BLOCK 4

NORMAL PROBABILITY DISTRIBUTION

BLOCK INTRODUCTION

The last block of this course is Normal Probability Distribution that is covered in **unit 8**, having the same title. The unit will deal with the concept of probability as well as the concepts related to probability. The concept, nature and properties of normal probability distribution will also be discussed with a focus on importance and properties of normal distribution. Yet another significant topic that will be explained is standard scores or z-scores. In this context, the concept, properties and uses of z-score will be covered. The computation of z-score will also be discussed. The last sub topic that will be discussed in this unit is divergence from normality:, viz., kurtosis and skewness.

UNIT 8 NORMAL PROBABILITY DISTRIBUTION*

Structure

8.0 OBJECTIVES

After reading this unit, you will be able to:

- describe the concept and nature of probability;
- discuss the concept, nature, properties and relevance of normal distribution curve;
- elucidate the concept, properties, uses and computation of z- scores; and
- explain divergence from normality.

8.1 INTRODUCTION

Let us understand the main concept of this unit, that is, normal probability curve with an example. Suppose we take variables like height or weight or psychological variables like intelligence or emotional intelligence and we

* Dr. Smita Gupta, Faculty, Discipline of Psychology, School of Social Sciences, IGNOU, New Delhi

measure the same across a large population, we are likely to obtain a graph which will indicate that the maximum scores lie in the middle of the distribution and lower scores at the extreme. With regard to height, for example, there will be more persons with average height and few persons who are taller or shorter compared to the average. And when this data is plotted on a graph, we will get a normal cure as can be seen in figure 8.1.

In the previous units, we discussed about the concept of statistics, data organisation and graphical representation, measures of central tendency and variability and we also discussed about correlation and its computation.

In the present unit, you will be introduced to the concept of probability, normal probability curve and other related aspects.

8.2 CONCEPT OF PROBABILITY

The term 'Probability' refers to chance or likelihood. For example, if you say that, "it will probably be hot tomorrow" or "probably the teacher might not come tomorrow", the sentences reveal that there are chances for the events to occur tomorrow but there is no certainty. This means that the events mentioned in the above examples are not certain to happen. Now, a question that might bother you is 'what is research and statistics to do with probability?' The probability of an event to occur due to certain reason helps the researchers to formulate hypotheses to continue with research and experiments. It serves as a baseline for the researchers to start as well as conclude their research. It is relevant in science and investigation and is also important in finance, gambling, artificial intelligence, mathematics and game theory. Researchers can draw conclusions on basis of probabilities.

In simple terms, probability refers to the chance, possibility or likelihood of an event to occur. In statistics, the term 'Probability' refers to the expected frequency (chance) of occurrence of an event among all possible similar events. The expected frequency of the occurrence of the event is based on knowledge or information about the conditions determining the occurrence of the event or phenomena. For example, if you toss a coin in air,there is an equal chance for head or tail to appear. Thus, the probability that head or tail will appear, when a coin is tossed in air, is ½ to appear. Similarly, a die (singular of dice) has six sides with dots ranging from one to six. The probability of a die to show any of its side is 1/6.

The probability is denoted in form of ratio in statistical terms, wherein a probability ratio is denoted as:

Probability Ratio = Desired Outcome/s or event(s) / Total Number of outcomes or events

To make it more clear, the probability ratio of any side of the coin to occur after each toss will be

= 1 (Either head or tail)/ 2 (Total Number of sides of coin)

And the probability ratio of any side of the dice to fall is

=1 (Any one side of the six sides)/ 6 (Total Number of sides of dice)

A probability ratio always ranges between the limit of 0.00 (impossibility of occurrence) to 1.00 (certainty of occurrence). We can say that the possibility of the Sun to set in east is 0.00 and the possibility for the Sun to set in west is 1.00. The other possible degrees of likelihood range between these limits (0.00-1.00) and are expressed in form of appropriate ratios. It is also worth mentioning here that if the probability of an event is higher, then, there are more chances that the event will occur. For example, predicting whether it will be hot the next day will have more chance to occur if on the present day the temperature might have been high enough. Similarly, if the teacher has fallen sick then it is more likely that s/he will not be coming to school on the next day.

8.2.1 Concepts Related to Probability

Probability is a statistical concept which can be measured and analysed. Due to its much scientific applications, there are certain related concepts which you need to know. These concepts are:

1) **Sample Space, Events and Power Set:** The collection or set of all possible results or outcome of an experiment or event is known as sample space. For example, in a die the all possible results to come when it is rolled, ranges between 1 to 6 (1, 2, 3, 4, 5, 6 are the number of dots assigned at each respective face of the dice). So the sample space for the outcome of die ranges from 1 to 6. The subset of the sample space which is a specified collection of possibilities from the overall possibilities is called power set. If we consider that a collection of possible results can be all even numbers (2, 4, 6) of the die, the subset (2, 4, 6) is an element of the power set of the sample space of dice rolls. These collections are called "events". In this case, {2, 4, 6} is the event that the die falls on some even number. If the results that actually occur fall in a given event, the event is said to have occurred. After considering all different collections of possible results of a sample space, a power set is formed.

2) **Mutually Exclusive/Disjoint Events:** Mutually exclusive event refers to incompatibility. Two events are said to be mutually exclusive if both cannot occur simultaneously in a single trial. In such circumstances, the occurrence of one event prevents the occurrence of other event. For example, when a single coin is tossed there is a chance of either head (H) to appear or tail (T) to appear in a single trial but both can not be up at a same time, then both H and T are called mutually exclusive events. Hence, if H an T are mutually exclusive events then, probability (HT) = 0.

3) **Exhaustive/Collective Events:** An event is said to be exhaustive when its totality includes all possible outcomes of a random experiment. Meaning thereby, that out of a set of events, atleast one event occurs in each trial. For example, if you throw a die, the outcomes will range in between 1,2,3,4,5 and 6. You can not get 7 because it is not there in the die at all. Therefore the outcomes 1, 2, 3, 4, 5, and 6 are collectively exhaustive, because they constitute the possible entire range of probable outcomes.

