Women University Press
Jayoti Publication Desk
ISBN No. : 978-93-90892-23-5

TEXT BOOK ON QUANTITATIVE TECHNIQUES

JV'n Dr. Vishal Saxena

JAYOTI VIDYAPEETH WOMEN'S UNIVERSITY, JAIPUR

UGC Approved Under 2(f) & 12(b) | NAAC Accredited | Recognized by Statutory Councils

Printed by :
JAYOTI PUBLICATION DESK

Published by :
Women University Press
Jayoti Vidyapeeth Women's University, Jaipur

Faculty of Agriculture & Veterinary Science

Title: Text Book on Quantitative Techniques

Author Name Dr. Vishal Saxena

Published By: Women University Press

Publisher's Address: Jayoti Vidyapeeth Women's University, Jaipur
Vedaant Gyan Valley,
Village-Jharna, Mahala Jobner Link Road, NH-8
Jaipur Ajmer Express Way,
Jaipur-303122, Rajasthan (INDIA)

Printer's Detail: Jayoti Publication Desk

Edition Detail: I

ISBN: 978-93-90892-23-5

INDEX

S.N.	CHAPTER	PAGE NO.
1	INTRODUCTION AND GRAPHICAL REPRESENTATION	1
2	MEASURE OF CENTRAL TENDENCY AND MEASURE OF DISPERSION	13
3	INDEX NUMBERS	29
4	SKEWNESS AND KURTOSIS	35
5	CORRELATION AND REGRESSION ANALYSIS	39
6	PROBABILITY AND ITS DISTRIBUTION	47
7	TIME SERIES ANALYSIS	56
8	REFERENCES	64

CHAPTER 1. INTRODUCTION AND GRAPHICAL REPRESENTATION

QUANTITATIVE TECHNIQUES:

Quantitative techniques are needed to process the information needed for effective planning, leading organizing and controlling. Qualitative and quantitative methods are productive tools in solving organizational problems. They are behavioral and mathematical techniques respectively that can provide a diversity of knowledge. Quantitative analysis concentrates on facts, data and numerical aspects associated with the problem.

The emphasis is on the development of mathematical expression to describe the objectives and constraints connected with the problem. Thus, the administrator's quantitative knowledge can help enhance the decision- making process. In this approach, past data is used in determining decisions that would prove most valuable in the future. The use of past data in a systematic manner and constructing it into a suitable model for future use comprises a major part of scientific management.

For example, consider a person investing in fixed deposit in a bank, or in shares of a company, or mutual funds, or in Life Insurance Corporation. The expected return on investments will vary depending upon the interest and time period. We can use the scientific management analysis to find out how much the investments made will be worth in the future. There are many scientific method software packages that have been developed to determine and analyze the problems.

In case of non-availability of past data where quantitative data is limited, qualitative factors play a major role in making decisions. Qualitative factors are important in situations like the introduction of breakthrough technologies. In today's complex and competitive global market, use of quantitative techniques with support of qualitative factors is paramount.

Application of scientific management and Analysis is more appropriate when there is not much of variation in problems due to external factors, and where input values are steady. In such cases, a model can be developed to suit the problem which helps us to take decisions faster. In today's complex and competitive global marketplace, use of Quantitative Techniques with support of qualitative factors is necessary.

Today, several quantitative techniques are available to solve managerial problems and use of these techniques helps managers to become explicit about their objectives and provides additional information to select on optimal decision. This approach starts with data like raw material for a factory which is manipulated or processed into information that is valuable to people making decision. This processing and manipulating of raw data into meaningful information is the heart of scientific management analysis.

APPLICATIONS FOR QUANTITATIVE TECHNIQUES IN BUSINESS DECISION MAKING

A small business owner is always making decisions under uncertainty. In the world of business, nothing is ever done with total confidence that you have made the right decision. Fortunately, numerous quantitative techniques are available to help organize and assess the risks of various issues.

Quantitative models give managers a better grasp of the problems so that they can make the best decisions based on the information available. Quantitative techniques are used by managers in practically all aspects of a business.

PROJECT MANAGEMENT

Quantitative methods have found wide applications in project management. These techniques are used for optimizing the allocation of manpower, machines, materials, money and time. Projects are scheduled with quantitative methods and synchronized with delivery of material and workforce.

PRODUCTION PLANNING AND SCHEDULING

Determining the size and location of new production facilities is a complex issue. Quantitative techniques aid in evaluating multiple proposals for costs, timing, location and availability of transportation. Product mix and scheduling get analyzed to meet customer demands and maximize profits.

PURCHASING AND INVENTORY

Predicting the amount of demand for a product is always dicey. Quantitative techniques offer guidance on how much materials to buy, levels of inventory to keep and costs to ship and store finished products.

MARKETING

Marketing campaigns get evaluated with large amounts of data. Marketers apply quantitative methods to set budgets, allocate media purchases, adjust product mix and adapt to customers' preferences.

Surveys produce data about viewers' responses to advertisements. How many people saw the ads, and how many purchased the products. All of this information is evaluated to get the return on investment of dollars in an advertising campaign.

FINANCE

Financial managers rely heavily on quantitative techniques. They evaluate investments with discounted cash flow models and return on capital calculations. Products get analyzed for profit contribution and cost of production. Workers are scrutinized for productivity standards and hiring or firing to meet changing workloads.

Predicting cash flow is always a critical concern for managers, and quantitative measurements help them to predict cash surpluses and shortfalls. They use probabilities and statistics to prepare annual profit plans.

RESEARCH AND DEVELOPMENT

Risking funds on research and development is always a best-guess scenario. The outcomes are never certain. So, managers look to mathematical projections about the probability of success and eventual profitability of products to make investment decisions.

AGRICULTURE

Operations research techniques have long been employed by farmers. They utilize decision trees and make assumptions about weather forecasts to decide which crops to plant. If forecasters predict cold weather, is it more profitable to plant corn or wheat? What happens if the weather is warm? These are all probabilities that farmers use to plan their crop rotations.

A variety of quantitative methods of analysis are finding more applications in business as managers learn how to use these techniques to provide more insight into problems and aid in daily decision-making.

STATISTICS

Statistics is a form of mathematical analysis that uses quantified models, representations and synopses for a given set of experimental data or real-life studies. Statistics studies methodologies to gather, review, analyze and draw conclusions from data.

Statistics is a term used to summarize a process that an analyst uses to characterize a data set. If the data set depends on a sample of a larger population, then the analyst can develop interpretations about the population primarily based on the statistical outcomes from the sample. Statistical analysis involves the process of gathering and evaluating data and then summarizing the data into a mathematical form.

Statistics is used in various disciplines such as psychology, business, physical and social sciences, humanities, government, and manufacturing. Statistical data is gathered using a sample procedure or other method. Two types of statistical methods are used in analyzing data: descriptive statistics and inferential statistics. Descriptive statistics are used to synopsize data from a sample exercising the mean or standard deviation. Inferential statistics are used when data is viewed as a subclass of a specific population.

DATA

Data can be defined as a systematic record of a particular quantity. It is the different values of that quantity represented together in a set. It is a collection of facts and figures to be used for a specific purpose such as a survey or analysis. When arranged in an organized form, can be called information. The source of data (primary data, secondary data) is also an important factor.

TYPES OF DATA

Data may be qualitative or quantitative. Once you know the difference between them, you can know how to use them.

- Qualitative Data: They represent some characteristics or attributes. They depict descriptions that may be observed but cannot be computed or calculated. For example, data on attributes such as intelligence, honesty, wisdom, cleanliness, and creativity collected using the students of your class a sample would be classified as qualitative. They are more exploratory than conclusive in nature.
- Quantitative Data: These can be measured and not simply observed. They can be numerically represented and calculations can be performed on them. For example, data on the number of students playing different sports from your class gives an estimate of how many of the total students play which sport. This information is numerical and can be classified as quantitative.

DATA COLLECTION

Depending on the source, it can classify as primary data or secondary data. Let us take a look at them both.

PRIMARY DATA

These are the data that are *collected for the first time* by an investigator for a specific purpose. Primary data are ‘pure’ in the sense that no statistical operations have been performed on them and they are original. An example of primary data is the Census of India.

SECONDARY DATA

They are the data that are *sourced from someplace* that has originally collected it. This means that this kind of data has already been collected by some researchers or investigators in the past and is available either in published or unpublished form. This information is impure as statistical operations may have been performed on them already. An example is an information available on the Government of India, the Department of Finance’s website or in other repositories, books, journals, etc.

CLASSIFICATION OF DATA

The process of arranging data into homogenous groups or classes according to some common characteristics present in the data is called classification.

For example: During the process of sorting letters in a post office, the letters are classified according to the cities and further arranged according to streets.

BASES OF CLASSIFICATION

Classification can be divided in following four bases:

(1) QUALITATIVE BASE

When the data are arranged by qualitative characteristics such as sex, literacy and intelligence, etc.

(2) QUANTITATIVE BASE

When the data are classified by quantitative characteristics like height, weight, age, income, etc.

(3) GEOGRAPHICAL BASE

When the data are classified by geographical regions or location, like states, provinces, cities, countries, etc.

(4) CHRONOLOGICAL OR TEMPORAL BASE

When the data are classified or arranged by their time of occurrence, such as years, months, weeks, days, etc.

For example: Time series data.

TABULATION OF DATA

The process of placing classified data into tabular form is known as tabulation. A table is a symmetric arrangement of statistical data in rows and columns. Rows are horizontal arrangements whereas columns are vertical arrangements. It may be simple, double or complex depending upon the type of classification.

TYPES OF TABULATION

(1) SIMPLE TABULATION OR ONE-WAY TABULATION

When the data are tabulated to one characteristic, it is said to be a simple tabulation or one-way tabulation.

For example: Tabulation of data on the population of the world classified by one characteristic like religion is an example of a simple tabulation.

(2) DOUBLE TABULATION OR TWO-WAY TABULATION

When the data are tabulated according to two characteristics at a time, it is said to be a double tabulation or two-way tabulation.

For example: Tabulation of data on the population of the world classified by two characteristics like religion and sex is an example of a double tabulation.

(3) COMPLEX TABULATION

When the data are tabulated according to many characteristics, it is said to be a complex tabulation.

