Ticks to the left are labeled by subtracting the value of σ from $\square$ for the three values.

6) For example, if M=52 and σ =12, then the middle value would be labeled with $(52+ 12)= 64$, then $+12 = 76$, and $+ 12 = 88$, and the three points to the left would have the values $(52 - 12) =40$, then 28, and then 16. An example is presented below:
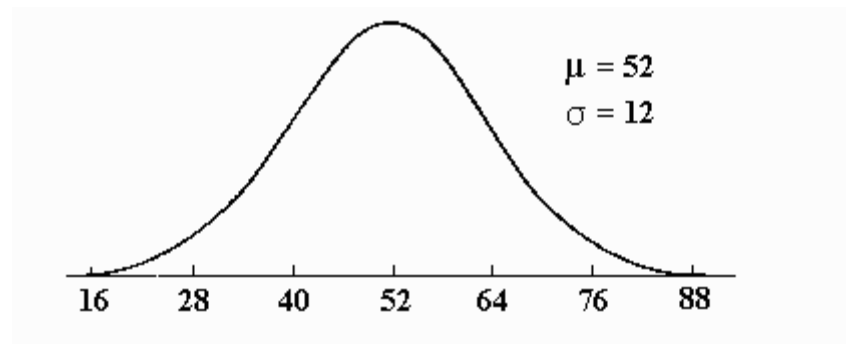
The two parameters, M and σ, each change the shape of the distribution in a different manner.

The first, M determines where the midpoint of the distribution falls.

Changes in M, without changes in σ, result in moving the distribution to the right or left.

That is, it depends on whether the new value of M was larger or smaller than the previous value.

At the same time, it does not change the shape of the distribution.

An example of how changes in M (μ) affect the normal curve are presented below:
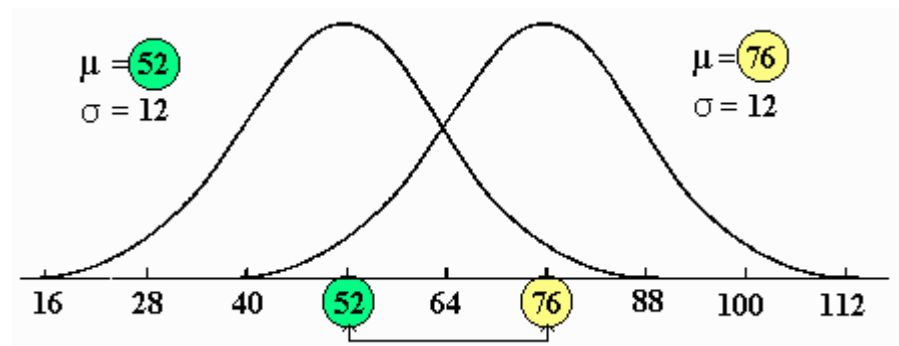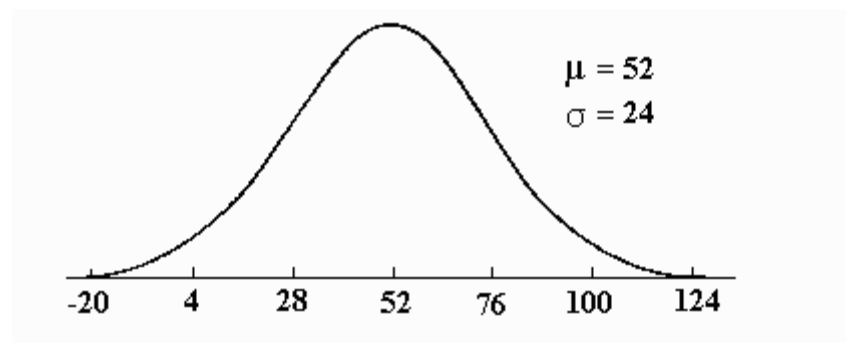

Changes in the value of σ , on the other hand, change the shape of the distribution without affecting the midpoint, because σ affects the spread or the dispersion of scores.

The larger the value of σ , the more dispersed the scores;

The smaller the value, the less dispersed.