4) **Independent and Dependent Events:** Two or more events are said to be independent events if the outcome of one event has no effect or is not affected by the outcome of the other event. For example, the results of

tossing a coin will not be affected by the outcome of throwing a die. While dependent events are those events in which either occurrence or non occurrence of one event in any trial affects the probability of other event in other trials. For example, we have five apples and five oranges in a basket. We take out one of the fruits, which could be apple or orange. If it was apple then there are four apples and five oranges in the bag. Thus, the probability of the next frute that we take from the basket being apple is 4/9. But if the first time the fruit was orange then the probability would be 5/9.

5) **Equally Likely Events:** Events are said to be equally likely events if each event has a fair enough chance to occur approximately the same number of times. Thus, none of the event is more likely to occur more often than others. For example, if an unbiased coin is tossed, each face of the coin may be expected to be observed for the same number of times.

6) **Complementary Events:** If two events are mutually exclusive and exhaustive, they both are said to complement each other. For example, if there are two events- X and Y, then X is called the complementary event of Y and vice versa. Let us understand it by throwing a die, the occurrence of even numbers (2, 4, 6) and odd numbers (1, 3, 5) are complementary events. They are the two possible outcomes of an event, where they are the only two possible outcomes.

Check Your Progress I

1) State whether the statements are True or False

Sr. No.	Statements	True/ False
A	Events are said to be equally likely events if each event has a fair enough chance to occur approximately the same number of times.	
B	A probability ratio always ranges between the limit of -1.00 to +1.00.	
C	Two events are said to be mutually exclusive if both occur simultaneously in a single trial.	
D	Two or more events are said to be independent events if the outcome of one event does not effect as well as is not affected by the outcome of the other event.	
E	If two events are mutually exclusive and exhaustive, they both are said to complement each other.	

8.3 CONCEPT, NATURE AND PROPERTIES OF NORMAL PROBABILITY DISTRIBUTION

A continuous probability distribution for a variable is called as normal probability distribution or simply normal distribution. It is also known as Gaussian/ Gauss or LAPlace – Gauss distribution. The normal distribution is determined by two parameters, mean and variance. The normal distributions are used to represent the real valued random variables whose distributions are unknown. They are used very frequently in the areas of natural sciences and social sciences. When the normal distribution is represented in form of a graph, it is known as normal probability distribution curve or simply normal curve. A normal curve is a bell shaped curve, bilaterally symmetrical and is continuous frequency distribution curve. Such a curve is formed as a result of plotting frequencies of scores of a continuous variable in a large sample. The curve is known as normal probability distribution curve because its y ordinates provides relative frequencies or the probabilities instead of the observed frequencies. A continuous random variable can be said to be normally distributed if the histogram of its relative frequency has shape of a normal curve (as represented in the below figure 8.1).

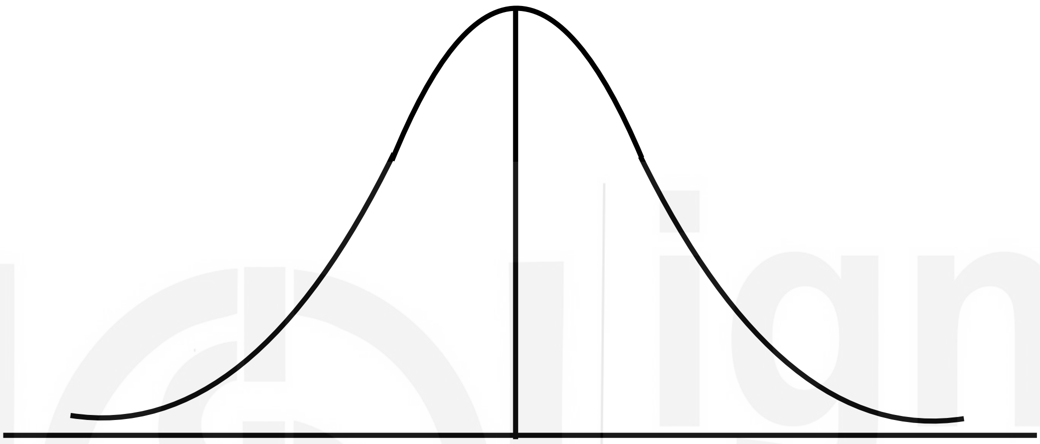


Fig. 8.1: Normal Curve

It is most important to understand the characteristics of frequency distribution of normal curve in the fields of mental measurement and experimental psychology.

8.3.1 Importance of Normal Distribution

As discussed earlier, the normal distribution plays a very significant role in the fields of natural science and other social sciences. Some of the relevance of the normal distribution are described below:

- The normal distribution is a continuous distribution and plays significant role in statistical theory and inference.
- The normal distribution has various mathematical properties which makes it convenient to express the frequency distribution in simplest form.
- It is a useful method of sampling distribution.
- Many of the variables in behavioral sciences like, weight, height, achievement, intelligence have distributions approximately like the normal curve.
- Normal distribution is a necessary component for many of the inferential statistics like z-test, t-test and F-test.

8.3.2 Properties of Normal Distribution Curve (NPC)

As discussed earlier, the representation of normal distribution of random variable in graphic form is known as normal probability distribution curve. The following are the properties of the normal curve:

- It is a bell shaped curve which is bilaterally symmetrical and has continuous frequency distribution curve.
- It is a continuous probability distribution for a random variable.
- It has two halves (right and left) and the value of mean, median and mode are equal (mean = median = mode), that is, they coincide at same point at the middle of the curve.
- The normal curve is asymptotic, that is, it approaches but never touches the x-axis, as it moves farther from mean.
- The mean lies in the middle of the curve and divides the curve in to two equal halves. The total area of the normal curve is within z $\pm$ 3 σ below and above the mean.
- The area of unit under the normal curve is said to be equal to one (N=1), standard deviation is one (σ =1), variance is one (σ^2=1) and mean is zero (μ=0).
- At the points where the curve changes from curving upward to curving downward are called inflection points.
- The z-scores or the standard scores in NPC towards the right from the mean are positive and towards the left from the mean are negative.
- About 68% of the curve area falls within the limit of plus or minus one standard deviation ($\pm$1 σ) unit from the mean; about 95% of the curve area falls within the limit of plus or minus two standard deviations ($\pm$2 σ) unit from the mean and about 99.7% of the curve area falls within the limit of plus or minus three standard deviations ($\pm$3 σ) unit from the mean (refer to figure 8.2).
- The normal distribution is free from skewness, that is, it's coefficient of skewness amounts to zero.
- The fractional areas in between any two given z-scores is identical in both halves of the normal curve, for example, the fractional area between the z-scores of +1 is identical to the z-scores of –1. Further, the height of the ordinates at a particular z-score in both the halves of the normal curve is same, for example, the height of an ordinate at +1z is equal to the height of an ordinate at –1z.