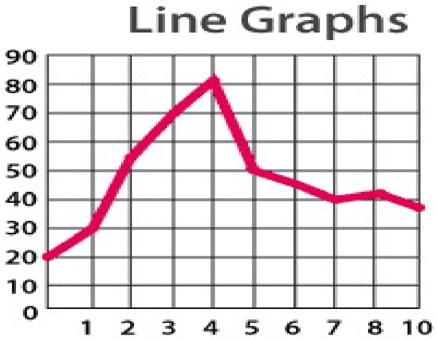
For example: Tabulation of data on the population of the world classified by three or morecharacteristics like religion, sex and literacy, etc. is an example of a complex tabulation.

Graphical Representation is a way of analysing numerical data. It exhibits the relation between data, ideas, information and concepts in a diagram. It is easy to understand and it is one of the most important learning strategies. It always depends on the type of information in a particular domain. There are different types of graphical representation. Some of them are as follows

(1) LINE GRAPHS

Linear graphs are used to display the continuous data and it is useful for predicting the future events over time.

Exp.



(2) BAR GRAPHS

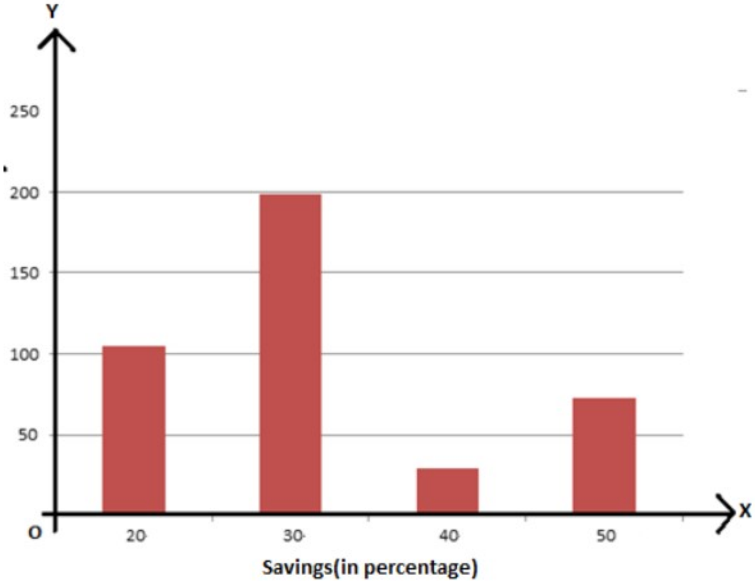
Bar Graph is used to display the category of data and it compares the data using solid bars to represent the quantities.

Exp.

In a firm of 400 employees, the percentage of monthly salary saved by each employee is given in the following table. Represent it through a bar graph.

Savings (in percentage)	20	30	40	50
Number of Employees(Frequency)	105	199	29	73

Sol.



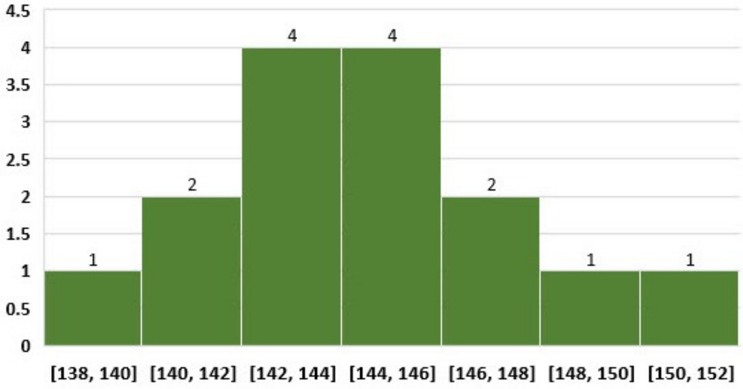
(3) HISTOGRAMS

The graph that uses bars to represent the frequency of numerical data that are organised into intervals. Since all the intervals are equal and continuous, all the bars have the same width.

Exp.

Mr. Larry, a famous doctor, is researching the height of the students studying in the 8th standard. He has gathered a sample of 15 students but wants to know which the maximum category is where they belong.

S.N.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Height	141	143	145	145	147	152	143	144	149	141	138	143	145	148	145

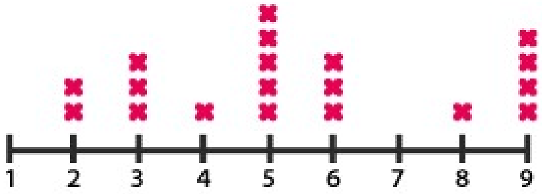


Here we can see the heights of the students on an average is in the range of 142 cm to 146 cm for 8th standard.

(4) LINE PLOT

It shows the frequency of data on a given number line. ‘ x ‘ is placed above a number line each time when that data occurs again.

Exp.



(5) CIRCLE GRAPH

Also known as pie chart that shows the relationships of the parts of the whole. The circle is considered with 100% and the categories occupied is represented with that specific percentage like 15%, 56% , etc.

Exp.

Let's construct a pie chart to visually display the favorite fruits of the students in the class based on the frequency table below.

Mango	Orange	Plum	Pineapple	Melon
45	30	15	30	30

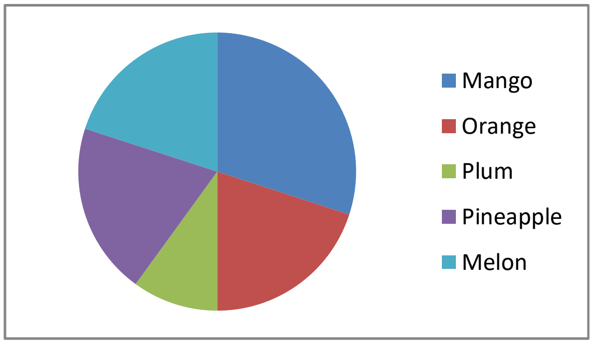
Category	Formula	Degrees
Mango	45/150×360	108
Orange	30/150×360	72
Plum	15/150×360	36
Pineapple	30/150×360	72
Melon	30/150×360	72

The total frequency sums up to 150.

Draw a circle and draw the radius. With the radius as the base, construct 108^0 using a protractor**.**

Subsequently, construct all other sectors with their respective angles.

Thus we get the pie chart as follows.



CHAPTER 2. MEASURE OF CENTRAL TENDENCY AND MEASURE OF DISPERSION

MEASURE OF CENTRAL TENDENCY

Measures of central tendency are the numbers which indicate the centre of a set of ordered numerical data. The most common measures of central tendency are the mean, median, and mode.

(A) ARITHMETIC MEAN

Arithmetic mean is the simple average of all items in a series. It is the simplest measure of central tendencies.

Basic formula: Arithmetic mean $=X_1 + X_2 + X_3 + \ldots\ldots.. + Xn / n = \sum X / n$

Exp. For the numbers 5, 6, 7, 8, 9

Arithmetic mean $= (5 + 6 + 7 + 8 + 9) / 5 = 35 /5=7$

METHODS OF CALCULATING SIMPLE ARITHMETIC MEAN

The three types of statistical series are

1. Individual series

2. Discrete series (Simple frequency distribution)

3. Continuous series (Grouped frequency distribution)

1. CALCULATION OF SIMPLE ARITHMETIC MEAN FOR INDIVIDUAL SERIES

For the individual series, arithmetic mean is calculated by

Mean$=\sum X / N$ = Total value of the items / No. of items

Exp.

Suppose the pocket allowance of 10 students in rupees are

15,20,30,22,25,18,40,50, 55,65.

Find out the average pocket allowance.

Sol.

Arithmetic mean = ∑X/n

= 15+20+30+22+25+18+40+50+55+65/10 = 340/10 = 34

Therefore, the average pocket allowance is = Rs 34

2. CALCULATION OF ARITHMETIC MEAN IN DISCRETE SERIES OR SIMPLE FREQUENCY DISTRIBUTION

i. Direct method

ii. Short-cut method

i. Direct method

Formula:- = ∑fX / ∑f

Exp. Following are the weekly wage earnings of 19 workers:

Wages (Rs) (X): 10 20 30 40 50

No. of workers (f): 4 5 3 2 5

Sol. Mean = ∑fX / ∑f = 560/19 = 29.47

Therefore mean wage earnings = Rs 29.47

ii. Short-cut method

In short-cut method, the following formula is to be applied to find mean

Mean =A+ ∑fd / ∑f

where A is the assumed mean, f denotes frequency and d=X-A (called deviation)

Exp. Find the mean for the following data

x:	900	950	1000	1100	1260	1440	1500
f:	26	22	18	19	15	3	2

Sol.

X	f	d = x-A	fd
900	26	-100	-2600
950	22	-50	-1100
1000=A(Let)	18	0	0
1100	19	100	1900
1260	15	260	3900
1440	3	440	1320
1500	2	500	1000
	$\sum f$ =105		$\sum fd$

Therefore, mean = $A+\dfrac{\sum fd}{\sum f}=1000+\dfrac{4420}{105}$ = 1042.1

3. CALCULATION OF SIMPLE ARITHMETIC MEAN IN CASE OF CONTINUOUS SERIES OR GROUPED FREQUENCY DISTRIBUTION

i. Direct method

ii. Short-cut method

iii. Step-deviation method

i. **Direct method Formula:**

M = ∑fX / ∑f

ii. **Short-cut method Formula**:

M = A+ (∑fd / ∑f)

where A is the assumed mean, f denotes frequency and d=X-A (called deviation).

iii. **Step-deviation method:** Formula:-

Mean= A+ (∑f.u / ∑f) h

where u = (X-A) / h=d / h, A is the assumed mean, h is the width of the class interval.

Exp. (By Direct method Formula):

Marks in Statistics of student of Class XI are given below. Find out arithmetic mean.

Sol. :

Marks:	0-10	10-20	20-30	30-40	40-50
No. of students :	5	12	14	10	8
Mid-value:	5	15	25	35	45
fx:	25	180	350	350	360

Mean= ∑fX / ∑f =1265/49=25.82

Arithmatic mean=25.82 marks

Exp. (By Step-deviation method):

Find the mean for the following data:

Class-Interval	100-150	150-200	200-250	250-300	300-350	350-400	400-450	450-500
Freq.	24	40	33	28	30	22	16	7

Sol.:

Class-Interval	Freq.	x mid-value	d = x-A	u = d/h	f.u
100-150	24	125	-150	-3	-72
150-200	40	175	-100	-2	-80
200-250	33	225	-50	-1	-33
250-300	28	275=A(let)	0	0	0
300-350	30	325	50	1	30
350-400	22	375	100	2	44
450-450	16	425	150	3	48
450-500	7	475	200	4	28
	$\sum f$ =200				$\sum f.u =$ -35

Therefore, mean = A+ (∑f.u / ∑f) h

$$= 275 + \frac{-35}{200} \times 50 \ = 266.25$$

(B) MEDIAN:

Median is a centrally located value in a series for which half of the values (or items) of the series are above it and rest of them are below it.