The distribution below demonstrates the effect of increasing the value of σ

Suppose the second distribution was drawn on a rubber sheet instead of a sheet of paper and stretched to twice its original length in order to make the

two scales similar. Drawing the two distributions on the same scale results in the following graph:
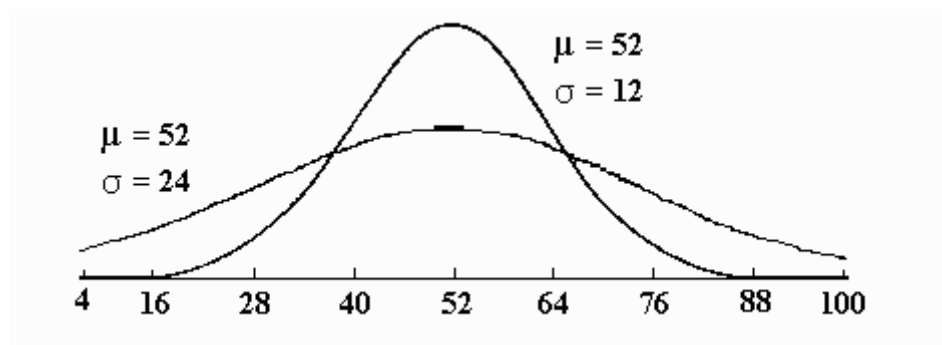
Note that the shape of the second distribution has changed dramatically, being much flatter than the original distribution.

It must not be as high as the original distribution because the total area under the curve must be constant, that is, 1.00.

The second curve is still a normal curve; it is simply drawn on a different scale on the X-axis.

A different effect on the distribution may be observed if the size of σ is decreased. Below the new distribution is drawn according to the standard procedure for drawing normal curves:
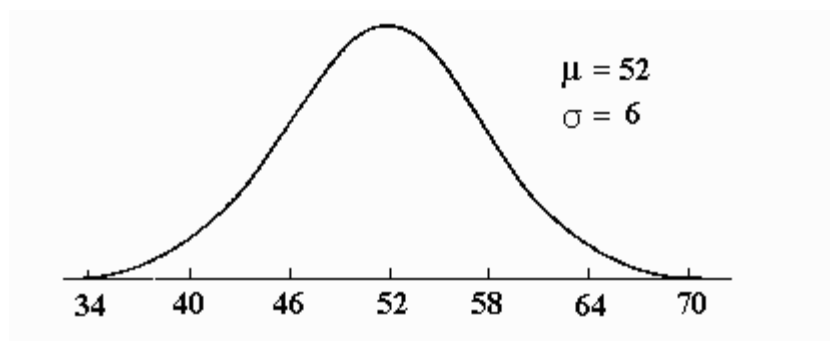
Now both distributions are drawn on the same scale, as outlined immediately above, except in this case the sheet is stretched before the distribution is drawn and then released in order that the two distributions are drawn on similar scales:
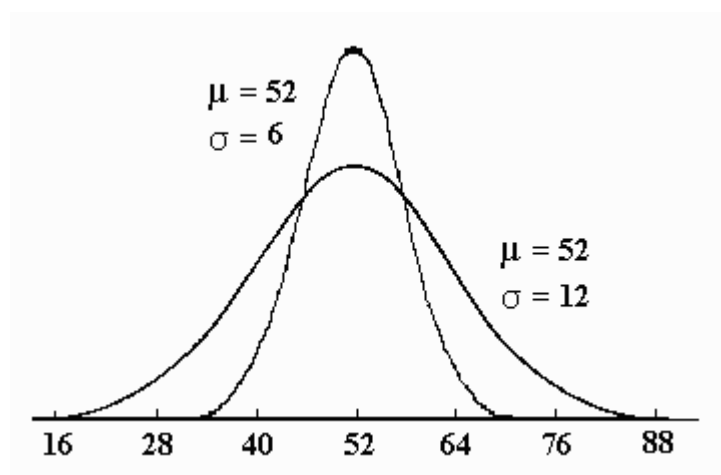
Note that the distribution is much higher in order to maintain the constant area of 1.00, and the scores are much more closely clustered around the value of σ, or the midpoint, than before.

Skewness in a given distribution may be computed by the following formula.

3 (Mean - Median)

Skewness =

Standard deviation

In case when the percentiles are known, the value of skewness may be computed from the following formula :

$P_{90} + P_{10}$

$S_k =$ $- P_{50}$

4.5.2 Kurtosis

The term Kurtosis refers to the peakedness or flatness of a frequency distribution as compared with the normal (Garrete 1981).

Kurtosis is usually of three types :

Platykurtic. A frequency distribution is said to be playkurtic, when it is flatter than the normal.

Leptokurtic. A frequency distribution is said to be leptokurtic, when it is more peaked than the normal.

Mesokurtic. A frequency distribution is said to be mesokurtic, when it almost resembles the normal curve (neither too flattened nor too peaked).