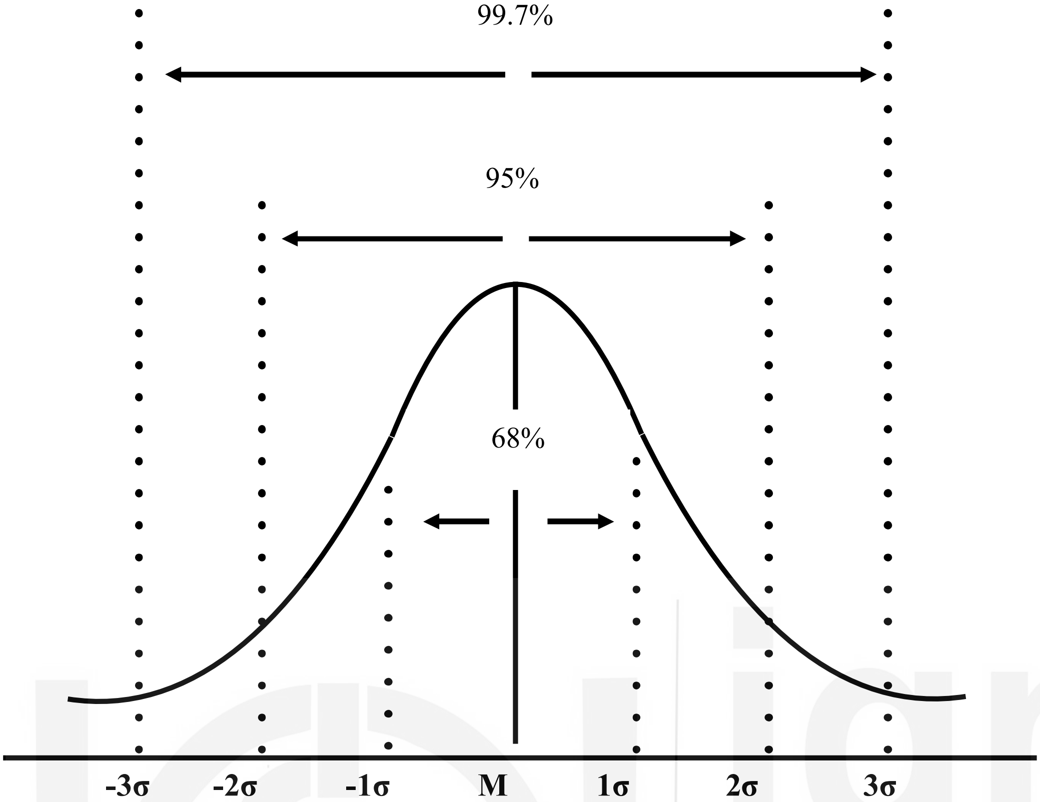


Fig. 8.2: Normal Probability Distribution Curve

Check Your Progress II

1) Fill in the Blanks

   a) The .... is considered as an ideal degree of peak or kurtosis.

   b) In a normal probability curve, the value of mean, median and mode are ....

   c) The normal distribution is a .... distribution and plays significant role in statistical theory and inference.

   d) The area of unit under the normal curve is said to be equal to ....

   e) The normal distribution is determined by two parameters................ and ....

8.4 STANDARD SCORES (Z-SCORES)

Standard score or z-score is a transformed score which shows the number of standard deviation units by which the value of observation (the raw score) is above or below the mean. The standard score helps in determining the probability of a score in the normal distribution. It also helps in comparing scores from different normal distributions.

8.4.1 Concept of Standard Score (z-score)

The standard score is a score that informs about the value and also where the value lies in the distribution. Typically, for example, if the value is 5 standard deviations above the mean then it refers to five times the average distance above the mean. It is a transformed score of a raw score. A raw score or sample value is the unchanged score or the direct result of measurement. A raw score (X) or sample value cannot give any information of its position within a distribution. Therefore, these raw scores are transformed in to z-scores to know the location of the original scores in the distribution. The z-scores are also used to standardise an entire distribution.

These scores (z) help compare the results of a test with the "normal" population. Results from tests or surveys have thousands of possible results and units. These results might not be meaningful without getting transformed. For example, if a result shows that height of a particular person is 6.5 feet; such findings can only be meaningful if it is compared to the average height. In such a case, the z-score can provide anidea about where the height of that person is in comparison to the average height of the population.

8.4.2 Properties of z-score

Following are some of the properties of the Standard (z) Score:

- The mean of the z-scores is always 0.
- It is also important to note that the standard deviation of the z-scores is always 1.
- Further, the graph of the z-score distribution always has the same shape as the original distribution of sample values.
- The z-scores above the value of 0 represent sample values above the mean, while z-scores below the value of 0 represent sample values below the mean.
- The shape of the distribution of the z-score will be similar or identical to the original distribution of the raw scores. Thus, if the original distribution is normal, then the distribution of the z-score will also be normal. Therefore, converting any data to z-score does not normalize the distribution of that data.

8.4.3 Uses of z-score

z-scores are useful in the following ways:

- **It helps in identifying the position of observation(s) in a population distribution:** As mentioned earlier, the z-scores helps in determining the position/distance of a value or an observation from the mean in the units of standard deviations. Further, if the distribution of the scores is like the normal distribution, then we are able to estimate the proportion of the population falling above or below a particular value. z-score has important implication in the studies related to diet and nutrition of children. It helps in estimating the values of height, weight and age of children with reference to nutrition.

- **It is used for standardising the raw data:** It helps in standardising or converting the data to enable standard measurements. For example, if you wish to compare your scores on one test with the scores achieved in another test, comparison on the basis of raw score is not possible. In such a situation, comparisons across tests can only be done when you standardise both sets of test scores.

- **It helps in comparing scores that are from different normal distributions:** As mentioned in the previous example, z-scores help in comparing scores from different normal distribution. Thus, z-scores can help in comparing the IQ scores received from two different tests.