OR

Median is the central value of the variable which divides the series into two equal parts such that half of the values (or items) of the series are above it and rest of them are below it. Median is defined as the value of the middle most term (or the mean of the values of the two middle terms) when the data are arranged in an ascending or descending order of magnitude.

METHODS OF CALCULATING MEDIAN

We know, there are three types of statistical series :

1. Individual series

2. Discrete series (Simple frequency distribution)

3. Continuous series (Grouped frequency distribution)

1. CALCULATION OF MEDIAN FOR INDIVIDUAL SERIES

The "Median" of a individual data set is dependent on whether the number of elements in the data set is odd or even. First reorder the data set from the smallest to the largest (i.e. ascending order).

For Odd series:

Formula:- Median = (n+1)/2th item

For even series:

Formula:- Median=Average of [(n/2)th item+(n/2+1)th item]

Exp. For odd Number of Elements

Data Set= 2, 6, 9, 3, 5, 4, 7

Reordered = 2, 3, 4, 5, 6, 7, 9

Median = (7+1)/2th item = 4th item = 5

Exp. For even Number of Elements

Data Set = 2, 6, 9, 3, 5, 4 Reordered = 2, 3, 4, 5, 6, 9

Median = (3rd item + 4th item) / 2 = (4+5)/2=4.5

2. CALCULATION OF MEDIAN FOR DISCRETE SERIES OR SIMPLE FREQUENCY DISTRIBUTION:

The "Median" of a discrete data set is dependent on whether the number of elements in the data set is odd or even. First reorder the data set from the smallest to the largest (i.e. ascending order).

The median is given by

For Odd series:

Formula:- Median = Size of (N+1)/2th item

For even series:

Formula:- Median=Average of [Size of (N/2)th item + Size of (N/2+1)th item]

Exp. from the following data calculate median

Marks	45	55	25	35	5	15
No. of students	40	30	30	50	10	20

Sol.:

Step I- First we will find out the commutative frequency

Marks in ascending order (x) :	5	15	25	35	45	55
No. of students (f) :	10	20	30	50	40	30
Commutative frequency C.f :	10	30	60	110	150	180

N = 180

N=180=even

Median = [size of (N/2)th item + size of (N/2+1)th item]/2

Step II – [Size of 90th item + size of 91st item]/2

=(35+35)/2

Median = 35

3. **CALCULATION OF MEDIAN FOR CONTINUOUS SERIES OR GROUPED FREQUENCY DISTRIBUTION**

In the case of grouped series, the median is calculated with the fallowing formula:

M=L+ [(N/2)-p.c.f.]/f x i

Where L= lower limit of median class interval (MCI)

p.c.f.=previous cumulative frequency of median class

f=frequency of median class

i= size of median class

N=total no of observation

Exp. From the following data, find median

Marks (x)	0-10	10-20	20-30	30-40	40-50	50-60
No. of Students (f)	10	20	30	50	40	30

Sol.:

I Commutative Frequency Table is

Marks (x)	No. of Students (f)	Cumulative frequency (cf)
0-10	10	10
10-20	20	30
20-30	30	60=p.c.f.
30-40 = MCI	50=f	110
40-50	40	150
50-60	30	180
	N=180	

II Size of N /2 item = size of 180/2 item = 90th item

III - Commutative frequency which includes 90^{th} item = 110 Class corresponding to 110 = 30-40, which is the median class

Now we applying the fallowing formula

M=L+ [(N/2)-p.c.f.]/f x i

M=30+[(90-60)/50]x10

Median = 36

(C). MODE

Mode is the value in a series which occurs most frequently or which has the greatest frequency. But it is not exactly true for every case or for every frequency distribution. it is that value around which the terms tend to concentrate most densely.

METHODS OF CALCULATING MEDIAN

Mode can be determined in the following three types of statistical series :

1. Individual series

2. Discrete series (Simple frequency distribution)

3. Continuous series (Grouped frequency distribution)

1. CALCULATION OF MODE FOR INDIVIDUAL SERIES

Exp. Find the mode of the following series: 8, 9, 11, 15, 16, 12, 15, 3, 7, 15

Sol. There are ten observations in the series, where the data 15 occurs maximum number of times (highest frequency). Therefore the mode is 15.

2. CALCULATION OF MODE FOR DISCRETE SERIES

EXP. Find the mode of the following series:

x:	**2**	**5**	**7**	**11**	**18**
f:	**5**	**12**	**19**	**7**	**5**

Sol.: There are five observations in the series, where the data 7 occurs maximum number of times (highest frequency). Therefore the mode is 7.

3. CALCULATION OF MODE FOR CONTINUOUS SERIES

In the case of grouped data, mode is determined by the following formula:

Mode = L + $[(f_1-f_0)/(2f_1-f_0-f_2)]$ x h

where

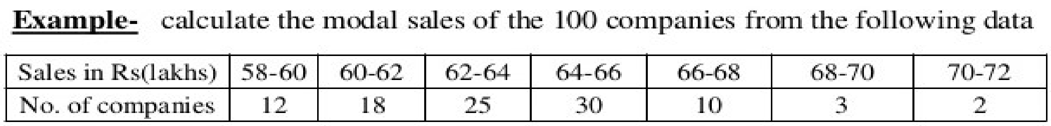
L = Lower limit of the modal class.

f_1 = modal class frequency

f_0 = frequency of preceding the modal class

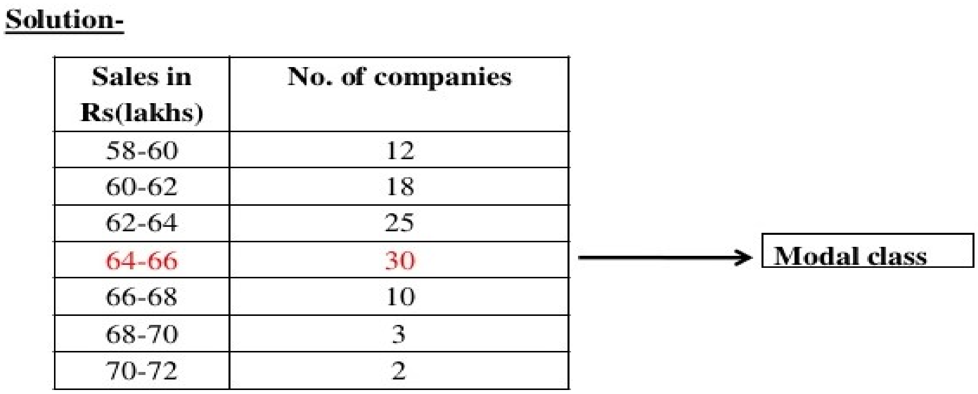
f_2 = frequency of succeeding the modal class

h= width of modal class



Example- calculate the modal sales of the 100 companies from the following data

Sales in Rs(lakhs)	58-60	60-62	62-64	64-66	66-68	68-70	70-72
No. of companies	12	18	25	30	10	3	2



Solution-

Sales in Rs(lakhs)	No. of companies
58-60	12
60-62	18
62-64	25
64-66	30
66-68	10
68-70	3
70-72	2

Modal class

Here the modal class interval is 64-66 (because it has the highest frequency)

Mode = $L + [(f_1-f_0)/(2f_1-f_0-f_2)] X h$

=64+[(30-25)/60-25-10] X 2= 64.4

RELATION BETWEEN MEAN, MEDIAN AND MODE

Mode = 3(median) – 2(mean)

MEASURE OF DISPERSION

The measure of dispersion shows the scatterings of the data. It tells the variation of the data from one another and gives a clear idea about the distribution of the data. The measure of dispersion shows the homogeneity or the heterogeneity of the distribution of the observations.

(A) RANGE

A range is the most common and easily understandable measure of dispersion. It is the difference between two extreme observations of the data set. If X_{max} and X_{min} are the two extreme observations then

Range = $X_{max} – X_{min}$

Coefficient of Range = $(X_{max} – X_{min})/(X_{max} + X_{min})$

Exp. For data

88, 89, 89, 89, 90, 91, 91, 91, 92

Range= 92-88 = 4

Exp.

71, 83, 85, 86, 90, 95, 100, 100, 100

Range=100-71 = 29

MERITS OF RANGE

- It is the simplest of the measure of dispersion
- Easy to calculate
- Easy to understand
- Independent of change of origin

DEMERITS OF RANGE

- It is based on two extreme observations. Hence, get affected by fluctuations
- A range is not a reliable measure of dispersion
- Dependent on change of scale

(B) QUARTILE DEVIATION

The values of the variate that divide the total frequency into four parts are called Quartile. The i^{th} Quartile is given by

$$Q_i = l + \left(\frac{iN/4 - C}{f} \right) \times h$$

For i=1, Q_1 is called lower (or first quartile).

For i=2, Q_2 is called median (or second quartile).

For i=3, Q_3 is called upper (or third quartile).

The difference between the upper and lower quartile is called the inter quartile range.

Quartile Deviation defines the absolute measure of dispersion. Whereas the relative measure corresponding to QD, is known as the coefficient of QD.

The Quartile Deviation(QD) is the product of half of the difference between the upper and lower quartiles. Mathematically we can define as:

Quartile Deviation = $(Q_3 – Q_1) / 2$

Coefficient of Quartile Deviation = $(Q_3 – Q_1) / (Q_3 + Q_1)$

A Coefficient of QD is used to study & compare the degree of variation in different situations.