Self Assessment Questions

1) Discuss the various deviations from normality.

    …………………………………………………………………………………..

    …………………………………………………………………………………..

    …………………………………………………………………………………..

    …………………………………………………………………………………..

2) What is kurtosis. Enumerate the different types of kurtosis

    …………………………………………………………………………………..

    …………………………………………………………………………………..

    …………………………………………………………………………………..

    …………………………………………………………………………………..

3) How does change in the mean affect the normal curve?

    …………………………………………………………………………………..

    …………………………………………………………………………………..

    …………………………………………………………………………………..

Significance of the Difference of Frequency

56

4.6   CHARACTERISTICS OF A NORMAL CURVE

The following are the characteristics of the normal curve.

Normal curves are of *symmetrical distribution*. It means that the left half of the normal curve is a mirror image of the right half. If we were to fold the curve at its highest point at the center, we would create two equal halves.

The first and third quartiles of a normal distribution are *equidistance from the median*.

For the curve the mean median and mode all have the same value.

In skewed distribution mean median and mode fall at different points.

The normal curve is *unimodal*, having only one peak or point of maximum frequency that point in the middle of the curve.

The curve is a *asymptotic*. It means starting at the centre of the curve and working outward, the height of the curve descends gradually at first then faster and finally slower. An important situation exists at the extreme of the curve. Although the curve descends promptly toward the horizontal axis it never actually touches it. It is therefore said to be asymptotic curve.

In the normal curve the *highest ordinate is at the centre*. All ordinate on both sides of the distribution are smaller than the highest ordinate.

A large number of scores fall relatively close to the mean on either side. As the distance from the mean increases, the scores become fewer.

The normal curve involves a *continuous distribution*.

4.7   PROPERTIES OF THE NORMAL DISTRIBUTION

In the following paragraphs we will discuss the properties of the normal distribution.

4.7.1 The Equation of the Normal Curve

$$y = N x^2 / \text{ s } 2p^e \ 2s^2$$

Here :

x = Scores (expressed as deviation from the mean) laid off along the base line or x axis.

y = the height of the curve above the x.

N = Number of cases.

s = standard deviation of the distribution.

p = 3.1416 (the ratio of the circumstances of a circle to its diameter).

e = 2.7183 (base of the Napierian system of logarithms)

When N and s are known, then with the help of above formula we can compute (1) the frequency (or Y) of a given value x; and (2) the number between the points. But these calculations are rarely necessary as tables are available from which this information may be readily obtained.

4.7.2   Area Under the Normal Curve

It is important to keep in mind that the normal curve is an ideal or theoretical distribution (that is, a probability distribution). Therefore, we denote its mean

57

by m and its standard deviation by s. The mean of the normal distribution is at its exact center. The standard deviation (s) is the distance between the mean (m) and the point on the base line just below where the reversed S-shaped portion of the curve shifts direction.

To employ the normal distribution in solving problems, we must acquaint ourselves with the area under the normal curve : the area that lies between the curve and the base line containing 100% or all of the cases in any given normal distribution.

When normally distributed, it is seen that 34.13% cases lie between the mean and 1 s above the mean. In the same way we can say that 47.72% of the cases under the normal curve lie between mean and 2 s above the mean and 49.87% lie between the mean and 3 s above the mean.

The symmetrical nature of the normal curve leads us to make another important point. Any given sigma distance above the mean contains the identical proportion of cases as the same sigma distance below the mean.

Thus, if 34.13% of the total area lies between the mean and 1s above the mean, then 34.13% of the total area also lies between the mean and 1s below the mean; if 47.72% lies between the mean and 2s above the mean, then 47.72% lies between the mean and 2s below the mean; if 49.87% lies between the mean and 3s above the mean, then 49.87% also lies between the mean and 3s below the mean. In other words, 68.26% of the total area of the normal curve (34.13% + 34.13%) falls between - 1s and + 1s from the mean; 95.44% of the area (47.72% + 47.72%) falls between - 2s and + 2s from the mean; and 99.74%, or almost all, of the cases (49.87% + 49.87%) falls between - 3s and + 3s from the mean. It can be said, then, that six standard deviations include practically all the cases (more than 99%) under any normal distribution.

For example an intelligence test was administered on large randomly selected sample of girls. The obtained mean (m) was 100 and standard deviation was 10. Then we can say that 68.26% of the population would have IQ scores that falls between 90 (100-10) and 110 (100 + 10).