8.4.4 Computation of z-score

As mentioned earlier, z-score refers to the distance of the sample value from the mean in the standard deviations. z-score can be computed for each value of the sample. The following formula is used to compute z-score of a sample value-

$$z = X - M / SD \text{ or}$$

$$z = X - M/\sigma$$

where,

X = a particular raw score

M = Sample mean

SD or σ = Standard Deviation

To illustrate, suppose the following are the marks obtained by students in mathematics. The marks obtained are expressed here in terms of raw scores. The mean, SD and z-scores can be then calculated accordingly:

Students	Raw Scores (X)	X- M	z
A	50	-15	-1.24
B	60	-5	-0.41
C	66	1	0.08
D	70	5	0.41
E	80	15	1.24
N= 5			
Sum	326		
Mean	65		
SD	12.04		

The above illustration shows the z-scores of the marks obtained by each student (A,B,C,D and E). In the above example, student A is -1.24 standard deviations, or 1.24 standard deviation units below the mean.Similarly the student E is 1.24 units above the mean. The standard deviation is used as unit of measurement in standard scores. The standard score helps innormalising or collapsing the data to a common standard based on how many standard deviations values lie from the mean.

The variation of z-scores range from -3 standard deviations (which would fall to the far left of the normal distribution curve) up to +3 standard deviations (which would fall to the far right of the normal distribution curve). Further, we need to know the values of the μ (mean) and also the σ (standard deviation) of the population.

Thus, if we want to compute z-score for X = 70, M= 65 and SD= 12.04, we willuse the formula

$$\mathbf{z = X - M/ SD}$$

$$= 70 - 65/ 12.04$$

$$= 5/ 12.04$$

$$= 0.42$$

Thus, the z-score is obtained as 0.42.

Check Your Progress III

1) Fill in the Blanks

a) The mean of the z-scores is always ....

b) .... is used for standardising the raw data.

c) The variation of z-scores ranges from .... to .... standard deviations.

d) The standard deviation of the z-scores is always ....

e) The standard score is a score that informs about the ....

8.5 DIVERGENCE FROM NORMALITY: SKEWNESS AND KURTOSIS

Many times, the frequency curve may be more peaked or flatter than the normal probability distribution curve. In such cases, the distribution is said to have diverged from normality. Basically these divergences are of two types: Kurtosis and Skewness. Kurtosis is a measure of "tailedness" of the probability distribution of a random variable. In other words, it is a measure whether a data is heavy tailed or light tailed in relation to normal distribution. On the other hand, Skewness is a measure of asymmetry of the probability distribution of a random variable about its mean. Let us discuss both the divergences one by one.

8.5.1 Kurtosis

Kurtosis deals with the tails of distribution curve and not its peak. It does not refer to the height of the curve. Kurtosis can be quantified in various ways for a particular distribution. The kurtosis of any variable in normal distribution is 3. This value is used for comparison with respect to the other types of Kurtosis. The kurtosis is classified as follows (refer to figure 8.3):

1) **Leptokurtosis/ Leptokurtic**

In a Leptokurtic distribution, the frequency curve is narrower than the NPC and the area of the curve shifts towards the center and has longer tails at both the ends. Usually they are referred to as positive kurtosis in which the value of distribution has heavier tails than the normal distribution (refer to figure 8.3).

2) **Mesokurtosis/ Mesokurtic**

A mesokurtic curve is not too flat or not too peaked and in a way resembles normal curve (Mangal, 2002).

3) **Platykurtosis/Platykurtic**

Platykurtic curve refers to the distribution having fewer and less extreme outliers than does the normal distribution. Usually they are referred to as negative kurtosis in which the value of distribution has lesser or fewer tails or outliers than the normal distribution. The curve in platykurtic is flatter when compared with normal curve (refer to figure 8.3).

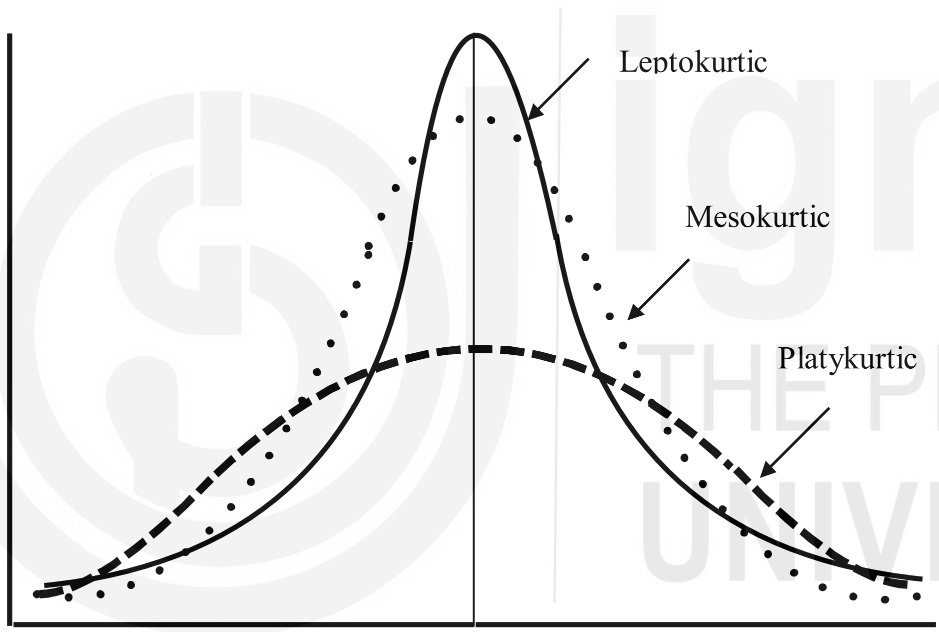


Fig. 8.3: Three Types of Kurtosis

Kurtosis can be computed with the help of the following formula:

$$K_u = Q/ P_{90} – P_{10}$$

Where,

K_u = Kurtosis

Q = Quartile Deviation

P_{90} = 90th Percentile

$P1_0$ = 10th Percentile

When kurtosis value is computed with the help of the above formula, the value for normal curve is obtained as 0.263. If the value obtained is below 0.263 then the curve can be termed as leptokurtic and if the value is above 0.263, then the curve can be termed as platykurtic (Mangal, 2002).