Exp. Find the Quartile Deviation from the given below data:

Marks	0-5	5-10	10-15	15-20	20-25	25-30
No. of students	4	6	8	12	7	2

Sol.: Cumulative frequency table is

Class Interval	No. of students	c.f.
0-5	4	4
5-10	6	10
10-15	8	18
15-20	12	30
20-25	7	37
25-30	2	39

$$N = \sum f = 39$$

$N = 39/4 = 9.75$

Class of Q_1 is 5-10, Now $Q_1 = l + \left(\frac{N/4 - C}{f}\right) \times h = 5 + \left(\frac{9.75 - 4}{6}\right) \times 5 = 9.79$

and 3N/4 =29.25, therefore class of Q_3 is 15-20,

$$Q_3 = l + \left(\frac{3N/4 - C}{f}\right) \times h = 15 + \left(\frac{29.25 - 18}{12}\right) \times 5 = 19.69$$

Quartile Deviation= $\left(\frac{Q_3 - Q_1}{2}\right) = 4.95$

(C) MEAN DEVIATION

Mean deviation is the arithmetic mean of the absolute deviations of the observations from a measure of central tendency. If $x_1, x_2, \ldots, x_n$ are the set of observation, then the mean deviation of x about the average A (mean, median, or mode) is

Mean deviation from average A = $1⁄n\ [\sum_i |x_i - A|]$

For a grouped frequency, it is calculated as:

Mean deviation from average A = $1⁄N\ [\sum_i\ f_i\, |x_i - A|],\ N = \sum f_i$

Exp. Find the mean deviation about the mean for 6, 7, 10, 12, 13, 4, 8, 12

Sol. Mean is given by

$$\bar{x} = \frac{\sum x}{n} = \frac{72}{8} = 9$$

x	$x - \bar{x}$	$\lvert x - \bar{x} \rvert$
6	-3	3
7	-2	2
10	1	1
12	3	3

13	4	4
4	-5	5
8	-1	1
12	3	3
		$\sum \lvert x - \overline{x} \rvert = 22$

Mean deviation $= \dfrac{\sum \lvert x - \overline{x} \rvert}{n} = \dfrac{22}{8}$

= 2.75

MERITS OF MEAN DEVIATION

- Based on all observations
- It provides a minimum value when the deviations are taken from the median
- Independent of change of origin

DEMERITS OF MEAN DEVIATION

- Not easily understandable
- Its calculation is not easy and time
- Dependent on the change of scale
- Ignorance of negative sign creates artificiality and becomes useless for further mathematical treatment

(D) STANDARD DEVIATION

A standard deviation is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean. It is denoted by a Greek letter sigma, σ. It is also referred to as root mean square deviation.

VARIANCE

The square of the standard deviation is the **variance**. It is also a measure of dispersion.

METHODS TO FIND STANDARD DEVIATION

The standard deviation is given as

$\sigma = [\Sigma_i (y_i - \bar{y})^2 / n]^{½} = [(\Sigma_i y_i^2 / n) - \bar{y}^2]^{½}$

For a grouped frequency distribution, it is

$\sigma = [\Sigma_i f_i (y_i - \bar{y})^2 / N]^{½} = [(\Sigma_i f_i y_i^2 / N) - \bar{y}^2]^{½}$

METHODS TO FIND VARIANCE

For a individual series

$\sigma^2 = [\Sigma_i (y_i - \bar{y})^2 / n] = [(\Sigma_i y_i^2 / n) - \bar{y}^2]$

For a grouped frequency distribution, it is

$\sigma^2 = [\Sigma_i f_i (y_i - \bar{y})^2 / N] = [(\Sigma_i f_i y_i^2 / N) - \bar{y}^2]$

Exp. Find the Variance and Standard Deviation of the Following Numbers: 1, 3, 5, 5, 6, 7, 9, 10.

Sol.

Step 1: The mean = 46/ 8 = 5.75

Step 2: Deviations from mean: $(y_i - \bar{y})$=

(1 – 5.75), (3 – 5.75), (5 – 5.75), (5 – 5.75), (6 – 5.75), (7 – 5.75), (9 – 5.75), (10 – 5.75)

= -4.75, -2.75, -0.75, -0.75, 0.25, 1.25, 3.25, 4.25

Step 3: Squaring the above values we get $(y_i - \bar{y})^2$ =22.563, 7.563, 0.563, 0.563, 0.063, 1.563, 10.563, 18.063

Step 4:

$\Sigma_i (y_i - \bar{y})^2$ =22.563 + 7.563 + 0.563 + 0.563 + 0.063 + 1.563 + 10.563 + 18.063 = 61.504

Step 4:

n = 8, therefore variance (σ^2) = $\Sigma_i (y_i - \bar{y})^2$ /n =61.504/ 8 = 7.69

Now, Standard deviation (σ) = 2.77

and variance = (σ^2) = 7.69

MERITS OF STANDARD DEVIATION

- Squaring the deviations overcomes the drawback of ignoring signs in mean deviations
- Suitable for further mathematical treatment
- Least affected by the fluctuation of the observations
- The standard deviation is zero if all the observations are constant
- Independent of change of origin

DEMERITS OF STANDARD DEVIATION

- Not easy to calculate
- Difficult to understand for a layman
- Dependent on the change of scale

CHAPTER 3. INDEX NUMBERS

MEANING OF INDEX NUMBERS

The value of money does not remain constant over time. It rises or falls and is inversely related to the changes in the price level. A rise in the price level means a fall in the value of money and a fall in the price level means a rise in the value of money. Thus, changes in the value of money are reflected by the changes in the general level of prices over a period of time. Changes in the general level of prices can be measured by a statistical device known as 'index number.'

Index number is a technique of measuring changes in a variable or group of variables with respect to time, geographical location or other characteristics. There can be various types of index numbers, but, in the present context, we are concerned with price index numbers, which measures changes in the general price level (or in the value of money) over a period of time.

Price index number indicates the average of changes in the prices of representative commodities at one time in comparison with that at some other time taken as the base period. According to L.V. Lester, "An index number of prices is a figure showing the height of average prices at one time relative to their height at some other time which is taken as the base period."

FEATURES OF INDEX NUMBERS

The following are the main features of index numbers

(i) Index numbers are a special type of average. Whereas mean, median and mode measure the absolute changes and are used to compare only those series which are expressed in the same units, the technique of index numbers is used to measure the relative changes in the level of a phenomenon where the measurement of absolute change is not possible and the series are expressed in different types of items.

(ii) Index numbers are meant to study the changes in the effects of such factors which cannot be measured directly. For example, the general price level is an imaginary concept and is not capable of direct measurement. But, through the technique of index numbers, it is possible to have an idea of relative changes in the general level of prices by measuring relative changes in the price level of different commodities.

(iii) The technique of index numbers measures changes in one variable or group of related variables. For example, one variable can be the price of wheat, and group of variables can be the price of sugar, the price of milk and the price of rice.

(iv) The technique of index numbers is used to compare the levels of a phenomenon on a certain date with its level on some previous date (e.g., the price level in 1980 as compared to that in 1960 taken as the base year) or the levels of a phenomenon at different places on the same date (e.g., the price level in India in 1980 in comparison with that in other countries in 1980).

STEPS OR PROBLEMS IN THE CONSTRUCTION OF PRICE INDEX NUMBERS

The construction of the price index numbers involves the following steps or problems

1. Selection of Base Year

The first step or the problem in preparing the index numbers is the selection of the base year. The base year is defined as that year with reference to which the price changes in other years are compared and expressed as percentages. The base year should be a normal year.

In other words, it should be free from abnormal conditions like wars, famines, floods, political instability, etc. Base year can be selected in two ways- (a) through fixed base method in which the base year remains fixed; and (b) through chain base method in which the base year goes on changing, e.g., for 1980 the base year will be 1979, for 1979 it will be 1978, and so on.

2. Selection of Commodities

The second problem in the construction of index numbers is the selection of the commodities. Since all commodities cannot be included, only representative commodities should be selected keeping in view the purpose and type of the index number.

In selecting items, the following points are to be kept in mind

(a) The items should be representative of the tastes, habits and customs of the people.

(b) Items should be recognizable,

(c) Items should be stable in quality over two different periods and places.

(d) The economic and social importance of various items should be considered

(e) The items should be fairly large in number.

(f) All those varieties of a commodity which are in common use and are stable in character should be included.

After selecting the commodities, the next problem is regarding the collection of their prices:

3. Collection of Prices

(a) From where the prices to be collected;

(b) Whether to choose wholesale prices or retail prices;

(c) Whether to include taxes in the prices or not etc.

While collecting prices, the following points are to be noted:

(a) Prices are to be collected from those places where a particular commodity is traded in large quantities.

(b) Published information regarding the prices should also be utilised,

(c) In selecting individuals and institutions who would supply price quotations, care should be taken that they are not biased.

(d) Selection of wholesale or retail prices depends upon the type of index number to be prepared. Wholesale prices are used in the construction of general price index and retail prices are used in the construction of cost-of-living index number.

4. Selection of Average

Since the index numbers are, a specialised average, the fourth problem is to choose a suitable average. Theoretically, geometric mean is the best for this purpose. But, in practice, arithmetic mean is used because it is easier to follow.

5. Selection of Weights

Generally, all the commodities included in the construction' of index numbers are not of equal importance. Therefore, if the index numbers are to be representative, proper weights should be assigned to the commodities according to their relative importance.

SOME OF THE USES OF INDEX NUMBERS ARE DISCUSSED BELOW

Index numbers possess much practical importance in measuring changes in the cost of living, production trends, trade, income variations, etc.

1. In Measuring Changes in the Value of Money:

Index numbers are used to measure changes in the value of money. A study of the rise or fall in the value of money is essential for determining the direction of production and employment to facilitate future payments and to know changes in the real income of different groups of people at different places and times. As pointed out by Crowther, “By using the technical device of an index number, it is thus possible to measure changes in different aspects of the value of money, each particular aspect being relevant to a different purpose.”

2. In Cost of Living:

Cost of living index numbers in the case of different groups of workers throw light on the rise or fall in the real income of workers. It is on the basis of the study of the cost of living index that money wages are determined and dearness and other allowances are granted to workers. The cost of living index is also the basis of wage negotiations and wage contracts.

3. In Analysing Markets for Goods and Services:

Consumer price index numbers are used in analysing markets for particular kinds of goods and services. The weights assigned to different commodities like food, clothing, fuel, and lighting, house rent, etc., govern the market for such goods and services.

4. In Measuring Changes in Industrial Production

Index numbers of industrial production measure increase or decrease in industrial production in a given year as compared to the base year. We can know from such as index number the actual condition of different industries, whether production is increasing or decreasing in them, for an industrial index number measures changes in the quantity of production.

5. In Internal Trade

The study of indices of the wholesale prices of consumer and industrial goods and of industrial production helps commerce and industry in expanding or decreasing internal trade.

6. In External Trade:

The foreign trade position of a country can be accessed on the basis of its export and import indices. These indices reveal whether the external trade of the country is increasing or decreasing.

7. In Economic Policies

Index numbers are helpful to the state in formulating and adopting appropriate economic policies. Index numbers measure changes in such magnitudes as prices, incomes, wages, production, employment, products, exports, imports, etc. By comparing the index numbers of these magnitudes for different periods, the government can know the present trend of economic activity and accordingly adopt price policy, foreign trade policy and general economic policies.

8. In Determining the Foreign Exchange Rate:

Index numbers of wholesale price of two countries are used to determine their rate of foreign exchange. They are the basis of the purchasing power parity theory which determines the exchange rate between two countries on inconvertible paper standard.