Moving away from the mean we would find that 99.74% of these cases would fall between score 70 and 30 (between - 3s to + 3s).

4.7.3   Table of Areas Under the Normal Curve

Suppose we want to determine the percent of total frequency that falls between the mean, and say a raw score located 1.40 s above the mean. A raw score 1.40s above the mean is obviously greater than 1s, but less than 2s from the mean. It means that this distance from the mean would include more than 34.13% but less than 47.72% of the total area under the normal curve. To determine the exact percentage within this interval we have to employ Table A given in any statistical book under the heading "area under the Normal curve".

In Table A the total area under the curve is taken arbitrarily to the 10,000 because of the greater case with which fractional parts of the total area may then be calculated.

The first column of the table, x/s gives the distance that lie tenth of s measured off on the base line of the normal curve from the mean as origin. We have seen the deviation from the mean as $x = x - M$. If x is divided by s, deviation from the mean is expressed in s units. Such s deviation scores are often called Z scores (Z = x/s).

To find the number of cases in the normal distribution between the mean and 1s from the mean, go down the x/s column until 1.0 is reached and in the next column under .00 take the entry opposite 1.0 viz 3413. This figure mean 34.13% of the total frequencies falls between the mean and 1s. To find out the percentage of the distribution between the mean and 1.57s, go down the x/s column to 1.5 then across horizontally to the column headed .07 and take the entry 4418. This means that in a normal distribution 44.18% of the N lies between mean and 1.57 s.

Since the curve is bilaterally symmetrical, the entries in Table A apply to s distance measured in the negative or positive direction, which ever we need. For example to find out the percentage of the distribution between the mean and -1.26s take the entry in the column headed .06, opposite 1.2 in the x/s column. The entry is 3962 it means that 39.62% of the cases fall between the mean and -1.26s.

Self Assessment Questions

1) What are the characteristics of a normal curve?

   ………………………………………………………………………..

   ………………………………………………………………………..

   ………………………………………………………………………..

   ………………………………………………………………………..

   ………………………………………………………………………..

2) What are the properties of a normal curve?

   ………………………………………………………………………..

   ………………………………………………………………………..

   ………………………………………………………………………..

   ………………………………………………………………………..

   ………………………………………………………………………..

4.8   APPLICATION OF THE NORMAL CURVE

In psychological researches the normal curve has the main practical application given below :

- A normal curve helps in transforming the raw scores into standard scores.

- With the help of normal curve we can calculate the percentile rank of the given scores.

- A normal curve is used to find the limits in any normal distribution which include a given percentage of the cases.

- We can compare two distributions in terms of overlapping with the help of normal curve.

- A normal curve is used to determining the relative difficulty of test questions, problems and other test items.

- When the trait is normally distributed normal curve is used to separate a given group into subgroups according to capacity.

4.9   LET US SUM UP

In this chapter we introduced the concept of probability, indicated by the number of times an event can occur relative to the total number of times. A frequency distribution is based on actual observation whereas probability distribution is theoretical idea. There are three types of probability distribution i.e. Binomial distribution, the poisson distribution and normal distribution. It is a symmetrical bell shaped curve. It is not skewed. Normal curve can be used to determine the percent of the total area under the normal curve associated with any given sigma distance from the mean. Any given sigma distance above the mean contains the identical properties of cases as the same sigma distances below the mean.

4.10   UNIT END QUESTIONS

1)   What is a normal curve ? Why is it named as Gaussion curve.

2)   What do you understand by the term divergence from normality ? Point out the main types of such divergent curve.

3)   Discuss the main characteristics of a normal curve.

4)   What are the application of normal distribution

Answer:

1)   (i) F (ii) F (iii) F (iv) T   (v) T

2)   (i) Identical (ii) Z score (iii) Standard deviation

(iv) 68.26% (v) Standard deviation

4.11   GLOSSARY

Continuous random variable	: A random variable than can assume any value within a given range.
Normal distribution	: A symmerical bell shaped curve. The two tail of the curve never touch the horizontal axis.
Random variable	: A variable that assume a unique numerical value for each of the outcomes is a sample space of a probability experiment.
Standard score	: Known as Z scores also can be obtained by taking the deviation from mean and divided by standard deviation.

4.12   SUGGESTED READINGS

Beri G.C. (2007), *Business Stastics,* (2nd ed.) New Delhi, Tata MCgraw Hill.

Levin, J. & Fox, J.A. (2006) *Elementary Statistics in Social Research* (10th ed.) India, Pearson Education.