8.5.2 Skewness

As informed to you earlier that in NPC, the mean, median and mode coincide together (fall at same point) and have equal values. In a Skewed distribution, the mean, median and mode fall at different points in the distribution, and the center of gravity is shifted to one side. So, the Skewness determines the lack of symmetry in the curve and since normal probability is a symmetrical curve, it has zero skewness. Skewness can be measured in terms of Pearson's measure and percentiles. Depending upon the distribution of scores, Skewness can be classified in to two types as mentioned below.

1) **Positive Skewness**

A distribution curve is said to be positively skewed when the distribution of scores are more at the left end (refer to figure 8.4) . In a positively skewed distribution, more individuals obtian scores that are less than the mean.

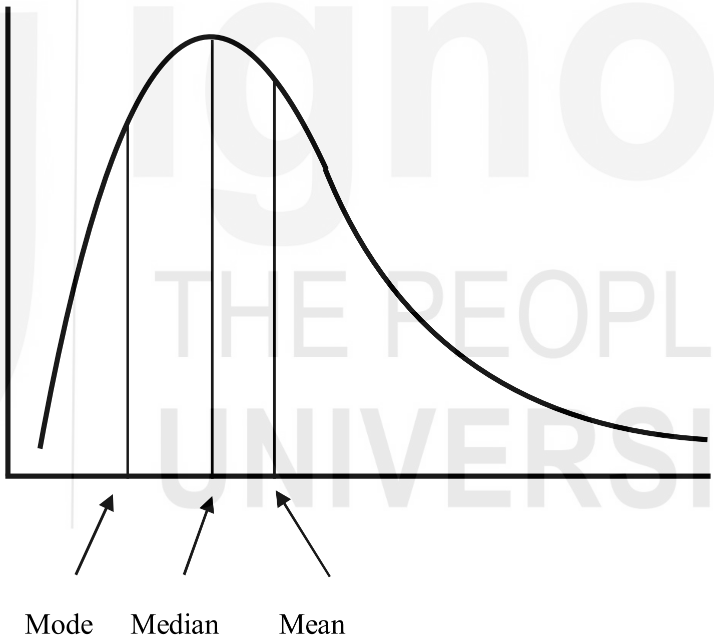


Fig. 8.4: Positive Skewness

2) **Negative Skewness**

On the other hand, if the distribution of the scores fall more towards the right side, the distribution is said to be negatively skewed (refer to figure 8.5). The median here is greater than the mean and that is why, the mean lies to the left of the median. Here, more individuals obtain scores that are higher than the mean.

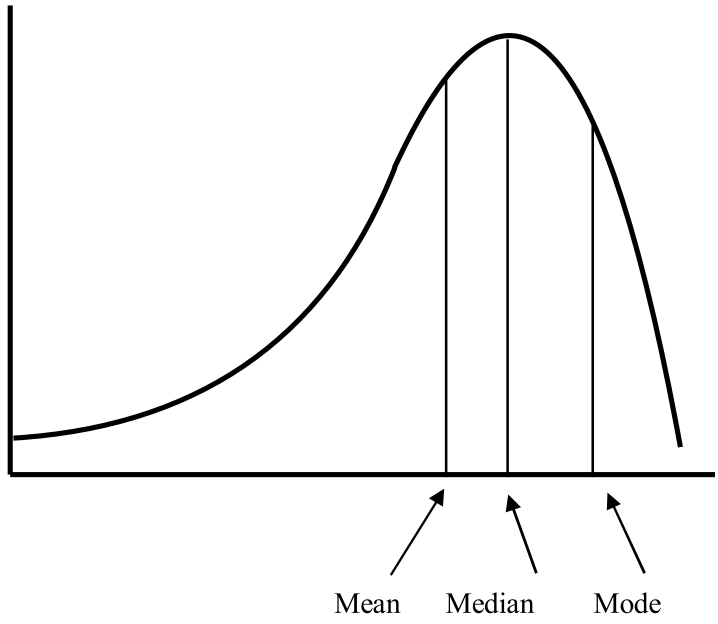


Fig. 8.5: Negative Skewness

Skewness can be computed with the help of the following formula:

$$S_k = 3\ (M- M_d)/\ SD$$

Where,

S_k= Skewness

M = Mean

M_d = Median

SD= Standard Deviation

There is another formula to compute skewness, that is used when the information about percentiles is available:

$$S_k = P_{90} + P_{10}/\ 2 - P_{50}$$

Where,

S_k= Skewness

P_{90} = 90th Percentile

$P1_0$ = 10th Percentile

P_{50} = 50th Percentile

Check Your Progress IV

1) What is the meaning of divergence from normality?

....

....

....

....

2) What is Kurtosis?

....

....

....

....

....

3) What is Skewness?

....

....

....

....

....

4) What is the difference between Kurtosis and Skewness?

....

....

....

....

....

5) What is the difference between positive skewness and negative skewness?

....

....

....

....

....

8.6 LET US SUM UP

To sum up, in the present unit we discussed about the concept of probability. Probability refers to chance or likelihood. For example, if you say that, "it will probably be hot tomorrow" or "Probably the teacher might not come tomorrow", the sentences reveal that there are chances for the events to occur tomorrow but there is no certainty. Some of the significant concepts related to probability were also discussed like, sample space, events and power set, mutually exclusive/ disjoint events, exhaustive/collective events independent and dependent events, equality likely events and complementary events were also discussed. Further, theconcept and nature of normal probability distribution were also highlightedwith the help of a figure showing normal curve.

A normal curve is bell shaped curve, bilaterally symmetrical and continuous frequency distribution curve. Such a curve is formed as a result of plotting frequencies of scores of a continuous variable in a large sample. The curve is known as normal probability distribution curve because its yordinates provides relative frequencies or the probabilities instead of the observed frequencies. Standard score or z-score was also discussed in detail and the discussion covered concpt, properties, uses and computation of z-score. Lastly, the unit explained divergence from normality, where kurtosis and skewness, with their types was discussed with help of figures.