CHAPTER 4.   SKEWNESS AND KURTOSIS

SKEWNESS

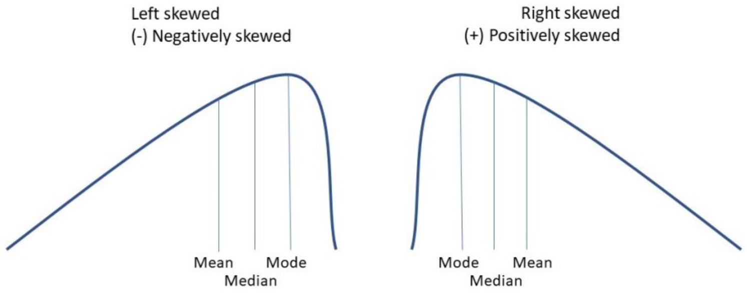
Skewness describes how much statistical data distribution is asymmetrical from the normal distribution, where distribution is equally divided on each side. If a distribution is not symmetrical or Normal, then it is skewed i.e. it is either the frequency distribution skewed to the left side or to the right side

If the distribution is symmetric then it has a skewness of 0 & its Mean = Median = Mode.

So basically, there are two types –

Positive: The distribution is positively skewed when most of the frequency of distribution lies on the right side of distribution & has a longer and fatter right tail. Where the distribution's Mean > median > Mode.

Negative: The distribution is negatively skewed when most of the frequency of distribution lies on the left side of distribution & has a longer and fatter left tail. Where the distribution's Mean < Median < Mode.



Skewness formula is represented as below –

$$\textbf{Skewness} = \frac{\sum_{i}^{N}(X_i - \bar{X})^3}{(N-1) * \sigma^3}$$

There are several ways to calculate the skewness of the data distribution. One of which is Pearson's first & second coefficients.

- Pearson's first coefficients (Mode Skewness): It is based on the Mean, Mode & Standard deviation of the distribution.

Formula: (Mean – Mode)/Standard Deviation.

- Pearson's second coefficients (Median Skewness): It is based on the Mean, Median & Standard deviation of the distribution.

Formula: (Mean – Median)/Standard Deviation.

As you can see above that Pearson's first coefficient of skewness has mode as its one variable to calculate it & it is useful only when data has more repetitive number into the data set, Like if there are only a few Repetitive data into data set which belong to mode, then Pearson's second coefficient of skewness is more reliable measure of central tendency as it considers median of the data set instead of mode.

LIMITS FOR SKEWNESS

For different limits of the two concepts, they are assigned different categories. For example, skewness is generally qualified as:

- Fairly symmetrical when skewed from -0.5 to 0.5

- Moderately skewed when skewed from -1 to -0.5 (left) or from 0.5 to 1 (right)
- Highly skewed when skewed from -1 (left) or greater than 1 (right)

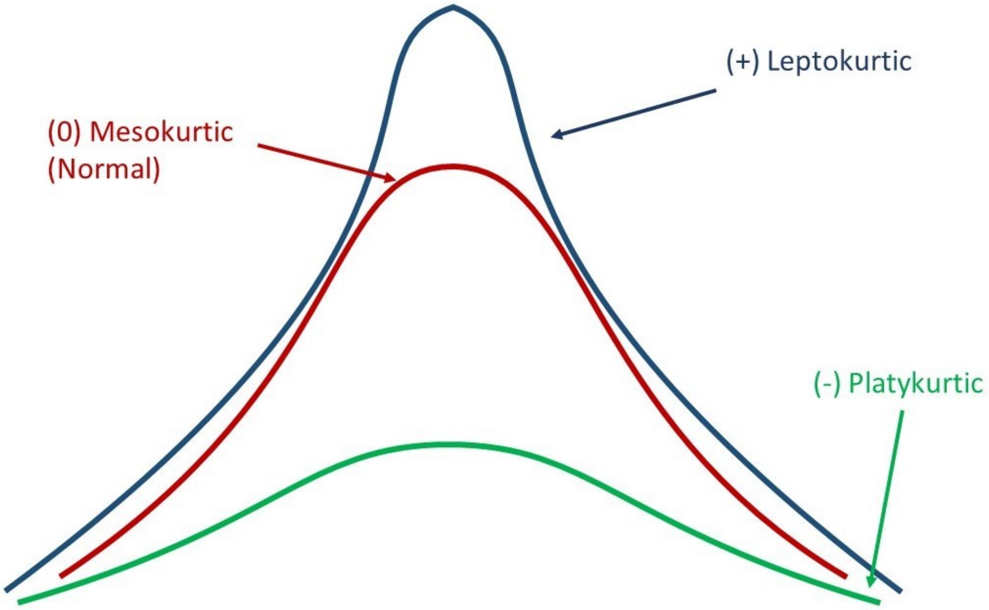
KURTOSIS

Kurtosis is descriptive or summary statistics and describes "peakedness" and frequency of extreme values in a distribution. Whereas skewness measures symmetry in a distribution, kurtosis measures the "heaviness" of the tails or the "peakedness".

Kurtosis is useful in statistics for making inferences, for example, as to financial risks in an investment: The greater the kurtosis, the higher the probability of getting extreme values.

So, the further the tails are from the mean the higher the risk of getting an extremely low return and the higher the chance of getting an extremely high return.

The degrees of kurtosis are labeled with leptokurtic, mesokurtic, platykurtic:



APPLICATIONS

Skewness is a descriptive statistic that can be used in conjunction with the histogram and the normal quantile plot to characterize the data or distribution.

Skewness indicates the direction and relative magnitude of a distribution's deviation from the normal distribution.

With pronounced skewness, standard statistical inference procedures such as a confidence interval for a mean will be not only incorrect, in the sense that the true coverage level will differ from the nominal (e.g., 95%) level, but they will also result in unequal error probabilities on each side.

Skewness can be used to obtain approximate probabilities and quantiles of distributions (such as value at risk in finance) via the Cornish-Fisher expansion.

Many models assume normal distribution; i.e., data are symmetric about the mean. The normal distribution has a skewness of zero. But in reality, data points may not be perfectly symmetric. So, an understanding of the skewness of the dataset indicates whether deviations from the mean are going to be positive or negative.

The sample kurtosis is a useful measure of whether there is a problem with outliers in a data set. Larger kurtosis indicates a more serious outlier problem, and may lead the researcher to choose alternative statistical methods.

D'Agostino's K-squared test is a goodness-of-fit normality test based on a combination of the sample skewness and sample kurtosis, as is the Jarque–Bera test for normality.

CHAPTER 5. CORRELATION AND REGRESSION ANALYSIS

The word correlation is used in everyday life to denote some form of association. We might say that we have noticed a correlation between foggy days and attacks of wheeziness. However, in statistical terms we use correlation to denote association between two quantitative variables. We also assume that the association is linear, that one variable increases or decreases a fixed amount for a unit increase or decrease in the other. The other technique that is often used in these circumstances is regression, which involves estimating the best straight line to summarise the association.

CORRELATION ANALYSIS

Correlation analysis is applied in quantifying the association between two continuous variables, for example, an dependent and independent variable or among two independent variables.

CORRELATION COEFFICIENT

The degree of association is measured by a correlation coefficient, denoted by r. It is sometimes called Pearson's correlation coefficient after its originator and is a measure of linear association. If a curved line is needed to express the relationship, other and more complicated measures of the correlation must be used.

TYPES OF CORRELATION:

In a bivariate distribution, the correlation may be:

1. Positive, Negative and Zero Correlation; and
2. Linear or Curvilinear (Non-linear).

1. POSITIVE, NEGATIVE OR ZERO CORRELATION:

When the increase in one variable (X) is followed by a corresponding increase in the other variable (Y); the correlation is said to be positive correlation. The positive correlations range from 0 to +1; the upper limit i.e. +1 is the perfect positive coefficient of correlation.

The perfect positive correlation specifies that, for every unit increase in one variable, there is proportional increase in the other. For example “Heat” and “Temperature” have a perfect positive correlation. If, on the other hand, the increase in one variable (X) results in a corresponding decrease in the other variable (Y), the correlation is said to be negative correlation.

The negative correlation ranges from 0 to – 1; the lower limit giving the perfect negative correlation. The perfect negative correlation indicates that for every unit increase in one variable, there is proportional unit decrease in the other.

Zero correlation means no relationship between the two variables X and Y; i.e. the change in one variable (X) is not associated with the change in the other variable (Y). For example, body weight and intelligence, shoe size and monthly salary; etc. The zero correlation is the mid-point of the range – 1 to + 1.

2. LINEAR OR CURVILINEAR CORRELATION:

Linear correlation is the ratio of change between the two variables either in the same direction or opposite direction and the graphical representation of the one variable with respect to other variable is straight line.

Consider another situation. First, with increase of one variable, the second variable increases proportionately upto some point; after that with an increase in the first variable the second variable starts decreasing.

METHODS OF COMPUTING COEFFICIENT OF CORRELATION:

1. KARL PEARSON'S CO-EFFICIENT OF CORRELATION:

The Karl Pearson's product-moment correlation coefficient (or simply, the Pearson's correlation coefficient) is a measure of the strength of a linear association between two variables and is denoted by *r* or r*xy*(x and y being the two variables involved). This method of correlation attempts to draw a line of best fit through the data of two variables, and the value of the Pearson correlation coefficient, *r*, indicates how far away all these data points are to this line of best fit.

The value of *r* always lies between +1 and -1. Depending on its exact value, we see the following degrees of association between the variables.

A value greater than 0 indicates a positive association i.e. as the value of one variable increases, so does the value of the other variable. A value less than 0 indicates a negative association i.e. as the value of one variable increases, the value of the other variable decreases.

KARL PEARSON CORRELATION COEFFICIENT FORMULA

The coefficient of correlation rxy between two variables x and y, for the bivariate dataset (xi,yi) where i = 1,2,3…..n; is given by

$$\rho_{X,Y} = \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

Where,

ρXY = Population correlation coefficient between X and Y

μX = Mean of the variable X

μY = Mean of the variable Y

σX = Standard deviation of X

σY = Standard deviation of Y

E = Expected value operator

Cov = Covriance

The above formulas can also be written as:

$$\rho_{X,Y} = \frac{\mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)}{\sqrt{\mathrm{E}(X^2) - \mathrm{E}(X)^2} \cdot \sqrt{\mathrm{E}(Y^2) - \mathrm{E}(Y)^2}}$$

The sample correlation coefficient formula is:

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\ \sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

The above are used to find the correlation coefficient for the given data. Based on the value obtained through these formulas, we can determine how much strong is the association between given two variables

Exp.

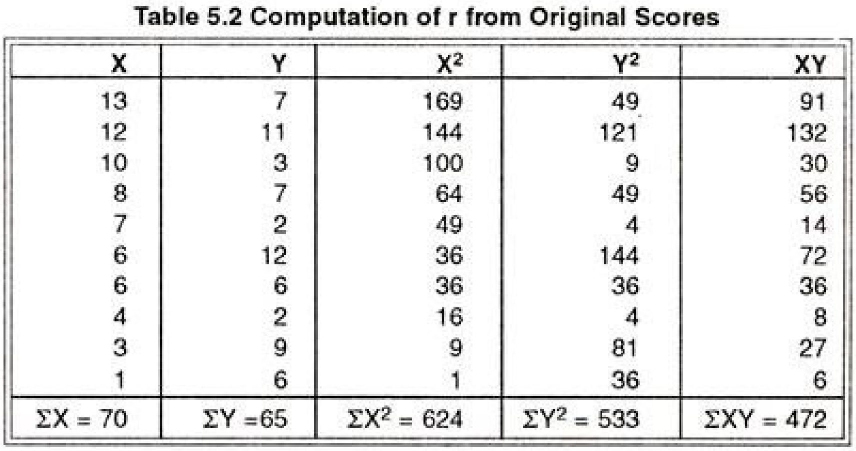
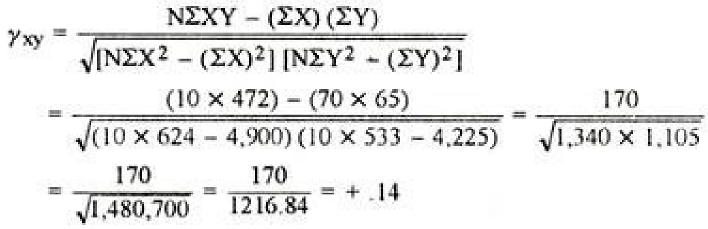


Table 5.2 Computation of r from Original Scores

X	Y	X^2	Y^2	XY
13	7	169	49	91
12	11	144	121	132
10	3	100	9	30
8	7	64	49	56
7	2	49	4	14
6	12	36	144	72
6	6	36	36	36
4	2	16	4	8
3	9	9	81	27
1	6	1	36	6
$\Sigma X = 70$	$\Sigma Y = 65$	$\Sigma X^2 = 624$	$\Sigma Y^2 = 533$	$\Sigma XY = 472$



$$\gamma_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2]\,[N\Sigma Y^2 - (\Sigma Y)^2]}}$$

$$= \frac{(10 \times 472) - (70 \times 65)}{\sqrt{(10 \times 624 - 4{,}900)(10 \times 533 - 4{,}225)}} = \frac{170}{\sqrt{1{,}340 \times 1{,}105}}$$

$$= \frac{170}{\sqrt{1{,}480{,}700}} = \frac{170}{1216.84} = +\,.14$$

2. SPEARMAN’S RANK ORDER CO-EFFICIENT OF CORRELATION

RANK CORRELATION

Sometimes there doesn’t exist a marked linear relationship between two random variables but a monotonic relation (if one increases, the other also increases or instead, decreases) is clearly noticed. Pearson’s Correlation Coefficient evaluation, in this case, would give us the strength and direction of the linear association only between the variables of interest. Herein comes the advantage of the Spearman Rank Correlation methods, which will instead, give us the strength and direction of the monotonic relation between the connected variables. This can be a good starting point for further evaluation.

THE SPEARMAN RANK-ORDER CORRELATION COEFFICIENT

The Spearman's Correlation Coefficient, represented by ρ or by r_R, is a nonparametric measure of the strength and direction of the association that exists between two ranked variables. It determines the degree to which a relationship is monotonic, i.e., whether there is a monotonic component of the association between two continuous or ordered variables.

Monotonicity is "less restrictive" than that of a linear relationship. Although monotonicity is not actually a requirement of Spearman's correlation, it will not be meaningful to pursue Spearman's correlation to determine the strength and direction of a monotonic relationship if we already know the relationship between the two variables is not monotonic.

SPEARMAN RANKING OF THE DATA

We must rank the data under consideration before proceeding with the Spearman's Rank Correlation evaluation. This is necessary because we need to compare whether on increasing one variable, the other follows a monotonic relation (increases or decreases regularly) with respect to it or not.

Thus, at every level, we need to compare the values of the two variables. The method of ranking assigns such 'levels' to each value in the dataset so that we can easily compare it.

- Assign number 1 to n (the number of data points) corresponding to the variable values in the order highest to lowest.
- In the case of two or more values being identical, assign to them the arithmetic mean of the ranks that they would have otherwise occupied.

The Formula for Spearman Rank Correlation

$$\rho = 1 - 6\frac{\sum_i d_i^2}{n\left(n^2 - 1\right)}$$

where *n* is the number of data points of the two variables and d_i is the difference in the ranks of the i^{th} element of each random variable considered. The Spearman correlation coefficient, ρ, can take values from +1 to -1.

- A ρ of +1 indicates a perfect association of ranks
- A ρ of zero indicates no association between ranks and
- ρ of -1 indicates a perfect negative association of ranks.
  The closer ρ is to zero, the weaker the association between the ranks.

Exp.

The following table provides data about the percentage of students who have free university meals and their CGPA scores. Calculate the Spearman's Rank Correlation between the two and interpret the result.

x:	14.4	7.2	27.5	33.8	38	15.9	4.9
y:	54	64	44	32	37	68	62

Sol.

d_X = Rank s_X	d_Y = Rank s_Y	d = (d_X – d_Y)	d^2
3	4	-1	1
2	6	-4	16
5	3	2	4

6	1	5	25
7	2	5	25
4	7	-3	9
1	5	-4	16
			$\Sigma d^2 = 96$

$$\rho = 1 - 6\frac{\sum_i d_i^2}{n\left(n^2 - 1\right)}$$

$$\rho = 1 - 6\frac{96}{7\left(7^2 - 1\right)}$$

$$\rho = -0.714$$

Therefore there is a strong negative coefficient.

REGRESSION ANALYSIS

Regression analysis refers to assessing the relationship between the outcome variable and one or more variables. The outcome variable is known as the dependent or response variable and the risk elements, and cofounders are known as predictors or independent variables. The dependent variable is shown by “y” and independent variables are shown by “x” in regression analysis.

The sample of a correlation coefficient is estimated in the correlation analysis. It ranges between -1 and +1, denoted by r and quantifies the strength and direction of the linear association among two variables. The correlation among two variables can either be positive, i.e. a higher level of one variable is related to a higher level of another or negative, i.e. a higher level of one variable is related to a lower level of the other.

The sign of the coefficient of correlation shows the direction of the association. The magnitude of the coefficient shows the strength of the association.

For example, a correlation of r = 0.8 indicates a positive and strong association among two variables, while a correlation of r = -0.3 shows a negative and weak association. A correlation near to zero shows the non-existence of linear association among two continuous variables.

LINEAR REGRESSION

Linear regression is a linear approach to modelling the relationship between the scalar components and one or more independent variables. If the regression has one independent variable, then it is known as a simple linear regression. If it has more than one independent variables, then it is known as multiple linear regression. Linear regression only focuses on the conditional probability distribution of the given values rather than the joint probability distribution. In general, all the real world regressions models involve multiple predictors. So, the term linear regression often describes multivariate linear regression.

DIFFERENCES BETWEEN CORRELATION AND REGRESSION:

- Correlation shows the quantity of the degree to which two variables are associated. It does not fix a line through the data points. You compute a correlation that shows how much one variable changes when the other remains constant. When r is 0.0, the relationship does not exist. When r is positive, one variable goes high as the other goes up. When r is negative, one variable goes high as the other goes down.
- Linear regression finds the best line that predicts y from x, but Correlation does not fit a line.
- Correlation is used when you measure both variables, while linear regression is mostly applied when x is a variable that is manipulated.

CHAPTER 6. PROBABILITY AND ITS DISTRIBUTIONS

PROBABILITY

Probability is a measure of the likelihood of an event to occur. Many events cannot be predicted with total certainty. We can predict only the chance of an event to occur i.e. how likely they are to happen, using it. Probability can range in from 0 to 1, where 0 means the event to be an impossible one and 1 indicates a certain event. Probability for Class 10 is an important topic for the students which explains all the basic concepts of this topic. The probability of all the events in a sample space adds up to 1.

For example, when we toss a coin, either we get Head OR Tail, only two possible outcomes are possible (H, T). But if we toss two coins in the air, there could be three possibilities of events to occur, such as both the coins show heads or both shows tails or one shows heads and one tail, i.e.(H, H), (H, T),(T, T).

FORMULA

The probability formula is defined as the possibility of an event to happen is equal to the ratio of the number of favourable outcomes and the total number of outcomes.

Probability of event to happen P(E) = Number of favourable outcomes/Total Number of outcomes

Exp.

There are 6 pillows in a bed, 3 are red, 2 are yellow and 1 is blue. What is the probability of picking a yellow pillow?

Ans.

The probability is equal to the number of yellow pillows in the bed divided by the total number of pillows, i.e. 2/6 = 1/3.

PROBABILITY OF AN EVENT

Assume an event E can occur in r ways out of a sum of n probable or possible equally likely ways. Then the probability of happening of the event or its success is expressed as;

P(E) = r/n

The probability that the event will not occur or known as its failure is expressed as:

P(E') = (n-r)/n = 1-(r/n)

E' represents that the event will not occur.

Therefore, now we can say;

P(E) + P(E') = 1

This means that the total of all the probabilities in any random test or experiment is equal to 1.

EQUALLY LIKELY EVENTS

When the events have the same theoretical probability of happening, then they are called equally likely events. The results of a sample space are called equally likely if all of them have the same probability of occurring. For example, if you throw a die, then the probability of getting 1 is 1/6. Similarly, the probability of getting all the numbers from 2,3,4,5 and 6, one at a time is 1/6. Hence, the following are some examples of equally likely events when throwing a die:

- Getting 3 and 5 on throwing a die
- Getting an even number and an odd number on a die
- Getting 1, 2 or 3 on rolling a die

are equally likely events, since the probabilities of each event are equal.