8.7 REFERENCES

https://faculty.elgin.edu/dkernler/statistics/ch07/7-2.html accessed on 14/11/17

http://esminfo.prenhall.com/samplechps/larson/pdfs/ch05.pdf accessed on 14/11/17

http://statistics.wikidot.com/ch7 accessed on 15/11/17

https://www.uth.tmc.edu/uth_orgs/educ_dev/oser/L1_6.HTM accessed on 4/6/18

http://influentialpoints.com/Training/z_scores-principles-properties-assumptions.htm accessed on 5/10/18

https://www.spcforexcel.com/knowledge/basic-statistics/are-skewness-and-kurtosis-useful-statistics accessed on 5/10/18

Comparison of Independent Component Analysis techniques for Acoustic Echo Cancellation during Double Talk scenario. - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/Types-of-kurtosis_fig4_275208180 [accessed 15 Oct, 2018].

A. K. Kurtz et al. (1979). Statistical Methods in Education and Psychology, New York: Springer-Verlag New York Inc.

Griffiths, D. et al. (1998). Understanding Data. Principles and Practice of Statistics. Wiley, Brisbane.

Heiman, G.W. (2001). Understanding research method and statistics. An integrated introduction for psychology. 2nd Edn.Houghton Miffin Company, Boston, NewYork.

8.8 KEY WORDS

Events: A set of outcomes of an experiment. It is the sub set of the sample space.

Kurtosis: Kurtosis is a measure of “tailedness” of the probability distribution of a random variable. In other words, it is a measure whether a data is heavy tailed or light tailed in relation to normal distribution.

Normal Probability Curve: A normal curve is a bell shaped curve, bilaterally symmetrical and continuous frequency distribution curve.

Normal Probability Distribution: A continuous probability distribution for a variable is called as normal probability distribution or simply normal distribution. It is also known as Gaussian/ Gauss or LAPlace – Gauss distribution.

Probability: In statistics, the term 'Probability' refers to the expected frequency (chance) of occurrence of an event among all possible similar events.

Skewness: Skewness is a measure of asymmetry of the probability distribution of a random variable about its mean.

Standard score: Standard score or z-score is a transformed score which shows the number of standard deviation units by which the value of observation (the raw score) is above or below the mean.

8.9 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress I

1. State whether the statements are True or False

Sr. No.	Statements	True/ False
A	Events are said to be equally likely events if each event has a fair enough chance to occur approximately the same number of times.	True
B	A probability ratio always ranges between the limit of -1.00 to +1.00.	False
C	Two events are said to be mutually exclusive if both occur simultaneously in a single trial.	False
D	Two or more events are said to be independent events if the outcome of one event does not effect as well as is not affected by the outcome of the other event.	True
E	If two events are mutually exclusive and exhaustive, they both are said to complement each other.	True

Check Your Progress II

1. **Fill in the Blanks**

   A. The Normal Probability Curve is considered as an ideal degree of peak or kurtosis.

   B. In a normal probability curve, the value of mean, median and mode are equal.

   C. The normal distribution is a continuous distribution and plays significant role in statistical theory and inference.

   D. The area of unit under the normal curve is said to be equal to one.

E. The normal distribution is determined by two parameters, mean and variance.

Check Your Progress III

1. Fill in the Blanks

A. The mean of the z-scores is always zero.

B. z- score is used for standardising the raw data.

C. The variation of z-scores ranges from -3 standard deviations (which would fall to the far left of the normal distribution curve) up to +3 standard deviations.

D. The standard deviation of the z-scores is always one.

E. The standard score is a score that informs about the value and also where the value lies in the distribution.

Check Your Progress IV

1. What is the meaning of divergence from normality?

   When the frequency curve is more peaked or flatter than the normal probability distribution curve, the distribution is said to be diverged from normality.

2. What is Kurtosis?

   Kurtosis is a measure of "tailedness" of the probability distribution of a random variable. It is a measure of whether a data is heavy tailed or light tailed in relation to normal distribution.

3. What is Skewness?

   Skewness determines the lack of symmetry in the curve and can be measured in terms of Pearson's measure and percentiles.

4. What is the difference between Kurtosis and Skewness?

   Kurtosis is a measure of "tailedness" of the probability distribution of a random variable. In other words, it is a measure of whether a data is heavy tailed or light tailed in relation to normal distribution. On the other hand, Skewness is a measure of asymmetry of the probability distribution of a random variable about its mean.

5. What is the difference between positive skewness and negative skewness?

   A distribution curve is said to be positively skewed when the distribution of scores are more at the left end. In a positively skewed distribution the median is less than the mean which means that the mean lies to the right of the median. On the other hand, if the distribution of the scores fall more towards the right side, the distribution is said to be negatively skewed. The median here is greater than the mean and that is why, the mean lies to the left of the median.

8.10 UNIT END QUESTIONS

1) Discuss the concept and related aspects of probability.
2) Differentiate between mutual exclusive and exhaustive events.
3) Discuss the concept, nature and properties of Normal Distribution Curve.
4) Discuss the computation of z-scores and its properties.
5) Discuss the different types of divergence from normality.

FORMULAE AT A GLANCE

Measure	Formulae
Percentile	$P = L + [(pN/ 100- F)/ f]$ X i
Percentile Ranks for ungrouped data	PR= 100-100R- 50/ N
Percentile Ranks for ungrouped data	PR= 100/ N [F + (X-L/ i) x f]
Mean for ungrouped data	$M = \sum X/ N$
Mean for grouped data	$M = \sum fX/ N$
Assumed Mean	$M = AM+(\sum fx'/N \times i)$
Median for ungrouped data (Even)	M_d= (N +1)/ 2^{th} score
Median for ungrouped data (Odd)	M_d = $(N/2)^{th}$ score + [$(N/2)^{th}$ score + 1]/ 2
Median for grouped data	$M_d=L+[(N/2)-F/f_m]$ x i
Mode for grouped data (first method)	M_o=3Mdn−2M
Mode for grouped data (Second method)	M_o= L + [d_1/ d_1 +d_2] x i
Range	R=H-L
Quartile Deviation	$QD=(Q_3 - Q_1)/2$
Average Deviation for ungrouped data	$AD=\sum \vert x \vert /N$
Average Deviation for grouped data	$AD= \sum \vert fx \vert /N$
Standard Deviation for ungrouped data	$SD= \sqrt{\sum x^2 /N}$
Standard Deviation for grouped data	$SD= \sqrt{\sum fx^2 /N}$
Standard Deviation for grouped data with assumed mean	$SD= i \sqrt{\sum fx'^2/N-(\sum(fx')/N)^2}$
Pearson product Moment Correlation	$\Sigma xy/ \sqrt{\Sigma x^2 \Sigma y^2}$
Pearson product Moment Correlation	$r = N\Sigma XY- \Sigma X \Sigma Y/ \sqrt{[N \Sigma X^2 -(\Sigma X)^2] [N \Sigma Y^2 - (\Sigma Y)^2]}$
Spearman's Rho	$p = 1- [(6\Sigma d^2) / [N (N^2 - 1)]$
Standard Score (z-score)	z = X- M/ SD

SUGGESTED READINGS

Cohen, B. H. (2008). *Explaining Psychological Statistics*. Hoboken, NJ: John Wiley & Sons, Inc.