COMPLEMENTARY EVENTS

The possibility that there will be only two outcomes which states that an event will occur or not. Like a person will come or not come to your house, getting a job or not getting a job, etc. are examples of complementary events. Basically, the complement of an event occurring in the exact opposite that the probability of it is not occurring. Some more examples are:

- It will rain or not rain today
- The student will pass the exam or not pass.
- You win the lottery or you don't.

PROBABILITY TERMS AND DEFINITION

Term	Definition	Example
Sample Space	The set of all the possible outcomes to occur in any trial	1. Tossing a coin, Sample Space (S) = {H,T} 2. Rolling a die, Sample Space (S) = {1,2,3,4,5,6 }
Sample Point	It is one of the possible results	In a deck of Cards: • 4 of hearts is a sample point. • the queen of clubs is a sample point.
Experiment or Trial	A series of actions where the outcomes are always uncertain.	The tossing of a coin, Selecting a card from a deck of cards, throwing a dice.
Event	It is a single outcome of an	Getting a Heads while tossing a

Term	Definition	Example
	experiment.	coin is an event.
Outcome	Possible result of a trial/experiment	T (tail) is a possible outcome when a coin is tossed.
Complimentary event	The non-happening events. The complement of an event A is the event, not A (or A')	Standard 52-card deck, A = Draw a heart, then A' = Don't draw a heart
Impossible Event	The event cannot happen	In tossing a coin, impossible to get both head and tail at the same time

PROBABILITY DENSITY FUNCTION

The Probability Density Function (PDF) is the probability function which is represented for the density of a continuous random variable lying between a certain range of values. Probability Density Function explains the normal distribution and how mean and deviation exists. The standard normal distribution is used to create a database or statistics, which are often used in science to represent the real-valued variables, whose distribution are not known.

Exp. Draw a random card from a pack of cards. What is the probability that the card drawn is a face card?

Sol.

A standard deck has 52 cards.

Total number of outcomes = 52

Number of favourable events = 4 x 3 = 12 (considered Jack, Queen and King only)

Probability, P = Number of Favourable Outcomes/Total Number of Outcomes = 12/52= 3/13.

Exp. Find the probability of 'getting 3 on rolling a die'.

Sol.

Sample Space = {1, 2, 3, 4, 5, 6}

Number of favourable event = 1

i.e. {3}

Total number of outcomes = 6

Thus, Probability, P = 1/6

PROBABILITY DISTRIBUTION

Probability distribution yields the possible outcomes for any random event. It is also defined based on the underlying sample space as a set of possible outcomes of any random experiment. These settings could be a set of real numbers or a set of vectors or set of any entities. It is a part of probability and statistics.

Random experiments are defined as the result of an experiment, whose outcome cannot be predicted. Suppose, if we toss a coin, we cannot predict, what outcome it will appear either it will come as Head or as Tail. The possible result of a random experiment is called an outcome. And the set of outcomes is called a sample point. With the help of these experiments or events, we can always create a probability pattern table in terms of variable and probabilities.

PROBABILITY DISTRIBUTION OF RANDOM VARIABLES

A random variable has a probability distribution, which defines the probability of its unknown values. Random variables can be discrete (not constant) or continuous or both. That means it takes any of a designated finite or countable list of values, provided with a probability mass function feature of the random variable’s probability distribution or can take any numerical value in an interval or set of intervals. Through a probability density function that is representative of the random variable’s probability distribution or it can be a combination of both discrete and continuous.

Two random variables with equal probability distribution can yet vary with respect to their relationships with other random variables or whether they are independent of these. The recognition of a random variable, which means, the outcomes of randomly choosing values as per the variable’s probability distribution function, are called **random variates.**

TYPES OF PROBABILITY DISTRIBUTION

There are two types of probability distribution which are used for different purposes and various types of the data generation process.

1. Discrete Probability Distribution (Binomial and Poisson Distribution)
2. Cumulative Probability Distribution (Normal Distribution)

Let us discuss now both the types along with its definition, formula and examples.

1. DISCRETE PROBABILITY DISTRIBUTION

A distribution is called a discrete probability distribution, where the set of outcomes are discrete in nature.

For example, if a dice is rolled, then all the possible outcomes are discrete and give a mass of outcomes. This is also known as probability mass functions.

So, the outcomes of binomial distribution consist of n repeated trials and the outcome may or may not occur. The formula for the binomial distribution is;

$$\mathrm{P(x)} = \frac{n!}{r!(n-r)!} \cdot p^r (1-p)^{n-r}$$

$$\mathrm{P(x)} = \mathrm{C\ (n, r)}.p^r (1-p)^{n-.r}$$

where,

- n = Total number of events

- r = Total number of successful events.
- p = Success on a single trial probability.
- ${}^{n}C_{r}$ = [n!/r!(n−r)]!
- 1 – p = Failure Probability

BINOMIAL DISTRIBUTION EXAMPLES

As we already know, binomial distribution gives the possibility of a different set of outcomes. In the real-life, the concept is used for:

- To find the number of used and unused materials while manufacturing a product.
- To take a survey of positive and negative feedback from the people for anything.
- To check if a particular channel is watched by how many viewers by calculating the survey of YES/NO.
- The number of men and women working in a company.
- To count the votes for a candidate in an election and many more.

Exp. If a coin is tossed 5 times, find the probability of:

(a) Exactly 2 heads

(b) At least 4 heads.

Sol.

(a) The repeated tossing of the coin is an example of a Bernoulli trial. According to the problem:

Number of trials: n=5

Probability of head: p= 1/2 and hence the probability of tail, q =1/2

For exactly two heads:

x=2

$P(x=2) = {}^{5}C2\ p^{2}\ q^{5-2} = 5!\ /\ 2!\ 3! \times (½)^{2} \times (½)^{3}$

P(x=2) = 5/16

(b) For at least four heads,

$x \geq 4$, $P(x \geq 4) = P(x = 4) + P(x=5)$

Hence,

$P(x = 4) = {}^5C4\ p^4\ q^{5-4} = 5!/4!\ 1! \times (½)^4 \times (½)^1 = 5/32$

$P(x = 5) = {}^5C5\ p^5\ q^{5-5} = (½)^5 = 1/32$

Therefore,

$P(x \geq 4) = 5/32 + 1/32 = 6/32 = 3/16$

POISSON PROBABILITY DISTRIBUTION

The Poisson probability distribution is a discrete probability distribution that represents the probability of a given number of events happening in a fixed time or space if these cases occur with a known steady rate and individually of the time since the last event. It was titled after French mathematician Siméon Denis Poisson. The Poisson distribution can also be practised for the number of events happening in other particularised intervals such as distance, area or volume. Some of the real-life examples are:

- A number of patients arriving at a clinic between 10 to 11 AM.
- The number of emails received by a manager between the office hours.
- The number of apples sold by a shopkeeper in the time period of 12 pm to 4 pm daily.

PROBABILITY DISTRIBUTION FUNCTION

A function which is used to define the distribution of a probability is called a Probability distribution function. Depending upon the types, we can define these functions. Also, these functions are used in terms of probability density functions for any given random variable.

In the case of **Normal distribution**, the function of a real-valued random variable X is the function given by;

$\mathbf{F_X(x) = P(X \leq x)}$

Where P shows the probability that the random variable X occurs on less than or equal to the value of x.

2. CUMULATIVE PROBABILITY DISTRIBUTION

The cumulative probability distribution is also known as a continuous probability distribution. In this distribution, the set of possible outcomes can take on values on a continuous range.

For example, a set of real numbers, is a continuous or normal distribution, as it gives all the possible outcomes of real numbers. Similarly, set of complex numbers, set of prime numbers, set of whole numbers etc. are the examples of Normal Probability distribution. Also, in real-life scenarios, the temperature of the day is an example of continuous probability. Based on these outcomes we can create a distribution table. A probability density function describes it.

The formula for the normal distribution is;

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

where,

- μ = Mean Value
- σ = Standard Distribution of probability.
- If mean(μ) = 0 and standard deviation(σ) = 1, then this distribution is known to be normal distribution.
- x = Normal random variable

NORMAL DISTRIBUTION EXAMPLES

Since the normal distribution statistics estimates many natural events so well, it has evolved into a standard of recommendation for many probability queries. Some of the examples are:

- Height of the Population of the world
- Rolling a dice (once or multiple times)
- To judge Intelligent Quotient Level of children in this competitive world
- Tossing a coin

CHAPTER 7. TIME SERIES ANALYSIS

A time series is a series of data points indexed (or listed or graphed) in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus it is a sequence of discrete-time data. Examples of time series are heights of ocean tides, counts of sunspots, and the daily closing value of the Dow Jones Industrial Average.

Time series are very frequently plotted via line charts. Time series are used in statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering, and largely in any domain of applied science and engineering which involves temporal measurements.

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. While regression analysis is often employed in such a way as to test theories that the current values of one or more independent time series affect the current value of another time series, this type of analysis of time series is not called "time series analysis", which focuses on comparing values of a single time series or multiple dependent time series at different points in time. Interrupted time series analysis is the analysis of interventions on a single time series.

Time series data have a natural temporal ordering. This makes time series analysis distinct from cross-sectional studies, in which there is no natural ordering of the observations (e.g. explaining people's wages by reference to their respective education levels, where the individuals' data could be entered in any order). Time series analysis is also distinct from spatial data analysis where the observations typically relate to geographical locations (e.g. accounting for house prices by the location as well as the intrinsic characteristics of the houses). A stochastic model for a time series will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values (see time reversibility.)

Time series analysis can be applied to real-valued, continuous data, discrete numeric data, or discrete symbolic data

Time series data is data that is collected at different points in time. This is opposed to cross-sectional data which observes individuals, companies, etc. at a single point in time. Because data points in time series are collected at adjacent time periods there is potential for correlation between observations. This is one of the features that distinguishes time series data from cross-sectional data.

Time series data can be found in economics, social sciences, finance, epidemiology, and the physical sciences.

Time Series Data		
Field	**Example topics**	**Example dataset**
Economics	Gross Domestic Product (GDP), Consumer Price Index (CPI), S&P 500 Index, and unemployment rates	U.S. GDP from the Federal Reserve Economic Data
Social sciences	Birth rates, population, migration data, political indicators	Population without citizenship from Eurostat
Epidemiology	Disease rates, mortality rates, mosquito populations	U.S. Cancer Incidence rates from the Center for Disease Control
Medicine	Blood pressure tracking, weight tracking, cholesterol measurements, heart rate monitoring	MRI scanning and behavioral test dataset
Physical sciences	Global temperatures, monthly sunspot observations, pollution levels.	Global air pollution from the Our World in Data

The statistical characteristics of time series data often violate the assumptions of conventional statistical methods. Because of this, analyzing time series data requires a unique set of tools and methods, collectively known as time series analysis.