Elmes, D. G., Kantrowitz, B. H. and Roedier, H. L. III. (2003). *Research Methods in Psychology*. Belmont, CA: Thomson/Wandsworth Learning.

Foster, J. J. (2001). *Data Analysis: Using SPSS for Windows*. London, UK: SAGE.

Guilford, J. P. and Fruchter, B. (1988). *Fundamental Statistics in Psychology and Education*. New Delhi: McGraw-Hill

Howell, D. (2010). *Statistical Methods for Psychology* (7th ed.). Belmont, CA: Wadsworth.

Jones, S. (2010). Statistics in Psychology. Basingstoke: Palgrave Macmillan

Kerlinger, F. R. (1986). Foundations of behavioural research. Bangalore: Prism Books.

Landau, S. and Everitt, B. S. (2004). A Handbook of Statistical Analyses using SPSS. London, UK: Chapman and Hall.

Levin, J. and Fox, J. A. (2006). Elementary Statistics in Social Research. Delhi: Pearson Education.

Mangal, S. K. (2002). Statistics in Psychology and Education. New Delhi: HI Learning Private Limited.

Mulhern, G., & Greer, B. (2011). Making Sense of Data and Statistics in Psychology. Basingstoke: Palgrave Macmillan.

Salkind, N. J. (2014). Statistics for People who (think that) hate statistics. New Delhi: SAGE.

Tate, M. W.(1955). Statistics in Education. New York: Macmillion Co.

Stephens, L. (2009). Statistics in psychology. New York: McGraw-Hill.

Veeraraghavan, V and Shetgovekar, S. (2016). Textbook of Parametric and Nonparametric Statistics. Delhi: Sage.

WORKBOOK: 'PRACTICE MAKES LEARNER PERFECT'

1) Compute 10^{th} Percentile for the following data

Table 2.7: Data for computation of Percentile	
Class Interval	***f***
25-29	16
20-24	7
15-19	8
10-14	5
5-9	10
0-4	4
	N= 50

2) Scores obtained by employees on organisational citizenship behaviour are given below. Compute percentile rank for the score 43.

45, 43, 46, 56, 66, 54, 50, 41, 34, 22

Scores	Rank order

3) Create a frequency distribution and cumulative frequency for the following data with the help of class interval 5:

10, 1, 2, 3, 4, 10, 12, 13, 14, 15, 3, 15, 6, 7, 9, 12, 10, 11, 12, 13, 15, 17, 18, 19

Class Interval	Talliies	Frequency (f)	Cumulative frequency

4) Compute mean, median and mode for the following data

a) 39, 41, 56, 72, 83, 92, 35, 36, 39

b) 7, 15, 14, 11, 21, 18, 11, 17, 32, 45, 15, 13, 12, 11, 20, 15, 11, 10, 6, 7, 5, 15, 12, 11, 10, 11, 5, 6, 11, 10, 7

5) Compute Mean, Median and Mode from the following data:

Scores	f	Mid point (X)	fX
25-29	2		
20-24	7		
15-19	1		
10-14	2		
5-9	3		
0-4	1		

6) Find the average deviation of the scores 19, 10, 6, 8, 12.

7) Compute average deviation from the following distribution:

Scores	***f***	**Mid Point X**	***f*X**	**x= (X- M)**	***fx***	**|fx|**
100-104	2					
95-99	0					
90-94	3					
75-89	1					
70-74	4					
65-69	2					
60-64	3					

8) Following are the marks of 10 students out of 50. Compute the Quartile Deviation (QD):

20, 36, 15, 40, 45, 32, 25, 35, 38, 46

9) Compute standard deviation for the following data:

Data (x)	*X*- M	x^2
20		
30		
12		
14		
16		
18		
30		
20		
20		

Formula

Computation

10) Compute Pearsons Product Moment Correlation for the following data:

Individuals	**Data 1 (X)**	**Data 2 (Y)**	**x**	***y***	***xy***	***x^2***	***y^2***
A	8	2					
B	4	7					
C	3	10					
D	2	10					
E	9	1					
F	8	3					
G	7	2					
H	4	8					
I	3	7					
J	2	10					

Formula

Computation

11) Compute Pearsons Product Moment Correlation for the following data:

Students	Marks in History (X)	Marks in English (Y)	x	*y*	*xy*	x^2	y^2
A	8	4					
B	7	4					
C	10	5					
D	15	10					
E	10	7					

Formula

Computation

12) Compute Pearsons Product Moment Correlation for the following data:

Students	**Marks in Mathematics (X)**	**Marks in Science (Y)**	**x**	y	xy	x^2	y^2
A	**4**	**10**					
B	**5**	**14**					
C	**3**	**12**					
D	**9**	**11**					
E	**7**	**15**					

Formula

Computation

13) Compute Spearman's rho for the following data:

Employees	Scores on Emotional Intelligence (X)	Scores on Work Motivation (Y)	Rank in Emotional Intelligence (R_1)	Rank in Work Motivation (R_2)	Difference in Ranks, irrespective of their signs ($R_1 - R_2 = d$)	Difference Squared (d^2)
A	20	45				
B	25	40				
C	22	39				
D	21	37				
E	29	30				
F	28	32				
G	34	34				

Formula

Computation

14) Compute Spearman's rho for the following data:

Employees	Scores on Test X (X)	Scores on Test Y (Y)	Rank in Test X (R_1)	Rank in Test Y (R_2)	Difference in Ranks, irrespective of their signs ($R_1 - R_2 = d$)	Difference Squared (d^2)
A	10	16				
B	15	10				
C	5	5				
D	15	16				
E	4	4				
F	20	16				
G	20	5				
H	12	25				
I	20	36				
J	38	20				

Formula

Computation

15) Compute Spearman's rho for the following data:

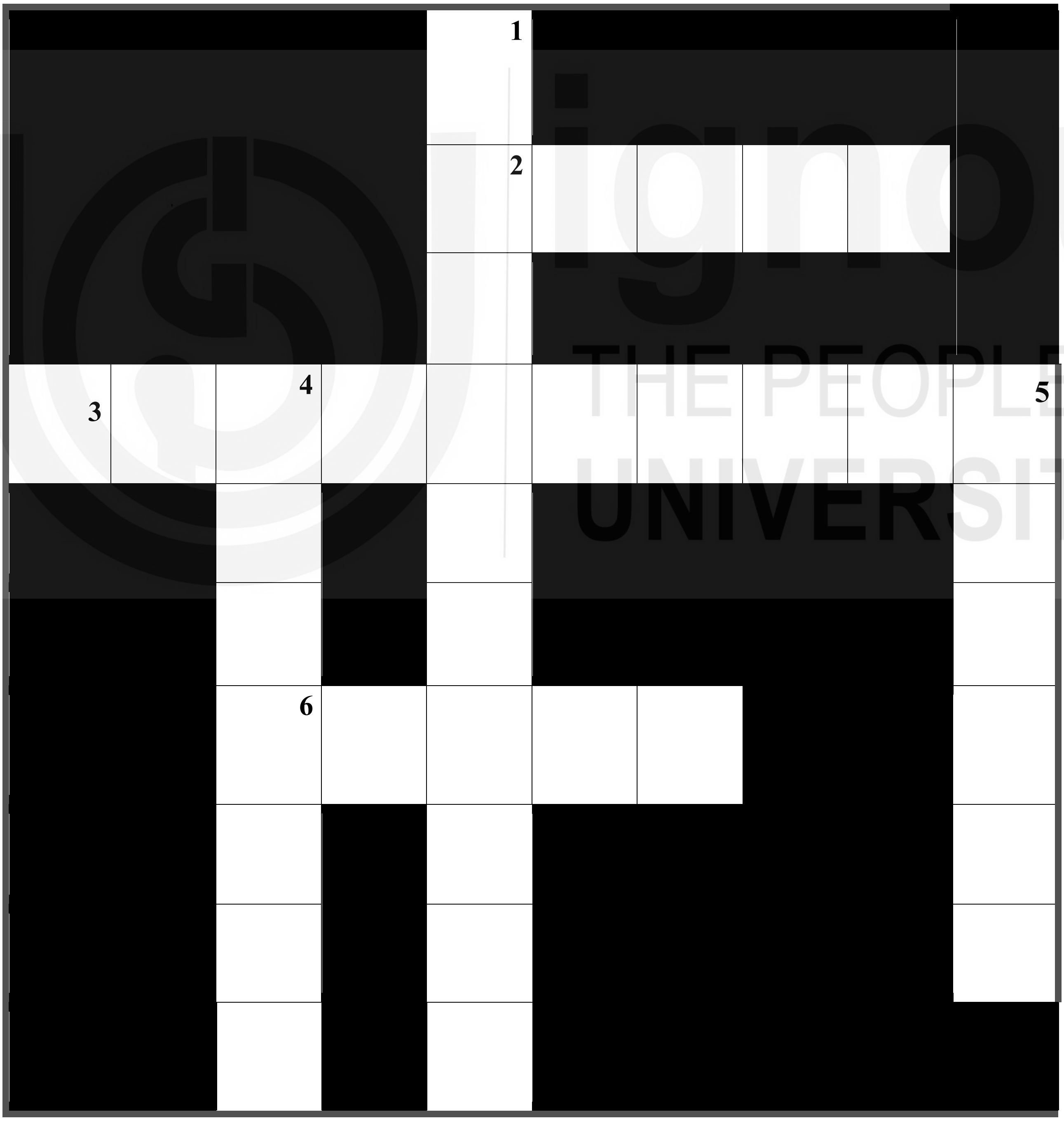
Individual	Scores on Perceived Parental Behaviour (X)	Scores on Self Concept (Y)	Rank in Perceived Parental Behaviour (R_1)	Rank in Self Concept (R_2)	Difference in Ranks, irrespective of their signs ($R_1 - R_2 = d$)	Difference Squared (d^2)
A	**36**	**23**				
B	**30**	**29**				
C	**25**	**20**				
D	**31**	**24**				
E	**40**	**40**				
F	**39**	**37**				
G	**41**	**30**				
H	**32**	**21**				

Formula

Computation

16) Crossword

1 Down: It can be can be described as a branch or sub field of mathematics that mainly deals with the organisation as well as analysis and interpretation of a group of numbers.	**2 Across:** ____________ error occurs when the null hypothesis is true but we reject it.
4 Down: Product moment coefficient of correlation was developed by ____________.	**3 Across:** _____________is a tentative statement that is tested with the help of scientific research
5 Down:__________ is a group of individuals who participate in the research.	**6 Across:** This scale has all the properties of all the scales, nominal, ordinal and interval scale, but also has an absolute zero, that indicates presence or absence of certain property or characteristics.



ANSWERS TO EXERCISES IN WORKBOOK

1) 5

2) PR = 30

3)

Class Interval	Talliies	Frequency (*f*)	Cumulative frequency
16- 20	///	3	24
11-15	//////////	10	21
6-10	///// /	6	11
1-5	/////	5	5

4) (a) Mean = 54. 78, Median = 41, Mode = 39

(b) Mean= 13. 35, Median = 11, Mode = 11

5) Mean= 17, Median = 20.21, Mode = 22.23

6) AD= 3.6

7) AD= 12.53

8) QD= 7.5

9) SD= 6.32

10) r = -0.97

11) r = 0.95

12) r = 0.14

13) p = - 0.75

14) p= 0.51

15) p = 0.71

16) Crossword

1 Down : STATISTICS

2 Across : TYPE I

4 Down : PEARSON

3 Across : HYPOTHESIS

5 Down : SAMPLE

6 Across : RATIO