This article covers the fundamental concepts of time series analysis and should give you a foundation for working with time series data.

TIME SERIES DATA

Time series data is a collection of quantities that are assembled over even intervals in time and ordered chronologically. The time interval at which data is collection is generally referred to as the time series frequency.

TIME SERIES GRAPH

A time series graph plots observed values on the y-axis against an increment of time on the x-axis. These graphs visually highlight the behavior and patterns of the data and can lay the foundation for building a reliable model.

More specifically, visualizing time series data provides a preliminary tool for detecting if data:

- Is mean-reverting or has explosive behavior;
- Has a time trend;
- Exhibits seasonality;
- Demonstrates structural breaks.

This, in turn, can help guide the testing, diagnostics, and estimation methods used during time series modeling and analysis.

STATISTICAL AVERAGES

Let’s start simple! Statistical averages. It’s an easy-to-understand concept, and very commonly used. The point of using averages is to get a *central value* of a dataset. Of course, there is more than one way to decide which value is the most central… That’s why we have more than one average type.

The three most common statistical averages are:

1. Mean
2. Median
3. Mode

MOVING AVERAGE (MA)

In statistics, a moving average is a calculation used to analyze data points by creating a series of averages of different subsets of the full data set. In finance, a moving average (MA) is a stock indicator that is commonly used in technical analysis. The reason for calculating the moving average of a stock is to help smooth out the price data by creating a constantly updated average price.

By calculating the moving average, the impacts of random, short-term fluctuations on the price of a stock over a specified time-frame are mitigated.

UNDERSTANDING MOVING AVERAGE (MA)

Moving average is a simple, technical analysis tool. Moving averages are usually calculated to identify the trend direction of a stock or to determine its support and resistance levels. It is a trend-following—or lagging—indicator because it is based on past prices.

The longer the time period for the moving average, the greater the lag. So, a 200-day moving average will have a much greater degree of lag than a 20-day MA because it contains prices for the past 200 days. The 50-day and 200-day moving average figures for stocks are widely followed by investors and traders and are considered to be important trading signals.

Moving averages are a totally customizable indicator, which means that an investor can freely choose whatever time frame they want when calculating an average. The most common time periods used in moving averages are 15, 20, 30, 50, 100, and 200 days. The shorter the time span used to create the average, the more sensitive it will be to price changes. The longer the time span, the less sensitive the average will be.

Investors may choose different time periods of varying lengths to calculate moving averages based on their trading objectives. Shorter moving averages are typically used for short-term trading, while longer-term moving averages are more suited for long-term investors.

There is no correct time frame to use when setting up your moving averages. The best way to figure out which one works best for you is to experiment with a number of different time periods until you find one that fits your strategy.

Predicting trends in the stock market is no simple process. While it is impossible to predict the future movement of a specific stock, using technical analysis and research can help you make better predictions.

A rising moving average indicates that the security is in an uptrend, while a declining moving average indicates that it is in a downtrend. Similarly, upward momentum is confirmed with a bullish crossover, which occurs when a short-term moving average crosses above a longer-term moving average. Conversely, downward momentum is confirmed with a bearish crossover, which occurs when a short-term moving average crosses below a longer-term moving average.

While calculating moving averages are useful in their own right, the calculation can also form the basis for other technical analysis indicators, such as the moving average convergence divergence (MACD).

The moving average convergence divergence (MACD) is used by traders to monitor the relationship between two moving averages. It is generally calculated by subtracting a 26-day exponential moving average from a 12-day exponential moving average.

When the MACD is positive, the short-term average is located above the long-term average. This an indication of upward momentum. When the short-term average is below the long-term average, this is a sign that the momentum is downward. Many traders will also watch for a move above or below the zero line. A move above zero is a signal to buy, while a cross below zero is a signal to sell.

TYPES OF MOVING AVERAGE

There are two types of Moving Averages

1. Simple Moving Average
2. Exponential Moving Average

SIMPLE MOVING AVERAGE

The Simple Moving Average is a form of the moving average which is calculated by adding the closing prices of stock during specific time periods and dividing the sum by the number of time periods in the calculation average; or in simple terms, the average price over a particular time period.

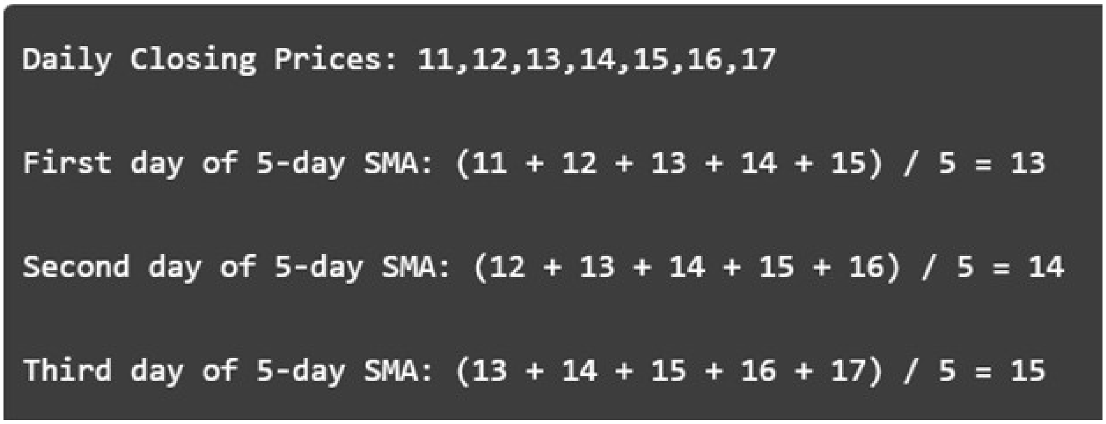
THE FORMULA OF SIMPLE MOVING AVERAGE

$$SMA = \frac{A_1 + A_2 + ... + A_n}{n}$$

In this formula An refers to the closing stock prices at period n, and n is the total number of periods.

Let’s try to understand this with an example.

Consider the daily closing prices in a 5-day moving average.



```
Daily Closing Prices: 11,12,13,14,15,16,17

First day of 5-day SMA: (11 + 12 + 13 + 14 + 15) / 5 = 13

Second day of 5-day SMA: (12 + 13 + 14 + 15 + 16) / 5 = 14

Third day of 5-day SMA: (13 + 14 + 15 + 16 + 17) / 5 = 15
```

Here, the first day of the moving average covers the last 5 days. As the days progress, the previous day's data point is dropped and replaced by a new data point. So the second day of the MA will drop the first data point (11) and add a new one (16). The third day continues by dropping the first data point (12) and adding a new one (17).

Thus, in the example above, prices rise from 11 to 17 in a span of 7 days. You can also notice that the MA rises from 13 to 15 within 3 days. Additionally, the price of each moving average is just below the last one. This is because prices in the previous four days were lower and this causes the moving average to lag.

EXPONENTIAL MOVING AVERAGE

While the Simple moving average is based on past data, the Exponential Moving average stresses more on recent prices. The weighting applied to recent prices depends on the number of periods. This difference between SMA and EMA, makes the latter more efficient in responding to new information. In an EMA, the calculation depends on the EMA calculations of all the previous days. You need at least more than 10 days of data to calculate Exponential moving average.

EXPONENTIAL MOVING AVERAGE FORMULA

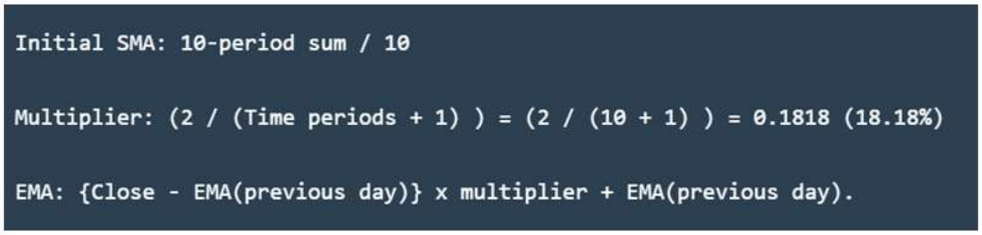
$$EMA_{Today} = (Value_{Today} * \left(\frac{Smoothing}{1 + Days}\right)) + EMA_{Yesterday} * (1 - \left(\frac{Smoothing}{1+Days}\right))$$

There are three steps to calculating an exponential moving average.

1. compute the simple moving average
2. calculate the multiplier for weighting the EMA. this is also called smoothening, with a formula of [2 ÷ (selected time period + 1)].
3. calculate the exponential moving average for each day between the first EMA value and current value, using the price, the multiplier, and the previous period's EMA value.

Let’s understand EMA with an example.

Consider a 10 days time period for calculating Exponential Moving average.



```
Initial SMA: 10-period sum / 10

Multiplier: (2 / (Time periods + 1) ) = (2 / (10 + 1) ) = 0.1818 (18.18%)

EMA: {Close - EMA(previous day)} x multiplier + EMA(previous day).
```

An exponential moving average for a 10-day time period applies an 18.18% weighting to the latest price. The weighting after applying to the most recent price is (2/(10+1) = 18.18%).

REFERENCES:

1. Quantitative Techniques for Management, Excel Books India, 2009
2. Quantitative Techniques: Theory and Problems, Pearson Education India, 2006
3. Statistics for Management, S. Chand & Company, New Delhi, 2003
4. S. P. Gupta, Statistical Methods, Sultan Chand & Sons, 1976.
5. B Antonisamy, Prasanna S. Premkumar, Solomon Christopher, Principles and Practice of Biostatistics, Elsevier, 2017.
6. https://www.bmj.com/
7. https://www.yourarticlelibrary.com/
8. https://byjus.com/maths/probability/
9. https://www.thoughtco.com/
10. http://www.brainkart.com/
11. https://www.academia.edu/36761015/Application_of_Quantitative_Methods_Techniques_in_Business_and_Economics
12. http://www.google.com/

Contact Us:
University Campus Address:

Jayoti Vidyapeeth Women's University

Vadaant Gyan Valley, Village-Jharna, Mahala Jobner Link Road,
Jaipur Ajmer Express Way, NH-8, Jaipur- 303122, Rajasthan (INDIA)
(Only Speed Post is Received at University Campus Address, No. any Courier Facility is available at Campus Address)

Pages	: 64
Book Price	: ₹ 150/-

JPD
Women University Press
Jayoti Publication Desk

Year & Month of Publication-3/3/2021